sciendo DOI: 10.2478/bile-2019-0003

Biometrical Letters
Vol. 56 (2019), No. 1, 17-43

The LMS for testing independence in two-way contingency tables

Piotr Sulewski

The Pomeranian University, Institute of Mathematics, 76-200 Słupsk, Poland,
e-mail: piotr.sulewski@apsl.edu.pl

SUMMARY

In the statistical literature there are proposed many test measures to determine the independence of two qualitative variables in contingency tables, in particular in two-way contingency tables larger than 2×2. For statistical analysis, three of the so-called "chi-squared tests"—the T_3 test, BP test and $|\chi|$ test—were selected. These tests were compared with a logarithmic minimum test, which is the author's proposal. Critical values for the tests were determined with the Monte Carlo method. To compare the tests, an appropriate measure of untruthfulness of H_0 was used and the power of the tests was calculated.

Key words: independence test; contingency table; Monte Carlo method; generating contingency tables.

1. Introduction

Independence tests are probably one of the most commonly used statistical tools. Test data are arranged in the form of contingency tables (CTs), in particular $w \times k$ CTs. The (Pearson's) χ^2 test and the log likelihood ratio G^2 test are the best-known and the most commonly used. Garside and Mack (1976) numerically compared the sizes of the χ^2 test and some of its corrected versions. The authors noted that, although the corrected versions are conservative in nature, the χ^2 test has the size closest to the nominal level α. For small CTs (not applicable to the 2×2 case) with small sample sizes, Lawal and Uptong (1984) suggested a modification to the χ^2 test to make the size closer to the nominal level α. There are numerous publications on CTs and the χ^2 test of independence—e.g. Meng and Chapman (1966); Diaconis and Efron (1985); Albert (1990); Andrés et al. (1995)—where the χ^2 test statistics are interpreted from various angles. Information about approximations of χ^2 and G^2 can be found in Cochran (1952); Cochran (1954); Koehler and Larntz (1980); and Cressie and Read (1989). The χ^2 and G^2 tests provide consistent and asymptotically unbiased tests of independence (Haberman, 1981). These test statistics belong to the power divergence statistics (PDS) family (Cressie and Read, 1984).

The Fisher exact test (Fisher, 1922) is also popular. It was independently developed by Irwin (1935) and is also known as the Fisher–Irwin (FI) test. The FI test is most commonly applied to 2×2 CTs, because it can be computationally time-consuming for tables larger than 2×2. Campbell (2007) recommended the use of the χ^2 test for large sample sizes and the FI test for small sample sizes. Some authors have argued that the FI test is conservative, i.e. that its actual rejection rate is below the nominal significance level (Liddell and Douglas, 1976; D'Agostino et al., 1988). Lydersen et al. (2009) recommended that the FI test should practically never be used. Berry and Mielke (1988) used Monte Carlo methods to assess how two asymptotic χ^2 tests, two asymptotic G^2 tests and a recently developed nonasymptotic χ^2 test fit the models specified by the null hypotheses of independence and homogeneity. The results of the study indicate that the nonasymptotic χ^2 test is superior in overall performance to the other analyzed tests. Lawal and Uptong (1990) compared the PDS with modified χ^2 test statistics (Lawal and Uptong, 1984) by means of the statistical power. Cohen and Nee (1990) used Monte Carlo methods and calculated the statistical power using the Rao F-test in CTs. Davis (1993) described a generalized chi-square approximation to the distribution of the χ^2 test statistics for testing independence in CTs. The new method consistently yields an estimated p-value approximate to the exact result. Yenigün et al. (2011) carried out a simulation study to observe the empirical power performance of the maximal correlation test and compared it with $\chi^2$and G^2 independence tests. When the underlying continuous variables are uncorrelated but dependent, the authors pointed out some cases for which the maximal correlation test appears to be more powerful. In the paper by Sulewski

(2013) a |χ| test, which is a modification of the χ^2 test for $w \times k$ CTs, was proposed. The |χ| test was compared with the PDS for selected sizes of a CT larger than 2×2 in terms of their power (Sulewski, 2016). Yu (2014) allows the margins to be random and compares the power of the G^2, the Bayes factor and the FI tests. Shan and Wilding (2015) extend the unconditional approach based on estimation and maximization to designs with a fixed total sum. The procedures based on the χ^2, Yates's corrected and G^2 test statistics are evaluated with regard to actual type I error rates and powers. Lipsitz et al. (2015) propose Wald and score test statistics for independence based on weighted least squares estimating equations. In contrast to the Rao–Scott test statistics, the proposed Wald and score test statistics exist unconditionally. Comparing the Rao–Scott test statistics, the score statistics and Wald statistics with respect to power, it was found that the Wald test statistics had the highest power. Vélez et al. (2016) propose and illustrate a new graphical method of performing diagnostic analyses in two-way CTs. In this method one observation is added or removed from each cell at a time, whilst the other cells are held constant, and the change in the test statistic of interest is represented graphically.

The bootstrap method is an indispensable tool for testing statistical hypotheses. Using resampling, bootstrapping approximates the sampling distribution of a statistic under the null (or the alternative) hypothesis. Amiri and von Rosen (2011) show that the nonparametric bootstrap method is more efficient than the χ^2 statistic, the χ^2 statistic with a Yates' correction and the FI test. Lin et al. (2015) explore the accuracy of the $\chi^2$and G^2 tests through an extensive simulation study and then propose bootstrap versions that appear to work better than the asymptotic tests in terms of adhering to the nominal level for small to large sample sizes as well as extreme cell frequencies. The proposed bootstrap tests are more convenient than the FI test, which is also often criticized for being too conservative. Amiri and Modarres (2017) proposed a bootstrap test statistic that provides more accurate inference for small sample sizes.

In this paper we propose the new logarithmic minimum statistic (LMS) for $w \times k$ CTs and compare it with six other statistics. The first is the well-known and commonly used χ^2 test statistic (Pearson, 1904). The second and third are the G^2 test statistic (Cressie and Read, 1984) and the Neyman modified χ^2 test statistic (Cressie and Read, 1984), which together with the χ^2 statistic represent the PDS. The fourth is the $|\chi|$ statistic (Sulewski, 2013). The fifth is the T_3 statistic (Amiri and von Rosen, 2011), and the sixth is the BP test (Amiri and Modarres, 2017). Critical values were determined by means of the Monte Carlo simulation method. For the above test statistics, the power of the tests (PoT) was determined. To calculate the PoT, we generate $w \times k$ CTs. At the same time, an appropriate measure of untruthfulness of H_0 (MoU) for six probability scenarios was defined. At the end of the paper, three examples are presented and discussed.

This article is organized as follows. Section 2 describes four variants of the presentation of CTs. Section 3 presents the new LMS and six other test statistics. Section 4 is devoted to CT modelling and presents six probability scenarios with a data flow parameter. Section 5 presents the measure of untruthfulness of H_0 (MoU) for the probability scenarios in question. The power of the considered tests is determined in section 6, and three examples are presented in section 7.

2. Variants of presentation of CTs

Let X, Y be two features of the same object and let them have respectively levels $X_1, \ldots, X_w$, $Y_1, \ldots, Y_k$. Testing these two features for independence with an appropriately arranged CT is probably one of the most common tasks performed by statisticians.

Nowadays there are four major variants of the presentation of CTs, each of which serves a specific purpose. These are detailed below:

- TP Variant (theoretical probabilities). Cells contain probabilities intrinsic to the phenomenon under investigation. The exact values of these probabilities are unknown to the investigator. In further sections of this paper CTs will be first simulated with the Monte Carlo method, and then we will apply the CT variant filled with probabilities arbitrarily set by the Monte Carlo experimenter.

- EC Variant (experimental counts). Cells contain counts observed on a sample drawn from the general population subject to investigation.
- TC Variant (theoretical counts). Cells contain expected theoretical counts. These counts are theoretical in the sense that they result from the TP variant.
- EP Variant (empirical probabilities). Cells that result from the EC variant and contain estimates of the unknown content of TP.

