Second Order Measures for Uncertainty Processing

Zdenek Zdrahal
Human Cognition Research Laboratory
The Open University
Walton Hall, Milton Keynes, MK7 6AA
United Kingdom
e-mail: Z.Zdrahal@open.ac.uk

Abstract

Uncertainty processing methods are analysed from the viewpoint of their sensitivity to small variations of certainty factors. The analysis makes use of the algebraic theory which defines the function for combining partial certainty factors by means of a group operation of the ordered Abelian group over the interval of uncertainty. Two approaches are introduced: (a) sensitivity analysis of the inference network and (b) calculation of second order probabilities. Sensitivity functions are defined as partial derivatives of the combining function with respect to their arguments. Based on the sensitivity functions, we define the path sensitivity which measures the sensitivity of a larger part of the inference network. If a set of samples of certainty factors is available instead of a single value, the second order probability distribution can be approximated by the distribution of an average value. It is shown that the parametric form of the distribution is completely determined by the combining function.

1 Introduction

Numerical values describing the uncertainty of knowledge and data in knowledge-based (KB) systems are usually imprecise due to the fact that they are almost always provided as the subjective assessments of experts or users. Nonetheless, these imprecise numbers are processed by some algorithm and the results are used to draw conclusions. Without a thorough KB verification which includes an analysis of the robustness of the uncertainty processing technique used, we must always be aware of die limited credibility of results. This paper aims to provide techniques for such an analysis. The methods described are based on compositional (extensional) calculation of uncertainty processing [Duda et id., 1976; Gashnig, 1980; Heckerman, 1986; Reiter, 1980; Wise, 1988] (see [Hajek et al., 1992; Pearl, 1988] for more detailed discussion). Although the current attention of the AI community is focused rather on intensional (model-based, probabilistic) approaches [Spiegelhalter, 1986; Lauritzen and Spiegelhalter, 1988; Pearl, 1988], the compositional methods are still popular due to their computational simplicity. The main objection to the compositional methods is that the results are not sound. In [Hajek et al., 1992] an attempt is made to revive these methods by replacing the original simple-minded interpretation of their results by a comparative one, thus improving their robustness as well as their soundness.

We will present two methods for assessing the imprecision of uncertainty measures in rule-based KB systems. The first approach is based on sensitivity evaluation. The idea of a *sensitivity analysis* of inference nets was explored in Prospector [Gashnig, 1980], where a uranium model was compiled and run for a large number of combinations of data and the sensitivity was calculated. Our approach is more analytical. We define sensitivity functions for particular methods of combining certainty factors and then in terms of these functions and rule sensitivities, we analyse the sensitivity of a path in the inference network. The second method proposed in this paper is based on the idea of *second order uncertainties,* i.e. the uncertainties of certainty factors. The concept of second order probabilities has already been suggested by [Cheeseman, 1985]. We will show that for certain statistics the parametric form of the second order probability density is completely determined by the method used for combining certainty factors. This property makes it possible to calculate the actual second order density functions. Moreover, the parametric form of this density function is invariant with respect to the combining function used.

2 Preliminaries

Regardless of their origin, the uncertainty measures will be called *certainty factors* throughout this paper. Our approach is based on the algebraic theory of uncertainty processing developed by [Hajek et al., 1992]. We will briefly summarise the relevant parts of Hajek's theory needed for the presentation our work (for the complete theory see [Hajek et al., 1992]). It is assumed that knowledge is expressed in terms of rules. A numerical certainty factor (weight) w from (-1,1) is associated with each rule, $E \rightarrow H(w)$. The extreme certainty factors correspond to "Hypothesis H is false" (value -1) and "Hypothesis H is true" (value 1) respectively. Certainty factor 0 (zero) stands for "There is no evidence concerning hypothesis H' If two or more rules bear on the same hypothesis, the overall certainty factor of the hypothesis is calculated by applying some combining function to individual contributions. We will call the result produced by the combining function "the global certainty factor", and the contributions, i.e. the arguments of the combining function, "partial" certainty factors. The combining function is defined by means of a binary operation ® for which the following axioms hold:

Let x, y and z be partial certainty factors of a hypothesis.

