Fully Exponential Laplace Approximations Using Asymptotic Modes

Yoichi MIYATA

Posterior means of positive functions can be expressed as the ratio of integrals which is called *fully exponential form*. When approximating the posterior means analytically, we usually use Laplace's method. Tierney and Kadane presented a second-order approximation by using the Laplace approximations in each of the numerator and denominator of the fully exponential form. However, Laplace's method requires the exact mode of the integrand. In this article we introduce the concept of asymptotic modes and present the Laplace method via an asymptotic mode under regularity conditions. Furthermore, we propose second-order approximations to the posterior means of positive functions without evaluating the third derivatives of a log-likelihood function and the exact modes of integrands. We also give an Edgeworth-like expansion for the random variable according to a posterior distribution using the Laplace method via an asymptotic mode.

KEY WORDS: Asymptotic expansions; Bayesian inference; Computation of integrals; Laplace's method; Predictive distributions.

1. INTRODUCTION

A simple and remarkable method of asymptotic expansion of integrals generally attributed to Laplace (1847) is used widely in applied mathematics. This method has been applied by many authors (Johnson 1970; Lindley 1980; Mosteller and Wallace 1964; Tierney and Kadane 1986). Tierney and Kadane (1986) applied Laplace's method in a special form, which we call the *fully exponential form*, that has the advantage of requiring only second derivatives of the log-likelihood function to achieve a second-order approximation to the expectation and variance of a strictly positive function $g(\boldsymbol{\theta})$. Moreover, Tierney, Kass, and Kadane (1989) obtained a second-order approximation to the expectation of a general function (not necessarily positive) by applying the fully exponential method to approximate the moment-generating function (MGF) and then differentiating. They called the MGF method.

Our main interest is in approximating the posterior expectation of a strictly positive function $g(\boldsymbol{\theta})$, which can be written as the ratio

$$E\big[g(\boldsymbol{\theta})|\mathbf{X}^{(n)}\big]=\frac{\int_{\Theta} g(\boldsymbol{\theta})p_n(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}}{\int_{\Theta} p_n(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}}, \tag{1}$$

where n is the sample size, $p_n(\mathbf{x}|\boldsymbol{\theta})$ is the likelihood, $\mathbf{X}^{(n)}$ is a random vector with density $p_n(\mathbf{x}|\boldsymbol{\theta})$, and π is the prior. We find it convenient to reexpress the integrals in (1) so that the expectation takes the form

$$E\big[g(\boldsymbol{\theta})|\mathbf{X}^{(n)}\big]=\frac{\int_{\Theta}\exp\{-nh_n^*(\boldsymbol{\theta})\}\,d\boldsymbol{\theta}}{\int_{\Theta}\exp\{-nh_n(\boldsymbol{\theta})\}\,d\boldsymbol{\theta}}, \tag{2}$$

where $h_n^*(\boldsymbol{\theta})=-n^{-1}\log[g(\boldsymbol{\theta})p_n(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})]$ and $h_n(\boldsymbol{\theta})=-n^{-1}\log[p_n(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})]$. For simplicity, we denote the density by $p_n=p$ and $E[g(\boldsymbol{\theta})|\mathbf{X}^{(n)}]=E[g(\boldsymbol{\theta})]$. If each of $-h_n^*$ and $-h_n$ has a dominant peak at its maximum, then Laplace's method will be suitable for application to both the numerator and denominator of (2). However, in practice it may be difficult to find the exact modes of $-h_n^*$ and $-h_n$, which are usually required in the fully exponential Laplace approximations.

The main purpose of this article is to give some approximations to (2) by using asymptotic modes of the numerator and denominator instead of the exact modes. Section 2 introduces the concept of asymptotic modes and give the Laplace method with an asymptotic mode under regularity conditions. Section 3 contains two kinds of approximations to (2). Theorem 2 gives an approximation to $E[g(\boldsymbol{\theta})]$ by applying the Laplace method to both the numerator and denominator of (2) with the same asymptotic mode. Theorems 3 and 4 present second-order approximations by applying the Laplace method to the numerator and denominator of (2) with different asymptotic modes. Section 4 gives an Edgeworth-like expansion for the cumulative distribution function of a standardized random variable, using Theorem 2. Section 5 contains the relation between standard-form approximations and the ones obtained by combining Theorem 2 and the MGF method. Section 6 briefly discusses the problem of approximating predictive densities. Section 7 illustrates these approximations with some special cases. Appendix A gives a brief introduction of standard-form approximations, Appendix B provides some lemmas concerning matrices that are needed in the proofs of the main results, and Appendix C presents the proofs of the main results in Sections 3–6.

2. LAPLACE'S APPROXIMATIONS

In this section we introduce the Laplace approximation for an integral of the form $\int_{\Theta} e^{-nh_n(\boldsymbol{\theta})}\,d\boldsymbol{\theta}$ with an asymptotic mode of $-h_n(\boldsymbol{\theta})$. Let $\boldsymbol{\Theta}$ be an open subset of $\mathbb{R}^d$ and $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_d)^T$. For convenience of exposition, we write

$$\frac{\partial^s}{\partial\theta_{i1}\cdots\partial\theta_{is}}h_n(\boldsymbol{\theta})=h_{i1\cdots is}(\boldsymbol{\theta}),\qquad \sum_{i1\cdots is}=\sum_{i1=1}^{d}\cdots\sum_{is=1}^{d},$$

and represent the first and second derivatives by

$$D^1h_n(\boldsymbol{\theta})=\frac{\partial}{\partial\boldsymbol{\theta}}h_n(\boldsymbol{\theta})=\left(\frac{\partial}{\partial\theta_1}h_n(\boldsymbol{\theta}),\ldots,\frac{\partial}{\partial\theta_d}h_n(\boldsymbol{\theta})\right)^T$$

and

$$D^2h_n(\boldsymbol{\theta})=\frac{\partial^2}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T}h_n(\boldsymbol{\theta})=\begin{pmatrix}\frac{\partial^2}{\partial\theta_1\partial\theta_1}h_n(\boldsymbol{\theta}) & \cdots & \frac{\partial^2}{\partial\theta_1\partial\theta_d}h_n(\boldsymbol{\theta})\\ \vdots & \ddots & \vdots\\ \frac{\partial^2}{\partial\theta_d\partial\theta_1}h_n(\boldsymbol{\theta}) & \cdots & \frac{\partial^2}{\partial\theta_d\partial\theta_d}h_n(\boldsymbol{\theta})\end{pmatrix}.$$

Yoichi Miyata is Teacher of Mathematics, Department of Mathematics, Waseda University Honjo Senior High School, Nishitomita, Honjo-shi Saitama, 367-0035, Japan (E-mail: *ymiyata@waseda.jp*). The author is grateful to the associate editor and two anonymous referees for very helpful comments that considerably improved the article.

Journal of the American Statistical Association
December 2004, Vol. 99, No. 468, Theory and Methods
DOI 10.1198/016214504000001673

We say that $\hat{\boldsymbol\theta}$ is an *asymptotic mode of* $-h_n$ if $\hat{\boldsymbol\theta}$ converges to the exact mode of $-h_n(\boldsymbol\theta)$ as the sample size n tends to infinity. Note that the exact mode of $-h_n(\boldsymbol\theta)$ is also an asymptotic mode for $-h_n$. In addition, we define the following asymptotic modes.

Definition 1. $\hat{\boldsymbol\theta}$ is an *asymptotic mode of order* n^{-1} *for* $-h_n$ if $\hat{\boldsymbol\theta}$ converges to the exact mode of $-h_n$ and $D^1 h_n(\hat{\boldsymbol\theta}) = O(n^{-1})$.

Remark 1. Let $h_n(\boldsymbol\theta) = -n^{-1}\log[p(\mathbf{x}|\boldsymbol\theta)\pi(\boldsymbol\theta)]$. The maximum likelihood estimator (MLE) $\hat{\boldsymbol\theta}_{\mathrm{ML}}$ for $p(\mathbf{x}|\boldsymbol\theta)$ is an asymptotic mode of order n^{-1} for $-h_n$ because $D^1 h_n(\hat{\boldsymbol\theta}_{\mathrm{ML}}) = -n^{-1}\pi(\hat{\boldsymbol\theta}_{\mathrm{ML}}) = O(n^{-1})$ and $\hat{\boldsymbol\theta}_{\mathrm{ML}}$ converges to the exact mode of $-h_n(\boldsymbol\theta)$, that is, the posterior mode, as n tends to infinity under suitable conditions (see, e.g., Heyde and Johnstone 1979).

Definition 2. $\hat{\boldsymbol\theta}$ is *an asymptotic mode of order* n^{-2} *for* $-h_n$ if $\hat{\boldsymbol\theta}$ converges to the exact mode of $-h_n$ and $D^1 h_n(\hat{\boldsymbol\theta}) = O(n^{-2})$.

Remark 2. Let $\hat{\boldsymbol\theta}_{\mathrm{ML}}$ be the MLE for $p(\mathbf{x}|\boldsymbol\theta)$, $h_n(\boldsymbol\theta) = -n^{-1}\log[p(\mathbf{x}|\boldsymbol\theta)\pi(\boldsymbol\theta)]$, and $\hat{\boldsymbol\theta} = \hat{\boldsymbol\theta}_{\mathrm{ML}} - [D^2 h_n(\hat{\boldsymbol\theta}_{\mathrm{ML}})]^{-1} \times D^1 h_n(\hat{\boldsymbol\theta}_{\mathrm{ML}})$. Then it follows from the same argument as in the proof of Theorem 3 that $D^1 h_n(\hat{\boldsymbol\theta}) = O(n^{-2})$.

Subsequently, we introduce the regularity conditions (A1), (A2), (A3), (A4), and (A5) for which the asymptotic expansions for $\int_{\boldsymbol\Theta} e^{-nh_n(\boldsymbol\theta)}\,d\boldsymbol\theta$ will be valid. Let $\|\mathbf{a}\| = (\mathbf{a}^T\mathbf{a})^{1/2}$ for any vector $\mathbf{a}$, and let $|\cdot|$ denote the determinant of a matrix. We use $B_\delta(\hat{\boldsymbol\theta})$ to denote the open ball of radius δ centered at $\hat{\boldsymbol\theta}$, namely $B_\delta(\hat{\boldsymbol\theta}) = \{\boldsymbol\theta \in \boldsymbol\Theta : \|\boldsymbol\theta - \hat{\boldsymbol\theta}\| < \delta\}$. Let $\{\hat{\boldsymbol\theta}\} = \{\hat{\boldsymbol\theta} : n = 1, 2, \dots\}$ be the sequence of asymptotic modes.

We list the following assumptions for $(\{h_n(\boldsymbol\theta)\}, \{\hat{\boldsymbol\theta}\})$:

(A1) $\{h_n(\boldsymbol\theta) : n = 1, 2, \dots\}$ is a sequence of six times continuously differentiable real functions on $\boldsymbol\Theta$.

There exists positive numbers ϵ, M, and ζ and an integer n_0 such that for the asymptotic mode $\hat{\boldsymbol\theta}$, $n \ge n_0$ implies the following:

(A2) The integral $\int_{\boldsymbol\Theta} e^{-nh_n(\boldsymbol\theta)}\,d\boldsymbol\theta$ is finite.

(A3) For all $\boldsymbol\theta \in B_\epsilon(\hat{\boldsymbol\theta})$ and all $1 \le j_1, \dots, j_m \le d$ with $m = 1, \dots, 6$, $\|\partial^d h_n(\boldsymbol\theta)/\partial\theta_{j_1}\cdots\partial\theta_{j_d}\| < M$.

(A4) $D^2 h_n(\hat{\boldsymbol\theta})$ is positive definite and $|D^2 h_n(\hat{\boldsymbol\theta})| > \zeta$.

(A5) For all δ for which $0 < \delta < \epsilon$, $B_\delta(\hat{\boldsymbol\theta}) \subseteq \boldsymbol\Theta$ and

$$|nD^2 h_n(\hat{\boldsymbol\theta})|^{1/2} C_n(\hat{\boldsymbol\theta})^{-1} \times \int_{\boldsymbol\Theta - B_\delta(\hat{\boldsymbol\theta})} \exp\{-n[h_n(\boldsymbol\theta) - h_n(\hat{\boldsymbol\theta})]\}\,d\boldsymbol\theta = O(n^{-2}).$$

The pair $(\{h_n\}, \{\hat{\boldsymbol\theta}\})$ is said to satisfy the *analytical assumptions for the asymptotic-mode Laplace method* if (A1), (A2), (A3), (A4), and (A5) are satisfied for the asymptotic mode $\hat{\boldsymbol\theta}$.

Our conditions (A1)–(A5) are analogous to those of Kass, Tierney, and Kadane (1990), except that $\hat{\boldsymbol\theta}$ is an asymptotic mode. If $\hat{\boldsymbol\theta}$ is an asymptotic mode of order n^{-1}, then $\int_{\boldsymbol\Theta - B_\delta(\hat{\boldsymbol\theta})} \exp\{-n[h_n(\boldsymbol\theta) - h_n(\hat{\boldsymbol\theta})]\}\,d\boldsymbol\theta = O(n^{-2-d/2})$ holds from (A5). Hence, (A5) means that the probability outside a neighborhood of the $\hat{\boldsymbol\theta}$ converges to 0 as the sample size n tends to infinity. Let $h_{j1\cdots jd}(\hat{\boldsymbol\theta})$ denote the dth partial derivative $\partial^d h_n(\boldsymbol\theta)/\partial\theta_{j1}\cdots\partial\theta_{jd}$ with respect to $\boldsymbol\theta$ evaluated at $\hat{\boldsymbol\theta}$; for example, $h_{112}(\hat{\boldsymbol\theta})$ means that $\partial^3 h_n(\hat{\boldsymbol\theta})/(\partial\theta_1^2\,\partial\theta_2)$. Let h^{ij} be the components of $[D^2 h(\hat{\boldsymbol\theta})]^{-1}$ and let $\mathbf{b} = (b_i) = -[D^2 h_n(\hat{\boldsymbol\theta})]^{-1} \times D^1 h_n(\hat{\boldsymbol\theta})$. Then the following expansion holds, the proof of which is given in Appendix C.

Theorem 1. Suppose that $\hat{\boldsymbol\theta}$ is an asymptotic mode of order n^{-1} for $-h_n$ and that the pair $(\{h_n(\boldsymbol\theta)\}, \{\hat{\boldsymbol\theta}\})$ satisfies the analytical assumptions for the asymptotic-mode Laplace method. Then it follows that for large n,

$$\int_{\boldsymbol\Theta} e^{-nh_n(\boldsymbol\theta)}\,d\boldsymbol\theta = (2\pi)^{d/2}|\boldsymbol\Sigma|^{1/2} e^{-nh_n(\hat{\boldsymbol\theta})} C_n(\hat{\boldsymbol\theta})\left(1 + \frac{a_n}{n} + O(n^{-2})\right), \quad (3)$$

where $C_n(\hat{\boldsymbol\theta}) = \exp\{(n/2) D^1 h_n(\hat{\boldsymbol\theta})^T [D^2 h_n(\hat{\boldsymbol\theta})]^{-1} D^1 h_n(\hat{\boldsymbol\theta})\}$,

$$a_n = -\frac{1}{2}\sum_{ijk} h_{ijk}(\hat{\boldsymbol\theta})(nb_i)h^{jk} - \frac{1}{8}\sum_{ijkq} h_{ijkq}(\hat{\boldsymbol\theta})h^{ij}h^{kq} + \frac{1}{72}\sum_{ijkqrs} h_{ijk}(\hat{\boldsymbol\theta})h_{qrs}(\hat{\boldsymbol\theta})\mu_{ijkqrs}n^3,$$

$\boldsymbol\Sigma = n^{-1}[D^2 h(\hat{\boldsymbol\theta})]^{-1}$, and μ_{ijkqrs} are the sixth central moments of a multivariate normal distribution having covariance matrix $\boldsymbol\Sigma$.