3. Tests of independence selected for the Monte Carlo study

Statistical science has been enriched with many statistics proposed as tests for independence. However, in practice and in statistical software, the most popular and important are the χ^2 statistics, especially for CTs larger than 2×2. For small sample sizes, the critical values for χ^2 statistics can be determined by simulation methods. In relation to $w \times k$ CTs, the χ^2 test statistic is defined as

$$Q_1 = \chi^2 = \sum_{i=1}^{w} \sum_{j=1}^{k} \frac{\left(n_{ij} - e_{ij}\right)^2}{e_{ij}} \tag{1}$$

where $e_{ij} = n_{i+} n_{+j}/n$ are expected counts and the plus signs denote summation over a row or a column. The statistic (1) asymptotically (i.e. sample size $n \to \infty$) follows the chi-square distribution with $(w-1)(k-1)$ degrees of freedom, provided that the hypothesis H_0 of the independence of X and Y is true.

Cressie and Read (1984) proposed the power divergence statistics (PDS). A PDS for $w \times k$ CTs is given by

$$P^2 = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{w} \sum_{j=1}^{k} n_{ij} \left[\left(\frac{n_{ij}}{e_{ij}} \right)^{\lambda} - 1 \right] (\lambda \neq -1, 0)$$

which is always positive, and can be defined by the limit of P^2 at $\lambda = -1$ and $\lambda = 0$. This is a rich class containing many test statistics, including:

- the χ^2 test statistic ($\lambda = 1$), see formula (1),
- the log likelihood ratio G^2 test statistic (the limit as λ goes to 0)

$$Q_2 = G^2 = 2 \sum_{i=1}^{w} \sum_{j=1}^{k} n_{ij} \, ln \left(\frac{n_{ij}}{e_{ij}} \right), \tag{2}$$

- the Freeman–Tukey test statistic ($\lambda = -0.5$)

$$FT = 4\sum_{i=1}^{w}\sum_{j=1}^{k}\left(\sqrt{n_{ij}} - \sqrt{e_{ij}}\right)^2,$$

- the modified G^2 test statistic (the limit as λ goes to –1)

$$KL = 2\sum_{i=1}^{w}\sum_{j=1}^{k} e_{ij}\, ln\left(\frac{e_{ij}}{n_{ij}}\right),$$

- the Neyman modified χ^2 test statistic ($\lambda = -2$)

$$Q_3 = N = \sum_{i=1}^{w}\sum_{j=1}^{k}\frac{\left(n_{ij}-e_{ij}\right)^2}{n_{ij}}, \qquad (3)$$

- the Cressie–Read test statistic ($\lambda = 2/3$)

$$CR = \frac{9}{5}\sum_{i=1}^{w}\sum_{j=1}^{k} n_{ij}\left[\left(\frac{n_{ij}}{e_{ij}}\right)^{2/3} - 1\right].$$

Sulewski (2016) showed that among the PDS the Neyman modified χ^2 test statistic and the χ^2 and CR tests have similar powers. Amiri and von Rosen (2011) considered the χ^2 test statistic (1) and the Neyman modified χ^2 test statistic (3), while Lin et al. (2015) considered the χ^2 test statistic (1) and the G^2 test statistic (2). Therefore, in this paper, statistics (1)–(3) from the PDS family were selected for the Monte Carlo study.

Sulewski (2013) proposed the $|\chi|$ statistic, which is a modification of the χ^2 statistic and is given by

$$Q_4 = |\chi| = \sum_{i=1}^{w}\sum_{j=1}^{k}\frac{\left|n_{ij}-e_{ij}\right|}{e_{ij}}$$

The $|\chi|$ statistic was compared in terms of power with the PDS for CTs larger than 2×2 (Sulewski, 2016) and for three-way CTs of small sizes (Sulewski, 2018). It was shown that the $|\chi|$ test is more powerful than the PDS.

The author's proposal is the logarithmic minimum statistic (LMS). The LMS for 2×2 CTs was introduced in (Sulewski, 2017). This statistic for two-way CTs is defined as follows:

$$Q_5 = LMS = -\sum_{i=1}^{w}\sum_{j=1}^{k} ln\left[\frac{min(n_{ij},e_{ij})}{max(n_{ij},e_{ij})}\right] \qquad (4)$$

Formula (4) shows that $n_{ij} \neq 0$ and $e_{ij} \neq 0$ for each $i = 1, \dots w;\ j = 1, .., k$. For this reason, the sample size cannot be too small to obtain the power of the test for different scenarios. Details appear in the following section.

It is well-understood that resampling must reflect the null hypothesis. It is essential to resample the CT, assuming that $p_{ij} = p_{i+}p_{+j}$ holds. When testing the independence of two categorical variables, Amiri and von Rosen (2011) and Lin et al. (2015) use the expectation of cells under the null hypothesis: $H_0: e_{ij} = n_{i+}n_{+j}/n$.

Amiri and von Rosen (2011) considered the χ^2 test statistic (1), the bootstrap version of the χ^2 test statistic, the Neyman modified χ^2 test statistic (3) and a test statistic defined as

$$Q_6 = T_3 = \sum_{i=1}^{w}\sum_{j=1}^{k} \frac{\left(n_{ij}-e_{ij}\right)^2}{wk}.$$

Lin et al. (2015) considered the χ^2 test statistic (1) and the G^2 test statistic (2) as well as their bootstrap versions. The main advantage of their methods is that the bootstrap methods give sizes of tests very close to the nominal level, especially for small sample sizes. Tables 5–10 show that sizes of the χ^2 and G^2 tests for the analyzed sample sizes are identical (to three decimal places) to the nominal level. More accurate sizes of the χ^2 and G^2 tests are e.g.:

a) 2×3 CT, scenario I, $n = 50$

- $\alpha = 0.05$: 0.0500063 (χ^2), 0.050008 (G^2)
- $\alpha = 0.1$: 0.100054 (χ^2), 0.1000017 (G^2)

b) 3×3 CT, scenario VI, $n = 40$

- $\alpha = 0.05$: 0.050005 (χ^2), 0.050003 (G^2)
- $\alpha = 0.1$: 0.10001 (χ^2), 0.100008 (G^2)

In this case, there is no need to apply the bootstrap methods in the Monte Carlo simulation.

We can convert the cell counts of the $w \times k$ CT $(n_{11}, \dots, n_{1k}; \dots; n_{w1}, \dots, n_{wk})$ to $(n_1, n_2, \dots, n_N)$, where n_u are the n_{ij} values indexed row by row. A new variable for each cell is $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_N)^t$ and the associated probabilities are $\boldsymbol{p} = (p_1, \dots, p_N)^t$. For a given CT, the variable $\boldsymbol{Z}$ and cell counts follow a multinomial distribution with $n = \sum_{u=1}^{N} n_k$ samples and probabilities $\boldsymbol{p}$. We can write this as $\boldsymbol{Z} \sim Multi(n, \boldsymbol{p})$.

Let $\boldsymbol{z} = (z_1 = n_1, z_2 = n_2, \dots, z_N = n_N)$ be a multinomial sample with $n = \sum_{i=1}^{N} n_i$. Estimates of the sample proportions are $\widehat{\boldsymbol{p}} = (\hat{p}_1, \dots, \hat{p}_N)$, where $\hat{p}_j = n_j / n$. The bootstrap resample is defined as sampling with replacement from the elements of $\mathbf{z}$ with size n. The bootstrap estimates of the proportions are $\boldsymbol{p}^* = (p_1^*, \dots, p_N^*)$ where $p_i^* = n_i^* / n$.

Amiri and Modarres (2017) proposed the BP test using a test statistic for the bootstrap sample defined as

$$Q_7 = BP = n(\boldsymbol{p}^* - \boldsymbol{p}_0)^t A (\boldsymbol{p}^* - \boldsymbol{p}_0)$$

where $\boldsymbol{p}_0$ is calculated under $H_0\colon p_{ij} = p_{i+} p_{+j}$, $\sum p = Diag(\boldsymbol{p}) - \boldsymbol{p}^t \boldsymbol{p}$, $A = \sum p^{-1}$ and $\mathbf{p}$ is the vector of observed proportions. Since the inverse of $\sum p$ does not exist ($\det(\sum p) = 0$), we use the Moore–Penrose generalized inverse, implemented in major programming environments such as R, Mathcad and Mathematica.

The above test statistics $Q_1 - Q_7$ were selected for the Monte Carlo study.

