Asymptotic distribution of inequality-restricted canonical correlation with application to tests for independence in ordered contingency tables

By Satoshi Kuriki

Institute of Statistical Mathematics, Tokyo, Japan

Running Head: Inequality-restricted canonical correlation

Address for correspondence:
Satoshi Kuriki, Institute of Statistical Mathematics, 4-6-7 Minami-azabu, Minato-ku, Tokyo 106-8569, Japan, Fax: +81-3-5421-8796, E-mail: kuriki@ism.ac.jp

Abstract

For two-way ordered categorical data, correspondence analysis and the RC association model (the row-column-effect association model) with order-restricted scores have been proposed mainly for descriptive purposes. In this paper, tests for independence in two-way ordered contingency tables based on these models are developed in a general framework of inequality-restricted canonical correlation analysis. The limiting null distributions are characterized as the maxima of Gaussian random fields and asymptotic expansions of their tail probabilities are derived by the tube method, an integral geometric approach. Some numerical techniques for fitting order-restricted models are discussed. An example of data analysis is given to demonstrate the practical usefulness of the proposed method.

AMS 2000 subject classifications: Primary 62H17; secondary 60D05, 62H10, 62H20.

Key words: correspondence analysis, Gaussian random field, level probability, RC association model, tube method, Wishart distribution.

1 Correspondence analysis and the RC association model in two-way ordered categorical data

The method of canonical correlation or correspondence analysis (CA) is one of the most popular tools for analyzing two-way contingency tables. Suppose that an $a \times b$ table $\{n_{ij}\}_{1\le i\le a,\,1\le j\le b}$, $\sum_{ij} n_{ij} = n$, is observed from a multinomial population with cell probabilities $\{p_{ij}\}_{1\le i\le a,\,1\le j\le b}$, $p_{ij} > 0$, $\sum_{ij} p_{ij} = 1$. The correspondence analysis in its simplest form fits the empirical frequency $\hat{p}_{ij} = n_{ij}/n$ to the model

$$p_{ij} = p_{i\cdot}p_{\cdot j}(1 + \phi\mu_i\nu_j) \tag{1.1}$$

with the side conditions

$$\textstyle\sum_i p_{i\cdot}\mu_i = \sum_j p_{\cdot j}\nu_j = 0, \qquad \sum_i p_{i\cdot}\mu_i^2 = \sum_j p_{\cdot j}\nu_j^2 = 1. \tag{1.2}$$

The weighted least square method is usually used for fitting with the weights defined by the covariance structure under the independence model $p_{ij} = p_{i\cdot}p_{\cdot j}$ (i.e., $\phi = 0$). Here the dot "$\cdot$" as a subscript means summation with respect to the corresponding subscript.

This method has the advantage that if the row and/or column variables are ordinal then the scores μ_i and ν_j are expected to reflect the levels of the ith row and the jth column, respectively.

The multiplicative model for the same purpose is the RC association model (the row-column-effect association model, RC model) proposed by Goodman [14], [15], [16]:

$$\log p_{ij} = \alpha_i + \beta_j + \phi\mu_i\nu_j. \tag{1.3}$$

To ensure identifiability, the same side conditions (1.2) are imposed. The RC association model can be regarded as a natural extension of the model by Johnson and Graybill [21] for two-way ANOVA in two-way categorical data analysis. In the RC association model the maximum likelihood method is usually used to estimate parameters. Numerical algorithms for maximizing likelihood are well developed (e.g., Goodman [16], Becker [3]).

In this paper we focus on the analysis of two-way contingency tables where the row and/or the column variables are ordinal. To analyze such ordered categorical data, we use the correspondence analysis or the RC association model with the order restrictions of scores

$$\phi \ge 0, \qquad \mu_1 \le \cdots \le \mu_a, \qquad \nu_1 \le \cdots \le \nu_b, \tag{1.4}$$

when both row and column variables are ordinal, or

$$\phi \ge 0, \qquad \nu_1 \le \cdots \le \nu_b, \tag{1.5}$$

when only column variables are ordinal. Note that, in the former case, reversing the order of either row or column categories gives a negatively correlated model.

Intuitively these order restrictions seem natural, because if the scores μ_i, ν_j reflect the actual levels of ordinal variables then the inequalities in (1.4) or (1.5) will be satisfied exactly. Another rationale is that under the model (1.1) with order restrictions in the column scores, $\nu_1 \le \cdots \le \nu_b$, a stochastic order exists between the two conditional probabilities $\{p_{j|i} = p_{ij}/p_{i\cdot}\}_{1\le j\le b}$ and $\{p_{j|i'} = p_{i'j}/p_{i'\cdot}\}_{1\le j\le b}$ for any $i \ne i'$ in the sense that

$$\sum_{j=1}^{l} p_{j|i} \underset{(\le)}{\ge} \sum_{j=1}^{l} p_{j|i'}, \qquad 1 \le \forall\, l \le b-1. \tag{1.6}$$

Similarly under the RC association model (1.3), the conditional probability satisfies the relation that, for any $i \ne i'$,

$$\frac{p_{j|i'}}{p_{j|i}} \underset{(\ge)}{\le} \frac{p_{j+1|i'}}{p_{j+1|i}}, \qquad 1 \le \forall\, j \le b-1, \tag{1.7}$$

which is a partial ordering in the sense of monotone likelihood ratio. In other words, by imposing the order restrictions the models (1.1) and (1.3) give simple models that embody the partial orders (1.6) and (1.7), respectively (also see Gilula and Ritov [13]).

For the above reasons, models with the order restrictions (1.4) or (1.5) have been proposed by many authors. See, for example, Nishisato and Arri [26], Tanaka [36], Saito and Otsu [30], and Ritov and Gilula [29] in the context of the correspondence analysis, and Goodman [16] and Ritov and Gilula [28] in the context of the RC association model. Douglas and Fienberg [8], and Etzioni, *et al.* [10] give excellent surveys of the relevant area.

However, almost all of these studies have treated fitting the model for descriptive purposes. From the viewpoint of statistical inference, there are at least two statistical problems of primary interest: one is testing the null hypothesis $H : \phi = 0$ that the row and column variables are independent, and the other is assessing the goodness of fit of the order restrictions when $\phi \ne 0$. For the latter problem, Ritov and Gilula [28] gave a clear answer. They derived the limiting null distribution of the likelihood ratio criterion for testing goodness of fit as a mixture of χ^2 distributions in the RC association model. In this paper we will tackle the former problem.

According to the method of correspondence analysis, when there are natural orderings in both row and column categories, the estimator of ϕ is given by

$$\begin{aligned}\hat\phi = \max\{\textstyle\sum_{ij} \hat p_{ij}\mu_i\nu_j \;|\; \sum_i \hat p_{i\cdot}\mu_i = \sum_j \hat p_{\cdot j}\nu_j = 0,\ \sum_i \hat p_{i\cdot}\mu_i^2 = \sum_j \hat p_{\cdot j}\nu_j^2 = 1,& \\ \mu_1 \le \cdots \le \mu_a,\ \nu_1 \le \cdots \le \nu_b\}.&\end{aligned} \tag{1.8}$$

If the order restriction was not imposed in the maximization (1.8), it is well known that $\sqrt{n}\,\hat\phi$ under the independence model $H : \phi = 0$ converges in distribution to the square root of the largest eigenvalue of an $(a-1)\times(a-1)$ Wishart random matrix with $(b-1)$ degrees of freedom, $W_{a-1}(b-1, I_{a-1})$ (O'Neill [27], Eaton and Tyler [9]). Haberman [17] proved that in the RC association model (1.3) the likelihood ratio criterion for testing

$H : \phi = 0$ has the same asymptotic distribution as the largest eigenvalue of the Wishart matrix $W_{a-1}(b-1, I_{a-1})$ under the null hypothesis.

In contrast to these cases, the null distribution of $\hat{\phi}$ under the order restrictions was completely unknown. Hirotsu [18] suggested the use of $\hat{\phi}$ as a test statistic for testing independence, but he pointed out difficulties in handling its distribution. In this paper we first show that the asymptotic distribution of $\hat{\phi}$ in (1.8) is characterized as a distribution of the maximum of a certain Gaussian random field. Recently an integral-geometric approach called the tube method has been developed for deriving the distribution of the maxima of Gaussian random fields (Sun [33], Kuriki and Takemura [23], Takemura and Kuriki [35]). With the help of the tube method, we derive an expression approximating the upper tail probabilities of "inequality-restricted canonical correlations", which includes $\hat{\phi}$ in (1.8) as a particular case.

Difficulties in the problem treated here come from a singularity in the models (1.1) or (1.3), such that the scores μ_i and ν_j are not identifiable under the independence model $\phi = 0$. For this reason our problem is crucially different from that of Das and Sen [7], who proved asymptotic normality of the inequality-restricted canonical correlation when the true canonical correlation is nonzero and maximizing scores (μ_i, ν_j, in our case) are identifiable.

The construction of the paper is as follows. In Section 2 we formulate the inequality-restricted canonical correlation analysis or correspondence analysis, and consider a class of distributions of the maxima of Gaussian random fields that appear as asymptotic distributions of the inequality-restricted canonical correlations including $\hat{\phi}$ in (1.8). The limiting null distribution of the likelihood ratio criterion for testing $H : \phi = 0$ in the RC association model with the order restrictions is proved to be the same as that of $n \max\{\hat{\phi}, 0\}^2$ using the theory of Chernoff [5]. A formula for approximating their tail probabilities is then given by the tube method. In Section 3 we give an example of data analysis. Some techniques for fitting order-restricted models are proposed there. Proofs of the main results are given in Section 4.

2 Tail probability of the inequality-restricted canonical correlation

2.1 Inequality-restricted canonical correlation

In this subsection we give a precise definition of "inequality- (or order-) restricted canonical correlation" and derive a canonical form of its asymptotic distribution.

Let $(x_t, y_t) \in R^p \times R^q,\ t = 1, \ldots, n$, be a sequence of i.i.d. random vectors from a population with finite cumulants up to the fourth order. The population and sample covariance matrices are denoted by

$$\begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma'_{xy} & \Sigma_{yy} \end{pmatrix}, \qquad \begin{pmatrix} \hat\Sigma_{xx} & \hat\Sigma_{xy} \\ \hat\Sigma'_{xy} & \hat\Sigma_{yy} \end{pmatrix}.$$

The population covariance matrix may be singular.

Let $K \subset R^p$ and $L \subset R^q$ be closed convex cones defined by a finite or infinite number of linear inequality constrains. In this paper the maximum

$$\hat\rho = \max_{v\in K,\ w\in L} \frac{v'\hat\Sigma_{xy}w}{\sqrt{v'\hat\Sigma_{xx}v}\sqrt{w'\hat\Sigma_{yy}w}} \tag{2.1}$$

is called the (sample) inequality-restricted canonical correlation. Our definition is an extension of that of Das and Sen [6], [7]. Note that the maximum $\hat\rho$ exists unless $v'\hat\Sigma_{xx}v = 0,\ \forall v \in K$ or $w'\hat\Sigma_{yy}w = 0,\ \forall w \in L$. Obviously when $K = R^p$ and $L = R^q$, (2.1) is reduced to the largest canonical correlation in the usual definition.

For the $a \times b$ contingency table, let $x_t \in R^a$ and $y_t \in R^b$ be a pair of independent random vectors consisting of zeros and ones such that

$$P\Big(x_t = (\delta_{ik})_{1\le k\le a},\ y_t = (\delta_{jk})_{1\le k\le b}\Big) = p_{ij},$$

where δ denotes the Kronecker delta. Let

$$K = \{\mu \in R^a \mid \mu_1 \le \cdots \le \mu_a\}, \qquad L = \{\nu \in R^b \mid \nu_1 \le \cdots \le \nu_b\}.$$

Then $\hat\rho$ in (2.1) is reduced to $\hat\phi$ in (1.8).

In this paper we consider the distribution of $\hat\rho$ in (2.1) in the null case $\Sigma_{xy} = 0$. In the context of the contingency table, this is equivalent to the independence model $p_{ij} = p_{i\cdot}p_{\cdot j}$. Let $\Sigma_{xx}^{\frac{1}{2}}$ be a $p \times p$ matrix satisfying $\Sigma_{xx}^{\frac{1}{2}}\Sigma_{xx}^{\frac{1}{2}\prime} = \Sigma_{xx}$. Let $\Sigma_{yy}^{\frac{1}{2}}$ be defined similarly. Define projection matrices by $R_x = (\Sigma_{xx}^{\frac{1}{2}})^+\Sigma_{xx}^{\frac{1}{2}}$, $R_y = (\Sigma_{yy}^{\frac{1}{2}})^+\Sigma_{yy}^{\frac{1}{2}}$, where "+" denotes the Moore–Penrose generalized inverse. Then, by the assumption of the finiteness of the fourth cumulants and the central limit theorem, it is easy to show that

$$Z_n = \sqrt{n}(\Sigma_{xx}^{\frac{1}{2}})^+\hat\Sigma_{xy}(\Sigma_{yy}^{\frac{1}{2}\prime})^+$$

converges in distribution to $R_xZR'_y$ as n goes to infinity, where $Z = (z_{ij}) \in R^{p\times q}$ is a $p \times q$ random matrix such that each component z_{ij} is an independent random variable distributed according to the standard normal distribution $N(0,1)$. The set of $p \times q$ real matrices is denoted by $R^{p\times q}$.