Here we refer to the term a_n. By assumption (A4), h^{ij} is of order $O(1)$. It is clear from (A3) that $h_{ijkq}(\hat{\boldsymbol\theta})$ and $h_{ijk}(\hat{\boldsymbol\theta})$ are of order $O(1)$. Hence, $\sum h_{ijkq}(\hat{\boldsymbol\theta})h^{ij}h^{kq}$ and $\sum h_{ijk}(\hat{\boldsymbol\theta})h_{qrs}(\hat{\boldsymbol\theta}) \times \mu_{ijkqrs}n^3$ are of order $O(1)$. Thus it is seen from $\mathbf{b} = O(n^{-1})$ that a_n is of order $O(1)$ if $\hat{\boldsymbol\theta}$ is an asymptotic mode of order n^{-1} for $-h_n$. Note that the order of $\sum h_{ijk}(\hat{\boldsymbol\theta})(nb_i)h^{jk}$ depends on that of the asymptotic mode $\hat{\boldsymbol\theta}$. Therefore, if $\hat{\boldsymbol\theta}$ is an asymptotic mode of order n^{-2}, then by dropping the term of order n^{-2} in a_n/n, the term of order n^{-1} in (3) becomes $-(8n)^{-1}\sum h_{ijkq}(\hat{\boldsymbol\theta})h^{ij}h^{kq} + (72n)^{-1}\sum h_{ijk}(\hat{\boldsymbol\theta}) \times h_{qrs}(\hat{\boldsymbol\theta})\mu_{ijkqrs}n^3$. As a final note, if $\hat{\boldsymbol\theta}$ is the exact mode of $-h_n(\boldsymbol\theta)$, then it follows from $D^1 h_n(\hat{\boldsymbol\theta}) = \mathbf{0}$ that $C_n(\hat{\boldsymbol\theta}) = 1$ and $\mathbf{b} = \mathbf{0}$. In this case, the approximation (3) is equivalent to the Tierney–Kadane approximation.

3. MAIN RESULTS

Let $g(\boldsymbol\theta)$ be a smooth and strictly positive function on the parameter space $\boldsymbol\Theta$. Recall the expression (2), $E[g(\boldsymbol\theta)] = \int_{\boldsymbol\Theta} e^{-nh_n^*(\boldsymbol\theta)}\,d\boldsymbol\theta / \int_{\boldsymbol\Theta} e^{-nh_n(\boldsymbol\theta)}\,d\boldsymbol\theta$, where n is the sample size, $\pi(\boldsymbol\theta)$ is the prior, $p(\mathbf{x}|\boldsymbol\theta)$ is the likelihood, $h_n(\boldsymbol\theta) = -n^{-1}\log p(\mathbf{x}|\boldsymbol\theta)\pi(\boldsymbol\theta)$, and $h_n^*(\boldsymbol\theta) = -n^{-1}\log g(\boldsymbol\theta)p(\mathbf{x}|\boldsymbol\theta)\pi(\boldsymbol\theta)$. This section contains mainly two kinds of approximations to (2), which are obtained by applying Theorem 1. Section 3.1 gives an approximation to MGFs that is superior to standard-form and Tierney–Kadane approximations. Section 3.2 gives second-order approximations to (2) without an evaluation of the exact mode of $-h_n^*$ and the third derivatives of h_n.

3.1 Approximations via the Same Asymptotic Mode

This section presents an approximation to (2) by applying Theorem 1 to the denominator and the numerator with the same asymptotic mode of order n^{-1} for $-h_n$. Although the approximation (4) requires the third derivatives of $h_n(\boldsymbol{\theta})$, Proposition 1 shows that it gives a good approximation to MGFs. The proof of Theorem 2 is given in Appendix C.

Theorem 2. Let $g(\boldsymbol{\theta})$ be a strictly positive function on $\boldsymbol{\Theta}$ and let $\hat{\boldsymbol{\theta}}$ be an asymptotic mode of order n^{-1} for $-h_n$. Suppose that $(\{h_n^*(\boldsymbol{\theta})\}, \{\hat{\boldsymbol{\theta}}\})$ and $(\{h_n(\boldsymbol{\theta})\}, \{\hat{\boldsymbol{\theta}}\})$ satisfy the analytical assumptions for the asymptotic-mode Laplace method. Then

$$E[g(\boldsymbol{\theta})] = g(\hat{\boldsymbol{\theta}})\left(\frac{|D^2h_n(\hat{\boldsymbol{\theta}})|}{|D^2h_n^*(\hat{\boldsymbol{\theta}})|}\right)^{1/2}\frac{C_n^*(\hat{\boldsymbol{\theta}})}{C_n(\hat{\boldsymbol{\theta}})} \times\left(1-\frac{1}{2n}\sum_{ijkq}h_{ijk}(\hat{\boldsymbol{\theta}})h^{iq}h^{jk}\frac{\partial}{\partial\theta_q}\log g(\hat{\boldsymbol{\theta}}) + O(n^{-2})\right), \quad (4)$$

where $C_n(\hat{\boldsymbol{\theta}}) = \exp\{(n/2)D^1h_n(\hat{\boldsymbol{\theta}})^T[D^2h_n(\hat{\boldsymbol{\theta}})]^{-1}D^1h_n(\hat{\boldsymbol{\theta}})\}$ and $C_n^*(\hat{\boldsymbol{\theta}}) = \exp\{(n/2)D^1h_n^*(\hat{\boldsymbol{\theta}})^T[D^2h_n^*(\hat{\boldsymbol{\theta}})]^{-1}D^1h_n^*(\hat{\boldsymbol{\theta}})\}$.

Proposition 1. Let $\hat{\boldsymbol{\theta}}$ be the exact mode of $-h_n(\boldsymbol{\theta})$, $M(\mathbf{t}) = E[\exp(\mathbf{t}^T\boldsymbol{\theta})]$ denote the MGF, $\mathbf{t} = (t_1,\dots,t_d)^T \in \mathbb{R}^d$, and $g(\boldsymbol{\theta}) = \exp(\mathbf{t}^T\boldsymbol{\theta})$. Under the assumptions of Theorem 2, it follows that

$$M(\mathbf{t}) = \exp\left\{\mathbf{t}^T\hat{\boldsymbol{\theta}} + \frac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}\right\} \times\left(1-\frac{1}{2n}\sum_{ijkq}h_{ijk}(\hat{\boldsymbol{\theta}})h^{iq}h^{ik}t_q + O^{\mathbf{t}}(n^{-2})\right), \quad (5)$$

where $\boldsymbol{\Sigma} = n^{-1}[D^2h(\hat{\boldsymbol{\theta}})]^{-1}$ and $O^{\mathbf{t}}(\cdot)$ indicates that the error term may depend on $\mathbf{t}$.

Proof of Proposition 1. It follows from $h^*(\boldsymbol{\theta}) = h(\boldsymbol{\theta}) - (1/n)\mathbf{t}^T\boldsymbol{\theta}$ that $D^2h_n^*(\boldsymbol{\theta}) = D^2h_n(\boldsymbol{\theta})$. Evaluating $C_n^*(\boldsymbol{\theta}) = \exp\{(n/2)D^1h_n^*(\boldsymbol{\theta})^T[D^2h^*(\boldsymbol{\theta})]^{-1}D^1h_n^*(\boldsymbol{\theta})\}$ at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, we have

$$C_n^*(\hat{\boldsymbol{\theta}}) = \exp\left\{\frac{1}{2n}\mathbf{t}^T[D^2h_n(\hat{\boldsymbol{\theta}})]^{-1}\mathbf{t}\right\},$$

because $D^1h_n^*(\hat{\boldsymbol{\theta}}) = -n^{-1}\mathbf{t}$. Hence, substituting these results into (4) yields (5).

As noted by Kass et al. (1990), the regularity conditions, for which usual Laplace approximations to posterior expectations are valid, are closely related to those of asymptotic normality of the posterior. Hence, it is reasonable to suppose that approximate MGFs are asymptotically equivalent to those of a multivariate normal distribution $\mathrm{N}(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$, namely $\exp\{\mathbf{t}^T\hat{\boldsymbol{\theta}} + (1/2)\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}\}$. However, the Tierney–Kadane and the standard-form approximations do not have the part $\exp\{(1/2)\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}\}$, because they do not have $C_n^*(\hat{\boldsymbol{\theta}})$. Therefore, the approximation (5) is more efficient. In Section 4 we give an Edgeworth-like expansion for the distribution functions of standardized random variables from the result of Proposition 1.

Incidentally, to obtain the expansion of Theorem 2, we need only assume that $(\{h_n^*\}, \{\hat{\boldsymbol{\theta}}\})$ and $(\{h_n\}, \{\hat{\boldsymbol{\theta}}\})$ satisfy the analytical assumptions for the asymptotic-mode Laplace method, that is, (A1), (A2), (A3), (A4), and (A5). In applications, it is often easier to check an alternative condition that is stronger than the assumption (A5):

(A5′) For all δ for which $0 < \delta < \epsilon$, $B_\delta(\hat{\boldsymbol{\theta}}) \subseteq \boldsymbol{\Theta}$ and

$$\limsup_{n\to\infty}\sup_{\boldsymbol{\theta}}\left\{\bar{h}_n(\hat{\boldsymbol{\theta}}) - \bar{h}_n(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta} - B_\delta(\hat{\boldsymbol{\theta}})\right\} < 0,$$

where $\bar{h}(\boldsymbol{\theta}) = -n^{-1}\log p(\mathbf{x}|\boldsymbol{\theta})$.

The assumptions (A1), (A2), (A3), (A4), and (A5′) are similar to those stated by Kass et al. (1990), except that $\hat{\boldsymbol{\theta}}$ is not the exact mode of $-h_n(\boldsymbol{\theta})$. The following proposition shows that (A1), (A2), (A3), (A4), and (A5′) are sufficient conditions for the asymptotic-mode Laplace method for $(\{h_n^*\}, \{\hat{\boldsymbol{\theta}}\})$ and $(\{h_n\}, \{\hat{\boldsymbol{\theta}}\})$. The proof of this proposition is given in Appendix C.

Proposition 2. Suppose that $(\{h_n^*(\boldsymbol{\theta})\}, \{\hat{\boldsymbol{\theta}}\})$ and $(\{h_n(\boldsymbol{\theta})\}, \{\hat{\boldsymbol{\theta}}\})$ satisfy the assumptions (A1), (A2), (A3), (A4), and (A5′) and that for some nonnegative integer n_0, the posterior density $p(\boldsymbol{\theta}|\mathbf{x}^{(n_0)})$, based on $\mathbf{x}^{(n_0)} = (x_1,\dots,x_{n_0})$ exists and has the finite expectation of $g(\boldsymbol{\theta})$. Then there exist constants c_1 and c_2 such that $\int_{\boldsymbol{\Theta}-B_\delta(\hat{\boldsymbol{\theta}})}\exp\{-nh_n(\boldsymbol{\theta})\}\,d\boldsymbol{\theta} = O(e^{-nc_1})$ and $\int_{\boldsymbol{\Theta}-B_\delta(\hat{\boldsymbol{\theta}})} g(\boldsymbol{\theta})\exp\{-nh_n(\boldsymbol{\theta})\}\,d\boldsymbol{\theta} = O(e^{-nc_2})$.

3.2 Approximations via Different Asymptotic Modes

Let $\hat{\boldsymbol{\theta}}$ be an asymptotic mode of $-h_n(\boldsymbol{\theta})$ and $\hat{\boldsymbol{\theta}}_N$ a single Newton step from $\hat{\boldsymbol{\theta}}$ toward the maximum of $-h_n^*$. In this section, we propose the following second-order approximations to the ratio of integrals obtained by expanding the numerator and denominator of (2) at $\hat{\boldsymbol{\theta}}_N$ and $\hat{\boldsymbol{\theta}}$.

Theorem 3. Let $g(\boldsymbol{\theta})$ be a strictly positive function on $\boldsymbol{\Theta}$, let $\hat{\boldsymbol{\theta}}$ be the exact mode of $-h_n$ and let $\hat{\boldsymbol{\theta}}_N$ be a single Newton step from $\hat{\boldsymbol{\theta}}$ toward the maximum of $-h_n^*(\boldsymbol{\theta})$, that is, $\hat{\boldsymbol{\theta}}_N = \hat{\boldsymbol{\theta}} - [D^2h_n^*(\hat{\boldsymbol{\theta}})]^{-1}D^1h_n^*(\hat{\boldsymbol{\theta}})$. If $(\{h_n\}, \{\hat{\boldsymbol{\theta}}\})$ and $(\{h_n^*\}, \{\hat{\boldsymbol{\theta}}_N\})$ satisfy the analytical assumptions for the asymptotic-mode Laplace method, then it follows that

$$E[g(\boldsymbol{\theta})] = \left(\frac{|D^2h_n(\hat{\boldsymbol{\theta}})|}{|D^2h_n^*(\hat{\boldsymbol{\theta}}_N)|}\right)^{1/2}C_n^*(\hat{\boldsymbol{\theta}}_N) \times\exp\left\{-n[h_n^*(\hat{\boldsymbol{\theta}}_N) - h_n(\hat{\boldsymbol{\theta}})]\right\}\left(1 + O(n^{-2})\right), \quad (6)$$

where $C_n^*(\hat{\boldsymbol{\theta}}_N) = \exp\{(n/2)D^1h_n^*(\hat{\boldsymbol{\theta}}_N)^T[D^2h_n^*(\hat{\boldsymbol{\theta}}_N)]^{-1} \times D^1h_n^*(\hat{\boldsymbol{\theta}}_N)\}$.

Equation (6) is simpler in the sense that evaluation of the third derivatives of log-likelihood functions is not required. In Theorem 3, the approximation with $C_n^*(\hat{\boldsymbol{\theta}}_N)$ replaced by 1 is due to Tierney and Kadane (1986). In fact, from $C_n^*(\hat{\boldsymbol{\theta}}_N) = 1 + O(n^{-3})$, this is also a second-order approximation. We denote the second Newton step by $\hat{\boldsymbol{\theta}}_{N2} = \hat{\boldsymbol{\theta}}_N - [D^2h_n^*(\hat{\boldsymbol{\theta}}_N)]^{-1}D^1h_n^*(\hat{\boldsymbol{\theta}}_N)$. Arguing as in the proof of Theorem 3, we have $D^1h_n^*(\hat{\boldsymbol{\theta}}_{N2}) = O(n^{-4})$. Therefore, replacing $\hat{\boldsymbol{\theta}}_N$ by the second Newton step $\hat{\boldsymbol{\theta}}_{N2}$ in (6) yields a second-order approximation. Finally, the extension of Theorem 3 is given without using either of the exact modes of the numerator and denominator of (2).

