Independence Relations in the Robustness Analysis of Multivariate Probabilistic Models

Fabio Gagliardi Cozman

Escola Politécnica, University of São Paulo
fgcozman@usp.br, http://www.cs.cmu.edu/~fgcozman

Abstract— A common technique in robustness analysis is to study the effect of parameter variations. For instance, a designer may be interested in the robustness of stability measures against parameter variations. This paper concentrates on the robustness of inferences and estimation with respect to multivariate probabilistic models. The idea is to represent multivariate structures through a class of models developed in Artificial Intelligence, called Bayesian networks, and to encode variations in these networks by convex sets of probability distributions. Walley's definition of independence is used to define the concept of type-1 extensions, which correspond to the usual concept of joint distribution. The paper presents an analysis of independence relations in type-1 extensions; the main contribution is a novel theorem that indicates all independence relations that can be detected in a type-1 extension by polynomial-time algorithms.

Key Words— Robust estimation; Graph theoretic models; Inference engines; Multivariate systems.

1 Introduction

To accommodate qualitative and quantitative errors in the modeling process, a reasonable design strategy is to investigate the *robustness* of a model to variations and perturbations. Robust Control and Robust Statistics are fields that investigate variations of output measures with respect to variations in model parameters. For instance, it may be important to know the extent to which variations in a system affect the system stability, both for small (differential) (Frank, 1978) or large (interval) variations (Weinman, 1991).

A common technique for robustness analysis is to consider the exhaustive collection of values that some parameters can assume, and obtain bounds on output values induced by such variations. For instance, this strategy is used in the Robust Control literature when bounded perturbations are assumed for parameters in a plant, and stability measures are calculated for all possible values of parameters.

Likewise, it is interesting to analyze variations in the stochastic aspects of a model. Most of the literature on robustness analysis for probabilistic models has appeared in the field of Robust Statistics, where the main concern is the calculation of expected utility for models with low-dimensionality (univariate and bivariate models) (Berger, 1990; Kadane, 1984; Wasserman, 1992).

The objective of this paper is to provide the basis for a comprehensive investigation of multivariate robustness analysis in probabilistic models. The idea is to start with a class of mathematical structures called Bayesian networks, that were developed in Artificial Intelligence for the representation of multivariate probabilistic models. Bayesian networks are then modified to allow for the presence of convex sets of probability distributions. Convex sets of probability distribution are used to model variations in standard probability distributions. The resulting structures are called *Quasi-Bayesian networks*.

The paper contributes with an analysis of the independence relations that exist in Quasi-Bayesian networks. Walley's definition of independence is used to build the concept of type-1 extensions, which corresponds to the usual concept of joint distribution. The main contribution of this paper is a theorem that clarifies all independence relations that can be detected in a type-1 extension by polynomial-time algorithms. Algorithms that benefit from this result are mentioned in the paper, and a system that implements these algorithms is described.

2 Paper roadmap

Sections 3 to 6 present the rationale that motivates the paper and review the background needed for the technical results. Section 3 reviews the ideas behind the multivariate models developed in Artificial Intelligence. Bayesian networks are briefly described in Section 4. Several approaches that expand or modify standard Bayesian networks are considered in Section 5; Section 5 also introduces the basic rationale for the use of convex sets of distributions. The basic theory of convex sets of distributions, which is used to model perturbations in probabilistic models, is presented in Section 6.

Section 7 introduces the concepts of Quasi-Bayesian networks and type-1 extensions; these structures are generated by associating convex sets of distributions with directed acyclic graphs. Section 8 presents the main result of the paper: a novel theorem that elucidates the close relationship between independence and d-separation in type-1 extensions. The algorithmic consequences of this result are discussed in Section 9, and closing remarks are left to Section 10.

3 Graphical models for probabilistic structures

There are several ways to build a probabilistic model for a collection of variables. The brute force approach is to explicitly specify the joint distribution for all variables. A more economical method, usually employed in the Kalman filtering literature, is to assume that random variables are Gaussian and specify a covariance matrix (Gelb, 1974).

