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 Jeffreys 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

---

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

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

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*

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

and

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

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