The Annals of Statistics
2001, Vol. 29, No. 2, 328–371

TAIL PROBABILITIES OF THE MAXIMA OF MULTILINEAR FORMS AND THEIR APPLICATIONS

BY SATOSHI KURIKI AND AKIMICHI TAKEMURA

Institute of Statistical Mathematics and University of Tokyo

Let Z be a k-way array consisting of independent standard normal variables. For column vectors $h_1, \dots, h_k$, define a multilinear form of degree k by $(h_1 \otimes \cdots \otimes h_k)' \mathrm{vec}(Z)$. We derive formulas for upper tail probabilities of the maximum of a multilinear form with respect to the h_i's under the condition that the h_i's are unit vectors, and of its standardized statistic obtained by dividing by the norm of Z. We also give formulas for the maximum of a symmetric multilinear form $(h_1 \otimes \cdots \otimes h_k)' \mathrm{vec}(\mathrm{sym}(Z))$, where $\mathrm{sym}(Z)$ denotes the symmetrization of Z with respect to indices. These classes of statistics are used for testing hypotheses in the analysis of variance of multiway layout data and for testing multivariate normality. In order to derive the tail probabilities we employ a geometric approach developed by Hotelling, Weyl and Sun. Upper and lower bounds for the tail probabilities are given by reexamining Sun's results. Some numerical examples are given to illustrate the practical usefulness of the obtained formulas, including the upper and lower bounds.

1. Introduction.

1.1. *Multilinear forms in statistical inference.* In this paper we study distribution theory of multilinear forms. Here we briefly mention various uses of multilinear forms in statistics and the relevance of our results to these applications. More detailed discussions of these applications are given in Section 2.

In the traditional ANOVA setting, multilinear forms are used to model higher order interactions. For the $I \times J$ two-way layout, Johnson and Graybill (1972) proposed to model the interaction by a bilinear form: $\phi u_i v_j$, $i = 1, \dots, I$, $j = 1, \dots, J$, where ϕ is a scalar. Their method was extended to a multiway layout by Boik and Marasinghe (1989). In a three-way layout, for example, they proposed the trilinear structure $\phi u_i v_j w_k$ as a model for interaction of the highest degree. Modeling of higher order interaction by this form is attractive because of its simplicity. For statistical inference it is important to test the null hypothesis $H_0\colon \phi = 0$. The distribution theory for testing this hypothesis will be provided by our results on multilinear forms. More general models of this type for multiway data are applied mainly in the field of psychometrics and chemometrics. However, appropriate distribution theory is lacking and at present these models are used for descriptive purposes only.

Received June 1998; revised April 2000.

AMS 2000 *subject classifications*. Primary 62H10, 62H15; secondary 53C65.

Key words and phrases. Gaussian field, Karhunen–Loève expansion, largest eigenvalue, multiple comparisons, multivariate normality, multiway layout, PARAFAC, projection pursuit, tube formula, Wishart distribution.

The results of this paper will be of basic importance for statistical inference with these models.

Other important multilinear structures are mixed cumulants of a random vector. For example, consider the third cumulant $\mathrm{cum}(u'x, v'x, w'x) = \sum u_i v_j w_k \,\mathrm{cum}(x_i, x_j, x_k)$ of a random vector $x = (x_i)$. Multilinearity is the basic property of cumulants. Note that this is a symmetric multilinear form since $\mathrm{cum}(x_i, x_j, x_k)$ is permutation invariant with respect to the indices. From the distributional point of view it is important to test the multiple hypotheses that $\mathrm{cum}(u'x, v'x, w'x) = 0$ for all u, v, w. A natural test statistic is the maximum of sample cumulants with respect to u, v, w normalized so that $u'x, v'x, w'x$ have unit variances. This statistic is shown to coincide with the test statistic for multivariate normality by Malkovich and Afifi (1973). Our results on symmetric multilinear form provide satisfactory distribution theory for this type of statistic.

1.2. *The problems.* Here we present the canonical forms of the statistics studied in this paper. Let $Z = (z_{j_1\cdots j_k})$, $j_i = 1, \dots, q_i$, $i = 1, \dots, k$, be a k-way random array whose components are distributed independently according to the standard normal distribution $N(0, 1)$. Let $h_i = (h_{i1}, \dots, h_{iq_i})' \in R^{q_i}$, $i = 1, \dots, k$, be coefficient vectors and consider a multilinear form of degree k, or k-linear form defined by

$$
\begin{aligned}
g_k(h_1, \dots, h_k; Z) &= \sum_{j_1=1}^{q_1} \cdots \sum_{j_k=1}^{q_k} h_{1j_1} \cdots h_{kj_k} z_{j_1\dots j_k} \\
&= (h_1 \otimes \cdots \otimes h_k)' z,
\end{aligned} \tag{1.1}
$$

where $\otimes$ denotes the Kronecker product and $z = \mathrm{vec}(Z) = (z_{11\cdots 1}, z_{11\cdots 2}, \dots, z_{q_1 q_2 \dots q_k})'$ is the $(\prod_{i=1}^k q_i)$-dimensional column vector consisting of the components of Z ordered lexicographically. We first consider the maximum of the k-linear form under the condition $\|h_i\| = 1$ for any i, that is,

$$
T_k = \max_{\|h_i\|=1, \forall i} g_k(h_1, \dots, h_k; Z) \tag{1.2}
$$

and its standardized statistic

$$
U_k = T_k / \|z\| = \max_{\|h_i\|=1, \forall i} g_k(h_1, \dots, h_k; Y), \tag{1.3}
$$

where $\|\cdot\|$ denotes the usual Euclidean norm and $Y = Z/\|z\| = (z_{j_1\cdots j_k}/\|z\|)$. Note that for the purpose of maximization the constraint $\|h_i\| = 1, \forall i$, in (1.2) and (1.3) is equivalent to the constraint $\|h_1 \otimes \cdots \otimes h_k\| = 1$, because g_k is linear in each h_i. $T_k \geq 0$ and $0 \leq U_k \leq 1$ since $\|h_1 \otimes \cdots \otimes h_k\| = \prod_i \|h_i\| = 1$.

Second, by imposing the additional condition that $q_1 = \cdots = q_k$ $(=q$, say), we consider the symmetric k-linear form, $g_k(h_1, \dots, h_k; \mathrm{sym}(Z))$, where $\mathrm{sym}(Z)$ is the k-way array with $(j_1, \dots, j_k)$th component

$$
\frac{1}{k!} \sum_{\pi \in S_k} z_{j_{\pi(1)} \cdots j_{\pi(k)}},
$$

and S_k denotes the set of permutations of $\{1, \dots, k\}$. The corresponding maxima are

$$\widetilde{T}_k = \max_{\|h_i\|=1, \forall i} g_k(h_1, \dots, h_k; \mathrm{sym}(Z)), \tag{1.4}$$

and its standardization

$$\widetilde{U}_k = \widetilde{T}_k / \|z\|. \tag{1.5}$$

Here the maximum in (1.4) is attained when $h_1 \otimes \cdots \otimes h_k = \pm(h \otimes \cdots \otimes h)$ for some $h \in R^q$. This is because $g_k(h_1, h_2, h_3, \dots, h_k; \mathrm{sym}(Z))$ is a symmetric bilinear form in h_1 and h_2 for fixed $h_3, \dots, h_k$, and hence its maximum is attained when $h_1 = h_2$ or $h_1 = -h_2$. Therefore we have

$$\widetilde{T}_k = \max_{\|h\|=1} \{\pm \tilde{g}_k(h; Z)\}, \tag{1.6}$$

where $h = (h_1, \dots, h_q)' \in R^q$ and

$$\tilde{g}_k(h; Z) = g_k(h, \dots, h; Z) = (\underbrace{h \otimes \cdots \otimes h}_{k})' z. \tag{1.7}$$

Note that $\widetilde{T}_k \geq 0$ and $0 \leq \widetilde{U}_k \leq 1$.

The primary purpose of this paper is to give some explicit formulas for the upper tail probabilities for T_k, $\widetilde{T}_k$, and their standardizations U_k, $\widetilde{U}_k$. More precisely, we shall give asymptotic series for $P(T_k \geq a)$ and $P(\widetilde{T}_k \geq a)$ when a is large and expressions for $P(U_k \geq a)$ and $P(\widetilde{U}_k \geq a)$ which hold exactly when the a's are greater than suitable constants.

When $k = 2$, the k-way array becomes a $q_1 \times q_2$ random matrix $Z = (z_{j_1 j_2})$, and some of the statistics introduced above were studied in the conventional framework of multivariate analysis. On the other hand for $k \geq 3$, except for Monte Carlo simulation, results have not been obtained. We summarize the facts about the case $k = 2$ here. Since T_2 is the largest singular value of Z, the squared statistic T_2^2 is the largest eigenvalue $\lambda_1(Z'Z)$ of the $q_2 \times q_2$ matrix $Z'Z$ [or $\lambda_1(ZZ')$ of the $q_1 \times q_1$ matrix ZZ'], where $Z'Z$ is distributed according to the Wishart distribution $W(I_{q_2}, q_1)$. When the parameter matrix of the Wishart matrix is the identity (i.e., the null case) and the matrix size is not large, the distribution of the largest eigenvalue can be obtained in principle by integrating out the other eigenvalues in the joint density of eigenvalues [e.g., Chapter 13 of Anderson (1984)]. Along this line some algorithms have been devised. See a survey paper by Pillai (1976). For the distribution of $U_2^2 = \lambda_1(Z'Z)/\mathrm{tr}(Z'Z)$, the largest eigenvalue divided by the trace of the same Wishart matrix, Davis (1972) proposed a method to obtain the cumulative distribution function by inverting a Laplace transformation symbolically and gave the explicit expressions for $\min(q_1, q_2) = 2, 3$. Using this method, Schuurmann, Krishnaiah and Chattopadhyay (1973) provided a table of quantiles. The maximum $\widetilde{T}_2$ equals $\max\{\lambda_1(A), -\lambda_q(A)\}$, where $\lambda_1(A)$ and $\lambda_q(A)$ are the largest and smallest eigenvalues of the symmetric matrix $A = \mathrm{sym}(Z) = (Z + Z')/2$. This random matrix is also well studied because this is the limiting distribution of standardized Wishart matrix as the degrees of freedom go to infinity. With the same technique as for the case of a Wishart matrix, the distribution function for $\widetilde{T}_2$ can be calculated numerically.

Although these approaches enable us to evaluate the distribution functions numerically, they are applicable only when $k = 2$ and the size of the matrix is not too large. Our approach will give simple and sufficiently accurate formulas for any k.

1.3. *The tube method: an integral-geometric approach.* In order to derive the tail probabilities of the maxima introduced above, we employ a geometric approach. Around sixty years ago, in order to give a significance level of a likelihood ratio test in a certain nonlinear regression model, Hotelling (1939) defined the one-dimensional tubes in the Euclidean space and the unit sphere, and derived formulas for the volume of tubes. Hotelling's tube formula was immediately generalized to general dimensional cases by Weyl (1939). For the history and applications to statistics, see Knowles and Siegmund (1989). More recently, Sun (1993) has developed a general theory of the tail probability of the maximum of a Gaussian random field with a Karhunen–Loève expansion which is not necessarily finite. Sun's theory states that the tail probability is expressed in terms of the geometric quantities which appear as the coefficients of Weyl's tube formula for a manifold defined by the Karhunen–Loève expansion. As we shall see later, evaluation of the tail probabilities for the standardized maxima U_k, $\widetilde{U}_k$ can be reduced to the evaluation of the volume of tubes. Derivation of the tail probabilities for the nonstandardized maxima T_k, $\widetilde{T}_k$ are within the scope of Sun (1993). However, it is in general difficult to give explicit expressions for all of the coefficients in Weyl's tube formula in each particular application. For example, Sun (1991) discussed the tail probability of a projection pursuit index in exploratory projection pursuit and gave only the first two terms in integral forms. In our paper, in order to evaluate all of the coefficients in the volume of tube, we use a tube formula represented in terms of the second fundamental form, whereas the tube formula described in Weyl (1939) is expressed in terms of the intrinsic curvature tensors. We believe that for evaluating the coefficients explicitly the tube formula based on the second fundamental form is often more helpful than that based on the curvature tensors. Indeed, for our problem we will obtain all of the coefficients of the volume of tube by integrating the second fundamental form.

The outline of this paper is as follows. In Section 2, applications of the distributions of the maxima mentioned briefly in Section 1.1 are examined in more detail. In Section 3, we first prepare geometric tools and then give our main results on k-linear forms and symmetric k-linear forms in Theorem 3.2 and Theorem 3.3, respectively. In the geometric preparation we summarize the theory by Weyl (1939) and Sun (1993) in a form comparable to the approach in our recent paper, Takemura and Kuriki (1997), where convexity was assumed. We also give a theorem to calculate the critical radius, the extreme radius for which Weyl's tube formula is valid. Sections 4 and 5 are devoted to discussion of some numerical examples and to the derivation of some geometric quantities needed for the tube formulas for k-linear forms and symmetric k-linear forms, respectively. Our numerical examples and Monte Carlo studies demonstrate that the obtained expressions are practical enough for calculating P-values. Some of the details on geometry and proofs are provided in the Appendix.

2. Applications to testing hypotheses. In this section we discuss testing problems where the distributions of the maxima of multilinear forms introduced in Section 1.2 are required for calculating their P-values.

2.1. *Tests for interaction in multiway data.* Let $x_{ij},\ i = 1, \dots, I,\ j = 1, \dots, J$, be observed as two-way layout data without replication. For such data Johnson and Graybill (1972) assumed a model:

$$x_{ij} = \alpha_i + \beta_j + \phi u_i v_j + \varepsilon_{ij}, \tag{2.1}$$

where $\alpha_i,\ \beta_j,\ \phi,\ u_i$ and v_j are unknown parameters and ε_{ij} is a random error distributed independently as $N(0, \sigma^2)$ with σ^2 unknown. They proposed a test for interaction effects, or nonadditivity, as a likelihood ratio test for testing $H_0\colon \phi = 0$. They showed that the critical region of the likelihood ratio test is given by

$$\lambda_1(Z'Z)/\mathrm{tr}(Z'Z) > c \tag{2.2}$$

for some constant c, where $Z = (z_{ij})$ is a $I \times J$ matrix with (i, j)th element,

$$z_{ij} = x_{ij} - x_{i\cdot} - x_{\cdot j} + x_{\cdot\cdot}\ ,$$

and $\lambda_1(Z'Z)$ is the largest eigenvalue of $Z'Z$. Here the dot means the arithmetic mean with respect to the corresponding subscript, for example, $x_{i\cdot} = (1/J)\sum_{j=1}^{J} x_{ij}$. Under the null hypothesis $H_0\colon \phi = 0$, the distribution of the likelihood ratio test statistic in (2.2) is shown to be that of U_2^2 in (1.3) with $q_1 = I - 1,\ q_2 = J - 1$.

As an extension of Johnson and Graybill (1972), Boik and Marasinghe (1989) considered a test for interaction in a k-way layout without replication. Their model in the case of a three-way layout is

$$x_{ijk} = (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \phi u_i v_j w_k + \varepsilon_{ijk},$$

where ε_{ijk} is distributed independently as $N(0, \sigma^2),\ i = 1, \dots, I,\ j = 1, \dots, J,\ k = 1, \dots, K$. Here as in (2.1) the parameters $(\alpha\beta)_{ij},\ (\alpha\gamma)_{ik},\ (\beta\gamma)_{jk},\ \phi,\ u_i,\ v_j,\ w_k$ and σ^2 are unknown. Using this model they proposed a test for the null hypothesis $H_0\colon \phi = 0$. The critical region of the likelihood ratio test is of the form

$$\max_{\|u\|=\|v\|=\|w\|=1} \Biggl(\sum_{i,j,k} u_i v_j w_k z_{ijk}\Biggr)^2 \Bigg/ \sum_{i,j,k} z_{ijk}^2 > c, \tag{2.3}$$

where $u = (u_1, \dots, u_I)'$, $v = (v_1, \dots, v_J)'$ and $w = (w_1, \dots, w_K)'$ are unit vectors and

$$z_{ijk} = x_{ijk} - x_{ij\cdot} - x_{i\cdot k} - x_{\cdot jk} + x_{i\cdot\cdot} + x_{\cdot j\cdot} + x_{\cdot\cdot k} - x_{\cdot\cdot\cdot}$$

is the residual under H_0. The distribution of the test statistic in (2.3) under H_0 is shown to be that of U_3^2 in (1.3) with $q_1 = I - 1$, $q_2 = J - 1$, $q_3 = K - 1$. Monte Carlo studies to estimate the distribution function of U_3^2 are found in Boik (1990) and Kawasaki and Miyakawa (1996).

In a similar fashion, this method can be extended to a multiway layout of higher order. It is easily proved that the distribution of U_k^2 arises as the null distribution of the likelihood ratio test statistics for testing interaction in a k-way layout.