A more general (but still compact and efficient) strategy is to specify a collection of variables and represent the relationships among them through conditional distributions. This strategy has been actively pursued in the Artificial Intelligence community (Jensen, 1996; Pearl, 1988). In this approach, the relationships among variables are represented explicitly through a graph. The objective of a graphical model is to summarize opinions, beliefs, data and experiments about a particular domain. A large model can be constructed from small blocks that are arbitrarily interconnected in a graphical structure. In this context, an *inference* is a calculation of posterior probabilities given some evidence.

The most popular graphical representation of probabilistic models in Artificial Intelligence is the *Bayesian network* formalism (presented in Section 4). Currently the main application of Bayesian networks is the representation of discrete, finite, static systems. The results discussed in this paper are completely general and do not assume special spaces or particular network topologies; results presented here are a step in the direction of a full theory of robustness analysis for general systems.

4 Bayesian Networks

A Bayesian network specifies a joint distribution over a set of variables $\tilde{X}$ through a directed acyclic graph (Jensen, 1996). Each node of a Bayesian network is associated with a variable X_i; the parents of X_i are denoted by $\mathrm{pa}(X_i)$. Each variable X_i has a finite set of values $\hat{X}_i$; the set of subsets of $\hat{X}_i$ is denoted $\mathcal{X}_i$.

Each variable in a Bayesian network is associated with a conditional distribution $p(X_i|\mathrm{pa}(X_i))$. Such a graphical structure defines a unique joint probability distribution through the following expression (Pearl, 1988):

$$p(\tilde{X}) = \prod_i p(X_i|\mathrm{pa}(X_i)). \qquad (1)$$

Bayesian networks represent many independence relations among the variables in the network; these relations can be analyzed through the concept of *d-separation*. The definition of d-separation is somewhat involved (Pearl, 1988, page 117). Given a variable X and a set of variables $\tilde{Z}$, take $\tilde{Y}$ to be the set of all variables that are d-separated from X given $\tilde{Z}$. This means that along every path between X and a node in $\tilde{Y}$, there is a node W such that: either W has both converging arrows and is not in $\tilde{Z}$ and none of its descendants are in $\tilde{Z}$, or W has no converging arrows and is in $\tilde{Z}$.

A property of Bayesian networks is that, if $\tilde{Y}$ d-separates X from $\tilde{Z}$, then X and $\tilde{Z}$ are independent given events in $\tilde{\mathcal{Y}}$. Note that d-separation is only a *graphical* relation, and independence is only a *probabilistic* relation; one of the fundamental properties of Bayesian networks is that these two relations are equivalent (Pearl, 1988). Any directed acyclic graph displays d-separation relations, which can be detected by a polynomial-time algorithm (Geiger et al., 1990).

5 Relaxing the Bayesian requirements

One difficulty with Bayesian networks is the requirement that all probability distributions must be precisely specified. Several non-probabilistic attempts have been made to relax the requirements of Bayesian networks. For example, there are multivariate representations based on possibility theory (Gebhardt and Kruse, 1995) and Dempster-Shafer theory (Shenoy and Shafer, 1990), and representations based on purely heuristic grounds (Shortliffe and Buchanan, 1985).

Another strategy is to associate interval-valued probability to variables in a network (Chrisman, 1996b; Fertig and Breese, 1990; H. E. Kyburg Jr., 1987). But interval-valued representations face two difficulties. First, it is not always possible to apply Bayes rule to an interval-valued distribution and obtain an interval-valued posterior distribution (Chrisman, 1995; H. E. Kyburg Jr., 1987). Second, there is no unique, accepted way to define independence for interval-valued distributions (Chrisman, 1996a). Given special definitions of conditionalization and independence, only a few classes of interval-valued distributions actually admit exact, efficient solutions (Chrisman, 1996b; Zaffalon, 1997).

Closed convex sets of distributions are also models for imprecision in probability values, and display several advantages when compared to interval-valued probability. Closed convex sets of distributions can be axiomatized (just as probability theory can be axiomatized (Savage, 1972)), and conditionalization and independence can be defined without technical difficulties. These interesting properties have been actively explored in the field of Robust Bayesian Statistics, with an emphasis on univariate and bivariate models (Berger, 1990; Kadane, 1984; Wasserman, 1992).

