Some Second Order Effects on Interval Based Probabilities

David Sundgren and
University of Gävle,
SE-801 76 Gävle, SWEDEN

Mats Danielson and **Love Ekenberg**
Dept. of Computer and Systems Sciences,
Stockholm University and KTH,
Forum 100, SE-164 40 Kista, SWEDEN

Abstract

In real-life decision analysis, the probabilities and values of consequences are in general vague and imprecise. One way to model imprecise probabilities is to represent a probability with the interval between the lowest possible and the highest possible probability, respectively. However, there are disadvantages with this approach, one being that when an event has several possible outcomes, the distributions of belief in the different probabilities are heavily concentrated to their centers of mass, meaning that much of the information of the original intervals are lost. Representing an imprecise probability with the distribution's center of mass therefore in practice gives much the same result as using an interval, but a single number instead of an interval is computationally easier and avoids problems such as overlapping intervals. Using this, we demonstrate why second-order calculations can add information when handling imprecise representations, as is the case of decision trees or probabilistic networks. We suggest a measure of belief density for such intervals. We also demonstrate important properties when operating on general distributions. The results herein 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

Imprecise probabilities are often modelled by intervals. There are a number of different approaches to deciding the interval boundaries; in (Choquet 1954) capacities were introduced, and then further developed in (Huber 1973) and (Huber & Strassen 1973). Capacities of order 2 can be used for interval probabilities (Denneberg 1994). Interval-valued probability functions have been based on classes of probability measures (Good 1962). The Dempster-Shafer theory (Dempster 1967), (Shafer 1976) provides a framework for modelling upper and lower probabilities. In (Walley 1991) interval based probabilities are thoroughly investigated. A geometric approach to interval-valued probabilities is taken in (Ha & Haddaway 1998).

In decision analysis, it is common to seek to maximize the expected utility. When the probabilities of the possible outcomes of an event are expressed using intervals $p_i \in [a_i, b_i]$, one can formulate the problem as a linear programming (LP) problem with the interval restrictions as constraints $p_i \geq a_i, p_i \leq b_i$. However, the optima of LP problems are found in the vertices of the feasible region, meaning that for each i, either interval boundary a_i or b_i will be used in the calculation of the optimal expected utility. The optimal solution then merely tells us what the expected utility would be if in fact $p_i = a_i$ or b_i, but the interval boundaries are not decision variables. Another issue is that often the decision-maker does not hold the interval boundary values to be very likely. It is for example rare that 20% is regarded as a reasonable probability while 19% is considered impossible.

Another approach for maximizing the expected utility is to compute the minimal and maximal expected utilities respectively, for each decision alternative, by using the upper and lower bounds for the probability intervals. If one decision alternative has a higher minimal expected utility than the maximal expected utility of another alternative, the latter is said to be dominated by the former. If one decision alternative dominates all the others, it is the one to choose. Typically, however, the respective alternatives' intervals of expected utility overlap.

To overcome this difficulty (Danielson & Ekenberg 1998) suggests *contracting* the probability intervals until one alternative dominates all the others. Contraction means that the intervals are narrowed towards the middle of the intervals, in the extreme case even to a single point between a_i and b_i. The underlying assumption here is that the decision maker believes that this contraction point is the best estimate of the probability with a symmetrical margin of error. This method can be modified by giving three numbers per probability, lower and upper bound and a value somewhere in between that is held to be the most likely, contraction is then made towards this latter value. There is, however, a certain degree of arbitrariness in the contraction procedure; the final probability interval used in the calculation of the expected utility is the one required to obtain an unambiguously optimal alternative.

Restricting the decision maker to just one value per probability is one way of solving the problems mentioned above. This approach might seem to disregard the reason for introducing intervals in the first place, namely the vagueness of the probability estimates. However, conceptually, imprecision may be reconciled with single-value probabilities if they are considered as the center of mass of a *second order* distribution; the single value demanded of the decision maker is simply the value he believes in the strongest. Employing the center of mass, or *centroid*, in this way is suggested in (Ekenberg & Thorbiörnson 2001). We argue that, in decision situations with many possible outcomes, most of the information in a second-order distribution is unnecessary and that it is enough to calculate the expected utility via the centroids, avoiding the (in general) cumbersome task of eliciting a second-order distribution through e.g. sampling (as in (Druzdzel & van der Gaag 1995)) and computing the corresponding distribution on the expected utility.

