Journal of Multivariate Analysis 171 (2019) 193–208

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Transformed statistics for tests of conditional independence in $J \times K \times L$ contingency tables

Nobuhiro Taneichi [a,*], Yuri Sekiya [b], Jun Toyama [c]

[a] *Hokkaido University of Education, Sapporo Campus, 1-5, Ainosato 5-3, Kita-ku, Sapporo 002-8502, Japan*
[b] *Hokkaido University of Education, Kushiro Campus, 15-55, Shiroyama 1, Kushiro 085-8580, Japan*
[c] *The Institute for the Practical Application of Mathematics, Sapporo 063-0001, Japan*

ARTICLE INFO

Article history:
Received 1 September 2017
Available online 7 December 2018

AMS 2010 subject classifications:
primary 62E20
secondary 62H10

Keywords:
Bartlett adjustment
Bartlett-type adjustment
Chi-square limiting distribution
Improved transformation
Test of conditional independence
Three-way contingency table

ABSTRACT

Tests of the hypothesis of conditional independence in $J \times K \times L$ contingency tables are considered. An expression to approximate the null distribution of the test statistics is derived. Using this expression, transformed statistics are obtained which converge to a chi-square limiting distribution faster than the original statistics do. Simulations are used to compare the transformed statistics with the original ones, and transformed statistics are proposed based on a Bartlett-type adjustment. Through this work and earlier ones, we cover testing hierarchical loglinear models in three-way tables except for a model having three interaction terms.

1. Introduction

In order to motivate this paper, consider the data in Table 1, arising from an investigation conducted at The Danish National Institute for Social Research in Copenhagen [6]. In this table, 223 employed men in the age group 18–67 years were cross-classified according to whether they live in a house or apartment (Mode: Rent, Own) and whether or not (Response: Yes, No) they had done any work in the preceding 12 months which, in earlier times, they would have paid a craftsman to do. One of the purposes of the investigation, for which the data were collected in 1978–79, was to estimate the extent of tax evasion in the construction industry.

In the table, there are many possible relations between the factors: (A1) Response, Age, and Mode are all independent; (B1) Response is independent of the pair (Age, Mode); (B2) Age is independent of the pair (Response, Mode); (B3) Mode is independent of the pair (Response, Age); (C1) Given any Mode, Response and Age are independent; (C2) Given any Age, Response and Mode are independent; (C3) Given any Response, Age and Mode are independent.

To test hypothesis (C1) in the data from Table 1, say, the following procedure would usually be carried out. First, the value of the log likelihood ratio test statistic G^2 would be computed; the statistic is a special case of Eq. (6). Here, the observed value of G^2 is 9.637. Since G^2 is asymptotically distributed as χ^2_4, (C1) is rejected at the asymptotic 5% level given that $\chi^2_4(0.05) = 9.488$; the asymptotic p-value for G^2 is 0.047.

Of course, such a conclusion is only valid on the basis of the assumption that the distribution of the test statistic G^2 is suitably approximated by a chi-square distribution with appropriate degrees of freedom. If the asymptotically approximated

* Corresponding author.
E-mail addresses: taneichi.nobuhiro@s.hokkyodai.ac.jp (N. Taneichi), sekiya.yuri@k.hokkyodai.ac.jp (Y. Sekiya), mandheling@nifty.com (J. Toyama).

https://doi.org/10.1016/j.jmva.2018.12.001

Table 1
A three-way table from Edwards and Kreiner [6].

Response	Mode	Rent			Own		
	Age	< 30	31–45	46–67	< 30	31–45	46–67
Yes		29	3	7	23	52	49
No		44	13	16	9	31	51

critical point $\chi^2_\nu(\alpha)$ and approximate p-value based on χ^2_ν are inaccurate, there is a risk that the results of the approximated test stated above will lead to the opposite conclusion of the test given by an exact critical point. We illustrate the application of our proposed method with the above data in Section 8.

In this paper, we consider approximations of the distribution of test statistics based on an asymptotic expansion that is more accurate than the asymptotic approximation. To this end, we develop approximations based on an asymptotic expansion of the law of test statistics whose distribution is, like G^2, discrete.

For goodness-of-fit tests for a multinomial distribution, Yarnold [24] obtained an asymptotic expansion for the null distribution of Pearson's X^2 statistic. The expansion consists of continuous and discontinuous terms. Yarnold [24] numerically examined the accuracy of (i) the chi-square approximation; (ii) an Edgeworth approximation for the null distribution of X^2; and (iii) an approximation based on Edgeworth expansion and discontinuous terms. He proposed the use of an approximation based on an Edgeworth expansion and discontinuous terms. In a fashion similar to the X^2 statistic, approximations based on asymptotic expansions for null distributions of the log likelihood ratio test statistic and the Freeman–Tukey statistic were obtained by Siotani and Fujikoshi [17]. An approximation of power divergence statistics was obtained by Read [14] and an approximation of ϕ-divergence statistics was obtained by Menéndez et al. [12]. The numerical accuracy of the approximation was shown by Yarnold [24] for the X^2 statistic and by Read [15] for power divergence statistics.

From the numerical results obtained by Yarnold [24], we notice that the chi-square approximation rarely performs better than the Edgeworth approximation. Thus, an Edgeworth approximation appears to be an effective approximation of the null distribution of the above statistics when the discontinuous term in the asymptotic expansion cannot be expressed in a simple form. When it can, Yarnold's recommendation applies. Under alternatives, it is very difficult to represent the discontinuous term in a simple form, both for the earlier mentioned statistics and for more general multinomial models such as contingency tables. A mathematical explanation is provided in Taneichi et al. [21] and Taneichi and Sekiya [19]. Edgeworth approximations of the distributions of specific multinomial goodness-of-fit statistics under alternative hypotheses were investigated in [16,20,21]. Taneichi and Sekiya [19] discussed approximations for the distribution of statistics used to test for independence in $r \times s$ contingency tables. Based on numerical investigations, it was found that omission of the discontinuous term does not lead to a serious error.

In a multidimensional contingency table, Taneichi et al. [22] derived approximation of the distribution of statistics for testing hypothesis (A1) based on an asymptotic expansion. In an $r \times s \times t$ contingency table, Kobe et al. [10] derived approximations of statistics for testing hypotheses (B1)–(B3) based on an asymptotic expansion. In this paper, we derive an approximation of the distribution of statistics for testing hypotheses (C1)–(C3) based on an asymptotic expansion. Using the continuous term of expansion (multivariate Edgeworth expansion), we construct transformed statistics that are more reliable than the original statistics.

2. Notation, test statistics and outline of paper

We consider a three-way $J \times K \times L$ contingency table. For any $j \in \{1, \ldots, J\}$, $k \in \{1, \ldots, K\}$, and $\ell \in \{1, \ldots, L\}$, let $X_{jk\ell}$ be frequency of cell (j, k, ℓ), and assume that the sum n of all frequencies is fixed. A generic $J \times K \times L$ contingency table is shown in Table 2. Assume that the random vector

$$X = (X_{111}, \ldots, X_{11L}, \ldots, X_{1K1}, \ldots, X_{1KL}, \ldots, X_{J11}, \ldots, X_{J1L}, \ldots, X_{JK1}, \ldots, X_{JKL})^\top$$

is distributed according to a multinomial distribution $\mathcal{M}_{JKL}(n, p)$, where

$$p = (p_{111}, \ldots, p_{11L}, \ldots, p_{1K1}, \ldots, p_{1KL}, \ldots, p_{J11}, \ldots, p_{J1L}, \ldots, p_{JK1}, \ldots, p_{JKL})^\top,$$

with $p_{jk\ell} \in (0, 1)$ for all $j \in \{1, \ldots, J\}$, $k \in \{1, \ldots, K\}$, and $\ell \in \{1, \ldots, L\}$. Let the marginal probabilities of rows, columns and layers be

$$p_{j\cdot\cdot} = \sum_{k=1}^{K}\sum_{\ell=1}^{L} p_{jk\ell}, \quad p_{\cdot k\cdot} = \sum_{j=1}^{J}\sum_{\ell=1}^{L} p_{jk\ell}, \quad p_{\cdot\cdot\ell} = \sum_{j=1}^{J}\sum_{k=1}^{K} p_{jk\ell}, \quad p_{jk\cdot} = \sum_{\ell=1}^{L} p_{jk\ell}, \quad p_{j\cdot\ell} = \sum_{k=1}^{K} p_{jk\ell}, \quad p_{\cdot k\ell} = \sum_{j=1}^{J} p_{jk\ell},$$

respectively. When $\Pr(\text{layer} = \ell) > 0$, the probability of row j and column k given that the layer is ℓ is

$$\Pr(\text{row} = j,\ \text{column} = k \mid \text{layer} = \ell) = \Pr(\text{row} = j,\ \text{column} = k,\ \text{layer} = \ell)/\Pr(\text{layer} = \ell) = p_{jk\ell}/p_{\cdot\cdot\ell}. \quad (1)$$

Table 2
A generic $J \times K \times L$ contingency table.

	Table of 1st layer			$\cdots$	Table of Lth layer			Total
	X_{111}	$\cdots$	X_{1K1}	$\cdots$	X_{11L}	$\cdots$	X_{1KL}	$X_{1\cdot\cdot}$
	$\vdots$		$\vdots$	$\cdots$	$\vdots$		$\vdots$	$\vdots$
	X_{j11}	$\cdots$	X_{jK1}	$\cdots$	X_{j1L}	$\cdots$	X_{jKL}	$X_{j\cdot\cdot}$
	$\vdots$		$\vdots$	$\cdots$	$\vdots$		$\vdots$	$\vdots$
	X_{J11}	$\cdots$	X_{JK1}	$\cdots$	X_{J1L}	$\cdots$	X_{JKL}	$X_{J\cdot\cdot}$
Total	$X_{\cdot 11}$	$\cdots$	$X_{\cdot K1}$	$\cdots$	$X_{\cdot 1L}$	$\cdots$	$X_{\cdot KL}$	n

When $\Pr(\text{layer} = \ell) > 0$ for all $\ell \in \{1, \ldots, L\}$, conditional independence of rows and columns for each layer means that for all $j \in \{1, \ldots, J\}$, $k \in \{1, \ldots, K\}$, and $\ell \in \{1, \ldots, L\}$,

$$\Pr(\text{row} = j,\ \text{column} = k \mid \text{layer} = \ell) = \Pr(\text{row} = j \mid \text{layer} = \ell) \times \Pr(\text{column} = k \mid \text{layer} = \ell) = \frac{p_{j\cdot\ell}}{p_{\cdot\cdot\ell}} \times \frac{p_{\cdot k\ell}}{p_{\cdot\cdot\ell}}. \tag{2}$$

From Eqs. (1)–(2), when $p_{\cdot\cdot\ell} > 0$ for all $\ell \in \{1, \ldots, L\}$, the null hypothesis that rows and columns are independent given the layers is

$$\mathcal{H}_0^{(1)} : \forall_{j\in\{1,\ldots,J\}}\forall_{k\in\{1,\ldots,K\}}\forall_{\ell\in\{1,\ldots,L\}} \quad p_{jk\ell} = p_{j\cdot\ell}p_{\cdot k\ell}/p_{\cdot\cdot\ell} \tag{3}$$

