Electronic Journal of Statistics
Vol. 5 (2011) 342–372
ISSN: 1935-7524
DOI: 10.1214/11-EJS610

Asymptotic independence of correlation coefficients with application to testing hypothesis of independence

Zhengjun Zhang

Department of Statistics, University of Wisconsin, Madison, WI 53706, USA
e-mail: zjz@stat.wisc.edu

Yongcheng Qi

Department of Mathematics and Statistics, University of Minnesota Duluth, Duluth, MN 55812, USA
e-mail: yqi@d.umn.edu

and

Xiwen Ma

Department of Statistics, University of Wisconsin, Madison, WI 53706, USA
e-mail: xiwenma@stat.wisc.edu

Abstract: This paper first proves that the sample based Pearson's product-moment correlation coefficient and the quotient correlation coefficient are asymptotically independent, which is a very important property as it shows that these two correlation coefficients measure completely different dependencies between two random variables, and they can be very useful if they are simultaneously applied to data analysis. Motivated from this fact, the paper introduces a new way of combining these two sample based correlation coefficients into maximal strength measures of variable association. Second, the paper introduces a new marginal distribution transformation method which is based on a rank-preserving scale regeneration procedure, and is distribution free. In testing hypothesis of independence between two continuous random variables, the limiting distributions of the combined measures are shown to follow a max-linear of two independent χ^2 random variables. The new measures as test statistics are compared with several existing tests. Theoretical results and simulation examples show that the new tests are clearly superior. In real data analysis, the paper proposes to incorporate nonlinear data transformation into the rank-preserving scale regeneration procedure, and a conditional expectation test procedure whose test statistic is shown to have a non-standard limit distribution. Data analysis results suggest that this new testing procedure can detect inherent dependencies in the data and could lead to a more meaningful decision making.

AMS 2000 subject classifications: Primary 62H20, 62G10; secondary 62G30.
Keywords and phrases: Dependence measures, rank-preserving scale regeneration, distribution free test, Fisher's Z-transformation tests, gamma tests, χ^2 tests, order statistics, conditional expectation.

Received August 2010.

Contents

1. Introduction

The determination of independence between variables enables us to measure variable associations and draw accurate conclusions in a quantitative way. Testing independence is a durable statistical research topic, and applications are very broad. The seminal nonparametric test of independence by [16, 3] has been widely used in many applications. The test (we shall use HBKR test throughout the paper) is based on Hoeffding's D-statistic which only uses ranks of the observed values. In the literature, [13] introduced the extremely useful Z-transformation function and test statistics, and [18] stated that the best present-day usage in dealing with correlation coefficients is based on [12, 13]. Recently, [14] showed that after Box-Cox transformation to approximate normal scales, the correlation based test is more efficient when testing independence between two positive random variables. Besides Fisher's Z-transformation test, many other testing procedures have been developed, for instance, testing independence in a two-way table ([9]), testing independence between two covariance stationary time series ([17]), testing the independence assumption in linear models ([5]), and others such as [1, 10, 29], as well as excellent reference books by [6, 7, 21] and [27], among others.

When alternatives are either linearly dependent or nonlinearly dependent, although much work has been done, there is no public consensus on which method is the best. Empirical evidences have shown that the gamma test ([30]) performs much better than Fisher's Z-transformation test when testing independence while the alternatives are nonlinear dependence. There are certainly other test statistics which can also achieve high detecting powers. In problems for testing independence, it is not hard to find examples such that one test performs better than another test, and vice versa. For example, Pearson's correlation coefficient based test statistics would have the best performance when the alternative hypothesis is two random variables being linearly dependent, especially when they are bivariate normal random variables.

Correlation is an extremely important and widely used analytical tool in statistical data analysis. The classical correlation coefficient is Pearson's product moment coefficient which indicates the strength and direction of a linear relationship between two random variables. Its history can be traced back to the 19th century when Sir Francis Galton introduced correlation and regression, while Karl Pearson provided the mathematical framework we are familiar with today. Since then, various (linear or nonlinear) correlation measures (coefficients) have been introduced in statistical inferences and applications. Examples include Spearman's rank correlation coefficient ([24]), Kendall tau rank correlation coefficient ([22]), the distance correlation coefficient ([26]), the quotient correlation coefficient ([30]), among many others. In the literature, there have been attentions on other dependence measures such as positive dependence, co-monotone dependence (or co-movement measures), negative dependence and setwise dependence. [11, 20] are excellent reference books related to correlations and multivariate dependencies like aforementioned ones. These dependence measures will not be considered in this paper.

