Taylor & Francis
Taylor & Francis Group

# Higher order asymptotic computation of Bayesian significance tests for precise null hypotheses in the presence of nuisance parameters

Stefano Cabras[a,b*], Walter Racugno[b] and Laura Ventura[c]

[a]*Department of Statistics, Carlos III University of Madrid, C/ Madrid 126, 28903 Getafe, Spain;*
[b]*Department of Mathematics, University of Cagliari, Via Ospedale 72, 09124 Cagliari, Italy;*
[c]*Department of Statistical Sciences, University of Padova, Via Cesare Battisti, 241, 35121 Padova, Italy*

The full Bayesian significance test (FBST) was introduced by Pereira and Stern for measuring the evidence of a precise null hypothesis. The FBST requires both numerical optimization and multidimensional integration, whose computational cost may be heavy when testing a precise null hypothesis on a scalar parameter of interest in the presence of a large number of nuisance parameters. In this paper we propose a higher order approximation of the measure of evidence for the FBST, based on tail area expansions of the marginal posterior of the parameter of interest. When in particular focus is on matching priors, further results are highlighted. Numerical illustrations are discussed.

## 1. Introduction

Consider a sampling model $f(y; \theta)$, with parameter $\theta \in \Theta \subseteq \mathbb{R}^d$, $d \geq 1$, and let $L(\theta) = L(\theta; y)$ denote the likelihood function for $\theta$ based on data $y$. Let $\pi(\theta)$ be a prior distribution for $\theta$ and let $\pi(\theta \mid y) \propto \pi(\theta)L(\theta)$ be the posterior distribution. In many applications it is usual to distinguish between quantities of primary interest and others not of direct concern by writing $\theta = (\psi, \lambda)$, with $\psi$ being scalar parameter of interest and $\lambda (d - 1)$-dimensional nuisance parameter. We are interested in testing the precise (or sharp) null hypothesis $H_0 : \psi = \psi_0$ versus $H_1 : \psi \neq \psi_0$. A possible example occurs in regression problems, when the parameter of interest $\psi$ is a regression coefficient, the null hypothesis is $H_0 : \psi = 0$, and the nuisance parameter is given by the remaining regression coefficients and possible variance parameters.

The usual Bayesian procedure for testing or model selection is based on the well-known Bayes factor (BF), which is defined as the ratio of the posterior to the prior odds in favour of $H_0$. We decide in favour of $H_0$ whenever the BF, or the corresponding weight of evidence log(BF), assumes high value. However, it is well known that, when the null hypothesis is precise and improper or vague priors are assumed, the BF can be undetermined, and it can lead to the so-called Jeffreys–Lindley's

---

*Corresponding author. Email: s.cabras@unica.it

paradox; see, e.g. Kass and Raftery.[1] Moreover, the BF is not calibrated, i.e. its finite sampling distribution is unknown and it may depend on the nuisance parameter.

Alternative to the BF, Pereira and Stern [2,3] provide an intuitive measure of evidence for the full Bayesian significance test (FBST) in favour of $H_0$. This measure is the posterior probability related to the less probable points of the parametric space, and it favours the null hypothesis whenever it is large; see, e.g. Madruga et al. [4,5] and Pereira et al.,[6] and references therein. Moreover, the FBST is based on a specific loss function,[4] and thus the decision made under this procedure is the action that minimizes the corresponding posterior risk.

When testing the null hypothesis $H_0 : \psi = \psi_0$, the FBST requires numerical optimization and multidimensional integration (see, e.g. [3,6]), which may be heavy or timely consuming in particular when the dimension of the nuisance parameter is large. These computational steps make the FBST a computationally intensive procedure.

In this paper we discuss a simple approximation of the FBST based on the higher order tail area approximation (HOTA) of the marginal posterior distribution of the parameter of interest (see, e.g. [7,8], and references therein), which requires little more than standard likelihood quantities for its implementation. In this respect, it is available at little additional computational cost over the first-order approximation.

Moreover, when in particular matching priors are used (see, e.g. [9], and references therein), the proposed approximation presents further advantages, since it does not require the elicitation of the prior on the nuisance parameters, it allows to perform accurate Bayesian inference even for small sample sizes, and it is shown that it is calibrated with respect to the Uniform$(0, 1)$ distribution.

The paper is organized as follows. Section 2 briefly reviews both the Pereira–Stern procedure and the tail area approximation for the marginal posterior distribution of $\psi$. Section 3 discusses the higher order approximation of the measure of evidence for the FBST. Some examples are discussed in Section 4. Finally, some concluding remarks are given in Section 5.

## 2. Statistical methods

### 2.1. *The Pereira–Stern measure of evidence*

Pereira and Stern [2] introduce a measure of evidence in favour of a null hypothesis $H_0$, which does not require explicitly elicitation of prior probabilities for the hypotheses $H_i$, $i = 0, 1$; see also Pereira and Stern,[3] Madruga et al.,[5] and Pereira et al.[6] Consider the following two hypotheses for $\theta$:

$$H_0 : \theta = \theta_0 \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \notin \Theta_0,$$

where $\Theta_0 = \{(\psi, \lambda) \in \Theta : \psi = \psi_0\}$ is a subset of the parametric space $\Theta$. To define the measure of evidence for the FBST, Pereira and Stern first consider the maximum of the posterior density over the null hypothesis, attained at $\theta^* = \arg\max_{\theta \in \Theta_0} \pi(\theta \mid y)$, given by the value $\pi(\theta^* \mid y)$. Second, they define $T^* = \{\theta \in \Theta : \pi(\theta \mid y) \geq \pi(\theta^* \mid y)\}$ as the set 'tangent' to the null hypothesis, whose credibility is $\kappa^* = \int_{T^*} \pi(\theta \mid y) \, d\theta$. The measure of evidence for the FBST is the complement of the probability of the set $T^*$, that is

$$EV = 1 - \kappa^*. \tag{1}$$

The procedure rejects the null hypothesis whenever EV is small.

The FBST is a Bayes test. Indeed, Madruga et al. [4] prove that the FBST procedure is the posterior minimization of an expected loss function defined by

$$\text{Loss}(\text{Accept } H_0, \theta) = b + cI(\theta \in T^*)$$
$$\text{Loss}(\text{Reject } H_0, \theta) = a[1 - I(\theta \in T^*)],$$

where $a$, $b$ and $c$ are real positive numbers. With respect to this loss function, the optimal Bayesian decision is to accept $H_0$ if

$$\text{EV} > \frac{b+c}{c+a} = p,$$

i.e. if EV is greater than a fixed critical level $0 < p < 1$.

