5-17-2013

Small Improvement to the Kolmogorov-Smirnov Test

Xing Dong
Georgia State University

SMALL IMPROVEMENT TO THE KOLMOGOROV-SMIRNOV TEST

by

XING DONG

Under the Direction of Dr. Yuanhui Xiao

ABSTRACT

The Kolmogorov-Smirnov (K-S) test is widely used as a goodness-of-fit test. This thesis consists of two parts to describe ways to improve the classical K-S test in both 1-dimensional and 2-dimensional data. The first part is about how to improve the accuracy of the classical K-S goodness-of-fit test in 1-dimensional data. We replace the *p*-values estimated by the asymptotic distribution with near-exact p-values. In the second part, we propose two new methods to increase power of the widely used 2-dimensional two-sample Fasano and Franceschini test. Simulation studies show the new methods are significantly more powerful than the Fasano and Franceschini's test.

INDEX WORDS: K-S test, Cramér-von Mises test, 2-Dimensional K-S test, Goodness-of-fit, Near-exact distribution of K-S statistics

SMALL IMPROVEMENT TO THE KOLMOGOROV-SMIRNOV TEST

by

XING DONG

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2013

SMALL IMPROVEMENT TO THE KOLMOGOROV-SMIRNOV TEST

by

XING DONG

Committee Chair: Dr. Yuanhui Xiao

Committee: Dr. Yu Xie

Dr. Yichuan Zhao

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2013

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

This thesis has two distinct sections. The first section focuses on how to improve the accuracy of the estimated distribution of the classical Kolmogorov-Smirnov (K-S) test statistics. The second section is dedicated to improve the power of the 2-dimensional two-sample K-S test with two proposed Crámer-von Mises (CVM) type test statistics. Throughout this paper, 1-dimensional and 2-dimensional data are abbreviated as 1D and 2D respectively.

1 Improving 1-Dimensional One-Sample K-S Test

1.1 Introduction

In practical research, there are needs to test hypothesis about the agreement between the underlying distribution of a sample and a hypothetical distribution. This type of test is frequently labeled as "goodness-of-fit" test, i.e. checking if data are normally distributed. Furthermore, there are two-sample tests about the hypothesis of the agreement of the underlying distributions of two samples. The K-S goodness-of-fit test was developed by the work from Kolmogorov (1933), Smirnov (1939a), Scheffe (1943), and Wolfowitz (1949), etc.

The one-sample K-S test can be defined as the follows. Given an independently identically distributed random sample $x_1,\ x_2, \dots ,x_N$ with unknown distribution *F*, to test if the distribution *F* is significantly different from a specified distribution F_0,

(2.1) $$H_0\colon F = F_0.$$

Kolgomorov (1933 and 1941) suggested a test of H_0 that is based on the test statistic

(2.2) $$Z_N = N^{\frac{1}{2}} D_N = N^{\frac{1}{2}} \cdot \sup_{-\infty<x<\infty} |F_N(x) - F(x)|.$$

Where $F_N(x)$ in (2.2) is defined as,

(2.3) $$F_N(x) = \frac{1}{N}\sum_{i=1}^{N} I(x_i \le x).$$

It is termed the empirical distribution function (EDF) based on the random sample $x_1,\ x_2, \dots ,x_N$. Obviously, $F_N(x)$ is the proportion of the sample points x_i such that $x_i \le x$. $|F_N(x) - F(x)|$ is actually the vertical difference between the hypothetical distribution function and the EDF. Therefore, in essence, the one-sample K-S test was based on the largest (vertical) difference,D_n, between these two distributions. Even though, the exact distribution of D_n is hard to track, Kolmogorov (1933) and Smirnov (1939b) proved that the limiting distribution of $Z_n$is (2.5),

(2.4) $$D_N = \sup_{-\infty<x<\infty} |F_N(x) - F(x)|,$$

(2.5) $$\lim \Pr\{Z_N < x|H_0\} = 1 + 2\sum_{\mathrm{j}=1}^{\infty}(-1)^j e^{-2j^2x^2}, 0 < x < \infty.$$

If Z_N is sufficiently large, H_0 will be rejected. Feller (1948) and Doob (1949) rederived (2.4) with a simpler and more general approach. The short table of this limiting distribution was first given by Smirnov (1939b) and later expanded by himself in 1948. The table was further modified by Kunisawa et al. (1951 and 1955).

The one-sample K-S test is very attractive due to its non-parametric nature. It is also generally regarded as more powerful than the well known Chi-Square test. Kolmogorov (1933) proved the limiting distribution (2.4) can even be applied to the two-sample K-S test statistic. However, the K-S test requires a continuous underlying distribution and the specified hypothetical distribution. The asymptotic approximation of the probabilities of Z_N works well with large sample. However, using it to approximate *p*-values of the probabilities of Z_N can be problematic when sample size is small. The direct impact is that the classical one-sample K-S test based on asymptotic approximation will be very conservative. The problem was later validated in the numerical study and the results are summarized in Figure 1.3 and Table 1.1.

To improve the accuracy of the approximated distribution of one-sample K-S test statistics, we propose a new approach to compute Z_N by replace $F(x)$ with an EDF based on a random sample drawn from the hypothetical distribution. The accurate estimation of $F(x)$ by an EDF requires a very large sample. This computing process may be time consuming. However, the computing power of contemporary computers makes it possible to compute the improved Z_N with large numbers of iterations and thus in turn to make the estimation of the near-exact distribution of Z_N possible.

1.2 Method and Numerical Study for Improving 1-Dimensional One-Sample K-S test

The proposed replacement of $F(x)$ in (2.4) with an EDF approximated from a large random sample of a hypothetical distribution is an application of the law of large numbers of EDFs. If the true underlying distribution of a random sample is $F(x)$, the law of large numbers implies that $F(x)$ is consistent: $F_N(x)$ converges to $F(x)$ as $(N \to \infty)$ almost surely for every value of x,

(2.6) $$F_N(x) \overset{a.s.}{\longrightarrow} F(x), \text{ every } x.$$

The *Glivenko-Cantelli* theorem extends the law of large numbers and states that the convergence in (2.5) happens uniformly over x (van der Varrt, 1998),

(2.7) $$||F_n - F||_\infty = \sup_{\mathrm{x} \in \mathrm{R}} |F_N(x) - F(x)| \overset{a.s.}{\longrightarrow} 0 \ .$$

The theorem basically points out that the distribution of random variable $sup_{x\in R}\,|F_N(x) - F(x)|$ converges in probability to zero when sample size is at infinity. In addition, if $F(x)$ is continuous then the distribution of random variable $sup_{x\in R}\,|F_N(x) - F(x)|$ does not depend on *F*.

For one-sample K-S test, to implement the proposed improvement, we can draw a random sample of size *m* from a hypothetical distribution. The resulting EDF is denoted by $F^{(m)}(x)$. $F^{(m)}(x)$ clearly depends on the random draws and is different from sample to sample. However, with the aid of the computing power of contemporary computers, we can quickly draw a large sample of size *m*, such as *m* = 10000, with ease. Thus, $F^{(m)}(x)$ is a good approximation of $F(x)$ as stated in (2.6) and (2.7) according to the law of large numbers. By replacing $F(x)$ with $F^{(m)}(x)$ in (2.2) and (2.4), we can compute

(2.8) $$D_N^{(m)} = \sup_{-\infty<x<\infty} |F_N(x) - F^{(m)}(x)|\,,$$

(2.9) $$Z_N^{(m)} = N^{\frac{1}{2}} D_N^{(m)}.$$

The distribution of $D_N^{(m)}$ can be computed exactly (Xiao et al., 2007), so is the normalized version $Z_N^{(m)} = N^{1/2} D_N^{(m)}$. The exact distribution of $N^{1/2} D_N^{(m)}$ is a good approximation of the K-S test statistics Z_N and it will be termed the near-exact distribution of Z_N.

