Bayesian Analysis (2025) **TBA**, Number TBA, pp. 1–28

A Differential Geometric Approach to Bayesian Marginalization

D. A. S. Fraser* and Mylène Bédard†

Abstract. The statistical tool box offers a wide range of inference techniques to choose from; regrettably, the reliability of these methods in the sense of 'reproducibility of frequency properties' can often be unclear or even ignored. We examine this issue for default Bayes methods and develop a prior that leads to full second-order inference for any regular scalar parameter of interest in presence of nuisance parameters; the new prior is Jeffreys based. Also, in parallel, we show that such second-order accuracy is widely unavailable for vector parameters of interest through the Bayesian framework, unless the interest parameter has a special linearity. Detailed examples, including simulations, are presented and discussed.

Keywords: coverage, exponential model, Jeffreys, mathematical prior, reproducibility.

MSC2020 subject classifications: Primary 62F15; secondary 62E20.

1 Introduction

As part of discussing Lindley (1975)'s view of a Bayesian 21st century, Efron (2013) proposed a simple classification for Bayes priors: 'genuine priors' where there is a valid frequency source for the parameter value, and 'uninformative or mathematical priors' for formal calculations without such a valid source. The conditional probability lemma of course says that when parameter values are actually sourced from a specified prior and the model itself is valid, then the posterior distribution exhibits the stated frequency property for the parameter. In cases where other priors describe the sourcing however, then the lemma does not say much in terms of frequency properties for the particular context. This is not to say that such posteriors might not have attractive properties; for example, if a quantile bound obtained from some posterior distribution exhibits the reproducibility-of-frequency-under-repetitions property, then that frequency is *de facto* confidence for the quantile bound.

Berger (2006) and Goldstein (2011) refer to the unification of Bayesian and frequentist procedures in terms of coverage matching. This property arises when a Bayes calculation is examined under repetitions and found to exhibit the stated posterior value. Such a concept is of course confidence with an alternative labelling, and indicates that a Bayes calculation can indeed provide a route to confidence. Over the years, various researchers have made progress on achieving reproducibility through Bayes. Tibshirani (1989), for instance, proposed a coverage-matching prior for the case where a

*Department of Statistical Science, University of Toronto, dfraser@utstat.toronto.edu

†Département de mathématiques et de statistique, Université de Montréal, mylene.bedard@umontreal.ca

 https://doi.org/10.1214/24-BA1503

scalar parameter of interest is orthogonal to the complementing nuisance parameter. More recently, Fraser et al. (2010) developed a prior based on matching Bayesian and frequentist higher-order approximations; we refer the reader to that article and the references therein for more information on the subject. Datta and Sweeting (2005) also provides a comprehensive review of available coverage-matching priors.

In this paper, we develop a new prior that leads to second-order accuracy in terms of frequency reproducibility for a scalar parameter of interest. The construction of this new prior, which was foreseen in the explorations of Fraser et al. (2016b), builds on regular models and likelihood asymptotics. Its development requires useful properties stemming from exponential models in their canonical form, but this does not jeopardize the wide applicability of the new prior: it can indeed be obtained from general regular models, not only exponential ones.

We record, in Section 2 and Section 3, the distributional results that are necessary for the development of the new prior. We then introduce the prior in Section 4; it is, in fact, Jeffreys (1946)'s prior used 'off-label', strictly on the one-dimensional profile contour for the interest parameter. In the case where the interest parameter is non-linear with respect to the canonical parameter, then a rotationally symmetric reparameterization of the model is used to express the required Jacobian. We present a clear and systematic approach for computing the prior, waving the need for potentially restrictive nuisance correction terms that are contingent on the geometry of the interest parameter with respect to the canonical parameterization, see Fraser et al. (2016b). The resulting one-dimensional posterior, which may be seen as emerging from the new Jeffreys-based prior combined to the profile likelihood of the interest parameter, can then be used for second-order Bayesian inference. Following this line of reasoning, further Bayesian calculations (such as predictive distributions) are then accessible through the one-dimensional statistical model that is proportional to the profile likelihood of interest (instead of the initial full-dimensional statistical model).

In parallel, we also show through a revealing example in Section 5 that posteriors for vector parameters quite generally do not have such confidence accuracy, particularly when marginalized to component parameters. In fact, priors featuring second-order accuracy are widely unavailable for vector parameters of interest, unless the parameter has a special linearity. Finally, in Section 6, we examine a spectrum of examples in detail, and find that the new prior gives remarkable accuracy for posterior quantile bounds and intervals.

2 Some results on Bayes-frequency equivalence

Let $Y_1, \ldots, Y_n$ be independent and identically distributed (i.i.d.) random variables from a statistical model with density $f(y_i; \theta)$ on $\mathbb{R}$, where θ has dimension p. We write $y = (y_1, \ldots, y_n)$ with observed data y^0, and assume that $f(y; \theta)$ is an exponential family model with continuity in the variable and parameter. Although this assumption might appear restrictive, we note that the methodology introduced in this paper is widely applicable. Indeed, for general regular statistical models (not in exponential form), we

may rely on a tangent exponential model that provides full third-order inference for the original model-data combination; see Section 8.

2.1 Bayes-frequency equivalence, location models

Under this framework, let s and φ be the canonical variable and parameter for the exponential family model, respectively. Specifically,

$$f(y;\theta) = \exp\{\varphi(\theta)^\top s(y) - \kappa[\varphi(\theta)]\}\mathtt{f}(y),$$

where $\varphi(\theta) \in \mathbb{R}^p$ is one-to-one-equivalent to θ and $s(y) \in \mathbb{R}^p$. The density in its canonical parameterization is therefore expressed as $g(s;\varphi) = \exp\{\varphi^\top s - \kappa(\varphi)\}\mathtt{g}(s)$.

Our goal is to find a prior density such that frequentist p-values match Bayesian posterior survivor values to some degree of accuracy. The exact p-value for φ_0 is defined here as the probability left to the observed data point s^0, $p(\varphi_0) = \int^{s^0} g(s;\varphi_0)\mathrm{d}s$, while the Bayesian posterior survivor value is the posterior probability right to φ_0, $\mathit{ɔ}(\varphi_0) = \int_{\varphi_0} \pi(\varphi|s^0)\mathrm{d}\varphi$.

In the particular case of location models, by seeing one region of integration as the reflection of the other, we directly have

$$p(\varphi_0) = \int^{s^0} g(s-\varphi_0)\mathrm{d}s = \int^{s^0-\varphi_0} g(z)\mathrm{d}z = \int_{\varphi_0} g(s^0-\varphi)\mathrm{d}\varphi = \mathit{ɔ}(\varphi_0).$$

This means that the p-value exactly matches the Bayesian posterior survivor value $\mathit{ɔ}(\varphi_0) = \int_{\varphi_0} \pi(\varphi|s^0)\mathrm{d}\varphi$, where the posterior density is obtained using the flat prior for location parameters, $\pi(\varphi) \propto 1$. Before introducing Welch and Peers (1963)'s result for scalar exponential models, we present a convenient approximation to the statistical density $g(s;\varphi)$.

2.2 Saddlepoint approximation

In recent years, many of the most productive developments for statistical analysis have come from the saddlepoint approximation promoted by Daniels (1954). This method offers an $\mathcal{O}(n^{-3/2})$ estimate of an exponential family density, hereafter referred to as third-order accurate. This highly accurate approximation presents itself in terms of very familiar statistical quantities:

$$\begin{aligned} g(s;\varphi) &= \exp\{\varphi^\top s - \kappa(\varphi)\}\ \mathtt{g}(s) \\ &= \frac{\mathrm{e}^{k/n}}{(2\pi)^{p/2}} \exp\{\ell(\varphi;s) - \ell(\hat\varphi;s)\}\ |\hat\jmath_{\varphi\varphi}|^{-1/2}\ \{1+\mathcal{O}(n^{-3/2})\}, \end{aligned} \tag{1}$$

where $\ell(\varphi;s) - \ell(\hat\varphi;s) = -r_\varphi^2/2$ is the negative log-likelihood ratio, $\hat\varphi = \hat\varphi(s)$ is the maximum likelihood estimator (MLE), and $\hat\jmath_{\varphi\varphi} = \jmath_{\varphi\varphi}(\hat\varphi)$ is the observed information matrix in the canonical parameterization. Each of these involves dependence on the

variable s, but only the first also has dependence on φ. The term k/n is a generic normalizing constant.

Hereafter, we let $\tilde{g}(s;\varphi_0) = \mathrm{e}^{k/n}(2\pi)^{-p/2}\mathrm{e}^{-r^2_{\varphi_0}/2}|\hat{\jmath}_{\varphi\varphi}|^{-1/2}$, keeping in mind that this expression is an $\mathcal{O}(n^{-3/2})$ approximation to the real statistical density; see (1). The high accuracy of the approximate density $\tilde{g}(s;\varphi_0)$ is important, but pales in contrast to its ability to extract measures of adequacy between data and parameter. It essentially replaces any use of sufficiency, ancillarity, and other reduction methods, yet retaining continuity. The approximation $\tilde{g}(s;\varphi_0)$ sort of gives the null distribution for computing the p-value in a single, unequivocal step.

2.3 Bayes-frequency equivalence, scalar case

Suppose we have a scalar-variable, scalar-parameter model in exponential form. A result from Cakmak et al. (1998) shows that $g(s;\varphi)$ can be rewritten as a location model, say $\bar{g}(t-\mu)$ with $t = t(s)$ and $\mu = \mu(\varphi)$, to second-order accuracy:

$$g(s;\varphi) = \bar{g}(t(s) - \mu(\varphi))\{1 + \mathcal{O}(n^{-1})\}.$$

This is achieved by making use of Taylor expansions and transformations on the variable and parameter spaces (see also §2.2 of Fraser et al., 2016b, for instance). Using this convenient approximate location property, it is then easy to show that the observed p-value function is equal, to second-order accuracy, to the Bayesian posterior survivor function under Jeffreys (1946) prior. This intriguing result was established by Welch and Peers (1963).

In particular, let $\psi = \psi(\varphi)$ be the interest parameterization. Using the saddlepoint approximation in (1) and applying a transformation from s to $\hat{\varphi}$, which is one-to-one with Jacobian $|\jmath_{\varphi\varphi}(\hat{\varphi})|$, leads to

$$\begin{aligned} p(\psi_0) &\approx \int^{s^0} \frac{\mathrm{e}^{k/n}}{(2\pi)^{1/2}}\mathrm{e}^{-r^2_{\psi_0}/2}\jmath^{-1/2}_{\varphi\varphi}(\hat{\varphi})\mathrm{d}s \\ &= \int^{\hat{\varphi}(s^0)} \frac{\mathrm{e}^{k/n}}{(2\pi)^{1/2}}\mathrm{e}^{-r^2_{\psi_0}/2}\jmath^{1/2}_{\varphi\varphi}(\hat{\varphi})\mathrm{d}\hat{\varphi}. \end{aligned} \tag{2}$$

We note that this saddlepoint approximation is itself an exponential family model. It can be expanded, as in Cakmak et al. (1998), and eventually approximated by a location model as follows

$$\begin{aligned} p(\psi_0) &\approx \int^{\hat{\varphi}(s^0)} \underbrace{\frac{\mathrm{e}^{k/n}}{(2\pi)^{1/2}}\mathrm{e}^{-r^2_{\psi_0}/2}}_{\bar{g}(t-\mu_0)\{1+\mathcal{O}(n^{-1})\}} \underbrace{\jmath^{1/2}_{\varphi\varphi}(\hat{\varphi})\mathrm{d}\hat{\varphi}}_{\mathrm{d}t} \;\approx\; \int^{t^0} \bar{g}(t-\mu_0)\mathrm{d}t \\ &= \int_{\mu_0} \bar{g}(t^0-\mu)\mathrm{d}\mu \;\approx\; \int_{\varphi_0} \underbrace{\frac{\mathrm{e}^{k/n}}{(2\pi)^{1/2}}\mathrm{e}^{-r^2_{\psi}/2}}_{\bar{g}(t^0-\mu)\{1+\mathcal{O}(n^{-1})\}} \underbrace{\jmath^{1/2}_{\varphi\varphi}(\varphi)\mathrm{d}\varphi}_{\mathrm{d}\mu}, \end{aligned} \tag{3}$$

where (3) is the Bayes survivor function $\jmath(\varphi_0) = \int_{\varphi_0} \pi(\varphi|s^0)\mathrm{d}\varphi$ as based on Jeffreys' prior $\pi(\varphi) \propto j_{\varphi\varphi}^{1/2}(\varphi)$. Both integrals are expressed in terms of $r_\psi^2/2$, the constrained signed log-likelihood ratio $\ell(\hat{\varphi}; s) - \ell(\hat{\varphi}_\psi; s)$, with $\hat{\varphi}_\psi$ the constrained maximum likelihood estimator given the interest parameterization ψ. On the first line, the interest ψ_0 is fixed and $r_{\psi_0}^2/2 = \ell(\hat{\varphi}; s) - \ell(\hat{\varphi}_{\psi_0}; s)$, whereas on the second line, the data is at its observed value and $r_\psi^2/2 = \ell(\hat{\varphi}; s^0) - \ell(\hat{\varphi}_\psi; s^0)$.

