An Update on Generalized Information Theory

GEORGE J. KLIR
Binghamton University (SUNY), USA

Abstract

The purpose of this paper is to survey recent developments and trends in the area of generalized information theory (GIT) and to discuss some of the issues of current interest in GIT regarding the measurement of uncertainty-based information for imprecise probabilities on finite crisp sets.

Keywords

uncertainty, uncertainty-based information, generalized information theory

1 Introduction

The term "Generalized Information Theory" (GIT) was introduced in the early 1990s to name a research program whose objective was to develop a broader treatment of uncertainty-based information, not restricted to the classical notions of uncertainty [6]. In GIT, the primary concept is uncertainty, and information is defined in terms of uncertainty reduction.

The basic tenet of GIT is that uncertainty can be formalized in many different ways, each based on some specific assumptions. To develop a fully operational theory for some conceived type of uncertainty, we need to address issues at four levels:

- LEVEL 1 – we need to find an appropriate mathematical representation of the conceived type of uncertainty
- LEVEL 2 – we need to develop a calculus by which this type of uncertainty can be properly manipulated
- LEVEL 3 – we need to find a meaningful way of measuring the amount of relevant uncertainty in any situation formalizable in the theory
- LEVEL 4 – we need to develop methodological aspects of the theory

GIT is an outgrowth of two classical uncertainty theories. The older one, which is also simpler and more fundamental, is based on the notion of possibility. The newer one, which has been considerably more visible, is based on the notion of probability. Proper ways of measuring uncertainty in these classical theories were established, respectively, by Hartley [5] and Shannon [12]. Basic features of the theories are outlined in [8].

The various nonclassical uncertainty theories in GIT are obtained by expanding the conceptual framework upon which the classical theories are based. At this time, the expansion is two-dimensional. In one dimension, the formalized language of the classical set theory is expanded to a more expressive language of *fuzzy set theory*, where further distinctions are based on various special types of fuzzy sets [10]. In the other dimension, the classical (additive) measures theory [4] is expanded to a less restrictive *fuzzy measure theory* [14], within which further distinctions are made by using fuzzy measures with various special properties. This expanded conceptual framework is a broad base for formulating and developing various theories of imprecise probabilities.

The subject of this paper is to discuss some of the issues of current interest regarding the measurement of uncertainty for imprecise probabilities on finite crisp sets. The various issues of possible fuzzifications of imprecise probabilities and of imprecise probablities on infinite sets are not addressed here. To facilitate the discussion, some common characteristics of imprecise probabilities on finite crisp sets are introduced in Section 2.

2 Imprecise Probabilities: Some Common Characteristics

One of the common characteristics of imprecise probabilities on finite crisp sets is that evidence within each theory is fully described by a *lower probability function (or measure)*, $\underline{g}$, or, alternatively, by an *upper probability function* (or measure) $\overline{g}$. These functions are always regular fuzzy measures that are superadditive and subadditive [14], respectively, and

$$\sum_{x\in X}\underline{g}(\{x\})\leq 1, \sum_{x\in X}\overline{g}(\{x\})\geq 1. \tag{1}$$

In the various special theories of uncertainty, they possess additional special properties.

When evidence is expressed (at the most general level) in terms of an arbitrary closed and convex set $\mathcal{D}$ of probability distribution functions p on a finite set X, functions $\underline{g}_{\mathcal{D}}$ and $\overline{g}_{\mathcal{D}}$ associated with $\mathcal{D}$ are determined for each $A\in\mathcal{P}(X)$ by the formulas

$$\underline{g}_{\mathcal{D}}(A)=\inf_{p\in\mathcal{D}}\sum_{x\in A}p(x) \text{ and } \overline{g}_{\mathcal{D}}(A)=\sup_{p\in\mathcal{D}}\sum_{x\in A}p(x).$$

Since

$$\sum_{x \in A} p(x) + \sum_{x \notin A} p(x) = 1,$$

for each $p \in \mathcal{D}$ and each $A \in \mathcal{P}(X)$, it follows that

$$\overline{g}_{\mathcal{D}}(A) = 1 - \underline{g}_{\mathcal{D}}(\overline{A}). \tag{2}$$

Due to this property, functions $\underline{g}_{\mathcal{D}}$ and $\overline{g}_{\mathcal{D}}$ are called *dual* (or *conjugate*). One of them is sufficient for capturing given evidence; the other one is uniquely determined by (2). It is common to use the lower probability function $\underline{g}_{\mathcal{D}}$ to capture the evidence.

