Discussion with José Bernardo on Bayesian reference analysis

Transcribed and edited by Luc Demortier (Rockefeller University)

Abstract

The discussion session that followed José Bernardo's talk was devoted to Bayesian reference analysis. Questions were asked about the use of sharp priors in Bayesian hypothesis tests, the relationship between objective Bayesian and frequentist testing, the construction, computation, and interpretation of reference posteriors, the combination of measurements, and sensitivity analysis. As the discussion unfolded, comments were made about more general issues such as the hierarchy of problem formulations, and somewhat esoteric topics such as the ratio of normal means.

Luc Demortier

Let's first take questions about José Bernardo's talk, and then later we'll open the floor to more general questions about reference priors and reference analysis.

Kyle Cranmer (New York University)

I have a question about invariance under reparametrization. I can specify a model parametrized, say, in θ, or in α. There is the procedure that you follow and you have all the various invariance properties. But what I am confused about is that, if you need to specify this ordering to construct the reference prior, what does it even mean to talk about invariance under reparametrization? Isn't that meaningless, especially if the transformation from θ to α is "weird"? So how do you talk about invariance, when you need ordering?

José Bernardo (University of Valencia)

Two answers. First of all, the real differences are hierarchical, in that it is important to select what the parameter of interest is. There are slight differences in the results depending on the ordering of the other parameters, but it is really the specified parameter of interest which matters, and the whole thing is defined in such a way that it is invariant under one-to-one transformations of this parameter of interest.

Apart from that I agree that often you have several parameters of interest, you can decide how many. Then you can try to get a global approximation, a reference prior which will not be *the* reference prior for each of these things as the parameter of interest, but will be such that the marginals of the joint posterior that you will get will not be far off of the corresponding reference posteriors. We have found in many examples that this actually works, the differences are so minute that when you have ten observations you don't see them. So in practical terms it's not that important. Of course if you have very few parameters then it does matter, but then it should matter. For instance if you are in a normal situation, with parameters μ and σ, and you are not interested in either μ or σ, but in μ/σ, so that the parameter of interest is the signal to noise ratio, then it does matter. In that case it doesn't make sense to use the global approximation because you can get the analytical answer for the quantity you want. However, in many-parameter problems with many quantities of interest, you'd better do something like a global approximation because otherwise the whole thing becomes unmanageable.

Glen Cowan (Royal Holloway)

When you look at what you call the objective Bayesian hypothesis testing and you write down this intrinsic test statistic, this d of H_0 given the data, is there an easy way of seeing how a test based on that statistic maps onto a classical test in terms of its power?

José Bernardo

Yes, this may be done. In fact, under asymptotic conditions, it's very easy, because if you have enough regularity to get a posterior distribution which is asymptotically normal, then what happens is that you are in an situation much like testing $\mu = \mu_0$ in a normal distribution. And in this particular example what you get is that the intrinsic statistic d is actually $(1 + t^2)/2$, where t is the number of standard deviations out. In asymptotic conditions, this will be approximately true in any problem, so if $t = 3$, then $d = 5$. Thus $d = 5$ is exactly equivalent to three standard deviations in the normal conditions. Of course this will be only an approximation if the data are not normal, but it will be a good approximation if the posterior of the quantity of interest is asymptotically normal, and you will get an immediate feeling for how it works. Actually you do get that d, the intrinsic statistic, is very often a (sometimes approximate) one-to-one transformation of some already known statistic. But the Bayesian analysis gives you an interpretation which is immediate and unique in terms of expected log-likelihood ratios, and you don't have to find the sampling distribution of anything. With the same argument, the often suggested five standard deviations gives $d = 13$, so if you get an intrinsic statistic around thirteen, it is as if you were five standard deviations off in the normal case.

Frederik Beaujean (Max Planck Institute for Physics)

I want to ask about your theorem 1 where you showed the example of how one can numerically construct the reference prior in one-dimensional problems. You said for large k this is good. Do you have any idea when k is large enough? If I have a few data points, is k large or not large?

