Kybernetika

Arthur Ramer

Fuzzy information and combinatorial inequalities

Kybernetika, Vol. 28 (1992), No. Suppl, 4--11

Persistent URL: http://dml.cz/dmlcz/124198

SUPPLEMENT TO KYBERNETIKA VOLUME 28 (1992), PAGES 4 – 11

FUZZY INFORMATION AND COMBINATORIAL INEQUALITIES

ARTHUR RAMER

Properties of information measures both in fuzzy theory and in the theory of evidence are closely related to combinatorial identities and inequalities. It is demonstrated how study of information functions leads to new combinatorial results. Three cases are presented
- metric property of fuzzy information distance
- approximations of continuous fuzzy information
- maximum for a combined Dempster-Shafer evidence measure

1. INTRODUCTION

In 1982 Higashi and Klir [3, 6] introduced the notion of fuzzy uncertainty measure and the related fuzzy information distance. They were expressed as function of the possibility assignments corresponding to fuzzy sets under consideration. Later, in [10] and [7] uncertainty measures were classified and characterized axiomatically. These characterizations were mostly in terms of certain identities, but also included at least one inequality, corresponding to the property of subadditivity. Following that development, we [8] obtained related characterizations for information distances. A closer analysis of their properties led to a novel combinatorial inequality. In particular, this inequality was responsible for the metric property of G-distance.

Another development was related to our recent paper showing how to define notions of fuzzy information on continuous domains of discourse. A natural question thus arose, whether the 'continuous' information can be approximated by the discrete ones. The answer is positive, leading to an interesting family of expressions for limits of sums, representing discrete information values.

Very recently, Klir [6] studying general properties of information across various modes of definition – probabilistic, fuzzy and Dempster-Shafer – introduced a principle of 'information preservation'. This study suggested a very intriguing inequality for Dempster-Shafer theory [11]. It gives an upper bound for a given evidence assignment.

In the remainder of our paper we outline these combinatorial results. The theorems about rearrangements and about approximations of continuous information will be only illustrated with representative examples, the proofs being published separately. The result about 'evidence' information is very recent and we outline its proof. In the closing we remark that our study is likely only a beginning of mutual influence of the developments in two very distinct disciplines – fuzzy reasoning and combinatorics.

2. TECHNICAL NOTE

In our discussion we often refer to the values of a finite sequence (p_i) as arranged in descending order of their values. We denote such rearrangement $(\tilde{p}_i)$. On occasion we assume tacitly that we have also defined $\tilde{p}_0 = \tilde{p}_{n+1} = 0$. We adopt [5] style of writing summation symbols without explicit limits if the maximum range is meant. Such range is usually 1 to n, but it may become 1 to $n-1$ or 2 to n to avoid terms like $\log 0$ or $\frac{1}{0}$. Logarithms are in base 2, except in the section on continuous information where natural logarithms are more convenient; we assume that $0 \log 0 = 0$.

Fuzzy sets [1] are defined by their *membership* grade with respect to a domain of discourse. We assume these domains to be finitely measurable sets. Most results will be stated only for finite domains; an obvious exception will be the definition and approximations of 'continuous' information functions.

Given a domain $X = \{x_1, \ldots, x_n\}$ and a membership grade $\pi : X \to [0,1]$, its values $p_1 = \pi(x_1), \ldots, p_n = \pi(x_n)$ form a *possibility* distribution. They can be interpreted as giving numerical expression of the likelihood of observing a specific instance x_i within the domain X.