The computation of Equation (1) is performed in two steps: (*a*) a numerical optimization and (*b*) a numerical integration.[2] The numerical optimization step consists in finding the argument $\theta^*$ that maximizes $\pi(\theta \mid y)$ under the null hypothesis. The numerical integration step consists of integrating the posterior density over the region where it is greater than $\pi(\theta^* \mid y)$, to obtain $\kappa^*$. Even if efficient computational algorithms are available for local and global optimization, as well as for numerical integration, the two steps may be heavy or timely consuming to perform, in particular when the dimension of the nuisance parameter $\lambda$ is large. Moreover, as pointed out in [6], the sophisticated numerical algorithms used in the computation of Equation (1) may be a serious obstacle to the popularization of the FBST.

## 2.2. *Bayesian higher order asymptotics*

Bayesian inference on $\psi$, in the presence of the nuisance parameter $\lambda$, is based on the marginal posterior distribution

$$\pi_{\mathrm{m}}(\psi \mid y) = \frac{\int \pi(\psi, \lambda) L(\psi, \lambda) \, \mathrm{d}\lambda}{\int \pi(\psi, \lambda) L(\psi, \lambda) \, \mathrm{d}\lambda \, \mathrm{d}\psi}. \tag{2}$$

In order to compute Equation (2) and the related tail area, it is possible to resort to higher order asymptotics, i.e. accurate approximations which provide very precise inferences even when the sample size is small (see, among others [7], and references therein). The basic regularity conditions for the approximations given in this section are that there exists a unique maximum likelihood estimate (MLE) or a unique posterior mode of $\theta$ (see, for instance, [10]).

Let $\ell_{\mathrm{p}}(\psi) = \log L(\psi, \hat{\lambda}_{\psi})$ be the profile loglikelihood for $\psi$, with $\hat{\lambda}_{\psi}$-constrained MLE of $\lambda$ given $\psi$. Moreover, let $(\hat{\psi}, \hat{\lambda})$ be the full MLE, and let $j_{\mathrm{p}}(\psi) = -\partial^2 \ell_{\mathrm{p}}(\psi)/\partial \psi^2$ be the profile observed information. The marginal posterior distribution (2) can be approximated by expanding the numerator $L(\psi, \lambda)$ as a function of $\lambda$ about $\hat{\lambda}_{\psi}$ and by using the Laplace formula for the denominator, see, e.g. Tierney and Kadane.[11] We get

$$\pi_{\mathrm{m}}(\psi \mid y) \doteq c |j_{\mathrm{p}}(\hat{\psi})|^{1/2} \exp\{\ell_{\mathrm{p}}(\psi) - \ell_{\mathrm{p}}(\hat{\psi})\} \frac{|j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|^{1/2}}{|j_{\lambda\lambda}(\psi, \hat{\lambda}_{\psi})|^{1/2}} \frac{\pi(\psi, \hat{\lambda}_{\psi})}{\pi(\hat{\psi}, \hat{\lambda})}, \tag{3}$$

where $c$ is the normalizing constant, $j_{\lambda\lambda}(\psi, \lambda)$ is the $(\lambda, \lambda)$-block of the observed Fisher information $j(\psi, \lambda)$ from $L(\psi, \lambda)$, and the symbol '$\doteq$' indicates that the approximation is accurate to $O(n^{-3/2})$. An application of the tail area argument gives the corresponding $O(n^{-3/2})$ approximation to the

marginal posterior tail area probability.[7] In particular, we have

$$\int_{\psi_0}^{\infty} \pi_m(\psi \mid y)\, d\psi \doteq \Phi(r_B^*(\psi_0)), \tag{4}$$

where $\Phi(\cdot)$ is the standard normal distribution function and

$$r_B^*(\psi) = r_p(\psi) + \frac{1}{r_p(\psi)} \log \frac{q_B(\psi)}{r_p(\psi)}, \tag{5}$$

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

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

When the particular class of matching priors is considered (see [9], and references therein), the marginal posterior distribution for $\psi$ can be expressed as

$$\pi_m(\psi \mid y) \propto L_{mp}(\psi)\pi_{mp}(\psi), \tag{6}$$

where $L_{mp}(\psi) = L_p(\psi)M(\psi)$ is the modified profile likelihood for a suitably defined correction term $M(\psi)$ (see, e.g. [12, Chapter 9]), and $\pi_{mp}(\psi) \propto i_{\psi\psi\cdot\lambda}(\psi, \hat{\lambda}_\psi)^{1/2}$ is the corresponding matching prior, with $i_{\psi\psi\cdot\lambda}(\psi, \lambda) = i_{\psi\psi}(\psi, \lambda) - i_{\psi\lambda}(\psi, \lambda)i_{\lambda\lambda}(\psi, \lambda)^{-1}i_{\lambda\psi}(\psi, \lambda)$ partial information, and $i_{\psi\psi}(\psi, \lambda)$, $i_{\psi\lambda}(\psi, \lambda)$, $i_{\lambda\lambda}(\psi, \lambda)$, and $i_{\lambda\psi}(\psi, \lambda)$ blocks of the expected Fisher information $i(\psi, \lambda)$ from $L(\psi, \lambda)$.

Accurate tail area probabilities are computable from Equation (6). In particular, we have

$$\int_{\psi_0}^{\infty} \pi_m(\psi \mid y)\, d\psi \doteq \Phi(r_p^*(\psi_0)), \tag{7}$$

where

$$r_p^*(\psi) = r_p(\psi) + \frac{1}{r_p(\psi)} \log \frac{q_F(\psi)}{r_p(\psi)} \tag{8}$$

is the modified profile likelihood root of see Barndorff-Nielsen and Chamberlin,[13] with

$$q_F(\psi) = \frac{\ell_p'(\psi)}{j_p(\hat{\psi})^{1/2}} \frac{i_{\psi\psi\cdot\lambda}(\hat{\psi}, \hat{\lambda})^{1/2}}{i_{\psi\psi\cdot\lambda}(\psi, \hat{\lambda}_\psi)^{1/2}} \frac{1}{M(\psi)}. \tag{9}$$

Thus, Equation (8) is a higher order pivotal quantity, which allows one to obtain frequentist *p*-values, confidence limits and accurate point estimators.

