RACSAM
Rev. R. Acad. Cien. Serie A. Mat.
VOL. **103** (1), 2009, pp. 137–139
Estadística e Investigación Operativa / Statistics and Operations Research
Comentarios / Comments

Comments on: Natural Induction: An Objective Bayesian Approach

F. Javier Girón and Elías Moreno

We would like to thank the editor of RACSAM, Professor Manuel López Pellicer, for the opportunity he is offering to us of discussing this paper, and to congratulate Berger, Bernardo and Sun for an interesting and thought provoking paper.

The paper is motivated by the observation that the uniform prior for R, say $\pi(R|N) = 1/(N+1)$, $R = 0, \ldots, N$, gives poor results. It is shown that the posterior probability that all the N elements of the population are conforming, conditional on the event that all the observed n elements in the sample are conforming, is very small for N large, whatever moderate the sample size n should be. Then, a more reasonable prior $\pi(R|M)$ is provided on the ground of being compatible with the Jeffreys prior for the parameter θ of the Binomial limiting distribution with parameters (n, θ), where $\theta = \lim_{R\to\infty, N\to\infty} R/N$. We enjoyed reading this clear argumentation.

However, in the abstract it is recognized that "*Bayesian solutions to this problem may be very sensitive to the choice of the prior, and there is no consensus as to the appropriate prior to use.* " It seems to us that the natural consequence of this assertion —that we share— is to consider a class of priors and reporting their posterior answers, instead of considering the posterior answer for the single reference prior for R. In this discussion we try to add the robustness analysis that we feel is missing in the paper.

For simplicity we will consider the limiting Binomial distribution $\mathrm{Bi}(r|n, \theta)$, and the two problems addressed in the paper. Firstly, the testing problem

$$H_0 : \theta = 1 \text{ versus } H_1 : \theta \in [0, 1],$$

conditional on the dataset $r = n$, the event that all the elements of the sample are $+$. Secondly, the computation of the posterior predictive probability that a new observation is $+$, conditional on $r = n$.

The naive objective model selection formulation of this testing problem is that of choosing between the reduced sampling model

$$M_0 : \mathrm{Bi}(n|n, \theta = 1)$$

and the full sampling model with the Jeffreys prior for θ, that is

$$M_1 : \left\{ \mathrm{Bi}(n|n, \theta),\ \pi^J(\theta) = \frac{1}{\pi}\theta^{-1/2}(1-\theta)^{-1/2} \right\}.$$

However, the Jeffreys prior does not concentrate its probability mass around the null with the consequence that those θ close to zero are privileged by the Jefrreys priors when being compared with the null $\theta = 1$. This is not reasonable, and many authors claim for a different prior to be used for testing that

Recibido / Received: 13 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.

should be concentrated around the null. See, for instance, Jeffreys (1961, Chapter 5) ([8]), Gûnel and Dickey (1974) ([9]), who note that this is the "Savage continuity condition", Berger and Sellke (1987) ([3]), Casella and Berger (1987) ([4]), Morris (1987) ([12]), Berger (1994) ([2]), Casella and Moreno (2009) ([5]).

The point is how to define an objective class of priors that concentrate mass around the null. Fortunately, an answer to this question is provided by the class of intrinsic priors (Berger and Pericchi 1996 ([1]), Moreno *et al.* 1998 ([10])). This objective class of priors has been proved to behave extremely well for model selection in different contexts (Casella and Moreno 2006 ([5]), Consonni and La Roca 2008 ([7]), Moreno and Girón 2008 ([11])). The intrinsic priors for θ depend on a hyperparameter m that controls the degree of concentration of the priors around the null, and it ranges from 1 to n, so as to not exceed the concentration of the likelihood of θ (Casella and Moreno 2009 ([5])). For the above model selection problem standard calculations render the intrinsic prior class as the set of beta distributions $\mathrm{Be}(m+1/2, 1/2)$, that is

$$\pi^I(\theta|m) = \frac{\Gamma(m+1)}{\Gamma(m+1/2)\Gamma(1/2)}\theta^{m-1/2}(1-\theta)^{-1/2}, \qquad m = 1, 2, \ldots, n.$$

Therefore, in the above model selection problem the Jeffreys prior should be replaced with the intrinsic prior, and M_0 should be compared with

$$M_1 : \{\mathrm{Bi}(n|n,\theta),\ \pi^I(\theta|m), m = 1, 2, \ldots, n\}.$$

We note that as m increases the intrinsic prior concentrates more around the null. Certainly, when the null is compared with models located in a small neighborhood of the null, one expects from the model selection problem an answer with more uncertainty than when the null is compared with models located far from it.

