

Bowker's Test for Symmetry and Modifications within the Algebraic Framework

Anne Krampe, Sonja Kuhnt

Fachbereich Statistik, Institut für Mathematische Statistik und industrielle
Anwendungen, Universität Dortmund, 44221 Dortmund, Germany

Abstract

Categorical data occur in a wide range of statistical applications. If the data are observed in matched pairs, it is often of interest to examine the differences between the responses. We concentrate on tests of axial symmetry in two-way tables. A commonly used procedure is the Bowker test which is a generalization of the McNemar test. The test decision is based on a χ^2 -approximation which might not be adequate, for example if the table is sparse. Therefore modifications of the test statistic have been proposed. We suggest a test of symmetry based on Bowker's test and Markov Chain Monte Carlo methods following the algorithm of Diaconis and Sturmfels (1998). We carry out a simulation study to determine and compare the performance of the simulation test, the Bowker test and two modifications.

Keywords: Computational commutative algebra, Diaconis-Sturmfels algorithm, matched-pairs data, MCMC, Metropolis-Hastings algorithm, test for symmetry.

1 Introduction

Categorical data occur in many different fields of applications and methods for the analysis of such data are necessary and useful, see e.g. Agresti (2002). In this article we consider data that consist of two dependent samples: each observation in one sample matches an observation of the other. According to Agresti (2002) we will call such data "matched-pairs data". As an example consider rater agreement studies. Suppose that two people, A and B , judge a sample of n subjects or objects into I different, predetermined categories. The resulting data is given in an $I \times I$ -contingency table. The cell entry n_{ij} is the number of subjects or objects that are classified into category i by rater A and to j by B , $i, j = 1, \dots, I$, see Rapallo (2002). Matched-pairs data occur also in genetics such as in associating-mapping

studies, e.g. to assess the role of candidate genes. Using the genome of individuals that are known to be affected with the disease of interest and of unaffected individuals, the counts of the corresponding alleles or genotypes are summarized in a two-dimensional contingency table, see Boehnke and Langefeld (1998) for details.

An interesting question is whether the contingency table is symmetric. For rater agreement studies this means to check if the number of subjects that are classified as (i, j) differ significantly from the number of individuals that are classified as (j, i) . For a 2×2 -table this corresponds to examine marginal homogeneity. McNemar (1947) provided a test for axial symmetry in this specific case. But often we have to deal not only with two but with I , $I > 2$, different categories. Bowker (1948) introduced a generalization of McNemar's test to check for symmetry in two-way tables. The decision is based on a χ^2 -approximation of the distribution of the test statistic. This might not be appropriate e.g. when the table is sparse. Edwards (1948) recommended to use a continuity correction. May and Johnson (2001) proposed a modified Wald statistic as an alternative to Bowker's test.

This article is organized as follows. In section 2 we describe tests for symmetry in a square contingency table and emphasize the need of an exact test for an $I \times I$ -table, $I > 2$. We suggest a test of symmetry based on Bowker's test using the results from computational commutative algebra in section 3. Diaconis and Sturmfels (1998) introduced an algorithm that establishes a relationship between computational commutative algebra and statistics. Using the Markov Chain Monte Carlo (MCMC) approach, the algorithm enables sampling from a conditional distribution of an exponential family given a sufficient statistic. In particular Gröbner bases are utilized to construct a Markov Chain. Hence computational commutative algebra provides a valuable alternative to traditional asymptotics for sufficiently large samples and exact inference for small samples. We describe the main concept of this algorithm as well as the used MCMC method to set up Bowker's test for symmetry within the algebraic framework. In addition we include two modifications of Bowker's test. We carry out a simulation study to determine and compare the performances of the approximate and the simulated tests in section 4. Furthermore we are interested in a rule of thumb for the adequacy of the approximation. In addition, we analyze data provided by the Research Network "Quiet Traffic".

2 Tests for symmetry in an $I \times I$ -table

Suppose the underlying data is represented in an $I \times I$ -contingency table. Let the cell counts n_{ij} come from random variables N_{ij} , $i, j = 1, \dots, I$, and assume that $\{N_{ij}\}$ follows a multinomial distribution with $\sum_i \sum_j N_{ij} = n$ and $N_{ij} \in \{0, 1, \dots, n\}$. We refer to π_{ij} as the probability of the occurrence of the event (i, j) , i.e. to be classified into category i by rater A and to j by B , respectively.

We start with the simple case of $I = 2$ different categories. The underlying data can be summarized in a 2×2 -table. One wants to know if there is an agreement in the classification of the subjects or objects. For a 2×2 -table this corresponds to testing for axial symmetry or equivalently for marginal homogeneity, see e.g. Agresti (2002) for details. The cells of interest are the cells that report a change of judgement, i.e. n_{12} and n_{21} . If the two raters agree about the classification of the n subjects, n_{12} and n_{21} are both equal to zero. In case of disagreement, these cell entries specify the different classifications. Thus we check whether the expected values of the corresponding cell entries differ significantly, i.e. we examine the hypothesis $H_0 : E(N_{12}) = E(N_{21})$ and the alternative $H_1 : E(N_{12}) \neq E(N_{21})$. McNemar (1947) introduced a test based on the usual χ^2 -test. The standardized sum of the squared differences between the cell entries N_{ij} and their expected values is used as a test statistic:

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(N_{ij} - E(N_{ij}))^2}{E(N_{ij})}.$$

Since we are testing for symmetry, the expected value of N_{ij} is $E(N_{ij}) = \frac{N_{ij} + N_{ji}}{2}$, $i, j = 1, 2$, and the test statistic becomes

$$\chi^2 = \frac{(N_{12} - N_{21})^2}{N_{12} + N_{21}}.$$

Under the hypothesis H_0 , χ^2 is approximately χ^2 -distributed with $\frac{1}{2} \cdot 2 \cdot (2 - 1) = 1$ degrees of freedom. But this approximation might not be adequate, e.g. when the table is sparse. As a rule of thumb, Agresti (2002) recommended a sample size of $n > 10$. If the approximation is inappropriate, we have to apply an exact test. Assuming H_0 , $N_{12} | N_{12} + N_{21}$ is binomially distributed with $P(N_{12} = n_{12}) = \frac{n_{12}}{n_{12} + n_{21}} = \frac{1}{2}$. Therefore the Binomial test is an intuitive choice for an exact test procedure, see e.g. Sheskin (2000) for a detailed description.

