Departament d'Estadística i I.O., Universitat de València.
Facultat de Matemàtiques, 46100–Burjassot, València, Spain.
Tel. 34.6.363.6048, Fax 34.6.363.6048 (direct), 34.6.386.4735 (office)
Internet: bernardo@uv.es, Web: http://www.uv.es/˜bernardo/

Printed on December 5, 1996
To appear in *J. Statist. Planning and Inference*

Noninformative Priors Do Not Exist: A Discussion with José M. Bernardo

PREAMBLE

José Bernardo, who is the originator and, with James Berger, a driving force behind the use of what are known as "reference priors", shared a ride with Telba Irony and Nozer Singpurwalla from Pittsburgh, Pa. to Washington, D.C., sometime during the latter part of 1995. After the usual gossip, which statisticians of all persuasions tend to relish, the discussion turned to the topic of "dishonest priors", and their increasing encroachment on the Empire of Chance. The discussion evolved in a Socratic tradition, with José playing the role of Socrates (albeit answering the questions in this particular dialog), and his driving partners, the pupils; fortunately José was not driving. This question and answer format of discussion turned out to be very fruitful, and even more enjoyable than gossip, because José's knowledge of the topic and its historic evolution, not to mention his passion for statistics (among other things), provided a useful perspective on a topic of much controversy. José was then asked to write up this perspective as a contribution for the Discussion Forum of JSPI, to which his suggestion was that we reproduce the dialogue and write it up as such. We thought that this suggestion was a great idea and so, when José visited Washington D.C. in May 1996, the dialogue was reconstructed. Given below is an account of this reconstruction; we feel that those of us who are not cognoscenti about dishonest priors may benefit from this conversational overview.

Telba Z. Irony
Nozer D. Singpurwalla

Question 1: It is often said that *noninformative* priors do not exist and yet under this label, plus many others, such as *conventional*, *default*, *flat*, *formal*, *neutral*, *non-subjective*, *objective* and *reference* priors, they do get used. Does this imply that the users of noninformative priors are really not honest Bayesians?

Answer: I would not say they that are dishonest but, too often, they are not precisely aware of the implications of its use. By definition, "non-subjective prior distributions" are *not* intended to describe personal beliefs, and in most cases, they are *not even proper* probability distributions in that they often do not integrate one. Technically, they are *only* positive functions to be formally used in Bayes theorem to obtain "non-subjective posteriors" which, for a *given model*, are supposed to describe whatever the data "have to say" about some *particular quantity*

José M. Bernardo is Professor of Statistics at the University of Valencia. Research partially funded with grant PB93-1204 of the DGICYT, Madrid, Spain. The author gratefully acknowledges stimulating comments by Jim Berger, Gauri Datta, Phil Dawid and Dennis Lindley to an earlier version of this manuscript.

of interest, either a function of the parameters in the model, or a function of future observations. Whether or not they achieve this goal or, indeed, whether or not this goal is at all achievable, is often a matter of debate.

Question 2: Are non-subjective posteriors always proper?

Answer: They *should* be! Indeed, a proper posterior for any minimum-size sample should be the first property to be required from any method of deriving non-subjective priors. But you should realise that a naïve use of "standard" non-subjective priors *may* lead to improper posteriors. Casella (1996) describes how, in a components of variance problem, this has lead to unaware users of Gibbs sampling to obtaining nice pictures of posterior distributions that do not exist! Actually, Dennis Lindley told me that he knew of this components of variance example since the 50's, and it is mentioned in Berger (1985, p. 187), but people seem to take for granted a crucial property, —the propriety of the posterior— which *must* however be carefully checked whenever an improper prior is used.

Question 3: Are there any non-subjective *priors* that are proper?

Answer: There are indeed. The simplest example is the Beta distribution $\mathrm{Be}(\theta|1/2, 1/2)$, widely regarded as the appropriate non-subjective prior to make inferences about the parameter θ of a binomial model. In fact, proper non-subjective priors are usually found whenever the parameter space is bounded, although this is not a general rule: the non-subjective prior typically recommended for the parameter θ of a negative binomial model is $\pi(\theta) = \theta^{-1}(1-\theta)^{-1/2}$, which is improper although the parameter space is bounded, while a sensible non-subjective prior to make inferences about the ratio of two multinomial parameters turns out to be proper even though the parameter space $]0, \infty[$ is not bounded. Actually, one could always work with *proper* non-subjective priors if the parameter spaces were taken to be appropriately chosen bounded sets. For example, the standard non-subjective prior for a real location parameter is uniform on $\Re$, which is improper; however, if given some experimental measures, inferences are made on the true value of some physical quantity a more realistic parameter space would be $[0, c]$, for some context-dependent constant c, and the presumably appropriate non-subjective prior, a uniform on $[0, c]$, would indeed be proper. Similarly, in the negative binomial setting where the probability of success must be strictly positive, a parameter space of the form $[\epsilon, 1]$, for some $\epsilon > 0$ would lead to a proper non-subjective prior.

Question 4: How do you interpret probability?

Answer: Naturally, as a measure of personal belief. Of course, this does not mean that I would systematically be prepared to bet in terms of non-subjective posterior distributions, because my *personal* beliefs may well be not closely approximated by any particular non-subjective prior.

Question 5: I agree that betting quotients may not reflect true belief, but do you subscribe to the axioms of probability?

Answer: Yes, I do. I am a strong believer in foundational arguments: the intellectual strength of the Bayesian argument comes directly from the fact that mathematical logic requires one to express *all* uncertainties by means of *probability* measures. In fact there are two independent arguments to support this claim:

(i) *Coherent decision theory*: If you try to guarantee that your decision making criteria are sensible —in that they meet some intuitively appealing axioms— or if you want to avoid "inadmissible" decisions, then you *must* express all your uncertainties by probabilities. For specific details, you can look at Savage (1954), Fishburn (1981, 1986), or Bernardo and Smith (1994, Ch. 2), and references therein.

