Generalised Risk Constraints in Decision Trees with Second-Order Probabilities

MATS DANIELSON  
Stockholm University  
Dept. of Computer and  
Systems Sciences  
Forum 100, SE-164 40 Kista  
SWEDEN  
mad@dsv.su.se

LOVE EKENBERG  
Stockholm University  
Dept. of Computer and  
Systems Sciences  
Forum 100, SE-164 40 Kista  
SWEDEN  
lovek@dsv.su.se

ARON LARSSON  
Stockholm University  
Dept. of Computer and  
Systems Sciences  
Forum 100, SE-164 40 Kista  
SWEDEN  
aron@dsv.su.se

Abstract: Recently, representations and methods analysing decision problems where probabilities and values (utilities) are associated with belief distributions over them (second order representations) have been suggested. In this paper we present an approach to how imprecise information can be modelled by means of second-order distributions and how a risk evaluation process can be elaborated by integrating procedures for numerically imprecise probabilities and utilities. We discuss some shortcomings in the use of the principle of maximising the expected utility and of utility theory in general, and offer remedies by the introduction of supplementary decision rules based on a concept of risk constraints taking advantage of second-order distributions.

Keywords: Decision analysis, risk analysis, interval probability, second-order probability, decision rule, utility theory.

1 Introduction

Methods and tools for analysing and evaluating decision problems under risk have been of great interest for centuries. During the latter decades, such methods have been more or less systematically integrated with risk management processes (cf., e.g., [1, 14] among others). The prevailing decision rule serving as an instrument for ensuring substantial rationality in decision making under risk is commonly referred to as the principle of maximising the expected utility (PMEU). The principle is inspired by early efforts in normative decision theory, e.g., [21, 13, 23]. It is derived from a number of different, although similar, axiom systems aiming to reflect the properties of a rational entity’s behaviour when discriminating between decision alternatives given that the consequence of an alternative is uncertain but the set of possible consequences for each alternative can be assigned with a numeric performance value (a utility value) and a probability of its occurrence. The initial formal verification of this rule is commonly credited to [13], and was initially intended as a model of the supposedly rational market actor’s behaviour in the game theory school of economic thought. The axioms (ordering axioms, independence axioms, continuity axioms, etc.) thus imply numerical representations of preferences and probabilities, and implied by the axioms are existence theorems stating that a utility function exists, and a uniqueness theorem stating that two utility functions, relative to a given preference ranking, are always affine transformations of each other. It is often argued that these results provide justification of PMEU.

However, the conclusion that the PMEU is logistically sufficient given the axioms, or whether or not the axioms themselves reflect the properties of a rational decision maker has not passed without criticism and debate. For instance, in [18] it is shown in a formal investigation that the relation between the utility principle and the axioms is not as strong as claimed, but rather that “an agent who endorses these axioms is not contradicting himself if he also accepts the utility principle” and that the PMEU is the simplest principle that is consistent with the axioms. Further, the use of a utility function for capturing all possible risk attitudes is not considered possible [25].

An important issue of PMEU based decision analysis is the elicitation of a decision-maker’s attitude towards risk. However, the elicitation of risk attitudes from human decision-makers is error prone and the result is highly dependent on the format and method used, see, e.g., [22]. This problem is even more evident when the decision situation involve catastrophic outcomes [20]. If not being able to elicit a properly reflecting risk attitude, we may have the situation that even if the evaluation of an alternative re-
sults in an acceptable expected utility, some consequences might be of a catastrophic kind so the alternative should be avoided in any case. Due to catastrophe aversion, this may be the case even if the probabilities of these consequences are very low. In such cases, the PMEU needs to be extended with other rules, and it has therefore been argued that a useful decision theory should permit a wider spectrum of risk attitudes than by means of a utility function only. A more pragmatic approach should give an agent the means for expressing risk attitudes in a variety of ways, as well as provide procedures for handling both qualitative and quantitative aspects.

