Nonparametric Versus Parametric Reasoning Based on 2×2 Contingency Tables

Piotr Sulewski

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

Abstract: This paper proposes scenarios of generating contingency tables (CTs) with the probability flow parameter (PFP). It also defines measures of untruthfulness of H_0 that involve PFP for all proposed scenarios. This paper is an attempt to replace a nonparametric statistical inference method by the parametric one. The paper applies the maximum likelihood method to estimate PFP and presents instructions to generate CTs by means of the bar method. The Monte Carlo method is used to carry out computer simulations.

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

I. INTRODUCTION

What we are focusing on is a particular phenomenon which - if occurs - causes the creation of one new object that belongs to a set called the general population. Each member of the general population has two features, namely *X* and *Y*, with two levels denoted by X_1, X_2 and Y_1, Y_2, respectively. The 2×2 contingency table (CT) is a perfect mathematical expression of the effect of several consecutive phenomena. Each object randomly falls into one of four cells of CT. It takes place because there are some random factors that are inherent to the phenomenon. Let p_{ij} $(i, j = 1, 2)$ be probabilities of falling into a CT_{ij} cell. These probabilities are determined by the random factors. Thus, CT expressed in terms of exact probabilities of falling into cells takes the form

$$\text{CT}i = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}. \tag{1}$$

Unfortunately, exact values of probabilities $p_{ij}(i, j = 1, 2)$ are and will remain unknown to the investigator. The only way for investigators to proceed is to draw a random sample of n items from the general population and form an empirical CT. This CT includes either counts n_{ij}^* that are numbers of sample objects which have fallen into particular CT cells

$$\text{CT}c = \begin{bmatrix} n_{11}^* & n_{12}^* \\ n_{21}^* & n_{22}^* \end{bmatrix} \tag{2}$$

or the empirical CT composed of probabilities p_{ij}^*

$$\text{CT}e = \begin{bmatrix} p_{11}^* & p_{12}^* \\ p_{21}^* & p_{22}^* \end{bmatrix} = \begin{bmatrix} n_{11}^*/n & n_{12}^*/n \\ n_{21}^*/n & n_{22}^*/n \end{bmatrix}. \tag{3}$$

The information about categorical data can be found e.g. in [1–3]. Statistical reasoning based on CTs is one of the most common tasks performed by statisticians. CTs can be applied in a wide variety of areas such as social sciences [4], genetics [5, 6], demography [7] and psychology [8]. Basic methods of testing for dependency in CTs in details is described e.g. in [9–12]. Other examples of applications may be found e.g. in [13–16].

One can recognize two general cases in which CTs may 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 a 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 the 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, Y_2 levels as sensible indicators of X_1, X_2 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:

1. 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).
2. Let us again retrace a way of the classic thinking. When a value of the test statistics does not exceed the appropriate critical value, it is said that "there is no reason to reject H_0". In case A, the decision maker is comfortable about independence. When test statistics exceeds the appropriately determined critical value, it means nothing but "there is a reason to reject 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.

This paper is organized as follows. CT 2×2 coming into being are presented in Section 2. Section 3 is devoted to measures of untruthfulness of H_0 including the measure that is defined by means of an absolute value. These measures are introduced under the scenarios in question. Statistic tests including the power divergence tests and the $|\chi|$ test under scenarios are defined in Section 4. In Section 5 the maximum likelihood method is applied to estimate the PFP. Section 6 is devoted to instructions on how to generate 2×2 CTs. Section 7 presents numerical examples and Section 8 presents closing remarks.

Monte Carlo simulation is performed in Visual Basic for Applications embedded in Microsoft Excel 2016.

II. A MODEL HOW 2X2 CTS COME 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 *XY* combinations and ascribes these probabilities to the cells of the table. Below is a "progenitor" of all the 2×2 CTs [35]:

$$T_p = \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}.$$

A variety of tables may be generated when portions of probability *a* flow from "maternal" cells of *Tp* to other cells. Obviously, the total probability always equals 1. In this paper four scenarios that seem fundamental are developed (Tab. 1).

In all the above scenarios the PFP satisfies the inequalities $0 \le a \le 0.25$. The scenarios are selected in such a way that they correspond to four levels of dependence: none (scenario 0), weak (scenario A), medium (scenario B), strong (scenario C) expressed by means of an appropriate measure (see Section 3).

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. 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/her and instantly repeat such simulations. All examples presented herein have a very precise algorithmic description in a form of a step list.

The research can be generalized by introducing several PFPs. This, however, causes a significant deterioration of? 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 present author does not know any practical applications of

Tab. 1. The contents of 2×2 CTs resulting from scenarios in question. (Source: author's own material)

Scenario 0

	Y_1	Y_2
X_1	$0.25 - a$	$0.25 - a$
X_2	$0.25 + a$	$0.25 + a$

Scenario A

	Y_1	Y_2
X_1	$0.25 - a$	0.25
X_2	0.25	$0.25 + a$

Scenario B

	Y_1	Y_2
X_1	$0.25 - a$	0.25
X_2	$0.25 + a$	0.25

Scenario C

	Y_1	Y_2
X_1	$0.25 - a$	$0.25 + a$
X_2	$0.25 + a$	$0.25 - a$

URG to describe the reliability results of technical objects. You can always add more parameters to the model; however, this might worsen their estimation.

III. MEASURES OF UNTRUTHFULNESS OF H_0

As it has already been stated in Section 1, some classes of feature *X* are ascribed to rows and some classes of feature *Y* are ascribed to columns. Features *X*, *Y* are independent, which means that H_0 is true, if $p_{ij} = p_{i\bullet} \times p_{\bullet j}$ $(i, j = 1, 2)$ for each pair of i, j. When this equality 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. Tab. 2 presents a list of MoUHs with relevant calculation formulas tailored for 2×2 CTs .These are valid provided CTs come into being under scenarios A-C. These measures are i.e.: the Pearson's φ, the Tschuprow's *T*, the Cramer's *V*, the corrected contingency *c*, the Goodman and Kruskal's τ, the Yule's *Q* and the odds ratio OR. The Pearson χ^2 statistics for 2×2 CTs is defined as

$$\chi^2 = \frac{n\left(p_{11}^* p_{22}^* - p_{12}^* \times p_{21}^*\right)}{\left(p_{11}^* + p_{12}^*\right)\left(p_{21}^* + p_{22}^*\right)\left(p_{11}^* + p_{21}^*\right)\left(p_{12}^* + p_{22}^*\right)}. \tag{4}$$

Formulas in Tab. 2 allow to calculate the MoUHs values for a given value of the PFP *a* under scenarios A-C (see Tab. 1). Therefore, these measures are functions of this parameter.

