The Annals of Statistics
1996, Vol. 24, No. 1, 353–364

HIGHER-ORDER APPROXIMATIONS TO CONDITIONAL DISTRIBUTION FUNCTIONS[1]

BY JOHN E. KOLASSA

University of Rochester Medical Center

This paper derives higher-order terms in the double-saddlepoint expansion of Skovgaard for a unidimensional conditional cumulative distribution function. Expansions for continuous and lattice random variables are derived. Results are applied to the sufficient statistic in logistic regression.

1. Introduction. Many authors use double-saddlepoint approximations to perform conditional inference in the presence of nuisance parameters. Among these are Bedrick and Hill (1992), Davison (1988) and Kolassa and Tanner (1994). This paper derives higher-order terms in this double-saddlepoint expansion for a unidimensional conditional cumulative distribution function.

Skovgaard (1987) derives a conditional distribution function approximation by approximating a multiple integral representing the product of the conditional distribution function and the density of statistics to be conditioned on. Jensen (1991) provides an alternate expansion using the method of Esscher.

The present work uses an extension of an expansion theorem of Temme (1982) to approximate the multivariate integral used by Skovgaard (1987). The resulting series is then factored into two factors, the first being an approximation to the density of statistics conditioned on. The second is the approximation to the conditional distribution function. This method is applied to continuous and lattice distributions.

2. Conditional cumulative distribution function expansions. Consider random vectors $\mathbf{X} = (X^1, \dots, X^d)$ arising as means of N independent and identically distributed random vectors $\mathbf{Y}_j$, each with cumulant generating function $K_{\mathbf{Y}}$. Define the quantity $D(x^1|x^2, \dots, x^d) = \int_{x^1}^{\infty} f(v, x^2, \dots, x^d)\, dv$; then, by the definition of the conditional tail probabilities,

$$P(X^1 \geq x^1 | X^2 = x^2, \dots, X^d = x^d) \times f_{X^2, \dots, X^d}(x^2, \dots, x^d)$$
$$= D(x^1|x^2, \dots, x^d).$$

Received September 1993; revised April 1995.
[1]Partially supported by NIH Grant CA 63050 and by IBM Corp.
AMS 1991 *subject classifications*. Primary 62E20; secondary 60E99.
Key words and phrases. Conditional probability, hypergeometric distribution, lattice variable, tail probability.

The cumulative distribution function approximation is derived by finding a series expansion for D and dividing by $f_{X^2,\dots,X^d}(x^2,\dots,x^d)$. The quotient is hence the desired series expansion.

A useful representation for D is derived from the standard cumulant generating function inversion formula for recovering a probability density,

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{N^d}{(2\pi i)^d}\int_{-i\infty}^{+i\infty}\cdots\int_{-i\infty}^{+i\infty}\exp\bigl(N\bigl[K_{\mathbf{Y}}(\boldsymbol{\beta}) - \beta_i x^i\bigr]\bigr)\,d\boldsymbol{\beta},$$

by replacing x^1 by a dummy integration variable w, and then integrating with respect to w between x^1 and ∞, to yield

$$D(x^1|x^2,\dots,x^d) = \frac{N^{d-1}}{(2\pi i)^d}\int_{-i\infty}^{+i\infty}\cdots\int_{-i\infty}^{+i\infty}\exp\bigl(N\bigl[K_{\mathbf{Y}}(\boldsymbol{\beta}) - \beta_j x^j\bigr]\bigr)\frac{d\boldsymbol{\beta}}{\beta_1}. \tag{1}$$

Unless otherwise noted, the presence of an index in a term both as a superscript and as a subscript indicates summation over that index. Skovgaard expands D in terms of the multivariate saddlepoint both for the full distribution of $\mathbf{X}$ and for the distribution of the shorter random vector $(X^2,\dots,X^d)$, and in terms of derivatives of $K_{\mathbf{Y}}$ at these saddlepoints. Define the saddlepoints $\hat{\boldsymbol{\beta}}$ solving $K^i_{\mathbf{Y}}(\hat{\boldsymbol{\beta}}) = x^i$, and define the reduced saddlepoint $\tilde{\boldsymbol{\beta}}$ to be the vector with d components satisfying $K^j_{\mathbf{Y}}(\tilde{\boldsymbol{\beta}}) = x^j$ for $j = 2,\dots,d$ and $\tilde{\beta}_1 = 0$, where $K^j_{\mathbf{Y}}$ denotes the derivative of $K_{\mathbf{Y}}$ with respect to component j of its argument. Factor D as above to obtain the conditional tail probability approximation of interest:

$$1 - \tilde{F}(x^1|x^2,\dots,x^d) = 1 - \Phi(\sqrt{N}\hat{w}) + \phi(\sqrt{N}\hat{w})\left(\frac{\sqrt{|K_{\mathbf{Y}''_{-1}}(\tilde{\boldsymbol{\beta}})|}}{\hat{\beta}_1\sqrt{N|K''_{\mathbf{Y}}(\hat{\boldsymbol{\beta}})|}} - \frac{1}{\sqrt{N}\hat{w}}\right), \tag{2}$$

where $\hat{\beta}_1$ is the first component of $\hat{\boldsymbol{\beta}}$,

$$\hat{w} = \operatorname{sgn}(\hat{\beta}_1)\sqrt{2\bigl[\hat{\beta}_j x^j - K_{\mathbf{Y}}(\hat{\boldsymbol{\beta}})\bigr] - 2\bigl[\tilde{\beta}_j x^j - K_{\mathbf{Y}}(\tilde{\boldsymbol{\beta}})\bigr]},$$

$K_{\mathbf{Y}''_{-1}}$ is the $(d-1)\times(d-1)$ submatrix of second derivatives of $K_{\mathbf{Y}}$, corresponding to all components of $\boldsymbol{\beta}$ except the first, and Φ and ϕ are the normal distribution function and density, respectively.