## 3. Higher order approximation for EV

We are interested in testing the precise (or sharp) null hypothesis $H_0 : \psi = \psi_0$ versus $H_1 : \psi \neq \psi_0$. In order to avoid the numerical optimization and multidimensional integration required for the FBST, in this section we discuss a simple approximation of the FBST based on the HOTA of the marginal posterior distribution of the parameter of interest. Moreover, when focus is on matching priors, further theoretical results are highlighted.
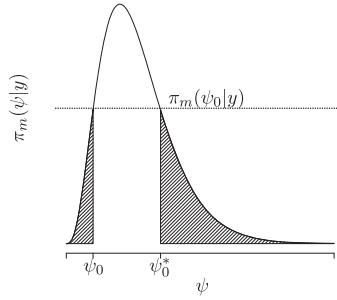
Figure 1.   Shaded area: EV for the precise hypothesis $H_0 : \psi = \psi_0$.

Consider the marginal posterior distribution (2) for the parameter of interest $\psi$, and consider the set

$$T(y) = \{\psi : \pi_{\mathrm{m}}(\psi \mid y) \geq \pi_{\mathrm{m}}(\psi_0 \mid y)\}. \tag{10}$$

Starting from $\pi_{\mathrm{m}}(\psi \mid y)$, the Pereira–Stern measure of evidence in favour of $H_0$ can be computed as (Figure 1)

$$\mathrm{EV} = 1 - \mathrm{Pr}_{\pi}(\psi \in T(y)), \tag{11}$$

where $\mathrm{Pr}_{\pi}(\cdot)$ denotes posterior probability, and the null hypothesis $H_0$ is accepted whenever EV is large enough.

A first-order approximation for Equation (11) is simply given by Pereira et al.,[6]

$$\mathrm{EV} \overset{O(n^{-1/2})}{=} 2\Phi\left(\frac{\psi_0 - \hat{\psi}}{\sqrt{j_{\mathrm{p}}(\hat{\psi})^{-1}}}\right). \tag{12}$$

In practice, it is well known that Equation (12) is often inaccurate, in particular when the dimension of $\lambda$ is large with respect to the sample size. Moreover, it forces the marginal posterior distribution to be symmetric.

The following theorem provides the higher order approximation for EV based on the tail area approximation (4).

THEOREM 1   *The third-order approximation of the measure of evidence (11) used in the FBST is*

$$\mathrm{EV} \doteq 1 - \Phi(r_{\mathrm{B}}^*(\psi_0)) + \Phi(r_{\mathrm{B}}^*(\psi_0^*)). \tag{13}$$

*Proof*   Let us assume, without loss of generality, that $\psi_0$ is smaller than the posterior mode of $\pi_{\mathrm{m}}(\psi \mid y)$ (as in Figure 1), and let $\psi_0^*$ be the value of the parameter such that $\pi_{\mathrm{m}}(\psi_0^* \mid y) = \pi_{\mathrm{m}}(\psi_0 \mid y)$. Then

$$\mathrm{EV} = \int_{-\infty}^{\psi_0} \pi_{\mathrm{m}}(\psi \mid y)\, \mathrm{d}\psi + \int_{\psi_0^*}^{+\infty} \pi_{\mathrm{m}}(\psi \mid y)\, \mathrm{d}\psi.$$

Using Equation (4), we can compute EV as in Equation (13), with $r_{\mathrm{B}}^*(\psi)$ defined in Equation (5).

Note that the higher order approximation (13) does not call for any condition on the prior $\pi(\psi, \lambda)$, i.e. it can be also improper, and on the corresponding marginal posterior $\pi_{\mathrm{m}}(\psi \mid y)$. ∎

When $\pi_{\mathrm{m}}(\psi \mid y)$ is symmetric, Equation (13) reduces to $\mathrm{EV} \doteq 2(1 - \Phi(r_{\mathrm{B}}^*(\psi_0)))$. Moreover,

$$\Phi(r_{\mathrm{B}}^*(\psi_0)) - \Phi(r_{\mathrm{B}}^*(\psi_0^*)) \doteq \int_{\psi_0}^{\psi_0^*} \pi_{\mathrm{m}}(\psi \mid y) \, \mathrm{d}\psi = \Pr_\pi(\psi \in T(y)) = 1 - \mathrm{EV},$$

gives the posterior probability of the HPD credible interval $(\psi_0, \psi_0^*)$.

To compute Equation (13) in practice, a simple simulation scheme based on HOTAs can be used.[8] The implementation of the HOTA sampling scheme is available at little additional computation cost over simple first-order approximations, and it has the advantage over Markov Chain Monte Carlo (MCMC) methods that samples are drawn independently in much lower computation time. Starting from $\pi_{\mathrm{m}}(\psi \mid y)$, the simulation algorithm can be summarized as follows. For $j = 1, \ldots, J$:

(1)  generate a pseudo-random number $z_j \sim N(0, 1)$, independently;
(2)  compute $\psi_j$ as the solution of $r_{\mathrm{B}}^*(\psi_j) = z_j$;

obtain thus a sample $(\psi_1, \ldots, \psi_J)$ from the marginal density $\pi_{\mathrm{m}}(\psi \mid y)$. The HOTA simulation procedure is essentially an inverse method of sampling and it provides independent samples from $\pi_{\mathrm{m}}(\psi \mid y)$ by inverting the cumulative distribution function approximation (4). In this respect, HOTA has an advantage over MCMC methods in that it is easier to implement and computationally faster. Moreover, it provides also a convenient approach for a sensitivity analysis with respect to the prior specification.[8]

When the class of matching priors is considered, then Equation (13) reduces to

$$\mathrm{EV} \doteq 1 - \Phi(r_{\mathrm{p}}^*(\psi_0)) + \Phi(r_{\mathrm{p}}^*(\psi_0^*)), \tag{14}$$

where $r_{\mathrm{p}}^*(\psi)$ is defined in Equation (8). Note that Equation (14) does not require the explicit elicitation on the nuisance parameters.

The following theorem shows that Equation (14) is calibrated to second order with respect to the Uniform$(0, 1)$ distribution.