An effective combination of correlation measures may be more informative and useful in measuring association (strength) of the relationship, or in testing independence, between two random variables. In some applications, one correlation measure may lead to a better interpretation of data than another correlation measure does, and vice versa. As a result, it is natural to consider a combination of as many measures as possible to get a maximal strength measure of the relationship between two variables. This strategy may be too general to apply since a limiting distribution of the combined measure may be difficult to derive, and hence it is not applicable in testing hypothesis of independence. One may argue that a randomized permutation test can overcome this drawback. However, a randomized permutation test does not guarantee a better performance. Our simulation examples of combining Pearson's correlation coefficient, Spearman's rank correlation coefficient, Kendall tau rank correlation coefficient demonstrate that combining more than two measures into a maximal strength measure does not result in a better performance. The performance really depends on which measures are included in the combination. Ideally, one would want a combined measure which uses as few (say ≥ 2) measures as possible and achieves relatively better performances when compared with commonly used measures.

Considering that Pearson's correlation coefficient and the quotient correlation coefficient are *asymptotically independent*, i.e., they measure completely different dependence between two random variables, we propose to first combine these two sample based correlation coefficients into a new association measure of the relationship between two continuous random variables. On the other hand, Pearson's correlation coefficient may be the best linear dependence measure to model dependence in the central region of the data, and the quotient correlation is a nonlinear dependence measure to describe dependence in extreme values. These facts indicate that a combination of these two coefficients will reveal more inherent dependence between random variables. Due to the fact that the sample based Pearson's correlation coefficient is $\sqrt{n}$ convergent, while the quotient correlation coefficient is n convergent under the null hypothesis of independence, the new associate measure is defined as the maximum of the squared root of the quotient correlation coefficient and the absolute value of Pearson's correlation coefficient.

A combination of Pearson's correlation coefficient and the quotient correlation coefficient introduces a new dependence measure which leads to a new test statistic for testing independence. We note that it is possible to find some unusual examples from which both calculated coefficients are zero. In order to overcome this drawback and to make a test statistic capable of testing every alternative hypothesis, at least as broad as HBKR test, we include Hoeffding's D measure in our combination. We shall see that the magnitude of D statistic is almost negligible in many examples. The asymptotic distribution of the new test statistic follows a max-linear of two χ^2 random variables with 1 and 4 degrees of freedom respectively. This new test statistic performs better than existing tests. For example, it performs as good as (or better than) existing tests when the alternatives are linear dependence and guarantees the same detecting powers as the gamma test when alternatives are nonlinear dependence. We shall see that the newly proposed test statistic is constructed to sufficiently utilize and maximize detecting powers of each component coefficient which may be used as a test statistic alone, not just simply achieving a new test statistic which outperforms a linear correlation based test statistic.

The rest of the paper is organized as follows. Main theoretical results dealing with asymptotic independence are presented in Sections 2. Section 3 introduces the new combined maximal strength measure. The limiting distribution of the new combined correlation coefficient under the null hypothesis of two random variables being independent is derived, and then a new max linear χ^2 test statistic is introduced. In Section 4, we first derive a new rank-preserving scale regeneration marginal transformation procedure using simulation which is distribution free. The transformation procedure gives advantages in applying extreme value distributions. The limiting distribution of the test statistic defined in Section 3 is again shown to be max linear χ^2 distributed when the test statistic is based on the transformed data. We further propose a new testing procedure, i.e. a **r**ank-preserving sc**a**le regeneration and **c**onditional **e**xpectation (R.A.C.E) test, from a practical perspective. The asymptotic distribution of R.A.C.E test statistics is derived. Type I errors of the new test are illustrated. In Section 5, power comparisons in simulation examples are presented for bivariate random variables. In Section 6, we first extend our combined dependence measure to a more general form and illustrate methods of data processing and variable scale transformations. We apply the newly introduced max linear χ^2 test in testing independencies among a group of random variables which are main index variables related to cigarette taxes, cigarette sales, revenues, smoking rates, etc. Section 7 offers some remarks regarding application of our new results to statistical data analysis. Finally, some technical proofs are summarized in Section 8.

2. Asymptotically independent sample statistics

2.1. The quotient correlation defined by two extreme order statistics

Suppose X and Y are identically distributed positive random variables satisfying $\mathrm{P}(X \geq Y) > 0$, $\mathrm{P}(X \leq Y) > 0$. Then the quotients between X and Y are Y/X and X/Y. Like the difference $X - Y$, the quotients can be used to measure the relative positions between two variables.

Suppose $\{(X_i, Y_i),\ i = 1, \ldots, n\}$ is a bivariate random sample of (X, Y). Let us consider two special cases: 1) $X = Y$, i.e. the perfect dependence, then $Y/X = 1$, $X/Y = 1$, $\max_{i \leq n}\{Y_i/X_i\} = 1$, and $\max_{i \leq n}\{X_i/Y_i\} = 1$; 2) $X \in (0, \infty)$, $Y \in (0, \infty)$, and X and Y are independent, then $\max_{i \leq n}\{Y_i/X_i\} \xrightarrow{a.s.} \infty$ and $\max_{i \leq n}\{X_i/Y_i\} \xrightarrow{a.s.} \infty$ as $n \to \infty$. When X and Y are neither perfectly dependent nor independent, it is easy to see that both $(\max_{i \leq n}\{X_i/Y_i\})$ and $(\max_{i \leq n}\{Y_i/X_i\})$ are asymptotically greater than 1 as $n \to \infty$, or we say that $(\max_{i \leq n}\{X_i/Y_i\})$ and $(\max_{i \leq n}\{Y_i/X_i\})$ are very likely to fall in $(1,\infty)$ for sufficiently large n.

