SOME NOTES TO THE ALGEBRAIZATION OF DEMPSTER-SHAFER TREATMENT OF UNCERTAINTY

TOMÁŠ HAVRÁNEK

Questions related with the centering operation used in ordering of uncertainty intervals are discussed. In consequence, the comparative treatment of uncertainties is recommended.

0. INTRODUCTION

Statistics is not a science; it is a scientific method [1]. There are nice theories supporting some statistical methods. These theories to be applied in real situations need a plenty of assumptions to be satisfied. Moreover, for many situations there are more theories applicable. Sometimes two distinct theories give identical or similar results even if, clearly, only one of them could be the "true" one; this is the good case. In other cases, theoretically recommended solutions are distinct.

An applied statistician has to decide which theory is the proper one and whether applicability assumptions are satisfied; the information available is usually too weak to support satisfactorily a decision. But the applied statistician has to reach a decision on methods used and to apply them. Results, in general, are not so bad as one can expect.

What is the reason for this practical performance of statistics? I think a written and/or unwritten experience with data, methods and computations, not the theories. In practical statistics it is necessary to analyse in detail how the methods work on data and what are real differences between these methods in practical and/or particular simple cases as well as how great is the sensitivity of methods with respect to breaking of their assumptions.

It seems that the situation in knowledge or uncertainty processing is similar. For construction of expert systems we have to analyse some things in detail, even on oversimplified cases. From this point of view one has to read the following notes.

1. INTERVAL UNCERTAINITIES AND CENTERING OPERATIONS

Uncertainty concerning a proposition can be expressed as an interval $\langle a, b \rangle$, $0 < a < b < 1$. An algebraic theory for this approach based on the Dempster combining rule was developed in [3]. For ordering of goals in a proposed expert system according to this theory, a representation of intervals using two algebraically definable mappings is used; the first one of them can be considered as a centering mapping, the second as expressing vagueness of an interval. An analysis of these mappings from the interpretational point of view is contained in [4]. Now we turn our attention to the centering operation (mapping): $p/(1 - p)$. Logistic transformation is defined as $\text{lo}(p) = \ln(p/(1 - p))$.

Fact 1. For an uncertainty interval $\langle a, b \rangle$, the Hájek-Valdés center c_H is given by $o(c_H) = b/(1 - a)$.

The odds of $c_H(a, b)$ are equal to the ratio of the upper limit of uncertainty of a proposition to the upper limit of uncertainty of its negation. Clearly, $\text{lo}(c_H) = \ln(b) - \ln(1 - a)$.

Perez [5], [6] considers the barycenter (minimax) principle for integrating "probabilistic" uncertainties to support decisions. Particularly he applies as a divergence measure the Shannon relative entropy. Under this approach the Perez barycenter c_P of $\langle a, b \rangle$ can be defined.

Fact 2. For an uncertainty interval a, b, the barycenter c_P is given by $\text{lo}(c_P) = K/(b - a)$ where $K = b \ln(b/(1 - b)) - a \ln(a/(1 - a)) + \ln((1 - b)/(1 - a))$.

Proof. The barycenter c_P is defined by $H(a, c_P) = H(b, c_P)$, i.e.

$$(*) \qquad a \ln\left(\frac{a}{c_P}\right) + (1 - a) \ln\left(\frac{1 - a}{1 - c_P}\right) = b \ln\left(\frac{b}{c_P}\right) + (1 - b) \ln\left(\frac{1 - b}{1 - c_P}\right)$$

Take K as above; then (*) is equivalent to

$$K = (b - a) \ln c_P - (b - a) \ln(1 - c_P) . \qquad \square$$

If $\mathsf{E}(p) = -(p \ln p + (1 - p) \ln(1 - p))$, then $o(c_P) = \exp((\mathsf{E}(a) - \mathsf{E}(b))/(b - a))$. Hence one has a closed form expression for c_P; in prevailing number of cases barycenter is to be searched by numerical iterative procedures. What seems to be importand that there is, apparently, no divergence D that can be used to define c_H by $D(a, c_H) = D(b, c_H)$. On the other hand; if one uses $D_1(x, y) = |x - y|$, then the condition $D_1(a, c) = D_1(b, c)$ leads to $\bar{c} = \frac{1}{2}(a + b)$; i.e. to the mean uncertainty.

Fact 3. The mean uncertainty $\bar{c}$ can be considered as a one-step approximation to c_P.