For the decision analysis field, having a more pragmatic approach than a purely normative theory of rational choice, the PMEU is deemed as sufficient in order to serve as a valuable tool for comparing decision alternatives, see, e.g., [12]. Also, in [19] the performance of a number of various decision rules is investigated, including that of [17], and it is concluded that from a decision analysis perspective there is really no other rule better suited to serve as the underlying basic decision rule.

In the context of catastrophic events, the partitioned multi-objective risk method (PMRM) and the use of conditional expected utilities for modelling of decision problems in low-likelihood severe-consequence domains have been suggested in [15, 11]. This approach puts emphasis on the tails of probability distributions over different kinds of values at risk in the case of a catastrophic scenario; still not discarding the unconditional probability of an extreme event actually occurring, but treating both the conditional and unconditional expected values as decision objectives. The approach outlined in this paper is nevertheless related to conditional expected utility approaches toward decision analysis in the face of extreme events, since we are concerned with consequences that have a very low probability of actually occurring. However, the approach taken herein is that although there is no rule deemed to be better than the PMEU, there is a need to allow the use of complementary rules in applications of decision analysis, rules also acknowledging that decision data often is subject to imprecision. It is argued that a useful decision theory should permit a wider spectrum of the modelling of risk attitudes than merely by means of a single utility function. A more pragmatic approach should give the decision-maker the means to express risk attitudes in a variety of ways, as well as provide procedures for handling both qualitative and quantitative aspects.

2 Modelling the Decision Problem

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. Further, the representational issues are of two kinds; a decision structure, modelled by means of a conventional decision analysis decision tree, i.e. a graph structure $(V, E)$ where $V$ is a set of nodes and $E$ is a set of node pairs (edges). We also have input statements, modelled by linear constraints.

Definition 1 A tree is a connected graph without cycles. A decision tree is a tree containing a finite set of nodes which 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 $\max(n)|\text{there exists a node at level } n|$.

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 $x_i$ are denoted $x_{i1}, x_{i2}, \ldots, x_{in}$ and the $m$ children of the node $x_{ij}$ are denoted $x_{ij1}, x_{ij2}, \ldots, x_{ijm}$ and so forth. For representational purposes, we will denote a consequence node of an alternative $A$, simply with $C_{ij}$.

Over each set of event nodes children probability distributions are defined, and over the set of consequence nodes utility functions are defined.

2.1 Interval Statements

For numerically imprecise decision situations, an elaborate modelling approach is to define sets of candidates of possible probability distributions and utility functions and then express these as points in polytopes that are solution sets to, so called, probability and utility bases [3]. For instance, the probability (or utility) of $C_{ij}$ being between the numbers $a_k$ and $b_k$ is expressed as $p_{ij} \in [a_k, b_k]$ (or $u_{ij} \in [a_k, b_k]$). Such an 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 2 Given a decision tree $T$, a utility base is a set of linear constraints of the types $u_{ij} \in [a_k, b_k]$,
where \( m \) note probability variables and \( p \) constraints as such a complementary decision rule. The next section discusses risk.

**Definition 3** Given an information frame \( \mathcal{I} = \langle T, \mathcal{P}, \mathcal{U} \rangle \) and an alternative \( A_i \in A \) the expression

\[
E(A_i) = \sum_{i_1=1}^{n_0} p_{i_1} \sum_{i_2=1}^{n_1} p_{i_1i_2} \cdots \sum_{i_{m-1}=1}^{n_{m-2}} p_{i_1i_2 \cdots i_{m-2}i_{m-1}} \sum_{i_m=1}^{n_{m-1}} p_{i_1i_2 \cdots i_{m-2}i_{m-1}i_m} u_{i_1i_2 \cdots i_{m-2}i_{m-1}i_m}
\]

where \( m \) is the depth of the tree corresponding to \( A_i \), \( n_{ij} \) is the number of possible outcomes following the event with probability \( p_{ij} \), \( j \in \{1, \ldots, m\} \), denote probability variables and \( u_{ij} \) denote utility variables as above, is the expected utility of alternative \( A_i \) in \( \mathcal{I} \).

