Submitted to Statistical Science

Bayes, Reproducibility, and the Quest for Truth

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

Abstract. We consider the use of default priors in the Bayes approach for obtaining information concerning the true value of a parameter. As initiated by Bayes (1763) and pursued by Laplace (1812), Jeffreys (1961), Bernardo (1979), and many more, this has recently been viewed as "potentially dangerous" (Efron, 2013), or "potentially useful" (Fraser, 2013). We obtain the existence of a Bayes type prior that does satisfy reproducibility properties as proposed by Fisher (1930), Neyman (1937), and implicitly Laplace (1812). For this, we use large-sample likelihood analysis and find that full information on a typical scalar parameter of interest can be found on a corresponding profile contour of the likelihood function; and then if possible priors are expanded to include discrete components, the full information can be extracted. The required prior however is usually data-dependent and interest-parameter dependent, which falls outside the common Bayes approach. In addition, the use of such priors typically involves substantially more analysis than direct frequency calculations, which in turn have higher accuracy.

We provide simple examples involving extensions of Jeffreys priors. These serve as counter-examples to the general claim that Bayes accomplishes statistical inference. To obtain more accurate results from Bayes, more effort is required compared to recent likelihood methods, while resulting in lesser accuracy. And for vector interest parameters, accuracy beyond first order is routinely not available, as an increase in parameter curvature causes Bayes and frequentist values to change in opposite direction, yet frequentist itself is fully reproducible!

An alternative is to view default Bayes as an exploratory technique and then can ask does it do as it overtly claims? Is it reproducible as understood in contemporary science? Does a β-posterior quantile in repetitions actually have β-coverage of the true? If so, then the approach is providing an approximate route to confidence. And if not, then what is it? No one has an answer although verbal 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 pa-

Département de mathématiques et de statistique, Univ. de Montréal, Montréal, Canada (e-mail: bedard@dms.umontreal.ca); Department of Mathematics and Statistics, York University, Toronto, Canada (e-mail: august@yorku.ca); Department of Statistical Sciences, Univ of Toronto, Toronto, Canada (e-mail: becky@utstat.toronto.edu); Department of Mathematics, The Univ of British Columbia, Vancouver, Canada (e-mail: afraser@math.ubc.ca).

rameter, two theories, Vioxx, Welch-Peers.

1. INTRODUCTION

1.1 Preview

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 would obtain a joint model for both parameter and variable. This procedure then makes available a 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 procedures. Making up a missing input to a theorem however leads to a legitimate concern about the validity of the conclusion arising 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 this subjectiveness?

Suppose we instigate a default Bayesian calculation with a prior $\pi(\theta)$ on the full parameter to obtain the β-level quantile $\hat{\psi}_\beta$ for ψ, the parameter of interest. We can certainly ask whether $pr\{\psi \leq \hat{\psi}_\beta\} = \beta$, in the common usage of probability, also called reproducibility. In other words, does the procedure do as it says? Indeed as the procedure is well defined and repeatable we can simulate and see whether and in what manner it is reproducible. In the eventuality that it is not, the procedure is subject to potentially serious consequences, such as the standard process of 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). This thus provides meaning to the "potentially dangerous" and "potentially useful" attributes discussed earlier.

1.2 Reproducibility.

Reproducibility is widely acknowledged and affirmed in the sciences; see for example, the editorial by Marcia McNutt (2014), the Editor-in-Chief of the prestigious journal Science and nominee as the next 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 had to administer the retraction of articles in Science (McNutt, 2015). And now, for a Bayesian who asserts that $\mathrm{pr}\{\psi \leq \hat{\psi}_\beta\} = \beta$ as mentioned in the Abstract, we can reasonably require that reproducibility apply: Thus the actual probability should be β as in ordinary usage, not the "made-up" usage offered by the improperly invoked conditional probability lemma.

