Distributions over Expected Utilities in Decision Analysis

Love Ekenberg
Dept. of Computer
and Systems Sciences
Stockholm University and
Royal Institute of Technology
lovek@dsv.su.se
Address: Forum 100
SE-164 40 Kista, Sweden

Mikael Andersson
Dept. of Mathematical
Statistics
Stockholm University
mikaela@math.su.se

Mats Danielson
Dept. of Computer
and Systems Sciences
Stockholm University
mad@dsv.su.se

Aron Larsson
Dept. of Information
Technology and Media
Mid Sweden University
aron.larsson@miun.se

Abstract

It is often recognised that in real-life decision situations, classical utility theory puts too strong requirements on the decision-maker. Various interval approaches for decision making have therefore been developed and these have been reasonably successful. However, a problem that sometimes appears in real-life situations is that the result of an evaluation still has an uncertainty about which alternative is to prefer. This is due to expected utility overlaps rendering discrimination more difficult. In this article we discuss how adding second-order information may 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.

Keywords. Decision analysis, Imprecise probabilities, Imprecise utilities, Hierarchical models.

1 Introduction

In classical types of utility theories, a widespread opinion is that utility theory captures the concept of rationality. However, the shortcomings of this standpoint are sometimes severe. Among other things, the question has been raised whether people are capable of providing the inputs that utility theory requires, when, for instance, most people cannot clearly distinguish between probabilities ranging over substantial intervals. Similar problems arise in the case of artificial decision-makers, since utility-based artificial agents usually base their reasoning on human assessments, for instance in the form of induced preference functions. Furthermore, even if a decision-maker is able to discriminate between different probabilities, very often complete, adequate, and precise information is missing.

Thus, the requirement to provide numerically precise information in such models has often been considered unrealistic for real-life decision situations and after quite intense activities in the area, particularly during recent years, a number of models with representations allowing imprecise probability statements have been suggested. Such models include possibility theory [4], capacities (of order 1 and 2) [3], [13], [5], evidence theory and belief functions [19], various kinds of logic [22], upper and lower probabilities [7], hierarchical models [21], [10], and sets of probability measures [15]. Some general approaches to evaluating imprecise decision situations include probabilities and utilities. [16] is an early example and more recently some other interesting approaches have been suggested, e.g., [17], [14], [1], [6], and [2].

2 Decision Trees

In this paper, we let an *information frame* represent a decision problem. The idea with such a frame is to collect all information necessary for the model into one structure. The representational issues are of two kinds, structure (trees) and constraints (statements).

Decisions under risk (probabilistic decisions) are often given a tree representation, cf. [18]. 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 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 tree is a connected graph without cycles. A decision tree is a tree containing a finite set of nodes and that has a dedicated node 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 $i+1$ that is adjacent to a node at level i is a child of the latter. A node at level 1 is an* alternative. *A node at level i is a leaf or* consequence *if it has no adjacent nodes at level $i+1$. A node that is at level 2 or more and has children is an* event *(an intermediary node). The depth of a rooted tree is* $\max(n|\textit{there exists a node at level } n)$.

Thus, a decision tree is a way of modelling a decision situation where the alternatives are nodes at level 1 and the set of final consequences are the set of nodes without children. Intermediary nodes are called events. For convenience we can, for instance, use the notation that the n children of a node c_i are denoted $c_{i1}, c_{i2}, \ldots, c_{in}$ and the m children of the node c_{ij} are denoted $c_{ij1}, c_{ij2}, \ldots, c_{ijm}$, etc.

Figure 1 shows a decision tree. Over the sets of events and consequences, different functions can be defined, such as probability measures and utility functions.

3 Intervals in Decision Making

For numerically imprecise decision situations, one option is to define probability and utility functions in the classical way. Another, more elaborate option is to define sets of candidates of possible probability and utility functions. For instance, in [7] such an approach is suggested. The possible functions are expressed as vectors in polytopes that are solution sets to, so called, *probability* and *utility bases* (see below).

For instance, the probability (or utility) of c_{ij} being between the numbers a_k and b_k is expressed as $p_{ij} \in [a_k, b_k]$ ($u_{ij} \in [a_k, b_k]$). This approach also includes relations: a measure (or function) of c_{ij} is greater than a measure (or function) of c_{kl} is expressed as $p_{ij} \geq p_{kl}$ and analogously $u_{ij} \geq u_{kl}$. Each statement can thus be represented by one or more constraints.