In this paper we use a MoUH which is given by [35]:

$$\text{MoU} = \frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{2}\left|n_{ij}^* - \frac{n_{i\bullet}^* \times \times n_{\bullet j}^*}{n}\right| =$$
$$= \sum_{i=1}^{2}\sum_{j=1}^{2}\left|p_{ij}^* - p_{i\bullet}^* \times p_{\bullet j}^*\right|. \tag{5}$$

The $|Q|$ and OR measures under scenarios A-C have values in interval $\langle 0, 1 \rangle$. The MoU also has values in interval $\langle 0, 1 \rangle$ and is used in the Monte Carlo simulation. This measure, doubtlessly, springs from the essence of H_0 and has a very simple form. The MoU formulas, the maximal MoU values (the minimal MoU values are equal to zero) and levels of dependency under scenarios A-C are presented in Tab. 3. Appropriate graphs of these formulas are shown in Fig. 1. The MoU, just like the measures presented in Tab. 2, is a function of the PFP *a*. Thanks to this the MoU values are very easy to calculate.
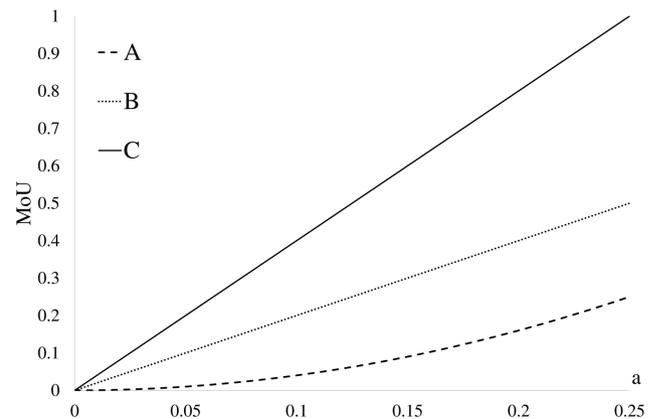

Fig. 1. The MoU under scenarios. (Source: author's own material)

IV. SELECTED TESTS OF INDEPENDENCE UNDER SCENARIOS

Statistical science has been enriched with many other statistics intended for research on/into test independency. Cressie

Tab. 2. MoUHs for 2×2 CTs under scenarios A-C, where χ^2 is given by (4). (Source: author's own material)

MoUH	General formula	Formula under scenario		
		A	B	C
φ, T, V	$\sqrt{\dfrac{\chi^2}{n}}$	$\dfrac{4a^2}{4a^2-1}$	$\sqrt{\dfrac{4a^2}{1-4a^2}}$	$4a$
c	$\sqrt{\dfrac{2\chi^2}{\chi^2+n}}$	$\sqrt{\dfrac{32a^4}{32a^4-8a^2+1}}$	$2a\sqrt{2}$	$\sqrt{\dfrac{32a^2}{16a^2+1}}$
τ	$\dfrac{\chi^2}{n}$	$\left(\dfrac{4a^2}{4a^2-1}\right)^2$	$\dfrac{4a^2}{1-4a^2}$	$16a^2$
Q	$\dfrac{p_{11}^*p_{22}^*-p_{12}^*p_{21}^*}{p_{11}^*p_{22}^*-p_{12}^*p_{21}^*}$	$\dfrac{8a^2}{8a^2-1}$	$-4a$	$\dfrac{-8a}{16a^2+1}$
OR	$\dfrac{p_{11}^*p_{22}^*}{p_{12}^*p_{21}^*}$	$1-16a^2$	$\dfrac{1-4a}{1+4a}$	$\left(\dfrac{4a-1}{4a+1}\right)^2$

and Read [18] propose the power divergence statistics (PDS). The PDS for 2×2 CTs is given by

$$P^2 = \frac{2}{\lambda(\lambda+1)}\sum_{i=1}^{2}\sum_{j=1}^{2}n_{ij}^*\left[\left(\frac{n_{ij}^*}{e_{ij}^*}\right)^\lambda - 1\right] =$$
$$= \frac{2n}{\lambda(\lambda+1)}\sum_{i=1}^{2}\sum_{j=1}^{2}p_{ij}^*\left[\left(\frac{p_{ij}^*}{p_{i\bullet}^* \times p_{\bullet j}^*}\right)^\lambda - 1\right], \quad (6)$$
$$-\infty < \lambda < \infty,$$

where $e_{ij} = n_{i\bullet}^* n_{\bullet j}^*/n = n \times p_{i\bullet}^* p_{\bullet j}^*$ are expected counts. Equation (6) 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 χ^2 statistics ($\lambda = 1$), the G^2 statistics (the limit as λ goes to 0), the Freeman-Tukey statistics ($\lambda = -0.5$), the modified G^2 statistics (the limit as λ goes

to -1), the Neyman modified χ^2 statistics ($\lambda = -2$) and the Cressie-Read statistics ($\lambda = 2/3$). If H_0 is true, statistics (6), for large n (i.e. asymptotically), follows the chi-square distribution with 1 degree of freedom.