THEOREM 2   *The sampling null distribution of the* EV *given in Equation (13) at* $\theta = \theta_0$ *is* Uniform$(0, 1)$ *to second order if and only if the prior is matching.*

*Proof*   Let us denote with $G_{\mathrm{m}}(\psi)$ the cumulative distribution function of the posterior $\pi_{\mathrm{m}}(\psi \mid y)$. Consider a function $B(\alpha) \subseteq [0, 1]$ for $\alpha \in [0, 1]$ with Lebesgue measure $\alpha$. Then, there exists the set $\bar{T}_\alpha(y) = G_{\mathrm{m}}^{-1}(B(\alpha))$, complementary to the set $T(y)$, such that $P_\pi\{\psi \in \bar{T}_\alpha(y)\} = \alpha$. Moreover, when considering the matching prior $\pi_{\mathrm{mp}}(\psi)$, it holds

$$P_\pi\{\psi \in \bar{T}_\alpha(y)\} = P_\theta\{\psi \in \bar{T}_\alpha(Y)\} + O_{\mathrm{p}}(n^{-1}) = \alpha + O_{\mathrm{p}}(n^{-1}),$$

where $P_\theta(\cdot)$ denotes probability under $f(y; \theta)$.

The if part can be shown as follows. Let $B(\alpha) = [0, \alpha)$, and assume that $\mathrm{EV} \sim \mathrm{Uniform}(0, 1)$ under $H_0$. Then

$$O_{\mathrm{p}}(n^{-1}) + \alpha = P_{\theta_0}\{\mathrm{EV} \in [0, \alpha)\}$$
$$= P_{\theta_0}\{\mathrm{EV} \in B(\alpha)\}$$
$$= P_{\theta_0}\{G_{\mathrm{m}}^{-1}(\mathrm{EV}) \in G_{\mathrm{m}}^{-1}(B(\alpha))\}$$
$$= P_{\theta_0}\{\psi \in \bar{T}_\alpha(Y)\},$$

where the last equality is the definition of matching prior. For the converse, assuming that the prior is matching, then

$$
\begin{aligned}
O_{\mathrm{p}}(n^{-1}) + \alpha &= P_{\theta_0}\{\psi \in \bar{T}_\alpha(Y)\} \\
&= P_{\theta_0}\{G_{\mathrm{m}}(\psi) \in G_{\mathrm{m}}(\bar{T}_\alpha(Y))\} \\
&= P_{\theta_0}\{\mathrm{EV} \in B(\alpha)\} \\
&= P_{\theta_0}\{\mathrm{EV} \in [0, \alpha)\}.
\end{aligned}
$$

∎

## 4. Monte Carlo studies and applications

The aim of this section is to illustrate the use and the accuracy, even for small sample sizes, of the approximation (13) of the EV in comparison to the original measure (1). The EV (11) from the marginal posterior distribution $\pi_{\mathrm{m}}(\psi \mid y)$ can be obtained exactly ($\mathrm{EV}^{\mathrm{ex}}$) if the marginal posterior distribution is known analytically, or from Equation (13) using the HOTA method ($\mathrm{EV}^{r_{\mathrm{B}}^*}$), or from Equation (12) when considering the first-order approximation ($\mathrm{EV}^{\mathrm{fo}}$). The HOTA method has been employed as explained in [8] that is approximating $r_{\mathrm{B}}^*(\psi)$ in a grid of 50 points and then using 10,000 Monte Carlo samples from the approximated marginal posterior.

In the examples different default priors have been considered together with the matching prior, in order to perform a sensitivity analysis of the EV with respect to the prior specification.

### 4.1. *Example 1. Inverse Gaussian distribution*

Let $y = (y_1, \dots, y_n)$ be a random sample from an Inverse Gaussian distribution, with likelihood function

$$
L(\psi, \lambda; y) = \psi^{n/2} \exp\left[\psi \left(\frac{n}{\lambda} - \frac{t}{2\lambda^2} - \frac{a}{2}\right)\right],
$$

where $t = \sum_{i=1}^{n} y_i = n\bar{y}$ and $a = \sum_{i=1}^{n} 1/y_i$. Let $\psi$ be the parameter of interest and suppose to test $H_0 : \psi = 1$ versus $H_1 : \psi \neq 1$.

We consider the reference prior [14] with parameter order $(\psi, \lambda)$, i.e. $\pi_{\mathrm{R}}(\psi, \lambda) \propto \psi^{-1}\lambda^{-1/2}$. Based on such prior, we calculate the original measure (1) and the first-order and higher order approximations (12) and (13), respectively. We further consider the matching prior $\pi_{\mathrm{mp}}(\psi)$,[15] for which the marginal posterior $\pi_{\mathrm{m}}(\psi \mid y)$ is a gamma distribution with mean $(n-1)/2s$ and variance $(n-1)/2s^2$, where $s = (n/2)(a/n - 1/\bar{y})$. Quantiles for the exact computation of Equation (11) can be thus obtained numerically.

For a sample of $n = 20$ observations drawn under the null model (with $\psi_0 = 1$ and $\lambda = 1$), Figure 2 highlights the original EV under the reference prior (a), according to the original definition in Pereira and Stern,[2] and of $\mathrm{EV}^{\mathrm{fo}}$ (d) and $\mathrm{EV}^{r_{\mathrm{B}}^*}$ (c) under the reference prior (denoted, respectively, with $\mathrm{EV}_{\pi_{\mathrm{R}}}$, $\mathrm{EV}_{\pi_{\mathrm{R}}}^{\mathrm{fo}}$ and $\mathrm{EV}_{\pi_{\mathrm{R}}}^{r_{\mathrm{B}}^*}$), and of $\mathrm{EV}^{\mathrm{ex}}$ under the matching prior (b), denoted with $\mathrm{EV}_{\pi_{\mathrm{mp}}}^{\mathrm{ex}}$. The EV in Equation (1) is about $\mathrm{EV}_{\pi_{\mathrm{R}}} = 0.999$, while $\mathrm{EV}_{\pi_{\mathrm{R}}}^{r_{\mathrm{B}}^*} = 0.923$ and $\mathrm{EV}_{\pi_{\mathrm{R}}}^{\mathrm{fo}} = 0.789$. When using the matching prior, we have $\mathrm{EV}_{\pi_{\mathrm{mp}}}^{\mathrm{ex}} = 0.823$. Note that $\mathrm{EV}_{\pi_{\mathrm{R}}}^{r_{\mathrm{B}}^*}$ is quite close to $\mathrm{EV}_{\pi_{\mathrm{R}}}$, while the first-order approximation $\mathrm{EV}_{\pi_{\mathrm{R}}}^{\mathrm{fo}}$ appears inaccurate.