The p-value essentially records the statistical position of the data relative to a parameter value φ_0 such that $\psi(\varphi_0) = \psi_0$. In the current context, (3) means that $p(\psi_0) = \jmath(\psi_0)\{1 + \mathcal{O}(n^{-1})\}$ and we can say that the root information (Jeffreys) prior gives Bayes-frequency equivalence.

3 Some results on likelihood asymptotics

Our goal is now to find a prior that leads to Bayes-frequency equivalence for a scalar parameter of interest in presence of nuisance parameters. For this, we use the concept of ancillarity to reparameterize the statistical model $g(s; \varphi)$.

3.1 Ancillary statistic and reparameterization, vector case

Suppose we have a p-dimensional exponential model and are interested in a scalar parameter $\psi = \psi(\varphi)$; let $\lambda = \lambda(\varphi)$ be a $(p-1)$-dimensional complementing nuisance parameter for ψ. We assume here that the nuisance λ is chosen orthogonal to the scalar interest ψ, in the sense of Cox and Reid (1987) (since ψ is a scalar, one can always reparameterize to obtain a nuisance parameter orthogonal to the interest parameter). Conveniently, this requirement will eventually be swallowed up in the theoretical developments of Section 4 and does not need to be accounted for in the examples or, more generally, in practice with general regular statistical models. Indeed, the orthogonality property is used below to obtain the marginal density of an ancillary variable that depends on the interest parameter ψ only. Once this is achieved, we reparameterize this one-dimensional density in terms of the canonical φ to pursue our developments. To this end, we use arguments from differential geometry to derive a Jacobian for this transformation from orthogonal parameters to canonical ones. The new prior we eventually obtain for ψ is one-dimensional and free of the nuisance parameter λ.

As it turns out, the unique null distribution for assessing a specific ψ value is directly available from asymptotic theory; we now provide the broad lines of the argument, see Fraser and Reid (1995) or Fraser (2016) for more details. With $\psi(\varphi)$ fixed at ψ_0, there exists an approximately ancillary statistic U for the nuisance parameter, i.e. a function of s whose distribution is second-order free of λ. This (scalar) statistic $U(s)$ takes values on a continuous contour in the sample space; $U(s)$ and the contour on which it is defined may not be unique, but its density is unique to third-order accuracy. We thus have

$$H(u; \psi_0, \lambda) = h(u; \psi_0)\{1 + \mathcal{O}(n^{-3/2})\}. \tag{4}$$

For convenience, let the continuous contour be the observed profile line $L_{\psi_0}^0 = \{s \in \mathbb{R}^p : \hat{\lambda}_{\psi_0} = \hat{\lambda}^0\}$ on which the constrained maximum likelihood estimator of λ is fixed at its

observed value. The statistic $U(s)$ thus takes values on $L^0_{\psi_0}$, where $\psi = \psi_0$ is fixed. We do not need to characterize $U(s)$, but an explicit expression for $h(u;\psi)$ is required.

With this in mind, we reparameterize the exponential model $g(s;\varphi)$ so that the couple $(u(s), v(s))$ acts as the full canonical variable, where $v(s)$ is some accompanying $(p-1)$-dimensional variable. We may face one of two cases. In the first one, the interest ψ is a linear function of φ. An appropriate change of variable then easily leads to an exponential model with canonical parameter $(\psi(\varphi), \lambda(\varphi))$ and canonical variable $(u(s), v(s))$; from there, a saddlepoint approximation $\tilde{g}(u, v; \psi, \lambda)$ as in (1) is then accessible.

In the second case, the interest ψ is not linear in φ; in lieu of ψ, we then define a new scalar parameter $\chi = \chi(\varphi)$ that is linear in φ and tangent to $\psi(\varphi)$ at $\hat{\varphi}_{\psi_0}$, the constrained maximum likelihood value of φ given $\psi(\varphi) = \psi_0$. Through a change of variable, we then obtain an exponential model with canonical variable $(u(s), v(s))$ and canonical parameter $(\chi(\varphi), \zeta(\varphi))$ (instead of $\psi(\varphi)$ and $\lambda(\varphi)$), where ζ is an accompanying $(p-1)$-dimensional parameter. Since the observed profile line is the contour of interest to us, we need not worry about a potential loss of accuracy from this manipulation. Indeed, we note that on $L^0_{\psi_0}$, the original exponential model $g(s;\varphi)$ exactly coincides with this tangent exponential model. Now, since a saddlepoint approximation $\tilde{g}(u, v; \chi, \zeta)$ is available for the latter, we then pursue the analysis with the approximation $\tilde{g}(u, v; \chi, \zeta)$. For simplicity however, we hold on to our usual notation ψ for the interest parameter.

3.2 Density of the ancillary statistic

Now that the exponential model has been reparameterized in terms of (u, v), we wish to obtain an expression for the marginal density $h(u;\psi)$. With $\psi(\varphi)$ fixed at ψ_0, the full density $g(u, v; \psi_0, \lambda)$ is factorized as

$$\begin{aligned} g(u, v; \psi_0, \lambda) &= q(v|u; \psi_0, \lambda)\ H(u; \psi_0, \lambda) \\ &= q(v|u; \psi_0, \lambda)\ h(u; \psi_0)\{1 + \mathcal{O}(n^{-3/2})\}, \end{aligned}$$

with a nuisance density $q(v|u; \psi_0, \lambda)$ and an interest density $h(u; \psi_0)$ that contains full third-order information about ψ_0; see (4). To obtain an expression for $h(u; \psi_0)$ on the ancillary contour $L^0_{\psi_0}$, we find expressions for the full model g and nuisance density q, both restricted to $L^0_{\psi_0}$, and then solve for h in the previous equation.

Remember that (u, v) now acts as the full canonical variable of the exponential model and consider a saddlepoint approximation as in (1):

$$\exp\{\psi u + \lambda v - \kappa(\psi, \lambda)\}\ \mathsf{g}(u, v) = \tilde{g}(u, v; \psi, \lambda)\ \{1 + \mathcal{O}(n^{-3/2})\},$$

where, using the orthogonality of ψ and λ to factorize the determinants,

$$\begin{aligned} &\tilde{g}(u, v; \psi, \lambda) \\ &\quad = \frac{\mathrm{e}^{k/n}}{(2\pi)^{p/2}} \exp\{\ell(\psi, \lambda; u, v) - \ell(\hat{\psi}, \hat{\lambda}; u, v)\}\ |\jmath_{\psi\psi}(\hat{\psi}, \hat{\lambda})|^{-1/2} |\jmath_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|^{-1/2}. \end{aligned}$$

On $L^0_{\psi_0}$, the interest parameter is fixed at ψ_0 and the constrainted MLE $\hat\lambda_{\psi_0}$ is fixed at its observed value $\hat\lambda^0$. From the exponential form of $g(u,v;\psi_0,\lambda)$, it easily follows that $\hat\lambda_{\psi_0}$ is a function of v only — not u (similarly, when χ is fixed, then $\hat\zeta_\chi$ is a function of v only). The observed profile line therefore corresponds to a contour on which v is fixed at the observed v^0. On that line, the saddlepoint approximation for the full density satisfies

$$\begin{aligned}&\tilde g(u,v^0;\psi_0,\lambda)\\ &\quad=\frac{\mathrm{e}^{k/n}}{(2\pi)^{p/2}}\exp\{\ell(\psi_0,\lambda;u,v^0)-\ell(\hat\psi,\hat\lambda;u,v^0)\}\ |\jmath_{\psi\psi}(\hat\psi,\hat\lambda)|^{-1/2}|\jmath_{\lambda\lambda}(\hat\psi,\hat\lambda)|^{-1/2}.\end{aligned}$$

The conditional density of v given the ancillary $U=u$, $q(v|u;\psi_0,\lambda)$, is a density on the $(p-1)$-dimensional variable space with parameter λ. Since $q(v|u;\psi_0,\lambda)\ \propto_v\ g(u,v;\psi_0,\lambda)$, it inherits exponential form from the full model and thus also admits a saddlepoint approximation, $\tilde q(v|u;\psi_0,\lambda)$. On the contour $L^0_{\psi_0}$, this conditional density is evaluated at v^0 with parameter λ:

$$\tilde q(v^0|u;\psi_0,\lambda)=\frac{\mathrm{e}^{k/n}}{(2\pi)^{(p-1)/2}}\exp\{\ell(\psi_0,\lambda;u,v^0)-\ell(\psi_0,\hat\lambda_{\psi_0};u,v^0)\}|\jmath_{\lambda\lambda}(\psi_0,\hat\lambda_{\psi_0})|^{-1/2}.$$

Dividing the full density $\tilde g(u,v^0;\psi_0,\lambda)$ by the conditional one $\tilde q(v^0|u;\psi_0,\lambda)$ leads to the marginal distribution $h(u;\psi_0)$ on $L^0_{\psi_0}$, with parameter ψ_0 and scalar differential $\mathrm{d}u$

$$\begin{aligned}h(u;\psi_0)=\frac{\mathrm{e}^{k/n}}{(2\pi)^{1/2}}&\exp\{\ell(\psi_0,\hat\lambda_{\psi_0};u,v^0)-\ell(\hat\psi,\hat\lambda;u,v^0)\}\\ &|\jmath_{\psi\psi}(\hat\psi,\hat\lambda)|^{-1/2}|\jmath_{\lambda\lambda}(\hat\psi,\hat\lambda)|^{-1/2}|\jmath_{\lambda\lambda}(\psi_0,\hat\lambda_{\psi_0})|^{1/2};\end{aligned}$$

we remind the reader that $\mathrm{e}^{k/n}$ is a generic normalizing constant. This density features third-order accuracy, and thus contains full third-order information on the interest parameter ψ_0.

If we now reexpress the density $h(u;\psi_0)$ in terms of the original canonical parameterization φ, we find

$$\begin{aligned}h(u;\psi_0)&=\frac{\mathrm{e}^{k/n}}{(2\pi)^{1/2}}\exp\{\ell(\hat\varphi_{\psi_0};u,v^0)-\ell(\hat\varphi;u,v^0)\}\\ &\qquad\qquad|\jmath_{(\psi\psi)}(\hat\varphi)|^{-1/2}|\jmath_{(\lambda\lambda)}(\hat\varphi)|^{-1/2}|\jmath_{(\lambda\lambda)}(\hat\varphi_{\psi_0})|^{1/2}\\ &=\frac{\mathrm{e}^{k/n}}{(2\pi)^{1/2}}\mathrm{e}^{-r^2_{\psi_0}/2}\ \left\{\frac{|\jmath_{(\lambda\lambda)}(\hat\varphi_{\psi_0})|}{|\jmath_{(\lambda\lambda)}(\hat\varphi)|}\right\}^{1/2}\ |\jmath_{(\psi\psi)}(\hat\varphi)|^{-1/2},\end{aligned}\tag{5}$$

with $-r^2_{\psi_0}/2=\ell(\hat\varphi_{\psi_0};u,v^0)-\ell(\hat\varphi;u,v^0)$. The parentheses around $\psi\psi$ and $\lambda\lambda$ indicate that the second derivatives must be rescaled with respect to the given exponential parameterization φ. The information determinants in the parameterization $(\psi),(\lambda)$ can

be obtained by applying Jacobians $\varphi_\psi = \partial\varphi/\partial\psi$ and $\varphi_\lambda = \partial\varphi/\partial\lambda$ to the determinants in the parameterization ψ, λ:

$$\begin{aligned}
|\jmath_{(\psi\psi)}(\hat\varphi)| &= |\jmath_{\psi\psi}(\hat\varphi)||\varphi_\psi^\top(\hat\varphi)\varphi_\psi(\hat\varphi)|^{-1},\\
|\jmath_{(\lambda\lambda)}(\hat\varphi)| &= |\jmath_{\lambda\lambda}(\hat\varphi)||\varphi_\lambda^\top(\hat\varphi)\varphi_\lambda(\hat\varphi)|^{-1},\\
|\jmath_{(\lambda\lambda)}(\hat\varphi_\psi)| &= |\jmath_{\lambda\lambda}(\hat\varphi_\psi)||\varphi_\lambda^\top(\hat\varphi_\psi)\varphi_\lambda(\hat\varphi_\psi)|^{-1}.
\end{aligned}$$

Equation (5) is a one-dimensional density for u expressed in terms of the original canonical parameter φ on the observed profile line $L^0_{\psi_0}$. The resulting expression is similar to (1), except that it involves an extra ratio of nuisance information determinants. We note that (5) also happens to be valid for vector ψ. For more information about the development of $h(u;\psi)$, and in particular the implications of using the tangent exponential model, see Fraser (2011).