(ii) *Representation theorems*: If you accept a probabilistic description of the behaviour of observables —as all statisticians presumably do—, and you want to describe mathematically the idea that some observations are "similar" in some sense —and hence some type of prediction is possible—, then the general representation theorem tells you, that these "exchangeable" observations are a random sample from some underlying model, indexed by some parameter defined as the limit of some function of the observations, and *there exists a prior* probability distribution over such parameter. Key references are Finetti (1937), Hewwit and Savage (1955), Smith (1981) and Diaconis (1988); you can get an overview from Bernardo and Smith (1994, Ch. 4), and references therein;

These are proven *existence results*; they imply that the common sentence "a prior does not exist" is a mathematical *fallacy*: for mathematical consistency one *must* be a Bayesian. However, these are *only* existence results; they leave open the question of *specifying* a particular prior in each problem.

Question 6: Don't the axioms imply that the probability of a tautology should be 1 and, thus, that priors have to be proper?

Answer: The natural axioms do not: they only lead to finite additivity, which is compatible with improper measures; however, the further natural assumption of *conglomerability* leads to σ-additivity and, hence, to proper measures; some signposts to this debate are Renyi (1962), Heath and Sudderth (1978, 1989), Hartigan (1983), Cifarelli and Regazzini (1987), Consonni and Veronese (1989) and Lindley (1996). Nevertheless, it must be stressed that what really matters is the posterior of the quantity of interest —which under σ-additivity must certainly be proper—, because this is what you have to use either in inference or in decision making.

Question 7: But those proper non-subjective posteriors are often derived from improper priors; how should one interpret the improper priors?

Answer: One should *not* interpret *any* non-subjective prior as a probability distribution. Non-subjective priors are merely technical devices to produce non-subjective posteriors by formal use of Bayes theorem, and "sensible" non-subjective posteriors are *always* proper.

Question 8: What do you mean by a "sensible" non-subjective posterior?

Answer: One that, —after careful scrutiny of its properties—, you would be prepared to use for scientific communication. Of course, there is no way to give a formal definition for this; indeed, an important part of the discussion on methods for deriving non-subjective priors is based on the analysis of the statistical properties of the posteriors they produce in specific, "test-case" examples.

Question 9: If the prior to posterior conversion describes the change in one's betting behaviour (and this is a debatable issue), should that prior also not be proper?

Answer: Non-subjective priors are *limits*. Any sensible non-subjective prior may be seen as some appropriately defined limit of a sequence of proper priors. In fact, as I have mentioned before, they are only improper because of the use of convenient, typically unbounded parameter spaces. If we tried to be more realistic, and worked with appropriately chosen bounded parameter spaces, non-subjective priors would always be proper; if one prefers to work with conventional parameter spaces, increasing sequences of bounded approximations to those parameter spaces may be used to provide sequences of *proper* priors which converge to the corresponding improper non-subjective priors in the precise sense that, for any data set, the sequence of posteriors thus produced converges to the corresponding non-subjective posterior.

Question 10: Why then have improper priors entered our lives? Can we not do away with them by concentrating only on compact spaces?

Answer: We could indeed, but this would probably be mathematically inconvenient. But, again, the main question is *not* whether non-subjective priors are proper or improper, but whether or not they lead to sensible non-subjective posteriors. Actually, if an improper prior leads to a posterior with undesirable properties, the posterior which would result from a proper approximation to that prior (like that obtained by truncation), will typically have the same undesirable properties: for instance the posterior of the sum of the squares of normal means $\phi = \sum \mu_i^2$ based on a joint uniform prior on the means $\pi(\mu_1, \ldots, \mu_k) \propto 1$ is extremely unsatisfactory as a non-subjective posterior for ϕ (Stein, 1959), but so it is that based on the *proper* multinormal prior $\pi(\mu_1, \ldots, \mu_k) \propto \prod_i \mathrm{N}(\mu_i|0, \sigma)$, for large σ. Proper or improper, we need non-subjective priors which appropriately represent a lack of relevant prior knowledge about the quantity of interest *relative* to that provided by the data.

Question 11: What do you mean by a prior describing a lack of knowledge? Given my utility function, the amount I am willing to bet signals the extent of my knowledge.

Answer: The contribution of the data in constructing the posterior of interest should be “dominant”. Note that this does *not* mean that a non-subjective prior is a mathematical description of “ignorance”. *Any* prior reflects some form of knowledge. Non-subjective priors try to make precise the type of prior knowledge which, for a *given model* and for a *particular inference problem* within this model, would make data dominant.

Question 12: We now see your point: namely, that non-subjective priors are those for which the contribution of the data are posterior dominant for the quantity of interest. This is sensible, but to construct non-subjective priors does one need to consider data?

Answer: Non-subjective priors do not typically depend on the data, but on the probabilistic model which is assumed to have produced them: as one would expect, the prior which makes the data posterior dominant for the quantity of interest usually depends both on the model assumed for the data and on the quantity of interest.

Question 13: It appears that, in using the prior as a mere technical device in a formal use of Bayes’ law, and with the aim of making the posterior data dominant, one is proposing a paradigm for inference which is not in the spirit of Bayes (and Laplace) nor is it in the spirit of Fisher, or Neyman-Pearson-Wald. If so, may one conclude that a non-subjective Bayesian is a sensible *pseudo-Bayesian*?

Answer: No, I think he is totally Bayesian. Non-subjective Bayesian analysis is just a part, —an important part, I believe—, of a healthy *sensitivity analysis* to the prior choice: it provides an answer a very important question in scientific communication, namely, what could one conclude from the data *if* prior beliefs were such that the posterior distribution of the quantity of interest were dominated by the data.

Question 14: What are the historical precedents of non-subjective priors?

Answer: The pioneers were Bayes (1763), who considered a uniform prior in a binomial setting, and Laplace (1812), who used an improper uniform prior in the normal case. Nobody considered using priors that were different from the uniform in those days when, by the way, a large part of statistical inference was based on inverse probability calculations.

Question 15: So Laplace did use uniform priors that were improper?

