Uncertain inference using interval probability theory

James W. Hall, David I. Blockley *, John P. Davis

Department of Civil Engineering, University of Bristol, Bristol, UK

Received 1 April 1997; received in revised form 1 December 1997; accepted 1 March 1998

Abstract

The use of interval probability theory (IPT) for uncertain inference is demonstrated. The general inference rule adopted is the theorem of total probability. This enables information on the relevance of the elements of the power set of evidence to be combined with the measures of the support for and dependence between each item of evidence. The approach recognises the importance of the structure of inference problems and yet is an open world theory in which the domain need not be completely specified in order to obtain meaningful inferences. IPT is used to manipulate conflicting evidence and to merge evidence on the dependability of a process with the data handled by that process. Uncertain inference using IPT is compared with Bayesian inference. © 1998 Elsevier Science Inc. All rights reserved.

Keywords: Interval probability theory; Theory of evidence; Inference networks; Process modelling; Bayesian inference

1. Introduction

Cui and Blockley [1] introduced interval probability theory (IPT) as a measure of evidential support in knowledge-based systems. Interval numbers are use to represent the probability measure in order to capture in a relatively simple manner, features of fuzziness and incompleteness. The idea of interval representation has attracted numerous researchers [2-4]. Cui and Blockley [1]...
developed previous work by introducing the parameter $\rho$ which represents the degree of dependence between evidence. Inference rules based on assumptions of dependence or independence are therefore special cases of IPT. The theory has since been used to model complex processes in the fields of earthquake engineering [5] and petroleum exploration [6]. The purpose of this paper is to describe theoretical developments which have been inspired by the experience of applying IPT in practice.

The methods discussed in this paper are suitable for complex inferences with sparse data and incomplete and possibly inconsistent knowledge. The intention is to provide decision-makers with information in a simple format which at the same time reflects the complexity of the inference problem and the richness of the available evidence. In practical decision-making it may be necessary to make use of very different types of uncertain information, from countable items of data to vague beliefs of domain experts. The theoretical background to the problem of merging different types of data is discussed in this paper and a solution based on the theorem of total probability is described.

2. A review of interval probability theory

IPT is founded on the axioms of probability theory but allows support for a conjecture to be separated from support for the negation of the conjecture. If $E$ is a proposition, an interval number is used as a probability measure, so that

$$P(E) = [S_n(E), S_p(E)],$$

where $S_n(E)$ is the lower bound and $S_p(E)$ is the upper bound of the probability $P(E)$. The negation is

$$P(\bar{E}) = [1 - S_p(E), 1 - S_n(E)].$$

An interval probability can be interpreted as a measure of belief, so that $S_n(E)$ represents the extent to which it is certainly believed that $E$ is true or dependable, $1 - S_p(E) = S_n(\bar{E})$ represents the extent to which it is certainly believed that $E$ is false or not dependable, and the value $S_p(E) - S_n(E)$ represents the extent of uncertainty of belief in the truth or dependability of $E$. Three extreme cases illustrate the meaning of this interval measure of belief.

- $P(E) = [0, 0]$ represents a belief that $E$ is certainly false or not dependable.
- $P(E) = [1, 1]$ represents a belief that $E$ is certainly true or dependable.
- $P(E) = [0, 1]$ represents a belief that $E$ is unknown.

The degree of dependence between two propositions $E_1$ and $E_2$ is defined by the parameter $\rho$

$$\rho = \frac{P(E_1 \cap E_2)}{\text{Min}(P(E_1), P(E_2))}.$$
Thus $\rho = 1$ indicates that $E_1 \subset E_2$ or $E_2 \subset E_1$, whilst if $E_1$ and $E_2$ are independent

$$\rho = \max(P(E_1), P(E_2))$$

so that

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2).$$

The minimum value of $\rho$ is given by

$$\rho = \max\left[\frac{P(E_1) + P(E_2) - 1}{\min(P(E_1), P(E_2))}, 0\right],$$

where $\rho = 0$ indicates that $E_1$ and $E_2$ are mutually exclusive.

If $\rho$ is defined as an interval number $[\rho_l, \rho_u]$ then

\begin{align*}
S_n(E_1 \cap E_2) &= \rho_l(S_n(E_1) \land S_n(E_2)), \\
S_p(E_1 \cap E_2) &= \rho_u(S_p(E_1) \land S_p(E_2)), \\
S_n(E_1 \cup E_2) &= S_n(E_1) + S_n(E_2) - \rho_l(S_n(E_1) \land S_n(E_2)), \\
S_p(E_1 \cup E_2) &= S_p(E_1) + S_p(E_2) - \rho_u(S_p(E_1) \land S_p(E_2)),
\end{align*}

where $\land, \lor$ refer to min and max, respectively.

The dependence parameter $\rho$ is a convenient means of exploring different dependence relationships between evidence when the exact nature of dependence is uncertain. The dependence parameter generalises other inference rules which assume a specific dependence relationship between evidence. The conventional (i.e. max and min) definitions of fuzzy union and intersection correspond to the special case when $\rho = 1$. Cui and Blockley [1] showed that the calculus of the Dempster–Shafer [3] theory of evidence is a special case of IPT. However, not all Dempster–Shafer models are probabilistic in nature. In particular the transfer of belief model proposed by Smets [7] is a belief-based interpretation of Dempster–Shafer that does not involve probabilities.

