AIMS Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 7(6): 10892–10916.
DOI: 10.3934/math.2022609
Received: 16 November 2021
Revised: 07 March 2022
Accepted: 22 March 2022
Published: 01 April 2022

Research article

Monotone set-valued measures: Choquet integral, f-divergence and Radon-Nikodym derivatives

Zengtai Gong[1,*] **and Chengcheng Shen**[1,2]

[1] College of Mathematics and Statistics, Northwest Normal University, Lanzhou 730070, China
[2] Department of Advance Mathematics Teaching, Lanzhou Technology and Business College, Lanzhou 730101, China

* **Correspondence:** Email: zt-gong@163.com; Tel: +869317971430.

Abstract: Divergence as a degree of the difference between two data is widely used in the classification problems. In this paper, f-divergence, Hellinger divergence and variation divergence of the monotone set-valued measures are defined and discussed. It proves that Hellinger divergence and variation divergence satisfy the triangle inequality and symmetry by means of the set operations and partial ordering relations. Meanwhile, the necessary and sufficient conditions of Radon-Nikodym derivatives of the monotone set-valued measures are investigated. Next, we define the conjugate measure of the monotone set-valued measure and use it to define and discuss a new version f-divergence, and we prove that the new version f-divergence is nonnegative. In addition, we define the generalized f-divergence by using the generalized Radon-Nikodym derivatives of two monotone set-valued measures and examples are given. Finally, some examples are given to illustrate the rationality of the definitions and the operability of the applications of the results.

Keywords: f-divergence; Hellinger divergence; variation divergence; Radon-Nikodym derivatives; Choquet integral
Mathematics Subject Classification: 26E50, 28E10

1. Introduction

As a mathematical model to measure the degree of difference between two types of information, Pearson [27] first described the mathematical model between two random distributions in 1900. Ernst Hellinger [21] introduced a distance to evaluate to which extent two probability distributions are similar in 1909. The definition is based on the Radon-Nikodym derivatives of the two probabilities with respect to a third probability measure and has been widely used in data privacy and data mining etc. (refer to [9,29]). Later, as a generalization of distance, Kullback and Leibler [24] introduced another function to evaluate the divergence between probability distributions and was named as Kullback-Leibler divergence. However, it is not a distance function since the function does not satisfy the symmetry of the distance function. In 1952, Chornoff [7] generalized the Kullback-Leibler divergence to produce a new divergence function, named α-divergence, and used this divergence function to measure error calculations in classification problems. Amari proves that the α-divergence constitute a unique class belonging to both classes when the space of positive measures or positive arrays is considered. They are the canonical divergences derived from the dually flat geometrical structure of the space of positive measures [3]. The development of α-divergence has formed a very famous theory: Entropy theory. Like non-additive measure and Choquet integral, entropy theory is also an important tool for dealing with uncertain problems, which has been widely used in machine learning [37], data fusion [41], intuitive fuzzy sets [33], biological mathematics [26] and many other fields [2,23]. Csiszar introduced f-divergence in references [10,11]. The Csiszar f-divergence is a unique class of divergences having information monotonicity, from which the dual alpha geometrical structure with the Fisher metric is derived. Friedrich Liese and Igor Vajda deal with the f-divergences of Csiszr generalizing the discrimination information of Kullback, the total variation distance, the Hellinger divergence, and the Pearson divergence in reference [25]. In fact, divergence was originally used to evaluate two probability distribution difference degrees, is widely used in classification problems [5,6,12,22,28,34]. However, most of the divergence used in these problems is concentrated in discrete cases. For continuous cases, Torra et al. [35] first defined two f-divergence of non-additive measures and used the Choquet integral as an alternative to the Lebesgue integral to consider the definition of the f-divergence for non-additive measures in 2016. Later, use the same theory, they consider the definition of the f-divergence for discrete non-additive measures [36]. However, in 2019, Hamzeh Agahi found that the f-divergence for non-additive measures of Torra is not always non-negative, so he defined a new version of f-divergence which is always non-negative [1].

It is well known that the probability measure is a mathematical index that describes the measurement problems of the error free condition, but in practical problems, the conditions of countable additivity are so strong that it is difficult to operate. Especially, the characteristics of the measurement problems are not additive when measurement errors are unavoidable, involving subjective judgments or non-repetitive experiments. Therefore, in 1954, French mathematician Choquet put forward a theory called capacity [8]. Choquet capacity refers to a monotone set function whose domain is the power set of the given space, and its value is in the set of real numbers and is continuous. In 1974, Sugeno [31] put forward the concept of fuzzy measure, which refers to a set functions that replace column additivity with weak monotonicity and continuity. Therefore, the fuzzy measure and Choquet integral theory based on the fuzzy measure have attracted the attention of many scholars. However, since Choquet integral has a wide range of applications in information fusion, machine learning, pattern recognition, decision analysis and other fields, the study of Choquet integral mainly focuses on discrete cases. The monotone measure means a monotone set function which maps to zero on the empty set. The capacities and fuzzy measure are special cases of monotone measures. Since 2004, Gavrilut [13–17], Gavrilut and Croitoru [18] extended the concepts of the monotone measure to the set-valued case and established various results concerning non-atomicity, fuzziness, regularity, integrability and many other problems in monotone set-valued measures framework.

As an important part of nonlinear analysis, set-valued and the integrals of set-valued functions are widely used in cybernetics and decision theory. The integrals of set-valued functions based on Lebesgue integrals were introduced by Aumann in 1965, which are usually called Aumann integrals [4]. Moreover, several kinds of other set-valued integrals have been suggested, such as set-valued fuzzy integrals [39,40], set-valued pseudo-integrals [20], Riemann integral on monotone set-valued measure space [38]. It is well known that a lot of information is highly uncertain in nature. For such highly uncertain systems, set-valued are more accurate than real-valued. Therefore, set-valued probabilities and monotone set-valued measures are good tools for dealing with such highly uncertain systems [30]. However, for monotone set-valued measures, there is still no method to express the degree of difference between two monotone set-valued measures. Because divergence can well represent the degree of difference between two information, one method is to use divergence to describe the degree of difference between two monotone set-value measures. Therefore, in this paper, f-divergence, Hellinger divergence and variation divergence of the monotone set-valued measures are defined and discussed. It proves that Hellinger divergence and variation divergence satisfies the triangle inequality and symmetry by means of the set operations and partial ordering relations. Meanwhile, the necessary and sufficient conditions of Radon-Nikodym derivatives of the monotone set-valued measures are investigated. The conjugate measure of monotone set-valued measure and the new f-divergence are defined. We prove that the new f-divergence is nonnegative. Finally, we define the generalized f-divergence by using the generalized Radon-Nikdoym derivatives of two monotone set-valued measures.

The structure of the paper is as follows: In Section 2, we review some definitions that we need for the rest of our work. They are mainly results on measures, Choquet integral and divergence. In Section 3, we state the main results of this paper. Meanwhile, some examples are given to illustrate the effectiveness of the definitions and results proposed. The paper finishes with some conclusions.

2. Preliminaries

In this section, we review some definitions that we need for the rest of our work. They are mainly results on measures, Choquet integral and f-divergence for non-additive measures.

2.1. Measure

In this paper, Ω is always a nonempty set, $\mathcal{A}$ is a σ-algebra over Ω. The binary $(\Omega, \mathcal{A})$ is called a measurable space. Let $P_0(R^+)$ denote the class of all non-empty subsets of R^+, $P_{kc}(R^+)$ denote the class of all non-empty compact convex subsets of R^+. Let $A,\ B \in P_0(R^+)$, the partial order relation of $P_0(R^+)$ denoted by $A \lesssim B$, or $B \gtrsim A$, if the following conditions hold:
(1) $\forall\ x_0 \in A, \exists\ y_0 \in B$ such that $x_0 \leq y_0$;
(2) $\forall\ y_1 \in B, \exists\ x_1 \in A$ such that $x_1 \leq y_1$.

For any $A, B \in P_0(R^+)$, we denote

$$A + B = \{a + b\ |a \in A,\ b \in B\};$$

$$A \cdot B = \{a \cdot b\ |a \in A,\ b \in B\};$$

$$\max\{A, B\} = \begin{cases} A, & if A \gtrsim B; \\ B, & if B \gtrsim A; \\ \{\phi\}, & otherwise. \end{cases}$$

Let $f : [0,\infty] \to R$ be a real convex function with the conventions $0f(\frac{0}{0}) = 0$ and $0f(\frac{c}{0}) = c\lim_{t\to\infty}\frac{f(t)}{t}$, $c > 0$ (see [10]). In this paper, all convex functions satisfy these conventions.

Definition 2.1. *Let $(\Omega, \mathcal{A})$ be a measurable space. A set function μ defined on $\mathcal{A}$ is called a non-additive measure if and only if*
(1) $0 \le \mu(A) \le \infty$ *for any* $A \in \mathcal{A}$;
(2) $\mu(\phi) = 0$;
(3) *If* $A \subset B$, *then* $\mu(A) \le \mu(B)$.
The triple $(\Omega, \mathcal{A}, \mu)$ is called a non-additive measure space.

Definition 2.2. *We say $\bar{\mu}$ is the conjugate of a non-additive measure μ, if*

$$\bar{\mu}(A) = \mu(\Omega) - \mu(\Omega/A),\ A \in \mathcal{A}.$$

Definition 2.3. *(see [35]) Let μ be a non-additive measure.*
(1) μ is said to be supermodular if $\mu(A\cup B) + \mu(A\cap B) \ge \mu(A) + \mu(B)$.
(2) μ is said to be submodular if $\mu(A\cup B) + \mu(A\cap B) \le \mu(A) + \mu(B)$.

Definition 2.4. *(see [38]) Let $(\Omega, \mathcal{A})$ be a measurable space. A set-valued function $\pi : \mathcal{A} \to P_{kc}(R^+)$ is called a monotone set-valued measure if and only if*
(1) $\pi(\emptyset) = \{0\}$;
(2) $\forall A, B \in \mathcal{A}$, if $A \subseteq B$, then $\pi(A) \lesssim \pi(B)$.
The triple $(\Omega, \mathcal{A}, \pi)$ is called a montone set-valued measure space.

Let π be monotone set-valued measure, μ be non-additive measure. we said that μ is a choice for π, if $\mu(A) \in \pi(A)$ for all $A \in \mathcal{A}$. Let $I(R^+)$ denote the closed interval on R^+, $F : R \to I(R^+)$ be closed interval value function. $F \circ G : R \to 2^R$: $F \circ G(x) = \bigcup_{y\in G(x)} F(y)$ for any $x \in R$. $\overline{F} = \sup F$, $\underline{F} = \inf F$. Obviously, if F and G are both closed interval valued functions, then $F \circ G$ is a closed interval valued function.

In this paper, without special instructions, $(c)\int$ denotes the Choquet integral of real-valued function with respect to non-additive set-valued measure, $(C)\int$ denotes the Choquet integral of real-valued function with respect to non-additive set-valued measure, and $(\bar{C})\int$ denotes the Choquet integral of set-valued function with respect to non-additive set-valued measure.

