LRP Discussion of Leo Held OBAYES 2022

Luis Raúl Pericchi Guerra
University of Puerto Rico Rio Piedras

OBAYES 2022, September 07, 2022

Debut and Farewell

"Law of Initial Results: *So often early promising results are followed by others that are less impressive. It is almost as if there is a law that states that first results are always spectacular, and subsequent ones are mediocre: The Law of Initial Results."* (Jeffreys (1961), Ioannides (2005))

Reverse Bayes and the Bayesian Principle

Reverse Bayes has been found useful in elicitation (even with only one study) but also in basic Bayesian judgement:
Berger and Pericchi (2001):`Principle:`
Testing and model selection methods should correspond... to actual Bayes factors, arising from reasonable default prior distributions
"One of the primary reasons that we are Bayesians is that we believe that the best **discriminator between procedures** *is study of the prior distribution giving rise to the procedures. Insights obtained from studying the overall properties of the procedures (eg consistency) are cruder (at least in parametric problems)". This principle opens the door for a Reverse Bayes handicap.*

Conjugate warning

Remember the archetypal anti "Bayesian" criticism: "With Bayes, You can get the results you want, by changing Your prior!", should read instead: "With Conjugate Bayes, You can get the results you want, by changing Your Conjugate prior".
Replication needs a very stringent assumption.

Influence of Priors

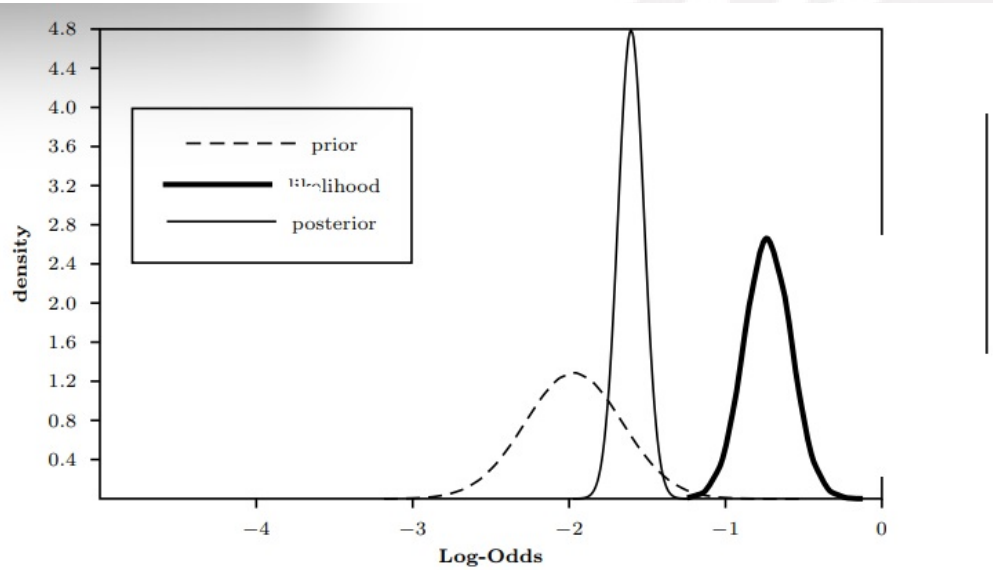

Figure 9: Prior(Finland), likelihood(Venezuela) and posterior distributions in the Bayesian analysis of a trial of the Rhesus Rotavirus-Based Quadrivalent Vaccine for the N/N model.

