Probabilistic Logics and Probabilistic Networks

Jan-Willem Romeijn, Philosophy, Groningen
Jon Williamson, Philosophy, Kent

ESSLLI 2008

Course Page:
http://www.kent.ac.uk/secl/philosophy/jw/2006/progicnet/ESSLLI.htm

Where are we?

Yesterday: evidential probability, probabilistic argumentation, and fiducial probability.
Plan for today:

Bayesian statistical inference: defining Bayesian inference, representation and interpretation in the Progicnet scheme.

Networks for Bayesian statistical inference: statistical models as credal networks, Extending statistical inference with credal networks.

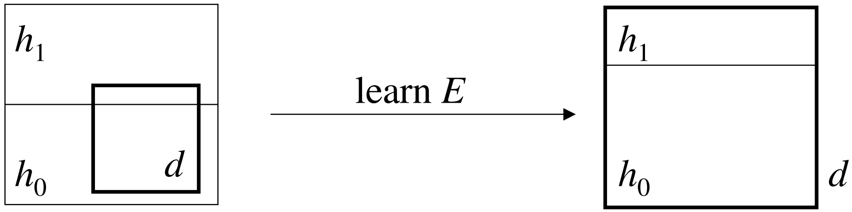
Application of Bayesian statistical inference: a psychometric case study, Bayesian inference in psychometrics.

And where we are heading tomorrow:

Objective Bayesian epistemology: choosing among the elements of the credal set by minimal information as an additional constraint on credal sets.

1 Bayesian statistical inference

Bayesian statistics is much more easily connected to the Progicnet schema than classical statistics. Its distinguishing feature is that it also employs probability assignments over statistical hypotheses.



A Bayesian statistical procedure is an indirect inference concerning these probability assignments over hypotheses.

Direct inference: Inference from a given statistical hypothesis to a probability assignments over a sample space, or particular data.

Indirect inference: Inference from particular data to a probability assignment over a set of hypotheses, collected in a statistical model.

Statistics is the first, and possibly the most important, application of Bayesian inference.

- The reach of Bayesian inference extends beyond the statistical domain, for example into philosophical and psychological modelling.
- Because it rests on a theorem of probability, Bayesian inference is very close to the inferences dealt with in the standard semantics and similar probabilistic logics.

1.1 Defining Bayesian inference

Let $\Omega_H \times \Omega_D$ be the combination of a partition of hypotheses and a sample space, and let P be a probability assignment over this space.

> **Definition 1.1 (Bayesian Statistical Inference)** *Assume $P(H_j)$, the prior probabilities assigned to a finite number of hypotheses H_j with $0 < j \leq n$, and $P(D|H_j)$, the probability assigned to the data D conditional on the hypotheses, called the likelihoods. Bayes' theorem determines that*
>
> $$P(H_j|D) = P(H_j)\frac{P(D|H_j)}{P(D)}. \quad (1)$$
>
> *Bayesian statistical inference is the derivation of the posterior $P(H_j|D)$ from the prior $P(H_j)$ and the likelihoods $P(D|H_j)$.*

Some remarks on Bayesian statistical inference:

- Both hypotheses H_j and data D refer to sets in the space $\Omega = \Omega_D \times \Omega_H$. So we associate each statistical hypothesis H_j with an entire sample space Ω_D.
- The probability of the data $P(D)$ can be hard to compute. One possibility is to use the law of total probability,

$$P(D) = \sum_j P(H_j)P(D|H_j).$$

- Often the interest is only in comparing the ratio of the posteriors of two hypotheses. By Bayes' theorem we have

$$\frac{P(H_1|D)}{P(H_0|D)} = \frac{P(H_1)P(D|H_1)}{P(H_0)P(D|H_0)},$$

  and if we assume equal priors $P(H_0) = P(H_1)$, we can use the ratio of the likelihoods of the hypotheses, the so-called Bayes factor, to compare the hypotheses.
- We distinguish between the above applications of Bayes' theorem and applications of what is often called Bayes' rule,

$$P_D(H_j) = P(H_j|D),$$

  The rule is an epistemological principle, relating two different probability functions that pertain to different epistemic states of an agent, at different points in time.

