Bernoulli **16**(4), 2010, 1191–1207
DOI: 10.3150/09-BEJ237

Multivariate saddlepoint approximations in tail probability and conditional inference

JOHN KOLASSA* and JIXIN LI**

Department of Statistics, Rutgers University, Hill Center, Busch Campus, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA. E-mail: *kolassa@stat.rutgers.edu; **jixli@stat.rutgers.edu

We extend known saddlepoint tail probability approximations to multivariate cases, including multivariate conditional cases. Our approximation applies to both continuous and lattice variables, and requires the existence of a cumulant generating function. The method is applied to some examples, including a real data set from a case-control study of endometrial cancer. The method contains less terms and is easier to implement than existing methods, while showing an accuracy comparable to those methods.

Keywords: conditional probability; saddlepoint approximation; tail probability; Watson's lemma

1. Introduction

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent and identically distributed random vectors from a density $f_{\mathbf{X}}(\cdot)$ on $\mathbf{R}^d$. We construct an accurate multivariate saddlepoint approximation of the tail probability of the mean random vector $\bar{\mathbf{X}} = (\mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n)/n$. We also develop a similar approximation for conditional tail probabilities. The approximation has a relative error of $\mathrm{O}(n^{-1})$, uniformly over a compact set of $\bar{\mathbf{x}}$, a realization of $\bar{\mathbf{X}}$, under some general conditions. Our method utilizes the likelihood ratio statistic, routinely calculated by standard software, which makes the approximation easy to implement.

The Edgeworth expansion is a natural competitor to the saddlepoint approximation. This expansion has a uniformly bounded absolute error and works well in the center of the distribution being approximated. However, the approximation deteriorates at the far tail of the distribution, where it can sometimes even attain negative values. [1] first applied saddlepoint techniques to the approximation of a probability density function. Saddlepoint approximation addresses the problem of degradation outside a region of radius $\mathrm{O}(n^{-1/2})$ about $E(\mathbf{X}_i)$ by bounding the relative error, rather than the absolute error, of the approximation over the admissible range.

[1] discussed approximating the density of $\bar{X}$ when the dimension $d = 1$, that is, the univariate case. The approximation achieved a relative error of $\mathrm{O}(n^{-1})$ uniformly over the whole admissible range of the variable, under some conditions. The method uses the Fourier inversion formula, which involves moment generating, or characteristic, functions and complex integration. In this approach, the path of integration is shifted so that it passes through the saddlepoint of the integrand and follows the steepest descent curve at the neighborhood of the saddlepoint. The asymptotic property follows from a lemma due to [14].

Extensions of univariate saddlepoint approximation of tail probabilities $P(\bar{X} > \bar{x})$ for the means of independent random variables have also been studied. This calculation is more dif-

1350-7265

ficult, in that, unlike the density function case, the integrand of the Fourier inversion integral for tail probabilities has a pole at zero.

[10] presented a general saddlepoint approximation technique that can be applied to tail probability approximation, based on Laplace approximation of the integrated saddlepoint density, with an error of $\mathrm{O}(n^{-1})$. Robinson used an argument involving a conjugate exponentially shifted distribution family and the Edgeworth expansion. The terms of the expansion can then be integrated termwise. There is no direct explicit formula for the integration of each term, but the terms may be computed recursively. This method applies when $\bar{x} \geq E(X)$. When $\bar{x} < E(X)$, Boole's law and reflection of the distribution must be used.

[8] provided an alternative approximation. [2] derived this technique, using a transformation of variables to directly address the local quadratic behavior of the numerator exponent. The integral is then split into two parts, one which contains a pole, but can be integrated exactly and explicitly, and the other which only has removable singularities and can be expanded and approximated accurately. The virtue of this method is that the approximation is compact and can be computed without recursion, and the formula is valid over the whole range of admissible $\bar{x}$.

[9] thoroughly discussed the usefulness of the saddlepoint method in a review of the method focusing on a variety of applications to statistical inference.

[5] generalized the univariate Robinson approach under the Daniels framework and achieved an error of size $\mathrm{O}(n^{-1})$. The method uses integral expressions for the tail probability in the multivariate case and presents a multivariate expansion of the numerator of the integrand and a termwise multivariate integration using recursion. This approach shares the drawback of Robinson's approach in that it requires a positivity constraint on the ordinate.

[13] generalized Lugannani and Rice's method to the case of a bivariate probability distribution function using variable transformations. [5] used a different method of proof and showed that the error term is of order $\mathrm{O}(n^{-1})$; his method is limited to $d = 2$. Furthermore, Wang's development involves an inversion integral in which the pole of one variable depends on the values of other variables in a fundamentally nonlinear way.

Wang's proof of the error rate in the neighborhood of the pole is incomplete. In this paper, a way of effectively extending Lugannani and Rice's method to the multivariate case, which uses a different transformation formula from Wang's and can be used in the case $d > 2$, is proposed. The method uses fewer terms and can be extended to multivariate conditional cases.

Our proposed saddlepoint approximation may be used to test null and alternative hypotheses concerning a multivariate parameter when the hypotheses are specified by systems of linear inequalities. [6] applied the method of [5], in conjunction with the adjusted profile likelihood, in such a case. For instance, [6] refers to data presented by [12] on 63 case-control pairs of women with endometrial cancer. The occurrence of endometrial cancer is influenced by explanatory variables including gall bladder disease, hypertension and non-estrogen drug use. The test of whether hypertension or non-estrogen drug use is associated with an increase in endometrial cancer will be performed, conditional on the sufficient statistic value associated with gall bladder disease.

The remainder of the paper is organized as follows. Section 2 provides the unified framework under which both unconditional and conditional tail probability approximations are considered. Section 3 derives formulas for multivariate unconditional distributions. Section 4 focuses on conditional distributions. Section 5 presents five examples and shows the approximation results.

2. Multivariate extension

The unconditional and conditional tail probability approximation share some common characteristics. We derive them in a unified way. Applying the Fourier inversion theorem and Fubini's theorem, as in [5], we find that both the unconditional and conditional tail probability approximations require evaluation of an integral of the form

$$\frac{n^{d-d_0}}{(2\pi \mathrm{i})^d}\int_{\mathbf{c}-\mathrm{i}\mathbf{K}}^{\mathbf{c}+\mathrm{i}\mathbf{K}} \frac{\exp(n[K(\boldsymbol{\tau})-\boldsymbol{\tau}^{\mathrm{T}}\mathbf{t}^*])}{\prod_{j=1}^{d_0}\rho(\tau_j)}\,\mathrm{d}\boldsymbol{\tau}, \tag{2.1}$$

where K is the cumulant generating function, which is the natural logarithm of the moment generating function, and $\mathbf{c}$ is any positive d-dimensional vector. This will be discussed in Section 4. In the unconditional case, for continuous variables, $\mathbf{K}$ is a vector of length d, with every entry infinity, $\mathbf{t}^* = \mathbf{t}$ and $\rho(\tau) = \tau$; for unit lattice, $\mathbf{K}$ is a vector of length d, with every entry π, $\mathbf{t}^*$ is $\mathbf{t}$ corrected for continuity, $\rho(\tau) = 2\sinh(\tau/2)$ and $d = d_0$. In the conditional case, the setting is the same, except that d_0 equals d minus the dimension of the conditioning variables.

[2] recast a great deal of the saddlepoint literature in terms of inversion integrals of the form (2.1), rescaled so that the exponent is exactly quadratic. This rescaling includes the multiplier for the linear term in the exponent; this linear term is the signed root of the likelihood ratio statistic. The idea of using the modified signed likelihood ratio statistic was proposed in [3]. [4] defines a multivariate version of this reparameterization and also defines the multiplier for the linear terms; again, these are signed roots of likelihood ratio statistics, but, this time, for a sequence of nested models:

$$-\frac{1}{2}\hat{\mathbf{w}}^{\mathrm{T}}\hat{\mathbf{w}} = \min_{\boldsymbol{\gamma}}\bigl(K(\boldsymbol{\gamma}) - \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{t}^*\bigr)$$

and

$$-\frac{1}{2}(\mathbf{w}-\hat{\mathbf{w}})^{\mathrm{T}}(\mathbf{w}-\hat{\mathbf{w}}) = K(\boldsymbol{\tau}) - \boldsymbol{\tau}^{\mathrm{T}}\mathbf{t}^* - \min_{\boldsymbol{\gamma}}\bigl(K(\boldsymbol{\gamma}) - \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{t}^*\bigr).$$

Further specification of $\hat{\mathbf{w}}$ and $\mathbf{w}$ is needed. For any vector $\mathbf{v}$ of length d, let $\mathbf{v}_j$ be the vector consisting of the first j elements, that is, $(v_1, v_2, \ldots, v_j)^{\mathrm{T}}$. For instance, $\boldsymbol{\gamma}_j = (\gamma_1, \gamma_2, \ldots, \gamma_j)^{\mathrm{T}}$, $\boldsymbol{\tau}_j = (\tau_1, \tau_2, \ldots, \tau_j)^{\mathrm{T}}$ and $\mathbf{0}_j$ is the zero vector $(0, 0, \ldots, 0)^{\mathrm{T}}$ with dimension j. Let $\mathbf{v}_{-j}$ be the vector consisting all but the first j elements of $\mathbf{v}$, that is, $(v_{j+1}, v_{j+2}, \ldots, v_d)^{\mathrm{T}}$. [4], Chapter 6 defines $\hat{\mathbf{w}}$ and $\mathbf{w}$ using

$$-\frac{1}{2}\hat{w}_j^2 = \min_{\boldsymbol{\gamma}, \boldsymbol{\gamma}_{j-1}=\mathbf{0}_{j-1}}\bigl(K(\boldsymbol{\gamma}) - \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{t}^*\bigr) - \min_{\boldsymbol{\gamma}, \boldsymbol{\gamma}_j=\mathbf{0}_j}\bigl(K(\boldsymbol{\gamma}) - \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{t}^*\bigr), \tag{2.2a}$$

$$-\frac{1}{2}(w_j - \hat{w}_j)^2 = \min_{\boldsymbol{\gamma}, \boldsymbol{\gamma}_{j-1}=\boldsymbol{\tau}_{j-1}}\bigl(K(\boldsymbol{\gamma}) - \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{t}^*\bigr) - \min_{\boldsymbol{\gamma}, \boldsymbol{\gamma}_j=\boldsymbol{\tau}_j}\bigl(K(\boldsymbol{\gamma}) - \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{t}^*\bigr). \tag{2.2b}$$

This definition is not invariant with regard to the order of the coordinates. Also, note that w_j is a function of only $\boldsymbol{\tau}_j$, but not of any element of $\boldsymbol{\tau}_{-j}$ $\forall j$. The same holds true for τ_j as a function of $\mathbf{w}$.

