



**HAL**  
open science

## On Elliptical Possibility Distributions

Charles Lesniewska-Choquet, Gilles Mauris, Abdourrahmane Atto, Grégoire  
Mercier

► **To cite this version:**

Charles Lesniewska-Choquet, Gilles Mauris, Abdourrahmane Atto, Grégoire Mercier. On Elliptical Possibility Distributions. IEEE Transactions on Fuzzy Systems, 2020, 28 (8), pp.1631-1639. 10.1109/TFUZZ.2019.2920803 . hal-02149345

**HAL Id: hal-02149345**

**<https://hal.science/hal-02149345>**

Submitted on 6 Jun 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On Elliptical Possibility Distributions

Charles LESNIEWSKA-CHOQUET<sup>1,\*</sup>, Gilles MAURIS<sup>1</sup>, Abdourrahmane M. ATTO<sup>1</sup>, Grégoire MERCIER<sup>2</sup>

**Abstract**—This paper aims to propose two main contributions in the field of multivariate data analysis through the possibility theory. The first proposition is the definition of a generalized family of multivariate elliptical possibility distributions. These distributions have been derived from a consistent probability-possibility transformation over the family of so-called elliptical probability distributions. The second contribution proposed by the paper is the definition of two divergence measures between possibilistic distributions. We prove that a symmetric version of the Kullback-Leibler divergence guarantees all divergence properties when related to the space of possibility distributions. We further derive analytical expressions of the latter divergence and of the Hellinger divergence for certain possibility distributions pertaining to the elliptical family proposed, especially the normal multivariate possibility divergence in two dimensions. Finally, this paper provides an illustration of the developed possibilistic tools in an application of bi-band change detection between optical satellite images.

**Keywords:** Multivariate possibility distributions; Joint possibility distributions; Elliptical probability distributions; Kullback-Leibler divergence; Hellinger divergence; Change detection.

## I. INTRODUCTION

Multivariate data analysis has long been an outstanding issue where the probability theory plays a major role. Recently, different attempts have been made in the literature in terms of possibility theory in order to suggest an alternative representation of multivariate data that can be convenient in cases where probabilistic features are weak or difficult to obtain.

The first way to deal with a multivariate issue is to directly define a multidimensional possibility distribution as proposed among others, by Tanaka in [1], who considers multivariate exponential distributions.

The second way consists in combining marginal possibility distributions by means of t-norm [2], in order to obtain joint possibility distribution [3][4][5][6].

The third way followed in this paper is to transform a multivariate probability distribution by extending the univariate probability-possibility transformation [7][8] to the n-dimensional case. In the first part of this paper, we will consider in the same framework:

- the probability-possibility transformation framework proposed in one-dimension
- the multivariate elliptical probability distributions [9]

- and the Mahalanobis distance relating multivariate data to their monivariate closeness measures

in order to derive a family of multivariate possibility distributions. These possibility distributions will be called elliptical, for terminology analogy with respect to their probabilistic counter parts. Thus our work will help to design a possibility distribution from any elliptical probability distribution thanks to a general analytical form and regardless of the dimension.

In the second part of this paper, we consider possibilistic divergence definition [10], called  $\Pi$ -divergence hereafter. Two divergences are studied: the Kullback-Leibler and the Hellinger divergence. We show that the probabilistic Kullback-Leibler divergence [11] does not satisfy positivity constraints when transposed to the possibility domain, and that its symmetric version does not suffer from this lack. On the other hand, due to its closeness to the Euclidean norm, the Hellinger divergence preserves its divergence properties when convert to the possibility domain. Then we derive some analytical closed forms from the previously mentioned  $\Pi$ -divergences applied to elliptical possibility distributions that can be used for dissemblance assessment in image processing applications. We also prove the relevancy of the joint possibility modeling and divergence evaluation in the context of a bi-band change detection application between optical satellite images.

This paper is organized as follows: Section 2 presents the proposed n-dimensional probability-possibility transformation for the derivation of elliptical possibility distributions. Some properties of this transformation are given as well as an illustration of different possibility distributions obtained in case  $n = 2$ . Section 3 is divided into two main steps. First, two possibilistic divergence measures are proposed and their analytical forms calculated for the joint possibility distribution obtained from the normal bivariate distribution are given. Secondly, an application using the possibilistic tools developed in this paper is presented in the form of an algorithm of bi-band change detection between optical satellite images. Section 4 provides an overall conclusion to this paper.

## II. ELLIPTICAL CONTINUOUS POSSIBILITY DISTRIBUTION

After recalling some definitions of the possibility theory and of the probability-possibility transformation in the continuous case for the unidimensional probability model, this section will present a new family of n-dimensional elliptical possibility distributions. At the end, a last subsection will propose some bivariate elliptical possibility distributions obtained from well-known probability distributions.

<sup>1</sup> Univ. Savoie Mont Blanc, LISTIC, F-74000 Annecy, France

<sup>2</sup> eXo maKina, Digital Technologies, Paris, France

\* charles.lesniewska-choquet@univ-smb.fr

### A. Unidimensional Probability-Possibility transformation

In the following, let us consider a continuous random variable  $X$  on the set of reals and its associated probability density function  $p$  with mode  $x^*$ ,  $F$  being its corresponding cumulative distribution function.

**Definition 1.** (Definition 3.1 in [8]) *The possibility distribution induced by a continuous strictly unimodal probability density  $p$  around  $x^*$  is the possibility (denoted  $\pi^*$ ) whose  $\alpha$ -cuts are the confidence interval  $I_\alpha^*$  of level  $P(I_\alpha^*) = \alpha$  around the mode  $x^*$  computed from  $p$ .*

$\pi^*$  can be defined as follows:

$$\pi^*(x) = 1 - P(\{y, p(y) \geq p(x)\}) \quad (1)$$

The possibility distribution  $\pi^*$  is continuous and encodes the whole set of confidence intervals and  $\pi^*(x^*) = 1$ .

