Hellinger distance for fuzzy measures

Vicenç Torra[1] Yasuo Narukawa[2] Michio Sugeno[3] Michael Carlson[4]

[1]IIIA-CSIC, Campus UAB s/n, 08193 Bellaterra (Catalonia, Spain)
[2]Toho Gakuen, Kunitachi, Tokyo 186-0004 (Japan)
[3]ECSC, c/ Gonzalo Gutiérrez Quirós s/n, 33600 Mieres (Spain)
[4]Dept. Statistics, U. Stockholm, SE-106 91 Stockholm (Sweden)

Abstract

Hellinger distance is a distance between two additive measures defined in terms of the Radon-Nikodym derivative of these two measures. This measure proposed in 1909 has been used in a large variety of contexts.

In this paper we define an analogous measure for fuzzy measures. We discuss them for distorted probabilities and give two examples.

Keywords: Hellinger distance, fuzzy measures, Radon-Nikodym derivative, Choquet integral, capacities.

1. Introduction

In 1909 Ernst Hellinger [7] introduced a distance to evaluate in which extent two probability distributions are similar. The definition is based on the Radon-Nikodym derivatives of the two probabilities with respect to a third probability measure. This distance has been used in a variety of contexts as e.g. data privacy [20], data mining [2].

In the area of fuzzy measures and capacities some research has been done to prove a Radon-Nikodym type theorem, as Graf [6] puts it. That is, researchers try to solve the question of when a given fuzzy measure can be expressed as the (Choquet) integral [1] of a function with respect to another given measure. When such relationship is found, we can say that this function is the Radon-Nikodym derivative. Graf [6] was one of the first authors to deal with this problem. He focuses on sub-additive fuzzy measures and gives (Theorem 4.3) necessary and sufficient conditions for this to happen. Sugeno [21] deals with the same problem but considering distorted probabilities. Rébillé [17] deals with the case of almost subadditive set functions of bounded sum.

In this paper we consider the definition of the Hellinger distance for fuzzy measures. To do so, we use the concept of derivative as used in [21], and the Choquet integral as an alternative to the Lebesgue integral. We will illustrate the definition with some examples using distorted probabilities, and prove some properties.

The calculation of our examples requires the computation of Choquet integrals. The problem of the calculation of the Choquet integral has been previously considered in e.g. [12, 13, 10, 14].

The structure of this paper is as follows. In Section 2 we review some concepts needed in the rest of this work. In Section 3 we introduce the Hellinger distance for fuzzy measures, give some examples and results. The paper finishes with some conclusions.

2. Preliminaries

This section reviews some results that are needed later on in the rest of this paper. We focus on the Hellinger distance and on the Choquet integral.

2.1. Measures

We begin with the definition of additive and non-additive (fuzzy) measures.

Definition 1 *A collection of subsets $\mathcal{F}$ of a set Ω is a σ-algebra if*

1. *$\Omega \in \mathcal{F}$ and $\emptyset \in \mathcal{F}$;*
2. *If $A \in \mathcal{F}$ then its complement $A^c \in \mathcal{F}$;*
3. *If $A_1, A_2, \ldots \mathcal{F}$, then their union $\cup A_i \in \mathcal{F}$.*

A pair $(\Omega, \mathcal{F})$ consisting of a set Ω and a σ-algebra $\mathcal{F}$ of subsets of Ω is called a measurable space.

Definition 2 *(see e.g. [19]) Let $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$ be measurable spaces and f a function from Ω to Λ. The function f is called a measurable function from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$ if and only if $f^{-1}(\mathcal{G}) \subset \mathcal{F}$.*

Here we use for a set $A \in \mathcal{G}$

$$f^{-1}(A) = \{\omega \in \Omega | f(\omega) \in A\},$$

and for a collection of subsets of $\mathcal{A}$ of Λ

$$f^{-1}(\mathcal{A}) = \{f^{-1}(A) | A \in \mathcal{A}\}$$

Definition 3 *Let $(\Omega, \mathcal{F})$ be a measurable space. A set function μ defined on $\mathcal{F}$ is called a (additive) measure if and only if*

- *$0 \leq \mu(A) \leq \infty$ for any $A \in \mathcal{F}$;*
- *$\mu(\emptyset) = 0$;*
- *If $A_1, A_2, \ldots$ are disjoint elements of $\mathcal{F}$ (i.e., $A_i \cap A_j = \emptyset$ for all $i \neq j$) then*

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i).$$

The triple $(\Omega, \mathcal{F}, \mu)$ is called a measure space.