Put

$$T_n = \sqrt{n}\max_{v\in K,\ w\in L} \frac{v'\hat\Sigma_{xy}w}{\sqrt{v'\Sigma_{xx}v}\sqrt{w'\Sigma_{yy}w}}.$$

Then

$$\frac{T_n}{\sqrt{\lambda_{\max}(\hat\Sigma_{xx}\Sigma_{xx}^+)\lambda_{\max}(\hat\Sigma_{yy}\Sigma_{yy}^+)}} \le \sqrt{n}\hat\rho \le \frac{T_n}{\sqrt{\lambda_{\min}(\hat\Sigma_{xx}\Sigma_{xx}^+)\lambda_{\min}(\hat\Sigma_{yy}\Sigma_{yy}^+)}},$$

where $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ are the maximum and nonzero minimum eigenvalues. Because $\lambda_{\max}(\hat{\Sigma}_{xx}\Sigma_{xx}^{+})$, $\lambda_{\min}(\hat{\Sigma}_{xx}\Sigma_{xx}^{+})$, $\lambda_{\max}(\hat{\Sigma}_{yy}\Sigma_{yy}^{+})$, and $\lambda_{\min}(\hat{\Sigma}_{yy}\Sigma_{yy}^{+})$ converge to one in probability, $\sqrt{n}\hat{\rho}$ has the same limiting distribution as T_n if that distribution exists.

T_n can be rewritten as

$$T_n = \max_{v \in P,\, w \in Q} v' Z_n w,$$

where

$$P = \left\{ \Sigma_{xx}^{\frac{1}{2}\prime} v \mid v \in K \right\} \cap S^{p-1}, \qquad Q = \left\{ \Sigma_{yy}^{\frac{1}{2}\prime} w \mid w \in L \right\} \cap S^{q-1}, \tag{2.2}$$

with S^{d-1} the $(d-1)$-dimensional unit sphere in R^d. By continuity of the map

$$X \ (\in R^{p \times q}) \ \mapsto \ \max_{v \in P,\, w \in Q} v' X w \ (\in R),$$

T_n is shown to converge in distribution to

$$T = \max_{v \in P,\, w \in Q} v'(R_x Z R_y') w = \max_{v \in P,\, w \in Q} v' Z w, \tag{2.3}$$

where $Z = (z_{ij}) \in R^{p \times q}$, $z_{ij} \sim N(0,1)$ is i.i.d. Now we have proved the following theorem.

Theorem 2.1 *Assume that $(x_t, y_t) \in R^p \times R^q$, $t = 1, \ldots, n$, is a sequence of i.i.d. random vectors from a population with finite fourth cumulants. Assume that x_t and y_t are uncorrelated. Then $\sqrt{n}$ times the inequality-restricted canonical correlation, $\sqrt{n}\hat{\rho}$, converges in distribution to T in (2.3) with P, Q defined in (2.2) as n goes to infinity.*

Asymptotic distributions of the order-restricted canonical correlations for two-way tables are summarized as follows.

Corollary 2.1 *(If both row and column variables are ordinal.)*
Let $\hat{\phi}$ in (1.8) be the order-restricted canonical correlation with order restrictions in both the row and column scores. Then under the independence model $p_{ij} = p_{i\cdot} p_{\cdot j}$, $\sqrt{n}\,\hat{\phi}$ has the limiting distribution T in (2.3) with

$$P = \{(v_1, \ldots, v_a)' \in S^{a-1} \mid \textstyle\sum_i \sqrt{p_{i\cdot}} v_i = 0,\ v_1/\sqrt{p_{1\cdot}} \le \cdots \le v_a/\sqrt{p_{a\cdot}}\}, \tag{2.4}$$

$$Q = \{(w_1, \ldots, w_b)' \in S^{b-1} \mid \textstyle\sum_j \sqrt{p_{\cdot j}} w_j = 0,\ w_1/\sqrt{p_{\cdot 1}} \le \cdots \le w_b/\sqrt{p_{\cdot b}}\}. \tag{2.5}$$

Corollary 2.2 *(If only the column variables are ordinal.)*
Let $\hat{\phi}$ be the order-restricted canonical correlation with order restrictions in the column scores, $\nu_1 \le \cdots \le \nu_b$. Then under the independence model $p_{ij} = p_{i\cdot} p_{\cdot j}$, $\sqrt{n}\,\hat{\phi}$ has the limiting distribution T in (2.3) with

$$P = \{(v_1, \ldots, v_a)' \in S^{a-1} \mid \textstyle\sum_i \sqrt{p_{i\cdot}} v_i = 0\},$$

and Q being given in (2.5).

The following theorem shows that these distributions arise as the limiting null distributions of the likelihood ratio criteria for testing independence. We will prove this by virtue of the theory of Chernoff [5], who discussed the asymptotic distribution of the likelihood ratio test statistic when the true parameter is on the boundary of the hypothesis parameter space (also see Self and Liang [31]). A proof is given in Section 4.

Theorem 2.2 *Assume that an $a\times b$ table $\{n_{ij}\}$ is a sample from the multinomial ($(a\times b)$-nominal) distribution under the RC model (1.3) with the order restrictions (1.4) in both the row and the column scores. Then the likelihood ratio criterion ($(-2)\times$ the maximum of the log likelihood ratio) for testing the hypothesis of independence $H:\phi=0$ converges in distribution to $\max\{T,0\}^2$ with T given in Corollary 2.1.*

The likelihood ratio criterion for testing the hypothesis of independence $H:\phi=0$ under the order restrictions (1.5) in column scores converges in distribution to T^2 with T given in Corollary 2.2.

The remainder of this section is devoted to deriving the distribution of

$$T=\max_{v\in P,\,w\in Q} v'Zw=\max_{v\in P,\,w\in Q}\mathrm{tr}((vw')'Z), \tag{2.6}$$

where $Z=(z_{ij})\in R^{p\times q}$, $z_{ij}\sim N(0,1)$ is i.i.d., and P and Q are arbitrary closed spherical convex regions, i.e., the intersection of the unit sphere and a closed convex cone. As we have seen above, this is a canonical form of the asymptotic distribution of the inequality-restricted canonical correlation. Note that $\{v'Zw \mid (v,w)\in P\times Q\}$ is a Gaussian random field of zero mean and unit variance with the index set $P\times Q$, and that T is the maximum of this random field.

2.2 The tube method

Put

$$P\otimes Q=\{v\otimes w\in R^{pq}\mid v\in P,\ w\in Q\},$$

where "$\otimes$" denotes the Kronecker product. Then $P\otimes Q$ is a subset of the unit sphere S^{pq-1} in R^{pq}. T in (2.6) can be rewritten as

$$T=\max_{u\in P\otimes Q} u'z, \tag{2.7}$$

where

$$z=\mathrm{vec}(Z)=(z_{11},z_{12},\ldots,z_{pq})'$$

is the lexicographically arranged vector of Z. For a given compact subset $M\subset S^{n-1}$, consider the maximum

$$\max_{u\in M} u'z, \qquad u'z=\sum_{i=1}^n u_i z_i, \tag{2.8}$$

where $z = (z_1, \ldots, z_n)' \sim N_n(0, I_n)$ is a Gaussian random vector. In (2.7), $n = pq$ and $M = P \otimes Q$. Note that $\{u'z \mid u \in M\}$ is a canonical form of the Gaussian random field of zero mean and unit variance having a finite Karhunen–Loève expansion.

It is in general difficult to derive the distribution of the maximum (2.8). However, recently it has been recognized that under mild regularity conditions the asymptotic expansion of the upper tail probability $P(\max_{u \in M} u'z \geq x)$ as x goes to infinity is expressed as a linear combination of upper probabilities of the χ^2 distributions with coefficients characterized by some geometric quantities of the index set M. This theory is called the *tube method*, originating from Hotelling [19], Weyl [37], and developed by Sun [33], Takemura and Kuriki [34], [35], and Kuriki and Takemura [22], [23]. In the following we give a brief summary of the tube method.

Define a geodesic distance between two points on the unit sphere S^{n-1} by the length of the great circle joining the two points:

$$\mathrm{dist}(u, v) = \cos^{-1}(u'v), \quad u, v \in S^{n-1}.$$

The subset of S^{n-1} consisting of points with distances from $M \subset S^{n-1}$ less than or equal to θ,

$$M_\theta = \Big\{ v \in S^{n-1} \mid \min_{u \in M} \cos^{-1}(u'v) \leq \theta \Big\},$$

is called the *tube* around M with radius θ.

We make assumptions on M:

Assumption 2.1 *M is an m-dimensional manifold with boundaries. M is divided disjointly as $M = \bigcup_{d=0}^{m} \partial M_d$, where ∂M_d is finite union of d-dimensional C^2-open manifold.*

Assumption 2.2 *At each point $u \in \partial M_d \subset M$, M has an $(n-d)$-dimensional tangent cone (support cone) $S_u(M)$. $S_u(M)$ is convex.*

For the definition of the tangent cone, see page 771 of Takemura and Kuriki [35].

The spherical projection point of $v \in M_\theta$ onto M, i.e., the point that attains the minimum $\min_{u \in M} \mathrm{dist}(u, v)$, is denoted by v_M. Although the projection point v_M is not necessarily determined uniquely, it is expected that for a sufficiently small $\theta > 0$ each $v \in M_\theta$ has the unique projection v_M. The supremum θ_c of such θ is called the *critical radius* of M. It can be proved that the assumptions of compactness and local convexity of M (Assumptions 2.1 and 2.2) ensure the positiveness of θ_c. θ_c can be evaluated by the following theorem, which is an extension of Proposition 4.3 of Johansen and Johnstone [20].

Theorem 2.3 *(Takemura and Kuriki [35], Lemma 2.1.)*
Let

$$N_v(M) = S_v(M)^* \cap \mathrm{span}\{v\}^\perp, \tag{2.9}$$

where

$$S_v(M)^* = \{x \in R^n \mid x'y \leq 0, \forall y \in S_v(M)\}$$

is the dual cone of $S_v(M)$ *in* R^n, $\mathrm{span}\{v\}^\perp$ *is the orthogonal complement space in* R^n *of the linear subspace spanned by* v. *Then*

$$\inf_{u,v\in M} \frac{\|u-v\|^2}{2\|\,\mathcal{P}(u-v \mid N_v(M))\|} = \begin{cases} \tan\theta_c, & \text{if } \theta_c < \pi/2, \\ \infty, & \text{if } \theta_c \ge \pi/2, \end{cases} \tag{2.10}$$

where $\mathcal{P}(\,\cdot \mid N_v(M))$ *is the orthogonal projector in* R^n *onto* $N_v(M)$.

The $(n-1)$-dimensional spherical volume of M_θ is denoted by $\mathrm{Vol}(M_\theta)$. Theorem 2.4 below gives a formula for the volume of the tube $\mathrm{Vol}(M_\theta)$, which is essentially given in Naiman [25], Theorem 3.3, or Takemura and Kuriki [34], Theorem 2.4. To state the theorem, we provide some notation.

Let $t=(t^1,\ldots,t^d)$ be a local coordinate system of the manifold ∂M_d such that $u\in \partial M_d$ has a local representation $u=\phi(t)$. The volume element of ∂M_d at u is given by $du=\sqrt{\det(g_{ij}(u))}\,dt^1\cdots dt^d$, where $g_{ij}(u)=(\partial\phi/\partial t^i)'(\partial\phi/\partial t^j)$ is the metric of ∂M_d at u. The second fundamental form of ∂M_d at u with respect to the direction v is defined as the $d\times d$ matrix $H(u,v)$ with (i,j)th element

$$H_i^j = -\sum_{k=1}^d v'\Big(\frac{\partial^2\phi}{\partial t^i\partial t^k}\, g^{kj}\Big),$$

where g^{ij} is the (i,j)th element of the inverse matrix of (g_{ij}). Note that the volume element du and the eigenvalues of $H(u,v)$ are independent of the choice of local coordinate system.

Let

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

be the volume of the unit sphere. Now we are ready to state the theorem.

Theorem 2.4 *Let* $N_u(M)$ *be defined as in (2.9). For each* $0\le\theta\le\min(\theta_c,\pi/2)$, *the volume of the tube* M_θ *is evaluated as*

$$\mathrm{Vol}(M_\theta) = \sum_{d=0}^m\sum_{k=0}^d \frac{\Omega_n}{\Omega_{d+1-k}\,\Omega_{n-d-1+k}} \int_{\partial M_d} du \int_{N_u(M)\cap S^{n-1}} dv\, \mathrm{tr}_k H(u,v) \times \bar{B}_{\frac12(d+1-k),\frac12(n-d-1+k)}(\cos^2\theta), \tag{2.11}$$

where du *is the volume element of* ∂M_d, dv *is the volume element of* $N_u(M)\cap S^{n-1}$, $H(u,v)$ *is the second fundamental form of* ∂M_d *at* u *with respect to the normal direction* v, tr_j *denotes the* j*th elementary symmetric function of eigenvalues, and* $\bar{B}_{a,b}(\cdot)$ *is the upper tail probability of the beta distribution with parameter* (a,b). *Let* $\bar{B}_{\frac12 n,0}\equiv 1$.

