Nonparametric versus parametric reasoning based on two-way and three-way contingency tables

1. INTRODUCTION

The analysis of contingency tables (CTs) is one of the most common tasks performed by statisticians. CTs display the frequency distribution of two (two-way CTs), three (three-way CTs) or more (multi-way CTs) categorical variables. The information about categorical data can be found e.g. in Bishop et al. (1975), Agresti (2002), Van Belle et al. (2014). Information presented as CTs features in a wide variety of areas such as the social sciences (Wickens, 1969), genetics (El Galta et al., 2008; Dickhaus et al., 2012), demography (Cung, 2013) and psychology (Iossifova et al., 2013). Basic methods of testing for dependency in CTs in details is described e.g. in Steinle et al. (2006), Bock (2003), Kaski et al. (2005), Allison, Liker (1982). Other examples of applications may be found in Ilyas et al. (2004), Oates, Cohen (1996), Schrepp (2003), Haas et al. (2007).

One can recognize two general cases in which CTs can be useful. This distinction between the cases is made with respect to the tasks which CTs are used for.

Case A. Dependency is unwanted. The general population is sought to be in its normal state or be under control when levels of feature $X$ are independent of levels of feature $Y$. Revealing dependency means revealing abnormality of members of the general population. If so, a large scale and very costly actions have to be obligatorily initiated. That is why a decision-maker tries to avoid false alarm. This case is typical, for instance, in security guarding. A classic statistical way of reasoning is tailored to case A. Please notice that the main hypothesis, commonly denoted by $H_0$, states that: $X$ and $Y$ are independent. Moreover, $H_0$ is guarded against rejection by setting significance level at 5% or less.

Case B. Dependency is wanted. The state of the general population is assessed upon feature $X$. Unfortunately, levels of feature $X$ are difficult to be determined e.g. determination is risky, costly or time consuming. In contrast, levels of another feature $Y$ are easy to be determined. Assessors are concerned with finding out whether there is a tie between $X$ and $Y$. In other words, whether $X$
and $Y$ are dependent or independent. Assessors use $Y_1, \ldots, Y_k$ levels as sensible indicators of $X_1, \ldots, X_w$ levels. Case B is typical in diagnostics, both medical and technical. In case B another way of statistical reasoning is needed, different from the classic way.

Conservativeness of the classic statistical way of reasoning often obstructs progress in numerous situations where rejecting $H_0$ means making a step ahead. This is a strong motivation for making a turnaround in statistical reasoning. In this new statistical reasoning there is no null hypothesis. In contrast to the classic way, there is a set of competing hypotheses. Moreover, the testing procedure warrants equality of all the alternatives when the test begins. The former null hypothesis is no longer the main one, but exists among the other ones of equal importance. Particular hypotheses relate to scenarios under which particular CTs are created. Details are presented in section 7. There are two reasons for which this likelihood based reasoning is developed and put forward:

a) Undoubtedly, CT-based classic statistical reasoning is the nonparametric reasoning. It is commonly known that parametric statistical reasoning, if applicable, is much more sensitive to untruthfulness of $H_0$ than nonparametric reasoning. In this paper we propose a parametric reasoning. Particular scenarios are parameterized with the probability flow parameter (PFP).

b) Let us again retrace a way of the classic thinking. A value of the test statistics is smaller than the appropriate critical value results in failing to reject $H_0$. In case A the decision maker is comfortable about independence. A value of the test statistics is no smaller than the appropriately determined critical value results in rejecting $H_0$. In case B the decision maker is comfortable about dependence because there is no word said what the reason of rejecting $H_0$ is. The most likely scenario is selected whereas reasons to reject $H_0$ or not are embedded in scenarios. One can say that the method put forward in this paper offers a transition from "unfathomable" to "fathomable" reasons.

Nonparametric and parametric reasoning based on $2 \times 2$ CT is presented in (Sulewski, 2018b), therefore this paper is devoted to bigger tables, e.g. $2 \times 3$, $2 \times 4$, $3 \times 3$, $3 \times 4$, $4 \times 4$, $2 \times 2 \times 2$, $3 \times 2 \times 2$ ones.

This paper is organized as follows. Variants of presentation of CT are described in section 2. CT coming into being are presented in section 3. Statistic tests including the power divergence tests and the $|\lambda|$ test are defined in section 4. Section 5 is devoted to measures of untruthfulness of $H_0$ including the measure that is defined by means of an absolute value. In section 6 maximum likelihood method is applied to estimate the PFP. Section 7 is devoted to instructions how to generate two-way and three-way CTs. Section 8 presents numerical examples and section 9 presents closing remarks.

Monte Carlo simulation is performed in Visual Basic for Applications embedded in Microsoft Excel 2016.
2. VARIANTS OF PRESENTATION OF CT

This section is devoted to the \( w \times k \times p \) CT. If \( p = 1 \), then we obviously have \( w \times k \) CT.

Let \( X, Y, Z \) be three features of the same object, respectively, have levels \( X_1, \ldots, X_w, Y_1, \ldots, Y_k, Z_1, \ldots, Z_p \). Testing these three features for independence with an appropriately arranged CT is probably one of the most common statisticians’ tasks. At the moment one can distinguish between four variants of presentation of CTs. It is because each variant is intended for a different purpose. Below details of particular variants are given:

— TP Variant (theoretical probabilities). Cells contain probabilities \( p_{ijt} \) intrinsic to the phenomenon being investigated (see table 1). The exact values of these probabilities are unknown to the investigator. This variant is introduced a little bit in advance since CTs will be simulated with the Monte Carlo method in further sections of this paper. And just then CT variant filled with probabilities arbitrarily set by Monte Carlo experimenter will be applied.

— TC Variant (theoretical counts). Cells contain theoretical expected counts \( n_{ijt} = np_{ijt} \). These counts are theoretical in this sense that they result from TP variant.

— EP Variant (empirical probabilities). Cells that result from EC variant and contain estimates \( p^*_{ijt} = n^*_{ijt}/n \) of the unknown content of TP.

— EC Variant (experimental counts). Cells contain \( n^*_{ijt} \) counts observed on a sample drawn from general population subjected to the investigation.

<table>
<thead>
<tr>
<th>( Z )</th>
<th>( z_1 )</th>
<th>...</th>
<th>( z_p )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X/Y )</td>
<td>( Y_1 )</td>
<td>...</td>
<td>( Y_k )</td>
<td>...</td>
</tr>
<tr>
<td>( Y_1 )</td>
<td>( p_{111} )</td>
<td>...</td>
<td>( p_{1k1} )</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>( X_w )</td>
<td>( p_{w11} )</td>
<td>...</td>
<td>( p_{wk1} )</td>
<td>...</td>
</tr>
<tr>
<td>Total</td>
<td>( p_{*11} )</td>
<td>...</td>
<td>( p_{*k1} )</td>
<td>...</td>
</tr>
</tbody>
</table>

3. ON HOW CT COMES INTO BEING

One can treat CTs as a mathematical expression of a certain phenomenon we deal with. This formulation suggests that there is an internal mechanism in this phenomenon that determines probabilities of particular \( X, Y \) or \( X, Y, Z \) combinations and ascribes these probabilities to the cells of the table. Below are “progenitors” of all the \( w \times k \) CTs.
and all the $w \times k \times p$ CTs

\[
W_p = \begin{bmatrix}
\frac{1}{wk(1)} & \cdots & \frac{1}{wk(p)} \\
\vdots & \ddots & \vdots \\
\frac{1}{wk(1)} & \cdots & \frac{1}{wk(p)} 
\end{bmatrix}
\]

A variety of tables may be generated when portions of PFP $a$ flow from "maternal" cells of (1) or (2) to other cells. Obviously, the total probability always equals 1. In this paper twenty eight scenarios that seem fundamental are developed (tables 2–3). These scenarios are created based on scenarios for the $2 \times 2$ CT that seem fundamental and describe different levels of dependence (Sulewski, 2018b).