Theorem 4. Let $g(\boldsymbol{\theta})$ be a strictly positive function on $\boldsymbol{\Theta}$, let $\hat{\boldsymbol{\theta}}$ be an asymptotic mode of order n^{-2} for $-h_n$ and let $\hat{\boldsymbol{\theta}}_N$ be the single Newton step, that is, $\hat{\boldsymbol{\theta}}_N = \hat{\boldsymbol{\theta}} - [D^2 h_n^*(\hat{\boldsymbol{\theta}})]^{-1} D^1 h_n^*(\hat{\boldsymbol{\theta}})$. If $(\{h_n\}, \{\hat{\boldsymbol{\theta}}\})$ and $(\{h_n^*\}, \{\hat{\boldsymbol{\theta}}_N\})$ satisfy the analytical assumptions for the asymptotic-mode Laplace method, then

$$E[g(\boldsymbol{\theta})] = \left(\frac{|D^2 h_n(\hat{\boldsymbol{\theta}})|}{|D^2 h_n^*(\hat{\boldsymbol{\theta}}_N)|}\right)^{1/2} \frac{C_n^*(\hat{\boldsymbol{\theta}}_N)}{C_n(\hat{\boldsymbol{\theta}})} \times \exp\{-n[h_n^*(\hat{\boldsymbol{\theta}}_N) - h_n(\hat{\boldsymbol{\theta}})]\}\big(1 + O(n^{-2})\big), \quad (7)$$

where $C_n^*(\hat{\boldsymbol{\theta}}_N) = \exp\{(n/2) D^1 h_n^*(\hat{\boldsymbol{\theta}}_N)^T [D^2 h_n^*(\hat{\boldsymbol{\theta}}_N)]^{-1} \times D^1 h_n^*(\hat{\boldsymbol{\theta}}_N)\}$.

An example of the asymptotic mode of order n^{-2} for $-h_n$ is given in Remark 2. The proof is provided in Appendix C.

4. APPROXIMATIONS TO CUMULATIVE DISTRIBUTION FUNCTIONS

This section presents an Edgeworth-like expansion for the random variable with a posterior distribution. Suppose that $\tilde{\Theta}$ is a random variable according to the posterior density $p(\theta|\mathbf{x})$ and that $P(\tilde{\Theta} \in A)$ denotes the posterior probability $\int_A p(\theta|\mathbf{x}) d\theta$. Let $n(\cdot|0,1)$ be the density of a normal variable with mean 0 and variance 1 and let $\Phi(\alpha_0)$ be the standard normal distribution function. Using Proposition 1 yields the following expansion.

Proposition 3. Suppose that Θ is an open interval of $\mathbb{R}$. Let $h_n(\theta) = -n^{-1}\log p(\mathbf{x}|\theta)\pi(\theta)$ and $h_n^*(\theta) = h_n(\theta) - (t\theta)/n$ and let $\hat{\theta}$ be exact mode of $-h_n(\theta)$. If $(\{h_n^*(\theta)\}, \{\hat{\theta}\})$ and $(\{h_n(\theta)\}, \{\hat{\theta}\})$ satisfy the analytical assumptions for the asymptotic-mode Laplace method, then it follows that

$$P\left(\frac{\tilde{\Theta} - \hat{\theta}}{\sqrt{1/(nh''(\hat{\theta}))}} \le \alpha_0\right) = \Phi(\alpha_0) + \frac{n(\alpha_0|0,1) h'''(\hat{\theta})}{2\sqrt{n}\cdot h''(\hat{\theta})^{3/2}} + O(n^{-1}), \quad (8)$$

where $h_n(\theta) = -n^{-1}\log p(\mathbf{x}|\theta)\pi(\theta)$, $h' = dh_n/d\theta$, $h'' = d^2 h_n/d\theta^2$ and $h''' = d^3 h_n/d\theta^3$.

The proof of Proposition 3, given in Appendix C, includes a heuristic argument. It is worth mentioning in passing that Johnson (1967, 1970) gave asymptotic expansions for the random variables according to posterior distributions. His asymptotic expansions are given by

$$P\left(\frac{\tilde{\Theta} - \hat{\theta}_{\mathrm{ML}}}{\sqrt{1/(n\bar{h}''(\hat{\theta}_{\mathrm{ML}}))}} \le \alpha_0\right) = \Phi(\alpha_0) + \frac{n(\alpha_0|0,1)}{6\sqrt{n}}\left\{\frac{\bar{h}'''(\hat{\theta}_{\mathrm{ML}})(\alpha_0^2+2)}{\bar{h}''(\hat{\theta}_{\mathrm{ML}})^{3/2}} - \frac{6}{\bar{h}''(\hat{\theta}_{\mathrm{ML}})^{1/2}}\frac{\partial}{\partial\theta}\log\pi(\hat{\theta}_{\mathrm{ML}})\right\} + O(n^{-1}), \quad (9)$$

where $\hat{\theta}_{\mathrm{ML}}$ is the MLE of θ, $\bar{h}(\theta) = -n^{-1}\log p(\mathbf{x}|\theta)$ and $\pi(\theta)$ is the prior. Note that the relative error of order $n^{-1/2}$ of (9) is not equal to that of (8) even if the prior π is proportional to a constant. We illustrate these two expansions in Example 2 of Section 7.

5. THE MOMENT–GENERATING FUNCTION METHOD

Tierney et al. (1989) obtained a second-order approximation to the expectation of a general function $g(\boldsymbol{\theta})$ (not necessarily positive) by applying the fully exponential method to approximate the MGF $E[\exp\{tg(\boldsymbol{\theta})\}]$ and then differentiating. This method gives a standard-form approximation without calculating the third derivatives of a log-likelihood. We now sketch the MGF method. First, we have the Tierney–Kadane approximation, $\hat{M}(t)$, to the MGF,

$$M(t) = \frac{\int_{\boldsymbol{\Theta}} e^{tg(\boldsymbol{\theta})} p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}} p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}} = \frac{\int_{\boldsymbol{\Theta}} e^{-nh_n^*(\boldsymbol{\theta})}\,d\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}} e^{-nh_n(\boldsymbol{\theta})}\,d\boldsymbol{\theta}}, \quad (10)$$

where $h_n(\boldsymbol{\theta}) = -n^{-1}\log p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ and $h_n^*(\boldsymbol{\theta}) = h_n(\boldsymbol{\theta}) - (t/n)g(\boldsymbol{\theta})$.
Second, differentiating $\hat{M}(t)$ at $t = 0$ yields a second-order approximation to $E[g(\boldsymbol{\theta})]$, that is,

$$E[g(\boldsymbol{\theta})] = \frac{d}{dt}\hat{M}(t)\bigg|_{t=0} + O(n^{-2}). \quad (11)$$

But it can be impossible to find the exact mode of $-h_n^*(\boldsymbol{\theta})$ when $\hat{M}(t)$ is approximated, because h_n^* has an indeterminate variable t. In this section we use the approximation (4) and discuss the relationship between a standard-form approximation and the MGF method.

Suppose that $\hat{\boldsymbol{\theta}}$ is an asymptotic mode of order n^{-1} for $-h_n$. Because $\exp(tg(\boldsymbol{\theta}))$ is always positive, applying Theorem 2 to $M(t) = E[\exp(tg(\boldsymbol{\theta}))]$ yields

$$E\big[e^{tg(\boldsymbol{\theta})}\big] = e^{tg(\hat{\boldsymbol{\theta}})}\left(\frac{|D^2 h_n(\hat{\boldsymbol{\theta}})|}{|D^2 h_n^*(\hat{\boldsymbol{\theta}})|}\right)^{1/2}\frac{C_n^*(\hat{\boldsymbol{\theta}})}{C_n(\hat{\boldsymbol{\theta}})} \times \left(1 - \frac{t}{2n}\sum_{ijkq} h_{ijk}(\hat{\boldsymbol{\theta}}) h^{iq} h^{jk}\frac{\partial}{\partial\theta_q} g(\hat{\boldsymbol{\theta}}) + O^t(n^{-2})\right), \quad (12)$$

where $C_n^*(\hat{\boldsymbol{\theta}}) = \exp\{(n/2) D^1 h_n^*(\hat{\boldsymbol{\theta}})^T [D^2 h_n^*(\hat{\boldsymbol{\theta}})]^{-1} D^1 h_n^*(\hat{\boldsymbol{\theta}})\}$ and $O^t(\cdot)$ indicates that the error term may depend on t. Note that $C_n^*(\hat{\boldsymbol{\theta}})$ depends on t, but $C_n(\hat{\boldsymbol{\theta}})$ does not. Let the part neglecting the error term $O^t(n^{-2})$ in (12) be $\tilde{M}(t)$. Then, putting $d\tilde{M}(t)/dt|_{t=0} = \tilde{E}[g(\boldsymbol{\theta})]$, we have the following relationship between $\tilde{E}[g(\boldsymbol{\theta})]$ and a standard-form approximation. The proof of this theorem is given in Appendix C.

Theorem 5. Let $g(\boldsymbol{\theta})$ be a general function (not necessarily positive) on $\boldsymbol{\Theta}$, and let $\hat{\boldsymbol{\theta}}$ be an asymptotic mode of order n^{-1} for $-h_n$. Suppose that $(\{h_n\}, \{\hat{\boldsymbol{\theta}}\})$ and $(\{h_n^*\}, \{\hat{\boldsymbol{\theta}}\})$ satisfy the analytical assumptions for the asymptotic-mode Laplace method. Then $\tilde{E}[g(\boldsymbol{\theta})]$ is expressed as

$$\tilde{E}[g(\boldsymbol{\theta})] = g(\hat{\boldsymbol{\theta}}) + \frac{1}{2n}\mathrm{tr}\big(D^2 g(\hat{\boldsymbol{\theta}})[D^2 h_n(\hat{\boldsymbol{\theta}})]^{-1}\big) - \frac{1}{2n}\sum_{ijkq} h_{ijk}(\hat{\boldsymbol{\theta}}) h^{iq} h^{jk}\frac{\partial}{\partial\theta_q} g(\hat{\boldsymbol{\theta}})$$

$$- D^1 g(\hat{\boldsymbol\theta})^T [D^2 h_n(\hat{\boldsymbol\theta})]^{-1} D^1 h_n(\hat{\boldsymbol\theta}) \quad (13)$$

$$+ \frac{1}{2} D^1 h_n(\hat{\boldsymbol\theta})^T [D^2 h_n(\hat{\boldsymbol\theta})]^{-1} \times D^2 g(\hat{\boldsymbol\theta}) [D^2 h_n(\hat{\boldsymbol\theta})]^{-1} D^1 h_n(\hat{\boldsymbol\theta}), \quad (14)$$

where $\mathrm{tr}(\cdot)$ indicates the trace of a matrix.

Note that because the term (14) is $O(n^{-2})$, dropping the term of order n^{-2} in Theorem 5 yields a standard-form approximation. Furthermore, if $\hat{\boldsymbol\theta}$ is the exact mode of $-h_n(\boldsymbol\theta)$, then $\tilde{E}[g(\boldsymbol\theta)]$ has a simpler form because the terms (13) and (14) vanish,

$$\tilde{E}[g(\boldsymbol\theta)] = g(\hat{\boldsymbol\theta}) + \frac{1}{2n} \mathrm{tr}\big(D^2 g(\hat{\boldsymbol\theta})[D^2 h(\hat{\boldsymbol\theta})]^{-1}\big) - \frac{1}{2n} \sum_{ijkq} h_{ijk}(\hat{\boldsymbol\theta}) h^{iq} h^{jk} \frac{\partial}{\partial \theta_q} g(\hat{\boldsymbol\theta}).$$

6. PREDICTIVE DENSITIES

The predictive density of an unknown random variable Y, given the observation $\mathbf{x}$, has an important role in statistical problems. The predictive density is written as

$$p(y|\mathbf{x}) = \frac{\int_\Theta p(y|\mathbf{x},\boldsymbol\theta) p(\mathbf{x}|\boldsymbol\theta) \pi(\boldsymbol\theta)\, d\boldsymbol\theta}{\int_\Theta p(\mathbf{x}|\boldsymbol\theta)\pi(\boldsymbol\theta)\, d\boldsymbol\theta} = E[p(y|\mathbf{x},\boldsymbol\theta)]. \quad (15)$$

Let $h_n^*(\boldsymbol\theta) = -n^{-1} \log p(y|\mathbf{x},\boldsymbol\theta) p(\mathbf{x}|\boldsymbol\theta)\pi(\boldsymbol\theta)$ and $h_n(\boldsymbol\theta) = -n^{-1} \log p(\mathbf{x}|\boldsymbol\theta)\pi(\boldsymbol\theta)$. When the Tierney–Kadane approximation is applied to (15), the exact mode of $-h_n^*(\boldsymbol\theta)$ is required. However, it may be difficult to express the exact mode of $-h_n^*$ explicitly, because $-h_n^*$ includes an unobserved variable y. In such a case we propose the approximate predictive density

$$p^{**}(y|\mathbf{x}) = c(\mathbf{x}) |\boldsymbol\Sigma_N^*|^{1/2} p(\mathbf{x}, y|\hat{\boldsymbol\theta}_N) \pi(\hat{\boldsymbol\theta}_N) C_n^*(\hat{\boldsymbol\theta}_N), \quad (16)$$

where $\hat{\boldsymbol\theta}$ is the posterior mode, $\hat{\boldsymbol\theta}_N = \hat{\boldsymbol\theta} - [D^2 \log p(\mathbf{x}, y|\hat{\boldsymbol\theta}) \times \pi(\hat{\boldsymbol\theta})]^{-1} D^1 \log p(y|\mathbf{x},\hat{\boldsymbol\theta})$, $\boldsymbol\Sigma_N^* = n^{-1}[D^2 h_n^*(\hat{\boldsymbol\theta}_N)]^{-1}$, and $c(\mathbf{x})$ is a normalizing constant. Equation (16) is easily given from the approximation (17). Note that $c(\mathbf{x}) = \int |\boldsymbol\Sigma_N^*|^{1/2} C_n^*(\hat{\boldsymbol\theta}_N) p(y|\mathbf{x},\hat{\boldsymbol\theta}_N)\pi(\hat{\boldsymbol\theta}_N)\, dy$ usually needs to be evaluated numerically. The approximation (16) requires a normalizing constant as well as any other approximate predictive densities, including the maximum likelihood predictive density of Lejeune and Faulkenberry (1982), the one considered by Leonard (1982), and so on. If the rigorous normalizing constant is not required, then we propose the following approximation (6) to (15):

$$\hat{p}(y|\mathbf{x}) = \left(\frac{|D^2 h_n(\hat{\boldsymbol\theta})|}{|D^2 h_n^*(\hat{\boldsymbol\theta}_N)|}\right)^{1/2} \frac{p(\mathbf{x}, y|\hat{\boldsymbol\theta}_N)\pi(\hat{\boldsymbol\theta}_N) C_n^*(\hat{\boldsymbol\theta}_N)}{p(\mathbf{x}|\hat{\boldsymbol\theta})\pi(\hat{\boldsymbol\theta})}. \quad (17)$$

This approximation is illustrated in Example 3 of the next section. In statistical prediction problem, the mode of a predictive density is a useful predictor for an unobserved random variable. It is seen that the mode of the approximation (16) can be found without calculating the normalizing constant.