Given two such distributions $\pi = (p_i)$ on X and $\varpi = (q_i)$ on Y, we can form a joint distribution $\rho = \pi \otimes \varpi$ on $X \times Y$ by putting

$$\rho : (x_i, y_j) \mapsto \min(p_i, q_j) \quad \text{or} \quad r_{ij} = \min(p_i, q_j).$$

Conversely, given an arbitrary distribution $\sigma = (s_{ij})$ on a product space $X \times Y$, we can construct its *marginal* projection σ' on X and σ'' on Y

$$\sigma'(x_i) = \max_j(s_{ij}) \qquad \sigma''(y_j) = \max_i(s_{ij}).$$

Our model of Dempster–Shafer theory [11] will consist of a domain of discourse given as a finite set X and of a *basic assignment* of evidence – an assignment of nonnegative weights m_i, $0 \le m_i \le 1$ to a family of subsets A_i of X. We stipulate that $\sum m_i = 1$. The subsets A_i are called *focal* subsets of evidence assignment, thus defining a *mapping* $\mu : A_i \mapsto m_i$. It is convenient to extend μ to all subsets of X by putting $\mu(B) = 0$ for a non-focal B. We do not presuppose any containment or intersection structure of sets A_i – they may, and usually will have non-empty intersections; if $A_i \subset A_j$ it is possible that $\mu(A_i) > \mu(A_j)$ and so on.

3. INEQUALITIES AND FUZZY INFORMATION

We shall conduct our discussion in terms of a finite domain of discourse X and possibility distributions, say $\pi : X \to [0,1]$, $\sup_{x \in X} \pi(x) = 1$, represented as a sequence $p_1, \ldots, p_n$, where $p_i = \pi(x_i)$. Here and in the following we assume that the cardinality $|X| = n$.

The basic information function [3, 4, 8] is U-uncertainty. Given π sorted in the descending order of values $\tilde{p}_1, \ldots, \tilde{p}_n$, we assign it an information value

$$U(\pi) = \sum (\tilde{p}_i - \tilde{p}_{i+1}) \log i = \sum \tilde{p}_i \log \frac{i}{i-1}.$$

Given two distributions $\pi = (p_i), \quad \rho = (r_i)$ such that $p_i \leq r_i, \quad i = 1, \ldots, n$ we define their information distance

$$g(\pi, \rho) = U(\rho) - U(\pi).$$

In a general case of arbitrary π and ρ we first define their lattice-theoretic supremum $\pi \vee \rho = (q_i'')$ where $q_i'' = \max(p_i, r_i)$. Now we can define a *metric* distance $G(\pi, \rho) = g(\pi, \pi \vee \rho) + g(\rho, \pi \vee \rho)$. Dual definition, using $\pi \wedge \rho$, leads to $H(\pi, \rho) = g(\pi \wedge \rho, \pi) + g(\pi \wedge \rho, \rho)$; this distance is not a metric, however it has an attractive property of being additive.

Closer analysis of $G(\pi, \rho)$ shows that its metric property is a direct consequence of an important inequality concerning rearrangements of sequences. Given a sequence (a_i) we shall denote its descending rearrangement by $\tilde{a} = (\tilde{a}_i)$. Given two sequences $a = (a_i)$ and $b = (b_i)$ we denote by $a \vee b$ the sequence $(\widetilde{\max(a_i, b_i)})$ and by $a \wedge b$ the sequence $(\min(a_i, b_i))$. We observe here that, for example, $\widetilde{a \vee b}$ is a very different sequence from $\tilde{a} \vee \tilde{b}$, as the former results from rearranging the pairwise maxima $\max(a_i, b_i)$ of the original a and b.

We can now state the inequalities. We shall state them only in the form directly applicable to the discussion of information metric; their general form is presented in [9]. We assume that $a = (a_i), b = (b_i)$ and $c = (c_i)$ are arbitrary sequences of n elements and that $w = (w_i)$ is a non-increasing sequence of weights $w_1 \geq \cdots \geq w_n$.

Theorem 1.

$$\sum w(a) + \sum w(b) \geq \sum w(\widetilde{a \vee b}) + \sum w(\widetilde{a \wedge b}).$$

Theorem 2.

$$\sum w(\widetilde{a \vee c}) + \sum w(\widetilde{b \vee c}) \geq \sum w(\widetilde{a \vee b}) + \sum w(\tilde{c}).$$

We now indicate how Theorem 2 leads to the metric property of G distance. Assume given three possibility distributions $\pi = (p_i), \quad \rho = (r_i)$ and $\sigma = (s_i)$. The inequality $G(\pi, \sigma) \leq G(\pi, \rho) + G(\rho, \sigma)$ can be expressed using U-uncertainties giving, after regrouping terms

$$U(\pi \vee \sigma) + U(\rho) \leq U(\pi \vee \rho) + U(\rho \vee \sigma).$$

It is effectively the statement of Theorem 2, where $w_i = \log \frac{i}{i-1}$.

