# Kullback-Leibler Divergence and Mutual Information of Experiments in the Fuzzy Case

## Dagmar Markechová

Department of Mathematics, Faculty of Natural Sciences, Constantine the Philosopher University in Nitra, A. Hlinku 1, SK-949 01 Nitra, Slovakia; dmarkechova@ukf.sk

**Abstract:** The main aim of this contribution is to define the notions of Kullback-Leibler divergence and conditional mutual information in fuzzy probability spaces and to derive the basic properties of the suggested measures. In particular, chain rules for mutual information of fuzzy partitions and for Kullback-Leibler divergence with respect to fuzzy P-measures are established. In addition, a convexity of Kullback-Leibler divergence and mutual information with respect to fuzzy P-measures is studied.

**Keywords:** fuzzy probability space; fuzzy partition; Shannon's entropy; Kullback-Leibler divergence; mutual information; conditional mutual information

---

## 1. Introduction

The notion of $\sigma-$algebra $S$ of random events and the concept of probability space $(\Omega,\ S,\ P)$ are a basis of the classical Kolmogorov probability theory [1]. A probability $P$ is a normalized measure defined on the $\sigma-$algebra $S$. The event in classical probability theory is understood as an exactly and clearly defined phenomenon and, from a mathematical point of view, it is a classical, ordinary set. Consider a finite measurable partition $A$ of $(\Omega,\ S,\ P)$ with probabilities $p_1, \ldots, p_n$ of the corresponding elements of $A$. We recall that the Shannon entropy [2] of $A$ is the number $H(A) = -\sum_{i=1}^{n} F(p_i)$, where the function $F : [0,\ \infty) \to \Re$ is defined by $F(x) = -x \log x$ if $x > 0$ and $F(0) = 0$. In [3], we have generalized this notion to situations when the considered probability space is a fuzzy probability space $(\Omega,\ M,\ \mu)$ defined by Piasecki [4]. Instead of probability $P$ it is considered a fuzzy P-measure $\mu$ defined on a fuzzy $\sigma-$algebra $M$ of fuzzy subsets of $\Omega$. Recall that by a fuzzy subset $f$ of a non-empty set $\Omega$ we mean a mapping $f : \Omega \to [0,\ 1]$ (Zadeh [5]). A fuzzy subset from the fuzzy $\sigma-$algebra $M$ is interpreted as a fuzzy event; the value $\mu(f)$ is interpreted as a probability of fuzzy event $f$. The structure $(\Omega,\ M,\ \mu)$ can serve as an alternative mathematical model of probability theory for the case where the observed events are described unclearly, vaguely, so-called fuzzy events. In [6], the mutual information of fuzzy partitions of a given fuzzy probability space $(\Omega,\ M,\ \mu)$ has been defined. It was shown that the entropy of fuzzy partitions introduced and studied by the author in [3] (see also [7]) can be considered as a special case of their mutual information. The proposed measures are fuzzy analogies of the relevant terms of the classical theory and they can be used whenever it is necessary to know the quantity of information received by the realization of experiments whose results are fuzzy events. Note that in [3] (see also [8]) using the concept of entropy of fuzzy partitions we define the entropy of the fuzzy dynamical system $(\Omega,\ M,\ \mu,\ U)$ (where $(\Omega,\ M,\ \mu)$ is a fuzzy probability space and $U : M \to M$ is a measure $\mu$ preserving $\sigma-$homomorphism). Analogies of our results for the case of logical entropy (see, e.g., [9]) are provided in our recently published papers [10,11]. Recall that the logical entropy of the probability distribution $P = (p_1, \ldots, p_n) \in \Re^n$ is defined as the number $h(P) = \sum_{i=1}^{n} p_i(1 - p_i)$. In [9] the author deals with historical aspects of the logical entropy formula $h(P)$ and investigates

the relationship between the logical entropy and Shannon's entropy. It should be noted that other fuzzy analogies of entropy are presented in [12–25]. It is known that there are many possibilities for defining operations with fuzzy sets; an overview can be found in [26]. While our approach is based on Zadeh's connectives [5], the authors of the cited papers used other connectives to define the fuzzy set operations.

In classical information theory [27] the mutual information is a special case of a more general quantity called Kullback-Leibler divergence, which was originally introduced by Kullback and Leibler in 1951 [28] as the divergence between two probability distributions. It is discussed in Kullback's historic text [29]. The Kullback-Leibler divergence is also called by many other different names, as K-L divergence, whether relative entropy, or information gain. It plays an important role, as a mathematical tool, in the stability analysis of master equations [30] and Fokker-Planck equations [31], and in isothermal equilibrium fluctuations and transient nonequilibrium deviations [32] (see also [31,33]).

The main aim of this contribution is to define, using our previous results on this issue, the notion of Kullback-Leibler divergence and conditional mutual information in fuzzy probability spaces and to study properties of the suggested measures. The paper is organized as follows. In the next section, we give the basic definitions and some known facts used in this paper. Our results are presented in Sections 3 and 4. In Section 3 we extend our study concerning the mutual information of fuzzy partitions. The notion of conditional mutual information of fuzzy partitions is introduced and subsequently chain rules for mutual information of fuzzy partitions are established. We also derive some more properties of this measure, e.g., data processing inequality. In Section 4 we define the Kullback-Leibler divergence and the conditional Kullback-Leibler divergence in fuzzy probability spaces. The basic properties of these measures are proved. Our results are summarized in Section 5.

## 2. Basic Definitions and Facts

We start by recalling some definitions and some known results which will be used in this contribution.

A fuzzy measurable space (Dvurečenskij [34]) is a couple $(\Omega, M)$, where $\Omega$ is a non-empty set, and $M$ is a fuzzy $\sigma-$algebra of fuzzy subsets of $\Omega$, i.e., $M \subset [0, 1]^{\Omega}$ containing $1_{\Omega}$, excluding $(1/2)_{\Omega}$, closed under the operation $\perp : f \mapsto 1_{\Omega} - f$ (i.e., if $f \in M$, then $f^{\perp} := 1_{\Omega} - f \in M$) and countable supremums (i.e., satisfying the implication if $f_n \in M$, $n = 1, 2, \ldots$, then $\cup_{n=1}^{\infty} f_n := \sup_n f_n \in M$). A fuzzy probability space (Piasecki [4]) is a triplet $(\Omega, M, \mu)$, where $(\Omega, M)$ is a fuzzy measurable space and the mapping $\mu : M \to [0, \infty)$ satisfies the following two conditions: (i) $\mu(f \cup f^{\perp}) = 1$ for all $f \in M$; (ii) if $\{f_n\}_{n=1}^{\infty}$ is a sequence of pairwise W-separated fuzzy subsets from $M$ (i.e., $f_i \leq f_j^{\perp}$ (point wisely) whenever $i \neq j$), then $\mu(\cup_{n=1}^{\infty} f_n) = \sum_{n=1}^{\infty} \mu(f_n)$.

The symbols $\cup_{n=1}^{\infty} f_n = \sup_n f_n$ and $\cap_{n=1}^{\infty} f_n := \inf_n f_n$ denote the fuzzy union and the fuzzy intersection of a sequence $\{f_n\}_{n=1}^{\infty} \subset M$, respectively, in the sense of Zadeh [5]. The complement of fuzzy subset $f$ of $\Omega$ is a fuzzy set $f^{\perp}$ defined by $f^{\perp}(\omega) = 1 - f(\omega)$, $\omega \in \Omega$. The following notions were defined by Piasecki in [35]. A fuzzy set $f \in M$ such that $f \geq f^{\perp}$ is called a W-universum; a fuzzy set $f \in M$ such that $f \leq f^{\perp}$ is called a W-empty fuzzy set. It can be proved that a fuzzy event $f \in M$ is a W-universum if and only if there exists a fuzzy event $g \in M$ such that $f = g \cup g^{\perp}$. A W-universum is interpreted as a certain event and a W-empty set as an impossible event. W-separated fuzzy events are interpreted as mutually exclusive events. Each mapping $\mu : M \to [0, \infty)$ having the properties (i) and (ii) is called, in the terminology of Piasecki, a fuzzy P-measure. Any fuzzy P-measure has all properties analogous to properties of a classical probability measure; the proofs and more details can be found in [4]. The monotonicity of fuzzy P-measure $\mu$ implies that this measure transforms $M$ into the interval $[0, 1]$. In the following we will use the following property of fuzzy P-measure $\mu$.

