Bayesian Analysis or Evidence Based Statistics

D.A.S. Fraser

Department of Statistics, University of Toronto, Canada M5S 3G3

Introduction

The original Bayes proposal leads to likelihood and confidence for many simple examples. More generally it gives approximate confidence but to achieve exact confidence reliability it needs refinement of the argument and needs more than just the usual minimum of the likelihood function from observed data. A general Bayes approach provides a flexible and fruitful methodology that has blossomed in contrast to the widely-based long-standing frequentist testing with focus on the 5% level. We examine some key events in the evolution of the Bayes approach promoted as an alternative to the present likelihood based frequentist analysis of data with model, the evidence-based approach of central statistics. And we are led to focus on the bane of Bayes: parameter curvature.

1. Bayes, 1763

Bayes (1763) examined the Binomial model $f(y;\theta) = \binom{n}{\theta}\theta^y(1-\theta)^{n-y}$ and proposed the flat prior $\pi(\theta) = 1$ on $[0, 1]$. Then with data y^0 he used a lemma from probability calculus to derive the posterior $\pi(\theta|y^0) = c\theta^{y^0}(1-\theta)^{n-y^0}$ on $[0, 1]$. And then for an interval say $(\theta, 1)$ he calculated the integral of the posterior,

$$s(\theta) = \int_\theta^1 \theta^{y^0}(1-\theta)^{n-y^0}d\theta \Big/ \int_0^1 \theta^{y^0}(1-\theta)^{n-y^0}d\theta$$

and referred to it as probability that the parameter belonged to the interval $(\theta, 1)$. Many endorsed the proposed calculation and many disputed it.

As part of his presentation he used an analogy. A ball was rolled on a level table, perhaps an available billiard table, and was viewed as having equal probability of stopping in any equal sized area. The table was then divided conceptually by a North-South line through the position where the ball stopped, with area θ to the West and $(1-\theta)$ to the East. The ball was then rolled n further times and the number y^0 of time that it stopped left of the line observed. In the analogy itself, the posterior probability calculation given data seems entirely appropriate.

2. The Economist, 2000

In an article entitled "In praise of Bayes", the Economist (2000) speaks of an "increasingly popular approach to statistics (but) not everyone is persuaded of its validity". The article mentions many areas of recent application of the Bayesian approach, and cites "the essence ... is to provide a mathematical rule explaining how you should change your existing beliefs in the light of new evidence". The indicated areas of application are wide spread and there is emphasis on attaining definitive answers. And this is set in full contrast to "traditional ways of presenting results" indicated to be the mid-twentieth-centure decision theoretic view of accepting a null view 95% to 5% on some departure scale. The article does offer some caution for "when used indiscriminently" in the form of a quotation from Larry Wasserman that it can become "more a religion than a science".

The mathematical rule cited as the essence of the Bayesian approach is a very broad expansion from Bayes original proposal where a statistical model $f(y;\theta)$ evaluated at an observed data value y^0 giving $f(y^0;\theta)$ is combined with a constant mathematical prior $\pi(\theta)=1$ and treated as a conditional density. The force of the rule is that with new model-data information the new likelihood would be folded with the old. But this is of course standard practice in statistics: use the up-to-date likelihood, and possibly refine such a procedure with meta-analysis. What is different is that the Bayesian method essentially overlooks evidence beyond the observed likelihood function and does so on principle.

3. Validity or Analogy

Bayes considered a uniform prior and a Binomial (n,p) model, and used analogy to justify combining them by a standard lemma from probability calculus. For the analogy involving balls on a billiard table, the calculations seem entirely proper and appropriate. The more generally interpreted Bayes approach has a statistical model $f(y;\theta)$ with data y^0 coupled with a mathematical prior $\pi(\theta)$ representing symmetries or other properties of the model or context. Analogies can be great for explaining an argument but not to be the argument itself: there is no billiard table equivalent in the typical binomial or more general context.

There is an explicit time line: There is a context with a true value θ_* for the parameter θ; there is an investigation $f(y;\theta)$ yielding an observed y^0 from the true value model $f(y;\theta_*)$; and possibilities for θ are then to be assessed. Thus in order: θ_* is realized but unknown; y^0 is observed; then assess θ. The values θ_* and y^0 are realized and are in the past. And the issue is what can be said about θ given the model $f(y;\theta)$ and data y^0.

