Statistical Science
2016, Vol. 31, No. 4, 578–590
DOI: 10.1214/16-STS573

Bayes, Reproducibility and the Quest for Truth

D. A. S. Fraser, M. Bédard, A. Wong, Wei Lin and A. M. Fraser

Abstract. We consider the use of default priors in the Bayes methodology for seeking information concerning the true value of a parameter. By default prior, we mean the mathematical prior as initiated by Bayes [*Philos. Trans. R. Soc. Lond.* **53** (1763) 370–418] and pursued by Laplace [*Théorie Analytique des Probabilités* (1812) Courcier], Jeffreys [*Theory of Probability* (1961) Clarendon Press], Bernardo [*J. Roy. Statist. Soc. Ser. B* **41** (1979) 113–147] and many more, and then recently viewed as "potentially dangerous" [*Science* **340** (2013) 1177–1178] and "potentially useful" [*Science* **341** (2013) 1452]. We do not mean, however, the genuine prior [*Science* **340** (2013) 1177–1178] that has an empirical reference and would invoke standard frequency modelling. And we do not mean the subjective or opinion prior that an individual might have and would be viewed as specific to that individual. A mathematical prior has no referenced frequency information, but on occasion is known otherwise to lead to repetition properties called confidence. We investigate the presence of such supportive property, and ask can Bayes give reliability for other than the particular parameter weightings chosen for the conditional calculation. Thus, does the methodology have reproducibility? Or is it a leap of faith.

For sample-space analysis, recent higher-order likelihood methods with regular models show that third-order accuracy is widely available using profile contours [In *Past, Present and Future of Statistical Science* (2014) 237–252 CRC Press].

But for parameter-space analysis, accuracy is widely limited to first order. An exception arises with a scalar full parameter and the use of the scalar Jeffreys [*J. Roy. Statist. Soc. Ser. B* **25** (1963) 318–329]. But for vector full parameter even with a scalar interest parameter, difficulties have long been known [*J. Roy. Statist. Soc. Ser. B* **35** (1973) 189–233] and with parameter curvature, accuracy beyond first order can be unavailable [*Statist. Sci.* **26**

D. A. S. Fraser is Professor Emeritus, Department of Statistical Sciences, University of Toronto, Toronto, Canada (e-mail: dfraser@utstat.toronto.edu). M. Bédard is Associate Professor, Département de mathématiques et de statistique, Université de Montréal, Montréal, Canada (e-mail: bedard@dms.umontreal.ca). A. Wong is Professor, Department of Mathematics and Statistics, York University, Toronto, Canada (e-mail: august@yorku.ca). Wei Lin is Assistant Professor, Department of Statistical Sciences, University of Toronto, Toronto, Canada (e-mail: becky@utstat.toronto.edu). A. M. Fraser is Professor, Department of Mathematics, University of British Columbia, Vancouver, Canada (e-mail: afraser@math.ubc.ca).

(2011) 299–316]. We show, however, that calculations on the parameter space can give full second-order information for a chosen scalar interest parameter; these calculations, however, require a Jeffreys prior that is used fully restricted to the one-dimensional profile for that interest parameter. Such a prior is effectively data-dependent and parameter-dependent and is focally restricted to the one-dimensional contour; these priors fall outside the usual Bayes approach and yet with substantial calculations can still give less than frequency analysis.

We provide simple examples using discrete extensions of Jeffreys prior. These serve as counter-examples to general claims that Bayes can offer accuracy for statistical inference. To obtain this accuracy with Bayes, more effort is required compared to recent likelihood methods, which still remain more accurate. And with vector full parameters, accuracy beyond first order is routinely not available, as a change in parameter curvature causes Bayes and frequentist values to change in opposite direction, yet frequentist has full reproducibility.

An alternative is to view default Bayes as an exploratory technique and then ask does it do as it overtly claims? Is it reproducible as understood in contemporary science? The posterior gives a distribution for an interest parameter and, thereby, a quantile for the interest parameter; an oracle could record whether it was left or right of the true value. If the average split in evaluative repetitions is in accord with the nominal level, then the approach is providing accuracy. And if not, then what is up, other than performance specific to the parameter frequencies in the prior. No one has answers although speculative claims abound.

Key words and phrases: Confidence, curved parameter, exponential model, gamma mean, genuine prior, Jeffreys, L'Aquila, linear parameter, opinion prior, regular model, reproducibility, risks, rotating parameter, two theories, Vioxx, Welch–Peers.

1. INTRODUCTION

1.1 Preview

Reproducibility has recently become prominent in science. What form of reproducibility might be available for Bayes methodology? And what is it? Or is Bayes above such verification of its approach? There are of course genuine priors as clarified by Efron (2013) which admit full frequency modelling; and there are subjective priors that represent an investigator's opinion. But otherwise there are default priors that claim to be objective and are called objective by those who promote them. As such we can reasonably ask what supports the claim of objectivity? Does the use of such methodology have some form of reproducibility as expected in science?

Being aware of conditional probability, Bayes realized that by combining the model for the data variable together with a hypothesized prior distribution for the parameter, he could obtain a joint model for both parameter and variable. This then provides a marginal posterior distribution for the parameter of interest. With this in mind, he then supposed the presence of a random source for his parameter, which led to the widely promoted Bayes approach. Making up a missing input to a theorem can lead to a legitimate concern about the validity of the conclusion from that theorem. Nonetheless, these worries aside, we can still wonder whether the Bayes procedure somehow works, or whether there exists a prior that cancels the effect of the subjectiveness?

Suppose we instigate a default Bayesian calculation with a prior $\pi(\theta)$ on the full parameter and obtain a distribution for the full parameter. Then for a scalar interest parameter $\psi(\theta)$ we can determine the marginal distribution and then invert to obtain say a β-level quantile for the interest parameter. We can certainly ask how that quantile relates to the true value of the parameter. The derivation asserts that if possible parameter values are in accord with the weighting in the prior then accuracy is at the specified β level; but if a different weighting represents the possible θ then the nominal β may be entirely erroneous. As this process is well defined and repeatable, we can certainly simulate and see whether and in what manner there is reproducibility. In the eventuality that the particular weighting in the prior does not work, then the procedure can be subject to potentially serious consequences. This provides meaning to the "potentially dangerous" and "potentially useful" attributes mentioned earlier. In other words, does the procedure do as it says? And it gives background to a standard process for publication retraction.

In some cases, however, we may uncover repetition properties, the reproducibility proposed later by Fisher (1930) and Neyman (1937), yet also implicitly present in Laplace (1812) and next described.

1.2 Reproducibility

Reproducibility is widely acknowledged and affirmed in the sciences; see, for example, the editorial by Marcia McNutt (2014), the former Editor-in-Chief of the prestigious journal Science and now president of the US National Academy of Sciences. She praises the role of reproducibility in science and more broadly the role of statistics in science, and in her role of Editor-in-Chief has recently administered the retraction of articles in Science (McNutt, 2015). And now, for a default Bayesian who asserts probabilities for an unknown parameter, we can reasonably require that reproducibility be verified: that the actual probability should be the asserted probabilities, not just those calculated from some speculative mathematical weighting of possible parameter values. If subjective, then state as subjective.

1.3 Bayes, Statistics and Science

Also in the journal Science, Efron (2013) discusses the role of Bayes theorem in the present century and offers a classification of prior densities: the "*genuine prior*," for those representing an empirical or theoretically based distribution that describes the sourcing of the true value of the parameter in the application; the "*Laplace prior*," for those providing some form of noninformative weight function, such as those of Laplace; and then, by omission, the "*opinion or subjective prior*" as sometimes promoted for applications. He describes the first as "genuine," the Laplace prior as "troublesome" or "potentially dangerous," and the opinion prior, by omission, as perhaps not deserving comment. In response, Fraser (2013) offers the view that the Laplace prior can on occasions provide "a route to approximate confidence." And then, separately, the above mentioned editorial in Science (2014 January 17) praises the role of reproducibility in science and more broadly the role of statistics in science.

