Robustness Analysis of Bayesian Networks with Global Neighborhoods

Fabio Cozman

CMU-RI-TR 96-42

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

January 23, 1997

Abstract

This paper presents algorithms for robustness analysis of Bayesian networks with global neighborhoods. Robust Bayesian inference is the calculation of bounds on posterior values given perturbations in a probabilistic model. We present algorithms for robust inference (including expected utility, expected value and variance bounds) with global perturbations that can be modeled by ϵ-contaminated, constant density ratio, constant density bounded and total variation classes of distributions.

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

1 Introduction

Robustness analysis employs sets of distributions to model perturbations in the parameters of a probability distribution [Berger, 1985; 1990; Huber, 1980; Kadane, 1984; Wasserman, 1992b]. Robust Bayesian inference is the calculation of bounds on posterior values given such perturbations. This paper focuses on perturbations that are imposed on the global structure of a Bayesian network. The goal is to find exact solutions for calculation of posterior bounds when global perturbations are present. We concentrate on bounds for marginal probabilities, expectations and variances.

No efficient *exact* solutions are available for calculation of posterior bounds in arbitrary Bayesian networks with perturbations modeled by sets of probability distributions (the algorithmic challenge presented by robust Bayesian inference is discussed in a companion technical report [Cozman, 1996]). However, this does not preclude the existence of models that are amenable to efficient analysis (for example, Markov-like lower probabilities [Chrisman, 1996b]). This paper proposes global models that lead to such efficient exact analysis.

Section 2 introduces the theory of inferences which validates our procedures. Section 3 introduces the definitions needed to develop global neighborhoods of Bayesian networks, and Sections 4, 5, 6 and 7 present solutions for the ϵ-contaminated, the constant density ratio, the constant density bounded and the total variation classes respectively. Section 8 extends the analysis to variance bounds. Section 9 discusses the relationship between global neighborhoods and independence relations in a Bayesian network.

2 Quasi-Bayesian theory

We use the theory of convex sets of probability distributions, referred to as Quasi-Bayesian theory [Giron & Rios, 1980]. Motivation for the use of this theory in inferences, and a summary of the theory, is available in a companion technical report [Cozman, 1996]. Here we outline the chief ideas behind Quasi-Bayesian theory.

Imprecision in probability assessments can be due either to difficulties in eliciting information from experts, or to difficulties in processing or combining data [Walley, 1991]. This imprecision is modeled by convex sets of distributions, called *credal sets*. To simplify terminology, we use the term credal set only when it refers to a set of distributions containing more than one element. Convex sets of conditional distributions are used to represent conditional beliefs. Inference is performed by applying Bayes rule to each distribution in a prior credal set; the posterior credal set is the union of all posterior distributions. 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.

Given a Quasi-Bayesian credal set $\mathcal{K}$, a probability interval is induced for every event A:

$$\underline{p}(A) = \inf_{p \in \mathcal{K}} p(A) \qquad \overline{p}(A) = \sup_{p \in \mathcal{K}} p(A).$$

A credal set always induces unique lower and upper bounds of probability, but arbitrary lower and upper probabilities do not induce a unique credal set.

Lower and upper expectations for a function $u(\cdot)$ are defined as:

$$\underline{E}[x_q] = \min_p E_p[u] \qquad \overline{E}[x_q] = \max_p E_p[u],$$

where $E_p[u]$ is the usual expectation for function $u(\cdot)$ with probability distribution $p(\cdot)$.

3 Global neighborhoods for Bayesian networks

A Bayesian network represents a joint distribution through a collection of locally defined probability distributions. Consider a set $\tilde{x}$ of discrete variables with a finite number of values. A Bayesian network defines a probability distribution through the expression [Pearl, 1988]:

$$p(\tilde{x}) = \prod_i p(x_i|\mathrm{pa}(x_i)), \qquad (1)$$

where $\mathrm{pa}(x_i)$ is a set of variables, the *parents* of variable x_i.