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: "*genuine priors*", for those representing an empirical or theoretically based distribution that describes the sourcing of the true value of the parameter in the application; "*Laplace priors*", for those providing some form of noninformative weight function, such as those of Laplace; and then, by omission, "*opinion or subjective priors*" as sometimes promoted for applications. He describes the first as "genuine", the Laplace priors as "troublesome" or "potentially dangerous", and the opinion priors, by omission, as perhaps not deserving comment. In response Fraser (2013) offers the view that Laplace priors can 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's 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 essential input but needs to be differentially applied in order to give reproducible information on the parameter of interest. These modified Bayes type priors are usually data dependent and interest parameter dependent, thus falling outside the usual Bayes framework. Although this modified Bayes 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 reproducibility verification?

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 equality says that the Bayes survivor value has merit in being just the lower confidence bound. Clearly we have here that frequency and Bayes are equivalent or that Laplace was just anticipating Fisher but didn't quite formulate his proposal for a confidence generalization. We now examine Laplace-based Bayes more generally in relation to reproducibility.

The preceding can be reexpressed in terms of corresponding quantile functions. Let $\hat\theta_\beta$ be the solution of $\beta = s(\theta)$ 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

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

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.

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 §3.3, we begin with a general exponential model with p-dimensional u and p-dimensional φ: $f(y;\theta) = \exp\{\varphi'(\theta)u(y) + k(\theta)\}H(y)$. This can be reexpressed in terms of the essential $u(y)$ and $\varphi(\theta)$ as

$$f(u;\varphi) = \exp\{\varphi' u - \kappa(\varphi)\}h(u) = \exp\{\ell(\varphi;u)\}h(u) \tag{2.3}$$

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

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

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):

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

This has observed information $\jmath(\varphi;s) = 1 + a\varphi/n^{1/2}$ and as written is normed 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

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

which now describes a location model to second order accuracy. And if we then switch from $d\hat{\beta}$ to $d\beta$ as from §2.1 to §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. 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 are now examining 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 gives 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 has full reproducibility from its construction. It follows then that Bayes or Jeffreys does not have reproducibilitiy. This is a shocking result! And the Bayes approach should not hide the failure. Earlier versions of this phenomenon (Dawid et al., 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 use a simple and familiar model, the gamma model, as counter example to Bayes, to illustrate the needed calculations and to see that they can only achieve second order accuracy and that this is at most for scalar interest parameters. More complex examples are not needed to demonstrate this 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 price of that appearance of probability may be too high, when faced with its failures.

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.

And Statistics itself has two theories (Fraser, 2014b) 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 have continuity in how interest parameters affect the model, and not all have clear data-size effects. But those models with these properties can reasonably be expected to have analyses that also respect these properties; otherwise they are not using important and relevant information. Recent likelihood sample-size methods show that models, in wide generality, can be analyzed at very high accuracy as if they were exponential models, see §3.4. Continuity shows that the assessment of components interest parameters of dimension d often $d = 1$ is fully and uniquely available in a specified marginal model; see §3.3. This has had profound 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 additive constant for log-likelihood is chosen so the representative has maximum value 0. Meanwhile the curvature $\hat{\jmath}_{\varphi\varphi}$ at the maximum value gives the observed information. These statistical quantities, $\{\ell(\varphi; u) - \ell(\hat{\varphi}; u), \hat{\jmath}_{\varphi\varphi}\}$ at points u make available the following highly accurate reexpression of the model (Daniels, 1954):

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

This approximation provides impressive third-order accuracy subject to the renormalization implied 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 it has the underlying density approximation $|\hat{\jmath}_{\varphi\varphi}|^{-1/2}$, 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. Also from recent likelihood theory, for example Fraser et al. (2010), we have that there is a uniquely defined marginal variable that is second-order free of φ given $\psi(\varphi) = \psi_0$. The corresponding conditional distribution can have a complementing parameter say λ with variable t. This allows a p^*-approximation

$$\tilde{h}(t;\lambda) = \frac{e^{k/n}}{(2\pi)^{(p-d)/2}} \exp\{\ell(\varphi;u) - \ell(\hat{\varphi}_{\psi_0};u)\}|\jmath_{(\lambda\lambda)}(\hat{\varphi}_{\psi_0})|^{-1/2} \tag{3.2}$$

that 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) we obtain the marginal model