4. CONTINUOUS INFORMATION FUNCTIONS

For the continuous case we consider interval $[0,1]$ as the prototypical domain of discourse and function $f : [0,1] \to [0,1]$ to represent a possibility distribution. We recall that in the discrete case

$$U(\pi) = \sum (\tilde{p}_i - \tilde{p}_{i+1}) \log i = \sum \tilde{p}_i \,\Delta \log i$$

and that suggests using a 'sorted' version $\tilde{f}(x)$ of the original function $f(x)$, together with the differential $d \ln x = \frac{1}{x}\, dx$. (In this section we use ln rather than log function.)

[5] provide such construction for an arbitrary measurable function f. The rearranged function is defined to 'stay' above any given level over the same 'space' as the original function. It is constructed as a descending function, and its formation generalizes the sorting process of possibility values.

We define $I(f) = \displaystyle\int_0^1 \frac{1-\tilde{f}(x)}{x}\, dx$. We use it instead of more obvious $\displaystyle\int_0^1 \frac{\tilde{f}(x)}{x}\, dx$ to avoid a singularity at 0. Our expression can in fact be viewed as the distance to the function which is identically 1 and represents a possibilistic uniform distribution. In the discrete case it would correspond to the distance to a distribution consisting of $|X| = n$ values 1. Thus for $\pi : X \to [0,1]$ we put $I(\pi) = U(1,\ldots,1) - U(\pi) = \log n - U(\pi)$.

(To define a general distance, we put for $f_1 \leq f_2$

$$g(f_1, f_2) = \int_0^1 \frac{\tilde{f}_2(x) - \tilde{f}_1(x)}{x}\, dx$$

and then extend it to arbitrary f_1, f_2 [3].) If the domain of discourse is an arbitrary set X the rearrangement results in a descending function over an interval corresponding to the measure of X. It is then only required to change limits of integration to correspond to the end-points of that interval.

Our preliminary results indicate that the complete theory, built for discrete distributions, carries to the setting of continuous (or measurable) distributions. It is significant that all the information values can be obtained through approximations by discrete distributions. It means that the discrete cases can be viewed as 'imperfect' approximations of an idealized continuous description.

Example 1. Let us consider possibility distributions represented by $f(x) = x^k, \quad k = 0, 1, \ldots$. Denoting $J_k = I(x^k)$ and remembering that $\tilde{x}^k = (1-x)^k$, we find

$$J_0 = I(1) = \int_0^1 \frac{1-1}{x} dx = 0$$

$$J_1 = I(x) = \int_0^1 \frac{1-(1-x)}{x} dx = 1$$

$$J_2 = I(x^2) = \int_0^1 \frac{1-(1-x)^2}{x} dx = \int_0^1 (2-x) dx = \frac{3}{2}$$

To find a general expression for J_k, we first compute $J_k - J_{k-1}$

$$\int_0^1 \frac{(1-(1-x)^k)-(1-(1-x)^{k-1})}{x}\,dx =$$

$$\int_0^1 \frac{(1-x)^{k-1}-(1-x)^k}{x}dx = \int_0^1 (1-x)^{k-1}dx = \frac{1}{k}$$

As $J_0 = 0$ we have $J_k = 1 + \frac{1}{2} + \cdots + \frac{1}{k} = H_k$, the kth harmonic number.

Finally, we remark that $I(f)$ can be approximated as a limit of $I(\pi^{(n)})$, where $\pi^{(n)}$ are discrete distributions approximating f. Already a non-trivial example is offered by a linear function.