Answer: Yes, although he seemed to be aware that this was only a sensible approximation to using a proper uniform prior in the *bounded* parameter space which would more closely reflect the underlying physical problem.

Question 16: How does this connect with Laplace's "principle of insufficient reason"?

Answer: The "rationale" for using a uniform prior was that any other prior would reflect specific knowledge. Of course, as I stated before, any prior reflects *some* knowledge; what happens here is that in any *location* problem, the uniform prior is precisely that which makes the data posterior dominant.

Question 17: When did problems with uniform priors surface?

Answer: By the early 20's it was widely realised that the universal use of a uniform prior did not make sense. Since most statisticians were not prepared to use personal priors in scientific work, alternative "objective" methods of statistical inference were produced, which only depended on the assumed probability model; this gave rise to both fiducial and frequentist inference. It was not until the 40's that Jeffreys (1946) produced an alternative to using the uniform as a non-subjective prior; however, Jeffreys was a physicist, barely integrated in the world of academic statistics, and his work did not achieve the impact it deserved.

Question 18: What did Jeffreys do?

Answer: He was motivated by invariance requirements and suggested, using differential geometry arguments, a solution which provides a non-subjective prior known as Jeffreys' prior. He then proceeded to a detailed investigation of the consequences of such an approach (Jeffreys, 1961).

Question 19: What are these invariance requirements? Are they really crucial?

Answer: They are invariance under one-to-one transformations, and invariance under sufficient statistics. I certainly believe that those properties are crucial for "sensible" non-subjective distributions, to the point that one should not seriously consider a proposal for non-subjective Bayesian inference which does not satisfy them:

(i) *Invariance under one-to-one transformations*. If $\theta = \theta(\phi)$ is a one-to-one function of ϕ, then $\pi(\phi|\boldsymbol{x})$is *logically equivalent* to $\pi(\theta|\boldsymbol{x}) = \pi(\phi|\boldsymbol{x})|d\phi/d\theta|$; hence, the non-subjective posterior for ϕ directly obtained from $p(\boldsymbol{x}|\phi)$ and the non-subjective posterior for θ obtained from the same model reparametrized in terms of θ *must* be related by that equation.

(ii) *Invariance under sufficient statistics*. If t is a *sufficient* statistic for the model $\pi(\phi|\boldsymbol{x})$, then the non-subjective posterior $\pi(\phi|\boldsymbol{x})$ obtained from model $p(\boldsymbol{x}|\phi)$ *must* be the same as the non-subjective posterior $\pi(\phi|t)$ obtained from $p(t|\phi)$.

Some pointers to the literature on the role of invariance in the selection of non-subjective priors are Hartigan (1964), Jaynes (1968) and Dawid (1983) and Yang (1995).

Question 20: Can you give an intuitive explanation of Jeffreys' prior?

Answer: His invariance requirements are easy to understand, but he did not offer an intuitively convincing explanation for his particular choice; however, a modern description by Kass (1989) offers a heuristic explanation based on the idea that "natural" volume elements, defined in terms of Fisher's matrix, should have equal prior probability. Moreover, when applicable (continuity and appropriate regularity conditions) the one-dimensional version of Jeffreys' prior has been justified from many different viewpoints: these include Perks (1947), Lindley (1961), Welch and Peers (1963), Hartigan (1965), Good (1969), Kashyap (1971), Box and Tiao (1973, Sec. 1.3), Bernardo (1979), Kass (1990), Wasserman (1991) and Clarke and Barron (1994). The derivation of the one-dimensional Jeffreys' prior by Welch and Peers (1963), as that prior for which the coverage probabilities of one-sided posterior credible intervals are asymptotically as close as possible to their posterior probabilities, may be specially appealing to frequentist trained statisticians: this means that under Jeffreys' one-parameter prior and for large sample sizes, an interval with posterior probability $1 - \alpha$ may be *approximately* be interpreted as a confidence interval, in the Neyman-Pearson sense, with significance level α.

Question 21: You have just mentioned that, from a technical point of view, Jeffreys prior is related to Fisher's information matrix. Is Fisher's information matrix related to the notion of "lack of information"?

Answer: Not directly. The connection comes from the role of Fisher's matrix in asymptotics. By the way, this should be simply called "Fisher's matrix", not "Fisher's information matrix": it is only directly related to general information measures under normal assumptions or in regular asymptotic conditions.

Question 22: Is Jeffreys' prior a flat prior?

Answer: I guess you mean "nearly uniform". Then no, generally it is not, unless the parameter is a location parameter. I will use your question to stress that "flat" priors, —typically uniform or log-uniform—, which are too often used as being synonymous with "noninformative" priors, may be *very* informative on non-location parameters. For instance, as I have mentioned before, a "flat" prior on the means of a multivariate normal implies strong knowledge about their sum of squares, thus producing Stein's (1959) paradox.

Question 23: Then in what sense is Jeffreys prior neutral; *i.e.*, it reflects a lack of knowledge?

Answer: It does *not* reflect "lack of knowledge", but it may be argued that —with one parameter and under regularity conditions— Jeffreys' prior describes the type of prior knowledge which would make the data as posterior dominant as possible. Thus, the corresponding posterior distribution may be argued to provide a benchmark, a "reference", for the class of the posterior distributions which may be obtained from other, possibly subjective, priors.

Question 24: What is the problem with Jeffreys' prior and in which cases it is appropriate?

Answer: Jeffreys himself realized that his proposal only works, and then under regularity conditions, in one-parameter continuous problems. He suggested a collection of *ad hoc* rules to deal with multiparameter problems, with mixed results. Moreover, he seemed to be convinced that a *unique* appropriate non-subjective prior could be defined for any given model, whatever the quantity of interest. This was later seen to be not true.

Question 25: What were the developments after Jeffreys'?

Answer: For a while, it was thought that clever analysis of multiparameter problems using some combination of Jeffreys' proposals would produce appropriate non-subjective priors for any problem. Lindley's (1965) book was an explicit attempt is this sense, and it *does* prove that most standard "textbook" inference problems have a non-subjective Bayesian solution within this framework, and one which produces credible intervals which are often *numerically* either identical or very close to their frequentist counterparts. But then, in the early 70's, the marginalization paradoxes emerged.