Definition 3. *Given a decision tree D, a utility base is a set of linear constraints of the types $u_{ij} \in [a_k, b_k]$, $u_{ij} \geq u_{kl}$ and, for all consequences $\{c_{ij}\}$ in D, $u_{ij} \in [0,1]$. A probability base has the same structure, but, for all nodes N (except the root node) in D, also includes $\sum_{j=1}^{m_i} p_{ij} = 1$ for the children $\{c_{ij}\}_{j=1,\ldots,m_i}$ of N.*

Since a vector in the polytope can be considered to represent a distribution, a probability base $\mathcal{P}$ can be interpreted as constraints defining the set of all possible probability measures over the consequences. Similarly, a utility base $\mathcal{U}$ consists of constraints defining the set of all possible utility functions over the consequences. The bases $\mathcal{P}$ and $\mathcal{U}$ together with the decision tree constitute the *information frame*.

Primary evaluation rules of a decision tree model are based on the expected utility. Since neither probabilities nor utilities are fixed numbers, the evaluation of the expected utility yields multi-linear expressions.

Definition 4. *Given a decision tree T and an alternative $A_i \in A$ the expression*

$$\mathrm{E}(A_i) = \sum_{i_1=1}^{n_{i_0}} p_{ii_1} \sum_{i_2=1}^{n_{i_1}} p_{ii_1i_2} \cdots \sum_{i_{m-1}=1}^{n_{i_{m-2}}} p_{ii_1i_2\ldots i_{m-2}i_{m-1}} \sum_{i_m=1}^{n_{i_{m-1}}} p_{ii_1i_2\ldots i_{m-2}i_{m-1}i_m} u_{ii_1i_2\ldots i_{m-2}i_{m-1}i_m}$$

where m is the depth of the tree corresponding to A_i, n_{i_k} is the number of possible outcomes following the event with probability p_{i_k}, $p_{\ldots i_j \ldots}$, $j \in [1, \ldots, m]$, denote probability variables and $u_{\ldots i_j \ldots}$ denote utility variables as above, is the expected utility of alternative A_i in T.

Maximisation of such non-linear objective functions subject to linear constraint sets (statements on probability and utility variables) are computationally demanding problems to solve for an interactive decision tool in the general case, using techniques from the area of non-linear programming. In, e.g., [7], [8], and [6], there are discussions about computational procedures reducing the evaluation of non-linear decision problems to systems with linear objective functions, solvable with ordinary linear programming methods. The approach taken is to model probability and utility intervals as constraint sets, containing statements on upper and lower bounds. Furthermore, normalisation constraints for the probabilities are added (representing that the consequences from a parent node are exhaustive and pairwise disjoint). Such constraints are always on the form $\sum_{j=1}^{n} p_{ij} = 1$.

The solution sets to probability and utility constraint sets are polytopes. The evaluation procedures then yield first-order interval estimates of the evaluations, i.e. upper and lower bounds for the expected utilities of the alternatives.

An advantage of approaches using upper and lower probabilities is that they do not require taking particular probability distributions into consideration. On the other hand, the expected utility range resulting from an evaluation is also an interval. To our experience, in real-life decision situations, it is then sometimes hard to discriminate between the alternatives. In effect, an interval based decision procedure keeps all alternatives with overlapping expected utility intervals, even if the overlap is quite small. Therefore, it is interesting to extend the representation of the decision situation using more information, such as distributions over classes of probability and utility measures, in pursuit of more discriminative power.

4 Including Second-Order Information

Basically, distributions have been used for expressing various beliefs over multi-dimensional spaces where each dimension corresponds to, for instance, possible probabilities or utilities of consequences. The distributions can consequently be used to express strengths of beliefs in different vectors in the polytopes.

Beliefs of such kinds are expressed using higher-order distributions (hierarchical models). Approaches for extending the interval representation using distributions over classes of probability and value measures have been developed into various hierarchical models, such as second-order probability theory. A quite early approach was suggested in [11] and [12]. A more recent example is [20] that provides a model for one-level trees similar to [9].