The alternatives in the tree are evaluated according to PMEU, and when so the resulting expected utilities yield a preference ordering of the alternatives such that \( A_i \) is not preferred to \( A_j \) if and only if \( E(A_i) \leq E(A_j) \). Note that interval statements can yield a partial order due to overlapping expected utility intervals. However, as discussed in the introduction, the use of utility functions to formalise the decision process seem to be an oversimplified idea, disregarding factors typically present in real-life applications of decision analysis. Therefore, there is a need to permit the use of additional ways to discriminate between alternatives. The next section discusses risk constraints as such a complementary decision rule.

### 3 Risk Constraints

The intuition behind risk constraints is that they express when an alternative is undesirable due to too risky consequences. It is intended as a pragmatic approach to model catastrophe aversion in decision analysis applications, and builds upon the idea of providing thresholds beyond which an alternative is deemed undesirable by the decision making agent. Thus, expressing risk constraints is analogous to expressing minimum requirements that should be fulfilled. A risk constraint can be viewed as a function stating a set of thresholds that may not be violated in order for an alternative to be acceptable with respect to risk [2].

Thus, a decision agent might regard an alternative as undesirable if it has consequences with too low a utility and with some probability of occurring, although if those particular consequences’ contribution to the expected utility is low. This procedure is fairly straightforward. Assuming a 1-level tree, an alternative \( A_i \) in an information frame \( \mathcal{I} \), given a utility threshold \( r' \) and a probability threshold \( s' \), then

\[
\sum_{i} p_{ij} \leq s'
\]

must be satisfied in order for \( A_i \) to be deemed an acceptable alternative. In this sense, a risk constraint can be considered a utility-probability pair \((r', s')\). A consequence \( C_{ij} \) is said to be violating \( r' \) if \( u_{ij} > r' \) does not always hold. Principles of this kind seem to be good prima facie candidates for evaluative principles in the literature, i.e., they conform well to established practices and enable a decision-maker to use qualitative assessments in a reasonable way. For a comprehensive treatment and discussion, see [8, 6].

However, when the information is numerically imprecise (i.e. probabilities and utilities are expressed as bounds or intervals), it is not obvious how to interpret such thresholds since a risk constraint may cease to be violated in sub-sets of the solution set. We have earlier suggested that the interval boundaries together with stability analyses could be considered in such cases [7].

**Example 1** An alternative \( A_1 \) is considered undesirable if the consequence \( c_{1j} \) belonging to \( A_1 \) has a possibility that the utility of \( c_{1j} \) is less than 0.45, and if the probability of \( c_{1j} \) is greater than 0.65. Assume that alternative \( A_1 \) has a consequence for which its utility lies in the interval \([0.40, 0.60]\). Further assume that the probability of this consequence lies in the interval \([0.20, 0.70]\). Since 0.45 is greater than the least possible utility of the consequence, and 0.65 is less than the greatest possible probability, \( A_1 \) violates the thresholds and is thus undesirable.

---

**Advances in Economics, Risk Management, Political and Law Science**

ISBN: 978-1-61804-123-4
Figure 1: Small decision tree referred to in Example 1.

Figure 2: Contraction analysis of risk constraints given in Example 1. Beyond a contraction level of 14%, the constraints are no longer violated for alternative $A_1$. The constraints for alternative $A_2$ are never violated. A decision making agent might nevertheless be inclined to accept the alternative since the constraints are violated in a small enough proportion of the possible values.