Question 26: What are the marginalization paradoxes?

Answer: They may collectively be seen as a proof that the original idea of a *unique* non-subjective prior for each model is untenable: we may only agree on a unique non-subjective prior for each quantity of interest within a model. A simple example of marginalization paradox is provided by the standarized mean $\phi = \mu/\sigma$ of a normal distribution: Stone and Dawid (1972) showed that the posterior distribution of ϕ only depends on the data through some statistics t, whose sampling distribution only depends on ϕ. Hence, one would expect that inferences derived from the model $p(t|\phi)$ would match those derived from the full model $N(x|\mu, \sigma)$, but they proved that this is not possible if one uses the "standard" non-subjective prior $\pi(\mu, \sigma) = 1/\sigma$, which everybody agrees it is the appropriate non-subjective prior to make inferences about either μ or σ. It was immediately seen (Dawid, Stone and Zidek, 1973) that marginalization paradoxes are ubiquitous in multiparameter problems: any future development of non-subjective Bayesian analysis would have to come to terms with them.

Question 27: Was your own work on *reference* distributions a reaction to this?

Answer: It was not a direct reaction, but I was certainly influenced by these results. In the mid 70's, as a part of my Ph.D. work on Bayesian design of experiments, I become interested in non-subjective, "noninformative" priors. The marginalization paradoxes made obvious to me that some new work in that area was necessary: reference analysis was the result.

Question 28: Can you explain what do you mean by reference analysis?

Answer: Reference analysis may be described as a method to derive model-based, non-subjective *posteriors*, based on information-theoretical ideas, and intended to describe the inferential content of the data for scientific communication. It is, to the best of my knowledge, the only general method available which has the required invariance properties and successfully deals with the marginalization paradoxes.

Question 29: This is a very convincing statement in favour of your paradigm, but what do you mean by the inferential content of the data? How may this be quantified?

Answer: In a sense, each possible answer to that pair of related questions may be the basis for a method to derive non-subjective posteriors. I personally believe that the inferential content of the data is appropriately measured by the *amount of information* they provide on the quantity of interest, where the word "information" is used in the technical sense of Shannon (1948) and Lindley (1956). This was the starting point for the definition of a *reference* posterior (Bernardo, 1979).

Question 30: Can you be more explicit on the relationship between reference distributions and information theoretic ideas?

Answer: Sure. The amount of information to be *expected* from the data is naturally a function of the prior knowledge, as described by the prior distribution: the more prior information available, the less information may one expect from the data. With only one real-valued parameter, one may unambiguously define a limit functional which measures, in terms of the prior distribution, the amount of *missing information* about the parameter which data from a given model could possibly be expected to provide; the reference prior is that which maximizes the missing information. The multiparameter case is handled by using recursively the one-parameter solution.

Question 31: How reference priors differ from other proposals? In particular, how do they differ from Jeffreys' priors?

Answer: The reference prior approach is totally general and, as far as I am aware, it includes within a single framework all generally accepted non-subjective solutions to specific cases. In one-parameter problems, the reference prior reduces to Jaynes (1968) *maximum entropy* prior if the parameter space has a finite number of points, and it reduces to Jeffreys' prior in the continuous regular case. In regular continuous multiparameter problems, one often obtains the solutions which Jeffreys suggested using *ad hoc* arguments rather than his general multivariate rule. Moreover, reference analysis can deal with non-regular cases which cause problems for other methods.

Question 32: Can you give some examples of reference priors in the one-parameter continuous case?

Answer: As I have just mentioned, under regularity conditions to guarantee asymptotic normality, the reference prior is simply Jeffreys' prior, namely

$$\pi(\theta) \propto \left(\boldsymbol{E}_{x|\theta} \left[-\frac{d^2}{d\theta} \log p(x|\theta) \right] \right)^{1/2},$$

but I will give you a couple of non-regular examples:

(i) *Uniform distribution on* $[\theta - a, \theta + a]$. The reference prior is then uniform on $\Re$, and the reference posterior is uniform over the set of θ values which remain feasible after the data have been observed; (Bernardo and Smith, 1994, p. 311)

(ii) *Uniform distribution on* $[0, \theta]$. The reference prior is then $\pi(\theta) \propto \theta^{-1}$, and the reference posterior is a Pareto distribution; (Bernardo and Smith, 1994, p. 438).

Question 33: You have sketched the derivation of reference posteriors associated to models with only one real-valued parameter, and stated that those are invariant under reparametrization, but how do you deal with nuisance parameters?

Answer: Recursively: the idea is very simple, although there are delicate technical issues involved. Consider the simplest case; suppose that you are interested in the reference posterior distribution $\pi(\phi|x_1, \ldots, x_n)$ of some quantity ϕ given a random sample from a model $p(x|\phi, \lambda)$, which contains one real-valued nuisance parameter $\lambda \in \Lambda \subset \Re$. Working conditionally on ϕ, this is a one-parameter problem, and hence the one-parameter solution may be used to provide a *conditional* reference prior $\pi(\lambda|\phi)$. If this is proper, then it may be used to integrate out the nuisance parameter λ and obtain a model with one real-valued parameter $p(x|\phi)$ to which the one-parameter solution is applied again to derive the *marginal* reference prior $\pi(\phi)$; the desired reference posterior is then simply

$$\pi(\phi|x_1, \ldots, x_n) \propto \pi(\phi) \int_\Lambda \prod\nolimits_{i=1}^n \{p(x_i|\phi, \lambda)\} \pi(\lambda|\phi) d\lambda.$$

If $\pi(\lambda|\phi)$ is not proper, the procedure is performed within an increasing sequence of bounded approximations $\{\Lambda_j,\, j = 1, 2, \ldots\}$ to the nuisance parameter space Λ, chosen such that $\pi(\lambda|\phi)$ is integrable within each of them; the reference posterior is then the limit of the resulting sequence $\{\pi_j(\phi|x_1, \ldots, x_n),\, j = 1, 2, \ldots\}$ of posterior distributions (Berger and Bernardo, 1989, 1992b).