1. If $x = 1$ and $y \neq -1$, or $y = 1$ and $x \neq -1$ then $x \oplus y = 1$.
   If $x = -1$ and $y \neq 1$, or $y = -1$ and $x \neq 1$ then $x \oplus y = -1$.
2. If $x = 1$ and $y = -1$, or $y = 1$ and $x = -1$ then $x \oplus y$ is not defined.
3. For $x, y, z \neq -1$ and $x, y, z \neq 1$
   (a) $x \oplus y = y \oplus x$ ... commutativity
   (b) $(x \oplus y) \oplus z = x \oplus (y \oplus z)$ ... associativity
   (c) $x \oplus 0 = x$ ... null element
   (d) $x \oplus -x = 0$ ... inverse element
   (e) If $x \leq y$ then $x \oplus z \leq y \oplus z$ ... ordering

These axioms describe the properties we intuitively expect from combining operations. The axioms 3 (a) – (e) define an ordered Abelian group (OAG) over (–1, 1) with the group operation $\oplus$ so far unspecified. This OAG is isomorphic with the additive OAG over $(-\infty, \infty)$; there are various isomorphisms which map $-1 \rightarrow -\infty$, $0 \rightarrow 0$, $1 \rightarrow \infty$, the group operation $\oplus$ being the normal addition and the inverse element being the number with the opposite sign. Similarly, the OAG over (-1, 1) is isomorphic with the multiplicative OAG over $\langle 0, \infty)$, i.e. $-1 \rightarrow 0$, $0 \rightarrow 1$, $1 \rightarrow \infty$, the group operation being the multiplication and the inverse element of x being $1/x$. Particular combining operations $\oplus$ over (–1, 1) can be defined by means of these isomorphisms. For example, if we use the additive OAG the partial weights x and y to be combined are mapped by some isomorphism from (–1, 1) to $(-\infty, \infty)$, then the isomorphic images are summed since addition is the group operation on $(-\infty, \infty)$, and finally the result is mapped back by the inverse isomorphism to (–1, 1). The group operation defines a combining function $g(x,y)$:

$$g(x,y) = x \oplus y = F^{-1}[F(x) + F(y)] \qquad (1)$$

There are several isomorphisms (–1, 1) to $(-\infty, \infty)$ suggested by [Hajek *et al.*, 1992]. We will mention only two of them: those used in Emycin and Prospector. They are defined for $\langle 0, 1) \rightarrow \langle 0, \infty)$ by the following formulae, for negative values we take an odd extension, $F(-x) = -F(x)$.

$$F(x) = \ln [1/(1-x)] \quad \text{... Emycin} \qquad (2)$$

$$F(x) = \ln [(1+x)/(1-x)] \quad \text{... Prospector} \qquad (3)$$

Substituting (2) to (1) gives Emycin formulae. Similarly, substituting (3) to (1) gives $(x + y)/(1 + xy)$ which is the Prospector odds-based formula recalculated into (–1, 1).

The isomorphisms between Prospector and Emycin were used in [Heckerman, 1986] to interpret MYCIN's certainty factors. Similarly, the AL/X uncertainty processing method [Reiter, 1980] is just an isomorphic image of the Prospector's. Graphs of combining functions defined by (2) and (3) are shown in Fig. 1 (a) – (b) for $y = -0.9, -0.6, -0.3, 0.0, 0.3, 0.6$ and 0.9.

In order to demonstrate that our intuition expressed by axioms 3 (a) – (e) is not sufficient to guarantee sensitivity properties required for processing of uncertain knowledge we have defined the combining function as follows:

$$F(x) = [x/(1-x)]^4 \qquad (4)$$

We will show this function to be a typical counter-example which does not meet acceptable sensitivity properties. Its graph is shown in Fig. 1 (c).

