Distribution of expected utility in decision trees

Mats Danielson a,*, Love Ekenberg a, b, Aron Larsson b

a Department of Computer and Systems Sciences, Stockholm University and Royal Institute of Technology (KTH), Forum 100, SE-164 40 Kista, Sweden
b Department of Information Technology and Media, Mid Sweden University, SE-851 70 Sundsvall, Sweden

Received 27 June 2005; received in revised form 27 February 2006; accepted 21 September 2006
Available online 21 November 2006

Abstract

Evaluation of decision trees in which imprecise information prevails is complicated. Especially when the tree has some depth, i.e. consists of more than one level, the effects of the choice of representation and evaluation procedures are significant. Second-order representation and evaluation may significantly increase a decision-maker’s understanding of a decision situation when handling aggregations of imprecise representations, as is the case in decision trees or influence diagrams, while the use of only first-order results gives an incomplete picture. Furthermore, due to the effects on the distribution of belief over the intervals of expected utilities, the $T$-maximin decision rule seems to be unnecessarily pessimistic as the belief in neighbourhoods of points near interval boundaries is usually lower than in neighbourhoods near the centre. Due to this, a generalized expected utility is proposed. The results in this paper apply also to approaches, which do not explicitly deal with second-order information, such as standard decision trees or probabilistic networks using only first-order concepts, for example upper and lower bounds. Furthermore, the results also apply to other, non-probabilistic weighted trees such as multi-criteria weight trees.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Decision analysis; Second-order probabilities; Utility theory; Decision evaluation
1. Introduction

In decision analysis, whether for agents or human decision-makers, it is often the case that complete, adequate, and precise information is missing. The requirement to provide numerically precise information in such models has often been considered unrealistic in real-life decision situations. Consequently, during recent decades of rather intense activities within the decision analysis community (see, e.g., [1–3] for overviews) several approaches have emerged. In particular, first-order approaches, i.e. suggestions based on sets of probability measures, upper and lower probabilities, and interval probabilities, have prevailed. As a representation, they have been successful in research and have gained acceptance among decision-makers. In evaluation of decision models, though, they often result in overlapping expected utilities because they are not able to use all information available in the decision model. That is the first reason to study second-order approaches. Sometimes, the decision-maker has an idea on the distribution of belief within intervals given. First-order representations do not admit for discrimination between different beliefs in different values of the decision parameters. This seems unnecessarily restrictive since a decision-maker does not necessarily believe with the same faith in all possible functions that the parameter vectors represent, i.e. in all points between the upper and lower bounds. That is the second reason to study second-order approaches. Furthermore, effects only evident from a second-order analysis may lead to warped evaluation results not discovered with an upper and lower bound analysis even in cases where a decision-maker originally does not discriminate between different beliefs. That is the third reason to study second-order approaches. By discussing evaluations based on expected utility, these reasons will be explored.

To allow for more efficient modelling of the decision-maker’s beliefs, representations of decision situations could involve beliefs in sets of epistemically possible probability and utility functions as well as relations between them. Beliefs of such kinds can be expressed using higher-order belief distributions. However, they have usually been dismissed for various reasons, conceptual and computational, and approaches based on sets of probability measures, upper and lower probabilities, or interval probabilities have traditionally been preferred. We demonstrate in this paper why second-order reasoning is useful and how information efficiency can be taken into account also when handling aggregations of first-order representations as they occur in decision trees or probabilistic networks. This paper handles probabilistic single-criterion decision trees. While not explicitly treated here, the same effects occur in multi-criteria weight trees. Interval weights are likewise normalized (i.e. adding up to 1) and are amenable to the same effects as interval probabilities.

This paper will initially consider the representation of decision information. The representation is discussed from two viewpoints, structure and constraints. Thereafter, the evaluation of decision trees is discussed in two steps. First, interval probabilities are introduced. Next, additional belief information is considered, and the sum of weighted values is handled, where the values are, for instance, utilities. The belief can (but need not) be second-order probabilities. This is a general framework where probabilities and utilities as presented below is a special case closely connected to probabilistic decision trees and networks.

2. Representation

Decisions under risk (probabilistic decisions) are often given a tree representation, cf. [4]. In this paper, we let a decision frame represent a decision problem. The idea with such
a frame is to collect all information necessary for the model in one structure. The representational issues are of two kinds, structure and constraints (statements).

2.1. Tree structure

One of the building blocks of a frame is a decision tree. Formally, a decision tree is a graph.

**Definition 1.** A graph is a structure $\langle V, E \rangle$ where $V$ is a set of nodes and $E$ is a set of node pairs (edges). A tree is a connected graph without cycles. A rooted tree is a tree with a dedicated node as a root. The root is at level 0. The adjacent nodes, except for the nodes at level $i - 1$, to a node at level $i$ is at level $i + 1$. A node at level $i$ is a leaf if it has no adjacent nodes at level $i + 1$. A node at level $i + 1$ that is adjacent to a node at level $i$ is a child of the latter. A (sub-)tree is symmetric if all nodes at level $i$ have the same number of adjacent nodes at level $i + 1$. The depth of a rooted tree is $\max(n|\text{there exists a node at level } n)$.

A general graph structure is, however, too permissive for representing a decision tree. Hence, we will restrict the possible degrees of freedom of expression in the decision tree.

**Definition 2.** A decision tree $T = \langle C \cup A \cup N \cup \{r\}, E \rangle$ is a tree where

- $r$ is the root;
- $A$ is the set of nodes at level 1;
- $C$ is the set of leaves;
- $N$ is the set of intermediary nodes in the tree except these in $A$;
- $E$ is the set of node pairs (edges) connecting nodes at adjacent levels.

A decision tree is a way of modelling a decision situation where $A$ is the set of alternatives and $C$ is the set of final consequences. For convenience we can, for instance, use the notation that the $n$ children of a node $x_i$ are denoted, $x_{i1}, x_{i2}, \ldots, x_{in}$ and the $m$ children of the node $x_{ij}$ are denoted $x_{ij1}, x_{ij2}, \ldots, x_{ijm}$, etc.

In a decision tree structure, the edges can be considered to be events, and thus be assigned probability variables. Likewise, the leaves can be considered to be consequence nodes and be assigned utility variables (Fig. 1).