Because $z/\|z\|$ is distributed uniformly on the unit sphere S^{n-1}, it holds by definition that

$$\mathrm{Vol}(M_\theta)/\Omega_n = P\Big(\max_{u\in M} u'z/\|z\| \ge \cos\theta\Big).$$

Noting the independence of $z/\|z\|$ and $\|z\|$, we have

$$P\Bigl(\max_{u\in M} u'z \ge x\Bigr) = P\Bigl(\max_{u\in M} u'z/\|z\| \ge x/\|z\|\Bigr) = E\Bigl[\mathrm{Vol}(M_{\cos^{-1}(x/\|z\|)})\Bigr]\Big/\Omega_n, \tag{2.12}$$

where we let $\cos^{-1}(x/\|z\|) = 0$ if $x > \|z\|$. If the expression of the volume formula $\mathrm{Vol}(M_\theta)$ in (2.11) were valid for all θ, the distribution of $\max_{u\in M} u'z$ could be obtained by taking an expectation. That is, substituting $\cos^2\theta := x^2/\|z\|^2$ into (2.11) and taking an expectation with respect to $\|z\|^2 \sim \chi^2_n$ according to the relation

$$E\Bigl[\bar B_{\frac{1}{2}j,\frac{1}{2}(n-j)}(x^2/\|z\|^2)\Bigr] = \bar G_j(x^2),$$

where $\bar G_j(\cdot)$ is the upper probability of the χ^2 distribution with j degrees of freedom. The resulting formula is not exact because the formula (2.11) is valid only for small θ. However, if x is large, then $\cos^{-1}(x/\|z\|)$ in the right hand side of (2.12) is small, and this formal method is expected to give an answer that is correct in some sense.

In fact, according to the arguments in Sun [33] and Theorem 3.1 of Kuriki and Takemura [23], the following result holds.

Theorem 2.5 *As $x \to \infty$,*

$$\begin{aligned} P\Bigl(\max_{u\in M} u'z \ge x\Bigr) = \sum_{d=0}^{m}\sum_{k=0}^{d} \frac{1}{\Omega_{d+1-k}\,\Omega_{n-d-1+k}} \int_{\partial M_d} du \int_{N_u(M)\cap S^{n-1}} dv\, \mathrm{tr}_k H(u,v) \\ \times \bar G_{d+1-k}(x^2) + O(\bar G_{n'}(x^2(1+\tan^2\theta'_c))), \end{aligned} \tag{2.13}$$

where $\theta'_c = \min(\theta_c, \pi/2)$, and $n' = \dim \mathrm{lin}(M)$ is the dimension of the linear hull of M.

Remark 2.1 *It should be noted that the accuracy of the asymptotic expansion depends on θ_c. Larger values of θ_c give a more accurate asymptotic expansion. In particular, when M is spherically convex, the critical radius is $\theta_c \ge \pi/2$ and hence the remainder term in (2.13) becomes zero. In this case the expression gives an exact upper probability.*

2.3 Volume of the tube and an approximation of the tail probability

Here we again consider the particular case of $M = P \otimes Q$. We assume for a while that P and Q are spherical polyhedra, i.e., the intersections of polyhedral cones and unit spheres.

Let E be (the relative interior of) an $(e-1)$-dimensional face of P. Let F be (the relative interior of) an $(f-1)$-dimensional face of Q. Then

$$E\otimes F = \{v\otimes w \in R^{pq} \mid v\in E,\ w\in F\}$$

is an $(e-1)(f-1)$-dimensional C^2-manifold, forming one of the connected components of $\partial M_{(e-1)(f-1)}$. Let v and w be relative interior points of E and F, respectively. Then

$v \otimes w$ is a relative interior point of $E \otimes F$, and the tangent cone of $P \otimes Q$ at $v \otimes w$ is given by

$$S_{v\otimes w}(P \otimes Q) = S_v(P) \otimes \{w\} \oplus \{v\} \otimes S_w(Q),$$

where $S_v(P) \subset R^p$ is the tangent cone of P at v, $S_w(Q) \subset R^q$ is the tangent cone of Q at w, and "$\oplus$" denotes the direct sum of vector spaces. Because both $S_v(P)$ and $S_w(Q)$ are convex, so is $S_{v\otimes w}(P \otimes Q)$. We have seen that Assumptions 2.1 and 2.2 are fulfilled.

Therefore, the tail probability of the inequality-restricted canonical correlation can be obtained by evaluating the volume of the tube around $P \otimes Q \subset S^{pq-1}$ and its critical radius θ_c, at least when P and Q are polyhedral. Indeed, we can reach results valid for non-polyhedral P and Q by considering approximating sequences of spherical polyhedra. The results are summarized as follows. The derivations are given in Section 4.

Theorem 2.6 *Let $P \subset S^{p-1}$ and $Q \subset S^{q-1}$ be spherical convex regions. Let P_θ be the tube around P in S^{p-1}. Assume that the $(p-1)$-dimensional volume of the tube P_θ is expressed in terms of the coefficients, $w_e(P)$, $1 \le e \le p$, as*

$$\mathrm{Vol}(P_\theta) = \Omega_p \sum_{e=1}^{p} w_e(P) \bar{B}_{\frac{1}{2}e,\frac{1}{2}(p-e)}(\cos^2\theta).$$

Let $w_f(Q)$, $1 \le f \le q$, be coefficients defined similarly to $w_e(P)$. Then as $x \to \infty$, it remains true that

$$\begin{aligned} &P\Big(\max_{v\in P,\, w\in Q} v'Zw \ge x\Big) \\ &\quad = \sum_{e=1}^{p}\sum_{f=1}^{q} w_e(P)\, w_f(Q) \sum_{k=0,\,k:\text{even}}^{2(\min(e,f)-1)} c_{e,f,k}\, \bar{G}_{e+f-1-k}(x^2) + O(\bar{G}_{n'}(x^2(1+\tan^2\theta_c))), \end{aligned} \tag{2.14}$$

where

$$c_{e,f,k} = (-1)^{k/2}\, 2^{e+f-1-k/2}\, \frac{\Gamma(\frac{1}{2}(e+1))\,\Gamma(\frac{1}{2}(f+1))\,\Gamma(\frac{1}{2}(e+f-1-k))}{\sqrt{\pi}\,\Gamma(e-k/2)\,\Gamma(f-k/2)\,(k/2)!}, \tag{2.15}$$

$n' = \dim \mathrm{lin}(P) \times \dim \mathrm{lin}(Q)$, and θ_c is the critical radius of $P \otimes Q$. Here if both P and Q are symmetric with respect to the origin, (i.e., if both P and Q are the unit spheres restricted to certain-dimensional linear subspaces) then multiply the right hand side of (2.14) by $1/2$.

Remark 2.2 *It is in general not easy to evaluate the coefficients $w_e(P)$ for arbitrarily given P. However, when P is the spherical polyhedron defined in (2.4) of Corollary 2.1, then the $w_e(P)$s are so-called level probabilities, explained below, and methods for evaluating them numerically are known.*

Consider a one-way ANOVA model $x_i \sim N(\theta_i, 1/n_i)$, $i = 1, \dots, k$. Denote by $\hat{\theta}_i$ the maximum likelihood estimator of θ_i under the simple order restriction $\theta_1 \le \cdots \le \theta_k$.

The level probability $P(l,k;n_1,\ldots,n_k)$, $1\le l\le k$, is defined to be the probability under $\theta_1=\cdots=\theta_k$ that the MLEs $\hat\theta_i$, $i=1,\ldots,k$, take exactly l distinct values. Then $w_e(P)$ with P defined in (2.4) is equal to

$$w_e(P)=\begin{cases} P(e+1,a;p_{1\cdot},\ldots,p_{a\cdot}), & 1\le e\le a-1,\\ 0, & e=a.\end{cases} \tag{2.16}$$

The expressions of $P(l,k;n_1,\ldots,n_k)$ for $k\le 4$ are given in Barlow, et al. [2], Section 3.3. For general k, Miwa, et al. [24] pointed out that $P(l,k;n_1,\ldots,n_k)$ can be evaluated numerically using a successive integration technique. From a practical point of view there is no difficulty in calculating the coefficients $w_e(P)$.

The following theorem gives the critical radius of $P\otimes Q$. Define the (spherical) diameter of $P\subset S^{p-1}$ by

$$\phi=\phi(P)=\sup_{u,v\in P}\cos^{-1}(u'v).$$

A proof of Theorem 2.7 is given in Section 4.

Theorem 2.7 *Let $P\subset S^{p-1}$ and $Q\subset S^{q-1}$ be spherically convex subsets of the unit spheres such that* $\dim P\ge 1$, $\dim Q\ge 1$. *If at least either $\phi(P)\le\pi/2$ or $\phi(Q)\le\pi/2$ holds, then the critical radius θ_c of $P\otimes Q$ is $\pi/4$. If $\phi(P)>\pi/2$ and $\phi(Q)>\pi/2$, then*

$$\theta_c=\tan^{-1}\sqrt{\frac{1-\cos\phi(P)\cos\phi(Q)}{1+\cos\phi(P)\cos\phi(Q)}}.$$

Corollary 2.3 *The critical radius of $P\otimes Q$ in Corollary 2.1 is $\pi/4$. The critical radius of $P\otimes Q$ in Corollary 2.2 is $\pi/4$.*

Proof. The set P defined in (2.4) is the intersection of the unit sphere S^{p-1} and the convex cone generated by $a-1$ edge vectors

$$e_i=\Bigl(-\frac{\sqrt{p_{1\cdot}}}{q_i},\ldots,-\frac{\sqrt{p_{i\cdot}}}{q_i},\frac{\sqrt{p_{i+1,\cdot}}}{1-q_i},\ldots,\frac{\sqrt{p_{a\cdot}}}{1-q_i}\Bigr)',\quad 1\le i\le a-1,$$

with $q_i=\sum_{j=1}^i p_{j\cdot}$. Because $(e_i)'e_j=1/(q_j(1-q_i))>0$ for $i<j$, it holds that $u'v\ge 0$ $\forall u,v\in P$. This is equivalent to $\phi(P)\le\pi/2$. ∎

Now we have determined the volume of tube and the critical radius of $P\otimes Q$. We summarize below the results in three important cases.

Corollary 2.4 *The tail probability $P(T\ge x)$ of T defined in Corollary 2.1 is given as (2.14) with $w_e(P)$ in (2.16),*

$$w_f(Q)=\begin{cases} P(f+1,b;p_{\cdot 1},\ldots,p_{\cdot b}), & 1\le f\le b-1,\\ 0, & f=b,\end{cases} \tag{2.17}$$

$n'=(a-1)(b-1)$, and $\theta_c=\pi/4$.

Corollary 2.5 *The tail probability $P(T \geq x)$ of T defined in Corollary 2.2 is given as (2.14) with*

$$w_e(P) = \begin{cases} 1, & e = a-1, \\ 0, & 1 \leq e \leq a-2,\ e = a, \end{cases}$$

$w_f(Q)$ in (2.17), $n' = (a-1)(b-1)$, and $\theta_c = \pi/4$.

Corollary 2.6 *(Kuriki and Takemura [23], Corollary 4.2)*
The tail probability of the square root of the largest eigenvalue of the Wishart matrix $W_p(q, I_p)$ is given by

$$P\Big(\max_{v \in S^{p-1},\, w \in S^{q-1}} v'Zw \geq x\Big) = \frac{1}{2} \sum_{k=0,\, k:\mathrm{even}}^{2(\min(p,q)-1)} c_{p,q,k}\, \bar{G}_{p+q-1-k}(x^2) + O(\bar{G}_{pq}(2x^2)) \qquad (2.18)$$

with the coefficient $c_{p,q,k}$ given in (2.15).

A numerical study to check the accuracy of the proposed approximation method is summarized in Figure 2.1. In both the left and the right figures, the approximate upper tail probabilities of $T = \max_{v \in P,\, w \in Q} v'Zw$ in (2.6) by Theorem 2.6 are plotted as solid lines. The tail probabilities estimated by Monte Carlo simulations with 100,000 replications are plotted by dashed lines. Upper and lower bounds of tail probabilities using the method of Theorem 3.1 of Kuriki and Takemura [23] are plotted by dotted lines.

The left figure depicts the case where $p = q = 2$,

$$P = Q = \{(\cos\theta, \sin\theta)' \mid 0 \leq \theta \leq \pi/3\}, \quad \theta_c = \pi/4.$$

This corresponds to a 3×3 table with $p_{i\cdot} \equiv 1/3$, $p_{\cdot j} \equiv 1/3$, with both row and column categories ordinal.

The right figure depicts the case where $p = 5$, $q = 2$,

$$P = S^{5-1}, \quad Q = \{(\cos\theta, \sin\theta)' \mid 0 \leq \theta \leq \pi/3\}, \quad \theta_c = \pi/4.$$

This corresponds to a 6×3 table with $p_{i\cdot} \equiv 1/6$, $p_{\cdot j} \equiv 1/3$, with only the column categories ordinal.

In each case one can see that the approximations using the tube method are sufficiently close to the tail probabilities estimated by Monte Carlo simulations. Therefore the proposed formula is accurate enough in practice for calculating relevant p-values of tests.

3 An example of data analysis

3.1 Fitting order-restricted models

In this section a contingency table is analyzed as an example. This is a cross-classified table of job satisfaction by income given in Table 3.1, which is cited from Agresti [1], Table

2.8. Everitt [11] reanalyzed the data using the (unrestricted) RC association model. This is a typical example of categorical data with both row and column variables ordinal.