For any function $f(\cdot)$, we use the abbreviation f_i for $f(x_i|\mathrm{pa}(x_i))$. In this notation, expression (1) can be rewritten as $p(\tilde{x}) = \prod_i p_i$. Evidence e refers to a set of values for some variables in the network.

From a robustness perspective, it is convenient to understand sets of probability distributions as neighborhoods for base probability distributions. Call $\Gamma(p)$ the neighborhood of distribution $p(\cdot)$ [Wasserman & Kadane, 1992]; $\Gamma(p)$ is a convex set of distributions induced by $p(\cdot)$.

The local structure of a Bayesian network has been explored in several attempts to encode probability intervals into a joint distribution [Breese & Fertig, 1991; Cano, Delgado, & Moral, 1993; Chrisman, 1996b; Shenoy & Shafer, 1990; Tessem, 1992]. A general solution for inferences with local neighborhoods in a Bayesian network has been given in a companion technical report, for most of the known neighborhoods used in robust Statistics [Cozman, 1996]. Here we depart from this local model and seek global neighborhoods for Bayesian networks, which are appropriate to model global perturbations such as the effect of invalid independence assumptions. We concentrate on neighborhoods $\Gamma(p(\tilde{}))$ for the joint distribution.

In the next sections expressions for upper and lower expectations are developed. Consider a function $u(\tilde{x})$; we assume that $u(\tilde{x}) = u(\breve{x})$ where $\breve{x}$ is a subset of the variables in $\tilde{x}$. The expressions indicate how to obtain upper and lower expectations for $u(\cdot)$ of this form.

4 ϵ-contaminated global neighborhoods

An ϵ-contaminated class is characterized by a distribution $p(\cdot)$ and a real number $\epsilon \in (0, 1)$ [Berger, 1985]:

$$\Gamma_\epsilon^C(p(x)) = \{r(x) : r(x) = (1-\epsilon)p(x) + \epsilon q(x)\}, \quad (2)$$

where $q(\cdot)$ is an arbitrary distribution. We are interested in ϵ-contaminated neighborhoods in the space of all Bayesian networks:

$$\Gamma_\epsilon^C(p(\tilde{x})) = \left\{ r(\tilde{x}) : r(\tilde{x}) = (1-\epsilon)\prod_i p_i + \epsilon q(\tilde{x}) \right\}.$$

The posterior expected value for $u(\tilde{x})$ is:

$$E[u] = \frac{U(e)}{p(e)}$$

where:

$$U(e) = \sum_{x \notin e} u^e(\tilde{x}) p^e(\tilde{x})$$

$$p(e) = \sum_{x \notin e} p^e(\tilde{x})$$

An ϵ-contaminated class is a finitely generated convex set of distributions [Cozman, 1996]. The vertices of this set are unitary point masses on each one of the possible configurations of the network. The maximum and minimum expected values for $u(\tilde{x})$ occur in these vertices, since we are optimizing a linear function over a convex set.

The upper expectation is:

$$\overline{E}[u] = \max\left(\frac{U(e)}{p(e)}, \frac{(1-\epsilon)U(e) + \epsilon\overline{u}^e}{(1-\epsilon)p(e) + \epsilon} \right),$$

where $\overline{u}^e = \max u^e(\tilde{x})$. The two terms inside parenthesis represent the two possibilities for a maximum. The first possibility is that $q(\tilde{x})$ places all the mass in a configuration for which the evidence has zero probability. The second possibility is that $q(\tilde{x})$ places all the mass in a configuration for which the evidence has non-zero probability.

The same reasoning leads to the lower expectation:

$$\underline{E}[u] = \min\left(\frac{U(e)}{p(e)}, \frac{(1-\epsilon)U(e) + \epsilon\underline{u}^e}{(1-\epsilon)p(e) + \epsilon} \right),$$

where $\underline{u}^e = \min u^e(\tilde{x})$.

