Metrical Approach to Measuring Uncertainty

Andrey G. Bronevich\textsuperscript{1(✉)} and Igor N. Rozenberg\textsuperscript{2}

\textsuperscript{1} National Research University Higher School of Economics, Myasnitskaya 20, Moscow 101000, Russia
brone@mail.ru

\textsuperscript{2} JSC “Research and Design Institute for Information Technology, Signalling and Telecommunications on Railway Transport”, Orlikov per. 5, building 1, Moscow 107996, Russia
I.Rozenberg@gismps.ru

Abstract. Many uncertainty measures can be generated by the corresponding divergences, like the Kullback-Leibler divergence generates the Shannon entropy. Divergences can evaluate the information gain obtained by knowing a posterior probability distribution w.r.t. a prior one, or the contradiction between them. Divergences can be also viewed as distances between probability distributions. In this paper, we consider divergences that satisfy a weak system of axioms. This system of axioms does not guaranty additivity of divergences and allows us to consider, for example, the $L_{α}$-metric on probability measures as a divergence. We show what kind of uncertainty measures can be generated by such divergences, and how these uncertainty measures can be extended to credal sets.

Keywords: Uncertainty measures · Divergences · Credal sets

1 Introduction

In our experience, we deal with various types of uncertainty. Probability theory allows us to describe conflict in information, other uncertainty theories can generalize it admitting imprecision or non-specificity into models like the theory of imprecise probabilities \cite{1,20} or the theory of belief functions \cite{12,17}. We also need to merge information from different sources, and during this process, it is important to analyze how these sources are contradictory or conflicting. Therefore, we need to measure conflict and non-specificity within each source of information and to measure contradiction among information sources. In probability theory, there are many functionals for evaluating conflict called entropies \cite{9,15,16} and there are many statistical distances called also divergences that can be used for measuring contradiction between probability models \cite{9,16,18,19}. One can notice the visible interactions between various types of uncertainty like contradiction, conflict and non-specificity.
In this paper, we argue that a measure of contradiction (or divergence) is the basic one, and show how other measures of uncertainty can be expressed through it. The paper has the following structure. In Sect. 2, we give some notations and definitions related to probability measures and credal sets. In Sect. 3, we formulate a system of axioms for conflict measures without the requirement of their additivity. In Sect. 4, we introduce a system of axioms for divergences and illustrate them by examples like the Kullback-Leibler divergence, the Rényi divergence, and the $g$-divergence introduced in the paper. Section 5 is devoted to the question: how such uncertainty measures can be extended on credal sets. The paper finishes with the discussion of obtained results and conclusions.

2 Some Notations and Definitions

Let $X$ be a finite non-empty set and let $2^X$ be the powerset of $X$. A set function $P : 2^X \to [0, 1]$ is called a probability measure on $2^X$ if 1) $P(\emptyset) = 0$ and $P(X) = 1$; 2) $P(A \cup B) = P(A) + P(B)$ for all disjoint sets $A, B \in 2^X$. A function $p(x) = P(\{x\})$, $x \in X$, is called the probability distribution. We see that $P(A) = \sum_{x \in A} p(x)$ for every non-empty set $A \in 2^X$. We say that probabilities are uniformly distributed on $X$ if $p(x) = 1/|X|$. The probability measure that corresponds to the uniform probability distribution is denoted by $P_u$. The set of all probability measures on $2^X$ is denoted by $M_{pr}(X)$, and we use the notation $M_{pr}$ if the basic set $X$ can be chosen arbitrary.

We use the following operations on $M_{pr}$:

a) $P = aP_1 + (1 - a)P_2$ is the convex sum of $P_1, P_2 \in M_{pr}(X)$ with $a \in [0, 1]$ if $P(A) = aP_1(A) + (1 - a)P_2(A)$ for all $A \in 2^X$;

b) let $X$ and $Y$ be non-empty finite sets, $\varphi : X \to Y$ be a mapping, and $P \in M_{pr}(X)$, then $P^\varphi \in M_{pr}(Y)$ is defined by $P^\varphi(B) = P(\varphi^{-1}(B))$, where $B \in 2^Y$ and $\varphi^{-1}(B) = \{x \in X | \varphi(x) \in B\}$.

We see that $M_{pr}(X)$ is a convex set and its extreme points are Dirac measures, such measures are defined by a probability distribution $p_x$, $x \in X$, for which $p_x(y) = 1$ if $x = y$, and $p_x(y) = 0$ otherwise. Let $P_x \in M_{pr}$ be a Dirac measure with the probability distribution $p_x$, then every $P \in M_{pr}(X)$ with a probability distribution $p$ is represented as the convex sum of Dirac measures: $P = \sum_{x \in X} p(x)P_x$.