José Bernardo

Well, I think this must be problem-dependent, but k around 1000 will often suffice. I have played with this in many different applications, like election forecasting and industrial quality control, and then I have always found that for the sample sizes that people use, this kind of approximation to the reference prior is more than enough. However, to be sure you would really have to do the numbers and see, by simulation probably, how far you would be. It should be possible to prove general things for conditions that guarantee that the approximation is good but, as far as I know, that has not been done.

Glen Cowan

So this is a point that went by quickly and I just wanted to make sure I understand this. You are saying that putting a sharp prior is bad?

José Bernardo

No, not necessarily bad, I am saying it is different, and that using a sharp prior might be dangerous in specific circumstances, that you should check. Putting a sharp prior is just one specific Bayesian solution, which is *not* non-informative, because by definition you are selecting a prior which concentrates on one point, and that's certainly not non-informative. Sharp priors *are* informative, by definition. To consider whether or not you want to use a sharp prior or a "flattish" one, just think in one dimension. There is obviously a huge difference between having a kind of flattish prior over the real line and a prior concentrated near one point. What I am saying is that, if you are interested in testing whether or not $\theta = 3$, and you believe, you really believe (and you want to make this part of the analysis) that the true value of θ is either 3 or close to 3, then of course you should use a prior that reflects this. And then you could do all that I have done, without any change. Now, this is complicated in some cases. A simple approximation is to replace this kind of presumably continuous prior by a point mass at $\theta = 3$ and something else around. And that is an approximation. You always have to be careful, because it is known to be an approximation, and it is known that under some conditions this approximation breaks down. So you have to check that you are not in those conditions. You have to check both that you are interested in a prior distribution which is concentrated on something, and you really want answers that depend on that, plus that the approximation is going to be all right mathematically, which it usually will be. The ESP example is a real example, but it is special because you have a hundred million observations. In general, the differences are not that large. But I insist that if you want a hypothesis testing procedure that does not depend on a prior distribution concentrated on the null value, then you have objective Bayesian alternatives, by just using a continuous loss and a continuous prior.

Glen Cowan

It seemed that, in the ESP example, when you put a point mass at 50%, then in some sense you got the answer that you didn't want, and I guess what confused me is you said that saying you should have a point mass at 50% is in some sense an approximation, whereas it seems to me that, no, that's exactly what you mean by ESP not existing. But I guess what you are saying is that in a real experiment, that you have perhaps a lack of ESP, but in addition some small bias, and so the more exact prior therefore includes a little bit of variation around 0.5 in order to account for the inevitable bias.

José Bernardo

In the ESP example, if rather than putting a probability mass at 0.5 (50% or any other), you use a continuous prior distribution centered on 0.5, then, unless you specify a hugely (really hugely) concentrated prior, then you will get precisely the same answer I did, because this prior will be actually wiped out by the data anyway. It will give practically the same posterior, and therefore you will see that the data suggest that the true value is *not* exactly 0.5, but about 0.5002 or whatever. Now, whether this is because of ESP, the mind being able to move about 0.02% of the particles, or more likely to my taste, because there is a small bias in the machine, well, it's not statistics that is going to determine this.

Diego Casadei (New York University)

When we choose the Bayes factor to make a decision, we have seen from Jim Berger's talk this morning what happens, and you also made a caveat: there are conditions in which this may give surprises. To avoid those problems, what is the alternative? Shall we consider the ratio of posteriors between the two hypotheses?

