

Article

# New Equivalence Tests for Approximate Independence in Contingency Tables

Vladimir Ostrovski 

Ampega Asset Management GmbH, Charles-de-Gaulle-Platz 1, 50679 Cologne, Germany;  
vladimir\_ostrovski@web.de

Received: 24 March 2019; Accepted: 21 April 2019; Published: 23 April 2019



**Abstract:** We introduce new equivalence tests for approximate independence in two-way contingency tables. The critical values are calculated asymptotically. The finite sample performance of the tests is improved by means of the bootstrap. An estimator of boundary points is developed to make the bootstrap based tests statistically efficient and computationally feasible. We compare the performance of the proposed tests for different table sizes by simulation. Then we apply the tests to real data sets.

**Keywords:** testing; approximate independence; contingency tables; bootstrap; equivalence; minimum distance; boundary point estimator

**MSC:** 62F03; 62G10

## 1. Introduction

Testing for approximate row-column independence in two-way contingency tables is a common task in statistical practice. The first publications on this topic go back to Hodges and Lehmann [1], Diaconis and Efron [2]. More recently, Liu and Lindsay [3] applied the semi-parametric tubular tolerance regions to the row-column independence model in two-way contingency tables. The method relies on the analytical properties of the LRT statistic to obtain a closed form estimator of boundary points. Wellek [4] develops a test for independence in multi-way contingency tables in Section 9.2. For this purpose, he applies a test for consistency with a fully specified multinomial distribution as follows. First, the marginal distributions of the contingency table are calculated. The test statistic is the Euclidean distance between the product measure of the marginal distributions and the contingency table. The critical value is calculated asymptotically.

Ostrovski [5] proposes a general method to test equivalence to families of multinomial distributions, which is based on the minimum distance

$$d(p, \mathcal{M}) = \inf_{q \in \mathcal{M}} d(p, q) \quad (1)$$

to a family  $\mathcal{M}$  of multinomial distributions. If  $d$  is Euclidean distance and  $\mathcal{M}$  is the independence model then the calculation of minimum distance Equation (1) requires numerical optimization. Generally, the method relies on the existence of a continuous minimizer of Equation (1). Unfortunately, it could not be shown if a continuous minimizer exists for the independence model. Instead, Ostrovski [5] assumes the existence of a continuous minimizer at all points and then applies the method to test for approximate independence. Additionally, numerical calculation of minimum distance Equation (1) makes the bootstrap test computationally intensive.

We follow the lines of [5], but avoid the numerical valuation of the minimum distance Equation (1) in the special case of independence testing. We also propose an efficient bootstrap test, which is based on the randomized estimator of the boundary points.

Any two-way contingency table of the size  $k_1 \times k_2$  corresponds to a probability matrix from  $\mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$ . Let  $p = (p_{ij})$  denote the probability matrix. Let  $\mathcal{M}$  be the independence model, which contains all product measures of the corresponding dimensions. The approximate row-column independence can be shown by testing

$$H_0 = \{d(p, \mathcal{M}) \geq \varepsilon\} \text{ against } H_1 = \{d(p, \mathcal{M}) < \varepsilon\}, \tag{2}$$

where  $\varepsilon > 0$  is a tolerance parameter.

Let  $r$  and  $c$  denote the probability vectors of the marginal distributions, which are defined by  $r_i = \sum_{j=1}^{k_2} p_{ij}$  and  $c_j = \sum_{i=1}^{k_1} p_{ij}$ . A probability matrix  $p$  belongs to  $\mathcal{M}$  iff the equality  $p_{ij} = r_i c_j$  is fulfilled for all  $p_{ij}$ . We consider the transformations  $h_a$  and  $h_r$  of the matrix  $p$ , which are defined by  $h_a(p) = (p_{ij} - r_i c_j)$  and  $h_r(p) = \left(\frac{p_{ij}}{r_i c_j}\right)$ .