REMARK 2.1. In the fields of psychometrics and chemometrics, three-way and higher multiway data analysis are extensively studied. The aim of this work is to extend the methods of principal component analysis or correspondence analysis which have been successfully developed in two-way data analysis into multiway data analysis. Leurgans and Ross (1992) with discussion comments by three authors are helpful for surveying multiway data analysis and related topics in mathematics. One of the most studied models among them is the PARAFAC model, which is called INDSCAL in the context of multidimensional scaling, where suitably preprocessed three-way data are modeled by a three-way array with (i, j, k)th cell of the structure $\sum_{r=1}^{R} u_i^{(r)} v_j^{(r)} w_k^{(r)}$. The model by Boik and Marasinghe (1989) is a particular case of a PARAFAC model where the "rank" R is equal to 1. Although there are quite a few papers on multiway data analysis, almost all of them are focused only on modeling and fitting and distributional results have not been given. Our geometric approach has the advantage that it enables us to tackle distribution theory and related statistical inferences for multiway data analysis.

2.2. *Tests for multivariate normality.* Let $x \in R^q$ be a random vector distributed according to a continuous distribution with unknown mean vector μ and non-degenerate covariance matrix Σ. Let $u_1, \dots, u_k$ be vectors normalized so that $u_i'\Sigma u_i = 1$ and hence $u_i'x$ has a unit variance. By Roy's union-intersection principle, we consider the maximum of the joint cumulant

$$B_k = \max_{u_i'\Sigma u_i = 1,\, \forall i} \operatorname{cum}(u_1'x, \dots, u_k'x) \tag{2.4}$$

as a nonnegative measure of the departure from multivariate normality. Note that (2.4) is independent of μ and Σ, and takes the values zero when the distribution of x is a multivariate normal. Since the joint cumulant in (2.4) is symmetric in $u_1, \dots, u_k$, the maximum is attained when $u_1 = \cdots = u_k$, or $u_1 = \cdots = u_{k-1} = -u_k$, and hence (2.4) is reduced to

$$B_k = \max_{u'\Sigma u = 1} |K_k(u)| = \max_{u \in S^{q-1}} \frac{|K_k(u)|}{K_2(u)^{k/2}}, \tag{2.5}$$

where

$$K_k(u) = \mathrm{cum}(\underbrace{u'x, \dots, u'x}_{k}).$$

Malkovich and Afifi (1973) called B_3 and B_4 multivariate skewness and kurtosis, respectively.

Assume that independent and identically distributed sample vectors $x_1, \dots, x_n \in R^q$ are observed. The sample version $\widehat{B}_k$ of B_k (2.5) is obtained by replacing $K_k(u)$ with the sample cumulant (cumulant with respect to the empirical distribution) $\widehat{K}_k(u)$ of $u'x_1, \dots, u'x_n$. Malkovich and Afifi (1973) proposed the tests for which the hypothesis of multivariate normality is rejected when $\widehat{B}_3$ or $\widehat{B}_4$ are greater than some critical points. From now on we consider the null distribution of $\widehat{B}_k$. We can assume $\mu = 0$ and $\Sigma = I_q$ without loss of generality.

Let $C(S^{q-1})$ be the Banach space of continuous function on the unit sphere S^{q-1} endowed with the supremum norm. Let $Z_k(u)$ be a Gaussian random field in $C(S^{q-1})$ with zero mean and covariance function $E[Z_k(u)Z_k(v)] = k!\,(u'v)^k$. The following theorem is an extension of Machado (1983) who only treated the cases $k = 3, 4$. The proof is given in Appendix A.1.

THEOREM 2.1. *Let $x_1, \dots, x_n \in R^q$ be independently distributed according to $N(0, I_q)$. Then $\sqrt{n}\widehat{K}_k(u)/\widehat{K}_2(u)^{k/2}$ converges in distribution to the Gaussian field $Z_k(u)$ as n goes to infinity in the space $C(S^{q-1})$.*

Here it is easily seen that $Z_k(u)$ has a representation

$$Z_k(u) = \sqrt{k!}\,\tilde{g}_k(u; Z),$$

where $\tilde{g}_k(u; Z)$ is defined in (1.7). By virtue of the continuous mapping theorem, $\sqrt{n}\widehat{B}_k$ converges to $\sqrt{k!}\,\widetilde{T}_k$ in distribution under the hypothesis of multivariate normality. Therefore, we can obtain approximate critical values for $\widehat{B}_k$ from the tail probability of $\widetilde{T}_k$. In Section 5.1 we will examine the accuracy of approximation by Monte Carlo studies.

REMARK 2.2. The test by Malkovich and Afifi (1973) can be regarded as a kind of projection pursuit for searching a direction $u \in S^{q-1}$ of nonnormality. Indeed the use of the standardized cumulant $|\widehat{K}_k(u)|/\widehat{K}_2(u)^{k/2}$ as a projection pursuit index was proposed by Huber (1985), Example 5.4. Although it has been pointed out that the standardized cumulant as a projection pursuit index is too sensitive with respect to tails of the distribution [Friedman (1987)], it still has an advantage that the approximate significance level can be calculated via the tail probability formula given by this paper.

3. Geometric preliminaries and main results. In this section we summarize geometric tools in a form suitable for our development and then give our main results in Theorem 3.2 and Theorem 3.3.

3.1. *Distribution of the projection onto a nonconvex smooth cone.* Here we summarize results mainly from Weyl (1939), Sun (1993) and Johansen and Johnstone (1990). Furthermore by re-examining Sun's derivation of the asymptotic expansion of the tail probability, we give upper and lower bounds for the tail probability $P(T \geq a)$ for the nonstandardized maximum [such as T_k or $\widetilde{T}_k$ in (1.2) or (1.4)], which are valid for each $a > 0$. We provide our own simplified proofs of these results in Appendix A.3.

Let $\{Z(t) \in R \mid t \in I\}$ be a Gaussian random field such that $E[Z(t)] = 0$, $E[Z(t)^2] = 1$ with the index set I. We assume that $Z(t)$ has a finite Karhunen–Loève expansion:

$$Z(t) = \sum_{i=1}^{p} \phi_i(t) z_i = \phi(t)' z, \qquad t \in I, \tag{3.1}$$

where $\phi(t) = (\phi_1(t), \dots, \phi_p(t))'$, $z = (z_1, \dots, z_p)'$ and z_i, $i = 1, \dots, p$, are independent standard normal random variables. Note that $E[Z(s)Z(t)] = \phi(s)'\phi(t)$, and that $\|\phi(t)\| = 1$ since $E[Z(t)^2] = 1$. Let

$$M = \phi(I) = \{\phi(t) \mid t \in I\} \subset S^{p-1}.$$

We put some assumptions on M.

ASSUMPTION 3.1. M is a compact C^2-submanifold without boundary of dimension d in S^{p-1}.

Define a closed cone $K \subset R^p$ associated with M by

$$K = \bigcup_{c \geq 0} cM = \{c\phi(t) \mid c \geq 0,\ t \in I\}, \tag{3.2}$$

which is smooth except for the origin. For $z \in R^p$ let $z_K \in K$ denote the projection of z onto K:

$$\|z - z_K\| = \min_{y \in K} \|z - y\|.$$

In

$$\begin{aligned} \min_{y \in K} \|z - y\|^2 &= \min_{r \geq 0,\, u \in M} \|z - ru\|^2 = \min_{r \geq 0,\, u \in M} \{\|z\|^2 - 2r(u'z) + r^2\} \\ &= \min_{r \geq 0} \Big\{ \|z\|^2 - 2r\Big(\max_{u \in M} u'z\Big) + r^2 \Big\}, \end{aligned}$$

the minimum is attained when $r = \max\{\max_{u \in M} u'z, 0\}$. Since $\|y\| = r$, this implies that

$$\|z_K\| = \max_{u \in M} u'z = \max_{t \in I} Z(t)$$

unless $\|z_K\| = 0$. See Figure 1 (left).

Note that z_K exists since K is closed. Here z_K may not be unique but $\|z_K\|$ and $\|z - z_K\|$ are uniquely determined. In Takemura and Kuriki (1997) we investigated properties of projections onto a convex cone K. In the case of the

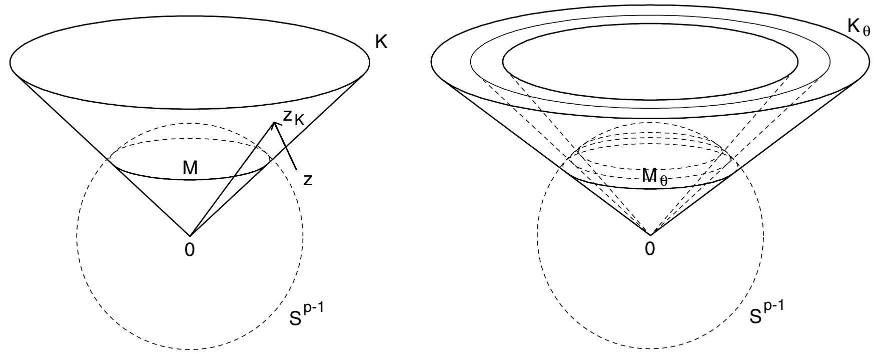

FIG. 1. *Index set M, cone K, projection z_K (left). Tube M_θ and associated cone K_θ (right).*

convex cone, z_K is always uniquely determined and its distribution is nicely characterized as a $\bar{\chi}^2$ distribution. By introducing a cone K in (3.2) it becomes clear that the results in this section are closely related to those in Takemura and Kuriki (1997).

For nonconvex K we need to be concerned with the uniqueness of the projection z_K. The essential notions are the tube around M and the critical radius (critical angle) of M with respect to the geodesic distance of S^{p-1}. Here the geodesic distance between two points $u, v \in S^{p-1}$ is given by $\arccos(u'v)$, which is the length of the part of the great circle joining u and v.

For $0 < \theta < \pi$ the tube of geodesic distance θ around M on S^{p-1} is defined by

$$M_\theta = \Big\{ v \in S^{p-1} \mid \max_{u \in M} u'v > \cos\,\theta \Big\}.$$

For each $u \in M$ let $T_u(M)$ denote the tangent space of M at u. Define a subset $C_\theta(u)$ of S^{p-1} by the set of points v with the geodesic distance less than θ from u and such that the geodesic from u to v is orthogonal to $T_u(M)$ at u. That is,

$$C_\theta(u) = \{ v \in S^{p-1} \mid u'v > \cos\theta \} \cap \{ u + T_u(M)^\perp \},$$

where $T_u(M)^\perp$ denotes the orthogonal complement of $T_u(M)$ in R^p. Since M is a closed submanifold of S^{p-1} without boundary we obviously have

$$M_\theta = \bigcup_{u \in M} C_\theta(u).$$

It is said that M_θ does not have self-overlap if $C_\theta(u),\ u \in M$, are disjoint. The supremum θ_c of θ for which M_θ does not have self-overlap is called the critical radius (or critical angle) of M,

$$\theta_c = \sup\{ \theta \mid M_\theta \text{ does not have self-overlap} \}.$$

Note that the critical radius never exceeds $\pi/2$, which is attained when $M = S^{d'-1} \subset S^{p-1},\ d' < p$.

For determining the critical radius of M the following lemma [Proposition 4.3 of Johansen and Johnstone (1990)] is very useful. Although Johansen and Johnstone (1990) stated their Proposition 4.3 for the case $\dim M = 1$ only, its statement and proof hold for $\dim M = d > 1$ almost verbatim and we omit the proof.

LEMMA 3.1. *The critical radius θ_c of M is given by*

$$\cot^2 \theta_c = \sup_{u,\, v \in M} \frac{1 - u' P_v u}{(1 - u'v)^2}, \tag{3.3}$$

where P_v is the orthogonal projection onto the tangent space $T_v(K)$ of K of (3.2) *at v.*

REMARK 3.1. Let

$$h(u, v) = \frac{\sqrt{1 - u' P_v u}}{1 - u'v} \tag{3.4}$$

be the square root of the argument of the supremum in (3.3). In Appendix A.2 we show that $h(u, v)$ can be defined also for $u = v$ by taking the appropriate supremum as $u \to v$, and the maximum over the compact set $M \times M$ exists and is finite. This implies that the critical radius θ_c is positive under our Assumption 3.1.

Let K_θ denote the cone associated with M_θ:

$$K_\theta = \bigcup_{c \ge 0} c M_\theta.$$

See Figure 1 (right). As before K denotes the cone associated with M. If $z \in K_{\theta_c}$ then the projection z_K of z onto K is unique. For $z \in K_{\theta_c}$ write

$$z = z_K + (z - z_K) = ru + sv,$$

where $r = \|z_K\|,\ s = \|z - z_K\|$ and

$$u = z_K / r \in M, \qquad v = (z - z_K)/s \in T_u(K)^\perp \cap S^{p-1}.$$

The one-to-one correspondence $z \leftrightarrow (r, u, s, v)$ is of class C^1 and Weyl (1939) derived its Jacobian. We state the Jacobian in the following lemma.

LEMMA 3.2. *Let $H(u, v)$ denote the second fundamental form of K at u with respect to the direction $v \in T_u(K)^\perp \cap S^{p-1}$. Then*

$$dz = \left| I_{d+1} + \frac{s}{r} H(u, v) \right| r^d dr\ du\ s^{p-d-2} ds\ dv, \tag{3.5}$$

where dz denotes the p-dimensional Lebesgue measure, du denotes the volume element of M and dv denotes the volume element of $T_u(K)^\perp \cap S^{p-1}$ [the $(p - d - 2)$-dimensional unit sphere restricted to the space $T_u(K)^\perp$].

A simple proof of Lemma 3.2 is given in Appendix A.1 of Kuriki and Takemura (2000).

Let $\mathrm{tr}_j\, H$ denote the jth trace, that is, the jth elementary symmetric function of the eigenvalues of $H = H(u, v)$. Let $\mathrm{tr}_0\ H \equiv 1$. Although $T_u(K)$ is of dimension $d+1$, $\mathrm{rank} H(u, v) \le d$ since $H(u, v)$ has at least one eigenvalue (principal curvature) equal to 0 with the eigenvector (principal direction) u. Therefore,

$$\left| I_{d+1} + \frac{s}{r} H(u, v) \right| r^d = \sum_{e=0}^{d} r^{d-e} s^e \ \mathrm{tr}_e\ H$$

and (3.5) can alternatively be written as

$$dz = \sum_{e=0}^{d} r^{d-e} s^{p-d-2+e}\ dr\ ds\ \mathrm{tr}_e\ H(u, v)\ du\, dv. \tag{3.6}$$

Moreover, as will be explained in Appendix A.2, the principal curvatures of K at u with respect to the principal directions orthogonal to u coincide with the principal curvatures of M at u. In other words, $H(u, v)$ appearing in (3.5) and (3.6) can be replaced by the second fundamental form of M at u with respect to v.

From Lemma 3.2 the volume of M_θ, $\theta \le \theta_c$, is obtained as follows. Let

$$\Omega_d = \mathrm{Vol}(S^{d-1}) = \frac{2\pi^{d/2}}{\Gamma(d/2)}$$

denote the total volume of S^{d-1} and let $\overline{B}_{m,n}(a)$ denote the upper tail probability of the beta distribution with parameter (m, n),

$$\overline{B}_{m,\,n}(a) = \int_a^1 \frac{1}{B(m, n)} \xi^{m-1} (1-\xi)^{n-1} d\xi.$$

LEMMA 3.3. *Let $z \in R^p$ be distributed according to the standard multivariate normal distribution $N(0, I_p)$. For $0 \le \theta \le \theta_c$,*

$$\mathrm{Vol}(M_\theta) = \Omega_p \cdot P(z \in K_\theta) = \Omega_p \sum_{\substack{e=0 \\ e:\mathrm{even}}}^{d} w_{d+1-e} \overline{B}_{\frac{1}{2}(d+1-e),\, \frac{1}{2}(p-d-1+e)}(\cos^2\theta),$$

where

$$w_{d+1-e} = \frac{1}{\Omega_{d+1-e}\Omega_{p-d-1+e}} \int_M \left[\int_{T_u(K)^\perp \cap S^{p-1}} \mathrm{tr}_e H(u, v)\, dv \right] du. \tag{3.7}$$

This formula was given by Weyl (1939). A simple proof is given in Appendix A.3. Note that w_{d+1-e} corresponds to the weight of the $\bar{\chi}^2$ distribution for a piecewise smooth cone given in Theorem 2.4 of Takemura and Kuriki (1997).

Now consider the tail probability of the standardized maximum statistic. Let $Z(t)$ be given as in (3.1) and consider

$$U = \max_{t \in I} \phi(t)' z / \|z\| = \max_{u \in M} u' z / \|z\|. \tag{3.8}$$

Because $z/\|z\|$ has a uniform distribution over S^{p-1}, for $-1 \le a \le 1$,

$$P(U \ge a) = \frac{1}{\Omega_p}\mathrm{Vol}(M_\theta), \qquad \theta = \theta(a) = \arccos(a).$$

If $a \ge \cos\theta_c$ then $\mathrm{Vol}(M_{\theta(a)})$ is given by Lemma 3.3. For convenience we state this as a lemma.