2.2. On the f-divergence for non-additive measures

In this section, we review the definition of the Choquet integral and the f-divergence for non-additive measures.

Definition 2.5. *(see [8]) Let $(\Omega, \mathcal{A}, \mu)$ be a non-additive measure space, μ be a non-additive measure, g be a real measurable function. The Choquet integral of g with respect to μ is defined by*

$$(c)\int_A g\mathrm{d}\mu = \int_{-\infty}^{0} [\mu(\{x|g(x) \ge \alpha\}\cap A) - \mu(A)]\mathrm{d}\alpha + \int_0^{\infty} \mu(\{x|g(x) \ge \alpha\}\cap A)\mathrm{d}\alpha.$$

Let g be a non-negative measurable function. The Choquet integral of g with respect to μ is defined by

$$(c)\int_A g\mathrm{d}\mu = \int_0^{\infty} \mu(\{x|g(x) \ge \alpha\}\cap A)\mathrm{d}\alpha.$$

Proposition 2.1. *(see [1,19]) Let μ, ν be two non-additive measure, f, f_1, f_2 be real value function, we have the following results.*
(1) If $\mu(A) = 0$, $A \in \mathcal{A}$, then $(c)\int_A f d\mu = 0$.
(2) If $\mu_1 \leq \mu_2$, then $(c)\int_A f d\mu_1 \leq (c)\int_A f d\mu_2$, where $\mu_1 \leq \mu_2$ if and only if $\mu_1(A) \leq \mu_2(A)$ for $\forall A \in \mathcal{A}$.
(3) If $f_1 \leq f_2$, then $(c)\int_A f_1 d\mu \leq (c)\int_A f_2 d\mu$.
(4) If $A \subset B$, then $(c)\int_A f d\mu \leq (c)\int_B f d\mu$.
(5) $(c)\int_A f d(a\mu_1 + b\mu_2) = a \cdot (c)\int_A f d\mu_1 + b \cdot (c)\int_A f d\mu_2$ for $a, b > 0$ and $A \in \mathcal{A}$, where $(a \cdot \mu_1 + b \cdot \mu_2)(B) = a \cdot \mu_1(B) + b \cdot \mu_2(B)$.
(6) $(c)\int_A -f d\mu = -(c)\int_A f d\bar{\mu}$.
(7) $(c)\int_A f + \alpha d\mu = (c)\int_A f d\mu + \alpha\mu(A)$ for any real α.
(8) $\mu_n \uparrow (\downarrow)\mu \Leftrightarrow$ for any f we have $(c)\int f d\mu_n \uparrow (\downarrow)(c)\int f d\mu$, where $\mu_n \uparrow (\downarrow)\mu$ if and only if $\mu_n(A) \uparrow (\downarrow)\mu(A)$ for all $A \in \mathcal{A}$.

Definition 2.6. *(see [35]) Let $(\Omega, \mathcal{A})$ be a measurable space, μ and ν be two non-additive measures. We say that ν is a Choquet integral of μ if there exists a measurable function $g : \Omega \to R^+$ with*

$$\nu(A) = (c)\int_A g d\mu$$

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

Definition 2.7. *(see [35]) Let μ and ν be two non-additive measures. If ν is a Choquet integral of μ, and g is a real function such that Definition 2.6 is satisfied. Then we write $g = d\nu/d\mu$, and we say that g is a Radon-Nikdoym derivative of ν with respect to μ.*

Definition 2.8. *(see [35]) Let μ_1 and μ_2 be two non-additive measures that are Choquet integrals of μ. Let f be a convex function with $f(1) = 0$. The f-divergence between μ_1 and μ_2 is defined as*

$$D_{f,\mu}(\mu_1, \mu_2) = (c)\int \frac{d\mu_2}{d\mu} f\left(\frac{d\mu_1/d\mu}{d\mu_2/d\mu}\right) d\mu.$$

Here, $d\mu_1/d\mu$ and $d\mu_2/d\mu$ are the derivatives of μ_1 and μ_2 with respect to μ according to Definition 2.7.

Definition 2.9. *(see [12]) Let $P = (p_1, ..., p_k)$ and $Q = (q_1, ..., q_k)$ be two probability distributions. Let f be a convex function with $f(1) = 0$. Then the discrete f-divergence is defined by*

$$D_f(P, Q) = \sum_{i=1}^{k} q_i f\left(\frac{p_i}{q_i}\right).$$

According to the discrete f-divergence. We review some of these expressions below [7].
(1) The Hellinger distances is defined by

$$H(P, Q) = \sqrt{\frac{1}{2}\sum_{i=1}^{k}(\sqrt{p_i} - \sqrt{q_i})^2}.$$

It corresponds to the case of $f(x) = (1 - \sqrt{x})^2$, and formally, $H(P, Q) = \sqrt{(1/2)D_f(P, Q)}$.

(2) The Kullback-Leibler distance corresponds to the f-divergence with $f(x) = x\log x$, and its expression is

$$KL(P,Q) = \sum_{i=1}^{k} p_i \log(\frac{p_i}{q_i}).$$

(3) The variation distance is defined by

$$\delta(P,Q) = \frac{1}{2}\sum_{i=1}^{k} |p_i - q_i|.$$

If $f(x) = |x-1|$, then $\delta(P,Q) = \frac{1}{2}D_f(P,Q)$.

(4) The α distance is defined by

$$D_\alpha(P,Q) = (\frac{1}{\alpha-1})\log(\sum_{i=1}^{k} p_i^\alpha q_i^{1-\alpha}).$$

When $f(x) = x^\alpha$ for $\alpha > 0$ and $\alpha \neq 1$, we have that $D_\alpha(P,Q) = (1/(\alpha-1))\log D_f(P,Q)$. The α distance generalizes the Kullback-Leibler distance because $\lim_{\alpha\to 1} D_\alpha(P,Q) = KL(P,Q)$.

(5) The χ^2 distance corresponds to the function $f(x) = (x-1)^2$, and its expression is

$$D_{\chi^2}(P,Q) = \sum_{i=1}^{k} \frac{(p_i - q_i)^2}{q_i}.$$

In general, f-divergence neither satisfies the triangular inequality nor the symmetry. Therefore, f-divergence is not a distance. For example, χ^2 distance is not symmetric, Kullback-Leibler distance does not satisfy the triangle inequality. Obviously, the Hellinger distance and variation distance satisfies symmetry and the triangular inequality.

Theorem 2.1. *(see [1]) Let μ_1 and μ_2 be two non-additive measures that are Choquet integrals of μ. If $f : [0,\infty) \to R$ is a convex function, then*

$$\max\{D_{f,\mu}(\mu_1,\mu_2), D_{f,\bar{\mu}}(\mu_1,\mu_2)\} \geq \mu(\Omega) f(\frac{1}{\mu(\Omega)}(c)\int \frac{d\mu_1}{d\mu} d\mu).$$

(1) If f is an increasing convex function, then

$$D_{f,\mu}(\mu_1,\mu_2) \geq \mu(\Omega) f(\frac{1}{\mu(\Omega)}(c)\int \frac{d\mu_1}{d\mu} d\mu).$$

(2) If f is an decreasing convex function, then

$$D_{f,\bar{\mu}}(\mu_1,\mu_2) \geq \mu(\Omega) f(\frac{1}{\mu(\Omega)}(c)\int \frac{d\mu_1}{d\mu} d\mu).$$

3. Main result

3.1. f-divergence of monotone set-valued measure with respect to a convex function

The f-divergence was defined for pairs of monotone set-valued measure. It is defined in terms of the Radon-Nikodym derivative. In this part, let $\mu_1(A) = \inf \pi(A)$, $\mu_2(A) = \sup \pi(A)$. μ_1 and μ_2 respectively are said to be the minimum choice and maximum choice of π. Obviously, $\mu_1 \in \pi,\ \mu_2 \in \pi$.

Definition 3.1. *Let $(\Omega, \mathcal{A}, \pi)$ be monotone set-valued measurable space, f be a measurable function, $A \in \mathcal{A}$. The Choquet integral of f with respect to π is defined by*

$$(C)\int_A f\mathrm{d}\pi = \{(c)\int_A f\mathrm{d}\mu, \mu \in \pi\}.$$

Definition 3.2. *Let $(\Omega, \mathcal{A}, \pi)$ be monotone set-valued measurable space, f be a measurable function, $A \in \mathcal{A}$, we say the Choquet integral of f with respect to π is bounded if $-\infty < (c)\int_A f\mathrm{d}\mu < \infty$ for any $\mu \in \pi$.*

Definition 3.3. *Let $(\Omega, \mathcal{A}, \pi)$ be monotone set-valued measurable space, π_1 and π_2 be two monotone set-valued measure. If there exists a nonnegative measurable function $g : \Omega \to R$ such that $\pi_1(A) = (C)\int_A g\mathrm{d}\pi_2$, $A \in \mathcal{A}$, then the function g is called the Radon-Nikodym derivative of π_1 with respect to π_2, denoted $g = \mathrm{d}\pi_1/\mathrm{d}\pi_2$ or $g = \frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_2}$.*

Definition 3.4. *Let $(\Omega, \mathcal{A}, \pi)$ be a monotone set-valued measurable space, π_1 and π_2 be two monotone set-valued measure and exists Radon-Nikodym derivatives with respect to π_3, f be a convex function with $f(1) = 0$. Then the f-divergence between π_1 and π_2 is defined as*

$$D_{f,\pi_3}(\pi_1, \pi_2) = (C)\int \frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_3} f(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}\pi_3.$$

Remark 3.1. *Specially, according to Definition 3.4, we can deduce the following conclusions.*

(1) The Hellinger divergence between π_1 and π_2 is defined as

$$H_{\pi_3}(\pi_1, \pi_2) = \sqrt{\frac{1}{2}(C)\int (\sqrt{\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_3}} - \sqrt{\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_3}})^2 \mathrm{d}\pi_3}.$$

Obviously, if $f(x) = (1 - \sqrt{x})^2$, then $H_{\pi_3}(\pi_1, \pi_2) = \sqrt{\frac{1}{2}D_{f,\pi_3}(\pi_1, \pi_2)}$.

(2) The variation divergence between π_1 and π_2 is defined as

$$\delta_{\pi_3}(\pi_1, \pi_2) = \frac{1}{2}(C)\int |\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_3} - \frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_3}|\mathrm{d}\pi_3.$$

If $f(x) = |x - 1|$, then $\delta_{\pi_3}(\pi_1, \pi_2) = \frac{1}{2}D_{f,\pi_3}(\pi_1, \pi_2)$.

(3) The Kullback-Leibler divergence corresponds to the f-divergence with $f(x) = x\log x$, and its expression is

$$KL_{\pi_3}(\pi_1, \pi_2) = (C)\int \frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_3}\log(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}\pi_3.$$

(4) The Kullback-Leibler divergence is not symmetric. The symmetric version of $KL_{\pi_3}(\pi_1, \pi_2)$ is given by

$$J_{\pi_3}(\pi_1, \pi_2) := KL_{\pi_3}(\pi_1, \pi_2) + KL_{\pi_3}(\pi_2, \pi_1),$$

which is called J-divergence.