3.3 Location form and standardization — density of ancillary statistic

Now that we have access to an explicit expression for $h(u;\psi)$, we may use it to compute an $\mathcal{O}(n^{-3/2})$ p-value function; evaluated at ψ_0, we write $p^*(\psi_0) = \int^{u^0} h(u;\psi_0)\mathrm{d}u$. In addition to this, the marginal null distribution $h(u;\psi_0)$ in (5) can be shown to have a location form to second-order accuracy; this will eventually lead to Bayes-frequency equivalence as in Section 2.3.

Suppose we temporarily ignore the factor $\{|\jmath_{(\lambda\lambda)}(\hat\varphi_{\psi_0})|/|\jmath_{(\lambda\lambda)}(\hat\varphi)|\}^{1/2}$ in (5); the contribution $|\jmath_{(\psi\psi)}(\hat\varphi)|^{-1/2}\mathrm{d}u$ then appears as the Welch and Peers (1963) differential on the observed profile line $L^0_{\psi_0}$ with respect to an underlying scalar exponential model; see (2). This therefore presents the expression $A_1 = \mathrm{e}^{k/n}(2\pi)^{-1/2}\cdot\exp\{-r^2_{\psi_0}/2\}|\jmath_{(\psi\psi)}(\hat\varphi)|^{-1/2}$ as a location model with variable t and parameter μ_0 to second-order accuracy, as argued in Section 2.3. Now, the factor $A_2 = \{|\jmath_{(\lambda\lambda)}(\hat\varphi_{\psi_0})|/|\jmath_{(\lambda\lambda)}(\hat\varphi)|\}^{1/2}$, already second-order accurate, can be expanded as a function $\exp\{a(t-\mu_0)/n^{1/2}\}$ with the same t and μ_0; see §A.1 of Fraser et al. (2016b). The product A_1A_2 is thus a function of $(t-\mu_0)$, providing a full location form for (5) to second order. A manipulation similar to (3) can then be applied to $p^*(\psi_0)$, leading to a prior density $\pi(\psi)$ that verifies Bayes-frequency equivalence for a scalar interest parameter in presence of a vector nuisance parameter. Before going forward with this, some technicalities however need attention.

The distribution (5) is on the line $L^0_{\psi_0}$ for a fixed ψ_0 and goes through the observed data (u^0, v^0). It is also perpendicular to the interest parameter contour, $P_\psi = \{\psi\in\mathbb{R} : \varphi = \hat\varphi_\psi\}$, at the constrained maximum likelihood value $\hat\varphi_{\psi_0}$ on the parameter space; see Figure 1. Now consider an arbitrary ψ value for the interest parameter; it turns out that $\psi(\varphi)$ often has certain rotation properties that cause the line L^0_ψ to change direction with ψ-change. This is the case, for instance, if $\varphi = (\varphi_1, \varphi_2)$ and $\psi = \varphi_1/\varphi_2$: as ψ varies on the parameter space (φ_1, φ_2), so does the direction of L^0_ψ. As a result, the observed information on L^0_ψ could also vary, as could the form of the underlying exponential distribution. In particular, if the observed information matrix is not proportional to the

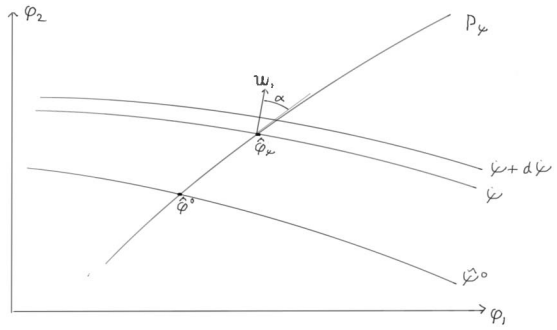

Figure 1: Exponential coordinates having symmetry ($\hat{\jmath}^0_{\varphi\varphi} = I$) at the point $\hat{\varphi}^0$; dψ is an increment in the parameter ψ; d$\hat{\varphi}_\psi$ is the corresponding vector increment for the point $\hat{\varphi}_\psi$ on the profile curve P_ψ; and d(ψ) is the corresponding increment in the symmetrized exponential coordinates.

identity matrix, then the scaling of underlying exponential distributions on L^0_ψ likely changes with ψ.

We can notationally avoid this complication by recalibrating the exponential coordinates φ to have an observed information matrix equal to the identity. This is not a change in substance, just notational so that what we have written as an exponential model is, under rotation, the same exponential model to second order. The recalibration is achieved by finding a right square root T of $\hat{\jmath}_{\varphi\varphi} = T^\top T$ and then using the modified canonical parameter $\bar{\varphi} = T\varphi$, which has acquired an identity observed information $\hat{\jmath}_{\bar{\varphi}\bar{\varphi}} = I$. This implies that the exponential distributions $h(u;\psi)$ through the data point are now, for various ψ, a single exponential distribution.

4 A Jeffreys-based prior featuring second-order reproducibility

We now combine the previous distributional results with Welch and Peers (1963)'s approach to derive a prior that achieves confidence, i.e. that possesses the reproducibility property. Hereafter, we use the modified exponential parameterization $\bar{\varphi}$ described at the end of Section 3.3; for notational simplicity, we however write φ and assume that the adjustment has been made.

4.1 Prior density

The density (5) can be integrated up to the observed $u = u^0$, leading to

$$p^*(\psi_0) = \int^{u^0} h(u;\psi_0)\mathrm{d}u$$

$$= \int^{u^0} \frac{\mathrm{e}^{k/n}}{(2\pi)^{1/2}} \mathrm{e}^{-r^2_{\psi_0}/2} \left\{ \frac{|\jmath_{(\lambda\lambda)}(\hat{\varphi}_{\psi_0})|}{|\jmath_{(\lambda\lambda)}(\hat{\varphi})|} \right\}^{1/2} |\jmath_{(\psi\psi)}(\hat{\varphi})|^{-1/2} \mathrm{d}u,$$

where $r^2_{\psi_0}$ and the information functions of course depend on the scalar u. From the developments in Section 3, recall that the integrand is a third-order approximation to the exact marginal density of the statistic $U(s)$, leading to a third-order p-value $p^*(\psi_0)$. Recall also the uniqueness of this density subject to retaining model continuity, and the fact that it contains full information about the interest parameter ψ_0.

As argued in Section 3.3, the density $h(u;\psi_0)$ in (5) can be expressed as a density that possesses location form to second order $\mathcal{O}(n^{-1})$. The implication of this property is that we can apply the location model result in (3) to the density $h(u;\psi_0)$. By first performing a change of variable from u to $\hat{\psi} = \hat{\psi}(u, v^0)$ on $L^0_{\psi_0}$ with Jacobian $|\jmath_{(\psi\psi)}(\hat{\psi}, \hat{\lambda}^0)| = |\jmath_{(\psi\psi)}(\hat{\varphi})|$, we find

$$p^*(\psi_0) = \int^{\hat{\psi}(u^0,v^0)} \underbrace{\frac{\mathrm{e}^{k/n}}{(2\pi)^{1/2}} \mathrm{e}^{-r^2_{\psi_0}/2} \left\{ \frac{|\jmath_{(\lambda\lambda)}(\hat{\varphi}_{\psi_0})|}{|\jmath_{(\lambda\lambda)}(\hat{\varphi})|} \right\}^{1/2}}_{\bar{g}(t-\mu_0)\{1+\mathcal{O}(n^{-1})\}} \underbrace{|\jmath_{(\psi\psi)}(\hat{\varphi})|^{1/2} \mathrm{d}\hat{\psi}}_{\mathrm{d}t}.$$

Then, by mimicking (3), the integrand may be approximated by the function $\bar{g}(t-\mu_0)$, a location model with variable t and fixed parameter μ_0. Recall that for such models, elementary calculus leads to $\int^{t^0} \bar{g}(t-\mu_0)\mathrm{d}t = \int_{\mu_0} \bar{g}(t^0-\mu)\mathrm{d}\mu$. With t^0 fixed and μ varying, the density on the right is

$$\begin{aligned} \bar{g}(t^0-\mu)\{1+\mathcal{O}(n^{-1})\} &= \frac{\mathrm{e}^{k/n}}{(2\pi)^{1/2}} \mathrm{e}^{-r^2_{\psi}/2} \left\{ \frac{|\jmath_{(\lambda\lambda)}(\hat{\varphi}_{\psi})|}{|\jmath_{(\lambda\lambda)}(\hat{\varphi})|} \right\}^{1/2} \\ &= \frac{\mathrm{e}^{k/n}}{(2\pi)^{1/2}} \exp\{\ell(\hat{\varphi}_\psi; u^0, v^0) - \ell(\hat{\varphi}; u^0, v^0)\} \left\{ \frac{|\jmath_{(\lambda\lambda)}(\hat{\varphi}_{\psi})|}{|\jmath_{(\lambda\lambda)}(\hat{\varphi})|} \right\}^{1/2}. \end{aligned}$$

We are left finding an expression for the differential $\mathrm{d}\mu$ in terms of the special parameterization (ψ); the latter indicates that we integrate with respect to φ, but along the contour P_ψ in the parameter space.

In Section 2.3, we went from an integral with respect to s on the variable space, to one with respect to φ on the parameter space. The transformations implied that $\mathrm{d}t = |\jmath_{\varphi\varphi}(\hat{\varphi})|^{-1/2}\mathrm{d}s = |\jmath_{\varphi\varphi}(\hat{\varphi})|^{1/2}\mathrm{d}\hat{\varphi}$ and $\mathrm{d}\mu = |\jmath_{\varphi\varphi}(\varphi)|^{1/2}\mathrm{d}\varphi$. In a similar fashion, we go here from an integral with respect to u on the line $L^0_{\psi_0} = \{(u, v^0) : \hat{\lambda}_{\psi_0} = \hat{\lambda}^0\}$ to one with respect to (ψ) on the ψ contour $P_\psi = \{\psi \in \mathbb{R} : \varphi = \hat{\varphi}_\psi\}$, where $\hat{\varphi}_\psi = \hat{\varphi}^0_\psi$ is based on observed data (u^0, v^0). The transformations then imply that $\mathrm{d}t = |\jmath_{(\psi\psi)}(\hat{\varphi})|^{-1/2}\mathrm{d}u = |\jmath_{(\psi\psi)}(\hat{\varphi})|^{1/2}\mathrm{d}\hat{\psi}$ and $\mathrm{d}\mu = |\jmath_{(\psi\psi)}(\hat{\varphi}_\psi)|^{1/2}\mathrm{d}(\psi)$. Indeed, given that v is fixed at its observed v^0 and that $\hat{\lambda}_\psi = \hat{\lambda}^0$, the differential is a function of the constrained MLE $\hat{\varphi}_\psi$.

To summarize, we have

$$p^*(\psi_0) = \int^{u^0} \underbrace{\frac{\mathrm{e}^{k/n}}{(2\pi)^{1/2}} \mathrm{e}^{-r^2_{\psi_0}/2} \left\{ \frac{|\jmath_{(\lambda\lambda)}(\hat{\varphi}_{\psi_0})|}{|\jmath_{(\lambda\lambda)}(\hat{\varphi})|} \right\}^{1/2}}_{\bar{g}(t-\mu_0)\{1+\mathcal{O}(n^{-1})\}} \underbrace{|\jmath_{(\psi\psi)}(\hat{\varphi})|^{-1/2} \mathrm{d}u}_{\mathrm{d}t} \tag{6}$$

$$\approx \int_{\psi_0} \underbrace{\frac{\mathrm{e}^{k/n}}{(2\pi)^{1/2}}\mathrm{e}^{-r_\psi^2/2} \quad \left\{\frac{|\jmath_{(\lambda\lambda)}(\hat\varphi_\psi)|}{|\jmath_{(\lambda\lambda)}(\hat\varphi)|}\right\}^{1/2}}_{\bar g(t^0-\mu)\{1+\mathcal{O}(n^{-1})\}} \underbrace{|\jmath_{(\psi\psi)}(\hat\varphi_\psi)|^{1/2}\mathrm{d}(\psi)}_{\mathrm{d}\mu}. \tag{7}$$

The integral on $L^0_{\psi_0}$ in (6) leads to $p^*(\psi_0)$, a third-order approximation to $p(\psi_0)$, while the integral on $P_\psi = \{\psi \in \mathbb{R} : \varphi = \hat\varphi_\psi\}$ in (7) is an $\mathcal{O}(n^{-1})$ approximation to $p(\psi_0)$. We can then approximate the p-value function using a posterior survivor function.

