Combining Second-Order Belief Distributions with Qualitative Statements in Decision Analysis

Ola Caster and Love Ekenberg

Abstract There is often a need to allow for imprecise statements in real-world decision analysis. Joint modeling of intervals and qualitative statements as constraint sets is one important approach to solving this problem, with the advantage that both probabilities and utilities can be handled. However, a major limitation with interval-based approaches is that aggregated quantities such as expected utilities also become intervals, which often hinders efficient discrimination. The discriminative power can be increased by utilizing second-order information in the form of belief distributions, and this paper demonstrates how qualitative relations between variables can be incorporated into such a framework. The general case with arbitrary distributions is described first, and then a computationally efficient simulation algorithm is presented for a relevant sub-class of analyses. By allowing qualitative relations, our approach preserves the ability of interval-based methods to be deliberately imprecise. At the same time, the use of belief distributions allows more efficient discrimination, and it provides a semantically clear interpretation of the resulting beliefs within a probabilistic framework.

1 Introduction

It is questionable whether people are capable of providing the inputs that utility theory requires, when most people cannot clearly distinguish between widely separated probabilities Shapira (1995). This indicates that precise numerical information does

O. Caster (✉)
Uppsala Monitoring Centre, WHO Collaborating Centre for International Drug Monitoring, Uppsala, Sweden

Department of Computer and Systems Sciences, Stockholm University
e-mail: ola.caster@who-umc.org

L. Ekenberg
Department of Computer and Systems Sciences, Stockholm University, Kista, Sweden
e-mail: lovek@dsv.su.se

Y. Ermoliev et al. (eds.), *Managing Safety of Heterogeneous Systems*, Lecture Notes in Economics and Mathematical Systems 658, DOI 10.1007/978-3-642-22884-1_4,

not make much sense in real-life decision making. Furthermore, even if a decision maker is able to discriminate between different probabilities, very often complete, adequate, and precise information is missing. Hence, decision problems frequently contain far less information than classical utility theory requires. In particular, quite often we might, at best, have access to some vague probability beliefs and qualitative preferences among the consequences, and very little more than that. This is the class of decision problems we aim at in this article.

It has since long been recognized that decision theory needs to accommodate imprecise probabilities (and utilities) and a vast amount of models with representations allowing imprecise probability statements have been suggested, including possibility theory, capacity theory, evidence theory and belief functions in the Dempster-Shafer sense, various kinds of logic, upper and lower probabilities, hierarchical models and sets of probability measures. A multitude of articles have been presented on various methods. For some early examples of these, see e.g. (Choquet 1954; Dempster 1967; Dubois and Prade 1988; Ellsberg 1961; Good 1962; Shafer 1976; Smith 1961). It is interesting to note that, during recent years, the activities within the area of imprecise probabilities have increased substantially and special conferences are now dedicated to contributions on this theme. An example of this is Jaffray (1999) from the first International Symposium on Imprecise Probabilities and Their Applications (ISIPTA).

Some general approaches to evaluating imprecise decision situations include both imprecise probabilities and utilities. We have earlier discussed various aspects on these issues in a sequence of articles and argued that there are strong arguments for modeling quantitative impreciseness as intervals (and similar constraints), enabling representation and modeling of qualitative information as constraint sets of relations. Cf. Danielson and Ekenberg (1998); Danielson et al. (2009); Ding et al. (2010); Ekenberg and Thorbiörnson (2001).

An obvious 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 then also an interval. To our experience, in real-life decision situations, it is hard to discriminate between the alternatives in a pure interval approach, even if various relations are added. In effect, an interval-based decision procedure preserves all alternatives with overlapping expected utility intervals, even if the overlap is quite small. Consequently, there is a need to extend the representation of the decision situation using more information, but keeping the requirement that the decision maker does not have to be more precise than what is possible. In pursuit of more discriminative power, we have also developed methods for handling belief distributions over sub-parts of the probabilities as well as the utilities involved, see e.g. Danielson et al. (2007); Ekenberg (2000). Furthermore, we have developed a calculus for aggregating these in various ways, predominantly using a generalization of the expected utility function Ekenberg et al. (2005).

Thus, we have developed computationally meaningful methods – that have also been implemented as software – for solving multi-linear expressions with respect to constraints sets. We have also developed – and implemented – methods for handling distributions over independent variables. Various aspects of these latter methods are provided in Ekenberg et al. (2007, 2006); Sundgren et al. (2009). However, a great and embarrassing dilemma has been to find the combination of these two approaches, i.e. to use qualitative statements and distributions at the same time. Dependencies of the kind that qualitative statements give rise to are not particularly straightforward to handle in the context of belief distributions. Nevertheless, such statements arise naturally in many decision situations. For instance, usually a decision maker has access only to local information and qualitative statements of relations between different parameters, in terms of constraints, and, consequently, has no explicit idea about the overall distribution.

This article presents a computationally meaningful method for solving this dilemma. We present how to solve expected utilities of quite complex structure, considering general decision trees and belief distributions over all the probabilities and utilities involved. Furthermore, and most importantly herein, we demonstrate a method for including qualitative statements, while still preserving the possibility to use these distributions efficiently. Starting from decision trees, we first solve the general case, followed by a computationally feasible method for handling this practically in a relevant sub-class of analyses. Compared to other approaches, this merging of second-order belief distributions and qualitative statements not only improves discrimination but also provides an easier interpretation within a probabilistic framework.

2 Decision Trees

A *decision tree* represents a decision problem, collecting all information necessary for the model into one structure.

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).

Definition 2. A tree is a connected graph without cycles. A rooted tree is a tree containing a finite set of nodes and that has a dedicated node at level 0. The adjacent nodes to a node at level i, except the nodes at level $i - 1$, are 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 $\max(n|\text{there exists a node at level n})$.

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.

Definition 3. Given a rooted tree, a decision tree T is formed by assigning a p symbol to each edge not starting in the root node, and a u symbol to each consequence node.

Generally, the p and u symbols can be given any meaning, but here they will represent probabilities and utilities, respectively. As such they are all constrained to [0, 1], and further the probabilities on edges from a common parent node (not the root) must sum to 1. Such a set of probabilities will henceforth be called a *probability group*.

Primary evaluation rules of a decision tree model are based on the expected utility.

Definition 4. Given a decision tree T and an alternative A_i, the expression

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

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.

This is a general representation and one option is thus to define probability distributions and utility functions in the classical way. Another option that also covers impreciseness is to define sets of possible probability distributions and utility functions. The possible functions are then conveniently expressed as vectors in polytopes that are solution sets to the constraints involved.

A number of evaluation procedures, earlier suggested by us, then yield first-order interval estimates of the evaluations, i.e. upper and lower bounds for the expected utilities of the alternatives Danielson and Ekenberg (2007). However, the expected utility range resulting from an evaluation now also becomes an interval. In real-life decision situations, it is then often hard to discriminate between the alternatives, i.e. an interval-based decision procedure will not separate out alternatives with overlapping expected utility intervals, even if the overlap is quite small. Furthermore, a decision maker does not necessarily believe with equal faith in all the epistemologically possible probability distributions, represented by a set of interval statements. 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.

3 Belief Distributions

The idea is now that distributions can be 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 solution sets.