(2.1) Let $g \in M$. Then $\mu(f \cap g) = \mu(f)$ for all $f \in M$ if and only if $\mu(g) = 1$.

Let $g \in M$ such that $\mu(g) > 0$. Then the mapping $\mu(\cdot/g) : M \to [0, 1]$ defined by the formula

$$\mu(f/g) = \frac{\mu(f \cap g)}{\mu(g)}, \; f \in M,$$

is a fuzzy P-measure on $M$; it is called a conditional probability.

**Definition 1** [36]. *By a fuzzy partition (of a space* $(\Omega, M, \mu)$*) we understand a finite collection* $\xi = \{f_1, \ldots, f_n\}$ *of pairwise W-separated fuzzy subsets from M such that* $\mu(\cup_{i=1}^n f_i) = 1$.

Every fuzzy partition $\xi = \{f_1, \ldots, f_n\}$ of $(\Omega, M, \mu)$ represents in the sense of the classical probability theory a random experiment with a finite number of outcomes $f_i$, $i = 1, 2, \ldots, n$ (which are fuzzy events) with a probability distribution $p_i = \mu(f_i)$, $i = 1, 2, \ldots, n$, since $p_i \geq 0$ for $i = 1, 2, \ldots, n$, and $\sum_{i=1}^n p_i = \sum_{i=1}^n \mu(f_i) = \mu(\cup_{i=1}^n f_i) = 1$. For that reason, we have defined in [3] the entropy of $\xi = \{f_1, \ldots, f_n\}$ by Shannon's formula:

$$H_\mu(\xi) = -\sum_{i=1}^n F(\mu(f_i)), \tag{1}$$

where:

$$F : [0, \infty) \to \Re, \ F(x) = \begin{cases} x \log x, & \text{if } x > 0; \\ 0, & \text{if } x = 0. \end{cases}$$

In accordance with the classical theory the log is to the base 2 and entropy is expressed in bits.

**Definition 2** [6]. *Let* $\xi = \{f_1, \ldots, f_n\}$ *and* $\eta = \{g_1, \ldots, g_m\}$ *be two fuzzy partitions of a fuzzy probability space* $(\Omega, M, \mu)$*. A conditional entropy of* $\eta$ *given a fuzzy event* $f_i \in \xi$ *is defined by:*

$$H_\mu(\eta/f_i) = -\sum_{j=1}^m F(\dot{\mu}(g_j/f_i)),$$

*where:*

$$\dot{\mu}(g_j/f_i) = \begin{cases} \mu(g_j/f_i), & \text{if } \mu(f_i) > 0; \\ 0, & \text{if } \mu(f_i) = 0. \end{cases}$$

*A conditional entropy of* $\eta$ *assuming a realization of the experiment* $\xi$ *is defined by the formula:*

$$H_\mu(\eta/\xi) = -\sum_{i=1}^n \mu(f_i) \cdot H_\mu(\eta/f_i) = -\sum_{i=1}^n \sum_{j=1}^m \mu(f_i) \cdot F(\dot{\mu}(g_j/f_i)). \tag{2}$$

In the following, we will use the convention (based on continuity arguments) that $x \log \frac{x}{0} = \infty$ if $x > 0$, and $0 \log \frac{0}{x} = 0$ if $x \geq 0$. It is easy to see that we can rewrite Equation (2) in the following way:

$$H_\mu(\eta/\xi) = -\sum_{i=1}^n \sum_{j=1}^m \mu(f_i \cap g_j) \cdot \log \frac{\mu(f_i \cap g_j)}{\mu(f_i)}. \tag{3}$$

As in [3] we define in the set of all fuzzy partitions of a fuzzy probability space $(\Omega, M, \mu)$ the relation $\prec$: Let $\xi, \eta$ be two fuzzy partitions of a fuzzy probability space $(\Omega, M, \mu)$. Then we write $\xi \prec \eta$ (and we say that the fuzzy partition $\eta$ is a refinement of the fuzzy partition $\xi$) iff for every $g \in \eta$ there exists $f \in \xi$ such that $g \leq f$. Given two fuzzy partitions $\xi = \{f_1, \ldots, f_n\}$ and $\eta = \{g_1, \ldots, g_m\}$ of a fuzzy probability space $(\Omega, M, \mu)$, their common refinement $\xi \vee \eta$ is defined as the system $\xi \vee \eta = \{f_i \cap g_j; i = 1, \ldots, n, \ j = 1, \ldots, m\}$. The fuzzy partition $\xi \vee \eta$ represents a joint experiment of experiments $\xi, \eta$. Evidently, $\xi \prec \xi \vee \eta$ and $\eta \prec \xi \vee \eta$. If $\xi_1, \xi_2, \ldots, \xi_n$ are fuzzy partitions of a fuzzy probability space $(\Omega, M, \mu)$, then we put $\vee_{i=1}^n \xi_i = \xi_1 \vee \xi_2 \vee \ldots \vee \xi_n$. Two fuzzy partitions $\xi = \{f_1, \ldots, f_n\}$ and $\eta = \{g_1, \ldots, g_m\}$ of a given fuzzy probability space $(\Omega, M, \mu)$ are called statistically independent, if $\mu(f_i \cap g_j) = \mu(f_i) \cdot \mu(g_j)$, for $i = 1, 2, \ldots, n, j = 1, 2, \ldots, m$.

**Example 1.** *Let $\Omega = [0, 1]$, and let $f : \Omega \to [0, 1]$ be defined by $f(\omega) = \omega$, $\omega \in \Omega$. Consider a fuzzy measurable space $(\Omega, M)$, where $M = \{1_\Omega, 0_\Omega, f, f^\perp, f \cup f^\perp, f \cap f^\perp\}$. Then it is easy to verify that the mapping $\mu : M \to [0, 1]$ defined by the equalities $\mu(1_\Omega) = \mu(f \cup f^\perp) = 1$, $\mu(0_\Omega) = \mu(f \cap f^\perp) = 0$, $\mu(f) = p$, $\mu(f^\perp) = 1 - p$, where $p \in (0, 1)$, is a fuzzy P-measure and the system $(\Omega, M, \mu)$ is a fuzzy probability space. The sets $\xi = \{f, f^\perp\}, \eta = \{f \cup f^\perp\}, \varsigma = \{1_\Omega\}$ are fuzzy partitions of $(\Omega, M, \mu)$ such that $\varsigma \prec \eta \prec \xi$. We can calculate their entropy: $H_\mu(\xi) = -p \log p - (1 - p) \log(1 - p)$, $H_\mu(\eta) = H_\mu(\varsigma) = 0$. This makes sense, because the partitions $\eta$ and $\varsigma$ represent experiments whose results are certain events. In particular, if $p = \frac{1}{2}$, then $H_\mu(\xi) = \log 2 = 1$ bit.*