The posterior probability of the null for the intrinsic priors is given by

$$\Pr(\text{All } + |n, m) = \left(1 + \frac{\Gamma(m+1)\Gamma(n+m+1/2)}{\Gamma(m+1/2)\Gamma(n+m+1)}\right)^{-1}, \qquad m = 1, ..., n.$$

Likewise, the posterior probability that a new observation is $+$, conditional on $r = n$, is given by the total probability theorem as

$$\Pr(+|n, m) = \sum_{i=0}^{1} \Pr(+|M_i, n, m)P(M_i|n, m),$$

where $\Pr(+|M_0, n, m) = 1$, and

$$\Pr(+|M_1, n, m) = \frac{n+m+1/2}{n+m+1}.$$

Example 1 *Assuming that the galápagos population in the island is large enough, we obtain that*

$$\min_{m=1,\ldots,55} \Pr(\text{All } + |n = 55, m) = \Pr(\textit{All } + |n = 55, m = 55) = 0.586,$$

and

$$\max_{m=1,\ldots,55} \Pr(\text{All } + |n = 55, m) = \Pr(\text{All } + |n = 55, m = 1) = 0.869,$$

while

$$\Pr(+|n = 55, m) \simeq 0.998$$

for $m = 1, 2, \ldots, 55$.

This example illustrates something about robustness that is well known: the posterior probability of an event is typically much less sensitive to the prior than the tests are. The posterior probability that a new observation is $+$, conditional on $r = n$, that we have obtained is similar to that given in the paper, but the report for the testing problem given in the paper and that given by us are rather different.

References

[1] BERGER, J. O. AND PERICCHI, L. R., (1996). The intrinsic Bayes factor for model selection and prediction, *Journal of the American Statistical Association*, **91**, 109–122.

[2] BERGER, J. O., (1994). An overview of robust Bayesian analysis (with discussion), *Test*, **3**, 5–124.

[3] BERGER, J. O. AND SELLKE, T. (1987). Testing a point null hypothesis: The irreconcilability of p-values and evidence (with discussion), *Journal of the American Statistical Association*, **82**, 112–122.

[4] CASELLA, G. AND BERGER, R. L., (1987). Reconciling Bayesian and Frequentist Evidence in the One-Sided Testing Problem (with discussion), *Journal of the American Statistical Association*, **82**, 106–111.

[5] CASELLA, G. AND MORENO E., (2006). Objective Bayesian variable selection. *Journal of the American Statistical Association*, **101**, 157–167.

[6] CASELLA, G. AND MORENO E., (2009). Assessing robustness of intrinsic test of independence in two-way contingency tables, *Journal of the American Statistical Association* (to appear).

[7] CONSONNI, G. AND LA ROCA, L., (2008). Tests Based on Intrinsic Priors for the Equality of Two Correlated Proportions, *Journal of the American Statistical Association*, **103**, 1260–1269.

[8] JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Oxford: University Press.

[9] GÛNEL, E. AND DICKEY, J., (1974). Bayes factors for independence in contingency tables, *Biometrika*, **61**, 545–557.

[10] MORENO, E., BERTOLINO, F. AND RACUGNO, W. (1998). An Intrinsic Limiting Procedure for Model Selection and Hypotheses Testing, *Journal of the American Statistical Association*, **93**, 1451–1460.

[11] MORENO, E. AND GIRÓN, F. J., (2008). Comparison of Bayesian objective procedure for variable selection in linear regression, *Test*. **17**, 472–492.

[12] MORRIS, C. N., (1987). Discussion of Berger/Sellke and Casella/Berger. *Journal of the American Statistical Association*, **82**, 106–111.

