Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions

LUKE TIERNEY, ROBERT E. KASS, and JOSEPH B. KADANE*

Tierney and Kadane (1986) presented a simple second-order approximation for posterior expectations of positive functions. They used Laplace's method for asymptotic evaluation of integrals, in which the integrand is written as $f(\theta)\exp(-nh(\theta))$ and the function h is approximated by a quadratic. The form in which they applied Laplace's method, however, was *fully exponential*: The integrand was written instead as $\exp[-nh(\theta) + \log f(\theta)]$; this allowed first-order approximations to be used in the numerator and denominator of a ratio of integrals to produce a second-order expansion for the ratio. Other second-order expansions (Hartigan 1965; Johnson 1970; Lindley 1961, 1980; Mosteller and Wallace 1964) require computation of more derivatives of the log-likelihood function. In this article we extend the fully exponential method to apply to expectations and variances of nonpositive functions. To obtain a second-order approximation to an expectation $E(g(\theta))$, we use the fully exponential method to approximate the moment-generating function $E(\exp(sg(\theta)))$, whose integrand is positive, and then differentiate the result. This method is formally equivalent to that of Lindley and that of Mosteller and Wallace, yet does not require third derivatives of the likelihood function. It is also equivalent to another alternative approach to the approximation of $E(g(\theta))$: We may add a large constant c to $g(\theta)$, apply the fully exponential method to $E(c + g(\theta))$, and subtract c; on passing to the limit as c tends to infinity we regain the approximation based on the moment-generating function. Furthermore, the second derivative of the logarithm of the approximation $E(\exp(sg(\theta)))$, which is an approximate cumulant-generating function, yields a simple second-order approximation to the variance. In deriving these results we omit rigorous justification of formal manipulations, which may be found in Kass, Tierney, and Kadane (in press). Although our point of view is Bayesian, our results have applications to non-Bayesian inference as well (DiCiccio 1986).

1. INTRODUCTION

A simple and remarkable method of asymptotic expansion of integrals generally attributed to Laplace (Laplace 1986; also see Stigler 1986) is widely used in applied mathematics. This method has been applied by many authors (DiCiccio 1986; Efron and Hinkley 1978; Hartigan 1965; Johnson 1970; Lindley 1961, 1980; Mosteller and Wallace 1964; Tierney and Kadane 1986) to find approximations to the ratios of integrals of interest, especially in Bayesian analysis. Tierney and Kadane (1986) applied the Laplace method in a special form, which we call fully exponential, 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 real function g of the vector parameter θ, but it has the disadvantage of requiring g to be positive. The purpose of this article is to extend the fully exponential method to expectations of arbitrary functions while retaining its advantageous simplicity.

Section 2 reviews Laplace's method and its application to ratios of integrals. In Section 3 we show that applying the fully exponential method to the moment-generating function, and then taking derivatives, leads to a second-order approximation to the expectation of a nonpositive function that is reasonably simple to compute. We also discuss the relationship of this approximation to others in the literature, and show its equivalence to another possible technique: adding a large constant, approximating the expectation of the modified function, and then subtracting the constant. In Section 4 we illustrate the method with some special cases and an example. Although our point of view is Bayesian, our results are formal and have applications to non-Bayesian inference as well. We do not give a careful treatment of precise conditions under which the expansions are valid. For such treatments, see Johnson (1967, 1970), Johnson and Ladalla (1979), and Kass, Tierney, and Kadane (in press). For a variation of Laplace's method using non-Normal exponential family kernels, see Morris (1988).

2. LAPLACE APPROXIMATIONS TO RATIOS OF INTEGRALS

The purpose of this section is to give a general form for the application of Laplace's method to the ratios of integrals, displaying both what we shall call the standard form (Lindley 1961, 1980; Mosteller and Wallace 1964) and the fully exponential form (Tierney and Kadane 1986) as special cases. For convenience of exposition, we assume that the parameter θ is one-dimensional.

The posterior expectation of a function g may be written as

$$E(g(\theta)) = \frac{\int g(\theta)L(\theta)\pi(\theta)\,d\theta}{\int L(\theta)\pi(\theta)\,d\theta}, \qquad (2.1)$$

where $\pi(\theta)$ is the prior and $L(\theta)$ is the likelihood, suppressing dependency on the data from the notation. We

* Luke Tierney is Associate Professor, School of Statistics, University of Minnesota, Minneapolis, MN 55455. Robert E. Kass is Associate Professor and Joseph B. Kadane is Leonard J. Savage Professor of Statistics and Social Sciences, Department of Statistics, Carnegie-Mellon University, Pittsburgh, PA 15213. The authors' research was supported by National Science Foundation Grants DMS-8503019 and DMS-8705646. The authors thank Elizabeth Slate, the referees, and the associate editor for unusually helpful comments.

Journal of the American Statistical Association
September 1989, Vol. 84, No. 407, Theory and Methods

find it convenient to reexpress the integrands in (2.1) so that the expectation takes the form

$$E(g(\theta)) = \frac{\int b_N(\theta)\exp\{-nh_N(\theta)\}\,d\theta}{\int b_D(\theta)\exp\{-nh_D(\theta)\}\,d\theta}, \qquad (2.2)$$

where $b_N(\theta)\exp\{-nh_N(\theta)\} = g(\theta)L(\theta)\pi(\theta)$ and $b_D(\theta)\exp\{-nh_D(\theta)\} = L(\theta)\pi(\theta)$. We assume that $b_N(\cdot)$ and $b_D(\cdot)$ do not depend on n and that $h_N(\theta)$ and $h_D(\theta)$ are constant-order functions of n, as $n \to \infty$. (In statistical applications, n is typically the sample size.) One example, considered next, consists of the choices $h_N(\theta) = h_D(\theta) = (1/n)\log\{L(\theta)\pi(\theta)\}$, $b_N(\theta) = g(\theta)$, and $b_D(\theta) = 1$. More generally, whenever $h_D(\theta) = h_N(\theta)$ we say that (2.2) is in *standard form*; when $b_N(\theta) = b_D(\theta)$ [in which case $h_N(\theta) \neq h_D(\theta)$] we say that (2.2) is in *fully exponential form*.