Approaches for extending the interval representation using distributions over classes of probability and value measures in this way have been developed into various hierarchical models, such as second-order probability theory, cf. Ekenberg et al. (2006).

In such an approach, it is possible to make use of distributions rather than intervals for expressing beliefs regarding the probabilities and utilities involved. However, since general distributions over the entire solution sets are very hard to imagine, already when handling just a few dimensions, the marginal distributions, and the relations between these, are of high importance. A more comprehensible distribution in the latter sense can straightforwardly be defined.

Definition 5. For a utility or probability variable x in a decision tree T, the continuous random variable $\tilde{X}$ is the belief distribution over x. $\tilde{X}$ is defined on $[a, b]$ with $a, b \in [0, 1]$ and $a < b$.

We will frequently use the density function $f_{\tilde{X}}(x)$ of $\tilde{X}$ to represent and visualize the belief distribution. In some cases we will be interested in the joint density function for the belief regarding a set of utilities or probabilities. Exemplifying with the joint belief distribution over the utilities $u_1, \ldots, u_n$, this density function will be denoted $f_{\tilde{U}_1,\ldots,\tilde{U}_n}(u_1, \ldots, u_n)$, or more compactly as $f_{\tilde{\mathbf{U}}}(\mathbf{u})$. [1]

3.1 Constrained Belief Distributions

Our main objective here is to extend the earlier approach to allow for comparative constraints. The type of constraints considered are linear relations between two variables, i.e. of the type $u_i \leq u_j$, which here translate to constraints for the corresponding belief distributions. These constraints, together with the specified belief distributions, make up the decision maker's perception; they are his or her statements about the decision situation.

We will allow comparative constraints between any two utilities and between any two probabilities in the same probability group[2].

Definition 6. Given a decision tree T, the total set of constraints for $\tilde{U}_1, \ldots, \tilde{U}_n$ is denoted by $A_{\mathbf{U}}$. Furthermore $B_{\mathbf{U}}$ is the corresponding subspace of $[0, 1]^n$ implied by $A_{\mathbf{U}}$. Analogously, the total set of constraints for the belief distributions $\tilde{P}_{k1}, \ldots, \tilde{P}_{kl}$ over probabilities from group k is denoted $A_{\mathbf{P}_k}$, and the corresponding l-dimensional subspace is denoted $B_{\mathbf{P}_k}$.

Note that $A_{\mathbf{P}_k}$ includes the implicit constraint $\sum_i \tilde{P}_{ki} = 1$.

The constrained belief distributions are obtained by conditioning the original belief distributions on the total set of constraints:

---

[1]Unless explicitly stated otherwise, bold face symbols denote vectors throughout this paper.

[2] It is implicit that these constraints are coherent so that, for example, if $\tilde{U}_1 < \tilde{U}_2$ and $\tilde{U}_2 < \tilde{U}_3$, then it cannot hold that $\tilde{U}_3 \leq \tilde{U}_1$.

Definition 7. The constrained belief distributions are given by

$$(U_1, \ldots, U_n)' = (\tilde{U}_1, \ldots, \tilde{U}_n)' | A_{\mathbf{U}}$$

for the utilities, and by

$$(P_{k1}, \ldots, P_{kl})' = (\tilde{P}_{k1}, \ldots, \tilde{P}_{kl})' | A_{\mathbf{P}_k}$$

for probabilities of group k.

Our real interest lies in the constrained variables, since these take into account both the originally defined belief distributions and the total set of constraints for those distributions. Because the constraints introduce dependencies, one needs to operate on the joint belief distributions. If the decision maker has some explicit beliefs concerning interdependencies, not captured by the constraint sets $A_{\mathbf{U}}$ and $A_{\mathbf{P}_k}$, joint belief distributions should be specified already from the outset. Otherwise, which should be the more common scenario, the constraint sets contain all available information on dependencies, and the marginal unconstrained belief distributions are independent of each other. This independence then allows for easy calculation of the required joint density function for the unconstrained belief distributions. Exemplifying with the utilities, one gets

$$f_{\tilde{U}_1,\ldots,\tilde{U}_n}(u_1, \ldots, u_n) = f_{\tilde{U}_1}(u_1) \cdots f_{\tilde{U}_n}(u_n) \; . \tag{1}$$

The joint density function for the constrained belief distributions is obtained by reducing the support of the original belief distributions to $B_{\mathbf{U}}$ (in the case of utilities), and scaling up the density function for all points in $B_{\mathbf{U}}$ so that the function integrates to one in its support. This is a multivariate equivalent to truncating a univariate random variable. The joint density function for $\mathbf{U} = (U_1, \ldots, U_n)'$ is therefore

$$f_{\mathbf{U}}(\mathbf{u}) = \frac{f_{\tilde{\mathbf{U}}}(\mathbf{u})}{\Pr(\tilde{\mathbf{U}} \in B_{\mathbf{U}})} = \frac{f_{\tilde{\mathbf{U}}}(\mathbf{u})}{\int \cdots \int_{B_{\mathbf{U}}} f_{\tilde{\mathbf{U}}}(\mathbf{u})\, d\mathbf{u}} \quad \text{for } \mathbf{u} \in B_{\mathbf{U}} \; . \tag{2}$$

The corresponding density for a probability group $\mathbf{P_k} = (P_{k1}, \ldots, P_{kl})'$ is

$$f_{\mathbf{P_k}}(\mathbf{p_k}) = \frac{f_{\tilde{\mathbf{P}}_\mathbf{k}}(\mathbf{p_k})}{\Pr(\tilde{\mathbf{P}}_\mathbf{k} \in B_{\mathbf{P}_k})} = \frac{f_{\tilde{\mathbf{P}}_\mathbf{k}}(\mathbf{p_k})}{\int \cdots \int_{B_{\mathbf{P}_k}} f_{\tilde{\mathbf{P}}_\mathbf{k}}(\mathbf{p_k})\, d\mathbf{p_k}} \quad \text{for } \mathbf{p_k} \in B_{\mathbf{P}_k} \; . \tag{3}$$

If the original belief distributions were defined in terms of marginal distributions rather than as a joint one, it is possible to express (2) and (3) even more explicitly. For the utilities one obtains by combining (1) and (2)

$$f_{\mathbf{U}}(\mathbf{u}) = \frac{f_{\tilde{U}_1}(u_1) \cdots f_{\tilde{U}_n}(u_n)}{\int \cdots \int_{B_{\mathbf{U}}} f_{\tilde{U}_1}(u_1) \cdots f_{\tilde{U}_n}(u_n)\, d\mathbf{u}} \quad \text{for } \mathbf{u} \in B_{\mathbf{U}} \; . \tag{4}$$

3.2 Marginal Constrained Belief Distributions

To see how the constraints altered the belief distribution over some variable, for example the j:th utility, it is necessary to compute the marginal constrained belief distribution U_j and compare this to the original marginal belief distribution $\tilde{U}_j$.