We now construct more explicit formulas for $\hat{\mathbf{w}}$ and $\mathbf{w}$. Let

$$\tilde{\boldsymbol{\tau}}_j(\boldsymbol{\gamma}_j) = (\gamma_1, \gamma_2, \ldots, \gamma_j, \tilde{\tau}_{j+1}(\boldsymbol{\gamma}_j), \tilde{\tau}_{j+2}(\boldsymbol{\gamma}_j), \ldots, \tilde{\tau}_d(\boldsymbol{\gamma}_j))$$

be the minimizer of $(K(\boldsymbol{\gamma}) - \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{t}^*)$ when the first j variables are fixed. The function $\tilde{\tau}_k(\boldsymbol{\gamma}_j)$ above is the minimizer for variable k when the first j variables are fixed, for $k > j$.

Using the above notation, the definitions of $\hat{\mathbf{w}}$ and $\mathbf{w}$ can be rewritten as

$$-\frac{1}{2}\hat{w}_j^2 = K(\tilde{\boldsymbol{\tau}}_{j-1}(\mathbf{0}_{j-1})) - \tilde{\boldsymbol{\tau}}_{j-1}(\mathbf{0}_{j-1})^{\mathrm{T}}\mathbf{t}^* - \big(K(\tilde{\boldsymbol{\tau}}_j(\mathbf{0}_j)) - \tilde{\boldsymbol{\tau}}_j(\mathbf{0}_j)^{\mathrm{T}}\mathbf{t}^*\big), \tag{2.3a}$$

$$-\frac{1}{2}(w_j - \hat{w}_j)^2 = K(\tilde{\boldsymbol{\tau}}_{j-1}(\boldsymbol{\tau}_{j-1})) - \tilde{\boldsymbol{\tau}}_{j-1}(\boldsymbol{\tau}_{j-1})^{\mathrm{T}}\mathbf{t}^* - \big(K(\tilde{\boldsymbol{\tau}}_j(\boldsymbol{\tau}_j)) - \tilde{\boldsymbol{\tau}}_j(\boldsymbol{\tau}_j)^{\mathrm{T}}\mathbf{t}^*\big), \tag{2.3b}$$

where $\tilde{\boldsymbol{\tau}}_{j-1}(\cdot)$ is set to $\hat{\boldsymbol{\tau}}$ when $j = 1$ for succinctness of expression.

By choosing a sign to make $\hat{\mathbf{w}}$ and $\mathbf{w}$ increasing functions of $\hat{\boldsymbol{\tau}}$ and $\boldsymbol{\tau}$, we can further specify them as follows:

$$\begin{aligned}\hat{w}_j &= \mathrm{sign}(\tilde{\tau}_j(\mathbf{0}_{j-1}))\\ &\quad\times\sqrt{-2\big[K(\tilde{\boldsymbol{\tau}}_{j-1}(\mathbf{0}_{j-1})) - \tilde{\boldsymbol{\tau}}_{j-1}(\mathbf{0}_{j-1})^{\mathrm{T}}\mathbf{t}^* - \big(K(\tilde{\boldsymbol{\tau}}_j(\mathbf{0}_j)) - \tilde{\boldsymbol{\tau}}_j(\mathbf{0}_j)^{\mathrm{T}}\mathbf{t}^*\big)\big]},\end{aligned} \tag{2.4a}$$

$$\begin{aligned}w_j &= \hat{w}_j + \mathrm{sign}\big(\tau_j - \tilde{\tau}_j(\boldsymbol{\tau}_{j-1})\big)\\ &\quad\times\sqrt{-2\big[K(\tilde{\boldsymbol{\tau}}_{j-1}(\boldsymbol{\tau}_{j-1})) - \tilde{\boldsymbol{\tau}}_{j-1}(\boldsymbol{\tau}_{j-1})^{\mathrm{T}}\mathbf{t}^* - \big(K(\tilde{\boldsymbol{\tau}}_j(\boldsymbol{\tau}_j)) - \tilde{\boldsymbol{\tau}}_j(\boldsymbol{\tau}_j)^{\mathrm{T}}\mathbf{t}^*\big)\big]}.\end{aligned} \tag{2.4b}$$

The derivation of the [8] approximation provided by [2] requires identification of the simple pole in the inversion integrand. We need to match zeros in the denominator of the multivariate integrand with functions of the variables in the new parameterization; the points at which this matching occurs will be denoted by a tilde. The quantities above, such as $\hat{\boldsymbol{\tau}}$, $\hat{\mathbf{w}}$, $\tilde{\tau}_j(\boldsymbol{\tau}_{j-1})$ and functional relationships between $\boldsymbol{\tau}$ and $\mathbf{w}$, etcetera, can be solved numerically by Newton–Raphson methods, or even analytically in some cases. Finally, we define a function $\tilde{w}_j(\mathbf{w}_{j-1})$ such that $\tau_j(w_1, w_2, \ldots, \tilde{w}_j(\mathbf{w}_{j-1})) = 0$ for $j > 1$.

It can be verified that the following properties hold:

$$\boldsymbol{\tau}_j = 0 \quad \text{if and only if} \quad \mathbf{w}_j = 0; \tag{2.5a}$$

$$\tilde{w}_j(\mathbf{0}_{j-1}) = 0 \qquad \text{for } j > 1; \tag{2.5b}$$

$$\tau_j = \tilde{\tau}_j(\boldsymbol{\tau}_{j-1}) \quad \text{if and only if} \quad w_j = \hat{w}_j \qquad \text{for } j > 1; \tag{2.5c}$$

$$\boldsymbol{\tau}_j = \hat{\boldsymbol{\tau}}_j \quad \text{if and only if} \quad \mathbf{w}_j = \hat{\mathbf{w}}_j. \tag{2.5d}$$

Below, the superscript of a function denotes differentiation with respect to the corresponding argument of the function. We will employ the same use of superscripts in the subsequent text of the paper, except that when the superscript is a set, it denotes difference, as defined at the end of this section. Also, a superscripted "T" denotes the transpose of matrix. We can obtain

$\breve{w}_j = \tilde{w}_j(\hat{\mathbf{w}}_{j-1})$ and $\breve{w}_j^k = \tilde{w}_j^k(\hat{\mathbf{w}}_{j-1})$, which will be used in later sections. Substituting $w_j = \breve{w}_j$, $\tau_j = 0$, $\boldsymbol{\tau}_{j-1} = \hat{\boldsymbol{\tau}}_{j-1}$ and $\boldsymbol{\tau}_j = (\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_{j-1}, 0)^{\mathrm{T}} = (\hat{\boldsymbol{\tau}}_{j-1}, 0)^{\mathrm{T}}$ into (2.4b), we obtain

$$\breve{w}_j = \hat{w}_j + \operatorname{sign}(0 - \hat{\tau}_j)\sqrt{-2\big[K(\hat{\boldsymbol{\tau}}) - \hat{\boldsymbol{\tau}}^{\mathrm{T}}\mathbf{t}^* - \big(K(\tilde{\boldsymbol{\tau}}_j(\hat{\boldsymbol{\tau}}_{j-1}, 0)) - (\hat{\boldsymbol{\tau}}_{j-1}, 0)^{\mathrm{T}}\mathbf{t}^*\big)\big]}. \tag{2.6}$$

Differentiating (2.3b) with respect to w_k and rearranging terms, we obtain

$$\breve{w}_j^k = \sum_{l=k}^{j-1}\left(K^l(\tilde{\boldsymbol{\tau}}_j(\hat{\boldsymbol{\tau}}_{j-1}, 0)) \cdot \frac{\mathrm{d}\tau_l}{\mathrm{d}w_k}\bigg|_{\hat{\mathbf{w}}_l} - t_l^*\right) \Big/ (\breve{w}_j - \hat{w}_j) \tag{2.7}$$

for $k < j$. The derivatives $\frac{\mathrm{d}\tau_l}{\mathrm{d}w_k}$ evaluated at the point $\mathbf{w}_l$ can be obtained by differentiating (2.3b) with respect to w_k once or twice, depending on whether or not $w_j = \hat{w}_j$, and solving the resulting system of equations. In particular, we are interested in

$$\frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\bigg|_{\mathbf{w}_j} = \begin{cases} \sqrt{\dfrac{1}{\sum_{l=j}^d K^{jl}(\tilde{\boldsymbol{\tau}}_{j-1}(\boldsymbol{\tau}_{j-1}))\tau_l^j([\tilde{\boldsymbol{\tau}}_{j-1}(\boldsymbol{\tau}_{j-1})]_j)}}, & \text{if } w_j = \hat{w}_j, \\ \dfrac{w_j - \hat{w}_j}{K^j(\tilde{\boldsymbol{\tau}}_j(\tau_j)) - t_j^*}, & \text{if } w_j \neq \hat{w}_j, \end{cases} \tag{2.8}$$

for $j \le d_0$, where $[\cdot]_j$ denotes the first j elements, and

$$\prod_{j=d_0+1}^{d} \frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\bigg|_{(\mathbf{w}_{d_0}, \hat{\mathbf{w}}_{-d_0})} = \prod_{j=d_0+1}^{d} \sqrt{\frac{1}{\sum_{l=j}^d K^{jl}(\tilde{\boldsymbol{\tau}}_{d_0}(\boldsymbol{\tau}_{d_0}))\tau_l^j([\tilde{\boldsymbol{\tau}}_{d_0}(\boldsymbol{\tau}_{d_0})]_j)}}, \tag{2.9}$$

where, for succinctness of expression, we define $\tau_l^j(\cdot)$ to be 1 when $l = j$. For $l > j$, we obtain $\tau_l^j(\cdot)$ by differentiating both sides of the definition of $\tau_l^j(\cdot)$, that is, $K^l(\cdot) = t_l^*$ with respect to τ_j $\forall l > j$, and solving the system of equations.