Clearly, we can identify every $P \in M_{pr}(X)$, where $X = \{x_1, ..., x_n\}$, with the point $(p(x_1), ..., p(x_n))$ in $\mathbb{R}^n$. A non-empty subset $P \subseteq M_{pr}(X)$ is called a credal set [1] if this subset is closed and convex. Further $Cr(X)$ denotes the family of all possible credal sets in $M_{pr}(X)$. We reserve the notation $Cr$ if the basic set $X$ can be chosen arbitrary. In the theory of imprecise probabilities, the model uncertainty based on credal sets is the general one. Some credal sets can be defined by monotone measures. A set function $\mu : 2^X \to [0, 1]$ is called a monotone measure [21] if 1) $\mu(\emptyset) = 0$ and $\mu(X) = 1$; 2) $\mu(A) \leq \mu(B)$ for every $A, B \in 2^X$ with $A \subseteq B$. A monotone measure $\mu$ on $2^X$ is called a coherent
upper probability [1,20] if there is a credal set $P \in Cr(X)$ such that $\mu(A) = \sup\{P(A) | P \in P\}$ for all $A \in 2^X$.

In the next, for the sake of notation simplicity, we will write $P(x)$ instead of $P(\{x\})$ omitting the figure brackets.

3 Axioms for a Conflict Measure on $M_{pr}$

In this paper, we will use axioms for a measure of conflict presented in [6]. Let us observe that in [6] such axioms have been formulated for belief functions and the authors show what it happens if the additivity axiom is dropped. We will formulate these axioms for probability measures.

A measure of conflict $U_C$ is the functional $U_C : M_{pr} \rightarrow [0, +\infty)$ that satisfies the following axioms.

Axiom 1. $U_C(P) = 0$ for $P \in M_{pr}(X)$ iff $P = P_x$ for some $x \in X$.

Axiom 2. Let $\varphi : X \rightarrow Y$ be an injection, then $U_C(P^\varphi) = U_C(P)$ for every $P \in M_{pr}(X)$.

Axiom 3. Let $\varphi : X \rightarrow Y$ be an arbitrary mapping, then $U_C(P^\varphi) \leq U_C(P)$ for every $P \in M_{pr}(X)$.

Axiom 4. Let $P_1, P_2 \in M_{pr}(X)$ and $a \in [0,1]$, then $U_C(aP_1 + (1-a)P_2) \geq aU_C(P_1) + (1-a)U_C(P_2)$.

Let us discuss the above axioms. Axiom 1 postulates that the conflict is equal to zero only for the information without uncertainty. Axiom 2 accumulates two known axioms for the Shannon entropy [16]. If $\varphi : X \rightarrow Y$ is the bijection, then Axiom 2 says that $U_C(P)$ does not depend on how elements of $X$ are labeled. Let $\varphi : X \rightarrow Y$ be an injection such that $X \subseteq Y$ and $\varphi(x) = x$ for all $x \in X$, then $P^\varphi$ has the following probability distribution: $P^\varphi(x) = P(x)$ if $x \in X$ and $P^\varphi(x) = 0$ otherwise. Thus, in this case Axiom 2 postulates that adding dummy elements to the set $X$ has no influence on $U_C$ values. Axiom 3 says that the conflict is not increases after a mapping. Notice that such a mapping can produce a loss of information, when two separate elements can map to the same element in $Y$. Axiom 4 shows the behavior of $U_C$ after merging sources of information using the mixture rule of aggregation.

In [6], a reader can find the theorem that fully characterizes a system of functions that defines $U_C$. In this paper, we only give some examples of $U_C$ discussed in [6].

Example 1. Let $f : [0,1] \rightarrow [0, +\infty)$ be a concave function with the following properties:

1) $f(0) = f(1) = 0$;
2) $f(t)$ is strictly decreasing at $t = 1$. 
Then \( U_C(P) = \sum_{x \in X} f(P(x)), \) where \( P \in M_{pr}(X), \) is a measure of conflict on \( M_{pr}. \)

Some notable examples of this class of conflict measures are the Shannon entropy, when \( f(t) = -t \log_2 t, \) and the Gini index, when \( f(t) = t(1-t). \)

**Example 2.** The functional \( U_C, \) defined by \( U_C(P) = 1 - \max_{x \in X} P(x), \) where \( P \in M_{pr}(X), \) is a conflict measure on \( M_{pr}. \)

We will establish later the connections between the conflict measure from Example 2 and the Rényi entropy of order infinity and other functionals for measuring conflict within belief functions.

