Imperial College London

MSC STATISTICS

Master Thesis

Analytic approximations in high-dimensional Bayesian inference

Author:
Daniel García Rasines

Supervisor:
Alastair Young

IMPERIAL COLLEGE LONDON

September 2017

The work contained in this thesis is my own work unless otherwise stated.

Signature:

I would like to thank Prof. A. Young for his guidance and constructive feedback during the completion of this work.

Abstract

The aim of this project is to present approximate analytic methods for Bayesian inference about a scalar parameter of interest. Approximations to posterior distributions that stem from Laplace's method are reviewed and their theoretical properties are discussed. These approximations are relatively straightforward to implement and are faster than available alternatives, but their high-dimensional performance is questionable. We illustrate the presented methodology with two examples and study its accuracy as the parameter dimensionality increases.

Contents

1 Introduction **4**

2 Analytic approximations in Bayesian inference **6**

2.1 Regularity conditions 6

2.2 Approximation of the marginal posterior density 9

2.3 Approximation of the marginal posterior distribution function 11

2.4 Sampling from the posterior 16

3 Assessing the effect of dimensionality **17**

3.1 Examples 17

3.2 Refined tail area approximation 22

4 Discussion **25**

A Proof of lemma 1 **26**

B Proof of lemma 2 **28**

C Proof of lemma 3 **29**

D Proof of lemma 4 **30**

E Code **32**

Chapter 1

Introduction

In this work we are concerned with inference about a scalar parameter of interest in a Bayesian framework. We consider a set of n observations $y = (y_1, \ldots, y_n)$ sampled from a parametric model with density $f(y;\theta)$ and parameter space $\Omega_\theta \subseteq \mathbb{R}^{p+1}$. The parameter is partitioned as $\theta = (\psi, \chi)$, with ψ one-dimensional and χ p-dimensional, and it is assumed that only the first component, ψ, is of interest, while χ has the role of nuisance parameter. For a given prior density $\pi(\theta)$, the posterior density of ψ conditioned on y is given by

$$\pi(\psi|y) = \frac{\int \pi(\psi,\chi)\exp\{l(\psi,\chi)\}d\chi}{\int \pi(\psi,\chi)\exp\{l(\psi,\chi)\}d\psi d\chi}, \tag{1.1}$$

where $l(\psi,\chi) = \log f(y;\psi,\chi)$ is the log-likelihood function and the domains of integration are $\{\chi\colon (\psi,\chi) \in \Omega_\theta\}$ in the numerator and Ω_θ in the denominator.

In many practical situations, analytic evaluation of the integrals involved in (1.1) is not possible, and numerical integration becomes infeasible as soon as $p \approx 4$. Markov Chain Monte Carlo methods (MCMC) are often used to overcome this problem; they provide a way of sampling from the posterior distribution of interest that avoids the need for complex integration. Unfortunately, they can require a long time to run before their samples are reliable, especially if the parameter dimension is large. In this work we explore an alternative approach to this problem based on accurate analytic approximations of the required integrals that is less time-consuming.

Application of Laplace's method of integration to (1.1) yields a very simple approximation of posterior density of ψ, which can in turn be integrated to produce equally simple approximations of the posterior distribution function. From a theoretical perspective, these approximations yield an error that decreases with the sample size at rate $n^{-3/2}$. In practice, various examples show that they can be very accurate even if n is small [18, 19], but their performance in high-dimensional settings is still unclear. A complete summary of analytic approximations in Bayesian inference can be found in [22].

This project has two main objectives. The first one is to introduce the analytic approximations and do a formal verification of their asymptotic properties. The literature on the topic contains heuristic derivations of these properties, and our aim here is to validate them under a suitable set of regularity conditions. We will do this in chapter 2. The second objective is to study the accuracy of the approximations with regards to the dimensionality of the nuisance parameter. This will be done in chapter 3 through the analysis of two particular

models.

We now introduce the notation. As before, the log-likelihood based on y is denoted by $l(\theta; y)$ or $l(\theta)$, as appropriate. Also, $(\hat{\psi}, \hat{\chi})$ denotes the maximum likelihood estimator (MLE) of (ψ, χ); $\hat{\chi}_\psi = \text{argmax}_\chi\, l(\psi, \chi)$ the constrained maximum likelihood estimator of χ given ψ; $j(\psi, \chi)$ the observed Fisher information matrix, with (i, j)th entry $-(\partial^2/\partial\theta_i\partial\theta_j)l(\psi, \chi)$; $j_{\chi\chi}(\psi, \chi)$ the $p \times p$ submatrix of $j(\psi, \chi)$ corresponding to the nuisance parameter; $l_p(\psi) = l(\psi, \hat{\chi}_\psi)$ the profile log-likelihood of ψ; and $j_p(\psi)$ the second derivative of $-l_p(\psi)$. In addition, we define the following quantities, which will be of great importance:

$$
\begin{aligned}
r(\psi) &= \text{sign}(\psi - \hat{\psi}) \left\{ 2(l_p(\hat{\psi}) - l_p(\psi)) \right\}^{1/2}; \\
q(\psi) &= -l_p'(\psi)|j_p(\hat{\psi})|^{-1/2} \frac{|j_{\chi\chi}(\psi, \hat{\chi}_\psi)|^{1/2}}{|j_{\chi\chi}(\hat{\psi}, \hat{\chi})|^{1/2}} \frac{\pi(\hat{\psi}, \hat{\chi})}{\pi(\psi, \hat{\chi}_\psi)}; \\
r^*(\psi) &= r(\psi) + \frac{1}{r(\psi)} \log\left(\frac{q(\psi)}{r(\psi)} \right).
\end{aligned}
$$

The first quantity is the signed root log-likelihood ratio, with the sign chosen for notational convenience. Regarded as a random variable, its posterior distribution is approximately normal. It constitutes the basis of the analytic approximations. The second one is a modification of the standardized profile score that is asymptotically equivalent to $r(\psi)$ and includes the prior information. It stems from the application of Laplace's method to the posterior density. The last quantity is a modification of the first one that as we shall see improves on the accuracy of the normal approximation. The dependence of all the previous quantities on the number of observations n is deliberately supressed to avoid heavy notation. However, when analysing the presented methods, we will be interested in their behaviour as $n \to \infty$, and hence the expressions involved should be thought of in a sequential way as the number of observations grows. Finally, another piece of notation that we will use frequently is the so called big-O notation. We will say that a real sequence $\{x_n\}$ is $O(n^a)$ for some $a \in \mathbb{R}$ if the sequence $\{|n^{-a}x_n|\}$ is bounded. Similarly, we will say that a sequence of real functions $\{g_n\}$ is $O(n^a)$ in the sets $\{R_n\}$ if $\{g_n(x_n)\}$ is $O(n^a)$ for any sequence $\{x_n\}$ such that $x_n \in R_n$ for all n.

Chapter 2

Analytic approximations in Bayesian inference

In this chapter we review higher-order analytic approximations of the density and distribution function of the interest parameter and establish their asymptotic accuracy under a set of mild regularity conditions. Heuristic derivations of some of the results can be found in [6, 7, 22]. One of our main aims here is to justify them formally. In order to facilitate the reading, the least illustrative parts of the derivations have been left for the appendix. In the last section of the chapter we show how approximate samples from the posterior distribution of interest can be obtained using the presented approximations.

2.1 Regularity conditions

Throughout this work we will assume that our model meets various regularity conditions. The requirements are very similar to those given by Kass *et al.* in [12] to ensure the validity of Laplace's method, which is the starting point of the presented approximations. It is claimed in [19] that most typical parametric models satisfy these assumptions. Here we show how to verify them for two important cases: linear regression models and members of full exponential families.

The conditions are the following: the parameter space factorizes as $\Omega_\theta = \Omega_\psi \times \Omega_\chi$, with Ω_ψ and Ω_χ open and connected; $\pi(\theta)$ is positive; $\pi(\theta)$ and $l(\theta)$ are six times continuously differentiable in Ω_θ; the posterior distribution based on $\pi(\theta)$ and $l(\theta)$ exists; and the log-likelihoods have a sequence of unique maxima $\{\hat{\theta}_n\} \equiv \{\hat{\theta}\}$ satisfying

(i) for some $\varepsilon > 0$ independent of n, all partial derivatives of $l(\theta)$ up to order 6 are $O(n)$ in $B_\varepsilon(\hat{\theta})$, the ε-ball centred at $\hat{\theta}$;

(ii) for some $M > 0$ independent of n, $|n^{-1} j(\hat{\theta})| > M$ for all n, where $|\cdot|$ denotes the determinant;

(iii) for all $0 < \delta < \varepsilon$, $B_\delta(\hat{\theta}) \subseteq \Omega_\theta$ and

$$\liminf_{n\to\infty} \inf_{\theta} \left\{ \frac{l(\hat{\theta}) - l(\theta)}{n} : \theta \in \Omega_\theta - B_\delta(\hat{\theta}) \right\} > 0.$$

Conditions **(i)** and **(ii)** are typically satisfied under independent and identically distributed (IID) sampling, since in this case the log-likelihood based on the full data is the sum of the log-likelihoods based on each individual observation. Condition **(iii)**, together with the other two, grants asymptotic normality of the posterior distribution [12]. An important consequence of these assumptions is the following.

Lemma 1. *If the previous regularity conditions are satisfied, for any $z > 0$ there exists a $\delta > 0$ such that, for any $\psi \in B_\delta(\hat{\psi})$, the constrained MLE $\hat{\chi}_\psi$ exists, is unique, $(\psi, \hat{\chi}_\psi) \in B_z(\hat{\theta})$, and the first six derivatives of $\hat{\chi}_\psi$ with respect to ψ are $O(1)$ in $B_\delta(\hat{\psi})$.*

Proof. See appendix A. □

An immediate consequence of this lemma is that a conditions analogous to **(i)** and **(ii)** also hold for the profile log-likelihood $l_p(\psi) = l(\psi, \hat{\chi}_\psi)$.

The following example shows how to verify conditions **(i)-(iii)** for the normal linear model.

Example 2.1.1. *Linear regression.* Consider the model $y_i = x_i^T \beta + \varepsilon_i$, with $x_i, \beta \in \mathbb{R}^p$ and $\varepsilon \sim N(0, \sigma^2)$ independently for $i = 1, 2, \ldots$ The parameters are β and σ^2, and x_i is assumed to be a known vector of observations. For this model conditions **(i)-(iii)** are satisfied provided the eigenvalues of $n^{-1}X^TX$ remain bounded and bounded away from zero as $n \to \infty$, where $X = (x_1, \ldots, x_n)^T$ is the design matrix. A proof can be found in [12] (pp. 479-480).

An obvious inconvenience of the previous requirements is that they involve an infinite sequence of observations, which makes them impossible to verify in practice. It is therefore useful to identify sampling models under which these conditions are met almost surely, so that any set of observations from them can be assumed to be valid. The following result, due to Kass *et al.* [12], is very helpful to this end.