Under this transformation of variables from $\boldsymbol{\tau}$ to $\mathbf{w}$, the Jacobian is just the product of the diagonal terms of the Jacobian matrix and (2.1) can be expressed as

$$\begin{aligned}
&\frac{n^{d-d_0}}{(2\pi \mathrm{i})^d}\int_{\hat{\mathbf{w}}-\mathrm{i}\mathbf{K}}^{\hat{\mathbf{w}}+\mathrm{i}\mathbf{K}} \frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w} - \hat{\mathbf{w}}^{\mathrm{T}}\mathbf{w}])}{\prod_{j=1}^{d_0}\rho(\tau_j(\mathbf{w}_j))} \prod_{j=1}^{d}\frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\,\mathrm{d}\mathbf{w} \\
&\quad = \frac{n^{d-d_0}}{(2\pi \mathrm{i})^d}\int_{\hat{\mathbf{w}}-\mathrm{i}\mathbf{K}}^{\hat{\mathbf{w}}+\mathrm{i}\mathbf{K}} \frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w} - \hat{\mathbf{w}}^{\mathrm{T}}\mathbf{w}])}{\prod_{j=1}^{d_0}(w_j - \tilde{w}_j(\mathbf{w}_{j-1}))} \cdot \prod_{j=1}^{d}\frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\frac{\prod_{j=1}^{d_0}(w_j - \tilde{w}_j(\mathbf{w}_{j-1}))}{\prod_{j=1}^{d_0}\rho(\tau_j(\mathbf{w}_j))}\,\mathrm{d}\mathbf{w} \qquad (2.10)\\
&\quad \sim \frac{n^{d-d_0}}{(2\pi \mathrm{i})^d}\int_{\hat{\mathbf{w}}-\mathrm{i}\mathbf{K}}^{\hat{\mathbf{w}}+\mathrm{i}\mathbf{K}} \frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w} - \hat{\mathbf{w}}^{\mathrm{T}}\mathbf{w}])}{\prod_{j=1}^{d_0}(w_j - \tilde{w}_j(\mathbf{w}_{j-1}))} G(\boldsymbol{\tau})\,\mathrm{d}\mathbf{w},
\end{aligned}$$

where

$$G(\boldsymbol{\tau}) = \frac{\prod_{j=1}^{d_0}(w_j - \tilde{w}_j(\mathbf{w}_{j-1}))}{\prod_{j=1}^{d_0}\rho(\tau_j(\mathbf{w}_j))}\prod_{j=1}^{d_0}\frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\cdot\prod_{j=d_0+1}^{d}\frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\bigg|_{(\mathbf{w}_{d_0},\hat{\mathbf{w}}_{-d_0})}$$

and, to simplify notation, we set $\tilde{w}_j(\mathbf{w}_{j-1})$ to zero for $j = 1$. For later convenience, we write $G(\boldsymbol{\tau})$ as a function $\boldsymbol{\tau}$ instead of $\mathbf{w}$. The relation $\sim$ in the last step indicates exact equality in the unconditional case, where $d = d_0$, but holds with a relative error of $\mathrm{O}(n^{-1})$ in the conditional case, which we will discuss in Section 4. Hereafter, we use $\sim$ to denote approximation with a relative error of $\mathrm{O}(n^{-1})$ of both the left-hand side and the tail probability, and we use $\dot{\sim}$ ($\sim$ with a dot above it) in the case where the right-hand side is an approximation with a relative error of $\mathrm{O}(n^{-1/2})$ of the left-hand side.

The last integral in (2.10) will be evaluated by splitting it into rather simple terms involving poles and more complicated terms involving analytic functions. We can decompose (2.10) into 2^{d_0} terms. Let $U = \{1, 2, \ldots, d_0\}$ be the index set of integers from 1 to d_0. For set $s \subseteq U$, define $G^s(\boldsymbol{\tau}) = G(\boldsymbol{\tau}^s)$, where the vector $\boldsymbol{\tau}^s$ is defined by

$$\tau_j^s = \begin{cases} \tau_j, & \text{if } j \in s, \\ 0, & \text{if } j \notin s. \end{cases}$$

For example, if $d_0 = 3$, then $G^{\{1,2\}}(\boldsymbol{\tau}) = G(\tau_1, \tau_2, 0)$. Now, for $t \subseteq U$, define $H^t = \sum_{s\subseteq t}(-1)^{|t-s|}G^s(\boldsymbol{\tau})$, where $|\cdot|$ denotes the cardinality, that is, the number of elements of a set. For example, $H^{\{1,2\}} = G^{\{1,2\}}(\boldsymbol{\tau}) - G^{\{1\}}(\boldsymbol{\tau}) - G^{\{2\}}(\boldsymbol{\tau}) + G^{\varnothing}(\boldsymbol{\tau}) = G(\tau_1, \tau_2, 0) - G(\tau_1, 0, 0) - G(0, \tau_2, 0) + G(0, 0, 0)$, where $\varnothing$ denotes the empty set. We conclude that $G(\boldsymbol{\tau}) = \sum_{t\subseteq U} H^t$. This decomposition holds by induction on d_0. Noting that $\forall s \subseteq U$ and $a \in s$, $H^s(\boldsymbol{\tau}^{\{a\}}) = 0$, we see that

$$\frac{H^t(\boldsymbol{\tau})}{\prod_{j\in t}(w_j - \tilde{w}_j(\mathbf{w}_{j-1}))}$$

is analytic. In other words, $|t|$ product terms in the denominator of the integrand in (2.10) are 'absorbed' by $H^t(\boldsymbol{\tau})$, leaving the remaining $(d_0 - |t|)$ product terms unabsorbed. As explained in [5], each term that is absorbed contributes a relative error of $\mathrm{O}(n^{-1/2})$. Therefore, if we let I^t be the integral corresponding to H^t, then we obtain

$$\frac{n^{d-d_0}}{(2\pi\mathrm{i})^d}\int_{\hat{\mathbf{w}}-\mathrm{i}\mathbf{K}}^{\hat{\mathbf{w}}+\mathrm{i}\mathbf{K}}\frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w} - \hat{\mathbf{w}}^{\mathrm{T}}\mathbf{w}])}{\prod_{j=1}^{d_0}(w_j - \tilde{w}_j(\mathbf{w}_{j-1}))}G(\boldsymbol{\tau})\,\mathrm{d}\mathbf{w} \sim \sum_{|t|\le 1, t\subseteq U} I^t. \tag{2.11}$$

In the next two sections, we compute the I^t, $|t| \le 1$, $t \subseteq U$, for distribution and conditional distribution, respectively.[1]

[1]More detailed derivations and formulae for bivariate distributions can be found at http://stat.rutgers.edu/resources/technical_reports10.html.

3. Multivariate distribution approximation

In the unconditional continuous case, we have $d = d_0$ and

$$G(\boldsymbol{\tau}) = \frac{\prod_{j=1}^{d_0}(w_j - \tilde{w}_j(\mathbf{w}_{j-1}))}{\prod_{j=1}^{d_0}\tau_j(\mathbf{w}_j)} \prod_{j=1}^{d_0} \frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}.$$

Therefore,

$$\begin{aligned} I^{\varnothing} &= \frac{1}{(2\pi \mathrm{i})^{d_0}} \int_{\hat{\mathbf{w}}-\mathrm{i}\infty}^{\hat{\mathbf{w}}+\mathrm{i}\infty} \frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w} - \hat{\mathbf{w}}^{\mathrm{T}}\mathbf{w}])}{\prod_{j=1}^{d_0}(w_j - \tilde{w}_j(\mathbf{w}_{j-1}))} G(\mathbf{0})\,\mathrm{d}\mathbf{w} \\ &= \frac{1}{(2\pi \mathrm{i})^{d_0}} \int_{\hat{\mathbf{w}}-\mathrm{i}\infty}^{\hat{\mathbf{w}}+\mathrm{i}\infty} \frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w} - \hat{\mathbf{w}}^{\mathrm{T}}\mathbf{w}])}{\prod_{j=1}^{d_0}(w_j - \tilde{w}_j(\mathbf{w}_{j-1}))}\,\mathrm{d}\mathbf{w} \end{aligned} \tag{3.1}$$

since $G(\mathbf{0}) = 0$ by properties (2.5a) and (2.5b).

