The Annals of Statistics
2003, Vol. 31, No. 1, 274–286

MULTIVARIATE SADDLEPOINT TAIL PROBABILITY APPROXIMATIONS[1]

BY JOHN E. KOLASSA

Rutgers University

This paper presents a saddlepoint approximation to the cumulative distribution function of a random vector. The proposed approximation has accuracy comparable to that of existing expansions valid in two dimensions, and may be applied to random vectors of arbitrary length, subject only to the requirement that the distribution approximated either have a density or be confined to a lattice, and have a cumulant generating function. The result is derived by directly inverting the multivariate moment generating function. The result is applied to sufficient statistics from a regression model with exponential errors, and compared to an existing method in two dimensions. The result is also applied to multivariate inference from a data set arising from a case-control study of endometrial cancer.

1. Introduction. This paper will develop saddlepoint approximations to multivariate tail probabilities for random vectors. The probabilities approximated are of form $\mathrm{P}[\mathbf{T} \geq \mathbf{t}]$, for a random vector $\mathbf{T}$, with vector inequalities understood to hold termwise. These approximations will apply to random vectors whose joint distributions have multivariate moment generating functions, and will be used for inference about parameters in a generalized linear model, specifically in the presence of order restriction. These approximations might also be used to perform approximate conditional inference of the sort suggested by Pierce and Peters (1999); this application of the current result is still in progress. These approximations will be generated by approximating multivariate complex integrals expressing conditional probabilities in terms of the cumulant generating function of the underlying distribution. Use of this approximation requires that $\mathbf{T}$ have a tractable cumulant generating function; any sufficient statistic vector associated with the canonical parameterization of a generalized linear model satisfies this requirement.

Suppose that a random vector $\mathbf{T}$ of length d has a density and a cumulant generating function $\mathcal{K}(\boldsymbol{\tau})$. The next section will demonstrate that

$$\mathrm{P}[\mathbf{T} \geq \mathbf{t}] = \int_{\mathbf{c}-iK}^{\mathbf{c}+iK} \frac{\exp(\mathcal{K}(\boldsymbol{\tau}) - \boldsymbol{\tau}^\top \mathbf{t}^*)}{(2\pi i)^d \prod_{j=1}^{d} \rho(\tau_j)} \, d\boldsymbol{\tau} \tag{1}$$

Received June 2001; revised February 2002.

[1]Supported in part by NSF Grant DMS 00-92-659.

AMS 2000 subject classifications. Primary 62E20; secondary 60E99.

Key words and phrases. Conditional probability, hypergeometric distribution, lattice variable, tail probability.

for any vector $\mathbf{c}$ of positive real numbers in the domain of $\mathcal{K}$, $K = \infty$, $\mathbf{t}^* = \mathbf{t}$, and $\rho(\tau) = \tau$, and that a similar relationship holds when $\mathbf{T}$ is supported on a lattice. Here $\int \cdots d\boldsymbol{\tau}$ represents the multiple complex integral with respect to the components of $\boldsymbol{\tau}$. This paper will present an asymptotic approximation to the right-hand side of (1).

Daniels (1954) presents approximations to densities; these approximations are derived by approximating integrals of the form (1) without the factor $\prod_{j=1}^{d} \rho(\tau_j)$ in the denominator. The quantity $\prod_{j=1}^{d} \rho(\tau_j)$ in the denominator of (1) presents a difficulty, in that as $\mathbf{t}$ moves so that one or more components of $\mathbf{c}$ approach zero, standard integral expansions of (1) become inaccurate. Authors including Skovgaard (1987) and Lugannani and Rice (1980) approached approximation of similar integrals in which only one factor of the form $\rho(\tau_j)$ appears in the denominator of the integrand. In the language of complex variables [Bak and Newman (1982)], this factor represents a simple pole of the otherwise analytic integrand, and the authors remove its effect by subtracting off a function with an identical simple pole at the same location whose integral may be calculated exactly as an evaluation of a standard normal CDF.

In the general case, with multiple factors in the denominator, no such simplification exists, because the resulting poles are not simple. Instead, a quantity will be isolated which will be shown to equal a multivariate normal CDF to relative error of $O(1/n)$. The remainder of the integral will be shown to have an expansion as products of normal densities and multivariate normal CDFs.

When $\mathbf{T}$ takes values on a unit lattice, (1) also holds, with $\rho(\tau) = 2\sinh(\tau/2)$ and $\mathbf{t}$ corrected for continuity. Technical details in approximating (1) in this case will be the same as in the continuous case.

Section 2 derives integral expressions for the probabilities of interest. Section 3 presents a Taylor expansion of the integrand, an argument demonstrating that truncation of the series after a few terms incurs an error of size no larger than $O(1/n)$, a termwise integration of the remainder, and a recursive representation for the terms that arise. The resulting approximation is a generalization of the univariate series of Robinson (1982). Section 4 explores the possibility of generalizing the expansion of Lugannani and Rice (1980) beyond the two-dimensional result of Wang (1990). Section 5 presents examples of the use of the series derived in Section 3, and in two dimensions compares the results with those of Wang (1990).

2. Inversion integrals. In this section I justify the integral relation (1), and reparameterize the integral into a form that facilitates asymptotic expansion. The proof is included primarily because it motivates an important regularity condition on the approximation derived in later sections.

LEMMA. *Suppose that $T^1, \ldots, T^d$ has a cumulant generating function $\mathcal{K}$, and is either continuous* [*i.e.*, $\mathbf{T}$ *has a bounded density, in which case $\rho(\tau)$ is set* *to* τ, $K=\infty$, *and* $\mathbf{t}=\mathbf{t}^*$] *or confined to an integer lattice* [*in which case* $\rho(\tau)$ *is set to* $2\sinh(\tau/2)$, $K=\pi$, *and* $\mathbf{t}=\mathbf{t}^*+\frac{1}{2}\mathbf{1}$]. *Choose* $\mathbf{c}>\mathbf{0}$ *in the domain of* $\mathcal{K}$. *Then relation* (1) *holds.*

PROOF. When $\mathbf{T}$ has a density, standard Fourier inversion techniques imply that for $K=\infty$,

$$f_{\mathbf{T}}(\mathbf{t})=\int_{\mathbf{c}-iK}^{\mathbf{c}+iK}(2\pi i)^{-d}\exp\big(\mathcal{K}(\boldsymbol{\tau})-\tau_j t^j\big)\,d\boldsymbol{\tau}$$

and

$$\begin{aligned}&\int_{\mathbf{c}-i\infty}^{\mathbf{c}+i\infty}\int_{\mathbf{u}\geq\mathbf{t}}\big|\exp\big(\mathcal{K}(\boldsymbol{\tau})-\tau_j u^j\big)\big|\,d\mathbf{u}\,d\boldsymbol{\tau}\\ &\qquad=\int_{\mathbf{c}-i\infty}^{\mathbf{c}+i\infty}\int_{\mathbf{u}\geq\mathbf{t}}|\exp(\mathcal{K}(\boldsymbol{\tau}))|\exp(-c_j u^j)\,d\mathbf{u}\,d\boldsymbol{\tau}\\ &\qquad=\left[\int_{\mathbf{c}-i\infty}^{\mathbf{c}+i\infty}|\exp(\mathcal{K}(\boldsymbol{\tau}))|\,d\boldsymbol{\tau}\right]\left[\int_{\mathbf{u}\geq\mathbf{t}}\exp(-c_j u^j)\,d\mathbf{u}\right].\end{aligned}$$