In general we consider I , $I > 2$, different categories. Thus the data can be presented in an $I \times I$ -table. We are interested in axial symmetry since it is more informative

than marginal homogeneity, see e.g. May and Johnson (2001). Analogously to the simple case we test $H_0 : E(N_{ij}) = E(N_{ji})$ for all $i, j = 1, \dots, I$ versus $H_1 : E(N_{ij}) \neq E(N_{ji})$ for at least one pair (i, j) . The test statistic of the standardized sum of the squared differences between N_{ij} and $E(N_{ij}) = \frac{N_{ij} + N_{ji}}{2}$, $i, j = 1, 2, \dots, I$ can be simplified to

$$\chi^2 = \sum_{j=i+1}^I \sum_{i=1}^{I-1} \frac{(N_{ij} - N_{ji})^2}{N_{ij} + N_{ji}}.$$

Assuming H_0 , χ^2 is approximately χ^2 -distributed with $\frac{1}{2} \cdot I \cdot (I - 1)$ degrees of freedom, see e.g. Bowker (1948). Furthermore we consider two modifications of the Bowker test. Edwards (1948) suggested a continuity correction for the McNemar test which can be extended for $I \times I$ -tables, $I > 2$

$$\chi_{\text{korr}}^2 = \sum_{j=i+1}^I \sum_{i=1}^{I-1} \frac{(|N_{ij} - N_{ji}| - 1)^2}{N_{ij} + N_{ji}}.$$

Assuming the underlying table is symmetric, χ_{korr}^2 is approximately $\chi_{\frac{1}{2}I(I-1)}^2$ distributed. May and Johnson (2001) introduced an alternative to Bowker's test by modifying the Wald test for symmetry with test statistic $\chi_{\text{W}}^2 = \delta'V^{-1}\delta$, with $\delta = \frac{N_{ij} - N_{ji}}{n}$. In particular they suggested using the modified covariance matrix

$$V_{\text{mw}} = \frac{1}{n} \begin{pmatrix} \lambda_{12} - \delta_{12}^2 & \delta_{13}\delta_{12} & \cdots & \delta_{(I-1)I}\delta_{12} \\ \delta_{12}\delta_{13} & \lambda_{13} - \delta_{13}^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \delta_{(I-1)I}\delta_{(I-2)(I-1)} \\ \delta_{12}\delta_{(I-1)I} & \cdots & \delta_{(I-2)(I-1)}\delta_{(I-1)I} & \lambda_{(I-1)I} - \delta_{(I-1)I}^2 \end{pmatrix},$$

with $\lambda_{ij} = \frac{N_{ij} + N_{ji}}{n}$. Thus the modified Wald test statistic becomes

$$\chi_{\text{mw}}^2 = \sum_{j=i+1}^I \sum_{i=1}^{I-1} \frac{n \cdot (N_{ij} - N_{ji})^2}{n \cdot (N_{ij} + N_{ji}) - (N_{ij} - N_{ji})^2}.$$

Supposing the considered $I \times I$ -table is symmetric, χ_{mw}^2 is approximately $\chi_{\frac{1}{2}I(I-1)}^2$ distributed.

It is an interesting question to determine and compare the performances of these tests. Connected with this is the inspection of the approximation. So far no rule of thumb seems to be available in the literature for the adequacy of the $\chi_{\frac{1}{2}I(I-1)}^2$ approximation. In addition we have not found an appropriate exact equivalent of the Bowker test for the general case of an $I \times I$ -table. In order to overcome this problem we suggest a test of symmetry based on the χ^2 -test and MCMC methods by using computational commutative algebra. Moreover we also employ this approach

for the continuity corrected test χ_{korrr}^2 as well as the modified Wald test χ_{mw}^2 to evaluate and to compare these tests of symmetry.

3 MCMC and algebra

Markov Chain Monte Carlo (MCMC) methods are a tool to sample from the distribution of interest, the "target distribution". For this purpose we construct a Markov Chain with stationary distribution proportional to the target distribution. We refer to Ewans and Grant (2001), Fahrmeir et al. (1981) and Sørensen and Gianola (2002) for an introduction to Markov Chain Monte Carlo methods. Diaconis and Sturmfels (1998) proposed an algorithm that establishes a relationship between computational commutative algebra and MCMC methods. We will briefly describe the main concept of the Diaconis-Sturmfels algorithm. For further details see e.g. Diaconis and Sturmfels (1998) or Rapallo (2003).

Let \mathcal{H} be a finite sample space. Densities that can be expressed as

$$P_{\theta}(X = x) = a(\theta)e^{\sum_{i=1}^d \theta_i T_i(x)} c(x), \quad x \in \mathcal{H}, \theta \in \Theta \subset \mathbb{R}^d$$

with normalizing constant $a(\theta)$, sufficient statistic $T := (T_1, \dots, T_d)'$, $T : \mathcal{H} \rightarrow \mathbb{N}^d$ and $c : \mathcal{H} \rightarrow \mathbb{N}$ belong to the exponential family.

In general we consider N random variables X_1, \dots, X_N . Suppose they are i.i.d. with density $P_{\theta}(X_k = x_k)$, $k = 1, \dots, N$. Thus their joint density is given by

$$\begin{aligned} P_{\theta}(X_1 = x_1, \dots, X_N = x_N) &= \prod_{k=1}^N a(\theta) \cdot e^{\sum_{i=1}^d \theta_i T_i(x_k)} \cdot c(x_k) \\ &= a(\theta)^N \cdot e^{\sum_{i=1}^d \theta_i \sum_{k=1}^N T_i(x_k)} \prod_{k=1}^N c(x_k). \end{aligned}$$

It can be shown that $\sum_{k=1}^N T(X_k)$ is a sufficient statistic for the parameter vector $\theta \in \Theta$ (Witting (1985), Korollar 3.20). We will be interested in the set of all data sets with the same value of the sufficient statistic $t = \sum_{k=1}^N T(x_k)$. Based on Diaconis Sturmfels (1998) we use the following notation:

$$\mathcal{Z}_t := \{z : \mathcal{H} \rightarrow \mathbb{N} \mid \sum_{x \in \mathcal{H}} z(x)T(x) = t\}.$$

If $z \in \mathcal{Z}_t$ every $z(x)$, $x \in \mathcal{H}$, is a data set whose sufficient statistic takes value t . Notice that \mathcal{Z}_t is finite, nonempty and that the probability function on \mathcal{Z}_t is hypergeometric, see Diaconis and Sturmfels (1998) and Rapallo (2003) for details.

