Statistical Test for Bivariate Uniformity

Zhenmin Chen and Tieyong Hu

Department of Mathematics and Statistics, Florida International University, Miami, FL 33199, USA

Correspondence should be addressed to Zhenmin Chen; chenzh@fiu.edu

Received 31 May 2014; Revised 17 September 2014; Accepted 1 October 2014; Published 19 October 2014

Copyright © 2014 Z. Chen and T. Hu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The purpose of the multidimension uniformity test is to check whether the underlying probability distribution of a multidimensional population differs from the multidimensional uniform distribution. The multidimensional uniformity test has applications in various fields such as biology, astronomy, and computer science. Such a test, however, has received less attention in the literature compared with the univariate case. A new test statistic for checking multidimensional uniformity is proposed in this paper. Some important properties of the proposed test statistic are discussed. As a special case, the bivariate statistic test is discussed in detail in this paper. The Monte Carlo simulation is used to compare the power of the newly proposed test with the distance-to-boundary test, which is a recently published statistical test for multidimensional uniformity. It has been shown that the test proposed in this paper is more powerful than the distance-to-boundary test in some cases.

1. Introduction

Testing uniformity in the univariate case has been studied by many researchers, whereas the multidimensional uniformity test seems to have received less attention in the literature. Testing whether a pattern of points in the multidimensional space is distributed uniformly has applications in many fields such as biology, astronomy, and computer science. A commonly used goodness-of-fit test for uniformity is the chi-square test [1]. Theoretically, the chi-square test can be applied for any multivariate distribution test. However, the problem for the chi-square test is the arbitrariness of cell limits determination. Another problem for the chi-square test is that the power of the chi-square test is usually low. Some other well-known methods for univariate goodness-of-fit tests are the Kolmogorov-Smirnov test [2, 3], Anderson-Darling test [4], and the Cramer-von Mises test [5]. Justel et al. [6] proposed a multivariate goodness-of-fit test based on the idea of the Kolmogorov-Smirnov test. By using the Rosenblatt’s transformation, they reduced the multivariate case to univariate case. The test statistic they used has distribution free property and can be applied to any dimensional case. The problem for that method is that the computation of test statistic is complicated especially for over two dimensions. Liang et al. [7] proposed several statistical tests for testing uniformity in multivariate case. Those tests used the number-theoretic and quasi-Monte Carlo method for measuring the discrepancy of the points in multidimensional unit. Berrendero et al. [8] proposed a test based on the idea of distance to the boundary. It was shown by Monte Carlo simulation that the distance-to-boundary test is more powerful than the tests proposed by Liang et al. [7]. Chen and Ye [9] developed an alternative test for uniformity in univariate case. In that paper, the authors proposed a test statistic based on the order statistics in support set [0, 1]. The test statistic proposed in that paper is

\[
G(X_1,X_2,\ldots,X_n) = \frac{n+1}{n} \sum_{i=1}^{n} \left( \frac{X_{(i)} - X_{(i-1)}}{n+1} \right)^2.
\]

The Monte Carlo simulation results showed that the proposed test in that paper is more powerful compared with the commonly used Kolmogorov-Smirnov test when the alternative distribution is a V-shape distribution or when the sample size is small. By applying the probability integral transformation, the uniformity test can be used to check whether the underlying distribution follows any specified distribution. The idea is adopted in this paper to develop a test for the multidimensional case.
The main purpose of this paper is to propose a new test statistic for testing multidimensional uniformity. It is expected that the newly proposed test may improve the power of the multidimensional uniformity tests. Since the distance-to-boundary test is a recently published test in multivariate analysis, the discussion will be mainly based on the bivariate case. Some techniques used in nonparametric statistics are adopted to modify the test statistic for the purpose of raising the power of the test for the bivariate case.