In order to study the asymptotic error in approximating the sampling null distributions of the EV with the Uniform(0, 1), a simulation study has been performed with 1000 independent samples of sizes $n = 5, 10, 20, 50$, with $\psi_0 = \lambda = 1$. For each sample, we evaluated $\mathrm{EV}_{\pi_{\mathrm{R}}}$, $\mathrm{EV}_{\pi_{\mathrm{R}}}^{r_{\mathrm{B}}^*}$, $\mathrm{EV}_{\pi_{\mathrm{R}}}^{\mathrm{fo}}$
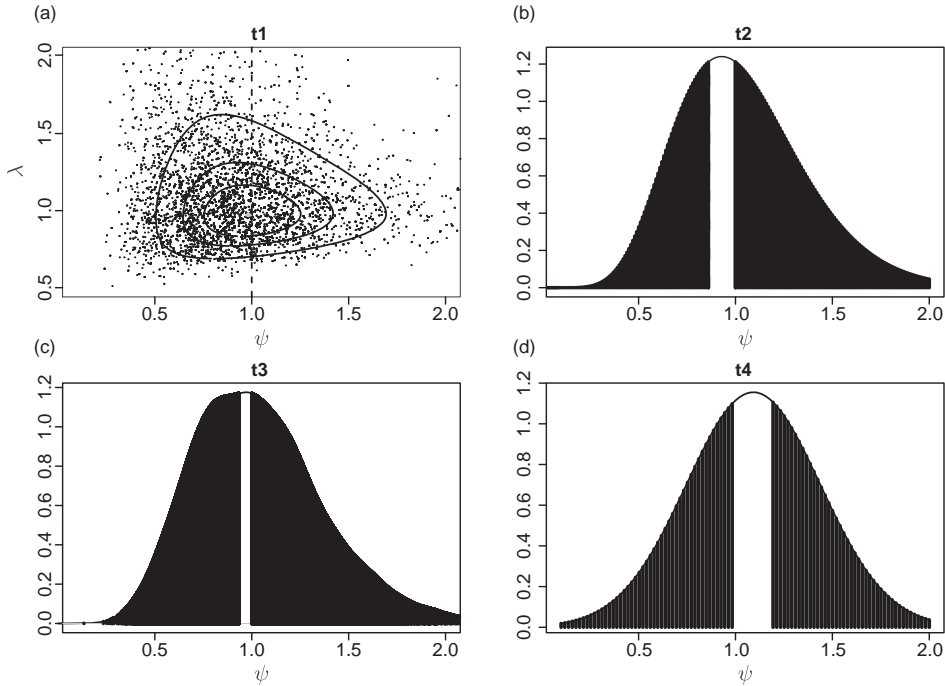
Figure 2.    Inverse Gaussian example: (a) tangent set $T^*$ under the reference prior (small grey area), (b) the tangent set $T$ under the matching prior, (c) tangent set $T$ with the third-order approximation and the reference prior and (d) tangent set $T$ with the first-order approximation and the reference prior.

and $EV_{\pi_{mp}}^{ex}$, and Figure 3 reports the QQ-plots against the Uniform$(0, 1)$ distribution. Moreover, Table 1 gives the nominal and empirical lower quantiles of the sampling null distribution of the different EV.

Figure 3 shows that $EV_{\pi_{mp}}^{ex}$ is uniformly distributed under the null model, as well as $EV_{\pi_R}^{r_B^*}$, and even is $EV_{\pi_R}^{fo}$ for larger sample sizes, under the reference prior. On the contrary, $EV_{\pi_R}$ appears to be conservative for the null model, more than expected under the uniform distribution; see also the quantiles in Table 1. For instance, from Table 1 we note that, for a sample of size $n = 5$, the frequency of observing a value of EV smaller than 1% under the null hypothesis tends to be larger for the reference analysis, while it is almost 1% under the matching analysis. The same occurs also for other nominal values of the EV and $n$.

### 4.2.    *Example 2. Extreme value regression*

Let $y_i$, $i = 1, \ldots, n$, be a random sample from the Weibull regression model, given by

$$\log(y_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \sigma \epsilon_i, \tag{15}$$

where $\epsilon_i$ has density $f(\epsilon) = \exp(\epsilon - e^\epsilon)$, i.e. the density of a log-Weibull variable, also called the extreme value density. Let $\psi = \beta_1$ be the parameter of interest, so that all other regression and scale parameters are nuisance parameters. For the null hypothesis $H_0 : \psi = 0$ versus $H_1 : \psi \neq 0$, the higher order approximation $r_p^*(\psi)$, which gives $EV_{\pi_{mp}}^{r_p^*}$ is illustrated in [16, p. 78] and [9].

The aim of this example is to illustrate the computational advantages of the higher order approximations of the EV with respect to the original expression (1), which requires integration and calculation of the tangential set $T^*$ over the full parameter space. In general, the computational
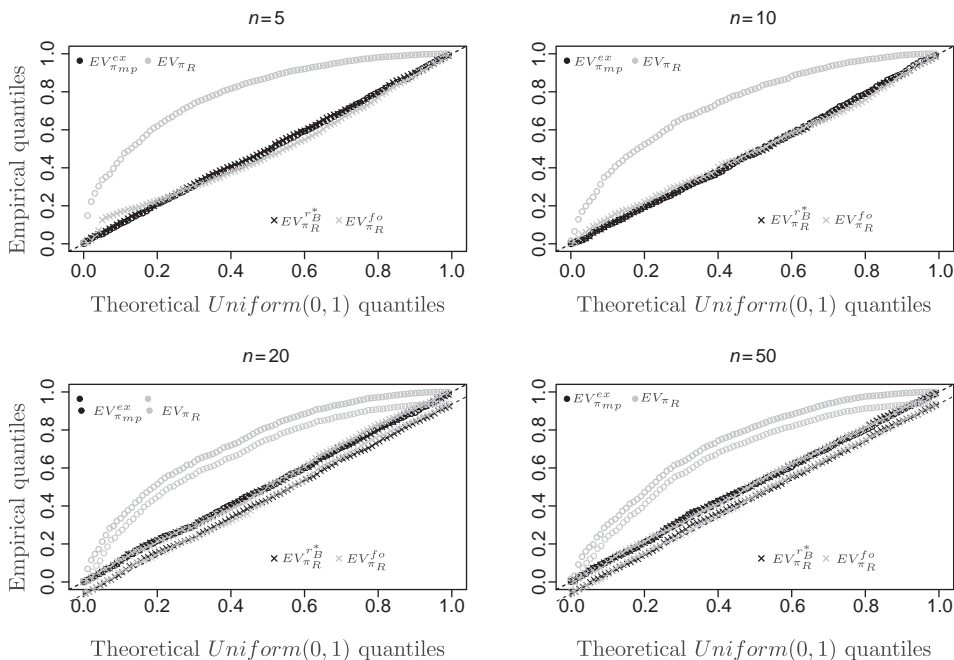
Figure 3. Inverse Gaussian example. Sampling null distribution of $EV_{\pi_R}$ (or), $EV_{\pi_R}^{r_B^*}$ (rb), $EV_{\pi_R}^{fo}$ (fo) and $EV_{\pi_{mp}}^{ex}$ (ex), for $n = 5, 10, 20$ and $50$.