1.4 It Is Tough to Make Bayes Reproducible

In this paper, we use large-sample likelihood theory to determine where and in what form the likelihood function provides information concerning a parameter of interest. We then determine how and to what degree that information can be extracted by Bayes-type arguments. As part of this, we find that the Jeffreys–Laplace prior is an essential input but needs to be differentially applied in order to give reproducible information on a parameter of interest. These modified Jeffreys-type priors are usually data dependent and interest parameter dependent, thus falling outside the usual Bayes framework. Although this modified prior is informed by large-sample likelihood methods, the frequency-based higher-order likelihood methods themselves produce parameter information with higher accuracy and lower computational overhead. So what does Bayes contribute other than an exploration option that separately needs its reproducibility verified?

2. BACKGROUND

2.1 The Scalar Location-Model with Flat Prior Gives Reproduciblility

For a location or measurement model $f(y-\theta)$ with observed data y^0, consider a comparison of the frequency approach and the Bayes approach using the flat prior favoured by Laplace. The frequency approach is essentially descriptive: it records in essence the statistical position of the data relative to a possible parameter value θ,

$$p(\theta) = \int_{-\infty}^{y^0} f(y-\theta)\,dy; \tag{2.1}$$

this is just $F(y^0;\theta) = F^0(\theta)$ or the observed distribution function. Meanwhile, the Laplace assessment based on transformation invariance or noninformative scaling uses the flat prior $\pi(\theta) = c$ and gives the nominal posterior survivor value

$$s(\theta) = \int_{\theta}^{\infty} f(y^0 - \theta')\,d\theta' \tag{2.2}$$

for the parameter value θ. These are numerically equal, $p(\theta) = s(\theta)$, as is obvious by elementary calculus, or by seeing one as a reflection of the other, or by looking left from the data or right from the parameter value and seeing the same functional shape. The technical equality says that the Bayes survivor value has merit in producing the lower confidence bound. Clearly, we have here that frequency and Bayes have formal equivalence or that Laplace was just anticipating Fisher but did not quite formulate his proposal in terms of the confidence generalization.

The preceding can be reexpressed in terms of corresponding quantile functions. Let $\hat{\theta}_\beta$ be the solution of $\hat{\beta} = s(\theta)$ for this special location case; then $\hat{\theta}_\beta = s^{-1}(\beta)$ is the β-level lower quantile of the posterior distribution with the frequency property that

$$\mathrm{pr}\{\hat{\theta}_\beta \le \theta; \theta\} = \beta,$$

thus just pure reproducibility. Indeed for say the Normal$(\mu; \sigma_0/n^{1/2})$ in obvious notation we have $s(\mu) = \Phi\{(\bar{y}^0 - \mu)/(\sigma_0/n^{1/2})\}$, $\hat{\mu}_\beta = \bar{y}^0 - z_\beta \sigma_0/n^{1/2}$ where z_β is the usual β-level quantile of the Normal(0, 1) with distribution function $\Phi(z)$, and $\bar{y}$ is the usual sample average. It follows routinely that $\hat{\mu}_\beta$ is the Bayes, the frequency, the confidence, the fiducial lower β-level quantile and has full reproducibility, call it confidence or call it probability or other appropriate term. We now consider Laplace-based Bayes more generally, in relation to reproducibility.

2.2 The Scalar Jeffreys, Where Bayes Gives Approximate Reproducibility

The location property can also arise as an approximation: Jeffreys (1946) recommended the use of an invariant prior, being the square root of the expected information or expected information determinant. For this in some wide generality indicated in Section 3.3, we can begin with a general exponential model $f(y;\theta) = \exp\{\varphi'(\theta)u(y) + k(\theta)\}H(y)$ with p-dimensional u and p-dimensional φ. This can be reexpressed in terms of the essential $u(y)$ and $\varphi(\theta)$ as

$$\begin{aligned} f(u;\varphi) &= \exp\{\varphi' u - \kappa(\varphi)\}h(u) \\ &= \exp\{\ell(\varphi; u)\}h(u), \end{aligned} \tag{2.3}$$

where the log-likelihood $\ell(\varphi; u) = a + \log f(u;\varphi)$ has the usual additive constant; the additive constant can then be replaced by a representative giving the log-likelihood $\log f(u;\varphi) - \log f(u;\hat{\varphi})$ which conveniently has maximum value 0. Let $j_{\varphi\varphi} = -\ell_{\varphi\varphi}(\varphi; u) = \kappa_{\varphi\varphi}(\varphi)$ be the observed information function with subscripts denoting differentiation; it is also the expected information. The standard Jeffreys prior is

$$\pi_J(\varphi) = |j_{\varphi\varphi}(\varphi)|^{1/2} \tag{2.4}$$

which is free of u; it also provides a measure element $\pi_J(\theta)\,d\theta$ that is parameterization invariant.

For the scalar parameter case, the role of the prior is easily seen from a second-order log-density expansion about the observed $(u^0, \hat{\varphi}^0)$ where coordinates have been re-centered at the observed data values and then rescaled with respect to root observed information (Cakmak et al., 1998):

$$\begin{aligned} g(s;\varphi) = (2\pi)^{-1/2} \exp\{&-(s-\varphi)^2/2 \\ &- a(\varphi^3 - s^3)/6n^{1/2}\}\{1 + O(n^{-1})\}. \end{aligned} \tag{2.5}$$

This has observed information $j(\varphi; s) = 1 + a\varphi/n^{1/2}$ and as written has been normalized to the second order. If we integrate the root information adjusted parameter increment, $(1 + a\varphi/n^{1/2})^{1/2}\,d\varphi = d\beta$, we obtain

$$\beta = \int_0^\varphi (1 + a\varphi/2n^{1/2})\,d\varphi = \varphi + a\varphi^2/4n^{1/2},$$

with inverse transformation $\varphi = \beta - a\beta^2/4n^{1/2}$. Calculating $\hat{\varphi}$ and $\hat{\beta}$ and substituting in (2.5) then gives

$$\begin{aligned} (2\pi)^{-1/2} \exp\{&-(\hat{\beta} - \beta)^2/2 \\ &- a(\hat{\beta} - \beta)^3/12n^{1/2}\}\,d\hat{\beta}, \end{aligned} \tag{2.6}$$

which now describes a location model to second-order accuracy. And if we then switch from $d\hat{\beta}$ to $d\beta$ as from Section 2.1 to Section 2.2, we find that the density for β is just the likelihood with the Jeffreys prior. It follows then that quantiles and intervals calculated using the scalar Jeffreys prior have second-order reproducibility; see Section 2.1. This was established by Welch and Peers (1963) using transforms and analysis in the complex plane. For vector parameters, however, Jeffreys (1961) indicated that there were problems with his prior in the regression model context and suggested an alternative; we now examine this problem.

2.3 Vector Laplace and Vector Jeffreys do Not Give Reproducibility

Consider a Normal location model on the plane, say $\phi(y_1 - \theta_1, y_2 - \theta_2)$ where $\phi(z_1, z_2)$ is the bivariate standard Normal; let $(y_1^0, 0)$ be the data and $\psi = \theta_1$ be the interest parameter; the Laplace or Jeffreys prior is the flat prior $\pi(\theta) = c$.

First, consider the linear parameter $\psi = \theta_1$. By the previous subsections, the Bayes posterior survivor value is $s(\psi) = \Phi(y_1^0 - \psi)$. This is in full accord with the usual confidence p-value, and thus has reproducibility.