2. New Test Statistic

Suppose

\[
X_1 = \begin{bmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{1p} \end{bmatrix}, \quad X_2 = \begin{bmatrix} X_{21} \\ X_{22} \\ \vdots \\ X_{2p} \end{bmatrix}, \quad \ldots, \quad X_n = \begin{bmatrix} X_{n1} \\ X_{n2} \\ \vdots \\ X_{np} \end{bmatrix}
\]

form a random sample from a \( p \)-dimensional population distribution with support set \([0,1]^p\). Here \([0,1]^p\) is the \( p \)-dimensional unit cube which is the set defined as

\[
\{ t_1, t_2, \ldots, t_p : 0 \leq t_i \leq 1 \ (i = 1, 2, \ldots, p) \}.
\]

Suppose also that \( X_{(1)1}, X_{(2)1}, \ldots, X_{(n)1} \) are the ordered values of \( X_{11}, X_{21}, \ldots, X_{n1} \) \((i = 1, 2, \ldots, p)\). The purpose is to test the following:

\[ H_0: \text{the population distribution is a uniform distribution on } [0,1]^p; \]

\[ H_1: \text{the population distribution is not uniform distribution on } [0,1]^p. \]

The test statistic proposed in this paper is defined as

\[
G_p (X_1, X_2, \ldots, X_n)
\]

\[
= \frac{(n+1)^p}{(n+1)^p - 1} \times \sum_{i_1=1}^{n+1} \sum_{i_1=1}^{n+1} \cdots \sum_{i_p=1}^{n+1} \left[ \prod_{j=1}^{p} (X_{(j)ij} - X_{(j-1)ij}) - \frac{1}{(n+1)^p} \right]^2.
\]

Here it is assumed that \( X_{(0)j} = 0 \) and \( X_{(n+1)j} = 1 \) \((j = 1, 2, \ldots, p)\). It can be seen that if the underlying distribution is a uniform distribution on \([0,1]^p\), then

\[
E \left[ \prod_{j=1}^{p} (X_{(j)ij} - X_{(j-1)ij}) \right] = \frac{1}{(n+1)^p} \quad (i_j = 1, 2, \ldots, n+1 \ (j = 1, 2, \ldots, p)).
\]

Therefore, if the value of \( G_p (X_1, X_2, \ldots, X_n) \) is too far away from zero, it could be an indication that the underlying distribution is not uniform distribution on \([0,1]^p\). This motivates the following test procedure. Under \( H_0 \), let \( G_{p,1-\alpha} \) be a number such that

\[
P \left( G_p (X_1, X_2, \ldots, X_n) > G_{p,1-\alpha} \right) = \alpha \quad (0 < \alpha < 1).
\]

Then \( H_0 \) should be rejected at significance level \( \alpha \) if \( G_p (X_1, X_2, \ldots, X_n) > G_{p,1-\alpha} \). It can be shown that \( G_p (X_1, X_2, \ldots, X_n) \) is always between 0 and 1. In fact,

\[
\sum_{i_1=1}^{n+1} \sum_{i_1=1}^{n+1} \cdots \sum_{i_p=1}^{n+1} \left[ \prod_{j=1}^{p} (X_{(j)ij} - X_{(j-1)ij}) \right]^2 - \frac{1}{(n+1)^p} + \frac{1}{(n+1)^p}
\]

\[
= \sum_{i_1=1}^{n+1} \sum_{i_1=1}^{n+1} \cdots \sum_{i_p=1}^{n+1} \left[ \prod_{j=1}^{p} (X_{(j)ij} - X_{(j-1)ij}) \right]^2 - \frac{1}{(n+1)^p}
\]

\[
\leq \sum_{i_1=1}^{n+1} \sum_{i_1=1}^{n+1} \cdots \sum_{i_p=1}^{n+1} \left[ \prod_{j=1}^{p} (X_{(j)ij} - X_{(j-1)ij}) \right]^2 - \frac{1}{(n+1)^p}.
\]

It can be found from above that \( G_p (X_1, X_2, \ldots, X_n) \) can also be rewritten as

\[
G_p (X_1, X_2, \ldots, X_n)
\]

\[
= \frac{(n+1)^p}{(n+1)^p - 1} \times \sum_{i_1=1}^{n+1} \sum_{i_1=1}^{n+1} \cdots \sum_{i_p=1}^{n+1} \left[ \prod_{j=1}^{p} (X_{(j)ij} - X_{(j-1)ij}) \right]^2 - \frac{1}{(n+1)^p}.
\]

