Inductive Logic and Confirmation

Kent – Paris Campus, 18 October 2013

Inductive Logic for Rich Languages

⋆

Jan-Willem Romeijn

University of Groningen

What this talk is about

Several items in statistics and inductive logic keep me busy:

- The analogical prediction rules pioneered by **Carnap**, **Jeffrey**, **Hintikka**, and many of his students and followers.
- The representation theorem by **de Finetti** linking prediction rules to **Bayes**ian statistical inference.
- The idea of rich languages from **Gaifman and Snir** and the convergence theorems that follow from that.
- The notion of a random sequence developed by **von Mises**, and their use in a frequentist theory of chance.

I take this talk as a good occasion to connect these dots.

What will emerge?

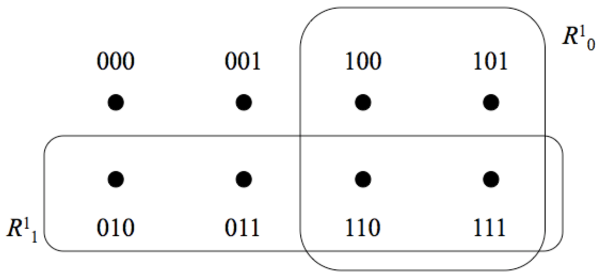
An enrichment of inductive logic that fits better with Bayesian statistics and its use of hypotheses.

Such an inductive logic can naturally accommodate analogy considerations and universal hypotheses.

Contents

1 Inductive logic

A Carnapian prediction rule is a probability distribution over an algebra of observation events $\mathcal{R}$. For t events the algebra is $\{0, 1\}^t$.



Results $s_1 s_2 \cdots s_i$ occurring after time t are denoted with the element $R_t^{s_1 s_2 \cdots s_i}$. If t is zero we omit it.

Conditioning on a given sequence

A prediction rule defines a probability distribution P over the observation algebra. Let t_q be the number of occurrences of q in the sequence $s_1 s_2 \cdots s_t$.

$$P(R_t^q | R^{s_1 \cdots s_t}) = \frac{t_q + \gamma_q \lambda}{t + \lambda}.$$

The prediction rule fully determines a probability over $\mathcal{R}$. We accommodate the sequence $s_1 s_2 \cdots s_t$ by simple conditioning.

Universal and analogical predictions

Several interesting classes of prediction rules were developed. Imagine the results are ternary, $q \in \{0, 1, 2\}$ for apple, pear, and banana.

- Hintikka systems factor in that one of the fruits may never be observed.
- Analogical predictions bring out that, e.g., observing apples may favor pears over banana's.

We can encode these inductive effects directly into a prediction rule and hence into a probability assignment over $\mathcal{R}$.

2 De Finetti's representation theorem

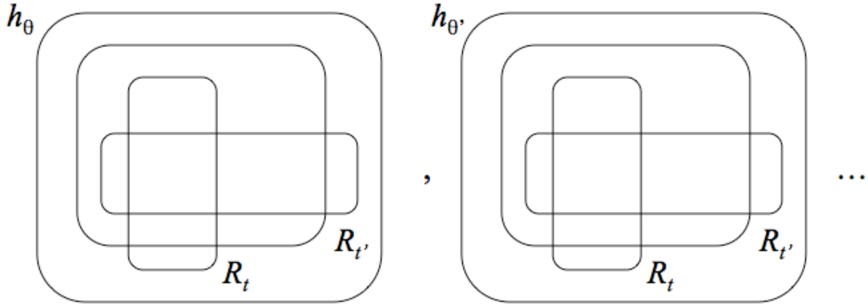
Any Bayesian inference over Bernoulli hypotheses corresponds to a rule whose predictions are invariant under permutations in the order of observations:

$$\text{Exchangeable } P(R_t^q | R^{s_1 \cdots s_t}) \quad \Leftrightarrow \quad \begin{cases} \text{Prior } P(h_\theta) \\ \text{Likelihoods } P_\theta(R_t^q | R^{s_1 \cdots s_t}) \\ \text{Bayesian updating with } P_\theta(\cdot) = P(\cdot | h_\theta). \end{cases}$$

De Finetti used this theorem to argue that we can dispose of the metaphysically suspicious story about hypotheses altogether.

Bayesian statistical inference

If we construct a probability model for Bayesian inference over statistical hypotheses, the latter indeed appear as supra-empirical.



The hypothesis h_θ shows up as a distribution P_θ over a tagged observation algebra, $h_\theta \times \mathcal{R}$.

3 Gaifman's rich language

Gaifman and Snir's [1982] paper is most well-known for the convergence theorems: Bayesian inference converges to truth values, priors wash out.

I will now focus on their use of rich languages: they show how to express statistical hypotheses in a space of possible observations.

Hypotheses as elements in $\sigma(\mathcal{R})$

We construct an idealised sample space consisting of infinitely long samples: $\{0,1\}^{\Omega}$.