Proof (or better, an explication). Using $\ln(x) = 2(x - 1)/(x + 1)$ in $H(a, c) = H(b, c)$ leads to the solution $c_P = \frac{1}{2}(a + b)$. $\square$

This observation is not, in fact, very important here, but it can play a role in more complex cases, where closed form expression for the entropic barycenter is not available. In the light of Fact 3, the following numeric results are not surprising:

a	b	c_P	c	c_H
0·25	0·5	0·372	0·375	0·400
0·25	0·35	0·299	0·300	0·318

What is the relation between c_P and c_H:

Fact 4. Under the logistic transformation we have

$$(**) \qquad \mathrm{lo}\,(c_P) - \mathrm{lo}\,(c_H) = a \ln\left(\frac{a}{b}\right) + (1 - b) \ln\left(\frac{1 - b}{1 - a}\right)$$

Particularly, for $c_H < 0{\cdot}5$, we have in consequence $c_P < c_H$, for $c_H > 0{\cdot}5$, $c_P > c_H$. (Proof. $c_H > 0{\cdot}5$ is equivalent to $b/(1 - a) > 1$; then $a/(1 - b) > 1$ as well and hence the left side of (**) is positive when $\ln b/a > \ln((1 - b)/(1 - a))$ which is equivalent to $b(1 - b) > a(1 - a)$). The relation between c_H and $\bar{c}$ can be expressed similarly:

Fact 5.

$$\mathrm{lo}\,(c_H) - \mathrm{lo}\,(\bar{c}) = \ln\left(1 + \frac{1 - b}{1 - a}\right) - \ln\left(1 + \frac{a}{b}\right).$$

For $c_H < 0{\cdot}5$, $\bar{c} < c_H$, for $c_H > 0{\cdot}5$, $\bar{c} > c_H$ (the sign of the difference depends on $b(1 - b) > a(1 - a)$).

From the numerical point of view, differences between c_P (and $\bar{c}$) on one side and c_H on the other side are not negligible. How it looks from the comparative point of view, does e.g. $c_P(a, b) < c_P(a', b')$ imply $c_H(a, b) < c_H(a', b')$?

2. THE QUESTION OF APPROXIMATION CRITERIA

Let us see the problem of uncertain knowledge treatment from another point of view. Consider that we have at our disposal a set of uncertainties $p_1, \ldots, p_n$ for a proposition. These uncertainties are to be represented, say approximated, by one number c to enable e.g. orddering of propositions (goals). Let now $d(x, y) = |x - y|$ is the usual distance between real numbers. Without any further assumptions we can apply the following three methods for finding c:

(i) $\min\limits_{c} \sum\limits_{i=1}^{n} d^2(p_i, c)$,

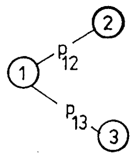
(ii) $\min\limits_{c} \sum\limits_{i=1}^{n} d(p_i, c)$ and

(iii) $\min\limits_{c} \max\limits_{i} d(p_i, c)$.

The method (i) (least squares) gives $\bar{c} = n^{-1} p_i$, the method (ii) (L_1-norm) gives the median c_M of $p_1, \ldots, p_n$ and (iii) (minimax) gives $\hat{c} = (p_{(1)} + p_{(n)})/2$ where $p_{(n)} = \max\{p_1, \ldots, p_n\}$ $p_{(1)} = \min\{p_1, \ldots, p_n\}$. The whole situation can look as follows:

What is the best method here? The minimax method seems apparently to be ruled out, but it can be reasonable, if one assumes that $p_1, \ldots, p_n$ are observations from the uniform distribution between a and b. We can conclude that no method is a panaceum, but that their applicability depends on underlying assumptions – and what we know about them in practice?

Let an expert draw a dependence graph between three propositions as follows:



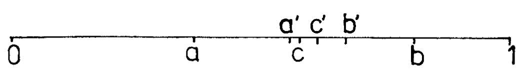
There are two extreme approaches that can be applied here. Under the first one, if there is no edge between 2 and 3, the 2 and 3 are conditionally independent and hence only probability distributions respecting this fact are to be considered. Under the second, all distributions with marginals p_{12}, p_{13} are to be taken into account. But perhaps the above knowledge pattern means that dependence between 2 and 3 is weak (since unobserved)? In general I hesitate whether it is appropriate to compare decision methods with respect to the worst possible case (or with respect to *one* case).