Question 34: Does this mean that, within a *single* model, you may have as many reference priors as possible parameters of interest?

Answer: It does indeed. Given a model, say $p(x \mid \theta_1, \theta_2)$, the reference algorithm provides a reference *posterior* distribution for *each* parameter of interest $\phi = \phi(\theta_1, \theta_2)$, and those may well correspond to different priors, because beliefs which maximize the missing information about $\phi = \phi(\theta_1, \theta_2)$ will generally differ from those which maximize the missing information about $\eta = \eta(\theta_1, \theta_2)$, unless η happens to be a one-to-one function of ϕ.

Note also that, as I mentioned before, using different priors for different parameters of interest is the *only* way to have non-subjective priors which avoid the marginalization paradoxes. For instance, in a normal model, $\mathrm{N}(x|\mu, \sigma)$, the reference posterior for μ is the Student distribution $\mathrm{St}(\mu|\overline{x}, (n-1)^{-1/2}s, n-1)$, obtained from the "conventional" improper prior $\pi(\sigma|\mu)\pi(\mu) = \sigma^{-1}$, while the reference posterior for $\phi = \mu/\sigma$ is obtained from $\pi(\sigma|\phi)\pi(\phi) = (2+\phi^2)^{-1/2}\sigma^{-1}$, a *different* improper prior, producing a reference posterior for ϕ which *avoids* the marginalization paradox that you would get if you used again the conventional prior (Bernardo, 1979).

Question 35: We now see how to deal with a single nuisance parameter, but how do you proceed when there are more than one?

Answer: The algorithm I have just described may easily be extended to any number $\{\lambda_1, \ldots, \lambda_m\}$ of *ordered* nuisance parameters: get the one-parameter conditional reference prior $\pi(\lambda_m|\phi, \lambda_1, \ldots, \lambda_{m-1})$ and use this to integrate out λ_m; get $\pi(\lambda_{m-1}|\phi, \lambda_1, \ldots, \lambda_{m-2})$ and use this to integrate out λ_{m-1}; continue until you get $\pi(\phi)$; then use

$$\pi(\lambda_m|\phi, \lambda_1 \ldots, \lambda_{m-1})\pi(\lambda_{m-1}|\phi, \lambda_1 \ldots, \lambda_{m-2}) \times \ldots \times \pi(\lambda_1|\phi)\pi(\phi)$$

in Bayes theorem, and marginalize to obtain the desired reference posterior $\pi(\phi|x_1, \ldots, x_n)$.

The result *might* possibly depend on the *order* in which the nuisance parameters are considered which, in that case, should reflect their order of importance in the problem analysed, the least important being integrated out first. We have found however that this is usually *not* the case: in most problems, the reference posterior of the quantity of interest is independent of the order in which the nuisance parameters are considered.

Question 36: Can you give some examples of this?

Answer: Sure. In a multinomial model, $\mathrm{Mu}(r_1, \ldots, r_m|\theta_1, \ldots, \theta_m, n)$, the reference posterior for, say, θ_1, is the Beta distribution $\mathrm{Be}(\theta_1|r_1+1/2, n-r_1+1/2)$ and this is independent of the order in which the other θ_i's are considered (Berger and Bernardo, 1992a); Note, by the way, that this does depends *not* on the irrelevant number of categories m as the posterior from Jeffreys' multivariate prior does; thus, the reference algorithm avoids the *agglomeration paradox* typically present in other proposals. Similarly, within the same model, the reference posterior for $\phi = \theta_1/\theta_2$ is

$$\pi(\phi|r_1, \ldots, r_m, n) \propto \frac{\phi^{r_1-1/2}}{(1+\phi)^{r_1+r_2+1}}\,,$$

(which, again, does not depend on m, but corresponds to a *different* prior), and this is independent of the order in which the nuisance parameters are considered (Bernardo and Ramón, 1996). Many more examples are referenced in Yang and Berger (1996).

Question 37: What has now happened to the invariance properties on which you insisted before?

Answer: They are still there. The reference posterior of any quantity of interest ϕ does not depend on whether one uses the full model or the joint sampling distribution of a set of sufficient statistics. Moreover, for any model $p(\boldsymbol{x} \mid \phi, \lambda_1, \ldots, \lambda_m)$, the reference posterior $\pi(\phi \mid \boldsymbol{x})$ does not depend on the particular parametrization chosen for each of the nuisance parameters and, besides, if $\theta = \theta(\phi)$ is a one-to-one transformation of ϕ, then $\pi(\theta|\boldsymbol{x}) = \pi(\phi|\boldsymbol{x})|d\phi/d\theta|$. Datta and Ghosh (1996) have recently shown that these invariance properties are often *not* shared by other proposed methods to derive non-subjective posteriors.

Question 38: How do you compute in practice reference distributions?

Answer: Reference priors only depend on the model through its asymptotic behaviour; essentially, if you know the asymptotics of your model, than you may easily find its associated reference priors. Under regularity conditions for asymptotic normality, any reference prior may be obtained from a relatively simple algorithm in terms of Fisher's matrix (Berger and Bernardo, 1992b). However, the derivation of reference priors in non-regular or complex models may be a difficult mathematical problem.

Of course, once you have obtained the appropriate reference prior for some quantity of interest, you simply use Bayes theorem and marginalize to derive the required reference posterior. It turns out that, within the exponential family, reference priors often correspond to some limiting form of the corresponding natural conjugate family and, in that case, the corresponding reference posteriors may often be obtained in closed form. When this is not the case, numerical reference posteriors may be efficiently obtained using MCMC sampling-resampling techniques, as described by Stephens and Smith (1992).

Question 39: We now have a procedure to derive reference posterior distributions when no prior information is available about the parameter of interest; however, even for scientific communication, one may sometimes want to use some *partial* information (possibly intersubjectively agreed). Can reference analysis deal with this situation?