The entropy and the conditional entropy of fuzzy partitions of a fuzzy probability space $(\Omega, M, \mu)$ satisfy all properties analogous to properties of Shannon's entropy of measurable partitions in the classical case; the proofs can be found in [3,6], respectively. We present some of them. If $\xi$, $\eta$, $\varsigma$ are fuzzy partitions of a fuzzy probability space $(\Omega, M, \mu)$, then:

(2.2)  $\xi \prec \eta$ implies $H_\mu(\xi) \leq H_\mu(\eta)$;

(2.3)  $H_\mu(\eta \vee \varsigma / \xi) = H_\mu(\varsigma / \xi \vee \eta) + H_\mu(\eta / \xi)$;

(2.4)  $\xi \prec \eta$ implies $H_\mu(\xi / \varsigma) \leq H_\mu(\eta / \varsigma)$;

(2.5)  $\xi \prec \eta$ implies $H_\mu(\varsigma / \xi) \geq H_\mu(\varsigma / \eta)$;

(2.6)  $H_\mu(\xi / \eta) \leq H_\mu(\xi)$ with the equality if and only if $\xi$, $\eta$ are statistically independent;

(2.7)  $H_\mu(\eta \vee \varsigma / \xi) \leq H_\mu(\eta / \xi) + H_\mu(\varsigma / \xi)$;

(2.8)  $H_\mu(\xi \vee \eta) = H_\mu(\xi) + H_\mu(\eta / \xi)$.

**Definition 3** [6]**.** *Let $\xi$, $\eta$ be two fuzzy partitions of a given fuzzy probability space $(\Omega, M, \mu)$. The mutual information of $\xi$ and $\eta$ is defined by the formula:*

$$I_\mu(\xi, \eta) = H_\mu(\xi) - H_\mu(\xi / \eta). \tag{4}$$

As a simple consequence of (2.8) we have:

$$I_\mu(\xi, \eta) = H_\mu(\xi) + H_\mu(\eta) - H_\mu(\xi \vee \eta). \tag{5}$$

It is evident that $I_\mu(\xi, \xi) = H_\mu(\xi)$, and $I_\mu(\xi, \eta) = I_\mu(\eta, \xi)$. Hence, we can write:

$$I_\mu(\xi, \eta) = H_\mu(\eta) - H_\mu(\eta / \xi). \tag{6}$$

The proofs of the following two theorems can be found in [6].

**Theorem 1** [6]**.** *Let $\xi = \{f_1, \ldots, f_n\}$ and $\eta = \{g_1, \ldots, g_m\}$ be two fuzzy partitions of a fuzzy probability space $(\Omega, M, \mu)$. Then:*

$$I_\mu(\xi, \eta) = \sum_{i=1}^{n} \sum_{j=1}^{m} \mu(f_i \cap g_j) \log \frac{\mu(f_i \cap g_j)}{\mu(f_i) \cdot \mu(g_j)}.$$

**Theorem 2** [6]**.** *$I_\mu(\xi, \eta) \geq 0$ with the equality if and only if $\xi$, $\eta$ are statistically independent.*

## 3. Mutual Information and Conditional Mutual Information in Fuzzy Probability Spaces

In this section by using our previous results we introduce the notion of conditional mutual information in fuzzy probability spaces. We derive chain rules for mutual information of fuzzy partitions and we will prove some more properties concerning these measures.

**Definition 4.** *Let $\xi$, $\eta$, $\varsigma$ be fuzzy partitions of a fuzzy probability space $(\Omega,\ M, \mu)$. Then the conditional mutual information of $\xi$ and $\eta$ given $\varsigma$ is defined by the formula:*

$$I_\mu(\xi, \eta/\varsigma) = H_\mu(\xi/\varsigma) - H_\mu(\xi/\eta \vee \varsigma). \tag{7}$$

**Remark 1.** *Since, according to (2.5), we have $H_\mu(\xi/\varsigma) \geq H_\mu(\xi/\eta \vee \varsigma)$, for the conditional mutual information the inequality $I_\mu(\xi, \eta/\varsigma) \geq 0$ holds.*

**Theorem 3.** *For fuzzy partitions $\xi$, $\eta$, $\varsigma$ of a fuzzy probability space $(\Omega,\ M, \mu)$, it holds:*

$$I_\mu(\xi, \eta \vee \varsigma) = I_\mu(\xi, \eta) + I_\mu(\xi, \varsigma/\eta) = I_\mu(\xi, \varsigma) + I_\mu(\xi, \eta/\varsigma).$$

**Proof.** By simple calculation we obtain:

$$\begin{aligned}
I_\mu(\xi, \eta) + I_\mu(\xi, \varsigma/\eta) &= H_\mu(\xi) - H_\mu(\xi/\eta) + H_\mu(\xi/\eta) - H_\mu(\xi/\eta \vee \varsigma) \\
&= H_\mu(\xi) - H_\mu(\xi/\eta \vee \varsigma) = I_\mu(\xi, \eta \vee \varsigma).
\end{aligned}$$

Analogously we obtain the second equality. □

Let $\xi$, $\eta$ be fuzzy partitions of a fuzzy probability space $(\Omega,\ M, \mu)$. Denote $\xi_0 = \{1_\Omega\}$. Evidently, we have $\xi_0 \vee \xi = \xi$, and $H_\mu(\xi/\xi_0) = H_\mu(\xi)$. Hence, we can write:

$$I_\mu(\xi, \eta/\xi_0) = H_\mu(\xi/\xi_0) - H_\mu(\xi/\eta \vee \xi_0) = H_\mu(\xi) - H_\mu(\xi/\eta) = I_\mu(\xi, \eta).$$

**Theorem 4** (Chain rules). *Let $\xi_1$, $\xi_2, \ldots, \xi_n$ and $\eta$ be fuzzy partitions of a fuzzy probability space $(\Omega,\ M, \mu)$. Then, for $n = 1, 2, \ldots$, the following equalities hold:*

(i)　　$H_\mu(\xi_1 \vee \xi_2 \vee \ldots \vee \xi_n) = \sum_{i=1}^{n} H_\mu(\xi_i / \vee_{k=0}^{i-1} \xi_k)$;
(ii)　　$H_\mu(\vee_{i=1}^{n} \xi_i / \eta) = \sum_{i=1}^{n} H_\mu(\xi_i / (\vee_{k=0}^{i-1} \xi_k) \vee \eta)$;
(iii)　　$I_\mu(\vee_{i=1}^{n} \xi_i, \eta) = \sum_{i=1}^{n} I_\mu(\xi_i, \eta / \vee_{k=0}^{i-1} \xi_k)$.

**Proof.** The first two equalities may be obtained using the principle of mathematical induction from the properties (2.8) and (2.3). By Equation (4), using the two previous equalities and Equation (7), we obtain:

$$\begin{aligned}
I_\mu(\vee_{i=1}^{n} \xi_i, \eta) &= H_\mu(\vee_{i=1}^{n} \xi_i) - H_\mu(\vee_{i=1}^{n} \xi_i / \eta) \\
&= \sum_{i=1}^{n} H_\mu(\xi_i / \vee_{k=0}^{i-1} \xi_k) - \sum_{i=1}^{n} H_\mu(\xi_i / (\vee_{k=0}^{i-1} \xi_k) \vee \eta) \\
&= \sum_{i=1}^{n} \left( H_\mu(\xi_i / \vee_{k=0}^{i-1} \xi_k) - H_\mu(\xi_i / (\vee_{k=0}^{i-1} \xi_k) \vee \eta) \right) = \sum_{i=1}^{n} I_\mu(\xi_i, \eta / \vee_{k=0}^{i-1} \xi_k). \quad □
\end{aligned}$$