<table>
<thead>
<tr>
<th>Table</th>
<th>Scenario</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 × 3</td>
<td>I</td>
<td>$p_{11} = 1/6 - a, p_{12} = p_{13} = p_{21} = p_{22} = 1/6, p_{23} = 1/6 + a$</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>$p_{11} = p_{12} = 1/6 - a, p_{13} = p_{21} = 1/6, p_{22} = p_{23} = 1/6 + a$</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>$p_{11} = p_{12} = 1/6 - a, p_{13} = p_{23} = 1/6, p_{21} = p_{22} = 1/6 + a$</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>$p_{11} = p_{23} = 1/6 - a, p_{12} = p_{22} = 1/6, p_{13} = p_{31} = 1/6 + a$</td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>$p_{11} = 1/8 - a, p_{12} = p_{13} = p_{24} = p_{22} = p_{23} = 1/8, p_{24} = 1/8 + a$</td>
</tr>
<tr>
<td></td>
<td>VI</td>
<td>$p_{11} = p_{12} = p_{13} = 1/8 - a, p_{14} = p_{21} = 1/8, p_{22} = p_{23} = p_{24} = 1/8 + a$</td>
</tr>
<tr>
<td></td>
<td>VII</td>
<td>$p_{11} = 1/8 - a, p_{12} = p_{13} = p_{14} = p_{22} = p_{23} = p_{24} = 1/8, p_{21} = 1/8 + a$</td>
</tr>
<tr>
<td></td>
<td>VIII</td>
<td>$p_{11} = p_{24} = 1/8 - a, p_{13} = p_{14} = p_{23} = p_{24} = 1/8, p_{22} = p_{21} = 1/8 + a$</td>
</tr>
<tr>
<td>2 × 4</td>
<td>IX</td>
<td>$p_{11} = 1/9 - a, p_{12} = p_{13} = p_{21} = p_{22} = p_{23} = p_{24} = 1/9, p_{33} = 1/9 + a$</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>$p_{11} = p_{12} = 1/9 - a, p_{13} = p_{21} = p_{22} = p_{23} = 1/9, p_{31} = p_{32} = 1/9 + a$</td>
</tr>
<tr>
<td></td>
<td>XI</td>
<td>$p_{11} = p_{23} = 1/9 - a, p_{12} = p_{21} = p_{22} = p_{23} = 1/9, p_{31} = p_{32} = 1/9 + a$</td>
</tr>
<tr>
<td></td>
<td>XII</td>
<td>$p_{11} = p_{21} = p_{23} = 1/9 - a, p_{12} = p_{22} = p_{23} = 1/9, p_{33} = p_{32} = 1/9 + a$</td>
</tr>
<tr>
<td>3 × 3</td>
<td>XIII</td>
<td>$p_{11} = 1/12 - a, p_{12} = p_{13} = p_{21} = p_{22} = p_{23} = p_{24} = p_{31} = p_{32} = p_{33} = 1/12, p_{34} = 1/12 + a$</td>
</tr>
<tr>
<td></td>
<td>XIV</td>
<td>$p_{11} = p_{12} = 1/12 - a, p_{13} = p_{21} = p_{22} = p_{23} = p_{24} = p_{31} = p_{32} = p_{33} = 1/12, p_{34} = 1/12 + a$</td>
</tr>
<tr>
<td></td>
<td>XV</td>
<td>$p_{11} = p_{12} = 1/12 - a, p_{21} = p_{22} = p_{23} = p_{24} = p_{31} = p_{32} = p_{33} = p_{34} = 1/12, p_{34} = 1/12 + a$</td>
</tr>
<tr>
<td></td>
<td>XVI</td>
<td>$p_{11} = p_{12} = 1/12 - a, p_{21} = p_{22} = p_{23} = p_{24} = 1/12, p_{13} = p_{24} = 1/12, p_{13} = p_{24} = 1/12 + a$</td>
</tr>
</tbody>
</table>
In all the above scenarios the PFP $\alpha$ takes values in $[0, \frac{1}{wp}]$ or $[0, \frac{1}{wpk}]$. The scenarios are selected in such a way that they correspond to different levels of dependence expressed by means of an appropriate measure of untruthfulness of $H_0$ (MoU). The MoU takes values on interval $[0,1]$. A simulation study is carried out for MoU values no bigger than $2/3$. It is obvious that the detection of a strong dependence is very simple. You can find more information about the MoU in section 5.

Obviously, scenarios do not cover all the cases. They may be locally mutated by reversing rows or columns to better fit the analyzed data. These are simple equal-portion scenarios. In the scenarios you can use a part of PFP, e.g. $a/2$, $a/3$,…. Surely, real scenarios can be more or less similar to these above. This is typical in relations between theory and real life. With the current availability of computers, the statistician can afford situations that interest him and instantly repeat such simulations. All examples presented here have a very precise algorithmic description in a form of a step list.
The researches can be generalized by introducing several PFPs. This, however, causes a significant deterioration in the properties of the parameter estimators. The Weibull distribution has a simple analytical form. For its generalization, the Generalized Gamma Distribution (URG) can be considered. Due to big problems with estimating URG parameters the author does not know any practical applications of URG to describe the reliability results of technical objects. You can always add more parameters to the model, however, this might worsen their estimation.

4. INDEPENDENCE TESTS

4.1. Two-way contingency table

Features \(X, Y\) are independent what means that \(H_0\) is true, if \(p_{ij} = p_i \cdot p_j (i,j = 1,2)\) for each pair of \(i,j\). The alternative hypothesis denoted \(H_1\) is such one that negates \(H_0\). Let \(e_{ij}\) be the expected counts

\[
e_{ij} = \frac{n_i \cdot n_j}{n} = np_i \cdot p_j (i = 1, ..., w; j = 1, ..., k).
\]

The expected counts \(e_{ij}\) have the same one-way marginal values as the observed table \(n_{ij}\) (Gokhale, Kullback, 1978).

Statistical science has been enriched with many other statistics intended for research on test independency. Cressie, Read (1984) propose the power divergence statistics (PDS). The PDS for \(w \times k\) CTs is given by

\[
p^2 = \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^{w} \sum_{j=1}^{k} n_{ij}^* \left[ \left( \frac{n_{ij}^*}{e_{ij}^*} \right)^{\lambda} - 1 \right] =
\]

\[
= \frac{2n}{\lambda(\lambda + 1)} \sum_{i=1}^{w} \sum_{j=1}^{k} p_{ij}^* \left[ \left( \frac{p_{ij}^*}{p_i^* \cdot p_j^*} \right)^{\lambda} - 1 \right] - \infty < \lambda < \infty,
\]

where \(e_{ij}\) values are given by (3). Equation (4) always takes positive values and is defined as a limit of \(p^2\) at \(-1\) and \(0\). \(p^2\) contains a very rich class of test statistics, for example: the \(\chi^2\) statistics (\(\lambda = 1\)), the \(G^2\) statistics (the limit as \(\lambda\) goes to 0), the Freeman-Tukey statistics (\(\lambda = -0.5\)), the modified \(G^2\) statistics (the limit as \(\lambda\) goes to \(-1\)), the Neyman modified \(\chi^2\) statistics (\(\lambda = -2\)) and the Cressie-Read statistics (\(\lambda = 2/3\)). If \(H_0\) is true, statistics (4), for large \(n\) (i.e. asymptotically), follows the chi-square distribution with \((w - 1)(k - 1)\) degrees of freedom.
The following PDS are selected to Monte Carlo study: the $\chi^2$ statistics (Pearson, 1900), the Freeman-Tukey $FT$ statistics (Freeman, Tukey, 1950), the Cressie-Read $CR$ statistics (Cressie, Read, 1984):

$$\chi^2 = \sum_{i=1}^{w} \sum_{j=1}^{k} \frac{(n_{ij}^* - e_{ij})^2}{e_{ij}},$$

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

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

The $G^2$ statistics (Sokal, Rohlf, 2012), the modified $G^2$ statistics (Kullback, 1959) and the Neyman modified $\chi^2$ statistics (Neyman, 1949) have not been subjects in the Monte Carlo study because they are applicable only in a case where all $n_{ij}$ ($i = 1, ..., w; j = 1, ..., k$) counts are not equal to zero.

The square used in the numerator of $\chi^2$ statistics (5) makes that large differences between expected and theoretical counts even bigger and the small differences even smaller. Another aim of the use of the square is to avoid that the differences are mutually exclusive. For this purpose one can use their absolute value instead of squared deviations. The $|\chi|$ statistics is selected to Monte Carlo study, too. It is an authorial modification of $\chi^2$ statistics and it has the form (Sulewski, 2013)

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

where $n_{ij}$ are experimental counts, $e_{ij}^*$ are expected counts and $p_{ij}^*$ are empirical probabilities. It is shown in (Sulewski, 2016) that $|\chi|$ test is more powerful than tests (5)–(7).

### 4.2. Three-way contingency table

In this paper the research has been limited only to complete independence. Features $X, Y, Z$ are completely independent from one another, and $H_0$ is true, if
\[
P_{ijt} = p_{i\ast}\cdot p_{j\ast}\cdot p_{\ast\ast}
\]  
(9)

for each \( i = 1, ..., w; j = 1, ..., k; t = 1, ..., p \). The alternative hypothesis \( H_1 \) negates \( H_0 \).

