RACSAM
Rev. R. Acad. Cien. Serie A. Mat.
VOL. **103** (1), 2009, pp. 145–148
Estadística e Investigación Operativa / Statistics and Operations Research
Comentarios / Comments

Comments on: Natural Induction: An Objective Bayesian Approach

Kalyan Raman

1 Introduction

Berger, Bernardo and Sun's thought-provoking paper offers a Bayesian resolution to the difficult philosophical problem raised by inductive inference. In a nutshell, the philosophical problem plaguing inductive inference is that no finite number of past occurrences of an event can prove its continuing occurrence in the future. It is thus natural to seek probabilistic reassurance for our instinctive feeling that an event repeatedly observed in the past must be more likely to recur than an event that happened only infrequently. Consequently, as the authors note, the "rule of succession" and the "natural law of induction" have engaged the attention of philosophers, scientists, mathematicians and statisticians for centuries. And rightly so because—despite philosophical qualms about induction—science cannot progress without inductive inferences. The vintage of the induction problem testifies to its difficulty and the pervasiveness of inductive inferences in science reinforces our ongoing efforts to strengthen its underlying logic and fortify its foundations through statistical reasoning. These circumstances necessitate diverse approaches to establish a rigorously justifiable framework for inductive inference.

Berger et al. have made a sophisticated contribution to the literature on rigorously justifying inductive inference, and they have innovatively illuminated an illustrious path blazed by none other than Laplace himself. At the risk of appearing mean-spirited, my main complaint with their solution is the technical virtuosity demanded by their methodology. The mathematical complexities of finding a reference prior are daunting enough to dissuade all but the most lion-hearted in venturing on the search. Given the importance of the problem that Berger et al. address, it may be worthwhile to dredge up an existing solution that seems to be unknown in the statistics literature. In that spirit, I will discuss an alternative approach that produces one of the key results that Berger et al. derive through their reference prior. My approach has the merit of being considerably simpler and more flexible at the expense of possibly not satisfying all the four desiderata listed in Bernardo (2005) ([2]) for objective posteriors, but it does quickly produce a central result in Berger et al. and offers insights into the value of additional replications—an issue that lies at the heart of inductive inference and scientific inquiry. First a few thoughts on the relevance of replications to the topic at hand.

Recibido / Received: 10 de marzo de 2009.
These comments refer to the paper of James O. Berger, José M. Bernardo and Dongchu Sun, (2009). **Natural Induction: An Objective Bayesian Approach**, *Rev. R. Acad. Cien. Serie A. Mat.*, **103**(1), 125–135.

2 Inductive inference and replications

Bernardo (1979) ([3]) defines a reference posterior in terms of limiting operations carried out on the amount of information about the unknown parameter, obtained from successive independent *replications* of an experiment. Bernardo's definition of reference priors through replications resonates well with a key guiding principle of good scientific research. Replications are the heart and soul of rigorous scientific work—findings that are replicated independently by investigators increase our confidence in the results (Cohen 1990 ([4])). Thus, replications play a fundamental role both in the mathematical definition of a reference posterior and in the scientific process. Clearly, replications are intimately related to inductive inference. It would thus seem conceptually attractive, if, as a by-product of modifying the Laplace Rule of Succession to strengthen its logical basis, we are also able to figure out the optimal informational role of replications.

3 Improving the Laplace rule of succession

Using a reference prior: The solution proposed by Berger et al. to the limitations of the Laplace Rule of succession is displayed in equations (20) and (27) of their paper. Using their notation, the authors' result is that:

$$\pi_u(E_n) = \frac{n + 1/2}{n + 1} \tag{1}$$

which yields faster convergence to unity than the Laplace Rule. The Laplace Rule yields the probability $\pi_u(E_n) = \frac{n+1}{n+2}$. To obtain equation (1), Berger et al. use a hypergeometric model (equation (4) in their paper) together with the reference prior shown in equation (13) of their paper. Equation (13) is obtained by using the Jeffreys prior (equation (12) in Berger et al.) in conjunction with an asymptotic argument which is justified on the basis of exchangeability, as the authors have shown elsewhere. Their logic is sophisticated and beautiful but the price paid for such beauty is that the resultant derivations are arduous. Indeed, Berger and Bernardo (1992) ([1]) themselves admit that the general reference prior method "is typically very hard to implement." Under these circumstances, perhaps the search for a simpler approach is defensible and meritorious of some attention.