(5) The α divergence is defined by

$$D_\alpha(\pi_1, \pi_2) = \{(\frac{1}{\alpha - 1})\log(\int (\frac{d\pi_1}{d\pi_3})^\alpha (\frac{d\pi_2}{d\pi_3})^{1-\alpha} dm) : m \in \pi_3\}.$$

(6) The χ^2 divergence corresponds to the function $f(x) = (x-1)^2$, and its expression is

$$D_{\chi^2,\pi_3}(\pi_1, \pi_2) = \int \frac{(d\pi_1/d\pi_3 - d\pi_2/d\pi_3)^2}{d\pi_2/d\pi_3} d\pi_3.$$

Remark 3.2. *If monotone set-valued measure in Definition 3.4 degenerate into the non-additive measure. Then f-divergence, Hellinger divergence and variation divergence of monotone set-valued measure degenerated into f-divergence, Hellinger distance and variation distance of non-additive measure respectively [35].*

The Definition 3.3 is not always well defined. First note that, in general, a pair of arbitrary monotone set-valued measures do not always have a Radon-Nikodym derivative. In case that derivatives exists, it is well defined when for π_1, π_2 and π_3 there is only one derivative g such that $g = d\pi_1/d\pi_3$ and there is only one function f such that $f = d\pi_2/d\pi_3$. We have the following theorem for Radon-Nikdoym derivative of the monotone set-valued measure.

Theorem 3.1. *Let $(\Omega, \mathcal{A}, \pi)$ be monotone set-valued measurable space, π_1 and π_2 be two monotone set-valued measure, μ_i^S, μ_i^I be the maximum choice and the minimum choice of π_i $(i = 1, 2)$ respectively, let $\frac{d\mu_1^S}{d\mu_2^S} = f$, $\frac{d\mu_1^I}{d\mu_2^I} = g$. There exists a measurable function h that is the Radon-Nikdoym derivative of π_1 with respect to π_2 if and only if f is almost everywhere equal to g.*

Proof. Suppose $\pi_1(A) = (C)\int_A h d\pi_2$. Then there exists a $\mu \in \pi_1$ such that $\mu(A) = (c)\int_A h d\mu_2^I$ for $\mu_2^I \in \pi_2$. Here, μ is obviously the minimum choice of π_1. That is, $\frac{d\mu_1^I}{d\mu_2^I} = h$. Similarly, we have $\frac{d\mu_1^S}{d\mu_2^S} = h$. Therefore, f is almost everywhere equal to g.

We just need to prove that

$$\pi_1(A) = (C)\int_A f d\pi_2.$$

First, since π_2 is a compact convex set-valued mapping, we have that $(C)\int_A f d\pi_2$ is a compact convex set. For any $x \in \pi_1(A)$, we may assume that

$$\alpha = \frac{\mu_1^S(A) - x}{\mu_1^S(A) - \mu_1^I(A)},$$

then we have

$$x = \alpha\mu_1^I(A) + (1-\alpha)\mu_1^S(A).$$

Let

$$\mu_1^S(A) = (c)\int_A f d\mu_2^S, \ \ \mu_1^I(A) = (c)\int_A f d\mu_2^I.$$

Since π_2 is a compact convex set-valued mapping, then there exist $\mu \in \pi_2$ and $\mu = \alpha\mu_2^I + (1-\alpha)\mu_2^S$ such that $x = (c)\int_A f \mathrm{d}\mu$. Therefore,

$$x \in (C)\int_A h \mathrm{d}\pi_2,$$

so we have

$$\pi_1(A) \subset (C)\int_A f \mathrm{d}\pi_2.$$

Second, for any $y \in (C)\int_A f\mathrm{d}\pi_2$, there exists $\mu \in \pi_2$ such that $y = (c)\int_A f\mathrm{d}\mu$. Obviously,

$$(c)\int_A f\mathrm{d}\mu_2^I \leqslant y \leqslant (c)\int_A f\mathrm{d}\mu_2^S.$$

Let

$$\alpha = \frac{(c)\int_A f\mathrm{d}\mu_2^S - y}{(c)\int_A f\mathrm{d}\mu_2^S - (c)\int_A f\mathrm{d}\mu_2^I},$$

we have

$$y = \alpha(c)\int_A f\mathrm{d}\mu_2^I + (1-\alpha)(c)\int_A f\mathrm{d}\mu_2^S.$$

Since $\pi_1(A)$ is a compact convex set, and

$$(c)\int_A f\mathrm{d}\mu_2^I \in \pi_1(A),\ (c)\int_A f\mathrm{d}\mu_2^S \in \pi_1(A).$$

So we have $y \in \pi_1(A)$. That is,

$$\pi_1(A) \supset (C)\int_A h\mathrm{d}\pi_2.$$

Therefore,

$$\pi_1(A) = (C)\int_A h\mathrm{d}\pi_2.$$

3.2. Properties and examples

We consider some additional properties below. They apply when the f-divergence is well defined.

Proposition 3.1. $D_{f,\pi_3}(\pi_1,\pi_2) = \{0\}$ *if* $\pi_1 = \pi_2$.

Proposition 3.2. *The Hellinger divergence and variation divergence satisfy symmetry.*

Proposition 3.3. *If the maximum choice and the minimum choice of* π_4 *are submodular and continuous from below, then we have*

$$H_{\pi_4}(\pi_1,\pi_2) + H_{\pi_4}(\pi_2,\pi_3) \gtrsim H_{\pi_4}(\pi_1,\pi_3).$$

Proposition 3.4. *If the maximum choice and the minimum choice of* π_4 *are submodular, then we have*

$$\delta_{\pi_4}(\pi_1,\pi_2) + \delta_{\pi_4}(\pi_2,\pi_3) \gtrsim \delta_{\pi_4}(\pi_1,\pi_3).$$

To prove the above four properties, we give the following two lemmas.

Lemma 3.1. *(see [35]) Let* $(\Omega, \mathcal{A}, \mu)$ *be non-additive measurable space,* f *and* g *be two non-negative measurable functions.*
(1) If μ *is submodular, then* $(c)\int_A (f+g)\mathrm{d}\mu \leq (c)\int_A f\mathrm{d}\mu + (c)\int_A g\mathrm{d}\mu$;
(2) If μ *is supermodular, then* $(c)\int_A (f+g)\mathrm{d}\mu \geq (c)\int_A f\mathrm{d}\mu + (c)\int_A g\mathrm{d}\mu$.

Lemma 3.2. *(see [35]) Let* $(\Omega, \mathcal{A}, \mu)$ *be non-additive measurable space,* f *and* g *be non-negative measurable functions. If* μ *is a submodular and continuous from blew, then*

$$[(c)\int (f+g)^2 \mathrm{d}\mu]^{\frac{1}{2}} \leq ((c)\int f^2 \mathrm{d}\mu)^{\frac{1}{2}} + ((c)\int g^2 \mathrm{d}\mu)^{\frac{1}{2}}.$$

Proof. Propositions 3.1 and 3.2 are obvious, we just need to prove Propositions 3.3 and 3.4.

Proof of Proposition 3.3: Let μ_1 be the minimum choice of π_4, μ_2 be the maximum choice of π_4. Suppose $a \in H_{\pi_4}(\pi_1, \pi_2) + H_{\pi_4}(\pi_2, \pi_3)$, then there exists $\mu_3, \mu_4 \in \pi_4$ such that

$$a = \sqrt{\frac{1}{2}(c)\int (\sqrt{\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}} - \sqrt{\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}})^2 \mathrm{d}\mu_3} + \sqrt{\frac{1}{2}(c)\int (\sqrt{\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}} - \sqrt{\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}})^2 \mathrm{d}\mu_4}.$$

Applying Lemma 3.2, we have

$$\begin{aligned} a &= \sqrt{\frac{1}{2}(c)\int (\sqrt{\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}} - \sqrt{\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}})^2 \mathrm{d}\mu_3} + \sqrt{\frac{1}{2}(c)\int (\sqrt{\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}} - \sqrt{\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}})^2 \mathrm{d}\mu_4} \\ &\geq \sqrt{\frac{1}{2}(c)\int (\sqrt{\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}} - \sqrt{\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}})^2 \mathrm{d}\mu_1} + \sqrt{\frac{1}{2}(c)\int (\sqrt{\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}} - \sqrt{\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}})^2 \mathrm{d}\mu_1} \\ &\geq \sqrt{\frac{1}{2}(c)\int (\sqrt{\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}} - \sqrt{\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}})^2 \mathrm{d}\mu_1} \\ &\in H_{\pi_4}(\pi_1, \pi_3). \end{aligned}$$

Suppose $b \in H_{\pi_4}(\pi_1, \pi_3)$, then there exists $\mu_5 \in \pi_4$ such that

$$\begin{aligned} b &= \sqrt{\frac{1}{2}(c)\int (\sqrt{\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}} - \sqrt{\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}})^2 \mathrm{d}\mu_5} \\ &\leq \sqrt{\frac{1}{2}(c)\int (\sqrt{\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}} - \sqrt{\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}})^2 \mathrm{d}\mu_2} \\ &\leq \sqrt{\frac{1}{2}(c)\int (\sqrt{\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}} - \sqrt{\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}})^2 \mathrm{d}\mu_2} + \sqrt{\frac{1}{2}(c)\int (\sqrt{\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}} - \sqrt{\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}})^2 \mathrm{d}\mu_2} \\ &\in H_{\pi_4}(\pi_1, \pi_2) + H_{\pi_4}(\pi_2, \pi_3). \end{aligned}$$

Therefore, $H_{\pi_4}(\pi_1, \pi_2) + H_{\pi_4}(\pi_2, \pi_3) \gtrsim H_{\pi_4}(\pi_1, \pi_3)$.

Proof of Proposition 3.4: Let μ_1 be the minimum choice of π_4, μ_2 be the maximum choice of π_4. Suppose $a \in \delta_{\pi_4}(\pi_1, \pi_2) + \delta_{\pi_4}(\pi_2, \pi_3)$, there are $\mu_3, \mu_4 \in \pi_4$ such that

$$a = \frac{1}{2}(c)\int |\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4} - \frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}| \mathrm{d}\mu_3 + \frac{1}{2}(c)\int |\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4} - \frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}| \mathrm{d}\mu_4.$$

Applying Lemma 3.1, we have

$$\begin{aligned}
a &= \frac{1}{2}(c)\int|\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}-\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}|\mathrm{d}\mu_3+\frac{1}{2}(c)\int|\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}-\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}|\mathrm{d}\mu_4\\
&\geq \frac{1}{2}(c)\int|\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}-\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}|\mathrm{d}\mu_1+\frac{1}{2}(c)\int|\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}-\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}|\mathrm{d}\mu_1\\
&\geq \frac{1}{2}(c)\int(|\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}-\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}|+|\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}-\frac{\mathrm{d}\mu_3}{\mathrm{d}\pi_4}|)\mathrm{d}\mu_1\\
&\geq \frac{1}{2}(c)\int|\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}-\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}|\mathrm{d}\mu_1\\
&\in \delta_{\pi_4}(\pi_1,\pi_3).
\end{aligned}$$

Suppose $b\in\delta_{\pi_4}(\pi_1,\pi_3)$, then there exists $\mu_5\in\pi_4$ such that

$$\begin{aligned}
b &= \frac{1}{2}(c)\int|\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}-\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}|\mathrm{d}\mu_5\\
&\leq \frac{1}{2}(c)\int|\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}-\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}|\mathrm{d}\mu_2\\
&\leq \frac{1}{2}(c)\int|\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}-\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}|\mathrm{d}\mu_2+\frac{1}{2}(c)\int|\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}-\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}|\mathrm{d}\mu_2\\
&\in \delta_{\pi_4}(\pi_1,\pi_2)+\delta_{\pi_4}(\pi_2,\pi_3).
\end{aligned}$$

Therefore, $\delta_{\pi_4}(\pi_1,\pi_2)+\delta_{\pi_4}(\pi_2,\pi_3)\gtrsim\delta_{\pi_4}(\pi_1,\pi_3)$.