2.2. Constraints

Constraints are first-order decision-maker information and structural information that delimit the solution spaces of the probability and utility variables. There are two sources for constraints, the first source being decision-maker statements of probabilities and utilities. The statements are translated into corresponding constraints. Such constraints can either be range constraints (containing only one variable) or various kinds of comparative constraints. Given consequences $c_i, c_j, c_k$, and $c_m$, denote their probabilities $p_i, p_j, p_k$,

---

1 Even the edge from an alternative can be considered an event with a probability in $\{0, 1\}$. 
and $p_m$ and denote their utilities $v_i$, $v_j$, $v_k$, and $v_m$. Then user statements can be of the following kinds for real numbers $a_1$, $a_2$, $d_1$, $d_2$, $d_3$, and $d_4$:

- **Range constraints**: “$v_i$ is between $a_1$ and $a_2$” is denoted $v_i \in [a_1, a_2]$ and translated into $v_i \geq a_1$ and $v_i \leq a_2$. Similarly for probabilities, “$p_i$ is between $b_1$ and $b_2$” is denoted $p_i \in [b_1, b_2]$ and translated into $p_i \geq b_1$ and $p_i \leq b_2$.

- **Comparative constraints**: Several possibilities; examples include “the difference between $v_i$ and $v_j$ is between $d_1$ and $d_2$” is denoted $v_i - v_j \in [d_1, d_2]$ and translated into $v_i - v_j \geq d_1$ and $v_i - v_j \leq d_2$. “The difference between $v_i - v_j$ and $v_k - v_m$ is between $d_3$ and $d_4$” is denoted $(v_i - v_j) - (v_k - v_m) \in [d_3, d_4]$ and translated into $(v_i + v_m) - (v_j + v_k) \geq d_3$ and $(v_i + v_m) - (v_j + v_k) \leq d_4$.

Constraints of the form $v_i/v_j \geq a_1$ are of interest for some applications. Such constraints depend on the utilities being defined on a ratio scale. This paper deals with constraints applicable to all cardinal utility scales, but the results are equally applicable to decision situations with quotient expressions on a ratio scale. The other source for constraints is implicit constraints. They emerge either from properties of the variables or from structural dependencies. Such constraints can either be default constraints (involving a single variable) or various kinds of structural constraints (involving more than one variable).

- **Default constraints**: A value scale range of $v_i$ from $s$ to $t$ is expressed as $v_i \in [s, t]$ and translated into $v_i \geq s$ and $v_i \leq t$.

- **Structural constraints**: For each parent node $x_i$ with children $x_{ij}$, the normalization constraint for probabilities is $\sum_j p_{ij} = 1$.

Combining these two sources, constraint sets are obtained. A constraint set can either be independent (containing only constraints involving a single variable each), or it can be dependent (also containing constraints involving more than one variable).

**Definition 3.** Given a decision tree $T$, let $P$ be a set of constraints in the variables $\{p_{i_1} \ldots i_n\}$. Substitute the intermediate node labels $x_{i_1 \ldots i_n}$ with $p_{i_1 \ldots i_n}$: $P$ is a...
probability constraint set for $T$ if, for all sets $\{p_{i1}, \ldots, p_{im}\}$ of all sub-nodes of nodes $p_{i}$ that are not leaves, the statements $p_{ij} \in [0, 1]$ and $\sum_j p_{ij} = 1$, $j \in [1, \ldots, m_{i}]$ are in $P$, where $m_{i}$ is the number of sub-nodes to $p_{i}$.

Thus, a probability constraint set relative to a decision tree can be seen as characterizing a set of discrete probability distributions, i.e. belief in discrete probabilities associated with, for example, edges in decision trees. Normalization constraints ($\sum_j p_{ij} = 1$) require the probabilities of sets of exhaustive and mutually exclusive nodes to sum to one.

**Definition 4.** Given a decision tree $T$, let $V$ be a set of constraints in $\{v_{i,j}\}$. Substitute the leaf labels $x_{i,j}$ with $v_{i,j}$. Then $V$ is a utility constraint set for $T$.

Similar to a probability constraint set, a utility constraint set can be seen as characterizing a set of utility functions. In this paper, we will without loss of generality assume that the utility variables’ ranges are $[0, 1]$. The elements above can be combined to create a decision frame, which constitutes a complete first-order description of the decision situation.

**Definition 5.** A decision frame is a structure $\langle T, P, V \rangle$, where $T$ is a decision tree, $P$ is a probability constraint set for $T$, and $V$ is a utility constraint set for $T$.

Thus, a decision frame combines structure with first-order constraints.

### 2.3. Beliefs

Approaches for extending interval representations by using distributions over classes of probability and utility measures have been developed into various hierarchical models such as second-order probability theory [5–8]. Gärdenfors and Sahlin consider global distributions of beliefs but restrict themselves to the probability case and to interval representations. Furthermore, they neither investigate the relation between global and local distributions, nor do they introduce methods for determining the consistency of user-asserted sentences [9]. The same criticism applies to [10–12].

To facilitate a better qualification of the various possible functions, second-order estimates, such as distributions expressing various beliefs, can be introduced over an $n$-dimensional space where each dimension corresponds to possible probabilities of events or utilities of consequences.

**Definition 6.** Let a unit cube be represented by $B = [0, 1]^k$. A belief distribution over $B$ is a positive distribution $F$ defined on $B$ such that

$$\int_B F(x) dV_B(x) = 1$$

where $V_B$ is a $k$-dimensional Lebesgue measure on $B$. The set of all belief distributions over $B$ is denoted by $BD(B)$. In some cases, we will denote a unit cube by $(b_1, \ldots, b_k)$ to make the number of dimensions and the labels of the dimensions clearer.

In this way, the distributions can be used to express varying strengths of belief in different first-order probability or utility vectors.
Example 1. Assume that the function

\[ f(x_1, x_2) = \begin{cases} 
3(x_1^2 + x_2^2) & \text{if } 1 \geq x_2 \geq x_1 \geq 0 \\
0 & \text{otherwise}
\end{cases} \]

represents beliefs in different vectors \((x_1, x_2)\). The volume under the graph of this function is 1.