### 4 Distances and Entropies in Probability Theory

Although, the Shannon entropy and that the Kullback-Leibler divergence (also called Kullback-Leibler distance or relative entropy) are very popular in probability theory, one can find many other functionals \([15,16,18,19]\) that can be used to measure conflict within and between probability distributions. It is important to say that distances (or more exactly statistical distances) in probability theory are not distances as a reader would expect. They do not always obey the triangular inequality and can be non-symmetric. Such statistical distances measure the conflict (or contradiction) between a prior probability distribution of a random variable and its posterior distribution. We will also consider another possible interpretation of contradiction in Sect. 4. The aim of this section is to illustrate of how such distances or divergences can generate entropies or conflict measures on \( M_{pr}. \)

We postulate that a statistical distance or divergence is the functional \( D: M_{pr} \times M_{pr} \rightarrow [0, +\infty] \) that satisfies the following axioms.

**Axiom 5.** \( D(P_1, P_2) = 0 \) for \( P_1, P_2 \in M_{pr}(X) \) iff \( P_1 = P_2. \)

**Axiom 6.** \( D(P_1, P_2) \in [0, +\infty) \) for \( P_1, P_2 \in M_{pr}(X) \) if \( P_1 \) is absolutely continuous w.r.t. \( P_2. \)

**Axiom 7.** Let \( \varphi: X \rightarrow Y \) be an injection, then \( D(P_1 \varphi, P_2 \varphi) = D(P_1, P_2) \) for every \( P_1, P_2 \in M_{pr}(X). \)

**Axiom 8.** Let \( P_u \) define the uniform probability distribution on \( 2^X, \) then

\[
\sup_{P \in M_{pr}(X)} D(P, P_u) = \max_{x \in X} D(P_x, P_u).
\]

**Axiom 9.** Let \( \varphi_n : X \rightarrow X_n \) be an injection, in which \( X = \{x_1, ..., x_m\}, \) \( X_n = \{x_1, ..., x_n\}, n > m, \) \( \varphi_n(x) = x \) for all \( x \in X, \) and let \( P_u^{(n)} \in M_{pr}(X_n) \) define the uniform probability distribution on \( X_n. \) Then the functional

\[
U_C(P) = \lim_{n \to \infty} \left( \sup_{P' \in M_{pr}(X)} D(P', P_u^{(n)}) - D(P^\varphi_n, P_u^{(n)}) \right), P \in M_{pr}(X),
\]

is the conflict measure on \( M_{pr}. \)
Let us discuss the introduced axioms. Axiom 5 reflects the behavior of $D$ like a distance or divergence. Axiom 6 allows us to cover the case of the Shannon entropy and the mentioned above interpretation of divergence. Axiom 7 is the similar to Axiom 2 for a measure of conflict. Axiom 8 postulates that the greatest distance between $P_u$ (that symbolizes the highest uncertainty in $M_{pr}(X)$) and a $P \in M_{pr}(X)$ is achieved on $P_x$ (the case, when uncertainty is absent). Axiom 9 establishes the main definition of $U_C$ through $D$.

We will show several choices of divergences satisfying Axioms 5–9 described in the next subsections.

4.1 Kullback-Leibler Divergence

We remind that the Kullback-Leibler divergence (distance) [16,19] is defined for probability measures $P_1, P_2 \in M_{pr}(X)$ by

$$D_1(P_1, P_2) = \sum_{x \in X} P_1(x) \log_2 \left( \frac{P_1(x)}{P_2(x)} \right).$$

Computing $U_C$ with $D = D_1$ and $|X| = n$, we get

$$H_1(P) = \lim_{n \to \infty} \left[ \log_2 n - \sum_{x \in X} P_1(x) \log_2 (nP_1(x)) \right] = -\sum_{x \in X} P_1(x) \log_2 P_1(x).$$

We see that $H_1$ is the Shannon entropy. We don’t check Axioms 5–9 for $D_1$, because Axioms 5–7 are well-known properties of the Kullback-Leibler divergence [16,19], Axiom 8 follows from the fact that $D(P, P_u)$ is a convex function of $P$.

4.2 Rényi Divergence

The Rényi divergence [16,19] is the parametrical generalization of the Kullback-Leibler divergence with the parameter $\alpha \in [0, +\infty]$ defined as

$$D_\alpha(P_1, P_2) = \frac{1}{\alpha - 1} \log_2 \left( \sum_{x \in X} \frac{P_1^\alpha(x)}{P_2^{\alpha-1}(x)} \right) \text{ for } \alpha \neq 0, 1, +\infty.$$

For special cases, when $\alpha = 0, 1, +\infty$, $D_\alpha$ is defined by taking the limit on $\alpha$. If $\alpha \to 1$, then we get the Kullback-Leibler divergence. Analogously,

$$D_0(P_1, P_2) = -\log_2 P_2(A), \text{ where } A = \{x \in X | P_1(x) > 0\},$$

$$D_\infty(P_1, P_2) = \max_{x \in X} \log_2 \frac{P_1(x)}{P_2(x)}.$$

The computation of $U_C$ for $D = D_\alpha$ with $\alpha \neq 0, 1, +\infty$ gives us the result

$$H_\alpha(P) = -\frac{1}{\alpha - 1} \log_2 \left( \sum_{x \in X} P_\alpha(x) \right).$$
H_\alpha is called the Rényi entropy of order \alpha. Let us consider also the special cases \alpha = 0 and \alpha = \infty. Substituting D by D_0 in the expression of U_C, we get the result

\[ H_0(P) = \log_2 |A|, \text{ where } A = \{x \in X | P(x) > 0\}. \]

