Laplace Approximation of High Dimensional Integrals

Zhenming Shun; Peter McCullagh

Journal of the Royal Statistical Society. Series B (Methodological), Volume 57, Issue 4 (1995), 749-760.

Stable URL:
http://links.jstor.org/sici?sici=0035-9246%281995%2957%3A4%3C749%3ALAOHDI%3E2.0.CO%3B2-Z

Journal of the Royal Statistical Society. Series B (Methodological)

J. R. Statist. Soc. B (1995)
57, *No.* 4, *pp.* 749–760

Laplace Approximation of High Dimensional Integrals

By ZHENMING SHUN and PETER McCULLAGH†

University of Chicago, USA

[Received March 1994. Revised December 1994]

SUMMARY

It is shown that the usual Laplace approximation is not a valid asymptotic approximation when the dimension of the integral is comparable with the limiting parameter n. The formal Laplace expansion for multidimensional integrals is given and used to construct asymptotic approximations for high dimensional integrals. One example is considered in which the dimension of the integral is $O(n^{1/2})$ and the relative error of the unmodified Laplace approximation is $O(1)$. Nevertheless, it is possible to construct a valid asymptotic expansion by regrouping terms in the formal expansion according to asymptotic order in n.

Keywords: ASYMPTOTIC APPROXIMATION; BIPARTITION; CONNECTED BIPARTITION; EXCHANGEABLE ARRAY; LAPLACE APPROXIMATION; POSTERIOR EXPECTATION; RANDOM EFFECTS MODEL

1. INTRODUCTION

In statistical work connected with Bayesian computations or mixture models, it is frequently necessary to evaluate integrals of the form

$$\int_{R^p} \exp\{-n\,g(x)\}\,\mathrm{d}x \tag{1}$$

in which n is the sample size. In the standard asymptotic regime p is fixed as $n \to \infty$. In the non-standard high dimensional limits considered here, p increases with n, although usually at a diminished rate such as $n^{1/2}$ or $n^{1/3}$. The examples that we have in mind are primarily the computation of marginal distributions for random effects models, examples of which can be found in Breslow and Clayton (1993), McCullagh and Nelder (1989), section 14.5, or Wolfinger (1993). Similar calculations also occur in the computation of posterior expectations when the parameter is high dimensional (Tierney and Kadane, 1986). Analytical expressions for such integrals are available only in very rare cases, so it is natural to resort to approximations. If all else fails, Monte Carlo methods can be used to produce numerical approximations. In this paper, however, we develop analytical approximations based on modifications to Laplace's method.

It does not seem feasible at present to develop useful general theorems for approximating arbitrary high dimensional integrals. As a compromise we develop a formal multivariate expansion. The asymptotic order in n of terms in this expansion depends on the structure of the derivative arrays and on the relationship between

†*Address for correspondence*: Department of Statistics, University of Chicago, 5734 University Avenue, Chicago, IL 60637, USA.
E-mail: pmcc@galton.uchicago.edu

 0035-9246/95/57749

p and n. For certain high dimensional integrals an asymptotic approximation can be developed by grouping together terms of similar asymptotic order. The example in Section 4 shows how this can be done for a non-standard case in which $p = O(n^{1/2})$. The main point of that example is that the correction term ϵ_0, ordinarily $O(n^{-1})$ for integrals of fixed dimension, is in fact $O(1)$ under the non-standard limits considered. The multiplicative correction $\exp \epsilon_0$ yields an approximation with relative error $o(1)$, whereas the more common correction factor, $1 + \epsilon_0$, has relative error $O(1)$. These conclusions are confirmed by a numerical example in Section 5.

The discussion in Section 6 suggests that the unmodified Laplace approximation is reliable provided that $p = o(n^{1/3})$. Beyond that point it is necessary to exploit peculiar characteristics of individual problems.

2. LAPLACE APPROXIMATION

We consider briefly the standard Laplace approximation to integral (1) in which p is constant as $n \to \infty$ and g is unimodal with a minimum at $\hat{x}$. Outside any fixed interval or open set containing $\hat{x}$ the integrand is exponentially small and the integral is likewise exponentially small. Consequently, for large n, the main contribution comes from x-values in an $o(1)$ neighbourhood of $\hat{x}$. At this stage we seek a one-to-one transformation $x \mapsto u(x)$ such that $g(x) - g(\hat{x}) = \frac{1}{2}\hat{g}'' u^2$, where $\hat{g}''$ is the second derivative at the minimum. The integral then becomes

$$\exp\{-n g(\hat{x})\} \int \exp(-n\hat{g}'' u^2/2)\, J(u)\, du$$

where $J(u)$, the Jacobian of the transformation, is $1 + O(u)$ in the neighbourhood of the origin. This integral yields an asymptotic expansion in the form

$$\left(\frac{n\hat{g}''}{2\pi}\right)^{-1/2} \exp(-n\hat{g}) \left\{1 + \frac{1}{2}\frac{J_2}{n\hat{g}''} + \frac{1}{8}\frac{J_4}{(n\hat{g}'')^2} + \ldots + \frac{J_{2r}}{(2n\hat{g}'')^r r!} + O(n^{-r-1})\right\}$$

where $J_r/r!$ is the rth coefficient in the Taylor expansion of $J(u)$ about the origin. In particular, $J_1 = -\hat{g}'''/3\hat{g}''$ and

$$12J_2 = 5\hat{g}'''^2/\hat{g}''^3 - 3\hat{g}^{\mathrm{iv}}/\hat{g}''.$$

See, for example, Bleistein and Handelsman (1986), p. 338, or Tierney and Kadane (1986), p. 86.