Similarly, the null hypothesis that rows and layers are independent given columns is

$$\mathcal{H}_0^{(2)} : \forall_{j\in\{1,\ldots,J\}}\forall_{k\in\{1,\ldots,K\}}\forall_{\ell\in\{1,\ldots,L\}} \quad p_{jk\ell} = p_{jk\cdot}p_{\cdot k\ell}/p_{\cdot k\cdot}, \tag{4}$$

and the null hypothesis that columns and layers are independent given rows is

$$\mathcal{H}_0^{(3)} : \forall_{j\in\{1,\ldots,J\}}\forall_{k\in\{1,\ldots,K\}}\forall_{\ell\in\{1,\ldots,L\}} \quad p_{jk\ell} = p_{jk\cdot}p_{j\cdot\ell}/p_{j\cdot\cdot}. \tag{5}$$

The unrestricted maximum likelihood estimator of $p_{jk\ell}$ is $\tilde{p}_{jk\ell} = X_{jk\ell}/n$. Maximum likelihood estimators of $p_{jk\ell}$ under $\mathcal{H}_0^{(m)}$ with $m \in \{1, 2, 3\}$ are

$$\hat{p}_{jk\ell}^{(1)} = X_{j\cdot\ell}X_{\cdot k\ell}/(nX_{\cdot\cdot\ell}), \quad \hat{p}_{jk\ell}^{(2)} = X_{jk\cdot}X_{\cdot k\ell}/(nX_{\cdot k\cdot}), \quad \hat{p}_{jk\ell}^{(3)} = X_{jk\cdot}X_{j\cdot\ell}/(nX_{j\cdot\cdot}),$$

respectively, where

$$X_{j\cdot\cdot} = \sum_{k=1}^{K}\sum_{\ell=1}^{L} X_{jk\ell}, \quad X_{\cdot k\cdot} = \sum_{j=1}^{J}\sum_{\ell=1}^{L} X_{jk\ell}, \quad X_{\cdot\cdot\ell} = \sum_{j=1}^{J}\sum_{k=1}^{K} X_{jk\ell}, \quad X_{jk\cdot} = \sum_{\ell=1}^{L} X_{jk\ell}, \quad X_{j\cdot\ell} = \sum_{k=1}^{K} X_{jk\ell}, \quad X_{\cdot k\ell} = \sum_{j=1}^{J} X_{jk\ell}.$$

The ϕ-divergence statistics for testing $\mathcal{H}_0^{(m)}$ with $m \in \{1, 2, 3\}$ are

$$C_\phi^{(m)} = 2n\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{\ell=1}^{L} \hat{p}_{jk\ell}^{(m)}\phi(\tilde{p}_{jk\ell}/\hat{p}_{jk\ell}^{(m)}),$$

where $\phi(t)$ is a convex function for $t > 0$ which satisfies $\phi(1) = \phi'(1) = 0$ and $\phi''(1) = 1$; see [13,25]. Let ϕ_a be

$$\phi_a(t) = \begin{cases} \{t^{a+1} - t + a(1-t)\}/\{a(a+1)\} & \text{if } a \notin \{0, -1\}, \\ t\ln t + 1 - t & \text{if } a = 0, \\ -\ln t - 1 + t & \text{if } a = -1. \end{cases}$$

Then for each $m \in \{1, 2, 3\}$, $C_{\phi_a}^{(m)}$ reduces to

$$R_{(m)}^a \equiv C_{\phi_a}^{(m)} = 2n\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{\ell=1}^{L} I^a(\tilde{p}_{jk\ell}, \hat{p}_{jk\ell}^{(m)}), \tag{6}$$

where

$$I^a(e, f) = \begin{cases} e\{(e/f)^a - 1\}/\{a(a+1)\} & \text{if } a \notin \{0, -1\}, \\ e\ln(e/f) & \text{if } a = 0, \\ f\ln(f/e) & \text{if } a = -1. \end{cases}$$

The statistics $R_{(1)}^a, R_{(2)}^a, R_{(3)}^a$ are based on power divergence [5], while $R_{(1)}^0, R_{(2)}^0, R_{(3)}^0$ are the log likelihood ratio statistics, and $R_{(1)}^1, R_{(2)}^1, R_{(3)}^1$ are Pearson's X^2 statistics. The statistics $R_{(1)}^{2/3}, R_{(2)}^{2/3}, R_{(3)}^{2/3}$ were recommended in [5] for goodness-of-fit testing. Under $\mathcal{H}_0^{(1)}, \mathcal{H}_0^{(2)}, \mathcal{H}_0^{(3)}$ given by Eqs. (3)–(5), it is known that the statistics $C_\phi^{(1)}, C_\phi^{(2)}, C_\phi^{(3)}$ have a chi-square limiting distribution with $(J-1)(K-1)L$, $(J-1)(L-1)K$ and $(K-1)(L-1)J$ degrees of freedom, respectively. The limiting distributions of test statistics for contingency tables can be found in [13].

In this paper, we use an asymptotic expansion to approximate the distribution of $C_\phi^{(m)}$ under $\mathcal{H}_0^{(m)}$ for $m \in \{1, 2, 3\}$. Using this approximation, we also construct transformed statistics and propose one of them when $C_\phi^{(m)}$ is $R^a_{(m)}$.

In Section 3, we derive a local Edgeworth approximation of the probability of X under $\mathcal{H}_0^{(m)}$. In Section 4, we consider an asymptotic approximation for the distribution of $C_\phi^{(m)}$ under $\mathcal{H}_0^{(m)}$ based on an asymptotic expansion. The approximation consists of continuous and discontinuous terms. In Section 5, we illustrate Bartlett's adjustment, a Bartlett-type adjustment, and an improved transformation. Next, by using the continuous term of the asymptotic approximation derived in Section 4, we obtain some new transformed statistics that increase the speed of convergence to a chi-square limiting distribution. In Section 6, the speeds of convergence to a chi-square limiting distribution of transformed statistics made by Bartlett adjustment, by Bartlett-type adjustment and by improved transformation are compared numerically with the speed of convergence of the original statistics $C_\phi^{(m)}$ when $C_\phi^{(m)}$ is $R^a_{(m)}$, i.e., statistics based on power divergence. Their powers are also compared numerically.

In Section 7, p-values of unconditional methods for a three-way contingency table are determined and the pros and cons of our method and competitors are considered. In Section 8, the proposed statistics are applied to the data from Section 1. In Section 9, we consider the correspondence between an independence model and a loglinear model in a three-way contingency table.

3. Local Edgeworth expansion

We consider the null hypothesis of conditional independence. In this case, the structure of $\mathcal{H}_0^{(m)}$ and statistics $R^a_{(m)}$ and $C_\phi^{(m)}$ for testing the hypothesis $\mathcal{H}_0^{(m)}$ are essentially the same for each $m \in \{1, 2, 3\}$. Therefore, it is sufficient to consider only $\mathcal{H}_0^{(1)}$, $R^a_{(1)}$ and $C_\phi^{(1)}$. Hereafter, we write $\mathcal{H}_0$, R^a and C_ϕ as $\mathcal{H}_0^{(1)}$, $R^a_{(1)}$ and $C_\phi^{(1)}$, respectively.

In this section, we derive a local Edgeworth approximation for the probability of X under the null hypothesis $\mathcal{H}_0$. Let X be distributed according to $\mathcal{M}_{JKL}(n, p_0)$, where

$$p_0 = (q_{111}, \ldots, q_{11L}, \ldots, q_{1K1}, \ldots, q_{1KL}, \ldots, q_{J11}, \ldots, q_{J1L}, \ldots, q_{JK1}, \ldots, q_{JKL})^\top$$

and $q_{jk\ell} = p_{j\cdot\ell}p_{\cdot k\ell}/p_{\cdot\cdot\ell}$ for all $j \in \{1, \ldots, J\}$, $k \in \{1, \ldots, K\}$, and $\ell \in \{1, \ldots, L\}$. For $j \in \{1, \ldots, J\}$, $k \in \{1, \ldots, K\}$, and $\ell \in \{1, \ldots, L\}$, let

$$U_{jk\ell} = (X_{jk\ell} - nq_{jk\ell})/\sqrt{n}. \tag{7}$$

Then $U = (U_{111}, \ldots, U_{11L}, \ldots, U_{1K1}, \ldots, U_{1KL}, \ldots, U_{J11}, \ldots, U_{J1L}, \ldots, U_{JK1}, \ldots, U_{JK,L-1})^\top$ is a lattice random vector that takes values in the set

$$S = \{u = (u_{111}, \ldots, u_{11L}, \ldots, u_{1K1}, \ldots, u_{1KL}, \ldots, u_{J11}, \ldots, u_{J1L}, \ldots, u_{JK1}, \ldots, u_{JK,L-1})^\top : u = (\tilde{x} - n\tilde{p}_0)/\sqrt{n},\ \tilde{x} \in S_0\},$$

where

$$\tilde{p}_0 = (I_{N-1}\ O_{N-1})p_0, \tag{8}$$

with I_{N-1} an $(N-1)\times(N-1)$ identity matrix, O_{N-1} an $(N-1)$-dimensional zero vector,

$$N = JKL, \tag{9}$$

and

$$S_0 = \left\{ \tilde{x} = (x_{111}, \ldots, x_{11L}, \ldots, x_{1K1}, \ldots, x_{1KL}, \ldots, x_{J11}, \ldots, x_{J1L}, \ldots, x_{JK1}, \ldots, x_{JK,L-1})^\top : \right.$$
$$\left. \text{the } x_{jk\ell}\text{s are nonnegative integers and } \sum_{j=1}^J \sum_{k=1}^K \sum_{\ell=1}^L x_{jk\ell} \le n + x_{JKL} \right\}.$$

The following theorem, proved in Appendix A.1, gives a local Edgeworth expansion.

Theorem 1. *Let $u = (\tilde{x} - n\tilde{p}_0)/\sqrt{n}$ for each $\tilde{x} \in S_0$. Then*

$$\Pr(U = u|\mathcal{H}_0) = n^{-(N-1)/2} f(u)\{1 + h_1(u)/\sqrt{n} + h_2(u)/n + h_3(u)/(n\sqrt{n}) + O(n^{-2})\},$$

where

$$f(u) = (2\pi)^{-(N-1)/2}|\Omega|^{-1/2}\exp(-u^\top \Omega^{-1} u/2), \tag{10}$$

$$\Omega = diag(q_{111}, \ldots, q_{JK,L-1}) - \tilde{p}_0\tilde{p}_0^\top, \tag{11}$$

$$h_1(u) = -M_1^1/2 + M_2^3/6, \tag{12}$$

$$h_2(u) = \{h_1(u)\}^2/2 + (1 - M_1^0)/12 + M_2^2/4 - M_3^4/12, \tag{13}$$

$$h_3(u) = -\{h_1(u)\}^3/3 + h_1(u)h_2(u) + M_2^1/12 - M_3^3/6 + M_4^5/20, \tag{14}$$

where $\tilde{p}_0$ *is given by* (8), $u_{JKL} = -(u_{111} + \cdots + u_{JK,L-1})$ *and*

$$M_b^a = \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{\ell=1}^{L} u_{jk\ell}^a / q_{jk\ell}^b.$$

4. An approximation based on asymptotic expansion

We consider an approximation for the distribution of C_ϕ under $\mathcal{H}_0$, viz. $\Pr(C_\phi \le x|\mathcal{H}_0) \approx W_1 + W_2$, where W_1 comes from a multivariate Edgeworth expansion and W_2 is a discontinuous term accounting for the discontinuity, which corresponds to an approximation for the distribution of Pearson's X^2 goodness-of-fit statistic for a multinomial distribution based on the asymptotic expansion given in [24]. This type of approximation for the distribution of test statistics for independence in a two-way table is discussed in [19]. The following result, proved in Appendix A.2, details how the term W_1 can be evaluated.