For any differentiable distance  $l$  on  $\mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$  we define two new distances  $d_a(p, q) = l(h_a(p), h_a(q))$  and  $d_r(p, q) = l(h_r(p), h_r(q))$ . It should be noted that  $d_a$  and  $d_r$  are only pseudo-metrics because  $d_r(p, q) = 0$  or  $d_a(p, q) = 0$  does not imply  $p = q$ . We put these distances in Equation (1) and obtain

$$d_a(p, \mathcal{M}) = \inf_{q \in \mathcal{M}} d_a(p, q) = \inf_{q \in \mathcal{M}} l(h_a(p), 0) = l(h_a(p), 0)$$

and  $d_r(p, \mathcal{M}) = l(h_r(p), 1)$ , where 0 denotes the zero matrix and 1 is the matrix of ones. The distances  $d_a(p, \mathcal{M})$  and  $d_r(p, \mathcal{M})$  can be interpreted respectively as the absolute deviation and the relative deviation between  $p$  and the product measure of the marginal distributions. The distances  $d_a(p, \mathcal{M})$  and  $d_r(p, \mathcal{M})$  are easy to calculate without optimization.

Therefore,  $d_a$  and  $d_r$  are good candidates for the general distance  $d$  in Definition (1) and we will use only these two specific distances in remainder of the paper.

We observe a contingency table  $p_n$  of relative frequencies, where  $n$  is the sample size and  $p$  is the true underlying probability matrix. Then the test statistic for Equation (2) is  $T_a(p_n) = \sqrt{n}(d_a(p_n, \mathcal{M}) - \varepsilon)$  or  $T_r(p_n) = \sqrt{n}(d_r(p_n, \mathcal{M}) - \varepsilon)$  depending on user preference. Below we write  $d_*$  instead of  $d_a$  and  $d_r$  if the statements are correct for both distances. We use the subscript  $*$  instead of  $a$  and  $r$ , if appropriate.

## 2. Asymptotic Tests

In this section, we derive the asymptotic distribution of the test statistic and give a detailed description of the asymptotic test.

Let  $v : \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} \rightarrow \mathbb{R}^{k_1+k_2}$  be the usual bijection  $v(p) = (p_{11}, p_{12}, \dots, p_{k_1 k_2})$ . Let  $\dot{d}_*$  denote the derivative of the function  $q \mapsto d_*(v^{-1}(q), \mathcal{M})$ , which can be easily calculated using the chain rule.

**Proposition 1.** *Let  $p$  be a boundary point of  $H_0$  and  $q = v(p)$ . Let  $D_q$  denote a square diagonal matrix, whose diagonal entries are  $q_1, \dots, q_{k_1+k_2}$ . Then the asymptotic distribution of  $T_*(p_n)$  is Gaussian with mean zero and variance  $\sigma_*(p) = \dot{d}_*(q) \Sigma(q) \dot{d}_*(q)^t$ , where  $\Sigma(q) = D_q - qq^t$  is a covariance matrix.*

**Proof.** Let  $q_n = v(p_n)$ . The normalized vector  $\sqrt{n}(q_n - q)$  converges weakly to a random variable, which is Gaussian with mean zero and covariance matrix  $\Sigma(q)$ , see [6], Theorem 14.3-4 for details. The assertion follows by the delta method, see [7], p. 26, Theorem 3.1.  $\square$

The asymptotic variance  $\sigma_*(p)$  is unknown and can be estimated by  $\sigma_*(p_n)$ . The estimator  $\sigma_*(p_n)$  is consistent by the continuous mapping theorem because  $p \mapsto \sigma_*(p)$  is a continuous function. Let  $l_\alpha$  denote the lower  $\alpha$ -quantile of the normal distribution. Then the critical value of the asymptotic test is  $l_\alpha \sigma_*(p_n)$ . Now we have all components of the asymptotic test, which can be carried out as follows:

1. Given are the contingency table  $p_n$  of relative frequencies, the tolerance parameter  $\varepsilon$  and the significance level  $\alpha$ .
2. Calculate the test statistic  $T_*(p_n) = \sqrt{n} (d_*(p_n, \mathcal{M}) - \varepsilon)$ .
3. Calculate the asymptotic variance  $\sigma_*(p_n)$ .
4. Reject  $H_0$  if  $T_*(p_n) \leq l_\alpha \sigma_*(p_n)$ .

