

Second Order Effects in Interval Valued Decision Graph Models

Love Ekenberg^{1,3}, Mats Danielson², and Aron Larsson³

¹Dept. of Computer and Systems Sciences, Stockholm University / KTH, Forum 100, SE-164 40 Kista, Sweden

²Dept. of Informatics / ESI, Örebro University, SE-701 82 Örebro, Sweden

³Dept. of Information Technology and Media, Mid Sweden University, SE-851 70 Sundsvall, Sweden

lovek@dsv.su.se, mad@dsv.su.se, aron.larsson@miun.se

Abstract

Second-order calculations may significantly increase a decision maker's understanding of a decision situation when handling aggregations of imprecise representations, as is the case in decision trees or influence diagrams, while the use of only first-order results gives an incomplete picture. The results apply also to approaches which do not explicitly deal with second-order distributions, instead using only first-order concepts such as upper and lower bounds.

Introduction

In decision analysis, it is often the case that complete, adequate, and precise information is missing. The requirement to provide numerically precise information in such models has often been considered unrealistic in real-life decision situations. Consequently, during recent years of rather intense activities in the decision analysis area (see, e.g., (Klir 1999; Cano and Moral 1999; Weichselberger 1999)) several approaches have emerged. In particular, first-order approaches, i.e., approaches based on sets of probability measures, upper and lower probabilities, and interval probabilities, have prevailed. However, the latter still do not admit for discrimination between different beliefs in different values. This seems unnecessarily restrictive since a decision maker does not necessarily believe with the same faith in all possible functions that the vectors represent, i.e., in all points between the upper and lower bounds. Furthermore, leaving out second-order information may lead to severely warped evaluation results not evident from an upper and lower bound analysis. This also troubles classical utility theory as well as the various relaxations that have been suggested.

To allow for estimates which closer models the decision maker's beliefs, representations of decision situations could involve beliefs in sets of epistemically possible value and probability functions, as well as relations between them. Beliefs of such kinds can be expressed using higher-order belief distributions. However, they have usually been dismissed for various reasons, computational and conceptual, and approaches based on sets of probability measures, upper and lower probabilities, or interval probabilities have traditionally been preferred. We demonstrate in this paper and in

(Larsson, Danielson, and Ekenberg 2005) why second-order reasoning is useful and how second-order effects can be taken into account also when handling aggregations of first-order representations, such as occurring in decision trees or probabilistic networks.

Preliminaries

Decisions under risk (probabilistic decisions) are often given a tree representation, c.f. (Raiffa 1968). In this paper, we let a decision frame represent a decision problem. The idea with such a frame is to collect all information necessary for the model in one structure. One of the building blocks of a frame is a decision tree. Formally, a decision tree is a graph.

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

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

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

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

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

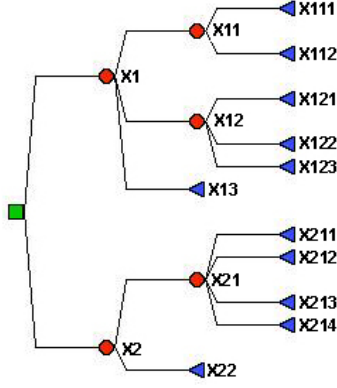


Figure 1 A decision tree.

There are two sources for constraints, the first source being decision maker statements (of probabilities and values). The statements are translated into corresponding constraints. Such constraints can either be range constraints (containing only one variable) or various kinds of comparative constraints. Given consequences c_i, c_j, c_k , and c_m , denote their values v_i, v_j, v_k , and v_m . Then user statements can be of the following kinds for real numbers a_1, a_2, d_1, d_2, d_3 , and d_4

- **Range constraints** “ v_i is between a_1 and a_2 ” is denoted $v_i \in [a_1, a_2]$ and translated into $v_i \geq a_1$ and $v_i \leq a_2$.
- **Comparative constraints** several possibilities; examples include “ v_i is between d_1 and d_2 larger than v_j ” is denoted $v_i - v_j \in [d_1, d_2]$ and translated into $v_i - v_j \geq d_1$ and $v_i - v_j \leq d_2$. The difference “ $v_i - v_j$ is between d_3 and d_4 larger than $v_k - v_m$ ” is denoted $(v_i - v_j) - (v_k - v_m) \in [d_3, d_4]$ and translated into $(v_i + v_m) - (v_j + v_k) \geq d_3$ and $(v_i + v_m) - (v_j + v_k) \leq d_4$.

