MDPI

Article

Objective Priors for Invariant $e-$Values in the Presence of Nuisance Parameters

Elena Bortolato † **and Laura Ventura *,†** 📷

Department of Statistical Sciences, University of Padova, 35121 Padova, Italy; elena.bortolato@unipd.it
* Correspondence: ventura@stat.unipd.it
† These authors contributed equally to this work.

Abstract: This paper aims to contribute to refining the *e*-values for testing precise hypotheses, especially when dealing with nuisance parameters, leveraging the effectiveness of asymptotic expansions of the posterior. The proposed approach offers the advantage of bypassing the need for elicitation of priors and reference functions for the nuisance parameters and the multidimensional integration step. For this purpose, starting from a Laplace approximation, a posterior distribution for the parameter of interest is only considered and then a suitable objective matching prior is introduced, ensuring that the posterior mode aligns with an equivariant frequentist estimator. Consequently, both Highest Probability Density credible sets and the *e*-value remain invariant. Some targeted and challenging examples are discussed.

Keywords: asymptotic expansions; adjusted score function; bias reduction; evidence; Full Bayesian Significance Test; higher-order asymptotics; matching priors; median bias reduction

1. Introduction

In this article, we discuss an objective matching prior that maintains the invariance of the posterior mode when testing specific hypotheses for parametric models, especially in the presence of nuisance parameters. These parameters are often introduced to establish flexible and realistic models, although the primary focus of inference is typically limited to a parameter of interest. The proposed approach offers the advantage of eliminating the need for eliciting information on the nuisance components and for conducting multidimensional integration, and it produces invariant $e-$values in the presence of nuisance parameters.

The parametric framework that we consider can be described as follows. Consider a random sample $y = (y_1, \ldots, y_n)$ of size n from a random variable Y with parametric model $f(y; \theta)$, indexed by a paramater θ, with $\theta \in \Theta \subseteq \mathbb{R}^d$. Given a prior $\pi(\theta)$ on θ, Bayesian inference for θ is based on the posterior density

$$\pi(\theta|y) \propto \pi(\theta)L(\theta), \tag{1}$$

where $L(\theta)$ represents the likelihood function based on $f(y; \theta)$. Interest is, in particular, in the situation in which $\theta = (\psi, \lambda)$, where ψ is a scalar parameter for which inference is required, and λ represents the remaining $(d-1)$ nuisance parameters. In such case, Bayesian inference for ψ is based on the marginal posterior density

$$\pi_m(\psi|y) = \int \pi(\psi, \lambda|y) \, d\lambda \propto \int \pi(\psi, \lambda)L(\psi, \lambda) \, d\lambda, \tag{2}$$

which for its computation requires both elicitation on the nuisance parameter λ and multidimensional integration.

Asymptotic arguments are widely used in Bayesian inference through (1) and (2), based on developments of so-called higher-order asymptotics (see, e.g., [1–3]). Indeed, the theory of asymptotic expansions provides very accurate approximations to posterior

distributions, and to various summary quantities of interest, including tail areas, credible regions and for the Full Bayesian Significance Test (see, e.g., [4,5]). Moreover, they are particularly useful for sensitivity analyses (see [6,7]) and also for the derivation of matching priors (see [8], and references therein). For instance, focusing on the presence of nuisance parameters, the Laplace approximation to (2) provides

$$\pi_m(\psi|y) \doteq \frac{1}{\sqrt{2\pi}} |j_p(\hat{\psi})|^{1/2} \exp\{\ell_p(\psi) - \ell_p(\hat{\psi})\} \frac{|j_{\lambda\lambda}(\hat{\psi},\hat{\lambda})|^{1/2}}{|j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)|^{1/2}} \frac{\pi(\psi,\hat{\lambda}_\psi)}{\pi(\hat{\psi},\hat{\lambda})}, \tag{3}$$

where $\ell_p(\psi) = \log L_p(\psi) = \log L(\psi,\hat{\lambda}_\psi)$ is the profile log-likelihood for ψ, with $\hat{\lambda}_\psi$ the constrained maximum likelihood estimate (MLE) of λ given ψ, $(\hat{\psi},\hat{\lambda})$ is the full MLE, and $j_p(\psi) = -\partial^2 \ell_p(\psi)/\partial\psi^2$ is the profile observed information. Moreover, $j_{\lambda\lambda}(\psi,\lambda)$ is the $(\lambda,\lambda)-$block of the observed information from the full log-likelihood $\ell(\psi,\lambda) = \log L(\psi,\lambda)$, and the notation $\doteq$ indicates that the approximation is accurate to order $O(n^{-3/2})$ in moderate deviation regions (see, e.g., [9], Chapter 2). One appealing feature of higher-order approximations like (3) is that they may routinely be applied in practical Bayesian inference, since they require little more than standard likelihood quantities for their implementation, and hence, they may be available at little additional computational cost over simple first-order approximations.

In the presence of nuisance parameters, starting from approximation (3), it is possible to define a general posterior distribution for ψ of the form

$$\pi^*(\psi|y) \propto \pi^*(\psi)L_p(\psi), \tag{4}$$

where $\pi^*(\psi)$ is now a prior distribution on ψ only. Bayesian inference based on pseudo-likelihood functions—i.e., functions of ψ only and of the data y with properties similar to those of a genuine likelihood function, such as the profile likelihood—have been widely used and discussed in the recent statistical literature. Moreover, it has been theoretically motivated in several papers (see, for instance, [10–12], and references therein), also focusing on the derivation of suitable objective priors. Especially when the dimension of λ is large, there are two advantages in using (4) instead of the marginal posterior distribution (2). First, the elicitation over λ is not necessary, and second, the computation of the integrals in (2) is circumvented.

Focusing on (4), in this paper, it is of interest to test the precise (or sharp) null hypothesis

$$H_0 : \psi = \psi_0 \quad \text{against} \quad H_1 : \psi \neq \psi_0 \tag{5}$$

using the measure of evidence for the Full Bayesian Significance Test (see, e.g., [4,5]). The Full Bayesian Significance Test (FBST) quantifies evidence by considering the posterior probability associated with the least probable points in the parameter space under H_0. Higher-order asymptotic computation of the FBST for precise null hypotheses in the presence of nuisance parameters has been discussed in [13].

The original measure of evidence for the FBST is not invariant under suitable transformations of the parameter, a property which has, however, been reached in the more recent definition of the $e-$value (see [14,15], and references therein). Neverthless, when working on a scalar parameter of interest, in the presence of nuisance parameters, the $e-$value is not invariant with respect to marginalizations of the nuisance parameter, and it must be used in the full dimensionality of the parameter space. This requires elicitation on the complete parameters, numerical optimization and numerical integration, that can be computationally heavy, especially when the dimension of λ is large.

The aim of this paper is to consider the $e-$value in the context of the pseudo-posterior distribution $\pi^*(\psi|y)$, suggesting in this respect a suitable objective prior $\pi^*(\psi)$ to be used in (4). More precisely, focus is on a particular matching prior, which ensures the invariance of the posterior mode of the pseudo-posterior distribution. As a consequence, Highest Probability Density credible (HPD) sets are also invariant, as well as the $e-$value.

This paper is organized as follows. Section 2 provides a short review on the FBST for testing precise null hypotheses and also illustrates asymptotic approximations for the $e-$value, extending results of [13]. Section 3 discusses the derivation of the objective matching prior for the parameter of interest only, called median matching prior, that produces invariant $e-$values. Also, several targeted and challenging examples are discussed. Finally, Section 4 closes with some concluding remarks.