We can simplify (7) by using the orthogonality of ψ and λ to affirm that $|\jmath_{\varphi\varphi}(\hat\varphi_\psi)| = |\jmath_{(\lambda\lambda)}(\hat\varphi_\psi)||\jmath_{(\psi\psi)}(\hat\varphi_\psi)|$. We can also absorb the information term $|\jmath_{(\lambda\lambda)}(\hat\varphi)|^{-1/2}$, which depends solely on the data, into the arbitrary constant k:

$$p(\psi_0) \approx \int_{\psi_0} \frac{\mathrm{e}^{k/n}}{(2\pi)^{1/2}}\mathrm{e}^{-r_\psi^2/2}\ |\jmath_{\varphi\varphi}(\hat\varphi_\psi)|^{1/2}\ \mathrm{d}(\psi) = \jmath(\psi_0). \tag{8}$$

As before, the parentheses around ψ indicate that the integrand is expressed in terms of φ, but that we integrate on the profile contour P_ψ; more details about the differential $\mathrm{d}(\psi)$ are provided in Section 4.2. Furthermore, to avoid notational difficulties with parameter rotation, recall that the exponential parameter φ is locally rotationally symmetric as described in the last paragraph of Section 3.3.

The $\mathcal{O}(n^{-1})$ version of the p-value in (8) has now been written as an integral of observed likelihood on the parameter space. This integral of likelihood is totally restricted to the profile curve P_ψ; as such, the integral is a contour integral, and not the usual full parameter space integral. The integrand in (8) can thus be seen as a posterior density for ψ obtained from the directional prior $\pi_D(\psi)\mathrm{d}\psi \propto |\jmath_{\varphi\varphi}(\hat\varphi_\psi)|^{1/2}\ \mathrm{d}(\psi)$, which is Jeffreys' prior used 'off-label', on just the profile P_ψ for the interest parameter. The resulting posterior survivor value $\jmath(\psi_0)$ in (8) has full second-order repetition accuracy.

4.2 Jacobian

We now study the differential $\mathrm{d}(\psi)$ based on the special exponential parameterization (ψ). The parentheses in (ψ) are to show that we integrate with respect to the canonical variable φ, but along a profile curve P_ψ developed in terms of ψ. On that curve, the constrained MLE $\hat\varphi_\psi$ varies as a function of the interest parameter ψ. Nonetheless, working in terms of φ was convenient as we did not need identifying a nuisance parameter λ orthogonal to ψ.

To make a change of variable from (ψ) to ψ, we need the Jacobian $|\mathrm{d}(\psi)/\mathrm{d}\psi|$, whose calculation involves a few steps. We first need to establish how a φ-change affects ψ along P_ψ. Once we have a standardized measure available for this, we can multiply it with $|\mathrm{d}\hat\varphi_\psi/\mathrm{d}\psi|$, the magnitude of a change in $\hat\varphi_\psi$ resulting from a ψ-change along P_ψ.

A change $\mathrm{d}\varphi$ generates a change $\mathrm{d}\psi$ along the gradient vector $\mathrm{d}\psi/\mathrm{d}\varphi$ of the ψ surface. If we are currently at the point $\hat\varphi_\psi$ on P_ψ, then a unit version of this vector is denoted $w_1 = \{\mathrm{d}\psi(\varphi)/\mathrm{d}\varphi\}/|\mathrm{d}\psi(\varphi)/\mathrm{d}\varphi|$, evaluated at $\hat\varphi_\psi$. In a similar fashion, a change $\mathrm{d}\psi$ generates a change $\mathrm{d}\hat\varphi_\psi$ along the gradient vector $\mathrm{d}\hat\varphi_\psi/\mathrm{d}\psi$ of the profile contour P_ψ. A unit version of this vector is denoted $w_2 = \{\mathrm{d}\hat\varphi_\psi/\mathrm{d}\psi\}/|\mathrm{d}\hat\varphi_\psi/\mathrm{d}\psi|$, also evaluated at $\hat\varphi_\psi$.

By looking at Figure 1, the standardized change w_1 (a unit change for ψ in the direction $\mathrm{d}\psi/\mathrm{d}\varphi$ viewed as originating from some change $\mathrm{d}\varphi$) obviously has a lesser effect along P_ψ. This effect is represented by the projection of w_1 onto w_2 (see Figure 1). In particular, if we let α be the angle between these unit vectors, then the cosine of α provides the magnitude of this projection onto w_2.

Now, recall that a ψ-change along the gradient vector of the profile contour P_ψ generates a $\hat{\varphi}_\psi$-change with magnitude $|\mathrm{d}\hat{\varphi}_\psi/\mathrm{d}\psi|$. The φ-change perpendicular to a ψ contour (along P_ψ) is then obtained by multiplying the magnitude $|\mathrm{d}\hat{\varphi}_\psi/\mathrm{d}\psi|$ and the cosine of α:

$$\begin{aligned}\mathrm{d}(\psi) &= \left|\frac{\mathrm{d}(\psi)}{\mathrm{d}\psi}\right| \mathrm{d}\psi = \cos\{\alpha\} \left|\frac{\mathrm{d}\hat{\varphi}_\psi}{\mathrm{d}\psi}\right| \mathrm{d}\psi \\ &= w_1 \cdot w_2 \left|\frac{\mathrm{d}\hat{\varphi}_\psi}{\mathrm{d}\psi}\right| \mathrm{d}\psi = w_1 \cdot \frac{\mathrm{d}\hat{\varphi}_\psi}{\mathrm{d}\psi}\, \mathrm{d}\psi, \end{aligned} \tag{9}$$

where $\cdot$ is the dot product.

Using this Jacobian, we obtain the Bayes posterior survivor function for a general scalar interest parameter ψ_0:

$$\mathcal{s}(\psi_0) = c \int_{\psi_0} \mathrm{e}^{-r_\psi^2/2} |\jmath_{\varphi\varphi}(\hat{\varphi}_\psi)|^{1/2} \left|\frac{\mathrm{d}\hat{\varphi}_\psi}{\mathrm{d}\psi}\right| \cos\{\alpha\}\, \mathrm{d}\psi,$$

where $|\mathrm{d}\hat{\varphi}_\psi/\mathrm{d}\psi| \cos\{\alpha\}$ represents the Jacobian for Jeffreys' prior on the profile curve, $|\jmath_{\varphi\varphi}(\hat{\varphi}_\psi)|^{1/2}$. This posterior survivor function $\mathcal{s}(\psi_0)$ has second-order reproductive accuracy as derived from the p-value function: $p(\psi_0) = \mathcal{s}(\psi_0)\{1 + \mathcal{O}(n^{-1})\}$.

Following (9), the implicit prior density is thus expressed as

$$\pi_D(\psi)\, \mathrm{d}\psi \propto |\jmath_{\varphi\varphi}(\hat{\varphi}_\psi)|^{1/2} \mathrm{d}(\psi) = |\jmath_{\varphi\varphi}(\hat{\varphi}_\psi)|^{1/2}\, w_1 \cdot \frac{\mathrm{d}\hat{\varphi}_\psi}{\mathrm{d}\psi}\, \mathrm{d}\psi \tag{10}$$

for ψ on P_ψ, and may be used with the profile log-likelihood $\ell(\hat{\varphi}_\psi; s^0)$ for further Bayesian developments. Note that the nuisance parameter nowhere appears in the prior, nor in the posterior; in practice, identifying λ orthogonal to ψ is thus not required. The posterior survivor function $\mathcal{s}(\cdot)$ in this section has second-order uniqueness and accuracy by its derivation from the p^*-value function, which has second-order uniqueness and accuracy by calculation respecting continuity. See Fraser (2014) for a related discussion.

4.3 Some thoughts on the new prior density

Tibshirani (1989) proposes a prior density featuring frequentist reproducibility properties; to this end, he considers a scalar parameter of interest ψ along with a nuisance parameter λ that is orthogonal to ψ (in the sense of Cox and Reid, 1987). The resulting prior is the product of Jeffreys' prior for the interest component ψ and an arbitrary function $g(\lambda)$ for the nuisance parameter. Tibshirani (1989) expounds his result by referring to Welch and Peers (1963), and therefore implicitly works under the assumption

that the statistical model is continuous in the variable and parameters, with i.i.d. data. Although he does not explicitly focus on full exponential models, the examples studied all involve the normal distribution.

The prior density he proposes is obviously not unique; while it is completely determined in the ψ direction, the function $g(\lambda)$ may be characterized using various approaches. Once the likelihood is marginalized with respect to λ however, the resulting inference is equivalent to working with the profile likelihood for ψ and the one-dimensional Jeffreys' prior for ψ.

The current paper may be seen as a generalization of Tibshirani (1989)'s approach to cases where a regular statistical model, not necessarily in exponential form, is expressed in its canonical parameterization (no need to find orthogonal interest and nuisance parameters). It provides a general expression for the prior density and Jacobian associated to any regular model, which then gives access to a posterior survivor function $\jmath(\cdot)$ along the profile P_ψ. The resulting inferences about the scalar interest parameter ψ are second-order reproducible and do not require the identification of an orthogonal nuisance parameter λ. The proposed one-dimensional prior is expressed in terms of $\hat{\varphi}_\psi$, but has not been obtained through optimization explicitly. It is rather seen as arising from a marginalization of the full model, itself approximated using Laplace expansions.

The proposed prior can be extended over the whole parameter space. To this end, we however need reverting to an orthogonal setting (ψ, λ). As in Tibshirani (1989), we would then use the prior $|\jmath_{\psi\psi}(\psi, \hat{\lambda}_\psi)|^{1/2}$ for ψ, choose an arbitrary prior $g(\lambda)$ for the orthogonal nuisance parameter, and then apply a change of variable to move from $\pi(\psi, \lambda) \propto |\jmath_{\psi\psi}(\psi, \hat{\lambda}_\psi)|^{1/2} g(\lambda)$ to the canonical parameterization φ. The interest in performing this last change of variable is not clear however: the previous steps require identifying an orthogonal nuisance parameter λ, so one might as well stick with the prior $\pi(\psi, \lambda)$ for the calculations.

If we use $\mathcal{L}(\psi, \lambda; u, v)$ with the prior $\pi(\psi, \lambda) = \pi(\lambda|\psi)\pi(\psi)$ to find the integrated likelihood function, we obtain

$$\begin{aligned}\mathcal{L}(\psi; u^0) &= \int \mathcal{L}(\psi, \lambda; u^0, v^0)\pi(\lambda|\psi)\, \mathrm{d}\lambda \\ &= \mathcal{L}(\psi, \hat{\lambda}_\psi; u^0)|\jmath_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{-1/2}\pi(\hat{\lambda}_\psi|\psi)\{1 + \mathcal{O}_p(n^{-1})\}.\end{aligned}$$

The second equality arises from the use of Laplace expansions (see Tierney and Kadane, 1986; Tierney et al., 1989), and the first two terms of that expression refer to Cox and Reid (1987)'s adjusted profile likelihood. In this context, the proposed approach is thus equivalent to choosing $\pi(\lambda|\psi) \equiv \pi(\lambda)$, integrating over λ, and then reexpressing the integrated inference problem in terms of φ (instead of ψ). The main strength of the new prior density is that it yields reproducible inferences without requiring orthogonal parameters. While used in several second-order approaches (see Tibshirani, 1989 and Severini, 2007, for instance), this orthogonality assumption requires solving differential equations Cox and Reid (1987), which often is a daunting task.

5 A prior for sets?

In Section 3 and Section 4, we considered a vector parameter φ and identified a scalar parameter of interest ψ. By focussing on the latter, we found a prior density $\pi_D(\psi)$ for ψ on the profile contour P_ψ that features second-order reproducibility properties. A reasonable concern is whether such second-order posterior accuracy might be available more generally. In particular, does a similar density exist for vector ψ? Or even for a scalar ψ that involves computations on a compact set rather than a one-dimensional contour?

We answer this question by presenting a simple core example where we can reasonably hope that things would be easy. Let $S = (S_1, S_2)$ on the plane be distributed according to a standard Normal centered at $\varphi = (\varphi_1, \varphi_2)$. This is a simple exponential model with canonical variable s and canonical parameter φ:

$$\begin{aligned} g(s;\varphi) &= (2\pi)^{-1} \exp\left\{-\tfrac{1}{2}[(s_1-\varphi_1)^2 + (s_2-\varphi_2)^2]\right\} \\ &= (2\pi)^{-1} \exp\left\{-\tfrac{1}{2}(s_1^2+s_2^2)\right\} \exp\left\{-\tfrac{1}{2}(\varphi_1^2+\varphi_2^2)\right\} \exp\left\{s_1\varphi_1 + s_2\varphi_2)\right\}. \end{aligned}$$

Let $\rho^2 = \varphi_1^2 + \varphi_2^2$ be the parameter of interest; here, ρ^2 represents the squared distance from the origin $(0,0)$ to the point (φ_1, φ_2) on the plane. Now, let $s^0 = (s_1^0, s_2^0)$ be the observed data point; the squared distance from the origin to this observation is $r^2 = (s_1^0)^2 + (s_2^0)^2$.