**Definition 5.** *Let $\xi$, $\eta$, $\varsigma$ be fuzzy partitions of a fuzzy probability space $(\Omega,\ M, \mu)$. We say that $\xi$ is conditionally independent to $\varsigma$ given $\eta$ (and write $\xi \to \eta \to \varsigma$) if $I_\mu(\xi, \varsigma/\eta) = 0$.*

**Theorem 5.** *For fuzzy partitions $\xi$, $\eta$, $\varsigma$ of a fuzzy probability space $(\Omega,\ M, \mu)$, it holds:*

$$\xi \to \eta \to \varsigma \text{ if and only if } \varsigma \to \eta \to \xi.$$

**Proof.** Let $\xi \to \eta \to \varsigma$. Then:

$$0 = I_\mu(\xi, \varsigma/\eta) = H_\mu(\xi/\eta) - H_\mu(\xi/\eta \vee \varsigma),$$

and according to the property (2.8) we can write:

$$H_\mu(\xi/\eta) = H_\mu(\xi/\eta \vee \varsigma) = H_\mu(\xi \vee \eta \vee \varsigma) - H_\mu(\eta \vee \varsigma).$$

Calculate:

$$
\begin{aligned}
I_\mu(\varsigma, \xi/\eta) &= H_\mu(\varsigma/\eta) - H_\mu(\varsigma/\xi \vee \eta) \\
&= H_\mu(\varsigma \vee \eta) - H_\mu(\eta) - H_\mu(\xi \vee \eta \vee \varsigma) + H_\mu(\xi \vee \eta) \\
&= H_\mu(\xi \vee \eta) - H_\mu(\eta) - H_\mu(\xi/\eta) = H_\mu(\xi/\eta) - H_\mu(\xi/\eta) = 0.
\end{aligned}
$$

It follows that $\varsigma \to \eta \to \xi$. The opposite implication is obvious. $\square$

According to Theorem 5, we may say that $\xi$ and $\varsigma$ are conditionally independent given $\eta$ and write $\xi \leftrightarrow \eta \leftrightarrow \varsigma$ instead of $\xi \to \eta \to \varsigma$.

**Theorem 6.** *For fuzzy partitions $\xi$, $\eta$, $\varsigma$ of a fuzzy probability space $(\Omega, M, \mu)$ such that $\xi \to \eta \to \varsigma$, we have:*

(i)  $I_\mu(\xi \vee \eta, \varsigma) = I_\mu(\eta, \varsigma)$;
(ii)  $I_\mu(\eta, \varsigma) = I_\mu(\xi, \varsigma) + I_\mu(\varsigma, \eta/\xi)$;
(iii)  $I_\mu(\xi, \eta/\varsigma) \leq I_\mu(\xi, \eta)$;
(iv)  $I_\mu(\xi, \eta) \geq I_\mu(\xi, \varsigma)$.

**Proof.**

(i)  Since by the assumption $I_\mu(\xi, \varsigma/\eta) = 0$, using the chain rule for logical mutual information, we obtain:
$$I_\mu(\xi \vee \eta, \varsigma) = I_\mu(\eta \vee \xi, \varsigma) = I_\mu(\eta, \varsigma) + I_\mu(\xi, \varsigma/\eta) = I_\mu(\eta, \varsigma).$$

(ii)  According to Theorem 3 we have $I_\mu(\xi \vee \eta, \varsigma) = I_\mu(\varsigma, \xi) + I_\mu(\varsigma, \eta/\xi)$. Hence, using the equality (i) of this theorem, we obtain:

$$I_\mu(\eta, \varsigma) = I_\mu(\xi \vee \eta, \varsigma) = I_\mu(\varsigma, \xi) + I_\mu(\varsigma, \eta/\xi).$$

(iii)  Since $I_\mu(\xi, \eta) \geq 0$, from (ii) it follows the inequality:

$$I_\mu(\varsigma, \eta/\xi) \leq I_\mu(\varsigma, \eta).$$

By Theorem 5 we can interchange $\xi$ and $\varsigma$. Doing so we obtain $I_\mu(\xi, \eta/\varsigma) \leq I_\mu(\xi, \eta)$.

(iv)  By Theorem 3, the mutual information $I_\mu(\xi, \eta \vee \varsigma)$ can be expressed in two different ways:

$$
\begin{aligned}
I_\mu(\xi, \eta \vee \varsigma) &= I_\mu(\xi, \eta) + I_\mu(\xi, \varsigma/\eta) \\
&= I_\mu(\xi, \varsigma) + I_\mu(\xi, \eta/\varsigma).
\end{aligned}
$$

Since $\xi \to \eta \to \varsigma$, we have $I_\mu(\xi, \varsigma/\eta) = 0$, and, therefore, it holds $I_\mu(\xi, \eta) = I_\mu(\xi, \eta \vee \varsigma)$. Using the second equality, we obtain:

$$I_\mu(\xi, \eta) = I_\mu(\xi, \eta \vee \varsigma) = I_\mu(\xi, \varsigma) + I_\mu(\xi, \eta/\varsigma).$$

Since $I_\mu(\xi, \eta/\varsigma) \geq 0$, we have $I_\mu(\xi, \eta) \geq I_\mu(\xi, \varsigma)$. $\square$

Note that, in the classical theory, the last assertion from the previous theorem is known as the data processing inequality.

## 4. Kullback-Leibler Divergence with Respect to Fuzzy P-Measures

In this part we define the Kullback-Leibler divergence in fuzzy probability spaces and its conditional version. We prove basic properties of these measures; in particular, Gibb's inequality. Finally, using the concept of conditional Kullback-Leibler divergence we establish a chain rule for Kullback-Leibler divergence with respect to fuzzy P-measures. In the proofs we use the following known log-sum inequality: for non-negative real numbers $a_1$, $a_2, \ldots, a_n$, $b_1$, $b_2, \ldots, b_n$, it holds:

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq (\sum_{i=1}^{n} a_i) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \tag{8}$$

with the equality if and only if $\frac{a_i}{b_i}$ is constant. Recall that we use the convention that $x \log \frac{x}{0} = \infty$ if $x > 0$, and $0 \log \frac{0}{x} = 0$ if $x \geq 0$.

**Definition 6.** *Let $\mu$, $\nu$ be fuzzy P-measures on a common fuzzy measurable space $(\Omega, M)$. Then, for any fuzzy partition $\xi$ of fuzzy probability spaces $(\Omega, M, \mu)$, $(\Omega, M, \nu)$, we define the Kullback-Leibler divergence $D_\xi(\mu \parallel \nu)$ by:*

$$D_\xi(\mu \| \nu) = \sum_{i=1}^{n} \mu(f_i) \, \log \frac{\mu(f_i)}{\nu(f_i)}.$$

**Remark 2.** *The Kullback-Leibler divergence is not a metric in a true sense since it is not symmetric, i.e., the equality $D_\xi(\mu \parallel \nu) = D_\xi(\nu \parallel \mu)$ is not necessarily true (as shown in the example that follows), and does not satisfy the triangle inequality.*

**Example 2.** *Consider the fuzzy measurable space $(\Omega, M)$ from Example 1 and the following two fuzzy P-measures $\mu$, $\nu$ defined on M: $\mu$ is defined as in Example 1 and $\nu$ is defined in a similar way: $\nu(1_\Omega) = \nu(f \cup f^\perp) = 1$, $\nu(0_\Omega) = \nu(f \cap f^\perp) = 0$, $\nu(f) = q$, $\nu(f^\perp) = 1 - q$, where $q \in (0, 1)$. Then, for the fuzzy partition $\xi = \{f, f^\perp\}$, we obtain:*

$$D_\xi(\mu \| \nu) = p \cdot \log \frac{p}{q} + (1 - p) \cdot \log \frac{1 - p}{1 - q}, \text{ and } D_\xi(\nu \| \mu) = q \cdot \log \frac{q}{p} + (1 - q) \cdot \log \frac{1 - q}{1 - p}.$$

