Uncertainty and information: Foundations to generalized information theory. George J. Klir. Hoboken, NJ: John Wiley; 2006: 499pp. $94.95 (ISBN: 0-471-74867-6).

The book studies "Generalized Information Theory" (GIT), being generalizations of, basically, two theories: possibility based uncertainty theory and probability based uncertainty theory. In the whole book, information has to be considered as the reduction of uncertainty, the latter to be defined.

Classical possibility based uncertainty theory is the simplest and the oldest of the two theories. It measures how possible an event is (given a universe X and a subset $E \subset X$ being the set of all possibilities) if we are in a subset $A \subset X$. One defines the function

$$r_E(x) = \begin{cases} = 1 & \text{if } x \in E \\ = 0 & \text{if } x \notin E \end{cases} \qquad (1)$$

(a more classical notation is $\chi_E(x)$, the characteristic function of E – also used in this book) and, for any subset $A \subset X$ one defines the possibility value $Pos_E(A)$ as follows

$$Pos_E(A) = \max_{x \in A} r_E(x) \qquad (2)$$

hence defining the possibility function Pos_E. Also a "necessity function" Nec_E is defined by

$$Nec_E(A) = 1 - Pos_E(\bar{A}) \qquad (3)$$

where A is the complement of A in X (for the time being, all sets are ordinary sets, also called crisp sets as contrasted with so-called fuzzy sets – see later). Possibility based uncertainty theory was developed by Hartley (1928) where the amount of uncertainty associated with the (finite) set E is (essentially):

$$H(r_E) = \log_2 |E|, \qquad (4)$$

where $|E|$ denotes the cardinality of E.

The function H is called the Hartley measure and one proves that this measure is unique (up to a multiplicative constant) based on the logical requirement that (with an abuse of notation: $H(r_E) = H(n)$ with $n = |E|$):

$$H(n.m) = H(n) + H(m) \tag{5}$$

It is well-known that (5), together with continuity of H, implies (4) (up to a constant), cf. Roberts (1979), based on elementary observations of functions satisfying functional relations of the type (5) – see also Egghe (2005), Appendix I. It is surprising that in the present book this classical argument is not used nor referenced and that a more intricate argument (of Rényi) is presented, hereby not referring to Roberts (1979).

Less elementary is the theory of probability based uncertainty. The notion of probability distribution function is well-known: it is a function

$$p : X \to [0,1] \tag{6}$$

on a general set X (taken here to be finite) with range the closed interval $[0,1]$ such that

$$\sum_{x \in X} p(x) = 1 \tag{7}$$

From this one defines a probability measure Pro as follows: for any $A \subset X$, define

$$Pro(A) = \sum_{x \in A} p(x) \tag{8}$$

It is clear that for any two disjoint subsets $A, B \subset X$ we have that

$$Pro(A \cup B) = Pro(A) + Pro(B) \tag{9}$$

a property called additivity.

Probability based uncertainty theory has been developed in 1948 by Shannon (cf. Shannon (1948)): in this case the amount of uncertainty is measured as follows:

$$S(p) = -\sum_{x \in X} p(x) \log_2 p(x) \tag{10}$$

the so-called entropy of the system (a more common notation is $\bar{H}$ being the average of the pointwise information contents $\log_2 p(x)$. Note that, when $|X| = n$ and when all probability values $p(x)$ are equal, hence

$$p = p(x) = \frac{1}{n} \tag{11}$$

for all $x \in X$, formula (10) reduces to

$$S = \log_2 n = -\log_2 p \tag{12}$$

which is essentially the same as (4) which is not surprising since, essentially, probability aspects have been eliminated. The notion of entropy is the most important notion of information theory. It measures the (average) amount of information one has (i.e. reduction of uncertainty) when events appear, given the probability distribution p. There are numerous applications of this measure ranging from coding theory to automatic indexation to information retrieval and linguistics. Most importantly, formulae (10) and (12) (the building blocks of (10)) form the basis for the notion of "bits", i.e. the number of 0s and 1s to binary represent the elements in a set of cardinality n (fixed or non-fixed length coding) – see e.g. the classical Heaps (1978). Also the measure entropy is unique (up to a constant) based on the requirements that it is continuous and that the amount of information in two independent events is the sum of the amounts of information in the separate events (comparable with requirement (5)).