Let $u_j = w_j - \tilde{w}_j(\mathbf{w}_{j-1})$, $\hat{\mathbf{u}}$ be such that $\mathbf{w}(\hat{\mathbf{u}}) = \hat{\mathbf{u}}$ and $g(\mathbf{u}) = \frac{1}{2}\mathbf{w}(\mathbf{u})^{\mathrm{T}}\mathbf{w}(\mathbf{u}) - \hat{\mathbf{w}}^{\mathrm{T}}\mathbf{w}(\mathbf{u})$. By changing variables, with Jacobian equal to 1, we have

$$I^{\varnothing} = \frac{1}{(2\pi \mathrm{i})^{d_0}} \int_{\hat{\mathbf{u}}-\mathrm{i}\infty}^{\hat{\mathbf{u}}+\mathrm{i}\infty} \frac{\exp(n[g(\mathbf{u})])}{\prod_{j=1}^{d_0} u_j}\,\mathrm{d}\mathbf{u}. \tag{3.2}$$

The integration in (3.2) cannot be integrated out exactly in general. However, using the same argument as in [5], we approximate it by expanding $g(\mathbf{u})$ about $\hat{\mathbf{u}}$ up to the third degree; after termwise integration, the resulting approximation to $I^{\varnothing}$ has relative error $\mathrm{O}(n^{-1})$. So, $I^{\varnothing}$ can be approximated by

$$\begin{aligned} I^{\varnothing} &= \frac{1}{(2\pi \mathrm{i})^{d_0}} \int_{\hat{\mathbf{u}}-\mathrm{i}\infty}^{\hat{\mathbf{u}}+\mathrm{i}\infty} \frac{\exp(n[\hat{g} + (1/2)\hat{g}^{jk}(u_j - \hat{u}_j)(u_k - \hat{u}_k)])}{\prod_{j=1}^{d_0} u_j} \\ &\quad \times \left(1 + \frac{n}{6}\hat{g}^{jkl}(u_j - \hat{u}_j)(u_k - \hat{u}_k)(u_l - \hat{u}_l)\right)\mathrm{d}\mathbf{u} \\ &= \frac{1}{(2\pi \mathrm{i})^{d_0}} \int_{\hat{\mathbf{u}}-\mathrm{i}\infty}^{\hat{\mathbf{u}}+\mathrm{i}\infty} \frac{\exp(n[\hat{g} + (1/2)\hat{g}^{jk}(u_j - \hat{u}_j)(u_k - \hat{u}_k)])}{\prod_{j=1}^{d_0} u_j}\,\mathrm{d}\mathbf{u} \\ &\quad + \frac{1}{(2\pi \mathrm{i})^{d_0}} \int_{\hat{\mathbf{u}}-\mathrm{i}\infty}^{\hat{\mathbf{u}}+\mathrm{i}\infty} \frac{\exp(n[\hat{g} + (1/2)\hat{g}^{jk}(u_j - \hat{u}_j)(u_k - \hat{u}_k)])}{\prod_{j=1}^{d_0} u_j} \\ &\quad \times \frac{n}{6}\hat{g}^{jkl}(u_j - \hat{u}_j)(u_k - \hat{u}_k)(u_l - \hat{u}_l)\,\mathrm{d}\mathbf{u}, \end{aligned} \tag{3.3}$$

where, for brevity, we write $\hat{g}^r$ for $g^r(\hat{\mathbf{u}})$. All derivatives of g evaluated at $\hat{\mathbf{u}}$ can be computed and, in particular, $\hat{g}^j = 0$. Here, we use tensor notation, that is, the use of superscripts and subscripts to denote summation over all possible combinations, by which we are able to omit the summation symbol. The computation of the second integral is addressed in [5]. The details involve partial derivatives of some functions up to the second or third degree; these are algebraically complicated and therefore omitted here. For the first integral, rearrange the terms in the numerator in the order of the degree of $\mathbf{u}$. The first integral is quadratic and can be computed as

$$\begin{aligned}
&\frac{1}{(2\pi \mathrm{i})^{d_0}} \int_{\hat{\mathbf{u}}-\mathrm{i}\infty}^{\hat{\mathbf{u}}+\mathrm{i}\infty} \frac{\exp(n[\hat{g} + (1/2)\hat{g}^{jk}(u_j - \hat{u}_j)(u_k - \hat{u}_k)])}{\prod_{j=1}^{d_0} u_j} \,\mathrm{d}\mathbf{u} \\
&= \frac{1}{(2\pi \mathrm{i})^{d_0}} \int_{\hat{\mathbf{u}}-\mathrm{i}\infty}^{\hat{\mathbf{u}}+\mathrm{i}\infty} \frac{\exp(n[(\hat{g} + (1/2)\hat{g}^{jk}\hat{u}_j\hat{u}_k) - \hat{g}^{jk}\hat{u}_k u_j + (1/2)\hat{g}^{jk}u_j u_k])}{\prod_{j=1}^{d_0} u_j} \,\mathrm{d}\mathbf{u} \qquad (3.4) \\
&= C^{\varnothing}\bar{\Phi}(\bar{\mathbf{y}}^{\varnothing}, \boldsymbol{\Sigma}^{\varnothing}),
\end{aligned}$$

where $C^{\varnothing} = \exp(n[\hat{g} + \frac{1}{2}\hat{g}^{jk}\hat{u}_j\hat{u}_k)])$, $\bar{\mathbf{y}}^{\varnothing}$ is a vector whose jth element is $\sqrt{n}\hat{g}^{jk}\hat{u}_k/\sqrt{\hat{g}^{jj}}$ and $\bar{\Phi}$ is the tail probability of a standard multivariate normal distribution with mean 0 and covariance matrix $\boldsymbol{\Sigma}^{\varnothing}$ with elements $\hat{g}^{jk}/\sqrt{\hat{g}^{jj}\hat{g}^{kk}}$. The last of the above equations can be obtained by changing variables to $\mathbf{v}$, where $v_j = u_j/\sqrt{\hat{g}^{jj}}$.

For I^t, $t = r$, we have

$$I^{\{r\}} = \frac{1}{(2\pi \mathrm{i})^{d_0}} \int_{\hat{\mathbf{w}}-\mathrm{i}\infty}^{\hat{\mathbf{w}}+\mathrm{i}\infty} \frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w} - \hat{\mathbf{w}}^{\mathrm{T}}\mathbf{w}])}{\prod_{j\neq r}(w_j - \tilde{w}(\mathbf{w}_{j-1}))} \cdot \frac{G^{\{r\}}(\boldsymbol{\tau}) - G(\mathbf{0})}{w_r - \tilde{w}(\mathbf{w}_{r-1})} \,\mathrm{d}\mathbf{w}. \qquad (3.5)$$

We perform a similar change of variable from $\mathbf{w}$ to $\mathbf{u}$ as in computing $I^{\varnothing}$, except that $u_r = w_r$. We then have

$$\begin{aligned}
I^{\{r\}} &= \frac{1}{(2\pi \mathrm{i})^{d_0}} \int_{\hat{\mathbf{u}}-\mathrm{i}\infty}^{\hat{\mathbf{u}}+\mathrm{i}\infty} \frac{\exp(n[g^{\{r\}}(\mathbf{u})])}{\prod_{j\neq r} u_j} h^{\{r\}}(\mathbf{u}) \,\mathrm{d}\mathbf{w} \\
&\mathrel{\dot\sim} \frac{1}{(2\pi \mathrm{i})^{d_0}} \int_{\hat{\mathbf{u}}-\mathrm{i}\infty}^{\hat{\mathbf{u}}+\mathrm{i}\infty} \frac{\exp(n[c_{00}^{\{r\}} + (1/2)(\mathbf{u}-\hat{\mathbf{u}})^{\mathrm{T}}\mathbf{cc}^{\{r\}}(\mathbf{u}-\hat{\mathbf{u}}) - \mathbf{u}^{\{r\}}(\mathbf{u}-\hat{\mathbf{u}}))])}{\prod_{j\neq r} u_j} h_{\mathbf{u}}^{\{r\}}(\mathbf{u}) \,\mathrm{d}\mathbf{w},
\end{aligned} \qquad (3.6)$$

where $g^{\{r\}}(\mathbf{u})$ is the exponent as a function of $\mathbf{u}$ after the change of variable, $h_{\mathbf{u}}^{\{r\}}(\mathbf{u}) = \frac{G^{\{r\}}(\boldsymbol{\tau})-G(\mathbf{0})}{w_r-\tilde{w}(\mathbf{w}_{r-1})}$. $c_{00}^{\{r\}} = g^{\{r\}}(\hat{\mathbf{u}})$, $\mathbf{cc}^{\{r\}}$ is the matrix with elements $cc_{ij}^{\{r\}} = [g^{\{r\}}]^{ij}(\hat{\mathbf{u}})$ and $\mathbf{c}^{\{r\}}$ is the vector such that $c_i^{\{r\}} = [g^{\{r\}}]^i(\hat{\mathbf{u}})$. We can perform a further change of variables $v_j = \sqrt{n}\sqrt{c_{jj}^{\{r\}}}u_j$ so that

$$I^{\{r\}} \mathrel{\dot\sim} \frac{C^{\{r\}}}{\sqrt{n}\sqrt{c_{rr}^{\{r\}}}} \int_{\hat{\mathbf{v}}-\mathrm{i}\infty}^{\hat{\mathbf{v}}+\mathrm{i}\infty} \frac{\exp((1/2)\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{v}}^{\{r\}}\mathbf{v} - \bar{\mathbf{y}}_{\mathbf{v}}^{\{r\}}\mathbf{v})}{(2\pi \mathrm{i})^{d_0}\prod_{j\neq r} v_j} h_{\mathbf{v}}^{\{r\}}(\mathbf{v}) \,\mathrm{d}\mathbf{v}, \qquad (3.7)$$

where $C^{\{r\}} = \exp(n[c_{00}^{\{r\}} + \frac{1}{2}\hat{\mathbf{u}}^{\mathrm{T}}\mathbf{cc}^{\{r\}}\hat{\mathbf{u}}])$, $\boldsymbol{\Sigma}_{\mathbf{v}}^{\{r\}}$ is the covariance matrix with elements $[\boldsymbol{\Sigma}_{\mathbf{v}}^{\{r\}}]_{ij} = c_{ij}^{\{r\}}/\sqrt{c_{ii}^{\{r\}}c_{jj}^{\{r\}}}$ and $[\bar{\mathbf{y}}_{\mathbf{v}}]_j^{\{r\}} = \frac{\sqrt{n}[\mathbf{cc}^{\{r\}}\hat{\mathbf{u}}]_j}{\sqrt{c_{jj}^{\{r\}}}}$. The function $h_{\mathbf{v}}^{\{r\}}(\mathbf{v})$ is analytic, but $\frac{h_{\mathbf{v}}^{\{r\}}(\mathbf{v})}{\prod_{j\neq r} v_j}$ is not analytic, and we cannot use Watson's lemma directly. We use the following technique. Let