For a stability analysis, it can be seen that the alternative in Example 1 ceases to be undesirable when the left end-point of the utility interval is increased by 0.05. A concept in line with such stability analyses is the concept of interval contraction, investigating to what extent the widths of the input intervals need be reduced in order for an alternative not to violate the risk constraints. The contractions of intervals are done toward a contraction point for each interval. Contraction points can either be given explicitly by the decision making agent or be suggested from, e.g., centre of mass calculations. The level of contraction is indicated as a percentage, where at 100% contraction all intervals have been replaced with their contraction points, see Figure 1 for a contraction analysis of the rudimentary problem in Example 1. Thus, one refinement is to provide a possibility for an agent to stipulate thresholds for proportions of the probability and utility bases, i.e., an alternative is considered unacceptable if it violates the risk constraints at a given contraction level [2, 16].

4 Second-Order Information

In interval-valued decision trees, the expected utility of an alternative will also become interval-valued. It is then in real-life cases often hard to discriminate between the alternatives since the intervals overlap, i.e., they are not narrow enough. For instance, an interval based decision procedure keeps all alternatives with overlapping expected utility intervals, even if the overlap is small. Therefore, it is worthwhile to extend the representation of the decision situation using more information, such as second-order distributions over classes of probability and utility measures.

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 points in the polytopes. Approaches for extending the interval representation using distributions over classes of probability and value measures have been developed into various such models, for instance second-order probability theory. In the following, we will pursue the idea of adding more information and discuss its implications on risk constraints.

4.1 Distributions over Information Frames

Interval estimates and relations 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 base 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 4 A second-order distribution over a unit cube $B$ is a positive distribution $F$ defined on $B$ 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 $BD(B)$.

We consider second-order probabilities to be an important sub-class of these distributions. Second-order probabilities will be used below as a measure of belief, i.e., a second-order joint probability distribution. Such distributions can then be defined over the information frame polytopes. 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 incoherent probability distribution over three mutually exclusive outcomes such as \((0.45, 0.25, 0.4)\). For this purpose, a suitable and general way of modelling random probabilities is the Dirichlet distribution.

**Definition 5** Let the notation be as above. Then the probability density function of the Dirichlet distribution is defined as

$$f_{Dir}(p, \alpha) = \frac{\Gamma \left( \sum_{i=1}^{n} \alpha_i \right)}{\prod_{i=1}^{n} \Gamma(\alpha_i)} \frac{\prod_{i=1}^{n} p_i^{\alpha_i - 1}}{1 - \prod_{i=1}^{n} p_i}$$

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 \(\alpha_i\) 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.

If the support of a belief distribution is restricted to a subset \(A\) of a unit cube \(B\), for instance when \(A\) is a polytope within \(B\) having equal number of dimensions as \(B\), beliefs in different points or subsets can be represented by a function defined on the specific subset. However, if we want to represent the belief on a subset which is of lower dimension than the unit cube itself we cannot use distributions that are upper bounded since a mass under such a distribution will be bound in a subset which is of lower dimension than the unit cube itself we cannot use distributions that are upper bounded since a mass under such a distribution will be

$$\int_{A} f(x) \, dV_B(x)$$

This issue can be solved by the characteristic distribution for \(A\).

**Definition 6** Let \(A\) be a subset of a unit cube \(B\), and let \(f\) be a belief distribution in \(A\). The natural extension \(f_A(x)\) of \(f\) with respect to \(A\) is defined by

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

**Definition 7** Let \(A\) be a subset of \(B\). A distribution \(g_A\) over \(B\) is called a characteristic distribution for \(A\) in \(B\) if

$$f(p) = \int_B \delta_p(x) f_A(x) g_A(x) \, dV_B(x)$$

for every belief distribution \(f\) over \(A\), and for every point \(p\) in \(A\), where \(\delta_p(x)\) is the Dirac delta distribution with pole at \(p\).

For a more comprehensive treatment of these properties, see [10]. Now with respect to second-order probabilities, 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 \(\hat{f}_A(x) \cdot g_A(x)\) is a belief distribution over \(B\) and equals 0 outside \(A\).

### 4.2 Marginal Distributions

In general, marginal distributions are obtained from the joint ones in the usual way.