After an Introductory Chapter 1, classical possibility based uncertainty theory is discussed in Chapter 2 and classical probability based uncertainty theory is discussed in Chapter 3.

The rest of the book is devoted to extensions of these theories. One can distinguish two different ways of doing this: first by generalizing the notion of probability measure to monotone measures μ that are non-additive. Here monotone measures are measures satisfying the implication

$$A,B\subseteq X,\ A\subseteq B\Rightarrow \mu(A)\leq \mu(B) \tag{13}$$

which generalizes the additivity property. This occupies the attention in Chapters 4, 5 and 6.

A second way of extending these theories is by replacing ordinary sets by fuzzy sets (Chapters 7 and 8).

In Chapter 4 one studies monotone non-additive measures such as Choquet capacities, i.e. monotone measures μ that satisfy

$$\mu\left(\bigcup_{j=1}^{k} A_j\right) \geq \sum_{K\subseteq \mathbb{N}_k} (-1)^{|K|+1}\mu\left(\bigcap_{j\in K} A_j\right) \tag{14}$$

for all families of k subsets of the universe X (here $\mathbb{N}_k = \{1,\ldots,k\}$). Monotone non-additive measures also appear in the theory of imprecise probabilities. A clear example (also given in Chapter 4) is as follows. Suppose we have a universe XxY being the Cartesian product of two sets $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2\}$ (hence both are doubletons).

Assume that we know the marginal probabilities

$p_X(x_1)$, $p_X(x_2) = 1 - p_X(x_1)$, $p_Y(y_1)$, $p_Y(y_2) = 1 - p_Y(y_1)$ and we want to use this information to determine the unknown joint probabilities $p_{ij} = p(x_i, y_j)$ $(i,j = 1,2)$. Note that one way of doing this is to define

$$p_{ij} = p_X(x_i)p_Y(y_j) \tag{15}$$

$(i,j = 1,2)$ but there is not evidence that this is always the case. The general solution of the above problem is by remarking that

$$\begin{aligned}
&(a) \quad p_{11} + p_{12} = p_X(x_1)\\
&(b) \quad p_{21} + p_{22} = 1 - p_X(x_1) = p_X(x_2)\\
&(c) \quad p_{11} + p_{21} = p_Y(y_1)\\
&(d) \quad p_{12} + p_{22} = 1 - p_Y(y_1) = p_Y(y_2)
\end{aligned} \tag{16}$$

Note that (d)= (a)+ (b)- (c) and hence can be deleted. The other equations are linearly independent in the four variables p_{11}, p_{12}, p_{21} and p_{22}, so one can be choosen as a free variable:

$$\begin{aligned} p_{12} &= p_X(x_1) - p_{11} \\ p_{21} &= p_Y(y_1) - p_{11} \\ p_{22} &= 1 - p_X(x_1) - p_Y(y_1) + p_{11} \end{aligned} \tag{17}$$

Of course p_{11} is limited to

$$\max(0, p_X(x_1) + p_Y(y_1) - 1) \leq p_{11} \leq \min(p_X(x_1), p_Y(y_1)) \tag{18}$$

This yields a whole range of solutions, e.g. for the case $p_X(x_1) = 0.8$ and $p_Y(y_1) = 0.6$:

$$\begin{aligned} p_{11} &\in [0.4, 0.6] \\ p_{12} &= 0.8 - p_{11} \\ p_{21} &= 0.6 - p_{11} \\ P_{22} &= p_{11} - 0.4 \end{aligned} \tag{19}$$

So we have a whole set of possible probability distributions, called a credal set $\mathcal{D}$. Going back to the general notation X for the universe we can, for every $A \subseteq X$, define a lower probability function $\underline{\mu}(A)$ as the infimum over the credal set of the values $\sum_{x \in A} p(x)$ $(p \in \mathcal{D})$ and an upper probability function $\overline{\mu}(A)$ as the supremum of these same values. This yields monotone non-additive measures for which there is a need for extension of the classical uncertainty theories. The general theory of imprecise probabilities (of which the above is an example) is further developed in Chapter 4.