The preceding expressions apply to the one-dimensional case but can readily be extended to the multivariate case by a simple change of notation employing the summation convention. The leading term, sometimes called the Laplace approximation, is

$$\det(n\hat{g}''/2\pi)^{-1/2} \exp(-n\hat{g}).$$

The first multiplicative correction term is $1 + \epsilon_0$ where

$$\epsilon_0 = -\frac{1}{24n}\hat{g}_{ijkl}\hat{g}^{ij}\hat{g}^{kl}[3] + \frac{1}{72n}\hat{g}_{ijk}\hat{g}_{rst}(\hat{g}^{ir}\hat{g}^{js}\hat{g}^{kt}[6] + \hat{g}^{ij}\hat{g}^{kr}\hat{g}^{st}[9]) \qquad (2)$$

and $\hat{g}^{ij}$ are the components of the inverse matrix of second derivatives. For details

of the derivation, see Barndorff-Nielsen and Cox (1989), section 6.2. Since the correction term above is formally identical with the Bartlett adjustment for full exponential family models, the computational techniques described by Cordeiro (1983, 1987) or McCullagh and Nelder (1989), section 15.3, which avoid multidimensional arrays, can be adapted.

The preceding line of development seems to be the most persuasive for establishing rigorously the asymptotic nature of the Laplace approximation. However, to develop an extended series expansion a more streamlined method is required.

3. BIPARTITIONS AND MULTIVARIATE EXPANSIONS

Let X be a p-dimensional random variable with components $X^1, \ldots, X^p$ having finite moments of all orders. In a slight departure from the notation of McCullagh (1987) the joint cumulants and joint moments are denoted as follows:

$$\kappa^i = E(X^i); \qquad \kappa^{ij} = \operatorname{cov}(X^i, X^j) = \operatorname{cum}_2(X^i, X^j); \qquad \kappa^{ijk} = \operatorname{cum}_3(X^i, X^j, X^k);$$

$$\kappa^{[ij]} = E(X^iX^j) = \kappa^{ij} + \kappa^i\kappa^j; \qquad \kappa^{[ijk]} = E(X^iX^jX^k) = \kappa^{ijk} + \kappa^{ij}\kappa^k[3] + \kappa^i\kappa^j\kappa^k.$$

It is convenient in this section to adopt the convention whereby any array with bracketed indices is defined to be the sum over all partitions of those indices of products of the related arrays with unbracketed indices. Thus, for instance, $a_{[ijkl]}$ is defined to be

$$a_{[ijkl]} = a_{ijkl} + a_{ijk}a_l[4] + a_{ij}a_{kl}[3] + a_{ij}a_ka_l[6] + a_ia_ja_ka_l,$$

the sum over all 15 partitions of four indices of products of arrays with unbracketed coefficients. A consequence of the notation is that the κs with bracketed indices denote moments; without brackets they denote cumulants.

Consider now the moment-generating function defined by the expected value of the expansion

$$M = E\exp(a_iX^i + a_{ij}X^iX^j/2! + a_{ijk}X^iX^jX^k/3! + \ldots),$$

which employs the usual implicit summation convention. We proceed formally ignoring all questions of convergence. From the relationship between moment-generating functions and cumulant-generating functions, we have

$$\begin{aligned} M &= E(1 + a_{[i]}X^i + a_{[ij]}X^iX^j/2! + a_{[ijk]}X^iX^jX^k/3! + a_{[ijkl]}X^iX^jX^kX^l/4! + \ldots) \\ &= 1 + a_{[i]}\kappa^{[i]} + a_{[ij]}\kappa^{[ij]}/2! + a_{[ijk]}\kappa^{[ijk]}/3! + a_{[ijkl]}\kappa^{[ijkl]}/4! + \ldots. \end{aligned}$$

In the term of order m, involving m dummy indices, $a_{[i_1 \ldots i_m]}$ is the sum over the partition lattice $\mathscr{P}_m$ of products of coefficients a with unbracketed indices. Likewise, the moment $\kappa^{[i_1 \cdots i_m]}$ is the sum over $\mathscr{P}_m$ of cumulant products, one cumulant for each block of the partition. Thus the complete term of algebraic order m in the expansion of M is a sum over $\mathscr{P}_m \times \mathscr{P}_m$ of coefficient products multiplied by cumulant products, each bipartition occurring exactly once.