A series of simulations was conducted to assess the proposed improvement to the one-sample K-S test. The random samples generation for the study was based on idea of integral probability transformation. If a random variable X has a continuous distribution for which the cumulative distributive function is F_X, the probability integral transformation $Y = F_X(X)$ has a uniform distribution on (0, 1). Thus, we may assume the hypothetical distribution is uniform on (0, 1), and regard the sample points $x_1,\ x_2, \dots, x_N$ as random draws from the uniform distribution on (0, 1) without loss of generality if the null hypothesis is true. From now on we will assume that the random sample $x_1,\ x_2, \ldots\ , x_N$ is uniform on (0, 1).

To investigate the relationship between the near-exact tail probability and the asymptotic distribution of Z_N at various sample sizes. $F_N(x)$ was approximated from the random samples of size *N* drawn from Uniform(0, 1), where *N*=5, 10, 15, 20, 25, 40 and 100. For each *N*, $F^{(m)}(x)$ was approximated from random samples of size *m* drawn from Uniform(0, 1), where *m* = 10,000. By repeating the process *k*=10,000 times, we can compute the asymptotic tail probability and the near-exact tail probability of Z_N. The results are presented in Table 1.1 and Figures 1.1.

To validate how well the exact distribution of $Z_N^{(m)}$ can approximate the near-exact distribution of Z_N, we did a simulation study with random samples drawn from Uniform(0, 1) of size *N*, where *N* = 5, 10, 20, and 40, to approximate $F_N(x)$. $F^{(m)}(x)$ was approximated with random samples drawn from Uniform(0, 1) of size *m*, where *m* = 1000 and 2500. The exact tail probability of $Z_N^{(m)}$ was computed with *k* = 1000 iterations. The near-exact distribution of Z_N was computed in the simulation described previously. The results are presented in Figures 1.2 to 1.3 and they confirmed that the exact distribution of $Z_N^{(m)}$ is a good approximation to the near-exact distribution of Z_N.

1.3 Results and Discussions of Improving 1-Dimensional One-Sample K-S test

In Figure 1.1, the differences between the near-exact and asymptotic distribution of improved Z_N are impacted by sample size. The differences between the two distributions become greater with the decrease of sample size and reach the maximum when *N* = 5. The differences become non-detectable when the sample size reaches 100, which implies that the asymptotic approximation works very well for sample size greater than 100. The simulation results confirm that using the asymptotic distribution tends to give greater *p*-values to Z_N than those from its near-exact distribution when sample size is less than 100. Apparently, using the asymptotic to approximate of *p*-values will result in overly conservative one-sample K-S test for small sample size. The proposed method will be able to facilitate more powerful test with the near-exact distribution of Z_N.

Table 1.1 summarized the critical values of Z_N from the near-exact distribution as a function of sample size at various α levels, 0.01, 0.05 and 0.10. The critical values are retrieved from the near-exact and asymptotic distributions of Z_N. In each alpha level, Z_n becomes greater with the increase of sample size and reaches the maximum value when it was from the asymptotic distribution. For instance, the critical values from the near-exact and asymptotic distributions are 1.242 and 1.358 respectively for $\alpha = 0.05$ and n = 5, which indicates that the effective significant region from the near-exact distribution is much larger than the one from the asymptotic approximation. The evidence exhibits the advantage of the near-exact distribution of Z_N over the asymptotic distribution of one-sample K-S statistics for small sample size.

Figures 1.2 and 1.3 clearly demonstrates that the exact distribution of $Z_N^{(m)}$ is a good approximation to the near-exact distribution of Z_N even when *m* and *k* are as low as 1000. The close similarity between the two distributions in Figures 1.2 and 1.3 is highly noticeable, which demonstrates that using $F^{(m)}(x)$ to replace $F(x)$ is a practical choice to compute the near-exact distribution of Z_N. If we use the random samples generated with *m* = 10000 and *k* = 10000 from the hypothetical distribution to compute $F^{(m)}(x)$, we can be highly certain that we will obtain the exact distribution of $Z_N^{(m)}$ that will real a good representation of the near-exact distribution of Z_N.

In summary, replacing $F(x)$ in (2.4) with an EDF approximated from a large random sample of the hypothetical distribution is an effective approach to compute the near-exact distribution of Z_N. The numerical study confirmed that the near-exact distribution of $Z_N^{(m)}$ is a good approximate to the near-exact distribution of Z_N for *m* as low as 1000. The near-exact distribution of Z_N can considerably improve the power of the one-sample K-S test for small sample size.

Table 1.1 The near-exact and asymptotic critical values of the classical 1D one-sample K-S test

Sample Size	α = 0.01	α = 0.05	α = 0.10
5	1.4904	1.2420	1.1395
10	1.5591	1.2993	1.1685
15	1.5608	1.3084	1.1803
20	1.5836	1.3159	1.1876
25	1.5899	1.3223	1.1829
40	1.5809	1.3240	1.1962
100	1.6101	1.3462	1.2028
Asymptotic	1.6275	1.3581	1.2238