Let \( e_{ijt} \) be the expected counts under complete independence of \( i, j, t \)

\[
e_{ijt} = \frac{n_{i\ast\ast}\cdot n_{j\ast\ast}\cdot n_{\ast\ast\ast}}{n^2} = n p_{i\ast}\cdot p_{j\ast}\cdot p_{\ast\ast\ast}(i = 1, ..., w; j = 1, ..., k; t = 1, ..., p).
\]  
(10)

The expected counts \( e_{ijt} \) under complete independence of \( i, j, t \) have the same one-way marginal values as the observed table \( n_{ijt} \) (Gokhale, Kullback, 1978)

To study the complete independence of the features \( X, Y, Z \) we use the statistics that are extensions of those for two-way CTs (Pardo, 1996)

\[
\begin{align*}
\chi_3^2 &= \sum_{i=1}^{w} \sum_{j=1}^{k} \sum_{t=1}^{p} \frac{(n_{ijt}^\ast - e_{ijt}^\ast)^2}{e_{ijt}^\ast}, \\
FT_3 &= 4 \sum_{i=1}^{w} \sum_{j=1}^{k} \sum_{t=1}^{p} \left( \frac{n_{ijt}^\ast}{\sqrt{e_{ijt}^\ast}} \right)^2, \\
CR_3 &= \frac{9}{5} \sum_{i=1}^{w} \sum_{j=1}^{k} \sum_{t=1}^{p} n_{ijt}^\ast \left[ \frac{(n_{ijt}^\ast)^{2/3}}{e_{ijt}^\ast} - 1 \right].
\end{align*}
\]  
(11)–(13)

Statistics (11)–(13) for CT, when \( H_0 \) is true, asymptotically follow the chi-square distribution with \( wkp - (w + k + p) + 2 \) degrees of freedom. Statistics

\[
\begin{align*}
G_3^2 &= 2 \sum_{i=1}^{w} \sum_{j=1}^{k} \sum_{t=1}^{p} n_{ijt}^\ast \ln \left( \frac{n_{ijt}^\ast}{e_{ijt}^\ast} \right), N_3 = \sum_{i=1}^{w} \sum_{j=1}^{k} \sum_{t=1}^{p} \frac{(n_{ijt}^\ast - e_{ijt}^\ast)^2}{n_{ijt}^\ast}, \\
KL_3 &= 2 \sum_{i=1}^{w} \sum_{j=1}^{k} \sum_{t=1}^{p} e_{ijt}^\ast \ln \left( \frac{e_{ijt}^\ast}{n_{ijt}^\ast} \right)
\end{align*}
\]  

also belong to the PDS. However, these statistics have not been applied in the Monte Carlo study, because they do not take into account the condition \( n_{ijt}^\ast = 0 \).

The \( |\chi_3^2| \) statistics is selected to Monte Carlo study, too. It is an authorial modification of \( \chi_3^2 \) statistics and it has the form (Sulewski, 2018a)
\[ |\chi_3| = \sum_{i=1}^{w} \sum_{j=1}^{k} \sum_{t=1}^{p} \frac{|n_{ijt}^* - e_{ijt}^*|}{e_{ijt}^*}, \quad (14) \]

where \( n_{ijt}^* \) are experimental counts and \( e_{ijt}^* \) are expected counts. It is shown in (Sulewski, 2018a) that \( |\chi_3| \) test is more powerful than tests (11)–(13).

5. MEASURES OF UNTRUTHFULNESS OF \( H_0 \)

5.1. Two-way contingency table

When the equality \( p_{ij} = p_i \cdot p_j \) is not fulfilled, \( H_0 \) is not true and an appropriate measure of untruthfulness of \( H_0 \) (MoUH) is needed. There are many different measures in literature, e.g.: the Pearson’s \( \varphi \), the Tschuprow’s \( T \), the Cramer’s \( V \), the corrected contingency \( c \), the Goodman and Kruskal’s \( \tau \).

In this paper we use a MoUH which is given by (Sulewski, 2016):

\[ MoU = \frac{1}{n} \sum_{i=1}^{w} \sum_{j=1}^{k} \left| \frac{n_{ij}^* - n_{i*}^* \cdot n_{*j}^*}{n} \right| = \sum_{i=1}^{w} \sum_{j=1}^{k} \left| \frac{p_{ij}^* - p_{i*}^* \cdot p_{*j}^*}{p_{ij}} \right|. \quad (15) \]

The \( MoU \) takes values in interval \((0,1)\). This measure, doubtlessly, springs from the essence of \( H_0 \) and has a very simple form. The \( MoU \) formulas and the maximal \( MoU \) values (the minimal \( MoU \) values are equal to zero) under scenarios I-XX are presented in table 4. The \( MoU \) is a function of the PFP \( a \). Owing to this the \( MoU \) values are very easy to calculate.

<table>
<thead>
<tr>
<th>Table</th>
<th>Scenario</th>
<th>MoU</th>
<th>( MoU_{max} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 x 3</td>
<td>I</td>
<td>0.2222</td>
<td></td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>0.3333</td>
<td></td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>0.4444</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>0.6667</td>
<td></td>
</tr>
<tr>
<td>3 x 3</td>
<td>IX</td>
<td>0.2469</td>
<td></td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>0.2963</td>
<td></td>
</tr>
<tr>
<td></td>
<td>XI</td>
<td>0.4444</td>
<td></td>
</tr>
<tr>
<td></td>
<td>XII</td>
<td>0.5926</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table</th>
<th>Scenario</th>
<th>MoU</th>
<th>( MoU_{max} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 x 4</td>
<td>V</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VI</td>
<td>** 0.3125</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VII</td>
<td>0.375</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VIII</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>3 x 4</td>
<td>XIII</td>
<td>0.2222</td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIV</td>
<td>0.3333</td>
<td></td>
</tr>
<tr>
<td></td>
<td>XV</td>
<td>0.4444</td>
<td></td>
</tr>
<tr>
<td></td>
<td>XVI</td>
<td>0.6667</td>
<td></td>
</tr>
</tbody>
</table>
5.2. Three-way contingency table

The theory devoted to MoUH for TT is not as rich as for the two-way contingency table, where the Goodman—Kruskal index plays an important role (Goodman, Kruskal, 1954). Numeric extensions of this index for three-way CT are: the Marcotorchino index $\tau_M$ (Marcotorchino, 1984), the delta index $\tau_L$ (Lombardo, 2011) and the Gray—Williams index $\tau_{GW}$ (Gray, Williams, 1975). Information about other less popular indices can be found in (Beh, Davy, 1998; Harshman, 1970; Lombardo, Beh, 2010; Trucker, 1963).

Based on the classical definition of independence of $X, Y, Z$, the $MoU_3$ in the form

$$MoU_3 = \frac{1}{n} \sum_{i=1}^{w} \sum_{j=1}^{k} \sum_{t=1}^{p} \left| n_{ijt}^{*} - \frac{n_{ij*}n_{jt*}n_{*jt}}{n^2} \right| = \sum_{i=1}^{w} \sum_{j=1}^{k} \sum_{t=1}^{p} \left| p_{ijt} - p_{i*}p_{j*t}p_{*jt} \right|,$$  (16)

is put forward in (Sulewski, 2018a). The measure (16) takes the value 0 when $H_0$ is true. The higher the $MoU_3$ value, the greater the possibility of $H_0$ falsity. More information about the measures $MoU_3$, $\tau_M$, $\tau_L$ and $\tau_{GW}$ defined under some scenarios for three-way CT you can find in (Sulewski, 2018a).

The $MoU_3$ as a natural measure, resulting from the definition of independence, is used in the Monte Carlo simulation. The $MoU_3$ formulas and the maximal $MoU_3$ values (the minimal $MoU_3$ values are equal to zero) under scenarios XXI–XXVIII are presented in table 5. The $MoU_3$ is a function of the PFP $a$. Owing to this the $MoU_3$ values are very easy to calculate.

<table>
<thead>
<tr>
<th>Table Scenario</th>
<th>$MoU_3$</th>
<th>$MoU_{3\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2×2×2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XXI</td>
<td>$16a^2$</td>
<td>0.25</td>
</tr>
<tr>
<td>XXII</td>
<td>*</td>
<td>0.359</td>
</tr>
<tr>
<td>XXIII</td>
<td>$3a$</td>
<td>0.375</td>
</tr>
<tr>
<td>XXIV</td>
<td>$4a$</td>
<td>0.5</td>
</tr>
<tr>
<td>3×2×2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XXV</td>
<td>$8a/3$</td>
<td>0.2222</td>
</tr>
<tr>
<td>XXVI</td>
<td>**</td>
<td>0.3889</td>
</tr>
<tr>
<td>XXVII</td>
<td>$6a$</td>
<td>0.5</td>
</tr>
<tr>
<td>XXVIII</td>
<td>$8a$</td>
<td>0.6667</td>
</tr>
</tbody>
</table>

Source: own elaboration.
6. APPLYING THE MAXIMUM LIKELIHOOD METHOD TO ESTIMATE THE PROBABILITY FLOW PARAMETER

This section is a simply attempt of replacing a nonparametric statistical inference method by the parametric one. Maximum likelihood method is applied to estimate the PFP.

Let us remember that in cells of two-way CT are values $n_{ij}$ ($i = 1, ..., w; j = 1, ..., k$), in cells of three-way CT are values $n_{ijt}$ ($i = 1, ..., w; j = 1, ..., k; t = 1, ..., p$). These values are components of the multinomial distribution. Thus the multinomial distribution was taken as a groundwork of the likelihood functions. A family of these likelihood functions is given below. Every function from this family has an index. Indices assign functions to particular scenarios presented in section 3 of the paper.

