Benedikt **Löwe**, Eric **Pacuit**, Jan-Willem **Romeijn** (*eds.*)
Foundations of the Formal Sciences VI
Reasoning about Probabilities and Probabilistic Reasoning

Logical Relations in a Statistical Problem

Rolf Haenni[1], Jan-Willem Romeijn[2],
Gregory Wheeler[3], Jon Williamson[4]

[1] Institut für Informatik und angewandte Mathematik, Universität Bern, Neubrückstrasse 10, 3012 Bern, Switzerland

[2] Faculteit Wijsbegeerte, Rijksuniversiteit Groningen, Oude Boteringestraat 52, 9712 GL Groningen, The Netherlands

[3] Centre for Research in Artificial Intelligence, Departamento de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516 Caparica, Portugal

[4] Philosophy Section, School of European Culture & Languages, The University of Kent, Canterbury, Kent, CT2 7NZ, United Kingdom

E-mail: `haenni@iam.unibe.ch, J.W.Romeijn@rug.nl, grw@fct.unl.pt, j.williamson@kent.ac.uk`

1 Introduction

While in principle probabilistic logics might be applied to solve a range of problems, in current practice they are rarely applied. This is perhaps because they seem disparate, complicated, and computationally intractable. In fact, as we shall illustrate in this paper, several approaches to probabilistic logic fit into a simple unifying framework. Furthermore, there is the potential to develop computationally feasible methods to mesh with this framework. A unified framework for dealing with logical relations may contribute to probabilistic methods in machine learning and statistics, much in the way that the notion of causality and its relation to Bayesian networks have contributed to advances in these fields.

The unifying framework is developed in detail in [6]. Here we shall very briefly describe the gist of the whole approach.

1.1 Probabilistic Logic

Probabilistic logic asks what probability (or set of probabilities) should attach to a conclusion sentence ψ, given premises which assert that certain

Received by the editors: 8 October 2007; 1 May 2008.
Accepted for publication: 5 May 2008.

probabilities (or sets of probabilities) attach to various sentences $\varphi_1, \ldots, \varphi_n$. That is, the fundamental question is to find a suitable set Y such that

$$\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \approx\!\!\!\models \psi^Y, \tag{1.1}$$

where $\approx\!\!\!\models$ is a notion of entailment, $X_1, \ldots, X_n, Y$ are sets of probabilities and $\varphi_1, \ldots, \varphi_n, \psi$ are sentences of some logical language $\mathcal{L}$. This is a *schematic* representation of probabilistic logic, inasmuch as the entailment relation $\approx\!\!\!\models$ and the logical language $\mathcal{L}$ are left entirely open.

1.2 The Progicnet Programme

What we call the *progicnet programme* consists of two basic claims:

Framework. A unifying framework for probabilistic logic can be constructed around Schema 1.1;

Calculus. Probabilistic networks can provide a calculus for probabilistic logic—in particular they can be used to find a suitable Y such that the entailment relation of Schema (1.1) holds.

These two claims offer a means of unifying various approaches to combining probability and logic in a way that seems promising for practical applications. We shall now take a look at these two claims in more detail.

1.2.1 Framework

The first claim is that a unifying framework for probabilistic logic can be constructed around Schema (1.1). This claim rests on the observation that several seemingly disparate approaches to inference under uncertainty can in fact be construed as providing semantics for Schema (1.1):

Standard Probabilistic Semantics. According to the standard semantics, the entailment $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \approx\!\!\!\models \psi^Y$ holds if all probability functions P which satisfy the premisses—i.e., for which $\mathrm{P}(\varphi_1) \in X_1$, …, $\mathrm{P}(\varphi_n) \in X_n$—also satisfy the conclusion $\mathrm{P}(\psi) \in Y$. The logical language may be a propositional or predicate language.

Bayesian Statistical Inference. Under this account, the probabilistic premisses contain information about prior probabilities and likelihoods which constitute a statistical model, the conclusion denotes posterior probabilities, and the entailment holds if, for every probability function subsumed by the statistical model of the premisses, the conclusion follows by Bayes's theorem. Again a propositional or predicate language may be used.

Evidential Probability. Here the language is a predicate language that can represent statistical statements of the form 'the frequency of S in reference class R is between l and u'. The φ_i capture the available evidence, which may include statistical statements. These evidential statements are uncertain and the X_i characterise their associated risk levels. The entailment holds if the conclusion follows from the premisses by the axioms of probability and certain rules for manipulating statistical statements.

Probabilistic Argumentation. Here the language is propositional and the entailment holds if Y contains the proportion of worlds for which the left-hand side forces ψ to be true.

Objective Bayesian Epistemology. This approach deals with a propositional or predicate language. The $\varphi_i^{X_i}$ are interpreted as evidential statements about empirical probability, and the entailment holds if the most *non-committal* (i.e., maximum entropy) probability function, from all those that that satisfy the premisses, satisfies the conclusion.

With the exception of the first, these different semantics for probabilistic logic are presented more fully in the subsequent sections of this paper.

1.2.2 Calculus

In order to answer the fundamental question that a probabilistic logic faces—i.e., in order to find a suitable Y—some computational machinery needs to be invoked. Rather than appealing to a proof theory as is usual in logic, the progicnet programme appeals to *probabilistic networks*. This is because determining Y is essentially a question of probabilistic inference, and probabilistic networks can offer a computationally tractable way of inferring probabilities. It turns out that under the different approaches to probabilistic inference outlined above, it is often the case that $X_1, \ldots, X_n, Y$ are single probabilities or intervals of probability. When that is the case, a *Bayesian network* (a tool for drawing inferences from a single probability function) or a *credal network* (which draws inferences from a closed convex set of probability functions) can be used to determine Y. The construction of the probabilistic network depends on the chosen semantics, but given the network the determination of Y is independent of semantics. Hence the progicnet programme includes a common set of tools for calculating Y [6]. Examples of the use of probabilistic networks will appear in the following sections; here we shall introduce the key features of probabilistic networks and their role in the progicnet programme.

A probabilistic network is based around a set of variables $\{A_1, \ldots, A_r\}$. In the context of probabilistic logic, these may be propositional variables, taking two possible values True or False; if the language $\mathcal{L}$ of the logic is a predicate language, the propositional variables may represent atomic propositions, i.e., propositions of the form Ut where U is a relation symbol and t is a tuple of constant symbols. A probabilistic network contains a directed acyclic graph whose nodes are $A_1, \ldots, A_r$. This graph is assumed to satisfy the *Markov condition*: each variable is probabilistically independent of its non-descendants, conditional on its parents in the graph. For instance, the following directed acyclic graph implies that A_3 is independent of A_1 conditional on A_2:

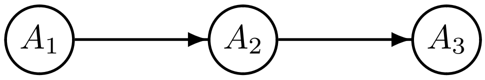

FIGURE 1. Example of a probabilistic network.

A probabilistic network also contains information about the probability distribution of each variable conditional on its parents in the graph. In a Bayesian network, these conditional probabilities are all fully specified; a Bayesian network then determines a joint probability distribution over $A_1, \ldots, A_r$ via the relation $\mathrm{P}(A_1, \ldots, A_r) = \prod_{i=1}^{r} \mathrm{P}(A_i|\mathrm{Par}_i)$ where Par_i is the set of parents of A_i. In our example, we might have

$$\mathrm{P}(A) = 0.7, \qquad \mathrm{P}(B|A) = 0.2, \qquad \mathrm{P}(C|B) = 0.9,$$
$$\mathrm{P}(B|\neg A) = 0.1, \qquad \mathrm{P}(C|\neg B) = 0.4,$$

from which we derive, for example,

$$\mathrm{P}(A \wedge \neg B \wedge C) = \mathrm{P}(A)\mathrm{P}(\neg B|A)\mathrm{P}(C|\neg B) = 0.224.$$

In a credal network, the conditional probabilities are only constrained to lie within closed intervals. A credal network then determines a set of joint probability distributions: the set of those distributions determined by Bayesian nets that satisfy the constraints. For example, a credal network might by satisfied by the above graph together with the following constraints:

$$\mathrm{P}(A) \in [0.7, 0.8], \qquad \mathrm{P}(B|A) = 0.2, \qquad \mathrm{P}(C|B) \in [0.9, 1],$$
$$\mathrm{P}(B|\neg A) \in [0.1, 1], \qquad \mathrm{P}(C|\neg B) \in [0.4, 0.45].$$

In the context of probabilistic logic, we are given premisses $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n}$, and a conclusion sentence ψ, and we need to determine an appropriate Y to attach to ψ. The idea is to build a probabilistic network that represents the set of probability functions satisfying the premisses, and use this network to calculate the range of probabilities that these functions give ψ. As mentioned above, the construction of the probabilistic network will depend on the chosen semantics, but common inference machinery may be used to calculate Y from this network. The approach taken in [6, §8.2] is to implement this common machinery as follows. First, *compile* this network: i.e., transform it into a different kind of network which is guaranteed to generate inferences in an efficient way. Second, use numerical *hill-climbing* methods in this compiled network to generate an approximation to Y.