In Chapter 5, special theories of imprecise probabilities are presented. First one extends possibility theory to graded possibilities (i.e. where the Pos measure can have values in $[0,1]$ instead of $\{0,1\}$ - somewhat comparable with the extension of crisp sets to fuzzy sets). Then

the monotone Sugeno λ-measures ${}^{\lambda}\mu$ are introduced and studied. These are characterized by the requirement for all $A,B \subseteq X$ such that $A \cap B = \emptyset$:

$$ {}^{\lambda}\mu(A \cup B) = {}^{\lambda}\mu(A) + {}^{\lambda}\mu(B) + \lambda\, {}^{\lambda}\mu(A)\, {}^{\lambda}\mu(B) \tag{20} $$

with $\lambda > -1$ (a parameter). Note the confusing notation (e.g. ${}^{\lambda}\mu$). Other monotone measures are the "belief measures" (Bel) (being special Choquet capacities) and plausibility measures (Pl) (being variants of Choquet capacities, called alternating capacities) are also studied in this chapter. The theory based on these dual pairs is called Dempster-Shafer theory (DST). In order to extend the uncertainty theories (as will be done in Chapter 6) we also need the so-called Möbius representation m of Bel being

$$ m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} Bel(B) \tag{21} $$

replacing the probability distribution on X (m now acts on $\mathcal{P}(X)$, the set of all subsets of X). Finally, a set $A \subseteq X$ such that $m(A) > 0$ is called a focal set and $\mathcal{F}$ denotes the set of all focal sets induced by m ($\mathcal{F}$ is called a body of evidence).

In Chapter 6 the generalized measures of uncertainty are presented. First one gives the generalized Hartley measure for graded possibilities. For a possibility profile $r = (r_1, r_2, ..., r_n)$ (ordered decreasingly) and sets $A_i = \{x_1, ..., x_i\}$ $(i \in \{1, ..., n\})$ one defines the U-uncertainty as

$$ U(r) = \sum_{i=1}^{n} (r_i - r_{i+1}) \log_2 |A_i| $$

$$ = \sum_{i=2}^{n} (r_i - r_{i+1}) \log_2 i \tag{22} $$

The general Hartley measure in DST looks as:

$$ GH(m) = \sum_{A \in \mathcal{F}} m(A) \log_2 |A| \tag{23} $$

while generalized entropy in DST is given by the pair of measures

$$E(m) = -\sum_{A \in \mathcal{F}} m(A) \log_2 Pl(A) \quad (24)$$

$$C(m) = -\sum_{A \in \mathcal{F}} m(A) \log_2 Bel(A) \quad (25)$$

The author, however, recognizes that none of the above extensions of entropy are mathematically satisfactory since the subadditivity property of entropy is violated. The present reviewer is unable to comment further on the (lack of) quality of these measures. This reviewer, however, has the impression that, certainly from Chapter 7 on, the book deteriorates into "unnecessary generalizations for generalization's sake". They do not have greater expressive power as claimed (admitted) in conclusion's section 10.4. Not only does one present (in Chapters 7 and 8) fuzzy set theory involving very general definitions of complement, union and intersections (assumably containing the "classical" Zadeh min-max definitions, but it is not indicated whether or not the useful probabilistic sum and algebraic product operations are included as well) but one also presents nonstandard fuzzy sets, e.g. where membership functions range in closed subintervals of $[0,1]$ or even in fuzzy intervals of $[0,1]$ (so-called fuzzy sets of type 2) or where membership functions are defined on fuzzy subsets of the universe X (so-called fuzzy sets of level 2); even higher types and levels are defined.

Although Chapter 8 gives fuzzy set interpretations of possibility theory and of probability theory one does not (apparently) present general fuzzified Hartley or Shannon measures. The fact that these theories are underdeveloped is also recognized in the conclusions Chapter 10. The "methodological" Chapter 9 is a mixture of philosophical and mathematical principles underlying uncertainty. The first two principles discuss the principle of "minimum" and "maximum" uncertainty. The latter one is mathematically formulated while the former one is not. Let us start with the latter one. As recognized in the book under review this principle is better known as the "maximum entropy principle" (MEP) as it is also studied in Egghe and

