Fall 1-31-2013

The goodness-of-fit tests for geometric models

Feiyan Chen
New Jersey Institute of Technology

ABSTRACT

THE GOODNESS-OF-FIT TESTS FOR GEOMETRIC MODELS

by
Feiyan Chen

We propose two types of goodness-of-fit tests for geometric distribution and for a bivariate geometric distribution called BGD(B&D), based on their probability generating function (PGF). The first type is a special-case application of the general testing procedure for discrete distributions proposed by Kocherlakota and Kocherlakota (1986). The second type utilizes the supremum of the absolute value of the standardized difference between the PGF's maximum likelihood estimator (MLE) and its empirical counterpart as the test statistic. We verify the asymptotic properties of the test statistics for the first type of test and explore the asymptotic behaviors of the test statistics for the second type of test by calculating the empirical critical points and constructing the density curves. We compare the proposed tests with Chi-square and the empirical distribution function (EDF) related tests proposed in the literature in terms of significance level and power. Based on the comparison results, we recommend the second type of goodness-of fit test for both geometric distribution and BGD(B&D) because of its robustness, efficiency in computation and no need for selecting t. Real data sets are used for illustration.

THE GOODNESS-OF-FIT TESTS FOR GEOMETRIC MODELS

by
Feiyan Chen

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology and
Rutgers, The State University of New Jersey – Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Mathematical Sciences

Department of Mathematical Sciences, NJIT
Department of Mathematics and Computer Science, Rutgers-Newark

January 2013

APPROVAL PAGE

THE GOODNESS-OF-FIT TESTS FOR GEOMETRIC MODELS

Feiyan Chen

Sunil K. Dhar, PhD, Dissertation Advisor — Date
Associate Professor, Department of Mathematical Sciences, NJIT

Sundarraman Subramanian, PhD, Committee Member — Date
Associate Professor, Department of Mathematical Sciences, NJIT

Aridaman K. Jain, PhD, Committee Member — Date
Senior University Lecturer, Department of Mathematical Sciences, NJIT

Wenge Guo, PhD, Committee Member — Date
Assistant Professor, Department of Mathematical Sciences, NJIT

Ganesh K. Subramanian, PhD, Committee Member — Date
Researcher, Research/ AT&T Lab

BIOGRAPHICAL SKETCH

Author: Feiyan Chen

Degree: Doctor of Philosophy

Date: January 2013

Undergraduate and Graduate Education:

- Doctor of Philosophy in Mathematical Sciences,
  New Jersey Institute of Technology, Newark, NJ, 2013
- Master of Applied Statistics,
  New Jersey Institute of Technology, Newark, NJ, 2009
- Master of Computer Sciences,
  Xi'an Jiaotong University, Xi'an, China, 2005
- Bachelor of Computer Sciences,
  Xi'an Polytechnic University, Xi'an, China, 2002

Major: Mathematical Sciences

Presentations and Publications:

Feiyan Chen, "Probability Generating Function based Goodness-of-fit Tests for Geometric Distribution," *57th Annual Meeting of New Jersey Academy of Science*, Seton Hall University, South Orange, April 21, 2012.

Feiyan Chen, "Comparison of Goodness-of-fit Tests for Geometric models," *the Ninth Annual Conference on Frontiers in Applied and Computational Mathematics (FACM '12)*, Department of Mathematical Sciences, NJIT, May 18-20, 2012.

Feiyan Chen, "Numerical Study of the Goodness-of-fit Tests for Geometric Models," *Applied Mathematics Seminar*, Department of Mathematical Sciences, NJIT, June 18, 2012.

Feiyan Chen, "The Goodness-of-fit Tests for BGD(B&D)," *VIIIth Annual Graduate Student Research Day*, NJIT, November 12, 2012.

Three Rules of Work: Out of clutter find simplicity. From discord find harmony. In the middle of difficulty lies opportunity.

—Albert Einstein

ACKNOWLEDGMENT

First of all, I would like to thank my dissertation advisor, Dr. Sunil K. Dhar for his tremendous guidance with this dissertation. Ever since I joined the PhD program of the Department of Mathematical Sciences in NJIT, he has been a steadfast source of information, ideas, support, and energy. I am deeply grateful for the patient, encouraging and resourceful guidance he has been giving me throughout the last few years, and I sincerely wish for him all the best in all his future endeavors.

I also would like to extend my gratitude to the other members in my committee for their kind encouragement and support to me throughout all these years while I was in the graduate school as well as throughout the process of working on my dissertation: Dr. Sundarraman Subramanian, Dr. Aridaman K Jain, Dr. Wenge Guo, and Dr. Ganesh K Subramanian. I would like to thank Professor Jonathan H. Luke, Chair of the Department of Mathematical Sciences, for his support.

I also want to thank all the other faculty, staff members and graduate students in the Department of Mathematical Sciences for all the help they have been giving me in the last four years. It has been a great honor to me to get to know all of them.

Last, but not the least, I would like to give the most sincere gratitude to my beloved husband, Peixin Zhang and my adorable son, Eric Zhang. While all has been encouraging and enlightening, I simply would not have accomplished even a little bit without the love, support and understanding from my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

In the literature regarding goodness-of-fitness test, most work is focused on continuous distributions. Here are some statistical transform based tests for normality. Epps and Pulley (1983) presented a test for normality based on its empirical characteristic function. Meintanis (2010) constructed the test for the skew-normal distribution based on its moment generating function (MGF). Zghoul (2010) proposed a comparatively higher power test for normality based on the true MGF and its empirical counterpart. Below is some work regarding the goodness-of-fit tests for some other continuous distributions. Csörgő (1989) developed a parameter independent method to test univariate and multivariate exponential distributions based on their empirical moment generating functions (EMGF). Meintanis (2004) advocated two tests for logistic distribution based on its empirical characteristic function and its EMGF, respectively. Kallioras et al. (2006) presented tests for two-parameter and three-parameter gamma distributions based on their MGF's.

Relative to continuous distributions, the work on developing hypothesis testing procedures for discrete distributions is rather limited. Among this limited amount of work, most are focused on testing Poisson distribution. Rueda et al. (1991) proposed a PGF related L2 test whose test statistic resembled Cramer-von Mises statistic and illustrated by Poisson distribution. Baringhaus and Henze (1992) derived a test statistic based on the property of the first differential equation of the PGF of Poisson distribution. Motivated by the fact that the second derivative of nature logarithm of Poisson PGF is equal to 0, Nakamura and Perez-Abreu (1993b) proposed a transformed test statistic which was shown approximately independent of the parameter. Epps (1995) developed a generalized test based on PGF for lattice distributions with particular illustrations on Poisson distribution. Henze (1996) and Klar (1999), among others, proposed methods for testing Poisson distribution based on its EDF. Other popularly used tests for Poisson distribution include Neyman's smooth tests, see Rayner and Best (1989), and Fisher's index of dispersion tests, e.g., Gart and Pettigrew

(1970) and Böhning (1994). Karlis and Xekalaki (2000) and Gürtler and Henze (2000), on separated work, accessed various types of tests for Poisson distribution via simulation studies and provided useful guides on their usage in practice.

Besides Poisson distribution, testing the goodness-of-fit for other discrete distributions is also of interest. Geometric distribution is one of the most important discrete distributions and has many useful applications. One of its important applications is on modeling discrete-time queuing, particularly in computer systems and Asynchronous Transfer Mode (ATM) telephone systems. For instance, Conti and Giovanni (2002) used geometric distribution to fit the inter-arrival time (number of time slots between two consecutive arrivals) of cell in ATM systems. Other important applications of geometric distribution can be found in survival analysis and reliability area largely by virtue of its well-known property of being the discrete analog of exponential distribution. As a special case of the geometric model, bivariate geometric distribution has also been studied in the literature. Hawkes (1971) introduced the most natural generalization to arrive at bivariate geometric distribution. Basu and Dhar (1995) derived an existing bivariate geometric distribution called BGD(B&D), which is a discrete analog of the Marshall-Oklin bivariate exponential distribution, see [26]. In particular, BGD(B&D) has been shown to have many useful applications, For example, BGD(B&D) models the life time to failure of a dual-component system, such as paired eyes and paired engines on airplanes, and also can be applied to compare two competing brands of a product and competition scores given by two different judges.

Under the setting of geometric models including univariate and bivariate geometric distributions, most work done so far is focused on investigating or deriving the distributional properties and characterizations. However, regarding the goodness-of-fit tests for geometric models, almost all the investigation concentrated on univariate case. Best and Rayner (1989) derived a Neyman-type smooth test based on Meixner orthonormal polynomials and recommended the test statistic of order four which is asymptotically Chi-square distributed with four degrees of freedom. Best and Rayner (2003) extended their work by comparing the powers of various tests, including Chi-square, smooth, Kolmogorov-Smirnov(K-S) and Anderson-Darling(A-D) tests, and suggested to use A-D for geometric distribution, since it performs well against all types of alternative distributions: underdispersed, equally dispersed and overdispersed. Furthermore, they recommended a data dependent Chernoff-Lehmann χ^2 test with the number of classes as large as possible and expected value for each class greater than unity. Conti (1997) constructed a method for testing geometric distribution against other lattice distributions with monotone hazard function, and applied it on modeling the arrival time of discrete-time queuing system. Given the amount and the quality of work done in the literature, we think that further effort is still worthwhile in order to find tests for univariate geometric distribution that are more powerful and more computationally efficient.

Relative to univariate discrete distributions, even fewer work has been done on the goodness-of-fit tests for its bivariate counterparts. The most widely used methods to test bivariate discrete distributions are Chi-square and two-dimensional K-S tests, despite their respective drawbacks. For Chi-square test, it is required to categorize the data into finite number of groups, a process that ignores the difference from elements in the same group, and thus leads to loss of information. Multi-dimensional K-S test is expectedly more powerful than Chi-square test since it takes the order of the data into account, especially when sample size is small. However, multi-dimension K-S test usually is computationally demanding. Particularly, if d denotes the number of dimensions, there are 2^{d-1} independent ways to define a cumulative distribution function of the test statistic, which makes it computationally challenging to adapt one-dimension K-S to a high-dimension case. Peacock (1983) developed a method by partitioning n data points into $4n^2$ quadrants and calculating the maximum absolute difference between the cumulative probability and data fraction for each quadrant, a process that was stated to be computationally demanding and time consuming. Based on Peacock's work, Fasano and Franceschini (1987) proposed a generalized version of K-S tests and considered $4n$ quadrants instead in the two-dimensional K-S test, which relatively improves the computational complexity but still leaves it with the order of $O(n^2)$. Therefore, it is of interest to explore a hypothesis testing procedure for BGD(B&D) that achieves both high power and efficiency in computation.

Our purpose in this research is to develop improved goodness-of-fit hypothesis tests for both univariate and bivariate geometric distributions, especially based on statistical transforms, which are popularly used for the purpose of making statistical inferences. Many authors have made efforts on developing testing procedures based on the statistical transforms, among those PGF, a special case of statistical transforms, is widely used largely due to its unique features such as simplicity and being a real valued continuous function which always exists in C[0,1] (see [29]). PGF has been particularly applied in dealing with goodness-of-fit tests for discrete distributions on counts, see [23], [34], [2], [30] and [29], among others. Nakamura and Perez-Abreu (1993a) gave an overview of empirical probability generating function (EPGF) and summarized EPGF as a tool for statistical inference of distribution for counts.

PGF was first used to test discrete distribution by Kocherlakota and Kocherlakota (1986). They discussed the general framework for testing goodness-of-fit of univariate and multivariate discrete distributions based on the difference between the MLE of PGF and its empirical counterpart. Their method, referred to as K&K later, was exemplified with Poisson-type distributions and Neyman Type A distribution. They suggest it is convenient to use the method when the number of parameters is small and t, at which PGF is evaluated, is close to zero. Their methods are the generalization of Epps et al. (1982) methods which are in terms of the MGF.

The geometric model goodness-of-fit test especially in the multivariate case has not be explored as much and K&K methods provide a tool to achieve this goal. Further, one of the main strengths of K&K methods is that they can be easily computed. In this paper, we evaluate the performance of K&K methods with single $\underline{t}$ and multiple $\underline{t}'s$ on testing both univariate geometric distribution and BGD(B&D).

The question on how to select t remains unresolved for the K&K tests. It is of interest to investigate in this issue, since as shown in previous and current work, the performance of the K&K tests would be different for different values on t. To resolve this issue, we propose a new type of test using the supremum norm over the region of t based on the PGF, and therefore there is no need in selecting t. By examining the performance of the proposed supremum test, we show that its testing procedure provides comparatively competing power and is computationally more efficient.

CHAPTER 2

REVIEW OF BACKGROUND

In this chapter we will give a brief description of the widely-used goodness-of-fit tests, Chi-square and EDF based tests, which will be compared with the proposed methods for testing geometric models. Next the K&K method goodness-of-fit tests will be introduced in detail.

2.1 Chi-square Tests for Geometric Models

One of the widely-used goodness-of-fit tests for geometric distribution is the Chi-square test. The Chi-square test is based on classifying the sample into categories and measuring the squared distance between the observed frequencies and the expected frequencies under the null distribution. The test statistic of Pearson Chi-square test introduced in elementary statistics course is $\sum_{1\le i\le k}(O_i - E_i)^2/E_i$, where O_i is the observed frequency in the ith category, E_i is the expected frequency in the ith category under the null hypothesis and k is the number of categories. The test statistic approximately follows a χ^2 distribution with $k-1$ degrees of freedom (d.f.), denoted by χ^2_{k-1}.

In most cases, the parameters of the distribution being tested are unknown. One way to estimate the parameters is to use maximum likelihood method. If the MLE is used in place of the true value of the parameters, the test statistic of Chi-square test becomes a Chernoff and Lehmann (1954) statistic, which does not follow an Chi-square distribution asymptotically, but is bounded between two Chi-square distributions. In particular for testing geometric distribution, the Chernoff and Lehmann statistic is bounded between χ^2_{k-2} and χ^2_{k-1}. Some authors just use χ^2_{k-2} for practical purposes provided that the sample size is fairly large and the expected frequencies are not too small. See Snedecor and Cocharan (1989, example 11.7.2, p.205) and Kimber (1987), who discussed the Chernoff and Lehmann Chi-square test on normal distribution. Similarly, when applied to test geometric distribution with unknown parameter given a large sample, the Chernoff and Lehmann Chi-square statistic can be approximated by a Chi-square distributed random variable with $k-2$ degrees of freedom. An alternative way to estimate the parameters is to minimize the Chi-square goodness-of-fit test statistic. More applications of some other Chi-square tests are described in Best and Rayner (2003) and Epps (1995).

The classical one dimensional Chi-square tests can be easily extended to multi-dimensional Chi-square tests. Hence, when testing BGD(B&D) with known parameters, we can use Pearson Chi-square test statistic, which is approximately Chi-square distribution with $k-1$ d.f., and when testing BGD(B&D) with unknown parameters, we can use the Chernoff and Lehmann Chi-square test statistic, which is approximately Chi-square distributed with $k-4$ d.f. (Note the number of parameters for BGD(B&D) is three) given a large sample.

2.2 EDF based Tests for Geometric Models

Another type of popularly-used tests for geometric distribution is EDF based tests such as Kolmogorov-Smirnov (K-S) and Anderson-Darling(A-D) tests. K-S test was originally developed to test the fit of a continuous distribution and the distribution of its test statistic is independent of the hypothesized distribution when the hypothesized distribution is continuous. It is well known that K-S is superior to Chi-square tests when the sample size is small. The test statistic of the classical one dimensional K-S test is the largest absolute value of the difference between the observed sample distribution function and the hypothesized distribution one.

Best and Rayner (2003) used the K-S test for testing geometric distribution described as follows. Let m is the maximum value of the data $(x_1, x_2, ..., x_n)$. The MLE of geometric parameter p is calculated from the data. Then the probability mass on 1 to $m-1$ via geometric distribution is calculated, and denoted by $\widehat{p}_1, ..., \widehat{p}_{m-1}$. Let $\widehat{p}_m = 1 - \widehat{p}_1 - ... - \widehat{p}_{m-1}$. The K-S test statistic can be calculated by $KS = \max(|D_1|, |D_2|, ...|D_m|)$ where $D_j = n_1 + n_1 + ... + n_j - n(\widehat{p}_1 + \widehat{p}_2 + ... + \widehat{p}_j)$ and n_j is the number of j-valued data for $j = 1, 2, ..., m$.

A-D, a modified version of K-S test, places more weight on observations in the tails of the distribution. A-D test is recommended by Best and Rayner (2003) for testing geometric distribution after they compare it with various other tests. The A-D test statistic described in their paper is $AD = n^{-1} \sum_{1 \leq j \leq m} D_j^2 \widehat{p}_j / H_j(1 - H_j)$ where $H_j = \widehat{p}_1 + \widehat{p}_2 + ... + \widehat{p}_j$. Notice that m is different from the above K-S test and should satisfy that $n_m = 0$ and $\widehat{p}_m < 10^{-3}/n$ in this A-D test.

Similar to Chi-square test, one would consider extending one-dimensional K-S test for the goodness-of-fit tests of BGD(B&D). Unfortunately, the adaption is challenging because of the complexity on defining the cumulative distribution function for high-dimension distributions. Studies on the multi-dimensional K-S tests appear to be quite limited. Fasano and Franceschini (1987) proposed a generalized version of classical K-S tests suitable to test two or three dimensional distributions, which is an improvement of the version proposed by Peacock (1983). Press et al. (2002) described the algorithms of the two dimensional K-S test from Fasano and Franceschini (1987) in Section 14.7, which is briefly presented as follows. For each data point, the (x,y) plane is divided into four quadrants. Then for each of these four quadrants, the difference between the fraction of points over the sample size and the null probability in this quadrant is calculated and then the test statistic is calculated as the maximum differences among four quadrants of each data point. We will implement this algorithm with parameters estimated by the MLE to testing the BGD(B&D) with unknown parameters and compare it with our proposed tests.

2.3 K & K Method of Goodness-of-fit Tests

Kocherlakota and Kocherlakota (1986) presented a general procedure for testing goodness-of-fit of discrete distributions based on PGF. They derived two types of test statistics in each case of univariate or multivariate composite null distributions. The first type is based on the difference between PGF's MLE and its empirical estimator, which has an asymptotic normal distribution as a function of t. The second test statistic, which is based on the several differences between PGF's MLE and its empirical estimator evaluated at several t's, has an asymptotic Chi-square distribution. Their goodness-of-fit test methods are applied to Poisson related distributions and Neyman type A distribution to verify their theories and the validity of test statistics.

The theoretical background of K&K methods are based on the following facts.

First,

$$\frac{G_n(t)-\hat{G}(t;\hat{\Theta})}{\sigma_\xi} \xrightarrow{D} \mathbb{N}(0,1) \text{ as } n\to\infty, \text{where } t\in(-1,1). \quad (2.1)$$

Here $G_n(t)$ is the EPGF for a univariate distribution, $G_n(t)=\frac{1}{n}\sum_{1\le i\le n} t^{X_i}$, and $\hat{G}(t;\hat{\Theta})$ is MLE of the PGF $G(t;\Theta)$, with Θ of dimension k replaced by $\hat{\Theta}$. Further,

$$\sigma_\xi^2=\frac{1}{n}[G(t^2;\Theta)-G^2(t;\Theta)]-\sum_{1\le s\le k}\sum_{1\le r\le k}\sigma_{rs}\frac{\partial G}{\partial\theta_r}\frac{\partial G}{\partial\theta_s}, \quad (2.2)$$

where σ_{rs} is the $(r,s)^{th}$ element of $(nI(\Theta))^{-1}$. Note that $I(\Theta)$ is the Fisher information matrix, which is equal to $\{E(\frac{\partial lnf(X;\Theta)}{\partial\theta_i}\frac{\partial lnf(X;\Theta)}{\partial\theta_j})\}$, where $i,j=1,2,...,k$, and $f(X;\Theta)$ is the probability mass function. In the case that Θ is known, in (2.1) MLE $\hat{\Theta}$ is not needed, and σ_ξ^2 is reduced to $\sigma_\xi^2=\frac{1}{n}[G(t^2;\Theta)-G^2(t;\Theta)]$.

Second, in the case of multiple $t's$, $\underline{t}=(t_1,t_2,...,t_q)$, $t_i\in(-1,1)$, $i=1,2,...,q$, they stated that

$$\text{let } \xi(\underline{t})=\Gamma^{-1/2}(\xi_1,\xi_2,...,\xi_q)' \xrightarrow{D} \mathbb{N}_q(0,\mathbf{1}) \text{ as } n\to\infty, \quad (2.3)$$

where $\mathbf{1}$ is the identity matrix with q dimension, $(\xi_1,\xi_2,...,\xi_q)=\langle G_n(t_1)-\hat{G}(t_1;\hat{\Theta}), G_n(t_2)-\hat{G}(t_2;\hat{\Theta}),...,G_n(t_q)-\hat{G}(t_q;\hat{\Theta})\rangle$, and $\Gamma=\{\Upsilon_{ij}\}$, for $i,j=1,2,...,q$,

$$\Upsilon_{ij}=\frac{1}{n}[G(t_it_j;\Theta)-G(t_i;\Theta)G(t_j;\Theta)]-\sum_{1\le r\le k}\sum_{1\le s\le k}\sigma_{rs}\frac{\partial G(t_i;\Theta)}{\partial\theta_s}\frac{\partial G(t_j;\Theta)}{\partial\theta_r}, \quad (2.4)$$

where σ_{rs} is the same as described in (2.2).