Also of interest are inversion techniques for conditional distributions supported on a lattice, or a set of form $\{a + j\Delta | j \text{ an integer}\}$. Often such conditional lattice distributions arise from multivariate distributions supported on a lattice of form $\{\boldsymbol{\alpha} + (j_1\Delta_1,\dots,j_d\Delta_d) | j_1,\dots,j_d \text{ integers}\}$. Simple examples of multivariate lattice distributions are binomial and Poisson distributions. More complicated examples are the multivariate distributions of cell counts in contingency tables, and multivariate distributions of sufficient statistics associated with indicator variables in logistic regression. Skovgaard (1987) derives a counterpart of (2) in the lattice case when $\Delta = 1$:

$$1 - \tilde{F}(x^1 | x^2, \dots, x^d) = 1 - \Phi(\sqrt{N}\hat{w})$$

$$(3) \qquad + \phi(\sqrt{N}\hat{w})\left(\frac{\sqrt{|K_{\mathbf{Y}''_{-1}}(\tilde{\boldsymbol{\beta}})|}}{2\sinh\left(\frac{1}{2}\hat{\beta}_1\right)\sqrt{N|K''_{\mathbf{Y}}(\hat{\boldsymbol{\beta}})|}} - \frac{1}{\sqrt{N}\hat{w}}\right).$$

Here x^1 is corrected for continuity when calculating $\hat{\boldsymbol{\beta}}$; that is, if possible, values for X^1 are one unit apart, $\hat{\boldsymbol{\beta}}$ solves $K'_{\mathbf{Y}}(\hat{\boldsymbol{\beta}}) = \tilde{\mathbf{x}}$, where $\tilde{x}^j = x^j$ if $j \neq 1$ and $\tilde{x}^1 = x^1 - \frac{1}{2}$. The relationship between expansions for continuous and lattice probability distributions is discussed by Kolassa and McCullagh (1990).

3. Plan for higher-order approximations. Skovgaard (1987) shows that in some extreme cases (3) behaves poorly. A refinement to this approximation may be derived by seeking a series expansion for D of (1) and by factoring this expansion in such a way that one of the factors is the expansion for the conditional tail probability.

Integration in (1) is performed by changing variables to a new set $\mathbf{w}$ so that the term that is exponentiated is quadratic. The integrand will then be of the form

$$(4) \qquad \exp\left(\frac{N}{2} w_j \delta^{jk} w_k - N w_j \delta^{jk} \hat{w}_k\right)\left|\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{w}}\right| \beta_1^{-1},$$

and D is the d-dimensional integral of (4), times $(2\pi i)^{-d}$. Here and below the symbol δ with superscripts or subscripts or both denotes the array with entries 1 where all indices agree, and zero elsewhere. Note specifically that $w_j \delta^{jk} w_k$ is the inner product of $\mathbf{w}$ with itself. Choose $\mathbf{w}$ so that

$$(5) \qquad \tfrac{1}{2}(w_j - \hat{w}_j)\delta^{jk}(w_k - \hat{w}_k) = K_{\mathbf{Y}}(\boldsymbol{\beta}) - \beta_i x^i - K_{\mathbf{Y}}(\hat{\boldsymbol{\beta}}) + \hat{\beta}_i x^i$$

along the path of integration, and so that w_1 depends only on β_1.

The function $|\partial \boldsymbol{\beta} / \partial \mathbf{w}| / \beta_1$ has a singularity at $w_1 = 0$. As noted by Skovgaard (1987) and others, the accuracy of the result obtained by applying integral expansion techniques directly to $|\partial \boldsymbol{\beta} / \partial \mathbf{w}| / \beta_1$ will deteriorate as $\hat{w}_1$ approaches 0. Express D as the sum of two integrals $D_1 + D_2$, where the integrand of D_2 has no singularity at $\hat{w}_1 = 0$ and where the integration with respect to w_1 in D_1 can be calculated exactly. Specifically, let

$$D_1 = \frac{N^{d-1}}{(2\pi i)^d} \int_{-i\infty}^{+i\infty} \cdots \int_{-i\infty}^{+i\infty} \exp\left(\frac{N}{2} w_j \delta^{jk} w_k - N w_j \delta^{jk} \hat{w}_k\right) \times \left|\frac{\partial \boldsymbol{\beta}_{-1}}{\partial \mathbf{w}_{-1}}\right|(0, \mathbf{w}_{-1}) w_1^{-1}\, d\mathbf{w}.$$

Here $|\partial \boldsymbol{\beta}_{-1}/\partial \mathbf{w}_{-1}|$ is the determinant of the matrix of first derivatives for $\beta_2, \dots, \beta_d$ as a function of $w_2, \dots, w_d$. Integration with respect to w_1 and then with respect to the remaining variables gives $(1 - \Phi(\sqrt{N}\hat{w}_1))f_{\mathbf{X}_{-1}}$. The task remains, then, of expressing D_2 as a series,

$$\begin{aligned} D_2 = {} & \frac{N^{d-1}}{(2\pi i)^d} \int_{-i\infty}^{+i\infty} \cdots \int_{-i\infty}^{+i\infty} \exp\left(\frac{N}{2} w_j \delta^{jk} w_k - N w_j \delta^{jk} \hat{w}_k \right)\Bigg) \\ & \times \left(\frac{|d\boldsymbol{\beta}/d\mathbf{w}|}{\beta_1} - \frac{|(d\boldsymbol{\beta}_{-1}/d\mathbf{w}_{-1})(0, \mathbf{w}_{-1})|}{w_1} \right) d\mathbf{w} \\ = {} & \frac{N^{d-1}}{(2\pi i)^d} \exp\left(-\frac{N}{2} \hat{w}_j \delta^{jk} \hat{w}_k \right) \sum_{s=0}^{m-1} A_s N^{-s} + O(N^{-m}), \end{aligned} \tag{6}$$

and factoring this series such that one factor is $f_{X_2, \dots, X_d}$. The quantities $\hat{w}_j$ are determined below.