The following PDS are selected to Monte Carlo study: the χ^2 statistics (7) [19], the Freeman-Tukey FT statistics (8) [20], the Cressie-Read *CR* statistics (9) [18]:

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij}^*-e_{ij})^2}{e_{ij}} = n\times\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(p_{ij}^*-p_{i\bullet}^* \times p_{\bullet j}^*)^2}{p_{i\bullet}^* \times p_{\bullet j}^*},$$
$$(7)$$

$$\text{FT} = 4\sum_{i=1}^{2}\sum_{j=1}^{2}\left(\sqrt{n_{ij}^*}-\sqrt{e_{ij}^*}\right)^2 =$$
$$= 4n\sum_{i=1}^{2}\sum_{j=1}^{2}\left(\sqrt{p_{ij}^*}-\sqrt{p_{i\bullet}^* p_{\bullet j}^*}\right)^2, \quad (8)$$

Tab. 3. The MoU under scenarios. (Source: author's own material)

Scenario	MoU	MoU$_{\max}$	Dependency
A	$4a^2$	0.25	Weak
B	$2a$	0.5	Medium
C	$4a$	1	Strong

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

$$= \frac{9n}{5} \sum_{i=1}^{2} \sum_{j=1}^{2} p_{ij}^{*} \left[\left(\frac{p_{ij}^{*}}{p_{i\bullet}^{*} p_{\bullet j}^{*}} \right)^{2/3} - 1 \right] . \qquad (9)$$

The G^2 statistics [21], the modified G^2 statistics [22] and the Neyman modified χ^2 statistics [23] have not been subjects in the Monte Carlo study because they are applicable only in a case where all n_{ij}^{*} $(i, j = 1, 2)$ counts are not equal to zero.

The square used in the numerator of χ^2 statistics (7) 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 χ^2 statistics and it has the form [24]

$$|\chi| = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\left| n_{ij}^{*} - e_{ij}^{*} \right|}{e_{ij}^{*}} = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\left| p_{ij}^{*} - p_{i\bullet}^{*} p_{\bullet j}^{*} \right|}{p_{i\bullet}^{*} p_{\bullet j}^{*}} . \quad (10)$$

V. APPLYING THE MAXIMUM LIKELIHOOD METHOD TO ESTIMATE THE PROBABILITY FLOW PARAMETER

This section is a simple attempt to replace a nonparametric statistical inference method by the parametric one. A maximum likelihood method is applied to estimate the PFP.

Let us remember that in the cells of CT there are values n_{ij} $(i, j = 1, 2)$, which are components of the multinomial distribution. Thus the multinomial distribution is taken as groundwork for 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 2. Let n_{ij}^{*} be the value of (i, j) cell and a is the PFP. Then likelihood functions are expressed as

$$L_0 (a) = D \times (0.25 - a)^{n_{11}^{*} + n_{12}^{*}}$$
$$\times (0.25 + a)^{n_{21}^{*} + n_{22}^{*}} , \qquad (11)$$

$$L_A (a) = D \times (0.25 - a)^{n_{11}^{*}}$$
$$\times (0.25)^{n_{12}^{*} + n_{21}^{*}} \times (0.25 + a)^{n_{22}^{*}} , \qquad (12)$$

$$L_B (a) = D \times (0.25 - a)^{n_{11}^{*}}$$
$$\times (0.25)^{n_{12}^{*} + n_{22}^{*}} \times (0.25 + a)^{n_{21}^{*}} , \qquad (13)$$

$$L_C (a) = D \times (0.25 - a)^{n_{11}^{*} + n_{22}^{*}}$$
$$\times (0.25 + a)^{n_{12}^{*} + n_{21}^{*}} . \qquad (14)$$

In formulas (11) - (14) $D = n! / (n_{11}^{*}! \times n_{12}^{*}! n_{21}^{*}! n_{22}^{*}!)$.

The logarithmic likelihood function under scenario 0 is given by

$$l_0 (a) = \ln L_0 (a) = \ln (D) + (n_{11}^{*} + n_{12}^{*}) \ln (0.25 - a)$$
$$+ (n_{21}^{*} + n_{22}^{*}) \ln (0.25 + a) .$$

Then

$$\frac{\partial l_0(a)}{\partial a} = \frac{n_{21}^{*} + n_{22}^{*}}{0.25 + a} - \frac{n_{11}^{*} + n_{12}^{*}}{0.25 - a} ,$$

$$\frac{\partial l_0(a)}{\partial a} = 0 \Rightarrow \hat{a} = \frac{n_{21}^{*} + n_{22}^{*} - (n_{11}^{*} + n_{12}^{*})}{4n} .$$

Let us check what kind of extremum we can find. As a result of a simple transformation we have

$$\frac{\partial^2 l(a)}{\partial a^2} = \frac{\partial}{\partial a} \left[\frac{n_{21}^{*} + n_{22}^{*}}{0.25 + a} - \frac{n_{11}^{*} + n_{12}^{*}}{0.25 - a} \right] =$$
$$= - \frac{n_{21}^{*} + n_{22}^{*}}{(0.25 + a)^2} - \frac{n_{11}^{*} + n_{12}^{*}}{(0.25 - a)^2} < 0 \qquad (15)$$

for all $a < 0.25$. 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 PFP. It may be proven that inequality (15) holds for all the scenarios considered in this paper.

Formulas for the maximum likelihood estimator of a under scenarios have the forms

$$\hat{a}_0 = \frac{n_{21}^{*} + n_{22}^{*} - n_{11}^{*} - n_{12}^{*}}{4n} , \quad \hat{a}_A = \frac{n_{22}^{*} - n_{11}^{*}}{4 (n_{11}^{*} + n_{22}^{*})} ,$$
$$\hat{a}_B = \frac{1}{4} \times \frac{n_{21}^{*} - n_{11}^{*}}{n_{11}^{*} + n_{21}^{*}} , \quad \hat{a}_C = \frac{n_{12}^{*} + n_{21}^{*} - n_{11}^{*} - n_{22}^{*}}{4n} . \qquad (16)$$

To decide which of the defined scenarios takes place, we follow the three steps below:

1. Calculate a^*, which is an estimate of parameter a for each scenario according to (16).
2. Calculate corresponding values of the maximum likelihood functions $L_0 (a^*)$, $L_A (a^*)$, $L_B (a^*)$, $L_C (a^*)$ by means of (11) – (14).
3. Choose a scenario for which the value $L (a^*)$ is the greatest.

VI. GENERATING 2×2 contingency tables

Generating CTs is very important in the simulation study. The approach in the literature for the generating two-way CTs is the Markov Chain Monte Carlo [25–29], the Sequential Importance Sampling [28–31], the probabilistic divide-and-conquer technique [32], the Generalized Gamma Distribution [33], the bar method [34].