As is well known [2, 3], any given lower probability function $\underline{g}_{\mathcal{D}}$ is uniquely represented by a set-valued function $m_{\mathcal{D}}$ for which $m_{\mathcal{D}}(\emptyset) = 0$ and

$$\sum_{A \in \mathcal{P}(X)} m_{\mathcal{D}}(A) = 1. \tag{3}$$

Any set $A \in \mathcal{P}(X)$ for which $m_{\mathcal{D}}(A) \neq 0$ is often called a *focal set*, and the family of all focal sets, $\mathcal{F}$, with the values assigned to them by function $m_{\mathcal{D}}$ is called a *body of evidence*. Function $m_{\mathcal{D}}$ is called a *Möbius representation* of $\underline{g}_{\mathcal{D}}$ when it is obtained for all $A \in \mathcal{P}(X)$ via the *Möbius transform*

$$m_{\mathcal{D}}(A) = \sum_{B|B \subseteq A} (-1)^{|A-B|} \underline{g}_{\mathcal{D}}(B), \tag{4}$$

where $|A-B|$ denotes the cardinality of the finite set $A-B$. The inverse transform is defined for all $A \in \mathcal{P}(X)$ by the formula

$$\underline{g}_{\mathcal{D}}(A) = \sum_{B|B \subseteq A} m_{\mathcal{D}}(B). \tag{5}$$

It follows directly from (2) that

$$\overline{g}_{\mathcal{D}}(A) = \sum_{B|B \cap A \neq \emptyset} m_{\mathcal{D}}(B). \tag{6}$$

for all $A \in \mathcal{P}(X)$.

Assume now that evidence is expressed in terms of a given lower probability function $\underline{g}$. Then, the set of probability distribution functions that are consistent with $\underline{g}$, $\mathcal{D}(\underline{g})$, which is always closed and convex, is defined as follows:

$$\mathcal{D}(\underline{g}) = \{p(x)|x \in X, p(x) \in [0,1], \sum_{x \in X} p(x) = 1, \text{ and}$$
$$\underline{g}(A) \leq \sum_{x \in A} p(x) \text{ for all } A \in \mathcal{P}(X)\}. \tag{7}$$

That is, each given function $\underline{\mathrm{g}}$ is associated with a unique set $\mathcal{D}$ and vice-versa.

3 Measures of Uncertainty

A *measure of uncertainty* of some conceived type in a given theory of imprecise probabilities is a functional, U, that assigns to each lower probability function in the theory a nonnegative real number. This number is supposed to measure, in an intuitively meaningful way, the amount of uncertainty of the considered type that is embedded in the lower probability function. To be acceptable as a measure of the amount of uncertainty, the functional U must satisfy several intuitively essential axiomatic requirements. Considering the most general level, when evidence is represented in terms of an arbitrary closed and convex set $\mathcal{D}$ of probability distribution functions p on finite set $X \times Y$, function U must satisfy the following requirements:

1. *Subadditivity*: $U(\mathcal{D}) \leq U(\mathcal{D}_X) + U(\mathcal{D}_Y)$, where

$$\mathcal{D}_X = \{p_X | p_X(x) = \sum_{y \in Y} p(x,y) \text{ for some } p \in \mathcal{D}\},$$

$$\mathcal{D}_Y = \{p_Y | p_Y(y) = \sum_{x \in X} p(x,y) \text{ for some } p \in \mathcal{D}\}.$$

2. *Additivity*: $U(\mathcal{D}) = U(\mathcal{D}_X) + U(\mathcal{D}_Y)$ if and only if $\mathcal{D}_X$ and $\mathcal{D}_Y$ are not interactive, which means that for all $A \in \mathcal{P}(X)$ and all $B \in \mathcal{P}(X)$, $m_{\mathcal{D}}(A \times B) = m_{\mathcal{D}_X}(A) \cdot m_{\mathcal{D}_Y}(B)$ and $m_{\mathcal{D}}(R) = 0$ for all $R \neq A \times B$.

