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 and just now: the standard semantics and evidential probability. Plan for today:

Probabilistic argumentation: logical and probabilistic information, representation and interpretations.

Fiducial probability: classical statistical inference, fiducial arguments, representation and interpretation in the Progicnet schema.

Networks for fiducial probability: functional models and networks, aiding fiducial inference with networks.

And where we are heading tomorrow:

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.

1 Probabilistic argumentation

Probabilistic argumentation weighs the arguments and counter-arguments for some hypothesis in terms of probability. These are turned into degrees of support and possibility, measuring the degree to which the hypothesis is proved, or supported, by the arguments.

1.1 Logical and probabilistic information

Probabilistic argumentation requires the available evidence to be encoded by a finite set

$$\Phi = \{\varphi_1, \ldots, \varphi_n\} \subset \mathcal{L}_V$$

of sentences in a logical language $\mathcal{L}_V$ over a set of variables V and a fully specified probability measure

$$P : 2^{\Omega_W} \to [0, 1],$$

where Ω_W denotes the finite sample space generated by a subset $W \subseteq V$ of so-called probabilistic variables.

Definition 1.1 (Probabilistic Argumentation System) *A* probabilistic argumentation system *is a quintuple*

$$\mathcal{A} = (V, \mathcal{L}_V, \Phi, W, P),$$

where V, $\mathcal{L}_V$, Φ, W, and P are as defined above.

Support and possibility sets

For a given probabilistic argumentation system $\mathcal{A}$, let another logical sentence $\psi \in \mathcal{L}_V$ represent the hypothesis in question. Consider the subset of Ω_W whose elements, if assumed to be true, are each sufficient to make ψ a logical consequence of Φ. Formally, this set of so-called *arguments* of ψ is defined by

$$Args_{\mathcal{A}}(\psi) = \{\omega \in \Omega_W : \Phi|\omega \models \psi\}, \quad (1)$$

where $\Phi|\omega$ denotes the set of sentences obtained from Φ by instantiating all the variables from W according to the partial truth assignment $\omega \in \Omega_W$. The elements of $Args_{\mathcal{A}}(\neg\psi)$ are sometimes called *counter-arguments* of ψ.

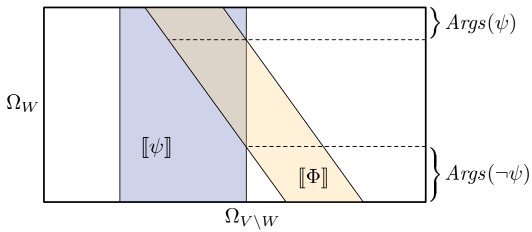


Figure 1: The sample space Ω_W is shown as the vertical sub-space of the complete space $\Omega_V = \Omega_W \times \Omega_{V\setminus W}$. The set of arguments $Args(\psi)$ is the horizontal projection of the area of Φ that lies entirely inside ψ.

Degrees of support and possibility

On the basis of these support and possibility sets and a probability assignment over W we can now define:

Definition 1.2 (Degree of Support) *The* degree of support *of ψ, denoted by $dsp_{\mathcal{A}}(\psi)$ or simply by $dsp(\psi)$, is the conditional probability of the event $Args(\psi)$ given the evidence Φ,*

$$dsp(\psi) = P(Args(\psi)|\Phi) = \frac{P(Args(\psi)) - P(Args(\perp))}{1 - P(Args(\perp))}. \tag{2}$$

Definition 1.3 (Degree of Possibility) *The* degree of possibility *of ψ, denoted $dps_{\mathcal{A}}(\psi)$ or simply by $dps(\psi)$, is defined by*

$$dps(\psi) = 1 - dsp(\neg\psi). \tag{3}$$

Note that these formal definitions imply $dsp(\psi) \leq dps(\psi)$ for all hypotheses $\psi \in \mathcal{L}_V$, but this inequality turns into an equality $dsp(\psi) = dps(\psi)$ for all $\psi \in \mathcal{L}_W$.

1.2 Representation and interpretations

We can represent a probabilistic argumentation system $\mathcal{A} = (V, \mathcal{L}_V, \Phi, W, P)$ in the form of the left hand side of the Progicnet framework accoring to

$$\varphi_1^{1.0}, \ldots, \varphi_n^{1.0}, \alpha_{\omega_1}^{x_1}, \ldots, \alpha_{\omega_m}^{x_m}, \text{ with } x_i = P(\{\omega_i\}),$$

where the α are complete valuations of all the $A_i \in W$. Now move to interpreting instances of the Progicnet schema as respective probabilistic argumentation systems.

Interpreting degrees of support and possibility

- An important property of degree of support is its consistency with pure logical and pure probabilistic inference. At the extreme cases of $W = \emptyset$ and $W = V$, degrees of support degenerate into classical logical entailment and entailment in the standard semantics.

