




Article

# Bayesian Discrepancy Measure: Higher-Order and Skewed Approximations

Elena Bortolato <sup>1</sup>, Francesco Bertolino <sup>2</sup>, Monica Musio <sup>2</sup> and Laura Ventura <sup>3,\*</sup><sup>1</sup> Barcelona School of Economics, Universitat Pompeu Fabra, 08005 Barcelona, Spain; elena.bortolato@bse.eu<sup>2</sup> Department of Mathematics and Computer Science, University of Cagliari, 09124 Cagliari, Italy; bertolin@unica.it (F.B.); mmusio@unica.it (M.M.)<sup>3</sup> Department of Statistical Sciences, University of Padova, 35121 Padova, Italy

\* Correspondence: ventura@stat.unipd.it

## Abstract

The aim of this paper is to discuss both higher-order asymptotic expansions and skewed approximations for the Bayesian discrepancy measure used in testing precise statistical hypotheses. In particular, we derive results on third-order asymptotic approximations and skewed approximations for univariate posterior distributions, including cases with nuisance parameters, demonstrating improved accuracy in capturing posterior shape with little additional computational cost over simple first-order approximations. For third-order approximations, connections to frequentist inference via matching priors are highlighted. Moreover, the definition of the Bayesian discrepancy measure and the proposed methodology are extended to the multivariate setting, employing tractable skew-normal posterior approximations obtained via derivative matching at the mode. Accurate multivariate approximations for the Bayesian discrepancy measure are then derived by defining credible regions based on an optimal transport map that transforms the skew-normal approximation to a standard multivariate normal distribution. The performance and practical benefits of these higher-order and skewed approximations are illustrated through two examples.

**Keywords:** Bayesian discrepancy measure; credible regions; evidence; higher-order asymptotics; matching priors; nuisance parameter; optimal transport map; precise null hypothesis; skewed approximations; skew-normal distribution; tail area probability



Academic Editor: Geert Verdoolaege

Received: 29 April 2025

Revised: 16 June 2025

Accepted: 17 June 2025

Published: 20 June 2025

**Citation:** Bortolato, E.; Bertolino, F.; Musio, M.; Ventura, L. Bayesian Discrepancy Measure: Higher-Order and Skewed Approximations. *Entropy* **2025**, *27*, 657. <https://doi.org/10.3390/e27070657>

**Copyright:** © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Bayesian inference often relies on asymptotic arguments, leading to approximate methods that frequently assume a parametric form for the posterior distribution. In particular, a Gaussian distribution provides a convenient density for a first-order approximation. This practice is formally justified under regularity conditions by the Bernstein–von Mises theorem. However, this approximation fails to capture potential skewness and asymmetry in the posterior distribution. To avoid this drawback, starting from third-order expansions of Laplace’s method for the posterior distributions (see, e.g., [1–3], and references therein), possible alternatives are as follows:

- Higher-order asymptotic approximations: These offer improved accuracy at minimal additional computational costs compared to first-order approximations, and are applicable to posterior distributions and quantities of interest such as tail probabilities and credible regions (see, e.g., [4], and references therein);

- To use skewed approximations for the posterior distribution, theoretically justified by a skewed Bernstein–von Mises theorem (see, e.g., [5,6], and references therein).

The aim of this contribution is to discuss higher-order expansions and skew-symmetric approximations for the Bayesian discrepancy measure (BDM) proposed in [7] for testing precise statistical hypotheses. Specifically, the BDM assesses the compatibility of a given hypothesis with the available information (prior and data). To summarize this information, the posterior median is used, providing a straightforward evaluation of the discrepancy with the null hypothesis. The BDM possesses desirable properties such as consistency and invariance under reparameterization, making it a robust measure of evidence.

For a scalar parameter of interest, even with nuisance parameters, computing the BDM involves evaluating tail areas of the posterior or marginal posterior distribution. A first-order Gaussian approximation can be used, but it may be inaccurate, especially with small sample sizes or many nuisance parameters, since it fails to account for potential posterior asymmetry and skewness. In this respect, the aim of this paper is to provide higher-order asymptotic approximations and skewed asymptotic approximations for the BDM. For the third-order approximations, connections with frequentist inference are highlighted when using objective matching priors.

Also, for multidimensional parameters, while a first-order Gaussian approximation of the posterior distribution can be used to calculate the BDM, it still fails to account for potential posterior asymmetry and skewness. In this respect, this paper also addresses higher-order asymptotic approximations and skewed approximations for the BDM. The latter ones are based on an optimal transport map (see [8,9]), which transforms the skew-normal approximation to a standard multivariate normal distribution.

This paper is organized as follows. Section 2 provides some background for the BDM for a scalar parameter of interest, even with nuisance parameters, and extends the definition to the multivariate framework. Section 3 illustrates higher-order Bayesian approximations for the BDM; connections with frequentist inference are highlighted when using objective matching priors. Section 4 discusses skewed approximations for the posterior distribution and for the BDM, theoretically justified by a skewed Bernstein–von Mises theorem, with new insights into the multivariate framework. Two examples are discussed in Section 5. Finally, some concluding remarks are given in Section 6.

## 2. Background

Consider a sampling model  $f(y; \theta)$ , indexed by a parameter  $\theta \in \Theta \subseteq \mathbb{R}^d$ ,  $d \geq 1$ , and let  $L(\theta) = L(\theta; y) = \exp\{\ell(\theta)\}$  be the likelihood function based on a random sample  $y = (y_1, \dots, y_n)$  of size  $n$ . Given a prior density  $\pi(\theta)$  for  $\theta$ , Bayesian inference for  $\theta$  is based on the posterior density  $\pi(\theta|y) \propto \pi(\theta)L(\theta)$ .

In several applications, it is of interest to test the precise (or sharp) null hypothesis

$$H_0 : \theta = \theta_0 \tag{1}$$

against  $H_1 : \theta \neq \theta_0$ . In Bayesian hypothesis testing, the usual approach relies on the well-known Bayes factor (BF), which measures the ratio of posterior to prior odds in favor of the null hypothesis  $H_0$ . Typically, a high BF, or the weight of evidence  $W = \log(\text{BF})$ , provides support for  $H_0$ . However, improper priors can lead to an undetermined BF, and in the context of precise null hypotheses, the BF can be subject to the Jeffreys–Lindley paradox. This paradox highlights a critical divergence between frequentist and Bayesian approaches, that is, as the sample size increases, a  $p$ -value can become arbitrarily small, leading to the rejection of the null hypothesis, while the BF can simultaneously provide overwhelming evidence in favor of the same precise null. This typically occurs when the alternative hypothesis is associated with a diffuse prior distribution for the parameter of

interest. With such priors, the BF tends to favor the simpler  $H_0$  because, while data might be unlikely under  $H_0$ , it may also be poorly supported by any specific value within the diffuse alternative space. Furthermore, the BF is not well-calibrated, as its finite sampling distribution is generally unknown and may depend on nuisance parameters. To address these limitations, recent research has explored alternative Bayesian measures of evidence for precise null hypothesis testing, including the  $e$ -value (see, e.g., [10–12] and references therein) and the BDM [7]. In the following, we focus on the Bayesian discrepancy measure of evidence proposed in [7] (see also [13]).

2.1. Scalar Case

The BDM gives an absolute evaluation of a hypothesis,  $H_0$ , in light of prior knowledge about the parameter and observed data. In the absolutely continuous case, for testing (1), the BDM is defined as

$$\delta_H = 1 - 2 \min \left\{ \int_{-\infty}^{\theta_0} \pi(\theta|y) d\theta, 1 - \int_{-\infty}^{\theta_0} \pi(\theta|y) d\theta \right\}. \tag{2}$$

The quantity  $\min \{ \int_{-\infty}^{\theta_0} \pi(\theta|y) d\theta, 1 - \int_{-\infty}^{\theta_0} \pi(\theta|y) d\theta \}$  can be interpreted as the posterior probability of a “tail” event concerning only the precise hypothesis  $H_0$ . Doubling this “tail” probability, related to the precise hypothesis  $H_0$ , one obtains a posterior probability assessment about how “central” hypothesis  $H_0$  is, and, hence, how it is supported by the prior and the data. This interpretation is related to an alternative definition for  $\delta_H$ . Let  $\theta_m$  be the posterior median and consider the interval defined as  $I_E = (\theta_0, +\infty)$  if  $\theta_m < \theta_0$  or as  $I_E = (-\infty, \theta_0)$  if  $\theta_0 < \theta_m$ . Then, the BDM of the hypothesis  $H_0$  can be computed as

$$\delta_H = 1 - 2 P(\theta \in I_E|y) = 1 - 2 \int_{I_E} \pi(\theta|y) d\theta. \tag{3}$$

Note that the quantity  $2P(\theta \in I_E|y)$  represents the posterior probability of an equi-tailed credible interval for  $\theta$ .

The Bayesian discrepancy test assesses hypothesis  $H_0$  based on the BDM. High values of  $\delta_H$  indicate strong evidence against  $H_0$ , whereas low values suggest data consistency with  $H_0$ . Under  $H_0$ , for large sample sizes,  $\delta_H$  is asymptotically uniformly distributed on  $[0, 1]$ . Conversely, when  $H_0$  is false,  $\delta_H$  tends to 1 in probability. While thresholds can be set to interpret  $\delta_H$ , in line with the ASA statement, we agree with Fisher that significance levels should be tailored to each case based on evidence and ideas.

The BDM remains invariant under invertible monotonic reparametrizations. Under general regularity conditions and assuming Cromwell’s rule for prior selection,  $\delta_H$  exhibits specific properties, as follows: (1) if  $\theta_0 = \theta_t$  (the true value of the parameter),  $\delta_H$  tends toward a uniform distribution as the sample size increases; (2) if  $\theta_0 \neq \theta_t$ ,  $\delta_H$  converges to 1 in probability. Furthermore, using a matching prior,  $\delta_H$  is exactly uniformly distributed for all sample sizes.