3. *Monotonicity*: if $\mathcal{D} \subseteq \mathcal{D}'$, then $U(\mathcal{D}) \subseteq U(\mathcal{D}')$; and similarly for $\mathcal{D}_X$ and $\mathcal{D}_Y$.

4. *Range*: if uncertainty is measured in bits, then $U(\mathcal{D}) \in [0, log_2|X \times Y|]$, and similarly for $\mathcal{D}_X$ and $\mathcal{D}_Y$.

The requirement of subadditivity and additivity, as stated here, are generalized counterparts of the classical requirements of subadditivity and additivity for probabilistic and possibilistic measures of uncertainty. The requirement of monotonicity (not applicable to classical probabilistic uncertainty) means that reducing the set of probability distributions consistent with a given lower (or upper) probability function cannot increase uncertainty. The requirement of range, which depends on the choice of measurement units, is defined by the two extreme cases: the full certainty and the total ignorance.

When distinct types of uncertainty coexist in a given uncertainty theory, it is not necessary that these requirements be satisfied by each uncertainty type. However, they must be satisfied by an overall uncertainty measure, which appropriately aggregates measures of the individual uncertainty types.

It is well established that two types of uncertainty coexist in all theories of imprecise probabilities [8, 9]. They are generalized counterparts of the classical possibilistic and probabilistic uncertainties. They are measured, respectively, by appropriate generalizations of the Hartley and Shannon measures.

4 Generalized Hartley Measures

An historical overview of efforts to generalize the classical Hartley measure of uncertainty can be found in [9]. Its full generalization (to arbitrary closed and convex sets of probability distributions) was completed fairly recently by Abellan and Moral [1]. They showed that the functional

$$GH(m_{\mathcal{D}}) = \sum_{A \in \mathcal{F}} m_{\mathcal{D}}(A) \log_2 |A|, \tag{8}$$

where $m_{\mathcal{D}}$ is the Möbius representation of the lower probability associated with a given closed and convex set $\mathcal{D}$ of probability distributions, satisfies all the essential axiomatic requirements defined in Sec. 3 (subadditivity, additivity, etc.). Moreover, this functional is also directly connected with the classical Hartley measure: it is the weighted average of the Hartley measure for each given body of evidence $(\mathcal{F}, m_{\mathcal{D}})$.

It is fairly obvious that the functional *GH* defined by (8) measures the lack of specificity in evidence. Large focal elements result in less specific predictions, diagnoses, etc., than their smaller counterparts. The type of uncertainty measured by *GH* is thus well characterized by the term nonspecificity.

Observe that $GH(m_{\mathcal{D}}) = 0$ for precise probabilities, where $\mathcal{D}$ consists of a single probability distribution function, which is expressed in (8) by function $m_{\mathcal{D}}$. All focal sets are in this case singletons. Evidence expressed by precise probabilities is thus fully specific.

Eq. (8) is clearly applicable only to functions $m_{\mathcal{D}}$ defined on finite sets. It must be properly modified when $m_{\mathcal{D}}$ is defined on the n-dimensional Euclidean space for some $n \geq 1$, as shown in [9]. However, this modification is not a subject of this paper.

5 Generalized Shannon Measures

There have been many promising, but eventually unsuccessful efforts to generalize the classical Shannon measure (usually referred to as the *Shannon entropy*). Virtually all these efforts were based on the recognition that the Shannon entropy measures the mean (expected) value of the conflict among evidential claims expressed by a single probability distribution function on a finite set of mutually exclusive alternatives [9]. An historical overview of most of these efforts is given in [9].