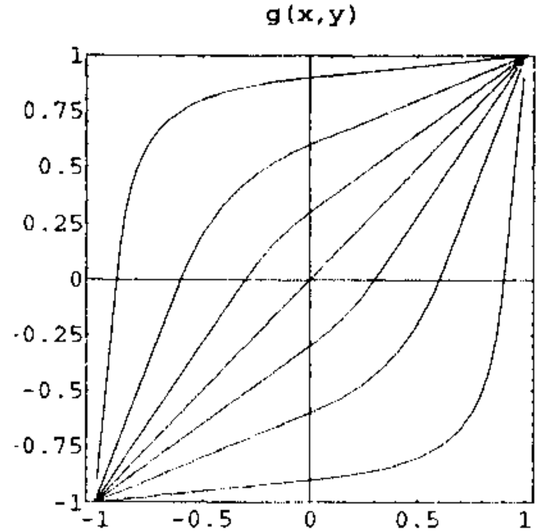

Fig. 1 (a) Combining functions - Emycin

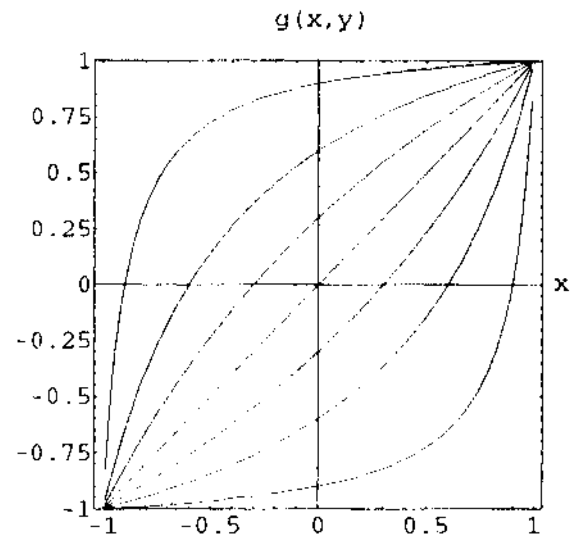

Fig. 1 (b) Combining functions - Prospector

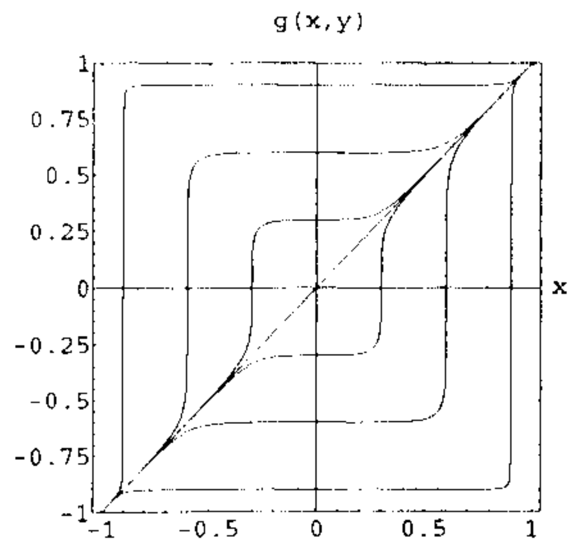

Fig. 1 (c) Combining functions - counter-example

3 Sensitivity functions

For applications, the exact values of certainty factors must not be crucial. Uncertainty processing methods must provide correct results regardless of small variations in numerical values. From this point of view we expect the processing methods to be insensitive to small changes. On the other hand the method must "weigh" the contributions; the global certainty factor must depend on the values of partial ones which means that it must not be too insensitive. Given a standard rule-based architecture it is reasonable to assume that no knowledge is represented in terms of the combining method, i.e. we will consider the same, *a priori* determined, combining function throughout the inference network. The sensitivity of the final hypothesis depends on the sensitivities of rules and the sensitivities of combining algorithms. The behaviour of the combining function for small variations of one variable is described by the first partial derivative with respect to this variable.

Definition 1

The sensitivity function $s_x(x,y)$ of a combining function $g(x,y)$ widi respect to x is

$$s_x = \frac{\partial g(x,y)}{\partial x}$$

Similarly we define the sensitivity function with respect to y as $s_y(x,y) = \partial g(x,y)/\partial y$. Since the operation $\oplus$ is associative, we can extend the concept of binary combining functions to n-ary ones as follows:

$$g(x_1, x_2 \dots x_n) = F^{-1}[F(x_1) + F(x_2) + \dots + F(x_n)] \qquad (5)$$

The concept of a sensitivity function can also be extended

$$s_{x_i} = \frac{\partial g(x_1, x_2, \dots, x_n)}{\partial x_i} \qquad (6)$$

The sensitivity functions have the following properties:

Proposition 1

For a and $b \in \langle -1,1 \rangle$, $s_X(a,b) = s_Y(b,a)$, i.e. the sensitivity functions are symmetric.