LEMMA 3.4. *For $a \ge \cos\,\theta_c$,*

$$P(U \ge a) = \sum_{\substack{e=0\\ e:\text{even}}}^{d} w_{d+1-e}\overline{B}_{\frac12(d+1-e),\frac12(p-d-1+e)}(a^2). \tag{3.9}$$

Now we consider the nonstandardized statistic. Let

$$T = \max_{t\in I}\phi(t)'z = \max_{u\in M} u'z. \tag{3.10}$$

Denote the density and the upper tail probability of the χ^2 distribution with m degrees of freedom by $g_m(a)$ and $\overline{G}_m(a)$, respectively. Furthermore, for $a, b > 0$ define

$$Q_{m,n}(a,b) = \int_a^\infty g_m(\xi)(1-\overline{G}_n(b\xi))\,d\xi = \overline{G}_m(a) - \int_a^\infty g_m(\xi)\,\overline{G}_n(b\xi)\,d\xi.$$

$Q_{m,n}(a,b)$ can be evaluated by numerical integration. It is also easy to obtain recurrence relations among $Q_{m,n}(a,b)$'s.

Now we can state the following theorem.

THEOREM 3.1. *Let w_{d+1-e} be given in* (3.7). *For $a > 0$,*

$$Q_L(a) \le P(T \ge a) \le Q_U(a),$$

where

$$Q_L(a) = \sum_{\substack{e=0\\ e:\text{even}}}^{d} w_{d+1-e}Q_{d+1-e,p-d-1+e}(a^2, \tan^2\theta_c) \tag{3.11}$$

and

$$Q_U(a) = Q_L(a) + \overline{G}_p(a^2(1+\tan^2\theta_c))\left(1 - \frac{\mathrm{Vol}(M_{\theta_c})}{\Omega_p}\right). \tag{3.12}$$

The proof is given in Appendix A.3. Furthermore, it is easy to see that

$$Q_U(a) - Q_L(a) \le \overline{G}_p(a^2(1+\tan^2\theta_c)) = o(\overline{G}_1(a^2))$$

and

$$\left| Q_L(a) - \sum_{\substack{e=0 \\ e:\text{even}}}^{d} w_{d+1-e} \overline{G}_{d+1-e}(a^2) \right|$$
$$\leq \sum_{\substack{e=0 \\ e:\text{even}}}^{d} |w_{d+1-e}| \int_{a^2}^{\infty} g_{d+1-e}(\xi)\, \overline{G}_{p-d-1+e}(\xi \tan^2 \theta_c)\, d\xi$$
$$\leq \sum_{\substack{e=0 \\ e:\text{even}}}^{d} |w_{d+1-e}|\ \overline{G}_p(a^2(1+\tan^2\theta_c)) = o(\overline{G}_1(a^2)).$$

As a corollary to Theorem 3.1 we have the following result by Sun (1993).

COROLLARY 3.1.

$$P(T \geq a) = \sum_{\substack{e=0 \\ e:\text{even}}}^{d} w_{d+1-e} \overline{G}_{d+1-e}(a^2) + o(\overline{G}_1(a^2)) \qquad \text{as } a \to \infty. \tag{3.13}$$

REMARK 3.2. Let $\mathrm{lin} K$ be the intersection of all linear subspaces containing the cone K. When $\mathrm{lin} K$ is a proper subset of R^p, there exists a Karhunen–Loève expansion of dimension $p' = \dim(\mathrm{lin} K) < p$, and p in Theorem 3.1 should be replaced with p' so as to improve the lower and upper bounds.

To conclude this subsection we point out a useful relationship between the coefficients w_{d+1-e}. Let $\chi(M)$ denote the Euler characteristic (Euler–Poincaré characteristic) of the index set M. The next lemma follows immediately from the Gauss–Bonnet theorem.

LEMMA 3.5. *For $d = \dim(M)$ even,*

$$\chi(M) = 2 \sum_{\substack{e=0 \\ e:\text{even}}}^{d} w_{d+1-e} = 2(w_1 + w_3 + \cdots + w_{d-1}).$$

A proof is given in Appendix A.3. Note that $\chi(M) = 0$ for d odd, since the Euler characteristic of a closed Riemannian manifold of odd dimension is zero.

3.2. *Main results.* We present our main results on the tail probability of the maxima of a multilinear form and a symmetric multilinear form in Theorem 3.2 and Theorem 3.3, respectively.

3.2.1. *Tail probability of the maximum of a multilinear form.* For a multilinear form let

$$M_k = \{h_1 \otimes \cdots \otimes h_k \mid h_i \in S^{q_i - 1},\ i = 1, \ldots, k\} \tag{3.14}$$

be the manifold of dimension $d = \sum_{i=1}^{k}(q_i - 1)$ in R^p with $p = \prod_{i=1}^{k} q_i$. Since $\|h_1 \otimes \cdots \otimes h_k\| = \prod_{i=1}^{k} \|h_i\| = 1$, it follows that $M_k \subset S^{p-1}$. It is easy to check that M_k is a submanifold of S^{p-1} satisfying Assumption 3.1. The statistics T_k and U_k in (1.2) and (1.3) are written as

$$T_k = \max_{u \in M_k} u'z, \qquad U_k = \max_{u \in M_k} u'z/\|z\|,$$

respectively, where z is a p-dimensional column vector distributed as $N(0, I_p)$. Then T_k and U_k are of the form of the random variables T and U in (3.10) and (3.8) whose tail probabilities can be derived by virtue of Lemma 3.4, Theorem 3.1 or Corollary 3.1.

In order to state the main theorem on multilinear forms, we need to introduce a combinatorial quantity.

DEFINITION 3.1. For nonnegative integers m and $d_1, \dots, d_k$, define a nonnegative integer $n_k(d_1, d_2, \dots, d_k; m)$ as follows. Let $A = \{1, \dots, d\}$ with $d = \sum_{i=1}^k d_i$. Put

$$A_i = \Big\{ a \in A \Big| \sum_{j=1}^{i-1} d_j + 1 \le a \le \sum_{j=1}^{i} d_j \Big\}, \qquad i = 1, \dots, k, \tag{3.15}$$

which form a partition of A. Consider a set of m pairings,

$$\begin{aligned} \big\{ (a_1, a_2), \dots, (a_{2m-1}, a_{2m}) \mid a_1 < a_3 < \cdots < a_{2m-1}, \\ a_1 < a_2, \dots, a_{2m-1} < a_{2m} \big\} \end{aligned} \tag{3.16}$$

such that:

(i) $2m$ indices $a_1, a_2, \dots, a_{2m}$ are distinct elements of $A = \{1, 2, \dots, d\}$.

(ii) For each pairing in (3.16), say $(a_{2l-1},\ a_{2l})$, a_{2l-1} and a_{2l} do not belong to the same set of (3.15), that is, if $a_{2l-1} \in A_i$ and $a_{2l} \in A_j$ then $i \neq j$.

Then $n_k(d_1, d_2, \dots, d_k; m)$ is defined as the total number of sets (3.16) of m pairings satisfying (i) and (ii).

A recurrence formula for calculating $n_k(d_1, \dots, d_k; m)$ is given in Lemma A.2 of Appendix A.4. Now we can state our result on multilinear forms. The proof is given in Section 4.2.

THEOREM 3.2. *The tail probabilities of T_k in* (1.2) *and U_k in* (1.3) *for $k \ge 2$ are given by Theorem* 3.1 (*or Corollary* 3.1) *and Lemma* 3.4, *respectively, where $d = \sum_{i=1}^k (q_i - 1)$, $p = \prod_{i=1}^k q_i$, and*:

(i) *The nonzero coefficients w_{d+1-e} are given by*

$$w_{d+1-e} = \frac{\pi^{\frac{1}{2}(k-1)}}{\prod_{i=1}^k \Gamma(\frac{1}{2} q_i)} \Big(-\frac{1}{2}\Big)^{e/2} \Gamma\Big(\frac{1}{2}(d+1-e)\Big)\, n_k(q_1 - 1, \dots, q_k - 1; e/2),$$

$e = 0, 2, \dots, [d/2] \times 2$, *with n_k given by Definition* 3.1.

(ii) *The critical radius* θ_c *is given by*

$$\theta_c = \cos^{-1}\sqrt{\frac{2k-2}{3k-2}}.$$

When $k=2$, $n_2(d_1, d_2; m)$ is the total number of m pairings of the form

$$\{(b_1, c_1), \dots, (b_m, c_m)\}, \ b_1, \dots, b_m \in A_1, \ c_1, \dots, c_m \in A_2.$$

There are $\binom{d_1}{m}$ ways of choosing m elements from $A_1 = \{1, \dots, d_1\}$ and there are $\binom{d_2}{m}$ ways of choosing m elements from $A_2 = \{d_1+1, \dots, d\}$. Furthermore there are $m!$ ways of forming pairs of the $2m$ chosen elements. Therefore we have

$$n_2(d_1, d_2; m) = \binom{d_1}{m}\binom{d_2}{m} m!.$$

The tail probability formula for $k=2$ is summarized in terms of the Wishart distribution as follows.

COROLLARY 3.2. *Let* W *be a* $q \times q$ *Wishart matrix distributed as* $W(I_q, \nu)$ *with* ν $(\geq q)$ *degrees of freedom, and let* $\lambda_1(W)$ *be the largest eigenvalue of* W. *Then the tail probabilities* $P(\lambda_1(W) \geq a^2)$ *and* $P(\lambda_1(W)/\mathrm{tr}(W) \geq a^2)$ *are given by* (3.13) *and* (3.9), *respectively, where* $d = q+\nu-2$, $p = q\nu$ *and the nonzero coefficients* w_{d+1-e} *are*

$$\begin{aligned} w_{d+1-e} &= w_{q+\nu-1-e} \\ &= (-1)^{e/2}\, 2^{q+\nu-2-e/2} \frac{\Gamma(\frac{1}{2}(q+1))\,\Gamma(\frac{1}{2}(\nu+1))\,\Gamma(\frac{1}{2}(q+\nu-1-e))}{\sqrt{\pi}\,\Gamma(q-e/2)\,\Gamma(\nu-e/2)\,(e/2)!} \end{aligned} \tag{3.17}$$

for $e = 0, 2, \dots, 2(q-1)$. *The critical radius is* $\theta_c = \pi/4$.

REMARK 3.3. As we will see in Lemma 4.2, the Euler characteristic of M_2 is $\chi(M_2) = 2$ if both q and ν are even, 0 otherwise. The Gauss–Bonnet theorem (Lemma 3.5) implies that when $\nu + q$ is an even integer there is a relation between the coefficients $w_{q+\nu-1-e}$ in (3.17),

$$\sum_{j=0}^{q-1} w_{q+\nu-1-2j} = \begin{cases} 1, & \text{if } q \text{ is odd}, \\ 0, & \text{if } q \text{ is even}. \end{cases} \tag{3.18}$$

Indeed, as we will prove in Appendix A.5, (3.18) holds even when $\nu + q$ is odd. [More precisely (3.18) holds for any real number ν such that $\nu \neq q-1, q-2, \dots$.] Noting this and the relation $\overline{G}_m(x) = 2g_m(x) + \overline{G}_{m-2}(x)$, we can rewrite the formula for the tail probability of $\lambda_1(W)$ given by Corollary 3.2 as

$$\begin{aligned} P(\lambda_1(W) \geq x) \sim & \sum_{j=0}^{q-2} \bar{w}_{q+\nu-1-2j}\, g_{q+\nu-1-2j}(x) \\ & + \begin{cases} \overline{G}_{\nu-q+1}(x), & \text{if } q \text{ is odd}, \\ 0, & \text{if } q \text{ is even}, \end{cases} \end{aligned} \tag{3.19}$$

where

$$\bar{w}_{q+\nu-1-2j} = 2\sum_{i=0}^{j} w_{q+\nu-1-2i}.$$

Hanumara and Thompson (1968) proposed an approximate tail probability formula for $\lambda_1(W)$ by modifying Pillai's approximation formula for the largest eigenvalue of a multivariate beta matrix. Their formula is shown to reduce to our formula (3.19), although it seems complicated at first glance. They concluded that this formula is accurate enough for calculating significance levels and made tables of quantiles based on it. However, Hanumara and Thompson (1968) did not give any mathematical justifications of (3.19). We have given a justification of (3.19) as an asymptotic expansion as x goes to infinity.

3.2.2. *Tail probability of the maximum of a symmetric multilinear form.* We now present our result on a symmetric multilinear form. The set

$$\widetilde{M}_k = \{\varepsilon \underbrace{h\otimes\cdots\otimes h}_{k} \mid h \in S^{q-1},\ \varepsilon = \pm 1\} \tag{3.20}$$

forms a manifold of dimension $d = q-1$ in S^{p-1} with $p = q^k$. As in the case of the manifold M_k in (3.14), it is easy to check that $\widetilde{M}_k$ is a submanifold of S^{p-1} satisfying Assumption 3.1. The statistics $\widetilde{T}_k$ and $\widetilde{U}_k$ in (1.4) and (1.5) can be written as

$$\widetilde{T}_k = \max_{u\in\tilde{M}_k} u'z, \qquad \widetilde{U}_k = \max_{u\in\tilde{M}_k} u'z/\|z\|,$$

respectively, where z is a p-dimensional column vector distributed as $N(0, I_p)$. Here it is to be noted that the representation $(h\otimes\cdots\otimes h)'z$ is not of minimal dimension. $\widetilde{M}_k$ or its associated cone $\widetilde{K}_k = \bigcup_{c\geq 0} c\widetilde{M}_k$ is degenerate. It is easily proved that

$$\dim \operatorname{lin}(\widetilde{K}_k) = \binom{q+k-1}{k}$$

[see, e.g., Takemura (1993)]. As stated in Remark 3.2 we have to be careful that the $p = q^k$ appearing in Theorem 3.1 is replaced with $p' = \binom{q+k-1}{k}$.

THEOREM 3.3. *The tail probabilities of $\widetilde{T}_k$ in* (1.4) *and $\widetilde{U}_k$ in* (1.5) *for $k \geq 2$ are given by Theorem* 3.1 (*or Corollary* 3.1) *and Lemma* 3.4, *respectively*, *where $d = q-1$, $p = \binom{q+k-1}{k}$, and*:

(i) *The nonzero coefficients w_{d+1-e} are given by*

$$w_{d+1-e} = w_{q-e} = k^{\frac{1}{2}(q-1)}\left(-\frac{k-1}{k}\right)^{e/2}\frac{\Gamma(\frac{1}{2}(q+1))}{\Gamma(\frac{1}{2}(q-e+1))\,(e/2)!},$$

$e = 0, 2, \ldots, [(q-1)/2]\times 2.$

(ii) *The critical radius* $\tilde{\theta}_c$ *is given by*

$$\tilde{\theta}_c = \cos^{-1}\sqrt{\frac{2k-2}{3k-2}}.$$

The proof of Theorem 3.3 is given in Section 5.2.

REMARK 3.4. When q is odd, the Gauss–Bonnet theorem $\sum_{e:\,\mathrm{even}} w_{q-e} = 1$ holds (see Lemma 5.2).

4. Multilinear forms: examples and proofs. In this section we first give numerical examples for Theorem 3.2. The rest of this section is devoted to the proof of Theorem 3.2.

4.1. *Examples.*

4.1.1. *The maximum of a bilinear form* (3×3). Consider the statistic T_2 in (1.2) with $q_1 = q_2 = 3$. Then T_2 is the square root of the largest eigenvalue of the Wishart matrix $W(I_3, 3)$. Then $p = q_1 q_2 = 9$ and $d = q_1 + q_2 - 2 = 4$. The approximate tail probability for T_2 is given by Corollary 3.2 as

$$P(T_2 \geq x) \sim 3\overline{G}_5(x^2) - 4\overline{G}_3(x^2) + 2\overline{G}_1(x^2). \tag{4.1}$$

Since the critical radius is $\theta_c = \pi/4$, the lower bound is

$$Q_L(x) = 3Q_{5,4}(x^2, 1) - 4Q_{3,6}(x^2, 1) + 2Q_{1,8}(x^2, 1). \tag{4.2}$$

Let M_c denote the tube of distance θ_c around M_2. The upper bound is

$$Q_U(x) = Q_L(x) + \overline{G}_9(2x^2)\,(1 - \mathrm{Vol}(M_c)/\Omega_9), \tag{4.3}$$

where

$$\mathrm{Vol}(M_c)/\Omega_9 = 3\overline{B}_{(5/2),2}(1/2) - 4\overline{B}_{(3/2),3}(1/2) + 2\overline{B}_{(1/2),4}(1/2) \doteq 0.990.$$

In Figure 2 the approximate tail probability by (4.1), the lower and upper bounds by (4.2) and (4.3), and the exact tail probability calculated by the Pfaffian method [Section 4.2 of Pillai (1976)] are plotted. The exact value and the upper bound are too close to be distinguished. We can conclude in this case that the approximation formula by asymptotic expansion is sufficiently accurate.

Also recalling that the value of $\mathrm{Vol}(M_c)/\Omega_p$ is the maximum P-value which can be calculated by Lemma 3.4, we can also conclude that Lemma 3.4 provides a practical method for calculating the P-values for U_2^2.

