Possibilistic Entropy Functions

Kramosil, Ivan
2007

Institute of Computer Science
Academy of Sciences of the Czech Republic

Possibilistic entropy functions

Ivan Kramosil

Technical report No. 1001

Pod Vodárenskou věží 2, 182 07 Prague 8, phone: +420 266 051 111, fax: +420 286 585 789,
e-mail:ics@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

Possibilistic entropy functions

Ivan Kramosil[1]

Technical report No. 1001

Abstract:

Given a real-valued possibilistic distribution π on a universe Ω of elementary events, the possibilistic entropy functions of this distribution is defined as the expected value of the complement function $1 - \pi$ over Ω, this expected value being given by the Sugeno integral of $1 - \pi$ over Ω. Applied to possibilistic decision making under uncertainty, possibilistic distribution with maximum entropy value plays the same role as the uniform probability distribution in bayesian statistical decision making. The entropy value defined by possibilistically independent products of finite or infinite systems of possibilistic distributions is proved to be identical with the supremun value of the entropies defined by particular possibilistic distributions.

Keywords:
Real-valued possibilistic distribution, possibilistically independent product of distributions, possibilistic entropy function, decision making under uncertainty, worst-case (minimax) principle

[1] Ivan Kramosil
Institute of Computer Science AS CR, v.v.i., Pod Vodárenskou věží 2, 182 07 Prague, Czech Republic, Fax: +420-268-585-789, E-mail: kramosil@cs.cas.cz

1 Introduction and Motivation

The notion of entropy plays on important role in the standard information theory and statistical decision functions theory based on Kolmogorov axiomatic probability theory. In this paper, our aim will be to suggest, if possible, an analogy of entropy within the framework of possibilistic measures and to analyze, whether the modified notion can play a role similar to that one played by the standard notion of entropy. First of all, let us sketch, very briefly, the probabilistic notion of entropy, restricting ourselves to the case of probabilities over finite spaces of elementary random events, as this approach seems to be quite sufficient for the purposes of this introductory section.

Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ be a finite set. An n-tuple $P = \langle p(\omega_1), p(\omega_2), \ldots, p(\omega_n)\rangle$, $p(\omega_i) \in [0,1]$ for each $i = 1, \ldots, n$, is called a *probabilistic distribution* on Ω, if $\sum_{i=1}^{n} p(\omega_i) = 1$ holds, we will write p_i for $p(\omega_i)$. For each $A \subset \Omega$, the *probability* $P_*(A)$ of A, induced by P, is defined by $P_*(A) = \sum_{\omega \in A} p(\omega)$. The *(total amount of) information* $I(P)$ contained in the probabilistic distribution P on Ω, and also called the *entropy* of P, is the real number defined by

$$I(P) = \sum_{i=1}^{n} (lg(1/p_i))p_i = -\sum_{i=1}^{n} (lg\, p_i)p_i, \tag{1.1}$$

where lg denotes the binary logarithm (logarithm to the base 2) and, for $p_i = 0$, the value $p_i\, lg\, p_i$ is taken as 0 by convention.

The intuition behind may be as follows. The amount of information obtained by a random sample giving ω_i, i.e., the result with probability p_i, is quantified by the real number $lg(1/p_i)$. It is a decreasing function of p_i, hence, the smaller is the probability of the obtained results (the more surprising is this result, we could say), the greater is the amount of information obtained. So, the total amount of information or the entropy $I(P)$ related to the probability distribution $P = \langle p_1, \ldots, p_n\rangle$ is then defined by the expected value of the function $lg(1/p)$ with respect to the probability distribution P.

The particular choice of the function $lg(1/p)$ is motivated by the well-known fact of elementary information theory according to which for the statistically independent product of two or more probabilistic distributions on Ω the entropy of the resulting probabilistic distribution on the corresponding Cartesian product Ω^n is the sum of the entropies of the particular probabilistic distributions on Ω. More formally, if $n = 2$ and $P^i = \langle p_1^i, \ldots, p_n^i\rangle, i = 1, 2$, are probabilistic distribution on Ω, and if P^{12}, $\Omega \times \Omega \to [0,1]$ is defined by $p_{i,j}^{12} = p^{12}(\omega_i, \omega_j) = p^1(\omega_i)p^2(\omega_j) = p_i^1 p_j^2$, then P^{12} obviously defines a probability distribution and $I(P^{12}) = I(P^1) + I(P^2)$.

However, because of our intention to modify the notion of entropy to the case when the phenomenon of uncertainty is quantified and processed by the tools offered by possibilistic measures (real-valued as well as lattice-valued ones) also other decreasing functions of the probability value are worth being considered, let us take the function $1 - p$.

It is a well-known fact of elementary mathematical analysis and information theory that $I(P)$ takes its maximum value $lg(n)$ iff P is the uniform probability distribution $P_0 = \langle 1/n, 1/n, \ldots, 1/n\rangle$. Interesting enough, the same is the case for the entropy I^* defined by

$$I^*(P) = \sum_{i=1}^{n} (1 - p_i)p_i = 1 - \sum_{i=1}^{n} p_i^2 \tag{1.2}$$

when the maximum value $1 - (1/n)$ is also taken iff $P = P_0$. Indeed, given $P = \langle p_1, \ldots, p_n\rangle$ and setting $\varepsilon_i = (1/n) - p_i$ for each $i = 1, \ldots, n$, so that $\sum_{i=1}^{n} \varepsilon_i = 0$, we obtain that

$$\sum_{i=1}^{n} p_i^2 = \sum_{i=1}^{n} ((1/n) - \varepsilon_i)^2 = \sum_{i=1}^{n} [(1/n)^2 - (2/n)\varepsilon_i + \varepsilon_i^2] = (1/n) + \sum_{i=1}^{n} \varepsilon_i^2 \geq 1/n, \tag{1.3}$$

so that $I^*(P) \leq 1 - (1/n)$ follows.