In this paper we introduce an algorithm for generating 2×2 CT using the bar method. The bar method is identical to the method that generates random numbers that follow the multinomial distribution. It is as follows:

1. Choose a scenario.
2. Set a value of the PFP a.
3. Calculate probabilities p_{ij} $(i, j = 1, 2)$ based on Tab. 1.
4. Set label cells of 2×2 tables p_i $(i = 1, 2, 3)$ according to the rules

$$p_1 = p_{11}, \ p_2 = p_1 + p_{12}, \ p_3 = p_2 + p_{21}. \quad (17)$$

5. Set a sample size n.
6. Repeat the following steps n times:
   (a) Set initial values of cell count i.e. $n_{ij} = 0$ $(i, j = 1, 2)$.
   (b) Generate random number rw uniformly distributed within $(0, 1)$.

(c) Increase cell count of 2×2 tables according to the rule: $rw \leq p_1 \Rightarrow n_{11} = n_{11} + 1$, $p_1 < rw \leq p_2 \Rightarrow n_{12} = n_{12} + 1$, $p_2 < rw \leq p_3 \Rightarrow n_{21} = n_{21} + 1$, $rw > p_3 \Rightarrow n_{22} = n_{22} + 1$, where p_i $(i = 1, 2, 3)$ are given by (17).

VII. PARAMETRIC REASONING PUT INTO PRACTICE

Example 1

One thousand CTs are generated under each scenario (see Tab. 1) for $a = 0.125$ and a sample size $n = 10 \times i$ $(i = 2, 3, ..., 10)$. We calculate the maximum likelihood estimator of a by means of (16) for each CT. Fig. 2 presents the standard deviation of a^* as a function of $t = 1/\sqrt{n}$ under each scenario. This figure shows that under each scenario the standard deviation of a^* depends linearly on $1/\sqrt{n}$.

Example 2

The algorithm describing this example is as follows:
1. Choose a scenario from a set $\{0, A, B, C\}$.
2. Set a value of PFP $a \in \{0.125, 0.15\}$.
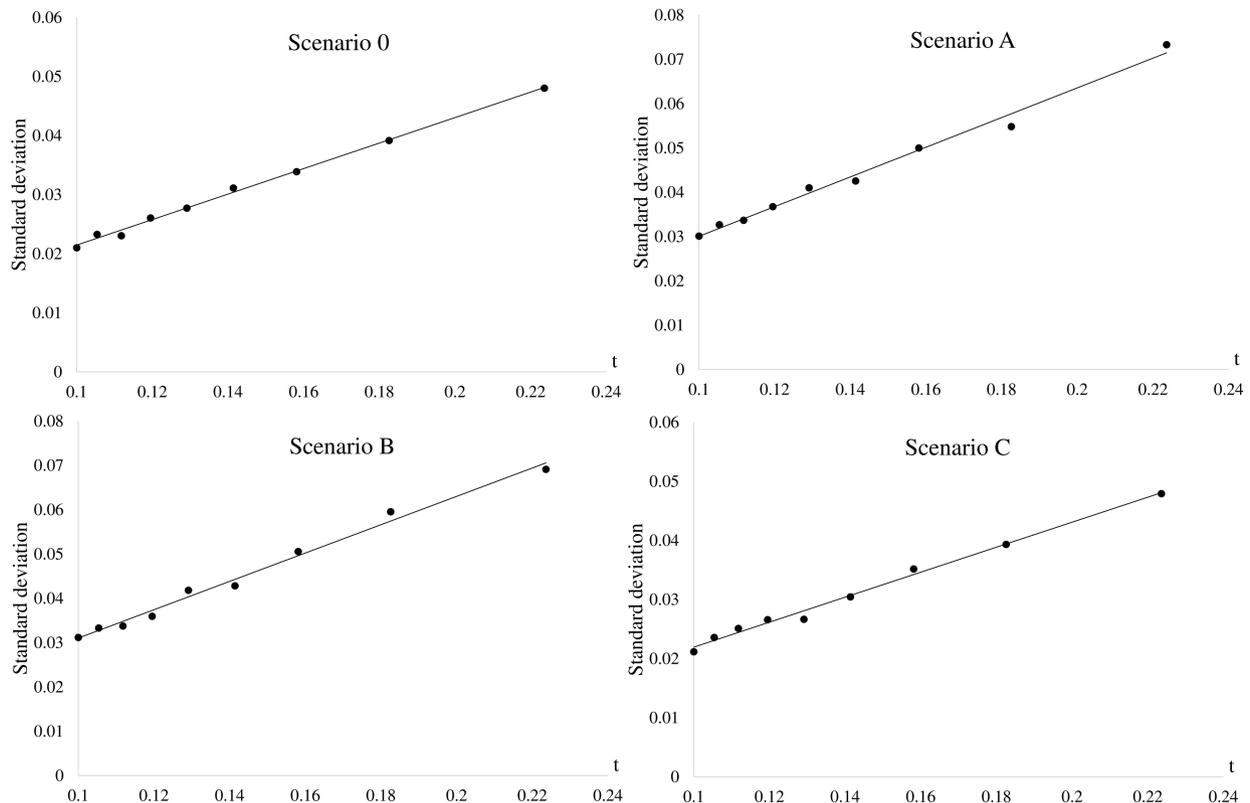3. Set a value of a sample size $n = 10 \times i$ $(i = 2, 3, ..., 10)$.

Fig. 2. The standard deviation of a^* as a function of $t = 1/\sqrt{n}$ under scenarios. (Source: author's own material)
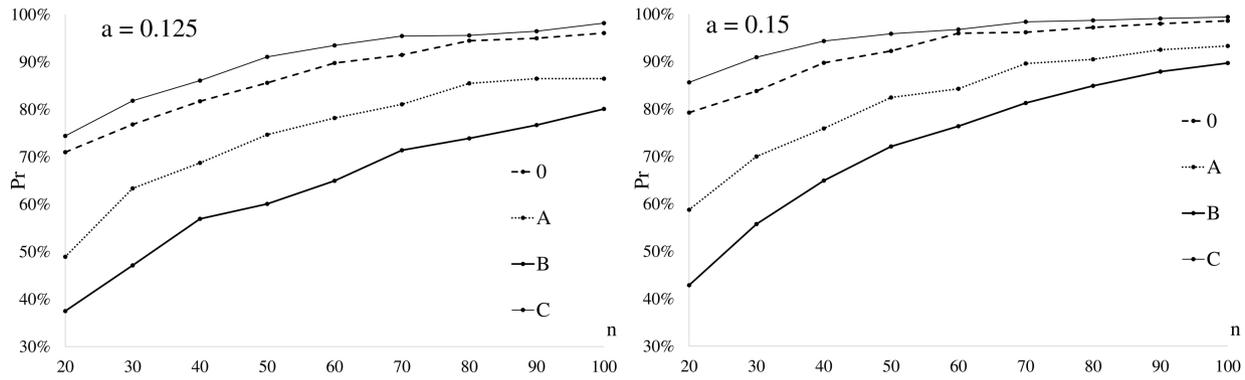
Fig. 3. Success ratio as a function of a simple size. (Source: author's own material)