REMARK 4.1. As mentioned in Section 1.3, the original tube formula by Weyl (1939) is represented in terms of the curvature tensor. Sun (1993) pointed out that for up to two terms the tube formula can be written in a relatively simple form by using the scalar curvature. Let $t = (t^i)$ be a local coordinate of the index set I, and let $(g_{ij}(t))$ and $R(t)$ be the metric tensor and the

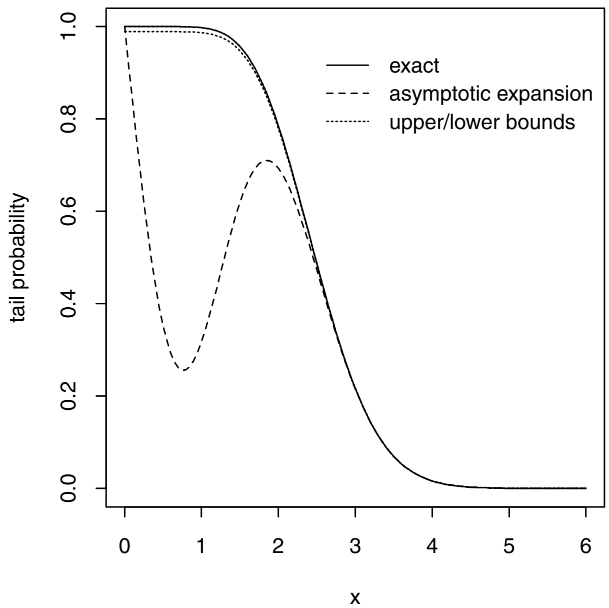

FIG. 2. *The maximum of a bilinear form* (3×3).

scalar curvature at t, respectively. Sun's two-term formula for the maxima $T = \max_{t \in I} Z(t)$ of the Gaussian field is

$$P(T \geq x) \sim \kappa_0 \psi_0(x) + \kappa_2 \psi_2(x), \tag{4.4}$$

where

$$\kappa_0 = \int_I \det(g_{ij}(t))^{1/2} \, dt^1 \cdots dt^d = \mathrm{Vol}(I),$$

$$\kappa_2 = \int_I \Big(\frac{R(t)}{2} - \frac{d(d-1)}{2} \Big) \det(g_{ij}(t))^{1/2} \, dt^1 \cdots dt^d,$$

$$\psi_e(x) = \frac{\Gamma(\frac{1}{2}(d+1-e))}{2^{1+e/2} \pi^{(d+1)/2}} \, \overline{G}_{d+1-e}(x^2).$$

In the following we confirm that the first two coefficients in (4.1) are obtained by (4.4).

Each element of $M_2 = \{h_1 \otimes h_2 \mid h_1, h_2 \in S^{3-1}\}$ is written as

$$(\cos t^1, \sin t^1 \cos t^2, \sin t^1 \sin t^2) \otimes (\cos t^3, \sin t^3 \cos t^4, \sin t^3 \sin t^4),$$

where $0 \leq t^1, t^2, t^3 \leq \pi$, $0 \leq t^4 < 2\pi$. The metric tensor is $(g_{ij}) = \mathrm{diag}(1, \sin^2 t^1, 1, \sin^2 t^3)$. Let g^{kl} be the (k, l)th element of the inverse matrix $(g_{ij})^{-1}$. The nonzero elements of the affine connections defined by

$$\Gamma_{ij}^k = \tfrac{1}{2} \sum_{l=1}^{d} g^{kl} \, (\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij}), \quad \partial_i = \partial / \partial t^i,$$

are $\Gamma^1_{22} = -\cos t^1 \sin t^1$, $\Gamma^2_{21} = \Gamma^2_{12} = \cot t^1$, $\Gamma^3_{44} = -\cos t^3 \sin t^3$, $\Gamma^4_{43} = \Gamma^4_{34} = \cot t^3$. The curvature tensor R_{ijkl} and the scalar curvature R are defined by

$$R_{ijkl} = \sum_{m=1}^{d} g_{im} R^m_{jkl}, \qquad R^l_{ijk} = \partial_j \Gamma^l_{ik} - \partial_k \Gamma^l_{ij} + \sum_{m=1}^{d} (\Gamma^m_{ik} \Gamma^l_{mj} - \Gamma^m_{ij} \Gamma^l_{mk}),$$

and $R = \sum_{ijkl=1}^{d} g^{ik} g^{jl} R_{ijkl}$. The nonzero elements of the curvature tensor are only $R_{2121} = R_{1212} = -R_{2112} = -R_{1212} = \sin^2 t^1$, $R_{4343} = R_{3434} = -R_{4334} = -R_{3443} = \sin^2 t^3$, and hence the scalar curvature is $R = 4$. Since the dimension is $d = 4$, we have $\kappa_0 = 8\pi^2$, $\kappa_2 = -32\pi^2$ and

$$\kappa_0 \psi_0(x) + \kappa_2 \psi_2(x) = 3\overline{G}_5(x^2) - 4\overline{G}_3(x^2),$$

as expected.

4.1.2. *The maximum of a trilinear form* $(2 \times 2 \times 2)$. As another example we consider the statistic T_3 in (1.2) with $q_1 = q_2 = q_3 = 2$. Then $p = \prod_i q_i = 8$ and $d = \sum_i (q_i - 1) = 3$. Since $n_3(1,1,1;0) = 1$ and $n_3(1,1,1;2/2) = 3$, we have $w_4 = \pi$, $w_2 = -3\pi/2$, and the other w's are 0. Therefore we have

$$P(T_3 \geq x) \sim \pi \overline{G}_4(x^2) - (3\pi/2)\overline{G}_2(x^2). \tag{4.5}$$

By Theorem 3.2 the critical radius θ_c of M_3 in (3.14) is given by $\cos^2 \theta_c = 4/7$. Then $\tan^2 \theta_c = 3/4$ and the lower and upper bounds for $P(T_3 \geq x)$ are given by

$$Q_L(x) = \pi Q_{4,4}(x^2, 3/4) - (3\pi/2) Q_{2,6}(x^2, 3/4) \tag{4.6}$$

and

$$Q_U(x) = Q_L(x) + \overline{G}_8(7x^2/4)(1 - \mathrm{Vol}(M_c)/\Omega_8), \tag{4.7}$$

where

$$\mathrm{Vol}(M_c)/\Omega_8 = \pi \overline{B}_{2,2}(4/7) - (3\pi/2)\overline{B}_{1,3}(4/7) \doteq 0.866.$$

These three functions (4.5), (4.6), (4.7) are plotted in Figure 3. In contrast to the Wishart matrix case, the exact distribution of T_3 is not known. Instead, the estimated tail probability by a Monte Carlo simulation with 100,000 replications are plotted there. We see that the asymptotic expansion (4.5) gives a fairly good approximation.

Also $\mathrm{Vol}(M_c)/\Omega_8$ is adequately large and in this case Lemma 3.4 is practical enough for calculating P-values for U_3^2.

4.2. *Proof of Theorem* 3.2. We prove here Theorem 3.2, one of the main theorems of this paper. The proof is divided into three parts. First, the geometric quantities of M_k such as the volume element and the second fundamental form of the manifold M_k are determined (Section 4.2.1). Second, the coefficients

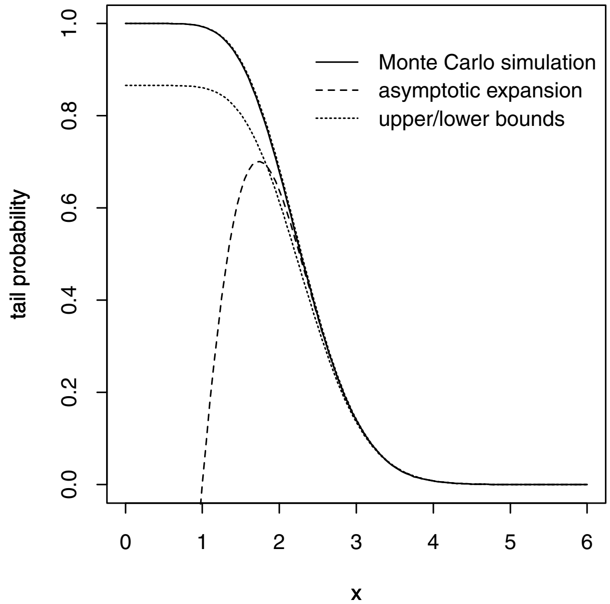

FIG. 3. *The maximum of a trilinear form* ($2 \times 2 \times 2$).

w_{d+1-e} are derived using combinatorial arguments (Section 4.2.2). Finally, the critical radius θ_c of M_k is obtained by virtue of Lemma 3.1 (Section 4.2.3).

4.2.1. *Volume element and second fundamental form.* We introduce a local coordinate system to make calculations simple. Let $t_i = (t_{i1}, \dots, t_{i,q_i-1})'$ be a local coordinate system of S^{q_i-1} so that $h_i \in S^{q_i-1}$ has a representation $h_i = h_i(t_i)$. Then $u = h_1 \otimes \cdots \otimes h_k \in M_k$ has a local representation $u = \phi(t)$, where

$$\phi(t) = h_1(t_1) \otimes \cdots \otimes h_k(t_k)$$

with parameter $t = (t_1', \dots, t_k')'$ of dimension $d = \sum_{i=1}^{k}(q_i - 1)$.

Taking a derivative of $\phi(t)$ with respect to t_{ia}, we have

$$\frac{\partial \phi}{\partial t_{ia}} = h_1 \otimes \cdots \otimes h_{i-1} \otimes \frac{\partial h_i}{\partial t_{ia}} \otimes h_{i+1} \otimes \cdots \otimes h_k.$$

The tangent space $T_u(M_k)$ at $u = \phi(t)$ is spanned by

$$\left\{ \frac{\partial \phi}{\partial t_{ia}} \in R^p \;\middle|\; i = 1, \dots, k,\ a = 1, \dots, q_i - 1 \right\},$$

and $T_u(K_k)$ is spanned by $T_u(M_k)$ and u. The (ia, jb)th element of the metric $G = G(u)$ at u is given by

$$\left(\frac{\partial \phi}{\partial t_{ia}}\right)' \frac{\partial \phi}{\partial t_{jb}} = \delta_{ij} \left(\frac{\partial h_i}{\partial t_{ia}}\right)' \frac{\partial h_i}{\partial t_{ib}} = \delta_{ij} \bar{g}_{i,ab}, \qquad (4.8)$$

where δ_{ij} is the Kronecker delta and

$$\bar{g}_{i,ab} = \left(\frac{\partial h_i}{\partial t_{ia}}\right)' \frac{\partial h_i}{\partial t_{ib}}$$

is the (a, b)th element of the metric $\overline{G}_i$ of S^{q_i-1} at $h_i = h_i(t_i)$. Therefore the metric of M_k is given by $G = \operatorname{diag}(\overline{G}_1, \ldots, \overline{G}_k)$ with $\overline{G}_i = (\bar{g}_{i,ab})$ a $(q_i - 1) \times (q_i - 1)$ matrix. The volume element at u is

$$du = |G|^{1/2} \prod_{i=1}^{k} \prod_{a=1}^{q_i-1} dt_{ia} = \prod_{i=1}^{k} \left\{ |\overline{G}_i|^{1/2} \prod_{a=1}^{q_i-1} dt_{ia} \right\},$$

which is a product of the volume elements of S^{q_i-1}, $i = 1, \ldots, k$.

LEMMA 4.1. *The volume element of M_k at $u = h_1 \otimes \cdots \otimes h_k$ is given by $du = \prod_{i=1}^{k} dS^{q_i-1}$, where dS^{q_i-1} denotes the volume element of S^{q_i-1} at h_i.*

Here we need to be careful about the fact that M_k and $S^{q_1-1} \times \cdots \times S^{q_k-1}$ are not one-to-one. Indeed, $h_1 \otimes \cdots \otimes h_k$ is invariant under an even number of sign changes $h_i \mapsto -h_i$. The multiplicity of the map g_k: $S^{q_1-1} \times \cdots \times S^{q_k-1} \to M_k$ is 2^{k-1}, since the signs of $h_1, \ldots, h_{k-1}$ can be arbitrarily chosen. Noting this fact, we have the following.

COROLLARY 4.1. *The total volume of M_k is*

$$\operatorname{Vol}(M_k) = \int_{M_k} du = \frac{1}{2^{k-1}} \prod_{i=1}^{k} \int_{S^{q_i-1}} dS^{q_i-1} = \frac{1}{2^{k-1}} \prod_{i=1}^{k} \Omega_{q_i}.$$

$S^{q_1-1} \times \cdots \times S^{q_k-1}$ is a 2^{k-1}-fold covering space of the index set M_k. Since $\chi(S^{q-1}) = 2$ if q is odd, 0 if q is even, we have the following.

LEMMA 4.2. *The Euler characteristic of the index set M_k is*

$$\chi(M_k) = \prod_{i=1}^{k} \chi(S^{q_i-1})/2^{k-1} = \begin{cases} 2, & \text{if } q_i\text{'s are all odd,} \\ 0, & \text{otherwise.} \end{cases}$$

Let H_i be a $q_i \times (q_i - 1)$ matrix such that (h_i, H_i) is $q_i \times q_i$ orthogonal. Let

$$\frac{\partial \phi}{\partial t_i} = \left(\frac{\partial \phi}{\partial t_{i1}}, \ldots, \frac{\partial \phi}{\partial t_{i,q_i-1}} \right)$$

be represented as a $p \times (q_i - 1)$ matrix, and let

$$\frac{\partial h_i}{\partial t_i} = \left(\frac{\partial h_i}{\partial t_{i1}}, \ldots, \frac{\partial h_i}{\partial t_{i,q_i-1}} \right)$$

be represented as a $q_i \times (q_i - 1)$ matrix. Then the columns of two $p \times (q_i - 1)$ matrices

$$B_i = h_1 \otimes \cdots \otimes h_{i-1} \otimes H_i \otimes h_{i+1} \otimes \cdots \otimes h_k$$

and

$$\frac{\partial \phi}{\partial t_i} = h_1 \otimes \cdots \otimes h_{i-1} \otimes \frac{\partial h_i}{\partial t_i} \otimes h_{i+1} \otimes \cdots \otimes h_k$$

span the same space, since $h_i'(\partial h_i/\partial t_i) = 0$ and $\operatorname{rank}(\partial h_i/\partial t_i) = q_i - 1$.

Any vector orthogonal to $u = h_1 \otimes \cdots \otimes h_k$ and the column spaces of B_i, $i = 1, \dots, k$, can be written as

$$\begin{aligned} v = {} & (H_1 \otimes H_2 \otimes h_3 \otimes \cdots \otimes h_k)\, e_{12} + (H_1 \otimes h_2 \otimes H_3 \otimes h_4 \otimes \cdots \otimes h_k)\, e_{13} \\ & + \cdots + (h_1 \otimes \cdots \otimes h_{k-2} \otimes H_{k-1} \otimes H_k)\, e_{k-1,\,k} \\ & + (H_1 \otimes H_2 \otimes H_3 \otimes h_4 \otimes \cdots \otimes h_k)\, e_{123} \\ & + \cdots \\ & + (H_1 \otimes H_2 \otimes \cdots \otimes H_k)\, e_{12\cdots k}, \end{aligned} \tag{4.9}$$

where e's are column vectors of appropriate sizes; for example, e_{12} is $(q_1 - 1)(q_2 - 1) \times 1$, e_{123} is $(q_1 - 1)(q_2 - 1)(q_3 - 1) \times 1$, $e_{12\cdots k}$ is $\prod_{i=1}^k (q_i - 1) \times 1$. Since the linear subspace spanned by the set of vectors v in (4.9) is of dimension $\prod_{i=1}^k q_i - \sum_{i=1}^k (q_i - 1) - 1 = p - d - 1$, it coincides with $T_u(K_k)^\perp$.

Now taking a second derivative we have

$$\frac{\partial^2 \phi}{\partial t_{ia} \partial t_{jb}} = \begin{cases} h_1 \otimes \cdots \otimes h_{i-1} \otimes \dfrac{\partial^2 h_i}{\partial t_{ia} \partial t_{jb}} \otimes h_{i+1} \otimes \cdots \otimes h_k, & \text{if } i = j, \\ h_1 \otimes \cdots \otimes h_{i-1} \otimes \dfrac{\partial h_i}{\partial t_{ia}} \otimes h_{i+1} \otimes \cdots \otimes h_{j-1} \otimes \dfrac{\partial h_j}{\partial t_{jb}} \otimes h_{j+1} \otimes \cdots \otimes h_k, & \text{if } i < j. \end{cases}$$

Then for v in (4.9),

$$v' \frac{\partial^2 \phi}{\partial t_{ia} \partial t_{jb}} = \begin{cases} 0, & \text{if } i = j, \\ e_{ij}' \left(H_i' \dfrac{\partial h_i}{\partial t_{ia}} \otimes H_j' \dfrac{\partial h_j}{\partial t_{jb}} \right), & \text{if } i < j. \end{cases}$$

For $i < j$ let E_{ij} be the $(q_i - 1) \times (q_j - 1)$ matrix defined by $\operatorname{vec}(E_{ij}) = e_{ij}$. There exists a $(q_i - 1) \times (q_i - 1)$ nonsingular matrix F_i such that

$$\frac{\partial h_i}{\partial t_i} = H_i F_i.$$

Then the $d \times d$ [$d = \sum_{i=1}^{k}(q_i - 1)$] matrix with (ia, jb)th element $v'(\partial^2\phi/\partial t_{ia}\partial t_{jb})$ is a block matrix with (i, j)th block

$$\begin{cases} O, & \text{if } i = j, \\ F_i' E_{ij} F_j, & \text{if } i < j, \\ F_i' E_{ij}' F_j, & \text{if } i > j, \end{cases}$$

$i, j = 1, \dots, k$.