Example 4 *There is a unique measure λ on $(\mathbb{R}, \mathcal{B})$ such that $\lambda([a, b]) = b - a$ for every finite interval $[a, b]$, $-\infty < a \leq b < \infty$. This is known as the Lebesgue measure.*

Here $\mathbb{R}$ is the real line and $\mathcal{B}$ is a Borel σ-algebra.

Definition 5 *Let $(\Omega, \mathcal{F})$ be a measurable space. A set function μ defined on $\mathcal{F}$ is called a fuzzy measure if an only if*

- $0 \leq \mu(A) \leq \infty$ *for any* $A \in \mathcal{F}$;
- $\mu(\emptyset) = 0$;
- *If* $A_1 \subseteq A_2 \subseteq \mathcal{F}$ *then*

$$\mu(A_1) \leq \mu(A_2)$$

Example 6 *Let $m : \mathbb{R}^+ \to \mathbb{R}^+$ be a continuous and increasing function such that $m(0) = 0$. Let λ be the Lebesgue measure. Let μ_m be the set function defined by*

$$\mu_m(A) = m(\lambda(A)) \tag{1}$$

for all A. The function μ_m is a fuzzy measure.

When $m(1) = 1$ and we restrict λ to the measurable space $([0, 1], \mathcal{B}_{[0,1]})$, then λ is a probability measure and μ_m is a distorted probability. Distorted probabilities have been studied in [5, 4, 11].

Example 7 *[21] Let $a > 0$. We define μ_a as the fuzzy measure of the form of Equation 1 with*

$$m_a(t) = t + at^2/2,$$

that is

$$\mu_a(A) = (\lambda(A)) + a(\lambda(A))^2/2.$$

Definition 8 Let μ be a fuzzy measure space.

1. μ is said to be submodular if

$$\mu(A) + \mu(B) \geq \mu(A \cup B) + \mu(A \cap B).$$

2. μ is said to be supermodular if

$$\mu(A) + \mu(B) \leq \mu(A \cup B) + \mu(A \cap B).$$

2.2. Hellinger distance

The Hellinger distance was defined for pairs of probabilities. It is defined in terms of the Radon-Nikodym derivative.

Definition 9 *(see e.g. [19, 8]) Let ν and μ be two additive measures in the same measurable space $(\Omega, \mathcal{F})$. Then,*

$$\text{if } \mu(A) = 0 \text{ implies } \nu(A) = 0$$

we say that ν is absolutely continuous with respect to μ. In this cas we write $\nu << \mu$.

Theorem 10 *Let μ and ν be two additive measures on $(\Omega, \mathcal{F})$ and μ be σ-finite. If $\nu << \mu$, then there exists a nonnegative measurable function f on Ω such that*

$$\nu(A) = \int_A f d\mu$$

The function f in this theorem is called the Radon-Nikodym derivative of ν with respect to μ, denoted

$$f = \frac{d\nu}{d\mu}.$$

The function f may not be unique, but if f_0 and f_1 are both Radon-Nikodym derivatives of ν, then $f_0 = f_1$ almost everywhere μ (see e.g. [8] p.7).

Definition 11 *Let P, Q be two probabilities that are absolutely continuous with respect to a third probability measure ν. The Hellinger distance between P and Q is defined as*

$$H(P, Q) = \sqrt{\frac{1}{2} \int \left(\sqrt{\frac{dP}{d\nu}} - \sqrt{\frac{dQ}{d\nu}} \right)^2 d\nu}$$

Here $dP/d\nu$ and $dQ/d\nu$ are the Radon-Nikodym derivatives of P and Q

The Hellinger distance does not depend on the measure ν (see [16]).

2.3. Choquet integral and derivatives with respect to fuzzy measures

In this section we review the definition of the Choquet integral as well as some results related to this integral. The Choquet integral integrates a function with respect to a fuzzy measure. When the fuzzy measure is additive, it corresponds to the Lebesgue integral.

Definition 12 *[1] Let X be a reference set, let $(X, \mathcal{A})$ be a measurable space, let μ be a fuzzy measure on $(X, \mathcal{A})$, and let g be a measurable function $g : X \to [0, 1]$; then, the Choquet integral of g with respect to μ is defined by*

$$(C) \int g d\mu := \int_0^\infty \mu_g(r) dr, \tag{2}$$

where $\mu_g(r) := \mu(\{x | g(x) > r\})$.