Given the distribution of S, the variable $S_1^2 + S_2^2$ is distributed as a Noncentral Chi-squared with $\nu = 2$ degrees of freedom and noncentrality parameter ρ_0^2; its distribution function is denoted $H_\nu(\cdot;\rho_0^2)$. This variable may be seen as the squared norm of a random vector with $\mathcal{N}(\varphi, I_2)$ distribution. The p-value for testing a specific parameter value ρ_0^2 therefore is the probability that (S_1, S_2) be closer to the origin than the observed s^0 given the parameter $(\varphi_1, \varphi_2)_0$, itself at a distance ρ_0 from the origin. We then look for the probability of the set $\{s \in \mathbb{R}^2 : s_1^2 + s_2^2 < r^2\}$ and the p-value satisfies $p(\rho_0^2) = \mathbb{P}(S_1^2 + S_2^2 < r^2; \rho_0^2) = H_2(r^2; \rho_0^2)$.

Now, since φ is a location parameter, the prior density for ρ^2 should be coherent with Jeffreys' flat prior for φ, $\pi(\varphi) \propto 1$. This leads to the following posterior density on the plane

$$\pi(\varphi|s^0) = (2\pi)^{-1} \exp\left\{-\tfrac{1}{2}[(\varphi_1 - s_1^0)^2 + (\varphi_2 - s_2^0)^2]\right\}.$$

The posterior density of φ is a standard Normal located at (s_1^0, s_2^0); it follows that ρ^2 is distributed as a Noncentral Chi-squared distribution with 2 degrees of freedom and noncentrality parameter r^2. The Bayes survivor function evaluated at ρ_0^2, which is the probability that the squared distance from the origin to ρ^2 be at least ρ_0^2, then satisfies $\jmath(\rho_0^2) = 1 - H_2(\rho_0^2; r^2)$.

Hence, both functions can easily be compared. First, let us look at $p(\rho_0^2)$ as a function of ρ_0^2. The p-value function evaluates the sample space probability within a disk of fixed radius r centered at the origin. To compute this probability, it uses a standard Normal density whose mode is initially located at $\rho_0 = 0$, and then gradually moves away from $(0,0)$. The function $p(\rho_0^2) = H_2(r^2; \rho_0^2)$ then starts at some value (< 1) for $\rho_0^2 = 0$

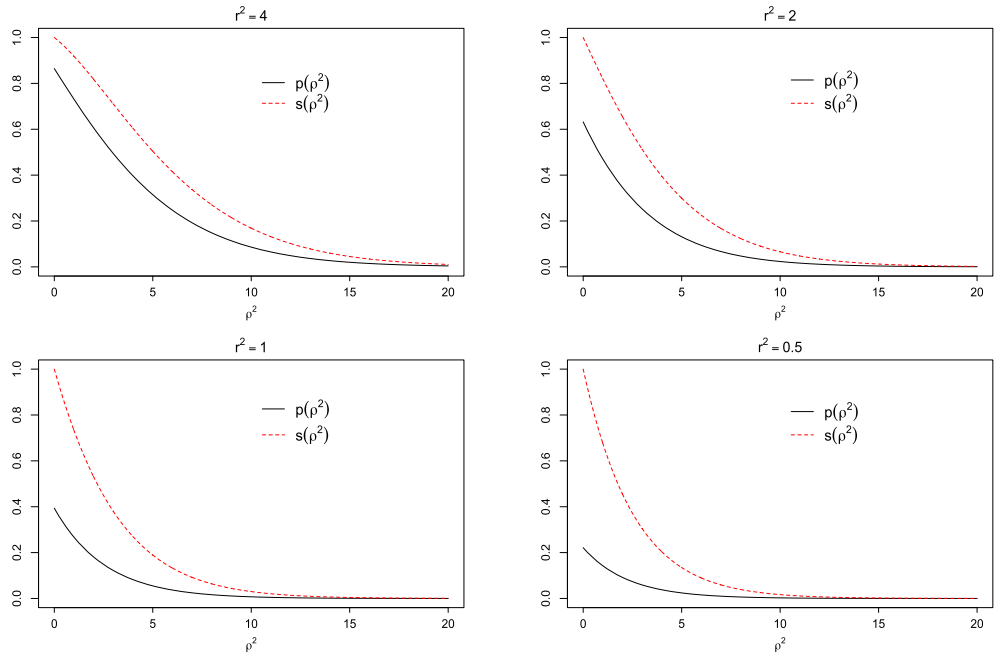

Figure 2: p-value and posterior survivor value functions for the parameter ρ^2; each graph has its own r^2 value: 4, 2, 1, and 0.5.

and decreases towards 0 as ρ_0^2 increases. By opposition, the Bayes survivor function evaluates the parameter space probability outside a circle of radius ρ_0 centered at the origin; the size of the circle gradually increases with ρ_0. The probability is computed using a standard Normal density centered at the fixed point s^0, at a distance r from the origin. The function $\mathfrak{s}(\rho_0^2) = 1 - H_2(\rho_0^2; r^2)$ therefore starts at 1 for $\rho_0 = 0$ and decreases towards 0 as ρ_0 increases.

When $r^2 = \rho_0^2$ for instance, the p-value $H_2(r^2; r^2)$ evaluates the probability within a disk of radius r using a standard Normal whose mode is located on the boundary of the disk. The Bayes survivor value $1 - H_2(r^2; r^2)$ evaluates the probability outside the same disk, using the same Normal distribution. It follows that the Bayes survivor function is larger than the p-value function, a familiar result in the presence of parameter curvature. And then if we decrease the value of r, the p-value moves towards 0 and the Bayes survivor function moves towards 1. In the extreme, the p-value can be close to 0 and the corresponding survivor value close to 1. As the p-value has repetition validity, it follows that the Bayes survivor probability in general does not, here to the extreme.

Figure 2 illustrates the behavior of the p-value and Bayes survivor functions of the parameter ρ^2 for various values of the observed radius ($r^2 = 4, 2, 1, 0.5$). In particular, the discrepancy between both approaches becomes larger as r^2 decreases, in which case the p-value function becomes closer to 0. The above should not be surprising given the behavior of the pivotal r/ρ in calculating confidence. Given that the usual flat prior for

location parameters does not yield a Bayes survivor function that matches the p-value function, we conclude that reproducibility is generally not attainable on sets.

6 Examples

We now present a range of examples based on simple exponential models in which the parameter of interest ψ increases in complexity. We begin with a parameter ψ that is linear in the canonical parameterization φ, then consider a rotational ψ, and follow with a ψ that is curved in terms of φ. We conclude with the Behrens-Fisher problem, which features a three-dimensional nuisance parameter λ. We detail how the new reproducibility prior is obtained in each of these cases, and graphically assess its performance by comparison to the exact p-value function. When the latter is not available, we instead present an exact conditional p-value based on Markov chain Monte Carlo (MCMC) methods, and also include the frequentist benchmark that is the third-order p-value.

6.1 Linear parameter

Consider an interest parameter ψ linear in the canonical parameterization φ, i.e. $\psi(\varphi) = v^\top\varphi = \sum v_i\varphi_i$. In this simple case, the line L_ψ^0 remains parallel to the vector v under ψ changes. Since L_ψ^0 does not rotate, there is no need to invoke rotational symmetry in the observed information matrix $\hat{j}_{\varphi\varphi}$, thus waiving the recalibration discussed at the end of Section 3.3. Users looking for an automated implementation of the method could nonetheless include a default use of this recalibration without altering results.

Specifically, consider a beta density with canonical parameter $\varphi = (\alpha, \beta)$:

$$f(y;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}, \quad y \in (0,1),$$

with $n = 5$ observed values $\mathbf{y}^0 = (0.36, 0.68, 0.44, 0.43, 0.34)$. The parameter α is of interest ($\psi = \alpha$), while β is a free nuisance ($\lambda = \beta$). The interest ψ is linear in the canonical parameterization, as $\psi(\varphi) = v^\top\varphi = \alpha$ with $v^\top = (1, 0)$.

We aim at comparing the p-value function $p(\alpha)$ to posterior survivor value functions $s(\alpha)$ arising from available uninformative priors. To this end, the signed log-likelihood root (SLR) approach is used as a simple approximation to $p(\alpha)$, while the third-order approach acts as a highly accurate one Fraser (2017); MCMC simulations replace the exact p-value when the latter is not available. These are then compared to posterior survivor value functions $s(\alpha)$ obtained using Jeffreys' prior and the new directional Jeffreys-style prior.

The beta model does not admit closed-form expressions for its maximum likelihood estimates (MLEs). Using the function `beta.mle` in the `R` package `Rfast` leads to $\hat{\varphi}^0 = (7.47, 9.03)$; this estimate is used in approximating $p(\alpha)$. The constrained MLE of β given α, $\hat{\beta}_\alpha$, is the solution of

$$D'(\hat{\beta}_\alpha) - D'(\alpha + \hat{\beta}_\alpha) = \frac{1}{n}\sum_{i=1}^{n} \log(1 - y_i^0),$$

where $D'(x) = \mathrm{d}\log\Gamma(x)/\mathrm{d}x$ is the digamma function. This equation is solved using the function `uniroot` in `R`; constrained MLEs $\hat{\beta}_\alpha$ are obtained for various values of the interest α, and then used in approximating $p(\alpha)$ and computing posterior survivor values $s(\alpha)$ based on the new directional Jeffreys-style prior.

The Fisher information function appears in every calculation (except in the SLR); it satisfies

$$j_{\varphi\varphi}(\varphi) = \begin{pmatrix} n\left(D''(\alpha) - D''(\alpha+\beta)\right) & -nD''(\alpha+\beta) \\ -nD''(\alpha+\beta) & n\left(D''(\beta) - D''(\alpha+\beta)\right) \end{pmatrix},$$

where $D''(x) = \mathrm{d}^2\log\Gamma(x)/\mathrm{d}x^2$ is the trigamma function. Jeffreys' prior does not distinguish between interest and nuisance parameters; it is defined on the full parameter space as the root of the Fisher information determinant:

$$\pi_J(\alpha,\beta) \propto \left\{D''(\alpha)D''(\beta) - D''(\alpha+\beta)\left[D''(\alpha) + D''(\beta)\right]\right\}^{1/2}, \quad \alpha,\beta > 0.$$

We note that the Bayesian benchmark prior, the reference prior of Bernardo (1979), is not easily available for a beta model in which an interest parameter is targeted. If it were, it would also lead to a prior on the full parameter space, but interest and nuisance parameters would have been treated differently in the derivation of this density.

As a new way of targeting the interest parameter, the directional Jeffreys-style prior restricts the usual Jeffreys' prior to the profile contour for the interest α. From (10), the new prior π_D satisfies

$$\pi_D(\alpha)\ \mathrm{d}\alpha \propto |j_{\varphi\varphi}(\hat{\varphi}_\psi)|^{1/2}\ \mathrm{d}(\psi) = \pi_J(\alpha,\hat{\beta}_\alpha)\ \mathrm{d}(\alpha).$$

In the current simple linear case, w_1 in (9) is the vector $(1,0)$, and thus

$$\mathrm{d}(\alpha) = w_1 \cdot \frac{\mathrm{d}(\alpha,\hat{\beta}_\alpha)^\top}{\mathrm{d}\alpha}\ \mathrm{d}\alpha = \mathrm{d}\alpha.$$

The posterior survivor value function $s_D(\alpha)$ is then obtained by integrating the one-dimensional posterior density

$$\pi_D(\alpha|\mathbf{y}^0)\ \mathrm{d}\alpha \propto \exp\left\{\ell(\alpha,\hat{\beta}_\alpha;\mathbf{y}^0)\right\} |j_{\varphi\varphi}(\alpha,\hat{\beta}_\alpha)|^{1/2}\ \mathrm{d}\alpha,$$

where $\ell(\alpha,\hat{\beta}_\alpha;\mathbf{y}^0) = \log f(\mathbf{y}^0;\alpha,\hat{\beta}_\alpha)$ denotes the profile log-likelihood function of the interest α.