In this paper we will illustrate the general approach of the progicnet programme by means of an example in which a number of applications can be exhibited. The example stems from psychology, more specifically from psychometrics, which studies the measurement of psychological attributes by means of tests and statistical procedures performed on test statistics. This example is constructed with the aim of bringing out the use of logical relations in probabilistic inference. In the next section we shall introduce the psychometric case study. In subsequent sections we shall see how the inferential procedures introduced above can be applied to this problem domain, and how they fit into a single framework within which the progicnet calculus can be utilized.

2 Applying the Progicnet Framework

We now illustrate the progicnet programme with an example on the measurement of psychological attributes. The first subsection introduces the example, and the second subsection indicates how each of the approaches that is covered by the progicnet framework can be employed to solve specific problems. At times, the example may come across as somewhat contrived. If so, this is because we illustrate all procedures with a single example. Straightforward applications of the framework and calculus will typically involve two procedures only.

2.1 A Psychometric Case Study

Psychometrics is concerned with the measurement of psychological attributes in individuals, for example to do with cognitive abilities, emotional states, and social strategies. Typically, such attributes cannot be observed directly. What we observe are the behavioural consequences of certain psychological attributes, such as a high score in a memory test, a certain reaction to emotionally charged images, or the characteristics of social interactions in some game. In many psychometric studies, the psychological attributes are taken as the hidden causes of these observable facts about subjects, or in short, they are taken as *latent variables*. The *observable variables*, and the correlational structure among them, are used to derive facts about these latent variables.

Notice that the general aim of psychometrics fits well with the general outlook of the progicnet framework. As in the progicnet framework, most psychometric questions start out with a number of probabilistic facts, deriving directly from the observations, and a number of logical and probabilistic relations among observable and latent variables, deriving from the psychological theory. The goal is then to find further logical and probabilistic facts concerning the latent variables, which satisfy the constraints determined by the observations and the psychological theory. Hence psychometrics lends itself well to a conceptualisation in terms of the progicnet framework.

Let us make this more concrete in the context of a version of a cognitive psychological experiment, which we concede is still rather abstract. Say that we have presented a number of subjects, indexed j, with three cognitive ability tasks, A, B, and C, which they can either pass or fail. We denote the corresponding test variables by A_j, B_j, and C_j, denoting the scores of subjects j on the three tests, respectively. Each test variable can be true or false, which, in the case of A_j, is denoted by the assignments a_j^1 (or a_j) and a_j^0 (or $\neg a_j$), respectively.

Imagine further that these tests are supposed to inform us about a psychological theory concerning three aspects of cognition, two of them to do with different developmental stages of the subjects and the other with processing speed. The corresponding latent variables are denoted by F_j, G_j, and H_j, respectively. Say that the categorical variables F_j and G_j each discern two developmental stages, and are thus binary. The processing speed $H_j \in [0, \infty)$ is continuous, but for convenience we may view H_j as categorical on some suitable scale, taking integer values n for $1 \leq n \leq N$ and N sufficiently large, say $N = 100$. The atomic statements in the language are then valuations of these variables for subjects. For example, b_5^0 or $\neg b_5$ mean that subject $j = 5$ failed test B, and h_3^{15} means that subject $j = 3$ has a latent processing speed $n = 15$. For convenience we collect the variables in $V_j = \{A_j, B_j, C_j, F_j, G_j, H_j\}$.

Imagine first that the psychological theory provides the following independence relation among the variables in the theory:

$$\forall j \neq k : \quad \mathrm{P}(V_j) = \mathrm{P}(V_k). \tag{2.1}$$

This relation expresses that all subjects are on the same footing, in the sense that they are each described by the same probability function over all the variables. Because of this the order in which the subjects are sampled does not matter to the conclusions we can draw from the sample. Moreover, unless we condition on observations of specific subjects and assignments, we can omit reference to the subjects j in the probability assignments to the variables.

Second, imagine that the developmental stages F and G and processing speed H are independent components in determining the test performance, and further that test scores are determined only by these latent variables, i.e., conditional on the latent variables, the performance on the tests is uncorrelated. The exact independence structure might be:

$$P(A, B, C, F, G, H) = P(F)P(G)P(H)P(A|F, G)P(B|G, H)P(C|H). \quad (2.2)$$

Both the independence among the subjects j, and the independence relations between the variables within each subject present strong simplifications to the psychometric example.

Next to the independence premises, psychological theory might determine the following relations between assignments to the latent and the observable variables. All these relations hold for all subjects j, and thus we omit again any such reference.

$$f \wedge g \rightarrow \neg a, \quad (2.3)$$

$$\neg g \rightarrow a, \quad (2.4)$$

$$P(b|g \wedge h^n) = \frac{n}{N}, \quad (2.5)$$

$$P(c|h^n) = \frac{N+n}{2N}. \quad (2.6)$$

Again these relations may be taken as premises in the progicnet framework, because each of these relations effectively restricts the set of probability assignments over both latent and observable variables. Or in terms more familiar to statisticians, the above premises determine a model: they fix the likelihoods of the hypotheses about subjects. Note, however, the available knowledge about the outcome of test A, as expressed in Equations (2.3) and (2.4), is purely logical and in this sense qualitative. One of the challenges is to combine such purely logical constraints with the probabilistic facts given in the other premises.

2.2 Various Approaches in a Unifying Framework

As signalled at the beginning of this section, the reader may feel that the psychometric example is unnecessarily complicated. We hope it will be apparent from subsequent sections why the example is so multi-faceted. One of the strengths of the progicnet framework is that it can accommodate a large variety of inferential problems, and we have chosen the example such that all these inferential problems find a natural place.

Of course a large number of problems on the psychometric example are essentially statistical. We may want to estimate the probability that a subject will pass test C given her performance on A and B, or how probable it is that her processing speed exceeds a certain threshold. Most of these problems will be dealt with in Bayesian statistical inference, which is sketched in Section 3. There we define a probability over the latent variables, by observing a number of subjects and then adapting the probability over latent variables accordingly. Because this type of inference is particularly well-suited for the example, we will pay a fair amount of attention to it.

Of course there are also statistical inference problems to which Bayesian statistical inference is not that easily applicable. For example, we might discover that an additional factor D influences the performance on the tests A, B, and C, so that we have to revise our predictions over these performances. Alternatively, we might be given further frequency information from various experimental studies on the variables already present in the example, say

$$\mathrm{P}(g|b) \in [0.2, 0.4], \tag{2.7}$$

$$\mathrm{P}(g|c) \in [0.3, 0.5]. \tag{2.8}$$

On the addition of such information, we can employ inferences that use so-called evidential probability. It tells us how to employ the discovery of the factor D in improving predictions, and how to adapt the predictions for G after learning the further frequency information. Section 4 introduces this approach.

Evidential probability provides solutions to a number of inferential problems on which Bayesian inference remains silent. But there are yet other problems for which both these statistical approaches are unsuited, for instance those concerned with logically complex combinations of observable and latent variables. Say that growing theoretical insight entails that

$$(a \wedge g) \vee b. \tag{2.9}$$

We might then ask what probability to attach to other complex formulae. As worked out in Section 5, probabilistic argumentation is able to provide answers on the basis of a strict distinction between logical and probabilistic knowledge, and by considering the probability of a hypothesis to be deducible from the given logical premises. However, answers to such questions will typically be intervals of probability, which makes actual computations less efficient. Here objective Bayesianism, as dealt with in Section 6, presents a technique to select a single probability assignment from all assignments that are consistent with the premises.

In the next few sections we show that inferential problems such as the above can be answered by the variety of approaches alluded to in the above, that these approaches can all be accommodated by the progicnet framework, and that their accommodation by the framework makes them amenable to the common calculus introduced in the foregoing. In this way we illustrate the use of this framework.

3 Bayesian Statistical Inference

This section introduces Bayesian statistical inference, illustrates how it is captured in the progicnet framework, and finally shows that it can be employed to solve inferential problems on the psychometric example. Bayesian statistical inference is a relatively important approach in this paper. It covers a fairly large number of the inferential problems in the example, because the example itself has a statistical nature. However, it also misses important aspects. In subsequent sections, we will show how each of the other approaches in this paper can be used to fill in these lacunas.

3.1 Simple Bayesian Inference in the Progicnet Framework

The key characteristic of Bayesian statistics is that it employs probability assignments over statistical hypotheses, next to probability assignments over data. More specifically, a Bayesian statistical inference starts by determining a model, or a *set of statistical hypotheses* that are each associated with a full probability assignment over the data, otherwise known as *likelihood functions*, and further a so-called *prior probability assignment* over the model. Relative to a model and a prior probability, the data then determine a so-called *posterior distribution* over the model, and from this posterior we can derive expectation values, predictions, credence intervals, and the like [1, 13].