An alternative notation for this integral is $C_\mu(f)$.

The Choquet integral of g with respect to a fuzzy measure μ on a set A is defined by:

$$(C) \int_A g d\mu := \int_0^\infty \mu(\{x | g(x) > r\} \cap A) dr. \tag{3}$$

which corresponds to

$$(C) \int_A g d\mu = (C) \int g \cdot 1_A d\mu$$

Theorem 13 [1, 3, 9] Let μ be a non-additive measure in on $(\mathbb{R}, \mathcal{B})$, and f, g be non negative measurable function.

1. If μ is submodular, then

$$(C) \int (f+g) d\mu \leq (C) \int f d\mu + (C) \int g d\mu.$$

2. If μ is supermodular, then

$$(C) \int (f+g) d\mu \geq (C) \int f d\mu + (C) \int g d\mu.$$

Definition 14 *Let $(\Omega, \mathcal{F})$ be a measurable space and let $\nu, \mu : \mathcal{F} \to \mathbb{R}^+$ be fuzzy measures. We say that ν is a Choquet integral of μ if there exists a measurable function $g : \Omega \to \mathbb{R}^+$ with*

$$\nu(A) = (C) \int_A g d\mu \qquad (4)$$

for all $A \in \mathcal{F}$.

Then, in Graf [6], Nguyen [15], and Sugeno [21] the inverse problem is considered. That is, given measures μ and ν, the problem is to know whether a g exists that satisfies Equation 4 and, if so, compute such g.

Definition 15 *Let μ and ν be two fuzzy measures. If μ is a Choquet integral of ν, and g is a function such that Equation 4 is satisfied we write*

$$d\nu/d\mu = g,$$

and we say that g is a derivative of ν with respect to μ.

Graf's definition [6] considers the case of two measures when these measures are subadditive capacities. Sugeno's [21] definition of derivative, which follows, is of a function with respect to a distorted Lebesgue measure, and, thus suitable for the derivative of a distorted Lebesgue measure with respect to another distorted Lebesgue measure.

In short, Graf and Sugeno prove some cases where g exists. I.e., there is a unique g that satisfies Equation 4 for all $A \in \mathcal{F}$. See e.g. Theorem 4.3. in Graf.

Definition 16 *(Definition 2 in [21]) For a continuous and increasing function $f(t)$ with $f(0) = 0$, the derivative of the function f with respect to a fuzzy measure μ_m is defined as the inverse operation of the Choquet integral by*

$$df/d\mu_m = g, \qquad (5)$$

if g is found to be continuous and increasing.

If $\mu(A)$ is a distorted Lebesgue measure then there exists a function m such that $\mu_m(A) = m(\lambda(A))$. Similarly if $\nu(A)$ is a distorted Lebesgue there exists a function n such that $\nu_n(A) = n(\lambda(A))$. Then, naturally, if $\lambda(A) = \lambda(B)$ it holds that

$\nu(A) = \nu(B)$. Therefore, when $\lambda(A) = x$, $\nu(A) = \nu([0, x])$ and we can define a function f as follows: $f(x) = \nu([0, x])$. So, if $\lambda(A) = x$ then

$$\nu(A) = \nu([0, x]) = f(x) = (C) \int_{[0,x]} g d\mu_m.$$

As stated in the introduction, Graf and Sugeno, among others have proven conditions on when a fuzzy measure ν is a Choquet integral of another fuzzy measure μ. Here, we illustrate these results with two examples adapted from [21]. We will use these examples later.

Example 17 *(See Example 7 in [21]) Let μ_m be as in Example 7 (i.e., a distorted Lebesgue measure with $m(t) = t + at^2/2$). Let ν_n be another distorted measure with $n(t) = t^2$. Then,*

$$\nu_n(A) = n(\lambda(A)) = (C) \int_{[0,x]} g d\mu_m$$

for all A such that $\lambda(A) = x$ with $g(t) = (2/a)(1 - e^{-at})$.

Example 18 *(See Example 8 in [21]) Let μ_m be as in Example 7 (i.e., a distorted Lebesgue measure with $m(t) = t + at^2/2$). Let ν_n be another distorted measure with $n(t) = e^{at} - 1$ for $a > 0$. Then,*

$$\nu_n(A) = n(\lambda(A)) = (C) \int_{[0,x]} g d\mu_m$$

for all A such that $\lambda(A) = x$ with $g(t) = \cosh(at)$.