Third, similarly K&K extended the theory to multivariate distributions for single $\underline{t}$ and multiple $\underline{t}'s$. In the case of single $\underline{t}=(t_1,t_2,...,t_m)$ where m is the numbers of variables and $t_j\in(-1,1), j=1,2,...,m$, and

$$\frac{G_n(\underline{t})-\hat{G}(\underline{t};\hat{\Theta})}{\sigma_\xi} \xrightarrow{D} \mathbb{N}(0,1) \text{ as } n\to\infty. \quad (2.5)$$

Here, $G_n(\underline{t})=G_n(t_1,t_2,...,t_m)=\frac{1}{n}\sum_{1\le i\le n} t_1^{x_{1i}}t_2^{x_{2i}}...t_m^{x_{mi}}$ and

$$\sigma_\xi^2=\frac{1}{n}[G(t_1^2,...,t_m^2;\Theta)-G^2(t_1,...,t_m;\Theta)]-\sum_{1\le r\le k}\sum_{1\le s\le k}\sigma_{rs}\frac{\partial G}{\partial\theta_r}\frac{\partial G}{\partial\theta_s}, \quad (2.6)$$

where σ_{rs} is the $(r,s)^{th}$ element of $(nI(\Theta))^{-1}$. Here the Fisher information matrix $I(\Theta)$ is equal to $\{E(\frac{\partial lnf(X_1,X_2,...,X_m;\Theta)}{\partial\theta_i}\frac{\partial lnf(X_1,X_2,...,X_m;\Theta)}{\partial\theta_j})\}$, where $i,j=1,2,...,k$, and $f(X_1,X_2,$ $...,X_m;\Theta)$ is the probability mass function.

Finally, when $\mathbf{t}=(\underline{t_1},...,\underline{t_q})$, $\underline{t_i}=(t_{1i},...,t_{mi})$ for $i=1,...,q$ and $t_{li}\in(-1,1), l=1,...,m$, they obtained that

$$\text{let } \xi(\mathbf{t})=\Gamma^{-1/2}(G_n(\underline{t_1})-\hat{G}(\underline{t_1},\hat{\Theta}),...,G_n(\underline{t_q})-\hat{G}(\underline{t_q},\hat{\Theta}))'\xrightarrow{D}\mathbb{N}_q(0,\mathbf{1}),\ \text{as } n\to\infty, \quad (2.7)$$

where $\mathbf{1}$ is identity matrix with q dimension, and $\Gamma=\{\Upsilon_{ij}\}$, for $i,j=1,2,...,q$, and

$$\Upsilon_{ij}=\frac{1}{n}[G(t_{1i}t_{1j},....,t_{mi}t_{mj};\Theta)-G(\underline{t_i};\Theta)G(\underline{t_j};\Theta)]-\sum_{1\le r\le k}\sum_{1\le s\le k}\sigma_{rs}\frac{\partial G(\underline{t_i};\Theta)}{\partial\theta_r}\frac{\partial G(\underline{t_j};\Theta)}{\partial\theta_s}, \quad (2.8)$$

where σ_{rs} is the same as described in (2.6).

Next, they described the generalized methods for testing the hypothesis $H_0: G(\underline{t};\Theta)=G_0(\underline{t};\Theta)$ with unknown Θ. In the case that $G_0(\underline{t};\Theta)$ is the PGF from a univariate distribution, the test statistic with single t for $t\in(-1,1)$ is $\frac{G_n(t)-\hat{G}_0(t;\hat{\Theta})}{\hat{\sigma_\xi}}$, which approximately follows the standard normal distribution. Here $\hat{G}_0(t;\hat{\Theta})$ is determined from $G_0(t;\Theta)$, replacing Θ with $\hat{\Theta}$, and $\hat{\sigma_\xi}$ is the MLE of σ_ξ from (2.2) under H_0. The test statistic with multiple $t's$ is $\hat{\xi}(\underline{t})'\hat{\xi}(\underline{t})$, which is asymptotically Chi-square distribution with q degrees of freedom, where $\underline{t}=(t_1,t_2,...,t_q)$, $t_i\in(-1,1)$, for $i=1,2,...,q$. Here $\hat{\xi}(\underline{t})$ is obtained from $\xi(\underline{t})$ in (2.3) with Γ replaced by its MLE $\hat{\Gamma}$ under H_0.

In the case that $G_0(\underline{t};\Theta)$ is the PGF from a multivariate distribution, the test statistic with single $\underline{t}=(t_1,t_2,...,t_m)$ is $(G_n(t_1,...,t_m)-\hat{G}_0(t_1,...,t_m;\hat{\Theta}))/\hat{\sigma}_\xi$, which is approximately the standard normal distribution, where $\hat{G}_0(t_1,...,t_m;\hat{\Theta})$ is determined from $G_0(t_1,...,t_m;\Theta)$, replacing Θ with $\hat{\Theta}$ and $\hat{\sigma}_\xi$ is estimator of σ_ξ under H_0, see (2.6). The second test statistic with multiple $t's$ is $\hat{\xi}(\mathbf{t})'\hat{\xi}(\mathbf{t})$, which is asymptotically Chi-square distributed with q degrees of freedom, where $\mathbf{t}=(\underline{t_1},...,\underline{t_q})$, $\underline{t_i}=(t_{1i},...,t_{mi})$ for $i=1,...,q$ and $t_{li}\in(-1,1), l=1,...,m$. Here $\hat{\xi}(\mathbf{t})$ is evaluated from $\hat{\xi}(\mathbf{t})$ in (2.7) with Γ replaced by its MLE $\hat{\Gamma}$ under H_0.

CHAPTER 3

TEST STATISTICS

In this chapter, we propose three test statistics for both geometric distribution and BGD(B&D). The first and second ones are the application of the K&K method with single t and multiple t's, respectively. Although the K&K methods are applied to testing geometric distribution by some authors (e.g., Epps (1995)), as far in the literature this research would be the first one exploring that the value of t determines the performance of the goodness-of-fit test for geometric distribution. Also this research would be the first one exploring their application on BGD(B&D). The third one is the supremum of absolute value of the standardized difference between the MLE of PGF and its empirical counterpart. Note the MLE of PGF is the estimator of PGF with the unknown parameters replaced by their MLE.

3.1 Goodness-of-fit Tests for Geometric Distribution

Let $X_1, X_2, ..., X_n$ be independently and identically distributed random variables taking positive integer values with the true PGF $G(t;p)$. We test the hypotheses H_0: $G(t;p) = \frac{tp}{1-t(1-p)}$, where $|t| < 1/(1-p)$, the PGF of geometric distribution with unknown p versus H_1: $G(t;\Theta)$ is not the PGF for geometric distribution. Note that in the whole article context, geometric distribution prefers to model the number of total trials before the first failure. That is, the geometric random variable here takes the values staring from one not zero. Here we focus on the case when p is unknown because the case when p is known is straightforward.

3.1.1 K&K Method with Single t

In this section, we apply the K&K method with single t to the goodness-of-fit of geometric distribution with unknown parameter p.

Let $\hat{p} = 1/\overline{x}$, the MLE for p. The MLE of PGF $G(t,p)$ under H_0 is given by $\hat{G}(t,\hat{p}) = \frac{t\hat{p}}{(1-t(1-\hat{p}))} = \frac{t/\overline{x}}{1-t(1-1/\overline{x})} = \frac{t}{\overline{x}-t(\overline{x}-1)}$. The EPGF is $G_n(t) = n^{-1}\sum_{1\leq j\leq n} t^{x_j}$. Let

$\xi(t) = \frac{G_n(t) - \hat{G}(t;\hat{p})}{\sigma_\xi}$, where

$$\begin{aligned}
\sigma_\xi^2 &= \frac{1}{n}(G(t^2;p) - G^2(t;p)) - \sum_i \sum_j \sigma_{ij} \frac{\partial G}{\partial \theta_i} \frac{\partial G}{\partial \theta_j} \\
&= \frac{1}{n}\left\{ \frac{pt^2}{1 - t^2(1-p)} - \frac{t^2 p^2}{(1 - t(1-p))^2} \right\} - \sigma_{11}(\partial G/\partial p)^2 \\
&= \frac{1}{n}\left\{ \frac{t^2 p}{1 - t^2(1-p)} - \frac{t^2 p^2[(1 - t + tp)^2 + (1-p)(1-t)^2]}{(1 - t + tp)^4} \right\}, \\
\partial G/\partial p &= \frac{t - t^2}{(1 - t(1-p))^2}, \\
\text{and} \quad \sigma_{11} &= \frac{1}{n}(E\{(\partial logF/\partial p)^2\})^{-1} = \frac{1}{n}(E\{(\frac{1-x}{1-p} + 1/p)^2\})^{-1} = \frac{p^2(1-p)}{n}. \qquad (3.1)
\end{aligned}$$

According to K&K method with single t, we know that $\xi(t) \xrightarrow{D} \mathbf{N}(0,1)$ as $n \to \infty$ with $|t| < 1$. Then the test statistic is

$$\mathcal{Z}(t) = \frac{G_n(t) - \hat{G}(t;\hat{p})}{\widehat{\sigma_\xi}}, |t| < 1,$$

where $\widehat{\sigma_\xi}$ is MLE of σ_ξ, determined by the plug-in estimator replacing p by $\hat{p} = 1/\overline{x}$. Therefore

$$\widehat{\sigma_\xi}^2 = \frac{t^2}{n}\left\{ \frac{1}{\overline{x} - t^2(\overline{x} - 1)} - \frac{1}{(\overline{x} - t(\overline{x} - 1))^2} - \frac{\overline{x}(\overline{x} - 1)(1-t)^2}{(\overline{x} - t(\overline{x} - 1))^4} \right\}. \qquad (3.2)$$

We reject the null hypothesis if $|\mathcal{Z}(t)| > z_{1-\alpha/2}$ at significance level α, where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)^{th}$ percentile of the standard normal distribution.

3.1.2 K&K Method with Multiple *t*'s

Now we consider another test statistic with several t values, say $t_1, t_2, ..., t_q$. Let the vector $\underline{t} = (t_1, t_2, ..., t_q)$. The vector of theoretical PGF for geometric distribution is evaluated at $\underline{t}$ as follows $G'(\underline{t};p) = \langle G(t_1;p), G(t_2;p), ..., G(t_q;p)\rangle$. Define $\hat{G}'(\underline{t};\hat{p}) = \langle \hat{G}(t_1;\hat{p}), \hat{G}(t_2;\hat{p}), ..., \hat{G}(t_q;\hat{p})\rangle$ from $G'(\underline{t};p)$ with p replaced by $\hat{p}$. The vector of the EPGF evaluated at multiple t's is $G_n'(\underline{t}) = \langle G_n(t_1), G_n(t_2), ..., G_n(t_q)\rangle$.

Now let $\xi'(\underline{t}) = G_n'(\underline{t}) - \hat{G}'(\underline{t};\hat{p})$. We know $\Gamma^{-1/2}\xi(\underline{t})$ is asymptotically $\mathbb{N}_q(\underline{0},\mathbf{1})$ as $n \to \infty$, according to the K & K method with multiple $t's$. Here $\Gamma = \{\Upsilon_{ij}\}$, for $i,j = 1,2,...,q$,

$$\begin{aligned}\Upsilon_{ij} &= \frac{1}{n}\left[G(t_it_j;p) - G(t_i;p)G(t_j;p)\right] - \sigma_{11}\frac{\partial G(t_i;p)}{\partial p}\frac{\partial G(t_j;p)}{\partial p} \\ &= \frac{1}{n}\left[\frac{t_it_jp}{1-t_it_j(1-p)} - \frac{t_ip}{1-t_i(1-p)}\frac{t_jp}{1-t_j(1-p)} - \frac{p^2(1-p)(t_i-t_i^2)(t_j-t_j^2)}{(1-t_i(1-p))^2(1-t_j(1-p))^2}\right].\end{aligned} \tag{3.3}$$

Therefore, the test statistic can be defined as

$$\mathcal{T}_q(\underline{t}) = \xi'(\underline{t})\hat{\Gamma}\xi(\underline{t}),$$

where $\underline{t} = (t_1, t_2, ..., t_q)$, $|t_i| < 1$, $i = 1,2,...,q$, $\hat{\Gamma} = \{\hat{\Upsilon}_{ij}\}$ is the MLE of Γ matrix evaluated by replacing p with $\hat{p} = 1/\overline{x}$, and

$$\begin{aligned}\hat{\Upsilon}_{ij} =& \frac{t_it_j}{n}\left(\frac{1}{\overline{x}-t_it_j(\overline{x}-1)} - \frac{1}{(\overline{x}-t_i(\overline{x}-1))(\overline{x}-t_j(\overline{x}-1))}\right) \\ &- \frac{t_it_j}{n}\left(\frac{\overline{x}(\overline{x}-1)(1-t_i)(1-t_j)}{(\overline{x}-t_i(\overline{x}-1))^2(\overline{x}-t_j(\overline{x}-1))^2}\right),\end{aligned} \tag{3.4}$$

for $i,j = 1,2,...,q$. Under the null hypothesis, $\mathcal{T}_q(t)$ is asymptotically χ^2 distributed with q degrees of freedom. We reject the null hypothesis if $\mathcal{T}_q(\underline{t})$ is greater than $100(1-\alpha)^{th}$ percentile of the Chi-square distribution with q degrees of freedom at significance level α. Note that when $\overline{x} = 1$, the data is noninformative and Γ is singular.

3.1.3 Supremum Test based on PGF

We develop a new testing procedure based on the K&K method to test the goodness-of-fit of geometric distribution. We propose the supremum of the absolute value of the test statistic of K&K method with single t over $(-1,1)$ as the test statistic. This new test is an improvement over both of the K&K methods since it does not require selecting specific values for t. The proposed test statistic is

$$\mathcal{SD}_n = \sup_{t\in(-1,1)}\left|[G_n(t) - \hat{G}(t;\hat{p})]/\hat{\sigma}_\xi\right| = \sup_{t\in(-1,1)}\left|\left(\frac{1}{n}\sum_{1\le j\le n} t^{x_j} - \frac{t}{\overline{x}-t(\overline{x}-1)}\right)/\hat{\sigma}_\xi\right|, \tag{3.5}$$

where $\hat{\sigma}_\xi$ is evaluated in (3.2). To explore the asymptotic distribution of $\mathcal{SD}_n$, we calculate the empirical critical points at $\alpha = 0.01$, 0.05, 0.1 and construct the density curves of $\mathcal{SD}_n$ at various sample size n and parameter p by Monte Carlo simulation.

In the real world application, the empirical critical points can be obtained by parametric bootstrap, which has been intensively examined and widely accepted (see, e.g., [36], [17], [20], [2], [5] and [28]). The bootstrap procedure of testing geometric distribution in this case is the following. (1) Calculate the MLE of geometric parameter $\hat{p} = 1/\overline{x}$ based on the data. (2) Generate 10,000 bootstrap samples of sample size n based on geometric distribution with parameter equal to $\hat{p}$. (3) Calculate $\mathcal{SD}_n$ for each of the 10,000 samples. (4) Sort the 10,000 $\mathcal{SD}_n$ values in the ascending order then determine the 90^{th}, 95^{th}, 99^{th} percentiles of $\mathcal{SD}_n$. (5) Calculate $\mathcal{SD}_n$ with the data and reject H_0 if this $\mathcal{SD}_n$ value is greater than the empirical $1-\alpha$ percentile at significance level α.

3.2 Goodness-of-fit Tests for BGD(B&D)

Let$\langle (X_{11}, X_{21}), (X_{12}, X_{22}), ..., (X_{1n}, X_{2n}) \rangle$ be a random sample from the true distribution with PGF $G(\underline{t}; \underline{p})$ and consider the hypotheses $H_0 : G(\underline{t}, \underline{p})$ is the PGF of BGD(B&D) with unknown parameters $\underline{p} = (p_1, p_2, p_3)$ against the alternative. We use the K&K methods with single $\underline{t}$ and multiple $\underline{t}$'s to evaluate these hypotheses and also develop a new supremum test similar to the one in the univariate case.

3.2.1 K&K Method with Single t

In this section, we apply the K&K general method with single $\underline{t} = (t_1, t_2)$ for bivariate distribution to evaluate the above hypotheses.

We estimate the PGF for $BGD(B\&D)$ $G(\underline{t}; \underline{p})$ by replacing $\underline{p} = (p_1, p_2, p_3)$ with its MLE $\underline{\hat{p}} = (\hat{p_1}, \hat{p_2}, \hat{p_3})$, and use $\hat{G}(\underline{t}; \underline{\hat{p}})$ to denote the estimator. So

$$\hat{G}(\underline{t}; \underline{\hat{p}}) = \hat{G}(t_1, t_2; \hat{p_1}, \hat{p_2}, \hat{p_3}) = \hat{A}_1 \hat{A}_2, \tag{3.6}$$

$$\text{where} \quad \hat{A}_1 = \frac{t_1 t_2}{1 - t_1 t_2 \hat{p_1} \hat{p_2} \hat{p_3}}$$

$$\hat{A}_2 = \frac{t_2 \hat{q_1} (1 - \hat{p_2}\hat{p_3}) \hat{p_2}\hat{p_3}}{1 - t_2 \hat{p_2}\hat{p_3}} + \frac{t_1 \hat{q_2} \hat{p_1}\hat{p_3} (1 - \hat{p_1}\hat{p_3})}{1 - t_1 \hat{p_1}\hat{p_3}} + 1 - \hat{p_1}\hat{p_3} - \hat{p_2}\hat{p_3} + \hat{p_1}\hat{p_2}\hat{p_3}.$$

The bivariate EPGF $G_n(\underline{t}) = G_n(t_1, t_2) = \frac{1}{n}\sum_{1\leq j\leq n} t_1^{x_{1j}} t_2^{x_{2j}}$. Let $\xi(\underline{t}) = \frac{G_n(t_1,t_2)-\hat{G}(t_1,t_2;\hat{p_1},\hat{p_2},\hat{p_3})}{\sigma_\xi}$, where

$$\sigma_\xi^2 = \frac{1}{n}[G(t_1^2, t_2^2; \underline{p}) - G^2(t_1, t_2; \underline{p})] - \sum_{1\leq i\leq 3}\sum_{1\leq j\leq 3} \sigma_{ij}\frac{\partial G}{\partial p_i}\frac{\partial G}{\partial p_j}, \tag{3.7}$$

$\Delta = \{\sigma_{ij}\}$, the inverse matrix of $n\mathbb{I}(\underline{p})$, and $\mathbb{I}(\underline{p})$ is the Fisher information matrix, which is equal to $\{-E(\frac{\partial^2 logL((x,y);\underline{p})}{\partial p_i \partial p_j})\}$ for $i, j = 1, 2, 3$. $\frac{\partial G}{\partial p_i}$ and $\frac{\partial G}{\partial p_j}$ can be obtained from the Appendix. According to the K&K method with single t, $\xi(\underline{t}) \xrightarrow{D} \mathbb{N}(0, 1)$ as $n \to \infty$, for $\underline{t} = (t_1, t_2)$, $|t_1|, |t_2| < 1$. So the test statistic is given by

$$\mathcal{MZ}(\underline{t}) = \frac{G_n(t_1, t_2) - \hat{G}(t_1, t_2; \hat{p_1}, \hat{p_2}, \hat{p_3})}{\hat{\sigma_\xi}}, \tag{3.8}$$

where $\hat{\sigma_\xi}$ is evaluated from σ_ξ in (3.7) with $\underline{p} = \underline{\hat{p}}$ and follows asymptotically standard normal distribution under H_0. We reject the null hypothesis if $|\mathcal{MZ}(\underline{t})| > z_{1-\alpha/2}$ at significance level α, where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)^{th}$ percentile of the standard normal distribution.

3.2.2 K&K Method with Multiple *t*'s

To describe this method, the following notations are needed. Let $\mathbf{t} = (\underline{t_1}, \underline{t_2}, ..., \underline{t_q})$ and $\underline{t_i} = (t_{1i}, t_{2i})$, for $i = 1, 2, ..., q$. The vector of PGF evaluated at $\mathbf{t}$ in this case is

$$\begin{aligned} G'(\mathbf{t}, p) &= G'(\underline{t_1}, \underline{t_2}, ..., \underline{t_q}; \underline{p}) \\ &= \langle G(\underline{t_1}; \underline{p}), G(\underline{t_2}; \underline{p}), ..., G(\underline{t_q}; \underline{p})\rangle \\ &= \langle G(t_{11}, t_{21}; p_1, p_2, p_3), G(t_{12}, t_{22}; p_1, p_2, p_3), ..., G(t_{1q}, t_{2q}; p_1, p_2, p_3)\rangle. \end{aligned} \tag{3.9}$$

The vector of q dimensional bivariate EPGF is denoted by

$$\begin{aligned} G_n'(\mathbf{t}) &= G_n'(\underline{t_1}, \underline{t_2}, ..., \underline{t_q}) \\ &= \langle G_n(\underline{t_1}), G_n(\underline{t_2}), ..., G_n(\underline{t_q})\rangle \\ &= \langle G_n(t_{11}, t_{21}), G_n(t_{12}, t_{22}), ..., G_n(t_{1q}, t_{2q})\rangle \\ &= \left(\frac{1}{n}\sum_{1\leq j\leq n} t_{11}^{x_{1j}} t_{21}^{x_{2j}}, \frac{1}{n}\sum_{1\leq j\leq n} t_{12}^{x_{1j}} t_{22}^{x_{2j}}, ..., \frac{1}{n}\sum_{1\leq j\leq n} t_{1q}^{x_{1j}} t_{2q}^{x_{2j}}\right). \end{aligned} \tag{3.10}$$

Let $\xi'_n(\mathbf{t}) = G'_n(\mathbf{t}) - \hat{G}'(\mathbf{t}, \underline{\hat{p}})$, where $\hat{G}'(\mathbf{t}; \underline{\hat{p}})$ is $G'(\mathbf{t}; \underline{p})$ with $\underline{p}$ substituted by $\underline{\hat{p}}$. The test statistic is given by

$$\mathcal{MT}_q(\mathbf{t}) = \xi'_n(\mathbf{t})\hat{\Gamma}^{-1}\xi_n(\mathbf{t}),$$

where $\hat{\Gamma}$ is calculated from $\Gamma = \{\Upsilon_{ij}\}$, the covariance matrix of $\xi'_n(\mathbf{t})$, by replacing $\underline{p}$ by $\underline{\hat{p}}$,

$$\begin{aligned}\Upsilon_{ij} = & \frac{1}{n}\left[G(t_{1i}t_{1j}, t_{2i}t_{2j}; \underline{p}) - G(t_{1i}, t_{2i}; \underline{p})G(t_{1j}, t_{2j}; \underline{p})\right] \\ & - \sum_{1\le r\le 3}\sum_{1\le s\le 3} \sigma_{ij} \frac{\partial G(t_{1i}, t_{2i}; \underline{p})}{\partial p_r}\frac{\partial G(t_{1j}, t_{2j}; \underline{p})}{\partial p_s}, \qquad \text{for } i, j = 1, 2, ..., q, \end{aligned} \tag{3.11}$$

and $\frac{\partial G(t_{1j}, t_{2j}; \underline{p})}{\partial p_r}$ and $\frac{\partial G(t_{1j}, t_{2j}; \underline{p})}{\partial p_s}$ are described as in the Appendix, and is asymptotically χ^2 distributed with q degrees of freedom under H_0. We reject the null hypothesis if $\mathcal{MT}_q(\mathbf{t})$ is greater than $100(1-\alpha)^{th}$ percentile of the Chi-square distribution with q degrees of freedom at significance level α.