4. Modeling how CTs are generated

Let us treat CT as a mathematical expression of a certain phenomenon being considered. This phenomenon makes features X and Y mutually dependent in a statistical sense. Saying this, we have in mind that there is an intrinsic mechanism behind this phenomenon. This mechanism not only makes the phenomenon occur, but also determines the probabilities of particular X, Y combinations. Figuratively speaking, the phenomenon fills cells of the relevant CT. Let us consider the tables shown below:

$$T_{2\times3} = \begin{bmatrix} 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 \end{bmatrix},\ T_{3\times3} = \begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix},$$

$$T_{w\times k} = \begin{bmatrix} 1/(wk) & \dots & 1/(wk) \\ \dots & \dots & \dots \\ 1/(wk) & \dots & 1/(wk) \end{bmatrix}$$

and take them as the "progenitors" of all possible $w \times k$ CTs.

Continuing this line of reasoning, we have to develop a scenario for a CT's offspring. A simple scenario that offers a prospect of wide applicability is one in which portions of probability equal to a flow between cells of $T_{w\times k}$. We can also conceptualize an advanced scenario where a is divided into two sub-portions that may flow independently between cells. In this and further sections we focus only on six scenarios related to $2 \times 3, 3 \times 3$ CTs. These scenarios are listed in Tables 1–2 and denoted by numbers from I to VI. One may, of course, anticipate a variety of modifications of these, as is common in statistics. The scenarios appear to be fundamental, so the corresponding PoT will be determined for further comparisons just for these scenarios.

Table 1. Contents of $\mathbf{2 \times 3}$ CTs resulting from given scenarios

		Scenario I	
	Y_1	Y_2	Y_3
X_1	$1/6-a$	$1/6$	$1/6$
X_2	$1/6$	$1/6$	$1/6+a$
		Scenario II	
	Y_1	Y_2	Y_3
X_1	$1/6-a$	$1/6-a$	$1/6$
X_2	$1/6+a$	$1/6+a$	$1/6$
		Scenario III	
	Y_1	Y_2	Y_3
X_1	$1/6-a$	$1/6$	$1/6+a$
X_2	$1/6+a$	$1/6$	$1/6-a$

Table 2. Contents of $\mathbf{3} \times \mathbf{3}$ CTs resulting from given scenarios

Scenario IV			
	Y_1	Y_2	Y_3
X_1	$1/9-a$	$1/9-a/2$	$1/9$
X_2	$1/9-a/2$	$1/9$	$1/9+a/2$
X_3	$1/9$	$1/9+a/2$	$1/9+a$
Scenario V			
	Y_1	Y_2	Y_3
X_1	$1/9-a$	$1/9-a/2$	$1/9$
X_2	$1/9$	$1/9$	$1/9$
X_3	$1/9+a$	$1/9+a/2$	$1/9$
Scenario VI			
	Y_1	Y_2	Y_3
X_1	$1/9-a$	$1/9$	$1/9+a$
X_2	$1/9$	$1/9$	$1/9$
X_3	$1/9+a$	$1/9$	$1/9-a$

In all of the above scenarios the inflow/outflow portion $|a| \leq 1/(wk)$. Scenarios may locally mutate, for example, by transposition of rows or columns.

The scenarios put forward here are very simple equal-portion scenarios. Of course, the real-life scenarios according to which particular CTs are generated may be similar to those presented above, as is typical for relations between theory and real life. The simple Exponential and Gaussian distributions turned out to be indispensable in practice.

Table 3 presents 3×3 CTs under scenario IV with empirical counts (EC) and empirical probabilities (EP), where $a = 1/15$.

If the n objects in the sample are independently and identically distributed, then the vector of cell counts $\boldsymbol{Z} = (n_{11}, \dots, n_{1k}; \dots; n_{w1}, \dots, n_{wk})^T$ has multinomial distribution as $\boldsymbol{Z} \sim Multi(n, a)$. Each scenario has a multinomial distribution of its own. Particular formulae are easy to obtain by substituting probabilities embedded in the multinomial distribution with probabilities taken from the relevant cells of the CT. One must then laboriously simplify the results of the substitutions. The following distributions reflect scenarios I–II:

Table 3. EC and EP variants of $\mathbf{3 \times 3}$ CTs under scenario IV and $\boldsymbol{n = 96}$

EC Variant

	Y_1	Y_2	Y_3	Total
X_1	4	7	11	22
X_2	7	11	14	32
X_3	11	14	17	42
Total	22	32	42	96

EP Variant

	Y_1	Y_2	Y_3	Total
X_1	4/90	7/90	10/90	21/90
X_2	7/90	10/90	13/90	30/90
X_3	10/90	13/90	16/90	39/90
Total	21/90	30/90	39/90	1

$$P_I(n_{11}, n_{12}, n_{13}, n_{21}, n_{22}, n_{23}n, a) = \frac{n!}{6^n n_{11}! n_{12}! n_{13}! n_{21}! n_{22}! n_{23}!} \cdot (1-6a)^{n_{11}} \cdot (1+6a)^{n_{23}} \cdot$$

$$P_{II}(n_{11}, n_{12}, n_{13}, n_{21}, n_{22}, n_{23}n, a) = \frac{n!}{6^n n_{11}! n_{12}! n_{13}! n_{21}! n_{22}! n_{23}!} \cdot (1-6a)^{n_{11}+n_{12}} \cdot (1+6a)^{n_{21}+n_{22}}.$$

To enable the wide applicability of the probabilistic model, its flexibility and ability to estimate have to be equilibrated. The ability to estimate means the effectiveness of estimation of parameters. Flexibility is mainly determined by the number of parameters embedded in a model (but also by the places which parameters occupy in a model formula). Increasing flexibility results in decreasing ability to estimate. Obviously one can add a_1, a_2 to the model (and even a_3) for good measure, but in this way an ineffective "sample glutton" will be created. Since the statistical inference method put forward in this paper is oriented rather towards small samples, the author is confident that a one-parameter model achieves the equilibrium.

5. Measure of untruthfulness of H_0

As we have already stated in section 2, certain classes of feature X are ascribed to rows and certain classes of feature Y are ascribed to columns. The features X, Y

are independent and H_0 is true, if $p_{ij} = p_{i+}p_{+j}$, of course. When this equality is not fulfilled, H_0 does not hold and an appropriate measure of untruthfulness of H_0 (MoUH) is needed. There are many different measures in the literature, inter alia: Pearson's φ, Tschuprow's T, Cramer's V, the corrected contingency C, and Goodman and Kruskal's τ.

Sulewski (2017) proposed an MoUH for 2×2 CTs denoted by MoU. This measure for $w \times k$ CTs is defined as

$$MoU = \sum_{i=1}^{w} \sum_{j=1}^{k} \left| p_{ij} - p_{i+}p_{+j} \right|.$$

Replacing theoretical probabilities by empirical ones (EP Variant), we obtain the sample MoU as

$$MoU_e = \frac{1}{n} \sum_{i=1}^{w} \sum_{j=1}^{k} \left| n_{ij}^* - \frac{n_{i+}^* n_{+j}^*}{n} \right| = \frac{1}{n} \sum_{i=1}^{w} \sum_{j=1}^{k} \left| n_{ij}^* - e_{ij}^* \right|.$$

The MoU measure takes values in $[0,1]$, and will be applied in the Monte Carlo simulation. This measure undoubtedly reflects the essence of H_0 and seems to be of a very simple form. In the statistical literature there are measures very similar to the MoU. These less-known measures are:

- D test statistic (Sulewski, 2014)

$$D = \sum_{i=1}^{w} \sum_{j=1}^{k} \left(p_{ij} - p_{i\bullet} p_{\bullet j} \right)^2,$$

- Belson test statistic (Marcotorchino, 1984)

$$B = n^2 \sum_{i=1}^{w} \sum_{j=1}^{k} \left(p_{ij} - p_{i\bullet} p_{\bullet j} \right)^2,$$

- Jordan test statistic (Marcotorchino, 1984)

$$J = n \sum_{i=1}^{w} \sum_{j=1}^{k} \left(p_{ij} - p_{i\bullet} p_{\bullet j} \right)^2,$$

- Variation of Squares test statistic (Marcotorchino, 1984)

$$E = n^2 \sum_{i=1}^{w} \sum_{j=1}^{k} \left(p_{ij} - p_{i\bullet} p_{\bullet j} \right)\left(p_{ij} + p_{i\bullet} p_{\bullet j} \right).$$

To go further, we need to take notice of the relations between the MoUH and test statistics (TS). MoUH and TS are functions of CT cell contents, although there is a fundamental difference between them: while MoUH takes appropriate values of cell probabilities, TS takes only relevant estimates of probabilities. As a result, when H_0 is true, MoUH is (by definition) equal to zero. Conversely, TS may be, and very often is, of non-zero value when H_0 is true. It may occasionally be equal to zero when H_0 is false. This is because the MoUH is a non-random variable while the test statistic is a random variable. Each test statistic can be converted to MoUH by replacing estimates with actual values. Conversely, each MoUH can, in a similar way, be converted to test statistics. Table 4 presents the MoU formulae for $w \times k$ CTs under scenarios I–VI and maximum values of MoU.