Theorem 2. *If we assume that* ϕ *is six-times differentiable and* $\phi^{(6)}$ *is continuous at* $t = 1$, *then*

$$W_1 = \Pr(\chi_M^2 \le x) + \frac{1}{n}\sum_{j=0}^{3} v_j^\phi \Pr(\chi_{M+2j}^2 \le x) + O(n^{-2}), \tag{15}$$

where

$$v_0^\phi = \gamma_0, \quad v_1^\phi = (\gamma_1 + \gamma_2) + 2\phi'''(1)\gamma_2 + \phi^{(4)}(1)\gamma_3 + \{\phi'''(1)\}^2\gamma_4,$$
$$v_2^\phi = 2(-\gamma_2 + \gamma_3) - 2\phi'''(1)(\gamma_2 + \gamma_4) - \phi^{(4)}(1)\gamma_3 - 2\{\phi'''(1)\}^2\gamma_4, \quad v_3^\phi = \gamma_4\{1 + \phi'''(1)\}^2,$$
$$\gamma_0 = -\Gamma_1/12 - \Gamma_2/12 + (\Gamma_3 + \Gamma_4)/12, \quad \gamma_1 = JK\Gamma_1/4 + \Gamma_2/4 - (K\Gamma_3 + J\Gamma_4)/4,$$
$$\gamma_2 = J^2K^2\Gamma_1/8 + \Gamma_2/8 - (K^2\Gamma_3 + J^2\Gamma_4)/8,$$
$$\gamma_3 = (-JK/2 - 1/8 + J/4 + K/4)\Gamma_1 - \Gamma_2/8 + (K\Gamma_3 + J\Gamma_4)/4 - (\Gamma_3 + \Gamma_4)/8,$$
$$\gamma_4 = (J^2K^2/8 + 3JK/4 + 1/3 - J/2 - K/2)\Gamma_1 + 5\Gamma_2/24 - (K^2\Gamma_3 + J^2\Gamma_4)/8 - (K\Gamma_3 + J\Gamma_4)/4 + (\Gamma_3 + \Gamma_4)/6,$$
$$\Gamma_1 = \sum_{\ell=1}^{L} 1/p_{\cdot\cdot\ell}, \quad \Gamma_2 = \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{\ell=1}^{L} 1/q_{jk\ell}, \quad \Gamma_3 = \sum_{j=1}^{J}\sum_{\ell=1}^{L} 1/p_{j\cdot\ell}, \quad \Gamma_4 = \sum_{k=1}^{K}\sum_{\ell=1}^{L} 1/p_{\cdot k\ell},$$

$M = (J-1)(K-1)L$, *and* χ_ν^2 *denotes a chi-square random variable with* ν *degrees of freedom.*

It is easily seen that $v_0^\phi + \cdots + v_3^\phi = 0$. By applying ϕ_a as ϕ in Theorem 2, we obtain the following corollary for the statistics based on power divergence.

Corollary 1. *When the statistic is* R^a *given by* (6),

$$W_1 = \Pr(\chi_M^2 \le x) + \frac{1}{n}\sum_{j=0}^{3} v_j^{(a)} \Pr(\chi_{M+2j}^2 \le x) + O(n^{-2}),$$

where for $j \in \{0, \ldots, 3\}$, $v_j^{(a)}$ *is defined as* v_j^ϕ *in the case of* $\phi'''(1) = a - 1$ *and* $\phi^{(4)}(1) = (a-1)(a-2)$, *respectively.*

We consider the evaluation of the discontinuous term W_2. As already defined in Section 2, u and $\tilde{p}_0$ are $(N-1)$-dimensional column vectors whose elements have three indices. To represent the elements of vectors concisely, we represent vectors u and $\tilde{p}_0$ as u^* and p^* by renumbering elements by one index, i.e., $u^* = (u_1^*, \ldots, u_{N-1}^*)^\top = u$ and $p^* = (p_1^*, \ldots, p_{N-1}^*)^\top = \tilde{p}_0$. Let $B_\phi(x) = \{u : C_\phi(u) \le x|\mathcal{H}_0\}$. The set $B_\phi(x)$ is a bounded extended convex set. For arbitrary $m \in \{1, \ldots, N-1\}$,

$$B_\phi(x) = \{u^* = (u_1^*, \ldots, u_{N-1}^*)^\top : \eta_m(\tilde{u}_m^*) \le u_m^* \le \theta_m(\tilde{u}_m^*),\ \tilde{u}_m^* = (u_1^*, \ldots, u_{m-1}^*, u_{m+1}^*, \ldots, u_{N-1}^*)^\top \in B_m\}, \tag{16}$$

where for $m \in \{1, \ldots, N-1\}$, $B_m \subset R_{N-2}$, and η_m and θ_m are real-valued continuous functions on R_{N-2}. Then, the discontinuous term W_2 is defined as

$$W_2 = -n^{-1/2}\sum_{m=1}^{N-1} n^{-(N-1-m)/2} \sum_{u_{m+1}^* \in L_{m+1}} \cdots \sum_{u_{N-1}^* \in L_{N-1}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \chi_{B_m}(\tilde{u}_m^*)\{S_1(\sqrt{n}u_m^* + np_m^*)f(u^*)\}_{\eta_m(\tilde{u}_m^*)}^{\theta_m(\tilde{u}_m^*)} du_1^* \cdots du_{m-1}^*,$$

where, for $m \in \{1, \ldots, N-1\}, j \in \{1, \ldots, J\}, k \in \{1, \ldots, K\}$ and $\ell \in \{1, \ldots, L\}$,

$$\{F(u^*)\}_{\eta_m(\tilde{u}_m^*)}^{\theta_m(\tilde{u}_m^*)} = F\{u_1^*, \ldots, u_{m-1}^*, \theta_m(\tilde{u}_m^*), u_{m+1}^*, \ldots, u_{N-1}^*\} - F\{u_1^*, \ldots, u_{m-1}^*, \eta_m(\tilde{u}_m^*), u_{m+1}^*, \ldots, u_{N-1}^*\},$$

$L_m = \{u_m^* : u_m^* = (n_m - np_m^*)/\sqrt{n},\ n_m \text{ is an integer}\},\ p_m^* = q_{jk\ell},\ m = (j-1)KL + (k-1)L + \ell,\ S_1(t) = t - [t] - 1/2,$ and χ_B is the indicator function of the set B. With regard to the evaluation of W_2, we have the following result.

Theorem 3. *If we assume that ϕ is four-times differentiable and $\phi^{(4)}$ is continuous at $t = 1$, then*

$$W_2 = \left\{(2\pi)^{N-1}\left(\prod_{j=1}^{J}\prod_{k=1}^{K}\prod_{\ell=1}^{L} q_{jk\ell}\right) q_{JKL}^{-1}\right\}^{-1/2} (A+B) - C + O(n^{-3/2}),$$

where A, B and C are

$$A = n^{-(N-1)/2} \sum_{u_1^* \in M_1^{\phi}} \cdots \sum_{u_{N-1}^* \in M_{N-1}^{\phi}} \exp\{-(u^*)^\top \Omega^{-1} u^*/2\},$$

$$B = \frac{1}{\sqrt{n}} \int \cdots \int_{B_\phi(x)} S_1(\sqrt{n}u_{N-1}^* + np_{N-1}^*)(u_{N-1}^*/p_{N-1}^* - u_N^*/p_N^*) \exp\{-(u^*)^\top \Omega^{-1} u^*/2\} du_1^* \cdots du_{N-1}^*$$
$$+ \frac{1}{n} \sum_{u_{N-1}^* \in M_{N-1}^{\phi}} \int \cdots \int_{B_{N-1}} S_1(\sqrt{n}u_{N-2}^* + np_{N-2}^*)(u_{N-2}^*/p_{N-2}^* - u_N^*/p_N^*) \exp\{-(u^*)^\top \Omega^{-1} u^*/2\} du_1^* \cdots du_{N-2}^*,$$

$$C = \Pr(\chi_M^2 \le x) + \frac{1}{n}\sum_{j=0}^{3} e_j^{\phi} \Pr(\chi_{M+2j}^2 \le x) + O(n^{-3/2}),$$

$$\begin{aligned}
e_0^{\phi} &= -(JKL)^2/8 - JKL^2/2 - L^2/8 + JKL/4 + L/4 + (J^2K + JK^2)L^2/4 \\
&\quad - (J^2 + K^2)L^2/8 + (J+K)L^2/4 - (J+K)L/4 + (-J^2K^2/8 + JK/4 - 5/8 + J/4 + K/4)\Gamma_1 \\
&\quad + (K^2\Gamma_3 + J^2\Gamma_4)/8 - (K\Gamma_3 + J\Gamma_4)/2 + 3(\Gamma_3 + \Gamma_4)/8, \\
e_1^{\phi} &= (JKL)^2/4 + JKL^2 + L^2/4 - (J^2K + JK^2)L^2/2 + (J^2 + K^2)L^2/4 \\
&\quad - (J+K)L^2/2 + (J^2K^2/4 + JK/2 + 5/4 - J - K)\Gamma_1 \\
&\quad - (K^2\Gamma_3 + J^2\Gamma_4)/4 + (K\Gamma_3 + J\Gamma_4)/2 - (\Gamma_3 + \Gamma_4)/4 + \phi^{(4)}(1)\gamma_3 + \{\phi'''(1)\}^2\gamma_4, \\
e_2^{\phi} &= -(JKL)^2/8 - JKL^2/2 - L^2/8 - JKL/4 - L/4 + (J^2K + JK^2)L^2/4 - (J^2 + K^2)L^2/8 \\
&\quad + (J+K)L^2/4 + (J+K)L/4 + (-J^2K^2/8 - 3JK/4 - 5/8 + 3J/4 + 3K/4)\Gamma_1 \\
&\quad + (K^2\Gamma_3 + J^2\Gamma_4)/8 - (\Gamma_3 + \Gamma_4)/8 - \phi^{(4)}(1)\gamma_3 - 2\{\phi'''(1)\}^2\gamma_4, \\
e_3^{\phi} &= \{\phi'''(1)\}^2\gamma_4.
\end{aligned}$$

$\gamma_3, \gamma_4, \Gamma_1, \Gamma_3$, and Γ_4 are defined in Theorem 2, and Ω and N are given by (11) and (9), respectively.

We can prove Theorem 3 essentially in a fashion similar to the proof of Theorem 5 in Taneichi and Sekiya [19]. The W_2 term is very complicated and difficult to handle in practice. Furthermore, as stated in the Introduction, as concerns the approximation of the distribution of test statistics based on an asymptotic expansion including the test statistic for independence in an $r \times s$ contingency table, the results of our numerical investigation suggest that an approximation based on the W_1 term, which means omitting the discrete term, does not lead to a serious error and already performs better than an approximation based on the asymptotic distribution. Therefore, in the next section, using the continuous term of the asymptotic expansion, we construct transformed statistics that converged to a chi-square limiting distribution faster than the original statistics do.

5. Some transformed statistics based on the continuous term of the asymptotic expansion

In order to improve small-sample accuracy of the chi-square approximation of the distribution of a test statistic, a Bartlett-type transformation and a transformation that is called improved transformation were considered.