The practical computation of  $\delta_H$  requires the evaluation of the tail areas of the following form:

$$P(\theta \geq \theta_0|y) = \int_{\theta_0}^{\infty} \pi(\theta|y) d\theta. \tag{4}$$

The derivation of a first-order tail area approximation is simple since it uses a Gaussian approximation. With this approximation, a first-order approximation for  $\delta_H$  when testing (1) is simply given by

$$\delta_H \doteq 2 \Phi \left( \left| \frac{\theta_0 - \hat{\theta}}{\sqrt{j(\hat{\theta})^{-1}}} \right| \right) - 1, \tag{5}$$

where  $\hat{\theta}$  is the maximum likelihood estimate (MLE) of  $\theta$ ,  $j(\theta) = -\ell^{(2)}(\theta) = -\partial^2 \ell(\theta) / \partial \theta^2$  is the observed information, 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,  $\delta_H$  agrees numerically with the  $1 - p$ -value based on the Wald statistic  $w(\theta) = (\hat{\theta} - \theta) / j(\hat{\theta})^{-1/2}$  and also with the first-order approximation of the  $e$ -value (see, e.g., [14]). In practice, the approximation (5) of  $\delta_H$  may be inaccurate, particularly for a small sample size, because it forces the posterior distribution to be symmetric.

### 2.2. Nuisance Parameters

In most applications,  $\theta$  is partitioned as  $\theta = (\psi, \lambda)$ , where  $\psi$  is a scalar parameter of interest and  $\lambda$  is a  $(d - 1)$ -dimensional nuisance parameter, and it is of interest to test the precise (or sharp) null hypothesis

$$H_0 : \psi = \psi_0 \tag{6}$$

against  $H_1 : \psi \neq \psi_0$ . In the absolutely continuous case, for testing (6) in the presence of nuisance parameters, the BDM is defined as

$$\delta_H = 1 - 2 \min \left\{ \int_{-\infty}^{\psi_0} \pi_m(\psi|y) d\psi, 1 - \int_{-\infty}^{\psi_0} \pi_m(\psi|y) d\psi \right\}, \tag{7}$$

where  $\pi_m(\psi|y)$  is the marginal posterior density for  $\psi$ , given by

$$\pi_m(\psi|y) = \int \pi(\psi, \lambda|y) d\lambda \propto \int \pi(\psi, \lambda)L(\psi, \lambda) d\lambda. \tag{8}$$

Also in this framework, the practical computation of  $\delta_H$  requires the evaluation of tail areas of the following form:

$$P_m(\psi \geq \psi_0|y) = \int_{\psi_0}^{\infty} \pi_m(\psi|y) d\psi. \tag{9}$$

The derivation of a first-order tail area approximation is still simple since it uses a Gaussian approximation. Let  $\ell_p(\psi) = \log L(\psi, \hat{\lambda}_\psi)$  be the profile log-likelihood for  $\psi$ , where  $\hat{\lambda}_\psi$  denotes the constrained MLE of  $\lambda$  given  $\psi$ . Moreover, let  $(\hat{\psi}, \hat{\lambda})$  be the full MLE, and let  $j_p(\psi) = -\ell_p^{(2)}(\psi) = -\partial^2 \ell_p(\psi) / \partial \psi^2$  be the profile observed information. A first-order approximation for  $\delta_H$  when testing (6) is simply given by

$$\delta_H \doteq 2 \Phi \left( \left| \frac{\psi_0 - \hat{\psi}}{\sqrt{j_p(\hat{\psi})^{-1}}} \right| \right) - 1. \tag{10}$$

Thus, to first order,  $\delta_H$  agrees numerically with  $1 - p$ -value based on the profile Wald statistic  $w_p(\psi) = (\hat{\psi} - \psi) / j_p(\hat{\psi})^{-1/2}$ . In practice, as in the scalar parameter case, the approximation (5) of  $\delta_H$  may be inaccurate, particularly for a small sample size or a large number of nuisance parameters, since it fails to account for potential posterior asymmetry and skewness.

### 2.3. The Multivariate Case

Extending the definition of the BDM to the multivariate setting, where  $\theta \in \Theta \subseteq \mathbb{R}^d$  with  $d > 1$ , presents some challenges. The core concepts of the univariate definition rely on the unique ordering of the real line and the uniquely defined median, which splits the probability mass into two equal halves (tail areas). In  $\mathbb{R}^d$ , with  $d > 1$ , there is no natural unique ordering, and concepts like the median and “tail areas” relative to a specific point  $\theta_0$  lack a single, universally accepted definition. Despite these challenges, the fundamental goal remains the same, that is, to quantify how consistent the hypothesized value  $\theta_0$  is with

the posterior distribution  $\pi(\theta|y)$ ; specifically measuring how “central” or, conversely, how “extreme”  $\theta_0$  lies within the posterior distribution.

Utilizing the notion of center-outward quantile functions ([8,9]), a concept from recent multivariate statistics, provides a theoretically appealing way to define the multivariate BDM. Let  $F_P^\pm : \mathbb{R}^d \rightarrow \mathbb{B}_d$  be the *center-outward distribution function* mapping the posterior distribution  $P_\theta$  (with density  $\pi(\theta|y)$ ) to the uniform distribution  $U_d$  on the unit ball  $\mathbb{B}_d$ . More precisely, the *center-outward distribution function*  $F_P^\pm : \mathbb{R}^d \rightarrow \mathbb{B}_d$  is defined as the *almost-everywhere unique gradient of a convex function* that pushes a distribution  $P_\theta$  forward to the uniform distribution  $U_d$  on the unit ball  $\mathbb{B}_d$  in  $\mathbb{R}^d$ . That is,

$$F_P^\pm := \nabla g, \quad \text{such that } F_P^\pm \# P_\theta = U_d.$$

The *center-outward quantile function*  $Q_P^\pm$  is defined as the (continuous) inverse of  $F_P^\pm$ , i.e.,

$$Q_P^\pm := (F_P^\pm)^{-1}.$$

It maps the open unit ball  $\mathbb{B}_d$  (minus the origin) to  $\mathbb{R}^d \setminus (F_P^\pm)^{-1}(\mathbf{0})$  and satisfies

$$Q_P^\pm \# U_d = P_\theta.$$

For  $\tau \in (0,1)$ , we define the *center-outward quantile region of order  $\tau$*  as

$$\mathcal{R}_P^\pm(\tau) := Q_P^\pm(\tau \mathbb{B}_d),$$

and the *center-outward quantile contour of order  $\tau$*  as

$$\mathcal{C}_P^\pm(\tau) := Q_P^\pm(\tau \mathbb{S}^{d-1}),$$

where  $\mathbb{S}^{d-1}$  is the unit sphere in  $\mathbb{R}^d$ . When  $d = 1$ , this coincides with the rescaled univariate cumulative distribution function  $F_P^\pm(x) = 2F_P(x) - 1$  and the BDM (7) can be expressed as

$$\delta_H = |F_P^\pm(\theta_0)|.$$

This measures the (rescaled) distance of the quantile rank of  $\theta_0$  from the center point (corresponding to rank 0). Generalizing this, we can define the multivariate BDM for the hypothesis  $H_0 : \theta = \theta_0$  as

$$\delta_H = \|F_P^\pm(\theta_0)\|, \tag{11}$$

where  $\|\cdot\|$  denotes the standard Euclidean norm in  $\mathbb{R}^d$ . Here,  $F_P^\pm(\theta_0)$  maps the point  $\theta_0$  to a location  $\mathbf{u}$  within the unit ball  $\mathbb{B}_d$ . This definition has desirable properties (see [8]):

- It yields a value between 0 and 1;
- $\delta_H = 0$  if  $\theta_0$  corresponds to the geometric center (or multivariate median) of the distribution (mapped to  $\mathbf{0}$  by  $F_P^\pm$ );
- $\delta_H$  increases as  $\theta_0$  moves away from the center toward the “boundary” of the distribution, approaching 1 for points mapped near the surface of the unit ball  $\mathbb{S}^{d-1}$ ;
- It is invariant under suitable classes of transformations (affine transformations if  $P_\theta$  is elliptically contoured, more generally under monotone transformations linked to an optimal transport map construction);
- It naturally reduces to the univariate definition  $\delta_H = |F_P^\pm(\theta_0)|$  when  $d = 1$ .

The primary practical difficulty lies in computing the center-outward distribution function  $F_P^\pm(\cdot)$  for an arbitrary posterior distribution  $\pi(\theta|y)$ , as it typically requires solving a complex optimal transport problem (see [15]).

### 3. Beyond Gaussian I: Higher-Order Asymptotic Approximations

#### 3.1. Scalar Case

In order to have more accurate evaluations of the first-order approximation (5) of  $\delta_H$ , it may be useful to resort to higher-order approximations based on tail area approximations (see, e.g., [3,4], and references therein). Using the tail area argument to the posterior density, we can derive the  $O(n^{-3/2})$  approximation:

$$P(\theta \geq \theta_0|y) \doteq \Phi(r_B^*(\theta_0)), \tag{12}$$

where the symbol “ $\doteq$ ” indicates that the approximation is accurate to  $O(n^{-3/2})$  and

$$r_B^*(\theta) = r(\theta) + \frac{1}{r(\theta)} \log \frac{q(\theta)}{r(\theta)},$$

with  $r(\theta) = \text{sign}(\hat{\theta} - \theta)[2(\ell(\hat{\theta}) - \ell(\theta))]^{1/2}$ , the likelihood root, and

$$q(\theta) = \ell^{(1)}(\theta) j(\hat{\theta})^{-1/2} \frac{\pi(\hat{\theta})}{\pi(\theta)}.$$

In the expression of  $q(\theta)$ ,  $\ell^{(1)}(\theta) = \partial\ell(\theta)/\partial\theta$  is the score function.

Using the tail area approximation (12), a third-order approximation of the BDM (2) can be computed as

$$\delta_H \doteq 1 - 2 \min\{\Phi(r_B^*(\theta_0)), 1 - \Phi(r_B^*(\theta_0))\} = 2\Phi(|r_B^*(\theta_0)|) - 1. \tag{13}$$

Note that the higher-order approximation (13) does not call for any condition on the prior  $\pi(\theta)$ , i.e., it can also be improper, and it is available at a negligible additional computational cost over the simple first-order approximation.