Lafouge (2006). (MEP) can be formulated in a mathematically exact way as follows (as is also done in this book but there is a confusion between the given effort constraints $(c_1,...,c_n)$ and $(x_1,...,x_n)$ - we will use $(c_1,...,c_n)$ here which is the same n-tuple as $(E_1,...,E_n)$ in Egghe and Lafouge (2006)): Maximize

$$S(p) = -\sum_{i=1}^{n} p_i \ln p_i \tag{26}$$

subject to the constraints

$$E = \sum_{i=1}^{n} p_i c_i \tag{27}$$

and $p_i \geq 0$ $(i = 1,...,n)$ and

$$\sum_{i=1}^{n} p_i = 1 \tag{28}$$

(intuitively: maximize the information content, e.g. of a text or speech, subject to a given effort value E – see Egghe and Lafouge (2006)). The method of the multiplicators of Lagrange correctly gives the solution (in the book under review and in Egghe and Lafouge (2006)): for $r = 1,...,n$

$$p_r = c\rho^{-c_r} \tag{29}$$

with $c > 0$ and $\rho > 0$. Note that the method of the multiplicators of Lagrange only gives a necessary condition for (MEP). But in Egghe and Lafouge (2006), by giving an extra proof we show that condition (29) (i.e. with $\rho > 0$) is necessary and sufficient. This implies that a "principle of minimum uncertainty" (i.e. a "minimum entropy principle") is nonexisting since this would imply the same necessary conditions ! We recognize that, in the book under review, the latter principle has not been formulated in a mathematically exact way but making two sections (9.2 and 9.3) on these principles at least presupposes that they both exist in a mathematically similar formulation, which is not the case !

What is lacking here is the "old" principle of least effort (PLE), well-known in linguistics (admittingly, less known in mathematics – see Egghe and Lafouge (2006)) and attributed to Zipf (1949) (but see also Rousseau (2002)) which states that (27) should be minimal, now subject to a constant value of (26) and (28). In Egghe and Lafouge (2006) it is shown that this principle is equivalent with (29) but for $\rho \geq 1$. This required for a principle that is equivalent with (29) for $0 < \rho < 1$. It was found in Egghe and Lafouge (2006) that the principle of most effort (PME) (introduced there) is the "missing link": here we require (27) to be maximal subject to a constant value of (26) and (28). This principle is shown to be equivalent to (29) for $0 < \rho \leq 1$ (and where (PLE) and (PME) coincide for $\rho = 1$, a degenerate case).

The last two principles in Chapter 9 are also "very philosophical" and deal with the way we can go from one uncertainty theory to another. These principles do not belong to a mathematical theory and, as recognized in the book under review, are underdeveloped. We think we can leave it here in view of the pitfalls of the former (philosophical) "principle" of minimum uncertainty which is, mathematically, non-existing and because of the underdeveloped state of many uncertainty theories (as recognized in this book).

A general conclusion is that this book has the merit to discuss some acceptable extensions of uncertainty theories e.g. to cases of non-additive probability measures and to cases of fuzzy sets but that the book suffers from a non-appropriate mixture of mathematical principles and philosophical principles (often as a substitute for not (yet) understood or even non-existing mathematical principles).

The book has a relatively fair price but it should only be recommended to researchers in this narrow field and certainly not to general researchers in information science (including informetrics researchers) as is the case for the JASIST readership.

REFERENCES

L. Egghe (2005). Power Laws in the Information Production Process: Lotkaian Informetrics. Elsevier, Oxford (UK). ISBN 0-12-088753-3.

L. Egghe and Th. Lafouge (2006). On the relation between the maximum entropy principle and the principle of least effort. Mathematical and Computer Modelling, 43, 1-8.

R.V.L. Hartley (1928). Transmission of information. The Bell System Technical Journal 7(3), 535-563.

H.S. Heaps (1978). Information Retrieval: computational and theoretical Aspects. Academic Press, New York, USA.

F.S. Roberts (1979). Measurement Theory with Applications to Decisionmaking, Utility and the social Sciences. Addison-Wesley, Reading (MA), USA.

R. Rousseau (2002). George Kingsley Zipf: life, ideas, his law and informetrics. Glottometrics 3, 11-18.

C.E. Shannon (1948). The mathematical theory of communication. The Bell System Technical Journal 27(3&4), 379-423.

G.K. Zipf (1949). Human Behavior and the Principle of least Effort. Addison-Wesley, Cambridge, USA. Reprinted: Hafner, New York, USA, 1965.

Leo Egghe

Hasselt University and University of Antwerp