**Theorem 1.** (Theorem 3.1 in [8]) *For any probability density  $p$ , the possibility distribution  $\pi^*$  in definition 1 is consistent with  $p$ , that is:  $\forall A$  measurable,  $\Pi^*(A) \geq P(A)$ ,  $\Pi^*$  and  $P$  being the possibility and probability measures associated to  $\pi^*$  and  $p$ .*

**Lemma 1.** (Lemma 3.1 in [8]) *For any continuous probability density  $p$  having a finite number of modes, any minimal length measurable subset  $I$  of the real line such that  $P(I) = \alpha \in (0, 1]$ , is of the form  $\{x, p(x) \geq \beta\}$  for some  $\beta \in [0, p_{max}]$  where  $p_{max} = \sup_x p(x)$ . It thus contains the modal value(s) of  $p$ .*

From Definition 1 and Theorem 1, it is stated that the maximally specific possibility distribution  $\pi^*$  obtained from a given probability distribution although remaining consistent is built from the set of confidence intervals around a nominal value. From Lemma 1, we ensure that we preserve the maximum amount of information during the transformation from  $p$  to  $\pi$  when the nominal value around which is built the possibility distribution is chosen to be the mode  $x^m$ . Moreover, in the one-dimensional case, it has been proved in [8] that from (1), the expression of the possibility distribution  $\pi^*$  corresponding to a unimodal symmetric PDF of mode  $x^*$  strictly increasing on the left and decreasing on the right can be written as [12],[13]:

$$\begin{aligned} \text{for } -\infty < x \leq x^*, \quad \pi^*(x) &= 2F(x) \\ \text{for } x^* \leq x < \infty, \quad \pi^*(x) &= 2(1 - F(x)) \end{aligned}$$

As an example we can consider the case of a Normal probability distribution with mean  $\mu$  and variance  $\sigma^2$ , the following possibility distribution is then:

$$\begin{aligned} \text{for } -\infty < x \leq \mu, \quad \pi(x) &= 1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right) \\ \text{for } \mu \leq x < \infty, \quad \pi(x) &= 1 - \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right) \end{aligned}$$

Where  $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$  is the Error function. Since the Error function is negative when  $x < 0$  and positive

when  $x > 0$ , we can then rewrite it as:

$$\pi(x) = 1 - \operatorname{erf}\left(\frac{|x-\mu|}{\sigma\sqrt{2}}\right) \quad (2)$$

As shown by (2), in one dimension, the continuous probability-possibility transformation can be applied to any symmetric unimodal probability distribution that has an analytical form for its PDF.

However, one can note that definition 1 and theorem 1, in which this transformation lies, is not limited to the unidimensional case but only to unimodal probability distributions. Lemma 1 can also be extended to the case of n-dimensional possibility distributions by taking a hyper-volume as a measurable length, which is a Lebesgue measure on  $\mathbb{R}^n$ . Different ways to define the hyper-volume are possible.

Thus, by taking as a starting point (1) from definition 1 and by applying it to the n-dimensional case, in the next section we will propose a new family of elliptical possibility distributions derived from the family of elliptical probability distributions.

### B. Probability-Possibility transformation for Multivariate elliptical probability distributions

For the sake of generality, we consider a family of probability which represents a general framework although presenting good performance in the modelization of correlated information. Thus our attention focuses on the family of elliptical probability distributions which has been widely studied especially in [14], [15] or [16], and has already shown its usefulness. This family includes notably the Normal, Laplace, Student's t, and Cauchy distributions.

Firstly, let's start by giving the considered framework.

Let  $X$  be an n-dimensional random vector elliptically contoured (denoted by  $X \sim EC_n(\Sigma)$ ), whose PDF has the following form

$$p_X(x, \mu, \Sigma) = \alpha_n |\Sigma|^{-\frac{1}{2}} g_n((x - \mu)^T \Sigma^{-1} (x - \mu)) \quad (3)$$

where  $\mu \in \mathbb{R}^n$  is the mean value,  $\Sigma \in \mathbb{R}^{n \times n}$  its positive and symmetric variance-covariance matrix,  $\alpha_n$  is a normalization scalar and  $|\Sigma|$  is the determinant of  $\Sigma$ . The function  $g_n : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is a density generator function of the elliptical distribution in  $\mathbb{R}^n$  with the following property [14]:

$$\int_{\mathbb{R}} g_n(x^T x) dx = 1.$$

Let  $d_{mah}$  be the Mahalanobis distance between two vectors  $x$  and  $y$  according to the weighting matrix  $\Sigma$ . Then  $d_{mah}$  is such as:

$$d_{mah}(x, y) = \|x - y\|_{\Sigma} = \sqrt{(x - y)^T \Sigma^{-1} (x - y)} \quad (4)$$

It is important to note that the equidensity contours of  $p_X$  are ellipsoids centered around the mean  $\mu$  (which is also the mode) and the axes of the ellipsoids are the eigenvectors of  $\Sigma$  [17]. Secondly, let's determine the probability accumulated in a region delimited by an equidensity contour of the PDF  $p$  given in (3).

**Proposition 1.** *Let  $a_x = d_{mah}(x, \mu)$ , for an n-dimensional random vector  $X \in \mathbb{R}^n$  with mean  $\mu \in \mathbb{R}^n$  and covariance*

matrix  $\Sigma \in \mathbb{R}^{n \times n}$ , whose PDF is (3), the probability accumulated in the region

$$A_\Sigma := \{z \in \mathbb{R}^n \text{ such that } d_{mah}(z, \mu) \leq a_x\} \quad (5)$$

is

$$P(A_\Sigma) := P(d_{mah}(z, \mu) \leq a_x) = \alpha_n \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \int_0^{a_x} r^{n-1} g_n(r^2) dr \quad (6)$$

where  $\alpha_n$  is a normalization scalar,  $\Gamma(u) = \int_0^{+\infty} t^{u-1} e^{-t} dt$  is the Gamma function and  $g_n$  is the density generator function.

*Proof.* Let's directly evaluate  $P(A_\Sigma) = \int_{A_\Sigma} p(z) dz$ . Firstly, as  $\Sigma$  is a definite positive matrix, we can proceed to a change of variable in the integral  $P(A_\Sigma)$  by setting  $y = \Sigma^{-\frac{1}{2}}(z - \mu)$  whose determinant of Jacobian is  $|\frac{\partial z}{\partial y}| = |\Sigma|^{\frac{1}{2}}$ . Hence, from (3)

$$\begin{aligned} \int_{A_\Sigma} p(z) dz &= \alpha_n |\Sigma|^{-\frac{1}{2}} \int_{\|y\| \leq a_x} g_n(\|y\|^2) |\Sigma|^{\frac{1}{2}} dy \\ &= \alpha_n \int_{\|y\| \leq a_x} g_n(\|y\|^2) dy \end{aligned} \quad (7)$$

where  $\|\cdot\|$  stands for the Euclidean Norm.

Then, by observing that there is a rotational invariance in the integral in (7) since it's the integral of an ellipsoid, we split the n-dimensional volume element  $dy$  into an area element  $dA$  and a radial element  $dr$ . Thus, by setting  $r = \|y\|$ , we obtain an expression of (7) in spherical coordinates such as

$$P(A_\Sigma) = \alpha_n \int_0^{a_x} \int_{S^{n-1}(r)} g_n(r^2) dA dr,$$

where  $S^{n-1}(r)$  is the  $(n-1)$ -sphere of radius  $r$ .