Note also that using  $r_B^*(\theta)$  an  $(1 - \alpha)$  equi-tailed credible interval for  $\theta$  can be computed as  $CI = \{\theta : |r_B^*(\theta)| \leq z_{1-\alpha/2}\}$ , where  $z_{1-\alpha/2}$  is the  $(1 - \alpha/2)$ -quantile of the standard normal distribution, and in practice, it can reflect asymmetries of the posterior. Moreover, from (12), the posterior median can be computed as the solution for  $\theta$  of the estimating equation  $r_B^*(\theta) = 0$ .

#### 3.2. Nuisance Parameters

When  $\theta$  is partitioned as  $\theta = (\psi, \lambda)$ , where  $\psi$  is a scalar parameter of interest and  $\lambda$  is a  $(d - 1)$ -dimensional nuisance parameter, in order to have more accurate evaluations of the first-order approximation (10) of  $\delta_H$ , using the tail area argument to the marginal posterior density, we can derive the  $O(n^{-3/2})$  approximation (see, e.g., [3,4]):

$$P_m(\psi \geq \psi_0|y) \doteq \Phi(r_{Bp}^*(\psi_0)), \tag{14}$$

where

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

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

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

In the expression of  $q_B(\psi)$ ,  $\ell_p^{(1)}(\psi)$  is the profile score function and  $j_{\lambda\lambda}(\psi, \lambda)$  represents the  $(\lambda, \lambda)$ -block of the observed information  $j(\psi, \lambda)$ .

Using the tail area approximation (14), a third-order approximation of the BDM (7) can be computed as

$$\delta_H \doteq 1 - 2 \min\{\Phi(r_{Bp}^*(\psi_0)), 1 - \Phi(r_{Bp}^*(\psi_0))\} = 2\Phi(|r_{Bp}^*(\psi_0)|) - 1. \tag{15}$$

Note that the higher-order approximation (15) does not call for any condition on the prior  $\pi(\psi, \lambda)$ , i.e., it can also be improper. Note also that, using  $r_{Bp}^*(\psi)$ , a  $(1 - \alpha)$  equi-tailed credible interval for  $\psi$  can be computed as  $CI = \{\psi : |r_{Bp}^*(\psi)| \leq z_{1-\alpha/2}\}$ . Moreover, from (14), the posterior median of (8) can be computed as the solution for  $\psi$  of the estimating equation  $r_{Bp}^*(\psi) = 0$ .

### Approximations with Matching Priors

The order of the approximations in the previous sections refers to the posterior distribution function and may depend, to varying degrees, on the choice of prior. A so-called strong-matching prior (see [16], and references therein) ensures that a frequentist  $p$ -value coincides with a Bayesian posterior survivor probability to a high degree of approximation, in the marginal posterior density (8).

Welch and Peers [17] showed that for a scalar parameter  $\theta$ , Jeffreys' prior is probability matching, in the sense that posterior survivor probabilities agree with frequentist probabilities, and credible intervals of a chosen width coincide with frequentist confidence intervals. With Jeffreys' prior, we have

$$q(\theta) = \ell^{(1)}(\theta)j(\hat{\theta})^{-1/2} \frac{i(\hat{\theta})^{1/2}}{i(\theta)^{1/2}}$$

and the corresponding  $r_B^*(\theta)$  coincides with the frequentist modified likelihood root as defined by [18]. In this case, using the tail area approximation (12), a third-order approximation of the BDM of the hypothesis  $H_0 : \theta = \theta_0$  coincides with  $1 - p^*$ , where  $p^*$  is the  $p$ -value based on the frequentist modified likelihood root. Thus, when using Jeffreys' prior and higher-order asymptotics in the scalar case, there is agreement between Bayesian and frequentist hypothesis testing.

In the presence of nuisance parameters, following [4], when using a strong matching prior, the marginal posterior density can be written as

$$\pi_m(\psi|y) \propto \exp\left(-\frac{1}{2}r_p^*(\psi)^2\right) \left| \frac{s_p(\psi)}{r_p(\psi)} \right|, \tag{16}$$

where  $s_p(\psi) = \ell_p^{(1)}(\psi)/j_p(\hat{\psi})^{1/2}$  is the profile score statistic, and  $r_p^*(\psi)$  is the modified profile likelihood root:

$$r_p^*(\psi) = r_p(\psi) + \frac{1}{r_p(\psi)} \log \frac{q_p(\psi)}{r_p(\psi)}, \tag{17}$$

which has a third-order standard normal null distribution. In (17), the quantity  $q_p(\psi)$  is a suitably defined correction term (see, e.g., [18] and [19], Chapter 9). Moreover, the tail area of the marginal posterior for  $\psi$  can be approximated to third order as

$$P_m(\psi \geq \psi_0|y) \doteq \Phi(r_p^*(\psi_0)), \tag{18}$$

A remarkable advantage of (16) and (18) is that their expressions automatically include the matching prior, without requiring its explicit computation.

Using (18), an asymptotic equi-tailed credible interval for  $\psi$  can be computed as  $CI = \{\psi : |r_p^*(\psi)| \leq z_{1-\alpha/2}\}$ , i.e., as a confidence interval for  $\psi$  based on (17) with approximate level  $(1 - \alpha)$ . Note from (18) that the posterior median of  $\pi_m(\psi|y)$  can be computed as the solution for  $\psi$  of the estimating equation  $r_p^*(\psi) = 0$ , and thus it coincides with the frequentist estimator defined as the zero-level confidence interval based on  $r_p^*(\psi)$ . Such an estimator has been shown to be a refinement of the MLE  $\hat{\psi}$ .

Using the tail area approximation (18), a third-order approximation of the BDM of the hypothesis  $H_0 : \psi = \psi_0$  is

$$\delta_H^* \doteq 1 - 2 \min\{\Phi(r_p^*(\psi_0)), 1 - \Phi(r_p^*(\psi_0))\} = 2 \Phi(|r_p^*(\psi_0)|) - 1. \tag{19}$$

In this case, (19) coincides with  $1 - p_r^*$ , where  $p_r^*$  is the  $p$ -value based on (17). Thus, when using strong matching priors and higher-order asymptotics, there is agreement between Bayesian and frequentist hypothesis testing, point estimation, and interval estimation.

From a practical point of view, the computation of (19) can be easily performed in practical problems using the `likelihoodAsy` package [20] of the statistical software R version 4.4.1. In practice, the advantage of using this package is that it does not require the function  $q_p(\psi)$  explicitly, but instead requires only the code for computing the log-likelihood function and for generating data from the assumed model. Some examples can be found in [14].

### 3.3. Multidimensional Parameters

When  $\theta$  is multidimensional, the derivation of a first-order tail area approximation and a first-order approximation for  $\delta_H$  remains straightforward, starting from the Laplace approximation of the posterior distribution. In particular, let  $W(\theta) = 2(\ell(\hat{\theta}) - \ell(\theta))$  be the log-likelihood ratio for  $\theta$ . Using  $W(\theta)$ , a first-order approximation of the BDM for the hypothesis  $H_0 : \theta = \theta_0$  can be obtained as follows:

$$\delta_H \doteq 1 - P(\chi_d^2 \geq W(\theta_0)), \tag{20}$$

where  $\chi_d^2$  is the Chi-squared distribution with  $d$  degrees of freedom. This approximation is asymptotically equivalent to the first-order approximation:

$$\delta_H \doteq 1 - P\left(\chi_d^2 \geq (\theta_0 - \hat{\theta})^T j(\hat{\theta})(\theta_0 - \hat{\theta})\right). \tag{21}$$

Higher-order approximations based on modifications of the log-likelihood ratios are also available for multidimensional parameters of interest, both with or without nuisance parameters (see [4,19,21], and references therein). As is the case with the approximations for a scalar parameter, the proposed results are based on the asymptotic theory of modified log-likelihood ratios [21], they require only routine maximization output for their implementation, and they are constructed for arbitrary prior distributions. For instance, paralleling the scalar parameter case, a credible region for a  $d$ -dimensional parameter of interest  $\theta$  with approximately  $100(1 - \alpha)\%$  coverage in repeat sampling, can be computed as  $CR = \{\theta : W^*(\theta) \leq \chi_{d;1-\alpha}^2\}$ , where  $W^*(\theta)$  is a suitable modification of the log-likelihood ratio  $W(\theta)$  or of the profile log-likelihood ratio (see [19,21]), and  $\chi_{d;1-\alpha}^2$  is the  $(1 - \alpha)$  quantile of the  $\chi_d^2$  distribution. In practice, the CR region can be interpreted as the extension to the multidimensional case of the equi-tailed CI set, i.e., the CR region is computed as a multidimensional case of the CI set based on the Chi-squared approximation. As in the scalar case, the CR region can reflect departures from symmetry with respect to the first-order approximation based on the Wald statistic. Some simulation studies on CR based on  $W^*(\theta)$  can be found in [22].

Using  $W^*(\theta)$ , a higher-order approximation of the BDM for the hypothesis  $H_0 : \theta = \theta_0$  can be obtained as

$$\delta_H \doteq 1 - P(\chi_d^2 \geq W^*(\theta_0)). \tag{22}$$

The major drawback with this approximation is that the signed root log-likelihood ratio transformation  $W^*(\theta)$  in general depends on the chosen parameter order. Moreover, its computation can be cumbersome when  $d$  is large.

#### 4. Beyond Gaussian II: Skewed Approximations

A major limitation of standard first-order Gaussian approximations, like (5) and (10), is their reliance on symmetric densities, which simplifies inference but can misrepresent key posterior features like skewness and heavy tails. Indeed, even simple parametric models can yield asymmetric posteriors, leading to biased and inaccurate approximations.

To overcome this, recent work has introduced flexible families of approximating posterior densities that can capture the shape and skewness [5,6,23]. In particular, [5] developed a class of closed-form deterministic approximations using a third-order extension of the Laplace approximation. This approach yields tractable, skewed approximations that better capture the actual shape of the target posterior while remaining computationally efficient.

Also, the skewed approximations, as well as the higher-order approximations discussed in Section 3, rely on higher-order expansions and derivatives. They start with a symmetric Gaussian approximation, but centered at the maximum a posteriori (MAP) estimate, and introduce skewness through the Gaussian distribution function combined with a cubic term influenced by the third derivative of the log-likelihood function.