The approximations (2) and (3) are the result of a first-order approximation to the integral of (6). To derive higher-order approximations, an asymptotic expansion for a multiple integral of an integrand of the form

$$\exp\left(\frac{N}{2} (w_j - \hat{w}_j) \delta^{jk} (w_k - \hat{w}_k) \right) \theta(\mathbf{w}), \tag{7}$$

with θ analytic, is needed. Note that (4) fails to be of this form, since the factor β_1^{-1} makes it not differentiable at $\beta_1 = 0$. A theorem concerning series expansions of integrals of the forms (4) and (7) will be cited in the next section. These will include a multivariate integral expansion theorem for integrals of the form (7) in terms of derivatives of $\theta(\mathbf{w})$. The following section will present two lemmas showing that the decomposition of D into $D_1 + D_2$ performs as desired; that is, that the singularity of (4) at $\boldsymbol{\beta} = \mathbf{0}$ is a simple pole and that the integrand of D_2 is analytic. These relationships will be usd to generate a series expansion for the quantity D of (1).

The resulting series will then be factored into two series, one of them being the series expansion for the conditioning probability. By the usual saddlepoint density arguments [McCullagh (1987), Chapter 6], the second factor then becomes the series expansion desired.

4. An extension of a theorem of Temme. As suggested by Skovgaard (1987), we begin our investigation of higher-order corrections to the conditional cumulative distribution function by proving a multivariate extension of a univariate theorem due to Temme (1982).

THEOREM 4.1. *Suppose that* $\mathbf{w}$ *is a vector with d components and that the function* θ *is analytic on an open set* Q *of the form* $\prod_j (I_j \times i\mathbf{R})$ *for* I_j *open* *real intervals, satisfying*

$$\frac{d^{v_1+\cdots+v_d}\theta(\mathbf{w})}{d^{v_1}w_1 \cdots d^{v_d}w_d} = O\left(|\mathbf{w}-\hat{\mathbf{w}}|^{\lambda}\exp(\omega|\mathbf{w}-\hat{\mathbf{w}}|^2)\right)$$

as $\min(|\Re(w_j)|) \to \infty$. *Let* $\mathbf{v}! = v_1! \times \cdots \times v_d!$, *and let*

$$\left[\frac{\partial^{\mathbf{v}}}{\partial^{\mathbf{v}}\mathbf{w}}\right] = \left[\frac{\partial^{v_1+\cdots+v_d}}{\partial^{v_1}w_1 \cdots \partial^{v_d}w_d}\right].$$

Let $\int_{-i\infty}^{+i\infty} \cdots \int_{-i\infty}^{+i\infty}$ *be the operator consisting of* d *integrations over smooth curves* $\gamma_j(t)$ *for* $t \in (-1,1)$ *such that* $\lim_{t\to\pm 1}\gamma_j(t) = \pm\infty$, *respectively*, $\gamma_j(0) = \hat{w}_j$, *and* $\Re(\gamma_j(t))$ *is bounded. Then*

$$(8)\qquad \begin{aligned}&\left(\frac{N}{2\pi}\right)^{d/2}\int_{-i\infty}^{+i\infty}\cdots\int_{-i\infty}^{+i\infty}\exp\left(\frac{N}{2}(w_j-\hat{w}_j)\delta^{jk}(w_k-\hat{w}_k)\right)\theta(\mathbf{w})\,d\mathbf{w}\\ &\quad = \sum_{s=0}^{m-1}A_s N^{-s} + E_{m,d}N^{-m},\end{aligned}$$

where

$$A_s = \sum_{\mathbf{v}\in\mathbf{S}(s)}\frac{(-2)^{-s}}{v!}\left[\frac{\partial^{2\mathbf{v}}}{\partial^{2\mathbf{v}}\mathbf{w}}\theta\right](\hat{\mathbf{w}}),$$

with $\mathbf{S}(s) = \{\mathbf{v}\in\mathbf{Z}^d | v_j \ge 0, \sum_j v_j = s\}$, *and*

$$E_{m,d} = \int_{-i\infty}^{+i\infty}\cdots\int_{-i\infty}^{+i\infty}\sum_{\mathbf{S}(m+1)}\frac{(-2)^{-(m+1)}}{\mathbf{v}!}\left[\frac{\partial^{2\mathbf{v}}}{\partial^{2\mathbf{v}}\mathbf{w}}\theta\right](\mathbf{w}^*(\mathbf{w}))\,d\mathbf{w},$$

for $\mathbf{w}^*$ *a convex combination of* $\mathbf{w}$ *and* $\hat{\mathbf{w}}$.