The basic Laplace approximation to integrals is

$$\int b(\theta)\exp\{-nh(\theta)\}\,d\theta = \sqrt{2\pi}\sigma \cdot \exp\{-n\hat{h}\}\left[\hat{b} + \frac{1}{2n}\left\{\sigma^2\hat{b}'' - \sigma^4\hat{b}'\hat{h}''' + \frac{5}{12}\hat{b}(\hat{h}''')^2\sigma^6 - \frac{1}{4}\hat{b}\hat{h}^{iv}\sigma^4\right\}\right] + O(n^{-2}), \qquad (2.3)$$

where hats on b, h, and their derivatives indicate evaluation at $\hat{\theta}$, the maximum of $-h$, and $\sigma^2 = [h''(\hat{\theta})]^{-1}$ (see Erdelyi 1956, pp. 36–39). Applying (2.3) to (2.2),

$$\frac{\int b_N \exp\{-nh_N\}\,d\theta}{\int b_D \exp\{-nh_D\}\,d\theta} = \frac{A(N)}{A(D)} + O(n^{-2}),$$

where

$$A(K) = \sigma_K\exp\{-n\hat{h}_K\}\left[\hat{b}_K + \frac{1}{2n}\left\{\sigma_K^2\hat{b}_K'' - \hat{h}_K'''\hat{b}_K'\sigma_K^4 + \frac{5}{12}\hat{b}_K(\hat{h}_K''')^2\sigma_K^6 - \frac{1}{4}\hat{b}_K\sigma_K^4\hat{h}_K^{iv}\right\}\right]$$

with $K = N, D$. Here the hats on b_K, h_K, and their derivatives indicate evaluation at $\hat{\theta}_K$, the maximum of $-h_K$; for example, $\hat{b}_K = b_K(\hat{\theta}_K)$ and $\sigma_K^2 = [h_K''(\hat{\theta}_K)]^{-1}$ for $K = N, D$. Assume that $\hat{b}_D \neq 0$. Then

$$\frac{\int b_N\exp\{-nh_N\}\,d\theta}{\int b_D\exp\{-nh_D\}\,d\theta} = \frac{\sigma_N\exp\{-n\hat{h}_N\}}{\sigma_D\exp\{-n\hat{h}_D\}}\left[\frac{\hat{b}_N}{\hat{b}_D} + \frac{1}{2n}\left(\frac{\sigma_N^2\hat{b}_D\hat{b}_N'' - \sigma_D^2\hat{b}_N\hat{b}_D''}{\hat{b}_D^2} - \frac{\sigma_N^4\hat{h}_N'''\hat{b}_D\hat{b}_N' - \sigma_D^4\hat{h}_D'''\hat{b}_N\hat{b}_D'}{\hat{b}_D^2} + \left(\frac{5}{12}\right)\frac{\sigma_N^6(\hat{h}_N''')^2\hat{b}_N - \sigma_D^6(\hat{h}_D''')^2\hat{b}_N}{\hat{b}_D} - \left(\frac{1}{4}\right)\frac{\sigma_N^4\hat{h}_N^{iv}\hat{b}_N - \sigma_D^4\hat{h}_D^{iv}\hat{b}_N}{\hat{b}_D}\right)\right] + O(n^{-2}). \qquad (2.4)$$

Now assume that the ith derivatives of h_K satisfy $\hat{h}_N^{(i)} - \hat{h}_D^{(i)} = O(n^{-1})$ for $i = 0, \ldots, 4$ (the 0th derivative of h_K at $\hat{\theta}_K$ being $\hat{h}_K$). This is the case for both the standard and fully exponential forms. We then have $\sigma_N = \sigma_D + O(n^{-1})$, and the last two terms in the order $O(n^{-1})$ part of Expression (2.4) are actually of order $O(n^{-2})$. Furthermore, σ_N and $\hat{h}_N^{(i)}$ in the other two terms multiplied by $1/2n$ can be replaced by σ_D and $\hat{h}_D^{(i)}$, to the same order of accuracy. Thus

$$\frac{\int b_N\exp\{-nh_N\}\,d\theta}{\int b_D\exp\{-nh_D\}\,d\theta} = \frac{\sigma_N\exp\{-n\hat{h}_N\}}{\sigma_D\exp\{-n\hat{h}_D\}} \times \left[\frac{\hat{b}_N}{\hat{b}_D} + \left(\sigma_D^2\frac{\hat{b}_D\hat{b}_N'' - \hat{b}_N\hat{b}_D''}{2n\hat{b}_D^2} - \sigma_D^4\hat{h}_D'''\frac{\hat{b}_D\hat{b}_N' - \hat{b}_N\hat{b}_D'}{2n\hat{b}_D^2}\right)\right] + O(n^{-2}). \qquad (2.5)$$