1.2 Representation in the Progicnet scheme

The derivation of the posterior $P(H_j|d)$ from the prior $P(H_j)$ and the likelihoods $P(d|H_j)$ for some particular sample d can be represented straightforwardly in the Progicnet schema. Abbreviating $P(d|H_j) = \theta_j(d)$, we write:

$$\forall j \leq n : \; H_j^{P(H_j)}, (d|H_j)^{\theta_j(d)} \models (H_j|d)^{P(H_j|d)}.$$

Infinitely many hypotheses

The representation of Bayesian statistical inference in the schema is not entirely unproblematic. An important restriction is the number of statistical hypotheses that can be considered.

- Many statistical applications do not employ a finite number of hypotheses H_j, but a continuum of hypotheses H_θ. We can write this down as

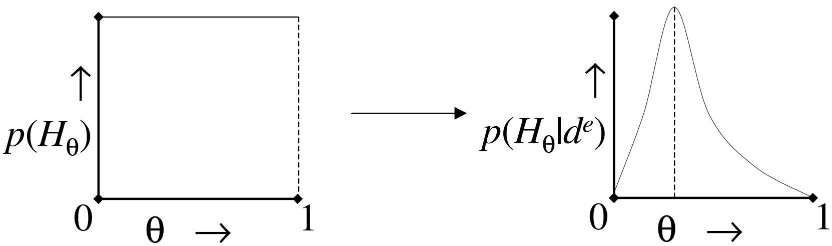
$$\forall \theta : \; H_\theta^{P(\theta)d\theta}, (d|H_\theta)^{\theta(d)} \models (H_\theta|d)^{P(\theta|d)d\theta}.$$

- We may employ a finite number n of statistical hypotheses H_j, each of which is composed of a set of hypotheses H_θ
- For example, we write $H_j = \left\{H_\theta : \theta \in \left[\frac{j-1}{n}, \frac{j}{n}\right)\right\}$. Choosing $\theta_j = \frac{2j-1}{2n}$ we can then approximate the continuous model arbitrarily close by increasing n.

De Finetti's representation

Another possible reaction to the uncountable infinity involved in using statistical hypotheses is to translate the schema with statistical hypotheses into something finite, using a result by de Finetti.

- We say that the data consist of assignments to binary propositional variables D_i, denoted $d^e = d_1^{e(1)} \cdots d_m^{e(m)}$.
- We restrict attention to statistical hypotheses that fix binomial distributions for the data, meaning that separate assignments $d_i^{e(i)}$ are associated with independent fixed chances: $P(d_{m+1}^1 | d^e, h_\theta) = \theta \in [0, 1]$.



We focus on specific predictions $P(d^{e'}|d^e)$, in which e' is a vector of length $m' > m$. These predictions can be derived from the posterior probability assignments $P(H_\theta|d^e)$ and the likelihoods $\theta(d^{e'})$ for $d^{e'}$ by the law of total probability:

$$\forall\theta : \; H_\theta^{P(H_\theta)}, \, (d^e|H_\theta)^{\theta(d^e)} \wedge (d^{e'}|H_\theta)^{\theta(d^{e'})} \models (d^{e'}|d^e)^{P(d^{e'}|d^e)}.$$

As de Finetti shows, these premises can be replaced by another set of premises, which refer to the so-called exchangeability of the probability P for samples d^e and $d^{e'}$. An exchangeable probability of the sample d^e is invariant under permutations of the assignments e. For any i,

$$\pi(\langle e(1), \ldots e(i), \ldots, e(m)\rangle) = \langle e(i), \ldots e(1), \ldots, e(m)\rangle$$

The new inference then becomes

$$\forall\pi : \; \left(d^{\pi(e')}\right)^{P(d^{e'})} \models \left(d^{e'}|d^e\right)^{P(d^{e'}|d^e)},$$

where the $\pi(e')$ refer to permutations of the elements in the vector e'. The salient point is that the restriction on the set of probability assignments induced by employing uncountably many hypotheses is also effected by this finite premises.