As mentioned above, this paper will mainly discuss the bivariate case. Suppose

\[
X_1 = \begin{bmatrix} X_{11} \\ X_{12} \end{bmatrix}, \quad X_2 = \begin{bmatrix} X_{21} \\ X_{22} \end{bmatrix}, \quad \ldots, \quad X_n = \begin{bmatrix} X_{n1} \\ X_{n2} \end{bmatrix}
\]

form a random sample from a bivariate population distribution with support set \([0,1] \times [0,1]\). Suppose also that \( X_{(1)1}, X_{(2)1}, \ldots, X_{(n)1} \) are the ordered values of
Advances in Statistics

\[X_{11}, X_{21}, \ldots, X_{n1}\] and \[X_{12}, X_{22}, \ldots, X_{n2}\] are the ordered values of \[X_{12}, X_{22}, \ldots, X_{n2}\]. The purpose is to test the following:

\[H_0; \text{the population distribution is a uniform distribution on } [0, 1] \times [0, 1];\]

\[H_a; \text{the population distribution is not uniform distribution on } [0, 1] \times [0, 1].\]

In order to raise the power of the test, the test statistic defined in (4) is adjusted by adopting the Kendall's \(r\) statistic. See, for example, Conover [10]. For any pair of points \([X_{ij}, X_{i'j'}]\) and \([X_{ij'}, X_{i'j}]\) in the two dimensional space, define the following items.

\([X_{ij}, X_{i'j'}]\) and \([X_{ij'}, X_{i'j}]\) are said to be concordant if \((X_{j2} - X_{j1})/(X_{i2} - X_{i1}) > 0.\)

\([X_{ij}, X_{i'j'}]\) and \([X_{ij'}, X_{i'j}]\) are said to be discordant if \((X_{j2} - X_{j1})/(X_{i2} - X_{i1}) < 0.\)

\([X_{ij}, X_{i'j'}]\) and \([X_{ij'}, X_{i'j}]\) are said to be half concordant and half discordant if \(X_{j2} - X_{j1} = 0.\)

No comparison is made if \(X_{j2} - X_{j1} = 0.\)

Let \(n_c\) be the total number of concordant pairs and let \(n_d\) be the total number of discordant pairs.

Define

\[G_2(X_{11}, X_{21}, \ldots, X_{n1}) = \frac{(n+1) \left( \begin{array}{c} n_c \times n_d + 1 \end{array} \right)}{n(n+2) \left( \begin{array}{c} n_c + n_d + 1 \end{array} \right)} \cdot \sum_{i,j=1}^{n+1} \sum_{i=1}^{n+1} \left( (X_{i(i-1)} - X_{j(i-1)}) \left( X_{j(i2) - X_{j(i-1)}} \right) \right)^2. \]

(10)

Then \(H_0\) should be rejected at significance level \(\alpha\) if \(G_2(X_{11}, X_{21}, \ldots, X_{n1}) > G_{2,1-\alpha}\).

It can be shown that \(0 \leq G_2(X_{11}, X_{21}, \ldots, X_{n1}) \leq 1\) for any \(X_{11}, X_{21}, \ldots, X_{n1} \in [0, 1] \times [0, 1].\)

To show \(G_2(X_{11}, X_{21}, \ldots, X_{n1}) \leq 1\), note that

\[\sum_{i,j=1}^{n+1} \sum_{i=1}^{n+1} \left( (X_{i(i-1)} - X_{j(i-1)}) \left( X_{j(i2) - X_{j(i-1)}} \right) \right)^2 = \frac{1}{(n+1)^2} \sum_{i,j=1}^{n+1} \sum_{i=1}^{n+1} \left( (X_{i(i-1)} - X_{j(i-1)}) \left( X_{j(i2) - X_{j(i-1)}} \right) \right)^2 = \frac{1}{(n+1)^2}. \]

Combining the above result with the fact that \((|n_c - n_d| + 1)/(n_c + n_d + 1)\) is always between 0 and 1, one can then conclude that \(0 \leq G_2(X_{11}, X_{21}, \ldots, X_{n1}) \leq 1\).

