stats

MDPI

Article

New Equivalence Tests for Hardy–Weinberg Equilibrium and Multiple Alleles

Vladimir Ostrovski

ERGO Group AG, ERGO-Platz 1, 40198 Düsseldorf, Germany; vladimir_ostrovski@web.de

Received: 18 January 2020; Accepted: 31 January 2020; Published: 5 February 2020

Abstract: We consider testing equivalence to Hardy–Weinberg Equilibrium in case of multiple alleles. Two different test statistics are proposed for this test problem. The asymptotic distribution of the test statistics is derived. The corresponding tests can be carried out using asymptotic approximation. Alternatively, the variance of the test statistics can be estimated by the bootstrap method. The proposed tests are applied to three real data sets. The finite sample performance of the tests is studied by simulations, which are inspired by the real data sets.

Keywords: test; testing; equivalence; Hardy; Weinberg; Equilibrium; asymptotic; bootstrap; simulation study

1. Introduction

Hardy–Weinberg Equilibrium (HWE) plays an important role in the field of the population genetics and related scientific domains. HWE is a common assumption in many areas of research so that assessing the compatibility of observed genotype frequencies with HWE is a basic step of a complete statistical analysis. There are two main approaches to this undertaking: goodness of fit tests and equivalence tests.

A vast amount of literature exists on the goodness of fit tests for HWE, which includes application of the asymptotic χ^2 and likelihood ratio tests. The specific exact goodness of fit tests for HWE are developed in [1–5] among others. The null hypothesis of all these tests is that the underlying population is exactly in HWE. Hence, the goodness of fit tests are tailored to establish lack of compatibility with HWE.

The equivalence tests are appropriate to establish sufficiently good agreement of the observed genotype frequencies with HWE. The exact and approximate equivalence tests for the biallelic case are developed recently in [6–8]. To our best knowledge, there are not any equivalence tests for HWE and multiple alleles. Two different equivalence tests are developed in this paper for the case of multiple alleles. The tests can be carried out using the asymptotic approximation or bootstrap method.

A distribution of diploid genotypes at a k-allele locus can be represented as a lower triangular matrix p, where $p\left(i,j\right)$ is the probability of the genotypes with alleles i and j. Let $a\left(p\right)$ denote the allele distribution under p. The probability of the allele i under p can be calculated as $a\left(p,i\right) = \frac{1}{2}\sum_{j=1}^{k}\left(p\left(i,j\right)+p\left(j,i\right)\right)$. If the population is in HWE, then the genotype distribution fulfills the conditions $p(i,j) = 2a(p,i)a(p,j)$ for $i < j$ and $p(i,i) = a(p,i)^2$. Let $e\left(p\right)$ denote the genotype distribution under the assumption of HWE, which is implied by the allele distribution $a\left(p\right)$.

Euclidean distance $l_2\left(p,e(p)\right)$ can be considered a conditional distance between the genotype distributions p and $e\left(p\right)$ under the joint allele distribution $a\left(p\right)$. The equivalence test problem is then defined by

$$H_0 = \left\{l_2\left(p,e\left(p\right)\right) \geq \varepsilon\right\} \quad \text{and} \quad H_1 = \left\{l_2\left(p,e\left(p\right)\right) < \varepsilon\right\} \tag{1}$$

where ε is a tolerance parameter.

Let $\mathcal{M}$ denote the family of all possible genotype distributions at HWE. The minimum distance between p and $\mathcal{M}$ is defined by $d(p,\mathcal{M}) = \min_{q\in\mathcal{M}} l_2(p,q)$. The corresponding equivalence test problem is given by

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

We observe the genotype frequencies p_n of the sample size n. The natural test statistic for (1) is

$$T_c(p_n) = \sqrt{n}\left(l_2^2(p_n, e(p_n)) - \varepsilon^2\right), \tag{3}$$

which can be easily computed. The appropriate test statistic for (2) is

$$T_m(p_n) = \sqrt{n}\left(d^2(p_n,\mathcal{M}) - \varepsilon^2\right), \tag{4}$$

which requires optimization for the calculation of $d(p_n,\mathcal{M})$. The test statistic $T_c(p_n)$ can be considered a numerically efficient approximation to $T_m(p_n)$ because of $l_2(p_n, e(p_n)) \geq d(p_n,\mathcal{M})$. The subscript $*$ will be used instead of c and m in the reminder of the paper, if statements are appropriate for both cases.