In this paper, closed convex sets of distributions are employed as the basic model for perturbations in a probabilistic structure. A set of distributions is understood as a fundamental entity that reflects somewhat imprecise beliefs or incomplete data about stochastic phenomena.

One axiomatization of closed convex sets of distributions that is particularly concise and powerful is the *Quasi-Bayesian* theory of Giron and Rios (Giron and Rios, 1980). This theory is summarized in the next section; several recent definitions and results, not present in the original theory by Giron and Rios, are incorporated in the presentation.

6 Quasi-Bayesian theory

Quasi-Bayesian theory uses convex sets of distributions to represent beliefs and to evaluate decisions (Giron and Rios, 1980). Several other theories use similar representations: inner/outer measures (Good, 1983; Halpern and Fagin, 1992; Ruspini, 1987; Suppes, 1974), lower probability theory (Breese and Fertig, 1991; Chrisman, 1996b; Fine, 1988; Smith, 1961), convex Bayesianism (H. E. Kyburg Jr., 1987), Dempster-Shafer theory (Shafer, 1976), probability/utility sets (Seidenfeld, 1993).

The convex set of distributions maintained by an agent is called the *credal* set, and its existence is postulated on the grounds of axioms about preferences (Giron and Rios, 1980).[a] To simplify terminology, the term credal set is used only when it refers to a set of distributions containing more than one element. A closed convex set of joint distributions is called a *joint credal set.*

6.1 Lower and upper values

Given a convex set K of probability distributions, a probability interval can be created for every event A by defining lower and upper bounds:

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

Lower and upper bounds are closely related by the expression $\underline{p}(A) = 1 - \overline{p}(A^c)$.

Lower and upper expectations for a function $f(X)$ are defined as:

$$\underline{E}[f] = \inf_{p \in K} E_p[f] \qquad \overline{E}[f] = \sup_{p \in K} E_p[f].$$

There is a one-to-one correspondence between lower (or upper) expectations and credal sets. Given a credal set, the set of all lower expectations for all arbitrary functions $f(X)$ is unique, and vice-versa.

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

6.2 Conditionalization

Convex sets of conditional distributions are used to represent conditional beliefs. Inference is performed by applying Bayes rule to each distribution in a jont credal set. The posterior credal set is the union of all posterior distributions obtained in this process. The vertices of the posterior credal set are obtained by applying Bayes rule to all vertices of the joint credal set (Walley, 1991).

Consider a variable Y with a set of values $\hat{Y}$, and denote by $\mathcal{Y}$ the set of all subsets of $\hat{Y}$. For any event A in $\mathcal{Y}$, take the conditional credal set $K(X|A)$ for a variable X. The collection of all such $K(X|A)$ is denoted by $K(X|Y)$.

6.3 Independence

There is no unique way to define independence relations with credal sets; in the most in-depth study of this matter, Campos and Moral have reviewed five different possible types of independence (de Campos and Moral, 1995). The results presented in this paper adopt Walley's definition of independence (Walley, 1991). This definition is stated in terms of lower expectations; to develop a theory of convex sets of distributions, it is important to recast Walley's definition using credal sets:

Definition 1. *Consider two sets of variables X and Y and the credal sets $K(X,Y)$, $K(X)$, $K(Y)$, $K(X|Y)$ and $K(Y|X)$. Note that distributions in $K(X)$ and $K(X|Y)$ are defined over the same algebra of events once Y is fixed; likewise, distributions in $K(Y)$ and $K(Y|X)$ are defined over the same algebra of events once X is fixed.*

Variable Y is irrelevant *to X if $K(X)$ is equal to $K(X|Y)$ regardless of the value of Y. Variables X and Y are* independent *if X is irrelevant to Y and Y is irrelevant to X.*

This concept of independence does not imply that joint credal sets contain only joint distributions with independent marginals, and it does not imply uniqueness for the joint credal set.