We consider a nonnegative random variable T that has an asymptotic expansion

$$\Pr(T \le x) = \Pr(\chi_\nu^2 \le x) + \frac{1}{n}\sum_{j=0}^{h} a_j \Pr(\chi_{\nu+2j}^2 \le x) + O(n^{-2}), \tag{17}$$

where h is a positive integer, and $a_0, \ldots, a_h$ do not depend on $n\,(>0)$ and satisfy $a_0 + \cdots + a_h = 0$.

When $h = 1$ in (17), the Bartlett adjustment T_B [2,3,11] is defined as

$$T_B(T; a_0, f, n) = \{1 + 2a_0/(nf)\}T. \tag{18}$$

When $h = 3$ in (17), improved transformation T_I [7–9] and Bartlett-type adjustment T_{CF} [4] are defined as

$$T_I(T; a_0, a_2, a_3, f, n) = (n\alpha + \beta)^2 \ln\left[1 + \frac{1}{(n\alpha)^2}\left\{T + \frac{1}{n\alpha}(T^2 + \gamma T^3) + \frac{1}{(n\alpha)^2}\left(\frac{1}{3}T^3 + \frac{3\gamma}{4}T^4 + \frac{9\gamma^2}{20}T^5\right)\right\}\right], \tag{19}$$

where $\alpha = -f(f+2)/\{2(a_2 + a_3)\}$, $\beta = -(f+2)a_0/\{2(a_2 + a_3)\}$ and $\gamma = a_3/\{(f+4)(a_2 + a_3)\}$, and

$$T_{CF}(T; a_0, a_1, a_2, f, n) = \left\{1 + \frac{2a_0}{nf} + \frac{2(a_0 + a_1)}{nf(f+2)}T + \frac{2(a_0 + a_1 + a_2)}{nf(f+2)(f+4)}T^2\right\}T, \tag{20}$$

respectively. Then, one has, for all $\xi \in \{B, I, CF\}$,

$$\Pr(T_\xi \le x) = \Pr(\chi_f^2 \le x) + O(n^{-2}), \tag{21}$$

respectively.

We can apply (18)–(20) to improve the approximation (15). For the statistic C_ϕ based on ϕ that satisfies

$$\phi'''(1) = -1 \quad \text{and} \quad \phi^{(4)}(1) = 2, \tag{22}$$

we propose transformed statistic $C_\phi^B = T_B(C_\phi; v_0^\phi, M, n)$, since $v_1^\phi = -v_0^\phi$, and $v_2^\phi = v_3^\phi = 0$ hold in (15). For the other statistics, we propose $C_\phi^I = T_I(C_\phi; v_0^\phi, v_2^\phi, v_3^\phi, M, n)$ and $C_\phi^{CF} = T_{CF}(C_\phi; v_0^\phi, v_1^\phi, v_2^\phi, M, n)$. However, when ϕ satisfies condition (22), we note that $C_\phi^B = C_\phi^{CF}$.

Now, $v_0^\phi, \ldots, v_3^\phi$ include $\Gamma_1, \ldots, \Gamma_4$, which are functions of unknown parameters $p_{\cdot\cdot\ell}$, $p_{j\cdot\ell}$ and $p_{\cdot k\ell}$ with $j \in \{1, \ldots, J\}$, $k \in \{1, \ldots, K\}$, $\ell \in \{1, \ldots, L\}$. Then, in practical application, we substitute the maximum likelihood estimates $\check{p}_{\cdot\cdot\ell} = x_{\cdot\cdot\ell}/n$, $\check{p}_{j\cdot\ell} = x_{j\cdot\ell}/n$, and $\check{p}_{\cdot k\ell} = x_{\cdot k\ell}/n$ for $p_{\cdot\cdot\ell}$, $p_{j\cdot\ell}$, and $p_{\cdot k\ell}$, respectively, where $x_{\cdot\cdot\ell}$, $x_{j\cdot\ell}$ and $x_{\cdot k\ell}$ are observed values of $X_{\cdot\cdot\ell}$, $X_{j\cdot\ell}$ and $X_{\cdot k\ell}$, respectively.

In the case of the power divergence statistic $R^a \equiv C_{\phi_a}$, $R_B^a \equiv C_{\phi_a}^B$ is defined when $a = 0$ (the log likelihood ratio statistic) while $R_I^a \equiv C_{\phi_a}^I$ is defined when $a \neq 0$, and $R_{CF}^a \equiv C_{\phi_a}^{CF}$ is defined when both $a \neq 0$ and $a = 0$ and $R_{CF}^0 = R_B^0$.

The approximation $\Pr(C_\phi^\xi < x) \approx \Pr(\chi_M^2 < x)$ ($\xi = B, I, CF$) is justified by Eq. (21). Therefore, for $\xi \in \{B, I, CF\}$,

$$\Pr(C_\phi^\xi \ge x) \approx \Pr(\chi_M^2 \ge x) \tag{23}$$

holds. This means that the p-value of the test statistics C_ϕ^ξ with $\xi \in \{B, I, CF\}$ is approximated by $\Pr(\chi_M^2 \ge x)$.

6. Numerical comparison of transformed statistics and the original statistics

We consider the statistics R^a based on power divergence as concrete statistics. We investigate the small-sample performance of the approximation to a chi-square distribution. We compare numerically the performance of transformed statistics $R_{CF}^0 = R_B^0$ ($a = 0$), R_I^a and R_{CF}^a ($a \neq 0$) with that of the original statistics R^a. We evaluate the performance using the following Monte Carlo procedure.

We generate N_1 multinomial contingency tables using multinomial random vectors under $\mathcal{H}_0$ and arrange the tables as $x^S(q)$ with $q \in \{1, \ldots, N_1\}$. For each $q \in \{1, \ldots, N_1\}$, let $T\{x^S(q)\}$ be the value of a statistic T at $x^S(q)$. Let $\chi_M^2(\alpha)$ be the upper α point of the chi-square distribution with M degrees of freedom, and let N_2 be the number of elements of the set of q that satisfies the condition $T\{x^S(q)\} > \chi_M^2(\alpha)$. Then the performance of the approximation for the distribution can be evaluated on the basis of the index $I(\alpha) = N_2/N_1 - \alpha$. For statistics $R_{CF}^0 = R_B^0$, R_I^a ($a \neq 0$) and R_{CF}^a ($a \neq 0$), we investigate $2 \times 2 \times 2$, $2 \times 2 \times 4$, $3 \times 3 \times 3$ and $2 \times 3 \times 4$ contingency tables. The following sets of marginal probabilities are considered.

Case I: The marginal probabilities are all equal, i.e., for all $j \in \{1, \ldots, J\}$, $k \in \{1, \ldots, K\}$, $\ell \in \{1, \ldots, L\}$,

$$p_{j\cdot\ell} = 1/(JL), \quad p_{\cdot k\ell} = 1/(KL), \quad p_{\cdot\cdot\ell} = 1/L.$$

Case II: The marginal probabilities are not equal, i.e., for all $j \in \{1, \ldots, J\}$, $k \in \{1, \ldots, K\}$, $\ell \in \{1, \ldots, L\}$,

$$p_{j\cdot\ell} = \frac{1}{2}\left\{\frac{1}{JL} + \frac{4j\ell}{J(J+1)L(L+1)}\right\}, \quad p_{\cdot k\ell} = \frac{1}{2}\left\{\frac{1}{KL} + \frac{4k\ell}{K(K+1)L(L+1)}\right\}, \quad p_{\cdot\cdot\ell} = \frac{1}{2}\left\{\frac{L+1+2\ell}{L(L+1)}\right\}.$$

We carry out the above Monte Carlo procedure for $N_1 = 10^6$ and sample size $n = sN$ ($s = 4$ and 6) for three-way contingency tables. We consider a statistic when $a = 0$ (the log likelihood ratio statistic), a statistic when $a = 0.2$, a statistic when $a = 2/3$ and a statistic when $a = 1$ (Pearson's X^2 statistic).

In our simulation, if $\tilde{p}_{jk\ell} \equiv x_{jk\ell}/n$ are 0, $I^a(\tilde{p}_{jk\ell}, \hat{p}_{jk\ell}^{(m)})$ in (6) are 0 for any $\hat{p}_{jk\ell}^{(m)}$. For a transformed statistic, if $x_{j\cdot\ell}$, $x_{\cdot k\ell}$, or $x_{\cdot\cdot\ell}$ are 0, we put $x_{j\cdot\ell} = 10^{-4}$, $x_{\cdot k\ell} = 10^{-4}$, or $x_{\cdot\cdot\ell} = 10^{-4}$, respectively.

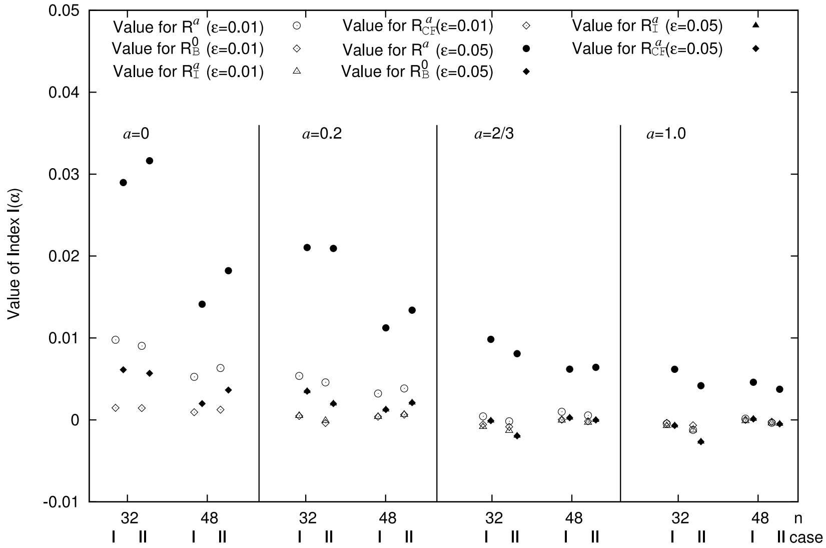

Fig. 1. Values of $I(\alpha)$ when the original statistic is R^a with $a \in \{0, 0.2, 2/3, 1\}$ for a $2 \times 2 \times 2$ contingency table with sample size $n \in \{32, 48\}$: ∘ and • are the values for R^a when $\alpha = 0.01$ and 0.05, respectively, △ and ▲ are the values for R_I^a when α = 0.01 and 0.05, respectively, and ♢ and ♦ are the values for R_{CF}^a when $\alpha = 0.01$ and 0.05, respectively. The first column is for Case I and the second column is for Case II.

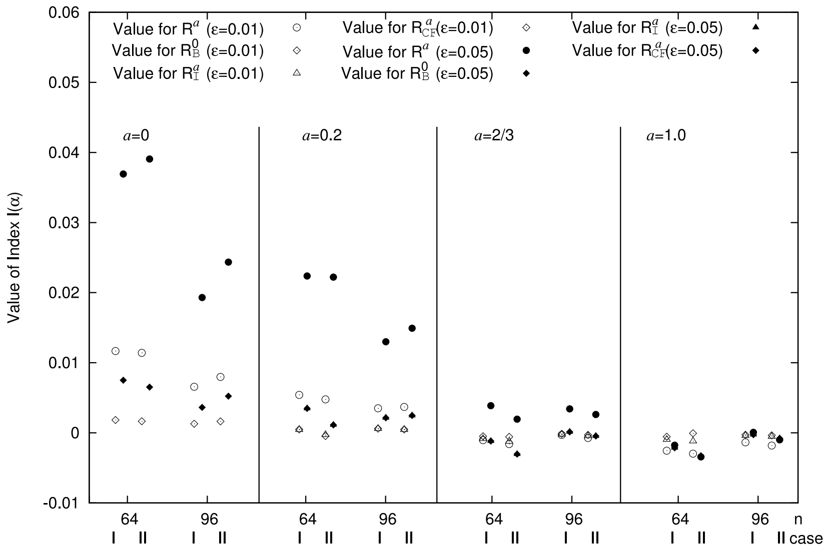

Fig. 2. Values of $I(\alpha)$ when the original statistic is R^a with $a \in \{0, 0.2, 2/3, 1\}$ for a $2 \times 2 \times 4$ contingency table with sample size $n \in \{64, 96\}$: ∘ and • are the values for R^a when $\alpha = 0.01$ and 0.05, respectively, △ and ▲ are the values for R_I^a when $\alpha = 0.01$ and 0.05, respectively, and ♢ and ♦ are the values for R_{CF}^a when $\alpha = 0.01$ and 0.05, respectively. The first column is for Case I and the second column is for Case II.