Remark 3.3. *In Propositions 3.3 and 3.4, the inequality is false when the maximum choice and the minimum choice are supermodular. For example:*

(1) Let $h_1=\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}-\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}$, $h_2=\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_4}-\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}$, $h_3=\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_4}-\frac{\mathrm{d}\pi_3}{\mathrm{d}\pi_4}$; μ_2 *be the maximum choice of* π_4; h_1,h_2,h_3 *be non-negative. If* μ_2 *is supermodular. According to Lemma 3.1, we have* $(c)\int_A h_1+h_2\mathrm{d}\mu_2\geq(c)\int_A h_1\mathrm{d}\mu_2+(c)\int_A h_2\mathrm{d}\mu_2$. *Then* $\delta_{\pi_4}(\pi_1,\pi_2)+\delta_{\pi_4}(\pi_2,\pi_3)\gtrsim\delta_{\pi_4}(\pi_1,\pi_3)$ *is false.*

(2) Let $h_1=1$, $h_2=-1$, $h_3=0$. *Since* $(c)\int_A h_3\mathrm{d}\mu_2\leq(c)\int_A|h_1|\mathrm{d}\mu_2+(c)\int_A|h_2|\mathrm{d}\mu_2$. *Therefore,* $\delta_{\pi_4}(\pi_1,\pi_3)\gtrsim\delta_{\pi_4}(\pi_1,\pi_2)+\delta_{\pi_4}(\pi_2,\pi_3)$ *is false.*

Theorem 3.2. *If* $D_{f,\pi_3}(\pi_1,\pi_2)$ *is bounded, then* $D_{f,\pi_3}(\pi_1,\pi_2)$ *is a compact convex set.*

Proof. It follows from Definition 3.4,

$$D_{f,\pi_3}(\pi_1,\pi_2)=(C)\int\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_3}f(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}\pi_3.$$

First, suppose $x,y\in D_{f,\pi_3}(\pi_1,\pi_2)$, then there exists $\mu_1,\mu_2\in\pi_3$ such that $x=(c)\int\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_3}f(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}\mu_1$, $y=(c)\int\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_3}f(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}\mu_2$. For $\forall\ \alpha\in(0,1)$, we have

$$\alpha x+(1-\alpha)y=(c)\int\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_3}f(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}\alpha\mu_1+(c)\int\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_3}f(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}(1-\alpha)\mu_2.$$

Applying Proposition 2.1, we have

$$\alpha x+(1-\alpha)y=(c)\int\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_3}f(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}[\alpha\mu_1+(1-\alpha)\mu_2].$$

Since π_3 is a compact convex set-valued mapping, so there exists $\mu_3 \in \pi_3$ such that $\mu_3 = \alpha\mu_1 + (1-\alpha)\mu_2$. Therefore,

$$\alpha x + (1-\alpha)y = (c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3}) d\mu_3.$$

That is, $\alpha x + (1-\alpha)y \in D_{f,\pi_3}(\pi_1, \pi_2)$.

Second, let $\{x_n\} \in D_{f,\pi_3}(\pi_1, \pi_2)$, $n = 1, 2, 3, \cdots$, there is $\{\mu_n\} \in \pi_3$ such that $x_n = (c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3}) d\mu_n$, $n = 1, 2, 3, \cdots$. Obviously, $\{\mu_n\}$ are bounded and limitless. Since π_3 is a compact convex set-valued function, therefore, there exists convergence sub columns $\{\mu_{n_k}\} \subseteq \{\mu_n\}$, we could assume that increase converges to γ, then, there exists $\{x_{n_k}\} \subset x_n$ such that $\{x_{n_k} = (c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3}) d\mu_{n_k}\}$. Applying Proposition 2.1, we have

$$\lim_{k\to\infty} x_{n_k} = (c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3}) d\gamma.$$

Therefore, $D_{f,\pi_3}(\pi_1, \pi_2)$ is a compact set. That is, $D_{f,\pi_3}(\pi_1, \pi_2)$ is a compact convex set.

Theorem 3.3. *Let μ_1 be the minimum choice for π_3 and μ_2 be the maximum choice,*

$$a = \min\{(c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3}) d\mu_1, (c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3}) d\mu_2\},$$

$$b = \max\{(c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3}) d\mu_1,\ (c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3}) d\mu_2\}.$$

If $a < \infty$, then

$$D_{f,\pi_3}(\pi_1, \pi_2) \subseteq [a, b].$$

If $b < \infty$, then

$$D_{f,\pi_3}(\pi_1, \pi_2) = [a, b].$$

Proof. If $a < \infty$, then $D_{f,\pi_3}(\pi_1, \pi_2)$ non-empty. Suppose $c \in D_{f,\pi_3}(\pi_1, \pi_2)$, then there exists $\mu_3 \in \pi_3$ such that $c = (c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3}) d\mu_3$. Applying Proposition 2.1, we have

$$a \le (c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3}) d\mu_3 \le b.$$

Therefore, $c \in [a, b]$. That is,

$$D_{f,\pi_3}(\pi_1, \pi_2) \subseteq [a, b].$$

If $b < \infty$, then $b \in D_{f,\pi_3}(\pi_1, \pi_2)$, therefore

$$D_{f,\pi_3}(\pi_1, \pi_2) = [a, b].$$

It is difficult to directly calculate the r-n derivatives of monotone set-valued measures, but Theorem 3.3 provides a method for calculating monotone set-valued measures that conform to the conditions of Theorem 3.3. Thus, it provides a basis for the calculation of f-divergence of monotone set-valued measure.

Example 3.1. (see [32]) Let $m : R^+ \to R^+$ 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

$$\mu_m(A) = m(\lambda(A)),$$

for all A. μ_m is said to be a distorted Lebesgue measure. Obviously, μ_m is a non-additive measure.

Lemma 3.3. *(see [32]) Let $f(t)$ be a continuous and increasing function with $f(0) = 0$, μ_m be a distorted Lebesgue measure. Then there exists an increasing function g so that $f(t) = (c)\int_{[0,t]} g\mathrm{d}\mu_m$, and the following holds:*

$$G(s) = F(s)/sM(s);$$

$$g(t) = L^{-1}(F(s)/sM(s)).$$

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

Example 3.2. Let μ be non-additive measure on $[a, b]$, $a > 0$, $\mathcal{A}$ be Borel-σ algebra on $[a, b]$, $\pi_3 : A \to [0, \mu(A)]$, $\pi_1 : A \to [0, (c)\int_A t\mathrm{d}\mu]$, $\pi_2 : A \to [0,\ (c)\int_A (t + 1)\mathrm{d}\mu]$, $A \in \mathcal{A}$. By the definition of the Choquet integral, we have π_1, π_2, π_3 are monotone set-valued measure and $\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_3} = t$, $\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_3} = t + 1$. Applying Theorem 3.3, we have

(1) If $(c)\int_a^b (t + 1)f(\frac{t}{t+1})\mathrm{d}\mu \geq 0$, then

$$D_{f,\pi_3}(\pi_1, \pi_2) = (C)\int_a^b (t + 1)f(\frac{t}{t + 1})\mathrm{d}\pi = [0,\ (c)\int_a^b (t + 1)f(\frac{t}{t + 1})\mathrm{d}\mu].$$

If $(c)\int_a^b (t + 1)f(\frac{t}{t+1})\mathrm{d}\mu \leq 0$, then

$$D_{f,\pi_3}(\pi_1, \pi_2) = (C)\int_a^b (t + 1)f(\frac{t}{t + 1})\mathrm{d}\pi = [(c)\int_a^b (t + 1)f(\frac{t}{t + 1})\mathrm{d}\mu, 0].$$

(2) $\delta_{\pi_3}(\pi_1, \pi_2) = [0,\ \frac{1}{2}(c)\int |\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_3} - \frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_3}|\mathrm{d}\mu] = [0,\ \frac{1}{2}\mu([a, b])]$.

(3) $H_{\pi_3}(\pi_1, \pi_2) = \left[0,\ \sqrt{\frac{1}{2}(c)\int (\sqrt{\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_3}} - \sqrt{\frac{\mathrm{d}\pi_2}{\mathrm{d}\pi_3}})^2\mathrm{d}\mu}\right] = \left[0,\ \sqrt{\frac{1}{2}(c)\int (\sqrt{t} - \sqrt{t + 1})^2\mathrm{d}\mu}\right]$.

Example 3.3. Let $\pi_1 = [\mu_{m_1}, \mu_{n_1}]$, $\pi_2 = [\mu_{m_2}, \mu_{n_2}]$, $\pi_3 = [\mu_{m_3}, \mu_{n_3}]$. Let μ_{m_i}, μ_{n_i} respectively be distorted Lebesgue measure with $m_i(t)$, $n_i(t)$, $i = 1, 2, 3$. $m_1(t) = t$, $n_1(t) = \frac{3}{2}t$, $m_2(t) = 2t^2$, $n_2(t) = 3t^2$, $m_3(t) = t$, $n_3(t) = \frac{3}{2}t$. Obviously, $\frac{d\mu_{m_1}}{d\mu_{m_3}} = 1$, $\frac{d\mu_{n_1}}{d\mu_{n_3}} = 1$. Let $\frac{d\mu_{m_2}}{d\mu_{m_3}} = g_1(t)$, $\frac{d\mu_{n_2}}{d\mu_{n_3}} = g_2(t)$. Applying Lemma 3.3, we have

$$g_1(t) = L^{-1}(\frac{M_2(s)}{sM_3(s)}) = 4t,$$

$$g_2(t) = L^{-1}(\frac{N_2(s)}{sN_3(s)}) = 4t.$$

Here, $M_2(s)$, $M_3(s)$, $N_2(s)$, $N_3(s)$ respectively are Laplace transformation of m_2, m_3, n_2, n_3. Applying Theorem 3.1, we have $\frac{d\pi_1}{d\pi_3} = 1$, $\frac{d\pi_2}{d\pi_3} = 4t$. Therefore,

(1) $D_{f,\pi_3}(\pi_1, \pi_2) = (C)\int 4tf(1/4t)\mathrm{d}\pi_3$.