3.2.3 Supremum Test based on PGF

We develop a new method based on the supremum of the absolute value of standardized difference between PGF's MLE and its EPGF, similar as the one for testing geometric distribution described in Section 3.1.3, to test the BGD(B&D) with unknown parameters. The proposed test statistic is given by

$$\begin{aligned}\mathcal{MSD}_n &= \sup_{t_1, t_2 \in (-1,1)} |((G_n(t_1, t_2) - \hat{G}(t_1, t_2; \hat{p_1}, \hat{p_2}, \hat{p_3}))/\hat{\sigma}_\xi| \\ &= \sup_{t_1, t_2 \in (-1,1)} \left| (\frac{1}{n}\sum_{1\le j\le n} t_1^{x_{1j}} t_2^{x_{2j}} - \hat{G}(t_1, t_2; \hat{p_1}, \hat{p_2}, \hat{p_3}))/\hat{\sigma}_\xi \right| \end{aligned} \tag{3.12}$$

where $\hat{G}(t_1, t_2; \hat{p_1}, \hat{p_2}, \hat{p_3})$ is obtained from (3.6) and $\hat{\sigma}_\xi$ is MLE of σ_ξ defined in (3.7). We explore the asymptotic behavior of $\mathcal{MSD}_n$ by calculating its empirical critical points and plotting the density curves for various $\underline{p}$ and n based on Monte-Carlo simulations. In the real world application, parametric bootstrap procedure should be applied for the goodness-of-fit of BGD(B&D)with unknown parameters $\underline{p}$, which is similar as described in Section 3.1.3 for the univariate case.

CHAPTER 4

SIMULATION STUDY FOR TESTING GEOMETRIC DISTRIBUTION

4.1 Empirical Distributions of Test Statistics

In order to verify the asymptotic behavior of $\mathcal{Z}(t)$, the test statistic of the K&K method with single t for testing geometric distribution described in Section 3.1.1, we calculate the empirical critical points of $\mathcal{Z}(t)$ for $t = \pm 0.01, \pm 0.5, \pm 0.9$ (t is in the neighbor of zero or is not) based on 10,000 samples of size $n = 50$ from geometric distribution with $p = 0.25$ by Monte-Carlo simulations, and compare them with their theoretical counterparts from the standard normal distribution (see Figure 4.1).

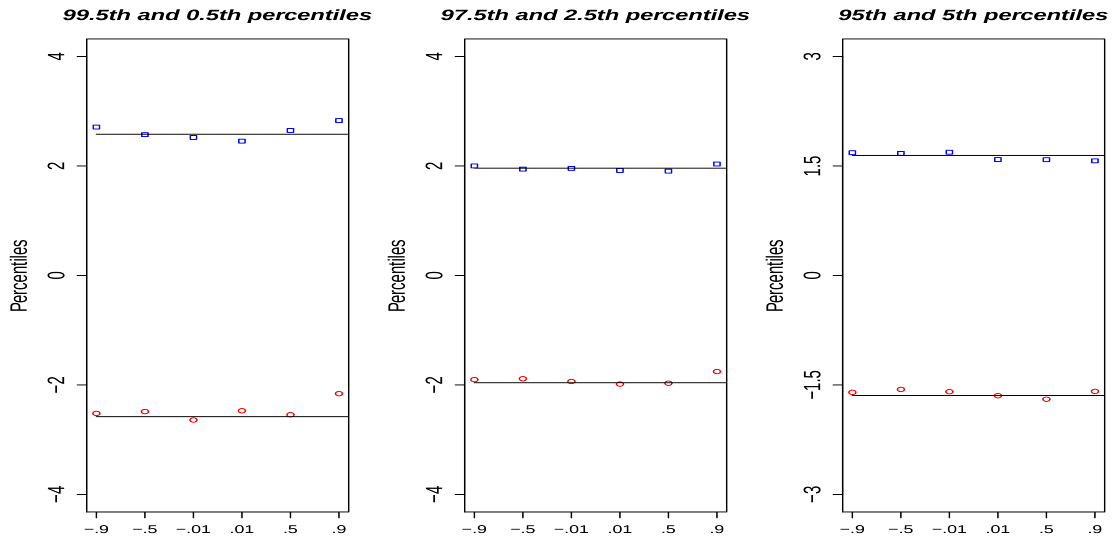

Figure 4.1 Empirical percentiles of $\mathcal{Z}(t)$ which is on the x-axis indicated by t value with $n = 50$ and $p = 0.25$.

Figure 4.1 shows that when t is closer to zero, the empirical critical points are closer to the corresponding standard normal percentiles. The empirical critical points are closer to the corresponding standard normal percentiles at the significance level $\alpha = 0.05$ and 0.10 than at $\alpha = 0.01$. Despite the difference in performance at different $t's$ and different significance levels, overall the empirical critical points are satisfactorily close to the corresponding theoretical ones.

Similarly, in order to verify the asymptotic behavior of $\mathcal{T}_q(\underline{t})$, the test statistic of the K&K method with multiple t's for testing geometric distribution described in Section 3.1.2, we compute the empirical critical points of $\mathcal{T}_q(\underline{t})$ by simulations based on 10,000 replications with $n = 50$ and $p = 0.25$, and compare them with their theoretical counterparts from the Chi-square distributions (see Table 4.1). Here in different cases, the number of t's, q is chosen to be variate, i.e. $q = 3$, 5 or 10, and the values of t are chosen to be either in the neighborhood of zero or well spanned in interval (-1,1).

Table 4.1 Empirical Critical Points of $\mathcal{T}_q(\underline{t})$ based on 10,000 Replications

Statistics	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$\mathcal{T}_3(-0.15, -0.05, 0.05)$	11.349	7.672	6.113
$\mathcal{T}_3(-0.90, 0.05, 0.85)$	11.483	7.606	6.061
χ^2_3	11.345	7.815	6.251
$\mathcal{T}_5(-0.25, -0.15, -0.05, 0.15, 0.25)$	14.647	10.759	9.018
$\mathcal{T}_5(-0.90, -0.55, -0.15, 0.25, 0.90)$	15.622	11.039	9.184
χ^2_5	15.080	11.070	9.236
$\mathcal{T}_{10}(-0.90, -0.75, -0.55, -0.35, -0.15, 0.2, 0.4, 0.6, 0.8, 0.95)$	25.547	17.580	14.971
χ^2_{10}	23.209	18.307	15.987

In our calculation, we find that sometimes the entries of $\hat{\Gamma}$, see (3.4), are too small, which causes $\hat{\Gamma}$ to be computationally singular. A threshold number of q exist for different type of t's to avoid this problem. The threshold is larger when the t values spread well in the whole interval (-1,1) than when the t values are close to 0. Also computationally singular arises more often when $q = 10$ than when $q = 3$ and $q = 5$. Table 4.1 shows that overall the empirical critical points perform well in terms of being close to their respective Chi-square percentiles. However, the empirical critical points are closer to the corresponding Chi-square percentiles when $q = 3$ and $q = 5$ than when $q = 10$.

Now in order to explore the asymptotic distributions of $\mathcal{SD}_n$, the supremum test statistic for testing geometric distribution described in Section 3.1.3, we calculate the empirical critical points of $\mathcal{SD}_n$ for various p and n by parametric bootstrap with 10,000 bootstrap samples, see Table 4.2 and plot the corresponding density curves (see Figure 4.2). Note that $p = 0.25$, 0.5 and 0.6, which are chosen corresponding to over dispersive, equally dispersive and under dispersive geometric distributions and that $n = 20$, 50, 100, 200, 500 so that the asymptotic behavior of the critical points of $\mathcal{SD}_n$ can be studied.

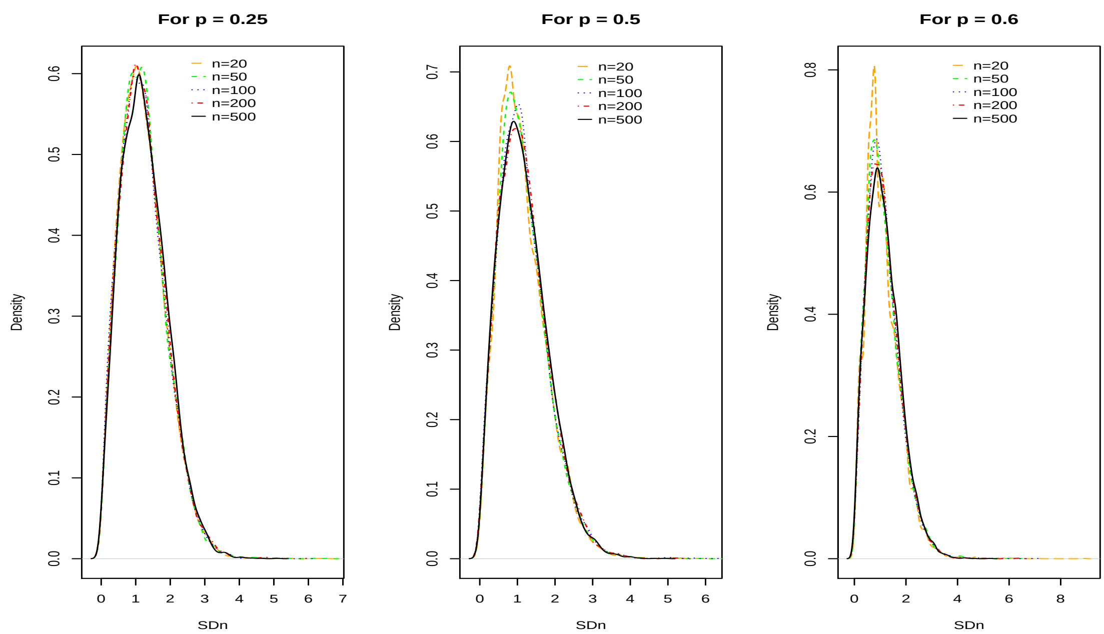

Figure 4.2 Density curves of $\mathcal{SD}_n$ with various p and n.

From Figure 4.2, we observe that for each p as n increases, the density curves of $\mathcal{SD}_n$ approach more and more close to each other, indicating a trend of achieving the limiting distributions. The curves attain a limiting density lot sooner for $p = 0.25$ case than for $p = 0.5$ and $p = 0.6$ cases. Table 4.2 indicates that the critical points of $\mathcal{SD}_n$ becomes fairly stable as n increases for each case of p.

Table 4.2 Empirical Critical Points of $\mathcal{SD}_n$ based on 10,000 Bootstrap Samples

	α	0.01	0.05	0.1
p=0.25	n=20	3.0646	2.4203	2.1047
	n=50	3.0961	2.4111	2.1284
	n=100	3.0628	2.4191	2.1186
	n=200	3.0988	2.4357	2.1197
	n=500	3.0500	2.4430	2.1443
p=0.5	n=20	3.0413	2.3384	2.0171
	n=50	3.0842	2.3786	2.0366
	n=100	3.1245	2.3812	2.0462
	n=200	3.0705	2.3801	2.0620
	n=500	3.0446	2.3650	2.0583
p=0.6	n=20	3.0687	2.3130	1.9660
	n=50	3.0392	2.3406	1.9995
	n=100	3.0367	2.3255	1.9900
	n=200	3.0381	2.3392	2.0095
	n=500	3.0001	2.3442	2.0164

4.2 Evaluating Tests by Comparison

Best and Rayner (2003) recommended A-D test for testing geometric distributions with $p =$ 0.25, 0.5 or 0.6. Here we propose to compare Type I Errors and powers of the above tests with those of A-D, K-S and Chi-square tests under the same goodness of fit tests with the null hypothesis distribution specified as geometric distribution. Note that the value of geometric distribution in this comparative study starts from one not zero. We generate 10,000 Monte Carlo simulation replications of size $n = 50$ from the alternative distributions with the first moment equal to the one of the null distribution. The power is the proportion of replications whose corresponding test statistic values greater than the 95^{th} percentile indicating the null hypothesis being rejected. Table 4.5, Table 4.6 and Table 4.7 show the power of various goodness-of-fit tests for geometric distributions with $p = 0.25$, 0.5 and 0.6, respectively.

In the Chernoff and Lehmann Chi-square test with test statistic χ^2_{cl1}, the data is categorized into four groups: [1, 2], [3, 4], [5, 6], [7, ∞). So χ^2_{cl1} is asymptotically Chi-square distributed with two degrees of freedom when the null hypothesis is true. Similarly, in the other two Chernoff and Lehmann Chi-square tests with test statistics χ^2_{cl2} and χ^2_{cl3}, we categorize the data into three groups: [1], [2], [3, ∞). So the degree of freedom of the asymptotic distribution of χ^2_{cl2} or χ^2_{cl3}, which is Chi-square distribution, is 1 when the null hypothesis is true. The empirical critical points of the Chernoff and Lehmann Chi-square test statistics are significantly deviating from their corresponding theoretical counterparts (see Table 4.3 in details). In order to have an accurate evaluation of the tests, we compare the goodness-of-fit tests based on their empirical critical points for χ^2_{cl1}, χ^2_{cl2} and χ^2_{cl3}.

Table 4.3 Empirical (Theoretical) Critical Points of χ^2_{cl1}, χ^2_{cl2} and χ^2_{cl3} based on 10,000 Replications

α	0.01	0.05	0.1
χ^2_{cl1}	8.4808 (9.2103)	5.5084 (5.9915)	4.1657 (4.6052)
χ^2_{cl2}	6.4526 (6.6349)	4.3026 (3.8415)	3.0415 (2.7055)
χ^2_{cl3}	5.9929 (6.6349)	3.6780 (3.8415)	2.7035 (2.7055)

We apply the methods described in Best and Rayner (2003) to calculate the values of test statistics denoted by KS and AD for K-S and A-D tests, respectively. By Monte-Carlo simulations, we calculate their empirical critical points at $\alpha = 0.01$, 0.05, 0.10 for $n = 50$ based on 10,000 replications. Table 4.4 shows the results. These critical points are used to compute powers and empirical α in Table 4.5, Table 4.6 and Table 4.7.

We choose the following six types of alternative distributions as used in Best and Rayner (2003) for our comparative study. (1) $BB(n, \alpha, \beta)$: a beta binomial(BB) distribution with parameter n, α and β, where n is the number of trials, and α and β are parameters from a standard beta distribution; (2) $NB(n, p)$: a negative binomial(NB) used to fit the number of

Table 4.4 Empirical Critical Points of KS and AD with $n = 50$ based on 10,000 Replications

	KS			*AD*		
p	α=0.01	α=0.05	α=0.1	α=0.01	α=0.05	α=0.1
0.25	7.4840	6.1018	5.4259	1.9545	1.3118	1.0460
0.5	6.4421	4.9615	4.2494	2.0482	1.2771	0.9740
0.6	5.8806	4.4118	3.6842	1.8466	1.1429	0.8846

trials before n successes with probability of success p for each trial; (3) $NA(\lambda_1, \lambda_2)$: a Neyman Type A (NA) distribution with probability mass function $e^{-\lambda_1}\lambda_2/x!\sum_{i=0}^{\infty}(\lambda_1 e^{-\lambda_2})^i i^x/i!$, for $x = 0, 1, 2, ...$; (4) $DU(i, j)$: a discrete uniform (DU) distribution with probability mess on $i, i + 1, ..., j$; (5) $B(n, p)$: a standard binomial distribution with n trials and probability of success p; (6) $PM(\omega_1, \omega_2, \lambda_1, \lambda_2)$: a Poisson mixture(PM) where λ_i is a standard Poisson distribution parameter for $i = 1, 2$, and ω_1 and ω_2 are weights obtained between zero and one, which add up to one.

We use the standard functions in R 2.14 to generate geometric, binomial and negative binomial distributed random numbers, and the function *rbetabinom.ab* in package $VGAM$ of R 2.14 to generate beta binomial distributed random numbers. Discrete uniformly distributed random numbers are obtained by inversing the cumulative distribution function. A Bernoulli distributed random value is created by using function *rbinom* in R 2.14 with the number of trial one and probability of success ω_1. If the value is one, we generate random numbers from $P(\lambda_1)$ as the PM random numbers, otherwise from $P(\lambda_2)$, where $P(\lambda_i)$ is the standard Poisson distribution with parameter λ_i for i = 1, 2. According to Best and Rayner (2003), a random sample from the Neyman Type A distribution is generated as follows. First, generate a random sample (n_1, n_2, ..., n_i) from $P(\lambda_1)$. Then generate i random samples from $P(\lambda_2)$, with size $n_1, n_2, ...,$ and n_i, respectively. Finally, sum up the data in each of the i random samples from $P(\lambda_2)$ to obtain a random sample of size i from Neyman Type A. Since here the value of geometric distributed random variable starts from one not zero, the final data from the alternative distributions are obtained by adding one to the generated random samples from the above six distributions.

In the case where the alternative distribution is $NB(1,\ p)$, the power is equivalent to the significance level since $NB(1,\ p)$ corresponds to geometric distribution with parameter p. See the highlighted rows in Table 4.5, Table 4.6 and Table 4.7.

Table 4.5 Powers(%) for Geometric Tests with $p = 0.25$ at $\alpha = 0.05$ based on 10,000 Samples of Size $n = 50$ and the Mean of the Alternative is 4

Alternative	$\mathcal{Z}(0.01)$	$\mathcal{Z}(-0.01)$	$\mathcal{Z}(0.5)$	$\mathcal{Z}(-0.5)$	$\mathcal{Z}(0.9)$	$\mathcal{Z}(-0.9)$	
NB(1, 0.25)	**5.05**	**4.97**	**4.87**	**4.82**	**3.76**	**4.70**	
BB(9, 1, 2)	23.97	22.78	52.95	10.16	63.68	7.21	
BB(9, 0.5, 1)	7.66	7.42	3.05	8.53	7.18	5.41	
NA(1, 3)	71.45	71.68	33.51	79.80	4.32	52.58	
NA(0.95, 3.16)	79.86	80.97	43.91	87.17	7.02	60.17	
NA(0.9, 3.33)	88.08	89.10	55.32	92.84	9.50	69.04	
Alternative	$\mathcal{T}_3(\underline{t_1})$	$\mathcal{T}_5(\underline{t_2})$	$\mathcal{T}_{10}(\underline{t_3})$	$\mathcal{SD}_n$	χ^2_{cl1}	KS	AD
NB(1, 0.25)	**5.01**	**5.13**	**3.90**	**5.12**	**4.94**	**4.98**	**5.15**
BB(9, 1, 2)	38.05	42.24	28.47	41.23	32.85	47.68	60.33
BB(9, 0.5, 1)	8.97	24.34	27.16	6.63	12.03	18.84	23.62
NA(1, 3)	74.80	66.87	45.69	67.51	12.82	65.07	67.62
NA(0.95, 3.16)	83.02	76.93	56.27	78.95	13.59	73.67	77.67
NA(0.9, 3.33)	89.81	84.96	66.89	86.10	15.73	83.03	84.81

where $\underline{t_1}$ = (-0.15, -0.05, 0.05), $\underline{t_2}$ = (-0.25, -0.15, -0.05, 0.15, 0.25), $\underline{t_3}$ = (-0.9, -0.75, -0.55, -0.35, -0.15, 0.2, 0.4, 0.6, 0.8, 0.95).