Now, M is the joint moment-generating function of the variables X^i, X^iX^j, $X^iX^jX^k$, ..., and $K = \log M$ is the joint cumulant-generating function of these same variables. Consequently K has a formal expansion similar to that for M in

which the term of order m is a sum over $\mathscr{P}_m \times \mathscr{P}_m$, but omitting bipartitions that are not connected (McCullagh (1987), p. 59). In symbols,

$$\log M = \sum_{m \geqslant 1} \frac{1}{m!} \sum_{P = p_1 | \ldots | p_\nu} \sum_{\substack{Q = q_1 | \ldots | q_\chi \\ P \vee Q = 1}} a_{p_1} \ldots a_{p_\nu} \kappa^{q_1} \ldots \kappa^{q_\chi}, \tag{3}$$

where P is a partition of m indices into $\nu \leqslant m$ blocks, Q is a partition of the same indices into $\chi \leqslant m$ blocks and the coefficient is 1 if the least upper bound of P and Q is maximal. In other words, all bipartitions (P, Q) are included in the expansion for M, but only connected bipartitions appear in the expansion for $\log M$. Otherwise the two expansions are formally identical.

3.1. *Application to Laplace Approximation*

Consider the formal Laplace approximation to the integral

$$M = \int_{R^p} \exp\{-g(x)\}\,dx$$

in which g has derivatives of all orders and is unimodal with a minimum at the origin. In the usual asymptotic development outlined in the preceding section, the sample size appears explicitly as a multiplicative factor in the exponent. Here, however, we ignore n but bear in mind that g and its derivatives may be $O(n)$ where n is some parameter to be considered large. The integrand may be factored as follows:

$$M = \exp\{-g(0)\} \int \exp(-\tfrac{1}{2} g_{ij} x^i x^j) \exp(-g_{ijk} x^i x^j x^k/3! - g_{ijkl} x^i x^j x^k x^l/4! - \ldots)\,dx$$

$$= |g''/2\pi|^{-1/2} \exp\{-g(0)\} E \exp(-g_{ijk} X^i X^j X^k/3! - g_{ijkl} X^i X^j X^k X^l/4! - \ldots).$$

In the notation of the preceding section, X is jointly normal with inverse covariance matrix g'', so all cumulants are 0 except for those of order 2. All coefficient arrays of order 1 and 2 are also 0. It follows then from equation (3) that

$$\log M = -g(0) - \frac{1}{2} \log\left\{ \det\left(\frac{g''}{2\pi}\right)\right\} + \sum_m \frac{1}{2m!} \sum_{\substack{\mathscr{P}_{2m} \times \mathscr{Q}_{2m} \\ P \vee Q = 1}} (-1)^\nu g_{p_1} \ldots g_{p_\nu} g^{q_1} \ldots g^{q_m} \tag{4}$$

in which $P = p_1 | \ldots | p_\nu$ is a partition of $2m$ indices into ν blocks each of size 3 or more and $Q = q_1 | \ldots | q_m$ is a partition of the same indices into m blocks of size 2. In equation (4), p_j is a set of indices, g_{p_1} is the associated array of partial derivatives and g^{q_1} is the inverse matrix of second derivatives at the origin.

The usual asymptotic order for fixed p of the term corresponding to the bipartition (P, Q) is $O(n^{\nu - m})$ regardless of whether the bipartition is connected. The standard Laplace expansion is obtained in logarithmic form when terms in equation (4) are grouped by asymptotic order. Alternatively, a multiplicative correction can be obtained by including disconnected bipartitions and grouping by asymptotic order. In either case, the first adjustment term, of order $O(n^{-1})$, includes three bipartitions of type $(4, 2^2)$ and 150 of type $(3^2, 2^3)$ split into two distinct subtypes as shown in equation (2).

The discussion leading up to equation (4) assumes that the Taylor expansion takes

place at the point that minimizes g. Although this choice is usually desirable to achieve good numerical accuracy using few terms, it is not necessary in the development of the formal series expansion. In fact, it is sometimes convenient in applications to expand g about a point near but not exactly equal to the minimum point. In symbols, we write

$$g(x) = h(x) + \epsilon(x)$$

where h is a quadratic approximation to g in the neighbourhood of the minimum, and ϵ has first and second derivatives that are small at the point of expansion. Then the formal expansion about an arbitrary point is

$$\log M = -g(0) - \frac{1}{2}\log\left\{\det\left(\frac{h''}{2\pi}\right)\right\} + \sum_m \frac{1}{2m!} \sum_{\substack{\mathscr{P}_{2m} \times \mathscr{Q}_{2m} \\ P \vee Q = 1}} (-1)^{\nu} \epsilon_{p_1} \ldots \epsilon_{p_\nu} h^{q_1} \ldots h^{q_m}, \tag{5}$$

where ϵ_{pj} are the partial derivatives of ϵ of order $|p_j|$ and h^{q_j} are the components of the inverse second-derivative matrix of h. Since the first and second derivatives of ϵ are not exactly 0, the block sizes of the partition P are unrestricted in equation (5). The grouping of terms in equation (5) to form an asymptotic expansion now depends critically on the magnitude in n of the first two derivatives of ϵ.