The other source for constraints is implicit constraints. They emerge either from properties of the variables or from structural dependencies. Such constraints can either be default constraints (involving a single variable) or various kinds of structural constraints (involving more than one variable).

- **Default constraints** “The range of v_i is between 0 and 1” is denoted $v_i \in [0, 1]$ and translated into $v_i \geq 0$ and $v_i \leq 1$.
- **Structural constraints** Examples include the constraint implied by the normalization constraint for probabilities, $\sum_j p_{ij} = 1$.

Combining these two sources, constraint sets are obtained. A constraint set can either be independent (containing only constraints involving a single variable each), or it can be dependent (also containing constraints involving more than one variable). In this paper, we will only treat independent constraint sets¹.

Definition 3 Given a decision tree T , let P be a set of constraints in the variables $\{p_{\dots i \dots j \dots}\}$. Substitute the intermediate node labels $x_{\dots i \dots j \dots}$ with $p_{\dots i \dots j \dots}$. P is a probability

¹For space reasons, we focus on interval probabilities and values and we do not explicitly cover relations in value constraint sets. However, the aggregated distributions over such constraint sets have the same properties as the distributions discussed herein.

constraint set for T if, for all sets $\{p_{\dots i1}, \dots, p_{\dots im}\}$ of all sub-nodes of nodes $p_{\dots i}$ that are not leaves, the statements $p_{\dots ij} \in [0, 1]$ and $\sum_j p_{\dots ij} = 1, j \in [1, \dots, m_{\dots i}]$ are in P , where $m_{\dots i}$ is the number of sub-nodes to $p_{\dots i}$.

Thus, a probability constraint set relative to a decision tree can be seen as characterizing a set of discrete probability distributions. The normalization constraints ($\sum_j p_{ij} = 1$) require the probabilities of sets of exhaustive and mutually exclusive nodes to sum to one.

Definition 4 Given a decision tree T , let V be a set of constraints in $\{v_{\dots 1}\}$. Substitute the leaf labels $x_{\dots 1}$ with $v_{\dots 1}$. Then V is a value constraint set for T .

Similar to a probability constraint set, a value constraint set can be seen as characterizing a set of value functions. In this paper, we will assume that the value variables’ ranges are $[0, 1]$. The elements above can be employed to create the decision frame, which constitutes a complete description of the decision situation.

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

Evaluation

The probability and value constraint sets are collections of linear inequalities. A minimal requirement for such a system of inequalities to be meaningful is that it is consistent, i.e., there must be some vector of variable assignments that simultaneously satisfies each inequality in the system.

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

Definition 6 Given a decision frame $\langle T, P, V \rangle$, $GEV(A_i)$ denotes the generalized expected value of alternative A_i and is obtained from

$$\sum_{i_1=1}^{n_{i_0}} p_{ii_1} \sum_{i_2=1}^{n_{i_1}} p_{ii_1 i_2} \dots \sum_{i_{m-1}=1}^{n_{i_{m-2}}} p_{ii_1 i_2 \dots i_{m-2} i_{m-1}} \dots \sum_{i_m=1}^{n_{i_{m-1}}} p_{ii_1 i_2 \dots i_{m-2} i_{m-1} i_m} \cdot v_{ii_1 i_2 \dots i_{m-2} i_{m-1} i_m}$$

where $p_{\dots i j \dots}, j \in [1, \dots, m]$, denote probability variables in P and $v_{\dots i j \dots}$ denote value variables in V .

Maximization of such non-linear expressions subject to linear constraints (the probability and value constraint sets) are computationally demanding problems to solve for an interactive tool in the general case, using techniques from the area of non-linear programming. In, e.g., (Danielson and Ekenberg 1998), (Ding, Danielson, and Ekenberg 2004), and (Danielson 2004), there are discussions about computational procedures that reduce such non-linear decision evaluation

problems to systems with linear objective functions, solvable with ordinary linear programming methods. The procedures yield interval estimates of the evaluations, i.e., upper and lower bounds of the expected values for the alternatives. They also include methods for separating alternatives resulting in overlapping expected value intervals. This is a first step in an analysis but more can be done. Regardless of the assumptions made on the decision maker's belief in the various parts of the input intervals, the evaluation should continue with a further analysis of the intervals obtained.

