'Stargazing' and *p*-hacking behaviours in social sciences: some insights from a developing country

The persistency of "stargazing," *p*-hacking and HARKing[4] are among the main causes for the severity of the irreproducibility problem in social sciences.[5] Facing this problem, Benjamin *et al.* argued for the redefinition of statistical significance: the *p*-value threshold is lowered to 0.005 "for claims of new discoveries".[1] Their proposal almost immediately meets objections from Amrhein & Greenland (2017) who demonstrated the high risk of overconfidence in mathematical results and underestimation of uncertainties and insisted on discarding the oversimplified reasoning based on judging "significance" with *p*-value.[2] However the debate unfolds, the key lies in the ability and courage of researchers to exercise intellectually honest judgement under uncertainty.

Bayesian statistics seems to offer a solution.[4,6,7] Social sciences where decisions are frequently made in spite of uncertainty may be in a good position to benefit from it.[6] Moreover, the ability to update plausibilities with new evidence reflects a key strength of "mathematics on top of common sense" the Bayesian approach represents,[8] especially in social sciences where mixing inconsistent research philosophies does not promise a solution to research irreproducibility. There has also been an abundance of evidence suggesting that refining the inferences, together with its being explicit about estimation biases, is one key strength of the Bayesian approach.[9]

These appeals make transition toward a more widespread application of Bayesian statistics sound inexorable, especially, when the main obstacle to adoption of Bayesian techniques in the past, that is, the computational costs, has now been basically wiped out by the increasing power of computers[8] and the availability of free and open software such as R, Stan, and JAGS.[6,7] Yet …

The persistence of *p*-hacking is quite strong in social sciences, where the 0.1 alpha level has still been considered acceptable in many sub-disciplines, including management and public administration. The persistent behaviours of social sciences researchers in "stargazing" may be attributed to what Scales & Snieder (1997) call the complicated justification of the mathematical representation of data,[8] which gives rise to the application of the principle of max entropy.[6,7] People tend to see "messy statistics" as a major obstacle to "pragmatic solutions" to concrete.[8] We offer some insight from interviews with five experienced authors in Vietnamese social sciences.

The fear of mathematical challenges

Even if persuaded that "you will not need to use any of this mathematics for applied data analysis..." most still view the Bayesian approach as blocked by math, ironically the genuine power of the method. Even people with statistical training may still incorrectly interpret p=.05 to mean that there is a 95% chance that the null hypothesis is wrong, and the alternative hypothesis correct.[9]

Albeit being experienced researchers, they all estimate a need for attending an intensive course on technical matters of the Bayesian method to help them overcome the challenges, running from 20 to 50 hours, roughly a semester or two.

The fear of computer code writing

A long-standing issue with social science researchers in adopting Bayesian statistics is the absence of plug-and-play software packages of the kind that have made frequentist approaches so easy to apply.[9]

With today's availability of open statistical programming language like R with versions running on Linux, Windows, Mac OS, and Bayesian computing engines such as BUGS, JAGS, Stan, operating on not just R, but also Mathematica, Matlab, the writing of computer codes per se should not be the problem.[6,7] However, getting over the fear itself is. The fear is further exacerbated with the fact that Fisherian methods have offered such practical advantages as ease of use, preinferential aspects of model building, and allowing for a division of labor in seeking solutions for complicated problems.[11]

Convenience of frequentist software packages and ready-made solutions available for numerous research problems may have been a major part of the reason for the deterioration of computer literacy, in terms of programming skills, and more importantly of declining willingness to learn new languages for the sake of science.

The fear of leaving their comfort zone

Malakoff (1999) is perhaps correct seeing that frequentist techniques bring about an advantage to researchers in social sciences: "They proved relatively easy to apply to real-world problems—unlike Bayesian methods".[5] At the end of the tunnel, there is a magic wand, *p*-value. Now, with the magic wand of *p*-value coupled with fast computing speed, the "stargazing" practice has become the main sports for those most capable Vietnamese social scientists, to the extent that these researchers even joke that they are of a "regression monkey" species.

The comfort zone, accommodating all sorts of frequentist habits and procedures, has thus been strongly protected and guarded against any possible intrusions of "alien concepts" such as Bayesianism. When asked if they have an interest in doing Bayesian analysis in the next 12 months, all interviewees respond that they do not have any plan for trying the Bayesian statistics in any foreseeable future. It is noteworthy that for this very "comfort," a highly accomplished (frequentist) researcher like Harrell has converted himself to Bayesian after a long-term struggle with the difficulty and discomfort of explaining the *p*-value and confidence intervals.[10]

First, open community dialogues in Bayesian analysis can be established for better science.[5] When surveying experienced researchers with accomplishments, we felt like dealing with a taboo. In fact, until researchers realize that outcome-based selective reporting is just a recipe for misleading results, the method is still regarded by the community as off-limits. In the same vein, warnings from experts in the field such as Gelman (2008) about an

overemphasis on "the convergence of the Gibbs sampler" and abuse of Bayesian statistics as the latest methodological fad should also be discussed openly.[15]

Secondly, pioneers in Bayesian methods will also be needed, and they should be the experienced researchers themselves.