We reanalyze the data using the order-restricted correspondence analysis and using the order-restricted RC association model. The results of the data analysis are summarized in Tables 3.2 and 3.3. Before examining the results we discuss the method of estimating parameters under the order restrictions. In the following we consider the case where there are natural orderings in row and column categories only for purposes of explanation.

In both the correspondence analysis and the RC association model, the parameters ϕ, μ_i and ν_j can be estimated from collapsed tables. Let $\mathcal{I} = I_1|I_2|\dots$ be a partition of $\{1, 2, \dots, a\}$ such that if $i \in I_k$ and $i' \in I_{k+1}$ then $i < i'$. For example $\mathcal{I} = 12|3|456$ is such a partition of $\{1, 2, 3, 4, 5, 6\}$. Let $\mathcal{J} = J_1|J_2|\dots$ be a partition of $\{1, 2, \dots, b\}$ such that if $j \in J_l$ and $j' \in J_{l+1}$ then $l < l'$. Given a pair of partitions $(\mathcal{I}, \mathcal{J})$, define a collapsed table with the (k, l)th cell

$$N_{kl} = \sum_{i \in I_k,\, j \in J_l} n_{ij}.$$

Let $\check{\phi}$, $\check{\mu}_k$ and $\check{\nu}_l$ be the estimates of ϕ, μ_k and ν_l in the correspondence analysis (1.1) or in the RC association model (1.3) under the side condition (1.2) without order restriction.

Lemma 3.1 *Let $\hat{\phi}$, $\hat{\mu}_i$ and $\hat{\nu}_j$ be the estimates in the correspondence analysis (or the maximum likelihood estimates in the RC association model) with order restrictions based on the non-collapsed data. Then there exists a pair of partitions $(\mathcal{I}, \mathcal{J})$ such that the estimates of the correspondence analysis (or the maximum likelihood estimates based on the RC association model, resp.) without order restriction based on the collapsed table $\{N_{kl}\}$ satisfy $\check{\phi} = \hat{\phi}$, $\check{\mu}_k = \hat{\mu}_i$, $\check{\nu}_l = \hat{\nu}_j$ for $i \in I_k$, $j \in J_l$.*

Lemma 3.1 for the RC association model is Lemma 4 of Ritov and Gilula [28]. The proof for correspondence analysis is parallel and omitted.

The desired order-restricted estimator can be obtained in principle by examining all the possible ways of collapsing. Here we must be careful about a particular partition $\mathcal{I} = 12\cdots a$ of $\{1, 2, \dots, a\}$ corresponding to $\mu_1 = \mu_2 = \cdots = \mu_a$. This is reduced to the independence model whenever $\mathcal{J}$ is. There are $(2^{a-1} - 1) \times (2^{b-1} - 1)$ ways of collapsing to be taken into account in addition to the independence model.

Moreover, the following branch and bound techniques can be used. For two partitions $\mathcal{I}$ and $\mathcal{I}'$, write $\mathcal{I} \preceq \mathcal{I}'$ if $\mathcal{I}$ is a subpartition of $\mathcal{I}'$. For example, $\mathcal{I} = 12|3|456 \preceq \mathcal{I}' = 12|3456$. Write $(\mathcal{I}, \mathcal{J}) \preceq (\mathcal{I}', \mathcal{J}')$ if $\mathcal{I} \preceq \mathcal{I}'$ and $\mathcal{J} \preceq \mathcal{J}'$.

Rule 1. Once a feasible (i.e., satisfying the order restrictions) solution corresponding to a pair of partitions, say $(\mathcal{I}, \mathcal{J})$, is found, it is not necessary to count other collapsings $(\mathcal{I}', \mathcal{J}')$ such that $(\mathcal{I}, \mathcal{J}) \preceq (\mathcal{I}', \mathcal{J}')$

Rule 2. If a solution (which may be feasible or nonfeasible) corresponding to a pair of partitions, say $(\mathcal{I}, \mathcal{J})$, gives a smaller canonical correlation (or likelihood) than the canonical correlation (or likelihood, resp.) given by the other feasible solution, it is not necessary to count other collapsings $(\mathcal{I}', \mathcal{J}')$ such that $(\mathcal{I}, \mathcal{J}) \preceq (\mathcal{I}', \mathcal{J}')$.

Therefore we can propose a procedure for examining $(2^{a-1}-1)\times(2^{b-1}-1)+1$ possibilities by starting with the smallest pair $(\mathcal{I},\mathcal{J})$, $\mathcal{I}=1|2|\cdots|a$, $\mathcal{J}=1|2|\cdots|b$, and searching other possibilities in an ascending direction in the sense of the partial order "$\preceq$".

This naive procedure seems unrealistic at first glance because the number of collapsed tables is of exponential order in $a+b$ as a and b increase. However, without order restriction, not only the correspondence analysis but also the maximum likelihood estimation of the RC association model can be performed at small computational cost (Becker [3]). According to preliminary numerical experiments, at least the case $a=b=10$ is manageable by standard personal computers even when Rules 1 and 2 above are not applied.

Finally it should be noted that, once the maximum likelihood estimates $\hat{\phi}$, $\hat{\mu}_i$ and $\hat{\nu}_j$ in the order-restricted RC model are obtained, then the maximum likelihood estimates of cell probabilities can be obtained by the iterative proportional scaling (IPS) procedure. The algorithm is as follows.

Step 1. Set $\hat{p}_{ij}=n_{ij}/n$, $q_{ij}:=e^{\hat{\phi}\hat{\mu}_i\hat{\nu}_j}$ as an initial value.

Step 2. Iterate the following: $q_{ij}:=(\hat{p}_{i\cdot}/q_{i\cdot})\times q_{ij}$, $q_{ij}:=(\hat{p}_{\cdot j}/q_{\cdot j})\times q_{ij}$.

Then the maximum likelihood estimates of cell probabilities are obtained as the limit in Step 2. The figures in parentheses in Table 3.1 are expected frequencies under the order-restricted RC association model obtained by this procedure.

3.2 Results of data analysis

Now we return to the analysis of Table 3.1. The estimates are summarized in Table 3.2. The row labeled "CA" indicates the estimates by correspondence analysis, and the row labeled "RC" indicates the maximum likelihood estimates based on the RC association model. The additional label "(ordered)" indicates when the order restrictions were imposed. In Table 3.2, it is evident that $\hat{\nu}_1=\hat{\nu}_2=\hat{\nu}_3$ in the correspondence analysis, whereas $\hat{\nu}_2=\hat{\nu}_3$ in the RC model approach.

The results of significance tests for independence are summarized in Table 3.3. In the row labeled "CA" the statistic $n\hat{\phi}^2$ is used as a test statistic, where $\hat{\phi}$ is the largest canonical correlation under the order restrictions. In the row labeled "RC" the likelihood ratio tests for independence against the RC association model are applied. The additional label "(ordered)" again indicates the order restrictions. The p-values of the test statistics without order restriction are calculated using (2.18) of Corollary 2.6 with $p=q=3$. The p-values of the test statistics with order restrictions were calculated by obtaining the level probabilities $w_e(P)$, $w_f(Q)$ first, and then substituting them into (2.14) of Theorem 2.6. The empirical marginal probabilities are $(\hat{p}_{i\cdot})=(206,289,235,171)/901$, $(\hat{p}_{\cdot j})=(62,108,319,412)/901$, and the corresponding level probabilities are

$$(w_e(P))=(0.451,0.268,0.049),\qquad (w_f(Q))=(0.450,0.271,0.050),$$

respectively. Because $\hat{p}_{i\cdot}$ and $\hat{p}_{\cdot j}$ are $\sqrt{n}$-consistent estimators of $p_{i\cdot}$ and $p_{\cdot j}$, and $w_e(P)$ and $w_f(Q)$ are differentiable with respect to $p_{i\cdot}$ and $p_{\cdot j}$, this method gives a $\sqrt{n}$-consistent estimator of the p-value.

One finds that imposing the order restrictions makes the p-values much smaller. We can interpret this reduction in the p-values as a reflection of the improved power of the tests.

4 Proofs of the main results

4.1 The proof of Theorem 2.2

We begin by summarizing Chernoff's theory for the distributions of likelihood ratio criterion applied to the multinomial distribution. For simplicity, the statements below are written in terms of vector-valued (not matrix-valued) multinomial random variables.

Theorem 4.1 *Let $x^{(t)} = (x_i^{(t)})_{1\le i\le k} \in R^k$, $t = 1,\ldots,n$, be an i.i.d. sequence of random vectors having all elements zero except one element one from a probability density $f(x,\theta) = \prod_{i=1}^k \theta_i^{x_i}$, $\theta \in \Omega$, where*

$$\Omega = \{\theta = (\theta_i) \in R^k \mid \theta_i > 0,\ \textstyle\sum_{i=1}^k \theta_i = 1\}.$$

Let $\theta^o \in \Omega$ be the true value. For $j = 0, 1$, let ω_j be a subset of Ω, and assume that both ω_js are locally compact and contain the true value θ^o. Let

$$T(\Omega) = \{\tilde{\theta} = (\tilde{\theta}_i) \in R^k \mid \textstyle\sum_{i=1}^k \tilde{\theta}_i = 0\}$$

be the tangent space of Ω at θ^o. Assume that, for $j = 0, 1$, ω_j has the approximating cone (tangent cone in the sense of Definition 2 of Chernoff [5] or Definition 1 of Shapiro [32]) $S(\omega_j) \subset T(\Omega)$ at θ^o. That is, for each $\omega = \omega_j$, when θ^o is an accumulating point of ω, a closed cone $S(\omega) \subset T(\Omega)$ exists, satisfying:

(i) for any sequence $y_l \in \omega \setminus \{\theta^o\}$ such that $\lim_{l\to\infty} y_l = \theta^o$, $\inf_{x\in S(\omega)} \|x - (y_l - \theta^o)\| = o(\|y_l - \theta^o\|)$, and

(ii) for any sequence $x_l \in S(\omega) \setminus \{0\}$ such that $\lim_{l\to\infty} x_l = 0$, $\inf_{y\in\omega} \|x_l - (y - \theta^o)\| = o(\|x_l\|)$;

when θ^o is an isolated point of ω, let $S(\omega) = \{0\}$. Then, for n sufficiently large, MLEs $\hat{\theta}_{n,0}$, $\hat{\theta}_{n,1}$ exist that maximize $\ell_n(\theta) = \sum_{t=1}^n \log f(x^{(t)},\theta)$ over the sets ω_0, ω_1, respectively, and the likelihood ratio test statistic for testing $H_0 : \theta \in \omega_0$ against $H_1 : \theta \in \omega_1$,

$$-2\log\Lambda_n = -2\{\ell_n(\hat{\theta}_{n,0}) - \ell_n(\hat{\theta}_{n,1})\},$$

converges in distribution to

$$\min_{\theta\in S(\omega_0)} (z-\theta)' D(\theta^o)^{-1}(z-\theta) - \min_{\theta\in S(\omega_1)} (z-\theta)' D(\theta^o)^{-1}(z-\theta) \qquad (4.1)$$

as $n \to \infty$*, where* $D(\theta) = \mathrm{diag}(\theta_i)_{1\le i\le k}$*,* z *is a* $k\times 1$ *random vector distributed as* $N_k(0, V(\theta^o))$ *with* $V(\theta) = D(\theta) - D(\theta)1_k 1_k' D(\theta)$*, and* $1_k = (1,\ldots,1)'$ *is a* $k\times 1$ *vector consisting of ones.*

Remark 4.1 *A multinomial random vector* $y = (y_i)_{1\le i\le k}$*,* $y_i = \sum_{t=1}^n x_i^{(t)}$*, is a sufficient statistic for* θ*, and hence the MLE and the likelihood ratio criterion based on the i.i.d. sequence* $x^{(t)}$*,* $t = 1,\ldots,n$*, are equivalent to those based on* y*.*

Proof. Under the assumptions that ω_js are locally compact and contain the true value, the MLEs $\hat\theta_{n,j} \in \omega_j$ under H_j exist for n sufficiently large, and converge to θ^o with probability one as $n\to\infty$ (Berk [4], Example 4). From this fact and the assumption of the existence of approximating cones, we can see that all of the assumptions of Theorem 1 of Chernoff [5] are fulfilled, and $-2\log\Lambda_n$ is proved to converge in distribution to

$$\min_{\theta\in S(\omega_0)} (\tilde z - i(\theta))' I(\theta^o)(\tilde z - i(\theta)) - \min_{\theta\in S(\omega_1)} (\tilde z - i(\theta))' I(\theta^o)(\tilde z - i(\theta)), \tag{4.2}$$

where $i(\theta) = (\theta_1,\ldots,\theta_{k-1})'$ is the first $k-1$ elements of θ, $I(\theta) = (\delta_{ij}/\theta_i + 1/\theta_k)_{1\le i,j\le k-1}$, $\theta_k = 1 - \sum_{i=1}^{k-1}\theta_i$, is the Fisher matrix of $i(\theta)$, and $\tilde z$ is a $(k-1)\times 1$ random vector distributed as $N_{k-1}(0, I(\theta^o)^{-1})$. It is easy to see that (4.1) and (4.2) have the same distribution. ∎

In applying Chernoff's theory, one crucial step is to find the approximating cone of the parameter set. The following lemma gives a useful sufficient condition for the approximating cone.