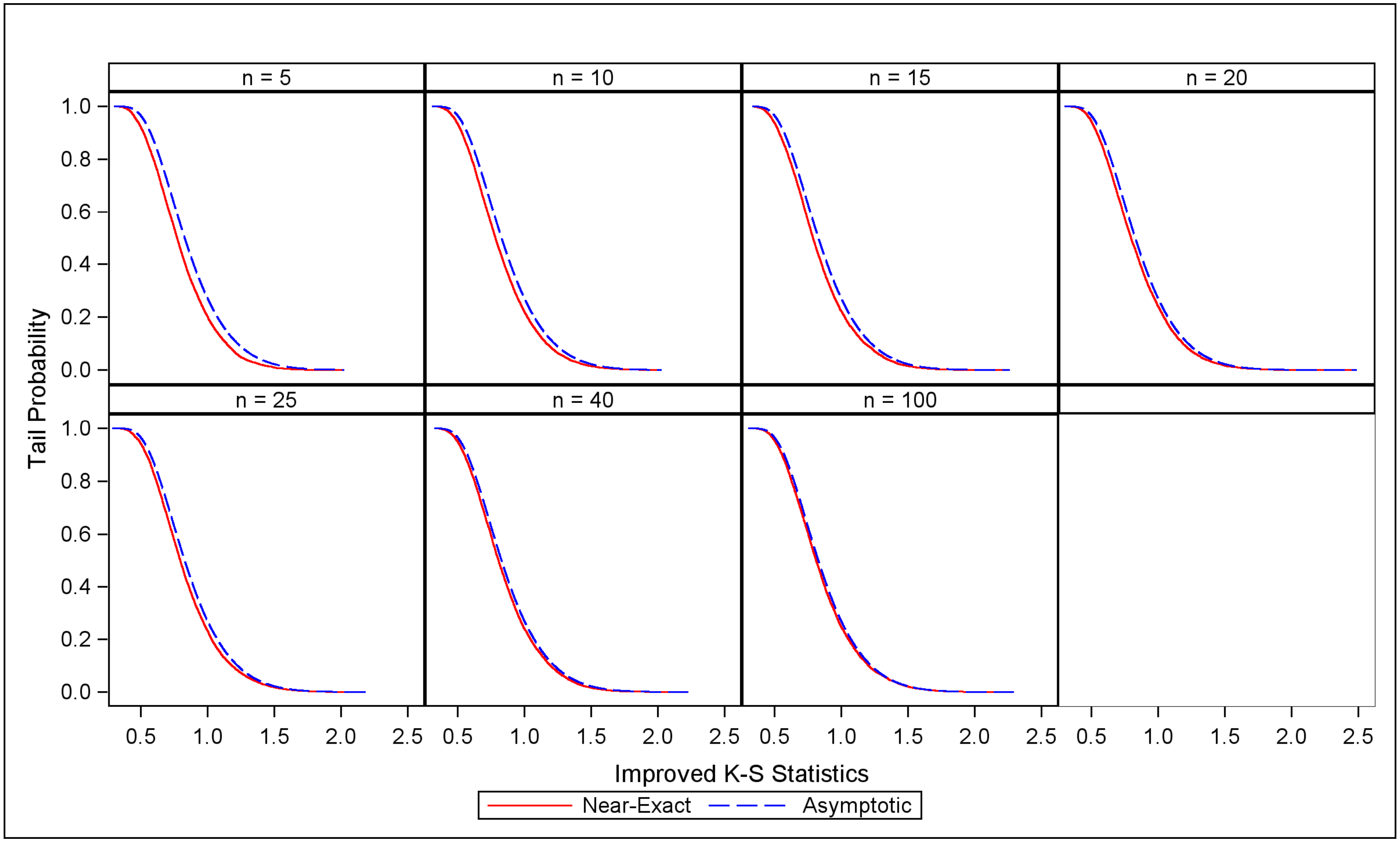

Figure 1.1 The near-exact and asymptotic tail probabilities of the modified 1D one-sample K-S test by sample size

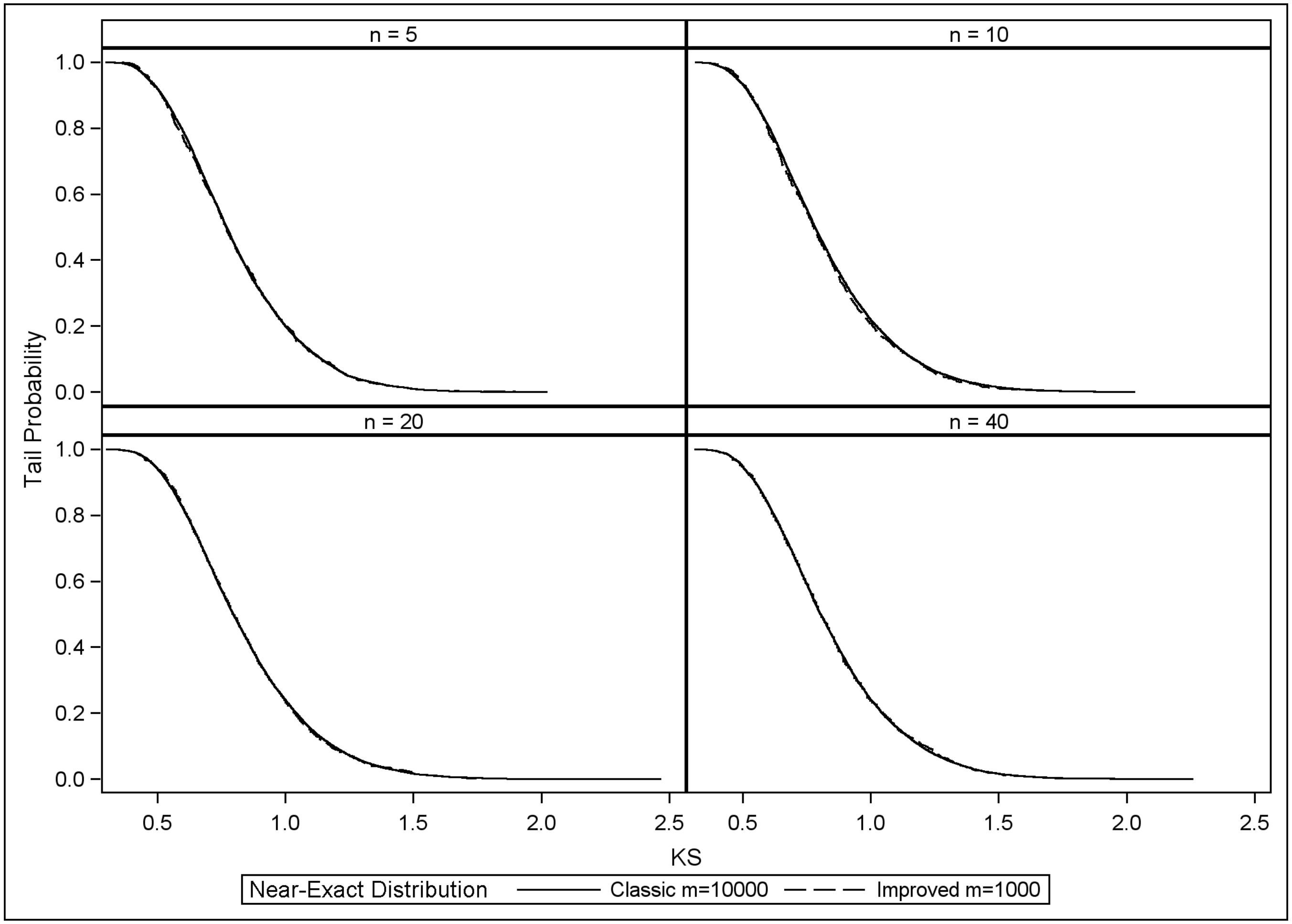

Figure 1.2 Tail probabilities of the classical and improved test statistics for one-sample K-S test (m= 1000 and k=1000)