$t_r = [\boldsymbol{\Sigma}_{\mathbf{v}}^{\{r\}}\mathbf{v}]_r$ and $t_j = \sqrt{1 - ([\Sigma_{\mathbf{v}}^{\{r\}}]_{rj})^2}v_j$ for $j \neq r$. Perform a change of variables to obtain

$$I^{\{r\}} \mathrel{\dot\sim} \frac{C^{\{r\}}}{\sqrt{n}\sqrt{c_{11}^{\{r\}}}} \int_{\hat{\mathbf{t}}-\mathrm{i}\infty}^{\hat{\mathbf{t}}+\mathrm{i}\infty} \frac{\exp(Q^{\{r\}}(\mathbf{t}))}{(2\pi\mathrm{i})^{d_0}\prod_{j\neq r} t_j} h_{\mathbf{t}}^{\{r\}}(\mathbf{t})\,\mathrm{d}\mathbf{t}, \tag{3.8}$$

where $Q^{\{r\}}(\mathbf{t}) = \frac{1}{2}\mathbf{t}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{t}}^{\{r\}}\mathbf{t} - \bar{\mathbf{y}}_{\mathbf{t}}^{\{r\}}\mathbf{t}$, here $\boldsymbol{\Sigma}_{\mathbf{t}}^{\{r\}}$ being the matrix with elements $[\boldsymbol{\Sigma}_{\mathbf{t}}^{\{r\}}]_{rj} = 0$ for $j \neq r$,

$$\big[\boldsymbol{\Sigma}_{\mathbf{t}}^{\{r\}}\big]_{jk} = \frac{[\Sigma_{\mathbf{v}}^{\{r\}}]_{jk} - [\Sigma_{\mathbf{v}}^{\{r\}}]_{rj}[\Sigma_{\mathbf{v}}^{\{r\}}]_{rk}}{\sqrt{[\Sigma_{\mathbf{v}}^{\{r\}}]_{rj}[\Sigma_{\mathbf{v}}^{\{r\}}]_{rk}}} \qquad \text{for } j,k \neq r,$$

$\bar{\mathbf{y}}_{\mathbf{t}}^{\{r\}}$ being the vector with elements

$$\big[\bar{\mathbf{y}}_{\mathbf{t}}^{\{r\}}\big]_r = \big[\bar{\mathbf{y}}_{\mathbf{v}}^{\{r\}}\big]_r \quad \text{and} \quad \big[\bar{\mathbf{y}}_{\mathbf{t}}^{\{r\}}\big]_j = \frac{[\bar{\mathbf{y}}_{\mathbf{v}}^{\{r\}}]_j - [\Sigma_{\mathbf{v}}^{\{r\}}]_{rj}[\bar{\mathbf{y}}_{\mathbf{v}}^{\{r\}}]_r}{\sqrt{1 - ([\Sigma_{\mathbf{v}}^{\{r\}}]_{rj})^2}} \qquad \text{for } j \neq r.$$

For a set s, let $\mathbf{t}^s$ denote the vector such that $[\mathbf{t}^s]_k = 0$ if $k \notin s$ and $[\mathbf{t}^s]_k = t_k$ if $k \in s$. We have

$$I^{\{r\}} \mathrel{\dot\sim} \frac{C^{\{r\}}}{\sqrt{n}\sqrt{c_{11}^{\{r\}}}} \int_{\hat{\mathbf{t}}-\mathrm{i}\infty}^{\hat{\mathbf{t}}+\mathrm{i}\infty} \frac{\exp(Q^{\{r\}}(\mathbf{t}))}{(2\pi\mathrm{i})^{d_0}\prod_{k\neq r} t_k} h_{\mathbf{t}}^{\{r\}}\big(\mathbf{t}^{\{r\}}\big)\,\mathrm{d}\mathbf{t}. \tag{3.9}$$

The argument that the above holds follows similar reasoning as in (2.11), except that we only need to consider the main term here. Now, because t_r can be separated after the change of variable and by Watson's lemma, we have

$$\begin{aligned} I^{\{r\}} &\mathrel{\dot\sim} \frac{C^{\{r\}}}{\sqrt{n}\sqrt{c_{11}^{\{r\}}}} \int_{\hat{\mathbf{t}}^{\{r\}}-\mathrm{i}\infty}^{\hat{\mathbf{t}}^{\{r\}}+\mathrm{i}\infty} \frac{\exp(Q^{\{r\}}(\mathbf{t}) - ((1/2)t_r^2 - [\bar{\mathbf{y}}_{\mathbf{t}}^{\{r\}}]_r t_r)}{(2\pi\mathrm{i})^{d_0-1}\prod_{k\neq r} t_k} \\ &\quad \times \int_{t_r-\mathrm{i}\infty}^{t_r+\mathrm{i}\infty} \frac{\exp((1/2)t_r^2 - [\bar{\mathbf{y}}_{\mathbf{t}}^{\{r\}}]_r t_r)}{2\pi\mathrm{i}} h_{\mathbf{t}}^{\{r\}}\big(\mathbf{t}^{\{r\}}\big)\,\mathrm{d}\mathbf{t} \\ &\mathrel{\dot\sim} \frac{C^{\{r\}}h_{\mathbf{t}}^{\{r\}}(\hat{\mathbf{t}}_r)}{\sqrt{nc_{11}^{\{r\}}}} \phi\big(\big[\bar{\mathbf{y}}_{\mathbf{t}}^{\{r\}}\big]_r\big)\bar{\Phi}\big(\bar{\mathbf{y}}^{\{r\}}, \Sigma^{\{r\}}\big), \end{aligned} \tag{3.10}$$

where $\bar{\mathbf{y}}^{\{r\}}$ is $\bar{\mathbf{y}}_{\mathbf{t}}^{\{r\}}$ with the rth element removed and $\boldsymbol{\Sigma}^{\{r\}}$ is $\boldsymbol{\Sigma}_{\mathbf{t}}^{\{r\}}$ with the rth row and column removed.

Multivariate tail probability approximations for unit lattice variables follow along the same lines, except that

$$G(\boldsymbol{\tau}) = \frac{\prod_{j=1}^{d_0}(w_j - \tilde{w}_j(\mathbf{w}_{j-1}))}{\prod_{j=1}^{d_0} 2\sinh(\tau_j/2)(\mathbf{w}_j)} \prod_{j=1}^{d_0} \frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}.$$

Since $\lim_{x\to 0}(2\sinh(x/2)/x)=1$, any analytic property in the continuous case still holds in the lattice case.

4. Multivariate conditional distribution approximation

Consider a multivariate canonical exponential family. In practice, we are often only interested in a subset of the parameters in a given statistical model, with the other model parameters usually treated as nuisance parameters. The distribution of the sufficient statistics associated with the parameters of interest, conditional on the sufficient statistics associated with the nuisance parameters, contains the parameters of interest and not the nuisance parameters. We can therefore use the conditional distributions instead of the original distribution in the study. For instance, in testing equality of proportions for a 2×2 contingency table, we condition on the row or column margins. Another example is logistic regression, where inference on some regression parameters is often performed conditionally on sufficient statistics associated with nuisance parameters.

Hypotheses involving parameters of interest may be tested by computing the tail probabilities for the conditional distribution $P(\mathbf{T}_{d_0}\ge \mathbf{t}_{d_0}|\mathbf{T}_{-d_0}=\mathbf{t}_{-d_0})$. [11] applies double saddlepoint approximation to the problem in the case where $d_0=1$, $d>1$ and $\mathbf{T}$ is the mean of independent and identically distributed random vectors. Here, we propose a method that extends the results to $d_0>1$ and $d>d_0$, using the idea from the previous sections.

First, consider $\mathbf{T}$, the mean of independent and identically distributed continuous random vectors. Then

$$P(\mathbf{T}_{d_0}\ge \mathbf{t}_{d_0}|\mathbf{T}_{-d_0}=\mathbf{t}_{-d_0})=\frac{\int_{\mathbf{t}_{d_0}}^{\infty} f_{\mathbf{T}}(y_1,\ldots,y_{d_0},t_{d_0+1},\ldots,t_d)\,\mathrm{d}\mathbf{y}_{d_0}}{f_{\mathbf{T}_{-d_0}}(\mathbf{t}_{-d_0})},$$

where $f_{\mathbf{T}}(\cdot)$ is the joint density and $f_{\mathbf{T}_{-d_0}}(\cdot)$ is the marginal density of $\mathbf{T}_{-d_0}$. Again, we use the Fourier inversion formula to obtain

$$P(\mathbf{T}_{d_0}\ge \mathbf{t}_{d_0}|\mathbf{T}_{-d_0}=\mathbf{t}_{-d_0})=\frac{n^{d-d_0}}{(2\pi\mathrm{i})^d}\int_{\mathbf{c}-\mathrm{i}\infty}^{\mathbf{c}+\mathrm{i}\infty}\frac{\exp(n[K(\boldsymbol{\tau})-\boldsymbol{\tau}^{\mathrm{T}}\mathbf{t}])}{\prod_{j=1}^{d_0}\tau_j}\,\mathrm{d}\boldsymbol{\tau}\Big/ f_{\mathbf{T}_{-d_0}}(\mathbf{t}_{-d_0}), \tag{4.1}$$

where $K(\boldsymbol{\tau})$ is the cumulant generating function of the random vector $\mathbf{T}$. The numerator is just a special case of (2.1).

