Accurate Approximations for Posterior Moments and Marginal Densities

LUKE TIERNEY and JOSEPH B. KADANE*

This article describes approximations to the posterior means and variances of positive functions of a real or vector-valued parameter, and to the marginal posterior densities of arbitrary (i.e., not necessarily positive) parameters. These approximations can also be used to compute approximate predictive densities. To apply the proposed method, one only needs to be able to maximize slightly modified likelihood functions and to evaluate the observed information at the maxima. Nevertheless, the resulting approximations are generally as accurate and in some cases more accurate than approximations based on third-order expansions of the likelihood and requiring the evaluation of third derivatives. The approximate marginal posterior densities behave very much like saddle-point approximations for sampling distributions. The principal regularity condition required is that the likelihood times prior be unimodal.

KEY WORDS: Bayesian inference; Laplace method; Asymptotic expansions; Computation of integrals.

1. INTRODUCTION

A user of Bayesian methods in practice needs to be able to evaluate various characteristics of posterior and predictive distributions, especially their densities, means, and variances. If the problem under consideration does not involve a conjugate prior-likelihood pair, these tasks cannot be performed in closed form; analytical or numerical approximation methods are needed. In these cases it would often be useful to have approximations that are more accurate than the usual normal approximation, yet not as computationally intensive as numerical integration methods (Naylor and Smith 1983) or Monte Carlo methods (Kloek and Van Dijk 1978; Zellner and Rossi 1982). Lindley (1980) has proposed approximations for moments that capture the first-order error terms of the normal approximation. This is generally accurate enough, but, as Lindley points out, the required evaluation of third derivatives of the posterior can be rather tedious—in particular, in problems with several parameters. Mosteller and Wallace (1964, sec. 4.6C) suggested a similar approach but introduced a transformation of the parameters to avoid the need for the direct use of third derivatives. The proposed transformation, however, depends on the second derivatives of the log-likelihood. A numerical maximization routine for locating the posterior mode of the transformed parameters will therefore require third derivatives of the log-likelihood unless a more complicated derivative-free algorithm is used.

In this article we introduce an easily computable approximation for the posterior mean and variance of a nonnegative parameter or, more generally, of a smooth function of the parameter that is nonzero on the interior of the parameter space. For definiteness we will take this function to be positive. In this introduction we give a brief, informal description of the basic approximation and its properties. Section 2 contains a more careful statement.

Let g be a smooth, positive function on the parameter space. The posterior mean of $g(\theta)$ can be written as

$$E_n[g] = E[g(\theta) \mid X^{(n)}] = \frac{\int g(\theta)e^{\ell(\theta)}\pi(\theta)d\theta}{\int e^{\ell(\theta)}\pi(\theta)d\theta}, \quad (1.1)$$

where ℓ is the log-likelihood function and π is the prior density. It is common practice to approximate the denominator integral in (1.1) by an integral of an approximating normal curve centered at the posterior mode and having variance equal to minus the inverse of the second derivative of the log posterior density at its mode. This approximation, used by Lindley and by Mosteller and Wallace, can be viewed as an application of the Laplace method for integrals, as described in De Bruijn (1961). It will produce reasonable results as long as the posterior is unimodal or at least dominated by a single mode; we will assume that the sample size is large enough for this to be the case. The new feature in the approximation proposed in the present article is in its approach to the numerator integral in (1.1). Rather than expand the integrand of this integral about the posterior mode, we propose to locate the mode of the numerator integrand itself, find the second derivative at this new mode, and then approximate this integral by a second application of Laplace's method.

The computational requirements of this approach are rather minimal: We only need to be able to evaluate first and second derivatives and maximize the two integrands, which can be viewed as slightly modified likelihood functions. These computations will be feasible for any problems in which maximum likelihood estimates and the observed information can be computed. Nevertheless, the resulting approximations are quite accurate. An intuitive explanation is this: If the function g is bounded away from zero near the posterior mode, then the two integrands will be similar in shape. Thus by applying the same approximation technique, Laplace's method, to the numerator and the denominator we will be making similar errors, and in taking the ratio some portion of these errors will cancel. In asymptotic terms, Laplace's method typically has an error of order $O(n^{-1})$. Used in this ratio form, the error is of order $O(n^{-2})$—the order $O(n^{-1})$ error terms in the numerator and the denominator cancel.