##### 4.1. Scalar Case

Let us denote with  $\ell^{(k)}(\theta)$  the  $k$ -th derivative of the log-likelihood  $\ell(\theta)$ , i.e.,  $\ell^{(k)}(\theta) = \partial^k \ell(\theta) / \partial \theta^k$ ,  $k = 1, 2, 3, \dots$ . Moreover, let  $\tilde{\theta} = \operatorname{argmax}_{\theta \in \Theta} \{\ell(\theta) + \log \pi(\theta)\}$  be the MAP estimate of  $\theta$  and let  $h = \sqrt{n}(\theta - \tilde{\theta})$  be the rescaled parameter. Using result (14) of [5] and all the regularity conditions there stated, the skew-symmetric (SKS) approximation for the posterior density for  $\theta$  is

$$\pi_{SKS}(\theta|y) \propto 2\phi(h; 0, \tilde{\omega}) \Phi(\tilde{\alpha}(h)), \tag{23}$$

where  $\phi(h; 0, \tilde{\omega})$  is the normal density function with mean 0 and variance  $\tilde{\omega} = nj(\tilde{\theta})^{-1}$  and

$$\tilde{\alpha}(h) = \frac{\ell^{(3)}(\tilde{\theta})\sqrt{2\pi}}{12n^{3/2}}h^3$$

is the skewness component, expressed as a cubic function of  $h$ , reflecting the influence of the third derivative of the log-likelihood on the shape of the posterior distribution.

Equation (23) provides a practical skewed second-order approximation of the target posterior density, centered at its mode. This approach is known as the SKS approximation or skew-modal approximation. Compared to the classical first-order Gaussian approximation derived from the Laplace method, the SKS approximation remains similarly tractable while providing significantly greater accuracy. Note that this approximation depends on the prior distribution through the MAP.

Using (23) and the approximation

$$2\phi(h; 0, \tilde{\omega}) \left( \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\tilde{\alpha}(h) \right) = 2\phi(h; 0, \tilde{\omega})\Phi(\tilde{\alpha}(h)) + O(n^{-1}),$$

we can derive the approximation

$$P_{SKS}(\theta \geq \theta_0|y) = \frac{\int_{h_0}^{\infty} 2\phi(h; 0, \tilde{\omega}) \left( \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\tilde{\alpha}(h) \right) dh}{\int_{-\infty}^{\infty} 2\phi(h; 0, \tilde{\omega}) \left( \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\tilde{\alpha}(h) \right) dh}$$

for the tail area for (4), where  $h_0 = \sqrt{n}(\theta_0 - \tilde{\theta})$ . Note that the denominator is simply equal to 1, due to the symmetry of  $\phi(\cdot)$  and the oddness of  $\tilde{\alpha}(h)$ . The numerator can be split into two integrals:

$$\int_{h_0}^{\infty} 2\phi(h; 0, \tilde{\omega}) \left( \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \tilde{\alpha}(h) \right) dh = \int_{h_0}^{\infty} \phi(h; 0, \tilde{\omega}) dh + \frac{\sqrt{2}}{\sqrt{\pi}} \int_{h_0}^{\infty} \phi(h; 0, \tilde{\omega}) \tilde{\alpha}(h) dh.$$

The first integral can be expressed as the standard Gaussian tail, as follows:

$$\int_{h_0}^{\infty} \phi(h; 0, \tilde{\omega}) dh = \left( 1 - \Phi \left( \frac{h_0}{\sqrt{\tilde{\omega}}} \right) \right),$$

while the second integral involves the skewness term and can be expressed as

$$\frac{\sqrt{2}}{\sqrt{\pi}} \int_{h_0}^{\infty} \phi(h; 0, \tilde{\omega}) \tilde{\alpha}(h) dh = \frac{\ell^{(3)}(\tilde{\theta})}{6n^{3/2}} \int_{h_0}^{\infty} h^3 \phi(h; 0, \tilde{\omega}) dh.$$

Substituting  $z = h/\sqrt{\tilde{\omega}}$  into the integral  $\int_{h_0}^{\infty} h^3 \phi(h; 0, \tilde{\omega}) dh$ , we have

$$\begin{aligned} \int_{h_0}^{\infty} h^3 \phi(h; 0, \tilde{\omega}) dh &= \int_{h_0}^{\infty} h^3 \frac{1}{\sqrt{2\pi\tilde{\omega}}} \exp\left(-\frac{h^2}{2\tilde{\omega}}\right) dh \\ &= \int_{h_0/\sqrt{\tilde{\omega}}}^{\infty} (\sqrt{\tilde{\omega}}z)^3 \frac{1}{\sqrt{2\pi\tilde{\omega}}} \exp\left(-\frac{(\sqrt{\tilde{\omega}}z)^2}{2\tilde{\omega}}\right) \sqrt{\tilde{\omega}} dz \\ &= \tilde{\omega}^{3/2} \int_{h_0/\sqrt{\tilde{\omega}}}^{\infty} z^3 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz. \end{aligned}$$

Using the identity  $\int_{z_0}^{\infty} z^3 \phi(z; 0, 1) dz = \phi(z_0; 0, 1)(z_0^2 + 2)$ , with  $z_0 = h_0/\sqrt{\tilde{\omega}}$ , and  $\int_{-\infty}^{z_0} z^3 \phi(z; 0, 1) dz = -\phi(z_0; 0, 1)(z_0^2 + 2)$ , we obtain

$$\begin{aligned} \int_{h_0}^{\infty} h^3 \phi(h; 0, \tilde{\omega}) dh &= \tilde{\omega}^{3/2} \phi\left(\frac{h_0}{\sqrt{\tilde{\omega}}}; 0, 1\right) \left( \left(\frac{h_0}{\sqrt{\tilde{\omega}}}\right)^2 + 2 \right) \\ &= \tilde{\omega}^{3/2} \phi\left(\frac{h_0}{\sqrt{\tilde{\omega}}}; 0, 1\right) \left( \frac{h_0^2}{\tilde{\omega}} + 2 \right). \end{aligned}$$

Then, the resulting SKS approximation to  $P(\theta \geq \theta_0|y)$  is

$$P_{SKS}(\theta \geq \theta_0|y) = \left( 1 - \Phi \left( \frac{h_0}{\sqrt{\tilde{\omega}}} \right) \right) + \frac{\ell^{(3)}(\tilde{\theta})}{6n^{3/2}} \tilde{\omega}^{3/2} \phi\left(\frac{h_0}{\sqrt{\tilde{\omega}}}; 0, 1\right) \left( \frac{h_0^2}{\tilde{\omega}} + 2 \right).$$

Finally, substituting this approximation into (7), we obtain the SKS approximation of the BDM, given by

$$\delta_H^{SKS} = 2 \Phi \left( \left| \frac{h_0}{\sqrt{\tilde{\omega}}} \right| \right) - 2 \text{sign}(h_0) \frac{\ell^{(3)}(\tilde{\theta})}{6n^{3/2}} \tilde{\omega}^{3/2} \phi\left(\frac{h_0}{\sqrt{\tilde{\omega}}}; 0, 1\right) \left( \frac{h_0^2}{\tilde{\omega}} + 2 \right) - 1. \tag{24}$$

Note that the first term of this approximation differs from that in (5) since it is evaluated at the MAP and not at the MLE.

#### 4.2. Nuisance Parameters

As in Section 2.2, suppose that the parameter is partitioned as  $\theta = (\psi, \lambda)$ , where  $\psi$  is a scalar parameter of interest and  $\lambda$  is a nuisance parameter of dimension  $d - 1$ .

Also, for the marginal posterior distribution  $\pi_m(\psi|y)$ , an SKS approximation is available (see [5], Section 4.2).

Adopting the index notation, let us denote by  $j(\theta) = -[\ell_{st}^{(2)}(\theta)]$  the observed Fisher information matrix, where  $\ell_{st}^{(2)}(\theta) = \frac{\partial^2 \ell(\theta)}{\partial \theta_s \partial \theta_t}$ ,  $s, t = 1, \dots, d$ , and let  $\Omega = (j(\tilde{\theta})/n)^{-1}$  be the inverse of the scaled observed Fisher information matrix evaluated at the MAP. We denote the elements of  $\Omega$  by  $\Omega_{st}$ ; in particular, let us denote by  $\Omega_{11}$  the element corresponding to the parameter of interest,  $\psi$ . Moreover, let us denote with  $\ell_{stl}^{(3)}(\theta) = \frac{\partial^3 \ell(\theta)}{\partial \theta_s \partial \theta_t \partial \theta_l}$  the elements of the third derivative of the log-likelihood, with  $s, t, l = 1, \dots, d$ . Finally, let us define the following two quantities:

$$v_{1,1} = 3 \sum_{i=1}^d \sum_{j=1}^d \ell_{1ij}^{(3)}(\tilde{\theta}) \Omega_{ij} + 3 \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \ell_{ijk}^{(3)}(\tilde{\theta}) \Omega_{ij} \Omega_{k1}$$

and

$$\begin{aligned} v_{3,111} &= \ell_{111}^{(3)}(\tilde{\theta}) + 3 \sum_{i=1}^d \ell_{11i}^{(3)}(\tilde{\theta}) \Omega_{i1} + 3 \sum_{i=1}^d \sum_{j=1}^d \ell_{1ij}^{(3)}(\tilde{\theta}) \Omega_{ij} \Omega_{j1} \\ &+ \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \ell_{ijk}^{(3)}(\tilde{\theta}) \Omega_{ij} \Omega_{k1} \Omega_{11}. \end{aligned}$$

Then, following formula (23) in [5], the SKS approximation of the marginal posterior density  $\pi_m(\psi|y)$  can be expressed as

$$\pi_{mSKS}(\psi|y) \propto 2 \phi(h_\psi; 0, \Omega_{11}) \Phi(\alpha_\psi(h_\psi)), \tag{25}$$

where  $h_\psi = \sqrt{n}(\psi - \tilde{\psi})$  is the rescaled parameter of interest,  $\phi(\cdot; 0, \Omega_{11})$  is the density of a Gaussian distribution with mean 0 and variance  $\Omega_{11}$ , and the skewness component  $\alpha_\psi(h_\psi)$  is defined as

$$\alpha_\psi(h_\psi) = \frac{\sqrt{2\pi}}{12n^{3/2}} (v_{1,1}h_\psi + v_{3,111}h_\psi^3).$$

Using (25), we can derive the SKS tail area approximation of (9), given by

$$P_{mSKS}(\psi \geq \psi_0|y) = \int_{h_{\psi_0}}^{\infty} 2 \phi(h_\psi; 0, \Omega_{11}) \Phi(\alpha_\psi(h_\psi)) dh_\psi,$$

where  $h_{\psi_0} = \sqrt{n}(\psi_0 - \tilde{\psi})$ . Finally, the marginal SKS approximation of the BDM is given by

$$\delta_H^{mSKS} = 1 - 2 \min\{P_{mSKS}(\psi \geq \psi_0|y), 1 - P_{mSKS}(\psi \geq \psi_0|y)\}. \tag{26}$$

The marginal SKS tail area approximation  $P_{mSKS}(\psi \geq \psi_0|y)$ , and, thus, also  $\delta_H^{mSKS}$ , can be derived numerically.