Theorem 1. *Let $y_1, y_2, \ldots$ be a sequence of IID observations from a density $f(y; \theta)$ with respect to a σ-finite measure, sample space $\mathcal{Y} \subseteq \mathbb{R}^q$ and open parameter space $\Omega_\theta \subseteq \mathbb{R}^m$ satisfying*

1. *$\theta_1 \neq \theta_2$ implies $f(\cdot|\theta_1) \neq f(\cdot|\theta_2)$;*
2. *$f(y; \theta) > 0$ for all $(y, \theta) \in \mathcal{Y} \times \Omega_\theta$;*
3. *$f(y; \cdot)$ is six-times continuously differentiable for all y;*
4. *for every θ_0 there exists a neighbourhood $B_1(\theta_0)$, a positive integer r and a measurable real function $M_1(y_1, \ldots, y_r)$ such that for all $\theta \in B_1(\theta_0)$,*

$$l(\theta; y_1, \ldots, y_r) - l(\theta_0; y_1, \ldots, y_r) < M_1(y_1, \ldots, y_r)$$

and M_1 has finite expected value under θ_0;

5. *for every θ_0 there exists a neighbourhood $B_2(\theta_0)$ and a measurable real function $M_2(y)$ such that for all $\theta \in B_1(\theta_0)$ and all $1 \leq j_1, \ldots, j_d \leq m$, with $d \leq 6$,*

$$\left| \frac{\partial^d}{\partial\theta_{j_1} \cdots \partial\theta_{j_d}} l(\theta; y) \right| < M_2(y)$$

and M_2 has finite expected value under θ_0;

6. *for all θ_0, $|g''(\theta_0)| > 0$, where*

$$g(\theta) = \mathbb{E}_{\theta_0} \{l(\theta_0; y) - l(\theta; y)\}$$

and $g''(\theta)$ denotes the Hessian of $g(\theta)$.

7. *for all θ_0 every sequence of maxima of the log-likelihood is strongly consistent.*

*Then the sequence of log-likelihoods satisfies conditions **(i)-(iii)** with probability one.*

Proof. See [12] (p. 486). □

For a more generic result not requiring IID sampling we refer to [12] (pp. 483-484). To finish this section we demonstrate how theorem 1 can be applied to members of the exponential family.

Example 2.1.2. *Exponential family.* Suppose the observations $y_1, y_2, y_3, \ldots$ are sampled independently from a density with respect to a σ-finite measure given by

$$f(y; \theta) = h(y) \exp\left\{s^T\theta - K(\theta)\right\},$$

where $s^T \equiv (s_1(y), \ldots, s_m(y))$ is the vector of natural statistics. We assume that the parameter space is $\Omega_\theta = \{\theta \in \mathbb{R}^m \colon \int h(y) \exp\{s^T\theta\} < \infty\}$, that Ω_θ is non-empty and open, and that the representation of the model is minimal in the sense that the components of $s(Y)$ do not satisfy any linear constrain with probability one. Under this assumptions the model is identifiable (condition 1), the MLE exists and is unique, $K(\theta)$ is strictly convex and the natural statistics have a moment generating function given by $\exp\{K(\theta + t) - K(\theta)\}$. Conditions 2 and 3 are immediately satisfied. For 4, fix any $\theta_0 \in \Omega_\theta$ and let $\varepsilon > 0$ such that $B_\varepsilon(\theta_0) \subseteq \Omega_\theta$. For any $\theta \in B_\varepsilon(\theta_0)$ we have

$$\begin{aligned} l(\theta; y_1) - l(\theta_0; y_1) &= s^T(\theta - \theta_0) - (K(\theta) - K(\theta_0)) \\ &\leq \|s^T\|\varepsilon + \sup\{K(\theta) - K(\theta_0) \colon \theta \in B_\varepsilon(\theta_0)\} \\ &\equiv M_1(y_1), \end{aligned}$$

where $\|\cdot\|$ denotes the Euclidean norm. Since $K(\theta)$ is continuous, the supremum of the second line is finite. The natural statistics have moments of all orders, so condition 4 is satisfied. Condition 5 is shown in a similar way. For 6, we have that, for any $\theta_0 \in \Omega_\theta$,

$$g(\theta) = \mathbb{E}_{\theta_0}\{s(Y_1)\}^T(\theta_0 - \theta) - (K(\theta_0) - K(\theta)),$$

so

$$g''(\theta_0) = K''(\theta_0),$$

which is the covariance matrix of the natural statistics and therefore is positive definite by minimality. Finally, differentiating the log-likelihood and using the Strong Law of Large Numbers we see that, under θ_0, the sequence of MLEs satisfies

$$K'(\hat{\theta}) = \frac{\sum_{i=1}^n s(Y_i)}{n} \xrightarrow{as} \mathbb{E}_{\theta_0}\{s(Y_1)\} = K'(\theta_0).$$

Since $K(\theta)$ is strictly convex, its gradient $K'(\theta)$ is injective, and therefore

$$\hat{\theta} \xrightarrow{as} \theta_0.$$

2.2 Approximation of the marginal posterior density

We begin our review with an approximation of the posterior density of the interest parameter, originally suggested in [14, 15, 20]. The simplest way to derive it is by application of Laplace's method, which we shall now briefly describe, to the two integrals involved in its computation. Despite its relative simplicity, the resulting estimates approximate the true density with a relative error of order $O(n^{-1})$, and their renormalized versions achieve an error of order $O(n^{-3/2})$.

Laplace's method, introduced in [13], is a technique for approximating integrals of the form

$$\int_D h(x) \exp\{ng(x)\}dx$$

for suitable smooth real functions h and g defined on a region $D \subseteq \mathbb{R}^m$ and n a large positive number. The essential requirements of the method are that g achieves a unique maximum $\hat{x}$ in the interior of D, the Hessian of g at $\hat{x}$ is negative definite and that h is positive in a neighbourhood of $\hat{x}$. The key idea being that, under these conditions, the value of the integral is dominated by the behaviour of g near its maximum. The approximation can be informally derived as follows:

$$\begin{aligned}
\int_D h(x) \exp\{ng(x)\}dx &\doteq \int_D h(x) \exp\{ng(\hat{x}) + (n/2)(x-\hat{x})^T g''(\hat{x})(x-\hat{x})\}dx \\
&\doteq \int_{\mathbb{R}^m} h(x) \exp\{ng(\hat{x}) + (n/2)(x-\hat{x})^T g''(\hat{x})(x-\hat{x})\}dx \\
&= h(\hat{x}) \exp\{ng(\hat{x})\} \frac{(2\pi)^{m/2}}{n^{m/2} \left|-g''(\hat{x})\right|^{1/2}}.
\end{aligned}$$

The last step follows from integrating a normal density with mean $\hat{x}$ and covariance matrix $-g''(\hat{x})$. A detailed analysis shows that the relative error of the approximation is generally of order $O(n^{-1})$ [12]. For more information on the Laplace approximation see *e.g.* [5].

Following Tierney and Kadane [20], the posterior density of ψ can be approximated by estimating the numerator and the denominator of (1.1) with Laplace's method. The regularity conditions outlined in the previous section ensure that the previous requirements for the application of the method are satisfied in a neighbourhood of $\hat{\psi}$. Therefore, in this neighbourhood, the approximation can be applied to the required integrals taking $h(\theta) = \pi(\theta)$ and $g(\theta) = n^{-1} l(\theta)$. The approximated density is

$$\begin{aligned}
\hat{\pi}(\psi|y) &= \frac{1}{\sqrt{2\pi}} \exp\left\{l_p(\psi) - l_p(\hat{\psi})\right\} \frac{|j(\hat{\psi},\hat{\chi})|^{1/2}}{|j_{\chi\chi}(\psi,\hat{\chi}_\psi)|^{1/2}} \frac{\pi(\psi,\hat{\chi}_\psi)}{\pi(\hat{\psi},\hat{\chi})} \\
&= \frac{1}{\sqrt{2\pi}} j_p(\hat{\psi})^{1/2} \exp\left\{l_p(\psi) - l_p(\hat{\psi})\right\} \frac{|j_{\chi\chi}(\hat{\psi},\hat{\chi})|^{1/2}}{|j_{\chi\chi}(\psi,\hat{\chi}_\psi)|^{1/2}} \frac{\pi(\psi,\hat{\chi}_\psi)}{\pi(\hat{\psi},\hat{\chi})}. \qquad (2.1)
\end{aligned}$$

In the second equality we have used that $|j(\hat{\psi},\hat{\chi})| = j_p(\hat{\psi})|j_{\chi\chi}(\hat{\psi},\hat{\chi})|$ [17]. A similar approximation, asymptotically equivalent to (2.1), can be derived by setting $h(\theta) = 1$ and

$g(\theta) = n^{-1}\{l(\theta) + n \log \pi(\theta)\}$, provided the necessary conditions for the application of Laplace's method are satisfied [19]. In general, approximation (2.1) does not integrate to one, but it can be easily renormalized numerically. The following theorem, due to Kass *et al.* [12], establishes the accuracy of $\hat{\pi}(\psi|y)$ in bounded neighbourhoods of $\hat{\psi}$.

Theorem 2. *Under the regularity conditions of section 2.1, there exists an $\delta > 0$ such that*

$$\pi(\psi|y) = \hat{\pi}(\psi|y)\{1 + E(\psi)\}$$

where $E(\psi)$ is $O(n^{-1})$ in $B_\delta(\hat{\psi})$.

Proof. See [12], pp. 482-483. □

Under the assumed regularity conditions, the posterior distribution of ψ is asymptotically normal with variance of order $O(n^{-1})$. Therefore, for any fixed probability p, the p-th quantile of $\psi|y$ shrinks to $\hat{\psi}$ at rate $n^{-1/2}$, and consequently the accuracy of the approximation is only relevant in shrinking regions of the form $\{\psi - \hat{\psi} \leq O(n^{-1/2})\}$. By the previous theorem, the error of the approximations in these sets is $O(n^{-1})$. However, a careful analysis of the error term shows that, in these regions, the functional approximation is in fact accurate to third-order. The following result is due to Tierney *et al.* [21].

Lemma 2. *Under the regularity conditions of section 2.1, there exists a $\delta > 0$ such that*

$$\pi(\psi|y) = \hat{\pi}(\psi|y)\left\{1 + \frac{R_1(\psi)}{n} + \frac{R_2(\psi)}{n^2}\right\},$$

where $R_1(\psi)$ is differentiable in $B_\delta(\hat{\psi})$, and $R_1(\psi), R_1'(\psi)$ and $R_2(\psi)$ are $O(1)$ in $B_\delta(\hat{\psi})$.

Proof. See appendix B. □

We can now establish the third-order accuracy of the renormalized approximation.

Corollary 1. *Under the regularity conditions of section 2.1,*

$$\pi(\psi|y) \propto \hat{\pi}(\psi|y)\left\{1 + \frac{R(\psi)}{n^{3/2}}\right\},$$

where $|R(\psi)| \leq C_1\sqrt{n}|\psi - \hat{\psi}| + C_2/\sqrt{n}$ in $B_\delta(\hat{\psi})$ for some $C_1, C_2, \delta > 0$. In particular, the relative error of the renormalized approximation is of order $O(n^{-3/2})$ when $\psi = \hat{\psi} + O(n^{-1/2})$.