Example 2. The functions \(f: b_1 \rightarrow b_1\) and \(h: b_2 \rightarrow b_2\) are belief distributions over the one-dimensional unit cubes \((b_1)\) and \((b_2)\), respectively defined by

\[ f(x_1) = \max(0, \min(-100x_1 + 20, 100x_1)) \]

and

\[ h(x_2) = \max\left(0, \min\left(-\frac{100}{3}x_2 + \frac{80}{3}, \frac{200}{3}x_2 - \frac{100}{3}\right)\right). \]

These have graphs given by triangles with bases on the \(x_1\)-axis and the \(x_2\)-axis, respectively, and with areas = 1. Therefore, \(g(x_1, x_2) = f(x_1) \cdot h(x_2)\) is a belief distribution over a unit cube \((b_1, b_2)\).

In a first-order representation, the belief is expressed only as an interval with no qualification within the interval. Thus, the decision-maker believes in points within the interval and does not believe in points outside of it. Sometimes, first-order information is the only information available, sometimes there is more. It is important to be able to model all information available, whether only first-order or also second-order.

2.4. Local distributions

Second-order information available in a given decision situation is often local over subsets of lower dimensions (most decision-makers are unable to perceive their global beliefs over, say, a 100-dimensional cube). An important issue is therefore to investigate the relationship between different types of distributions. A reasonable semantics for this relationship, i.e. what do beliefs over some subset of a unit cube mean with respect to beliefs over the entire cube, is provided by summing up all possible belief values of the vectors with some components fixed. This is captured by the concept of S-projections.

Definition 7. Let \(B = (b_1, \ldots, b_k)\) and \(A = (b_{i_1}, \ldots, b_{i_l})\), \(i_j \in \{1, \ldots, k\}\) be unit cubes. Let \(F \in BD(B)\), and let

\[ f_A(x) = \int_{B \setminus A} F(x) \, dV_{B \setminus A}(x). \]

Then \(f_A\) is the S-projection of \(F\) on \(A\). This projection is denoted \(f_A = \text{proj}_A(F)\).

Thus, an S-projection of a belief distribution is also a belief distribution. A special case of projection is when belief distributions over the axes of a unit cube \(B\) are S-projections of a belief distribution over \(B\).

Definition 8. Given a unit cube \(B = (b_1, \ldots, b_k)\) and a distribution \(F \in BD(B)\). Then the distribution \(f(x_i)\) obtained by
\[ f_i(x_i) = \int_{B_i} F(x) dV_{B_i}(x) \]

where \( B_i = (b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_k) \) is a belief distribution over the \( b_i \)-axis. Such a distribution will be referred to as a local distribution.²

The rationale behind local distributions is that the resulting belief in, e.g., the point 0.1 is, in a sense, the sum of all beliefs over the vectors where 0.1 is the first component, i.e. the totality of belief in this point.

Example 3. Let a unit cube be given. Assume that the vectors in this cube are represented by pairs. Each of these pairs is assigned a belief, e.g., \( g(0.1,0.4) = 0.2, g(0.1,0.7) = 0.3 \), etc.

Example 4. Given a unit cube \([0,1]^3\) with positive uniform belief on the surface where \( \sum_{i=1}^{3} x_i = 1 \), the S-projection \( f(x_i) \) on the axes is \( f(x_i) = 2 - 2x_i \), i.e.

\[ f(x_i) = \int_0^{1-x_i} \frac{2}{3} \sqrt{3} \, dy = 2 - 2x_i \]

In a first-order representation, belief cannot be assigned to specific points or sets of points within intervals. Thus, it seems that the decision-maker has to implicitly believe in the same way in all points within the upper and lower bounds. In Example 4, the bounds for each variable correspond to the interval \([0,1]\). But it is mathematically impossible to believe equally in the entire range due to the warp effect. This effect will be discussed in Section 3.

2.5. Centroids

Intuitively, the centroid of a distribution is a point in space where some of the geometrical properties of the distribution can be regarded as concentrated. This is, in some respects, analogous to the centre of mass of physical bodies. The centroid will turn out to be a good representative of the distribution in expected utility calculations.

Definition 9. Given a belief distribution \( F \) over a cube \( B \), the centroid \( F_c \) of \( F \) is

\[ F_c = \int_B xF(x) dV_B(x) \]

where \( V_B \) is some \( k \)-dimensional Lebesgue measure on \( B \).

Centroids are invariant under projections on subsets of the unit cubes in the sense that the S-projections of a centroid on a subset have the same coordinates as the centroids of the corresponding S-projections [8]. Thus, a local distribution of a belief distribution preserves the centroid in that dimension.

² Note that belief is more general than second-order probability. For the special case when the belief distributions are second-order probability distributions, global distributions would be called joint distributions and local distributions would be called marginal distributions. Similarly, centroids (defined in the next section) would be called expectations.
Example 5. The centroid \( f_c \) of the local distribution given in Example 4 is
\[
f_c = \int_{0}^{1} x \cdot (2 - 2x) \, dx = \frac{1}{3}
\]
i.e. the center of mass in an ordinary triangle.

In Example 4 above, the upper and lower first-order bounds correspond to the interval \([0, 1]\). The midpoint of the interval, 0.5, is not a good single-point representative of the decision-maker’s belief. The vector \((0.5, 0.5, 0.5)\) is not even consistent with the normalization constraint.

The second-order information is collected together with first-order information (from the previously defined decision frame) into a composite frame holding all information connected to a particular decision situation. We will, in the sequel, use the notation of a general decision frame \(\langle T, P^*, V^* \rangle\) as a (first-order) decision frame \(\langle T, P, V \rangle\) augmented with second-order information (global distributions or sets of local distributions) over the probabilities and utilities in \(P\) and \(V\), respectively.

