Bayesian and Likelihood Methods in Statistics and Econometrics
S. Geisser, J.S. Hodges, S.J. Press and A. Zellner (Editors)

THE VALIDITY OF POSTERIOR EXPANSIONS BASED ON LAPLACE'S METHOD

Robert E. Kass, Luke Tierney, and Joseph B. Kadane
Carnegie-Mellon University, University of Minnesota, and Carnegie-Mellon University

Abstract

We present methods for justifying heuristic derivations of asymptotic expansions for predictive densities, odds factors, marginal posterior densities, and posterior moments given by various authors. In applications, an observed sample is imbedded in an infinite sequence to which an expansion is applied. We therefore begin by specifying what constitutes a well-behaved sequence using an approach similar to that of Chen (1985). We then go on to give conditions under which sequences are well-behaved with probability one. The latter results are closely related to those of Johnson (1970).

Key Words and Phrases: Bayesian inference, second-order theory, asymptotics.

1. Introduction

Laplace's method for the asymptotic evaluation of integrals (Laplace, 1847) is a basic technique of mathematical analysis (described, for instance, in Erdelyi, 1956), which has been used frequently in statistical theory. In Bayesian inference Laplace's method has been used by various authors to derive approximate posterior expectations and marginal densities, predictive densities, and Bayes factors. The purpose of the present paper is to present methods for justifying these approximations. Our approach is similar to that used by Chen (1985) in discussing asymptotic posterior Normality, and our results are closely related to those of Johnson (1970). For additional references and discussion, with emphasis on the role of asymptotic expansions in interactive data analysis, see Kass, Tierney, and Kadane (1988).

The primary problem we concern ourselves with is that of approximating the expectation of a function $g(\theta)$, where θ is a parameter vector having a posterior probability density based on a sample of n observations. The posterior density will be proportional to a likelihood function $L = L_n$ and a prior density π so that the expectation we seek to evaluate asymptotically will have the form

$$E(g(\theta)) = \frac{\int g(\theta)L(\theta)\pi(\theta)d\theta}{\int L(\theta)\pi(\theta)d\theta}. \tag{1.1}$$

If the likelihood function is well-behaved, it will have a dominant peak at its maximum. Laplace's method will then be suitable for application to both the numerator and denominator of (1.1): each may be written in the form

$$\mathcal{I} = \int b(\theta)\exp[-nh_n(\theta)]d\theta \tag{1.2}$$

where $-h_n$ has a maximum at a point $\hat{\theta}$ and b and $h = h_n$ may be expanded in a Taylor series about $\hat{\theta}$. The method consists of approximating the integrand by the product of a low-order polynomial and the factor

$$\exp[-\frac{n}{2}\Sigma h_{ij}(\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j)]$$

where h_{ij} is the (i,j)-component of the Hessian of h at $\hat{\theta}$; the resulting integral is then evaluated in terms of the moments of a multivariate Normal distribution.

Note that there are infinitely many choices for the functions b and h in any particular application. For example, in writing the denominator of (1.1) in the form of (1.2), two obvious choices for h are $h_n(\theta) = n^{-1}\log[L(\theta)]$ and $h_n(\theta) = n^{-1}\log[L(\theta)\pi(\theta)]$, which determine $\hat{\theta}$ to be, respectively, the MLE and the posterior mode. Tierney and Kadane (1986) used the latter in the denominator and then, assuming $g(\theta)$ was positive, took $h_n(\theta) = n^{-1}\log[g(\theta)L(\theta)\pi(\theta)]$ in the numerator. They showed that for this pair of choices, which we will call *fully exponential*, first-order approximations in each of the numerator and denominator of (1.1), having error of order $O(n^{-1})$, lead to a second-order approximation of the ratio, having error of order $O(n^{-2})$. This may be contrasted with the methods of Lindley (1961, 1980) and Mosteller and Wallace (1964), in which h was the same function in both numerator and denominator of (1.1), and which involved third derivative terms in the second-order approximations.

In a recent paper, Tierney, Kass, and Kadane (1989b) obtained a second-order expansion of the expectation of a general function $g(\theta)$ (which need not be positive), by applying the fully exponential method to approximate the moment generating function $E(\exp[sg(\theta)])$ and then differentiating. They showed that this expansion is formally equivalent to the one used by Lindley and Mosteller and Wallace, while maintaining the computational simplicity of the fully exponential method for positive functions.

The arguments given by many authors who have applied Laplace's method to the calculation of features of posterior distributions have usually been heuristic. They have essentially relied on two presumptions: first, the well-known conditions for the rigorous justification of Laplace's method in producing *analytical* expansions (see Erdelyi, 1956) may be replaced by suitable extensions to cover the cases of interest in statistics; that is, for well-behaved sequences of data, the posterior distribution will be sufficiently well-behaved that Laplace's method will apply. Second, under certain general conditions, these well-behaved data sequences will be the rule rather than the exception in a *stochastic* sense. Using arguments related to those of Le Cam (1953), Johnson (1967, 1970) effectively showed that in certain cases these presumptions are correct, by providing a set of conditions under which the expansions will be valid with probability one. Johnson treated the i.i.d. case in detail, and noted how his results could be extended to cover the Markov case. (See also Johnson and Ladalla, 1979, who treated multiparameter problems, and Ghosh, Sinha, and Joshi, 1982, Joshi, 1984, and Bickel, Götze, and van Zwet (1985) who provided rigorous treatments of uniform expansions.)

