RAC$\mathbb{SAM}$
Rev. R. Acad. Cien. Serie A. Mat.
VOL. **103** (1), 2009, pp. 157–159
Estadística e Investigación Operativa / Statistics and Operations Research
Réplica / Rejoinder

Rejoinder on: Natural Induction: An Objective Bayesian Approach

J. O. Berger, J. M. Bernardo and D. Sun

We appreciate the positive general comments of most of the discussants. And, of course, we are grateful for the interesting and thought-provoking additional insights and comments that they have provided. We provide below a response to these comments.

Girón and Moreno. We certainly agree with Professors Girón and Moreno on the interest in sensitivity of *any* Bayesian result to changes in the prior. That said, we also consider of considerable pragmatic importance to be able to single out a *unique*, particular prior which may reasonably be proposed as the reference prior for the problem under study, in the sense that the corresponding posterior of the quantity of interest could be routinely used in practice when no useful prior information is available or acceptable. This is precisely what we have tried to do for the twin problems of the rule of succession and the law of natural induction.

The discussants consider the limiting binomial version of the Law of Natural Induction, and focused on the version that can be stated in the language of hypothesis testing involving $H_0 \equiv \{p = 1\}$. They then noted that a popular objective Bayesian approach to hypothesis testing is to use intrinsic priors on the alternative, which tend to be more concentrated about the null value than the $\mathrm{Be}(p \mid 1/2, 1/2)$ prior we use. The notion is that, if a problem is posed as that of testing a 'privileged' null hypothesis, then realistic alternatives will tend to be close to the null value, and the prior distribution – even in supposedly objective procedures – should reflect this. Thus a strong case can be made for use of intrinsic priors in that setting.

The natural induction problem, however, is not a problem with a privileged null hypothesis in this sense; there is no a-priori notion that values of p near 1 are more believable than other values. Hence we would argue that, for the natural induction problem, the analysis we propose is the preferred objective procedure.

Lindley. As one would expect from a well known subjectivist, Professor Lindley questions our use of the word *objective*. Although we certainly agree that *any* statistical analysis is really subjective (for the crucial model assumptions are typically made from subjective choices) we use the term objective in the precise sense that the Bayesian result obtained only depends on the model assumed and the data obtained and, therefore, has precisely the same claim to objectivity that is often made of frequentist statistics. Unfortunately, Bayesian methods are still often disregarded because of the completely wrong impression that they *must* use subjective priors; by use of the word objective we simply want to stress that Bayesianism need not be rejected for this reason.

Professor Lindley is certainly right in pointing out that we have assumed that R and n are independent given N, so that a more appropriate notation would have been $\Pr(R \mid n, N)$ rather than $\Pr(R \mid N)$. Indeed,

Recibido / Received: 25 de marzo de 2009.
This rejoinder 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.

an inverse sampling procedure would have produced different results. We have certainly worked conditional on N so that, if N is not really known, a range of results must be quoted, as in the galápagos example. That said, there are many important applications where N is actually known. For instance, in industrial production, one may be interested in the probability that all items in a production of size N are conforming, given that the elements of a random sample of size n are all conforming, and the value of N is typically known in those cases.

As Professor Lindley interestingly points out, our rule of succession gives the same result as Laplace, but for twice the sample size. Considering the practical implications of this fact (for instance when setting insurance policies in the industrial situations described above), the situation is a good example of the possible rewards in careful choice of the reference prior.

Liseo. As Professor Liseo points out, model choice priors (or precise hypothesis testing priors in the language we use in the paper) must be different from those of estimation. Proposed hypothesis typically have a precise physical meaning, and its possible validity has to be reflected in the prior. One may choose to do this subjectively, or one may try an objective approach using a reference prior where the quantity of interest is whether or not the hypothesis is true, as we have tried to do in the problem of natural induction.

Professor Liseo poses the question of what to do if the application suggests a hypothesis of the type "R is close to N". In line with the argument above we would suggest eliciting a value $0 < \alpha < 1$ such that the closest integer to αN is close enough to N for the application in mind, and then use a marginal reference prior with $\Pr(\alpha N \leq R \leq N) = 1/2$, and a conditional reference prior for R given $0 \leq R < \alpha N$ which would be the conditional reference prior (14) renormalized to the set $0 \leq R < \alpha N$.

Raman. Professor Raman is certainly right when he points out the fact that the derivation of reference priors is often less than trivial (although, their *use* is often trivial, once someone has done the relevant research and produced the appropriate answers.) That said, it is certainly interesting to be able to derive some specific instances of the reference analysis results using simpler techniques. For instance, invariance arguments provide simple and powerful techniques to provide objective priors (which are typically found to be the same as those derived from reference analysis), but many important problems do not have the required structure to use them; reference analysis provides a general argument, which does not depend of particular properties of the problem under investigation.