We may illustrate the general idea of Bayesian inference with the psychometric example of the previous section. In the example, $\{h_j^1, \ldots, h_j^N\}$ is a model with a finite number of hypotheses concerning the latent speed of some subject j, and Equation (2.6) determines the likelihoods $\mathrm{P}(c_j^1|h_j^n) = \frac{N+n}{2N}$ of the hypotheses h_j^n for c_j^1, the event of subject j passing the test C_j. Finally, we might take a uniform distribution $\mathrm{P}(h_j^n) = \frac{1}{N}$ as prior probabilities. With Bayes's theorem it follows that

$$\mathrm{P}(h_j^n|c_j^1) = \mathrm{P}(h_j^n)\frac{\mathrm{P}(c_j^1|h_j^n)}{\mathrm{P}(c_j^1)} = \frac{2(N+n)}{N(3N+1)}. \tag{3.1}$$

That is, upon learning that subject j passed test C_j, we may adapt the probability assignment over processing speeds for that subject to the values on the right hand side. This transition from the prior $\mathrm{P}(h_j^n)$ to the posterior $\mathrm{P}(h_j^n|c_j^1)$ is at the heart of all Bayesian statistical inferences.

It may be noted that the probability of the datum $\mathrm{P}(c_j^1)$ appears in Bayes's theorem. This probability may seem hard to determine directly. However, by the law of total probability we have

$$\mathrm{P}(c_j^1) = \sum_{n=1}^{N} \mathrm{P}(h_j^n)\mathrm{P}(c_j^1|h_j^n) = \frac{3N+1}{4N}. \tag{3.2}$$

So, relative to a model, the probability of c_j^1 is easily determined. We simply need to weigh the likelihoods of the hypotheses with the prior over the hypotheses in the model.

We can represent the transition from prior and likelihoods to posterior in the progicnet framework, as it was introduced in Section 1. Recall that in Schema (1.1), all premises take the form of restrictions to a probability of a logical expression, $\varphi_i^{X_i}$. However, the likelihoods $\mathrm{P}(c_j^1|h_j^n) = \frac{N+n}{2N}$ cannot be identified directly with probability assignments to specific statements, because $c_j^1|h_j^n$ does not correspond to a specific proposition. They do represent restrictions to the probability assignments, but rather they are restrictions of a different type. Since

$$\mathrm{P}(c_j^1|h_j^n) = \frac{\mathrm{P}(c_j^1 \wedge h_j^n)}{\mathrm{P}(h_j^n)},$$

we may write out this restriction in terms of two related and direct restrictions to the probability assignments, as follows:

$$\left(c_j^1|h_j^n\right)^{\frac{N+n}{2N}} \Leftrightarrow \forall\gamma \in [0,1] : \left(h_j^n\right)^{\gamma} \text{ and } \left(c_j^1 \wedge h_j^n\right)^{\gamma\frac{N+n}{2N}}. \tag{3.3}$$

The left side of this equivalence is the likelihood in the notation of Schema (1.1), while the right side fixes the probability of two related propositions in parallel. In words, we restrict the set of probability functions over the algebra to those functions for which the ratio of the probabilities of the two propositions $c_j^1 \wedge h_j^n$ and h_j^n is $\frac{N+n}{2N}$.

With this notation in place, all expressions in Equation (3.1) are seen to be restrictions to a class of probability assignments, or models for short. More specifically, the restrictions together determine the set of models uniquely: only one probability assignment over the h_j^n's and c_j^1's satisfies the restrictions on the left hand side. But this is not to say that the complete credal set, as introduced in Section 1, is a singleton. The one probability assignment over the h_j^n's and c_j^1's can still be combined with any probability assignment over the other propositional variables.

Still restricting attention to the transition from prior to posterior for the hypotheses h_j^n and the data c_j^1, the Bayesian inference can now be represented straightforwardly in the form of Schema (1.1):

$$\forall n \in \{1,\ldots,N\} : (h_j^n)^{\frac{1}{N}}, (c_j^1|h_j^n)^{\frac{N+n}{2N}} \models (h_j^n|c_j^1)^{\frac{2(N+n)}{N(3N+1)}}. \tag{3.4}$$

Equation (3.4) is a representation of the Bayesian statistical inference, starting with a model of hypotheses h_j^n, their priors and likelihoods

$$\mathrm{P}(h_j^n) = \frac{1}{N}, \qquad \mathrm{P}(c_j^1|h_j^n) = \frac{N+n}{2N},$$

and ending with a posterior

$$\mathrm{P}(h_j^n|c_j^1) = \frac{2(N+n)}{N(3N+1)}.$$

The derivation of the posterior employs standard probability theory and concerns credal sets. It is therefore amenable to the calculus introduced in Section 1.

In sum, provided we supply the relevant premises, we can also interpret inferences within the progicnet framework as Bayesian statistical inferences. One type of premise concerns the statistical model, the other type of premise determines the prior probability assignment over the model. From these two sets of restrictions we can derive, by using the progicnet calculus, a further restriction on the posterior probability $\mathrm{P}(h_j^n|c_j^1)$.

3.2 Bayesian Inference across Subjects

The above makes explicit what Bayesian statistical inference is, and how it relates to the progicnet framework. In the remainder of this section, we will show that we can accommodate the psychometric example in its entirety in a Bayesian statistical inference. That is, we extend Bayesian inference to apply to all variables and subjects, and we include all probabilistic restrictions presented in the example. It is noteworthy that this involves additional assumptions to do with a prior over latent and observable variables. If we want to do without such assumptions, we must move to one of the other approaches for incorporating logical and probabilistic relations that this paper deals with.

Recall that the idea of statistical inference is not just that we can learn about values of variables *within subjects*, but that we can learn about them *across subjects*. For example, from observing the value of C_j for a subject j we should be able to derive something about the probability assignment over the values H_k for a different subject k. The independence expressed in Equation (2.2) determines in what way this learning across subjects can take place. It expresses that each subject has a valuation over both latent and observable variables, that is drawn from the same multinomial distribution $\mathrm{P}(V)$ with $V = \{A, B, C, F, G, H\}$. By learning valuations and expectations over these variables for some subjects, we therefore also learn the expectations over variables for other, as yet unobserved subjects. Moreover, the valuations of the variables are not drawn from just any multinomial distribution over the variables. Because we only have access to the observable variables, the latter would mean we could never learn anything about the latent variables. Fortunately the psychometric example offers a number of relations among latent and observable variables, and these relations restrict the set of multinomial distributions from which the valuations of the observable variables are drawn.

To make this specific, consider again the relation between the observable variables C_j and the latent variables H_j. To keep things manageable we choose $N = 3$, so that we have $3 \times 2 = 6$ complete valuations of C_j

and H_j together. Without further restrictions, we thus have a multinomial distribution determined by 6 parameters, namely probabilities for each full valuation $\mathrm{P}(c_j^i \wedge h_j^n) = \theta_k$ with $k = ni + n$, and a restriction that these probabilities add up to 1, leading to 5 degrees of freedom. We can also parameterise this distribution differently, with a probability $\mathrm{P}(h_j^n) = \theta_{h^n}$ for the latent variables h^n, a restriction that these sum to 1, and next to that three conditional probabilities $\mathrm{P}(c_j^1|h_j^n) = 1 - \mathrm{P}(c_j^0|h_j^n) = \theta_{C^n}$. In either case we have a set of multinomial distributions from which valuations of observed and latent variables may be drawn.

As suggested in the foregoing, we have some additional restrictions to this set of distributions deriving from the likelihoods of H_j for C_j: $\mathrm{P}(c_j^1|h_j^n) = \frac{N+n}{2N}$. In the latter parametrisation of the multinomial distributions, these restrictions can be accommodated very easily, because they come down to setting parameters θ_{C^n} to specific values, namely

$$\theta_{C^n} = \frac{N+n}{2N}. \tag{3.5}$$

Once the restrictions given by the likelihoods $\mathrm{P}(c_j^1|h_j^n)$ are put in place, all remaining degrees of freedom in the parameter space derive from the freedom in the probability over the hypotheses $\mathrm{P}(h_j^n)$. Every point in the parameter space $\theta_h = \langle \theta_{h^1}, \theta_{h^2}, \theta_{h^3} \rangle$ is associated with a particular value for the probability of the observable variable C_j, according to

$$P_{\theta_h}(c_j^1) = \sum_{n=1}^{3} \mathrm{P}(c_j^1|h_j^n)\mathrm{P}(h_j^n) = \sum_{n=1}^{3} \frac{N+n}{2N}\theta_{h^n}. \tag{3.6}$$

Note that these values need not be unique: it may happen, and indeed it does happen in the example, that several probability assignments over the h_j^n, or points θ_h in the parameter space, lead to the same overall probability for c_j^1. Hence observing the relative frequency of values for the variables C_j may not lead to a unique probability over the hypotheses $\mathrm{P}(h_j^n)$. In any case, the main insight is that learning the relative frequency of values for the variables C_j does tell us something about the probabilities of h_k^n for some as yet unobserved subject k.

3.3 Setting up the Statistical Model

The foregoing concludes the introduction into Bayesian statistical inference for the psychometric example. We will now fill in the details of this approach. The aim is to specify a Bayesian inference for F, G and H from the observation of A, B, and C and the relations (2.3) to (2.6), along the lines just sketched for H and C. Readers who are more interested in the complementary tools provided by the other approaches can skip the present subsection.