In the following, we will pursue the idea of adding more information and discuss some interesting properties that appear when evaluating second-order models as well as the effects of aggregating such distributions over expected utilities. The main conclusion here is that the actual deep and breadth of the decision tree under consideration is of large importance for the interpretation of the result. We will also see that the detailed shapes of the distributions are not utterly important compared with this and approximates are sufficient.

4.1 Distributions over Information Frames

Interval estimates can be considered as special cases of representations based on distributions over polytopes. For instance, a distribution can be defined to have a positive support only for $x_i \leq x_j$. More formally, the solution set to a probability or utility constraint set is a subset of a unit cube since both variable sets have $[0, 1]$ as their ranges. This subset can be represented by the support of a distribution over the cube.

Definition 5. *Let a unit cube be represented by $B = (b_1, \ldots, b_n)$. The b_i can be explicitly written out to make the labelling of the dimensions clearer. (More rigorously, the unit cube should be represented by all the tuples $(x_1, \ldots, x_n)$ in $[0, 1]^n$.)*

Definition 6. *By a second-order distribution over B, we denote a positive distribution F defined on the unit cube B such that*

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

where V_B is the n-dimensional Lebesque measure on B. The set of all second-order distributions over B is denoted by $BD(B)$.

For our purposes here, second-order *probabilities* are an important sub-class of these distributions and will be used below as a measure of belief, i.e. a second-order joint probability distribution. Marginal distributions are obtained from the joint ones in the usual way.

Definition 7. *Let a unit cube $B = (b_1, \ldots, b_n)$ and $F \in BD(B)$ be given. Furthermore, let $B_i^- = (b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_n)$. Then*

$$f_i(x_i) = \int_{B_i^-} F(x)\, dV_{B_i^-}(x)$$

is a marginal distribution over the axis b_i.

Such distributions can then straightforwardly be defined over the information frames. However, regardless of the actual shapes of the distributions involved, constraints such as $\sum_{i=1}^n x_i = 1$ must be satisfied since it is not reasonable to believe in an inconsistent point such as $(0.15, 0.25, 0.4, 0.3)$ if the vector is supposed to represent a probability distribution over four mutually exclusive outcomes. Therefore, a convenient and general way of modelling random weights in $[0, 1]$ is the *Dirichlet distribution.*

Definition 8. *Let the notation be as above. Then the probability density function of the Dirichlet distribution is defined as*

$$f_{Dir}(p, \alpha) = \frac{\Gamma(\sum_{i=1}^n \alpha_i)}{\prod_{i=1}^n \Gamma(\alpha_i)} p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} \cdots p_n^{\alpha_n - 1}$$

on a set $\{p = (p_1, \ldots, p_n) \mid p_1, p_2, \ldots, p_n \geq 0, \sum p_i = 1\}$, where $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is a parameter vector in which each α_i is a positive parameter and $\Gamma(\alpha_i)$ is the Gamma function.

This distribution is particularly popular among Bayesian statisticians because it is conjugate with respect to the multinomial distribution, i.e. if we choose the prior to be the Dirichlet distribution then the posterior will also become Dirichlet. It is also convenient in the sense that it is not hard to choose parameters to reflect our prior knowledge about the weights $p_1, p_2, \ldots, p_n$. If we choose large values for $\alpha_1, \alpha_2, \ldots, \alpha_n$ we obtain small variances, which reflect a large measure of certainty about the probabilities involved.

Formally, this probability density function does not fulfil our requirement for a belief distribution, but as demonstrated in [9], the issue with the dimension loss can be solved using the *the Dirac distribution*, $\delta_p(x)$, with pole at the point p.

Definition 9. Let A be a subset of a unit cube B, and let f be a belief distribution in A. The natural extension $\tilde{f}_A(x)$ of f with respect to A is defined by

$$\tilde{f}_A(x) = \begin{cases} f(x) & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

Definition 10. Let A be a subset of B. A distribution g_A over B is called a characteristic distribution for A in B if

$$f(p) = \int_B \delta_p(x)\tilde{f}_A(x)g_A(x)\,dV_B(x)$$

for every belief distribution f over A, and for every point p in A.