2. The FBST Measure of Evidence

Suppose that we need to decide between two hypotheses: the null H_0 and the alternative H_1. The usual Bayesian testing procedure is based on the well-known Bayes factor (BF), defined as the ratio of the posterior odds to the prior odds in favor of the null hypothesis. A high BF or its logarithm suggests evidence in favor of H_0. However, it is well known that, when improper priors are used, the BF can be undetermined, and when the null hypothesis is precise (as specified in (5)), the BF can lead to the so-called Jeffreys–Lindley's paradox (see, e.g., [16]). Moreover, the BF is not calibrated, i.e., its finite sampling distribution is unknown and it may depend on the nuisance parameter.

To avoid these drawbacks, in recent years, an alternative Bayesian procedure, called FBST, has been introduced by [5] in case of sharp hypothesis H_0 identified by the null set Θ_0, a submanifold of Θ of lower dimension. The FBST quantifies evidence by considering the posterior probability associated with the least probable points in the parameter space Θ_0. When this probability is high, it favors the null hypothesis, providing a clear and interpretable measure of support for H_0 (see, e.g., [4,15,17], and references therein). The FBST is based on a specific loss function, and thus, the decision made under this procedure is the action that minimizes the corresponding posterior risk.

The FBST operates by determining the $e-$value, a representation of Bayesian evidence associated to H_0. To construct the $e-$value, the authors introduced the *posterior surprise function* and its supremum given, respectively, by

$$\pi_s(\theta|y) = \frac{\pi(\theta|y)}{r(\theta)} \quad \text{and} \quad s^* = \pi_s(\theta^*|y) = \sup_{\theta \in \Theta_0} \pi_s(\theta|y),$$

where $r(\theta)$ is a suitable *reference function* to be chosen. The surprise function was introduced in the context of statistical inference also by [18] (see [15], and references therein). Then, they introduce the *tangential set* $T_y(\theta^*)$ defined as the set of parameter values for which the posterior surprise function exceeds the supremum s^*, that is

$$T_y(\theta^*) = \{\theta \in \Theta : \pi_s(\theta|y) > s^*\}.$$

This set, often referred to as the Highest Relative Surprise Set, includes parameter values with higher surprise than those within the null set Θ_0. The $e-$value is then computed as

$$ev = 1 - \int_{T_y(\theta^*)} \pi_s(\theta|y)\, d\theta,$$

and H_0 is rejected for *small* values of ev.

The original FBST, as proposed by [5,19], relies on a flat reference function $r(\theta) \propto 1$, so that this first version involved the determination of the tangential set $T_y(\theta)$ starting only from the posterior distribution $\pi(\theta|y)$. However, this initial version lacked invariance under reparameterizations. Subsequent refinements of the FBST introduced the importance of reference density functions, making the $e-$value explicitly invariant under appropriate transformations of the parameter. Common choices for the reference function include uninformative priors, like the uniform distribution, maximum entropy densities, or Jeffreys' invariant prior. In [20], the use of the Jeffreys' prior, $\pi(\theta) \propto |i(\theta)|^{1/2}$, where $i(\theta)$ is the Fisher information derived from $L(\theta)$, is discussed as the reference function to derive invariant HPD sets and Maximum A Posteriori (MAP) estimators that are invariant under

reparameterizations. Note that the *ev* uses the full dimensionality of the parameter space. Moreover, this measure is not invariant with respect to transformations of the nuisance parameters, and the use of high posterior densities to construct credible sets may produce inconsistencies.

Concerning the asymptotic behavior of the *ev*, it can be proven that, under suitable regularity conditions as the sample size increases, with θ_0 representing the true parameter value (see [15]), it holds:

- If H_0 is false, i.e., $\theta_0 \notin H_0$, then *ev* converges in probability to 1.
- If H_0 is true, i.e., $\theta_0 \in H_0$, then, denoting by $V(c) = Pr(ev \leq c)$ the cumulative distribution function of *ev*, we have that $V(c) \approx Q(d - h, Q^{-1}(d, c))$, with $d = dim(\Theta)$, $h = dim(\Theta_0)$ and $Q(k, x)$ the cumulative chi$-$square distribution with k degrees of freedom.

In practice, the computation of *ev* is performed in two steps: (a) a numerical optimization and (b) a numerical integration. The numerical optimization step consists of finding the maximizer θ^* of $\pi_s(\theta|y)$ under the null hypothesis. The numerical integration step consists of integrating the posterior surprise function over the region where it is greater than $\pi_s(\theta^*|y)$, to obtain the $e-$value. These computational steps make the FBST a computationally intensive procedure. Despite efficient computational algorithms for local and global optimization, as well as numerical integration, obtaining precise results for hypotheses like (5) is highly demanding, especially with large nuisance parameter dimensions. Numerical integration can be tackled by resorting to higher-order tail area approximations, as reviewed in the Bayesian framework in [3]. An application of asymptotic approximation to the FBST in its first formulation, i.e., with reference function $r(\theta) \propto 1$, has been discussed in [13].

Asymptotic Approximations for the e $-$ Value

A first-order approximation for the $e-$value, when testing (5), is simply given by (see, e.g., [21,22])

$$ev \doteq 2\left(1 - \Phi\left(\left|\frac{\psi_0 - \hat{\psi}}{\sqrt{j_p(\hat{\psi})^{-1}}}\right|\right)\right), \tag{6}$$

where the symbol "$\doteq$" indicates that the approximation is accurate to $O(n^{-1/2})$, and $\Phi(\cdot)$ is the standard normal distribution function. Thus, to first-order, *ev* agrees with the $p-$value based on the profile Wald statistic

$$w_p(\psi) = \frac{(\hat{\psi} - \psi_0)}{\sqrt{j_p(\hat{\psi})^{-1}}}. \tag{7}$$

In practice, the approximation (6) of *ev* may be inaccurate, in particular when the dimension of λ is large with respect to the sample size, because it forces the marginal posterior distribution to be symmetric.

The practical computation of *ev* requires the evaluation of integrals of the marginal posterior distribution. In order to have more accurate evaluations of *ev*, it may be useful to resort to higher-order asymptotics based on tail area approximations (see, e.g., [2,3], and references therein). Indeed, the measure of evidence involves integrals of the marginal surprise posterior density $\pi_{ms}(\psi|y)$. In particular, extending the application of the tail area argument to the marginal surprise posterior density, we can derive a $O(n^{-3/2})$ approximation to the marginal surprise posterior tail area probability, given by

$$\int_{\psi_0}^{\infty} \pi_{ms}(\psi|y)\, d\psi \doteq \Phi(r_B^*(\psi_0)), \tag{8}$$

where

$$r_B^*(\psi) = r_p(\psi) + \frac{1}{r_p(\psi)} \log \frac{q_B(\psi)}{r_p(\psi)},$$

with

$$r_p(\psi) = \text{sign}(\hat{\psi} - \psi)[2(\ell_p(\hat{\psi}) - \ell_p(\psi))]^{1/2}$$

profile likelihood root and

$$q_B(\psi) = \ell_p'(\psi)|j_p(\hat{\psi})|^{-1/2} \frac{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2}}{|j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|^{1/2}} \frac{\pi(\hat{\psi}, \hat{\lambda})}{\pi(\psi, \hat{\lambda}_\psi)} \frac{r(\psi, \hat{\lambda}_\psi)}{r(\hat{\psi}, \hat{\lambda})}.$$

In the expression of $q_B(\psi)$, $\ell_p'(\psi) = \partial \ell_p(\psi)/\partial \psi$ is the profile score function.