Table 4. MoU under scenarios I–III (CT 2×3), IV–VI (CT 3×3), where $\boldsymbol{a} \in [\mathbf{0}, \mathbf{1}/(\boldsymbol{wk})]$

Scenario	MoU	MoU_{max}	Scenario	MoU	MoU_{max}
I	4a/3	2/9	IV	$9a^2$	1/9
II	8a/3	4/9	V	2a	2/9
III	4a	6/9	VI	4a	4/9

6. Determining the power of the test

In this paper an algorithm generating two-way CTs using the bar method is applied. The bar method is similar to generating random numbers that follow the multinomial distribution. Details of the bar method applied to two-way and three-way CTs may be found in Sulewski and Motyka (2015) and Sulewski (2018) respectively.

Different scenarios determine different intervals of achievable MoU values. Statistics (2), (3) and (4) can be calculated when $n_{ij}^* \neq 0$ for each $i = 1, \dots w$; $j = 1, \dots, k$. The PoT is not calculated for the maximum MoU value under a given scenario (see Table 4) because in this case $n_{ij}^* = 0$ for any $i = 1, \dots w$; $j = 1, \dots, k$. We need the minimal sample size n for each particular scenario to guarantee $e_{ij}^* \neq 0$ for each $i = 1, \dots w$; $j = 1, \dots, k$. Example 3, based on real data, shows how to use statistic (4) when the CT has zero cells.

An algorithm calculating the critical value of test and the PoT is presented in Sulewski (2017) and Sulewski (2018).

Tables 5–7 show sizes and powers of tests for 2×3 CTs, scenarios I–III and sample size n. Tables 8–10 show sizes and powers of tests for 3×3 CTs, scenarios IV–VI and sample size n.

Table 5. Sizes and powers of tests for $\mathbf{2} \times \mathbf{3}$ CTs and scenario I

Test	MoU									
	0	0.022	0.044	0.067	0.089	0.111	0.133	0.156	0.178	0.200
$\alpha = 0.05, n = 50$										
χ^2	0.050	0.051	0.055	0.062	0.081	0.105	0.135	0.174	0.221	0.267
G^2	0.050	0.052	0.057	0.067	0.081	0.102	0.135	0.175	0.245	0.279
N	0.050	0.051	0.059	0.070	0.088	0.114	0.160	0.210	0.305	0.376
$\lvert\chi\rvert$	0.050	0.051	0.056	0.067	0.091	0.123	0.170	0.228	0.306	0.391
LMS	0.050	0.051	0.057	0.069	0.096	0.134	0.190	0.266	0.366	0.488
T_3	0.051	0.054	0.056	0.070	0.073	0.091	0.115	0.135	0.166	0.178
BP	0.055	0.060	0.080	0.142	0.230	0.437	0.674	0.882	0.986	1.000
$\alpha = 0.05, n = 100$										
χ^2	0.050	0.056	0.069	0.096	0.136	0.200	0.294	0.422	0.586	0.766
G^2	0.050	0.559	0.058	0.093	0.116	0.187	0.286	0.415	0.597	0.793
N	0.050	0.055	0.058	0.093	0.116	0.195	0.306	0.452	0.649	0.863
$\lvert\chi\rvert$	0.050	0.056	0.070	0.101	0.148	0.228	0.350	0.510	0.705	0.892
LMS	0.050	0.056	0.071	0.102	0.152	0.237	0.367	0.540	0.742	0.924
T_3	0.050	0.056	0.058	0.087	0.103	0.151	0.210	0.287	0.410	0.513
BP	0.051	0.067	0.119	0.262	0.437	0.751	0.938	0.995	1.000	1.000
$\alpha = 0.1, n = 50$										
χ^2	0.100	0.100	0.107	0.121	0.144	0.180	0.232	0.287	0.340	0.414
G^2	0.100	0.114	0.122	0.136	0.153	0.209	0.238	0.288	0.371	0.428
N	0.100	0.115	0.123	0.137	0.154	0.215	0.265	0.323	0.431	0.520
$\lvert\chi\rvert$	0.100	0.101	0.109	0.128	0.158	0.206	0.271	0.352	0.439	0.551
LMS	0.100	0.101	0.109	0.129	0.162	0.215	0.290	0.385	0.495	0.638
T_3	0.100	0.111	0.121	0.129	0.141	0.187	0.197	0.231	0.275	0.289
BP	0.101	0.106	0.136	0.216	0.324	0.542	0.760	0.920	0.992	1.000
$\alpha = 0.1, n = 100$										
χ^2	0.100	0.108	0.128	0.162	0.223	0.297	0.406	0.549	0.711	0.860
G^2	0.100	0.108	0.133	0.146	0.228	0.292	0.398	0.541	0.731	0.871
N	0.100	0.108	0.133	0.150	0.235	0.300	0.413	0.573	0.774	0.909
$\lvert\chi\rvert$	0.100	0.108	0.131	0.170	0.239	0.330	0.462	0.625	0.801	0.935
LMS	0.100	0.108	0.131	0.171	0.242	0.336	0.473	0.644	0.823	0.952
T_3	0.100	0.106	0.130	0.142	0.215	0.255	0.335	0.435	0.558	0.668
BP	0.102	0.127	0.168	0.253	0.361	0.583	0.789	0.941	0.995	1.000

Table 6. Sizes and powers of tests for $\mathbf{2 \times 3}$ CTs and scenario II

Test	MoU									
	0.000	0.044	0.089	0.133	0.178	0.222	0.267	0.311	0.356	0.400
$\alpha = 0.05, n = 50$										
χ^2	0.050	0.063	0.082	0.131	0.204	0.312	0.444	0.591	0.725	0.841
G^2	0.050	0.055	0.090	0.135	0.216	0.305	0.426	0.566	0.701	0.804
N	0.050	0.054	0.090	0.134	0.216	0.300	0.424	0.556	0.702	0.812
$\lvert\chi\rvert$	0.050	0.064	0.085	0.142	0.228	0.360	0.524	0.699	0.845	0.946
LMS	0.050	0.064	0.085	0.141	0.227	0.357	0.520	0.692	0.839	0.942
T_3	0.050	0.057	0.082	0.120	0.194	0.253	0.344	0.456	0.542	0.630
BP	0.055	0.070	0.128	0.269	0.452	0.713	0.904	0.983	0.999	1.000
$\alpha = 0.05, n = 75$										
χ^2	0.050	0.062	0.101	0.178	0.301	0.463	0.647	0.811	0.924	0.979
G^2	0.050	0.066	0.098	0.177	0.296	0.464	0.648	0.809	0.921	0.974
N	0.050	0.069	0.101	0.175	0.299	0.460	0.643	0.809	0.917	0.977
$\lvert\chi\rvert$	0.050	0.063	0.104	0.190	0.332	0.517	0.721	0.882	0.969	0.997
LMS	0.050	0.063	0.104	0.188	0.327	0.509	0.712	0.877	0.967	0.997
T_3	0.050	0.065	0.097	0.171	0.272	0.409	0.561	0.699	0.813	0.889
BP	0.053	0.079	0.169	0.366	0.626	0.885	0.985	0.999	1.000	1.000
$\alpha = 0.1, n = 50$										
χ^2	0.100	0.110	0.147	0.218	0.313	0.436	0.571	0.702	0.820	0.912
G^2	0.100	0.107	0.138	0.216	0.306	0.433	0.559	0.682	0.789	0.896
N	0.100	0.110	0.130	0.216	0.310	0.431	0.558	0.691	0.795	0.895
$\lvert\chi\rvert$	0.100	0.110	0.151	0.232	0.339	0.483	0.639	0.785	0.900	0.972
LMS	0.100	0.110	0.150	0.231	0.338	0.482	0.634	0.780	0.899	0.972
T_3	0.100	0.105	0.128	0.204	0.288	0.378	0.484	0.584	0.666	0.753
BP	0.101	0.123	0.199	0.361	0.551	0.788	0.940	0.991	1.000	1.000
$\alpha = 0.1, n = 75$										
χ^2	0.100	0.121	0.179	0.281	0.423	0.592	0.756	0.885	0.959	0.991
G^2	0.100	0.117	0.183	0.279	0.420	0.584	0.750	0.883	0.958	0.988
N	0.100	0.118	0.181	0.281	0.418	0.580	0.748	0.878	0.955	0.989
$\lvert\chi\rvert$	0.100	0.122	0.182	0.292	0.450	0.637	0.809	0.927	0.983	0.999
LMS	0.100	0.121	0.181	0.292	0.447	0.632	0.804	0.925	0.982	0.998
T_3	0.100	0.118	0.178	0.263	0.392	0.528	0.679	0.805	0.891	0.939
BP	0.101	0.138	0.253	0.478	0.724	0.927	0.993	1.000	1.000	1.000