On the other hand, as we have seen in (4.8), the metric G of M_k can be written as a diagonal block matrix with (i, i)th block $F_i' F_i$, $i = 1, \dots, k$. This implies the following lemma.

LEMMA 4.3. *In an appropriate coordinate system, the second fundamental form of M_k at u with respect to the direction v in* (4.9) *can be written as*

$$H(u, v) = -\begin{pmatrix} O & E_{12} & E_{13} & \cdots & E_{1k} \\ E_{12}' & O & E_{23} & \cdots & E_{2k} \\ E_{13}' & E_{23}' & O & \cdots & E_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ E_{1k}' & E_{2k}' & E_{3k}' & \cdots & O \end{pmatrix}. \tag{4.10}$$

4.2.2. *Evaluation of the coefficient w_{d+1-e}.* We now proceed to the evaluation of the coefficient w_{d+1-e} in (3.7). For fixed $u \in M_k$, we first evaluate the integral

$$\int_{T_u(K_k)^\perp \cap S^{p-1}} \mathrm{tr}_e H(u, v)\, dv, \tag{4.11}$$

where dv is the volume element of $T_u(K_k)^\perp \cap S^{p-1}$, the unit sphere restricted to $T_u(K_k)^\perp$. We introduce a random variable and replace the integration with an expectation.

Let $y \in R^p$ be a singular Gaussian vector distributed as $N(0, P_u^\perp)$, where $P_u^\perp$ is the orthogonal projection matrix onto the linear subspace $T_u(K_k)^\perp$. Then $r = \|y\|$ and $v = y/\|y\|$ are independently distributed. r^2 has the χ^2 distribution with $p-d-1$ degrees of freedom and v has the uniform distribution over $T_u(K_k)^\perp \cap S^{p-1}$. Since $H(u, v)$ is linear in v, we have

$$\begin{aligned} E[\mathrm{tr}_e H(u, y)] &= E[\mathrm{tr}_e H(u, rv)] = E[r^e \mathrm{tr}_e H(u, v)] \\ &= E[r^e] \cdot E[\mathrm{tr}_e H(u, v)] \\ &= E[(\chi^2_{p-d-1})^{e/2}] \cdot \frac{1}{\Omega_{p-d-1}} \int \mathrm{tr}_e H(u, v)\, dv, \end{aligned}$$

where

$$E[(\chi^2_{p-d-1})^{e/2}] = 2^{e/2} \frac{\Gamma(\frac{1}{2}(p-d-1+e))}{\Gamma(\frac{1}{2}(p-d-1))}.$$

Hence we have a representation of the integral (4.11) as

$$\int_{T_u(K_k)^{\perp}\cap S^{p-1}} \operatorname{tr}_e H(u, v)\, dv = \frac{\Omega_{p-d-1}}{E[(\chi^2_{p-d-1})^{e/2}]} \cdot E[\operatorname{tr}_e H(u, y)]. \tag{4.12}$$

Note that the random vector y can be written as $Q_u \bar{y}$, where Q_u is a $p \times (p-d-1)$ matrix such that $Q_u Q_u' = P_u^{\perp}$ and $\bar{y}$ is a $(p-d-1)$-dimensional random vector distributed as $N(0, I_{p-d-1})$.

Now we return to the problem of multilinear forms of degree k. As we saw, $T_u(K_k)^{\perp}$ is spanned by the vectors of the form of v in (4.9). In this parameterization, the squared norm of v in (4.9) is

$$\|v\|^2 = \sum_{1\le i<j\le k} \|e_{ij}\|^2 + \sum_{1\le i<j<l\le k} \|e_{ijl}\|^2 + \cdots + \|e_{12\cdots k}\|^2,$$

which means that elements of the vectors

$$e_{ij}\ (i<j),\ e_{ijl}\ (i<j<l), \ldots,\ e_{12\cdots k}$$

form an orthonormal basis of $T_u(K_k)^{\perp}$. If we suppose that every element of these vectors $e_{ij}, e_{ijl}, \ldots, e_{12\cdots k}$ is an independent random variable distributed as $N(0,1)$, for example, e_{12} is a $(q_1-1)(q_2-1)$-dimensional standard multivariate normal random vector, then v defined in (4.9) has the distribution $N(0, P_u^{\perp})$. Therefore the problem is reduced to evaluating the expectation $E[\operatorname{tr}_e H]$ with $H = H(u,v)$ in (4.10), where each component of E_{ij} $(i<j)$ is independently distributed as $N(0,1)$.

LEMMA 4.4. *Let y be distributed as $N(0, P_u^{\perp})$. Then*

$$E[\operatorname{tr}_e H(u, y)] = \begin{cases} (-1)^{e/2} n_k(q_1-1, \ldots, q_k-1; e/2), & \text{for } e \text{ even},\\ 0, & \text{for } e \text{ odd},\end{cases}$$

where n_k is defined in Definition 3.1.

PROOF. Note first that the generalized trace $\operatorname{tr}_e H$ of H can be written as

$$\operatorname{tr}_e H = \sum_{\substack{A\subset\{1,\ldots,d\}\\ \operatorname{card}(A)=e}} \det H[A],$$

where $H[A]$ with $A = \{1 \le a_1 < \cdots < a_e \le d\}$ denotes the $e\times e$ submatrix of H formed by deleting all but columns and rows of H numbered $a_1, \ldots, a_e$ [Muirhead (1982), Appendix A.7]. The cardinality of A is denoted by $\operatorname{card}(A)$. Consider the termwise expectation

$$E[\det H[A]] = \sum_{\pi\in S(A)} \operatorname{sgn}(\pi)\, E[h_{a_1\pi(a_1)} \cdots h_{a_e\pi(a_e)}], \tag{4.13}$$

where $S(A)$ is the set of permutations of the elements of A.

Since $H = (h_{ab})_{1\le a,\, b\le d}$ is a symmetric random matrix whose diagonal and upper off-diagonal elements are zero mean independent random variables (maybe a constant 0), $E[h_{a_1\pi(a_1)} \cdots h_{a_e\pi(a_e)}] = 0$ unless e is even and $\pi(a) \ne a$, $\pi(\pi(a)) = a$, $\forall a$. In this case $\mathrm{sgn}(\pi) = (-1)^{e/2}$, and by relabeling the indices of a's, nonvanishing terms in (4.13) can be written uniquely as

$$(-1)^{e/2} E[h^2_{a_1a_2} h^2_{a_3a_4} \cdots h^2_{a_{e-1}a_e}]$$

with $a_1 < a_3 < \cdots < a_{m-1}$, $a_1 < a_2, \ldots, a_{e-1} < a_e$. Moreover, $h_{a_{2l-1}a_{2l}} = 0$ iff

$$\exists\, i, \quad \sum_{j=1}^{i-1}(q_j - 1) + 1 \le a_{2l-1} < a_{2l} \le \sum_{j=1}^{i}(q_j - 1).$$

Therefore, for e even we have

$$E[\mathrm{tr}_e\, H(u, y)] = (-1)^{e/2} \sum\nolimits^* E[h^2_{a_1a_2} h^2_{a_3a_4} \cdots h^2_{a_{e-1}a_e}],$$

where the summation $\sum^*$ is taken over all sets of $m = e/2$ pairings (3.16) satisfying (i) and (ii) of Definition 3.1. Since the expectation in the right-hand side is 1, we have proved the lemma. $\square$

Now we proceed to integrate (4.11) with respect to du:

$$\int_{M_k} \left[\int_{T_u(K_k)^\perp \cap S^{p-1}} \mathrm{tr}_e\, H(u, v)\, dv \right] du.$$

As we already saw, the integrand does not depend on u. Therefore, the integration with respect to du over M_k reduces to multiplying by a constant $\int_{M_k} du = \mathrm{Vol}(M_k)$ obtained in Corollary 4.1.

Then from (4.12) the coefficient in (3.7) for M_k is

$$\begin{aligned} w_{d+1-e} &= \frac{1}{\Omega_{d+1-e}\Omega_{p-d-1+e}} \cdot \mathrm{Vol}(M_k) \cdot \frac{\Omega_{p-d-1}}{E[(\chi^2_{p-d-1})^{e/2}]} E[\mathrm{tr}_e\, H(u, y)] \\ &= \frac{\mathrm{Vol}(M_k)}{\Omega_{d+1}} \cdot \frac{\Gamma(\frac{1}{2}(d+1-e))}{2^{e/2}\,\Gamma(\frac{1}{2}(d+1))} E[\mathrm{tr}_e\, H(u, y)]. \end{aligned}$$

Summarizing the above calculations, we obtain the proof of (i) of Theorem 3.2.

4.2.3. *Critical radius.* In this subsection we obtain the critical radius θ_c of the manifold M_k in (3.14) by virtue of Lemma 3.1.

Fix a point $v = h_1 \otimes \cdots \otimes h_k \in M_k$ with $h_i \in S^{q_i-1}$. Let H_i, $i = 1, \ldots, k$, be $q_i \times (q_i - 1)$ matrices such that (h_i, H_i) is $q_i \times q_i$ orthogonal. Let $K_k = \bigcup_{c\ge 0} cM_k$ be the cone associated with M_k. The tangent space $T_v(K_k)$ at v is spanned by $v = h_1 \otimes \cdots \otimes h_k$ and the column spaces of

$$B_i = h_1 \otimes \cdots \otimes h_{i-1} \otimes H_i \otimes h_{i+1} \otimes \cdots \otimes h_k, \quad i = 1, \ldots, k.$$

Then the orthogonal projection matrix onto $T_v(K_k)$ is given by

$$\begin{aligned} P_v &= vv' + \sum_{i=1}^{k} B_i B_i' \\ &= \sum_{i=1}^{k} h_1 h_1' \otimes \cdots \otimes h_{i-1} h_{i-1}' \otimes I_{q_i} \otimes h_{i+1} h_{i+1}' \otimes \cdots \otimes h_k h_k' \\ &\quad - (k-1) h_1 h_1' \otimes \cdots \otimes h_k h_k'. \end{aligned}$$

Let $\tilde{v} = \tilde{h}_1 \otimes \cdots \otimes \tilde{h}_k \in M_k$. Then $\tilde{v}'v = \prod_{i=1}^{k} (\tilde{h}_i' h_i)$ and

$$\tilde{v}' P_v \tilde{v} = \sum_{i=1}^{k} \prod_{j \neq i} (\tilde{h}_j' h_j)^2 - (k-1) \prod_{i=1}^{k} (\tilde{h}_i' h_i)^2.$$

Note that both $\tilde{v}' P_v \tilde{v}$ and $\tilde{v}'v$ depend on $\tilde{v}$ and v through $\tilde{h}_i' h_i (= x_i$, say) which takes values $-1 \le x_i \le 1$. Then by Lemma 3.1,

$$\cot^2 \theta_c = \sup_{\tilde{v}, v \in M} \frac{1 - \tilde{v}' P_v \tilde{v}}{(1 - \tilde{v}'v)^2} = \sup_{-1 < x_i < 1, \forall i} \frac{1 - \sum_i \prod_{j \neq i} x_j^2 + (k-1) \prod_i x_i^2}{(1 - \prod_i x_i)^2}.$$

Here we take the supremum in two steps. First, take the supremum under the restriction that $\prod_i x_i\,(= y,$ say) is fixed. Second, take the supremum with respect to $-1 < y < 1$. By the inequality between the arithmetic and geometric means, we have

$$\sum_{i=1}^{k} \prod_{j \neq i} x_j^2 \ge k \left(\prod_{i=1}^{k} \prod_{j \neq i} x_j^2 \right)^{1/k} = k\,|y|^{2(k-1)/k},$$

where the equality holds if and only if $x_1^2 = \cdots = x_k^2$. Then we have

$$\cot^2 \theta_c = \sup_{-1 < y < 1} \frac{1 - k\,|y|^{2(k-1)/k} + (k-1) y^2}{(1-y)^2}. \tag{4.14}$$

Note that in (4.14) we can restrict y to be nonnegative. Here we give a lemma, whose proof is given in Appendix A.6.

LEMMA 4.5.

$$\sup_{0 \le z < 1} \frac{1 - k z^{2(k-1)} + (k-1) z^{2k}}{(1 - z^k)^2} = \frac{2(k-1)}{k}, \tag{4.15}$$

where the supremum is attained when $z \uparrow 1$.

Then by making a change of variable $y = z^k$ in (4.14), we have by Lemma 4.5 that $\cot^2 \theta_c = 2(k-1)/k$. The proof of (ii) of Theorem 3.2 is complete.

5. Symmetric multilinear forms: examples and proofs. In this section we first give numerical examples for Theorem 3.3. Monte Carlo studies to determine the necessary sample sizes for the asymptotic approximation of the Malkovich–Afifi statistics are also given. The rest of this section is devoted to the proof of Theorem 3.3.

5.1. *Examples: the maxima of symmetric* 3- *and* 4-*linear forms.* Consider the statistics $\widetilde{T}_k$, $k = 3, 4$, with $q = 2$. Then $d = q - 1 = 1$ and $p' = \binom{q+k-1}{k} = k + 1 = 4, 5$. The approximate tail probabilities are given by

$$P(\widetilde{T}_k \geq x) \sim \sqrt{k}\,\overline{G}_2(x^2) = \sqrt{k}\,e^{-x^2/2}. \tag{5.1}$$

In Figure 4, the approximate tail probability for $\widetilde{T}_3$, the estimated tail probability by a Monte Carlo simulation with 100,000 replications, as well as the upper and lower bounds

$$Q_L(x) = \sqrt{3}\,Q_{2,2}(x^2, 4/7), \qquad Q_U(x) \doteq Q_L(x) + (1 - 0.742)\,\overline{G}_4(7x^2/4)$$

are plotted.

Moreover, we examine the convergence speed of the Malkovich–Afifi statistics. Let $x_1, \ldots, x_n$ be n i.i.d. samples from the two-dimensional normal distribution $N(0, I_2)$, and let

$$\widehat{B}_k = \max_{u \neq 0} \left| \frac{(1/n)\sum_{i=1}^n (u'x_i - u'\hat{\mu})^k}{(u'\widehat{\Sigma}u)^{k/2}} - 3\,\delta_{k,4} \right|, \qquad k = 3, 4,$$

be the Malkovich–Afifi statistics, where $\hat{\mu} = (1/n)\sum_{i=1}^n x_i$, $\widehat{\Sigma} = (1/n)\sum_{i=1}^n (x_i - \hat{\mu})(x_i - \hat{\mu})'$. We estimate the type I error rates $P(\widehat{B}_k \geq c_{k,\alpha})$ by Monte Carlo simulations with 50,000 replications, where $c_{k,\alpha}$ is the approximate $100\alpha\%$ critical point of $\sqrt{k!/n}\,\widetilde{T}_k$ based on (5.1). The results are summarized in Table 1. The last row labeled "s.e." indicates the standard error $\sqrt{\alpha(1-\alpha)/50{,}000}$. We see from this table that:

1. The formulas (5.1) give fairly precise critical points for the limiting distributions $\widetilde{T}_k$, $k = 3, 4$ even when $\alpha = 0.25$ (see the rows $n = \infty$).
2. The convergence of $\widehat{B}_3$ is faster than that of $\widehat{B}_4$. Suppose that about 20% inflation or deflation of the type I error are acceptable. Then the required sample sizes n for $\widehat{B}_3$ are $n \sim 50$ ($\alpha \geq 0.025$), $n \sim 200$ ($\alpha = 0.01$), whereas for $\widehat{B}_4$ $n \sim 100$ or 500 ($\alpha \geq 0.05$), $n \sim 5000$ ($\alpha = 0.025,\ 0.01$).

5.2. *Proof of Theorem* 3.3. We give here a proof of Theorem 3.3. The construction of this subsection is the same as that of Section 4.2.