In accordance with the literature we use t to denote the test statistic as well as its outcome. For illustration we will apply the theory to the special case we are interested in: testing symmetry in an $I \times I$ -contingency table. If the cell entries $\{N_{ij}\}$ are multinomially distributed, their joint density is given by

$$f(n_{11}, n_{12}, \dots, n_{II}) = \frac{n!}{n_{11}! n_{12}! \dots n_{II}!} \prod_{i=1}^I \prod_{j=1}^I p_{ij}^{n_{ij}}.$$

Under H_0 this density can be rewritten in terms of an $I \cdot (I - 1)$ parametric exponential family with sufficient statistic $t = (N_{ii}, i = 1, \dots, (I - 1); (N_{ij} + N_{ji}), j = (i + 1), \dots, I, i = 1, \dots, (I - 1))'$. \mathcal{Z}_t is the set of all $I \times I$ -tables with cell entries $n_{ij} \in \mathbb{N}$ and value t of the sufficient statistic given above. The finite sample space of such a table is $\mathcal{H} = \{(i, j), i, j \in \{1, \dots, I\}\}$. Supposing the underlying table is symmetric, $T((i, j))$ is a vector of length $(I - 1) + 2 \cdot \sum_{v=1}^{I-1} (I - v)$. A detailed description of $T((i, j))$ is given in the appendix. Assuming H_0 , the density function on \mathcal{Z}_t is hypergeometric:

$$H(z) = \frac{n!}{|\mathcal{Z}_t|} \prod_{x \in \mathcal{H}} \frac{1}{z(x)!},$$

see Diaconis and Sturmfels (1998).

To investigate the symmetry of the underlying $I \times I$ -table, we sample from $H(z)$. For this purpose we will apply the Metropolis-Hastings algorithm as a MCMC method. We will briefly introduce the main concept of this algorithm, a detailed description can be found e.g. in Chib and Greenberg (1995). Subsequently we will combine computational commutative algebra and the Metropolis-Hastings algorithm by using the Diaconis-Sturmfels algorithm.

Metropolis-Hastings algorithm

The procedure of the Metropolis-Hastings algorithm can be divided into two parts:

- i) generate a candidate as a potential new state of the Markov Chain and
- ii) accept or reject the proposed candidate.

Let π denote the invariant distribution of the Markov Chain with density π^* , E the state space of the Markov Chain and $q(r, s)$, $r, s \in E$ the proposal distribution (candidate generating distribution) with $\int q(r, s) ds = 1$. Assuming the Markov Chain is presently in state r , state s is proposed in the next step with probability $q(r, s)$. In this setting it might happen that the Markov Chain moves more often

in one direction, i.e. from r to s than vice versa. In order to adjust this potential imbalance, the acceptance probability α is introduced. Thus the chain moves from r to s , $r \neq s$, $r, s \in E$ with probability $p(r, s) = q(r, s) \cdot \alpha(r, s)$. Using the reversibility condition $\pi^*(r)p(r, s) = \pi^*(s)p(s, r)$ it can be shown that

$$\alpha(r, s) = \begin{cases} \min\left(\frac{\pi^*(s)q(s,r)}{\pi^*(r)q(r,s)}, 1\right), & \text{if } \pi^*(r)q(r, s) > 0 \\ 1, & \text{otherwise,} \end{cases}$$

see Chib and Greenberg (1995) for details. The generated Markov Chain is reversible due to the design of this algorithm. If also some mild regularity conditions (aperiodicity and irreducibility) are fulfilled, π^* is the invariant density of the Markov Chain. Recall that in our case π^* equals H , the density function on \mathcal{Z}_t . The difficult task now is to find a suitable proposal distribution $q(\cdot, \cdot)$. For this purpose Diaconis and Sturmfels (1998) introduced the Markov Basis. A Markov Basis is a set of functions $m_1, m_2, \dots, m_L : \mathcal{H} \rightarrow \mathbb{Z}$, called moves, such that

$$\text{i) } \sum_{x \in \mathcal{H}} m_i(x)T(x) = 0 \quad \text{for all } 1 \leq i \leq L \quad \text{and}$$

ii) for any t and $z, z' \in \mathcal{Z}_t$ there is a sequence of moves $(m_{i_1}, \dots, m_{i_A})$ as well as a sequence of directions $(\epsilon_1, \dots, \epsilon_A)$ with $\epsilon_j = \pm 1$ such that

$$z' = z + \sum_{j=1}^A \epsilon_j m_{i_j} \quad \text{and} \quad z + \sum_{j=1}^a \epsilon_j m_{i_j} \geq 0 \quad 1 \leq a \leq A.$$

Both conditions ensure that the sufficient statistic t remains the same for the new state z' . The second constraint also guarantees the irreducibility of the chain. Using this definition, we can set up an appropriate Markov Chain for the hypergeometric probability function H on \mathcal{Z}_t :

Construction of the Markov Chain

Suppose a Markov Basis m_1, \dots, m_L is given. Select a move m_U by choosing U uniformly in $\{1, \dots, L\}$ and the direction of the move $\epsilon = \pm 1$ with probability $\frac{1}{2}$ independently of U . Therefore the proposal distribution $q(\cdot, \cdot)$ is symmetric and corresponds to a random walk. If the chain is currently in state $z \in \mathcal{Z}_t$, it moves to $z' = z + \epsilon m_U \in \mathcal{Z}_t$ with probability

$$\alpha = \min \left(\frac{\prod_{x \in \mathcal{H}} z(x)}{\prod_{x \in \mathcal{H}} (z(x) + \epsilon m_U(x))}, 1 \right),$$

see Rapallo (2003) for a detailed derivation. If the proposed new state z' is not element of \mathcal{Z}_t , e.g. if an entry of z' is negative, the chain stays at z .

In this manner we get an aperiodic, irreducible, reversible Markov Chain on \mathcal{L}_t with stationary probability function proportional to H , see Diaconis and Sturmfels (1998) for a proof and a brief discussion for the rates of convergence. Thus the problem to ascertain the proposal distribution is reduced to propose an appropriate Markov Basis. Diaconis and Sturmfels (1998) suggested to apply results from computational commutative algebra to solve this problem. In particular they showed that the Markov Basis is equal to the reduced Gröbner Basis of an ideal \mathcal{I}_T which will be specified later. In the following we will briefly describe the algebraic background and refer to Cox et al. (1992) and Pistone et al. (2001) for an introduction to computational commutative algebra.

Computational commutative algebra

Recall that \mathcal{H} is a finite set. For each element $x \in \mathcal{H}$ we introduce an indeterminate also denoted x . The basis for our consideration is $k[\mathcal{H}]$, the ring of polynomials in $x \in \mathcal{H}$. Note that any function $g : \mathcal{H} \rightarrow \mathbb{N}$ can be represented as a monomial $\prod_{x \in \mathcal{H}} x^{g(x)}$. In general a monomial involves more than one indeterminate. To achieve comparability of monomials, we introduce a monomial ordering which will be symbolized by \succ , see e.g. Cox et al. (1992) § 2.2.