But now suppose we add curvature to the interest parameter, so $\psi^c = \theta_1 + \gamma\theta_2^2/2$ and have γ positive so that the contours of ψ^c are cupped to the left. Then with increasing γ the p-value *decreases* from that $s(\psi) = \Phi(y_1^0 - \psi)$ under linearity, and the Bayes survivor s-value *increases* from that under linearity (Fraser, 2011). They change in opposite directions from the neutral linearity. Of course, the frequency p-value retains full reproducibility from its construction. It follows then that Bayes or Jeffreys does not have reproducibility. This is a shocking result. And the Bayes approach should not hide the failure. Earlier versions of this phenomenon (Dawid, Stone and Zidek, 1973) were attributed to marginalization, but the present example is more specific and attributes it to marginalization in the presence of a curved interest parameter.

In this paper, we determine where the information concerning an interest parameter is to be found in the likelihood function and in what form. This leads us to determine what sort of prior would extract this information concerning an interest parameter. We then use a simple and familiar model, the gamma model, as a counter-example to Bayes, to illustrate the needed calculations and to see that they can only achieve second-order accuracy, in general. More complex examples are not needed to demonstrate the failure. And, in addition to this mitigated accuracy, the method requires intensive analysis and greater computational overhead than the routine frequency procedures. Of course, the Bayesian calculations lead to nominal probabilities for a parameter and such does have appeal. But the subjective derivation seems in conflict with reproducibility.

2.4 Statistics and Highest Professional Standards

Statistics, at the centre of science and community, deserves the highest professional standards for accuracy, precision and reliability, as appropriate to the context. Of course, there have been huge professional developments in methods for exploration and for discovery, and this is of immense value. But also there has been false discovery, and a need for verifications, along with the potential risks. Can these be serious? And is it more than just having liability insurance? Can things go wrong with statistics centrally involved?

The risks can be serious and the consequences immense. An earthquake at L'Aquila, Italy, on January 5, 2009, caused an estimated 300 deaths. But it had been preceded by many small seismic shocks that alarmed people. A government authority appointed a committee of seismologists with statistical expertise that reported that there was no strong reason for a major quake. The people were reassured and returned to their usual activities but the major quake arrived and a legal court charged the committee members with manslaughter.

The pain killer Vioxx was approved by the US Food and Drug Administration (FDA) in 1999 and then withdrawn by the pharmaceutical company Merck in 2004 after an acknowledged excess of cardiovascular thrombotic (CVT) events with Vioxx, in a placebo controlled study. However, the available evidence for life-threatening risks had long been overwhelming and some 40,000 died as indicated by an FDA estimate; and Merck paid over five billion dollars in penalties and in settlements to benefit the injured and their survivors.

Statistics itself has two theories (Fraser, 2014) that can give contradictory results and each is strongly promoted: this could provide powerful fuel for any legal action concerning disputed results. Should the basics of statistical inference then be decided in a court of law? Or should science with reproducibility, and mathematics with logic directly address the lack of coherence in the discipline of statistics? We start by examining this in the context of a regular model with observed data.

3. HOW MODEL CHARACTERISTICS AFFECT ANALYSIS

3.1 Continuity and Sample Size Effects

Not all statistical models show continuity in how parameters affect the model, and not all are amenable to data-size effects. But models with these properties can reasonably be expected to have analyses that respect these properties; otherwise, they are not incorporating important and relevant information. Recent likelihood methods show that models, in wide generality, can be analyzed at very high accuracy as if they were exponential models, see Section 3.4. And continuity shows that the assessment of components interest parameters of dimension d often $d = 1$ is clearly and uniquely available in an available marginal model; see Section 3.3. This has had substantial effects on the directions of recent inference theory, and striking results for default Bayes analysis.

3.2 Exponential Models

Consider an exponential model (2.3). For any data value u, the likelihood function with arbitrary additive constant can of course be replaced by the representative $\ell(\varphi; u) - \ell(\hat{\varphi}; u)$ where the usual arbitrary constant for likelihood is chosen so the representative log-likelihood has maximum value 0. Meanwhile, the curvature $\hat{j}_{\varphi\varphi}$ at the maximum value gives observed information. These statistical quantities, $\{\ell(\varphi; u) - \ell(\hat{\varphi}; u),$

$\hat{j}_{\varphi\varphi}\}$ at points u make available the highly accurate re-expression of the model (Daniels, 1954):

$$\tilde{f}(u;\varphi) = \frac{e^{k'/n}}{(2\pi)^{p/2}} \cdot \exp\{\ell(\varphi;u) - \ell(\hat{\varphi};u)\}|\hat{j}_{\varphi\varphi}|^{-1/2}. \tag{3.1}$$

This approximation provides impressive third-order accuracy widely unaffected by the renormalization indicated by the constant $e^{k'/n}$. It also has the highly attractive property that at each point u it offers the same likelihood as the initial model; and in addition quite strikingly has the underlying density approximation $|\hat{j}_{\varphi\varphi}|^{-1/2}$, a simple highly accurate Fourier inverse.

3.3 What Continuity Says About Component Parameters

To find a prior to extract information on a component parameter $\psi(\varphi)$, we should want to know where the relevant information is located in an observed likelihood function. For this in wide generality consider an interest parameter $\psi(\varphi)$ of dimension d, initially with a particular interest value ψ_0. When $\psi(\varphi) = \psi_0$, we have of course the approximation (3.1) for u. And from recent likelihood theory, say Fraser, Fraser and Staicu (2010), there is a uniquely determined marginal distribution that is second order free of φ given $\psi(\varphi) = \psi_0$; for this, the needed conditional distribution with complementing parameter say λ and nominal variable t has a p^*-approximation

$$\tilde{h}(t;\lambda) = \frac{e^{k''/n}}{(2\pi)^{(p-d)/2}} \cdot \exp\{\ell(\varphi;u) - \ell(\hat{\varphi}_{\psi_0};u)\}|\jmath_{(\lambda\lambda)}(\hat{\varphi}_{\psi_0})|^{-1/2} \tag{3.2}$$

which uses the nuisance information $|\jmath_{(\lambda\lambda)}(\hat{\varphi}_{\psi_0})| = |\jmath_{\lambda\lambda}(\hat{\varphi}_{\psi_0})||\varphi_{\lambda}(\hat{\varphi}_{\psi_0})|^{-2}$ where the Jacobian φ_{λ} of φ with respect to λ for fixed $\psi = \psi_0$ in effect gives a reexpressed nuisance parameter that is locally scaled, designated as (λ), and is in accord with the full canonical variable u.

Then dividing the joint distribution (3.1) by the conditional distribution (3.2) on the profile contour we obtain the marginal model

$$\tilde{g}(s;\psi_0) = \frac{e^{k/n}}{(2\pi)^{d/2}} \exp\{\ell(\hat{\varphi}_{\psi_0};u) - \ell(\hat{\varphi};u)\}|\hat{j}_{\varphi\varphi}|^{-1/2}|\jmath_{(\lambda\lambda)}(\hat{\varphi}_{\psi_0})|^{1/2} \tag{3.3}$$

$$= \frac{e^{k/n}}{(2\pi)^{d/2}} \exp\{\ell(\hat{\varphi}_{\psi_0};u) - \ell(\hat{\varphi};u)\}|\hat{j}^{\mathrm{P}}_{(\psi\psi)}|^{-1/2} \frac{|\jmath_{(\lambda\lambda)}(\hat{\varphi}_{\psi_0})|^{1/2}}{|\jmath_{(\lambda\lambda)}(\hat{\varphi})|^{1/2}}. \tag{3.4}$$

The interest parameter profile information $\hat{j}^{\mathrm{P}}_{(\psi\psi)}$ uses the interest parameter ψ but in a rescaled form (ψ) that is in accord with the canonical variable u and implied by the two versions (3.3) and (3.4). The preceding is available in Fraser (2016).