Now let $A = \{(p_1, \ldots, p_n) \mid \sum_{i=1}^n p_i = 1\}$ and let g_A be a Dirichlet distribution. From distribution theory follows that for every measurable subset A in a unit cube B, there exists a characteristic distribution for A in B. It also follows that $\tilde{f}_A(x) \cdot g_A(x)$ is a belief distribution over B and equals 0 outside A.

4.2 Marginal Distributions

A marginal distribution of a Dirichlet distribution is a beta distribution. For instance, if the distribution is uniform, the resulting marginal distribution (over an axis) is a polynomial of degree $n-2$, where n is the dimension of a cube B: let $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 1$. Then the Dirichlet distribution is uniform and the marginal distribution is

$$f(x_i) = \int\limits_{B_i^-} dV_{B_i^-}(x) = (n-1)(1-x_i)^{n-2}\,.$$

Example 1. *The marginal distribution $f(x_i)$ of the uniform Dirichlet distribution in a 4-dimensional cube is*

$$f(x_i) = \int\limits_0^{1-x_i}\int\limits_0^{1-y-x_i} 6\,dz\,dy = 3(1-2x_i+x_i^2)$$

$$= 3(1-x_i)^2\,.$$

This tendency is the result of a general phenomenon that becomes more emphasised as the dimension increases. As it will be discussed in the next section, this observation of marginal probabilities is important for the analysis of expected values in decision trees and similar structures.

4.3 The Expected Value and its Variance

Consider a decision tree with only one level of events and n alternatives. Let p_i denote probabilities and u_i utilities of the consequences of an alternative A_j. We assume that $u_1, u_2, \ldots, u_n$ can be considered as independent random variables and we denote the mean and the variance of u_i by μ_i and σ_i^2, respectively. We also assume that $p_1, p_2, \ldots, p_n$ are random variables in the interval $[0,1]$ satisfying the condition $\sum_i p_i = 1$.

Using the Dirichlet distribution, the expected value of $\sum_{i=1}^n p_i u_i$ can be calculated straightforwardly. Let y below represent the (uncertain) expected utility of the alternative A_j such that $y = \sum_{i=1}^n p_i u_i$. Then

$$\mathrm{E}(y) = \mathrm{E}\left(\sum_{i=1}^n p_i u_i\right) = \sum_{i=1}^n \mathrm{E}(p_i)\mathrm{E}(u_i) = \sum_{i=1}^n \frac{\alpha_i}{\alpha}\mu_i$$

When calculating the variance, we have to take the dependence of the p_i-variables into account.

We use the convenient formula

$$\mathrm{Var}(y) = \mathrm{E}(y^2) - \mathrm{E}(y)^2$$

where

$$\begin{aligned}
\mathrm{E}(y^2) &= \mathrm{E}\left(\left(\sum_{i=1}^n p_i u_i\right)^2\right) \\
&= \mathrm{E}\left(\sum_{i=1}^n p_i^2 u_i^2\right) + 2\mathrm{E}\left(\sum_{i<j} p_i p_j u_i u_j\right) \\
&= \sum_{i=1}^n \mathrm{E}(p_i^2)\mathrm{E}(u_i^2) + 2\sum_{i<j}\mathrm{E}(p_i p_j)\mathrm{E}(u_i)\mathrm{E}(u_j) \\
&= \sum_{i=1}^n (\mathrm{E}(p_i)^2 + \mathrm{Var}(p_i))(\mathrm{E}(u_i)^2 + \mathrm{Var}(u_i)) \\
&+2\sum_{i<j}(\mathrm{E}(p_i)\mathrm{E}(p_j) + \mathrm{Cov}(p_i,p_j))\mu_i\mu_j \\
&= \sum_{i=1}^n \left(\frac{\alpha_i^2}{\alpha^2} + \frac{\alpha_i(\alpha-\alpha_i)}{\alpha^2(\alpha+1)}\right)(\mu_i^2+\sigma_i^2) \\
&+2\sum_{i<j}\left(\frac{\alpha_i\alpha_j}{\alpha^2} - \frac{\alpha_i\alpha_j}{\alpha^2(\alpha+1)}\right)\mu_i\mu_j \\
&= \sum_{i=1}^n \frac{\alpha_i(\alpha_i+1)}{\alpha(\alpha+1)}(\mu_i^2+\sigma_i^2) + 2\sum_{i<j}\frac{\alpha_i\alpha_j}{\alpha(\alpha+1)}\mu_i\mu_j
\end{aligned}$$