Example 2. We select $f(x) = 1 - x$ and approximate it using the values at $\frac{1}{n}, \frac{2}{n}, \ldots 0$. The approximating distributions are $\pi^{(n)} = (\frac{n-1}{n}, \frac{n-2}{n}, \ldots 1)$ and have as their information measures

$$\begin{aligned} U(\pi^{(n)}) &= \sum (p_i - p_{i+1}) \ln i = \\ &= \sum \left(\frac{n-i}{n} - \frac{n-i-1}{n} \right) \ln i = \\ &= \frac{1}{n} \sum \ln i = \frac{1}{n} \ln n! \end{aligned}$$

Invoking Stirling's formula, we find

$$\ln n! = n \ln n - n + O(\ln n),$$

and therefore

$$U(\pi^{(n)}) \sim \ln n - 1, \qquad n \to \infty.$$

This gives

$$I(\pi^{(n)}) = \ln n - U(\pi^{(n)}) \sim 1$$

which agrees with $I(f)$.

A more general $f(x) = x^k$ leads to an interesting identity involving limits of logarithmic sums. Approximating as above we get

$$\begin{aligned} U(\pi^{(n)}) &= \sum (\tilde{p}_i^k - \tilde{p}_{i+1}^k) \ln i \\ &= \sum \frac{(n-i)^k - (n-i-1)^k}{n^k} \ln i \end{aligned}$$

Expanding with respect to $n - i$ and retaining only the highest order terms gives

$$U(\pi^{(n)}) \sim \sum \frac{k(n-i)^{k-1}}{n^k} \ln i$$

Substituting into $\lim_{n\to\infty} I(\pi^{(n)}) = I(x^k)$ finally leads to

$$\lim_{n\to\infty} \left(\ln n - \sum \frac{k(n-i)^{k-1}}{n^k} \ln i \right) = H_k$$

5. INFORMATION INEQUALITIES IN DEMPSTER–SCHAFER THEORY

We consider a finite domain of discourse X, with cardinality $|X| = n$, endowed with a *basic evidence assignment* μ concentrated on the *focal subsets* $A_i \subset X$. We put $\mu(A_i) = m_i$, where $0 \le m_i \le 1$ and $\sum m_i = 1$.

For this theory there were developed essentially two separate information measures [1]. The first one – *nonspecificity* $N(\mu)$ expresses how specific is the assignment. It is represented by a sum

$$N(\mu) = \sum m_i \log |A_i|.$$

The other one – $S(\mu)$ addresses the question of how much overall information results from mutual support lend by the focal subsets. Given a focal set B we first define its *plausibility* $Pl(B) = \sum_{A_i \cap B \neq \emptyset} m_i$. Now $S(\mu) = -\sum m_i \log Pl(A_i)$.

Klir [6] has proposed studying a combined information measure

$$J(\mu) = N(\mu) + S(\mu) = \sum m_i \log \frac{|A_i|}{Pl(A_i)}.$$

He in particular suggested that this measure has as its maximum value $\log |X|$ (for a given overall domain X), and asked for a description of evidence assignments which have this information value. The answer is given in the next two theorems. We let F be a concave, monotonically increasing function which satisfies $\lim_{x\to 0} xF(x) = 0$ and is defined for all positive real numbers. Thus if $F(x)$ is twice differentiable, then $F'(x) \ge 0$, $F''(x) \le 0$. An example of such function is $\log x$, and conclusions apply to the question as posed.

Theorem 3.

$$\sum m_i F\left(\frac{|A_i|}{Pl(A_i)}\right) \le F\left(\sum m_i \frac{|A_i|}{Pl(A_i)}\right) \le F(|X|).$$

Proof. Function F being upward concave and m_i's summing to 1, we can apply Jensen's inequality [5] and obtain the first inequality. We note that the equality occurs here only if all terms $\dfrac{|A_i|}{Pl(A_i)}$ are equal.