Fig. 1 shows the values of $I(\alpha)$ in Cases I–II when significance levels are $\alpha = 0.01$ and 0.05 for a $2 \times 2 \times 2$ contingency table and the original statistics are R^0, $R^{0.2}$, $R^{2/3}$ and R^1. Figs. 2–4 show those as shown in Fig. 1 for $2 \times 2 \times 4$, $3 \times 3 \times 3$ and $2 \times 3 \times 4$ contingency tables, respectively.

From Figs. 1–4, we find that the transformed statistic $R_B^0 = R_{CF}^0$ always performs much better than the original statistic R^0. The transformed statistics $R_I^{0.2}$ and $R_{CF}^{0.2}$ also perform much better than the original statistic $R^{0.2}$. Furthermore, the transformed statistics $R_I^{2/3}$ and $R_{CF}^{2/3}$ perform better than $R^{2/3}$ in most cases, and $R_{CF}^{2/3}$ performs better than $R_I^{2/3}$. Moreover, we find that the transformed statistic R_{CF}^1 performs better than Pearson's X^2 statistic R^1 in most cases. However, R_I^1 sometimes does not perform better than R^1.

For the test statistics based on power divergence for conditional independence in a $J \times K \times L$ contingency table, in the case of the usual setting, the transformed statistic R_{CF}^a generally improves the speed of convergence to a chi-square distribution very well. For the case of an extreme setting, we consider the following cell probability type models in Tables 3–4. Table 3

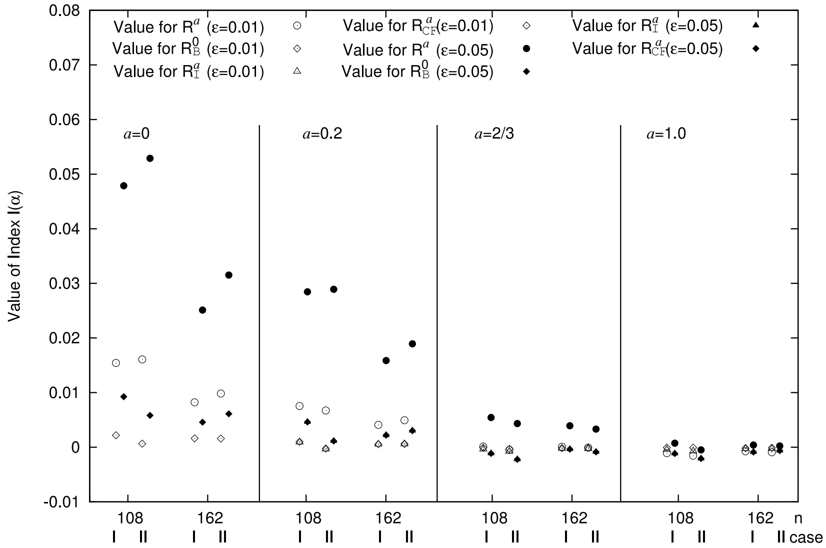

Fig. 3. Values of $I(\alpha)$ when the original statistic is R^a with $a \in \{0, 0.2, 2/3, 1\}$ for a $3 \times 3 \times 3$ contingency table with sample size $n \in \{108, 162\}$: $\circ$ and $\bullet$ are the values for R^a when $\alpha = 0.01$ and 0.05, respectively, $\triangle$ and $\blacktriangle$ are the values for R_I^a when $\alpha = 0.01$ and 0.05, respectively, and $\lozenge$ and $\blacklozenge$ are the values for R_{CF}^a when $\alpha = 0.01$ and 0.05, respectively. The first column is for Case I and the second column is for Case II.

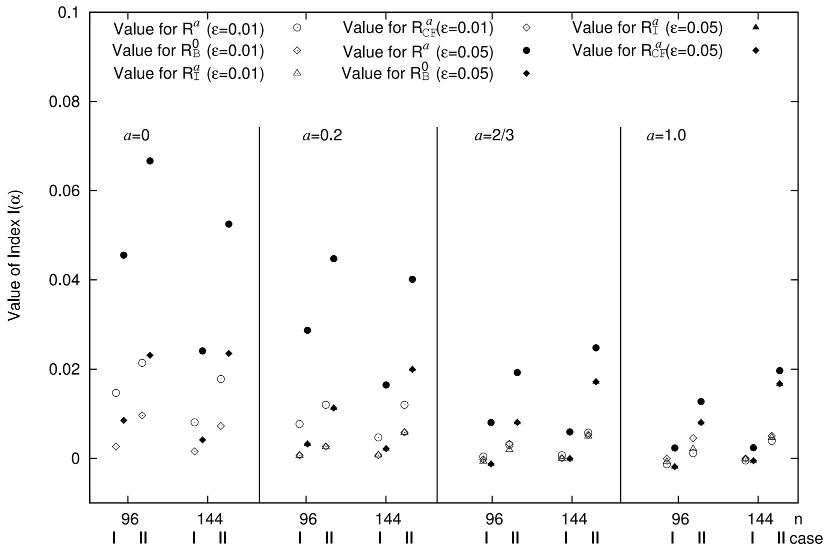

Fig. 4. Values of $I(\alpha)$ when the original statistic is R^a with $a \in \{0, 0.2, 2/3, 1\}$ for a $2 \times 3 \times 4$ contingency table with sample size $n \in \{96, 144\}$: $\circ$ and $\bullet$ are the values for R^a when $\alpha = 0.01$ and 0.05, respectively, $\triangle$ and $\blacktriangle$ are the values for R_I^a when $\alpha = 0.01$ and 0.05, respectively, and $\lozenge$ and $\blacklozenge$ are the values for R_{CF}^a when $\alpha = 0.01$ and 0.05, respectively. The first column is for Case I and the second column is for Case II.

Table 3
Model (A) in which cell probability includes 0.01.

j	ℓ	1		2	
	k	1	2	1	2
1		$p_{111} = 0.01$	$p_{121} = 0.04$	$p_{112} = 0.03$	$p_{122} = 0.12$
2		$p_{211} = 0.04$	$p_{221} = 0.16$	$p_{212} = 0.12$	$p_{222} = 0.48$

shows a model (model (A)) that includes a cell probability of 0.01, and Table 4 shows a model (model (B)) that includes a cell probability of 0.99. Table 5 shows the value of $I(\alpha)$ for model (A) for each statistic and significance levels $\alpha = 0.01$ and 0.05. Table 6 shows the value of $I(\alpha)$ for model (B) for each statistic and significance levels $\alpha = 0.01$ and 0.05. From Tables 5–6, for the extreme model, transformed statistics exhibit a performance that is almost the same as that of the original statistics.

Table 4
Model (B) in which cell probability includes 0.99.

j	ℓ	1		2	
	k	1	2	1	2
1		$p_{111} = 0.000011$	$p_{121} = 0.003300$	$p_{112} = 0.00084725$	$p_{122} = 0.00084725$
2		$p_{211} = 0.003300$	$p_{221} = 0.990000$	$p_{212} = 0.00084725$	$p_{222} = 0.00084725$

Table 5
Values of $I(\alpha)$ for Model (A).

n	a	$\alpha = 0.01$			$\alpha = 0.05$		
		R^a	R_I^a	R_{CF}^a	R^a	R_I^a	R_{CF}^a
32	0	−0.00227		0.17983	−0.00362		0.17595
	0.2	−0.00384	0.11240	0.06129	−0.01136	0.13740	0.08947
	2/3	−0.00340	−0.00675	0.00332	−0.01146	−0.03358	−0.01093
	1	0.00146	−0.00688	0.00471	−0.00537	−0.03291	−0.00125
48	0	−0.00085		0.05975	0.00313		0.04660
	0.2	−0.00251	0.03359	0.00157	−0.00603	0.03250	0.00006
	2/3	−0.00118	−0.00429	0.00395	−0.00833	−0.02407	−0.00743
	1	0.00269	−0.00440	0.00162	−0.00257	−0.02222	0.00311
64	0	0.00005		0.01735	0.00963		0.00376
	0.2	−0.00177	0.00851	−0.00410	−0.00193	−0.00101	−0.01670
	2/3	−0.00049	−0.00223	0.00305	−0.00522	−0.01666	−0.00412
	1	0.00272	−0.00272	0.00147	−0.00081	−0.01423	0.00492
80	0	0.00086		0.00260	0.01223		−0.00855
	0.2	−0.00111	0.00041	−0.00388	0.00159	−0.00917	−0.01561
	2/3	−0.00048	−0.00119	0.00248	−0.00469	−0.01193	−0.00315
	1	0.00232	−0.00162	0.00127	−0.00140	−0.00925	0.00390

Table 6
Values of $I(\alpha)$ for Model (B).

n	a	$\alpha = 0.01$			$\alpha = 0.05$		
		R^a	R_I^a	R_{CF}^a	R^a	R_I^a	R_{CF}^a
32	0	−0.01000		−0.00886	−0.04967		−0.04879
	0.2	−0.00970	0.00061	−0.00995	−0.04965	−0.03933	−0.04916
	2/3	−0.00966	−0.00969	−0.00995	−0.04965	−0.04967	−0.04910
	1	−0.00967	−0.00970	−0.00995	−0.04967	−0.04970	−0.04992
48	0	−0.00960		−0.00946	−0.04948		−0.04992
	0.2	−0.00960	−0.00664	−0.00999	−0.04949	−0.02837	−0.04988
	2/3	−0.00949	−0.00956	−0.00993	−0.04948	−0.04956	−0.04991
	1	−0.00948	−0.00955	−0.00992	−0.04947	−0.04955	−0.04991
64	0	−0.00956		−0.00930	−0.04934		−0.04926
	0.2	−0.00955	−0.00450	−0.00999	−0.04931	−0.03882	−0.04996
	2/3	−0.00929	−0.00942	−0.00987	−0.04929	−0.04942	−0.04987
	1	−0.00929	−0.00942	−0.00987	−0.04929	−0.04942	−0.04987
80	0	−0.00947		−0.00914	−0.04916		−0.04913
	0.2	−0.00919	−0.00435	−0.00999	−0.04912	−0.04067	−0.04999
	2/3	−0.00913	−0.00934	−0.00979	−0.04912	−0.04934	−0.04979
	1	−0.00914	−0.00934	−0.00980	−0.04914	−0.04934	−0.04980