6.1. Two-way contingency table

Let $n_{ij}^*$ be the value of $(i, j)$ cell and $a$ is the probability flow parameter. Then likelihood functions under CT $w \times k$ have the form

a) table $2 \times 3$

$$L_I(a) = C(1/6 - a)^{n_{11}}(1/6 + a)^{n_{23}}(1/6)^{n_{12} + n_{13} + n_{21} + n_{22}}, \quad (17)$$

$$L_{II}(a) = C(1/6 - a)^{n_{12} + n_{13}}(1/6 + a)^{n_{22} + n_{23}}(1/6)^{n_{12} + n_{13} + n_{21}}, \quad (18)$$

$$L_{III}(a) = C(1/6 - a)^{n_{11} + n_{12}}(1/6 + a)^{n_{22} + n_{23}}(1/6)^{n_{13} + n_{23}}, \quad (19)$$

$$L_{IV}(a) = C(1/6 - a)^{n_{11} + n_{12} + n_{13}}(1/6 + a)^{n_{22} + n_{23}}(1/6)^{n_{13} + n_{21} + n_{22}}, \quad (20)$$

b) table $2 \times 4$

$$L_V(a) = C(1/8 - a)^{n_{11}}(1/8 + a)^{n_{24}}(1/8)^{n_{12} + n_{13} + n_{14} + n_{21} + n_{22} + n_{23}}, \quad (21)$$

$$L_{VI}(a) = C(1/8 - a)^{n_{11} + n_{12}}(1/8 + a)^{n_{22} + n_{23} + n_{24}}(1/8)^{n_{14} + n_{21}}, \quad (22)$$

$$L_{VII}(a) = C(1/8 - a)^{n_{12}}(1/8 + a)^{n_{21}}(1/8)^{n_{11} - n_{21}}, \quad (23)$$

$$L_{VIII}(a) = C(1/8 - a)^{n_{11} + n_{12}}(1/8 + a)^{n_{21} + n_{22}}(1/8)^{n_{13} + n_{14} + n_{23} + n_{24}}, \quad (24)$$
c) table $3 \times 3$

\[
L_{IX}(a) = C(1/9 - a)^{n_{i1}^*}(1/9 + a)^{n_{i3}^*}(1/9)^{n_i - n_{i1}^* - n_{i3}^*},
\]

\[
L_X(a) = C(1/9 - a)^{n_{i1}^* + n_{i2}^*}(1/9 + a)^{n_{i3}^* + n_{i3}^*}(1/9)^{n_i - n_{i1}^* - n_{i2}^* - n_{i3}^*},
\]

\[
L_{XI}(a) = C(1/9 - a)^{n_{i1}^* + n_{i3}^*}(1/9 + a)^{n_{i1}^* + n_{i3}^*}(1/9)^{n_i - n_{i1}^* - n_{i2}^* - n_{i3}^*},
\]

\[
L_{XII}(a) = C(1/9 - a)^{n_{i1}^* + n_{i2}^* + n_{i3}^*}(1/9 + a)^{n_{i1}^* + n_{i2}^* + n_{i3}^*}(1/9)^{n_i - n_{i1}^* - n_{i2}^* - n_{i3}^*},
\]

d) table $3 \times 4$

\[
L_{XIII}(a) = C(1/12 - a)^{n_{i1}^*}(1/12 + a)^{n_{i4}^*}(1/12)^{n_i - n_{i1}^* - n_{i4}^*},
\]

\[
L_{XIV}(a) = C(1/12 - a)^{n_{i1}^* + n_{i2}^*}(1/12 + a)^{n_{i3}^* + n_{i4}^*}(1/12)^{n_i - n_{i1}^* - n_{i2}^* - n_{i3}^* - n_{i4}^*},
\]

\[
L_{XV}(a) = C(1/12 - a)^{n_{i1}^* + n_{i2}^*}(1/12 + a)^{n_{i3}^* + n_{i4}^*}(1/12)^{n_i - n_{i1}^* - n_{i2}^* - n_{i3}^* - n_{i4}^*},
\]

\[
L_{XVI}(a) = C(1/12 - a)^{n_{i1}^* + n_{i3}^* + n_{i4}^*}(1/12 + a)^{n_{i1}^* + n_{i4}^* + n_{i3}^* + n_{i2}^*}(1/12)^{n_i - n_{i1}^* - n_{i2}^* - n_{i3}^* - n_{i4}^*},
\]

e) table $4 \times 4$

\[
L_{XVII}(a) = C(1/16 - a)^{n_{i1}^* + n_{i2}^* + n_{i3}^* + n_{i4}^*}(1/16 + a)^{n_{i4}^* + n_{i4}^*}(1/16)^{n_i - n_{i1}^* - n_{i2}^* - n_{i4}^* - n_{i3}^* - n_{i4}^*},
\]

\[
L_{XVIII}(a) = C(1/16 - a)^{n_{i1}^* + n_{i2}^* + n_{i3}^* + n_{i4}^*}(1/16)^{n_i - n_{i1}^* - n_{i2}^* - n_{i3}^* - n_{i4}^* - n_{i4}^*},
\]

\[
L_{XIX}(a) = C(1/16 - a)^{n_{i1}^* + n_{i2}^* + n_{i3}^* + n_{i4}^*}(1/16)^{n_i - n_{i1}^* - n_{i2}^* - n_{i3}^* - n_{i4}^* - n_{i4}^*},
\]

\[
L_{XX}(a) = C(1/16 - a)^{n_{i1}^* + n_{i2}^* + n_{i3}^* + n_{i4}^*}(1/16)^{n_i - n_{i1}^* - n_{i2}^* - n_{i3}^* - n_{i4}^* - n_{i4}^*},
\]

In formulas (17)–(36) $C = n! / \prod_{i=1}^{w} \prod_{j=1}^{k} n_{i,j}!$.

The logarithmic likelihood function under scenario I in CT $2 \times 3$ is given by

\[
l_t(a) = 1n L_t(a) = 1n(C) + n_{i1}^* 1n(1/6 + a) + (n_{i2}^* + n_{i3}^* n_{i22}^* n_{i22}^*) 1n(1/6),
\]

then

\[
\frac{\partial l_t(a)}{\partial a} = \frac{n_{i3}^*}{1/6 + a} - \frac{\partial l_t(a)}{\partial a} = 0 \Rightarrow \hat{a} = \frac{n_{i3}^* - n_{i1}^*}{6(n_{i3}^* + n_{i1}^*)}.
\]
Let us check what kind of extremum we can found. As a result of a simple transformation we have

\[
\frac{\partial l(x)}{\partial a^2} = \frac{\partial}{\partial a} \left[ \frac{n_{23}^*}{1/6 + a} - \frac{n_{11}^*}{1/6 - a} \right] = -\frac{n_{23}^*}{(1/6 + a)^2} - \frac{n_{11}^*}{(1/6 - a)^2} < 0, \tag{37}
\]

for \( a < 1/6 \). It means that the logarithmic likelihood function has always a maximum at \( a = \hat{a} \). So, \( \hat{a} \) is the maximum likelihood estimator of \( a \) which is the probability flow parameter. It may be proven that inequality (37) holds for all scenarios considered in this paper.

Formulas for the maximum likelihood estimator of \( a \) under scenarios I-XX for two-way CT are given by:

a) Table 2 x 3:

\[
\hat{a}_I = \frac{n_{23}^* - n_{11}^*}{6(n_{23}^* + n_{11}^*)}, \quad \hat{a}_{II} = \frac{n_{22}^* + n_{23}^* - (n_{11}^* + n_{12}^*)}{6(n_{11}^* + n_{12}^* + n_{22}^* + n_{23}^*)},
\]

\[
\hat{a}_{III} = \frac{n_{21}^* + n_{22}^* - (n_{11}^* + n_{12}^*)}{6(n_{11}^* + n_{12}^* + n_{21}^* + n_{22}^*)}, \quad \hat{a}_{IV} = \frac{n_{13}^* + n_{21}^* - (n_{11}^* + n_{23}^*)}{6(n_{11}^* + n_{12}^* + n_{21}^* + n_{23}^*)}.
\tag{38}
\]

b) Table 2 x 4:

\[
\hat{a}_V = \frac{n_{24}^* - n_{11}^*}{8(n_{24}^* + n_{11}^*)}, \quad \hat{a}_{VI} = \frac{n_{22}^* + n_{23}^* + n_{24}^* - (n_{11}^* + n_{12}^* + n_{13}^*)}{8(n - n_{14}^* - n_{21}^*)},
\]

\[
\hat{a}_{VII} = \frac{n_{21}^* - n_{11}^*}{8(n_{11}^* + n_{21}^*)}, \quad \hat{a}_{VIII} = \frac{n_{21}^* + n_{22}^* - (n_{11}^* + n_{12}^*)}{8(n_{11}^* + n_{12}^* + n_{21}^* + n_{22}^*)}.
\tag{39}
\]

c) Table 3 x 3:

\[
\hat{a}_{XI} = \frac{n_{13}^* - n_{11}^*}{9(n_{11}^* + n_{33}^*)}, \quad \hat{a}_X = \frac{n_{31}^* + n_{32}^* - (n_{11}^* + n_{12}^*)}{9(n_{11}^* + n_{12}^* + n_{31}^* + n_{32}^*)},
\]

\[
\hat{a}_{XI} = \frac{n_{13}^* + n_{31}^* - (n_{11}^* + n_{33}^*)}{9(n_{11}^* + n_{13}^* + n_{31}^* + n_{33}^*)}, \quad \hat{a}_{XII} = \frac{n_{13}^* + n_{23}^* + n_{31}^* - (n_{11}^* + n_{12}^* + n_{33}^*)}{9(n - n_{12}^* - n_{22}^* - n_{32}^*)}.
\tag{40}
\]

d) Table 3 x 4:

\[
\hat{a}_{XIII} = \frac{n_{34}^* - n_{11}^*}{12(n_{11}^* + n_{34}^*)}, \quad \hat{a}_{XIV} = \frac{n_{31}^* + n_{32}^* - (n_{11}^* + n_{12}^*)}{12(n_{11}^* + n_{12}^* + n_{31}^* + n_{32}^*)},
\]

\[
\hat{a}_{XV} = \frac{n_{13}^* + n_{14}^* - (n_{11}^* + n_{12}^*)}{12(n_{11}^* + n_{12}^* + n_{13}^* + n_{14}^*)}, \quad \hat{a}_{XVI} = \frac{n_{13}^* + n_{14}^* + n_{31}^* + n_{32}^* - (n_{11}^* + n_{12}^* + n_{31}^* + n_{32}^*)}{12(n - n_{21}^* - n_{22}^* - n_{23}^* - n_{24}^*)}.
\tag{41}
\]
e) table 4 × 4

\[
\hat{a}_{XVIII} = \frac{n_{i41} + n_{i42} + n_{i43} - (n_{i11} + n_{i12} + n_{i13})}{16(n_{i11} + n_{i12} + n_{i13} + n_{i41} + n_{i42} + n_{i43})},
\]
\[
\hat{a}_{XIX} = \frac{n_{i32} + n_{i33} + n_{i42} + n_{i43} - (n_{i11} + n_{i12} + n_{i13} + n_{i21} + n_{i22} + n_{i23})}{16(n_{i11} + n_{i12} + n_{i13} + n_{i41} + n_{i42} + n_{i43} + n_{i44})},
\]
\[
\hat{a}_{XX} = \frac{n_{i13} + n_{i14} + n_{i41} + n_{i42} - (n_{i11} + n_{i12} + n_{i31} + n_{i44})}{16(n_{i11} + n_{i12} + n_{i13} + n_{i41} + n_{i42} + n_{i43} + n_{i44})}.
\]

To decide which of the defined scenarios takes place, you use the following algorithm:

1. Find dimension of CT in question and read out related values according to the rule:
   - 2 × 3 (q = 1), 2 × 4 (q = 2), 3 × 3 (q = 3), 3 × 4 (q = 4), 4 × 4 (q = 5)
2. Set a set of scenario indices \( \{4q-3, 4q-2, 4q-1, 4q\} \).
3. Calculate \( a^* \), which is an estimate of parameter \( a \) for each scenario from step 2 by means of (38)–(42).
4. Calculate corresponding values of the maximum likelihood functions \( L_{4q-3}(a^*), L_{4q-2}(a^*), L_{4q-1}(a^*), L_{4q}(a^*) \) by means of (17)–(36).
5. Choose a scenario for which the value \( L(a^*) \) is the greatest.

This algorithm will be used in section 8.1, Example 1.

6.2. Three-way contingency table

Let \( n_{ijt} \) be the value of \( (i, j, t) \) cell and \( a \) is the PFP. Then likelihood functions for selected three-way CT have the form:

a) table 2 × 2 × 2

\[
L_{XXI}(a) = D(1/8 - a)^{n_{i11} + n_{i12} (1/8 + a)^{n_{j21} + n_{j22} (1/8)^{n_{k21} + n_{k22} n_{i11} + n_{i12} + n_{i21} + n_{i22} + n_{i23}},}
\]
\[
L_{XXII}(a) = D(1/8 - a)^{n_{i11} + n_{i12} (1/8 + a)^{n_{j21} + n_{j22} (1/8)^{n_{k21} + n_{k22} n_{i11} + n_{i12} + n_{i21} + n_{i22} + n_{i23},}
\]
\[
L_{XXIII}(a) = D(1/8 - a)^{n_{i11} (1/8 + a)^{n_{j21} (1/8)^{n_{k21} + n_{k22} n_{i11} + n_{i12} + n_{i21} + n_{i22} + n_{i23},}
\]
\[
L_{XXIV}(a) = D(1/8 - a)^{n_{i11} + n_{i12} (1/8 + a)^{n_{j21} + n_{j22} (1/8)^{n_{k21} + n_{k22} n_{i11} + n_{i12} + n_{i21} + n_{i22} + n_{i23},}
\]

b) table 3 × 2 × 2

\[
L_{XXV}(a) = D(1/8 - a)^{n_{i11} + n_{i12} (1/8 + a)^{n_{j321} + n_{j322} (1/8)^{n_{k311} - n_{k312} - n_{k321} - n_{k322},}
\]
\[
L_{XXVI}(a) = D(1/8 - a)^{n_{i11} + n_{i12} (1/8 + a)^{n_{j321} (1/8)^{n_{k311} - n_{k312} - n_{k321} - n_{k322},}
\]
\[
L_{XXVII}(a) = D(1/8 - a)^{n_{i11} + n_{i12} + n_{j21} (1/8 + a)^{n_{j11} + n_{j12} + n_{j21} + n_{j22} + n_{j23},}
\]
\[
L_{XXVIII}(a) = D(1/8 - a)^{n_{i11} + n_{i12} + n_{j21} (1/8)^{n_{j11} + n_{j12} + n_{j21} + n_{j22} + n_{j23},}
\]
\[
L_{XXVIII}(a) = D(1/8 - a)^{n_{111} + n_{112} + n_{321} + n_{322}} (1/8 + a)^{n_{121} + n_{122} + n_{311} + n_{312}}
(1/8)^{n_{211} + n_{221} + n_{212} + n_{222}}. \tag{50}
\]

In formulas (43)–(50) \(D = n! / \prod_{i=1}^{w} \prod_{j=1}^{k} \prod_{t=1}^{p} n_{ijt}!\).

Formulas for the maximum likelihood estimator of \(a\) under scenarios XXIV–XXVIII for CT \(w \times k \times p\) are given by

a) table \(2 \times 2 \times 2\)
\[
\hat{a}_{XXI} = \frac{n_{221}^* + n_{222}^* - (n_{111}^* + n_{112}^*)}{8(n_{111}^* + n_{112}^* + n_{221}^* + n_{222}^*)},
\hat{a}_{XXII} = \frac{n_{221}^* + n_{212}^* + n_{222}^* - (n_{111}^* + n_{121}^* + n_{122}^*)}{8(n - n_{122}^* - n_{211}^*)},
\hat{a}_{XXIII} = \frac{n_{211}^* - n_{111}^*}{8(n_{111}^* + n_{211}^*)}, \hat{a}_{XXIV} = \frac{n_{211}^* + n_{212}^* - (n_{111}^* + n_{112}^*)}{8(n_{111}^* + n_{112}^* + n_{211}^* + n_{212}^*)}.
\tag{51}
\]

a) table \(3 \times 2 \times 2\)
\[
\hat{a}_{XXV} = \frac{n_{321}^* + n_{322}^* - (n_{111}^* + n_{112}^*)}{12(n_{111}^* + n_{112}^* + n_{321}^* + n_{322}^*)},
\hat{a}_{XXVI} = \frac{n_{122}^* + n_{321}^* - (n_{111}^* + n_{112}^*)}{12(n_{111}^* + n_{112}^* + n_{321}^* + n_{322}^*)},
\hat{a}_{XXVII} = \frac{n_{121}^* + n_{122}^* + n_{211}^* - (n_{111}^* + n_{112}^* + n_{221}^*)}{8(n_{111}^* + n_{112}^* + n_{211}^* + n_{221}^*)},
\hat{a}_{XXVIII} = \frac{n_{121}^* + n_{112}^* + n_{311}^* + n_{312}^* - (n_{111}^* + n_{112}^* + n_{312}^* + n_{321}^*)}{12(n - n_{211}^* - n_{221}^* - n_{212}^* - n_{222}^*)}.
\tag{52}
\]

To decide which of the defined scenarios takes place, you use the following algorithm:
1. Find dimension of CT in question and read out related \(q\) according to the rule:
   \[
   2 \times 2 \times 2 (q = 6), 3 \times 2 \times 2 (q = 7).
   \]
2. Set a set of scenario indices \(\{4q - 3, 4q - 2, 4q - 1, 4q\}\).
3. Calculate \(a^*\), which is an estimate of parameter \(a\) for each scenario from step 2 according to (51)–(52)
4. Calculate corresponding values of the maximum likelihood functions \(L_{4q-3}(a^*), L_{4q-2}(a^*), L_{4q-1}(a^*), L_4(a^*)\) by means of (43)–(50).
5. Choose a scenario for which the value \(L(a^*)\) is the greatest.