The main purpose of developing the formal expansions (4) and (5) is not so much to carry Laplace approximation out to higher order under standard asymptotic conditions but to use it under non-standard conditions in which a non-standard reordering of the terms may be called for. In particular, our expansion is formally correct even when the dimension of the integral is equal to n. Whether it is then possible to group terms in a useful manner with appropriate asymptotic behaviour as $n \to \infty$ is a problem to be examined in each case.

4. APPLICATION

4.1. *Exchangeable Binary Array Model*

Let Y_{ij} be the components of a binary random array of order $r \times c$ whose joint distribution is generated as follows. Conditionally on the values of row and column effects α_i and β_j, the components of Y are independent with probabilities π_{ij} satisfying the linear logistic model

$$\text{logit}(\pi_{ij}) = \mu + \alpha_i + \beta_j. \tag{6}$$

The row and column logistic effects are taken to be independent and normally distributed random variables with variances σ_a^2 and σ_b^2 respectively. The unconditional joint probability of the observations y_{ij} is then proportional to

$$\sigma_a^{-r}\sigma_b^{-c} \int_\alpha \int_\beta \exp\left\{\mu y.. + \sum \alpha_i y_{i\cdot} + \sum \beta_j y_{\cdot j} - K(\alpha \oplus \beta) - \alpha^{\mathrm{T}}\alpha/2\sigma_a^2 - \beta^{\mathrm{T}}\beta/2\sigma_b^2\right\} d\beta\, d\alpha \tag{7}$$

where the dot subscript denotes summation and $K(\alpha \oplus \beta) = \Sigma \log\{1 + \exp(\mu + \alpha_i + \beta_j)\}$ is the conditional cumulant function. This joint distribution is row exchangeable and column exchangeable.

In this example the number of 'observations' is $r \times c$ whereas the dimension of the integral in expression (7) is $r + c$. Our asymptotic approximation is based on

a formal limit in which $r = c \to \infty$, so the dimension of the integral increases at a slower rate than the number of observations.

Although the integral arising in expression (7) is in some ways rather special, its form is not limited to exchangeable binary arrays. In fact, if the components Y_{ij} are conditionally independent in a natural exponential family model whose canonical parameter satisfies model (6), then the unconditional distribution has exactly the form (7), with cumulant function appropriate to the exponential family in question. Such a formulation makes sense only if the canonical parameter space coincides with $R^r \times R^c$.

4.2. *Asymptotics and Derivatives*

Since K is convex, the exponent in expression (7) is concave with a unique maximum. However, the observations are not identically distributed so the sample size does not enter explicitly as a multiplicative factor as in integral (1). To make the formal connection with expression (1), therefore, we take $r = c$ and incorporate the sample size ($n = r^2$) into the function g. In other words, $x = (\alpha, \beta)$ and $-g$ is the exponent in expression (7). The partial derivatives of g of order 3 and higher are the derivatives of K, which are the joint conditional cumulants of $(Y_{i\cdot}, Y_{\cdot j})$. In the cumulant array of order $s + t$ the only non-zero elements are the $r + c$ 'diagonal' elements $\kappa_{s+t}(Y_{i\cdot})$ and $\kappa_{s+t}(Y_{\cdot j})$, and the two-component 'mixed' cumulants $\kappa_{st}(Y_{i\cdot}, Y_{\cdot j}) = \kappa_{s+t}(Y_{ij})$. The $r + c$ diagonal elements are $O(r)$; the $2rc$ non-zero off-diagonal elements are $O(1)$ in the formal limit considered here.

A slight complication arises from the fact that $\Sigma Y_{i\cdot} - \Sigma Y_{\cdot j}$ is identically 0. The $(r + c)$-component vector $J = (1_r, -1_c)$ lies in the null space of each cumulant tensor of order 2 and higher. For example, $K_{abcd}J^d = 0$ for all a, b and c. The second-derivative matrix has the partitioned form

$$g'' = K'' + \Sigma^{-1} = \begin{pmatrix} \operatorname{diag}\{V_{i\cdot}\} & V \\ V^{\mathrm{T}} & \operatorname{diag}\{V_{\cdot j}\} \end{pmatrix} + \begin{pmatrix} \Sigma_a^{-1} & 0 \\ 0 & \Sigma_b^{-1} \end{pmatrix}$$

where $V = \{V_{ij}\}$ is the array of conditional variances, and Σ_a and Σ_b are the covariance matrices of α and β, here assumed to be multiples of the identity matrix. Ordinarily, to justify the Laplace formula as an asymptotic approximation it is necessary to show that g'' is large in the sense that $n^{-1}g''$ or $r^{-1}g''$ has a positive definite limit. Clearly, such an approach cannot work in our case because the dimension of g'' increases without limit. Also, at least one of the eigenvalues of g'' is $O(1)$, which further complicates the asymptotic analysis. To see this, consider the case $\sigma_a = \sigma_b = \sigma$. Then J is an eigenvector of g'' with eigenvalue σ^{-2}. All eigenvectors orthogonal to J correspond to non-trivial contrasts of the row and column totals and therefore have eigenvalues of order $O(r)$.