Knowing that the area of the surface of this  $(n-1)$ -sphere can be written as  $\mathcal{A}_n(r) = r^{n-1} \mathcal{A}_{n-1}(1)$  where  $\mathcal{A}_{n-1}(1) = 2\pi^{\frac{n}{2}}/\Gamma(\frac{n}{2})$  is the area of the unit sphere, we have

$$P(A_\Sigma) = \alpha_n \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \int_0^{a_x} r^{n-1} g_n(r^2) dr \quad (8)$$

□

Finally, from (1) and proposition 1 we are able to obtain the elliptical possibility distribution associated with any elliptical probability distribution:

$$\begin{aligned} \pi_\Sigma(x) &= 1 - P(A_\Sigma) \\ &= 1 - \alpha_n \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \int_0^{a_x} r^{n-1} g_n(r^2) dr \end{aligned} \quad (9)$$

with  $a_x = d_{mah}(x, \mu)$ . The following proposition provides a stochastic dominance property between possibility distributions.

**Proposition 2.** *Let  $X$  be an n-dimensional random vector on the set of reals with mean  $\mu \in \mathbb{R}^n$  and  $\Sigma_X \in \mathbb{R}^{n \times n}$  its semi-definite positive covariance matrix.*

*Let  $Y$  be another n-dimensional random vector on the set of reals with the same mean  $\mu \in \mathbb{R}^n$  and  $\Sigma_Y \in \mathbb{R}^{n \times n}$  its semi-definite positive covariance matrix.*

Lastly, let  $\pi_X$  and  $\pi_Y$  be the two possibility distributions associated respectively to  $X$  and  $Y$ .

If

the matrix  $M = \Sigma_Y - \Sigma_X$  is semi-definite positive.

Then

$$\pi_{\Sigma_X} \leq \pi_{\Sigma_Y}.$$

*Proof.* Let's denote  $\mathcal{C}_1$  the class of all convex centrally symmetric sets in  $\mathbb{R}^n$  and let's  $\Sigma_X$  and  $\Sigma_Y$  be two definite positive matrices. Finally, let's denote  $P_{\Sigma_X}$  and  $P_{\Sigma_Y}$  as two multivariate elliptical probability distributions.

As a consequence of Theorem 5.1 in [18] we have the following statement: Assume that  $X \sim EC_n(\Sigma)$ . If  $C \in \mathcal{C}_1$  and  $C$  is closed, and if  $\Sigma_Y - \Sigma_X$  is a definite positive matrix then  $P_{\Sigma_X}(C) \geq P_{\Sigma_Y}(C)$ .

Hence, we can conclude that

$$\begin{aligned} P_{\Sigma_X}(C) \geq P_{\Sigma_Y}(C) &\Leftrightarrow 1 - P_{\Sigma_X}(C) \leq 1 - P_{\Sigma_Y}(C) \\ &\Leftrightarrow \pi_{\Sigma_X}(x) \leq \pi_{\Sigma_Y}(x) \end{aligned}$$

□

Note that in the case of an uncorrelated random variable, the associated covariance matrix is diagonal definite positive (i.e.  $\Sigma_X = \text{diag}(\sigma_{X,1}, \dots, \sigma_{X,n})$  and  $\Sigma_Y = \text{diag}(\sigma_{Y,1}, \dots, \sigma_{Y,n})$ ). Hence the matrix  $M = \Sigma_Y - \Sigma_X$  is definite positive if and only if the eigenvalues of the diagonal matrix  $M$  are positive or null meaning that

$$\forall i \in \mathbb{N}^*, \sigma_{X,i} \leq \sigma_{Y,i}.$$

Although (9) may seem complicated to handle in its general form, it has to be applied to some well-known elliptical probability models such as the Normal multivariate probability distribution, the Cauchy multivariate distribution or the Student's t-distribution, for example.

### C. Special case of elliptical possibility distributions

In the previous section, we have determined a general form of elliptical possibility distributions from the n-dimensional family of elliptical probability distributions, let's now apply this transformation onto some well-known probability distributions.

In order to illustrate the use of the probability-possibility transformation given by (9), we will firstly consider the case of one of the most famous elliptical probability distributions: the Normal distribution. Firstly, we recall that the Normal multivariate distribution possesses a PDF of the following form:

$$p_N(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}.$$

It corresponds to (3) when  $\alpha_n = \frac{1}{(2\pi)^{n/2}}$  and  $g_n(t) = e^{-\frac{1}{2}t}$ . From (6) the expression of the probability accumulated in the region  $A_\Sigma$  becomes

$$P_N(A) = \frac{2\pi^{n/2}}{(2\pi)^{n/2} \Gamma(n/2)} \int_0^{a_x} r^{n-1} e^{-\frac{r^2}{2}} dr. \quad (10)$$

Moreover, it has been proved (Result 2 in [19]), that (10), which represents the probability accumulated in a region delimited by an equidensity contour of the PDF, can be computed in this way thanks to the substitution  $t = r^2/2$ :

$$P_N(A_\Sigma) := P_N(d_{mah}(z, \mu) \leq a_x) = \frac{\gamma\left(\frac{n}{2}, \frac{a_x^2}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \quad (11)$$

$\gamma$  being the lower incomplete Gamma function defined as

$$\gamma(a, b) = \int_0^b t^{a-1} e^{-t} dt.$$

Finally, the possibility distribution associated to the multivariate Normal distribution takes the following form:

$$\forall x \in \mathbb{R}^n, \quad \pi_N(x) = 1 - \frac{\gamma\left(\frac{n}{2}, \frac{a_x^2}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}. \quad (12)$$

In the case of  $n = 1$ , the Mahalanobis distance is equivalent to  $d_{mah}(x, \mu) = \|x - \mu\|_\Sigma = \frac{|x - \mu|}{\sigma}$  and  $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ . Thus, we find an expression of the possibility distribution exactly similar to (2).

In table I, we give the analytical expressions of joint possibility distributions obtained for different probability distributions for  $n = 2$ .