A function $T := (T_1, \dots, T_d)'$, $T : \mathcal{H} \rightarrow \mathbb{N}^d \setminus \{0\}$ can be described by the homomorphism

$$\begin{aligned} \varphi_T : k[\mathcal{H}] &\rightarrow k[t_1, \dots, t_d] \\ x &\rightarrow t_1^{T_1(x)} t_2^{T_2(x)} \dots t_d^{T_d(x)}, \end{aligned}$$

see e.g. Rapallo (2003) for details. We concentrate our further study on the kernel of this homomorphism which is called an ideal, $\mathcal{I}_T = \{p \in k[\mathcal{H}] : \varphi_T(p) = 0\}$. Diaconis and Sturmfels (1998) provided a link between a Markov Basis and computational commutative algebra. Using that each function $m : \mathcal{H} \rightarrow \mathbb{Z}$ can be represented by the difference $m(x) = m^+(x) - m^-(x)$ with $m^+(x) := \max(m(x), 0)$ and $m^-(x) := \max(-m(x), 0)$, they have shown that a Markov Basis corresponds to a basis of the ideal \mathcal{I}_T .

Theorem

A set of functions m_1, \dots, m_L is a Markov Basis iff the set

$$\mathcal{H}^{m^+} - \mathcal{H}^{m^-}, \quad 1 \leq i \leq L$$

generates the ideal \mathcal{I}_T , where $m^+ m^- : \mathcal{H} \rightarrow \mathbb{N}$ with $m^+(x)$ and $m^-(x)$ as given above.

Proof: Diaconis and Sturmfels (1998), Theorem 3.1.

According to the Hilbert Basis Theorem (see e.g. Cox et al. (1992), pg. 75), every ideal \mathcal{I} in a polynomial ring has a finite generating set called basis. Hence we can restate our principle task: in order to define an appropriate proposal distribution we look for a Markov Basis or equivalently for a basis of the ideal $\mathcal{I}_T \subset k[x_1, \dots, x_n]$, $x_1, \dots, x_n \in \mathcal{H}$. But unlike the concept of a basis in linear algebra, an ideal in a polynomial ring can have more than one basis. Therefore we seek for a unique basis that exists for every ideal: the reduced Gröbner Basis, see Cox et al. (1992), pp. 76 for details. Using the implicitation algorithm, Diaconis and Sturmfels (1998) have specified a procedure to determine this Gröbner Basis.

Diaconis-Sturmfels algorithm

Assume that \mathcal{H} is a finite set and $\mathcal{T} = \{t_1, \dots, t_d\}$. Let $T : \mathcal{H} \rightarrow \mathbb{N}^d$ with $T = (T_1, \dots, T_d)'$ be given as well as a monomial ordering \succ for \mathcal{H} . This ordering will be extended for $\mathcal{H} \cup \mathcal{T}$ such that $t \succ x$ for all $x \in \mathcal{H}$ and $t \in \mathcal{T}$ in $k[\mathcal{H}, \mathcal{T}]$. Basically Diaconis and Sturmfels create an ancillary ideal $\mathcal{I}^* = \{x - \mathcal{T}^{T(x)}, x \in \mathcal{H}\}$ with $\mathcal{T}^{T(x)} := t_1^{T_1(x)} \cdot t_2^{T_2(x)} \cdot \dots \cdot t_d^{T_d(x)}$ and compute the reduced Gröbner Basis G^* for \mathcal{I}^* . Since $\mathcal{I}_T = \mathcal{I}^* \cap k[\mathcal{H}]$, the reduced Gröbner Basis G for \mathcal{I}_T contains those polynomials of G^* which only involve \mathcal{H} , see Diaconis and Sturmfels (1998), Theorem 3.1 and Cox et al. (1992), § 3.1, § 3.3 for details.

4 Simulation study and data example

We carry out a simulation study to examine the use of the Diaconis-Sturmfels algorithm for analyzing matched-pairs data. In particular we compare the performance of the symmetry tests introduced in section 2 for 5×5 -tables, i.e. we test $H_0 : E(N_{ij}) = E(N_{ji})$ against $H_1 : E(N_{ij}) \neq E(N_{ji})$ for at least one pair (i, j) , $i, j \in \{1, \dots, 5\}$ with the Bowker test, the modified Wald test and the continuity corrected test. Related to that question we assess the adequacy of the χ^2 -approximation and propose an alternative test based on Bowker's test and computational commutative algebra. We construct a Markov Chain based on the reduced Gröbner Basis conform to the constraints of the symmetry tests. We focus

on sparse tables to learn about the applicability of the approximation with the χ^2 -distribution with $\frac{1}{2} \cdot 5 \cdot (5 - 1) = 10$ degrees of freedom.

Consider again the finite sample space $\mathcal{H} = \{(i, j), 1 \leq i, j \leq 5\}$. Assuming the table is symmetric, it can be shown that the sufficient statistic is given by $(N_{ii}, i = 1, \dots, 5 - 1; ((N_{ij} + N_{ji}), j = (i + 1), \dots, 5), i = 1, \dots, 5 - 1)'$. Rewriting the the sufficient statistic in terms of \mathcal{Z}_t , the set of all data sets whose sufficient statistic takes value $t = \sum_{x \in \mathcal{H}} z(x)T(x)$, the corresponding $T((i, j))$, $i, j \in \{1, \dots, I\}$ is a vector of length 24. We refer to the appendix for a detailed description of $T((i, j))$.

The design of our simulation study is as follows: for each data set we generate a Markov Chain with 500000 states, i.e. 5×5 -contingency tables that are element of the corresponding \mathcal{Z}_t . To ensure that the chain is independent of the original table, we disregard the first 100000 tables in the so called burn-in-phase. We calculate the values of the three test statistics for the underlying contingency table and compare them to the corresponding values of each 100th table for a test decision. Diaconis and Sturmfels (1998) considered briefly the rate of convergence for a two-dimensional table. A more detailed investigation can be found in Diaconis and Saloff-Coste (1995).

We present the analysis of 5×5 -tables with specifically interesting results and refer to the appendix for the examination of additional tables. Firstly we consider a case where more than 20% of the expected values $E(N_{ij}), i \neq j, i, j \in \{1, \dots, 5\}$, are smaller than 5. This is in accordance with a common rule of thumb for the adequacy of the χ^2 -test of independence (see e.g. Agresti (2002)). It turns out that even if 50% of the expected values $E(N_{ij}), i \neq j$, are less than 5, the approximation is reasonable. For illustration we consider a data set given in table 1.

$$\begin{pmatrix} 35 & 4 & 6 & 4 & 7 \\ 2 & 47 & 3 & 8 & 2 \\ 4 & 5 & 25 & 3 & 7 \\ 5 & 2 & 3 & 23 & 3 \\ 3 & 6 & 5 & 8 & 11 \end{pmatrix}$$

Table 1: Data set 1; 10 expected values $E(N_{ij}), i \neq j$, are smaller than 5

We carry out the approximate and the corresponding simulated tests of symmetry.

To get an idea of the adequacy of the χ^2 -approximation we display our results in figure 1. The simulated values are characterized by the histogram, the χ^2_{10} -density is specified by the red line.