Equation (2.5) has been applied in the literature in both the standard form (Lindley 1961, 1980; Mosteller and Wallace 1964) and the fully exponential form (Tierney and Kadane 1986). For the standard form, $h_N = h_D$ implies that $b_N = b_D g$; then,

$$\frac{\hat{b}_D\hat{b}_N' - \hat{b}_N\hat{b}_D'}{2n\hat{b}_D^2} = \frac{\hat{b}_D(\hat{b}_D'\hat{g} + \hat{b}_D\hat{g}') - \hat{b}_D\hat{g}\hat{b}_D'}{2n\hat{b}_D^2} = \frac{\hat{g}'}{2n}$$

and

$$\frac{\hat{b}_D\hat{b}_N'' - \hat{b}_N\hat{b}_D''}{2n\hat{b}_D^2} = \frac{\hat{b}_D(\hat{b}_D''\hat{g} + 2\hat{b}_D'\hat{g}' + \hat{b}_D\hat{g}'') - \hat{b}_D\hat{g}\hat{b}_D''}{2n\hat{b}_D^2} = \frac{\hat{b}_D'\hat{g}'}{n\hat{b}_D} + \frac{\hat{g}''}{2n},$$

where $\hat{g} = g(\hat{\theta})$ and $\hat{\theta}$ is the maximum of $-h_N = -h_D$. Both terms are $O(n^{-1})$ and have to be evaluated in obtaining the $O(n^{-2})$ approximation,

$$\frac{\int b_N\exp\{-nh_N\}}{\int b_D\exp\{-nh_D\}} = \hat{g} + \frac{\sigma_D^2\hat{b}_D'\hat{g}'}{n\hat{b}_D} + \frac{\sigma_D^2\hat{g}''}{2n} - \frac{\sigma_D^4\hat{h}_D'''\hat{g}'}{2n} + O(n^{-2}). \qquad (2.6)$$

For the fully exponential form, $b_N = b_D$, which implies that $h_N = h_D + (1/n)\log g$, and we now must assume that g is positive. Tierney and Kadane (1986) treated the case $b_N = b_D = 1$. In the fully exponential form, $b_N(\cdot) = b_D(\cdot)$, but $\hat{b}_N = b_N(\hat{\theta}_N)$ is not necessarily equal to $\hat{b}_D = b_D(\hat{\theta}_D)$. If the sequence of values $g(\hat{\theta}_D)$ is bounded away from 0, then $\hat{\theta}_N - \hat{\theta}_D = O(n^{-1})$ and hence the ith derivatives of b_K satisfy $\hat{b}_N^{(i)} - \hat{b}_D^{(i)} = O(n^{-1})$ for $i = 0, 1, 2$. Thus

$$\frac{\hat{b}_N\hat{b}_D' - \hat{b}_D\hat{b}_N'}{2n\hat{b}_D^2} = O(n^{-2}), \qquad \frac{\hat{b}_N\hat{b}_D'' - \hat{b}_D\hat{b}_N''}{2n\hat{b}_D^2} = O(n^{-2});$$

therefore, both terms can be ignored to give the $O(n^{-2})$

approximation

$$\frac{\int b_N \exp\{-nh_N\}}{\int b_D \exp\{-nh_D\}} = \frac{\hat{b}_N \sigma_N \exp\{-n\hat{h}_N\}}{\hat{b}_D \sigma_D \exp\{-n\hat{h}_D\}} + O(n^{-2}). \quad (2.7)$$

Both (2.6) and (2.7) have analogous multivariate versions, which are given in the references.

One might think that the restriction that g be positive in the fully exponential method would be unimportant, since if g were negative then $-g$ would be positive, so the fully exponential method would apply. Then a general g could be written as

$$g = g^+ - g^-, \quad (2.8)$$

where $g^+ = \max(g, 0)$ and $g^- = -\min(g, 0)$, the fully exponential method might be applied to g^+ and g^+, and the results could be combined using (2.8). Neither g^+ nor g^- is bounded away from 0, however, so $\log g^+$ and $\log g^-$ are undefined at some points in the parameter space. Consequently, although the decomposition (2.8) is natural and useful in the measure-theoretic treatment of integration, it is neither for the problem here, and application of the fully exponential method using (2.8) will produce poor approximations in practice. The next section takes up more promising methods for the extension of the fully exponential method to the expectation of functions that are unrestricted in sign.

3. EXPECTATIONS AND VARIANCES OF GENERAL FUNCTIONS

The fully exponential approximation is suited only for functions g bounded away from 0. In this section we consider the following three methods of approximating expectations of general functions:

1. Approximating the moment-generating function (MGF), which is an expectation of a positive function, using the fully exponential form, and differentiating the result.
2. Adding a large constant to g, using the fully exponential approximation on the sum, and then subtracting the constant. The limit of such approximations as the constant tends to infinity is the second method.
3. The standard-form approximation (2.6).

We show that these three methods give identical approximations and that natural extensions of the first two to the approximation of variances are also identical. We also establish the appropriate order of the approximations.

3.1 Approximation Using the Moment-Generating Function

Since $\exp\{sg(\theta)\}$ is always positive, $M(s) = E(\exp\{sg(\theta)\})$ may be approximated according to (2.7), with $b_N = b_D = b > 0$, $-h_D = -h = (l(\theta) + \log\{\pi(\theta)\} - \log\{b(\theta)\})/n$, and $h_N = h_D - sg(\theta)/n$, where $l(\theta)$ is the log-likelihood and $\pi(\theta)$ is the prior. We denote the resulting approximation by $\hat{M}(s)$. The symbol ^ here, and in Equations (3.1) and (3.2), indicates approximation rather than evaluation at a maximum, as in Section 2. Since interest centers on the mean and variance of g, we find it convenient to use derivatives of the approximate cumulant-generating function, $\log \hat{M}(s)$. Thus, let

$$\hat{E}(g) = (d/ds)\log \hat{M}(s)\,|_{s=0} \quad (3.1)$$

and

$$\hat{V}(g) = (d^2/ds^2)\log \hat{M}(s)\,|_{s=0}. \quad (3.2)$$

Since $\hat{M}(0) = 1$, we have $\hat{E}(g) = \hat{M}'(0)$ and $\hat{V}(g) = \hat{M}''(0) - (\hat{M}'(0))^2$. The following theorem establishes the order of approximation for (3.1) and (3.2).