The distribution $\tilde{g}(s;\psi_0)$ is defined on the plane $\mathcal{L}^0$ that goes through the data point u^0 and is perpendicular to $\psi(\varphi) = \psi_0$ at the constrained $\hat{\varphi}_{\psi_0}$; the variable s provides d rotated coordinates obtained from u on $\mathcal{L}^0$. At a point u on $\mathcal{L}^0$, the exponent is the profile log-likelihood for $\psi = \psi_0$ and has profile information obtained from $|\hat{j}_{\varphi\varphi}| = |\jmath_{(\lambda\lambda)}(\hat{\varphi})||\hat{j}^{\mathrm{P}}_{(\psi\psi)}|$. The density $\tilde{g}(s;\psi_0)$ gives full third-order information for $\psi = \psi_0$ and has uniqueness given the requirement that the model be continuous in the parameter and the variable.

The preceding distribution for assessing $\psi = \psi_0$ is a marginal distribution of an ancillary under $\psi = \psi_0$, and is unique although the expression for the ancillary variable itself is not unique; the uniqueness derives from respecting the parameter continuity in the initial model (Fraser, Fraser and Staicu, 2010).

3.4 What Continuity Says About Regular Models with Data

More generally consider a regular model $f(y;\theta)$ with continuous parameter and observed y^0. The observed log-likelihood is widely available $\ell(\theta) = \log f(y^0;\theta)$. Also, the coordinate distribution functions are often available and can be inverted to give quantile functions, and then combined to give a vector quantile function say $y(z;\theta)$. The latter can be used for simulations, of course, but also to examine how changes in θ at the observed maximum likelihood value $\hat{\theta}^0$ affect data points near y^0:

$$V = (v_1, \dots, v_p) = \left.\frac{\partial y(z;\theta)}{\partial\theta}\right|_{y^0,\hat{\theta}^0}. \tag{3.5}$$

This shows that a change $d\theta$ at $\hat{\theta}^0$ produces a change $dy = V d\theta$ at the data y^0; or equivalently the change dy corresponds to the related change $d\theta$ at the maximum likelihood value. It follows that there is an ancillary contour through the data of dimension p and the conditional distribution on the contour is the indicated distribution for assessing the parameter θ (Fraser, Fraser and Staicu, 2010, Brazzale, Davison and Reid, 2007); then the gradient of likelihood on the ancillary contour $\varphi(\theta) = d\ell(\theta;y)/dV|_{y^0}$ gives the canonical parameter for the exponential model which is fully equivalent to the given model for third-order inference. We thus have that the exponential model $\{\ell(\theta), \varphi(\theta)\}$ provides full third-order inference for the initial model (Fraser and Reid, 1995, Reid and Fraser, 2010); we call this model the tangent exponential model. It follows that very general regular models can be examined entirely within the framework of the exponential model yet retain third-order accuracy.

4. A SCALAR WELCH-PEERS EXAMPLE FOR BAYES

As a simple example with an extremely small sample size consider the scalar parameter gamma model with density $f(y;\alpha) = \Gamma^{-1}(\alpha)y^{\alpha-1}\exp\{-y\}$ on $(0,\infty)$ plus an observation $y^0 = 0.5$. Exact frequency inference gives the p-value function, $p(\alpha) = F^0(\theta)$, as described after (2.1). A quick and dirty approximation can be obtained from first-order Normal approximations using say the maximum likelihood departure or the signed likelihood root (SLR) departure. And Bayes survivor probability functions $s(\alpha)$ can be obtained from say the Jeffreys (1946) prior discussed in Section 2.2, and from the reference prior (Bernardo, 1979). Both involve targeting the parameter of interest, but achieve the goal differently: the Jeffreys uses the parameterization invariant prior $\pi(\varphi) = |-\ell_{\varphi\varphi}(\varphi;u)|^{1/2}$, while the reference prior aims at maximizing the Kullback–Leibler divergence between prior and posterior. In this simple scalar parameter example, these two priors are the same and given by $\pi(\alpha) = \{d^2\log\Gamma(\alpha)/d\alpha^2\}^{1/2}$, leading to a common posterior distribution, $\pi(\alpha|y) \propto \Gamma^{-1}(\alpha)y^{\alpha}\{d^2\log\Gamma(\alpha)/d\alpha^2\}^{1/2}$.

Figure 1 compares the exact p-value function $p(\alpha)$ (solid line) to popular frequentist evaluations (the maximum-likelihood departure represented by points, and the signed log-likelihood root r depicted by a dash-dotted line). It also features a posterior survivor function obtained with Jeffreys prior (dashed line). The p-value function has been obtained exactly in R, while the posterior survivor values were obtained by running 100,000 iterations of a random walk Metropolis algorithm with a Gaussian proposal distribution having standard deviation of $\sigma = 1.5$.

As expected from the Welch and Peers (1963) result, the Bayes approach with Jeffreys prior features second-order reproducibility.

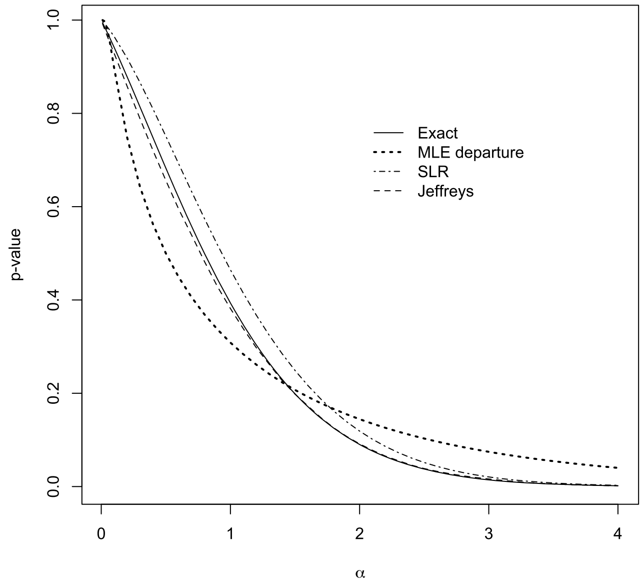

FIG. 1. *Comparison of p-value functions, $p(\alpha)$, and survivor posterior functions, $s(\alpha)$, in terms of α for the scalar parameter distribution $\Gamma(\alpha, 1)$. The exact p-value function is represented by the solid line, the mle departure by points and the SLR approximation by the dash-dotted line. The dashed line represents the survivor posterior function obtained with Jeffreys prior.*

5. VECTOR PARAMETER: REPRODUCIBILITY WITH BAYES

Now consider a regular model $f(u;\psi,\lambda)$ as recorded at (3.1); we seek a prior to extract the information concerning a scalar interest parameter ψ free of λ, and from Section 3 have that this information is fully available on the profile contour for ψ. For this, we have from Section 3 that the model can be expressed as

$$f(u;\varphi) = h(t|s;\lambda,\psi_0)\,g(s;\psi_0), \tag{5.1}$$

with a nuisance density $h(t|s;\lambda,\psi_0)$ at (3.2) and an interest density $g(s;\psi)$ at (3.4) that contains full third-order information on ψ. We determine the prior density that does the extraction from the profile. To eliminate the first factor in (5.1), the prior must have a contribution $|j_{(\lambda\lambda)}(\hat{\varphi}_\psi)|^{1/2}$ to cancel $|j_{(\lambda\lambda)}(\hat{\varphi}_\psi)|^{-1/2}$ and no contribution concerning the exponential factor which this is just 1 on the profile $C^0_\psi = \{\hat{\varphi}^0_\psi\}$. To enable the second factor in (5.1) as displayed at (3.4), we need the Welch–Peers contribution $\{j^{\mathrm{P}}_{(\psi\psi)}(\hat{\varphi}_\psi)\}^{1/2}$ to address the profile information factor $\{\hat{j}^{\mathrm{P}}_{(\psi\psi)}(\hat{\varphi}_\psi)\}^{-1/2}$ to give the needed location form; of course this works with the profile information, and Appendix A.1 shows that the marginalization factor $|j_{(\lambda\lambda)}(\hat{\varphi}_\psi)|^{1/2}/|j_{[\lambda\lambda]}(\hat{\varphi}_\psi)|^{1/2}$ has the needed location form without further help.