Here and below, a product containing the same index as a subscript and as a superscript denotes summation over that index. Furthermore, superscripts on t indicate component rather than power, and superscripts on functions denote differentiation with respect to the corresponding component of the argument. The first factor above is finite since $\mathbf{T}$ has a density, and the second is $\exp(c_j t^j)/\prod_{j=1}^d c_j$. By Fubini's theorem, the result follows by interchanging the order of integration with respect to $\mathbf{t}$ and with respect to $\boldsymbol{\tau}$, as long as all components of $\mathbf{c}$ are positive. When $\mathbf{T}$ is confined to an integer lattice, $K=\pi$ and the integration with respect to $\mathbf{u}$ is replaced by summation. □

The requirement that $\mathbf{c}>\mathbf{0}$, needed to justify Fubini's theorem, is important; because of this requirement, tail probabilities associated with $\mathbf{t}$ for which one or more of the components of $\hat{\boldsymbol{\tau}}$ are negative must be calculated by applying the approximation to random vectors with some of their components negated, and differencing.

Consider the special case of the normal distribution. Let $\bar{\mathfrak{N}}(\mathbf{t};\boldsymbol{\Sigma})$ be the upper corner probability for a random variable with a multivariate normal distribution, with mean $\mathbf{0}$ and variance $\boldsymbol{\Sigma}$. From (1),

$$\bar{\mathfrak{N}}(\mathbf{t},\boldsymbol{\Sigma})=\frac{1}{(2\pi i)^d}\int_{c_1-i\infty}^{c_1+i\infty}\cdots\int_{c_d-i\infty}^{c_d+i\infty}\exp(\tau_j\Sigma^{jk}\tau_k/2-\tau_j t^j)\,d\boldsymbol{\tau}\Big/\prod_{j=1}^d\tau_j, \tag{2}$$

for any positive vector $\mathbf{c}$.

3. Integral expansion. Suppose that $\mathbf{T}$ is the mean of n independent random vectors, each with cumulant generating function $\mathcal{K}$. Then $\mathbf{T}$ has as its cumulant generating function $n\mathcal{K}(\boldsymbol{\tau}/n)$. Choose a compact subset $\mathcal{C}$ of the range of $\mathcal{K}'$. Let $c = \sup_{\mathbf{t}\in\mathcal{C}} \hat{\tau}_j \mathcal{K}^{jk}(\hat{\boldsymbol{\tau}})\hat{\tau}_k$. Existence of a bounded density corresponding to the cumulant generating function $\mathcal{K}(\boldsymbol{\tau}) - \tau_j t^j - \mathcal{K}(\hat{\boldsymbol{\tau}}) + \hat{\tau}_j t^j$ implies integrability of $|\exp(\mathcal{K}(\boldsymbol{\tau}) - \tau_j t^j - \mathcal{K}(\hat{\boldsymbol{\tau}}) + \hat{\tau}_j t^j)|$; see Kolassa [(1997), Theorem 2.4.2] for the univariate case. Hence one might choose $\varepsilon > 0$ so that $\sup_{\mathbf{t}\in\mathcal{C}, \|\boldsymbol{\tau}-\hat{\boldsymbol{\tau}}(\mathbf{t})\|>\varepsilon} |\exp(\mathcal{K}(\boldsymbol{\tau}) - \tau_j t^j)| < \exp(-c)$; here $\|\cdot\|$ is the sup norm, and $\hat{\boldsymbol{\tau}}(\mathbf{t})$ is the solution to $\mathcal{K}'(\hat{\boldsymbol{\tau}}) = \mathbf{t}$. Below the dependence of $\hat{\boldsymbol{\tau}}$ on $\mathbf{t}$ is not generally made explicit.

Hence, using (1),

$$\text{(3)}\quad \mathrm{P}[\mathbf{T} \ge \mathbf{t}] = \int_{\hat{\boldsymbol{\tau}}-i\varepsilon}^{\hat{\boldsymbol{\tau}}+i\varepsilon} \frac{\exp(n[\mathcal{K}(\boldsymbol{\tau}) - \tau_j t^j])}{(2\pi i)^d \prod_{j=1}^d \rho(\tau_j)}\, d\boldsymbol{\tau} + \exp\bigl(n[\mathcal{K}(\hat{\boldsymbol{\tau}}) - \hat{\tau}_j t^j]\bigr) E_1,$$

with $|E_1| \le \exp(-nc)$ for some $c > 0$. Expand $\mathcal{K}$ about $\hat{\boldsymbol{\tau}}$, to obtain

$$\begin{aligned}
\mathrm{P}[\mathbf{T} \ge \mathbf{t}] &= \exp\bigl(n[\mathcal{K}(\hat{\boldsymbol{\tau}}) - \hat{\tau}_j t^j]\bigr) \\
&\quad \times \Biggl\{ \int_{\hat{\boldsymbol{\tau}}-i\varepsilon}^{\hat{\boldsymbol{\tau}}+i\varepsilon} \frac{\exp(n[(\tau_j - \hat{\tau}_j)\mathcal{K}^{jk}(\hat{\boldsymbol{\tau}})(\tau_k - \hat{\tau}_k)/2])}{(2\pi i)^d \prod_{j=1}^d \rho(\tau_j)} \\
&\qquad \times \biggl[1 + \frac{n}{6}\mathcal{K}^{klm}(\hat{\boldsymbol{\tau}})(\tau_k - \hat{\tau}_k)(\tau_l - \hat{\tau}_l)(\tau_m - \hat{\tau}_m) \\
&\qquad\quad + \frac{n}{24}\mathcal{K}^{jklm}(\boldsymbol{\tau}^\dagger)(\tau_j - \hat{\tau}_j)(\tau_k - \hat{\tau}_k) \\
&\qquad\quad\quad \times (\tau_l - \hat{\tau}_l)(\tau_m - \hat{\tau}_m) \\
&\qquad\quad + \frac{n^2}{72}\mathcal{K}^{jkl}(\boldsymbol{\tau}^\dagger)\mathcal{K}^{mpq}(\boldsymbol{\tau}^\dagger)(\tau_j - \hat{\tau}_j) \\
\text{(4)}\qquad &\qquad\quad\quad \times (\tau_k - \hat{\tau}_k)(\tau_l - \hat{\tau}_l) \\
&\qquad\qquad\qquad \times (\tau_m - \hat{\tau}_m)(\tau_p - \hat{\tau}_p)(\tau_q - \hat{\tau}_q)\biggr] d\boldsymbol{\tau} + E_1 \Biggr\} \\
&= \exp\bigl(n[\mathcal{K}(\hat{\boldsymbol{\tau}}) - \hat{\tau}_j t^j + \hat{\tau}_j \mathcal{K}^{jk}(\hat{\boldsymbol{\tau}})\hat{\tau}_k/2]\bigr) \\
&\quad \times \Biggl[\int_{\hat{\boldsymbol{\tau}}-i\varepsilon}^{\hat{\boldsymbol{\tau}}+i\varepsilon} \frac{\exp(n[\tau_j \mathcal{K}^{jk}(\hat{\boldsymbol{\tau}})\tau_k/2 - \tau_j \mathcal{K}^{jk}(\hat{\boldsymbol{\tau}})\hat{\tau}_k])}{(2\pi i)^d \prod_{j=1}^d \rho(\tau_j)} \\
&\qquad \times \biggl[1 + \frac{n}{6}\mathcal{K}^{klm}(\hat{\boldsymbol{\tau}})(\tau_k - \hat{\tau}_k)(\tau_l - \hat{\tau}_l)(\tau_m - \hat{\tau}_m)\biggr] d\boldsymbol{\tau} \\
&\qquad\qquad + \exp\bigl(-n\hat{\tau}_j \mathcal{K}^{jk}(\hat{\boldsymbol{\tau}})\hat{\tau}_k/2\bigr) E_2/n \Biggr],
\end{aligned}$$

