The Annals of Statistics
2011, Vol. 39, No. 6, 3357–3368
DOI: 10.1214/11-AOS945

SADDLEPOINT APPROXIMATIONS FOR LIKELIHOOD RATIO LIKE STATISTICS WITH APPLICATIONS TO PERMUTATION TESTS

BY JOHN KOLASSA[1] AND JOHN ROBINSON[2]

Rutgers University and University of Sydney

We obtain two theorems extending the use of a saddlepoint approximation to multiparameter problems for likelihood ratio-like statistics which allow their use in permutation and rank tests and could be used in bootstrap approximations. In the first, we show that in some cases when no density exists, the integral of the formal saddlepoint density over the set corresponding to large values of the likelihood ratio-like statistic approximates the true probability with relative error of order $1/n$. In the second, we give multivariate generalizations of the Lugannani–Rice and Barndorff-Nielsen or r^* formulas for the approximations. These theorems are applied to obtain permutation tests based on the likelihood ratio-like statistics for the k sample and the multivariate two-sample cases. Numerical examples are given to illustrate the high degree of accuracy, and these statistics are compared to the classical statistics in both cases.

1. Introduction.

In parametric problems where distributions are specified exactly, the likelihood ratio is generally used for hypothesis testing whenever possible. In multiparameter problems, the distribution of twice the log likelihood ratio is approximated by a chi-squared distribution. Refinements of this approximation were obtained by Barndorff-Nielsen [2] for parametric problems. In a nonparametric setting the empirical exponential likelihood is described in Chapter 10 of [6] and discussed in a number of references cited there. Saddlepoint approximations for empirical exponential likelihood statistics based on multiparameter M-estimates are given, for example, in [12] and for tests of means in [10], under the strong assumption that the density of the M-estimate exists and has a saddlepoint approximation. They used methods based on those of [3] to obtain an approximation analogous to the Lugananni–Rice approximation for the one-dimensional case.

It is the purpose of this paper to show that, under conditions which will allow the application of the approximations in bootstrap, permutation and rank statistics used for multiparameter cases, the integral of the formal saddlepoint density

Received April 2011; revised August 2011.

[1]Supported in part by NSF Grant DMS-09-06569.

[2]Supported in part by ARC DP0451722.

MSC2010 subject classifications. Primary 62G20, 62G09; secondary 60F10.

Key words and phrases. Randomization tests, nonparametric tests, large deviations.

approximation can be used to give an approximation with relative error of order n^{-1} to the tail probability of a likelihood ratio-like statistic. This then permits the approximation to be put in the Lugananni–Rice form as in [12] and also in a form analogous to the r^* or Barndorff-Nielsen form given in [2] and [8] for the one-dimensional case. These results are then applied to two multiparameter nonparametric cases. We require the existence of a moment generating function. This may be too strong an assumption in the case of tests concerning means considered here, but robust versions of these, as in [12], can be used to make the results widely applicable.

The statistic used is obtained by using the conjugate distribution approach of [5] and is the log likelihood ratio in the parametric case of exponential families. It can be written as a convex function of $\bar{\mathbf{X}}$, the mean of n independent random variables. This statistic can be approximated to first order by a quadratic form in the means $\bar{\mathbf{X}}$. However, it does not seem to be possible to approximate tail probabilities for quadratic forms with relative errors of order n^{-1}, as are obtained for our statistic. Cramér large deviation results for the case of quadratic forms in multivariate means were obtained by [9] and a number of earlier authors cited in that paper, but the relative errors for the approximation to the probability of the statistic, a random variable of order $1/n$, exceeding λ is of order $\sqrt{n\lambda}n^{-1/4}$. So the relative error is at best of order $n^{-1/4}$. The same problem arises in the case of an empirical likelihood statistic, where we know of no saddlepoint approximation.

In the next section we introduce the notation and assumptions necessary to obtain the likelihood ratio-like statistic, tail probabilities of which can be used for hypothesis testing in multivariate nonparametric settings. We reduce certain conditional cases given lattice variables to a more convenient notation and state the main result in a theorem showing that tail probabilities for the statistic can be approximated, to relative order n^{-1}, by an integral of a formal saddlepoint density. We then state and prove a theorem giving the integrals in forms like those of Lugananni–Rice and Barndorff-Nielsen in the one-dimensional case. In Section 3 we consider two examples of permutation tests, for the k-sample problem and for a two sample multivariate permutation test, using the results of the previous section to obtain explicit formulas for test statistics and for the approximations of the tail probabilities of these statistics under permutations. We then present numerical examples illustrating the accuracy of the approximations and comparing results to those obtained using the standard sum of squares test statistics for the k-sample permutation and rank tests and the Mahalanobis D^2 test for the 2-sample multivariate test. In the final section we give the proof of the main result.