Using the tail area approximation (8), a third-order approximation of the measure of evidence *ev* can be derived. The approximation, assuming without loss of generality that ψ_0 is smaller than the MAP of $\pi_{ms}(\psi|y)$, is given by

$$ev(\psi) \doteq 1 - \Phi(r_B^*(\psi_0)) + \Phi(r_B^*(\psi_0^*)), \tag{9}$$

with ψ_0^* the value of the parameter such that $\pi_{ms}(\psi_0^*|y) = \pi_{ms}(\psi_0|y)$. Note that

$$\Phi(r_B^*(\psi_0)) - \Phi(r_B^*(\psi_0^*)) \doteq \int_{\psi_0^*}^{\psi_0} \pi_{ms}(\psi|y) \, d\psi = 1 - ev$$

in (9) gives the posterior probability of the HPD credible interval (ψ_0, ψ_0^*). Note also that the higher-order approximation (9) does not call for any condition on the prior $\pi(\psi, \lambda)$, i.e., it can be also improper. Finally, when $\pi_{ms}(\psi|y)$ is symmetric, Equation (9) reduces to $ev \doteq 2(1 - \Phi(r_B^*(\psi_0)))$.

While tail area approximations require little more than standard likelihood quantities for their implementation and, in this respect, they are available at little additional computational cost over the first-order approximation, they require elicitation on the complete parameter θ and to choose the reference function $r(\theta)$.

3. An Invariant Objective Prior

The aim of this section is to derive a default prior $\pi^*(\psi)$ to be used in (4). To this end, following [8], we use the shrinkage argument, which is a crucial procedure in the development of matching priors, i.e., priors that ensure, up to the desired order of asymptotics, an agreement between Bayesian and frequentist procedures. Examples of matching priors are (see [8]) for posterior quantiles, for credible regions and for prediction. Here, we focus on a specific matching prior that ensures the invariance of the posterior mode in the posterior distribution (4). As a consequence, the invariance extends to HPDs, as well as the $e-$value, achieved incorporating the reference function within the prior.

The proposed choice of the prior $\pi^*(\psi)$, which makes the MAP and thus also HPDs and the $e-$value invariant under 1-1 reparameterization, will depend on the log-likelihood $\ell(\theta)$ and on its derivatives. In regular parametric estimation problems, both the MLE and the score-estimating function exhibit an asymptotically symmetric distribution centered at the true parameter value and at zero, respectively. However, these asymptotic behaviors may poorly reflect exact sampling distributions, particularly in cases with small or moderate sample information, sparse data, or complex models. Several proposals have been developed to correct the estimate or the estimating function. Most available methods are aimed at approximate bias adjustment, either of the MLE or of the profile score function, also when nuisance parameters are present (see [23] for a review of bias reduction for the MLE and [24] and subsequent literature for bias correction of the profile score). Lack of equivariance impacts the so-called implicit bias reduction methods, which achieve first-order bias correction by modifying the score equation (see [23,25]). To avoid this drawback, in this paper we focus on the median modification of the score, or profile score equation,

whose solution respects equivariance under monotone reparameterizations ([26]). Similar to Firth's implicit method ([25]), the median modification of the score, or profile score, does not rely on finiteness of the MLE, thereby effectively preventing infinite estimates.

In practice, to derive the median matching prior $\pi^*(\psi)$, we impose that the MAP of $\pi^*(\psi|y)$ coincides with a refined version of the MLE, obtained as the solution of the median modified score function ([26]). To introduce this new invariant prior, we initially explore the scenario without nuisance parameters and then the situation in which nuisance parameters are present.

3.1. No Nuisance Parameters

Let us explore first the scenario where θ is scalar. In order to obtain median bias reduction of the MLE, it is possible to resort to a modified version of the score function of the form

$$t(\theta) = \ell_\theta(\theta) + m(\theta), \tag{10}$$

where $\ell_\theta(\theta) = \ell_\theta(\theta; y) = \partial \ell(\theta; y)/\partial \theta$ is the score function, and $m(\theta)$ is a suitable correction term of order $O(1)$. In particular, the median modified score function assumes for $m(\theta)$ the expression

$$m(\theta) = -\frac{E(\ell_\theta(\theta)^3)}{6\, i(\theta)}.$$

The solution $\tilde{\theta}$ to the equation $t(\theta) = 0$ not only upholds equivariance under component-wise monotone reparameterizations, but also approximates median unbiasedness ([26]). Note that likelihood inference based on (10) does not depend explicitly on the MLE. Indeed, the modified score function has been found to overcome infinite estimate problems. Likewise, the MLE and also $\tilde{\theta}$ is asymptotically $N(\theta, i(\theta)^{-1})$, so that the Wald-type statistics only differ in location.

Since Bayes' theorem is a statement of adittivity on the log scale $\log \pi(\theta|y) = \log \pi(\theta) + \log L(\theta) +$ constant, we observe that in the Bayesian framework, $m(\theta)$ can be interpreted as the derivative of the logarithm of a prior, that is, $m(\theta) = \partial \log \pi(\theta)/\partial \theta$. We are thus looking for a matching prior $\pi^*(\theta)$ such that

$$\frac{\partial \log \pi^*(\theta)}{\partial \theta} = -\frac{E(\ell_\theta(\theta)^3)}{6\, i(\theta)}.$$

In the scalar parameter case, it is straightforward to show that the proposed *median matching prior* takes the form

$$
\begin{aligned}
\pi^*(\theta) \quad &\propto \quad \exp\left(-\frac{1}{6}\int i(\theta)^{-1} E(\ell_\theta(\theta)^3)\, d\theta\right) \\
&\propto \quad \exp\left(\frac{1}{6}\int i(\theta)^{-1}(3E(\ell_{\theta\theta}(\theta)\ell_\theta(\theta)) + E(\ell_{\theta\theta\theta}(\theta)))\, d\theta\right),
\end{aligned}
$$

with $\ell_{\theta\theta}(\theta) = \partial\ell_\theta(\theta)/\partial\theta$ and $\ell_{\theta\theta\theta}(\theta) = \partial\ell_{\theta\theta}(\theta)/\partial\theta$, where the second expression for $\pi^*(\theta)$ follows from the Bartlett's identities. The posterior based on the median matching prior is thus

$$\pi^*(\theta|y) \propto \exp\left(\ell(\theta) - \frac{1}{6}\int i(\theta)^{-1} E(\ell_\theta(\theta)^3)\, d\theta\right).$$

A first-order approximation for the $e-$value, when testing $H_0 : \theta = \theta_0$, is simply given by

$$ev \doteq 2\left(1 - \Phi\left(\left|\frac{\theta_0 - \tilde{\theta}}{\sqrt{i(\theta_0)^{-1}}}\right|\right)\right), \tag{11}$$

which differs in location with respect to the classical first-order approximation for the $e-$value based on the MLE. A second approximation for the $e-$value, when testing H_0 :

$\theta = \theta_0$, can be obtained from the asymptotic distribution of the modified score function (10), that is

$$ev \doteq 2\left(1 - \Phi\left(\left|\frac{t(\theta_0)}{\sqrt{i(\theta_0)}}\right|\right)\right). \tag{12}$$

Although the first-order equivalence between (11) and (12), note that (11) is based on an easily understandable comparison between estimated value and hypotetical value, taking estimation error into account, and is widely used in applications but does not satisfy the principle of parameterization invariance. On the other hand, $t(\theta)/\sqrt{i(\theta)}$ is parameterization invariant.