(2) $H_{\pi_3}(\pi_1, \pi_2) = \sqrt{\frac{1}{2}(C)\int(1-\sqrt{4t})^2 \mathrm{d}\pi_3}$.
(3) $\delta_{\pi_3}(\pi_1, \pi_2) = \frac{1}{2}(C)\int|1-4t|\mathrm{d}\pi_3$.

Example 3.4. Let $\pi_1 = [\mu_{m_1}, \mu_{n_1}]$, $\pi_2 = [\mu_{m_2}, \mu_{n_2}]$, $\pi_3 = [\mu_{m_3}, \mu_{n_3}]$. Let μ_{m_i}, μ_{n_i} respectively be distorted Lebesgue measure with $m_i(t),\ n_i(t), i = 1, 2, 3$. $m_1(t) = t,\ n_1(t) = \frac{3}{2}t,\ m_2(t) = \frac{1}{2}t^2, n_2(t) = \frac{3}{4}t^2, m_3(t) = t$, $n_3(t) = \frac{3}{2}t$. Obviously, $\frac{d\mu_{m_1}}{d\mu_{m_3}} = 1$, $\frac{d\mu_{n_1}}{d\mu_{n_3}} = 1$. Let $\frac{d\mu_{m_2}}{d\mu_{m_3}} = g_1(t)$, $\frac{d\mu_{n_2}}{d\mu_{n_3}} = g_2(t)$. Applying Lemma 3.3, we have

$$g_1(t) = L^{-1}(\frac{M_2(s)}{sM_3(s)}) = t,$$

$$g_2(t) = L^{-1}(\frac{N_2(s)}{sN_3(s)}) = t.$$

Here, $M_2(s),\ M_3(s),\ N_2(s),\ N_3(s)$ respectively are Laplace transformation of $m_2,\ m_3,\ n_2,\ n_3$. Applying Theorem 3.1, we have $\frac{d\pi_1}{d\pi_3} = 1$, $\frac{d\pi_2}{d\pi_3} = t$. Let$f(x) = x\log_2 x$, then

$$KL_{\pi_3}(\pi_1, \pi_2) = (C)\int \frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_3}\log(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}\pi_3.$$

Let $a = (c)\int \frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_3}\log(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}\mu_{m_3}$, $b = (c)\int \frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_3}\log(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}\mu_{n_3}$.

According to Proposition 2.1 and Choquet integral, we have

$$\begin{aligned} a &= (c)\int \frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_3}\log(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}\mu_{m_3} \\ &= (c)\int -\log_2 t\mathrm{d}\mu_{m_3} \\ &= \int_0^{+\infty} \mu_{m_3}(\{-\log_2 t > \alpha\})\mathrm{d}\alpha \\ &= \int_0^{+\infty} \mu_{m_3}(\{t < 2^{-\alpha}\})\mathrm{d}\alpha = \int_0^{+\infty} \lambda([0, 2^{-\alpha}])\mathrm{d}\alpha \\ &= \int_0^{+\infty} 2^{-\alpha}\mathrm{d}\alpha = \frac{1}{\ln 2}. \end{aligned}$$

Similarly, we have

$$b = \frac{3}{2\ln 2}.$$

Applying Theorem 3.3, we have

$$KL_{\pi_3}(\pi_1, \pi_2) = [\frac{1}{\ln 2}, \frac{3}{2\ln 2}].$$

3.3. A new version f-divergence of the monotone set-valued measures

Hamzeh Agahi found that the f-divergence for non-additive measures of Torra is not always non-negative. He solved this problem by replacing Torra's f-divergence with the maximum value of the f-divergence of the non-additive measure and its conjugate measure in reference [2]. There is also such a problem in the f-divergence of monotone set-valued measure. In the solution, since the conjugate measure of monotone set-valued measure has not been proposed. Therefore, the conjugate measure of monotone set-valued measure is defined by conjugate selection of monotone set-valued measure, and it is proved that the conjugate measure of monotone set-valued measure is also a monotone set-valued measure and has similar properties to non-additive measure. Finally, the f-divergence in Definition 3.4 is replaced by the maximum value of f-divergence of monotone set-valued measure and its conjugate measure.

The following Example 3.5 shows that f-divergence of in Definition 3.4 is not always non-negative.

Example 3.5. Let $\Omega = [0,1]$. Let $\pi_1 = [\mu_{m_1}, \mu_{n_1}]$, $\pi_2 = [\mu_{m_2}, \mu_{n_2}]$, $\pi_3 = [\mu_{m_3}, \mu_{n_3}]$. Let μ_{m_i} and μ_{n_i} respectively be distorted Lebesgue measure with $m_i(t)$, $n_i(t)$, $i = 1,2,3$. $m_1(t) = t$, $n_1(t) = \frac{3}{2}t$, $m_2(t) = 2t^2$, $n_2(t) = 3t^2$, $m_3(t) = t$, $n_3(t) = \frac{3}{2}t$. Applying Example 3.3, we have $\frac{d\pi_1}{d\pi_3} = 1$, $\frac{d\pi_2}{d\pi_3} = 4t$.

$$KL_{\pi_3}(\pi_1, \pi_2) = (C)\int \frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_3} \log(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}\pi_3.$$

Let $a = (c)\int \frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_3} \log(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}\mu_{m_3}$, $b = (c)\int \frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_3} \log(\frac{\mathrm{d}\pi_1/\mathrm{d}\pi_3}{\mathrm{d}\pi_2/\mathrm{d}\pi_3})\mathrm{d}\mu_{n_3}$.

According to Proposition 2.1 and Choquet integral, we have

$$\begin{aligned}
a &= (c)\int -\log_2 4t\mathrm{d}\mu_{m_3} = (c)\int(-2-\log_2 t)\mathrm{d}\mu_{m_3} \\
&= -2\mu_{m_3}([0,1]) + (c)\int -\log_2 t\mathrm{d}\mu_{m_3} = -2 + (c)\int -\log_2 t\mathrm{d}\mu_{m_3} \\
&= -2 + \int_0^{+\infty} \mu_{m_3}(\{-\log_2 t > \alpha\})\mathrm{d}\alpha \\
&= -2 + \int_0^{+\infty} \mu_{m_3}(\{t < 2^{-\alpha}\})\mathrm{d}\alpha = -2 + \int_0^{+\infty} \lambda([0, 2^{-\alpha}])\mathrm{d}\alpha \\
&= -2 + \int_0^{+\infty} 2^{-\alpha}\mathrm{d}\alpha = -2 + \frac{1}{\ln 2}.
\end{aligned}$$

Similarly, we have

$$b = \frac{3}{2}(-2 + \frac{1}{\ln 2}).$$

According to Theorem 3.3, we have

$$KL_{\pi_3}(\pi_1, \pi_2) = [\frac{3}{2}(-2 + \frac{1}{\ln 2}), -2 + \frac{1}{\ln 2}].$$

On the light of this example, we pose the question: Can we define a modified version of f-divergence for monotone set-valued measures which is always non-negative? In this section, we answer this question. First, we define the conjugate measure of the monotone set-valued measure and prove that the conjugate measure of the monotone set-valued measure is also a monotone set-valued measure.

Definition 3.5. *Let A be a set-valued, A is said to be non-negative, if* $A \gtrsim \{0\}$.

Definition 3.6. *Let* $(\Omega, \mathcal{A}, \pi)$ *be monotone set-valued measurable space. We say* $\bar{\pi}$ *is the conjugate of a monotone set-valued measure* π*, if*

$$\bar{\pi}(A) = \{\bar{\mu}(A) : \bar{\mu}(A) = \mu(\Omega) - \mu(\Omega/A), \mu \in \pi\}.$$

Theorem 3.4. *$\bar{\pi}$ is a monotone set-valued measure.*

Proof. We only need to prove the following three points.

(1) $\bar{\pi}(\phi) = \{\bar{\mu}(\phi) : \bar{\mu}(\phi) = \mu(\Omega) - \mu(\Omega - \phi) = 0, \mu \in \pi\} = \{0\}$.

(2) Let $A, B \in \mathcal{A}$, $\bar{\pi}(A) = \{\bar{\mu}(A) : \bar{\mu}(A) = \mu(\Omega) - \mu(\Omega/A), \mu \in \pi\}$, $\bar{\pi}(B) = \{\bar{\mu}(B) : \bar{\mu}(B) = \mu(\Omega) - \mu(\Omega/B), \mu \in \pi\}$. If $A \subseteq B$, suppose $x \in \bar{\pi}(A)$, then there exists a $\mu \in \pi$ such that $x = \bar{\mu}(A)$, we just have to take $y = \bar{\mu}(B)$, then $y \geq x$. Similarly, for any $y_1 \in \bar{\pi}(B)$, there exist a $\mu_1 \in \pi$ such that $y_1 = \bar{\mu_1}(B)$, we just have to take $x_1 = \bar{\mu_1}(A)$, then $y_1 \geq x_1$. Therefore, $\bar{\mu}(A) \lesssim \bar{\mu}(B)$.

(3) We need to proof that $\bar{\pi}$ is a compact convex set-value mapping. We just need to prove that $\bar{\pi}(A)$ is a compact convex set for all $A \in \mathcal{A}$.

Suppose $x, y \in \bar{\pi}(A)$, then there exists $\mu_1, \mu_2 \in \pi$ such that $x = \bar{\mu}_1(A)$, $y = \bar{\mu}_2(A)$. Let $\alpha \in [0, 1]$, we have

$$\begin{aligned}\alpha x + (1-\alpha)y &= \alpha\bar{\mu}_1(A) + (1-\alpha)\bar{\mu}_2(A)\\ &= \alpha\mu_1(\Omega) + (1-\alpha)\mu_2(\Omega) - (\alpha\mu_1(\Omega/A) + (1-\alpha)\mu_2(\Omega/A))\\ &= \alpha\mu_1(\Omega) + (1-\alpha)\mu_2(\Omega) - [(\alpha\mu_1(\Omega/A) - (1-\alpha)\mu_2(\Omega/A))].\end{aligned}$$

Since π is a compact convex set-value mapping. Therefore, there exist $\mu_3 \in \pi$ such that

$$\alpha\mu_1(\Omega) + (1-\alpha)\mu_2(\Omega) - [\alpha\mu_1(\Omega/A) + (1-\alpha)\mu_2(\Omega/A)] = \mu_3(\Omega) - \mu_3(\Omega/A).$$

Let $\bar{\mu_3}(A) = \mu_3(\Omega) - \mu_3(\Omega/A)$, then $\bar{\mu_3}(A) \in \bar{\pi}(A)$. Therefore, $\bar{\pi}(A)$ is a convex set.