Proposition 2

$s_X(a,b) = s_X(-a,-b)$, i.e. the sensitivity functions are even functions of their arguments.

Proposition 3

$s_X(a,b) \geq 0$, i.e. the sensitivity functions are non-negative.

The following two propositions show that n-ary sensitivity functions can be composed of binary ones.

Proposition 4

$$s_{x_i}(x_1, x_2, \dots, x_n) = s_{x_i}(x_i, g(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n))$$

Proposition 5

$$s_{x_i}(x_1, x_2, \dots, x_n) = s_{u_1}(u_1, t_n) \cdot s_{u_2}(u_1, t_{n-1}) \cdots s_{u_{n-1}}(u_{n-1}, t_n)$$

where $t_1 = x_i$, $t_j = x_{j-1}$ for $i \geq j$, $t_j = x_j$ for $i < j$, and for u_k we substitute $u_k = g(t_1, \dots, t_{n-k})$ for $k = 1, \dots, n-2$ and $u_{n-1} = t_1$.

The sensitivity functions corresponding to (2), (3) and (4) are shown in Fig. 2 for y = 0, 0.3, 0.6 and 0.9. The completion for $y < 0$ describes Proposition 2.

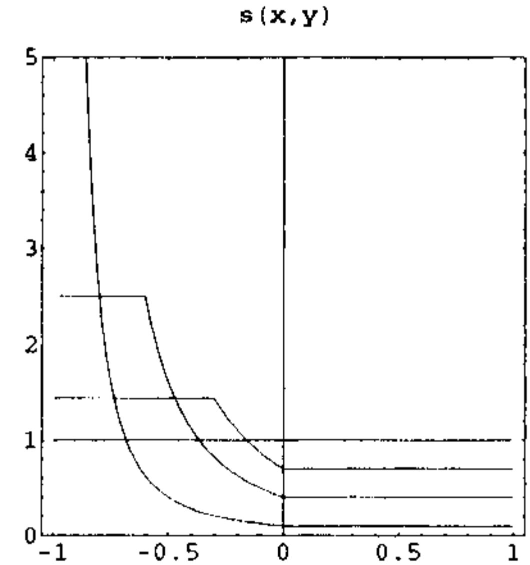

Fig. 2 (a) Sensitivity functions – Emycin

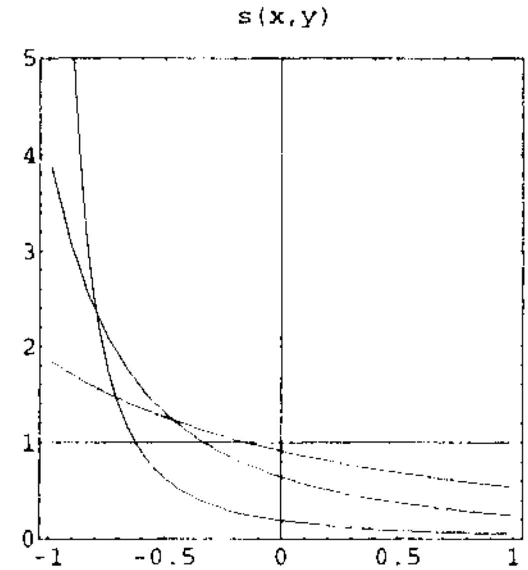

Fig. 2 (a) Sensitivity functions - Prospector

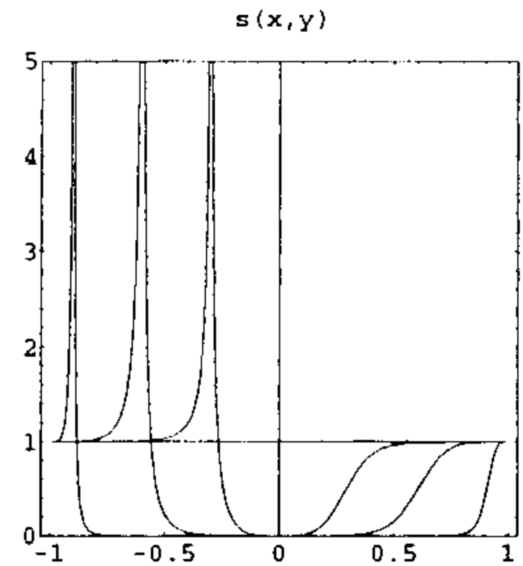

Fig. 2 (c) Sensitivity functions – counter-example

For proofs of all the propositions introduced in this paper see [Zdrahal, 1991a].