When the variables $x_i, i = 1, \ldots, n$ are restricted by intervals $[a_i, b_i]$, a n-dimensional polytope is formed. In decision analysis, where the maximal expected utility is sought, the x_i are either the probabilities p_i or utilities u_i of the possible outcomes of an event. In the case of probabilities the $n-1$-dimensional polytope formed by $x_i \in [a_i, b_i], i = 1, \ldots, n-1, (x_n = 1 - \sum_{i=1}^{n-1} x_i)$ is cut by the plane $\sum_{i=1}^{n} x_i = 1$.

A point in the polytope represents a probability distribution over the possible outcomes. Thus, the polytope is a subspace of the space of all possible probability distributions.

Below we shall see that in some cases there are strong second-order effects when all interval based probabilities $p_i \in [a_i, b_i]$ of the n possible outcomes of an event are considered under the fundamental restriction (normalization) $\sum_{i=1}^{n} p_i = 1$. When n is large and some of the intervals are wide, the resulting second-order distribution is warped towards the lower bound and most of the information of the intervals is lost. We suggest that the centroid is particularly suitable for representing a probability in such cases. When the intervals become more narrow as n grows, the second-order effects are more subtle and the centers of mass are closer to the midpoints of the intervals. In either case, the centroid is a reasonable choice for a single value representative of a probability.

Belief Distributions and some Properties

The basic idea of belief distributions is that a decision-maker does not necessarily have to believe as strongly in all possible functions that the points of the polytope represent. Distributions expressing various beliefs enable a differentiation of functions.

Definition 1. *Let a unit cube* $[0,1]^n$ *be given. By a* belief distribution over B, *we mean a positive distribution* g *defined on the unit cube* B *such that*

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

where V_B *is some* k*–dimensional Lebesque measure on* B.

Example 1. *Let*

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

be a function defined on the unit cube $[0,1]^2$. f *is a belief distribution over* $[0,1]^2$ *since the volume under the function surface* $x_3 = f(x_1, x_2)$ *is one.*

In this paper, we assume a belief distribution corresponding to equal belief in all points of the polytope. This 'uniform belief' assumption will be used for presentational purposes and does not in any way mean that it is the only valid interpretation of boundary approaches. This issue has been thoroughly debated over the years. Other approaches, such as preferences for and valuation of a gamble, lead to a decision-maker's support of the gamble, in which case the second-order distribution could be interpreted as support (either varying or uniform). Thus, while the concept 'belief' is used throughout the paper, it can be thought of as the decision-maker's support. And the existence of the projection effects discussed does not depend on uniform distributions; they are merely less complicated from a presentational point of view.

Belief distributions can also be used to represent subsets of a unit cube by considering the support of the distributions. However, when representing a subset of lower dimension than the unit cube itself, distributions that are upper bounded cannot be used, since a mass under such a distribution will be 0 while integrating with respect to some Lebesque measure defined on the unit cube. This particular problem is solved in detail in (Ekenberg & Thorbiörnson 2001).

An important task is to investigate the relationship between different distributions. In particular, we need to study the relationship between the, typically high-dimensional, background (global) distribution and projections of this on various sub-spaces. I.e., we need a semantics for this relationship – what do beliefs over some subset of a unit cube mean with respect to beliefs over the entire cube. One reasonable candidate for providing this semantics is provided by the concept of S-projections.

Definition 2. *Let* $B = [0,1]^k$ *and* $A = [0,1]^{i_s}, i_j \in \{1, \ldots, k\}$ *be unit cubes. Let* F *be a belief distribution on* 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.

The S-projection can be regarded as a second-order distribution on A.

In the sequel, we will only compute S-projections with respect to a single variable, i.e. A will have dimension one.

Definition 3. *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 Lebesque measure on* B.