Proof. Writing the error as in lemma 2 and expanding $R_1(\psi)$ in a Taylor series gives, after some manipulation,

$$\begin{aligned}\pi(\psi|y) &= \hat{\pi}(\psi|y)\left\{1 + \frac{R_1(\hat{\psi})}{n} + \frac{R_1'(x_\psi)(\psi - \hat{\psi})}{n} + \frac{R_2(\psi)}{n^2}\right\} \\ &\propto \hat{\pi}(\psi|y)\left\{1 + \frac{R_1'(x_\psi)(\psi - \hat{\psi})}{n + R_1(\hat{\psi})} + \frac{R_2(\psi)}{n(n + R_1(\hat{\psi}))}\right\},\end{aligned}$$

where x_ψ lies between ψ and $\hat{\psi}$. Taking the same δ as in lemma 2 it is straightforward to conclude the result by bounding the terms inside the brackets. □

From a computational perspective approximation (2.1) is straightforward to compute as long as the parameter dimension is small to moderate, since its implementation only requires computing the MLE and constrained MLE at a given ψ and the Hessians at these points. For high-dimensional parameters complications could arise in two different ways: the modes may be difficult to locate, particularly if the sample size is small; and the error from approximating the determinant of the Hessians may be considerable, as the number of second-derivative computations increases quadratically with the dimension.

We conclude the section with an illustration of the method for a simple model, for which numerical algorithms are not needed.

Example 2.2.1. *Inverse Gaussian (density).* Suppose that n observations $y_1, \ldots, y_n$ are independently sampled from the Inverse Gaussian distribution with pdf

$$f(y_i; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi y_i^3}} \exp\left\{\frac{-\lambda(y_i - \mu^2)}{2\mu^2 y_i}\right\}, \qquad \mu, \lambda > 0.$$

We take λ to be the interest parameter and consider the non-informative prior proposed in [1], given by $\pi(\mu, \lambda) \propto (\mu^2\lambda)^{-1}$. One can easily see that for this model approximation (2.1) takes the form

$$\hat{\pi}(\lambda|y) = \frac{1}{2}\sqrt{\frac{n}{\pi}} \exp\left\{\frac{n}{2}\left(\log\left(\frac{\lambda}{\hat{\lambda}}\right) - \frac{\lambda}{\hat{\lambda}} + 1\right)\right\} \sqrt{\frac{\hat{\lambda}}{\lambda^3}},$$

where $\hat{\lambda}$ is the MLE of λ. Remarkably, the approximation does not depend on the whole sufficient statistic, given by the MLE of the full parameter $(\hat{\mu}, \hat{\lambda})$, even though its true posterior density does depend on it [1]. Figure 2.1 displays the posterior density of λ and its unnormalized and renormalized approximations for $n = 5$, $\hat{\mu} = 1$ and $\hat{\lambda} = 2$. As we can see the renormalized approximation is very accurate despite the small sample size.

2.3 Approximation of the marginal posterior distribution function

For many purposes one is usually more interested in the posterior distribution function of the interest parameter rather than in the density itself, as the former allows explicit computation of probabilities. Starting from approximation (2.1) it is possible to construct two higher-order approximations of the conditional distribution function of ψ with no additional computational cost. The first of these, which we shall refer to as the Lugannani-Rice (LR) approximation, was derived in [6, 7]. It is closely related to the so-called Lugannani-Rice approximation in a frequentist framework, hence the name. The second one will be referred to as the Barndorff-Nielsen (BN) approximation, as it was derived, again in a frequentist setting, by Barndorff-Nielsen in [2, 3, 4]. The approximations are, respectively,

$$\mathbb{P}(\psi \leq \psi_0|y) = \Phi(r_0) + \phi(r_0)\left(\frac{1}{r_0} - \frac{1}{q_0}\right) + O(n^{-3/2}); \tag{2.2}$$

$$\mathbb{P}(\psi \leq \psi_0|y) = \Phi(r_0^*) + O(n^{-3/2}); \tag{2.3}$$

for $\psi_0 = \hat{\psi} + O(n^{-1/2})$, where $r_0 = r(\psi_0)$, $q_0 = q(\psi_0)$, and $r_0^* = r^*(\psi_0)$. These approximations are frequently referred to as tail area approximations.

A common way of deriving (2.2) and (2.3) is by integration of $\hat{\pi}(\psi|y)$. The following lemma tells us that in doing so we only make an error of order $O(n^{-3/2})$.

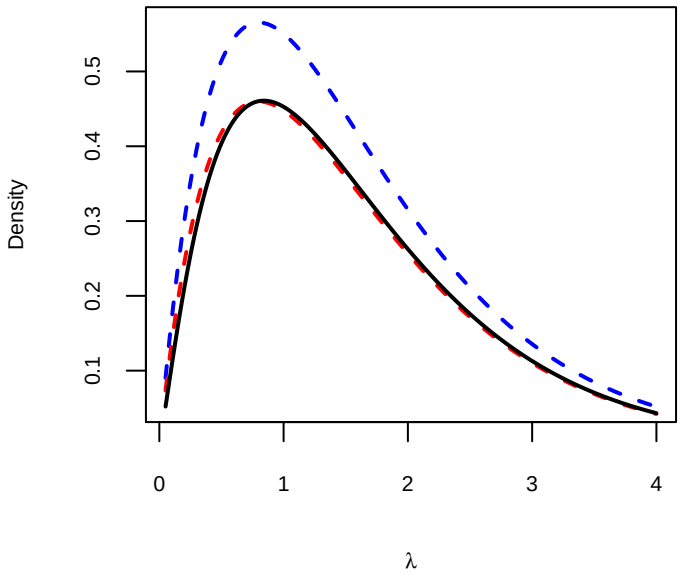

Figure 2.1: Posterior density of λ (solid black), unnormalized approximation (dashed blue) and renormalized approximation (dashed red) for $n = 5$, $\hat{\mu} = 1$ and $\hat{\lambda} = 2$.

Lemma 3. *Suppose $\Omega_\psi = (a, b)$. Under the regularity conditions of section 2.1, for $\psi_0 = \hat{\psi} + O(n^{-1/2})$,*

$$\int_a^{\psi_0} \pi(\psi|y)d\psi = \int_a^{\psi_0} \hat{\pi}(\psi|y)d\psi + O(n^{-3/2}).$$

Proof. See appendix C. □

Now, if $l_p(\psi) \to -\infty$ as $\psi \to a^+$ and r is a one-to-one function of ψ, a change of variables from ψ to r gives

$$\int_a^{\psi_0} \hat{\pi}(\psi|y)d\psi = K \int_{-\infty}^{r_0} \phi(r)\frac{r}{q(r)}dr,$$

where K is a normalizing constant. Note that $\psi_0 = \hat{\psi} + O(n^{-1/2})$ is equivalent to $r_0 = O(1)$. This follows because $r(\psi) = (\psi - \hat{\psi})j_p(x_\psi)^{1/2}$ for some x_ψ between ψ and $\hat{\psi}$. The desired approximations can then be obtained by considering the first terms of the asymptotic expansion of $r/q(r)$, which can be found in the following lemma.

Lemma 4. *Under the regularity conditions of section 2.1, the quotient $r/q(r)$ admits the expansion*

$$\frac{r}{q(r)} = 1 + p_1 r + p_2 r^2 + S(r),$$

where p_i is $O(n^{-i/2})$ and $n^{3/2}|S(r)|$ is bounded by $p(|r|)$ in the intervals $(-u\sqrt{n}, u\sqrt{n})$ for some $u > 0$ and some polynomial $p(r)$, both independent of n.

Proof. See appendix D. □

Using this results we can formally derive approximations (2.2) and (2.3).

Theorem 3. *Under the regularity conditions of section 2.1, if* $l_p(\psi) \to -\infty$ *as* $\psi \to a^+$ *and* r *is a one-to-one function of* ψ*, approximations* (2.2) *and* (2.3) *hold for* $\psi_0 = \hat{\psi} + O(n^{-1/2})$.

Proof. We need to evaluate the integral

$$K \int_{-\infty}^{r_0} \phi(r) \frac{r}{q(r)} dr,$$

where

$$K = \exp\{l_p(\hat{\psi})\} \frac{\pi(\hat{\psi}, \hat{\chi})}{|j_{\chi\chi}(\hat{\psi}, \hat{\chi})|^{1/2}} \left(\int_{\Omega_\psi} \exp\{l_p(\psi)\} \frac{\pi(\psi, \hat{\chi}_\psi)}{|j_{\chi\chi}(\psi, \hat{\chi}_\psi)|^{1/2}} d\psi \right)^{-1}.$$

Since the first part of the previous expression is the leading term of the Laplace approximation of integral in the last factor, we have that $K = 1 + O(n^{-1})$. Using lemma 4 and integration by parts, we also have

$$\begin{aligned}
\int_{-\infty}^{r_0} \phi(r) \frac{r}{q(r)} &= \int_{-\infty}^{r_0} \phi(r) \left\{1 + p_1 r + p_2 r^2 + S(r)\right\} dr \\
&= \int_{-\infty}^{r_0} \phi(r) \{1 + p_1 r + p_2 r^2\} dr + O(n^{-3/2}) \\
&= \Phi(r_0)(1 + p_2) + \phi(r_0)(-p_1 - p_2 r_0) + O(n^{-3/2}) \\
&= \Phi(r_0)(1 + p_2) + \phi(r_0) \left(\frac{1}{r_0} - \frac{1}{q_0} \right) + O(n^{-3/2}). \qquad (2.4)
\end{aligned}$$

The second step is follows from partitioning the integration domain into $(-\infty, -u\sqrt{n})$ and $(-u\sqrt{n}, r_0)$ with u chosen as in lemma 4, and noting that in the first interval the integral of $\phi(r)S(r)$ is $O(n^{-3/2})$ by lemma 2 of [12] and in the second one the integrand is $O(n^{-3/2})$ by lemma 4. To show the LR approximation it suffices to prove that

$$\begin{aligned}
K(1 + p_2) &= 1 + O(n^{-3/2}); \\
K \left(\frac{1}{r_0} - \frac{1}{q_0} \right) &= \left(\frac{1}{r_0} - \frac{1}{q_0} \right) + O(n^{-3/2}).
\end{aligned}$$

Letting $r_0 \to \infty$ in (2.4) yields $K^{-1} = 1 + p_2 + O(n^{-3/2})$, which shows the first equality. The second one follows immediately from the fact that $K^{-1} = 1 + O(n^{-1})$ and $r_0^{-1} - q_0^{-1} = O(n^{-1/2})$ for $r_0 = O(1)$.