Figure 3 examines the third-order function $p(\alpha)$ (solid line) and the normal approximation for the signed log-likelihood root r_α (dash-dotted line). The graph also features a comparison with posterior survivor values obtained under Jeffreys' prior (dotted line) and the new directional Jeffreys (red dashed line). 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 featuring a scaling $\sigma^2 = 4$ (also in `R`). In the current context, the

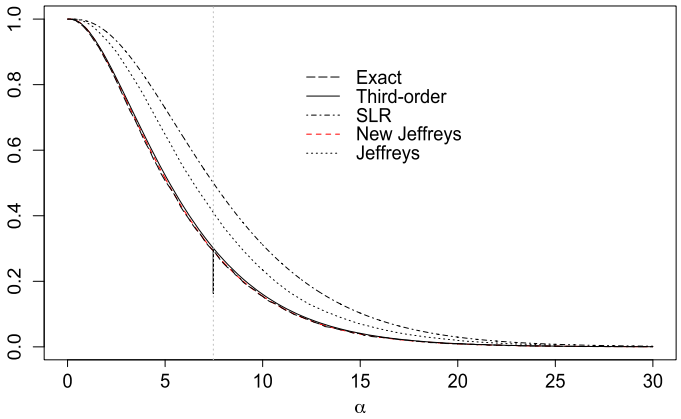

Figure 3: p-value and Bayes survivor functions for the parameter α in the beta model; the MLE of α is identified by a pale vertical line.

directional Jeffreys offers second-order reproducibility; this is not available from the regular Jeffreys, which treats both parameters as of equal importance.

In this example, we studied the simplistic case where $\varphi = (\alpha, \beta) = (\psi, \lambda) \equiv \theta$. Linear examples with $\varphi \neq \theta$ are easy to find and in such cases, the development of the new prior is similar to that expounded in the current section. Consider, for instance, the same beta model and let $\psi = \alpha + \beta$, $\lambda = \beta$; this yields an interest parameter that is still linear in φ, expressed as $\psi(\varphi) = v^\top \varphi$ with $v^\top = (1, 1)$. The constrained MLE of β given ψ, $\hat{\beta}_\psi$, is now the solution of

$$D'(\psi - \hat{\beta}_\psi) - D'(\hat{\beta}_\psi) = \frac{1}{n}\sum_{i=1}^{n} \log y_i^0 - \frac{1}{n}\sum_{i=1}^{n} \log(1 - y_i^0).$$

The vector w_1 in (9) is $w_1 = (1, 1)/\sqrt{2}$, leading to

$$\mathrm{d}(\psi) = w_1 \cdot \frac{\mathrm{d}(\psi - \hat{\beta}_\psi, \hat{\beta}_\psi)^\top}{\mathrm{d}\psi}\, \mathrm{d}\psi. \tag{11}$$

In practice, an analytical expression for $\mathrm{d}\hat{\varphi}_\psi/\mathrm{d}\psi$ is not always available. In such cases, $\mathrm{d}(\psi)$ is simply reexpressed as $\mathrm{d}(\psi) = w_1 \cdot \mathrm{d}\hat{\varphi}_\psi$ and posterior survivor values are then easily obtained using numerical integration, by selecting an appropriately small lag h and letting $\mathrm{d}\hat{\varphi}_\psi \approx \hat{\varphi}_{\psi+h} - \hat{\varphi}_\psi$.

From (10) and (11), the new prior satisfies $\pi_D(\psi)\mathrm{d}\psi \propto |j_{\varphi\varphi}(\psi - \hat{\beta}_\psi, \hat{\beta}_\psi)|^{1/2} w_1 \cdot \mathrm{d}\hat{\varphi}_\psi$ and is combined to the profile likelihood to obtain posterior survivor values, as explained above. Figure 4 provides a comparison similar to that found in Figure 3, outlining virtually parallel performances amongst implemented methods.

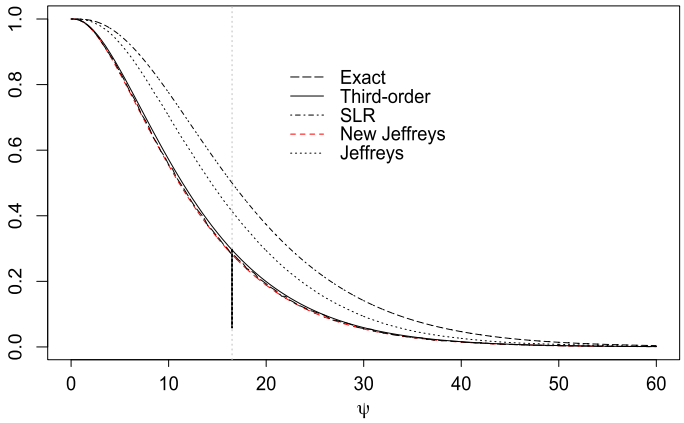

Figure 4: p-value and posterior survivor value functions for the parameter $\psi = \alpha + \beta$ in the beta model; the MLE of ψ is identified by a pale vertical line.

6.2 Rotating parameter

In several cases, the direction of the line L_ψ^0 may vary under ψ changes. Although this does not happen in linear cases, more generally, L_ψ^0 may rotate through an $\mathcal{O}(n^{-1/2})$ angle. This even happens in very simple settings and with classical distributions, as the following example illustrates.

Consider a normal model in which $Y \sim \mathcal{N}(\mu, \sigma^2)$ and let $\theta = (\psi, \lambda) = (\mu, \sigma^2)$. For a vector of n observations, the log-likelihood function of this model satisfies

$$\ell(\mu, \sigma^2; \mathbf{y}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n y_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^n y_i - \frac{n\mu^2}{2\sigma^2}. \tag{12}$$

From (12), the canonical parameters are $\varphi(\theta) = (\mu/\sigma^2, -1/\sigma^2)$. The interest parameter thus satisfies $\psi(\varphi) = -\varphi_1/\varphi_2 = \mu$, which is obviously not linear in φ. The maximum likelihood estimates in the canonical parameterization are $\hat\varphi = (n\bar y/S^2, -n/S^2)$, where $\bar y = \sum_{i=1}^n y_i/n$ and $S^2 = \sum_{i=1}^n y_i^2 - n\bar y^2$. The constrained MLE for σ^2 given μ is $\hat\sigma_\mu^2 = \{S^2 + n(\bar y - \mu)^2\}/n$, leading to $\hat\varphi_\psi = (\mu/\hat\sigma_\mu^2, -1/\hat\sigma_\mu^2)$. The Fisher information function is expressed as

$$\jmath_{\varphi\varphi}(\varphi) = \begin{pmatrix} -n/\varphi_2 & n\varphi_1/\varphi_2^2 \\ n\varphi_1/\varphi_2^2 & n/2\varphi_2^2 - n\varphi_1^2/\varphi_2^3 \end{pmatrix}. \tag{13}$$

Using (13), Jeffreys' prior is $\pi_J(\varphi)\mathrm{d}\varphi \propto |\jmath_{\varphi\varphi}(\varphi)|^{1/2}\mathrm{d}\varphi \propto (-\varphi_2)^{-3/2}\,\mathrm{d}\varphi$ on $\mathbb{R}\times\mathbb{R}^-$. Furthermore, Bernardo (1979)'s reference prior satisfies $\pi_R(\varphi)\mathrm{d}\varphi \propto \mathrm{d}\varphi/\varphi_2^2$.

We now proceed to determine the new Jeffreys-style prior, based on an observed sample $\mathbf{y}^0 = (0.00, 1.10, -0.50, 0.25, -0.95, -0.60, 0.35)$. Since the angle of L_ψ^0 rotates under ψ changes, we apply the recalibration $\bar\varphi = T\varphi$ with $\jmath_{\varphi\varphi}(\hat\varphi) = T^\top T$ and $\hat\varphi = \hat\varphi^0$,

as mentioned in Section 3.3. For simplicity and interchangeability in the use of T and its transpose, we find the eigenvalues and eigenvectors of $\jmath_{\varphi\varphi}(\hat{\varphi})$ with the function `eigen` in `R`, and then use these quantities to define a symmetrical matrix T.

In practice, this change from φ to $\bar{\varphi}$ only affects (10) through the differential $\mathrm{d}(\psi)$. Indeed, since $\jmath_{\bar{\varphi}\bar{\varphi}} = (T^{-1})^\top \jmath_{\varphi\varphi} T^{-1}$, then $|\jmath_{\bar{\varphi}\bar{\varphi}}| = |\jmath_{\varphi\varphi}|/|T|^2$; when evaluated at $\theta = (\psi, \hat{\lambda}_\psi)$ along the profile curve, these determinants are proportional with respect to ψ and therefore interchangeable in terms of Bayesian computations. The determinant of (13) evaluated at $\hat{\varphi}_\psi$ is then computed as

$$|\jmath_{\varphi\varphi}(\hat{\varphi}_\psi)|^{1/2} \propto \left(\hat{\sigma}_\mu^2\right)^{3/2} \propto \{S^2 + n(\bar{y}-\mu)^2\}^{3/2}. \tag{14}$$

We finally develop the differential term $\mathrm{d}(\psi)$ in (9). It is crucial to explicitly take account of the recalibration T in this Jacobian. The term $\mathrm{d}\psi/\mathrm{d}\bar{\varphi}$ in w_1 (Section 4.2) satisfies

$$\frac{\mathrm{d}\psi}{\mathrm{d}\bar{\varphi}} = \frac{\mathrm{d}\psi(\varphi)}{\mathrm{d}\varphi} \cdot \frac{\mathrm{d}\varphi}{\mathrm{d}\bar{\varphi}} = \left(-\frac{1}{\varphi_2}, \frac{\varphi_1}{\varphi_2^2}\right) \cdot T^{-1} = \sigma^2(1,\mu) \cdot T^{-1},$$

which can then be normalized to the unit vector $w_1 = \{\mathrm{d}\psi/\mathrm{d}\bar{\varphi}\}/|\mathrm{d}\psi/\mathrm{d}\bar{\varphi}|$ and evaluated at $\hat{\varphi}_\psi$. The term $\mathrm{d}(T\hat{\varphi}_\psi)/\mathrm{d}\psi$ in w_2 (Section 4.2) is obtained as

$$\begin{aligned} T \cdot \frac{\mathrm{d}\hat{\varphi}_\psi}{\mathrm{d}\psi} &= T \cdot \frac{\mathrm{d}}{\mathrm{d}\mu}\left(\frac{\mu}{\hat{\sigma}_\mu^2}, -\frac{1}{\hat{\sigma}_\mu^2}\right)^\top \\ &= \frac{1}{(\hat{\sigma}_\mu^2)^2}\, T \cdot \left(\hat{\sigma}_\mu^2 + 2\mu(\bar{y}-\mu), -2(\bar{y}-\mu)\right)^\top. \end{aligned}$$

This finally leads to

$$\begin{aligned} \mathrm{d}(\psi) &= \left|\frac{\mathrm{d}\psi}{\mathrm{d}\bar{\varphi}}\right|^{-1} \frac{\mathrm{d}\psi}{\mathrm{d}\varphi} \cdot T^{-1} \cdot T \cdot \frac{\mathrm{d}\hat{\varphi}_\psi}{\mathrm{d}\psi}\, \mathrm{d}\psi \\ &= \frac{1}{|(1,\mu)\cdot T^{-1}|(\hat{\sigma}_\mu^2)^2}\, (1,\mu) \cdot \left(\hat{\sigma}_\mu^2 + 2\mu(\bar{y}-\mu), -2(\bar{y}-\mu)\right)^\top \mathrm{d}\mu \\ &= \frac{1}{|(1,\mu)\cdot T^{-1}|\hat{\sigma}_\mu^2}\, \mathrm{d}\mu, \end{aligned} \tag{15}$$

and the matrix T conveniently appears in the norm of the vector w_1 only. Using the latter along with (14), we obtain the directional Jeffreys-style prior

$$\pi_D(\mu)\, \mathrm{d}\mu \propto \{S^2 + n(\bar{y}-\mu)^2\}^{3/2}\, \mathrm{d}(\mu) \propto \frac{\{S^2 + n(\bar{y}-\mu)^2\}^{1/2}}{|(1,\mu)\cdot T^{-1}|}\, \mathrm{d}\mu;$$

the resulting posterior survivor value evaluated at μ_0 is

$$\jmath_D(\mu_0) = \int_{\mu_0} \exp\left\{\ell(\mu, \hat{\sigma}_\mu^2; \mathbf{y}^0)\right\}\, \pi_D(\mu)\, \mathrm{d}\mu.$$

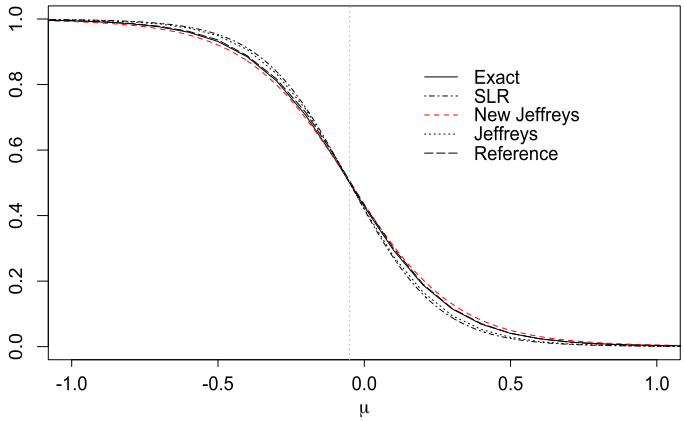

Figure 5: p-value and posterior survivor value functions for the parameter μ in the normal model; the MLE of μ is identified by a pale vertical line.