F. Javier Girón
Departamento de Estadística e Investigación Operativa,
Universidad de Málaga,
España
fj_giron@uma.es

Elías Moreno
Departamento de Estadística e Investigación Operativa
Universidad de Granada,
España
emoreno@ugr.es

RACSAM
Rev. R. Acad. Cien. Serie A. Mat.
VOL. **103** (1), 2009, pp. 141–141
Estadística e Investigación Operativa / Statistics and Operations Research
Comentarios / Comments

Comments on: Natural Induction: An Objective Bayesian Approach

Dennis V. Lindley

This very fine paper is valuable because it produces challenging and interesting results in a problem that is at the heart of statistics. The real populations that occur in the world are all finite and the infinities that we habitually invoke are constructs, albeit useful but essentially artificial. Another reason for being excited about the work is that it opens the way to further Bayesian studies of the practice of sampling procedures, where several features, and not just one as here, are being investigated.

The authors make much use of the term 'objective'; what does this mean? My dictionary gives at least two, rather different, meanings: "exterior to the mind" and "aim". The authors would appear to use both meanings in the same sentence when they say, at the end of Section 1, "A formal objective Bayesian solution ... is the main objective of this paper". My opinion is that all statistics is subjective, the subject being the scientist analysing the data, so that the contrary position needs clarification. There is also a confusion for me with the term 'reference prior', a term that I have queried in earlier discussions.

An unstated assumption that R and n are independent, given N, has crept into (2) where $\Pr(R \mid N)$ should be $\Pr(R \mid n, N)$. The assumption may not be trivial, as when the sampling procedure is to continue until the first non-conforming element is found. Another assumption made is that N is fixed, despite the fact that, in the example of the tortoises, it is unknown. I would welcome some clarification of the role of the sampling procedure.

Perhaps the most interesting section in the paper is 3, where the use of Jeffreys's prior (12), or (13), superficially very close to the uniform (roughly $1/2$ a confirmation and $1/2$ non-confirmation) gives such different results from it. For example (20) can be written $\pi_r(E_n) = \frac{2n+1}{2n+2}$. Thus Jeffreys gives the same result as Laplace but for *twice* the sample size. Again in the hierarchical model $\pi_r(\text{All} + \mid n, N)$ is about $\sqrt{n/N}$, equation (13), whereas with the uniform it is about n/N, equation (9), the larger value presumably being due to the prior on θ attaching higher probability than the uniform to values near 1. We therefore have the unexpected situation where an apparently small change in the prior results in an apparently large change in at least some aspects of the posterior.

There are many issues here that merit further study and we should be grateful to the authors for the stimulus to employ their original ideas to do this.

Dennis V. Lindley
Royal Statistical Society's Guy Medal in Gold in 2002.
University College London, UK
ThomBayes@aol.com

Recibido / Received: 24 de febrero 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.

RACSAM
Rev. R. Acad. Cien. Serie A. Mat.
VOL. **103** (1), 2009, pp. 143–144
Estadística e Investigación Operativa / Statistics and Operations Research
Comentarios / Comments

Comments on: Natural Induction: An Objective Bayesian Approach

Brunero Liseo

I really enjoyed reading the paper. It shed new and clear light to some issues which stand at the core of the statistical reasoning.

In standard statistical models, where the parameter space is a subset of $\mathbb{R}^k$ for some integer k, *reference priors*, and to some extent, *Jeffreys' priors*, offer a way to find a compromise between Bayesian and classical goals of statistics. Usually such a solution lies at the boundary of the Bayesian world, i.e. the objective priors to be used in order to get *good* frequentist behaviour are in general improper. A remarkable exception is however the objective prior for the probability of success θ in a sequence of Bernoulli trials.

Finite populations problems can hardly be approximated by an "infinite population" scenario and, apart from the computational burden, difficulties arise in figuring out what the "boundary of the Bayesian world" would be in these situations. In other words, it is not clear whether a compromise between Bayesian and frequentist procedures is at all possible in finite populations. This paper is then welcome in providing some evidence that, at least, an objective Bayesian analysis of such class of problems is indeed meaningful.

In the rest of the discussion I will focus on the *Law of natural induction*, that is how to evaluate the probability that all the N elements in a population are conforming, given that all the n elements in the sample are. Let R be the unknown number of conforming elements in the population.