In Table 4.5, $\mathcal{Z}(0.01)$, $\mathcal{Z}(-0.01)$ $\mathcal{Z}(-0.5)$, $\mathcal{T}_3(\underline{t_1})$, $\mathcal{SD}_n$ have higher powers than KS and AD for NA alternatives while they all maintain Type I error probability close to 5%. However, AD performs the best overall for BB alternatives. The χ^2_{cl1} performs the worst for NA alternatives and moderately for BB alternatives. Obviously, the choice of goodness-of-fit

Table 4.6 Powers(%) for Geometric Tests with $p = 0.5$ at $\alpha = 0.05$ based on 10,000 Samples of Size $n = 50$ and the Mean of the Alternative is 2

Alternative	$\mathcal{Z}(0.01)$	$\mathcal{Z}(-0.01)$	$\mathcal{Z}(0.5)$	$\mathcal{Z}(-0.5)$	$\mathcal{Z}(0.9)$	$\mathcal{Z}(-0.9)$	
NB(1, 0.5)	**5.13**	**5.23**	**4.71**	**5.13**	**3.30**	**5.02**	
DU(0, 2)	95.88	95.40	99.43	50.03	98.20	8.72	
BB(1, 2, 3)	53.31	53.96	71.53	31.60	51.49	22.03	
PM(0.5, 0.5, 0.5, 1.5)	41.87	42.01	42.55	31.82	20.42	21.70	
PM(0.5, 0.5, 0.0, 2.0)	21.02	20.94	6.71	32.73	1.96	28.38	
NA(1, 1)	6.81	6.52	4.03	8.72	2.15	8.88	
NA(0.625, 1.6)	53.94	53.74	36.33	57.25	19.08	44.29	
NA(0.5, 2)	85.51	86.10	68.80	87.19	41.95	71.24	
NA(0.4, 2.5)	98.57	98.75	93.14	98.66	70.87	91.17	
Alternative	$\mathcal{T}_3(\underline{t_1})$	$\mathcal{T}_5(\underline{t_2})$	$\mathcal{T}_{10}(\underline{t_3})$	$\mathcal{SD}_n$	χ^2_{cl2}	KS	AD
NB(1, 0.5)	**4.62**	**4.75**	**5.03**	**4.80**	**4.94**	**5.50**	**5.04**
DU(0, 2)	99.91	99.45	90.16	97.14	79.67	97.24	99.83
BB(1, 2, 3)	54.38	46.54	8.48	49.50	26.31	58.23	68.13
PM(0.5, 0.5, 0.5, 1.5)	27.50	15.85	2.07	31.54	24.90	41.21	41.31
PM(0.5, 0.5, 0.0, 2.0)	34.14	27.60	6.62	21.12	35.44	25.77	29.47
NA(1, 1)	9.00	7.80	3.79	5.85	9.61	7.89	7.95
NA(0.625, 1.6)	42.18	38.23	24.93	46.82	43.57	52.75	55.08
NA(0.5, 2.0)	75.16	68.59	50.94	80.20	66.37	84.09	85.60
NA(0.4, 2.5)	95.48	92.45	81.53	97.16	85.33	97.69	98.66

where $\underline{t_1}$ = (-0.15, -0.05, 0.05), $\underline{t_2}$ = (-0.25, -0.15, -0.05, 0.15, 0.25), $\underline{t_3}$ = (-0.9, -0.75, -0.55, -0.35, -0.15, 0.2, 0.4, 0.6, 0.8, 0.95).

tests in this context would depend upon what the alternative distribution is for a given problem.

In Table 4.6, the first three alternatives following $NB(1, 0.5)$ are under dispersive, the next two alternatives are equally dispersive and the last three alternatives are over dispersive. We observe that $\mathcal{Z}(0.5)$ and AD has the greatest powers for under dispersive alternatives, χ^2_{cl2}, $\mathcal{T}_3(\underline{t_1})$ and $\mathcal{Z}(-0.5)$ have the greatest powers for equally dispersive alternatives, and $\mathcal{Z}(-0.5)$ has the greatest powers for over dispersive alternatives. Meanwhile all of these tests mentioned above have the Type I error rates close to 5% within 0.05±0.005. Also we find that AD performs the best overall for all the alternatives, and $\mathcal{Z}(0.01)$ and $\mathcal{Z}(-0.01)$ control

Table 4.7 Powers(%) for Geometric Tests with $p = 0.6$ at $\alpha = 0.05$ based on 10,000 Samples of Size $n = 50$ and the Mean of the Alternative is 5/3

Alternative	$\mathcal{Z}(0.01)$	$\mathcal{Z}(-0.01)$	$\mathcal{Z}(0.5)$	$\mathcal{Z}(-0.5)$	$\mathcal{Z}(0.9)$	$\mathcal{Z}(-0.9)$	
NB(1, 0.6)	**4.94**	**5.19**	**4.20**	**5.30**	**3.07**	**5.27**	
B(4, 0.17)	83.02	83.97	80.53	73.26	54.61	58.47	
PM(0.5, 0.5, 0.33, 1.0)	27.67	28.46	22.41	24.51	7.23	18.79	
PM(0.5, 0.5, 0.0, 1.33)	6.05	6.23	3.32	10.02	1.52	12.54	
NA(1, 0.67)	4.35	4.58	3.58	5.11	2.11	6.05	
NA(0.67, 1)	18.88	19.43	15.08	21.37	9.50	19.27	
NA(0.33, 2)	93.97	94.17	87.59	93.51	69.57	83.96	
Alternative	$\mathcal{T}_3(\underline{t_1})$	$\mathcal{T}_5(\underline{t_2})$	$\mathcal{T}_{10}(\underline{t_3})$	$\mathcal{SD}_n$	χ^2_{cl3}	KS	AD
NB(1, 0.6)	**4.38**	**4.49**	**5.16**	**4.56**	**5.23**	**4.46**	**5.69**
B(4, 0.17)	62.55	38.69	8.79	71.22	68.03	78.94	82.34
PM(0.5, 0.5, 0.33, 1.0)	13.04	4.98	0.57	18.46	16.11	25.04	27.44
PM(0.5, 0.5, 0.0, 1.33)	12.71	8.07	1.51	7.62	19.32	7.81	10.11
NA(1, 0.67)	5.50	4.62	2.77	4.38	6.78	4.34	5.74
NA(0.67, 1)	17.13	15.93	9.48	18.10	19.78	18.37	22.59
NA(0.33, 2)	86.95	83.32	60.86	91.19	73.53	89.73	94.85

where $\underline{t_1}$ = (-0.15, -0.05, 0.05), $\underline{t_2}$ = (-0.25, -0.15, -0.05, 0.15, 0.25), $\underline{t_3}$ = (-0.9, -0.75, -0.55, -0.35, -0.15, 0.2, 0.4, 0.6, 0.8, 0.95).

Type I error well and are comparative to AD. The KS controls Type I error poorly and AD has higher powers than KS for all alternatives considered here.

In Table 4.7, we can see that $\mathcal{Z}(0.01)$ and $\mathcal{Z}(-0.01)$ have the greatest powers for the first two alternatives following $NB(1, 0.6)$. The χ^2_{cl3} has the greatest powers for the next two alternatives. The AD has the greatest powers for the last two alternatives. However, $\mathcal{Z}(0.01)$, $\mathcal{Z}(-0.01)$ and χ^2_{cl3} control Type I error rates well while the Type I error rate of AD is 5.69%, which is relatively off the mark. Also we find $\mathcal{Z}(0.01)$ and $\mathcal{Z}(-0.01)$ perform the best overall.

From the above analysis based on Table 4.5, Table 4.6 and Table 4.7, we can see that $\mathcal{Z}(0.01)$, $\mathcal{Z}(-0.01)$ perform well overall while for the other $\mathcal{Z}(t)$ tests with t far away from zero, their performance varies depending on different alternatives, and in some cases, they do not control Type I Error very well. For example, the Type I Error rates of $\mathcal{Z}(0.9)$ for all the cases of p are smaller than 5% by more than 1%. The performances of $\mathcal{T}_3(\underline{t_1})$, $\mathcal{T}_5(\underline{t_2})$ and $\mathcal{T}_{10}(\underline{t_3})$ also vary according to the value of p and the alternative hypotheses. The $\mathcal{SD}_n$

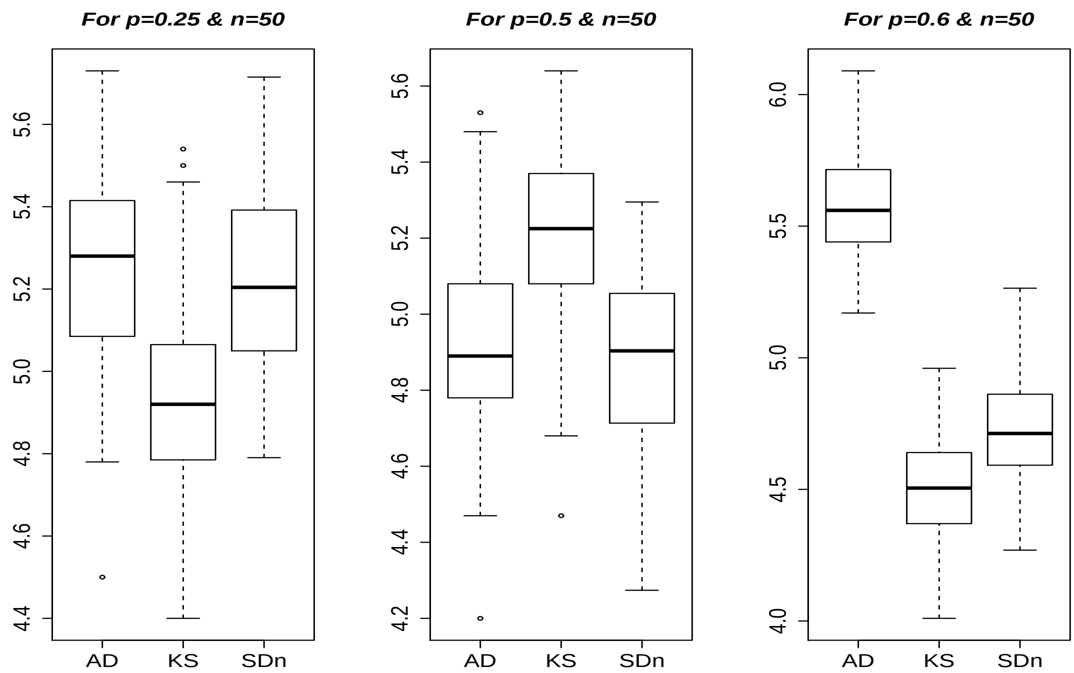

Figure 4.3 Empirical Type I error rates of $\mathcal{SD}_n$, KS and AD boxplots comparison.

has tolerably lower powers than KS and AD for most of the alternatives and higher powers for some alternatives, but has the Type I error rate better controlled (See Figure 4.3 for the Type I Error boxplot comparison of $\mathcal{SD}_n$, KS and AD for each p case based on 100 replications on the empirical Type I error rates).

In conclusion, we recommend $\mathcal{Z}(0.01)$ and $\mathcal{Z}(-0.01)$, the K&K method with single t with t close to zero for the goodness-of-fit tests of geometric distribution against unknown alternatives. This result is consistent with the finding in Kocherlakota and Kocherlakota (1986), who recommended $\mathcal{Z}(t)$ where t is in the neighborhood of zero to test Poisson distribution in particular. Moreover, we recommend $\mathcal{SD}_n$ for the goodness-of-fit test of geometric distribution because it has robust performances over all dispersive cases and control Type I error well. In addition, $\mathcal{SD}_n$ does not need a selection of t.

4.3 Real Example Analysis

In order to illustrate the proposed goodness-of-fit tests for geometric distribution, we apply the above-mentioned tests to a library circulation data presented in Best and Rayner (1989). The data describe the frequency of books checked out (See Table 4.8 for details). Best and Rayner (1989) suggested that geometric distribution is a proper fit to the data by smooth tests.

Table 4.8 Library Circulation

Number of Checkout	1	2	3	4	5	6	7
Number of Books	65	26	12	10	5	3	1

The MLE of geometric parameter $\hat{p} = 0.502$. We apply our recommended tests $\mathcal{Z}(0.01)$, $\mathcal{Z}(-0.01)$, $\mathcal{SD}_n$ and the best test for the case of $p = 0.5$ according to Table 4.6 AD to test whether this library circulation data follows geometric distribution, and calculate the values of the test statistics from the data and the corresponding p-values. The bootstrap p-values of $\mathcal{SD}_n$ and AD are calculated as follows. (1) Generate 10,000 sample from the geometric distribution with $p = 0.5$ and compute the corresponding 10,000 values of test statistic; (2)

Calculate the the value of test statistic from the data. (3) The p-value is the proportion of the values of test statistics, which are greater than or equal to the one calculated from the data, out of 10,000 bootstrap samples (see Best and Rayner (2003) and Gulati and Neus (2003)). Table 4.9 shows the results.

Table 4.9 Goodness-of-fit Tests Results based on the Real Data

	$\mathcal{Z}(0.01)$	$\mathcal{Z}(-0.01)$	$\mathcal{SD}_n$	AD
Statistic value	0.9600	-0.9638	0.9710	0.4198
P-value	0.1685	0.1676	0.5774	0.3806

The p-values of all the tests in Table 4.9 are greater than 0.05 so that we do not reject the null hypothesis that this library circulation data is from an equally dispersive geometric distribution. This real data analysis result is consistent with the result based on Table 4.6 which is that $\mathcal{Z}(0.01)$, $\mathcal{Z}(-0.01)$, $\mathcal{SD}_n$ and AD perform well for the case of $p = 0.5$.

CHAPTER 5

SIMULATION STUDY FOR TESTING BGD(B&D)

In this chapter, we evaluate the performance of the three goodness-of-fit tests for BGD(B&D) described in Section 3.2.1, 3.2.2 and 3.2.3, respectively, by simulation studies and analyze an accident records from Arbous and Kerrich (1951) by using those tests for illustration. We explore the empirical distributions of the new test statistics $\mathcal{MZ}(\underline{t})$, $\mathcal{MT}_q(\underline{t})$ and $\mathcal{MSD}_n$ and compare their Type I error rates and powers with those of the Chi-square and the two-dimensional K-S tests.

5.1 Empirical Distribution of Test Statistics

In order to analyze the test statistic $\mathcal{MZ}(\underline{t})$, we select multiple tests depending on the value of $\underline{t}$. In this analysis, the value of each element of $\underline{t}$ is chosen from ± 0.01, ± 0.5 and ± 0.9 so that some of the $\underline{t}$'s are close to zero and others are not, similar to the study in the univariate case. Therefore, there are 36 different test statistics in total as described in Table 5.1. In this section, we analyze the asymptotic behavior of the percentiles of $\mathcal{MZ}(\underline{t})$. Although these percentiles should be asymptotically correspond to those from the standard normal

Table 5.1 $\mathcal{MZ}(\underline{t})$ Tests

$\mathcal{MZ}(\underline{t_1})$ $\mathcal{MZ}$(0.01, -0.5)	$\mathcal{MZ}(\underline{t_2})$ $\mathcal{MZ}$(0.01, -0.9)	$\mathcal{MZ}(\underline{t_3})$ $\mathcal{MZ}$(0.5, 0.01)	$\mathcal{MZ}(\underline{t_4})$ $\mathcal{MZ}$(0.5, -0.01)	$\mathcal{MZ}(\underline{t_5})$ $\mathcal{MZ}$(0.5, 0.9)	$\mathcal{MZ}(\underline{t_6})$ $\mathcal{MZ}$(0.5, -0.9)
$\mathcal{MZ}(\underline{t_7})$ $\mathcal{MZ}$(-0.5, 0.01)	$\mathcal{MZ}(\underline{t_8})$ $\mathcal{MZ}$(-0.5, -0.01)	$\mathcal{MZ}(\underline{t_9})$ $\mathcal{MZ}$(-0.01, 0.01)	$\mathcal{MZ}(\underline{t_{10}})$ $\mathcal{MZ}$(-0.5, 0.5)	$\mathcal{MZ}(\underline{t_{11}})$ $\mathcal{MZ}$(-0.9, 0.9)	$\mathcal{MZ}(\underline{t_{12}})$ $\mathcal{MZ}$(0.9, 0.01)
$\mathcal{MZ}(\underline{t_{13}})$ $\mathcal{MZ}$(0.9, -0.01)	$\mathcal{MZ}(\underline{t_{14}})$ $\mathcal{MZ}$(0.9, -0.5)	$\mathcal{MZ}(\underline{t_{15}})$ $\mathcal{MZ}$(-0.9, 0.01)	$\mathcal{MZ}(\underline{t_{16}})$ $\mathcal{MZ}$(-0.9, -0.01)	$\mathcal{MZ}(\underline{t_{17}})$ $\mathcal{MZ}$(-0.9, 0.5)	$\mathcal{MZ}(\underline{t_{18}})$ $\mathcal{MZ}$(-0.9, -0.5)
$\mathcal{MZ}(\underline{t_{19}})$ $\mathcal{MZ}$(0.01, 0.01)	$\mathcal{MZ}(\underline{t_{20}})$ $\mathcal{MZ}$(0.01, -0.01)	$\mathcal{MZ}(\underline{t_{21}})$ $\mathcal{MZ}$(-0.01, -0.01)	$\mathcal{MZ}(\underline{t_{22}})$ $\mathcal{MZ}$(0.5, 0.5)	$\mathcal{MZ}(\underline{t_{23}})$ $\mathcal{MZ}$(0.5, -0.5)	$\mathcal{MZ}(\underline{t_{24}})$ $\mathcal{MZ}$(-0.5, -0.5)
$\mathcal{MZ}(\underline{t_{25}})$ $\mathcal{MZ}$(0.9, 0.9)	$\mathcal{MZ}(\underline{t_{26}})$ $\mathcal{MZ}$(0.9, -0.9)	$\mathcal{MZ}(\underline{t_{27}})$ $\mathcal{MZ}$(-0.9, -0.9)	$\mathcal{MZ}(\underline{t_{28}})$ $\mathcal{MZ}$(0.01, 0.9)	$\mathcal{MZ}(\underline{t_{29}})$ $\mathcal{MZ}$(-0.01, 0.9)	$\mathcal{MZ}(\underline{t_{30}})$ $\mathcal{MZ}$(-0.01, -0.9)
$\mathcal{MZ}(\underline{t_{31}})$ $\mathcal{MZ}$(0.01, 0.5)	$\mathcal{MZ}(\underline{t_{32}})$ $\mathcal{MZ}$(-0.01, 0.5)	$\mathcal{MZ}(\underline{t_{33}})$ $\mathcal{MZ}$(-0.01, -0.5)	$\mathcal{MZ}(\underline{t_{34}})$ $\mathcal{MZ}$(0.9, 0.5)	$\mathcal{MZ}(\underline{t_{35}})$ $\mathcal{MZ}$(-0.5, 0.9)	$\mathcal{MZ}(\underline{t_{36}})$ $\mathcal{MZ}$(-0.5, -0.9)

distribution, their empirical behavior can be quite erratic due to the sample size and the choice of parameters.

We compute the empirical critical points of $\mathcal{MZ}(\underline{t})$ at $\alpha = 0.01$, 0.05 and 0.10 based on 10,000 Monte-Carlo simulation replications with $n = 100$ or 500, for each of the following cases, $\underline{p} = (0.8, 0.7, 0.9)$, $\underline{p} = (0.50, 0.55, 0.45)$, and $\underline{p} = (0.25, 0.20, 0.15)$ and compared them with the corresponding standard normal percentiles in Figure 5.1, Figure 5.2 and Figure 5.3, respectively.

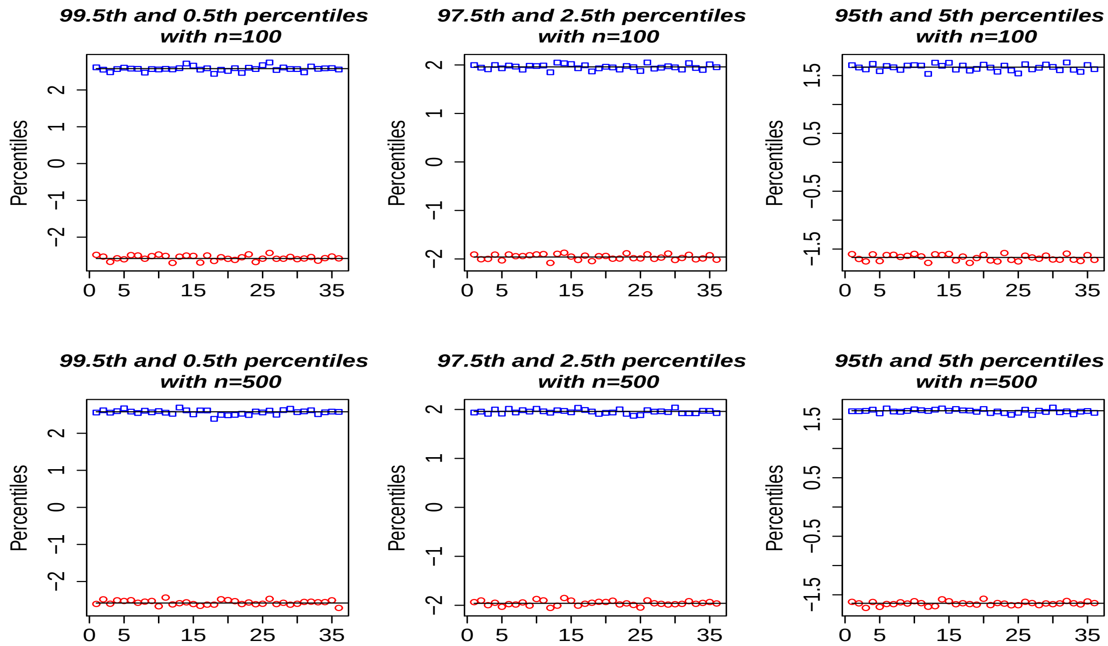

Figure 5.1 Empirical percentiles of $\mathcal{MZ}(\underline{t})$, which is identified by the subscript of $\underline{t}$ on the x-axis, for $\underline{p} = (0.8, 0.9, 0.7)$.

In Figures 5.1 and 5.2, one can observe that the empirical critical points of all the $\mathcal{MZ}(\underline{t})$ tests scatter tightly around their corresponding standard normal percentiles, which are denoted in solid lines on the graphs at all the significance levels considered here. As expected, the empirical points gather closer to the corresponding theoretical percentiles as n increasing from 100 to 500. In Figure 5.2, we notice that $\mathcal{MZ}(t_{25})$ or $\mathcal{MZ}(0.9, 0.9)$ performs the worst among all the tests since its empirical critical points locate farther to the corresponding theoretical percentiles than those of any other tests at all significance levels.