All the proposed generalizations of the Shannon entropy were intuitively promising as measures of conflict among evidential claims in general bodies of evidence, but each of them was eventually found to violate the essential requirement of subadditivity. In fact, no generalized Shannon entropy can be subadditive on its own, as is shown in [13]. The subadditivity may be obtained only in terms of the total uncertainty — an aggregate of the two coexisting types of uncertainty (nonspecifivity and conflict). However, when the total uncertainty is viewed as the sum of the generalized Hartley measure with the various candidates for the generalized Shannon entropy, none of these aggregated uncertainty measures is still subadditive, as demonstrated by relevant counterexamples in each case [13].

The latest promising candidate (not previously analyzed in terms of the requirement of subadditivity) is based on the so-called Shapley index, which plays an important role in game theory [11, 15]. For any given finite universal set X, this candidate for the generalized Shannon entropy, GS, is defined as the average Shannon entropy of differences in a given lower probability (or, alternatively, an upper probability) for all maximal chains in the lattice $(\mathcal{P}(X), \subseteq)$. Unfortunately, the sum $GH + GS$ does not satisfy in this case again the requirement of subadditivity. This can be demonstrated by the following counterexample.

Let $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2\}$, and let us consider a body of evidence on $X \times Y$ whose Möbius representation is:

$$\begin{aligned} m(\{(x_1,y_1),(x_2,y_2),(x_2,y_1)\}) &= a, \\ m(X \times Y) &= 1-a, \end{aligned}$$

where $a \in [0,1]$. Then, $m_X(X) = m_Y(Y) = 1$, and, hence, $GS_X(m_X) = GS_Y(m_Y) = 0$ and $GH_X(m_X) + GH_Y(m_Y) = 2$. Furthermore,

$$\begin{aligned} GS(m) &= [-a\log_2 a - (1-a)log_2(1-a)]/4, \\ GH(m) &= alog_2 3 + 2 - 2a \end{aligned}$$

For subadditivity of $GH + GS$, the difference

$$\begin{aligned} \Delta &= (GH_X + GH_Y + GS_X + GS_Y) - (GH + GS) \\ &= [a\log_2 a + (1-a)log_2(1-a)]/4 + 2a - alog_2 3 \end{aligned}$$

is required to be nonnegative for all values $a \in [0,1]$. However, Δ is negative in this case for any value $a \in (0, 0.58)$ and it reaches its minimum, $\Delta = -0.1$, at $a = 0.225$.

6 Total Uncertainty Measures

Generalized Shannon measure, GS, was eventually defined indirectly, via an *aggregated uncertainty*, AU, covering both nonspecificity and conflict, and the well established generalized Hartley measure of nonspecificity, GH, defined by (8). Since it must be that $GH + GS = AU$, the generalized Shannon measure can be defined as

$$GS = AU - GH \tag{9}$$

Using this definition, the unsuccessful effort to find *GS* directly is replaced with the effort to find *AU* and define *GS* indirectly via Eq. (9). The latter effort was successful in the mid 1990s, when a functional *AU* satisfying all essential requirements was established in evidence theory [9]. However, this functional is applicable to all the other theories of imprecise probabilities as well, which follows from the common properties shared by these theories (Sec. 2). Given any lower probability function $\underline{g}_{\mathcal{D}}$ associated with a closed convex set $\mathcal{D}$ of probability distributions (or vice versa), $AU(\underline{g}_{\mathcal{D}})$ is defined by the formula

$$AU(\underline{g}_{\mathcal{D}}) = \max_{p\in\mathcal{D}}[-\sum_{x\in X} p(x)\log_2 p(x)]. \tag{10}$$

It is the maximum Shannon entropy within $\mathcal{D}$. An efficient algorithm for computing this maximum, which was proven correct for belief functions of evidence theory [9], is applicable without any change when belief functions are replaced with arbitrary lower probability functions of any other kind.