### 4.3. Multidimensional Parameters

While the SKS approximation is theoretically elegant, similar to the higher-order modification of the log-likelihood ratio  $W^*(\theta)$ , it has two main drawbacks. The first one is that it relies only on local information around the mode. The second is that it is computationally intensive because it relies on third-order derivatives (i.e., a tensor of derivatives) of the log-likelihood. The size of this derivative tensor increases cubically with the number of parameters, leading to substantial memory and computational demands, particularly in models with many parameters. Furthermore, quantities such as the moments, marginal distributions, and quantiles of the SKS approximation are not available in closed form, even in the scalar case.

To address these challenges, [6] proposed a class of approximations based on the standard skew-normal (SN) distribution. Their method matches posterior derivatives, aiming to preserve the ability to model skewness while employing more computationally tractable structures. It utilizes local information around the MAP by matching the mode  $m$ , the negative Hessian at the mode, i.e.,  $j(\tilde{\theta})$ , and the third-order unmixed derivatives vector  $t \in \mathbb{R}^d$  of the log-posterior. Moreover, modern tools for automatic differentiation can greatly facilitate the computation of such higher-order derivatives without manual derivation. The goal is to find the parameters of the multivariate SN distribution  $\text{SN}_d(\xi, \Omega, \alpha)$  that best match these quantities. The notation  $\text{SN}_d(\xi, \Omega, \alpha)$  indicates a  $d$ -dimensional SN distribution (see e.g., [24], and references therein), with location parameter  $\xi$ , scale matrix  $\Omega$ , and shape parameter  $\alpha$ . The matching equations are given by

$$\begin{aligned} 0 &= -\Omega^{-1}(m - \xi) + \zeta_1(\kappa)\alpha, \\ j(\tilde{\theta}) &= \Omega^{-1} - \zeta_2(\kappa)\alpha\alpha^\top, \\ t &= \zeta_3(\kappa)\alpha^{\circ 3}, \\ \kappa &= \alpha^\top(m - \xi), \end{aligned}$$

where  $\zeta_k(\kappa)$  denotes the  $k$ -th derivative of  $\log \Phi(\kappa)$ , and  $\circ 3$  represents the Hadamard (element-wise) product. The solution proceeds by reducing the system to a one-dimensional root-finding problem in  $\kappa$ , after which  $\alpha$ ,  $\Omega$ , and  $\xi$  can be obtained analytically. Ultimately, the marginal distributions are available in closed form as well. Given its tractability, we adopt the derivative matching approach proposed by [6] to derive SKS approximations for models with multidimensional parameters. For the SN model, we can instead easily define the multivariate quantiles.

As suggested in [8,9], an effective approach to defining quantiles in the multidimensional case is to identify the optimal transport (OT) map between the spherical uniform distribution and the target multivariate SN distribution. Considering the inherent relationship between the standard multivariate Gaussian distribution and the spherical uniform distribution, we explore the OT map linking a multivariate SN distribution to a multivariate standard normal distribution. Indeed, given a multivariate standard normal  $S$  in  $\mathbb{R}^d$ , it is well known that  $U = S/\|S\|$  is uniformly distributed on the sphere of radius  $\sqrt{d}$  in  $\mathbb{R}$ . Furthermore,  $2(\Phi(\|S\|) - 0.5)$  is uniform in  $(0,1)$ . Thus, the OT map and the quantiles of the multivariate standard Gaussian are coherently defined as a bijection of the norm of the multivariate standard normal vector  $S$  (the distance from the origin). In particular, we utilize the canonical multivariate SN distribution, derived from applying a rotational transformation, and we consider a component-wise transformation using the univariate SN distribution function and the standard normal quantile function, which delineates a transport map represented as the gradient of a convex function.

From  $X \sim \text{SN}_d(\xi, \Omega, \alpha)$ , let  $\delta = \Omega(\alpha/\sqrt{1 + \alpha^\top \Omega \alpha})$ . We define a rotation  $T_1(X) = QX$  by means of the matrix  $Q \in \mathbb{R}^{d \times d}$  such that

- $Z = Q^\top(X - \xi)$  aligns the skewness with the first coordinate;
- in the rotated space,  $Z_1 \sim \text{SN}_1(0, \omega^2, \|\alpha\|)$ , with  $\omega^2 = [Q^\top \Omega Q]_{1,1}$ , and  $Z_{2:d}$  are Gaussian.

The matrix  $Q$  is obtained by applying a (rectangular) QR decomposition to the  $\alpha$  vector. The vector of means is  $E(Z) = Q^\top \delta \sqrt{2/\pi}$  and the covariance matrix is  $V = Q^\top(\Omega - \frac{2}{\pi}(Q^\top \delta)^\top Q^\top \delta)$ . Moreover, the scale parameter of  $Z_1$  is  $\sigma = \sqrt{Q^\top \Omega Q}$  and we denote its mean and variance as  $\mu_1 = E[Z_1]$  and  $V_1 = \text{Var}(Z_1)$ , respectively.

We define the transport map  $T_2(X)$  in the rotated space as

$$T_2(X) = \begin{bmatrix} \Phi^{-1}(F_{SN}(X_1, 0, \sigma^2, Q^\top \alpha), \mu_1, V_1) \\ X_2 \\ \vdots \\ X_d \end{bmatrix},$$

where  $F_{SN}(\cdot)$  is the univariate SN cumulative distribution function and  $\Phi^{-1}(\cdot)$  is the standard normal quantile function. In practice, we transform the first component using the univariate SN cumulative distribution function ( $F_{SN}$ ) and the standard normal quantile function ( $\Phi^{-1}$ ) to remove its skewness, while leaving other components unchanged. Note that the SN distribution is closed under linear transformations. In particular, after the rotation, the skewness of the variable  $Z$  becomes  $Q^\top \alpha$  (see [24]). The variable  $Z' = T_2(Z)$  is now approximately multivariate normal. Finally, we apply an affine transformation to standardize the result. More precisely, we consider  $T_3(X) = V^{-1/2}(X - Q^\top \delta \sqrt{2/\pi})$ , and set  $U = T_3(Z')$ . The resulting  $U$  is distributed as a standard normal (see Figure 1).

It follows that, using the SN approximation  $\pi_{SN}(\theta|y)$  for the posterior distribution of  $\theta$ , the SN approximation of the BDM can be expressed as

$$\delta_H^{SN} = 1 - \Pr(\chi_d^2 \geq \|T(\theta_0)\|), \tag{27}$$

where  $T(x) = T_3 \circ T_2 \circ T_1$ . The map  $T(x) = T_3 \circ T_2 \circ T_1$  is the OT map as it is the gradient of a convex function. In particular,  $T_1$  and  $T_3$  are affine transformations, and the function  $\Phi^{-1}(F_{SN}(z, \zeta, \omega, \alpha))$  is monotonically increasing in  $z$ , hence its integral is convex. Defining

$$g(Z) = \int_0^{Z_1} \Phi^{-1}(F_{SN}(t, \zeta, \omega, \alpha)) dt + \frac{1}{2} \sum_{i=2}^d Z_i^2,$$

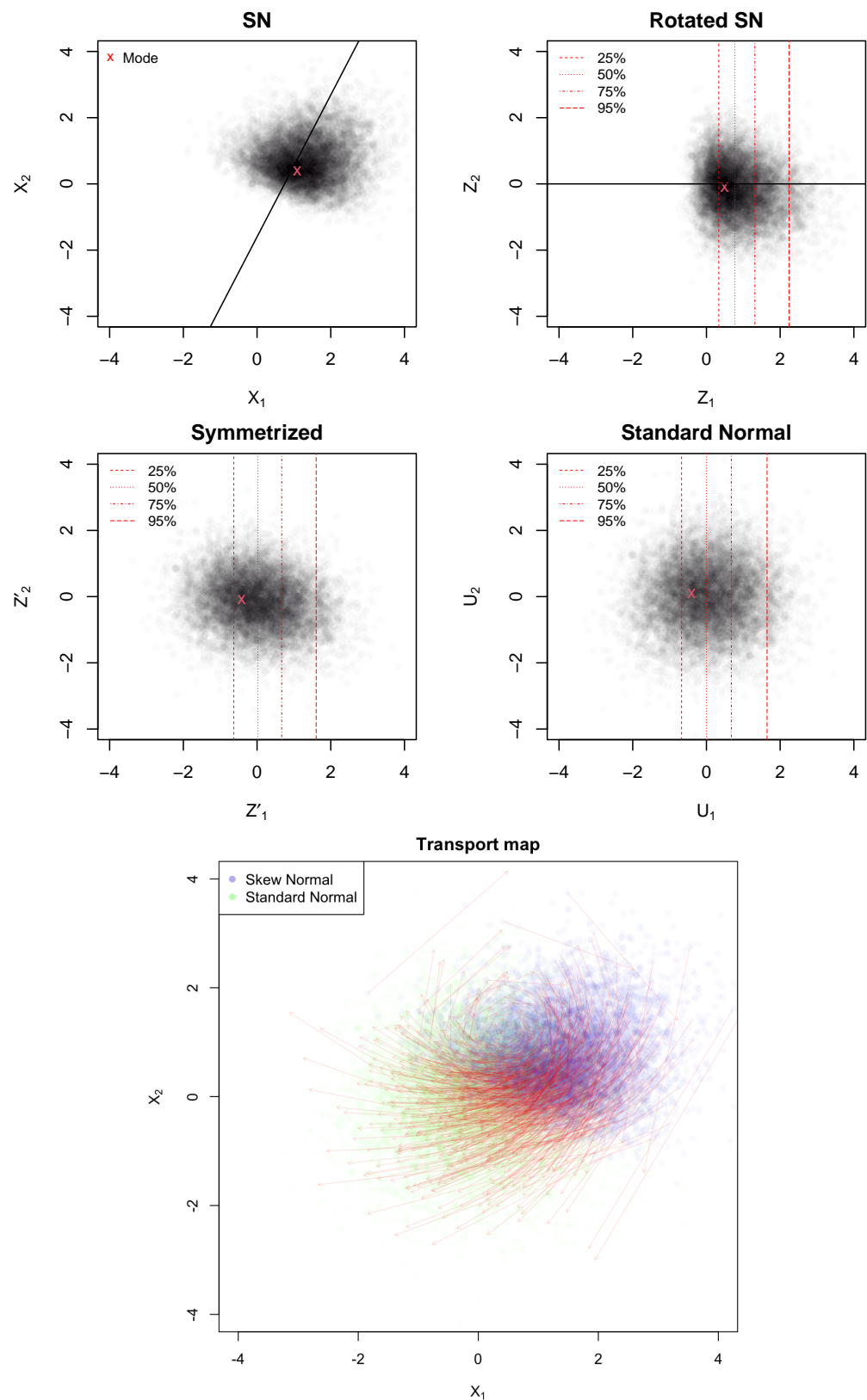
then  $T_2(Z) = \nabla g(Z)$ . The composite map  $T(\cdot)$ , used in (27), is the gradient of a convex function and, thus, it represents the optimal transport map (under quadratic cost) from an SN distribution to a standard normal. The procedure is summarized in Algorithm 1.

