Düsseldorf 2007

Abducted by Bayesians?

Jan-Willem Romeijn
Faculty of Philosophy
University of Groningen

Puzzle

☝ Hypotheses that assign equal probability to the data are only theoretically distinct.

☝ For empirical purposes, such theoretical distinctions between hypotheses seem redundant.

💣 Science sometimes employs hypotheses that are only theoretically distinct, for example in terms of causality or simplicity.

Plan

❶ Bayesian statistical inference

❷ The use of theoretical distinctions

❸ Example: determining causal structure

❹ Abductive inference

❺ Conclusions

❶ Bayesian statistical inference

Statistical inference aims at determining the most likely chance hypothesis from the data.

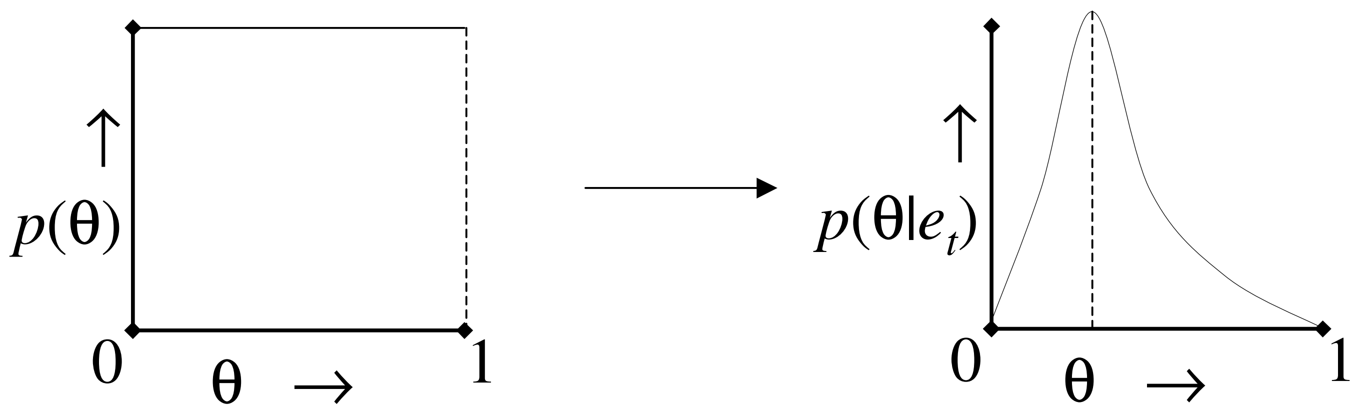
hypothesis θ: $p(q_t = 1 \mid \theta, e_{t-1}) = \theta.$

model Θ: $\Theta = \{\theta : \theta \in [0,1]\},$

A statistical model is a set of hypotheses from which we may choose the most likely one.

Probability over hypotheses

In Bayesian statistical inference, we assign a prior probability over chance hypotheses.



We adapt this probability to the data to arrive at estimations of chance parameters, and predictions.

Carnapian prediction rules

By choosing the appropriate prior probability $p(\theta)$ we can derive Carnapian prediction rules c_λ:

$$p(q_{t+1} = 1 \mid e_t) = \frac{t_q + \lambda/2}{t + \lambda}.$$

The larger λ, the slower the predictions for q change from a preconceived probability $1/2$ to the observed relative frequency t_q / t.

De Finetti representation

The use of a continuum of hypotheses θ results in exchangeable predictions,

$$p(q_{t+1} = 1 \mid e_t) = f(t_q, t).$$

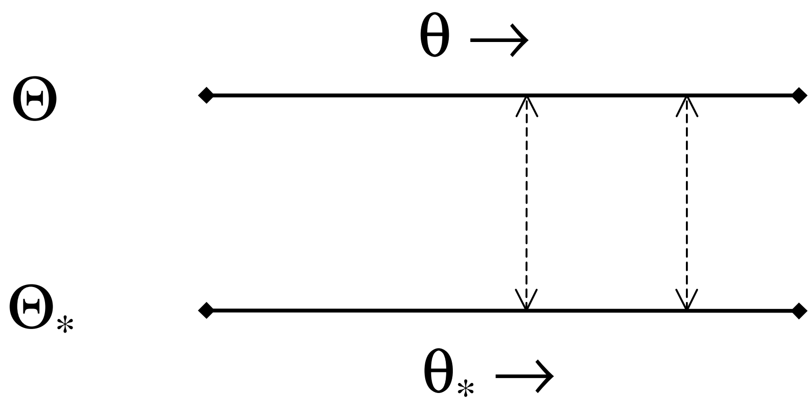
If we define a Dirichtlet prior over these hypotheses,

$$p(\theta)d\theta \quad \propto \quad \theta^{\lambda/2-1}(1\text{-}\theta)^{\lambda/2-1},$$

we effectively define the Carnapian rule c_λ.

❷ Theoretical distinctions

The hypotheses θ all have different likelihoods. So they have different observational content.

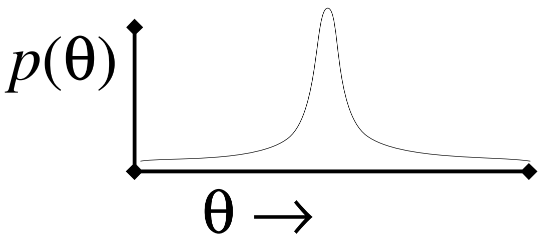


Can there be any use for distinguishing between hypotheses that have the same likelihoods?

Using prior knowledge

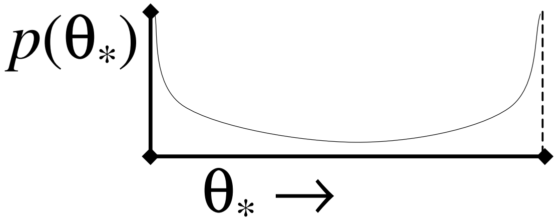
Here is an example. We can express the difference between a normal and a magical coin in a prior probability over the submodels Θ and Θ_* as follows:

Θ:



→ predictions c_λ with $\lambda = 12$

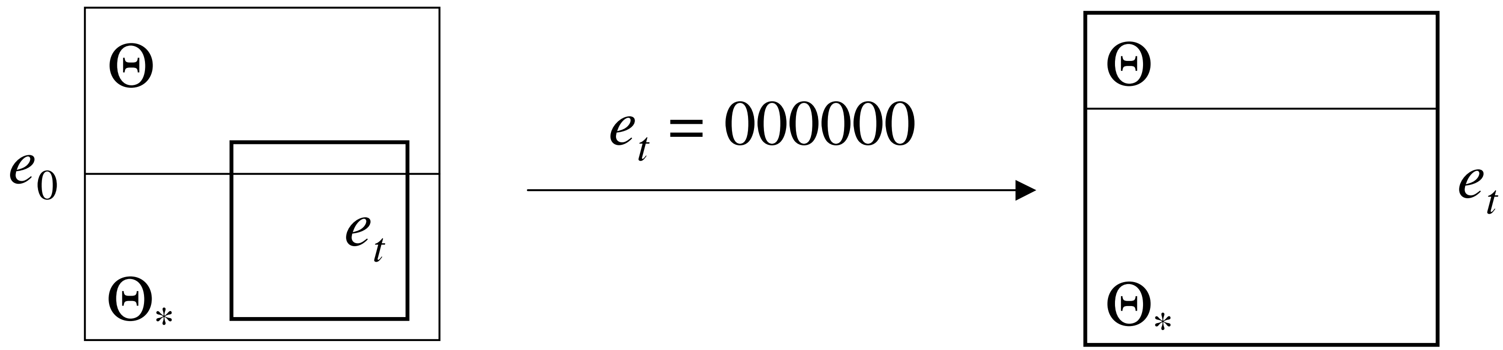
Θ_*:



→ predictions c_λ with $\lambda = 1/3$

Hypercarnapian rules

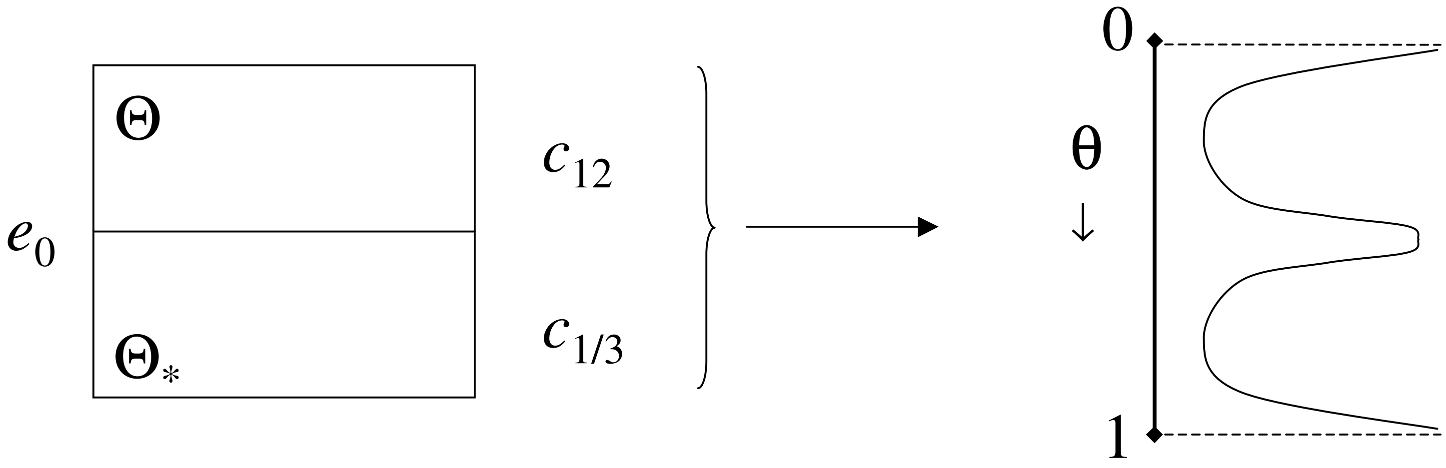
The c_λ function as likelihoods for the submodels Θ and Θ_*.



The degenerate model results in estimates that are a weighted average of the Carnapian rules. They show faster convergence and better short-term accuracy.

Using theoretical distinctions

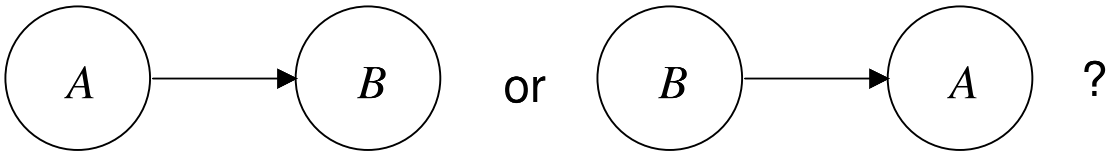
There is a prior over a single model Θ that generates the same predictions as this degenerate model.



But dealing with Θ and Θ_* separately makes for an easier motivation of the priors, and less complicated updating.

❸ Determining causal structure

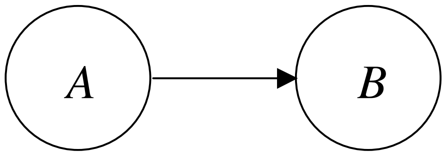
Several theoretical distinctions may be associated with differing prior probability assignments:



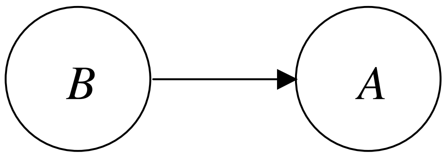
An important example of a theoretical distinction concerns the distinction between opposing causal directions.

Causal models

The direction of the arrow in a causal model suggests a specific parametrisation of the hypotheses space:

Θ_A : 

so $p(A=1) = \alpha$ and $p(B=1|A=i) = \beta_i$

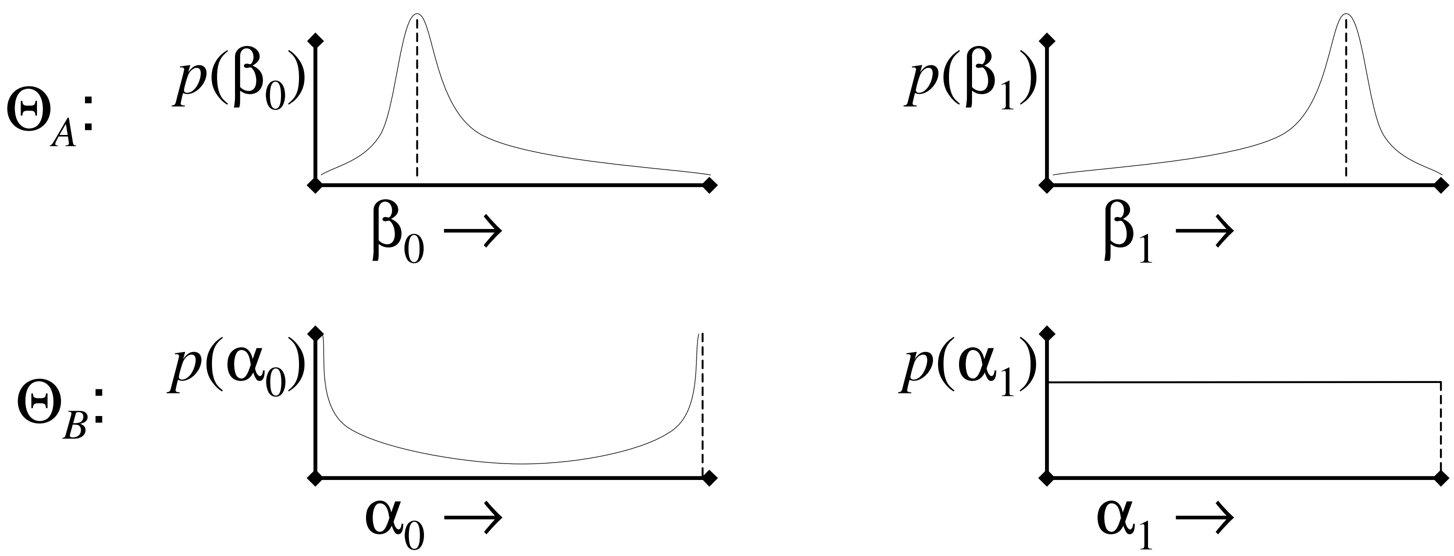
Θ_B : 

so $p(B=1) = \beta$ and $p(A=1|B=i) = \alpha_i$

Note: a non-informative prior over the space (α,β_0,β_1) is different from such a prior over $(\beta,\alpha_0,\alpha_1)$.

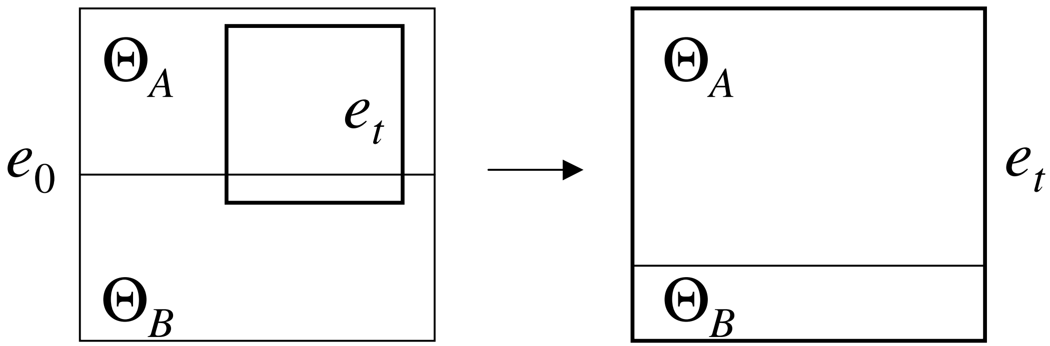
Additional prior information

We may have further expectations concerning the parameters. Encoding these in the priors will sharpen the distinctions between the two submodels. For example:



Updating over causal models

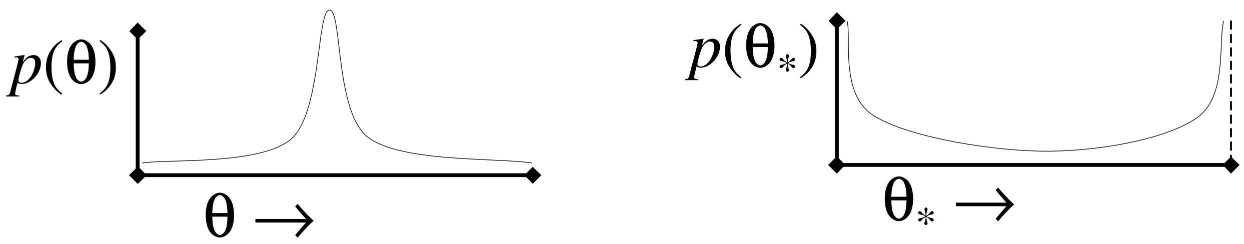
The submodel with the prior that expresses the true causal relation gains more posterior probability.



This submodel will, like in the coin case, show faster convergence and more accurate short-term estimations.

Theoretically informed priors

The above example shows that degenerate models are statistically useful, when told apart by considerations on underlying structures, like causality.



The theoretically informed priors over the submodels allow us to tell apart the submodels, and improve convergence properties and estimations.

❹ Abductive inference

A first conclusion is that we can use theoretical distinctions between hypotheses in statistical inference.

- Theoretical distinctions can be useful for expressing knowledge in a prior probability assignment.

- If they can indeed be used in that way, conditioning allows us to tell apart the theoretically distinct hypotheses after all.

But are these distinctions theoretical or empirical?

Theoretical and empirical

The distinction between the hypotheses θ and θ_* in the coin example may be called theoretical, because they have identical likelihoods.

$$e_t = 1011001001011100 \longrightarrow \text{normal}$$

$$e_t = 0000000001000000 \longrightarrow \text{trick}$$

But the priors make the submodels Θ and Θ_* empirically distinct. The submodels may only be called theoretically distinct relative to the hypotheses θ.

Observations

This empirical content of theoretical notions may be considered a formal explication of the fact that there is no sharp distinction between observation and theory.

- The nature and meaning of observations is in part determined by the theoretical structure in which they find a place.

- The theoretical considerations on underlying processes, causality and simplicity are in this sense also empirical.

Abducted by Bayesians

The inferences presented in the Bayesian framework of the above employ theoretical virtues, and can therefore be called abductive.

- The inferences illustrate how theoretical considerations such as causality can play a role in the probability kinematics.

- These considations enter the inferences in the motivation of priors over hypotheses, which derive from the stories associated with the hypotheses.

❺ Conclusions

☆ Theoretical distinctions can in certain cases be used to express knowledge on underlying processes in a prior probability assignment.

☆ In such cases, updating the probability assignment by conditioning can tell apart hypotheses that are only theoretically distinct.

☆ This presents us with a Bayesian model of a kind of abductive inference.

Disclaimers

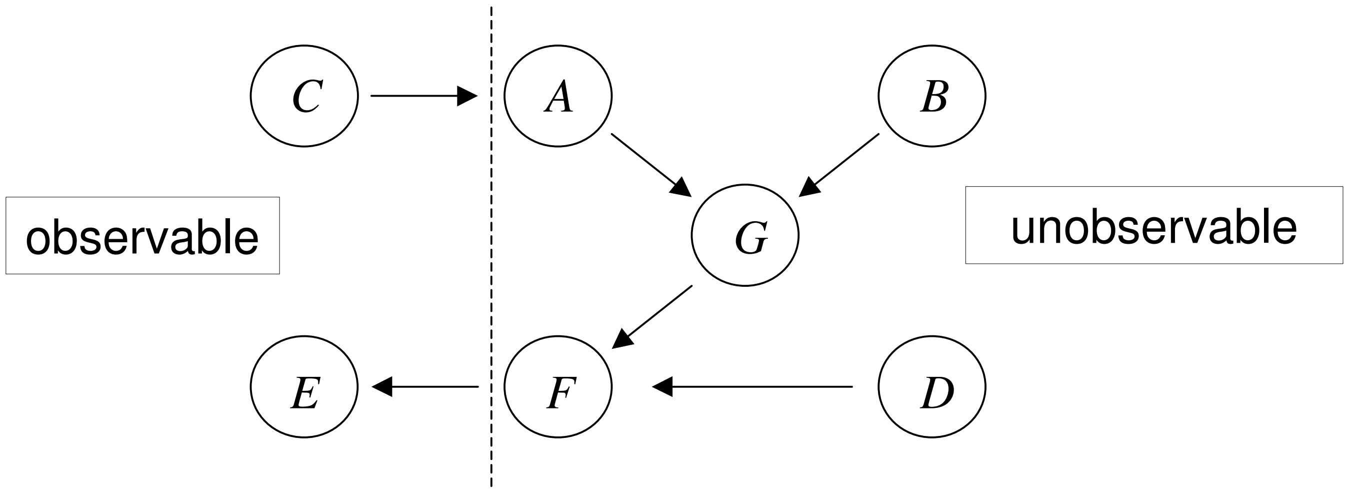
I do not claim that the Bayesian setting can accommodate all types of abductive inference.

- Not all useful theoretical distinctions will be translatable into a specific shape of the prior. The beauty of a scientific theory may be a case in point.

- There may be cases of abductive inference in which scientists do not follow any rule, and ignore the higher posterior of some hypotheses for completely external reasons.

Back to science

I hope that reasoning in experimental science can be understood in this way.



Theoretical notions are simply useful, and abductive inference naturally follows from that.

Thanks

j.w.romeijn@rug.nl

www.fmg.uva.nl/romeijn

dr. J.W. Romeijn
RuG/Philosophy
Oude Boteringestraat 52
9712 GL Groningen
The Netherlands