7. EXAMPLES

In this section we explore the performance of some approximations in Sections 3–6.

Example 1. For a beta posterior distribution with density proportional to $\theta^{\alpha-1}(1-\theta)^{\beta-1}$, the exact mean is $E[\theta] = \alpha/(\alpha+\beta)$. Let

$$E[\theta] = \frac{\int \theta \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta}{\int \theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta} = \frac{\int e^{-nh_n^*(\theta)}\, d\theta}{\int e^{-nh_n(\theta)}\, d\theta}, \quad (18)$$

where $h_n(\theta) = -(1/n)\log \theta^{\alpha-1}(1-\theta)^{\beta-1}$ and $h_n^*(\theta) = h(\theta) - (1/n)\log\theta$. The posterior mode is given by $\hat\theta = (\alpha-1)/(\alpha+\beta-2)$. Let $\hat{E}[\theta]$ and $\tilde{E}[\theta]$ denote the approximation (4) and the approximation in Theorem 5. First, we compare the two approximations asymptotically. Because the beta random variable is a positive random variable, the approximation (4) may be applied directly with the posterior mode $\hat\theta$, giving

$$\hat{E}[\theta] = \frac{\alpha-1}{\alpha+\beta-2}\left(\frac{(\alpha+\beta-2)(\alpha-1)}{(\alpha-1)(\alpha+\beta-2)+(\beta-1)}\right)^{1/2} \times \exp\left\{\frac{\beta-1}{2(\alpha-1)(\alpha+\beta-1)+2(\beta-1)}\right\} \times \left(1 + \frac{\beta-\alpha}{(\alpha+\beta-2)(\alpha-1)}\right).$$

Applying Theorem 5 to (18) with the posterior mode $\hat\theta$, we have

$$\tilde{E}[\theta] = \frac{\alpha^2+\alpha\beta-4\alpha+2}{(\alpha+\beta-2)^2}. \quad (19)$$

Detailed calculations are available from the author. To compare the errors in these two approximations, set $p = \alpha/(\alpha+\beta)$ and $n = \alpha+\beta$. Then, for fixed p, the error in the approximation $\tilde{E}[\theta]$ is

$$\tilde{E}[\theta] - E[\theta] = \frac{\alpha^2+\alpha\beta-4\alpha+2}{(\alpha+\beta-2)^2} - p = \frac{2(1-2p)}{n^2} + O(n^{-3}), \quad (20)$$

whereas the error in the direct approximation (4) is

$$\hat{E}[\theta] - E[\theta] = \frac{1}{n^2}\left(\frac{-17p^2+10p-1}{4p}\right) + O(n^{-3}). \quad (21)$$

Figure 1 shows a comparison of the two leading error terms multiplied by n^2.

Subsequently, to compare some methods in previous sections numerically, we would like to give a table with some approximations to the beta posterior means for $p = 1/4$ and $3/5$. In Table 1, MGF indicates the approximation in Theorem 5 with the posterior mode $\hat\theta$, and FEAM indicates the fully exponential asymptotic-mode approximation (4) with the posterior mode $\hat\theta$. One Newton iteration and two Newton iterations denote the approximation (6) and the approximation with $\hat\theta_N$ replaced by the second Newton step $\hat\theta_{2N}$ in (6). Both of these approximations are also based on the posterior mode $\hat\theta$. Neglect C_n^* indicates the approximation with $C_n^*(\hat\theta_N) = 1$ in one Newton iteration.

We see from Table 1 that the MGF method and two Newton iterations method are stable and effective, especially when n is not large. We also see that for $p = 1/4$ and $3/5$, the approximation (6) is superior to the one with $C_n^*(\hat\theta_N) = 1$ in the

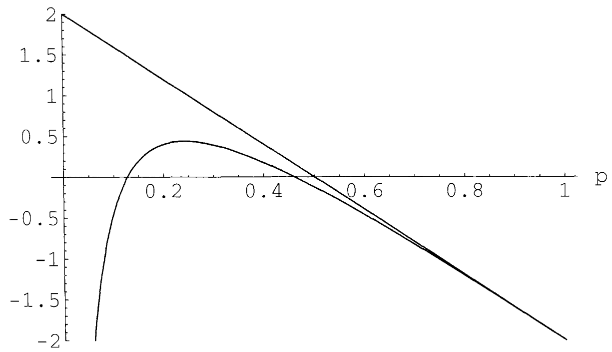

Figure 1. Leading Terms of the Error in Approximating the Mean of the Beta(np, n(1 − p)) Distribution. The horizontal axis is p, the straight line is the MGF approximation, and the curved line is the approximation $\hat{E}[\theta]$.

one Newton iteration method. Note that the MGF method is the same as a standard-form approximation in this case.

Example 2. Suppose that $X_1, \dots, X_n$ are iid according to a gamma distribution with density $p(x_i|\theta) = \Gamma(\alpha)^{-1}\theta^\alpha x_i^{\alpha-1} \times \exp\{-\theta x_i\}$, where θ is unknown and α is known. Let the Jeffreys prior be $\pi(\theta) \propto \theta^{-1}$. Because the posterior is given by

$$p(\theta|\mathbf{x}) \propto \theta^{n\alpha-1}\exp\left\{-\theta\sum_{i=1}^n x_i\right\}, \tag{22}$$

the posterior mode is $\hat\theta = (n\alpha - 1)/\sum_{i=1}^n x_i$. Hence the expansion of Proposition 2 is given by

$$P\left(\sqrt{nh''(\hat\theta)}(\tilde\Theta - \hat\theta) \le \alpha_0\right) = \Phi(\alpha_0) - \frac{1}{\sqrt{n\alpha - 1}}n(\alpha_0|0,1) + O(n^{-1}), \tag{23}$$

where $nh''(\hat\theta) = (n\alpha - 1)/\hat\theta^2$. In contrast, using the MLE $\hat\theta_{\mathrm{ML}} = n\alpha/\sum_{i=1}^n x_i$ of θ, Johnson's expansion is expressed as

$$P\left(\sqrt{n\bar h''(\hat\theta_{\mathrm{ML}})}(\tilde\Theta - \hat\theta_{\mathrm{ML}}) \le \alpha_0\right) = \Phi(\alpha_0) - \frac{\alpha_0^2 - 1}{3\sqrt{n\alpha}}n(\alpha_0|0,1) + O(n^{-1}), \tag{24}$$

Table 1. The Numerical Results in Some Approximations for the Beta Posterior Mean

Sample size n	FEAM	MGF	One Newton iteration	Two Newton iterations	Neglect C_n^*
$p = 1/4$					
6	.4521	.3125	.1955	.2502	.1762
8	.3112	.2778	.2272	.2518	.2201
10	.2805	.2656	.2372	.2512	.234
12	.2684	.26	.2417	.2508	.2401
14	.2623	.2569	.2442	.2505	.2432
$p = 3/5$					
6	.5774	.575	.5695	.5785	.5691
8	.5902	.5889	.5858	.5892	.5857
10	.5945	.5938	.5918	.5935	.5918
12	.5905	.596	.5947	.5957	.5946
14	.5976	.5972	.5963	.5969	.5962

where $\bar h_n = -n^{-1}\sum_{i=1}^n \log p(x_i|\theta)$ and $\bar h''(\hat\theta_{\mathrm{ML}}) = (n\alpha)/\hat\theta_{\mathrm{ML}}^2$.

Example 3. Suppose that $X_1, X_2, \dots, X_m$ are independent random variables according to a Weibull distribution with density

$$p(x|\beta) = \beta\alpha x^{\alpha-1}\exp\{-\beta x^\alpha\}, \tag{25}$$

where $\alpha > 0$ is known, $\beta > 0$ is unknown, and $x \ge 0$. We denote the Weibull distribution with the parameters α and β by $WEIB(\alpha,\beta)$. Let $X_{(1)} < \cdots < X_{(n)} < \cdots < X_{(m)}$ be the order statistics corresponding to $X_1, X_2, \dots, X_m$. Suppose that $\mathbf{X} = (X_{(1)}, \dots, X_{(n)})$ $(n < m)$ is an observed random vector and that $Y = X_{(m)}$ is an unobserved random variable. First, we calculate the predictive density of Y, given the observed random vector $\mathbf{X}$. Assume that the prior of β is flat, namely $\pi(\beta) \propto$ constant. Then the density functions of $(\mathbf{X}, Y)$ and $\mathbf{X}$ are given by

$$p(\mathbf{x}|\beta) = \frac{m!\alpha^n\beta^n}{(m-n)!}\left(\prod_{j=1}^n x_{(j)}\right)^{\alpha-1} \times \exp\left\{-\beta\left[\sum_{j=1}^n x_{(j)}^\alpha + (m-n)x_{(n)}^\alpha\right]\right\} \tag{26}$$

and

$$p(\mathbf{x}, y|\beta) = \frac{m!\alpha^{n+1}\beta^{n+1}}{(m-n-1)!}\left(\prod_{j=1}^n x_{(j)}\right)^{\alpha-1} y^{\alpha-1} \times \exp\left\{-\beta\left[\sum_{j=1}^n x_{(j)}^\alpha + y^\alpha\right]\right\}. \tag{27}$$

(For the properties of order statistics, see David and Nagaraja 2003.) Hence the exact predictive density of Y given $\mathbf{X}$ is

$$p(y|\mathbf{x}) = (m-n)(n+1)\left[\sum_{j=1}^n x_{(j)}^\alpha + (m-n)x_{(n)}^\alpha\right]^{n+1}\alpha y^{\alpha-1} \times \sum_{i=0}^{m-n-1}\binom{m-n-1}{i} \times \frac{(-1)^i}{[(m-n-1)x_{(n)}^\alpha + \sum_{j=1}^n x_{(j)}^\alpha + (i+1)y^\alpha]^{n+2}}, \tag{28}$$

where $y > x_{(n)}$. Subsequently, we approximate the predictive density which is written as

$$p(y|\mathbf{x}) = \frac{\int p(\mathbf{x}, y|\beta)\pi(\beta)\,d\beta}{\int p(\mathbf{x}|\beta)\pi(\beta)\,d\beta} = \frac{\int \exp\{-nh_n^*(\beta)\}\,d\beta}{\int \exp\{-nh_n(\beta)\}\,d\beta}, \tag{29}$$

where $h_n^*(\beta) = -n^{-1}\log p(\mathbf{x}, y|\beta)$, $h_n(\beta) = -n^{-1}\log p(\mathbf{x}|\beta)$ and the integrals are taken over $\beta > 0$. Although the exact mode of $-h_n(\beta)$ can be easily found, that of $-h_n^*(\beta)$ cannot be given in an explicit form. Hence, using Theorem 3, we have the second-order approximation

$$\hat{p}(y|\mathbf{x}) = \left(\frac{h''(\hat{\beta})}{h^{*''}(\hat{\beta}_N)}\right)^{1/2} \frac{(m-n)\alpha y^{\alpha-1}}{\hat{\beta}^n \exp\{-n\}} \times \left[\exp\{-\hat{\beta}_N x_{(n)}^\alpha\} - \exp\{-\hat{\beta}_N y^\alpha\}\right]^{m-n-1} \times \exp\left\{-\hat{\beta}_N\left(\sum_{j=1}^n x_{(j)}^\alpha + y^\alpha\right)\right\} C_n^*(\hat{\beta}_N), \quad (30)$$

where $\hat{\beta} = n/[\sum_{j=1}^n x_{(j)}^\alpha + (m-n)x_{(n)}^\alpha]$ is the exact mode of $-h_n$, $h''(\beta) = \beta^{-2}$,

$$h^{*'}(\beta) = \frac{n+1}{n\beta} + \frac{(m-n-1)(y^\alpha \exp\{-\beta y^\alpha\} - x_{(n)}^\alpha \exp\{-\beta x_{(n)}^\alpha\})}{n(\exp\{-\beta x_{(n)}^\alpha\} - \exp\{-\beta y^\alpha\})} - \sum_{j=1}^n x_{(j)}^\alpha - y^\alpha,$$

$$h^{*''}(\beta) = \frac{n+1}{n\beta^2} + \frac{(m-n-1)(y^\alpha - x_{(n)}^\alpha)^2 \exp\{-\beta(y^\alpha + x_{(n)}^\alpha)\}}{n(\exp\{-\beta x_{(n)}^\alpha\} - \exp\{-\beta y^\alpha\})^2},$$

$$\hat{\beta}_N = \hat{\beta} - h^{*''}(\hat{\beta})^{-1} h^{*'}(\hat{\beta}),$$

and

$$C_n^*(\hat{\beta}_N) = \exp\{nh^{*'}(\hat{\beta}_N)^2/(2h^{*''}(\hat{\beta}_N))\}.$$

In contrast, using the MLE $\hat{\beta}_{\mathrm{ML}}$, the standard-form approximation is given by

$$p(y|\mathbf{x}) = p(y|\mathbf{x}, \hat{\beta}_{\mathrm{ML}}) + \frac{\hat{\beta}_{\mathrm{ML}}}{n} \cdot \frac{d}{d\beta} p(y|\mathbf{x}, \hat{\beta}_{\mathrm{ML}}) + \frac{\hat{\beta}_{\mathrm{ML}}^2}{2n} \cdot \frac{d^2}{d\beta^2} p(y|\mathbf{x}, \hat{\beta}_{\mathrm{ML}}), \quad (31)$$

where $\hat{\beta}_{\mathrm{ML}} = n/[\sum_{j=1}^n x_{(j)}^\alpha + (m-n)x_{(n)}^\alpha]$ is the MLE for $p(\mathbf{x}|\beta)$ and

$$p(y|\mathbf{x}, \beta) = (m-n)\beta\alpha y^{\alpha-1} \exp\{-\beta[y^\alpha - (m-n)x_{(n)}^\alpha]\} \times \left[\exp\{-\beta x_{(n)}^\alpha\} - \exp\{-\beta y^\alpha\}\right]^{m-n-1}. \quad (32)$$

To compare the approximate predictive densities numerically, we carry out a simulation. Let $x_1, x_2, \ldots, x_{40}$ be a sample of size 40 from *WEIB*(2, 4) and let $x_{(1)} < x_{(2)} < \cdots < x_{(40)}$ denote the ordered sample. Let $\mathbf{x} = (x_{(1)}, \ldots, x_{(20)})$ be the data from the minimum value to the 20th smallest value in the ordered sample. We then calculate $\hat{p}(y|\mathbf{x})$ based on the observations $\mathbf{x}$. Figure 2 shows the approximation (30), the standard-form approximation in Appendix A and the exact predictive density produced by substitution of the statistics $\sum_{j=1}^{20} x_{(j)} = 1.8817$ and $x_{(20)} = .474482$. Although the data used in calculating

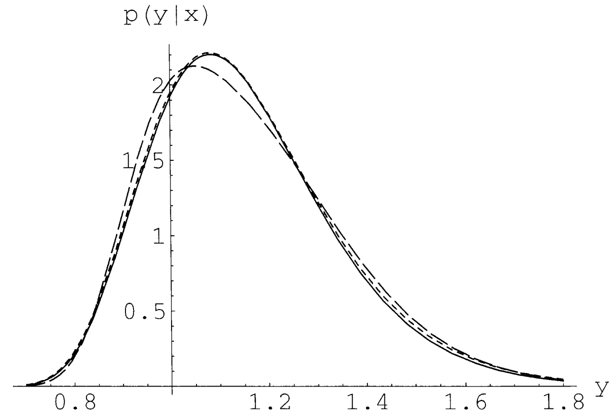

Figure 2. The Exact and Approximate Predictive Densities: —— Approximation (30); - - - - - - Exact Predictive Density; – – – Standard-Form Approximation (31).

these two statistics are omitted, they are available from the author on request.