Concepts of conditional irrelevance and independence can be defined as well. Consider a collection of variables with a set of joint distributions K, and three groups of variables in this collection, $\tilde{X}$, $\tilde{Y}$ and $\tilde{Z}$. Say that $\tilde{Y}$ is irrelevant to $\tilde{X}$ given $\tilde{Z}$ if $K(\tilde{X}|\tilde{Z})$ is equal to $K(\tilde{X}|\tilde{Y},\tilde{Z})$, regardless of the value of $\tilde{Z}$. Say that $\tilde{X}$ is independent of $\tilde{Y}$ given $\tilde{Z}$ if $\tilde{X}$ is irrelevant to $\tilde{Y}$ given $\tilde{Z}$ and $\tilde{Y}$ is irrelevant to $\tilde{X}$ given $\tilde{Z}$.

7 Quasi-Bayesian networks and type-1 extensions

The previous sections discussed various graphical representations for multivariate models, and presented tools for representation of perturbations in probabilistic models. The goal of this section is to unify these ideas into a framework for robustness analysis of multivariate models. This goal is achieved through the definition of a new type of network, called locally defined Quasi-Bayesian network.

A locally defined Quasi-Bayesian network is a directed acyclic graph where each variable is associated with a conditional credal set. It may happen that a variable is in fact associated with a single conditional distribution $p(X_i|\mathrm{pa}(X_i))$, but in general a conditional credal set $K(X_i|\mathrm{pa}(X_i))$ is given for each variable X_i.

The key point is that Walley's definition of independence does not induce a single joint credal set from a collection of conditional credal sets. The quantitative structure defined by conditional credal sets is not sufficient to fully specify a Quasi-Bayesian network, because there are several ways to form a joint credal set from a collection of conditional credal sets (Section 6.3).

There are several strategies to generate a joint credal set from conditional credal sets. For example, Walley calls *natural extension* the largest possible convex set of joint distributions that are compatible with all conditional credal sets (Walley, 1991, pages 453, 455). To fully specify a natural extension, it is necessary to indicate which irrelevance and independence relations must be satisfied by the joint credal set; there may be several different natural extensions for the same quantitative structure.

Joint credal sets can also be generated by strategies that do not state independence relations explicitly. The most common approach is to use *type-1 extensions* (Cano et al., 1994; Cano et al., 1993; Tessem, 1992; Zaffalon, 1997). The idea is to cross-multiply distributions in conditional and marginal sets, and take the convex hull of all these joint distributions.This convex hull is the type-1 extension of the conditional and marginal sets. A type-1 extension is the convex hull of all the joint distributions formed by cross-multiplication of *extreme points* of local credal sets; consequently, a type-1 extension is the largest joint credal set where all extreme points satisfy Expression (1) (Section 4).

Type-1 extensions are intuitive generalizations for standard joint distributions, and they are commonly used to *define* the meaning of independence (de Campos and Moral, 1995). The difficulty is that type-1 extensions are not obtained from considerations of independence, and it is not clear what is obtained in terms of independence relations from a type-1 extension. The goal of this paper is to address these questions and provide a theorem that answers them; this theorem is presented in the next section.

8 Main result

The appeal of type-1 extensions comes from their intuitive similarity with standard Bayesian networks. The main result of this paper is the following theorem, which provides a formalization for this intuitive similarity through Walley's definition of independence:

Theorem 1. *Every graphical d-separation relation in a Quasi-Bayesian network corresponds to a valid conditional independence relation in the type-1 extension of the network.*

Proof. Consider three arbitrary disjoint sets of variables in the network, $\tilde{X}$, $\tilde{Y}$ and $\tilde{Z}$, such that $\tilde{X}$ is d-separated from $\tilde{Z}$ given $\tilde{Y}$. Take the type-1 extension $K(\tilde{X}, \tilde{Y}, \tilde{Z})$ and obtain, by conditionalization (elementwise application of Bayes rule), the sets $K(\tilde{X}|\tilde{Y}, \tilde{Z})$ and $K(\tilde{X}|\tilde{Y})$. Call $\mathrm{ext}K$ the set of extreme points of K.