As can be easily seen, both I and I^* take their minimal value 0 iff $P = \langle p_1, \ldots, p_n\rangle$ is "degenerated" in the sense that $p_i = 1$ for some $i \leq n$, hence, $p_j = 0$ for every $j \neq i$. Indeed, if $p_i > 0$, $p_j > 0$ holds for $i \neq j$, then $lg(1/p_i) > 0$, $lg(1/p_j) > 0$ follows, so that $I(P) \geq p_i\, lg(1/p_i) + p_j\, lg(1/p_j) > 0$ results. Moreover, in this case $\sum_{i=1}^{n} p_i^2 < \sum_{i=1}^{n} p_i = 1$ holds, so that $I^*(P) = 1 - \sum_{i=1}^{n} p_i^2 > 0$ is valid.

2 Laplace Principle and Maximum Entropy

The uniform probability distribution P_0 may enter the scene also when applying the so called Laplace principle to statistical decision making under uncertainty. In its most general philosophical setting this principle declares that two or more alternative solutions (to a decision problem, say) can be taken as equivalent (equally good, acceptable,...) with respect to the criterion under consideration (profit or loss resulting when applying this or that solution), if we have not at hand any argument in favor of the hypothesis that one of these alternatives is better than the other one(s). Hence, considering a system with possible internal states $\omega_1, \omega_2, \dots, \omega_n$, just one ω_i being the actual one, supposing that this actual internal state results from a random sample (taken by God or the Nature, say), but not having at hand any arguments in favor of the idea that one state is more probable than the other ones, the Laplace principle brings us to the conclusion that P_0 should be taken as our hypothesis concerning the apriori probabilities of particular states. This approach is often applied in the so called Bayesian statistical decision functions when some apriori probabilistic distribution on states is a necessary part of the mathematical model, but we have no idea how to specify this distribution.

The choice of $P_0 = \langle 1/n, \dots, 1/n \rangle$ as the apriori probability distribution on Ω is rational from the following minimax or worst-case point of view. Consider the most pessimistic situation when we have no relevant information (neither a statistical one) enabling to specify, at least partially, the actual internal state, but we still have to take a decision which of $\omega_1, \omega_2, \dots, \omega_n$ is the case. Let us also assume that the loss $\lambda(\omega_i, \omega_j)$, suffered when $\omega_{\ddot{\imath}}$ is the actual internal state and ω_j is our decision, is defined in the most simple way: $\lambda(\omega_i, \omega_j) = 0$, if $i = j$, $\lambda(\omega_i, \omega_j) = 1$, if $\ddot{\imath} \neq j$. Our decision making will consist in the random sampling from Ω using a probability distribution $P = \langle p_1, p_2, \dots, p_n \rangle$, hence, the expected loss suffered when ω_i is the actual state is given by the value $P_*(\{\omega \in \Omega : \omega \neq \omega_i\}) = 1 - p_i$. Applying the worst-case principle we take the value $\max_{i=1}^{n}\{1 - p_i\} = 1 - \min_{i=1}^{n} p_i$ as the minimax quality criterion of the decision procedure based on P. This value takes the optimal, i.e., the minimum value just when $P = P_0$, the distribution with the maximum entropy value $I(P)$ as well as $I^*(P)$.

3 Real-Valued Possibilistic Measures and Entropies

In what follows, we will investigate an alternative model of uncertainty quantification and processing based on possibilistic distributions and measures. In particular, we will seek for analogies of the notions of entropy, Laplace principle and uniform probability distribution, if any, analyzing in more detail their properties.

Let Ω be a nonempty set. A mapping $\pi : \Omega \to [0,1]$ is called *(real-valued) possibilistic distribution* on Ω, if $\bigvee_{\omega \in \Omega} \pi(\omega) = 1$ holds, here and below $\bigvee, \vee (\bigwedge, \wedge$, resp.) denotes the supremum (infimum, resp.) operation defined in $[0,1]$ by the standard linear ordering $\leq$. The *(real-valued) possibilistic measure* induced by π is the mapping $\Pi : \mathcal{P}(\Omega) \to [0,1]$ such that $\Pi(A) = \bigvee_{\omega \in A} \pi(\omega)$ for each $\emptyset \neq A \subset \Omega, \Pi(\emptyset) = 0$ by convention. Given a mapping $f : \Omega \to [0,1]$, the expected value $Ef(\cdot)d\Pi$ of f w.r.t. Π and w.r.to the t-norm $t = \wedge$ (infimum) on $[0,1] \times [0,1]$ is defined by

$$Ef(\cdot)d\Pi = \bigvee_{t \in [0,1]} [t \wedge \Pi(\{\omega \in \Omega : f(\omega) \geq t\})]. \tag{3.1}$$

This is a particular case of the Sugeno integral analyzed in [2] under a mode general setting with lattice-valued partial possibilistic measures and mappings f defined on some proper subsystems of $\mathcal{P}(\Omega)$ and for general t-norms on $[0,1] \times [0,1]$. According to the results obtained in [2] (cf. formula (5), p. 229), in our particular case the relation

$$Ef(\cdot)d\Pi = \bigvee_{\omega \in \Omega} [f(\omega) \wedge \pi(\omega)] \tag{3.2}$$

holds. Indeed,

$$\bigvee_{\omega\in\Omega}(f(\omega)\wedge\pi(\omega)) = \bigvee_{t\in[0,1]}\ \bigvee_{\omega\in\Omega, f(\omega)=t}(f(\omega)\wedge\pi(\omega)) =$$
$$= \bigvee_{t\in[0,1]}[t\wedge\Pi(\{\omega\in\Omega : f(\omega)=t\})] =$$
$$= \bigvee_{t\in[0,1]}[t\wedge\Pi(\{\omega\in\Omega : f(\omega)\geq t\})] = Ef(\cdot)d\Pi. \tag{3.3}$$

In the following definition we take an inspiration from the probabilitic entropy I^* introduced above keeping also the symbol I^*, as perhaps no misunderstanding menaces.

Definition 3.1 Let π be a real-valued possibilistic distribution on a nonempty space Ω. The *possibilistic entropy* $I^*(\pi)$ *of* π is the real number from $[0,1]$ defined by

$$I^*(\pi) = E(1-\pi(\cdot))d\Pi. \tag{3.4}$$

Lemma 3.1 Let π be as in Definition 3.1. Then $I^*(\pi)=0$ iff π takes only the values 0 or 1.