for $\boldsymbol{\tau}^\dagger$ between $\hat{\boldsymbol{\tau}}$ and $\boldsymbol{\tau}$, and

$$E_2 = \int_{\hat{\boldsymbol{\tau}}-i\varepsilon}^{\hat{\boldsymbol{\tau}}+i\varepsilon} \frac{\exp(n[(\tau_j-\hat{\tau}_j)\mathcal{K}^{jk}(\hat{\boldsymbol{\tau}})(\tau_k-\hat{\tau}_k)/2])}{(2\pi i)^d \prod_{j=1}^d \rho(\tau_j)}$$
$$\times \left[\frac{n}{24}\mathcal{K}^{jklm}(\boldsymbol{\tau}^\dagger)(\tau_j-\hat{\tau}_j)(\tau_k-\hat{\tau}_k)(\tau_l-\hat{\tau}_l)(\tau_m-\hat{\tau}_m)\right.$$
$$+\frac{n^2}{72}\mathcal{K}^{jkl}(\boldsymbol{\tau}^\dagger)\mathcal{K}^{mpq}(\boldsymbol{\tau}^\dagger)(\tau_j-\hat{\tau}_j)(\tau_k-\hat{\tau}_k)$$
$$\left.\times(\tau_l-\hat{\tau}_l)(\tau_m-\hat{\tau}_m)(\tau_p-\hat{\tau}_p)(\tau_q-\hat{\tau}_q)\right]d\boldsymbol{\tau} + E_1.$$

Consider one of the terms constituting $\mathcal{K}^{jklm}(\boldsymbol{\tau}^\dagger)(\tau_j-\hat{\tau}_j)(\tau_k-\hat{\tau}_k)(\tau_l-\hat{\tau}_l)\times(\tau_m-\hat{\tau}_m)$, without including the leading factor n. Let $\mathcal{S}$ denote the set of unique superscripts of $\mathcal{K}^{jklm}(\boldsymbol{\tau}^\dagger)$. Let

$$g(\boldsymbol{\tau}) = \mathcal{K}^{jklm}(\boldsymbol{\tau}^\dagger)\prod_{r\in\mathcal{S}}(\tau_r-\hat{\tau}_r)\prod_{r\in\mathcal{S}^c}\tau_r \Big/ \prod_{r=1}^d \rho(\tau_r).$$

Let $\mathcal{R}$ be the duplicated superscripts of $\mathcal{K}^{jklm}(\boldsymbol{\tau}^\dagger)$. For example, if the superscripts are 1, 2, 2 and 3, then $\mathcal{S}=\{1,2,3\}$, and $\mathcal{R}=(2)$, and if the superscripts are 1, 1, 1 and 2, then $\mathcal{S}=\{1,2\}$, and $\mathcal{R}=(1,1)$. Let $G=\sup_{\mathbf{t}(\Re(\boldsymbol{\tau}))\in\mathcal{C},\Im(\boldsymbol{\tau})\in[-\varepsilon,\varepsilon]^d}|g(\boldsymbol{\tau})|<\infty$. Choose V such that $\mathcal{K}''(\boldsymbol{\tau})-V$ is positive definite for all $\mathbf{t}(\boldsymbol{\tau})\in\mathcal{C}$. Then the error term is bounded by

$$\int_{\hat{\boldsymbol{\tau}}-i\varepsilon}^{\hat{\boldsymbol{\tau}}+i\varepsilon}\frac{\exp(n[(\tau_j-\hat{\tau}_j)V^{jk}(\tau_k-\hat{\tau}_k)/2])}{(2\pi i)^d}G\prod_{r\in\mathcal{R}}(\tau_r-\hat{\tau}_r)\Big/\prod_{r\in\mathcal{S}^c}\tau_r\,d\boldsymbol{\tau}.$$

If V is diagonal, the above integral may be factored into d univariate integrals. Integrals with respect to components with indices in $\mathcal{S}^c$ are bounded; this may be verified by noting that the integral is equal to the ratio of a normal tail probability to a normal density and using standard results involving Mills' ratio or by noting that $\int_{\hat{\tau}_i-i\varepsilon}^{\hat{\tau}_i+i\varepsilon}d\tau_i/\tau_i=\int_{-\varepsilon}^{\varepsilon}\hat{\tau}_i\,d\tau_i/(\tau_i^2+\hat{\tau}_i^2)$. The latter integral may be bounded by dividing the range of integration into bins of length $|\hat{\tau}_i|$ starting at zero. The maximum of the integrand in a bin of form $(k|\hat{\tau}_i|,(k+1)|\hat{\tau}_i|]$ is $|\hat{\tau}_i|/[(1+k^2)\hat{\tau}_i^2]$, and the contribution to the integral from this bin is $\hat{\tau}_i^2/[(1+k^2)\hat{\tau}_i^2]$. The negative portion of the range of integration behaves symmetrically, and so the entire integral is bounded by $2\sum_{k=0}^{\infty}1/(1+k^2)$, a finite constant independent of $\hat{\tau}_i$. Each of the remaining integrals contributes a factor of $1/\sqrt{n}$, plus an additional factor of $1/\sqrt{n}$ for each time the index appears in $\mathcal{R}$. There are four of these factors. Hence the term is of size $1/n^2$. Taking into account the leading n, the sum of terms of this sort contributes an error of size $1/n$. The terms with six superscripts on $\mathcal{K}$ are handled similarly. This term absorbs $\exp(n[\hat{\tau}_j\mathcal{K}^{jk}(\hat{\boldsymbol{\tau}})\hat{\tau}_k/2])E_1$, which has absolute value less than $\exp(-nc/2)$, by our choice of ε. Hence E_2 is uniformly of size $O(1/n)$.