2. Notation and main result. For a sample of size n with mean vector $\bar{\mathbf{x}}$ from a parametric canonical exponential family with density $f_{\boldsymbol{\tau}}(x) = \exp(\boldsymbol{\tau}^\top \mathbf{x} - \kappa(\boldsymbol{\tau}))g(\mathbf{x})$, the maximum likelihood estimate of $\boldsymbol{\tau}$ is $\hat{\boldsymbol{\tau}}$, the solution of $\kappa'(\boldsymbol{\tau}) = \bar{\mathbf{x}}$, and, taking $\kappa'(\mathbf{0}) = \mathbf{0}$, the log likelihood ratio statistic is $\Lambda(\bar{\mathbf{x}}) = \hat{\boldsymbol{\tau}}^\top \bar{\mathbf{x}} - \kappa(\hat{\boldsymbol{\tau}})$. This is used to test the hypothesis that $\boldsymbol{\tau} = \mathbf{0}$, or equivalently, that $\kappa'(\boldsymbol{\tau}) = \mathbf{0}$. For the nonparametric case an empirical exponential family is taken, and it is shown, for example, in [12], page 1163, that the empirical exponential likelihood ratio statistic for a test that the expectation is zero is $\Lambda(\bar{\mathbf{x}}) = -\beta_0^\top \bar{\mathbf{x}} + \kappa_n(\beta_0)$, where $\kappa_n(\beta) = \log[\sum_{i=1}^n \exp(\beta^\top \mathbf{x}_i)]/n$ and β_0 is the solution of $\kappa_n'(\beta) = \mathbf{0}$. In [12] a bootstrap approximation can be based on the statistic $\Lambda(\bar{\mathbf{x}}^*) = \hat{\boldsymbol{\tau}}^\top \bar{\mathbf{x}}^* - \kappa_n(\beta_0 + \hat{\boldsymbol{\tau}}) + \kappa_n(\beta_0)$, where the bootstrap is taken from the tilted empirical distribution $\hat{F}_0(\mathbf{x}) = \sum_{i=1}^n \exp(\beta_0^\top \mathbf{x}_i - \kappa_n(\beta_0)) I\{\mathbf{x}_i \le \mathbf{x}\}/n$. A saddlepoint approximation to this bootstrap is given, but it is noted that the relative errors of this approximation could not be proven from the theorem of that paper. The theorems of this section permit this proof. We use an analogous approach to give the likelihood ratio-like statistics for the two permutation test examples in the next section.

Consider independent d-dimensional random vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$, with the first d_0 components $X_{1j}, \dots, X_{d_0 j}$ confined to a lattice with unit spacings, for $d_0 < d$, and with the average cumulant generating function

$$\kappa(\boldsymbol{\tau}) = n^{-1} \log(E e^{\boldsymbol{\tau}^\top \mathbf{S}_n}) = n^{-1} \sum_{i=1}^n \log(E e^{\boldsymbol{\tau}^\top \mathbf{X}_i}), \tag{1}$$

where $\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$. For some $\mathbf{x}$ we can define

$$\Lambda(\mathbf{x}) = \hat{\boldsymbol{\tau}}^\top \mathbf{x} - \kappa(\hat{\boldsymbol{\tau}}) \tag{2}$$

for $\hat{\boldsymbol{\tau}}$ satisfying

$$\kappa'(\hat{\boldsymbol{\tau}}) = \mathbf{x} \tag{3}$$

and

$$r(\mathbf{x}) = e^{-n\Lambda(\mathbf{x})} (2\pi n)^{-d_0/2} (2\pi/n)^{-d_1/2} |V_{\hat{\boldsymbol{\tau}}}|^{-1/2}. \tag{4}$$

In the case when the last d_1 components of $\mathbf{X}_1, \dots, \mathbf{X}_n$ have densities, this is the saddlepoint density approximation for $\bar{\mathbf{X}} = \mathbf{S}_n/n$, obtained in the case of identically distributed random vectors in [4]. In many cases when these last components lack a density, the theorem below will imply that their distribution may be well approximated by a continuous distribution.

Let μ denote the distribution of $\bar{\mathbf{X}} = \mathbf{S}_n/n$, let $\boldsymbol{\Theta}^* = \{\boldsymbol{\tau} : \kappa(\boldsymbol{\tau}) < \infty\}$, and let $\mu_{\boldsymbol{\tau}}(d\mathbf{y}) = \exp(-n(\kappa(\boldsymbol{\tau}) - \boldsymbol{\tau}^\top \mathbf{y}))\mu(d\mathbf{y})$ define the distribution of $\bar{\mathbf{X}}_{\boldsymbol{\tau}}$, the mean of $\mathbf{X}_{1\boldsymbol{\tau}}, \dots, \mathbf{X}_{n\boldsymbol{\tau}}$, the associated independent random vectors. These conjugate distributions, first introduced in [5], permit us to consider large deviations. Let $V_{\boldsymbol{\tau}} = \kappa''(\boldsymbol{\tau})$, and, taking $\|\mathbf{x}\| = (\mathbf{x}^\top \mathbf{x})^{1/2}$, let

$$\eta_j(\boldsymbol{\tau}) = n^{-1} \sum_{i=1}^n E\big[\| V_{\boldsymbol{\tau}}^{-1/2} (\mathbf{X}_{i\boldsymbol{\tau}} - E[\mathbf{X}_{i\boldsymbol{\tau}}]) \|^j\big].$$

Let

$$\begin{aligned} q_{\boldsymbol{\tau}}(T) = \sup\big\{ \big| e^{\kappa(\boldsymbol{\tau} + i\boldsymbol{\xi}) - \kappa(\boldsymbol{\tau})} \big| : \| V_{\boldsymbol{\tau}}^{1/2} \boldsymbol{\xi} \| > (3/4) \eta_3(\boldsymbol{\tau})^{-1}, \\ |\xi_i| < \pi \text{ for } i \le d_0, |\xi_i| < T, i > d_0 \big\}. \end{aligned} \tag{5}$$

We consider the following conditions, essentially from [11], where, throughout, c and C are generic positive constants, and $|A|$ denotes the determinant of a square matrix A. The complexity of these conditions is due to the fact that we need to consider conditional distributions of independent, but not identically distributed, random variables.

- (A1) There is a compact subset, $\boldsymbol{\Theta}$, of the interior of $\boldsymbol{\Theta}^*$, with $\mathbf{0}$ in the interior of $\boldsymbol{\Theta}$.
- (A2) $|V_{\boldsymbol{\tau}}| > c > 0$ for $\boldsymbol{\tau} \in \boldsymbol{\Theta}$.
- (A3) $\eta_j(\boldsymbol{\tau}) < C$ for $j = 1, \dots, 5$ and $\boldsymbol{\tau} \in \boldsymbol{\Theta}$.
- (A4) $n^{2d_1+2} q_{\boldsymbol{\tau}}(n^{-2}) < C$.