*If $p = q$, then $D_\xi(\mu \parallel \nu) = D_\xi(\nu \parallel \mu) = 0$. If $p = \frac{1}{2}$, $q = \frac{1}{3}$, then we have:*

$$D_\xi(\mu \| \nu) = \frac{1}{2} \cdot \log \frac{\frac{1}{2}}{\frac{1}{3}} + \frac{1}{2} \cdot \log \frac{\frac{1}{2}}{\frac{2}{3}} = \frac{1}{2} \cdot \log \frac{3}{2} + \frac{1}{2} \cdot \log \frac{3}{4} = 0.084963 \text{ bit}$$

*and:*

$$D_\xi(\nu \| \mu) = \frac{1}{3} \cdot \log \frac{\frac{1}{3}}{\frac{1}{2}} + \frac{2}{3} \cdot \log \frac{\frac{2}{3}}{\frac{1}{2}} = \frac{1}{3} \cdot \log \frac{2}{3} + \frac{2}{3} \cdot \log \frac{4}{3} = 0.0817036 \text{ bit}.$$

*This means that $D_\xi(\mu \parallel \nu) \neq D_\xi(\nu \parallel \mu)$, in general.*

The following result suggests interpretation of Kullback-Leibler divergence as a measure of how different two fuzzy P-measures (over the same fuzzy partition) are.

**Theorem 7.** *Let $\xi = \{f_1, \ldots, f_n\}$ be a fuzzy partition of fuzzy probability spaces $(\Omega, M, \mu)$, $(\Omega, M, \nu)$. Then $D_\xi(\mu \parallel \nu) \geq 0$ (Gibb's inequality) with the equality if and only if $\mu(f_i) = \nu(f_i)$, for $i = 1, 2, \ldots, n$.*

**Proof.** Putting $a_i = \mu(f_i)$ and $b_i = \nu(f_i)$, for $i = 1, 2, \ldots, n$, in the log-sum inequality, we obtain $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} \mu(f_i) = \mu(\cup_{i=1}^{n} f_i) = 1$; analogously we obtain $\sum_{i=1}^{n} b_i = 1$. Therefore, by Equation (8):

$$D_{\xi}(\mu \| \nu) = \sum_{i=1}^{n} \mu(f_i) \, \log \frac{\mu(f_i)}{\nu(f_i)}$$

$$\geq \left( \sum_{i=1}^{n} \mu(f_i) \right) \, \log \frac{\sum_{i=1}^{n} \mu(f_i)}{\sum_{i=1}^{n} \nu(f_i)} = 1 \cdot \log \frac{1}{1} = 0$$

with the equality if and only if $\frac{\mu(f_i)}{\nu(f_i)} = \alpha$ for $i = 1, 2, \ldots, n$, where $\alpha$ is constant. Taking the sum for all $i = 1, 2, \ldots, n$, we obtain $\alpha = 1$. This means that $D_{\xi}(\mu \| \nu) = 0$ if and only if $\mu(f_i) = \nu(f_i)$, for $i = 1, 2, \ldots, n$. $\square$

**Theorem 8.** *Let $\mu, \nu$ be fuzzy P-measures on a common fuzzy measurable space $(\Omega, M)$. Then, for any fuzzy partition $\xi = \{f_1, \ldots, f_n\}$ of fuzzy probability spaces $(\Omega, M, \mu)$, $(\Omega, M, \nu)$, it holds:*

$$H_{\mu}(\xi) = \log n - D_{\xi}(\mu \| \nu),$$

*where $n$ is the number of members of $\xi$, and $\nu$ is the uniform probability distribution over $\xi$, i.e., $\nu(f_i) = \frac{1}{n}$, for $i = 1, 2, \ldots, n$.*

**Proof.** Calculate:

$$D_{\xi}(\mu \| \nu) = \sum_{i=1}^{n} \mu(f_i) \, \log \frac{\mu(f_i)}{\nu(f_i)} = \sum_{i=1}^{n} \mu(f_i) \, \log \frac{\mu(f_i)}{\frac{1}{n}}$$

$$= \sum_{i=1}^{n} \mu(f_i) \left( \log \mu(f_i) - \log n^{-1} \right)$$

$$= \sum_{i=1}^{n} \mu(f_i) \, \log \mu(f_i) + \log n = \log n - H_{\mu}(\xi). \ \square$$

As a consequence of the previous two theorems we obtain the following property of entropy of fuzzy partitions:

**Corollary 1.** *For any fuzzy partition $\xi$ of a fuzzy probability space $(\Omega, M, \mu)$, it holds $H_{\mu}(\xi) \leq \log n$, where $n$ denotes the number of members of $\xi$, with the equality if and only if $\mu$ is the uniform probability distribution over $\xi$.*

**Proof.** Let $\nu$ be the uniform probability distribution over $\xi = \{f_1, \ldots, f_n\}$. Then, according to the previous theorem and Gibb's inequality (Theorem 7) we have:

$$0 \leq D_{\xi}(\mu \| \nu) = \log n - H_{\mu}(\xi),$$

which implies the inequality $H_{\mu}(\xi) \leq \log n$, where $n$ is the number of members of $\xi$. Since by Theorem 7 $D_{\xi}(\mu \| \nu) = 0$ if and only if $\mu(f_i) = \nu(f_i)$, for $i = 1, 2, \ldots, n$, the equality $H_{\mu}(\xi) = \log n$ holds if and only if $\mu$ is the uniform probability distribution over $\xi$. $\square$

In the following, a concavity of entropy $H_{\mu}(\xi)$ as a function of $\mu$ and convexity of K-L divergence with respect to fuzzy P-measures are shown. We recall, for the convenience of the reader, the definitions of convex and concave function:

A real-valued function $f$ is said to be convex over an interval $[a, b]$ if for every $x_1, x_2 \in [a, b]$ and for any real number $\alpha \in [0, 1]$,

$$f(\alpha x_1 + (1 - \alpha) x_2) \leq \alpha f(x_1) + (1 - \alpha) f(x_2).$$

A real-valued function $f$ is said to be concave over an interval $[a, b]$ if for every $x_1, x_2 \in [a, b]$ and for any real number $\alpha \in [0, 1]$,

$$f\left(\alpha x_1 + (1 - \alpha)x_2\right) \geq \alpha f(x_1) + (1 - \alpha)f(x_2).$$

**Remark 3.** *In the proofs of some of the next assertions, the following known properties are used:*

(i)   *A function $f$ is concave over an interval if and only if the function $-f$ is convex over the interval.*
(ii)   *The sum of two concave functions is itself concave; the sum of two convex functions is itself convex.*
(iii)   *Every real-valued affine function, i.e., each function of the form $f(x) = ax + b$, $a, b \in \Re$, is simultaneously convex and concave.*

It is easy to prove the following proposition.

