Special Issue on Combining Probability and Logic
Introduction

Fabio Cozman, Rolf Haenni, Jan-Willem Romeijn,
Federica Russo, Gregory Wheeler and Jon Williamson

Draft of November 6, 2007

This volume arose out of an international, interdisciplinary academic network on *Probabilistic Logic and Probabilistic Networks*, called *progicnet* and funded by the Leverhulme Trust from 2006–8. Many of the papers in this volume were presented at an associated conference, the Third Workshop on Combining Probability and Logic (Progic 2007), held at the University of Kent on 5–7 September 2007. The papers in this volume concern either the special focus on the connection between probabilistic logic and probabilistic networks or the more general question of the links between probability and logic. Here we introduce probabilistic logic, probabilistic networks, current and future directions of research and also the themes of the papers that follow.

1 What is Probabilistic Logic?

Probabilistic logic, or *progic* for short, can be understood in a broad sense as referring to any formalism that combines aspects of both probability theory and logic, or in a narrow sense as a particular kind of logic, namely one that incorporates probabilities in the language or metalanguage.

In the latter case, if the probabilities are incorporated directly into the logical language we have what might be called an *internal* progic. An example is a first-order language where one or more of the function symbols are intended to refer to probability functions. Thus one can form expressions like $(P_1(Fa) = 0.2 \wedge P_2(Rab) \geq 0.5) \rightarrow Gb$. This kind of language is suitable for *reasoning about probabilities* and is explored by Halpern (2003), for instance. If, on the other hand, the probabilities are incorporated into the meta-language we have an *external* progic. For example, one might attach probabilities to sentences of a propositional language: $(p \wedge q) \rightarrow r^{0.95}$. This kind of language is suitable for *reasoning under uncertainty*, and maintains a stricter distinction between the level of logic and the level of probability (see, e.g., Paris, 1994). A logic that incorporates probabilities both within the language and the metalanguage is a *mixed* progic.

The central question facing a progic is what conclusions should be drawn from given premisses. For an internal progic the question is which ψ to conclude from given premisses $\varphi_1, \ldots, \varphi_n$, where these are sentences of a language involving probabilities. This is analogous to the question facing classical logic. But for an external (or mixed) progic, the question is rather different. Instead of asking what ψ^Y to conclude from given premisses ${\varphi_1}^{X_1}, \ldots, {\varphi_n}^{X_n}$ one would normally ask what Y to attach to a given ψ, when also given premisses ${\varphi_1}^{X_1}, \ldots, {\varphi_n}^{X_n}$. Note that, depending on the semantics in question, the $X_1, \ldots, X_n, Y$ might be probabilities or sets of probabilities.

Since the fundamental question of an external probabilistic logic differs from that of a non-probabilistic logic, different techniques may be required to answer this question. In non-probabilistic logics one typically appeals to a *proof theory* to determine which ψ to conclude from given $\varphi_1, \ldots, \varphi_n$. This is not always appropriate in the case of a probabilistic logic. An external progic requires machinery for manipulating probabilities, not just tools for handling sentences. In Haenni et al. (2008) it is suggested that the machinery of *probabilistic networks* can fruitfully be applied here.

2 What are Probabilistic Networks?

Probabilistic networks (or *probabilistic graphical models*) is a general term for various mathematical models in which probabilistic information is linked to network-based structural information. The network structure is usually formalized by a directed graph with nodes and arrows, where an arrow between two nodes is meant to represent some sort of *influence* or *dependency* between the variables associated with those nodes. This in turn means that the absence of an arrow between two nodes implies some sort of *independence* among the associated variables. As an example, we could thus represent our knowledge about the positive correlation between smoking and lung cancer by two network nodes S and L and an arrow from S towards L. And we could then enhance the network by nodes and corresponding arrows for further smoking-related diseases or for other possible (but possibly independent) causes of lung cancer. To avoid circular dependencies, directed graphs are normally assumed to be acyclic.

Depending on the available probabilistic information, it is common to distinguish different types of probabilistic networks. In the most simplest case of so-called *Bayesian* (or *belief*) *networks*, it is assumed that the conditional probability of each network variable given its parents is fully specified. In the example above, we could meet this requirement by specifying $P(L{=}yes|S{=}yes) = 0.05$ and $P(L{=}yes|S{=}no) = 0.01$ for the network variable L (which means that smoking increases the risk of lung cancer by a factor 5) and by assuming a prior probability $P(S{=}yes) = 0.3$ for the network variable S. What makes Bayesian networks particularly attractive is the fact that the included structural and probabilistic information is sufficient to specify a unique *joint probability function* over all involved variables, which in turn can be used to answer all sorts of probabilistic queries. Note that an explicit specification of a joint probability function would require exponentially many parameters. Bayesian networks are thus useful to specify large joint probability functions efficiently and to reduce the complexity of respective computations.