Note that, when using a *predictive matching prior*, i.e., a prior ensuring asymptotic equivalence of higher-order frequentist and Bayesian predictive densities (see, e.g., [8]), the term $m(\theta)$ in (10) corresponds to the Firth's adjustment ([25])

$$m_F(\theta) = -\frac{(E(\ell_\theta(\theta)^3) + E(\ell_{\theta\theta}(\theta)\ell_\theta(\theta)))}{2i(\theta)}.$$

In view of this, for general regular models, Firth's estimate coincides with the mode of the posterior distribution obtained using the default predictive matching prior. However, lack of invariance affects this kind of adjustment ([23]), unless dealing with linear transformations.

Example 1 (One parameter exponential family). *For a one-parameter exponential family with canonical parameter θ, i.e., with density*

$$f(y;\theta) = \exp\{\theta a(y) - K(\theta)\}b(y),$$

the median modified score function has the form

$$t(\theta) = \ell_\theta(\theta) + \frac{K_{\theta\theta\theta}}{6K_{\theta\theta}},$$

where $K_{\theta\theta\theta} = \partial^3 K(\theta)/\partial\theta^3$ and $K_{\theta\theta} = \partial^2 K(\theta)/\partial\theta^2 = i(\theta)$. In this parameterization, $t(\theta)$ can be seen as the first derivative of the log-posterior

$$\log \pi(\theta|y) = \ell(\theta) + \log i(\theta)/6.$$

On the other hand, Firth's modified score takes the form $t_F(\theta) = \ell_\theta(\theta) + K_{\theta\theta\theta}/(2K_{\theta\theta})$. The effect of the median modification is to consider the median matching prior $\pi^*(\theta) \propto i(\theta)^{1/6}$, while $t_F(\theta)$ implies a Jeffreys' prior $\pi_J(\theta) \propto i(\theta)^{1/2}$. Note that, for a one-parameter exponential family with canonical parameter θ, both $\pi^*(\theta)$ and $\pi_J(\theta)$ belong to the family of invariant priors discussed in [27,28].

Example 2 (Scale model). *Consider the scale model $f(y;\theta) = (1/\theta)p_0(y/\theta)$, where $p_0(\cdot)$ is a given function. Let $g(\cdot) = -\log p_0(\cdot)$. We have $E(\ell_\theta^3) = c_1/\theta^3$, $E(\ell_\theta^3\ell_\theta) = c_2/\theta^3$ and $i(\theta) = c_3/\theta^2$, with $c_1 = \int(y^3 g'''(y) + 6y^2 g''(y) + 6y g'(y) - 2)p_0(y)dy$, $c_2 = \int(3y g'(y) + y^2 g''(y) - 2y^2 g'(y)^2 - y^3 g'(y)g''(y) - 1)p_0(y)dy$ and $c_3 = \int(2y g'(y) + y^2 g''(y) - 1)p_0(y)dy$. The median matching prior is thus $\pi^*(\theta) \propto \theta^{-c_1/6c_3}$, while the Jeffreys' prior for a one-parameter scale model is $\pi_J(\theta) \propto \theta^{-1}$ and the prior associated to the Firth's adjustment is $\pi_F(\theta) \propto \theta^{-(c_1+c_2)/2c_3}$.*

Example 3 (Skew–normal distribution). *Consider a skew-normal distribution with shape parameter $\theta \in \mathbb{R}$, with density $f(y;\theta) = 2\phi(y)\Phi(y\theta)$, where $\phi(\cdot)$ is the standard normal density function. The median correction term for the score function associated to the median matching prior is (see [26,29])*

$$m(\theta) = \frac{E(y^3 \phi(y\theta)^3/\Phi(y\theta)^3)}{6E(y^2 \phi(y\theta)^2/\Phi(y\theta)^2)}.$$

Numerical integration must be performed to obtain the expected values involved in $m(\theta)$.

In order to illustrate the proposed prior, we consider draws from the skew–normal model with true parameter $\theta_0 = 3$ and increasing sample sizes $n = 20, 30, 50, 200$ (Figure 1). The posterior distributions are obtained with the method by [30], i.e., drawing 10^5 values and accepting the best 5%. The $e-$values associated to the null (true) hypothesis $H_0 : \theta = 3$ and the (false) hypothesis $H_0 : \theta = 4$ are reported in Table 1. For comparison, the Jeffreys' prior ([31]), the predictive matching prior ([29]) and the flat prior, with uniform reference function, are also considered. Progressive agreement among evidence values obtained with the proposed median matching prior and the other priors for larger sample size is shown. Also, as expected, when progressively increasing n, the evidence values indicate agreement with the true hypothesis $H_0 : \theta_0 = 3$ and disagreement with $H_0 : \theta_0 = 4$ for all the priors used. Anyway, note that the posterior distribution obtained with a flat prior and a uniform reference function is proportional to the likelihood function that can be monotone. In view of this, while the MAPs of the posterior based on the default priors are always finite, in some samples the MAP of the posterior with the non-informative prior may be infinite. An example of this effect is illustrated in Figure 2.
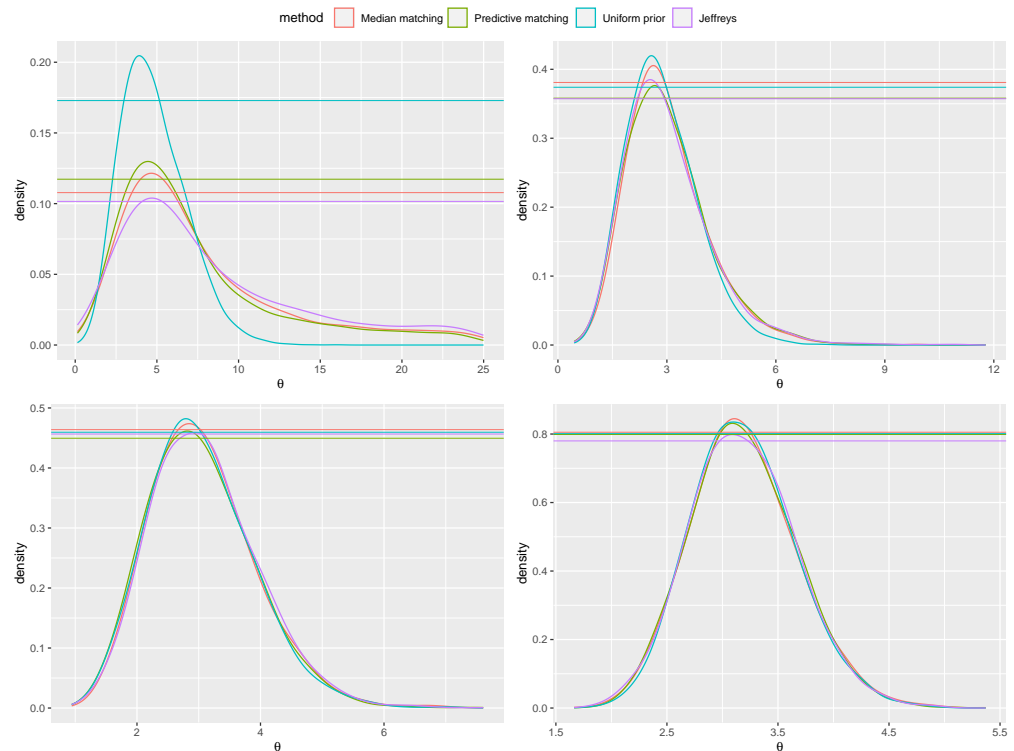

Figure 1. Inference for the scalar parameter θ of the skew-normal model with sample sizes $n = 20, 30, 50, 200$ (**top-left**, **top-right**, **bottom-left** and **bottom-right** panels, respectively). The red line is used for the posterior obtained from the median matching prior, the green one for the predictive matching prior, the violet one for the Jeffreys' prior and the blue one from an improper flat prior. The horizontal lines identify the corresponding tangential sets associated to the hypothesis $H_0 : \theta = 3$.