What kind of sensitivity function do we prefer? Whereas in control theory the minimum sensitivity with respect to parameter variation is a clear goal, here it is more difficult to define an ideal sensitivity behaviour since we are interested in an *input sensitivity*. Therefore the answer necessarily is rather speculative.

If both arguments are of the same sign, the sensitivity should be approximately constant. The combining function weighs partial certainty factors. As both arguments approach the same extreme value (+1 or –1) the sensitivity function should go to zero. We are almost certain that the hypothesis holds (for +1) or does not hold (for –1) and small variations do not have any impact. The sensitivity increases as the partial certainty factors are approximately of the same absolute value but with different signs. The partial certainty factors compensate each other and the result is close to zero.

The combining functions of both Emycin and Prospector satisfy our intuitive insight. The required weighing effect prevails. On the contrary, function (4) almost does not weigh at all. Instead, it has a switching character. For some $y_0 > 0$ and $-y_0 < x < y_0$ the impact of x is negligible (see Fig. 2 (c)). The value of x becomes important only near $-y_0 = x$. At this point the combining function completely switches to the global certainty factor of the opposite sign. The narrow band around $-y_0 = x$ is overly sensitive and unstable. In this case $s_X(x,y)$ has even got a singularity point at $-y_0 = x$.

Out of interest we have also evaluated the conditional probability $P(H \mid E_1, E_2)$ calculated from $P(H \mid E_1)$ and $P(H \mid E_2)$ by means of the maximum entropy criterion [Zdrahal, 1991b]. Both hypothesis H and evidences E_1 and E_2 were binary valued, probabilities were recalculated into $\langle -1,1 \rangle$ and the values were expressed in terms of a "combining function". We are fully aware that this function is based on a completely different (intensional) philosophy: it has nothing to do with axioms 3 (a) – (e), cannot be expanded by (5) and thus it is impossible to carry out a fair comparison. However the corresponding sensitivity function proved to be very similar to that of Emycin and Prospector (Fig. 2 (a) and (b)) and very different from that of Fig. 2 (c).

From a sensitivity point of view, the formulae (2) and (3) on the one hand and (4) on the other hand produce very different types of combining functions. The properties, however, become more obvious if we take into account the fact that all of them must obey the following proposition.

Proposition 6

The average value of sensitivity over interval $\langle -1,1 \rangle$ is constant and equals 1, i.e.

$$\tfrac{1}{2}\int_{-1}^{1} s_x(x,y)\,dx = 1$$

This proposition in combination with Proposition 3 implies that an excessive sensitivity of the combining function on some subinterval must be compensated by insensitivity in some other subinterval. This effect can be observed in Fig. 2. In cases (a) and (b) there is an even sensitivity on the major part of the certainty interval. On the contrary, in case (c) there is extreme sensitivity around the switching point which is necessarily compensated by almost zero sensitivity in a large subinterval of (–1, 1).

4 Path sensitivity in the inference network

For the purpose of sensitivity analysis the KB is represented as a directed acyclic graph (DAG). We will suppose the DAG to be a tree. The nodes correspond to propositions and the edges to rules. The combining function (5) calculates the global node certainty factor given partial ones. The partial certainty factor corresponding to a single rule combines the certainty factor of the evidence with the weight of the rule. It is usually calculated as their product for a positive certainty factor of the evidence and zero for the negative one. Thus, given rule R: $A \rightarrow B(w_R)$ and certainty factors a of A and b of B, we write

$$b = w_R \cdot a \quad \text{for } a > 0 \qquad b = 0 \quad \text{otherwise.} \tag{7}$$

In terms of the DAG, equation (7) describes the edge propagation. The same or a similar formula is used in [Duda et al., 1976; Hajek et al., 1992; Heckerman, 1986; Reiter, 1980] . The sensitivity of the rule is $\partial b / \partial a = w_R$.

