Reply to “Comment on ‘Reference priors for high energy physics’ ”

Luc Demortier, Supriya Jain, and Harrison B. Prosper

Phys. Rev. D **84**, 118502 — Published 2 December 2011

DOI: 10.1103/PhysRevD.84.118502

Reply to “Comment on ‘Reference priors for high energy physics’ ”

Luc Demortier
Laboratory of Experimental High Energy Physics,
The Rockefeller University, New York, New York 10065

Supriya Jain
Department of Physics, State University of New York, Buffalo, New York 14260

Harrison B. Prosper
Department of Physics, Florida State University, Tallahassee, Florida 32306
(Dated: November 3, 2011)

In their Comment, Grientschnig and Lira derive a Bayesian reference prior for cross section measurements that turns out to be identical to one obtained by L. Demortier, S. Jain, and H. B. Prosper [Phys. Rev. D **80**, 034002 (2010)]. Since the new derivation uses a different sequence of compact sets to normalize the reference prior, this result points to a certain robustness of posterior inferences in this approach. However, Grientschnig and Lira make some additional claims about their methodology, which we show to be unwarranted.

PACS numbers: 06.20.Dk,02.50.Tt

I. INTRODUCTION

In the abstract and introduction of their Comment, Grientschnig and Lira assert that the reference prior and posterior derived in our paper [1] were found to depend on the sequence of compact subsets of parameter space used for a normalization procedure. This is an incorrect reading of our paper. Indeed, we merely discovered that a simple, plausible choice of subsets leads to a posterior that is improper for some realizations of the underlying model *and must therefore be dismissed*. We then proposed a different sequence of subsets, found the resulting posterior to be proper, and presented the latter as a usable reference posterior for the problem at hand. We did not address the notoriously difficult question of the uniqueness of this posterior with respect to changes in compact set sequence. The fact that Grientschnig and Lira have now derived the same posterior from a different sequence is of course an encouraging result from the point of view of uniqueness.

In our opinion however, Grientschnig and Lira go one step too far by stating that their analysis “questions the manner in which [our result] was achieved.” They base this statement on the fact that we did not make use of “a sufficient condition for uniquely defining the reference prior”, and on their belief that “if [the reference prior] were not unique an unsatisfactory ambiguity of the interpretation of measurement results would arise”. They subsequently address the uniqueness problem by making two claims, (1) that in every problem of inference one can identify an “original parametrization”, and (2) that it is this original parametrization that should be used to construct the compact sets. They refer to this second claim as a “principle”, and as the “original parametrization rule”. As we will show, these claims are neither compelling nor supported by the literature on reference priors. We briefly review what this literature has to say about compact sets in Sec. II, expand on the motivation for our choice of compact sets in Sec. III, and return to the claims of Grientschnig and Lira in Sec. IV.

II. COMPACT SETS IN THE LITERATURE ON REFERENCE PRIORS

Although compact sets are technically not needed to obtain reference priors for single-parameter models [2], there seems to be no way to avoid them in multi-parameter settings. This feature can be traced back to the stepwise nature of the general reference prior construction [3]. In brief, when multiple parameters are present, they must first be sorted by order of inferential importance. Next, at each step of the algorithm one constructs a conditional reference prior for one parameter given all the remaining ones in the ordered sequence. If the parameter space Θ is unbounded, this conditional prior is typically improper, and therefore only defined up to a proportionality constant that may depend on the parameters that were conditioned on. To fix this dependence, one introduces a sequence of nested compact sets Θ_ℓ that converges to the whole space Θ, and normalizes the conditional prior over an arbitrary Θ_ℓ. After all steps are completed, the reference posterior on Θ_ℓ is obtained from Bayes' theorem, and the limit $\Theta_\ell \to \Theta$ is taken.

Whereas the compact set normalization procedure arises as a purely mathematical device to resolve an indeterminacy in the reference prior construction, the compact sets themselves do have a more general meaning, regardless of whether the model has one parameter or many. Indeed, in all practical measurement problems, the physical parameter space that one is trying to model is actually bounded. However, it is usually not easy to specify its boundary, and one therefore uses the unbounded space as an approximation to the physical space [2]. Thus Bernardo writes that "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" [4, p. 187 (emphasis in the original)].