Lemma 4.1 *Suppose that there exist a neighborhood of* $U\subset\Omega$ *around* θ^o*, and* C^1*-diffeomorphism* $\varphi: U\to\varphi(U)\subset T(\Omega)$ *such that* $\varphi(\theta^o) = 0$*,* $\varphi(\overline{\omega\cap U}) = S\cap\varphi(\bar U)$*, and the differential map of* φ *at* θ^o*,*

$$d\varphi|_{\theta^o} : T(\Omega) \to T(\Omega),$$

is the identity map. Then S *is the approximating cone of* ω *at* θ*.*

Proof. Write the inverse of φ as φ^{-1}. For $y\in\omega\cap U$, $\inf_{x\in S}\|x - (y-\theta^o)\| \le \inf_{y'\in\omega\cap U}\|\varphi(y') - (y-\theta^o)\| \le \|\varphi(y) - (y-\theta^o)\| = o(\|y-\theta^o\|)$. For $x\in S\cap V$, $V = \varphi(U)$, $\inf_{y\in\omega}\|x - (y-\theta^o)\| \le \inf_{x'\in S\cap V}\|x - (\varphi^{-1}(x') - \theta^o)\| \le \|x - (\varphi^{-1}(x) - \theta^o)\| = o(\|x\|)$. ∎

Now we return to our problem of the $a\times b$ contingency table. We first treat the case of the RC model with both row and column variables ordinal.

The ambient parameter space is

$$\Omega = \{(p_{ij})_{1\le i\le a,\ 1\le j\le b} \in R^{a\times b} \mid \textstyle\sum_{ij} p_{ij} = 1,\ p_{ij} > 0\},$$

which is of dimension $ab-1$. The true parameter is denoted by $(p_{ij}^o) = (p_{i\cdot}^o p_{\cdot j}^o) \in \Omega$. The null parameter space ω_0 is the set of $(p_{ij})\in\Omega$ satisfying (1.3) with $\phi = 0$. The alternative parameter space ω_1 is the set of $(p_{ij})\in\Omega$ satisfying (1.3) with the order restriction (1.4). Because both ω_0 and ω_1 are locally compact and contain the true parameter (p^o_{ij}), the MLEs under H_0 and H_1 exist and are consistent. We will proceed by determining the approximating cones. The tangent space of Ω at (p^o_{ij}) is defined by $T(\Omega) = \{(p_{ij})_{1\le i\le a,\,1\le j\le b} \in R^{a\times b} \mid \sum_{ij} p_{ij} = 0\}$. Define a map $\varphi : \Omega \to T(\Omega)$ by

$$(p_{ij})_{1\le i\le a,\,1\le j\le b} \quad \mapsto \quad \Big(p^o_{ij}\log\frac{p_{ij}}{p^o_{ij}} - \sum_{ij} p^o_{ij}\log\frac{p_{ij}}{p^o_{ij}}\Big)_{1\le i\le a,\,1\le j\le b}.$$

Without loss of generality, we put the side conditions for the parameters in (1.3) as

$$\textstyle\sum_i p^o_{i\cdot}\mu_i = 0, \quad \sum_j p^o_{\cdot j}\beta_j = \sum_j p^o_{\cdot j}\nu_j = 0 \tag{4.3}$$

instead of (1.2). Let $p^o_{ij} = \exp(\alpha^o_i + \beta^o_j)$, $\sum_j p^o_{\cdot j}\beta^o_j = 0$. It is easy to see that the map φ is one-to-one, $\varphi(p^o_{ij}) = 0$, and its differential map $d\varphi$ at (p^o_{ij}) is the identity map. Therefore, the approximating cone of ω_0 at p^o_{ij} is obtained as the cone generated by the set $\varphi(\omega_0)$ (the smallest cone containing the set $\varphi(\omega_0)$). Noting the side conditions (4.3),

$$\begin{aligned} S(\omega_0) &= \text{the cone generated by } \varphi(\omega_0) \\ &= \{p^o_{i\cdot}p^o_{\cdot j}(\alpha_i - \alpha^o_i + \beta_j - \beta^o_j) - \textstyle\sum_{ij} p^o_{i\cdot}p^o_{\cdot j}(\alpha_i - \alpha^o_i + \beta_j - \beta^o_j) \mid \sum_j p^o_{\cdot j}\beta_j = 0\} \\ &= \{(\tilde\alpha_i p^o_{\cdot j} + p^o_{i\cdot}\tilde\beta_j) \mid \textstyle\sum_i \tilde\alpha_i = \sum_j \tilde\beta_j = 0\}. \end{aligned}$$

In the last equation we put $\tilde\alpha_i = p^o_{i\cdot}(\alpha_i - \alpha^o_i) - \sum_i p^o_{i\cdot}(\alpha_i - \alpha^o_i)$, $\tilde\beta_{\cdot j} = p^o_j(\beta_j - \beta^o_j)$. Similarly the approximating cone of ω_1 at p^o_{ij} is obtained as

$$\begin{aligned} S(\omega_1) &= \{p^o_{i\cdot}p^o_{\cdot j}(\alpha_i - \alpha^o_i + \beta_j - \beta^o_j + \phi\mu_i\nu_j) - \textstyle\sum_{ij} p^o_{i\cdot}p^o_{\cdot j}(\alpha_i - \alpha^o_i + \beta_j - \beta^o_j + \phi\mu_i\nu_j) \\ &\qquad \mid \text{side conditions (4.3)}\} \\ &= \{(\tilde\alpha_i p^o_{\cdot j} + p^o_{i\cdot}\tilde\beta_j + \tilde\phi(p^o_{i\cdot}\tilde\mu_i)(p^o_{\cdot j}\tilde\nu_j)) \mid \textstyle\sum_i \tilde\alpha_i = \sum_j \tilde\beta_j = \sum_i p^o_{i\cdot}\tilde\mu_i = \sum_j p^o_{\cdot j}\tilde\nu_j = 0, \\ &\qquad \textstyle\sum_i p^o_{i\cdot}\tilde\mu_i^2 = \sum_j p^o_{\cdot j}\tilde\nu_j^2 = 1,\ \tilde\phi \ge 0,\ \tilde\mu_1 \le \cdots \le \tilde\mu_a,\ \tilde\nu_1 \le \cdots \le \tilde\nu_b\}. \end{aligned}$$

Let $p = \mathrm{vec}(p_{ij}) = (p_{11}, p_{12}, \ldots, p_{ab})'$ be the lexicographically arranged parameter. At the true value $p^o = \mathrm{vec}(p^o_{i\cdot}p^o_{\cdot j})$,

$$\begin{aligned} D(p^o) &= \mathrm{diag}(p_{ij})_{ab\times ab}\Big|_{p_{ij}=p^o_{ij}} = P \otimes Q, \\ V(p^o) &= D(p) - D(p)1_{ab}1'_{ab}D(p)\Big|_{p_{ij}=p^o_{ij}} = P\otimes Q - P1_a1'_aP \otimes Q1_b1'_bQ, \end{aligned}$$

where $P = \mathrm{diag}(p^o_{i\cdot})_{a\times a}$, and $Q = \mathrm{diag}(p^o_{\cdot j})_{b\times b}$. The inverse of $D(p^o)$ is given by

$$D(p^o)^{-1} = P^{-1} \otimes Q^{-1}.$$

Now we are ready to derive the limiting null distribution of the likelihood ratio statistic with the help of Theorem 4.1. Let Z be an $a \times b$ Gaussian random matrix with mean matrix M such that

$$\mathrm{vec}(Z) \sim N_{ab}(\mathrm{vec}(M), P\otimes Q - P1_a1'_aP \otimes Q1_b1'_bQ),$$

and let

$$\begin{aligned} L(M) &= \operatorname{vec}(Z-M)' D(p^o)^{-1} \operatorname{vec}(Z-M) \\ &= \operatorname{tr}(P^{-1}(Z-M)Q^{-1}(Z-M)') = \|P^{-\frac{1}{2}}(Z-M)Q^{-\frac{1}{2}}\|^2, \end{aligned}$$

where $P^{-\frac{1}{2}} = \operatorname{diag}(1/\sqrt{p^o_{i\cdot}})$, $Q^{-\frac{1}{2}} = \operatorname{diag}(1/\sqrt{p^o_{\cdot j}})$, and $\|\cdot\|$ is the matrix norm. Then our required limiting distribution is expressed as

$$\min_{M \in S(\omega_0)} L(M) - \min_{M \in S(\omega_1)} L(M).$$

Note that $M \in S(\omega_1)$ is written as

$$M = \alpha 1_b' Q + P 1_a \beta' + \phi P \mu \nu' Q$$

with $\alpha = (\alpha_1, \ldots, \alpha_a)'$, $\beta = (\beta_1, \ldots, \beta_b)'$, $\mu = (\mu_1, \ldots, \mu_a)'$, and $\nu = (\nu_1, \ldots, \nu_b)'$. Using the relations $1_a'\alpha = 1_b'\beta = 1_a' P\mu = 1_b' Q\nu = 0$, $\mu' P \mu = \nu' Q \nu = 1$, $L(M)$ is decomposed as

$$\begin{aligned} L(M) &= \|P^{-\frac{1}{2}}(Z_2 - \alpha 1_b' Q)Q^{-\frac{1}{2}}\|^2 + \|P^{-\frac{1}{2}}(Z_3 - P1_a\beta')Q^{-\frac{1}{2}}\|^2 \\ &\quad + \|P^{-\frac{1}{2}}(Z_1 - \phi P\mu\nu' Q)Q^{-\frac{1}{2}}\|^2 + \|P^{-\frac{1}{2}} Z_4 Q^{-\frac{1}{2}}\|^2, \end{aligned}$$

where

$$\begin{aligned} &Z_1 = (I_a - P1_a 1_a') Z (I_b - 1_b 1_b' Q), \qquad Z_2 = (I_a - P1_a1_a')Z(1_b1_b'Q), \\ &Z_3 = (P1_a1_a')Z(I_b - 1_b1_b'Q), \qquad Z_4 = (P1_a1_a')Z(1_b1_b'Q). \end{aligned}$$

Therefore

$$\begin{aligned} \min_{M \in S(\omega_0)} L(M) - \min_{M \in S(\omega_1)} L(M) &= \|P^{-\frac{1}{2}} Z_1 Q^{-\frac{1}{2}}\|^2 - \min_{\phi \ge 0,\, \mu,\, \nu} \|P^{-\frac{1}{2}}(Z_1 - \phi P\mu\nu' Q)Q^{-\frac{1}{2}}\|^2 \\ &= \|\tilde{Z}_1\|^2 - \min_{\phi\ge 0,\, \tilde\mu,\, \tilde\nu} \|\tilde{Z}_1 - \phi\tilde\mu\tilde\nu'\|^2 \\ &= \max_{\tilde\mu,\, \tilde\nu}\Big(\max\{\tilde\mu'\tilde{Z}_1\tilde\nu, 0\}^2\Big) \\ &= \max\Big\{\max_{\tilde\mu,\,\tilde\nu}(\tilde\mu'\tilde{Z}_1\tilde\nu), 0\Big\}^2 \qquad (4.4) \end{aligned}$$

with $P^{-\frac{1}{2}} Z_1 Q^{-\frac{1}{2}} = \tilde{Z}_1$, $P^{\frac{1}{2}}\mu = \tilde\mu$, $Q^{\frac{1}{2}}\nu = \tilde\nu$. The constraints of the maximization in (4.4) are

$$1_a' P^{\frac{1}{2}}\tilde\mu = 1_b' Q^{\frac{1}{2}}\tilde\nu = 0, \qquad \tilde\mu'\tilde\mu = \tilde\nu'\tilde\nu = 1$$

and

$$\tilde\mu_1/\sqrt{p^o_{1\cdot}} \le \cdots \le \tilde\mu_a/\sqrt{p^o_{a\cdot}}, \qquad \tilde\nu_1/\sqrt{p^o_{\cdot 1}} \le \cdots \le \tilde\nu_b/\sqrt{p^o_{\cdot b}}.$$

Because

$$\operatorname{vec}(\tilde{Z}_1) \sim N_{ab}\Big(0, (I_a - P^{\frac{1}{2}} 1_a 1_a' P^{\frac{1}{2}}) \otimes (I_b - Q^{\frac{1}{2}} 1_b 1_b' Q^{\frac{1}{2}})\Big),$$

the distribution of the maximum $\max_{\tilde\mu,\tilde\nu}(\tilde\mu'\tilde{Z}_1\tilde\nu)$ in (4.4) is shown to be reduced to the distribution of T in (2.6) with P, Q given in Corollary 2.1.

The proof for the order restrictions (1.4) is completed. The proof for the order restrictions (1.5) is completely similar and is omitted.

4.2 The proof of Theorem 2.6

The proof is divided into four parts. In the first three sections (Sections 4.2.1–4.2.3) we prove the theorem when P and Q are polyhedral. Then, in Section 4.2.4, the results are extended to the non-polyhedral case by considering sequences of spherical polyhedra approximating P and Q. Throughout this section we denote by "$\oplus$" the orthogonal direct sum.