Sensitivity of a larger part of the inference net can be expressed in terms of the sensitivity of edge propagations and sensitivity functions. Sensitivity along a path in the inference net is defined as follows.

Definition 2

Let $h(x_1,\ldots,x_n)$ be the global certainty factor of a root proposition H, where $x_1,\ldots,x_n$ are certainty factors of all leaves $X_1,\ldots,X_n$ of the tree with root H. Path sensitivity S_{h,x_i} of h with respect to x_i is defined as

$$S_{h,x_i} = \frac{\partial h(x_1, x_2, \ldots, x_n)}{\partial x_i}$$

Similarly we can define the path sensitivity of subtrees in the KB. The path sensitivity can be calculated recursively in accordance with the following proposition.

Proposition 7

Let us denote X_i, X_j nodes of an inference network. Let path p of length $l(p) \geq 0$ exist from X_i to X_j. If $l(p) > 0$ let X_k be the immediate predecessor of X_j, which is on the path p.

Let us denote: w_{kj} ... weight of the rule $X_k \rightarrow X_j$,

y_k ... partial certainty factor corresponding to the rule $X_k \rightarrow X_j$, $y_k = w_{kj} \cdot x_j$,

$s_{y_k}(y_1,\ldots,y_k,\ldots)$, sensitivity function of x_j with respect to y_k.

Path sensitivity S_{x_j,x_i} equals:

(a) 1 if $X_i = X_j$, i.e. if $l(p) = 0$.

(b) $w_{kj} \cdot s_{y_k}(y_1,\ldots,y_k,\ldots) \cdot S_{x_k,x_i}$ if $x_k \geq 0$, and

(c) 0 if $x_k < 0$.

Proposition 7 describes a recursive algorithm for calculation of path sensitivities which is evaluated as a function of certainty factors of some nodes (possibly leaves) of the inference network. This function is considerably non-linear due to the non-linearity of both sensitivity functions and equation (7) which complicates its calculation. However, in some cases the exact values of path sensitivity are not so important. Rather the aim of the analysis is to find extreme cases. We are usually interested in an upper bound of the path sensitivity. This can be estimated if all sensitivity functions are substituted by their upper estimates. The upper estimates of the sensitivity functions are not calculated over the whole interval $\langle -1,1 \rangle$, since the range of values of the partial certainty factors which are the arguments of sensitivity functions get systematically restricted by repeated application of (7) for each rule.

The following simple example demonstrates the calculation of the path sensitivity: assume that the inference network consists of four rules $X \to V(w_x)$, $Y \to V(w_y)$, $V \to H(w_v)$ and $Z \to H(w_z)$, i.e. H is the root, X, Y and Z are the leaves and V is the intermediate node. Let us denote as $p = x \cdot w_x$ the contribution of rule $X \to V(w_x)$ to the value of V, similarly $q = y \cdot w_y$ $v = p \oplus q = g(p,q)$, $r = v \cdot w_v$, $t = z \cdot w_z$ and $h = g(r,t)$. Assuming that the values x, w_x, y, w_y, z and w_z are positive (in order to enable the complete edge propagation) we can calculate for example the path sensitivity S_{kx} as $S_{kx} = s_r(r,t) \cdot w_v \cdot s_p(p,q) \cdot w_x$, where $r = w_v \cdot g(w_x \cdot x, w_y \cdot y)$. With respect to the possible values of p and q we evaluate $s_p(p,q)$ only over $p \in \langle 0, w_x \rangle$ and $q \in \langle 0, w_y \rangle$. Similar restrictions apply to the other variables.

5 Second order probabilities

The second way of characterising the imprecision of certainty factors is by means of second order probabilities. We are going to use the term probability although the first order uncertainty is expressed in terms of general, not necessarily probability-based certainty factors. While the sensitivity analysis estimates the potential errors produced by variations of certainty factors the second order probabilities estimate the chances that the error really occurs. They express our doubts concerning the given value of certainty factors.