Let $\{x_n\} \in \bar{\pi}(A)$, $(n = 1, 2, 3, ...)$, there is $\{\mu_n\} \in \pi$ such that $x_n = \mu_n(\Omega) - \mu_n(\Omega/A)$, $n = 1, 2, 3, \cdots$. Obviously, $\{\mu_n\}$ are bounded and limitless. Since π is a compact convex set-valued mapping. Therefore, there exists convergence sub columns $\{\mu_{n_k}\} \subseteq \{\mu_n\}$, we could assume that increase converges to μ, then

$$\lim_{k\to\infty} x_{n_k} = \lim_{k\to\infty}(\mu_{n_k}(\Omega) - \mu_{n_k}(\Omega/A)) = \mu(\Omega) - \mu(\Omega/A).$$

Since $\mu \in \pi$, therefore $\mu(\Omega) - \mu(\Omega/A) \in \bar{\pi}(A)$. That is, $\bar{\pi}(A)$ is a compact convex set. Therefore, $\bar{\pi}$ is a monotone set-valued measure.

Theorem 3.5. $\bar{\pi}(\Omega) = \pi(\Omega)$ *and* $\bar{\bar{\pi}}(A) = \pi(A)$.

Proof. (1) $\bar{\pi}(\Omega) = \{\bar{\mu}(\Omega) : \bar{\mu}(\Omega) = \mu(\Omega) - \mu(\Omega - \Omega) = \mu(\Omega), \mu \in \pi\} = \pi(\Omega)$.

(2)$\bar{\bar{\pi}}(A) = \{\bar{\bar{\mu}}(A) : \bar{\bar{\mu}}(A) = \bar{\mu}(\Omega) - \bar{\mu}(\Omega/A) = \mu(A), \bar{\mu} \in \bar{\pi}\} = \pi(A)$.

Definition 3.7. *Let $(\Omega, \mathcal{A}, \pi)$ be monotone set-valued measurable space, π_1 and π_2 be two monotone set-valued measure and π_1 exists Radon-Nikodym derivatives with respect to π_2, f be a convex function with $f(1) = 0$.*

(1) If f is a non-negative convex function, then the f-divergence between π_1 and π_2 is defined as

$$D_f^+(\pi_1, \pi_2) = (C)\int \frac{d\pi_2}{d\pi_1} f(\frac{d\pi_1}{d\pi_2})d\pi_1.$$

(2) If f is a real convex function, then the f-divergence between π_1 and π_2 is defined as

$$D_f^+(\pi_1, \pi_2) = \max\{(C)\int \frac{d\pi_2}{d\pi_1} f(\frac{d\pi_1}{d\pi_2})d\pi_1, (C)\int \frac{d\pi_2}{d\pi_1} f(\frac{d\pi_1}{d\pi_2})d\bar{\pi}_1\}.$$

Theorem 3.6. *Let $(\Omega, \mathcal{A}, \pi)$ be monotone set-valued measurable space.*
(1) If there exists a $\mu \in \pi$ such that $(c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3})d\mu \geq 0$, then $(C)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3})d\pi_3 \gtrsim \{0\}$.
(2) If there exists a $\mu \in \pi$ such that $(c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3})d\mu \leq 0$, then $(C)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3})d\pi_3 \lesssim \{0\}$.

Proof. Let μ_1 be the minimum choice and μ_2 be the maximum choice for π_3. Suppose there exists $\mu \in \pi_3$ such that $(c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3})d\mu \geq 0$. Let's suppose that $\mu' = \{0\}$. We have

$$0 = (c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3})d\mu' \leq (c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3})d\mu.$$

Since $\mu' \leq \mu_1 \leq \mu$. According to Proposition 2.1, we have

$$0 \leq (c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3})d\mu_1 \leq (c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3})d\mu.$$

Therefore,

$$(c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3})d\mu_1 \geq 0.$$

That is,

$$(C)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3})d\pi_3 \gtrsim \{0\}.$$

Similarly, if there exists a $\mu \in \pi$ such that $(c)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3})d\mu \leq 0$, then $(C)\int \frac{d\pi_2}{d\pi_3} f(\frac{d\pi_1/d\pi_3}{d\pi_2/d\pi_3})d\pi_3 \lesssim \{0\}$.

Theorem 3.7. $D_f^+(\pi_1, \pi_2) \gtrsim \{0\}$.

Proof. (1) If f is a non-negative convex function, we obviously have $D_f^+(\pi_1, \pi_2) \gtrsim \{0\}$.

(2) If f is a real convex function. According to Theorem 3.6, we just have to prove that there exists a $\mu \in \pi$ such that

$$\max\{(c)\int \frac{d\pi_2}{d\pi_1} f(\frac{d\pi_1}{d\pi_2})d\mu, (c)\int \frac{d\pi_2}{d\pi_1} f(\frac{d\pi_1}{d\pi_2})d\bar{\mu}\} \geq 0.$$

Let $\mu = \mu_1$ be the maximum choice of π_1, μ_2 be the maximum choice of π_2. Applying Theorem 3.1, we have $\frac{d\pi_2}{d\pi_1} = \frac{d\mu_2}{d\mu_1}$. Therefore,

$$\max\{(c)\int \frac{d\pi_2}{d\pi_1} f(\frac{d\pi_1}{d\pi_2})d\mu, (c)\int \frac{d\pi_2}{d\pi_1} f(\frac{d\pi_1}{d\pi_2})d\bar{\mu}\} = \max\{(c)\int \frac{d\mu_2}{d\mu_1} f(\frac{d\mu_1}{d\mu_2})d\mu_1, (c)\int \frac{d\mu_2}{d\mu_1} f(\frac{d\mu_1}{d\mu_2})d\bar{\mu}_1\}.$$

Applying Theorem 2.1,

$$\begin{aligned}
&\max\{(c)\int \frac{d\mu_2}{d\mu_1} f(\frac{d\mu_1}{d\mu_2})d\mu_1, (c)\int \frac{d\mu_2}{d\mu_1} f(\frac{d\mu_1}{d\mu_2})d\bar{\mu}_1\} \\
&\geq \mu_1(\Omega) f(\frac{1}{\mu_1(\Omega)}(c)\int \frac{d\mu_1}{d\mu_1}d\mu_1) \\
&= \mu_1(\Omega) f(\frac{1}{\mu_1(\Omega)}(\mu_1(\Omega)) = \mu_1(\Omega) f(1) = 0.
\end{aligned}$$

Therefore, $\max\{(C)\int \frac{d\pi_2}{d\pi_1} f(\frac{d\pi_1}{d\pi_2})d\pi_1, (C)\int \frac{d\pi_2}{d\pi_1} f(\frac{d\pi_1}{d\pi_2})d\bar{\pi}_1\} \gtrsim \{0\}$. That is, $D_f^+(\pi_1, \pi_2) \gtrsim \{0\}$.

3.4. A generalized f-divergence of the monotone set-valued measures

In the above problems, the monotone set-valued measures of Radon-Nikodym derivative is defined as a real value function. Although it can solve many problems, it seems too strict. Because under such a condition, many monotone set-valued measure of Radon-Nikodym derivative does not exist. Therefore, the Radon-Nikodym derivative defined above is extended to a set-valued function, and the set-valued function is used instead of the real valued function to express the Radon-Nikdoym derivative of two monotone set-valued measures.

Definition 3.8. *Let $(\Omega, \mathcal{A}, \pi)$ be monotone set-valued measurable space, F be a set-valued function, $A \in \mathcal{A}$, then, the Choquet integral of F with respect to π is defined by*

$$(\bar{C}) \int_A F \mathrm{d}\pi = \{(C) \int_A f \mathrm{d}\pi, f \in F\}.$$

Definition 3.9. *Let $(\Omega, \mathcal{A}, \pi)$ be monotone set-valued measurable space, π_1 and π_2 be two monotone set-valued measure. If there exists a closed interval value function $F : \Omega \to I(R^+)$ such that $\pi_1(A) = (\bar{C}) \int_A F \mathrm{d}\pi_2$, $A \in \mathcal{A}$, then, the interval function F is called the generalized Radon-Nikodym derivative of π_1 with respect to π_2, denoted $F = \mathrm{d}\pi_1/\mathrm{d}\pi_2$ or $F = \frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_2}$.*

Here, the generalized Randon Nikodym-derivative is not unique, for example: Let $\pi_1 = [0, \mu_1]$, $\pi_2 = [0, \mu_2]$. There exists a non-negative measurable function h such that $h = \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}$. Let $F_g = [g, h]$, here, g is a non-negative measure function and $g \leq h$. According to Definition 3.9, we have F_g is π_1 generalized Radon-Nikodym derivatives with respect to π_2. Since g is not the only one, so F_g too.

Definition 3.10. *Let $(\Omega, \mathcal{A}, \pi)$ be monotone set-valued measurable space, π_1 and π_2 be two monotone set-valued measure and π_1 exists generalized Radon-Nikodym derivatives with respect to π_2, f be a convex function with $f(1) = 0$. Then the generalized f-divergence between π_1 and π_2 is defined as*

$$D_f(\pi_1, \pi_2) = (\bar{C}) \int f \circ \frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_2} \mathrm{d}\pi_1.$$

Example 3.6. Let $\pi_1 = [0, \mu_1]$, $\pi_2 = [0, \mu_2]$. There exists a non-negative measurable function h such that $h = \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}$. $F_g = [g, h]$ is π_1 generalized Radon-Nikodym derivatives with respect to π_2. Then the generalized f-divergence between π_1 and π_2 is defined as

$$D_f(\pi_1, \pi_2) = (\bar{C}) \int f(\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_2}) \mathrm{d}\pi_1 = \bigcup_{h_i \in F} (C) \int f(h_i) \mathrm{d}\pi_1 = \bigcup_{h_i \in F} [0, (C) \int f(h_i) \mathrm{d}\mu_1] = [0, (C) \int f(h) \mathrm{d}\mu_1].$$

Obviously, F_g is not the only one, but the f-divergence between π_1 and π_2 is only one.

Theorem 3.8. *Let $(\Omega, \mathcal{A}, \pi)$ be monotone set-valued measurable space, π_1 and π_2 be two monotone set-valued measure, μ_i^S, μ_i^I be the maximum choice and the minimum choice of $\pi_i (i = 1, 2)$, respectively. If $\frac{\mathrm{d}\mu_1^S}{\mathrm{d}\mu_2^S} = f$, $\frac{\mathrm{d}\mu_1^I}{\mathrm{d}\mu_2^I} = g$ and $g \leq f$. Then, there is a set-valued function $F : \mathcal{A} \to [g, f]$ is the generalized Radon-Nikdoym derivative of π_1 with respect to π_2.*

Proof. We just need to prove $\pi_1(A) = (\bar{C}) \int_A F \mathrm{d}\pi_2$ for $\forall A \in \mathcal{A}$.

First, let $x \in (\bar{C})\int_A F\mathrm{d}\pi_2$. Since $\pi_1(A)$ is a convex set and $(c)\int_A g\mathrm{d}\mu_2^I \leq x \leq (c)\int_A f\mathrm{d}\mu_2^s$. Therefore, $x \in \pi_1(A)$, that is

$$\pi_1(A) \supset (\bar{C})\int_A F\mathrm{d}\pi_2.$$

Second, we need to prove $\pi_1(A) \subset (C)\int_A F\mathrm{d}\pi_2$.