4.2.1 Tangent cones and their duals

Let E be a face of the spherical polyhedron P of dimension $e-1$, $0 \le e-1 \le p-1$. Fix a relative interior point v of E. The tangent cone of P at v is given by

$$S_v(P) = T_v(E) \oplus C_v,$$

where $T_v(E)$ is the tangent space of E at v, and C_v is a convex cone contained in the orthogonal compliment space

$$N_v(E) = (\mathrm{span}\{v\} \oplus T_v(E))^\perp \subset R^p.$$

The dual cone of $S_v(P)$ in R^p is

$$S_v(P)^* = \mathrm{span}\{v\} \oplus \tilde{C}_v \quad \text{with} \quad \tilde{C}_v = \{y \in N_v(E) \mid y'x \le 0,\ \forall x \in C_v\}.$$

Let F be a face of Q of dimension $f-1$, $0 \le f-1 \le q-1$. Fix a relative interior point w of F. Define $T_w(F)$ and

$$N_v(F) = (\mathrm{span}\{w\} \oplus T_w(F))^\perp \subset R^q$$

as above. Then the tangent cone of Q at w and its dual in R^q are written as

$$S_w(Q) = T_w(F) \oplus D_w \quad \text{and} \quad S_w(Q)^* = \mathrm{span}\{w\} \oplus \tilde{D}_w,$$

respectively.

Let

$$E \otimes F = \{v \otimes w \in R^{pq} \mid v \in E,\ w \in F\}.$$

$P \otimes Q$ is a disjoint union of smooth manifolds of the form of $E \otimes F$. The tangent space of $E \otimes F$ at a relative interior point $v \otimes w$ is

$$T_{v \otimes w}(E \otimes F) = T_v(E) \otimes \{w\} \oplus \{v\} \otimes T_w(F),$$

which is of dimension $e + f - 2$.

Noting that for two points $v_1 \otimes w_1 \in P \otimes Q$ and $v \otimes w \in E \otimes F$ that are close to each other,

$$v_1 \otimes w_1 - v \otimes w \doteq (v_1 - v) \otimes w + v \otimes (w_1 - w),$$

we see that the tangent cone of $E \otimes F$ at $v \otimes w$ is

$$\begin{aligned}
S_{v\otimes w}(P \otimes Q) &= S_v(P) \otimes \{w\} \oplus \{v\} \otimes S_w(Q) \\
&= T_v(E) \otimes \{w\} \oplus C_v \otimes \{w\} \oplus \{v\} \otimes T_w(F) \oplus \{v\} \otimes D_w \\
&= T_{v\otimes w}(E \otimes F) \oplus C_v \otimes \{w\} \oplus \{v\} \otimes D_w.
\end{aligned}$$

The dual cone is given by

$$S_{v\otimes w}(P \otimes Q)^* = \operatorname{span}\{v \otimes w\} \oplus N_{v\otimes w}(P \otimes Q),$$

where

$$\begin{aligned}
&N_{v\otimes w}(P \otimes Q) \\
&\quad = T_v(E) \otimes T_w(F) \oplus T_v(E) \otimes N_w(F) \oplus N_v(E) \otimes T_w(F) \oplus N_v(E) \otimes N_w(F) \\
&\qquad \oplus \tilde{C}_v \otimes \{w\} \oplus \{v\} \otimes \tilde{D}_w. \qquad (4.5)
\end{aligned}$$

This is of dimension

$$\begin{aligned}
&\dim\, N_{v\otimes w}(P \otimes Q) \\
&\quad = (e-1)(f-1) + (e-1)(q-f) + (p-e)(f-1) + (p-e)(q-f) \\
&\qquad + (p-e) + (q-f) \\
&\quad = pq - e - f + 1. \qquad (4.6)
\end{aligned}$$

4.2.2 The second fundamental form

Let $v = v(t) \in E$, $t = (t^i)_{1\le i\le e-1}$, be a local coordinate system of E, and let $w = w(u) \in F$, $u = (u^i)_{1\le i\le f-1}$, be a local coordinate system of F. Then $v(t) \otimes w(u)$ gives a local coordinate system of $E \otimes F$ around the relative interior point $v \otimes w$.

Let $v_i = v_i(t) = \partial v(t)/\partial t^i$, and $w_i = w_i(u) = \partial w(u)/\partial u^i$. Then the tangent space of $E \otimes F$ is given by

$$\begin{aligned}
T_{v\otimes w}(E \otimes F) &= T_v(E) \otimes \{w\} \oplus \{v\} \otimes T_w(F) \\
&= \operatorname{span}\{v_i \otimes w \mid 1 \le i \le e-1\} \oplus \operatorname{span}\{v \otimes w_i \mid 1 \le i \le f-1\}.
\end{aligned}$$

The metric

$$G = \begin{pmatrix} (v_i' v_j)_{1\le i,j\le e-1} & 0 \\ 0 & (w_k' w_l)_{1\le k,l\le f-1} \end{pmatrix} \qquad (4.7)$$

of $E \otimes F$ is an $(e+f-2) \times (e+f-2)$ block diagonal matrix. This implies that the volume element of $E \otimes F$ at $v \otimes w$ is $dv\,dw$, where dv is the volume element of E at v, and dw is the volume element of F at w.

The second derivatives of $v(t) \otimes w(u)$ are written as an $(e+f-2) \times (e+f-2)$ matrix with elements $p \times q$ matrices

$$\begin{pmatrix} (v_{ij} \otimes w)_{1\le i,j\le e-1} & (v_i \otimes w_l)_{1\le i\le e-1,\,1\le l\le f-1} \\ (v_j \otimes w_k)_{1\le k\le f-1,\,1\le j\le e-1} & (v \otimes w_{kl})_{1\le k,l\le f-1} \end{pmatrix}. \qquad (4.8)$$

Let $\tilde{v}_a$, $1 \le a \le p-e$, be a basis of $N_v(E)$, and let $\tilde{w}_b$, $1 \le b \le q-f$, be a basis of $N_w(F)$. Consider inner products of the elements of (4.8) and the elements of $N_{v\otimes w}(P \otimes Q)$. Note that the bases of $N_{v\otimes w}(P \otimes Q)$ are $v_i \otimes w_j$, $v_i \otimes \tilde{w}_a$, $\tilde{v}_a \otimes w_i$, $\tilde{v}_a \otimes \tilde{w}_b$, $\tilde{v}_a \otimes w$, and $v \otimes \tilde{w}_a$ (see (4.5)). Here it is true by definition that $v'v_i = 0$, $v'\tilde{v}_a = 0$, $v_i'\tilde{v}_a = 0$, and $w'w_i = 0$, $w'\tilde{w}_a = 0$, $w_i'\tilde{w}_a = 0$. Because P is a spherical polyhedron, $T_v(E) \oplus \mathrm{span}\{v\}$ becomes the linear hull of E, which is invariant with respect to a choice of the relative interior point $v \in E$. Then $N_v(E)$ is independent of v, and hence we can choose the bases $\tilde{v}_a = \tilde{v}_a(t)$ independent of t. Therefore

$$v_{ij}'\tilde{v}_a = (\partial v_i/\partial t^j)'\tilde{v}_a = \partial(v_i'\tilde{v}_a)/\partial t^i = 0.$$

Similarly $w_{ij}'\tilde{w}_a = 0$. Only the inner products of the form

$$(v_i \otimes w_l)'(v_j \otimes w_k) = (v_i'v_j)\,(w_k'w_l)$$

must be considered. (This is because principal directions of $E \otimes F$ are restricted in $T_v(E) \otimes T_w(F)$.)

Let $\bar{v}_i$, $1 \le i \le e-1$, be an orthonormal basis of $T_v(E)$ such that $\bar{v}_i'\bar{v}_j = \delta_{ij}$. Then a nonsingular matrix A exists such that $(v_1, \dots, v_{e-1}) = (\bar{v}_1, \dots, \bar{v}_{e-1})A$. Let $\bar{w}_i$, $1 \le i \le f-1$, be an orthonormal basis of $T_w(F)$. A nonsingular matrix B exists such that $(w_1, \dots, w_{f-1}) = (\bar{w}_1, \dots, \bar{w}_{f-1})B$. The metric G in (4.7) is written as

$$G = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}' \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}.$$

Write an element of $N_{v\otimes w}(P \otimes Q)$ as

$$r = \sum_{m=1}^{e-1}\sum_{n=1}^{f-1} r(m,n)(\bar{v}_m \otimes \bar{w}_n) + \tilde{r}, \tag{4.9}$$

where $\tilde{r}$ is an element orthogonal to $T_v(E) \otimes T_w(F)$. Let $R = (r(m,n))_{1\le m\le e-1,\,1\le n\le f-1}$ be an $(e-1)\times(f-1)$ matrix. Then the inner product of elements of (4.8) and r in (4.9) is

$$\begin{aligned}&\begin{pmatrix} 0 & \left(\sum_{m,n} r(m,n)(v_i'\bar{v}_m)(w_l'\bar{w}_n)\right)_{\substack{1\le i\le e-1\\ 1\le l\le f-1}} \\ \left(\sum_{m,n} r(m,n)(v_j'\bar{v}_m)(w_k'\bar{w}_n)\right)_{\substack{1\le k\le f-1\\ 1\le j\le e-1}} & 0 \end{pmatrix} \\ &\quad = \begin{pmatrix} 0 & A'RB \\ B'R'A & 0 \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}' \begin{pmatrix} 0 & R \\ R' & 0 \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}.\end{aligned}$$

Therefore the second fundamental form of $E \otimes F$ at $v \otimes w$ with respect to the normal direction r in (4.9) is

$$H(v \otimes w; r) = \begin{pmatrix} 0 & R \\ R' & 0 \end{pmatrix}.$$

4.2.3 Evaluation of integrals

Finally we evaluate the integral (2.13). In (2.13) we consider $E\otimes F$ instead of ∂M_d with $d=e+f-2$. The contribution of the integral with respect to $E\otimes F$ is

$$\sum_{k=0}^{e+f-2} \frac{1}{\Omega_{e+f-1-k}\Omega_{pq-e-f+1+k}} \int_{s\in E\otimes F} ds \int_{t\in N_s(P\otimes Q)\cap S^{pq-1}} dt\,\mathrm{tr}_k H(s,t)\times \bar{G}_{e+f-1-k}(x^2). \tag{4.10}$$

We begin by evaluating the integral

$$\int_{s\in E\otimes F} ds = \mathrm{Vol}(E\otimes F).$$

As we have seen, the volume element of $E\otimes F$ at $v\otimes w$ is the product measure of the volume elements dv and dw at $v\in E$ and $w\in F$. Hence if either P or Q is not symmetric with respect to the origin, then the map $(v,w)\,(\in P\times Q)\ \mapsto\ v\otimes w\,(\in P\otimes Q)$ is one-to-one, and hence

$$\int_{s\in E\otimes F} ds = \int_E dv \times \int_F dw = \Omega_e\,\Omega_f\,\beta(E)\,\beta(F), \tag{4.11}$$

where

$$\beta(E)=\frac{\mathrm{Vol}(E)}{\mathrm{Vol}(\mathrm{lin}(E)\cap S^{p-1})}, \qquad \beta(F)=\frac{\mathrm{Vol}(F)}{\mathrm{Vol}(\mathrm{lin}(F)\cap S^{q-1})}$$

are internal angles. When both P and Q are symmetric with respect to the origin, then the map $(v,w)\ \mapsto\ v\otimes w$ is two-to-one, and a multiplier $1/2$ is required in the middle and right hand sides of (4.11).

Next we consider the integral

$$\int_{t\in N_s(P\otimes Q)\cap S^{pq-1}} dt\,\mathrm{tr}_k H(s,t).$$

As in Kuriki and Takemura [23], we evaluate this by taking expectations with respect to normal random variables.

Assume that $t\in R^{pq}$ is distributed uniformly on the $(pq-e-f)$-dimensional unit sphere restricted in the linear hull $\mathrm{lin}(N_s(P\otimes Q))$ (see (4.6)). The density of t is $dt/\Omega_{pq-e-f+1}$, where dt is the volume element of $\mathrm{lin}(N_s(P\otimes Q))\cap S^{pq-1}$ at t. Assume that y^2 is a random variable distributed as $\chi^2_{pq-e-f+1}$ independent of t. Then $y\times t$ has a multivariate standard normal distribution restricted in the linear subspace $\mathrm{lin}(N_s(P\otimes Q))$. Therefore

$$\begin{aligned}
&\int_{t\in N_s(P\otimes Q)\cap S^{pq-1}} dt\,\mathrm{tr}_k H(s,t) \\
&\quad = \Omega_{pq-e-f+1}\,E[I(t\in N_s(P\otimes Q))\,\mathrm{tr}_k H(s,t)] \\
&\quad = \Omega_{pq-e-f+1}\frac{E[I(y\times t\in N_s(P\otimes Q))\,\mathrm{tr}_k H(s,y\times t)]}{E[y^k]} \\
&\quad = \frac{\Omega_{pq-e-f+1+k}}{(2\pi)^{\frac{k}{2}}} E[I(y\times t\in N_s(P\otimes Q))]\,E[\mathrm{tr}_k H(s,y\times t)],
\end{aligned} \tag{4.12}$$

where $I(\cdot)$ is the indicator function. We used $\Omega_d/E[\chi_d^k] = \Omega_{d+k}/(2\pi)^{\frac{k}{2}}$ in the last equality. Moreover,

$$E[I(y \times t \in N_s(P \otimes Q))] = \gamma(E)\gamma(F), \tag{4.13}$$

where

$$\gamma(E) = \frac{\mathrm{Vol}(E^*)}{\mathrm{Vol}(\mathrm{lin}(E)^\perp \cap S^{p-1})}, \qquad \gamma(F) = \frac{\mathrm{Vol}(F^*)}{\mathrm{Vol}(\mathrm{lin}(F)^\perp \cap S^{q-1})},$$

and

$$\begin{aligned} E^* &= \{v \in \mathrm{lin}(E)^\perp \cap S^{p-1} \mid v'\tilde{v} \le 0,\, \forall \tilde{v} \in P\}, \\ F^* &= \{w \in \mathrm{lin}(F)^\perp \cap S^{q-1} \mid w'\tilde{w} \le 0,\, \forall \tilde{w} \in Q\}, \end{aligned}$$

are external angles. Let $R = (r_{ij})$ be an $(e-1) \times (f-1)$ random matrix such that all elements r_{ij}s are independent standard normal random variables. Then

$$\begin{aligned} E[\mathrm{tr}_k H(s, y \times t)] &= E\Bigl[\mathrm{tr}_k \begin{pmatrix} O & R \\ R' & 0 \end{pmatrix}\Bigr] \\ &= \begin{cases} (-1)^{k/2}\binom{e-1}{k/2}\binom{f-1}{k/2}(k/2)!, & \text{for } k \text{ even}, \\ 0, & \text{for } k \text{ odd} \end{cases} \end{aligned} \tag{4.14}$$

(Kuriki and Takemura [23]).