As can be seen from the definition, a centroid is a center of mass generalized to arbitrary dimensions.

Here, we study distributions where $\sum_{i=1}^{n} x_i = 1$ and $x_i \in [a_i, b_i]$ for $i = 1, \ldots, n-1$, $x_n = 1 - \sum_{i=1}^{n-1}$. I.e., we can interpret $x_i, i = 1, \ldots, n$ as probabilities and we assume that belief is uniformly distributed over the polytope.

The Warp Effect of S-Projections

In order to see which second-order distributions on the single probabilities comply with a uniform distribution on the polytope $x_i \in [a_i, b_i](i = 1, \dots, n-1), \sum_{i=1}^{n} x_i = 1$, we will project the uniform distribution on one of the axis. That is, we will compute the S-projection $f(x_1)$ and centroid f_c of x_1. We choose x_1 w.l.o.g. since if one wishes to compute, say, the S-projection $f(x_3)$ or the corresponding centroid one can either switch x_1 and x_3 or replace x_1 in the formulas below with x_3. Since in our formulation only $x_1, \dots x_{n-1}$ have interval boundaries, we work with the $(n-1)$-dimensional polytope formed by $x_i \in [a_i, b_i]$ for $(i = 1, \dots, n-1)$ and $\sum_{i=1}^{n-1} x_i \le 1$.

We denote the $k \le 2^{n-1}$ vertices of the rectangular parallelepiped defined by $x_i \in [a_i, b_i], i = 1, \dots, n-1$ that fall inside the (hyper-)pyramid with apex in $(0, 0, \dots, 0)$ and base $\sum_{i=1}^{n-1} x_i = 1$ by $\sigma_j = \sum_{i=1}^{n-1} c_{i,j}, c_{i,j} \in \{a_i, b_i\}$, ordered such that $\delta_1 \ge \delta_2 \ge \dots \ge \delta_k$, where $\delta_j = 1 - \sigma_j + c_{1,j} = 1 - \sum_{i=2}^{n-1} c_{i,j}$. Further, for each vertex σ_j, we let P_j be the pyramid with apex in $(c_{1,j}, c_{2,j}, \dots, c_{n-1,j})$ and base $\sum_{i=1}^{n-1} x_i = 1$.

$f(x_1)$ is produced by integrating over the solid with respect to all variables except x_1 and dividing by its volume. f_c is then obtained by integrating $x_1 f(x_1)$ with respect to x_1. First we compute the solid's volume.

Lemma 1. *The volume of the solid resulting from the rectangular parallelepiped defined by $x_i \in [a_i, b_i], i = 1, \dots, n-1$ being cut by the plane $\sum_{i=1}^{n-1} x_i = 1$ is*

$$\frac{\sum_{i=1}^{k} (-1)^{m_i} (1-\sigma_i)^{n-1}}{(n-1)!},$$

where m_i is the number of terms b_j in σ_i.

Proof We first compute the volume of the pyramid P_1 with apex in $(a_1, a_2, \dots, a_{n-1})$ and then subtract the volumes of all the other $P_i, i = 2, \dots k$ since they are outside of the solid. But the pyramids P_i corresponding to σ_i with an even number m of b_j:s will have to be added, they are already removed twice since they are subsets of $\sum_{i=0}^{m-1} \binom{m}{i}$ removed pyramids; if $\sigma_i \le 1$ has a term b_j, there must be a $\sigma_p = \sigma_i + a_j - b_j \le 1$.

And the volume of the pyramid P_i is

$$\int_{c_{1,i}}^{1-c_{2,i}-c_{3,i}-\dots-c_{n-1,i}} \int_{c_{2,i}}^{1-x_1-c_{3,i}-\dots-c_{n-1,i}} \dots \int_{c_{n-1,i}}^{1-x_1-x_2-\dots-x_{n-1}} dx_{n-1} \dots dx_2\, dx_1 = \frac{(1-\sigma_i)^{n-1}}{(n-1)!},$$

completing the proof. □

Theorem 1. *With $\delta_{k+1} = a_1$ the S-projection $f(x_1)$ equals*

$$\frac{\sum_{i \in I_j} (-1)^{m_i} (n-1)(\delta_i - x_1)^{n-2}}{\sum_{i=1}^{k} (-1)^{m_i} (1-\sigma_i)^{n-1}},$$

where $I_j = \{i \in \mathbb{N} : 1 \le i \le j-1, c_{1,i} \ne b_1\}$, for $\delta_j \le x_1 \le \min(\delta_{j-1}, b_1), j = 2, \dots, k+1$.