The formal asymptotic argument proceeds as follows. First, $g'' > K''$ in the sense that the difference $g'' - K'' = \Sigma^{-1}$ is positive definite with eigenvalues that are $O(1)$. Second, in all scalars of the type (2), g'' may be replaced by K'' without affecting the asymptotic order. Finally K'' may be replaced by $K'' + \lambda JJ^{\mathrm{T}}$ without affecting the numerical value. This augmented matrix is positive definite for $\lambda > 0$. Equivalently, scalars of the type (2) are unaffected by the choice of generalized inverse provided that all cumulant arrays have a common null space. It is convenient here to choose $\lambda = \bar{V}$, the mean of V_{ij}, in which case all eigenvalues of the

augmented matrix are strictly $O(r)$. For determining the asymptotic order of various scalars, therefore, it is sufficiently good to proceed as if $g'' = rI_{2r}$. For serious numerical approximation, however, it is essential to use the exact matrix g'' or a good approximation thereof.

The asymptotic order of the first scalar in equation (2) is determined as follows:

$$\begin{aligned} g_{ijkl}g^{ij}g^{kl} &\sim r^{-2}K_{ijkl}\delta^{ij}\delta^{kl} \\ &\sim r^{-2}\sum_i K_{iiii} + r^{-2}\sum_{ij}{}^{*} K_{iijj} \\ &\sim \bar{K}_4/r + \bar{K}_{22}, \end{aligned}$$

in an obvious notation showing that this scalar is $O(1)$ and not $O(n^{-1})$ as in the standard version of the Laplace expansion. Note that $\bar{K}_4$, the average of the diagonal elements, is $O(r)$. Similar investigations for the remaining scalars in equation (2) show that both are $O(1)$. For example the final scalar is

$$\begin{aligned} g_{ijk}g_{rst}g^{ij}g^{kr}g^{st} &\sim r^{-3}K_{ijk}K_{rst}\delta^{ij}\delta^{kr}\delta^{st} \\ &\sim r^{-3}\sum_i K_{iii}^2 + 2r^{-3}\sum_{ij}{}^{*} K_{iii}K_{ijj} + r^{-3}\sum_{ijk}{}^{*} K_{iij}K_{jkk}, \end{aligned}$$

where Σ^* denotes summation over distinct values of the indices. The first sum involves $2r$ terms of order $O(r^2)$; the second involves roughly r^2 terms of order $O(r)$; the third involves $2r^3$ terms of order $O(1)$. All three expressions are $O(1)$.

4.3. *Modified Laplace Expansion*

The results of the preceding section show that the standard Laplace formula does not provide a valid asymptotic approximation with relative error $o(1)$ for integral (7). We now seek a modification of the Laplace expansion by making use of the formal expansion (4). The general idea is to regroup in decreasing asymptotic order the scalars that appear in that expansion and to truncate the resulting series at an appropriate point.

The terms of algebraic order $2m$ in expansion (4) are in one-to-one correspondence with connected bipartitions (P, Q) of $2m$ indices with $P = p_1|\ldots|p_\nu$ and $Q = q_1|\ldots|q_m$ satisfying $|p_j| \geqslant 3$ and $|q_j| = 2$. In other words, Q is a 2^m-partition, and P has blocks of size not less than 3. For each such bipartition, it is required to determine the asymptotic order in n, or in $r = n^{1/2}$, of the scalar

$$K_{p_1} \ldots K_{p_\nu} g^{q_1} \ldots g^{q_m} \tag{8}$$

bearing in mind that such an expression involves implicit summation over $(2r)^{2m}$ terms.