The dependence parameter $\rho$ can be interpreted in terms of triangular norms (T-norms) [8,9] in which case the minimum value of $\rho$ corresponds to

$$T_1(a, b) = \max(0, a + b - 1);$$

the independence value of $\rho$ corresponds to

$$T_2(a, b) = a \cdot b;$$

and $\rho = 1$ corresponds to

$$T_3(a, b) = \min(a, b).$$

Intermediate values of $\rho$ correspond to other T-norms.

### 3. Support for compound propositions

Consider two propositions $E_1$ and $E_2$ with dependency between them $[\rho_l, \rho_u]$. The probability assignments to the power set of the universe of
discourse, i.e. \( E_1 \cap E_2, E_1 \cap E_2^c, E_1^c \cap E_2, E_1^c \cap E_2^c \) are illustrated in tabular form in Fig. 1 so that in terms of interval probabilities:

\[
P(E_1 \cap E_2) = [m_{11}, m_{11} + m_{13} + m_{31} + m_{33}],
\]

\[
P(E_1 \cap E_2^c) = [m_{12}, m_{12} + m_{13} + m_{32} + m_{33}],
\]

\[
P(E_1^c \cap E_2) = [m_{21}, m_{21} + m_{23} + m_{31} + m_{33}],
\]

\[
P(E_1^c \cap E_2^c) = [m_{22}, m_{22} + m_{23} + m_{32} + m_{33}].
\]

The values of \( m_{ij} \) on the interval \((0, 1)\) are by convention constrained so that

\[
m_{11} + m_{12} + m_{13} = S_n(E_1),
\]

\[
m_{21} + m_{22} + m_{23} = 1 - S_p(E_1),
\]

\[
m_{11} + m_{21} + m_{31} = S_n(E_2),
\]

\[
m_{12} + m_{22} + m_{32} = 1 - S_p(E_2),
\]

\[
m_{11} + m_{12} + m_{13} + \cdots + m_{33} = 1.
\]

From Eqs. (1) and (3)

\[
m_{11} = \rho_i(S_n(E_1) \land S_n(E_2)).
\]

From Eq. (6)

\[
m_{22} = S_n(E_1 \cap E_2^c) = 1 - S_p(E_1 \cup E_2),
\]

\[
\begin{array}{c|c|c|c}
S_n(E_1) & S_n(E_2) & 1 - S_p(E_2) & S_p(E_2) - S_n(E_2) \\
E_1 & E_2 & \overline{E}_2 & \overline{E}_2 U \\
\hline
S_n(E_1) & m_{11} & m_{12} & m_{13} \\
E_1 & m_{21} & m_{22} & m_{23} \\
1 - S_p(E_1) & m_{31} & m_{32} & m_{33} \\
\end{array}
\]

Fig. 1. Representation of compound propositions.
so from Eq. (2)

\[ m_{22} = 1 - S_p(E_1) - S_p(E_2) + \rho_p(S_p(E_1) \wedge S_p(E_2)) \]  (13)

Whilst \( P(E_1 \cap E_2) \) and \( P(E_1' \cap E_2') \) are uniquely defined, the constraints of Eqs. (7)-(13) do not result in unique intervals for \( P(E_1 \cap E_2) \) and \( P(E_1' \cap E_2') \) under all values of \( P(E_1), P(E_2) \) and \( \rho \). To obtain unique values of \( P(E_1 \cap E_2') \) and \( P(E_1' \cap E_2) \) would require specific knowledge about the dependency between \( E_1 \) and \( E_2 \) and between \( E_1' \) and \( E_2' \). Because this knowledge can be difficult to articulate it is preferable to calculate the family of permissible intervals for \( P(E_1 \cap E_2) \) and \( P(E_1' \cap E_2') \). An example of this procedure is shown in Fig. 2.

If \( E_1 \) and \( E_2 \) are items of the same evidence derived from different sources then the sum \( m_{12} + m_{21} \) is the conflict between the two items of evidence. This measure \( m_{12} + m_{21} \) is of great use in locating areas of conflicting evidence so that it may, where possible, be reconciled. Conflict is sometimes an unavoidable characteristic of the evidence in a knowledge base and if so will be reflected in the compound proposition. This is unlike Dempster's rule of combination where conflict is removed altogether by renormalization, leading to the familiar assertion that it generates counter-intuitive results [10].

Although the above procedure represents a restriction on the general assignment method [11] it is more general than the multiplication rule adopted in FRILL [12]. Indeed the multiplication rule in FRILL and in support logic programming [4,13] is one of the possible solutions when \( \rho \) is set to the independence value (Fig. 2(b)). The minimum rule in support logic programming is one of the possible solutions when \( \rho \) is set to unity (Fig. 2(a)).