4. Repeat the following steps $u = 10^4$ times:
   (a) Let $Counter = 0$.
   (b) Generate CT under the scenario that you have chosen in Step 1.
   (c) Calculate $a_0^*, a_A^*, a_B^*, a_C^*$ by means of (16).
   (d) Calculate $L_0(a_0^*), L_A(a_A^*), L_B(a_B^*), L_C(a_C^*)$ by means of (11) – (14).
   (e) If $MaxL = L_s(\hat{a}_s)$, where s is an index of scenario which has been chosen in Step 1, then $Success\_Counter = Success\_Counter + 1$.
5. Calculate the success ratio
   $\Pr = (Success\_Counter/u) \times 100\%$.

Fig. 3 presents the values of success ratio (in percent) under scenarios for a sample size $n = 10 \times i$ $(i = 2, 3, ..., 10)$ and $a \in \{0.125, 0.15\}$. This figure shows that the success ratio is the increasing function of a sample size n. It is noteworthy that the success ratio function related to scenario C predominates success ratio functions related to other scenarios.

Example 3

This example compares decision making by means of classic statistic testing with likelihood-based decision making. Tab. 4 shows a set of example CTs filled one by one accordingly to all the scenarios considered. The table is divided into two parts. The left hand side part is related to likelihood-based decisions. The right hand side part is related to classic statistical testing. The null hypothesis 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 guarantees very precise results. When reading rows A-C of Tab. 4 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 conservativeness of classic testing is a reason why we suggest making a turnaround in this domain. 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 CT is generated accordingly to particular scenarios".

Fig. 4 shows sets of associated likelihood functions for $a = 0.04$, $n = 100$ and explains why particular decisions are made. It is noteworthy (see Fig. 4) that in all the four sets of likelihood curves, the scenario related curve predominates over the others. You can read values that maximize likelihood functions from an appropriate figure or calculate them from appropriate formulas (16). It happens that the curves corresponding to scenarios overlap.

Example 4

The algorithm describing this example is as follows:
1. Choose a scenario from a set of scenarios $\{A, B, C\}$.
2. Set a value of PFP $a \in \{0.025, 0.05, ..., 0.25\}$.
3. Set a value of a sample size n.
4. Repeat the following steps $u = 10^4$ times:
   (a) Let $ScA = 0$, $ScB = 0$, $ScC = 0$
   (b) Generate CT under the scenario that you have chosen in Step 1.
   (c) Calculate $L_A(a), L_B(a), L_C(a)$ by means of (12) – (14).
   (d) If $MaxL = L_A(a)$, then $ScA = ScA + 1$;
       if $MaxL = L_B(a)$, then $ScB = ScB + 1$;
       if $MaxL = L_C(a)$, then $ScC = ScC + 1$.
5. Calculate the probability for actual scenario of recognizing generated scenario by means of formulas $\Pr_A = ScA/u$, $\Pr_B = ScB/u$, $\Pr_C = ScC/u$.

Fig. 5–8 present the probabilities of recognizing scenario A-C for PFP $a_i = 0.025 \times i$ $(i = 1, 2, ..., 10)$ and for MoU (see Tab. 3).

Fig. 5 and 7 show that even when samples are small (e.g. 20 items), probabilities of proper recognition (actual A as A, actual C as C) are greater than probabilities of improper recognitions, regardless of how small the PFP is. Fig. 6 con-

Tab. 4. The classic statistical testing versus likelihood-based decisions. (Source: author's own material)

Scenario of creation	Contents		X and Y dependent	Decision	χ^2	FT	CR	$|\chi|$	Ho true	Ho rejected
					3,962	4,010	3,967	0,800		
0	21	21	No	0	0	0	0	0	Yes	No
	29	29								
A	21	25		A	0,004	0,004	0,004	0,026	No	No
	25	29								
B	21	25	Yes	B	0,644	0,645	0,644	0,322	No	No
	29	25								
C	21	29		C	2,560	2,581	2,562	0,640	No	No
	29	21								