The Authors criticize the use of a uniform prior for R and argue that a version of the reference prior, based on the idea of embedding (Berger, Bernardo and Sun, 2009 ([1])), provides more appropriate results. I agree with this conclusion, although the differences are not dramatic. Both uniform and reference priors for R are "symmetric around $N/2$"; besides that, the hypothesis $R = N$ does not play a special role: for instance, the two hypotheses $R = N$ and $R = 0$ are given the same weight under both priors; also the cases $R = N$ and $R = N - 1$ have approximately the same prior (and posterior...) probability both under the uniform and the reference prior. These conclusions are perfectly reasonable for an estimation problem when no prior information on R is available. However, the Authors argue that the small value of $\Pr(\text{All} + |n, N)$ *"clearly conflicts with the common perception from scientists that, as* n *increases,* $\Pr(\textit{All} + |n, N)$ *should converge to one, whatever the value of* N *might be"*. This is the crucial point and brings into the discussion the role of models in Statistics. The uniform and the reference prior approaches are not able to catch the idea that $R = N$ and R close to N may be two dramatically different descriptions of the phenomenon: if we are interested in the number of individuals in a population which do not show a genetic mutation, $R = N$ would imply the absence of the mutation with completely different scientific implications from those related to any other value of R.

If the hypothesis $R = N$ has a "physical meaning" then I would have no doubt that the correct analysis to perform is the one described in Section 4. This analysis would make Jeffreys and other objective

Recibido / Received: 24 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.

Bayesians happy, since it clearly distinguishes between the statistical "meaning" of the hypothesis $R = N$ and the meaning of other hypotheses, such as $R = 0$ or $R = N - 1$.

In such a situation, formula (28) (or 26) seems a perfectly reasonable "objective Bayesian" answer to the Law of Natural Induction: it is monotonically increasing in n for fixed N and monotonically decreasing in N for fixed n.

So the final question is: can we consider all the scientific questions equivalent to those leading to formula (28)? Should not we take into account the *common perception from scientists* as a guide to choice the best statistical formalization of the problem? To make the point, what happens if a reasonable working model in a specific application, is of the type "R close to N"? This is not an infrequent situation; consider, for example, surveys on human or animal populations in order to detect the presence of rare events. In such cases, strong prior information about R might be available and one would rather prefer to perform a reference analysis conditional on some partial prior information, along the lines of Sun and Berger (1998 ([2]), Reference priors with partial information, *Biometrika* [2]).

References

[1] BERGER, J., BERNARDO, J. M. AND SUN, D. (2009b). Reference priors for discrete parameter spaces. *Submitted*.

[2] SUN, D. AND BERGER, J. O. (1998). Reference priors with partial information. *Biometrika*, **85**, 55–71.

Brunero Liseo
Dipartimento di studi geoeconomici, linguistici, statistici e storici per l'analisi regionale
Sapienza Università di Roma,
Italy
brunero.liseo@uniroma1.it

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

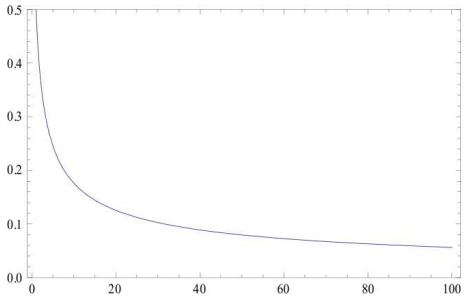

Figure 1. $\pi_r(N \mid N)$ as a function of N.

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`

RACSAM
Rev. R. Acad. Cien. Serie A. Mat.
VOL. **103** (1), 2009, pp. 149–150
Estadística e Investigación Operativa / Statistics and Operations Research
Comentarios / Comments

Comments on: Natural Induction: An Objective Bayesian Approach

Christian P. Robert

This is a quite welcomed addition to the multifaceted literature on this topic of natural induction that keeps attracting philosophers and epistemologists as much as statisticians. The authors are to be congratulated on their ability to reformulate the problem in a new light that makes the law of natural induction more compatible with the law of succession. Their approach furthermore emphasize the model choice nature of the problem.

First, I have always been intrigued by the amount of attention paid to a problem which, while being formally close to Bayes' own original problem of the binomial posterior, did seem quite restricted in scope. Indeed, the fact that the population size N is supposed to be known is a strong deterrent to see the problem as realistic, as shown by the (neat!) Galapagos example. My first question is then to wonder how the derivation of the reference prior by Berger, Bernardo, and Sun extends to the case when N is random, in a rudimentary capture-recapture setting. An intuitive choice for $\pi_r(N)$ is $1/N$ (since N appears as a scale parameter), but is

$$\frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}\,\frac{\Gamma(R+1/2)\Gamma(N-R+1/2)}{R!(N-R)!}\,\frac{1}{N}$$

summable in both R and N? (Obviously, improperness of the posterior does not occur for a fixed N.)