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

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

The interest parameter profile information $\hat{\jmath}^{\mathrm{P}}_{(\psi\psi)}$ uses the interest parameter ψ but in a rescaled form (ψ) that is in accord with the canonical variable u, as implied by the two versions (3.3) and (3.4). The preceding is available in Fraser (2014a).

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 observed profile information such that $|\hat{\jmath}_{\varphi\varphi}| = |\jmath_{(\lambda\lambda)}(\hat{\varphi})||\hat{\jmath}^{\mathrm{P}}_{(\psi\psi)}|$. The density $\tilde{g}(s;\psi_0)$ provides 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 follows from requiring the parameter continuity in the initial model to be respected in the derived model (Fraser et al., 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 the observed value y^0:

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

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 et al., 2010), (Brazzale et al., 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 first example with an extremely small sample size consider a 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, $p(\alpha)$, as described at (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 posterior survivor probability functions $s(\alpha)$ can be obtained from say the Jeffreys (1946) prior discussed in §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 present 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 and a proposal 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.

5. VECTOR PARAMETER: REPRODUCIBILITY WITH BAYES.

Consider a regular statistical model $f(u; \psi, \lambda)$ as recorded at (3.1). We seek a prior to extract information concerning a scalar interest parameter ψ and use

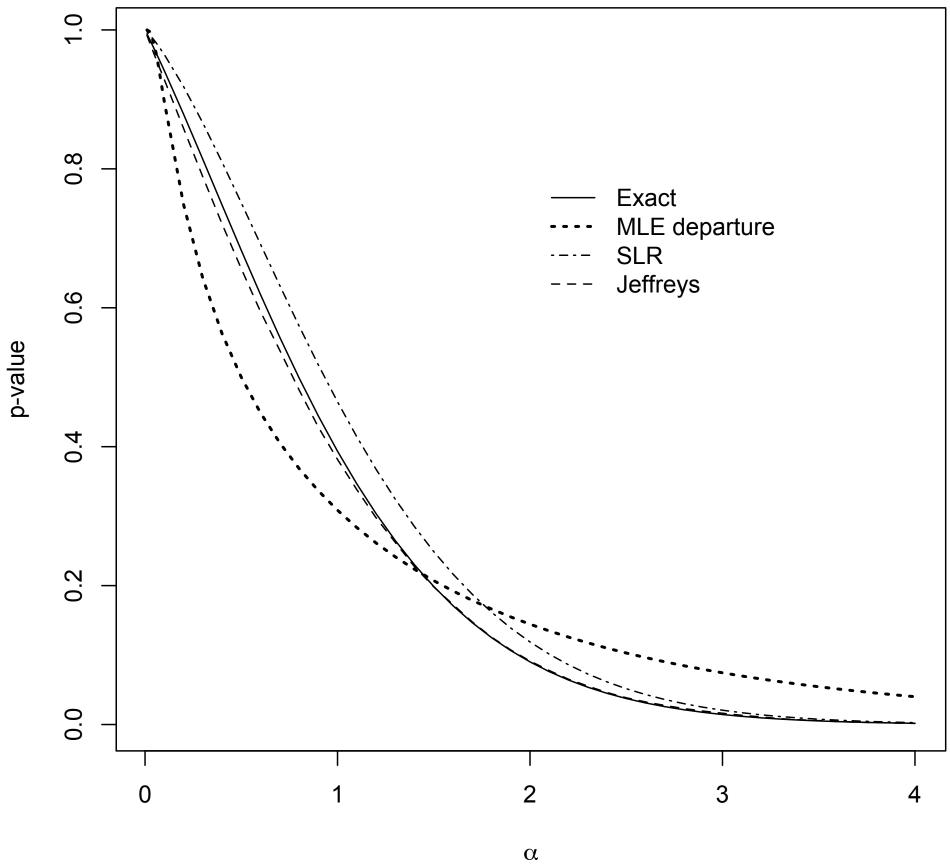

FIG 1. *Comparison of p-value functions, $p(\alpha)$, and survivor posterior functions, $s(\alpha)$, in terms of α for the scalar parameter distribution $\Gamma(\alpha, 1)$. The exact p-value function is represented by the solid line, the mle departure by points, and the SLR approximation by the dash-dotted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

theory from §3 to inform us as to where this information can be found in the observed likelihood function. First we have from §3 that the full model can loosely be expressed as

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

involving a nuisance parameter distribution $h(t|s; \lambda)$ at (3.2) and an interest component $g(s; \psi)$ at (3.4); this focuses on testing $\psi = \psi_0$; further details are given in §3.3.