Next, we compare the power of transformed statistics $R_{CF}^0 = R_B^0, R_I^a$ and R_{CF}^a with $a \in \{0.2, 2/3, 1\}$ with that of the original statistics. For a $2 \times 2 \times 4$ table, against the null hypothesis $\mathcal{H}_0$, we consider the alternative hypothesis

$$\mathcal{H}_1 : \forall_{j\in\{1,2\}}\forall_{k\in\{1,2\}}\forall_{\ell\in\{1,\ldots,4\}} \quad p_{jk\ell} = p_{j\cdot\ell}p_{\cdot k\ell}/p_{\cdot\cdot\ell} + \varepsilon_{jk\ell}, \tag{24}$$

where $\varepsilon_{111} = -r$, $\varepsilon_{112} = 2r$, $\varepsilon_{113} = 3r$, $\varepsilon_{114} = 0$, $\varepsilon_{121} = 3r$, $\varepsilon_{122} = 0$, $\varepsilon_{123} = -r$, $\varepsilon_{124} = 2r$, $\varepsilon_{211} = -r$, $\varepsilon_{212} = r$, $\varepsilon_{213} = 3r$, $\varepsilon_{214} = 2r$, $\varepsilon_{221} = r$, $\varepsilon_{222} = 3r$, $\varepsilon_{223} = -3r$, $\varepsilon_{224} = -14r$. Also, for a $2 \times 3 \times 4$ table, against the null hypothesis $\mathcal{H}_0$, we consider the alternative hypothesis

$$\mathcal{H}_1 : \forall_{j\in\{1,2\}}\forall_{k\in\{1,2,3\}}\forall_{\ell\in\{1,\ldots,4\}} \quad p_{jk\ell} = p_{j\cdot\ell}p_{\cdot k\ell}/p_{\cdot\cdot\ell} + \varepsilon_{jk\ell}, \tag{25}$$

where $\varepsilon_{111} = r$, $\varepsilon_{112} = r$, $\varepsilon_{113} = -r$, $\varepsilon_{114} = -r$, $\varepsilon_{121} = r$, $\varepsilon_{122} = r$, $\varepsilon_{123} = -r$, $\varepsilon_{124} = -r$, $\varepsilon_{131} = r$, $\varepsilon_{132} = r$, $\varepsilon_{133} = -r$, $\varepsilon_{134} = -r$, $\varepsilon_{211} = -r$, $\varepsilon_{212} = -r$, $\varepsilon_{213} = r$, $\varepsilon_{214} = r$, $\varepsilon_{221} = -r$, $\varepsilon_{222} = -r$, $\varepsilon_{223} = r$, $\varepsilon_{224} = r$, $\varepsilon_{231} = -r$, $\varepsilon_{232} = -r$, $\varepsilon_{233} = r$, $\varepsilon_{234} = r$. We calculated the simulated average power P_o against the alternative hypotheses (24) and (25) with $r = 0.01$ by using simulated exact critical values of the statistics.

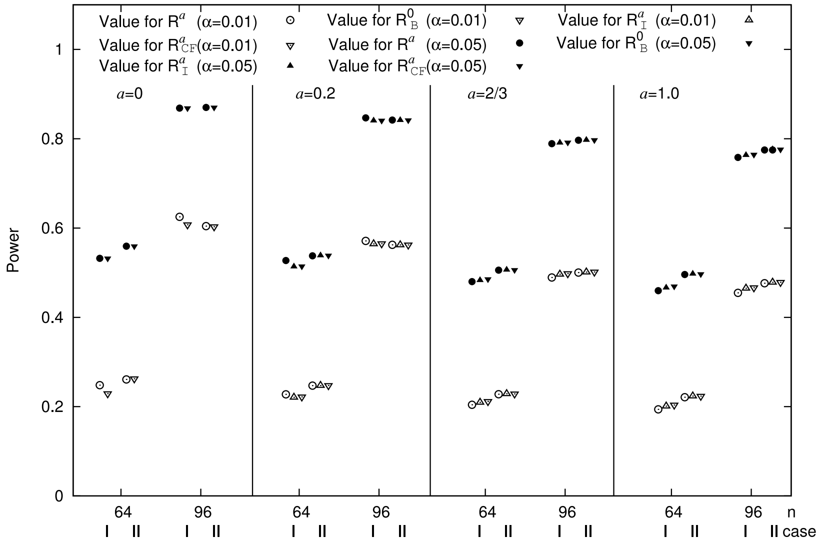

Fig. 5. Simulated average power P_0 against an alternative model (24) when original statistics are R^0, $R^{0.2}$, $R^{2/3}$, and R^1 for a $2 \times 2 \times 4$ contingency table with sample size $n \in \{64, 96\}$: ○ and ● are the values for R^a with $a \in \{0, 0.2, 2/3, 1\}$ when $\alpha = 0.01$ and 0.05, respectively, △ and ▲ are the values for R_I^a with $a \in \{0.2, 2/3, 1\}$ when $\alpha = 0.01$ and 0.05, respectively, and ▽ and ▼ are the values for $R_B^0 (= R_{CF}^0)$ and R_{CF}^a with $a \in \{0.2, 2/3, 1\}$ when $\alpha = 0.01$ and 0.05, respectively. The first column is for Case I and the second column is for Case II.

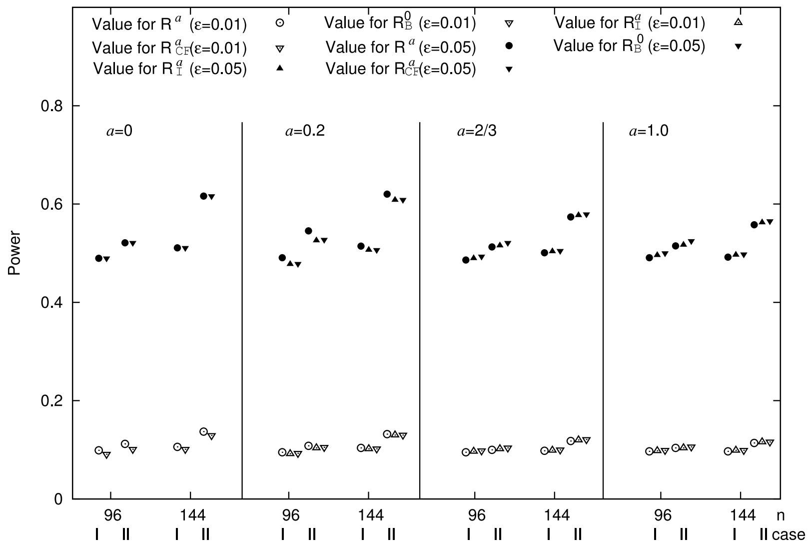

Fig. 6. Simulated average power P_0 against an alternative model (25) when the original statistics are R^0, $R^{0.2}$, $R^{2/3}$, and R^1 for a $2 \times 3 \times 4$ contingency table with sample size $n \in \{96, 144\}$: ○ and ● are the values for R^a with $a \in \{0, 0.2, 2/3, 1\}$ when $\alpha = 0.01$ and 0.05, respectively, △ and ▲ are the values for R_I^a with $a \in \{0.2, 2/3, 1\}$ when $\alpha = 0.01$ and 0.05, respectively, and ▽ and ▼ are the values for $R_B^0 (= R_{CF}^0)$ and R_{CF}^a with $a \in \{0.2, 2/3, 1\}$ when $\alpha = 0.01$ and 0.05, respectively. The first column is for Case I and the second column is for Case II.

Fig. 5 shows the power of the statistics for a $2 \times 2 \times 4$ contingency table in Cases I–II when the significance level is $\alpha = 0.01$ and 0.05 and the original statistics are R^a for $a \in \{0, 0.2, 2/3, 1\}$. The number of repetitions is 10^6, the ratio of irregular cases is 0.0, and the sample sizes are 64 and 96.

Fig. 6 shows the power of the statistics for a $2 \times 3 \times 4$ contingency table. The number of repetitions is 10^6, the ratio of irregular cases is 0.0, and the sample sizes are 96 and 144. From Figs. 5–6, we conclude that the power against $\mathcal{H}_1$ given by (24) and (25) of the transformed statistics $R_{CF}^0 = R_B^0$, R_I^a and R_{CF}^a with $a \in \{0.2, 2/3, 1\}$ is not so different from that of the original statistics in $2 \times 2 \times 4$ and $2 \times 3 \times 4$ contingency tables, respectively.

These power results were expected since it is known that a transformed statistic by a monotone function that includes a Bartlett adjustment has the same power as that of the original statistic.

Table 7
Three-way data #1.

A	C	Yes			No		
	B	1	2	3	1	2	3
Yes		30	23	21	25	19	40
No		8	5	1	4	2	2

Table 8
Three-way data #2.

A	C	Yes		No	
	B	Yes	No	Yes	No
Yes		5	9	21	1
No		7	17	9	4

Table 9
Three-way data #3.

A	C	Yes		No	
	B	Yes	No	Yes	No
Yes		7	14	9	2
No		3	14	16	3

Table 10
Three-way data #4.

A	C	Yes		No	
	B	Yes	No	Yes	No
Yes		4	13	5	4
No		0	11	14	4

7. Unconditional methods and our methods

For a 2×2 contingency table, Storer and Kim [18] and Tang and Tang [23] proposed and developed exact methods (unconditional methods). Since our methods are improved asymptotic methods, their performance is inferior to that of unconditional methods in a small sample size. For a conditional independent model of a three-way contingency table, in order to compute the p-value of statistic T by an approximate unconditional method (AU-method), we consider the probability structure

$$\Pr(X = x) = n! \left\{ \prod_{j=1}^{J} \prod_{k=1}^{K} \prod_{\ell=1}^{L} \left(\frac{x^*_{j\cdot\ell} x^*_{\cdot k\ell}}{n x^*_{\cdot\cdot\ell}} \right)^{x_{jk\ell}} \right\} \Big/ \left\{ \prod_{j=1}^{J} \prod_{k=1}^{K} \prod_{\ell=1}^{L} (x_{jk\ell})! \right\},$$

where $x^*_{\cdot\cdot\ell}$, $x^*_{j\cdot\ell}$ and $x^*_{\cdot k\ell}$ are the marginal observation calculated from observation $x^* = (x^*_{111}, \ldots, x^*_{JKL})^\top$. Let the observed value of statistic be $T(x^*)$. We consider every lattice $x = (x_{111}, \ldots, x_{JKL})^\top$ with non-negative integers $x_{jk\ell} \geq 0$ for all $j \in \{1, \ldots, J\}$, $k \in \{1, \ldots, K\}$, and $\ell \in \{1, \ldots, L\}$. Suppose that the counts add up to n. The number L_A of lattices is then $\binom{n+JKL-1}{n}$. For every x, compute the value of statistic $T(x)$. Then, the p-value $p^{AU}(x^*)$ of T given by AU-method is given by

$$p^{AU}(x^*) = \sum_{x \in \{x | T(x) \geq T(x^*)\}} \Pr(X = x).$$

However, in a three-way contingency table, the number L_A of lattices increases rapidly with n. For example, for a $2 \times 2 \times 2$ contingency table, $L_A = 245{,}157$ when $n = 16$ and $L_A = 2{,}629{,}575$ when $n = 24$. The computation of $p^{AU}(x^*)$ is thus very difficult in practice except for a very small sample size.

Let p_A be the p-value of the statistic R^a_{CF} given by the approximated AU-method by using the Monte Carlo method. The value of p_A is considered to be very accurate. Let p_B be the p-value of statistic R^a_{CF} given by (23).