The BN approximation can be proved by noting that it is related to the LR formula by a Taylor series. For $r_0 = O(1)$ we have

$$\begin{aligned}
\Phi(r_0^*) &= \Phi(r_0) + \phi(r_0) \frac{1}{r_0} \log \left(\frac{q_0}{r_0} \right) - \phi(r_0) \frac{1}{2r_0} \log \left(\frac{q_0}{r_0} \right)^2 + O(n^{-3/2}) \\
&= \Phi(r_0) + \phi(r_0) \frac{1}{r_0} \left\{ \log \left(\frac{q_0}{r_0} \right) - \frac{1}{2} \log \left(\frac{q_0}{r_0} \right)^2 \right\} + O(n^{-3/2}).
\end{aligned}$$

Since $\exp\{-x\} = 1 - x + x^2 + ...$ and $\log(q_0/r_0) = O(n^{-1/2})$ by lemma 4, we have that

$$\begin{aligned}\log\left(\frac{q_0}{r_0}\right) - \frac{1}{2}\log\left(\frac{q_0}{r_0}\right)^2 &= 1 - \exp\left\{-\log\left(\frac{q_0}{r_0}\right)\right\} + O(n^{-3/2}) \\ &= 1 - \frac{r_0}{q_0} + O(n^{-3/2}).\end{aligned}$$

Substituting in the Taylor expansion we conclude that, for $r_0 = O(1)$,

$$\Phi(r_0^*) = \Phi(r_0) + \phi(r_0)\left(\frac{1}{r_0} - \frac{1}{q_0}\right) + O(n^{-3/2}).$$

□

As we have just seen, both approximations agree asymptotically to third order. Regarding finite sampling performance, there is no clear consensus as to which one is more accurate. Often the BN version is favoured for a couple of reasons. First, its approximations are guaranteed to be actual probabilities, in that they lie in the interval $[0, 1]$. This is not necessarily the case for the LR formula. Another reason for preferring the BN version is that the r^* statistic can be used as a pivotal quantity to construct posterior credible intervals. Indeed, for any $0 < \alpha < 1$, if $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of a standard normal, then an approximate $(1 - \alpha)$ credible interval for ψ is given by $\{\psi\colon r^*(\psi) \in (-z_{\alpha/2}, z_{\alpha/2})\}$.

We now return to the Inverse Gaussian example to illustrate these approximations.

Example 2.3.1. *Inverse Gaussian (distribution function).* With the same setting as before, r and q admit the simple expressions

$$\begin{aligned}r(\lambda) &= \text{sign}(\lambda - \hat{\lambda})\left\{-n\left(\log\left(\frac{\lambda}{\hat{\lambda}}\right) - \frac{\lambda}{\hat{\lambda}} + 1\right)\right\}^{1/2}; \\ q(\lambda) &= (\hat{\lambda}^{-1} - \lambda^{-1})\left\{\frac{n\lambda^3}{2\hat{\lambda}}\right\}^{1/2}.\end{aligned}$$

Figure 2.2 contains the posterior distribution function of λ for the same choice of parameters as in example 2.2.1 together with its LR and BN approximations. As we can see they are virtually perfect, even though the sample size is very small.

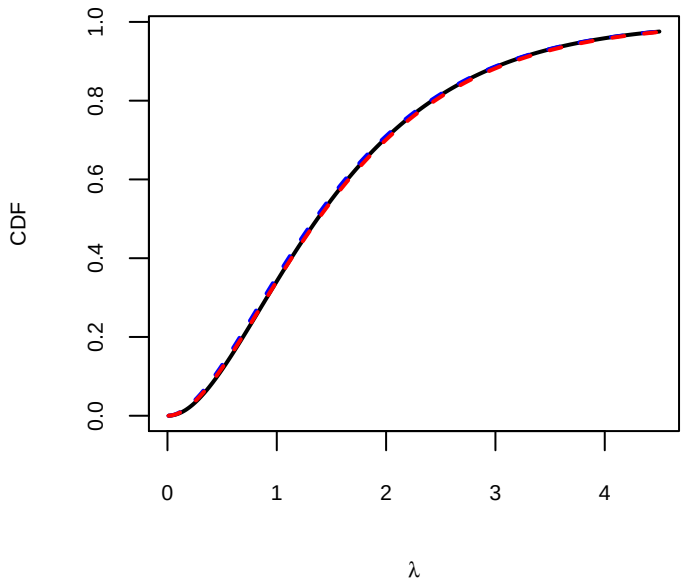

Figure 2.2: Posterior distribution function of λ (solid black), LR approximation (dashed blue) and BN approximation (dashed red) for $n = 5$, $\hat{\mu} = 1$ and $\hat{\lambda} = 2$.

2.4 Sampling from the posterior

Although the posterior density and distribution function may be of interest in their own right, statistical practice often requires computation of posterior summaries such as the posterior mean and variance of the interest parameter. Unfortunately, the approximations presented in this chapter do not generally admit closed-form expressions for these quantities. A possible solution would be to integrate numerically approximation (2.1), but as noted in [19], this has the drawback of requiring a large number of evaluations of the approximation, which can be time-demanding. Alternatively, Ruli *et al.* [19] propose an algorithm that allows efficient sampling from the approximate posterior distribution, from which Monte Carlo estimates of posterior quantities can be readily obtained. Their sampling scheme has a very the simple form. For $t \in \{1, 2, \ldots, T\}$:

1. Draw $z_t \sim N(0, 1)$;
2. Solve $r^*(\psi_t) = z_t$ and store ψ_t.

The resulting set $\{\psi_1, \psi_2, \ldots, \psi_T\}$ is a sample of size T from the higher-order approximation given by the BN formula. For the implementation of step 2, Ruli *et al.* [19] suggest an interpolation method that does not entail a big computational cost, requiring only a moderate number of evaluations of $r^*(\psi)$ independent of T.

The previous scheme has the attractive feature of producing independent samples, which means that the Monte Carlo error of the estimates decreases reasonably fast. This contrasts with MCMC methods, which may require a long time to produce accurate approximations due to dependence in the sample. On the negative side, this scheme produces samples from an approximation, and therefore it is not asymptotically exact. Furthermore, there is, to the best of our knowledge, no method for accurately quantifying the error of the approximation, which therefore cannot be controlled.

Chapter 3

Assessing the effect of dimensionality

In this chapter we study the accuracy of the tail area approximations introduced in section 1.3 as the dimension of the nuisance parameter increases. Until this point we have considered the usual asymptotic setting where p is kept fixed and n tends to infinity. This allowed us to investigate the performance of the approximations as the available sample becomes larger. Now we let p change as well as n, and assess the effect of this on the accuracy of the methods.

Problems involving high-dimensional parameters have become the norm in many areas of application, and the accuracy of techniques relying on Laplace's method is questionable in such contexts. Recall that Laplace's method (section 2.2) relies on the similarity of the posterior likelihood, $\pi(\theta)\exp\{l(\theta)\}$, with a normal density. Although this approximation is guaranteed to hold asymptotically under weak conditions, it may be very poor if p and n are of comparable size. Besides, computational difficulties can arise in the implementation of the approximations and numerical errors may be no longer be negligible.

As opposed to chapter 2, here we take an example-based approach. We consider two models for which the dimension of the nuisance parameters can be increased in a natural way. In both cases the posterior distribution of interest is available, either exactly or with very high accuracy, and the tail area approximations admit closed-form expressions. In the last section we introduce a refined approximation and study its performance on the considered models. The R code used in both examples can be found in appendix E.

3.1 Examples

The first of the examples is concerned with estimation of the common variance of several normal populations. In this case both the posterior distribution function and its higher-order approximations are available analytically.

Example 3.1.1. *Variance of normal populations.* Let $y_1, \ldots, y_n$ be n independent observations from the distribution $N_p(\mu, \sigma^2 I)$, where $\mu = (\mu_1, \ldots, \mu_p)^T$, $\sigma^2 > 0$ and I is the $p \times p$ identity matrix. Both μ and σ^2 are unkown, and we assume that σ^2 is of interest and that

the p-dimensional vector of means is a nuisance parameter.

Under the prior $\pi_{\sigma^2}(\sigma^2) \propto 1/\sigma^2$, $\pi_{\mu|\sigma^2}(\mu) \propto 1$, the marginal posterior distribution of σ^2 is Inverse Gamma with shape and scale parameters

$$\alpha = \frac{(n-1)p}{2};$$
$$\beta = \frac{np\hat{\sigma}^2}{2};$$

where $\hat{\sigma}^2$ is the MLE of σ^2. Its posterior density is therefore given by

$$\pi(\sigma^2|y) = \frac{\beta^\alpha}{\Gamma(\alpha)}\sigma^{2(-\alpha-1)}\exp\left\{-\frac{\beta}{\sigma^2}\right\}.$$

Also, we have

$$r(\sigma^2) = \text{sign}(\sigma^2 - \hat{\sigma}^2)\left\{np\left(\frac{\hat{\sigma}^2}{\sigma^2} - \log\left(\frac{\hat{\sigma}^2}{\sigma^2}\right) - 1\right)\right\}^{1/2};$$
$$q(\sigma^2) = \left(\frac{np}{2}\right)^{1/2}\left(1 - \frac{\hat{\sigma}^2}{\sigma^2}\right)\left(\frac{\hat{\sigma}^2}{\sigma^2}\right)^{p/2}.$$

The approximations are therefore straightforward to implement. Figure 3.1 contains plots of the true posterior distribution function together the BN approximation when $\hat{\sigma}^2 = 1$, $n = 5$, and $p = 50, 250, 500$. The LR approximation yielded very inaccurate results, with estimated probabilities consistently outside the interval $[0, 1]$ for moderate values of p, so is excluded from the study. As we can see, the BN approximation seems to get worse as the dimensionality of the nuisance parameter increases, but the effect is not catastrophic. This is ratified by the numerical evidence in table 3.1. For every fixed value of n the approximation loses accuracy as p grows, being particularly unreliable for $n = 3$.

Note that in this example the r^* statistic is a function of $\hat{\sigma}^2/\sigma^2$, and by standard properties of the Inverse Gamma distribution, the posterior distribution of σ^2 given $\hat{\sigma}^2$ satisfies

$$\sigma^2|\lambda\hat{\sigma}^2 = \lambda(\sigma^2|\hat{\sigma}^2)$$

for all $\lambda > 0$, with equality in law. As a consequence, the results of the table are in fact independent of $\hat{\sigma}^2$.

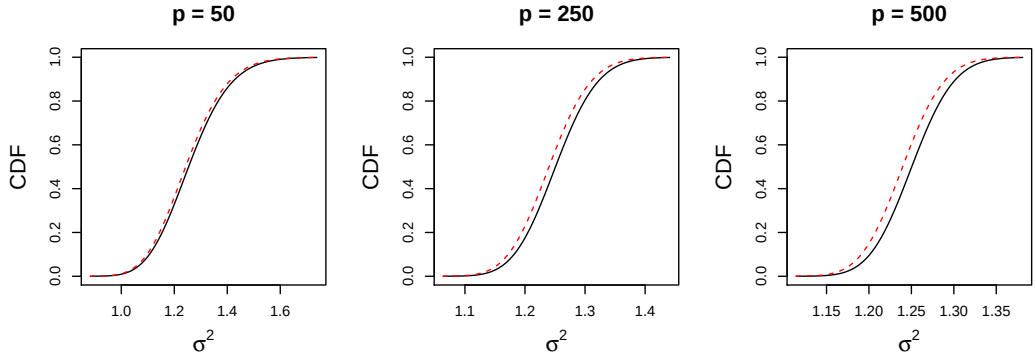

Figure 3.1: In solid black, posterior CDF of σ^2 for $\hat{\sigma}^2 = 1$, $n = 5$, and $p = 50, 250, 500$. In dashed red, its BN approximation.