As indicated in Section 1, to make actual inferences in the psychometric example it is convenient to build up a so-called *credal network*, a graphical representation of the probability assignment over all the variables, and to build up the parametrisation of the multinomial distribution, from which observations are drawn, on the basis of this network. By the independence relation of Equation (2.2) we have the following network:

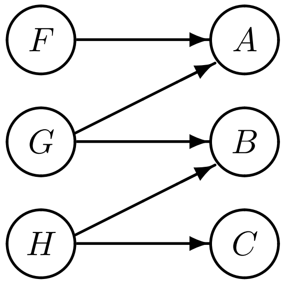

FIGURE 2. The network for the psychometric case study.

This network captures the independence relations for each subject j separately. It expresses exactly the independencies brought out by Equation (2.2): conditional on certain latent variables certain test variables are independent of each other, and the three latent variables are independent of each other as well.

Now that we have pinned down this overall structure of the model, we can fill in some of the details by means of the relations between latent and observable variables. More specifically, from Equation (2.4) we can derive that

$$g_j^0 \wedge a_j^0$$

is false, so that we have $\mathrm{P}(a_j^0|g_j^0) = 0$ and hence

$$\mathrm{P}(a_j^0|g_j^0 \wedge f_j^i) = 0$$

for $i = 0, 1$. Similarly, from Equation (2.3) we can derive that $f_j^1 \wedge g_j^1 \wedge a_j^1$ is false, so that we have

$$\mathrm{P}(a_j^1|g_j^1 \wedge f_j^1) = 0.$$

Equations (2.5) and (2.6) provide input to the Bayesian inference even more straightforwardly: they fix the values for $\mathrm{P}(b_j^1|g_j^1 \wedge h_j^n)$ and $\mathrm{P}(c_j^1|h_j^n)$ respectively. The nice thing about the above network representation is that its parametrisation, in terms of probabilities for latent variables and probabilities of observable variables conditional on these latent variables, allows us to include these restrictions directly. All the relations between latent and observable variables restrict the space of multinomial probability distributions, by setting one or more of its parameters to specific values.

After all these relations have been incorporated, we have narrowed down the set of multinomial distributions to a specific set, which we may denote $\mathbb{P}$. Within this specific set, we have the following degrees of freedom left:

$$\mathrm{P}(a_j^1|f_j^0 \wedge g_j^1) = \theta_{A^1|F^0G^1}, \tag{3.7}$$

$$\mathrm{P}(b_j^1|g_j^0 \wedge h_j^n) = \theta_{B^1|G^0H^n}, \tag{3.8}$$

$$\mathrm{P}(f_j^1) = \theta_{F^1}, \tag{3.9}$$

$$\mathrm{P}(g_j^1) = \theta_{G^1}, \tag{3.10}$$

$$\mathrm{P}(h_j^n) = \theta_{h^n}. \tag{3.11}$$

So for $N = 3$ we have 7 degrees of freedom left in the space of multinomial distributions. Note that the uncertainty of the likelihoods, Equations (3.7) and (3.8), is quite different from the uncertainty over the latent variables, Equations (3.9) to (3.11). The former uncertainty concerns the evidential bearing that the observable variables have on the latent variables, while the latter uncertainties concern the latent variables themselves.

For each point within the above space of multinomial distributions, we can derive likelihoods for the observable variables A and B, analogously to Equation (3.6) for C:

$$\mathrm{P}(a_j^1) = (1 - \theta_{F^1})\,\theta_{G^1}\,\theta_{A^1|F^0G^1}, \tag{3.12}$$

$$\mathrm{P}(b_j^1) = \sum_{n=1}^{3} \theta_{h^n} \left(\theta_{G^1}\,\frac{n}{N} \,+\, (1 - \theta_{G^1})\,\theta_{B^1|G^0H^n}\right). \tag{3.13}$$

Because Equations (2.3) to (2.6) do not pin down all evidential relations, the likelihoods for A_j and B_j will also depend on the values of $\theta_{A^1|F^0G^1}$ and $\theta_{B^1|G^0H^n}$. One possible reaction to this is that we stipulate specific values for the latter parameters, for instance by the maximum entropy principle. This approach is developed further in Section 6.

The fully Bayesian reaction, however, is to include the unknown likelihoods in the space of multinomial distributions, and to work with a second-order probability assignment over the entire space, which includes parameters pertaining to the probability of latent variables, and parameters pertaining to observable variables conditional on latent variables. We then assign a prior probability assignment to each point in the space of multinomial distributions. And once we have provided a prior probability over all parameters, we can integrate the parameters $\theta_{A^1|F^0G^1}$ and $\theta_{B^1|G^0H^n}$ out, and come up with a marginal likelihood for A_j and B_j of all probability assignments over latent variables.

3.4 Bayesian Inference and Beyond

With these last specifications, we are ready to apply the machinery of Bayesian statistical inference. We have a model, namely the space of multinomial distributions over observable and latent variables, suitably restricted by Equations (2.1) to (2.6). And we have a prior probability over this model. So from a sample of subjects with their scores on the observable variables, we can derive a posterior probability distribution over the possible multinomial distributions, which entails expectations for the latent variables and test scores of as yet unobserved subjects. This completes the exposition of a Bayesian statistical inference for the psychometric example.

But can we accommodate this full Bayesian inference in the progicnet framework? Recall that this framework only takes finite numbers of probability assignments as input. However, the space of multinomial distributions used in the foregoing comprises of a continuum of statistical hypotheses. Fortunately, this can be solved by making the θ-parameters of the above vary discretely, exactly like we made the hypotheses H_j on processing speed vary discretely in order to fit it into the progicnet framework. With this discretisation of the probability space, we can indeed accommodate the advanced version of Bayesian statistical inference in the progicnet framework, and use the common calculus to the inference problems.

There are, however, shortcomings of the Bayesian approach that invite us to supplement it with other approaches. It depends on the details of the relations between latent and observable variables whether inferences such as the above can guide us to a unique probability assignment over latent variables. As repeatedly indicated in the foregoing, different points in the space of multinomial distributions may have the same marginal likelihoods for the observable variables, and in such cases the statistical model is simply not identified. For example, setting aside the extreme cases, there will always be several probability assignments over the latent variables h_j^n that have maximal likelihood for the observed relative frequency of c_j^1. Unfortunately, this paper is too short to include a discussion of the exact conditions under which this occurs. But we are sure that if it does occur, the results of the statistical analysis crucially depend on the prior probability assignment over the model, and in a way that cannot be resolved by collecting more data.

Shortcomings of this kind call for different approaches to the problem presented by the psychometric example. To improve on the estimations we might, for example, try and employ statistical knowledge on test and latent variables for slightly different classes of subjects. In the next section we will show how evidential probability enables us to employ such knowledge, and furthermore how this approach is covered by the progicnet framework. Alternatively, we might try and avoid the use of priors over the model altogether and simply work with the set of probability assignments determined by the input. This is the approach of probabilistic argumentation, which is dealt with in Section 5. Finally, we may also take the preferred element in the set of allowed distributions under some preference ordering of probability distributions. This objective Bayesian approach, finally, is dealt with in Section 6.

4 Evidential Probability

The first of the above suggestions is nicely accommodated by evidential probability (EP). We will first briefly review EP and then illustrate it in the context of the psychometric example.

4.1 Introduction into EP

The theory of evidential probability rests on two central ideas [10, 12, 7]: probability assessments should be based upon relative frequencies, to the extent that we know them, and the assignment of probability to specific events should be determined by everything that is known about that event.

The crux of the difference between evidential probability and Bayesian statistical inference is how approximate joint statistical distributions are handled. Bayesian statistical methods assume that there are always joint distributions available for use, whereas evidential probability does not. Instead, EP maintains that there must be empirical grounds for assigning a joint frequency probability and that we must accept the uncertainty that attends our incomplete knowledge of statistical regularities. There are of course many inference problems where the two approaches perfectly align: both theories agree that Bayes's theorem is a theorem. But the two accounts differ sharply in their assessment of the range of reasonable applications of Bayesian inference structures, and whether the alternative evidential probability methods are appropriate. See Seidenfeld [14] and Kyburg [11] for a succinct comparison.

Evidential probability is conditional in the sense that the probability of a sentence ψ is relative to a finite set of sentences Γ_δ, which represent background knowledge. The evidential probability of $\psi(j)$ given Γ_δ, written as $\%j(\psi(j), \Gamma_\delta)$,[1] is an interval, $[l, u]$, in view of our limited knowledge of relative frequencies. $\text{Prob}(\psi(t), \Gamma_\delta) = [l, u]$ expresses that the evidential probability that individual j is a ψ given the relevant statistical information

[1]Syntactically, '$\tau(\vec{x}), \rho(\vec{x}), [l, u]$' is an open formula schema, where '$\tau(\cdot)$' and '$\rho(\cdot)$' are replaced by open first-order formulas, '$\vec{x}$' is replaced by a sequence of propositional variables, and '$[l, u]$' is replaced by a specific sub-interval of $[0, 1]$. The binding operator '%' is similar to the ordinary binding operators ($\forall, \exists$) of first-order logic, except that '%' is a 3-place binding operator over the propositional variables appearing within the *target formula* $\tau(\vec{x})$ and the *reference formula* $\rho(\vec{x})$, and binds those formulas to an interval.