3. Evaluation

In the basic model of Bayesian decision analysis, a decision-maker has to choose an alternative (action) from a non-empty, finite set \(\mathcal{A} = \{A_1, \ldots, A_n\}\) of possible alternatives. Each alternative may end up in a finite set of consequences, and the resulting consequence of each alternative depends on the true (but probably unknown) state of nature \(\theta \in \Theta = \{\theta_1, \ldots, \theta_n, \ldots, \theta_k\}\). The corresponding outcome is then evaluated by means of a utility function \(u\) satisfying
\[
\begin{align*}
(\mathcal{A} \times \Theta) &\to \mathbb{R} \\
(A, \theta) &\to u(A, \theta)
\end{align*}
\]

Since the true state of nature is unknown, the model asserts the knowledge of the probability distribution \(P(\cdot)\) on \((\Theta, P(\Theta))\) where \(P(\Theta)\) is the power set of \(\Theta\). The most common selection procedure is referred to as the principle of maximizing the expected utility (PMEU), and is argued for in [14] and [15]. It is implied from widely accepted axiom systems defining formal models of rationality. Other selection procedures, for example thresholds and regret functions, are discussed in [20]. In PMEU, the alternative \(A\) to prefer is the alternative which maximizes the expected utility for all \(A_i \in \mathcal{A}\).

Definition 10. The principle of maximizing the expected utility is accepted if a decision-making agent chooses the alternative \(A^*\) from a set \(\mathcal{A}\) whenever
\[
A^* = \arg \max_A (E(A))
\]
where
\[
E(A) = \sum_{\theta \in \Theta} u(A, \theta) \cdot P(\theta)
\]
for all \(A \in \mathcal{A}\).

A preference ordering relation \(\succeq\) on \(\mathcal{A}\) is implied from the magnitudes of the different alternatives’ expected utility.
3.1. Interval evaluation

An important relation in Bayesian decision analysis is \( A_i \succeq A_j \iff E(A_i) \geq E(A_j) \) for any two alternatives \( A_i, A_j \in \mathcal{A} \). This relation is useful if complete, adequate, and precise information is available. In many real-life decision situations, this is not the case. As a consequence, much work in the decision analysis community has been focused on developing models and frameworks handling imprecision. Some well-known approaches are based on upper and lower probabilities [16], sets of probability measures [17,18], and interval probabilities [3]. For instance, generalizations of probability and utility estimates can straightforwardly be expressed by the following.

**Definition 11.** Given a finite sample space \( \Theta \) and a \( \sigma \)-field \( \Gamma \) of random events in \( \Theta \), the probability \( P(\theta_i) \) of state \( \theta_i \) is expressed as the variable \( p_i \) bounded by the following constraints
\[
\sup P(\theta_i) = 1 - \inf P(\neg \theta_i)
\]
\[
\sum_{\theta_i \in \Theta} p_i = 1
\]

**Definition 12.** Let \( L \) be a set of mappings \( L = \{u(\mathcal{A} \times \Theta) \rightarrow [0,1]\} \) where all \( u \in L \) are increasing. Given a subset \( U \subset L \) such that \( u_{ij} = \{u(A_i, \theta_j): u \in U\} \) is a closed interval, then the interval-valued utilities are defined in terms of the closed intervals \( u_{ij} \).

The expected utility of the alternatives represented by a classical decision tree are straightforwardly calculated when all components are numerically precise. When the domains of the terms are solution sets to probability and utility constraint sets, this is not as straightforward, but there are efficient methods available [19,20]. The probability and utility constraint sets are collections of linear inequalities. A minimal requirement for such a system of inequalities to be meaningful is that it is consistent, i.e. there must be some vector of variable assignments that simultaneously satisfies each inequality in the system.

The first step in an evaluation procedure is to calculate the meaningful (consistent) constraint sets in the sense above. Ensuing consistency, the primary evaluation rule for a decision tree model is based on a generalized expected utility. Since neither probabilities nor utilities are fixed numbers, evaluation of the expected utility yields multi-linear objective functions.

**Definition 13.** Given a decision frame \( \langle T, P, V \rangle \), \( GEV(A_i) \) denotes the generalized expected utility of alternative \( A_i \) and is obtained from
\[
\sum_{i_1=1}^{n_0} p_{i_1} \sum_{i_2=1}^{u_{i_1}} p_{i_2} \cdots \sum_{i_{m-1}=1}^{u_{i_{m-2}}} p_{i_{m-1}i_{m-2}} \cdots \sum_{i_{m-1}=1}^{u_{i_{m-2}}} p_{i_{m-1}i_{m-2}} \cdots \sum_{i_{m-1}=1}^{u_{i_{m-2}}} p_{i_{m-1}i_{m-2}} \cdots \sum_{i_{m-1}=1}^{u_{i_{m-2}}} p_{i_{m-1}i_{m-2}} \cdots \sum_{i_{m-1}=1}^{u_{i_{m-2}}} p_{i_{m-1}i_{m-2}} \cdots v_{i_1i_2\cdots i_{m-2}i_{m-1}i_m}
\]
where \( p_{i_1j}, j \in [1, \ldots, m] \), denote probability variables in \( P \), \( v_{ij} \), denote utility variables in \( V \), and \( n_{ij} \) is the number of children nodes of the node at depth \( j \).

Methods for evaluating GEV in imprecise domains mainly involve finding lower and upper bounds on expected utilities in decision trees. Optimisation of non-linear expressions, such as the expected utility, subject to linear constraints (the probability and utility constraint sets) are computationally demanding problems to solve for an interactive or
real-time computer-based tool in the general case, using techniques from the area of non-linear programming. In, e.g., [13,24,25] there are discussions about computational procedures that reduce such non-linear decision evaluation problems to systems with linear objective functions, solvable with ordinary linear programming methods. The procedures yield interval estimates of the evaluations, i.e. upper and lower bounds of the expected utilities for the alternatives, yielding interval-valued expected utilities to be compared. However, when trying to compare the expected utility intervals directly, many decision situations will lead only to a partial preference order on \( A \) since the intervals are overlapping. Due to this complication, some authors suggest the \( \Gamma \)-maximin\(^3\) principle [21–23] for evaluation of interval-valued decision problems.

**Definition 14.** The \( \Gamma \)-maximin principle is accepted if a decision-making agent chooses the alternative \( A^* \) from a set \( \mathcal{A} \) whenever

\[
A^* = \arg\max_A (\inf(E(A)))
\]

where

\[
\inf E(A) = \inf \sum_{\theta \in \Theta} u(A, \theta) \cdot P(\theta)
\]

for all \( A \in \mathcal{A} \).