Nonerror terms in expression (4) will be integrated termwise. For $\mathbf{r}, \mathbf{m} \in \mathfrak{Z}^d$, and for $\mathbf{\Sigma}$ a positive definite matrix, let

$$I(\mathbf{\Sigma}, \mathbf{r}, \mathbf{m}, \mathbf{t}, \hat{\boldsymbol{\tau}}) = \int_{\hat{\boldsymbol{\tau}}-i\infty}^{\hat{\boldsymbol{\tau}}+i\infty} \frac{\exp(\tau_j \mathbf{\Sigma}^{jk} \tau_k/2 - \tau_j \delta^{jk} t_k)}{(2\pi i)^d} \prod_{l=1}^{d} [\tau_l^{r_l} (\tau_l - \hat{\tau}_l)^{m_l}]\, d\boldsymbol{\tau}.$$

Here δ^{jk} is 1 if $j = k$ and zero otherwise. Temporarily restrict attention to the case when $T_1, \dots, T_d$ has a continuous distribution, and hence $\rho(\tau) = \tau$. Let $\mathbf{s}$ be the vector such that $s_j = -1$ for all j. Let

$$\begin{aligned} Q(\mathbf{t}) &= \exp\big(n[\mathcal{K}(\hat{\boldsymbol{\tau}}) - \hat{\tau}_j t^j + \hat{\tau}_j \mathcal{K}^{jk}(\hat{\boldsymbol{\tau}})\hat{\tau}_k/2]\big) \\ &\quad \times \big[I\big(n\mathcal{K}''(\hat{\boldsymbol{\tau}}), \mathbf{s}, \mathbf{0}, n\mathcal{K}''(\hat{\boldsymbol{\tau}})\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\tau}}\big) \\ &\qquad + \tfrac{1}{6}\mathcal{K}^{jkl}(\hat{\boldsymbol{\tau}}) I\big(n\mathcal{K}''(\hat{\boldsymbol{\tau}}), \mathbf{s}, \mathbf{e}_j + \mathbf{e}_k + \mathbf{e}_l, n\mathcal{K}''(\hat{\boldsymbol{\tau}})\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\tau}}\big)/\sqrt{n}\big], \end{aligned}$$

where $\mathbf{e}_j$ is the vector with every component 0 except for component j, which is 1. Then

$$(5)\qquad \mathrm{P}[\mathbf{T} \geq \mathbf{t}] = Q(\mathbf{t}) + \exp\big(n[\mathcal{K}(\hat{\boldsymbol{\tau}}) - \hat{\tau}_j t^j]\big) E_2(\mathbf{t})/n \qquad \text{for } \sup_{\mathbf{t}\in\mathfrak{C}} |E_2(\mathbf{t})| < \infty.$$

Saddlepoint approximations to densities typically yield an error term that is relative; that is, the ratio of the true density to the approximation may be expressed as one plus a negative power of the sample size times a term that is uniformly bounded as n increases and $\mathbf{t}$ varies, at least within a compact set such as $\mathfrak{C}$. Many authors, including Routledge and Tsao (1995), describe such results. Achieving a uniform relative bound on tail probability approximations is more difficult, even in one dimension. These approximations are typically of form $a(\mathbf{t}, n)\bar{\mathfrak{N}}(\sqrt{n}v(\mathbf{t})) + b(\mathbf{t}, n)\mathfrak{n}(\sqrt{n}v(\mathbf{t}))$, for functions a, b and v; see, for example, Robinson (1982), Daniels (1987) and Kolassa (1998). The error term typically has a bound of form $n^{-\alpha}C\mathfrak{n}(\sqrt{n}v(\mathbf{t}))$, for some constant C. Unfortunately, $\bar{\mathfrak{N}}(\sqrt{n}v(\mathbf{t}))/\mathfrak{n}(\sqrt{n}v(\mathbf{t})) \to 0$ as $n \to \infty$, and so uniformity of the relative error fails. The error bound in $Q(\mathbf{t})$ is of this form; the error is uniformly exponentially small, but not strictly speaking both relative and uniform.

Evaluation of $Q(\mathbf{t})$ requires evaluation of $I(\mathbf{\Sigma}, \mathbf{r}, \mathbf{m}, \mathbf{t}, \hat{\boldsymbol{\tau}})$ for $\mathbf{r}$ a vector of integers no smaller than -1. For any $\mathbf{r}, \mathbf{m} \in \mathfrak{Z}^d$, and $j \in \{1, \dots, d\}$,

$$(6)\quad I(\mathbf{\Sigma}, \mathbf{r}, \mathbf{m}, \mathbf{t}, \hat{\boldsymbol{\tau}}) = I(\mathbf{\Sigma}, \mathbf{r} + \mathbf{e}_j, \mathbf{m} - \mathbf{e}_j, \mathbf{t}, \hat{\boldsymbol{\tau}}) - \hat{\tau}_j I(\mathbf{\Sigma}, \mathbf{r}, \mathbf{m} - \mathbf{e}_j, \mathbf{t}, \hat{\boldsymbol{\tau}}).$$

This recursion may be continued until for each j, $m_j = 0$. Alternatively, one might expand each of the factors $(\tau_j - \hat{\tau}_j)^{m_j}$ using the binomial theorem and integrate termwise. Manipulating (2),

$$(7)\qquad I(\mathbf{\Sigma}, \mathbf{r}, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}}) = \prod_{j=1}^{d} (-1)^{r_j+1} \frac{d^{r_j+1}}{(dt_j)^{r_j+1}} \bar{\mathfrak{N}}(\mathbf{t}, \mathbf{\Sigma}).$$

The integrals $I(\boldsymbol{\Sigma}, \mathbf{r}, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}})$ will be evaluated by expressing them in closed form when $\mathbf{r}$ takes on only values in $\{-1, 0\}$, and by providing a recursive representation for other values of $\mathbf{r}$. For $\mathcal{A}, \mathcal{B} \subset \mathfrak{F} = \{1, \ldots, d\}$, and for vector $\mathbf{t}$ of length d and $d \times d$ matrix $\boldsymbol{\Sigma}$, let $\mathbf{t}^{\mathcal{A}}$ be the components of $\mathbf{t}$ with indices in $\mathcal{A}$, let $\boldsymbol{\Sigma}^{\mathcal{A}}$ the elements of $\boldsymbol{\Sigma}$ with row and column indices in $\mathcal{A}$, if any, and $\boldsymbol{\Sigma}^{\mathcal{A},\mathcal{B}}$ the elements of $\boldsymbol{\Sigma}$ with row indices in $\mathcal{A}$ and column indices in $\mathcal{B}$, if any. Let $\boldsymbol{\Sigma}_{\mathcal{A},\mathcal{B}}$ be the corresponding entries in $\boldsymbol{\Sigma}^{-1}$. Let $\mathcal{A} = \{j | r_j = 0\}$.