Proof We divide the interval $a_1 \le x_1 \le 1 - a_1 - \dots - a_{n-1}$ into k segments, where k is the number of σ_i that are less than or equal to one, $\delta_i \le x_1 \le \delta_{i-1}, i = 2, \dots k+1$. In each such segment we proceed as in Lemma 1 and integrate over the pyramids P_j, adding and subtracting the results as appropriate. Here, though, when $\delta_i \le x_1 \le \delta_{i-1}$, only P_1 through P_{i-1} come into consideration.

And integrating over the pyramid P_i with respect to all variables except x_1 results in

$$\int_{c_{2,i}}^{1-x_1-c_{3,i}-\dots-c_{n-1,i}} \int_{c_{3,i}}^{1-x_1-x_2-c_{4,i}-\dots-c_{n-1,i}} \dots \int_{c_{n-1,i}}^{1-x_1-x_2-\dots-x_{n-1}} dx_{n-1} \dots dx_3\, dx_2 = \frac{(\delta_i - x_1)^{n-2}}{(n-2)!}.$$

Dividing by the volume from Lemma 1 gives the result. □

Corollary 1. *The centroid f_c of x_1 is*

$$\frac{\sum_{i=1}^{k} (-1)^{m_i} \left((1-\sigma_i)^{n-1} \left((n-1)c_{1,i} + \delta_i\right)\right)}{n \left(\sum_{i=1}^{k} (-1)^{m_i} (1-\sigma_i)^{n-1}\right)}.$$

Proof The corollary follows from integrating $x f(x_1)$ with $f(x_1)$ as in Theorem 1 from a_1 to $\min(\delta_1, b_1)$. □

Exploring the Warp Effect

The warp effect is important for real-life applications of uncertain interval reasoning and interval decision analysis. Whenever a decision-maker makes statements of probability using intervals as representations of second-order uncertainty (i.e. not uncertainty of the event itself occurring but rather of the probability of the event occurring), the upper and lower pair of boundaries must be handled. As the warp effect shows, the results are far from the everyday intuition. In order to make the discussion of the effect clearer, we present some examples highlighting the nature of the warp effect.

Example 2. *We have $x_1 \in [0.2, 0.6], x_2 \in [0.1, 0.2]$ and $x_3 \in [0.2, 0.7]$, $n = 4, k = 3$, $\sigma_1 = a_1 + a_2 + a_3 = 0.5, \delta_1 = 1 - a_2 - a_3 = 0.7, \sigma_2 = b_1 + a_2 + a_3 = 0.9, \delta_2 = 1 - a_2 - a_3 = 0.7, \sigma_3 = a_1 + b_2 + a_3 = 0.6$, and $\delta_3 = 1 - b_2 - a_3 = 0.6$. Then*

$$f(x_1) = \frac{3\left((0.7 - x_1)^2 - (0.6 - x_1)^2\right)}{0.5^3 - 0.4^3 - 0.1^3}$$

if $0.2 \le x_1 \le 0.6$,
and $f_c = \frac{(1-\sigma_1)^3(3a_1+\delta_1) - (1-\sigma_2)^3(3b_1+\delta_2) - (1-\sigma_3)^3(3a_1+\delta_3)}{4((1-\sigma_1)^3 - (1-\sigma_2)^3 - (1-\sigma_3)^3)} = 0.347$. *We see the graph of $f(x_1)$ in Figure 1.*