Combining these components gives the new prior (5.2), which is the Jeffreys prior $|j_{\varphi\varphi}(\varphi)|^{1/2}$ but now just on the profile contour for ψ. This comes also with an adjustment factor soon seen to involve a measure of interest parameter curvature, and of course with a Jacobian $k(\psi)$ that arises with parameter rotation, as described in Section 6.3 and Appendix A.2:

$$\pi_N(\psi)\, d\varphi_{dir} \tag{5.2}$$
$$= \left|j_{(\lambda\lambda)}(\hat{\varphi}_\psi)\right|^{1/2} \left\{j^{\mathrm{P}}_{(\psi\psi)}(\hat{\varphi}_\psi)\right\}^{1/2} k(\psi)\, d\psi$$

$$= \left|j_{\varphi\varphi}(\hat{\varphi}_\psi)\right|^{1/2} \left\{\frac{|j_{(\lambda\lambda)}(\hat{\varphi}_\psi)|}{|j_{[\lambda\lambda]}(\hat{\varphi}_\psi)|}\right\}^{1/2} k(\psi)\, d\psi. \tag{5.3}$$

Here, $|j_{[\lambda\lambda]}(\hat{\varphi}_\psi)| = |j_{\varphi\varphi}(\hat{\varphi}_\psi)|/j^{\mathrm{P}}_{(\psi\psi)}(\hat{\varphi}_\psi)$ is the nuisance information determinant given the linear parameter χ tangent to ψ at the profile point $\hat{\varphi}_\psi$; this can be obtained by expressing negative log-likelihood in terms of the standardized parameters $(\tilde{\chi}, \tilde{\lambda})$ and differentiating twice with respect to $\tilde{\lambda}$ for fixed $\tilde{\chi}$; see Section 6.3.

This prior is targeted on ψ and is defined on the one-dimensional profile contour C^0_ψ using directed increments in the standardized version of φ; see Section 6.3. In nonlinear cases, it needs a Jacobian $k(\psi)$ to accommodate the parameter change of variable from the directed φ to the interest parameter ψ itself. The curvature adjustment $\{|j_{(\lambda\lambda)}(\hat{\varphi}_\psi)|/|j_{[\lambda\lambda]}(\hat{\varphi}_\psi)|\}^{1/2}$ is evaluated for the observed data and depends on ψ along the profile contour for ψ.

This is a remarkable simplification, essentially back to Jeffreys but used with an indicator function to restrict to the relevant profile contour; in other words, use the historic prior but precisely where the full relevant information is known to be located, on the appropriate profile contour. Of course, there are minor technical details concerning change of variable and rotation of parameter that need attention, but change of variable is reasonably to be expected in any marginalization; see Section A.2. These details do not arise for the linear interest parameter case, first to be examined.

6. EXAMPLES: NEW JEFFREYS WITH REPRODUCIBILITY

6.1 Linear Parameter

Now suppose that $\psi(\varphi) = a'\varphi = \Sigma a_i \varphi_i$ is linear in the canonical parameterization φ. All the sample space contours for assessing ψ are then parallel to the vector a, and thus the line $\mathcal{L}^0$ is given as $u^0 + \mathcal{L}(a)$ which is fixed in direction, that is, does not rotate under ψ_0 change.

6.2 Linear Parameter Example

Let us consider a gamma model with shape α and rate β, both canonical and both unknown, and take α as the parameter of interest and β as a free nuisance parameter. The density model is

$$f(y; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp\{-\beta y\},$$

with observed values say $\mathbf{y}^0 = (1, 4)$; thus, $n = 2$, the minimum number for identifying two parameters. The Fisher information function is

$$\begin{pmatrix} nD''(\alpha) & -n/\beta \\ -n/\beta & n\alpha/\beta^2 \end{pmatrix},$$

where

$$D''(x) = \frac{d^2 \log \Gamma(x)}{d^2 x}. \tag{6.1}$$

is the trigamma function, the second derivative of $\log \Gamma(x)$.

For the p-value function $p(\alpha)$, we use the signed log-likelihood root approach for a simple approximation and the third order as a very accurate approximation. These are then compared to posterior survivor functions, $s(\alpha)$, obtained using three prior distributions: the regular Jeffreys, the reference and the new Jeffreys-style prior.

The regular Jeffreys prior treats both parameters as of equal interest; it is obtained as the root Fisher information determinant $\pi_J(\alpha, \beta) \propto \{\alpha D''(\alpha) - 1\}^{1/2}/\beta$. The reference prior targets the interest parameter α and is expressed as $\pi_R(\alpha, \beta) \propto \{D''(\alpha) - 1/\alpha\}^{1/2}/\beta$; see Yang and Berger (1996), for instance.

The new Jeffreys prior targets the interest parameter α by using the usual Jeffreys prior but fully restricted to the profile contour for the interest α. For a given α, the constrained maximum likelihood estimate for β is $\tilde{\beta}_\alpha = n\alpha/\sum_{i=1}^n y_i$; this leads to the prior

$$\pi_N(\alpha) = \pi_J(\alpha, \tilde{\beta}_\alpha) \propto \{\alpha D''(\alpha) - 1\}^{1/2}/\alpha,$$

but on the profile only; the Jacobian $k(\alpha)$ is of course constant. The posterior distribution is obtained by combining the latter prior with the profile log-likelihood function

$$\ell^{\mathrm{P}}(\alpha|\mathbf{y}) = \alpha \sum_{i=1}^n \log(y_i) - n\alpha - n\log\Gamma(\alpha)$$
$$+ n\alpha \log\alpha - n\alpha \log\left(n \Big/ \sum_{i=1}^n y_i\right),$$

and is given as

$$\pi_N(\alpha|\mathbf{y}) \propto \exp\{\ell^{\mathrm{P}}(\alpha|\mathbf{y})\}\pi_N(\alpha),$$

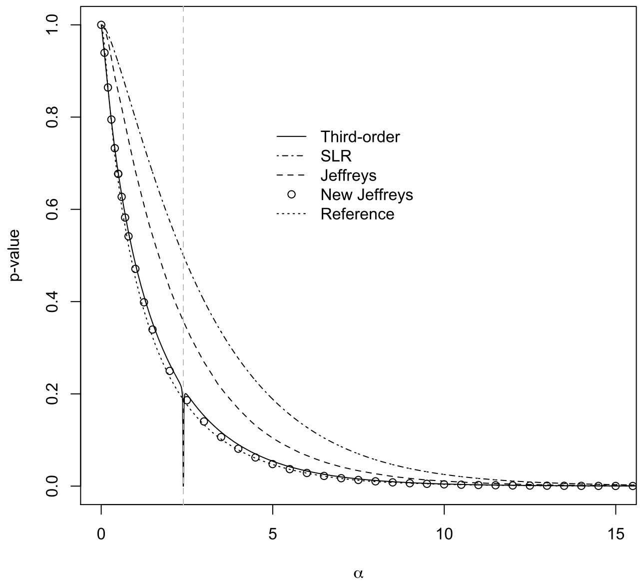

FIG. 2. *Comparison of p-value functions, $p(\alpha)$, and survivor posterior functions, $s(\alpha)$, for the interest α using a $\Gamma(\alpha, \beta)$ model. The third-order p-value function is represented by the solid line and the SLR approximation by the dash-dotted line. Survivor posterior values obtained with Jeffreys, reference and new prior are represented, in order, by dashes, dots and discs. The maximum likelihood value for α is also depicted.*

but calculated strictly on the profile curve for the parameter of interest.