- When it comes to quantitatively judging the truth of a hypothesis ψ, it is possible to interpret degrees of support and possibility as respective lower and upper bounds of a corresponding credal set.

- There is a strict mathematical analogy between degrees of support/possibility and belief/plausibility in the Dempster-Shafer theory of evidence.

Interpreting the inferences in probabilistic argumentation

- As in the standard semantics, we interpret the probability sets attached to premises as constraints on the possible probability measures P.

- Let us further assume Φ as given in addition to some premises $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n}$. If we then fix W to be the set of variables appearing in $\varphi_1, \ldots, \varphi_n$, we can apply the standard semantics to obtain the set

$$\mathbb{P} = \{P : P(\varphi_i) \in X_i, \ \forall i = 1, \ldots, n\}$$

  of all admissible probability measures w.r.t. to the sample space Ω_W.

- The result is what could be called an *imprecise probabilistic argumentation system* $\mathcal{A} = (V, \mathcal{L}_V, \Phi, W, \mathbb{P})$.

- Alternatively, we may look at each probability measure from $\mathbb{P}$ individually and consider the family $\mathbb{A} = \{(V, \mathcal{L}_V, \Phi, W, P) : P \in \mathbb{P}\}$ of all such probabilistic argumentation systems, each of which with its own degree of support and degree of possibility function.

- Instead of using the sets X_i as constraints directly for P, we may also interpret them as respective constraints for corresponding degrees of support or possibility.

2 Fiducial probability

An important application of probability theory is the use of statistics in science, in particular classical statistics as devised by Fisher and Neyman and Pearson. Here we discuss one way in which it can be accommodated by the Progicnet framework.

2.1 Classical statistical inference

Classical statistical procedures concern probability assignments over data relative to a statistical hypothesis H.

- Let Ω_H denote a statistical model, consisting of statistical hypotheses H_j with $0 \leq j < n$ that are mutually exclusive.
- The assignment that H_j is true is written h_j. Let Ω_D be the sample space, consisting of observations of binary variables D_i. We write d for a specific sample.
- An essential property of these procedures is that they only involve the derivation of the probability of some event for a given statistical hypothesis, $P_{h_0}(d)$.

The crucial difference with Bayesian statistics is that no probability is assigned to the hypotheses. Decisions to reject or accept a hypothesis are based on a probability for the data, conditional on the hypothesis. Hence we might say the procedures do not concern inferences about hypotheses.

Hypotheses testing

Restricting Ω_H to $n = 2$, we can compare the hypotheses h_0 and h_1 by means of a Neyman-Pearson test function.

> **Definition 2.1 (Neyman-Pearson Hypothesis Test)** *Let T be a set function over the sample space Ω_D,*
>
> $$T(D) = \begin{cases} 1 & \text{if } \frac{P_{h_1}(D)}{P_{h_0}(D)} > t, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$
>
> *where P_{h_j} is the probability over the sample space determined by the statistical hypothesis h_j. If $T = 1$ we decide to reject the null hypothesis h_0, else we reject the alternative h_1.*

The decision to reject is associated with a significance and a power of the test:

$$\text{Significance}_T(D) = \int_{\omega \in \Omega_D} T(D) P_{h_0}(D) d\omega, \qquad \text{Power}_T(D) = \int_{\Omega_D} (1 - T(D)) P_{h_1}(D) d\omega.$$

Neyman and Pearson proved that test functions depending only on the likelihood-ratio have optimal significance and power. Evidential probability is designed for capturing the dynamics of the significance levels when combining hypotheses.

Estimation of parameters

Among a larger set of statistical hypotheses, we may also choose the best performing one according to an estimation procedure, e.g., the maximum likelihood estimator

> **Definition 2.2 (Maximum Likelihood Estimation)** *Let Ω_H be a model with hypotheses H_θ, where $\theta \in [0, 1]$. Then the maximum likelihood estimator is*
>
> $$\hat{\theta}(D) = \{\theta : \forall\theta' \big(P_{h_{\theta'}}(D) \leq P_{h_{\theta'}}(D)\big)\}, \tag{5}$$
>
> *where D is again generic data. So the estimator is a set, typically a singleton, of those values of θ for which the likelihood of h_θ is maximal. The associated best hypothesis we denote with $h_{\hat{\theta}}$.*

We may further compute the so-called confidence interval. Define a region R of parameter values for which the data are not highly unlikely, $R(D) = \{\theta : P_{h_\theta}(D) > 1\%\}$. The so-called symmetric 95% confidence interval is slightly more complicated:

$$C_{95}(D) = \left\{ \theta : |\theta - \hat{\theta}| < \lambda \text{ , and } \int_{\hat{\theta}-\lambda}^{\hat{\theta}+\lambda} P_{H_\theta}(D)d\theta = .95 \right\}.$$

Every element of the sample space is assigned a region of parameter values, which expresses the quality of the estimate.