PROOF. Choose $\delta > 0$. Expand θ using a Taylor series:

$$\theta(\mathbf{w}) = \sum_{s=0}^{2m}\sum_{\mathbf{S}(s)}\frac{1}{\mathbf{v}!}\left[\frac{\partial^{\mathbf{v}}}{\partial^{\mathbf{v}}\mathbf{w}}\theta\right](\hat{\mathbf{w}})w_1^{v_1}\cdots w_d^{v_d} + R_{m+1}(\mathbf{w}),$$

for $\mathbf{w}^*$ a linear combination of $\mathbf{w}$ and $\hat{\mathbf{w}}$. Deforming the path of integration to run vertically through each $\hat{w}_j$, and integrating terms not involving $R_{m+1}(\mathbf{w})$ gives the sum in (8). By Taylor's theorem, the function $R_{m+1}(\mathbf{w})$ can be expressed as

$$\sum_{\mathbf{S}(m+1)}\frac{(-2)^{-(m+1)}}{\mathbf{v}!}\left[\frac{\partial^{2\mathbf{v}}}{\partial^{2\mathbf{v}}\mathbf{w}}\theta\right](\mathbf{w}^*),$$

for some $\mathbf{w}^*$ between $\hat{\mathbf{w}}$ and $\mathbf{w}$. □

5. Integral parameterization. A careful choice of the parameterization of $\boldsymbol{\beta}$ in terms of $\mathbf{w}$ to satisfy (5) makes analysis much easier. For real $\boldsymbol{\beta}$ and for $1 \le m \le d$, let

$$
\begin{aligned}
-\tfrac{1}{2}(w_m - \hat{w}_m)^2 &= \min\bigl(K_{\mathbf{Y}}(\boldsymbol{\gamma}) - \gamma_j x^j | \gamma_j = \beta_j \ \forall\, j < m\bigr) \\
&\qquad - \min\bigl(K_{\mathbf{Y}}(\boldsymbol{\gamma}) - \gamma_j x^j | \gamma_j = \beta_j \ \forall\, j \le m\bigr), \\
-\tfrac{1}{2}\hat{w}_m^2 &= \min\bigl(K_{\mathbf{Y}}(\boldsymbol{\gamma}) - \gamma_j x^j | \gamma_j = 0 \ \forall\, j < m\bigr) \\
&\qquad - \min\bigl(K_{\mathbf{Y}}(\boldsymbol{\gamma}) - \gamma_j x^j | \gamma_j = 0 \ \forall\, j \le m\bigr).
\end{aligned} \tag{9}
$$

This parameterization makes proving the following lemmas easy.

Lemma 5.1. *The function* $\boldsymbol{\beta}(\mathbf{w})$ *is analytic in* $\mathbf{w}$ *at* $\hat{\mathbf{w}}$.

Proof. For each m the function $(\gamma_{m+1}, \dots, \gamma_k)$ of $(\beta_1, \dots, \beta_m)$ solving

$$K_{\mathbf{Y}}^{j}(\gamma_1, \dots, \gamma_m, \beta_{m+1}, \dots, \beta_k) = x^j,$$

for $j > m$, exists and is differentiable by the implicit function theorem of complex variables. Here $K_{\mathbf{Y}}^{j}(\boldsymbol{\beta}) = \partial K_{\mathbf{Y}}(\boldsymbol{\beta})/\partial\beta_j$. This choice of $\boldsymbol{\gamma}$ achieves the minimum in (9), and so the left-hand side of (9) is an analytic function of $\boldsymbol{\beta}$ for all m [Bochner and Martin (1948)]. The right-hand side of (9) has zero constant and first-order terms when viewed as a function of β_m at $\hat{\boldsymbol{\beta}}$ and so it can be factored as $(\beta_m - \hat{\beta}_m)^2$ times a function $q_m(\boldsymbol{\beta})$ of $\boldsymbol{\beta}$ analytic at $\hat{\boldsymbol{\beta}}$ and which does not vanish at $\hat{\boldsymbol{\beta}}$. Then w_m satisfies

$$\sqrt{\tfrac{1}{2}}\,(w_m - \hat{w}_m) = \bigl(\beta_m - \hat{\beta}_m\bigr)\sqrt{q_m(\boldsymbol{\beta})}\,.$$

The quantity $q_m(\hat{\boldsymbol{\beta}})$ is easily shown to be positive, and the branch of the square root function assigning a positive result may be used. Since the equation for w_m does not contain $\beta_1, \dots, \beta_{m-1}$, the determinant of the matrix of derivatives of the right-hand sides with respect to $\boldsymbol{\beta}$ is the product of the diagonals; one can easily show that these derivatives, evaluated at $\hat{\boldsymbol{\beta}}$, are nonzero. Hence the requirements of the inverse function theorem for complex variables are satisfied, and the lemma follows. □

Lemma 5.2. *Let* $\hat{\beta}_i^m$ *be the derivative of* β_i *with respect to* w_m, *evaluated at* $\hat{\mathbf{w}}$. *Let* $\hat{\kappa}_{ij}$ *be the generic entry in the inverse of second derivatives of* $K_{\mathbf{Y}}$ *evaluated at* $\hat{\boldsymbol{\beta}}$. *Then* $\hat{\beta}_i^m \delta_{mn} \hat{\beta}_j^n = \hat{\kappa}_{ij}$.