Fig. 1 presents different possibility distributions as well as

PDF	Associated joint possibility distributions for $n = 2$
Cauchy	$\pi_C(x) = \frac{1}{\sqrt{1 + (x - \mu)^T \Sigma^{-1} (x - \mu)}} \quad (13)$
Student's t	$\pi_S(x) = \left(1 + \frac{(x - \mu)^T \Sigma^{-1} (x - \mu)}{\nu}\right)^{-\frac{\nu}{2}} \quad (14)$
Normal	$\pi_N(x) = e^{-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)} \quad (15)$

TABLE I: Analytic forms of possibility distributions obtained from well-known elliptical probability distributions.

their associated probability distributions. In each case the mean vector is set to  $\mu = (5 \ 20)$ , the variance along the  $x$  and  $y$  components are  $\sigma_x = 8$  and  $\sigma_y = 6$  respectively and the correlation coefficient is equal to 0.8. For the Student's t distribution, the degree of freedom  $\nu$  is equal to 3. Note that (15) is close to the analytical expression found by H.Tanaka [1] although his exponential possibility function holds for any dimension whereas (15) is derived from a consistent probability-possibility transformation and it is specific to the case  $n = 2$  and differs otherwise. Finally, as can be seen, (15) is remarkable as it is a covariance based normalization of the joint probability distribution for  $n = 2$  such as

$$\pi_N(x) = 2\pi|\Sigma|^{\frac{1}{2}} p_N(x). \quad (16)$$

Thus, in the following section, we will focus on the consequences of this normalization based possibility distribution

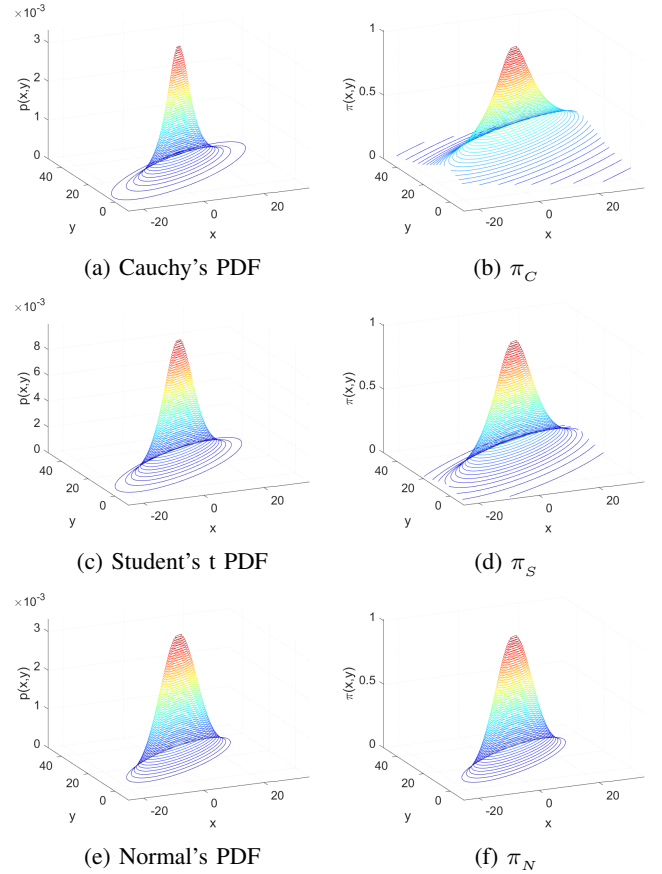


Fig. 1: Examples of PDFs (left column) and their associated possibility distributions (right column).

through a change detection application under a high level of imprecision.

### III. CHANGE DETECTION OVER OPTICAL SATELLITE IMAGES

In this section, we will firstly present the two divergences we derived in order to compare possibility distributions. The possibility distributions are obtained thanks to the previously proposed multivariate probability-possibility transformation applied to the Normal distribution on the special case of  $n = 2$ . Then we will illustrate through a change detection application on real optical satellite images the use of the proposed possibilistic framework.

#### A. Possibilistic Divergences

Let's first introduce the possibilistic divergences ( $\Pi$ -divergence) as a comparison function between two possibility distributions.

The first divergence proposed is inspired (though slightly different) from the extension of the probabilistic Kullback-Leibler divergence by replacing the probability distributions by possibility distributions [20]. This approach has proved its interest in many applications and is of particular value due to the exponential form of the possibility distribution in (15).

As a reminder, for two PDF  $p_1$  and  $p_2$ , the probabilistic Kullback-Leibler divergence is defined as

$$\forall X \in \mathbb{R}^2 \quad \text{KL}(p_1 \| p_2) = \int_{\mathbb{R}} p_1(x) \ln \left( \frac{p_1(x)}{p_2(x)} \right) dx. \quad (17)$$

**Remark 1.** *Let's begin with an example. In the case where  $\pi_2 = \pi_1^2$ , a possibilistic Kullback-Leibler divergence would take the following form*

$$\text{KL}(\pi_1 \| \pi_2) = \int_{\mathbb{R}} \pi_1 \log \frac{\pi_1}{\sqrt{\pi_1}} = \frac{1}{2} \int_{\mathbb{R}} \pi_1 \log (\pi_1).$$

*But it has been demonstrated in (25) in appendix that if  $\pi$  has the form of (15) then we have*

$$\int_{\mathbb{R}} \pi_1 \log (\pi_1) = -2\pi |\Sigma_1|^{\frac{1}{2}}.$$

*Thus  $\text{KL}(\pi_1 \| \pi_2) = -\pi |\Sigma_1|^{\frac{1}{2}}$  which is negative and so a possibilistic Kullback-Leibler divergence does not respect the positivity.*

*Moreover, when the possibility distribution has the form of (15) we can determine the lowest value of  $\text{KL}(\pi_1 \| \pi_2) = \int_{\mathbb{R}} \pi_1 \log \frac{\pi_1}{\pi_2} = -\int_{\mathbb{R}} \pi_1 \log \frac{\pi_2}{\pi_1}$ . Indeed, since the logarithm is concave and thanks to the Jensen's inequality we can write*

$$\int_{\mathbb{R}} \pi_1 \log \frac{\pi_2}{\pi_1} \leq \log \left( \int_{\mathbb{R}} \pi_1 \frac{\pi_2}{\pi_1} \right) = \log \int_{\mathbb{R}} \pi_2.$$

*From (16), we can rewrite this inequality as*

$$\int_{\mathbb{R}} \pi_1 \log \frac{\pi_2}{\pi_1} \leq \log \left( 2\pi |\Sigma_2|^{\frac{1}{2}} \int_{\mathbb{R}} p_2 \right).$$

*Thus, we have  $\text{KL}(\pi_1 \| \pi_2) \geq -\log \left( 2\pi |\Sigma_2|^{\frac{1}{2}} \right)$ .*

In order to conserve the properties of a divergence, it is necessary to consider a symmetrical form of the  $KL$ -divergence. We have called this the divergence  $\Pi D_{KL}$  and it is expressed for all  $X$  in  $\mathbb{R}^2$  as

$$\Pi D_{KL}(\pi_1, \pi_2) = \int_{\mathbb{R}} \pi_1(x) \log \left( \frac{\pi_1(x)}{\pi_2(x)} \right) dx + \int_{\mathbb{R}} \pi_2(x) \log \left( \frac{\pi_2(x)}{\pi_1(x)} \right) dx \quad (18)$$

Note that (18) has to be close to the probabilistic Jeffrey's divergence [21].