Figure 5 examines p-value and posterior survivor value functions obtained with the observed sample $\mathbf{y}^0$. The exact p-value function $p(\alpha)$ is obtained using a Student-t distribution with $n-1$ degrees of freedom and is represented on the graph by a solid line. The normal approximation for the signed likelihood root is also included (dash-dotted line). The graph features a comparison with posterior survivor values obtained under Jeffreys' prior (dotted line), the reference prior (long-dash), and the new directional Jeffreys (red dashed line). Exact and approximated p-value functions have been obtained in `R`, while the posterior survivor values (based on Jeffreys and reference) were obtained by running 200,000 iterations of a random walk Metropolis algorithm with a Gaussian proposal distribution featuring a scaling $\sigma^2 = 0.40$ (also in `R`). Posterior survivor values using the new directional Jeffreys were obtained through numerical integration. Results from the new Jeffreys-style prior are as convincing as those based on the Bayesian benchmark, the reference prior.

6.3 Curved parameter

As an example with curvature, consider a gamma model with canonical parameters $\alpha, \beta > 0$. We are interested in the variance $\psi = \alpha/\beta^2$, which is curved in terms of $\varphi = (\alpha, \beta)$, and we choose to work with the free nuisance parameter $\lambda = \beta$. The density of the model is

$$f(y; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp\{-\beta y\}, \quad y > 0,$$

with $n = 5$ observed values $\mathbf{y}^0 = (0.20, 0.45, 0.78, 1.28, 2.28)$ as used in Brazzale et al. (2007) on page 13. The maximum likelihood estimates of the canonical parameters, $\hat{\varphi} = (\hat{\alpha}, \hat{\beta})$, are the solution of the equations

$$\hat{\alpha}/\hat{\beta} = \bar{y},$$

$$D'(\hat{\alpha}) - \log \hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} \log y_i - \log \bar{y}.$$

By re-expressing the log-likelihood function in terms of interest and nuisance as $\ell(\psi, \lambda; \mathbf{y})$, we find the constrained MLE $\hat{\lambda}_\psi$ to be the solution of

$$2\psi\lambda \left[\log \lambda + \frac{1}{2} - D'(\psi\lambda^2) + \frac{1}{n}\sum_{i=1}^{n} \log y_i \right] = \bar{y}.$$

The Fisher information function in the canonical parameterization is

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

and so Jeffreys' prior $|\jmath_{\varphi\varphi}(\varphi)|^{1/2}$, which treats both parameters as of equal interest, is $\pi_J(\varphi)\mathrm{d}\varphi \propto \{\alpha D''(\alpha) - 1\}^{1/2}/\beta \ \mathrm{d}\varphi$. The reference prior for this specific context would target the interest parameter $\psi = \alpha/\beta^2$, but is not widely available is this case.

Since the model studied does not satisfy the linearity constraint, a recalibration $\bar{\varphi} = T\varphi$ of the canonical parameter is required, where T is such that $\jmath_{\varphi\varphi}(\hat{\varphi}^0) = T^\top T$. From (15) in Section 6.2, recall that this recalibration only impacts the differential $\mathrm{d}(\psi)$ through the term $|\mathrm{d}\psi/\mathrm{d}\bar{\varphi}|$. Jeffreys' prior evaluated on the profile contour, i.e. at $\hat{\varphi}_\psi = (\hat{\alpha}_\psi, \hat{\beta}_\psi) = (\psi\hat{\beta}_\psi^2, \hat{\beta}_\psi)$, is

$$|\jmath_{\bar{\varphi}\bar{\varphi}}(\hat{\bar{\varphi}}_\psi)|^{1/2} \propto |\jmath_{\varphi\varphi}(\hat{\varphi}_\psi)|^{1/2} \propto \{\hat{\alpha}_\psi D''(\hat{\alpha}_\psi) - 1\}^{1/2}/\hat{\beta}_\psi.$$

The term $\mathrm{d}\psi/\mathrm{d}\bar{\varphi}$ in w_1 of Section 4.2 is

$$\frac{\mathrm{d}\psi}{\mathrm{d}\bar{\varphi}} = \left(\frac{1}{\beta^2}, -\frac{2\alpha}{\beta^3}\right) \cdot T^{-1};$$

evaluated at $\hat{\varphi}_\psi$, it becomes $(1/\hat{\beta}_\psi^2, -2\psi/\hat{\beta}_\psi)\cdot T^{-1}$. Since we did not obtain a closed-form expression for $\hat{\varphi}_\psi$, the differential term $\mathrm{d}\hat{\varphi}_\psi/\mathrm{d}\psi$ in (9) cannot be computed explicitly; we simply use $\mathrm{d}(\psi) = |\mathrm{d}\psi/\mathrm{d}\bar{\varphi}|^{-1} \ \mathrm{d}\psi/\mathrm{d}\varphi \ \mathrm{d}\hat{\varphi}_\psi$ and numerically evaluate this expression for an appropriately small lag h by letting $\mathrm{d}\hat{\varphi}_\psi \approx \hat{\varphi}_{\psi+h} - \hat{\varphi}_\psi$. This leads to the directional Jeffreys-style prior satisfying

$$\begin{aligned} \pi_D(\psi) \ \mathrm{d}\psi &\propto \pi_J(\hat{\varphi}_\psi) \ \mathrm{d}(\psi) \\ &\propto \frac{1}{\hat{\beta}_\psi}\{\hat{\alpha}_\psi D''(\hat{\alpha}_\psi) - 1\}^{1/2} \frac{1}{|\mathrm{d}\psi/\mathrm{d}\varphi \cdot T^{-1}|} \frac{\mathrm{d}\psi}{\mathrm{d}\varphi} \cdot \mathrm{d}\hat{\varphi}_\psi \\ &\propto \frac{1}{\hat{\beta}_\psi}\{\hat{\alpha}_\psi D''(\hat{\alpha}_\psi) - 1\}^{1/2} \frac{1}{|(1, -2\psi\hat{\beta}_\psi) \cdot T^{-1}|} (1, -2\psi\hat{\beta}_\psi) \cdot \mathrm{d}(\hat{\alpha}_\psi, \hat{\beta}_\psi)^\top, \end{aligned}$$

and to a posterior survivor function

$$\jmath_D(\psi_0) = \int_{\psi_0} \exp\left\{\ell(\hat{\alpha}_\psi, \hat{\beta}_\psi; \mathbf{y}^0)\right\} \ \pi_D(\psi) \ \mathrm{d}\psi.$$

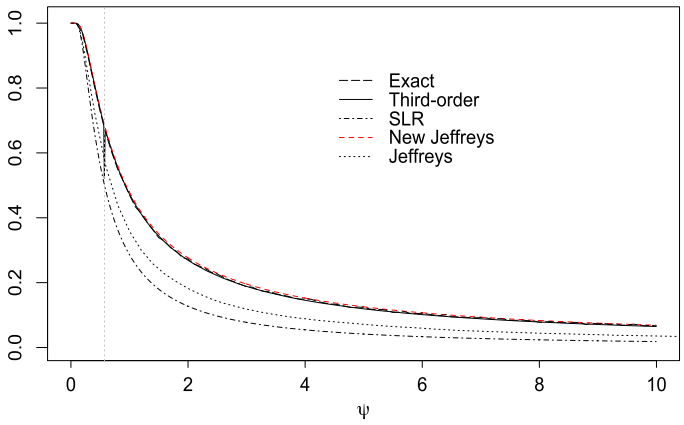

Figure 6: p-value and posterior survivor value functions for the parameter $\psi = \alpha/\beta^2$ in the gamma model; the MLE of ψ is identified by a pale vertical line.

Figure 6 compares approximations of the p-value function (SLR, third-order) and posterior survivor value functions under different priors (regular Jeffreys and new directional Jeffreys). The new directional prior is again extremely close to the third-order p-value function, while Jeffreys' prior now significantly underestimates the latter. The exact conditional p-value is obtained here by making use of MCMC on the tangent exponential model.

6.4 Behrens-Fisher problem

Consider two independent variables $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ with data $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$, where $\mathbf{y}_i$ is of size n_i from Y_i $(i = 1, 2)$. The interest parameter is $\psi = \mu_1 - \mu_2$ and we let the nuisance be $\lambda = (\mu_2, \sigma_1^2, \sigma_2^2)$; the full parameter is then $\theta = (\psi, \lambda)$. The log-likelihood function satisfies

$$\begin{aligned}\ell(\theta; \mathbf{y}) = & -\frac{n_1}{2}\log(2\pi\sigma_1^2) - \frac{n_2}{2}\log(2\pi\sigma_2^2) \\ & - \frac{1}{2\sigma_1^2}\{n_1(\bar{y}_1 - \psi - \mu_2)^2 + S_1^2\} - \frac{1}{2\sigma_2^2}\{n_2(\bar{y}_2 - \mu_2)^2 + S_2^2\},\end{aligned}$$

with $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$ and $S_i^2 = \sum_{j=1}^{n_i} y_{ij}^2 - n_i(\bar{y}_i)^2$, $i = 1, 2$. This leads to the full MLE $\hat{\theta} = (\bar{y}_1 - \bar{y}_2, \bar{y}_2, S_1^2/n_1, S_2^2/n_2)$. To obtain the constrained MLE of λ given ψ, we solve the following system of equations:

$$\begin{aligned}\hat{\mu}_2 &= \frac{n_1\hat{\sigma}_2^2(\bar{y}_1 - \psi) + n_2\hat{\sigma}_1^2\bar{y}_2}{n_1\hat{\sigma}_2^2 + n_2\hat{\sigma}_1^2}, \\ \hat{\sigma}_i^2 &= (\bar{y}_i - \psi\mathbb{1}_{(i=1)} - \hat{\mu}_2)^2 + \frac{S_i^2}{n_i}, \quad i = 1, 2;\end{aligned} \tag{16}$$

plugging $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ into $\hat{\mu}_2$, we numerically solve for $\hat{\mu}_2$ and then work backwards for the variances.

The canonical parameter of this model is $\varphi(\theta) = ((\psi+\mu_2)/\sigma_1^2, \mu_2/\sigma_2^2, 1/\sigma_1^2, 1/\sigma_2^2)$. The MLE is $\hat{\varphi} = \varphi(\hat{\theta}) = (n_1\bar{y}_1/S_1^2, n_2\bar{y}_2/S_2^2, n_1/S_1^2, n_2/S_2^2)$ and the constrained MLE given ψ is $\hat{\varphi}_\psi = ((\psi+\hat{\mu}_2)/\hat{\sigma}_1^2, \hat{\mu}_2/\hat{\sigma}_2^2, 1/\hat{\sigma}_1^2, 1/\hat{\sigma}_2^2)$, using estimates in (16). The log-likelihood function can be reexpressed as $\ell(\varphi; \mathbf{y})$, which leads to the information matrix

$$\jmath_{\varphi\varphi}(\varphi) = \begin{pmatrix} \frac{n_1}{\varphi_3} & 0 & -\frac{n_1\varphi_1}{\varphi_3^2} & 0 \\ 0 & \frac{n_2}{\varphi_4} & 0 & -\frac{n_2\varphi_2}{\varphi_4^2} \\ -\frac{n_1\varphi_1}{\varphi_3^2} & 0 & \frac{n_1}{2\varphi_3^2} + \frac{n_1\varphi_1^2}{\varphi_3^3} & 0 \\ 0 & -\frac{n_2\varphi_2}{\varphi_4^2} & 0 & \frac{n_2}{2\varphi_4^2} + \frac{n_2\varphi_2^2}{\varphi_4^3} \end{pmatrix}$$

with determinant $|\jmath_{\varphi\varphi}(\varphi)| = n_1^2 n_2^2/\{4\varphi_3^3\varphi_4^3\}$. Jeffreys' prior for this problem is therefore $\pi_J(\varphi)\mathrm{d}\varphi \propto |\jmath_{\varphi\varphi}(\varphi)|^{1/2}\mathrm{d}\varphi \propto (\varphi_3\varphi_4)^{-3/2}\mathrm{d}\varphi$, while the reference prior satisfies $\pi_R(\varphi)\mathrm{d}\varphi \propto (\varphi_3\varphi_4)^{-2}\mathrm{d}\varphi$.