Carnapian inductive logic

The general idea of inductive logic is that invariances under the permutations π of this schema can be motivated by independent rationality criteria or principles of logic.

- Exchangeability, as discussed above, and symmetry among the possible outcomes.
- Relevance relations among the outcomes, as captured by various principles of instantial relevance.
- Symmetries pertaining to relations among individuals rather than their predicates.

Carnap and co-workers derived the so-called continuum of inductive methods, which assumes exchangeablility and symmetry among the outcomes,

$$P(d^1_{m+1}|d^e) = \frac{m_1 + \frac{\lambda}{2}}{m + \lambda},$$

where m_1 is the observed relative frequency of 1's and λ is a free parameter. The value of λ is then determined by the distribution $P(H_\theta)d\theta$. Paris and his group have recntly developed a number of extensions and alternatives to the original Carnapian logic.

Interval-valued priors and posteriors

Until now we have focused on sharp probability assignments. But we may also represent restrictions in terms of intervals of probability assignments. We can use this to represent a wider class of Bayesian statistical inferences.

- Walley discusses interval restrictions in the context of the above multinomial hypotheses and their associated Carnapian predictions.
- A sharp distribution over multinomial hypotheses leads to an exchangeable prediction rule. We can also allow for interval-valued assignments to statistical hypotheses.
- Such valuations can be dealt with adequately by considering a class of prior distributions over the statistical hypotheses instead of a single and sharp-valued prior distribution.
- These interval-valued assignments are useful in studying the sensitivity of statistical results to the prior probability that is chosen.

Classes of prior distributions are associated with ranges of prediction rules. As Skyrms and Festa show, we can use such ranges of rules to constitute so-called hyper-Carnapian prediction rules. The resulting prediction rules can be used for deriving predictions that incorporate relevance relations among the possible outcomes.

1.3 Interpretation

We can interpret the schema as a Bayesian statistical inference as long as the inference concerns a probability assignment over a sample space and a model, $\Omega_D \times \Omega_H$. From the restrictions laid down by the model, a prior, and the assignment d to the data variable D we can derive a posterior probability over the model.

Interpretation of probabilities
Bayesian statistics is very often associated with the so-called subjective interpretation of probability. We briefly comment on the issue of interpreting probability here.

- If we are interpreting probabilities, we should distinguish between the probability assigned to statistical hypotheses H_j and to samples or data d.
- Probability assignments over hypotheses seem most naturally understood as epistemic, meaning that they pertain to degrees of belief.
- This does not yet mean that probabilities of hypotheses are subjective, because we may also try to determine the epistemic probability assignment by means of objective criteria.
- As for the probability assigned to samples, the likelihoods, we may give it either a physical interpretation, meaning that the probabilities refer to aspects of the physical world, or an epistemic interpretation.

- On the one hand, hypotheses are statements about the world, but on the other, their likelihoods seem best interpreted as epistemic since they fulfill an evidential role.

As probabilistic logicians we are not tied to any particular interpretation of the probability functions, just like classical deductive logicians are not necessarily tied to Tarski's or some other truth definition as an interpretation of the truth values in their inferential schemes.

Bayesian confidence intervals
in the context of Bayesian statistical inference, a rather natural interpretation of interval probabilities is as credence intervals.

> **Example 1.2 (Credence intervals)** *Consider a continuum of Bernoulli hypotheses H_θ concerning binary data D, with likelihoods $\theta \in [0, 1]$ and a uniform prior. We can then define Bayesian confidence intervals, or credence intervals, for the values of θ. Each interval $\theta \in [l, u]$ is associated with a posterior probability, or credence,*
>
> $$P(h_{[l,u]}|D) = \int_l^u P(H_\theta|D)d\theta.$$

Example 1.3 (Credence intervals (continued)) *Now we may fix u and l such that*

$$\int_0^l P(H_\theta|D)d\theta = \int_u^1 P(H_\theta|D)d\theta = 2.5\%.$$

The corresponding interval-valued assignment is $\theta \in [.025, .975]$. After incorporating a specific sample d with an observed relative frequency of 0.1, we obtain a posterior density function $P(H_\theta|d)$, leading to the corresponding interval-valued assignment $\theta \in [.07, .13]$.