Our approach is somewhat different than Johnson's in that we explicitly separate the analytical and stochastic parts of the problem. From the Bayesian point of view this seems desirable since, in applications, the use of Laplace's method effectively imbeds the *given* set of observations into an infinite sequence to which the method applies. We state sufficient conditions on the behavior of the sequence of loglikelihood functions for validity of analytical expansions, and call a sequence of loglikelihood functions satisfying these conditions *Laplace regular*. In linear regression, Laplace regularity becomes an assumption on the sequence of design matrices and residual vectors: the eigenvalues of $X^T X/n$ must be bounded and bounded away from zero, and $\hat{\sigma}_n$ must be bounded away from zero. No probability statements are required. We then go on to consider the family of sampling densities constituting the model and provide conditions to guarantee that Laplace regularity of the loglikelihood functions will hold with probability one. When a model satisfies these conditions we refer to it as a *Laplace regular model*. In linear regression, the model is Laplace regular if the eigenvalues of $X^T X/n$ are bounded and bounded away from zero. It is apparent from an examination of Johnson's arguments that demonstrating Laplace regularity of a model is essentially what is needed in extending his results. We have not tried to establish a very general theorem guaranteeing Laplace regularity of stochastic process models, but instead intend Laplace regularity to be verified for interesting special families as the need arises in practice.

We treat second-order approximations, that is, expansions that incorporate the first correction term beyond the MLE or modal approximation for expectations and the Normal approximation for marginal densities. Higher-order theory is analogous. Our analytical results are presented in section 2, and our results relating Laplace regularity of models to Laplace regularity of sequences of loglikelihood functions are given in section 3.

We should emphasize the technical nature of this work on two accounts. First, although our point of view is Bayesian, our results may be useful in non-Bayesian theory as well. Second, to say that an expansion is valid asymptotically is not to say that it provides a good approximation in any particular instance. Although these expansions have been shown to be quite useful (in the references cited by Kass, Tierney, and Kadane, 1988), it is not hard to construct finite-sample examples in which the approximations perform poorly. Numerical techniques including Gaussian quadrature and Monte Carlo remain essential alternatives to asymptotic methods in practical work. See Naylor and Smith (1982) and Smith et al. (1985) for discussion of these. The choice of parameterization in which Laplace's method is applied is also an important practical concern, as shown by the examples in Tierney, Kass, and Kadane (1988), and Achcar and Smith (1989). Alternative expansions, based on Pearson family kernels that replace the Normal kernel of Laplace's method, have been shown by Morris (1988) to offer improvements in many one-parameter problems.

2. Analytical Expansions

Suppose Θ is an open subset of R^m, $\{h_n : n = 1, 2, \ldots\}$ is a sequence of six-times continuously differentiable real functions on Θ, having local minima $\{\hat{\theta} = \hat{\theta}_n : n = 1, 2, \ldots\}$, and b is a four-times continuously differentiable real function on Θ. We use $B_\delta(\theta)$ to denote the open ball of radius δ centered at θ. The Hessian of $h = h_n$ at θ will be denoted by $D^2h[\theta]$, and its (i, j)-component will be written either as $\partial_{ij}h[\theta]$ or h_{ij} while the components of its inverse will be written h^{ij}.

The pair $(\{h_n\}, b)$ will be said to satisfy the *analytical assumptions for Laplace's method* if there exist positive numbers ε, M and η, and an integer n_0 such that $n \geq n_0$ implies

(i) for all $\theta \in B_\varepsilon(\hat{\theta})$ and all $1 \leq j_1, \ldots, j_d \leq m$ with $0 \leq d \leq 6$, $|\,\partial_{j_1 \cdots j_d} h_n[\theta]\,| < M$;

(ii) $\det(D^2 h_n[\hat{\theta}]) > \eta$;

and the integral in (1.2) exists and is finite, and

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

$$[\det(nD^2 h_n[\hat{\theta}])]^{1/2} \cdot \int_{\Theta - B_\delta(\hat{\theta})} b(\theta) \exp[-n(h_n(\theta) - h_n(\hat{\theta}))] d\theta = O(n^{-2}).$$

THEOREM 1: *If $(\{h_n\}, b)$ satisfy the analytical assumptions for Laplace's method then*

$$\int_\Theta b(\theta)\exp[-nh(\theta)]d\theta = (2\pi)^{m/2}[\det(nD^2h)]^{-1/2}\exp[-nh(\hat{\theta})]$$

$$\cdot\Bigg(b(\hat{\theta}) + \frac{1}{n}\Bigg\{\frac{1}{2}\Sigma h^{ij}b_{ij} - \frac{1}{6}\Sigma h_{ijk}b_s\mu^4_{ijks}$$

$$+\frac{1}{72}b(\hat{\theta})\Sigma h_{ijk}h_{qrs}\mu^6_{ijkqrs} \tag{2.1}$$

$$-\frac{1}{24}b(\hat{\theta})\Sigma h_{ijks}\mu^4_{ijks}\Bigg\} + O(n^{-2})\Bigg)$$

where μ^4_{ijks} and μ^6_{ijkqrs} are the fourth and sixth central moments of a multivariate Normal distribution having covariance matrix $(D^2h)^{-1}$, i.e.,

$$\begin{aligned}
\mu^4_{ijks} &= \sigma_{ij}\sigma_{ks} + \sigma_{ik}\sigma_{js} + \sigma_{is}\sigma_{jk}\\
\mu^6_{ijkqrs} &= \sigma_{ij}\sigma_{kq}\sigma_{rs} + \sigma_{ij}\sigma_{kr}\sigma_{qs} + \sigma_{ij}\sigma_{ks}\sigma_{qr}\\
&+ \sigma_{ik}\sigma_{jq}\sigma_{rs} + \sigma_{ik}\sigma_{jr}\sigma_{qs} + \sigma_{ik}\sigma_{js}\sigma_{qr}\\
&+ \sigma_{iq}\sigma_{jk}\sigma_{rs} + \sigma_{iq}\sigma_{jr}\sigma_{ks} + \sigma_{iq}\sigma_{js}\sigma_{kr}\\
&+ \sigma_{ir}\sigma_{jh}\sigma_{qs} + \sigma_{ir}\sigma_{jq}\sigma_{ks} + \sigma_{ir}\sigma_{js}\sigma_{kq}\\
&+ \sigma_{is}\sigma_{jk}\sigma_{qr} + \sigma_{is}\sigma_{jq}\sigma_{kr} + \sigma_{is}\sigma_{jr}\sigma_{kq}
\end{aligned}$$

where $\sigma_{ij} = h^{ij}$ and all derivatives b_s, etc., are evaluated at $\hat{\theta}$.