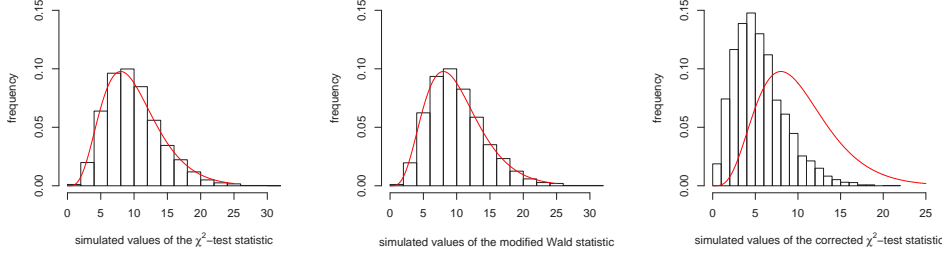


Figure 1: Histograms of the simulated values of the three symmetry test statistics for data set 1

The approximation for the Bowker test and the modified Wald test almost coincides with our simulated results. But the approximation for the continuity corrected test is not suitable (too conservative). Inspecting the values of the test statistic as well as the approximate and simulated p-values for this data set in table 2 leads to an analogous conclusion. We get almost the same approximate and simulated p-values for the Bowker test ($p_{\text{approx}} = 0.321$, $p_{\text{simu}} = 0.305$) and the modified Wald test ($p_{\text{approx}} = 0.313$, $p_{\text{simu}} = 0.307$). Hence we cannot reject the hypothesis of symmetry at level $\alpha = 0.05$ for these tests. For the continuity corrected test we get a different test decision for the simulated and approximate test, i.e. the p-value for the simulated test is $p_{\text{simu}} = 0.04$ but the approximate p-value is $p_{\text{approx}} = 0.304$. Thus we reject H_0 at level $\alpha = 0.05$ for the simulated test in contrast to the approximate test.

| | test statistic | approximate p-value | simulated p-value |
|-----------------------|----------------|---------------------|-------------------|
| Bowker test | 11.484 | 0.321 | 0.305 |
| modified Wald test | 11.596 | 0.313 | 0.307 |
| continuity correction | 11.717 | 0.304 | 0.040 |

Table 2: simulated and approximate results for data set 1

In the next step we analyzed a data set whose expected values $E(N_{ij})$, $i \neq j$, are all less than 5. The data are given in table 3.

As for data set 1, the results are diagramed in figure 2 to get an impression for the adequacy of the χ^2_{10} -approximation. Even in this case, the approximation is appropriate for the Bowker test and the modified Wald test. And again, the

$$\begin{pmatrix} 35 & 4 & 2 & 4 & 5 \\ 2 & 47 & 3 & 5 & 2 \\ 4 & 5 & 25 & 3 & 3 \\ 5 & 2 & 3 & 23 & 1 \\ 3 & 6 & 5 & 8 & 11 \end{pmatrix}$$

Table 3: data set 2; all expected values $E(N_{ij}), i \neq j$, are smaller than 5

χ^2_{10} -approximation for the continuity corrected test does not seem to be sensible at all. Looking at the values of the test statistics as well as the approximate and simulated p-values in table 4 we get results that are consistent with the results for data set 1.

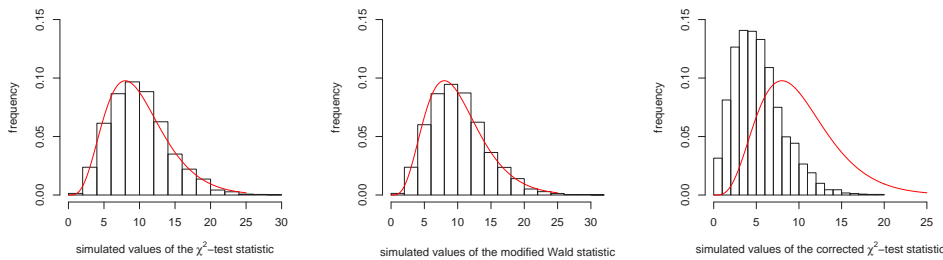


Figure 2: Histograms of the simulated values of the three symmetry test statistics for data set 2

| | test statistic | approximate p-value | simulated p-value |
|-----------------------|----------------|---------------------|-------------------|
| Bowker test | 11.675 | 0.307 | 0.303 |
| modified Wald test | 11.849 | 0.295 | 0.302 |
| continuity correction | 15.294 | 0.122 | 0.004 |

Table 4: simulated and approximate results for data set 2

Subsequently we consider a smaller data set given in table 5 with all $E(N_{ij}) < 5$ for all $i \neq j$ and structural zeroes: $n_{34} = n_{43} = 0$.

The results of our analysis are displayed in figure 3, the values of the test statistics with the corresponding p-values are specified in table 6. The approximation for Bowker's test and the modified Wald test is reasonable in the lower tail of the distribution but gets worse quickly (at test statistic value 7 for both tests).

Inspecting the p-values of these tests given in table 6 we come to different con-

$$\begin{pmatrix} 10 & 2 & 1 & 1 & 3 \\ 6 & 17 & 4 & 0 & 4 \\ 1 & 0 & 23 & 0 & 2 \\ 1 & 2 & 0 & 14 & 0 \\ 0 & 5 & 0 & 3 & 31 \end{pmatrix}$$

Table 5: data set 3; 16 expected values $E(N_{ij}), i \neq j$, are less than 3, 4 expected values $E(N_{ij}), i \neq j$, are less than 5

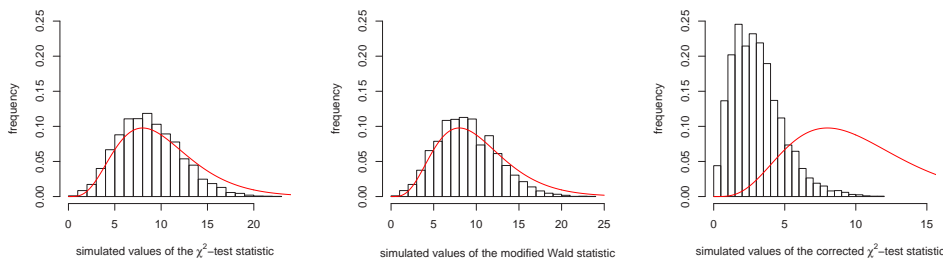


Figure 3: Histograms of the simulated values of the three symmetry test statistics for data set 3

clusions for the simulated and the approximate method, i.e. we reject H_0 at level $\alpha = 0.05$ for Bowker's test and the modified Wald test. Thus the approximate tests are more conservative than the simulated equivalent. The adequacy of the χ^2_{10} -approximation for the continuity corrected test for data set 3 is even worse than before and will not be discussed further.

| | test statistic | approximate p-value | simulated p-value |
|-----------------------|----------------|---------------------|-------------------|
| Bowker test | 16.111 | 0.096 | 0.027 |
| modified Wald test | 16.434 | 0.087 | 0.049 |
| continuity correction | 17.486 | 0.064 | 0.0001 |

Table 6: simulated and approximate results for data set 3

So far we examined contingency tables with cell entries $n_{ij}, i \neq j, i, j \in \{1, \dots, 5\}$ all less than 10. A data set with a bigger range for the values of n_{ij} is more realistic and will be analyzed in the following. The data are given in table 7, the results are displayed in figure 4 and table 8.