José Bernardo

Bayes factors by themselves are not a choice. Their use is a consequence of the fact that if one uses a sharp type of prior, the posterior probability of the null depends on the data only through the Bayes factor. This is just a mathematical fact. But if you do not do this sharp prior approximation, this is no longer true. So it's not that Bayes factors are good or not, it's just that they do not appear as the relevant quantities to compute. What appears in place of the Bayes factor is the kind of expression that I have mentioned: you get a continuous posterior distribution. The simplification that the posterior probability of the null depends on the data only through the Bayes factor occurs only if you have a sharp prior. And that has the implications that if you don't believe this sharp prior to be sensible for one reason or another, then of course you should not be using the corresponding result. A lot of people tend to use only Bayes factors (rather than posterior probabilities), because of course then you don't have to argue with all the prior probabilities of the hypotheses. However, what Bayesian statistics tells you, is that you should really use posterior probabilities. Now, because of these mathematical equivalencies it is tempting to forget about it. This would be too long to discuss here, but if you insist on using only Bayes factors, as opposed to using posterior probabilities, you are going to get a number of problems, because you are easily getting outside foundations. For instance, the kind of approximate Bayes factors that you are forced to introduce may sometimes not be transitive, so you have that the Bayes factor of H_1 over H_2 times the Bayes factor of H_2 over H_3 is not the Bayes factor of H_1 over H_3, so these are not real Bayes factors. All these things may happen because all the time you are using approximations. There is nothing wrong with approximations, except that you have to realize that they are approximations, and therefore to be careful. That's all I am saying. And the advantage of using the sort of procedure that I have described is that you don't have to do that at all. I mean, you can always use the standard, continuous reference prior, precisely the same as in estimation, and I find that attractive.

Luc Demortier

We have about half an hour left before the coffee break, so maybe we should switch to a discussion of reference priors. Louis suggested that we ask professional statisticians first if they have any comments about reference priors. If that's the case, let's hear them first and then go to questions.

Jim Berger (Duke University)

I think I'd rather answer questions.

Luc Demortier

So let's open the floor to everybody then.

Harrison Prosper (Florida State University)

I hope this is not too technical, but in one of your most recent papers [1], with Jim Berger, you go through a formal definition of reference priors, and you arrive at the form that you illustrated on your slides. There is a sentence after that formula, to the effect that one no longer needs to use this compact set argument to make sure that, every step along the way, the reference prior is normalized. Did I read that correctly?

José Bernardo

If I remember correctly (I guess you refer to the recent Annals paper) what we meant is that, if you have only one parameter, it just happens that once you have created the structure, the particular choice of the compact sets doesn't matter, so the answer is unique. Unfortunately this is not true with more than one parameter. So even if you have just two parameters, we have examples where you can see that the answer that you get might depend on the particular sequence of compactification that you use, and following one of the questions that Luc has posed this morning, we don't have an answer as yet, as to what the standard procedure would be to have more or less automatic choice of the compacts. We have answers for Luc's other two questions, but not for the one about compacts. Except in one dimension. In one dimension you need the compacts to make the mathematics work, but then, the specific compact thing doesn't happen, so whatever compact set you choose, you are going to get the same answer.

Harrison Prosper

So presumably that means that we can use the numerical algorithm to calculate the reference prior, without having to worry about compact sets.

José Bernardo

In one dimension you don't have any problem. But if you do it in more than one dimension, then because it is a conditional argument, the problem will come up.

Harrison Prosper

Yes, the reason that this important is because of the sort of thing we are looking at right now, problems in which you have, say, one parameter of interest, and you may have many many nuisance parameters. And so the question is whether in that circumstance, supposing you wanted to calculate a conditional reference prior probability for θ given a whole bunch of nuisance parameters, whether in that case one has to use the compact set argument or one can just immediately calculate, for that particular parameter value of θ, using the algorithm you have in your paper.

José Bernardo

Well, this is an important research challenge, and I am not totally sure what may be done, but my attitude in that kind of problem would be to integrate out all the nuisance parameters using some approximate reference prior on a particular set of compacts, and then do a strict reference analysis with the parameter of interest with the resulting integrated model. It might be very complicated, in fact it *will* typically be very complicated, but because you have a numerical procedure you can obtain the reference posterior from this (approximate) integrated model.