where $\alpha = \sum_i \alpha_i$, and

$$\mathrm{E}(y)^2 = \left(\sum_{i=1}^n \frac{\alpha_i}{\alpha}\mu_i\right)^2 = \sum_{i=1}^n \frac{\alpha_i^2}{\alpha^2}\mu_i^2 + 2\sum_{i<j}\frac{\alpha_i\alpha_j}{\alpha^2}\mu_i\mu_j$$

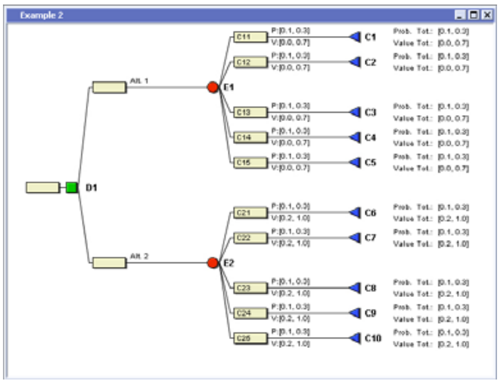

Figure 1: The decision tree in Example 2.

Combining these results yields the variance

$$\mathrm{Var}(y) = \frac{1}{\alpha^2(\alpha+1)}\left(\sum_{i=1}^{n}\alpha_i((\alpha-\alpha_i)\mu_i^2+\alpha(\alpha_i+1)\sigma_i^2)-2\sum_{i<j}\alpha_i\alpha_j\mu_i\mu_j\right)$$

For the uniform case, we obtain

$$\mathrm{E}(y) = \sum_{i=1}^{n}\frac{1}{n}\mu_i = \bar{\mu}$$

and

$$\mathrm{Var}(y) = \frac{1}{n^2(n+1)}\left(\sum_{i=1}^{n}((n-1)\mu_i^2+2n\sigma_i^2)-2\sum_{i<j}\mu_i\mu_j\right)$$

Example 2. *Let an information frame contain a decision tree with two alternatives* A_1 *and* A_2. *Assume that each have five consequences* $C_{i1},\ldots,C_{i5}$ *with probabilities* $p_{ij}\in[0.1,0.3], j=1,\ldots,5, i=1,2$ *and with utilities* $u_{1j}\in[0,0.7], j=1,\ldots,5, u_{2j}\in[0.2,1], j=1,\ldots,5$. *This tree is shown in Figure 1. An interval analysis yields* $\mathrm{E}(A_1)\in[0,0.7]$ *and* $\mathrm{E}(A_2)\in[0.2,1]$. *The major overlap between the two alternatives' expected utility intervals,* $[0.2,0.7]$, *makes it difficult to supply the decision-maker with any advice. If, for example, the distributions over the information frame are uniform, we can see that the distribution of mass over the expected utility clearly discriminates the alternatives. The expected values are* 0.35 *and* 0.6 *and the variances are around* 0.015. *Furthermore, in Figure 2 and Figure 3 the alternatives are entirely separated already for* 75% *of the belief mass (the darker areas). A comparison of the two alternatives is further demonstrated in Figure 4, showing the distribution over the difference* $\mathrm{E}(A_1)-\mathrm{E}(A_2)$.

If we do not know any specifics of the underlying distributions, we can utilise *Chebyshev's inequality* which

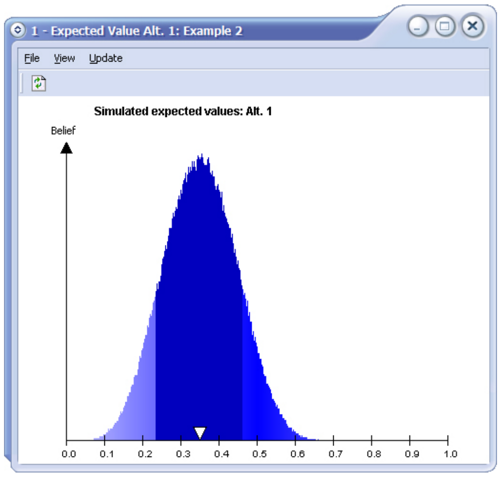

Figure 2: Distribution over $\mathrm{E}(A_1)$ in Example 2.