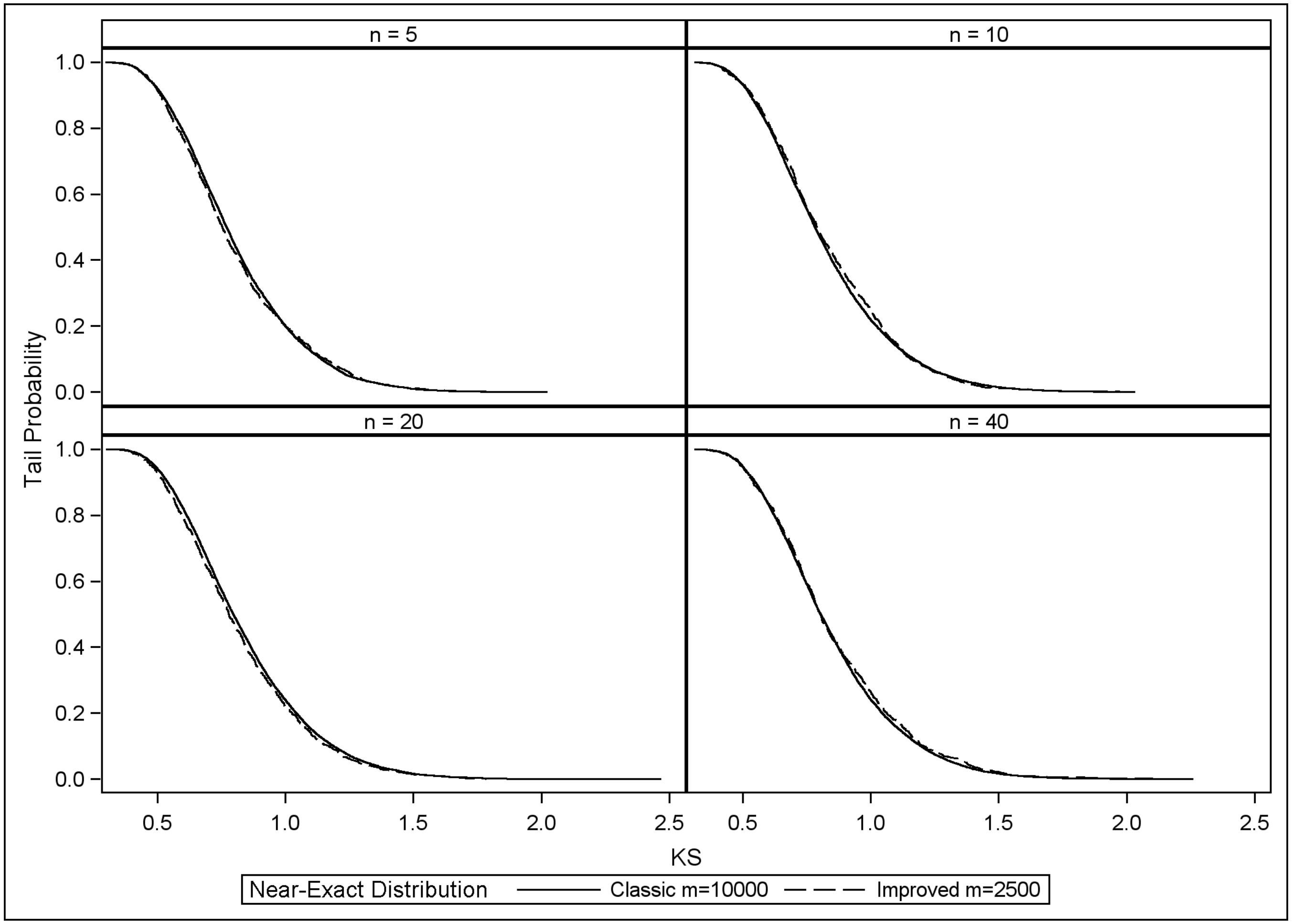

Figure 1.3 Tail probabilities of the classical and improved test statistics for one-sample K-S test (m=2500 and k=1000)

2 Improving 2-Dimensional Two-Sample K-S Test

2.1 Introduction

In Section I of this thesis, we discussed the properties of and improvement to the classical one-sample K-S test. These discussions are under the domain of one-dimensional distribution. If the research objects have to be characterized by two random quantities or comes (x, y), comparing the similarity of the empirical bivariate distribution of (x, y) in (x, y) plane to a hypothetical bivariate distribution is the one-sample case of 2-dimensional (2D) K-S test. The comparison of the similarity of two bivariate distributions of 2D data is the two-sample case of 2D two-sample K-S test. Let's denote the 2D random quantities for the two samples case as,

(2.1) $$\left\{\left(x_i^k, y_i^k\right),\ \text{where } k = 1 \text{ and } 2, \text{and } 1 \le i \le n_k\right\} \text{ for two-sample,}$$

with unknown bivariate distribution *F* and *G*. The goal is to test the null hypothesis whether *F* is the same as *G*. One difficulty of the 2D K-S test is to define the cumulative probability distributions since there is more than one direction to define EDFs. For instance, there are 3 independent ways to define cumulative distribution for 2D data.

Peacock (1983) developed an approach to compute the absolute maximum difference for 2D data between two bivariate distributions. Peacock's method considers all combinations of (x, y) and uses the four natural quadrants created by coordinates using $\left(x_i, y_j\right)$ as the origin. The approach uses the four quadrants in turn to integrate the cumulative distributions and compute the maximum of the four differences between the EDF and hypothetical cumulative distribution over all data points. Since Peacock's method will consider all combinations of

(x_i, y_j), where $1 \le i, j \le n_1 + n_2$, it is very expensive in terms of computing time if sample size gets large. Fasano and Franceschini (1987) modified Peacock's method by only considering observed data points (x_i, y_i), where $1 \le i \le n_1 + n_2$ for the two-sample test.

To describe the EDFs of 2D data for the Fasano and Franceschini method, we start with the denotation of *l* for $x \le x_i$ or $y < y_i$ and g for $x > x_i$ or $y \ge y_i$ and follow that by defining the EDF of each quadrant as these,

(2.2)
$$Pgg^k(x,y) = \left\{\sum I(\, x^k > x_i, y^k \ge y_i, 1 \le i \le n_k)\right\}/n_k,$$
$$Pgl^k(x,y) = \left\{\sum I(\, x^k > x_i, y^k < y_i, 1 \le i \le n_k)\right\}/n_k,$$
$$Plg^k(x,y) = \left\{\sum I(\, x^k \le x_i, y^k \ge y_i, 1 \le i \le n_k)\right\}/n_k,$$
$$Pll^k(x,y) = \left\{\sum I(\, x^k \le x_i, y^k < y_i, 1 \le i \le n_k)\right\}/n_k,$$

where $k = 1$ and 2. After the pooled sample is defined to be

(2.3)
$$\{(x_t^0, y_t^0),\ \text{where and } 1 \le t \le n_1 + n_2\},$$

we can define the distance between two EDFs of each quadrant as

(2.4)
$$D_{nsgg}^t = |Pgg^1(x_t^0, y_t^0) - Pgg^2(x_t^0, y_t^0)|,$$
$$D_{nsgl}^t = |Pgl^1(x_t^0, y_t^0) - Pgl^2(x_t^0, y_t^0)|,$$
$$D_{nslg}^t = |Plg^1(x_t^0, y_t^0) - Plg^2(x_t^0, y_t^0)|,$$
$$D_{nsll}^t = |Pll^1(x_t^0, y_t^0) - Pll^2(x_t^0, y_t^0)|.$$

where $1 \le t \le n_1 + n_2$. Thus, the test statistics of the 2D two-sample K-S test is,

(2.5)
$$D_{ns} = \max\{D_{nspp}^t, D_{nspl}^t, D_{nslp}^t, D_{nsll}^t\}.$$

In this section, two new test statistics, T_1 and $T_{2,}$ are proposed to improve the power of the 2D two-sample K-S test. Based on (2.4), T_1 and T_2 can then be expressed as,

(2.6a) $$T_1 = \sum_{1 \le t \le n_1+n_2} D^t_{nspp} + D^t_{nspl} + D^t_{nslp} + D^t_{nsll} \,,$$

(2.6b) $$T_2 = \sum_{1 \le t \le n_1+n_2} {D^t_{nspp}}^2 + {D^t_{nspl}}^2 + {D^t_{nslp}}^2 + {D^t_{nsll}}^2 \,.$$

2.2 Method and Numerical Study for Improving 2-Dimensional Two-Sample K-S Test

For 1D data, the classical K-S test is usually cited for lack of power and thus in practice it needs large sample size to reject the null hypothesis. CVM test is one of the distribution free tests that is proved to be more powerful than the classical K-S test. In this paper, we intend to apply CVM type approach to introduce two new statistics as described in (2.6a) and (2.6b) to the 2D two-sample K-S test and assess any improvements can be achieved in terms of testing power.