If Hypothesis (1) or (2) of the non-equivalence can be rejected for some appropriate value of ε then the true underlying genotype distribution is close to HWE with the probability greater than $1-\alpha$, where α is the nominal level of the test. The appropriate value of ε depends on the application and the available sample size. The value of the parameter ε can be found by simulation as shown in Section 3.2. Alternatively, the minimum tolerance parameter ε, for which H_0 can be rejected, can be computed and reported, see Section 2 for details.

2. Equivalence Tests

In this section, we derive the asymptotic distributions of the test statistics $T_c(p_n)$ and $T_m(p_n)$. We provide also an algorithm for the asymptotic and bootstrap-based tests.

Let v be the usual bijective mapping of the matrix p to the vector $(p(1,1), p(1,2), \ldots, p(k,k))$. Let $\mathring{d}_c$ denote the derivative of the function $q \mapsto l_2^2\left(v^{-1}(q), e\left(v^{-1}(q)\right)\right)$, where q is a vector of length k^2. The derivative $\mathring{d}_c$ can be derived using the chain rule. Let $p_0 \in H_0$ fulfill the boundary condition $l_2(p_0, e(p_0)) = \varepsilon$ and let $q_0 = v(p_0)$. Then the asymptotic distribution of $T_c(p_n)$ under p_0 is Gaussian with mean zero and variance $\sigma_c^2(p_0) = \mathring{d}_c(q_0)\,\Sigma(q_0)\,\mathring{d}_c(q_0)^t$, where $\Sigma(q_0) = D_q - qq^t$ is a covariance matrix and D_q is a square diagonal matrix, whose diagonal entries are elements of q. The proof of the statement can be found in [9].

The test statistic $T_m(p_n)$ converges weakly under the assumption that there exists a continuous function h on an open neighborhood of p_0 such that $h(p) \in \mathcal{M}$ and $d(p,\mathcal{M}) = l_2(p, h(p))$. The existence of a continuous minimizer h is also an important requirement for the numerical computation of $d(p,\mathcal{M})$. We assume the existence of a continuous minimizer h on an open neighborhood of p_0 for the reminder of the paper. Let $\mathring{d}_m$ denote the derivative of the function $q \mapsto l_2^2\left(v^{-1}(q), h(p_0)\right)$. Then the asymptotic distribution of $T_m(p_n)$ under p_0 is Gaussian with mean zero and variance $\sigma_m^2(p_0) = \mathring{d}_m(q_0)\,\Sigma(q_0)\,\mathring{d}_m(q_0)^t$, see [10] for details.

The asymptotic variance $\sigma_*^2(p_0)$ is unknown and can be estimated by $\sigma_*^2(p_n)$. The asymptotic test can be carried out as follows:

1. Given are the genotype frequencies p_n, the tolerance parameter ε and the significance level α.
2. Compute the tests statistic $T_*(p_n)$.
3. Estimate the asymptotic variance by $\sigma_*^2(p_n)$.
4. Reject H_0 if $T_*(p_n) \leq c_\alpha \sigma_*(p_n)$, where c_α is the lower α-quantile of the normal distribution.

The minimum tolerance parameter ε, for which the asymptotic test can reject H_0, can be computed as $\sqrt{l_2^2\left(p_n, e\left(p_n\right)\right) - n^{-\frac{1}{2}} c_\alpha \sigma_c\left(p_n\right)}$ or $\sqrt{d\left(p_n, \mathcal{M}\right) - n^{-\frac{1}{2}} c_\alpha \sigma_m\left(p_n\right)}$ correspondingly.

To improve the finite sample performance of the proposed tests, the bootstrap method is applied to estimate the variance of $T_*\left(p_n\right)$, see [11], Section 6 for details. The estimator $\sigma_*\left(p_n\right)$ is then replaced by the bootstrap estimator of the variance. Otherwise, everything stays the same.

3. Simulation Study

The proposed tests are implemented in R and are freely available on GitHub under https://github.com/TestingEquivalence/HardyWeinbergEquilibriumR. All simulations are performed in R-Studio on a usual scientific workstation.