Hereafter we relax notation and simply use an arbitrary variable 'x' for '$\vec{x}$'.

in Γ_δ is $[l, u]$, where relevant information in Γ_δ *includes*

- the relative frequency information that the proportion of a reference set R that is also $\psi(j)$ is between l and u percent, and
- the information that the individual j is a member of R,

but *excludes*

- the relative frequency information of rival reference sets R^* to which j belongs that are no stronger than R, and
- all other frequency information about ψ *except* those from sets R' that j belongs to that are larger than R, i.e., $R \subset R'$.

There may well be several classes that satisfy these conditions with respect to $\psi(j)$, each with conflicting statistics to associate to j, but there is nevertheless a unique evidential probability assigned to $\psi(j)$ given Γ_δ: it is the smallest cover of the intervals associated with the set of undominated reference formulas.

There are two types of inference in EP, corresponding to *direct inference* and *indirect inference*. First, *direct inference*, the inference from known frequencies of ψ in a population that are R to a member t of that population, is effected in EP by each canonical statement. The statement $\text{Prob}(\psi(j), \Gamma_\delta) = [l, u]$ is an instance of direct inference. It is straightforward to accommodate this inference in the progicnet framework, because it essentially relies on a fixed set of probability assignments. The other type is *indirect inference*, the inference from an interval valued probability that an individual j is ψ to an interval valued probability assignment of ψ in a population R. It is effected in EP by its rules for adjudicating between strength and conflict among potential reference classes.

EP is much less easily accommodated in the progicnet framework than other semantics we consider, because EP employs probability distributions that are defined over different populations and the semantics for the entailment relation are determined primarily by rules for resolving conflict among relevant reference statistics. However, as is further worked out in the progicnet programme, the error probabilities that are associated with this type of inference can still be treated within in the progicnet framework.

4.2 Illustration in the Psychometric Example

Since all probability assessments in EP are based upon observed relative frequencies, the probabilistic components of our psychological theory—relations (2.5) and (2.6)—do not have direct expression within EP: there is no place for a 'latent' random variable within the theory. Nevertheless, the sentences representing the psychological theory within EP may include the bi-conditionals

$$f_j \leftrightarrow \rho$$
$$g_j \leftrightarrow \rho'$$
$$h_j \leftrightarrow \rho''$$

for all j, where each ρ^i is an open reference formula occurring in some or another closed direct inference statements in Γ_δ that effect the constraints described by (2.5) and (2.6). There may be several statistical statements in Γ_δ in which each open reference formula appears, of course. We are simply specifying the potential statistics for our inference problem, and pointing out that the list of potential statistics are determined by knowledge in Γ_δ.

Suppose that we have a particular subject, $j = 5$. We said at the outset that EP uses two sources of knowledge for assigning probabilities that concern subject 5: it draws upon knowledge of relevant statistical regularities known to affect subject 5, and it draws upon everything that is known about *that* individual, subject 5. We now demonstrate how each of these features is exercised in EP, and how this is represented in terms of the fundamental question of the progicnet framework.

Imagine that we have the medical files on our subjects and that what warrants accepting constraint (2.5) is that none of them have a record of adverse exposure to lead during childhood, which is taken to be a quantity greater than 10 micrograms of lead per deciliter of blood. However, news reaches us that any exposure to lead greater than 5 micrograms per deciliter is adverse, and a review of files reveals that there are subjects in the study who have had exposure above this threshold. Thus a new parameter is introduced, D, for exposure to lead.

Our theory says that adverse exposure to lead reduces the pass rates for task B of late development subjects. In other words, (2.5) is now available in leaded (d) or unleaded ($\neg d$) grades:

$$\%j(b_j, \rho'_j \wedge h^n_j \wedge d_j) = \frac{n-m}{N}, \text{ for some positive } m < n \quad (4.1)$$

$$\%j(b_j, \rho'_j \wedge h^n_j \wedge \neg d_j) = \frac{n}{N} \quad (4.2)$$

So if we know that subject 5 was a late development subject exposed to lead as a child, we would discount his expected performance category H by m in predicting his success at task B, and if we know all this about subject 5 but that he was not poisoned as a child then we would predict his success at B to be $\frac{n}{N}$.

And what if we had no pediatric records for subject 5? Here we would expect a prediction of success on B to be within the interval $[\frac{n-m}{N}, \frac{n}{N}]$, since leaded and unleaded are values of a binary variable and thus represent mutually exclusive categories. Still we do not know which state subject 5 is in, and it won't do to pick some point in between: subject 5 is either a leaded or unleaded subject. Thus, the evidential probability assigned to the direct inference b_5 given that (4.1) and (4.2) are in Γ_δ, and that *no other relevant statistics* are known, is the interval $[\frac{n-m}{N}, \frac{n}{N}]$.

Suppose now that we want to know the developmental category G that subject 5 belongs to, and that Γ_δ is fixed. We know that there are replacements for (2.7) and (2.8) in Γ_δ, of the form

$$\%j(\rho', b_j, [0.2, 0.4]), \tag{4.3}$$

$$\%j(\rho', c_j, [0.3, 0.5]) \tag{4.4}$$

respectively. Sentence (4.3) expresses that a proportion between 0.2 and 0.4 of the subjects who pass B belong to observable class ρ', which has the same truth value as category 1 of G. Sentence (4.4) expressed that between 0.3 and 0.5 of the subjects who pass C also belong to observable class ρ, which has the same truth value in our theory as category 1 of G. Suppose subject 5 has passed B and has also passed C. What is the probability that he is in category 1? Subject 5 belongs to two references sets, B and C, that yield conflicting probabilities regarding subject 5's membership to category 1 of G. There are no reference sets to which j belongs that offer stronger frequency information, nor are there larger sets to which either B or C belong. Thus, B and C represent undominated relevant reference statistics for ρ'. Therefore, EP assigns the shortest cover to $\rho', [0.2, 0.5]$. Thus $\text{Prob}(g(j), \Gamma_\delta) = [0.2, 0.5]$.

Each of these inferences may be represented as an instance of the basic question,

$$\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \mathrel{|\!\approx} \psi^Y,$$

by substituting $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n}$ by Γ_δ on the left hand side and ψ by an ordered pair, $\langle \chi, [l, u] \rangle$, on the right hand side, which expresses that the evidential probability of formula χ is $[l, u]$. So, the inference towards $\text{Prob}(g(j), \Gamma_\delta) = [0.2, 0.5]$ would be represented as

$$\bigwedge_i \ulcorner \%x(\tau(x), \rho(x), [l', u'])_i^1 \urcorner \bigwedge_j \varphi_j^1 \mathrel{|\!\approx} \langle g(j), [0.2, 0.5] \rangle^1,$$

where the left hand side consists of the conjunction of all direct inference statements ($\ulcorner \%x(\tau(x), \rho(x), [l, u])^1 \urcorner$) and all logical knowledge about relationships between classes (φ^1), the entailment relation $\mathrel{|\!\approx}$ is non-monotonic, the right hand side asserts that the target sentence $g(j)$ is assigned $[0.2, 0.5]$. That $g(j)$ is $[0.2, 0.5]$ just means that the proportion of EP models of

$$\bigwedge_i \ulcorner \%x(\tau(x), \rho(x), [l', u'])_i^1 \urcorner \bigwedge_j \varphi_j^1$$

that also satisfy $g(j)$ is between $[0.2, 0.5]$. Since the semantics for $\approx\!\!\!\mid$ are given by the rules for resolving conflict rather than by probabilistic coherence, we assign 1 to all premises and also to $\psi = \langle g(j), [0.2, 0.5]\rangle$.

This shows that EP fits into the progicnet framework. For statistical information that is fully certain the application of the common calculus is uninteresting, since the semantics for $\approx\!\!\!\mid$ is determined by the EP rules for resolving conflicts among reference statistics. Nevertheless, we can pose a question about the robustness of an EP inference, where error probabilities are assigned to the statistical premises. This 'second-order' EP inference does utilize the calculus, and we refer to the joint progicnet paper [6] for details.

5 Probabilistic Argumentation

In the above we have concentrated on statistical questions concerning the psychometric example. Probabilistic argumentation tackles a different set of questions that we might ask about subjects and psychological attributes, concerning the logical relations between the attributes. To some extend such logical relations can be accommodated by Bayesian statistical inference, as was illustrated in Section 3. But probabilistic argumentation provides tools for dealing with logical and probabilistic relations without taking recourse to prior probability assignments.

5.1 Introduction into Probabilistic Argumentation

In the theory of probabilistic argumentation [3, 4, 5, 9], the available knowledge is partly encoded as a set of logical premises Φ and partly as a fully specified probability space $(\Omega, 2^{\Omega}, \mathrm{P})$. Variables which constitute the multivariate state space Ω are called *probabilistic*. This setting gets particularly interesting when some of the logical premises include *non-probabilistic* variables, i.e., variables that are not contained in the probability space. The two classical questions of the probability and the logical deducibility of a hypothesis ψ can then be replaced by the more general question of the probability of a hypothesis being logically deducible from the premises. In other words, we use the given logical constraints to carry the probability measure P from Ω into the state space of all variables involved.