The expected values $E(N_{ij}), i \neq j$, for data set 4 range from 45.5 (for cell entries (2,3) and (3,2)) to 0 (for cell entries (1,5),(5,1),(2,5) and (5,2)). The approximation of the distribution of the test statistic is inappropriate for all considered tests,

$$\begin{pmatrix} 83 & 18 & 5 & 7 & 0 \\ 20 & 67 & 37 & 5 & 0 \\ 9 & 54 & 76 & 15 & 10 \\ 4 & 9 & 23 & 70 & 8 \\ 0 & 0 & 3 & 27 & 65 \end{pmatrix}$$

Table 7: data set 4, expected values $E(N_{ij})$, $i \neq j$, differ from 0 to 45.5

particularly for the continuity corrected test. Compared to the simulated results, the approximation is conservative although all tests reject the hypothesis of symmetry at level $\alpha = 0.05$.

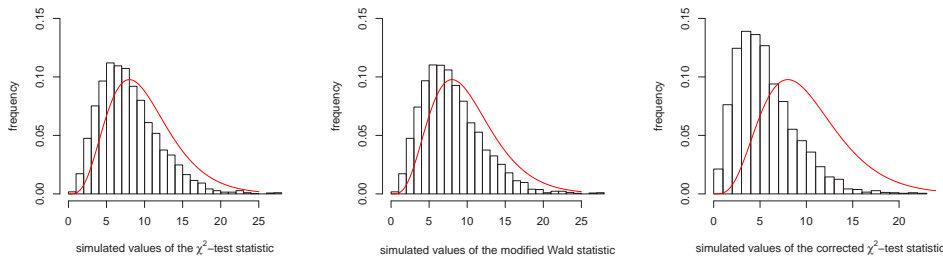


Figure 4: Histograms of the simulated values of the three symmetry test statistics for data set 4

| | test statistic | approximate p-value | simulated p-value |
|-----------------------|----------------|---------------------|-------------------|
| Bowker test | 22.153 | 0.014 | 0.005 |
| modified Wald test | 22.378 | 0.013 | 0.005 |
| continuity correction | 24.062 | 0.007 | 0.000 |

Table 8: simulated and approximate results for data set 4

Data example

Mobility is an important precondition for many aspects of modern life like economic growth. But associated with increasing mobility is increasing traffic noise. The research network "Quiet traffic" sponsored by the "Bundesministerium für Bildung und Forschung" (BMBF) examines among others things the effect of transportation noise on humans, specifically the mental and physical health (Griefahn et al. 2005). We analyze a data set from the 1st preliminary study. 72 subjects are exposed

twice to two different noise sources (i.e. road and rail noise) with four different noise levels (40, 52, 70 and 82 [dB]) (Kuhnt et al. 2004). The subjects judge the level of annoyance for the traffic noise on a scale which can be subsumed into five different levels (i.e. very high, high, moderate, low, very low). The resulting data set is given in table 9.

| | | annoyance level 2 | | | | |
|-------------------|-----------|-------------------|-----|-----|-----|-----------|
| | | very low | ... | ... | ... | very high |
| annoyance level 1 | very low | 51 | 28 | 3 | 0 | 0 |
| | ... | 15 | 68 | 40 | 5 | 1 |
| | ... | 0 | 29 | 77 | 21 | 1 |
| | ... | 0 | 4 | 19 | 80 | 14 |
| | very high | 0 | 1 | 5 | 26 | 88 |

Table 9: Data set provided by the research network "Quiet traffic" for the annoyance level of traffic noise

We are interested in whether the subjects classify noise differently when they are exposed to it twice, thus we test for symmetry. Several expected values of this data are very small or even equal zero. Hence the χ^2 -approximation of the distribution of the test statistic should be questioned. For the analysis of this data we follow the same procedure as in the simulation study, i.e. we generate a Markov Chain with 500000 states, disregard the first 100000 tables in the burn-in-phase and calculate the values of the test statistics (introduced in section 2) of each 100th table. The test decision is based on the comparison of this values with the value of the corresponding test statistic of the original data set. The simulation results as well as the χ^2_{10} -approximation are displayed in figure 5.

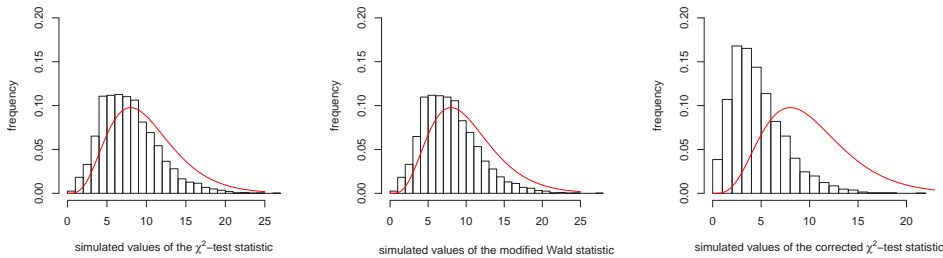


Figure 5: Histograms of the simulated values of the three symmetry test statistics for the data set provided by "Quiet traffic"

It turns out that the Bowker test and the modified Wald test yield almost the same results, but the χ^2_{10} -approximation of the distribution of the test statistics doesn't seem to fit very well. The adequacy of the χ^2_{10} -approximation is even worse for the corrected test of symmetry. The simulated and approximate results of the tests are given in table 10.

| | test statistic | approximate p-value | simulated p-value |
|-----------------------|----------------|---------------------|-------------------|
| Bowker test | 15.162 | 0.127 | 0.040 |
| modified Wald test | 15.245 | 0.123 | 0.041 |
| continuity correction | 14.548 | 0.149 | 0.005 |

Table 10: values of the test statistic of symmetry tests with approximate and simulated p-values for the "Quiet traffic" data set

According to table 10 we can reject the hypothesis of symmetry at level $\alpha = 0.05$. Looking at the data we find that the judgement of the subjects is almost constant and that there is hardly any change over more than two categories. In the cases where the individuals classify the level of annoyance differently, the changes are mostly towards a more moderate judgement.