Differentiation of the d dimensional normal tail probability with respect to components whose indices are in $\mathcal{A}$ yields the marginal density for components in $\mathcal{A}$ times the conditional tail probability of components in $\mathcal{A}^{\mathrm{c}}$ conditional on those in $\mathcal{A}$, all evaluated at $\mathbf{t}$, and hence

$$I(\boldsymbol{\Sigma}, \mathbf{r}, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}}) = \mathfrak{n}(\mathbf{t}^{\mathcal{A}}, \boldsymbol{\Sigma}^{\mathcal{A}})\bar{\mathfrak{N}}\big(\mathbf{t}^{\mathcal{A}^{\mathrm{c}}} + (\boldsymbol{\Sigma}_{\mathcal{A}^{\mathrm{c}}})^{-1}\boldsymbol{\Sigma}_{\mathcal{A}^{\mathrm{c}},\mathcal{A}}\mathbf{t}^{\mathcal{A}}, \boldsymbol{\Sigma}_{\mathcal{A}^{\mathrm{c}}}\big) \quad \text{for } r_j \in \{0, -1\}\ \forall\, j,\ \mathcal{A} = \{j | r_j = 0\}. \tag{8}$$

If $\mathbf{r} \in \mathfrak{Z}^d$ such that $r_l \in \{-1, 0\}\ \forall\, l$ and $r_j = r_k = 0$, then

$$\begin{aligned} I(\boldsymbol{\Sigma}, \mathbf{r} + \mathbf{e}_j, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}}) &= -\frac{d}{dt_j} I(\boldsymbol{\Sigma}, \mathbf{r}, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}}) \\ &= [(\boldsymbol{\Sigma}^{\mathcal{A}})^{-1}\mathbf{t}^{\mathcal{A}}]_j I(\boldsymbol{\Sigma}, \mathbf{r}, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}}) \\ &\quad + \sum_{l \in \mathcal{A}^{\mathrm{c}}} I(\boldsymbol{\Sigma}, \mathbf{r} + \mathbf{e}_l, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}})\big[(\boldsymbol{\Sigma}_{\mathcal{A}^{\mathrm{c}}})^{-1}\boldsymbol{\Sigma}_{\mathcal{A}^{\mathrm{c}},\{j\}}\big]_l, \end{aligned} \tag{9}$$

$$\begin{aligned} I(\boldsymbol{\Sigma}, \mathbf{r} + \mathbf{e}_j + \mathbf{e}_k, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}}) &= -\frac{d}{dt_k} I(\boldsymbol{\Sigma}, \mathbf{r} + \mathbf{e}_j, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}}) \\ &= [(\boldsymbol{\Sigma}^{\mathcal{A}})^{-1}\mathbf{t}^{\mathcal{A}}]_j I(\boldsymbol{\Sigma}, \mathbf{r} + \mathbf{e}_k, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}}) - (\boldsymbol{\Sigma}^{\mathcal{A}})^{-1}_{jk} I(\boldsymbol{\Sigma}, \mathbf{r}, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}}) \\ &\quad + \sum_{m \in \mathcal{A}^{\mathrm{c}}} I(\boldsymbol{\Sigma}, \mathbf{r} + \mathbf{e}_m + \mathbf{e}_k, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}})\big[(\boldsymbol{\Sigma}_{\mathcal{A}^{\mathrm{c}}})^{-1}\boldsymbol{\Sigma}_{\mathcal{A}^{\mathrm{c}},\{m\}}\big]_j \end{aligned} \tag{10}$$

and

$$\begin{aligned} I(\boldsymbol{\Sigma}, \mathbf{r} + \mathbf{e}_j + \mathbf{e}_k + \mathbf{e}_l, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}}) &= -\frac{d}{dt_l} I(\boldsymbol{\Sigma}, \mathbf{r} + \mathbf{e}_j + \mathbf{e}_k, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}}) \\ &= (\boldsymbol{\Sigma}^{\mathcal{A}})^{-1}_{jl} I(\boldsymbol{\Sigma}, \mathbf{r} + \mathbf{e}_k, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}}) \\ &\quad - [(\boldsymbol{\Sigma}^{\mathcal{A}})^{-1}\mathbf{t}^{\mathcal{A}}]_j I(\boldsymbol{\Sigma}, \mathbf{r} + \mathbf{e}_k + \mathbf{e}_l, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}}) + (\boldsymbol{\Sigma}^{\mathcal{A}})^{-1}_{jk} I(\boldsymbol{\Sigma}, \mathbf{r} + \mathbf{e}_l, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}}) \\ &\quad - \sum_{m \in \mathcal{A}^{\mathrm{c}}} I(\boldsymbol{\Sigma}, \mathbf{r} + \mathbf{e}_m + \mathbf{e}_k + \mathbf{e}_l, \mathbf{0}, \mathbf{t}, \hat{\boldsymbol{\tau}})\big[(\boldsymbol{\Sigma}_{\mathcal{A}^{\mathrm{c}}})^{-1}\boldsymbol{\Sigma}_{\mathcal{A}^{\mathrm{c}},\{m\}}\big]_j. \end{aligned} \tag{11}$$

Hence (6) and (8)–(11) allow for the recursive calculation of the quantities in $Q(\mathbf{t})$.

Expanding $\mathcal{K}$ about $\mathbf{0}$ rather than $\hat{\boldsymbol{\tau}}$ causes (4) to be replaced by

$$\begin{aligned} Q^*(\mathbf{t}) &= \int_{-i\infty}^{+i\infty} \frac{\exp(n[\tau_j \mathcal{K}^{jk}(\mathbf{0}) t_k/2 - \tau_j \delta^{jk} \hat{\tau}_k])}{(2\pi i)^d \prod_{j=1}^d \rho(\tau_j)} [1 + n\mathcal{K}^{klm}(\mathbf{0})\tau_k\tau_l\tau_m]\, d\boldsymbol{\tau} \\ &\qquad + O(1/n), \\ &= \Big[I\big(n\mathcal{K}''(\mathbf{0}), \mathbf{s}, \mathbf{0}, n\mathbf{t}, \mathbf{0}\big) \\ &\qquad + \frac{\mathcal{K}^{jkl}(\mathbf{0})}{6\sqrt{n}} I\big(n\mathcal{K}''(\mathbf{0}), \mathbf{s} + \mathbf{e}_j + \mathbf{e}_k + \mathbf{e}_l, \mathbf{0}, n\mathbf{t}, \mathbf{0}\big) + O(1/n) \Big]. \end{aligned} \tag{12}$$