* Luke Tierney is Associate Professor, School of Statistics, University of Minnesota, Minneapolis, MN 55455. Joseph B. Kadane is Professor, Departments of Statistics and Social Sciences, Carnegie-Mellon University, Pittsburgh, PA 15213. Kadane's research was sponsored in part by Office of Naval Research Contract N00014-82-K-0622. The research of both authors was sponsored in part by National Science Foundation Grant DMS-8503019. The authors thank Rob Kass for many helpful conversations, for suggesting the argument leading to (4.3), and for bringing the results of Leonard and of Mosteller and Wallace to their attention.

In the next section we give a more formal statement of the proposed approximation for the posterior mean of a positive function g and an approximation for its variance. Section 3 briefly discusses the approximation of predictive densities. The use of Laplace's method for approximating marginal posterior densities is discussed in Section 4, and an application to data from the Stanford heart-transplant experiment is given in Section 5. Section 6 presents some closing remarks, and the Appendix contains some technical details on the asymptotic error rates.

In the derivations given in this article we have chosen to sacrifice formal rigor for clarity of exposition. Explicit statements of regularity conditions have been omitted. A technical report including some of the omitted details is available from us.

2. POSTERIOR MEANS AND VARIANCES

Setting $L = \log \pi + \mathfrak{L}/n$ and $L^* = \log g + \log \pi + \mathfrak{L}/n$, expression (1.1) for the posterior mean of g can be written as

$$E_n[g] = \int e^{nL^*}d\theta \Big/ \int e^{nL}d\theta. \tag{2.1}$$

Let $\hat{\theta}$ be the posterior mode, the mode of L, and let $\sigma^2 = -1/L''(\hat{\theta})$. Then Laplace's method produces the approximation

$$\begin{aligned}\int e^{nL(\theta)}d\theta &\approx \int \exp[nL(\hat{\theta}) - n(\theta - \hat{\theta})^2/(2\sigma^2)]d\theta \\ &= \sqrt{2\pi}\sigma n^{-1/2}e^{nL(\hat{\theta})}\end{aligned} \tag{2.2}$$

for the denominator integral in (2.1). Similarly, if $\hat{\theta}^*$ is the mode of L^* and $\sigma^{*2} = -1/L^{*\prime\prime}(\hat{\theta}^*)$, then the Laplace approximation to the numerator integral in (2.1) is $\sqrt{2\pi}\sigma^* n^{-1/2}\exp\{nL^*(\hat{\theta}^*)\}$. Taking the ratio of these two approximations produces the approximation

$$\hat{E}_n[g] = (\sigma^*/\sigma)\exp\{n(L^*(\hat{\theta}^*) - L(\hat{\theta}))\} \tag{2.3}$$

to $E_n[g]$. We will refer to approximation (2.3) as the Laplace approximation.

A similar approximation applies in the multiparameter case: Set

$$\hat{E}_n[g] = \left(\frac{\det \Sigma^*}{\det \Sigma}\right)^{1/2} \exp\{n(L^*(\hat{\theta}^*) - L(\hat{\theta}))\},$$

where $\hat{\theta}^*$ and $\hat{\theta}$ maximize L^* and L, respectively, and Σ^* and Σ are minus the inverse Hessians of L^* and L at $\hat{\theta}^*$ and $\hat{\theta}$, respectively.

The asymptotic accuracy of the Laplace approximation (2.3) is investigated in the Appendix. To summarize the results: The errors in the approximations to the two integrals in (2.1) are of order $O(n^{-1})$. The leading terms in the two errors are identical, however, and cancel when the ratio is taken. As a result, the error in (2.3) is of order $O(n^{-2})$; that is,

$$E_n[g] = \hat{E}_n[g](1 + O(n^{-2})). \tag{2.4}$$

Having obtained an approximation for the posterior mean of g, we would next like to approximate the posterior variance. The simplest way to obtain such an approximation is to set

$$\hat{V}_n[g] = \hat{E}_n[g^2] - \hat{E}_n[g]^2. \tag{2.5}$$

That is, we use (2.3) to approximate the posterior means of g and g^2, and then we insert these approximations into the standard computational formula for the variance.