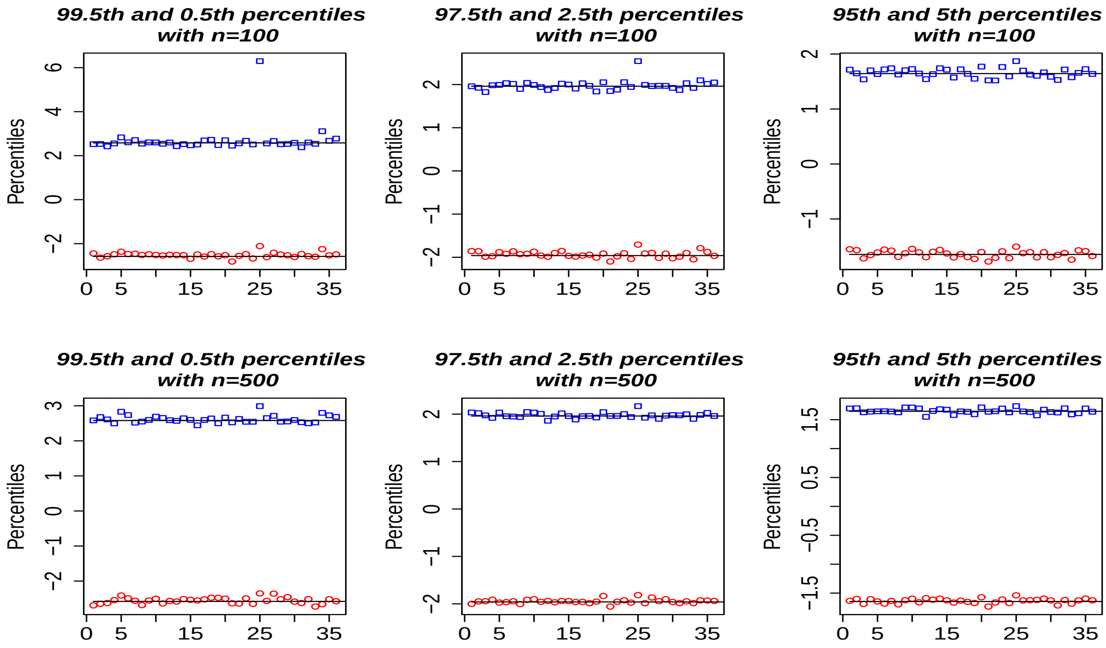

Figure 5.2 Empirical percentiles of $\mathcal{MZ}(\underline{t})$, which is identified by the subscript of $\underline{t}$ on the x-axis, for $\underline{p} = (0.50, 0.55, 0.45)$.

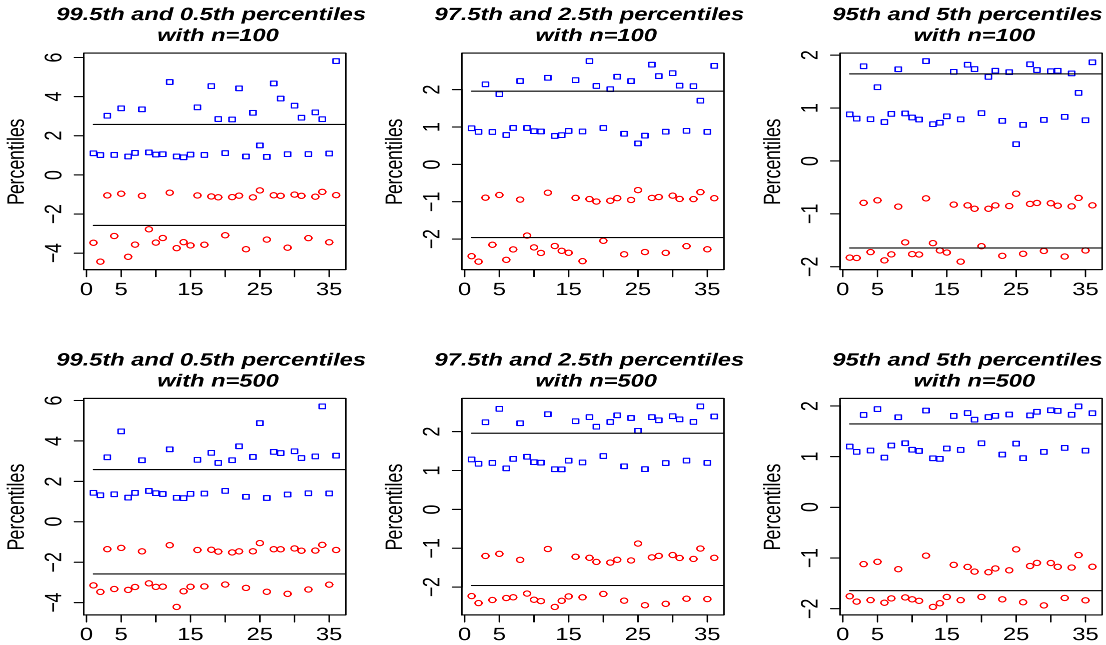

Figure 5.3 Empirical percentiles of $\mathcal{MZ}(\underline{t})$, which is identified by the subscript of $\underline{t}$ on the x-axis, for $\underline{p} = (0.25, 0.20, 0.15)$.

In Figure 5.3 for $\underline{p} = (0.25,\ 0.20,\ 0.15)$, the empirical points are widely off the lines. We find that quite a large sample size (> 5000) is needed to make empirical critical points be around the corresponding theoretical percentiles.

Now we explore the asymptotic behavior of the distributions of $\mathcal{MT}_q(\mathbf{t})$. We calculate 10,000 values of $\mathcal{MT}_q(\mathbf{t})$ based on Monte Carlo simulations of BGD(B&D), replication size equal to 10,000 and sample size $n = 100$ or 500, with $\underline{p} = (0.8,\ 0.7,\ 0.9)$, $\underline{p} = (0.5,\ 0.55,\ 0.45)$ or $\underline{p} = (0.25,\ 0.20,\ 0.15)$. Here we choose $q = 3$, 5 or 10 and three types of $\mathbf{t}$ values: all the elements of $\mathbf{t}$, are around zero, are close to 1 or -1 or are bivariate uniformly distributed in $(-1, 1) \times (-1, 1)$ (see Table 5.2 for the list of these test statistics). Therefore there are nine tests for each combination of q and the three types of $\mathbf{t}$ values. Then we plot the empirical density curves of the nine test statistics for each combination of $\underline{p}$ and n based on their corresponding 10,000 test statistic values, and compare them with the corresponding theoretical density curves of Chi-square distributions. Some of these graphs appear in Figure 5.4, 5.5 and 5.6.

Table 5.2 $\mathcal{MT}_q(\mathbf{t})$ Tests

$\mathcal{MT}_3(\mathbf{t}_1)$	$\mathbf{t}_1$=[(0.01, 0.02),(-0.015, 0.01),(-0.02, -0.025)]
$\mathcal{MT}_3(\mathbf{t}_2)$	$\mathbf{t}_2$=[(0.9, -0.92),(-0.85, 0.9),(0.87, 0.91)]
$\mathcal{MT}_3(\mathbf{t}_3)$	$\mathbf{t}_3$=[(-0.7, -0.2),(-0.9, 0.1),(0.07, 0.6)]
$\mathcal{MT}_5(\mathbf{t}_1)$	$\mathbf{t}_1$=[(0.03, 0.016)(0.01, -0.01)(-0.025, 0.01)(-0.005, -0.035)(0.02, 0.015)]
$\mathcal{MT}_5(\mathbf{t}_2)$	$\mathbf{t}_2$=[(0.89, 0.91)(0.85, -0.9)(-0.9, 0.87)(-0.91, -0.9)(0.88, -0.9)]
$\mathcal{MT}_5(\mathbf{t}_3)$	$\mathbf{t}_3$=[(-0.65, -0.17)(0.3, -0.09)(-0.06, 0.5)(0.4, 0.05)(0.93, -0.32)]
$\mathcal{MT}_{10}(\mathbf{t}_1)$	$\mathbf{t}_1$=[(0.01, 0.015),(0.005, 0.02),(-0.01, 0.01),(-0.01, 0.025),(0.01, -0.02), (-0.01, -0.02),(-0.01, -0.015),(0.025, -0.01),(-0.03, -0.01),(0.02, 0.03)]
$\mathcal{MT}_{10}(\mathbf{t}_2)$	$\mathbf{t}_2$=[(0.91, 0.9),(0.87, 0.85),(0.85, 0.8),(-0.94, -0.9),(-0.81, -0.89), (-0.85, -0.86),(0.9, -0.87),(0.85, -0.9),(0.88, -0.85),(-0.9, 0.9)]
$\mathcal{MT}_{10}(\mathbf{t}_3)$	$\mathbf{t}_3$=[(0.5, 0.03),(-0.0045, -0.5),(-0.85,-0.25),(-0.81, -0.77),(-0.4, 0.45), (0.43, 0.33),(-0.62, 0.004),(-0.23, -0.03),(-0.55, -0.4),(0.44, -0.78)]

When $\underline{p} = (0.8,\ 0.7,\ 0.9)$, the empirical density curves of all the nine tests with $n = 100$ or 500 match well with their corresponding theoretical ones. For example, Figure 5.4 shows the histograms of $\mathcal{MT}_3(\mathbf{t}_1)$, $\mathcal{MT}_3(\mathbf{t}_2)$ and $\mathcal{MT}_3(\mathbf{t}_3)$ with $n = 100$ and $n = 500$, and they match well with the density curve, denoted in solid line, of Chi-square distribution with three degrees of freedom. When $\underline{p} = (0.50,\ 0.55,\ 0.45)$, the singularity of the covariance matrix $\hat{\Gamma}$ occurs when we calculate the value of $\mathcal{MT}_{10}(\mathbf{t}_1)$. Therefore, $\mathcal{MT}_{10}(\mathbf{t}_1)$ is not included into

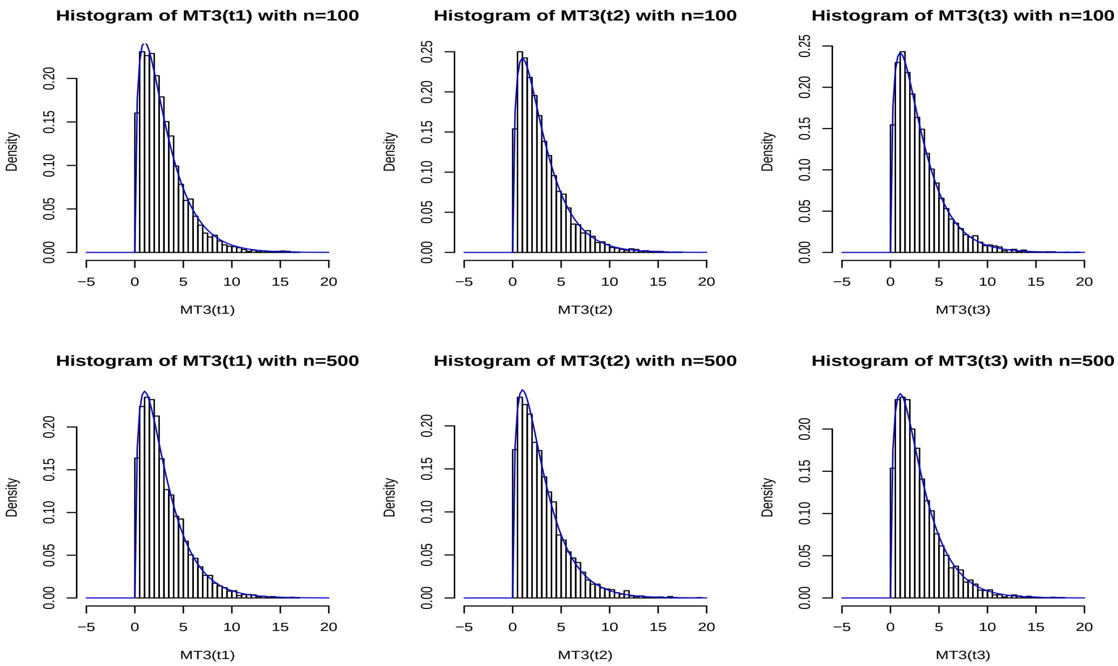

Figure 5.4 Histograms of $\mathcal{MT}_3(\mathbf{t}_1)$, $\mathcal{MT}_3(\mathbf{t}_2)$ and $\mathcal{MT}_3(\mathbf{t}_3)$ for $\underline{p} = (0.8,\ 0.7,\ 0.9)$ compared with the density curve of Chi-square distribution with 3 d.f.

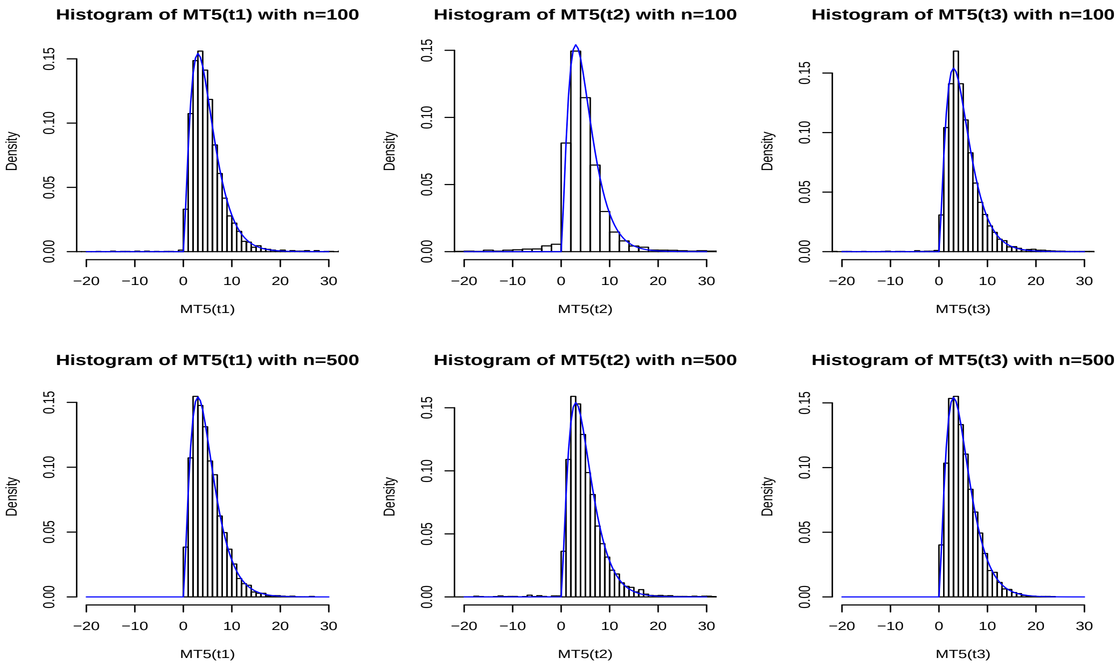

Figure 5.5 Histograms of $\mathcal{MT}_5(\mathbf{t}_1)$, $\mathcal{MT}_5(\mathbf{t}_2)$ and $\mathcal{MT}_5(\mathbf{t}_3)$ for $\underline{p} = (0.50,\ 0.55,\ 0.45)$ compared with the density curve of Chi-square distribution with 5 d.f.

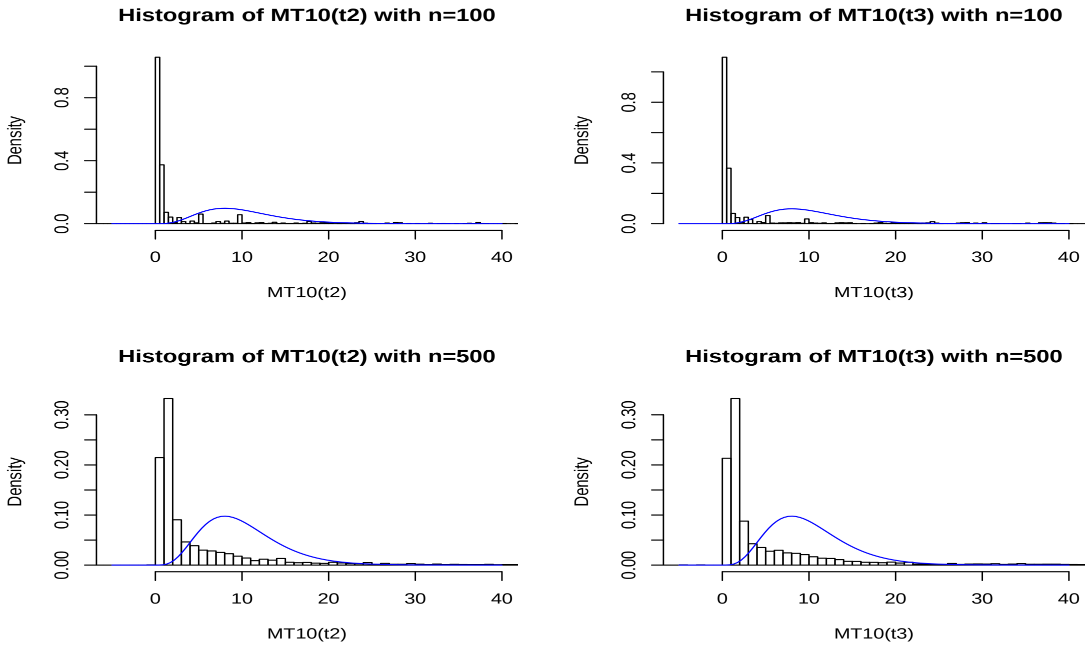

Figure 5.6 Histograms of $\mathcal{MT}_{10}(\mathbf{t}_2)$ and $\mathcal{MT}_{10}(\mathbf{t}_3)$ for $\underline{p}$ = (0.25, 0.20, 0.15) compared with the density curve of Chi-square distribution with 10 d.f.

the comparison study in this case of $\underline{p}$. For the rest of the tests, the empirical density curves of the test statistics have good matches with their corresponding theoretical ones except $\mathcal{MT}_5(\mathbf{t}_2)$ and $\mathcal{MT}_{10}(\mathbf{t}_2)$. As shown in Figure 5.5 for $\mathcal{MT}_5(\mathbf{t}_2)$, there are a large number of negative values when $n = 100$ and the histogram does not match well with the density curve of Chi-square distribution with five degrees of freedom. When $\underline{p}$ = (0.25, 0.20, 0.15), $\mathcal{MT}_{10}(\mathbf{t}_1)$ is not included in the comparative study again because the covariance matrix $\hat{\Gamma}$ is singular. In this case, none of the empirical density curves of the other eight tests is close in shape to their corresponding theoretical ones. For instance, Figure 5.6 shows the empirical density curves of $\mathcal{MT}_{10}(\mathbf{t}_2)$ and $\mathcal{MT}_{10}(\mathbf{t}_3)$ with $n = 100$ and $n = 500$, all of which do not match well with the density curve of Chi-square distribution with 10 degrees of freedom. The empirical density curves of the tests shift to the left with respect to the theoretical ones.

One can observe that the empirical density curve is closer to its theoretical one when $n = 500$ than when $n = 100$ for all the tests in each case on $\underline{p}$. The $\mathcal{MT}_q(\mathbf{t})$ type tests perform better when all elements of $\mathbf{t}$ are around zero or generated from bivariate uniform distribution in (-1, 1) $\times$ (-1, 1) than when close to 1 or -1. The singularity of the covariance matrix $\hat{\Gamma}$ occurs more often when all elements of $\mathbf{t}$ are around zero than in the other cases and more often when $q = 10$ than when $q = 5$ or $q = 3$. We have similar finding when testing the goodness-of-fit of univariate geometric distribution. Furthermore, the empirical density curves match better with their corresponding theoretical ones for $q = 3$ or $q = 5$ than for $q = 10$ in all the cases on $\underline{p}$ considered here.

We study the null distribution of $\mathcal{MSD}_n$ by constructing a table of its empirical critical points and plotting the empirical density curves for various $\underline{p}$ and n based on simulations. We calculate 10,000 values of $\mathcal{MSD}_n$ based on 10,000 BGD(B&D) random samples of size n = 50, 100, 200, 500, 750, 1000 or 1200 for $\underline{p}$ = (0.8, 0.7, 0.9), $\underline{p}$ = (0.50, 0.55, 0.45) or $\underline{p}$ = (0.25, 0.20, 0.15). Then based on the 10,000 $\mathcal{MSD}_n$ values, we compute the empirical critical points at α=0.01, 0.05 and 0.10 for each combination of n and $\underline{p}$. Table 5.3 shows the results. Meanwhile, we construct the density curves of $\mathcal{MSD}_n$ with the 10,000 $\mathcal{MSD}_n$ values for each combination of n and $\underline{p}$. See Figure 5.7 for the results. We evaluate the supremum over the region $(-1, 1) \times (-1, 1)$ by the function *optim* in R 2.14 with option *"L-BFGS-B"* method. In order to make the calculation stable with the function *optim*, we set the supremum over a narrower region $(-0.5, 0.5) \times (-0.5, 0.5)$.

We observe from Figure 5.7 that for each $\underline{p}$ as n increases, the density curves approach more and more close to each other, indicating that a limiting distribution of $\mathcal{MSD}_n$ is being attained. The curves are more clustered with each other for $\underline{p} = (0.8, 0.7, 0.9)$ and $\underline{p} = (0.50, 0.55, 0.45)$ than for $\underline{p} = (0.25, 0.20, 0.15)$. Table 5.3 indicates that the critical points of $\mathcal{MSD}_n$ becomes more and more stable as n increases for each case on $\underline{p}$.