**Proposition 1.** *Let $\mu_1$, $\mu_2$ be fuzzy P-measures on a given fuzzy measurable space $(\Omega, M)$. Then, for every real number $\alpha \in [0, 1]$, the mapping $\alpha\mu_1 + (1 - \alpha)\mu_2$ is a fuzzy P-measure on $(\Omega, M)$.*

**Theorem 9** (Concavity of entropy). *Let $\xi$ be a fuzzy partition of fuzzy probability spaces $(\Omega, M, \mu_1)$, $(\Omega, M, \mu_2)$. Then, for any real number $\alpha \in [0, 1]$, the following inequality holds:*

$$\alpha\, H_{\mu_1}(\xi) + (1 - \alpha)\, H_{\mu_2}(\xi) \leq H_{\alpha\,\mu_1 + (1-\alpha)\mu_2}(\xi).$$

**Proof.** Assume that $\xi = \{f_1, f_2, \ldots, f_n\}$. Since the function $F$ is convex, we obtain:

$$\alpha\, H_{\mu_1}(\xi) + (1 - \alpha)\, H_{\mu_2}(\xi) = -\alpha\sum_{i=1}^{n} F(\mu_1(f_i)) - (1 - \alpha)\sum_{i=1}^{n} F(\mu_2(f_i))$$

$$= -\sum_{i=1}^{n}\left(\alpha\, F(\mu_1(f_i)) + (1 - \alpha)\, F(\mu_2(f_i))\right) \leq -\sum_{i=1}^{n} F(\alpha\mu_1(f_i) + (1 - \alpha)\mu_2(f_i))$$

$$= -\sum_{i=1}^{n} F((\alpha\mu_1 + (1 - \alpha)\mu_2)(f_i)) = H_{\alpha\mu_1 + (1-\alpha)\mu_2}(\xi),$$

which proves the concavity of entropy $H_\mu(\xi)$ as a function of fuzzy P-measure $\mu$. $\square$

**Theorem 10** (Convexity of K-L divergence). *The K-L divergence $D_\xi(\mu \| \nu)$ is convex in the pair $(\mu, \nu)$, i.e., for fuzzy P-measures $\mu_1, \nu_1$ and $\mu_2, \nu_2$ on a common fuzzy measurable space $(\Omega, M)$, we have, for any fuzzy partition $\xi$ of fuzzy probability spaces $(\Omega, M, \mu_i)$, $(\Omega, M, \nu_i)$, $i = 1, 2$, and for any real number $\alpha \in [0, 1]$, the following inequality:*

$$D_\xi(\alpha\mu_1 + (1 - \alpha)\mu_2 \| \alpha\nu_1 + (1 - \alpha)\nu_2) \leq \alpha D_\xi(\mu_1 \| \nu_1) + (1 - \alpha)D_\xi(\mu_2 \| \nu_2).$$

**Proof.** Let $\xi = \{f_1, \ldots, f_n\}$ be a fuzzy partition of fuzzy probability spaces $(\Omega, M, \mu_1)$, $(\Omega, M, \mu_2)$, $(\Omega, M, \nu_1)$, $(\Omega, M, \nu_2)$. Fix $i \in \{1, 2, \ldots, n\}$. Putting $a_1 = \alpha\mu_1(f_i)$, $a_2 = (1 - \alpha)\mu_2(f_i)$, $b_1 = \alpha\nu_1(f_i)$, $b_2 = (1 - \alpha)\nu_2(f_i)$ in the log-sum inequality, we obtain:

$$(\alpha\mu_1(f_i) + (1 - \alpha)\mu_2(f_i)) \cdot \log\frac{\alpha\mu_1(f_i) + (1-\alpha)\mu_2(f_i)}{\alpha\nu_1(f_i) + (1-\alpha)\nu_2(f_i)}$$

$$\leq \alpha\mu_1(f_i) \cdot \log\frac{\alpha\mu_1(f_i)}{\alpha\nu_1(f_i)} + (1 - \alpha)\mu_2(f_i) \cdot \log\frac{(1-\alpha)\mu_2(f_i)}{(1-\alpha)\nu_2(f_i)}.$$

Summing these inequalities over all $i$, we obtain what was claimed. $\square$

**Theorem 11.** *Let $\xi = \{f_1, \ldots, f_n\}$ and $\eta = \{g_1, \ldots, g_m\}$ be two fuzzy partitions of a fuzzy probability space $(\Omega, M, \mu)$. Then the mutual information $I_\mu(\xi, \eta)$ of $\xi$ and $\eta$ is a concave function of $\mu$ for fixed $\mu(g_j/f_i)$,*

$j = 1, 2, \ldots, m$, $i = 1, \ldots, n$. *The mutual information $I_\mu(\xi, \eta)$ of $\xi$ and $\eta$ is a convex function of $\mu(g_j/f_i)$, for fixed $\mu(f_i)$, $i = 1, \ldots, n$.*

**Proof.** Let us prove the first assertion. By Equations (6) and (2) we have:

$$I_\mu(\xi, \eta) = H_\mu(\eta) - \sum_{i=1}^{n} \sum_{j=1}^{m} \mu(f_i) \cdot F(\dot{\mu}(g_j/f_i)). \tag{9}$$

Let $g_j \in M$. Since $f_k \cap g_j \le f_k \le f_l^\perp \le f_l^\perp \cup g_j^\perp = (f_l \cap g_j)^\perp$ whenever $k \ne l$, the system $\{f_1 \cap g_j, \ldots, f_n \cap g_j\}$ is a system of pairwise W-separated fuzzy subsets from $M$. Due to the assumption $\mu(\cup_{i=1}^{n} f_i) = 1$, the property (2.1) and additivity of the fuzzy P-measure, we get:

$$\mu(g_j) = \mu\big((\cup_{i=1}^{n} f_i) \cap g_j\big) = \mu\big(\cup_{i=1}^{n}(f_i \cap g_j)\big) = \sum_{i=1}^{n} \mu(f_i \cap g_j) = \sum_{i:\mu(f_i)>0} \mu(f_i)\mu(g_j/f_i).$$

This means that, for fixed $\mu(g_j/f_i)$, $\mu(g_j)$ is a linear function of $\mu(f_i)$. Therefore, $H_\mu(\eta)$, which is a concave function of $\mu(g_j)$, is a concave function of $\mu(f_i)$. The second term of the difference in Equation (9) is a linear function of $\mu(f_i)$. Hence, (see Remark 3) the difference in Equation (9) is a concave function of $\mu(f_i)$. Thus, the first part of theorem is proved.

Now, let us prove the second part. We fix $\mu$ over $\xi$ and consider two different conditional fuzzy P-measures $\mu_1(g_j/f_i)$ and $\mu_2(g_j/f_i)$, defined for $f_i \in \xi$, $\mu(f_i) > 0$, and $j = 1, 2, \ldots, m$. Then, for $i = 1, \ldots, n$, $j = 1, 2, \ldots, m$, we have:

$$\mu_1(f_i \cap g_j) = \mu(f_i)\,\dot{\mu}_1(g_j/f_i), \text{ and } \mu_2(f_i \cap g_j) = \mu(f_i)\,\dot{\mu}_2(g_j/f_i).$$

According to Proposition 1 we can define, for every real number $\alpha \in [0, 1]$, the following conditional fuzzy P-measure:

$$\mu_\alpha(g_j/f_i) = \alpha\mu_1(g_j/f_i) + (1 - \alpha)\,\mu_2(g_j/f_i).$$

Then we have:

$$\mu_\alpha(f_i \cap g_j) = \alpha\mu_1(f_i \cap g_j) + (1 - \alpha)\,\mu_2(f_i \cap g_j), \text{ and}$$
$$\mu_\alpha(g_j) = \alpha\mu_1(g_j) + (1 - \alpha)\,\mu_2(g_j).$$

Therefore, if we put $\nu_\alpha(f_i \cap g_j) = \mu(f_i) \cdot \mu_\alpha(g_j)$, we obtain:

$$\nu_\alpha(f_i \cap g_j) = \alpha\mu(f_i) \cdot \mu_1(g_j) + (1 - \alpha)\mu(f_i) \cdot \mu_2(g_j).$$

According to Theorem 1 and Theorem 10 we can write:

$$I_{\mu_\alpha}(\xi, \eta) = \sum_{i=1}^{n} \sum_{j=1}^{m} \mu_\alpha(f_i \cap g_j) \log \frac{\mu_\alpha(f_i \cap g_j)}{\mu(f_i) \cdot \mu_\alpha(g_j)} = D_{\eta/\xi}(\mu \| \nu_\alpha)$$

$$\le \alpha\sum_{i=1}^{n} \sum_{j=1}^{m} \mu_1(f_i \cap g_j) \log \frac{\mu_1(f_i \cap g_j)}{\mu(f_i) \cdot \mu_1(g_j)} + (1 - \alpha)\sum_{i=1}^{n} \sum_{j=1}^{m} \mu_2(f_i \cap g_j) \log \frac{\mu_2(f_i \cap g_j)}{\mu(f_i) \cdot \mu_2(g_j)}$$

$$= \alpha\, I_{\mu_1}(\xi, \eta) + (1 - \alpha)I_{\mu_2}(\xi, \eta). \,\square$$

Finally, we define a conditional version of the Kullback-Leibler divergence and, subsequently, we will prove the chain rule for K-L divergence.