By (2.4) we might expect that (2.5) will have an absolute error of order $O(n^{-2})$ or a relative error of order $O(n^{-1})$ (since $V_n[g] \sim \sigma/n$). Another cancellation occurs, however: From the expression (A.2) given in the Appendix for the error of (2.3), we can see that the leading error terms in $\hat{E}_n[g^2]$ as an approximation to $E_n[g^2]$ and in $\hat{E}_n[g]^2$ as an approximation to $E_n[g]^2$ are identical. As a result, they cancel in computing (2.5). So $\hat{V}_n[g]$ has an absolute error of order $O(n^{-3})$ or a relative error of order $O(n^{-2})$; that is,

$$\hat{V}_n[g] = V_n[g](1 + O(n^{-2})). \tag{2.6}$$

A similar argument shows that if we approximate the posterior covariance $C_n[g, h]$ of two positive functions $g(\theta)$ and $h(\theta)$ by

$$\hat{C}_n[g, h] = \hat{E}_n[gh] - \hat{E}_n[g]\hat{E}_n[h], \tag{2.7}$$

then

$$C_n[g, h] = \hat{C}_n[g, h] + O(n^{-3}) \tag{2.8}$$

as well.

As a practical point, it is worth mentioning that approximation (2.5) should be used with caution if n is very large, since it involves the computation of a small number as the difference between two large numbers. If computations are done with sufficient precision, however, then for most practical sample sizes this will not cause any problems. At the other end of the spectrum, if n is very small, then it is possible for the variance approximation (2.5) to be negative and for a covariance matrix computed from (2.7) not to be positive semidefinite. This should be checked in any application, but in most cases it does not seem to be a problem even for moderate sample sizes. More work is needed to see if modified variance approximations can be obtained that are guaranteed to be positive. A second point concerns the maximizations required for (2.3), (2.5), and (2.7): Once $\hat{\theta}$, the posterior mode, has been determined, it can be used as a starting value for a numerical search for $\hat{\theta}^*$, the maximum of L^*. Generally, the number of iterations needed to find $\hat{\theta}^*$ from $\hat{\theta}$ will be quite small. In fact, since the asymptotic statement of (2.4) only depends on the leading asymptotic term in $\hat{\theta}^* - \hat{\theta}$, it remains valid if we replace $\hat{\theta}^*$ by a single Newton step from $\hat{\theta}$ towards $\hat{\theta}^*$. For (2.6) and (2.8) to remain valid, two Newton steps are sufficient.

In concluding this section, a comment on the assumed positivity of g is appropriate. This assumption is needed to insure that the numerator and denominator integrands in (2.1) are similar in shape. This similarity in shape, in turn, is responsible for the cancellation of error terms in the approximation to the ratio (2.1). Thus for the approximation to be accurate for a function g taking both positive and negative values, the posterior distribution of g must be concentrated almost entirely to one side of the origin. If this is not the case, then this approach is not directly applicable. Work on extensions that will retain the computational simplicity of (2.3) and (2.5) but are applicable, for example, for computing posterior moments of regression parameters is currently in progress.

3. PREDICTIVE DISTRIBUTIONS

Consider approximating the predictive density

$$f_n(z) = f(z \mid x^{(n)}) = E_n[f(z \mid \theta)]$$

at specific values of z. Direct application of (2.3) produces the approximation

$$\hat{f}_n(z) = \hat{E}_n[f(z \mid \theta)],$$

which, by (2.4), has an error of order $O(n^{-2})$. This is similar to the approximation considered by Leonard (1982) in his comment on the paper by Lejeune and Faulkenberry (1982), who considered predictive densities from the frequentist point of view.

With its error of order $O(n^{-2})$, the approximation $\hat{f}_n(z)$ has a lower-order error than Dunsmore's (1976) modification of the simple approximation $f(z \mid \hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimator (MLE). Dunsmore's approximation includes some but not all terms of order $O(n^{-1})$. In particular, he drops the integral of the second term in the integrand of his equation (1); Lindley's (1980) equation (7) shows that this integral is generally of order $O(n^{-1})$ but no smaller.