This approach for establishing the assignments to the power set of the universe of discourse can be extended to apply to three or more propositions. For \( n \) propositions the tableau will occupy \( n \)-dimensional space.

4. Logical inference

4.1. Single item of evidence

Consider a conjecture \( H \) to which pertains evidence \( E \). To establish the support \( P(H) \) on the basis of the available evidence we require \( P(E) \) and some knowledge of the relationship between \( E \) and \( H \) which is defined by the conditional measures \( P(H|E) \) and \( P(H|\overline{E}) \). \( P(H) \) is obtained from the theorem of total probability

\[ P(H) = P((H \cap E) \cup (H \cap \overline{E})). \]

If \( H \cap E \) and \( H \cap \overline{E} \) are exclusive

\[ P(H) = P(H|E)P(E) + P(H|\overline{E})P(\overline{E}) \]  (14)
which can be rewritten as

\[ P(H) = P(H|E)P(E) + P(H|\overline{E})(1 - P(E)). \]

Dubois and Prade [14] showed that when all the terms are expressed as interval numbers, the bounds on \( P(H) \) are
and

\[
S_n(H) = S_n(H|E)S_n(E) + S_n(H|\bar{E})(1 - S_n(E)); \quad S_n(H|E) \geq S_n(H|\bar{E}),
\]

\[
S_n(H) = S_n(H|E)S_p(E) + S_n(H|\bar{E})(1 - S_p(E)); \quad \text{otherwise},
\]

(15)

\[
S_p(H) = S_p(H|E)S_p(E) + S_p(H|\bar{E})(1 - S_p(E)); \quad S_p(H|E) \geq S_p(H|\bar{E}),
\]

\[
S_p(H) = S_p(H|E)S_n(E) + S_p(H|\bar{E})(1 - S_n(E)); \quad \text{otherwise}.
\]

(16)

The relationship between \( E \) and \( H \) is a feature of the structure of the inference problem. For example \( E \) may be a necessary condition for \( H \) (Fig. 3(a)), in which case

\[
P(H|E) \leq [1, 1], \quad P(H|\bar{E}) = [0, 0],
\]

or \( E \) may be a sufficient condition for \( H \) (Fig. 3(b)), in which case

\[
P(H|E) = [1, 1], \quad P(H|\bar{E}) \leq [1, 1].
\]

In the sufficient condition there may not be specific evidence relating to \( P(H|\bar{E}) \) so, using an interval number, it can be set to the "unknown" interval of \([0, 1]\).

In the special case when \( E \) is a necessary and sufficient condition for \( H \)

\[
P(H|E) = [1, 1], \quad P(H|\bar{E}) = [0, 0].
\]

A weaker and more general condition is when \( E \) is relevant to \( H \) (Fig. 3(c)), in which case

\[
[0, 0] < P(H|E) \leq [1, 1], \quad [0, 0] \leq P(H|\bar{E}) [1, 1].
\]

4.2. Two items of evidence

Suppose now that there are two items of evidence \( E_1 \) and \( E_2 \) which pertain to \( H \). The potential sample space of \( H \) can now be partitioned into four mutually exclusive subsets, so

\[
P(H) = P(H|E_1 \cap \bar{E}_2)P(E_1 \cap \bar{E}_2) + P(H|\bar{E}_1 \cap E_2)P(\bar{E}_1 \cap E_2)
\]

\[
+ P(H|E_1 \cap E_2)P(E_1 \cap E_2) + P(H|\bar{E}_1 \cap \bar{E}_2)P(\bar{E}_1 \cap \bar{E}_2),
\]

(17)

![Fig. 3. Venn diagrams of (a) \( E \) necessary for \( H \); (b) \( E \) sufficient for \( H \); (c) \( E \) relevant to \( H \).](image-url)
where \( P(H|E_1 \cap \overline{E}_2), P(H|\overline{E}_1 \cap E_2), P(H|E_1 \cap E_2), P(H|\overline{E}_1 \cap \overline{E}_2) \) define the relationship between \( H \) and \( E_1 \) and \( E_2 \). The most general relationship is illustrated in Fig. 4.

So, for example, if \( E_1 \) and \( E_2 \) are both necessary conditions for \( H \) then

\[
P(H|E_1 \cap \overline{E}_2) = P(H|\overline{E}_1 \cap E_2) = P(H|\overline{E}_1 \cap \overline{E}_2) = [0, 0]
\]

so

\[
P(H) = P(H|E_1 \cap E_2) \cdot P(E_1 \cap E_2),
\]

and if \( E_1 \) and \( E_2 \) are necessary and sufficient conditions for \( H \) then

\[
P(H|E_1 \cap E_2) = [1, 1]
\]

so

\[
P(H) = P(E_1 \cap E_2)
\]

i.e. the assignments reduce to a logical AND operator.