In the context of integral (7), the arguments presented in the preceding section show that, as regards asymptotic order in n, we may take each superscripted g to be $g^q = r^{-1}I_{2r}$. Each of the $2r$ diagonal elements in expression (8), for which all $2m$ indices are equal, is therefore $O(r^{\nu-m})$, giving a diagonal contribution of order $O(r^{\nu-m+1})$. The off-diagonal elements for which the $2m$ indices take on exactly two distinct values are $O(r^{\nu-m-1})$ at most, again giving a contribution of order $O(r^{\nu-m+1})$. By extension, the off-diagonal elements for which the indices

take on $k \leqslant \nu + 1$ distinct values are $O(r^{\nu+1-k-m})$, but there are $O(r^k)$ such terms giving a total contribution of order $O(r^{\nu-m+1})$. If the indices take on more than $\nu + 1$ distinct values, connectivity of the partitions implies that one of the Ks must have three distinct subscripts. Such terms are necessarily 0 by the argument given in the preceding section. Thus the asymptotic order in n of the scalar corresponding to the connected bipartition (P, Q) is $O(r^{\nu-m+1})$. The value of any scalar on a disconnected bipartition is equal to the product of lower order scalars over the connected components of the bipartition. Bipartitions that are not connected generally give rise to scalars that are $O(1)$. Thus, truncation of the expansion for $\log M$ in equation (4) gives a valid asymptotic approximation, but truncation of the series expansion for M does not.

An asymptotic expansion for the *logarithm* of integral (7) can therefore be obtained by grouping together bipartitions from equation (4) in descending order of $\nu - m + 1$. The usual approximation with multiplicative correction (2) for the integral is therefore not asymptotically correct in the limit considered here. Instead, the correct leading term in the Laplace approximation has the exponential form

$$\det(\hat{g}''/2\pi)^{-1/2}\exp(-\hat{g} + \epsilon_0) \tag{9}$$

where

$$\epsilon_0 = -\frac{1}{24}\hat{g}_{ijkl}\hat{g}^{ij}\hat{g}^{kl}[3] + \frac{1}{72}\hat{g}_{ijk}\hat{g}_{rst}(\hat{g}^{ir}\hat{g}^{js}\hat{g}^{kt}[6] + \hat{g}^{ij}\hat{g}^{kr}\hat{g}^{st}[9]).$$

The zero-order 'correction term' $\exp\epsilon_0$ is equivalent to $1 + \epsilon_0$ under standard limits but differs by $O(1)$ under the non-standard limits considered here. The error term in the exponent of expression (9) is $O(r^{-1})$, or $O(n^{-1/2})$.

5. NUMERICAL EXAMPLE

We consider the problem of evaluating integral (7) on the 10×10 binary array shown in Table 1. For $\mu = \sigma_a = \sigma_b = 1$, the 'exact' value obtained by simulation is $(2\pi)^{10}\exp(-72.2938 \pm 0.0032)$. The uncorrected Laplace approximation gives $(2\pi)^{10}\exp(-72.6796)$, whereas formula (9) with the 'zero-order correction' gives $(2\pi)^{10}\exp(-72.2612)$. The zero-order correction term in this example is equal to $\exp 0.4184 = 1.520$, or 0.837 in twice-log-likelihood units. The zero-order correction reduces the logarithmic error of the Laplace approximation from -0.386 to 0.0326,

TABLE 1
10 × 10 binary array for the example

1	1	1	1	0	1	1	1	0	1
0	0	1	0	1	0	1	1	0	0
0	1	1	1	1	1	1	1	1	1
1	0	1	0	1	1	0	1	1	1
0	0	1	0	1	1	1	1	1	1
1	0	0	1	1	1	0	0	1	0
1	0	0	0	0	1	0	0	1	1
1	1	0	0	0	1	1	0	1	0
1	0	1	0	1	0	1	1	0	0
1	1	1	1	0	1	1	0	0	0

a slight overcorrection. The multiplicative $1 + \epsilon_0$ correction reduces the error to -0.0363, a slight undercorrection, but not markedly inferior to the exponentiated correction.

More extensive numerical investigation reveals that ϵ_0 depends more on the values of σ_a and σ_b than on μ. For $\mu = 1$, and for several values of $\sigma = \sigma_a = \sigma_b$, Table 2 compares the various Laplace approximations for the log-likelihood with the value obtained by simulation. The column labelled 'Laplace' is the estimated error in the logarithm of the approximation to integral (7). It is readily apparent that ϵ_0 varies appreciably with σ, even over the range of statistical interest. Further, the exponentiated correction (additive on the log-likelihood scale) generally overcorrects, whereas $1 + \epsilon_0$ undercorrects, at least in this example in which ϵ_0 is positive. Although there is little to choose between the two corrections for $\sigma \leqslant 1$, the exponentiated correction is clearly superior for large σ where the correction is greatest.

TABLE 2

Log-likelihood approximations compared with Monte Carlo simulations for various values of $\sigma_a = \sigma_b$ with $\mu = 1$

σ	*Monte Carlo results*†		ϵ_0	*Estimated error in approximation*		
	Log-likelihood	*Standard error*		*Laplace*	*Laplace* + ϵ_0	*Laplace* + *log*$(1 + \epsilon_0)$
0.25	−70.5109	0.0002	0.0091	−0.0092	−0.0002	−0.0002
0.50	−70.0090	0.0009	0.1062	−0.1024	0.0038	−0.0015
0.75	−70.7847	0.0018	0.2682	−0.2509	0.0173	−0.0133
1.00	−72.2938	0.0032	0.4184	−0.3858	0.0326	−0.0363
1.50	−75.9453	0.0048	0.6349	−0.5901	0.0448	−0.0985
2.00	−79.5089	0.0087	0.7680	−0.7320	0.0360	−0.1622
3.00	−85.6111	0.0104	0.9071	−0.8525	0.0547	−0.2069
5.00	−94.3584	0.0162	1.0054	−0.9497	0.0557	−0.2538

†Based on 50000 integrand evaluations.