For the prior to acquire full information on ψ from the likelihood based on (5.1) it needs in part to eliminate any contribution from the first factor in (5.1). As seen from (3.2) the first factor has a constant $|\jmath_{(\lambda\lambda)}(\hat{\varphi}_\psi)|^{-1/2}$ involving ψ which can be removed by a prior that includes the corresponding reciprocal $|\jmath_{(\lambda\lambda)}(\hat{\varphi}_\psi)|^{1/2}$. The first factor also has an exponential, that is equal to 1 if the prior is restricted to the profile contour $C_\psi^0 = \{\hat{\varphi}_\psi^0\}$, the trajectory of the constrained maximum likelihood value under various ψ. Then combining these components concerning λ gives just $|\jmath_{(\lambda\lambda)}(\hat{\varphi}_\psi)|^{1/2}$, but fully restricted of course to the profile contour $C_\psi^0 = \{\hat{\varphi}_\psi^0\}$ for the parameter of interest ψ and clearly avoiding likelihood points that involve the nuisance λ. Alternatively one could use a Jeffreys type prior to eliminate the nuisance parameter and be more conventionally Bayesian but this is unnecessary when the nuisance parameter is already eliminated by the restriction to the profile; and this occurs with no loss of information to third order. We choose not to pursue this alternative here.

For the second factor in (5.1), as displayed at (3.4) concerning ψ, a Welch-Peers prior contribution can address the profile information factor $\{\hat{\jmath}^{\mathrm{P}}_{(\psi\psi)}(\hat{\varphi}_\psi)\}^{-1/2}$ as well as the marginalization factor $|\jmath_{(\lambda\lambda)}(\hat{\varphi}_\psi)|^{1/2}$. In the absence of the marginalization factor the model is exponential and thus subject to Welch-Peers; this requires the prior contribution $\{\jmath^{\mathrm{P}}_{(\psi\psi)}(\hat{\varphi}_\psi)\}^{1/2}$. But there is also the marginalization factor, and from the Appendix §8.1 we have that the Welch-Peers contribution remains second-order reproducible in the presence of such factor.

Combining these components gives the new prior (5.2), which is the Jeffreys prior $|\jmath_{\varphi\varphi}(\varphi)|^{1/2}$ but now just on the profile contour for ψ. This comes 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 §6.3 and Appendix §8.2:

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

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

Here $|\jmath_{[\lambda\lambda]}(\hat{\varphi}_\psi)| = |\jmath_{\varphi\varphi}(\hat{\varphi}_\psi)|/\jmath^{\mathrm{P}}_{(\psi\psi)}(\hat{\varphi}_\psi)$ 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 §6.3.

This prior is targeted on ψ and is defined on the one-dimensional profile contour C_ψ^0 using directed increments in the standardized version of φ; see §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 $\{|\jmath_{(\lambda\lambda)}(\hat{\varphi}_\psi)|/|\jmath_{[\lambda\lambda]}(\hat{\varphi}_\psi)|\}^{1/2}$ 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 just where the relevant information is fully 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 §8.2. These details do not arise for the linear interest parameter case, first to be examined.