Example 3. *Let $\sigma_1 = a_1 + a_2 + a_3 + a_4 = 0.2 + 0.1 + 0.05 + 0.15 = 0.5$ and $\sigma_2 = a_1 + b_2 + a_3 + a_4 = 0.2 + 0.3 + 0.05 + 0.15 = 0.7$ be the only sums $c_1 + c_2 + c_3 + c_4$ that are less*

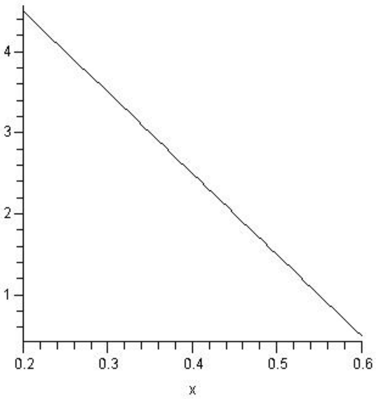

Figure 1: $f(x_1)$ when $x_1 \in [0.2, 0.6], x_2 \in [0.1, 0.2]$ and $x_3 \in [0.2, 0.7]$

than or equal to one. Then $n = 5, k = 2, \delta_1 = 0.7, \delta_2 = 0.5$ *and*

$$f(x_1) = \begin{cases} \frac{4\left((0.7-x_1)^3-(0.5-x_1)^3\right)}{0.5^4-0.3^4}, & 0.2 \leq x_1 \leq 0.5 \\ \frac{4(0.7-x_1)^3}{0.5^4-0.3^4}, & 0.5 \leq x_1 \leq 0.7 \end{cases};$$

the graph of $f(x_1)$ *is shown in Figure 2, and* $f_c = 0.306$.

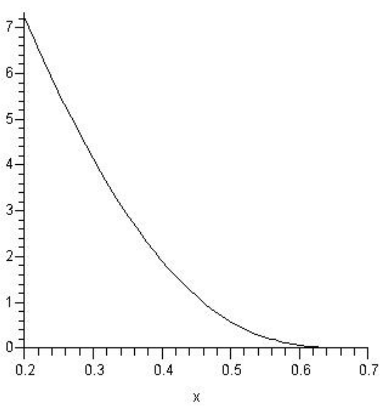

Figure 2: Example of S-projection $f(x_1)$ when $n = 5$.

Example 4. *If we let* $n = 6$ *and* $x_1 \in [0.15, 0.7], x_2 \in [0.1, 0.2], x_3 \in [0, 0.25], x_4 \in [0.2, 0.55]$ *and* $x_5 \in [0.25, 0.7]$, $k = 3, \sigma_1 = 0.7, \sigma_2 = 0.8, \sigma_3 = 0.95, \delta_1 = 0.45, \delta_2 = 0.35$*, and* $\delta_3 = 0.2$*. Then*

$$f(x_1) = \frac{5\left((0.45-x_1)^4 - (0.35-x_1)^4 - (0.2-x_1)^4\right)}{0.3^5 - 0.2^5 - 0.05^5}$$

if $0.15 \leq x_1 \leq 0.2$,

$$f(x_1) = \frac{5\left((0.45-x_1)^4 - (0.35-x_1)^4\right)}{0.3^5 - 0.2^5 - 0.05^5}$$

if $0.2 \leq x_1 \leq 0.35$ *and*

$$f(x_1) = \frac{5(0.45-x_1)^4}{0.3^5 - 0.2^5 - 0.05^5}$$

if $0.35 \leq x_1 \leq 0.45$.

And $f_c = 0.203$*. We see the graph of* $f(x_1)$ *in Figure 3.*

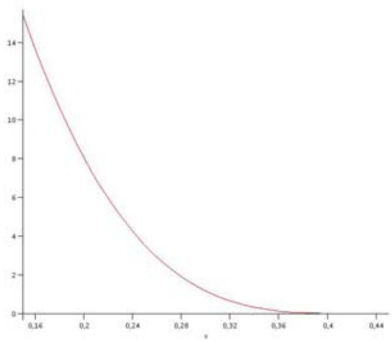

Figure 3: Example of S-projection $f(x_1), 0.15 \leq x_1 \leq 0.45$ when $n = 6$.