Given an arbitrary lower probability function $\underline{g}$ on $\mathcal{P}(X)$, the generalized version of this algorithm consists of the following seven steps:

Step 1. Find a non-empty set $A \subseteq X$, such that $\underline{g}(A)/|A|$ is maximal. If there are more such sets than one, take the one with the largest cardinality.

Step 2. For all $x \in A$, put $p(x) = \underline{g}(A)/|A|$.

Step 3. For each $B \subseteq X - A$, put $\underline{g}(B) = \underline{g}(B \cup A) - \underline{g}(A)$.

Step 4. Put $X = X - A$.

Step 5. If $X \neq \emptyset$ and $\underline{g}(X) > 0$, then go to Step 1.

Step 6. If $\underline{g}(X) = 0$ and $X \neq \emptyset$, then put $p(x) = 0$ for all $x \in X$.

Step 7. Calculate $AU = -\sum_{x\in X} p(x)\log_2 p(x)$.

Although functional *AU* is a well-justified measure of total uncertainty in the various theories of uncertainty, it is highly insensitive to changes in evidence due to its aggregated nature. It is an aggregate of the two coexisting types of uncertainty, nonspecificity and conflict. It is thus desirable to express the total uncertainty, TU, in a disaggregated form

$$TU = (GH, GS), \tag{11}$$

where *GH* is defined by (8) and *GS* is defined by (9) and (10). It is assumed here that the axiomatic requirements are defined in terms of the sum of the two functionals involved, which is always the well-justified aggregate measure *AU*. In this sense the measure satisfies trivially all the requirements. Its advantage is that measures of both types of uncertainty that coexist in uncertainty theory employed (nonspecificity and conflict) are expressed explicitly and, consequently, the measure is sensitive to changes in evidence.

To appreciate the difference between AU and TU, let us consider three simple examples of given evidence within a finite universal set X and let $|X| = n$ for convenience: (i) in the case of total ignorance (when $m(X) = 1$), we obtain $AU = \log_2 n$ and $TU = (\log_2 n, 0)$; (ii) when evidence is expressed by the uniform probability distribution on X, then again we have $AU = \log_2 n$, but $TU = (0, \log_2 n)$; (iii) when evidence is expressed by $m(\{x\}) = a$ for all $x \in X$ and $m(X) = 1 - na$, then again $AU = \log_2 n$ for all values $a \leq 1/n$, while

$$TU = ((1-na)\log_2 n, na\log_2 n).$$

It is clear that TU defined by (11) possesses all the required properties in terms of the sum of its components, since $GH + GS = AU$. Moreover, as was proven by Smith [13], $GS \geq 0$ for all bodies of evidence. Additional properties of GS defined by (9) can be determined by employing the algorithm for computing AU, as shown for some properties in Section 7.

It is also reasonable to express the generalized Shannon entropy by the interval $[\underline{S}, \overline{S}]$, where $\underline{S}$ and $\overline{S}$ are, respectively, the minimum and maximum values of the Shannon entropy within the set of all probability distributions that are consistent with a given lower probability function. Clearly $\overline{S} = AU$ and $\underline{S}$ is defined by replacing max with min in Eq. (10). Then, the total uncertainty, TU', has the form

$$TU' = (GH, [\underline{S}, \overline{S}]). \tag{12}$$

Let us define a partial ordering of these total uncertainties as follows:

$$TU'_1 \leq TU'_2 \text{ iff } GH_1 \leq GH_2 \text{ and } [\underline{S}_1, \overline{S}_1] \subseteq [\underline{S}_2, \overline{S}_2].$$

Then, due to subadditivity of $\overline{S}$, subadditivity of TU' is guaranteed. Indeed,

$$[\underline{S}_X + \underline{S}_Y, \overline{S}_X + \overline{S}_Y] \not\subset [\underline{S}, \overline{S}]$$

for any joint and associated marginal bodies of evidence. However, no algorithm for computing $\underline{S}$ that has been proven correct is available as yet.

