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