Let $m(t)$, $g(t)$ and $f(t)$ be continuously differentiable. Let $\mu([\tau, t])$ differentiable with respect to τ on $[0, t]$ for every $t > 0$. We require the regularity condition that $\mu(\{t\}) = 0$ holds for every $t \geq 0$. Let $\mu'([\tau, t])$ denote $(\partial/\partial\tau)\mu([\tau, t])$, where we note that $\mu'([\tau, t]) \leq 0$ for $\tau \leq t$. If μ is a distorted Lebesgue measure (Definition 1) in terms of m, i.e. $\mu = \mu_m$, then $\mu'([\tau, t]) = -m'(t - \tau)$ where $m'(t) = dm(t)/t$.

Theorem 19 *(Theorem 1 in [21]) Let $\mathcal{F}^+$ be the class of measurable, non-negative, continuous and increasing (non-decreasing) functions such that $g : \mathbb{R}^+ \to \mathbb{R}^+$. Let $g \in \mathcal{F}^+$, then the Choquet integral of g with respect to μ on $[0, t]$ is represented as:*

$$(C) \int_{[0,t]} g d\mu = \int_0^\infty \mu(\{\tau | g(\tau) \geq r\} \cap [0,t]) dr$$

$$= -\int_0^t \mu'([\tau, t]) g(\tau) dt, \qquad (6)$$

and when the measure is a distorted Lebesgue measure $\mu = \mu_m$ then

$$\int_0^\infty \mu(\{\tau | g(\tau) \geq r\} \cap [0,t]) dr = \int_0^t m'(t-\tau) g(\tau) d\tau.$$

For a function $h : \mathbb{R}^+ \to \mathbb{R}^+$, its Laplace transformation is denoted by $H(s)$ and its inverse Laplace transformation by $h(t) = L^{-1}[H(s)]$.

Proposition 20 *(Proposition 3 in [21]) Let $\mathcal{F}^+$, $g \in \mathcal{F}^+$, and $\mu = \mu_m$ be as in Theorem 19, let*

$$f(t) = (C) \int_{[0,t]} g d\mu_m = \int_0^t m'(t-\tau)g(\tau)d\tau.$$

Then, the Laplace transformation of $f(t)$ and the expression of $f(t)$ in terms of the inverse Laplace transformation are as follows:

$$F(s) = L(f(t)) = sM(s)G(s) \qquad (7)$$
$$f(t) = L^{-1}[sM(s)G(s)] \qquad (8)$$

where $M(s)$ is the Laplace transformation of m and $G(s)$ is the Laplace transformation of g.

Proposition 21 *(Proposition 4 in [21]) Let $f(t)$ be a continuous and increasing function with $f(0) = 0$, let μ_m be a distorted Lebesgue measure, then there exists an increasing (non-decreasing) function g so that $f(t) = (C) \int_{[0,t]} g(\tau)d\mu_m$ and the following holds:*

$$G(s) = F(s)/sM(s) \qquad (9)$$
$$g(t) = L^{-1}[F(s)/sM(s)]. \qquad (10)$$

Here, $F(s)$ is the Laplace transformation of f, M the Laplace transformation of m, and G the Laplace transformation of g.

3. Variation of Hellinger distance

We introduce now our definition of the Hellinger distance for non-additive measures.

Definition 22 *Let μ_1 and μ_2 be two fuzzy measures that are Choquet integrals of ν. The Hellinger distance between μ_1 and μ_2 is defined as*

$$H_\nu(\mu_1, \mu_2) = \sqrt{\frac{1}{2}(C) \int \left(\sqrt{\frac{d\mu_1}{d\nu}} - \sqrt{\frac{d\mu_2}{d\nu}} \right)^2 d\nu}$$

Here $d\mu_1/d\nu$ and $d\mu_2/d\nu$ are the derivatives of μ_1 and μ_2 according to Definition 15.

The first proposition below establishes that this definition is a proper generalization of standard Hellinger distance.

Proposition 23 *Let ν, μ_1, and μ_2 be three additive measures such that μ_1 and μ_2 are absolutely continuous with respect to ν. Then, $H_\nu(\mu_1, \mu_2)$ is the standard Hellinger distance of Definition 11.*

This proposition follows from the fact that the Choquet integral corresponds to the Lebesgue integral for additive measures.