Some special cases are important. When $u(\tilde{x}) = x_q$, then $\overline{E}[u] = \overline{E}[x_q]$, the upper expected value of variable x_q. The most important special case is $u(\tilde{x}) = \delta_a(x_q)$, where $\delta_a(x_q)$ is one if $x_q = a$ and zero otherwise. In this case $\overline{E}[u] = \overline{p}(x_q = a|e)$, the posterior probability for x_q.

The posterior probability bounds are obtained from the previous expressions through simple substitutions:

$$\overline{p}(x_q = a|e) = \frac{(1-\epsilon)p(x_q = a, e) + \epsilon}{(1-\epsilon)p(e) + \epsilon},$$

$$\underline{p}(x_q = a|e) = \frac{(1-\epsilon)p(x_q = a, e)}{(1-\epsilon)p(e) + \epsilon},$$

where

$$p(x_q = a, e) = \sum_{x \notin \{x_q, e\}} p^{\{x_q = a, e\}}(\tilde{x}).$$

Notice that standard Bayesian network algorithms can be used to produce the elements of $p(x_q, e)$ which are required in this expression [Cannings & Thompson, 1981; Dechter, 1996; Zhang & Poole, 1996].

5 Constant density ratio global neighborhoods

A density ratio class consists of all probability densities $p(A)$ so that for any event A [DeRobertis & Hartigan, 1981]:

$$\Gamma^R_{l,u}(p(A)) = \{r(A) : l(A) \leq \alpha r(A) \leq u(A)\},$$

where $l(A)$ and $u(A)$ are arbitrary positive measures such that $l(\cdot) \leq u(\cdot)$ and α is some positive real number.

A global neighborhood can be constructed with a subclass of the density ratio class. Take a base Bayesian network and a positive constant $k > 1$. Consider the set of all joint distributions such that for some α:

$$\Gamma^R_k(p(\tilde{x})) = \left\{ r(\tilde{x}) : (1/k) \prod_i p_i \leq \alpha r(\tilde{x}) \leq k \prod_i p_i \right\}.$$

Call this class a *constant density ratio* class. This class is invariant to marginalization and application Bayes rule; in fact it is the *only* class that has these two properties [Wasserman, 1992a]. Another way to characterize this class is to define it as the set of all distributions $r(\cdot)$ that obey the inequalities (valid for all events A and B):

$$\frac{r(A)}{r(B)} \leq k^2 \frac{p(A)}{p(B)}.$$

Since the class is marginalization and conditionalization invariant, the first step is to obtain $p(\breve{x}|e)$, the marginal posterior for the base distribution of the class. Now we can set up a linear programming problem:

$$\max \left(\sum_{x \in \breve{x}} u(\breve{x}) r(\breve{x}|e) \right)$$

subject to

$$r(x|e) \leq k^2 \frac{p(x|e)}{p(y|e)} r(y|e),$$

where x, y are arbitrary elements of $\breve{x}$ (if $\breve{x}$ has n elements, there are $n(n-1)$ inequalities). This procedure produces the upper bound; the lower bound is obtained by minimization.

For expected value calculations, maximization/minimization of expectation for $u(\tilde{x}) = x_q$ can be easily performed. For calculation of posterior marginals, take $u(\tilde{x}) = \delta_a(x_q)$, where $\delta_a(x_q)$ is one if $x_q = a$ and zero otherwise. In this case $\overline{E}[u] = \overline{p}(x_q = a|e)$, the posterior probability for x_q. The linear programming problem can be solved in closed-form [Seidenfeld & Wasserman, 1993]:

$$\overline{r}(x_q = a|e) = \frac{kp(x_q = a|e)}{kp(x_q = a|e) + p(x_q = a^c|e)},$$

$$\underline{r}(x_q = a|e) = \frac{p(x_q = a|e)}{p(x_q = a|e) + kp(x_q = a^c|e)}.$$

The linear programming problem above can be intractable if $\breve{x}$ has too many variables. In this case a Gibbs sampling procedure can be applied to the problem. Consider a sample of the posterior distribution $p(\tilde{x}|e)$ with N elements X_j, which can be produced through Gibbs sampling techniques [York, 1992]. The following expression converges to the upper expectation of a function $u(\cdot)$ [Wasserman & Kadane, 1992]:

$$\max_j \left(\frac{1}{1 + (1 - (j/N))(k-1)} \left(\frac{(k-1)Z_j}{N} + \frac{Z_0}{N} \right) \right).$$

where

$$Z_0 = \sum_j u(X_j)$$

$$Z_j = \sum_{l \geq j} u_{(l)}$$

The value $u_{(l)}$, used here and in the next sections, is the lth value of $u(X_j)$ as the N values are ordered from smallest to largest.