Combining (4.11)–(4.14), the contribution (4.10) of $E \otimes F$ becomes

$$\begin{aligned} \sum_{k=0,\,k:\text{even}}^{2(\min(e,f)-1)} & \frac{\Omega_e\, \Omega_f}{(2\pi)^{\frac{k}{2}}\, \Omega_{e+f-1-k}}\, \beta(E)\, \beta(F)\, \gamma(E)\gamma(F) \\ & \times (-1)^{k/2}\binom{e-1}{k/2}\binom{f-1}{k/2}(k/2)!\, \bar{G}_{e+f-1-k}(x^2). \end{aligned} \tag{4.15}$$

Summing (4.15) over $0 \le e \le p$ and $0 \le f \le q$, and noting that

$$w_e(P) = \sum_{E:\,\dim E = e-1} \beta(E)\, \gamma(E), \qquad w_f(Q) = \sum_{F:\,\dim F = f-1} \beta(F)\, \gamma(F)$$

(Takemura and Kuriki [34]), we obtain the expression (2.14).

4.2.4 Approximation by sequences of spherical polyhedra

Finally we prove that (2.14) still holds when P and Q are non-polyhedral.

Define a distance between two subsets $M_1, M_2 \in S^{n-1}$ by

$$\delta(M_1, M_2) = \inf\{\theta \ge 0 \mid M_1 \subset (M_2)_\theta,\, M_2 \subset (M_1)_\theta\}.$$

This is the Hausdorff distance in R^n between $\mathrm{cone}(M_i) \cap B^n$, $i = 1, 2$, where $\mathrm{cone}(M_i)$ is the smallest cone containing M_i, and B^n is the unit ball in R^n (Takemura and Kuriki [34]). The following theorem states the continuity of the volume of the tube with respect to the distance $\delta(\cdot, \cdot)$. This is a spherical version of Theorems 5.6 and 5.9 of Federer [12]. The proof is parallel to that of Federer [12] and is omitted.

Theorem 4.2 *Suppose that $\epsilon > 0$. Let A_i, $i = 1, 2, \ldots$, and B be closed subsets of S^{n-1} with the critical radii $\theta_c(A_i), \theta_c(B) \geq \epsilon$ such that $\delta(A_i, B) \to 0$ as $i \to \infty$. Assume that the volume of the tube around A_i is written as*

$$\mathrm{Vol}((A_i)_\theta) = \Omega_n \sum_{e=1}^{n} w_e(A_i) \bar{B}_{\frac{1}{2}e, \frac{1}{2}(n-e)}(\cos^2\theta), \quad \theta \leq \theta_c(A_i).$$

Then the limits $w_e(A_i) \to w_e(B)$, $i \to \infty$ exist, and the volume of tube around B is written as

$$\mathrm{Vol}(B_\theta) = \Omega_n \sum_{e=1}^{n} w_e(B) \bar{B}_{\frac{1}{2}e, \frac{1}{2}(n-e)}(\cos^2\theta), \quad \theta \leq \theta_c(B).$$

Lemma 4.2 *Let $P, \tilde{P} \subset S^{p-1}$, $Q, \tilde{Q} \subset S^{q-1}$. Then*

$$\delta(P \otimes Q, \tilde{P} \otimes \tilde{Q}) \leq 2 \max\{\delta(P, \tilde{P}), \delta(Q, \tilde{Q})\}.$$

Proof. Assume that $\delta(P, \tilde{P}) \leq c$, $\delta(Q, \tilde{Q}) \leq c$. Let $x \in P_c$, $y \in Q_c$. Because $x'v \geq \cos(c)$, $y'w \geq \cos(c)$, $\exists v \in P$, $\exists w \in Q$, it is true that $(x \otimes y)'(v \otimes w) = (x'v)(y'w) \geq \cos^2(c) \geq \cos(2c)$. This implies $x \otimes y \in (P \otimes Q)_{2c}$. Therefore $(\tilde{P} \otimes \tilde{Q}) \subset (P_c \otimes Q_c) \subset (P \otimes Q)_{2c}$, $(P \otimes Q) \subset (\tilde{P} \otimes \tilde{Q})_{2c}$, and $\delta(P \otimes Q, \tilde{P} \otimes \tilde{Q}) \leq 2c$ follows. ∎

By the spherical convexity of P and Q, sequences of spherical polyhedra P_i, Q_i, $i = 1, 2, \ldots$ exist such that $\delta(P_i, P) \to 0$, $\delta(Q_i, Q) \to 0$ as $i \to \infty$ (Takemura and Kuriki [34], Lemma 1.2). By Lemma 4.2, $P_i \otimes Q_i \subset R^{pq}$, $i = 1, 2, \ldots$, is a sequence such that $\delta(P_i \otimes Q_i, P \otimes Q) \to 0$ as $i \to \infty$.

Let $\phi(P_i)$ and $\phi(Q_i)$ be the diameters of P_i and Q_i. Because $|\phi(P_i) - \phi(P)| \leq 2\delta(P_i, P)$, $|\phi(Q_i) - \phi(Q)| \leq 2\delta(Q_i, Q)$, it is true that $\phi(P_i) \to \phi(P)$, $\phi(Q_i) \to \phi(Q)$. It follows from Theorem 2.7 that the critical radius $\theta_c(P_i \otimes Q_i)$ converges to $\theta_c(P \otimes Q)$ (> 0).

We have seen that the assumptions of Theorem 4.2 are fulfilled. We can conclude that the expression (2.14) is valid for non-polyhedral cones P and Q.

4.3 The proof of Theorem 2.7

We evaluate the critical radius θ_c by (2.10). In this proof, the point $v \otimes w \in R^{pq}$ is represented as a matrix $vw' \in R^{p \times q}$. With this change, we introduce a new symbol "$\odot$" defined by

$$A \odot B = \{vw' \in R^{p \times q} \mid v \in A,\ w \in B\} \qquad \text{for } A \subset R^p,\ B \subset R^q.$$

As in Section 4.2, we denote the orthogonal direct sum by "$\oplus$".

Let Z be a point in $R^{p \times q}$. Let $N_v(P) = S_v(P)^* \cap \mathrm{span}\{v\}^\perp \subset R^p$ and $N_w(Q) = S_w(Q)^* \cap \mathrm{span}\{w\}^\perp \subset R^q$. By (4.5),

$$N_{vw'}(P \odot Q) = (\mathrm{span}\{v\}^\perp \odot \mathrm{span}\{w\}^\perp) \oplus N_v(P)w' \oplus vN_w(Q)',$$

where

$$N_v(P)w' = \{\tilde{v}w' \mid \tilde{v} \in N_v(P)\}, \qquad vN_w(Q)' = \{v\tilde{w}' \mid \tilde{w} \in N_w(Q)\}.$$

Therefore the orthogonal projection of Z onto $N_{vw'}(P \odot Q)$ is given by

$$\mathcal{P}(Z \mid N_{vw'}) = (I - vv')Z(I - ww') + \mathcal{P}(Z \mid N_v(P)w') + \mathcal{P}(Z \mid vN_w(Q)'). \tag{4.16}$$

Here the second term in the right hand side of (4.16) above is rewritten as

$$\mathcal{P}(Zw \mid N_v(P))w', \tag{4.17}$$

because for $x \in N_v(P)$,

$$\begin{aligned} \|Z - xw'\|^2 &= \|Z(I - ww')\|^2 + \|(Zw - x)w'\|^2 \\ &= \|Z(I - ww')\|^2 + \|Zw - x\|^2. \end{aligned}$$

($\|\cdot\|$ appearing in the second term of the right hand side means the norm for vectors. The other $\|\cdot\|$s mean the norms for matrices.) Similarly the third term in the right hand side of (4.16) is $v\,\mathcal{P}(Z'v \mid N_v(Q))'$.

Set $Z = s - t = \tilde{v}\tilde{w}' - vw'$, where $\tilde{v}\tilde{w}', vw' \in P \odot Q$. Then (4.17) is reduced to $\mathcal{P}((\tilde{w}'w)\tilde{v} - v \mid N_v(P))w'$. Noting that

$$\|(I - vv')(\tilde{v}\tilde{w}' - vw')(I - ww')\|^2 = (1 - (\tilde{v}'v)^2)(1 - (\tilde{w}'w)^2),$$

$$\|\tilde{v}\tilde{w}' - vw'\|^2 = 2(1 - (\tilde{v}'v)(\tilde{w}'w)),$$

the left hand side of (2.10) is reduced to

$$\inf_{\tilde{v},v\in P,\,\tilde{w},w\in Q} \frac{(1 - \tau\rho)^2}{(1 - \tau^2)(1 - \rho^2) + \|\,\mathcal{P}(\rho\tilde{v} - v \mid N_v(P))\|^2 + \|\,\mathcal{P}(\tau\tilde{w} - w \mid N_w(Q))\|^2} \tag{4.18}$$

with $\tau = \tilde{v}'v$, $\rho = \tilde{w}'w$.

Case I. Consider the case $\phi(P) \le \pi/2$. This is equivalent to $\tau = \tilde{v}'v \ge 0$, $\forall \tilde{v}, v \in P$. The smallest cone containing Q is denoted by cone(Q). Because $\tau \ge 0$, $\tau\tilde{w} \in$ cone(Q). On the other hand, because $N_w(Q)$ is the normal cone of the convex cone(Q) at w, we have

$$\mathcal{P}(\tau\tilde{w} - w \mid N_w(Q)) = 0.$$

Therefore, (4.18) is reduced to

$$\inf_{\tilde{v},v\in P,\,\tilde{w},w\in Q} \frac{(1 - \tau\rho)^2}{(1 - \tau^2)(1 - \rho^2) + \|\,\mathcal{P}(\rho\tilde{v} - v \mid N_v(P))\|^2}. \tag{4.19}$$

We evaluate the infimum in (4.19) by examining the cases $\rho > 0$ and $\rho \le 0$ separately.

The case $\rho > 0$. Because $\mathcal{P}(\rho\tilde{v} - v \mid N_v(P)) = 0$, the argument of the infimum in (4.19) becomes

$$\frac{(1-\tau\rho)^2}{(1-\tau^2)(1-\rho^2)}.$$

The infimum of this is one, which is attained as $|\tau - \rho| \to 0$ with $\tau, \rho \neq 1$. (Lemma 4.5 of Kuriki and Takemura [23]). Hence if a sequence $\tilde{v}'\tilde{w} \to v'w > 0$ with $\tilde{v} \neq v$, $\tilde{w} \neq w$ exists, then the infimum is attained. This is possible when $\dim P \geq 1$ and $\dim Q \geq 1$.

The case $\rho \leq 0$. Because $N_v(P) \subset \mathrm{span}\{v\}^{\perp}$,

$$\begin{aligned} \| \mathcal{P}(\rho\tilde{v} - v \mid N_v(P))\|^2 \leq \| \mathcal{P}(\rho\tilde{v} - v \mid \mathrm{span}\{v\}^{\perp})\|^2 &= \|(I - vv')(\rho\tilde{v} - v)\|^2 \\ &= \rho^2(1-\tau^2). \end{aligned} \tag{4.20}$$

Therefore the argument of the infimum in (4.19) is bounded below by

$$\frac{(1-\tau\rho)^2}{(1-\rho^2)(1-\tau^2) + \rho^2(1-\tau^2)} = \frac{(1-\tau\rho)^2}{1-\tau^2} \geq \frac{1}{1-\tau^2} \geq 1.$$

Now we have proved that (4.19) is equal to one, and hence $\tan^2\theta_c = 1$, $\theta_c = \pi/4$.