Table 5.3 Empirical Critical Points of $\mathcal{MSD}_n$ based on 10,000 Replications

	α	0.01	0.05	0.1
$\underline{p} = (0.8, 0.7, 0.9)$	n=50	2.974048	2.449002	2.161685
	n=100	3.091729	2.452455	2.132254
	n=200	3.025133	2.445564	2.131091
	n=500	2.999105	2.457129	2.166694
	n=750	3.025595	2.450285	2.140294
	n=1000	3.016462	2.443400	2.141058
	n=1200	3.017077	2.425687	2.133596
$\underline{p} = (0.50, 0.55, 0.45)$	n=50	2.894928	2.315283	2.032061
	n=100	2.928199	2.343221	2.067772
	n=200	2.942843	2.362074	2.05612
	n=500	2.972083	2.380423	2.076762
	n=750	2.985481	2.395410	2.076814
	n=1000	2.984349	2.408860	2.084174
	n=1200	2.927579	2.371872	2.056918
$\underline{p} = (0.25, 0.20, 0.15)$	n=50	3.279629	2.207581	1.544216
	n=100	3.806083	2.463971	1.78076
	n=200	3.896545	2.480614	1.904846
	n=500	3.456446	2.448386	1.89607
	n=750	3.369802	2.417712	1.907302
	n=1000	3.333352	2.349017	1.894366
	n=1200	3.288807	2.339390	1.899354

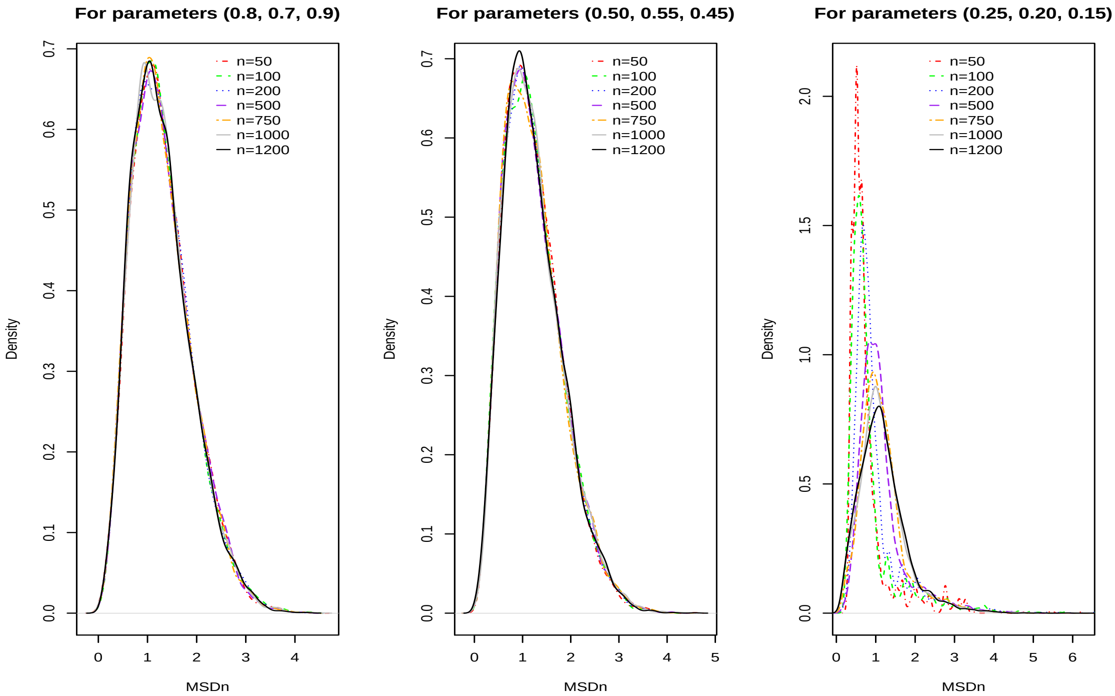

Figure 5.7 Empirical density curves of $\mathcal{MSD}_n$ with various $\underline{p}$ and n.

5.2 Evaluating Tests by Comparison

In order to evaluate our proposed goodness-of-fit tests for the BGD(B&D), we compare their Type I error rates and powers with those of Chernoff and Lehmann Chi-square and K-S tests using various bivariate discrete distributions as the alternatives, for $\underline{p}$ = (0.8, 0.7, 0.9) and $\underline{p}$ = (0.50, 0.55, 0.45) and with sample size n = 100 and n = 500. We do not consider $\underline{p} = (0.25, 0.20, 0.15)$ here for the following reasons. First we know from Section 5.1 that in this case the empirical distributions of $\mathcal{MZ}(\underline{t})$ and $\mathcal{MT}_q(\mathbf{t})$ type tests depart greatly from the theoretical ones when n = 100 and n = 500. Also, this case arises very rarely in the real world since the component failure probability $1 - p_i$ for i = 1, 2, 3 are too high in this case.

Three types of bivariate discrete distributions, which are presented in Kocherlakota and Kocherlakota (1992), are chosen as the alternatives.

(1) BVP(λ_1, λ_2, λ_3): a bivariate Poisson(BVP) distribution, which can be constructed as follows. Let U, V, W be three independent random variables and follow poisson distribution

with the corresponding parameters μ_1, μ_2, μ_3. Let $X = U + V$ and $Y = V + W$. Then random variable (X, Y) have BVP(λ_1, λ_2, λ_3) with $\lambda_1 = \mu_1 + \mu_3$, $\lambda_2 = \mu_2 + \mu_3$ and $\lambda_3 = \mu_3$.

(2) $BVB(n, p_{11}, p_{10}, p_{01})$: a Type I bivariate binomial(BVB) distribution described in Kocherlakota and Kocherlakota (1992), which can be derived from a sequence of n bivariate Bernoulli (U_i, V_i) trials with the joint probability mass function $P(U_i = r, V_i = s) = p_{rs}$ for r, $s = 0$, 1 and $i = 1$, 2,..., n where $p_{00} + p_{01} + p_{10} + p_{11} = 1$. Let $X = \sum_{i=1}^{n} U_i$ and $Y = \sum_{i=1}^{n} V_i$. The random variable (X, Y) is distributed by $BVB(n, p_{11}, p_{10}, p_{01})$. The marginal distributions on X and Y are binomial distributions with the same number of trials n and the probability of success $p_{1+} = p_{10} + p_{11}$ and $p_{+1} = p_{01} + p_{11}$, respectively. Bivariate bernoulli random number (u_i, v_i) can be simulated based on the marginal distribution on U_i and the conditional distribution on V_i given u_i. The marginal distribution of U_i is bernoulli distribution with probability of success $p_{1+} = p_{10} + p_{11}$. If $u_i = 0$, $P(V_i|0) = p_{0s}/(1 - p_{1+})$, else $P(V_i|1) = p_{1s}/p_{1+}$. Then v_i is obtained by inversing the cumulative distribution function of $V_i|u_i$.

(3) $BVNB(r, p_1, p_2)$: a bivariate negative binomial (BVNB) distribution. Consider n independent trials, each of which has three possible outcomes with corresponding probabilities p_1, p_2 and $1 - p_1 - p_2$. Let X and Y denote the number of the first and the second outcomes occurring before the rth occurrence of the third outcome, respectively. Then random variable (X, Y) is BVNB distributed with parameters r, p_1, and p_2. Notice that here r is extended to take any positive real number. The marginal distribution on Y is negative binomial distribution with parameters r and $(1 - p_1 - p_2)/(1 - p_1)$. The conditional distribution on X given y is also negative binomial distribution with parameters $r + y$ and $1 - p_1$. Obviously, BVNB random sample can be generated based on the the marginal distribution on Y and the conditional distribution on X given y.

More detailed descriptions of the three bivariate distributions can be found in Kocherlakota and Kocherlakota (1992). We choose the values of the parameter of the alternative distributions based on the constraint that the first moments, namely EX, EY and EXY, of the null and the alternative distributions are equal. Meanwhile, since the value of BGD(B&D) variable starts from (1, 1), we need to take $(X+1, Y+1)$ as the random variables from the alternative distributions, where (X, Y) follows any of the above-mentioned three distributions. Table 5.4 displays the notations of the alternative distributions.

Table 5.4 Notations of Alternative Distributions

	Alternative	Notation
$\underline{p} = (0.8, 0.7, 0.9)$	BVP(1.59, 0.72, 0.98)+(1, 1)	BVP_1
	BVB(10, 0.14, 0.11, 0.03)+(1, 1)	BVB_1
	BVNB(2.5, 0.37, 0.24)+(1, 1)	$BVNB_1$
$\underline{p} = (0.50, 0.55, 0.45)$	BVP(0.20, 0.16, 0.13)+(1, 1)	BVP_2
	BVB(5, 0.03, 0.04, 0.03) +(1, 1)	BVB_2
	BVNB(1.1, 0.20, 0.18)+(1, 1)	$BVNB_2$
$\underline{p} = (0.64, 0.56, 0.86)$	BVP(0.96, 0.66, 0.27)+(1, 1)	BVP_3
	BVB(2, 0.42, 0.19, 0.05)+(1, 1)	BVB_3
	BVNB(2.83, 0.24, 0.18)+(1, 1)	$BVNB_3$

In order to do the comparison with Chi-square test for $\underline{p} = (0.8, 0.7, 0.9)$, we construct the test statistic, which is denoted by $\mathcal{M}\chi^2_{cl1}$, based on the following six regions: $0 < X \leq 2$, $0 < Y \leq 2$; $(2 < X \leq 3,\ 0 < Y \leq 3) \cup (0 < X \leq 2,\ 2 < Y \leq 3)$; $(3 < X \leq 4,\ 0 < Y \leq 4) \cup (0 < X \leq 3,\ 3 < Y \leq 4)$; $(4 < X \leq 5,\ 0 < Y \leq 5) \cup (0 < X \leq 4,\ 4 < Y \leq 5)$; $(5 < X \leq 7,\ 0 < Y \leq 7) \cup (0 < X \leq 5,\ 5 < Y \leq 7)$; otherwise. Clearly $\mathcal{M}\chi^2_{cl1}$ can be asymptotically Chi-square distributed with two degrees of freedom when the null hypothesis is true and sample size is fairly large. Similarly, in the Chi-square test for $\underline{p} = (0.50, 0.55, 0.45)$, the test statistic, which is denoted by $\mathcal{M}\chi^2_{cl2}$, is constructed based on the following five regions: $X = 1, Y = 1$; $X = 2, Y = 1$; $X = 1, Y = 2$; $X = 2, Y = 2$; otherwise. Hence, $\mathcal{M}\chi^2_{cl2}$ can be asymptotically Chi-square distributed with one degree of freedom when the null hypothesis is true and sample size is fairly large. Table 5.5 shows the the empirical critical points of $\mathcal{M}\chi^2_{cl1}$ and $\mathcal{M}\chi^2_{cl2}$. From this table, we can see that the empirical and theoretical critical points differ greatly with the empirical values always dominating the theoretical ones. For the sake of fair comparison of tests, we use the empirical critical points for $\mathcal{M}\chi^2_{cl1}$ and $\mathcal{M}\chi^2_{cl2}$.

Table 5.5 Empirical Critical Points of $\mathcal{M}\chi^2_{cl1}$ and $\mathcal{M}\chi^2_{cl2}$ based on 10,000 Replications

	α	0.01	0.05	0.1
$\mathcal{M}\chi^2_{cl1}$	n=100	13.26010	9.48909	7.79239
	n=500	12.89645	9.50713	7.77927
χ^2_2		9.21034	5.99147	4.60517
$\mathcal{M}\chi^2_{cl2}$	n=100	7.83930	5.12474	4.04465
	n=500	7.99846	5.21191	4.08826
χ^2_1		6.63490	3.84146	2.70554

We implement the algorithm of the two dimensional K-S test described in Section 14.7 of Press et al. (2002) to calculate the value of K-S test statistic denoted by $\mathcal{MKS}$. Table 5.6 displays the critical points for $\mathcal{MKS}$. These critical points are used to compute the empirical α and power of the two-dimensional K-S test for comparison.

Table 5.6 Empirical Critical Points of $\mathcal{MKS}$ based on 10,000 Replications

	α	0.01	0.05	0.1
$\underline{p} = (0.8, 0.7, 0.9)$	n=100	0.09773	0.10891	0.13171
	n=500	0.06025	0.05051	0.04510
$\underline{p} = (0.50, 0.55, 0.45)$	n=100	0.05141	0.05873	0.07474
	n=500	0.03214	0.02612	0.02315

Now we calculate the empirical significance level of each test in the comparison study at $\alpha = 0.05$ based on 10,000 BGD(B&D) simulation replications of $n = 100$ or $n = 500$ for $\underline{p}$ = (0.8, 0.7, 0.9) or $\underline{p}$ = (0.50, 0.55, 0.45). Figure 5.8 presents the empirical Type I error rates of the 36 $\mathcal{MZ}(\underline{t})$ tests at $\alpha = 0.05$. Figure 5.9 displays the empirical Type I error rates of the $\mathcal{MT}_q(\mathbf{t})$ tests at $\alpha = 0.05$. Table 5.7 shows the empirical Type I error rates of $\mathcal{MSD}_n$, $\mathcal{MKS}$, $\mathcal{M}\chi^2_{cl1}$ and $\mathcal{M}\chi^2_{cl2}$ at $\alpha = 0.05$.

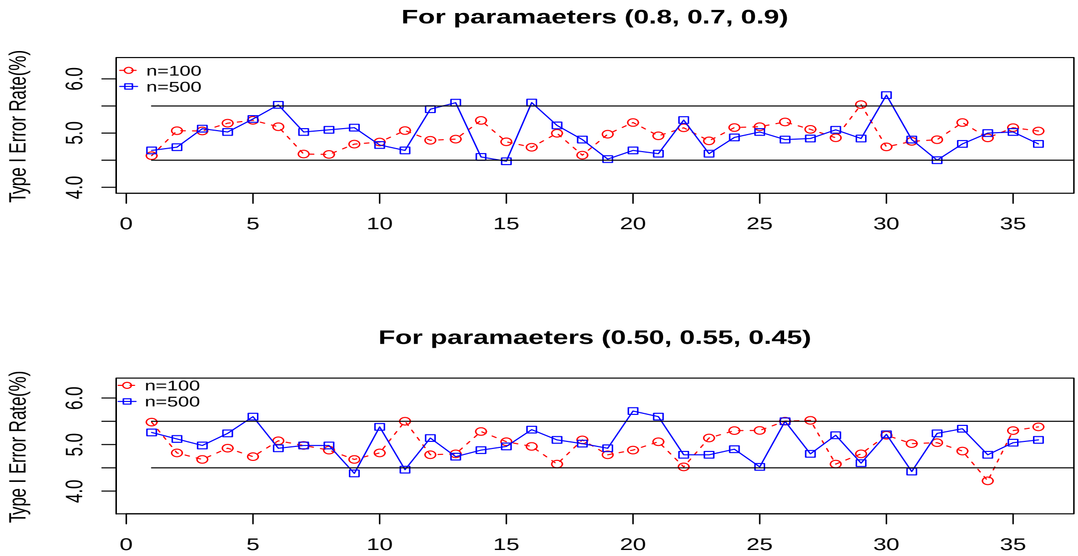

Figure 5.8 Empirical Type I error rates (%) of $\mathcal{MZ}(\underline{t})$ at $\alpha = 5\%$ with the subscript of $\underline{t}$ on the x-axis identifying the test.

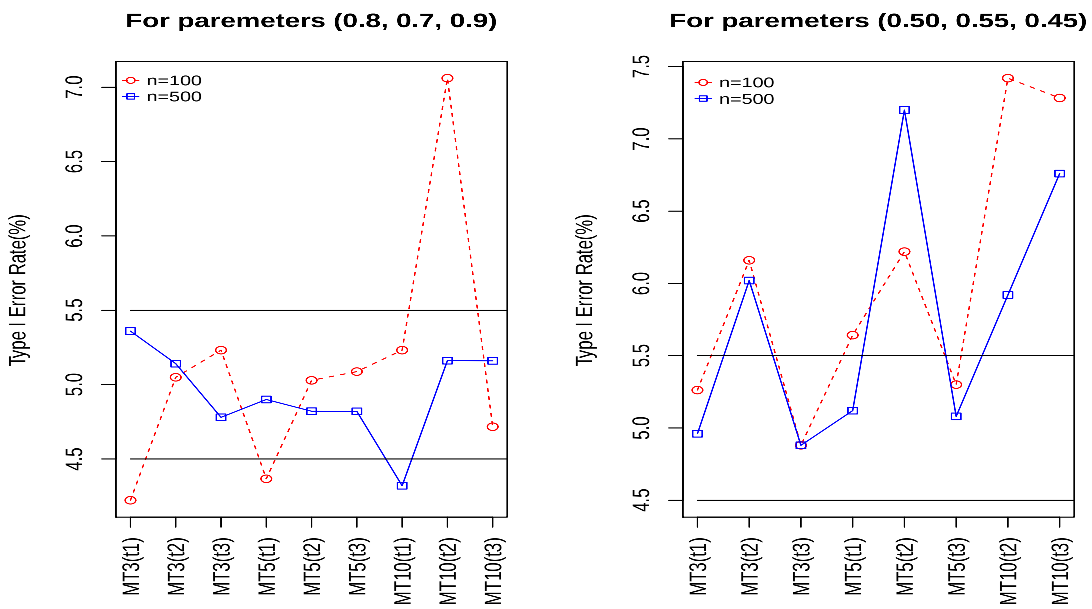

Figure 5.9 Empirical Type I error rates (%) of $\mathcal{MT}_q(\mathbf{t})$ at $\alpha = 5\%$.

Table 5.7 Empirical Type I Error Rates (%) of $\mathcal{MSD}_n$, $\mathcal{MKS}$, $\mathcal{M}\chi^2_{cl1}$ and $\mathcal{M}\chi^2_{cl2}$ at $\alpha = 5\%$ based on 10,000 Replications

	Tests	$\mathcal{M}\chi^2_{cl1}$	$\mathcal{M}\chi^2_{cl2}$	$\mathcal{MKS}$	$\mathcal{MSD}_n$
$\underline{p}$ = (0.8, 0.7, 0.9)	n=100	4.89	*NA*	5.76	5.11
	n=500	4.70	*NA*	6.40	4.94
$\underline{p}$ = (0.50, 0.55, 0.45)	n=100	*NA*	4.95	5.94	5.27
	n=500	*NA*	5.04	6.00	5.13

In Figures 5.8 and 5.9 we can see that it is not necessarily true that Type I errors have been controlled better as n increases from 100 to 500, and in some case even worse. Meanwhile, we find that $\mathcal{MZ}(\underline{t})$ tests with single $\underline{t}$ perform better in controlling the Type I Error than $\mathcal{MT}_q(\mathbf{t})$ tests with multiple $\underline{t}$'s. Table 5.7 shows that the two dimensional K-S test $\mathcal{MKS}$ has an unacceptably higher Type I error rates than $\mathcal{MSD}_n$, $\mathcal{M}\chi^2_{cl1}$ and $\mathcal{M}\chi^2_{cl2}$. There are missing Type I error rates in Table 5.7 since $\mathcal{M}\chi^2_{cl1}$ and $\mathcal{M}\chi^2_{cl2}$ are only constructed to reflect the cases of $\underline{p}$ = (0.8, 0.7, 0.9) and $\underline{p}$ = (0.50, 0.55, 0.45), respectively.

Next, we calculate the powers of the 36 $\mathcal{MZ}(\underline{t})$ tests based on 10,000 simulation replications of size n for $\underline{p}$, where $n = 100$ or 500 and $\underline{p}$ = (0.8, 0.7, 0.9) or $\underline{p}$ = (0.50, 0.55, 0.45). Figure 5.10 and Figure 5.11 show the result for $\underline{p}$ = (0.8, 0.7, 0.9) and $\underline{p}$ = (0.50, 0.55, 0.45), respectively.

In Figure 5.10, we can see that $\mathcal{MZ}(\underline{t_5})$ and $\mathcal{MZ}(\underline{t_{25}})$ perform the best overall when $n = 100$ and also they are among the top performers when $n = 500$ while controlling Type I error well as shown in Figure 5.8 for $\underline{p}$ = (0.8, 0.7, 0.9). Therefore, we recommend $\mathcal{MZ}(\underline{t_5})$ and $\mathcal{MZ}(\underline{t_{25}})$ among the 36 $\mathcal{MZ}(\underline{t})$ tests for the goodness-of-fit tests of BGD(B&D) with (0.8, 0.7, 0.9). On the other hand, we do not recommend $\mathcal{MZ}(\underline{t_6})$, $\mathcal{MZ}(\underline{t_{11}})$, $\mathcal{MZ}(\underline{t_{17}})$ and $\mathcal{MZ}(\underline{t_{26}})$ since they provide lower powers regardless of the sample size. Furthermore, we notice that none of the $\underline{t}$ elements of all the above investigated tests are close to zero. The tests $\mathcal{MZ}(\underline{t_9})$, $\mathcal{MZ}(\underline{t_{19}})$, $\mathcal{MZ}(\underline{t_{20}})$ and $\mathcal{MZ}(\underline{t_{21}})$, whose $\underline{t}$ are in the neighborhood of zero, have similar performance to each other and are the second best performers while controlling Type I error well regardless of n as shown in Figure 5.8 for $\underline{p}$ = (0.8, 0.7, 0.9).

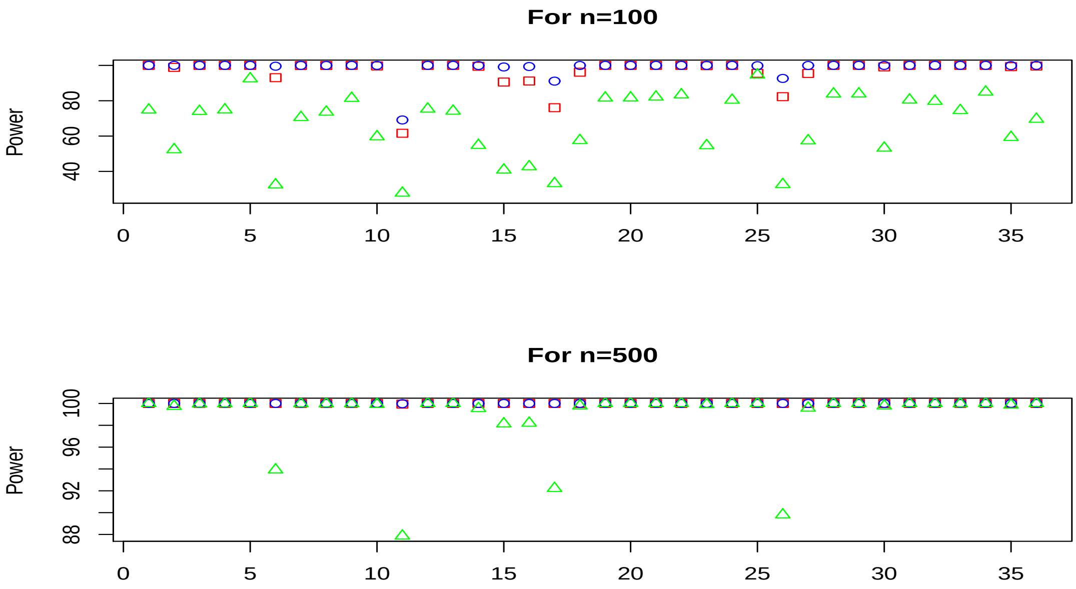

Figure 5.10 Power (%) of $\mathcal{MZ}(\underline{t})$ at $\alpha = 0.05$ for $\underline{p} = (0.8, 0.7, 0.9)$ with the subscript of $\underline{t}$ on the x-axis identifying the test where $\square$: BVP_1, $\bigcirc$: BVB_1, $\triangle$: $BVNB_1$.