3.1. Real Data Sets

The equivalence tests are applied to the following data sets, which are already analyzed in the literature on the goodness of fit tests: 1. from rheumatoid arthritis study [12]; 2. from the documentation included with the GENEPOP software package [13]; 3. genotype frequency data at Rhesus locus [14]. The genotype distributions of the data sets are given in Table 1.

The minimum tolerance parameters ε, for which the tests can reject the corresponding H_0, are displayed in Table 2. The distances $d\left(p_n, \mathcal{M}\right)$ and $l_2\left(p_n, e\left(p_n\right)\right)$ are close to each other in all cases so that $l_2\left(p_n, e\left(p_n\right)\right)$ provides a good approximation to $d\left(p_n, \mathcal{M}\right)$. The test results are also similar for T_c and T_m. The bootstrap tests are slightly more conservative than the asymptotic tests in all cases.

It could not be shown that data sets 1 and 2 are close to HWE. All goodness of fit tests in [15] reject also the null hypothesis of HWE for data set 2 at the nominal level 0.05. Data set 3 is very close to HWE. This observation corresponds to the results of the goodness of fit tests in [5,15].

Table 1. The data sets: (1) from rheumatoid arthritis study, [12]; (2) from the documentation included with the GENEPOP software package, [13]; (3) genotype frequency data at Rhesus locus, [14].

(1)

5			
40	12		
6	32	2	
30	55	15	33

(2)

2			
12	24		
30	34	54	
22	21	20	10

(3)

1236								
120	3							
18	0	0						
982	55	7	249					
32	1	0	12	0				
2582	132	20	1162	29	1312			
6	0	0	4	0	4	0		
2	0	0	0	0	0	0	0	
115	5	2	53	1	149	0	0	4

Table 2. Minimum tolerance parameter ε, for which H_0 can be rejected at the nominal level 0.05. A stands for the asymptotic test and B stands for the bootstrap test.

Data Set	n	$l_2\left(p_n, e\left(p_n\right)\right)$	$d\left(p_n, \mathcal{M}\right)$	T_c A	T_c B	T_m A	T_m B
1	230	0.102	0.101	0.130	0.134	0.130	0.132
2	229	0.126	0.118	0.159	0.164	0.149	0.153
3	8295	0.013	0.013	0.017	0.019	0.018	0.018

3.2. Test Power

In this subsection we study the test power at HWE. We restrict yourself to the genotype distributions at HWE, which are implied by the real data sets from Section 3.1, because the family $\mathcal{M}$ is very large. To shed some light on the appropriate values of the tolerance parameter ε, the test power is computed for different values of ε, see Table 3. The value of ε may be considered appropriate if the test power is approximately 0.9. Hence, the appropriate value of ε is 0.1 for data set 1, 0.1 for data set 2 and 0.018 for data set 3.

The observed genotype frequencies p_n are subjected to the sampling error. It is important for the test efficiency that the sampling error has a small influence on the test power at HWE. The test power is computed at 100 random genotype frequencies, where the corresponding random samples of size n are drawn from the implied genotype distribution $e(p_n)$. The simulation results are summarized in Table 4. The power of all considered tests varies little from point to point. Hence, the impact of the sampling error on the test power at HWE is very small.

Table 3. Simulated rejection probability of the equivalence tests at the nominal level 0.05. The rejection probability is simulated for different values of the tolerance parameter ε at the HWE distributions $e(p_n)$, which are implied by data sets 1, 2 and 3. The sample size equals the size of the corresponding data set. The number of replications is 1000 for each experiment. A stands for the asymptotic test and B stands for the bootstrap test.

	Data set 1				**Data set 2**				**Data set 3**			
ε	**0.07**	**0.08**	**0.09**	**0.10**	**0.07**	**0.08**	**0.09**	**0.10**	**0.012**	**0.014**	**0.016**	**0.018**
T_c, A	0.56	0.75	0.87	0.95	0.53	0.74	0.87	0.94	0.67	0.82	0.91	0.96
T_c, B	0.40	0.63	0.79	0.90	0.40	0.62	0.80	0.90	0.50	0.71	0.83	0.91
T_m, A	0.54	0.74	0.87	0.94	0.58	0.78	0.89	0.96	0.61	0.77	0.88	0.95
T_m, B	0.49	0.70	0.85	0.93	0.53	0.74	0.86	0.94	0.58	0.75	0.86	0.94

Table 4. Summary of the simulated rejection probabilities at the nominal level 0.05. The rejection probabilities are simulated at the 100 random samples from the HWE distributions $e(p_n)$, which are implied by data sets 1, 2 and 3. The sample size equals the size of the corresponding data set. The number of replications is 1000 for each experiment. A stands for the asymptotic test and B stands for the bootstrap test.