Second order probability density $p(x,m)$ has a meaning for $x, m \in$ (-1, 1). The parameter m is a value of the certainty factor into which the density function is located. Intuitively, we expect $p(x,m)$ to peak at both ends of interval (-1, 1), i.e. for m close to -1 and 1 respectively, since if we are certain about some proposition we do not allow any doubts. In the middle of the certainty factor interval $p(x,m)$ is flattest as the lack of evidence allows maximal doubts. We can however hardly assume that the second order densities are *a priori* known. Therefore we will attempt to replace them by some statistic estimated from potentially available data.

Let us assume that instead of a single user providing certainty factors there is a user poll which gives n independent answers to each query. Therefore when investigating some evidence instead of a single certainty factor x we get a set $\chi = \{x_1, \dots, x_n\}$ of certainty factors. This set characterises the distribution $p(x,m)$ of x. Instead of calculating n independent inferences and then evaluating n results we characterise the set χ by a single value, a statistic of χ. Let us use the sample average value. We cannot simply take the arithmetic average since the average must take into account the properties of uncertainty processing group. We will use the group average which is defined as follows:

Definition 3
Group average $\bar{x}$ is a number for which the equation $\bar{x} \oplus \bar{x} \oplus \dots \bar{x} = x_1 \oplus x_2 \oplus \dots x_n$ holds.

In the additive group on $(-\infty, \infty)$ the group average is the well-known arithmetic average, similarly in the multiplicative group on $(0, \infty)$ the group average is the geometric average. In the certainty factor group on $(-1, 1)$ we must calculate the group average in three steps: (i) mapping certainty factors by means of the isomorphism from $(-1, 1)$ to $(-\infty, \infty)$, (ii) calculating the arithmetic average in the additive group and (iii) transforming the result by means of the inverse isomorphism back to $(-1, 1)$, i.e.

$$\bar{x} = F^{-1}[\frac{1}{n}\sum_{i=1}^{n} F(x_i) = F(\bar{X}), \text{ where } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} F(x_i) \qquad (8)$$

In accordance with the assumptions χ is a set of independent answers, therefore also $\{F(x_i)\}$ is a set of independent samples. Let us assume that there exists a mean value μ and a dispersion σ^2 of distribution of $F(x)$. According to the Central Limit Theorem the term

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} F(x_i) \qquad (9)$$

has the normal distribution $N(\mu, \sigma / \sqrt{n})$ From (8) it follows that the average value $\bar{x}$ has, up to the mean and dispersion of $\bar{X}$, completely determined its density function by the isomorphism F. We denote the density function of the group average as $q(\bar{x}, \mu)$ and summarise the result in the following proposition.

Proposition 8
For a given set $\chi = \{x_1, \dots, x_n\}$ the group average has a distribution with probability density

$$q(\bar{x}, \mu) = F^{-1}[N(\mu, \sigma / \sqrt{n}]$$

i.e. the parametric form of the second order density $q(\bar{x}, \mu)\mu)$ is determined by the isomorphism F.

Given a set χ and the parametric form of the density, the actual values of μ and σ can be estimated by any standard parametric technique. For the group averages the following propositions hold (for proofs see [Zdrahal, 1991a]).

Proposition 9
Let $\{x_i\}$ and $\{y_i\}$, $i = 1, 2, \dots, n$ be sets of certainty values. Let $\bar{x}$ and $\bar{y}$ denote the group averages of $\{x_i\}$ and $\{y_i\}$ respectively. Then $\overline{g(x_i, y_i)} = g(\bar{x}, \bar{y})$.

Proposition 10
Let $\{x_i\}$, $i = 1, 2, \dots, n$ and $\{y_j\}$, $j = 1, 2, \dots, m$, be sets of independent certainty factors, let $\bar{x}$ and $\bar{y}$ denote the group averages of $\{x_i\}$ and $\{y_i\}$ respectively. Then the parametric form of the second order density of $g(\bar{x}, \bar{y})$ is the same as that of $\bar{x}$ and $\bar{y}$ respectively.

Proposition 9 states that the combining function preserves sample averages. In accordance with the Proposition 10 the combining function preserves the parametric form of the second order density of group averages. However the calculation of the second order density for group averages across all inference network cannot be carried out automatically by recursively repeating results of Proposition 10 since there is still the non-linear edge propagation described by (7). It is necessary to split the inference network into simple parts and cases and analyse them individually.