As exposed in the paper, one reason for this special focus on natural induction may be that it leads to such a different outcome when compared with the binomial situation. Another reason is certainly that Laplace succession's rule seems to summarise in the simplest possible problem the most intriguing nature of inference. And to attract its detractors, from the classical Hume's (1748) ([1]) to the trendy Taleb's (2007) ([2]) "black swan" argument (which is not the issue here, since the "black swan" criticism deals with the possibility of model changes).

Second, the solution adopted in the paper follows Jeffreys' approach and I find this perspective quite meaningful for the problem at hand. Indeed, while N can be seen as $(N-1)+1$, i.e. as one of the $N+1$ possible values for R, the consequence of having R equal to either N or 0 lead to atomic distributions for the number of successes. Thus, to distinguish those two values from the other makes sense even outside a testing perspective. In Jeffreys' (1939) original formulation, both extreme values, 0 and N, are kept separate, with a prior probability k between $1/3$ and $1/2$. I thus wonder why the authors moved away from this original perspective. The computation for this scenario does not seem much harder since $\pi_r(0|N) = f(N)$ as well and the equivalent of (22) would then be

Recibido / Received: 11 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.

$$\pi_\phi(\text{All} + |n, N) = \left(1 + \frac{k}{1-2k}\,\frac{f(n) - 2f(N)}{1-2f(N)}\right)^{-1},$$

which is then $(1 + 0.5f(n))^{-1}$ for N large. In this case, (24) is replaced with $\sqrt{n}/(\sqrt{n} + 2/\sqrt{\pi})$, not a considerable difference.

In conclusion, I enjoyed very much reading this convincing analysis of a hard "simple problem"! It is unlikely to close the lid on the debate surrounding the problem, especially by those more interested in the philosophic side of it, but rephrasing natural induction as a model choice issue and advertising the relevance of Jeffreys' approach to this very problem have bearings beyond the "simple" hypergeometric model.

References

[1] HUME, D., (1748). *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*, (2000 version: Oxford University Press).

[2] TALEB, N. N., (2007). *The Black Swan: The Impact of the Highly Improbable*. Penguin Books, London.

Christian P. Robert
CEREMADE
Université Paris Dauphine,
France
xian@ceremade.dauphine.fr

RACSAM
Rev. R. Acad. Cien. Serie A. Mat.
VOL. **103** (1), 2009, pp. 151–151
Estadística e Investigación Operativa / Statistics and Operations Research
Comentarios / Comments

Comments on: Natural Induction: An Objective Bayesian Approach

Raúl Rueda

According to the authors: "The conventional use of a uniform prior for discrete parameters can ignore the structure of the problem under consideration". This motivates the introduction of a hierarchical structure

$$p(R \mid N) = \int_0^1 \mathrm{Bi}(R \mid N, \theta)\, \mathrm{Be}(\theta \mid 1/2, 1/2)\, \mathrm{d}\theta,$$

where $\mathrm{Bi}(R \mid N, \theta)$ is the binomial ditribution with parameters (N, θ) and $\mathrm{Be}(\theta \mid 1/2, 1/2)$ is the reference prior for θ in this case.

However, it must be pointed out that the assumption of exchangeability to justify the hierarchical structure is also valid for the uniform prior, by replacing the $\mathrm{Be}(\theta \mid 1/2, 1/2)$ distribution with a uniform in $(0, 1)$ yielding as a prior $p[R] = 1/(N+1)$.

Anyway, the reference *Rule of Succession* is essentially the same as Laplace's, but there is a difference in $\pi[\mathrm{All} + \mid n, N]$ when n is small compared with N. This difference disappears when $n \to N$.

Even though the authors find equations (9) and (11) to be "dramatically different", suggesting a contradictory behaviour, this is perfectly possible for example, in the case of rare events, such a finding a person who suffers from a desease with a prevalence of one in a million. If the conforming event is the absence of the desease, then there is a high probability that we observe another conforming element given that all elements in the sample are conforming. At the same time, the probability that all are conforming is close to zero, so in this case (9) and (11) are both valid, and the behaviour of (22) becomes more difficult to accept.