Case II. Consider the case $\phi(P) > \pi/2$ and $\phi(Q) > \pi/2$. Both $\tau = \tilde{v}'v$ and $\rho = \tilde{w}'w$ can take minus values. According to (4.20), the argument of the infimum in (4.18) with τ, $\rho < 0$ is bounded below by

$$\frac{(1-\tau\rho)^2}{(1-\rho^2)(1-\tau^2) + \rho^2(1-\tau^2) + \tau^2(1-\rho^2)} = \frac{1-\tau\rho}{1+\tau\rho} \geq \frac{1-\cos\phi(P)\cos\phi(Q)}{1+\cos\phi(P)\cos\phi(Q)}.$$

The right hand side is less than one. Together with the examinations in Case I, we have

$$\tan^2\theta_c \geq \frac{1-\cos\phi(P)\cos\phi(Q)}{1+\cos\phi(P)\cos\phi(Q)}. \tag{4.21}$$

In the following we see that the equality in (4.21) holds.

Let $v, \tilde{v} \in P$ be a pair of points such that $\cos^{-1}(\tilde{v}'v) = \phi(P)$. Let $w, \tilde{w} \in Q$ be a pair of points such that $\cos^{-1}(\tilde{w}'w) = \phi(Q)$. Then $\tau = \cos\phi(P) < 0$, $\rho = \cos\phi(Q) < 0$. Here we claim that

$$\mathcal{P}(\rho\tilde{v} - v \mid \mathrm{span}\{v\}^{\perp}) = \rho(\tilde{v} - \tau v) \in N_v(P).$$

Assume that $\rho(\tilde{v} - \tau v) \notin N_v(P)$. Because $\rho < 0$ and $N_v(P) = S_v(\mathrm{cone}(P))^*$, $u \in \mathrm{cone}(P)$ exists such that $(\tilde{v} - \tau v)'u < 0$. Because $\mathrm{cone}(P)$ is convex,

$$v_\epsilon = \frac{(1-\epsilon)v + \epsilon u}{\|(1-\epsilon)v + \epsilon u\|} \in P, \quad \text{for} \;\; 0 \leq \epsilon \leq 1.$$

Let $h(\epsilon) = (\tilde{v}'v_\epsilon)^2$ be a function in $\epsilon \in [0, 1]$. Then $(d/d\epsilon)h(0+) = 2\tau(\tilde{v} - \tau v)'u > 0$. This implies that for a sufficiently small $\epsilon > 0$, $0 > \tilde{v}'v > \tilde{v}'v_\epsilon$, and $\cos^{-1}(\tilde{v}'v) < \cos^{-1}(\tilde{v}'v_\epsilon)$. This contradicts the assumption that $\cos^{-1}(\tilde{v}'v) = \phi(P)$.

Therefore, we have

$$\mathcal{P}(\rho\tilde{v} - v \mid N_v(P)) = \mathcal{P}(\mathcal{P}(\rho\tilde{v} - v \mid \mathrm{span}\{v\}^\perp \mid N_v(P))) = \rho(\tilde{v} - \tau v),$$

and $\mathcal{P}(\tau\tilde{w} - w \mid N_w(Q)) = \tau(\tilde{w} - \rho w)$. Substituting them into the argument of the infimum in (4.18), we see that the lower bound in (4.21) is really attained.

Acknowledgments

The author thanks Akimichi Takemura for pointing out that Theorem 2.6 holds for non-polyhedral cones, Stephen E. Fienberg and Tadahiko Maeda for providing several important references, Takemi Yanagimoto and Hidetoshi Shimodaira for helpful comments on the exposition, and Chihiro Hirotsu for bringing the author's attention to the correspondence analysis under order restrictions. He is also grateful for the constructive comments of an anonymous referee, which improved the original manuscript.

References

[1] A. Agresti, "Categorical Data Analysis," 2nd ed., Wiley, New York, 2002.

[2] R. E. Barlow, D. J. Bartholomew, J. M. Bremner, and H. D. Brunk, "Statistical Inference Under Order Restrictions," Wiley, London, 1972.

[3] M. P. Becker, Maximum likelihood estimation of the RC(M) association model, *Appl. Statist.*, **39** (1990), 152–167.

[4] R. H. Berk, Consistency and asymptotic normality of MLE's for exponential models, *Ann. Math. Statist.*, **43** (1972), 193–204. Acknowledgement of priority and correction, *Ann. Statist.*, **1** (1973), 593.

[5] H. Chernoff, On the distribution of the likelihood ratio, *Ann. Math. Statist.*, **25** (1954), 573–578.

[6] S. Das and P. K. Sen, Restricted canonical correlations, *Linear Algebra Appl.*, **210** (1994), 29–47.

[7] S. Das and P. K. Sen, Asymptotic distribution of restricted canonical correlations and relevant resampling methods, *J. Multivariate Anal.*, **56** (1996), 1–19.

[8] R. Douglas and S. E. Fienberg, An overview of dependency models for cross-classified categorical data involving ordinal variables, in "Topics in statistical dependence" (H. W. Block, A. R. Sampson, and T. H. Savits, Eds), 167–188, IMS Lecture Notes Vol. 16, Inst. Math. Statist., Hayward, 1990.

[9] M. Eaton and D. Tyler, The asymptotic distribution of singular values with applications to canonical correlations and correspondence analysis, *J. Multivariate Anal.*, **50** (1994), 238–264.

[10] R. D. Etzioni, S. E. Fienberg, Z. Gilula, and S. J. Haberman, Statistical models for the analysis of ordered categorical data in public health and medical research, *Statistical Methods in Medical Research*, **3** (1994), 179–204.

[11] B. S. Everitt, "The Analysis of Contingency Tables," 2nd ed, Chapman & Hall, London, 1992.

[12] H. Federer, Curvature measures, *Trans. Amer. Math. Soc.*, **93** (1959), 418–491.

[13] Z. Gilula and Y. Ritov, Inferential ordinal correspondence analysis: motivation, derivation and limitations, *Internat. Statist. Rev.*, **58** (1990), 99–108.

[14] L. A. Goodman, Simple models for the analysis of association in cross-classifications having ordered categories, *J. Amer. Statist. Assoc.*, **74** (1979), 537–552.

[15] L. A. Goodman, Association models and canonical correlation in the analysis of cross-classifications having ordered categories, *J. Amer. Statist. Assoc.*, **76** (1981), 320–334.

[16] L. A. Goodman, The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries, *Ann. Statist.*, **13** (1985), 10–69.

[17] S. J. Haberman, Tests for independence in two-way contingency tables based on canonical correlation and on linear-by-linear interaction, *Ann. Statist.*, **9** (1981), 1178–1186.

[18] C. Hirotsu, Use of cumulative efficient scores for testing ordered alternatives in discrete models, *Biometrika*, **69** (1982), 567–577.

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

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

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

[22] S. Kuriki and A. Takemura, Distribution of the maximum of Gaussian random field: tube method and Euler characteristic method, *Proc. Inst. Statist. Math.*, **47** (1999), 199–219 (in Japanese).

[23] S. Kuriki and A. Takemura, Tail probabilities of the maxima of multilinear forms and their applications, *Ann. Statist.*, **29** (2001), 328–371.

[24] T. Miwa, A. J. Hayter and W. Liu, Calculations of level probabilities for normal random variables with unequal variances with applications to Bartholomew's test in unbalanced one-way models, *Comput. Statist. Data Anal.*, **34** (2000), 17–32.

[25] D. Q. Naiman, Volumes of tubular neighborhoods of spherical polyhedra and statistical inference, *Ann. Statist.*, **18** (1990), 685–716.

[26] S. Nishisato and P. S. Arri, Nonlinear programming approach to optimal scaling of partially ordered categories, *Psychometrika*, **40** (1975), 525–548.

[27] M. E. O'Neill, Asymptotic distributions of the canonical correlations from contingency tables, *Austral. J. Statist.*, **20** (1978), 75–82.

[28] Y. Ritov and Z. Gilula, The order-restricted RC model for ordered contingency tables: estimation and testing for fit, *Ann. Statist.*, **19** (1991), 2090–2101.

[29] Y. Ritov and Z. Gilula, Analysis of contingency tables by correspondence models subject to order constraints, *J. Amer. Statist. Assoc.*, **88** (1993), 1380–1387.

[30] T. Saito and T. Otsu, A method of optimal scaling for multivariate ordinal data and its extensions, *Psychometrika*, **53** (1988), 5–25, 593 (errata).

[31] S. G. Self and K-Y. Liang, Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *J. Amer. Statist. Assoc.*, **82** (1987), 605–610.

[32] A. Shapiro, On differentiability of metric projections in R^n. I. Boundary case. *Proc. Amer. Math. Soc.*, **99** (1987), 123–128.

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

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

[35] A. Takemura and S. Kuriki, On the equivalence of the tube and Euler characteristic methods for the distribution of the maximum of Gaussian fields over piecewise smooth domains, *Ann. Appl. Probab.*, **12** (2002), 768–796.

[36] Y. Tanaka, Optimal scaling for arbitrarily ordered categories, *Ann. Inst. Statist. Math.*, **31** (1979), 115–124.

[37] H. Weyl, On the volume of tubes, *Amer. J. Math.*, **61** (1939), 461–472.

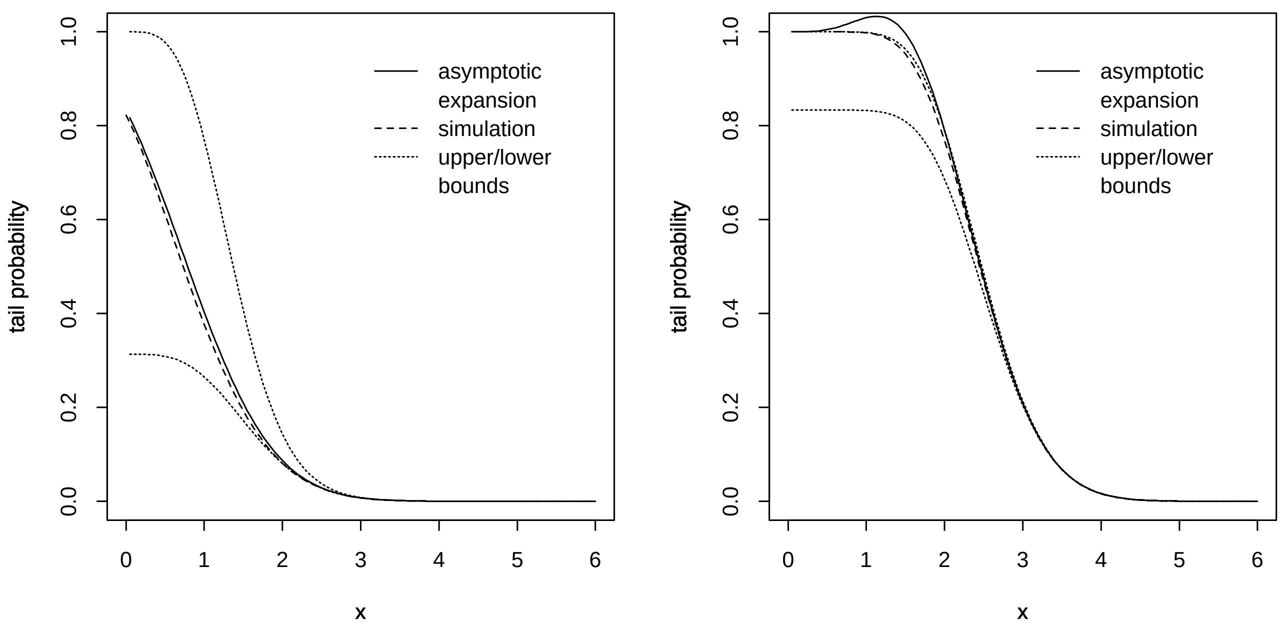

Figure 2.1. The upper tail probability $P(T \geq x)$.

Table 3.1. Income and job satisfaction ($n = 901$).

Income (dollars) \ Job satisfaction	Very dissatisfied	Little dissatisfied	Moderately satisfied	Very satisfied
< 6000	20 (19.7)	24 (26.5)	80 (78.2)	82 (81.6)
$6000 - 15000$	22 (22.2)	38 (35.8)	104 (105.7)	125 (125.3)
$15000 - 25000$	13 (13.3)	28 (27.4)	81 (80.9)	113 (113.4)
> 25000	7 (6.8)	18 (18.4)	54 (54.2)	92 (91.7)

Note: From Agresti [1]. Estimated frequencies are in parentheses.

Table 3.2. Estimates of ϕ, μ_i, ν_j.

	$\hat{\phi}$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	$\hat{\nu}_1$	$\hat{\nu}_2$	$\hat{\nu}_3$	$\hat{\nu}_4$
CA	0.1125	–1.31	–0.44	0.56	1.56	–2.55	0.89	–1.05	0.39
CA (ordered)	0.0979	–1.21	–0.51	0.48	1.65	–0.92	–0.92	–0.92	1.09
RC	0.1161	–1.26	–0.48	0.51	1.62	–2.63	–0.37	–0.57	0.93
RC (ordered)	0.1160	–1.24	–0.49	0.51	1.63	–2.63	–0.52	–0.52	0.93

Table 3.3. Tests for independence.

	test statistic	df	p-value
CA	11.40		0.0945
CA (ordered)	8.64		0.0254
RC	11.59		0.0882
RC (ordered)	11.55		0.0069
Saturated	12.04	9	0.2112
Pearson χ^2	11.99	9	0.2140

CA : Test based on $n\,\hat{\phi}^2$
RC : LRT against the RC model
Saturated : LRT against the saturated model