Table 1. Nominal and empirical quantiles for different sample sizes of the sampling null distribution of the measures of evidence.

| | Matching prior | | | | | | Reference prior | | | | | |
| | $EV_{\pi_{mp}}^{ex}$ | | | $EV_{\pi_R}$ | | | $EV_{\pi_R}^{r_B^*}$ | | | $EV_{\pi_R}^{fo}$ | | |
| $n$ | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 0.02 | 0.05 | 0.11 | 0.15 | 0.34 | 0.44 | 0.02 | 0.05 | 0.11 | 0.00 | 0.12 | 0.17 |
| 10 | 0.01 | 0.04 | 0.10 | 0.07 | 0.22 | 0.36 | 0.01 | 0.04 | 0.10 | 0.01 | 0.08 | 0.13 |
| 20 | 0.01 | 0.05 | 0.11 | 0.07 | 0.21 | 0.33 | 0.01 | 0.05 | 0.11 | 0.01 | 0.06 | 0.12 |
| 50 | 0.01 | 0.06 | 0.11 | 0.03 | 0.18 | 0.30 | 0.01 | 0.06 | 0.11 | 0.01 | 0.05 | 0.10 |

issues under Equation (1) are not cumbersome with MCMC and small $d$. However, when the dimension of the nuisance parameter is large, the computation of the tangential set $T^*$ over the full parameter space may be problematic as well as the elicitation on the nuisance parameters.

Consider a real data set concerning a clinical study on malignant mesothelioma (MM).[17] This data set reports survival times for 77 individuals, with other covariates, like the gender, the type of MM, i.e. type epithelioid (37 cases), biphasic (18 cases) or sarcomatoid (22 cases), and a set of genetical markers. Consider the Weibull regression (15) of the survival times with all the covariates for a total of 36 regression parameters and 77 observations. The effect of the histotype, with respect to the baseline epithelioid, is modelled by two scalar regression parameters, which are here regarded as separately parameters of interest. In the following, we consider $EV_{\pi_{mp}}^{r_p^*}$ from Equation (14), $EV^{fo}$ from Equation (12) and the original (1), where the latter two are both based on the noninformative prior $\pi(\beta_1, \ldots, \beta_{36}, \sigma) \propto \sigma^{-1}$.
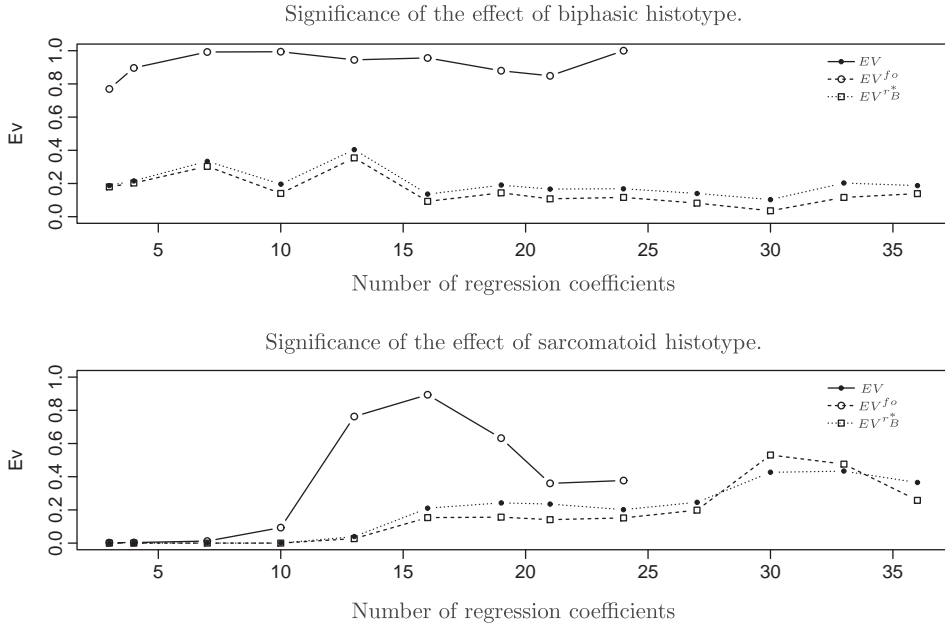
Figure 4. Extreme value regression. Evidence for the significance of the histotype effect on the survival from malignant mesotheliomas.[17] Evidence has been calculated using the original EV, the first-order EV$^{\text{fo}}$ and the third-order EV$^{r_B^*}$ approximations for nested regression models.

In order to illustrate the effect of the dimensionality of the nuisance parameter, in the evaluation of the measure of evidence for the histotype biphasic and for the histotype sarcomatoid, we considered different nested regression models all containing the initial regressor histotype. Figure 4 reports, for all the considered models, the value (1) of the original EV, the first-order approximation EV$^{\text{fo}}$ and the third-order approximation EV$_{\pi_{\text{mp}}}^{r_p^*}$. The latter two have been obtained with the HOTA sampling scheme, while Equation (1) has been computed using a random walk Metropolis over the full parameter space with a multivariate normal proposal. In both cases, $10^5$ samples were used, but HOTA has the advantage over MCMC methods that it samples independently. Moreover, using the same MCMC setup for the regression analyses up to 25 coefficients, we were not able to obtain a satisfactory approximation of the posterior with more than 25 coefficients and the corresponding values (1) have been not presented.