Figure 2 examines the third-order p-value function $p(\alpha)$ (solid line) taken as the exact and the Normal approximation for the signed log-likelihood root r (dash-dotted line). The graph also features a comparison with posterior survivor values obtained with Jeffreys prior (dashed line), the reference prior (dotted line) and the new Jeffreys (discs). Approximations of the p-value function have been obtained in R, while the posterior survivor values were obtained by running 100,000 iterations of a random walk Metropolis algorithm with a Gaussian proposal distribution (also in R). In the current example, the new Jeffreys offers second-order reproducibility, which is not available from the regular Jeffreys. Results from the new Jeffreys prior are as convincing as those based on the present Bayesian benchmark which is the reference prior.

6.3 Rotating Parameter

The line $\mathcal{L}^0$ in some examples can change direction with different ψ_0 values under test. As just noted, this does not happen in the special case with $\psi(\varphi)$ linear in φ, where the sample space contours for various fixed $\psi(\varphi)$ values are all parallel, and thus the corresponding lines $\mathcal{L}^0$ all have the same direction. More generally, however, $\mathcal{L}^0$ can rotate through an angle of order $O(n^{-1/2})$, and thus the model scaling on the line can also change $O(n^{-1/2})$; this arises when $\hat{j}_{\varphi\varphi}$ is not an identity matrix or a constant times such. We refer to such parameters as *rotating*, and this even happens with μ in a Normal$(\mu; \sigma^2)$ analysis. We examine this in this section, and then examine *curved* parameters in the next Section 6.5.

Toward determining effects from a lack of rotational symmetry, let B be a $p \times p$ right square root of the observed information $\hat{j}^0_{\varphi\varphi} = B'B$ and define a new canonical parameter as $\bar{\varphi} = B\varphi$. Then in the new parameterization the observed information $\hat{j}^0_{\bar{\varphi}\bar{\varphi}} = I$ is the identity, and the related information scaling of the distribution under different ψ_0 remains constant. We then also have that the cubic term of order $O(n^{-1/2})$ is constant when examined just to the second order. Thus, the model to that order is fully unaffected by the rotation coming from the direction change of $\mathcal{L}^0$; and thus we have a single underlying reference model for the data, to the given order $O(n^{-1})$. It follows that any Bayes procedure with second-order accuracy must be free of the rotational characteristics of parameters. For some similar considerations, see Fraser (2003).

6.4 Rotating Parameter Example

As a third example, we still consider the gamma model with shape α and rate β, but this time with interest in the mean $\mu = \alpha/\beta$. The density in terms of the parameter of interest μ and nuisance α is thus

$$f(y; \alpha, \mu) = \Gamma^{-1}(\alpha)\left(\frac{\alpha}{\mu}\right)^{\alpha} y^{\alpha-1} \exp\{-\alpha y/\mu\}.$$

We consider a sample of $n = 5$ observations, $\mathbf{y}^0 = (0.20, 0.45, 0.78, 1.28, 2.28)$ as used in Brazzale, Davison and Reid (2007) on page 13. As in Example 2, the third-order and signed log-likelihood root versions of the p-value functions are compared to the Bayesian posterior survivor functions obtained with three different prior distributions.

Jeffreys prior, which is invariant under bivariate parameter transformations, can be obtained from $\pi_J(\alpha, \beta)\, d\alpha\, d\beta$ in Example 2 by change of variable:

$$\pi_J(\alpha, \mu) \propto \frac{1}{\mu}\{\alpha D''(\alpha) - 1\}^{1/2},$$

where $D''(\alpha)$ is as in (6.1).

Finally, the new prior is the full regular Jeffreys prior calculated in the rotationally symmetric ordinates $\bar{\varphi}$ but examined exclusively on the profile curve $C^0_\mu = \{\hat{\varphi}_\mu\}$

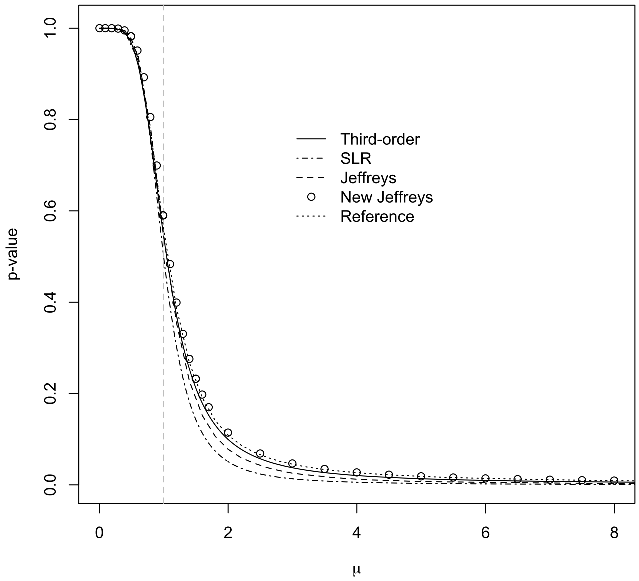

FIG. 3. *Comparison of p-value functions, $p(\mu)$, and survivor posterior functions, $s(\mu)$, in terms of μ for a $\Gamma(\alpha,\mu)$ with interest in the parameter μ. The third-order p-value function is represented by the solid line and the SLR approximation by the dash-dotted line. Survivor posterior values obtained with Jeffreys, reference and new Jeffreys priors are represented, in order, by dashes, dots and discs. The maximum likelihood value for μ is also depicted.*

and with a Jacobian $k(\mu)$ that gives the change-of-variable from $\bar{\varphi}$ to μ as recorded in Appendix A.2:

$$\pi_N(\mu) = \frac{1}{\mu}\{\hat{\alpha}_\mu D''(\hat{\alpha}_\mu) - 1\}^{1/2} k(\mu).$$

As explained in Section 5, the new posterior distribution is then obtained by combining this prior with the profile likelihood function, $L^{\mathrm{P}}(\mu)$ and integrating on the one dimensional profile contour for the parameter μ of interest. For comparison, the reference prior targeting μ is given (Ghosh, 2011) as

$$\pi_R(\alpha,\mu) \propto \frac{1}{\mu}\{D''(\alpha) - 1/\alpha\}^{1/2}.$$

Figure 3 compares the third-order p-value function $p(\mu)$ (solid line) to the signed log-likelihood root r (dash-dotted line). The graph also features a comparison with posterior survivor values obtained with the regular Jeffreys prior (dashed line), the reference prior (dotted line) and the new Jeffreys (discs). Approximations of the p-value function have been obtained in R, while the posterior survivor values were obtained by running 100,000 iterations of random walk Metropolis algorithms with a Gaussian proposal distribution (also in R). Once again, the new Jeffreys offers results that compete with the reference prior and that are much more accurate than those obtained with the regular Jeffreys and of course the SLR.

6.5 Curved Parameter Example

As a very simple example with curvature, we now consider two independent variables $\mathcal{N}(\chi,1)$ and $\mathcal{N}(\lambda,1)$ with observed data say $(0,0)$ and curved interest parameter $\psi = \chi + \frac{1}{2}a\lambda^2$ with fixed curvature a. The log-likelihood function from the pair of observations (y_1, y_2) is

$$\ell(\chi,\lambda) = -\frac{1}{2}\chi^2 - \frac{1}{2}\lambda^2 + \chi y_1 + \lambda y_2;$$

the corresponding maximum likelihood estimate is $\hat{\theta} = (\hat{\chi},\hat{\lambda}) = (y_1, y_2)$.

It is possible to reparameterize from (χ,λ) to $(\psi - \frac{1}{2}a\lambda^2, \lambda)$ and obtain the log-likelihood function in terms of ψ and λ:

$$\ell(\psi,\lambda) = -\frac{1}{2}\left(\psi - \frac{1}{2}a\lambda^2\right)^2 - \frac{1}{2}\lambda^2 + \left(\psi - \frac{1}{2}a\lambda^2\right)y_1 + \lambda y_2,$$

with information matrix

$$j(\psi,\lambda) = \begin{pmatrix} 1 & -a\lambda \\ -a\lambda & ay_1 - a\psi + \frac{3}{2}a^2\lambda^2 + 1 \end{pmatrix}. \tag{6.2}$$

The particularity of this model lies in the curvature of the parameter ψ, and yet the profile log-likelihood for ψ, given the observations $\mathbf{y}^0 = (0,0)$, is just $\ell_{\mathrm{P}}(\psi) = -\frac{1}{2}\psi^2$.