The use of logical operators AND, OR, XOR, and NOT gives rise to special cases of the general Eq. (17). In these special cases each element of the power set of the universe of discourse is either wholly included or wholly excluded from \( H \) (see Table 1). However, experts in practical evidential situations manipulate ideas of necessity, sufficiency and relevance in richer and more flexible ways than can be expressed using the logical operators shown in Table 1. The natural language used in these situations reflects their complexity. An expert may, for example, explain that "to convincingly demonstrate hypothesis \( H \), I would do test \( E_1 \) and test \( E_2 \), but carrying out only one of the tests may be enough for me to be quite confident in hypothesis \( H \); of the two, test \( E_1 \) would probably tell me more about hypothesis \( H \) than test \( E_2 \); without test \( E_1 \) or test

![Fig. 4. Illustration of the power set of \( H \).](image-url)
Table 1

|                | $P(H|E_1 \cap E_2)$ | $P(H|\bar{E}_1 \cap E_2)$ | $P(H|E_1 \cap \bar{E}_2)$ | $P(H|\bar{E}_1 \cap \bar{E}_2)$ |
|----------------|---------------------|-----------------------------|---------------------------|-------------------------------|
| $E_1 \text{ AND } E_2$ | [0, 0]              | [0, 0]                      | [1, 1]                    | [0, 0]                        |
| $E_1 \text{ OR } E_2$    | [1, 1]              | [1, 1]                      | [1, 1]                    | [0, 0]                        |
| $E_1 \text{ XOR } E_2$   | [1, 1]              | [1, 1]                      | [0, 0]                    | [0, 0]                        |
| NOT $E_1$                | [0, 0]              | [1, 1]                      | [0, 0]                    | [1, 1]                        |
| NOT $E_2$                | [1, 1]              | [0, 0]                      | [0, 0]                    | [1, 1]                        |
| NOT ($E_1 \text{ OR } E_2$)| [0, 0]             | [0, 0]                      | [0, 0]                    | [1, 1]                        |

$E_2$, I would not have any idea about hypothesis $H''$. After further interrogation it may be established that the structure of the evidential situation can be described by the following assignments,

\[
P(H|E_1 \cap E_2) = [1.0, 1.0], \quad P(H|E_1 \cap \bar{E}_2) = [0.4, 0.9],
\]

\[
P(H|\bar{E}_1 \cap E_2) = [0.2, 0.6], \quad P(H|\bar{E}_1 \cap \bar{E}_2) = [0.0, 1.0].
\]

Establishing the relevance of evidence is a delicate empirical process. Great care is required in mapping from natural language to the mathematical structure of the problem. For example when in the above testimony the expert states that “... carrying out only one of the tests may be enough for me to be quite confident in hypothesis $H$; of the two, test $E_1$ would probably tell me more about hypothesis $H$ than test $E_2$...” she would normally be referring to $P(H|E_1)$ and $P(H|E_2)$ and would require further interrogation in order to establish $P(H|E_1 \cap \bar{E}_2)$ and $P(H|\bar{E}_1 \cap E_2)$.

Ancillary evidence may be employed to help establish relevance and to help structure the problem. Ancillary evidence itself may be the product of an inference network. The structure of the inference problem should be recognised as being dynamic. New ancillary evidence uncovered during the inference process may suggest revised structure and relevance measures.

Measures of relevance or dependency which do not take proper account of redundancy, where it exists, will over or under-value the force of evidence. Redundancy is a consequence of hidden dependencies between items of evidence. There is a limit to the amount of relevant independent evidence which can be adduced to support a particular hypothesis $H$. If $P(H) = 1$ any additional evidence is irrelevant to $H$ or completely dependent on the evidence already adduced in support of $H$. Rarely, however, can one be so specific about the relevance or dependency of evidence.

Bounds on $P(H)$ in Eq. (17) can be found by testing each of the permissible factorisations and the family of values for $P(E_1 \cap \bar{E}_2)$ and $P(\bar{E}_1 \cap E_2)$ (recall that $P(E_1 \cap \bar{E}_2)$ and $P(\bar{E}_1 \cap E_2)$ will not generally be uniquely defined):
\( S_n(H) = \inf_{m_{12}, m_{21}} \{ S_n(H|E_1 \cap E_2)S_n(E_1 \cap E_2) + S_n(H|E_1 \cap \overline{E_2})S_n(E_1 \cap \overline{E_2}) \\
+ S_n(H|\overline{E_1} \cap E_2)S_n(\overline{E_1} \cap E_2) + S_n(H|\overline{E_1} \cap \overline{E_2})S_n(\overline{E_1} \cap \overline{E_2}) \}, \)

\( S_p(H) = \sup_{m_{12}, m_{21}} \{ S_p(H|E_1 \cap E_2)S_p(E_1 \cap E_2) + S_p(H|E_1 \cap \overline{E_2})S_p(E_1 \cap \overline{E_2}) \\
+ S_p(H|\overline{E_1} \cap E_2)S_p(\overline{E_1} \cap E_2) + S_p(H|\overline{E_1} \cap \overline{E_2})S_p(\overline{E_1} \cap \overline{E_2}) \}, \)

where \( S_i = S_1 \ldots S_{16} \) are the permissible factorisations (see Table 2), each representing a different permutation of the nine assignments of the compound proposition, such that \( S_n \leq S_i \leq S_p \) and

\[ S_i(E_1 \cap E_2) + S_i(E_1 \cap \overline{E_2}) + S_i(\overline{E_1} \cap E_2) + S_i(\overline{E_1} \cap \overline{E_2}) = 1. \]

Inspection of Eqs. (18) and (19) together with Table 2 demonstrates that the most general values of \( S_n(H) \) and \( S_p(H) \) will be found when \( m_{12} \) or \( m_{21} \) are at a minimum.