6. MODIFIED JEFFREYS GIVES 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 change 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

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

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 computing the usual Jeffreys prior but using it fully restricted to the profile contour for 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

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

it is given as

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

but examined 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

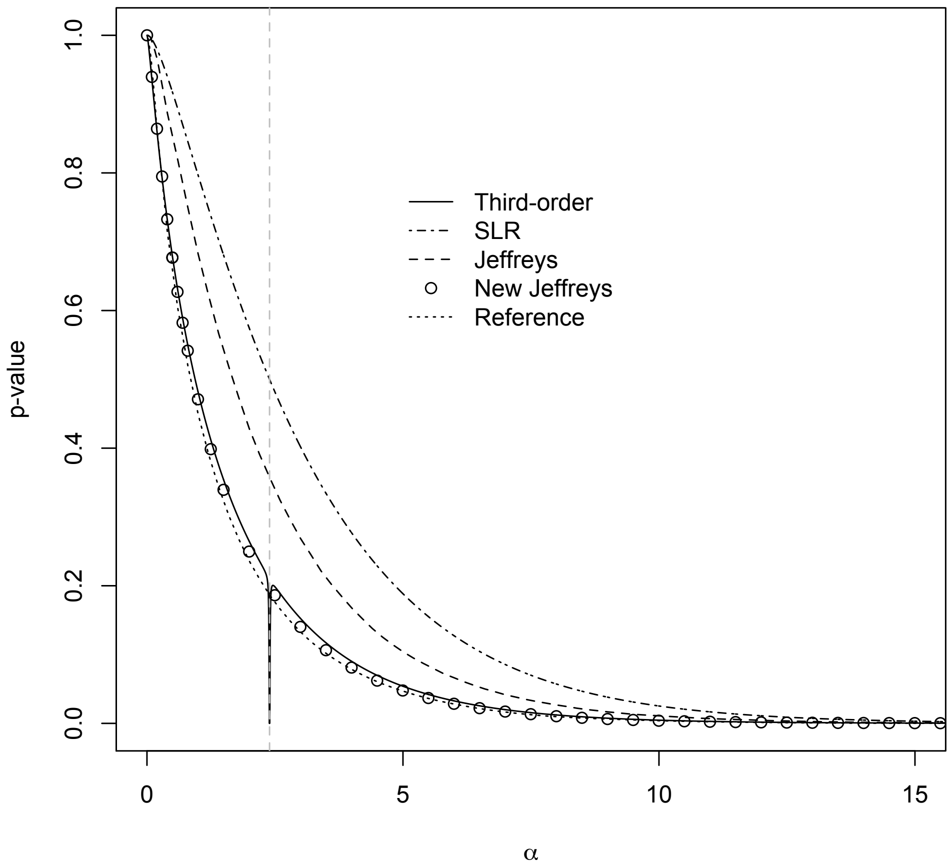

FIG 2. *Comparison of p-value functions, $p(\alpha)$, and survivor posterior functions, $s(\alpha)$, for the interest α using a $\Gamma(\alpha, \beta)$ model. The third-order p-value function is represented by the solid line and the SLR approximation by the dash-dotted line. Survivor posterior values 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.*

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 $\hat{\jmath}_{\varphi\varphi}$ 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.