Proof. If $\pi(\omega)\in\{0,1\}$ for every $\omega\in\Omega$, then either $\pi(\omega)=0$ or $1-\pi(\omega)=0$ for every $\omega\in\Omega$, so that, due to (3.2), $I^*(\pi)=\bigvee_{\omega\in\Omega}((1-\pi(\omega))\wedge\pi(\omega))=0$ follows. If $0<\pi(\omega)<1$ holds for some $\omega\in\Omega$, then $0<1-\pi(\omega)<1$ and $0<(1-\pi(\omega))\wedge\pi(\omega)\leq I^*(\pi)$ follows and the assertion is proved. □

A probabilistic (possibilistic, resp.) distribution on Ω is called *positive,* if $p(\omega)>0(\pi(\omega)>0$, resp.) holds for each $\omega\in\Omega$. Supposing that Ω is finite or countable, each non-positive probabilistic distribution on Ω can be replaced, without any loss of generality, by its reduction to the subset $\Omega_0=\{\omega\in\Omega : p(\omega)>0\}$. For possibilistic distribution such a reduction is possible for each Ω. Consequently, the only positive probabilistic distribution P_0 for which $I(P_0)=I^*(P_0)=0$ holds is the case when $\Omega=\{\omega_0\}$, so that $p(\omega_0)=1$, and the only positive possibilistic distribution π_0 on Ω with $I^*(\pi_0)=0$ is the unit or maximum one with $\pi_0(\omega)=1$ for every $\omega\in\Omega$.

In the probabilistic case the standard intuition behind is easy to be accepted. If $p(\omega)=1$ for some $\omega_0\in\Omega$, then we can predict "with the probability one" that ω_0 will be the result when making a random sample from the corresponding distribution, so that no new information is achieved. When considering a possibilistic distribution with values only 0 or 1, the interpretation behind may read as follows. If $\pi(\omega_{a_0})=1$, we have *absolutely no argument against* the idea or explanation that ω_{i_0} will occur, so that the actual appearance of this ω_{i_0} says to us nothing new about the world and the system under consideration. Hence, no new information is obtained just as it is the case of the random sample giving the result the apriori probability of which is just one. Contrary to this case, if the possibility degree $\pi(\omega_0)$ is zero, we may take it as the case when we have at hand a sufficient amount of data or arguments to deduce, beyond any doubt, that ω_{i_0} cannot occur. Finally, if $0<\pi(\omega_{i_0})<1$ holds, there are some nonnegligible arguments *in favor* of the idea that ω_{i_0} occurs, but there are also (different) nonnegligible arguments *against* this expectation. Hence, none of the two alternatives can be deduced beyond any doubts and the actual appearance of ω_{i_0} can be taken as a new piece of information enriching our knowledge (reducing our uncertainty) as far as the world and system under consideration are concerned.

As $(1-x)\wedge x\leq 1/2$ holds for each $0\leq x\leq 1$, the equality being the case when $x=1/2$, we obtain easily that

$$I^*(\pi) = E(1-\pi(\cdot))d\Pi = \bigvee_{\omega\in\Omega}((1-\pi(\omega))\wedge\pi(\omega)) \leq 1/2 \tag{3.5}$$

holds for every real-valued possibilistic distribution π on Ω. If there exists $\omega_0\in\Omega$ such that $\pi(\omega_0)=1/2$ holds, then $I^*(\pi)=1/2$ follows, this condition being also the necessary one when Ω is finite. It would be perhaps more elegant to define the possibilistic entropy as a *normalized* mapping setting $I^*(\pi)=2E(1-\pi(\cdot))d\Pi$, but let us keep the definition 3.1 as it stands.

4 Possibilistic Decision Functions and Extremum Entropy Values

Let us sketch, very briefly, a particular case of possibilistic decision functions proposed and developed in [7], namely that one with real-valued possibilistic measures and normalized loss functions and with the infimum operation in the role of t-norm on $[0,1] \times [0,1]$. Let S be the set of all states of a system (solutions to a problem, ...) just one state being the actual one (just one solution being the correct or acceptable one,...). Let D be the set of all possible decisions, let E be the set of possible empirical values (data, observations,...), let $\delta : E \to D$ be a decision function, let $\lambda : S \times D \to [0,1]$ be the loss function, so that $\lambda(s,d)$ defines the loss suffered by the subject (a manager who controls the system under consideration, say) when $s \in S$ is the actual state of the system and $d \in D$ is the decision taken by the subject. Hence, if $e \in E$ is the empirical value being at the subject's disposal, and δ is the decision function applied, the suffered loss reads as $\lambda(s,\delta(e)) \in [0,1]$.

The phenomenon of uncertainty enters our model when supposing that the empirical value $e \in E$ and the actual state s are the values of variables η and σ, both being charged by uncertainty, this time quantified and processed by possibilistic tools. Hence, let Ω be a nonempty space and let π be a possibilistic distribution on Ω, so that $\Pi(A) = \bigvee_{\omega \in A} \pi(\omega)$ defines, for each $A \subset \Omega$, the possibilistic measure induced by π on $\mathcal{P}(\Omega)$. Let $\sigma : \Omega \to S$, $\eta : \Omega \to E$ be mappings with the intuition behind as above, obviously, under our setting, the values $\Pi(\{\omega \in \Omega : \sigma(\omega) \in S_1\})$ and $\Pi(\{\omega \in \Omega : \eta(\omega) \in E_1\})$ are defined for every $S_1 \subset S$ and $E_1 \subset E$. The possibilistic distribution π_S on S, defined by $\pi_S(s) = \Pi(\{\omega \in \Omega : \sigma(\omega) = s\})$ for each $s \in S$, will be called the *apriori possibilistic distribution on S*.

Within the formal framework just introduced, the loss suffered when $\sigma(\omega)$ is the actual state, $\eta(\omega)$ is the empirical value being at the disposal, and δ is the applied decision function, converts into the function $\lambda(\sigma(\omega),\delta(\eta(\omega)))$, taking Ω into $[0,1]$, and the subject's aim is to minimize this loss by an appropriate choice of the decision function δ. As can be easily seen, supposing that the variables σ and η are fixed, the best decision function $\delta_0 : E \to D$ would be such one with the inequality $\lambda(\sigma(\omega),\delta_0(\eta(\omega))) \leq \lambda(\sigma(\omega),\delta(\eta(\omega)))$ holding for each $\delta : E \to D$ and each $\omega \in \Omega$, however, such δ_0 does not exists up to the most trivial cases, cf. [7], e.g. Hence, some weaker demands must be imposed on the optimality of the chosen decision function.