Here without loss of generality I take $\mathcal{K}(0) = 0$. Expansion (12) is the well-known Edgeworth expansion for Q^*, and is valid even when $\mathcal{K}(\boldsymbol{\tau})$ exists only for pure imaginary arguments. It is also valid when $\mathbf{T}$ is confined to a unit lattice and $t_1, \ldots, t_d$ is evaluated at continuity corrected points [Kolassa (1989)]. In this case $\rho(\tau) = 2\sinh(\tau/2)$, and linear terms generated by expanding $\tau/\rho(\tau)$ are zero. The counterpart of $Q(\mathbf{t})$ when $n(T_1, \ldots, T_d)$ are supported on a unit lattice is

$$\begin{aligned} Q(\mathbf{t}) &= \exp\big(n[\mathcal{K}(\hat{\boldsymbol{\tau}}) - \hat{\tau}_j t^j]\big) \\ &\quad \times \Big[\exp\big(n\hat{\tau}_j \mathcal{K}^{jk}(\hat{\boldsymbol{\tau}})\hat{\tau}_k/2\big) \\ &\qquad \times \int_{\hat{\boldsymbol{\tau}}-i\infty}^{\hat{\boldsymbol{\tau}}+i\infty} \frac{\exp(n[\tau_j\mathcal{K}^{jk}(\hat{\boldsymbol{\tau}})\tau_k/2 - \tau_j\mathcal{K}^{jk}(\hat{\boldsymbol{\tau}})\hat{\tau}_k])}{(2\pi i)^d \prod_{j=1}^d [\rho(\hat{\tau}_j)/\hat{\tau}_j]} \\ &\qquad \times \left\{ 1 + \sum_{j=1}^d \left[\frac{1}{\hat{\tau}_j} - \frac{\cosh(\hat{\tau}_j/2)}{2\sinh(\hat{\tau}_j/2)} \right] (\tau_j - \hat{\tau}_j) \right\} \\ &\qquad \times \big[1 + n\mathcal{K}^{klm}(\hat{\boldsymbol{\tau}})(\tau_k - \hat{\tau}_k)(\tau_l - \hat{\tau}_l)(\tau_m - \hat{\tau}_m)\big]\, d\boldsymbol{\tau} \Big] \\ &= \exp\big(n[\mathcal{K}(\hat{\boldsymbol{\tau}}) - \hat{\tau}_j t^j]\big) \\ &\quad \times \prod_{j=1}^d \frac{\hat{\tau}_j}{2\sinh(\hat{\tau}_j/2)} \\ &\qquad \times \Big[\exp\big(n\hat{\tau}_j\mathcal{K}^{jk}(\hat{\tau})\hat{\tau}/2\big) I\big(n\mathcal{K}''(\hat{\boldsymbol{\tau}}), \mathbf{s}, \mathbf{0}, n\mathcal{K}''(\hat{\boldsymbol{\tau}})\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\tau}}\big) \\ &\qquad + \sum_{j=1}^d \left[\frac{1}{\hat{\tau}_j} - \frac{\cosh(\hat{\tau}_j/2)}{2\sinh(\hat{\tau}_j/2)} \right] I\big(n\mathcal{K}''(\hat{\boldsymbol{\tau}}), \mathbf{s}, \mathbf{e}_j, n\mathcal{K}''(\hat{\boldsymbol{\tau}})\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\tau}}\big) \\ &\qquad + \frac{1}{6}\mathcal{K}^{jkl}(\hat{\boldsymbol{\tau}}) I\big(n\mathcal{K}''(\hat{\boldsymbol{\tau}}), \mathbf{s}, \mathbf{e}_j + \mathbf{e}_k + \mathbf{e}_l, n\mathcal{K}''(\hat{\boldsymbol{\tau}})\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\tau}}\big)/\sqrt{n} \Big]. \end{aligned} \tag{13}$$

The argument justifying (5) was general enough to justify (5) in this case as well. Note that $1/\hat{\tau} - \cosh(\hat{\tau}/2)/2\sinh(\hat{\tau}/2)$ evaluated at $\hat{\tau} = 0$ is zero.

Approximation $Q(\mathbf{t})$ holds only when all components of $\hat{\boldsymbol{\tau}}$ are positive. Approximations for other $\mathbf{t}$ may be calculated recursively. Specifically, suppose that a vector with a negative subscript denotes that vector with the indicated component omitted, and suppose that $\mathbf{t}$ corresponds to a multivariate saddlepoint $\hat{\boldsymbol{\tau}}$ with $\hat{\tau}_j < 0$. Then let $\mathbf{u} = (t_1, \dots, t_{j-1}, -t_j, t_{j+1}, \dots, t_d)$ and define $\mathbf{U}$ analogously. Then $\mathrm{P}[\mathbf{T} \geq \mathbf{t}] = \mathrm{P}[\mathbf{T}_{-j} \geq \mathbf{t}_{-j}] - \mathrm{P}[\mathbf{U} \geq \mathbf{u}]$, and the saddlepoints associated with $\mathbf{u}$ have one fewer negative entry than does $\hat{\boldsymbol{\tau}}$, and $Q(\mathbf{t})$ may be applied to $\mathbf{U}$ and $\mathbf{T}_{-j}$.

The vector $\hat{\boldsymbol{\tau}}$ might be interpreted as the maximum likelihood estimator for $\boldsymbol{\tau}$ when $\mathbf{t}$ is embedded in the exponential family with density $f_{\mathbf{T}}(\mathbf{t})\exp(\boldsymbol{\tau}^\top \mathbf{t} - \mathcal{K}(\boldsymbol{\tau}))$. Lugannani and Rice (1980), Skovgaard (1987) and Wang (1990) have developed approximations built around modifications of signed roots of likelihood ratio statistics when $d \leq 2$; the next section will review the two-dimensional version for comparison with $Q(\mathbf{t})$ and argue why this approach is infeasible for higher dimensions.

4. Uniformization. Lugannani and Rice (1980) present an approximation to $\mathrm{P}[\mathbf{T} \geq \mathbf{t}]$ in the case $d = 1$, by reparameterizing the integrand of (1) to make the argument to the exponent exactly quadratic. The resulting integral has a simple pole in its integrand, which is removed by a technique known in the applied mathematics literature on saddlepoint approximation as uniformization. This resulted in a particularly simple approximation. The approximation avoids the leading exponential factor in $Q(\mathbf{t})$ and hence is valid regardless of the signs of components of $\hat{\boldsymbol{\tau}}$. This section explores the possibility of extending this argument to higher dimensions, introduces the related expansion of Wang (1990), and explains why extensions to $d > 2$ will not be presented.