Jim Berger

Harrison, tell me if I am wrong, but I think you are thinking of a situation where you have evidence-based priors for the nuisance parameters. So they actually have priors for all the nuisance parameters.

José Bernardo

Real priors?

Jim Berger

Real priors. And so his question is simply, would we trust the fact that the one-dimensional numerical algorithm works, so then we look at the problem of the one parameter of interest conditioned on the nuisance parameters, and would we need compact sets there? And I think not. I mean, because all of the examples where we needed compact sets were examples where we were doing more than one parameter at a time, and it was the interaction among them. So, barring finding an example tomorrow where it doesn't work, I'd be quite happy with evidence-based priors for all the nuisance parameters and the numerical algorithm for the parameter of interest.

Louis Lyons (Imperial College)

So, can I ask a question then? Say we are convinced that we want to use reference priors in our analyses. The only paper I knew of up till this present conference was by Luc and Harrison and Supriya. [2] If I want to do a reference prior analysis, is there somebody I can send an e-mail to, tell them what my problem is, and get a reference prior back? Or do I have to do it myself?

José Bernardo

I think the answer is that, if your problem belongs to a set of textbook examples, then yes I can give you the reference prior, or Jim, or other people, or even in the internet you can probably find a reference to a paper. If the model is not standard, then you have to do it.

Günter Zech (University of Siegen)

I have a non-expert question. If you have two experiments measuring the same parameter, and they have different acceptances, then they have different reference priors, right? When I measure the same parameter with different priors, how do you combine or compare these two results?

José Bernardo

That's a very interesting question, and one that has many sides to it. First of all I think you have to make sure that from a statistical point of view, the parameter is the same. From a Bayesian point of view, a parameter is the limit, a frequentist limit actually, of a function of the observations. So for a Bayesian, the parameter p in a binomial situation has a precise definition, namely it's the limit of the relative frequencies when n goes to infinity. Now, if you have two different models, you have to make sure that if you call θ, the same θ in two models, this θ is really the same thing. If you take one limit and you take the other limit, you get the same thing. And that's not trivial, you have actually to do it. If you have done that, then you have the same parameter.

Even then, you will get two different posteriors, even though it's the same parameter. Now these posteriors, these reference posteriors, are the answer to a "what if" question, namely what would I say about θ if my prior beliefs were as — let me get this sentence right — if my prior beliefs were those under which the data from this particular experiment would give me the highest possible information, that's the formal definition. Now two different experiments may give you the highest possible information for two different priors. And because they are conditioned answers, they are mathematically compatible, just different answers. Whether you would use one, or the other, or a mixture, you have these two models, you still can have, say a finite mixture of the two of them, either multiplicative or additive, and you have yet a third model and a slightly different answer. And all those answers are conditional answers to the particular model.

There is no such thing as *the* objective answer, it's a conditional answer, to a particular model, and you want to be precise about the assumption that your prior is such that the information provided by the data would be maximum. And somehow often there are underlying assumptions there, for instance the best known case is the binomial, where you have a fixed number of tosses or a fixed number of successes, binomial or inverted binomial. Now, in the standard binomial case, the reference prior, which is Jeffreys' prior, is Beta(1/2,1/2), and in the other one it's Beta(0,1/2), which is different. But the point is that in the inverted binomial situation you are assuming that the probability of success is strictly positive, because otherwise you would never get the required number of successes, and in the other one you are assuming that it might be zero. This slight difference is a radical assumption, which is reflected in the prior and therefore in the posterior.

Of course, if you have a lot of data the thing is not going to matter, but for very small datasets it *is* going to matter. If you do not know what the mechanism has been, binomial or inverted binomial, or "I am just tired" (you don't really know), neither of these answers is objective. I think English people have this saying that there no such thing as a free lunch, so there is no such thing as a free answer: you have to put in assumptions. Reference analysis is very specifically tailored to give an answer to a very precise, conditional question: what could I say about the parameter if my prior was that prior that maximizes the information from the data drawn from a specified model. And for this question we have an answer which I think is useful.