Second-Order Belief Distributions

Approaches for extending the interval representation using distributions over classes of probability and value measures have developed into various hierarchical models, such as second-order probability theory (Gärdenfors and Sahlin 1982; Gärdenfors and Sahlin 1983; Ekenberg and Thorbiörnson 2001; Ekenberg, Thorbiörnson, and Baidya 2005). Gärdenfors and Sahlin consider global distributions of beliefs, but restrict themselves to the probability case and to interval representations. Other limitations are that they neither investigate the relation between global and local distributions, nor do they introduce methods for determining the consistency of user-asserted sentences (Ekenberg 2000). The same criticism applies to (Hodges and Lehmann 1952), (Hurwicz 1951), and (Wald 1950).

To facilitate better qualification of the various possible functions, second-order estimates, such as distributions expressing various beliefs, can be defined over an n -dimensional space, where each dimension corresponds to possible probabilities of events or utilities of consequences. In this way, the distributions can be used to express varying strength of beliefs in different first-order probability or utility vectors.

Definition 7 Let a unit cube be represented by $B = [0, 1]^k$. A belief distribution over B is a positive distribution F defined on B such that

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

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

Example 1 Assume that the function

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

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

Example 2 The functions $f_{b_1} \rightarrow b_1$ and $h_{b_2} \rightarrow b_2$ are belief distributions over the one-dimensional unit cubes (b_1) and (b_2) respectively defined by

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

and

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

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

Local Distributions

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

Definition 8 Let $B = (b_1, \dots, b_k)$ and $A = (b_{i_1}, \dots, b_{i_s}), i_j \in \{1, \dots, k\}$ be unit cubes. Let $F \in BD(B)$, and let

$$f_A(x) = \int_{B \setminus A} F(x) dV_{B \setminus A}(x)$$

Then f_A is the S-projection of F on A .

Theorem 1 Given a unit cube $B = (b_1, \dots, b_k)$ and a belief distribution $F \in BD(B)$, let $f_A = Pr_A(F)$, then $f_A(x') \in BD(b_{i_1}, \dots, b_{i_s}), i_j \in \{1, \dots, k\}$.

Thus, Theorem 1 shows that an S-projection of a belief distribution is also a belief distribution. A special kind of projection is when belief distributions over the axes of a unit cube B are S-projections of a belief distribution over B .

Definition 9 Given a unit cube $B = (b_1, \dots, b_k)$ and a distribution $F \in BD(B)$. Then the distribution $f_i(x_i)$ obtained by

$$f_i(x_i) = \int_{\bar{B}_i} F(x) dV_{\bar{B}_i}(x)$$

where $\bar{B}_i = (b_1, \dots, b_{i-1}, b_{i+1}, \dots, b_k)$, is a belief distribution over the b_i -axis. Such a distribution will be referred to as a local distribution.

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

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

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

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

Centroids

In the sequel, we will use the concept of centroids. Intuitively, the centroid of a distribution is a point in space where some of the geometrical properties of the distribution can be regarded as concentrated. This is, in some respects, analogous to the center of mass of physical bodies. It will below turn out to be a good representative of the distribution in various calculations.

Definition 10 Given a belief distribution F over a cube B , the centroid F_c of F is

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

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

Centroids are invariant under projections on subsets of the unit cubes in the sense that the S-projections of a centroid on a subset have the same coordinates as the centroids of the corresponding S-projections (Ekenberg, Thorbiörnson, and Baidya 2005). Thus, a local distribution of a belief distribution preserves the centroid in that dimension.

Example 5 The centroid f_c of the local distribution given in Example 4 is

$$f_c = \int_0^1 x \cdot (2 - 2x) dx = \frac{1}{3}$$

i.e., the center of mass in an ordinary triangle.

