CAS seminar

LMU March 2026

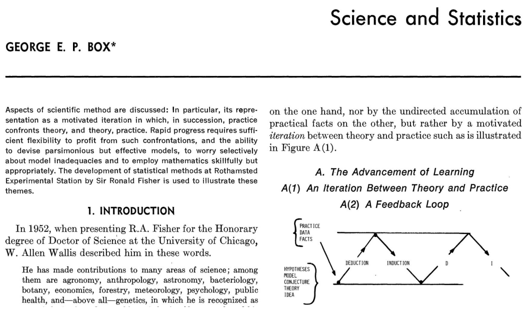
Can a statistical model be true?

Jan-Willem Romeijn

University of Groningen

All models are wrong. . .

In his 1976 paper on scientific method, Box claims that statistical models are never more than approximations.



Science and Statistics

GEORGE E. P. BOX*

Aspects of scientific method are discussed: In particular, its representation as a motivated iteration in which, in succession, practice confronts theory, and theory, practice. Rapid progress requires sufficient flexibility to profit from such confrontations, and the ability to devise parsimonious but effective models, to worry selectively about model inadequacies and to employ mathematics skillfully but appropriately. The development of statistical methods at Rothamsted Experimental Station by Sir Ronald Fisher is used to illustrate these themes.

1. INTRODUCTION

In 1952, when presenting R.A. Fisher for the Honorary degree of Doctor of Science at the University of Chicago, W. Allen Wallis described him in these words.

> He has made contributions to many areas of science; among them are agronomy, anthropology, astronomy, bacteriology, botany, economics, forestry, meteorology, psychology, public health, and—above all—genetics, in which he is recognized as

on the one hand, nor by the undirected accumulation of practical facts on the other, but rather by a motivated *iteration* between theory and practice such as is illustrated in Figure A(1).

A. The Advancement of Learning

A(1) An Iteration Between Theory and Practice

A(2) A Feedback Loop

The paper alludes to the idealisation or deliberate falsehood of statistical models. So models are not factive.

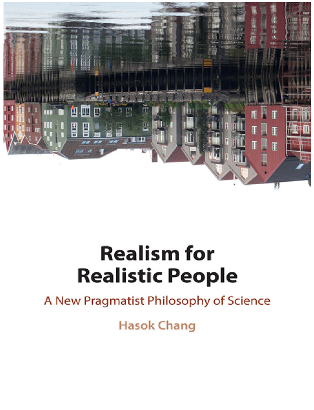
Truth-aptness of models

Can we interpret models as propositions capable of being true? I develop this idea and clarify some debates in statistical methodology with it.

- Presenting statistical models as propositions allows us to interpret probability assignments over models.
- We can deploy the formal semantics of statistical models to spell out Bayesian statistics as a specimen of logic.
- Ideas from this semantics can be used to specify when statistical models are true, in the sense of corresponding to the world.

“Probability does not exist”

We need not endorse a fully developed metaphysics of chance to reap these conceptual benefits.

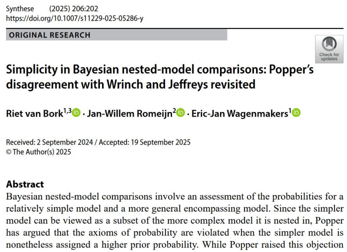


The truth of statistical models ultimately resides in their usefulness.

Contents

1 Popper’s problem

In a recent paper, van Bork et al. review the responses to Popper’s problem on statistical model evaluation.


Synthese (2025) 206:202
https://doi.org/10.1007/s11229-025-05286-y

ORIGINAL RESEARCH

Check for updates

Simplicity in Bayesian nested-model comparisons: Popper’s disagreement with Wrinch and Jeffreys revisited

Riet van Bork[1,3] · Jan-Willem Romeijn[2] · Eric-Jan Wagenmakers[1]

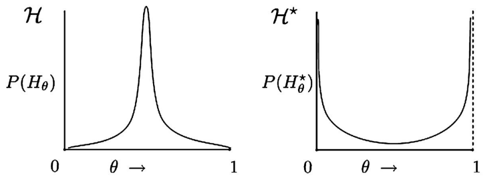
Received: 2 September 2024 / Accepted: 19 September 2025

Abstract
Bayesian nested-model comparisons involve an assessment of the probabilities for a relatively simple model and a more general encompassing model. Since the simpler model can be viewed as a subset of the more complex model it is nested in, Popper has argued that the axioms of probability are violated when the simpler model is nonetheless assigned a higher prior probability. While Popper raised this objection

Can we prefer a nested statistical model over the statistical encompassing it? This seems to violate basic probability axioms.