Raúl Rueda
Departamento de Probabilidad y Estadística,
Universidad Autonoma Nacional de Mexico,
Mexico
pinky@sigma.iimas.unam.mx

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.

RACSAM
Rev. R. Acad. Cien. Serie A. Mat.
VOL. **103** (1), 2009, pp. 153–155
Estadística e Investigación Operativa / Statistics and Operations Research
Comentarios / Comments

Comments on: Natural Induction: An Objective Bayesian Approach

S. L. Zabell

Berger, Bernardo, and Sun (BBS) briefly allude to exchangeability in their paper. Since I personally find this the most natural way to view these types of questions, I begin by discussing their results from this point of view.

Suppose $X_1, \ldots, X_N$ is a finite exchangeable sequence of 0s and 1s with respect to a probability P. The simplest such probability assignment is the one corresponding to an urn $U_{R,N}$ with N balls, R 1s and $N - R$ 0s, where we drawn the balls out at random without replacement. Denote this hypergeometric probability assignment by $H_{R,N}$. If $S_N = X_1 + \cdots + X_N$ and $p_R = P(S_N = R)$, then by exchangeability $P = \sum_R p_R H_{R,N}$. Call this the finite de Finetti representation.

In general a finite exchangeable sequence $X_1, \ldots, X_N$ cannot be extended (and remain exchangeable), but if it can be indefinitely extended then it admits an integral representation: for some probability measure Q on the unit interval one has

$$P(S_N = R) = \int_0^1 \binom{N}{R} p^R (1-p)^{N-R} \, \mathrm{d}Q(p);$$

this is the celebrated de Finetti representation theorem.

Thus P can be thought of arising in two apparently different, but actually stochastically equivalent ways: 1) choose a p-coin according to Q, and toss it N times, or 2) choose an urn $U_{R,N}$ according to p_R, and then draw the balls out at random from it without replacement.

In Laplace's famous 1774 paper he took the first route, adopting the flat $\mathrm{d}Q(p) = \mathrm{d}p$. This special prior has, as BBS note, the interesting properties that $P(S_N = R) = 1/(N+1)$, for $0 \le R \le N$; and for any $n < N$, $P(X_{n+1} = 1 | S_n = r) = (r+1)/(n+2)$; the "rule of succession".

The classical Laplacean analysis raises a number of questions: the nature of p (presumably some form of "physical probability"); the implicit presence of an (at least in principle) infinitely extendable sequence; and exactly what is meant by Laplace's idea of sampling with replacement from an infinite population. So it was perhaps inevitable that someone would eventually ask about the corresponding state of affairs if you sample without replacement from a finite population and make the natural assumption that all possible fractions of 0s and 1s are equally likely.

This is what C. D. Broad did in 1918. But as the Bible tells us, "there is nothing new under the sun". The analysis of sampling without replacement from a finite population, using a uniform prior on the fraction of "conformable elements", had already been carried out more than a century earlier, by Prevost and L'Huilier in 1797! In their direct attack on the problem it is necessary to prove a not entirely trivial combinatorial identity in order to establish the rule of succession; see Todhunter (1865 ([2]), pp. 454–457), Zabell (1988).

Recibido / Received: 6 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.

The result was a surprise: as Prevost and L'Huilier note, the rule of succession $(r+1)/(n+2)$ (for a sample of size n from a population of size N) does not depend on N, and is the same of Laplace's! As Todhunter remarks, "The coincidence of the results obtained on the two different hypotheses is remarkable" (1865, p. 457).

But in fact it is not remarkable; from the perspective of the finite de Finetti representation one does not need to evaluate a tricky combinatorial sum, nor is the coincidence of the rules of succesion in any way surprising. A finite exchangeable sequence $X_1, \ldots, X_n$ is completely characterized by the probabilities $p_R = P(S_N = R)$; and so whether the process is generated by first picking p at random from the unit interval and then tossing a p-coin N times, or by randomly selecting an urn $U_{R,N}$ and then drawing its balls out one at a time without replacement, *you get stochastically identical processes* (because in both cases $P(S_N = R) = 1/(1+N)$); it's not just that the rules of succession coincide but *everything* is the same!

It is interesting to trace intellectual dependencies. Broad (1924 ([1]), Section 3) attributes his "interest in the problems of probability and induction" to W. E. Johnson (a Cambridge logician who derived Carnap's "continuum of inductive methods" more than 20 years before Carnap did); and Broad's 1918 analysis was in turn an important influence on Sir Harold Jeffreys, who tells us that "It was the profound analysis in this paper that led to the work of Wrinch and myself" (Jeffreys, 1961, p. 128).