The outlined test is locally asymptotically most powerful, see [8], Proposition 3.

**Remark 1.** The minimum tolerance parameter  $\varepsilon$ , for which the asymptotic test rejects  $H_0$ , can be calculated as  $d_*(p_n, \mathcal{M}) - \frac{1}{\sqrt{n}} l_\alpha \sigma_*(p_n)$ .

**Remark 2.** The asymptotic test can be straightforward generalized for the multi-way contingency tables.

### 3. Bootstrap Tests

The parametric bootstrap is an efficient method to improve the finite sample performance of the proposed tests. Let  $\partial H_0$  denote the boundary of  $H_0$ . Let  $\tilde{p}_n$  denote an estimator of  $p$ , which fulfills the condition  $\tilde{p}_n \in \partial H_0$ . The critical value  $c(\alpha, p)$  can be estimated by  $c(\alpha, \tilde{p}_n) = \sup_{x \in \mathbb{R}} \{x : P(T_*(\tilde{p}_n) \leq x) \leq \alpha\}$  because the critical value should be estimated so as if  $H_0$  were true. The estimator  $c(\alpha, \tilde{p}_n)$  can be computed by the Monte Carlo method to any degree of accuracy.

The minimum distance estimator of  $p$  would be difficult to compute because the boundary  $\partial H_0$  cannot be parameterized to apply common optimization techniques. Therefore, we propose a computationally feasible estimator of  $p$ , which is based on the randomized approximation to the minimum distance estimator.

Let  $q$  be some probability matrix such that  $d_*(q, \mathcal{M}) > \varepsilon$ . If  $d_*(p_n, \mathcal{M}) \leq \varepsilon$ , then let  $a_n$  be the largest number from  $[0, 1]$  such that  $d_*(a_n p_n + (1 - a_n) q, \mathcal{M}) = \varepsilon$ . Otherwise let  $a_n = 1$ . The linear combination  $c(p_n, q) = a_n p_n + (1 - a_n) q$  is a consistent estimator of the boundary point  $p$  under additional requirements as shown below.

**Proposition 2.** Assume that  $d_*(ap + (1 - a) q, \mathcal{M}) > \varepsilon$  for all  $a \in [0, 1)$ . Then  $c(p_n, q) \rightarrow p$  a.e. for  $n \rightarrow \infty$ .

**Proof.** We show that  $a_n \rightarrow 1$  for  $n \rightarrow \infty$ . Let  $d_*(p_n, \mathcal{M}) < \varepsilon$  because  $a_n = 1$  otherwise. The function  $f : a \mapsto d_*(ap_n + (1 - a) q, \mathcal{M})$  is continuous on  $[0, 1]$  and  $f(0) = d_*(q, \mathcal{M}) > \varepsilon$  as well as  $f(1) = d_*(p_n, \mathcal{M}) \leq \varepsilon$ . Therefore, there exists a largest number  $a_n \in [0, 1]$  such that  $f(a_n) = \varepsilon$ . It is worth mentioning that  $a_n$  is a function of  $p_n$ .

Let  $E = \{\lim_{n \rightarrow \infty} p_n = p\}$ . By the strong law of large numbers,  $p_n$  converges to  $p$  a.e. for  $n \rightarrow \infty$  and therefore  $P(E) = 1$ . Let  $\omega \in E$  be an arbitrary point and let  $a_n(\omega)$  denote  $a_n(p_n(\omega))$ . The sequence of  $a_n(\omega)$  is bounded. Hence, there exists a convergent sub-sequence  $a_{n_j}(\omega) \rightarrow a_0(\omega)$  for  $j \rightarrow \infty$ . We obtain

$$a_{n_j}(\omega) p_{n_j}(\omega) + (1 - a_{n_j}(\omega)) q \rightarrow a_0(\omega) p + (1 - a_0(\omega)) q$$

for  $j \rightarrow \infty$  and consequently  $d_*(a_0(\omega) p + (1 - a_0(\omega)) q, \mathcal{M}) = \varepsilon$ . We conclude  $a_0(\omega) = 1$  due to the assumption  $d_*(ap + (1 - a) q, \mathcal{M}) > \varepsilon$  for all  $a \in [0, 1)$ . Overall, we have shown that  $a_n(\omega) \rightarrow 1$  for all  $\omega \in E$ .  $\square$