5 Conclusion

We presented the Bowker test and two modifications, i.e. the modified Wald test and the continuity corrected test to test for symmetry for matched-pairs data. The test decisions are based on an approximation of the distribution of the test statistic which might be inappropriate. To overcome this problem, we introduced tests for symmetry using MCMC methods, computational commutative algebra and the test statistics of the Bowker test and the two modifications. Based on our simulation study we examined the adequacy of the approximation and a potential cut-off point to justify the approximation. It turns out that the approximation works reasonably for the Bowker test and the modified Wald test even if the expected values $E(N_{ij})$, $i \neq j$, are all small. But analyzing a data set whose expected values differ by more than 40, the approximation fails. For the continuity corrected test the approximation works inappropriately (is conservative) for all considered data sets in our simulation study. In order to avoid test decisions based on an inadequate approximation of the distribution of the test statistic, we suggest Bowker's test for symmetry using MCMC methods and the Diaconis-Sturmfels algorithm. Further

modifications of the Bowker test may also be included in the future.

Acknowledgements

The financial support of the Deutsche Forschungsgemeinschaft (SFB 475: "Reduction of Complexity for Multivariate Data Structures" and Graduiertenkolleg "Statistical modelling") is gratefully acknowledged.

Appendix

A description of $T((i, j))$

An important condition for the use of the Diaconis-Sturmfels algorithm is the possibility to rewrite the set of all data sets with given sufficient statistic in terms of $\mathcal{Z}_t = \{z : \mathcal{H} \rightarrow \mathbb{N} \mid \sum_{x \in \mathcal{H}} z(x)T(x) = t\}$ with \mathcal{H} the finite sample space. We shortly review the theoretical background. Assuming symmetry for the underlying $I \times I$ -table and the multinomial distribution for the cell counts $\{N_{ij}\}$, the joint density is given by $f(n_{11}, n_{12}, \dots, n_{II}) = \frac{n!}{n_{11}!n_{12}!\dots n_{II}!} \prod_{i=1}^I \prod_{j=1}^I p_{ij}^{n_{ij}}$. Assuming symmetry, this density can be described in terms of an $I \cdot (I-1)$ parametric exponential family with sufficient statistic $t = (N_{ii}, i = 1, \dots, (I-1); (N_{ij} + N_{ji}), j = (i+1), \dots, I, i = 1, \dots, (I-1))'$. \mathcal{Z}_t is the set of all $I \times I$ -tables with cell entries $n_{ij} \in \mathbb{N}$ and realization t of the sufficient statistic given above. The finite sample space of such a table is $\mathcal{H} = \{(i, j), i, j \in \{1, \dots, I\}\}$. For rewriting t in terms of \mathcal{Z}_t , we have to extend t to t^* with $t^* = (N_{ii}, i = 1, \dots, (I-1); (N_{ij} + N_{ji}), (N_{ji} + N_{ij}), j = (i+1), \dots, I, i = 1, \dots, (I-1))'$.

$T((i, j)), i, j \in \{1, \dots, I\}$ has the same length than t^* which is $(I-1) + 2 \cdot \sum_{v=1}^{I-1} (I-v)$, it has cell entries 0 and 1 and can be divided into I parts. The first part has length $(I-1)$ and represents the sufficient statistic that corresponds to the diagonal cells $N_{ii}, i = 1, \dots, (I-1)$. In particular the i th entry of $T((i, i))$ is 1, all other entries are 0. $T((i, j)), i, j = 1, \dots, I, i \neq j$ is defined as follows: the first $I-1$ entries are 0, the remaining parts of $T((i, j))$ have each length $2 \cdot (I-v), v = 1, \dots, I-1$. Particularly they represent the $(N_{ij} + N_{ji}), (N_{ji} + N_{ij}), j = (i+1), \dots, I, i = 1, \dots, (I-1)$ by two cell entries equal to 1 and the other entries equal to 0.

Example

We want to test for symmetry in a 5×5 -table. The sufficient statistic t is given by $t = (N_{11}, N_{22}, N_{33}, N_{44}, N_{12} + N_{21}, N_{13} + N_{31}, N_{14} + N_{41}, N_{15} + N_{51}, N_{23} + N_{32}, N_{24} + N_{42}, N_{25} + N_{52}, N_{34} + N_{43}, N_{35} + N_{53}, N_{45} + N_{54})'$.

We extend t to t^* with

$$t^* = (N_{11}, N_{22}, \dots, N_{44}, N_{12} + N_{21}, N_{21} + N_{12}, \dots, N_{45} + N_{54}, N_{54} + N_{45})'$$

Thus $T((i, j))$ is a vector of length 24 given by

$$\begin{aligned}
 T((1, 1)) &= (1, 0, 0, \dots, 0)' \\
 &\vdots \\
 T((4, 4)) &= (0, 0, 0, 1, 0, \dots, 0)' \\
 T((1, 2)) &= (0, 0, 0, 0, 1, 1, 0, \dots, 0)' \\
 T((2, 1)) &= (0, 0, 0, 0, 1, 1, 0, \dots, 0)' \\
 &\vdots \\
 T((5, 4)) &= (\underbrace{0, \dots, 0}_{\text{length: } 5-1=4}, \underbrace{0, \dots, 0}_{\text{length: } 2 \cdot (5-1)=8}, \underbrace{0, \dots, 0}_{\text{length: } 2 \cdot (5-2)=6}, \dots, \underbrace{1, 1}_{\text{length: } 2})'.
 \end{aligned}$$

B Additional results of the simulation study

$$\begin{pmatrix}
 35 & 4 & 2 & 4 & 3 \\
 2 & 47 & 3 & 5 & 2 \\
 3 & 3 & 25 & 3 & 3 \\
 1 & 2 & 3 & 23 & 1 \\
 3 & 0 & 5 & 4 & 11
 \end{pmatrix}$$

Table 11: data set 5, all expected values $E(N_{ij})$, $i \neq j$, are less than 4

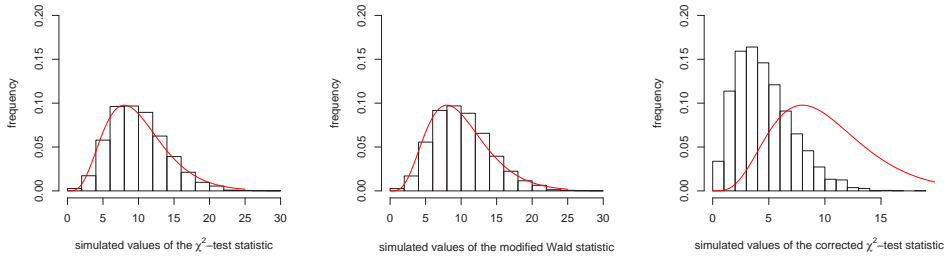


Figure 6: Histograms of the simulated values of the three symmetry test statistics for the data set 5

| | test statistic | approximate p-value | simulated p-value |
|-----------------------|----------------|---------------------|-------------------|
| Bowker test | 8.252 | 0.604 | 0.632 |
| modified Wald test | 8.318 | 0.598 | 0.632 |
| continuity correction | 7.163 | 0.710 | 0.159 |

