Invited Discussion of J.O. Berger: Four Types of Frequentism and Their Interplay with Bayesianism[☆]

Luis PERICCHI

1. A CONVERGENCE OF THE SCHOOLS OF STATISTICS?

One of the merits of this far reaching article is to show that not all "Frequentisms" are equal. Furthermore that there are frequentist approaches which are compelling scientifically, notably the "Empirical Frequentist" (EP), which can be paraphrased as *"The proof of the pudding is in the eating"*. Somewhat surprisingly to some (but anticipated in Wald's admissibility Theorems in Decision Theory), is the conclusion that the easiest and best way to achieve the EP property is through Bayesian reasoning, perhaps more exactly, through Objective Bayesian reasoning. (I am avoiding the expression Empirical Bayesian reasoning which would be appropriate if it wasn't associated with a very particular group of methods. It is argued below that a better name would be "Bayes Empirical") I concentrate on Hypothesis Testing since that is the most challenging area of deeper disagreement among schools.

From this substantive classification of Frequentisms, emerges the opportunity for a convergence, which is even more satisfying than a compromise, between schools. This may only be fully achieved if the prior probabilities are known, which is not usually the case. However, particularly in Hypothesis Testing, prior probabilities can and should be estimated and its uncertainty acknowledged in a Bayesian way. This may be termed perhaps, Bayes Empirical: The systematic empirical study of Prior Possibilities based on relevant data, acknowledging its uncertainty.

1.1 A General Standard for Most (If Not All) of Statistics

A striking and enlightening bold affirmation in the paper, that will be remembered is:

> The empirical frequentist principle seems compelling to most statisticians

Jim Berger, De Finetti's Lecture, ISBA 2021, and 2022 (this article)

1.1.1 We focus on Hypothesis Testing

In this respect, perhaps the best indication of the crisis of the bad versions of frequentism, is the reaction against the

practitioners upside-down interpretation of p-values, what is called the prosecutor's fallacy, in this case taking a p-value as the probability of the Null.

One of the important messages of the paper: it is NOT empirical frequentist the correct interpretation of p-values. It is stated in the article: *"reporting the p-value as the error probability is terrible according to the empirical frequentist principle, it is reasonable only when the (unknown) π_0 is very small"*. This phrase leads us to two major conclusions:

i) p-values needs calibration from an empirical point of view and ii) the paramount importance of prior probabilities to all schools of statistics.

1.2 The World of Statistics Is Changing Regarding Significance Testing, Why?

The timing of the growing awareness about the flaws of current practices in Significance Testing suggest: The change in attitudes is not due to the Mathematics: Neither because very good mathematical reasons as

i) Completeness Wald's Theorems of Decision Theory.
ii) Obedience of Likelihood Principle.
iii) Conditional (superior) inference.
iv) Not even Stein's paradox, etc.

It is because of the Science: "Why most published research findings are false?", the famous **Empirical Frequentist** apothegm coined by Ioannidis (2005) [1].

Next we insist in a theme of the foremost importance, which can be derived from the paper: Prior Probabilities are at least as important as Power.

I denote as False Discovery Rate (fdr) equation (10) of the article, with known prior probabilities π_0, and Type I Error α and Power β:

$$fdr = \frac{\pi_0 \cdot \alpha}{\pi_0 \cdot \alpha + (1 - \pi_0) \cdot \beta},$$

from which we may construct a table changing priors and power to check their influence on fdr (Table 1).

The conclusion is: fdr is more sensitive about π_0 than to β. **Science needs Baselines or "prevalences" as they are called in epidemiology, i.e. prior probabilities of hypothesis**. In the following figure, I graph fdr versus prior

Table 1. Prior π_0, Power β and False Discovery Rate (fdr), for $\alpha = 0.05$.

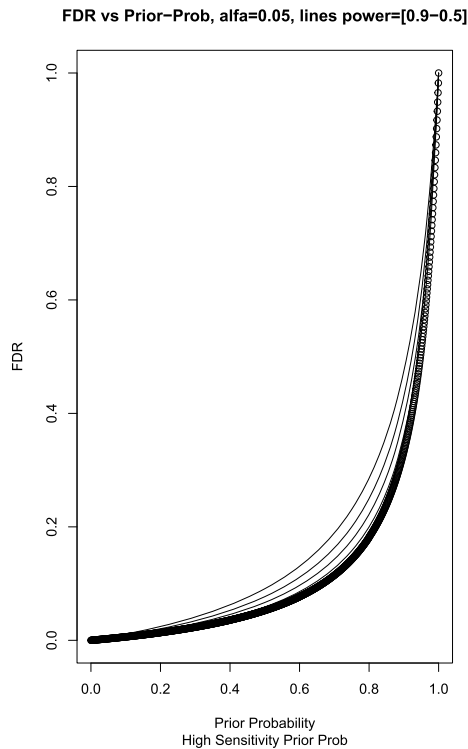| π_0	β	fdr
0.9	0.9	0.333
0.9	0.5	0.47
0.8	0.8	0.2
0.5	0.9	0.05
0.5	0.5	0.091

Figure 1: fdr versus prior probabilities, for different powers.

probability, for different power lines $(0.5 < \beta < 0.9)$ Power Versus Prior, showing High Sensitivity to the Prior, and less to power. This suggest a change of emphasis in Statistics to achieve the Empirical Frequentist synthesis.Admittedly, this strong conclusion is based on just a few examples, but the reasoning of the article seems compelling.