**Proposition 3.** *The divergence  $\Pi D_{KL}$  is a particular case of divergence.*

The demonstration that the divergence  $\Pi D_{KL}$  is a  $f$ -divergence can be found in the appendix. Lastly, the analytical expression of the  $\Pi$ -divergence between two possibility distributions whose expression is of the form in (15) is given by:

$$\begin{aligned} \Pi D_{KL}(\pi_1, \pi_2) = & \pi \sqrt{|\Sigma_1|} \left[ \text{tr}(\Sigma_2^{-1} \Sigma_1) + (\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2) - 2 \right] \\ & + \pi \sqrt{|\Sigma_2|} \left[ \text{tr}(\Sigma_1^{-1} \Sigma_2) + (\mu_2 - \mu_1)^T \Sigma_1^{-1} (\mu_2 - \mu_1) - 2 \right] \end{aligned} \quad (19)$$

where  $|A|$  and  $\text{tr}(A)$  are respectively the determinant and the trace of the matrix  $A$ . The calculation steps needed to determine (19) are given in the appendix.

The second  $\Pi$ -divergence proposed is an extension of the Hellinger divergence by replacing probability distributions by

possibility distributions.

As a reminder, for two PDF  $p_1$  and  $p_2$ , the square of the Hellinger divergence is defined as

$$H_{prob}^2(p_1, p_2) = \frac{1}{2} \int_{\mathbb{R}} \left( \sqrt{p_1(x)} - \sqrt{p_2(x)} \right)^2 dx. \quad (20)$$

Note that, due to its form closely related to the Euclidean norm, the Hellinger divergence still preserves its divergence properties while replacing PDF by possibility distributions.

After some straightforward calculus, following [22], the square of the analytical expression of the Hellinger divergence (called  $\Pi D_H$ ) between two possibility distributions in the form of (15) is given by:

$$\begin{aligned} \Pi D_H^2(\pi_1, \pi_2) = & \pi \left( \sqrt{|\Sigma_1|} + \sqrt{|\Sigma_2|} \right) \\ & - 2\pi \left| \frac{1}{2} \Sigma_1^{-1} + \frac{1}{2} \Sigma_2^{-1} \right|^{-\frac{1}{2}} \times \\ & e^{-\frac{1}{8} (\mu_1 - \mu_2)^T \left( \frac{1}{2} \Sigma_2 + \frac{1}{2} \Sigma_1 \right)^{-1} (\mu_1 - \mu_2)}. \end{aligned} \quad (21)$$

Another approach to extend probabilistic divergence is to use a measure-based approach as proposed by V.Torra, Y.Narukawa, M.Sugeno in [23]. But unlike the distribution-based approach analytical expressions are not available in a lot of cases. Therefore we have not considered this approach in this paper but this could be the object of future works.

## B. Application

1) *Data description:* The application of change detection has been carried out over 14 pairs of images from the training set of the Onera Satellite Change Detection (OSCD) dataset where the groundtruth is available. The OSCD dataset has been presented in the paper "Urban Change Detection for Multispectral Earth Observation Using Convolutional Neural Network" presented at the international conference IGARSS'2018 [24]. The original OSCD dataset comprises 24 pairs of multispectral images taken from the Sentinel-2 satellite between 2015 and 2018. Locations are picked all over the world, in South America, USA, Europe, Middle East and Asia. For each location, registered pairs of 13-band multispectral satellite images obtained by the Sentinel-2 satellites are provided. Images vary in spatial resolution between 10m, 20m and 60m. Fig. 2 presents an example of a pair of images used in the application. The pair considered hereafter consists of two optical images of the city of Beihai taken between December 2016 and March 2018. The ground truth mask of the changes is provided as well as an example of change map obtained after running the algorithm of change detection.

2) *Experimental setup:* Let's now expose in more details the algorithm functioning of the change detection application.

Firstly, a preliminary step is necessary before running the core of the algorithm of change detection. The images are represented as double precision numbers in the ranges 0 to 1. Next, in order to avoid any border effects, a symmetric extension is applied to each satellite image proportionally to the size of the sliding window considered after. Then, the images are altered by the application of a Gaussian noise



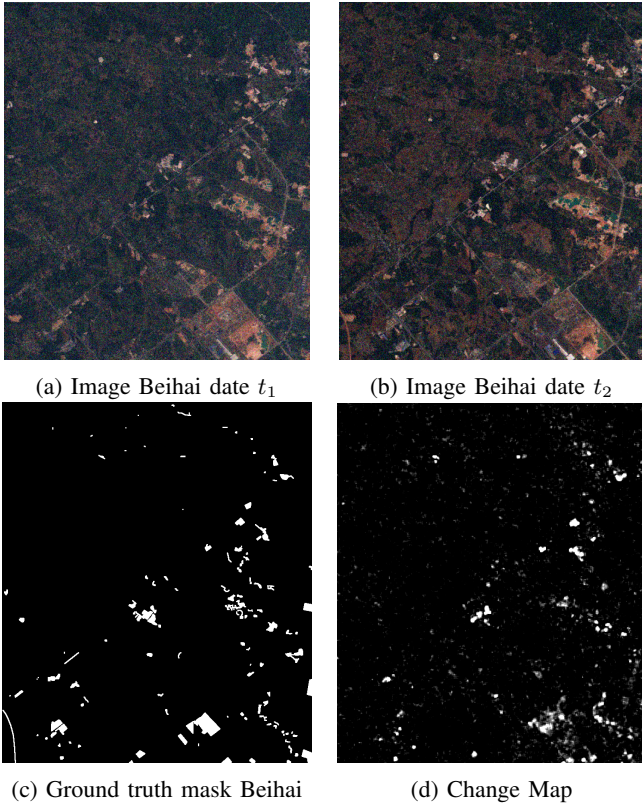


Fig. 2: Example of a pair of images with its associated Ground truth and a change map obtained.