Table 1. Skew-normal: $e-$values associated to the hypotheses $H_0 : \theta = 3$ and $H_0 : \theta = 4$.

Hypothesis H_0	n	Flat Prior	Median Matching Prior	Predictive Matching Prior	Jeffreys' Prior
$\theta = 3$	20	0.59	0.62	0.56	0.65
	30	0.65	0.70	0.73	0.64
	50	0.81	0.84	0.82	0.91
	200	0.79	0.79	0.81	0.82
$\theta = 4$	20	0.82	0.91	0.98	0.84
	30	0.17	0.22	0.22	0.22
	50	0.20	0.20	0.20	0.21
	200	0.07	0.08	0.08	0.09
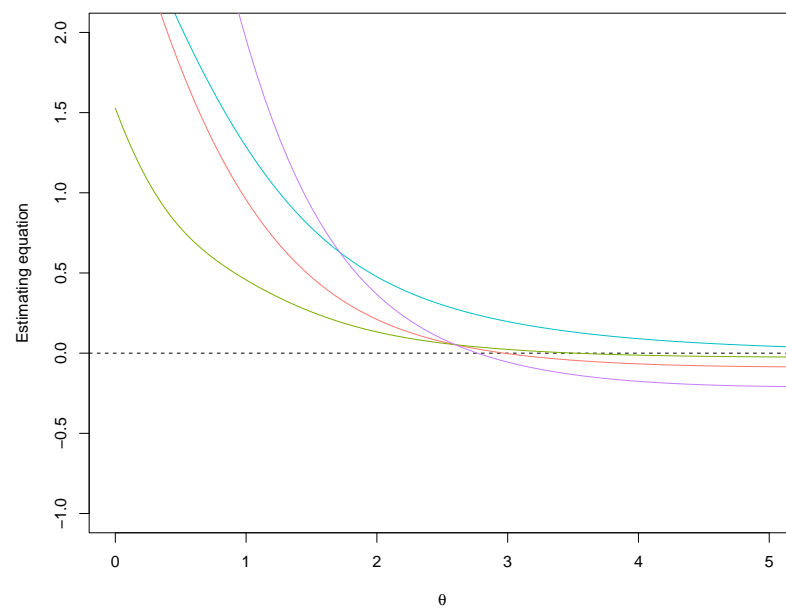
Figure 2. Skew-normal model: An example of $\partial \log \pi(\theta|y)/\partial\theta$ (estimating equation) with a flat prior (blue line), the median matching prior (red line) the predictive matching prior (green line) and the Jeffreys' prior (violet line) in a sample where all the observations are positive.

The properties of first-order approximations of the $e-$values have been investigated by a simulation study, with sample sizes $n = 20, 30, 50, 200$. Results are displayed in Figure 3. Distributions of the $e-$value from the posterior based on the median matching prior are better, both for small and moderate sample sizes, in terms of convergence to the uniform distribution. Moreover, score-type $e-$values (12) are also preferable over Wald-type $e-$values (11). For the results with the posterior distribution obtained with a flat prior, we found $4.3\%, 4.2\%, 0.9\%, 0\%$ of infinite estimates for the sample sizes considered in the simulation study, and in these cases the $e-$value was considered as 0.
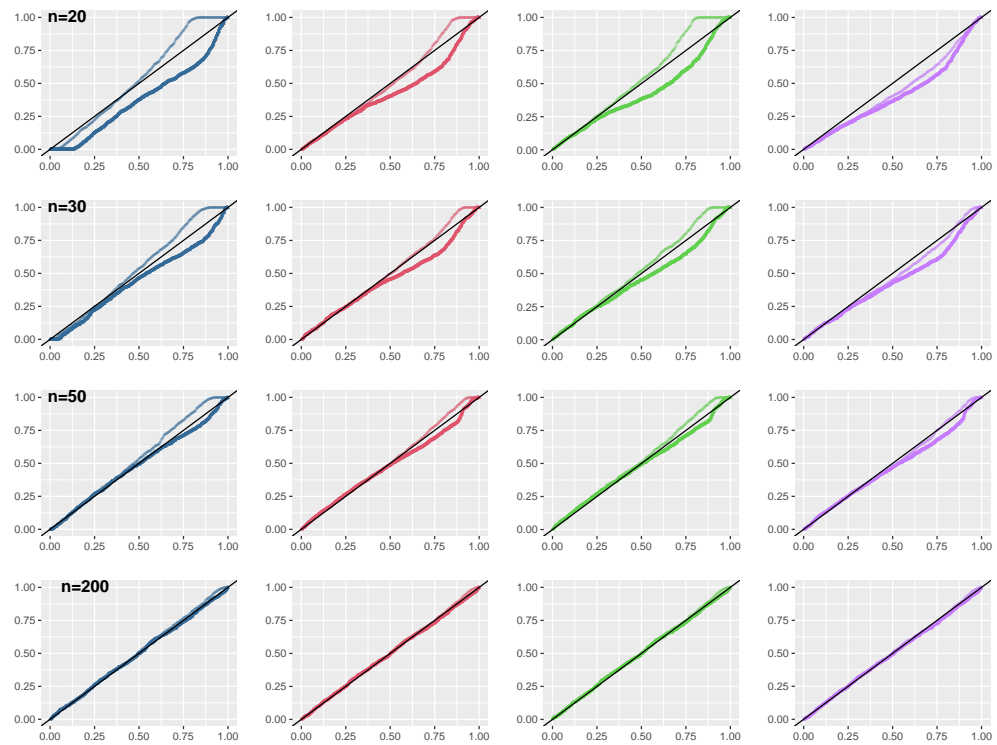
Figure 3. Skew-normal model: Distributions of $e-$values from a simulation study under the null hypothesis $H_0 : \theta = 3$, using a flat prior (blue line), the median matching prior (red line), the predictive matching prior (green line), and the Jeffreys' prior (violet line). The darker line is used for the approximation (11), while the lighter is for that based on (12).

3.2. Presence of Nuisance Parameters

In the presence of nuisance parameters, in order to obtain median bias reduction of the MLE, it is possible to resort to a modified version of the profile score function of the form

$$t_p(\psi) = \ell_p'(\psi) + m(\psi, \hat{\lambda}_\psi), \tag{13}$$

where $m(\psi, \lambda)$ is a suitable correction term of order $O(1)$. In particular, for the median modified profile score function, the adjustment $m(\psi, \lambda)$ assumes the expression

$$m(\psi, \lambda) = -\kappa_{1\psi} + \frac{\kappa_{3\psi}}{6\kappa_{2\psi}},$$

where $\kappa_{1\psi}$, $\kappa_{2\psi}$ and $\kappa_{3\psi}$ are the first three cumulants of $\ell_p'(\psi)$ (see [26], Section 2.2, for their expression). For the estimator $\tilde{\psi}_p$, defined as the solution of $t_p(\psi) = 0$, parameterization equivariance holds under interest respecting reparameterizations ([26]).