For example suppose that \( P(E_1) = [0.3, 0.7] \) and \( P(E_2) = [0.2, 0.5] \) and \( \rho = [0.3, 0.7] \) as shown in Fig. 2(b). A range of permissible values for \( m_{12} \) and \( m_{21} \) and the corresponding assignments \( m_{12} \ldots m_{33} \) are listed in Table 3.

Now suppose that the inference problem has a rather general structure so that \( E_1 \cap E_2, E_1 \cap \overline{E_2}, \overline{E_1} \cap E_2, \) and \( \overline{E_1} \cap \overline{E_2} \) are all relevant to \( H \) and

\[ P(H|E_1 \cap E_2) = [0.5, 0.9], \quad P(H|E_1 \cap \overline{E_2}) = [0.7, 0.9], \]
\[ P(H|\overline{E_1} \cap E_2) = [0.2, 0.6], \quad P(H|\overline{E_1} \cap \overline{E_2}) = [0.0, 1.0]. \]

The calculation of the values of \( S_n(H) \) and \( S_p(H) \) according to Table 2 using the assignments in Table 3 is illustrated in Table 4. In this case the lower bound on \( S_n(H) = 0.18 \) (indicated by bold entries) which is found when \( m_{12} \) is at its minimum and the upper bound on \( S_p(H) = 0.96 \) (indicated by bold entries) which is found when \( m_{21} \) is at its minimum. Thus the most general inference is that \( P(H) = [0.18, 0.96] \). However, these bounds are conservative because they do not co-exist in the same solution. Whilst it is acceptable to proceed on the basis of these bounds the conservatism can reduce the precision of the inference. Increased precision is achievable by testing the co-existent bounds for which \( S_n(H) \) and \( S_p(H) \) are minimum and maximum, respectively.

4.3. \( n \) items of evidence

It there are \( n \) items of evidence \( E_1 \ldots E_n \) then the potential sample space of \( H \) can be partitioned into a power set with \( j \) elements \( H|\theta_1 \ldots H|\theta_j \) and
Table 2
Permissible combinations of $E_1$ and $E_2$

<table>
<thead>
<tr>
<th>$i$</th>
<th>$S_i(E_1 \cap E_2)$</th>
<th>$S_i(E_1 \cap \overline{E_2})$</th>
<th>$S_i(\overline{E_1} \cap E_2)$</th>
<th>$S_i(\overline{E_1} \cap \overline{E_2})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$m_{11} + m_{13} + m_{31} + m_{33}$</td>
<td>$m_{12} + m_{32}$</td>
<td>$m_{21} + m_{23}$</td>
<td>$m_{22}$</td>
</tr>
<tr>
<td>2</td>
<td>$m_{11} + m_{13} + m_{31} + m_{33}$</td>
<td>$m_{12}$</td>
<td>$m_{21} + m_{23}$</td>
<td>$m_{22} + m_{32}$</td>
</tr>
<tr>
<td>3</td>
<td>$m_{11} + m_{13} + m_{31} + m_{33}$</td>
<td>$m_{12} + m_{32}$</td>
<td>$m_{21}$</td>
<td>$m_{22} + m_{23}$</td>
</tr>
<tr>
<td>4</td>
<td>$m_{11} + m_{13} + m_{31} + m_{33}$</td>
<td>$m_{12}$</td>
<td>$m_{21} + m_{23}$</td>
<td>$m_{22} + m_{32} + m_{33}$</td>
</tr>
<tr>
<td>5</td>
<td>$m_{11} + m_{31}$</td>
<td>$m_{12} + m_{13} + m_{32} + m_{33}$</td>
<td>$m_{21} + m_{23} + m_{31}$</td>
<td>$m_{22} + m_{32}$</td>
</tr>
<tr>
<td>6</td>
<td>$m_{11} + m_{31}$</td>
<td>$m_{12} + m_{13} + m_{32} + m_{33}$</td>
<td>$m_{21} + m_{23} + m_{31}$</td>
<td>$m_{22} + m_{32}$</td>
</tr>
<tr>
<td>7</td>
<td>$m_{11}$</td>
<td>$m_{12} + m_{13} + m_{32} + m_{33}$</td>
<td>$m_{21} + m_{23} + m_{31}$</td>
<td>$m_{22} + m_{32}$</td>
</tr>
<tr>
<td>8</td>
<td>$m_{11}$</td>
<td>$m_{12} + m_{13} + m_{32} + m_{33}$</td>
<td>$m_{21} + m_{23} + m_{31}$</td>
<td>$m_{22} + m_{32}$</td>
</tr>
<tr>
<td>9</td>
<td>$m_{11} + m_{13}$</td>
<td>$m_{12}$</td>
<td>$m_{21} + m_{23}$</td>
<td>$m_{22} + m_{32}$</td>
</tr>
<tr>
<td>10</td>
<td>$m_{11} + m_{13}$</td>
<td>$m_{12} + m_{32}$</td>
<td>$m_{21} + m_{23}$</td>
<td>$m_{22} + m_{32} + m_{33}$</td>
</tr>
<tr>
<td>11</td>
<td>$m_{11}$</td>
<td>$m_{12} + m_{13}$</td>
<td>$m_{21} + m_{23}$</td>
<td>$m_{22} + m_{32} + m_{33}$</td>
</tr>
<tr>
<td>12</td>
<td>$m_{11}$</td>
<td>$m_{12} + m_{13}$</td>
<td>$m_{21} + m_{23}$</td>
<td>$m_{22} + m_{32} + m_{33}$</td>
</tr>
<tr>
<td>13</td>
<td>$m_{11} + m_{13}$</td>
<td>$m_{12}$</td>
<td>$m_{21} + m_{23}$</td>
<td>$m_{22} + m_{32} + m_{33}$</td>
</tr>
<tr>
<td>14</td>
<td>$m_{11} + m_{13} + m_{31}$</td>
<td>$m_{12}$</td>
<td>$m_{21} + m_{23}$</td>
<td>$m_{22} + m_{32} + m_{33}$</td>
</tr>
<tr>
<td>15</td>
<td>$m_{11}$</td>
<td>$m_{12} + m_{13}$</td>
<td>$m_{21} + m_{23}$</td>
<td>$m_{22} + m_{32} + m_{33}$</td>
</tr>
<tr>
<td>16</td>
<td>$m_{11} + m_{31}$</td>
<td>$m_{12} + m_{13}$</td>
<td>$m_{21} + m_{23}$</td>
<td>$m_{22} + m_{32} + m_{33}$</td>
</tr>
</tbody>
</table>
Table 3
Example of some permissible assignments to $m_{11}$...$m_{33}$ using example shown in Fig. 2(b)