This algorithm will be used in section 8.2, Example 4.

7. GENERATING CONTINGENCY TABLE

Generating CT is very important in the simulation study. The approach in the literature for the generating two-way CTs is the Markov Chain Monte Carlo (Diaconis, Sturmfels, 1998; Cryan, Dyer, 2003; Cryan et al., 2006; Chen et al., 2005; Fishman, 2012), the Sequential Importance Sampling (Chen et al., 2005; Chen et al., 2006; Blitzstein, Diaconis, 2011; Yoshida, 2011), the probabilistic

IN THIS PAPER WE USE AN ALGORITHM FOR GENERATING TWO-WAY AND THREE-WAY CT'S USING THE BAR METHOD. THE BAR METHOD IS IDENTICAL TO THE METHOD THAT GENERATES RANDOM NUMBERS THAT FOLLOW THE MULTINOMIAL DISTRIBUTION.

8. PARAMETRIC REASONING PUT INTO PRACTICE

8.1. TWO-WAY CONTINGENCY TABLE

EXAMPLE 1. THIS EXAMPLE COMPARES DECISION MAKING BY MEANS OF CLASSIC STATISTIC TESTING WITH LIKELIHOOD BASED DECISION MAKING. TABLES 6-10 SHOW A SET OF EXAMPLE TWO-WAY CT'S FILLED ONE BY ONE ACCORDINGLY TO THE SCENARIOS I–XX. THE TABLE IS DIVIDED INTO TWO PARTS. THE LEFT HAND SIDE IS RELATED TO LIKELIHOOD BASED DECISIONS. TO DECIDE WHICH OF THE DEFINED SCENARIOS TAKES PLACE, SEE ALGORITHM IN SECTION 6.1. THE RIGHT HAND SIDE IS RELATED TO CLASSIC STATISTICAL TESTING AND PRESENTS VALUES OF TEST STATISTICS. THE $H_0$ STATES THAT $X$ AND $Y$ ARE INDEPENDENT. CRITICAL VALUES, INDICATED BY UNDERLINING, ARE DETERMINED BY MONTE CARLO METHOD BASED ON $10^6$ ORDER STATISTICS. SUCH A LARGE NUMBER OF REPETITIONS GUARANTIES VERY PRECISE RESULTS. WHEN READING ROWS OF THE TABLE IT TURNS OUT THAT ALL THE DECISIONS MADE IN A CLASSIC WAY ARE WRONG. IT IS BECAUSE UNTRUE $H_0$ HYPOTHESES HAVE NOT BEEN REJECTED. BUT IT DOES NOT REVEAL ANYTHING NEW. THIS IS JUST ONE MORE CONFIRMATION OF WHAT IS COMMONLY KNOWN – THE CLASSIC STATISTICAL TEST IS VERY CONSERVATIVE. THE PFP $a$ IS A MAXIMAL VALUE OF THIS PARAMETER FOR WHICH UNTRUE $H_0$ IS REJECTED IN NO SCENARIOS.

TABLES 6–10 SHOW THAT ALL THE DECISIONS MADE IN A CLASSIC WAY ARE WRONG UNDER THE SCENARIOS IN QUESTION. IT IS BECAUSE UNTRUE $H_0$ HYPOTHESES HAVE NOT BEEN REJECTED EVEN IN SITUATIONS WHERE THE MOU DOES NOT HAVE SUCH SMALL VALUES, E.g. $MOU = 0.3$ IN SCENARIO XX. IN TURN, THE PARAMETRIC APPROACH DETECTED A DEPENDENCY BETWEEN FEATURES.

EXAMPLE 2. THE CONSERVATIVENESS OF CLASSIC TESTING IS A REASON WHY WE SUGGEST MAKING A TURNAROUND IN THIS DOMAIN. THE NEW PROPOSAL IS A METHOD OF STATISTICAL INFERENCE AND NOT A CLASSICAL PARAMETRIC TEST. NOW THERE WILL BE NO NULL HYPOTHESIS, BUT THERE WILL BE A SET OF COMPETING ALTERNATIVE HYPOTHESES INSTEAD. THE FORMER $H_0$ IS NO LONGER THE MAIN ONE, BUT EXISTS AMONG THE COMPETITORS OF AN EQUAL RANK. ALL THE HYPOTHESES STATE: "THE CONSIDERED TWO-WAY CT IS GENERATED ACCORDINGLY TO PARTICULAR SCENARIOS". EACH FIGURE FROM 1 TO 5 SHOWS SETS OF FOUR LIKELIHOOD CURVES. EACH CURVE HAS ITS FOUR ATTRIBUTES, NAMELY $\tilde{a}$, $n$, TABLE DIMENSIONS AND, OF COURSE, THE ACTUAL GENERATION SCENARIO THAT IS SPECIFIED IN THE FIGURE'S TITLE. THE LEGEND LISTS ALL COMPETING SCENARIOS INCLUDING THE ACTUAL ONE.
| Scenario (MoU) | Content | X and Y dependent | Scenario of maximum likelihood | $\chi^2$ | FT | CR | $|x|$ | $H_0$ true | $H_0$ rejected |
|---------------|---------|-------------------|-------------------------------|-------|-----|-----|-----|-----------|---------------|
| I(0.067)      | 7       | 10                | Yes                           | 1     | 0.324 | 0.324 | 0.324 | 0.418     | No            | No            |
|               | 10      | 10                |                               |       |      |      |      |           |               |               |
| II(0.067)     | 7       | 7                 | Yes                           | 1     | 0.334 | 0.338 | 0.335 | 0.402     | No            | No            |
|               | 10      | 13                |                               |       |      |      |      |           |               |               |
| III(0.133)    | 7       | 7                 | Yes                           | 1     | 1.250 | 1.237 | 1.246 | 0.833     | No            | No            |
|               | 13      | 13                |                               |       |      |      |      |           |               |               |
| IV(0.200)     | 7       | 10                | Yes                           | 1     | 3.600 | 3.706 | 3.612 | 1.200     | No            | No            |
|               | 13      | 10                |                               |       |      |      |      |           |               |               |

Experiment settings: $n = 60, \alpha = 0.05, \alpha = 0.05$

Source: own elaboration.
### Table 7. THE CLASSIC STATISTICAL TESTING VERSUS LIKELIHOOD BASED DECISIONS, TABLE 2 × 4

Experiment settings: $n = 80, \alpha = 0.0625, \alpha = 0.05$

| Scenario (MoU) | Scenario of maximum likelihood | $\chi^2$ | $FT$ | $CR$ | $|x|$ | $H_0$ true | $H_0$ rejected |
|----------------|--------------------------------|---------|-----|-----|------|------------|-------------|
| V(0.125)       | Yes V                          | 8.058   | 8.657 | 8.131 | 2.283 | No         | No          |
|                | 10 10 10 15                     |         |      |      |      |            |             |
| VI(0.125)      | Yes VI                         | 8.657   | 8.131 | 2.283 |      | No         | No          |
|                | 10 15 15 15                     |         |      |      |      |            |             |
| VII(0.188)     | Yes VII                        | 8.131   | 2.283 |      |      | No         | No          |
|                | 15 10 10 10                     |         |      |      |      |            |             |
| VIII(0.250)    | Yes VIII                       | 2.283   |      |      |      | No         | No          |
|                | 15 15 10 10                     |         |      |      |      |            |             |

Source: own elaboration.
| Scenario (MoU) | Content | $X$ and $Y$ dependent | Scenario of maximum likelihood | $\chi^2$ | $FT$ | $CR$ | $|x|$ | $H_0$ true | $H_0$ rejected |
|---------------|---------|-----------------------|-------------------------------|---------|-------|-------|-------|-------------|----------------|
| IX(0.093)     | 6       | 10                    | 10                            | 9.416   | 10.067| 9.458 | 2.544 | No          | No             |
|               | 10      | 10                    | 10                            |         |       |       |       |             |                |
|               | 10      | 10                    | 14                            |         |       |       |       |             |                |
| X(0.119)      | 6       | 6                     | 10                            |         |       |       |       |             |                |
|               | 10      | 10                    | 10                            |         |       |       |       |             |                |
|               | 14      | 14                    | 10                            |         |       |       |       |             |                |
| XI(0.178)     | 6       | 10                    | 14                            |         |       |       |       |             |                |
|               | 10      | 10                    | 10                            |         |       |       |       |             |                |
|               | 14      | 10                    | 6                             |         |       |       |       |             |                |
| XII(0.237)    | 6       | 10                    | 14                            |         |       |       |       |             |                |
|               | 14      | 10                    | 6                             |         |       |       |       |             |                |