4. MARGINAL POSTERIOR DENSITIES

Laplace's method can also be used to approximate marginal posterior densities of individual parameters in multiparameter settings. The resulting approximation, first suggested by Leonard (1982), provides a useful alternative to generally more time-consuming numerical or Monte Carlo integration techniques. In its use of Laplace's method to integrate out one or more variables from a multivariate function to obtain a density, this approach is similar to the saddle-point method introduced by Daniels (1954) and studied further, for example, in Barndorff-Nielsen and Cox (1979) and Daniels (1980).

To obtain the approximation, set $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p) = (\theta_1, \boldsymbol{\theta}_2)$; that is, partition the p vector $\boldsymbol{\theta}$ into its first component and the $(p - 1)$ vector of the remaining components. Suppose $\hat{\theta} = (\hat{\theta}_1, \hat{\boldsymbol{\theta}}_2)$ maximizes πe^{ℓ} (i.e., $\hat{\boldsymbol{\theta}}$ is the posterior mode), and let $\hat{\Sigma}$ be minus the inverse of the Hessian of $(\ell + \log \pi)/n$ at $\hat{\boldsymbol{\theta}}$; thus $\hat{\Sigma}$ is a $p \times p$ matrix. For a given θ_1, let the $(p - 1)$ vector $\hat{\boldsymbol{\theta}}_2^* = \hat{\boldsymbol{\theta}}_2^*(\theta_1)$ maximize the function $h(\cdot) = \pi(\theta_1, \cdot)e^{\ell(\theta_1, \cdot)}$, the function πe^{ℓ} with θ_1 held fixed, and let $\hat{\Sigma}^* = \hat{\Sigma}^*(\theta_1)$ be minus the inverse of the Hessian of $(\log h(\cdot))/n$, a $(p - 1) \times (p - 1)$ matrix. Applying Laplace's method to the integrals in the numerator and denominator of the expression

$$\pi_1(\theta_1 \mid x^{(n)}) = \pi_{n,1}(\theta_1) = \frac{\int \pi(\theta_1, \boldsymbol{\theta}_2)e^{\ell(\theta_1, \boldsymbol{\theta}_2)}d\boldsymbol{\theta}_2}{\int \pi(\boldsymbol{\theta})e^{\ell(\boldsymbol{\theta})}d\boldsymbol{\theta}},$$

for the marginal posterior density of θ_1 we obtain the approximation

$$\hat{\pi}_{n,1}(\theta_1) = \left(\frac{\det \hat{\Sigma}^*(\theta_1)}{2\pi n \det \hat{\Sigma}}\right)^{1/2} \frac{\pi(\theta_1, \hat{\boldsymbol{\theta}}_2^*)e^{\ell(\theta_1, \hat{\boldsymbol{\theta}}_2^*)}}{\pi(\hat{\boldsymbol{\theta}})e^{\ell(\hat{\boldsymbol{\theta}})}}. \quad (4.1)$$

Using calculations similar to those presented in the Appendix, it is possible to show that

$$\pi_{n,1}(\theta_1) = \hat{\pi}_{n,1}(\theta_1)(1 + O_{\theta_1}(n^{-1})), \quad (4.2)$$

where $O_{\theta_1}(n^{-1})$ is of order $O(n^{-1})$ but depends on $\boldsymbol{\theta}_1$. If $\boldsymbol{\theta}_0$ is the true parameter vector, then under regularity conditions similar to those given in Walker (1969), it is possible to show that the error term $O_{\theta_1}(n^{-1})$ will be uniformly of order $O(n^{-1})$ for θ_1 in some fixed neighborhood of $\theta_{0,1}$, the first component of $\boldsymbol{\theta}_0$. By contrast, the absolute error of the usual normal approximation is only of order $O(n^{-1/2})$. Moreover, the relative error of the normal approximation, or any Edgeworth-type approximation for that matter, usually only tends to zero on neighborhoods that shrink toward $\theta_{0,1}$ at rate $n^{-1/2}$.