In this case, \( H_0 \) is called the Hartley entropy. Computing \( U_C \) with \( D = D_\infty \), we get result

\[ H_\infty(P) = -\log_2 \left( \max_{x \in X} P(x) \right), \]

and \( H_\infty \) is called the min-entropy. Again, we do not check axioms, because they follow from the known properties of the Rényi divergence and the Rényi entropy.

4.3 \textit{g}-divergence

Let \( g : [0, 1] \rightarrow [0, 1] \) be a convex, strictly increasing and continuous function on [0,1] such that \( g(0) = 0 \) and \( g(1) = 1 \). Then the \textit{g}-divergence is the functional

\[ D_g(P_1, P_2) = \sum_{x \in X} g(|P_1(x) - P_2(x)|), \]

where \( P_1, P_2 \in M_{pr}(X) \). Let us compute the functional \( U_C \) generated by \( D_g \). Assume that \( X_n = \{x_1, ..., x_n\} \) and \( P_u^{(n)} \) defines the uniform probability distribution on \( 2^X \). Then \( \sup_{P' \in M_{pr}(X_n)} D(P', P_u^{(n)}) = D(P, P_u^{(n)}) \) for every \( x \in X_n \) and, for our case,

\[ D_g(P_x, P_u) = g(1 - 1/n) + (n - 1)g(1/n). \]

Without decreasing generality, consider \( P \in M_{pr}(X) \), where \( X = \{x_1, ..., x_m\} \), such that \( P(x_i) > 0, i = 1, ..., m \). Let us choose \( n \) such that \( P(x_i) > 1/n, i = 1, ..., m \). Then

\[ D_g(P_{\phi^n}, P_u^{(n)}) = \sum_{i=1}^{m} g(P(x_i)) + (n - m)g(1/n), \]

and

\[ U_C(P) = \lim_{n \to \infty} \left( g(1 - 1/n) + (n - 1)g(1/n) - \sum_{i=1}^{m} g(P(x_i)) - (n - m)g(1/n) \right) \]

\[ = \lim_{n \to \infty} \left( g(1 - 1/n) + (m - 1)g(1/n) - \sum_{i=1}^{m} g(P(x_i)) \right) = 1 - \sum_{i=1}^{m} g(P(x_i)). \]

Let us denote

\[ H_g(P) = 1 - \sum_{x \in X} g(P(x)), \]

where \( P \in M_{pr}(X) \), and call it the \textit{g}-entropy. Next result directly follows from Example 1.
Proposition 1. $H_g$ is a conflict measure on $M_{pr}$ if the function $f(t) = t - g(t)$, $t \in [0, 1]$, is strictly decreasing at $t = 1$.

Proof. Obviously, the expression of $H_g$ can be rewritten using the function $f$ as $H_g(P) = \sum_{x \in X} f(P(x))$. We see that $f(0) = 0$, $f(1) = 1$, $f$ is a concave function on $[0, 1]$, i.e. the all necessary conditions are fulfilled for $H_g$ to be a conflict measure.

Theorem 1. $D_g$ satisfies Axioms 5–9 on $M_{pr} \times M_{pr}$ if the conditions of Proposition 1 are fulfilled.

Proof. Let us check axioms. We see that the truth of Axioms 5–7 follows from the definition of $D_g$. Let us check Axiom 8. It is easy to see that $D_g(P, P_u)$ is a convex function of $P$, i.e. $D_g(aP_1 + (1-a)P_2, P_u) \leq aD_g(P_1) + (1-a)D_g(P_2, P_u)$, for every $a \in [0, 1]$, and $P_1, P_2 \in M_{pr}(X)$. Thus, $\sup_{P \in M_{pr}(X)} D_g(P, P_u)$ is achieved on extreme points of $M_{pr}(X)$, i.e.

$$\sup_{P \in M_{pr}(X)} D_g(P, P_u) = \max_{x \in X} D_g(P_x, P_u).$$

Axiom 9 follows from Proposition 1. The theorem is proved.