It should be mentioned that the lower and upper bounds of the above inequality cannot be improved. In fact, one may construct bivariate data sets easily such that the values of \(G_2(X_{11}, X_{21}, \ldots, X_{n1})\) will reach 0 and 1, respectively.

### 3. Critical Values and Power Comparison

Monte Carlo simulation is used to find the critical values of the test statistic described in (10). To accomplish this, \(k = 10,000,000\) pseudo random samples of size \(n\) are generated from the two-dimensional uniform distribution on \([0, 1] \times [0, 1]\) for \(n = 5, 6, \ldots, 50\). The critical values of the test statistic are tabulated in Table 1. The first column is for the sample sizes and the first row is for the significance levels. The values inside the tables are the critical values corresponding to the sample size and the significance level.

The power of the test statistic proposed in this paper is compared with the recently published distance-to-boundary test by Berrendero et al. [8]. This is because the distance-to-boundary test has been shown to possess good performance in many cases. For convenience, the test statistic proposed in this paper is denoted as the \(G_2\) test for the rest of the paper. Several alternative distributions are selected for the power comparison purpose. The selected alternative distributions can be classified into two types. The first type of alternative
Table 1: Critical values of $G_2$ test statistic.

| n  | $G_{2,0.1}$ | $G_{2,0.05}$ | $G_{2,0.025}$ | $G_{2,0.01}$ |
|----|-------------|-------------|---------------|-------------|
| 5  | 0.05194     | 0.06722     | 0.08300       | 0.10584     |
| 6  | 0.03555     | 0.04323     | 0.05385       | 0.06894     |
| 7  | 0.02321     | 0.02997     | 0.03695       | 0.04726     |
| 8  | 0.01676     | 0.02154     | 0.02678       | 0.03430     |
| 9  | 0.02125     | 0.01610     | 0.01990       | 0.02562     |
| 10 | 0.00973     | 0.01250     | 0.01525       | 0.01955     |
| 11 | 0.00773     | 0.00984     | 0.01210       | 0.01509     |
| 12 | 0.00618     | 0.00790     | 0.00972       | 0.01212     |
| 13 | 0.00515     | 0.00657     | 0.00802       | 0.01002     |
| 14 | 0.00426     | 0.00540     | 0.00657       | 0.00829     |
| 15 | 0.00359     | 0.00455     | 0.00552       | 0.00698     |
| 16 | 0.00308     | 0.00388     | 0.00468       | 0.00585     |
| 17 | 0.00265     | 0.00334     | 0.00404       | 0.00502     |
| 18 | 0.00231     | 0.00292     | 0.00356       | 0.00439     |
| 19 | 0.00201     | 0.00254     | 0.00308       | 0.00379     |
| 20 | 0.00177     | 0.00222     | 0.00270       | 0.00333     |
| 21 | 0.00158     | 0.00200     | 0.00241       | 0.00298     |
| 22 | 0.00140     | 0.00176     | 0.00212       | 0.00262     |
| 23 | 0.00126     | 0.00158     | 0.00191       | 0.00236     |
| 24 | 0.00113     | 0.00141     | 0.00169       | 0.00211     |
| 25 | 0.00102     | 0.00127     | 0.00153       | 0.00188     |
| 26 | 0.00093     | 0.00116     | 0.00139       | 0.00171     |
| 27 | 0.00085     | 0.00106     | 0.00127       | 0.00156     |
| 28 | 0.00077     | 0.00096     | 0.00114       | 0.00139     |
| 29 | 0.00071     | 0.00088     | 0.00106       | 0.00129     |
| 30 | 0.00065     | 0.00081     | 0.00096       | 0.00117     |
| 31 | 0.00060     | 0.00074     | 0.00089       | 0.00108     |
| 32 | 0.00055     | 0.00069     | 0.00082       | 0.00099     |
| 33 | 0.00052     | 0.00064     | 0.00077       | 0.00094     |
| 34 | 0.00048     | 0.00059     | 0.00070       | 0.00085     |
| 35 | 0.00044     | 0.00055     | 0.00065       | 0.00079     |
| 36 | 0.00041     | 0.00051     | 0.00061       | 0.00075     |
| 37 | 0.00039     | 0.00048     | 0.00057       | 0.00070     |
| 38 | 0.00036     | 0.00045     | 0.00053       | 0.00064     |
| 39 | 0.00034     | 0.00042     | 0.00050       | 0.00060     |
| 40 | 0.00032     | 0.00040     | 0.00047       | 0.00057     |
| 41 | 0.00030     | 0.00037     | 0.00044       | 0.00053     |
| 42 | 0.00028     | 0.00035     | 0.00041       | 0.00050     |
| 43 | 0.00027     | 0.00033     | 0.00039       | 0.00047     |
| 44 | 0.00025     | 0.00031     | 0.00037       | 0.00044     |
| 45 | 0.00024     | 0.00029     | 0.00035       | 0.00041     |
| 46 | 0.00023     | 0.00028     | 0.00033       | 0.00040     |
| 47 | 0.00021     | 0.00026     | 0.00031       | 0.00037     |
| 48 | 0.00020     | 0.00025     | 0.00029       | 0.00035     |
| 49 | 0.00019     | 0.00024     | 0.00028       | 0.00034     |
| 50 | 0.00018     | 0.00022     | 0.00026       | 0.00032     |