Theorem 1. (a) $E(g) = \hat{E}(g) + O(n^{-2})$. (b) $V(g) = \hat{V}(g) + O(n^{-3})$.

Proof. By analogy with (A.2) in the appendix of Tierney and Kadane (1986), we can write

$$M(s) = \hat{M}(s)(1 + sc_n/n^2 + O_s(n^{-3})), \quad (3.3)$$

where O_s indicates that the error term may depend on s, and where the c_n are bounded and do not depend on s. Since the numerator integrand of $M(s)$ depends smoothly on s, the error term and its first and second derivatives are uniformly $O(n^{-3})$ for s in a neighborhood of the origin. Thus we can replace $O_s(n^{-3})$ in (3.3) with $e_n(s) \cdot n^{-3}$, where the $e_n(s)$ are uniformly bounded and have uniformly bounded first and second derivatives for s near 0. Making this substitution, taking logarithms, and differentiating both sides of (3.3) with respect to s we obtain

$$\frac{M'(s)}{M(s)} = \frac{\hat{M}'(s)}{\hat{M}(s)} + \frac{c_n/n^2 + e_n'(s)/n^3}{1 + sc_n/n^2 + e_n(s)/n^3}. \quad (3.4)$$

Evaluating (3.4) at $s = 0$ and using the fact that $M(0) = \hat{M}(0) = 1$ gives

$$E(g) = \hat{E}(g) + \frac{c_n}{n^2} + \frac{e_n'(0)}{n^3} = \hat{E}(g) + O(n^{-2}),$$

which is part (a). Differentiating (3.4) once more yields

$$\frac{M(s)M''(s) - M'(s)^2}{M(s)^2} = \frac{\hat{M}(s)\hat{M}''(s) - \hat{M}(s)^2}{\hat{M}(s)^2}$$
$$+ \frac{(e_n''(s)/n^3)(1 + sc_n/n^2 + e_n(s)/n^3) - (c_n/n^2 + e_n'(s)/n^3)^2}{(1 + sc_n/n^2 + e_n(s)/n^3)^2}.$$

By evaluating both sides at $s = 0$ we obtain $V(g) = \hat{V}(g) + O(n^{-3})$. This establishes part (b) and completes the proof.

As the following theorem shows, (3.1) and (3.2) can be rewritten in a suggestive form: Let $\hat{\theta}$ be the maximum of $-h$ and $\sigma = [h''(\hat{\theta})]^{-1/2}$. Let $-nh_s(\theta) = -nh(\theta) + sg(\theta)$, and let θ_s be the maximum of $-h_s$ and $\sigma_s = [h_s''(\theta_s)]^{-1/2}$.

Theorem 2. (a) $\hat{E}(g) = g(\hat{\theta}) + (d/ds)\log \sigma_s\,|_{s=0} + (d/ds)\log b(\theta_s)\,|_{s=0}$. (b) $\hat{V}(g) = [\sigma g'(\hat{\theta})]^2/n + (d^2/ds^2)\log \sigma_s\,|_{s=0} + (d^2/ds^2)\log b(\theta_s)\,|_{s=0}$.

Proof. Suppose that $f(\theta, s)$ is an arbitrary smooth function. Let θ_s minimize $f(\cdot, s)$. Then

$$\partial f/\partial\theta \,|_{(\theta_s,s)} = 0. \tag{3.5}$$

Hence

$$\frac{d}{ds}\min_\theta f(\theta, s) = \frac{d}{ds} f(\theta_s, s) = \left.\frac{\partial f}{\partial s}\right|_{(\theta_s,s)}. \tag{3.6}$$

Since

$$\hat{M}(s) = \frac{b(\theta_s)\sigma_s\exp\{-nh_s(\theta_s)\}}{b(\hat\theta)\sigma\exp\{-nh(\hat\theta)\}},$$

we apply (3.6), where $f(\theta, s) = nh_s(\theta) = nh(\theta) - sg(\theta)$, to yield

$$\frac{d}{ds}\log \hat{M}(s) = g(\theta_s) + \frac{d}{ds}[\log\sigma_s + \log b(\theta_s)].$$

Part (a) follows immediately, and

$$\frac{d^2\log\hat{M}(s)}{ds^2} = g'(\theta_s)\frac{d\theta_s}{ds} + \frac{d^2}{ds^2}\log\sigma_s + \frac{d^2}{ds^2}\log b(\theta_s). \tag{3.7}$$

Now, using (3.5),

$$0 = \left.\frac{\partial f}{\partial\theta}\right|_{(\theta_s,s)} = \frac{\partial}{\partial\theta}[nh(\theta) - sg(\theta)]\,|_{(\theta_s,s)} = nh'(\theta_s) - sg'(\theta_s).$$

Differentiation of this equation at $s = 0$ yields $0 = nh''(\hat\theta)(d\theta_s/ds)|_{s=0} - g'(\hat\theta)$. Hence

$$d\theta_s/ds\,|_{s=0} = g'(\hat\theta)\sigma^2/n, \tag{3.8}$$

and substitution into (3.7) yields part (b).