Let us analyze the range of $D_g$. We see that $D_g(P_1, P_2)$, where $P_1, P_2 \in M_{pr}(X)$, is convex on both arguments $P_1$ and $P_2$, therefore, the maximum is achieved on extreme points of $M_{pr}(X)$ and

$$\sup_{P_1, P_2 \in M_{pr}(X)} D_g(P_1, P_2) = \max_{x, y \in X} D_g(P_x, P_y) = 2.$$ 

In some cases, it is convenient that $D_g$ should be normalized. Then we use the normalized $D_g$ defined by

$$D_g(P_1, P_2) = 0.5 \sum_{x \in X} g(|P_1(x) - P_2(x)|).$$

The value $D_g(P_1, P_2) = 1$ manifests the highest contradiction between $P_1$ and $P_2$. As a rule, in information theory, entropies are normalized. This means that $H(P_u) = 1$ for an entropy $H : M_{pr} \to [0, +\infty)$, where $P_u$ defines the uniform probability distribution on $X$ and $|X| = 2$. Thus, we can introduce the normalized entropy $H_g(P)$ as

$$H_g(P) = \frac{1 - \sum_{x \in X} g(P(x))}{1 - 2g(0.5)}.$$

4.4 Divergence and Entropy Based on $L_\alpha$-metric

Let us consider the case, when $g(x) = x^\alpha$, where $\alpha > 1$, then we denote $D_g$ by $D_{L_\alpha}$ and

$$D_{L_\alpha}(P_1, P_2) = 0.5 \sum_{x \in X} |P_1(x) - P_2(x)|^\alpha, P_1, P_2 \in M_{pr}(X).$$
We use this notation, because \( d_{L_\alpha}(P_1, P_2) = \left( 2D_{L_\alpha}(P_1, P_2) \right)^{1/\alpha} \) is the \( L_\alpha \)-metric on \( M_{pr} \). In this case, the corresponding \( H_g \) for \( g(x) = x^\alpha \) is

\[
H_{L_\alpha}(P) = \frac{1 - \sum_{x \in X} |P(x)|^{\alpha}}{1 - 2^{1-\alpha}}.
\]

Consider the case \( \alpha = 1 \). We see that

\[
D_{L_1}(P_1, P_2) = 0.5 \sum_{x \in X} |P_1(x) - P_2(x)|,
\]

and we can find the expression for \( H_{L_1} \) taking the limit for \( \alpha \to 1 \). Using l’Hôpital’s rule, we get

\[
H_{L_1}(P) = - \sum_{x \in X} \frac{\ln(P(x)) (P(x))^{\alpha}}{2^{1-\alpha} \ln 2} = - \sum_{x \in X} P(x) \log_2 P(x).
\]

Thus, \( H_{L_1}(P) \) is the Shannon entropy.

### 4.5 Concluding Remarks

We see that any \( g \)-divergence has the properties that any metric has, and the entropies generated by \( D_{L_\alpha}, \alpha \geq 1 \), looks identical to the Rényi entropy \( H_\alpha \) of order \( \alpha \):

\[
H_\alpha(P) = - \frac{1}{\alpha - 1} \log_2 \left( \sum_{x \in X} P^\alpha(x) \right),
\]

\[
H_{L_\alpha}(P) = \frac{1}{1 - 2^{1-\alpha}} \left( 1 - \sum_{x \in X} P^\alpha(x) \right), \alpha > 0;
\]

\[
H_{L_1}(P) = H_1(P) = - \sum_{x \in X} P(x) \log_2 P(x), \alpha = 0.
\]

We see that \( H_{L_1} = H_1 \) is the Shannon entropy. If we denote \( t = \sum_{x \in X} P^2(x) \), then \( H_\alpha(P) = - \frac{1}{1-\alpha} \log_2 t \) and \( H_{L_\alpha}(P) = \frac{1}{1-2^{1-\alpha}} (1 - t) \). We see that both functions \( \varphi_1(t) = - \frac{1}{1-\alpha} \log_2 t \) and \( \varphi_2(t) = \frac{1}{1-2^{1-\alpha}} (1 - t) \) are decreasing on \([0, 1]\), therefore, \( H_\alpha \) and \( H_{L_\alpha} \) similarly discriminate uncertainty. Formally, we can take the \( L_\alpha \)-metric

\[
d_{L_\alpha}(P_1, P_2) = \left( \sum_{x \in X} |P_1(x) - P_2(x)|^\alpha \right)^{1/\alpha}, \alpha \geq 1,
\]

as a divergence. The divergence \( d_{L_\alpha} \) generates the entropy

\[
h_{L_\alpha}(P) = \frac{1}{1 - 2^{1-\alpha}} \left( 1 - \left( \sum_{x \in X} P^\alpha(x) \right)^{1/\alpha} \right), \alpha > 1,
\]
that is identical to the Shannon entropy if $\alpha \to 1$. We can also define $d_{L_\alpha}(P_1, P_2)$ and $h_{L_\alpha}(P)$ if $\alpha \to +\infty$. Then

$$d_{L_\infty}(P_1, P_2) = \max_{x \in X} |P_1(x) - P_2(x)|, h_{L_\infty}(P) = 2 \left(1 - \max_{x \in X} P(x)\right).$$