The well-known *Bayes principle* and the *worst-case* or *minimax principle* are the two most often used alternative approaches how to optimize the decision function δ. According to the Bayes principle the user's aim is to minimize the expected value of the loss function $\lambda(\sigma(\omega),\delta(\eta(\omega)))$, hence, processing the uncertainty by possibilistic tools, to minimize the value

$$\begin{aligned} \chi_\sigma^B(\delta) &= E\lambda(\sigma(\cdot),\delta(\eta(\cdot)))d\Pi = \bigvee_{t \in [0,1]} [t \wedge \Pi(\{\omega \in \Omega : \lambda(\sigma(\omega),\delta(\eta(\omega))) \geq t\})] = \\ &= \bigvee_{\omega \in \Omega} [\pi(\omega) \wedge \lambda(\sigma(\omega),\delta(\eta(\omega)))], \end{aligned} \tag{4.1}$$

when applying (3.2). For the worst-case of minimax approach we define, for each $s \in S$, the conditional expected value of the loss $\lambda(\sigma(\omega),\delta(\eta(\omega)))$ supposing that $\sigma(\omega) = s$, and we take the *supremum* of these values for s ranging over S as the quality criterion of the decision function δ. Hence, we try to choose δ minimizing the value

$$\begin{aligned} &\chi^{MM}(\delta) = \bigvee_{s \in S} E\lambda(s,\delta(\eta(\cdot)))d\Pi = \bigvee_{s \in S} \bigvee_{\omega \in \Omega} [\lambda(s,\delta(\eta(\omega))) \wedge \pi(\omega)] = \\ = &\bigvee_{\omega \in \Omega} \bigvee_{s \in S} [\lambda(s,\delta(\eta(\omega))) \wedge \pi(\omega)] \geq \bigvee_{\omega \in \Omega} [\lambda(\sigma(\omega),\delta(\eta(\omega))) \wedge \pi(\omega)] = \\ = &\chi_\sigma^B(\delta). \end{aligned} \tag{4.2}$$

So, the well-known and easy to understand inequality between the Bayes and the minimax risk of the decision function δ, valid in the case of statistical decision functions, is valid also in the case of possibilistic decision functions.

Lemma 4.1 Let $S = D$, let $\pi(\omega) \in \{0,1\}$ for each $\omega \in \Omega$, let $\lambda(s,d) = 0$, if $s = d, \lambda(s,d) = 1$, if $s \neq d$ hold for each $s, d \in E(= D)$. Then $\chi^B_\sigma(\delta) = \chi^{MM}(\delta) = 1$, if there is $\omega_0 \in \Omega$ such that $\pi(\omega_0) > 0$ and $\sigma(\omega_0) \neq \delta(\eta(\omega_0))$ holds, $\chi^B_\sigma(\delta) = \chi^{MM}(\delta) = 0$ otherwise.

Proof. Let $\omega_0 \in \Omega$ be such that $\pi(\omega_0) > 0$ (hence, $\pi(\omega_0) = 1$) and $\sigma(\omega_0) \neq \delta(\eta(\omega_0))$ holds. Then $\lambda(\sigma(\omega_0, \delta(\eta(\omega_0))) = 1$ follows, so that the relation

$$\begin{aligned} \chi^B_\sigma(\delta) &= \bigvee_{\omega \in \Omega} [\lambda(\sigma(\omega), \delta(\eta(\omega))) \wedge \pi(\omega)] \geq \lambda(\sigma(\omega_0), \delta(\eta(\omega_0))) \wedge \pi(\omega_0) = \\ &= 1 \wedge 1 = 1 = \chi^{MM}(\delta) \end{aligned} \tag{4.3}$$

is valid. If $\sigma(\omega) = \delta(\eta(\omega))$ for each $\omega \in \Omega$ such that $\pi(\omega) > 0$ holds, then obviously $\langle(\sigma(\omega), \delta(\eta(\omega))) \wedge \pi(\omega) = 0$ for every $\omega \in \Omega$, so that $\chi^B_\sigma(\delta) = \chi^{MM}(\delta) = \oslash$ follows and the assertion is proved. $\square$

Consequently, if π is a possibilistic distribution on a finite space Ω and if $I^*(\pi) = 0$, then $\chi^B_\sigma(\delta) = \chi^{MM}(\delta) = 1$ for each apriori possibilistic distribution σ on S and for each decision function δ which is not absolutely correct, i.e., which decides wrongly for some $\omega \in \Omega$ with $\pi(\omega) > 0$.

The following alternative variant of Lemma 4.1 seems to be perhaps more close to the intuition behind the notion of possibilistic degrees and measures.

Lemma 4.2 Let the notations and conditions of Lemma 4.1 hold, but this time with loss function defined by $\lambda(s,d) = |\pi_S(s) - \pi_S(d)|$, let us recall that $\pi_S : S \to [0,1]$ is the apriori possibilistic distribution on the set S of states. Then $\chi^B_\sigma(\delta) = \chi^{MM}(\delta) = 1$, if there is $\omega_0 \in \Omega$ such that $\pi(\omega_0) > 0$ and $\pi_S(\sigma(\omega_0)) \neq \pi_S(\delta(\eta(\omega_0)))$ holds.