A mereological response

Mathematical sets come for free. The probabilistic incoherence can be avoided by a simple model duplication.



The problem is interpretative: can we view such models as expressing different matters of fact?

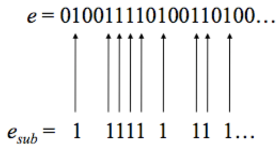
Coin factories

We can make intuitive sense of such a duplicated statistical model if the priors themselves are based on frequencies.

But what if we are tossing a single, unique coin? Can we make sense of the truth of an expectation?

2 Frequentist Bayesianism

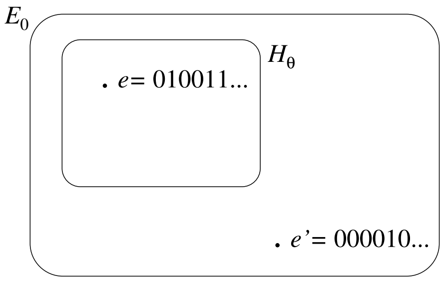
Following ideas of von Mises, we can represent statistical hypotheses as tail events in a σ-algebra of observation sequences.



In their seminal 1986 paper, Gaifman and Snir develop this idea into a “rich language” that includes chance propositions.

Probabilities over chance propositions

By Carathéodory's extension theorem, a probability assignment over observation sequences is automatically an assignment over statistical hypotheses, $P(H_\theta)$.



After adopting the above semantics for hypotheses, we can claim: if you are a frequentist, you are a Bayesian too!

3 Conceptual benefits

Using this semantics, we can draw a parallel between logical completeness and the convergence theorems for Bayesian inference.

- A logic is complete precisely when, if a proposition is true, it can be proven by the inference rules of the logic.
- The convergence results show that, if a statistical hypothesis in the model is true, its posterior probability will converge to 1.

Notice that the theorems do not establish soundness, and that the observations have to “separate” the model.

De Finetti's representation theorem

Also, assuming the formal semantics, the representation theorems of De Finetti can be understood as reduction axioms.

- In logic a reduction axiom relates a complex expression to composites of more basic expressions.
- The representation theorems show that there is a correspondence between direct prediction rules and predictions from a statistical model.

Statistical hypotheses are theoretical terms that allow us to express inductive assumptions more economically.

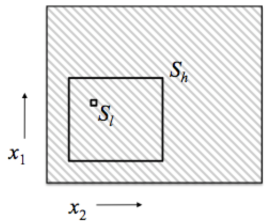
4 Real chances?

We have located chances within the epistemic domain, as part of a formal semantics for statistical inference.

What about chances as pertaining to the world out there? Can we give chance a realist interpretation?

Deterministic systems

A central problem for physical probability is the reference class problem: different descriptions of an event lead to different chances.



Chances thus become perspectival, and ultimately epistemic: they result from a lack of information on deterministic states.

Arbitrary functions

We can give an objective motivation for a certain perspective, relying on Hopf's method of arbitrary functions.

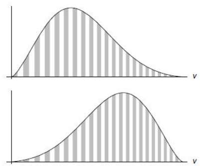
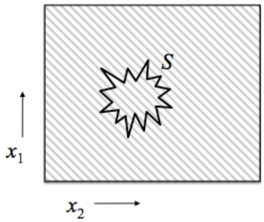


Figure 2: Different spin speed distributions induce the same probability for *red*, equal to the strike ratio for *red* of one-half

This idea can be developed further using ideas from the frequentist semantics of statistical hypotheses.

Radical multiple realizability

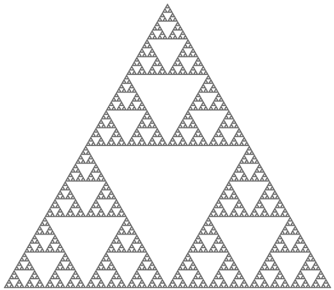
The key idea is a radical form of multiple realizability: some propositions principally resist translation to a different theoretical level.



More precisely, conditioning on micro-level sets cannot alter the macro-level probability of a proposition.

Relation to fractals?

The radical multiple realizability rests on the infinite intricacy of the event S in relation to the refinements of $\mathcal{X}$.



Fractals provide an illustration of such infinitely intricate sets. Notice that such sets can be constructed through a very simple mechanical procedure.