**Definition 7.** *Let $\xi = \{f_1, \ldots, f_n\}$, $\eta = \{g_1, \ldots, g_m\}$ be two fuzzy partitions of fuzzy probability spaces $(\Omega,\ M, \mu)$, $(\Omega,\ M, \nu)$. Then we define the conditional Kullback-Leibler divergence $D_{\eta/\xi}(\mu \parallel \nu)$ by:*

$$D_{\eta/\xi}(\mu \| \nu) = \sum_{i=1}^{n} \mu(f_i) \sum_{j=1}^{m} \dot{\mu}(g_j/f_i) \, \log \frac{\dot{\mu}(g_j/f_i)}{\dot{\nu}(g_j/f_i)}.$$

**Theorem 12** (Chain rule for K-L divergence)**.** *If $\xi$, $\eta$ are two fuzzy partitions of fuzzy probability spaces $(\Omega,\ M, \mu)$, $(\Omega,\ M, \nu)$, then:*

$$D_{\xi \vee \eta}(\mu \| \nu) = D_{\xi}(\mu \| \nu) + D_{\eta/\xi}(\mu \| \nu).$$

**Proof.** Let $\xi = \{f_1, \ldots, f_n\}$, $\eta = \{g_1, \ldots, g_m\}$. Using the property (2.1), we obtain:

$$\sum_{j=1}^{m} \mu(f_i \cap g_j) = \mu\left( f_i \cap (\cup_{j=1}^{m} g_j) \right) = \mu(f_i), \text{ for } i = 1, 2, \ldots, n.$$

Therefore:

$$D_{\xi \vee \eta}(\mu \| \nu) = \sum_{i=1}^{n} \sum_{j=1}^{m} \mu(f_i \cap g_j) \, \log \frac{\mu(f_i \cap g_j)}{\nu(f_i \cap g_j)}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \mu(f_i \cap g_j) \, \log \frac{\mu(f_i) \, \dot{\mu}(g_j/f_i)}{\nu(f_i) \, \dot{\nu}(g_j/f_i)}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \mu(f_i \cap g_j) \, \log \frac{\mu(f_i)}{\nu(f_i)} + \sum_{i=1}^{n} \sum_{j=1}^{m} \mu(f_i \cap g_j) \, \log \frac{\dot{\mu}(g_j/f_i)}{\dot{\nu}(g_j/f_i)}$$

$$= \sum_{i=1}^{n} \mu(f_i) \, \log \frac{\mu(f_i)}{\nu(f_i)} + \sum_{i=1}^{n} \mu(f_i) \sum_{j=1}^{m} \dot{\mu}(g_j/f_i) \, \log \frac{\dot{\mu}(g_j/f_i)}{\dot{\nu}(g_j/f_i)} = D_{\xi}(\mu \| \nu) + D_{\eta/\xi}(\mu \| \nu). \ \square$$

## 5. Discussion

In this paper we extend our study concerning mutual information of fuzzy partitions in fuzzy probability spaces. In Section 3, using our previous results, the notion of conditional mutual information of fuzzy partitions is defined and chain rules for mutual information of fuzzy partitions are established. Subsequently, using the notion of conditional mutual information of fuzzy partitions we have defined the notion of conditional independence of fuzzy partitions and we have proved, inter alia, data processing inequality for the studied situation. In Section 4 the notion of Kullback-Leibler divergence for fuzzy P-measures is introduced and the basic properties of this measure are shown. In particular, a convexity of Kullback-Leibler divergence with respect to fuzzy P-measures is proved and a convexity of mutual information of fuzzy partitions is studied. Finally, a conditional version of the Kullback-Leibler divergence is defined and chain rules for Kullback-Leibler divergence with respect to fuzzy P-measures are established.

As noted previously, logical versions of some results concerning the entropy and mutual information of fuzzy partitions presented in Sections 2 and 3 are given in [10]. In [9] Ellerman studied, in addition to the logical entropy and logical mutual information, the concept of logical Kullback-Leibler divergence. The aim of our next study will be to provide a logical version of Kullback-Leibler divergence for fuzzy probability measures.

Let us mention, finally, the fuzzy set theory has been shown to be useful in many applications of mathematics and it is continually developing. Currently, algebraic structures, based on fuzzy set theory, as MV-algebras (cf. [37,38]), D-posets (cf. [39,40]), effect algebras (cf. [41,42]), and IF-sets (cf. [43–47]) are subject of intensive research and, of course, there are also many results about the entropy on these structures. Some of them can be found, e.g., in [48–58].