Proof. Again, if $\pi(\omega_0) > 0$, then $\pi(\omega_0) = 1$ holds, so that (4.3) yields that

$$\chi^B_\sigma(\delta) \geq \lambda(\sigma(\omega_0), \delta(\eta(\omega_0))) \wedge \pi(\omega_0) = |\pi_S(\sigma(\omega_0)) - \pi_S(\delta(\eta(\omega_0)))| \tag{4.4}$$

holds. As $\pi(\omega) \in \{0,1\}$ for each $\omega \in \Omega$, we obtain immediately that $\Pi(A) = \bigvee_{\omega \in A} \pi(\omega) \in \{0,1\}$ is valid for each $A \subset \Omega$, in particular, also $\pi_S(s) \in \{0,1\}$ for each $s \in S$. Consequently, if $\pi_S(s) \neq \pi_S(d)$ for some $s, d \in S$ $(= D)$, then $|\pi_S(s) - \pi_S(d)| = 1$ follows. So, $\chi^B_\sigma(\delta) = \chi^{MM}(\delta) = 1$ and the assertion is proved. $\square$

When considering the $0-1$-loss function as in Lemma 4.1, what matters when taking the decision $d = \delta(\eta(\omega))$ is the *identity* of this state with the actual state $s = \sigma(\omega)$. We could generalize this approach when replacing this demand by that of *indistinguishability.* So, given a general loss function $\lambda : S \times D(= S) \to [0,\infty)$, we could define states s and d as indistinguishable, if $\lambda(s,d) = 0$ (including the case when $s = d$ as a particular one). Such states can be taken as being identical as far as the quantitative characteristics of qualities of the decision functions under consideration are concerned. In the possibilistic framework, states s and d can be taken as indistinguishable, if the arguments (reasons) putting in question that $s(d$, resp.) is the actual state are the same, in particular, that there are no such reasons as it is the case when $\pi_S(s) = \pi_S(d) = 1$. Dually, $\pi_S(s) = \pi_S(d) = 0$ is the case when the reasons against s as well as against d are so strong that they exclude ultimately both these possibilisties. The loss function $\lambda(s,d)$ defined by $|\pi_S(s) - \pi_S(d)|$ may be taken as an attempt to formalize this idea more precisely.

An apriori possibilistic distribution π_S on S is called S_0-*almost uniform,* where S_0 is a fixed nonempty proper subset of S, if $\pi_S(s) = 1/2$ for each $s \in S - S_0$. Obviously, $S_0 \neq \emptyset$ must hold, as the mapping $\pi_S : S \to [0,1]$ such that $\pi_S(s) = 1/2$ for each $s \in S$ does not meet the condition that $\bigvee_{s \in S} \pi_S(s) = 1$. Hence, for each S_0-almost uniform apriori possibilistic distribution π_S on S the relation

$$I^*(\pi_S) = \bigvee_{s \in S} ((1 - \pi_S(s)) \wedge \pi_S(s)) = 1/2 \tag{4.5}$$

holds, so that $I^*(\pi_S)$ takes the maximal possible value for π_S ranging over possibilistic distributions on S. Given a fixed $s_0 \in [0,1]$, the value $\sup\{|x - x_0| : x \in [0,1]\} = |0 - x_0| = x_0$, if $x_0 \geq 1/2$, this value being $1 - x_0 \geq 1/2$, if $x_0 \leq 1/2$ holds, hence, $\inf_{x_0 \in [0,1]} \sup_{x \in [0,1]} |x - x_0| = 1/2$. So, if π_S is an S_0-almost uniform apriori possibilistic distribution on S, then for each $\omega \in \Omega$ such that $\sigma(\omega) \in S - S_0$ holds the loss function $\lambda(\sigma(\omega), \delta(\eta(\omega)))$ defined by $|\pi_S(\sigma(\omega)) - \pi_S(\delta(\eta(\omega)))|$ takes the value $1/2$, hence, for each $\delta(\eta(\cdot)) : \Omega \to S$,

$$\bigvee_{\omega \in \Omega, \sigma(\omega) \in S - S_0} \lambda(\sigma(\omega), \delta(\eta(\omega))) = 1/2 \tag{4.6}$$

holds, so that the apriori possibilistic distribution π_S is optimal in the minimax sense on $\Omega_0 = \{\omega \in \Omega : \sigma(\omega) \in S - S_0\}$, like it is the case of the uniform apriori probability distribution $\langle 1/n, 1/n, \ldots, 1/n \rangle$ on the finite state space S of cardinality n.

It is perhaps worth noting explicitly that for π_S on S such that $I^*(\pi_S) < 1/2$ holds the relation (4.6) is not valid for no matter which proper subset S_0 of S. Indeed, if $I^*(\pi_S) = (1/2) - \varepsilon$ holds for some $\varepsilon > 0$, then $\pi_S(s) \leq (1/2) - \varepsilon$ or $\pi_S(s) \geq (1/2) + \varepsilon$ follows for each $s \in S$, hence,

$$\bigvee_{x \in [0,1]} |x - \pi_S(s)| \geq (1/2) + \varepsilon \tag{4.7}$$

is the case for each $s \in S$. Consequently, $\bigvee_{\omega \in \Omega, \sigma(\omega) \in S - S_0} \lambda(\sigma(\omega), \delta(\eta(\omega))) = (1/2) + \varepsilon$ is valid for each $\emptyset \neq S_0 \subset S$, so that π_S is not the optimal (in the minimax sense) apriori possibilistic distribution on S.

5 Independent Products of Real-Valued Possibilistic Distributions

An important property of the notion of standard probabilistically based entropy function consists in the fact that the entropy of independent product of two or more probability distributions is equal to the sum of the entropy values defined by the particular probability distributions. In more detail, let us limit ourselves to finite spaces Ω_1 and Ω_2, let $p_i : \Omega_i \to [0,1]$ be a probability distribution on Ω_i, $i = 1, 2$, and let us define the probability distribution p_{12} on $\Omega_1 \times \Omega_2$, setting $p_{12}(\omega_1, \omega_2) = p_1(\omega_1) p_2(\omega_2)$ for every $\langle \omega_1, \omega_2 \rangle \in \Omega_1 \times \Omega_2$. For entropy values $I(p_i) = -\sum_{\omega_i \in \Omega_i} p_i(\omega_i) lg_2\, p_i(\omega_i)$ for $i = 1, 2$, and $I(p_{12}) = -\sum_{\langle \omega_1, \omega_2 \rangle \in \Omega_1 \times \Omega_2} p_{12}(\omega_1, \omega_2)\ lg_2\, p_{12}(\omega_1, \omega_2)$ (with the convention $0\, lg_2\, 0 = 0$ applied) the relation $I(p_{12}) = I(p_1) + I(p_2)$ is valid. Let us prove that for possibilistically defined independent product π_{12} of real-valued possibilistic distributions π_1, π_2 and for entropy function $I^*(\pi) = E(1 - \pi(\cdot)) d\Pi$ (cf. (3.4)) the maxitive version of the additivity property is valid. The following more or less trivial assertions of standard mathematical analysis are perhaps worth being stated and proved explicitly.