As for the original question, if you have two sets of data x and y from different experiments measuring the same parameter, it should be possible to specify a single experiment from which (x, y) may be assumed to have been drawn, and then derive the corresponding reference posterior for the common parameter.

Günter Zech

If you measure for instance the lifetime of the Lambda particle, and two experiments do it, and they use two different reference priors, it's a mess afterwards, to compare these data.

José Bernardo

Well, I see it as an example of sensitivity analysis. You have to do such a sensitivity analysis with respect to change in the model, with respect to change in the prior, with respect to everything, and in a sense, the reference posteriors that you would get for all kinds of different models that you can think of, would precisely result in a robust answer, in that my posterior must be in the convex set defined by all these posteriors. And if you are not able to select more precisely which of the experiments you prefer, well, you have a range of posteriors. I don't think there is anything wrong with that.

Luc Demortier

I think you just answered the first question that was on my list. I don't know if you want to say anything more about that. So the question was, what is the correct probabilistic interpretation of reference posteriors. A lot of my colleagues in high energy physics have this question, and they always come back to it. And related to that is, when you calculate a reference posterior, is it OK to just report that result or should you make it part of a sensitivity analysis, and if so, how do you choose other priors? Because reference priors have a special status, so is it necessary to go look for other priors, and if so, what other priors?

José Bernardo

OK first things first. The idea that you need to make reference analysis part of a global analysis, I don't think you have to, but obviously if you have the time and the resources to do it, it's always better. With respect to interpretation, I think that that is more or less clear. The interpretation is that a reference posterior is a posterior, an expression of beliefs in the parameter of interest, given the data. But as a posterior, it depends on the prior. Because of the way the prior has been constructed, it is the solution to what should I believe if I did not have any relevant information, in the precise sense defined by that prior that would maximize the information from the data. And there is of course a second answer, which is essentially asymptotic, namely that it is also true that if you give a credible interval with posterior probability 0.95, you will cover the true value, under most circumstances, with probability close to 0.95, and in some examples with exact 0.95 probability. So the two interpretations are out there. Indeed, in most cases credible intervals are approximate confidence intervals, but in some problems there are not, and for a good reason. If you insist in getting a frequentist interval for the ratio of means problem, from a frequentist point of view you get a disaster, you get the whole real line with probability 0.95. Of course you do not want to reproduce that, and indeed the Bayesian reference posterior does not reproduce that. [3] By and large, it is always true that you should be able to bet on credible intervals, if your own real priors are kind of non-informative with respect to the parameter of interest. Besides, they mostly have a rough sort of interpretation in frequentist terms; but this is not a general result, it does require conditions.

Glen Cowan

I have a comment for Günter Zech, and then a naïve question, and that is that if you had two experiments with independent data, so they are each characterized by a certain likelihood function, but they are measuring the same underlying parameter, then is it not true that there would be a reference prior for experiment A, a reference prior for experiment B, but in addition if you were simply to consider both experiments together, they are characterized by a single likelihood function which is the product of the two, so there is a unique reference prior that characterizes in some sense the combined experiment. Would that not solve your problem, Günter?

Günter Zech

You would have to publish the acceptances of the two experiments because the reference prior is not simply a function of the likelihoods.

Glen Cowan

Ah, how you publish is another problem! I think the thing is, if you really want to get together and compare results and combine results, that cannot be done if each guy shows up with his own posterior distribution, you have to take a step back. I guess we knew that already.

David Cox (Nuffield College, Oxford)

I'd like to make a couple of comments and just a historical point. I find the ideas of reference priors highly attractive, but... And the but is largely an issue of interpretation. It's the posterior I would get if this was my prior, but I don't want it, it's not my prior normally. And in particular if it's improper... That's the issue of interpretation.

