Alternatives to Lavine's algorithm for calculation of posterior bounds given convex sets of distributions

Fabio Cozman
Robotics Institute, Carnegie Mellon University
fgcozman@cs.cmu.edu — http://www.cs.cmu.edu/~fgcozman/home.html

CMU-RI-TR 97-38

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

October 21, 1997

This research is supported in part by NASA under Grant NAGW-1175. Fabio Cozman was supported under a scholarship from CNPq, Brazil.

Abstract

This paper presents alternatives to Lavine's algorithm, currently the most popular method for calculation of expectation bounds induced by sets of probability distributions. The White-Snow algorithm is first analyzed and demonstrated to be superior to Lavine's algorithm in a variety of situations. The calculation of posterior bounds is then reduced to a fractional programming problem. From the unifying perspective of fractional programming, Lavine's algorithm is identical to Dinkelbach's algorithm, and the White-Snow algorithm is essentially identical to the Charnes-Cooper transformation. A novel algorithm for expectation bounds is given for the situation where both prior and likelihood functions are specified as convex sets of distributions.

1 Introduction

This paper presents alternatives to Lavine's algorithm, currently the most popular method for calculation of expectation bounds induced by sets of probability distributions. Models based on convex sets of distributions have been advocated as more realistic, meaningful and flexible [15, 24, 42] than standard probability models.

The most important practical application of convex sets of distributions is *robustness analysis* [3, 19]. The idea is employ sets of distributions to represent perturbations and variations in a probabilistic model. The goal of robustness analysis is to produce bounds on expected values. The interval between the upper and lower bounds induced by a convex set of distributions reflects the quality of model and data; small intervals indicate *robustness* to perturbations.

Lavine has proposed a bracketing algorithm, aimed at robustness analysis, that has captured great attention in recent years [23]. The purpose of this paper is to present other algorithms, with better performance than Lavine's.

The White-Snow algorithm is first analyzed and demonstrated to be superior to Lavine's algorithm in a variety of situations. The calculation of posterior bounds is then reduced to a fractional programming problem. From the unifying perspective of fractional programming, Lavine's algorithm is identical to Dinkelbach's algorithm, and the White-Snow algorithm is essentially identical to the Charnes-Cooper transformation. A novel algorithm for expectation bounds is given for the situation where both prior and likelihood functions are specified as convex sets of distributions.

2 Quasi-Bayesian theory

This section presents the basic technical results used in this paper. The mathematical ideas are taken from Giron and Rios' axiomatization, which they refer to as *Quasi-*

Bayesian theory [15]. Quasi-Bayesian theory employs a closed convex sets of distributions and a single utility function to represent the beliefs and to evaluate the preferences of a rational decision-maker. Several other theories use similar representations: inner/outer measures [16, 18, 31, 39], lower probability theory [5, 8, 13, 36], convex Bayesianism [17], probability/utility sets [34]. Some theories, like Dempster-Shafer theory [35], use representations that can be recast in terms of convex sets of distributions [18, 45].

A closed convex set of distributions is called a *credal set* [24]. A credal set containing distributions $p(X)$ for a set of variables X is denoted $K(X)$. To simplify terminology, the term *credal set* is used only when it refers to a set of distributions containing more than one element.

2.1 Lower and upper envelopes

Given a credal set K, a probability interval is induced for every event A:

$$\underline{p}\,(A) = \min_{p \in K} p(A) \qquad \overline{p}\,(A) = \max_{p \in K} p(A). \tag{1}$$

The set-functions $\underline{p}(A)$ and $\overline{p}(A)$ are called lower and upper envelopes respectively. Lower envelopes can be obtained from upper envelopes through the expression $\underline{p}(A) = 1 - \overline{p}(A^c)$.

A credal set always creates unique lower and upper bounds of probability, but a set of lower and upper bounds of probability does not define a unique credal set [42, section 2.7].

2.2 Lower and upper expectations

Lower and upper expectations for a function f are defined as:

$$\underline{E}\,[f] = \min_{p \in K} E[f] \qquad \overline{E}\,[f] = \max_{p \in K} E[f], \tag{2}$$

Lower expectations can be obtained from upper expectations through the expression $\underline{E}[f] = -\overline{E}[-f]$. The lower envelope for event A is obtained by taking the lower expectation for the indicator function $\delta_A(X)$, which is one if $X \in A$ and zero otherwise.