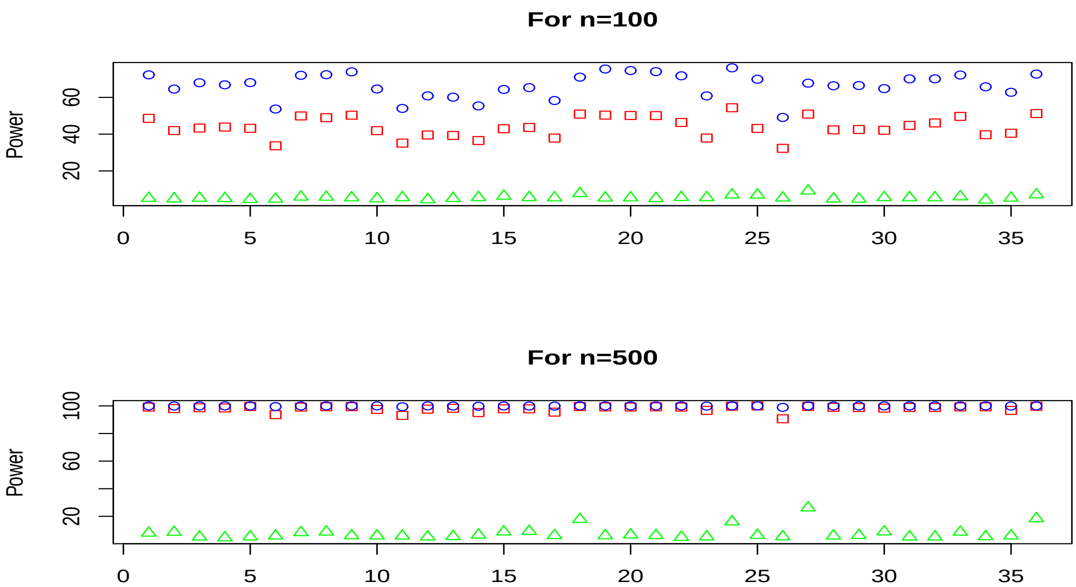

Figure 5.11 Power (%) of $\mathcal{MZ}(\underline{t})$ at $\alpha = 0.05$ for $\underline{p} = (0.50, 0.55, 0.45)$ with the subscript of $\underline{t}$ on the x-axis identifying the test where $\square$: BVP_2, $\bigcirc$: BVB_2, $\triangle$: $BVNB_2$.

In the Figure 5.11, we can see that all the tests perform almost equally well. However $\mathcal{MZ}(\underline{t_{24}})$ and $\mathcal{MZ}(\underline{t_{36}})$ perform the best overall and also control Type I error well as shown in Figure 5.8 for $\underset{\sim}{p}$ = (0.50, 0.55, 0.45) while $\mathcal{MZ}(\underline{t_6})$, $\mathcal{MZ}(\underline{t_{11}})$, $\mathcal{MZ}(\underline{t_{14}})$, $\mathcal{MZ}(\underline{t_{17}})$ and $\mathcal{MZ}(\underline{t_{26}})$ perform the worst for both $n = 100$ and $n = 500$. Moreover, The $\underset{\sim}{t}$ elements of all these tests do not include any value in the neighborhood of zero, which is the same finding as in the case of $\underset{\sim}{p} = (0.8, 0.7, 0.9)$. Consider the tests $\mathcal{MZ}(\underline{t_9})$, $\mathcal{MZ}(\underline{t_{19}})$, $\mathcal{MZ}(\underline{t_{20}})$ and $\mathcal{MZ}(\underline{t_{21}})$ with $\underset{\sim}{t}$ around zero. The test $\mathcal{MZ}(\underline{t_{19}})$ are among the second best tests overall while controlling Type I error well for both of n, Although $\mathcal{MZ}(\underline{t_9})$, $\mathcal{MZ}(\underline{t_{20}})$ and $\mathcal{MZ}(\underline{t_{21}})$ have similar performances as $\mathcal{MZ}(\underline{t_{19}})$ in terms of power, they control Type I error poorly when $n = 500$ with the rates 4.38%, 5.72% and 5.60%, respectively, which are out of the range (4.5%, 5.5%).

Based on the analysis of Figure 5.10 and Figure 5.11 we can see that $\mathcal{MZ}(\underline{t_6})$, $\mathcal{MZ}(\underline{t_{11}})$, $\mathcal{MZ}(\underline{t_{17}})$ and $\mathcal{MZ}(\underline{t_{26}})$ perform the worst in both cases of $\underset{\sim}{p}$ while there is not one $\mathcal{MZ}(\underset{\sim}{t})$ test that performs the best overall. However, we recommend $\mathcal{MZ}(\underline{t_{19}})$ where $(\underline{t_{19}})$ = (0.01, 0.01) because it performs comparatively well with the most powerful tests for both cases of $\underset{\sim}{p}$ while controlling Type I error well regardless of sample size.

Similarly, we calculate the powers of the $\mathcal{MT}_q(\mathbf{t})$ tests based on 10,000 simulation replications of size n for $\underset{\sim}{p}$ where $n = 100$ or $n = 500$ and $\underset{\sim}{p}$ = (0.8, 0.7, 0.9) or $\underset{\sim}{p}$ = (0.50, 0.55, 0.45). Figure 5.12 and Figure 5.13 show the results.

In Figure 5.12, $\mathcal{MT}_3(\mathbf{t}_2)$ performs the best overall in terms of power, while in Figure 5.9 for $\underset{\sim}{p}$ = (0.8, 0.7, 0.9) we can see that it controls Type I error well regardless of the sample size. All the tests have greater power, equal to or close to 100%, as n increases, except for $\mathcal{MT}_{10}(\mathbf{t}_2)$. Therefore, we do not recommend $\mathcal{MT}_{10}(\mathbf{t}_2)$ since it perform the worst regardless of the value of n. Tests $\mathcal{MT}_{10}(\mathbf{t}_1)$, $\mathcal{MT}_{10}(\mathbf{t}_3)$ and $\mathcal{MT}_3(\mathbf{t}_3)$ have almost the same performances as $\mathcal{MT}_3(\mathbf{t}_2)$. However, $\mathcal{MT}_{10}(\mathbf{t}_1)$ controls Type I error poorly. Thus we conclude $\mathcal{MT}_{10}(\mathbf{t}_3)$ and $\mathcal{MT}_3(\mathbf{t}_3)$ as the second best tests. In Figure 5.13, we observe that $\mathcal{MT}_5(\mathbf{t}_2)$ has the greatest powers overall for both $n = 100$ and $n = 500$, however, it has intolerably high Type I error rates as shown in Figure 5.9 for $\underset{\sim}{p}$ = (0.50, 0.55, 0.45). Meanwhile we can see that $\mathcal{MT}_3(\mathbf{t}_3)$ performs second to the best in terms of power while

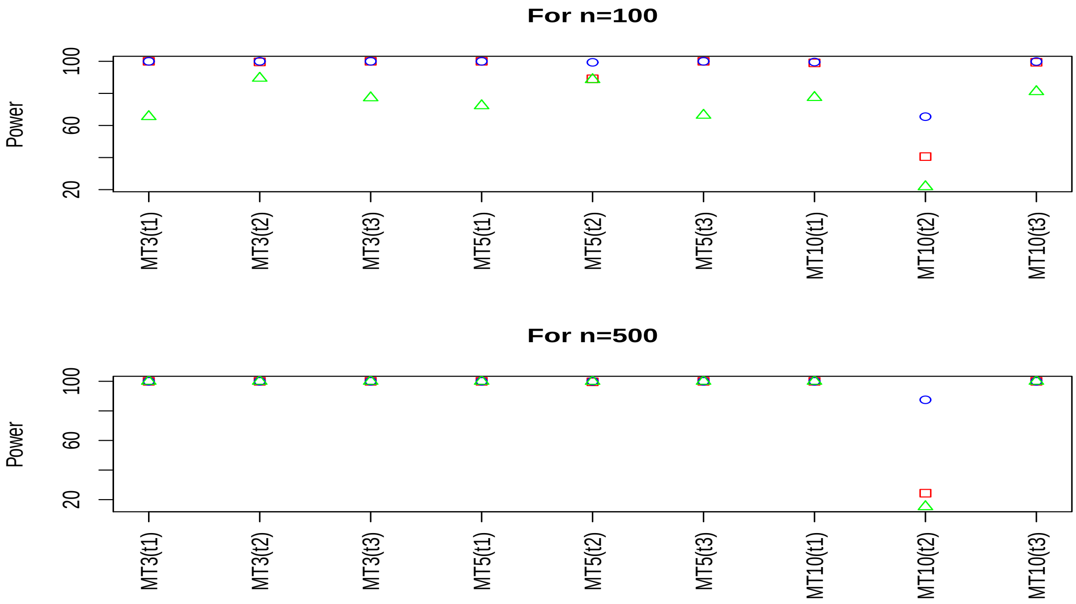

Figure 5.12 Power (%) of $\mathcal{MT}_q(\mathbf{t})$ at $\alpha = 0.05$ for $\underline{p} = (0.8,\ 0.7,\ 0.9)$ where □: BVP_1, ◯: BVB_1, △: $BVNB_1$.

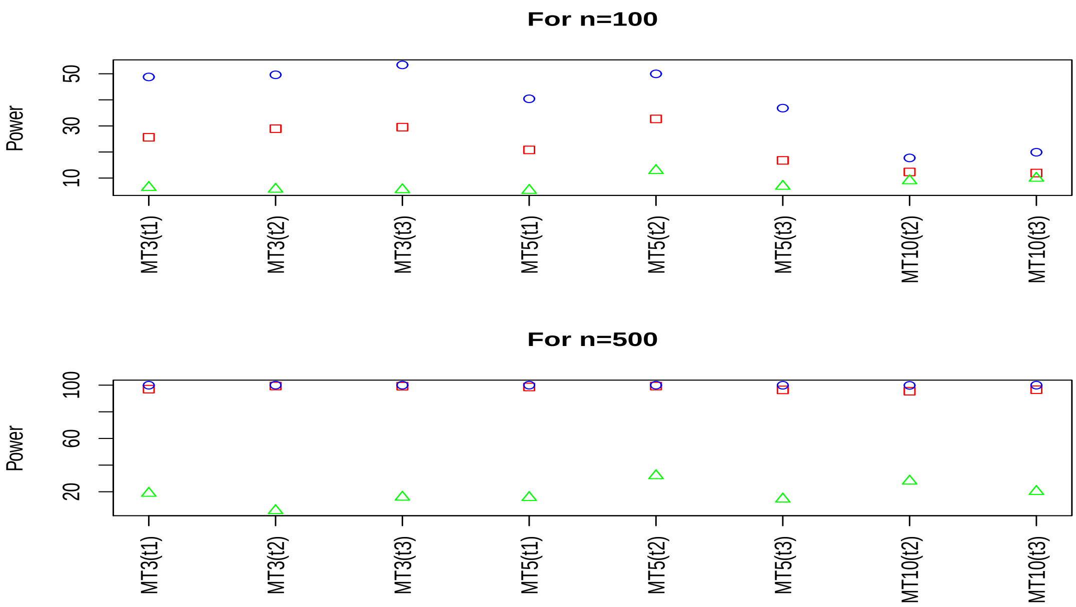

Figure 5.13 Power (%) of $\mathcal{MT}_q(\mathbf{t})$ at $\alpha = 0.05$ for $\underline{p} = (0.50,\ 0.55,\ 0.45)$ where □: BVP_2, ◯: BVB_2, △: $BVNB_2$.

they control Type I error well for both $n = 100$ and $n = 500$ as displayed in Figure 5.9.

Based on the analysis of Figure 5.12 and Figure 5.13 we recommend $\mathcal{MT}_3(\mathbf{t}_3)$ where $\mathbf{t}_3$ = [(-0.7, -0.2),(-0.9, 0.1),(0.07, 0.6)] for both cases of $\underline{p}$.

Finally, we calculate the powers of $\mathcal{MSD}_n$, $\mathcal{MKS}$, $\mathcal{M}\chi^2_{cl1}$ and $\mathcal{M}\chi^2_{cl2}$ based on 10,000 simulation replications of size n for $\underline{p}$ where n = 100 or n = 500 and $\underline{p}$ = (0.8, 0.7, 0.9) or $\underline{p}$ = (0.50, 0.55, 0.45), and compare them with those of the top performers among the $\mathcal{MZ}(\underline{t})$ and $\mathcal{MT}_q(\mathbf{t})$ tests considered here. Table 5.8 and Table 5.9 show the powers for $\underline{p}$ = (0.8, 0.7, 0.9) and $\underline{p}$ = (0.50, 0.55, 0.45), respectively.

Table 5.8 Powers(%) of Tests for BGD(B&D) with $\underline{p}$ = (0.8, 0.7, 0.9) at α = 0.05 based on 10,000 Replications

		BVP_1	BVB_1	$BVNB_1$
	$\mathcal{MZ}(\underline{t_5})$	100	100	92.44
	$\mathcal{MZ}(\underline{t_{19}})$	100	100	81.57
	$\mathcal{MZ}(\underline{t_{25}})$	95.50	99.82	94.62
n=100	$\mathcal{MT}_3(\mathbf{t}_2)$	99.62	100	89.44
	$\mathcal{MT}_3(\mathbf{t}_3)$	100	100	77.24
	$\mathcal{MSD}_n$	100	100	74.45
	$\mathcal{M}\chi^2_{cl1}$	100	100	49.38
	$\mathcal{MKS}$	100	100	95.92
	$\mathcal{MZ}(\underline{t_5})$	100	100	100
	$\mathcal{MZ}(\underline{t_{19}})$	100	100	100
	$\mathcal{MZ}(\underline{t_{25}})$	100	100	100
n=500	$\mathcal{MT}_3(\mathbf{t}_2)$	100	100	100
	$\mathcal{MT}_3(\mathbf{t}_3)$	100	100	100
	$\mathcal{MSD}_n$	100	100	100
	$\mathcal{M}\chi^2_{cl1}$	100	100	99.73
	$\mathcal{MKS}$	100	100	100

In Table 5.8, $\mathcal{MKS}$ has the best performance in terms of power. However, it control Type I error the worst among all the tests listed in the table. All of its Type I error rates as shown in Table 5.7 are greater than 5% by more than 0.76%. Test $\mathcal{MZ}(\underline{t_5})$ performs comparatively well to $\mathcal{MKS}$ in terms of power while it control Type I error well for both cases of n. In Table 5.9, $\mathcal{MZ}(\underline{t_{24}})$ and $\mathcal{MZ}(\underline{t_{36}})$ perform the best overall, and $\mathcal{MSD}_n$ and $\mathcal{MZ}(\underline{t_{19}})$ are the second best performers while all of these four tests control Type I error well for both cases of n.

Table 5.9 Powers(%) of Tests for BGD(B&D) with $\underline{p} = (0.50, 0.55, 0.45)$ at $\alpha = 0.05$ based on 10,000 Replications

	Alternative	BVP_2	BVB_2	$BVNB_2$
	$\mathcal{MZ}(\underline{t_{19}})$	50.36	75.44	5.30
	$\mathcal{MZ}(\underline{t_{24}})$	54.38	76.10	6.96
	$\mathcal{MZ}(\underline{t_{36}})$	51.20	72.70	7.04
n=100	$\mathcal{MT}_3(\mathbf{t}_3)$	29.52	53.40	5.52
	$\mathcal{MSD}_n$	47.85	71.85	6.10
	$\mathcal{MKS}$	36.01	58.76	6.58
	$\mathcal{M}\chi^2_{cl2}$	34.46	60.26	6.50
	$\mathcal{MZ}(\underline{t_{19}})$	99.28	100	5.98
	$\mathcal{MZ}(\underline{t_{24}})$	99.76	100	16.04
	$\mathcal{MZ}(\underline{t_{36}})$	99.72	100	18.34
n=500	$\mathcal{MT}_3(\mathbf{t}_3)$	99.22	99.98	15.90
	$\mathcal{MSD}_n$	99.43	100	10.11
	$\mathcal{MKS}$	95.54	99.92	13.98
	$\mathcal{M}\chi^2_{cl2}$	98.58	100	9.80

In both Table 5.8 and Table 5.9 one observes that the power increases as n increases. Chi-square tests perform poorly overall maybe because they are sensitive to the choice of categories. The K-S test $\mathcal{MKS}$ controls Type I error the worst (see Table 5.7) and can not beat our proposed tests in Table 5.8 and Table 5.9. The $\mathcal{MZ}(\underline{t})$ type tests generally perform better than the $\mathcal{MT}_q(\mathbf{t})$ type tests. The $\mathcal{MZ}(\underline{t})$ tests with $\underline{t}$ elements far from zero perform the best for the goodness-of-fit tests of BGD(B&D). However, the value of $\underline{t}$ that provides the best performance varies according to $\underline{p}$. Furthermore, $\mathcal{MZ}(\underline{t_{19}})$, the recommended $\mathcal{MZ}(\underline{t})$ type test, performs relatively well expect for alternative $BVNB_2$ when $n = 500$. Therefore, overall we recommend the supremum test $\mathcal{MSD}_n$ because it has robust performance in terms that it controls Type I error well and has comparative powers with respect to the top performers. In addition, there is no need to select the value of $\underline{t}$.

5.3 Robustness Study

In order to examine the robustness of our tests, we conduct one more simulation study with $\underline{p} = (0.64, 0.56, 0.86)$ and sample size $n = 122$. We compare the 36 $\mathcal{MZ}(\underline{t})$ (see Table 5.1), the nine $\mathcal{MT}_q(\mathbf{t})$ (see Table 5.2), $\mathcal{MSD}_n$ and $\mathcal{MKS}$ tests in terms of power and empirical Type I error rate at $\alpha = 0.05$. Here Chi-square test is not included in the comparison study because it varies according to the grouping of data and shows consistently weak performance as shown in the previous study.

First of all, by parametric bootstrap we calculate the empirical critical points of $\mathcal{MSD}_n$ and $\mathcal{MKS}$ at $\alpha = 0.05$ with $\underline{p} = (0.64, 0.56, 0.86)$ and $n = 122$ bases on 10,000 bootstrap samples, and they are 2.4674 and 0.0894, respectively. These two critical points will be used to compute the power and empirical α of $\mathcal{MSD}_n$ and $\mathcal{MKS}$.

Next, the empirical α at the significance level of 0.05 are computed for each of the tests in the study based on 10,000 replications. Figure 5.14 shows the results of $\mathcal{MZ}(\underline{t})$ and $\mathcal{MT}_q(\mathbf{t})$ tests. Here the empirical significance level of $\mathcal{MSD}_n$ and $\mathcal{MKS}$ at $\alpha = 5\%$ are 4.76% and 4.86%, respectively, which mean they control Type I error well.

Finally, the power of all the competing tests are calculated based on 10,000 replications. The following three alternative distributions are chosen for the power comparison: BVP_3, BVB_3 and $BVNB_3$ (see Table 5.4), which are designed to have the same first moments as the null distribution BGD(B&D) with (0.64, 0.56, 0.86). The power of $\mathcal{MZ}(\underline{t})$ and $\mathcal{MT}_q(\mathbf{t})$

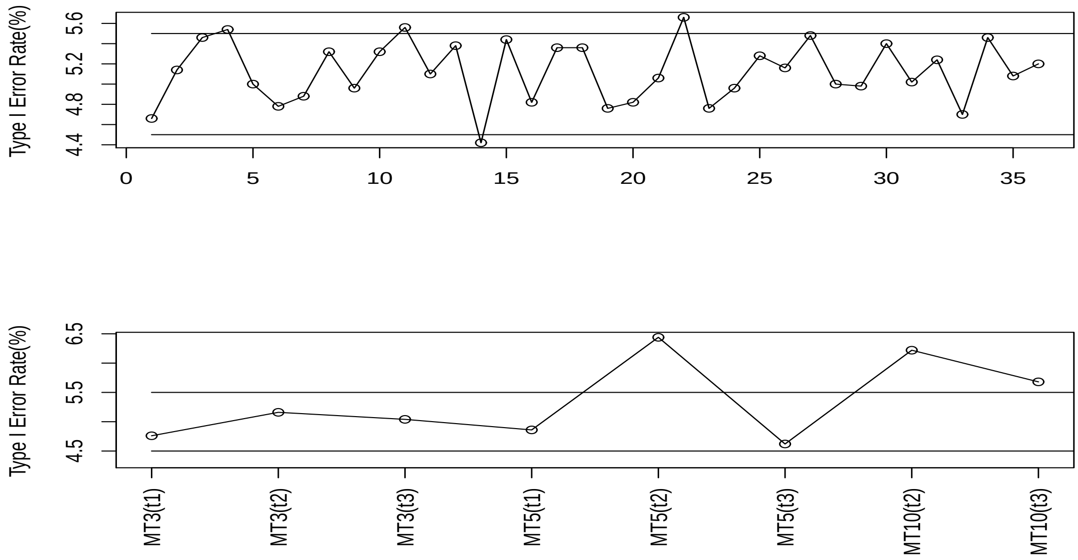

Figure 5.14 Type I error rates (%) of $\mathcal{MZ}(\underline{t})$ and $\mathcal{MT}_q(\mathbf{t})$ at $\alpha = 5\%$ for $\underline{p} = (0.64, 0.56, 0.86)$ and $n = 122$. In the first figure, the subscript of $\underline{t}$ on the x-axis identifies the test.