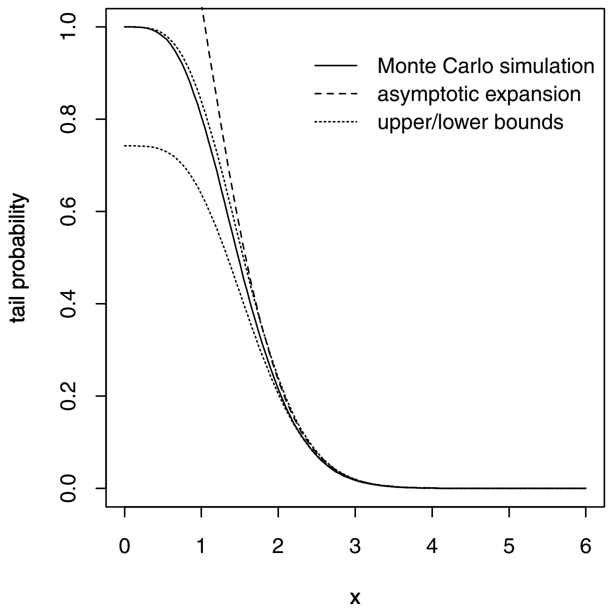

FIG. 4. *The maximum of a symmetric trilinear form* ($2 \times 2 \times 2$).

5.2.1. *Volume element and second fundamental form.* First of all, we introduce a local coordinate system for the sake of convenience of calculation. Let $t = (t_1, \dots, t_{q-1})'$ be a local coordinate system of S^{q-1} so that $h \in S^{q-1}$ has a representation $h = h(t)$. Then $u = \varepsilon h \otimes \cdots \otimes h \in \widetilde{M}_k$, $\varepsilon \in \{1, -1\}$, has a local representation $u = \varphi(t)$ where

$$\varphi(t) = \varepsilon \underbrace{h(t) \otimes \cdots \otimes h(t)}_{k}.$$

Taking a derivative of $\varphi(t)$ with respect to t_i, we have

$$\frac{\partial \varphi}{\partial t_i} = \varepsilon \sum_{l=1}^{k} \underbrace{h \otimes \cdots \otimes h}_{l-1} \otimes \frac{\partial h}{\partial t_i} \otimes \underbrace{h \otimes \cdots \otimes h}_{k-l}.$$

The tangent space $T_u(\widetilde{M}_k)$ at $u = \varphi(t)$ is spanned by

$$\left\{ \frac{\partial \varphi}{\partial t_i} \in R^p \;\middle|\; i = 1, \dots, d \right\}.$$

The tangent space $T_u(\widetilde{K}_k)$ of $\widetilde{K}_k$ is spanned by $T_u(\widetilde{M}_k)$ and u. The (i, j)th element of the metric $G = G(u)$ at u is given by

$$\left(\frac{\partial \varphi}{\partial t_i}\right)' \frac{\partial \varphi}{\partial t_j} = k \left(\frac{\partial h}{\partial t_i}\right)' \frac{\partial h}{\partial t_j} = k \bar{g}_{ij}, \tag{5.2}$$

TABLE 1
Estimation of $P(\widehat{B}_k \geq c_{k,\alpha})$[1]

k	n	$\alpha = 0.25$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
3	10	0.0807	0.0212	0.0074	0.0021	0.0001
	20	0.1444	0.0606	0.0330	0.0175	0.0094
	50	0.1962	0.0848	0.0466	0.0268	0.0139
	100	0.2156	0.0919	0.0498	0.0275	0.0134
	200	0.2298	0.0956	0.0502	0.0278	0.0120
	500	0.2342	0.0965	0.0496	0.0255	0.0105
	1000	0.2369	0.0982	0.0497	0.0247	0.0104
	2000	0.2400	0.0995	0.0500	0.0257	0.0105
	5000	0.2417	0.0994	0.0506	0.0257	0.0104
	10000	0.2411	0.0983	0.0492	0.0249	0.0096
	∞	0.2395	0.0993	0.0498	0.0256	0.0104
4	10	0.0100	0.0021	0.0006	0.0000	0.0000
	20	0.0639	0.0394	0.0284	0.0208	0.0154
	50	0.1139	0.0701	0.0540	0.0415	0.0301
	100	0.1397	0.0841	0.0624	0.0474	0.0338
	200	0.1689	0.0883	0.0626	0.0470	0.0319
	500	0.2067	0.0906	0.0568	0.0401	0.0254
	1000	0.2212	0.0971	0.0585	0.0355	0.0213
	2000	0.2323	0.0963	0.0528	0.0320	0.0161
	5000	0.2438	0.0968	0.0499	0.0282	0.0122
	10000	0.2427	0.0975	0.0499	0.0270	0.0118
	∞	0.2400	0.0996	0.0499	0.0246	0.0097
s.e.		0.0019	0.0013	0.0010	0.0007	0.0004

[1]Monte Carlo simulations with 50,000 replications.

where

$$\bar{g}_{ij} = \left(\frac{\partial h}{\partial t_i}\right)' \frac{\partial h}{\partial t_j}$$

is the (i, j)th element of the metric $\overline{G}$ of S^{q-1} at $h = h(t)$. Therefore we have $G = k\overline{G}$, and hence the volume element at u is

$$du = |G|^{1/2} \prod_{i=1}^{q-1} dt_i = k^{\frac{1}{2}(q-1)} |\overline{G}|^{1/2} \prod_{i=1}^{q-1} dt_i.$$

LEMMA 5.1. *The volume element of* $\widetilde{M}_k$ *at* $u = \varepsilon \underbrace{h \otimes \cdots \otimes h}_{k}$ *is given by* $du = k^{\frac{1}{2}(q-1)} dS^{q-1}$, *where* dS^{q-1} *denotes the volume element of* S^{q-1} *at* h.

When k is even, $\widetilde{M}_k$ consists of two disjoint sets $\widetilde{M}_k^+ = \{\underbrace{h \otimes \cdots \otimes h}_{k} \mid h \in S^{q-1}\}$ and $-\widetilde{M}_k^+$. The multiplicity of the map $\tilde{g}_k$: $S^{q-1} \to \widetilde{M}_k^+$ is 2, and hence

$\mathrm{Vol}(\widetilde{M}_k) = 2\,\mathrm{Vol}(\widetilde{M}_k^+) = 2 \times (1/2)\int_{S^{q-1}} du$. On the other hand when k is odd, $\widetilde{M}_k^+ = \widetilde{M}_k$ and the multiplicity of the map $\tilde{g}_k$ is 1. Therefore the following holds in each case.

COROLLARY 5.1. *The total volume of $\widetilde{M}_k$ is*

$$\mathrm{Vol}(\widetilde{M}_k) = k^{\frac{1}{2}(q-1)} \int_{S^{q-1}} dS^{q-1} = k^{\frac{1}{2}(q-1)}\,\Omega_q.$$

By similar consideration, we get the following.

LEMMA 5.2. *The Euler characteristic of the index set is*

$$\chi(\widetilde{M}_k) = \chi(S^{q-1}) = \begin{cases} 2, & \text{if } q \text{ is odd}, \\ 0, & \text{if } q \text{ is even}. \end{cases}$$

Let H be a $q \times (q-1)$ matrix such that (h, H) is $q \times q$ orthogonal. Using H, any vector $v \in R^p$ orthogonal to $u = \varphi(t)$ can be written as

$$\begin{aligned} v = {} & (H \otimes \underbrace{h \otimes \cdots \otimes h}_{k-1})\, e_1 + (h \otimes H \otimes \underbrace{h \otimes \cdots \otimes h}_{k-2})\, e_2 \\ & + \cdots + (\underbrace{h \otimes \cdots \otimes h}_{k-1} \otimes H)\, e_k \\ & + (H \otimes H \otimes \underbrace{h \otimes \cdots \otimes h}_{k-2})\, e_{12} + (H \otimes h \otimes H \otimes \underbrace{h \otimes \cdots \otimes h}_{k-3})\, e_{13} \\ & + \cdots + (\underbrace{h \otimes \cdots \otimes h}_{k-2} \otimes H \otimes H)\, e_{k-1,k} \\ & + (H \otimes H \otimes H \otimes \underbrace{h \otimes \cdots \otimes h}_{k-3})\, e_{123} \\ & + \cdots + (\underbrace{H \otimes H \otimes \cdots \otimes H}_{k})\, e_{12\cdots k}, \end{aligned} \tag{5.3}$$

where $e_{i_1 \cdots i_l}$ $(1 \le i_1 < \cdots < i_l \le k)$ is a $(q-1)^l \times 1$ column vector.

Suppose that $v \in T_u(\widetilde{K}_k)^{\perp}$. Then it follows that

$$v' \frac{\partial \varphi}{\partial t_i} = \varepsilon \sum_{l=1}^{k} e_l' H' \frac{\partial h}{\partial t_i} = 0.$$

Since the $q \times (q-1)$ matrix

$$\frac{\partial h}{\partial t} = \left(\frac{\partial h}{\partial t_1}, \ldots, \frac{\partial h}{\partial t_{q-1}} \right)$$

is of rank $q-1$ and its columns are orthogonal to h, it holds that $\sum_{l=1}^k e_l = 0$. Since the linear subspace spanned by v in (5.3) with $\sum_{l=1}^k e_l = 0$ is of dimension $q^k - q = p - d - 1$, it coincides with $T_u(\widetilde{K}_k)^\perp$.

Now taking a second derivative we have

$$\begin{aligned}\frac{\partial^2 \varphi}{\partial t_i \partial t_j} &= \varepsilon \sum_{l=1}^k \underbrace{h \otimes \cdots \otimes h}_{l-1} \otimes \frac{\partial^2 h}{\partial t_i \partial t_j} \otimes \underbrace{h \otimes \cdots \otimes h}_{k-l} \\ &\quad + \varepsilon \sum_{1 \le l < m \le k} \underbrace{h \otimes \cdots \otimes h}_{l-1} \otimes \frac{\partial h}{\partial t_i} \otimes \underbrace{h \otimes \cdots \otimes h}_{m-l-1} \otimes \frac{\partial h}{\partial t_j} \otimes \underbrace{h \otimes \cdots \otimes h}_{k-m} \\ &\quad + \varepsilon \sum_{1 \le l < m \le k} \underbrace{h \otimes \cdots \otimes h}_{l-1} \otimes \frac{\partial h}{\partial t_j} \otimes \underbrace{h \otimes \cdots \otimes h}_{m-l-1} \otimes \frac{\partial h}{\partial t_i} \otimes \underbrace{h \otimes \cdots \otimes h}_{k-m}.\end{aligned}$$

Then for v in (5.3) with $\sum_{l=1}^k e_l = 0$,

$$v' \frac{\partial^2 \varphi}{\partial t_i \partial t_j} = \varepsilon \sum_{1 \le l < m \le k} e'_{lm} \left(H' \frac{\partial h}{\partial t_i} \otimes H' \frac{\partial h}{\partial t_j} + H' \frac{\partial h}{\partial t_j} \otimes H' \frac{\partial h}{\partial t_i} \right).$$

For $l < m$ let E_{lm} be the $(q-1) \times (q-1)$ matrix defined by $\mathrm{vec}(E_{lm}) = e_{lm}$. There exists a $(q-1) \times (q-1)$ nonsingular matrix F such that

$$\frac{\partial h}{\partial t} = HF.$$

It follows that $v'(\partial^2 \varphi / \partial t_i \partial t_j)$ is the (i, j)th element of

$$\varepsilon F' \left\{ \sum_{1 \le l < m \le k} (E_{lm} + E'_{lm}) \right\} F.$$

On the other hand, as we have seen in (5.2), the metric G of $\widetilde{M}_k$ can be written as $k F' F$. Therefore we have the following lemma.

LEMMA 5.3. *In an appropriate coordinate system, the second fundamental form of* $\widetilde{M}_k$ *at* u *with respect to the direction* v *in* (5.3) *with* $\sum_{l=1}^k e_l = 0$ *is written as*

$$H(u, v) = -\frac{\varepsilon}{k} \sum_{1 \le l < m \le k} (E_{lm} + E'_{lm}). \tag{5.4}$$

5.2.2. *Derivation of the coefficient* w_{d+1-e}. Now let us proceed to the evaluation of the integral

$$\int_{T_u(\widetilde{K}_k)^\perp \cap S^{p-1}} \mathrm{tr}_e\, H(u, v)\, dv. \tag{5.5}$$

As in Section 4.2 we calculate this integral by taking an expectation.

Let R_k be a $k \times (k-1)$ matrix such that

$$R'_k R_k = I_{k-1} \quad \text{and} \quad 1'_k R_k = 0,$$

where 1_k is a $k \times 1$ vector consisting of 1's. Then the $q \times 1$ vectors $e_1, \dots, e_k$ satisfying $\sum_{l=1}^k e_l = 0$ can be reparameterized as

$$(e_1, \dots, e_k) = (\bar{e}_1, \dots, \bar{e}_{k-1}) R'_k,$$

where $\bar{e}_i$ is $(q-1) \times 1$. Using this parameterization, the squared norm of v defined in (5.3) with $\sum_{l=1}^k e_l = 0$ can be written as

$$\|v\|^2 = \sum_{1 \le i \le k-1} \|\bar{e}_i\|^2 + \sum_{1 \le i < j \le k} \|e_{ij}\|^2 + \sum_{1 \le i < j < l \le k} \|e_{ijl}\|^2 + \cdots + \|e_{12\cdots k}\|^2,$$

which means that elements of the vectors

$$\bar{e}_i, \ e_{ij} \ (i < j), \ e_{ijl} \ (i < j < l), \dots, \ e_{12\cdots k} \tag{5.6}$$

form an orthonormal basis of $T_u(\widetilde{K}_k)^\perp$. Now suppose that every element of these vectors (5.6) is an independent random variable distributed as $N(0,1)$ and take the expectation $E[\mathrm{tr}_e H(u,v)]$ with respect to v. Then the integral (5.5) can be evaluated as

$$\int_{T_u(\widetilde{K}_k)^\perp \cap S^{p-1}} \mathrm{tr}_e H(u,v)\, dv = \frac{\Omega_{p-d-1}}{E[(\chi^2_{p-d-1})^{e/2}]} \cdot E[\mathrm{tr}_e H(u,v)].$$

Rewrite $H(u,v)$ in (5.4) as

$$H(u,v) = \sqrt{\frac{2(k-1)}{k}} C_d,$$

where

$$C_d = -\frac{\varepsilon}{\sqrt{2(k-1)k}} \sum_{1 \le l < m \le k} (E_{lm} + E'_{lm}).$$

We have assumed that each component of E_{lm} $(l < m)$ is independently distributed as $N(0,1)$. $C_d = (c_{ij})$ is a $d \times d$ symmetric random matrix whose diagonal element c_{ii} and upper off-diagonal element c_{ij} $(i < j)$ are distributed independently as $N(0,1)$ and $N(0,1/2)$, respectively.

Consider

$$E[\mathrm{tr}_e H] = \sum_{\substack{A \subset \{1,\dots,d\} \\ \mathrm{card}(A) = e}} E[\det H[A]].$$

For e odd $E[\mathrm{tr}_e H] = 0$ holds because any central moment of odd degrees is 0. Now suppose that e is even. Since $H[A]$ is equivalent in distribution to

$\sqrt{2(k-1)/k}\,C_e$, we have

$$E[\mathrm{tr}_e\, H] = \binom{d}{e}\left\{\frac{2(k-1)}{k}\right\}^{e/2} E[\det\, C_e]. \tag{5.7}$$

Here for $C_e = (c_{ij})$,

$$E[\det\, C_e] = \sum_{\pi\in S_e} \mathrm{sgn}(\pi) E[c_{1\pi(1)}c_{2\pi(2)}\cdots c_{e\pi(e)}]. \tag{5.8}$$

The expectation of the right-hand side of (5.8) above does not vanish if and only if $\pi(i)\neq i$ and $\pi(\pi(i)) = i$ for any i. In this case $\mathrm{sgn}(\pi) = (-1)^{e/2}$, and nonvanishing terms of the right-hand side of (5.8) can be written uniquely in the form

$$(-1)^{e/2} E[c^2_{i_1 i_2} c^2_{i_3 i_4}\cdots c^2_{i_{e-1} i_e}]$$

with $i_1 < i_3 < \cdots < i_{e-1}$, $i_1 < i_2, \dots, i_{e-1} < i_e$. Counting the number of ways of forming $e/2$ pairings from $\{1, 2, \dots, e\}$,

$$\{(i_1, i_2), (i_3, i_4), \dots, (i_{e-1}, i_e) \mid i_1 < i_3 < \cdots < i_{e-1},\ i_1 < i_2, \dots, i_{e-1} < i_e\},$$

we have for e even that

$$E[\det\, C_e] = (-1)^{e/2}\frac{e!}{2^{e/2}(e/2)!}\,(1/2)^{e/2}.$$

Hence from (5.7),

$$E[\mathrm{tr}_e\, H] = \left(-\frac{k-1}{2k}\right)^{e/2}\frac{d!}{(d-e)!\,(e/2)!}.$$

Now it remains to evaluate the integral

$$\int_{\widetilde{M}_k}\left[\int_{T_u(\widetilde{K}_k)^{\perp}\cap S^{p-1}} \mathrm{tr}_e H(u,v)\,dv\right]du.$$

As in the case of a multilinear form in Section 4, the integrand does not depend on u, and the integration with respect to du over $\widetilde{M}_k$ reduces to multiplying by a constant $\int_{\widetilde{M}_k} du = \mathrm{Vol}(\widetilde{M}_k)$ obtained in Corollary 5.1. Then the coefficient in (3.7) is given by

$$\begin{aligned} w_{d+1-e} &= \frac{1}{\Omega_{d+1-e}\Omega_{p-d-1+e}}\cdot \mathrm{Vol}(\widetilde{M}_k)\cdot\frac{\Omega_{p-d-1}}{E[(\chi^2_{p-d-1})^{e/2}]} E[\mathrm{tr}_e\, H(u,v)] \\ &= \frac{\mathrm{Vol}(\widetilde{M}_k)}{\Omega_{d+1}}\cdot\frac{\Gamma(\frac{1}{2}(d+1-e))}{2^{e/2}\,\Gamma(\frac{1}{2}(d+1))} E[\mathrm{tr}_e\, H(u,v)]. \end{aligned}$$

The proof of (i) of Theorem 3.3 is complete.