When $d > 1$, one might attempt to develop an expansion for $\mathrm{P}[\mathbf{T} \geq \mathbf{t}]$ by defining functions $\mathbf{w}$ such that

$$\tfrac{1}{2}(\mathbf{w} - \hat{\mathbf{w}})^\top (\mathbf{w} - \hat{\mathbf{w}}) = \mathcal{K}(\boldsymbol{\tau}) - \boldsymbol{\tau}^\top \mathbf{t} - \mathcal{K}(\hat{\boldsymbol{\tau}}) + \hat{\boldsymbol{\tau}}^\top \mathbf{t}, \tag{14}$$

and changing variables in (3) to find

$$\begin{aligned} \mathrm{P}[\mathbf{T} \geq \mathbf{t}] &= (2\pi i)^{-d} \int_{\hat{w}_1 - i\varepsilon}^{\hat{w}_1 + i\varepsilon} \cdots \int_{\hat{w}_d - i\varepsilon}^{\hat{w}_d + i\varepsilon} \exp\bigl(n(\mathbf{w}^\top \mathbf{w}/2 - \hat{\mathbf{w}}^\top \mathbf{w})\bigr) g(\mathbf{w})\, d\mathbf{w} \\ &\quad + O\bigl(\exp(-cn)\bigr) \end{aligned} \tag{15}$$

for $g(\mathbf{w}) = \det \frac{d\boldsymbol{\tau}}{d\mathbf{w}} / \prod_{j=1}^{d} \rho(\tau_j)$. The parameterization (14) is specified uniquely by requiring that w_j not depend on τ_k if $k > j$, and by requiring that w_j be an increasing function of τ_j. Kolassa (1997) proved that $\boldsymbol{\tau}(\mathbf{w})$ is analytic in $\mathbf{w}$ at $\hat{\mathbf{w}}$.

One might approximate (15) by expanding $g(\mathbf{w})$ as a power series, and applying (2) termwise to eliminate the effect of poles. Unfortunately, the resulting series has terms in components of $\mathbf{w}$ of unbounded negative order, making termwise inversion impossible. In the literature on multiple complex variables, $h(\mathbf{w})$ is in general not regular in $w_2, \dots, w_d$, because the value of w_j making $\tau_j = 0$ may depend on $w_1, \dots, w_{j-1}$; furthermore, this problem generally can not be repaired using a linear transformation of $\mathbf{w}$. If one defines $\tilde{w}_i(\mathbf{w})$ to satisfy

$$\tau_i\big(w_1, \dots, w_{i-1}, \tilde{w}_i(\mathbf{w})\big) = 0, \tag{16}$$

and approximates $\tilde{w}_j$ as approximately linear in $\mathbf{w}$ when $d \le 2$, one obtains the approximation

$$\begin{aligned} \mathrm{P}[\mathbf{T} \ge \mathbf{t}] = {} & \bar{\mathfrak{N}}(\sqrt{n}\mathbf{A}\hat{\mathbf{w}}, \mathbf{A}\mathbf{A}^\top) \\ & + \sum_{j=1}^{d} \bar{\mathfrak{N}}\big(\sqrt{n}[\hat{w}_{-j} - \tilde{w}_{-j}(\hat{\mathbf{w}})]\big)\mathfrak{n}(\sqrt{n}\hat{w}_j) \\ & \qquad \times \big\{1/(\hat{w}_j - \tilde{w}_j(\hat{\mathbf{w}})) - \sigma_j/\hat{\tau}_j\big\}/\sqrt{n} \\ & + O(1/n), \end{aligned} \tag{17}$$

where the generic element of matrix $\mathbf{A}$ is

$$A_{ij} = \begin{cases} (\hat{w}_i - \tilde{w}_i(\hat{\mathbf{w}}))/\hat{w}_j, & \text{if } i > j, \\ 1, & \text{if } i = j, \\ 0, & \text{if } i < j, \end{cases}$$

and σ_i are the diagonal elements of $\boldsymbol{\Upsilon}$, for $\boldsymbol{\Upsilon}$ the lower triangular matrix such that $\boldsymbol{\Upsilon}^\top\boldsymbol{\Upsilon} = \mathcal{K}''(\hat{\boldsymbol{\tau}})$. Lugannani and Rice (1980) derived this approximation when $d = 1$, and Wang (1990) derived this approximation for $d = 2$, using a different method of proof. He included an additional term of order $O(1/n)$, and demonstrated that the error term is of the same order. Call the resulting approximation $W(\mathbf{t})$. Tedious calculation shows that this approximation is valid to $O(1/n)$ for $d = 2$, but the method of proof fails for $d > 2$. Table 1 describes the association between notation used by Wang (1990) and the notation in this manuscript.

TABLE 1
Symbols used in Wang's paper

Notation of Wang (1990)	**Notation of the present manuscript**
v_0	$\hat{w}_1$
(u_0, t_0)	$\hat{\boldsymbol{\tau}}$
$w(-v_0) = w_u\|_{u=0}$	$\hat{w}_2$
$w(0) = w_{u_0}$	$\hat{w}_2 - \tilde{w}_2$
x_1	$\hat{w}_2$
y_1	$(\hat{w}_1 - b\hat{w}_2)/\sqrt{1+b^2}$

5. Examples. Wang (1990) considers a distribution of $\mathbf{T} = \sum_{j=1}^{n} \mathbf{Z}^\top \mathbf{X}^i$, for $\mathbf{X}^i$ independent vectors of independent unit exponentials and $\mathbf{Z}^\top = \binom{1\ 1\ 0}{0\ 1\ 1}$. In this case, $\mathcal{K}(\boldsymbol{\tau}) = -\sum_{j=1}^{3} \log(1 - \mathbf{z}_j \boldsymbol{\tau})$, for $\mathbf{z}_j$ row j of $\mathbf{Z}$. Figure 1 presents a comparison of the behavior of $W(\mathbf{t})$ and $Q(\mathbf{t})$. Both of these approximations require solution to the saddlepoint equations; these solutions are obtained using the Newton–Raphson method. Figure 1 presents $(|W(\mathbf{t}) - \mathrm{P}[\mathbf{T} \geq \mathbf{t}]|)/(|Q(\mathbf{t}) - \mathrm{P}[\mathbf{T} \geq \mathbf{t}]| + |W(\mathbf{t}) - \mathrm{P}[\mathbf{T} \geq \mathbf{t}]|)$, for the distribution of the sum of ten independent copies of $\mathbf{T}$. In both cases, $\mathrm{P}[\mathbf{T} \geq \mathbf{t}]$ is estimated using one million Monte Carlo samples. Calculations are performed in FORTRAN, using the IMSL subroutines for matrix manipulation and random number generation. I applied a mild loess smoother to the Monte Carlo approximation, and then plotted contours of the ratio of the absolute value of error in $Q(\mathbf{t})$ to the sum of the absolute errors of $W(\mathbf{t})$ and $Q(\mathbf{t})$. Neither $Q(\mathbf{t})$ nor $W(\mathbf{t})$ is clearly superior in this example.