siderably contrasts those two. When PFP is small, curves interlace. Moreover, probabilities of improper recognition (B as A and B as C) are then greater that the probability of proper recognition! The picture normalizes when a sample becomes large as it is exemplified in Fig. 8. In the classic statistical testing (see Tab. 3) untrue H_0 has not been rejected even if $n = 100$ and PFP $a = 0.04$. In a likelihood-based decision dependence between features is visible already for $n = 20$ and PFP $a < 0.04$ (see Fig. 5–7).
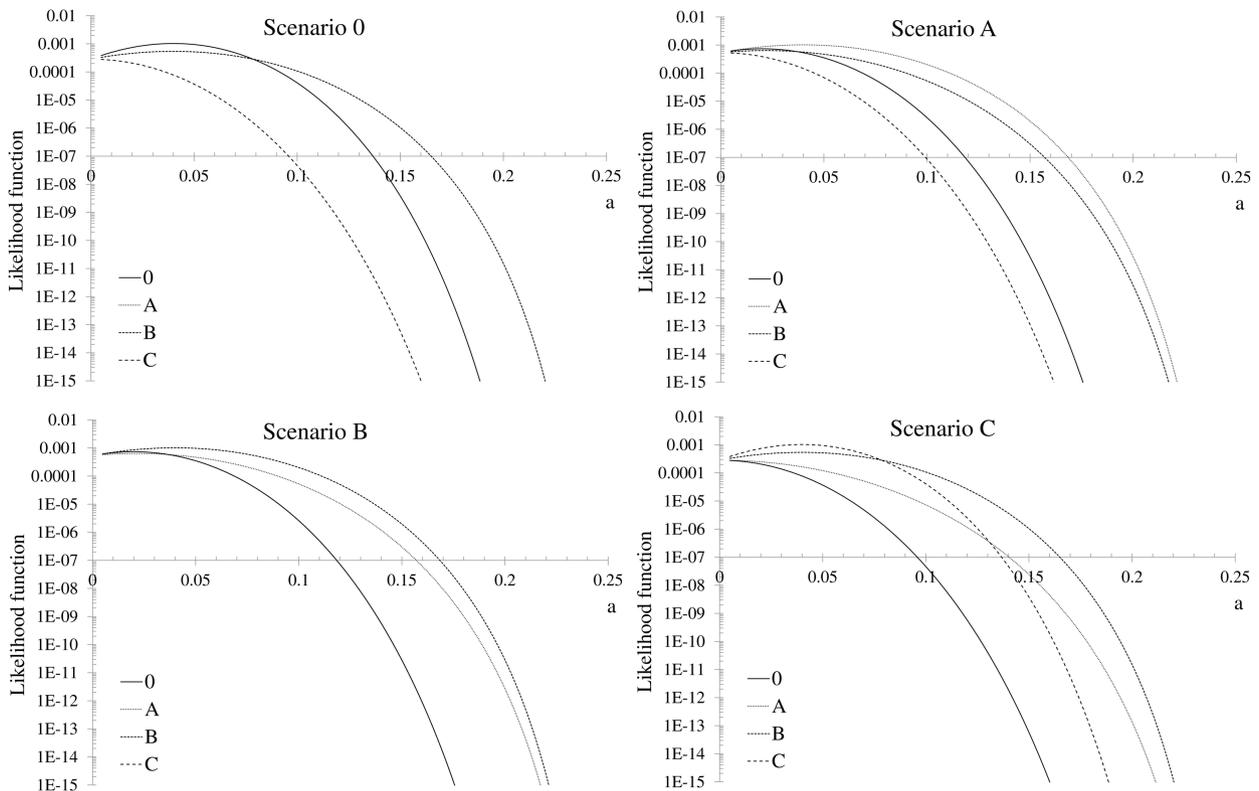

Fig. 4. The likelihood function under scenarios for $a = 0.04$ and $n = 100$. (Source: author's own material)
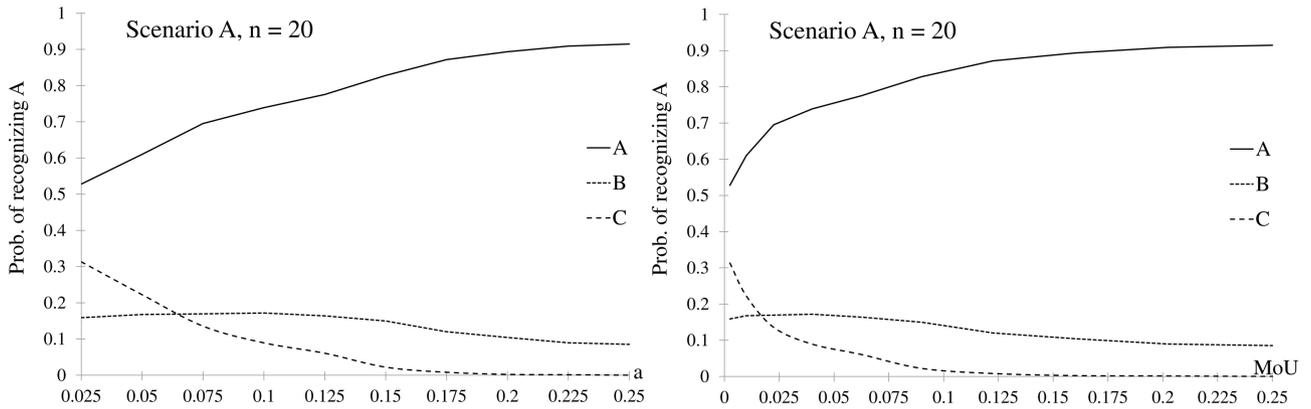
Fig. 5. The probability of recognizing scenario A (actual) as one of A-C scenarios. (Source: author's own material)
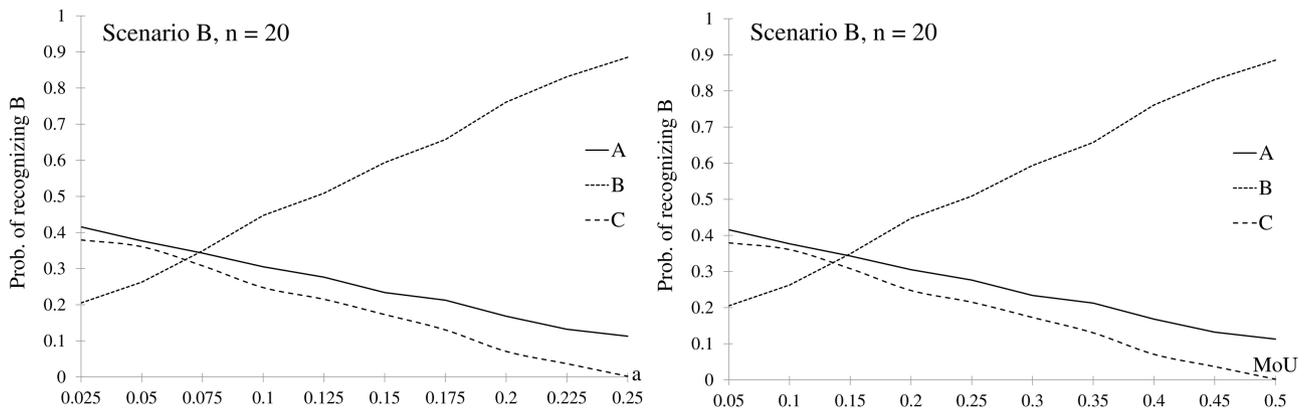