	Data set 1, $\varepsilon = 0.1$				**Data set 2,** $\varepsilon = 0.1$				**Data set 3,** $\varepsilon = 0.018$			
	min	**max**	**mean**	**dev**	**min**	**max**	**mean**	**dev**	**min**	**max**	**mean**	**dev**
T_c, A	0.93	0.97	0.95	0.008	0.93	0.97	0.94	0.008	0.94	0.98	0.97	0.007
T_c, B	0.87	0.92	0.90	0.010	0.88	0.94	0.90	0.012	0.89	0.94	0.91	0.010
T_m, A	0.92	0.96	0.95	0.009	0.93	0.99	0.96	0.012	0.92	0.97	0.95	0.007
T_m, B	0.90	0.95	0.93	0.010	0.91	0.99	0.95	0.015	0.91	0.96	0.94	0.009

3.3. Type I error

We study the type I error rates of the proposed tests in this subsection. The boundary of H_0 is so complex that it is very difficult to find boundary points, which have the largest rejection probability. We consider therefore randomly selected boundary points of H_0, which are based on the three real data sets from Section 3.1. The boundary points are generated using the following algorithm:

1. Given are p_n and ε.
2. Draw a sample of size n from p_n and compute the sample genotype frequency $\widetilde{p}_n$.
3. If $T_*(\widetilde{p}_n) < 0$ then reject $\widetilde{p}_n$ and repeat step 2. Otherwise accept $\widetilde{p}_n$.
4. Consider the linear combination $a\widetilde{p}_n + (1-a)\,e(p_n)$ for $a \in [0,1]$. Find $a_n \in [0,1]$ such that $T_*(a_n\widetilde{p}_n + (1-a_n)\,e(p_n)) = 0$. The value of a_n can be found using any line search method.
5. Return $a_n\widetilde{p}_n + (1-a_n)\,e(p_n)$, which is a random boundary point of H_0.

The tolerance parameter ε is close to $l_2\left(p_n, e\left(p_n\right)\right)$ for each data set under consideration so that a_n is usually not far from 1. The corresponding random boundary point is then close to p_n. Hence, we explore the boundary of H_0 in the neighborhood of the given data set. The test power at 100 random boundary points is summarized in Table 5. The test power varies considerable from point to point. The asymptotic test based on T_c is not conservative for all three data sets. The asymptotic test based on T_m shows some anti-conservative tendencies for data sets 2 and 3. The bootstrap test based on T_c is conservative for all three data sets. The bootstrap test based on T_m shows slight non conservative tendencies.

The power at the boundary points is larger than the nominal level due to the following reasons. If the number of observations $np_n\left(i,j\right)$ is too small for some i and j then the distribution of $T_*\left(p_n\right)$ may be far away from the normal approximation and also may have considerable jumps. The critical values of the asymptotic and bootstrap tests are then incorrect. If the vector $\mathring{d}_*\left(v\left(p_n\right)\right)$ contains zero elements then the power of the asymptotic tests tends to be above the nominal level α. The power of the bootstrap tests is closer to α in this case because the vector $\mathring{d}_*\left(v\left(p_n\right)\right)$ is not used for the variance estimation by the bootstrap method.

Table 5. Summary of the simulated rejection probabilities at the nominal level 0.05. The rejection probabilities are simulated at the 100 randomly selected boundary points of H_0. The sample size equals the size of the corresponding data set. The number of replications is 1000 for each experiment. A stands for the asymptotic test and B stands for the bootstrap test.