The main reason that the error in (4.2) is as large as $O(n^{-1})$ is that the dimensionalities of the two integrals in the numerator and the denominator of $\pi_{n,1}$ are different. In fact, most of this error is due to the constant of integration; the error in the approximation of the functional form of $\pi_{n,1}(\theta_1)$ is only of order $O(n^{-3/2})$ in $n^{-1/2}$ neighborhoods of $\hat{\theta}_1$. To see this, fix u, let $\theta_1 = \hat{\theta}_1 + n^{-1/2}u$, and note that the error $O_{\theta_1}(n^{-1})$ is usually a smooth function of θ_1. Thus we can write

$$O_{\theta_1}(n^{-1}) = O_{\hat{\theta}_1}(n^{-1}) + (u/n^{-1/2})O_u(n^{-1})$$

and

$$\pi_{n,1}(\theta_1) = \hat{\pi}_{n,1}(\theta)(1 + O_{\hat{\theta}_1}(n^{-1}))(1 + O_u(n^{-3/2})). \quad (4.3)$$

Since the term $O_{\hat{\theta}_1}(n^{-1})$ does not depend on u, it is an error in the constant of integration, and the error in the approximation $\hat{\pi}_{n,1}$ to the functional form of $\pi_{n,1}$ is of order $O(n^{-3/2})$, as claimed.

A similar result was pointed out by Daniels (1956) for the saddle-point approximation. As in that case, the main implication is that an approximate marginal posterior calculated using (4.1) should be renormalized by numerical integration to integrate to one.

To appreciate just how accurately (4.1) can capture the functional form of $\pi_{n,1}(\theta_1)$, consider, for example, the normal-gamma conjugate distribution for normal data with unknown mean and precision. Thus the joint posterior for the mean m and the precision r is of the form

$$\pi(m, r) \propto r^{\alpha - 1/2}e^{-r(\beta + \tau(m-\mu)^2/2)}$$

for some α, β, μ, and τ. Leonard (1982) pointed out that (4.1) is remarkably accurate in this case. In fact, a simple calculation shows that (4.1) produces the correct functional forms!

Another joint distribution for which (4.1) produces the exact functional forms of the marginals is the Dirichlet distribution. It would be interesting to obtain a characterization of all joint distributions for which this occurs. A similar phenomenon occurs for the saddle-point approximation; that is, there are certain distributions for which that approximation produces the exact functional forms. For the saddle-point approximation, Daniels (1980) has obtained a characterization of all cases in which the approximation produces exact results. It may be possible to modify Daniels's approach to characterize the joint distributions for which (4.1) produces exact functional forms.

5. AN APPLICATION

As an example we consider a three-parameter model used by Turnbull, Brown, and Hu (1974) to describe data from the

Stanford heart-transplant program and referred to by them as the Pareto model. This model, described in section 4.3 of their paper, views individual patients in the nontransplant group as having exponential lifetimes with mean ϕ, where ϕ is itself a random variable drawn independently for each patient from a gamma distribution with density proportional to $\phi^{p-1}e^{-\lambda\phi}$. Patients in the transplant group have a similar distribution but with $\tau\phi$ in place of ϕ for the residual lifetime after the transplant. The resulting likelihood function of the three parameters τ, λ, p is

$$\prod_{i=1}^{n} \frac{p\lambda^p}{(\lambda + x_i)^{p+1}} \prod_{i=n+1}^{N} \left(\frac{\lambda}{\lambda + x_i}\right)^p \times \prod_{j=1}^{m} \frac{\tau p\lambda^p}{(\lambda + y_j + \tau z_j)^{p+1}} \prod_{j=m+1}^{M} \left(\frac{\lambda}{\lambda + y_j + \tau z_j}\right)^p,$$

where the x_i are the survival times in days of the $N = 30$ nontransplant patients, $n = 26$ of whom died, and y_j, z_j are the times to transplant and survival times after transplant, respectively, of the $M = 52$ transplant patients, $m = 34$ of whom died.

Naylor and Smith (1982) used this model with an improper uniform prior on the parameters τ, λ, p to illustrate their computational approach based on Gauss–Hermite quadrature, and we use the same improper prior distribution for the present illustration. Naylor and Smith point out the possibility of integrating out the parameter p analytically, but, following their example, we have chosen not to do this and to apply the Laplace approximations directly.