PROOF: For simplicity, we treat the case $m = 1$, the general case involving obvious modifications. Let $\hat{\theta}$ be the maximum of $-h_n$, let $u \equiv n^{1/2}(\theta - \hat{\theta})$ so that for a given u, $(\theta - \hat{\theta})^k = (n^{-1/2}u)^k = O(n^{-k/2})$, and consider the expansion of the integrand $b(\theta)\exp[-nh_n(\theta)]$ about $\hat{\theta}$ that includes terms of order greater than $O(n^{-2})$. These terms involve the cubic expansion of b, the quadratic expansion of $\exp[-x]$, and the quintic expansion of h_n, as follows. We have

$$nh(\theta) = \{nh(\hat{\theta})\} + \left\{\frac{1}{2}h''(\hat{\theta})u^2\right\} + \left\{\sum_{k=3}^{5}(n^{(k-2)/2}k!)^{-1}h^{(k)}(\hat{\theta})u^k\right\} + r_n(u)$$

where $r_n(u)$ is bounded over $B_\varepsilon(\hat{\theta})$ by a polynomial in u having coefficients that are of order $O(n^{-2})$. Applying the expansion of $\exp[-x]$ to the third bracketed term above, and expanding b, we get

$$
\begin{aligned}
b(\theta)\exp[-nh(\theta)] = \{\exp[-nh(\hat\theta)]\}&\left\{\exp[-\frac{1}{2}h''(\hat\theta)u^2]\right\}\\
&\cdot\left\{1-\frac{1}{6}n^{-1/2}h^{(3)}(\hat\theta)u^3+\frac{1}{72}n^{-1}[h^{(3)}(\hat\theta)]^2u^6-\frac{1}{24}n^{-1}h^{(4)}(\hat\theta)u^4\right.\\
&\left.-\frac{1}{120}n^{-3/2}h^{(5)}(\hat\theta)u^5+\frac{1}{72}n^{-3/2}h^{(3)}(\hat\theta)h^{(4)}(\hat\theta)u^7+R_{1n}(\theta,\hat\theta)\right\}\\
&\cdot\left\{b(\hat\theta)+n^{-1/2}b'(\hat\theta)u+\frac{1}{2}n^{-1}b''(\hat\theta)u^2\right.\\
&\left.+\frac{1}{6}n^{-3/2}b^{(3)}(\hat\theta)u^3+R_{2n}(\theta,\hat\theta)\right\}
\end{aligned}
\tag{2.2}
$$

where $R_{jn}(\theta,\hat\theta)=O(n^{-2})$ uniformly on $B_\varepsilon(\hat\theta)$, j=1,2. A heuristic argument would conclude by using $d\theta=n^{-1/2}du$, and noting that the odd moments of a Normal distribution vanish so that formal integration of (2.2) would yield (2.1).

Let us rewrite (2.2) as

α

$$
b(\theta)\exp[-nh(\theta)]=\{\exp[-nh(\hat\theta)]\}\left\{\exp[-\frac{1}{2}h''(\hat\theta)u^2]\right\}\left\{I_n(\theta,\hat\theta)+R_n(\theta,\hat\theta)\right\}
$$

where $R_n(\theta,\hat\theta)$ is the sum of all terms involving R_{1n}, R_{2n}, or u^s with $s\geq 7$. This expansion holds (for $n\geq n_0$) on a neighborhood $B_\delta(\hat\theta)$ with $\delta<\varepsilon$. Condition (iii) reduces consideration to the integral of $b(\theta)\exp[-nh(\theta)]$ over $B_\delta(\hat\theta)$, on which region expansion (2.2) holds; there are thus two integrals to evaluate, one involving I_n and one involving R_n.

The integral of $\{\exp[-\frac{1}{2}h''(\hat\theta)u^2]\}I_n(\theta,\hat\theta)$ over $B_\delta(\hat\theta)$ is evaluated by first noting that on changing variables from θ to u the domain of integration becomes $u(B_\delta(\hat\theta))=B_{\delta(n)}(0)$ where $\delta(n)=n^{1/2}\delta$. Since this domain is expanding at the rate $O(n^{1/2})$, and I_n is a polynomial in u, which is being integrated against a Normal density in u, the replacement of this domain by the whole real line incurs an error of exponentially decreasing order. Using $d\theta=n^{-1/2}du$, writing $\mu^4=3[h''(\hat\theta)]^{-2}$ and $\mu_6=15[h''(\hat\theta)]^{-3}$ as the fourth and sixth central moments of a Normal distribution having precision $h''(\hat\theta)$, and noting that the odd moments vanish then yields the right-hand side of (2.1) as the integral of $\{\exp[-\frac{1}{2}h(\hat\theta)]\}\cdot\{\exp[-\frac{1}{2}h''(\hat\theta)u^2]\}\cdot I_n(\theta,\hat\theta)$ over $B_\delta(\hat\theta)$.

The terms comprising R_n may be represented explicitly using the mean value form of the remainders in terms of higher derivatives of b and h evaluated at points between $\hat{\theta}$ and θ, e.g., one such term is $(1/120)n^{-2}h^{(6)}(\xi)u^6$, where $\xi = \xi(u)$ is between $\hat{\theta}$ and θ; it is one piece of the error term appearing as $R_{1n}(\theta, \hat{\theta})$. Now $|h^{(6)}(\theta)| \leq M$ on $B_\delta(\hat{\theta})$ implies $|\int \exp\{-\frac{1}{2}h''(\hat{\theta})u^2\}n^{-2}h^{(6)}(\xi)u^6 du| \leq Mn^{-2}\int \exp\{-\frac{1}{2}h''(\hat{\theta})u^2\}u^6 du = O(n^{-2})$ (the integrals being taken over $B_{\delta(n)}(0)$ where $\delta(n) = n^{1/2}\delta$). The other terms are similar. Thus, the integral of $\{\exp[-nh(\hat{\theta})]\} \cdot \{\exp[-\frac{1}{2}h''(\hat{\theta})u^2]\} \cdot R_n(\theta, \hat{\theta})$ over $B_\delta(\hat{\theta})$ is of order $O(n^{-2})$. □