How then should the compact sets be chosen? Berger and Bernardo admit they "have no clear-cut answer to this question, precisely because a dependence of the solution on the Θ_ℓ is essentially an indication that some subjective input is needed; one cannot unambiguously define a reference prior" [5, p. 205]. Nevertheless, in Ref. [6, p. 42] they provide the following guideline: "Choosing the Θ_ℓ to be natural sets in the original parametrization has always worked well in our experience. Indeed, the way we think of the Θ_ℓ is that there is some large compact set on which we are really noninformative, but we are unable to specify the size of this set. We might, however, be able to specify a shape, Ω, for this set, and would then choose $\Theta_\ell = \ell\,\Omega \cap \Theta$, where $\ell\,\Omega$ consists of all points in Ω multiplied by ℓ." In situations more complex than location-scale problems however, Bernardo introduces approximate location parametrizations, which he defines as parametrizations for which the reference prior is uniform. He then suggests that "reparametrization to asymptotically independent parameters and approximate location reparametrizations [. . .] may be combined to choose appropriate approximating sequences" [3, §3.8] (see also Ref. [4, p. 187], which points to [7]).

As the above summary indicates, the problem of choosing compact sets, while not a major issue in the theory of reference priors, is far from settled. Describing directions for further research in Ref. [4, p. 171], Bernardo lists in first place the need to obtain "a general definition of the appropriate bounded approximation". Later in the same paper he adds that "the necessary approximation of open parameter spaces by convergent compact sequences in order to derive the reference distributions certainly requires further work" [4, p. 187].

III. THE CHOICE OF COMPACT SETS IN OUR PAPER

In [1] we wrote: "The theory of reference priors currently does not provide guidelines for choosing the compact sets Θ_ℓ, other than to require that the resulting posterior be proper. In most cases this choice makes no difference and one is free to base the choice of compact sets on considerations of simplicity and convenience." Simplicity and convenience were our way to summarize Berger and Bernardo's suggestion that "most frequently it would probably be natural to choose the Θ_ℓ to be simple sets (rectangles, spheres, etc.) in the original parametrization of the problem; initially chosen parametrizations are often ones in which the analyst is roughly noninformative over natural sets" [5, p. 205]. Since our initial parametrization was in terms of the signal cross section σ, the effective integrated luminosity ϵ, and the background contamination μ, we simply constructed rectangular sets of the form

$$\Theta_\ell \;=\; \Big\{(\sigma, \epsilon, \mu) :\; \sigma \in [0, u_\ell],\; \epsilon \in [0, v_\ell],\; \mu \in [0, w_\ell]\Big\}, \tag{1}$$

where $\{u_\ell\}$, $\{v_\ell\}$, and $\{w_\ell\}$ are increasing sequences of positive constants. This led to the following conditional reference prior for σ given ϵ and μ:

$$\pi_R(\sigma \,|\, \epsilon, \mu)\, d\sigma \;\propto\; \sqrt{\frac{\epsilon}{\epsilon\,\sigma + \mu}}\, d\sigma. \tag{2}$$

When we discovered that this prior leads to an improper posterior for some form of the evidence-based prior for ϵ, we asked Berger and Sun, the authors of [8], for advice. They did not find fault with our interpretation of "original parametrization", but noted that the resulting prior (2) depends on the parameter of interest σ only through the product $\epsilon\sigma$. Any proper prior with this property would be invariant under multiplication of σ and division of ϵ by the same constant, that is, it would be of the form $\epsilon\, f(\epsilon\sigma)\, d\sigma$, which (2) isn't. Berger and Sun then suggested that we construct rectangular sets in the parametrization $\{\tilde\sigma, \epsilon, \mu\}$, where $\tilde\sigma \equiv \epsilon\sigma$. As indicated in our paper, the resulting prior,

$$\pi_{R1}(\sigma \,|\, \epsilon, \mu)\, d\sigma \;\propto\; \frac{\epsilon}{\sqrt{\epsilon\sigma + \mu}}\, d\sigma, \tag{3}$$

has the desired invariance property and leads to a posterior that is proper regardless of the form of the evidence-based prior for the problem.

IV. EVALUATION OF THE CLAIMS MADE IN THE COMMENT

Our overview of the literature on compact sets failed to find a compelling and unambiguous "principle" or "general rule" to settle their choice. What we did find is an admission that some subjective input may be needed, a statement that further work is required, and a couple of guidelines, namely to try a "natural" parametrization, or perhaps an approximate location-scale parametrization. The exact meaning of "natural" is left largely unspecified. On the other hand, Grientschnig and Lira claim to have identified a family of models where the concept of natural parametrization is unambiguous. They write, unequivocally, "If the parameter of interest is a function of parameters originating from different sampling distributions that have no parameters in common, the original parametrization consists of these parameters." This is the basis of their derivation of the reference prior for cross section measurements.