The values of EV given in Figure 4 indicate that according to EV$_{\pi_{\text{mp}}}^{r_p^*}$ there is a significant effect on the survival of sarcomatoid and biphasic histotypes with respect to the epithelioid, since the corresponding values are fairly below 0.5. Similar results are obtained with EV$^{\text{fo}}$ and such findings are in line with those obtained in [17]. On the contrary, when considering the original EV for the biphasic effect and also for the sarcomatoid effect, along with a large number of covariates, the results are quite different. Note also that the analyses with EV tends to diverge with respect to those with EV$^{\text{fo}}$ and EV$^{r_B^*}$ as the dimensionality of the integration space increases and this is mainly related to the computational problems in approximating the full posterior distribution with MCMC.

In order to study the asymptotic error in approximating the sampling null distribution of EV$_{\pi_{\text{mp}}}^{r_p^*}$, a simulation study has been performed under the null hypothesis of no effect of histotypes in survival. The simulation study is conducted using the same full design matrix as in [17], by simulating the response with all coefficients and scale equal to those estimated for this data set except those of sarcomatoid and biphasic which have been forced to be 0. The QQ-plot of the
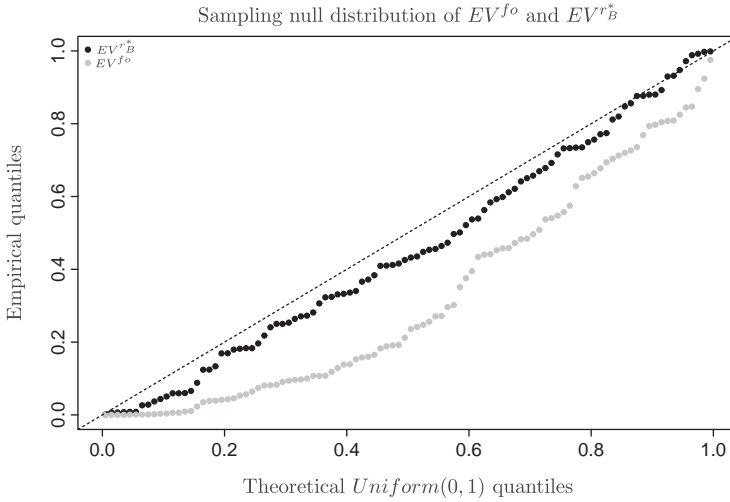
Figure 5. Extreme value regression: sampling null distribution of $EV^{fo}$ and of $EV_{\pi_{mp}}^{r_p^*}$ for the significance of the effect of biphasic histotype. The corresponding 100 measures of evidence have been simulated under no effect of histotypes on the survival.

sampling null distribution of $EV_{\pi_{mp}}^{r_p^*}$ and $EV^{fo}$ for 100 simulated samples, for the biphasic histotype, is reported in Figure 5. Note that $EV_{\pi_{mp}}^{r_p^*}$ exhibits a clear improvement over $EV^{fo}$.

### 4.3. *Example 3. Logistic regression*

Consider a logistic regression model applied to the `urine` data set given in [18]; see also Brazzale et al.[16, Chapter 4] The data set concerns calcium oxalate crystals in $n = 77$ samples of urine. The binary response $y$ indicates the presence of such crystals, and there are six explanatory variables: specific gravity (`gravity`), i.e. the density of urine relative to water; pH (`ph`); osmolarity (`osmo`, mOsm); conductivity (`conduct`, mMho); urea concentration (`urea`, millimoles per litre) and calcium concentration (`calc`, millimoles per litre). Let $X$ be the $n \times 7$ fixed design matrix composed by a first column of ones and the six covariates, as listed above, and let $\beta = (\beta_0, \ldots, \beta_6)$ be the vector of coefficients including the intercept. The log-likelihood function for $\beta$ is

$$\ell(\beta) = y^{\mathsf{T}} X \beta - \sum_{i=1}^{n} \log\{1 + \exp\{x_i^{\mathsf{T}} \beta\}\},$$

where $x_i$ represents the $i$th row of $X$, $i = 1, \ldots, n$, and $y$ is the vector of binary responses.

In order to assess the significance of a scalar regression coefficient, we consider the EV computed under the flat prior $\pi_F(\beta) \propto 1$, the noninformative G-prior $\pi_G(\beta)$ given in [19, p.101], and the EV defined in Equation (14), i.e. under the matching prior $\pi_{mp}(\psi)$. The higher order approximations of the EV have been computed according to the HOTA simulation scheme,[8] while Equation (1) has been obtained, for each prior, using a separately random walk Metropolis–Hastings with 50,000 samples after thinning a chain made of $10^6$ steps.

Figure 6 gives the marginal posterior distributions of the coefficients of interest in the analysis: $\beta_4$ (`conduct`), $\beta_5$ (`urea`) and $\beta_6$ (`calc`), as well as the values of the EV for $H_0 : \beta_4 = 0$, $H_0 : \beta_5 = 0$ and $H_0 : \beta_6 = 0$.

The original EV defined in Equation (1) in general seems to provide larger evidences for the null hypotheses, as also reported in Table 2. The evidence for the significance of $\beta_6$ is still strong,
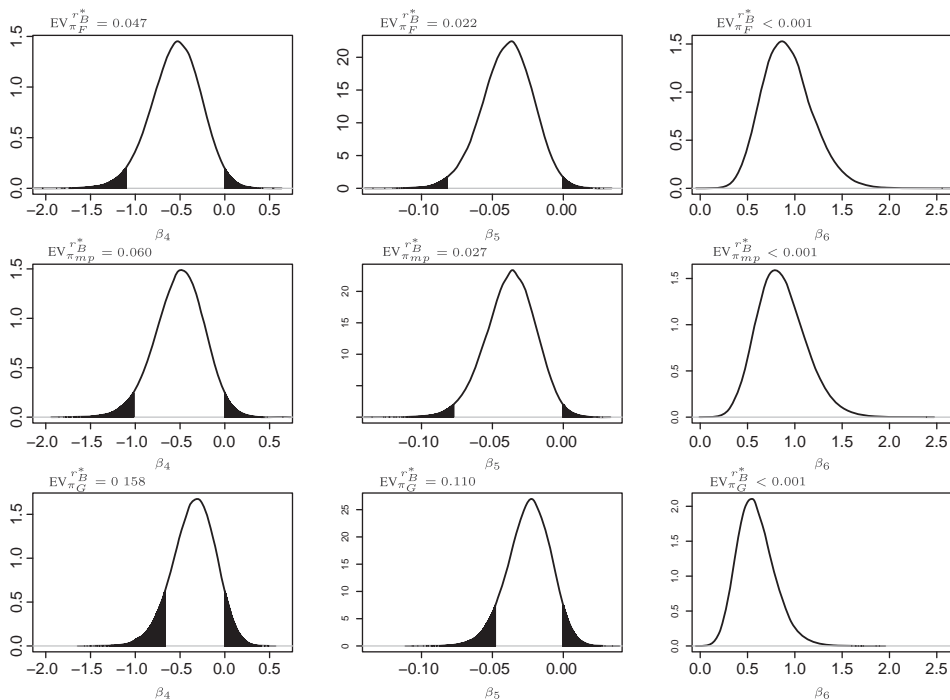
Figure 6. Logistic regression example: marginal posterior distributions obtained with the HOTA algorithm along with the corresponding measures of evidence for $\beta_4$, $\beta_5$ and $\beta_6$ under the flat prior (first row), the matching prior (second row) and the Zellner G-prior (last row).