bivariate distributions is based on two independent univariate Beta distributions. The second type of alternative distributions is, so called, the metatype uniform distribution.

3.1. Alternative Bivariate Distributions Based on Independent Beta Distribution. The probability density function of the univariate Beta distribution is

$$f(x) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} & 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

The Beta distribution family is quite flexible to get different shapes by selecting parameters $\alpha$ and $\beta$. The bivariate alternative distributions used in the power comparison are formed by two independent univariate Beta distributions.

3.1.1. Alternative Distribution 1. The bivariate Beta distribution is formed by two independent Beta $(5,2)$ marginal distributions. Figure 1 shows the power comparison between the $G_2$ test and distance-to-boundary test. It can be seen that the $G_2$ test is more powerful than the distance-to-boundary test in this case.

3.1.2. Alternative Distribution 2. The bivariate Beta distribution is formed by two independent Beta $(5,1)$ marginal distributions. Figure 2 shows the power comparison between the $G_2$ test and distance-to-boundary test. It can be seen that the $G_2$ test is more powerful when sample size is less than 25. When the sample size increases, the power of two tests is pretty close. The distance-to-boundary test is slightly better.

3.1.3. Alternative Distribution 3. The bivariate Beta distribution is formed by two independent Beta $(0.5,0.5)$ marginal distributions. Figure 3 shows the power comparison between the $G_2$ test and distance-to-boundary test. It can be seen that the distance-to-boundary test performs better in this case. This is probably because the Beta $(0.5,0.5)$ is a symmetric distribution. The symmetric situation is discussed in Berrendero et al. [8]. It has been shown that the power of distance-to-boundary is higher in this case. After changing the symmetric condition, the result seems different.
3.2. Metatype Uniform Distribution. The Metatype uniform distribution was mentioned in the papers of Liang et al. [7] and Berrendero et al. [8]. They introduced this distribution for the power comparison purpose.

The basic idea for creating metatype multivariate distribution is as follows. Let the random vector $X = (X_1, X_2)$ have a distribution function $F(X)$. We define $F_1(X_1)$ as the marginal distribution function of $X_1$ and $F_2(X_2)$ as the marginal distribution function of $X_2$. Then define random vector $Y$ as $Y = (F_1(X_1), F_2(X_2))^T$. As we know, $F_1(X_1)$ and $F_2(X_2)$ are uniformly distributed in the support set $[0,1]$ and the joint distribution is different from the uniform distribution since $F_1(X_1)$ and $F_2(X_2)$ are not independent. This kind of multivariate distribution is easily generated by any software and is useful to check the multivariate uniform distribution. Specifically, we have considered two of the metatype uniform distributions in power study.