For this, the state space Ω is divided into an area $\mathrm{Args}(\psi) = \{\omega \in \Omega : \Phi_\omega \models \psi\}$ of so-called *arguments*, whose elements are each sufficient to make the hypothesis ψ a logical consequence of the premises, and another area $\mathrm{Args}(\neg\psi) = \{\omega \in \Omega : \Phi_\omega \models \neg\psi\}$ of so-called *counter-arguments*, whose elements are each sufficient to make the complementary hypothesis $\neg\psi$ a logical consequence of the premises (by Φ_ω we denote the set of premises obtained from instantiating the probabilistic variables in Φ according to ω). Note that the premises themselves may restrict the possible states in the probability space, and thus serves as evidence to turn the given prior probability measure P into a (conditional) posterior probability measure P′.

The so-called *degree of support* of ψ is then the posterior probability of the event $\mathrm{Args}(\psi)$,

$$\mathrm{dsp}(\psi) = \mathrm{P}'(\mathrm{Args}(\psi)) = \frac{\mathrm{P}(\mathrm{Args}(\psi)) - \mathrm{P}(\mathrm{Args}(\bot))}{1 - \mathrm{P}(\mathrm{Args}(\bot))}, \tag{5.1}$$

and its dual counterpart, the so-called *degree of possibility* of ψ, is 1 minus the posterior probability of the event $\mathrm{Args}(\neg\psi)$,

$$\mathrm{dps}(\psi) = 1 - \mathrm{P}'(\mathrm{Args}(\neg\psi)) = 1 - \mathrm{dsp}(\neg\psi). \tag{5.2}$$

Intuitively, degrees of support measure the presence of evidence supporting the hypothesis, whereas degrees of possibility measure the absence of evidence refuting the hypothesis. Probabilistic argumentation is thus concerned with probabilities of a particular type of event of the form "*the hypothesis is deducible*" rather than "*the hypothesis is true*". Apart from that, they are classical additive probabilities in the sense of Kolmogorov's axioms. In principle, degrees of support and possibility can therefore be accommodated in the progicnet framework.

When it comes to quantitatively evaluate the truth of a hypothesis ψ, it is possible to interpret degrees of support and degrees of possibility as respective lower and upper bounds of an interval. The fact that such bounds are obtained without effectively dealing with probability intervals or probability sets distinguishes the theory from most other approaches to probabilistic logic. Note that the use of probability intervals or sets of probabilities is by no means excluded in the context of probabilistic argumentation. This would simply lead to respective intervals or sets of degrees of support and degrees of possibility. Indeed, in order to solve the psychometrical example from Section 2.1, it turns out that we need to introduce such intervals of support and possibility.

5.2 Illustration in the Psychometric Example

Looking at the example from Section 2.1 from the probabilistic argumentation perspective, we first observe that the probabilistic constraints (2.5) to (2.8) affect the variables B, C, G, and H only, whereas variables A and F are tied to variable G by (2.3) and (2.4) on a purely logical basis. This allows us to consider a set of premises $\Phi = \{f \wedge g \rightarrow \neg a, \neg g \rightarrow a\}$ and a restricted state space Ω which includes the variables B, C, G, and H, but not A and F. If further logical constraints are observed, for example $(a \wedge g) \vee b$ from (2.9) or any other complex formula, they can be easily incorporated by extending Φ accordingly. The multi-faceted psychometric example is thus a nice illustration of the setting on which probabilistic argumentation operates. It also underlines the large variety of inferential problems the progicnet framework accommodates.

Since the probabilistic constraints in the example do not sufficiently restrict the possible probability measures relative to Ω to a single function P, we must cope with a whole set $\mathbb{P}$ of such probability measures. Recall that we specified this set in Section 3, where we identified the space of multinomial distributions that is consistent with the relations provided in the psychometric example, Equations (3.7) to (3.11). Recall further that for Bayesian inference, even when it came to inference about a single subject, we needed to define a prior probability over the model. But probabilistic argumentation does not need any such prior. Relative to what we have already learnt about a subject, for example that she passed test A, each $P \in \mathbb{P}$ in the remaining set of probability assignments leads to respective degrees of support and possibility for a given hypothesis, for example the hypothesis that the subject passes test C.

Moreover, from the fact that all given probabilistic constraints are either point-valued or intervals, we know that the resulting sets for degrees of support and possibility will also be point-valued or intervals. Note that hypotheses involving only probabilistic variables B, C, G, or H have equal degrees of support and possibility, i.e., the two intervals will coincide in those cases, but this does not hold for hypotheses involving A or F. In general, we may interpret the numerical difference between respective degrees of support and possibility as a quantification of the amount of available evidence that is relevant to the hypothesis in question. Besides the usual interpretation of probabilities as additive degrees of belief, which is central to the Bayesian account of rational decision making, classical Bayesian inference is not designated to provide such a separate notion of *evidential strength* relative to the resulting degrees of belief.

From a computational point of view, however, the step from a fixed probability measure to a set of probability measure, as required in our example, makes the inferential procedure of probabilistic argumentation much more challenging. As suggested in Subsection 1.2, one solution would be to incorporate the given constraints over the probabilistic variables into a credal network [2], and to use that network to compute lower and upper probabilities for the events $\mathrm{Args}(\psi)$ and $\mathrm{Args}(\neg\psi)$ to finally obtain respective bounds for degrees of support and possibility. Thus, the progicnet framework neatly accommodates inferences in probabilistic argumentation that employ interval-valued degrees of support and possibility (for corresponding algorithms and technical technical details we refer to [6]).

As inference in credal networks still gets extremely costly, even for small or mid-sized networks, the solution sketched above is not always a satisfactory way out. More promising is the idea of choosing (according to some principles) the "best" probability measure among the ones in $\mathbb{P}$, and then proceed as in the default case. The next section proposes a possible strategy for this.

6 Objective Bayesianism

To some extent the previous sections have had the idea of the progicnet framework as an epistemological scheme in the background: the inferences in the psychometric example tell us what to believe on the basis of the input provided. In objective Bayesianism, this perspective is brought to the fore. To answer the questions posed at the end of Section 2.1, they are recast explicitly in terms of the strengths of one's beliefs. For example, given background knowledge, assumptions and data—such as Equations (2.1) to (2.6)—and the observed performance of a subject on tests A and B, how strongly should one believe that the subject will pass test C? By reformulating the questions this way, one can invoke the machinery of Bayesian epistemology.

6.1 Bayesian Epistemology and Objective Bayesianism

According to the Bayesian view of epistemology, the strengths of our beliefs should be representable by real numbers in the unit interval, and these numbers should satisfy the axioms of probability: an agent should believe a tautology to degree 1 and her degree of belief in a disjunction of mutually exclusive propositions should equal the sum of her degrees of beliefs in those individual propositions. Thus the strengths of the agent's beliefs should be representable by a probability function P. Moreover, an agent's degrees of belief should be compatible with her background knowledge, assumptions, data and evidence (which we shall collectively call her *epistemic background* or simply *evidence* $\mathcal{E}$). The notion of compatibility can be explicated by principles of the following kind:

1. If a proposition is in her evidence, then the agent should fully believe it.

2. The agent's degrees of belief should match her best estimates of the physical probabilities: if the agent knows that 70% of subjects who pass A and B also pass C, and she knows that the subject in question has passed A and B, but no other relevant facts, then she should believe that the subject will pass C to degree 0.7.

3. If no probability function fits the evidence using the above principles—the evidence is inconsistent—then some consistency maintenance strategy should be invoked. E.g., deem a probability function to be compatible with the evidence if it is compatible with a maximal consistent subset of the evidence.

4. If two probability functions are compatible with the evidence then so is any function that lies between them; if a sequence of probability functions are compatible with the evidence then so is the limit of that sequence.

Via principles 1 and 2 the evidence $\mathcal{E}$ imposes constraints χ on the agent's degrees of belief. The set of probability functions that satisfy these constraints will be denoted by $\mathbb{P}_\chi$. If this set is empty we may need to consider a set $\mathbb{P}'_\chi$ that is obtained by a consistency maintenance procedure (principle 3). Invoking principle 4 we consider the convex closure $[\mathbb{P}'_\chi]$ of this set of probability functions. Then $\mathbb{E}$, the set of probability functions that are compatible with the evidence $\mathcal{E}$, is just $[\mathbb{P}'_\chi]$. See [15, §5.3] for a more detailed discussion of these principles and their motivation.