Stokes, Davis and Koch (1995) present data on 63 case-control pairs of women with endometrial cancer. They seek to explain the occurrence of endometrial cancer on various explanatory variables, among them the presence of three risk factors, gall bladder disease, hypertension and nonestrogen drug use. They model

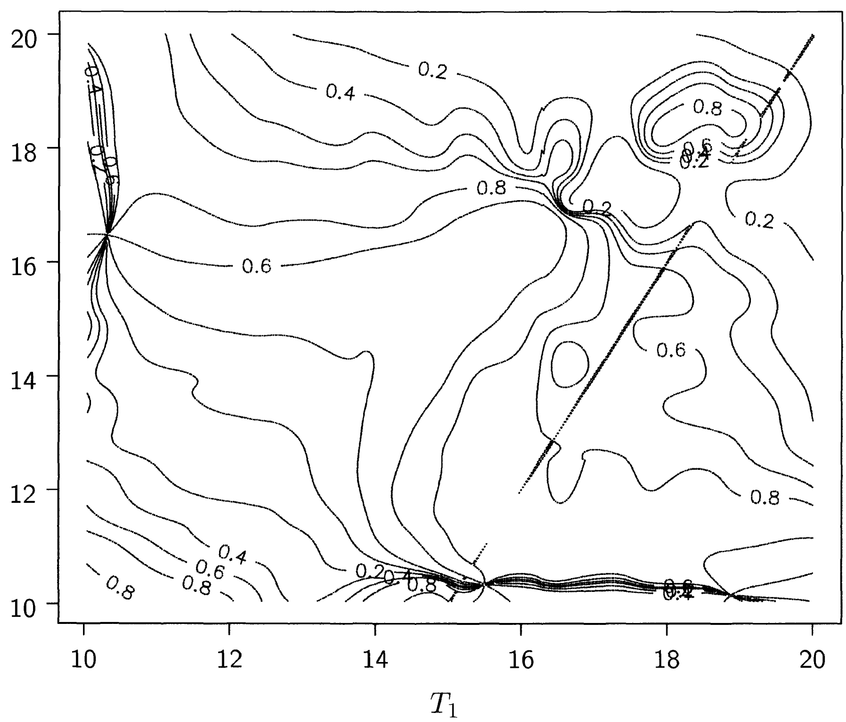

FIG. 1. *Contours of the ratio of the absolute value of error in $Q(\mathbf{t})$ to the sum of the absolute errors of $W(\mathbf{t})$ and $Q(\mathbf{t})$. Example of correlated Γ variables. Calculations are for lower quadrant probabilities, generated by applying $Q(\mathbf{t})$ to $-\mathbf{T}$.*

TABLE 2
Differences between cases and controls for endometrial cancer data

Gall bladder disease	−1	−1	−1	0	0	0	0	0	0	1	1	1	1	1	1	1
Hypertension	−1	0	1	−1	−1	0	0	1	1	−1	−1	0	0	0	1	1
Nonestrogen drug use	0	−1	0	−1	0	0	1	0	1	0	1	−1	0	1	0	1
Number of pairs	1	1	1	2	6	14	10	12	4	3	1	1	4	1	1	1

the probability of endometrial cancer π_j using logistic regression: $\pi_j = \exp(\theta_0 + \sum_{i=1}^3 \theta_i W_{ij})/(1+\exp(\theta_0+\sum_{i=1}^3 \theta_i W_{ij}))$, where W_{1j}, W_{2j} and W_{3j} are indicators for gall bladder disease, hypertension and nonestrogen drug use in individual j, respectively. Their data were obtained from a case-control study, and hence the independent variable is case-control status and the dependent variables are presence of the risk factors. Stokes, Davis and Koch (1995) note that the likelihood for these data is equivalent to that of a logistic regression arising from a prospective study, in which the units of observation are the matched pairs, the explanatory variables are those of the case member minus those of the control member, and the response variable may be taken to be unity. Table 2 contains the number of pairs with each configuration of differences of the three explanatory variables. Let $\mathbf{z}_j$ be the row vector formed by the top three entries in column j of Table 2, let m_j be the bottom entry in column j, and let $\mathbf{Z}$ be the matrix whose rows are $\mathbf{z}_j$. Let $\mathbf{T} = \mathbf{Z}^\top \mathbf{1}$, for $\mathbf{1}$ a column vector with as many entries as there are columns in Table 2, whose entries are all 1. Then $\mathcal{K}(\boldsymbol{\tau}) = \sum_j m_j[\log(1+\exp(\mathbf{z}_j\boldsymbol{\tau})) - \log(2)]$. Again, the saddlepoint $\hat{\boldsymbol{\tau}}$ is computed using the Newton–Raphson method.

None of these risk factors is likely to have a protective effect, and so the alternative hypothesis to $\boldsymbol{\theta} = \mathbf{0}$ is $\theta_j \geq 0\ \forall\, j$ and $\theta_j > 0$ for some j. This test will be performed by comparing the minimum p-value arising from the three univariate one-sided conditional tests to its null distribution. The three univariate one-sided p-values are 0.0175, 0.1885 and 0.0133. The observed critical region is $\{\mathbf{t} \mid \mathrm{P}[T_j \geq t_j] \leq 0.0133 \text{ for some } j\}$ and, applying (17) for $d = 1$, is approximated by $\mathcal{T} = \{\mathbf{t} \mid T_1 \geq 10 \text{ or } T_2 \geq 10 \text{ or } T_3 \geq 13\}$. Using Boole's law, $\mathrm{P}[\mathcal{T}] = 0.01709$. By comparison, 10,000,000 independent draws from the distribution of $\mathbf{T}$ yields the 95% confidence interval $\mathrm{P}[\mathcal{T}] \in (0.01702, 0.01718)$.

Acknowledgments. The author thanks Larry Shepp, Richard Gundy, Ovidiu Costin, Yodit Seifu, two anonymous referees, and an anonymous Associate Editor for helpful suggestions.

REFERENCES

BAK, J. and NEWMAN, D. J. (1982). *Complex Analysis*. Springer, New York.
DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25** 631–650.
DANIELS, H. E. (1987). Tail probability approximations. *Internat. Statist. Rev.* **55** 37–46.

KOLASSA, J. E. (1989). Topics in series approximations to distribution functions. Ph.D. dissertation, Univ. Chicago.

KOLASSA, J. E. (1997). Infinite parameter estimates in logistic regression, with application to approximate conditional inference. *Scand. J. Statist.* **24** 523–530.

KOLASSA, J. E. (1998). Uniformity of double saddlepoint conditional probability approximations. *J. Multivariate Anal.* **64** 66–85.

LUGANNANI, R. and RICE, S. (1980). Saddle point approximation for the distribution of the sum of independent random variables. *Adv. in Appl. Probab.* **12** 475–490.

PIERCE, D. A. and PETERS, D. (1999). Improving on exact tests by approximate conditioning. *Biometrika* **86** 265–277.

ROBINSON, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. *J. Roy. Statist. Soc. Ser. B* **44** 91–101.

ROUTLEDGE, R. and TSAO, M. (1995). Uniform validity of saddlepoint expansion on compact sets. *Canad. J. Statist.* **23** 425–431.

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

STOKES, M. E., DAVIS, C. S. and KOCH, G. G. (1995). *Categorical Data Analysis Using the SAS System.* SAS Institute, Cary, NC.

WANG, S. (1990). Saddlepoint approximations for bivariate distributions. *J. Appl. Probab.* **27** 586–597.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
HILL CENTER, BUSCH CAMPUS
110 FRELINGHUYSEN ROAD
PISCATAWAY, NEW JERSEY 08854-8019
E-MAIL: kolassa@stat.rutgers.edu