	$p = 50$								
Prob.	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99
$n = 3$	0.017	0.075	0.141	0.321	0.585	0.815	0.934	0.969	0.995
$n = 10$	0.011	0.053	0.105	0.259	0.512	0.759	0.905	0.953	0.991
$n = 17$	0.010	0.051	0.102	0.254	0.505	0.754	0.902	0.951	0.990

	$p = 250$								
Prob.	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99
$n = 3$	0.030	0.117	0.205	0.416	0.680	0.874	0.961	0.983	0.998
$n = 10$	0.012	0.057	0.112	0.271	0.525	0.770	0.911	0.956	0.992
$n = 17$	0.011	0.053	0.105	0.259	0.511	0.759	0.905	0.953	0.991

	$p = 500$								
Prob.	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99
$n = 3$	0.046	0.159	0.264	0.492	0.745	0.910	0.974	0.990	0.999
$n = 10$	0.013	0.060	0.117	0.279	0.536	0.778	0.915	0.959	0.992
$n = 17$	0.011	0.054	0.107	0.263	0.516	0.762	0.907	0.954	0.991

Table 3.1: BN approximation at true posterior quantiles of several orders for $\hat{\sigma}^2 = 1$, $n = 3, 10, 17$, and $p = 50, 250, 500$.

In the second example we consider again several normal populations with common variance, but we take the quantity of interest to be the length of the vector of means. As before, closed-form expressions for the tail area approximations are straightforward to derive, but the true posterior distribution of interest is no longer tractable analytically.

Example 3.1.2. *Normal circle.* For this example we consider the same setting as Reid and Sun in [18] (section 3). The sampling model is identical to example 3.1.1, with the simplification that the variance of the populations is assumed to be known and equal to 1.

In this case we take the quantity of interest to be length of μ, this is, $\psi = \|\mu\|$. Since ψ is not a coordinate of the model parameter, a change of variables is needed prior to the implementation of the tail area approximations. The most natural choice is to consider p-dimensional spherical coordinates $(\psi, \phi_1, \ldots, \phi_{p-1})$, where

$$\phi_i = \arccos \frac{\mu_i}{\sqrt{\sum_{j=i}^{n} \mu_j^2}}, \quad i = 1, 2, \ldots, p-1.$$

The prior distribution is also different from the one in example 3.1.1. As mentioned in [18], the flat prior on the vector of means is actually very informative for inference about its length. For the latter task we can use the improper prior $\pi(\mu) \propto \|\mu\|^{-(p-1)}$, which yields a flat prior for ψ. Note that the Jacobian of the transformation from μ to $(\psi, \phi_1, \ldots, \phi_{p-1})$ is given by

$$\psi^{p-1} g(\phi_1, \ldots, \phi_{p-1})$$

for some function $g(\cdot)$, so the prior density of the transformed parameter is proportional to $g(\phi_1, \ldots, \phi_{p-1})$, which does not depend on ψ.

To derive the LR and BN formulas, we note that the log-likelihood for μ is given, up to an additive constant, by

$$\begin{aligned} l(\mu) &= -\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^T(y_i-\mu) \\ &= -\frac{1}{2}\left(\sum_{i=1}^{n}\|y_i\|^2 + n\|\mu\|^2 - 2n\|\mu\|\|\hat{\mu}\|\cos\{\varphi(\mu)\}\right), \end{aligned}$$

where $\varphi(\mu)$ is the angle between μ and $\hat{\mu}$. Clearly, $\varphi(\mu)$ is independent of $\|\mu\|$, in the sense that $\varphi(c\mu) = \varphi(\mu)$ for any $c > 0$, and therefore it is a function of the angles $(\phi_1, \ldots, \phi_{p-1})$ only. As a consequence, the constrained MLE of $(\phi_1, \ldots, \phi_{p-1})$ is independent of ψ, and by the invariance property of the MLE we have

$$\cos\{\varphi(\mu(\hat{\psi}, \hat{\phi}_1, \ldots, \hat{\phi}_{p-1}))\} = \cos\{\varphi(\hat{\mu})\} = 1.$$

Hence, the profile log-likelihood for ψ is

$$\begin{aligned} l_p(\psi) &= K_1 - \frac{n}{2}\left(\psi^2 - 2\psi\|\hat{\mu}\|\right) \\ &= K_2 - \frac{n}{2}(\psi - \|\hat{\mu}\|)^2, \end{aligned}$$

for some constants K_1 and K_2. Letting $\hat{\psi} = \|\hat{\mu}\|$, it is easy to see that

$$\begin{aligned} r(\psi) &= \sqrt{n}(\psi - \hat{\psi}); \\ q(\psi) &= \sqrt{n}(\psi - \hat{\psi})\left(\frac{\psi}{\hat{\psi}}\right)^{(p-1)/2}. \end{aligned}$$

In order to assess the accuracy of the approximations we estimated the posterior distribution of ψ via Metropolis Hastings algorithm with a large number of samples. In figure 3.2 we see that the three approximations agree very closely when $\hat{\psi} = 1$, $n = 20$ and $p = 5$. However, when p is increased to 10, their shape is very different, and the higher-order approximations behave rather poorly as $\psi \to 0$.

In this model, the accuracy of the LR and BN formulas is comparable and there is no clear favourite, as can be appreciated in table 3.2. The MCMC estimates displayed on the table were ensured to have an estimated standard error not exceeding 2×10^{-3}, so their first two decimal places can be trusted with high confidence. This limited the combinations of n and p considered, as the number of samples needed to meet the condition increases significantly for small values of n and large values of p. As in the previous example, it is apparent that the accuracy of the analytic approximations is negatively affected by the dimensionality of the nuisance parameter.

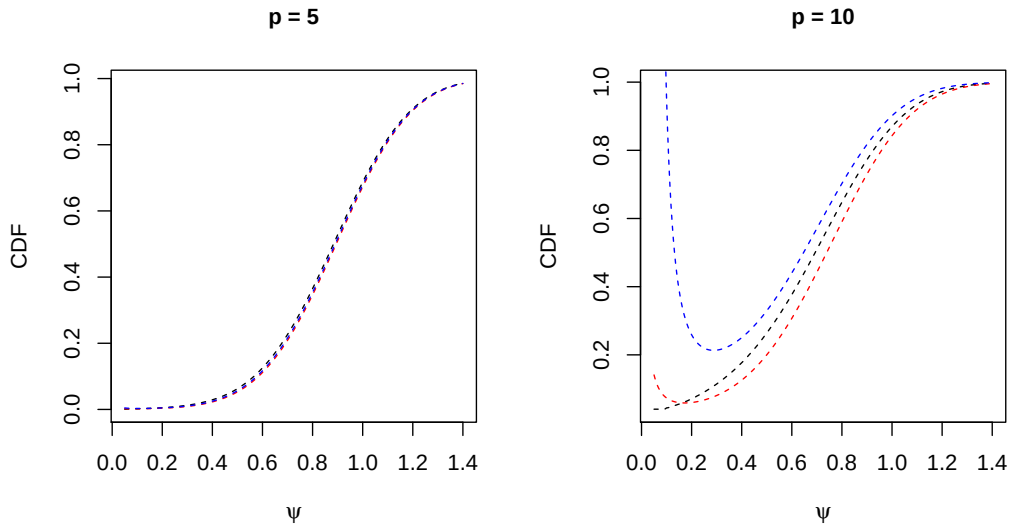

Figure 3.2: In dashed black, MCMC approximation of the posterior distribution of ψ for $\hat{\psi} = 1$, $n = 20$, and $p = 5, 10$; in dashed blue, LR approximation; in dashed red, BN approximation.

	method	$p = 5$								
	MCMC	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99
$n = 15$	LR	0.025	0.044	0.087	0.229	0.479	0.737	0.894	0.947	0.989
	BN	0.015	0.037	0.078	0.217	0.467	0.729	0.890	0.945	0.989
$n = 20$	LR	0.008	0.044	0.092	0.239	0.489	0.743	0.897	0.948	0.990
	BN	0.007	0.041	0.088	0.232	0.482	0.738	0.894	0.947	0.989
$n = 25$	LR	0.009	0.047	0.095	0.243	0.494	0.746	0.898	0.949	0.990
	BN	0.008	0.045	0.092	0.239	0.489	0.743	0.896	0.948	0.990

	method	$p = 10$								
	MCMC	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99
$n = 25$	LR	0.071	0.069	0.122	0.282	0.539	0.782	0.918	0.961	0.993
	BN	0.013	0.036	0.076	0.213	0.462	0.726	0.888	0.944	0.989
$n = 30$	LR	0.015	0.059	0.114	0.275	0.530	0.774	0.913	0.958	0.992
	BN	0.008	0.041	0.086	0.228	0.477	0.734	0.892	0.946	0.989
$n = 35$	LR	0.011	0.056	0.111	0.270	0.524	0.769	0.910	0.956	0.992
	BN	0.007	0.043	0.090	0.236	0.485	0.739	0.895	0.947	0.989

Table 3.2: LR and BN approximations compared to MCMC for $\hat{\psi} = 1$, $n = 15, 20, 25$, and $p = 5$; and for $\hat{\psi} = 1$, $n = 25, 30, 35$, and $p = 10$.

3.2 Refined tail area approximation

In unpublished work, DiCiccio and Young suggest a variant of the BN formula derived by constructing versions of the r and q statistics based a modified profile log-likelihood. The technical details behind its derivation will be omitted. Broadly, the motivation comes from writing the density approximation (2.1) as

$$\hat{\pi}(\psi|y) \propto \exp\{l_p(\psi) - \log T_{NP}(\psi)\} \frac{\pi^{\psi}(\hat{\psi})}{\pi^{\psi}(\psi)},$$

where

$$T_{NP}(\psi) = \frac{|j_{\chi\chi}(\psi, \hat{\chi}_{\psi})|^{1/2}}{\pi^{\chi|\psi}(\psi, \hat{\chi}_{\psi})}$$

is a function that reflects the presence of nuisance parameters in the model, $\pi^{\chi|\psi}(\theta)$ is the conditional prior density of χ given ψ, and $\pi^{\psi}(\psi)$ is the marginal prior density of ψ. Writing the exponent as $\breve{l}_p(\psi) = l_p(\psi) - \log T_{NP}(\psi)$, the refined approximation is defined by

$$\mathbb{P}(\psi \leq \psi_0|y) \quad \doteq \quad \Phi\left(\breve{r}_0 + \frac{1}{\breve{r}_0} \log\left(\frac{\breve{q}_0}{\breve{r}_0}\right)\right),$$

where

$$\breve{r}_0 = \text{sign}(\psi_0 - \breve{\psi})\sqrt{2\left\{\breve{l}_p(\breve{\psi}) - \breve{l}_p(\psi_0)\right\}};$$

$$\breve{q}_0 = -\frac{\breve{l}_p^{(1)}(\psi_0)}{\{-\breve{l}_p^{(2)}(\breve{\psi})\}^{1/2}} \frac{\pi^{\psi}(\psi_0)}{\pi^{\psi}(\breve{\psi})};$$

and $\breve{\psi}$ is the mode of $\breve{l}_p(\psi)$.