5 Uncertainty Measures on Credal Sets

The main idea is to use divergences or distances on probability measures and express through them other measures of uncertainty. As it was established by many papers (see [2] and the references there), imprecise probabilities describe two types of uncertainty: non-specificity and conflict. A pure conflict is described by the classical probability theory, non-specificity corresponds to the choice of a probability model among possible ones. It is also important to evaluate the degree of total uncertainty that aggregates conflict and non-specificity. In previous sections, we have described the choice of two functionals $U_C : \mathcal{M}_{pr} \to [0, +\infty)$ and $D : \mathcal{M}_{pr} \times \mathcal{M}_{pr} \to [0, +\infty]$. The functional $U_C$ evaluates the amount conflict if uncertainty in information described by a probability measure, and $D$ describes the contradiction between two sources of information in a probabilistic setting. It is important to distinguish two possible interpretations of divergence. The first one related to the Rényi and Kullback-Leibler divergences that evaluate the contradiction between a prior probability distribution and a posterior one. Thus, in this case $D$ is not symmetrical, $D(P_1, P_2) = +\infty$ iff $P_1$ is not absolutely continuous w.r.t. $P_2$. If $D$ is a $g$-divergence or the $L_\alpha$-metric, then $D$ can be used for evaluating difference between probability models, or for finding the degree of how information obtained from two sources is the same. Thus, values of $U_C$ can be understood differently. In the case of the Rényi and Kullback-Leibler divergences, the part $D(P^{\varphi_n}, P_u(n))$ of $U_C^{(1)}(P)$ gives us the amount of information gain after obtaining $P^{\varphi_n}$ w.r.t. the most uncertain information described by $P_u(n)$ and the part $\sup_{P' \in \mathcal{M}_{pr}(X_n)} D(P', P_u(n))$ gives us the maximal information gain, therefore, the difference $\sup_{P' \in \mathcal{M}_{pr}(X_n)} D(P', P_u(n)) - D(P^{\varphi_n}, P_u(n))$ evaluates the deficiency of information in $P^{\varphi_n}$. If $D$ is understood as a distance, then $\sup_{P' \in \mathcal{M}_{pr}(X_n)} D(P', P_u(n))$ is the distance between exact and the most uncertain information, and $D(P^{\varphi_n}, P_u(n))$ characterizes how far $P^{\varphi_n}$ is located from $P_u(n)$.

Assume that $P \in Cr(X)$ and $P \in \mathcal{M}_{pr}(X)$. Let us introduce the following functionals:

$$\Phi_1(P, P) = \inf\{D(P', P)|P' \in P\}, \Phi_2(P, P) = \sup\{D(P', P)|P' \in P\}.$$

We see that $\Phi_1(P, P)$ and $\Phi_2(P, P)$ give the smallest and highest information gains if the posterior probability distribution is described by the credal set $P$. According to the Laplace principle the prior information, described by $P$, has the highest uncertainty if $P = P_u$, where $P_u$ is the uniform probability distribution.
on $2^X$. Thus, $\Phi_1(P, P_u)$ gives us the amount of information in $P$. Using the same logic as before, we define the measure $U_T$ of total uncertainty on $Cr$ as

$$U_T(P) = \lim_{n \to \infty} \left( \sup_{P' \in M_{pr}(X_n)} D(P', P_u^{(n)}) - \inf \{ D(P^\varphi_n, P_u^{(n)})| P \in P \} \right),$$

where $P \in Cr(X)$ and we use notations and definitions from Axiom 9. If $P = \{P\}$, then the expression for $U_T(P)$ is simplified to

$$U_T(P) = \lim_{n \to \infty} \left( \sup_{P' \in M_{pr}(X_n)} D(P', P_u^{(n)}) - D(P^\varphi_n, P_u^{(n)}) \right),$$

i.e. $U_T$ is the conflict measure $U_C$ defined in Axiom 9. We see that $U_T$ on $Cr$ can be seen as an extension of $U_T$ on $M_{pr}$ by the following formula:

$$U_T(P) = \sup \{ U_T(P) | P \in P \}.$$  

Observe that the functional $U_T$ is called the upper or maximal entropy \cite{[2,14]} if $U_T$ is the Shannon entropy on $M_{pr}$. Let us analyze the functional

$$\Phi(P) = \sup \{ D(P, P_u)| P \in P \} - \inf \{ D(P, P_u)| P \in P \},$$

where $P \in Cr(X)$ and $P_u$ is the uniform probability distribution on $2^X$. We see that it characterizes the amount of uncertainty in choosing a true probability measure in $P$. Thus, we can choose as a measure of non-specificity the functional