5.2.3. *Critical radius.* We obtain here the critical radius $\tilde{\theta}_c$ of the manifold $\widetilde{M}_k$ in (3.20) by virtue of Lemma 3.1.

Fix a point $v = \varepsilon h \otimes \cdots \otimes h \in \widetilde{M}_k$ with $h \in S^{q-1}$, $\varepsilon \in \{1, -1\}$. Let H be a $q \times (q-1)$ matrix such that (h, H) is $q \times q$ orthogonal. Let $\widetilde{K}_k = \bigcup_{c \geq 0} c\widetilde{M}_k$ be the cone associated with $\widetilde{M}_k$. Then the tangent space $T_v(\widetilde{K}_k)$ at v is spanned by $v = \varepsilon h \otimes \cdots \otimes h$ and the column spaces of

$$B = \varepsilon \sum_{l=1}^{k} \underbrace{h \otimes \cdots \otimes h}_{l-1} \otimes H \otimes \underbrace{h \otimes \cdots \otimes h}_{k-l}.$$

The orthogonal projection matrix onto $T_v(\widetilde{K}_k)$ is easily shown to be

$$P_v = vv' + \frac{1}{k} BB'.$$

Let $\tilde{v} = \tilde{\varepsilon}\tilde{h} \otimes \cdots \otimes \tilde{h} \in \widetilde{M}_k$, $\tilde{h} \in S^{q-1}$, $\tilde{\varepsilon} = \pm 1$. Then $\tilde{v}'v = (\tilde{\varepsilon}\varepsilon)(\tilde{h}'h)^k$, $B'\tilde{v} = (\tilde{\varepsilon}\varepsilon)k(\tilde{h}'h)^{k-1}H'\tilde{h}$, and

$$\begin{aligned}
\tilde{v}'P_v\tilde{v} &= (\tilde{v}'v)^2 + \frac{1}{k}(B'\tilde{v})'(B'\tilde{v}) \\
&= (\tilde{h}'h)^{2k} + k(\tilde{h}'h)^{2(k-1)}\tilde{h}'HH'\tilde{h} \\
&= k(\tilde{h}'h)^{2(k-1)} - (k-1)(\tilde{h}'h)^{2k}.
\end{aligned}$$

Put $x = \tilde{h}'h$. Then by Lemma 3.1 we have

$$\cot^2\theta_c = \sup_{\tilde{v}, v \in \widetilde{M}_k} \frac{1 - \tilde{v}'P_v\tilde{v}}{(1 - \tilde{v}'v)^2} = \sup_{-1<x<1} \frac{1 - kx^{2(k-1)} + (k-1)x^{2k}}{(1 - x^k)^2} = \frac{2(k-1)}{k}.$$

The last equality follows from Lemma 4.5. The proof of (ii) of Theorem 3.3 is complete.

APPENDIX

A.1. Proof of Theorem 2.1 Let $H_k(x) = e^{x^2/2}(-d/dx)^k e^{-x^2/2}$ be the Hermite polynomial of degree k. The generating function is given by

$$e^{tx - t^2/2} = \sum_{k=0}^{\infty} \frac{t^k}{k!} H_k(x).$$

Let (y, z) be distributed as the two-dimensional normal distribution with zero mean and covariance structure $E[y^2] = E[z^2] = 1$, $E[yz] = \rho$. We claim that

$$E[H_j(y)\, H_k(z)] = \delta_{jk}\, k!\, \rho^k. \tag{A.1}$$

Indeed (A.1) is proved by comparing the coefficients of $s^j t^k$ of the identity

$$\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{s^j}{j!}\frac{t^k}{k!}E[H_j(y)\,H_k(z)]$$
$$= E[e^{sy-s^2/2}e^{tz-t^2/2}] = e^{st\rho} = \sum_{k=0}^{\infty}\frac{(st)^k}{k!}\rho^k.$$

From the i.i.d. sequence $x_1, \dots, x_n \in R^q$, define an empirical field in $C(S^{q-1})$ by

$$\widehat{Z}_k(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} H_k(u'x_i).$$

Then (A.1) implies immediately that the finite-dimensional distributions of $\widehat{Z}_k(\cdot)$ converge to the corresponding finite-dimensional distributions of $Z_k(\cdot)$. Moreover, we can prove the convergence in distribution in the sense of $C(S^{q-1})$ by applying Corollary 7.17 of Araujo and Giné (1980) as in Theorem 2.1 of Baringhaus and Henze (1991).

Thus, to complete the proof, it is sufficient to show that

$$\sup_{u\in S^{q-1}} |\sqrt{n}\widehat{K}_k(u)/(u'\widehat{\Sigma}u)^{k/2} - \widehat{Z}_k(u)| = o_p(1).$$

Since $\widehat{Z}_k(u) = O_p(1)$ and $(u'\widehat{\Sigma}u)^{-k/2} = 1 + o_p(1)$ uniformly in u, we only have to show that

$$\sup_{u\in S^{q-1}} |\sqrt{n}\widehat{K}_k(u) - \widehat{Z}_k(u)| = o_p(1).$$

We will prove this by using the generating function again.

Note first that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e^{t(u'x_i)-t^2/2} = \sum_{k\ge 0}\frac{t^k}{k!}\widehat{Z}_k(u).$$

The left-hand side of the above expression is rewritten as

$$\begin{aligned} &\sqrt{n}e^{-t^2/2}(1/n)\sum_{i=1}^{n} e^{t(u'x_i)} \\ &\quad = \sqrt{n}\exp\left((u'\hat{\mu})t + (u'\widehat{\Sigma}u - 1)\frac{t^2}{2} + \sum_{k\ge 3}\frac{t^k}{k!}\widehat{K}_k(u)\right) \end{aligned} \tag{A.2}$$

with $\hat{\mu} = (1/n)\sum_{i=1}^{n} x_i$ and $\widehat{\Sigma} = (1/n)\sum_{i=1}^{n}(x_i - \hat{\mu})(x_i - \hat{\mu})'$, because $(1/n)\times \sum_{i=1}^{n} e^{t(u'x_i)}$ is the empirical moment generating function of $u'x_i$, $i = 1, \dots, n$. By expanding (A.2) and comparing the coefficients of t^k, we see that

$$\widehat{Z}_k(u) = \sqrt{n}\widehat{K}_k(u) + \sqrt{n}R_k(u), \tag{A.3}$$

where $R_k(u)$ is a finite summation of the products of at least two of $u'\hat{\mu}$, $u'\widehat{\Sigma}u - 1$, or $\widehat{K}_j(u)$, $3 \leq j < k$. [The relation (A.3) is just a well-known relation between moments and cumulants.] Noting that $\widehat{Z}_k(u) = O_p(1)$ and $u'\hat{\mu} = O_p(n^{-1/2})$, $u'\widehat{\Sigma}u = 1 + O_p(n^{-1/2})$ uniformly in u, we can prove by mathematical induction that $\widehat{K}_k(u) = O_p(n^{-1/2})$ and $\widehat{Z}_k(u) - \sqrt{n}\widehat{K}_k(u) = O_p(n^{-1/2})$ uniformly in u. The proof is complete. □

A.2. Critical radius and local radius of curvature. Here we investigate the relation between the global critical radius and the local radius of curvature. In Section 3.1 we considered the tube of $M \subset S^{p-1}$ with respect to the geodesic distance of S^{p-1}. For clarity and completeness of argument we first consider the tube in R^p with respect to the ordinary Euclidean distance. It will be shown that geodesic curvature of M is closely related to the curvature of the cone $K = \bigcup_{c \geq 0} cM$.

Let N be a compact C^2-submanifold without boundary of dimension d in R^p. The tube around N with radius ρ is defined as

$$N_\rho = \{y \mid \|y - y_N\| < \rho\}, \tag{A.4}$$

where y_N is the projection of y onto N. As in Section 3.1, for $x \in N$ we define

$$C_\rho(x) = \{x + y \mid y \in T_x(N)^\perp,\ \|y\| < \rho\},$$

where $T_x(N)^\perp$ denotes the orthogonal complement of the tangent space of N at x. Then $N_\rho = \bigcup_{x \in N} C_\rho(x)$. It is said that N_ρ does not have self-overlap if $C_\rho(x)$, $x \in N$, are disjoint. The critical radius ρ_c of N is defined as

$$\rho_c = \sup\{\rho \mid N_\rho \text{ does not have self-overlap}\}.$$

Note that if $N \subset S^{p-1}$ then $N_\rho \cap S^{p-1}$ is a tube of N with respect to the geodesic distance of S^{p-1}. The problem is that N_ρ may have self-overlap in R^p even if $N_\rho \cap S^{p-1}$ does not have self-overlap in S^{p-1}. For this reason we make distinction between tube with respect to Euclidean distance and tube with respect to the geodesic distance on S^{p-1}.

The following lemma [Proposition 4.1 of Johansen and Johnstone (1990)] holds for the case $\dim N = d > 1$. We omit the proof for the same reason as given for Lemma 3.1.

LEMMA A.1. *The critical radius ρ_c of N is given by*

$$\rho_c = \inf_{x,\, y \in N} \frac{\|x - y\|^2}{2\|P_y^\perp(x - y)\|}, \tag{A.5}$$

where $P_y^\perp$ is the orthogonal projection onto the orthogonal complement of the tangent space $T_y(N)$ of N at y.

Here we discuss the property of

$$h(x, y) = \frac{2\|P_y^{\perp}(x - y)\|}{\|x - y\|^2} \tag{A.6}$$

appearing in (A.5). Since $P_y^{\perp}$ is continuous in y, $h(x, y)$ is continuous on $\{(x, y) \in N \times N \mid x \neq y\}$. Then we investigate the behavior of $h(x, y)$ as $\|x - y\| \to 0$. Since we are considering local property of N we can take d-dimensional local coordinates $t = (t^1, \ldots, t^d)$ and express x, y in terms of t. For the sake of convenience we use the Einstein convention of indices.

Write $y = \phi(t)$ and $x = \phi(t + dt)$. Then

$$\|x - y\| = \|\phi(t + dt) - \phi(t)\|^2 = g_{ij} dt^i dt^j + o(\|dt\|^2),$$

where

$$g_{ij} = \left(\frac{\partial \phi}{\partial t^i}\right)' \frac{\partial \phi}{\partial t^j}, \quad i, j = 1, \ldots, d,$$

are the elements of the first fundamental form at $y = \phi(t)$. On the other hand,

$$\begin{aligned} P_y^{\perp}(\phi(t + dt) - \phi(t)) &= P_y^{\perp} \frac{\partial \phi}{\partial t^i} dt^i + \frac{1}{2} P_y^{\perp} \frac{\partial^2 \phi}{\partial t^i \partial t^j} \, dt^i dt^j + o(\|dt\|^2) \\ &= \frac{1}{2} P_y^{\perp} \frac{\partial^2 \phi}{\partial t^i \partial t^j} \, dt^i \, dt^j + o(\|dt\|^2) \end{aligned}$$

and

$$2\|P_y^{\perp}(\phi(t + dt) - \phi(t))\| = \left\| P_y^{\perp} \frac{\partial^2 \phi}{\partial t^i \partial t^j} \, dt^i \, dt^j \right\| + o(\|dt\|^2).$$

Let

$$w^* \propto -P_y^{\perp} \frac{\partial^2 \phi}{\partial t^i \partial t^j} dt^i dt^j$$

such that $\|w^*\| = 1$. (If the right-hand side is the zero vector, let $w^* = 0$.) Then

$$2\|P_y^{\perp}(\phi(t + dt) - \phi(t))\| = H_{ij}(w^*) \, dt^i \, dt^j + o(\|dt\|^2),$$

where

$$H_{ij}(w) = -w' \frac{\partial^2 \phi}{\partial t^i \partial t^j}, \qquad i, j = 1, \ldots, d.$$

Therefore we have

$$h(x, y) = \frac{H_{ij}(w^*) \, dt^i \, dt^j}{g_{ij} dt^i dt^j} + o(\|dt\|^2).$$

The $d \times d$ matrix with (i, j)th element $H_i^j(w) = H_{ik}(w) g^{kj}$ is called the second fundamental form of N at y with respect to the direction w. The eigenvalues of the second fundamental form are called the principal curvatures and their associated eigenvectors are called principal directions. Note that w^* depends on dt through the direction $dt/\|dt\|$.

Fix dt. Then

$$\begin{aligned} H_{ij}(w^*)dt^i dt^j &= -w^{*\prime}\frac{\partial^2\phi}{\partial t^i\partial t^j}\,dt^i\,dt^j = -w^{*\prime}P_y^{\perp}\frac{\partial^2\phi}{\partial t^i\partial t^j}\,dt^i\,dt^j \\ &= \max_{\|w\|=1}\Big(-w' P_y^{\perp}\frac{\partial^2\phi}{\partial t^i\partial t^j}\,dt^i\,dt^j\Big) \\ &= \max_{w\in T_y(N)^{\perp},\ \|w\|=1}\Big(-w'\frac{\partial^2\phi}{\partial t^i\partial t^j}\,dt^i\,dt^j\Big) \\ &= \max_{w\in T_y(N)^{\perp}\cap S^{p-1}} H_{ij}(w)\,dt^i\,dt^j. \end{aligned}$$

Taking the maximum with respect to the direction $dt/\|dt\|$ we have

$$\begin{aligned} \limsup_{x\to y} h(x, y) &= \max_{\|dt\|=1}\frac{H_{ij}(w^*)\,dt^i\,dt^j}{g_{ij}\,dt^i\,dt^j} \\ &= \max_{w\in T_y(N)^{\perp}\cap S^{p-1}}\ \max_{\|dt\|=1}\frac{H_{ij}(w)\,dt^i\,dt^j}{g_{ij}\,dt^i\,dt^j} \\ &= \max_{w\in T_y(N)^{\perp}\cap S^{p-1}} |\lambda_{\max}(w)|, \end{aligned}$$

where $|\lambda_{\max}(w)|$ denotes the principal curvature having the largest absolute value. $1/|\lambda_{\max}(w)|$ is the local radius of curvature at y with respect to the direction $\pm w$.

Write

$$h(y, y) = \limsup_{x\to y} h(x, y)$$

so that $h(x, y)$ is defined and finite for all $(x, y)\in N\times N$. By continuity of the radius of curvature it is easy to see that as $x, z\to y$,

$$h(y, y) = \limsup_{x, z\to y} h(x, z).$$

Now by a simple compactness argument h attains a finite maximum over $N\times N$. To prove this let (x_i, y_i), $i = 1, 2, \ldots$, be a sequence of points of $N\times N$ such that $h(x_i, y_i)\uparrow \bar h = \sup_{x, y\in N} h(x, y)$. By compactness we can assume without loss of generality that $(x_i, y_i)\to(x_0, y_0)$. If $x_0\neq y_0$ then $h(x_0, y_0) = \bar h$ by continuity. If $x_0 = y_0$ then $h(x_0, y_0) = \limsup_{(x, y)\to(x_0, y_0)} h(x, y)\geq \lim_{i\to\infty} h(x_i, y_i) = \bar h$. However, obviously $h(x_0, y_0)\leq \bar h$. This proves that h attains a finite maximum over $N\times N$, and hence the critical radius ρ_c is positive under our assumptions.