The CVM test was originally proposed by Cramer (1928) and von Mises (1931), which uses the summation of squared distances between an EDF and a hypothetical distribution (or another EDF) as its test statistics ${W^2}_N$. The critical values of CVM test were tabulated by Anderson and Darling (1954). Anderson and Darling (1954) also extended (2.7) into a more general form.

(2.7) $${W^2}_N = \int_{-\infty}^{\infty} (F_N(x) - F(x))^2 [F(x)] dF(x) \text{ for one-sample case.}$$

For a hypothesis about whether two random samples with sample size *N* and *M* respectively are from the same unspecified continuous distribution, the test statistics can be expressed as

(2.8) $$W^2{}_N = NM/(N+M)\int_{-\infty}^{\infty}(F_N(x) - G_M(x))^2 dH_{N+M}(x)\,,$$

where $F_N(x)$ and $G_M(x)$ are the EDFs of two random samples and $H_{N+M}(x) = (NF_N(x) - MG_M(x))/(N+M)$ is the pooled distribution of the two samples. The critical values of the two-sample CVM test were tabled by Smirnov (1948) and improved and expanded by Massey (1951b), and Anderson and Darling (1952).

The CVM test is more powerful than classical K-S test, and Chi-Square test. The distribution free is also a very appealing trait. However, the implementation of CVM test is limited by lack of efficient algorithm and programming. Xiao et al. (2007) developed an efficient algorithm and C++ program package to compute the CVM test efficiently. Leveraging the algorithm improved by Xiao et al., we compute the two proposed T_1 and T_2 statistics by using the method proposed by Fasano and Franceschini (1987) to integrate cumulative probabilities of the four natural quadrants formed by each observed data points.

To confirm if these two new test statistics can improve the power of the 2D two-sample K-S test, a numerical study was conducted using the random samples generated from two standard bivariate normal distributions, one uncorrelated and one correlated between two data dimensions. The random samples were generated with $n_1 = n_2 = 10$ to 100 by increment of 10 from the un-correlated distribution. For each n, random samples were generated from the correlated distribution with correlation coefficients range from 0.1 to 0.9 by increment of 0.1. Each random sample had 5000 pairs of data points from both distributions. Because multi-dimensional K-S and CVM test are no longer distribution-free, 1000 permutations were used to estimate p-values. The power was estimated for the 2D two-sample K-S test with T_1, T_2 and D_{ns} test statistics at significant levels 0.05 and 0.10.

2.3 Results and Discussions of Improving 2-Dimensional Two-Sample K-S Test

The simulation results apparently show that T_1 and T_2 statistics are superior to D_{ns} in terms of increase the power of the 2D two-sample K-S test under given setting. Figures 2.4 and 2.5 present the comparisons of testing power of three test statistics. Both figures display similar trends and patterns of test power as a function of correlation coefficients and sample size (Figure 2.4 will be discussed). Throughout the range of the correlation coefficients from 0.1 to 0.9, T_1 and T_2 statistics significantly improve the power of the 2D two-sample K-S test over D_{ns} over all sample size tested. Both T_1 and T_2 raise the power more steeply than D_{ns} does when the correlation coefficients are greater than 50%. The power curves of T_1 and T_2 start to converge to 100% at the high end of the correlation scale for the sample size 80 or greater, whereas D_{ns} shows no signs of such convergence. One unexpected finding is that T_1 statistics performs better than T_2 statistics in raising the power of the 2D two-sample K-S test regardless the correlation coefficients, sample size, and alpha levels evaluated.

The power of the 2D two-sample K-S test highly dependent on both sample size and correlation coefficient regardless the test statistics involved (Figure 2.1, Figure 2.2 and Figure 2.3). Such dependency shown by D_{ns} is presented in Figure 2.1. The classical 2D two-sample K-S test has very low power and it can’t reach meaningful power, such as 80%, until the sample size is greater than or equal to n = 70 at high correlation level. The power increases with the increase of correlation regardless of the sample sizes and alpha level simulated. We also observed accelerated increase of power between correlation coefficients 0.4 and 0.8 for sample sizes 20 or greater.

Figure 2.2 and Figure 2.3 summarizes the effect of T_1 and T_2 test statistics on the power in relationship to the correlation coefficients and sample size. The patterns and trends of power from T_1 and T_2 test statistics are similar to those from D_{ns}. T_1 and T_2 statistics also promote the power of the 2D two-sample K-S test to above 80% at the high correlation level for sample size as small as 40, which is a substantial improvement over the D_{ns}. T_1 and T_2 show signs of convergence to the power level of 100% at the 90% correlation for the sample size 70 and 80 respectively over the two alpha levels analyzed.

The dependence of the power of the test on the correlation is evident for the 2D two-sample K-S test. The fast increase of testing power at the higher end of the correlation scale has been observed for all three test statistics and the phenomenon may be from the methodology itself. The power is more sensitive to the change of correlation when the test uses T_1 and T_2 as test statistics than uses D_{ns}. Fasano and Franceschini's (1987) algorithm to integrate probability of the four quadrants around the observed data pairs was conceived to be capable to reduce the number of quadrants contributing to computing Z_{ns} to 2 as the correlation between data points increase. If the perfect correlation exists, the distribution of the 2D test statistics collapses into a one-dimensional case. Such reduction in contributing quadrants and data dimensions might dependency of power and correlation.

In general, T_1 and T_2 statistics have been shown by the numerical study are far more superior to D_{ns} in terms of increasing power of the 2D two-sample K-S test regardless of sample size and correlation coefficients tested. T_1 performs even better than T_2 in increasing the testing power for the 2D two-sample K-S test.