---

**Algorithm 1** Optimal transport from  $SN_d(\zeta, \Omega, \alpha)$  to  $\mathcal{N}_d(0, I_d)$

---

- Input**  $X \in \mathbb{R}^d \sim SN_d(\zeta, \Omega, \alpha)$
  - Output**  $U \sim \mathcal{N}_d(0, I_d)$
  - Compute  $\delta \leftarrow \Omega \alpha / \sqrt{1 + \alpha^\top \Omega \alpha}$
  - Compute QR decomposition:  $Q \leftarrow \text{QR\_decomposition}(\alpha)$
  - Compute mean and covariance:  $\mu_Z \leftarrow Q^\top \delta \cdot \sqrt{2/\pi}$ ,  $V \leftarrow Q^\top \left( \Omega - \frac{2}{\pi} (Q^\top \delta)(Q^\top \delta)^\top \right) Q$
  - Compute variance of the rotated first component:  $\sigma^2 \leftarrow [Q^\top \Omega Q]_{1,1}$
  - Set  $\alpha_{\text{rot}} \leftarrow Q^\top \alpha$
  - Set  $\mu_1 \leftarrow (\mu_Z)_1$  and  $V_1 \leftarrow V_{1,1}$
  - Apply rotation:  $Z \leftarrow Q^\top (X - \zeta)$  {T1}
  - Compute  $u \leftarrow F_{SN}(Z_1; 0, \sigma^2, \alpha_{\text{rot}})$  Compute  $z'_1 \leftarrow \Phi^{-1}(u; \mu_1, V_1)$
  - Set  $Z' \leftarrow (z'_1, Z_2, \dots, Z_d)$  {T2}
  - Compute  $U \leftarrow V^{-1/2}(Z' - \mu_Z)$  {T3}
-



**Figure 1.** *First panel:* Original SN approximation of a bivariate posterior distribution, with the mode in red and skewness direction indicated by the black line. *Second panel:* Rotated SN distribution aligning the skewness with the first coordinate; red dashed lines show quantiles of the first rotated component. *Third panel:* Symmetrized distribution after applying a univariate marginal transformation. *Fourth panel:* Final standardized and centered normal distribution. *Bottom panel:* Visualization of the optimal transport (OT) map (red arrows).

### 5. Examples of Higher-Order and Skewed Approximations

In the following, we focus on assessing the performance of the higher-order approximations and of the skewed approximations of the BDM in two examples, discussed also in [5,7].

#### 5.1. Exponential Model

We revisit Example 1 in [7], where the model for data  $y_1, \dots, y_n$  is an exponential distribution with scale parameter  $\theta$ , meaning  $E(Y) = \theta$ . By employing Jeffrey’s prior, which is  $\pi(\theta) \propto \theta^{-1}$ , the resulting posterior distribution is an inverse gamma, characterized by shape and rate parameters equal to  $n$  and  $t_n$ , respectively, with  $t_n = \sum_i y_i$ . The quantities for the SKS approximation of the posterior distribution are available in [5] (see Section 3.1), while for the higher-order approximation, we have that  $q(\theta)$  coincides with the score statistic, i.e.,  $q(\theta) = \ell^{(1)}(\theta)/i(\theta)^{1/2}$ . We analyze how well the two approximations align with the true BDM under the growing sample size ( $n = 6, 12, 20, 40$ ) while keeping the MLE fixed at  $\hat{\theta} = 1.2$ . The MAP values are 1.03 ( $n = 6$ ), 1.11 ( $n = 12$ ), 1.14 ( $n = 20$ ), and 1.17 ( $n = 40$ ).

Figures 2 and 3 and Table 1 report the approximations of the BDM for several candidate values for  $\theta_0$ . In particular, the first-order (IO) approximation (5), the higher-order (HO) approximation (13), the SKS approximation (24), a direct numerical tail area calculation (SKS-num) of (23) and the SN approximation (27) are considered. Figures 2 and 3 also display the approximations to the corresponding posterior distributions, where the HO approximation is derived numerically by inverting the tail area. Also, note that the SKS approximation of the posterior distribution is not guaranteed to lie within the interval (0,1), so we practically bounded the BDM in this interval.

The results confirm that the HO and SKS approximations yield remarkable improvements over the first-order counterpart for any  $n$ . Moreover, they show that the HO approximation of the BDM is almost perfectly over-imposed on the true BDM, especially for values of  $\theta_0$  far from the MLE. When the value under the null hypothesis is closer to the MLE, the SKS approximation and the numerical tail areas derived from the SKS and SN approximations better approximate the true BDM. Furthermore, the SN approximation more accurately captures the tail behavior of the posterior distribution than the SKS approximation.

**Table 1.** BDM for a series of values  $\theta_0$  for the parameter and increasing sample sizes in the exponential example. The values of the true BDM and the best approximation(s) in each configuration are highlighted in bold.

	$\theta_0$	0.3	0.6	0.9	1.2	1.5	1.8	2.1	2.4
$n = 6$	IO	0.93	0.78	0.46	0.00	0.46	0.78	0.93	0.99
	HO	<b>1.00</b>	<b>0.96</b>	<b>0.62</b>	0.00	<b>0.30</b>	<b>0.57</b>	<b>0.73</b>	<b>0.83</b>
	SKS	<b>1.00</b>	1.00	0.80	0.20	0.32	0.73	0.94	0.99
	SKS-num	<b>1.00</b>	0.94	0.53	<b>0.07</b>	0.58	0.91	0.99	1.00
	SN	<b>1.00</b>	0.94	0.52	0.06	0.51	0.78	0.91	0.97
	<b>BDM</b>	<b>1.00</b>	<b>0.96</b>	<b>0.62</b>	<b>0.11</b>	<b>0.30</b>	<b>0.57</b>	<b>0.73</b>	<b>0.83</b>
$n = 12$	IO	0.99	0.92	0.61	0.00	0.61	0.92	0.99	1.00
	HO	<b>1.00</b>	<b>0.99</b>	<b>0.75</b>	0.00	<b>0.48</b>	<b>0.78</b>	<b>0.91</b>	<b>0.96</b>
	SKS	<b>1.00</b>	1.00	0.74	−0.00	0.61	0.91	0.99	1.00
	SKS-num	<b>1.00</b>	0.99	0.72	0.01	0.64	0.95	1.00	1.00
	SN	<b>1.00</b>	1.00	0.77	<b>0.04</b>	0.62	0.89	0.98	1.00
	<b>BDM</b>	<b>1.00</b>	<b>0.99</b>	<b>0.75</b>	<b>0.08</b>	<b>0.48</b>	<b>0.78</b>	<b>0.91</b>	<b>0.96</b>

Table 1. Cont.

	$\theta_0$	0.3	0.6	0.9	1.2	1.5	1.8	2.1	2.4
$n = 20$	IO	1.00	0.97	0.74	0.00	0.74	0.97	1.00	1.00
	HO	1.00	<b>1.00</b>	<b>0.85</b>	0.00	<b>0.62</b>	<b>0.90</b>	<b>0.97</b>	<b>0.99</b>
	SKS	1.00	<b>1.00</b>	0.91	<b>0.08</b>	0.66	0.96	1.00	1.00
	SKS-num	1.00	<b>1.00</b>	0.84	0.02	0.73	0.98	1.00	1.00
	SN	1.00	<b>1.00</b>	0.94	0.02	0.72	0.95	1.00	1.00
	<b>BDM</b>	<b>1.00</b>	<b>1.00</b>	<b>0.85</b>	<b>0.06</b>	<b>0.62</b>	<b>0.90</b>	<b>0.97</b>	<b>0.99</b>
$n = 40$	IO	1.00	1.00	0.89	0.00	0.89	1.00	1.00	1.00
	HO	1.00	1.00	<b>0.95</b>	0.00	<b>0.81</b>	<b>0.98</b>	1.00	1.00
	SKS	1.00	1.00	0.99	<b>0.05</b>	0.83	1.00	1.00	1.00
	SKS-num	1.00	1.00	0.96	0.03	0.87	1.00	1.00	1.00
	SN	1.00	1.00	1.00	0.02	0.87	0.99	1.00	1.00
	<b>BDM</b>	<b>1.00</b>	<b>1.00</b>	<b>0.95</b>	<b>0.04</b>	<b>0.81</b>	<b>0.98</b>	<b>1.00</b>	<b>1.00</b>