If θ is understood in fact to come from an objective source $\pi(\theta)$ with realized value θ_*; then the time line is longer. Accordingly: $\pi(\theta)$ produces θ_*; $f(y;\theta_*)$ produces y^0; and the issue is to assess θ. In this situation $\pi(\theta)$ is properly an objective prior. And an option is of course is to examine and present the composite model $\pi(\theta)f(y;\theta)$ with observed y^0. But an even more compelling option is to examine and present $\pi(\theta)$ and to separately examine and present $f(y;\theta)$ with y^0.

Now consider the model $f(y;\theta)$ with data y^0; and the mathematical prior $\pi(\theta)$ as proposed by Bayes. The lemma from the probability calculus has two probability inputs say $\pi(x)$ and $f(y|x)$ and it has one probability output $\pi(x|y^0)$; the output records the behavior of x that is associated with the observed value $y=y^0$.

For the Bayes case $\pi(x)$ would be $\pi(\theta)$ and $\pi(x|y^0)$ would be $\pi(\theta|y^0)$. Is the lemma applicable or relevant in the Bayes case? In the Bayes case there is just one probability input $f(y;\theta)$; and the other nominal input is $\pi(\theta)$, a mathematical object that refers to symmetry or patterns in the model and has no probability status whatsoever. Thus the assumptions of the lemma do not hold, and consequently the output of the lemma is ... by analogy ... not by derivation. The usage of the lemma in the proposed argument is not proper and can be viewed as fraudulent logic.

The standard frequentist would refer to $f(y^0;\theta)$ as likelihood $L(\theta;y^0) = L^0(\theta)$. An exploration with weighted likelihood $\pi(\theta)L^0(\theta)$ can be a very natural, obvious and sensible procedure ... for just that, for exploring possibilities for θ. But for obtaining probabilities, perhaps a pipe dream!

4. Likelihood and Confidence

Bayes' (1763) original approach suggested a density $c\pi(\theta)f(y^0;\theta)$ as a description of an unknown θ in the presence of observed data y^0. As such it records likelihood $L^0(\theta)$ or weighted likelihood. And this was long before the formal introduction (Fisher, 1922) of likelihood. Both viewpoints record the same formal information concerning the parameter; the differences are in the color or flavor associated with the particular argument; and with properties attributed to the output.

Bayes (1763) also offered a distribution as a summary of information concerning the parameter θ; the distribution had density $c\pi(\theta)f(y^0;\theta) = c\pi(\theta)L^0(\theta)$. The majority of models at that time had least-squares location structure and for such models the posterior $\pi(\theta)L^0(\theta)$ using a natural prior just reproduces what is now called confidence (Fisher, 1930, 1935).

It thus seems appropriate to acknowledge that Bayes introduced the primary concepts of likelihood and confidence long before Fisher (1922, 1930) and long before the refinement offered by Neyman (1937). For likelihood he offered the extra flexibility of the weight function but for confidence he did not have the statistical refinement that later provided the logical extension to non-location models; this latter can be viewed as a matter of reasonable fine tuning of the argument, of intellectual evolution, and of the familiar iterative processes of science.

5. Laplace and Venn

Laplace (1812) seems to have fully endorsed the proposals arising from Bayes (1763). And Venn (1886) seems equally to have rejected them. Certainly asserting the conclusions of a theorem or lemma when one of the premises does not hold is unacceptable from a mathematical or logical viewpoint. Nonetheless the results were impressive and filled a substantial need, but indeed with downstream risks. And it does have, as is now becoming apparent, the support of approximate confidence (Fraser, 2010). At present Bayes and confidence lead a coexistence, perhaps an uneasy unstable coexistence!

6. Priors and Priors

The original Bayes prior was a flat prior $\pi(\theta) = 1$ for a probability θ that then in sequence becomes the parameter in a Binomial (n, θ) model; the resulting posterior is $\pi(\theta)L^0(\theta)$, which focally uses the observed likelihood from the Binomial context. Some aspects of invariance were invoked to support the particular choice. The possible plausible extensions are immense.