As shown in Sec. III, we started from a different interpretation of "natural parametrization", but eventually obtained the same result as Grientschnig and Lira. They claim that their derivation questions the manner in which our result was achieved. For this to be true, it seems to us that Grientschnig and Lira should demonstrate that their definition of original parametrization has special merits, for example that it can be derived from foundational considerations or that it is guaranteed to deliver a posterior with good properties within the family of models they identified. Unfortunately they do not go beyond re-deriving our original result.

In the final analysis, if two sequences of compact sets were to lead to different posteriors for the same problem (see Ref. [5] for an example), how should one decide which sequence is best? Would it be acceptable to prefer one sequence simply because it is based on a general principle? Or should the decision be based instead on an objective assessment of the properties of the resulting posteriors: whether they are proper, have acceptable repeated sampling properties, yield posterior expectations that are consistent, etc. It is the latter approach that Berger and Bernardo followed in Ref. [5] and that we took in our paper. Thus, while the adoption of a single general rule for the construction of compact sets would guarantee the uniqueness of the resulting posterior, it would be premature to do this without a better theoretical understanding of the effect of various rules on posterior inferences. Given the current lack of such understanding, we caution users of reference priors against the mindless application of general guidelines.

Finally, we are puzzled by the criticism Grientschnig and Lira direct against a prior of the form $\exp(-\epsilon)/\sqrt{\pi\epsilon}$ for the effective luminosity ϵ (Sec. III of the Comment). They correctly point out that the prior probability *density* of ϵ is infinite at $\epsilon = 0$, and that the cross section σ cannot be measured when $\epsilon = 0$. However, the occurrence of $\epsilon = 0$ has zero prior *probability* and will therefore *almost surely* not happen. In any case the density $\exp(-\epsilon)/\sqrt{\pi\epsilon}$ is a valid prior. For an evidence-based interpretation, one can think of it as the reference posterior of an auxiliary measurement of ϵ, derived from the Poisson likelihood $\exp(-\epsilon)\epsilon^x/x!$ and Jeffreys' rule $1/\sqrt{\epsilon}$, in the (admittedly undesirable) situation where no events are observed in the time allotted to the measurement ($x = 0$). Once the acceptability of $\exp(-\epsilon)/\sqrt{\pi\epsilon}$ as a prior is recognized, any method that claims to be broadly applicable and general ought to be able to deal with it. As we showed in [1], this is the case of the reference prior method, provided one uses the compact set sequence we proposed.

Grientschnig and Lira subsequently express concern that a different choice of evidence-based prior may require a change in compact set sequence in order to achieve posterior propriety, and that this "contradicts a basic property of reference priors with partial information, namely that they do not depend on the probability distribution encoding the partial information." This objection puts the cart before the horse: the basic property it refers to only applies *after* an appropriate sequence of compact sets has been found [8]; hence the supposed contradiction cannot take place.

In summary, we still do not have a general rule for constructing a compact set sequence that *guarantees* a proper and well-behaved posterior. Fortunately, given the large number of problems to which the reference prior methodology has been successfully applied, this compact set difficulty appears to be of minor consequence in practice.

Acknowledgments

This work was supported in part by the U.S. Department of Energy under grant Nos. DE-FG02-91ER40651 and DE-FG02-97ER41022, and by the National Science Foundation under grant No. PHY-0855637.

[1] L. Demortier, S. Jain, and H. B. Prosper, Phys. Rev. D **82**, 034002 (2010).
[2] J. O. Berger, J. M. Bernardo, and D. Sun, Ann. Statist. **37**, 905 (2009), http://www.uv.es/~bernardo/2009Annals.pdf.
[3] J. M. Bernardo, in *Handbook of Statistics 25*, edited by D. K. Dey and C. R. Rao (Elsevier, Amsterdam, 2005), pp. 17-90, http://www.uv.es/~bernardo/RefAna.pdf.
[4] J. M. Bernardo, J. Statist. Planning and Inference **65**, 159 (1997).
[5] J. O. Berger and J. M. Bernardo, J. Amer. Statist. Assoc. **84**, 200 (1989); in this and other quotations from the literature we have taken the liberty to relabel the compact sets as Θ_ℓ, for consistency within this Reply.
[6] J. O. Berger and J. M. Bernardo, in *Bayesian Statistics 4*, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Oxford University Press, Oxford, 1992), pp. 35-60, http://www.uv.es/~bernardo/1992Valencia4Ref.pdf.
[7] J. M. Bernardo, Comment on "Exponential and Bayesian conjugate families: review and extensions," by E. Gutiérrez-Peña and A. F. M. Smith, Test **6**, 70 (1997).
[8] D. Sun and J. O. Berger, Biometrika **85**, 55 (1998), http://www.stat.duke.edu/~berger/papers/sun.html.