7 Some Properties of Generalized Shannon Entropy

The purpose of this section is to examine the generalized Shannon entropy defined by (9). To facilitate this examination, let

$$\mathcal{F} = \{A_i | A_i \in \mathcal{P}(X), i \in \mathbb{N}_q\}$$

denote the family of all focal sets of a given body of evidence, where $\mathbb{N}_q = \{1,2,\ldots,q\}$ for some integer q, and let $m_i = m(A_i)$ for convenience. Moreover, let

$$E = \bigcup_{i \in \mathbb{N}_q} A_i.$$

The algorithm for computing $\overline{S}(= AU)$ produces a partition,

$$\mathcal{E} = \{E_k | k \in \mathbb{N}_r, r \leq q\}$$

of E. For convenience, assume that block E_k of this partition was produced in k-th iteration of the algorithm and let $e_k = |E_k|$. Then

$$\overline{S}(m) = -\sum_{k \in \mathbb{N}_r} \underline{g}_k \log_2(\underline{g}_k / e_k)$$

where $\underline{g}_k$ denotes the lower probability of E_k in k-th iteration of the algorithm. This equation can be rewritten as

$$\overline{S}(m) = -\sum_{k \in \mathbb{N}_r} \underline{g}_k \log_2 \underline{g}_k + \sum_{k \in \mathbb{N}_r} \underline{g}_k \log_2 e_k.$$

It follows from this equation and from Eq. (9) that

$$GS(m) = S(\underline{g}_k | k \in \mathbb{N}_r) + GH(\underline{g}_k | k \in \mathbb{N}_r) - GH(m), \tag{13}$$

where S denotes the Shannon entropy.

Assume now that $\mathcal{F}$ consists of pair-wise disjoint focal sets. Then, the Möbius representation, m, is a positive function since any negative value m_i for some $A_i \in \mathcal{F}$ would clearly violate in this case the requirement that values of the associated lower probability function must be in $[0,1]$. When applying the algorithm for computing $\overline{S}$ to our case, it turns out that the values m_i for all $A_i \in \mathcal{F}$ are uniformly distributed among elements of each focal set A_i. This only requires to prove that

$$\sum_{i \in I} m_i / \sum_{i \in I} a_i \leq m_k / a_k$$

for each $k \in I$ and all nonempty sets $I \subseteq \mathbb{N}_q$, where $a_k = |A_k|$. The proof of this inequality, which is omitted here due to limited space, can be obtained by the method of contradiction. The maximum entropy probability distribution function, p, for the given body of evidence is thus defined for all $x_{i_k} \in A_i (k \in \mathbb{N}_{|A_i|})$ and all

$A_i \in \mathcal{F}$ by the formula $p(x_{i_k}) = m_i/a_i$ where $a_i = |A_i|$. Hence,

$$\begin{aligned}
\overline{S}(m) &= -\sum_{i=1}^{q}\sum_{k=1}^{a_i} p(x_{i_k})\log_2 p(x_{i_k}) \\
&= -\sum_{i=1}^{q} m_i \log_2(m_i/a_i) \\
&= -\sum_{i=1}^{q} m_i \log_2 m_i + \sum_{i=1}^{q} m_i \log_2 a_i \\
&= -\sum_{i=1}^{q} m_i \log_2 m_i + GH(m).
\end{aligned}$$

Consequently,

$$GS(m) = -\sum_{i=1}^{q} m_i \log_2 m_i.$$

This is clearly a property that we would expect, on intuitive grounds, the generalized Shannon entropy to satisfy.

To examine some properties of the generalized Shannon entropy for nested bodies of evidence, let $X = \{x_i | i \in \mathbb{N}_n\}$ and assume that elements of X are ordered in such a way that the family

$$\mathcal{A} = \{\mathcal{A}_i = \{x_1, x_2, \ldots, x_i\} | i \in \mathbb{N}_n\}$$

contains all focal sets. That is, $\mathcal{F} \subseteq \mathcal{A}$. For convenience, let $m_i = m(A_i)$ for all $i \in \mathbb{N}_n$.