Towards determing effects from a lack of rotational symmetry, let B be a $p \times p$ right square root of the observed information $\hat{\jmath}^0_{\varphi\varphi} = B'B$ and define a new canonical parameter as $\bar{\varphi} = B\varphi$. Then in the new parameterization the observed information $\hat{\jmath}^0_{\bar{\varphi}\bar{\varphi}} = I$ 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 et al. (2007) at 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 $C^0_\mu = \{\hat{\varphi}_\mu\}$ and with a Jacobian $k(\mu)$ that gives the change-of-variable from $\bar{\varphi}$ to μ as recorded in §8.2:

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

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

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

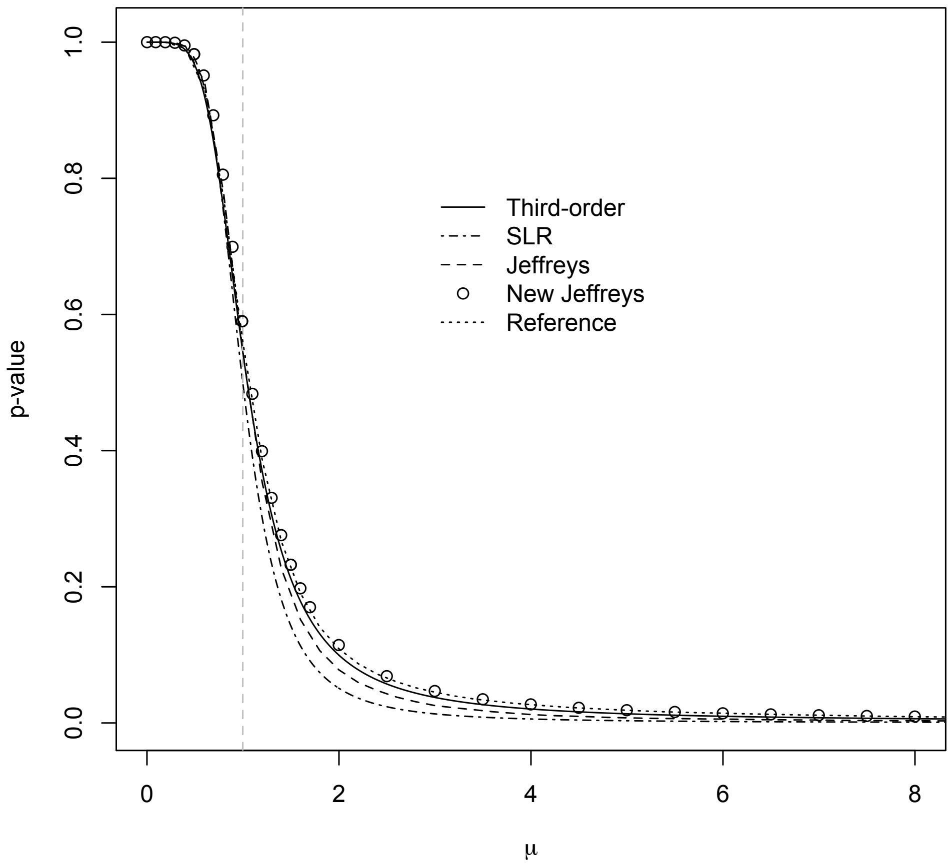

FIG 3. *Comparison of p-value functions, $p(\mu)$, and survivor posterior functions, $s(\mu)$, in terms of μ for a $\Gamma(\alpha, \mu)$ with interest in the parameter μ. The third-order p-value function is represented by the solid line and the SLR approximation by the dash-dotted line. Survivor posterior 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.*