Any marginal distribution can be obtained by integrating the joint distribution over all variables except the one of interest. If $\mathbf{u_j^-} = (u_1, \ldots, u_{j-1}, u_{j+1}, \ldots, u_n)'$, then the marginal distribution of U_j is given by

$$f_{U_j}(u_j) = \int \cdots \iint \cdots \int_{B_U^{-j}} \left[\frac{f_{\tilde{\mathbf{U}}}(\mathbf{u})}{\int \cdots \int_{B_{\mathbf{U}}} f_{\tilde{\mathbf{U}}}(\mathbf{u})\, \mathrm{d}\mathbf{u}} \right] \mathrm{d}\mathbf{u_j^-} \quad \text{for } u_j \in B_U^j \ , \qquad (5)$$

where B_U^{-j} is the $n-1$-dimensional subspace of $B_{\mathbf{U}}$ that arises by removing its j:th dimension.

Example 1. Consider a decision tree where the belief distributions over two utilities u_a and u_b are both the standard uniform distribution U(0, 1):

$$f_{\tilde{U}_j}(u_j) = 1 \quad \text{for } 0 \le u_j \le 1 \text{ and } j \in \{a, b\} \ ,$$

and where there is only one constraint $\tilde{U}_a \le \tilde{U}_b$. Combining (1) and (5), the marginal belief distribution over u_b under the constraint can be computed:

$$f_{U_b}(u_b) = \int_0^{u_b} \left[\frac{1 \times 1}{\int_0^1 \int_0^{u_b} 1 \times 1\, \mathrm{d}u_a \mathrm{d}u_b} \right] \mathrm{d}u_a = \int_0^{u_b} \frac{\mathrm{d}u_a}{1/2} = 2u_b \quad \text{for } u_b \ge u_a \ .$$

We recognize that $U_b \sim \text{Beta}(2, 1)$, which is a special case of a more general result (see Theorem 3). The shift in marginal belief distribution over u_b, imposed by the constraint, is depicted in Fig. 1. This marked alteration of the marginal belief imposed by the constraint is a demonstration that constraints carry a substantial amount of information about the decision situation at hand. □

3.3 Belief Distribution Over Expected Utility

The quantity of main interest here is the expected utility for a given decision alternative, or the difference in expected utility between two alternatives. Within our proposed framework, the specified belief distributions over utilities and probabilities, in combination with the sets of comparative constraints, will all affect the resulting belief distribution over the expected utility.

The resulting distributions tend to become more and more warped around the mean as the depth and breadth of the tree increases. This phenomenon is due in part

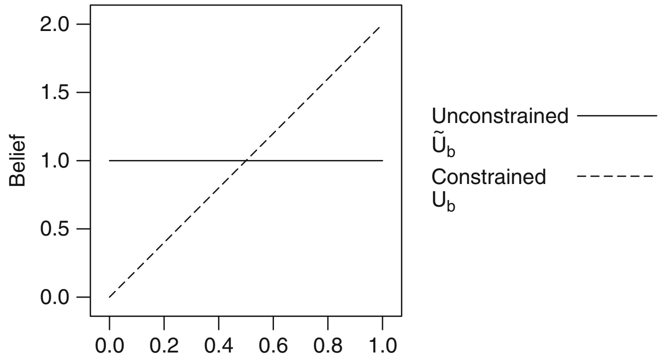

Fig. 1 Results from Example 1, where the considered decision tree contains two utilities whose unconstrained belief distributions were both standard uniform. Following the constraint $\tilde{U}_a \leq \tilde{U}_b$, the resulting (constrained) belief U_b over the second utility is now Beta(2, 1)

to the multiplication of distributions that takes place from root to leaf, and in part to the effects of convolution of the resulting leaf distributions Sundgren et al. (2009).

In view of Definition 4, if we impose belief distributions over the variables, the expected utility is really a transformation of random vectors. Therefore, to be able to analytically derive the resulting belief distribution over the expected utility of a given alternative, we need the following central result from probability theory:

Theorem 1 (The transformation theorem). *Let* $\mathbf{X} = (X_1, \ldots, X_n)'$ *be a continuous random vector with density function* $f_{\mathbf{X}}(\mathbf{x})$ *and domain* $V \subset \mathbb{R}^n$*. Let* $g = (g_1, \ldots, g_n)$ *be a bijection from* V *to a set* $W \subset \mathbb{R}^n$*, and define* $\mathbf{Y} = g(\mathbf{X})$*. Assume that* g *and its inverse* h *are both continuously differentiable. Then, the density function of* $\mathbf{Y}$ *is*

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(h_1(\mathbf{y}), \ldots, h_n(\mathbf{y})) \times |\det(\mathbf{J})| \quad \textit{for } \mathbf{y} \in W \ ,$$

where

$$\mathbf{J} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix}$$

is the Jacobian of the transformation. □

Theorem 1 can be applied to the expected utility of some alternative A_i:

Theorem 2. *Given a decision tree* T*, for the branch corresponding to* A_i*, label the (constrained) belief distributions over all utilities by* $\mathbf{U} = (U_1, \ldots, U_s)'$*, and label the belief distributions over all probabilities by* $\mathbf{P} = (P_1, \ldots, P_r)'$*. Further, let* $\mathbf{U}^- = (U_1, \ldots, U_{s-1})'$*. Finally, denote the index set for probabilities leading up* *to u_j by C_j, and let $\Psi_j = \prod_{i \in C_j} P_i$*[3]*. Then the belief distribution over* $\mathrm{E}(A_i)$ *is given by*

$$f_{\mathrm{E}(A_i)}(z) = \int \cdots \int f_{\mathbf{P},\mathbf{U}}\left(\mathbf{p}, \mathbf{u}^{-}, \frac{z - \sum_{i=1}^{s-1} \psi_i u_i}{\psi_s}\right) \frac{1}{\psi_s} \, \mathrm{d}\mathbf{p} \, \mathrm{d}\mathbf{u}^{-} \, .$$

Proof. Let $Z = \mathrm{E}(A_i)$ and consider the following transformation:

$$\begin{cases} (Y_1, \ldots, Y_r)' & = \mathbf{P} \\ (Y_{r+1}, \ldots, Y_{r+s-1})' & = \mathbf{U}^{-} \\ Y_{r+s} & = Z = \Psi_s U_s + \sum_{i=1}^{s-1} \Psi_i U_i \end{cases}$$

which has the following inverse:

$$\begin{cases} \mathbf{P} = (Y_1, \ldots, Y_r)' \\ \mathbf{U} = \left(Y_{r+1}, \ldots, Y_{r+s-1}, \frac{Y_{r+s} - \sum_{i=1}^{s-1} \Psi_i Y_{r+i}}{\Psi_s}\right)' \end{cases} .$$

The Jacobian of this transformation has the following determinant:

$$\det(\mathbf{J}) = \begin{vmatrix} \frac{\partial p_1}{\partial y_1} & \cdots & \frac{\partial p_1}{\partial y_{r+s}} \\ \vdots & \ddots & \vdots \\ \frac{\partial p_r}{\partial y_1} & \cdots & \frac{\partial p_r}{\partial y_{r+s}} \\ \frac{\partial u_1}{\partial y_1} & \cdots & \frac{\partial u_1}{\partial y_{r+s}} \\ \vdots & \ddots & \vdots \\ \frac{\partial u_s}{\partial y_1} & \cdots & \frac{\partial u_s}{\partial y_{r+s}} \end{vmatrix} = \begin{vmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ e_1 & e_2 & \ldots & 1/\psi_s \end{vmatrix} = 1/\psi_s$$

where $e_1, \ldots, e_{r+s-1}$ are partial derivatives that do not contribute to the determinant and therefore do not need to be calculated. According to Theorem 1, the joint density of $\mathbf{Y} = (Y_1, \ldots, Y_{r+s})'$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{P},\mathbf{U}}\left(\mathbf{p}, \mathbf{u}^{-}, \frac{z - \sum_{i=1}^{s-1} \psi_i u_i}{\psi_s}\right) \frac{1}{\psi_s}$$