These relations clearly tell that $\max_{i \leq n}\{Y_i/X_i\}$ and $\max_{i \leq n}\{X_i/Y_i\}$ can be used in defining a dependence measure between two random variables. When $\max_{i \leq n}\{Y_i/X_i\}$ and $\max_{i \leq n}\{X_i/Y_i\}$ are close to 1 for any given large sample size, we can say that there are co-movements between $\{X_i\}$ and $\{Y_i\}$, i.e. there are strong correlation between X and Y. When $\max_{i \leq n}\{Y_i/X_i\}$ or $\max_{i \leq n}\{X_i/Y_i\}$ is close to infinity or relatively very large, we can say that X and Y are near independent.

Following the classical way of defining a correlation coefficient in a range of (0,1), [30] introduced the sample based quotient correlation coefficient:

$$\mathrm{q}_n = \frac{\max_{i \leq n}\{Y_i/X_i\} + \max_{i \leq n}\{X_i/Y_i\} - 2}{\max_{i \leq n}\{Y_i/X_i\} \times \max_{i \leq n}\{X_i/Y_i\} - 1}. \tag{2.1}$$

Properties and geometric interpretation of (2.1) are illustrated in [30]. [30] also showed that $n\mathrm{q}_n \xrightarrow{\mathcal{L}} \zeta$ (a gamma random variable), assuming that X and Y are independent unit Fréchet random variables, i.e. $F_X(x) = F_Y(x) = e^{-1/x},\ x > 0$. For practical convenience, we make a convention that the marginal distributions of X and Y in (2.1) are transformed to be unit Fréchet distributed. Under this convention, quotient correlation coefficient can be applied to continuous random variables, but not discrete random variables.

One of the most usages of the quotient correlation coefficient is to test the hypothesis of independence between two random variables X and Y using the gamma test statistic $n\mathrm{q}_n$. The testing hypotheses are

$$H_0 :\ X \text{ and } Y \text{ are independent } \mathit{versus}\ H_1 :\ X \text{ and } Y \text{ are dependent.} \tag{2.2}$$

We shall see, for example, in a later section that the gamma test statistic $n\mathrm{q}_n$ does not achieve the best detecting power given that the alternative hypothesis is bivariate normal. The reason is that q_n measures the dependence among the tail part (extreme values) of the data, so it may lose its detecting power if the dependence mainly comes from the center part of the data. We note that a more efficient test statistic related to q_n for testing tail independence is the tail quotient correlation coefficient. Our focus in this paper is testing independence, not only testing tail independence.

As discussed in the introduction section, our goal is to find a test statistic which works for every alternative hypothesis of dependence and leads to a better performance than the existing test statistics. In the following sub-sections, we study asymptotic independence of order statistics, and asymptotic independence of sample correlations coefficients, which is very important in constructing a new maximal strength dependence measure and a powerful test statistic.

2.2. Asymptotic independence of order statistics and extended definition of quotient correlation coefficients

Now assume that $\{X_i, Y_i\}$, $i \geq 1$ are unit Fréchet distributed random variables. Let $Z_i = \frac{Y_i}{X_i}$ for $1 \leq i \leq n$ and $Z_{(1)} \leq Z_{(2)} \leq \cdots \leq Z_{(n)}$ be their order statistics. We extend (2.1) to a general class of quotient correlation coefficients based on the $(n-i+1)$th largest ratio and the jth smallest ratio:

$$q_n(i,j) = \frac{Z_{(n-i+1)} + Z_{(j)}^{-1} - 2}{Z_{(n-i+1)} \times Z_{(j)}^{-1} - 1}.$$

We note that the asymptotic independence of order statistics was first established in [19]. Their results are very general. In Section 8, we re-state the asymptotic independence of order statistics in two lemmas (Lemmas 8.1 and 8.2) in which certain specific marginal distributions are assumed, and the limit distributions of the order statistics are characterized. Then for any fixed (i,j), from Lemma 8.2, the limiting distribution of $nq_n(i,j)$ can be obtained. We state the results in the following theorem whose proof is straightforward using Lemmas 8.1 and 8.2.

Theorem 2.1. *Suppose $\{X_i, Y_i\}$, $i \geq 1$ are iid unit Fréchet distributed random variables. Then*

$$nq_n(i,j) \xrightarrow{\mathcal{L}} \zeta,$$

where ζ is a random variable with the gamma($i+j$,1) density given by $((i+j-1)!)^{-1}x^{i+j-1}e^{-x}$, $x>0$.