3. ORDERINGS OF INTERVALS

If an evaluation of goals in terms of uncertainties is obtained, then it is necessary to order them with respect to the corresponding "degrees of belief". This is the fundamental paradigma concerning the presentation of results to users usually used in constructing expert systems, particularly in the proposals of Hájek and Valdés [3]. If $c(a, b)$ is a centering mapping and $v(a, b)$ the vagueness mapping, then the lexicographic ordering of intervals is proposed:

$$\langle a, b\rangle <_1 \langle a', b'\rangle \quad \text{if} \quad c(a, b) < c(a', b') \quad \text{or} \quad (c(a, b) = c(a', b') \\ \text{and} \quad v(a, b) > v(a', b'))\,.$$

The underlying idea is that the resulting order must be linear, e.g. each two goals have to be comparable (let us omit here the theoretical algebraic reasons). As a result we obtain that $\langle a, b\rangle <_l \langle a', b'\rangle$ (or goal $G <_l$ goal G') even in case when



If the only knowledge is that the "probability" of G is somewhere in $\langle a, b\rangle$ and the "probability" of G' is somewhere in $\langle a', b'\rangle$, then the decision that $G <_1 G'$ seems to be, without some additional assumptions, rather risky. The possibility of an error can have the chance close to 1. By interval uncertainties only a partial ordering can be reliably generated:

$$G \leqq_p G' \quad \text{if} \quad a \leqq a' \quad \text{and} \quad b \leqq b' .$$

For each pair of intervals $\langle a, b \rangle$ and $\langle a', b' \rangle$ there is an infimum $\langle a, b \rangle \wedge \langle a', b' \rangle = \langle \min(a, a'), \min(b, b') \rangle$ and a supremum $\langle a, b \rangle \vee \langle a', b' \rangle = \langle \max(a, a'), \max(b, b') \rangle$; a lattice is formed. Clearly, $G <_p G'$ implies $G <_1 G'$. Trivially, if $\langle a, b \rangle \leqq_p \langle a', b' \rangle$ then $c_H(a, b) \leqq c_H(a, b')$ and $\bar{c}(a, b) \leqq \bar{c}(a', b')$: slightly less trivially $c_P(a, b) \leqq c_P(a', b')$. From the naive point of view, $\wedge$ is a candidate to be used in the contribution function for evaluating of rules (if $\langle a, b \rangle$ is the uncertainty of the antecedent, $\langle v, w \rangle$ the weight of the rule, then the uncertainty of succedent can be defined by $\langle a, b \rangle \wedge \langle v, w \rangle$ etc.).

What is substantial here, is that a result of *numerical* manipulation with uncertainties only a (natural) *partial ordering* of goals is obtained.

4. CONCLUSIONS

Observations similar to the facts presented in the paper seem to lead to a conclusion that insisting on numerical uncertainties (probabilities) is not the proper way for modelling expert decision making and, if we agree that expert systems have to model expert decision making, for construction of expert systems. It is necessary to investigate nonnumeric comparative probabilities or uncertainties, but moreover without usual assumption (axiom) of comparability of each two propositions. Similarly as in statistical decisions (cf. [2]) we must learn that some choice between possibilities has to be left to the users and their experience.

REFERENCES

[1] F. E. Croxton and P. J. Cowden: Applied General Statistics. Prentice-Hall, New York 1946.

[2] D. Edwards and T. Havránek: A fast model selection procedure for large families of models. J. Amer. Statist. Assoc. *82* (1987), 205—213.

[3] P. Hájek and J. Valdés: Algebraic Foundations of Uncertainty Processing in Rule-Based Expert Systems. Res. Rep. 28/1987, Mathematical Institute, Czechoslovak Academy of Science, Prague 1987.

[4] T. Havránek: An interpretation of Hájek-Valdés results on the Dempster semigroup. In: AI '88 (V. Mařík, ed.), Ústav pro informační systémy v kultuře, Prague 1988, pp. 69—76.

[5] A. Perez: Probability approach in integration partial knowledge for medical decision making (in Czech). In: Souhrn přednášek 1, I. celostátní sjezd biomedicinského inženýrství BMI 83, 83, Mariánské Lázně 1983, pp. 221—226.

[6] A. Perez: "Barycenter" of a set of probability measures and its application in statistical decisions. In: COMPSTAT 84 (T. Havránek, Z. Šidák, M. Novák, eds.), Physical-Verlag, Wien 1984, pp. 154—159.

RNDr. Tomáš Havránek, CSc., Středisko výpočetní techniky ČSAV (Institute of Computer Science — Czechoslovak Academy of Sciences), Pod vodárenskou věží 2, 182 07 Praha 8. Czechoslovakia.