We consider some additional properties below. The next proposition is obvious from the definition.

Proposition 24 $h_\nu(\mu_1, \mu_2) = 0$ *if $\mu_1 = \mu_2$.*

The next proposition follows from Theorem 13.

Proposition 25 *If ν is submodular, then we have*

$$h_\nu(\mu_1, \mu_2) + h_\nu(\mu_2, \mu_3) \geq h_\nu(\mu_1, \mu_3).$$

As h is symmetric and $h_\nu(\mu_1, \mu_2) \geq 0$, we have that our definition is a distance for submodular ν.

Corollary 26 *If ν is submodular, Definition 22 defines a distance.*

This definition is illustrated with an example based on the measures given above.

Example 27 *Let $a > 0$. Let μ_m be as in Example 7 (i.e., a distorted Lebesgue measure with $m(t) = t + at^2/2$). Let ν_{n_1} be the measure in Example 17 (i.e., a distorted Lebesgue measure with $n_1(t) = t^2$), and let ν_{n_2} be the measure in Example 18 (i.e., a distorted Lebesgue measure with $n_2(t) = e^{at} - 1$). Then, the Hellinger distance of μ_{n_1} and μ_{n_2} with respect to μ_m is defined as follows:*

$$H_{\mu_m}(\mu_1, \mu_2) = \sqrt{\frac{1}{2}(C) \int \left(\sqrt{\frac{d\mu_1}{d\mu_m}} - \sqrt{\frac{d\mu_2}{d\mu_m}} \right)^2 d\mu_m} =$$

$$= \sqrt{\frac{1}{2}(C) \int \left(\sqrt{(2/a)(1 - e^{-at})} - \sqrt{\cosh(at)} \right)^2 d\mu_m}$$

where

$$(C) \int \left(\sqrt{(2/a)(1 - e^{-at})} - \sqrt{\cosh(at)} \right)^2 d\mu_m =$$

$$\int \mu_m \left(\left\{ t \mid \left(\sqrt{(2/a)(1 - e^{-at})} - \sqrt{\cosh(at)} \right)^2 \geq r \right\} \right) dr$$

Now we consider another example of the new Hellinger distance.

Example 28 *Let us consider μ_m be the distorted Lebesgue measure with $m(t) = t^2$ and let ν^p be the distorted Lebesgue measure with distortion $n(t) = t^p$. Therefore, $\nu^p(A) = (\lambda(A))^p$ for $p \geq 2$, and $\nu^p([0,t]) = t^p$.*

In this example we compute the Hellinger distance between ν^2 and ν^p with respect to μ_m. That is, $H_{\mu_m}(\nu^2, \nu^p)$ using Definition 22.

It is known (see e.g. [18]) that for all $r > -1$, $s > 0$:

$$L[t^r] = \frac{\Gamma(r+1)}{s^{r+1}}$$
$$L^{-1}\left[\frac{1}{s^r}\right] = \frac{t^{r-1}}{\Gamma(r)}$$

where Γ is Euler's gamma function.

Therefore, the Laplace transforms of ν^p and μ are:

$$N^p(s) = L[\nu^p] = \frac{\Gamma(p+1)}{s^{p+1}}$$
$$M(s) = L[\mu] = \frac{2}{s^3} \qquad (11)$$

Now, let us consider $L[\frac{d\nu^p}{d\mu_m}]$. Applying Proposition 21, we have that:

$$L[\frac{d\nu^p}{d\mu_m}] = \frac{N^p(s)}{sM(s)} = \frac{\Gamma(p+1)}{2s^{p-1}}.$$

Using the inverse Laplace transform on this expression we obtain:

$$\begin{aligned} \frac{d\nu^p}{d\mu_m} &= L^{-1}[\frac{\Gamma(p+1)}{2s^{p-1}}] = \frac{\Gamma(p+1)}{2\Gamma(p-1)}t^{p-2} \\ &= \frac{p(p-1)}{2}t^{p-2}. \end{aligned}$$

Let us denote by $g(t)$ the following expression to be used later to compute the Hellinger distance between ν^2 and μ^p with respect to μ_m:

$$g(t) = \left(\sqrt{\frac{d\nu^2}{d\mu_m}} - \sqrt{\frac{d\nu^p}{d\mu_m}}\right)^2.$$