Lemma 5.1 Let $[0,1]$ be the unit interval of real numbers equipped by their standard linear ordering $\leq$, let $\vee, \bigvee (\wedge, \bigwedge$, resp.) denote the standard supremum (infimum, resp.) operation in $[0,1]$. Then for each $\emptyset \neq A \subset [0,1]$ and each $s \in [0,1]$ the relation

$$s \wedge \left(\bigvee A \right) = \bigvee_{t \in A} (s \wedge t) \tag{5.1}$$

holds, where $\bigvee A$ abbreviates $\bigvee_{t \in A} t$.

Proof. The relation $s \wedge (\bigvee A) \geq s \wedge t$ holds for each $t \in A$, so that the inequality $s \wedge (\bigvee A) \geq \bigvee_{t \in A} (s \wedge t)$ immediately follows. As far as the inverse inequality is concerned, if $s \geq \bigvee A$ is the case, then $s \wedge (\bigvee A) = \bigvee A$, $s \wedge t = t$ for each $t \in A$, so that $\bigvee_{t \in A} (s \wedge t) = \bigvee A$. Finally, $s \wedge t \leq s$ holds for each $t \in A$, but if $s < \bigvee A$ is the case, then there exists $t_0 \in A$ such that $s < t_0$ holds, hence, the relation $s \wedge t_0 = s = \bigvee_{t \in A} (s \wedge t) = s \wedge (\bigvee A)$ easily follows and the assertion is proved. □

Lemma 5.2 For each $\emptyset \neq I \subset [0,1]$ the inequality

$$\left(\bigwedge_{x\in I} x\right) \wedge \left(1 - \bigwedge_{x\in I} x\right) \leq \bigvee_{x\in I} (x \wedge (1-x)) \tag{5.2}$$

holds.

Proof. As $1 - \bigwedge_{x\in I} x = \bigvee_{s\in I}(1-x)$ obviously holds, (5.2) reduces to

$$\left(\bigwedge_{x\in I} x\right) \wedge \left(\bigvee_{x\in I} (1-x)\right) \leq \bigvee_{s\in I} (x \wedge (1-x)) \tag{5.3}$$

Applying Lemma 5.1 to $s = \bigwedge_{x\in I} x$ and $A = \{1-x : x \in I\}$, we obtain that

$$\begin{aligned} \left(\bigwedge_{x\in I} x\right) \wedge \left(\bigvee_{x\in I} (1-x)\right) &= \bigvee_{x\in I} \left(\left(\bigwedge_{x\in I} x\right) \wedge (1-x)\right) \leq \\ \leq \quad & \bigvee_{x\in I} (x \wedge (1-x)) \end{aligned} \tag{5.4}$$

holds, as $\bigwedge_{x\in I} x \leq x$ holds trivially for each $x \in I$. The assertion is proved. □

Lemma 5.3 For each $x, y \in [0,1]$ such that $|x-y| \leq \varepsilon$ holds, the inequality

$$K(x,y) = |((1-x) \wedge x) - ((1-y) \wedge y)| \leq \varepsilon \tag{5.5}$$

holds as well.

Proof. If $x, y \leq 1/2 (x, y \geq 1/2$, resp.) is the case, then $K(x,y) = |x-y| (K(x,y) = |(1-x)-(1-y)| = |y-x| = |x-y|$, resp.) follows, so that $K(x,y) \leq \varepsilon$ holds. Otherwise we may suppose, without any loss of generality, that $x < 1/2$ and $y > 1/2$ is the case, hence, as $|x-y| \leq \varepsilon$ holds, the inclusions $x \in (1/2 - \varepsilon, 1/2)$ and $y \in (1/2, 1/2 + \varepsilon)$ follow. In this case, however, $(1-x) \wedge x = x$, $(1-y) \wedge y = 1-y \in (1/2 - \varepsilon, 1/2)$ holds, so that $K(x,y) = |x-(1-y)| \leq \varepsilon$ results and the assertion is proved. □

Let Ω be a nonempty space, let S be a set of (real-valued) possibilistic measures on Ω, i.e., for each $i \in S, \pi_i$ takes Ω into $[0,1]$ in such a way that $\bigvee_{\omega\in\Omega} \pi_i(\omega) = 1$ holds. Let Ω^S denote the space of all mappings taking S into Ω, hence, for each $\omega^S \in \Omega^S$ and each $i \in S$, $\omega^S(i) \in \Omega$ holds. Let π^S be the mapping taking Ω^S into $[0,1]$ and defined in this way: for each $\omega^S \in \Omega^S$,

$$\pi^S(\omega^S) = \bigwedge_{t\in S} \pi_i(\omega^S(i)). \tag{5.6}$$

Lemma 5.4 The mapping π^S defines a (real-valued) possibilistic distribution on ω^S.

Proof. As $\bigvee_{\omega\in\Omega} \pi_i(\omega) = 1$ holds for each $i \in S$, there exists, for each $i \in S$ and each $\varepsilon > 0$, an element $\omega_{0,i} \in \Omega$ such that $\pi_i(\omega_{0,i}) > 1 - \varepsilon$ holds. For $\omega_0^S \in \Omega^S$ defined by $\omega_0^S(i) = \omega_{0,i}$, (5.6) yields that $\pi^S(\omega_0^S) \geq 1 - \varepsilon$ holds, hence,

$$\bigvee_{\omega^S\in\Omega^S} \pi^S(\omega^S) = 1 \tag{5.7}$$

follows and the assertion is proved. □

The possibilistic distribution π^S on Ω^S will be called *(possibilistically) independent product of possibilistic distributions* π_i *on* Ω *with* i *ranging over* S. For each particular possibilistic distribution $\pi_i, i \in S$, as well as for their product π^S, we can define their entropy functions applying (3.1) and (3.2) and we obtain that

$$I(\pi_i) = \bigvee_{\omega \in \Omega} ((1 - \pi_i(\omega)) \wedge \pi_i(\omega)), \tag{5.8}$$

$$I(\pi^S) = \bigvee_{\omega^S \in \Omega^S} ((1 - \pi^S(\omega^S)) \wedge \pi^S(\omega^S)), \tag{5.9}$$

The following theorem can be taken as a possibilistic variant of the additivity property of statistically independent products of probability distributions.