We will now test this approximation on the previous examples. In the case of example 3.1.1, it is easy to see that the modified profile log-likelihood $\breve{l}_p(\sigma^2)$ takes the form

$$-\frac{np}{2}\left\{\log\sigma^2 + \frac{\hat{\sigma}^2}{\sigma^2}\right\} + \left(\frac{p}{2} - 2\right)\log\sigma^2,$$

while for example 3.1.2 it is given by

$$-\frac{n}{2}(\psi - \hat{\psi})^2 - \frac{p-1}{2}\log\psi.$$

Tables 3.3 and 3.4 are the same as tables 3.1 and 3.2 with added rows for the refined approximation. In the first example the improvement yielded by the refinement is remarkable. For all cases considered the approximation is virtually perfect across all the quantiles, and its performance does not seem to degrade for large values of p. In the second example, however, there is no clear improvement over the standard approximations. In fact, for the lower quantiles the refinement is less reliable than the alternatives. This happens because for fixed n and p, $\breve{l}_p(\psi) \to \infty$ as $\psi \to 0$, which causes $\breve{r}(\psi_0)$ to be decreasing near 0 and even undefined when $\breve{l}_p(\psi_0) > \breve{l}_p(\breve{\psi})$. Asymptotically, however, this does not constitute a problem, since in regions that are bounded away from zero the second term of $\breve{l}_p(\psi)$ becomes negligible relative to the first one.

		$p = 50$								
	Prob.	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99
$n = 3$	BN	0.017	0.075	0.141	0.321	0.585	0.815	0.934	0.969	0.995
	Ref	0.010	0.050	0.100	0.251	0.501	0.751	0.900	0.950	0.990
$n = 10$	BN	0.011	0.053	0.105	0.259	0.512	0.759	0.905	0.953	0.991
	Ref	0.010	0.050	0.100	0.250	0.500	0.750	0.900	0.950	0.990
$n = 17$	BN	0.010	0.051	0.102	0.254	0.505	0.754	0.902	0.951	0.990
	Ref	0.010	0.050	0.100	0.250	0.500	0.750	0.900	0.950	0.990

		$p = 250$								
	Prob.	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99
$n = 3$	BN	0.030	0.117	0.205	0.416	0.680	0.874	0.961	0.983	0.998
	Ref	0.010	0.050	0.100	0.250	0.500	0.750	0.900	0.950	0.990
$n = 10$	BN	0.012	0.057	0.112	0.271	0.525	0.770	0.911	0.956	0.992
	Ref	0.010	0.050	0.100	0.250	0.500	0.750	0.900	0.950	0.990
$n = 17$	BN	0.011	0.053	0.105	0.259	0.511	0.759	0.905	0.953	0.991
	Ref	0.010	0.050	0.100	0.250	0.500	0.750	0.900	0.950	0.990

		$p = 500$								
	Prob.	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99
$n = 3$	BN	0.046	0.159	0.264	0.492	0.745	0.910	0.974	0.990	0.999
	Ref	0.010	0.050	0.100	0.250	0.500	0.750	0.900	0.950	0.990
$n = 10$	BN	0.013	0.060	0.117	0.279	0.536	0.778	0.915	0.959	0.992
	Ref	0.010	0.050	0.100	0.250	0.500	0.750	0.900	0.950	0.990
$n = 17$	BN	0.011	0.054	0.107	0.263	0.516	0.762	0.907	0.954	0.991
	Ref	0.010	0.050	0.100	0.250	0.500	0.750	0.900	0.950	0.990

Table 3.3: BN and refined approximation (Ref) evaluated a true posterior quantiles for $\hat{\sigma}^2 = 1$, $n = 3, 10, 17$, and $p = 50, 250, 500$.

	method	$p = 5$								
	MCMC	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99
$n = 15$	LR	0.025	0.044	0.087	0.229	0.479	0.737	0.894	0.947	0.989
	BN	0.015	0.037	0.078	0.217	0.467	0.729	0.890	0.945	0.989
	Ref	-	0.078	0.118	0.262	0.509	0.755	0.902	0.951	0.990
$n = 20$	LR	0.008	0.044	0.092	0.239	0.489	0.743	0.897	0.948	0.990
	BN	0.007	0.041	0.088	0.232	0.482	0.738	0.894	0.947	0.989
	Ref	0.013	0.052	0.103	0.254	0.504	0.753	0.902	0.951	0.990
$n = 25$	LR	0.009	0.047	0.095	0.243	0.494	0.746	0.898	0.949	0.990
	BN	0.008	0.045	0.092	0.239	0.489	0.743	0.896	0.948	0.990
	Ref	0.010	0.051	0.101	0.252	0.503	0.753	0.901	0.951	0.990

	method	$p = 10$								
	MCMC	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99
$n = 25$	LR	0.071	0.069	0.122	0.282	0.539	0.782	0.918	0.961	0.993
	BN	0.013	0.036	0.076	0.213	0.462	0.726	0.888	0.944	0.989
	Ref	-	0.144	0.172	0.318	0.558	0.786	0.916	0.959	0.992
$n = 30$	LR	0.015	0.059	0.114	0.275	0.530	0.774	0.913	0.958	0.992
	BN	0.008	0.041	0.086	0.228	0.477	0.734	0.892	0.946	0.989
	Ref	0.028	0.072	0.128	0.287	0.536	0.773	0.911	0.956	0.991
$n = 35$	LR	0.011	0.056	0.111	0.270	0.524	0.769	0.910	0.956	0.992
	BN	0.007	0.043	0.090	0.236	0.485	0.739	0.895	0.947	0.989
	Ref	0.014	0.061	0.117	0.275	0.526	0.768	0.908	0.954	0.991

Table 3.4: LR, BN and refined (Ref) approximations compared to MCMC for $\hat{\psi} = 1$, $n = 15, 20, 25$, and $p = 5$; and for $\hat{\psi} = 1$, $n = 25, 30, 35$, and $p = 10$. In two of the cases the refined approximation is not well-defined.

Chapter 4

Discussion

In this project we have discussed an approach to approximate Bayesian inference about a scalar parameter of interest based on accurate analytic approximations. In the first part we proved several properties concerning the asymptotic accuracy of the presented methods, taking the work of Kass *et al.* [12] as a starting point. The regularity conditions we required to this end are not be straightforward to verify in general, but hold for several widely-used models. In addition to the examples considered here (linear regression and full exponential families), Kass *et al.* [12] provide a way to verify them for non-linear regression models, and claim that they can also be verified for curved exponential families and various mixture models. In the second part of the work we considered two models and analysed the performance of the approximations as a function the parameter dimensionality. We saw that, for similar sample sizes, the approximations were less reliable as the dimensionality increased. We then tested a potential refinement of the approximations which proved to be particularly effective for one of the models but yielded no apparent improvement for the other. Naturally, the conclusions that can be drawn from this study are limited, as we only considered two examples, both of which were relatively simple. A more complete analysis of this issue is thus desirable.

Here we have only treated the scalar parameter case for simplicity. If the interest parameter happens to be multidimensional, Laplace's approximation can be applied in a similar way to get approximations of the posterior density, and a multivariate version of the r^* statistic can be used to obtain confidence regions for the parameter. A detailed treatment of the multidimensional case is given in [23].

Appendix A

Proof of lemma 1

Let $h(\theta) = n^{-1}l(\theta)$. It follows from conditions **(i)** and **(ii)** that $h(\theta)$ is strictly concave in $B_{\varepsilon'}(\hat{\theta})$ for some $0 < \varepsilon' < \varepsilon$. Indeed, let $g(\theta) = |-h''(\theta)|$. We know that the third-order derivatives of $h(\theta)$ are $O(1)$ in $B_\varepsilon(\hat{\theta})$, so the gradient of $g(\theta)$, $\nabla g(\theta)$, is also $O(1)$ in these regions. Let $S = \limsup\sup\{\|\nabla g(\theta)\| \colon \theta \in B_\varepsilon(\hat{\theta})\}$ and take $\varepsilon' = (1/2)\min\{\varepsilon, M/S\}$, with M taken from condition **(ii)**. For $\theta \in B_{\varepsilon'}(\hat{\theta})$,

$$g(\theta) = g(\hat{\theta}) + \nabla g(x_\theta)^T(\theta - \hat{\theta}),$$

where $x_\theta \in B_{\varepsilon'}(\hat{\theta})$. From condition **(ii)** we know that $g(\hat{\theta}) > M$ for all n, and by Cauchy-Schwarz inequality the second term is bounded in absolute value by $S\varepsilon' < M$. Therefore $g(\theta)$ is bounded away from zero in $B_{\varepsilon'}(\hat{\theta})$, and hence $-h(\theta)$ is strictly convex in these regions.

Now, take any $0 < z < \varepsilon'$ and let

$$\begin{aligned} m_1 &= \liminf\inf\{h(\hat{\theta}) - h(\theta) \colon \theta \notin B_z(\hat{\theta})\} > 0; \\ m_2 &= \limsup\sup\{\|\nabla h(\theta)\| \colon \theta \in B_z(\hat{\theta})\} < \infty; \\ \delta &= \min\left\{\frac{m_1}{m_2}, z\right\}. \end{aligned}$$

As before, a first-order Taylor expansion shows that, for $\theta \in B_\delta(\hat{\theta})$,

$$h(\hat{\theta}) - h(\theta) < m_2\delta \leq m_2\frac{m_1}{m_2} = m_1.$$

For any $\theta_1 \in B_\delta(\hat{\theta})$ and $\theta_2 \notin B_z(\hat{\theta})$ we have $h(\hat{\theta}) - h(\theta_2) \geq m_1 > h(\hat{\theta}) - h(\theta_1)$, which implies $h(\theta_1) > h(\theta_2)$. Now, take $\psi \in B_\delta(\hat{\psi})$. By the previous consideration $h(\psi, \cdot)$ has a maximum in $B_z(\hat{\chi})$ and all possible maximums must lie in this region. Furthermore, the maximum has to be unique because $h(\psi, \cdot)$ is strictly concave in $B_z(\hat{\chi})$. Clearly this choice of δ is also valid for larger values of z, so the first three claims follow.