First note $\frac{d\nu^2}{d\mu_m} = 1$, therefore,

$$\begin{aligned} g(t) &= \left(1 - \sqrt{\frac{p(p-1)}{2}}t^{(p-2)/2}\right)^2 \\ &= 1 - \sqrt{2p(p-1)}t^{(p-2)/2} + \frac{p(p-1)}{2}t^{p-2}. \end{aligned}$$

From this follows that its Laplace transform is:

$$G(s) = \frac{1}{s} - \sqrt{2p(p-1)}\frac{\Gamma(p/2)}{s^{p/2}} + \frac{p(p-1)}{2}\frac{\Gamma(p-1)}{s^{p-1}}. \tag{12}$$

Now, let us consider the Laplace transform of the Choquet integral. Using Proposition 20, and the expressions for $M(s)$ and $G(s)$ in Equations 11 and 12, we have

$$L[(C)\int_0^t g d\mu_m] = \frac{2}{s^2}G(s) =$$

$$2[\frac{1}{s^3} - \sqrt{2p(p-1)}\Gamma(p/2)\frac{1}{s^{p/2+2}} +$$

$$\frac{\Gamma(p+1)}{2}\frac{1}{s^{p+1}}].$$

Therefore, using the inverse of the Laplace transform, we have that

$$(C)\int_0^t g d\mu_m =$$

$$2[\frac{t^2}{2} - \sqrt{2p(p-1)}\Gamma(p/2)\frac{t^{p/2+1}}{\Gamma(p/2+2)}$$

$$+\frac{\Gamma(p+1)}{2}\frac{t^p}{\Gamma(p+1)}] =$$

$$t^2 - \frac{8\sqrt{2p(p-1)}}{(p+2)p}t^{p/2+1} + t^p$$

Let us consider now ν^p and μ_m as distorted probabilities and take the integration on [0,1]. Then,

$$(C)\int_0^t g(\tau)d\mu_m(\tau) = 2 - \frac{8\sqrt{2p(p-1)}}{(p+2)p}.$$

Therefore, the Hellinger distance between ν^n and ν^p with respect to μ_m is calculated as

$$\begin{aligned} H_{\mu_m}(\nu^2, \nu^p) &= \sqrt{\frac{1}{2}(C)\int_0^1\left(\sqrt{\frac{d\nu^2}{\mu_m}} - \sqrt{\frac{d\nu^p}{\mu_m}}\right)^2 d\mu_m(\tau)} \\ &= \sqrt{1 - \frac{4\sqrt{2p(p-1)}}{(p+2)p}} \end{aligned} \tag{13}$$

We note that $H_{\mu_m}(\nu^2, \nu^p) = 0$ for $p = 2$.

Remark 29 *The Hellinger distance depends on μ_m.*

To illustrate the previous remark, let us consider the following example.

Example 30 *Let $\mu_{m'}$ be a distorted Lebesgue measure with $m'(t) = t$. Let ν^p be the distorted Lebesgue measure of Definition 28. Then,*

$$\frac{df}{d\mu_{m'}} = \frac{df}{dt}$$

and

$$(C)\int_0^t g d\mu_{m'} = \int_0^t g(\tau)d\tau.$$

We have

$$\frac{d\nu^p}{dt} = pt^{p-1}$$

and let us define $g(t)$ as follows

$$\begin{aligned} g(t) &= \left(\sqrt{\frac{d\nu^2}{dt}} - \sqrt{\frac{d\nu^p}{dt}}\right)^2 \\ &= \left(\sqrt{2t} - \sqrt{pt^{p-1}}\right)^2 = 2t - 2\frac{2p^{p/2}}{t} + pt^{p-1}. \end{aligned}$$

We obtain

$$\int_0^t g(t)dt = t^2 - \frac{2\sqrt{2p}}{p/2+1}t^{p/2+1} + t^p$$

and, hence,

$$\int_0^1 g(t)dt = 2 - \frac{4\sqrt{2p}}{p+2}.$$

Therefore, the Hellinger distance with respect to $\mu_{m'}$ with $m'(t) = t$ is

$$H_{\mu_{m'}}(\nu^2, \nu^p) = \sqrt{1 - \frac{2\sqrt{2p}}{p+2}}. \tag{14}$$

It is clear that the Hellinger distance between ν^2 and ν^p depends on the measure μ_m and $\mu_{m'}$ as Equation 13 and Equation 14 are, in general, different (only equivalent for $p = 2$ when both distances are zero).