$$U_N(P) = \lim_{n \to \infty} \left( \sup \{ D(P^\varphi_n, P_u^{(n)})| P \in P \} - \inf \{ D(P^\varphi_n, P_u^{(n)})| P \in P \} \right),$$

where $P \in Cr(X)$ and we use notations and definitions from Axiom 9. Assume that the measure of total uncertainty aggregates non-specificity and conflict additively, i.e. for every $P \in Cr$

$$U_C(P) + U_N(P) = U_T(P),$$

where $U_C$ is a measure of conflict on $Cr$. This assumption implies that

$$U_C(P) = \lim_{n \to \infty} \left( \sup_{P' \in M_{pr}(X_n)} D(P', P_u^{(n)}) - \sup \{ D(P^\varphi_n, P_u^{(n)})| P \in P \} \right),$$

and we can express $U_C(P)$ through its values on $M_{pr}$ as

$$U_C(P) = \inf \{ U_C(P) | P \in P \}.$$ 

Note that $U_C$ is called the minimal or lower entropy \cite{[2,14]} on $Cr$ if $U_C$ is the Shannon entropy on $M_{pr}$. 
6 Discussion and Conclusion

Let us notice that the axioms for the Rényi divergence can be found in [16]. Although, in this paper, the author considers these axioms as evident ones, the interpretation of some of them seems to be problematic, because they are based on so-called generalized probability distributions that are not necessarily normalized. In our approach, divergences and entropies are not necessarily additive, that allows, for example, to use $L_\alpha$-metrics as such divergences.

The results of the paper allow us to resolve some problems in the theory of belief functions. For example, in this theory [17] the conflict between two sources of information is evaluated using Dempster’s rule of aggregation. If sources of information are described by probability measures $P_1, P_2 \in M_{pr}(X)$, then this evaluation is produced by the formula:

$$k(P_1, P_2) = 1 - \sum_{x \in X} P_1(x)P_2(x).$$

We see that $k(P, P) = 1 - \sum_{x \in X} P(x)P(x) = H_{L_2}(P)$ for $P \in M_{pr}(X)$, i.e. $k(P, P)$ is the Gini index that can be interpreted as entropy or conflict measure. Another representation

$$D_{L_2}(P_1, P_2) = \sum_{x \in X} (P_1(x) - P_2(x))^2 = -H_{L_2}(P_1) - H_{L_2}(P_2) + 2k(P_1, P_2)$$

implies that $k(P_1, P_2) = 0.5(H_{L_2}(P_1) + H_{L_2}(P_2) + D_{L_2}(P_1, P_2))$ consists of two parts: the part $0.5(H_{L_2}(P_1) + H_{L_2}(P_2))$ measures conflict within information sources and the part $0.5D_{L_2}(P_1, P_2)$ measures conflict (contradiction) between information sources. Let us notice that this problem of disaggregating of $k$ on two parts for belief functions is investigated in [11].

We also pay attention on using $h_{L_\infty}$ for defining $U_C$ on credal sets. In this case,

$$U_C(P) = \inf\{h_{L_\infty}(P) | P \in \mathbf{P}\} = 2 \left(1 - \max_{x \in X} \sup\{P(x) | P \in \mathbf{P}\}\right).$$

If a credal set $\mathbf{P}$ is described by a coherent upper probability $\mu$ on $2^X$, then the expression for $U_C$ can be rewritten as $U_C(\mu) = 2 \left(1 - \max_{x \in X} \mu(x)\right)$. Such a functional (without the factor 2) is proposed for measuring inner conflict in the theory of belief functions [10].

Formally, in this paper we propose to analyze uncertainty by choosing a divergence $D$, and then to take the compatible with $D$ measures of uncertainty. Let us discuss our approach in detail if we choose $d_{L_1}$ as divergence. We see that we can use $d_{L_1}$ with the Shannon entropy. Let us also look at the following functional:

$$U'_C(P) = \min_{x \in X} D(P_x, P), \text{ where } P \in M_{pr}(X).$$
We see that $U'_C(P)$ evaluates the distance between $P$ and the exact information without uncertainty, thus, it can be considered as a candidate for the measure of conflict. If we take $D = d_{L_1}$, then it is possible to show that

$$d_{L_1}(P_1, P_2) = 2 \left(1 - \sum_{x \in X} \min \{P_1(x), P_2(x)\}\right).$$

This implies that

$$U'_C(P) = 2 \min_{x \in X} (1 - P(x)) = 2 \left(1 - \max_{x \in X} P(x)\right) = h_{L_\infty}(P).$$

Doing in the same way for $D = D_\alpha$, we get that $U'_C$ coincides in this case with the min-entropy.