Subjective Bayesian epistemology holds that an agent should set her degrees of belief according to any probability function in $\mathbb{E}$—she can subjectively choose which function to follow. Objective Bayesian epistemology, on the other hand, holds that while an agent's degrees of belief should be compatible with her evidence, her degrees of belief should equivocate on issues that are not decided by this evidence. Thus the agent's degrees of belief should be set according to a function $\mathrm{P}_\mathcal{E}$ in $\mathbb{E}$ that is maximally equivocal. Where the domain is specified by a finite set Ω of elementary outcomes, the function in $\mathbb{E}$ that is maximally equivocal is the function in $\mathbb{E}$ that is closest to function $\mathrm{P}_=$ which gives the same probability $1/|\Omega|$ to each elementary outcome. ($\mathrm{P}_=$ is called the *equivocator*.) Distance from the equivocator is measured by cross entropy $d(\mathrm{P},\mathrm{P}_=) = \sum_{\omega\in\Omega}\mathrm{P}(\omega)\log\mathrm{P}(\omega)/\mathrm{P}_=(\omega) = \sum_{\omega\in\Omega}\mathrm{P}(\omega)\log(|\Omega|\mathrm{P}(\omega))$. Distance from the equivocator is minimised when entropy $-\sum_{\omega\in\Omega}\mathrm{P}(\omega)\log\mathrm{P}(\omega)$ is maximised, and so this procedure is often called the *Maximum Entropy Principle* or *maxent* for short. On a finite domain, there will be a unique function $\mathrm{P}_\mathcal{E}$ that is closest to $\mathrm{P}_=$ in $\mathbb{E}$, so the agent has no choice about what degrees of belief to adopt—they are objectively determined by her evidence. (On an infinite domain—such as that determined by an infinite predicate language—there are cases in which degrees of belief are not objectively determined; nevertheless, $\mathrm{P}_\mathcal{E}$ tends to be very highly constrained, leaving little room for subjective choice.)

Note that this equivocation requirement yields a substantial difference between subjective and objective Bayesian epistemology. If a doctor knows nothing about a particular patient, she is perfectly entitled, on the subjective Bayesian account, to fully believe that the patient does not have particular ailment A. On the objective Bayesian account, however, the doctor should equivocate—i.e., she should believe that the patient has A to degree $\frac{1}{2}$. This equivocation constraint is motivated by considerations of risk. More extreme degrees of belief tend to be associated with riskier actions: with a full belief in $\neg A$ the doctor is likely to dismiss the patient, who may then deteriorate or perish, but with degree of belief $\frac{1}{2}$ the doctor is likely to seek further evidence. Now one should not take on more risk than the evidence demands: if the evidence forces a full belief then so be it; if not, it would be rash to adopt a full belief. Thus one should equivocate as far as evidence allows. This line of argument is developed in [17].

The objective Bayesian approach fits into the progicnet programme as follows. First, objective Bayesian epistemology provides a semantics for the probabilistic logic framework of Schema (1.1): $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \approx\!\!\!\!| \; \psi^Y$. According to this semantics, the premisses $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n}$ are construed as characterising the agent's evidence $\mathcal{E}$. Here $\varphi_i^{X_i}$ is understood as saying that the physical probability of φ_i is in X_i (perhaps as determined by appropriate frequency information). This evidence imposes constraints χ on an agent's degrees of belief, where $\chi = \{\mathrm{P}(\varphi_1) \in X_1, \ldots, \mathrm{P}(\varphi_n) \in X_n\}$. The set of probability functions compatible with this evidence is $\mathbb{E} = [\mathbb{P}'_\chi]$. An agent with this evidence should adopt degrees of belief represented by a function $\mathrm{P}_\mathcal{E}$ in $\mathbb{E}$ that is maximally equivocal. The question arises as to what value $\mathrm{P}_\mathcal{E}$ gives to ψ, and one can take $Y = \{\mathrm{P}_\mathcal{E}(\psi) : \mathrm{P}_\mathcal{E} \in \mathbb{E}$ is maximally equivocal$\}$. On a finite domain Y will be a singleton. Thus objective Bayesianism provides a natural semantics for Schema (1.1). Now according to the progicnet programme, probabilistic networks might be used to calculate Y. Indeed, as we shall now see, *objective Bayesian nets* can be used to calculate Y.

6.2 Illustration in the Psychometric Example

Returning to the psychometric case study, the objective Bayesian approach provides the following recipe. Equations (2.1) to (2.6) and the subject's performance on tests A and B constitute the evidence $\mathcal{E}$. We should then believe that the subject will pass C to degree $\mathrm{P}_\mathcal{E}(C)$, where $\mathrm{P}_\mathcal{E}$ is the maximally equivocal probability function out of all those that are compatible with $\mathcal{E}$.

In general, *objective Bayesian nets* can be used to calculate objective Bayesian probabilities [16] and [15, §§5.6–5.8]. The idea here is that the objective Bayesian probability function $\mathrm{P}_\mathcal{E}$ can be represented by a Bayesian net, now called an objective Bayesian net, and standard Bayesian network algorithms can be invoked to calculate the required probabilities, such as $\mathrm{P}_\mathcal{E}(C)$. Because this probability function is a maximum entropy probability function it will automatically satisfy certain probabilistic independencies and the graph in the Bayesian network that represents these independencies is rather straightforward to construct. Join two variables by an undirected edge if they occur in the same constraint of $\mathcal{E}$. Then separation in the resulting undirected graph implies independence in $\mathrm{P}_\mathcal{E}$: if X separates Y from Z in the graph then it is a fact that $\mathrm{P}_\mathcal{E}$ renders Y and Z probabilistically independent conditional on X. This undirected graph can easily be transformed into a directed acyclic graph that is required in a Bayesian net.

The example of Subsection 2.1 is actually a very special case. Here Equation (2.1) is a consequence of the objective Bayesian procedure: since there are no known connections between different subjects in $\mathcal{E}$, $\mathrm{P}_{\mathcal{E}}$ will render the features of different subjects probabilistically independent. In this example we also have a causal picture in the evidence, namely that depicted in Figure 2, where the latent variables F, G and H are causes of the test results. When we have a causal graph, the graph in the objective Bayesian network is just this graph [15, §5.8], and hence the factorisation of Equation (2.2) is also a consequence of the objective Bayesian procedure. The evidence can thus be viewed as the causal graph Figure 2 together with the constraints Equations (2.3) to (2.9). Since we have the graph in the objective Bayesian net, it remains to determine the conditional probability distributions, i.e., the distributions $\mathrm{P}_{\mathcal{E}}(F)$, $\mathrm{P}_{\mathcal{E}}(G)$, $\mathrm{P}_{\mathcal{E}}(H)$, $\mathrm{P}_{\mathcal{E}}(A|F,G)$, $\mathrm{P}_{\mathcal{E}}(B|G,H)$, $\mathrm{P}_{\mathcal{E}}(C|H)$. Since the causal structure is known, these distributions can be determined iteratively: first determine the distribution $\mathrm{P}_{\mathcal{E}}(F)$ that is maximally equivocal, then $\mathrm{P}_{\mathcal{E}}(G)$, and so on up to $\mathrm{P}_{\mathcal{E}}(C|H)$ [15, §5.8]. By iteratively maximising entropy we obtain:

$$\begin{array}{ccc}
\mathrm{P}_{\mathcal{E}}(f) = 1/2, & \mathrm{P}_{\mathcal{E}}(a|f,g) = 0, & \mathrm{P}_{\mathcal{E}}(b|g,h^n) = n/N, \\
\mathrm{P}_{\mathcal{E}}(g) = 1/2, & \mathrm{P}_{\mathcal{E}}(a|f,\neg g) = 1, & \mathrm{P}_{\mathcal{E}}(b|\neg g,h^n) = 0.4, \\
\mathrm{P}_{\mathcal{E}}(h^n) = 1/N, & \mathrm{P}_{\mathcal{E}}(a|\neg f,g) = 1/2, & \mathrm{P}_{\mathcal{E}}(c|h^n) = (N+n)/2N, \\
 & \mathrm{P}_{\mathcal{E}}(a|\neg f,\neg g) = 1. &
\end{array}$$

With these probability distributions and the directed acyclic graph we have a Bayesian network and can use standard Bayesian network methods to answer probabilistic questions. For example, how strongly should we believe that subject j will pass C given that she has passed tests A and B?

$$\begin{aligned}
\mathrm{P}_{\mathcal{E}}(c_j|a_j,b_j) &= \\
&= \frac{\sum_{f_j,g_j,h_j} \mathrm{P}_{\mathcal{E}}(c|h_j)\mathrm{P}_{\mathcal{E}}(b|g_j,h_j)\mathrm{P}_{\mathcal{E}}(a|f_j,g_j)\mathrm{P}_{\mathcal{E}}(f_j)\mathrm{P}_{\mathcal{E}}(g_j)\mathrm{P}_{\mathcal{E}}(h_j)}{\sum_{f_j,g_j,h_j} \mathrm{P}_{\mathcal{E}}(b|g_j,h_j)\mathrm{P}_{\mathcal{E}}(a|f_j,g_j)\mathrm{P}_{\mathcal{E}}(f_j)\mathrm{P}_{\mathcal{E}}(g_j)\mathrm{P}_{\mathcal{E}}(h_j)} \\
&= \frac{\sum_{f_j,g_j,h_j} \mathrm{P}_{\mathcal{E}}(c|h_j)\mathrm{P}_{\mathcal{E}}(b|g_j,h_j)\mathrm{P}_{\mathcal{E}}(a|f_j,g_j)}{\sum_{f_j,g_j,h_j} \mathrm{P}_{\mathcal{E}}(b|g_j,h_j)\mathrm{P}_{\mathcal{E}}(a|f_j,g_j)} \\
&= \frac{24N(3N+1)+(N+1)(5N+1)}{6N(21N+5)} = 0.61 \text{ as } N \longrightarrow \infty.
\end{aligned}$$

With the more extensive evidence of Equations (2.1) to (2.9), the procedure is just the same, though of course the conditional distributions and final answer differ from those calculated above.