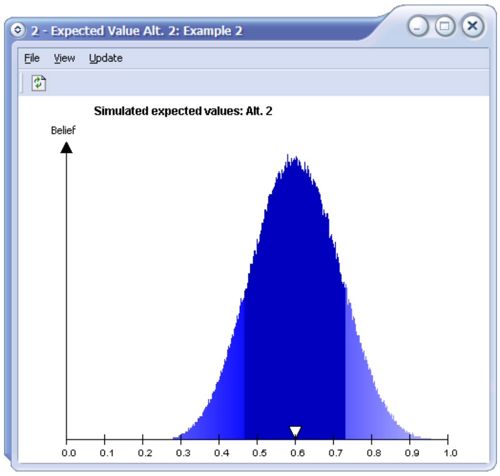

Figure 3: Distribution over $\mathrm{E}(A_2)$ in Example 2.

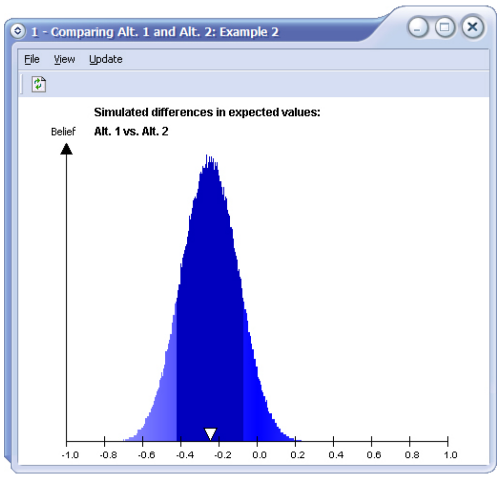

Figure 4: Distribution over $\mathrm{E}(A_1)-\mathrm{E}(A_2)$ in Example 2.

can be formulated in a number of different ways depending on the application. The most common and useful version is

$$\mathrm{P}(|X-\mu| > c\sigma) \leq \frac{1}{c^2}$$

where X is a random variable with mean μ and standard deviation σ and c is an arbitrary constant. For instance, if we want to determine a symmetric 95 % interval around μ, we choose $c = \sqrt{20} = 4.47$. For many classical distributions, this approximation is unfortunately quite rough, even if it is possible to find distributions where equality is attained. For instance, the normal distribution satisfies $\mathrm{P}(|X-\mu| > 1.96\sigma) = 0.05$, which yields an interval being less than half as wide as the Chebyshev approximation.

In any case, it should be noted that we can add information to the decision tree by utilising second-order information. Moreover, the distributions resulting from multiplications generally have shapes very different from their marginal components and we will further investigate this effect below. As will be seen, this has some implications for trees deeper than one level.

4.4 Aggregations

The characteristic of a decision tree is that the marginal (or conditional) probabilities of the event nodes are multiplied in order to obtain the joint probability of a combined event, i.e. of a path from the root to a leaf. In the evaluation of a decision tree the operations involved are multiplications and additions. There are therefore two effects present at the same time when calculating expected utilities in decision trees. Those are additive effects (for joint probabilities aggregated together with the utilities at the leaf nodes) and multiplicative effects (for intermediate probabilities).

One important effect is that multiplied distributions become considerably warped compared to the corresponding component distributions. Such multiplications occur in obtaining the expected utility in decision trees and probabilistic networks, enabling discrimination while still allowing overlap. Properties of additions of components follow from ordinary convolution, i.e. there is a strong tendency towards the middle.

We will now investigate the combined effect and consider how to put second-order information into use to further discriminate between alternatives. The main idea is not to require a total lack of overlap but rather allowing overlap by interval parts carrying little belief mass, i.e. representing a very small part of the decision-maker's belief. Then, the non-overlapping parts can be thought of as being the core of the decision-maker's appreciation of the decision situation, thus allowing discrimination. In addition, effects from varying belief (i.e. differing forms of belief distribution) should be taken into account.

Evaluations of expected utilities in trees lead to multiplication of probabilities using a type of "multiplicative convolution" of two densities.