Fig. 6. The probability of recognizing scenario B (actual) as one of A-C scenarios. (Source: author's own material)
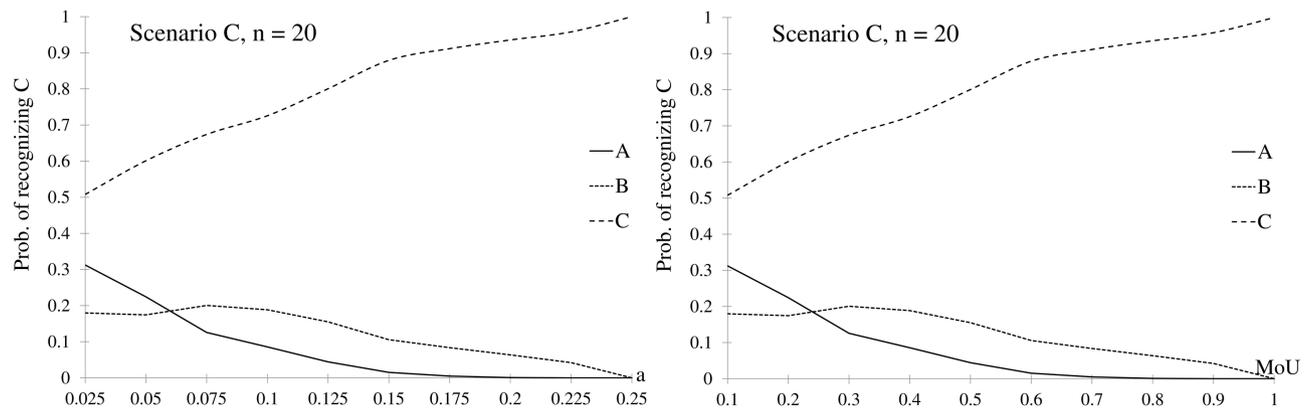

Fig. 7. The probability of recognizing C scenario (actual) as one of A-C scenarios. (Source: author's own material)
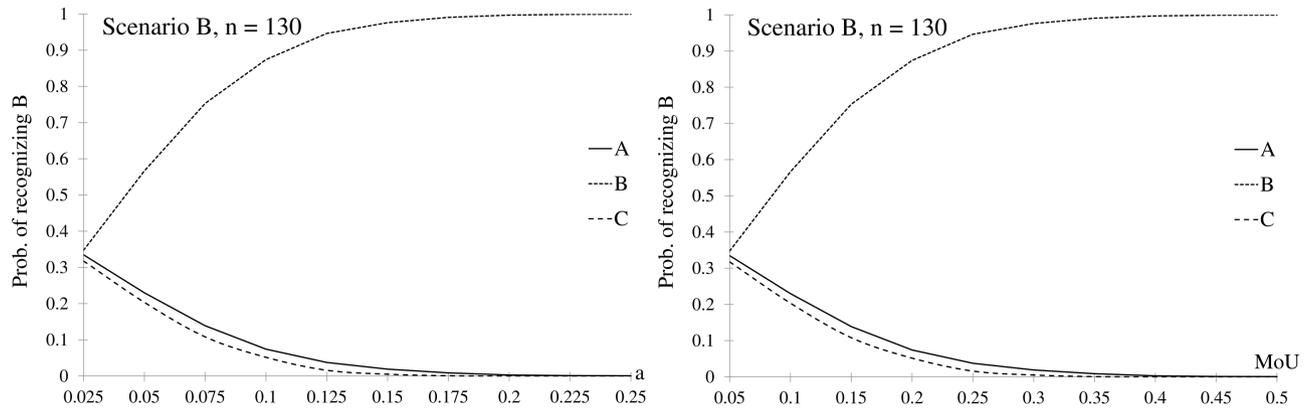
Fig. 8. The probability of recognizing B scenario (actual) as one of A-C scenarios. (Source: author's own material)

VIII. CLOSING REMARKS

There are two new elements in the method of statistical reasoning from 2×2 CTs presented in this paper. Firstly, 2×2 CTs are parameterized with the probability flow parameters. Parametric reasoning turns out to be much more sensitive in revealing dependency between X and Y features than classic reasoning. Secondly, we suggest a scenario (i.e. internal mechanism) under which particular 2×2 CT comes into being.

Thanks to precisely described algorithms it becomes possible to create and use two-way CTs bigger than 2×2 using appropriate scenarios, see e.g. [35]. Moreover, applying scenarios for generating the three-way CTs are presented in 36, 37].

Figuring up more and more general scenarios does not seem very difficult. The researches can be generalized by introducing several PFPs to the model. However, this causes a significant deterioration in the properties of the parameter estimators and in the estimation itself. Hence, inflated generalizations should be avoided.

Acknowledgments

References

[1] A. Agresti, *Categorical Data Analysis*, Wiley, New Jersey 2002.

[2] G. Van Belle, L.D. Fisher, P.J. Heagerty, T. Lumley, *Categorical Data: Contingency Tables*, John Wiley & Sons 2004.

[3] Y. Bishop, S. Fienberg, P. Holland, *Discrete Multivariate Analysis – Theory and Practice*, Cambridge, MA: MIT Press 1975.

[4] T.D. Wickens, *Multiway Contingency Tables Analysis for The Social Sciences*, Psychology Press 2014.

[5] T. Dickhaus, K. Straßburger, D. Schunk, C. Morcillo-Suarez, T. Illig, A. Navarro, *How to Analyze Many Contingency Tables Simultaneously in Genetic Association Studies*, Stat. Appl. Genet. Mol. Biol. **11**(4), 12 (2012).

[6] R. El Galta, T. Stijnen, J.J. Houwing-Duistermaat. *Testing for Genetic Association: A Powerful Score Test*, Statistics in Medicine **27**(22), 4596–4609 (2008).

[7] B. Cung, *Crime and Demographics: An Analysis of LAPD Crime Data*, UCLA Electronic Theses and Dissertations 2013.

[8] R. Iossifova, F. Marmolejo-Ramos, *When The Body Is Time: Spatial and Temporal Deixis in Children with Visual Impairments and Sighted Children*, Research in developmental disabilities **34**(7), 2173–2184, (2013).

[9] P.D. Allison, J.K. Liker, *Analyzing Sequential Categorical Data on Dyadic Interaction: A comment on Gottman*, Psychological Bulletin, **91**(2) 393–403 (1982).