To express $GS(m)$, we need to express $GH(m)$ and $\overline{S}(m)$. Clearly,

$$GH(m) = \sum_{i=1}^{n} m_i \log_2 i \tag{14}$$

To express $\overline{S}(m)$, three cases must be distinguished in terms of values m_i:

(a) $m_i \geq m_{i+1}$ for all $i \in \mathbb{N}_{n-1}$;

(b) $m_i \leq m_{i+1}$ for all $i \in \mathbb{N}_{n-1}$;

(c) neither (a) nor (b).

Following the algorithm for computing $\overline{S}$, we obtain the formula

$$GS_a(m) = -\sum_{i=1}^{n} m_i \log_2(m_i i) \tag{15}$$

for any function m that conforms to Case (a). By applying the method of Lagrange multipliers, we can readily find out that the maximum, $GS_a^*(n)$, of this functional for some $n \in \mathbb{N}$ is obtained for

$$m_i = (1/i)2^{(-1/\ln 2+\alpha)}(i \in \mathbb{N}_n), \tag{16}$$

where the value of α is determined by solving the equation

$$2^{-(1/\ln 2+\alpha)} \sum_{i=1}^{n} (1/i) = 1.$$

Let $s_n = \sum_{i=1}^{n} (1/i)$. Then,

$$\alpha = -\log_2(1/s_n) - (1/\ln 2)$$

and, hence,

$$\begin{aligned} m_i &= (1/i)2^{\log_2(1/s_n)} \\ &= 1/(is_n). \end{aligned}$$

Substituting this expression for m_i in (15), we obtain

$$\begin{aligned} GS_a^*(n) &= \sum_{i=1}^{n} (1/i)(1/s_n)\log_2 s_n \\ &= [(1/s_n)\log_2 s_n] \sum_{i=1}^{n} (1/i). \end{aligned}$$

Consequently,

$$GS_a^*(n) = \log_2 s_n. \tag{17}$$

In Case (b), $\overline{S} = \log_2 n$ and GH is given by (8). Hence,

$$GS_b(m) = \log_2 n - \sum_{i=1}^{n} m_i \log_2 i.$$

The maximum, $GS_b^*(n)$, of this functional for some $n \in \mathbb{N}$ subject to the inequalities that are assumed in Case (b), is obtained for $m_i = 1/n$. Hence,

$$GS_b^*(n) = \log_2 \frac{n}{n!^{1/n}}. \tag{18}$$

Employing Stirling's formula for approximating $n!$, it can be shown that

$$\begin{aligned} \lim_{n\to\infty} \log_2 \tfrac{n}{n!^{1/n}} &= \log_2 e \\ &= 1.442695. \end{aligned}$$

GS_b^* is thus bounded, contrary to $GS_a^*(n)$. Moreover, $GS_b^*(n) < GS_a^*(n)$, for all $n \in \mathbb{N}$.

Case (c) is more complicated for a general analytic treatment since it covers a greater variety of bodies of evidence with respect to the computation of *GS*. This follows from the algorithm for computing $\overline{S}$. For each given body of evidence, the algorithm partitions the universal set in some way, and distributes the value of the lower probability in each block of the partition uniformly. For nested bodies of evidence, the partitions preserve the induced order of elements of X. There are 2^{n-1} order preserving partitions. The most refined partition and the least refined one are represented by Cases (a) and (b), respectively. All the remaining $2^{n-1}-2$ partitions are represented by Case (c). A conjecture, based on a complete analysis for $n=3$ and extensive simulation experiments for $n>3$, is that the maxima of *GS* for all these partitions are for all $n \in \mathbb{N}$ smaller than the maximum GS_a^* for Case (a). According to this plausible conjecture, whose proof is an open problem, the difference between the maximum nonspecificity, $GH^*(n)$, and maximum conflict, $GS_a^*(n)$, grows rapidly with n. For example, $GH^*(2)=1$ and $GS_a^*(2)=0.585$, while $GH^*(10^4)=13.29$ and $GS_a^*(10^4)=3.29$. Similarly, the maximum value of conflict is 36.9% of the maximum value of total uncertainty for $n=2$, but it reduces to 19.8% for $n=10^4$. For nested (consonant) bodies of evidence, this feature makes intuitively a good sense.