Here the first condition asserts that there is an open neighborhood of the origin where the cumulative generating function exists. The second condition bounds the average variance of the associated random variables away from zero, and the third gives upper bounds the first 5 standardized moments in this neighborhood. The fourth condition is a smoothness condition introduced first for the univariate case in [1] and which is sufficient to allow Edgeworth expansions for many statistics based on ranks and applications to bootstrap and permutation statistics when the original observations are from a continuous distribution.

Let $\mathcal{X} = \kappa'(\boldsymbol{\Theta})$; then we are able to obtain equations (2), (3) and (4) for $\mathbf{x} \in \mathcal{X}$. Also, if $d_0 > 0$, let $\Lambda_0(\mathbf{x}_0) = \hat{\boldsymbol{\tau}}_0^\top \mathbf{x}_0 - \kappa_0(\hat{\boldsymbol{\tau}}_0)$ for $\hat{\boldsymbol{\tau}}_0$ satisfying $\kappa_0'(\hat{\boldsymbol{\tau}}_0) = \mathbf{x}_0$, where the subscript 0 denotes a reduction to the first d_0 elements of the d-vectors, and we will use the subscript 1 to denote the last d_1 elements. If $r_0(\mathbf{x}_0) = (2\pi n)^{-d_0/2} |V_{\hat{\boldsymbol{\tau}}_0}|^{-1/2} \exp(-n\Lambda_0(\mathbf{x}_0))$, then from [4], we have $\mathrm{P}(\bar{\mathbf{X}}_0 = \mathbf{x}_0) = \mu_0(\mathbf{x}_0) = r_0(\mathbf{x}_0)(1 + O(1/n))$. For $d_0 > 0$, we will consider $\mathbf{x}^\top = (\mathbf{x}_0^\top, \mathbf{x}_1^\top)$ and replace $r(\mathbf{x})$ by

$$r(\mathbf{x}_1|\mathbf{x}_0) = r(\mathbf{x})/r_0(\mathbf{x}_0) = \frac{|V_{\hat{\boldsymbol{\tau}}_0}|^{1/2} e^{-n(\Lambda(\mathbf{x}) - \Lambda_0(\mathbf{x}_0))}}{(2\pi/n)^{d_1/2} |V_{\hat{\boldsymbol{\tau}}}|^{1/2}}, \tag{6}$$

and replace μ by the distribution of $\bar{\mathbf{X}}_1$ conditional on $\bar{\mathbf{X}}_0 = \bar{\mathbf{x}}_0$, so that we consider conditional probabilities of $\bar{\mathbf{X}}_1$ given $\bar{\mathbf{X}}_0 = \mathbf{x}_0$, associated with sets of form

$$\mathcal{F} = \{\mathbf{x}_1 : \Lambda(\mathbf{x}) - \Lambda_0(\mathbf{x}_0) \geq \lambda, \mathbf{x}^\top = (\mathbf{x}_0^\top, \mathbf{x}_1^\top)\}.$$

The main result is the following theorem, whose proof is deferred to a later section.

THEOREM 1. *Under conditions* (A1)–(A4),

$$\left|\mu(\mathcal{F}) - \int_{\mathcal{F}} r(\mathbf{x}_1|\mathbf{x}_0)\, d\mathbf{x}_1\right| = \left[\int_{\mathcal{F}} r(\mathbf{x}_1|\mathbf{x}_0)\, d\mathbf{x}_1\right] O(1/n). \tag{7}$$

Note that if the nonlattice subvectors of $\mathbf{X}_1, \dots, \mathbf{X}_n$ have densities, the variables are identically distributed and (A1) and (A2) hold, then the theorem follows from Theorem 1 of [4].

The following theorem is a corollary whose derivation we include here. This is the form that will be used in examples.

THEOREM 2. *Under the conditions of Theorem 1, if* $u = \sqrt{2\lambda}$,

$$\int_{\mathcal{F}} r(\mathbf{x}_1|\mathbf{x}_0)\,d\mathbf{x}_1 = \bar{Q}_{d_1}(nu^2)[1 + O(1/n)] + \frac{c_n}{n} u^{d_1} e^{-nu^2/2} \frac{G(u) - 1}{u^2} \tag{8}$$

and

$$\int_{\mathcal{F}} r(\mathbf{x}_1|\mathbf{x}_0)\,d\mathbf{x}_1 = \bar{Q}_{d_1}(nu^{*2})[1 + O(1/n)], \tag{9}$$

where $\bar{Q}_d(x) = \mathrm{P}(\chi^2_d \geq x)$,

$$u^* = u - \log(G(u))/nu, \tag{10}$$

$$c_n = \frac{n^{d_1/2}}{2^{d_1/2-1}\Gamma(d_1/2)}, \tag{11}$$

$$\delta(\sqrt{2\lambda}, s) = \frac{\Gamma(d_1/2)|V_{\hat{\boldsymbol{\tau}}_0}|^{1/2}|V_{\hat{\boldsymbol{\tau}}}|^{-1/2}|V_0|^{1/2} r^{d_1-1}}{2\pi^{d_1/2} u^{d_1-2} |\mathbf{s}^\top V_0^{1/2} \hat{\boldsymbol{\tau}}_1|}, \tag{12}$$

$$G(u) = \int_{S_{d_1}} \delta(u, s)\,ds \tag{13}$$

for S_{d_1} *the* d_1*-dimensional unit sphere centered at zero, and where, for each* $\mathbf{s} \in S_{d_1}$, r *is chosen so* $\Lambda(\mathbf{x}_0, r\mathbf{s}) - \Lambda_0(\mathbf{x}_0) = \lambda$ *and* $V_0^{-1} = [\kappa''(\mathbf{0})^{-1}]_{11}$, *with the final subscripts denoting the lower right* $d_1 \times d_1$ *submatrix.*