If $(c)\int_A g\mathrm{d}\mu_2^S \geq (c)\int_A f\mathrm{d}\mu_2^I$, since $(c)\int_A g\mathrm{d}\mu_2^I = \mu_1^I(A)$, $(c)\int_A f\mathrm{d}\mu_2^S = \mu_1^S(A)$ and $(C)\int_A f\mathrm{d}\pi_2$ and $(C)\int_A g\mathrm{d}\pi_2$ are convex set. Therefore,

$$\pi_1(A) = (C)\int_A f\mathrm{d}\pi_2 \bigcup (C)\int_A g\mathrm{d}\pi_2.$$

That is,

$$\pi_1(A) \subset (C)\int_A F\mathrm{d}\pi_2.$$

Otherwise, there is $h_1 \in [g, f]$ such that

$$(c)\int_A g\mathrm{d}\mu_2^S \geq (c)\int_A h_1\mathrm{d}\mu_2^I.$$

If $(c)\int_A h_1\mathrm{d}\mu_2^S \geq (c)\int_A f\mathrm{d}\mu_2^I$. Then

$$\pi_1(A) = (C)\int_A f\mathrm{d}\pi_2 \bigcup (C)\int_A h_1\mathrm{d}\pi_2 \bigcup (C)\int_A g\mathrm{d}\pi_2.$$

Otherwise, there is $h_2 \in [h_1, f]$ such that

$$(c)\int_A h_1\mathrm{d}\mu_2^S \geq (c)\int_A h_2\mathrm{d}\mu_2^I.$$

If $(c)\int_A h_1\mathrm{d}\mu_2^S \geq (c)\int_A h_1\mathrm{d}\mu_2^I$. Then

$$\pi_1(A) = (C)\int_A f\mathrm{d}\pi_2 \bigcup (C)\int_A h_1\mathrm{d}\pi_2 \bigcup (C)\int_A g\mathrm{d}\pi_2.$$

Otherwise, there is $h_3 \in [h_2, f]$. Repeating the above procedure, we can obtain a column of functions $\{h_k\}$, $k = 1, 2, ...$ such that

$$\pi_1(A) = \bigcup_{h_k} (C)\int_A h_k\mathrm{d}\pi_2 \bigcup (C)\int_A f\mathrm{d}\pi_2 \bigcup (C)\int_A g\mathrm{d}\pi_2.$$

That is,

$$\pi_1(A) \subset (\bar{C})\int_A F\mathrm{d}\pi_2.$$

Therefore, $F = [g, f]$ is the generalized Radon-Nikdoym derivative of π_1 with respect to π_2.

Example 3.7. Let $\pi_1 = [\mu_{m_1}, \mu_{n_1}]$, $\pi_2 = [\mu_{m_2}, \mu_{n_2}]$, μ_{m_i}, μ_{n_i} respectively be distorted Lebesgue measure with $m_i(t)$, $n_i(t)$, $i = 1, 2$. $m_1(t) = \frac{1}{2}t^2$, $n_1(t) = \frac{3}{4}t^2$, $m_2(t) = t$, $n_2(t) = 3t$. Let $\frac{d\mu_{m_2}}{d\mu_{m_3}} = g_1(t)$, $\frac{d\mu_{n_2}}{d\mu_{n_3}} = g_2(t)$. According to Lemma 3.3, we have

$$g_1(t) = L^{-1}\left(\frac{M_1(s)}{sM_2(s)}\right) = t,$$

$$g_2(t) = L^{-1}(\frac{N_1(s)}{sN_2(s)}) = 2t.$$

Here, $M_1(s)$, $M_2(s)$, $N_1(s)$, $N_2(s)$ respectively are Laplace transformation of m_1, m_2, n_1, n_2. According to Theorem 3.8, we have $\frac{d\pi_2}{d\pi_3} = [t, 2t]$.

Theorem 3.9. *Let $(\Omega, \mathcal{A}, \pi)$ be monotone set-valued measurable space, π_1 and π_2 be two monotone set-valued measure and exists a closed interval value function $F : \mathcal{A} \to I(R^+)$ such that $F = \frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_2}$, μ_1 and μ_2 be the minimum choice and the maximum choice of π_2 respectively. If $(c)\int \underline{f\circ F}\mathrm{d}\mu_1 \le (c)\int \overline{f\circ F}\mathrm{d}\mu_2$, then*

$$D_{f,\pi_3}(\pi_1,\pi_2) = [(c)\int \underline{f\circ F}\mathrm{d}\mu_1, (c)\int \overline{f\circ F}\mathrm{d}\mu_2].$$

Proof. According to Definition 3.10, we have $D_f(\pi_1,\pi_2) = (\bar{C})\int f(\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_2})\mathrm{d}\pi_1$, we just need to prove $(\bar{C})\int f(\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_2})\mathrm{d}\pi_1 = [(c)\int \underline{f\circ F}\mathrm{d}\mu_1, (c)\int \overline{f\circ F}\mathrm{d}\mu_2]$.

Suppose $m \in D_f(\pi_1,\pi_2)$, there exists $h_m \in F$, $\mu_m \in \pi_1$ such that $m = (c)\int f\circ h_m \mathrm{d}\mu_m$. According to Proposition 2.1, we have

$$(c)\int \underline{f\circ F}\mathrm{d}\mu_1 \le m \le (c)\int \overline{f\circ F}\mathrm{d}\mu_2.$$

That is,

$$D_f(\pi_1,\pi_2) \subset [(c)\int \underline{f\circ F}\mathrm{d}\mu_1, (c)\int \overline{f\circ F}\mathrm{d}\mu_2].$$

We just need to prove

$$D_f(\pi_1,\pi_2) \supset [(c)\int \underline{f\circ F}\mathrm{d}\mu_1, (c)\int \overline{f\circ F}\mathrm{d}\mu_2].$$

If $(c)\int \underline{f\circ F}\mathrm{d}\mu_2 \ge (c)\int \overline{f\circ F}\mathrm{d}\mu_1$, then

$$D_f(\pi_1,\pi_2) \supset [(c)\int \underline{f\circ F}\mathrm{d}\mu_1, (c)\int \overline{f\circ F}\mathrm{d}\mu_2].$$

Otherwise, there is $h_1 \in [\underline{f\circ F}, \overline{f\circ F}]$ such that

$$(c)\int_A \underline{f\circ F}\mathrm{d}\mu_2 \ge (c)\int_A h_1\mathrm{d}\mu_1.$$

If $(c)\int_A h_1\mathrm{d}\mu_2 \ge (c)\int_A \overline{f\circ F}\mathrm{d}\mu_1$. Then

$$[(c)\int \underline{f\circ F}\mathrm{d}\mu_1, (c)\int \overline{f\circ F}\mathrm{d}\mu_2] = (C)\int_A \underline{f\circ F}\mathrm{d}\pi_2 \bigcup (C)\int_A h_1\mathrm{d}\pi_2 \bigcup (C)\int_A \overline{f\circ F}\mathrm{d}\pi_2.$$

Otherwise, there is $h_2 \in [h_1, \overline{f\circ F}]$ such that

$$(c)\int_A h_2\mathrm{d}\mu_2 \ge (c)\int_A h_1\mathrm{d}\mu_1.$$

If $(c)\int_A h_1 \mathrm{d}\mu_2 \geq (c)\int_A h_1 \mathrm{d}\mu_1$. Then

$$\begin{aligned}&[(c)\int \underline{f\circ F}\mathrm{d}\mu_1,(c)\int \overline{f\circ F}\mathrm{d}\mu_2]\\ =&(C)\int_A \underline{f\circ F}\mathrm{d}\pi_2\bigcup(C)\int_A h_1\mathrm{d}\pi_2\bigcup(C)\int_A h_2\mathrm{d}\pi_2\bigcup(C)\int_A \overline{f\circ F}\mathrm{d}\pi_2.\end{aligned}$$

Otherwise, there is $h_3 \in [h_2, \overline{f\circ F}]$. Repeating the above procedure, we can obtain a column of functions $\{h_k\}, (k = 1, 2, \cdots)$ such that

$$[(c)\int \underline{f\circ F}\mathrm{d}\mu_1,(c)\int \overline{f\circ F}\mathrm{d}\mu_2] = (C)\int_A \underline{f\circ F}\mathrm{d}\pi_2 \bigcup_{h_k\in[\underline{f\circ F},\overline{f\circ F}]}(C)\int_A h_k\mathrm{d}\pi_2\bigcup(C)\int_A \overline{f\circ F}\mathrm{d}\pi_2.$$

Therefore,

$$D_f(\pi_1,\pi_2)\supset[(c)\int \underline{f\circ F}\mathrm{d}\mu_1,(c)\int \overline{f\circ F}\mathrm{d}\mu_2].$$

That is,

$$D_f(\pi_1,\pi_2)=[(c)\int \underline{f\circ F}\mathrm{d}\mu_1,(c)\int \overline{f\circ F}\mathrm{d}\mu_2].$$

According to Theorem 3.9, we can obtain the following Theorem 3.10.

Theorem 3.10. *Let $(\Omega, \mathcal{A}, \pi)$ be monotone set-valued measurable space, π_1 and π_2 be two monotone set-valued measure and exists a closed interval value function $F : \mathcal{A} \to I(R^+)$ such that $F = \frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_2}$, μ_1 and μ_2 be the minimum choice and the maximum choice of π_2 respectively.*

If f is monotone increasing on Ω, then

$$D_f(\pi_1,\pi_2)=[(c)\int f\circ\underline{F}\mathrm{d}\mu_1,(c)\int f\circ\overline{F}\mathrm{d}\mu_2].$$

If f is monotone decreasing on Ω, then

$$D_f(\pi_1,\pi_2)=[(c)\int f\circ\overline{F}\mathrm{d}\mu_1,(c)\int f\circ\underline{F}\mathrm{d}\mu_2].$$

Example 3.8. Let m be Lebesgue measure on $[0, 1]$, $\mathcal{A}$ be Borel-σ algebra on $[0, 1]$. $m_1 : A \to \int_A t\mathrm{d}m$, $m_2 : A \to \int_A (t + 1)\mathrm{d}m$, $A \in \mathcal{A}$. According to the property of Lebesgue integral, m_1, m_2 are additive measure. Let $\pi_2(A) = \{m(A)\}$, $\pi_1(A) = [m_1(A), m_2(A)]$. It is not difficult to verify that π_1, π_2 are monotone set-valued measures. Since $\frac{\mathrm{d}m_1}{\mathrm{d}m} = t$, $\frac{\mathrm{d}m_2}{\mathrm{d}m} = t + 1$ and $t + 1 > t$ on $[0, t]$. Therefore, the generalized Radon-Nikodym derivative of π_1 with respect to π_2 is $F(t) = [t, t + 1]$. Then, the f-divergence of π_1 between and π_2 is

$$D_f(\pi_1,\pi_2)=[\int_0^1 \underline{f\circ F}\mathrm{d}m,\ \int_0^1 \overline{f\circ F}\mathrm{d}m].$$