The more technical question concerns the issue that Harrison raised, I think really, that if the number of nuisance parameters is large, what's going on? The asymptotics can be taken to be that as the number of observations tends to infinity, so does the number of nuisance parameters. Now I know you have considered one of Charles Stein's examples, where you get a nice answer from the reference priors, a correct answer about non-central chisquared and so forth. But I am unclear whether your results apply more generally than that. If we have a problem in which the number of nuisance parameters is quite large in some sense, do things tend to go wrong?

Another issue where I do very strongly disagree with the answers you get, which you mentioned about a minute or so ago, is about the ratio of two normal means, where I think the traditional answer is the right one, and the answer you give is not, from a purely common sense point of view.

The historical point is, just by chance I mentioned Renyi this morning. Renyi did give an axiomatization of probability in which improper priors where allowed, in the spirit of Kolmogorov, and Kolmogorov did approve this in some sense. It might possibly be interesting to go back to that. That was, I shudder to think, sixty years ago that he gave this.

Jim Linnemann (Michigan State University)

I'd like to provoke a little more conversation amongst the statisticians, because for me coming here is certainly about learning about their views on these issues. What we physicists have always wanted from statisticians, is to give us a unique answer. We have gotten used to the fact that we are not going to get a unique answer out of discussions of what Bayesian techniques can do, and what frequentist techniques can do. But maybe, at least the (objective) Bayesians could give us a unique answer. There are two kinds of uniqueness. One is mathematical uniqueness, for a sufficiently well-posed problem. The other is practical uniqueness: something sufficiently convincing that it sweeps the field. So my question is: there are many attractive features to this procedure, but why isn't everyone using it? And if all statisticians aren't convinced, why should we use it?

Louis Lyons

So, any Bayesian answers?

Jim Berger

Another sort of question, or answer of uniqueness, is in the cases we have been talking about where the reference prior is not unique, how different are the answers? And it will be negligibly different. I mean in the cases that we talked about, where there are two different experiments, if both people use one reference prior or both use another, the answers are not going to be distinguishable, in a confidence limit sense. So the answer will be, not exactly unique, but more or less the scientific conclusion will be unique. So if the choice of the reference prior doesn't matter that much, why do we do them? Well, it's because there are other priors, like vague, constant priors and what not, that we know are dangerous, that they may work well, but we know that there are scenarios where they just collapse. We don't know scenarios, well we know of only one, where a reference prior doesn't work, in the sense of giving a sensible answer, that's usually very close to a frequentist answer, except in the problems where it's impossible to do both, like the mu over sigma problem. And I would argue with David that it's not necessarily possible to say that the unknown Poisson mean is between 0 and infinity with probability 0.95. I mean the frequentist answer in that case, to get a genuine frequentist answer, sometimes there are theorems that say you have to say the whole real set with probability one minus alpha. Oh, you have other frequentist answers? We'll talk about that offline. But there are problems where the frequentist and the Bayesian simply can't agree, because of conditioning issues, and in those problems there is, you know, a serious debate as to what you should do. But in the ones where the frequentist and Bayesian can reach agreement, it seems like reference priors get to that agreement better than anything.

David Cox

I would say that the answer to the crucial question about uniqueness is, how deep a formulation are you willing to give of a problem? I mean, problems come in many different kinds, and some have only weak specification. The ideal situation is that you have your priors based on evidence, you have your model, you have your likelihood, and then the Bayesian solution, clearly, is the right one. And nobody, as far as I know, has ever argued with that. The issue is, how far along that route of specification are you willing to go? And the reference priors are a very beautiful attempt, and successful, perhaps to a large extent successful, of evading part of that question by saying, well no we can't look down a very specific, probabilistic statement of the nature of the evidence external to the data. So we do something else instead.