For the last part, differentiate the equality $l_\chi(\psi, \hat{\chi}_\psi) = 0$ with respect to ψ. Using obvious notation, we have

$$\begin{aligned} l_{\chi\psi}(\psi, \hat{\chi}_\psi) + l_{\chi\chi}(\psi, \hat{\chi}_\psi)\hat{\chi}'_\psi = 0 \iff \\ l_{\chi\chi}(\psi, \hat{\chi}_\psi)^{-1}l_{\chi\psi}(\psi, \hat{\chi}_\psi) + \hat{\chi}'_\psi = 0. \end{aligned}$$

By condition **(i)**, $\|l_{\chi\psi}(\psi, \hat{\chi}_\psi)\|$ is $O(n)$ in $B_\delta(\hat{\psi})$. To conclude that $\|\hat{\chi}'_\psi\|$ is $O(1)$ in $B_\delta(\hat{\psi})$ we need to show that $\|l_{\chi\chi}(\psi, \hat{\chi}_\psi)^{-1}\|$ is $O(n^{-1})$ in $B_\delta(\hat{\psi})$ for some matrix norm. This is equivalent to $\|nl_{\chi\chi}(\psi, \hat{\chi}_\psi)^{-1}\|$ being $O(1)$, which holds if the largest eigenvalue of $nl_{\chi\chi}(\psi, \hat{\chi}_\psi)^{-1}$ is bounded, or equivalently if the smallest eigenvalue of $n^{-1}l_{\chi\chi}(\psi, \hat{\chi}_\psi)$ is bounded away from zero. This holds if we take a small enough δ, such that the eigenvalues of $n^{-1}l_{\chi\chi}(\psi, \hat{\chi}_\psi)$ are bounded by condition **(i)** and their product is bounded away from zero by conditions **(i)** and **(ii)**. To see the boundedness of the subsequent derivatives we just differentiate the equality 5 more times and use the same argument.

Appendix B

Proof of lemma 2

Note that the error from approximating the denominator of (1.1) is independent of ψ. Now, from the proof of theorem 1 of [12], it follows that the error of the numerator approximation is

$$1 + \frac{R_1(\psi)}{n} + \frac{R_2(\psi)}{n^2},$$

where $R_1(\psi)$ and $R_2(\psi)$ are $O(1)$ in $B_\delta(\hat{\psi})$ for some $\delta > 0$. This is shown in Kass *et al.* [12] (p. 483). Essentially, the error terms come from integrating higher-order terms of the asymptotic expansion of $\pi(\psi, \cdot)\exp\{l(\psi, \cdot)\}$ around $\hat{\chi}_\psi$, and by condition **(i)** and lemma 1, these can be bounded uniformly in $B_\delta(\hat{\psi})$ for some $\delta > 0$.

To prove that the derivative of $R_1(\psi)$ is also bounded, we use its explicit expression as given in the theorem, which is a combination of sums and products of $\pi(\hat{\theta})^{-1}$, $\pi_i(\psi, \hat{\chi}_\psi)$, $\pi_{ij}(\psi, \hat{\chi}_\psi)$, $n^{-1}l_{ijk}(\psi, \hat{\chi}_\psi)$, $n^{-1}l_{ijkl}(\psi, \hat{\chi}_\psi)$ and $nl^{ij}(\psi, \hat{\chi}_\psi)$, for $1 \leq i, j, k, l \leq p + 1$, where the subscripts denote partial differentiation and $l^{ij}(\psi, \hat{\chi}_\psi)$ is the (i, j)th entry of $l''(\psi, \hat{\chi}_\psi)^{-1}$. Clearly, the derivatives of $\pi(\psi, \hat{\chi}_\psi)$ and $n^{-1}l(\psi, \hat{\chi}_\psi)$ up to order 6 are $O(1)$ in some δ-ball around $\hat{\psi}$. Hence, we only need to show that the derivatives of $nl^{ij}(\psi, \hat{\chi}_\psi)$ are also $O(1)$ in some δ-ball around $\hat{\psi}$.

Let $g(\theta) = |n^{-1}j(\theta)|$, which by the previous proof we know is positive and bounded away from zero in $B_\delta(\hat{\theta})$ for some $0 < \delta < \varepsilon$. By Cramer's rule, the entries of $nj(\theta)^{-1}$ are

$$-nl^{ij}(\theta) = \frac{R(\theta)}{g(\theta)},$$

where $R(\theta)$ is a linear combination of the entries of $n^{-1}j(\theta)$. In particular $R(\theta)$ and its partial derivatives are $O(1)$ in $B_\delta(\hat{\theta})$. Taking derivatives with respect to θ_k we have

$$-nl^{ij}_k(\theta) = \frac{R_k(\theta)g(\theta) - R(\theta)g_k(\theta)}{g(\theta)^2}.$$

Clearly, the numerator is $O(1)$ in $B_\delta(\hat{\theta})$, so, as before, since $g(\theta)$ is bounded from below, $-nl^{ij}_k(\theta)$ is $O(1)$ in $B_\delta(\hat{\theta})$. Application of lemma 1 concludes the result.

Appendix C

Proof of lemma 3

Recall that $\hat{\pi}(\psi|y)$ is given by

$$\hat{\pi}(\psi|y) = \frac{1}{\sqrt{2\pi}} j_p(\hat{\psi})^{1/2} \exp\left\{l_p(\psi) - l_p(\hat{\psi})\right\} \frac{|j_{\chi\chi}(\hat{\psi},\hat{\chi})|^{1/2}}{|j_{\chi\chi}(\psi,\hat{\chi}_\psi)|^{1/2}} \frac{\pi(\psi,\hat{\chi}_\psi)}{\pi(\hat{\psi},\hat{\chi})}.$$

By conditions **(i)** and **(ii)** and lemma 1, $\sqrt{n^{-1}j_p(\hat{\psi})}$ and the last two factors are $O(1)$ in $B_\delta(\hat{\psi})$ for some $0 < \delta < \varepsilon$. Hence,

$$\hat{\pi}(\psi|y) \leq C\sqrt{n} \exp\left\{l_p(\psi) - l_p(\hat{\psi})\right\}$$

in $B_\delta(\hat{\psi})$ for some constant C. Also, by theorem 4 of [12],

$$\int_a^{-\delta} \pi(\psi|y) = O(n^{-3/2}),$$

so the domain of integration can be taken to be $(-\delta, \psi)$. If δ is chosen to be smaller than in corollary 1, then, using the notation of the corollary, for n large enough,

$$\begin{aligned}\int_{-\delta}^{\psi_0} \hat{\pi}(\psi|y)R(\psi)d\psi &\leq \int_{-\delta}^{\psi_0} C'n \exp\left\{l_p(\psi) - l_p(\hat{\psi})\right\} |\psi - \hat{\psi}|d\psi + O(n^{-1/2}) \\ &= \int_{-\delta}^{\psi_0} C'n \exp\left\{-\frac{1}{2} j_p(x_\psi)(\psi - \hat{\psi})^2\right\} |\psi - \hat{\psi}|d\psi + O(n^{-1/2}),\end{aligned}$$

for some constant C', where x_ψ lies between ψ and $\hat{\psi}$. Let $u = \sqrt{n}(\psi - \hat{\psi})$ and note that, if δ is small enough, $n^{-1}j_p(x_\psi)$ is bounded from below in the integration domain by some constant C'' to get

$$\int_{-\delta}^{\psi_0} \hat{\pi}(\psi|y)R(\psi)d\psi \leq \int_{-\infty}^{\infty} C' \exp\left\{-\frac{1}{2} C''u^2\right\} |u|du + O(n^{-1/2}) = O(1),$$

from which the result follows.

Appendix D

Proof of lemma 4

Put $l_i \equiv l_p^{(i)}(\hat{\psi})$ for $i = 1, 2, \ldots$ and let $z = (\psi - \hat{\psi})(-l_2)^{1/2}$. Clearly z is in one-to-one correspondence with ψ, and therefore with r. Also, note that

$$\psi \in B_\delta(\hat{\psi}) \iff z \in (-\delta(-l_2)^{1/2}, \delta(-l_2)^{1/2}).$$

A Taylor expansion of $l_p(\psi)$ gives

$$l_p(\psi) = l_0 + \frac{l_2}{2}(\psi - \hat{\psi})^2 + \frac{l_3}{6}(\psi - \hat{\psi})^3 + \frac{l_4}{24}(\psi - \hat{\psi})^4 + \frac{l_5}{120}(\psi - \hat{\psi})^5 + \frac{l_p^{(6)}(x_\psi)}{720}(\psi - \hat{\psi})^6,$$

where x_ψ lies between ψ and $\hat{\psi}$. In terms of z, with a slight abuse of notation,

$$\begin{aligned} l_p(z) &= l_0 - \frac{1}{2}z^2 + \frac{l_3}{6(-l_2)^{3/2}}z^3 + \frac{l_4}{24(-l_2)^2}z^4 + \frac{l_5}{120(-l_2)^{5/2}}z^5 + \frac{l_p^{(6)}(x_z)}{720(-l_2)^3}z^6 \\ &\equiv l_0 - \frac{1}{2}z^2 + \frac{l_3}{6(-l_2)^{3/2}}z^3 + \frac{l_4}{24(-l_2)^2}z^4 + e_1(z), \end{aligned}$$

where x_z lies between 0 and z. Note that the derivative of $e_1(z)$ is

$$\frac{l_5}{24(-l_2)^{5/2}}z^4 + \frac{l_p^{(6)}(x_z')}{120(-l_2)^{5/2}}z^5,$$

where x_z' lies between 0 and z. Rearranging terms gives

$$r^2 = z^2\left\{1 - \frac{l_3}{3(-l_2)^{3/2}}z - \frac{l_4}{12(-l_2)^2}z^2 - \frac{e_1(z)}{z^2}\right\}.$$

By lemma 1 there exists a $\delta > 0$ such that $e_1(z)/z^2$ is bounded in absolute value by $n^{-3/2}p_1(|z|)$ and its derivative by $n^{-3/2}p_2(|z|)$ in $(-\delta(-l_2)^{1/2}, \delta(-l_2))$, where $p_1(z), p_2(z)$ are polynomials independent of n. Also, taking a suitable δ the factor in brackets is bounded away from zero in these regions.

Now we use that r and z have the same sign, and that for all $x > -1$,

$$(1 + x)^{1/2} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + h(x),$$

where $h'(x) = -(1+a_x)^{-3/2}x^2/16$ for some a_x between 0 and x. In particular $h'(x)$ is bounded by a polynomial in x^2 if x is bounded away from zero. Applying it to the previous expression we get

$$r = z\left\{1 - \frac{l_3}{6(-l_2)^{3/2}}z - \left(\frac{l_4}{24(-l_2)^2} - \frac{l_3^2}{72l_2^3}\right)z^2 + e_2(z)\right\},$$

where $e_2(z)$ is a sum of terms of the form

- az^j, with $a = O(n^{-3/2})$ and $j \geq 3$;
- $bz^j e_1(z)^k$, with $b = O(1)$, $j \geq 0$ and $k \geq 1$;
- $h(r^2/z^2 - 1)$.

By all the previous considerations, there exists a polynomial $p_3(z)$ independent of n such that $e_2'(z) \leq n^{-3/2}p_3(|z|)$ for $z \in (-\delta(-l_2)^{1/2}, \delta(-l_2)^{1/2})$. For simplicity, let us write $r = z + a_1z^2 + a_2z^3 + e_3(z)$, with a_i being $O(n^{-i/2})$ and $e_3(z) = ze_2(z)$. Inversion of the expansion gives

$$z = r + b_1r^2 + b_2r^3 + e_4(z),$$

where b_i is $O(n^{-i/2})$. Plugging this expression in the one for r shows that $e_4(z)$ can be written as a sum of terms of the form az^j, with $a = O(n^{-3/2})$ and $j \geq 3$, and $bz^j e_3(z)^k$, with $b = O(1)$, $j \geq 0$ and $k \geq 1$. This implies that $|e_4'(z)| \leq n^{-3/2}p_4(|z|)$ in $(-\delta(-l_2)^{1/2}, \delta(-l_2)^{1/2})$. Hence

$$\frac{dz}{dr} = 1 + 2b_1r + 3b_2r^2 + e_4'(z)\frac{dz}{dr}.$$

But straightforward calculations show that dz/dr is bounded when $z \in (-\delta(-l_2)^{1/2}, \delta(-l_2)^{1/2})$ if δ is small enough, so the last term is bounded in absolute value by $n^{-3/2}p_5(|z|)$ for some polynomial $p_5(z)$ in these intervals.