The question arises whether we can somehow interpret the interval-valued probabilities in the Progic schema as such credence intervals. Unfortunately, we cannot.

- The inferences of interval-valued probabilities, when interpreted as credence intervals, are elliptic, that is, they omit certain premises.
- More precisely, the fact that $\theta \in [l, u]$ does not fix the detailed shape of the prior probability density $P(H_\theta)$.
- We need this detailed shape to arrive at the specific conditional credence interval $\theta \in [l', u']$. It is not possible to rely just on the rules of inference of the schema for these credence intervals.

2 Networks for Bayesian statistical inference

We present two ways in which credal networks may be employed in Bayesian inference. The first presents a computational improvement of the inferences as such, while the second supplements the inference with further tools, deriving from the common machinery of Progicnet.

2.1 Statistical models as credal networks

We first spell out how a credal network can be related to a statistical model, i.e. a set of statistical hypotheses.

- A credal network is associated with a credal set, a set of probability functions over some designated set of variables.
- A credal set may be viewed as a statistical model: each element of the credal set is a probability function over the set of variables, and this probability may be read as a likelihood of some hypothesis for observations of valuations of the network.
- Conversely, any statistical model concerns inter-related trials of some specific set of variables, so that we can identify any statistical model with a credal network containing these variables.

Construction of the credal network

To construct the credal network that represents a statistical model, we first list all the independence relations that obtain between the variables in the sample space Ω_D.

> **Example 2.1 (Credal set as model)** *Consider subsequent observations, at times i, of three binary variables $V_i = \{A_i, B_i, C_i\}$. The independence relations I are as follows:*
>
> $$\forall i : A_i \perp\!\!\!\perp B_i | C_{i-1},$$
> $$\forall i' \neq i-1 : A_i \perp\!\!\!\perp C_{i'},\ B_i \perp\!\!\!\perp C_{i'},$$
> $$\forall i' \neq i : C_i \perp\!\!\!\perp C_{i'},\ A_i \perp\!\!\!\perp A_{i'},\ B_i \perp\!\!\!\perp B_{i'},\ A_i \perp\!\!\!\perp B_{i'},\ B_i \perp\!\!\!\perp B_{i'}.$$
>
> *These independence relations lead to the following graph:*

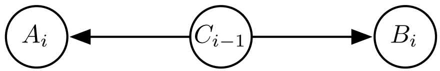


Figure 1: The graph capturing the independence relations among the propositional variables A_i, B_i, and C_i. This graph holds for each and every value of i.

To determine the credal network, we now run the following algorithm, which also allows us to include further constraints $\varphi_i^{X_i}$ on the probability assignments of the variables.

Algorithm 2.2 (Model to credal net) Construction of a credal network for the semantics of Bayesian statistical inference, based on a given model of statistical hypotheses.

Input: a model Ω_H of statistical hypotheses H_θ, a sample space Ω_D concerning a set $V = \{A_1, \ldots, A_M\}$ of propositional variables, and further premises $\varphi_1^{X_1}, \ldots, \varphi_N^{X_N}$ involving those variables.