Table 7. Sizes and powers of tests for $\mathbf{2 \times 3}$ CTs and scenario III

Test	MoU									
	0.000	0.067	0.133	0.200	0.267	0.333	0.400	0.467	0.533	0.600
$\alpha = 0.05, n = 30$										
χ^2	0.050	0.066	0.116	0.202	0.311	0.467	0.647	0.802	0.930	0.984
G^2	0.050	0.066	0.116	0.198	0.306	0.466	0.642	0.799	0.924	0.983
N	0.050	0.065	0.111	0.191	0.291	0.447	0.624	0.792	0.921	0.984
$\lvert\chi\rvert$	0.050	0.066	0.113	0.194	0.303	0.445	0.633	0.793	0.922	0.985
LMS	0.050	0.067	0.114	0.189	0.298	0.441	0.632	0.790	0.916	0.984
T_3	0.050	0.067	0.115	0.199	0.304	0.454	0.627	0.781	0.888	0.962
BP	0.055	0.059	0.092	0.146	0.234	0.411	0.616	0.828	0.955	0.998
$\alpha = 0.05, n = 50$										
χ^2	0.050	0.074	0.168	0.316	0.522	0.749	0.902	0.980	0.998	1.000
G^2	0.050	0.074	0.167	0.313	0.522	0.745	0.899	0.979	0.999	1.000
N	0.050	0.070	0.161	0.306	0.509	0.733	0.891	0.978	0.999	1.000
$\lvert\chi\rvert$	0.050	0.073	0.164	0.299	0.515	0.737	0.891	0.976	0.999	1.000
LMS	0.050	0.070	0.158	0.298	0.504	0.726	0.888	0.976	0.999	1.000
T_3	0.050	0.073	0.168	0.316	0.517	0.740	0.892	0.973	0.996	1.000
BP	0.055	0.068	0.133	0.247	0.423	0.679	0.879	0.980	0.998	1.000
$\alpha = 0.1, n = 30$										
χ^2	0.100	0.125	0.189	0.298	0.443	0.612	0.777	0.890	0.963	0.996
G^2	0.100	0.123	0.188	0.295	0.438	0.608	0.774	0.889	0.961	0.995
N	0.100	0.122	0.186	0.286	0.423	0.587	0.759	0.885	0.959	0.996
$\lvert\chi\rvert$	0.100	0.123	0.188	0.287	0.431	0.599	0.768	0.883	0.959	0.993
LMS	0.100	0.118	0.187	0.279	0.424	0.586	0.757	0.881	0.957	0.993
T_3	0.100	0.126	0.190	0.295	0.438	0.597	0.758	0.871	0.943	0.985
BP	0.108	0.115	0.157	0.238	0.347	0.538	0.720	0.889	0.971	0.999
$\alpha = 0.1, n = 50$										
χ^2	0.100	0.133	0.253	0.436	0.662	0.841	0.949	0.991	0.999	1.000
G^2	0.100	0.132	0.253	0.433	0.660	0.839	0.949	0.991	0.999	1.000
N	0.100	0.131	0.249	0.429	0.654	0.836	0.947	0.992	0.999	1.000
$\lvert\chi\rvert$	0.100	0.131	0.245	0.430	0.651	0.833	0.945	0.991	0.999	1.000
LMS	0.100	0.130	0.245	0.424	0.649	0.830	0.943	0.990	0.999	1.000
T_3	0.100	0.133	0.258	0.440	0.657	0.833	0.942	0.989	0.998	1.000
BP	0.101	0.111	0.200	0.341	0.533	0.767	0.922	0.988	0.999	1.000

Table 8. Sizes and powers of tests for $\mathbf{3 \times 3}$ CTs and scenario IV

Test	MoU									
	0.000	0.001	0.004	0.010	0.018	0.028	0.040	0.054	0.071	0.090
$\alpha = 0.05, n = 75$										
χ^2	0.050	0.052	0.048	0.053	0.051	0.052	0.055	0.058	0.062	0.069
G^2	0.050	0.045	0.044	0.048	0.047	0.049	0.054	0.053	0.061	0.068
N	0.050	0.045	0.045	0.054	0.054	0.060	0.068	0.072	0.080	0.090
$\|\chi\|$	0.050	0.052	0.053	0.059	0.064	0.072	0.086	0.099	0.130	0.159
LMS	0.050	0.051	0.054	0.059	0.067	0.080	0.098	0.123	0.157	0.202
T_3	0.050	0.051	0.049	0.047	0.046	0.046	0.048	0.051	0.059	0.068
BP	0.052	0.060	0.088	0.157	0.278	0.479	0.736	0.926	0.995	1.000
$\alpha = 0.05, n = 100$										
χ^2	0.050	0.050	0.050	0.051	0.058	0.057	0.062	0.073	0.100	0.107
G^2	0.050	0.048	0.051	0.055	0.052	0.060	0.058	0.079	0.088	0.120
N	0.050	0.049	0.055	0.058	0.064	0.074	0.085	0.128	0.152	0.223
$\|\chi\|$	0.050	0.051	0.054	0.056	0.068	0.079	0.100	0.135	0.186	0.258
LMS	0.050	0.051	0.054	0.060	0.076	0.087	0.124	0.179	0.271	0.391
T_3	0.050	0.047	0.049	0.053	0.049	0.053	0.051	0.064	0.075	0.085
BP	0.052	0.060	0.103	0.200	0.365	0.620	0.861	0.980	0.999	1.000
$\alpha = 0.1, n = 75$										
χ^2	0.100	0.095	0.095	0.096	0.098	0.105	0.102	0.115	0.122	0.137
G^2	0.100	0.100	0.102	0.108	0.095	0.106	0.111	0.115	0.136	0.140
N	0.100	0.104	0.106	0.109	0.109	0.119	0.136	0.147	0.172	0.187
$\|\chi\|$	0.100	0.095	0.099	0.109	0.116	0.132	0.149	0.182	0.216	0.264
LMS	0.100	0.096	0.100	0.111	0.124	0.146	0.170	0.211	0.261	0.324
T_3	0.100	0.105	0.103	0.101	0.093	0.090	0.104	0.104	0.119	0.123
BP	0.106	0.118	0.166	0.259	0.408	0.611	0.826	0.960	0.997	1.000
$\alpha = 0.1, n = 100$										
χ^2	0.100	0.098	0.098	0.104	0.106	0.112	0.120	0.137	0.172	0.202
G^2	0.100	0.100	0.097	0.108	0.106	0.116	0.117	0.154	0.186	0.236
N	0.100	0.099	0.095	0.114	0.119	0.128	0.147	0.192	0.264	0.355
$\|\chi\|$	0.100	0.102	0.102	0.110	0.127	0.149	0.169	0.228	0.290	0.396
LMS	0.100	0.104	0.104	0.114	0.137	0.163	0.194	0.282	0.380	0.533
T_3	0.100	0.099	0.096	0.098	0.092	0.107	0.098	0.122	0.143	0.169
BP	0.103	0.118	0.186	0.321	0.498	0.742	0.919	0.990	1.000	1.000

Table 9. Sizes and powers of tests for $\mathbf{3} \times \mathbf{3}$ CTs and scenario V