Example 5. *When* $n = 6, x_1 \in [0.15, 0.7], x_2 \in [0.1, 0.2], x_3 \in [0, 0.25], x_4 \in [0.2, 0.55]$ *and* $x_5 \in [0.1, 0.7]$, $k = 4$ *and* $\sigma_1 = 0.55, \sigma_2 = 0.65, \sigma_3 = 0.8, \sigma_4 = 0.9, \delta_1 = 0.6, \delta_2 = 0.5, \delta_3 = 0.35$*, and* $\delta_4 = 0.25$*, we get*

$$f(x_1) = \frac{5\left((0.6-x_1)^4 - (0.5-x_1)^4 - (0.35-x_1)^4 + (0.25-x_1)^4\right)}{0.45^5 - 0.35^5 - 0.2^5 + 0.1^5}$$

if $0.15 \leq x_1 \leq 0.25$,

$$f(x_1) = \frac{5\left((0.6-x_1)^4 - (0.5-x_1)^4 - (0.35-x_1)^4\right)}{0.45^5 - 0.35^5 - 0.2^5 + 0.1^5}$$

if $0.25 \leq x_1 \leq 0.35$,

$$f(x_1) = \frac{5\left((0.6-x_1)^4 - (0.5-x_1)^4\right)}{0.45^5 - 0.35^5 - 0.2^5 + 0.1^5}$$

if $0.35 \leq x_1 \leq 0.5$,

$$f(x_1) = \frac{5(0.6-x_1)^4}{0.45^5 - 0.35^5 - 0.2^5 + 0.1^5}$$

if $0.5 \leq x_1 \leq 0.6$,
and $f_c = 0.233$*. The graph of the S-projection* $f(x_1)$ *is shown in Figure 4.*

Example 6. *If* $a_i = 0$ *and* $b_i = 1$, $f(x_1) = (n-1)(1-x_1)^{n-2}$*, and* $f_c = \frac{1}{n}$*. In Figure 5 we see the graphs of* $f(x_1)$ *for* $n = 3, 4, 5, 6, 7,$ *and* 8.

Example 7. *If* $a_i = 0$ *and* $b_i = b = \frac{\frac{1}{n-1}+\frac{1}{n-2}}{2}$*, every possible sum* $\sum_{i=1}^{n-1} c_i$ *is less than one, except* $\sum_{i=1}^{n-1} b_i$*. Then we can express* $f(x_1)$ *as*

$$\frac{(n-1)\left(\sum_{i=0}^{n-2}(-1)^i\binom{n-2}{i}(1-ib-x_1)^{n-2}\right)}{\sum_{i=0}^{n-2}(-1)^i\binom{n-1}{i}(1-ib)^{n-1}}$$

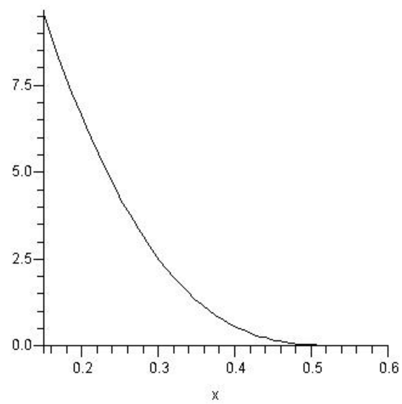

Figure 4: $f(x_1)$ from Example 5, $0.15 \leq x_1 \leq 0.6$ when $n = 6, k = 4$

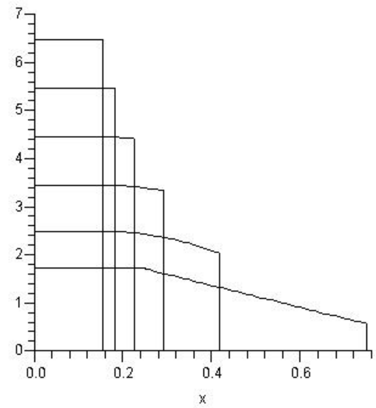

Figure 6: $x_i \in [0, \frac{1}{2(n-1)} + \frac{1}{2(n-2)}], n = 3, 4, 5, 6, 7$ and 8