Let  $Q = \{q_1, \dots, q_m\}$  be a finite set of probability matrices, such that any  $q_i$  fulfills  $d_*(q_i, \mathcal{M}) > \varepsilon$ . We define the estimator  $\tilde{p}_n$  as a minimum distance estimator among the linear combinations  $c(p_n, q)$  for all  $q \in Q$ . Formally, the estimator  $\tilde{p}_n$  equals  $c(p_n, q_i)$ , which fulfills the conditions  $q_i \in Q$  and  $l(c(p_n, q_i), p_n) = \min_{q \in Q} l(c(p_n, q), p_n)$ . Note that the distance  $l$  is used to define the estimator  $\tilde{p}_n$  because  $d_*$  is a pseudo-metric only.

**Corollary 1.** Let at least one  $q \in Q$  satisfy  $d_*(ap + (1 - a)q, \mathcal{M}) > \varepsilon$  for all  $a \in [0, 1)$ . Then  $\tilde{p}_n \rightarrow p$  a.e. for  $n \rightarrow \infty$ .

**Proof.** By definition of  $\tilde{p}_n$ , we obtain

$$l(\tilde{p}_n, p_n) \leq l(c(p_n, q), p_n) \leq l(c(p_n, q), p) + l(p, p_n),$$

where  $l(c(p_n, q), p) \rightarrow 0$  a.e. by Proposition 2 and  $l(p, p_n) \rightarrow 0$  a.e. by the strong law of large numbers.  $\square$

The bootstrap test can be carried out as follows:

1. Given are the contingency table  $p_n$  of relative frequencies, the tolerance parameter  $\varepsilon$ , the number of exterior points  $m$  and the significance level  $\alpha$ .
2. Calculate the test statistic  $T_*(p_n) = \sqrt{n}(d_*(p_n, \mathcal{M}) - \varepsilon)$ .
3. If  $T_*(p_n) \geq 0$  then set  $\tilde{p}_n = p_n$  and go to step 7.
4. Find  $m$  different points  $q \in H_0$  such that  $d_*(q, \mathcal{M}) > \varepsilon$ . The following rejection algorithm can be applied for the search:
  - (a) Simulate a random matrix  $w$  whose entries are independently uniformly distributed on  $[0, 1]$ .
  - (b) Normalize  $w$  to a probability matrix  $q$ .
  - (c) Add  $q$  to  $Q$  if  $d_*(q, \mathcal{M}) > \varepsilon$  or reject  $q$  otherwise.
  - (d) Repeat previous steps until all exterior points are found.
5. Solve the equation  $d_*(a_n p_n + (1 - a_n)q, \mathcal{M}) = \varepsilon$  for  $a_n$  using some root finding method. Repeat for all  $q \in Q$ .
6. Find the minimum distance estimator  $\tilde{p}_n$  among all linear combinations  $c(p_n, q)$ , where  $q \in Q$ .
7. Estimate the critical value  $c(\alpha, \tilde{p}_n)$  using Monte Carlo simulation.
8. Reject  $H_0$  if  $T_*(p_n) \leq c(\alpha, \tilde{p}_n)$ .

**Remark 3.** The bootstrap test is asymptotically consistent, see [9], Theorem 15.6.1. Consequently, the test is also locally asymptotically most powerful see [8], Proposition 3.

**Remark 4.** The appropriate number of exterior points  $m$  can be found empirically. We found that  $m = (k_1 + k_2) * 50$  is sufficient and scales well with the table size.