PROOF. The derivation of (8), given Theorem 1, is given in [12]. To get (9), we use a related method. After making the transformations $\mathbf{y} = V_{\mathbf{0}}^{-1/2}\mathbf{x}_1$, $\mathbf{y} \to (r, \mathbf{s})$ and $(r, \mathbf{s}) \to (u, \mathbf{s})$, where the first is the polar transformation with $\|\mathbf{x}\| = r$ and $\mathbf{s} \in S_d$, the unit sphere in d-dimensions, and the second has $u = \sqrt{2(\Lambda(\mathbf{x}_0, r\mathbf{s}) - \Lambda_0(\mathbf{x}_0))}$, we have

$$\begin{aligned}\int_{\mathcal{F}} r(\mathbf{x}_1|\mathbf{x}_0)\,d\mathbf{x}_1 &= c_n \int_u^\infty v^{d-1} e^{-nv^2/2} G(v)\,dv \\ &= c_n \int_u^\infty v^{d-1} e^{-n(v - \log G(v)/nv)^2/2}\,dv\big(1 + O(1/n)\big).\end{aligned}$$

Then make the transformation $v^* = v - \log G(v)/nv$. The final equality follows since $G(v) = 1 + v^2 k(v)$ and $G'(v) = vk^*(v)$, where $k(v)$ and $k^*(v)$ are bounded as shown in [12]. □

REMARK. The integral (13) can be approximated by a Monte Carlo method, for example, by approximating $\int_{S_d} h(s)\,ds$ as

$$\frac{2\pi^{d/2}}{\Gamma(d/2)}\frac{1}{M}\sum_{\ell=1}^{M} h(U_\ell),$$

where $U_1, \dots, U_M$ are i.i.d. uniformly distributed on S_d. Here the number of replicates in the Monte Carlo simulation can be small with little loss of accuracy. We discuss this in the examples where it was found that $M = 10$ was sufficient. It would be possible to use a method such as that in [7] to get a numerical approximation to the integral, but the Monte Carlo method is much simpler to use and easily gives the required accuracy.

3. Two examples of permutation tests. We consider a k sample permutation test in a one-way design and a multivariate two-sample permutation test. In both cases we consider hypotheses that the populations of random variables or vectors are exchangeable. In the first case the observations are generated either by sampling $n_1, \dots, n_k$ independent random variables from distributions $F_1, \dots, F_k$, and we test $H_0 : F_1 = \cdots = F_k$, or they are generated from an experiment in which k treatments are allocated at random to groups of sizes $n_1, \dots, n_k$, and we test H_0: treatments have equal effects. We choose a statistic suitable for testing with respect to differences in means. The standard choices of test statistic are the F-statistic from the analysis of variance or, for a nonparametric test based on ranks, the Kruskal–Wallis statistic. In the second case the observations are generated by sampling from two populations of l-dimensional random vectors, and we test for equality of the distributions, or they are generated by experimental randomization, and we test for equality of two treatments. Here the test statistic arising from an assumption of multivariate normality is the Mahalanobis D^2 test.

3.1. *Permutation tests for k samples.* Suppose that $a_1, \dots, a_N$ are the elements of a finite population, such that $\sum_{m=1}^N a_m = 0$ and $\sum_{m=1}^N a_j^2 = N$. Let $n_1, \dots, n_k$ be integers, such that $N = \sum_{i=1}^k n_i$. Suppose that $R_1, \dots, R_N$ is an equiprobable random permutation of $1, \dots, N$. Let $X_{ij} = a_{R_{n_1+\cdots+n_{i-1}+j}}$, and let $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$.

For $i = 1, \dots, k-1$, let $\mathbf{e}_i$ have $k-1$ components, with component i equal to 1, and other components zero. Let $\mathbf{I}_m, m = 1, \dots, N$ be independent and identically distributed random vectors with $\mathrm{P}(\mathbf{I}_m = \mathbf{e}_i) = n_i/N = p_i$ for $i < k$ and $\mathrm{P}(\mathbf{I}_m = \mathbf{0}) = n_k/N = p_k$. Let $\mathbf{S}^\top = (\sum_{m=1}^N \mathbf{I}_m^\top, \sum_{m=1}^N a_m \mathbf{I}_m^\top) = (\mathbf{S}_0^\top, \mathbf{S}_1^\top)$. We have

$$\mathrm{P}(n\bar{\mathbf{X}} \le \mathbf{x}) = \mathrm{P}(\mathbf{S}_1 \le \mathbf{x} | \mathbf{S}_0 = N\mathbf{p}),$$

where $\bar{\mathbf{X}}, \mathbf{x}, \mathbf{p}$ are $k-1$ vectors corresponding to the first $k-1$ samples. Under H_0, the cumulant generating function of $\mathbf{S}$ is

$$N\kappa(\boldsymbol{\tau}_0, \boldsymbol{\tau}_1) = \log E e^{\sum_{m=1}^{N}(\boldsymbol{\tau}_0^\top \mathbf{I}_m + \boldsymbol{\tau}_1^\top \mathbf{I}_m a_m)} = \sum_{m=1}^{N} \log\left(p_k + \sum_{i=1}^{k-1} p_i e^{\tau_{0i}+\tau_{1i}a_m}\right).$$

Let $(\hat{\boldsymbol{\tau}}_0^\top, \hat{\boldsymbol{\tau}}_1^\top)$ be the solution of

$$\kappa'(\boldsymbol{\tau}_0, \boldsymbol{\tau}_1) = (\mathbf{p}, \mathbf{x}).$$