[10] H.H. Bock, *Two-Way Clustering for Contingency Tables: Maximizing A Dependence Measure*, Between data science and applied data analysis, 143–154 (2003).

[11] S. Kaski, J. Nikkila, J. Sinkkonen, L. Lahti, J.E. Knuuttila, C. Roos, *Associative Clustering for Exploring Dependencies Between Functional Genomics Data Sets*, IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), **2**(3), 203–216 (2005).

[12] M. Steinle, K. Aberer, S. Girdzijauskas, C. Lovis, *Mapping Moving Landscapes by Mining Mountains of Logs: Novel Techniques for Dependency Model Generation*, [in:] Proceedings of the 32nd international conference on Very large data bases VLDB Endowment, pp. 1093–1102 (2006).

[13] P.J. Haas, F. Hueske, V. Markl, *Detecting Attribute Dependencies from Query Feedback*, [in:] Proceedings of the 33rd international conference on Very large data bases VLDB Endowment, pp. 830–841 (2007)

[14] I.F. Ilyas, V. Markl, P. Haas, P. Brown, A. Aboulnaga, *CORDS: Automatic Discovery of Correlations and Soft Functional Dependencies*, [in:] Proceedings of the 2004 ACM SIGMOD international conference on Management of data, ACM, pp. 647–658 (2004).

[15] T. Oates, P.R. Cohen, *Searching for Structure in Multiple Streams of Data*, [in:] ICML 96, pp. 346 (1996).

[16] M. Schrepp, *A Method for the Analysis of Hierarchical Dependencies Between Items of a Questionnaire*, Methods of Psychological Research Online **19**, 43–79 (2003).

[17] J. Blitzstein, P. Diaconis, *A Sequential Importance Sampling Algorithm for Generating Random Graphs with Prescribed Degrees*, Internet mathematics **6**(4), 489–522 (2011).

[18] N. Cressie, T.R. Read, *Multinomial Goodness-Of-Fit Tests*, Journal of the Royal Statistical Society. Series B (Methodological), 440–464 (1984).

[19] K. Pearson, *On the Criterion that a given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be Reasonably Supposed to have Arisen from Random Sampling*, Philosophy Magazine Series **50**, 157–172 (1900).

[20] M.F. Freeman, J.W. Tukey, *Transformations Related to the Angular and the Square Root*, Annals of Mathematical Statistics **21**, 607–611 (1950).

[21] R.R. Sokal, F.J. Rohlf, *Biometry: the Principles and Practice of Statistics in Biological Research*, Freeman, New York 2012.

[22] S. Kullback, *Information Theory and Statistics*, Wiley, New York 1959.

[23] J. Neyman, *Contribution to the Theory of the Chi-Square Test*, Proc. (First) Berkeley Symp. on Math. Statist. and Prob. (Univ. of Calif. Press), 239–273 (1949).

[24] P. Sulewski, *Modyfikacja testu niezależności (Modification of the Independence Test)*, Statistical News **10**, 1–19 (2013) (in Polish).

[25] P. Diaconis, B. Sturmfels, *Algebraic Algorithms for Sampling from Conditional Distributions*, The Annals of Statistics **26**(1), 363–397 (1998).

[26] M. Cryan, M. Dyer, *A Polynomial-Time Algorithm to Approximately Count Contingency Tables When the Number of Rows Is Constant*, Journal of Computer and System Sciences **67**(2), 291–310 (2003).

[27] M. Cryan, M. Dyer, L.A. Goldberg, M. Jerrum, R. Martin, *Rapidly Mixing Markov Chains for Sampling Contingency Tables with A Constant Number of Rows*, SIAM Journal on Computing **36**(1), 247–278 (2006).

[28] Y. Chen, P. Diaconis, S.P. Holmes, J.S. Liu. *Sequential Monte Carlo Methods for Statistical Analysis of Tables*, Journal of the American Statistical Association **100**(469), 109–120 (2005).

[29] G.S. Fishman, *Counting Contingency Tables Via Multistage Markov Chain Monte Carlo*, Journal of Computational and Graphical Statistics **21**(3), 713–738 (2012).

[30] Y. Chen, I.H. Dinwoodie, S. Sullivant *Sequential Importance Sampling for Multiway Tables*, The Annals of Statistics, 523–545 (2006).

[31] R. Yoshida, J. Xi, S. Wei, F. Zhou, D. Haws, *Semigroups and Sequential Importance Sampling for Multiway Tables*, arXiv preprint arXiv:1111.6518 (2011).

[32] S. DeSalvo, J.Y. Zhao, *Random Sampling of Contingency Tables Via Probabilistic Divide-And-Conquer*. arXiv preprint arXiv:1507.00070 (2015).

[33] P. Sulewski, *Two-By-Two Contingency Table as a Goodness-Of-Fit Test*, CMST **15**(2), 203–211 (2009).

[34] P. Sulewski, R. Motyka, *Power Analysis of Independence Testing for Contingency Tables*, Scientific Journal of Polish Naval Academy **56**, 37–46 (2015).

[35] P. Sulewski, *Moc testów niezależności w tablicy dwudzielczej większej niż 2x2 (Power Analysis of Independence Testing for Two-way Contingency Tables bigger than 2x2)*, Przegląd Statystyczny **63**(2), 191–209 (2016) (in Polish).

[36] P. Sulewski, *Moc testów niezależności w tablicy trójdzielczej 2x2x2 (Power Analysis of Independence Testing for Three-way Contingency Table 2x2x2)*, Przegląd Statystyczny **63**(4), 431–447 (2016) (in Polish).

[37] P. Sulewski, *Power Analysis of Independence Testing for the Three-way Contingency Tables of Small Sizes*, Journal of Applied Statistics, http://dx.doi.org/10.1080/02664763.2018.1424122 (2018).

Piotr Sulewski (www.sulewski.apsl.edu.pl) graduated in Mathematics in 1996. Since then he has been working at the Institute of Mathematics at Pomeranian Academy in Słupsk. He received the PhD in reliability theory in 2001 from the Systems Research Institute of Polish Academy of Sciences in Warsaw. His research interests concern mathematical statistics, computational methods in statistics and reliability mathematics. He is the author of 3 books and 28 publications.