As Professor Raman points out, the continuous version of our rule of succession is the particular case which obtains if a conjugate Beta prior is used, and the particular case $\alpha = \beta = 1/2$ is chosen; he further outlines how this can interestingly be extended to the use of beta mixtures. The argument by Polya provides some additional reasons to choose a conjugate prior in this problem, beyond mathematical expediency, but one would need some special argument to select precisely $\alpha = \beta = 1/2$, if one wanted to argue that $(n + 1/2)/(n + 1)$ is indeed the appropriate rule of succession.

Professor Raman seems surprised by the fast convergence to zero of the reference probability $\pi_r(N \mid N)$. This is a reflection of the fact that, as N increases, its precise value does not matter much. Besides, reference priors should not really be analyzed as prior probabilities (indeed, in continuous parameter problems those are typically improper) but as technical devices to produce posterior probabilities. And it this the reference posterior probabilities what should be carefully discussed for any possible data set.

Robert. Professor Robert focuses attention on extending the problem to consider the case where N is unknown, a situation also mentioned by Professor Lindley. The objective Bayesian answer is indeed to specify an objective prior for N and, as Professor Robert points out, the intuitive choice for an objective prior for the scale-like parameter N would be $\pi(N) = 1/N$. Given the time restrictions imposed by the Journal, we have not had time to explore the consequences of this choice, let alone to derive the appropriate conditional reference prior for N, $\pi(N \mid R)$, when R is the quantity of interest, but we certainly take note of this problem for future research.

As Professor Robert points out, it is not difficult to implement Jeffreys suggestion for the prior probabilities of $R = 0$ and $R = N$, together with the conditional reference prior for $\pi(R \mid n, N)$ but, if one is to follow the reference analysis approach, one should use a marginal reference prior of the quantity of interest, which here is whether or not $R = N$, and this leads to $\pi(R = N) = \pi(R \neq N) = 1/2$ rather than to the values which Jeffreys suggested, namely $\pi(R = 0) = \pi(R = N) = k$, with $1/3 \leq k \leq 1/2$. As already raised in the response to Professors Girón and Moreno, the issue is partly one of when and how one should formally admit subjective knowledge into objective procedures. By incorporating into the reference analysis the knowledge that $R = N$ is a clearly important possibility, we went part way to Jeffreys solution. It seemed to us, however, that the contexts in which the Laws are discussed are not contexts in which $R = 0$ has any special believability.

Rueda. Professor Rueda is certainly right when he points out that the exchangeability argument only implies a hierachical model with a a prior distribution $\pi(\theta)$, which could be the $\mathrm{Be}(\theta \mid 1/2, 1/2)$ we use, the uniform prior leading to Laplace rule, or any other proper prior for θ. Our particular choice for $\pi(\theta)$ is motivated by the fact that this is *the* reference prior which corresponds to the *implied* integrated model which, as argued in the paper, is the Binomial $\mathrm{Bi}(r \mid n, \theta)$.

The very different behavior of (9) and (11) —and, correspondingly, (18) and (20)— as the population size N increases is mathematically driven by nature of the priors used, with a prior probability for $R = N$ which depends on N and goes to 0 as N increases. If R is the only parameter of interest, these are indeed the appropriate results for the problems considered. If, however, the quantity of interest is whether or not $R = N$, then a prior probability for $R = N$ which does not depend on N is required, and the results, (25) and (22), are very noticeably different.

Zabell. We are very grateful to Professor Zabell for his authorative insights into the history of the problems considered in this paper: we were not aware of the many precedents to Broad's derivation of the Laplace rule of succession for the finite case. We are similarly not aware of any previous derivation of the reference rule of succession for the finite case given by Equation (19) (a result given by Perks (1947) for the continuous case), but then, we would not be surprised if there is one!

As also noted by Professor Robert, Professor Zabell wonders about the advantages of using a symmetric prior for $\Pr(R = 0)$ and $\Pr(R = N)$. As mentioned in our response to Professor Robert, such decisions are indeed context dependent; there is subjectivity in terms of how one chooses to formalize a reference analysis and, in a situation such as drug testing, we would certainly propose a different formulation than that which we felt to be reasonable for the Laws.

As Professor Zabell correctly points out, one should only use a reference prior in the absence of public, commonly accepted substantial background information. If there is such information, this should be explicitly stated and be made part of the model. Often, this will take the form of a hierarchical model, and then one would need an objective prior for the hyperparameter, possibly the reference prior associated to the corresponding integrated model.

The extension of the results presented to more than two possibilities in the population is a very interesting possible generalization. We expect that the same basic reference analysis techniques could be used for this generalization.

Acknowledgement. We would like to conclude by thanking again our discussants and, last but not least, by thanking Professor Pellicer and the staff of RACSAM for inviting us to present this work, and for processing the paper in a record time.

James O. Berger
Duke University,
USA
berger@stat.duke.edu

José M. Bernardo
Universitat de València,
Spain
jose.m.bernardo@uv.es

Dongchu Sun
University of Missouri-Columbia,
USA
sund@missouri.edu