For a location model $f\{y-\beta(\theta)\}$ the natural prior would be $\pi(\theta)d\theta = d\beta(\theta) = \beta'(\theta)d\theta$. Thus for $f(y - X\beta)$ we would have $\pi(\beta)d\beta = dX\beta = cd\beta$, giving a flat prior for the regression coefficients. Motivation would come by noting that $y - \beta(\theta)$ has a fixed θ-free distribution.

Extensions are possible by seeking approximate θ-free distributions. This was initiated by Jeffreys (1939) and then fine-tuned to acknowledge various types of parameters (Jeffreys, 1946). These extensions use expected information $i(\theta) = E\{-\ell_{\theta\theta}(\theta; y); \theta\}$ in the model to calibrate the scale for θ; here $-\ell_{\theta\theta}(\theta)$ is the negative second derivative of likelihood and the initial Jeffreys prior is $\pi(\theta)d\theta = |i(\theta)|^{1/2}d\theta$, and it is parameterization invariant. For the regression model, where $y = X\beta + \sigma$ with $N(0, 1)$ error, the Jeffreys (1939) prior is $\pi(\theta)d\theta = d\beta d\sigma/\sigma^{r+1}$ where r is the column rank of X. The second or modified Jeffreys (1946) is $\pi(\theta)d\theta = d\beta d\sigma/\sigma$ and gives generally more acceptable results, often in agreement with confidence.

The approximate approach can be modified (Fraser et al, 2010) to work more closely with the location invariance indicated by the initial Bayes (1763) approach. In many regular problems continuity within the model leads to a relationship $d\hat{\theta} = W(\theta)d\theta$ where $W(\theta)$ is a $p \times p$ matrix; the $d\hat{\theta}$ refers to an increment at the data y^0 and the $d\theta$ refers to an increment $d\theta$ at θ. This immediately indicates the prior $\pi(\theta)d\theta = |W(\theta)|d\theta$ based on simple extension of the translation invariance $dy = \beta'(\theta)d\theta$ for the model $f(y-\beta(\theta))d\theta$; and it widely agrees with preferred priors in many problems. But the parameter must not have curvature: the bane of Bayes!

The approximate approach can also be modified to make use of an asymptotic result that to second order the statistical model can be treated as an exponential model (Reid & Fraser, 2010; Fraser, Reid, Marras, Yi, 2010). This uses continuity to obtain a nominal reparameterization $\varphi(\theta)$ that yields second and third order inference by acting as if the model were just $g(s; \theta) = \exp\{\ell(\theta) + \varphi(\theta)s\}h(s)$ with data $s^0 = 0$. This allows information to be calculated within the approximating model using the information function $j_{\varphi\varphi}(\theta; s) = -\ell_{\varphi\varphi}\{\theta(\varphi)\}$; this draws attention to marginalization and to curvature effects that are not usually apparent in the search for default priors (Fraser, Reid, Marras, Yi, 2010).

The preceding can also be viewed as a somewhat natural evolution from the original Bayes proposal with some reference to location invariance. The evolution has been assisted by the fact that many posterior distributions have appealing and sensible properties. It is our view here that these sensible properties are precisely the approximate confidence properties that have become evident quite separately. In any case the priors just described can all be classified as default priors, priors that one might choose to use as a default without strong arguments for something different.

The Bayesian approach is committed to using a weight function applied to an observed likelihood and thus to formally omitting other properties of the model. Within this approach the default priors are widely called objective priors. But the term objective means objective reference, and this as a property is specifically absent here; there is a strong flavour of deception. Thus using the term objective for default priors seems highly inappropriate, but could be viewed as just a seeking for a wider area of application. The author was present at the Bayesian convention when the term was being chosen and did not register an objection, being perhaps somewhat of an outsider, not a good defense! We will however explicitly refer to them as default priors, and keep the term objective for contexts where the prior does have an explicit reference in context.

A difficulty with the use of default priors is that a posterior probability obtained by marginalization from a full posterior distribution may not be equal the posterior probability calculated directly from the appropriate marginal model; this was given prominence by Dawid et al (1973) and applies equally to confidence distributions and other attempts to present model-data information as a distribution for the parameter. The complication in any of these cases derives from parameter curvature: for some discussion see Fraser, Reid, Marras, Yi (2010) and Fraser & Sun (2010).