Although the theorem has been proved here only for univariate θ, the same proof works for vector-valued θ. In the multivariate statement of the theorem, σ_s would be replaced by $\det(\Sigma_s)^{1/2}$, where $\Sigma_s = [D^2h_s\,|_{\theta_s}]^{-1}$, and $[\sigma g'(\hat\theta)]^2$ would be replaced by $(Dg)^T\Sigma(Dg)$, where Dg is the derivative of g evaluated at $\hat\theta$ and $\Sigma = [D^2h\,|_{\hat\theta}]^{-1}$.

Both the MGF method of (3.1) and the standard method of (2.6) give second-order approximations. The next theorem, however, shows that for expectations the approximations are arithmetically identical.

Theorem 3. Using the definitions of g, σ_D, and $\hat{h}_D$ from the first paragraph of this section, and $\hat{g} = g(\hat\theta_D)$, we have

$$\hat{E}(g) = \hat{g} + \frac{\sigma_D^2\hat{g}''}{2n} - \frac{\sigma_D^4\hat{h}_D'''\hat{g}'}{2n} + \frac{\sigma^2\hat{b}_D'\hat{g}'}{n\hat{b}_D}.$$

Proof. From Theorem 2(a),

$$\hat{E}(g) = g(\hat\theta) + \frac{d}{ds}\log\sigma_s\,|_{s=0} + \frac{d}{ds}\log b(\theta_s)|_{s=0}.$$

Now

$$\log\sigma_s = -\tfrac{1}{2}\log[h_s''(\theta_s)] = -\tfrac{1}{2}\log\{h''(\theta_s) - (s/n)g''(\theta_s)\}.$$

Then, using (3.8),

$$\frac{d}{ds}\log b(\theta_s) = \frac{b_D'}{b_D}\frac{d\theta_s}{ds} = \frac{b_D'}{b_D}\frac{\sigma^2\hat{g}'}{n}$$

and

$$d\log\sigma_s/ds\,|_{s=0}$$

$$= -\frac{1}{2}\left.\frac{h'''(\theta_s)\dfrac{d\theta_s}{ds} - g''(\theta_s)/n - (s/n)g'''(\theta_s)\dfrac{d\theta_s}{ds}}{h''(\theta_s) - (s/n)g''(\theta_s)}\right|_{s=0}$$

$$= -\frac{\sigma^2}{2}[h'''(\theta_s)g'(\hat\theta)\sigma^2/n - g''(\theta_s)/n]\,|_{s=0}$$

$$= \frac{\sigma^2\hat{g}''}{2n} - \frac{\sigma^4}{2n}\hat{h}'''\hat{g}'.$$

Note that Theorem 3 furnishes an alternate proof of Theorem 1(a).

3.2 Addition of a Large Constant

The second method listed at the beginning of this section is to add a large constant to g, use the fully exponential method on the sum, and then subtract the constant; the approximation results from letting the constant become infinite. We now show that this approximation is the same (for both means and variances) as that resulting from the fully exponential method applied to the MGF. We begin with a heuristic argument.

Since the fully exponential method is based on maximization of $-nh + \log(g)$, it is unaffected by multiplication of g by a constant. Thus we get the same result by approximating the expectation of $1 + g/c$ and then multiplying by c, as we do by approximating $g + c$. For large c, meanwhile, $1 + g/c$ is roughly equal to $\exp(g/c)$, the expectation of which is $M(1/c)$. Thus, for large c, the fully exponential approximation to the expectation of $1 + g/c$ should be roughly equal to the fully exponential approximation $\hat{M}(1/c)$ to $M(1/c)$. If we now multiply $\hat{M}(1/c)$ by c and pass to the limit as c becomes infinite, we arrive once again at the derivative at 0 of the MGF approximation. That is, from this brief argument it would appear that adding a large constant to g and then applying the fully exponential approximation would produce essentially the same result as Approximation (3.1). One way to verify this equivalence is to compute the derivative at $s = 0$ of the approximation to $E(1 + sg(\theta))$, given in Equation (2.7). After some manipulation along the lines of the argument leading to Theorem 2(a) one obtains (3.1). We present here, instead, a simpler derivation, which is effective in deriving the variance approximation equivalence as well.

We let $\tilde{E}$ and $\tilde{V}$ denote approximate fully exponential expectation and variance operators based on (2.7) (with $b_N = b_D$), where $\tilde{V}(g) = \tilde{E}(g^2) - (\tilde{E}(g))^2$, as in Tierney and Kadane (1986); $\hat{E}$ and $\hat{V}$ will be the operators of (3.1) and (3.2). We define $\tilde{E}_c(g) = \tilde{E}(c + g) - c$ and $\tilde{V}_c(g) = \tilde{E}\{(c + g)^2\} - \{\tilde{E}_c(g)\}^2$.

Theorem 4. (a) $\lim_{c\to\infty} \tilde{E}_c(g) = \hat{E}(g)$. (b) $\lim_{c\to\infty} \tilde{V}_c(g) = \hat{V}(g)$.