Let $B = \{\mathbf{x} : \Lambda(\mathbf{x}) \geq u^2/2\}$, where $\Lambda(\mathbf{x}) = \hat{\boldsymbol{\tau}}_0^\top \mathbf{p} + \hat{\boldsymbol{\tau}}_1^\top \mathbf{x} - \kappa(\hat{\boldsymbol{\tau}}_0, \hat{\boldsymbol{\tau}}_1)$, and note that $\kappa'(0, \mathbf{0}) = (p, \mathbf{0})$ and $\kappa(0, \mathbf{0}) = 0$. Now from Theorem 1, if $q_{\boldsymbol{\tau}}(n^{-2}) = O(n^{-2k})$,

$$\mathrm{P}\big(\Lambda(\bar{\mathbf{X}}) \geq u^2/2\big) = \int_B r(\mathbf{x}|\mathbf{p})\, d\mathbf{x}\big(1 + O(1/N)\big),$$

where

$$r(\mathbf{x}|\mathbf{p}) = (2\pi/N)^{-(k-1)/2} |\kappa_{00}(0, \mathbf{0})|^{1/2} |\kappa''(\hat{\boldsymbol{\tau}}_0, \hat{\boldsymbol{\tau}}_1)|^{-1/2} e^{-N\Lambda(\mathbf{x})}.$$

Then from Theorem 2, $G(u)$ is given in (12) and (13) with $d_0 = d_1 = k-1$. Now we can use (8) and (9) to get the two approximations.

3.2. *Numerical results for k-sample test.* Consider first the rank test based on the statistic $\Lambda(\bar{\mathbf{X}})$ where $a_1 = 1, \ldots, a_N = N$, with $N = 20$ for 4 groups of size 5; in the standard case the Kruskal–Wallis test would be used. The following table gives the results of tail probabilities from a Monte Carlo simulation of $\Lambda(\bar{\mathbf{X}})$ (MC Λ) and of the Kruskal–Wallis statistic (MC K–W) using 100,000 permutations, the chi-squared approximation (χ_3^2) and the saddlepoint approximations using (8) (SP LR Λ) and (9) (SP BN Λ), using $M = 1000$ Monte Carlo samples from S_3. Inspection of the table comparing the saddlepoint Lugananni–Rice and Barndorff-Nielsen approximations with the Monte Carlo approximation for Λ shows the considerable accuracy of these approximations throughout the range. The chi square approximations to the distribution of the Kruskal–Wallis statistic does not have this degree of accuracy. We note that good approximations for the saddlepoint approximations are achieved by M as small as 10. We obtained the standard deviation of individual random values of the integrand and noted that for Table 1 this was 0.003 for $\hat{u} = 0.6$ and 0.0007 for $\hat{u} = 0.9$, indicating that $M = 10$ gives sufficient accuracy in this example.

Also consider the permutation test based on a single sample of 40 in 4 groups of 10 from an exponential distribution, comparing as above each of the saddlepoint approximations with the Monte Carlo approximations in this case and with the standard test based on the sum of squares from an analysis of variance. The same pattern of accuracy as reported above is apparent from inspection of Table 2.

TABLE 1
The 4*-sample rank tests with* $n_i = 5$

$\hat{u}$	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
MC Λ	0.6758	0.4328	0.2365	0.1087	0.0423	0.0142	0.0041
MC K–W	0.6583	0.4027	0.1921	0.0652	0.0135	0.0012	0.0000
χ_3^2	0.6149	0.3618	0.1718	0.0658	0.0203	0.0051	0.0010
SP LR Λ	0.6811	0.4446	0.2454	0.1151	0.0464	0.0164	0.0052
SP BN Λ	0.6753	0.4380	0.2387	0.1101	0.0434	0.0148	0.0045

3.3. *A two-sample multivariate permutation test.* Let $\mathbf{a}_1, \dots, \mathbf{a}_N$ be l-vectors regarded as elements of a finite population such that $\sum_{i=1}^N \mathbf{a}_i = 0$ and $\sum_{i=1}^N \mathbf{a}_i \mathbf{a}_i^T = NI$. Let $R_1, \dots, R_N$ be obtained by an equiprobable random permutation of $1, \dots, N$, let $\mathbf{X}_j = \mathbf{a}_{R_j}$, $j = 1, \dots, N$ and let $\bar{\mathbf{X}}_1 = \sum_{j=1}^n \mathbf{X}_j / n$ for $n = Np$ with $0 < p < 1$. Let $I_1, \dots, I_N$ be i.i.d. Bernoulli variables with $EI_1 = p$. If $\mathbf{S}^T = (S_0, \mathbf{S}_1^T)$ with $S_0 = \sum_{i=1}^N I_i$ and $\mathbf{S}_1 = \sum_{i=1}^N \mathbf{a}_i I_i$, then for any Borel set $\mathcal{F}$,

$$P(\bar{\mathbf{X}} \in \mathcal{F}) = P(\mathbf{S}_1/N \in \mathcal{F} | S_0/N = p). \tag{14}$$

Let $\boldsymbol{\tau}^\top = (\tau_0, \boldsymbol{\tau}_1^\top)$ with $\tau_0 \in \Re$ and $\boldsymbol{\tau}_1 \in \Re^d$ and let

$$\begin{aligned} \kappa(\boldsymbol{\tau}) &= N^{-1} \log E \exp(\tau_0 S_0 + \boldsymbol{\tau}_1^\top \mathbf{S}_1) \\ &= N^{-1} \sum_{i=1}^N \log(q + p e^{\tau_0 + \boldsymbol{\tau}_1^\top \mathbf{a}_i}). \end{aligned}$$

Let $\hat{\boldsymbol{\tau}}$ be the solution of $\kappa'(\boldsymbol{\tau}) = (p, \mathbf{x}^\top)^\top$, and let $\Lambda(p, \mathbf{x}) = \hat{\tau}_0 p + \hat{\boldsymbol{\tau}}_1^\top \mathbf{x} - \kappa(\hat{\boldsymbol{\tau}})$. Consider sets $\mathcal{F} = \{\mathbf{x} : \Lambda(p, \mathbf{x}) \geq \lambda\}$. Then from Theorem 2, we can approximate (14) by (8) or (9) where $G(u)$ is given by (12) and (13) with $d_0 = 1$ and $d_1 = l$.