TABLE 3

Log-likelihood approximations compared with Monte Carlo simulations for a (20 × 20)-table for various values of $\sigma_a = \sigma_b$ with $\mu = 1$

σ	*Monte Carlo results*†		ϵ_0	*Estimated error in approximation*		
	Log-likelihood	*Standard error*		*Laplace*	*Laplace* + ϵ_0	*Laplace* + *log*$(1 + \epsilon_0)$
0.50	−244.2616	0.0016	0.2228	−0.2159	0.0069	−0.0148
0.75	−228.9482	0.0031	0.4056	−0.3896	0.0160	−0.0491
1.00	−221.7176	0.0055	0.5576	−0.5416	0.0160	−0.0984
1.50	−217.6652	0.0085	0.7960	−0.7546	0.0414	−0.1690
2.00	−218.9059	0.0126	0.9746	−0.9080	0.0667	−0.2276
3.00	−225.3109	0.0183	1.2329	−1.1210	0.1119	−0.3177
5.00	−238.6189	0.0371	1.6165	−1.3664	0.2502	−0.4045

†Based on 50000 integrand evaluations.

Similar calculations are given in Table 3 for a 20 × 20 binary array in which the marginal totals are moderately dispersed as shown below:

$y_{i\cdot}$ = (11, 7, 16, 13, 10, 17, 6, 12, 10, 10, 15, 5, 16, 10, 6, 11, 12, 16, 12, 10),

$y_{\cdot j}$ = (17, 4, 18, 11, 10, 15, 13, 4, 16, 18, 9, 7, 9, 20, 11, 12, 4, 6, 20, 1).

A consequence of this excess dispersion is that the estimate of σ is around 1.5. The overall picture is that ϵ_0 is not negligible. Even by the standards of approximation required for applied statistics, it varies substantially over the region of interest. Although it overcorrects, the exponentiated correction is clearly superior to the $(1 + \epsilon_0)$-correction. The accuracy of both approximations decreases for very large values of σ and for tables with highly dispersed marginal totals. For the values considered here, the error in the modified Laplace approximation is comparable with the Monte Carlo standard error based on 2000 integrand evaluations.

Even though the dimension of the integrals in Table 3 is twice the dimension in Table 2, the errors in the modified approximation are larger for the higher dimensional integrals, in apparent contradiction of the claim made in Section 4. The reason for this is that the error in the new approximation depends on the variability of the observed marginal totals, $\{y_{i\cdot}, y_{\cdot j}\}$. The marginal totals for the integrals in Table 3 are considerably more dispersed than those in Table 2. To check on the asymptotics, we generated a 25 × 25 binary array with $\sigma_a = \sigma_b = 0.5$ and evaluated integral (7) for $\sigma_a = \sigma_b = 1$ on a nested sequence of square subtables with sizes ranging from $r = 3$ to $r = 25$. The variability of the marginal frequencies is thereby kept constant, at least in a probabilistic sense. The error in the approximate log-likelihood based on the modified Laplace formula (9) increased to a maximum of 0.0267 ± 0.0012 at $r = 10$, thereafter decreasing to 0.0182 ± 0.0019 at $r = 25$. The value of ϵ_0 increased with r from 0 to 0.411 at $r = 10$ and 0.695 at $r = 25$.

Monte Carlo simulations were carried out by generating multivariate normal random variables with mean $(\hat{\alpha}, \hat{\beta})$ and inverse covariance matrix equal to $\hat{g}''$, the second derivative of the exponent in equation (3). This procedure ensures that the ratio of the integrand to the normal density is nearly constant in the centre. The integral is estimated by the average value of this ratio, and variability is thus kept to a minimum. Effectively, the Monte Carlo procedure mimics by simulation the steps that are approximated by expansion and integration in the modified Laplace formula. Monte Carlo simulation has the advantage over asymptotic methods that greater accuracy can be achieved by increasing the number of simulations. However, despite the precautions taken here to improve the efficiency of the simulation, roughly 100 modified Laplace approximations could be computed in the time devoted to one Monte Carlo simulation with 4000 integrand evaluations.

6. POSTERIOR EXPECTATIONS FOR EXPONENTIAL MODELS

Let $l(\beta; y)$ be the log-likelihood for the p-dimensional parameter β based on data y. Our asymptotic development is such that y has n independent components, not necessarily identically distributed, and that $p = o(n)$ but not necessarily small. For definiteness, assume that the log-likelihood has the exponential family form

$$l(\beta; y) = y\theta - K(\theta) = yX\beta - K(X\beta),$$

where the observation-specific canonical parameters satisfy the regression model $\theta = X\beta$ for some model matrix X of order $n \times p$.

