*Departament d'Estadística i I.O., Universitat de València.*
*Facultat de Matemàtiques, 46100–Burjassot, València, Spain.*
*Tel. 34.6.386.4314 (direct), 34.6.386.4362 (office); Fax 34.6.386.4735,*
*Internet: bernardo@uv.es, Web: http://www.uv.es/˜bernardo/*

Printed on February 26, 1997
To appear in *J. Statist. Planning and Inference*

# Rejoinder

JOSÉ M. BERNARDO

*Universitat de València, Spain*

This is a fine set of discussions, and I am very grateful to their authors for their illuminating contributions to our understanding of a very polemic topic. Predictably, "truth" is partially perceived by some discussants from a different perspective than mine, but their refreshingly sincere attitude is bound to help to clarify some of the more relevant issues. I will try to answer individually the queries which have been raised.

### *Reply to Professor Cox*

I totally agree with Professor Cox on the importance of Jeffreys' work; not only did he pioneered a successful use of nonsubjective prior distributions, but he produced a rule which, in the regular one-parameter case, —the only case for which he strongly recommended its use—, it is still regarded as 'the' appropriate solution. An scholarly account of his developments previous to his 1946 famous paper would certainly be very welcome by the statistical community.

It would have been a surprise if Professor Cox gave to foundational arguments the considerable weight Bayesians believe they deserve and, although the topic is clearly too deep to be adequately dealt with here, I welcome the opportunity he gives me to expand on it: (i) I fail to see the need for different types of uncertainty: indeed, I see as one of the strengths of the Bayesian approach that its notion of probability encompasses its semantic use, as well as those related to symmetries or to replications; (ii) it is certainly true that axiomatics of rational behaviour typically include a clearly idealized assumption on the comparability of probabilities, but then no physicist would claim the ability to measure with infinite precision, and yet physical measurements are assumed to be real numbers: I believe on the usefulness of a *prescriptive* theory which assumes precise probabilities, —taken with a large pinch of salt and a great deal of sensitivity analysis—; (iii) the representation theorems are mathematical facts which only depend on the exchangeability assumption and do *not* depend of the particular view one might have on probability; although it is true that not all problems may be represented within this structure, *all statistical analysis which assume a random sample* from one model or another, —and pragmatically those are an overwhelming majority—, *are a particular case of the exchangeability structure* and, hence, they *require* a prior distribution for its logically correct analysis.

---

José M. Bernardo is Professor of Statistics at the University of Valencia. Research partially funded with grant PB93-1204 of the DGICYT, Madrid, Spain.

As Professor Cox remarks, I regard as crucial the role played by the logarithmic concept of information in statistics; not only may it be used to define nonsubjective priors, but much more generally, it provides a foundational basis to encompass statistical inference within decision theory (Bernardo, 1979), and a natural definition of the goodness of a probabilistic approximation in the very many instances where such a concept is required in statistics (Bernardo, 1987).

Finally, I would like to thank Professor Cox for drawing my attention to the problem posed by Box and Cox (1964) transformation families; this is indeed a very interesting problem and I am including the derivation of the corresponding reference posteriors within my 'to do' file.

*Reply to Professor Dawid*

Professor Dawid's main criticism to the use of reference priors is foundational: since the reference priors depend on the parameter of interest, focusing on different aspects of the problem would lead to different priors and, hence, to inconsistent results. The argument would indeed be devastating if reference priors were supposed to describe unique, personal, possibly 'diffuse' beliefs, but there are *not*! Reference analysis *must* be regarded as part of a *sensitivity* analysis to the choice of the prior. Reference analysis clearly establishes that you cannot *simultaneously* have a prior which is minimally informative with respect to, say the $\mu_i$'s of a multinormal model *and* with respect to the sum of its squares, $\sum \mu_i^2$, hence the title of this paper. Consequently, a reference posterior *must* be regarded as the answer to a precise question on sensitivity: *if* I wanted to use a prior minimally informative with respect to $\phi$, *then* $\pi(\phi \mid \boldsymbol{z})$ would encapsulate my inferences abut $\phi$. In a decision situation, where a unique prior must indeed be used, a class of reference priors for several parameters of interest may usefully be considered to help to understand the implications of the particular prior one is going to use, by precisely making explicit the possibly important judgements which such a prior implies about specific functions of the parameters.

Professor Dawid is certainly right when he mentions that it has not been *proven* that reference analysis always avoids the marginalisation paradoxes; the problem is that the marginalisation paradoxes are described by a set of examples, with no unifying theory on the general conditions which may produce them. The fact remains however, that 25 years after they were discovered, no marginalization paradox has ever been encountered using reference priors, and that reference analysis is the only method to derive nonsubjective priors which successfully avoids the paradoxes.

Finally, I appreciate Professor Dawid's warning on the delicate aspects involved in the approximation of improper priors by a sequence of proper priors. This is precisely the reason behind the apparently involved definition of a reference *posterior*, —as the limit, in the logarithmic divergence sense, of the sequence of posterior distributions obtained using Bayes theorem on a sequence of proper priors—, rather than attempting a direct definition in terms of a limit of the priors themselves.

*Reply to Professor Ghosh*

As Professor Ghosh mentions, reference priors in multiparameter settings may formally be defined with respect to any ordered set of parameter subgroups, but it is only by sequentially using the one-parameter solution, —the 'one-at-a-time' reference prior—, that satisfactory results are obtained. This could be expected both from the analysis of the two-parameter problems and from information-theoretical arguments, and has been precisely argued in Datta and Ghosh (1995) in problems which are invariant under an amenable group of transformations.