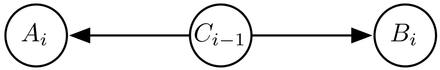
1. **Derivation of the independence relations:** using the (conditional) independence relations inherent to the likelihoods of all the H_θ.
   (a) If $P(A_i \wedge A_j \mid H_\theta) = P(A_i \mid H_\theta)P(A_j \mid H_\theta)$, add $A_i \perp\!\!\!\perp A_j$ to the set of independence constraints I.
   (b) If $P(A_i \wedge A_j \mid \bigwedge_k A_k \wedge H_\theta) = P(A_j \mid \bigwedge_k A_k \wedge H_\theta)P(A_j \mid \bigwedge_k A_k \wedge H_\theta)$, add $A_i \perp\!\!\!\perp A_j \mid \bigwedge_k A_k$ to the set of independence constraints I.
2. **Running the algorithm of the standard semantics:** based on the set I of independence relations and the premises $\varphi_1^{X_1}, \ldots, \varphi_N^{X_N}$.

Output: a graph $\mathcal{G}$ and a set of extremal points v_i in terms of the network coordinates.

Dimension reduction of the credal set

Representing a statistical model as a credal network highlights the reduction in computational load effected by networks.

Example 2.3 (Credal set as model (continued)) *Even if we restrict attention to Markov processes with these variables, a statistical hypothesis on them is determined by* $2^3(2^3-1)=56$ *probabilities for the transitions between valuations of the variables. Now consider the network that captures the independence relations.*



The complete extension of this network consists of the probability functions arrived at by filling in the probabilities on the nodes and edges. The corresponding model Θ_{net} *is determined by a vector* $\eta = \langle \gamma, \alpha_0, \beta_0, \alpha_1, \beta_1 \rangle$*, where*

$$P(c_i | H_\eta) = \gamma,$$
$$P(a_{i+1} | c_i^k \wedge H_\eta) = \alpha_k,$$
$$P(b_{i+1} | c_i^k \wedge H_\eta) = \beta_k.$$

The space Θ_{net} *is a strict subset of the space* Θ*. The above credal network may be characterised by a prior probability assignment that is nonzero only on this subset.*

Computational advantages of using the credal network

The dimension reduction and the choice of a particular coordinate system for the credal set, or statistical model, has a number of advantages.

- The graph reduces the dimensions of the statistical model, which can entail major reductions in computational load, parallel to the reduction in computational load effected by using Bayesian networks.
- Another important consequence concerns the coordinate system of the statistical model. By replacing the product of simplexes Θ with the space Θ_{net}, we are also availing ourselves of coordinates that are orthogonal to each other. Therefore, determination of the posteriors over the parameters of the model run entirely separately.
- This computational advantage shows up in the predictive probabilities that derive from the posterior probability assignments over the model. If the probability density over statistical hypotheses is a Dirichlet distribution, then the predictions for valuations of the separate network variables are Carnapian prediction rules.

So credal networks can be used to improve computation in Bayesian statistical inferences. However, this is not a direct application of the common inference machinery of Progicnet.

2.2 Extending statistical inference with credal networks

In this section we include the statistical hypotheses in a credal network representation, and investigate some of the evidential and logical relations between hypotheses and observed variables.

Interval-valued likelihoods

We have associated statistical hypotheses with probability assignments in the credal set. But we may also add the hypotheses as nodes in the network.

> **Example 2.4 (Fuzzy likelihoods)** *Consider a credal network consisting of a hypothesis node H_j with $j \in \{0, 1\}$ and a large number of instantiation nodes of the propositional variables C_i, labelled with i.*

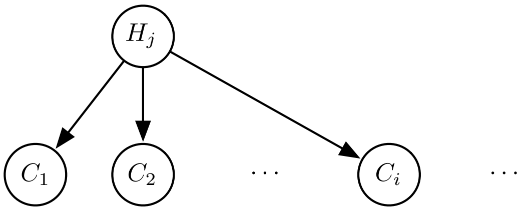


Figure 2: A graph containing the statistical hypothesis H_j as the top node. Conditional on the hypothesis, the propositional variables C_i for $i = 1, 2, \ldots$ are independent.