**Definition 8** 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 \(b_i\).

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

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

**Example 2** The marginal distribution \(f(x_i)\) of the uniform Dirichlet distribution in a 4-dimensional cube is

$$f(x_i) = \int_{0}^{1-x_i} \int_{0}^{1-y-x_i} 6 \, dz \, dy = 3(1 - 2x_i + x_i^2)$$

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

This tendency is the result of a general phenomenon that becomes more emphasised as the dimension increases.
4.3 Aggregations and Expected Utility

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 by means of PMEU 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).

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. $x_{ij1}, x_{ij2}, \ldots, x_{ijn}$ (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 9** The cumulative distribution of the two belief distributions $f(x)$ and $g(y)$ is

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

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

$$h(z) = \frac{d}{dz} \int_0^z f(x)G(z/x) \, dx = \int_0^1 \left( \frac{f(x)g(z/x)}{x} \right) dx.$$

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

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

Informally, it means that the beliefs of $x$ and $y$ are multiplied and then added for all values $z = x \cdot y$. The addition of such products is analogous to the product rule for first-order (standard) probabilities. Similarly, addition is the ordinary convolution of two densities.

Using these combination rules, there are two main cases. The linearly independent case (utility variables) and the linearly dependent case (probability variables).

Assume that the assertions (statements) involved are made through intervals and that the constraint sets are linearly independent. If the marginal belief distributions are uniform, the multiplications of uniform distributions over intervals $[0, 1]$ result in the following theorem.

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

$$h_m(z_m) = \frac{(-1)^{m-1}(\ln(z_m))^{m-1}}{(m-1)!}.$$

The mass of the resulting belief distributions becomes more concentrated to the lower values, the more factors are involved. Indeed, already after one multiplication, this effect is clearly seen $(-\ln(z))$.

From initially uniform distributions, the resulting distribution has quite differing properties. It still has a non-zero support on the entire interval $[0, 1]$, but the relative beliefs in the various feasible points are shifted towards the lower bound. For instance, the belief distribution $-\ln(z)/3$ has about 97% of its belief mass over the sub-interval $[0, 0.3]$, a result that deviates from the initial equal belief in the entire interval $[0, 1]$.

One important effect is that multiplied distributions become considerably warped compared to the corresponding component distributions [9]. Such multiplications occur in obtaining the expected utility in decision trees, 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, see [24].

This warp effect can be exploited in decision evaluation, and evaluation of decision trees with respect to PMEU using second-order distributions is discussed in [5]. 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. Comparing two alternatives, we can study the belief mass for $E(A_i) - E(A_j) > 0$ and compare it with the belief
mass for \( E(A_j) - E(A_i) > 0 \). In addition, effects from varying belief, i.e. differing forms of belief distribution can be taken into account. The result is a method that can offer more discriminating power in selecting alternatives where overlap prevails, as the method may compare expected utility sub-intervals where the second-order belief mass is kept under control.

The generalisation of risk constraints in second-order decision analysis is rather straightforward. The basic idea is to consider the actual proportions of the resulting distributions that the thresholds cut off. As an analysis using risk constraints is done investigating one alternative at a time, we let a utility cube with respect to an alternative be denoted by \( B_{U_i} \) and a probability unit cube with respect to \( A_i \) be denoted by \( B_{P_i} \). Hence, \( B_{U_i} \) is represented by all the tuples \((u_{i1}, \ldots, u_{in})\) in \([0, 1]^n\) and \( B_{P_i} \) is represented by all the tuples \((p_{i1}, \ldots, p_{in})\) in \([0, 1]^n\) when \( A_i \) has \( n \) consequences. The normalisation constraint for probabilities imply that for a belief distribution over \( B_{P_i} \) there can be positive support only for tuples where \( \sum p_{ij} = 1 \).

**Definition 10** A probability unit cube for alternative \( A_i \) is a unit cube \( B_{P_i} = (p_{i1}, \ldots, p_{in}) \) where \( F_i(p_{i1}, \ldots, p_{in}) > 0 \Rightarrow \sum_{j=1}^{\infty} p_{ij} = 1 \). A utility unit cube for \( A_i, B_{U_i} \) lacks this latter normalisation.