Now suppose $-nh_n$ is the log of a posterior density. Our conditions (i)-(iii) are closely related to those of Chen (1985). Chen assumes $\Theta = R^m$. Then (iii) with $b = 1$ and the right-hand side being $o(1)$ becomes Chen's condition (C3). It follows from Chen's Theorem 2.1 that under (i) and (ii) this condition is necessary and sufficient for asymptotic posterior Normality. To obtain an expansion of the expectation (1.1) we need only assume that both $(\{h_n\}, \pi)$ and $(\{h_n\}, g\pi)$ satisfy the analytical assumptions for Laplace's method, where $-nh_n$ is the loglikelihood function. In applications, however, it is often easier to verify an alternative condition, which is stronger than condition (iii) when $b = 1$:

(iii′) for all δ for which $0 < \delta < \varepsilon$, $B_\delta(\hat{\theta}) \subseteq \Theta$ and

$$\limsup_{n\to\infty} \sup_{\theta}\{h_n(\hat{\theta}) - h_n(\theta) : \theta \in \Theta - B_\delta(\hat{\theta})\} < 0.$$

An additional convenience of this condition is that we will be able to treat at once the posterior expectations of all four-times continuously differentiable functions g based on four-times continuously differentiable priors π. When a sequence of loglikelihood functions $-nh_n(\theta) = \ell_n(\theta) = \log L_n(\theta)$ satisfies conditons (i), (ii), and (iii′) we will say it is *Laplace regular.*

EXAMPLE: *Linear regression.* The Normal linear model is $Y_i = x_i^T\beta + \varepsilon_i$ where ε_i are i.i.d. Normal$(0, \sigma^2)$, β is an unknown vector in R^p and x_i is a given vector in R^p, for i = 1, 2, ..., n. When the vectors $x_1^T, \ldots, x_n^T$ are collected into a matrix X_n, the loglikelihood function based on $y = (y_1, \ldots, y_n)$ becomes

$$\ell(\beta, \sigma) - \ell(\hat{\beta}, \hat{\sigma}) = -\frac{1}{2\sigma^2} \parallel X_n(\beta - \hat{\beta}) \parallel^2 - \frac{n}{2}\left(\frac{\hat{\sigma}^2}{\sigma^2} - 1 - \log\left(\frac{\hat{\sigma}^2}{\sigma^2}\right)\right)$$

with $\hat{\beta}$ the least-squares estimator and $\hat{\sigma}^2 = n^{-1} \parallel y - X_n\hat{\beta} \parallel^2$. Let λ_{n1} and λ_{nm} be the largest and smallest eigenvalues of $n^{-1}X_n^TX_n$. From the inequality $\log u \leq u - 1$, which holds for all $u \geq 0$, condition (iii′) of Laplace regularity holds if there exists $A > 0$ such that for all sufficiently large n, $A < \hat{\sigma}_n$ and for all $\delta > 0$

$$\sup_{\theta}\{-\frac{1}{n} \parallel X_n(\beta - \hat{\beta}) \parallel^2 : \beta \in B_\delta(\hat{\beta})^c\} < 0.$$

δ β $\longrightarrow$

The latter condition, in turn, is satisfied if there exists $a > 0$ such that for all sufficiently large n, $a < \lambda_{nm}$. (In the nonsingular case, the condition $(X_n^T X_n)^{-1} \to 0$ is necessary and sufficient for consistency of $\hat{\beta}$ under the much weaker assumption that ε_i are i.i.d. with mean zero and variance $\sigma^2 < \infty$; see Lai, Robbins, and Wei, 1978.) Condition (i) requires, in addition, that there exist $b < \infty$ such that, for all sufficiently large n, $\lambda_{n1} < b$. Condition (ii) is immediate. □

Before stating our theorems for Laplace-regular sequences, we first present a lemma which will be needed and a corollary which is of some interest on its own. We let $y^{(n)} = (y_1, \ldots, y_n)$ denote the sample on which the loglikelihood ℓ_n is based.

LEMMA 2: *Suppose $\{\ell_n\}$ is a Laplace-regular sequence of loglikelihood functions, π is a four-times continuously differentiable positive real function on Θ, and for some nonnegative integer n_0 the posterior $P\{\cdot \mid y^{(n_0)}\}$ based on $y^{(n_0)}$ and the prior π exists (i.e., the integral of $\exp\{\ell_{n_0}\}\pi$ is finite); then, with ε given in the definition of Laplace regularity, for any δ for which $0 < \delta < \varepsilon$, there exists a positive number c_1 such that for all sufficiently large n,*

$$\int_{\Theta - B_\delta(\hat{\theta})} \exp\{\ell_n(\theta) - \ell_n(\hat{\theta})\}\pi(\theta)d\theta < \exp\{-nc_1\}. \tag{2.3}$$

COROLLARY 3: *Under the conditions of Lemma 2, for any δ for which $0 < \delta < \varepsilon$, there exists a positive number c_2 such that for all sufficiently large n,*

$$P(\Theta - B_\delta(\hat{\theta}) \mid y^{(n)}) < \exp\{-nc_2\}. \tag{2.4}$$

PROOFS: Suppose first that π is a proper prior (i.e., it integrates to one); then, by condition (iii′), there exists $c_1 > 0$ such that (2.3) holds. Now note that Theorem 1 may be applied with $h_n = -n^{-1}\ell_n$ and $b = \pi$. This provides the normalizing constant for the posterior: retaining only the first term in the expansion, we have

$$\int_\Theta \exp\{\ell_n(\theta) - \ell_n(\hat{\theta})\}\pi(\theta)d\theta = (2\pi)^{m/2}[\det(-D^2\ell_n(\hat{\theta}))]^{-1/2}\pi(\hat{\theta}) \cdot \{1 + O(n^{-1})\}.$$

Combining this with (2.3), we obtain (2.4) for the case in which π is proper.