	Data set 1, $\varepsilon = 0.1$				**Data set 2, $\varepsilon = 0.1$**				**Data set 3, $\varepsilon = 0.018$**			
	min	**max**	**mean**	**dev**	**min**	**max**	**mean**	**dev**	**min**	**max**	**mean**	**dev**
T_c, A	0.017	0.060	0.035	0.009	0.023	0.065	0.044	0.009	0.051	0.114	0.090	0.011
T_c, B	0.005	0.036	0.018	0.006	0.013	0.048	0.029	0.006	0.012	0.052	0.033	0.008
T_m, A	0.019	0.049	0.031	0.006	0.025	0.077	0.046	0.009	0.034	0.075	0.051	0.008
T_m, B	0.011	0.042	0.022	0.006	0.016	0.064	0.034	0.008	0.025	0.064	0.038	0.008

4. Summary

Two different test statistics are proposed to establish equivalence of the genotype distributions to HWE. The critical values of the tests are calculated using the asymptotic approximation by the normal distribution. The variance of the test statistic is estimated asymptotically or by the bootstrap method. The minimum tolerance parameter ε, for which H_0 can be rejected, is derived. The tests are successfully applied to three real data sets, which are frequently considered in the literature. The test power at HWE and the type I error rates are studied at a large number of points, which are inspired by the real data sets. The asymptotic tests have anti-conservative tendencies and should be used with caution. The bootstrap-based tests are sufficiently conservative for the most practical situations. If more conservative tests are required then the nominal level may be halved or the tolerance parameter ε may be reduced. We recommend to perform all proposed tests in any case and compare the results. The appropriate value of ε depends on the application and the available sample size. The reasonable values of the parameter ε can be found by simulation as shown in Section 3.2. Additionally, the rejection probabilities at the close random boundary points may be studied as shown in Section 3.3.

Funding: This research received no external funding.

Acknowledgments: The author would like to thank the editors and reviewers for the excellent work.

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

References

1. Levene, H. On a matching problem arising in genetic. *Ann. Math. Stat.* **1949**, *20*, 91–94. [CrossRef]
2. Haldane, J.B.S. An exact test for randomness of mating. *J. Genet.* **1954**, *52*, 631–635. [CrossRef]
3. Chapco, W. An exact test of the Hardy-Weinberg law. *Biometrics* **1976**, *32*, 183–189. [CrossRef]

4. Louis, E.J.; Dempster, E.R. An exact test for Hardy-Weinberg and multiple alleles. *Biometrics* **1987**, *43*, 805–811. [CrossRef] [PubMed]
5. Guo, S.; Thompson, E. Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics* **1992**, *48*, 361–372. [CrossRef]
6. Wellek, S. Tests for establishing compatibility of an observed genotype distribution with Hardy Weinberg equilibrium in the case of biallelic locus. *Biometrics* **2004**, *60*, 694–703. [CrossRef]
7. Wellek, S. *Testing Statistical Hypotheses of Equivalence and Noninferiority*, 2nd ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2010.
8. Wellek, S.; Goddard, K.; Ziegler, A. A Confidence-Limit-Based Approach to the Assessment of Hardy Weinberg Equilibrium. *Biom. J.* **2010**, *52*, 253–270. [CrossRef]
9. Ostrovski, V. New Equivalence Tests for Approximate Independence in Contingency Tables. *Stats* **2019**, *2*, 239–246. [CrossRef]
10. Ostrovski, V. Testing equivalence to families of multinomial distributions with application to the independence model. *Stat. Probab. Lett.* **2018**, *139*, 61–66. [CrossRef]
11. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*, 1st ed.; Chapman & Hall: New York, NY, USA, 1993.
12. Wordsworth, P.; Pile, K.; Buckley, J.; Lanchbury, J.; Ollier, B.; Lathrop, M.; Bell, J. HLA heterozygosity contributes to susceptibility to rheumatoid arthritis. *Am. J. Hum. Genet.* **1992**, *51*, 585–591. [PubMed]
13. Rousset, F. A complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Res.* **2008**, *8*, 103–106. [CrossRef] [PubMed]
14. Cavalli-Sforza, L.; Bodmer, W. *The Genetics of Human Populations*; W. H. Freeman: San Francisco, CA, USA, 1971.
15. Engels, W.R. Exact Tests for Hardy-Weinberg Proportions. *Genetics* **2009**, *183*, 1431–1441. [CrossRef] [PubMed]

© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).