This theorem is very useful in constructing a robust statistical measure based on observed values. For example, we can also consider the limiting distribution for any weighted sum of $q_n(i,j)$ for $1 \leq i,j \leq k$, where $k \geq 1$ is a fixed integer. However, this is beyond the scope of the current paper, and we shall consider the choice of i and j and the weights as a future research project. Throughout the paper, we will deal with the case of $k=1$, and the results can be generated to cases with an arbitrary k.

2.3. Asymptotic independence of Pearson's correlation coefficient and the quotient correlation coefficient

We first prove that under some mild conditions the sample mean vector of a random sample of a random vector $\boldsymbol{\xi}$ is asymptotically independent of a sample vector of componentwise maxima of a random vector $\boldsymbol{\eta}$ which is dependent of $\boldsymbol{\xi}$.

Lemma 2.2. *Let $\{(\boldsymbol{\xi}_j, \boldsymbol{\eta}_j),\ \ 1 \leq j \leq n\}$ be a sequence of independent random vectors having the same distribution as that of vector $(\boldsymbol{\xi}, \boldsymbol{\eta})$ which are defined on the same probability space, $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_p)$ is a random vector with finite covariance matrix Σ_ξ, i.e. all elements in Σ_ξ are finite, and $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_q)$ is a random vector which belongs to the max-domain of attraction of a multivariate extreme value distribution $G(x_1, \ldots, x_q)$, i.e.*

$$\lim_{n\to\infty}[P(\eta_1 \leq a_{1n}x_1 + b_{1n}, \ldots, \eta_q \leq a_{qn}x_q + b_{qn})]^n = G(x_1, \ldots, x_q) \text{ for } \mathbf{x} \in \mathbb{R}^{\mathbf{q}} \tag{2.3}$$

where a_{in} are positive normalizing constants, $\mathbf{x} = (x_1, \ldots, x_q)$.

Set $\boldsymbol{\xi}_i = (\xi_{1i}, \xi_{2i}, \ldots, \xi_{pi}),\ i = 1, \ldots, n,\ \boldsymbol{\eta}_i = (\eta_{1i}, \eta_{2i}, \ldots, \eta_{qi}),\ i = 1, \ldots, n$, and write $\boldsymbol{\mu} = E(\boldsymbol{\xi})$, $\bar{\boldsymbol{\xi}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\xi}_i = \frac{1}{n}(\sum_{i=1}^n \xi_{1i}, \sum_{i=1}^n \xi_{2i}, \ldots, \sum_{i=1}^n \xi_{pi}) = (\bar{\xi}_1, \bar{\xi}_2, \ldots, \bar{\xi}_p)$, and $M_{\boldsymbol{\eta}} = (\max_{1\leq i\leq n}\eta_{1i}, \ldots, \max_{1\leq i\leq n}\eta_{qi})$. Then no matter how random variables between $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are dependent on each other, $\bar{\boldsymbol{\xi}}$ and $M_{\boldsymbol{\eta}}$ are asymptotically independent, i.e.

$$P\left(\sqrt{n}(\bar{\boldsymbol{\xi}} - \boldsymbol{\mu}) \leq \mathbf{y}, \qquad \frac{\max_{1\leq i\leq n}\eta_{ki} - b_{kn}}{a_{kn}} \leq x_k,\ k = 1, \ldots, q\right) \to \Phi(\Sigma_\xi^{-1/2}\mathbf{y}^\tau)G(x_1, \ldots, x_q) \tag{2.4}$$

as $n \to \infty$, where Φ is the standard multivariate normal distribution function, p and q are fixed integer numbers.

Suppose now that $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3, \xi_4, \xi_5)$, where $\xi_3 = \xi_1^2$, $\xi_4 = \xi_2^2$, $\xi_5 = \xi_1\xi_2$; $\boldsymbol{\eta} = (\eta_1, \eta_2)$, where $\eta_1 = g_1(\xi_1)$ and $\eta_2 = g_2(\xi_2)$ are unit Fréchet random variables. The sample Pearson correlation coefficient of ξ_1 and ξ_2 is given by

$$r_n^{(\xi_1,\xi_2)} = \frac{\bar{\xi}_5 - \bar{\xi}_1\bar{\xi}_2}{\sqrt{(\bar{\xi}_3 - \bar{\xi}_1^2)(\bar{\xi}_4 - \bar{\xi}_2^2)}}. \tag{2.5}$$

Then we have the following theorem.