Answer: It surely can; you define the reference prior under partial information as that which maximizes the missing information subject to whatever constraints are imposed by the information assumed. If the restrictions take the form of expected values, then explicit forms for the corresponding restricted reference priors are readily obtained (Bernardo and Smith, 1994, pp 316–320). Note that restricted reference analysis typically leads to *proper* priors; for instance, in a location model, the reference prior which corresponds to the partial information provided by the first two moments of the unknown parameter is the *normal* distribution with those moments.

Question 40: Please, talk about the new developments on reference priors.

Answer: I can see several directions in which further research is needed:

(i) *Bounded approximations*: I have mentioned before that to implement the reference algorithm in multiparameter problems when the conditional reference priors are not proper, a bounded approximation to the parameter space in which the conditional reference priors are integrable is required. It may be seen (Berger and Bernardo, 1989) that the result *may* depend on the bounded approximation chosen. Although in specific applications it is usually clear what the "natural" bounded approximation is, —and this should be the *same* for all parameters of interest within the same model—, a general definition of the appropriate bounded approximation is needed.

(ii) *Grouping*: With many parameters, one may apply the reference prior algorithm by lumping the parameters in just two groups (parameters of interest and nuisance parameters), or one may lump them into any number of groups and proceed sequentially (Berger and Bernardo, 1992a, 1992b, 1992c). There is evidence to suggest, however, that one should *not* group the parameters but proceed recursively using the one-parameter solution as I have described to you before; this seems to guarantee both admissible coverage properties and the absence of marginalization paradoxes, but further research is needed to substantiate this point.

(iii) *Prediction and hierarchical models*: A reference prior is technically defined for an ordered parametrization suggested by the problem of interest. What are the appropriate ordered parametrizations to use in prediction and hierarchical model problems? Again, although some answers are available in specific cases, the general strategy is not clear.

(iii) *Model choice*: Reference distributions are not directly applicable to model choice between models of different dimensionalities: indeed, reference distributions are typically only defined up to proportionality constant, and those constants become relevant in this case. Nice results are available however (Bernardo, 1996) by posing the question as a decision problem, and working with the reference posterior of the quantity of interest implied by the corresponding utility function.

(iv) *Numerical reference analysis*: The derivation of reference priors may sometimes be a difficult mathematical problem, but numerical reference posteriors may be obtained, in principle, by simulation methods. This is easily done in one or two parameters, but the general problem is not trivial (Efstathiou, 1996, Ch. 5), as computational explosion has to be avoided in higher dimensions.

Question 41: Is there anything to be said about the long-term frequentist properties of non-subjective posteriors? I realize that good Bayesians should not be raising this type of a question but, politically speaking, it may be wise to raise the issue.

Answer: Politics aside, this is a very interesting issue, and one that is central to discussions on *comparative* statistical inference. Interest on the frequentist coverage probabilities of credible intervals derived from non-subjective posteriors has a long history; key references include the pioneering work by Welch and Peers (1963) that I have already mentioned, Peers (1965), Hartigan (1966), Tibshirani (1989), Ghosh and Mukerjee (1992), Mukerjee and Day (1993), Nicolau (1993), and Datta and Ghosh (1995). The coverage probabilities of credible regions has often been an important element in arguing among competing non-subjective posteriors, as in Berger and Bernardo (1989) or Ye and Berger (1991) and, indeed, discussions on the coverage properties of non-subjective priors were prominent in the workshop on "default" Bayesian methodology recently held in the States (Purdue University, November 1–3, 1996). Reference posteriors have consistently been found to have attractive coverage properties, —what may be seen as a form of *calibration*— but, as far as I am aware, no general results have been established.

Question 42: Is there any other issue on comparative inference over which non-subjective priors may have a bearing?

Answer: A very important one is the procedure used to eliminate nuisance parameters. The marginalization paradox examples may be used to demonstrate than non-Bayesian methods to eliminate nuisance parameters (plug-in estimates, profile likelihood, naïve integrated likelihood and the like) are often *inconsistent within their own paradigms* in that the resulting 'marginal' likelihood may differ from the simplified likelihood. For example, the sampling distribution of the sample coefficient of correlation r in a bivariate normal model only depends on the population coefficient of correlation ρ, so that non-Bayesian statisticians would presumably consider $p(r \mid \rho)$ to be an 'exact' 'marginal' likelihood from which inferences about ρ could be made. Yet, the more sophisticated non-Bayesian techniques to eliminate nuisance parameters fail to derive $p(r \mid \rho)$ (Efron, 1993), while integration of the nuisance parameters with the conditional reference priors (taken in any order) easily produces the 'exact' marginal likelihood (Bayarri, 1981; Lindley, 1965, pp. 215–219, Bernardo and Smith, 1994, pp. 363–364).

Question 43: What about admissibility?

Answer: Non-subjective priors are sometimes criticized on the grounds that, since they are often improper, they may lead, for instance, to inadmissible estimates. We have seen, however, that sensible non-subjective priors are, in an appropriate sense, limits of proper priors; regarded as a "baseline" for admissible inferences, non-subjective posteriors need not be themselves admissible, but only arbitrarily close to admissible posteriors. That said, admissibility is not really the relevant concept: truncating the parameter space may lead to technically admissible but very unsatisfactory posteriors, as in the sum of squares of normal means example I described before. Besides, admissibility crucially depends on the loss function; thus, if one is really interested in estimation, one should explicitly work in terms of the corresponding decision problem, with an appropriate, context dependent, loss function. To deal with those problems, reference analysis may be extended to define *reference decisions*, as those optimal under the prior which maximizes the *missing utility* (Bernardo, 1981, Bernardo and Smith, 1994, Sec. 5.4.1). This is, by the way, another very promising area for future research.

Question 44: Could you talk about general criticisms to non-subjective priors?