**Remark 5.** The minimum tolerance parameter  $\varepsilon$ , for which the bootstrap test rejects  $H_0$ , can be found numerically. For this purpose, the equation  $T_*(p_n) = c(\alpha, \tilde{p}_n)$  should be solved for the tolerance parameter  $\varepsilon$  using some root finding algorithm. The exterior points and bootstrap samples should remain unchanged during optimization.

#### 4. Simulation Study of Finite Sample Performance

We study the finite sample performance of the proposed tests by the Monte Carlo simulation for different sample sizes and table sizes. The tests are implemented in VB.NET and available online, see <https://github.com/TestingEquivalence/TestingApproximateIndependence>.

The distance  $l$  is scaled Euclidean distance  $l_2$ , where the scale factor is necessary to obtain comparable test results for different table sizes. We use  $l = \frac{1}{\sqrt{k_1 k_2}} l_2$  in case of  $d_r$  and  $l = \sqrt{k_1 k_2} l_2$  in case of  $d_a$ . Alternatively the smoothed total variation distance would be a good choice, see [8].

The minimum  $\varepsilon$ , for which the test power equals 0.9, is calculated for different table sizes and sample sizes at the uniform probability matrices for the purpose of throwing some light on the appropriate values of  $\varepsilon$  and the effective sample sizes. Table 1 shows the minimum  $\varepsilon$  for the distance  $d_r$ . The minimum  $\varepsilon$  for  $d_a$  can be found in Table A1 because the results are very similar for  $d_a$  and  $d_r$ . The minimum  $\varepsilon$  decreases with the increasing sample size at the rate  $n^{-\frac{1}{2}}$ . The minimum parameter  $\varepsilon$  climbs with the increasing table size at the rate  $k_1 + k_2$ . Thus, the test power falls slowly with the

increasing table size. The bootstrap tests have a smaller minimum  $\epsilon$  than the asymptotic tests and the difference increases considerably with the table size.

**Table 1.** Minimum tolerance parameter  $\epsilon$ , for which the test power equals 0.9 at nominal level  $\alpha = 0.05$ , is calculated at the uniform probability matrices using the distance  $d_r$ . The sample size is 100, . . . , 10,000.

Asymptotic Test							
table size	100	200	500	1000	2000	5000	10,000
2 × 4	0.390	0.272	0.171	0.120	0.085	0.054	0.038
3 × 3	0.418	0.295	0.185	0.130	0.092	0.058	0.041
3 × 4	0.474	0.331	0.208	0.146	0.104	0.066	0.046
4 × 4	0.540	0.375	0.236	0.166	0.117	0.074	0.052
4 × 5	0.593	0.413	0.258	0.181	0.128	0.081	0.057
5 × 5	0.655	0.453	0.283	0.200	0.141	0.089	0.063
Bootstrap Test							
2 × 4	0.382	0.271	0.171	0.121	0.085	0.054	0.038
3 × 3	0.400	0.284	0.179	0.126	0.089	0.056	0.040
3 × 4	0.431	0.302	0.190	0.134	0.095	0.060	0.042
4 × 4	0.464	0.322	0.203	0.143	0.101	0.064	0.045
4 × 5	0.488	0.339	0.212	0.149	0.106	0.067	0.047
5 × 5	0.519	0.356	0.223	0.158	0.111	0.071	0.050

We study the type I error rates at 100 randomly selected points from  $\partial H_0$  because the boundary of  $H_0$  is a very complex set and it is difficult to identify particularly interesting boundary points. The points are found using steps 4 and 5 of the algorithm at the end of Section 3. The sample size  $n$  equals  $100 * (k_1 + k_2)$  to maintain similar test power for different table sizes because test power falls with increasing table size. The simulation results are summarized in Table 2. The power of all tests varies considerably from point to point. The averaged power of the asymptotic tests decreases quickly with the table size. The asymptotic tests are not conservative for the small tables and become very conservative for the larger tables. The averaged power of the bootstrap tests is very close to the nominal level for all table sizes. However, the bootstrap tests are not conservative for all table sizes. Particularly, the  $d_r$  based bootstrap test shows strong anti-conservative tendency.