Test	MoU									
	0.000	0.022	0.044	0.067	0.089	0.111	0.133	0.156	0.178	0.200
	$\alpha = 0.05, n = 75$									
χ^2	0.050	0.052	0.068	0.091	0.124	0.173	0.241	0.327	0.427	0.533
G^2	0.050	0.055	0.070	0.089	0.119	0.171	0.245	0.330	0.413	0.539
N	0.050	0.057	0.073	0.084	0.125	0.175	0.260	0.363	0.477	0.639
$\|\chi\|$	0.050	0.051	0.068	0.092	0.129	0.183	0.264	0.366	0.485	0.613
LMS	0.050	0.051	0.068	0.093	0.132	0.191	0.279	0.398	0.539	0.694
T_3	0.050	0.052	0.070	0.088	0.115	0.156	0.224	0.279	0.334	0.429
BP	0.052	0.056	0.082	0.137	0.237	0.418	0.663	0.888	0.989	1.000
	$\alpha = 0.05, n = 100$									
χ^2	0.050	0.056	0.075	0.103	0.160	0.234	0.346	0.477	0.615	0.751
G^2	0.050	0.048	0.073	0.109	0.163	0.253	0.332	0.488	0.653	0.783
N	0.050	0.051	0.074	0.109	0.160	0.253	0.353	0.527	0.703	0.844
$\|\chi\|$	0.050	0.056	0.077	0.106	0.169	0.250	0.376	0.521	0.680	0.825
LMS	0.050	0.056	0.075	0.107	0.171	0.258	0.394	0.555	0.731	0.886
T_3	0.050	0.050	0.079	0.102	0.154	0.226	0.299	0.414	0.525	0.611
BP	0.052	0.063	0.095	0.165	0.317	0.554	0.799	0.961	0.999	1.000
	$\alpha = 0.1, n = 75$									
χ^2	0.100	0.108	0.124	0.161	0.208	0.276	0.364	0.461	0.566	0.675
G^2	0.100	0.111	0.122	0.148	0.209	0.268	0.358	0.472	0.570	0.697
N	0.100	0.111	0.122	0.149	0.206	0.272	0.368	0.505	0.614	0.772
$\|\chi\|$	0.100	0.107	0.125	0.164	0.216	0.294	0.393	0.506	0.630	0.755
LMS	0.100	0.108	0.125	0.165	0.219	0.301	0.408	0.534	0.674	0.815
T_3	0.100	0.112	0.121	0.145	0.198	0.254	0.319	0.414	0.474	0.567
BP	0.103	0.112	0.146	0.231	0.358	0.549	0.769	0.931	0.994	1.000
	$\alpha = 0.1, n = 100$									
χ^2	0.100	0.109	0.141	0.186	0.250	0.349	0.471	0.605	0.740	0.853
G^2	0.100	0.111	0.141	0.191	0.245	0.368	0.460	0.625	0.737	0.846
N	0.100	0.107	0.133	0.195	0.256	0.378	0.478	0.646	0.774	0.899
$\|\chi\|$	0.100	0.108	0.143	0.191	0.259	0.366	0.501	0.651	0.792	0.906
LMS	0.100	0.108	0.144	0.193	0.264	0.376	0.517	0.675	0.825	0.939
T_3	0.100	0.103	0.138	0.176	0.237	0.333	0.421	0.529	0.637	0.706
BP	0.103	0.119	0.173	0.274	0.446	0.677	0.870	0.980	1.000	1.000

Table 10. Sizes and powers of tests for $\mathbf{3 \times 3}$ CTs and scenario VI

Test	MoU									
	0.000	0.044	0.089	0.133	0.178	0.222	0.267	0.311	0.356	0.400
	$\alpha = 0.05, n = 40$									
χ^2	0.050	0.061	0.089	0.140	0.193	0.310	0.429	0.582	0.740	0.836
G^2	0.050	0.068	0.084	0.137	0.200	0.308	0.447	0.604	0.753	0.867
N	0.050	0.075	0.079	0.128	0.199	0.307	0.433	0.608	0.774	0.892
$\|\chi\|$	0.050	0.065	0.082	0.132	0.183	0.295	0.382	0.525	0.664	0.750
LMS	0.050	0.071	0.083	0.136	0.195	0.302	0.420	0.573	0.726	0.851
T_3	0.050	0.062	0.087	0.144	0.200	0.313	0.421	0.562	0.711	0.798
BP	0.053	0.071	0.099	0.147	0.251	0.425	0.675	0.899	0.991	1.000
	$\alpha = 0.05, n = 75$									
χ^2	0.050	0.070	0.062	0.237	0.411	0.632	0.821	0.941	0.989	1.000
G^2	0.050	0.072	0.063	0.240	0.414	0.634	0.826	0.943	0.991	1.000
N	0.050	0.069	0.062	0.226	0.393	0.618	0.813	0.941	0.991	1.000
$\|\chi\|$	0.050	0.069	0.060	0.227	0.383	0.592	0.781	0.917	0.983	0.999
LMS	0.050	0.070	0.063	0.227	0.388	0.602	0.795	0.931	0.987	0.999
T_3	0.050	0.069	0.064	0.239	0.411	0.622	0.805	0.926	0.981	0.996
BP	0.052	0.073	0.100	0.195	0.391	0.661	0.904	0.991	1.000	1.000
	$\alpha = 0.1, n = 40$									
χ^2	0.100	0.113	0.113	0.218	0.294	0.435	0.588	0.735	0.851	0.937
G^2	0.100	0.114	0.111	0.216	0.299	0.441	0.592	0.746	0.858	0.944
N	0.100	0.116	0.109	0.210	0.294	0.436	0.580	0.745	0.858	0.953
$\|\chi\|$	0.100	0.113	0.119	0.208	0.278	0.409	0.540	0.692	0.809	0.906
LMS	0.100	0.116	0.115	0.211	0.287	0.425	0.567	0.728	0.841	0.940
T_3	0.100	0.113	0.110	0.231	0.306	0.430	0.569	0.715	0.828	0.890
BP	0.100	0.125	0.168	0.233	0.363	0.544	0.772	0.938	0.995	1.000
	$\alpha = 0.1, n = 75$									
χ^2	0.100	0.119	0.202	0.349	0.548	0.744	0.895	0.972	0.997	1.000
G^2	0.100	0.121	0.203	0.349	0.552	0.746	0.899	0.974	0.997	1.000
N	0.100	0.118	0.200	0.340	0.541	0.736	0.897	0.974	0.998	1.000
$\|\chi\|$	0.100	0.117	0.196	0.333	0.522	0.711	0.869	0.962	0.995	1.000
LMS	0.100	0.118	0.197	0.332	0.526	0.717	0.881	0.968	0.996	1.000
T_3	0.100	0.121	0.200	0.350	0.544	0.730	0.884	0.964	0.994	1.000
BP	0.103	0.126	0.180	0.311	0.526	0.775	0.947	0.996	1.000	1.000

Tables 5–10 show that the sizes of tests, except BP, are identical to the nominal level (to three decimal places). The BP test is also distinguished in terms of the power. BP is surprisingly the most powerful test under scenarios I, II, IV and V, characterized by a low maximum MoU value and strong dependence

within a given scenario. For example, under scenario I, $\alpha = 0.05$, $n = 50$ and $MoU = 0.2$, the power of the BP test equals 1, the power of the χ^2 test is only 0.267, and the power of the LMS test equals 0.488. Not including the BP test, the LMS is the most powerful test under scenarios I, II, IV and V, characterized by a low maximum MoU value. The situation changes under scenarios III and VI, characterized by a high maximum MoU value. The χ^2 test is the most powerful test under scenario III and $MoU \leq 0.467$. The BP is the most powerful test under scenario VI and $n = 40$ as well as under scenario VI, $n = 75$ and $MoU \geq 0.222$.

7. Examples

Example 1. We carried out a test of independence with regard to features X and Y. The null hypothesis H_0 states that X, Y are independent. An alternative hypothesis, denoted H_1, negates H_0. The empirical probabilities p_{ij}^* are presented in Tables 11 and 13. In these cases H_0 does not hold and the sample MoU, i.e. the measure of untruthfulness of H_0, is equal to 0.248 (2×3) and 0.091 (3×3). Critical values for $Q_j (j = 1, \dots, 7)$ tests were determined with the Monte Carlo method for $\alpha = 0.05, 0.1, n = 100$ $(2 \times 3), n = 300$ (3×3). These critical values and values of test statistics are presented in Tables 12 and 14.