[3] That is, Ψ_j is the aggregated belief distribution over the probability at the leaf of the j:th utility.

for some domain W. The marginal belief distribution of $\mathrm{E}(A_i) = Z$ is derived by integrating out all other variables:

$$f_{\mathrm{E}(A_i)}(z) = \int \cdots \int f_{\mathbf{P},\mathbf{U}}\left(\mathbf{p}, \mathbf{u}^{-}, \frac{z - \sum_{i=1}^{s-1} \psi_i u_i}{\psi_s}\right) \frac{1}{\psi_s} \, \mathrm{d}\mathbf{p} \, \mathrm{d}\mathbf{u}^{-}$$

where integration is over some subspace of $[0, 1]^{r+s-1}$ and the domain of $f_{\mathrm{E}(A_i)}$ can be called W'. □

A few remarks are called upon. Firstly, $\mathrm{E}(A_i)$ could have been expressed in terms of any U_j, however the choice of U_s is notationally convenient. Secondly, it should be noted that each ψ_j is not a constant, but rather a product of p_is. Finally, since $f_{\mathbf{P},\mathbf{U}}$ is the joint (constrained) density for the beliefs over all utilities and probabilities, it can be factorized into groups of independent variables. Specifically, all utilities will be independent of all probabilities, and probabilities from different groups will also be independent.

The following example shows how the transformation works in practice for an apparently simple situation.

Example 2. Consider the decision tree for some alternative A_i given in Fig. 2. The belief distribution over the expected utility is given by

$$\begin{aligned} \mathrm{E}(A_i) &= P_{11}P_{111}U_{111} + P_{11}P_{112}U_{112} + P_{12}U_{12} \\ &= P_{11}P_{111}U_{111} + P_{11}(1 - P_{111})U_{112} + (1 - P_{11})U_{12} \, . \end{aligned}$$

For simplicity, since it reduces the dimensionality of the problem, we assume that the reformulation of P_{112} and P_{12} is coherent with the constraints used. The transformation becomes

$$\begin{cases} (Y_1, Y_2, Y_3, Y_4)' = (P_{11}, P_{111}, U_{111}, U_{112})' \\ Y_5 \qquad\qquad\quad\;\; = P_{11}P_{111}U_{111} + P_{11}(1 - P_{111})U_{112} + (1 - P_{11})U_{12} \end{cases}$$

with inverse

$$\begin{cases} (P_{11}, P_{111}, U_{111}, U_{112})' = (Y_1, Y_2, Y_3, Y_4)' \\ U_{12} \qquad\qquad\qquad\qquad\;\; = \dfrac{Y_5 - Y_1Y_2Y_3 - Y_1(1 - Y_2)Y_4}{1 - Y_1} \end{cases} .$$

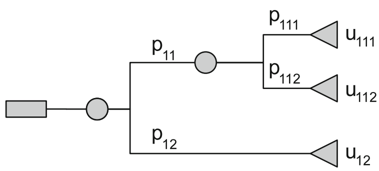

Fig. 2 Decision tree considered in Example 2

The determinant of the Jacobian is given by

$$\det(\mathbf{J}) = \begin{vmatrix} \frac{\partial p_{11}}{\partial y_1} & \frac{\partial p_{11}}{\partial y_2} & \frac{\partial p_{11}}{\partial y_3} & \frac{\partial p_{11}}{\partial y_4} & \frac{\partial p_{11}}{\partial y_5} \\ \frac{\partial p_{111}}{\partial y_1} & \frac{\partial p_{111}}{\partial y_2} & \frac{\partial p_{111}}{\partial y_3} & \frac{\partial p_{111}}{\partial y_4} & \frac{\partial p_{111}}{\partial y_5} \\ \frac{\partial u_{111}}{\partial y_1} & \frac{\partial u_{111}}{\partial y_2} & \frac{\partial u_{111}}{\partial y_3} & \frac{\partial u_{111}}{\partial y_4} & \frac{\partial u_{111}}{\partial y_5} \\ \frac{\partial u_{112}}{\partial y_1} & \frac{\partial u_{112}}{\partial y_2} & \frac{\partial u_{112}}{\partial y_3} & \frac{\partial u_{112}}{\partial y_4} & \frac{\partial u_{112}}{\partial y_5} \\ \frac{\partial u_{12}}{\partial y_1} & \frac{\partial u_{12}}{\partial y_2} & \frac{\partial u_{12}}{\partial y_3} & \frac{\partial u_{12}}{\partial y_4} & \frac{\partial u_{12}}{\partial y_5} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ e_1 & e_2 & e_3 & e_4 & \frac{1}{(1-y_1)} \end{vmatrix} .$$

This evaluates to $\frac{1}{1-y_1} = \frac{1}{1-p_{11}}$, which coincides with the general description above, for $\psi_s = p_{12} = 1 - p_{11}$. The joint density of $\mathbf{Y}$ is given, for some domain W, by

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{P_{11},P_{111},U_{111},U_{112},U_{12}}\left(p_{11}, p_{111}, u_{111}, u_{112}, \frac{z-v}{1-p_{11}}\right)\frac{1}{1-p_{11}} ,$$

with $v = p_{11}p_{111}u_{111} + p_{11}(1-p_{111})u_{112}$. This density can be factorized with respect to independent variables. Assuming that the utilities are not independent under the given constraints, one obtains

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{P_{11}}(p_{11}) f_{P_{111}}(p_{111}) f_{U_{111},U_{112},U_{12}}\left(u_{111}, u_{112}, \frac{z-v}{1-p_{11}}\right)\frac{1}{1-p_{11}} .$$

Finally, the resulting belief distribution over $\mathrm{E}(A_i)$ can be calculated from

$$f_{\mathrm{E}(A_i)}(z) = \iiiint f_{\mathbf{Y}}(\mathbf{y})\, \mathrm{d}p_{11}\, \mathrm{d}p_{111}\, \mathrm{d}u_{111}\, \mathrm{d}u_{112} . \qquad \square$$

However, the complexity of this operation is very high since both the integrand itself as well as the integration limits might be very difficult to derive. And even if these steps have been carried out, the sheer dimensionality of the integration might be prohibitive in practice. So in real cases more efficient methods of calculating this must be utilized. As we shall see, in certain cases this can be done at relative computational ease, even for moderately large trees.

4 Simulation from Expected Utilities

As shown in Theorem 2, the resulting distribution over the expected utility of an alternative can be expressed in terms of a multidimensional integral. However, in general such integrals should rarely be possible to compute analytically, and so to be able to benefit from the theoretical results presented thus far, approximate methods are called upon. One possibility would be to make use of numerical integration techniques. We have, however, opted for another solution, namely to use simulation.