Using a beta prior: In Raman (1994) ([7]), I show that the following rule of succession generalizes the Laplace Rule. Suppose that p is the probability that a scientific theory is true, and assume that the prior for p is $\mathrm{Be}(p \mid \alpha, \beta)$; if we subsequently obtain 'n' confirmations of the theory, then, using the notation $b_n(E_n)$ to suggest its beta-binomial roots, the probability of observing an additional confirmation is given by,

$$b_n(E_n) = \frac{\alpha + n}{\alpha + \beta + n} \tag{2}$$

Equation (2) follows easily from a result in DeGroot (1975 ([5]), p. 265) guaranteeing equivalence of the sequential updating of $\mathrm{Be}(p \mid \alpha, \beta)$ with the updating of $\mathrm{Be}(p \mid \alpha, \beta)$, conditional on having observed "n" successes. The Jeffreys prior $f(p) = \frac{1}{\pi}\frac{1}{\sqrt{p(1-p)}}$, $0 < p < 1$, is a special case resulting from the choice $\alpha = \beta = \frac{1}{2}$ in the prior $\mathrm{Be}(p \mid \alpha, \beta)$. For that choice of prior, equation (2) reduces to the equation (20) of the Berger et al. paper:

$$\text{For } \alpha = \beta = 1/2, \qquad b_n(E_n) = \frac{n + 1/2}{n + 1} \tag{3}$$

Polya (1954) ([6]) recommends a number of properties that an "induction-justifying" rule ought to have—and the beta-binomial rule (equation (2) above) exhibits those desiderata.

Using a general prior, not necessarily beta: It would be natural to object that the above derivation is driven by a specific prior—the Beta distribution. However, in Raman (2000) ([8]), I show that a generalized rule of succession can be obtained for a general class of priors which includes the Beta distribution as a special case. The generalized rule of succession includes as special cases, the original Laplace Rule, the Beta-Binomial rule and the rule derived in Berger et al. through a reference prior. The exact result is the following: if $g(p)$ is a prior density function with a convergent Maclaurin series representation $g(p) \sim \sum_{i\geq 0} a_i p^i$, then, using the notation g_n to denote the rule of succession under this general prior density,

$$g_n = \sum_{i\geq 0} a_i \, \frac{i+1+n}{i+2+n} \tag{4}$$

As special cases, $a_0 = 1$, $a_i = 0$, $i \geq 1$, yields the Laplace rule of succession, the choice of a_i as the coefficients in a power-series expansion of $\mathrm{Be}(p \,|\, \alpha, \beta)$ results in the beta-binomial rule, which includes, as a special case, the rule of succession for the Jeffreys' prior derived in Berger et al. through a reference prior. Clearly, g_n may be viewed as a linear combination of beta-binomial rules of succession or, with equal right, as a linear combination of Laplacian rules of succession.

From an applied perspective, the Beta density's flexibility and tractability make it an attractive choice for a prior; from a theoretical perspective, the above results show that it suffices for the purpose of generating a more plausible rule of succession than the Laplacian rule, and, in fact, yields results that are identical to Berger et al. Finally, although I do not delve into the topic here, the Beta prior permits derivation of an adaptive controller that shows the value of performing an additional replication as a function of our prior beliefs about the theory, the accumulated evidence in favor of the theory, the precision deemed necessary and the cost of the replication (Raman 1994) ([7]).

Using the Jeffreys' reference prior in Berger et al.: I should remark on the following property of the Jeffreys' reference prior which appears somewhat odd to me. When $N = 1$, it assigns a probability of 0.50, for R, which makes sense. Furthermore, as $N \to \infty$, the probability $\pi_r(R \,|\, N)$ for $R = N$, tends to 0 —a result which is attractive. However as N increases, at intermediate values of N, the behavior of $\pi_r(R \,|\, N)$ is somewhat odd for $R = N$. Let me explain.