The general case is treated by applying the above argument with the prior π replaced by the posterior density based on $y^{(n_0)}$ (and π). This replacement is possible because condition (iii′) of Laplace regularity also holds when applied to the "loglikelihood" $\ell_{n,n_0}(\theta) = \log p(y_{n_0+1}, \ldots, y_n \mid y^{(n_0)}, \theta)$. This is so because $\limsup_{n\to\infty} \sup_\theta\{n^{-1}[\ell_{n_0}(\theta) - \ell_{n_0}(\hat{\theta})] : \theta \in \Theta - B_\delta(\hat{\theta})\} = 0$. □

We now present the main results of this section, establishing second-order approximations to the posterior expectation of a real function g on Θ: first in a standard form, then in a *fully exponential* form used by Tierney and Kadane (1986); finally we give the basic approximation for a marginal density.

Given a prior π, loglikelihood ℓ_n, positive function γ, and real function g we define h_n and b by $-nh_n = \ell_n + \log \gamma$ and $b = \pi/\gamma$. We then have

$$E(g(\theta)) = \frac{\int g(\theta)b(\theta)\exp[-nh_n(\theta)]d\theta}{\int b(\theta)\exp[-nh_n(\theta)]d\theta} \tag{2.5}$$

where the integrals are taken over Θ. The approximation we justify, in the notation of Theorem 1, is

$$E(g(\theta)) = g(\hat{\theta}) + \frac{1}{n}\Sigma g_i h^{ij}\left\{(b_j/b(\hat{\theta})) - \frac{1}{2}\Sigma h^{rs}h_{rsj}\right\} + \frac{1}{2n}\Sigma h^{ij}g_{ij} + O(n^{-2}). \tag{2.6}$$

Approximation (2.6) was used and discussed by Mosteller and Wallace (1964). The special case $\gamma = 1$ makes $b = \pi$ while $\hat{\theta}$ becomes the MLE, and the choice $\gamma = \pi$ makes $b = 1$ with $\hat{\theta}$ becoming the posterior mode. The latter was used by Lindley (1961, 1980).

THEOREM 4: *Under the conditions of Lemma 2 if, in addition, γ is a six-times and g is a four-times continuously differentiable real function on θ, γ is positive, and the posterior expectation of g, based on $y^{(n_0)}$ and π, is finite, then the posterior expectation of g has the expansion (2.6).*

PROOF: Letting $p(\theta \mid y^{(n_0)})$ be the posterior density based on $y^{(n_0)}$ and π, using (2.3) and arguing as in the proof of Lemma 2, for any $\delta > 0$ there is a positive number c such that

$$\int_{\Theta - B_\delta(\hat{\theta})} g(\theta)\exp\{\ell_n(\theta) - \ell_n(\hat{\theta})\}\pi(\theta)d\theta < \exp\{-nc\}.$$

Thus, the order of error will be unaffected by replacing Θ with $B_\delta(\hat{\theta})$ in the numerator and denominator of (2.5) and Theorem 1 may be applied to both integrals. □

When g is positive, an alternative to approximating the expectation in (2.5) using (2.6) is to define $-nh_n = \ell_n + \log \pi$ and $-nh_n^* = -nh_n + \log g$ so that

$$E(g(\theta)) = \frac{\int \exp[-nh_n^*(\theta)]d\theta}{\int \exp[-nh_n(\theta)]d\theta}\,. \tag{2.7}$$

Tierney and Kadane (1986) showed that first-order approximations in the numerator and denominator of (2.7) yield a second-order approximation

$$E(g(\theta)) = \frac{\det(\Sigma^*)^{1/2}\exp[-nh_n^*(\theta^*)]}{\det(\tilde{\Sigma})^{1/2}\exp[-nh_n(\tilde{\theta})]} + O(n^{-2}) \tag{2.8}$$

where $\tilde{\Sigma} = (nD^2h[\tilde{\theta}])^{-1}$ and $\Sigma^* = (nD^2h^*[\theta^*])^{-1}$, with $\tilde{\theta}$ and θ^* maximizing $-h_n$ and $-h_n^*$.

THEOREM 5: Approximation of expectations based on the fully exponential method. *Suppose the sequence of loglikelihood functions $\{\ell_n\}$ is Laplace-regular, π and g are six times continuously differentiable positive real functions on Θ, and for some nonnegative integer n_0 the posterior based on $y^{(n_0)}$ and π exists and the posterior expectation of g is finite; then the posterior expectation of g has the expansion (2.8).*

PROOF: The proof of Theorem 5 is essentially the same as that of Theorem 4. □

Suppose now that $\Theta = \Theta_1 \times \Theta_2$, with Θ_1 and Θ_2 being of dimension m_1 and m_2, γ is positive, and h_n and b are defined by $-nh_n = \ell_n + \log\gamma$ and $b = \pi/\gamma$. The marginal posterior density of θ_1 becomes

$$p(\theta_1 \mid y^{(n)}) = \frac{\int_{\Theta_2} b(\theta_1,\theta_2)\exp[-nh_n(\theta_1,\theta_2)]d\theta_2}{\int_{\Theta} b(\theta_1,\theta_2)\exp[-nh_n(\theta_1,\theta_2)]d\theta_1 d\theta_2} \tag{2.9}$$

and a first-order approximation is

$$p(\theta_1 \mid y^{(n)}) = (2\pi)^{-m_1/2} \cdot \frac{\det(\Sigma(\theta_1))^{1/2}\exp[-nh_n(\theta_1,\hat{\theta}_2(\theta_1))]b(\theta_1,\hat{\theta}_2(\theta_1))}{\det(\hat{\Sigma})^{1/2}\exp[-nh_n(\hat{\theta})]b(\hat{\theta})} \cdot \{1 + 0(n^{-1})\} \tag{2.10}$$

where $\hat{\Sigma}$ and $\Sigma(\theta_1)$ are the inverses of the Hessian matrices of nh_n and $nh_n(\theta_1,\cdot)$ at $\hat{\theta}$ and $(\theta_1,\hat{\theta}_2(\theta_1))$ with $\hat{\theta}$ and $\hat{\theta}_2(\theta_1)$ maximizing $-h_n$ and $-h_n(\theta_1\ ,\cdot)$. We write the marginal posterior mode, i.e., the maximum of $p(\theta_1 \mid y^{(n)})$, as $\hat{\theta}_1$. As noted by Tierney and Kadane (1986), approximation (2.10) holds uniformly on bounded regions about $\hat{\theta}_1$.