This is why in multiparameter situations, I refer, to the one-at-a-time reference prior as 'the' reference prior (see *e.g.*, the answer to question 35).

The question of the conditions under which an improper prior leads to a proper posterior has not yet found a general answer. As Professor Ghosh mentions, in all examples reference posteriors *given a minimum size sample* have been found to be proper. It may reasonably be expected that general results could be established from the general definition of reference posteriors as limits of posteriors obtained from proper priors which I have just mentioned above, but I am not aware of any. In the example he mentions, the minimum size sample is clearly one non-zero observation, for zeros do not provide any information about $\theta$; hence, as one would expect, the reference posterior of $\theta$ will not be proper until such a non-zero observation has been obtained, precisely indicating that there is nothing to be said about $\theta$ solely based on those data and the assumed model.

I very much welcome Professor Ghosh's insightful comments on the mathematics which, in the regular case, operate behind the maximization of the missing information required by the definition of a reference prior. However, I would like to stress that such a maximization may *also* be performed in non-regular cases, —where Fisher's matrix may not even be defined—, and that, even in regular cases, it is often simpler, —as in the example above—, to derive directly the form of the asymptotic posterior distribution from first principles, that it is to check the regularity conditions and obtain its form from Fisher's matrix.

The frequentist validation of reference posterior statements is important both theoretically, —to guarantee that no inconsistencies may exist—, and pragmatically, —to establish bridges with non-Bayesian statisticians. As he mentions, more work remain to be done; I would specially like to draw attention to the need for further work with *small* samples; indeed available results only explain the good coverage properties of reference posterior regions in *asymptotic* conditions and, yet, simulations repeatedly suggest that, with continuous parameters, very good coverage properties are indeed obtained with *any* sample size.

Finally, the necessary approximation of open parameter spaces by convergent compact sequences in order to derive the reference distributions certainly requires further work. I believe one should always consider a probability model *endowed with an appropriate compact approximation* to its parameter space, which should then be kept *fixed*, via the appropriate transformations, for *all* inference problems considered within that model. A good candidate for such 'canonical' compact endowment could be the natural uniform approximation in the corresponding variance-stabilizing transformations (Bernardo, 1997).

*Reply to Professor Lindley*

Professor Lindley stresses the basic foundational arguments of Bayesian decision theory to argue that a prior is an expression of someone's beliefs and should therefore be independent of the model used. I certainly accept the formal argument, and I agree that in a decision making situation, assessing a prior reflecting the his or herknowledge is precisely what any decision-maker should *try* to do. However, a reliable direct probabilistic description of complex multivariate information is next to impossible, so that, when data may be expected to dominate the prior, one may be prepared to approximately describe one's prior information as minimally informative with respect to some specific aspects of the problem, if only as an insurance policy against a multivariate prior unsuspectedly overwhelming the information from the data in specific directions of interest. In that case, mathematics, —the reference algorithm—, could be used to transform this prior statement into a prior distribution (and the procedure could presumably be included within a project on multivariate assessment funded under Professor Lindley's guidance!). Moreover, as we move away form personal decision making and concentrate in scientific reporting, it is obvious to me that one is *forced* to perform some form of sensitivity analysis with respect to changes in the prior, with special emphasis in minimally-informative prior situations, as I have tried to describe in my reply to Professor Dawid's comments; I believe that reference analysis does provide an appropriate mechanism for this type of work.

Professor Lindley further insists on the *physical*, real-world meaning of the parameters. I believe that the existence of such a meaning is the exception, not the rule; what is the physical meaning of, say, the many parameters with appear in a complex hierarchical log-linear model? The representation theorems guarantee that exchangeable observations may be regarded as a random sample from some probability model, whose parameters are *defined* as a limit of observables, and hence, are unobservable themselves; direct assessment a probability distribution on an unobservable quantity cannot be made operational and, hence, a programme on subjective probability assessment about the parameters of a model is fraught with difficulties.

Professor Lindley also stresses the required propriety of the prior; as I have argued above, the whole theory of reference distributions is actually based on *proper* priors; it is only in the last step that limits are taken for mathematical tractability. It is a fact, however, that if the reference analysis is kept proper by working within the appropriate compact subsets and performing by simulation the required integrations, the results are numerically indistinguishable from those analytically available using the limiting form.

I can only agree with Professor Lindley's request for scientific repeatibility, but I disagree with his conclusions: the coverage probabilities of reference posterior regions have the kind of repeating properties that scientists often require, and this is something you cannot possibly obtain from subjective priors. With respect to sequential experimentation, it is less than obvious to me that one would always want to use as a prior the posterior from the last experiment; indeed there is here a problem of temporal coherence: too many things would have typically changed between experiments for the same 'small world' contemplated by Bayes theorem to remain valid. Again, I would prefer an analysis of the present experiment from a 'what if', sensitivity analysis perspective, whithin which, exploring the consequences of assuming minimal information about the main quantity of interest may well provide the more useful results.

I would like to close by thanking again all the discussants by their thought provoking comments, and by thanking Professor Singpurwalla for offering me this opportunity for a stimulating debate.

## REFERENCES

Bernardo, J. M. (1979). Expected information as expected utility. *Ann. Statist.* **7**, 686–690.

Bernardo, J. M. (1987). Approximations in statistics from a decision-theoretical viewpoint. *Probability and Bayesian Statistics* (R. Viertl, ed.). London: Plenum, 53–60.

Bernardo, J. M. (1997). Comment to 'Exponential and Bayesian conjugate families: review and extensions', by E. Gutiérrez-Peña and A. F. M. Smith. *Test* **6** (to appear).

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *J. Roy. Statist. Soc. B* **26**, 211–252 (with discussion).