



Introduction

Combining Probability and Logic

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

United Kingdom

Available online 28 January 2008

This volume arose out of an international, interdisciplinary academic network on *Probabilistic Logic and Probabilistic Networks* involving four of us (Haenni, Romeijn, Wheeler and Williamson), called *Prolognet* 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 (Prolog 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 [5], for instance. If, on the other hand, the probabilities are incorporated into the metalanguage 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., 10]. A logic that incorporates probabilities both within the language and the metalanguage is a *mixed progic*.

The central question facing a progic is which conclusions to draw from given premisses. For an internal progic the question is which ψ to conclude from given premisses $\varphi_1, \dots, \varphi_n$, 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 $\varphi_1^{X_1}, \dots, \varphi_n^{X_n}$ one would normally ask what Y to attach to a given ψ , when also given premisses $\varphi_1^{X_1}, \dots, \varphi_n^{X_n}$. Note that, depending on the semantics in question, the X_1, \dots, 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

* Corresponding author.

E-mail address: j.williamson@kent.ac.uk (J. Williamson).

proof theory to determine which ψ to conclude from given $\varphi_1, \dots, \varphi_n$. 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. [4] 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 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 other (possibly independent) causes of lung cancer, for further smoking-related diseases, or for possible symptoms 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 network. In the simplest case of so-called *Bayesian* (or *belief*) *networks* [11], it is assumed that the conditional probability of each network variable given its parents is fully known. In the example above, we could meet this requirement by specifying $P(L = \text{yes} \mid S = \text{yes}) = 0.05$ and $P(L = \text{yes} \mid S = \text{no}) = 0.01$ for the network variable L (which indicates that smoking increases the risk of lung cancer by a factor of 5) and by assuming a prior probability $P(S = \text{yes}) = 0.3$ for the network variable S (by which we express the assumption that about 30% of the population smokes). What makes Bayesian networks particularly attractive is the fact that the included structural and probabilistic information is sufficient to induce a unique *joint probability function* over all involved variables, which in turn can be used to answer all sorts of probabilistic queries. A non-trivial query in the example would be to compute the probability $P(S = \text{yes} \mid L = \text{yes})$, which can easily be derived from the joint probability function. Note that the explicit specification of a joint probability function would require exponentially many parameters, which is not feasible for practical applications. Bayesian networks are thus useful to specify large joint probability functions efficiently and to reduce the complexity of respective computations. The simplicity and efficiency of Bayesian networks explains their vast success in the last two decades in many different areas.

A second type of probabilistic network arises when the uniqueness assumption for the given probability values is relaxed. The resulting *credal networks* [1] are thus similar to Bayesian networks, except that they expect sets of probabilities instead of point-valued probabilities for each network variable. In the lung cancer example above, we could for example express our lack of knowing the percentage of smokers precisely by $0.27 \leq P(S = \text{yes}) \leq 0.32$. In the simplest case of binary variables with two possible outcomes, the required sets of probabilities are always intervals, like $[0.27, 0.32]$, but in general it is assumed that they are closed convex sets of probabilities, so-called *credal sets*. The advantage of credal networks over Bayesian networks is their increased generality and flexibility, but the price for this is added computational complexity. Credal networks have been around for about 10 years now, but a remarkable increase of attention has only been observed in the last few years.

Another type of probabilistic network results from relaxing the restriction to directed graphs. One class of such *undirected* probabilistic networks is known as *Markov networks* (or *Markov random fields*) [8,12]. They are similar to Bayesian networks in their representation of dependencies, but they can represent certain dependencies that a Bayesian network cannot (such as cyclic dependencies), and they can't represent certain dependencies that a Bayesian network can. The probabilistic information of a Markov network comes in form of so-called *potential functions* ϕ_k , where k is a clique (set of pairwise connected nodes) of the undirected graph. The collection of potential functions represents a factorization of the joint probability function of all involved variables and can thus be used to answer arbitrary probabilistic queries. Both Bayesian networks and Markov networks are special cases of *valuation systems* [7] (or *factor graphs* [6]), a more general type of (probabilistic or non-probabilistic) graphical model.

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 attach probabilities 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 [9], 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 Ng 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 de Campos 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 [2]. Second, the research often focuses 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 [3]. The emphasis on single-model progics is particularly strong when learning from data is considered.

4. This volume

In the first paper of this volume, *Assembling a consistent set of sentences in relational probabilistic logic with stochastic independence*, Cassio Polpo de Campos, Fabio G. Cozman, and José E. Ochoa Luna examine network-based representations for independence relations in probabilistic logics. The paper starts with a comprehensive survey of the literature on probabilistic logic and thus provides an ideal entry point for the topic of this volume. Next, it discusses independence relations in probabilistic logics with a focus on networks for relational probabilistic logic, and then it explores a possible application. The paper ends by pointing out some methodological care that should be taken when assembling knowledge bases.