From a computational point of view, the objective Bayesian approach is relatively straightforward for two reasons. First, there is only a single probability function $P_{\mathcal{E}}$ under consideration. As we have seen, other approaches deal with sets of probability functions. Second, since this function is obtained by maximising entropy, we get lots of independencies for free; these independencies permit the construction of a relatively sparse Bayesian net, which in turn permits relatively quick inferences.

Computationally feasibility is one reason for preferring the objective Bayesian approach over the Bayesian statistical methods of Section 3, but there are others. A second reason is that the whole approach is simpler under the objective Bayesian account: instead of defining (higher-order) probabilities over statistical models one only needs to define probabilities over the variables of the domain. It may be argued that the move to higher-order probabilities is only warranted when the evidence includes specific information about these higher-order probabilities. Such information is generally not available.

A third argument for preferring the objective Bayesian approach appeals to epistemological considerations. Since Bayesian statistics defines probabilities over statistical hypotheses, these probabilities must be interpreted epistemologically, in terms of degrees of belief—it makes little sense to talk of the chance or frequency of a statistical model being true. Hence the Bayesian statistical approach naturally goes hand in hand with Bayesian epistemology. Typically, Bayesian statisticians advocate a subjective Bayesian approach to epistemology—probabilities should fit the evidence but are otherwise a matter of subjective choice. As we have seen, however, there are good reasons for thinking that this is too lax: such an approach condones degrees of belief that are more extreme than the evidence warrants, and degrees of belief that are too extreme subject the believer to unjustifiable risks and so are irrational.

Hence Bayesian statistics should minimally be accompanied by a principled way to determine reasonable priors, such as is provided by objective Bayesian epistemology. While there is a growing movement of statisticians who advocate such a move, it is well recognised that objective Bayesian epistemology is much harder to implement on the uncountable domains of Bayesian statistics than the finite domain considered here. This is because there may be no natural equivocator on an uncountable domain (cf. the discussion of the wine-water paradox in [8]), unless we can provide an argument to favour a particular parameterisation of the domain.

For lack of a preferred parameterisation, we have a dilemma: Bayesian statistics needs to be accompanied by a Bayesian epistemology; if a subjective Bayesian epistemology is chosen then Bayesian statistics is flawed for normative reasons; on the other hand if an objective Bayesian epistemology is chosen then there are implementational difficulties; moreover, the move to higher-order probabilities should only be made where absolutely necessary. Such a move is not absolutely necessary in the example of this paper. It may be argued, therefore, that in the context of the case study considered here, the objective Bayesian approach outlined in this section is more appropriate than the Bayesian statistical approach of Section 3. Minimally, it will provide a valuable addition to the statistical treatment considered there.

7 Conclusion

In this paper we have sketched a number of different approaches to combining logical and probabilistic inference. We showed how each of these approaches can be used to answer questions in the context of a toy example from psychometrics, how each approach can be subsumed under a unifying framework, thereby making them amenable to a common underlying calculus. But what exactly did we gain in doing so? We give a number of reasons for saying that the formulation of framework and calculus, as part of an overarching progicnet programme, amounts to progress.

First of all, we hope to have shown that the standard statistical treatment of the psychometric example, in this case using Bayesian statistics, can be supplemented in various ways by other approaches to logical and probabilistic inference. The progicnet programme provides a way to unify these approaches systematically. More specifically, and as illustrated in the psychometric example, the progicnet framework allows us to supplement the statistical inference that is standard in the psychometric context with some powerful inference tools from logic, all subject to the same calculus. We believe that there are many cases, in the sciences and in machine learning, in which the context provides a lot of logical background knowledge. The psychometric example is one of them, but many more such examples can be found in data mining, bioinformatics, computational linguistics, and sociological modelling. In all of these fields the existing statistical techniques cannot optimally employ the logical background knowledge. The progicnet framework may provide the means to use logical and statistical background knowledge simultaneously, and in a variety of problem domains.

More specifically, let us reiterate the conclusions on the use of the different approaches, that were reached in the preceding sections.

Bayesian statistical inference allows for dealing with the standard inferential problems of the psychometric example. In this paper it serves as a backdrop against which the merits of the other approaches covered by the progicnet framework can be made precise. Note that this is not to say that Bayesian statistical inference occupies a central place in the progicnet framework more generally.

Evidential probability is particularly suited if we learn further statistical information that conflicts with the given statistical model or introduces further constraints on it. It provides us with the tools to incorporate this new information and find trade-offs, where Bayesian inference must remain silent.

Probabilistic argumentation can be employed to derive upper and lower bounds on the probability assignments on the basis of the statistical model and the logical relations between the variables in the model only, without presupposing any prior probability assignments. This is very useful for investigating the properties of the model and the probabilistic implications of logical relations.

Objective Bayesianism offers a principled technique for reducing a set of probability assignments, such as the statistical model of the example, to a single probability assignment. For complicated models with many parameters, this provides a powerful simplification, and thus efficient inferential procedures.

Other reasons for using a common framework are more internal to the philosophical debate. The field of probabilistic inference is rather disparate, and discussions over interpretation and applications frequently interfere with discussions to do with formalisation and validity. Perfectly valid inferences in one approach may appear invalid in another approach, and even while all approaches somehow employ Kolmogorov's measure theoretic notion of probability, what is being measured by probability, and consequently the treatment of probability in the approaches, varies wildly. We hope that by providing a common framework for probabilistic logic, we help to structure the discussions, and determine more clearly which disagreements are meaningful and which are not.

Finally, the existence of a common framework also proves useful on a more practical level. Now that we have described a common framework, we can apply the common calculus of credal networks to it. As indicated in Section 1, and roughly illustrated in Section 3 and 6, credal networks can play an important part in keeping inferences manageable in probabilistic logic. More generally, the application of these networks will lead to more efficient inferences within each of the approaches involved. We must admit, however, that in the confines of the present paper, we have not explained the advantages of using networks in detail. For the exact use of credal networks in the progicnet programme, we again refer the reader to the central progicnet paper [6].

References

[1] Vic Barnett. *Comparative Statistical Inference.* John Wiley, 1999.

[2] Fabio G. Cozman. Credal Networks. *Artificial Intelligence*, 120(2):199–233, 2000.

[3] Rolf Haenni. Towards a Unifying Theory of Logical and Probabilistic Reasoning. In Fabio Cozman, Robert Nau, and Teddy Seidenfeld, editors, *ISIPTA'05, 4th International Symposium on Imprecise Probabilities and Their Applications*, pages 193–202, 2005.

[4] Rolf Haenni. Probabilistic Argumentation. *Journal of Applied Logic*, 7:155–176, 2009.

[5] Rolf Haenni, Jürg Kohlas, and Norbert Lehmann. Probabilistic Argumentation Systems. In Dov M. Gabbay and Philippe Smets, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 5: Algorithms for Uncertainty and Defeasible Reasoning, pages 221–288. Kluwer Academic Publishers, 2000.

[6] Rolf Haenni, Jan-Willem Romeijn, Gregory Wheeler, and Jon Williamson. *Probabilistic logic and probabilistic networks.* Springer-Verlag, to appear.

[7] William Harper and Gregory Wheeler, editors. *Probability and Inference: Essays In Honor of Henry E. Kyburg, Jr.*, volume 2 of *Texts in Philosophy.* College Publications, 2007.

[8] John Maynard Keynes. *A treatise on probability.* Macmillan, London, 1921.

[9] Jürg Kohlas. Probabilistic Argumentation Systems: A New Way to Combine Logic With Probability. *Journal of Applied Logic*, 1(3–4):225–253, 2003.

[10] Henry E. Kyburg, Jr. *Probability and the Logic of Rational Belief.* Wesleyan University Press, 1961.

[11] Henry E. Kyburg, Jr. Bayesian Inference with Evidential Probability. In William Harper and Gregory Wheeler, editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.*, volume 2 of *Texts in Philosophy*, pages 281–296. College Publications, 2007.

[12] Henry E. Kyburg, Jr. and Choh Man Teng. *Uncertain Inference.* Cambridge University Press, 2001.

[13] James Press. *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications.* John Wiley, 2003.

[14] Teddy Seidenfeld. Forbidden Fruit: When Epistemic Probability may not take a bite of the Bayesian apple. In William Harper and Gregory Wheeler, editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.*, volume 2 of *Texts in Philosophy*, pages 267–279. College Publications, 2007.

[15] Jon Williamson. *Bayesian nets and causality: philosophical and computational foundations.* Oxford University Press, 2005.

[16] Jon Williamson. Objective Bayesian nets. In Sergei Artemov, Howard Barringer, Artur S. d'Avila Garcez, Luis C. Lamb, and John Woods, editors, *We Will Show Them! Essays in Honour of Dov Gabbay. Volume 2*, volume 2 of *Tributes*, pages 713–730. College Publications, 2005.

[17] Jon Williamson. Motivating objective Bayesianism: from empirical constraints to objective probabilities. In William L. Harper and Gregory R. Wheeler, editors, *Probability and Inference: Essays in Honour of Henry E. Kyburg Jr.*, volume 2 of *Texts in Philosophy*, pages 151–179. College Publications, 2007.