<table>
<thead>
<tr>
<th>Case</th>
<th>$m_{11}$</th>
<th>$m_{12}$</th>
<th>$m_{13}$</th>
<th>$m_{21}$</th>
<th>$m_{22}$</th>
<th>$m_{23}$</th>
<th>$m_{31}$</th>
<th>$m_{32}$</th>
<th>$m_{33}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.06</td>
<td>0.00</td>
<td>0.24</td>
<td>0.09</td>
<td>0.15</td>
<td>0.06</td>
<td>0.05</td>
<td>0.35</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>0.06</td>
<td>0.00</td>
<td>0.24</td>
<td>0.14</td>
<td>0.15</td>
<td>0.01</td>
<td>0.00</td>
<td>0.35</td>
<td>0.05</td>
</tr>
<tr>
<td>3</td>
<td>0.06</td>
<td>0.09</td>
<td>0.15</td>
<td>0.00</td>
<td>0.15</td>
<td>0.15</td>
<td>0.14</td>
<td>0.26</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>0.06</td>
<td>0.24</td>
<td>0.00</td>
<td>0.15</td>
<td>0.15</td>
<td>0.14</td>
<td>0.11</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.06</td>
<td>0.24</td>
<td>0.00</td>
<td>0.14</td>
<td>0.15</td>
<td>0.01</td>
<td>0.00</td>
<td>0.11</td>
<td>0.29</td>
</tr>
</tbody>
</table>

\[ P(H) = \sum_{i=1}^{j} P(H|\theta_i)P(\theta_i) \quad \text{where } j = 2^n. \quad (20) \]

The first term in the summation in Eq. (20) determines the relationship between the body of evidence and $H$ is referred to as the \textit{relevance}. The second term is calculated from the evidence $E_1...E_n$ and the dependencies between the various items of evidence.

5. Integrating evidence on process dependability with probabilistic data

Eq. (14) can be applied to the important problem of how evidence concerning the dependability of a process can be integrated with the data handled by that process. The process may, for example, involve use of some data model $dm$ to manipulate some input parameters and obtain some prediction $x$ which may be expressed as a probability distribution $P(X = x|dm)$. Now the domain experts will recognise that, because the abstractions involved in its construction, model $dm$ is not a complete representation of reality. Suppose that the domain experts therefore construct a process model $pm$ to establish the dependability of $dm$. The evidence concerning the dependability of $dm$ may be vague, incomplete or contradictory so it would be appropriate to construct $pm$ using IPT [5]. Using the calculus outlined in Section 4 a measure of the dependability of $dm$, $P(dm = True|pm)$, is obtained from the process model. The situation is illustrated in Fig. 5.