Proof. Differentiating (5) twice with respect to $\mathbf{w}$, evaluating at $\hat{\mathbf{w}}$ and noticing that $K_{\mathbf{Y}}^{i}(\hat{\boldsymbol{\beta}}) = x^i$, gives $\delta^{mn} = \hat{\beta}_i^m \hat{\kappa}^{ij} \hat{\beta}_j^n$. Hence the result follows upon examination of the product $\hat{\beta}_i^m \delta_{mn} \hat{\beta}_j^n \hat{\kappa}^{jk}$. □

Lemma 5.3. *We have*

$$\lim_{w_1 \to 0} \frac{\beta_1}{w_1} \left| \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{w}} \right| = \left| \frac{\partial \boldsymbol{\beta}_{-1}}{\partial \mathbf{w}_{-1}} \right| (0, \mathbf{w}_{-1}).$$

PROOF. For each m, w_m is a function of $\beta_1,\dots,\beta_m$, demonstrating that $|\partial\boldsymbol{\beta}/\partial\mathbf{w}|$ is triangular, and hence the Jacobian is the product of the derivatives; specifically, $|\partial\boldsymbol{\beta}/\partial\mathbf{w}| = |\partial\boldsymbol{\beta}_{-1}/\partial\mathbf{w}_{-1}|\,|\partial\beta_1/\partial w_1|$. The lemma follows by noting that

$$\lim_{w_1\to 0}\frac{\beta_1}{w_1} = \frac{\partial\beta_1}{\partial w_1}(0,\mathbf{w}_{-1}). \qquad \square$$

6. The expansion of the inversion integral. Expressing the integrand of D as the sum of an analytic term and a term that can be integrated exactly, and integrating these terms separately, allows the use of the theorems of the previous section. Let $\mathbf{w}_{-1}$ denote the vector $(w_2,\dots,w_d)$. The quantity D_1 may be expressed as

$$\begin{aligned}&(2\pi i)^{-1}\int_{\hat{w}_1-i\infty}^{\hat{w}_1+i\infty}\exp\left(-\frac{N}{2}w_1^2 - Nw_1\hat{w}_1\right)\frac{dw_1}{w_1}\\ &\quad\times\frac{N^{d-1}}{(2\pi i)^{d-1}}\int_{\hat{w}_2-i\infty}^{\hat{w}_2+i\infty}\cdots\int_{\hat{w}_d-i\infty}^{\hat{w}_d+i\infty}\exp\left(-\frac{N}{2}\sum_{j=2}^{d}\left(w_j^2 - 2w_j\hat{w}_j\right)\right)\\ &\qquad\times\left|\frac{\partial\boldsymbol{\beta}_{-1}}{\partial\mathbf{w}_{-1}}\right|(0,\mathbf{w}_{-1})\,d\mathbf{w}_{-1}.\end{aligned}$$

By reversing the logic used to derive (5), the second factor is exactly the unconditional density of $(x^2,\dots,x^d)$; the first is $1-\Phi(\hat{w}_1)$.

THEOREM 6.1. *The second-order saddlepoint approximation to the conditional cumulative distribution function of the mean of N independent and identically distributed continuous random vectors with a density, accurate to $O(N^{-5/2})$, is*

$$\begin{aligned}&1-\tilde{F}(x^1|x^2,\dots,x^d)\\ &\quad= 1-\Phi\left(\sqrt{N}\hat{w}_1\right)+\phi\left(\sqrt{N}\hat{w}_1\right)\\ &\qquad\times\left(\frac{1}{\left(\sqrt{N}\hat{w}_1\right)^3}-\frac{1}{\sqrt{N}\hat{w}_1}\right.\\ &\qquad\quad+\frac{1}{\sqrt{N}\hat{z}}\left(1+\frac{1}{N}\left[\frac{1}{8}\left(\hat{\rho}_4-\tilde{\rho}_4\right)-\frac{1}{8}\left(\hat{\rho}_{13}^2-\tilde{\rho}_{13}^2\right)\right.\right.\\ &\qquad\qquad\left.\left.\left.-\frac{1}{12}\left(\hat{\rho}_{23}^2-\tilde{\rho}_{23}^2\right)-\frac{1}{2}\frac{\hat{\kappa}_{1k}\hat{\kappa}^{ijk}\hat{\kappa}_{ij}}{\hat{\beta}_1}-\frac{\hat{\kappa}^{11}}{\left(\hat{\beta}_1\right)^2}\right]\right)\right),\end{aligned}\tag{10}$$

where

$$\hat{z} = \hat{\beta}_1 \hat{\sigma} \quad \textit{and} \quad \hat{\sigma} = \sqrt{\det K''_{\mathbf{Y}}(\hat{\boldsymbol{\beta}}) / \det K''_{\mathbf{Y}_{-1}}(0, \hat{\boldsymbol{\beta}}_{-1})}\,.$$

The invariants are given by $\hat{\rho}_{13}^2 = \hat{\kappa}^{gij} \times \hat{\kappa}^{hkl} \hat{\kappa}_{gh} \hat{\kappa}_{ij} \hat{\kappa}_{kl}$, $\hat{\rho}_{23}^2 = \hat{\kappa}^{gij} \hat{\kappa}^{hkl} \hat{\kappa}_{gh} \hat{\kappa}_{il} \hat{\kappa}_{jl}$ *and* $\hat{\rho}_4 = \hat{\kappa}^{ijkl} \hat{\kappa}_{ij} \hat{\kappa}_{kl}$; $\tilde{\rho}_4$, $\tilde{\rho}_{13}^2$ *and* $\tilde{\rho}_{23}^2$ *are the corresponding quantities calculated from* $\tilde{\kappa}_{ij}$, $\tilde{\kappa}^{ijk}$ *and* $\tilde{\kappa}^{ijkl}$. *If these random vector summands lie in the lattice of Section* 2, *the second-order saddlepoint approximation to the conditional cumulative distribution function of a random vector on a lattice, accurate to* $O(N^{-5/2})$, *is*