In Figure 2, the solid line is the approximation (30), the long-dashed line is the standard-form approximation (31), and the short-dashed line is the exact predictive density. It is seen that the approximation (30) is very close to the exact predictive density, because the solid line lies almost upon the broken line. However, the approximation (30) for large y may not be available, because $\hat{\beta}_N$ becomes negative. Note in passing that the respective modes of the exact predictive density, (30), and (31) are 1.08162, 1.08359, and 1.05126.

8. CONCLUDING REMARKS

In this article we have presented second-order approximations to the posterior expectations of positive functions without calculating the third derivatives of a log-likelihood function and the exact modes of integrands, using the Laplace method with an asymptotic mode. To verify the accuracy of the approximation (30), we repeated the simulation of Example 3 20 times. We then obtained the results similar to those depicted in Figure 2. However, we should emphasize that to say that an expansion is valid asymptotically is not to say that it provides a good approximation in any particular instance. In practice, it is not hard to construct examples in which the approximations perform poorly.

In contrast, a new Edgeworth-like expansion for the random variable with a posterior distribution was given in Section 4. However, when demonstrating the first-order accuracy of the Edgeworth-like expansion in Appendix C, we make the heuristic argument that the purely imaginary number is was substituted into the approximate MGF (C.18) to obtain the approximate characteristic function. Thus further discussions will be needed to demonstrated the validity of our Edgeworth-like expansion. Incidentally, the posterior mode is required when Laplace approximations are applied to a posterior mean. For complex models, it is easier to find the estimated posterior mode, such as the multivariate median or L^1-center (see, e.g., Lewis and Raftery 1997). We believe that the approximations (6) and (7) with the asymptotic modes replaced by the estimated posterior means also perform well. However, further work will also be needed in these cases.

APPENDIX A: STANDARD–FORM APPROXIMATIONS

Here we sketch out standard-form approximations for the posterior means. (For details, see Lindley 1967, 1980; Kass et al. 1990.)

Given a prior π, log-likelihood $\log p(\mathbf{x}|\boldsymbol{\theta})$, positive function γ, and real function g, we define h_n and ρ by $-nh_n(\boldsymbol{\theta}) = \log p(\mathbf{x}|\boldsymbol{\theta}) + \log \gamma$ and $\rho = \pi/\gamma$. We then have

$$E[g(\boldsymbol{\theta})] = \frac{\int_{\boldsymbol{\Theta}} g(\boldsymbol{\theta})\rho(\boldsymbol{\theta})\exp\{-nh_n(\boldsymbol{\theta})\}\,d\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}} \rho(\boldsymbol{\theta})\exp\{-nh_n(\boldsymbol{\theta})\}\,d\boldsymbol{\theta}}. \tag{A.1}$$

Then

$$E[g(\boldsymbol{\theta})] = g(\hat{\boldsymbol{\theta}}) + \frac{1}{n}\sum_{ij}\frac{\partial}{\partial\theta_i}g(\hat{\boldsymbol{\theta}})h^{ij}\left\{\frac{\rho_j(\hat{\boldsymbol{\theta}})}{\rho(\hat{\boldsymbol{\theta}})} - \frac{1}{2}\sum_{rs}h^{rs}h_{rsj}(\hat{\boldsymbol{\theta}})\right\} + \frac{1}{2n}\sum_{ij}h^{ij}g_{ij} + O(n^{-2}),$$

where $\rho_j(\hat{\boldsymbol{\theta}}) = \partial\rho(\hat{\boldsymbol{\theta}})/\partial\theta_j$. Note that $\sum_{ij} h^{ij}g_{ij}$ can be reexpressed as $\mathrm{tr}(D^2g(\hat{\boldsymbol{\theta}})[D^2h(\hat{\boldsymbol{\theta}})]^{-1})$. The foregoing approximation was used and discussed by Mosteller and Wallace (1964). The special case where $\gamma = 1$ yields $\rho = \pi$ while $\hat{\boldsymbol{\theta}}$ becomes the MLE, and the choice $\gamma = \pi$ yields $\rho = 1$ with $\hat{\boldsymbol{\theta}}$ becoming the posterior mode. The latter was used by Lindley (1980).

APPENDIX B: LEMMAS ABOUT MATRICES

Lemma B.1. Suppose that $\mathbf{a}$ is a $1 \times d$ matrix and that $\mathbf{M}$ is a $d \times d$ symmetric matrix. Then

$$\mathbf{a}^T(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\mathbf{M}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = (\boldsymbol{\theta} - \mathbf{y})^T\mathbf{M}(\boldsymbol{\theta} - \mathbf{y}) - \frac{1}{4}\mathbf{a}^T\mathbf{M}^{-1}\mathbf{a},$$

where $\mathbf{y} = \hat{\boldsymbol{\theta}} - (1/2)\mathbf{M}^{-1}\mathbf{a}$.

Proof. The proof is omitted because it is easily verified.

Lemma B.2. Suppose that $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^T$ is distributed according to $\mathrm{N}_d(\hat{\boldsymbol{\theta}} + \mathbf{b}, \boldsymbol{\Omega})$, where $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_d)^T$, $\mathbf{b} = (b_1, \dots, b_d)^T$, and the covariance matrix $\boldsymbol{\Omega} = (\omega_{ij})$. Let the expectation in $\mathrm{N}_d(\hat{\boldsymbol{\theta}} + \mathbf{b}, \boldsymbol{\Omega})$ denote $E^N[\cdot]$. Then the following equations hold:

(a) $$E^N[(\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j)(\theta_k - \hat{\theta}_k)] = b_i\omega_{jk} + b_j\omega_{ik} + b_k\omega_{ij} + b_ib_jb_k,$$

(b) $$E^N[(\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j)(\theta_k - \hat{\theta}_k)(\theta_q - \hat{\theta}_q)] = \omega_{ij}\omega_{kq} + \omega_{ik}\omega_{jq} + \omega_{iq}\omega_{kj} + \sum_{[m_1,m_2]} b_{m1}b_{m2}\omega_{m3m4} + b_ib_jb_kb_q,$$

(c) $$E^N\big[(\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j)(\theta_k - \hat{\theta}_k)(\theta_q - \hat{\theta}_q)(\theta_r - \hat{\theta}_r)(\theta_s - \hat{\theta}_s)\big] = \mu_{ijkqrs} + \sum_{[m_1,m_2]} b_{m1}b_{m2}\mu_{m1m2m3m4} + \sum_{[l_1,l_2,l_3,l_4]} b_{l1}b_{l2}b_{l3}b_{l4}\omega_{l5l6} + b_ib_jb_kb_qb_rb_s,$$

where $\mu_{m1m2m3m4}$ and μ_{ijkqrs} are the fourth and sixth central moments of a multivariate normal distribution with covariance matrix $\boldsymbol{\Omega}$. The foregoing notation $[m_1, m_2]$ indicates a tuple chosen among i, j, k, and q. Hence, the summation $\sum_{[m_1,m_2]}$ is over ${}_4C_2$ terms. Similarly, $[l_1, l_2, l_3, l_4]$ denotes a quadruple chosen among i, j, k, q, r, and s. Hence, the summation $\sum_{[l_1,l_2,l_3,l_4]}$ is over ${}_6C_4$ terms.

Proof. The proofs are omitted because it is easy to check these equations.

Lemma B.3. Let $\mathbf{I}_d$ be a $d \times d$ unit matrix. For any $d \times d$ matrix $\mathbf{A}$ and any scalar x,

$$|\mathbf{A} + x\mathbf{I}_d| = \sum_{r=0}^{d} x^r \sum_{\{i_1,\dots,i_r\}} \left|\mathbf{A}^{\{i_1,\dots,i_r\}}\right|, \tag{B.1}$$

where $|\cdot|$ denotes the determinant of a matrix, $\{i_1, \dots, i_r\}$ is an r-dimensional subset of the first d positive integers $1, \dots, d$ (and the second summation is over all ${}_dC_r$ such subsets), and $\mathbf{A}^{\{i_1,\dots,i_r\}}$ is the $(d-r) \times (d-r)$ principal submatrix of $\mathbf{A}$ obtained by striking out the $i_1, \dots, i_r$th rows and columns. (For $r = d$, the sum $\sum_{\{i_1,\dots,i_r\}} |\mathbf{A}^{\{i_1,\dots,i_r\}}|$ is to be interpreted as 1.) Note that the expression (B.1) is a polynomial in x, and that the coefficient of x^0 (i.e., the constant term of the polynomial) equals $|\mathbf{A}|$, and the coefficient of x^{d-1} equals $\mathrm{tr}(\mathbf{A})$.

Proof. See Harville (1997; corollary 13.7.4, p.197).

Lemma B.4. For $d \times d$ nonsingular matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\frac{d}{dt}\left(\frac{|\mathbf{A} - (t/n)\mathbf{B}|}{|\mathbf{A}|}\right)\bigg|_{t=0} = -\frac{1}{n}\,\mathrm{tr}(\mathbf{B}\mathbf{A}^{-1}).$$

Proof. Using Lemma B.3 and the relation between an inverse matrix and the cofactors, we have

$$\begin{aligned}
&\frac{d}{dt}\left(\frac{|\mathbf{A} - (t/n)\mathbf{B}|}{|\mathbf{A}|}\right)\bigg|_{t=0} \\
&= \frac{d}{dt}\left(|\mathbf{B}\mathbf{A}^{-1}|\left|\mathbf{A}\mathbf{B}^{-1} - \frac{t}{n}\mathbf{I}_d\right|\right)\bigg|_{t=0} \\
&= \frac{d}{dt}\left(|\mathbf{B}\mathbf{A}^{-1}|\left[|\mathbf{A}\mathbf{B}^{-1}| - \frac{t}{n}\sum_{\{i_1\}}|(\mathbf{A}\mathbf{B}^{-1})^{\{i_1\}}| + \frac{t^2}{n^2}\xi + \cdots + \frac{(-t)^d}{n^d}\right]\right)\bigg|_{t=0} \\
&= -\frac{1}{n}\frac{1}{|\mathbf{A}\mathbf{B}^{-1}|}\sum_{\{i_1\}}|(\mathbf{A}\mathbf{B}^{-1})^{\{i_1\}}| \\
&= -\frac{1}{n}\,\mathrm{tr}(\mathbf{A}\mathbf{B}^{-1})^{-1} \\
&= -\frac{1}{n}\,\mathrm{tr}(\mathbf{B}\mathbf{A}^{-1}),
\end{aligned}$$

where $\xi = \sum_{\{i_1,i_2\}} |(\mathbf{A}\mathbf{B}^{-1})^{\{i_1,i_2\}}|$ and $\{i_1\}$ and $\{i_1, i_2\}$ are the same notations as in Lemma B.3. Hence, this lemma is proved.

Lemma B.5. For $d \times d$ nonsingular matrices $\mathbf{A}$ and $\mathbf{B}$ and any scalar x, it follows that

$$(\mathbf{A} + x\mathbf{B})^{-1} - \mathbf{A}^{-1} = O(x). \tag{B.2}$$

Proof. Multiplying $(\mathbf{A} + x\mathbf{B})^{-1}(\mathbf{A} + x\mathbf{B}) = \mathbf{I}_d$ by $\mathbf{A}^{-1}$ from the right side yields

$$(\mathbf{A} + x\mathbf{B})^{-1} - \mathbf{A}^{-1} = -x(\mathbf{A} + x\mathbf{B})^{-1}\mathbf{B}\mathbf{A}^{-1}. \tag{B.3}$$

Hence the lemma is proved.

Lemma B.6. For $d \times d$ nonsingular matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\frac{d}{dt}\left(\mathbf{A} - \frac{t}{n}\mathbf{B}\right)^{-1}\bigg|_{t=0} = \frac{1}{n}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}.$$

Proof. Using (B.3),

$$\frac{(\mathbf{A} - (t/n)\mathbf{B})^{-1} - \mathbf{A}^{-1}}{t} = \frac{1}{n}\left(\mathbf{A} - \frac{t}{n}\mathbf{B}\right)^{-1}\mathbf{B}\mathbf{A}^{-1}. \tag{B.4}$$

Hence, letting $t \to 0$ yields the lemma.

APPENDIX C: PROOFS OF THE MAIN RESULTS

Here we prove some results stated in the article. Throughout this appendix, for simplicity, let $h_n = h$ and $h_n^* = h^*$.

Proof of Theorem 1

We let $\hat{\boldsymbol\theta} = (\hat\theta_1, \dots, \hat\theta_d)^T$ and $z_i = \theta_i - \hat\theta_i$. First, we approximate the integrand of $\int_{B_\delta(\hat{\boldsymbol\theta})} e^{-nh(\boldsymbol\theta)}\, d\boldsymbol\theta$. By expanding $h(\boldsymbol\theta)$ about $\hat{\boldsymbol\theta}$, it follows that

$$\begin{aligned}
h(\boldsymbol\theta) &= h(\hat{\boldsymbol\theta}) + \sum_i h_i(\hat{\boldsymbol\theta}) z_i + \frac{1}{2}\sum_{ij} h_{ij}(\hat{\boldsymbol\theta}) z_i z_j + \frac{1}{6}\sum_{ijk} h_{ijk}(\hat{\boldsymbol\theta}) z_i z_j z_k \\
&\quad + \frac{1}{24}\sum_{ijkq} h_{ijkq}(\hat{\boldsymbol\theta}) z_i z_j z_k z_q \\
&\quad + \frac{1}{120}\sum_{ijkqr} h_{ijkqr}(\hat{\boldsymbol\theta}) z_i z_j z_k z_q z_r + r_{1n}, \\
&= h(\hat{\boldsymbol\theta}) + \sum_i h_i(\hat{\boldsymbol\theta}) z_i + \frac{1}{2}\sum_{ij} h_{ij}(\hat{\boldsymbol\theta}) z_i z_j + K(\boldsymbol\theta, \hat{\boldsymbol\theta}) + R_{1n},
\end{aligned}$$

where r_{1n} is bounded over $B_\epsilon(\hat{\boldsymbol\theta})$ by a polynomial in $z_i z_j z_k z_q z_r z_s$, $K(\boldsymbol\theta, \hat{\boldsymbol\theta}) = (1/6)\sum h_{ijk}(\hat{\boldsymbol\theta}) z_i z_j z_k + (1/24)\sum h_{ijkq}(\hat{\boldsymbol\theta}) z_i z_j z_k z_q$, and R_{1n} is the sum of $(1/120)\sum h_{ijkqr}(\hat{\boldsymbol\theta}) z_i z_j z_k z_q z_r$ and r_{1n}. Using Lemma B.1, we have

$$\begin{aligned}
&\sum_i h_i(\hat{\boldsymbol\theta}) z_i + \frac{1}{2}\sum_{ij} h_{ij}(\hat{\boldsymbol\theta}) z_i z_j \\
&\quad = D^1 h(\hat{\boldsymbol\theta})^T (\boldsymbol\theta - \hat{\boldsymbol\theta}) + \frac{1}{2}(\boldsymbol\theta - \hat{\boldsymbol\theta})^T D^2 h(\hat{\boldsymbol\theta})(\boldsymbol\theta - \hat{\boldsymbol\theta}) \\
&\quad = \frac{1}{2}(\boldsymbol\theta - \mathbf{y})^T D^2 h(\hat{\boldsymbol\theta})(\boldsymbol\theta - \mathbf{y}) \\
&\qquad - \frac{1}{4} D^1 h(\hat{\boldsymbol\theta})^T \left[\frac{1}{2} D^2 h(\hat{\boldsymbol\theta})\right]^{-1} D^1 h(\hat{\boldsymbol\theta}),
\end{aligned}$$

where $\mathbf{y} = \hat{\boldsymbol\theta} - [D^2 h(\hat{\boldsymbol\theta})]^{-1} D^1 h(\hat{\boldsymbol\theta})$.