However, this is quite a pessimistic decision rule, disregarding much of the information present in the decision model. Below we will present what we consider a more balanced and information efficient candidate.

Other decision rules available are based on the concept of admissibility or derivations thereof [20]. The common denominator is that the rules consider some combination of upper and lower bounds of the expected utility. Such a first-order rule based procedure is a first step in an interval decision analysis, but it cannot really solve the overlap situation. More can be done, however, using the information available if the evaluation continues with second-order (belief) analyses.

### 3.2. Belief evaluation

Belief evaluation in decision trees proceeds in three steps. In the first step, global belief is generated from local belief. Global belief distributions are obtained by multiplying the local belief distributions as explained below. In the second step, the global belief is used in obtaining the expected utility by term-wise multiplication with the belief in utilities and subsequent summation of the generated terms. In the final step, the alternatives are analysed using the belief information. Below, we will in separate subsections discuss the three steps.

There are three major kinds of second-order analysis (the third step). In a **centroid analysis**, the centroid is employed as a single-point representative of the distributions. The representatives are compared in order to generate a preference order on the alternatives. In a **contraction analysis**, the centroid is employed in a generalisation of interval analysis in which the first-order intervals are contracted by a procedure allowing points closer to the centroid to be emphasized in the evaluation [13]. In a **distribution analysis**, the resulting

\(^3\) When loss functions are used instead of utility functions, the rule is labelled \( \Gamma \)-minimax.
distributions of expected utility are compared by assessing the fractions of belief overlapping between alternatives being evaluated.

3.3. Global probability belief

Obtaining global belief from local is the first step in a belief evaluation. It corresponds to generating global probability variables \( q_{ijk} \) from a multiplication of local probability variables \( p_i \cdot p_{ij} \cdot p_{ijk} \) in a first-order analysis. Let \( G \) be a belief distribution over the two cubes \( A \) and \( B \). Assuming that \( G \) has a positive support on the feasible probability distributions at level \( i \) in a decision tree, i.e. is representing them (the support of \( G \) in cube \( A \)), as well as on the feasible probability distributions of the children of a node \( x_{ij} \), i.e. \( x_{ij1}, x_{ij2}, \ldots, x_{ijm} \) (the support of \( G \) in cube \( B \)). Let \( f = \text{proj}_A(G) \) and \( g = \text{proj}_B(G) \) as in Definition 7. Then the functions \( f \) and \( g \) are belief distributions according to [7]. Furthermore, there are no relations between two probabilities at different levels (having different parent nodes) so the distributions \( f \) and \( g \) are independent. Consequently, the following combination rule for the distribution over the product of the distributions \( f \) and \( g \) has a well-defined semantics, i.e. it expresses a convolution of two densities which is the density of the product under an independence assumption.

Definition 15. The cumulative distribution \( H \) of a product of two belief distributions \( f(x) \) and \( g(y) \) is

\[
H(z) = \int \int \Gamma_z f(x)g(y) \, dx \, dy = \int_0^1 \int_0^{y/x} f(x)g(y) \, dy \, dx = \int_0^1 f(x)G \left( \frac{z}{x} \right) \, dx
\]

where \( G \) is the antiderivative of \( g \), \( \Gamma_z = \{(x,y):x \cdot y = z\} \), and \( 0 \leq z \leq 1 \). Let \( h(z) \) be the corresponding density function. Then

\[
h(z) = \frac{d}{dz} \int_z^1 f(x)G \left( \frac{z}{x} \right) \, dx = \int_z^1 \frac{f(x)g(z/x)}{x} \, dx
\]

Let us now consider the relation to traditional (first-order) interval calculus. When aggregating interval probabilities and utilities in a tree as above, there are two main classes of belief distributions to consider, centred and non-centred.

- **Centred**: Belief distributions with at least as much mass toward the centre as toward the boundaries. Examples include triangular, trapezoid, uniform, and bell-shaped distributions.
- **Non-centred**: Belief distributions with less mass toward the centre than toward the boundaries. Examples include convex and bimodal distributions.

The uniform distribution is the limiting case of the centred class, having an equal distribution of belief mass over the interval. The centred class contains most implicit distributions, i.e. when there is no explicit information on the distribution of belief. If there is explicit information on, e.g., the belief in only boundary points, then the distribution will belong to the non-centred class. In Theorems 1 and 2 below, we will discuss the *warp effect*.
on centred distributions by using one of the class members least affected by warp, i.e. the uniform distribution. The effect is still highly visible. Most other centred distributions are more affected, e.g. symmetric distributions.

**Theorem 1.** Let \( f_1(x_1), \ldots, f_m(x_m) \) be independent uniform belief distributions over the intervals \([0, 1]\). The product \( h_m(z_m) \) over these \( m \) factors is the distribution

\[
h_m(z_m) = \frac{(-1)^{m-1}(\ln(z_m))^{m-1}}{\Gamma(m)}
\]

where \( \Gamma(m) = (m-1)! \) is the Euler Gamma function.

**Proof.** By induction:

\[
h_1(z_1) = \frac{(-1)^{1-1}(\ln(z_1))^{1-1}}{\Gamma(1)} = 1
\]

\[
h_{m+1}(z_{m+1}) = \int_{z_{m+1}}^{1} \frac{h_m(x)f_{m+1}(\frac{z_{m+1}}{x})}{x} \, dx
\]

\[
= \int_{z_{m+1}}^{1} \frac{(-1)^{m-1}(\ln(x))^{m-1}}{(m-1)!} \, dx
\]

\[
= \frac{(-1)^{m}(\ln(z_{m+1}))^{m}}{(m)!}
\]

The example shows the distributions at increasing levels of depth. The warp effect can be clearly seen as the distributions tend to lower values in deeper trees. The interval boundaries are still \([0, 1]\) for all distributions. This illustrates that first-order information does not give the complete picture.

**Example 6.** The distributions \( h_m(z_m) \) in **Theorem 1** are belief distributions, and Fig. 2 below shows, from right to left, the plots of the functions of depth 2–7, i.e. \(- \ln(x), \frac{-\ln^2(x)}{2}, \frac{-\ln^3(x)}{6}, \frac{-\ln^4(x)}{24}, \frac{-\ln^5(x)}{120}, \frac{-\ln^6(x)}{720}\).