The wealth of possibilities available from a weight-function combined with likelihood is well documented in the development of the Bayesian methods as just described. Its success can amply be supported as "approximate confidence" but derived by a route that is typically much easier. Approximate confidence provides full support for the acceptable, often meritorious behavior of Bayes posterior probabilities. We address later whether there can be anything beyond approximate confidence in support of the Bayesian approach.

Another approach, somewhat different from the original Bayes way of obtaining a weight function is derived from Kullback-Leibler distance on measure spaces (Bernardo, 1971): this chooses a prior to maximize the statistical distance from prior to posterior. Modifications of this distance approach have been developed to obtain specialized priors for different component parameters of interest, often parameters that have a statistical curvature.

The richness available from using just a likelihood function is clearly evident to Bayesians if not to frequentists; but is not widely acknowledged. Much of recent likelihood theory divides on whether or not to use more than the observed likelihood, specifically sampling properties that are associated with likelihood characteristics but are not widely or extensively available. In many ways central statistics has ignored the extra in going beyond likelihood, and indeed has ignored the wealth available from just likelihood alone.

Meanwhile those committed to using just the weighted likelihoods, those associated with the development of the Bayes approach as we have just described, have aggressively sought to use the weighted likelihood approach as a general approach to updating information and to producing decisions. Central to this direction is the subjective approach with a major initiative coming from Savage (1972). This takes a prior to represent the views, the understanding, the personal probabilities concerning the true value of the parameter; these might come from highly personal thoughts, from detailed elicitation from knowledgeable people, from gut feelings as one approaches a game at a casino; and they can have the benefit of intuition or the merits of a seasoned gambler, with or without insider information. But who should use it? Certainly the chronic gambler will. But from the statistical perspective here there is nothing of substance to say that such prior 'information' $\pi(\theta)$ should be combined with likelihood. With due respect it can be presented as $\pi(\theta)$ alongside a presentation of the evidence-based well calculated confidence. If a user would like to combine them, it would certainly be plausible for him to do so but it would not be an imperative despite Bayesian persuasion. Certainly place than both to be seen and available. In wide generality combining them is not a necessary statistical step, although sometimes expedient.

7. Lindley and Territory

Fisher's (1930, 1935) proposal for confidence with effective support from Neyman (1937) offered strong alternatives to a prominent sympathy for the Bayesian approach. Then Jeffreys (1939, 1946) with great prominence in geophysics provided reinforcement for the use of the Bayesian approach in the physical sciences. Meanwhile the confidence approach gained strength both in mathematics departments and in scientific applications. Both approaches lead from model and data to a distribution for the parameter, but the results were often in conflict. Both sides clearly felt threatened, and each side in a practical sense had territory to defend.

Lindley (1958) focussed on the very basic case, a scalar parameter and a scalar variable, say with distribution function $F(y;\theta)$. The Bayesian answer with prior $\pi(\theta)$ is given by the posterior distribution $c\pi(\theta)F_y(y;\theta)d\theta$ where the subscript y denotes differentiation with respect to the argument y thus giving the density or likelihood function. By contrast the Fisher (1930, 1935) approach gives the confidence distribution $|F_\theta(y;\theta)|d\theta$. Lindley examined when these would be equal and solved for $\pi(\theta)$:

$$\pi(\theta) = c\frac{F_{;\theta}(y;\theta)}{F_y(y;\theta)} = c\frac{\partial}{\partial\theta}y(u;\theta);$$

the right hand expression records the derivative of the quantile function for fixed p-value $u = f(y;\theta)$ as pursued in Fraser, Reid, Marras, Yi (2010). The equation is actually a differential equation that asserts that the model must be a location model, the form of model actually found in Section 6 to have good Bayesian answers. In Fraser, Reid, Marras, Yi (ibid) the equation is used to determine the data dependent priors that give posterior probabilities having objective validation.

Lindley was concerned that the confidence approach did not follow the primal Bayesian concept that a probability statement concerning a parameter should be updated by multiplication by new likelihood and his criticism had a profound effect suggesting that the confidence distribution approach was defective. We now know that the defect is the attempt to use a distribution as the summary, either by Bayes or by confidence. And if there were to be a lesser of two evils then calling confidence probability and calling Bayes approximate confidence would be safer.