3.4. *Numerical results for two-sample test.* Consider the test based on two samples of size 40 from a 3-variate exponential distribution with mean 1 and covariance matrix I. After standardizing the combined sample we consider tests

TABLE 2
The 4*-sample permutation tests with exponentially distributed errors and* $n_i = 10$

$\hat{u}$	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
MC Λ	0.9456	0.6837	0.3434	0.1160	0.0275	0.0043	0.0004
MC ANOV	0.9455	0.6784	0.3273	0.0971	0.0164	0.0015	0.0004
χ_3^2	0.9402	0.6594	0.3080	0.0937	0.0186	0.0024	0.0002
SP LR Λ	0.9491	0.6888	0.3456	0.1174	0.0272	0.0043	0.0004
SP BN Λ	0.9486	0.6877	0.3441	0.1164	0.0268	0.0042	0.0004

TABLE 3
The 3-dimensional two sample parmutation test

$\hat{u}$	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**
MC Λ	0.3543	0.1249	0.0276	0.0041	0.0006
χ_3^2	0.3080	0.0937	0.0186	0.0024	0.0002
SP LR Λ	0.3528	0.1207	0.0282	0.0045	0.0005
SP BN Λ	0.3507	0.1194	0.0278	0.0043	0.0005
Quadratic	0.3325	0.0939	0.0135	0.0004	0.0001

based on the statistic $\Lambda(\bar{\mathbf{X}})$ or $\bar{\mathbf{X}}\bar{\mathbf{X}}^T$, equivalent to the usual normal theory based statistic. We calculate the tail probabilities based on Theorem 2 in this case and Monte Carlo approximations to the permutation tests based on 10,000 random permutations. Table 3 demonstrates the accuracy of the two saddlepoint approximations throughout the range. It also shows that the chi-squared approximation is not satisfactory either for Λ or for the classical quadratic form statistic. However, while we have accurate tail probability approximations for the new statistic, such approximations are not available for the classical quadratic form.

4. Proofs of the main results. For notational convenience we will restrict attention to the case $d_0 = 0$, as details of the case conditional on lattice variables follow in a straightforward manner. The following theorem is a simplified version of Theorem 1 of [11], taking $s = 5$, $d_0 = 0$ and $\mathcal{A}$ as a d-dimensional cube in $\mathcal{X}$ with center $\mathbf{a}$ and side $\delta = n^{-1}$. As in (1.10) of [11], let

$$e_2(\mathbf{y}, \boldsymbol{\mu}_\tau) = \big(1 + Q_1(\mathbf{y}^*) + Q_2(\mathbf{y}^*)\big)(2\pi/n)^{-d/2}|V_{\boldsymbol{\tau}}|^{-1/2}e^{-\mathbf{y}^{*\top}\mathbf{y}^*/2}$$

with $\mathbf{y}^* = n^{1/2}V_{\boldsymbol{\tau}}^{-1/2}(\mathbf{y} - \kappa'(\boldsymbol{\tau}))$, be the formal Edgeworth expansion of order 2 for $\bar{\mathbf{X}}_{\boldsymbol{\tau}} = \sum_{i=1}^n \mathbf{X}_{i\boldsymbol{\tau}}/n$, and let

$$e_2\big(\boldsymbol{\tau}, \mathcal{E}, \mathbf{x} - \kappa(\boldsymbol{\tau})\big) = \int_{\mathcal{E}} e^{n\boldsymbol{\tau}^\top(\mathbf{x}-\mathbf{y})}e_2(\mathbf{y}, \boldsymbol{\mu}_\tau)\,d\mathbf{y}.$$

The terms Q_1 and Q_2 are given explicitly in (1.11) of [11], and are terms of order $n^{-1/2}$ and n^{-1}, respectively.

THEOREM 3. *For any set $\mathcal{E} \subset \mathcal{A}$ and $\varepsilon > 0$, take $\mathcal{E}_\varepsilon = \{\mathbf{z} : \exists \mathbf{y} \in \mathcal{E}, \|\mathbf{z} - \mathbf{y}\| < \varepsilon\}$. Choose $\varepsilon \in (0, c/n^2)$, and let $T = 1/\varepsilon$. For $\mathbf{x} \in \mathcal{E} \subset \mathcal{X}$,*

$$\big|\mu(\mathcal{E}) - e^{-n(\boldsymbol{\tau}^\top\mathbf{x}-\kappa(\boldsymbol{\tau}))}e_2\big(\boldsymbol{\tau}, \mathcal{E}, \mathbf{x} - m(\boldsymbol{\tau})\big)\big| \le e^{-n(\boldsymbol{\tau}^\top\mathbf{x}-\kappa(\boldsymbol{\tau}))}|V_{\boldsymbol{\tau}}|^{-1/2}R$$

for

$$R = C\big[\mathrm{Vol}(\mathcal{E}_{2\varepsilon})\big(\eta_5(\boldsymbol{\tau})n^{-3/2} + |V_{\boldsymbol{\tau}}|^{1/2}n^{1/2}T^d q_{\boldsymbol{\tau}}(T)\big) + \mathrm{Vol}(\mathcal{E}_{2\varepsilon} - \mathcal{E}_{-2\varepsilon})\big].$$

Note that this follows, since

$$\hat{\chi}_{\boldsymbol{\tau},\mathcal{E}_{2\varepsilon}}(\mathbf{0}) = \int_{\mathcal{E}_{2\varepsilon}} e^{n\boldsymbol{\tau}^\top(\mathbf{u}-\mathbf{a})}\,du < C\,\mathrm{Vol}(\mathcal{E}_{2\varepsilon})$$

as $\mathcal{E} \subset \mathcal{A}$ implies that $\|\mathbf{u}-\mathbf{a}\| < c(\delta + 2\varepsilon) < cn^{-1}$.