6 Constant density bounded global neighborhoods

A density bounded class is the set of all distributions such that [Lavine, 1991]

$$\Gamma^B_{l,u}(p(x)) = \{r(x) : l(x) \leq r(x) \leq u(x)\},$$

where $l(\cdot)$ and $u(\cdot)$ are arbitrary non-negative functions such that $\sum_x l(x) \leq 1$ and $\sum_x u(x) \geq 1$.

A global neighborhood can be constructed with a subclass of the density bounded class. Take a base Bayesian network and a positive constant $k > 1$. Consider the set of all joint distributions such that:

$$\Gamma_k^B(p(\tilde{x})) = \left\{ r(\tilde{x}) : (1/k) \prod_i p_i \leq r(\tilde{x}) \leq k \prod_i p_i \right\}.$$

Call this class a *constant density bounded* class. We consider calculation of the upper bound; the lower bound can be obtained through similar methods.

The constant density bounded class is invariant to marginalization, but not to conditionalization [Wasserman & Kadane, 1992]. To obtain posterior bounds, we must resort to Lavine's algorithm [Cozman, 1996]. The algorithm brackets the value of the posterior upper expectation of $u(\cdot)$ by successive calculations of a *prior* upper expectation:

$$\overline{E}_k[u] = \max \left[\sum_{x \notin e} \left((u(\breve{x}) - k) \prod_i p_i^e \right) \right],$$

where k is a given real number. The key observation is that the value of $\overline{E}_k[u]$ is zero only when k is equal to the posterior upper expectation. The algorithm starts at an arbitrary k and increases or decreases it until the expression above is zero.

To obtain $\overline{E}_k[u]$, we use the marginalization invariant property of the constant density bounded class. First marginalize $p(\tilde{x})$ to $p^e(\breve{x})$ with any standard algorithm for Bayesian networks. Now we can set up a linear programming problem:

$$\max \left(\sum_{x \in \breve{x}} (u(\breve{x}) - k) r^e(\breve{x}) \right)$$

subject to

$$(1/k)p^e(x) \leq r^e(x) \leq kp^e(x),$$

where the x are arbitrary elements of $\breve{x}$.

For expected value calculations, maximization/minimization of expectation for $u(\tilde{x}) = x_q$ can be easily performed. For calculation of posterior, $u(\tilde{x}) = \delta_a(x_q)$, where $\delta_a(x_q)$ is one if $x_q = a$ and zero otherwise. In this case $\overline{E}[u] = \overline{p}(x_q = a|e)$, the posterior probability for x_q. The linear programming problem can be solved in closed-form [Wasserman, 1990]:

$$\overline{r}^e(x_q = a) = \min \left(kp^e(x_q = a), 1 - (1/k) \sum_{x_q \neq a} p^e(x_q) \right)$$

$$\underline{r}^e(x_q = a) = \max \left((1/k)p^e(x_q = a), 1 - k \sum_{x_q \neq a} p^e(x_q) \right).$$

The linear programming problem above can be intractable if $\breve{x}$ has too many variables. In this case a Monte Carlo sampling procedure can be applied to the problem [Wasserman & Kadane, 1992]. Consider a sample of $p(\tilde{x})$ with N elements X_j. The following expression converges to the upper expectation of a function $u(\cdot)$:

$$(1/k)\frac{Z_1}{N} + k\frac{Z_2}{N}$$

where

$$Z_1 = \sum_{l=1}^{(nk/(k+1))} u_{(l)}$$

$$Z_2 = \sum_{l=(nk/(k+1))+1}^{N} u_{(l)}.$$

7 Total variation class

The total variation class is defined as the set of distributions such that:

$$\Gamma_k^T(p(\tilde{x})) = \{ r(\tilde{x}) : \forall A, \ |p(A) - r(A)| \leq \epsilon \}.$$