$$\begin{aligned}
1 - \tilde{F}&(x_1 | x_2, \dots, x_d) \\
&= 1 - \Phi(\sqrt{N}\hat{w}_1) + \phi(\sqrt{N}\hat{w}_1) \\
&\quad \times \Bigg(\frac{1}{(\sqrt{N}\hat{w}_1)^3} - \frac{1}{\sqrt{N}\hat{w}_1} \\
&\quad + \frac{1}{\sqrt{N}\hat{z}} \Bigg(1 + \frac{1}{n} \Bigg[\frac{1}{8}(\hat{\rho}_4 - \tilde{\rho}_4) - \frac{1}{8}(\hat{\rho}_{13}^2 - \tilde{\rho}_{13}^2) \\
&\qquad - \frac{1}{12}(\hat{\rho}_{23}^2 - \tilde{\rho}_{23}^2) - \frac{1}{4}\Delta_1 \hat{\kappa}_{1k} \hat{\kappa}^{ijk} \hat{\kappa}_{ij} \coth\left(\frac{1}{2}\hat{\beta}_1 \Delta_1\right) \\
&\qquad - \left(\frac{1}{4} \coth\left(\frac{1}{2}\hat{\beta}_1\Delta_1\right)^2 - \frac{1}{8} \right) \Delta_1^2 \hat{\kappa}_{11} \Bigg] \Bigg) \Bigg).
\end{aligned} \tag{11}$$

Here $\hat{z} = 2\sinh(\frac{1}{2}\hat{\beta}_1\Delta_1)\hat{\sigma}$, *and* x_1 *is corrected for continuity before calculating* $\hat{\boldsymbol{\beta}}$.

PROOF. Skovgaard (1987) suggests performing the inner integrals to obtain the standard higher-order saddlepoint approximation to the density of the conditioning variables and then applying Temme's theorem to the remaining univariate integral. The present approach of using the extended version of Temme's theorem is analytically more straightforward. Theorem 4.1 is applied once to D_2, and the resulting series is factored. The first will be the series expansion for the density of $\mathbf{X}_{-1}$; the second will then be what is added to $1 - \Phi(\hat{w}_1)$ to produce the desired expansion for the conditional cumulative distribution function.

Using Theorem 4.1, $D_2 = (N/2\pi)^{d/2} \exp(-(N/2)\hat{w}_j \delta^{jk} \hat{w}_k) \sum_{s=0}^{\infty} A_s N^{-s}$, where

$$A_s = \sum_{\mathbf{S}(s)} \frac{(-2)^{-s}}{\mathbf{v}!} \left[\frac{\partial^{2\mathbf{v}}}{\partial^{2\mathbf{v}} \mathbf{w}} \theta \right] (\hat{\mathbf{w}}),$$

where $\theta(\mathbf{w}) = |\partial\boldsymbol{\beta}/\partial\mathbf{w}|(\mathbf{w})/\beta_1 - |\partial\boldsymbol{\beta}_{-1}/\partial\mathbf{w}_{-1}|(0,\mathbf{w}_{-1})/w_1$. Let $h(\mathbf{w})$ be the function of $\mathbf{w}$ whose power series about $\hat{\mathbf{w}}$ is constructed from that of $|\partial\boldsymbol{\beta}_{-1}/\partial\mathbf{w}_{-1}|(0,\mathbf{w}_{-1})$, with all terms involving components of $\mathbf{w}-\hat{\mathbf{w}}$ to odd powers omitted. Let

$$g(\mathbf{w}) = \frac{\theta(\mathbf{w})}{h(\mathbf{w})} = \frac{|\partial\boldsymbol{\beta}/\partial\mathbf{w}|(\mathbf{w})}{h(\mathbf{w})\beta^1} - \frac{|\partial\boldsymbol{\beta}_{-1}/\partial\mathbf{w}_{-1}|(0,\mathbf{w}_{-1})}{h(\mathbf{w})w_1}.$$

Then

$$\begin{aligned}\frac{(-2)^{-s}}{\mathbf{v}!}&\left[\frac{\partial^{2\mathbf{v}}}{\partial^{2\mathbf{v}}\mathbf{w}}\theta\right](\hat{\mathbf{w}})\\ &= \frac{1}{\mathbf{v}!}\sum_{0\le u_j\le 2v_j}\frac{2^{-s}(2\mathbf{v})!}{\mathbf{u}!(2\mathbf{v}-\mathbf{u})!}\left[\frac{\partial^{\mathbf{u}}}{\partial^{\mathbf{u}}\mathbf{w}}h\right](\hat{\mathbf{w}})\left[\frac{\partial^{2\mathbf{v}-\mathbf{u}}}{\partial^{2\mathbf{v}-\mathbf{u}}\mathbf{w}}g\right](\hat{\mathbf{w}}).\end{aligned}$$

Recalling that h has only even-order terms,

$$\begin{aligned}A_s = \sum_{\mathbf{S}(s)}\frac{(2\mathbf{v})!}{\mathbf{v}!}&\sum_{0\le u_j\le v_j}\frac{(-2)^{-s}}{(2\mathbf{u})!(2\mathbf{v}-2\mathbf{u})!}\\ &\times\left[\frac{\partial^{2\mathbf{u}}}{\partial^{2\mathbf{u}}\mathbf{w}}h\right](\hat{\mathbf{w}})\left[\frac{\partial^{2\mathbf{v}-2\mathbf{u}}}{\partial^{2\mathbf{v}-2\mathbf{u}}\mathbf{w}}g\right](\hat{\mathbf{w}}).\end{aligned}\tag{12}$$