Theorem 5.1 *Under the notations and conditions introduced above, the relation*

$$I(\pi^S) = \bigvee_{i \in S} I(\pi_i) \tag{5.10}$$

holds.

Proof. As shown in the proof of Lemma 5.4, for each $i_0 \in S$ and each $\varepsilon > 0$ there exists $\omega(i, \varepsilon) \in \Omega$ such that $\pi_i(\omega(i, \varepsilon)) > 1 - \varepsilon$ holds. Define, for each $\omega \in \Omega$, the mapping $\omega^S[i_0, \varepsilon, \omega] : S \to \Omega$ in this way:

$$\omega^S[i_0, \varepsilon, \omega](i_0) = \omega \tag{5.11}$$

$$\omega^S[i_0, \varepsilon, \omega](j) = \omega(j, \varepsilon) \text{ for each } j \in S, j \neq i_0. \tag{5.12}$$

Consequently,

$$\pi^S(\omega^S[i_0, \varepsilon, \omega]) = \bigwedge_{j \in S} \pi_j(\omega^S(i_0, \varepsilon, \omega](j)) =$$

$$= \pi_{i_0}(\omega) \wedge \left(\bigwedge_{j \in I, j \neq i_0} \pi_j(\omega(j, \varepsilon)) \right). \tag{5.13}$$

Hence, if $\pi_{i_0}(\omega) \leq 1 - \varepsilon$ holds, then $\pi^S(\omega^S[i_0, \varepsilon, \omega]) = \pi_{i_0}(\omega)$ follows, as $\pi_j(\omega(j, \varepsilon)) > 1 - \varepsilon$ is valid for each $j \in I, j \neq i_0$. If $\pi_{i_0}(\omega) > 1 - \varepsilon$ is the case, i.e., of $\pi_{i_0}(\omega) \in (1 - \varepsilon, 1]$ holds, then (5.13) yields that also

$$\pi^S(\omega^S[i_0, \varepsilon, \omega]) \geq 1 - \varepsilon, \text{ i.e. } \pi^S(\omega^S[i_0, \varepsilon, \omega]) \in [1 - \varepsilon, 1] \tag{5.14}$$

holds. So, in both the cases the inequality

$$|\pi^S(\omega^S[i_0, \varepsilon, \omega]) - \pi_{i_0}(\omega)| \leq \varepsilon \tag{5.15}$$

is valid. Due to Lemma 5.3, also the inequality

$$\begin{aligned} & K(\pi^S(\omega^S[i_0, \varepsilon, \omega]), \pi_{i_0}(\omega)) = \\ = \; & ([\pi^S(\omega^S[i_0, \varepsilon, \omega]) \wedge (1 - \pi^S(\omega^S[i_0, \varepsilon, \omega]))] - \\ - \; & [\pi_{i_0}(\omega) \wedge (1 - \pi_{i_0}(\omega))]) \leq \varepsilon \end{aligned} \tag{5.16}$$

is valid. Hence, for each $i \in S$ and $\omega \in \Omega$, the value $\pi_i(\omega) \wedge (1 - \pi_i(\omega))$ can be approximated, up to an arbitrarily small difference $\varepsilon > 0$, by the value $\pi^S(\omega^S) \wedge (1 - \pi^S(\omega^S))$ for some $\omega^S \in \Omega^S$. So, we may conclude that the inequality

$$
\begin{aligned}
& I(\pi^S) = \bigvee_{\omega^S \in \Omega^S} ((1-\pi^S(\omega^S)) \wedge \pi^S(\omega^S)) \geq \\
\geq \quad & \bigvee_{i \in S, \omega \in \Omega} ((1-\pi_i(\omega)) \wedge \pi_i(\omega)) = \bigvee_{i \in S} \left(\bigvee_{\omega \in \Omega} ((1-\pi_i(\omega)) \wedge \pi_i(\omega)) \right) = \\
= \quad & \bigvee_{i \in S} I(\pi_i)
\end{aligned}
\tag{5.17}
$$

is valid.

Let us prove the inverse inequality $I(\pi^S) \leq \bigvee_{i \in S} I(\pi_i)$, i.e., the inequality

$$
\bigvee_{\omega^S \in \Omega^S} ((1-\pi^S(\omega^S)) \wedge \pi^S(\omega^S)) \leq \bigvee_{i \in S} \left(\bigvee_{\omega \in \Omega} ((1-\pi(\omega)) \wedge \pi_i(\omega)) \right). \tag{5.18}
$$

What obviously suffices is to prove this inequality for each $\omega^S \in \Omega^S$ in particular, hence, to prove that for each $\omega^S \in \Omega$ the relation

$$
\begin{aligned}
& \pi^S(\omega^S) \wedge (1-\pi^S(\omega^S)) = \left(\bigwedge_{i \in S} \pi_i(\omega^S(i)) \right) \wedge \left(1 - \bigwedge_{i \in S} \pi_i(\omega^S(i)) \right) \\
\leq \quad & \bigvee_{i \in I} \left(\bigvee_{\omega \in \Omega} (\pi_i(\omega) \wedge (1-\pi_i(\omega))) \right)
\end{aligned}
\tag{5.19}
$$

holds. Due to Lemma 5.2 we obtain that

$$
\begin{aligned}
& \left(\bigwedge_{i \in S} \pi_i(\omega^S(i)) \right) \wedge \left(1 - \bigwedge_{i \in S} \pi_i^S(\omega^S(i)) \right) = \\
= \quad & \left(\bigwedge_{i \in S} \pi_i(\omega^S(i)) \right) \wedge \left(\bigvee_{i \in S} (1 - \pi_i(\omega^S(i)) \right) \leq \\
\leq \quad & \bigvee_{i \in S} [\pi_i(\omega^S(i)) \wedge (1 - \pi_i(\omega^S(i)))] \leq \\
\leq \quad & \bigvee_{\omega \in \Omega} \left(\bigvee_{i \in S} (\pi_i(\omega) \wedge (1-\pi_i(\omega))) \right) = \\
= \quad & \bigvee_{i \in S} \left(\bigvee_{\omega \in \Omega} (\pi_i(\omega) \wedge (1-\pi_i(\omega))) \right) = \bigvee_{i \in S} I(\pi_i).
\end{aligned}
\tag{5.20}
$$

The inequality inverse to (5.17) as well as whole the assertion are proved. □

Theorem 5.1 extends the result from [8], where the equality (5.10) was proved only for possibilistically independent products of finite or countable sequences of possibilistic distributions, i.e., for finite or countable parametric set S, in our notation, and the main idea of the proof consisted in an application of the principle of mathematical induction. However, this proof technique can be more of less routinely applied to the case of non-numerical (in particular, lattice-valued) possibilistic distributions and their possibilistically independent products supposing that the complete lattice in question meets certain condition. On the other side, in the proof of Theorem 5.1 as introduced above we applied some specific properties of the structure $\langle [0,1], \leq \rangle$ of real numbers, namely the fact that the supremum of each nonempty $A \subset [0,1]$ can be reached when restricting ourselves to an appropriate countable subset A_0 of A.