Answer: The major criticism usually comes from subjectivist Bayesians: the prior should be an honest expression of the analyst's prior knowledge, not a function of the model, specially if this involves integration over the sample space and hence violates the likelihood principle. I believe there are two complementary answers to this:

(i) *Foundational*: A non-subjective posterior is the answer to a *what if* question, namely what could be said about the quantity of interest given the data, if one's prior knowledge was dominated by the data; if the experiment is changed, or a different quantity of interest is considered, the non-subjective prior may be expected to change correspondingly. If subjective prior information is specified, the corresponding subjective posterior could be compared with the non-subjective posterior in order to assess the relative importance on the initial opinions in the final inference.

(ii) *Pragmatic*: In the complex multiparameter models which are now systematically used as a consequence of the availability of numerical MCMC methods, there is little hope for a detailed assessment of a huge personal multivariate prior; the naïve use of some "tractable" prior may then hide important unwarranted assumptions which may easily dominate the analysis (see *e.g.*, Casella, 1996, and references therein). Careful, responsible choice of a non-subjective prior is possibly the best available alternative.

Question 45: Could we end with some signposts for those interested in pursuing this discussion at a more technical level?

Answer: The classic books by Jeffreys (1961), Lindley (1965) and Box and Tiao (1973) are a must for anyone interested in non-subjective Bayesian inference; other relevant books are Zellner (1971) and Geisser (1993).

The construction of non-subjective priors has a very interesting history, which dates back to Laplace (1812), and includes Jeffreys (1946, 1961), Perks (1947), Lindley (1961), Geisser and Cornfield (1963), Welch and Peers (1963), Hartigan (1964, 1965), Novick and Hall (1965), Jaynes (1968, 1971), Good (1969), Villegas (1971, 1977), Box and Tiao (1973, Sec. 1.3), Zellner (1977, 1986), Bernardo (1979), Rissanen (1983), Tibshirani (1989) and Berger and Bernardo (1989, 1992c) as some of the more influential contributions.

For a general overview of the subject, see Bernardo and Smith (1994, Sec. 5.6.2), Kass and Wasserman (1996), and references therein. Yang and Berger (1996) is a problem-specific (partial) catalog of the many non-subjective priors which have been proposed in the literature.

For someone specifically interested in reference priors, the original paper, Bernardo (1979) is easily read and it is followed by a very lively discussion; Berger and Bernardo (1989, 1992b) contain crucial extensions; Bernardo and Smith (1994, Sec. 5.4) provide a description of reference analysis at a textbook level; Bernardo and Ramón (1996) offer a modern elementary introduction.

REFERENCES

Bayarri, M. J. (1981). Inferencia Bayesiana sobre el coeficiente de correlación de una población normal bivariante. *Trab. Estadist.* **32**, 18–31.

Berger, J. O. and Bernardo, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.* **84**, 200–207.

Berger, J. O. and Bernardo, J. M. (1992a). Ordered group reference priors with applications to a multinomial problem. *Biometrika* **79**, 25–37.

Berger, J. O. and Bernardo, J. M. (1992b). On the development of reference priors. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 35–60 (with discussion).

Berger, J. O. and Bernardo, J. M. (1992c). Reference priors in a variance components problem. *Bayesian Analysis in Statistics and Econometrics* (P. K. Goel and N. S. Iyengar, eds.). Berlin: Springer, 323–340.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. B* **41**, 113–147 (with discussion). Reprinted in *Bayesian Inference* (N. G. Polson and G. C. Tiao, eds.), Brookfield, VT: Edward Elgar, (1995), 229–263.

Bernardo, J. M. (1981). Reference decisions. *Symposia Mathematica* **25**, 85–94.

Bernardo, J. M. (1996). Bayesian hypothesis testing: a reference analysis. *Workshop on Default Bayesian Statistical Methodology*. Purdue University, November 1–3, 1996.

Bernardo, J. M. and Ramón J. M. (1996). An elementary introduction to Bayesian reference analysis. *Tech. Rep.*, Universitat de València, Spain.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. Chichester: Wiley.

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley.

Casella, G. (1996). Statistical inference and Monte Carlo algorithms. *Test* **5**, 249–340, (with discussion).

Clarke, B. and Barron, A. R. (1994). Jeffreys' prior is asymptotically least favorable under entropy risk. *J. Statist. Planning and Inference* **41**, 37–60.

Cifarelli, D. M. and Regazzini, E. (1987). Priors for exponential families which maximize the association between past and future observations. *Probability and Bayesian Statistics* (R. Viertl, ed.). London: Plenum, 83–95.

Consonni, G. and Veronese, P. (1989). Some remarks on the use of improper priors for the analysis of exponential regression problems. *Biometrika* **76**, 101–106.

Datta, G. S. and Ghosh, J. K. (1995). On priors providing a frequentist validity for Bayesian inference. *Biometrika* **82**, 37–45.

Datta, G. S. and Ghosh, M. (1996). On the invariance of noninformative priors. *Ann. Statist.* **24**, 141–159.

Dawid, A. P. (1983). Invariant prior distributions. *Encyclopedia of Statistical Sciences* **4** (S. Kotz, N. L. Johnson and C. B. Read, eds.). New York: Wiley, 228–236.

Dawid, A. P., Stone, M. and Zidek, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference. *J. Roy. Statist. Soc. B* **35**, 189–233 (with discussion).

de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* **7**, 1–68. Reprinted in 1980 as 'Foresight; its logical laws, its subjective sources' in *Studies in Subjective Probability* (H. E. Kyburg and H. E Smokler, eds.). New York: Dover, 93–158.

Diaconis, P. (1988a). Recent progress on de Finetti's notion of exchangeability. *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.). Oxford: University Press, 111-125 (with discussion).

Efron, B. (1993). Bayes and likelihood calculations from confidence intervals. *Biometrika* **80**, 3–26.

Efstathiou, M. (1996). *Some Aspects of Approximation and Computation for Bayesian Inference*. Ph.D. Thesis, Imperial College London, UK.

Fishburn, P. C. (1981). Subjective expected utility: a review of normative theories. *Theory and Decision* **13**, 139–199.

Fishburn, P. C. (1986). The axioms of subjective probability. *Statist. Sci.* **1**, 335–358 (with discussion).