5 Randomness and robustness

An event S is *random* relative to some algebra $\mathcal{X}$ iff for all refinements X_i from this algebra we have

$$P(S|R \cap X_i) = P(S|R).$$

We call the resulting probability assignment *robust* (cf. Lewis and Skyrms): adding further information from $\mathcal{X}$ to a reference class R does not refine or alter the chances for S.

Using ideas from frequentism

We can provide a formal underpinning of robustness by relying on ideas from frequentism. Recall the relative frequency of a place selection x_i on sequence s:

$$F(s;x_i) = \lim_{n\to\infty} \frac{\Sigma_{t=1}^{n} s(t)x_i(t)}{\Sigma_{t=1}^{n} x_i(t)}.$$

The sequence s is random relative to selection rules $\mathcal{X}$ if for all $x_i \in \mathcal{X}$ we have:

$$F(s) = F(s;x_i).$$

We might say that the set of sequences $\mathcal{X}$ contains “no information” about the original sequence s.

Autonomous chances

The ultimate version of robustness has the event S distributed uniformly over the space X. Consider the σ-algebra $\mathcal{X}$ generated by

$$B[x, \delta] = \{x' : x' \in (x, x + \delta)\}$$

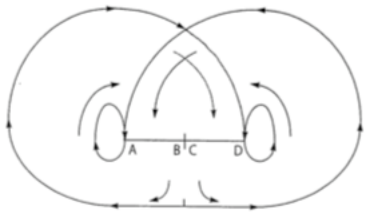
where the δ can be arbitrarily small, and stipulate that

$$P(S|B[x, \delta]) = P(S).$$

That is, no amount of fine-graining will offer additional information on the event S, making the chances fully robust or *autonomous*.

Fully random set: example

We can construct such a random set by employing Von Mises' notion of place selections in a simple ergodic system. Consider a set R consisting of elements $R_t = 2^t R(0) \bmod 1$,



where $R(0) \in [0,1]$ is the starting position generating the set. For a given initial state $R(0)$, we label all the points $R(t)$ by 0 if $R(t-1) \leq 1/2$ and by 1 if $R(t-1) > 1/2$. The set of points labelled 1 we call S.

Autonomous chance: example

Assuming that $R(0)$ is a collective with a relative frequency σ, the limiting relative frequency of points of S within R is

$$P(S|R) = \sigma.$$

The set R is dense everywhere in $[0, 1]$. Moreover, for any interval $X_i \subset [0, 1]$, however small, we have

$$P(S|R \cap X_i) = P(S|R).$$

But every point x in the space X, or in the class R, is or is not a member of S so we do not violate determinism.

True chance propositions

We call a chance proposition $P(S|R)$ true if it is autonomous in the above sense. Notice:

- There are many variations on the randomness of the event S depending on the details of the algebra $\mathcal{X}$. This is formally equivalent to the randomness of sequences in Von Mises' Kollektivs.
- It seems that a natural line is drawn by the algebra $\mathcal{X}_{\text{Schnorr}}$ corresponding to "Schnorr randomness": whether or not a point x is a member of an element $X \in \mathcal{X}$ must be computable.
- This allows that the random events S and R are constructed by an effective procedure.

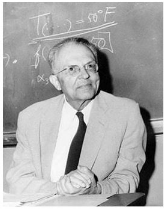
6 True models

A statistical model is true if one of its hypotheses describes chances that are sufficiently robust for the target system.

- This explains what the aim of statistical model selection is: to find the partition $\mathcal{R}$ for which the $P(S|R)$ are autonomous.
- Nothing guarantees that random events, like S in R, actually obtain. The foregoing offers an extreme case of autonomous chance but the reality of chances may fall short of this.
- The autonomy of chance is relative to the algebra $\mathcal{X}$, which is contextually determined.

Usefulness

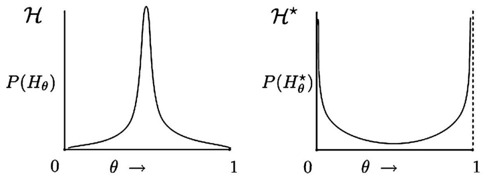
Recall that, by De Finetti's representation theorem, statistical hypotheses offer a useful access to predictions of observations.


Eventually the model is true if the predictions generated by it converge towards robust and optimal probabilities.

Semantics for hierarchical Bayes?

It seems that we cannot claim truth-aptness for hierarchical statistical models in the sense of correspondence.



However, we can still claim they are truth-apt in a pragmatic sense, insofar as they assist in making optimal predictions.

Thanks for your attention

Slides of the talk will be available at romeijn.web.rug.nl. For comments and questions, email j.w.romeijn@rug.nl.