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1.  Kolmogorov, A.N. *Foundations of the Theory of Probability*; Chelsea Press: New York, NY, USA, 1950.
2.  Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
3.  Markechová, D. The entropy of fuzzy dynamical systems and generators. *Fuzzy Sets Syst.* **1992**, *48*, 351–363. [CrossRef]
4.  Piasecki, K. Probability of fuzzy events defined as denumerable additive measure. *Fuzzy Sets Syst.* **1985**, *17*, 271–284. [CrossRef]
5.  Zadeh, L.A. Fuzzy Sets. *Inf. Control* **1965**, *8*, 338–358. [CrossRef]
6.  Markechová, D. Entropy and mutual information of experiments in the fuzzy case. *Neural Netw. World* **2013**, *23*, 339–349. [CrossRef]
7.  Markechová, D. Entropy of complete fuzzy partitions. *Math. Slovaca* **1993**, *43*, 1–10.
8.  Markechová, D. A note to the Kolmogorov-Sinai entropy of fuzzy dynamical systems. *Fuzzy Sets Syst.* **1994**, *64*, 87–90. [CrossRef]
9.  Ellerman, D. An Introduction to Logical Entropy and Its Relation to Shannon Entropy. *Int. J. Seman. Comput.* **2013**, *7*, 121–145. [CrossRef]
10. Markechová, D.; Riečan, B. Logical Entropy of Fuzzy Dynamical Systems. *Entropy* **2016**, *18*, 157. [CrossRef]
11. Ebrahimzadeh, A.; Giski, Z.E.; Markechová, D. Logical Entropy of Dynamical Systems—A General Model. *Mathematics* **2017**, *5*, 4. [CrossRef]
12. Mesiar, R.; Rybárik, J. Entropy of Fuzzy Partitions—A General Model. *Fuzzy Sets Syst.* **1998**, *99*, 73–79. [CrossRef]
13. Mesiar, R. The Bayes formula and the entropy of fuzzy probability spaces. *Int. J. Gen. Syst.* **1990**, *4*, 67–71.
14. Dumitrescu, D. Fuzzy measures and the entropy of fuzzy partitions. *J. Math. Anal. Appl.* **1993**, *176*, 359–373. [CrossRef]
15. Dumitrescu, D. Entropy of a fuzzy dynamical system. *Fuzzy Sets Syst.* **1995**, *70*, 45–57. [CrossRef]
16. Riečan, B. An entropy construction inspired by fuzzy sets. *Soft Comput.* **2003**, *7*, 486–488.
17. Hu, Q.; Yu, D.; Xie, Z.; Liu, J. Fuzzy probabilistic approximation spaces and their information measures. *IEEE Trans. Fuzzy Syst.* **2006**, *14*, 191–201.
18. Yu, D.; Hu, Q.; Wu, C. Uncertainty measures for fuzzy relations and their applications. *Appl. Soft Comput.* **2007**, *7*, 1135–1143. [CrossRef]
19. Rahimi, M.; Riazi, A. On local entropy of fuzzy partitions. *Fuzzy Sets Syst.* **2014**, *234*, 97–108. [CrossRef]
20. Rahimi, M.; Assari, A.; Ramezani, F. A Local Approach to Yager Entropy of Dynamical Systems. *Int. J. Fuzzy Syst.* **2015**, *1*, 1–10. [CrossRef]
21. Markechová, D.; Riečan, B. Entropy of Fuzzy Partitions and Entropy of Fuzzy Dynamical Systems. *Entropy* **2016**, *18*, 19. [CrossRef]
22. Khare, M. Fuzzy σ-algebras and conditional entropy. *Fuzzy Sets Syst.* **1999**, *102*, 287–292. [CrossRef]
23. Khare, M.; Roy, S. Conditional entropy and the Rokhlin metric on an orthomodular lattice with Bayessian state. *Int. J. Theor. Phys.* **2008**, *47*, 1386–1396. [CrossRef]
24. Srivastava, P.; Khare, M.; Srivastava, Y.K. *M*-Equivalence, entropy and *F*-dynamical systems. *Fuzzy Sets Syst.* **2001**, *121*, 275–283. [CrossRef]
25. Criado, F.; Gachechiladze, T. Entropy of fuzzy events. *Fuzzy Sets Syst.* **1997**, *88*, 99–106. [CrossRef]
26. Dubois, D.; Prade, M. A review of fuzzy set aggregation connectives. *Inf. Sci.* **1985**, *36*, 85–121. [CrossRef]
27. Gray, R.M. *Entropy and Information Theory*; Springer: Berlin/Heidelberg, Germany, 2009.
28. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
29. Kullback, S. *Information Theory and Statistics*; John Wiley & Sons: New York, NY, USA, 1959.
30. Schnakenberg, J. Network theory of microscopic and macroscopic behavior of master equation systems. *Rev. Mod. Phys.* **1976**, *48*, 571–585. [CrossRef]
31. Risken, H. *The Fokker-Planck Equation, Methods of Solution and Applications*; Springer: New York, NY, USA, 1984.
32. Qian, H. Relative entropy: Free energy associated with equilibrium fluctuations and nonequilibrium deviations. *Phys. Rev. E* **2001**, *63*, 042103. [CrossRef] [PubMed]
33. Ellis, R.S. *Entropy, Large Deviations, and Statistical Mechanics*; Springer: New York, NY, USA, 1985.
34. Dvurečenskij, A. On the existence of probability measures on fuzzy measurable spaces. *Fuzzy Sets Syst.* **1991**, *43*, 173–181. [CrossRef]

35. Piasecki, K. New concept of separated fuzzy subsets. In Proceedings of the Polish Symposium on Interval and Fuzzy Mathematics, Poznan, Poland, 26–29 August 1983; Albrycht, J., Wiśniewski, H., Eds.; Wydaw. Politechniki Poznańskiej: Poznan, Poland, 1985; pp. 193–195.

36. Piasecki, K. Fuzzy partitions of sets. *BUSEFAL* **1986**, *25*, 52–60.

37. Riečan, B.; Mundici, D. Probability on MV-algebras. In *Handbook of Measure Theory*; Pap, E., Ed.; Elsevier: Amsterdam, The Netherlands, 2002; pp. 869–910.

38. Di Nola, A.; Dvurečenskij, A. Product MV-algebras. *Multiple-Valued Log.* **2001**, *6*, 193–215.

39. Kôpka, F.; Chovanec, F. D-posets. *Math. Slovaca* **1994**, *44*, 21–34.

40. Frič, R. On D-posets of fuzzy sets. *Math. Slovaca* **2014**, *64*, 545–554. [CrossRef]

41. Foulis, D.J.; Bennett, M.K. Effect algebras and unsharp quantum logics. *Found. Phys.* **1994**, *24*, 1325–1346. [CrossRef]

42. Dvurečenskij, A.; Pulmannová, S. *New Trends in Quantum Structures*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2000.

43. Atanassov, K. *Intuitionistic Fuzzy Sets: Theory and Applications*; Physica Verlag: New York, NY, USA, 1999.

44. Atanassov, K. Intuitionistic fuzzy sets. *Fuzzy Sets Syst.* **1986**, *20*, 87–96. [CrossRef]

45. Atanassov, K. More on intuitionistic fuzzy sets. *Fuzzy Sets Syst.* **1989**, *33*, 37–45. [CrossRef]

46. Atanassov, K.; Riečan, B. On two operations over intuitionistic fuzzy sets. *J. Appl. Math. Stat. Inform.* **2006**, *2*, 145–148. [CrossRef]

47. Riečan, B. *Algebraic and Proof-Theoretic Aspects of Non-Classical Logics: Papers in Honor of Daniele Mundici on the Occasion of His 60th Birthday*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 290–308.

48. Petrovičová, J. On the entropy of partitions in product MV-algebras. *Soft Comput.* **2000**, *4*, 41–44. [CrossRef]

49. Petrovičová, J. On the entropy of dynamical systems in product MV-algebras. *Fuzzy Sets Syst.* **2001**, *121*, 347–351. [CrossRef]

50. Riečan, B. Kolmogorov–Sinaj entropy on MV-algebras. *Int. J. Theor. Phys.* **2005**, *44*, 1041–1052. [CrossRef]

51. Ďurica, M. Entropy on IF-events. *Notes Intuit. Fuzzy Sets* **2007**, *13*, 30–40.

52. Di Nola, A.; Dvurečenskij, A.; Hyčko, M.; Manara, C. Entropy on Effect Algebras with the Riesz Decomposition Property I: Basic Properties. *Kybernetika* **2005**, *41*, 143–160.

53. Di Nola, A.; Dvurečenskij, A.; Hyčko, M.; Manara, C. Entropy on Effect Algebras with the Riesz Decomposition Property II: MV-Algebras. *Kybernetika* **2005**, *41*, 161–176.

54. Giski, Z.E.; Ebrahimi, M. Entropy of Countable Partitions on Effect Algebras with the Riesz Decomposition Property and Weak Sequential Effect Algebras. *Cankaya Univ. J. Sci. Eng.* **2015**, *12*, 20–39.

55. Farnoosh, R.; Rahimi, M.; Kumar, P. Removing noise in a digital image using a new entropy method based on intuitionistic fuzzy sets. In Proceedings of the International Conference on Fuzzy Systems, Vancouver, BC, Canada, 24–29 July 2016.

56. Burillo, P.; Bustince, H. Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets. *Fuzzy Sets Syst.* **1996**, *78*, 305–316. [CrossRef]

57. Szmidt, E.; Kacprzyk, J. Entropy for intuitionistic fuzzy sets. *Fuzzy Sets Syst.* **2001**, *118*, 467–477. [CrossRef]

58. Ebrahimzadeh, A.; Giski, Z.E.; Markechová, D. Logical Entropy on Effect Algebras with the Riesz Decomposition Property. *Commun. Theor. Phys.* **2017**. under review.