Frequentist statisticians of course don't disregard external information, they simply say, we can't formulate it probabilistically, so we have to incorporate it qualitatively, with a confidence interval or significance test or whatever. And the weakest form of such a procedure is the simple significance test, which I talked mostly about this morning, where the formulation is extremely weak, it's just a null hypothesis and an indication of which direction you are looking in, nothing else is formulated probabilistically. And if that's all as far as you can go quantitatively, that's as far as you can go. And it's far from an ideal situation. Nobody, I think, really likes it too much. And then, you may have an idea of an alternative, you may have a detailed model for the alternative, confidence intervals, or posterior intervals, you may have your prior distribution and so forth, there is this hierarchy.

R.A. Fisher, who really invented most of the more traditional, the non-Bayesian side of the subject, repeatedly emphasized, especially in his last book, that there are hierarchies of formulation, including — he very explicitly, and he used occasionally — Bayesian formulations, when he thought they were appropriate in genetical problems. And he emphasized, there really might well be forms that are not already discovered. And speaking purely personally, I first learned about statistics from Jeffreys, and Jeffreys' work is very much extensive about that. I find that appealing, but I have my reservations, as I tried to indicate.

Nikolai Krasnikov (Institute for Nuclear Research, Moscow)

For some priors Bayesian results coincide with frequentist results, for instance upper limits for Poisson distributions. So maybe they use these priors as the best priors, which coincide with frequentist results.

David Cox

I have said more than I intended to already, but on this issue of the ratio of the means, the point is that, supposing we have an answer 0.5 with a standard error of 5 in the numerator, and 0.5 with a standard error of 5 in the denominator, the frequentist answer is, any value of the ratio is consistent with the data. Now Neyman, who strongly emphasized taking the confidence interval interpretation very very literally, never resolved the issue. And if you start to talk about 95% confidence intervals, then you are in the trouble that Jim emphasizes. But I don't regard that as the frequentist answer. The frequentist answer is that, at any level of significance, up to some maximum, or down to some minimum, any value might be consistent with the data. And in another context, it may be that any value outside certain intervals is consistent with the data. And the data tell you which of three possible answers you get: a confidence interval for a ratio, the exterior of an interval for a ratio, or that the whole real line is consistent with the data, up to whatever level of probability seems appropriate. The data tell you that. If you force yourself to make an interval statement, when interval statements are inappropriate, then you may get quite the wrong answer.

Jim Berger

To come back to what a discussion like this ends up saying, one has to start being precise about what frequentism means. I have gotten the feeling that when this body is talking about confidence sets in a frequentist way, it interprets it in a literal Neyman way, where, if I have a 95% confidence set, by gosh that thing has to cover the true parameter 95% of the time no matter what the true parameter is. And so the discussion about this ratio of means problem that we are having is that there is this theorem that says that the only way a frequentist can do that is by saying some of the time for some data, the whole real line. David points out that perhaps this is perfectly reasonable to say, but my only complaint about it is saying the parameter is in the whole real line, and I have 95% confidence in that statement. Because obviously I am sure that the parameter is in the whole real line. So it's the attachment of the 95% confidence statement, which is demanded by strict frequentist, Neyman-type thinking, that is the issue in my mind.

Luc Demortier

Well, if there are no more questions, there was a third question on my list this morning, it's a quick one, whether you think it is possible to combine ABC (Approximate Bayesian Computation) methods with the reference prior algorithm.

José Bernardo

I do not have a real answer because I have never tried to. My hunch is that you can always do something, either exactly or approximately, but I don't know enough about the subject to make a serious comment. But Jim?

Jim Berger

I don't think so, because the ABC methods are when you don't have an explicit form for the likelihood, and given that the computation of reference priors is a delicate asymptotic computation, that depends very much on the form of the likelihood, I don't think that any kind of ABC approximation could be combined with that. So I think the answer is no.