So far we have considered the tube with respect to the Euclidean distance. We proceed to discuss the tube in the unit sphere S^{p-1} with respect to the geodesic distance. $h(u, v)$ in (3.4) can be written as

$$h(u, v) = \frac{\sqrt{1-u'P_v u}}{1-u'v} = \frac{2\|P_v^{\perp}(u-v)\|}{\|u-v\|^2},$$

which is identical to $h(x, y)$ in (A.6) with N replaced with K except that u is restricted to $M \subset S^{p-1}$. However, as $u \to v$, $(u-v)/\|u-v\|$ becomes orthogonal to v. On the other hand, since K is a cone, one of the principal directions of K at v is v itself and the other principal directions are orthogonal to v. Therefore the calculation involving the second fundamental form of M at $v \in M$ can be replaced with the calculation of the second fundamental form of K at $v \in K$. In particular, $h(v, v) = \limsup_{u \to v} h(u, v)$ is similarly defined and $h(u, v)$ attains a finite maximum over $M \times M$. This proves the claims of Remark 3.1. □

A.3. Proof of Lemmas 3.3, 3.5 and Theorem 3.1. Let $z \in R^p$ be distributed as $N(0, I_p)$, and let $r = \|z_K\|$, $s = \|z - z_K\|$. By (3.6),

$$
\begin{aligned}
P(z \in K_\theta) &= P(s < r \tan \theta) \\
&= \frac{1}{(2\pi)^{p/2}} \sum_{e=0}^{d} \iint_{0 \le s < r \tan \theta} e^{-(r^2+s^2)/2} r^{d-e} s^{p-d-2+e} \, dr ds \\
&\qquad \times \int_M \left[\int_{T_u(K)^\perp \cap S^{p-1}} \mathrm{tr}_e \, H(u, v) \, dv \right] du.
\end{aligned}
$$

By a simple change of variables we obtain

$$
\begin{aligned}
&\iint_{0 \le s < r \tan \theta} e^{-(r^2+s^2)/2} r^{d-e} s^{p-d-2+e} \, dr \, ds \\
&\quad = \overline{B}_{\frac{1}{2}(d+1-e), \frac{1}{2}(p-d-1+e)}(\cos^2 \theta) \cdot 2^{p/2-2} \, \Gamma\!\left(\frac{d+1-e}{2}\right) \Gamma\!\left(\frac{p-d-1+e}{2}\right).
\end{aligned}
$$

Note that

$$
\int_{T_u(K)^\perp \cap S^{p-1}} \mathrm{tr}_e \, H(u, v) \, dv = 0
$$

if e is odd, since $\mathrm{tr}_e \, H(u, v)$ is an odd degree polynomial in the elements of v. This proves Lemma 3.3.

Now we proceed to the proof of Theorem 3.1.

$$
P(T \ge a) = P(T \ge a, \; z \in K_{\theta_c}) + P(T \ge a, \; z \notin K_{\theta_c}).
$$

We bound the second term on the right-hand side from above. Note that the projection z_K always exists and we can write

$$
z = r \frac{z_K}{\|z_K\|} + s \frac{z - z_K}{\|z - z_K\|}
$$

and $z \in K_{\theta_c}$ if and only if

$$
s < r \tan \theta_c.
$$

Since $r = \max(T, 0)$, we have for $z \notin K_{\theta_c}$ and $T \ge 0$,

$$
\|z\|^2 = r^2 + s^2 \ge r^2(1 + \tan^2 \theta_c) \ge T^2(1 + \tan^2 \theta_c).
$$

Therefore for $a > 0$,

$$P(T \geq a,\ z \notin K_{\theta_c}) \leq P(\|z\|^2 \geq a^2(1+\tan^2\ \theta_c),\ z \notin K_{\theta_c})$$
$$= \overline{G}_p(a^2(1+\tan^2\theta_c))\, P(z \notin K_{\theta_c})$$

and

$$P(T \geq a) \leq P(T \geq a, z \in K_{\theta_c}) + \overline{G}_p(a^2(1+\tan^2\theta_c))\, P(z \notin K_{\theta_c}).$$

Furthermore,

$$P(T \geq a) \geq P(T \geq a,\ z \in K_{\theta_c}).$$

Therefore it remains to show that $P(T \geq a,\ z \in K_{\theta_c})$ for $a > 0$ can be written as $Q_L(a)$ of (3.11). Now

$$P(T \geq a,\ z \in K_{\theta_c}) = \frac{1}{(2\pi)^{p/2}} \sum_{\substack{e=0\\ e:\text{even}}}^{d} \iint_{\substack{a \leq r < \infty \\ 0 \leq s < r\tan\theta_c}} e^{-(r^2+s^2)/2} r^{d-e} s^{p-d-2+e} dr\, ds$$
$$\times \int_M \left[\int_{T_u(K)^\perp \cap S^{p-1}} \mathrm{tr}_e\, H(u,v)\, dv \right] du.$$

Integrating the right-hand side with respect to s first we see that $P(T \geq a,\ z \in K_{\theta_c}) = Q_L(a)$. This proves the theorem.

Finally we prove Lemma 3.5. Consider a tube M_ρ in R^p around M with respect to the Euclidean distance as defined in (A.4). Then the p-dimensional volume $\mathrm{Vol}(M_\rho)$ is a polynomial in ρ of degree p unless M_ρ has self-overlap. Moreover, the Gauss–Bonnet theorem states that the coefficient of $\mathrm{Vol}(M_\rho)$ of the highest degree p is $\omega_p\, \chi(M)$, where $\omega_p = \Omega_p/p$ is the volume of the unit ball in R^p [e.g., Theorem 5.9 of Gray (1990)]. On the other hand, using the coordinate system (r, u, s, v), the volume of M_ρ is evaluated as

$$\mathrm{Vol}(M_\rho) = \int_{(r-1)^2+s^2<\rho^2,\ s\geq 0} dz$$
$$= \sum_{e=0}^{d} \iint_{(r-1)^2+s^2<\rho^2,\ s\geq 0} r^{d-e} s^{p-d-2+e} dr\, ds$$
$$\times \int_M \left[\int_{T_u(K)^\perp \cap S^{p-1}} \mathrm{tr}_e H(u,v)\, dv \right] du$$
$$= 2\omega_p\, \rho^p \sum_{\substack{e=0\\ e:\text{even}}}^{d} w_{d+1-e} + (\text{terms of lower degrees in } \rho).$$

The proof is complete. □

A.4. Recurrence formula for n_k of Definition 3.1. $n_k(d_1, \ldots, d_k; m)$ of Definition 3.1 can be easily calculated by the recurrence formula in Lemma A.2. Since $n_k(d_1, \ldots, d_k; m)$ is symmetric in $d_1, \ldots, d_k$, we can restrict our attention to the case when $d_1 \geq \cdots \geq d_k$.

LEMMA A.2. *For $k \geq 2$, $d_1 \geq \cdots \geq d_k \geq 0$, and $m \geq 0$, it holds*

$$n_k(d_1, \ldots, d_k; m) = \begin{cases} 1, & \textit{if } m = 0, \\ 0, & \textit{if } m > 0,\ d_k = 0,\ k = 2, \\ n_{k-1}(d_1, \ldots, d_{k-1}; m), & \textit{if } m > 0,\ d_k = 0,\ k \geq 3, \\ n_k(d_1 - 1, d_2, \ldots, d_k; m) & \\ \quad + \sum_{j=2}^{k} d_j\, n_k(d_1 - 1, d_2, \ldots, d_{j-1}, & \\ \qquad d_j - 1, d_{j+1}, \ldots, d_k; m - 1), & \textit{otherwise.} \end{cases}$$

Here in the last expression the arguments of n_k should be reordered so that $d_1 \geq d_2 \geq \cdots \geq d_k \geq 0$ (if necessary). For example, if $d_2 > d_1 - 1 \geq d_3$, $n_k(d_1 - 1, d_2, \ldots, d_k; m)$ should be replaced with $n_k(d_2, d_1 - 1, d_3, \ldots, d_k; m)$.

PROOF. Consider the first element '1' of $A_1 = \{1, \ldots, d_1\}$. Among the $n_k(d_1, \ldots, d_k; m)$ possible m pairings, there are $n_k(d_1 - 1, d_2, \ldots, d_k; m)$ ways where '1' does not appear in the m pairings; while there are

$$d_j \times n_k(d_1 - 1, d_2, \ldots, d_{j-1}, d_j - 1, d_{j+1}, \ldots, d_k; m - 1)$$

ways where "1" makes a pairing with one element of A_j. Then the last equation of the recurrence formula follows when $m \geq 1$ and $d_k \geq 1$. The other three equations are obvious boundary conditions for the recursion. □

A.5. Proof of (3.18). In order to prove (3.18) we use the following.

LEMMA A.3. *For a real number a such that $a \neq 0, -1, -2, \ldots$, and a nonnegative integer $n = 0, 1, 2, \ldots$, define*

$$I_n(a) = \sum_{k=0}^{n} (-1)^{n-k}\, 2^k \frac{\Gamma(a/2 + k)}{\Gamma(a + k)} \binom{n}{k}$$

and

$$J_n(a) = \sum_{k=0}^{n} (-1)^{n-k}\, 2^k \frac{\Gamma(a/2 + k + 1)}{\Gamma(a + k + 1)} \binom{n}{k}.$$

Then

$$I_n(a) = \begin{cases} c_n(a), & \textit{if } n \textit{ is even}, \\ 0, & \textit{if } n \textit{ is odd}, \end{cases} \tag{A.7}$$

and

$$J_n(a) = \begin{cases} \frac{1}{2} c_n(a), & \textit{if } n \textit{ is even}, \\ \frac{1}{2} c_{n+1}(a) & \textit{if } n \textit{ is odd}, \end{cases} \tag{A.8}$$

where

$$c_n(a) = 2^n \frac{\Gamma((n+1)/2)\,\Gamma((n+a)/2)}{\sqrt{\pi}\,\Gamma(n+a)} = \frac{\Gamma((n+1)/2)}{2^{a-1}\,\Gamma((n+a+1)/2)}.$$

PROOF. We use induction on n. The claims (A.7) and (A.8) are easily checked for $n = 0, 1$. Assume that they are true for $n-1$ and n.

Making use of the identity

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1},$$

we have

$$\begin{aligned} I_{n+1}(a) &= \sum_{k=0}^{n+1} (-1)^{n+1-k}\, 2^k \frac{\Gamma(a/2+k)}{\Gamma(a+k)} \left\{ \binom{n}{k} + \binom{n}{k-1} \right\} \\ &= -I_n(a) + 2J_n(a) \\ &= \begin{cases} 0, & \text{if } n \text{ is even,} \\ c_{n+1}(a), & \text{if } n \text{ is odd.} \end{cases} \end{aligned} \tag{A.9}$$

Similarly we have

$$J_{n+1}(a) = -J_n(a) + 2\sum_{k=0}^{n} (-1)^{n-k}\, 2^k \frac{\Gamma(a/2+k+2)}{\Gamma(a+k+2)} \binom{n}{k}.$$

Noting that

$$\frac{\Gamma(a/2+k+2)}{\Gamma(a+k+2)} \binom{n}{k} = \frac{\Gamma((a+2)/2+k)}{\Gamma((a+2)+k)} \left\{ \frac{a+2}{2} \binom{n}{k} + n\binom{n-1}{k-1} \right\},$$

we have

$$\begin{aligned} J_{n+1}(a) &= -J_n(a) + (a+2)\, I_n(a+2) + 4n\, J_{n-1}(a+2) \\ &= \begin{cases} \frac{1}{2}\, c_{n+2}(a), & \text{if } n \text{ is even,} \\ \frac{1}{2}\, c_{n+1}(a), & \text{if } n \text{ is odd.} \end{cases} \end{aligned} \tag{A.10}$$

(A.9) and (A.10) imply that (A.7) and (A.8) hold for $n \geq 2$. The proof is complete. □

The relation (3.18) is equivalent to (A.7) with $n = q-1$, $k = q-1-j$ and $a = \nu - q + 1$.

A.6. Proof of Lemma 4.5. Let $f(z) = 1 - kz^{2(k-1)} + (k-1)z^{2k}$ and $g(z) = (1-z^k)^2$ be the numerator and denominator of the argument of the supremum in (4.15). When $k = 2$, $f(z) \equiv g(z)$ and the statement holds trivially. Consider the case $k \geq 3$. We claim that

$$\frac{d}{dz}\left(\frac{f(z)}{g(z)}\right) > 0 \text{ for } 0 < z < 1. \tag{A.11}$$

In fact, simple calculation yields that

$$\frac{d}{dz}\left(\frac{f(z)}{g(z)}\right) = \frac{2k(1-z^k)z^{k-1}}{g(z)^2} \cdot h(z),$$

where

$$\begin{aligned} h(z) &= 1-(k-1)z^{k-2}+(k-1)z^k - z^{2k-2} \\ &= (1-z^2)\{1+z^2+\cdots+(z^2)^{k-2}-(k-1)z^{k-2}\}. \end{aligned}$$

By the convexity of the map $\xi \mapsto (z^2)^\xi$, we have

$$\frac{1+z^2+\cdots+(z^2)^{k-2}}{k-1} \geq (z^2)^{(0+1+\cdots+(k-2))/(k-1)} = |z|^{k-2},$$

and the equality holds if and only if $|z|=1$. Therefore $h(z)>0$ for $|z|<1$, which implies (A.11). Therefore we have the supremum in (4.15) as

$$\sup_{0\leq z<1} \frac{f(z)}{g(z)} = \lim_{z\uparrow 1}\frac{f(z)}{g(z)} = \lim_{z\uparrow 1}\frac{(d^2/dz^2)f(z)}{(d^2/dz^2)g(z)} = \frac{2(k-1)}{k}.$$

Acknowledgments. The authors are grateful to the two referees and Associate Editor for constructive comments and suggestions. They also thank Anthony J. Hayter for careful proofreading and Yoshiyuki Ninomiya for his help in preparing Table 1.

REFERENCES

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.

ARAUJO, A. and GINÉ, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley, New York.

BARINGHAUS, L. and HENZE, N. (1991). Limit distributions for measures of multivariate skewness and kurtosis based on projections. *J. Multivariate Anal.* **38** 51–69.

BOIK, R. J. (1990). A likelihood ratio test for three-mode singular values: upper percentiles and an application to three-way ANOVA. *Comput. Statist. Data Anal.* **10** 1–9.

BOIK, R. J. and MARASINGHE, M. G. (1989). Analysis of nonadditive multiway classifications. *J. Amer. Statist. Assoc.* **84** 1059–1064.

DAVIS, A. W. (1972). On the ratios of the individual latent roots to the trace of a Wishart matrix. *J. Multivariate Anal.* **2** 440–443.

FRIEDMAN, J. H. (1987). Exploratory projection pursuit. *J. Amer. Statist. Assoc.* **82** 249–266.

GRAY, A. (1990). *Tubes*. Addison–Wesley, Redwood City, CA.

HANUMARA, R. C. and THOMPSON, JR., W. A. (1968). Percentage points of the extreme roots of a Wishart matrix. *Biometrika* **55** 505–512.

HOTELLING, H. (1939). Tubes and spheres in n-spaces, and a class of statistical problems. *Amer. J. Math.* **61** 440–460.

HUBER, P. J. (1985). Projection pursuit. *Ann. Statist.* **13** 435–475.

JOHANSEN, S. and JOHNSTONE, I. (1990). Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. *Ann. Statist.* **18** 652–684.

JOHNSON, D. E. and Graybill, F. A. (1972). An analysis of a two-way model with interaction and no replication. *J. Amer. Statist. Assoc.* **67** 862–868.

KAWASAKI, H. and Miyakawa, M. (1996). A test of three-factor interaction in a three-way layout without replication. *J. Japanese Society for Quality Control* **26** 97–108 (in Japanese).

KNOWLES, M. and SIEGMUND, D. (1989). On Hotelling's approach to testing for a nonlinear parameter in regression. *Internat. Statist. Rev.* **57** 205–220.

KURIKI, S. and TAKEMURA, A. (2000). Shrinkage estimation towards a closed convex set with a smooth boundary. *J. Multivariate Anal.* **75** 79–111.

LEURGANS, S. and ROSS, R. T. (1992). Multilinear models: applications in spectroscopy (with discussion). *Statist. Sci.* **7** 289–319.

MACHADO, S. G. (1983). Two statistics for testing for multivariate normality. *Biometrika* **70** 713–718.

MALKOVICH, J. F. and Afifi, A. A. (1973). On tests for multivariate normality. *J. Amer. Statist. Assoc.* **68** 176–179.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

PILLAI, K. C. S. (1976). Distributions of the characteristic roots in multivariate analysis I. Null distribution. *Canad. J. Statist. Sec. A and B* **4** 157–184.

SCHUURMANN, F. J., KRISHNAIAH, P. R. and CHATTOPADHYAY, A. K. (1973). On the distribution of the ratios of the extreme roots to the trace of the Wishart matrix. *J. Multivariate Anal.* **3** 445–453.

SUN, J. (1991). Significance levels in exploratory projection pursuit. *Biometrika* **78** 759–769.

SUN, J. (1993). Tail probabilities of the maxima of Gaussian random fields. *Ann. Probab.* **21** 34–71.

TAKEMURA, A. (1993). Maximally orthogonally invariant higher order moments and their application to testing elliptically-contouredness. In *Statistical Science and Data Analysis* (K. Matushita, M. L. Puri and T. Hayakawa, eds.) 225–235. VSP, Utrecht.

TAKEMURA, A. and KURIKI, S. (1997). Weights of $\bar{\chi}^2$ distribution for smooth or piecewise smooth cone alternatives. *Ann. Statist.* **25** 2368–2387.

WEYL, H. (1939). On the volume of tubes. *Amer. J. Math.* **61** 461–472.

INSTITUTE OF STATISTICAL MATHEMATICS
4-6-7 MINAMI-AZABU
MINATO-KU
TOKYO 106-8569
JAPAN
E-MAIL: kuriki@ism.ac.jp

FACULTY OF ECONOMICS
UNIVERSITY OF TOKYO
7-3-1 HONGO
BUNKYO-KU
TOKYO 113-0033
JAPAN
E-MAIL: takemura@e.u-tokyo.ac.jp