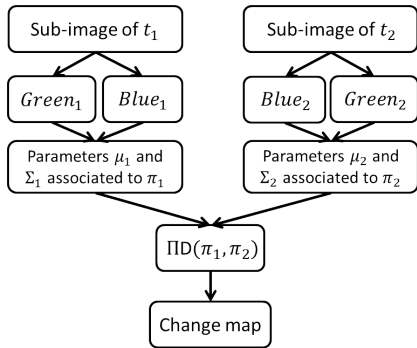


Fig. 3: Flowchart of proposed Change detection algorithm

with zero mean and fixed variance. This guarantees realistic covariance matrix estimation in presence of rank-deficient correlation structures that can be due to missing pixels. Note that the higher the value of the variance, the blurrier the image is. An example of the pair images of the city of Beihai altered by a Gaussian noise with zero mean and variance equal to 0.1 is given in fig. 4. Secondly, we run the algorithm of change detection over a pair of images. The principle of the algorithm is to calculate a measure of dissimilarity at each iteration between the image before ( $t_1$ ) and the image after ( $t_2$ ) for a pixel and its surrounding neighborhood. A high dissimilarity measure signifies a possible change. At each iteration, a patch (sliding window) is extracted for the two images and we consider two color channels out of three available for the

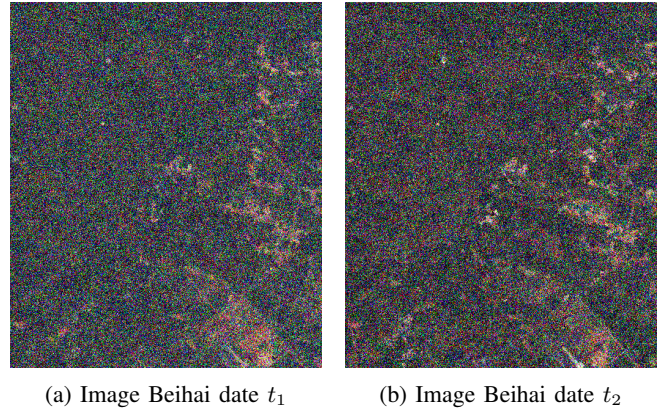


Fig. 4: Example of the pair image of the city of Beihai altered by a Gaussian noise with zero mean and a variance of 0.1.

sake of performing a bi-band Change Detection. Then, for the two color channels selected, we extract the mean vector  $\mu$  and the variance-covariance matrix  $\Sigma$  thanks to the Sample Covariance Matrix method (SCM) [25]. The parameters  $\mu$  and  $\Sigma$  being the parameters of the possibility distribution in (15). Then, with the parameters estimated, one of the previously proposed analytical possibilistic divergences (either in the form of (19) or (22)) is computed and its result is stored into the change map. The last step consists in shifting the sliding window by one pixel before the next iteration. Fig. 3 presents the functioning of the algorithm for any possibilistic divergence proposed and when the color channels selected for the parameters extraction are the green and the blue ones.

Next, when the change map is fully completed, the performance of the algorithm are determined by computing the Receiving Operator Characteristic (ROC) curve by comparing the change map and the ground truth mask. Then by looking at the area under the curve (AUC) we can calculate the overall performance of the algorithm for all possible thresholds of detection. See [26] for more details.

Finally, in order to highlight the interest of a possibilistic framework, we compare the performance of the possibilistic model to the corresponding probabilistic one. We repeat exactly the same steps with the analytical form of the symmetrical Kullback-Leibler divergence (called  $KL_{prob}$ ) or with the Hellinger divergence (called  $H_{prob}$ ) associated with the normal bivariate probability distribution.

3) *Experimental results:* Two different test protocols have been done with the change detection algorithm. In the first series of tests, we set a specific variance of 0.01 for the Gaussian noise applied to each image. Then we run the algorithm several times by increasing the size of the sliding window by 2 pixels each time from 3 by 3 pixels to 41 by 41 pixels. The results of this first series of tests are presented in fig. 5 for the two pairs of images. In the second series of tests, we fix the size of the sliding window to 5 by 5 pixels and we increase the variance of the applied noise at each running of the algorithm. The results of this second series of tests are presented in fig. 6. In fig. 5, the two divergences considered obtain similar results on the 14 pairs of images considered,

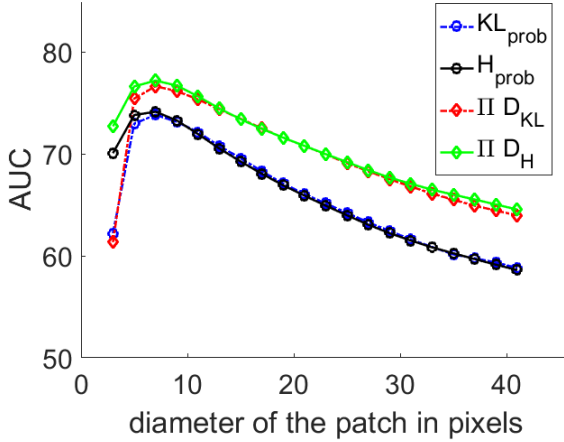


Fig. 5: Graphs showing the average AUC over 14 pairs of images as a function of the diameter of the patch used.

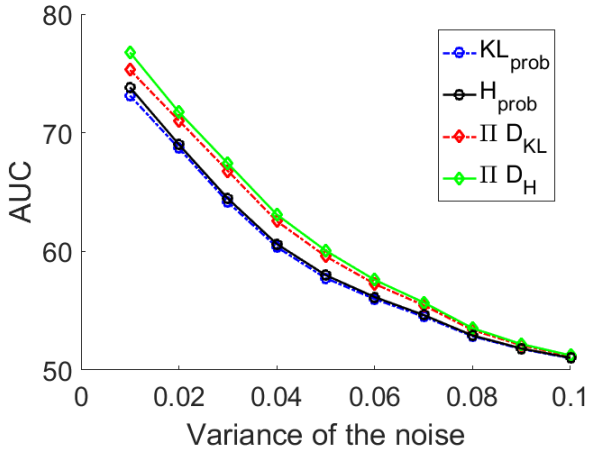


Fig. 6: Graphs showing the average AUC over 14 pairs of images according to the variance of the applied noise for a 5x5 pixels sliding window.

either by comparing possibility distributions or probability distributions. When the size of the patch is 3 by 3 pixels, the SCM estimator is affected by the presence of outliers in the patch. Therefore the overall performance of each method is affected and the results are below the optimal performance. On the other side, when the patch size is big enough (more than 10 pixels width), the SCM estimator is more consistent. Nevertheless, since most of the change size are about few pixels width, each method has its AUC decreasing as the size of the patch becomes bigger. Finally, the best compromise seems to be found for a 7x7 pixels patch although it depends highly on the size of the changes considered. Note that the possibilistic algorithm shows an interesting robustness to the noise observable both in the case of the Kullback-Leibler divergence or in the Hellinger divergence. This property of robustness needs to be confirmed in the second series of tests. Fig. 6 presents the average AUC obtained while testing the noise resilience of the possibilistic methods and of the prob-

abilistic methods over 14 pairs of images. As expected their performance decrease as the variance of the Gaussian noise applied increases. However, as we could begin to observe in the first series of tests, an interesting property of robustness to the noise emerges from the possibilistic methods compared to the probabilistic methods. This property seems to be confirmed by the fact that both divergences give similar results.