We give a preliminary lemma before proceeding to the proof of Theorem 1, using the notation $\kappa(\boldsymbol{\tau}(\mathbf{x})) = \mathbf{x}$ and $\Lambda(\mathbf{x}) = \boldsymbol{\tau}(\mathbf{x})^\top \mathbf{x} - \kappa(\boldsymbol{\tau}(\mathbf{x}))$ for $\mathbf{x} \in \mathcal{X}$.

LEMMA 1. *For* $\mathbf{x} \in \mathcal{E} \subset \mathcal{A} \subset \mathcal{X}$,

$$\int_{\mathcal{E}} r(\mathbf{y})\,d\mathbf{y} - e^{n(\kappa(\boldsymbol{\tau}(\mathbf{x}))-\boldsymbol{\tau}(\mathbf{x})^\top\mathbf{x})} e_2(\boldsymbol{\tau}, \mathcal{E}, 0) = \int_{\mathcal{E}} r(\mathbf{y})\,d\mathbf{y}\,O(1/n). \tag{15}$$

PROOF. Ignoring for the moment the terms involving Q_1 and Q_2, the left-hand side in (15) is

$$\int_{\mathcal{E}} r(\mathbf{y})\left[1 - \frac{e^{n(\Lambda(\mathbf{y})-\Lambda(\mathbf{x})-\boldsymbol{\tau}(\mathbf{x})^\top(\mathbf{y}-\mathbf{x})-(\mathbf{y}-\mathbf{x})^\top V^{-1}_{\boldsymbol{\tau}(\mathbf{x})}(\mathbf{y}-\mathbf{x})/2)}}{|V_{\boldsymbol{\tau}(\mathbf{x})}|^{1/2}/|V_{\boldsymbol{\tau}(\mathbf{y})}|^{1/2}}\right] d\mathbf{y}. \tag{16}$$

Noting that $\|\mathbf{y}-\mathbf{x}\| = O(1/n)$, and using a Taylor series expansion about $\mathbf{x}$, we see that the exponent in (16) is $O(1/n^2)$, and the denominator is $1 + O(1/n)$. So in (15) the first term on the left is as given by the expression on the right. Noting that $Q_1(\mathbf{0}) = 0$, we see that the term involving Q_1 is of the same form. The proof is completed by noting that the term Q_2 is also of this form. □

PROOF OF THEOREM 1. The proof will proceed by dividing $\mathcal{X}$ into small rectangles, applying Theorem 3 on each of these rectangles, and summing the results in a manner similar to that of [9]. For $\mathbf{j} \in \mathbb{Z}_1^d$, let $\mathcal{A}_{\mathbf{j}} = \{\mathbf{x} \in \mathbb{R}^d : x_l \in ((j_{l-d_0} - \frac{1}{2})\delta, (j_{l-d_0} + \frac{1}{2})\delta]\}$, and let $\mathcal{E}^{\mathbf{j}} = \mathcal{A}_{\mathbf{j}} \cap \mathcal{F}$. By the intermediate value theorem, on each $\mathcal{E}^{\mathbf{j}}$, there is an $\mathbf{x}_{\mathbf{j}}$ such that

$$\int_{\mathcal{E}^{\mathbf{j}}} r(\mathbf{x})\,d\mathbf{x} = r(\mathbf{x}(\mathbf{j}))\,\mathrm{Vol}(\mathcal{E}^{\mathbf{j}}).$$

Note that $\mathcal{F} = \bigcup_{\mathbf{j}\in\mathbb{J}} \mathcal{E}^{\mathbf{j}}$ and $\mathcal{E}^{\mathbf{j}}$ are disjoint. Define $\hat{\boldsymbol{\tau}}_{\mathbf{j}}$ so that $\mathbf{x}_{\mathbf{j}} = \kappa'(\hat{\boldsymbol{\tau}}_{\mathbf{j}})$. Write $\mathbb{J} = \{j : \mathrm{Vol}(\mathcal{E}^{\mathbf{j}}) > 0\}$. Then

$$\begin{aligned}\mu(\mathcal{F}) - \int_{\mathcal{F}} r(\mathbf{x})\,d\mathbf{x} &= \sum_{\mathbf{j}\in\mathbb{J}} [\mu(\mathcal{E}^{\mathbf{j}}) - r(\mathbf{x}_{\mathbf{j}})\,\mathrm{Vol}(\mathcal{E}^{\mathbf{j}})]\\ &= E_1 + E_2,\end{aligned}$$

where

$$E_1 = \sum_{\mathbf{j}\in\mathbb{J}} [\mu(\mathcal{E}^{\mathbf{j}}) - r(\mathbf{x}_{\mathbf{j}})(2\pi/n)^{d/2}|V_{\hat{\boldsymbol{\tau}}_{\mathbf{j}}}|^{1/2} e_2(\hat{\boldsymbol{\tau}}_{\mathbf{j}}, \mathcal{E}^{\mathbf{j}}, 0)]$$

and

$$E_2 = -\sum_{\mathbf{j}\in\mathbb{J}}\left[\int_{\mathcal{E}^{\mathbf{j}}} r(\mathbf{y})\,d\mathbf{y} - e^{n\kappa(\hat{\boldsymbol{\tau}}_{\mathbf{j}})-\hat{\boldsymbol{\tau}}_{\mathbf{j}}^{\top}\mathbf{x}_{\mathbf{j}}} e_2(\hat{\boldsymbol{\tau}}_{\mathbf{j}}, \mathcal{E}^{\mathbf{j}}, 0)\right].$$