Table 11. CT $\mathbf{2 \times 3}$ with empirical probabilities $\boldsymbol{p_{ij}^*} (\boldsymbol{i = 1, 2, j = 1, 2, 3})$

X/Y	Y_1	Y_2	Y_3	Total
X_1	0.240	0.120	0.160	0.520
X_2	0.110	0.230	0.140	0.480
Total	0.350	0.350	0.300	1.000

Table 12. Critical values and values of test statistics for α=0.05, 0.1, $\boldsymbol{MoU_e = 0.248}$

Test		Related critical value		Related value
Name	Symbol	0.05	0.1	of test statistic
χ^2	Q_1	6.028	4.644	8.272
G^2	Q_2	6.195	4.752	8.436
N	Q_3	6.790	5.080	9.356
$\lvert\chi\rvert$	Q_4	1.351	1.183	1.427
LMS	Q_5	1.393	1.212	1.484
T_3	Q_6	16.627	12.747	24.080
BP	Q_7	11	9.2	9.254

All tests in question reject H_0 for $MoU_e = 0.248$ and at the significance level$s$ $\alpha = 0.05, 0.1$ except BP at $\alpha = 0.05$.

Table 13. CT $\mathbf{3 \times 3}$ with empirical probabilities $\boldsymbol{p^*_{ij}}(\boldsymbol{i, j = 1, 2, 3})$

X/Y	Y_1	Y_2	Y_3	Total
X_1	0.120	0.127	0.190	0.437
X_2	0.057	0.020	0.083	0.160
X_3	0.117	0.120	0.166	0.403
Total	0.294	0.267	0.439	1.000

Table 14. Critical values and values of test statistics for $\boldsymbol{\alpha = 0.05, 0.1, MoU_e = 0.091}$

Test		Related critical value		Related value
Name	Symbol	0.05	0.1	of test statistic
χ^2	Q_1	9.507	7.803	5.996
G^2	Q_2	9.634	7.907	6.798
N	Q_3	10.187	8.256	9.833
$\lvert\chi\rvert$	Q_4	1.376	1.245	1.275
LMS	Q_5	1.399	1.262	1.462
T_3	Q_6	35.166	28.897	12.245
BP	Q_7	15.48	13.38	159.137

LMS and BP are the preferred tests at $\alpha = 0.05$ because they reject H_0, whereas the other tests uphold H_0. LMS, BP and |χ| are the preferred tests at $\alpha = 0.1$.

Example 2. This example is described by means of the following algorithm:

Step 1. Set a sample size n, number of rows w, number of columns k, significance level α.

Step 2. Set theoretical probabilities $p_{ij}(i = 1, \dots, w;\ j = 1, \dots, k)$.

Step 3. Calculate the MoU for the $p_{ij}(i = 1, \dots, w;\ j = 1, \dots, k)$ set in Step 2.

Step 4. Set critical values $cv_j(j = 1, \dots, 7)$ for $Q_j(j = 1, \dots, 7)$ tests equal to those given in Tables 12, 14 and 18.

Step 5. Set initial values of a counter of rejected hypotheses $CF_j = 0\ (j = 1, \dots, 7)$.

Step 6. Repeat the following steps m times:

Step 6.1. Generate a $w \times k$ CT according to the bar method.

Step 6.2. Calculate values of test statistics $Q_j (j = 1, \dots, 7)$.
Step 6.3. If $Q_j > cv_j (j = 1, \dots, 7)$ then $CF_j = CF_j + 1$.
Step 7. Calculate rejection probabilities $FR_j^* = CF_j / m \ (j = 1, \dots, 7)$ (Tables 15, 19)

Specific values were set in this example: for a 2×3 CT ($n = 100, MoU = 0.248$), for a 3×3 CT ($n = 300, MoU = 0.091$).

Table 15. Rejection probabilities estimated from $\boldsymbol{m = 10^6}$ repetitions

Test		Rejection probabilities FR^*			
		2×3 CT		3×3 CT	
Name	Symbol	$MoU = 0.248$		$MoU = 0.091$	
		$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
χ^2	Q_1	0.741	0.833	0.484	0.631
G^2	Q_2	0.736	0.830	0.537	0.667
N	Q_3	0.730	0.827	0.647	0.742
$\lvert\chi\rvert$	Q_4	0.723	0.822	0.683	0.789
LMS	Q_5	0.722	0.821	0.753	0.832
T_3	Q_6	0.748	0.837	0.207	0.342
BP	Q_7	0.648	0.760	0.996	0.998

Since the number of repetitions was equal to one million, there is no need to carry out a formal test for equality of two proportions. The T_3, χ^2 and G^2 tests take the highest values of FR_j^* under 2×3 CT, when $MoU = 0.248$. The BP and LMS tests win against other tests under 3×3 CT, when $MoU = 0.091$.

Example 3 (real data). We consider the dataset given by Koch and Edwards (1988) to study the performance of the proposed method with real data (Table 16). This table compares a treatment for rheumatoid arthritis with a placebo. The outcome reflects whether individuals show no improvement, some improvement, or marked improvement. We cannot select the G^2, N and LMS statistics here because of the zero cell in Table 17. For the analyzed real data we can use a modified version of LMS, which is very similar to the original version (4). It is defined as

$$LMS_m = -\sum_{i=1}^{w} \sum_{j=1}^{k} ln \left[\frac{min(n_{ij}, e_{ij})}{max(n_{ij}, e_{ij})} + 0.00001 \right]$$

Table 16 shows that adding 0.00001 to the logarithmic value does not affect the size and power of the LMS_m tests.

Table 16. The power of the $\boldsymbol{LMS}$ and $\boldsymbol{LMS}_{\boldsymbol{m}}$ tests, $\boldsymbol{\alpha}$ = 0.05, scenario I

	MoU	0.000	0.022	0.044	0.067	0.089	0.111	0.133	0.156	0.178	0.200
$n = 50$	LMS	0.050	0.054	0.063	0.074	0.096	0.130	0.188	0.274	0.375	0.487
	LMS_m	0.050	0.054	0.063	0.074	0.096	0.130	0.188	0.274	0.375	0.487
$n = 75$	LMS	0.050	0.064	0.073	0.102	0.137	0.208	0.305	0.447	0.635	0.814
	LMS_m	0.050	0.064	0.073	0.102	0.137	0.208	0.305	0.447	0.635	0.814

Values of test statistics and critical values are presented in Table 18. Table 19 presents values of rejection probabilities FR_j^* ($j = 1, \dots, 10^6$) as defined in Example 2.

Table 17. The effect of a treatment for rheumatoid arthritis vs. a placebo

$Treatment/Outcome$	None	Some	Marked	Total
Active	7	2	5	14
Placebo	10	0	1	11
Total	17	2	6	25

Table 18. Values of test statistics and critical values for $\boldsymbol{\alpha = 0.05, 0.1}$ and $\boldsymbol{MoU = 0.4032}$

Test Name	Related critical value $\alpha = 0.05$	$\alpha = 0.1$	Related value of test statistics
χ^2	5.252	4.233	4.907
$\lvert\chi\rvert$	2.633	2.338	3.497
LMS_m	2.949	2.580	14.059
T_3	3.665	2.826	3.271
BP	11.24	9.320	23.571

LMS_m, BP and $|\chi|$ are the preferred tests at $\alpha = 0.05$ because they reject H_0, whereas the χ^2 and T_3 tests uphold H_0. All of the tests reject H_0 at $\alpha = 0.1$.

The LMS_m test wins against other tests under 2×3 CT with real data for $\alpha = 0.05, 0.1$ and $MoU = 0.4032$.