Table 1 lists the posterior means and standard deviations computed by the Laplace approximation method and by Naylor and Smith using Gauss–Hermite integration applied to an orthogonalized reparameterization. As can be seen from this table, the largest relative difference between any of the Laplace approximations and the results of Naylor and Smith is about 4%.

In approximating the marginal densities, we selected a set of 60 equally spaced points for each parameter and then at each point computed approximations to the marginal densities by formula (4.1). A simple rectangular integration of these approximate densities produced integrals of approximately 1.2 in all three cases; thus renormalization was necessary. We then obtained plots of spline interpolations of the renormalized densities. We used 60 points for increased accuracy; however, 30 points produced identical pictures. In performing the maximizations for the individual grid points, we proceeded outward from the MLE's, using each current set of optimal values as the starting values for the next maximization.

As a basis for comparison, for each of the 60 grid points selected for a given parameter we orthogonalized the other two

Table 1. Posterior Means and Standard Deviations for the Pareto Model

	Posterior Means			Posterior Standard Deviations		
Method	τ	λ	p	τ	λ	p
Laplace	1.044	32.11	.4926	.4944	16.09	.1381
Naylor and Smith	1.04	32.5	.50	.47	16.2	.14

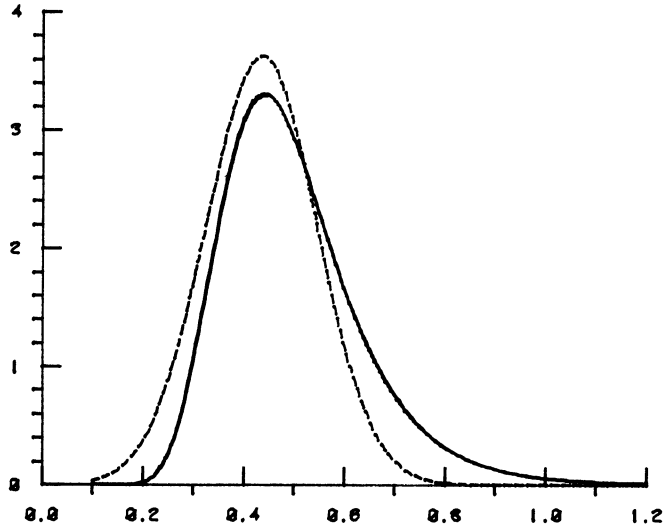

Figure 1. Marginal Posterior Densities for p. ——: Laplace and 20-point adaptive Gauss–Hermite approximations; ---: asymptotic normal approximation.

parameters using the $\hat{\Sigma}$ matrix computed for the Laplace approximation at that point. We then integrated with respect to each of the two orthogonalized parameters using a 20-point Gauss–Hermite quadrature. The results were renormalized using a rectangular integration formula and plots were obtained. In all three cases the resulting plots were indistinguishable from the renormalized Laplace approximations. Figure 1 shows the results for the marginal density of p. The solid line is the superposition of the renormalized Laplace approximation and the Gauss–Hermite calculations; the broken line is the asymptotic normal approximation.

As a final note on this example we mention the computing time requirements. On the University of Minnesota's Cyber 630 computer, computing the Laplace approximations to the moments took about .15 seconds of central processing unit (CPU) time; the Gauss–Hermite method took about 3 seconds. For marginals, the Laplace calculations took about 3 seconds and the Gauss–Hermite calculations about 60 seconds of CPU time.

6. CONCLUDING REMARKS

Several open questions remain to be investigated. The most important is to obtain approximations for moments of quantities taking on both positive and negative values, such as regression coefficients. Another is to determine the exact conditions under which the approximate marginalization approach produces exact results. It would also be useful to determine whether the approximations proposed here remain accurate when numerical derivatives are used in place of analytic ones in cases where closed-form derivatives are not available.

The approximations of this article may also prove helpful in certain theoretical problems such as developing tractable Bayesian approaches to log-linear models and to experimental design for nonlinear models. An extension that is currently being explored is to multimodal posteriors such as the poly-t distributions of Dickey (1968) and Dreze (1977).