To be able to simulate from the resulting belief distributions over the expected utilities of the alternatives, or from some function thereof, we need to sample from the respective constrained belief distributions over probabilities and utilities. Any utility is independent of any probability, and any two probabilities from separate groups are independent, because of the restricted set of constraints allowed. Therefore we can sample from the joint belief distribution over the utilities and from the belief distributions over the various probability groups separately.

The most straightforward approach would be to utilize (2) and (3) through a simple form of rejection sampling von Neumann (1963). Specifically, exemplifying with utilities, we could simply apply the following scheme:

Algorithm 1 (Rejection sampling).

Repeat until m samples are retained in total:
Sample $\tilde{\mathbf{u}} = (\tilde{u}_1, \ldots, \tilde{u}_n)'$ from $(\tilde{U}_1, \ldots, \tilde{U}_n)'$ and retain the sample if $\tilde{\mathbf{u}} \in B_{\mathbf{U}}$.
□

Even if $\tilde{U}_1, \ldots, \tilde{U}_n$ are independent, so that sampling is straightforward, this approach has one major drawback: As n grows, the probability that a sample is accepted approaches 0, which in effect means that the real number of samples m' required to collect a nominal number of m samples approaches infinity. If the intended application is not interactive, however, this straightforward simulation scheme might suffice.

Another generic approach to sampling from a multivariate distribution is to factorize the joint density function into a series of univariate, conditional distributions. For our joint distribution of beliefs over the utilities this would correspond to the following:

$$f_{U_1,\ldots,U_n}(u_1, \ldots, u_n) = f_{U_n}(u_n) f_{U_{n-1}|U_n=u_n}(u_{n-1}) \cdots f_{U_1|U_2=u_2,\ldots,U_n=u_n}(u_1) \,. \quad (6)$$

For each of the univariate distributions, sampling can be performed according to the inverse transformation method Devroye (1986). This method requires draws from a standard uniform distribution to be inserted into the inverse of the distribution function. Therefore, in practice, this approach can only be really effective if the distribution functions that correspond to the density functions in (6) can be analytically inverted. The complete sampling scheme is summarized in the following:

Algorithm 2 (Multivariate inverse transform sampling).

---

1. Split up $f_{U_1,\ldots,U_n}$ according to (6).
2. Derive the distribution functions $F_{U_n}, F_{U_{n-1}|U_n=u_n}, \ldots, F_{U_1|U_2=u_2,\ldots,U_n=u_n}$.
3. Derive the inverse distribution functions $F^{-1}_{U_n}, F^{-1}_{U_{n-1}|U_n=u_n}, \ldots, F^{-1}_{U_1|U_2=u_2,\ldots,U_n=u_n}$.
4. Draw n vectors $\mathbf{x}_i = (x_i^1, \ldots, x_i^m)'$ of samples from $X \sim \mathrm{U}(0, 1)$.

5a. Sample from U_n:

$$(u_n^1, \ldots, u_n^m)' = (F_{U_n}^{-1}(x_n^1), \ldots, F_{U_n}^{-1}(x_n^m))'$$

5b. Sample from U_{n-1}:

$$(u_{n-1}^1, \ldots, u_{n-1}^m)' = (F_{U_{n-1}|U_n=u_n^1}^{-1}(x_{n-1}^1), \ldots, F_{U_{n-1}|U_n=u_n^m}^{-1}(x_{n-1}^m))'$$

⋮

5n. Sample from U_1 :

$$(u_1^1, \ldots, u_1^m)' = (F_{U_1|U_2=u_2^1,\ldots,U_n=u_n^1}^{-1}(x_1^1), \ldots, F_{U_1|U_2=u_2^m,\ldots,U_n=u_n^m}^{-1}(x_1^m))' \qquad \square$$

The following example is supposed to delineate the fundamental principles of Algorithm 2:

Example 3. Consider again the situation in Example 1. According to (4), the joint constrained belief distribution is given by

$$f_{U_a,U_b}(u_a,u_b) = \frac{f_{\tilde{U}_a}(u_a) f_{\tilde{U}_b}(u_b)}{\int_0^1 \int_0^{u_b} f_{\tilde{U}_a}(u_a) f_{\tilde{U}_b}(u_b)\, \mathrm{d}u_a \mathrm{d}u_b} = \frac{1 \times 1}{\int_0^1 \int_0^{u_b} 1 \times 1\, \mathrm{d}u_a \mathrm{d}u_b}$$
$$= \frac{1}{1/2} = 2$$

for $u_b \geq u_a$. As will be demonstrated in Sect. 4.1, f_{U_a,U_b} can be split up as follows:

$$f_{U_a,U_b}(u_a,u_b) = f_{U_b}(u_b) f_{U_a|U_b=u_b}(u_a) = 2u_b \times (1/u_b) \; .$$

The corresponding distribution functions and their inverses are given by

$$\begin{cases} F_{U_b}(u_b) &= u_b^2 \text{ for } 0 \leq u_b \leq 1 \\ F_{U_a|U_b=u_b}(u_a) &= \frac{u_a}{u_b} \text{ for } 0 \leq u_a \leq u_b \end{cases} \text{ and } \begin{cases} F_{U_b}^{-1}(x) &= \sqrt{x} \text{ for } 0 \leq x \leq 1 \\ F_{U_a|U_b=u_b}^{-1}(x) &= x u_b \text{ for } 0 \leq x \leq 1 \end{cases}.$$

Assume we want $m = 5$ draws, and that we sample $\mathbf{x}_b = (0.94, 0.13, 0.83, 0.47, 0.55)'$ and $\mathbf{x}_a = (0.18, 0.70, 0.57, 0.17, 0.94)'$ from $X \sim \mathrm{U}(0,1)$. This then yields

$$\begin{bmatrix} u_b^1 \\ u_b^2 \\ u_b^3 \\ u_b^4 \\ u_b^5 \end{bmatrix} = \begin{bmatrix} \sqrt{0.94} \\ \sqrt{0.13} \\ \sqrt{0.83} \\ \sqrt{0.47} \\ \sqrt{0.55} \end{bmatrix} = \begin{bmatrix} 0.97 \\ 0.36 \\ 0.91 \\ 0.69 \\ 0.74 \end{bmatrix} \text{ and } \begin{bmatrix} u_a^1 \\ u_a^2 \\ u_a^3 \\ u_a^4 \\ u_a^5 \end{bmatrix} = \begin{bmatrix} 0.18 \times u_b^1 \\ 0.70 \times u_b^2 \\ 0.57 \times u_b^3 \\ 0.17 \times u_b^4 \\ 0.94 \times u_b^5 \end{bmatrix} = \begin{bmatrix} 0.17 \\ 0.25 \\ 0.52 \\ 0.12 \\ 0.70 \end{bmatrix}.$$

Clearly, to obtain samples that accurately represent the $(U_a, U_b)'$ distribution, one would need to set m considerably higher than in this illustrative example. □

An implicit assumption in Algorithm 2 is that each distribution function be strictly increasing, since otherwise its inverse would not exist. However, this criterion is fulfilled for the vast majority of distributions and should not present any issues in practice. There are other complicating matters of greater practical concern, such that for an arbitrary joint distribution, even step 1 in the scheme might be difficult to accomplish. Should step 1 be successful, step 3 might still prove difficult or impossible. Should this be the case, one could use numerical techniques for inverting the distribution functions, though this would slow down the procedure quite dramatically. We shall therefore make use of one particular distribution, the uniform distribution, where it is possible to carry out the complete scheme. The uniform distribution is appealing since it can be thought of as a direct probabilistic equivalent to using intervals. In other words, if the decision maker would choose an interval $[a, b]$, a natural choice in the context of the current approach would be the $\mathrm{U}(a, b)$ distribution.