In contrast, using $e^x = 1 + x + x^2/2! + x^3/3! + e^{\tau_1} x^4/4!$, where τ_1 is a point between 0 and x, the expansion of $\exp(-n(K(\boldsymbol\theta, \hat{\boldsymbol\theta}) + R_{1n}))$ is given by

$$\begin{aligned}
&\exp\left\{-n\left[\frac{1}{6}\sum_{ijk} h_{ijk}(\hat{\boldsymbol\theta}) z_i z_j z_k + \frac{1}{24}\sum_{ijkq} h_{ijkq}(\hat{\boldsymbol\theta}) z_i z_j z_k z_q + R_{1n}\right]\right\} \\
&\quad = 1 - n\left(\frac{1}{6}\sum_{ijk} h_{ijk}(\hat{\boldsymbol\theta}) z_i z_j z_k + \frac{1}{24}\sum_{ijkq} h_{ijkq}(\hat{\boldsymbol\theta}) z_i z_j z_k z_q + R_{1n}\right) \\
&\qquad + \frac{n^2}{2}\left\{\left(\frac{1}{6}\sum_{ijk} h_{ijk}(\hat{\boldsymbol\theta}) z_i z_j z_k\right)^2 + R_{2n}\right\} \\
&\qquad - \frac{n^3}{6}\big(K(\boldsymbol\theta, \hat{\boldsymbol\theta}) + R_{1n}\big)^3 + R_{3n} \\
&\quad = J_n(\boldsymbol\theta, \hat{\boldsymbol\theta}) + R_n,
\end{aligned} \tag{C.1}$$

where R_{2n} is the rest of the terms squared in the bracket, R_{3n} is a term with polynomials of degree greater than 12, R_n is the sum of all terms involving R_{1n}, R_{2n}, R_{3n}, and $K(\boldsymbol\theta, \hat{\boldsymbol\theta})^3$, and where $J_n(\boldsymbol\theta, \hat{\boldsymbol\theta}) = 1 - (n/6)\sum h_{ijk}(\hat{\boldsymbol\theta}) z_i z_j z_k - (n/24)\sum h_{ijkq}(\hat{\boldsymbol\theta}) z_i z_j z_k z_q + (n^2/72) \times \sum h_{ijk}(\hat{\boldsymbol\theta}) h_{qrs}(\hat{\boldsymbol\theta}) z_i z_j z_k z_q z_r z_s$.

Hence, combining the foregoing results, we have

$$\begin{aligned}
&\int_{B_\delta(\hat{\boldsymbol\theta})} e^{-nh(\boldsymbol\theta)}\, d\boldsymbol\theta \\
&= e^{-nh(\hat{\boldsymbol\theta})} \cdot \exp\left\{\frac{n}{4} D^1 h(\hat{\boldsymbol\theta})^T \left[\frac{1}{2} D^2 h(\hat{\boldsymbol\theta})\right]^{-1} D^1 h(\hat{\boldsymbol\theta})\right\} \\
&\quad \times \int_{B_\delta(\hat{\boldsymbol\theta})} \exp\left(-\frac{1}{2}(\boldsymbol\theta - \mathbf{y})^T \left(\frac{1}{n} D^2 h(\hat{\boldsymbol\theta})^{-1}\right)^{-1} (\boldsymbol\theta - \mathbf{y})\right) \\
&\qquad \times \exp\left\{-n\left(\frac{1}{6}\sum_{ijk} h_{ijk}(\hat{\boldsymbol\theta}) z_i z_j z_k \right.\right. \\
&\qquad\qquad \left.\left. + \frac{1}{24}\sum_{ijkq} h_{ijkq}(\hat{\boldsymbol\theta}) z_i z_j z_k z_q + R_{1n}\right)\right\} d\boldsymbol\theta \\
&= e^{-nh(\hat{\boldsymbol\theta})} C_n(\hat{\boldsymbol\theta}) \\
&\quad \times \int_{B_\delta(\hat{\boldsymbol\theta})} \exp\left(-\frac{1}{2}(\boldsymbol\theta - \mathbf{y})^T \boldsymbol\Sigma^{-1} (\boldsymbol\theta - \mathbf{y})\right) J_n(\boldsymbol\theta, \hat{\boldsymbol\theta})\, d\boldsymbol\theta \qquad \text{(C.2)} \\
&\quad + e^{-nh(\hat{\boldsymbol\theta})} C_n(\hat{\boldsymbol\theta}) \\
&\quad \times \int_{B_\delta(\hat{\boldsymbol\theta})} \exp\left(-\frac{1}{2}(\boldsymbol\theta - \mathbf{y})^T \boldsymbol\Sigma^{-1} (\boldsymbol\theta - \mathbf{y})\right) R_n\, d\boldsymbol\theta, \qquad \text{(C.3)}
\end{aligned}$$

where $C_n(\hat{\boldsymbol\theta}) = \exp\{(n/2) D^1 h(\hat{\boldsymbol\theta})^T [D^2 h(\hat{\boldsymbol\theta})]^{-1} D^1 h(\hat{\boldsymbol\theta})\}$ and $\boldsymbol\Sigma = n^{-1}[D^2 h(\hat{\boldsymbol\theta})]^{-1}$.

Second, we evaluate (C.3). The terms composing R_n may be represented explicitly using the mean value form of the remainders in terms of higher derivatives of h evaluated at points between $\hat{\boldsymbol\theta}$ and $\boldsymbol\theta$; for example, one such term is $(-n/720)\sum h_{ijkqrs}(\boldsymbol\gamma_1) z_i z_j z_k z_q z_r z_s$, where $\boldsymbol\gamma_1$ is a point between $\boldsymbol\theta$ and $\hat{\boldsymbol\theta}$. It is one piece of the error term appearing as R_n. Because it follows from (A3) that $\|h_{ijkqrs}(\boldsymbol\gamma_1)\| < M$ on $B_\delta(\hat{\boldsymbol\theta})$, we have

$$\begin{aligned}
&\left\| \frac{-n}{720} \int_{B_\delta(\hat{\boldsymbol\theta})} \exp\left(-\frac{1}{2}(\boldsymbol\theta - \mathbf{y})^T \boldsymbol\Sigma^{-1} (\boldsymbol\theta - \mathbf{y})\right) \right. \\
&\qquad \left. \times h_{ijkqrs}(\boldsymbol\gamma_1) z_i z_j z_k z_q z_r z_s\, d\boldsymbol\theta \right\| \\
&\quad \le \frac{nM}{720} \int_{B_\delta(\hat{\boldsymbol\theta})} \exp\left(-\frac{1}{2}(\boldsymbol\theta - \mathbf{y})^T \boldsymbol\Sigma^{-1} (\boldsymbol\theta - \mathbf{y})\right) \|z_i z_j z_k z_q z_r z_s\|\, d\boldsymbol\theta \\
&\quad = |\boldsymbol\Sigma|^{1/2} \times O(n^{-2}).
\end{aligned}$$

The other terms are similar. Thus (C.3) becomes $(2\pi)^{d/2} e^{-nh(\hat{\boldsymbol\theta})} \times |\boldsymbol\Sigma|^{1/2} C_n(\hat{\boldsymbol\theta}) \times O(n^{-2})$.

Third, we evaluate the integral (C.2). Because there exists a symmetric matrix $\mathbf{A}^{1/2}$ such that $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = D^2 h(\hat{\boldsymbol\theta})$, putting $n^{1/2}\mathbf{A}^{1/2}(\boldsymbol\theta - \mathbf{y}) = \mathbf{u}$, we have

$$\begin{aligned}
&\boldsymbol\Theta - B_\delta(\hat{\boldsymbol\theta}) \\
&\quad = \{\boldsymbol\theta : (\mathbf{u} - n^{1/2}\mathbf{A}^{1/2}\mathbf{b})^T [D^2 h(\hat{\boldsymbol\theta})]^{-1} (\mathbf{u} - n^{1/2}\mathbf{A}^{1/2}\mathbf{b}) > n\delta^2\} \\
&\quad \subseteq \left\{\boldsymbol\theta : \frac{1}{\lambda_1}(\mathbf{u} - n^{1/2}\mathbf{A}^{1/2}\mathbf{b})^T (\mathbf{u} - n^{1/2}\mathbf{A}^{1/2}\mathbf{b}) > n\delta^2\right\} \\
&\quad \subseteq \{\boldsymbol\theta : 2\mathbf{u}^T\mathbf{u} + 2n\mathbf{b}^T\mathbf{A}^{1/2}\mathbf{A}^{1/2}\mathbf{b} > n\lambda_1\delta^2\} \\
&\quad = \{\boldsymbol\theta : \mathbf{u}^T\mathbf{u} > nc_2\} \\
&\quad = \{\boldsymbol\theta : (\boldsymbol\theta - \mathbf{y})^T \boldsymbol\Sigma^{-1} (\boldsymbol\theta - \mathbf{y}) > nc_2\},
\end{aligned}$$

where λ_1 is the smallest eigenvalue of $D^2 h(\hat{\boldsymbol\theta})$, $\mathbf{b} = -[D^2 h(\hat{\boldsymbol\theta})]^{-1} \times D^1 h(\hat{\boldsymbol\theta})$, and $c_2 = \lambda_1\delta^2/2 - D^1 h(\hat{\boldsymbol\theta})^T [D^2 h(\hat{\boldsymbol\theta})]^{-1} D^1 h(\hat{\boldsymbol\theta})$. Note that $c_2 > 0$ for large n, because it holds from the assumption (A4) that

$\lambda_1 > 0$. Hence, once again putting $n^{1/2}\mathbf{A}^{1/2}(\boldsymbol{\theta}-\mathbf{y}) = \mathbf{u} = (u_i)$, we have

$$\int_{\boldsymbol{\Theta}-B_\delta(\hat{\boldsymbol{\theta}})} \exp\left(-\frac{1}{2}(\boldsymbol{\theta}-\mathbf{y})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\mathbf{y})\right) J_n(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})\,d\boldsymbol{\theta} \tag{C.4}$$
$$\le \int_{(\boldsymbol{\theta}-\mathbf{y})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\mathbf{y})>nc_2} \exp\left(-\frac{1}{2}(\boldsymbol{\theta}-\mathbf{y})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\mathbf{y})\right) \times J_n(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})\,d\boldsymbol{\theta}$$
$$= |\boldsymbol{\Sigma}|^{1/2}\int_{\mathbf{u}^T\mathbf{u}>nc_2} \exp\left(-\frac{1}{2}\mathbf{u}^T\mathbf{u}\right)\mathrm{Po}(\mathbf{u})\,d\mathbf{u},$$

where $\mathrm{Po}(\mathbf{u})$ denotes a term with multivariate polynomials in $u_1,\ldots,u_d$ of finite degree. As a result, (C.4) becomes a term of exponentially decreasing order by the same argument as given by Kass et al. (1990, p. 478). This result allows replacement of the domain in (C.2) by the whole Euclidean space. We denote the expectation in $\mathrm{N}(\hat{\boldsymbol{\theta}}+\mathbf{b}, n^{-1}[D^2h(\hat{\boldsymbol{\theta}})]^{-1})$ by $E^N[\cdot]$. Then, using Lemma B.2, we have

$$nE^N[(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j)(\theta_k-\hat{\theta}_k)] = b_ih^{jk}+b_jh^{ik}+b_kh^{ij}+nb_ib_jb_k. \tag{C.5}$$

Hence from (C.5) and $D^1h(\hat{\boldsymbol{\theta}}) = O(n^{-1})$, it holds that

$$-\frac{n}{6}E^N\left[\sum_{ijk}h_{ijk}(\hat{\boldsymbol{\theta}})z_iz_jz_k\right] = -\frac{1}{2n}\sum_{ijk}h_{ijk}(\hat{\boldsymbol{\theta}})(nb_i)h^{jk}+O(n^{-2}). \tag{C.6}$$

Similarly, applying Lemma B.2 to the expanded terms $\sum h_{ijkq}(\hat{\boldsymbol{\theta}})\times z_iz_jz_kz_q$ and $[(1/6)\sum h_{ijk}(\hat{\boldsymbol{\theta}})z_iz_jz_k]^2$ in $J_n(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})$, we have

$$\int_{B_\delta(\hat{\boldsymbol{\theta}})} e^{-nh(\boldsymbol{\theta})}\,d\boldsymbol{\theta} = (2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}e^{-nh(\hat{\boldsymbol{\theta}})}C_n(\hat{\boldsymbol{\theta}})\left(1+\frac{a_n}{n}+O(n^{-2})\right), \tag{C.7}$$

where $C_n(\hat{\boldsymbol{\theta}}) = \exp\{(n/2)D^1h(\hat{\boldsymbol{\theta}})^T[D^2h(\hat{\boldsymbol{\theta}})]^{-1}D^1h(\hat{\boldsymbol{\theta}})\}$,

$$a_n = -\frac{1}{2}\sum_{ijk}h_{ijk}(\hat{\boldsymbol{\theta}})(nb_i)h^{jk} - \frac{1}{8}\sum_{ijkq}h_{ijkq}(\hat{\boldsymbol{\theta}})h^{ij}h^{kq} + \frac{1}{72}\sum_{ijkqrs}h_{ijk}(\hat{\boldsymbol{\theta}})h_{qrs}(\hat{\boldsymbol{\theta}})\mu_{ijkqrs}n^3$$

and μ_{ijkqrs} are the sixth central moments of a multivariate normal distribution with covariance matrix $\boldsymbol{\Sigma} = [D^2h(\hat{\boldsymbol{\theta}})]^{-1}$. In contrast, it follows from the assumption (A5) of Section 2 that

$$\int_{\boldsymbol{\Theta}-B_\delta(\hat{\boldsymbol{\theta}})} e^{-nh(\boldsymbol{\theta})}\,d\boldsymbol{\theta} = (2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}e^{-nh(\hat{\boldsymbol{\theta}})}C_n(\hat{\boldsymbol{\theta}})\times O(n^{-2}).$$

Therefore, by combining (C.7) and the foregoing result, the theorem is proved.