The above can be used to determine the SLR and third-order p-value functions. In the current case, these functions respectively are $\Phi(-\psi)$ and $\Phi(-\psi - a/2)$. Also from the information matrix, it is not difficult to verify that the posterior survivor function under Jeffreys prior is $\Phi(-\psi + a/2)$, as $\psi = \chi$ when the constrained maximum likelihood for χ is 0. The new prior (5.3) simply consists of the usual Jeffreys on the profile contour together with the nuisance information adjustment factor but with $k(\psi) = 1$ thus vanishing; also the root information adjustment factor simplifies to $\exp\{-\operatorname{tr} A\psi/2\}$ which is just $\exp\{-a\psi/2\}$ on the profile line; see Appendix A.3. The resulting posterior density for ψ is then

$$\pi(\psi|\mathbf{y}^0) \propto L_p(\psi)|j_{\lambda\lambda}(\psi,0)|^{1/2} 1 = c\exp\left\{-\frac{1}{2}(\psi^2 + a\psi)\right\},$$

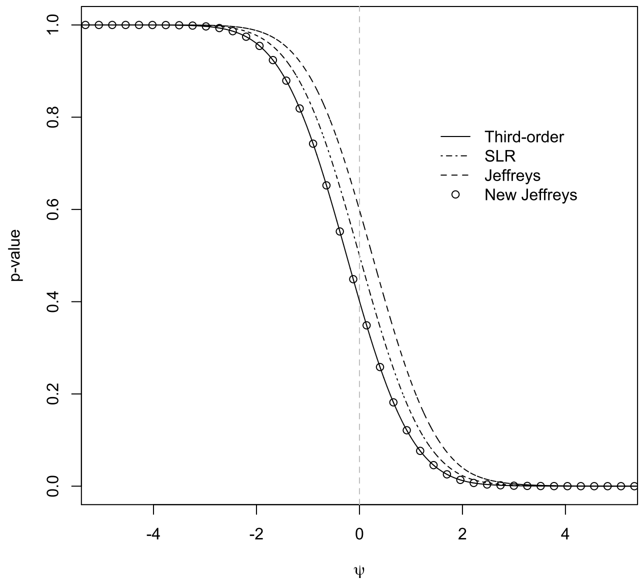

FIG. 4. *Comparison of p-value functions, $p(\psi)$, and posterior survivor functions, $s(\psi)$, in terms of ψ for a bivariate Normal model with interest in the parameter ψ. The third-order p-value function is represented by the solid line and the SLR approximation by the dash-dotted line. Posterior survivor values obtained with Jeffreys and new priors are respectively represented by dashes and circles. The maximum likelihood value for ψ is also depicted.*

which gives a posterior survivor value that is identical to that of the third-order p-value, $\Phi(-\psi - a/2)$.

Figure 4, which is similar to the figures presented in the preceding examples, features a comparison for a curvature parameter $a = 0.5$. From the previous developments, the third-order p-value and posterior survivor function obtained with the new Jeffreys prior can be seen to exactly match. Whether reference priors can accommodate parameter curvature would be of interest.

7. REMARKS

The genuine prior. In his classification of prior densities Section 1.3, Efron (2013) emphasizes genuine priors, priors that describe the sourcing of the true value of the parameter in the application, and thus have a theoretical or empirical basis. The term "genuine" is to indicate that the prior is describing a true objective sourcing, not an exploration or subjective opinion. Some earlier consideration of these priors may be found in Fisher (1956), page 18, and in references therein. In this genuine context, we have two supported models and we have the option of combining them; this is the long-standing frequentist issue of *statistical modelling*.

Recommendation: Record probabilistic information from the sourcing and investigate reliability; separately record information for the model with data; and then as appropriate present results for the combined model. This would be in agreement with scientific practice, and has no Bayes content.

The Laplace prior. Efron (2013) also discusses the mathematical priors proposed by Bayes, and then promoted by Laplace (1812) as uninformative priors. For this, the prior has no objective frequency background but is viewed as a device to explore and nominally use the conditional probability lemma. Efron remarks that during his editorship of an applied statistics journal almost a quarter of the processed manuscripts involved Bayes conditioning and almost all of these then used uninformative Laplace type prior, thus not the genuine prior previously mentioned. The function of a default prior is to check the consequences of the particular weightings in the chosen prior, and the consequences from other weightings are usually not examined. This brings us again directly to reproducibility.

Recommendation: Any use of the Laplace type prior can be viewed as exploratory and subjective, to be assessed by simulations to determine performance, thus reproducibility (Fraser, 2013).

The opinion prior. Opinions and subjective views are sometimes assembled as a subjective prior; see, for example, Savage (1972). There are perhaps good arguments why these are inappropriate in scientific contexts: the user can certainly try his luck at a casino and even explore, but this has no part otherwise in the process for developing valid information and knowledge.

Recommendation: Avoid opinion priors, you could be held legally or otherwise responsible.

Summary. A mathematical prior is of use only if it works, and it thus needs checking for repetition validity: in other words, confidence and reproducibility. Otherwise, the nominal probabilities are subjective and provide nothing without the leap of faith.

APPENDIX

A.1 Scalar Jeffreys and an Adjustment Factor

Consider an exponential model $g(s;\chi) = (2\pi)^{-1/2} \cdot \exp\{\ell(\chi;s) - \ell(\hat{\chi};s)\}\hat{j}_{\chi\chi}^{-1/2}$ to second order, and suppose a model of interest has the form $f(s;\chi) = g(s;\chi)A(s,\chi)$ where the adjustment factor A is constant to first order. For the exponential model alone, the standard Jeffreys prior combined with likelihood from the exponential model gives a survivor probability that is reproducible second order for that exponential model; as part of this it gives a location model say $h(t-\tau)$ as demonstrated at (2.6). Then if that same prior is used with the composite model $f(s;\chi)$ it gives of course the posterior $h(t-\tau)$ as just described together with the factor $A(s,\chi)$; this factor in turn can be expanded as $\exp\{a(t-\tau)/n^{1/2}\}$ in terms of the t and τ. The combination then is a function of $(t-\tau)$, and thus is also a location model and Jeffreys works to second order for the adjusted model $f(s;\chi)=g(s;\chi)A(s,\chi)$.

A.2 Jacobian Concerning Parameter Rotation

Consider an exponential model with canonical parameter φ and a scalar interest parameter ψ. If ψ is linear in φ as discussed briefly in Section 6.1 then the sample space model is defined on a line $\mathcal{L}^0$, and this line from the observed data is fixed in direction under variation in ψ_0. More generally, if $\psi(\varphi)=\psi_0$ is not linear then the line $\mathcal{L}^0$ can change direction under variation in ψ_0. If we then substitute and use a symmetric parameterization $\bar{\varphi}=B\varphi$, we find that the new version of the model in the newly defined variable remains the same to second order on the various lines $\mathcal{L}^0$ from the observed data point. Accordingly, we now consider and analyze in terms of the rotationally symmetric coordinates and have the rewritten model second-order invariant under change in ψ_0.