6 Conclusions

Keeping in mind the fact that within the framework of the standard probability and information theory the notion of entropy defines an important characteristic of probability distributions and measures, our aim and goal throughout this paper was to propose and analyze an alternative idea of entropy definable in the case when the uncertainty phenomenon entering our formal model, and degrees of uncertainty in particular, are defined, quantified and processed by the tools offered by possibilistic distributions and measures. For this sake we have defined a possibilistic variant of the alternative version of probabilistic entropy function resulting when replacing the logarithmic function $\log(1/p(\omega))$ by the function $1 - p(\omega)$, hence, we define possibilistic entropy of a possibilitic distribution function as the possibilistic expected value (defined by Sugeno integral) of the value $1 - \pi(\omega), \omega \in \Omega$, where $\pi(\omega)$ defines a real-valued possibilistic distribution over the space Ω.

The possibilistic entropy has been proved to conserve two important properties of probabilistic entropy functions. First, the possibilistic distributions with the minimum (i.e., zero, in our case) entropy value are just those when the realiation of the random (in the possibilistic sense) sample does not bring any new information, hence, it is like the case of degenerated probability distribution with one result occurring with the probability one, i.e., beyond any doubts. However, we must keep in mind the interpretation according to which the occurrence of any result with possibility degree one also does not bring any new information. On the other side, when applying the idea of possibilistic decision function under the bayesian setting, we may conclude that in the case when no specification of the actual apriori possibilistic distribution is at hand, the hypothesis of the apriori possibilistic distribution with the maximum possibilistic entropy value is rational and reasonable as this hypothesis minimizes, from the worst-case point of view, the possibilistically expected loss related to the decision making in question (remember the analogous role of uniform probability distribution over finite state spaces in bayesian statistical decision functions).

Second, the possibilistic entropy functions as defined in this paper copy also the additivity property of entropies for statistically independent products of two or more probability distributions just with additivity replaced by maxitivity. Hence, combining two or more possibilistic distributions in possibilistically independent way into one possibilistic distribution, the possibilistic entropy of the resulting distribution is identical with the supremum value of the possibilistic entropies defined by particular distributions. The result is valid also for the independent product of an infinite system of possibilistic distributions.

As far as some future research is concerned, at least the two following directions seem to be worth being pursued. First, we could investigate other decreasing (or at least nonincreasing) functions taking $[0, 1]$ into $[0, 1]$ and replacing the function $1 - x$ applied in our approach. Second, we could try to propose a lattice-valued entropy function defined by lattice-valued possibilistic measures. A serious problem consists in the fact that our approach takes substantial profit of the specific properties of the complement operation $1 - x$ in $[0, 1]$ so that its more or less formal application to the complements defined in the boolean or residual sense, definable in complete lattices, does not lead to sophisticated enough results, so that a new and qualitatively different idea seems to be highly desirable.

The list of references introduced below deserves perhaps a very short comment. Elementary ideas dealing with information and entropy quantification and processing can be found in [5], one of the earliest monographs on information theory. Decision making under probabilistically quantified and processed uncertainty, which served as an inspiration and a challenge for our alternative possibilistic approach, was conceived by [9] and [1]. Elementary ideas and results on possibilistic measures can be found in [3], their detailed mathematical formalization including the notion of Sugeno integral, important for our purposes, are analyzed in [2]. References [4] and [6] are of surveyal character, emphasizing the philosophical and methodological aspects of various mathematical models for uncertainty quantification and processing, as well as the mutual relations among the corresponding notions, properties and the achieved results. An attempt to propose a possibilistic variant of the theory of decision making under uncertainty is presented in [7]. The author believes that all these references could be of use for a reader attempting to penetrate more deeply into the problems touched and perhaps at least partially solved in this paper.

Acknowledgement

This work was partially supported by grant No. IAA100300503 of GA AS CR and by the Institutional Research Plan AV0Z10300504 “Computer Science for the Informational Society: Models, Algorithms, Applications.”

Bibliography

[1] Blackwell D., Qirshick M. Q.: Theory of Games and Statistical Decisions, New York, Wiley, 1954.

[2] DeCooman G.: Possibility theory I, II, III, International Journal of General Systems, **25**, 1997, pp. 291-323, pp. 325-351, pp. 353-371.

[3] Dubois D., Prade H.: Possibility Theory: An Approach to Computerized Processing of Uncertainty, New York, Plenum Press, 1988.

[4] Dubois D., et al.: Possibility theory, probability theory and fuzzy sets: misunderstandings, bridges and gaps. In: The handbook of Fuzzy Sets Series – Fundamentals of Fuzzy Sets, D. Dubois and H. Prade (Eds.), Boston, Kluwer Academic Press, 2000, pp. 343-438.

[5] Khinchin A. J.: Mathematical Foundations of Information Theory. New York, Dover Publications, 1957.

[6] Klir G. J, Folger T. O.: Fuzzy Sets, Uncertainty and Information, New Jersey, Prentice Hall, Englewood Cliffs, 1988.

[7] Kramosil I.: Decision-Making under uncertainty processed by lattice-valed possibilistic measures, Kybernetika, **42**, 2006, pp. 629-646.

[8] Kramosil I.: Entropy functions defined by possibilistic distributions, submitted for publication to the Proceedings of the 5th International Symposium on Imprecise Probabilities and Their Applications, Prague, 2007.

[9] Wald A.: Statistical Decision Functions, New York, Wiley, 1950.