If $\{C_s\}$ are coefficients such that $\sum_{s=0}^{m-1}A_sN^{-s} = \sum_{s=0}^{m-1}B_sN^{-s}\times\sum_{s=0}^{m-1}C_sN^{-s} + O(N^{-m})$, where B_s are coefficients in the asymptotic expansion of the density for the conditioning variables,

$$B_s = \sum_{\substack{v_2,\dots,v_k\ge 0\\ \sum_{j=2}^k v_j = s}}\frac{(-2)^{-\sum_{j=2}^d v_j}}{v_2!\cdots v_d!}\left[\frac{\partial^{2v_2+\cdots+2v_d}}{\partial^{2v_2}w_2\cdots\partial^{2v_d}w_d}h\right](\hat{\mathbf{w}})$$

such that $f_{X_2,\dots,X_d} = \sum_{s=0}^{m-1}B_sN^{-s} + O(N^{-m})$, then

$$1-\Phi\big(\sqrt{N}\hat{w}_1\big) + \phi\big(\sqrt{N}\hat{w}_1\big)N^{-1/2}\sum_{s=0}^{m-1}C_sN^{-s}$$

is the required expansion for the conditional tail probability. Skovgaard (1987) provides such a decomposition for $m=1$; that is, he factors the lead term $A_0 = |\partial\boldsymbol{\beta}_{-1}/\partial\mathbf{w}_{-1}|(0,\hat{\mathbf{w}}_{-1})\times g(\hat{\mathbf{w}})$ into the lead term $B_0 = |\partial\boldsymbol{\beta}_{-1}/\partial\mathbf{w}_{-1}|(0,\hat{\mathbf{w}}_{-1})$ in the asymptotic expansion for the density of the conditioning density times a factor $C_0 = 1/\hat{z} - 1/\hat{w}_1$, which consequently contributes to the lead term in the asymptotic expansion for the conditional tail probability desired.

When $m=2$, we are concerned with A_0 and A_1. When $s=0$ or $s=1$, $(2\mathbf{v})!$ in (12) is equal either to $(2\mathbf{u})!$ or $(2\mathbf{v}-2\mathbf{u})!$, or to both. The other quantity is 1. Also, $\mathbf{u}!, \mathbf{v}!$ and $(\mathbf{v}-\mathbf{u})!$ are all 1. Hence, when $s=0$ or $s=1$,

$$A_s = \sum_{\mathbf{S}(s)} \sum_{0 \le u_j \le v_j} \frac{(-2)^{-s}}{(\mathbf{u})!(\mathbf{v}-\mathbf{u})!} \left[\frac{\partial^{2\mathbf{u}}}{\partial^{2\mathbf{u}} \mathbf{w}} h \right] (0, \mathbf{w}_{-1}) \left[\frac{\partial^{2\mathbf{v}-2\mathbf{u}}}{\partial^{2\mathbf{v}-2\mathbf{u}} \mathbf{w}} g \right] (\hat{\mathbf{w}}),$$

indicating that

$$C_1 = \sum_{j=1}^{d} \frac{(-2)^{-1}}{1!} \frac{\partial^2}{\partial w_j^2} g(\hat{\mathbf{w}}) = -\frac{1}{2} \delta_{mn} g^{mn}(\hat{\mathbf{w}}),$$

where superscripts on g indicate derivatives with respect to components of $\mathbf{w}$.

Implicitly differentiating (5) to obtain derivatives of $\boldsymbol{\beta}$ with respect to $\mathbf{w}$, and using standard matrix determinant differentiation formulae [Hocking (1985), Appendix A.II.1.2] and Lemma 5.2,

$$C_1 = -\frac{1}{\hat{z}} \left(\frac{1}{8} \left(\hat{\rho}_{13}^2 - \tilde{\rho}_{13}^2 - \hat{\rho}_4 + \tilde{\rho}_4 \right) + \frac{1}{12} \left(\hat{\rho}_{23}^2 - \tilde{\rho}_{23}^2 \right) \right.$$
$$\left. + \frac{1}{2} \frac{\hat{\kappa}_{1j} \hat{\kappa}^{ijk} \hat{\kappa}_{ik}}{\hat{\beta}_1} + \frac{\hat{\kappa}_{11}}{\left(\hat{\beta}_1 \right)^2} \right) - \frac{1}{\left(\hat{w}_1 \right)^3},$$

completing the proof for the continuous case. See Kolassa (1994) for some details. The lattice case follows by noting that the counterpart to (1) is

$$D(x_1 | x_2, \dots, x_d) = \frac{N^d - 1}{(2\pi i)^d} \int_{-i\pi/\Delta_1}^{i\pi/\Delta_1} \cdots \int_{-i\pi/\Delta_d}^{i\pi/\Delta_d} \exp\left(N \left[K_{\mathbf{X}}(\boldsymbol{\beta}) - \boldsymbol{\beta}^T \mathbf{x} \right] \right)$$
$$\times \left(\frac{\Delta_1}{2} \right) \frac{d\boldsymbol{\beta}}{\sinh(\Delta_1 \beta_1 / 2)}$$

and by expanding the factor involving $\sinh(\Delta_1 \beta_1/2)$. □

7. A logistic regression example. Consider the logistic regression model for binary outcomes X_{ij}:

$$(13) \qquad P(X_{ij} = 1) = \frac{\exp(\mathbf{z}_i \boldsymbol{\theta})}{1 + \exp(\mathbf{z}_i \boldsymbol{\theta})}; \qquad P(X_{ij} = 0) = \frac{1}{1 + \exp(\mathbf{z}_i \boldsymbol{\theta})},$$

for $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, n_i\}$. The quantities $\mathbf{z}_i$ are row vectors of covariates, with d components. Let $Y_i = \sum_j X_{ij}$. Let $\mathbf{Y}$ and $\mathbf{n}$ be the vectors of the number of successes Y_i and the number of binary trials n_i, each with m components. The sufficient statistics $\mathbf{T}$ associated with $\boldsymbol{\beta}$ are $\mathbf{T} = \mathbf{Y}^T \mathbf{Z}$, where $\mathbf{Z}$ is the $m \times d$ matrix whose rows are the covariate vectors $\mathbf{z}_i$ associated with the various groups. Assume that $\mathbf{Z}$ is of full rank. Let $\mathbf{y}$ and $\mathbf{t}$ be observed values of $\mathbf{Y}$ and $\mathbf{T}$, respectively. For any $j \in \{1, \dots, d\}$, a p-value for