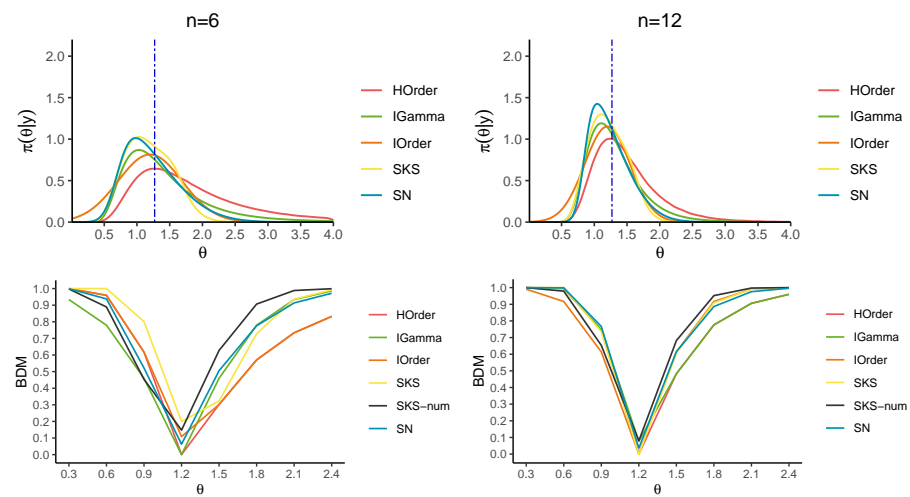


Figure 2. Exact posterior (in green) and approximate posteriors for  $n = 6, 12$  in the exponential model (top panels). The blue vertical line indicates the posterior median. BDM for a series of parameters (lower panels).

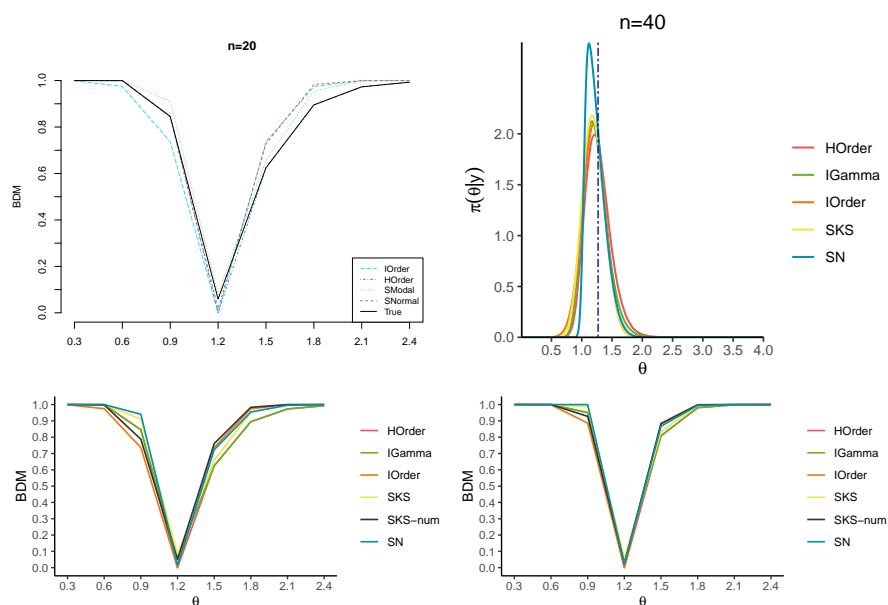


Figure 3. Exact posterior (in green) and approximate posteriors for  $n = 20, 40$  in the exponential model (top panels). The blue vertical line indicates the posterior median. BDM for a series of parameters (lower panels).

5.2. Logistic Regression Model

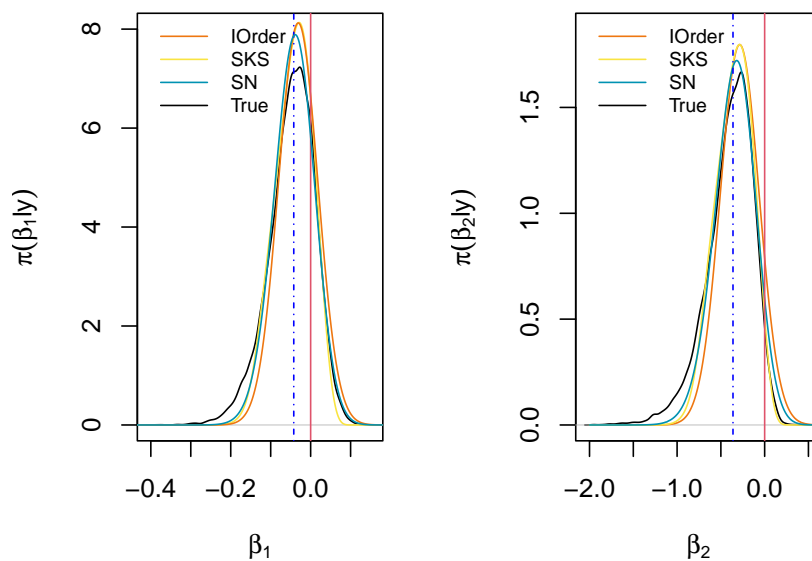
We now consider a real-data application using the Cushing’s dataset (see [5], Section 5.2), which is openly available in the R library MASS. The data were obtained from a medical study on  $n = 27$  individuals, aimed at investigating the relationship between Cushing’s syndrome and two steroid metabolites, namely Tetrahydrocortisone and Pregnanetriol.

We define a binary response variable  $Y$ , which takes the value 1 when the patient is affected by bilateral hyperplasia, and 0 otherwise. The two observed covariates  $x_1$  and  $x_2$  are two dummy variables representing the presence of the metabolites. We focus on the most popular regression model for binary data, namely, logistic regression with mean function  $\text{logit}^{-1}(\beta_0 + \beta_1x_1 + \beta_2x_2)$ . As in [5], Bayesian inference is carried out by employing independent, weakly informative Gaussian priors  $N(0, 25)$  for the coefficients  $\beta = (\beta_0, \beta_1, \beta_2)$ .

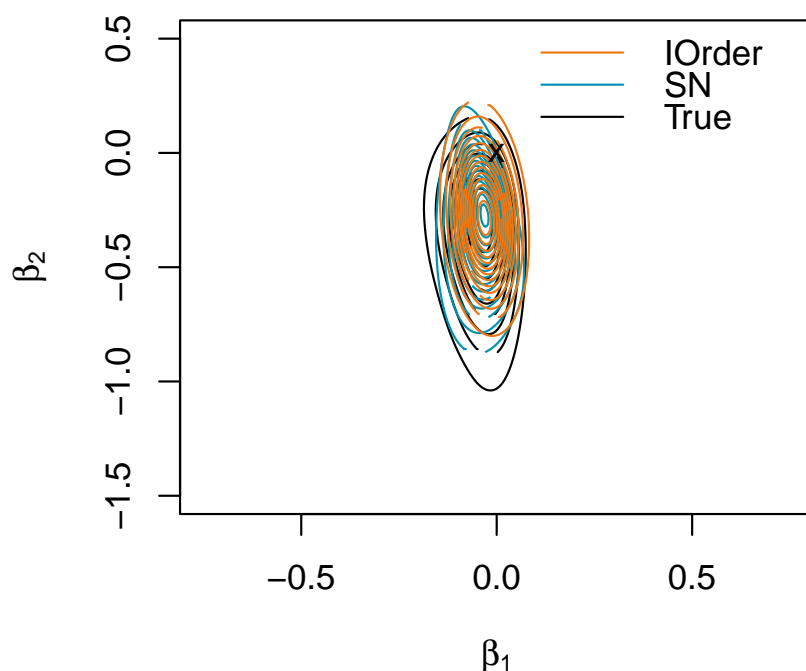
Figure 4 displays the marginal posterior distributions for  $\beta_1$  and  $\beta_2$  obtained via MCMC sampling (black curves), along with the first-order, the SKS, and the SN approximations. The MAP values for the two parameters are  $-0.031$  and  $-0.286$ , respectively.

We aim to test the two null hypotheses  $H_0 : \beta_1 = 0$  and  $H_0 : \beta_2 = 0$ , corresponding to the null effect of the metabolites’ presence in determining Cushing’s syndrome (indicated by red vertical lines in Figure 4). The exact BDM gives the values 0.592 and 0.932, respectively, indicating that the hypothesized value may support the null hypothesis for the first parameter  $\beta_1$ , whereas the second value suggests a weak disagreement with the assumed value for  $H_0 : \beta_2 = 0$ . The SKS approximations of the BDM for the considered hypotheses are 0.612 and 0.935, respectively; the SN approximations are 0.584 and 0.870, respectively; the first-order approximations are 0.512 and 0.891, respectively; while the higher-order approximations provide 0.611 and 0.998, respectively. Finally, the approximations based on the matching priors are 0.477 and 0.862, respectively. Thus, the skewed approximations (SKS, SN) provide the best results.

For the composite hypothesis  $H_0 : \beta_1 = \beta_2 = 0$ , the ground truth is not available; however, in the presence of low correlation between the components, one can approximate it as the geometric mean of the two marginal measures, which is 0.743. The first-order approximation for the BDM gives 0.300, while the SN approximation gives 0.760, revealing that the value under the null is more extreme (see also Figure 5).



**Figure 4.** Marginal posterior distributions for the regression parameters of the logistic regression example. The marginal medians are indicated in blue, while the parameters under the null hypothesis are indicated in red.



**Figure 5.** Joint posterior for  $(\beta_1, \beta_2)$  in the logistic regression example with the first-order (IOrder) and skew-normal (SN) approximations. The point  $(0,0)$  is marked with a cross.

## 6. Concluding Remarks

Although the higher-order and skewed approximations described in this paper are derived from asymptotic considerations, they perform well in moderate or even small sample situations. Moreover, they represent an accurate method for computing posterior quantities and approximating  $\delta_H$ , and they make it quite straightforward to assess the effect of changing priors (see, e.g., [25]). When using objective Bayesian procedures based on strong-matching priors and higher-order asymptotics, there is agreement between Bayesian and frequentist point and interval estimation, as well as in significance measures. This is not true, in general, with the  $e$ -value, as discussed in [14].