Almost all the discussion for the constant density bounded class carries to the total variation class since both classes are invariant to marginalization but not to conditionalization. We perform calculation of the upper bound through Lavine's algorithm. First marginalize $p(\tilde{x})$ to $p^e(\breve{x})$ with any standard algorithm for Bayesian networks. Now we can set up a linear programming problem:

$$\max \left(\sum_{x \in \breve{x}} (u(\breve{x}) - k) r^e(\breve{x}) \right)$$

subject to

$$p^e(A) - \epsilon \leq r^e(A) \leq p(A) + \epsilon,$$

where the A are arbitrary elements of $\breve{x}$.

For expected value calculations, maximization/minimization of expectation for $u(\tilde{x}) = x_q$ the can be easily performed. For calculation of posterior, $u(\tilde{x}) = \delta_a(x_q)$, where $\delta_a(x_q)$ is one if $x_q = a$ and zero otherwise. In this case $\overline{E}[u] = \overline{p}(x_q = a|e)$, the posterior probability for x_q. The linear programming problem can be solved in closed-form [Wasserman, 1990]:

$$\overline{r}^e(x_q = a) = \min\left(1, p^e(x_q = a) + \epsilon\right)$$

$$\underline{r}^e(x_q = a) = \max\left(0, p^e(x_q = a) - \epsilon\right).$$

The linear programming problem above can be intractable if $\breve{x}$ has too many variables. In this case a Monte Carlo sampling procedure can be applied to the problem [Wasserman & Kadane, 1992]. Consider a sample of $p(\tilde{x})$ with N elements X_j. The following expression converges to the upper expectation of a function $u(\cdot)$:

$$\epsilon \max u^e(\breve{x}) + \frac{Z_0}{N},$$

where

$$Z_0 = \sum_{l=\epsilon N}^{N} u_{(l)}.$$

8 Calculation of variance

Probabilistic inference uses variance $V_p[x_q] = E_p[x_q^2] - (E_p[x_q])^2$ of a variable x_q, for a fixed probability distribution $p(x_q)$.

Lower and upper variances are defined as:

$$\underline{V}[x_q] = \min_p V_p[x_q] \qquad \overline{V}[x_q] = \max_p V_p[x_q]$$

Calculation of bounds for variances in a Quasi-Bayesian network is a great challenge because the expression for $V_p[x_q]$ is quadratic on the probability values.

To produce a convergent algorithm for calculation of lower and upper variances, we can use Walley's variance envelope theorem [Walley, 1991, Theorem G2],

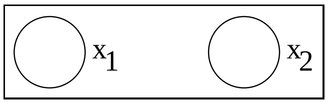

Figure 1: Bayesian network with two independence variables

which demonstrates that $\underline{V}[x_q] = \min_\mu(\underline{E}[(x_q - \mu)^2])$ and $\overline{V}[x_q] = \min_\mu(\overline{E}[(x_q - \mu)^2])$. The calculation of lower and upper variances becomes a unidimensional optimization problem, which can be solved by discretizing μ (note that μ must be larger than zero and smaller than the square of the largest value of x_q). The computational burden of this procedure is very intense since for each value of μ it is necessary to obtain the bounds for expected value of $u(x_q) = (x_q - \mu)^2$.

9 Independence relations

A Bayesian network exhibits several independence relations in a joint distribution. In general, neighborhoods of a Bayesian network do not preserve all the decompositions in the original joint distribution [Berger & Moreno, 1994; Walley, 1991]. This fact has led to considerable discussion in the literature [Chrisman, 1996a; Seidenfeld & Wasserman, 1993]. Here it should not disturb us, as we are interested in creating neighborhoods of probabilistic models: it is not surprising that a neighborhood of a model should contain some distributions which do not obey exactly the same properties of the model.