4.1 Special Case of Uniform Distributions

One important sub-case is when the distributions are uniform. We describe here an efficient way of sampling from the resulting belief distribution over the expected utility if one (a) uses arbitrary independent uniform belief distributions over all utilities; (b) uses an ordering constraint within each of an arbitrary number of disjoint subsets of the utilities[4]; and (c) uses no other constraints. Because there are no relations between utilities from different subsets, independence allows sampling from each subset separately. We can therefore, without loss of generality, describe the special case where there is just one subset $\{u_1, \ldots, u_n\}$ containing all utilities.

For the probabilities, the implicit sum constraint is a complicating factor. We will here use the Dirichlet distribution to sample from a probability group, and not impose any relational constraints. The Dirichlet distribution has been previously used in this context Ekenberg et al. (2007) and it does relate to the above: If we use standard uniform distributions over $p_{k1}, \ldots, p_{kl}$ and impose no further constraints than the implicit sum constraint, then $(P_{k1}, \ldots, P_{kl})'$ will be distributed according to the Dirichlet$(\alpha_1 = 1, \ldots, \alpha_l = 1)$ distribution. Sampling from a Dirichlet distribution is straightforward using standard statistical software.

Assume, for brevity, that we have named the utilities after their ordering. Thus, the constraint under consideration is $A_{\mathbf{U}} : \tilde{U}_1 \leq \ldots \leq \tilde{U}_n$. The following result, given with a standard proof, is useful:

[4] The union of these subsets need not equate the set of all utilities.

Theorem 3. *Let $X_1, \ldots, X_n$ be a sample from a distribution with density function f and distribution function F. Then the density function for the largest observation $X_{(n)}$ is $f_{\mathrm{Beta}(n,1)}(F(x))f(x)$.*

Proof. The distribution function for $X_{(n)}$ is given by

$$F_{X_{(n)}}(x) = \Pr(X_1 \le x, \ldots, X_n \le x) = \prod_{k=1}^{n} \Pr(X_k \le x) = (F(x))^n \ .$$

Differentiation then yields the density function:

$$f_{X_{(n)}}(x) = n(F(x))^{n-1} f(x) = f_{\mathrm{Beta}(n,1)}(F(x))f(x) \ . \qquad \square$$

If all unconstrained belief distributions $\tilde{U}_j$ are $\mathrm{U}(a,b)$, then the resulting belief distribution U_n is equivalent to $X_{(n)}$ in Theorem 3. It follows that

$$f_{U_n}(u_n) = n\left(\frac{u_n - a}{b-a}\right)^{n-1} \frac{1}{b-a} = \frac{n(u_n-a)^{n-1}}{(b-a)^n} \ .$$

Conditional on $U_n = u_n$, all remaining belief distributions are $\mathrm{U}(a,b)$ truncated to the interval $[a, u_n]$, which means that $\tilde{U}_j | U_n = u_n \sim \mathrm{U}(a, u_n)$ for $j < n$. Since $\tilde{U}_{n-1}$ is now the largest of the remaining variables, Theorem 3 gives

$$f_{U_{n-1}|U_n=u_n}(u_{n-1}) = \frac{(n-1)(u_{n-1}-a)^{n-2}}{(u_n-a)^{n-1}} \ .$$

By repeating the same argument for all belief distributions down to $\tilde{U}_1$, the joint density for the constrained belief distribution can be factorized, as required:

$$\begin{aligned} f_{U_1,\ldots,U_n}(u_1,\ldots,u_n) &= \frac{n!}{(b-a)^n} \\ &= \frac{n(u_n-a)^{n-1}}{(b-a)^n} \frac{(n-1)(u_{n-1}-a)^{n-2}}{(u_n-a)^{n-1}} \cdots \frac{2(u_2-a)}{(u_3-a)^2} \frac{1}{u_2-a} \\ &= f_{U_n}(u_n) f_{U_{n-1}|U_n=u_n}(u_{n-1}) \cdots f_{U_2|U_3=u_3}(u_2) f_{U_1|U_2=u_2}(u_1) \ . \end{aligned}$$

All corresponding distribution functions can be derived, and it turns out that they are all analytically invertible:

$$\begin{cases} F^{-1}_{U_n}(x) & = x^{1/n}(b-a) + a \quad \text{for } k = n \\ F^{-1}_{U_k|U_{k+1}=u_{k+1}}(x) & = x^{1/k}(u_{k+1}-a) + a \ \text{for } k = n-1, n-2, \ldots, 1 \end{cases} . \qquad (7)$$

Thus, all prerequisites for applying Algorithm 2 are fulfilled.

It would be a severe limitation to require that $\tilde{U}_1, \ldots, \tilde{U}_n$ follow the same uniform distribution. We can overcome this by once again making use of the fact that a truncated uniform variable still is uniform. Let each variable have its own uniform distribution, $\tilde{U}_j \sim \mathrm{U}(a_j, b_j)$, and consider the following scheme:

Algorithm 3.

1. Order the set $\{0, a_1, \ldots, a_n, b_1, \ldots, b_n, 1\}$, and define all intervals $I_j = [c_j, d_j]$ from adjacent points in this ordered set. (Note that some points might be identical, in which case no interval results.) Denote the total number of intervals by N^I.
2. Construct all *configurations* C_i, i.e. all ways to distribute $\tilde{U}_1, \ldots, \tilde{U}_n$ between these intervals. Denote by N^C the total number of configurations where $A_{\mathbf{U}}$ can hold, and by N_{ij}^U the number of variables $\tilde{U}_k$ in interval I_j under configuration C_i.
3. FOR $i = 1$ TO $i = N^C$

    FOR $j = 1$ TO $j = N^I$

        FOR $k = 1$ TO $k = N_{ij}^U$

            Calculate $\bar{p}_{ijk} = \Pr(\tilde{U}_k \in I_j) = F_{\tilde{U}_k}(d_j) - F_{\tilde{U}_k}(c_j)$

        END FOR

        Calculate $\bar{p}_{ij} = \Pr(\tilde{U}_1, \ldots, \tilde{U}_{N_{ij}^U} \in I_j) = \prod_{k=1}^{N_{ij}^U} \bar{p}_{ijk}$

        Calculate $\widehat{p}_{ij} = \Pr(A_{\mathbf{U}} \text{ holds in } I_j | C_i) = 1/N_{ij}^U!$

    END FOR

    Calculate $\bar{p}_i = \Pr(C_i) = \prod_{j=1}^{N^I} \bar{p}_{ij}$

    Calculate $\widehat{p}_i = \Pr(A_{\mathbf{U}} | C_i) = \prod_{j=1}^{N^I} \widehat{p}_{ij}$