Figure: Normal Prior

Robust Cauchy Prior

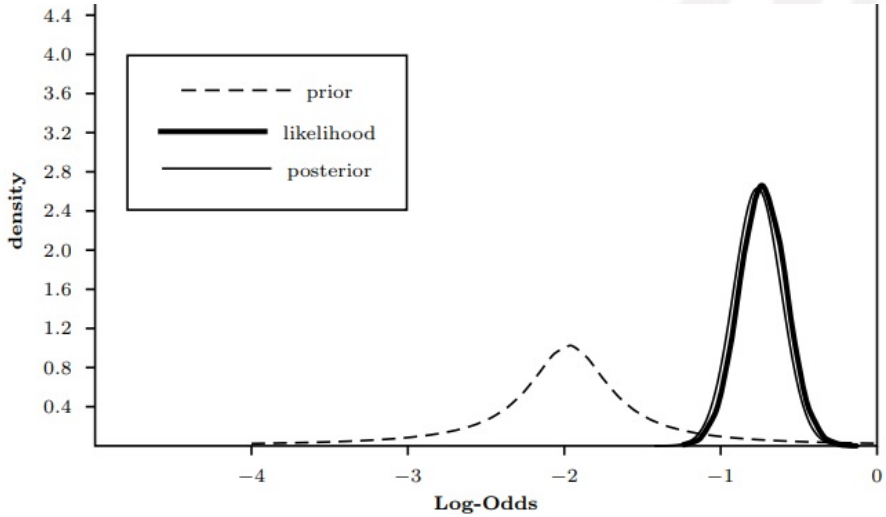

Figure 10: Prior(Finland), likelihood(Venezuela) and posterior distributions in the Bayesian analysis of a trial of the Rhesus Rotavirus-Based Quadrivalent Vaccine for the C/N model.

Figure: Cauchy Prior

Care with p-values: Effect of Sample Sizes on p-values: Freeman (1993)

		Sample Sizes Versus Effect Sizes	
Number of Patients	Numbers preferring A to B	percent preferring A	Two Sided p-value
20	15:5	75	0.04
200	115:86	57.50	0.04
2000	1046:954	52.30	0.04
2000000	1001445: 998555	50.07	0.04

Table: Sample Sizes and Effect Sizes, for constant p-values

Posterior probabilities of the Null will be skyrocketing, with fixed p-value and growing sample size.

Spiegelhalter et al (2004): The Role of "Scepticism" in confirmatory studies

Use Sceptical priors dealing with "regression to the mean", on which early extreme results tend to return to average over time.
""Bayesian Approaches to CLinical Trials and Health Care Evaluation" (2004) Spiegelhalter, Abrams and Myles.

Scepticism about the Sceptical p-value

Comparison with "Two Trials Rule":

- SIMPLICITY
- Ubiquity of p-values
- BUT they are not the probabilities we need, multiplying them as probabilities may compound the confusion.
- **Naive Question: What about "Two Trials Posterior Probability Rule" (using Bayes Factors and Prior Probabilities), at least as a first step?**

The (transformed) Sceptical p-value

A lot of ingenuity has come into the development of sceptical p-value like "re-calibration by the Golden Ratio" etc., etc. ("p-values are too familiar to ditch!", Spiegelhalter (2018))
Good in situations of high publication bias (low p-values, perhaps).

Newer more developed approaches

I commented to the original Held (2019): "The interesting procedures of Matthews and of Held, are based on intervals and not on the probability of hypotheses. If it is desired to stick with intervals in order to achieve consistency it is necessary to make thresholds decrease with the sample size, as in Pérez and Pericchi (2014). Let me propose a general alternative and based on probabilities of hypothesis and the corresponding Bayes Factors."
Thus I prefer approaches based on Bayes Factors.
"The need to evaluate signal size, and not only p-values: Peter Diggle, cited".

The Sceptical Bayes Factor for the assessment of replication success: Pawel and Held (2022)

This needs further study but looks a default promising starting point.
Very useful extensions to non-normal models.
Assessment of prior-data conflict and Advocacy prior, attractive.
Depends on compatibility Q among studies, **serendipity**.

General Comments

"Compared to other methods, the sceptical Bayes factor poses more stringent requirements but also allows for stronger statements about replication success. It ensures that both studies provide sufficient evidence against a null effect, while also penalising incompatibility of their effect estimates".
On the other hand the author avoids a single method and propose a Framework.
I think it needs experiment. But I rather prefer methods based on Probabilities of Hypothesis.

Epilogue: Are you a Bayesian?Are you a Frequentist?

Answer to both:
Depends! The are very different kinds of Frequentists!
Interestingly in his De-Finetti Lecture, Jim Berger classifies (at least) 4 kinds of Frequentists and advocates for:
The Empirical Frequentist. That may be thought of a well calibrated scientist on which we all can converge.
(The Bayesian objective-empirical route seems the safer one, even if it is not the shortest).