Another view might be that this was just a territorial dispute as to who had the rights to provide a distributional description of the parameter in the model data context. But the social conflict aspects were not in evidence. Rather there was a wide spread perception that giving a distribution of confidence was wrong. Neyman (1937) of course had provided a route around. But nonetheless, the judgement stuck: a confidence distribution was wrong and a Bayesian analysis was all right. Of course, in Dawid, Stone, and Zidek (1973), there is a clear message that neither approach can handle vector parameters without special fine-tuning. Clearly Lindley had focused on a substantive issue but the arguments invoked had not quite attained the point of acknowledging that an effective prior must in general be data dependent; for some current discussion see Fraser, Reid, Marras, Yi (2010).

8. Bayesian analysis and imperatives

Bayesian (1763) analysis has been around for a long time, but alternative views perhaps now identified as frequentist are perhaps older although somewhat less formalized. These approaches have cross dialogued and often been in open conflict. Each has made various appeals to holding the truth. And they have actively sought territorial advantage. In particular Fisher's (1930) initial steps towards confidence were directly to provide an alternative to inverse probability, the name at the time attached to the Bayesian approach. So it is not surprising that there would be a very focal reverse criticism (Lindley, 1958) of the confidence approach.

Those favoring the Bayesian approach have frequently felt they were underdogs, often for example having their articles rejected by journals for just being Bayesian. It thus seems rather natural that the Bayesian supporters would seek to broaden their methodology and their community. The subjective approach as strongly initiated by Savage (1954) has led to a powerful following in an area where prior probabilities are extended to include personal feelings, elicited feelings, and betting view points. Certainly such extensions are a guide for gambling and much more. But there is nothing of substance to assert that they should be ... the imperative ... used for the analysis. The prior subjective assessment and the objective evidence-based assessment can be placed side by side for anyone to see and to use as deemed appropriate. And the Bayes combination of these can also be presented for any one to use if so inclined. Perhaps the Bayesian expansion was ill advised to promote the imperative: that the proper analysis was that of the Bayes paradigm.

What is perhaps even more dangerous is the widely promoted hierarchical model where each parameter is given a prior distribution, and then parameters in the prior distributions are themselves given priors, perhaps then multilevel. Often an impressive edifice that seems only equalled by the lack of evidence for the various introduced elements and the impressive resort to McMC. The resort to multilevel Bayes modeling would seemingly be best viewed as one of expediency, to extend the base of Bayes without supporting evidence.

And then of course there are model data situations where the true parameter has come from a source with a known frequency distribution. In such cases the obvious name for the prior would be objective prior. But assembled Bayesian as mentioned earlier have adopted that name for the opposite situation, where there is in fact no objective reference, and the prior is purely a mathematical construct. But what about the multitude of cases where there is an identified source for the true parameter value? These can arise widely when the entity being examined has been obtained by sampling from some identified subpopulation; or they can arise by genetics or by Mendel or perhaps by updated genetic laws. Or much more. In reality this is just a modeling issue: what aspect of the context, of the immediate environment, or the more extended environment should be modeled. It is a modeling issue. It is perhaps only natural that Bayesian promotion should seek to subsume wider and wider contexts as part of the evolution of the approach. Especially when traditional statistics has been widely immersed in technical criteria connected with some global optimization or with decision rules to reject at some 5% level or accept at some 19/20 level, even when it was becoming abundantly apparent that these rule for scientific publication have serious defects.

But if there is an objective source $\pi(\theta)$ for a true value in a model-data context, there is nothing that says it should be folded into a combined model for analysis. The prior source $\pi(\theta)$ can be set in parallel with the more directly evidence-based analysis of the model-data combination. And of course even the combined model-data-prior analysis presented. But again there is no substantive precept that says the combined analysis is the statistical inference. Such a step would be purely an assertion of extended territory for Bayesian analysis.

9. Curvature: The Bane of Bayes

Contours of a parameter can have obvious curvature. A simple example can throw light on the effects of such curvature.

Consider (y_1, y_2) with a Normal $\{(\theta_1, \theta_2); I\}$ distribution on the plane. With data (y_1^0, y_2^0) the basic original Bayes approach would say that (θ_1, θ_2) was Normal $\{(y_1^0, y_2^0); I\}$. First we examine an obviously linear parameter $\psi = \theta_1$ and assess say the value $\psi = 0$ on the basis of data, say $(y_1^0, y_2^0) = (0, 0)$.