As mentioned, one candidate for serving as a belief distribution over \( B_{P_i} \) is the Dirichlet distribution.

Given an information frame \( I \), a prima facie solution is to let \( f_{ij}^k(p_{ik}) \) and \( g_{ij}(u_{ij}) \) be marginal second-order distributions over the probabilities and utilities in \( I \), where the superscript \( k \) denotes the \( k \)th probability base in the frame. Furthermore, given a path \( R \) in the tree to the consequence \( c_{ij} \) and that \( p_{ij} = \prod_{k \in R} p_{ik} \) and assuming \( f_{ij}^k \) is the repeated multiplicative convolution over the distributions in \( R \), then, given thresholds \( r' \) and \( s' \) and second-order thresholds \( r'' \) and \( s'' \), where \( s', r', r'', s'' \in [0, 1] \), if

\[
\int_0^{r'} g_{ij}(u_{ij}) \, du_{ij} \geq r''
\]

and

\[
\int_{s'}^{1} f_{ij}(p_{ij}) \, dp_{ij} \geq s''
\]

are satisfied the alternative is deemed undesirable. Note that \( r' \) and \( s' \) are limits on actual utilities and probabilities respectively but \( r'' \) and \( s'' \) are limits on their distributions.

However, as for ordinary risk constraints, it is also necessary to take into account the way in which subsets of consequences, i.e. events, together can make an alternative undesirable. If we would have independent distributions in the probability base, this would be accomplished by using standard convolution, utilizing the product rule for standard probabilities. Due to inequality constraints, this approach must be modified.

Let \( \{g_{ij}(u_{ij})\}_{j=1}^n \) be marginal second-order distributions with respect to consequences \( \{C_{ij}\} \) of an alternative \( A_i \) in an information frame \( I = (T, \mathcal{P}, \mathcal{U}) \). Let \( \Phi_i \) be a consequence set such that

\[
C_{ij} \in \Phi_i \iff \int_0^{r'} g_{ij}(u_{ij}) \, du_{ij} \geq r''
\]

Further, let \( R_{ij}^s \) be the paths from alternative \( A_i \), in a decision tree, let \( f_{ij}^s \) be the resulting belief distribution over \( p_{ij}^s \), where \( C_{ij} \in \Phi_i \). Then

\[
t'' = \text{Convolute}(f_{ij}^s \in R_{ij}^s)
\]

where

\[
\Gamma_s = \{P_i : \sum_{C_{ij} \in \Phi_i} p_{ij} \geq s\}
\]

Then the inequality

\[
t'' \leq s''
\]

must be satisfied for the alternative to be acceptable. This is a straightforward generalisation of the risk constraint concept utilising second-order information. In addition to the utility-probability threshold pair \((r', s')\), we also use a pair \((r'', s'')\) acting as thresholds on the belief mass violating \( r' \) and \( s' \) respectively.

### 5 Summary and Conclusions

The various axiomatic systems proposed supporting the principle of maximising the expected utility are insufficient and have been subject to severe criticism. Due to this and from pragmatic issues in employing this principle as the rule for rational choice, it is worthwhile to supplement frameworks based on the utility principle with other decision rules taking a wider spectrum of risk attitudes into account. One such supplement is the inclusion of thresholds in the form of risk constraints.

Using the concept of risk constraint violation, a general model can be constructed for representing various risk attitudes and providing alternative means for expressing such. The definitions are computationally
meaningful, and are therefore also well suited to automated decision making. Rules have been suggested for sorting out undesirable decision alternatives, rules which should also serve as a tool for guaranteeing that certain norms are not violated.

Acknowledgements: The research was funded in part by the Swedish research council Formas, grant 2011-3313-20412-31, and by Strategic funds from the Swedish government within ICT – The Next Generation.

References: