Core Logic 2006

Statistical Inference, Induction and Realism

J.W. Romeyn

Psychological Methods

University of Amsterdam

Outline of presentation

❶ The problem of induction

❷ The Carnapian solution: logical probability

❸ Bayesian inductive logic

❹ Statistical hypotheses and frequentism

❺ Empiricism and subjectivism

❻ Statistical underdetermination

❶ Problem of induction

Past observations do not entail anything about future observations.

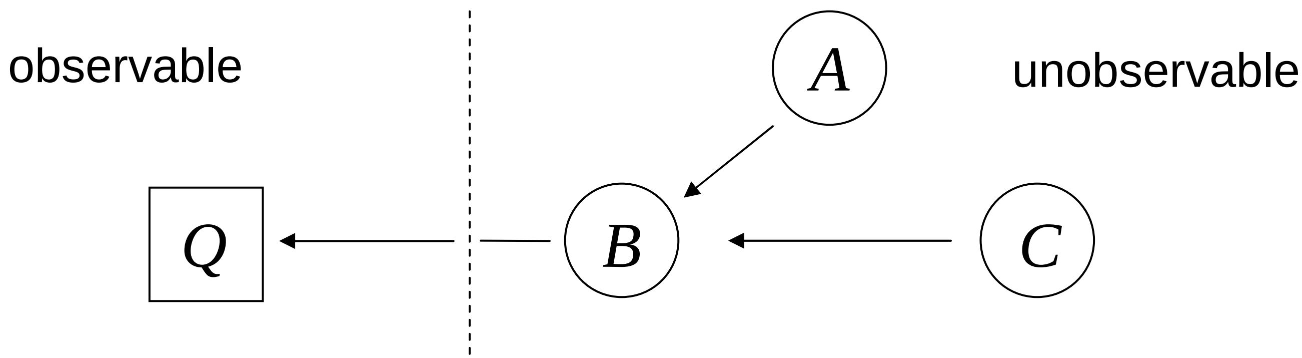
$$e_t = 10100001010101010000\ ? \quad \begin{array}{l} \nearrow\ q = 1 \\ \searrow\ q = 0 \end{array}$$

Rather dramatically, there is no justification for predictions from science or common sense.

❶

Underdetermination

Past observations do not entail anything about the structure underlying the observations either.



Limited stocks of data are always consistent with a large number of distinct scientific theories.

❷ Probabilistic solution?

Perhaps we can derive probabilistic predictions on the observations from past observations.

$$e_t = 10100001010101010000 \begin{array}{l} \nearrow \; p\,(q{=}0 \mid e_t) = (1 - \alpha) \\ \searrow \; p\,(q{=}1 \mid e_t) = \alpha \end{array}$$

For a solution the predictions need only be comparative, for example $\alpha < 1/2$.

❷

Carnapian logic

For Carnap, probabilistic solutions can be based solely on the choice of an observation language.

$$e_t = 10100001010101010000$$

$$e'_t = 00000000000000111111$$

$$p\,(q{=}1 \mid e_t) = \alpha$$

Specifically, he took relative frequencies as the basis for predictions. Such predictions are called exchangeable.

❷

Carnapian logic

Using some further assumptions on symmetry and inductive relevance, Carnap derived the rule c_λ:

$$c_\lambda(e_t, q) = \left(\frac{\lambda}{t+\lambda}\right)\frac{1}{2} + \left(\frac{t}{t+\lambda}\right)\frac{t_q}{t}$$

Parameter λ determines how fast predictions change from $1/2$ to the observed relative frequency t_q / t.

❷

Implicit assumptions

In the Carnapian picture, inductive assumptions are part of the logical famework.

$$e_t = 10100001010101010000$$

rule with implicit inductive assumption

predictions $p\,(q \mid e_t) \;=\; c_\lambda(e_t, q)$

Many procedures in classical statistics have the same implicit character.

❷

A logic of induction

The following aims to improve on this framework in two ways.

- The inference rule that brings us from observations to predictions must not be carrying implicit assumptions.

- Inductive logic must provide the tools for choosing the assumptions that underly the predictions, the so-called projectability assumptions.

❸ Bayesian logic

The likelihood principle provides a neutral way of adapting probability assignments to observations.

inductive assumptions

$$e_t = 10100001010101010000$$

likelihood principle

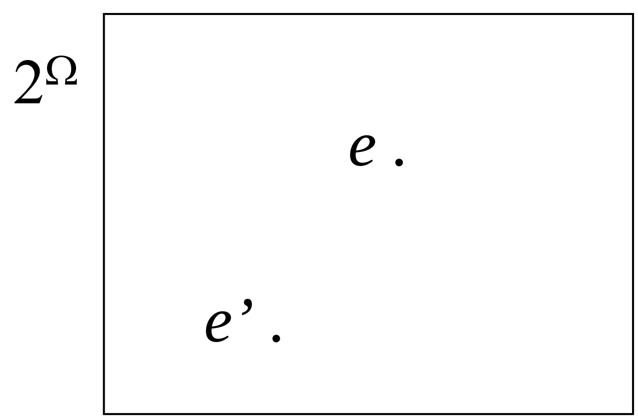
predictions $p\,(q \mid e_t) = c_\lambda(e_t, q)$

Statistical inference can in this way be cast in a logical form. Assumptions can be made explicit.

❸

Observations

Observations are represented as sets of infinite sequences of binary observations, e.g., $e = 101000010...$

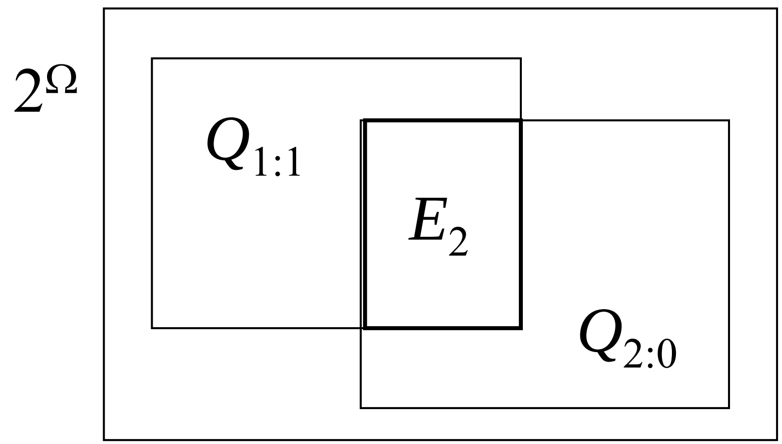


Every point in the rectangle represents a separate infinite sequence.

❸

Cylindrical algebra

The observation of a 0 at time 2 is the set of all those sequences e that have a 0 at position 2, denoted $Q_{2:0}$.

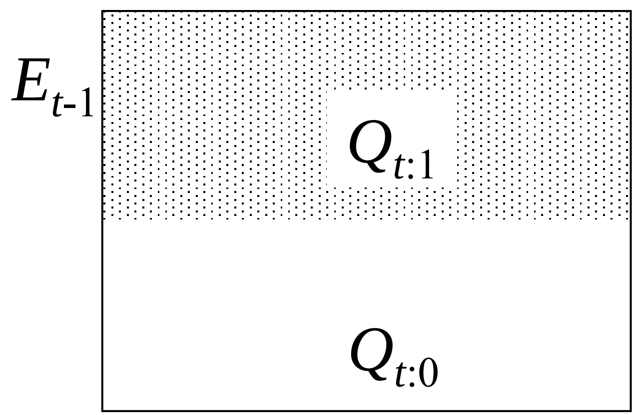


Similarly, the finite sequence $e_2 = 10$ is given by the set $E_2 = Q_{1:1} \cap Q_{2:0}$. The whole space is $E_0 = 2^{\Omega}$.

❸

Predictions

Predictions are determined by a probability function p over the observations. The areas in the diagrams refer to the size of the probability.

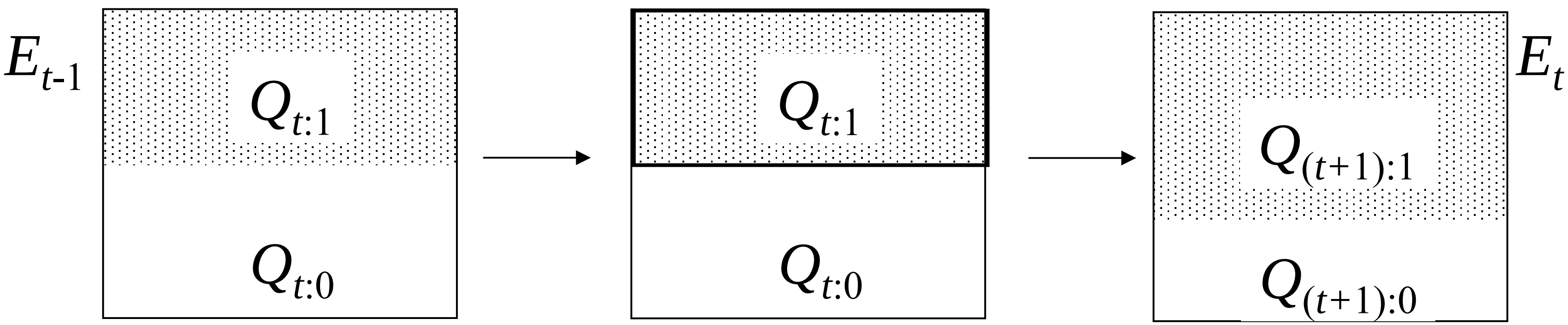


So the above means that $p\ (\ Q_{t:1}\ |\ E_{t\text{-}1}\ ) = p\ (\ Q_{t:0}\ |\ E_{t\text{-}1}\ )$.

❸

Likelihood principle

Conditioning on observation $Q_{t:1}$ is like zooming in on the probability p within the subset $Q_{t:1}$.



For Carnapian prediction rules we must choose the initial probability to conform to $p\,(\,Q_{(t+1):q}\,|\,E_t\,) = c_\lambda(\,e_t\,,\,q\,)$.

❸

Bayesian inference

The inductive assumptions can be expressed in the prior probability.

prior over observations conforming to c_λ

$$e_t = 10100001010101010000$$

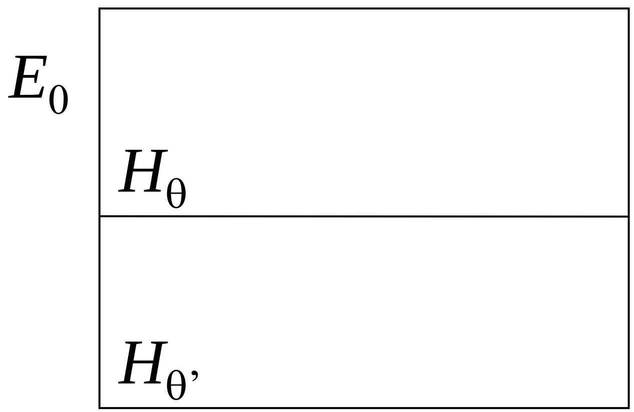
likelihood principle

predictions $p\,(Q_{(t+1):q} \,|\, E_t) \;=\; c_\lambda(e_t\,, q)$

Note that the prior does not provide tools for controling the inductive assumptions.

❹ Statistical hypotheses

Statistical hypotheses offer an alternative way for determining a prior over the observation algebra.



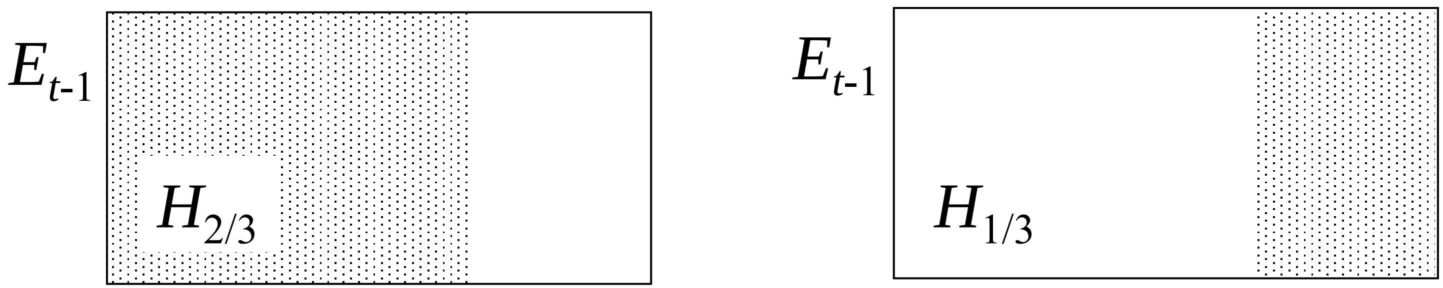
Considerations on statistical models can therefore be captured in the prior probability.

❹

Chance processes

Statistical hypotheses may be associated with chance processes underlying the observations:

$$p(Q_{t:1} \mid H_\theta \cap E_{t\text{-}1}) = \theta$$



This is where statistical inference becomes relevant for underdetermination.

❹

Frequentism

Hypotheses can be seen as elements in an extended observation algebra using the frequentist interpretation.

$$e = 101000010101010...$$

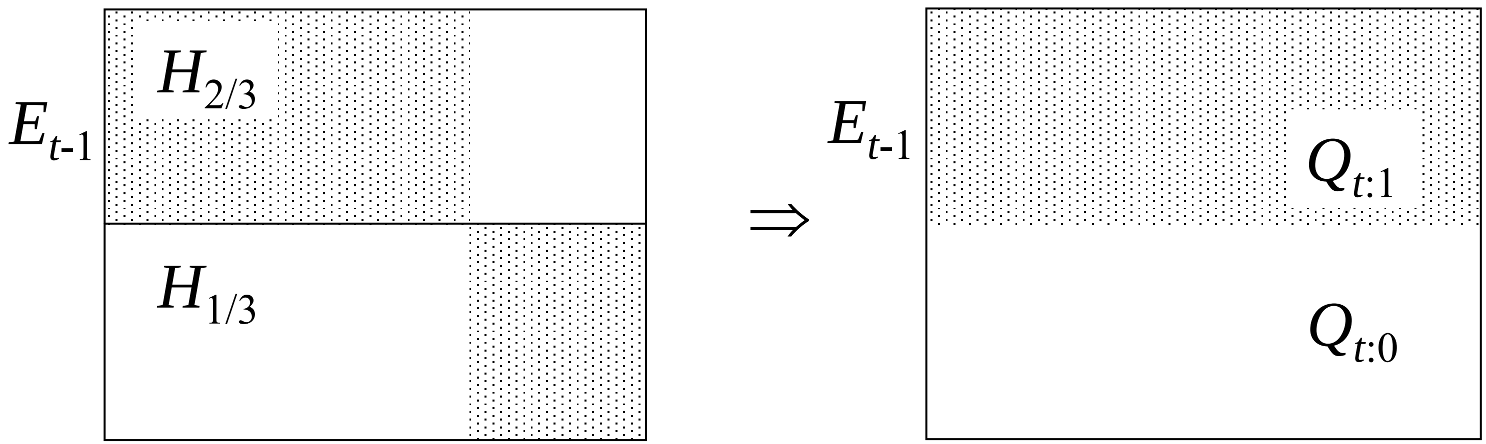
$$H_\theta = \left\{ e : \mathrm{freq}_1(e) = \theta \right\}.$$

The hypotheses are collections of so-called Kollektivs, and as such they are tail events in an extended observation algebra.

❹

Subjective probability

To these hypotheses we can assign a probability that expresses belief.



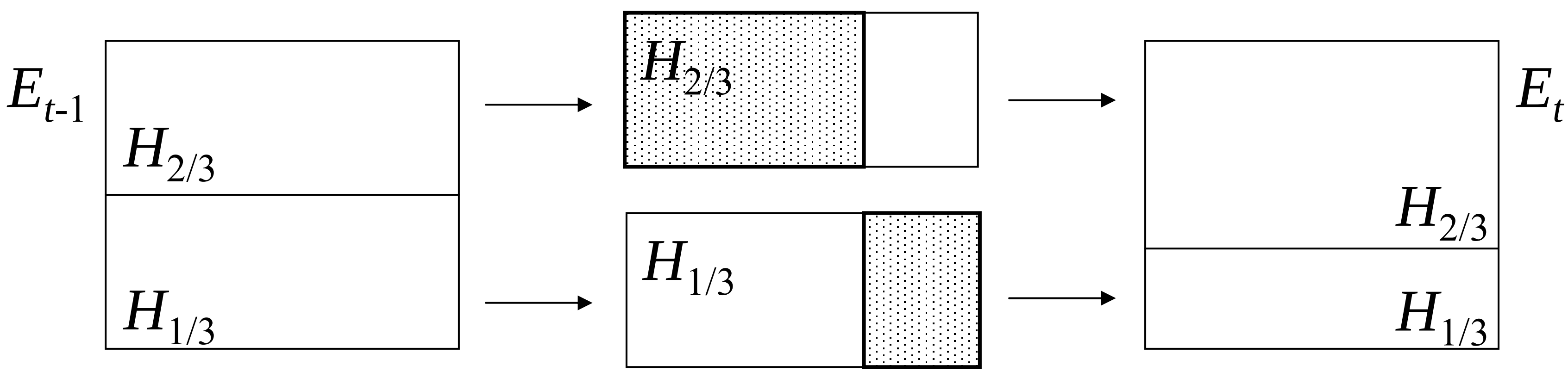
Predictions can be derived by weighing the objectivist likelihoods with these subjective beliefs.

❹

Inference over hypotheses

Conditioning determines changes in the probability over the hypotheses due to accumulating data:

$$p(H_\theta \mid E_{t-1}) \quad \times \quad \frac{p(Q_{t:1} \mid H_\theta \cap E_{t-1})}{p(Q_{t:1} \mid E_{t-1})} \quad = \quad p(H_\theta \mid E_t).$$



❹

Bayesian statistical inference

This completes the following alternative logic of inductive predictions.

hypotheses H_θ associated with chance processes

prior over the hypotheses H_θ

$$e_t = 10100001010101010000$$

likelihood principle

posterior over H_θ, and predictions $p\,(Q_{(t+1):q} \mid E_t)$

❹

Well-calibrated Bayesians

With the accummulation of data, the subjective probability will converge onto the true hypothesis.

$$t \to \infty \quad \Rightarrow \quad \exists\, \theta : \quad p(H_\theta \mid E_t) \to 1 .$$

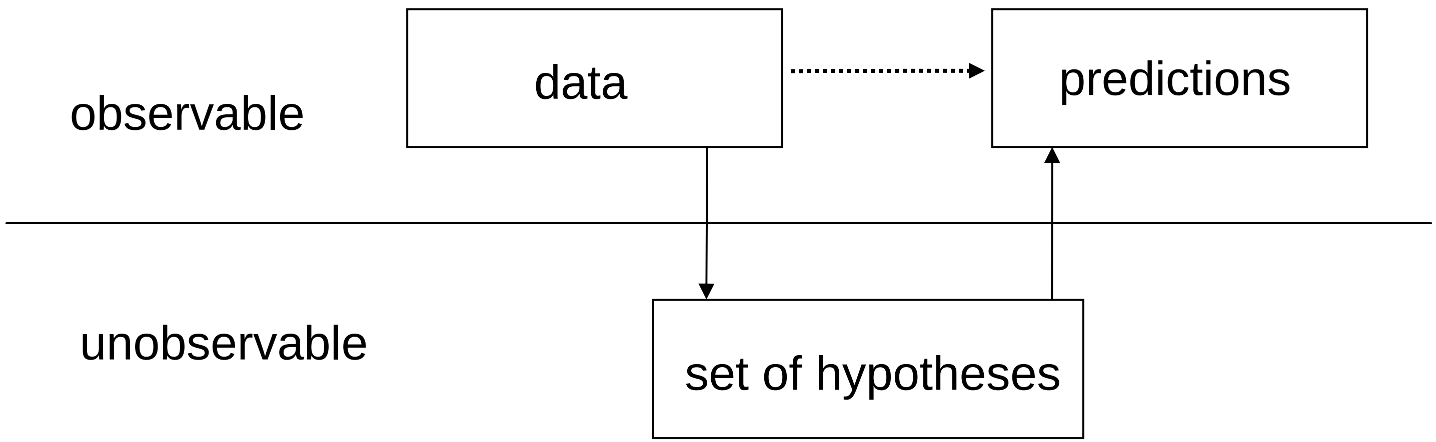
The set of hypotheses thus present an inductive assumption:

$$\bigcup_\theta H_\theta \text{ is true} .$$

This connects well to statistical inference being logical.

❺ Empiricism and subjectivism

In controlling inductive assumptions, hypotheses turn out to be useful tools.



But their use has long been suspect for empiricist and subjectivist reasons.

❺

Representation theorem

Consider the continuum of hypotheses H_θ associated with process with fixed chances θ:

$$p(Q_{(t+1):1} \mid H_\theta \cap E_t) = \theta.$$

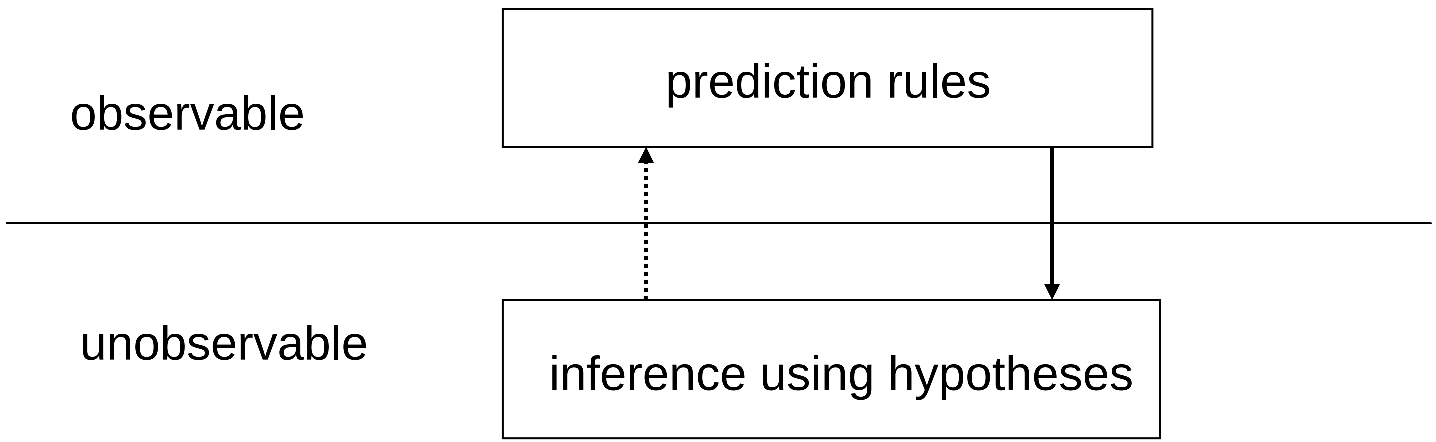
De Finetti proved that the class of priors over this continuum results in the class of exchangeable prediction rules.

$$p(H_\theta)d\theta = f(\theta) \quad \Leftrightarrow \quad p(Q_{t+1:q} \mid E_t) = f(t_q, t).$$

❺

Strict empiricism?

The representation theorem is usually interpreted as a reason for doing away with statistical hypotheses.



Bayesian statistical inference employs the representation theorem in opposite direction.

❺

Using hypotheses

By using hypotheses we can tackle many problems in inductive logic and methodology.

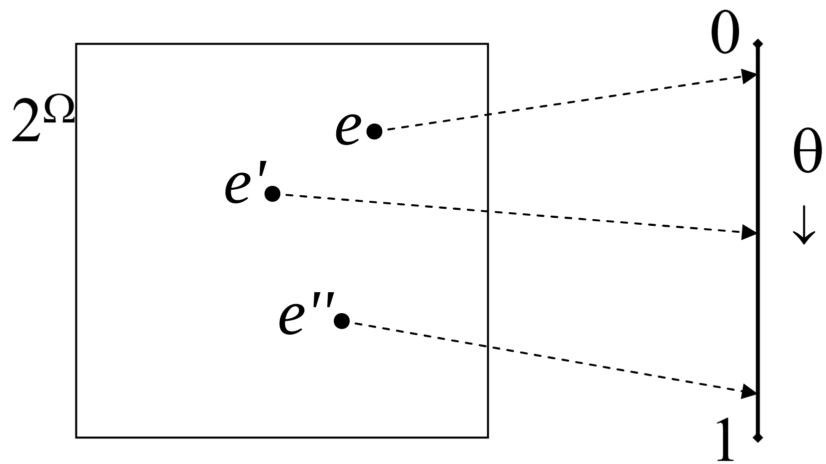
- Analogical reasoning: formalizing inductive relevance between variables.

- Conceptual change: defining a distance function between statistical models.

- Realism debate: the use of distinguishing empirically equivalent models.

❻ Statistical underdetermination

The hypotheses H_θ prescribe different likelihoods θ. Every e belongs in exactly one such hypothesis.

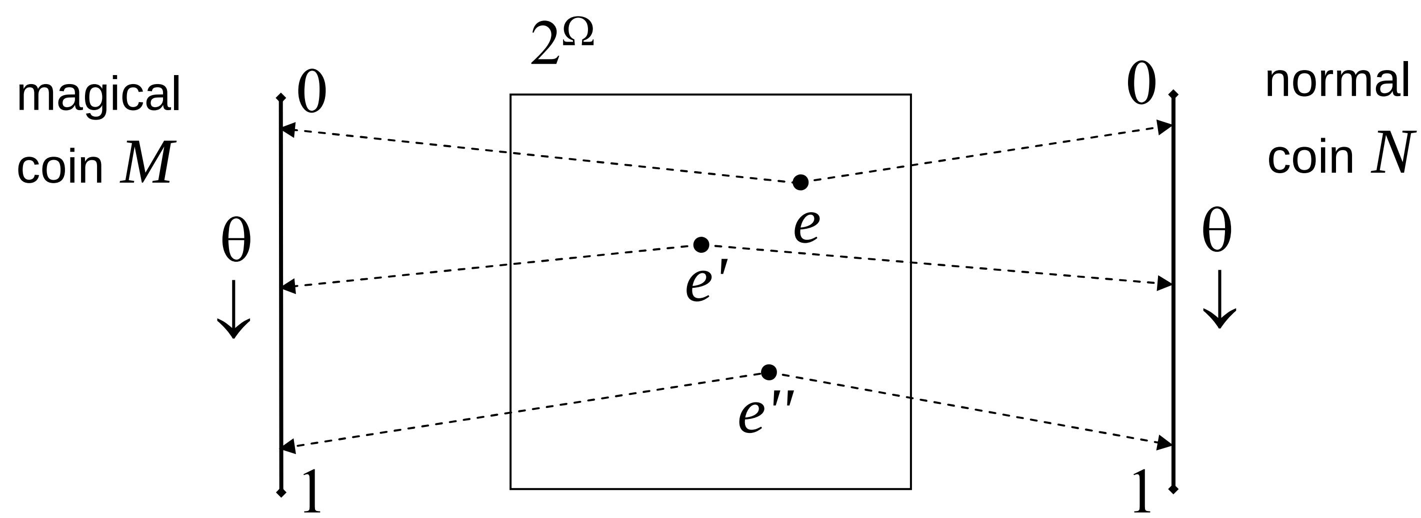


Can there be any use for distinguishing between hypotheses that have the same likelihoods?

❻

Magical coin

Imagine that the observations are coin tosses, but that we are not sure whether the coin is from a wallet or a magic box.

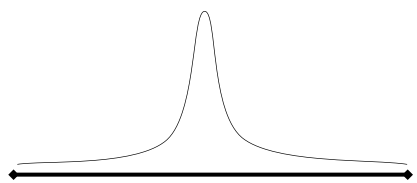


❻

Prior knowledge

We can express the difference between the normal and magical coin in a prior probability over the hypotheses $N \cap H_\theta$ and $M \cap H_\theta$:

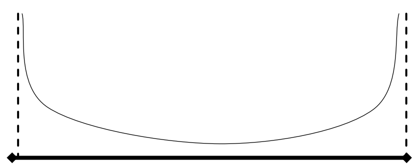
N



$\theta \rightarrow$

$\longrightarrow$ c_λ with $\lambda = 8$

M



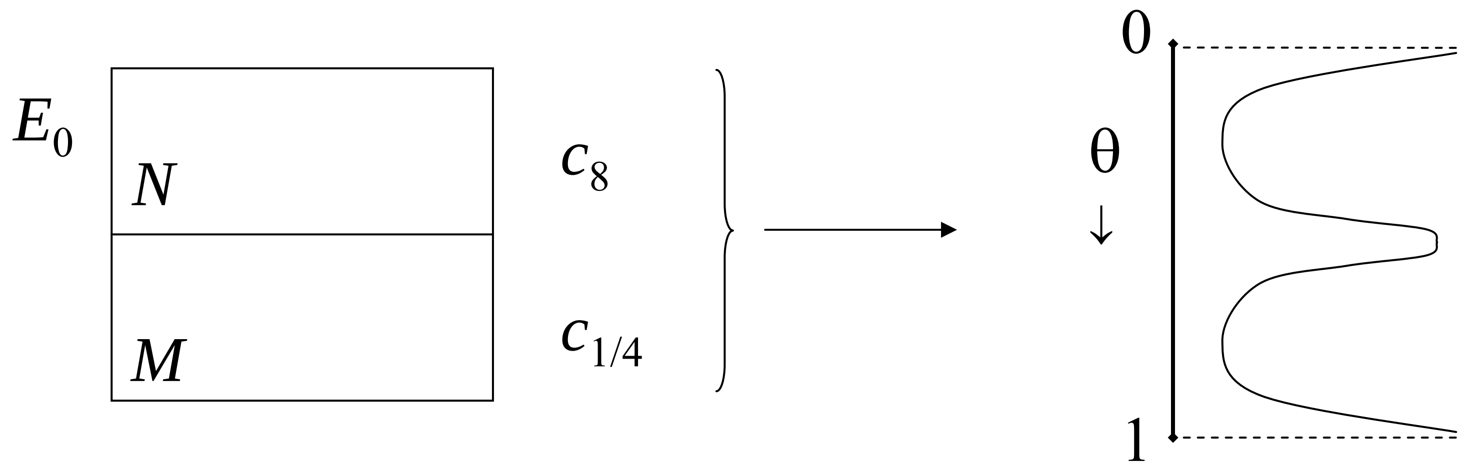
$\theta \rightarrow$

$\longrightarrow$ c_λ with $\lambda = 1/4$

❻

The use of underdetermination

There is a prior over a single partition into H_θ that generates the same predictions.

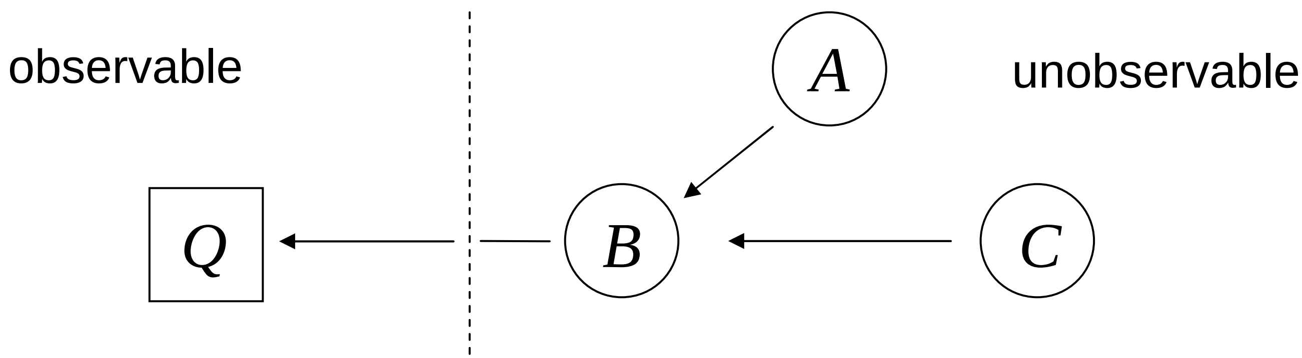


But dealing with N and M separately allows for an easier integration of prior information and inferences.

❻

Back to science

The more general claim is that much of experimental science employs underdetermined distinctions for exactly these reasons.



There are pragmatic reasons for using latent structures.

❼ Abducted by Bayesians?

Combining the identical partitions with differing prior probabilities results in a mixture of Carnapian rules:

$$p(Q_{t+1:q} \mid E_t) =$$
$$p(N \mid E_t)\, c_8(e_t, q) + p(M \mid E_t)\, c_{1/4}(e_t, q)$$

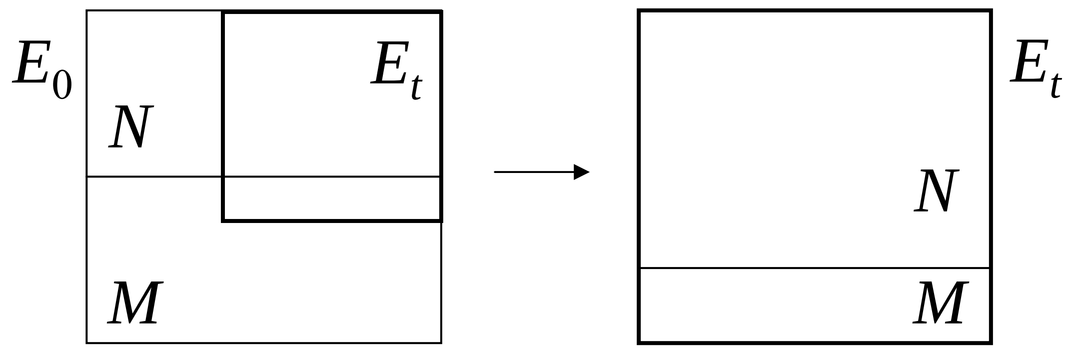
The probability assignment is updated in two parts: first over the two partitions of H_θ separately, and then over the partition into N and M.

❼

As-if confirmation

The Carnapian predictions c_λ function as likelihoods for the hypotheses N and M.

$$e_t = 1011001001$$



It is therefore possible to update over these hypotheses, which nevertheless consist of the very same H_θ.

❼

Observation and theory

The above update for N and M is less magical than it may seem.

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

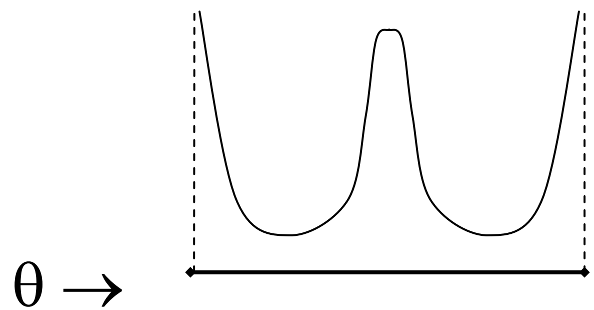
$$e_t = 0000001000010000 \longrightarrow \text{magical}$$

The hypotheses N and M do have different observable content. This difference is simply not expressed in the statistical hypotheses H_θ.

❼

Theoretical distinctions

The distinction between N and M is only theoretical relative to the hypotheses H_θ. Here it serves to motivate and manipulate a prior probability over the H_θ.



The prior probability expresses knowledge on the underlying chance processes that the hypotheses H_θ do not capture.

Conclusion

Conclusions on induction, Bayesian logic and the use of theoretical concepts.

- Inductive logic must be based on a justifiable rule, and reveal the underlying inductive assumptions.

- Bayesianism provides the justified rule, and hypotheses schemes can reveal the assumptions.

- Distinguishing empirically equivalent models may have computational advantages.