The values of p_A and p_B for some data (Tables 1, 7–10) are shown in Table 11. From Table 11, p_A and p_B are very close, and the approximated p-value based on (23) therefore performed well for not so small sample sizes. Also, this approximation is not necessary for a large computational power.

8. Application to real data

Using the results in Table 11, we test (C1) for the data in Table 1. The value of the log likelihood ratio statistic $R^0 = G^2$ is 9.637. The value of the transformed log likelihood ratio statistic R^0_{CF} is 9.425. Since $\chi^2_4(0.05) = 9.488$, (C1) cannot be rejected

Table 11
Values of p_A and p_B for each data set.

a	Table 1		Table 7		Table 8		Table 9		Table 10	
	p_A	p_B	p_A	p_B	p_A	p_B	p_A	p_B	p_A	p_B
0	0.05226	0.05130	0.30512	0.28683	0.11950	0.11582	0.57098	0.56378	0.07205	0.07560
0.2	0.05183	0.05128	0.30381	0.29393	0.11202	0.11240	0.56953	0.56399	0.08139	0.08720
2/3	0.05179	0.05176	0.31155	0.31159	0.09958	0.10327	0.56735	0.56477	0.09834	0.10627
1	0.05251	0.05253	0.32235	0.32422	0.09265	0.09570	0.56675	0.56552	0.10828	0.11496

at the 5% level using the transformed statistic R^0_{CF}. The asymptotic p-value of R^0 is 0.0470, and the simulated p-value based on the AU-method (which is more precise) for R^0 is 0.05254. In contrast, the asymptotic p-value of R^0_{CF} is 0.05130, and the simulated p-value based on the AU-method for R^0_{CF} is 0.05226. Therefore, for the statistic R^0, the conclusion of the test based on an asymptotic distribution is opposite to that of the test based on an exact distribution at the 5% level. In contrast, for the statistic R^0_{CF}, the conclusion of the test based on an asymptotic distribution agrees with that of the test based on the exact distribution at the 5% level.

9. Relation to testing a loglinear model for three-way contingency tables

Here we consider the correspondence between testing independence and testing a loglinear model. Throughout this section, we use the notations of Agresti [1] for loglinear models in three-way contingency tables.

Taneichi et al. [22] obtained an approximation for the distribution of test statistics for complete independence in multi-way contingency tables. Using the approximation, we proposed transformed statistics to improve the speed of convergence to a chi-square distribution. For the loglinear model of a three-way table, testing complete independence corresponds to testing the model

$$\ln \mu_{ijk} = \lambda + \lambda^X_i + \lambda^Y_j + \lambda^Z_k.$$

Kobe et al. [10] obtained an approximation for the distribution of test statistics for the independence of one factor to two other factors in three-way tables and proposed transformed statistics using the approximation. For the loglinear model of a three-way table, testing the independence of one factor to two other factors corresponds to testing the model

$$\ln \mu_{ijk} = \lambda + \lambda^X_i + \lambda^Y_j + \lambda^Z_k + \lambda^{XY}_{ij}.$$

We obtained an approximation for the distribution of C_ϕ for a test of conditional independence in three-way contingency tables, and we proposed a transformed statistic using the approximation. For the loglinear model of a three-way table, testing conditional independence corresponds to testing the model

$$\ln \mu_{ijk} = \lambda + \lambda^X_i + \lambda^Y_j + \lambda^Z_k + \lambda^{XY}_{ij} + \lambda^{YZ}_{jk}.$$

By using tests of three types of independence model, we have covered for testing the loglinear model in three-way tables except for the model

$$\ln \mu_{ijk} = \lambda + \lambda^X_i + \lambda^Y_j + \lambda^Z_k + \lambda^{XY}_{ij} + \lambda^{YZ}_{jk} + \lambda^{XZ}_{ik}.$$

10. Concluding remarks

We derived an expression of asymptotic expansion for the distribution of test statistics C_ϕ (based on ϕ-divergence) for conditional independence in $J \times K \times L$ contingency tables. Using the continuous terms of the expression, we obtained transformations that improve the speed of convergence to the chi-square asymptotic distribution of C_ϕ.

As a special case of C_ϕ, we considered power divergence statistics R^a. By numerical comparison, we found that the Bartlett-type transformed R^0 and $R^{0.2}$ statistics, namely R^0_{CF} and $R^{0.2}_{CF}$, perform very well and are recommendable when $n \geq 4JKL$. Also, the power of transformed statistics is almost the same as that of the original statistics.

We can state with certitude that for small-sample $J \times K \times L$ contingency tables with non-zero marginal frequencies and sample size $n \geq 4JKL$, the proposed transformed statistics, especially R^0_{CF} and $R^{0.2}_{CF}$, are more reliable than the original statistics.

There are four standard independence models in a $J \times K \times L$ contingency table. We obtained improved transformed test statistics for the third case. Therefore, by the results of this work and previous works [10,22], we constructed and covered the transformed test statistics for three out of four possible independence models.

Acknowledgments

The authors are very grateful to the Editor-in-Chief, Prof. Christian Genest, an Associate Editor and two referees for their valuable comments and suggestions. This research was partially supported by Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (C), 24540133. Any remaining error is ours.

Appendix

A.1. Proof of Theorem 1

Introduce the notation

$$\sum^{[JKL-1]} a_{jk\ell} = \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{\ell=1}^{L} a_{jk\ell} - a_{JKL}.$$

If we set $\tilde{X} = (I_{N-1} O_{N-1})X$, the characteristic function G of $\tilde{X}$ is given by

$$G(s) = \sum_{\tilde{x}\in S_x} \exp(is^\top \tilde{x}) \Pr(\tilde{X} = \tilde{x} \mid \mathcal{H}_0) = \left\{ \sum^{[JKL-1]} q_{jk\ell} \exp(is_{jk\ell}) + q_{JKL} \right\}^n,$$

where $s = (s_{111}, \ldots, s_{JK,L-1})^\top$. For each $u = (\tilde{x} - n\tilde{p}_0)/\sqrt{n} \in S$, we find

$$\Pr(U = u \mid \mathcal{H}_0) = \Pr(\tilde{X} = \tilde{x} \mid \mathcal{H}_0) = (2\pi)^{-(N-1)} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} G(s) \exp(-is^\top \tilde{x}) ds = (2\pi\sqrt{n})^{-(N-1)} I,$$

where

$$I = \int_{-\sqrt{n}\pi}^{\sqrt{n}\pi} \cdots \int_{-\sqrt{n}\pi}^{\sqrt{n}\pi} \gamma(s) \exp(-is^\top u) ds \tag{A.1}$$

and $\gamma(s) = G(s/\sqrt{n}) \exp(-i\sqrt{n}s^\top \tilde{p}_0)$. We can write

$$\gamma(s) = \{\exp(-s^\top \Omega s/2)\} \left\{ 1 + \sum_{m=1}^{3} n^{-m/2} b_m(s) + O(n^{-2}) \right\} \tag{A.2}$$

for large n and fixed s, where

$$b_1(s) = \frac{i^3}{6} \left\{ \sum^{[JKL-1]} q_{jk\ell} s_{jk\ell}^3 - 3(s^\top \tilde{p}_0) s^\top \Omega s - (s^\top \tilde{p}_0)^3 \right\},$$

$$b_2(s) = \frac{1}{2}\{b_1(s)\}^2 + \frac{i^4}{24} \left\{ \sum^{[JKL-1]} q_{jk\ell} s_{jk\ell}^4 - 4(s^\top \tilde{p}_0) \sum^{[JKL-1]} q_{jk\ell} s_{jk\ell}^3 - 3(s^\top \Omega s)^2 + 6(s^\top \tilde{p}_0)^2 s^\top \Omega s + 3(s^\top \tilde{p}_0)^4 \right\}$$

and

$$b_3(s) = -\frac{1}{3}\{b_1(s)\}^3 + b_1(s) b_2(s) + \frac{i^5}{120} \left\{ \sum^{[JKL-1]} q_{jk\ell} s_{jk\ell}^5 - 5(s^\top \tilde{p}_0) \sum^{[JKL-1]} q_{jk\ell} s_{jk\ell}^4 - 10(s^\top \Omega s) \sum^{[JKL-1]} q_{jk\ell} s_{jk\ell}^3 \right.$$
$$\left. + 10(s^\top \tilde{p}_0)^2 \sum^{[JKL-1]} q_{jk\ell} s_{jk\ell}^3 + 30(s^\top \tilde{p}_0)(s^\top \Omega s)^2 - 6(s^\top \tilde{p}_0)^5 \right\}.$$

From (A.2), I given by (A.1) is divided into three parts, say $I = I_1 + I_2 - I_3$, where

$$I_1 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\exp\{-(s^\top \Omega s + 2is^\top u)/2\}] \left\{ 1 + \sum_{m=1}^{3} n^{-m/2} b_m(s) \right\} ds,$$

$$I_2 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\exp\{-(s^\top \Omega s + 2is^\top u)/2\}] O(n^{-2}) ds,$$

$$I_3 = \int \cdots \int_{Q^c} [\exp\{-(s^\top \Omega s + 2is^\top u)/2\}] \left\{ 1 + \sum_{m=1}^{3} n^{-m/2} b_m(s) + O(n^{-2}) \right\} ds$$

and $Q = [-\sqrt{n}\pi, \sqrt{n}\pi] \times \cdots \times [-\sqrt{n}\pi, \sqrt{n}\pi]$. Evaluation $I = I_1 + O(n^{-2})$ is derived by using $I_2 = O(n^{-2})$ and $I_3 = o(n^{-2})$. Therefore, we find

$$\Pr(U = u|\mathcal{H}_0) = (2\pi\sqrt{n})^{-(N-1)} \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\exp\{-(s^\top \Omega s + 2is^\top u)/2\}] \left\{ 1 + \sum_{m=1}^{3} n^{-m/2} b_m(s) \right\} ds + O(n^{-2}) \right].$$

By calculating the integral of the above expression, we find the results stated in Theorem 1. □