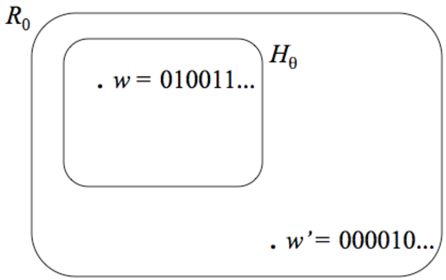
$$w = 010011011001010\ldots$$

We can identify the hypothesis h_θ, and its distribution P_θ, with a particular set of sequences w, so-called tail events in $\sigma(\mathcal{R}) \setminus \mathcal{R}$:

$$H_\theta = \{w : \text{Relative frequency}(w) = \theta \text{ and } w \text{ otherwise random}\}.$$

Hypotheses as sets of *Kollektivs*

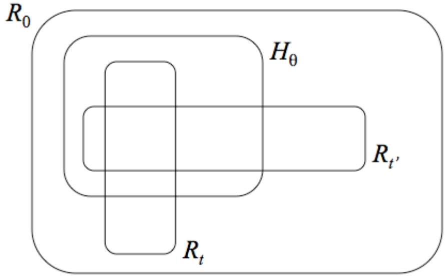
Note that the elements $w = 00110111\ldots$ of H_θ are von Mises collectives that instantiate the probability distribution P_θ.



This is frequentism in reverse: we presuppose a distribution and use frequentism to relate it to a model of empirical fact.

Events as distributions

The distribution P_θ of hypothesis h_θ is thus associated with a particular set H_θ that lie *inside* the sample space.



Each set H_θ intersects with every observation R that is assigned some probability by P_θ.

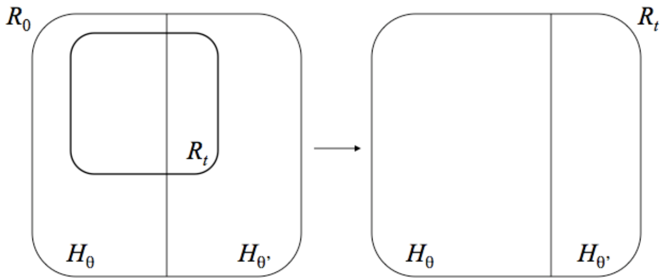
4 Frequentism as formal semantics

You can be both a Bayesian and a frequentist, much in line with Jeffrey's mixed Bayesianism.

The association of hypotheses and events offers many conceptual advantages.

Unique extension and convergence

The assignment of probability to a distribution, $P(H_\theta)$, becomes automatic: P extends uniquely from $\mathcal{R}$ to $\sigma(\mathcal{R})$.



The convergence theorems show up as a matter of course: if the observations are separating, they will zoom in on one of the sets H_θ.

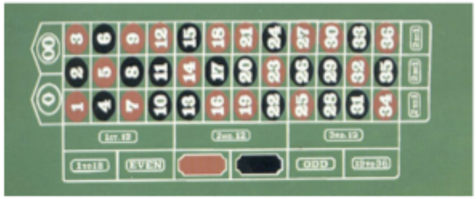
Hypotheses fix inductive dependence

Recall that prediction rules fix inductive relations between observations by constraints on the probability over the observation algebra $\mathcal{R}$.

For any rule we can find hypotheses that provide an alternative route to fixing the constraints: they enrich the language of inductive logic.

Why statisticians use hypotheses

In the sciences we hardly ever find statistical analyses that employ inductive relations among observations directly.



One explanation is that statistical hypotheses are a succinct, and perhaps more expressive way of fixing inductive relations among observations.

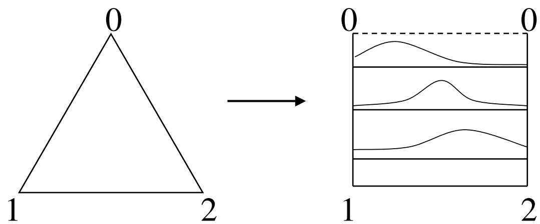
5 Analogical reasoning

The efficiency of using hypotheses can be illustrated nicely in the context of exchangeable analogical predictions.

By the foregoing, we are looking for a prior over Bernoulli hypotheses that brings out the salient inductive relevances.

Apples, bananas, pears

Basic idea: among hypotheses that give a high chance to apples, give higher prior probability to the ones that favor pears over bananas.



Observing a single apple will make apples more probable (PIR). But pears will loose less of their probability than bananas, and may even benefit!

Analogical and universal prediction rules

Several classes of prediction rules can be understood in terms of particular classes of priors in this way.

- Hintikka systems: non-zero probability mass on the extremities of the space of statistical hypotheses.
- Skyrms' hyper-Carnapian inductive rules: mixtures of Dirichlet priors over the space of statistical hypotheses.
- Paris and Hill on analogical reasoning: Dirac-delta priors over the space of statistical hypotheses.

6 Conclusion

I have shown how De Finetti's representation theorem and Gaifman and Snir's idea of rich languages can help to align inductive logic and Bayesian statistics.

- This elucidates statistical inference by specifying an observational content for statistical hypotheses.
- The idea of frequentist chance is thereby wedded to the Bayesian idea that we assign probability to hypotheses.
- It provides insight into the convergence theorems for Bayesian inference.
- And it suggests why statistical hypotheses are being used in the first place: they make inductive logic more succinct and manageable.

Thanks!

This talk will be available at http://www.philos.rug.nl/~romeyn. For comments and questions, email j.w.romeijn@rug.nl.