    Calculate $p_i = \widehat{p}_i \bar{p}_i$

   END FOR

   Calculate $p = \sum_{i=1}^{N^C} p_i$
4. FOR $r = 1$ TO $r = m$

    Sample $\mathbf{x^r} = (x_1^r, \ldots, x_{N^C}^r)'$ from $(X_1, \ldots, X_{N^C})' \sim$ Multinom(1; $p_1/p, \ldots, p_{N^C}/p$)

    FOR the single i corresponding to $x_i^r = 1$

        FOR $j = 1$ TO $j = N^I$

            Draw a sample of the variables in I_j under configuration C_i, by setting $a = c_j$ and $b = d_j$ in (7), and using Algorithm 2

        END FOR

    END FOR

   END FOR □

In other words, Algorithm 3 simulates from $f_{\mathbf{U}}(\mathbf{u})$ by treating it as a mixture density over the configurations, with suggested mixture parameters p_i/p. The validity of this approach is asserted by the following theorem:

Theorem 4. *Consider a decision tree T with a set of utility variables whose corresponding belief distributions $\tilde{U}_1, \ldots, \tilde{U}_n$ are independent and distributed as $\tilde{U}_j \sim \mathrm{U}(a_j, b_j)$. Under the constraint $A_{\mathbf{U}} : \tilde{U}_1 \leq \ldots \leq \tilde{U}_n$, Algorithm 3 can be used to sample from the resulting belief distribution $\mathbf{U} = (U_1, \ldots, U_n)'$.*

Proof. By the law of total probability and Bayes' theorem, and following the notation of Algorithm 3:

$$
\begin{aligned}
f_{\mathbf{U}}(\mathbf{u}) &= \sum_{i=1}^{N^C} \left[f_{\tilde{\mathbf{U}}|A_{\mathbf{U}},C_i}(\mathbf{u}) \times \Pr(C_i|A_{\mathbf{U}}) \right] \\
&= \sum_{i=1}^{N^C} \left[f_{\tilde{\mathbf{U}}|A_{\mathbf{U}},C_i}(\mathbf{u}) \times \frac{\Pr(A_{\mathbf{U}}|C_i)\Pr(C_i)}{\Pr(A_{\mathbf{U}})} \right] \\
&= \sum_{i=1}^{N^C} \left[f_{\tilde{\mathbf{U}}|A_{\mathbf{U}},C_i}(\mathbf{u}) \times \frac{\Pr(A_{\mathbf{U}}|C_i)\Pr(C_i)}{\sum_{i=1}^{N^C} \Pr(A_{\mathbf{U}}|C_i)\Pr(C_i)} \right] \\
&= \sum_{i=1}^{N^C} \left[f_{\tilde{\mathbf{U}}|A_{\mathbf{U}},C_i}(\mathbf{u}) \times \frac{p_i}{p} \right],
\end{aligned}
$$

which shows that p_i/p is the correct mixture parameter for configuration C_i.

Finally the claim follows by considering the distributions $f_{\tilde{\mathbf{U}}|A_{\mathbf{U}},C_i}(\mathbf{u})$. Under C_i, but not yet considering $A_{\mathbf{U}}$, all variables in I_j are distributed as $\mathrm{U}(c_j, d_j)$. Thus, since they are equidistributed, $A_{\mathbf{U}}$ can be introduced through Algorithm 2 and (7). Further, because the relative order of any two variables from different intervals is fixed given C_i, they are independent, and the complete distribution $(U_1, \ldots, U_n)'$ can be obtained by repeated interval-wise sampling from configurations drawn according to a Multinom$(1;\ p_1/p, \ldots, p_{N^C}/p)$ distribution. □

Note that the algorithm only includes configurations where $A_{\mathbf{U}}$ can hold. This is for practical reasons, and the validity follows immediately since $\Pr(A_{\mathbf{U}}|C_i) = 0$ for any configuration not fulfilling this criterion.

If there are n variables and N^I resulting intervals, the total number of configurations where $A_{\mathbf{U}}$ can hold is $\binom{n+N^I-1}{N^I-1}$. Therefore, in practice, Algorithm 3 is only efficient if N^I is kept relatively low, which means that the variables need to share a few common endpoints. In our experience, it is very feasible to allow a resolution of 0.2, meaning that endpoints are chosen from $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$, even for n relatively large (about 30). Obviously the computational complexity will also depend on how many samples that are desired in the simulation. Note that the tree as a whole can be much larger than n; it is really the size of the largest set of ordered utility variables that matters.

Example 4. Consider the decision tree with two alternatives given in Fig. 3. Assume that the decision maker has specified the following:

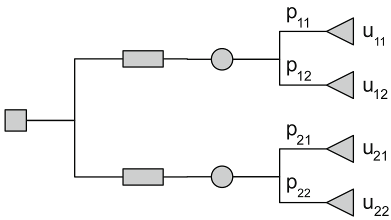

Fig. 3 Decision tree considered in Example 4

$$\begin{cases} u_{11}, u_{12}, u_{21} \in [0, 1] \text{ and } u_{22} \in [0, 0.5] \\ p_{11}, p_{12}, p_{21}, p_{22} \in [0, 1] \\ u_{11} \geq u_{21} \text{ and } u_{12} \geq u_{22} \end{cases}.$$

The intuitive interpretation of this situation is that alternative 1 is superior to alternative 2. However, an interval analysis would yield $\mathrm{E}(A_1) - \mathrm{E}(A_2) \in [-1, 1]$, indicating no discrimination whatsoever.

Now consider the approach described here, and assume that the decision maker instead specifies

$$\begin{cases} \tilde{U}_{11}, \tilde{U}_{12}, \tilde{U}_{21} \sim \mathrm{U}(0, 1) \text{ and } \tilde{U}_{22} \sim \mathrm{U}(0, 0.5) \\ (\tilde{P}_{11}, \tilde{P}_{12})' \sim \text{Dirichlet}(1, 1) \text{ and } (\tilde{P}_{21}, \tilde{P}_{22})' \sim \text{Dirichlet}(1, 1) \\ \tilde{U}_{11} \geq \tilde{U}_{21} \text{ and } \tilde{U}_{12} \geq \tilde{U}_{22} \end{cases}.$$

The resulting belief distribution over $(\mathrm{E}(A_1), \mathrm{E}(A_2))'$ was simulated using Algorithm 3 separately for $(\tilde{U}_{11}, \tilde{U}_{21})'$ and $(\tilde{U}_{12}, \tilde{U}_{22})'$[5]. $(\tilde{P}_{11}, \tilde{P}_{12})'$ and $(\tilde{P}_{21}, \tilde{P}_{22})'$ were also simulated separately. 1 million samples were drawn, and the results are displayed in Fig. 4. It is apparent from these figures that the belief in alternative 1 exceeds that of alternative 2 quite clearly, even though there is some overlap. Further, the simulation provides us with a straightforward quantification of the degree of discrimination. For example, we can extract a 90% probability interval for $\mathrm{E}(A_1) - \mathrm{E}(A_2)$ from the 5:th and 95:th percentiles of the simulated values, which happens to be $[0.03, 0.71]$. The interpretation of this interval is direct: Under the beliefs and constraints specified by the decision maker, there is a 90% probability that the difference in expected utility between alternatives 1 and 2 lies between 0.03 and 0.71. Based on this analysis we can confidently discriminate between the alternatives, and the result is coherent with intuition. □

A desirable extension of the proposed method would be to allow for other classes of distributions than the uniform. There are two inherent obstacles in the framework that precludes this. First, as can be realized from Theorem 3, the class of density functions for the distribution in question must obey rather strict form conventions, in

[5] For $(\tilde{U}_{11}, \tilde{U}_{21})'$, there is only one interval to consider, and so it really suffices with Algorithm 2. Note that $\tilde{U}_{11}$ and $\tilde{U}_{21}$ are precisely $\tilde{U}_b$ and $\tilde{U}_a$, respectively, from Examples 1 and 3.