Remark C.1. We explain why the Taylor expansion with the six terms and the remainder was applied to h in the proof of Theorem 1. If we use the expansion with five terms and the remainder, then a term $nh_{ijkqr}(\boldsymbol{\gamma}_2)z_iz_jz_kz_qz_r$ appears in R_n. Arguing as in the evaluation of (C.3), we have

$$\left\|n\int_{B_\delta(\hat{\boldsymbol{\theta}})}\exp\left(-\frac{1}{2}(\boldsymbol{\theta}-\mathbf{y})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\mathbf{y})\right)h_{ijkqr}(\boldsymbol{\gamma}_2)z_iz_jz_kz_qz_r\,d\boldsymbol{\theta}\right\| = |\boldsymbol{\Sigma}|^{1/2}\times O(n^{-3/2}).$$

Therefore, the term may not be of order $|\boldsymbol{\Sigma}|^{1/2}\times O(n^{-2})$.

We now give an expansion for (2) that is used in the proofs of Theorems 2 and 3. Let $\tilde{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}$ be asymptotic modes of order n^{-1} for $-h_n^*$ and $-h_n$, let, $\tilde{\mathbf{b}} = (\tilde{b}_i) = -[D^2h^*(\tilde{\boldsymbol{\theta}})]^{-1}D^1h^*(\tilde{\boldsymbol{\theta}})$, let $\tilde{h}^{*ij}$ denote the components of $[D^2h^*(\tilde{\boldsymbol{\theta}})]^{-1}$, and let $\tilde{\boldsymbol{\Sigma}}^* = n^{-1}[D^2h^*(\tilde{\boldsymbol{\theta}})]^{-1}$. Assume that $h^*(\tilde{\boldsymbol{\theta}}) - h(\hat{\boldsymbol{\theta}}) = O(n^{-1})$ and $h^*_{i1\cdots is}(\tilde{\boldsymbol{\theta}}) - h_{i1\cdots is}(\hat{\boldsymbol{\theta}}) = O(n^{-1})$ for $s = 2,3,4$. Using Theorem 1 and the same argument as given by Tierney and Kadane (1986, p. 86), we have

$$\frac{\int_{\boldsymbol{\Theta}} e^{-nh^*(\boldsymbol{\theta})}\,d\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}} e^{-nh(\boldsymbol{\theta})}\,d\boldsymbol{\theta}} = \left(\frac{|\tilde{\boldsymbol{\Sigma}}^*|}{|\boldsymbol{\Sigma}|}\right)^{1/2} \times \frac{\exp\{-nh^*(\tilde{\boldsymbol{\theta}})\}}{\exp\{-nh(\hat{\boldsymbol{\theta}})\}}\frac{C_n^*(\tilde{\boldsymbol{\theta}})}{C_n(\hat{\boldsymbol{\theta}})}\left(1+\frac{\tilde{a}_n^*-a_n}{n}+O(n^{-2})\right). \tag{C.8}$$

First, we evaluate the term $(\tilde{a}_n^*-a_n)/n$. Then

$$\frac{\tilde{a}_n^*-a_n}{n} = -\frac{1}{2n}\sum_{ijk}\left[h^*_{ijk}(\tilde{\boldsymbol{\theta}})(n\tilde{b}_i^*)\tilde{h}^{*jk} - h_{ijk}(\hat{\boldsymbol{\theta}})(nb_i)h^{jk}\right] \tag{C.9}$$
$$-\frac{1}{8n}\sum_{ijkq}\left[h^*_{ijkq}(\tilde{\boldsymbol{\theta}})\tilde{h}^{*ij}\tilde{h}^{*kq} - h_{ijkq}(\hat{\boldsymbol{\theta}})h^{ij}h^{kq}\right] \tag{C.10}$$
$$+\frac{1}{72n}\sum_{ijkqrs}\left[h^*_{ijk}(\tilde{\boldsymbol{\theta}})h^*_{qrs}(\tilde{\boldsymbol{\theta}})\mu_{ijkqrs}n^3 - h_{ijk}(\hat{\boldsymbol{\theta}})h_{qrs}(\hat{\boldsymbol{\theta}})\mu_{ijkqrs}n^3\right]. \tag{C.11}$$

In the foregoing terms, it is verified from the same argument as that of Tierney and Kadane (1986, p. 86) that (C.10) and (C.11) are $O(n^{-2})$. Here the evaluation of (C.9) is the key to the proof of Theorems 2 and 3.

Proof of Theorem 2

Let h^{*ij} be the components of $[D^2h^*(\hat{\boldsymbol{\theta}})]^{-1}$, $\mathbf{b}^* = (b_i^*) = -[D^2h^*(\hat{\boldsymbol{\theta}})]^{-1}D^1h^*(\hat{\boldsymbol{\theta}})$. Let a_n^* denote the term with h_{ijk}, b_i, and h^{ij} replaced by h^*_{ijk}, b_i^*, and h^{*ij} in a_n of Theorem 1. Setting $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}$ in (C.8), the notations $\tilde{\mathbf{b}}$, $\tilde{h}^{*ij}$, and $\tilde{a}_n^*$ can be replaced by $\mathbf{b}^*$, h^{*ij}, and a_n^*. (C.10) and (C.11) are of order $O(n^{-2})$ because $\hat{\boldsymbol{\theta}}$ is an asymptotic mode of order n^{-1} for both $-h$ and $-h^*$. Because $h^*(\hat{\boldsymbol{\theta}}) = h(\hat{\boldsymbol{\theta}}) - (1/n)\log g(\hat{\boldsymbol{\theta}})$, it follows from (B.3) that $h^{*ij} - h^{ij} = O(n^{-1})$. Hence

$$n(\mathbf{b}^*-\mathbf{b}) = n\left(-[D^2h^*(\hat{\boldsymbol{\theta}})]^{-1}+[D^2h(\hat{\boldsymbol{\theta}})]^{-1}\right)D^1h(\hat{\boldsymbol{\theta}}) + [D^2h^*(\hat{\boldsymbol{\theta}})]^{-1}D^1\log g(\hat{\boldsymbol{\theta}})$$
$$= [D^2h(\hat{\boldsymbol{\theta}})]^{-1}D^1\log g(\hat{\boldsymbol{\theta}}) + O(n^{-1}). \tag{C.12}$$

From (C.12), (C.9) is given by

$$-\frac{1}{2n}\sum_{ijkq}h_{ijk}(\hat{\boldsymbol{\theta}})h^{iq}h^{jk}\frac{\partial}{\partial\theta_q}\log g(\hat{\boldsymbol{\theta}})+O(n^{-2}).$$

Moreover, it is obvious that $\exp\{-nh^*(\hat{\boldsymbol{\theta}})\}/\exp\{-nh(\hat{\boldsymbol{\theta}})\} = g(\hat{\boldsymbol{\theta}})$ and $|\boldsymbol{\Sigma}^*|/|\boldsymbol{\Sigma}| = |D^2h(\hat{\boldsymbol{\theta}})|/|D^2h^*(\hat{\boldsymbol{\theta}})|$. Therefore, Theorem 2 is proved.

Proof of Theorem 3

Let $\hat{\boldsymbol{\theta}}$ be the exact mode of $-h$, $\hat{\boldsymbol{\theta}}_N = \hat{\boldsymbol{\theta}} - [D^2h^*(\hat{\boldsymbol{\theta}})]^{-1}D^1h^*(\hat{\boldsymbol{\theta}})$ and $\mathbf{b}_N^* = -[D^2h^*(\hat{\boldsymbol{\theta}}_N)]^{-1}D^1h^*(\hat{\boldsymbol{\theta}}_N)$. We denote the components of $[D^2h^*(\hat{\boldsymbol{\theta}}_N)]^{-1}$ by h_N^{*ij}. In the expansion (C.8), we replace $\tilde{\boldsymbol{\theta}}$, $\tilde{\mathbf{b}}$, and $\tilde{h}^{*ij}$ by $\hat{\boldsymbol{\theta}}_N$, $\mathbf{b}_N^* = (b_{Ni}^*)$, and h_N^{*ij}. First, we prove that $D^1h^*(\hat{\boldsymbol{\theta}}_N) = O(n^{-2})$. We can expand $\partial h^*(\hat{\boldsymbol{\theta}}_N)/\partial\theta_i$ around the exact mode $\hat{\boldsymbol{\theta}}$ of $-h$, giving for $i = 1, \ldots, d$,

$$\frac{\partial}{\partial\theta_i}h^*(\hat{\boldsymbol{\theta}}_N) = \frac{\partial}{\partial\theta_i}h^*(\hat{\boldsymbol{\theta}}) + \left(D^1\frac{\partial}{\partial\theta_i}h^*(\hat{\boldsymbol{\theta}})\right)^T(\hat{\boldsymbol{\theta}}_N - \hat{\boldsymbol{\theta}}) + (\hat{\boldsymbol{\theta}}_N - \hat{\boldsymbol{\theta}})^T\left[D^2\frac{\partial}{\partial\theta_i}h^*(\boldsymbol{\xi}_i)\right](\hat{\boldsymbol{\theta}}_N - \hat{\boldsymbol{\theta}}), \quad \text{(C.13)}$$

where $\boldsymbol{\xi}_i$ is an interior point on the line from $\hat{\boldsymbol{\theta}}$ to $\hat{\boldsymbol{\theta}}_N$.

Writing (C.13) in a matrix form gives

$$D^1h^*(\hat{\boldsymbol{\theta}}_N) = D^1h^*(\hat{\boldsymbol{\theta}}) + D^2h^*(\hat{\boldsymbol{\theta}})^T(\hat{\boldsymbol{\theta}}_N - \hat{\boldsymbol{\theta}}) + R^*(\boldsymbol{\xi}), \quad \text{(C.14)}$$

where $R^*(\boldsymbol{\xi}) = ((\hat{\boldsymbol{\theta}}_N - \hat{\boldsymbol{\theta}})^T[D^2\partial h^*(\boldsymbol{\xi}_1)/\partial\theta_1](\hat{\boldsymbol{\theta}}_N - \hat{\boldsymbol{\theta}}), \ldots, (\hat{\boldsymbol{\theta}}_N - \hat{\boldsymbol{\theta}})[D^2\partial h^*(\boldsymbol{\xi}_d)/\partial\theta_d](\hat{\boldsymbol{\theta}}_N - \hat{\boldsymbol{\theta}}))^T$.

Substituting $\hat{\boldsymbol{\theta}}_N - \hat{\boldsymbol{\theta}} = -[D^2h^*(\hat{\boldsymbol{\theta}})]^{-1}D^1h^*(\hat{\boldsymbol{\theta}})$ into (C.14), we have

$$D^1h^*(\hat{\boldsymbol{\theta}}_N) = D^1h^*(\hat{\boldsymbol{\theta}}) - D^2h^*(\hat{\boldsymbol{\theta}})[D^2h^*(\hat{\boldsymbol{\theta}})]^{-1}D^1h^*(\hat{\boldsymbol{\theta}}) + O(n^{-2}) = O(n^{-2}).$$

Next, we evaluate the term (C.9). Because it follows from $D^1h^*(\hat{\boldsymbol{\theta}}_N) = O(n^{-2})$ that $\mathbf{b}_N^* = O(n^{-2})$, the left-side term of (C.9) is of order n^{-2}, namely

$$-\frac{1}{2n}\sum_{ijk}h_{ijk}^*(\hat{\boldsymbol{\theta}}_N)(nb_{Ni}^*)h_N^{*jk} = O(n^{-2}).$$

In addition, because it follows from $D^1h(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ that $\mathbf{b} = \mathbf{0}$, the right-side term of (C.9) is 0, namely $(1/2n)\sum_{ijk}h_{ijk}(\hat{\boldsymbol{\theta}})(nb_i)h^{jk} = 0$. Hence, (C.9) is of order $O(n^{-2})$. Therefore, Theorem 3 is proved.

Proof of Theorem 4

Expanding $D^1h^*(\hat{\boldsymbol{\theta}}_N)$ around $\hat{\boldsymbol{\theta}}$ and arguing as in the proof of Theorem 3, we have $D^1h^*(\hat{\boldsymbol{\theta}}_N) = O(n^{-2})$. The rest of the proof is essentially the same as that of Theorem 3.

Proof of Proposition 2

First, we show that $\int_{\boldsymbol{\Theta}-B_\delta(\hat{\boldsymbol{\theta}})}e^{-nh(\boldsymbol{\theta})}\,d\boldsymbol{\theta} = O(e^{-nc_1})$. When π is proper, by assumption (A5′),

$$\int_{\boldsymbol{\Theta}-B_\delta(\hat{\boldsymbol{\theta}})}e^{-nh(\boldsymbol{\theta})}\,d\boldsymbol{\theta} = e^{-n\bar{h}(\hat{\boldsymbol{\theta}})}\int_{\boldsymbol{\Theta}-B_\delta(\hat{\boldsymbol{\theta}})}\pi(\boldsymbol{\theta})e^{-n[\bar{h}(\boldsymbol{\theta})-\bar{h}(\hat{\boldsymbol{\theta}})]}\,d\boldsymbol{\theta}. \le e^{-n\bar{h}(\hat{\boldsymbol{\theta}})}e^{-nc_1}\int_{\boldsymbol{\Theta}-B_\delta(\hat{\boldsymbol{\theta}})}\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta} = O(e^{-nc_1}).$$

When $\pi(\boldsymbol{\theta})$ is improper, it is also proved by applying the preceding argument with the prior π replaced by the posterior density based on $\mathbf{x}^{(n_0)}$ (see Kass et al. 1990). Similarly, $\int_{\boldsymbol{\Theta}-B_\delta(\hat{\boldsymbol{\theta}})}g(\boldsymbol{\theta})e^{-nh(\boldsymbol{\theta})}\,d\boldsymbol{\theta}$ can be proved.