In the following paragraphs, the authors will propose a hypothesis that could explain the robustness to the noise of the possibilistic framework. First, it is important to recall that the final results in terms of AUC are impacted by three factors: the parameters estimator (SCM here), the statistical model (probabilistic or possibilistic) and the divergences used. In each case the parameter estimator and the divergences used are the same. Thus, the real difference between the probabilistic and the possibilistic framework is the statistical model chosen. As both statistical models are elliptical distributions, we have to look at their intrinsic properties to give an explanation on the behavior observed. The elliptical possibility distribution is homogeneous to a cumulative function (but slightly different) thanks to the probability-possibility transformation which integrates the probability density function. Thus the possibilistic divergence is a comparison between cumulative functions whereas the probabilistic framework compares density functions. That could explain the difference in terms of performance.

#### IV. CONCLUSION

In this paper we have proposed a new family of elliptical possibility distributions derived from the well-known elliptical family of probability distributions thanks to the extension to the n-dimension of the continuous probability-possibility transformation in one dimension.

Finally, through a change detection application, we have emphasized the consequences of a possibilistic framework in the field of multivariate data analysis especially when the data are noisy or their amount is insufficient to allow evaluating their characteristic parameters accurately.

In future works, two main aspects will be investigated; firstly, elliptical possibility distributions of higher dimensions will be considered for the sake of higher multivariate data analysis, and secondly, other divergences will be studied to propose useful analytical forms in application of continuous processing.

#### APPENDIX A

##### KULLBACK-LEIBLER DIVERGENCE

###### A. positivity & symmetry

Considering the following equality we have for  $0 < x \leq 1$ :

$$\begin{aligned} \log(x) \leq x - 1 &\Leftrightarrow \log\left(\frac{1}{x}\right) \leq \frac{1}{x} - 1 \\ &\Leftrightarrow \log(x) \geq 1 - \frac{1}{x}. \end{aligned}$$

then, by replacing  $x$  by  $\frac{x}{y}$  with  $0 < y \leq 1$ , and by multiplying the whole expression by  $x$  we obtain

$$x \log\left(\frac{x}{y}\right) \geq x - y.$$



Let's now replace  $x$  and  $y$  by any classical continuous possibility distribution  $\pi_1$  and  $\pi_2$ , which have values in the interval  $[0, 1]$ . Thus, it can become

$$\pi_1 \log \left( \frac{\pi_1}{\pi_2} \right) \geq \pi_1 - \pi_2$$

or

$$\pi_2 \log \left( \frac{\pi_2}{\pi_1} \right) \geq \pi_2 - \pi_1.$$

Finally by adding the two expressions together we have

$$\pi_1 \log \left( \frac{\pi_1}{\pi_2} \right) + \pi_2 \log \left( \frac{\pi_2}{\pi_1} \right) \geq 0.$$

And so  $\text{IID}(\pi_1, \pi_2) \geq 0$ .

The symmetry property is direct. Indeed by setting  $\pi_1 = \pi_2$ , we find  $\text{IID}(\pi_1, \pi_1) = 0$ .

### B. analytical form

First, let's consider the first term in (18) that we will call  $A$  and  $B$  will be the second term. Then replace  $\pi_1(x)$  and  $\pi_2(x)$  by their respective expressions given in (15). We then obtain the following form with

$$\forall X \in \mathbb{R}^2 \quad \pi_1(x) = e^{-\frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1)}$$

and  $\pi_2(x) = e^{-\frac{1}{2}(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2)}$ ;

$$\begin{aligned} A &= \int_{\mathbb{R}} \pi_1(x) \log \left( \frac{\pi_1(x)}{\pi_2(x)} \right) dx \\ &= \int_{\mathbb{R}} \pi_1(x) \log(\pi_1(x)) dx - \int_{\mathbb{R}} \pi_1(x) \log(\pi_2(x)) dx \\ &= \int_{\mathbb{R}} \frac{1}{2}(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2) \pi_1(x) dx \\ &\quad - \int_{\mathbb{R}} \frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1) \pi_1(x) dx. \end{aligned}$$

From (16) we can then rewrite it as

$$\begin{aligned} A &= 2\pi \sqrt{|\Sigma_1|} \left( \int_{\mathbb{R}} \frac{1}{2}(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2) p_1(x) dx \right. \\ &\quad \left. - \int_{\mathbb{R}} \frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1) p_1(x) dx \right). \end{aligned} \quad (22)$$

Let's call  $\alpha$  and  $\beta$  the two scalars such as

- $\alpha = \int_{\mathbb{R}} \frac{1}{2}(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2) p_1(x) dx$
- $\beta = \int_{\mathbb{R}} \frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1) p_1(x) dx.$

By noting  $\mathbb{E}(x)$  the expected value of the random variable  $X \in \mathbb{R}^2$ , we have:

$$\begin{aligned} \beta &= \int_{\mathbb{R}} \frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1) p_1(x) dx \\ &= \mathbb{E}_1 \left[ \text{tr} \left( \frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1) \right) \right] \\ &= \frac{1}{2} \mathbb{E}_1 \left[ \text{tr} \left( (x-\mu_1)(x-\mu_1)^T \Sigma_1^{-1} \right) \right] \\ &= \frac{1}{2} \text{tr} \left[ \mathbb{E}_1 \left( (x-\mu_1)(x-\mu_1)^T \right) \Sigma_1^{-1} \right] \\ &= \frac{1}{2} \text{tr} \left( \Sigma_1 \Sigma_1^{-1} \right) \\ &= \frac{1}{2} \text{tr} (I_2) \\ &= 1 \quad . \end{aligned}$$

and

$$\begin{aligned} \alpha &= \int_{\mathbb{R}} \frac{1}{2}(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2) p_1(x) dx \\ &= \int_{\mathbb{R}} \frac{1}{2} [(x-\mu_1)(\mu_1-\mu_2)]^T \Sigma_2^{-1}(x-\mu_1)(\mu_1-\mu_2) p_1(x) dx \\ &= \int_{\mathbb{R}} \frac{1}{2} \left( (x-\mu_1)^T \Sigma_2^{-1}(x-\mu_1) + 2(x-\mu_1)^T \Sigma_2^{-1}(\mu_1-\mu_2) \right. \\ &\quad \left. + (\mu_1-\mu_2)^T \Sigma_2^{-1}(\mu_1-\mu_2) \right) p_1(x) dx \\ &= \frac{1}{2} \left( \mathbb{E}_1[(x-\mu_1)^T \Sigma_2^{-1}(x-\mu_1)] + \mathbb{E}_1[2(x-\mu_1)^T \Sigma_2^{-1}(\mu_1-\mu_2)] \right. \\ &\quad \left. + \mathbb{E}_1[(\mu_1-\mu_2)^T \Sigma_2^{-1}(\mu_1-\mu_2)] \right) \\ &= \frac{1}{2} \left[ \text{tr}(\Sigma_2^{-1} \Sigma_1) + (\mu_1-\mu_2)^T \Sigma_2^{-1}(\mu_1-\mu_2) \right] \end{aligned}$$

The analytical expression of the first term of (18) then becomes:

$$A = \pi \sqrt{|\Sigma_1|} \left[ \text{tr}(\Sigma_2^{-1} \Sigma_1) + (\mu_1-\mu_2)^T \Sigma_2^{-1}(\mu_1-\mu_2) - 2 \right] \quad (23)$$

Finally, we obtain (19) by adding  $A$  and  $B$  where  $B$  is calculated in the same way as  $A$ .