An important property of the distribution above is its centroid. It experiences the same warp effect, tending to lower values as the level of depth increases.

**Theorem 2.** The centroid of the distribution \( h_m(z_m) \) in **Theorem 1** is \( 2^{-m} \).

**Proof**

\[
\int_0^1 z_m \frac{(-1)^{m-1}(\ln(z_m))^{m-1}}{\Gamma(m)} \, dz_m = \frac{(-1)^{m-1}}{\Gamma(m)} \int_0^1 z_m (\ln(z_m))^{m-1} \, dz_m
\]

\[
= \frac{(-1)^{m-1}}{\Gamma(m)} \left( -\left( \frac{-1}{2} \right)^m \right) \Gamma(m) = 2^{-m}
\]
The important observations above are that the mass of the resulting belief distributions become more concentrated to the lower values the deeper the tree is and, as a consequence, the centroid tends towards lower values. Already after one multiplication, this effect is significant. This is called the warp effect. It should be regarded as additional information by any method employing an interval calculus of weights (such as probabilities or multi-criteria weights). As mentioned above, these effects are not dependent on the use of uniform distributions. Rather, they are even more emphasized for most other centre-focused (e.g. triangular or bell-shaped) belief distributions. For this reason, belief evaluation is important also for decision problems containing only first-order statements.

3.4. Belief in expected utility

Obtaining the belief in expected utilities from global belief is the second step in a belief evaluation. It consists of two sub-steps, first multiplying belief in global probability and utility, and then adding all the multiplied belief. In a first-order analysis, the first sub-step corresponds to multiplying global probability variables $q_{ijk}$ with utility variables $u_{ijk}$. The second sub-step corresponds to adding the multiplied terms $q_{ijk} \cdot u_{ijk}$.

Let $B_F = [0,1]^k$ be a unit cube with an associated belief distribution $F$. The local distributions $f_i(x_i)$ can be calculated through the concept of S-projections. Then the following rule for the distribution over a sum has well-defined semantics.

**Definition 16.** The distribution $h_2$ on a sum $y_2 = x_1 + x_2$ of two independent variables associated with belief distributions $f_1(x_1)$ and $f_2(x_2)$ is given by a convolution

$$h_2(y_2) = \int_{-\infty}^{\infty} f_1(x_1) f_2(y_2 - x_1) \, dx_1.$$

When letting $y_k = x_1 + \cdots + x_k$ the belief distribution $h_k(y_k)$ on the sum $y_k = \sum_{i=1}^{k} x_i$ is obtained by induction applying **Definition 16**.

Centroids are preserved under projections, in the sense that the centroid of an S-projection of $F$ on $A \subset \text{im} F$ share coordinates with the centroid of $F$ [8]. Thus, a local distribution of a belief distribution preserves the centroid in that dimension. From [8], it is clear...
that the centroid is multiplicative, i.e. for two local belief distributions \( f \) and \( g \) on independent variables \( x \) and \( y \) with centroids \( f_c \) and \( g_c \), the centroid of the distribution on their product \( x \cdot y \) is given by \( f_c \cdot g_c \). An equally important property is that the centroid is additive as well.

**Lemma 1.** The horizontal centroid \( h_c \) of \( h \), where \( h \) is defined as in **Definition 16**, is the sum of the horizontal centroids of each \( f_i \), i.e. the horizontal centroid is additive.

**Proof.** This follows directly from the property of convolution in which the horizontal centroid adds. □

When calculating expected utilities, each term is a product of a probability and a utility. From this reasoning, the following rule for the distribution over the line segment of possible expected utilities follows.

**Definition 17.** Given a probability unit cube \( B_P = (p_1, \ldots, p_k), \sum p_i = 1 \), and a utility unit cube \( B_U = (u_1, \ldots, u_k) \), an expected utility unit cube, denoted \( B_{EU} \), is the cross product \( B_{EU} = B_P \times B_U \).

Thus, given any point \( e = (p_1, u_1, p_2, u_2, \ldots, p_k, u_k) \in B_{EU} \), there is an expected utility \( z \in [0, 1] \) such that \( z = p_1 \cdot u_1 + p_2 \cdot u_2 + \cdots + p_k \cdot u_k \) whenever \( \sum_{i=1}^k p_i = 1 \).

**Definition 18.** Let \( \langle T, P^*, V^* \rangle \) be a general decision frame. The function \( h(z) \) on a sum \( z = \sum_{i=1}^k p_i u_i \) of a set of products of two variables \( \{p_i \cdot u_i\}_{i=1}^k \), associated with global belief distributions \( F(p_1, \ldots, p_k) \) and \( G(u_1, \ldots, u_k) \) in \( \langle T, P^*, V^* \rangle \), is obtained by evaluating the integral

\[
h(z) = \int_{S_z} F(p)G(u)dS_z(p, u)
\]

where \( S_z = \{ (p_1, u_1, \ldots, p_k, u_k) : z = \sum_{i=1}^k p_i \cdot u_i \in [0, 1]^{2k}, 0 \leq z \leq 1 \} \), and \( dS_z(p, u) \) is the surface area element.\(^4\)

**Lemma 2.** Let \( \langle T, P^*, V^* \rangle \) be a general decision frame. The function \( h(z) \) in **Definition 18** is a belief distribution.

**Proof.** Since \( \int_{B_{EU}} |F(p)G(u)|dV_{B_{EU}}(p, u) < \infty \), applying Fubini’s theorem yields

\[
\int_0^1 h(z)dz = \int_0^1 \left( \int_{S_z} F(p)G(u)dS_z(p, u) \right)dz = \int_{B_{EU}} F(p)G(u)dV_{B_{EU}}(p, u)
= \left( \int_{B_P} F(p)dV_{B_P}(p) \right) \left( \int_{B_U} G(u)dV_{B_U}(u) \right) = 1 \cdot 1 \quad \Box
\]

\(^4\) In this definition we use \( p \) to denote the vector components \( (p_1, \ldots, p_k) \) which is a subset of the vector components included in \( e = (p_1, u_1, \ldots, p_k, u_k) \). The same denotation applies to the vector \( u \).
Theorem 3. Let $(T, P^*, V^*)$ be a general decision frame. The horizontal centroid $h_c$ of $h$, where $h$ is defined as in Definition 18, is the sum of the centroid products $f_{i_c} \cdot g_{i_c}$, i.e. the horizontal centroid is additive and multiplicative.