THEOREM 6: Approximation of marginal densities. *Suppose the sequence of loglikelihood functions $\{\ell_n\}$ is Laplace-regular, $\Theta = \Theta_1 \times \Theta_2$ with Θ_1 and Θ_2 being open subsets of R^{m_1} and R^{m_2} , π and γ are six times continuously differentiable positive* *real functions on Θ, and for some n_0 the posterior based on $y^{(n_0)}$ and π exists; then, with ε given in the definition of Laplace regularity, for any δ for which $0 < \delta < \varepsilon$, the marginal posterior density of θ_1 has the expansion (2.10) uniformly for all θ_1 in $B_\delta(\hat{\theta}_1)$,*

PROOF: As in the proof of Lemma 2, the denominator of (2.9) may be approximated by the denominator of (2.10) (multiplied by $(2\pi)^{m/2}$), with multiplicative error of order $O(n^{-1})$, which does not depend on θ_1. For any $\delta < \varepsilon$, Theorem 1 may be applied to approximate the numerator in (2.9) for each $\theta_1 \in B_\delta(\hat{\theta}_1)$. To obtain uniformity we return to the proof of Theorem 1. Repeating the argument following (2.2), we pick $\delta^* < (\varepsilon^2 - \delta^2)^{1/2}$ and then note that Θ_2 may be replaced by $B_{\delta^*}(\hat{\theta}_2(\theta_1)) \subseteq \Theta_2$, and each error incurred in the subsequent steps may be replaced by its supremum over $\theta_1 \in B_\delta(\hat{\theta}_1)$. □

A more general form of approximation (2.10) is presented by Tierney, Kass, and Kadane (1989a).

3. Stochastic Expansions

In this section we list assumptions on a family of distributions that guarantee the sequence of loglikelihood functions will be Laplace regular with probability one. We then briefly relate our approach to that of Johnson (1970) and to the problem of guaranteeing consistency of the MLE.

Suppose Θ is an open subset of R^m, $(\Omega, \mathcal{A})$ is a measurable space, $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is a family of probability distributions on $(\Omega, \mathcal{A})$, and $\{Y_i : i = 1, 2, \ldots\}$ is a stochastic process on $(\Omega, \mathcal{A})$ with the Y_i's taking values in $(\mathcal{Y}, \mathcal{B})$ where $\mathcal{Y}$ is a subset of R^k and $\mathcal{B}$ is the class of Borel subsets of $\mathcal{Y}$. Letting $Y^{(n)} = (Y_1, \ldots, Y_n)$, we will assume that for all n the n-dimensional distributions of $Y^{(n)}$ are dominated by a σ-finite measure, and we will denote a density of $Y^{(n)}$ under P_θ by $p(y^{(n)} \mid \theta)$; we will denote the loglikelihood function by ℓ_n, i.e., $\ell_n(\theta) = \log p(y^{(n)} \mid \theta)$.

The family $\mathcal{P}$ will be called *Laplace-regular* if there exist densities $p(y^{(n)} \mid \theta)$ such that

(i) for all $y^{(n)}$ and θ, $p(y^{(n)} \mid \theta) > 0$ and for all $y^{(n)}$, ℓ_n is six times continuously differentiable;

(ii) for all $\theta_0 \in \Theta$ there exist $\varepsilon > 0$ and $M < \infty$ such that $B_\varepsilon(\theta_0) \subseteq \Theta$ and for all $1 \leq j_1, \ldots, j_d \leq m$ with $d \leq 6$,

$$\limsup_{n\to\infty} \sup_{\theta}\{n^{-1} \mid \partial_{j_1}\cdots_{j_d} \ell_n[\theta] \mid : \theta \in B_\varepsilon(\theta_0)\} < M,$$

with P_{θ_0}-probability one;

(iii) for all $\theta_0 \in \Theta$ and some $\varepsilon > 0$,

$$\limsup_{n\to\infty} \sup_{\theta}\{n^{-1} \det(D^2\ell_n[\theta]) : \theta \in B_\varepsilon(\theta_0)\} < 0$$

with P_{θ_0}-probability one;

(iv) for all $\theta_0 \in \Theta$ and all $\delta > 0$,

$$\limsup_{n\to\infty} \sup_{\theta}\{n^{-1}[\ell_n(\theta) - \ell_n(\theta_0)] : \theta \in \Theta - B_\delta(\theta_0)\} < 0$$

with P_{θ_0}-probability one.

THEOREM 7: *If a family $\mathcal{P}$ is Laplace regular its sequence of loglikelihood functions is Laplace regular with P_{θ_0}-probability one, for all θ_0.*

PROOF: This is essentially immediate. We need only note that (iv) is a familiar consistency condition for the MLE, and it entails condition (iii′) of Laplace regularity for ℓ_n. □

EXAMPLE: *Linear and nonlinear regression.* Consider again the linear regression setting given in the previous section. The condition that the eigenvalues of $n^{-1}X_n^TX_n$ be bounded and bounded away from zero is precisely what is needed for Laplace regularity of the family.