NB: Note that the term  $\beta$  is actually similar to a Shanon entropy of a 2D-joint possibility distribution, such as:

$$\begin{aligned} \beta &= \int_{\mathbb{R}} \frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1) p_1(x) dx \\ &= -\frac{1}{2\pi \sqrt{|\Sigma_1|}} \int_{\mathbb{R}} \pi_1(x) \ln(\pi_1(x)) dx \end{aligned}$$

Since  $\beta = 1$ , we then have:

$$-\int_{\mathbb{R}} \pi_1(x) \ln(\pi_1(x)) dx = 2\pi \sqrt{|\Sigma_1|}. \quad (24)$$

Moreover, we also have:

$$\int_{\mathbb{R}} \pi_1(x) dx = 2\pi \sqrt{|\Sigma_1|} \int_{\mathbb{R}} p_1(x) dx = 2\pi \sqrt{|\Sigma_1|} \quad (25)$$

Since  $\int p_1(x) dx = 1$ .

Thus, an invariance property appears and can be written as:

$$\int_{-\infty}^{+\infty} \pi(x) (1 + \ln(\pi(x))) dx = 0 \quad (26)$$

## REFERENCES

- [1] H. Tanaka and H. Ishibuchi, "Evidence theory of exponential possibility distributions," *International Journal of Approximate Reasoning*, vol. 8, no. 2, pp. 123–140, 1993.
- [2] D. Dubois and H. Prade, "A class of fuzzy measures based on triangular norms a general framework for the combination of uncertain information," *International Journal of General Systems*, vol. 8, no. 1, pp. 43–61, 1982.
- [3] J. Vejnarová, "On interpretation of t-product extensions of possibility distributions," *Fuzzy Sets and Systems*, vol. 232, pp. 3–17, 2013.
- [4] E. L. Esmi, G. Barroso, L. C. Barros, and P. Sussner, "A family of joint possibility distributions for adding interactive fuzzy numbers inspired by biomathematical models," in *IFSA-EUSFLAT*, 2015.
- [5] C. Baudrit, I. Couso, D. Dubois *et al.*, "Joint propagation of probability and possibility in risk analysis: Towards a formal framework," *International Journal of Approximate Reasoning*, vol. 45, no. 1, pp. 82–105, 2007.
- [6] L. Coroianu and R. Fullér, "Nguyen type theorem for extension principle based on a joint possibility distribution," *International Journal of Approximate Reasoning*, vol. 95, pp. 22–35, 2018.
- [7] A. Ferrero, M. Prioli, S. Salicone, and B. Vantaggi, "2d probability-possibility transformations," in *Synergies of Soft Computing and Statistics for Intelligent Data Analysis*. Springer, 2013, pp. 63–72.
- [8] D. Dubois, L. Foulloy, G. Mauris, and H. Prade, "Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities," *Reliable computing*, vol. 10, no. 4, pp. 273–297, 2004.
- [9] S. Cambanis, S. Huang, and G. Simons, "On the theory of elliptically contoured distributions," *Journal of Multivariate Analysis*, vol. 11, no. 3, pp. 368–385, 1981.
- [10] C. Lesniewska-Choquet, A. M. Atto, G. Mauris, and G. Mercier, "Image change detection by possibility distribution dissemblance," in *Fuzzy Systems (FUZZ-IEEE), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–6.
- [11] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.
- [12] G. Mauris, V. Lasserre, and L. Foulloy, "A fuzzy approach for the expression of uncertainty in measurement," *Measurement*, vol. 29, no. 3, pp. 165–177, 2001.
- [13] G. Mauris, "Possibility distributions: A unified representation of usual direct-probability-based parameter estimation methods," *International Journal of Approximate Reasoning*, vol. 52, no. 9, pp. 1232–1242, 2011.
- [14] S. Bhattacharyya and P. J. Bickel, "Adaptive estimation in elliptical distributions with extensions to high dimensions," *Preprint*, 2014.
- [15] J. Owen and R. Rabinovitch, "On the class of elliptical distributions and their applications to the theory of portfolio choice," *The Journal of Finance*, vol. 38, no. 3, pp. 745–752, 1983.
- [16] K. W. Fang, *Symmetric Multivariate and Related Distributions: 0*. Chapman and Hall/CRC, 2017.
- [17] R. Blanc, E. Syrkinina, and G. Székely, "Estimating the confidence of statistical model based shape prediction," in *International Conference on Information Processing in Medical Imaging*. Springer, 2009, pp. 602–613.
- [18] M. D. Perlman, "Concentration inequalities for multivariate distributions: II. elliptically contoured distributions," *Lecture Notes-Monograph Series*, pp. 284–308, 1992.
- [19] G. Gallego, C. Cuevas, R. Mohedano, and N. Garcia, "On the mahalanobis distance classification criterion for multidimensional normal distributions," *IEEE Transactions on Signal Processing*, vol. 61, no. 17, pp. 4387–4396, 2013.
- [20] A. Ohlan, "An overview of recent developments of fuzzy divergence measures and their generalizations," *International Journal of Current Research and Review*, vol. 9, no. 7, p. 1, 2017.
- [21] H. Jeffreys, *The theory of probability*. OUP Oxford, 1998.
- [22] L. Pardo, *Statistical inference based on divergence measures*. Chapman and Hall/CRC, 2005.
- [23] V. Torra, Y. Narukawa, and M. Sugeno, "On the f-divergence for non-additive measures," *Fuzzy Sets and Systems*, vol. 292, pp. 364–379, 2016.
- [24] R. Caye Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral earth observation using convolutional neural networks," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2018.
- [25] S. T. Smith, "Covariance, subspace, and intrinsic crame/spl acute/r-rao bounds," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1610–1630, 2005.
- [26] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.