Note that, also in the context of nuisance parameters, we are in the situation in which the proposed prior $\pi^*(\psi)$ is known through its first derivative; this is typically the situation with matching priors (see, e.g., [8]). Since the parameter of interest is scalar, the posterior based on the median matching prior can be written as

$$\pi^*(\psi|y) \propto \exp\left(\ell_p(\psi) + \int m(\psi, \hat{\lambda}_\psi) \, d\psi \right). \tag{14}$$

A simple analytical way of approximating to first-order the posterior distribution (14) based on the median matching prior is to resort to a quadratic form of $t_p(\psi)$. In particular, the approximate posterior distribution for ψ takes the form

$$\pi^*(\psi|y) \dot\propto \exp\left(-\frac{1}{2}s_p(\psi; y)\right), \tag{15}$$

where $s_p(\psi) = t_p(\psi)^2 j_p(\psi)^{-1}$ is a Rao score-type statistic based on (13), and the symbol "$\dot\propto$" means asymptotic proportionality to first-order. In this case, a first-order approximation of the $e-$value, when testing $H_0 : \psi = \psi_0$, is given by

$$ev \doteq 2\left(1 - \Phi\left(\left|\frac{t_p(\psi_0)}{\sqrt{j_p(\psi_0)}}\right|\right)\right). \tag{16}$$

In this case, an higher-order approximation via (3) would be impractical since a closed-form prior is not available. As an alternative, simulation-based approaches may be used to derive the implied posterior distribution (14) based on the median matching prior. The first one relies on Approximate Bayesian Computation (ABC) techniques, using $\tilde\psi_p$ or the modified profile score function $t_p(\psi)$ as summary statistics; see [32] for the modification of the algorithm of [30] by using a profile estimating equation. This first method introduces an approximation at the level of the posterior estimation. The second one still relies on (13) but considers use of Manifold MCMC methods (see ,e.g., [33]) to conditioning exactly on the profile equation and not up to a tolerance level, as in ABC (see [34,35]). The algorithm moves on the constrained space $\{(y, \psi) \in \mathcal{Y} \times \ominus | t_p(\tilde\psi_p) = 0\}$, where $\tilde\psi_p$ is the solution of the estimating equation on the original data. For the latter method, we need minimal regularity assumptions on $m(\psi, \lambda)$, which is assumed to be continuous, differentiable and available in closed form expression. Note, for instance, that in the skew-normal example in Section 3.1 these conditions are not met.

Example 4 (**Exponential family**). *If $f(y; \theta)$ is an exponential family of order d with canonical parameter (ψ, λ), i.e., $f(y; \psi, \lambda) = \exp\{\psi t(y) + \lambda^T s(y) - K(\psi, \lambda)\}h(y)$, quantities involved in $m(\psi, \lambda)$ are simply obtained from derivatives of $K(\psi, \lambda)$ ([26]). Note that, in this framework, $\ell'_p(\psi) - \kappa_{1\psi}$ is an approximation with error of order $O(n^{-1})$ of the score for ψ in the conditional model given $s(y)$ (see e.g., [36], Section 10.2). Then, in the continuous case, the MAP $\tilde\psi_p$ is an approximation of the optimal conditional median unbiased estimator, and $\pi^*(\psi|y)$ is related to the conditional likelihood for ψ given by $L_c(\psi) = \exp(\psi t(y) - K_s(\psi))$; see [37] for a Bayesian interpretation of such pseudo-likelihoods.*

Example 5 (**Multivariate regression model**). *Consider a regression model of the form*

$$Y_{ij} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, 2,$$

where it is assumed that $\epsilon_i \sim N_2(0, \Sigma)$, with $\Sigma = \sigma^2\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ positive definite matrix, and $(\beta_0, \beta_1, \beta_2, \sigma^2, \rho)$ are unknown parameters. This model is widely used for instance in time series analysis, in which as regression covariates the past of the observables y are used. We focus on the problem of testing hypothesis on the correlation coefficient ρ.

Consider a draw with true parameter $\rho_0 = 0.95$ and $n = 20$. For obtaining the proposed posterior (14) for ρ, we first compute the MAPs with the median matching prior and also, for comparison, with the predictive matching prior, which are, respectively, 0.953 and 0.92. Note that the expression of the predictive matching prior for (ψ, λ) corresponds to the Firth's adjustment to the score function. The expressions of the modified profile estimating functions $t_p(\psi)$ and $t_F(\psi)$ are obtained from [38] and are available in closed form expressions. Hence, the Manifold MCMC

method can be used to obtain the implied posteriors, whose approximation is comparable to that of any MCMC sampler. In particular we used 20,000 iterations.

We compare the posterior distribution based on the proposed median matching prior, with those obtained with the predictive matching prior and with an inverse-Wishart prior for the covariance matrix Σ with one degree of freedom and identity position, and uniform prior on the regression parameters. The posterior distributions are displayed in Figure 4. The hypothesis of interest is $H_0 : \rho = 0.9$, and a smaller e-value indicating disagreement with the hypothesis should be preferable. The e-values are 0.25 with the median matching prior, 0.36 for the predictive matching prior and 0.60 with the inverse-Wishart prior. Note that the e-value based on the inverse-Wishart prior involves the constrained maximization and multidimensional integration and thus is not directly readable in Figure 4. Indeed, one crucial difference is that the original e-value formulation links the evidence of the null hypothesis to the evidence of a more refined hypothesis, choosing the MAP under the null hypothesis for all the nuisance parameters, while in the alternative (tangential) set, all values are used, and integration is performed on the full dimensionality of the space. On the contrary, in the proposed posterior based on the median matching prior, the maximizer of nuisance parameters are taken both in the null and non-null sets.

Finally, for the posterior based on the inverse-Wishart prior, we also computed the e-value based on the high-order tail area approximation (9) of the marginal surprise posterior, which is equal to 0.27. This procedure still avoids the multidimensional integration, but the result is not invariant to changes in parametrization.
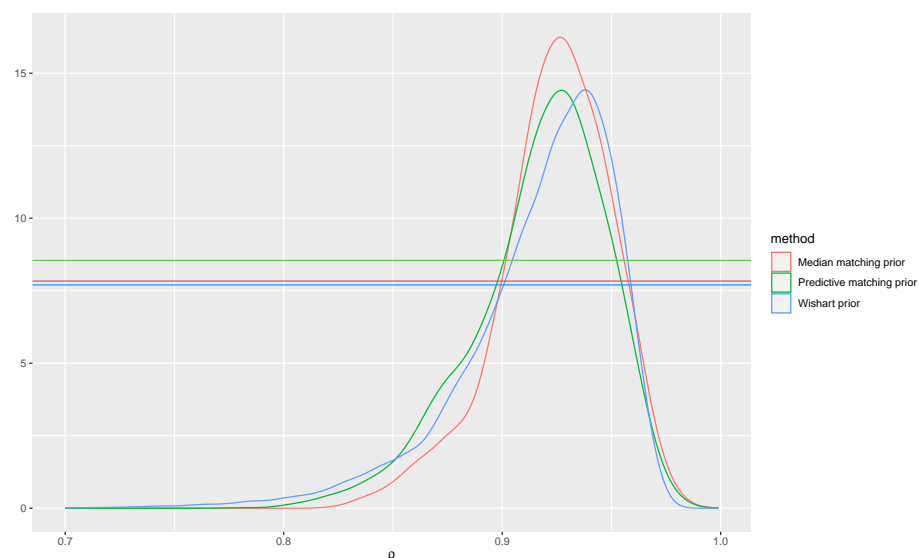
Figure 4. Posterior distributions for the correlation parameter ρ of the bivariate regression model obtained from MCMC draws and the three different priors.