Theorem 2.3. *With the established notations, suppose* $\boldsymbol{\xi}$ *and* $\boldsymbol{\eta}$ *satisfy the conditions in Lemma 2.2. Then*

$$P\left(\sqrt{n}(r_n^{(\xi_1,\xi_2)} - r) \le z, \frac{\max_{1\le i\le n}\eta_{ki} - b_{kn}}{a_{kn}} \le x_k,\ k = 1,2\right) \to \Phi\left(\frac{z}{\sigma}\right)G(x_1,x_2) \tag{2.6}$$

as $n \to \infty$, *where* r *is Pearson correlation coefficient of* ξ_1 *and* ξ_2, *and* $\mu_1 = E(\xi_1)$, $\mu_2 = E(\xi_2)$, $\sigma_1 = \sqrt{Var(\xi_1)}$, $\sigma_2 = \sqrt{Var(\xi_2)}$,

$$\sigma^2 = \left(1 + \frac{r^2}{2}\right)E\left(\frac{\xi_1-\mu_1}{\sigma_1}\right)^2\left(\frac{\xi_2-\mu_2}{\sigma_2}\right)^2 + \frac{r^2}{4}E\left(\left(\frac{\xi_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{\xi_2-\mu_2}{\sigma_2}\right)^2\right)^2 - rE\left(\frac{\xi_1-\mu_1}{\sigma_1}\right)\left(\frac{\xi_2-\mu_2}{\sigma_2}\right)\left(\left(\frac{\xi_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{\xi_2-\mu_2}{\sigma_2}\right)^2\right),$$

and Φ *is the standard normal distribution function.*

Furthermore, when ξ_1 *and* ξ_2 *are independent, and denote* $X_i = \eta_{1i}$, $Y_i = \eta_{2i}$ *in (2.1). Then we have*

$$\lim_{n\to\infty} P(\sqrt{n}r_n^{(\xi_1,\xi_2)} \le z,\ nq_n \le x) = \Phi\left(\frac{z}{\sigma}\right)F_g(x) \tag{2.7}$$

with $F_g(x)$ *is a gamma(2,1) distribution function.*

3. Integrating two correlation coefficients

The asymptotic independence of r_n (the same as $r_n^{(\xi_1,\xi_2)}$ throughout the paper) and q_n reveals that these two sample based correlation coefficients measure completely different dependencies between two random variables. We can see that the most popular sample based Pearson's correlation coefficient defined by (2.5) is mainly determined by the center part of data. It is an antithetical and complementary measure of the quotient correlation, which is a very important and desirable property to satisfy when one combines measures together. A natural choice of the combination is simply taking the maximum of the two correlation coefficients. We note that Pearson's correlation coefficient requires each marginal distribution having a finite second moment. Hence we may need to perform scale transformation in calculating r_n. In the rest of the paper, we will first perform marginal transformation of $\{(X_i,\ Y_i),\ i = 1,\dots,n\}$ to

(i) unit Fréchet scale when q_n is implemented;
(ii) pre-specified marginal with finite second moment, for example, a standard normal distribution when r_n is implemented;

Such transformation is straightforward if the marginal distributions of X and Y are known. In Section 4, a distribution free scale transformation procedure is proposed to deal with the case when the marginal distributions are unknown.

3.1. How to combine two correlation coefficients: A view from testing hypotheses

Suppose two unit Fréchet random variables X and Y satisfy the following properties:

$$P(X/Y < z) = \left(\frac{z}{1+z}\right)^{1+\lambda}, \ z > 0, \ 0 \le \lambda < 1, \tag{3.1}$$

$$\operatorname{corr}\big(\Phi^{-1}(e^{-1/X}), \Phi^{-1}(e^{-1/Y})\big) = \rho. \tag{3.2}$$

Then q_n can be used to test the following hypotheses:

$$H_0^\lambda: \ \lambda = 0 \ versus \ H_1^\lambda: \ \lambda > 0, \tag{3.3}$$

and r_n can be used to test the following hypotheses:

$$H_0^\rho: \ \rho = 0 \ versus \ H_1^\rho: \ \rho \neq 0. \tag{3.4}$$

Notice that H_0 of (2.2) implies both H_0^λ and H_0^ρ, but not reversely. There is no direct connection between H_0^λ and H_0^ρ. It is not known yet that whether H_0^λ can imply H_0. Nevertheless, if we simply use (3.3) or (3.4) to make inferences about (2.2), the conclusion may not be optimal. A better inference problem is to test

$$H_0^c: \ \max(\lambda, |\rho|) = 0 \ versus \ H_1^c: \ \max(\lambda, |\rho|) > 0. \tag{3.5}$$

This motivates us to consider a maximal variable association measure by just simply taking the maximum of two measures. We note that this idea is not something new at all in the literature. One may argue that if we can take the maximum of two measures, why not taking the maximum of as many measures as possible. Theoretically, when we deal with sample based measures, taking the maximum of many measures is not a good strategy since the limiting distribution under H_0^c can not be characterized even with the maximum of two coefficients if they are not appropriately chosen, not to mention the derivations under the alternatives. We aim to develop such a combination under which a limit distribution is characterized under H_0, and a limit distribution and asymptotic powers are also derived under local alternatives.

We also note that (3.3) is only one of applications of q_n, and we shall not consider that q_n is derived for a family of distributions satisfying (3.1) only.