Table 12: values of the test statistic of symmetry tests with approximate and simulated p-values for data set 5

$$\begin{pmatrix} 38 & 0 & 5 & 4 & 1 \\ 28 & 16 & 13 & 4 & 20 \\ 0 & 28 & 5 & 1 & 1 \\ 13 & 27 & 16 & 48 & 24 \\ 0 & 10 & 0 & 18 & 29 \end{pmatrix}$$

Table 13: data set 6, expected values $E(N_{ij})$, $i \neq j$, vary from 0.5 to 21

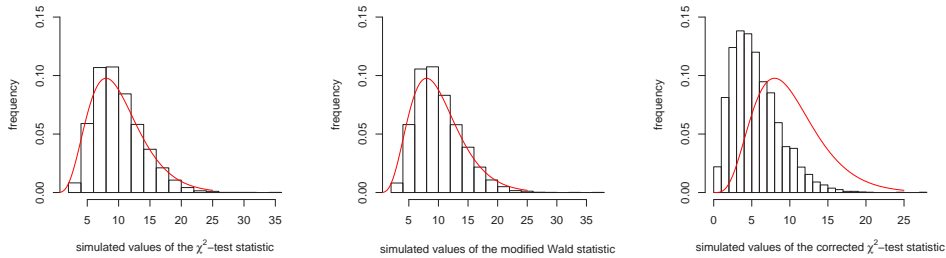


Figure 7: Histograms of the simulated values of the three symmetry test statistics for the data set 6

| | test statistic | approximate p-value | simulated p-value |
|-----------------------|----------------|---------------------|-------------------|
| Bowker test | 79.743 | 0 | 0 |
| modified Wald test | 83.850 | 0 | 0 |
| continuity correction | 82.297 | 0 | 0 |

Table 14: values of the test statistic of symmetry tests with approximate and simulated p-values for data set 6

$$\begin{pmatrix} 35 & 0 & 2 & 4 & 0 \\ 2 & 47 & 1 & 0 & 2 \\ 2 & 3 & 25 & 2 & 3 \\ 1 & 2 & 3 & 23 & 1 \\ 2 & 1 & 0 & 3 & 11 \end{pmatrix}$$

Table 15: data set 7, expected values $E(N_{ij})$, $i \neq j$, are all less than 2.5

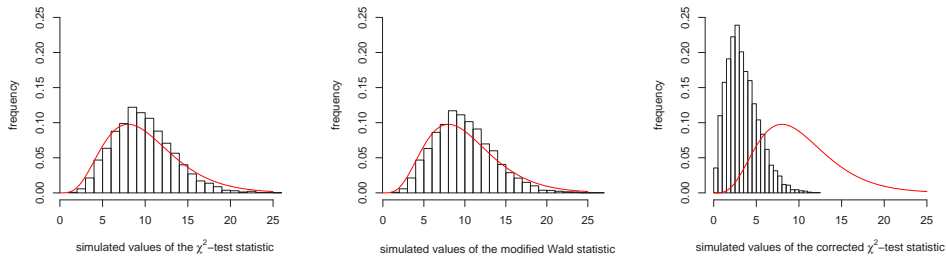


Figure 8: Histograms of the simulated values of the three symmetry test statistics for the data set 7

| | test statistic | approximate p-value | simulated p-value |
|-----------------------|----------------|---------------------|-------------------|
| Bowker test | 13.333 | 0.206 | 0.166 |
| modified Wald test | 13.486 | 0.198 | 0.107 |
| continuity correction | 20.933 | 0.022 | 0.000 |

Table 16: values of the test statistic of symmetry tests with approximate and simulated p-values for data set 7

References

- AGRESTI, A. (2002): Categorical Data Analysis (second edition). John Wiley, New York.
- BOEHNKE, M. LENGFELD, C.D. (1998): Genetic Association Mapping Based on Discordant Sib Pairs: The Discordant-Alleles Test. *Am. J. Hum. Genet.*, 62, 950-961.
- BOWKER, A.H. (1948): A Test for symmetry in Contingency Tables. *Journal American Statistical Association*, 43, 572-574.
- CHIB, S., GREENBERG, E. (1995): Understanding the Metropolis-Hastings-Algorithm. *The American Statistician*, 49, 327-335.
- COX, D., LITTLE, J., O'SHEA, D. (1992): Ideals, Varieties, and Algorithms. Springer, New York.
- DIACONIS, P., SALOFF-COSTE, L. (1995): What do we know about the Metropolis algorithm? *Technical Report. Dept. Mathematics, Harvard University.*
- DIACONIS, P., STURMFELS, B. (1998): Algebraic Algorithms for Sampling from Conditional Distributions. *The Annals of Statistics*, 26 , 363-397.
- EDWARDS, A.L. (1948): Note on the "Correction for Continuity" in Testing the Significance of the Difference Between Correlated Proportions. *Psychometrika*, 13, 185-187.
- EWENS, W.J., GRANT, G.R. (2001): Statistical Methods in Bioinformatics. An Introduction. Springer, New York.
- FAHRMEIR, L., KAUFMANN, H.L., OST, F. (1981): Stochastische Prozesse. Carl Hanser Verlag, München.
- GRIEFAHN, B., MARKS, A., ROBENS, S. (2005) Noise emitted from road, rail and air traffic and their effects on sleep. Submitted for publication.
- KUHNT, S., SCHÜRMAN, C. GRIEFAHN, B. (2004): Annoyance from Multiple Transportation Noise: Statistical Models and Outlier Detection. *Methods of Information in Medicine*, 5, 510-515.
- MAY, W.L., JOHNSON, W.D. (2001): Symmetry in Square Contingency Tables: Tests of Hypotheses and Confidence Interval Construction. *Journal of Biopharmaceutical Statistics*, 11, 23-33.

MCNEMAR, Q. (1947): Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages. *Psychometrika*, 12, 153-157.

PISTONE, G., RICCOMAGNO, E., WYNN, H.P. (2001): Algebraic Statistics. Chapman & Hall CRC Press, Boca Raton, FL.

RAPALLO, F. (2002): Algebraic Exact Inference for Rater Agreement Models. *Preprint #467 Dipartimento di Matematica, Università di Genova*.

RAPALLO, F.(2003): Algebraic Markov Bases and MCMC for Two-Way Contingency Tables. *Scandinavian Journal of Statistics*, 30, 385-397.

SHESKIN, D.J. (2000): Handbook of Parametric and Nonparametric Statistical Procedures. Chapman & Hall, Boca Raton.

SØRENSEN, D., GIANOLA, D. (2002): Likelihood, Bayesian, and MCMC Methods in Qualitative Genetics. Springer, New York.

WITTING, H. (1985): Mathematische Statistik I, Teubner, Stuttgart.