Source: own elaboration.
### Table 9. THE CLASSIC STATISTICAL TESTING VERSUS LIKELIHOOD BASED DECISIONS, TABLE 3 × 4

Experiment settings: \( n = 80, \alpha = 0.0625, \alpha = 0.05 \)

| Scenario (MoU) | Content | Scenario and \( Y \) dependent | \( \chi^2 \) | \( FT \) | \( CR \) | \( |\chi| \) | \( H_0 \) true | \( H_0 \) rejected |
|----------------|---------|---------------------------------|-------------|---------|--------|----------|---------------|-----------------|
| XIII(0.067)    | 7 10 10 10 | Yes | 12.565 | 13.463 | 12.638 | 3.325 | No | No |
|                | 10 10 10 10 |                   |            |         |        | | | |
|                | 10 10 10 10 |                   |            |         |        | | | |
| XIV(0.100)     | 7 7 10 10 | Yes | 12.565 | 13.463 | 12.638 | 3.325 | No | No |
|                | 10 10 10 10 |                   |            |         |        | | | |
|                | 10 10 10 10 |                   |            |         |        | | | |
| XV(0.133)      | 7 7 13 13 | Yes | 12.565 | 13.463 | 12.638 | 3.325 | No | No |
|                | 10 10 10 10 |                   |            |         |        | | | |
|                | 10 10 10 10 |                   |            |         |        | | | |
| XVI(0.200)     | 7 7 13 13 | Yes | 12.565 | 13.463 | 12.638 | 3.325 | No | No |
|                | 10 10 10 10 |                   |            |         |        | | | |
|                | 10 10 7 7 |                   |            |         |        | | | |

Source: own elaboration.
### Table 10. THE CLASSIC STATISTICAL TESTING VERSUS LIKELIHOOD BASED DECISIONS, TABLE 4 × 4

| Scenario (MoU) | Content | $X$ and $Y$ dependent | Scenario of maximum likelihood | $\chi^2$ | $FT$ | $CR$ | $|\chi|$ | $H_0$ true | $H_0$ rejected |
|---------------|---------|------------------------|-------------------------------|--------|-----|-----|--------|------------|--------------|
| XVII(0.113)   | 2 2 2 5 | Yes                    | XVII                          | 16.798 | 22.675 | 16.965 | 6.370 | No         | No          |
|               | 5 5 5 5 |                        |                               |        |      |      |        |            |              |
|               | 5 5 5 5 |                        |                               |        |      |      |        |            |              |
|               | 8 8 8 5 |                        |                               |        |      |      |        |            |              |
| XVIII(0.150)  | 2 2 5 5 | Yes                    | XVIII                         | 2.206  | 2.277 | 2.217 | 2.698 | No         | No          |
|               | 5 5 5 5 |                        |                               |        |      |      |        |            |              |
|               | 5 5 5 5 |                        |                               |        |      |      |        |            |              |
|               | 5 5 8 8 |                        |                               |        |      |      |        |            |              |
| XIX(0.150)    | 2 2 2 5 | Yes                    | XIX                           | 2.704  | 2.677 | 2.690 | 2.712 | No         | No          |
|               | 5 8 8 8 |                        |                               |        |      |      |        |            |              |
|               | 5 8 8 8 |                        |                               |        |      |      |        |            |              |
| XX(0.300)     | 2 2 8 8 | Yes                    | XX                            | 14.400 | 16.421 | 14.613 | 4.800 | No         | No          |
|               | 5 5 5 5 |                        |                               |        |      |      |        |            |              |
|               | 5 5 5 5 |                        |                               |        |      |      |        |            |              |
|               | 8 8 2 2 |                        |                               |        |      |      |        |            |              |

Source: own elaboration.
Figure 1. The likelihood functions for $\hat{a} = 0.05, n = 60$, table $2 \times 3$

Source: own elaboration.

Figure 2. The likelihood functions for $\hat{a} = 0.0625, n = 80$, table $2 \times 4$

Source: own elaboration.
Figure 3. The likelihood functions for $\alpha = 0.0444$, $n = 90$, table $3 \times 3$

Source: own elaboration.

Figure 4. The likelihood functions for $\alpha = 0.0250$, $n = 120$, table $3 \times 4$

Source: own elaboration.
Figure 5. The likelihood functions for $\hat{a} = 0.0375, n = 80$, table $4 \times 4$

Source: own elaboration.

It is noteworthy (see figures 1–5) that in each set of likelihood curves, the curve related to the actual scenario predominates over the others. It is instructive to read out a value of $a_{\text{max}}$ that maximizes likelihood function on a particular figure and notice that this value is close to assumed $\hat{a}$.

**Example 3.** This example is carried out in accordance with the following algorithm:

1. Find dimension of CT in question and read out related $q$ according to the rule:
   
   \[ 2 \times 3 (q = 1), 2 \times 4 (q = 2), 3 \times 3 (q = 3), 3 \times 4 (q = 4), 4 \times 4 (q = 5) \]

2. Set a set of scenario indices $\{4q - 3, 4q - 2, 4q - 1, 4q\}$.

3. Calculate values of PFP by means of $a_i = 0.1i/(\text{wk})$ $(i = 1, 2, ..., 10)$.

4. Set a sample size $n$.

5. Repeat the following steps $u = 10^4$ times:

   5.1. Let $Sc_{4q-3} = 0, Sc_{4q-2} = 0, Sc_{4q-1} = 0, Sc_{4q} = 0$

   5.2. Generate CT under the scenarios that you have chosen in Step 2.

   5.3. Calculate $L_{4q-3}(a), L_{4q-2}(a), L_{4q-1}(a), L_{4q}(a)$ by means of (17)–(36).

   5.4. If $\text{MaxL} = L_{4q-3}(a)$, then $Sc_{4q-3} = Sc_{4q-3} + 1$,
   
   If $\text{MaxL} = L_{4q-2}(a)$, then $Sc_{4q-2} = Sc_{4q-2} + 1$, 

   

If $MaxL = L_{4q-1}(a)$, then $Sc_{4q-1} = Sc_{4q-1} + 1$.
If $MaxL = L_{4q}(a)$, then $Sc_{4q} = Sc_{4q} + 1$.

6. Calculate probabilities of recognizing (PoR) scenarios. These are probabilities for the actual scenario to be recognized as one of scenarios in question.

$$Pr_{4q-3} = Sc_{4q-3}/u, Pr_{4q-2} = Sc_{4q-2}/u, Pr_{4q-1} = Sc_{4q-1}/u, Pr_{4q} = Sc_{4q}/u.$$  

Figures 6-10 present PoR(a) functions for two-way CT in question. Sample sizes for a given CT are different because maximal MoU values under the scenarios are different (see table 4). The minimal sample sizes are chosen in such a way that probabilities of proper recognition (actual I as I, ..., actual XX as XX) are greater than probabilities of improper recognitions for all the $a$ values.

Figure 6. The PoR actual scenario as one of I–IV scenarios, table $2 \times 3$

Source: own elaboration.
Figure 7. The PoR actual scenario as one of V–VIII scenarios, table $2 \times 4$

Source: own elaboration.

Figure 8. The PoR actual scenario as one of IX–XII scenarios, table $3 \times 3$

Source: own elaboration.
Figure 9. The PoR actual scenario as one of XIII–XVI scenarios, table 3 × 4

Source: own elaboration.

Figure 10. The PoR actual scenario as one of XVII–XX scenarios, table 4 × 4

Source: own elaboration.
Figures 6–10 show that even when samples are small (e.g. 15 items), probabilities of proper recognition are greater than probabilities of improper recognitions, regardless how small the PFP is. The bigger PFP $a$, the bigger PoR actual scenario. In the classic statistical testing related to $2 \times 3$ CT (see table 6) untrue $H_0$ has not been rejected even if $n = 60$, PFP $a = 0.05$ and $MoU = 0.2$. In a likelihood based decision dependence between features in $2 \times 3$ CT is visible already for $n = 15$ and PFP $a < 0.05$ (see figure 6). In the classic statistical testing related to $2 \times 4$ CT (see table 7) untrue $H_0$ has not been rejected even if $n = 80$, PFP $a = 0.0625$ and $MoU = 0.25$. In a likelihood based decision dependence between features in $2 \times 4$ CT is visible already for $n = 25$ and PFP $a = 0.0625$ (see figure 7). In the classic statistical testing related to $3 \times 3$ CT (see table 8) untrue $H_0$ has not been rejected even if $n = 90$, PFP $a = 0.0444$ and $MoU = 0.237$. In a likelihood based decision dependence between features in $3 \times 3$ CT is visible already for $n = 30$ and PFP $a < 0.0444$ (see figure 8). A similar situation occurs related to $3 \times 4$ and $4 \times 4$ CTs.

8.2. Three-way contingency table

Example 4. This example compares decision making by means of classic statistic testing with likelihood based decision making. Tables 11–12 show a set of example three-way CTs filled one by one accordingly to the scenarios XXI–XXVIII. The description of these tables has been presented in the example 1. To decide which of the defined scenarios takes place, see algorithm in section 6.2.