Table 2. Logistic regression example: values of the evidence across different priors, and high-order $p$-values from Brazzale et al.[16]

| Null hypothesis | Original | | HOTA approximations | | | | |
|---|---|---|---|---|---|---|---|
| | $EV_{\pi_F}$ | $EV_{\pi_G}$ | $EV_{\pi_F}^{r_B^*}$ | $EV_{\pi_M}^{r_B^*}$ | $EV_{\pi_G}^{r_B^*}$ | $EV^{fo}$ | High-order $p$-value |
| $H_0 : \beta_4 = 0$ | 0.833 | 0.963 | 0.047 | 0.060 | 0.158 | 0.084 | .085 |
| $H_0 : \beta_5 = 0$ | 0.648 | 0.968 | 0.022 | 0.027 | 0.110 | 0.047 | .047 |
| $H_0 : \beta_6 = 0$ | 0.002 | 0.063 | <0.001 | <0.001 | <0.001 | 0.001 | .001 |

but much less than that provided by $EV_{\pi_{mp}}^{r_B^*}$, which in general agrees also with the evidence of the higher order frequentist analysis illustrated in [16] (see the $p$-values in Table 2).

## 5.  Final remarks

This paper discusses higher order asymptotics for the measure of evidence EV for the FBST, originally proposed in [2]. The computation of the EV considered here differs from the original one since the maximization in Equation (10) and integration in Equation (11) is done only in the dimension of the scalar parameter of interest, instead of the full parameter space $\Theta$. In this respect, in particular when the dimension $d$ is large, the proposed EV is computationally lighter and is preferable to the original EV which requires heavy or timely consuming computations. In particular, as illustrated in Examples 2 and 3 of Section 4, this may occur in regression problems, when the parameter of interest $\psi$ is a regression coefficient, and the nuisance parameter is given

by the remaining regression coefficients and possible variance parameters. Examples in Section 4 show also that the HOTA method can be successfully employed to compute Equation (13) in practice, with the advantage over MCMC methods that it samples independently, and to perform sensitivity analyses with respect to the prior specification.

With respect to the existing literature on procedures for testing or model selection, we note that, on the contrary of the BF, the higher order approximation of the EV is always defined and does not call for any condition on the prior $\pi(\psi, \lambda)$. Indeed, in the computation of Equation (13) the prior enters in $r_B^*(\psi)$ as $\pi(\hat\psi, \hat\lambda)/\pi(\psi, \hat\lambda_\psi)$, and thus it can be also improper. Finally, when in particular matching priors are used, we note that the approximate EV does not require the explicit elicitation on the nuisance parameters and is calibrated to second order with respect to the Uniform(0, 1) distribution.

## Funding

## References

[1] Kass R, Raftery A. Bayes factors. J Am Statist Assoc. 1995;90:773–795.
[2] Pereira C, Stern J. Evidence and credibility: full Bayesian significance test for precise hypotheses. Entropy. 1999;1:99–110.
[3] Pereira C, Stern J. Model selection: full Bayesian approach. Envitonmetrics. 2001;12:559–568.
[4] Madruga M, Esteves L, Wechslerz S. On the Bayesianity of Pereira–Stern tests. Test. 2001;10:291–299.
[5] Madruga M, Pereira C, Stern J. Bayesian evidence test for precise hypotheses. J Statist Plan Infer. 2003;117:185–198.
[6] Pereira C, Stern J, Wechsler S. Can a significance test be genuinely Bayesian? Bayesian Anal. 2008;3:79–100.
[7] Reid N. The 2000 wald memorial lectures: asymptotics and the theory of inference. Ann Stat. 2003;31:1695–1731.
[8] Ruli E, Sartori N, Ventura L. Marginal posterior simulation via higher-order tail area approximations. Bayesian Anal. 2014;9:129–146.
[9] Ventura L, Sartori N, Racugno W. Objective bayesian higher-order asymptotics in models with nuisance parameters. Comput Stat Data Anal. 2013;60:90–96.
[10] Kass R, Tierney L, Kadane J. The validity of posterior expansions based on Laplace's method. In: Geisser S, Hodges J, Press S, Zellner A, editors. Bayesian and likelihood methods in statistics and econometrics: essays in honor of George A. Barnard. North-Holland: Elsevier Science Publishers B.V.; 1990. p. 473–487.
[11] Tierney L, Kadane JB. Accurate approximations for posterior moments and marginal densities. J Am Statist Assoc. 1986;81:82–86.
[12] Severini TA. Likelihood methods in statistics. Oxford: Oxford University Press; 2000.
[13] Barndorff-Nielsen O, Chamberlin S. Stable and invariant adjusted directed likelihoods. Biometrika. 1994;81:485–499.
[14] Liseo B. Elimination of nuisance parameter with reference priors. Biometrika. 1993;80:295–304.
[15] Ventura L, Cabras S, Racugno W. Prior distributions from pseudo-likelihoods in the presence of nuisance parameters. J Am Statist Assoc. 2009;104(486):768–774.
[16] Brazzale AR, Davison AC, Reid N. Applied asymptotics: case studies in small-sample statistics. Cambridge: Cambridge University Press; 2007.
[17] Fassina A, Cappellesso R, Guzzardo V, Dalla Via L, Piccolo S, Ventura L, Fassan M. Epithelial-mesenchymal transition in malignant mesothelioma. Mod Pathol. 2012 Jan;25(1):86–99.
[18] Andrews DF, Herzberg AM. Data: a collection of problems from many fields for the student and research worker. New York: Springer; 1985.
[19] Marin J-M, Robert CP. Bayesian core: a practical approach to computational Bayesian statistics. Springer texts in statistics. New York: Springer; 2007.