Given any function $f(\tilde{X})$ solely of $\tilde{X}$, obtain its lower expectation $\underline{E}[f(\tilde{X})|\tilde{Y}, \tilde{Z}]$ by the calculation:

$$\min_{p \in \mathrm{ext}K(\tilde{X}|\tilde{Y},\tilde{Z})} \left(\sum_{\tilde{X}} f(\tilde{X}) p(\tilde{X}|\tilde{Y}, \tilde{Z}) \right).$$

The minimum is attained at an extreme point of the type-1 extension. Because every such extreme point satisfies Expression (1), $p(\tilde{X}|\tilde{Y}, \tilde{Z}) = p(\tilde{X}|\tilde{Y})$ (by d-separation). The lower expectation $\underline{E}[f(\tilde{X})|\tilde{Y}, \tilde{Z}]$ is then:

$$\min_{p \in \mathrm{ext}K(\tilde{X}|\tilde{Y})} \left(\sum_{\tilde{X}} f(\tilde{X}) p(\tilde{X}|\tilde{Y}) \right)$$
$$= \underline{E}[f(\tilde{X})|\tilde{Y}].$$

Because a lower expectation uniquely defines a convex set of distributions (Section 6), the lower expectation $\underline{E}[f(\tilde{X})|\tilde{Y}]$ uniquely defines $K(\tilde{X}|\tilde{Y})$ and the lower expectation $\underline{E}[f(\tilde{X})|\tilde{Y}, \tilde{Z}]$ uniquely defines $K(\tilde{X}|\tilde{Y}, \tilde{Z})$. Because both lower expectations are equal for arbitrary f, the underlying credal sets are the same.

This argument guarantees that $\tilde{Z}$ is irrelevant to $\tilde{X}$ given $\tilde{Y}$; the same argument proves that $\tilde{X}$ is irrelevant to $\tilde{Z}$ given $\tilde{Y}$. So $\tilde{X}$ is independent of $\tilde{Z}$ given $\tilde{Y}$. $\square$

This result demonstrates that type-1 extensions mimic the properties of standard Bayesian networks as independence-maps (Pearl, 1988, page 119). The theorem also complements results by Cano et al. (Cano et al., 1993). They give conditions on independence concepts that satisfy d-separation in type-1 extensions, but they do not provide any definition of independence to illustrate their result. The theorem demonstrates that Walley's independence relations exhibit the desired correspondence with d-separation.

9 Inferences with type-1 extensions and the *JavaBayes* system

The importance of theorem 1 lies in the fact that d-separation has important algorithmic consequences. Graphical operations that are guaranteed by d-separation can be performed in a type-1 extension in the same way that they can normally be applied to a standard Bayesian network. In particular, consider a query involving a variable X_q and evidence E. All variables that do not affect computation of $p(X_q|E)$ can be detected and discarded through d-separation computations (Geiger et al., 1990).

Several algorithms for type-1 extensions can be improved using this fact, because variables can be analyzed and potentially discarded through d-separation before any inference algorithm is run. This greatly reduces the computational effort in Quasi-Bayesian inferences with type-1 extensions.

This fact is used in an implementation of Quasi-Bayesian inferences in the *JavaBayes* system. The *JavaBayes* system is a portable and freely distributed inference engine for graphical models, developed as a testbed for robustness analysis through Quasi-Bayesian models. Documentation, code and examples for *JavaBayes* can be downloaded from http://www.cs.cmu.edu/~javabayes/.

The main inference and estimation engine in *JavaBayes* uses standard algorithms to perform calculation of posterior marginals, expectations, maximum a posteriori explanations and maximum a posteriori expectations. The *JavaBayes* system can be used to construct type-1 extensions when credal sets are associated to variables in a network. D-separation relations are always used by the inference engine to detect variables that can be discarded in a given computation.