Using Lemma 1 on each $\mathcal{E}^{\mathbf{j}}$ and summing, we have

$$E_2 = \int_{\mathcal{F}} r(\mathbf{y})\,d\mathbf{y}\,O(1/n).$$

Now consider E_1. Apply Theorem 3 to each $\mathcal{E}^{\mathbf{j}}$, and sum to get

$$|E_1| \le \sum_{\mathbf{j}\in\mathbb{J}} r(\mathbf{x_j})(R_{1\mathbf{j}} + R_{2\mathbf{j}}), \tag{17}$$

where

$$R_{1j} = C\,\mathrm{Vol}(\mathcal{E}^{\mathbf{j}}_{2\varepsilon})[\eta_5(\boldsymbol{\tau}_{\mathbf{j}})n^{-1} + |V_{\hat{\boldsymbol{\tau}}_{\mathbf{j}}}|^{1/2} n^{d/2} T q_{\hat{\boldsymbol{\tau}}_{\mathbf{j}}}(T)]$$

and

$$R_{2\mathbf{j}} = \mathrm{Vol}(\mathcal{E}^{\mathbf{j}}_{2\varepsilon} - \mathcal{E}^{\mathbf{j}}_{-2\varepsilon}).$$

The summation of these terms is complicated by the fact that the sets are not disjoint and not all are subsets of $\mathcal{F}$. So introduce sets $\mathcal{H}_{\mathbf{j}} = \mathcal{A}_{\mathbf{j}} \cap \mathcal{F}_{2\varepsilon}$. Consider the set $\mathcal{H}^*_{\mathbf{j}}$, the union of $\mathcal{H}_{\mathbf{j}}$ and the $3^d - 1$ sets formed by reflections of $\mathcal{H}_{\mathbf{j}}$ in each of the lower-dimensional faces of $\mathcal{A}_{\mathbf{j}}$. Then $\mathcal{E}_{\mathbf{j}} \subset \mathcal{H}^*_{\mathbf{j}}$ so

$$\mathrm{Vol}(\mathcal{E}_{\mathbf{j}2\varepsilon})/\,\mathrm{Vol}(\mathcal{H}_{\mathbf{j}}) \le 3^d.$$

So

$$\sum_{\mathbf{j}\in\mathbb{J}} r(\mathbf{x_j})R_{1\mathbf{j}} \le \sum_{\mathbf{j}\in\mathbb{J}} r(\mathbf{x_j})\,\mathrm{Vol}(\mathcal{H}_{\mathbf{j}})O(1/n) = \int_{\mathcal{F}_{2\varepsilon}} r(\mathbf{y})\,d\mathbf{y}\,O(1/n).$$

Also

$$\mathrm{Vol}(\mathcal{E}^{\mathbf{j}}_{2\varepsilon} - \mathcal{E}^{\mathbf{j}}_{-2\varepsilon})/\,\mathrm{Vol}(\mathcal{H}_{\mathbf{j}}) \le C\varepsilon/\delta = O(1/n).$$

Using this to bound the second sum on the right-hand side of (17) and the previous bound for the first term gives

$$|E_1| = \int_{\mathcal{F}_{2\varepsilon}} r(\mathbf{y})\,d\mathbf{y}\,O(1/n).$$

Note that for any $\mathbf{x}$ such that $\Lambda(\mathbf{x}) = \lambda$ and any $\mathbf{z} \in \mathcal{F}_{2\varepsilon}$,

$$\Lambda(\mathbf{z}) \ge \Lambda(\mathbf{x}) - |(\mathbf{z}-\mathbf{x})^{\top}\Lambda'(\mathbf{x})| \ge \lambda - C\varepsilon.$$

So the theorem follows by noting that

$$\int_{\mathcal{F}_{2\varepsilon}} r(\mathbf{y})\,d\mathbf{y} = \int_{\mathcal{F}} r(\mathbf{y})\,d\mathbf{y}\big(1 + O(1/n)\big). \qquad \square$$

REFERENCES

[1] ALBERS, W., BICKEL, P. J. and VAN ZWET, W. R. (1976). Asymptotic expansions for the power of distribution free tests in the one-sample problem. *Ann. Statist.* **4** 108–156. MR0391373

[2] BARNDORFF-NIELSEN, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73** 307–322. MR0855891

[3] BARNDORFF-NIELSEN, O. E. and COX, D. R. (1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J. Roy. Statist. Soc. Ser. B* **46** 483–495. MR0790634

[4] BOROVKOV, A. A. and ROGOZIN, B. A. (1965). On the multi-dimensional central limit theorem. *Theory Probab. Appl.* **10** 55–62.

[5] CRAMÉR, H. (1938). Sur un noveau théorème–limite de la théorie des probabilitiés. *Actualités Schi. Indust* **736** 5–23.

[6] DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application. Cambridge Series in Statistical and Probabilistic Mathematics* **1**. Cambridge Univ. Press, Cambridge. MR1478673

[7] GENZ, A. (2003). Fully symmetric interpolatory rules for multiple integrals over hyper-spherical surfaces. *J. Comput. Appl. Math.* **157** 187–195. MR1996475

[8] JING, B. Y., FEUERVERGER, A. and ROBINSON, J. (1994). On the bootstrap saddlepoint approximations. *Biometrika* **81** 211–215. MR1279668

[9] OSIPOV, L. V. (1981). On large deviations for sums of random vectors in R^k. *J. Multivariate Anal.* **11** 115–126. MR0618780

[10] ROBINSON, J. (2004). Multivariate tests based on empirical saddlepoint approximations. *Metron* **62** 1–14. MR2089164

[11] ROBINSON, J., HÖGLUND, T., HOLST, L. and QUINE, M. P. (1990). On approximating probabilities for small and large deviations in R^d. *Ann. Probab.* **18** 727–753. MR1055431

[12] ROBINSON, J., RONCHETTI, E. and YOUNG, G. A. (2003). Saddlepoint approximations and tests based on multivariate M-estimates. *Ann. Statist.* **31** 1154–1169. MR2001646

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
110 FRELINGHUYSEN RD
PISCATAWAY, NEW JERSEY 08854-8019
USA
E-MAIL: kolassa@stat.rutgers.edu

SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF SYDNEY
NSW 2006
AUSTRALIA
E-MAIL: john.robinson@sydney.edu.au