A detailed analysis of the boundary points shows that the test power is far above the nominal level at the points, where  $r_i c_j$  is close to zero for some  $i$  and  $j$ . Therefore, the test results should be treated with caution, if the marginal probability is close to zero for at least one category.

**Table 2.** Summary of the simulated exact rejection probability of the equivalence tests at nominal level  $\alpha = 0.05$  and tolerance parameter  $\epsilon = 0.2$ . The rejection probability is simulated at 100 randomly selected boundary points. The sample size is  $(k_1 + k_2) * 100$  and the number of replications is 10.000 for each experiment.

	Asymptotic Test Based on $d_r$						Asymptotic Test Based on $d_a$					
	2 × 4	3 × 3	3 × 4	4 × 4	4 × 5	5 × 5	2 × 4	3 × 3	3 × 4	4 × 4	4 × 5	5 × 5
Minimum	0.02	0.01	0.01	0.00	0.00	0.00	0.04	0.03	0.02	0.01	0.00	0.00
Maximum	0.12	0.10	0.06	0.05	0.03	0.01	0.08	0.08	0.06	0.03	0.02	0.01
Average	0.05	0.04	0.02	0.01	0.01	0.00	0.06	0.05	0.03	0.02	0.01	0.00
Deviation	0.02	0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00
	Bootstrap Test Based on $d_r$						Bootstrap Test Based on $d_a$					
	Minimum	0.00	0.00	0.01	0.03	0.03	0.04	0.04	0.03	0.04	0.04	0.04
Maximum	0.07	0.08	0.10	0.13	0.10	0.11	0.09	0.08	0.07	0.07	0.07	0.06
Average	0.04	0.05	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.05
Deviation	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00

The conservative tests can be obtained by shrinking the tolerance parameter  $\varepsilon$ . Table A5 summarizes the simulation results for  $\varepsilon = 0.18$ , where the test power is calculated at the same points as in Table 2, i.e.,  $d_*(p, \mathcal{M}) = 0.2$  at all considered points. Then the  $d_a$  based tests are conservative at all points and the  $d_r$  based tests are still non conservative at some points.

The type II error rates are studied at 100 randomly selected product measures for each table size, see Table 3. It should be noted that Table 3 contains test power and the type II error rate equals 1 minus test power. The sample size equals  $100 * (k_1 + k_2)$  to be comparable to the type I error analysis. The power of  $d_r$ -based tests changes very strongly from point to point. Given the fixed table size, the power of the  $d_a$ -based tests is almost constant at all considered points. The averaged power of the asymptotic tests decreases slightly with the increasing table size. The averaged power of the bootstrap tests does not change with the table size.

**Table 3.** Summary of the simulated test power at nominal level  $\alpha = 0.05$  and tolerance parameter  $\varepsilon = 0.2$ . The rejection probability is simulated at 100 randomly selected product measures. The sample size is  $(k_1 + k_2) * 100$  and the number of replications is 10.000 for each experiment.

	Asymptotic Test Based on $d_r$						Asymptotic Test Based on $d_a$					
	2 × 4	3 × 3	3 × 4	4 × 4	4 × 5	5 × 5	2 × 4	3 × 3	3 × 4	4 × 4	4 × 5	5 × 5
Minimum	0.20	0.09	0.13	0.10	0.16	0.26	0.99	0.98	0.96	0.94	0.93	0.90
Maximum	0.99	0.98	0.97	0.95	0.92	0.87	1.00	1.00	0.99	0.97	0.96	0.92
Average	0.84	0.80	0.80	0.77	0.70	0.64	0.99	0.99	0.98	0.96	0.95	0.91
Deviation	0.17	0.21	0.18	0.16	0.18	0.15	0.00	0.00	0.00	0.00	0.01	0.00
	Bootstrap Test Based on $d_r$						Bootstrap Test Based on $d_a$					
	2 × 4	3 × 3	3 × 4	4 × 4	4 × 5	5 × 5	2 × 4	3 × 3	3 × 4	4 × 4	4 × 5	5 × 5
Minimum	0.13	0.13	0.22	0.28	0.56	0.73	0.98	0.97	0.97	0.98	0.98	0.99
Maximum	0.99	0.99	0.99	0.99	0.99	0.99	1.00	1.00	0.99	0.99	1.00	1.00
Average	0.82	0.80	0.86	0.89	0.90	0.94	0.99	0.99	0.99	0.99	0.99	0.99
Deviation	0.20	0.22	0.17	0.11	0.10	0.06	0.00	0.01	0.00	0.00	0.00	0.00