Thirdly, a transitioning from rigid ready-made software to flexible, open and re-packageable computer codes should be preceded by easy-to-use and graphics-intuitive introduction to Bayesian concepts. They should be computer programs, better stand-alone ones, which visualize the Bayesian concepts from simple to complex providing some utilities of changing parameters and specifications so that a novice learner can observe the shifting trends, and related behaviour of statistics

In our experience of working with the data structures and the documents of McElreath[6] and Kruschke,[7] we found several limitations in terms of pedagogy. First, these books use different datasets for giving instructions on how to apply the Bayesian approach, which caused certain barriers for the learners to move from one set of data to the others. Second, different problems require different ways of performing statistical checks on prior and posterior estimations, using routines placed in different files. Finally, the fact that the problems and the programmed routines are in different places, and it is not uncommon that they are calling others or called by others, in different problems, created the (incorrect) feeling that they are separate parts and come out of the blue.

We are currently building a real-world database using 5,300 questionnaires to investigate the tendencies among junior high school students in a Northern province of Vietnam regarding book reading, STEM learning, and GPA. We aim to perform the Bayesian statistical analyses in a pedagogical manner so that we can show that working with Bayesian statistics is not that intimidating, through steps from organizing data, parameterization, choosing functions, building models, checking estimates on priors and posteriors, to interpreting the results. Such a manual is hoped to help the researchers, both the senior and novice, to overcome three kinds of fears discussed in the previous section.

All scientific debates ultimately boiled down to the hope for a spark of curiosity among their audiences. The debate over statistical significance[1,2] is the inspiration for the interview of this paper, which in turn, gives us the idea of creating a tutorial program that could help the transition from the frequentist world to the Bayesian world smoother. If the efforts would spark enough curiosity among researchers, we can hope that a smoother transition to the Bayesian may happen and gradually help cure the science disease of stargazing, *p*-hacking, HARKing, and all sorts of variations. Even though the hope may be regarded as too naive, having no hope at all is truly the disaster.

Quan-Hoang Vuong, Tung Manh Ho, Phuong Viet La
Centre for Interdisciplinary Social Research, Phenikaa University, Hanoi, Vietnam
hoang.vuongquan@phenikaa-uni.edu.vn

References

1 Benjamin DJ, Berger JO, Zinman J, *et al*. Redefine statistical significance. *Nature Human Behaviour* 1, doi:10.1038/s41562-017-0189-z (2017).

2 Amrhein V, Greenland S. Remove, rather than redefine, statistical significance. *Nature Human Behaviour* 1, doi:10.1038/s41562-017-0224-0 (2017).

3 Ho TM, Nguyen HKT, Vuong T-T, *et al*. On the sustainability of co-authoring behaviors in Vietnamese social sciences: A preliminary analysis of network data. *Sustainability*, 9, 2142. (2017)

4 Editorial. Promoting reproducibility with registered reports. *Nature Human Behaviour* 1, 34, doi:10.1038/s41562-016-0034 (2017).

5 Ho MT, Vuong QH. The values and challenges of 'openness' in addressing the reproducibility crisis and regaining public trust in social sciences and humanities. *European Science Editing*, 45(1): 14-17. DOI: https://doi.org/10.20316/ESE.2019.45.17021. (2019).

6 McElreath R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (2nd ed.) (CRC Press, 2016).

7 Kruschke JK. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS and Stan* (2nd ed.) (Academic Press, 2015).

8 Scales JA, Snieder R. To Bayes or not to Bayes?. *Geophysics* 62(4), 1045-1046 (1997).

9 Malakoff D. Bayes offers a 'new' way to make sense of numbers. *Science* 286, 1460-1464, doi:10.1126/science.286.5444.1460 (1999).

10 Harrell F. My journey from frequentist to Bayesian statistics. *Statistical Thinking*; http://www.fharrell.com/2017/02/my-journey-from-frequentist-to-bayesian.html (2017).

11 Efron B. Why isn't everyone a Bayesian?. *The American Statistician* 40, 1-5 (1986).

12 Rupp AA, Dey DK, Zumbo BD. To Bayes or not to Bayes, from whether to when: applications of Bayesian methodology to modeling. *Structural Equation Modeling* 11, 424-451 (2004).

13 Wetzels R, Matzke D, Lee MD, *et al*. Statistical evidence in experimental psychology: an empirical comparison using 855 t-tests. *Perspectives on Psychological Science* 6, 291-298 (2011).

14 Perezgonzalez JD, Frías-Navarro MD. Retract $p < 0.005$ and propose using JASP, instead. *F1000Research* 6, 2122, v1, doi:10.12688/f1000research.13389.1 (2017)

15 Gelman A. Objection to Bayesian statistics. *Bayesian Analysis* 3, 445-450 (2008).