Let $f(x) = (1 - \sqrt{x})^2$. Since $f(x)$ is monotone decrease on $[0, 1]$. Applying Theorem 3.10, we have

$$\begin{aligned} H(\pi_1, \pi_2) &= [\int_0^1 f \circ \overline{F} \mathrm{d}m \,, \int_0^1 f \circ \underline{F} \mathrm{d}m] \\ &= \Big[\int_0^1 (1 - \sqrt{t+1})^2 \mathrm{d}m \,, \int_0^1 (1 - \sqrt{t})^2 \mathrm{d}m \Big] \\ &= \Big[\int_0^1 (1 - \sqrt{t+1})^2 \mathrm{d}t \,, \int_0^1 (1 - \sqrt{t})^2 \mathrm{d}t \Big] \\ &= [\frac{23}{6} - \frac{8\sqrt{2}}{3}, \frac{7}{6}]. \end{aligned}$$

Let $f(x) = |x - 1|$. Since $f(x)$ is monotone decrease on $[0, 1]$. Applying Theorem 3.10, we have

$$\delta(\pi_1, \pi_2) = [\int_0^1 t \mathrm{d}m \,, \int_0^1 (1 - t) \mathrm{d}m] = \{0.5\}.$$

Let $f(x) = x \log_2 x$. Since $f(x)$ is monotone increasing on $[0, 1]$. Applying Theorem 3.10, we have

$$KL(\pi_1, \pi_2) = [(c) \int_0^1 t \log_2 t \mathrm{d}m, (c) \int_0^1 (t+1) \log_2(t+1) \mathrm{d}m].$$

In this section, compared with the generalized f-divergence defined in Definition 3.10 and that in Definitions 3.4 and 3.7, the range of Radon-Nikdoym derivative of two monotone set-valued measures is extended, thus the range of adaptation of f-divergence is extended.

4. Conclusions

In this paper, we have defined and discussed the f-divergence for monotone set-valued measure. Some basic properties of this divergence have been studied. Meanwhile, some examples have been given to illustrate the effectiveness of the definitions and results proposed. We have discussed the non-negativity of f-divergence for monotone set-valued. We have also proposed an example which shows that the f-divergence is not always non-negative in general. Then the modified versions of f-divergence have been introduced and we have proved it is always non-negatives. Finally, we define the generalized f-divergence by using the generalized Radon-Nikdoym derivatives of two monotone set-valued measures.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (12061607).

Conflict of interest

The authors declare that they have no conflicts of interest.

References

1. H. Agahi, A modified Kullback-Leibler divergence for non-additive measures based on Choquet integral, *Fuzzy Sets Syst.*, **367** (2019), 107–117. https://doi.org/10.1016/j.fss.2019.01.021
2. H. Agahi, Fundamental properties of relative entropy and Lin divergence for Choquet integral, *Int. J. Approx. Reason.*, **134** (2021) 15–22. https://doi.org/10.1016/j.ijar.2021.03.009
3. S. Amari, α-Divergence is Unique, belonging to both f-divergence and Bregman divergence classes, *IEEE Trans. Inform. Theory*, **55** (2009), 4925–4931. https://doi.org/10.1109/TIT.2009.2030485
4. R. J. Aumann, Integrals of set-valued functions, *J. Math. Anal. Appl.*, **12** (1965), 1–12.
5. R. Beran, Minimum hellinger distance estimates for parametric models, *Ann. Stat.*, **5** (1997), 445–463.
6. J. Burbea, C. Rao, On the convexity of some divergence measures based on entropy functions, *IEEE Trans. Inform. Theory*, **28** (1982), 489–495. https://doi.org/10.1109/TIT.1982.1056497
7. H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Stat.*, **23** (1952), 493–507.
8. G. Choquet, Theory of capacities, *Ann. Inst. Fourier*, **5** (1954), 131–295. https://doi.org/10.5802/aif.53
9. D. A. Cieslak, T. R. Hoens, N. V. Chawla, W. P. Kegelmeyer, Hellinger distance decision trees are robust and skew-insensitive, *Data Min. Knowl. Disc.*, **24** (2012), 136–158. https://doi.org/10.1007/s10618-011-0222-1
10. I. Csiszár, Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.*, **2** (1967), 299–318.
11. I. Csiszár, Information measures: A critical survey, In: *Transactions of the seventh Prague conference on information theory, statistical decision functions, Random processes*, 1974, 73–86.
12. S. S. Dragomir, V. Glušččević, C. E. M. Pearce, Csiszár f-divergence, Ostrowski's inequality and mutual information, *Nonlinear Anal.*, **47** (2001), 2375–2386. https://doi.org/10.1016/S0362-546X(01)00361-3
13. A. C. Gavriluţ, Properties of regularity for multisubmeasures, *An. Ştiinţ. Univ. Al. I. Cuza Iaşi. Mat.*, 2004, 373–292.
14. A. C. Gavriluţ, Non-atomicity and the Darboux property for fuzzy and non-fuzzy Borel/Baire multivalued set functions, *Fuzzy Sets Syst.*, **160** (2009), 1308–1317. https://doi.org/10.1016/j.fss.2008.06.009
15. A. C. Gavriluţ, Rugularlity and autocontinuity of set multifunctions, *Fuzzy Sets Syst.*, **161** (2010), 681–693. https://doi.org/10.1016/j.fss.2009.05.007
16. A. C. Gavriluţ, Abstract rugular null-null-additive set multifunctions in Hausdorff topology, *An. Ştiinţ. Univ. Al. I. Cuza Iaşi. Mat.*, **59** (2013), 129–147.
17. A. C. Gavriluţ, Alexandroff theorem in Hausdorff topology for null-null-additive set multifunctions, *An. Ştiinţ. Univ. Al. I. Cuza Iaşi. Mat.*, 2013, 237–251. https://doi.org/10.2478/v10157-012-0046-3

18. A. C. Gavriluţ, A. Croitoru, Pseudo-atoms and Darboux property for set multifunctions, *Fuzzy Sets Syst.*, **161** (2010), 2897–2908. https://doi.org/10.1016/j.fss.2010.06.007

19. Z. T. Gong, X. Y. Kou, T. Xie, Interval-valued Choquet integral for set-valued mappings: Definitions, integral representations and primitive characteristics, *AIMS Math.*, **5** (2020), 6277–6297. https://doi.org/10.3934/math.2020404

20. T. Grbić, I. Štajner-Papuga, L. Nedović, Pseudo-integral of set-valued functions, In: *New dimensions in fuzzy logic and related technologies*, **1** (2007), 221–225.

21. E. Hellinger, Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen, *J. Reine Angew. Math.*, **1909** (1909), 210–271. https://doi.org/10.1515/crll.1909.136.210

22. K. C. Jain, A. Srivastava, On symmetric information divergence measures of Csiszar's f-divergence class, *J. Appl. Math. Stat. Inform.*, **3** (2007), 85–102.

23. R. Kadian, S. Kumar, Jensen-Renyi's-Tsallis fuzzy divergence information measure with its applications, *Commun. Math. Stat.*, 2021, 1–32. https://doi.org/10.1007/s40304-020-00228-1

24. S. Kullback, R. A. Leibler, On information and sufficiency, *Ann. Math. Stat.*, **22** (1951), 79–86.

25. F. Liese, I. Vajda, On divergences and informations in statistics and information theory, *IEEE Trans. Inform. Theory*, **52** (2006), 4394–4412. https://doi.org/10.1109/TIT.2006.881731

26. E. Nikita, P. Nikitas, Measures of divergence for binary data used in biodistance studies, *Archaeol. Anthropol. Sci.*, **13** (2021), 1–14. https://doi.org/10.1007/s12520-021-01292-6

27. K. Pearson, On the criterion that a given system of deviations from the probable in the case of correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, **50** (1900), 157–175. https://doi.org/10.1080/14786440009463897

28. V. V. Prelov, On the maximum values of f-divergence and Rényi divergence under a given variational distance, *Prob. Inform. Transm.*, **56** (2020), 1–12. https://doi.org/10.1134/S0032946020010019

29. N. Shlomo, L. Antal, M. Elliot, Measuring disclosure risk and data utility for flexible table generators, *J. Off. Stast.*, **31** (2015), 305–324. https://doi.org/10.1515/JOS-2015-0019

30. M. Stojaković, Set valued probability and its connection with set valued measure, *Stat. Probab. Lett.*, **82** (2012), 1043–1048. https://doi.org/10.1016/j.spl.2012.02.021

31. M. Sugeno, *Theory of fuzzy integral and its applications*, Tokyo Institute of Thconology, 1974.

32. M. Sugeno, A note on derivatives of functions with respect to fuzzy measures, *Fuzzy Sets Syst.*, **222** (2013), 1–17. https://doi.org/10.1016/j.fss.2012.11.003

33. N. X. Thao, Some new entropies and divergence measures of intuitionistic fuzzy sets based on Archimedean t-conorm and application in supplier selection, *Soft Comput.*, **25** (2021), 5791–5805. https://doi.org/10.1007/s00500-021-05575-x

34. F. Topsoe, Some inequalities for information divergence and related measures of discrimination, *IEEE Trans. Inform. Theory*, **46** (2000), 1602–1609. https://doi.org/10.1109/18.850703

35. V. Torra, Y. Narukawa, M. Sugeno, On the f-divergence for non-additive measures, *Fuzzy Sets Syst.*, **292** (2016), 364–379. https://doi.org/10.1016/j.fss.2015.07.006

36. V. Torra, Y. Narukawa, M. Sugeno, On the f-divergence for discrete non-additive measures, *Inform. Sci.*, **512** (2020), 50–63. https://doi.org/10.1016/j.ins.2019.09.033

37. A. Umar, R. N. Saraswat, Decision-making in machine learning using novel picture fuzzy divergence measure, *Neural Comput. Appl.*, **34** (2022), 457–475. https://doi.org/10.1007/s00521-021-06353-4

38. J. R. Wu, X. W. Kai, J. J. Li, Atoms of monotone set-valued measures and integrals, *Fuzzy Sets Syst.*, **304** (2016), 131–139. https://doi.org/10.1016/j.fss.2016.05.006

39. D. L. Zhang, C. M. Guo, Fuzzy integrals of set-valued mappings and fuzzy mappings, *Fuzzy Sets Syst.*, **75** (1995), 103–109. https://doi.org/10.1016/0165-0114(94)00342-5

40. D. L. Zhang, C. M. Guo, D. Y. Liu, Set-valued Choquet integrals revisited, *Fuzzy Sets Syst.*, **147** (2004), 475–485. https://doi.org/10.1016/j.fss.2004.04.005

41. K. Y. Zhao, R. T. Sun, L. Li, M. M. Hou, G. Yuan, R. Z. Sun, An optimal evidential data fusion algorithm based on the new divergence measure of basic probability assignment, *Soft Comput.*, **25** (2021), 11449–11457. https://doi.org/10.1007/s00500-021-06040-5