To illustrate this point, consider a network with two variables, x_1 and x_2, with no arrow between them (Figure 1). x_1 is associated with a probability distribution $p_1(x_2)$ and x_2 is associated with a probability distribution $p_2(x_2)$. A global density ratio neighborhood is constructed:

$$(1/k)p_1(x_1)p(x_2) \leq \alpha r(x_1, x_2) \leq k p_1(x_1)p(x_2).$$

This neighborhood includes many models $r(x_1, x_2)$ where x_1 and x_2 are related and dependent.

10 Conclusion

We presented exact solutions for Bayesian networks associated with global neighborhoods. Such neighborhoods specify perturbations in the joint probabilistic model, and are useful when one seeks to assess the global influence of parameter variations and independence assumptions. Inference bounds must be carefully analyzed, since global perturbations over multivariate structures may lead to very large intervals of probability.

Bayesian networks have not been analyzed with respect to robustness to structure, i.e., how the inferences degrade as the structure of the network loses accuracy. Such analysis must be part of inference procedures, but so far the classes that admit exact inference have been restrictive or inefficient. This paper offers the first analysis of global neighborhoods for Bayesian networks, demonstrating that some classes of models are tractable for inferences.

Acknowledgements

We thank Eric Krotkov for substantial support during the research that led to this work, and Lonnie Chrisman for reading an earlier draft and suggesting substantial improvements.

References

[Berger & Moreno, 1994] Berger, J., and Moreno, E. 1994. Bayesian robustness in bidimensional model: Prior independence. *Journal of Statistical Planning and Inference* 40:161–176.

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

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

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

[Cannings & Thompson, 1981] Cannings, C., and Thompson, E. A. 1981. *Genealogical and Genetic Structure.* Cambridge: Cambridge University Press.

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

[Chrisman, 1996a] Chrisman, L. 1996a. Independence with lower and upper probabilities. *Proc. Twelfth Conference Uncertainty in Artificial Intelligence* 169–177.

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

[Cozman, 1996] Cozman, F. 1996. Robust analysis of bayesian networks with finitely generated convex sets of distributions. Technical Report CMU-RI-TR96-41, Carnegie Mellon University.

[Dechter, 1996] Dechter, R. 1996. Bucket elimination: A unifying framework for probabilistic inference. *Proc. of the Twelfth Conference Uncertainty in Artificial Intelligence* 211–219.

[DeRobertis & Hartigan, 1981] DeRobertis, L., and Hartigan, J. A. 1981. Bayesian inference using intervals of measures. *The Annals of Statistics* 9(2):235–244.

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

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

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

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

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

[Seidenfeld & Wasserman, 1993] Seidenfeld, T., and Wasserman, L. 1993. Dilation for sets of probabilities. *The Annals of Statistics* 21(9):1139–1154.

[Shenoy & Shafer, 1990] Shenoy, P. P., and Shafer, G. 1990. Axioms for probability and belief-function propagation. In Shachter, R. D.; Levitt, T. S.; Kanal, L. N.; and Lemmer, J. F., eds., *Uncertainty in Artificial Intelligence 4*. North-Holland: Elsevier Science Publishers. 169–198.

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

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

[Wasserman & Kadane, 1992] Wasserman, L., and Kadane, J. B. 1992. Computing bounds on expectations. *Journal of the American Statistical Association* 87(418):516–522.

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

[Wasserman, 1992a] Wasserman, L. 1992a. Invariance properties of density ratio priors. *The Annals of Statistics* 20(4):2177–2182.

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

[York, 1992] York, J. 1992. Use of the gibbs sampler in expert systems. *Artificial Intelligence* 56:115–130.

[Zhang & Poole, 1996] Zhang, N. L., and Poole, D. 1996. Exploiting causal independence in Bayesian network inference. *Journal of Artificial Intelligence Research* 301–328.