We now work on finding the new prior. The interest parameter ψ is not a linear function of φ, as $\psi(\varphi) = \varphi_1/\varphi_3 - \varphi_2/\varphi_4$. We therefore need to recalibrate and work with $\bar{\varphi} = T\varphi$, where $\jmath_{\varphi\varphi}(\hat{\varphi}) = T^\top T$. This transformation only has an impact on the differential $d(\psi)$ and does not affect the term $\pi_J(\hat{\varphi}_\psi) = |\jmath_{\varphi\varphi}(\hat{\varphi}_\psi)|^{1/2}$. The differential $\mathrm{d}\psi/\mathrm{d}\bar{\varphi}$ is

$$\frac{\mathrm{d}\psi}{\mathrm{d}\bar{\varphi}} = \left(\frac{1}{\varphi_3}, -\frac{1}{\varphi_4}, -\frac{\varphi_1}{\varphi_3^2}, \frac{\varphi_2}{\varphi_4^2}\right) \cdot T^{-1},$$

which can then be normalized to the unit vector $w_1 = \{\mathrm{d}\psi/\mathrm{d}\bar{\varphi}\}/|\mathrm{d}\psi/\mathrm{d}\bar{\varphi}|$ and evaluated at $\hat{\varphi}_\psi$. Since we did not obtain a closed-form expression for $\hat{\varphi}_\psi$, the differential term $\mathrm{d}\hat{\bar{\varphi}}_\psi/\mathrm{d}\psi = T \cdot \mathrm{d}\hat{\varphi}_\psi/\mathrm{d}\psi$ in (9) cannot be computed explicitly. In that case, we simply use the differential $\mathrm{d}(\psi) = w_1 \cdot T \cdot \mathrm{d}\hat{\varphi}_\psi$, and numerically evaluate this expression by letting $\mathrm{d}\hat{\varphi}_\psi \approx \hat{\varphi}_{\psi+h} - \hat{\varphi}_\psi$ for an appropriately small lag h. This leads to the directional Jeffreys-style prior satisfying

$$\pi_D(\psi)\ \mathrm{d}\psi \propto \pi_J(\hat{\varphi}_\psi)\ \mathrm{d}(\psi) = (\hat{\sigma}_1^2\hat{\sigma}_2^2)^{3/2} \frac{1}{|\mathrm{d}\psi/\mathrm{d}\varphi\ \cdot T^{-1}|} \frac{\mathrm{d}\psi}{\mathrm{d}\varphi} \cdot \mathrm{d}\hat{\varphi}_\psi,$$

and to a posterior survivor function

$$\jmath_D(\psi_0) = \int_{\psi_0} \exp\{\ell(\hat{\varphi}_\psi; \mathbf{y})\}\ \pi_D(\psi)\ \mathrm{d}\psi.$$

Figure 7 provides a comparison of p-value and posterior survivor value functions similar to previous examples; it is based on the dataset $\mathbf{y}_1^0 = (1.02, 0.82, -0.37, 0.40, 1.29, 1.39, -0.21)$, $\mathbf{y}_2^0 = (-0.86, -2.13, -0.76, 0.60, 0.26, -0.74, 0.49)$. For the Behrens-Fisher problem, it is well-known that Jeffreys' prior leads to a second-order reproducible p-value function. Since the reference prior differs from the latter, it now over- or under-estimates the p-value, depending of the specific ψ tested. As expected, the new directional Jeffreys-based prior is extremely close to the third-order p-value, which illustrates its robustness across various contexts.

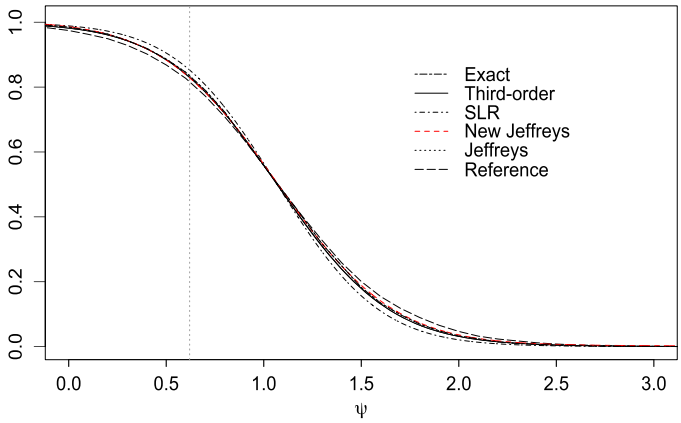

Figure 7: p-value and posterior survivor value functions for the parameter $\psi = \mu_1 - \mu_2$ in the Behrens-Fisher problem; the MLE of ψ is identified by a pale vertical line.

7 Discussion

Efron (2013) offered a classification of Bayes priors, mentioning 'genuine priors' when there is an objective random source for the actual parameter value and 'uninformative priors' for formal calculations, sometimes referred to as mathematical priors. For the non-genuine priors, Berger (2006) and Goldstein (2011) recommend unifying Bayesian and frequentist procedures, by which they mean reproducibility, that is repetition under identical conditions. Repetition reliability has had extensive discussion in the frequency literature and leads to third-order accuracy for scalar parameters with most regular models. With this as a benchmark under repetitions, we have developed a second-order accurate prior for scalar parameters and find it to be essentially Jeffreys' prior but confined to the profile contour for the scalar parameter of interest; this indicates that the ordinary use of Jeffreys does what might be viewed as a double overlapping calculation.

According to the theory exposed, Bayesian inference should then use the profile likelihood with parameter ψ, along with the new Jeffreys-based prior. Although the resulting one-dimensional posterior appears to rely on plug-in estimators, we note that it arises from usual Bayesian arguments such as marginalization. Indeed an ancillary statistic u, whose distribution is free of the nuisance parameter, was first identified; an expression for the density of this statistic was then obtained by marginalizing the joint density with respect to λ. Therefore, the issue of the theoretical developments is not to be mistaken for a deliberate plug-in approach.

Examples have been investigated under increasing complexity: linear, rotating, curved, and they clearly support the claimed second-order accuracy. Vector interest parameters, however, do not generally have repetition reliability; this was investigated by Dawid et al. (1973) as marginalization paradoxes, and by the present discussion under parameter curvature in Section 5. Second-order frequency-based p-values for vector parameters are available from Fraser et al. (2016a).

8 From exponential to general models

The results discussed in this paper were presented for regular exponential models, but they are available for quite general regular models. For this, consider an n-dimensional variable with a p-dimensional full parameter, plus continuity for parameter effects. In the simple scalar variable and parameter case, the distribution function $F(y;\theta) = z$ (say) can be inverted to give the quantile function $y = y(z;\theta)$. This allows easy simulations for the variable y using an underlying uniform distribution for z. The same is widely available for the vector variable case, by determining an $n \times p$ matrix

$$V = (v_1, \ldots, v_p) = \frac{\partial y}{\partial \theta},$$

where the differentiation is for fixed pivotal $z = z(y;\theta) = z(y^0;\hat{\theta}^0)$. Differentiating the log-model $\ell(\theta;y)$ in the directions V then gives the needed canonical parameter

$$\varphi = \frac{\partial \ell(\theta;y)}{\partial V}\Big|_{y^0},$$

which is used with an observed canonical variable $s = 0$. This leads to an exponential model using (φ, s), called the tangent exponential model, which then provides full third-order inference for the original model-data combination.

Funding

We acknowledge support from the Natural Sciences and Engineering Research Council of Canada, and the Senior Scholars Funding of York University.

References

Berger, J. (2006). "The case for objective Bayesian analysis." *Bayesian Analysis*, 1: 385–402. MR2221271. doi: https://doi.org/10.1214/06-BA115. 1, 25

Bernardo, J. M. (1979). "Reference posterior distributions for Bayesian inference." *Journal of the Royal Statistical Society. Series B*, 41: 113–147. MR0547240. 17, 19

Brazzale, A. R., Davison, A. C., and Reid, N. (2007). *Applied Asymptotics Case Studies in Small-Sample Statistics*. Cambridge, UK: Cambridge University Press. MR2342742. doi: https://doi.org/10.1017/CBO9780511611131. 21

Cakmak, S., Fraser, D. A. S., McDunnough, P., Reid, N., and Yuan, X. (1998). "Likelihood centered asymptotic model: exponential and location model versions." *Journal of Statistical Planning and Inference*, 66: 211–222. MR1614476. doi: https://doi.org/10.1016/S0378-3758(97)00085-2. 4

Cox, D. R. and Reid, N. (1987). "Parameter orthgonality and approximate conditional inference (with Discussion)." *Journal of the Royal Statistical Society. Series B*, 49: 1–39. MR0893334. 5, 12, 13

Daniels, H. E. (1954). "Saddlepoint approximations in statistics." *Annals of Mathematical Statistics*, 46: 21–31. MR0066602. doi: https://doi.org/10.1214/aoms/1177728652. 3

Datta, G. S. and Sweeting, T. J. (2005). "Probability matching priors." *Handbook of Statistics*, 25: 91–114. MR2490523. doi: https://doi.org/10.1016/S0169-7161(05)25003-4. 2

Dawid, A. P., Stone, M., and Zidek, J. V. (1973). "Marginalization paradoxes in Bayesian and structural inference." *Journal of the Royal Statistical Society. Series B*, 35: 189–233. MR0365805. 25

Efron, B. (2013). "Bayes' theorem in the 21st century." *Science*, 340: 1177–1178. MR3087705. doi: https://doi.org/10.1126/science.1236536. 1, 25

Fraser, D. (2011). "Assessing a value for an interest parameter and a definitive reference distribution." *Biometrika*. 8

Fraser, D., Reid, N., and Sartori, N. (2016a). "Accurate directional inference for vector parameters." *Biometrika*, 103(3): 625–639. MR3551788. doi: https://doi.org/10.1093/biomet/asw022. 25

Fraser, D. A. S. (2014). "Why does statistics have two theories?" In Lin, X., Genest, C., Banks, D. L., Molenberghs, G., Scott, D. W., and Wang, J.-L. (eds.), *Past, Present and Future of Statistical Science*, 237–252. Florida: CRC Press. MR3289623. doi: https://doi.org/10.1201/b16720. 12

Fraser, D. A. S. (2016). "Definitive testing of an interest parameter: Using parameter continuity." *Journal of Statistical Research*, 47: 153–165. 5

Fraser, D. A. S. (2017). "The p-value function: The core concept of modern statistical inference." *Annual Reviews in Statistics and Its Applications*, 4: 153–165. 16

Fraser, D. A. S., Bédard, M., Wong, A., Lin, W., and Fraser, A. (2016b). "Bayes, reproducibility and the quest for truth." *Statistical Science*, 31: 578–590. MR3598740. doi: https://doi.org/10.1214/16-STS573. 2, 4, 8

Fraser, D. A. S. and Reid, N. (1995). "Ancillaries and third order significance." *Utilitas Mathematica*, 47: 33–53. MR1330888. 5

Fraser, D. A. S., Reid, N., Marras, E., and Yi, G. (2010). "Default priors for Bayesian and frequentist inference." *Journal of the Royal Statistical Society. Series B*, 75: 631–654. MR2758239. doi: https://doi.org/10.1111/j.1467-9868.2010.00750.x. 2

Goldstein, M. (2011). "Subjective Bayesian analysis: Principles and practice." *Bayesian Analysis*, 26: 187–202. MR2221272. doi: https://doi.org/10.1214/06-BA116. 1, 25

Jeffreys, H. (1946). "An invariant form for the prior probability in estimation problems." *Proceedings of the Royal Society of Edinburgh. Section A*, 186: 453–461. MR0017504. doi: https://doi.org/10.1098/rspa.1946.0056. 2, 4

Lindley, D. (1975). "The future of statistics: A Bayesian 21st century." *Advances in Applied Probability*, 7: 106–115. MR0488402. 1

Severini, T. A. (2007). "Integrated likelihood functions for non-Bayesian inference."

Biometrika, 94(3): 529–542. MR2410006. doi: https://doi.org/10.1093/biomet/asm040. 13

Tibshirani, R. (1989). "Noninformative priors for one parameter of many." *Biometrika*, 76(3): 604–608. MR1040654. doi: https://doi.org/10.1093/biomet/76.3.604. 1, 12, 13

Tierney, L. and Kadane (1986). "Accurate approximations for posterior moments and marginal densities." *Journal of the American Statistical Association*, 81: 82–86. MR0830567. 13

Tierney, L., Kass, R., and Kadane, J. (1989). "Fully exponential Laplace approximations to expectations and variances of nonpositive functions." *Journal of the American Statistical Association*, 84: 710–716. MR1132586. 13

Welch, B. L. and Peers, H. W. (1963). "On formulae for confidence points based on intervals of weighted likelihoods." *Journal of the Royal Statistical Society. Series B*, 25: 318–329. MR0173309. 3, 4, 8, 9, 12