There is a one-to-one correspondence between lower (or upper) expectations and credal sets. Any credal set generates a unique lower (or upper) expectation and vice-versa [42].

2.3 Conditionalization

Credal sets are used to represent the beliefs held by a decision maker *given* an event. A credal set containing conditional distributions $p(X|Y)$ for sets of variables X and Y is indicated by $K(X|Y)$.

Quasi-Bayesian inference is performed by applying Bayes rule to each distribution in a credal set; the posterior credal set is the union of all posterior distributions[1]. Two well-known results about posterior credal sets are used in this paper. First, to obtain a posterior credal set, one has to apply Bayes rule only to the vertices of a credal set and take the convex hull of the resulting posterior distributions [15]. Second, to obtain maximum and minimum values of posterior probabilities, one must look only at the vertices of the posterior credal sets [42, section 6.4.2].

The functionals $\overline{E}[f|A]$ and $\overline{E}[f]$ are closely related by the *generalized Bayes rule* (first proposed by Walley [42, section 6.4.1]). For a non-null event A, $\overline{E}[f|A]$ is the only solution of the equation:

$$\overline{E}[(f - \lambda)\,\delta_A] = 0. \tag{3}$$

In many situations, a joint credal set is not directly specified; instead, marginal and conditional credal sets are provided. The following result is important in this situation.

[1] An introduction to technical aspects of Quasi-Bayesian theory, with a larger list of references, can be found at http://www.cs.cmu.edu/~fgcozman/qBayes.html.

Suppose a marginal credal set $K(X)$ and a collection of conditional credal sets $K(X|Y)$ are defined. Define the *lower likelihood* $L(A|Y) = \underline{p}(A|Y)$ and the *upper likelihood* $U(A|Y) = \overline{p}(A|Y)$.

The following result shows that the lower and upper likelihoods summarize all the information in $K(X|Y)$:

Theorem 1 (Walley [42, section 8.5.3]). *Consider a function $f(Y)$ and an event A such that $\underline{p}(A) > 0$. There is a number λ such that the value of $\overline{E}[f|A]$ is attained when $p(A|Y)$ is:*

$$p(A|Y) = \begin{cases} U(A|Y) & \text{if} \quad f(Y) < \lambda, \\ L(A|Y) & \text{if} \quad f(Y) \geq \lambda. \end{cases}$$

3 Probabilistic logic and robust Statistics

A Quasi-Bayesian model can be understood as a collection of constraints on probability distributions. This is a view emphasized in probabilistic logic [28] and its variants [2, 10, 14, 11, 25, 41]. The purpose of the work is to start from a collection of linear constraints on the probability of propositions, and obtain probability bounds through linear programming. Conditional and posterior constraints can be handled to a limited extent [11, 26].

Another line of research has focused on the properties of 2-monotone and infinite monotone Choquet capacities; for those types of credal sets, closed-form expressions exist for posterior quantities when a single measurement is performed [18, 46]. Chrisman has modified these closed-form expressions to handle sequences of measurements [7].

Most work in Bayesian robust Statistics focuses on a single conditional distribution $p(X|\theta)$ and a prior credal set $K(\theta)$ [3, 4, 22, 44]. Inference is equated to calculation of the posterior credal set $K(\theta|X)$ [4]. A number of results are available in special but important cases: 2-monotone Choquet capacities [46], density ratio classes [4] and unimodal and symmetric constraints [4].

Posterior quantities can always be generated through the generalized Bayes rule (expression (3); Walley has proposed an iterative descent algorithm to obtain posterior quantities by calculating a series of upper expectations [43, page 17]. This idea is closely related to Lavine's bracketing algorithm (section 4), the most well-known method for calculation of upper posterior expectations.

Only a few authors consider the possibility that prior *and* conditional credal sets be defined [23, 30, 42]. A solution for this problem is derived by linear fractional programming methods in section 6.