Proof. Let $F_c = (p_{i_1}, p_{i_2}, \ldots, p_{i_k}) \in B_P$ and let $G_c = (u_{i_1}, u_{i_2}, \ldots, u_{i_k}) \in B_V$. Then we have

$$\int_{B_P} p_i F(p) dV(p_1) dV(p_2), \ldots, dV(p_k) = p_{i_c}$$

and analogously

$$\int_{B_V} u_i G(u) dV(u_1) dV(u_2), \ldots, dV(u_k) = u_{i_c}.$$

Since $F$ and $G$ are independent

$$\int_{B_P U} p_i F(p) \cdot u_i G(u) dV_B dV_B = p_i \cdot u_i.$$

Consequently $h_c = \sum_{i=1}^k f_{i_c} g_{i_c}$. □

This makes the centroid attractive for evaluation purposes. The centroid of the expected utility is composed from the component centroids in a straightforward manner.

3.5. Example of belief in expected utility

The following illustrative example deals with an expected utility calculation where $\Theta = \{\theta_1, \theta_2\}$, i.e. there are only two uncertain outcomes. This example is provided to give a geometrical understanding of belief in a decision situation. The belief is uniformly distributed over all possible probability and utility distributions. In this case we have two two-dimensional unit cubes, $B_V = (u_1, u_2)$ and $B_P = (p_1, p_2)$. Over $B_V$ the belief distribution is $G(u) = 1$ since the belief is uniformly distributed. Regarding $B_P$, the belief distribution over the surface $p_1 + p_2 = 1$ is $F(p) = \frac{1}{\sqrt{2}}$ (this surface is a line with length $\sqrt{2}$). Through the transformation $(p_1, p_2) = P(p) = \left(\frac{p}{\sqrt{2}}, \frac{\sqrt{2} - p}{\sqrt{2}}\right)$, $0 \leq p \leq \sqrt{2}$, $B_P$ may be replaced with the line segment $[0, \sqrt{2}]$.

Consider the three-dimensional space obtained by $B_V \times [0, \sqrt{2}]$. In this space, each vector of points $(p, u_1, u_2)$ now represents a possible expected utility. Given any point in this space there is an expected utility $z \in [0, 1]$ such that $\frac{p}{\sqrt{2}} u_1 + \frac{\sqrt{2} - p}{\sqrt{2}} u_2 = z$. The belief in a given $z$ is then obtained by summing up all beliefs of the vectors in the space $[0, 1]^2 \times [0, \sqrt{2}]$ which satisfy $z = \frac{p}{\sqrt{2}} u_1 + \frac{\sqrt{2} - p}{\sqrt{2}} u_2$ (Figs. 3 and 4).

According to Definition 18, the belief in a particular $z$ is derived from the area of each such surface when the component distributions are uniform. For $z = 0$ and $z = 1$, the surface will have an area of zero, and the cumulative belief distribution of $z$ is given by the volume bounded by the surface $S_z$ and the planes where $p = 0, u_1 = 0, u_2 = 0, p = \sqrt{2}, u_1 = 1, u_2 = 1$ weighted by $F(p) G(u) = \frac{1}{\sqrt{2}}$. It can be shown that this cumulative distribution $H_2(z)$ is
\[ H_2(z) = z + (z - 1)^2 \ln(1 - z) - z^2 \ln(z). \]

Hence

\[ h_2(z) = \frac{d}{dz} H_2(z) = 2(z - 1) \ln(1 - z) - 2z \ln(z) \]

and

\[ \int_0^1 2(z - 1) \ln(1 - z) - 2z \ln(z) dz = 1. \]
As can be seen, the belief is more concentrated around the centroid, i.e. the distribution is concave. The belief in expected utility becomes centroid-focused compared to the component-wise initial beliefs in probability and utility (Fig. 5).

3.6. Decision rules

Making a decision based on a decision rule is the third step in a belief evaluation. It corresponds to applying a decision rule such as $\Gamma$-maximin in a first-order evaluation. To illustrate the need for a new decision rule within this framework, we will consider the following decision situation under risk. In the decision matrix below, $p_{ij}$ corresponds to $P(h_j|A_i)$ and $u_{ij}$ corresponds to $u(h_j|A_i)$ (Fig. 6).

To obtain a weak preference order on the three alternatives, one proposed candidate is the $\Gamma$-maximin decision rule. Applying $\Gamma$-maximin will give us the order $A_3 \succ A_2 \succ A_1$. Thus, the decision-maker is obliged to choose $A_3$ if accepting $\Gamma$-maximin.

Let $z_1$, $z_2$, and $z_3$ be variables representing possible expected utilities of $A_1$, $A_2$, and $A_3$, respectively, and let $h_1(z_1)$, $h_2(z_2)$, and $h_3(z_3)$ denote the belief distributions on the possible expected utilities of each alternative. Through Monte Carlo simulations\(^5\) we obtain the shapes of the belief distributions on the possible expected utilities. From Theorem 3, it is clear that the horizontal centroid is additive, yielding

\(^5\) Simulations performed with the software Crystal Ball, 300,000 trials.

As can be seen, the belief is more concentrated around the centroid, i.e. the distribution is concave. The belief in expected utility becomes centroid-focused compared to the component-wise initial beliefs in probability and utility (Fig. 5).

3.6. Decision rules

Making a decision based on a decision rule is the third step in a belief evaluation. It corresponds to applying a decision rule such as $\Gamma$-maximin in a first-order evaluation. To illustrate the need for a new decision rule within this framework, we will consider the following decision situation under risk. In the decision matrix below, $p_{ij}$ corresponds to $P(h_j|A_i)$ and $u_{ij}$ corresponds to $u(h_j|A_i)$ (Fig. 6).

To obtain a weak preference order on the three alternatives, one proposed candidate is the $\Gamma$-maximin decision rule. Applying $\Gamma$-maximin will give us the order $A_3 \succ A_2 \succ A_1$. Thus, the decision-maker is obliged to choose $A_3$ if accepting $\Gamma$-maximin.