A second type of probabilistic networks, so-called *credal networks*, arise when the uniqueness assumption for the given probability values is relaxed. They are thus similar to Bayesian networks, except that they expect sets of probabilities instead of point-valued probabilities. In the example above, we could

The price of this additional

but instead of requiring conditional probabilities, they require so-called *credal* *sets* for each network variable. In the particular case of

Markov nets = undirected

This type of graphical model is known as a directed graphical model, Bayesian network, or belief network. Classic machine learning models like hidden Markov models, neural networks and newer models such as variable-order Markov models can be considered as special cases of Bayesian networks.

Graphical models with undirected edges are generally called Markov random fields or Markov networks.

A third type of graphical model is a factor graph, which is an undirected bipartite graph connecting variables and factor nodes. Each factor represents a probability distribution over the variables it's connected to. In contrast to a Bayesian network, a factor may be connected to more than two nodes.

3 Current Directions

Research on probabilistic logic has, since its beginnings, explored trade-offs between expressivity and complexity in various kinds of internal and external progics. Generally these progics let probabilities be attached to logical constructs in a very flexible manner, possibly letting *many* measures satisfy a given set of probabilistic assessments (this "single-measure vs multiple-measure" debate is further discussed in Section 5). The current literature continues to investigate the merits and the applications of progics that do not impose any special structure on probabilistic assessments. Take for instance, the recent work by Ognjanović (2006), and several papers on philosophical and psychological topics in this special issue—by Howson, by Leuridan, by Pfeifer and Kleiter, and by Sprenger.

The last fifteen years have witnessed the arrival of many progics where graphs are used to structure sentences and assessments, so as to lower the complexity of inference. Until a decade ago the work on "progics + graphs" focused on a few narrow strands; a noticeable change took place around 2000, and since then the literature has been growing at a staggering pace. Most of this recent literature employs the technology of Bayesian and Markov networks to produce progics where any set of well-formed formulas is satisfied by a *single* measure. The paper by Kee and Lloyd in this special issue gives examples where Bayesian networks are encoded through logic.

A key reason for the dramatic growth of interest in single-model progics based on graphs, particularly in the artificial intelligence literature, is the maturity of inference algorithms for Bayesian and Markov networks. Today one can employ graphs so as to produce a progic with the power to handle problems of practical significance in reasonable computing time. Relatively little attention has been given to progics that are based on graphs and yet depart from the single-measure assumption (however see the papers by Cozman et al and by Haenni in this issue).

Two further characteristics of the current work on "progics + graphs" deserve to be mentioned. First, research on progics based on probabilistic networks is heavily oriented towards applications—for instance, applications in data mining, because data are relational or structured (an example is textual data), or because it is advisable to use existing domain knowledge expressed as logical sentences (Domingos, 2007). Second, the research often focus on *learning* logical and probabilistic sentences from data. Indeed, a key reason for the surge of such progics around 2000 is that, at that point, researchers began to combine Bayesian networks and logic so as to process relational data (Friedman et al, 1999). The emphasis on single-model progics is particularly strong when learning from data is considered.

4 This Volume

A brief paragraph on each paper by the editor responsible Cozman Fabio, Cassio de Campos, Jose Eduardo Ochoa, Assembling a consistent set of sentences in relational probabilistic logic with independence

Haenni Rolf, Probabilistic Argumentation

In *Can logic be combined with probability? Some observations*, Colin Howson addresses one of the most fundamental questions underlying this special issue and more generally the Progicnet project. As Howson says, a positive answer to this question is of course possible, but that would turn out to be very trivial, as "logic can be combined with anything". A slightly more sophisticated answer would be that logic and probability can indeed be combined because of their both being formal languages: logic and probability have a common set of concept and methods, and some of their semantic aspects also turn out to be significantly similar. But the interesting question becomes whether such a formal similarity is simply *formal*, or whether there is a deeper degree of conceptual affinity. This is exactly the challenge that Howson takes up in the paper. In the first part of the paper, the author closely investigates the extent to which Gaifman, Scott and Krauss succeeded in combining probability and logic. This attempt consisted in adapting the concepts, tools and procedures of model theory in modern logic—in particular, the notions of consistency and of consquence—to provide a corresponding model theory for the probability language. Howson shows that this account isn't fully successful because expressiveness—i.e., the consideration of language systems whose expressive power is closer to the σalgebra of mathematical probability—and effectiveness—i.e., the possibility of developing a proof theory—eventually pull in opposite directions. The *pars construens* of the paper develops along the same line of Scott's and Krauss' ideas, but it differs in that it develops a formalism of epistemic probability, based on set theoretic algebras, that is a generalisation of the classical logical concepts of model, consistency, and consequence. In particular, for assignments of real-valued probabilities to elements of a field or σ-field F of sets, Howson shows that three theorems, which have their analogue (meta)results in first order logic, follow: (i) Absoluteness of consistency, (ii) Non-ampliativity of consequence, and (iii) Compactness.