Geisser, S. (1993). *Predictive Inference: an Introduction*. London: Chapman and Hall.

Ghosh, J. K. and Mukerjee, R. (1992). Non-informative priors. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 195–210 (with discussion).

Good, I. J. (1969). What is the use of a distribution? *Multivariate Analysis* **2** (P. R. Krishnaiah, ed.). New York: Academic Press, 183–203.

Hartigan, J. A. (1964). Invariant prior distributions. *Ann. Math. Statist.* **35**, 836–845.

Hartigan, J. A. (1965). The asymptotically unbiased prior distribution. *Ann. Math. Statist.* **36**, 1137–1152.

Hartigan, J. A. (1966). Note on the confidence prior of Welch and Peers. *J. Roy. Statist. Soc. B* **28**, 55-56.

Hartigan, J. A. (1983). *Bayes Theory*. Berlin: Springer.

Heath, D. L. and Sudderth, W. D. (1978). On finitely additive priors, coherence and extended admissibility. *Ann. Statist.* **6**, 333–345.

Heath, D. L. and Sudderth, W. D. (1989). Coherent inference from improper priors and from finitely additive priors. *Ann. Statist.* **17**, 907–919.

Hewitt, E. and Savage, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80**, 470–501.

Jaynes, E. T. (1968). Prior probabilities. *IEEE Trans. Systems, Science and Cybernetics* **4**, 227–291.

Jaynes, E. T. (1971). The well posed problem. *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.). Toronto: Holt, Rinehart and Winston, 342–356 (with discussion).

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. A* **186**, 453–461.

Jeffreys, H. (1961). *Theory of Probability* (3rd ed.). Oxford: University Press.

Kashyap, R. L. (1971). Prior probability and uncertainty. *IEEE Trans. Information Theory* **14**, 641–650.

Kass, R. E. (1989). The geometry of asymptotic inference. *Statist. Sci.* **4**, 188–234.

Kass, R. E. (1990). Data-translated likelihood and Jeffreys' rule. *Biometrika* **77**, 107–114.

Kass, R. E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *J. Amer. Statist. Assoc.* **91**, (to appear).

Lindley, D. V. (1956). On a measure of information provided by an experiment. *Ann. Math. Statist.* **27**, 986–1005.

Lindley, D. V. (1961). The use of prior probability distributions in statistical inference and decision. *Proc. Fourth Berkeley Symp.* **1** (J. Neyman and E. L. Scott, eds.). Berkeley: Univ. California Press, 453–468.

Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint. Vol. 2: Statistics.* Cambridge: University Press.

Lindley, D. V. (1996). Some comments on Bayes Factors. *J. Statist. Planning and Inference* (to appear).

Mukerjee, R. and Dey, D. K. (1993). Frequentist validity of posterior quantiles in the presence of a nuisance parameter: Higher order asymptotics. *Biometrika* **80**, 499–505.

Nicolau, A. (1993). Bayesian intervals with good frequentist behaviour in the presence of nuisance parameters. *J. Roy. Statist. Soc. B* **55**, 377–390.

Novick, M. R. and Hall, W. K. (1965). A Bayesian indifference procedure. *J. Amer. Statist. Assoc.* **60**, 1104–1117.

Peers, H. W. (1965). On confidence points and Bayesian probability points in the case of several parameters. *J. Roy. Statist. Soc. B* **27**, 9–16.

Perks, W. (1947). Some observations on inverse probability, including a new indifference rule. *J. Inst. Actuaries* **73**, 285–334 (with discussion).

Renyi, A. (1962). *Wahrscheinlichkeitsrechnung*. Berlin: Deutscher Verlag der Wissenschaften. English translation in 1970 as *Probability Theory*. San Francisco, CA: Holden-Day.

Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length. *Ann. Statist.* **11**, 416–431.

Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley. Second edition in 1972, New York: Dover.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Tech. J.* **27** 379–423 and 623–656. Reprinted in *The Mathematical Theory of Communication* (Shannon, C. E. and Weaver, W., 1949). Urbana, IL.: Univ. Illinois Press.

Stephens, D. A. and Smith, A. F. M. (1992). Sampling-resampling techniques for the computation of posterior densities in normal means problems. *Test* **1**, 1–18.

Smith, A. F. M. (1981). On random sequences with centred spherical symmetry. *J. Roy. Statist. Soc. B* **43**, 208–209.

Stein, C. (1959). An example of wide discrepancy between fiducial and confidence intervals. *Ann. Math. Statist.* **30**, 877–880.

Stone, M. and Dawid, A. P. (1972). Un-Bayesian implications of improper Bayesian inference in routine statistical problems. *Biometrika* **59**, 369–375.

Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika* **76**, 604–608.

Villegas, C. (1971). On Haar priors. *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.). Toronto: Holt, Rinehart and Winston, 409–414 (with discussion).

Villegas, C. (1977). Inner statistical inference. *J. Amer. Statist. Assoc.* **72**, 453–458.

Wasserman L. (1991). An inferential interpretation of default priors. *Tech. Rep.* **516**, Carnegie Mellon University, USA.

Welch, B. L. and Peers, H. W. (1963). On formulae for confidence points based on intervals of weighted likelihoods. *J. Roy. Statist. Soc. B* **25**, 318–329.

Yang, R. (1995). Invariance of the reference prior under reparametrization. *Test* **4**, 83–94.

Yang, R. and Berger, J. O. (1996). A catalog of noninformative priors. *Tech. Rep.*, Purdue University, USA.

Ye, K. and Berger, J. O. (1991). Non-informative priors for inferences in exponential regression models. *Biometrika* **78**, 645–656.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley. Reprinted in 1987, Melbourne, FL: Krieger.

Zellner, A. (1977). Maximal data information prior distributions. *New Developments in the Applications of Bayesian Methods* (A. Aykaç and C. Brumat, eds.). Amsterdam: North-Holland, 211–232.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti* (P. K. Goel and A. Zellner, eds.). Amsterdam: North-Holland, 233–243.