Approximation (2.10) holds because of the following lemma which will allow us to apply previous unconditional results by substituting components of $\hat{\mathbf{w}}$ for components of $\mathbf{w}$ when the components correspond to variables in the conditioning event.

Lemma 4.1.

$$\begin{aligned}&\frac{n^{d-d_0}}{(2\pi\mathrm{i})^d}\int_{\hat{\mathbf{w}}-\mathrm{i}\mathbf{K}}^{\hat{\mathbf{w}}+\mathrm{i}\mathbf{K}}\frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w}-\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{w}])}{\prod_{j=1}^{d_0}(w_j-\tilde{w}_j(\mathbf{w}_{j-1}))}\cdot\prod_{j=1}^{d}\frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\frac{\prod_{j=1}^{d_0}(w_j-\tilde{w}_j(\mathbf{w}_{j-1}))}{\prod_{j=1}^{d_0}\rho(\tau_j(\mathbf{w}_j))}\,\mathrm{d}\mathbf{w}\\&\quad=\frac{n^{d-d_0}}{(2\pi\mathrm{i})^d}\int_{\hat{\mathbf{w}}-\mathrm{i}\mathbf{K}}^{\hat{\mathbf{w}}+\mathrm{i}\mathbf{K}}\frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w}-\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{w}])}{\prod_{j=1}^{d_0}(w_j-\tilde{w}_j(\mathbf{w}_{j-1}))}G(\boldsymbol{\tau})\,\mathrm{d}\mathbf{w}\left(1+\mathrm{O}(n^{-1})\right),\end{aligned} \tag{4.2}$$

where

$$G(\boldsymbol{\tau}) = \frac{\prod_{j=1}^{d_0} (w_j - \tilde{w}_j(\mathbf{w}_{j-1}))}{\prod_{j=1}^{d_0} \rho(\tau_j(\mathbf{w}_j))} \cdot \prod_{j=1}^{d} \frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\bigg|_{(\mathbf{w}_{d_0}, \hat{\mathbf{w}}_{-d_0})}.$$

Proof. By Watson's lemma, given fixed $\mathbf{w}_{d_0}$, we have

$$\int_{\hat{\mathbf{w}}_{-d_0}-\mathrm{i}\mathbf{K}}^{\hat{\mathbf{w}}_{-d_0}+\mathrm{i}\mathbf{K}} \exp\biggl(n\biggl[\frac{1}{2}\mathbf{w}_{-d_0}^{\mathrm{T}}\mathbf{w}_{-d_0} - \hat{\mathbf{w}}_{-d_0}^{\mathrm{T}}\mathbf{w}_{-d_0}\biggr]\biggr) \prod_{j=d_0+1}^{d} \frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\, \mathrm{d}\mathbf{w}_{-d_0}$$

$$= \int_{\hat{\mathbf{w}}_{-d_0}-\mathrm{i}\mathbf{K}}^{\hat{\mathbf{w}}_{-d_0}+\mathrm{i}\mathbf{K}} \exp\biggl(n\biggl[\frac{1}{2}\mathbf{w}_{-d_0}^{\mathrm{T}}\mathbf{w}_{-d_0} - \hat{\mathbf{w}}_{-d_0}^{\mathrm{T}}\mathbf{w}_{-d_0}\biggr]\biggr) \prod_{j=d_0+1}^{d} \frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\bigg|_{(\mathbf{w}_{d_0}, \hat{\mathbf{w}}_{-d_0})}$$

$$\times \biggl(1 + \frac{E(\mathbf{w}_{d_0})}{n}\biggr)\mathrm{d}\mathbf{w}_{-d_0},$$

for some analytic function $E(\mathbf{w}_{d_0})$ of $\mathrm{O}(1)$. Therefore,

$$LHS = \frac{n^{d-d_0}}{(2\pi \mathrm{i})^d} \int_{\hat{\mathbf{w}}_{d_0}-\mathrm{i}\mathbf{K}}^{\hat{\mathbf{w}}_{d_0}+\mathrm{i}\mathbf{K}} \frac{\exp(n[(1/2)\mathbf{w}_{d_0}^{\mathrm{T}}\mathbf{w}_{d_0} - \hat{\mathbf{w}}_{d_0}^{\mathrm{T}}\mathbf{w}_{d_0}])}{\prod_{j=1}^{d_0} \rho(\tau_j(\mathbf{w}_j))} \prod_{j=1}^{d_0} \frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}$$

$$\times \int_{\hat{\mathbf{w}}_{-d_0}-\mathrm{i}\mathbf{K}}^{\hat{\mathbf{w}}_{-d_0}+\mathrm{i}\mathbf{K}} \exp\biggl(n\biggl[\frac{1}{2}\mathbf{w}_{-d_0}^{\mathrm{T}}\mathbf{w}_{-d_0} - \hat{\mathbf{w}}_{-d_0}^{\mathrm{T}}\mathbf{w}_{-d_0}\biggr]\biggr) \prod_{j=d_0+1}^{d} \frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\bigg|_{(\mathbf{w}_{d_0}, \hat{\mathbf{w}}_{-d_0})}$$

$$\times \biggl(1 + \frac{E(\mathbf{w}_{d_0})}{n}\biggr)\mathrm{d}\mathbf{w}_{-d_0}\,\mathrm{d}\mathbf{w}_{d_0}$$

$$= A\biggl(1 + \frac{1}{n}\frac{B}{A}\biggr),$$

where

$$A = \frac{n^{d-d_0}}{(2\pi \mathrm{i})^d} \int_{\hat{\mathbf{w}}-\mathrm{i}\mathbf{K}}^{\hat{\mathbf{w}}+\mathrm{i}\mathbf{K}} \frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w} - \hat{\mathbf{w}}^{\mathrm{T}}\mathbf{w}])}{\prod_{j=1}^{d_0} (w_j - \tilde{w}_j(\mathbf{w}_{j-1}))} G(\boldsymbol{\tau})\,\mathrm{d}\mathbf{w}$$

and

$$B = \frac{n^{d-d_0}}{(2\pi \mathrm{i})^d} \int_{\hat{\mathbf{w}}-\mathrm{i}\mathbf{K}}^{\hat{\mathbf{w}}+\mathrm{i}\mathbf{K}} \frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w} - \hat{\mathbf{w}}^{\mathrm{T}}\mathbf{w}])}{\prod_{j=1}^{d_0} (w_j - \tilde{w}_j(\mathbf{w}_{j-1}))} G(\boldsymbol{\tau}) E(\mathbf{w}_{d_0})\,\mathrm{d}\mathbf{w}.$$

If A and B are expanded according to [5], each integral is approximated by a tilting term times a normal multivariate tail probability, up to relative order $\mathrm{O}(1/\sqrt{n})$. The expression for B is also multiplied by the leading term of E. Hence, $A/B = \mathrm{O}(1)$ and, therefore, the left-hand side equals $A(1 + \mathrm{O}(n^{-1}))$. □

To deal with the denominator in (4.1), [7] demonstrates that

$$\left(\frac{n}{2\pi \mathrm{i}}\right)^{d-d_0} \int_{\hat{\mathbf{w}}_{-d_0}-\mathrm{i}\infty}^{\hat{\mathbf{w}}_{-d_0}+\mathrm{i}\infty} \exp\biggl(n\biggl[\frac{1}{2}\mathbf{w}_{-d_0}^{\mathrm{T}}\mathbf{w}_{-d_0} - \hat{\mathbf{w}}_{-d_0}^{\mathrm{T}}\mathbf{w}_{-d_0}\biggr]\biggr) \prod_{j=d_0+1}^{d} \frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\bigg|_{(\mathbf{0}_{d_0},\hat{\mathbf{w}}_{-d_0})} \mathrm{d}\mathbf{w}_{-d_0}$$
$$= f_{\mathbf{T}_{-d_0}}(\mathbf{t}_{-d_0})\bigl(1+\mathrm{O}(n^{-1})\bigr). \tag{4.3}$$

This development is similar to that of [4], page 147.

With continuous variables, we can decompose A according to (2.11) with

$$G(\boldsymbol{\tau}) = \prod_{j=1}^{d_0}\biggl(\frac{w_j - \tilde{w}_j(\mathbf{w}_{j-1})}{\tau_j}\frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\biggr)\prod_{j=d_0+1}^{d}\frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\bigg|_{(\mathbf{w}_{d_0},\hat{\mathbf{w}}_{-d_0})}. \tag{4.4}$$