Let G be a distribution over the two cubes A and B. Assume that G has a positive support on the feasible probability distributions at level i in a decision tree, i.e. is representing these (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 $f(x)$ and $g(y)$ be the marginal distributions of $G(z)$ on A and B, respectively.

Definition 11. *The cumulative distribution of the two belief distributions $f(x)$ and $g(y)$ is*

$$H(z) = \iint_{\Gamma_z} f(x)g(y)\,dx\,dy = \int_0^1 \int_0^{z/x} f(x)g(y)\,dy\,dx = \int_0^1 f(x)G(z/x)\,dx = \int_z^1 f(x)G(z/x)\,dx\,,$$

where G is a primitive function to g, $\Gamma_z = \{(x,y) \mid x \cdot y \leq z\}$, and $0 \leq z \leq 1$.

Let $h(z)$ be the corresponding density function. Then

$$h(z) = \frac{d}{dz}\int_z^1 f(x)G(z/x)\,dx = \int_z^1 \frac{f(x)g(z/x)}{x}\,dx\,.$$

The addition of such products is analogous to the product rule for standard probabilities and we can use the ordinary convolution of two densities restricted to the cubes. The distribution h on a sum $z = x + y$ of two independent variables associated with belief distributions $f(x)$ and $g(y)$ is therefore given by

$$h(z) = \int_0^z f(x)g(z-x)\,dx\,.$$

Example 3. *Consider an information frame containing an alternative A_1 with depth 3 and with 3 consequences at each event node. Let $p_{1i} \in [0,1], p_{1ij} \in$*

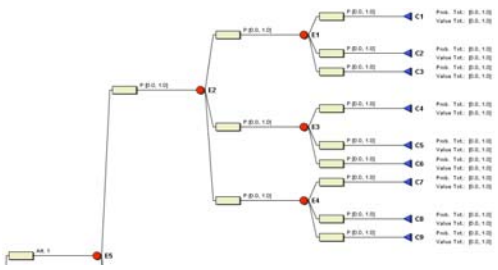

Figure 5: The upper one third of the decision tree in Example 3.

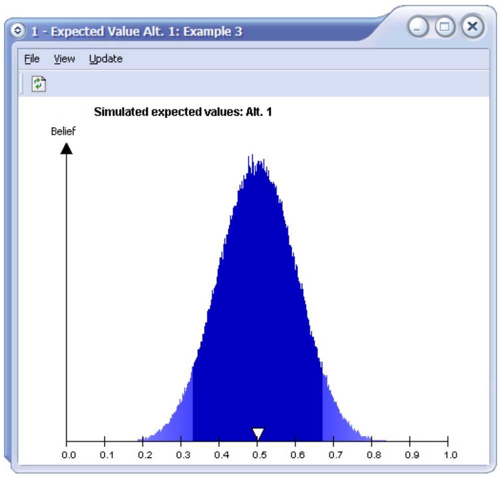

Figure 6: Distribution over $\mathrm{E}(A_1)$ in Example 3.

$[0,1], p_{1ijk} \in [0,1], u_{1ijk} \in [0,1]$*, and* $i,j,k \in \{1,2,3\}$. *This means that no numerical information, except for the trivial constraints, is provided. Part of the tree is shown in Figure 5. Looking at the upper and lower bounds for the expected utility, we find that* $\mathrm{E}(A_1) \in [0,1]$. *If, for example, the second-order distributions over the information frame are uniform, we find that the resulting distribution from each path is* $-4(-12+12z-6\ln(z)-6z\ln(z)-\ln(z)^2+z\ln(z)^2)$ *and that, e.g., 90% of the mass is over the interval* $[0.33, 0.67]$*, see Figure 6.*

As can be seen, second-order data, for instance in terms of Dirichlet distributions, may provide important information in decision evaluation. The example above is taking a particular distribution into account, but as in the previous discussion, these results apply for all types of distributions.

5 Summary and Conclusions

In classic decision theory it is assumed that a decision-maker can assign precise numerical values corresponding to the true value of each consequence, as well as precise numerical probabilities for their occurrences. In attempting to address real-life problems, where uncertainty in the input data prevails, some kind of representation of imprecise information is important and several have been proposed. In particular, representations such as sets of probability measures, upper and lower probabilities, and interval probabilities and utilities of various kinds have been perceived as enabling a better representation of the input sentences for a subsequent decision analysis. However, higher-order analysis can sometimes add important information to the analysis, enabling further discrimination between alternatives.