### 5. Real Data Sets

To demonstrate the application of the proposed tests, three examples with real data sets are considered: gender and nitrendipine therapy (Nitrendipine); eye color and hair color (Color); children number and income (Children). The corresponding two way contingency tables are given in Appendix A, Tables A2, A3 and A4. Table 4 displays the minimum tolerance parameter  $\varepsilon$ , for which  $H_0$  can be rejected at the nominal level  $\alpha = 0.05$ . The three examples are also used in [5], such that a direct comparison is possible. The results for distance  $d_a$  are similar to those presented in [5] after appropriate re-scaling. However, we avoid the unproven assumptions and the extensive use of the numeric optimization, which are necessary in [5].

The first example concerns with the question if the treatment outcome on nitrendipine mono-therapy in patients suffering from mild arterial hypertension depends on gender. The data set is also an example for approximate independence in [4]. The asymptotic and bootstrap test results for  $d_r$  are very close to each other. The results for  $d_a$  differ considerably for the asymptotic and bootstrap test. Given the small sample size, the treatment outcome and gender can be considered approximately independent.

A common example for independence testing is the cross-classification of eye color and hair color, see [2,3]. The test results in Table 4 reflect the well known fact that eye color and hair color are not independently distributed. All tests behave very similarly and can reject  $H_0$  only for very large values of  $\varepsilon$ .

The cross-classification of the number of children by the annual income has a large sample size. However, the category, where the number of children is larger than or equal 4, is sparsely populated. Therefore, the  $d_r$  based tests can reject  $H_0$  only for comparatively large values of  $\varepsilon$  and the test power is low. The  $d_a$  based tests show that the number of children and annual income may be considered approximately independent, but the approximation is very inaccurate.

**Table 4.** Minimum tolerance parameter  $\varepsilon$ , for which  $H_0$  can be rejected at the nominal level  $\alpha = 0.05$ .

Data Set	$n$	Tests Based on $d_r$		Tests Based on $d_a$	
		Asymptotic	Bootstrap	Asymptotic	Bootstrap
Nitrendipine	217	0.23	0.22	0.15	0.18
Color	592	0.59	0.58	0.59	0.59
Children	25263	0.28	0.31	0.18	0.18

**Funding:** This research received no external funding.

**Conflicts of Interest:** The author declares no conflict of interest.

**Appendix A**

**Table A1.** Minimum tolerance parameter  $\varepsilon$ , for which the test power equals 0.9 at nominal level  $\alpha = 0.05$ , is calculated at the uniform probability matrices using the distance  $d_a$ . The sample size is 100, . . . , 10,000.

Asymptotic Test							
table size	100	200	500	1000	2000	5000	10,000
2 × 4	0.377	0.267	0.169	0.120	0.085	0.054	0.038
3 × 3	0.405	0.290	0.183	0.130	0.092	0.058	0.041
3 × 4	0.458	0.327	0.207	0.146	0.104	0.065	0.046
4 × 4	0.520	0.368	0.234	0.165	0.117	0.074	0.052
4 × 5	0.567	0.403	0.255	0.180	0.128	0.081	0.057
5 × 5	0.623	0.444	0.281	0.199	0.141	0.089	0.063
Bootstrap Test							
2 × 4	0.378	0.268	0.170	0.120	0.085	0.054	0.038
3 × 3	0.396	0.282	0.178	0.125	0.089	0.056	0.040
3 × 4	0.423	0.302	0.190	0.134	0.095	0.060	0.042
4 × 4	0.457	0.320	0.203	0.143	0.101	0.064	0.045
4 × 5	0.478	0.337	0.212	0.149	0.106	0.067	0.047
5 × 5	0.501	0.354	0.222	0.158	0.111	0.070	0.050

**Table A2.** Contingency table relating gender and treatment outcome on nitrendipine mono-therapy in patients suffering from mild arterial hypertension.