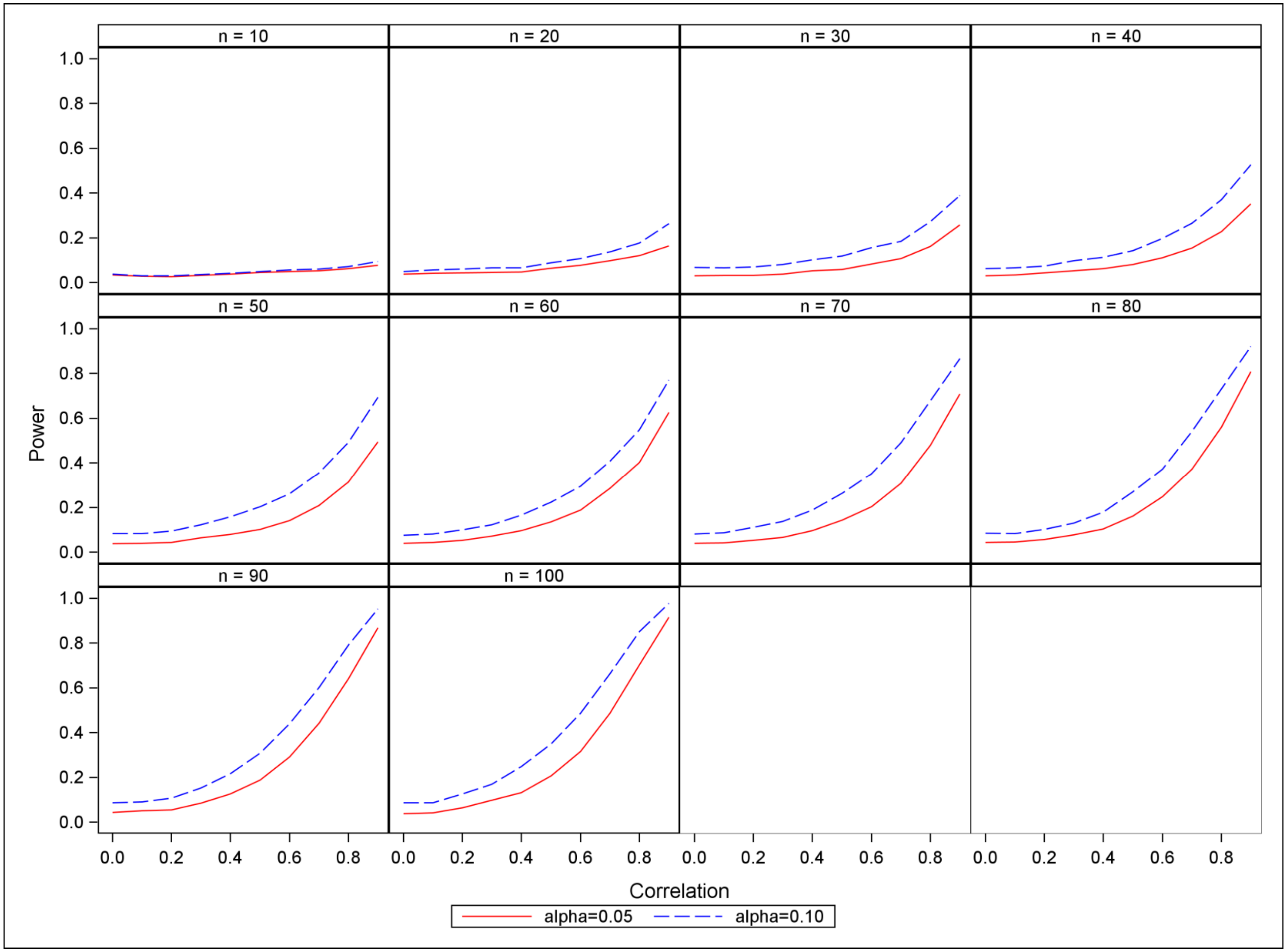

Figure 2.1 The power of the classical 2D two-sample K-S test as a function of correlation by sample size and alpha level

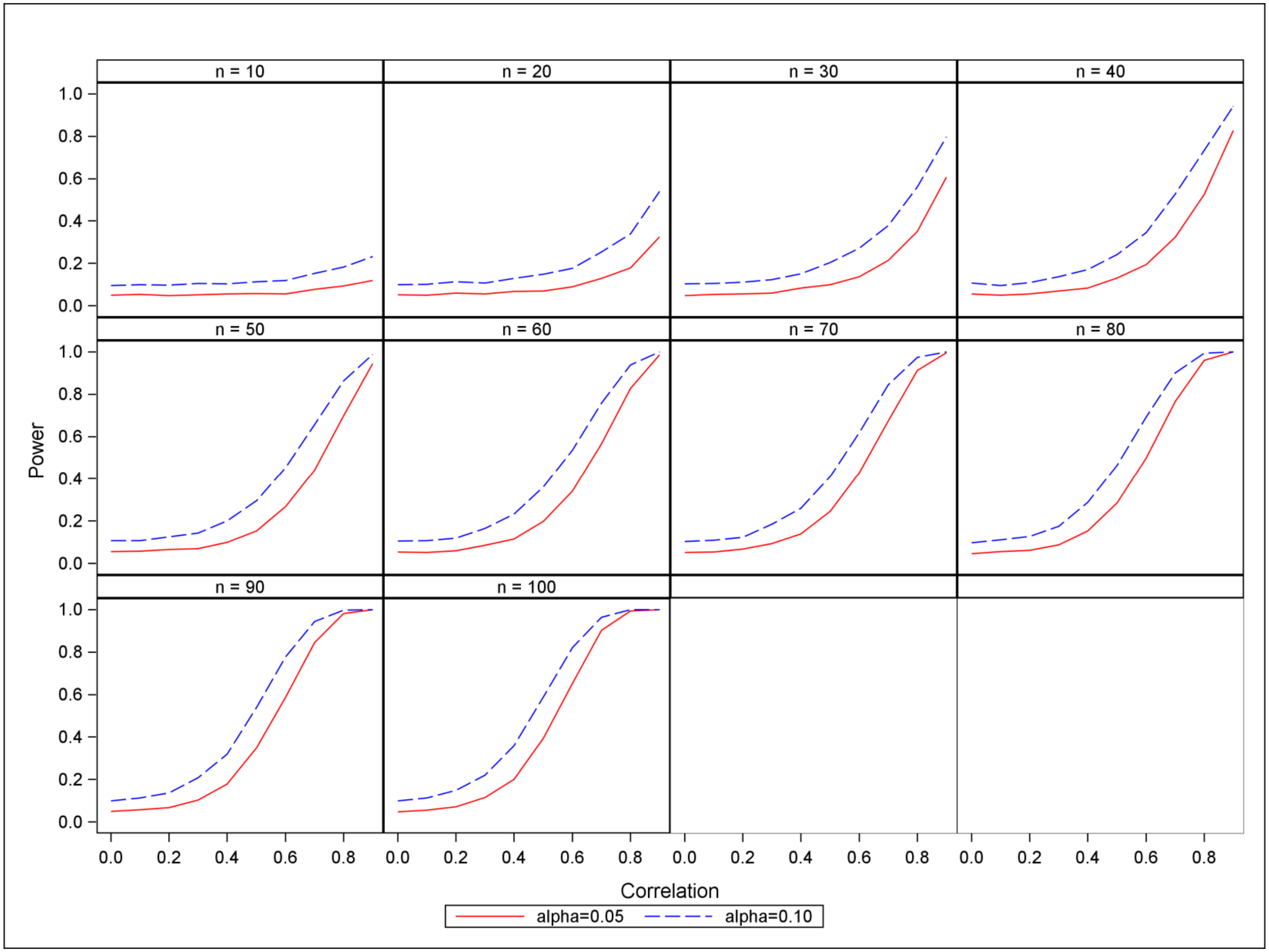

Figure 2.2 The power of 2D two-sample K-S test with T_1 statistics as a function of correlation by sample size and alpha level