Table 19. Rejection probabilities for real data estimated from $\mathbf{10^6}$ repetitions

Test Name	Rejection probabilities FR^* $\alpha = 0.05$	$\alpha = 0.1$
χ^2	0.516	0.642
$\|\chi\|$	0.866	0.930
LMS_m	1.000	1.000
T_3	0.430	0.565
BP	0.96	0.98

8. Conclusion

The Monte Carlo simulations show that the LMS test is (excluding the BP test) the most powerful in the sense of the proposed measure MoU under scenarios characterized by a smaller variability range of MoU and for strong dependence within a given scenario (see scenarios I, II, IV and V). The advantage of the BP test over the other analyzed tests is surprising for the scenarios in question. For example, under scenario V, $\alpha = 0.05$, $n = 75$ and $MoU = 0.178$, the power of the BP test equals 0.989 and the power of the LMS test equals 0.539. The dominance of the BP test is not so clear for scenarios III and VI, characterized by a wider range of variability of MoU. Similar results were obtained for 2×4, 3×4 and 4×4 CTs; however, they are not presented in this paper due to its limited size.

Example 1 is characterized by stronger dependency between features than Example 2. Example 1 shows that all of the tests reject H_0, except the BP test. The rejection probability estimated from $m = 1\,000\,000$ repetitions takes the highest values for the T_3, χ^2 and G^2 tests. Example 2 shows that LMS and BP are the preferred tests. They reject H_0, whereas the other tests uphold this hypothesis. The rejection probability takes the highest values for the BP and LMS tests.

Example 3 presents a strong dependency between features. The LMS_m, BP and $|\chi|$ tests reject H_0, whereas the χ^2 and T_3 tests uphold this hypothesis. The rejection probability takes the highest values for the LMS and BP tests.

The paper shows that LMS and BP are more effective than the other tests considered, as they detect dependency even for low MoU values. However, it should be emphasized that the LMS test is simpler than the BP test.

Acknowledgments

We gratefully acknowledge the constructive comments and suggestions of the anonymous referees and the associate editor.

REFERENCES

Albert, J.H. (1990): A Bayesian test for a two-way contingency table using independence. Prior, Canadian Journal of Statistics 18(4): 347–363.

Amiri, S., Modarres, R. (2017): Comparison of tests of contingency tables, Journal of Biopharmaceutical Statistics: 1–13.

Amiri, S., von Rosen, D. (2011): On the efficiency of bootstrap method into the analysis contingency table. Computer Methods and Programs in Biomedicine 104(2): 182–187.

Andrés, A.M., Tejedor, I.H., Mato, A.S. (1995): The Wilcoxon, Spearman, Fisher, χ^2, Student and Pearson Tests and 2x2 Tables, The Statistician: 441–450.

Berry, K.J., Mielke, P.W. (1988): Monte Carlo comparisons of the asymptotic chi-square and likelihood-ratio tests with the no asymptotic chi-square tests for sparse r×c tables, Psychological Bulletin 103(2): 256.

Campbell, I. (2007): Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations, Statistics in Medicine 26(19): 3661–3675.

Cochran, W.G. (1952): The χ^2 test of goodness of fit, The Annals of Mathematical Statistics 23(3): 315–345.

Cochran, W.G. (1954): Some methods for strengthening the common χ^2 tests, Biometrics 10(4): 417–451.

Cohen, J., Nee, J.C. (1990): Robustness of Type I Error and Power in Set Correlation Analysis of Contingency Tables, Multivariate Behav. Res. 25(3): 341–350.

Cressie, N., Read, T. (1984): Multinomial Goodness-of-Fit Tests, J R Stat Soc Ser B Stat Methodol 46: 440–464.

Cressie, N., Read, T.R. (1989): Pearson's χ^2 and the log likelihood ratio statistics G2: a comparative review, International Statistical Review/Revue Internationale de Statistique: 19–43.

Davis, C.S. (1993): A new approximation to the distribution of Pearson's chi-square, Statistica Sinica: 189–196.

D'Agostino, R.B., Chase, W., Belanger, A. (1988): The Appropriateness of Some Common Procedures for Testing Equality of Two Independent Binomial Proportions. The American Statistician 42(3): 198–202.

Diaconis, P., Efron, B. (1985): Testing for independence in a two-way table: new interpretations of the chi-square statistics, The Annals of Statistics 13(3): 845–874.

Fisher, R.A. (1922): On the interpretation of χ^2 from contingency tables, and the calculation of P, Journal of the Royal Statistical Society 85(1): 87–94.
Garside, G.R., Mack, C. (1976): Actual type 1 error probabilities for various tests in the homogeneity case of the 2×2 contingency table, The American Statistician 30(1): 18–21.
Haberman, S.J. (1981): Tests for independence in two-way contingency tables based on canonical correlation and on linear-by-linear interaction, The Annals of Statistics 9(6): 1178–1186.
Irwin, J.O. (1935): Tests of significance for differences between percentages based on small numbers, Metron 12(2): 84–94.
Koch, G., Edwards, S. (1988): Clinical efficiency trials with categorical data. In K.E. Peace (Ed.), Biopharmaceutical Statistics for Drug Development. New York, NY: Marcel Dekker: 403–451.
Koehler, K.J., Larntz, K. (1980): An empirical investigation of goodness-of-fit statistics for sparse multinomials, Journal of the American Statistical Association 75(370): 336–344.
Lawal, H.B., Uptong, G.J.G. (1984): On the use of χ^2 as a test of independence in contingency tables with small cell expectations, Australian Journal of Statistics 26: 75–85.
Lawal, H.B., Uptong, G.J.G. (1990): Comparisons of Some Chi-squared Tests for the Test of Independence in Sparse Two-Way Contingency Tables, Biometrical Journal 32(1): 59–72.
Liddell, D. (1976): Practical tests of 2x2 contingency tables. The Statistician 25(4): 295–304.
Lin, J.J., Chang, C.H., Pal, N. (2015): A revisit to contingency table and tests of independence: bootstrap is preferred to Chi-Square approximations as well as Fisher’s exact test, Journal of Biopharmaceutical Statistics 25(3): 438–458.
Lipsitz, S.R., Fitzmaurice, G.M., Sinha, D., Hevelone, N., Giovannucci, E., Hu, J.C. (2015): Testing for independence in J×K contingency tables with complex sample survey data, Biometrics 71(3): 832–840.
Lydersen, S., Fagerland, M.W., Laake, P. (2009): Recommended tests for association in 2×2 tables, Statistics in Medicine 28(7): 1159–1175.
Marcotorchino F. (1984): Utilisation des Comparaisons par Paires en Statistique des Contingences, Partie (1), Etude IBM F069, France.
Meng, R.C., Chapman, D.G. (1966): The power of chi square tests for contingency tables, Journal of the American Statistical Association 61(316): 965–975.
Pearson, K. (1904): On the theory of contingency and its relation to association and normal correlation, K. Pearson, Early Papers.
Shan, G., Wilding, G. (2015): Unconditional tests for association in 2×2 contingency tables in the total sum fixed design, Statistica Neerlandica 69(1): 67–83.
Sulewski, P. (2013): Modyfikacja testu niezależności [Modification of the independence test], (The paper is in Polish. The abstract in English is available on scholar.google.pl), Statistical News – Central Statistical Office 10: 1–19.
Sulewski, P. (2014): Statystyczne badanie współzależności cech typu dyskretne kategorie, (in Polish), Akademia Pomorska, Słupsk.

Sulewski, P., Motyka, R. (2015): Power analysis of independence testing for contingency tables, Scientific Journal of Polish Naval Academy 56(1): 37-46.
Sulewski, P. (2016): Moc testów niezależności w tablicy dwudzielczej większej niż 2x2 [Power of independence test in the 2×2 contingency tables bigger than 2x2], (The paper is in Polish. The abstract in English is available on scholar.google.pl), Statistical Review 63(2): 191–209.
Sulewski, P. (2017): A new test for independence in 2x2 contingency tables, Acta Universitatis Lodziensis. Folia Oeconomica 4(330): 55–75.
Sulewski, P. (2018): Power analysis of independence testing for the three-way contingency tables of small sizes, Journal of Applied Statistics 45(13): 2481–2498.
Vélez, J.I., Marmolejo-Ramos, F., Correa, J.C. (2016): A Graphical Diagnostic Test for Two-Way Contingency Tables, Revista Colombiana de Estadística 39(1): 97–108.
Yenigün, C.D., Székely, G.J., Rizzo, M.L. (2011): A Test of Independence in Two-Way Contingency Tables Based on Maximal Correlation, Communications in Statistics—Theory and Methods 40(12): 2225–2242.
Yu, Y. (2014): Tests of independence in a single 2×2 contingency table with random margins, Doctoral dissertation, Worcester Polytechnic Institute, Worcester.