We then need the connection between the symmetrized coordinates $\bar{\varphi}$ and the ψ parameter as part of the iterative numerical calculation of the posterior distribution. For this, let $\psi_0=\hat{\psi}^0$ be the observed maximum likelihood value, and let d be a suitable small increment for the iterative calculations using $\psi_{i+1}=\psi_i+d$. For each ψ_i, let $\bar{\varphi}_i$ be the constrained maximum likelihood value for $\bar{\varphi}$ given $\psi(\varphi)=\psi_i$, and let $\delta_i=\bar{\varphi}_{i+1}-\bar{\varphi}_i$ be the vector increment in the symmetrized canonical parameter $\bar{\varphi}$. We also need the unit gradient vector $u(\bar{\varphi})$ of ψ with respect to $\bar{\varphi}$ at each point $\bar{\varphi}_i$: for this let $g_i=g(\bar{\varphi}_i)=d\psi/d\bar{\varphi}$ be the gradient vector; then $u_i=g_i/|g_i|$ is the corresponding unit vector and is perpendicular to $\psi(\varphi)=\psi_i$ in the $\bar{\varphi}$ coordinates at $\bar{\varphi}_i$. Let $k_i=\delta_i u_i$. Then k_i gives the Jacobian at $\bar{\varphi}_i$ from the $\bar{\varphi}$ coordinates to the ψ coordinates for the iterative calculations on the profile curve C_ψ.

A.3 Curvature and Information

Consider a surface defined in explicit form as $y=\psi_0-\Sigma a_{ij}x_ix_j/2n^{1/2}$ above a $p-1$ dimensional space, and suppose that interest focuses on properties near $x=0$. The matrix $A=\{a_{ij}\}$ records curvature properties of the surface at $x=0$ and is called the curvature matrix of the surface at $x=0$. The determinant of the curvature matrix is called the Gaussian curvature; and the trace of the curvature matrix is called the mean curvature which will be of particular interest to us. The surface can also be presented in implicit form as $\psi(x)=y+\Sigma a_{ij}x_ix_j/2n^{1/2}=\psi_0$. We are interested in curvature properties of a surface when it is presented in the implicit form, properties that are relevant to the adjustment factors in (3.4) and (5.2).

We use the symmetrized model say $f(u;\varphi)$ that has fixed form relative to the symmetrized coordinates, and let $\ell(\varphi)$ be the corresponding observed log-likelihood function with $\psi(\varphi)$ as the scalar parameter of interest. For a particular value of the parameter, say ψ, we seek an expression for the adjustment factors in (3.4) and (5.2), and relate them to the curvature matrix of the surface $\psi(\varphi)=\psi$ at the constrained maximum likelihood value $\varphi=\hat{\varphi}_\psi$. At $\hat{\varphi}=\varphi(\hat{\psi}^0)$, we let χ be a canonical parameter coordinate that is tangent to $\psi(\varphi)=\psi$ at the point $\hat{\varphi}_\psi$ and let λ be a complementing parameter now taken to be orthogonal to χ at $\hat{\varphi}^0$; accordingly, we take $\varphi=(\psi,\lambda)$ to be the symmetrized canonical parameter, and for convenience assume that these coordinates have been centred at the observed data as well as the symmetrized scaling. The interest parameter ψ can be expanded in terms of φ as

$$\psi=\chi+\Sigma a_{ij}\lambda_i\lambda_j/2n^{1/2} \tag{A.1}$$

with $\chi=\psi-\Sigma a_{ij}\lambda_i\lambda_j/2n^{1/2}$, to the second order. The log-likelihood in terms of φ will be $-\chi^2/2-\Sigma\lambda_i^2/2$ to first order. The above change to ψ will replace the preceding by $-\psi^2/2-\Sigma\lambda_i^2/2$ plus the term $\psi\Sigma a_{ij}\lambda_i\lambda_j/2n^{1/2}$. An element of the nuisance information matrix given χ when changed into an element of the nuisance information given ψ will then acquire an extra term $\psi a_{ij}/n^{1/2}$ and then the ratio $|j_{(\lambda\lambda)}(\hat{\varphi}_{\psi_0})|/|j_{(\lambda\lambda)}(\hat{\varphi})|$ will have the form $(I-\psi A/n^{1/2})$ and then the root determinant ratio becomes $1-\operatorname{tr}A\psi/2n^{1/2}$ to first order where the $n^{1/2}$ is just a formality to keep track of data-size effects.

ACKNOWLEDGEMENTS

We thank the Editor, the Associate Editor and referees for very helpful support and recommendations, We also acknowledge support from the Natural Sciences and Engineering Research Council of Canada, and the Senior Scholars Funding of York University. The material here builds strongly on Fraser and Reid (2002) and Fraser et al. (2010) and joint work with Nancy Reid.

REFERENCES

BAYES, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philos. Trans. R. Soc. Lond.* **53** 370–418.

BERNARDO, J.-M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. Ser. B* **41** 113–147. MR0547240

BRAZZALE, A. R., DAVISON, A. C. and REID, N. (2007). *Applied Asymptotics: Case Studies in Small-Sample Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **23**. Cambridge Univ. Press, Cambridge. MR2342742

CAKMAK, S., FRASER, D. A. S., MCDUNNOUGH, P., REID, N. and YUAN, X. (1998). Likelihood centered asymptotic model exponential and location model versions. *J. Statist. Plann. Inference* **66** 211–222. MR1614476

DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25** 631–650. MR0066602

DAWID, A. P., STONE, M. and ZIDEK, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference. *J. Roy. Statist. Soc. Ser. B* **35** 189–233. MR0365805

EFRON, B. (2013). Bayes' theorem in the 21st century. *Science* **340** 1177–1178. MR3087705

FISHER, R. (1930). Inverse probability. *Proc. Camb. Philos. Soc.* **26** 528–535.

FISHER, R. A. (1956). *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh.

FRASER, D. A. S. (2003). Likelihood for component parameters. *Biometrika* **90** 327–339. MR1986650

FRASER, D. A. S. (2011). Is Bayes posterior just quick and dirty confidence? *Statist. Sci.* **26** 299–316. MR2918001

FRASER, D. A. S. (2013). Bayes' confidence. *Science* **341** 1452.

FRASER, D. A. S. (2014). Why does statistics have two theories? In *Past, Present and Future of Statistical Science* (X. Lin, C. Genest, D. L. Banks, G. Molenberghs, D. W. Scott and J.-L. Wang, eds.) 237–252. CRC Press, Boca Raton, FL.

FRASER, D. A. S. (2016). Definitive testing of an interest parameter: Using parameter continuity. *J. Statist. Res.* **47** 193–169.

FRASER, A. M., FRASER, D. A. S. and STAICU, A.-M. (2010). Second order ancillary: A differential view from continuity. *Bernoulli* **16** 1208–1223. MR2759176

FRASER, D. A. S. and REID, N. (1995). Ancillaries and third order significance. *Util. Math.* **47** 33–53. MR1330888

FRASER, D. A. S. and REID, N. (2002). Strong matching of frequentist and Bayesian parametric inference. *J. Statist. Plann. Inference* **103** 263–285. MR1896996

FRASER, D. A. S., REID, N., MARRAS, E. and YI, G. Y. (2010). Default priors for Bayesian and frequentist inference. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 631–654. MR2758239

GHOSH, M. (2011). Objective priors: An introduction for frequentists. *Statist. Sci.* **26** 187–202. MR2858380

JEFFREYS, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. London, Ser. A.* **186** 453–461. MR0017504

JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Clarendon Press, Oxford. MR0187257

LAPLACE, P. S. (1812). *Théorie Analytique des Probabilités*. Courcier, Paris.

MCNUTT, M. (2014). Reproducibility. *Science* **343** 229.

MCNUTT, M. (2015). Editorial retraction. *Science* **348** 1100.

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.

REID, N. and FRASER, D. A. S. (2010). Mean loglikelihood and higher-order approximations. *Biometrika* **97** 159–170. MR2594424

SAVAGE, L. J. (1972). *The Foundations of Statistics*, revised ed. Dover Publications, New York. MR0348870

WELCH, B. L. and PEERS, H. W. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *J. Roy. Statist. Soc. Ser. B* **25** 318–329. MR0173309

YANG, R. and BERGER, J. O. (1996). A catalog of noninformative priors. Institute of Statistics and Decision Sciences, Duke Univ., Durham, DC.