3.2.1. Alternative Distribution 4. MNU is obtained from bivariate normal distribution with $\mu = [0, 0]$ and $\Sigma = [1, 0.5; 0.5, 1]$ and 5 degrees of freedom. The power comparison result under such a metatype uniform distribution is shown in Figure 4. The $G_2$ test is more powerful than distance-to-boundary test in this case. When the sample size increases, the power of the $G_2$ test increases and the power of distance-to-boundary test does not change too much.

3.2.2. Alternative Distribution 5. MTU is obtained from bivariate Student’s-$t$ distribution with $\mu = [0, 0]$ and $\Sigma = [1, 0.5; 0.5, 1]$ and 5 degrees of freedom. The power comparison result under such a metatype uniform distribution is shown in Figure 5. The $G_2$ test is more powerful than distance-to-boundary test in this case. When the sample size increases, the power of the $G_2$ test increases and the power of distance-to-boundary test does not change too much.

4. Conclusion and Discussion

In this paper, the new multidimensional uniformity test is proposed. The basic idea is from univariate uniform distribution test in the paper of Chen and Ye [9]. The method is extended to the multidimensional case and the bivariate case is discussed in detail. The new test can be used to test whether an underlying multivariate probability distribution differs from a uniform distribution. The critical value of bivariate uniformity test is calculated and the power study performed by comparing with the recently published multivariate uniformity test.

The distance-to-boundary is a recently published multivariate uniformity test by Berrendero et al. [8]. The result of
the power comparison shows that the new test proposed in this paper is more powerful than the distance-to-boundary test in some cases. In particular, when the marginal distribution of alternative distribution is not symmetric and all the marginal distributions are independent and identically distributed, the new test is more powerful. The metatype uniform distribution is used in this paper for power comparison purpose. The metatype uniform distributions are generated from the normal distribution and the Student’s-\(t\) distribution. The power study shows that new test is more powerful than distance-to-boundary test in this case.

**Conflict of Interests**

The authors declare that there is no conflict of interests regarding the publication of this paper.

**Acknowledgment**

The authors sincerely thank two anonymous referees for their comments and suggestions that greatly improved the quality of the paper.

**References**

[1] E. S. Pearson, “The probability transformation for testing goodness of fit and combining independent tests of significance,” *Biometrika*, vol. 30, pp. 134–148, 1939.

[2] A. N. Kolmogorov, “Sulla determinazione empirica di una legge di distribuzione,” *Giornale dell’Istituto Italiano degli Attuari*, vol. 4, pp. 83–91, 1933.

[3] N. V. Smirnov, “Estimate of deviation between empirical distributions,” *Bulletin Moscow University*, vol. 2, pp. 3–16, 1939 (Russian).

[4] T. W. Anderson and D. A. Darling, “A test of goodness of fit,” *Journal of the American Statistical Association*, vol. 49, pp. 765–769, 1954.

[5] H. Cramér, “On the composition of elementary errors,” *Skandinavisk Aktuarietidskrift*, vol. 11, pp. 13–74, 1928.

[6] A. Justel, D. Peña, and R. Zamar, “A multivariate Kolmogorov-Smirnov test of goodness of fit,” *Statistics and Probability Letters*, vol. 35, no. 3, pp. 251–259, 1997.

[7] J.-J. Liang, K.-T. Fang, F. J. Hickernell, and R. Li, “Testing multivariate uniformity and its applications,” *Mathematics of Computation*, vol. 70, no. 233, pp. 337–355, 2001.

[8] J. R. Berrendero, A. Cuevas, and F. Vjosáquez-grande, “Testing multivariate uniformity: the distance-to-boundary method,” *The Canadian Journal of Statistics*, vol. 34, no. 4, pp. 693–707, 2006.

[9] Z. Chen and C. Ye, “An alternative test for uniformity,” *International Journal of Reliability, Quality and Safety Engineering*, vol. 16, no. 4, pp. 343–356, 2009.

[10] W. J. Conover, *Practical Nonparametric Statistics*, Wiley, 3rd edition, 1999.