Let $z_1$, $z_2$, and $z_3$ be variables representing possible expected utilities of $A_1$, $A_2$, and $A_3$, respectively, and let $h_1(z_1)$, $h_2(z_2)$, and $h_3(z_3)$ denote the belief distributions on the possible expected utilities of each alternative. Through Monte Carlo simulations\(^5\) we obtain the shapes of the belief distributions on the possible expected utilities. From Theorem 3, it is clear that the horizontal centroid is additive, yielding

\(^5\) Simulations performed with the software Crystal Ball, 300,000 trials.
\[ h_1 = 0.525 \times 75 + 0.475 \times 50 = 63.125, \]
\[ h_2 = 0.35 \times 75 + 0.65 \times 50 = 58.75, \]
\[ h_3 = 0.4 \times 75 + 0.6 \times 50 = 60. \]

As can be seen in Figs. 7–9 below, when basing a decision on lower bounds (which is the case in the \( \Gamma \)-maximin decision rule), the decision is based on points, which do not represent the decision-maker's implicit belief in an adequate manner. Furthermore, since the belief distributions over the expected utility intervals are not uniform, even in cases where

![Fig. 7. Belief distribution on \( z_1 \).](image1)

![Fig. 8. Belief distribution on \( z_2 \).](image2)

![Fig. 9. Belief distribution on \( z_3 \).](image3)
we started with uniform component-wise belief, a decision rule should take account of the fact that belief, in the cases discussed in this paper, tend to concentrate in sub-intervals containing the centroid.

In the light of the discussion above, basing a decision only on extreme points such as lower bounds seems to be too conservative and unnecessarily pessimistic. If a weak preference order is desired (avoiding situations like, e.g., indecision or incomparability) the following decision rule based on a generalized expected utility is suggested, which basically can be described as the expected utility vector at the centroid of a belief distribution $h$ as defined in Definition 18.

**Definition 19.** Let $(T, P^*, V^*)$ be a general decision frame. The principle of maximizing the generalized expected utility is accepted if a decision-making agent chooses the alternative $A^*$ from $(T, P^*, V^*)$ whenever

$$A^* = \arg \max_A (G(A))$$

for all $A \in \mathcal{A}$, where $G(A) = h_c$ as defined in Definition 18.

This decision rule is a good candidate when the belief distributions have a major part of the belief mass concentrated to some neighbourhood of each centroid. The variance of the belief is a measure of appropriateness, where a high variance indicates the decision rule being less appropriate. Furthermore, a centroid analysis is very attractive from a computational viewpoint due to the properties of the centroid. Note that applying this rule yields the preference order $A_1 \succ A_3 \succ A_2$ in the example given in this section. Needless to say, the single-centroid decision rule may not be appropriate to employ in cases where the belief distribution on the expected utility interval is convex, but due to the warp and summation effects, such situations are infrequent.

Compare the centroid based decision rule in Definition 19 to the pointwise rule in Definition 10 and the lower bound rule in Definition 14. The pointwise rule is unable to handle imprecision at all. While the lower bound rule tries to handle first-order imprecision, it is a very pessimistic rule. The centroid rule is a more balanced rule, handling first-order as well as second-order imprecision.

A further use of the centroid is in contraction analysis, a generalisation of first-order interval analysis in which the intervals are contracted towards a focal point by a procedure allowing points closer to the focal point to be emphasized in the evaluation [13]. The focal point is the single-point representative of each interval, and by using the centroid as the focal point, second-order information is used to enhance the analysis.

As can be seen from the results above, in general, the effects are considerable when evaluating imprecise decision problems. Inevitably, in a distribution analysis the most important sub-intervals to consider are the supports of the distributions where the most mass is concentrated. This can be compared to the ordinary (first-order) multiplication of extreme points (bounds) which would generate an interval without any more information. Consequently, an important component in a distributional method for decision tree analysis is the possibility of determining belief-dense sub-intervals.

These observations do not imply that the extreme points (interval boundaries) themselves are wrong, only that interval methods in general are unable to use all information available. This does neither imply that algorithms for determining upper and lower bounds in trees are inappropriate, but the results are incomplete and should be supplemented by
second-order calculations. While a complete use of second-order information is somewhat complicated, the centroid is a very good candidate for practical purposes, either stand-alone or in combination with contractions.

4. Conclusion

There are several reasons to look at second-order approaches for the efficient use of information in interval decision analysis. The key idea is to use the information available in efficient evaluation of decision structures. In this paper, we show effects of employing the principle of maximizing the expected utility using interval estimates and belief distributions in decision trees. Using only interval estimates often does not provide enough discrimination power to generate a preference order among alternatives considered. We have demonstrated that second-order belief adds decision information when handling aggregations of interval representations, such as in decision trees or probabilistic networks, and that interval estimates (upper and lower bounds) in themselves are incomplete. The handling of second-order information can be amended to a first-order tree model such as the one suggested in this paper where a traditional decision frame is enlarged and the extended model is handled by a set of well-defined concepts. Interpreting probabilities as general weights, the results also apply to other weighted structures such as multi-criteria weight trees.

The results apply equally well to approaches, which do not explicitly deal with belief distributions. Focusing only on first-order concepts in evaluating decision situations does not provide the complete picture. Second-order effects are still present regardless of the expressed or implied beliefs of the decision-maker. The rationale behind this fact is the demonstration that multiplied distributions warp significantly compared to their component distributions. The multiplied (global belief) distributions concentrate their masses towards the lower values compared to their component distributions. Furthermore, the sum of utilities weighted by global belief concentrates around the centroids of the linear combination of the centroid components.

There are essentially three ways of handling the effects shown in this paper. The first way (centroid analysis) is to use the centroid as the best single-point representative of the distributions. The centroid is additive and multiplicative. Thus, the centroid of the distribution of expected utility is the expected utility of the centroids of the projections. A centroid analysis gives a good overview of a decision situation. The second way (contraction analysis) is to use the centroid as the focal point (contraction point) towards which the intervals are decreased while studying the overlap in first-order expected utility intervals. The third way (distribution analysis) is more elaborated, involving the analysis of the resulting distributions of expected utility and calculating the fraction of belief overlapping between alternatives being evaluated.

References