Writing $r = (\psi - \hat{\psi})j_p(x_r)^{1/2}$ for some x_r between 0 and r, it is immediate to see that $z \in (-\delta(-l_2)^{1/2}, \delta(-l_2)^{1/2}) \iff r \in (-u_1\sqrt{n}, u_1\sqrt{n})$ for some $u_1 > 0$, and that the last term, as a function of r, is bounded in absolute value by by $n^{-3/2}p_6(|r|)$ for some polynomial $p_6(r)$ in these intervals.

Now, differentiating the equality $-r^2/2 = l_p(\psi) - l_0$ with respect to ψ shows that

$$\frac{r}{q} = \frac{dz}{dr}\frac{|j_{\chi\chi}(\psi, \hat{\chi}_\psi)|^{1/2}}{|j_{\chi\chi}(\hat{\psi}, \hat{\chi})|^{1/2}}\frac{\pi(\hat{\psi}, \hat{\chi})}{\pi(\psi, \hat{\chi}_\psi)}.$$

Let $G(\psi) = rq^{-1}(dz/dr)$. By our assumptions $G(\psi)$ is three-times continuously differentiable and its first two derivatives are $O(1)$ in a δ-ball around $\hat{\psi}$. Also, $G(\hat{\psi}) = 1$. Hence, proceeding as before, we can see that

$$G(r) = 1 + c_1r + c_2r^2 + e_5(r),$$

where c_i is $O(n^{-i/2})$ and $|e_5(r)| \leq n^{-3/2}p_7(|r|)$ for some polynomial $p_7(r)$ in $(-u_2\sqrt{n}, u_2\sqrt{n})$ for some $u_2 > 0$.

Taking $u = \min\{u_1, u_2\}$ and multiplying out the expressions for dz/dr and $G(r)$ to get r/q concludes the proof.

Appendix E

Code

```
library(invgamma)
library(mcmc)
library(mcmcse)
library(numDeriv)

############################################################
######################### EXAMPLE 1 ########################
############################################################
rm(list = ls())

# Parameters
p = 500
n = 10
sigma2_hat = 1
alpha = p*(n-1)/2
beta = p*n*sigma2_hat/2

# Standard analytic approximations
r = function(sigma2){
  sign(sigma2-sigma2_hat)*
  sqrt(n*p*(sigma2_hat/sigma2-log(sigma2_hat/sigma2)-1))
}

qB = function(sigma2){
  sqrt(n*p/2)*(1-sigma2_hat/sigma2)*(sigma2_hat/sigma2)^(p/2)
}

rstar = function(sigma2){
  r(sigma2) + log(qB(sigma2)/r(sigma2))/r(sigma2)
}

LR = function(x) pnorm(r(x))+dnorm(r(x))*(1/r(x) - 1/qB(x))
BN = function(x) pnorm(rstar(x))

# Refined approximation
mod.ll = function(sigma2){
```

```
  -(n*p/2)*(log(sigma2)+sigma2_hat/sigma2)+(p/2-2)*log(sigma2)
} # modifield log-likelihood

sigma2.mod = optimize(mod.ll, c(0.01, 50), maximum = TRUE)$maximum

r.mod = function(sigma2){
  sign(sigma2-sigma2.mod)*sqrt(2*(mod.ll(sigma2.mod)-mod.ll(sigma2)))
}

q.mod = function(sigma2){
  -(grad(mod.ll, sigma2)/sqrt(-hessian(mod.ll, sigma2.mod)))*
    sigma2.mod/sigma2
}

r.star.mod = function(sigma2){
  r.mod(sigma2)+log(q.mod(sigma2)/r.mod(sigma2))/r.mod(sigma2)
}

Ref = function(sigma2) pnorm(r.star.mod(sigma2))

# Approximations evaluated at true posterior quantiles.
# The rate parameter of the Inv. Gamma in R corresponds
# with the shape parameter in our notation
x = c(0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.99)
quants = qinvgamma(x, alpha, rate = beta)
sapply(quants, LR)
sapply(quants, BN)
sapply(quants, Ref)


############################################################
######################### EXAMPLE 2 ########################
############################################################
rm(list = ls())

# Parameters
p = 5
n = 15
psi.hat = 1

# Posterior log-likelihood
mu.hat = c(psi.hat, rep(0, p-1))
logposterior = function(mu) {
  x = sqrt(sum((mu-mu.hat)^2))
  y = sqrt(sum(mu^2))
  -(n/2)*x^2-(p-1)*log(y)
}

# MCMC with scale parameter adjusted for each case
# based on diagnostics and number of batchs large
# enough to meet low standard error restriction
set.seed(4461)
MCMC = metrop(logposterior, initial = mu.hat, nbatch = 1e7, scale = 0.25)
```

```
MCMC$accept
acf(MCMC$batch[, 1])

# Posterior sample of interest parameter
SAMPLE = apply(MCMC$batch, 1, function(x) sqrt(sum(x^2)))

# MCMC approximation of posterior distribution function
mcmc.app = function(psi) mean(SAMPLE<=psi)

# Check small standard error for all quantiles considered
prob = 0.5
quant = quantile(SAMPLE, prob)
g = function(x) (sqrt(sum(x^2))<=quant)
mcse.multi(MCMC$batch, g = g)

# Standard analytic approximations
r = function(psi) sqrt(n)*(psi-psi.hat)
qB = function(psi) r(psi)*(psi/psi.hat)^((p-1)/2)
r.star = function(psi) r(psi)+log(qB(psi)/r(psi))/r(psi)

LR = function(psi) pnorm(r(psi))+dnorm(r(psi))*(1/r(psi)-1/qB(psi))
BN = function(psi) pnorm(r.star(psi))

# Refinement
mod.ll = function (psi){
  -(1/2)*r(psi)^2-((p-1)/2)*log(psi)
}# modified profile log-likelihood

psi.mod = optimize(mod.ll, c(0.01, 50), maximum = TRUE)$maximum

r.mod = function(psi){
  sign(psi-psi.mod)*sqrt(2*(mod.ll(psi.mod)-mod.ll(psi)))
}

q.mod = function(psi){
  -grad(mod.ll, psi)/sqrt(-hessian(mod.ll, psi.mod))
}

r.star.mod = function(psi){
  r.mod(psi)+log(q.mod(psi)/r.mod(psi))/r.mod(psi)
}

Ref = function(psi) pnorm(r.star.mod(psi))

# Approximations evaluated at MCMC posterior quantiles
x = c(0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.99)
quants = quantile(SAMPLE, x)
sapply(quants, LR)
sapply(quants, BN)
sapply(quants, Ref)
```

Bibliography

[1] Banerjee, A. K., and Bhattacharyya, G. K. (1979). Bayesian results for the Inverse Gaussian distribution with an application. *Technometrics* **21**(2) 247-251.

[2] Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**(2) 307–322.

[3] Barndorff-Nielsen, O. E. (1990). Approximate interval probabilities. *J. Roy. Statist. Soc. Ser. B* **52**(3) 485–496.

[4] Barndorff-Nielsen, O. E. (1991). Modified signed log likelihood ratio. *Biometrika* **78**(3) 557–563.

[5] Barndorff-Nielsen, O. E., and Cox, D. R. (1989). *Asymptotic techniques for use in statistics*. London: Chapman & Hall.

[6] Diciccio, T. J., Field, C. A., and Fraser, D. A. S. (1990). Approximations of marginal tail probabilities and inference for scalar parameters. *Biometrika* **77**(1) 77–95

[7] Diciccio, T. J., and Martin, M. A. (1991). Approximations of marginal tail probabilities for a class of smooth functions with applications to bayesian and conditional inference. *Biometrika* **78**(4) 891-902.

[8] Flegal J. M., Hughes J., Vats D., and Dai, N. (2017). mcmcse: Monte Carlo Standard Errors for MCMC. R package version 1.3-2. Riverside, CA, Denver, CO, Coventry, UK, and Minneapolis, MN.

[9] Geyer, C. J., and Johnson, L. T. (2017). mcmc: Markov Chain Monte Carlo. R package version 0.9-5. `https://CRAN.R-project.org/package=mcmc`

[10] Gilbert, P., and Varadhan, R. (2016). numDeriv: Accurate Numerical Derivatives. R package version 2016.8-1. URL: `https://CRAN.R-project.org/package=numDeriv`

[11] Kahle., D., and Stamey, J. (2017). invgamma: The Inverse Gamma Distribution. R package version 1.1.URL: `https://CRAN.R-project.org/package=invgamma`

[12] Kass, R. E., Tierney, L., and Kadane, J. B. (1990). The validity of posterior expansions based on Laplace's method, Essays in Honor of George Bernard, eds. S. Geisser, J.S. Hodges, S.J. Press, and A. Zellner, Amsterdam: North Holland, 473-488. URL: `http://www.stat.cmu.edu/~kass/papers/validity.pdf`

[13] Laplace, P. S. (1774). Memoir on the probability of causes of events. *Mémoires de Mathématique et de Physique, Tome Sixième.*

[14] Leonard, T. (1982). Comment on "A simple predictive density function", by M. Lejeune and G. D. Faulkenberry. *J. Amer. Statist. Assoc.* **77**(379) 657-658.

[15] Phillips, P. C. B. (1983). Marginal densities of instrumental variable estimators in the general single equation case. *Advances in Econometrics* **2** 1-24.

[16] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: `http://www.R-project.org`

[17] Reid, N. (1996). Likelihood and Bayesian approximation methods. In *Bayesian Statistics* **5**, 351–368. Oxford University Press.

[18] Reid, N., and Sun, Y. (2010). Assessing sensitivity to priors using higher order approximations. *Commun. Stat. Theory Methods* **39**(8-9) 1373-1386.

[19] Ruli, E., Sartori, N., and Ventura, L. (2014). Marginal posterior simulation via higher-order tail area approximations. *Bayesian Anal.* **9**(1) 129-146.

[20] Tierney, L. J., and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* **81**(393) 82–86.

[21] Tierney, L. J., Kass, R. E., and Kadane, J. B. (1989). Approximate marginal densities of nonlinear functions. *Biometrika* **76**(3) 425-433.

[22] Ventura L., and Reid, N. (2014). Approximate Bayesian computation with modified log-likelihood ratios. *Metron* **72**(2), 231-245.

[23] Ventura, L., Ruli, E., and Racugno, W. (2013). A note on approximate Bayesian credible sets based on modified log-likelihood ratios. *Stat. Prob. Lett.* **83**(11), 2467–2472.