In the second paper, on *Probabilistic argumentation*, Rolf Haenni assumes that the available knowledge is partly encoded as a set of logical premisses and partly as a fully specified probability space. This setting gets particularly interesting when some of the logical premisses include variables that are not part of the probability space. The two classical questions of the probability and the logical deducibility of a hypothesis can then be replaced by the more general question of the probability of a hypothesis being logically deducible from the premisses. The theory is thus an example of an external progic, which does not explicitly deal with sets of probabilities, but still allows a hypothesis to be judged by a pair of probability values.

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 Progenicnet project. As Howson says, a positive answer to this question is of course possible, but could turn out to be 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 consequence—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 σ -algebra

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 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 \mathcal{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.

The paper by Bert Leuridan, *Causal discovery and the problem of ignorance*, deals with the use of logic in determining the structure of a Bayesian network on the basis of a probability assignment. It discusses the problem that Pearl’s IC algorithm, which determines a causal graph on the basis of a probability assignment over variables, requires us to have full knowledge of the probability assignment, in particular of all the conditional (in)dependence relations among variables. Clearly, in many applications we do not have a full frequency table over all variables available, and hence we cannot take a unique probability assignment over all the variables as starting point. The paper proposes a solution to this problem: it defines a so-called adaptive logic, which allows us to work with default settings for variables whose relation is unknown, and indeed to revise the default settings once more information is obtained. The paper provides a proof theory, a semantics, and finally a soundness proof for this logic.

In the next paper, *Framing human inference by coherence based probability logic*, Niki Pfeifer and Gernot Kleiter establish a link from inference in probabilistic logic to psychological aspects of human deductive reasoning. The approach they adopt is to take probability logic based on the coherence approach of subjective probability as the basic reference theory. The paper first gives a brief overview of the recent developments of combining logic and probability to build models of human deductive reasoning. It then proposes the coherence approach and highlights its advantages for psychological model building. The main claims of the paper are supported by results of experimental studies, which seem to be most consistent with coherence based probability logic.

The paper by Kee Siong Ng 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. Ng and Lloyd indeed present a language that does so, by allowing functions to return whole densities over domains of types. Ng and Lloyd comment extensively on expressivity and computational complexity of higher-order logics, 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.

Finally, the paper by Jan Sprenger, *Statistics between inductive logic and empirical science*, is concerned with the logical status of statistical inference. More specifically, it poses the question of whether we can isolate a distinctly logical part to statistical inference that is universal to all statistical inferences, as separate from the rather more application-dependent choices of a model and, in the case of Bayesian inference, certain prior probabilities. By means of two examples, one on parameter estimation and one on model selection, Sprenger argues that statistical practice provides no grounds for any such separation between a unifying logical and a diverse application-oriented part of statistical inference. Rather the image emerges of statistics as closely tied up with the empirical sciences, requiring case-by-case optimisation instead of blanket solutions. Inductive logic may be interesting in its own right; portraying statistical inference as inductive logic misses out on the juiciest parts of the latter.

5. Future directions

Prolog 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 Section 1, 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 logics and external logics. 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 this representation, and why it does so. 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 of 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

- [1] F.G. Cozman, Credal networks, *Artificial Intelligence* 120 (2) (2000) 199–233.
- [2] P. Domingos, Toward knowledge-rich data mining, *Data Mining and Knowledge Discovery* 15 (1) (2007) 21–28.
- [3] N. Friedman, L. Getoor, D. Koller, A. Pfeffer, Learning probabilistic relational models, in: *International Joint Conference on Artificial Intelligence*, 1999, pp. 1300–1309.
- [4] R. Haenni, J.-W. Romeijn, G. Wheeler, J. Williamson, Probabilistic logic and probabilistic networks, 2008, in press.
- [5] J.Y. Halpern, *Reasoning about Uncertainty*, MIT Press, Cambridge MA, 2003.
- [6] F.R. Kschischang, B.J. Frey, H.A. Loeliger, Factor graphs and the sum–product algorithm, *IEEE Transactions on Information Theory* 47 (2) (2001) 498–519.
- [7] J. Kohlas, *Information Algebras: Generic Structures for Inference*, Springer, London, 2003.
- [8] S.L. Lauritzen, *Graphical Models*, Oxford University Press, 1996.
- [9] Z. Ognjanović, Discrete linear-time probabilistic logics: Completeness, decidability and complexity, *Journal of Logic and Computation* 16 (2) (2006) 257–285.
- [10] J.B. Paris, *The Uncertain Reasoner's Companion*, Cambridge University Press, Cambridge, 1994.
- [11] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Mateo, CA, 1988.
- [12] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*, John Wiley and Sons, 1990.