Kyle Cranmer

I am just trying to think of something a little bit more physics motivated but still about reference priors. If we go to something like searches for the Higgs, and we wanted to, say, set limits on the Higgs cross section, or the Higgs mass or something like that, there we may or may not have relationships between various parameters. For instance the Higgs mass is the only free parameter in the standard model, and we could try to come up with a reference prior for it. But we don't have to work hard to go to a different class of theoretical physics models where the relationship between how the Higgs might decay in various different ways gets broken, and it's no longer this one-dimensional model but it's now two or three or four of five. A reasonable question for us is to imagine that someone handed us all of the machinery for using reference analysis, and we are doing something like the Higgs, we need for instance the normal standard model Higgs and then maybe a Higgs where the cross section is independent of the mass, and maybe one where we have multiple Higgs bosons, each with different masses and cross sections, and different production modes and different branching ratios, and so there we could quickly get like 16, 20 parameters of interest in that model. We will somehow need to figure out how to navigate through that. So that's really a question for the physicists but still along the lines that if we had this kind of machinery available, maybe that prompts some thought for discussion.

Harrison Prosper

This is a very interesting, important problem, actually, how to deal with these multi-parameter models in this framework. In fact, just to give you a concrete example that my colleague and I are working on. We use the standard Poisson model, but this time there are two parameters of interest, which enter linearly in the mean. So you have some constant times θ_1, plus another constant times θ_2, plus a nuisance parameter, the background. And so one of the questions I have is that right now we were talking about this one-dimensional parameter problem, but if you have two parameters that seem to enter the problem in a symmetric way, can we treat those two parameters simultaneously, that is, to apply the reference algorithm on two parameters at the same time, or does one still have to do the sequential algorithm?

José Bernardo

I really think you should do it sequentially, not because of the technicalities, but because of the results to be obtained. I'll give you an example: under regularity conditions, you can easily extend the argument by which the reference prior becomes a Jeffreys prior, to the argument that in a multivariate situation you get a multivariate Jeffreys prior; but we know that this doesn't work, in fact it works very badly. The reason is that you are maximizing the amount of information simultaneously on all the parameters. It's just a fact that if you want to be able to get good properties of the marginal reference posterior for the parameter of interest, you just cannot do it globally, you have to do it sequentially, and in order. If you reverse the order, as some people have tried, you don't get the right answer either. The Stein's paradox, the problem that David Cox mentioned before, is a very dramatic example of that. When you have many parameters, if you try to do anything globally, with all of them, you are going to get very bad results. The answer I think is, it would work technically, but you would not get the right answer. So I believe you have to do it sequentially.

Jim Berger

This is a sort of comment on this question of reference priors when you have many parameters. Just when I sat down with Kyle, he reminded me that with his Asimov datasets [4] he can compute Fisher information matrices. And the numerical reference prior algorithms are very, very general, but everything simplifies a great deal if one is in a regular situation, where the Fisher information exists and makes sense. Then, in one of our earlier papers [5], José and I have a much simpler algorithm, based only on the Fisher information matrix, for computing the sequential, iterative reference prior, and I think it's worth taking a look at that algorithm, combined with Kyle and collaborators' numerical computation of the Fisher information matrix, that might end up being a much easier way to implement the numerical algorithm.

References

[1] J. O. Berger, J. M. Bernardo, and D. Sun, "The formal definition of reference priors," Ann. Statist. **37**, 905 (2009); `http://www.uv.es/~bernardo/2009Annals.pdf`.

[2] L. Demortier, H. B. Prosper, and S. Jain, "Reference priors for high energy physics," Phys. Rev. D **82**, 034002 (2010); arXiv:1002.1111v2 [stat.AP] (2010).

[3] See section 5.2 in J. M. Bernardo, "Reference posterior distributions for Bayesian inference," J. R. Statist. Soc. B **41**, 113 (1979); `http://www.uv.es/~bernardo/1979JRSSB.pdf`.

[4] G. Cowan *et al.*, "Asymptotic formulae for likelihood-based tests of new physics," arXiv:1007.1727v2 [physics.data-an] (2010).

[5] J. O. Berger and J. M. Bernardo, "Ordered group reference priors with application to the multinomial problem," Biometrika **79**, 25 (1992); `http://www.uv.es/~bernardo/1992Biometrika.pdf`.