Denote the left-hand side of (4.3) by J_{-d_0}. Note that $G(\mathbf{0}) = \prod_{j=d_0+1}^{d}\frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}|_{(\mathbf{0}_{d_0},\hat{\mathbf{w}}_{-d_0})}$. The main term is then

$$I^{\varnothing} = \int_{\hat{\mathbf{w}}-\mathrm{i}\infty}^{\hat{\mathbf{w}}+\mathrm{i}\infty} \frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w} - \hat{\mathbf{w}}^{\mathrm{T}}\hat{\mathbf{w}}])}{(2\pi \mathrm{i})^{d_0}\prod_{j=1}^{d_0}(w_j - \tilde{w}_j(\mathbf{w}_{j-1}))} \frac{n^{d-d_0}}{(2\pi\mathrm{i})^{d-d_0}} \prod_{j=d_0+1}^{d}\frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\bigg|_{(\mathbf{0}_{d_0},\hat{\mathbf{w}}_{-d_0})} \mathrm{d}\mathbf{w}$$
$$= \int_{\hat{\mathbf{w}}-\mathrm{i}\infty}^{\hat{\mathbf{w}}+\mathrm{i}\infty} \frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w} - \hat{\mathbf{w}}^{\mathrm{T}}\hat{\mathbf{w}}])}{(2\pi \mathrm{i})^{d_0}\prod_{j=1}^{d_0}(w_j - \tilde{w}_j(\mathbf{w}_{j-1}))} \mathrm{d}\mathbf{w}_{d_0}\cdot J_{-d_0} \tag{4.5}$$
$$\sim \int_{\hat{\mathbf{w}}-\mathrm{i}\infty}^{\hat{\mathbf{w}}+\mathrm{i}\infty} \frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w} - \hat{\mathbf{w}}^{\mathrm{T}}\hat{\mathbf{w}}])}{(2\pi \mathrm{i})^{d_0}\prod_{j=1}^{d_0}(w_j - \tilde{w}_j(\mathbf{w}_{j-1}))} \mathrm{d}\mathbf{w}_{d_0}\cdot f_{\mathbf{T}_{-d_0}}(\mathbf{t}_{-d_0}),$$

where

$$\int_{\hat{\mathbf{w}}-\mathrm{i}\infty}^{\hat{\mathbf{w}}+\mathrm{i}\infty} \frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w} - \hat{\mathbf{w}}^{\mathrm{T}}\hat{\mathbf{w}}])}{(2\pi \mathrm{i})^{d_0}\prod_{j=1}^{d_0}(w_j - \tilde{w}_j(\mathbf{w}_{j-1}))} \mathrm{d}\mathbf{w}_{d_0} \tag{4.6}$$

can be obtained by formula (3.4).

Using the same technique as in (3.5)–(3.10), we have

$$I^{\{r\}} \mathrel{\dot{\sim}} \frac{n^{d-d_0}}{(2\pi d)^d}\int_{\hat{\mathbf{w}}-\mathrm{i}\infty}^{\hat{\mathbf{w}}+\mathrm{i}\infty} \frac{\exp(n[(1/2)\mathbf{w}^{\mathrm{T}}\mathbf{w} - \hat{\mathbf{w}}^{\mathrm{T}}\hat{\mathbf{w}}])}{\prod_{j\neq r}(w_j - \tilde{w}(\mathbf{w}_{j-1}))} \frac{G^{\{r\}}(\boldsymbol{\tau}) - G(\mathbf{0})}{w_r - \tilde{w}(\mathbf{w}_{r-1})}\mathrm{d}\mathbf{w}$$
$$\mathrel{\dot{\sim}} \frac{C^{\{r\}}h_{\mathbf{t}}^{\{r\}}(\hat{\mathbf{t}}_r)}{\sqrt{nc_{11}^{\{r\}}}\prod_{j=d_0+1}^{d}\mathrm{d}\tau_j/\mathrm{d}w_j|_{(\mathbf{0}_{d_0},\hat{\mathbf{w}}_{-d_0})}} \phi\bigl(\bigl[\bar{\mathbf{y}}_{\mathbf{t}}^{\{r\}}\bigr]_r\bigr)\bar{\Phi}\bigl(\bar{\mathbf{y}}^{\{r\}}, \Sigma^{\{r\}}\bigr)\cdot J_{-d_0} \tag{4.7}$$

at $\mathrm{O}(n^{-1})$. The computation involves $\prod_{j=d_0+1}^{d}\frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}|_{(w_1,w_2,\hat{\mathbf{w}}_{-2})}$, which can be obtained using (2.9).

In summary, in the conditional case, $P(\mathbf{T}_{d_0} > \mathbf{t}_{d_0}|\mathbf{T}_{-d_0} > \mathbf{t}_{-d_0}) \sim \sum_{|s|\le 1, s\subseteq U} I^s/f_{-d_0}(\mathbf{t}_{-d_0})$, where $U = \{1, 2, \ldots, d_0\}$.

Similarly to the unconditional case, in the case of unit lattice variables, we have

$$G(\boldsymbol{\tau}) = \prod_{j=1}^{d_0}\left(\frac{w_j - \tilde{w}_j(\mathbf{w}_{j-1})}{2\sinh(\tau_j/2)}\frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\right)\prod_{j=d_0+1}^{d}\frac{\mathrm{d}\tau_j}{\mathrm{d}w_j}\bigg|_{(\mathbf{w}_{d_0},\hat{\mathbf{w}}_{-d_0})}. \tag{4.8}$$

Other analytic properties and formulae still hold.

5. Five examples

We present five examples here. The fourth example is based on real data.

In the first example, we consider the bivariate random vector (Y_1, Y_2), with $Y_1 = X_1 + X_2$ and $Y_2 = X_2 + X_3$, where X_1, X_2 and X_3 are independent and identically distributed random variables following the exponential distribution, which has a density function $f(x) = \mathrm{e}^{-x}$ for $x > 0$. The results for approximating $P(\bar{Y}_1 \geq \bar{y}_1, \bar{Y}_2 \geq \bar{y}_2)$ when $n = 5$ are listed in Table 1, where "P. approx." stands for the saddlepoint approximation proposed in this paper, "K. approx." stands for the saddlepoint approximation presented in [5] and "N. approx." stands for bivariate normal approximation. The "exact" column shows the exact tail probability values computed in [15]. The "relative error" column shows the relative error of "P. approx." The results for the cases $(\bar{y}_1, \bar{y}_2) = (2.5, 3.0)$ and $(\bar{y}_1, \bar{y}_2) = (3.0, 4.0)$ are the special cases where $\hat{w}_1 = 0$, which we have mentioned, but which are omitted here because of the removable singularity. The normal approximation deteriorates at the far tail, while both saddlepoint approximations show much better and more stable relative errors. In almost all cases, the new method shows smaller relative errors than those in [5].

In the second example, we consider the bivariate random vector (Y_1, Y_2), with $Y_1 = X_1 + X_2$ and $Y_2 = X_2 + X_3$, where X_1, X_2 and X_3 are independent and identically distributed random variables following the binomial distribution, which has a mass function $\binom{N}{x}p^x(1-p)^{N-x}$ for $0 \leq x \leq N$. The results for approximating $P(\bar{Y}_1 \geq \bar{y}_1, \bar{Y}_2 \geq \bar{y}_2)$ when $N = 10$, $p = 0.2$ and $n = 8$ are displayed in Table 2. We can again see from the table that the normal approximation (with adjustment for continuity) deteriorates at the far tail, while the saddlepoint approximations show

Table 1. Results of saddlepoint approximation compared with other approximations in the continuous case

$\bar{y}_1$	$\bar{y}_2$	P. approx.	K. approx.	N. approx.	Exact	Relative error
2.5	2.5	9.12×10^{-2}	8.98×10^{-2}	9.65×10^{-2}	9.22×10^{-2}	−1.08%
2.5	3.5	1.41×10^{-2}	1.41×10^{-2}	6.54×10^{-3}	1.41×10^{-2}	0.00%
2.5	4.0	3.91×10^{-3}	3.99×10^{-3}	6.69×10^{-3}	3.93×10^{-3}	−0.51%
3.0	3.0	2.20×10^{-2}	2.14×10^{-2}	1.46×10^{-2}	2.22×10^{-2}	−0.90%
3.0	3.5	8.97×10^{-3}	8.73×10^{-3}	3.52×10^{-3}	8.96×10^{-3}	0.11%
3.5	3.5	4.40×10^{-3}	4.25×10^{-3}	1.09×10^{-3}	4.40×10^{-3}	0.00%
3.5	4.0	1.67×10^{-3}	1.61×10^{-3}	1.78×10^{-4}	1.66×10^{-3}	0.60%
4.0	4.0	7.69×10^{-4}	7.34×10^{-4}	3.88×10^{-5}	7.58×10^{-4}	1.45%

Table 2. Results of saddlepoint approximation compared with other approximations in the unit lattice case

$\bar{y}_1$	$\bar{y}_2$	P. approx.	K. approx.	N. approx.	Exact	Relative error
4.5	4.5	1.15×10^{-1}	1.16×10^{-1}	1.16×10^{-1}	1.15×10^{-1}	0.00%
4.5	5.0	4.43×10^{-2}	4.51×10^{-2}	4.28×10^{-2}	4.44×10^{-2}	−0.23%
4.5	5.5	1.04×10^{-2}	1.05×10^{-2}	8.73×10^{-3}	1.04×10^{-2}	0.00%
4.5	6.0	1.46×10^{-3}	1.45×10^{-3}	9.50×10^{-4}	1.46×10^{-3}	0.00%
5.0	5.0	2.07×10^{-2}	2.12×10^{-2}	1.92×10^{-2}	2.08×10^{-2}	−0.48%
5.0	5.5	5.89×10^{-3}	6.04×10^{-3}	4.85×10^{-3}	5.91×10^{-3}	−0.34%
5.0	6.0	9.91×10^{-4}	1.01×10^{-3}	6.40×10^{-4}	9.94×10^{-4}	−0.30%
5.5	5.5	2.11×10^{-3}	2.16×10^{-3}	1.57×10^{-3}	2.11×10^{-3}	0.00%
5.5	6.0	4.45×10^{-4}	4.56×10^{-4}	2.69×10^{-4}	4.47×10^{-4}	−0.45%
6.0	6.0	1.21×10^{-4}	1.24×10^{-4}	6.14×10^{-5}	1.21×10^{-4}	0.00%

much better and more stable relative errors. In most cases, the new approximation shows better accuracy than that of [5].