10 Conclusion

Robustness analysis of multivariate probabilistic models is an area of practical and theoretical importance. One particular technique to conduct robustness analysis, emphasized in the present work, is to employ convex sets of probability distributions to represent variations in probability values. This paper provides a collection of tools and results that advances our understanding of multivariate convex sets of probability distributions.

The paper contributes by applying Walley's definition of independence to the study of type-1 extensions. Theorem 1 is a novel result that characterizes the relationship between properties of directed acyclic graphs and properties of Quasi-Bayesian networks. The theorem establishes a close connection between standard Bayesian networks and the Quasi-Bayesian counterparts; this connection indicates that many important algorithms for Bayesian networks can produce inferences with Quasi-Bayesian networks.

Many questions of theoretical significance remain for Quasi-Bayesian networks. What is the status of d-separation when other types of extensions or other definitions of independence are used? How to introduce time and dynamic behavior into the structure of a directed acyclic graph? These questions can be best understood and answered from the perspective of type-1 extensions developed in this paper.

Acknowledgements

I thank my former advisor, Eric Krotkov, for substantial support during the research that led to this work.

This research was conducted in the Robotics Institute, Carnegie Mellon University. The project was partially supported by NASA under Grant NAGW-1175; the author was supported under a scholarship from CNPq, Brazil.

References

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

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

Cano, A., Cano, J. E. and Moral, S. (1994). Convex sets of probabilities propagation by simulated annealing, *Fifth IPMU* pp. 4–8.

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

Chrisman, L. (1996a). Independence with lower and upper probabilities, *XII Uncertainty in Artificial Intelligence Conference* pp. 169–177.

Chrisman, L. (1996b). Propagation of 2-monotone lower probabilities on an undirected graph, *XII Uncertainty in Artificial Intelligence Conference* pp. 178–186.

de Campos, L. and Moral, S. (1995). Independence concepts for convex sets of probabilities, *XI Uncertainty in Artificial Intelligence* .

Fertig, K. W. and Breese, J. S. (1990). Interval influence diagrams, *in* M. Henrion, R. D. Shachter, L. N. Kanal and J. F. Lemmer (eds), *Uncertainty in Artificial Intelligence 5*, Elsevier Science Publishers, North-Holland, pp. 149–161.

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

Frank, P. M. (1978). *Introduction to System Sensitivity Theory*, Academic Press, New York.

Gebhardt, J. and Kruse, R. (1995). Learning possibilistic networks from data, *Fifth International Workshop on Artificial Intelligence and Statistics* .

Geiger, D., Verma, T. and Pearl, J. (1990). d-separation: from theorems to algorithms, *Uncertainty in Artificial Intelligence 5* .

Gelb, A. (1974). *Applied Optimal Estimation*, MIT Press, Cambridge.

Giron, F. J. and Rios, S. (1980). 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 (eds), *Bayesian Statistics*, University Press, Valencia, Spain, pp. 17–38.

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

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

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

Jensen, F. V. (1996). *An Introduction to Bayesian Networks*, Springer Verlag, New York.

Kadane, J. B. (1984). *Robustness of Bayesian Analyses*, Vol. 4 of *Studies in Bayesian econometrics*, Elsevier Science Pub. Co., New York.

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

Ruspini, E. H. (1987). The logical foundations of evidential reasoning, *Technical Report SRIN408*, SRI International.

Savage, L. J. (1972). *The Foundations of Statistics*, Dover Publications, Inc, New York.

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

Shafer, G. (1976). *A mathematical theory of evidence*, Princeton University Press.

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

Shortliffe, E. H. and Buchanan, B. G. (1985). *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project*, The Addison-Wesley series in artificial intelligence, Addison-Wesley, Reading, Mass.

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

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

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

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, New York.

Wasserman, L. (1992). Recent methodological advances in robust Bayesian inference, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds), *Bayesian Statistics 4*, Oxford University Press, pp. 483–502.

Weinman, A. (1991). *Uncertain Models and Robust Control*, Springer-Verlag, New York.

Zaffalon, M. (1997). *Inferenze e Decisioni in Condizioni di Incertezza con Modelli Grafici Orientati*, PhD thesis, Università di Milano.