Suppose instead that $Y_i = f(x_i, \beta) + \varepsilon_i$, with ε_i again being i.i.d. Normal$(0, \sigma^2)$ and f being a smooth function of β. To check condition (iii′) of Laplace regularity of the sequence of loglikelihood functions we could use

$$\begin{aligned}\ell(\beta,\sigma) - \ell(\hat\beta,\hat\sigma) &= -\frac{1}{\sigma^2}\Sigma(y_i - f(x_i,\hat\beta))(f(x_i,\hat\beta) - f(x_i,\beta)) \\ &\quad - \frac{1}{\sigma^2}\Sigma(f(x_i,\hat\beta) - f(x_i,\beta))^2 - \frac{n}{2}\left(\frac{\hat\sigma^2}{\sigma^2} - 1 - \log\left(\frac{\hat\sigma^2}{\sigma^2}\right)\right) \\ &\le -\frac{1}{\sigma^2}\Sigma(y_i - f(x_i,\hat\beta))(f(x_i,\hat\beta) - f(x_i,\beta)) \\ &\quad - \frac{1}{2\sigma^2}\Sigma(f(x_i,\hat\beta) - f(x_i,\beta))^2\end{aligned}$$

the inequality following from $u - 1 \ge \log u$, as in the linear regression case.

Alternatively, Laplace regularity of the family may be checked as follows. Suppose, in addition, that β lies in a compact subset B of R^p, $D_n(\beta,\beta') = \sum_{i=1}^n (f(x_i,\beta) - f(x_i,\beta'))^2$ satisfies

(A) $n^{-1}D_n(\beta,\beta')$ converges uniformly to a continuous function $D(\beta,\beta')$ and $D(\beta,\beta') = 0$ if and only if $\beta = \beta'$,

and $D_\beta f = (\partial_{\beta_1} f[x_i,\beta], \ldots, \partial_{\beta_p} f[x_i,\beta])$ satisfies

(B) $n^{-1}\sum_{i=1}^n (D_\beta f)^T(D_\beta f)$ converges uniformly to a matrix $A(\beta)$, which is continuous in β and positive definite,

while for all $1 \le j_1, \ldots, j_d \le p$ with $1 \le d \le 6$

(C) $n^{-1}\sum_{i=1}^n (\partial_{j_1} f)\ldots(\partial_{j_d} f)$ converges uniformly.

Condition (A) corresponds to conditions (a) and (b) of Jennrich (1969) while (B) and (C) together correspond to his conditions (c) and (d) when $d = 1$ or 2. Assuming these when ε_i are i.i.d. with mean zero and variance $\sigma^2 < \infty$, Jennrich derived the asymptotic consistency and Normality of the nonlinear least-squares estimator. From the proof of Jennrich's Theorem 6 it may be seen that conditions (A)-(C) imply Laplace regularity for Normal nonlinear regression models. Alternative conditions for asymptotic consistency and Normality are given by Wu (1981). □

We conclude by relating our approach to that of Johnson (1970). We have given Laplace regularity conditions on the family of densities, and have derived asymptotic expectations and marginal densities. Our approach easily yields an asymptotic approximation to the distribution function as well: the first term gives asymptotic Normality, as in the work of LeCam (1953), Walker (1969), and Heyde and Johnstone (1979). Compare also the discussion of Chen (1985). Johnson (1970) provides additional conditions primarily to guarantee the requisite tail behavior of our regularity condition (iv); they are essentially consistency conditions for the MLE. We now present a result that relates the problem of checking condition (iv) to that of checking consistency of the MLE.

THEOREM 8: (Laplace regularity in the i.i.d. case.) *Suppose $\mathcal{P}^{(1)} = \{P_\theta^{(1)} : \theta \in \Theta\}$ is a family of probability measures dominated by a σ-finite measure on a measurable space $(\mathcal{Y},\mathcal{B})$ with $\mathcal{Y}$ in R^k, $\mathcal{B}$ the Borel sets in $\mathcal{Y}$ and Θ an open subset of R^m, and let $\{Y_i : i = 1,2,\ldots\}$ be a sequence of i.i.d. observations from $P_\theta^{(1)}$ with density $p(y \mid \theta)$. The family $\mathcal{P} = \mathcal{P}^{(\infty)}$ of infinite product measures on $(\mathcal{Y}^{(\infty)}, \mathcal{B}^{(\infty)})$ is Laplace-regular if the following conditions are satisfied:*

(i) $\theta \ne \theta'$ implies $P_\theta^{(1)} \ne P_{\theta'}^{(1)}$;

(ii) for all y and $\theta, p(y \mid \theta) > 0$ and for all y, $p(y \mid \cdot)$ is six times continuously differentiable;

(iii) *for all* θ_0, *there exist a neighborhood* $N_1(\theta_0)$ *of* θ_0, *a positive integer* r, *and a measurable real function* M_1 *on the r-fold product space* $(\mathcal{Y}^{(r)}, \mathcal{B}^{(r)})$ *such for all* $\theta \in N_1(\theta_0)$,

$$r^{-1}\sum_{i=1}^{r}[\log p(y_i \mid \theta) - \log p(y_i \mid \theta_0)] < M_1(y_1, \ldots, y_r)$$

and, under $P_{\theta_0}, EM_1(Y_1, \ldots, Y_r) < \infty$;

(iv) *for all* θ_0, *there exist a neighborhood* $N_2(\theta_0)$ *of* θ_0 *and a measurable real function* M_2 *on such that for all* $\theta \in N_2(\theta_0)$ *and all* $1 \leq j_1, \ldots, j_d \leq m$, *with* $d \leq 6$,

$$\mid \partial_{j_1} \cdots_{j_d} \log p(y \mid \theta) \mid < M_2(y)$$

and, under $P_{\theta_0}, EM_2(Y) < \infty$;

(v) *for all* $\theta_0, \det(D^2 f[\theta_0]) > 0$, *where* $f(\theta) = E[\log p(y \mid \theta_0) - \log p(y \mid \theta)]$ *with the expectation taken under* P_{θ_0} ;

(vi) *for all* θ_0 *every sequence of maxima of the loglikelihood function* ℓ_n *is strongly consistent.*

PROOF: This follows from the strong law of large numbers and Theorem 4.1 of Perlman (1972), together with a version of Huzurbazar's Theorem given in Proposition 4.2 in Perlman (1983), and the accompanying discussion on page 352 of Perlman (1983; the result used here being applicable in the multiparameter case). □

REMARKS:

(1) When the integral of $p(y \mid \theta)$ may be differentiated twice under the integral sign, condition (v) becomes positive-definiteness of the Fisher information matrix.