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

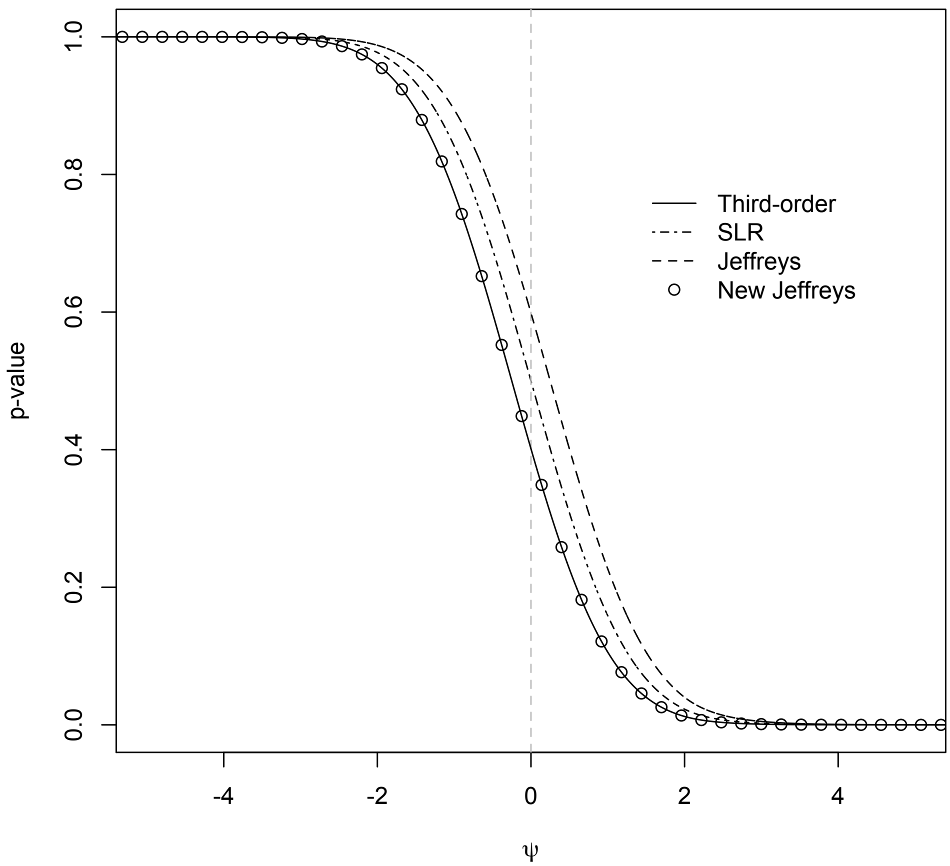

FIG 4. *Comparison of p-value functions, $p(\psi)$, and posterior survivor functions, $s(\psi)$, in terms of ψ for a bivariate Normal model with interest in the parameter ψ. The third-order p-value function is represented by the solid line and the SLR approximation by the dash-dotted line. Posterior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

log-likelihood function in terms of ψ and λ:

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

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 $\exp\{-\mathrm{tr}A\psi/2\}$ which is just $\exp\{-a\psi/2\}$ on the profile line; see §8.3. The resulting posterior density for ψ is then

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

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.

7. CONCLUDING REMARKS

The genuine prior. In his classification of prior densities §1.3, Efron (2013) emphasizes genuine priors, priors that describe the sourcing of the true value of the parameter in the application and have a valid theoretical or empirical basis. The term 'genuine' is to indicate that the prior is real, truly objective, not just labeled 'objective' as has become common practise in recent Bayesian literature for the mathematical priors of Bayes. Some earlier recognition of such priors may be found in Fisher (1956), page 18, and in references therein. In this genuine context for the prior we clearly have two models and we have the option of combining them. There is no Bayes to this; entirely, it is a frequentist issue of **statistical modelling.**

Recommendation: Record probabilistic information obtained from the objective prior; and separately record confidence information from the model with data; and when appropriate record the confidence information from the combined model. This would be in agreement with what is accepted scientific practice. And has no Bayes content other than the routine use of the frequentist conditional probability lemma.

The Laplace prior. Efron (2013) also discusses the mathematical priors proposed by Bayes, and then promoted by Laplace (1812) in a context of uninformative priors. Here the prior has no objective frequency background but is viewed as a device to nominally enable the conditional probability lemma. Efron remarks that during his editorship of an applied statistics journal almost a quarter of the manuscripts processed invoked Bayes conditioning and almost all of these then used the uninformative Laplace type prior, not the genuine priors just mentioned. As we have noted the conditional probability formula does not apply to this mathematical prior context. However frequency properties may be present and these could support confidence.

Recommendation: Any use of the Laplace type prior should be viewed as exploratory and to be assessed by simulations to determine whether the confidence property holds (Fraser, 2013).