In an obvious way y_1 measures ψ and has the Normal$(\psi; 1)$ distribution. Accordingly the p-value for ψ from the observed data is

$$p(\psi) = \Phi\{(y_1^0 - \psi)/1\} = \Phi(-\psi).$$

And for assessing the value $\psi = 0$ we have $p(0) = 50\%$.

Now consider the Bayesian assessment of the value ψ. the marginal posterior distribution of ψ is $N(y_1^0, 1)$ and the corresponding posterior survivor value is

$$s(\psi) = 1 - \Phi((\psi - y_1^0)/1) = 1 - \Phi(\psi)$$

at the observed data. In particular for assessing $\psi = 0$ we would have $s(0) = 50\%$. the Bayesian and frequentist values are equal for the special $\psi = 0$ and also for general ψ.

Now consider a clearly curved parameter, the distance ψ on the parameter space from the point $(-1, 0)$ to the parameter value (θ_1, θ_2),

$$\psi = \{(\theta_1 + 1)^2 + \theta_2^2\}^{1/2}.$$

An obvious way to measure this parameter is by using the distance r from the point $(-1, 0)$; thus $r = \{(y_1 + 1)^2 + y_2^2\}^{1/2}$. The distribution of r^2 is noncentral Chi-square with 2 degrees of freedom and noncentrality $\delta^2 = \psi^2$. The indicated p-value for assessing ψ is then

$$p(\psi) = H_2(r^2; \psi^2)$$

where H_2 is the noncentral Chi-square distribution function with 2 degrees of freedem and noncentrality $\delta^2 = \psi^2$. This is readily available in R. In particular for assessing $\psi = 1$ we would have

$$p(1) = H_2(1; 1) = 26.7\%$$

which is substantially less than 50%.

Now consider the Bayesian assessment of the curved parameter ψ. The posterior distribution of ψ^2 from the observed data is noncentral Chi-square with 2 degrees of freedom and noncentrality $\delta^2 = 1$. If follows that the posterior survivor value for assessing $\psi = 1$ is

$$s(1) = 1 - H_2(1; 1) = 73.3\%$$

which is substantially larger than 50%.

For this simple example we have seen that the p-value and the survivor value are equal for a linear parameter. This happens generally for linear parameters (Fraser & Reid, 2002). And with the introduction of a curvature change to the parameter, the Bayesian and frequentist values go in opposite directions. This happens widely with curved parameters: as a parameter contour is changed from linear to curved, the Bayesian survivor changes in the opposite direction from the frequentist. Thus the Bayesian can be viewed as correcting negativity, that is making an adjustment opposite to what is appropriate in a context. For some recent discussion see Fraser (2010). The example above suggests that curvature is precisely the reason that Bayes fails to correctly assess parameters.

Consider y with a Normal $\{\theta, \sigma^2(\theta)\}$ distribution where the variance $\sigma^2(\theta)$ depends weakly on the mean θ. Precise p-values are available for assessing θ:

$$p(\theta) = \Phi\{(y - \theta)/\sigma(\theta)\}$$

with a clear frequency interpretations. The confidence inversion is well established (Fisher; 1930, 1935). The Bayesian inversion does not seem to have an obvious prior that targets the parameter θ.

How does one assess the merits of a proposed distribution for a parameter? The use of two-sided intervals provides a slippery slope. Strange tradeoffs can be made between two interval bounds; see for example Fraser, Reid & Wong (2004) on statistics for discovering new particles in High Energy Physics. A more direct approach is to examine a particular quantile of a proposed distribution, say the β-th quantile $\hat{\theta}_\beta$ which has posterior probability β to the right and $(1-\beta)$ to the left. One can certainly simulate or have an oracle and determine what proportion of the time the true value is larger than the particular quantile being considered; and determine whether the true proportion bears a sensible relation to the alleged value β. This has been addressed at length in Fraser (2010).

In particular for the Normal$\{\theta, \sigma^2(\theta)\}$ example there is no determination of a prior that will give the third order accuracy that is available from the confidence approach unless the prior is directly specific to the observed data value. This result holds wide generality: the use of a default or Bayesian prior can not lead to the third order accuracy readily available from the evidence-based procedures of frequentist inference. And parameter curvature is the number one culprit.