The third example involves conditional distribution functions. Let X_i, $i = 1, 2, 3$, be independent and identically distributed random variables following the exponential distribution, as in the first example. Consider the random vector (Y_1, Y_2, Y_3) with $Y_1 = X_2$, $Y_2 = X_3$ and $Y_3 = X_1 + X_2 + X_3$. The results for approximating $P(\bar{Y}_1 \geq \bar{y}_1, \bar{Y}_2 \geq \bar{y}_2 | \bar{Y}_3 = \bar{y}_3)$ when $n = 10$ are shown below in Table 3. The case where $\bar{y}_1 = 2.0$, $\bar{y}_2 = 2.5$ and $\bar{y}_3 = 7.0$ is the special case where both $\tilde{\tau}_2(0) = 0$ and $\hat{w}_2 = 0$, as discussed in Section 3, and is omitted here. The cases where $\bar{y}_1 = 2.0$, $\bar{y}_2 = 3.0$ and $\bar{y}_3 = 7.0$, and $\bar{y}_1 = 2.0$, $\bar{y}_2 = 2.5$ and $\bar{y}_3 = 6.5$, are the cases where $\hat{w}_1 = 0$; these are also omitted. The exact values are computed in [15].

The fourth example was used in [5] and [6], which refers to data presented in [12]. The data consist of 63 case-control pairs of women with endometrial cancer. The relationship between

Table 3. Results of saddlepoint approximation compared with bivariate normal approximation in the conditional continuous case

$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	P. approx.	N. approx.	Exact	Relative error
2.0	2.0	7.0	4.42×10^{-1}	8.04×10^{-2}	4.38×10^{-1}	0.91%
2.5	2.5	7.0	6.25×10^{-2}	2.04×10^{-2}	6.32×10^{-2}	−1.11%
2.5	3.0	7.0	8.00×10^{-3}	4.14×10^{-5}	8.54×10^{-3}	−6.32%
3.0	3.0	7.0	3.02×10^{-4}	1.00×10^{-8}	3.46×10^{-4}	−12.7%
2.0	2.0	6.5	2.93×10^{-1}	1.16×10^{-1}	2.91×10^{-1}	0.69%
2.0	3.0	6.5	1.09×10^{-2}	6.48×10^{-5}	1.14×10^{-2}	−4.39%
2.5	2.5	6.5	1.49×10^{-2}	6.96×10^{-4}	1.56×10^{-2}	−4.49%
2.5	3.0	6.5	5.25×10^{-4}	1.57×10^{-7}	6.09×10^{-4}	−13.8%
3.0	3.0	6.5	9.63×10^{-7}	3.67×10^{-12}	1.10×10^{-6}	12.5%

Table 4. Differences between cases and controls for endometrial cancer data

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

the occurrence of endometrial cancer and explanatory variables including gall bladder disease, hypertension and non-estrogen drug use is modeled with logistic regression. [12] noted that the likelihood for these data is equivalent to that of a logistic regression 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 is 1.

The number of pairs with each configuration of differences of the three variables are shown in Table 4. Let $\mathbf{z}_j$, $j = 1, 2, \ldots, 63$ denote the differences of covariates between cases and controls, as given in Table 4. Consider the situation under the null hypothesis, where the linear coefficients are zero. Let $\mathbf{Z}_j$, $j = 1, 2, \ldots, 63$, be the random vectors that take value $\mathbf{z}_j$ with a probability of $\frac{1}{2}$ and $\mathbf{0}$ with a probability of $\frac{1}{2}$. Let $\mathbf{Z}$ be matrix whose rows are $\mathbf{Z}_j$ and where $\mathbf{T} = \mathbf{Z}'\mathbf{1}$, where $\mathbf{1}$ is a column vector with dimension 63. We then have $K(\boldsymbol{\tau}) = \sum_j m_j[\log(\frac{1+\exp(\mathbf{z}_j\boldsymbol{\tau})}{2})]$. [6] tested the association of hypertension or non-estrogen drug use with an increase in endometrial cancer, conditional on the sufficient statistic value associated with gall bladder disease. The test required evaluating the quantity $P(T_2 \geq 10 \text{ or } T_3 \geq 13 | T_1 = 9)$ for $\mathbf{T} = (T_1, T_2, T_3)$. By Boole's law, this probability can be computed using

$$P(T_2 \geq 10 | T_1 = 9) + P(T_3 \geq 13 | T_1 = 9) - P(T_2 \geq 10, T_3 \geq 13 | T_1 = 9).$$

The results for approximating $P(T_2 \geq 10, T_3 \geq 13 | T_1 = 9)$ compared to those listed in [6] are shown in Table 5, where "N. app." stands for normal approximation, "E. app." stands for Edgeworth approximation, "K. app." stands for the approximation presented in [6] and "P. app." is the

Table 5. Endometrial cancer results for some (t_2, t_3) instances

Method	(10, 13)	(9, 12)	(8, 11)	(7, 10)	(6, 9)
N. app.	3.50×10^{-4}	1.78×10^{-3}	7.26×10^{-3}	2.39×10^{-2}	6.39×10^{-2}
E. app.	3.31×10^{-4}	1.72×10^{-3}	7.13×10^{-3}	2.37×10^{-2}	6.37×10^{-2}
K. app.	1.51×10^{-4}	1.07×10^{-3}	5.37×10^{-3}	2.01×10^{-2}	5.84×10^{-2}
P. app.	1.62×10^{-4}	1.13×10^{-3}	5.60×10^{-3}	2.08×10^{-2}	6.00×10^{-2}
Exact	1.52×10^{-4}	1.09×10^{-3}	5.48×10^{-3}	2.05×10^{-2}	5.95×10^{-2}

Table 6. Results of saddlepoint approximation compared with normal approximations for a multivariate gamma distribution

y_1	y_2	y_3	P. approx.	N. approx.	Simulation	Std. err.	Relative error
5.5	5.5	5.5	4.93×10^{-2}	9.64×10^{-2}	5.58×10^{-2}	1.42×10^{-3}	−11.6%
5.5	5.5	6.5	3.65×10^{-2}	7.17×10^{-2}	4.16×10^{-2}	1.24×10^{-3}	−12.3%
5.5	6.5	6.5	2.70×10^{-2}	5.34×10^{-2}	3.26×10^{-2}	1.10×10^{-4}	−17.2%
6.5	6.5	6.5	1.98×10^{-2}	3.99×10^{-2}	2.45×10^{-2}	9.58×10^{-4}	−19.2%
6.5	6.5	7.5	1.45×10^{-2}	2.78×10^{-2}	1.83×10^{-2}	8.31×10^{-4}	−20.8%
6.5	7.5	7.5	1.05×10^{-2}	1.94×10^{-2}	1.34×10^{-2}	7.11×10^{-4}	−21.6%
7.5	7.5	7.5	0.76×10^{-2}	1.36×10^{-3}	1.04×10^{-3}	1.89×10^{-4}	−26.9%

proposed approximation. Approximation results of $P(T_2 \geq t_2, T_3 \geq t_3 | T_1 = 9)$ for other values of t_2 and t_3 are also listed in the table. We can see that the proposed method achieves better results than other methods, except for the method of [6], which is far more complicated computationally.

In the fifth example, we consider a multivariate gamma distribution, which is the diagonal of a Wishart distribution, formed from a 3-variate normal distribution, with covariance matrix

$$V = \begin{pmatrix} 1 & 0.25 & 0.25 \\ 0.25 & 1 & 0.25 \\ 0.25 & 0.25 & 1 \end{pmatrix}$$

and $n = 5$. The results are listed in Table 6, where "P. approx." stands for the saddlepoint approximation proposed in this paper and "N. approx." stands for bivariate normal approximation. The "simulation" and "std. err." column shows the simulation results and 5% standard error. The "relative error" column shows the relative error of "P. approx." compared with the simulation results. We can see that the proposed approximation performs better than the normal approximation.

Acknowledgements

This research was supported in part by NSF Grant DMS-0505499. The authors would like to thank the referees and the Associate Editor for their helpful suggestions.

References

[1] Daniels, H.E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25** 631–645. MR0066602

[2] Daniels, H.E. (1987). Tail probability approximations. *Internat. Statist. Rev.* **55** 37–46. MR0962940

[3] Jensen, J.L. (1992). The modified signed likelihood statistic and saddlepoint approximations. *Biometrika* **79** 693–703. MR1209471

[4] Kolassa, J.E. (2006). *Series Approximation Methods in Statistics*, 3rd ed. *Lecture Notes in Statistics* **88**. New York: Springer. MR1487639

[5] Kolassa, J.E. (2003). Multivariate saddlepoint tail probability approximations. *Ann. Statist.* **31** 274–286. MR1962507

[6] Kolassa, J.E. (2004). Approximate multivariate conditional inference using the adjusted profile likelihood. *Canad. J. Statist.* **32** 5–14. MR2060541

[7] Li, J. (2008). Multivariate saddlepoint tail probability approximations, for conditional and unconditional distributions, based on the signed root of the log likelihood ratio statistic. Ph.D. thesis, Rutgers Univ., Dept. Statistics and Biostatistics.

[8] Lugannani, R. and Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of indepednent random variables. *Adv. in Appl. Probab.* **12** 475–490. MR0569438

[9] Reid, N. (1988). Saddlepoint methods and statistical inference (with discussion). *Statist. Sci.* **3** 213–238. MR0968390

[10] Robinson, J. (1982). Saddlepoint approximation for permutation tests and confidence intervals. *J. R. Statist. Soc. Ser. B* **44** 91–101. MR0655378

[11] Skovgaard, I.M. (1987). Saddlepoint expansions for conditional distributions. *J. Appl. Probab.* **24** 875–887. MR0913828

[12] Stokes, M.E., David, C.S. and Koch, G.G. (1995). *Categorical Data Analysis Using the SAS System*. Cary, NC.

[13] Wang, S. (1991). Saddlepoint approximation for bivariate distribution. *J. Appl. Probab.* **27** 586–597. MR1067024

[14] Watson, G.N. (1948). *Theory of Bessel Functions*. Cambridge: Cambridge Univ. Press.

[15] Wolfram Research, Inc. (2005). *Mathematica. Version 5.0*. Champaign, IL: Wolfram Research, Inc.

Received September 2008 and revised August 2009