The following can be deduced from the above expressions.

Lemma 31 *When $p \to \infty$, both $H_{\mu_m}(\nu^2, \nu^p) = 1$ and $H_{\mu_{m'}}(\nu^2, \nu^p) = 1$.*

Lemma 32 *Distances* $H_{\mu_m}(\nu^2, \nu^p)$ *and* $H_{\mu_{m'}}(\nu^2, \nu^p)$ *are increasing with respect to p, and the following holds*

- $H_{\mu_m}(\nu^2, \nu^p) \in [0,1]$ *for all* $p \geq 2$,
- $H_{\mu_{m'}}(\nu^2, \nu^p) \in [0,1]$ *for all* $p > 0$.

4. Conclusions

In this paper we have extended the definition of the Hellinger distance, which was initially defined for additive measures, to fuzzy measures. This extension relies on a Radon-Nikodym-type derivative for fuzzy measure.

As future work we plan to study some properties of this distance, and also study how this extension can be applied to other f-divergences, and if a general definition of f-divergence can also be given.

5. Acknowledgments

References

[1] Choquet, G. (1953/54) Theory of capacities, Ann. Inst. Fourier, 5, 131-295.

[2] Cieslak, D. A., Hoens, T. R., Chawla, N. V., Kegelmeyer, W. P. (2012) Hellinger distance decision trees are robust and skew-insensitive, Data Min Knowl Disc 24 136-158.

[3] Denneberg,D. (1994) Non Additive Measure and Integral, Dordorecht:Kluwer Academic Publishers.

[4] Edwards, W. (1953) Probability-preferences in gambling, American Journal of Psychology 66 349-364.

[5] Edwards, W. (1962) Subjective probabilities inferred from decisions, Psychological Review 69 109-135.

[6] Graf, S. (1980) A Radon-Nikodym theorem for capacities, Journal für die reine und angewandte Mathematik, 1980 192-214.

[7] Hellinger, E. (1909) Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen, Journal für die reine und angewandte Mathematik 136: 210-271.

[8] Keener, R. W. (2010) Theoretical Statistics, Springer.

[9] Klement, E. P., Mesiar, R., Pap, E. (2000) Triangular Norms, Kluwer Academic Publisher.

[10] Narukawa, Y. (2012) Choquet Integral on the Real Line as a Generalization of the OWA Operator. Proc. MDAI 2012, Lecture Notes in Artificial Intelligence 7647 56-65.

[11] Narukawa, Y., Torra, V. (2005) Fuzzy measure and probability distributions: distorted probabilities, IEEE Trans. on Fuzzy Systems 13:5 617-629.

[12] Narukawa, Y., Torra, V. (2009) Aggregation operators on the real line, Proc. 3rd International Workshop on Soft Computing Applications (SOFA 2009) (ISBN: 978-1-4244-5056-5), Szeged, Hungary and Arad, Romania, 185-188.

[13] Narukawa, Y., Torra, V. (2009) Continuous OWA operator and its calculation, Proc. IFSA-EUSFLAT (ISBN:978-989-95079-6-8), Lisbon, Portugal, 1132-1135.

[14] Narukawa, Y., Torra, V., Sugeno, M. Choquet integral with respect to a symmetric fuzzy measure of a function on the real line, Annals of Operations Research, in press. DOI: 10.1007/s10479-012-1166-6

[15] Nguyen, H.T. (2006) An Introduction to Random Sets, Chapman and Hall, CRC Press.

[16] Pollard, D. (2008) Distances and affinities between measures, from:
`http://www.stat.yale.edu/~pollard/`
`Books/Asymptopia/Metrics.pdf`

[17] Rébillé, R. (2012) A Radon-Nikodym derivative for almost subadditive set functions, Int. J. of Unc., Fuzziness and Knowledge-Based Systems, in press.

[18] Schiff, J. L. (1999) The Laplace Transform, Springer.

[19] Shao, J. (2010) Mathematical Statistics, Springer.

[20] Shlomo, N., Antal, L., Elliot, M. (2013) Disclosure Risk and Data Utility in Flexible Table Generators, Proc. NTTS 2013.

[21] Sugeno, M. (2013) A note on derivatives of functions with respect to fuzzy measures, Fuzzy Sets and Systems 222 1-17.