3.2. The maximum of two transformed and asymptotically independent correlation coefficients

Under H_0 of (2.2), the quotient coefficient q_n is shown to be n convergence ([30]), and it is well known that r_n is squared root n convergence. Therefore, q_n and

$|r_n|$ are not at the same scales, and q_n is more likely to be smaller than $|r_n|$ for a finite sample size. If we simply use the maximum of these two coefficients as a new test statistic, q_n may not be useful at all. To overcome this impediment, we define a new combined variable association measure:

$$C_n = \max(\mathrm{q}_n^{1/2},\ |r_n|). \tag{3.6}$$

When two random variables follow a bivariate normal distribution with the correlation coefficient $|\rho| < 1$, we have C_n tending to $|\rho|$ almost surely. In many cases of two random variables being linearly dependent, we shall have C_n tending to linear correlation coefficient $|\rho|$ almost surely, i.e. for sufficiently large n, C_n equals $|r_n|$. In the cases of $\rho = 0$, the limit of C_n is the same as the limit of $\mathrm{q}_n^{1/2}$. In the case of co-monotone or positive dependent, the limit of C_n can be either from $\mathrm{q}_n^{1/2}$ or from $|r_n|$ depending on which one is in dominance. Therefore, for a finite sample size n, C_n can be regarded as a nonlinear dependence measure. Here we shall regard 'linear' as a special case of 'nonlinear'. An extended and practically useful nonlinear dependence measure is defined by (6.1) in Section 6.1 and is used in our real data analysis.

Notice that in the definition of the new association measure, we have chosen not to perform a power transformation of Pearson's correlation coefficient r_n, i.e. not to use $|r_n|^2$. The reason for this construction is two folds. First, a linear correlation is easy to implement and interpret, and it serves as the first association measure to compute in practice. It is a bench mark in checking variable associations. Second, there does not exist a commonly used nonlinear dependence measure which is as popular as Pearson's correlation coefficient.

3.3. The limiting distributions of C_n under H_0

Under H_0 defined in (2.2), we have the following proposition:

Proposition 3.1. *Under H_0, the limiting distributions of C_n have the following forms:*

$$nC_n^2 \xrightarrow{\mathcal{L}} \max\left(\chi_1^2, \frac{1}{2}\chi_4^2\right),$$

where χ_1^2 and χ_4^2 are two independent χ^2 random variables with 1 and 4 degrees of freedom, respectively. We denote this limit random variable as $\chi^2_{\max}$.

In Section 4 and the Appendix section, we will show that the limiting distribution of q_n is the same as a χ_4^2 when q_n is derived based on nonparametric marginal transformations. This is an important property of the quotient correlation coefficient. In this paper, we propose a test statistic mainly based on Proposition 3.1, which controls empirical Type I error probabilities being less than their pre-specified nominal levels and still gives high detecting powers in testing independence between two dependent random variables.

The limiting distribution of nC_n^2 leads to the following testing procedure (a new max linear χ^2 test):

- If $nC_n^2 > \chi^2_{\max,\alpha}$, H_0 of (2.2) is rejected; otherwise it is retained. Here $\chi^2_{\max,\alpha}$ is the upper α percentile of the random variable $\max(\chi_1^2, \frac{1}{2}\chi_4^2)$.

We note that the new max linear χ^2 test achieves asymptotic power 100% when two random variables under the alternative hypothesis are dependent and either r_n or q_n has a nonzero limit. When both r_n^2 and q_n tend to zero with a slower rate than n, the new max linear χ^2 test still achieves asymptotic power 100%.

It is worth noting that for example q_n is $\sqrt{n}$ convergence when X_i and Y_i follow a bivariate Gumbel copula distribution, see [28] for the derivation, and hence the max linear χ^2 test will have asymptotic power one under the alternative hypothesis of Gumbel copula dependence.

3.4. Asymptotic powers under local alternatives

It can easily be shown that under any fixed alternative hypothesis H_1^λ with $\lambda = \lambda_1 > 0$, nq_n is an asymptotic power one test, and hence nC_n^2 is also an asymptotic power one test under any fixed alternative hypothesis H_1^c. We now study the asymptotic powers under a class of local alternatives:

$$\begin{aligned} H_0^c: \quad & \lambda = \lambda_0 = 0, \rho = \rho_0 = 0 \\ & \textit{versus } H_1^{cLn}: \ \rho = \rho_n = \Delta/\sqrt{n}, \lambda = \lambda_n = \Delta/\sqrt{n}, \ \Delta > 0. \end{aligned} \tag{3.7}$$

Note that the distribution function of random variable $\max(\chi_1^2, \frac{1}{2}\chi_4^2)$ is given by

$$F_{\max}(x) = 2\left(\Phi(\sqrt{x}) - \frac{1}{2}\right)\left(1 - e^{-x} - \frac{1}{2}xe^{-x}\right), \quad x \geq 0,$$

where Φ is the standard normal distribution function.