In a notation closer to that of Section 3, let g be the function

$$g(\beta) = -yX\beta + K(X\beta) + \epsilon(\beta)$$

in which $\epsilon = O(p)$ and all its partial derivatives are $O(1)$ as $p \to \infty$. For statistical applications $\exp\{-\epsilon(\beta)\}$ is the product of the prior and that function of the parameter whose posterior expectation is required. For a slightly different statistical interpretation of a similar integral, see Wong and Li (1992). The partial derivatives of g take the form

$$g_{rs} = \sum_i X_{ir}X_{is}\kappa_{2i} + \epsilon_{rs},$$

$$g_{rst} = \sum_i X_{ir}X_{is}X_{it}\kappa_{3i} + \epsilon_{rst}$$

and so on, where κ_{ri} is the rth cumulant of Y_i. Under typical limiting conditions on the matrix X, the components of these arrays are $O(n)$, but it should be borne in mind that the arrays themselves are of order $p^2, p^3, \ldots$ in terms of the number of components. For example, if the rows of X are formally independent and identically distributed random variables with zero mean, covariances λ_{rs}, higher order cumulants $\lambda_{rst}, \ldots$ then

$$\left.\begin{aligned} g_{rs} &\simeq n\kappa_2\lambda_{rs}\{1 + O(n^{-1})\}, \\ g_{rst} &\simeq n\kappa_3\lambda_{rst}\{1 + O(n^{-1})\}, \\ g_{rstu} &\simeq n\kappa_4(\lambda_{rstu} + \lambda_{rs}\lambda_{tu}[3])\{1 + O(n^{-1})\} \end{aligned}\right\} \tag{10}$$

under the simplifying assumption that the higher order cumulants of Y are approximately constant. Whatever the mechanism generating the matrix X, we assume that the derivatives of g behave according to conditions (10). Without loss of generality we may take $g_{rs} = n\delta_{rs}$: this can be achieved by linear transformation of the columns of X.

The scalar $g_{rstu}g^{rs}g^{tu}$ is easily seen to be of order $O(p^2/n)$ since it involves summation over p^2 terms each of order $O(n^{-1})$. By the same argument, the scalars $g_{ijk}g_{rst}g^{ir}g^{js}g^{kt}$ and $g_{ijk}g_{rst}g^{ij}g^{kr}g^{st}$ are of order $O(p^3/n)$. In general, a bipartition (P, Q) in expansion (4) has asymptotic order $O(p^m n^{\nu - m})$. Since $3\nu \leqslant 2m$, those bipartitions of type $(3^\nu, 2^m)$ have asympotic order $O(p^m n^{-m/3})$. Thus if $p = o(n^{1/3})$ the standard Laplace formula has relative error $o(1)$: the correction term in expression (9) is $O(p^3/n)$ and does generally improve the approximation. If $n^{-1/3}p \not\to 0$ neither the standard Laplace formula nor the modified version (9) is asymptotically valid under the limiting conditions (10).

A posterior expectation is the ratio of two rather similar integrals. Under standard conditions the errors in the two Laplace approximations tend to cancel, so that the error in the ratio is typically smaller than the errors in the individual integrals (Tierney and Kadane, 1986). There is reason to expect the same phenomenon to occur here, but we have not investigated the extent of such cancellation under nonstandard limits. It is conceivable that the relative error in the ratio might in some

circumstances be $o(1)$ even when the relative error in the individual integrals is $O(1)$.

REFERENCES

Barndorff-Nielsen, O. E. and Cox, D. R. (1989) *Asymptotic Techniques for Use in Statistics*. London: Chapman and Hall.

Bleistein, N. and Handelsman, R. (1986) *Asymptotic Expansions of Integrals*. New York: Dover Publications.

Breslow, N. E. and Clayton, D. G. (1993) Approximate inference in generalized linear mixed models. *J. Am. Statist. Ass.*, **88**, 9–25.

Cordeiro, G. M. (1983) Improved likelihood-ratio statistics for generalized linear models. *J. R. Statist. Soc.* B, **45**, 404–413.

——— (1987) On the corrections to the likelihood-ratio statistics. *Biometrika*, **74**, 265–274.

McCullagh, P. (1987) *Tensor Methods in Statistics*. London: Chapman and Hall.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.

Tierney, L. and Kadane, J. B. (1986) Accurate approximations for posterior moments and marginal densities. *J. Am. Statist. Ass.*, **81**, 82–90.

Wolfinger, R. (1993) Laplace's approximation for non-linear mixed models. *Biometrika*, **80**, 791–795.

Wong, W. H. and Li, B. (1992) Laplace expansion for posterior densities of non-linear functions of parameters. *Biometrika*, **79**, 393–398.