One of the reasons Broad's paper made such a splash at the time was his noting that although (under the uniform prior) the probability that the next crow will be black, given all n crows to date have been black, is nearly one if n is large, $(n+1)/(n+2)$, the probability that *all* crows (in the finite population of N) are black, given the n so far are black, is small for $n \ll N$, namely $(n+1)/(N+2)$.

This property was seen as a problem for any attempt at a mathematical explication of induction, and led Wrinch and Jeffreys to write their papers. As BBS note, in Section 3.2 of his book Jeffreys makes the natural suggestion to allocate some initial probability *independent of* N to natural laws. BBS say "The simplest choice is to let $Pr(R = N) = 1/2$"; but as far as I can tell, Jeffreys usually puts the cases $R = N$ and $R = 0$ on an equal footing. So I would have liked to have seen some further discussion of this suggestion, which clearly treats the cases asymmetrically (since of course if $P(R = N) = P(R = 0) = 1/2$, this would account for all the probability).

There is, however, an interesting historical precedent for viewing matters from such an asymmetric perspective. The Reverend Dr. Richard Price, in his discussion at the end of Bayes's famous essay (Bayes, 1764), considers the application of Bayes's results to the problem of induction. Bayes's version of the rule of succession is different from Laplace's (Bayes's rule is $1 - 2^{-(n+1)}$, a different answer to a different question), but the point here is when Price thinks one should start counting. Bayes's results, he tells us, apply to

> [A]n event about the probability of which, antecedently to trials, we know nothing, that it has happened *once*, and that it is enquired what conclusion we may draw from hence with respect to the probability of it's happening on a second trial.

Note the requirement that the event will have already occurred *once*. Why? Imagining "a solid or die or whose number of sides and constitution we know nothing", Price explains:

> The first throw only shews that *it has* the side then thrown ... . It will appear, therefore, that *after* the first throw and not before, we should be in the circumstances required by the conditions of the present problem, and that the whole effect of this throw would be to bring us into these circumstances. That is: the turning the side first thrown in any subsequent single trial would be an event about the probability or improbability of which we could form no judgment, and of which we should know no more than that it lay somewhere between nothing and certainty. With the second trial then our calculations must begin ... .

This leads Price to consider the famous (or infamous example?) of the rising of the sun:

> Let us imagine to ourselves the case of a person just brought forth into this world and left to collect from his observations the order and course of events what powers and causes take

> place in it. The Sun would, probably, be the first object that would engage his attention; but after losing it the first night he would be entirely ignorant whether he should ever see it again. He would therefore be in the condition of a person making a first experiment about an event entirely unknown to him. But let him see a second appearance or one *return* of the Sun, and an expectation would be raised in him of a second return ... .

I take it that one would adopt a reference prior only absent substantial background information. The interest of Price's remarks is they address in a serious way just how, that being the case, epistemic asymmetry might still be natural. So, as indicated earlier, I would have been interested to see further discussion in BBS of when the assignment $P(R = N) = 1/2$ is appropriate. If, for example, I am a doctor trying out a new procedure or drug, then would I not want some "Jeffreys-like" prior probability assigned to both extremes? Is the reference prior assignment of $1/2$ most appropriate in "Price-like" situations?

One final question for BBS. Suppose there are $t \geq 3$ possibilities (say a, b, c) rather than just the two of conforming and non-conforming. Just as seeing all a thus far should increase the probability that all elements of the population are a, so too seeing, say, only a and b but no c should increase the probability that there are no c in the population. In general there are 2^{t-1} sub-simplexes to which one would like to assign some positive probability. What would be the reference prior approach in this case?

In any case I would like to complement the authors on a most interesting and stimulating paper.

References

[1] BROAD, C. D., (1924). *Philosophical Autobiography*, Contemporary British Philosophy: Personal Statements, First Series (ed. J. H. Muirhead), G. Allen and Unwin, London, 77–100.

[2] TODHUNTER, I., (1865). *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace*, Macmillan, London.

S. L. Zabell
Departments of Mathematics and Statistics,
Northwestern University,
IL, USA
zabell@math.northwestern.edu