Having obtained the second order density we can calculate average values of various characteristics which depend on values of certainty factors. As an example we can combine both measures introduced in this paper and calculate an average sensitivity function. It will be defined as

$$\bar{s}_m(m, y) = \int_{-1}^{1} s_x(x, y) \cdot q(x, m) dx$$

The averaging is a kind of smoothing procedure. If the dispersion a is very small, i.e. our knowledge of certainly factors is very certain, the second order density is a very high and narrow peak which takes a very local sample of the averaged function (of the sensitivity function in the case above). For a large c the average sensitivity function is very smooth. The doubts concerning the correct value of the certainty factor helps to solve the sensitivity problem. Similarly, the second order density can be used to average other useful characteristics.

6 Conclusions

We have presented two different tools for the analysis of uncertainty processing methods in rule-based systems and shown some of their properties. The first method - sensitivity analysis - is focussed on properties of the knowledge base with uncertainty. Sensitivity functions evaluate the sensitivity of the combining formula while the path sensitivity makes it possible to assess the sensitivity of the inference network as a whole. The second method - second order probability - is concerned with the impact of uncertainty values from outside the knowledge base, i.e. from the user. Both techniques, which are intended mainly as an off-line analysis, can be used independently or in combination. They allow a deeper insight into properties of inference networks which is important from an application point of view.

Acknowledgements

Thanks to Renato Barrera, Marc Eisenstadt, Radim Jirousek, John Reiter, Arthur Stutt, Stuart Watt and two of the three anonymous referees for useful comments on earlier versions of the paper.

References

[Cheeseman, 1985] Peter Cheeseman. In defence of probability. In Proc. 9th IJCAI '85, pages 1002-1009, Los Angeles, 1985. International Joint Committee on Artificial Intelligence.

[Duda et ai, 1976] R.O. Duda, P.E. Hart and NJ. Nilsson. Subjective Bayesian methods for rule-based inference systems. Technical note 134, SRI International, 1976.

[Gashnig, 1980] John Gashnig. Development of Uranium Exploration Models for Prospector Consultant Systems. Final Report SRI Project 7856, SRI International, March 1980.

[Hajek et ai, 1992] Petr Hajek, Tomas Havranek and Radim Jirousek. Uncertain Information Processing in Expert Systems. Boca Raton, CRC Press, 1992.

[Heckcrman, 1986] D. lleckerman. Probabilistic interpretations for MYCIN'S certainty factors. In Uncertainty in Artificial Intelligence, (L.N. Kanal & J.F. Lemmer eds.), pp. 167-196, 1986. The Netherlands: North-Holland.

[Lauritzen and Spiegelhalter, 1988] S.L Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graph structures and their applications to expert systems. Journal of the Royal Statist. Soc. B., Vol. 50, No. 2, pp. 157-224, 1988.

[Pearl, 1988] Judea Pearl. Probabilistic reasoning in intelligent systems: Network of Plausible Inference. San Mateo, California: Morgan Kaufmann, 1988.

[Reiter, 1980] John Reiter. AL/X: An Expert System Using Plausible Inference. Oxford: Intelligent Terminals Ltd., 1980.

[Spiegelhalter, 1986] D.L Spiegelhalter. Probabilistic reasoning in predictive expert systems. In Uncertainty in Artificial Intelligence, (L.N. Kanal & J.F. Lemmer eds.), pp. 47-67, The Netherlands: North-Holland, 1986.

[Wise, 1988] B.P. Wise. Experimentally comparing uncertain inference systems to probability. In Uncertainty in Artificial Intelligence, (J.F. Lemmer & L.N. Kanal eds.), 1988, pp. 89-99, The Netherlands: North-Holland.

[Zdrahal, 1991a] Z. Zdrahal. Sensitivity Analysis of Uncertainty Processing in Rule-based Systems. Technical Report, CICESE / Computer Science, Ensenada, Baja Calif., 1991.

[Zdrahal, 1991b] Z. Zdrahal. Sensitivity models of inference networks. 8th Mexican National Conference on Artificial Intelligence, Ensenada, Baja Calif., August 1991.