In conclusion, we would like to emphasize that we do not

think of these approximations as replacements for exact calculations in situations where extremely accurate results are needed. Instead, we consider them to be simple first approximations that are easy to obtain, are often sufficiently accurate in their own right, and generally provide good starting points for exact computations, should these be required.

APPENDIX: ASYMPTOTIC ERRORS OF THE LAPLACE APPROXIMATIONS

Laplace's method for integrals as described, for example, in De Bruijn (1961) provides an approximation for integrals of the form $\int e^{nL(\theta)}d\theta$ when n is large. The idea is that if L has a unique maximum at $\hat{\theta}$, then for large n the value of this integral depends only on the behavior of the function L near its maximum. Thus if we set $\sigma^2 = -1/L''(\hat{\theta})$, then we can replace $L(\theta)$ by $L(\hat{\theta}) - (\theta - \hat{\theta})^2/(2\sigma^2)$. This produces the approximation

$$\int e^{nL(\theta)}d\theta \approx e^{nL(\hat{\theta})}\int \exp\{-[n(\theta - \hat{\theta})^2/2\sigma^2]\}d\theta$$
$$= \sqrt{2\pi}\sigma n^{-1/2}\exp\{nL(\hat{\theta})\}.$$

By expanding $n(L(\theta) - L(\hat{\theta}) + (\theta - \hat{\theta})^2/(2\sigma^2))$ about $\hat{\theta}$ and e^x about zero, it is possible to obtain the more refined result

$$\int e^{nL(\theta)}d\theta = \sqrt{2\pi}\sigma n^{-1/2}e^{nL(\hat{\theta})}\left(1 + \frac{a}{n} + \frac{b}{n^2} + O(n^{-3})\right), \quad (A.1)$$

where, setting $L_k = (d/d\theta)^k L(\hat{\theta})$, the constants a and b are given by

$$a = \frac{1}{8}\sigma^4 L_4 + \frac{5}{24}\sigma^6 L_3^2$$

and

$$b = \frac{1}{48}\sigma^6 L_6 + \frac{35}{384}\sigma^8 L_4^2 + \frac{7}{48}\sigma^8 L_3 L_5 + \frac{35}{64}\sigma^{10} L_3^2 L_4 + \frac{385}{1{,}152}\sigma^{12} L_3^4.$$

Result (A.1) remains valid if L is replaced by a sufficiently well-behaved sequence $L^{(n)}$ of functions. In this case the coefficients a and b may depend on n, but this dependence will be suppressed. If a and b do indeed depend on n, we will assume regularity conditions for the sequence $L^{(n)}$ that insure that a and b are bounded in n.

The posterior expectation $E_n[g]$ of a positive function g is a ratio of two integrals of the form $\int \exp\{nL^*(\theta)\}d\theta / \int \exp\{nL(\theta)\}d\theta$, where the difference $L^* - L$ is of order $O(n^{-1})$. Thus if we write $L^* = G/n + L$, with $G = \log g$, then

$$\frac{\int e^{nL^*(\theta)}d\theta}{\int e^{nL(\theta)}d\theta} = \frac{\sigma^* e^{nL^*(\hat{\theta}^*)}}{\sigma e^{nL(\hat{\theta})}}\left[1 + \frac{c}{n^2} + O(n^{-3})\right], \quad (A.2)$$

where $\hat{\theta}^*$ maximizes L^*, $\sigma^{*2} = -1/L^{*\prime\prime}(\hat{\theta}^*)$,

$$c = G_1 d_1 + G_2 d_2 + G_3 d_3 + G_4 d_4,$$

and the d_i's are given by

$$d_1 = \frac{1}{8}\sigma^6 L_5 + \frac{1}{4}\sigma^8 L_3 L_4 + \frac{5}{12}\sigma^8 L_3 L_4 + \frac{5}{8}\sigma^{10} L_3^3$$

$$d_2 = \frac{1}{4}\sigma^6 L_4 + \frac{5}{8}\sigma^8 L_3^2$$

$$d_3 = \frac{5}{12}\sigma^6 L_3$$

$$d_4 = \frac{1}{8}\sigma^4.$$

To see this, apply (A.1) to $\int \exp\{nL^*(\theta)\}d\theta$ and $\int \exp\{nL(\theta)\}d\theta$ to obtain