Leuridan Bert, Causal Discovery and the Problem of Ignorance

Pfeifer Niki, Gernot Kleiter, Framing human inference by coherence based probability logic

The paper by Siong Kee and John Lloyd, *Probabilistic Reasoning in a Classical Logic*, explores the combination of logical and probabilistic reasoning through higher-order logics. The paper argues against the view that one must add "special" features to classical logic so as to obtain a probabilistic logic—they contend that one can handle probabilistic reasoning within classical logic through higher-order machinery. See and Lloyd indeed present a language that does so, by allowing functions to return whole densities over domains of types. See and Lloyd comment extensively on expressivity and computational complexity of higher-order progics, and make connections with recent developments in logic and artificial intelligence research. The paper also discusses several examples, and in particular an extended example that mixes individuals, relations, and Bayesian networks.

Sprenger Jan, Statististics between inductive logic and empirical science

5 Future Directions

Progic 2007 was the third in a series of workshops on combining probability and logic. The special focus of the Kent workshop was the relationship between probabilistic logic and probabilistic networks, and a number of proposals to use network structures to simplify probabilistic calculation were discussed. The sheer variety of approaches naturally raises the question of which to choose, and why. In addressing this question at the workshop three themes emerged, which suggests future directions of research.

The first issue, touched upon in the introduction, concerns the type of uncertainty to be managed. There is a natural distinction between reasoning about uncertainty and reasoning under conditions of uncertainty, which have formal analogues to internal progics and external progics. As we observed, each asks very different things from a probability logic and some disagreements over approach trace to different types of representation and reasoning problems to solve.

Second, there is a familiar trade-off in logic between expressive capacity of the representational language and inferential power of a logic: an increase in expressive capacity nearly always accompanies a decrease in capacities for the logic to effectively draw out consequences. One place this tension appears within probability logic is when discussing the relative advantages of sharp probabilities versus interval-valued probabilities. A single distribution is an easier object to compute with than a set of distributions, so there is little surprise that point-valued approaches are favored from a computational point of view. But deferring purely to computational considerations without check would yield ditching probability for propositional logic. The second issue concerns an inventory of advantages and disadvantages for adopting sharp probabilities, and comparing those to the advantages and disadvantages to using interval-valued probabilities.

In addition to practical and philosophical reasons for favoring imprecise or interval valued probabilities to sharp values (e.g., sometimes you don't have enough information to construct a full joint distribution; then what?), there are also strong theoretical reasons for paying attention to results and methods developed within interval-valued frameworks. One theme of the workshop was to demonstrate that an interval-valued framework that has as a special case one or another sharp-valued approach often gives a vivid picture of what is happening within his representation, and why. The upshot is that even those who are unmoved by the philosophical reasons for favoring interval-valued probabilities over sharp probabilities may nevertheless benefit from several mathematical insights that are gained from the shift.

A third theme concerns how to judge the correctness of a representation of uncertainty within a probabilistic logic. For many who take de Finetti as a starting point, the criteria for the representation of uncertain opinion derive from the behavioral consequences of opinion, such as buying a bet. One of the motivations for imprecise probability is that the lowest selling price for a bet may be and often is higher than the highest buying price, suggesting that the associated probability can be determined up to an interval. However, this invites the question whether a probabilistic logic should perhaps be supplemented with a decision theory, or even designed in conjunction with a decision theory from scratch. It may be that as a normative theory, a stand-alone probabilistic logic is without a firm foundation.

References

Domingos, P. (2007). Toward knowledge-rich data mining. *Data Mining and Knowledge Discovery*, 15(1):21–28.

Friedman, N., Getoor, L., Koller, D., and Pfeffer, A. (1999). Learning probabilistic relational models. In *International Joint Conference on Artificial Intelligence*, pages 1300–1309.

Haenni, R., Romeijn, J.-W., Wheeler, G., and Williamson, J. (2008). Probabilistic logic and probabilistic networks. *Forthcoming.*

Halpern, J. Y. (2003). *Reasoning about uncertainty.* MIT Press, Cambridge MA.

Ognjanović, Z. (2006). Discrete linear-time probabilistic logics: Completeness, decidability and complexity. *Journal of Logic Computation*, 16(2):257–285.

Paris, J. B. (1994). *The uncertain reasoner's companion.* Cambridge University Press, Cambridge.