Multiplication of Distributions

The expected utility of the alternatives represented by a classical decision tree are straightforwardly calculated when all components are numerically precise. When the domains of the terms are solution sets to probability and value constraint sets, this is not as straightforward, but there are other methods available (Danielson et al. 2003), (Danielson 2005). Either the decision maker has belief distributions in mind or not when making interval assignments. In the presentation below, we assume that the beliefs in the feasible values are uniformly distributed and we show the effects when multiplying variables in trees as in the calculation of the expected value $GEV(A_i)$.²

Let G be a belief distribution over the two cubes A and B . Assuming that G has a positive support on the feasible probability distributions at level i in a decision tree, i.e., is representing 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., $x_{ij1}, x_{ij2}, \dots, x_{ijm}$ (the support of G in cube B). Let $f = Pr_A(G)$ and $g = Pr_B(G)$. Then the functions f and g are belief distributions according to Theorem 1. Furthermore, there are no relations between two probabilities at different levels (having different parent nodes) so the distributions f and g are independent. Consequently, the following combination rule for the distribution over the product of the distribution f and g has a well-defined semantics.

²Other conceivable distributions have similar properties, but in order to keep the presentation clear, we focus on uniform belief.

Definition 11 The product of two belief distributions $f(x)$ and $g(x)$ is

$$h(z) = \int_{\Gamma_z} f(x)g(y) ds$$

where $\Gamma_z = \{(x, y) \mid x \cdot y = z\}$ and $0 \leq z \leq 1$.

Let us now consider the relation to traditional (1st order) interval calculus. When aggregating interval probabilities and values in a tree as above, there are two main cases to consider.

- The constraints can be linearly independent as in a value constraint set without equality statements.
- The other interesting case is linear dependencies, such as in a probability constraint set where the probabilities of disjoint events must add 1.³

In the first case, by the assumption the distributions over the intervals (i.e. the local distributions) are considered to be uniform over the respective axes. Assume that we have constraint sets where the constraints are linearly independent. If the assertions (statements) are made through intervals (range constraints) there are several options for the distributions over them and usually they are not well known. Below, we will discuss the case when the belief in all feasible points could be considered equal, i.e., the local belief distributions are uniform, $f(x) = g(y) = 1$, over the intervals $[0, 1]$. Needless to say, these are not at all the true distributions. However, we use these to illustrate the general tendencies, which will be the same in all reasonable cases.

Theorem 2 Let $f_1(x_1), \dots, f_m(x_m)$ be belief distributions over the intervals $[0, 1]$. The product $h_m(z_m)$ over these m factors is the distribution

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

Proof.

$$\begin{aligned} h_m(z_m) &= \int_{\Gamma_m} f_m(x_m) \cdot \\ &\left(\dots \left(\int_{\Gamma_3} f_3(x_3) \left(\int_{\Gamma_2} f_2(x_2) f_1(x_1) ds_2 \right) ds_3 \right) \dots \right) ds_m = \\ &\int_{\Gamma_m} \dots \int_{\Gamma_3} \int_{\Gamma_2} ds_2 ds_3 \dots ds_m = \frac{(-1)^{m-1}(\ln(z_m))^{m-1}}{(m-1)!} \end{aligned}$$

□

Theorem 3 The centroid of the distribution $h_m(z_m)$ in Theorem 2 is 2^{-m} .

Proof.

$$\int_0^1 z_m \frac{(-1)^{m-1}(\ln(z_m))^{m-1}}{(m-1)!} dz_m = 2^{-m}$$

□

³This case is beyond the scope of this paper.

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

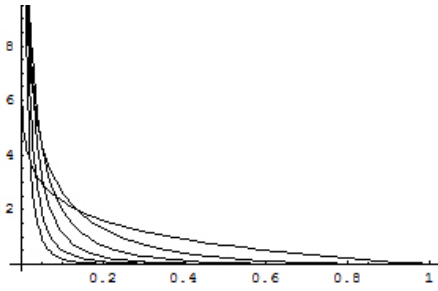


Figure 2 Multiplication of distributions of 7, 6, 5, 4, 3, and 2 consecutive node values.

As mentioned above, this effect is not dependent on the assumption of uniform distribution.

The important observation above is that the mass of the resulting belief distributions becomes dramatically more concentrated to the lower values the deeper the tree is and the more factors that are aggregated in the expected value (the dual warp effect). Already after one multiplication, this effect is significant. It should be regarded as additional information by any method employing an interval calculus.