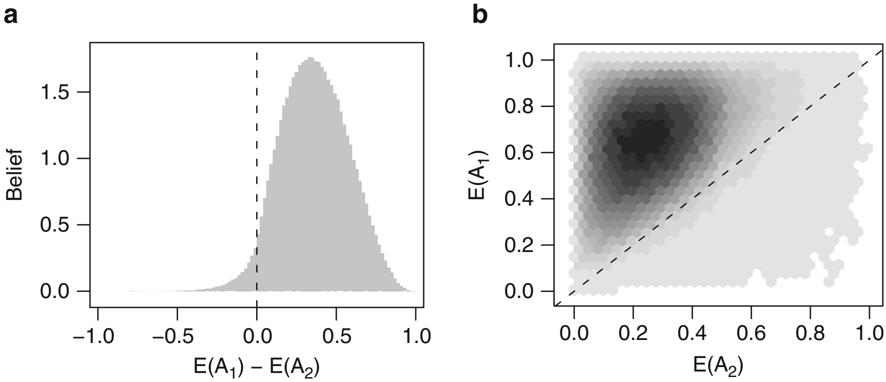

Fig. 4 Results from the simulation of Example 4. **(a)** Histogram over the simulated belief for the difference $\mathrm{E}(A_1) - \mathrm{E}(A_2)$. **(b)** Two-dimensional histogram over the simulated beliefs for $\mathrm{E}(A_1)$ and $\mathrm{E}(A_2)$. Darker color indicates a higher density of points

order to yield analytically invertible distribution functions for the ordered variables. Second, (7), as well as the fundamental idea behind Algorithm 3, rest on a subtle, yet very powerful property of the uniform distribution: A truncated uniform variable is still uniform, with parameters equal to the endpoints of the truncated interval. As far as we are aware, no other continuous distribution shares this property. Therefore, to be able to derive similar methods for other classes of distributions, fundamentally different approaches would be needed.

5 Summary and Conclusions

There is often a need in real-life decision analyses to allow for imprecise statements regarding probabilities and utilities. Various interval-based approaches have been suggested, but the resulting aggregations range within intervals as well, causing an, often unnecessary, information loss. There is therefore a need to extend the representation of the decision situation using more information, but keeping the requirement that the decision maker does not have to be more precise than what is possible.

To this end, we have proposed a solution where qualitative statements, i.e. relations between variables, are combined with second-order information in the form of belief distributions over probability and utility variables. Within the framework of belief distributions, we have demonstrated how qualitative statements translate to constraints, and how these constraints affect both marginal belief distributions as well as the resulting belief distributions over the expected utilities of the alternatives. Further, and of clear practical benefit, we have presented a computationally meaningful method where arbitrary uniform distributions can be used for the utility variables, in combination with ordering relations among these variables. This method rests on a series of fairly subtle, yet fundamental, theoretical arguments.

The results presented here have two distinct advantages. First, while preserving the ability to be deliberately imprecise and including qualitative relations, our approach allows a high degree of discrimination between alternatives compared to what interval-based approaches can accomplish. At the same time, we can make a semantically clear interpretation of the resulting beliefs within a probabilistic framework.

Acknowledgements The authors wish to thank Alina Kuznetsova for valuable suggestions that helped improve the contents of this paper.

References

Choquet, G. (1954). Theory of capacities. *Annales de l'institut Fourier*, *5*, 131–295.

Danielson, M., & Ekenberg, L. (1998). A framework for analysing decisions under risk. *European Journal of Operational Research*, *104*(3), 474–484.

Danielson, M., & Ekenberg, L. (2007). Computing upper and lower bounds in interval decision trees. *European Journal of Operational Research*, *181*(3), 808–816.

Danielson, M., Ekenberg, L., & Larsson, A. (2007). Distribution of expected utility in decision trees. *International Journal of Approximate Reasoning*, *46*(2), 387–407.

Danielson, M., Ekenberg, L., & Riabacke, A. (2009). A prescriptive approach to elicitation of decision data. *Journal of Statistical Theory and Practice*, *3*(1), 157–168.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, *38*(2), 325–399.

Devroye, L. (1986). Non-Uniform Random Variate Generation. New York, USA: Springer.

Ding, X., Danielson, M., & Ekenberg, L. (2010). Disjoint programming in computational decision analysis. *Journal of Uncertain Systems*, *4*(1), 4–13.

Dubois, D., & Prade, H. (1988). *Possibility Theory*. New York: Plenum Press.

Ekenberg, L. (2000). The logic of conflicts between decision making agents. *Journal of Logic and Computation*, *10*(4), 583–602.

Ekenberg, L., Andersson, M., Danielsson, M., & Larsson, A. (2007). Distributions over Expected Utilities in Decision Analysis. In 5th International Symposium on Imprecise Probability: Theories and Applications, Prague, Czech Republic.

Ekenberg, L., Danielson, M., & Thorbiörnson, J. (2006). Multiplicative properties in evaluation of decision trees. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *14*(3), 293–316.

Ekenberg, L., & Thorbiörnson, J. (2001). Second-order decision analysis. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *9*(1), 13–37.

Ekenberg, L., Thorbiörnson, J., & Baidya, T. (2005). Value differences using second-order distributions. *International Journal of Approximate Reasoning*, *38*(1), 81–97.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, *75*(4), 643–669.

Good, I. J. (1962). Subjective probability as the measure of a non-measurable set. In *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress*, pp. 319–329. Stanford University Press.

Jaffray, J. Y. (1999). Rational Decision Making With Imprecise Probabilities. In *1st International Symposium on Imprecise Probabilities and Their Applications, Ghent, Belgium*.

von Neumann, J. (1963). Various techniques used in connection with random digits. In A. H. Taub (Ed.), *John von Neumann, Collected Works*, Vol. V. New York, USA: MacMillan.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.

Shapira, Z. (1995). *Risk Taking: A Managerial Perspective*. Russel Sage Foundation.

Smith, C. A. B. (1961). Consistency in statistical inference and decision. *Journal of the Royal Statistical Society. Series B (Methodological) 23*(1), 1–25.

Sundgren, D., Danielson, M., & Ekenberg, L. (2009). Warp effects on calculating interval probabilities. *International Journal of Approximate Reasoning*, *50*(9), 1360–1368.