Example 6 (Logistic regression model). *Let y_i, $i = 1, \ldots, n$, be independent realizations of binary random variables with probability π_i, where $\log(\pi_i/(1 - \pi_i)) = \eta_i = x_i\beta$ and $x_i = (x_{i1}, \ldots, x_{ip})$ is a row vector of covariates. We assume that a generic scalar component of β is of interest, and we treat the remaining components as nuisance parameters.*

As an example, we consider the endometrial cancer grade *dataset analyzed, among others, in [39]. The aim of the clinical study was to evaluate the relationship between the histology of the endometrium (HG), the binary response variable, of $n = 79$ patients and three risk factors: 1. Neovascularization (NV), that indicates the presence or extent of new blood vessel formation; 2. Pulsatility Index (PI), that measures blood flow resistance in the endometrium; 3. Endometrium Height (EH), that indicates the thickness or height of the endometrium. A logistic model for HG, including an intercept and using all the covariates (NV, PI, EH), has been fitted, but a maximum likelihood leads to the infinite MLE of the coefficient β_2 related to NV, due to quasi-complete separation. This phenomenon prohibits the use of diffuse priors for β_2, since the*

corresponding posterior would not concentrate. Moreover, the $e-$value with non-informative priors cannot be obtained also for any hypothesis concerning parameters different from β_2.

If we consider β_2 as the parameter of interest, while the remaining regression coefficients are treated as nuisance parameters, the analysis with the median matching prior allows us to obtain a global proper posterior, with MAP equal to 3.86, open to interpretation both in the original scale and in terms of odds ratios. Similarly, the posterior based on the predictive matching prior, which in this model coincides with Jeffreys' prior $\pi(\beta) \propto |i(\beta)|^{1/2}$ is proper, with the MAP set at 2.92. The latter suffers from a lack of interpretability on different scales, since a different parametrization in the estimation phase would affect the results.

If we consider β_3 as the parameter of interest, related to the risk factor PI, the MAPs are -0.038 when using the median matching prior and -0.035 when using the predictive matching prior. The $e-$values for the hypothesis $\mathcal{H}_0 : \beta_3 = 0$ are 0.60 and 0.55, respectively (see Figure 5). Likewise, the interpretation of e-values remains consistent and independent of parametrization solely in the first case.
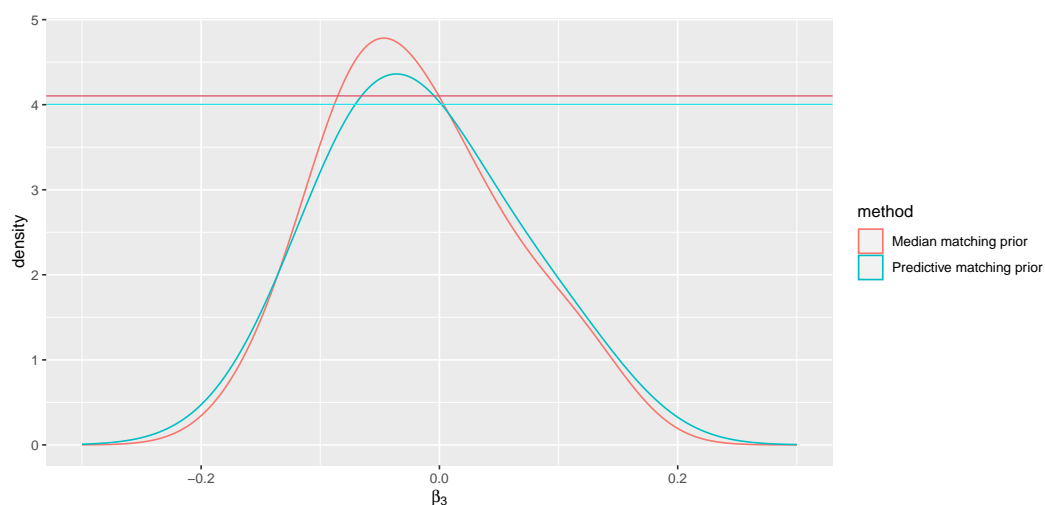
Figure 5. Median matching posterior distribution and predictive matching posterior distributions for β_3 in the logistic regression model.

4. Conclusions

Although (14) cannot always be considered as orthodox in a Bayesian setting, the use of alternative likelihoods is nowadays widely shared, and several papers focus on the Bayesian application of some well-known pseudo-likelihoods. In particular, the proposed posterior $\pi^*(\psi|y)$ has the advantages of avoiding the elicitation on the nuisance parameter λ and of the computation of multidimensional integrations. Moreover, it provides invariant MAPs, HPDs and $e-$values, without the adoption of a reference function. Finally, we remark that frequentist properties of the MAP of the posterior based off the proposed median matching prior in comparison with the MAP of the posterior based of the predictive matching prior have been investigated in [26,38] for some of the examples discussed in this paper.

For inference on a full vector parameter θ, with $d > 1$ components, a direct extension of the rationale leading to (10) does not seem to be practicable due to lack of a manageable definition of the multivariate median. Actually, in [26,40], it is shown how the method can be extended to a vector parameter of interest in the presence of nuisance parameters by simultaneously solving median bias-corrected score equations for all parameter components. This leads to componentwise third-order median unbiasedness and parameterization equivariance. Moreover, the use of default priors involving all parameter components, also the nuisance, becomes necessary to regularize likelihoods in case of monotonicity. We note that among the possible objective priors that ensures invariance of the posterior, we did not focus on the Jeffreys' in the multidimensional case, since it often exhibits poor convergence

properties. Conversely, the default matching priors considered in this paper are easily generalizable to the multidimensional case [40] preserving good convergence properties.

As a final remark, we highlight that this paper opens several topics of future research. In particular, from a computational point of view, it could be of interest:

- To develop a library of computational routines exploring the methods proposed in this paper for a wide range of statistical models of interest;
- To develop semi-automated procedures for further expanding this library, as is done for point estimation for Generalized Linear Models in the R package brglm2 [41].

Moreover, from a theoretical point of view, it could be of interest:

- To further explore the theoretical connections between the $e-$value invariance properties and matching priors;
- To explore the existence of similar connections in other classes of pseudo-likelihoods, in particular in the context of empirical and profile empirical likelihoods, with a large number of nuisance parameters ([42]),
- To apply and extend the methodology to consider other objective priors used in Bayesian inference, such as those obtained from scoring rules, as proposed in [10], that are expressed as solutions of differential equations.

Author Contributions: Conceptualization, E.B. and L.V.; methodology, E.B. and L.V.; software, E.B.; validation, E.B. and L.V.; formal analysis, E.B. and L.V.; investigation, E.B. and L.V.; resources, E.B. and L.V.; data curation, E.B.; writing—original draft preparation, E.B. and L.V.; writing—review and editing, E.B. and L.V.; visualization, E.B. and L.V.; supervision, E.B. and L.V.; project administration, E.B. and L.V.; funding acquisition, L.V. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data and code used in the examples are available in https://github.com/elenabortolato/invariant_evalues (accessed on 28 November 2023).