2.2 Fiducial arguments

Fisher's fiducial argument allows us more than choosing the best among the available hypotheses: we can derive a probability assignment over hypotheses. Thus classical statistics can be reconciled with an inferential attitude towards hypotheses after all.

Example 2.3 (Fiducial argument) *Suppose that it is known that a quantity, F, is distributed normally with an unknown mean, μ, and a known variance σ^2 of 1.*

- *Since $\mu - r$ is known to be normally distributed, we can look up the table value of the probability that $|\mu - r|$ exceeds any given amount.*
- *For instance, if $r = 10$ is observed, we can, by direct inference, infer that the probability that μ is between 9 and 11 is 0.68.*

This example is an illustration of Fisher's fiducial inference. The argument draws an inference from observable data (that F takes r in the sample) to a statistical hypothesis (the mean μ of F) without following the form of inverse inference.

The fiducial argument is controversial and its exact formulation is a subject of debate. The controversy stems from Fisher setting out goals for fiducial inference that do not appear mutually consistent, and is compounded by Fisher's informal account of fiducial probability.

Functional models: probabilistic arguments

Dawid and Stone provide a general characterisation of the set of statistical problems that allow for application of the fiducial argument, using so-called functional models.

Definition 2.4 (Functional Model) *Let Ω_H be a statistical model, and Ω_W a set of stochastic elements ω. A functional model consists of a function*

$$f(H_\theta, \omega) = D, \tag{6}$$

and a probability assignment $P(\omega)$ over the stochastic elements in Ω_W.

A functional model relates every combination of a statistical hypothesis H_θ and an assumed stochastic element ω to data D. On the assumption of a functional model, we can derive the likelihoods of the hypotheses h_θ:

$$P(d|h_\theta) = \int_{\omega \in V_d(h_\theta)} P(\omega) d\omega. \tag{7}$$

A functional model is smoothly invertible iff there is a function $f^{-1}(H_\theta, D) = \omega$, so that relative to the sample D, each hypothesis H_θ is associated with exactly one stochastic element ω. The fiducial argument is applicable iff the functional model is smoothly invertible.

Support and possibility for statistical hypotheses
Consider the hypothesis $h_I = \cup_{\theta \in I} h_\theta$. The set $U_d(\omega) = \{h_\theta : f(h_\theta, \omega) = d\}$ covers exactly those hypotheses that point to the sample d under the assumption of the stochastic element ω. Now define:

$$Sup_{h_I}(d) = \{\omega : U_d(\omega) \subset h_I\} \qquad Pos_{h_I}(d) = \{\omega : U_d(\omega) \cap h_I \neq \emptyset\}.$$

The former are arguments for h_I, the latter are the complement of arguments against h_I. We define the degrees of support and possibility for h_I as

$$dsp(h_I) = \int_{\omega \in Sup_{h_I}(d)} P(\omega) d\omega, \qquad dps(h_I) = \int_{\omega \in Pos_{h_I}(d)} P(\omega) d\omega.$$

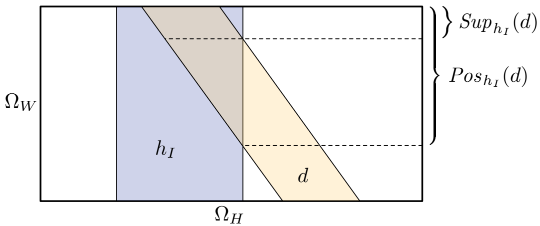


Figure 2: The rectangle shows all combinations of H_θ and ω. The combinations for which $f(H_\theta, \omega) = d$ are included in the shaded area.

2.3 Representation and interpretation

Using functional models, we can represent fiducial degrees of support and possibility in the Progicnet framework, much like we captured probabilistic argumentation:

$$(\{(h_\theta,\omega) : f(h_\theta,\omega) = d\})^1, \omega^{P(\omega)} \models h_I^{[dsp,dps]}. \tag{8}$$

Here h_I refers to the statistical hypothesis of interest, as defined above, and ω to the stochastic element of the functional model, over which the probability assignment $P(\omega)$ is given. The function f determines how these elements relate to possible data D.

- The set $\{(H_\theta,\omega) : f(H_\theta,\omega) = d\}$ refers to all combinations of hypothesis and stochastic element that lead to d as evidence.
- The set of these combinations (H_θ,ω) is given probability 1, meaning that the evidence occurred.
- In the case that the function f is smoothly invertible the interval-valued assignment to h_I collapses to a sharp probability value.