Consider equation (13) in Berger et al.

$$\pi_r(R \,|\, N) = \frac{1}{\pi} \, \frac{\Gamma(R+\frac{1}{2})\,\Gamma(N-R+\frac{1}{2})}{\Gamma(R+1)\,\Gamma(N-R+1)}\,, \qquad R \in \{0, 1, \ldots, N\}, \tag{13}$$

so $R = N$ implies

$$\pi_r(R \,|\, N) = \frac{1}{\pi} \, \frac{\Gamma(N+\frac{1}{2})\,\Gamma(\frac{1}{2})}{\Gamma(N+1)}\,.$$

Consider the behavior of the above function as N grows large. The first derivative of $\pi_r(N \,|\, N)$ is a complicated expression involving the polygamma function, but if we plot $\pi_r(N \,|\, N)$ as a function of 'N', then we obtain insights. Plotting the function in Mathematica as a function of N (see Figure 1), we find that $\pi_r(N \,|\, N)$ at first drops very steeply but that the rate of decline slows down dramatically for $N > 20$. For example, for $100 \leq N \leq 200$, the probability drops from 0.056 at $N = 100$ to 0.039 at $N = 200$.

Thus $\pi_r(N \,|\, N)$ is insensitive to new information for large but finite values of N, which is the case that would be of greatest pragmatic interest in scientific theory-testing. It would be useful if the authors could comment on the significance of this property for natural induction.

4 Conclusion

My thoughts on the elegant analysis of Berger et al. are driven by an entirely applied perspective. Consequently, I seek the most parsimonious and mathematically tractable route to model-building. The alternative approach I have described lacks the technical sophistication and mathematical rigor of the authors' reference prior approach—its primary justification is its ease of use and pliability at addressing a broader set

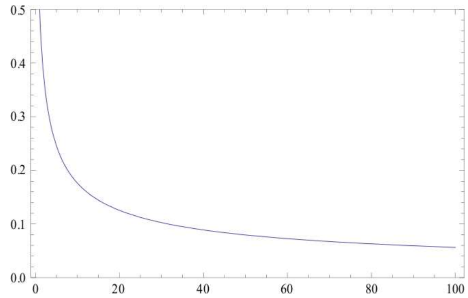

Figure 1. $\pi_r(N \mid N)$ as a function of N.

of issues (such as the development of an optimal controller to balance the tradeoffs involved in replicating experiments). I realize that these broader issues are not necessarily relevant to the authors—but even so, I would argue that the authors may benefit from thinking about how reference priors can address these questions better than my naïve approach based on a mathematically convenient family of conjugate priors, because their reflection on the applied concerns I have raised could lead to new results that would broaden the scope and scientific impact of reference priors on researchers across multiple disciplines.

In conclusion, I applaud the authors for their innovative application of a powerful new technique to an important and vexing problem of ancient vintage, and hope that some of their future work on reference priors makes the methodology less mysterious, thereby disseminating their ideas to a wider audience and paving the way for new applications based on reference priors.

References

[1] Berger, J. O. and Bernardo, J. M., (1992). On the development of reference priors. in *Bayesian Statistics*, **4**. (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) Oxford: University Press, 35–60 (with discussion).

[2] Bernardo, J. M., (2005). Reference analysis, in *Handbook of Statistics*, **25**, 17–90. D. K. Dey and C. R. Rao, (eds.) Amsterdam: Elsevier.

[3] Bernardo, J. M., (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. B*, **41**, 113–147 (with discussion). Reprinted in *Bayesian Inference*, **1** (G. C. Tiao and N. G. Polson, eds.) Oxford: Edward Elgar, 229–263.

[4] Cohen, Jacob, (1990). Things I have Learned So Far, *American Psychologist*, **45**, (December), 1304–1312.

[5] DeGroot, Morris H., (1975). *Probability and Statistics*, Reading, MA: Addison-Wesley.

[6] Polya, George, (1968). *Mathematics and Plausible Reasoning*, Vol. **2**, Princeton, NJ: Princeton University Press.

[7] Raman, Kalyan, (1994). Inductive Inference And Replications:A Bayesian Perspective, *Journal of Consumer Research*, March, **20**, 633–643.

[8] Raman, Kalyan, (2000). The Laplace Rule of Succession Under A General Prior, *Interstat*, June, **1**, http://interstat.stat.vt.edu/interstat/articles/2000/abstracts/u00001.html-ssi.

Kalyan Raman
Medill IMC Department,
Northwestern University,
USA
k-raman@northwestern.edu