Proof. Let $f(x, y) = \tilde{E}\{[1 + yg(\theta)]^{x/y}\}$, where the convention $z^{x/y} = 1$ for $z \le 0$ is employed. Also let $f(x, 0) = \lim_{y\downarrow 0} f(x, y)$. Then $\hat{M}(s) = f(s, 0)$, so $\hat{E}(g) = \partial f/\partial x\,|_{(0,0)}$ and $\hat{V}(g) = [\partial^2 f/\partial x^2 - (\partial f/\partial x)^2]\,|_{(0,0)}$. Also,

$$\tilde{E}_c(g) = \tilde{E}(c + g) - c$$
$$= c\tilde{E}(1 + g/c) - c = cf\left(\frac{1}{c}, \frac{1}{c}\right) - c.$$

Then, since $f(0, 0) = 1$,

$$\lim_{c\to\infty} \tilde{E}_c(g) = [\partial f/\partial x + \partial f/\partial y]\,|_{(0,0)},$$

and, since $f(0, y) = 1$, $\partial f/\partial y\,|_{(0,0)} = 0$ and part (a) follows. Similarly,

$$\tilde{V}_c(g) = \tilde{E}(c + g)^2 - [\tilde{E}(c + g)]^2$$
$$= c^2\left[f\left(\frac{2}{c}, \frac{1}{c}\right) - f\left(\frac{1}{c}, \frac{1}{c}\right)^2\right]$$
$$= c^2\left[f\left(\frac{2}{c}, \frac{1}{c}\right) - 2f\left(\frac{1}{c}, \frac{1}{c}\right) + 1 - (f(1/c, 1/c) - 1)^2\right],$$

so

$$\lim_{c\to\infty} \tilde{V}_c(g) = [\partial^2 f/\partial x^2 - (\partial f/\partial x)^2]\,|_{(0,0)} = \hat{V}(g),$$

proving part (b).

4. EXAMPLES

In this section we explore the performance of the MGF approximation for several examples. We begin with four cases in which the approximations and the correct means and standard deviations can be evaluated explicitly: one-dimensional Normal, gamma, t, and beta distributions.

For the Normal and gamma distributions the results are quite simple: If the posterior density is a Normal density, the approximate MGF $\hat{M}(s)$ is equal to the exact MGF $M(s)$ for all s. If the posterior density is a gamma density proportional to $\theta^{\alpha-1}\exp\{-\lambda\theta\}$, then the approximate MGF $\hat{M}(s)$ is equal to the exact MGF for $s < \lambda$ and is undefined for $s > \lambda$, when the exact MGF is infinite. Thus, in both cases, the approximate mean and variance are exact. It remains to be determined whether there are other univariate distributions for which the Laplace approximation to the MGF is exact; the methods of Daniels (1980) may prove useful to this end.

Next, suppose the posterior density is Student's t on k degrees of freedom, proportional to $(1 + x^2/k)^{-(k+1)/2}$. Here the degrees of freedom k play the role of n in the asymptotics, and we did the computations with $b_N \equiv b_D \equiv 1$ and used the MGF approach. We write the random variable as X, with a typical value of x, in place of θ. The standardized t case is important because the MGF approximation is location and scale equivariant; that is, $\hat{E}(aX + b) = a\hat{E}(X) + b$. The MGF approximation is, of course, not directly applicable, since the MGF $M(s)$ is infinite for all $s \ne 0$; however, the MGF of a t variable truncated, say, at $\pm 1{,}000\, k^{1/2}$ does exist, and for $k > 2$ the truncated mean and variance differ from the untruncated mean and variance by an error that is of exponentially decreasing order as k tends to ∞. Thus the approximations $\hat{E}(X)$ and $\hat{V}(X)$ obtained by (3.1) and (3.2) can be justified as approximations based on the truncated variable. The resulting approximate mean is 0, which is the correct value for $k > 1$. The approximate variance is

$$\hat{V}(X) = \frac{k}{k+1}\left(1 + \frac{3}{k+1}\right),$$

with the correct value being $k/(k - 2)$ for $k > 2$. Thus $\hat{V}(X) = V(X) - 7k/(k + 1)^2(k - 2)$. The first-order approximation to the variance, $k/(k + 1)$, is always below 1, whereas the correct and MGF approximate variances are above 1. Detailed calculations are available from us. Figure 1 shows a comparison of the correct, first-order, and MGF approximate variances.

For a beta posterior distribution with density proportional to $x^{\alpha-1}(1 - x)^{\beta-1}$, the exact and approximate means and variances are

$$E(X) = \frac{\alpha}{\alpha + \beta}, \qquad \hat{E}(X) = \frac{\alpha^2 + \alpha\beta + 2 - 4\alpha}{(\alpha + \beta - 2)^2}$$

and

$$V(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)},$$
$$\hat{V}(X) = \frac{\alpha\beta^2 + \alpha^2\beta - 9\alpha\beta + 6\alpha + 6\beta - 5}{(\alpha + \beta - 2)^4}.$$

Again, calculations are available from us. For a fixed value of $\alpha/(\alpha + \beta) \in (0, 1)$ both relative errors are $O((\alpha + \beta)^{-2})$. Since the beta random variable is a positive random variable, the fully exponential Laplace approximation (2.5) may be applied directly, giving

$$\tilde{E}(X) = \frac{\alpha}{\alpha + \beta - 1}\left(\frac{\alpha}{\alpha - 1}\right)^{\alpha - 1/2}\left(\frac{\alpha + \beta - 2}{\alpha + \beta - 1}\right)^{\alpha+\beta-1/2}.$$

To compare the errors in these two approximations set $p = \alpha/(\alpha + \beta)$ and $n = \alpha + \beta$. Then for fixed p the error