In this paper, we have discussed the effects of employing second-order information in decision trees. As was seen from Definition 11, the multiplicative effects on probabilities in decision trees increase with tree depth. We have also shown that the multiplicative and additive effects strongly influence the resulting distribution over the expected values.

These effects combined yield a method that sometimes can offer more discriminative power in selecting alternatives in decision trees. The main idea of the method is to allow a small overlap where the belief mass is kept under control. While the discussion focuses on probabilistic decision trees, the results also apply to other formalisms involving products of probabilities, such as probabilistic networks, and to formalisms dealing with other products of interval entities such as interval weight trees in hierarchical multi-criteria decision models.

References

[1] T. Augustin. On Decision Making under Ambiguous Prior and Sampling Information. *Proceedings of ISIPTA'01*, 2001.

[2] T. Augustin. On the Suboptimality of the Generalized Bayes Rule and Robust Bayesian Procedures from the Decision Theoretic Point of View - A Cautionary Note on Updating Imprecise Priors. *Proceedings of ISIPTA'03*, 2001.

[3] G. Choquet. Theory of Capacities. *Ann. Inst. Fourier*, 5: 131–295, 1953/54.

[4] G. de Cooman. Possibility Theory. *International Journal of General Systems*, 25(4): 291–371, 1997.

[5] V. Cutello and J. Montero. Fuzzy Rationality Measures. *Fuzzy Sets and Systems*, 62: 39–54, 1994.

[6] M. Danielson and L. Ekenberg. Computing Upper and Lower Bounds in Interval Decision Trees. *Eu-*

ropean Journal of Operational Research, 181(2): 808–816, 2007.

[7] M. Danielson and L. Ekenberg. A Framework for Analysing Decisions under Risk. *European Journal of Operational Research*, 104(3): 474–484, 1998.

[8] X. S. Ding, M. Danielson and L. Ekenberg. Non-linear Programming Solvers for Decision Analysis Support Systems. *Operations Research Proceedings 2003 - Selected Papers of the International Conference on Operations Research (OR 2003)*, 475–482, 2004.

[9] L. Ekenberg and J. Thorbiörnson. Second-order Decision Analysis. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(1): 13–38, 2001.

[10] L. Ekenberg, J. Thorbiörnson and T. Baidya. Value Differences using Second Order Distributions. *International Journal of Approximate Reasoning*, 38(1): 81–97, 2005.

[11] P. Gärdenfors and N.-E. Sahlin. Unreliable Probabilities, Risk Taking, and Decision Making. *Synthese*, 53: 361–386, 1982.

[12] P. Gärdenfors and N.-E. Sahlin. Decision Making with Unreliable Probabilities. *British Journal of Mathematical and Statistical Psychology* 36: 240–251, 1983.

[13] P. J. Huber. The Case of Choquet Capacities in Statistics. *Bulletin of the International Statistical Institute*, 45: 181–188, 1973.

[14] J.-Y. Jaffray. Rational Decision Making with Imprecise Probabilities. *Proceedings of ISIPTA'99*, 1999.

[15] I. Levi. *The Enterprise of Knowledge*, MIT Press, 1980.

[16] I. Levi. On Indeterminate Probabilities. *The Journal of Philosophy*, 71: 391–418, 1974.

[17] R. F. Nau. The Aggregation of Imprecise Probabilities. *Journal of Statistical Planning and Inference*, 105: 265–282, 2002.

[18] H. Raiffa. *Decision Analysis*, Addison Wesley, 1968.

[19] P. Smets. Practical Uses of Belief Functions. *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, 1999.

[20] L. Utkin and T. Augustin. Decision Making with Imprecise Second-order Probabilities. *Proceedings of ISIPTA'03*, 547–561, 2003.

[21] L. Utkin. Imprecise Second-order Hierarchical Uncertainty Model. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11(3): 301–317. 2003.

[22] N. Wilson. A Logic of Extended Probability. *Proceedings of ISIPTA'99*, 1999.