$$\int e^{nL^*(\theta)}d\theta \Big/ \int e^{nL(\theta)}d\theta$$
$$= \frac{\sigma^*}{\sigma}\exp\{n(L^*(\hat{\theta}^*) - L(\hat{\theta}))\}\frac{\left(1 + \frac{a^*}{n} + \frac{b^*}{n^2} + O(n^{-3})\right)}{\left(1 + \frac{a}{n} + \frac{b}{n^2} + O(n^{-3})\right)}$$
$$= \frac{\sigma^*}{\sigma}\exp\{n(L^*(\hat{\theta}^*) - L(\hat{\theta}))\} \times \left(1 + \frac{a^* - a}{n} + \frac{b^* - b - a(a^* - a)}{n^2} + O(n^{-3})\right),$$

again suppressing any dependence of a^*, b^* and a, b on n.

Then observe that $\hat{\theta}^*$ solves $L^{*\prime}(\theta) = 0$ and $\hat{\theta}$ solves $L'(\theta) = 0$; thus

$$0 = L^{*\prime}(\hat{\theta}^*) = L'(\hat{\theta}^*) + (1/n)G'(\hat{\theta}^*)$$
$$\approx L'(\hat{\theta}) + (\hat{\theta}^* - \hat{\theta})L''(\hat{\theta}) + (1/n)G'(\hat{\theta})$$
$$= -(\hat{\theta}^* - \hat{\theta})/\sigma^2 + (1/n)G_1.$$

So $\hat{\theta}^* - \hat{\theta} = (1/n)G_1\sigma^2 + O(n^{-2})$. Together with the fact that $L^*(\theta) - L(\theta) = (1/n)G(\theta) = O(n^{-1})$ for any θ, this implies that $a^* - a$ and $b^* - b$ are both of order $O(n^{-1})$.

[*Received June 1984. Revised September 1985.*]

REFERENCES

Barndorff-Nielsen, O., and Cox, D. R. (1979), "Edgeworth and Saddle-point Approximations With Statistical Applications," *Journal of the Royal Statistical Society*, Ser. B, 41, 270–312.

Daniels, H. E. (1954), "Saddle Point Approximations in Statistics," *Annals of Mathematical Statistics*, 25, 631–650.

——— (1956), "The Approximate Distribution of Serial Correlation Coefficients," *Biometrika*, 67, 335–349.

——— (1980), "Exact Saddlepoint Approximations," *Biometrika*, 67, 59–63.

De Bruijn, N. G. (1961), *Asymptotic Methods in Analysis*, Amsterdam: North-Holland.

Dickey, J. (1968), "Three Multidimensional Integral Identities With Bayesian Applications," *Annals of Mathematical Statistics*, 39, 1615–1627.

Dreze, J. (1977), "Bayesian Regression Analysis Using Poly-*t* Densities," *Journal of Econometrics*, 6, 329–354.

Dunsmore, T. R. (1976), "Asymptotic Prediction Analysis," *Biometrika*, 63, 627–630.

Kloek, T., and Van Dijk, H. K. (1978), "Bayesian Estimation of Equation System Parameters: An Application of Integration by Monte Carlo," *Econometrica*, 46, 1–19.

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.

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

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

Naylor, J. C., and Smith, A. F. M. (1982), "Applications of a Method for the Efficient Computation of Posterior Distributions," *Applied Statistics*, 31, 214–225.

——— (1983), "A Contamination Model in Clinical Chemistry: An Illustration of a Method for the Efficient Computation of Posterior Distributions," *The Statistician*, 32, 82–87.

Turnbull, B. W., Brown, B. W., Jr., and Hu, M. (1974), "Survivorship Analysis of Heart Transplant Data," *Journal of the American Statistical Association*, 69, 74–80.

Walker, A. M. (1969), "On the Asymptotic Behaviour of Posterior Distributions," *Journal of the Royal Statistical Society*, Ser. B, 31, 80–88.

Zellner, A., and Rossi, P. E. (1982), "Bayesian Analysis of Dichotomous Quantal Response Models," technical report, University of Chicago, Graduate School Business.