Thus from the process modelling, the prediction $P(X = x|pm)$ can be expressed as an interval number obtained from Eq. (14) as follows.

\[
P(X = x|pm) = P(X = x|dm = True, pm) \cdot P(dm = True|pm) \\
+ P(X = x|dm = False, pm) \cdot P(dm = False|pm). \quad (21)
\]

$(X = x|dm = False, pm)$ is the value of the prediction given that $dm$ is not true or dependable. Whilst the expert may recognise that events outside the data model could be as important as those inside, it is very difficult to make any estimate of $(X = x|dm = False, pm)$. It may therefore in practice be assigned an interval number $[0, 1]$. 
### Table 4

Example of calculation of bounds on $S_n(H)$ and $S_p(H)$ using assignments shown in Table 3

<table>
<thead>
<tr>
<th>$i$</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Case 4</th>
<th>Case 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$S_n(H)$</td>
<td>$S_p(H)$</td>
<td>$S_n(H)$</td>
<td>$S_p(H)$</td>
<td>$S_n(H)$</td>
</tr>
<tr>
<td>1</td>
<td>0.45</td>
<td>0.87</td>
<td>0.45</td>
<td>0.87</td>
<td>0.45</td>
</tr>
<tr>
<td>2</td>
<td>0.21</td>
<td>0.91</td>
<td>0.21</td>
<td>0.91</td>
<td>0.27</td>
</tr>
<tr>
<td>3</td>
<td>0.44</td>
<td>0.89</td>
<td>0.45</td>
<td>0.87</td>
<td>0.42</td>
</tr>
<tr>
<td>4</td>
<td>0.19</td>
<td>0.93</td>
<td>0.20</td>
<td>0.91</td>
<td>0.24</td>
</tr>
<tr>
<td>5</td>
<td>0.19</td>
<td>0.89</td>
<td>0.19</td>
<td>0.89</td>
<td>0.23</td>
</tr>
<tr>
<td>6</td>
<td>0.44</td>
<td>0.86</td>
<td>0.44</td>
<td>0.86</td>
<td>0.41</td>
</tr>
<tr>
<td>7</td>
<td>0.24</td>
<td>0.89</td>
<td>0.24</td>
<td>0.89</td>
<td>0.26</td>
</tr>
<tr>
<td>8</td>
<td>0.48</td>
<td>0.86</td>
<td>0.48</td>
<td>0.86</td>
<td>0.44</td>
</tr>
<tr>
<td>9</td>
<td>0.50</td>
<td>0.87</td>
<td>0.51</td>
<td>0.87</td>
<td>0.48</td>
</tr>
<tr>
<td>10</td>
<td>0.49</td>
<td>0.89</td>
<td>0.51</td>
<td>0.87</td>
<td>0.45</td>
</tr>
<tr>
<td>11</td>
<td>0.48</td>
<td>0.86</td>
<td>0.51</td>
<td>0.87</td>
<td>0.44</td>
</tr>
<tr>
<td>12</td>
<td>0.47</td>
<td>0.88</td>
<td>0.51</td>
<td>0.87</td>
<td>0.41</td>
</tr>
<tr>
<td>13</td>
<td>0.18</td>
<td>0.91</td>
<td>0.18</td>
<td>0.91</td>
<td>0.20</td>
</tr>
<tr>
<td>14</td>
<td>0.19</td>
<td>0.93</td>
<td>0.18</td>
<td>0.91</td>
<td>0.24</td>
</tr>
<tr>
<td>15</td>
<td>0.23</td>
<td>0.91</td>
<td>0.23</td>
<td>0.91</td>
<td>0.23</td>
</tr>
<tr>
<td>16</td>
<td>0.24</td>
<td>0.93</td>
<td>0.23</td>
<td>0.91</td>
<td>0.27</td>
</tr>
</tbody>
</table>
It may be argued that the measure of process support $P(dm|pm)$ and the model result $P(X=x|dm)$ are essentially too distinct to be related in the same equation. If this is the case the decision maker will be inclined to use the evidence from the process model in some heuristic way, perhaps adopting a satisficing strategy by which the model has to achieve some minimum dependability before a decision is taken. On the other hand there must be some boundary (albeit fuzzy) to possible deviations from model predictions. Eq. (21) enables some limits to be put on that boundary in a way which is not possible with existing probabilistic approaches.

Suppose, for example, that a civil engineer is designing a harbour. The harbour operator has specified a maximum frequency with which a certain wave height may be exceeded within the harbour. The engineer has carried out some analysis of how effective the proposed design is at limiting wave heights and now wishes to decide whether the design dependably satisfies the harbour operator's specification. The situation is illustrated in Fig. 6.

A discussion of hierarchical process modelling may be found in Ref. [15]. A simplified version of a process model for establishing the frequency distribution of wave heights in the harbour is illustrated in Fig. 7. The model consists of processes ordered according to their precision of definition. The dependability of the process of establishing wave heights in the harbour can be estimated by evaluating the evidence for and the relevance of the sub-processes.