Sadly, the inferential representation and interpretation of this classical statistical procedure runs into difficulty with dynamic coherence, as will become apparent from the step-by-step fiducial argument.

3 Networks for fiducial probability

Fiducial probability can thus be given a modest place in the Progicnet framework. We now present a way in which credal networks may be used to aid the computation of fiducial probability.

3.1 Functional models and networks

We first provide a network representation for the fiducial inference represented by functional models. We restrict attention to fiducial probability using a smoothly invertible functional relation.

- The relation between the hypotheses H_θ, the stochastic elements ω, and the data D is such that the stochastic elements and the hypotheses are probabilistically independent: $P(H_\theta, \omega) = P(H_\theta)P(\omega)$.

- Given a hypothesis h_θ and a stochastic element ω, the occurrence of the data d is completely determined, $P(d|h_\theta, \omega) = I_d(h_\theta, \omega)$ where $I_d(h_\theta, \omega) = 1$ if $f(h_\theta, \omega) = d$ and $I_d(h_\theta, \omega) = 0$ otherwise.

If we condition on the observed data d then, because of the network structure and the further fact that the relation $f(H_\theta, \omega)$ is deterministic, the variables ω and H_θ become perfectly correlated: each h_θ is associated with some $\omega = f^{-1}(h_\theta, d)$.

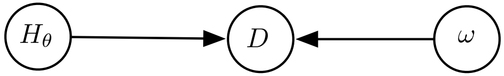


Figure 3: A network representing the conditional independencies between H_θ, ω and D in the functional model $f(H_\theta, \omega) = D$ that generates the fiducial probability for H_θ from the distribution over the stochastic elements ω.

3.2 Aiding fiducial inference with networks

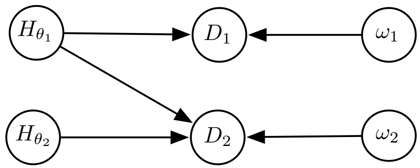
We now turn to the use of networks in aiding fiducial inference.

Step-by-step fiducial argument
Say that there are many statistical parameters θ_j in the model, and equally many observed variables D_j. We may then employ independence relations that obtain between the parameters and the data to speed up the fiducial inference.

Example 3.1 (Step-by-step fiducial argument) *Consider hypotheses H_θ determined by two variables, $\theta = \langle \theta_1, \theta_2 \rangle$, stochastic elements $\omega = \langle \omega_1, \omega_2 \rangle$, and two propositions D_1 and D_2 representing the data. We have a number of independence relations between these variables and data sets:*

$$P(D_1 \wedge D_2 | H_\theta \wedge \omega) = P(D_1 | H_{\theta_1} \wedge \omega_1) P(D_2 | H_{\theta_1} \wedge H_{\theta_2} \wedge \omega_2),$$
$$P(H_\theta) = P(H_{\theta_1}) P(H_{\theta_2}),$$
$$P(\omega) = P(\omega_1) P(\omega_2).$$

Finally, we have smoothly invertible functional relations $f(H_{\theta_1}, \omega_1) = D_1$ and $f_{\theta_1}(H_{\theta_2}, \omega_2) = D_2$, meaning that for each fixed value of θ_1 the function f_{θ_1} is smoothly invertible.



We first derive a fiducial probability $P(H_{\theta_1} | d_1)$, assuming we do not have knowledge of D_2 or H_{θ_2}. Then we derive a fiducial probability for H_{θ_2} from d_2, first computing a fiducial probability over H_{θ_2} for each value of θ_1 separately. We then use the law of total probability to arrive at $P(H_{\theta_2} | d_2)$.

An algorithm for setting up a network for a step-by-step fiducial argument is provided in the reading material.

Trouble with the step-by-step fiducial argument

Fiducial probabilities are controversial, and this is all the more so for the step-by-step procedure. We explain briefly why we recommend caution in the application of the step-by-step fiducial argument.

- We can only apply the fiducial argument on the condition that, in the absence of knowledge about H_{θ_2}, the data D_1 indeed allow for transferring the probability assignment over ω_1 to the hypotheses H_{θ_1}.
- There is something a bit awkward about the assumption of the absence of knowledge about H_{θ_2} and the use of the function f.
- From the above graph we can deduce that if there is some definite probability assignment $P(H_{\theta_2})$, learning about D_2 will be relevant to the probability of H_{θ_1}.
- By the assumption of a particular distribution $P(H_{\theta_2})$ we might destroy the smoothly invertible relation $f(H_{\theta_1}, \omega_1) = D_1$.

A lot hinges on the absence of knowledge about H_{θ_2}. This makes step-by-step fiducial inference more problematic than its simple and direct application.