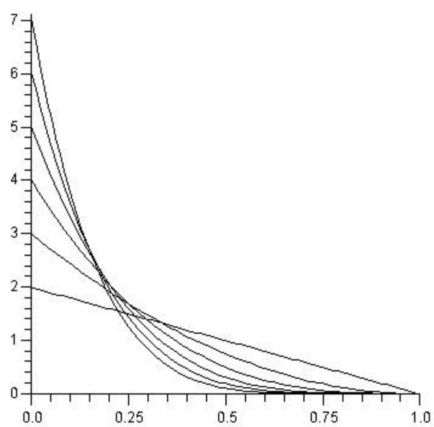

Figure 5: $x_i \in [0, 1], n = 3, 4, 5, 6, 7$ and 8

for $0 \leq x_1 \leq 1 - (n-2)b$ *and*

$$\frac{(n-1)\left(\sum_{i=0}^{n-3}(-1)^i\binom{n-2}{i}(1-ib-x_1)^{n-2}\right)}{\sum_{i=0}^{n-2}(-1)^i\binom{n-1}{i}(1-ib)^{n-1}}$$

for $1 - (n-2)b \leq x_1 \leq b$.

The graphs of $f(x_1)$ *for* $n = 3, 4, 5, 6, 7,$ *and* 8 *are shown in Figure 6. We see that when, as in this case, the upper bound* b_i *is adjusted to the fact that* $\sum_{i=1}^{n} x_i = 1$*, the second order distribution of* x_i *approaches the uniform distribution and the centroid goes to the middle point of the interval as* n *grows.*

Conclusions

We have demonstrated that second-order belief can supply important insights to the decision-maker when handling interval representations, such as in decision trees or probabilistic networks, and that interval estimates (upper and lower bounds) in themselves are not complete. The results apply also to approaches which do not explicitly deal with belief distributions.

The main second-order effect on interval based probabilities is the centroid tending towards the lower bound when n grows as the upper part of the interval contains a shrinking part of the total belief. This effect is dramatic when all or some of the intervals are wide but less pronounced when the intervals are allowed to shrink in inverse proportion to n.

Therefore, it is mainly when there are relatively many possible outcomes to an event and the probability intervals are fairly wide that using the centroid of a probability's second-order distribution as single-value representation is crucial. When the intervals are narrower, the centroid is closer to the intervals' mid-points, thus using either centroids or mid-point contraction of intervals in these cases give roughly the same result.

This is an important observation for reasoning with interval probabilities, either in the form of decision analysis or by other means of inference. The interval boundaries (upper and lower) do not carry the same information. Neither do points in between, and disregarding this information leads to warp effects in the results of probabilistic interval computations.

References

Choquet, G. 1954. Theory of capacities. *Ann. Inst. Fourier* 5:131–295.

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

Dempster, A. P. 1967. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics, vol xxxviii* 325–399.

Denneberg, D. 1994. *Non-Additive Measure and Integral*. Kluwer.

Druzdzel, M. J., and van der Gaag, L. C. 1995. Elicitation of probabilities for belief networks: Combining qualitative and quantitative information. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-95)* 141–148.

Ekenberg, L., and Thorbiörnson, J. 2001. Second-order decision analysis. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 9, No 1* 13–37.

Good, I. 1962. *Subjective Probability as the Measure of a Non-measurable Set*. Stanford University Press. 319–329.

Ha, V., and Haddaway, P. 1998. Geometric foundations for interval-based probabilities. *14th Conference in Uncertainty in Artificial Intelligence (UAI '98)* 193–201.

Huber, P., and Strassen, V. 1973. Minimax tests and the neyman-pearsons lemma for capacities. *Annals of Statistics, 1* 251–263.

Huber, P. 1973. The case of choquet capacities in statistics. *Bulletin of the International Statistical Institute, 45* 181–188.

Shafer, G. 1976. *A Mathematical theory of evidence*. Princeton University Press.

Walley, P. 1991. *Statistical reasoning with Imprecise Probabilities*. Chapman and Hall.