Let us consider the functional

$$Con(P_1, P_2) = \inf \{D(P_1, P_2) | P_1 \in P_1, P_2 \in P_2\}, P_1, P_2 \in Cr(X).$$

We see that $Con(P_1, P_2) = 0$ iff $P_1 \cap P_2 = \emptyset$, and $Con$ can be used for measuring contradiction between information sources described by credal sets. This measure of contradiction for $D = d_{L_1}$ is well-known in the theory of belief functions [3,8,13], its axiomatic can be found in [4,7] and its extension to imprecise probabilities based on generalized credal sets is given in [5,7].

Finalizing our paper, we can conclude that there is the variety of divergences and the corresponding uncertainty measures. The choice one of them can depend on the problem statement or on the complexity of realization, or on their additional desirable properties. We aware that in this paper we do not investigate in detail ways of evaluating conflict, non-specificity and contradiction in information presented by credal sets with the help of divergences. This can be the topic of our next research.

References

1. Augustin, T., Coolen, F.P.A., de Cooman, G., Troffaes, F.C.M. (eds.): Introduction to Imprecise Probabilities. Wiley Series in Probability and Statistics. Wiley, New York (2014)
2. Bronevich, A.G., Klr, G.J.: Measures of uncertainty for imprecise probabilities: an axiomatic approach. Int. J. Approx. Reason. 51, 365–390 (2010)
3. Bronevich, A.G., Rozenberg, I.N.: The choice of generalized Dempster-Shafer rules for aggregating belief functions. Int. J. Approx. Reason. 56(Part A), 122–136 (2015)
4. Bronevich, A.G., Rozenberg, I.N.: Conjunctive rules in the theory of belief functions and their justification through decisions models. In: Vejnarová, J., Kratochvíl, V. (eds.) BELIEF 2016. LNCS (LNAI), vol. 9861, pp. 137–145. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45559-4_14
5. Bronevich, A.G., Rozenberg, I.N.: Modelling uncertainty with generalized credal sets: application to conjunction and decision. Int. J. Gen. Syst. 27(1), 67–96 (2018)
6. Bronevich, A.G., Lepskiy, A.E.: Clustering a body of evidence based on conflict measures. In: Štěpnička, M. (ed.) Proceedings of the 11th Conference of the European Society for Fuzzy Logic and Technology, vol. 1, pp. 328–333, Atlantis Press, Paris (2019). https://doi.org/10.2991/eusflat-19.2019.47
7. Bronevich, A.G., Rozenberg, I.N.: The contradiction between belief functions: its description, measurement, and correction based on generalized credal sets. Int. J. Approx. Reason. 112, 119–139 (2019)
8. Cattaneo, M.E.G.V.: Combining belief functions issued from dependent sources. In: Bernard, J.-M., Seidenfeld, T., Zaffalon, M. (eds.), ISIPTA 2003: Proceedings in Informatics, vol. 18, pp. 133–147, Carleton Scientific, Waterloo (2003). https://doi.org/10.3929/ethz-a-004531249
9. Csiszár, I.: Axiomatic characterizations of information measures. Entropy 10, 261–273 (2008)
10. Daniel, Milan: Conflicts within and between belief functions. In: Hüllermeier, Eyke, Kruse, Rudolf, Hoffmann, Frank (eds.) IPMU 2010. LNCS (LNAI), vol. 6178, pp. 696–705. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14049-5_71
11. Daniel, Milan: Conflict between belief functions: a new measure based on their non-conflicting parts. In: Cuzzolin, Fabio (ed.) BELIEF 2014. LNCS (LNAI), vol. 8764, pp. 321–330. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11191-9_35
12. Dempster, A.P.: Upper and lower probabilities induced by multivalued mapping. Ann. Math. Stat. 38, 325–339 (1967)
13. Destercke, S., Burger, T.: Toward an axiomatic definition of conflict between belief functions. IEEE Trans. Cybern. 43(2), 585–596 (2013)
14. Klir, G.J.: Uncertainty and Information: Foundations of Generalized Information Theory. Wiley, Hoboken (2006)
15. Morales, D., Pardo, L., Vajda, I.: Uncertainty of discrete stochastic systems: general theory and statistical theory. IEEE Trans. Syst. Man Cybern. 26, 1–17 (1996)
16. Rényi, A.: On measures of entropy and information. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Contributions to the Theory of Statistics, vol. 1, pp. 547–561, University of California Press, Berkeley (1961)
17. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
18. Taneja, I.J.: Generalized symmetric divergence measures and the probability of error. Commun. Stat. - Theory Methods 42(9), 1654–1672 (2013)
19. Van Erven, T., Harremos, P.: Rényi divergence and Kullback-Leibler divergence. IEEE Trans. Inf. Theory 60(7), 3797–3820 (2014)
20. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)
21. Wang, Z., Klir, G.J.: Generalized Measure Theory. Springer, Heidelberg (2009). https://doi.org/10.1007/978-0-387-76852-6