1.3 What If the Prior Probabilities Are Unknown? Then fdr Is a Random Variable Based on a Random Prior Probability Based on Surveys

Bayes empirical approach should try to acknowledge all sources of variability, including the information on which the prior is based empirically. In the equation (10) quoted above, π_0 is now not assumed precisely known. But we can, and should!, organize a survey. Suppose then, that our knowledge is based on a small survey on which $n = 100$ and $S =$

90, so that $\hat{\pi}_0 = 0.9$, thus if, the initial prior for π_0 is say Jeffreys prior, then the posterior of π_0 is $Beta(\pi_0|S + 1/2, n - S + 1/2)$.

Now fdr is a random variable, and it may have large dispersion when the variability on priors is acknowledged.
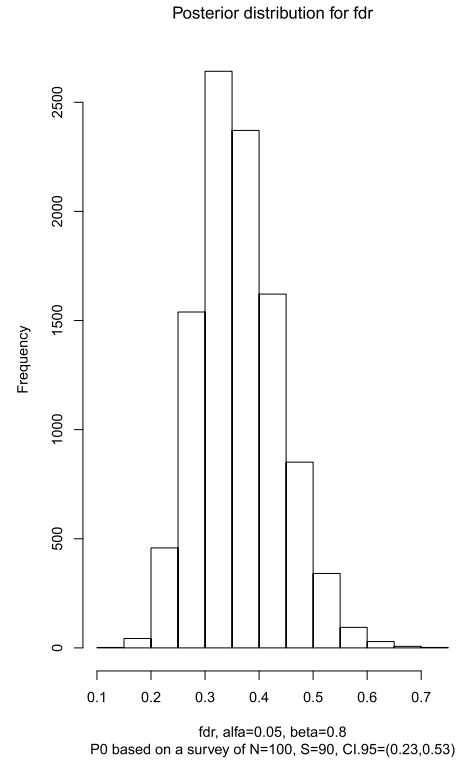

Figure 2: Histogram of fdr as a function of random prior probabilities.

If the "priors are unknown" (and usually they are) is not the end of Bayes, on the contrary we model the priors as random variables and estimate its distribution, Bayesianly. In doing so we respect the variability of the Information about the Prior, see for example Mossman and Berger (2001) [3].

2. MAKING THE P-VALUE, LOWER BOUND CLOSER TO A BAYES FACTOR

Shall we forget p-values? *"p-values are just too familiar and useful to ditch"* David Spiegelhalter (2017) [5].

The paper study the well known calibration of p-values:

$$LowerBound(p) = -e \cdot p \cdot log(p),$$

which has the advantage of depending only on p, but "as a lower bound lack strict empirical frequentist justification", as stated in the article. Another problem of the bound is that it does not change with n. But we may invoke the Bayes Factor as a function of a p-value.

The $LowerBound(p)$ can be simply modified to approximate a Bayes Factor as (Pericchi and Perez (2017) [4])

$$ABF(p,n) = LowerBound(p) \cdot \sqrt{\frac{2\pi n}{e^2 \cdot (\chi^2(1-p) + log(n))}}.$$

This modification has been published in [6].

3. IS THIS PAPER SHOWING THE FUTURE OF STATISTICS AS AN UNIFIED FIELD? I HOPE SO!

A field like statistics divided in fighting schools is not good. The present paper, presents implicitly a rout for convergence of schools. This contrasts with other predictions that anticipated a fully Bayesian world:

> *THE FUTURE OF STATISTICS- A BAYESIAN 21ST CENTURY. "It had originally been my intention to follow Orwell and use 1984 in the title, but de Finetti (1974) suggests 2020".*

> Dennis Lindley, "Advances in Applied Probability",
> (1975) [2].

3.1 Conclusion

> *Perhaps, instead, the direction of growth of Statistics for the rest of 21st century,*
> *will be Bayesian... and Empirical Frequentist.*

REFERENCES

[1] IOANNIDIS, J. P. (2005). Why most published research findings are false. *PLoS medicine* **2**(8) 124. https://doi.org/10.1080/09332480.2005.10722754. MR2216666

[2] LINDLEY, D. (1975). Advances in Applied Probability. *Supplement: Proceedings of the Conference on Directions for Mathematical Statistics.* **7** 106–115.

[3] MOSSMAN, D. and BERGER, J. O. (2001). Intervals for posttest probabilities: a comparison of 5 methods. *Medical Decision Making* **21**(6) 498–507.

[4] PERICCHI, L. R. and PEREZ, M. Q. E. (2017). Converting P-Values in Adaptive Robust Lower Bounds of Posterior Probabilities to increase the reproducible Scientific "Findings". *arXiv preprint arXiv:1711.06219*.

[5] SPIEGELHALTER, D. (2017). Too familiar to ditch. *Significance* **14**(2) 41.

[6] VELEZ RAMOS, D., GUERRA, L. R. and PEREZ HERNANDEZ, M. E. (2023). From *p*-Values to Posterior Probabilities of Null Hypotheses. *Entropy* **25**(4) 618. https://doi.org/10.3390/e25040618.

Luis Pericchi. Department of Mathematics, University of Puerto Rico Rio Piedras, Puerto Rico.
E-mail address: luarpr@gmail.com