Gender	Outcome Category			
	1	2	3	4
female	9	13	13	48
male	24	18	20	72

**Table A3.** Cross-classification of eye color and hair color.

Eye Color	Hair Color			
	Black	Brunette	Red	Blonde
Brown	68	119	26	7
Blue	20	84	17	94
Hazel	15	54	14	10
Green	5	29	14	16

**Table A4.** Cross-classification of number of children by annual income.

No. of Children	Annual Income			
	0–1	1–2	2–3	3+
0	2161	3577	2184	1636
1	2755	5081	2222	1052
2	936	1753	640	306
3	225	419	96	38
4+	39	98	31	14

**Table A5.** Summary of the simulated exact rejection probability of the equivalence tests at nominal level  $\alpha = 0.05$  and shrunk tolerance parameter  $\varepsilon = 0.18$ . The rejection probability is simulated at 100 randomly selected boundary points of  $H_0 = \{d_*(p, \mathcal{M}) \geq 0.2\}$ . The sample size is  $(k_1 + k_2) * 100$  and the number of replications is 10.000 for each experiment.

	Asymptotic Test Based on $d_r$						Asymptotic Test Based on $d_a$					
	$2 \times 4$	$3 \times 3$	$3 \times 4$	$4 \times 4$	$4 \times 5$	$5 \times 5$	$2 \times 4$	$3 \times 3$	$3 \times 4$	$4 \times 4$	$4 \times 5$	$5 \times 5$
Minimum	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00
Maximum	0.08	0.06	0.03	0.02	0.01	0.00	0.04	0.04	0.02	0.01	0.00	0.00
Average	0.02	0.02	0.01	0.00	0.00	0.00	0.02	0.02	0.01	0.00	0.00	0.00
Deviation	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00
	Bootstrap Test Based on $d_r$						Bootstrap Test Based on $d_a$					
	Minimum	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Maximum	0.05	0.06	0.09	0.12	0.06	0.07	0.04	0.04	0.03	0.03	0.02	0.02
Average	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01
Deviation	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00

## References

- Hodges, J.; Lehmann, E. Testing the approximate validity of statistical hypotheses. *J. R. Stat. Soc. Ser. B* **1954**, *16*, 261–268. [CrossRef]
- Diaconis, P.; Efron, B. Reply to comments on “Testing for independence in a two-way table: New interpretations of the chi-square statistic”. *Ann. Stat.* **1985**, *13*, 845–913. [CrossRef]
- Liu, J.; Lindsay, B.G. Building and using semiparametric tolerance regions for parametric multinomial models. *Ann. Stat.* **2009**, *37*, 3644–3659. [CrossRef]
- Wellek, S. *Testing Statistical Hypotheses of Equivalence and Noninferiority*; CRC Press: Boca Raton, FL, USA, 2010.
- Ostrovski, V. Testing equivalence to families of multinomial distributions with application to the independence model. *Stat. Probab. Lett.* **2018**, *139*, 61–66. [CrossRef]
- Bishop, Y.M.; Fienberg, S.E.; Holland, P.W. *Discrete Multivariate Analysis: Theory and Practice*; MIT Press: Cambridge, MA, USA, 1975.
- van der Vaart, A. *Asymptotic Statistics*; Cambridge University: Cambridge, UK, 1998.
- Ostrovski, V. Testing equivalence of multinomial distributions. *Stat. Probab. Lett.* **2017**, *124*, 77–82. [CrossRef]
- Lehmann, E.; Romano, J. *Testing Statistical Hypotheses*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2008.