A.2. Proof of Theorem 2

By the transformation (7), the statistic C_ϕ can be rewritten as

$$C_\phi(U) = 2n \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{\ell=1}^{L} q_{jk\ell} G_{jk\ell} \phi(D_{jk\ell}),$$

where for each $j \in \{1, \dots, J\}$, $k \in \{1, \dots, K\}$, $\ell \in \{1, \dots, L\}$,

$$G_{jk\ell} = \left(1 + \frac{U_{j\cdot\ell}}{\sqrt{n} p_{j\cdot\ell}}\right)\left(1 + \frac{U_{\cdot k\ell}}{\sqrt{n} p_{\cdot k\ell}}\right)\left(1 + \frac{U_{\cdot\cdot\ell}}{\sqrt{n} p_{\cdot\cdot\ell}}\right)^{-1}, \quad D_{jk\ell} = \left(1 + \frac{U_{jk\ell}}{\sqrt{n} q_{jk\ell}}\right) G_{jk\ell}^{-1},$$

with $U_{j\cdot\ell} = U_{j1\ell} + \cdots + U_{jK\ell}$, $U_{\cdot k\ell} = U_{1k\ell} + \cdots + U_{Jk\ell}$, and $U_{\cdot\cdot\ell} = \sum_{j=1}^{J} \sum_{k=1}^{K} U_{jk\ell}$. Since W_1 is multivariate Edgeworth expansion, W_1 is represented as

$$W_1 = \int \cdots \int_{B_\phi(x)} f(u)\{1 + h_1(u)/\sqrt{n} + h_2(u)/n + h_3(u)/(n\sqrt{n})\} du + O(n^{-2}),$$

where $B_\phi(x)$ is defined by (16), $du = du_{111} \cdots du_{JK,L-1}$, and $f(u)$, $h_1(u)$, $h_2(u)$ and $h_3(u)$ are defined by (10) and (12)–(14), respectively. We derive an approximation of the characteristic function of $C_\phi(U)$, i.e.,

$$\psi_\phi(t) = \int_{\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\{itC_\phi(u)\} f(u)\{1 + h_1(u)/\sqrt{n} + h_2(u)/n + h_3(u)/(n\sqrt{n})\} du \tag{A.3}$$

up to order $n^{-3/2}$. By the assumption on $\phi(t)$, we obtain

$$C_\phi(u) = \tau_0(u) + \tau_1^\phi(u)/\sqrt{n} + \tau_2^\phi(u)/n + \tau_3^\phi(u)/(n\sqrt{n}) + O(n^{-2}), \tag{A.4}$$

where $\tau_0(u) = u^\top \Omega^{-1} u + M(1)_2 - M(2)_{0,2}$,

$$\begin{aligned}\tau_1^\phi(u) &= \phi'''(1) M_2^3/3 + \{4\phi'''(1)/3 + 1\} M(1)_3 + \{2\phi'''(1)/3 + 1\} M(2)_{0,3} \\ &\quad + \{\phi'''(1) + 1\}\{-2M(2)_{1,2} + 2M(3)_{0,1,1} + M(3)_{1,0,2} - M(4)_{0,0,1,2}\},\end{aligned}$$

$$\begin{aligned}\tau_2^\phi(u) = \phi^{(4)}(1) M_3^4/12 &+ \{3\phi^{(4)}(1)/4 + 4\phi'''(1)/3\} M(1)_4 - \{\phi^{(4)}(1)/4 + 4\phi'''(1)/3 + 1\} M(2)_{0,4} \\ &- \{3\phi^{(4)}(1)/2 + 4\phi'''(1) + 1\} M(2)_{2,2} \\ &+ \{\phi^{(4)}(1)/2 + 2\phi'''(1) + 1\} \left\{2M(2)_{1,3} + M(3)_{0,2,0} + M(4)_{0,0,2,2} - 2M(4)_{0,1,2,1}\right\} \\ &+ \{\phi^{(4)}(1) + 2\phi'''(1)\}\{M(3)_{2,0,2}/2 + M(3)_{1,0,3}/3 - M(4)_{0,0,1,3}/3\} \\ &+ \{\phi^{(4)}(1) + 3\phi'''(1) + 1\}\{M(3)_{0,1,2} + 2M(3)_{1,1,1} - M(4)_{1,0,1,2}\},\end{aligned}$$

$$M(1)_a = \sum_{\ell=1}^{L} u_{\cdot\cdot\ell}^a / p_{\cdot\cdot\ell}^{a-1}, \quad M(2)_{a,b} = \sum_{j=1}^{J} \sum_{\ell=1}^{L} (u_{\cdot\cdot\ell}/p_{\cdot\cdot\ell})^a \, u_{j\cdot\ell}^b / p_{j\cdot\ell}^{b-1} + \sum_{k=1}^{K} \sum_{\ell=1}^{L} (u_{\cdot\cdot\ell}/p_{\cdot\cdot\ell})^a \, u_{\cdot k\ell}^b / p_{\cdot k\ell}^{b-1},$$

$$M(3)_{a,b,c} = \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{\ell=1}^{L} (u_{\cdot\cdot\ell}/p_{\cdot\cdot\ell})^a \, \{(u_{j\cdot\ell} u_{\cdot k\ell})/(p_{j\cdot\ell} p_{\cdot k\ell})\}^b u_{jk\ell}^c / q_{jk\ell}^{c-1},$$

$$M(4)_{a,b,c,d} = \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{\ell=1}^{L} (u_{\cdot\cdot\ell}/p_{\cdot\cdot\ell})^a \{(u_{j\cdot\ell}/p_{j\cdot\ell})^b (u_{\cdot k\ell}/p_{\cdot k\ell})^c + (u_{j\cdot\ell}/p_{j\cdot\ell})^c (u_{\cdot k\ell}/p_{\cdot k\ell})^b\} u_{jk\ell}^d / q_{jk\ell}^{d-1},$$

where $\tau_3^\phi(u)$ is a homogeneous polynomial of degree 5 with respect to the variables $u_{111}, \dots, u_{JK,L-1}$. By substituting the expanded expression of $\exp\{itC_\phi(u)\}$ obtained by using (A.4) for $\exp\{itC_\phi(u)\}$ in (A.3), we find

$$\psi_\phi(t) = (|\Omega|/|\Lambda|)^{-1/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_0(u) g_\phi(u) du + O(n^{-2}), \tag{A.5}$$

where

$$h_0(u) = (2\pi)^{-(N-1)/2} |\Lambda|^{-1/2} \exp(-u^\top \Lambda^{-1} u/2),$$

$$g_\phi(u) = 1 + \frac{1}{\sqrt{n}}\{h_1(u) + (it)\tau_1^\phi(u)\} + \frac{1}{n}[h_2(u) + (it)\tau_1^\phi(u) h_1(u) + (it)\tau_2\phi(u) + (it)^2\{\tau_1^\phi(u)\}^2/2] + \frac{1}{n\sqrt{n}} Q_0(u),$$

$$\Lambda = (1 - 2it)^{-1}(\Omega - 2it\Omega\Xi\Omega),$$

$$\Xi = (I_{N-1}, -1_{N-1}) \left((1_K 1_K^\top) \otimes D_1^{-1} \oplus \cdots \oplus (1_K 1_K^\top) \otimes D_J^{-1} + (1_J 1_J^\top) \otimes D_{(2)}^{-1} - (1_J 1_J^\top) \otimes (1_K 1_K^\top) \otimes D_{(3)}^{-1} \right) (I_{N-1}, -1_{N-1})^\top ,$$

and for all $j \in \{1, \ldots, J\}$, $D_j = \mathrm{diag}(p_{j\cdot 1}, \ldots, p_{j\cdot L})$, $D_{(2)} = \mathrm{diag}(p_{\cdot 11}, \ldots, p_{\cdot 1L}, \ldots, p_{\cdot K1}, \ldots, p_{\cdot KL})$ and $D_{(3)} = \mathrm{diag}(p_{\cdot\cdot 1}, \ldots, p_{\cdot\cdot L})$, where N is given by (9), $\otimes$ and $\oplus$ denote the Kronecker product and direct sum of matrices, and the degrees of all terms of polynomial $Q_0(u)$ are odd.

We compute the integral in (A.5) by using the equation $|\Omega|/|\Lambda| = (1-2it)^M$. Then,

$$\psi_\phi(t) = (1-2it)^{-M/2} \left\{ 1 + \frac{1}{n} \sum_{j=0}^{3} (1-2it)^{-j} v_j^\phi + O(n^{-2}) \right\}, \tag{A.6}$$

where $v_0^\phi, \ldots, v_3^\phi$ are defined in (15). The results of Theorem 2 are derived by inverting (A.6).

References

[1] A. Agresti, Categorical Data Analysis, third ed., Wiley, Hoboken, NJ, 2013.
[2] O.E. Barndorff-Nielsen, D.R. Cox, Bartlett adjustments to the likelihood ratio statistic and the distribution of maximum likelihood estimator, J. R. Stat. Soc. Ser. B Stat. Methodol. 46 (1984) 483–495.
[3] O.E. Barndorff-Nielsen, P. Hall, On the level-error after Bartlett adjustment of the likelihood ratio statistic, Biometrika 75 (1988) 374–378.
[4] G.M. Cordeiro, S.L.P. Ferrari, A modified score test statistic having chi-squared distribution to order n^{-1}, Biometrika 78 (1991) 573–582.
[5] N. Cressie, T.R.C. Read, Multinomial goodness-of-fit tests, J. R. Stat. Soc. Ser. B Stat. Methodol. 46 (1984) 440–464.
[6] D. Edwards, S. Kreiner, The analysis of contingency tables by graphical models, Biometrika 70 (1983) 553–565.
[7] Y. Fujikoshi, A method for improving the large-sample chi-squared approximations to some multivariate test statistics, Amer. J. Math. Manag. Sci. 17 (1997) 15–29.
[8] Y. Fujikoshi, Transformations with improved chi-squared approximations, J. Multivariate Anal. 72 (2000) 249–263.
[9] Y. Kakizawa, Higher order monotone Bartlett-type adjustment for some multivariate test statistics, Biometrika 83 (1996) 923–927.
[10] T. Kobe, N. Taneichi, Y. Sekiya, Improved transformed statistics for the test of one factor independence from the other two in an $r \times s \times t$ contingency table, J. Japan Statist. Soc. 45 (2015) 77–98.
[11] D.N. Lawley, A general method for approximating to the distribution of the likelihood ratio criteria, Biometrika 43 (1956) 295–303.
[12] M.L. Menéndez, J.A. Pardo, L. Pardo, M.C. Pardo, Asymptotic approximations for the distributions of the (h, ϕ)-divergence goodness-of-fit statistics: Application to Rényi's statistic, Kybernetes 26 (1997) 442–452.
[13] L. Pardo, Statistical Inference Based on Divergence Measures, Chapman & Hall/CRC, Boca Raton, FL, 2006.
[14] T.R.C. Read, Closer asymptotic approximations for the distributions of the power divergence goodness-of-fit statistics, Ann. Inst. Statist. Math. 36 (1984) 59–69.
[15] T.R.C. Read, Small-sample comparisons for the power divergence goodness-of-fit statistics, J. Amer. Statist. Assoc. 79 (1984) 929–935.
[16] Y. Sekiya, N. Taneichi, Improvement of approximations for the distributions of multinomial goodness-of-fit statistics under nonlocal alternatives, J. Multivariate Anal. 91 (2004) 199–223.
[17] M. Siotani, Y. Fujikoshi, Asymptotic approximations for the distributions of multinomial goodness-of-fit statistics, Hiroshima Math. J. 14 (1984) 115–124.
[18] B.E. Storer, C. Kim, Exact properties of some exact test statistics for comparing two binomial proportions, J. Amer. Statist. Assoc. 85 (1990) 146–155.
[19] N. Taneichi, Y. Sekiya, Improved transformed statistics for the test of independence in $r \times s$ contingency tables, J. Multivariate Anal. 98 (2007) 1630–1657.
[20] N. Taneichi, Y. Sekiya, A. Suzukawa, An asymptotic approximation for the distribution of ϕ-divergence multinomial goodness-of-fit statistic under local alternatives, J. Japan Statist. Soc. 31 (2001) 207–224.
[21] N. Taneichi, Y. Sekiya, A. Suzukawa, Asymptotic approximations for the distributions of the multinomial goodness-of-fit statistics under local alternatives, J. Multivariate Anal. 81 (2002) 335–359.
[22] N. Taneichi, Y. Sekiya, J. Toyama, Test of complete independence in multi-dimensional contingency tables based on transformed statistics, submitted for publication.
[23] N.S. Tang, M.L. Tang, Exact unconditional inference for risk ratio in a correlated 2×2 table with structural zero, Biometrics 58 (2002) 972–980.
[24] J.K. Yarnold, Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set, Ann. Math. Stat. 43 (1972) 1566–1580.
[25] K. Zografos, Asymptotic properties of ϕ-divergence statistic and its application in contingency tables, Internat. J. Math. Statist. Sci. 2 (1993) 5–21.