TABLE 1
Logistic regression results

	Continuity-corrected one-sided p-values
Normal approximation	0.0260
Double-saddlepoint approximation	0.0315
Higher-order double-saddlepoint approximation	0.0321
Exact	0.0352

a two-sided test that β_j takes on a prespecified value can be calculated by doubling the smaller of

$$(14) \qquad P\big(T^j \geq t^j | \mathbf{T}_{-j} = \mathbf{t}_{-j}; \beta_j\big) \quad \text{and} \quad P\big(T^j \leq t^j | \mathbf{T}_{-j} = \mathbf{t}_{-j}; \beta_j\big),$$

where $\mathbf{T}_{-j}$ and $\mathbf{t}_{-j}$ represent the random vector $\mathbf{T}$ with component j removed and its observed value, respectively. Both the true conditional probability and the Skovgaard approximation are independent of β_{-j}. The approximation depends on β_j through $\hat{w}_j$. Davison (1988) approximates the probabilities in (14) using first-order terms in (11). The cumulant generating function for the sufficient statistic vector $\mathbf{T}$ is given by $K_{\mathbf{T}}(\boldsymbol{\theta}) = \varphi(\boldsymbol{\theta} + \boldsymbol{\beta}) - \varphi(\boldsymbol{\theta})$, where $\varphi(\boldsymbol{\theta}) = \sum_i n_i \log(1 + \exp(\boldsymbol{\theta}\mathbf{z}_i))$. Derivatives of $K_{\mathbf{T}}$ evaluated at $\hat{\boldsymbol{\beta}}$ are equal to the corresponding derivative of φ at $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + \hat{\boldsymbol{\beta}}$. In particular, $\hat{\boldsymbol{\theta}}$ satisfies $\varphi'(\hat{\boldsymbol{\theta}}) = \mathbf{t}$. Similarly, derivatives of $K_{\mathbf{T}}$ evaluated at $\tilde{\boldsymbol{\beta}}$ are equal to the corresponding derivative of φ at $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} + \tilde{\boldsymbol{\beta}}$. In particular, $\hat{\boldsymbol{\theta}}$ satisfies $\varphi^{(r)}(\tilde{\boldsymbol{\theta}}) = t^r$ for $r \neq j$ and $\tilde{\theta}_j = \theta_j$. Mehta, Patel and Senchaudhuri (1993) present an alternative Monte Carlo approach to these calculations.

These methods are applied to a data set of Gordon and Foss (1966). Using this data set, Cox and Snell (1989) model the probability of a baby crying as a function of day and treatment status, using (13), and perform a test of the hypothesis that treatment has a nonzero effect, calculating the probabilities in (14) exactly. Table 1 contains various approximations to these results, including the higher-order double-saddlepoint approximation.

REFERENCES

BEDRICK, E. J. and HILL, J. R. (1992). An empirical assessment of saddlepoint approximations for testing a logistic regression parameter. *Biometrics* **48** 529–544.

BOCHNER, S. and MARTIN, W. T. (1948). *Several Complex Variables*. Princeton Univ. Press.

COX, D. R. and SNELL, E. J. (1989). *Analysis of Binary Data*. Chapman and Hall, London.

DAVISON, A. C. (1988). Approximate conditional inference in generalized linear models. *J. Roy. Statist. Soc. Ser. B* **50** 445–461.

GORDON, T. and FOSS, B. M. (1966). The role of stimulation in the delay of onset of crying in the new-born infant. *Journal of Experimental Psychology* **16** 79–81.

HOCKING, R. R. (1985). *The Analysis of Linear Models*. Brooks/Cole, Monterey, CA.

JENSEN, J. L. (1991). Uniform saddlepoint approximations and log-concave densities. *J. Roy. Statist. Soc. Ser. B* **53** 157–172.

KOLASSA, J. E. (1994). *Series Approximation Methods in Statistics*. Springer, New York.

KOLASSA, J. E. and MCCULLAGH, P. (1990). Edgeworth series for lattice distributions. *Ann. Statist.* **18** 981–985.

KOLASSA, J. E. and TANNER, M. A. (1994). Approximate conditional inference in exponential families via the Gibbs sampler. *J. Amer. Statist. Assoc.* **89** 697–702.

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

MEHTA, C. R., PATEL, N. R. and SENCHAUDHURI, P. (1993). Smart Monte Carlo methods for conditional logistic regression. Unpublished manuscript.

SKOVGAARD, I. M. (1987). Saddlepoint expansions for conditional distributions. *J. Appl. Probab.* **24** 875–887.

TEMME, N. M. (1982). The uniform asymptotic expansion of a class of integrals related to cumulative distribution functions. *SIAM J. Math. Anal.* **13** 239–252.

DEPARTMENT OF STATISTICS
NORTHWESTERN UNIVERSITY
2006 SHERIDAN ROAD
EVANSTON, ILLINOIS 60208