Tables 11–12 show that all the decisions made in a classic way are wrong under the scenarios in question. It is because untrue $H_0$ hypotheses have not been rejected even in situations where the MoU does not have such small values, e.g. $MoU = 0.267$ in scenario XXVIII. In turn, the parametric approach detected a dependency between features.

Example 5. This example is very similar to the Example 2. The new proposal is a method of statistical inference, not a classical parametric test. All the hypotheses state: “the considered three-way CT is generated accordingly to particular scenarios”. Each of figures from 11 to 12 shows a sets of four likelihood curves. Each curve has its four attributes, namely $\hat{a}$, $n$, table dimensions and, of course, the actual generation scenario that is specified in the figure’s title. The legend lists all competing scenarios including the actual one.
Table 11. THE CLASSICAL STATISTICAL TESTING VERSUS LIKELIHOOD BASED DECISIONS, TABLE 2 × 2 × 2

| Scenario (MoU) | Content | $X$ and $Y$ dependent | Scenario of maximum likelihood | $\chi^2$ | $FT$ | $CR$ | $|\chi|$ | $H_0$ true | $H_0$ rejected |
|---------------|---------|-----------------------|--------------------------------|--------|-----|------|-------|------------|-------------|
| XXI(0.130)    | 5 9 5 14 | Yes XXI               | 1.505                          | 1.469  | 1.496 | 0.979 | No    | No         |
|               | 9 15 9 14 |                       |                                |        |      |       |       |            |             |
| XXII(0.168)   | 5 5 5 10  | Yes XXII              | 1.669                          | 1.661  | 1.665 | 1.088 | No    | No         |
|               | 10 15 15 15 |                |                                |        |      |       |       |            |             |
| XXIII(0.188)  | 5 10 10 10 | Yes XXIII             | 3.810                          | 4.104  | 3.855 | 1.524 | No    | No         |
|               | 15 10 10 10 |                           |                                |        |      |       |       |            |             |
| XXIV(0.250)   | 5 10 5 10  | Yes XXIV              | 5.333                          | 5.482  | 5.350 | 2.133 | No    | No         |
|               | 15 10 15 10 |                          |                                |        |      |       |       |            |             |

Source: own elaboration.
Table 12. THE CLASSIC STATISTICAL TESTING VERSUS LIKELIHOOD BASED DECISIONS, TABLE 3 × 2 × 2

Experiment settings: $n = 120$, $\alpha = 0.0333$, $\alpha = 0.05$

| Scenario (MoU) | Content | $X$ and $Y$ dependent | Scenario of maximum likelihood | $\chi^2$ | $FT$ | $CR$ | $|\chi|$ | $H_0$ true | $H_0$ rejected |
|----------------|---------|------------------------|-------------------------------|---------|------|------|-------|-------------|---------------|
| XXV(0.089)     | 6 10 6 10 | Yes                    | XXV                           | 13.989  | 15.069 | 14.043 | 3.489 | No          | No            |
|                | 6 10 10 10 |                       |                               |         |       |       |       |             |               |
|                | 10 10 6 14 |                       |                               |         |       |       |       |             |               |
| XXVI(0.149)    | 6 10 6 14  | Yes                    | XXVI                          | 3.697   | 3.629 | 3.674 | 1.815 | No          | No            |
|                | 6 10 14 10 |                       |                               |         |       |       |       |             |               |
| XXVII(0.200)   | 6 10 6 14  | Yes                    | XXVII                         | 5.000   | 5.078 | 5.009 | 2.500 | No          | No            |
|                | 6 10 14 10 |                       |                               |         |       |       |       |             |               |
| XXVIII(0.267)  | 6 10 6 14  | Yes                    | XXVIII                        | 12.800  | 13.500 | 12.879 | 3.200 | No          | No            |
|                | 10 10 14 6 |                       |                               |         |       |       |       |             |               |

Source: own elaboration.
Figure 11. The likelihood functions for $d = 0.0625$, $n = 80$, table $2 \times 2 \times 2$

Source: own elaboration.

Figure 12. The likelihood functions for $d = 0.0333$, $n = 120$, table $3 \times 2 \times 2$

Source: own elaboration.
It is noteworthy (see figures 11–12) that in each set of likelihood curves, the curve related to the actual scenario predominates over the others. It is instructive to read out a value of $a_{max}$ that maximizes likelihood function on a particular figure and notice that this value is close to assumed $\hat{a}$.

**Example 6.** This example is carried out in accordance with the following algorithm:
1. Find dimension of CT in question and read out related $q$ according to the rule:

   $$2 \times 2 \times 2 \ (q = 6), 3 \times 2 \times 2 \ (q = 7)$$

2. Set a set of scenario indices $\{4q - 3, 4q - 2, 4q - 1, 4q\}$.
3. Calculate a value of PFP by means of $a_i = 01\i/(wkp) \ (i = 1, 2, ..., 10)$.

   Steps 4–6 are the same as in the example 3.

Figure 13–14 present PoR(a) functions for three-way CT in question. Sample sizes for a given CT are different because a maximal MoU values under the scenarios are different (see table 5). The minimal sample sizes are chosen in such a way that probabilities of proper recognition (actual XXI as XXI, ..., actual XXVIII as XXVIII) are greater than probabilities of improper recognitions for all the $a$ values.

![Figure 13](image-url)
Figures 13–14 show that even when samples are small (e.g. 30 items), probabilities of proper recognition are greater than probabilities of improper recognitions, regardless how small the PFP is. The bigger PFP $\alpha$, the bigger PoR actual scenario. In the classic statistical testing related to $2 \times 2 \times 2$ CT (see table 11) untrue $H_0$ has not been rejected even if $n = 80$, $PFP \ a = 0.0625$ and $MoU = 0.25$. In a likelihood based decision dependence between features in $2 \times 2 \times 2$ CT is visible already for $n = 30$ and $PFP \ a < 0.0625$ (see figure 13). In the classic statistical testing related to $3 \times 2 \times 2$ CT (see table 12) untrue $H_0$ has not been rejected even if $n = 120$, $PFP \ a = 0.0333$ and $MoU = 0.267$. In a likelihood based decision dependence between features in $3 \times 2 \times 2$ CT is visible already for $n = 30$ and $PFP \ a < 0.0333$ (see figure 14).

9. CONCLUSIONS

There are two new elements in the method of statistical reasoning from CTs presented in this paper. Firstly, CTs are parameterized with the probability flow parameters. Parametric reasoning turns out to be much more sensitive in revealing dependency between features than classic reasoning. Secondly, we suggest a scenario (i.e. internal mechanism) under which particular CT comes into being.
Figuring up more and more general scenarios does not seem very difficult. The researches can be generalized by introducing a part of flow parameter, e.g. $a/2, a/3, \ldots$ also remembering about the condition of normalization. The researches can also be generalized by introducing several flow parameters. This, however, causes a significant deterioration in the properties of the parameter estimators. You can always add more parameters to the model, however, this might worsen their estimation. Hence, inflated generalizations should be avoided.

REFERENCES


Oates T., Cohen P. R., (1996), Searching for Structure in Multiple Streams of Data, ICML 96.


WNIOSKOWANIE PARAMETRYCZNE I NIEPARAMETRYCZNE
W TABLICACH DWUDZIELCZYCH I TRÓJDZIELCZYCH

Streszczenie

W artykule proponowane są scenariusze generowania tablic dwudzielczych (TD) z parametrem przepływu prawdopodobieństwa i zdefiniowane są miary nieprawdziwości $H_0$. W artykule wykorzystywane są statystyki z rodziny $X^2$ oraz statystyka modułowa $|X|$. Niniejsza praca jest prostą próbą zastąpienia nieparametrycznej metody wnioskowania statystycznego metodą parametryczną. Metoda największej wiarygodności jest wykorzystana do oszacowania parametru przepływu prawdopodobieństwa. W pracy opisane są także instrukcje generowania TD za pomocą metody słupkowej. Symulacje komputerowe przeprowadzono metodami Monte Carlo.

Słowa kluczowe: wnioskowanie statystyczne, funkcja największej wiarygodności, tablice kontyngencji, test parametryczny, parametr przepływu prawdopodobieństwa

NONPARAMETRIC VERSUS PARAMETRIC REASONING BASED ON TWO-WAY AND THREE-WAY CONTINGENCY TABLES

Abstract

This paper proposes scenarios of generating two-way and three way contingency tables (CTs). A concept of probability flow parameter (PFP) plays a crucial role in these scenarios. Additionally, measures of untruthfulness of $H_0$ are defined. The power divergence statistics and the $|X|$ statistics are used. This paper is a simple attempt to replace a nonparametric statistical inference from CTs by the parametric one. Maximum likelihood method is applied to estimate PFP and instructions of generating CTs according to scenarios in question are presented. The Monte Carlo method is used to carry out computer simulations.

Keywords: statistical inference, likelihood function, contingency tables, parametric test, probability flow parameter