10. Why Bayes?

Linear approximations are widely used throughout statistics, mathematics, physics, the sciences generally, and much more. They provide a local replica of something that might be intangible otherwise and when used iteratively can provide exploration of something unknown otherwise. There is substantial evidence that the Bayes procedure provides an excellent first order approximation for the analysis of a statistical model. There are also ample warnings that global acceptance of Bayes results can be extremely hazardous. Use but be cautious!

The Bayes calculus asserts that the posterior results are probabilities. And the name itself is assertive. The Bayesian supporters have also been vocal, asserting that confidence results do not have the status of probabilities calculated by the Bayes paradigm; some indication of this is implicit in Lindley (1958); and further indication is found in the active broadening of the application area for Bayesian analysis. From an evidence-based approach it is clear that the direct use of the likelihood function provides substantial information, first order information. And higher order results are available with the careful choice of prior. But beyond that, the Bayes procedure comes up short, unless the priors become data dependent and the calculations are carefully targetted using an evidence-based formulation.

Thus linear approximations can be hugely useful but they can carry substantial risks. The assertion of probability status is directly contradicted by reality! And no indications seem available that a Bayesian calculation could yield more than just approximate confidence. The promotional assertiveness that accompanies current Bayes development is misleading and misleading to the extent of being fraudulent.

Of course there can be contexts where there is an objective prior $\pi(\theta)$ that records how the true value was generated. The Bayes paradigm can be applied but it is inappropriate; the direct approach is a matter of modeling, of what aspect of the context is appropriate to include. From this viewpoint the indicated methodology predates Bayes; it is just probablity analysis. Even then it allows that the prior information and the evidence based information to be presented separately, thus of course allowing the end user to combine if needed or wanted.

There is no imperative that says the prior and the evidence-based should be combined. It is an option. And it is an option with risks!

REFERENCES

1. Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc.* London **53**, 370-418; **54**, 296-325. Reprinted in *Biometrika* **45** (1958), 293-315.
2. Bernardo, J. M. (1971). Reference posterior distributions for Bayesian inference (with discussion). *J. Roy. Statist. Soc.* B **41**, 113-147.
3. David, A. P., Stone, M. and Zidek, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference. *J. Roy. Statist. Soc.* B **35** 189-233
4. Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans. Royal Soc.* London A 222, 309-368.
5. Fisher, R. A. (1930). Inverse probability. *Proc. Camb. Phil. Soc.*, **26**, 528-535.
6. Fisher, R. A. (1935). The fiducial argument in statistical inference. *Annals of Eugenics*, B, 391-398.
7. Fraser, D.A.S. (2010). Is Bayes posterior just quick and dirty confidence? *Statistical Science* in review.
8. Fraser, D.A.S., and Reid, N. (2002) Strong matching of frequentist and Bayesian parametric inference. *Journal of Statistical Planning and Inference.* **103**, 263-285.
9. Fraser, D. A. S., Reid, N., Wong, A. (2004). Inference for bounded parameters. *Physics Review* D **69** 033002.
10. Fraser, D. A. S., Reid, N., Marras, E., and Yi, G. Y. (2010). Default prior for Bayesian and frequentist inference. *J. Roy. Statist. Soc.* B, to appear.
11. Fraser, D. A. S. and Sun, Y. (2010). Some corrections for Bayes curvature. *Pak. J. Statist.*, to appear.
12. Jeffreys, H. (1939). *Theory of Probability.* Oxford: Oxford University Press. Third Edition.
13. Jeffreys, H. (1946). An invariant form by the prior probabilities in estimation problem. *Proc. Roy. Soc. A.* **186**, 453-461.
14. Laplace, P. S. (1812). Théorie Analytique des Probabilités. Paris: Courcier.
15. Lindley, D.V. (1958). Fiducial distribution and Bayes theorem. *J. Roy. Statist. Soc.* B **20**, 102-107.
16. Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Phil.Trans. Roy. Soc. A***237**, 333-380.
17. Reid, N. and Fraser, D. A. S. (2010). Mean likelihood and higher order inference. *Biometrica*, to appear.
18. Savage, L. J. (1972). *The Foundations of Statistics.* New York: John Wiley.