4 Lavine's algorithm

The posterior upper expectation for a function f conditional on event A is:

$$\overline{E}\left[f|A\right] = \max\left[\frac{\sum_{X \notin A} f(X)p(X)}{\sum_{X \notin A} p(X)}\right], \tag{4}$$

where the maximization is with respect to all the distributions in the joint credal set. To simplify the presentation, only upper expectations are considered; lower expectations can be obtained through the similar computations.

Notice that $\overline{E}\left[f|A\right] > \mu$ if and only if:

$$\max\left[\sum_{x \notin A}(f(x) - \mu)p(x)\right] > 0. \tag{5}$$

The idea of the bracketing algorithm is the following. Pick a real number μ and check whether $\overline{E}\left[f|A\right]$ is larger, smaller or equal to μ, and respectively increase μ, decrease μ or stop. Repeat this until the bracketing interval is small enough.

Expression (5) is equal to:

$$\max E_p[(f - \mu)\delta_A] = \overline{E}\left[(f - \mu)\delta_A\right];$$

Lavine's algorithm is simply a bracketing iteration based on the generalized Bayes rule (expression (3)).

Lavine's algorithm is a viable choice if the maximization in expression (5) is in fact easier than the nonlinear maximization in expression (4). This happens for several special models, such as density ratio classes and 2-monotone Choquet capacities [4]. Apart from these special cases, Lavine's algorithm is advantageous when a single conditional distribution $p(X|\theta)$ and a prior credal $K(\theta)$ are specified; in this case each iteration of the algorithm (expression 5) becomes a linear program.

5 White-Snow algorithm

Despite its popularity, Lavine's algorithm is *not* the most efficient way to obtain the upper expectation for a function f when:

- a single conditional distribution $p(X|\theta)$ is specified and
- a prior credal $K(\theta)$ is specified through linear constraints:

  $$\mathbf{A}[p(\theta_1) \dots p(\theta_n)]^T \leq \mathbf{B},$$

  where $\mathbf{A}$ and $\mathbf{B}$ are matrices of appropriate dimensions.

A more efficient algorithm, developed by White [47] and improved by Snow [37] should be used in these cases. This fact has not been noticed before; the White-Snow algorithm was developed in the context of expert systems and no connection to Lavine's algorithm has been mentioned. The White-Snow algorithm is summarized below in uniform notation.

Define:

- $\alpha_i = p(\theta_i)$ and $\beta_i = p(X|\theta_i)$;

- $f_i = f(\theta_i)$;
- the matrix $\mathbf{C} = \mathbf{A} - \mathbf{B}\mathbf{1}$, where $\mathbf{1}$ is a row vector of ones;
- the matrix $\mathbf{D} = \mathbf{A} \times \ \mathrm{diag}\left[\beta_1^{-1}, \ldots, \beta_n^{-1}\right]$.

Suppose a measurement A is obtained. The calculation of the posterior upper expectation is:

$$\overline{E}\left[f|A\right] = \max_{\alpha} \left[\frac{\sum_i f_i \alpha_i \beta_i}{\sum_j \alpha_j \beta_j}\right],$$

subject to:

$$\mathbf{C}\alpha \leq 0, \qquad \sum_i \alpha_i = 1, \qquad \alpha_i \geq 0.$$

White proposed the following change of variables [47]:

$$\gamma_i = \frac{\alpha_i \beta_i}{\sum_j \alpha_j \beta_j},$$

which reduces the calculation of the posterior upper expectation to a linear program:

$$\overline{E}\left[f|X\right] = \max_{\gamma} \left[\sum_i f_i \gamma_i\right],$$

subject to:

$$\mathbf{D}\gamma \leq 0, \qquad \sum_i \gamma_i = 1, \qquad \gamma_i \geq 0.$$

To obtain the matrix $\mathbf{D}$, all β_i must be larger than zero. When $\beta_i = 0$, set the variable γ_i to zero and discard it from the equations.

Snow has proved that the solution of this linear program produces the correct posterior upper envelope [37]. Notice that this solution is much better than Lavine's since it is exact and does not require multiple iterations.