As can be seen from the results above, in general, the effects of this are considerable when evaluating imprecise decision problems. Inevitably, the most important sub-intervals to consider are the supports of the distributions where the most mass is concentrated. This can be compared to the ordinary multiplication of extreme points (bounds) which would generate an interval $[0, 1]$. Consequently, an important component in any method for decision tree analysis is the possibility of determining belief-dense sub-intervals.

This warp effect does not imply that the extreme points themselves are wrong. Interval methods in general are unable to give any qualification on the belief distribution. This does neither imply that algorithms for determining upper and lower bounds in trees are inappropriate, but the results should be supplemented by second-order and centroid calculations.

Conclusion

In the literature, there has been a debate whether or not various kinds of second-order approaches are better suited than first-order ones for modelling incomplete knowledge. In this paper, we show effects of multiplying interval estimates in decision trees. We have demonstrated that second-order belief adds information when handling aggregations of interval representations, such as in decision trees or probabilistic networks, and that interval estimates (upper and lower bounds) in themselves are incomplete.

The results apply equally to all approaches which do not explicitly deal with belief distributions. Focussing only on first-order concepts does not provide the complete picture. The second-order effects are still present regardless of the

precise beliefs of the decision maker. The rationale behind this fact is that we have demonstrated that multiplied distributions sharpen (warp) significantly compared to their component distributions. Secondly, the multiplied distributions dramatically concentrate their mass to the lower values compared to their component distributions. This also means that due to the dual warp effect, calculations using the centroid, instead of the complete intervals, provide a very good estimate already at quite shallow tree depths. Thus, while a complete use of second order information is complicated, the centroid is a very good candidate for practical purposes.

References

- Cano, A. and Moral, S. 1999. A Review of Propagation Algorithms for Imprecise Probabilities. Proceedings of ISIPTA'99.
- Danielson, M. 2004. Handling Imperfect User Statements in Real-Life Decision Analysis. *International Journal of Information Technology and Decision Making* 3/3:513-534.
- Danielson, M. 2005. Generalized Evaluation in Decision Analysis. *European Journal of Operational Research* 162/2:442-449.
- Danielson, M. and Ekenberg, L. 1998. A Framework for Analysing Decisions under Risk. *European Journal of Operational Research* 104/3:474-484.
- Danielson, M., Ekenberg, L., Johansson J., and Larsson, A. 2003. The *DecideIT* Decision Tool. Proceedings of ISIPTA'03:204-217. Carleton Scientific.
- Ding, X. S., Danielson, M., and Ekenberg, L. 2004. Non-linear Solvers for Decision Analysis. Selected Papers of the International Conference on Operations Research (OR 2003):475-482.
- Ekenberg, L. 2000. The Logic of Conflict between Decision Making Agents. *Journal of Logic and Computation* 10/4:583-602.
- Ekenberg, L. and Thorbiörnson, J. 2001. Second-Order Decision Analysis. *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems* 9/1:13-38.
- Ekenberg, L., Thorbiörnson, J., and Baidya, T. 2005 Value Differences using Second Order Distributions. *International Journal of Approximate Reasoning* 38/1:81-97.
- Gärdenfors P. and Sahlin, N-E. 1982. Unreliable Probabilities, Risk Taking, and Decision Making. *Synthese* 53:361-386.
- Gärdenfors P. and Sahlin N-E. 1983. Decision Making with Unreliable Probabilities. *British Journal of Mathematical and Statistical Psychology* 36:240-251.
- Hodges, J. L. and Lehmann, E. L. 1952. The Use of Previous Experience in Reaching Statistical Decisions. *The Annals of Mathematical Statistics* 23:396-407.
- Hurwicz, L. 1951. Optimality Criteria for Decision Making under Ignorance. Cowles Commission Discussion Paper 370.

Klir, G. J. 1999. Uncertainty and Information Measures for Imprecise Probabilities An Overview. Proceedings of ISIPTA'99.

Larsson, A., Danielson, M., and Ekenberg, L. 2005. Non-Uniform Belief in Expected Utilities in Interval Decision Analysis. Proceedings of FLAIRS'05. AAAI Press.

Raiffa, H. 1968. *Decision Analysis*. Addison Wesley.

Wald, A. 1950. *Statistical Decision Functions*. John Wiley and Sons.

Weichselberger, K. 1999. The Theory of Interval-Probability as a Unifying Concept for Uncertainty. Proceedings of ISIPTA'99.