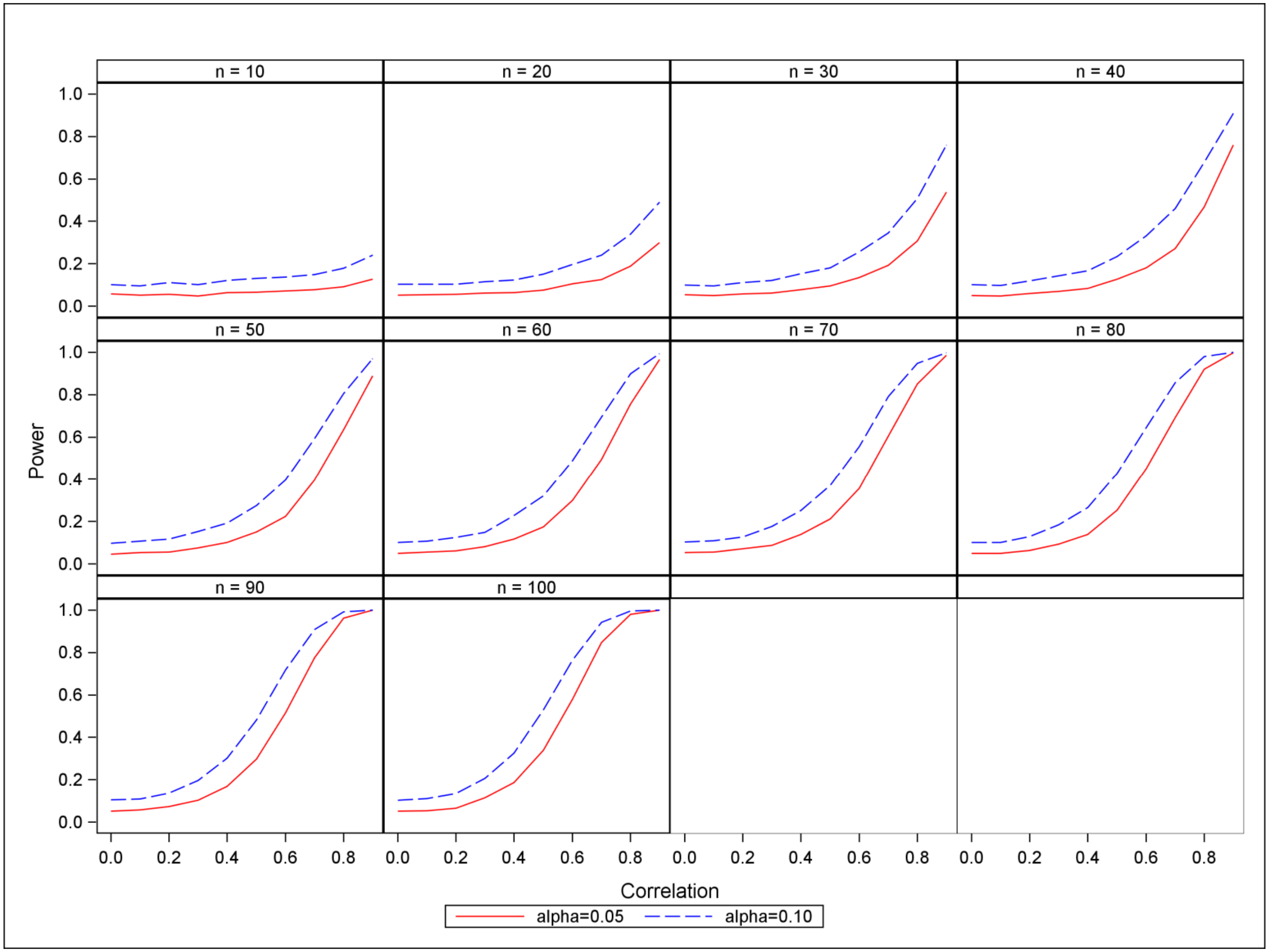

Figure 2.3 The power of 2D two-sample K-S test with T_2 statistics as a function of correlation by sample size and alpha level

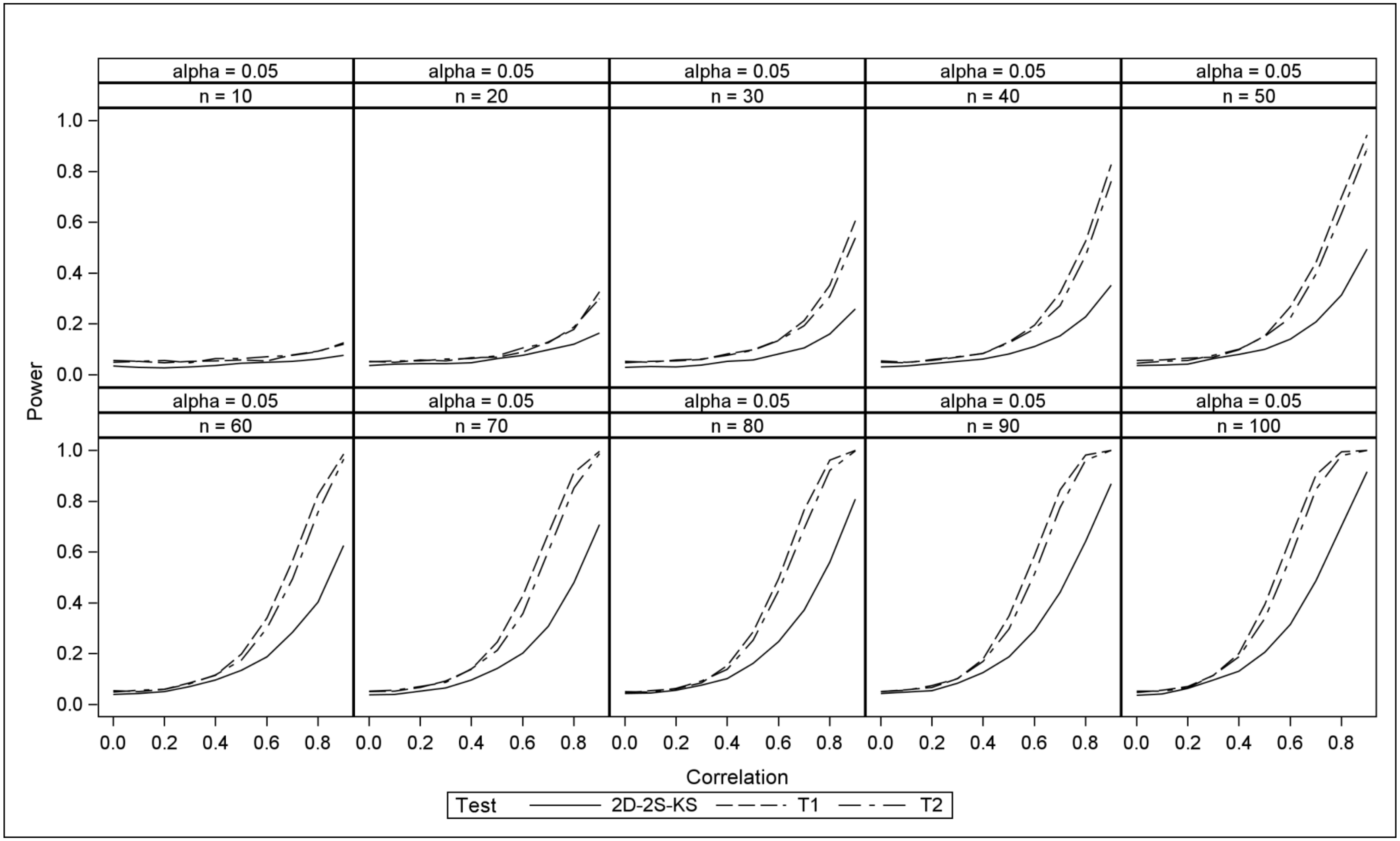

Figure 2.4 The power of 2D two-sample K-S test with three test statistics as functions of correlation by sample size for alpha = 0.05