Abbreviations

The following abbreviations are used in this manuscript:

ABC	Approximate Bayesian Computation
BF	Bayes Factor
FBST	Full Bayesian Significance Test
HPD	Highest Probability Density
MAP	Maximum A Posteriori
MLE	Maximum Likelihood Estimator

References

1. Brazzale, A.R.; Davison, A.C.; Reid, N. *Applied Asymptotics. Case-Studies in Small Sample Statistics*; Cambridge University Press: Cambridge, UK, 2007.
2. Reid, N. The 2000 Wald memorial lectures: Asymptotics and the theory of inference. *Ann. Stat.* **2003**, *31*, 1695–1731.
3. Ventura, L.; Reid, N. Approximate Bayesian computation with modified loglikelihood ratios. *Metron* **2014**, *7*, 231–245.
4. Madruga, M.; Pereira, C.; Stern, J. Bayesian evidence test for precise hypotheses. *J. Stat. Plan. Inference* **2003**, *117*, 185–198.
5. Pereira, C.; Stern, J.M. Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy* **1999**, *1*, 99–110.
6. Kass, R.E.; Tierney, L.; Kadane, J. Approximate methods for assessing influence and sensitivity in Bayesian analysis. *Biometrika* **1989**, *76*, 663–674.
7. Reid, N.; Sun, Y. Assessing sensitivity to priors using higher order approximations. *Commun. Stat. Theory Methods* **2010**, *39*, 1373–1386.
8. Datta, G.S.; Mukerjee, R. *Probability Matching Priors: Higher-Order Asymptotics*; Lecture Notes in Statistics; Springer: Berlin/Heidelberg, Germany, 2004.
9. Severini, T.A. *Likelihood Methods in Statistics*; Oxford University Press: Oxford, UK, 2000.

10. Leisen, F.; Villa, C.; Walker, S.G. On a class of objective priors from scoring rules (with discussion). *Bayesian Anal.* **2020**, *15*, 1345–1423.

11. Miller, J.W. Asymptotic normality, concentration, and coverage of generalized posteriors. *J. Mach. Learn. Res.* **2021**, *22*, 1.

12. Ventura, L.; Racugno, W. Pseudo-likelihoods for Bayesian inference. In *Topics on Methodological and Applied Statistical Inference, Series Studies in Theoretical and Applied Statistics*; Springer: Cham, Switzerland, 2016; pp. 205–220, ISBN: 978-3-319-44093-4.

13. Cabras, S.; Racugno, W.; Ventura, L. Higher-order asymptotic computation of Bayesian significance tests for precise null hypotheses in the presence of nuisance parameters. *J. Stat. Comput. Simul.* **2015**, *85*, 2989–3001.

14. Diniz, M.B.; Pereira, C.; Stern, J.M. Cointegration and unit root tests: A fully Bayesian approach. *Entropy* **2020** *22*, 968.

15. Pereira, C.; Stern, J.M. The $e-$value: A fully Bayesian significance measure for precise statistical hypotheses and its research program. *Sao Paulo J. Math. Sci.* **2022**, *16*, 566–584.

16. Robert, C.P. On the Jeffreys Lindley paradox. *Philos. Sci.* **2014**, *81*, 216–232.

17. Madruga, M.R.; Esteves, L.G.; Wechsler, S. On the bayesianity of pereira-stern tests. *Test* **2001**, *10*, 291–299.

18. Evans, M. Bayesian inference procedures derived via the concept of relative surprise. *Commun. Statist. – Theory and Methods* **1997**, *26*, 1125–1143.

19. Pereira, C.; Stern, J.M. Model Selection: Full Bayesian Approach. *Environmetrics* **2001**, *12*, 559–568.

20. Druilhet, P.; Marin, J.M. Invariant HPD credible sets and MAP estimators. *Bayesian Anal.* **2007**, *2*, 681–692.

21. Pereira, C.; Stern, J.M.; Wechsler, S. Can a significance test be genuinely Bayesian? *Bayesian Anal.* **2008**, 79–100.

22. Diniz, M.A.; Pereira, C.; Polpo, A.; Stern, J.M.; Wechsler, S. Relationship between Bayesian and frequentist significance indices. *Int. J. Uncertain. Quantif.* **2012**, *2*, 161–172.

23. Kosmidis, I. Bias in parametric estimation: Reduction and useful side effects. *Wiley Interdiscip. Rev. Comput. Stat.* **2014**, *6*, 185–196.

24. Stern, S.E. A second-order adjustment to the profile likelihood in the case of a multidimensional parameter of interest. *J. R. Stat. Ser. B* **1997**, *59*, 653–665.

25. Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika* **1993**, *80*, 27–38.

26. Kenne Pagui, E.C.; Salvan, A.; Sartori, N. Median bias reduction of maximum likelihood estimates. *Biometrika* **2017**, *104*, 923–938.

27. Hartigan, J.A. Invariant prior densities. *Ann. Math. Statist.* **1964**, *35*, 836–845.

28. Hartigan, J.A. The asymptotically unbiased density. *Ann. Math. Statist.* **1965**, *36*, 1137–1152.

29. Sartori, N. Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *J. Stat. Plan. Inference* **2006**, *136*, 4259–4275.

30. Ruli, E.; Sartori, N.; Ventura, L. Robust approximate Bayesian inference. *J. Stat. Plan. Inference* **2020**, *205*, 10–22.

31. Liseo, B.; Loperfido, N. A note on reference priors for the scalar skew-normal distribution. *J. Statist. Plann. Inference* **2006**. *136*, 373–389.

32. Bortolato, E.; Ventura, L. On approximate robust confidence distributions. *Econom. Stat.* **2023**, ISSN 2452-3062. https://doi.org/10.1016/j.ecosta.2023.04.006.

33. Brubaker, M.; Salzmann, M.; Urtasun, R. A family of MCMC methods on implicitly defined manifolds. *Artif. Intell. Stat.* **2012**, *22*, 161–172.

34. Lewis, J.R.; MacEachern, S.N.; Lee, Y. Bayesian restricted likelihood methods: Conditioning on insufficient statistics in Bayesian regression (with discussion). *Bayesian Anal.* **2021**, *104*, 1393–1462.

35. Graham, M.; Storkey, A. Asymptotically exact inference in differentiable generative models. *Artif. Intell. Stat.* **2017**, 499–508.

36. Pace, L.; Salvan, A. *Principles of Statistical Inference*; World Scientific: Singapore, 1997.

37. Severini, T.A. On the relationship between Bayesian and non-Bayesian elimination of nuisance parameters. *Stat. Sin.* **1999**, *9*, 713–724.

38. Bortolato, E.; Kenne Pagui, E.C. Bias reduction and robustness in gaussian longitudinal data analysis. *J. Stat. Comput. Simul.* **2023**, 1–16. https://doi.org/10.1080/00949655.2023.2248334.

39. Agresti, A. *Foundations of Linear and Generalized Linear Models*; John Wiley & Sons: New York, NY, USA, 2015.

40. Kosmidis, I.; Kenne Pagui, E.C.; Sartori, N. Mean and median bias reduction in generalized linear models. *Stat. Comput.* **2020**, *30*, 43–59.

41. Kosmidis, I. brglm2: Bias Reduction in Generalized Linear Models. R Package Version 0.9.2. 2023. Available online: https://CRAN.R-project.org/package=brglm2 (accessed on 28 November 2023).

42. Bedoui, A.; Lazar, N.A. Bayesian empirical likelihood for ridge and lasso regressions. *Comput. Stat. Data Anal.* **2020**, *145*, 106917.