The engineer’s beliefs in the dependability, relevance and inter-dependencies of the sub-processes can be propagated using the calculus described in Section 4 to obtain an overall measure of the dependability of the top process in Fig. 7, that of establishing the frequency distribution of wave heights in the harbour. Suppose for example that this process has a support interval...
[0.78, 0.93] and the data model indicates that the specified wave height $x$ in the harbour will be exceeded in any year with a probability of 0.36 and the engineer has no knowledge of the wave conditions in the harbour given that the model is not dependable, then it follows that:
\[ P(dm = \text{True}|pm) = [0.78, 0.93], \]
\[ P(X > x|dm = \text{True}, pm) = [0.36, 0.36], \]
\[ P(dm = \text{False}|pm) = [0.07, 0.22], \]
\[ P(X > x|dm = \text{False}, pm) = [0.0, 1.0], \]

and so, applying Eq. (21) (expanded as in Eqs. (15) and (16))

\[ S_n(X > x|pm) = 0.36 \times 0.78 + 0.0 \times 0.22 = 0.28, \]
\[ S_p(X > x|pm) = 0.36 \times 0.78 + 1.0 \times 0.22 = 0.50. \]

So the evidence that the specified wave height will be exceeded is \([0.28, 0.50]\).

6. A comparison with bayesian inference

The approach described in this paper is one which attributes as much importance to the structure and relevance of evidence as it does to the evidence itself. This coincides with current research into networks of logical inference [16–18] which has explored the use of directed acyclic graphs for structuring inference problems. Emphasis on structure represents a significant shift away from traditional discussions of probability which have started with the assumption that the probability structure and even probability numbers are available at the outset of the analysis [19].

However, whilst our emphasis on structure coincides with current practice in the Bayesian school, the approach advocated in this paper differs from Bayesian inference in important respects. The first is that the method is aimed at inferring \(P(H)\) (see Eq. (20)) and not \(P(H|\theta)\) i.e. the inference is defined on the universe of discourse and is not necessarily conditioned on the evidence \(\theta\) (though it is of course conditional on the assumptions of the inference process). Bayesian inference, by contrast, is conditioned on \(\theta\) and so neglects \(P(H|\theta)\), a part of the universe of discourse which may be of the utmost relevance. Naturally there are situations when only \(P(H|\theta)\) is of interest in which case Bayesian conditioning is appropriate. The approach advocated here is designed for situations when this is not the case. Nevertheless, if particular parts of the universe of discourse are not relevant to a particular inference Eq. (20) can still be usefully employed by setting irrelevant areas of the universe of discourse to zero by appropriate assignments to \(P(H|\theta)\).

Bayesian inference relies on likelihood functions \(P(\theta|H)\). In some contexts, for example medical diagnosis or evidential reasoning on legal matters, it may be natural to assign likelihood functions to each item of relevant evidence. In other circumstances calculating likelihood functions could be very difficult. Consider for example a complex human process \(H\) which to be successful requires a number of sub-processes \(\theta_1, \ldots, \theta_i\) to be successful. It is very hard to
estimate the likelihood of each individual sub-process being successful given that the super-process is successful. The reverse approach of estimating the relevance of the sub-processes to the super-process is more natural.

Bayesian inference relies on a complete model of the problem domain. Yet in many complex modelling situations it is very difficult indeed to construct a complete model. The approach described in this paper explicitly recognises the issue of incompleteness and can function even if the domain is incomplete. Aspects of the inference problem which are unknown can be assigned an interval number $[0, 1]$. This uncertainty will propagate through the inference hierarchy and will be reflected in the overall outcome. A situation where the domain is incomplete can be referred to as an “open world” problem, so IPT has been described as an open world theory [1]. Bayesian inference, by contrast, is a “closed world” theory. An interval treatment of Bayesian belief networks [20] enables the sensitivity to uncertainty in the probability assignments to be expressed and explored. Nonetheless, the inference mechanism is still Bayes theorem.

The use of interval numbers enables a straightforward treatment of problems where evidence is inconsistent or conflicting. The approach described in Section 3 is generally applicable for all evidence be it consistent or conflicting. The degree of conflict can be readily calculated and communicated to the decision maker. In Bayesian inference networks contradictory or conflicting evidence is manipulated so as to reduce it to point probability values, an approach which does not do justice to the richness of some inferential situations.

7. Conclusions

A general approach to logical inference based on interval probability theory has been presented. A review of interval probability theory has stressed the significance of the dependence parameter $\rho$ which generalises other inference rules which assume a specific dependence relationship between evidence. Conflict between items of evidence, which can be unavoidable but also informative, is measured and propagated by the calculus.

The proposed inference method partitions the available evidence amongst the power set of the universe of discourse and assigns a relevance to each member of the power set. The same equation can therefore embody the available evidence and the structural relationship between the items of evidence and the hypothesis of interest. This approach recognises that the structure of an inference problem is as important as the evidence itself. The structural situations of necessary or sufficient evidence are special cases of the proposed approach.

Uncertain inference using interval probability theory is based on an open world view in which the problem domain need not be completely specified in order to obtain meaningful inferences. It is therefore capable of integrating data from imperfect models, recognising that events outside the model domain
may be as important as those inside it. It has been demonstrated how interval probability theory can be used to integrate the data handled by models with evidence about the dependability of those models. The proposed approach is particularly applicable to analysis of complex processes where Bayesian likelihood functions and prior probabilities are very difficult to establish. It can therefore be used to provide decision makers with information in a simple format which at the same time reflects the complexity of the inference problem and the richness of available evidence.

References