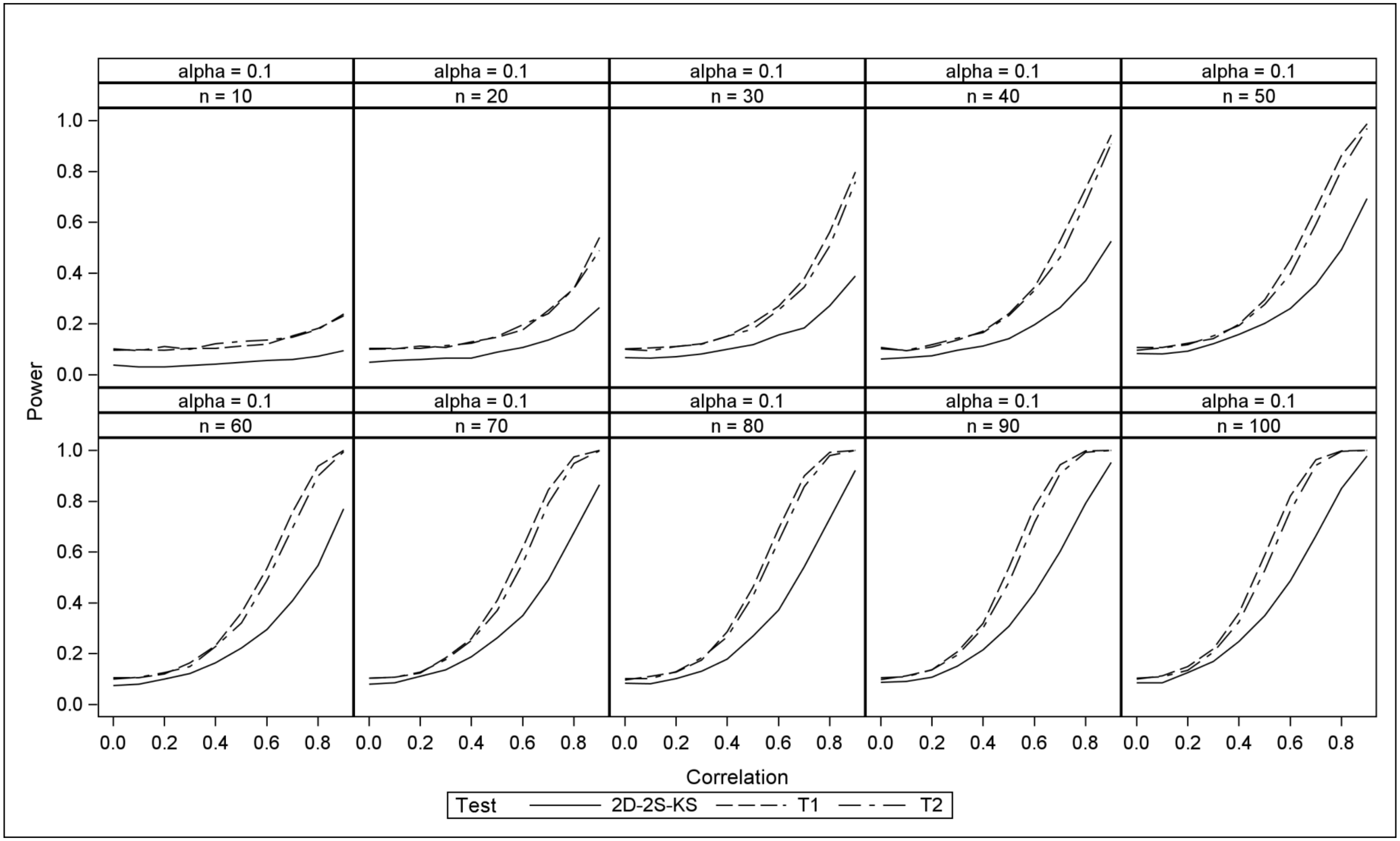

Figure 2.5 The power of 2D two-sample K-S test with three test statistics as functions of correlation by sample size for alpha = 0.10

REFERENCES

Cramér, H. (1928), "On the composition of elementary errors," Skand. Aktuarietids., Vol. 11, pp. 13-74 and pp. 141-180

Mises, von. (1931), Wahrscheinlichkeitsrechnung und Ihre Anwendung in der Statistik und Theoretischen Physik, F. Deuticke, Leipzig (6.1)

Kolmogorov, A. (1933), "Sulla Determinazione Empirica di una Legge di Duistributione," Giornale dell' Istituto Ialiano delgli Attuar, 4, pp. 1-11

Smirnov, N. (1939a), "Sur les Ecarts de la Courbe de Distribution Empirique,", Recueil Mathematique (Matematiceskii Sbornik), N. S. 6, pp. 3-26

Smirnov, N. (1939b), "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," Bull. Math. Univ. Moscow, Vol. 2, No. 2, pp. 3-14

Kolmogorov, A. (1941), "Confidence limits for an unknown distribution function," Annals of Mathematical Statistics, Vol. 12, pp. 461-463

Scheffe, H. (1943), "Statistical Inference in the Non-parametric Case," Annals of Mathematical Statistics, 14, 305-342

Feller, W. (1948), "On the K-S limited theorems for empirical distributions," The annual of Mathematical Statistics, Vol. 19, No. 2, pp. 177-189

Smirnov, N. (1948), "Table for Estimating the Goodness-of-fit of Empirical Distributions," Annls of Mathematical Statistics, 19, pp. 279-281

Doob, J. L. (1949), "Heuristic Approach to the K-S Theorems," Annals of Mathematicals, 20, 393-403

Wolfowitz, J. (1949), "Non-parametric Statistical Inference," Proceedings of the Berkeley Symposium of Mathematical Statistics and Probability, Berkeley (University of California Press), pp. 93-113

Kunisawa, K., Makabe, H., and Morimura, H. (1951), "Table of confidence bands for the population distribution function," Rerports of Stat. App. Res. JUSE, Vol. 1, pp. 23-44

Massey, F. M. (1951b), "The distribution of the maximum deviation between two samples cumulative step functions," the Annuals of Mathematical Statistics, Vol. 22, No. 1, pp. 125-128

Anderson, T. W. and Darling, D. A. (1952), “Asymptotic theory of certain ‘goodness-of-fit’ critera based on stochastic processes,” Ann. Math. Statist., 23, pp. 193-212

Anderson, T. W. and Darling, D. A. (1954), “A Test of Goodness-of-fit,” Journal of the American Statistical Association, Vol. 49, No. 268, pp. 765-769

Kunisawa, K., Makabe, H., and Morimura, H. (1955), “Notes on the confidence bands of population distributions,” Rerports of Stat. App. Res. JUSE, Vol. 4, pp. 18-20

Peacock, J. A., (1983), “Two-dimensional goodness-of-fit testing in astronomy,” Monthly Notices of the Royal Astronomical Society, 202, pp. 615-627

Fasano, G., and Franceschini, A., (1987), “A multidimensional version of the K-S test,” Monthly Notices of the Royal Astronomical Society, 225, pp. 155-170

van der Vaart, A.W. (1998). Asymptotic statistics. Cambridge University Press. ISBN 978-0-521-78450-4.

Xiao, Y., Gordon A., and Yakovlev A, (2007), “A C++ Programming for the Cramer-von Mises Two-Sample Test,” Journal of Statistical Software, Vol. 17, No. 8, pp. 1-15