6 Linear fractional programming

The algorithms presented above were derived through special properties of upper expectations. Both algorithms can be derived in a more direct and general way as algorithms for linear fractional programming. This fact has not been previously noticed in the literature. Only recent references associate linear fractional programs to upper expectations [21, 27, 48], but none connects linear fractional programming to previously known algorithms.

Linear fractional programming studies the of maximization of ratios of linear functions [33]. Consider the linear fractional program:

$$\max_{\alpha} \left[\frac{\sum_i f_i \alpha_i \beta_i}{\sum_j \alpha_j \beta_j} \right],$$

for given β_i and f_i, and α_i subject to linear constraints.

There are two main algorithms for linear fractional programming [20]:

- Create a "parameterized" problem for a parameter μ:

  $$\max_{\alpha} \left[\sum_i (f_i - \mu) \alpha_i \beta_i \right],$$

  subject to the same constraints on the original problem; and search for the value of μ such that the maximum is zero. This method is variously called Dinkelbach or Jagannatham algorithm, and dates back to the sixties [20]. Notice that this method is identical to Lavine's algorithm.

- Transform the problem by a change of variables:

  $$\gamma'_i = \frac{\alpha_i}{\sum_j \alpha_j \beta_j},$$

  which reduces the calculation of the posterior upper expectation to a linear program:

  $$\max_{\gamma'} \left[\sum_i f_i \beta_i \gamma'_i \right],$$

subject to:

$$\mathbf{C}\gamma' \leq 0, \qquad \sum_i \beta_i \gamma'_i = 1, \qquad \gamma'_i \geq 0.$$

This is called the Charnes-Cooper method, and also dates back to the sixties [20]. Notice that this method is essentially identical to the White-Snow algorithm; the only difference is that $\gamma_i = \gamma'_i \beta_i$. The Charnes-Cooper method has the advantage of automatically handling the case $\beta_i = 0$.

6.1 Posterior upper expectation for prior and conditional credal sets

To illustrate the convenience of the linear fractional programming approach, consider a situation where both prior and conditional credal sets, $K(\theta)$ and $K(X|\theta)$ respectively, are specified. Assume that each $K(X|\theta)$ is separately specified: no constraint on $K(X|\theta_{k_1})$ involves values $K(X|\theta_{k_2})$ for $k_1 \neq k_2$. This requirement is discussed in section 7.

A simplified version of this problem, where an interval is associated with each value of $p(X|\theta)$, is studied by Snow [38]; his methods are used in the following derivation. Approximate bounds for the same problem are provided by Salo [32] using different techniques.

Consider a direct, exact solution through fractional linear programming, which extends previous work by allowing for arbitrary linear constraints over each $K(X|\theta)$.

By theorem 1, the upper expectation is attained only by distributions that assume either $U(X|\theta_i)$ or $L(X|\theta_i)$ for each θ_i [42, section 8.5.4]. The lower and upper likelihoods are functions such that:

$$L(X|\theta_i) \leq p(X|\theta_i) \leq U(X|\theta_i).$$

Define two vectors, α' and α'', with the same length as α. Consider the following linear fractional program:

$$\max_{\alpha',\alpha''}\left[\frac{f_i L(X|\theta_i)\alpha'_i + f_i U(X|\theta_i)\alpha''_i}{\sum_j (L(X|\theta_j)\alpha'_j + U(X|\theta_j)\alpha''_j)}\right],$$

subject to:

$$\mathbf{C}(\alpha' + \alpha'') \leq 0, \qquad \sum_i (\alpha'_i + \alpha''_i) = 1, \qquad \alpha'_i \geq 0, \qquad \alpha''_i \geq 0.$$

For each θ_i, either $\alpha'_i = 0$ or $\alpha''_i = 0$. This automatically selects the correct likelihood value. Now the Charnes-Cooper transformation can be applied and the lower expectation can be obtained through a linear program. The final algorithm is slightly more general than the procedure derived by Snow [38], and the derivation is shorter, simpler and more elegant; notice that the algorithm automatically handles situations where the likelihoods are zero, while Snow has to consider a variety of special cases.