As F is monotonically increasing, the other inequality will follow if we show that

$$\sum m_i \frac{|A_i|}{Pl(A_i)} \le |X|$$

We rewrite the left-hand side as

$$\sum |A_i| \frac{m(A_i)}{Pl(A_i)}$$

and consider it a counting expression for a multiset, where elements are those from all A_i's separately, and where every $x \in X$ is entered as many times as the number of sets A_i to which it belongs. However, we shall count each element only with a *weighted* multiplicity – if it is contributed by A_i we give it weight $\frac{m_i}{Pl(A_i)}$.

Now the combined multiplicity of an element $x \in X$ is

$$\sum_{x \in A_i} \frac{m_i}{Pl(A_i)}$$

Let the summation be over the sets $A_1, \ldots, A_k$. As x is simultaneously in all $A_1, \ldots, A_k$, they all have pairwise non-empty intersections and all contribute to one another plausibilities. In particular $Pl(A_i) \geq m(A_1) + \cdots + m(A_k)$ and the combined multiplicity of any given element $x \in X$ is

$$\sum_{i=1}^{k} \frac{m_i}{Pl(A_i)} \leq \sum_{i=1}^{k} \frac{m_i}{m_1 + \cdots + m_k} = 1$$

They are n elements in X and thus the sum of their multiplicities is less or equal to n – the cardinality of X. We remark that the equality here requires that (in the notation above) $Pl(A_i) = m_1 + \cdots + m_k$ for all A_i. This would imply that any set from among all A_j that intersects A_i must contain the selected element x. As x could be any element of A_i it means that A_i would be a subset of every A_j it intersects. This is possible only if A_i's are all disjoint. Finally, all $x \in X$ must be accounted for; therefore A_i's form a partition of X.

Theorem 4. Combined information value of the evidence assignment $J(\mu)$ reaches its maximum value $\log |X|$ exactly when it consists of disjoint focal subsets A_i covering X and assigns weights m_i proportional to cardinalities $|A_i|$. Thus

$$A_i \cap A_j = \emptyset, \quad i \neq j, \qquad \bigcup A_i = X, \qquad m_i = \mu(A_i) = \frac{|A_i|}{|X|}.$$

Proof. Immediate from the previous discussion. □

ACKNOWLEDGEMENT

The author wishes to thank the anonymous reviewer for a meticulous study of the paper.

REFERENCES

[1] D. Dubois and H. Prade: Possibility Theory. Plenum Press, New York 1988.

[2] D. Dubois and H. Prade: Properties of measures of information in evidence and possibility theories. Fuzzy Sets Systems *24* (1987), 161 – 182.

[3] M. Higashi and G. Klir: On the notion of distance representing information closeness. Internat. J. General Systems *9* (1983), 103 – 115.

[4] M. Higashi and G.Klir: Measures of uncertainty and information based on possibility distributions. Internat. J. General Systems *8* (1982), 43 – 58.

[5] G. Hardy, J. Littlewood and G. Polya: Inequalities. Cambridge University Press, Cambridge 1934.

[6] G. Klir: A principle of uncertainty and information invariance. Internat. J. General Systems *17* (1990), 249 – 276 .

[7] G. Klir and M. Mariano: On the uniqueness of possibilistic measure of uncertainty and information. Fuzzy Sets Systems *2* (1987), 197 – 220.

[8] A. Ramer: Structure of possibilistic information metrics and distances. Internat. J. General Systems *17* (1990), 21 – 32, and *18* (1990), 1 – 10.

[9] A. Ramer: Certain Inequalities Related to Rearrangements. Technical Report OU-PPI-TR-89-01, University of Oklahoma, Norman, OK 1989 (submitted to Eur. J. Combinatorics, 1991.)

[10] A. Ramer and L. Lander: Classification of possibilistic uncertainty and information functions. Fuzzy Sets Systems *24* (1987), 221 – 230.

[11] G. Shafer: A Mathematical Theory of Evidence. Princeton University Press, Princeton 1976.

Prof. Dr. Arthur Ramer, School of Electrical Engineering and Computer Science, University of Oklahoma, 200 Felgar St., Room 114, Norman, OK 73019. U. S. A.