Denote σ in Theorem 2.3 under H_0 and H_1^{cLn} as σ_0 and σ_n respectively, and let $u_\alpha = \chi^2_{\max,\alpha}$ which is the α level critical value of $F_{\max}(x)$. Assuming nq_n under the alternative has the same limit as it has under H_0, then under H_1^{cLn}, the asymptotic power is

$$\begin{aligned} \beta_n &= P_n(nC_n^2 > u_\alpha) = 1 - P_n(nq_n < u_\alpha, \ nr_n^2 < u_\alpha) \\ &= 1 - P_n\left(nq_n < u_\alpha, \ -\frac{\sqrt{u_\alpha}}{\sigma_0} < \frac{\sqrt{n}(r_n - \rho_n)}{\sigma_n}\frac{\sigma_n}{\sigma_0} + \frac{\sqrt{n}(\rho_n - \rho_0)}{\sigma_n} < \frac{\sqrt{u_\alpha}}{\sigma_0}\right) \\ &\to 1 - \left[\Phi\left(-\frac{\Delta}{\sigma_0} + \frac{\sqrt{u_\alpha}}{\sigma_0}\right) - \Phi\left(-\frac{\Delta}{\sigma_0} - \frac{\sqrt{u_\alpha}}{\sigma_0}\right)\right]\left(1 - e^{-u_\alpha} - \frac{1}{2}u_\alpha e^{-u_\alpha}\right). \end{aligned}$$

It is easy to see that if we fix the cutoff value as u_α, then the asymptotic power of using test statistic nr_n^2 alone is $1-[\Phi(-\frac{\Delta}{\sigma_0}+\frac{\sqrt{u_\alpha}}{\sigma_0})-\Phi(-\frac{\Delta}{\sigma_0}-\frac{\sqrt{u_\alpha}}{\sigma_0})]$, while the asymptotic power of using test statistic nq_n alone is $1-(1-e^{-u_\alpha}-\frac{1}{2}u_\alpha e^{-u_\alpha})$. As a result, the combination of the two test statistics increases the asymptotic detecting power.

3.5. Incorporating Hoeffding's D statistic

There are some unusual situations that both r_n^2 and q_n tend to zeros with the rate n, and hence nC_n^2 as a test statistic does not have detecting power under such unusual situations of alternative hypothesis of dependence. We propose to include Hoeffding's D statistic in our proposed maximal dependence measure as:

$$C_n^\delta = \max(\mathrm{q}_n^{1/2},\ |r_n|,\ |D|^{1+\delta}), \tag{3.8}$$

where δ is a small positive number, say $\delta = 10^{-16}$, and D is defined as follows ([16]). Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a population with the d.f. $F(x, y)$, $n \geq 5$, and let

$$D = D_n = \frac{1}{n(n-1)\cdots(n-4)} \Sigma'' \phi(X_{\alpha_1}, Y_{\alpha_1}; \ldots; X_{\alpha_5}, Y_{\alpha_5}),$$

where Σ'' denotes summation over all α such that

$$\alpha_i = 1, \ldots, n;\ \alpha_i \neq \alpha_j \text{ if } i \neq j,\ (i, j = 1, \ldots, 5),$$

$$\phi(x_1, y_1; \ldots; x_5, y_5) = \frac{1}{4}\psi(x_1, x_2, x_3)\psi(x_1, x_4, x_5)\psi(y_1, y_2, y_3)\psi(y_1, y_4, y_5)$$

and

$$\psi(x, y, z) = I_{(x \geq y)} - I_{(x \geq z)}$$

where $I_{(x\geq y)}$ is an indicator function. It is shown in [16] that $-\frac{1}{60} \leq D_n \leq \frac{1}{30}$. We have $P(\chi^2_{\max} < 1/30) = 7.8716 \times 10^{-5}$, which tells that under H_0, the value of D is almost negligible. Under H_0 and with $\delta > 0$, $n|D_n|^{1+\delta}$ converges to 0 in probability, and hence C_n^δ has the same limit distribution as C_n has. We still call nC_n^δ as the new max linear χ^2 test statistic. Under the alternative hypothesis, D has a positive limit, and hence nC_n^δ is an asymptotic power one test.

The use of D as a nonparametric test (HBKR test) of independence has been well-known due to [16, 3]. The HBKR test has been embedded in many statistical softwares such as SAS and R. We note that D statistic uses only rank information of the data, and the magnitudes of the data are ignored. As a result, D statistic may not be as powerful as a statistic which uses both ranks and magnitudes. We aim to develop such a test statistic, and our simulation results show that nC_n^δ has a better performance, and the computation of nC_n^δ is very simple and fast.

4. Distribution free approaches for unknown marginals

4.1. Marginal transformation via rank-preserving scale regeneration

In the previous discussions, we assume that $\{(X_i,\ Y_i),\ i = 1, \ldots, n\}$ can be transformed to unit Fréchet scale and any other margins with finite second moment. Such a transformation is straightforward if the marginal distributions of X and Y are known. In practical applications, however, we often face the case that the marginal distributions are unknown. In this section, we propose a rank-preserving scale regeneration marginal transformation procedure to overcome this impediment.