To illustrate the method, consider the following example (taken from White's original paper [47]):

Example 1. *Consider a variable X with four states $\{\theta_1, \theta_2, \theta_3, \theta_4\}$, and the following constraints on the marginal (prior) distribution of X:*

$$2.5p(\theta_1) \geq p(\theta_4) \geq 2p(\theta_1), \qquad 10p(\theta_3) \geq p(\theta_2) \geq 9p(\theta_3), \qquad p(\theta_2) = 5p(\theta_4).$$

Suppose the following lower and upper likelihoods are given:

$$\begin{array}{llll} L(X|\theta_1) = 0.9, & L(X|\theta_2) = 0.1125, & L(X|\theta_3) = 0.05625, & L(X|\theta_4) = 0.1125, \\ U(X|\theta_1) = 0.95, & U(X|\theta_2) = 0.1357, & U(X|\theta_3) = 0.1357, & U(X|\theta_4) = 0.1357. \end{array}$$

Consider the calculation of $\underline{p}(X = \theta_1|X)$:

$$\underline{p}(X = \theta_1|X) = \min_{\alpha',\alpha''}(\alpha'_1 + \alpha''_1),$$

where α' and α'' are vectors with four elements, subject to:

- $\mathbf{C}[0.9\alpha' + 0.95\alpha''] \leq 0$, where the matrix $\mathbf{C}$ is:

$$\mathbf{C} = \begin{bmatrix} -2.5 & 0 & 0 & 1 & -2.5 & 0 & 0 & 1 \\ 2 & 0 & 0 & -1 & 2 & 0 & 0 & -1 \\ 0 & -1 & 0 & 5 & 0 & -1 & 0 & 5 \\ 0 & 1 & 0 & -5 & 0 & 1 & 0 & -5 \\ 0 & -1 & 9 & 0 & 0 & -1 & 9 & 0 \\ 0 & 1 & -10 & 0 & 0 & 1 & -10 & 0 \end{bmatrix}.$$

- $[0.9, 0.1125, 0.0562, 0.1125]\alpha' + [0.95, 0.1357, 0.1357, 0.1357]\alpha'' = 1$,

- $\alpha'_i \geq 0$ and $\alpha''_i \geq 0$.

The value $\underline{p}(X = \theta_1|X) = 0.2881$ is obtained through linear programming. The minimizing α' is $[0.3201, 0, 0, 0]$ and the minimizing α'' is $[0, 4.0013, 0.4446, 0.8003]$.

The bounds obtained through linear fractional programming are only valid if the conditional credal sets are specified separately for each value of θ. White's original example specified the conditional distributions through linear inequalities:

$$p(X|\theta_2) = p(X|\theta_4), \qquad p(X|\theta_3) \leq p(X|\theta_2) \leq 2p(X|\theta_3),$$

$$7p(X|\theta_2) \leq p(X|\theta_1) \leq 8p(X|\theta_2), \qquad p(X|\theta_3) \geq 0.01, \qquad 0.9 \leq p(X|\theta_1) \leq 0.95.$$

In this case the bounds produced by linear fractional programming are not tight, because theorem 1 does not apply.

6.2 Interval-valued prior and conditional distributions

Suppose the prior credal set is defined by lower and upper bounds [43, page 18]. For a prior credal set $K(\theta) = \left\{p(\theta)|\underline{p}(\theta) \leq p(\theta) \leq \overline{p}(\theta)\right\}$, such that:

$$\begin{aligned} \underline{p}(\theta_i) &\geq 1 - \sum_{j \neq i} \overline{p}(\theta_j), \\ \overline{p}(\theta_i) &\leq 1 - \sum_{j \neq i} \underline{p}(\theta_j) \end{aligned}$$

for all θ_i, define:

- A distribution p', generated by fixing $p'(\theta_i) = \underline{p}(\theta_i)$ and by placing as much probability mass as possible in the values of θ with largest upper likelihood.
- A distribution p'', generated by fixing $p''(\theta_i) = \overline{p}(\theta_i)$ and by placing as much probability mass as possible in the values of θ with smallest lower likelihood.