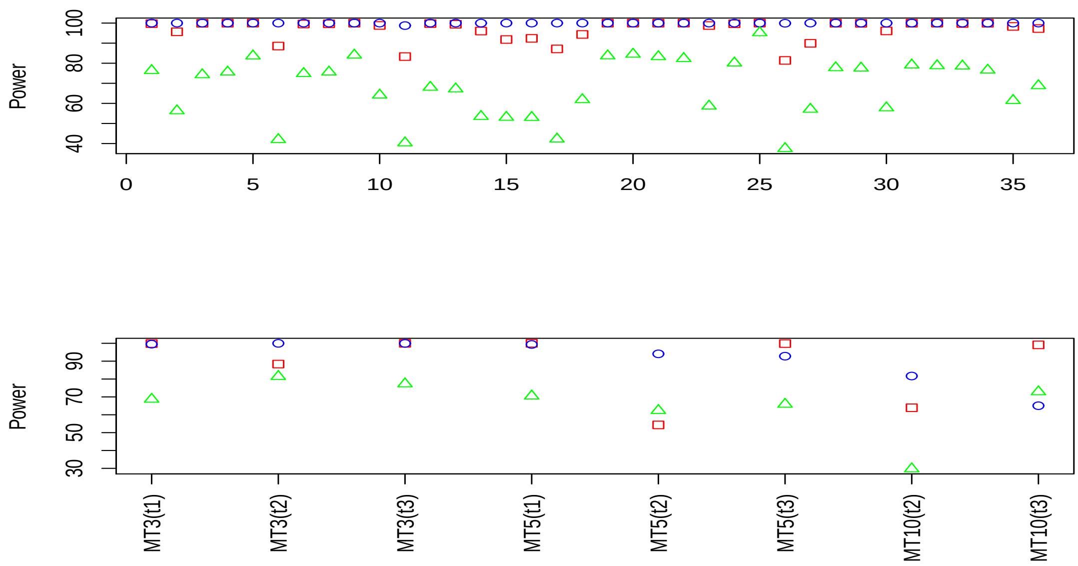

Figure 5.15 Power (%) of $\mathcal{MZ}(\underline{t})$ and $\mathcal{MT}_q(\mathbf{t})$ at $\alpha = 0.05$ for $\underline{p} = (0.64, 0.56, 0.86)$ and $n = 122$ where □: BVP_3, ◯: BVB_3, △: $BVNB_3$. In the first figure, the subscript of $\underline{t}$ on the x-axis identifies the test.

tests are displayed in Figure 5.15. $\mathcal{MT}_{10}(\mathbf{t}_1)$ is removed from comparison because of the singularity of the covariance matrix $\hat{\Gamma}$.

In Figure 5.14 and Figure 5.15, among all the 36 $\mathcal{MZ}(\underline{t})$ tests we can see $\mathcal{MZ}(\underline{t_{25}})$ performs the best and $\mathcal{MZ}(\underline{t_9})$, $\mathcal{MZ}(\underline{t_{19}})$, $\mathcal{MZ}(\underline{t_{20}})$, $\mathcal{MZ}(\underline{t_{21}})$ with $\underline{t}$ in the neighborhood of zero and $\mathcal{MZ}(\underline{t_5})$ are the second best performers. All of the above-mentioned tests control Type I error well with rates in (4.5%, 5.5%). Meanwhile, we can find that $\mathcal{MZ}(\underline{t_6})$, $\mathcal{MZ}(\underline{t_{11}})$, $\mathcal{MZ}(\underline{t_{17}})$ and $\mathcal{MZ}(\underline{t_{26}})$ perform the worst among all the 36 $\mathcal{MZ}(\underline{t})$ tests. This result verifies the previous findings regarding the $\mathcal{MZ}(\underline{t})$ tests: (1) The $\underline{t}$ values of the best tests vary according to $\underline{p}$ and the $\underline{t}$ elements of the best tests are not close to zero; (2) The common worst tests are $\mathcal{MZ}(\underline{t_6})$, $\mathcal{MZ}(\underline{t_{11}})$, $\mathcal{MZ}(\underline{t_{17}})$ and $\mathcal{MZ}(\underline{t_{26}})$; (3) Test $\mathcal{MZ}(\underline{t_{19}})$, the recommended test among the 36 $\mathcal{MZ}(\underline{t})$ tests in Section 5.2, also performs comparatively well in this robustness study.

Furthermore, in Figure 5.14 and Figure 5.15, we observe that $\mathcal{MT}_3(\mathbf{t}_3)$ performs the best among the eight $\mathcal{MT}_q(\mathbf{t})$ tests while controlling Type I error well. We recommend $\mathcal{MT}_3(\mathbf{t}_3)$ among all of the $\mathcal{MT}_q(\mathbf{t})$ tests, which is consistent with the results for both of cases $\underline{p} = (0.8, 0.7, 0.9)$ and $\underline{p} = (0.50, 0.55, 0.45)$ in Section 5.2.

Table 5.10 Powers (%) of Tests for BGD(B&D) with $\underline{p} = (0.64, 0.56, 0.86)$ at $\alpha = 0.05$

	BVP_3	BVB_3	$BVNB_3$
$\mathcal{MZ}(\underline{t_{19}})$	99.98	100	83.66
$\mathcal{MZ}(\underline{t_{25}})$	100	100	95.17
$\mathcal{MT}_3(\mathbf{t}_3)$	99.98	99.95	77.28
$\mathcal{MSD}_n$	99.94	100	76.62
$\mathcal{MKS}$	99.76	100	66.98

In summary, the powers of $\mathcal{MZ}(\underline{t_{19}})$, $\mathcal{MZ}(\underline{t_{25}})$, $\mathcal{MT}_3(\mathbf{t}_3)$, $\mathcal{MSD}_n$ and $\mathcal{MKS}$ are displayed in Table 5.10. From this table we can see that (1) $\mathcal{MZ}(\underline{t})$ tests generally perform better than $\mathcal{MT}_q(\mathbf{t})$ tests; (2) The $\mathcal{MKS}$ test has the lowest powers among the tests in Table 5.10; (3) The supremum test $\mathcal{MSD}_n$ has comparative powers while controlling Type

I error well so far. Therefore, we recommend $\mathcal{MSD}_n$ when the alternative distributions are unknown.

5.4 Remarks

Here are some remarks on the simulation and computation using R.

1. Each estimated critical point is based on 10,000 bootstrap samples. Each computed empirical significance level or power is based on 10,000 Monte Carlo replications. In the univariate case, three cases of p are considered. 13 different tests are compared for each of case on p and there are totally 22 alternative distributions chosen. In the bivariate case, we consider two sample size $n = 100$ or 500 and three cases of $\underline{p}$. For each combination of n and $\underline{p}$ there are 47 or 48 different tests and three alternative distributions. Therefore, the simulations were extremely computer intensive. All the simulations and computations are implemented in R version 2.14.1.

2. We implement the following two steps to generate BGD(B&D) sample: (1) Generate three independent geometric samples U, V, W with the same sample size n and the probability of success $1 - p_1$, $1 - p_2$, $1 - p_3$, respectively. (2) Let $X = Min(U, V)$ and $Y = Min(V, W)$. Then (X,Y) follows the BGD(B&D) with parameter $\underline{p} = (p_1, p_2, p_3)$.

3. Since there are no closed form for the MLE $\hat{p}$ in the BGD(B&D), we use the *dfsane* function in package *BB* of R to calculate $\underline{\hat{p}}$ numerically, and add option $NM = TRUE$ to get a better estimator of $\underline{p}$ which also speeds up the running time and provides better convergence rate. We apply *hessian* function in *numDeriv* package of R to get the Hessian matrix for estimating n Fisher information matrix, and use *optim* function of R with method option "$L - BFGS - B$" to find the supremum of the functions.

4. We do not use the simulation replications if *dfsane* function does not converge, *optim* function fails, *hessian* function can not find the matrix, or the determinant of the matrix is zero when calculating the critical points, empirical α or powers of the test statistics.

5.5 Real Example Analysis

For illustration, our proposed tests for BGD(B&D) are applied to a real data from Arbous and Kerrich (1951) that consists of the accident records of 122 experienced shunters on the South African Railways over the eleven-year periods from 1937 to 1947. The data is split into two periods: 1937-1942 and 1943-1947. Let x denote the number of accidents in the first period and y denote the number of accidents in the second period. Table 5.11 shows the data.

Table 5.11 Accidents Records of 122 Shunters during 1937-1947

	y								
x	0	1	2	3	4	5	6	7	Total
0	21	13	4	2	0	0	0	0	40
1	18	14	5	1	0	0	0	1	39
2	8	10	4	3	1	0	0	0	26
3	2	1	2	2	1	0	0	0	8
4	1	4	1	0	0	0	0	0	6
5	0	1	0	1	0	0	0	0	2
6	0	0	1	0	0	0	0	0	1
Total	50	43	17	9	2	0	0	1	122

Arbous and Kerrich (1951) found that BVNB distribution has a satisfactory fit to the data. Now we consider a test that the data agree to BGD(B&D) against BVNB distribution. The MLE of $\underline{p}$, $\hat{\underline{p}} = (0.64, 0.56, 0.86)$ is calculated from the data. The tests used here include $\mathcal{MZ}(\underline{t_{25}})$ (the best test for the this case of $\underline{p}$), $\mathcal{MZ}(\underline{t_{19}})$ (the recommended $\mathcal{MZ}(\underline{t})$ type test), $\mathcal{MSD}_n$ and $\mathcal{MKS}$. We calculate the values of test statistics and the corresponding p-values. The bootstrap p-values of $\mathcal{MSD}_n$ and $\mathcal{MKS}$ are calculated similarly as described in Section 4.3. The results are displayed in Table 5.12.

Table 5.12 Statistic Values and P-values for Accident Records Example

	Test Statistic	P-value
$\mathcal{MZ}(\underline{t_{19}})$	-3.80415	1.42E-04
$\mathcal{MZ}(\underline{t_{25}})$	-5.39301	6.93E-08
$\mathcal{MSD}_n$	3.83634	7.65E-04
$\mathcal{MKS}$	0.12780	1.25E-03

In Table 5.12, all the p-values are less than 0.05. Therefore, we reject the null hypothesis that the data arise from BGD(B&D). However, $\mathcal{MZ}(\underline{t_{25}})$ has the smallest p-value and $\mathcal{MKS}$ has the largest one, which is consistent with the result in Table 5.10 that $\mathcal{MZ}(\underline{t_{25}})$ is the most powerful test while $\mathcal{MKS}$ is the least powerful one.

CHAPTER 6

CONCLUSIONS

Our research focuses on the goodness-of-fit tests of geometric models based on the PGF and its empirical counterpart against the alternative distributions from different families. We have applied PGF based K&K methods to realize two types of test statistics for testing geometric distribution and BGD(B&D). In addition, and more importantly, we propose new test statistics which are the supremum of the absolute value of standardized difference between the estimator of PGF with parameters replaced by their MLE's and its empirical counterpart for testing the above-mentioned distributions. We compare the empirical critical points of the K & K method-based test statistics with their corresponding theoretical ones, and use Monte Carlo simulation to explore the limiting distributions of the supremum test statistics. Moreover, we compare the performances of the K&K method-based tests and the supremum test with competing tests in the literature, such as Chi-square test and EDF based tests. We find comparative and satisfactory results on the K&K method with single t and the supremum tests in terms of controlling Type I error rate and power for both geometric distribution and BGD(B&D) goodness-of-fit tests. However, we recommend the supremum tests, especially when the alternatives are unknown, because they are robust performers and have no requirements on selecting t, a step that affects the performances of K&K method-based tests.

APPENDIX

BGD(B&D)

Suppose X=(X_1,X_2) follows the BGD(B&D). The joint probability mass function of (X_1,X_2) is given by

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2) = &\, p_2^{x_2-1}(p_1p_3)^{x_1-1}q_2(1-p_1p_3)\mathbf{1}_{\{x_1>x_2\}} \\ &+ p_1^{x_1-1}(p_2p_3)^{x_2-1}q_1(1-p_2p_3)\mathbf{1}_{\{x_1<x_2\}} \\ &+ (p_1p_2p_3)^{x_1-1}(1-p_2p_3-p_1p_3+p_1p_2p_3)\mathbf{1}_{\{x_1=x_2\}}, \end{aligned} \tag{A.1}$$

for $0 < p_1,\ p_2 < 1,\ 0 < p_3 \leq 1,\ q_1 = 1-p_1,\ q_2 = 1-p_2$, and $1 - p_1p_3 - p_2p_3 + p_1p_2p_3 > 0$, where $x_1, x_2 = 1, 2, ...$ and $\mathbf{1}_A$ is indicator function, which is 1 if $x_1, x_2 \in A$ and 0 otherwise. The PGF evaluated at $\underline{t} = (t_1, t_2)$ is

$$\begin{aligned} &G(\underline{t};\underline{p}) = G(t_1, t_2; p_1, p_2, p_3) = A_1A_2, \\ &\text{where} \quad A_1 = \frac{t_1t_2}{1-t_1t_2p_1p_2p_3} \quad \text{and} \\ &A_2 = \frac{t_2q_1(1-p_2p_3)p_2p_3}{1-t_2p_2p_3} + \frac{t_1q_2p_1p_3(1-p_1p_3)}{1-t_1p_1p_3} + 1 - p_1p_3 - p_2p_3 + p_1p_2p_3. \end{aligned} \tag{A.2}$$

Through some algebra, we get the first derivative of PGF with respective to p_1, p_2, p_3.

$$\begin{aligned} &\frac{\partial G(\underline{t};\underline{p})}{\partial p_1} = A_1\frac{\partial A_2}{\partial p_1} + A_2\frac{\partial A_1}{\partial p_1}, \\ &\quad \text{where} \quad \frac{\partial A_1}{\partial p_1} = \frac{t_1^2t_2^2p_2p_3}{(1-t_1t_2p_1p_2p_3)^2} \\ &\quad \text{and} \quad \frac{\partial A_2}{\partial p_1} = p_3[\frac{t_2p_2(p_2p_3-1)}{1-t_2p_2p_3} + \frac{(1-p_2)(t_1-1)}{(1-t_1p_1p_3)^2}] \end{aligned} \tag{A.3}$$

$$\begin{aligned} &\frac{\partial G(\underline{t};\underline{p})}{\partial p_2} = A_1\frac{\partial A_2}{\partial p_2} + A_2\frac{\partial A_1}{\partial p_2}, \\ &\quad \text{where} \quad \frac{\partial A_1}{\partial p_2} = \frac{t_1^2t_2^2p_1p_3}{(1-t_1t_2p_1p_2p_3)^2} \\ &\quad \text{and} \quad \frac{\partial A_2}{\partial p_2} = p_3[\frac{t_1p_1(p_1p_3-1)}{1-t_1p_1p_3} + \frac{(1-p_1)(t_2-1)}{(1-t_2p_2p_3)^2}]. \end{aligned} \tag{A.4}$$

$$\frac{\partial G(\underline{t};\underline{p})}{\partial p_3} = A_1 \frac{\partial A_2}{\partial p_3} + A_2 \frac{\partial A_1}{\partial p_3},$$

where $\dfrac{\partial A_1}{\partial p_3} = \dfrac{t_1^2 t_2^2 p_1 p_2}{(1 - t_1 t_2 p_1 p_2 p_3)^2}$

and $\dfrac{\partial A_2}{\partial p_3} = \dfrac{q_1 t_2 p_2 (1 - 2p_2 p_3 + t_2 p_2^2 p_3^2)}{(1 - t_2 p_2 p_3)^2} + \dfrac{q_2 t_1 p_1 (1 - 2p_2 p_3 + t_1 p_1^2 p_3^2)}{(1 - t_1 p_1 p_3)^2} - p_1 - p_2 + p_1 p_2.$ (A.5)

There is no closed form for $\underline{\hat{p}} = (\hat{p_1}, \hat{p_2}, \hat{p_3})$. We solve the score equations (Li and Dhar (2013)) numerically to obtain the maximum likelihood estimators of the parameters. Another approach is using *optm* function in *R* to get the MLE.

REFERENCES

[1] A. G. Arbous and Kerrich. J. E. Accident Statistics and the Concept of Accident-Proneness. *Biometrics*, 7:340–432, 1951.

[2] L. Baringhaus and N. Henze. A goodness of fit test for Poisson distribution based on the empirical probability generating function. *Statistics and Probability letters*, 13:269–274, 1992.

[3] A. P. Basu and S. K. Dhar. Bivariate Geometric Distribution. *Journal of Applied Statistical Science*, 2:33–44, 1995.

[4] D. J. Best and J. C. W. Rayner. Goodness of fit for geometric distribution. *Biometrical journal*, 31:307–311, 1989.

[5] D. J. Best and J. C. W. Rayner. Tests of Fit for the Geometric Distribution. *Communications in Statistics:Simulation and Computation*, 32:1065–1078, 2003.

[6] D. Böhning. A note on a test for Poisson overdispersion. *Biometrika*, 81:418–419, 1994.

[7] H. Chernoff and E.L. Lehmann. The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *The Annals of Mathematical Statistics*, 25:579–586, 1954.

[8] P. L. Conti. An asymptotic test for a geometric process against a lattice distribution with monotone hazard. *Journal of the Italian Statistical Society*, 6(3):213–231, 1997.

[9] P. L. Conti and L. D. Giovanni. Queueing models and statistical analysis for ATM based networks. *The Indian Journal of Statistics*, 64:50–57, 2002.

[10] S. Csörgő. Testing for Exponential and Marshall-Olkin Distribution. *Journal of Statistical Planning and Inference*, 23:287–300, 1989.

[11] T. W. Epps. A Test of Fit for Lattice Distribution. *Communications in Statistics:theory and methods*, 24(6):1455–1479, 1995.

[12] T. W. Epps and L. B. Pulley. A test for normality based on the empirical characteristic function. *Biometrika*, 70:732–736, 1983.

[13] T. W. Epps, K. J. Singleton, and L. B. Pulley. A test of separate families of distributions based on the empirical moment generating function. *Biometrika*, 69:391–399, 1982.

[14] G Fasano and A. Franceschini. A multidimensional of the Kolmogorov-Smirnov test. *Monthly Notices Royal Astronomy Society*, 225:155170, 1987.

[15] J. Gart and H. Pettigrew. On the conditional moments of the K-statistics for the Poisson distribution. *Biometrika*, 57:661–664, 1970.

[16] S. Gulati and J. Neus. Goodness of Fit Statistics for the Exponential Distribution When the Data Are Grouped. *COMMUNICATIONS IN STATISTICS: Theory and Methods*, 32:681700, 2003.

[17] N. Gürtler and N. Henze. Recent and classical goodness-of-fit tests for the Poisson distribution. *Journal of Statistical Planning and Inference*, 90:207–225, 2000.

[18] N. Henze. Empirical-distribution-function goodness-of-fit tests for discrete models. *Canadian Journal of Statistics*, 24:81–93, 1996.

[19] A. G. Kallioras, I. A. Koutrouvelis, and G. C. Cavavos. Testing the Fit of Gamma Distribution Using the Empirical Moment Generating Function. *Communications in Statistics:theory and methods*, 35:527–540, 2006.

[20] D. Karlis and E. Xekalaki. A simulation comparison of several procedures for testing the Poisson assumption. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 49:355–382, 2000.

[21] A. C. Kimber. When is a χ^2 not a χ^2? *Teaching Statistics*, 9:74–77, 1987.

[22] B. Klar. Goodness-of-fit tests for discrete models based on the integrated distribution function. *Metrika*, 49:53–69, 1999.

[23] S. Kocherlakota and K. Kocherlakota. Goodness of fit tests for discrete distribution. *Communications in Statistics:theory and methods*, 15:815–829, 1986.

[24] S. Kocherlakota and K. Kocherlakota. *Bivariate Discrete Distributions.* Marcel Dekker Inc., New York, NY, 1992.

[25] J. Li and S. K. Dhar. Modeling with Bivariate Geometric Distribution. *Communications in Statistics:theory and methods*, Vol 42, 2013.

[26] A. W. Marshall and I. Olkin. A Multivariate Exponential Distribution. *Journal of the American Statistical Association*, 63:30–44, 1967.

[27] S. G. Meintanis. Goodness-of-fit Tests For The Logistic Distribution Based on Empirical Transforms. *The Indian Journal of Statistics*, 66:304–324, 2004.

[28] S. G. Meintanis. Testing Skew Normality via the Moment Generating Function. *Mathematical Methods of Statistics*, 19:64–72, 2010.

[29] M. Nakamura and V. Pérez-Abreu. Empirical Probability Generating Function An overview. *Mathematics and Economics*, 12:287–295, 1993a.

[30] M. Nakamura and V. Pérez-Abreu. Using of an Empirical Probability Generating Function for Testing a Poisson Model. *The Canadian Journal of Statistics*, 21:149–156, 1993b.

[31] J. A. Peacock. Two-dimensional goodness-of-fit testing in astronomy. *Monthly Notices Royal Astronomy Society*, 202:615627, 1983.

[32] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C.* Cambridge University Press, New York, NY, 2002.

[33] J. C. W. Rayner and D. J. Best. *Smooth Tests of Goodness of Fit.* Oxford University Press, New York, NY, 1989.

[34] R. Rueda, V. Pérez-Abreu, and F. O'Reilly. Goodness of Fit for the Poisson Distribution Based on Probability Generating Function. *Communications in Statistics:theory and methods*, 20(10), 1991.

[35] G. W. Snedecor and W. G. Cochran. *Statistical Methods.* Iowa State University Press, Ames, Iowa, 8th edition, 1989.

[36] W. Stute and M. Manteiga, W.and Quindimil. Bootstrap based Goodness-of-Fit tests. *Metrika*, 40, 1993.

[37] A. A. Zghoul. A Goodness of Fit Test for Normality Based on the Empirical Moment Generating Function. *Communications in Statistics:Simulation and Computation*, 39:1292–1304, 2010.