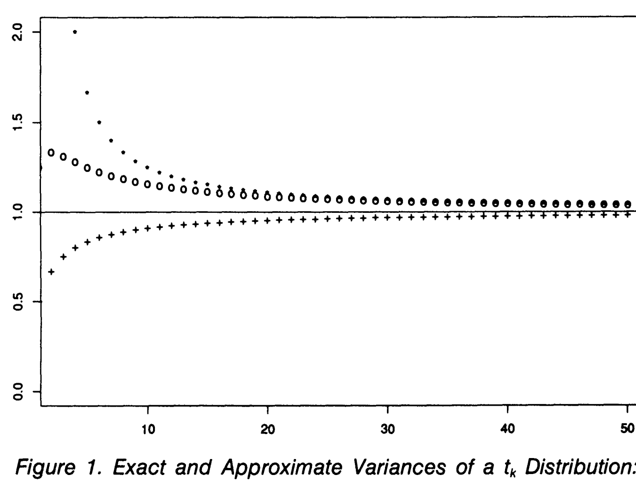

*Figure 1. Exact and Approximate Variances of a t_k Distribution: *, Exact Variance; +, First-Order Approximation; 0, MGF Approximation. The horizontal axis is k, for k > 2.*

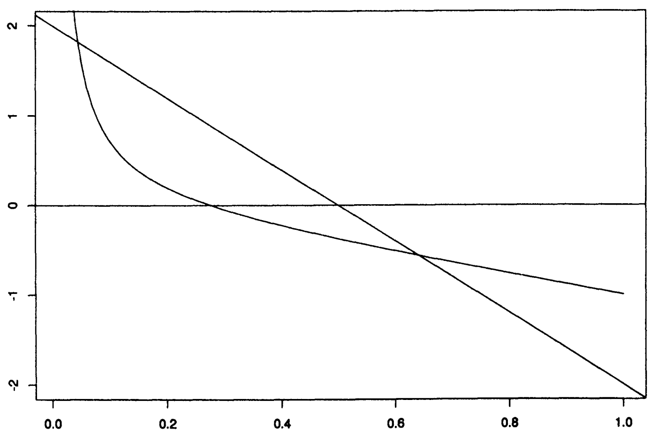

Figure 2. Leading Terms of Error in Approximating the Mean of a Beta(np, n(1 − p)) Distribution. The horizontal axis is p, the straight line is the MGF approximation, and the curved line is the fully exponential approximation.

in the MGF approximation is

$$\hat{E}(X) - E(X) = 2(1 - 2p)/n^2 + O(n^{-3}),$$

whereas the error in the direct fully exponential approximation is

$$\tilde{E}(X) - E(X) = -\frac{13p^2 - 1}{12pn^2} + O(n^{-3}).$$

Figure 2 shows a comparison of the two leading error terms multiplied by n^2. It would appear that, except for values of p that are very close to 0 or $\frac{1}{2}$, the direct fully exponential approximation is superior to the MGF approximation. This phenomenon has been observed in all examples considered so far. In many of these cases, however, the performance of the MGF approximations and the direct approximations can be improved by transforming to a parameterization in which the posterior may be better approximated by a Normal distribution. For the beta distribution, the MGF approximation and the first-order approximation to the posterior mean $E(X)$ are exact when the transformed parameter $y = \log(x/(1 - x))$ is used. The MGF approximation to the posterior variance $V(X)$ has a relative error of $1/(\alpha + \beta)^2$.

To illustrate these methods in a data-analytic context, we turn now to the example considered in Tierney and Kadane (1986). This example is based on a Pareto model proposed by Turnbull, Brown, and Hu (1974) for data from the Stanford heart-transplant experiment. Each patient has an exponential residual lifetime after entering the study. The hazard rates vary from patient to patient according to a gamma distribution with parameters p and λ. The effect of a transplant, if and when it occurs, is to multiply the patient's hazard rate by a factor τ. This produces the three-parameter likelihood

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

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

Following Naylor and Smith (1982) we compute a posterior distribution for τ, λ, and p based on a flat prior proportional to $d\lambda\, d\tau\, dp$. Table 1 lists the exact posterior means and standard deviations as computed by Gauss–Hermite quadrature, the direct fully exponential approximation using (2.7), the MGF approximation based on the original parameterization in terms of τ, λ, and p, and the first-order approximations using the posterior mode. The MGF approximation was computed using symmetric difference quotients with s equal to $\pm.01$ times the first-order approximate standard deviations and two Newton steps in the maximization; varying these choices had very little effect on the results. As can be seen from Table 1, the MGF approximation represents a considerable improvement over the first-order approximations; however, with errors on the order of 10% to 12%, they do not perform as well as the direct fully exponential approximations that produced errors on the order of 4%. On the other hand, if the MGF and direct approximations are applied to the parameterization ($\log \tau$, $\log \lambda$, $\log p$), then (as shown by the remaining rows of Table 1) both methods produce results that differ from the exact answers by only 3% to 4%.

5. CONCLUSION

We have extended the fully exponential approximation of Tierney and Kadane (1986) to nonpositive functions with the MGF method (3.1) and (3.2), and we have noted the arithmetic equivalence of (3.1) and the "standard" second-order expectation approximation (2.6). Although identical in their exact representations, (3.1) and (2.6)

Table 1. Approximate Moments for the Pareto Model

Method	Parameterization	Mean			Standard deviation		
		τ	λ	p	τ	λ	p
Exact	—	1.04	32.5	.50	.47	16.2	.14
Direct	(τ, λ, p)	1.044	32.11	.493	.494	16.09	.138
MGF	(τ, λ, p)	1.007	30.21	.487	.437	13.97	.131
First order	(τ, λ, p)	.813	21.87	.434	.332	10.25	.110
Direct	($\log \tau$, $\log \lambda$, $\log p$)	1.042	32.56	.496	.496	16.55	.142
MGF	($\log \tau$, $\log \lambda$, $\log p$)	1.026	32.2	.495	.457	15.78	.138
First order	($\log \tau$, $\log \lambda$, $\log p$)	.947	29.04	.478	.405	13.80	.129

differ in the numerical computation they suggest and, from a methodological point of view, this is what distinguishes these alternative approximations.