Then:

$$\underline{p}(\theta_i|X) = \frac{L(X|\theta_i)p'(\theta_i)}{L(X|\theta_i)p'(\theta_i) + \sum_{j \neq i} U(X|\theta_j)p'(\theta_j)},$$
$$\overline{p}(\theta_i|X) = \frac{U(X|\theta_i)p''(\theta_i)}{U(X|\theta_i)p''(\theta_i) + \sum_{j \neq i} L(X|\theta_j)p''(\theta_j)}.$$

7 Conclusion

The algorithms presented in this paper can be divided in two groups. The first group focuses on boolean variables, linked by logical propositions, and searches for bounds on probability values. There is large freedom on the logical connections among variables, but minimal expressivity in terms of probabilities. The other group of algorithms involves a few variables, and concentrates on posterior quantities. Variables can be very expressive, but reasoning chains are of limited length. In neither case are independence relations handled. Together, these two groups of algorithms cover the majority of known methods for Quasi-Bayesian inferences.

A different, more general approach is to sketch a dependency structure among variables and specify conditional credal sets for this structure. The idea is to mimic graphical methods for standard probability theory; this path has been pursued by several authors in Artificial Intelligence [1, 6, 12, 9, 40, 48].

Two new results are noteworthy in this paper. First, a unified perspective for Lavine's algorithm and the White-Snow algorithm is derived, based on linear fractional programming. Second, the main contribution is the algorithm for calculation of posterior quantities for prior and conditional credal sets presented in section 6.1. The algorithm provides a complete solution for conditional credal under the assumption that conditional credal sets separately specified. The assumption is reasonable, since each conditional credal set must be specified *given* a particular, fixed value of the conditioning variable.

References

[1] K. A. Andersen and J. N. Hooker. Bayesian logic. *Decision Support Systems*, 11:191–210, 1994.

[2] F. Bacchus. *Representing and Reasoning with Probabilistic Knowledge: A Logical Approach*. MIT Press, Cambridge, 1990.

[3] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, 1985.

[4] J. O. Berger. Robust Bayesian analysis: Sensitivity to the prior. *Journal of Statistical Planning and Inference*, 25:303–328, 1990.

[5] J. S. Breese and K. W. Fertig. Decision making with interval influence diagrams. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 467–478. Elsevier Science, North-Holland, 1991.

[6] J. Cano, M. Delgado, and S. Moral. An axiomatic framework for propagating uncertainty in directed acyclic networks. *International Journal of Approximate Reasoning*, 8:253–280, 1993.

[7] L. Chrisman. Incremental conditioning of lower and upper probabilities. *International Journal of Approximate Reasoning*, 13(1):1–25, 1995.

[8] L. Chrisman. Propagation of 2-monotone lower probabilities on an undirected graph. *XII Uncertainty in Artificial Intelligence Conference*, pages 178–186, 1996.

[9] F. Cozman. Robustness analysis of Bayesian networks with local convex sets of distributions. *Proc. Thirteen Conference Uncertainty in Artificial Intelligence*, 1997.

[10] D. Dubois, H. Prade, and P. Smets. Representing partial ignorance. *IEEE Transactions on Systems, Man and Cybernetics A*, 26(3):361–377, May 1996.

[11] R. Fagin, J. Y. Halpern, and N. Megiddo. A logic for reasoning about probabilities. *Information and Computation*, 87:78–128, 1990.

[12] K. W. Fertig and J. S. Breese. Interval influence diagrams. In M. Henrion, R. D. Shachter, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5*, pages 149–161. Elsevier Science Publishers, North-Holland, 1990.

[13] T. L. Fine. Lower probability models for uncertainty and nondeterministic processes. *Journal of Statistical Planning and Inference*, 20:389–411, 1988.

[14] A. M. Frisch and P. Haddawy. Anytime deduction for probabilistic logic. *Artificial Intelligence*, 69:93–122, 1994.

[15] F. J. Giron and S. Rios. Quasi-Bayesian behaviour: A more realistic approach to decision making? In J. M. Bernardo, J. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics*, pages 17–38. University Press, Valencia, Spain, 1980.

[16] I. J. Good. *Good Thinking: The Foundations of Probability and its Applications*. University of Minnesota Press, Minneapolis, 1983.

[17] H. E. Kyburg Jr. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, 31:271–293, 1987.

[18] J. Y. Halpern and R. Fagin. Two views of belief: Belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54:275–317, 1992.