Example 2.5 (Fuzzy likelihoods (continued)) *These instantiation variables are independent of each other conditional on H_j, and each value j may be associated with a conditional probability of each of the instantiations. These conditional probability assignments can be filled in on the edges leading from the hypothesis node H_j to the separate instantiations of C_i.*

The idea to include a hypothesis in a credal network leads quite naturally to the idea of interval-valued likelihoods. That is, we may assign a probability interval to the edges connecting a statistical hypothesis with the data. Illustrating this by the above example, we replace the sharp probability values for c_i with

$$P(c_i|H_0) \in [0.3, 0.7],$$
$$P(c_i|H_1) \in [0.6, 0.8].$$

In words, this expresses that the statistical hypotheses are not exactly clear on the probability of c_t, although they do differ on it. The use of this formal possibility is in a sense complementary to the well-known method of Jeffrey conditioning.

We can generate the credal networks associated with a model having interval-valued likelihoods, as follows:

Algorithm 2.6 Construction of a credal network for Bayesian statistical inference in which the statistical hypotheses are included.

Input: a finite set of credal sets $\mathbb{K}_j$ for $j = 1, \ldots N$, each based on a given set of propositional variables $V = \{A_1, \ldots A_M\}$, independence relations I, and premises $\varphi_i^{X_i}$, such that the sets are all associated with a unique credal network $\mathcal{G}$.

1. **Addition of hypotheses nodes:** add the node H_j with possible valuations $j = 1, \ldots N$, and connect the node H_j with each variable A_j in the graph $\mathcal{G}$.
2. **Fixing the likelihoods:** associate each credal set $\mathbb{K}_j$ with a statistical hypothesis H_j, by setting the credal set for the variables V to $\mathbb{K}_j$ for the valuation j of the hypothesis node H_j.

Output: a statistical model in which each hypothesis H_j is associated with a credal set over the variables V.

Logically complex statements with statistical hypotheses
The idea to include hypotheses in the credal network invites a further application that derives from the standard semantics.

- Second-order probabilities over statistical hypotheses allow one to infer various useful quantities, such as expectation values and predictions.

- By including the statistical hypotheses in the credal network, we can also infer probability assignments over logically complex propositions concerning statistical hypotheses.

- We must be careful in interpreting the resulting probability assignments over logically complex propositions involving hypotheses, which are typically interval-valued: they pertain to a single statistical hypothesis and not to a range of values of statistical parameters.

- Statisticians may nevertheless make good use of credal networks that include statistical hypotheses. Often enough the proposition of interest in a scientific enquiry is logically complex, and very often the background knowledge contains logically complex propositions.

3 Application of Bayesian statistical inference

We now illustrate Bayesian inference in the Progicnet schema by means of an example on the measurement of psychological attributes.

3.1 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.

Latent variables: the psychological attributes are taken as the hidden causes of the observable facts about subjects.

Observable variables: the correlational structure among the values of observed variables are used to derive facts about these latent variables.

In the example we present subjects j with binary ability tasks A_j, B_j, and C_j, valued as α_j^i with $i = 0, 1$, etc. The binary latent variables F_j and G_j each discern two developmental stages. The latent processing speed H_j has N levels.

Modelling assumptions

Assume that all subjects are described by the same probability function over the variables, and that latent variables are independent components in determining the test performance:

$$\forall j \neq k : \quad P(V_j) = P(V_k),$$

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

Next to the independence premises, psychological theory might determine the following relations between assignments to the latent and the observable variables, for all j:

$$f \wedge g \rightarrow \neg a, \qquad \neg g \rightarrow a,$$

We might further have certain probabilistic relations between observed and latent variables:

$$P(b|g \wedge h^n) = \frac{n}{N}, \qquad P(c|h^n) = \frac{N+n}{2N}.$$

3.2 Bayesian inference in psychometrics

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, such that

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

We take a uniform distribution as prior and derive

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

The Bayesian inference can thus be represented straightforwardly in the form of the Progicnet schema:

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

We now extend this inference to apply to all variables and subjects, including all probabilistic restrictions presented above.