(2) Conditions for consistency of the MLE are discussed at length in Perlman (1972).

(3) It is straightforward to verify Laplace regularity for exponential and curved exponential families. See Kass and Fu (1988). Crawford (1988) has also verified Laplace regularity for certain mixture models.

Acknowledgments

This work was supported by the National Science Foundation under Grant Nos. NSF/DMS-850319 and NSF/DMS-8705646.

We thank Yves Thibaudeau, Arieh Epstein and the referees for their comments on the manuscript.

References

Achcar, J.A. and A.F.M. Smith (1989). "Aspects of Reparameterization in Approximate Bayesian Inference," (this volume).

Bickel, P.J., F. Götze, and W.R. van Zwet, (1985). "A Simple Analysis of Third-Order Efficiency of Estimates," in *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II, L. LeCam and R.A. Olshen, eds. Belmont, CA: Wadsworth.

Chen, C.F. (1985). "On Asymptotic Normality of Limiting Density Functions with Bayesian Implications," *J. Roy. Statist. Soc., B*, Vol. 47, 540-546.

Crawford, S. (1988). "An Approximate Bayesian Analysis of Finite Mixture Distributions," Ph.D. Thesis, Department of Statistics, Carnegie Mellon University.

Erdelyi, A. (1956). *Asymptotic Expansions*. New York: Dover Publications.

Ghosh, J.K., B.K. Sinha, and S.N. Joshi (1982). "Expansions for Posterior Probability and Integrated Bayes Risk," *Statistical Decision Theory and Related Topics, III*, Vol. 1, Shanti S. Gupta and James O. Berger, eds. New York: Academic Press, 403-456.

Heyde, C.C. and I.M. Johnstone (1979). "On Asymptotic Posterior Normality for Stochastic Processes," *J. Roy. Statist. Soc., B*, Vol. 41, 184-189.

Jennrich, R.I. (1969). "Asymptotic Properties of Nonlinear Least Squares Estimators," *Ann. Math. Statist.*, Vol. 40, 633-643.

Johnson, R.A. (1967). "An Asymptotic Expansion for Posterior Distributions," *Ann. Math. Statist.*, Vol. 38, 1899-1907.

Johnson, R.A. (1970). "Asymptotic Expansions Associated with Posterior Distributions," *Ann. Math. Statist.*, Vol. 41, 851-864.

Johnson, R.A. and J.N. Ladalla (1979). "The Large Sample Behavior of Posterior Distributions When Sampling from Multiparameter Exponential Family Models, and Allied Results," *Sankyha, B* Vol. 41, 196-215.

Joshi, S.N. (1984). "Expansion of Bayes Risk in the Case of Double Exponential Family," *Sankhya, A*, Vol. 46, 64-74.

Kass, R.E., and J.C. Fu (1988). "Posterior Large Deviations," Department of Statistics Technical Report, Carnegie Mellon University.

Kass, R.E., L. Tierney, and J.B. Kadane (1988). "Asymptotics in Bayesian Computation," in *Bayesian Statistics 3*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, eds. Oxford: Oxford University Press, 261-278.

Lai, T.L., H. Robbins, and C.Z. Wei (1978). "Strong Consistency of Least Squares Estimates in Multiple Regression," *Proc. Nat. Acad. Sci.* Vol. 75, 3034-3036.

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

LeCam, L. (1953). "On Some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes Estimates," *Univ. Calif. Publ. Statist.*, Vol. 1, 277-330.

Leonard, T. (1982). "Comment: Simple Predictive Density Function," *J. Amer. Statist. Assoc.*, Vol. 77, 657-658.

Lindley, D.V. (1961). "The Use of Prior Probability Distributions in Statistical Inference and Decisions," *Proc. 4th Berkeley Symp.* Vol. 1, 453-468.

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

Morris, C.N. (1988). "Approximating Posterior Distributions and Posterior Moments," in *Bayesian Statistics 3*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, eds. Oxford: Oxford University Press, 327-344.

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

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

Perlman, M.D. (1972). "On the Strong Consistency of Approximate Maximum Likelihood Estimators," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 263-282.

Perlman, M.D. (1983). "The Limiting Behavior of Multiple Roots of the Likelihood Equation," in *Recent Advances in Statistics: Papers in Honor of Herman Chernoff*, M.H. Rizvi, J. Rustagi, and D. Siegmund, eds. New York: Academic Press.

Smith, A.F.M. A.M. Skene, J.E.H. Shaw, J.C. Naylor, and M. Dransfield (1984). "The Implementation of the Bayesian Paradigm," *Commun. Statist. Theor. Meth.*, Vol. 14, No. 5, 1079-1102.

Tierney, L. and J.B. Kadane (1986). "Accurate Approximations for Posterior Moments and Marginal Densities," *J. Amer. Statist. Assoc.*, Vol. 81, 82-86.

Tierney, L., R.E. Kass, and J.B. Kadane (1989a). "Approximate Marginal Densities of Nonlinear Functions," *Biometrika*, to appear.

Tierney, L., R.E. Kass, and J.B. Kadane (1989b). "Fully Exponential Laplace Approximations to Expectations and Variances of Non-Positive Functions," *J. Amer. Statist. Assoc.*, to appear.

Walker, A.M. (1969). "On the Asymptotic Behavior of Posterior Distributions," *J. Roy. Statist. Soc., B*, Vol. 31, 80-88.

Wu, C.F. (1981). "Asymptotic Theory of Nonlinear Least Squares Estimation," *Ann. Statist.*, Vol. 9, 501-513.