[19] P. J. Huber. *Robust Statistics*. Wiley, New York, 1980.

[20] T. Ibaraki. Solving mathematical programming problems with fractional objective functions. In S. Schaible and W. T. Ziemba, editors, *Generalized Concavity in Optimization and Economics*, pages 440–472. Academic Press, 1981.

[21] B. Jaumard, P. Hansen, and M. P. de Aragão. Column generation methods for probabilistic logic. *ORSA Journal on Computing*, 3(2):135–148, Spring 1991.

[22] J. B. Kadane. *Robustness of Bayesian Analyses*, volume 4 of *Studies in Bayesian econometrics*. Elsevier Science Pub. Co., New York, 1984.

[23] M. Lavine. Sensitivity in Bayesian statistics, the prior and the likelihood. *Journal of the American Statistical Association*, 86(414):396–399, June 1991.

[24] I. Levi. *The Enterprise of Knowledge*. The MIT Press, Cambridge, Massachusetts, 1980.

[25] T. Lukasiewicz. Uncertain reasoning in concept lattices. *Proc. of the 3rd European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 293–300, 1995.

[26] T. Lukasiewicz. *Precision of Probabilistic Deduction under Taxonomic Knowledge*. PhD thesis, Universitat Augsburg, 1996.

[27] C. Luo, C. Yu, J. Lobo, G. Wang, and T. Pham. Computation of best bounds of probabilities from uncertain data. *Computational Intelligence*, 12(4):541–566, 1996.

[28] N. J. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28:71–87, 1986.

[29] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kauffman, San Mateo, CA, 1988.

[30] L. R. Pericchi and M. E. Perez. Posterior robustness with more than one sampling model. *Journal of Statistical Planning and Inference*, 40:279–294, 1994.

[31] E. H. Ruspini. The logical foundations of evidential reasoning. Technical Report SRIN408, SRI International, 1987.

[32] A. A. Salo. Tighter estimates for the posteriors of imprecise prior and conditional probabilities. *Transactions on Systems, Man, and Cybernetics A*, 26(6):820–825, November 1996.

[33] S. I. Schaible and W. T. Ziemba. *Generalized Concavity in Optimization and Economics*. Academic Press, 1981.

[34] T. Seidenfeld. Outline of a theory of partially ordered preferences. *Philosophical Topics*, 21(1):173–188, Spring 1993.

[35] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, 1976.

[36] C. A. B. Smith. Consistency in statistical inference and decision. *Journal Royal Statistical Society B*, 23:1–25, 1961.

[37] P. SNow. Improved posterior probability estimates from pirior and conditional linear constraint systems. *Transactions on Systems, Man, and Cybernetics A*, 21(2):464–469, March/April 1991.

[38] P. Snow. The posterior probabilities of linearly constrained priors and interval-bounded conditionals. *IEEE Transactions on Systems, Man and Cybernetics*, 26A(5):655–659, September 1996.

[39] P. Suppes. The measurement of belief. *Journal Royal Statistical Society B*, 2:160–191, 1974.

[40] B. Tessem. Interval probability propagation. *International Journal of Approximate Reasoning*, 7:95–120, 1992.

[41] H. Thone, U. Guntzer, and W. Kieβling. Towards precision of probabilistic bounds propagation. *VIII Uncertainty in Artificial Intelligence Conference*, pages 315–322, 1992.

[42] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York, 1991.

[43] P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996.

[44] L. Wasserman. Recent methodological advances in robust Bayesian inference. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 4*, pages 483–502. Oxford University Press, 1992.

[45] L. A. Wasserman. Prior envelopes based on belief functions. *The Annals of Statistics*, 18(1):454–464, 1990.

[46] L. A. Wasserman and J. B. Kadane. Bayes' theorem for choquet capacities. *The Annals of Statistics*, 18(3):1328–1339, 1990.

[47] C. C. White III. A posteriori representations based on linear inequality descriptions of a priori and conditional probabilities. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-16(4):570–573, July/August 1986.

[48] M. Zaffalon. *Inferenze e Decisioni in Condizioni di Incertezza con Modelli Grafici Orientati*. PhD thesis, Università di Milano, February 1997.