8 Conclusions

For the last two decades or so, research in GIT has been focusing on developing justifiable ways of measuring uncertainty and the associated uncertainty-based information in the various emerging uncertainty theories. This objective is now, by and large, achieved. However, some research in this direction is still needed to improve our understanding of the generalized Shannon entropy, defined either by (9) or by the interval $[\underline{S},\overline{S}]$. Results presented in this paper are intended to contribute a little to this understanding.

In the years ahead, the focus of GIT will likely divide into two branches of research. One of them will focus on developing methodological tools based on our capability to measure uncertainty in the various established theories of uncertainty. Methodological tools for making the principles of uncertainty maximization, minimization, and invariance operational will in particular be sought due to the broad utility of these principles [7, 9]. The other branch of research will pursue the development of additional uncertainty theories. One direction in this research area will undoubtedly include a comprehensive investigation of the various ways of fuzzifying existing uncertainty theories.

References

[1] Abellán, J. and S. Moral, "A non-specificity measure for convex sets of probability distributions." *Intern. J. of Uncertainty, Fuzziness, and Knowledge-Based Systems*, **8**(3): 357–367, 1999.

[2] Chateauneuf, A. and Jaffray, J. Y., "Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion." *Mathematical Social Sciences*, **17**: 263–283, 1989.

[3] Grabisch, M. "The interaction and Möbius representations of fuzzy measures on finite spaces, k-additive measures: a survey." Grabish, M. et al., eds., *Fuzzy Measures and Integrals: Theory and Applications*. Springer-Verlag, New York, pp. 70–93.

[4] Halmos, P. R., *Measure Theory*. D. Van Nostrand, Princeton, NJ, 1950.

[5] Hartley, R. V. L., "Transmission of information." *The Bell System Technical J.*, **7**(3): 535–563, 1928.

[6] Klir, G. J., "Generalized information theory." *Fuzzy Sets and Systems*, **40**(1): 127–142, 1991.

[7] Klir, G. J., "Principles of uncertainty: What are they? Why do we need them?" *Fuzzy Sets and Systems*, **74**(1): 15–31, 1995.

[8] Klir, G. J., "Uncertainty-based information." Teodorescu, H. and P. Melo, eds, *Systemic Organization of Information in Fuzzy Systems*. IOS Press, Amsterdam, pp. 21–52, 2003.

[9] Klir, G. J. and Wierman M. J., *Uncertainty-Based Information: Elements of Generalized Information Theory* (Second Edition). Physica-Verlag/Springer-Verlag, Heidelberg and New York, 1999.

[10] Klir, G. J. and Yuan, B., *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, PTR, Upper Saddle River, NJ, 1995.

[11] Marichal, J. and Roubens, M., "Entropy of discrete fuzzy measures." *Intern. J. of Uncertainty, Fuzziness, and Knowledge-Based Systems*, **8**(6): 625–640, 2000.

[12] Shannon, C. E., "The mathematical theory of communication." *The Bell System Technical J.*, **27**(3&4): 379–423, 623–656, 1948.

[13] Smith, R. M., *Generalized Information Theory: Resolving Some Old Questions and Opening Some New Ones*. PhD. Dissertation, Binghamton University (SUNY), Binghamton, NY, 2000.

[14] Wang, Z. and Klir, G. J., *Fuzzy Measure Theory*. Plenum Press, New York, 1992.

[15] Yager, R. R., "On the entropy of fuzzy measures." *IEEE Trans. on Fuzzy Systems*, **8**(4): 453–461, 2000.

George J. Klir is with the Center for Intelligent Systems, Binghamton University (SUNY), Binghamton, New York, 13902 USA