Since (3.1) does not require explicit third derivatives, it is easier to use, and this is the basis for our preference for it in our current software implementation (Tierney, Kass, and Kadane 1987; Kass, Tierney, and Kadane 1988). From the point of view of computational efficiency, however, neither (3.1) nor (2.6) is better than the other in every situation, so the choice between them should depend on the application. To consider this statement in a bit more detail, let us assume the parameter space Θ is m-dimensional, analytical rather than numerical second derivatives are used in both methods, and the derivative in (3.1) is obtained from a single difference quotient.

The computationally costly parts of each approximation are the function evaluations for the log-likelihood and its derivatives. Both methods require calculation of a mode and a Hessian (of the log posterior and, thus, the log-likelihood). The standard form requires an additional $2m(m + 1)(m + 2)/6$ third derivatives. If a forward difference quotient is used, the fully exponential form will require an additional $m(m + 1)/2$ second derivatives, and m first derivatives for each iteration in the maximization of h_N. As Tierney and Kadane (1986) noted, only one Newton-like iteration toward the maximum of $-nh_N$ (using the Hessian of nh_D rather than nh_N) is required for the second-order accuracy of fully exponential expectation approximations. Using only one iteration, we thus find that (3.1) will be more efficient than (2.6) when $m(m + 1)(m + 2)/6$ exceeds $m + m(m + 1)/2$. This occurs when m is greater than 2.

On the other hand, in the Stanford heart-transplant example we used a central difference quotient with two iterations to obtain (3.1) (two iterations are required for asymptotic accuracy of the variance approximation). In this case, the number of additional function evaluations required for (3.1) increases by a factor of 4, and (3.1) does not become more efficient than (2.6) until m is greater than 10. Furthermore, when (2.6) is applied to more than one function g, it requires no further function evaluations. Thus it is generally more efficient when many different g functions are of interest [as in the hierarchical modeling context of Deely and Lindley (1981), Kass and Steffey (1988), Mosteller and Wallace (1964), and Tsutakawa (1985)].

We have also shown that the MGF method is equivalent to the addition to g of a large constant c, followed by approximation of the expectation of $g + c$, and subtraction of c from the result. The choice of c, however, poses a problem analogous to the choice of s in the difference quotient required in the MGF method. The MGF approximations are exact in some cases and perform well in others, but can sometimes behave poorly; the parameterization in which they are applied is very important. Sensitivity of Laplace's method to choice of parameterization is also illustrated by the examples of Achcar and Smith (in press). Further work that would assist a data analyst in finding good parameter transformations would be valuable.

[Received December 1986. Revised January 1989.]

REFERENCES

Achcar, J. A., and Smith, A. F. M. (in press), "Aspects of Reparameterization in Approximate Bayesian Inference," in *Essays in Honor of George Barnard,* eds. S. Geisser, J. S. Hodges, S. J. Press, and A. Zellner, Amsterdam: North-Holland.

Daniels, H. E. (1980), "Exact Saddlepoint Approximations," *Biometrika,* 67, 59–63.

Deely, J. J., and Lindley, D. V. (1981), "Bayes Empirical Bayes," *Journal of the American Statistical Association,* 76, 833–841.

DiCiccio, T. J. (1986), "Approximate Conditional Inference for Location Families," *The Canadian Journal of Statistics,* 14, 5–18.

Efron, B., and Hinkley, D. V. (1978), "Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information," *Biometrika,* 65, 457–487.

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

Hartigan, J. A. (1965), "The Asymptotically Unbiased Prior Distribution," *The Annals of Mathematical Statistics,* 36, 1137–1152.

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.

Johnson, R. A., and Ladalla, J. N. (1979), "The Large Sample Behavior of Posterior Distributions With Sampling From Multiparameter Exponential Family Models, and Allied Results," *Sankhyā,* Ser. B, 41, 196–215.

Kass, R. E., and Steffey, D. (1988), "Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models)," Technical Report 386, Carnegie-Mellon University, Dept. of Statistics.

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

——— (in press), "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.

Laplace, P. S. (1986), "Memoir on the Probability of Causes of Events" (translation by S. Stigler of the 1774 original), *Statistical Science,* 1, 364–378.

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: 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.

Morris, C. N. (1988), "Approximating Posterior Distributions and Posterior Moments," in *Bayesian Statistics 3,* eds. J. Bernardo, M. DeGroot, D. V. Lindley, and A. F. M. Smith, Oxford, U.K.: Oxford University Press, pp. 327–345.

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

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

Stigler, S. M. (1986), "Laplace's 1774 Memoir on Inverse Probability," *Statistical Science,* 1, 359–362.

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. J., Kass, R. E., and Kadane, J. B. (1987), "Interactive Bayesian Analysis Using Accurate Asymptotic Approximations," in *Computer Science and Statistics: Nineteenth Symposium on the Interface,* ed. R. Heiberger, Alexandria, VA: American Statistical Association, pp. 15–21.

Tsutakawa, R. K. (1985), "Estimation of Cancer Mortality Rates: A Bayesian Analysis of Small Frequencies," *Biometrics,* 41, 69–79.

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.