A Heuristic Argument About Approximate Characteristic Functions

To prove Proposition 3, we rewrite the error terms of the approximate MGF (5) in another form and give an approximate characteristic function. Let $\tilde{\Theta}$ be a one-dimensional random variable with the posterior density $p(\theta|x)$, let $\hat{\theta}$ be the exact mode of $-h$, and let $M(t) = E[\exp\{t\theta\}]$ denote the MGF of $\tilde{\Theta}$. Recall the result of Proposition 1,

$$M(t) = \exp\left\{t\hat{\theta} + \frac{t^2}{2nh''(\hat{\theta})}\right\}\left(1 - \frac{h'''(\hat{\theta})}{2nh''(\hat{\theta})^2}t + O^t(n^{-2})\right), \quad \text{(C.15)}$$

where $O^t(\cdot)$ indicates that the error term may depend on t. The variable t in $b^* = -h^{*\prime\prime}(\hat{\theta})^{-1}h^{*\prime}(\hat{\theta}) = t/(nh''(\hat{\theta}))$ is the key to Proposition 3. Hence we write $b^* = b^*[t]$. By $b^*[t] = O(t/n)$, the error term can be written as

$$O^t\left(\frac{1}{n^2}\right) = O\left(\frac{a_{20}}{n^2} + \frac{a_{21}b^*[t]}{n} + a_{22}b^*[t]^2\right) + O^t(n^{-3}) = O\left(\sum_{j=0}^{2}a_{2j}\left(\frac{1}{n}\right)^{2-j}b^*[t]^j\right) + O^t(n^{-3}), \quad \text{(C.16)}$$

where a_{2j} $(j = 0, 1, 2)$ are of order $O(1)$ and do not depend on t. Similarly, we have

$$O^t\left(\frac{1}{n^2}\right) = O\left(\sum_{j=0}^{2}\frac{a_{2j}}{n^{2-j}}b^*[t]^j + \frac{a_{30}}{n^3} + \frac{a_{31}b^*[t]}{n^2} + \frac{a_{32}b^*[t]^2}{n} + a_{33}b^*[t]^3\right) + O^t(n^{-4}).$$

Repeating this treatment inductively, $O^t(n^{-2})$ is expressed as

$$O^t\left(\frac{1}{n^2}\right) = O\left(\sum_{k=2}^{N_1}\sum_{j=0}^{k}\frac{a_{kj}}{n^{k-j}}b^*[t]^j\right), \quad \text{(C.17)}$$

where a_{kj} $(j = 0, 1, \ldots, k)$ are of order $O(1)$ and do not depend on t and N_1 is a finite number because R_n is bounded above by a term with polynomials of finite degree in (C.3). Hence (C.15) is reexpressed as

$$M(t) = \exp\left\{t\hat{\theta} + \frac{t^2}{2nh''(\hat{\theta})}\right\} \times \left(1 - \frac{h'''(\hat{\theta})}{2nh''(\hat{\theta})^2}t + O\left(\sum_{k=2}^{N_1}\sum_{j=0}^{k}\frac{a_{kj}}{n^{k-j}}b^*[t]^j\right)\right). \quad \text{(C.18)}$$

Therefore, by substituting $t = is$ into (C.18) formally, the approximate characteristic function is given by

$$\Psi_\theta(s) = \exp\left\{is\hat{\theta} - \frac{s^2}{2nh''(\hat{\theta})}\right\} \times \left(1 - \frac{h'''(\hat{\theta})}{2nh''(\hat{\theta})^2}is + O\left(\sum_{k=2}^{N_1}\sum_{j=0}^{k}\frac{a_{kj}}{n^{k-j}}b^*[is]^j\right)\right). \quad \text{(C.19)}$$

Next we present two lemmas that are required in the proof of Proposition 3.

Lemma C.1 (Hermite polynomials). It follows that for $k = 0, 1, 2, \ldots,$

$$\frac{d^k}{d\alpha_0^k}n(\alpha_0|0,1) = (-1)^kH_k(\alpha_0)n(\alpha_0|0,1).$$

$H_0(\alpha_0), H_1(\alpha_0), \ldots$ are called Hermite polynomial. The first few are $H_0(\alpha_0) = 1$, $H_1(\alpha_0) = \alpha_0$, $H_2(\alpha_0) = \alpha_0^2 - 1$, and $H_3(\alpha_0) = \alpha_0^3 - 3\alpha_0$.

Proof. The proof is omitted because Hermite polynomials are well known. (For the detailed proof, see Petrov 1975, pp. 136–139.)

Lemma C.2. Let $\phi(s)$ be the characteristic function of a standard normal distribution N(0, 1). Then it follows that for $k = 1, 2, \ldots,$

(a) $$\int_{-\infty}^{\alpha_0}\int_{-\infty}^{\infty}\frac{1}{2\pi}(-is)^k e^{-is\theta}\phi(s)\,ds\,d\theta = (-1)^{k-1}H_{k-1}(\alpha_0)n(\alpha_0|0,1)$$

and

(b) $$\int_{-\infty}^{\alpha_0}\int_{-\infty}^{\infty}\frac{1}{2\pi}b^*\big[-i\sqrt{nh''(\hat\theta)}s\big]^k e^{-is\theta}\phi(s)\,ds\,d\theta = O\big(n^{-k/2}H_{k-1}(\alpha_0)n(\alpha_0|0,1)\big),$$

where $b^*[t] = t/(nh''(\hat\theta))$.

Proof. First, we prove (a). Using Lemma C.1, we have

$$\begin{aligned}
&\int_{-\infty}^{\alpha_0}\int_{-\infty}^{\infty}\frac{1}{2\pi}(-is)^k e^{-is\theta}\phi(s)\,ds\,d\theta \\
&\quad= \int_{-\infty}^{\alpha_0}\int_{-\infty}^{\infty}\frac{1}{2\pi}\frac{d^k}{d\theta^k}e^{-is\theta}\phi(s)\,ds\,d\theta \\
&\quad= \int_{-\infty}^{\alpha_0}\frac{d^k}{d\theta^k}n(\theta|0,1)\,d\theta \\
&\quad= \frac{d^{k-1}}{d\theta^{k-1}}n(\theta|0,1)\Big|_{\theta=\alpha_0} \\
&\quad= (-1)^{k-1}H_{k-1}(\alpha_0)n(\alpha_0|0,1).
\end{aligned}$$

By using the result (a), (b) is easily verified.

Proof of Proposition 3

Suppose that $\tilde\Theta$ is a random variable according to the posterior density $p(\theta|x)$ and the characteristic function of $\tilde\Theta$ is $\Psi_\theta(s) = E[\exp\{is\theta\}]$. Let $\Psi_\eta(s)$ be a characteristic function of the standardized random variable:

$$\eta = \frac{\tilde\Theta-\hat\theta}{\sqrt{1/(nh''(\hat\theta))}} = \sqrt{nh''(\hat\theta)}(\tilde\Theta-\hat\theta).$$

Then, using the property of characteristic functions, we have the following relation

$$\Psi_\eta(s) = \exp\big\{-i\sqrt{nh''(\hat\theta)}\hat\theta s\big\}\Psi_\theta\big(\sqrt{nh''(\hat\theta)}s\big). \tag{C.20}$$

Recall the expansion (C.19) of the characteristic function $\Psi_\theta(s)$. Then, replacing the variable s by $(nh''(\hat\theta))^{1/2}s$ in (C.19) and substituting the replaced one into (C.20), we get

$$\begin{aligned}
\Psi_\eta(s) &= \exp\left\{-\frac{s^2}{2}\right\} \\
&\quad\times\left(1+\frac{h'''(\hat\theta)\sqrt{h''(\hat\theta)}}{2nh''(\hat\theta)^2}(-is)\right. \\
&\qquad\left.+\,O\left(\sum_{k=2}^{N_1}\sum_{j=0}^{k}\frac{a_{kj}}{n^{k-j}}b^*\big[i(nh''(\hat\theta))^{1/2}s\big]^j\right)\right).
\end{aligned} \tag{C.21}$$

To obtain the approximate distribution function of η, we apply the inversion formula to (C.21). Using Lemma C.2, we have

$$\begin{aligned}
&P\left(\frac{\tilde\Theta-\hat\theta}{\sqrt{1/(nh''(\hat\theta))}}\le\alpha_0\right) \\
&\quad= \int_{-\infty}^{\alpha_0}\left(\int_{-\infty}^{\infty}\frac{1}{2\pi}e^{-is\eta}\Psi_\eta(s)\,ds\right)d\eta \\
&\quad= \int_{-\infty}^{\alpha_0}\int_{-\infty}^{\infty}\frac{1}{2\pi}e^{-is\eta}\phi(s) \\
&\qquad\times\left(1+\frac{h'''(\hat\theta)\sqrt{h''(\hat\theta)}}{2nh''(\hat\theta)^2}(-is)\right. \\
&\qquad\left.+\,O\left(\sum_{k=2}^{N_1}\sum_{j=0}^{k}\frac{a_{kj}}{n^{k-j}}b^*\big[i(nh''(\hat\theta))^{1/2}s\big]^j\right)\right)ds\,d\eta \\
&\quad= \Phi(\alpha_0)+n^{-1/2}\frac{h'''(\hat\theta)\sqrt{h''(\hat\theta)}}{2h''(\hat\theta)^2}n(\alpha_0|0,1) \\
&\qquad+\,O\big(n^{-1}\alpha_0 n(\alpha_0|0,1)\big).
\end{aligned}$$

Therefore, Proposition 3 is proved.

Proof of Theorem 5

We first differentiate (12). Then, using Lemma B.4, we have

$$\begin{aligned}
&\frac{d}{dt}\left(\frac{|D^2h_n(\hat{\boldsymbol\theta})|}{|D^2h_n^*(\hat{\boldsymbol\theta})|}\right)^{1/2}\Bigg|_{t=0} \\
&\quad= \frac{d}{dt}\left(\frac{|D^2h(\hat{\boldsymbol\theta})-(t/n)D^2g(\hat{\boldsymbol\theta})|}{|D^2h(\hat{\boldsymbol\theta})|}\right)^{-1/2}\Bigg|_{t=0} \\
&\quad= \frac{1}{2n}\operatorname{tr}\big(D^2g(\hat{\boldsymbol\theta})[D^2h(\hat{\boldsymbol\theta})]^{-1}\big).
\end{aligned} \tag{C.22}$$

Subsequently, we evaluate $(d/dt)C_n^*(\hat{\boldsymbol\theta})|_{t=0}$. It follows from Lemma B.6 that $(d/dt)[D^2h^*(\hat{\boldsymbol\theta})]^{-1}|_{t=0} = n^{-1}[D^2h(\hat{\boldsymbol\theta})]^{-1}\times D^2g(\hat{\boldsymbol\theta})[D^2h(\hat{\boldsymbol\theta})]^{-1}$. Using the result, we obtain

$$\begin{aligned}
&\frac{d}{dt}\big(D^1h^*(\hat{\boldsymbol\theta})^T[D^2h^*(\hat{\boldsymbol\theta})]^{-1}D^1h^*(\hat{\boldsymbol\theta})\big)\Big|_{t=0} \\
&\quad= \frac{d}{dt}\left(D^1h(\hat{\boldsymbol\theta})^T[D^2h^*(\hat{\boldsymbol\theta})]^{-1}D^1h(\hat{\boldsymbol\theta})\right. \\
&\qquad-\frac{2t}{n}D^1g(\hat{\boldsymbol\theta})^T[D^2h^*(\hat{\boldsymbol\theta})]^{-1}D^1h(\hat{\boldsymbol\theta}) \\
&\qquad\left.+\frac{t^2}{n^2}D^1g(\hat{\boldsymbol\theta})^T[D^2h^*(\hat{\boldsymbol\theta})]^{-1}D^1g(\hat{\boldsymbol\theta})\right)\Bigg|_{t=0} \\
&\quad= \frac{1}{n}D^1h(\hat{\boldsymbol\theta})^T[D^2h(\hat{\boldsymbol\theta})]^{-1}D^2g(\hat{\boldsymbol\theta})[D^2h(\hat{\boldsymbol\theta})]^{-1}D^1h(\hat{\boldsymbol\theta}) \\
&\qquad-\frac{2}{n}D^1g(\hat{\boldsymbol\theta})^T[D^2h(\hat{\boldsymbol\theta})]^{-1}D^1h(\hat{\boldsymbol\theta}).
\end{aligned} \tag{C.23}$$

Therefore,

$$\begin{aligned}
\frac{d}{dt}C_n^*(\hat{\boldsymbol\theta})\Big|_{t=0} &= C_n(\hat{\boldsymbol\theta})\Big(-D^1g(\hat{\boldsymbol\theta})^T[D^2h(\hat{\boldsymbol\theta})]^{-1}D^1h(\hat{\boldsymbol\theta}) \\
&\qquad+\frac{1}{2}\mathbf{b}^TD^2g(\hat{\boldsymbol\theta})\mathbf{b}\Big),
\end{aligned} \tag{C.24}$$

where $\mathbf{b} = [D^2h(\hat{\boldsymbol\theta})]^{-1}D^1h(\hat{\boldsymbol\theta})$. Combining (C.22), (C.23), and (C.24) yields

$$\begin{aligned}
\frac{d}{dt}\hat M(t)\Big|_{t=0} &= g(\hat{\boldsymbol\theta})+\frac{1}{2n}\operatorname{tr}\big(D^2g(\hat{\boldsymbol\theta})[D^2h(\hat{\boldsymbol\theta})]^{-1}\big) \\
&\quad-\frac{1}{2n}\sum_{ijkq}h_{ijk}(\hat{\boldsymbol\theta})h^{iq}h^{jk}\frac{\partial}{\partial\theta_q}g(\hat{\boldsymbol\theta}) \\
&\quad-D^1g(\hat{\boldsymbol\theta})^T[D^2h(\hat{\boldsymbol\theta})]^{-1}D^1h(\hat{\boldsymbol\theta})+\frac{1}{2}\mathbf{b}^TD^2g(\hat{\boldsymbol\theta})\mathbf{b}.
\end{aligned}$$

Hence, Theorem 5 is proved.

[Received January 2003. Revised March 2004.]

REFERENCES

David, H. A., and Nagaraja, H. N. (2003), *Order Statistics*, New York: Wiley-Interscience.

Harville, D. A. (1997), *Matrix Algebra From a Statistician's Perspective*, New York: Springer-Verlag.

Heyde, C. C., and Johnstone, I. M. (1979), "On Asymptotic Posterior Normality for Stochastic Processes," *Journal of the Royal Statistical Society*, Ser. B, 41, 184–189.

Johnson, R. A. (1967), "An Asymptotic Expansion for Posterior Distributions," *The Annals of Mathematical Statistics*, 38, 1266–1272.

——— (1970), "Asymptotic Expansions Associated With Posterior Distributions," *The Annals of Mathematical Statistics*, 41, 851–864.

Kass, R. E., Tierney, L., and Kadane, J. B. (1990), "The Validity of Posterior Expansions Based on Laplace's Method," in *Essays in Honor of George Barnard*, eds. S. Geisser, J. S. Hodges, S. J. Press, and A. Zellner, Amsterdam: North-Holland, pp. 473–488.

Laplace, P. S. (1847), *Oeuvres*, Vol. 7, Paris: Imprimerie Royale.

Lejeune, M., and Faulkenberry, G. D. (1982), "A Simple Predictive Density Function," *Journal of the American Statistical Association*, 77, 654–657.

Leonard, T. (1982), Comment on "A Simple Predictive Density Function," by M. Lejeune and G. D. Faulkenberry, *Journal of the American Statistical Association*, 77, 657–658.

Lewis, S. M., and Raftery, A. E. (1997), "Estimating Bayes Factors via Posterior Simulation With the Laplace-Metropolis Estimator," *Journal of the American Statistical Association*, 92, 648–655.

Lindley, D. V. (1961), "The Use of Prior Probability Distributions in Statistical Inference and Decisions," in *Proceedings of the Fourth Berkeley Symposium*, Vol. 1, Berkeley, CA: University of California Press, pp. 453–468.

——— (1980), "Approximate Bayesian Methods," in *Bayesian Statistics*, eds. J. M. Bernardo, M. H. Degroot, D. V. Lindley, and A. F. M. Smith, Valencia, Spain: University Press, pp. 223–245.

Mosteller, F., and Wallace, D. L. (1964), *Inference and Disputed Authorship: The Federalist Papers*, Reading, MA: Addison-Wesley.

Petrov, V. V. (1975). *Sums of Independent Random Variables*, Berlin: Akademie-Verlag, Berlin.

Tierney, L., and Kadane, J. B. (1986), "Accurate Approximations for Posterior Moments and Marginal Densities," *Journal of the American Statistical Association*, 81, 82–86.

Tierney, L., Kass, R. E., and Kadane, J. B. (1989), "Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions," *Journal of the American Statistical Association*, 84, 710–716.