Setting up the statistical model

The idea of statistical inference is not just to learn values of variables within subjects, but to learn across subjects. Learning across subjects can take place because each subject has a valuation over both latent and observable variables that is drawn from the same multinomial distribution

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

Consider again the relation between the observable variables C_j and the latent variables H_j and choose $N = 3$, so that we have $3 \times 2 = 6$ complete valuations of C_j and H_j together. We can parameterise the distribution over these valuations with a probability

$$P(h_j^n) = \theta_{h^n} \qquad P(c_j^1|h_j^n) = 1 - P(c_j^0|h_j^n) = \theta_{C^n}.$$

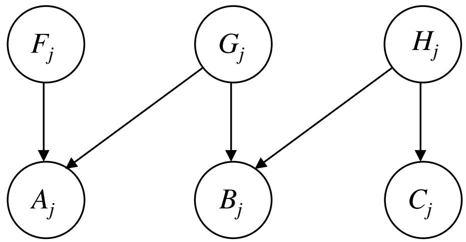
The additional restrictions to this set of distributions, deriving from the likelihoods of H_j for C_j, can be expressed in terms of these parameters:

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

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^n, or points θ in the parameter space, lead to the same overall probability for C.

Building the credal network

It is convenient to build up a credal network for the variables. By the independence relations given above we have the following network, which captures the independence relations for each subject *j* separately.



As indicated in the foregoing, this network dramatically reduces the number of parameters involved in the statistical inference. Instead of the $3 \times 2^5 - 1 = 95$, we can parameterise the whole space now by means of 1 parameter for F, 1 for G, 2 parameters over H, 4 for A, 4 for B, and 3 for C, totalling 15.

From logical relations to parameter restrictions

We can restrict the parameters of the credal network even further, by means of the logical relations between latent and observable variables. More specifically, we have that

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

is false, so that $P(a_j^0|g_j^0) = 0$ and hence

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

for $i = 0, 1$. Similarly, we have $f_j^1 \wedge g_j^1 \wedge a_j^1$ is false, so that

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

Other modelling assumptions provide restrictions even more straightforwardly, fixing $P(b_j^1|g_j^1 \wedge h_j^n)$ and $P(c_j^1|h_j^n)$ respectively.

Coordinates and likelihoods for the resulting model

With these restrictions we narrow down the set of multinomial distributions $P(A,...,H)$ to the credal set $\mathbb{P}$. It has the following degrees of freedom for the likelihoods,

$$P(a_j^1|f_j^0 \wedge g_j^1) = \theta_{A^1|F^0G^1}, \qquad P(b_j^1|g_j^0 \wedge h_j^n) = \theta_{B^1|G^0H^n},$$

and for the latent variables,

$$P(f_j^1) = \theta_{F^1}, \qquad P(g_j^1) = \theta_{G^1}, \qquad P(h_j^n) = \theta_{h^n}.$$

So for $N = 3$ we have 7 degees of freedom in the space of multinomial distributions. For each point within the space of multinomial distributions, we can derive likelihoods for the observable variables A and B, analogously to $P(c_j^1|h_j^n) = \theta_{C^n}$ for C:

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

Thus the likelihoods for A_j and B_j will also depend on the values of $\theta_{A^1|F^0G^1}$ and $\theta_{B^1|G^0H^n}$.

Bayesian inference

With these last specifications, we are ready to apply the machinery of Bayesian statistical inference. We have:

A model: the space of multinomial distributions over observable and latent variables, suitably restricted.

A prior probability over the model: typically the uniform prior, but more information may be available.

A sample of subjects: in particular their scores on observable variables.

We thus derive a posterior probability over the possible multinomial distributions, and thereby expectations for the latent variables and as yet unobserved subjects. The credal network present a number of improvements:

- The use of a network effects a reduction of the computational load of calculating with the model.
- We can combine logical information with statistical assumptions, and determine the probability of logical combinations of hypotheses.
- Interval-valued likelihoods provide a new way of observed variables to have bearing on the latent variables.
- We can perform a robustness analysis using interval-valued prior probabilities, following Walley.