A significant contribution of this work is the extension to multivariate hypotheses. We propose a formal definition of the multivariate BDM based on center-outward optimal transport maps, providing a theoretically sound generalization of the univariate concept. By utilizing either the multivariate normal or multivariate SN approximations of the posterior distribution, we can formulate the multivariate quantiles in a closed form, thereby allowing us to derive the BDM for composite hypotheses. Nonetheless, precisely determining or defining these quantiles on the true posterior is challenging, as the transport map may not be available in a closed form and requires solving a complex optimization problem. However, the SN approximation as well as the derived OT map continue to be manageable in high-dimensional settings, whereas typical OT methods generally do not scale efficiently with increasing dimensions.

As a final remark, the high-order procedures proposed and described are tailored to continuous posterior distributions, and their extension to models with discrete or mixed-type parameters warrants further study. Moreover, although the higher-order and skewed methods, alongside SN-based OT maps, offer a useful means for approximating the posterior distributions and computing tail areas, their application might fail in handling complex or irregular posterior landscapes. In such cases, employing integrated computational procedures to find the transport map [26] and utilizing the direct definition of the multivariate BDM could be more appropriate. Furthermore, the utility of the higher-order and skew-normal approximation techniques developed here is not restricted to the Bayesian

discrepancy measure. These methods hold considerable promise for approximating other Bayesian measures of evidence. For example, applying these procedures to quantities like the e-value is a natural and compelling direction for future work.

**Author Contributions:** Conceptualization, E.B., M.M. and L.V.; Methodology, E.B., M.M. and L.V.; Software, E.B.; Validation, E.B. and L.V.; Formal analysis, E.B., M.M. and L.V.; Investigation, E.B. and L.V.; Data curation, E.B.; Writing—original draft, E.B., M.M. and L.V.; Writing—review & editing, E.B. and M.M.; Supervision, F.B., M.M. and L.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** Elena Bortolato acknowledges funding from the European Union under the ERC grant project number 864863.

**Institutional Review Board Statement:**

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations and Notation Glossary

The following abbreviations are used in this manuscript:

Symbol	Meaning / Definition
BDM	Bayesian discrepancy measure
BF	Bayes factor
LR	Log-likelihood ratio
MAP	Maximum a posteriori
MLE	Maximum likelihood estimate
OT	Optimal transport
SKS	Skew-symmetric
SN	Skew-normal
$H_0$	Sharp (precise) null hypothesis
$\theta$	Scalar parameter or parameter vector in the multivariate case
$\theta_0$	Specific hypothesized value of the parameter $\theta$ under $H_0$
$\psi$	Scalar parameter of interest
$\lambda$	Nuisance parameter (scalar or vector)
$y$	Observed data
$d$	Dimension of the full parameter vector
$n$	Sample size
$\hat{\theta}$	MLE of $\theta$
$\tilde{\theta}$	MAP of $\theta$
$\hat{\lambda}_\psi$	Constrained MLE of $\lambda$ given $\psi$
$\ell(\theta), \ell(\psi, \lambda)$	Log-likelihood function
$\ell_p(\psi)$	Profile log-likelihood function for $\psi$
$\ell^{(1)}(\theta), \ell_p^{(1)}(\psi)$	Score function, profile score function
$j(\theta), j_p(\psi)$	Observed information matrix, profile observed information
$j_{\psi\psi}, j_{\lambda\lambda}, j_{\psi\lambda}$	Submatrices of $j(\theta)$ for parameter partitions
$w(\theta), w_p(\psi)$	Wald statistic, profile Wald statistic
$s(\theta), s_p(\psi),$	Score statistic, profile score statistic
$W(\theta)$	Log-likelihood ratio statistic
$r(\theta), r_p(\psi)$	Likelihood root, profile likelihood root
$\pi(\theta), \pi(\theta   y)$	Prior density of $\theta$ , posterior density of $\theta$
$\delta_H$	Bayesian discrepancy measure, quantifying evidence against $H_0$
$r_B^*(\theta), r_{Bp}^*(\psi)$	Bayesian modified likelihood root statistic (scalar, with nuisance parameters)
$\Phi(\cdot)$	Standard normal cumulative distribution function

$\pi_m(\psi   y)$	Marginal posterior density of $\psi$
$\mathbf{F}_P^\pm$	Center-outward distribution function mapping posterior to the unit ball
$\mathbf{Q}_P^\pm$	Center-outward quantile function (inverse of $\mathbf{F}_P^\pm$ )
$U_d$	Uniform distribution on the unit ball $\mathbb{B}_d \subset \mathbb{R}^d$
$\mathbb{B}_d, \mathbb{S}^{d-1}$	Unit ball in $\mathbb{R}^d$ , unit sphere in $\mathbb{R}^d$
$\mathcal{R}_P^\pm(\tau), \mathcal{C}_P^\pm(\tau)$	Center-outward quantile regions and quantile contours of order $\tau$
$\ \cdot\ $	Euclidean norm
$\doteq, \dot{=}$	Approximate equality to first- or third-order (e.g., $O(n^{-1/2})$ or $O(n^{-3/2})$ accuracy)

## References

- Kass, R.E.; Tierney, L.; Kadane, J. The validity of posterior expansions based on Laplace's method. In *Bayesian and Likelihood Methods in Statistics and Econometrics*; Elsevier: Amsterdam, The Netherlands, 1990; pp. 473–488.
- Reid, N. Likelihood and Bayesian approximation methods. *Bayesian Stat.* **1995**, *5*, 351–368.
- Reid, N. The 2000 Wald memorial lectures: Asymptotics and the theory of inference. *Ann. Stat.* **2003**, *31*, 1695–1731. [[CrossRef](#)]
- Ventura, L.; Reid, N. Approximate Bayesian computation with modified log-likelihood ratios. *Metron* **2014**, *7*, 231–245. [[CrossRef](#)]
- Durante, D.; Pozza, F.; Szabo, B. Skewed Bernstein–von Mises theorem and skew-modal approximations. *Ann. Stat.* **2024**, *52*, 2714–2737. [[CrossRef](#)]
- Zhou, J.; Grazian, C.; Ormerod, J.T. Tractable skew-normal approximations via matching. *J. Statist. Comput. Simul.* **2024**, *94*, 1016–1034. [[CrossRef](#)]
- Bertolino, F.; Manca, M.; Musio, M.; Racugno, W.; Ventura, L. A new Bayesian discrepancy measure. *Stat. Methods Appl.* **2024**, *33*, 381–405. [[CrossRef](#)]
- Hallin, M.; Del Barrio, E.; Cuesta-Albertos, J.; Matrán, C. Distribution and quantile functions, ranks and signs in dimension  $d$ : A measure transportation approach. *Ann. Stat.* **2021**, *49*, 1139–1165. [[CrossRef](#)]
- Hallin, M.; Konen, D. Multivariate Quantiles: Geometric and Measure-Transportation-Based Contours. In *Applications of Optimal Transport to Economics and Related Topics*; Springer Nature: Cham, Switzerland, 2024; pp. 61–78.
- Madrugá, M.; Pereira, C.; Stern, J. Bayesian evidence test for precise hypotheses. *J. Stat. Plan. Inf.* **2003**, *117*, 185–198. [[CrossRef](#)]
- Pereira, C.; Stern, J.M. Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy* **1999**, *1*, 99–110. [[CrossRef](#)]
- 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. [[CrossRef](#)]
- Bertolino, F.; Columbu, S.; Manca, M.; Musio, M. Comparison of two coefficients of variation: A new Bayesian approach. *Commun. Stat. Simul. Comput.* **2024**, *53*, 6260–6273. [[CrossRef](#)]
- Ruli, E.; Ventura, L. Can Bayesian, confidence distribution and frequentist inference agree? *Stat. Methods Appl.* **2021**, *30*, 359–373. [[CrossRef](#)]
- Peyré, G.; Cuturi, M. Computational optimal transport: With applications to data science. *Found. Trends Mach. Learn.* **2019**, *11*, 355–607. [[CrossRef](#)]
- Fraser, D.A.S.; Reid, N. Strong matching of frequentist and Bayesian parametric inference. *J. Stat. Plan. Inf.* **2002**, *103*, 263–285. [[CrossRef](#)]
- Welch, B.L.; Peers, H.W. On formulae for confidence points based on integrals of weighted likelihoods. *J. Roy. Statist. Soc. B* **1963**, *25*, 318–329. [[CrossRef](#)]
- Barndorff-Nielsen, O.E.; Chamberlin, S.R. Stable and invariant adjusted directed likelihoods. *Biometrika* **1994**, *81*, 485–499. [[CrossRef](#)]
- Severini, T.A. *Likelihood Methods in Statistics*; Oxford University Press: Oxford, UK, 2000.
- Pierce, D.A.; Bellio, R. Modern likelihood-frequentist inference. *Int. Stat. Rev.* **2017**, *85*, 519–541. [[CrossRef](#)]
- Skovgaard, I.M. Likelihood Asymptotics. *Scand. J. Stat.* **2001**, *28*, 3–32. [[CrossRef](#)]
- Ventura, L.; Ruli, E.; Racugno, W. A note on approximate Bayesian credible sets based on modified log-likelihood ratios. *Stat. Prob. Lett.* **2013**, *83*, 2467–2472. [[CrossRef](#)]
- Tan, L.S.; Chen, A. Variational inference based on a subclass of closed skew normals. *J. Stat. Comput. Simul.* **2024**, *34*, 1–15. [[CrossRef](#)]
- Azzalini, A.; Capitanio, A. Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc. B* **1999**, *61*, 579–602. [[CrossRef](#)]

25. Reid, N.; Sun, Y. Assessing Sensitivity to Priors Using Higher Order Approximations. *Commun. Stat. Simul. Comput.* **2010**, *39*, 1373–1386. [[CrossRef](#)]
26. Li, K.; Han, W.; Wang, Y.; Yang, Y. Optimal Transport-Based Generative Models for Bayesian Posterior Sampling. *arXiv* **2025**, arXiv:2504.08214.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.