The opinion prior. Opinions and subjective views are sometimes assembled as a subjective or opinion prior; see for example , Savage (1953). There are perhaps good arguments why these are inappropriate in scientific contexts; the user can certainly try his luck at a casino, but not as part of the process of developing new knowledge.

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

Summary. Priors for a conditional probability calculation: Certainly the genuine priors; but possibly the mathematical priors, provided their performance has been verified by simulations, thus providing confidence.

8. APPENDIX

8.1 Scalar Jeffreys and an adjustment factor

Consider an exponential model $g(s;\chi) = (2\pi)^{-1/2}\exp\{\ell(\chi;s) - \ell(\hat{\chi};s)\}\hat{\jmath}_{\chi\chi}^{-1/2}$ 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 $f(s;\chi) = g(s;\chi)A(s,\chi)$.

8.2 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 §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 $\bar{\varphi} = B\varphi$ as in §7.1 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.

8.3 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

$$\psi = \chi + \Sigma a_{ij}\lambda_i\lambda_j/2n^{1/2} \tag{8.1}$$

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 $|\jmath_{(\lambda\lambda)}(\hat\varphi_{\psi_0})|/|\jmath_{(\lambda\lambda)}(\hat\varphi)|$ will have the form $(I - \psi A/n^{1/2})$ and then the root determinant ratio becomes $1 - \mathrm{tr}A\psi/2n^{1/2}$ to first order where the $n^{1/2}$ is just a formality to keep track of data-size effects.

ACKNOWLEDGEMENTS

We thank the Editor and referees for very helpful 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 other joint work with Nancy Reid.

REFERENCES

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. R. Statist. Soc. B 41*, 113–147.

Brazzale, A. R., A. C. Davison, and N. Reid (2007). *Applied Asymptotics Case Studies in Small-Sample Statistics.* Cambridge, UK: Cambridge University Press.

Cakmak, S., D. A. S. Fraser, P. McDunnough, N. Reid, and X. Yuan (1998). Likelihood centered asymptotic model: exponential and location model versions. *J. Statist. Planning and Inference 66*, 211–222.

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Annals Math. Statist. 46*, 21–31.

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

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

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

Fraser, A. M., D. A. S. Fraser, and A. M. Staicu (2010). Second order ancillary: A differential view with continuity. *Bernoulli 16*, 1208–1223.

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

Fraser, D. A. S. (2011). Is Bayes posterior just quick and dirty confidence? (with discussion). *Statistical Science 26*, 299–316.

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

Fraser, D. A. S. (2014a). Definitive testing of an interest parameter: using parameter continuity. *Jour. Statist. Res. 47*, to appear.

Fraser, D. A. S. (2014b). Why does statistics have two theories? In X. Lin, C. Genest, D. L. Banks, G. Molenberghs, D. W. Scott, and J.-L. Wang (Eds.), *Past, Present and Future of Statistical Science*, pp. 237–252. Florida: CRC Press.

Fraser, D. A. S. and N. Reid (1995). Ancillaries and third order significance. *Utilitas Mathematica 47*, 33 – 53.

Fraser, D. A. S. and N. Reid (2002). Strong matching of frequentist and Bayesian inference. *J. Statist. Plan. Infer. 103*, 263 – 285.

Fraser, D. A. S., N. Reid, E. Marras, and G. Yi (2010). Default priors for Bayesian and frequentist inference. *J. Royal Statist. Soc., Ser B 75*, 631–654.

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

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

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

Laplace, P. S. d. (1812). *Théorie Analytique des Probabilités.* Paris: Courcier.

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

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

Reid, N. and D. A. S. Fraser (2010). Mean likelihood and higher order approximations. *Biometrika 97*, 159–170.

Savage, L. J. (1953). *The Foundations of Statistics.* New York: Dover.

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

Yang, R. and J. O. Berger (1996). *A catalog of noninformative priors.* Institute of Statistics and Decision Sciences, Duke University.