We first generate two unit Fréchet samples of size n and denote $Z^X_{(1)} < Z^X_{(2)} < \cdots < Z^X_{(n)}$, $Z^Y_{(1)} < Z^Y_{(2)} < \cdots < Z^Y_{(n)}$ as the ordered samples respectively, and the rank-preserving scale regeneration marginal transformation procedure results in the following transformed data

$$X_i^* = Z^X_{(\text{rank}[X_i])},\ Y_i^* = Z^Y_{(\text{rank}[Y_i])},\ i = 1, \dots, n. \tag{4.1}$$

Note that with this procedure, the ranks of the original data are preserved, and the transformed scales have the desired distribution.
The rank-preserving scale regeneration quotient correlation is defined as

$$\mathrm{q}_n^{\mathfrak{R}} = \frac{\max_{i\le n}\{Y_i^*/X_i^*\} + \max_{i\le n}\{X_i^*/Y_i^*\} - 2}{\max_{i\le n}\{Y_i^*/X_i^*\} \times \max_{i\le n}\{X_i^*/Y_i^*\} - 1}. \tag{4.2}$$

One immediate advantage of this definition is that the transformation does not depend on any correction term (for example $1/n$ as in an empirical distribution function based transformation). The transformation procedure gives advantages in applying extreme value theory. The following Lemma 4.1 and Theorem 4.2 are evidences. We note that the simulation procedure here is similar to the procedure in [30] which uses one simulated sample in two populations X and Y. Here we simulate one sample for each population. We have the following important lemma whose proof is given in Section 8.

Lemma 4.1. *Suppose X and Y are independent absolutely continuous random variables (not necessarily being unit Fréchet), $\{(X_i,\ Y_i),\ i = 1, \dots, n\}$ is a random sample from (X, Y), and $\{(X_i',\ Y_i'),\ i = 1, \dots, n\}$ is a different random sample from (X, Y). Denote $Z^X_{(1)} < Z^X_{(2)} < \cdots < Z^X_{(n)}$, $Z^Y_{(1)} < Z^Y_{(2)} < \cdots < Z^Y_{(n)}$ as the ordered samples of $\{(X_i',\ Y_i'),\ i = 1, \dots, n\}$ respectively. Define*

$$X_i^* = Z^X_{(rank[X_i])},\ Y_i^* = Z^Y_{(rank[Y_i])},\ i = 1, \dots, n.$$

Then $\{(X_i^*,\ Y_i^*),\ i = 1, \dots, n\}$ is a random sample from (X, Y).

Theorem 4.2. *Suppose X and Y are independent continuous random variables (not necessarily being unit Fréchet), and $\{(X_i,\ Y_i),\ i = 1, \dots, n\}$ is a random sample from (X, Y). Then $\mathrm{q}_n^{\mathfrak{R}}$ is asymptotically χ^2 distributed - i.e., $2a_n q_n^{\mathfrak{R}} \xrightarrow{\mathcal{L}} \chi^2_4$, as $a_n/n \to 1,\ n \to \infty$, where χ^2_4 is a χ^2 random variable with 4 degrees of freedom.*

A rigorous proof of Theorem 4.2 is given in Section 8. We note that Theorem 3.2 in [30] was proved under a stronger condition than necessary. The proof of the above theorem does not need to satisfy that condition.

We find that the form (4.2) is easy to implement and the simulation results are very close to q_n which is based on a parametric transformation. Theorem 4.2 delivers the following testing procedure:

- If $n(\max\{(\mathrm{q}_n^{\mathfrak{R}})^{1/2},\ |r_n^{\mathfrak{R}}|,\ |D|^{1+\delta}\})^2 > \chi^2_{\max,\alpha}$, H_0 of (2.2) is rejected; otherwise it is retained. Here $r_n^{\mathfrak{R}}$ is computed using (2.5) based on the transformed data (X_i^*, Y_i^*), $i = 1, 2, \ldots, n$ in (4.1) and their corresponding marginally transformed normal values.

One can immediately see that Type I error probabilities of the above defined test are the same as the testing procedure introduced in Section 3.3. One can also see that asymptotic powers of the above defined test under the alternative hypothesis are equal to asymptotic powers of the testing procedure introduced in Section 3.3.

4.2. The rank-preserving scale regeneration and conditional expectation test

When the procedure (4.1) is used, one may argue that a different variable transformation may result in a different conclusion since the test is based on simulated unit Fréchet random variates. This issue can be resolved by performing a large number of max linear χ^2 tests and making a convincing conclusion based on rates of rejecting the null hypothesis.

Our testing procedure is:

- For a given random sample $\{(X_i,\ Y_i),\ i = 1, \ldots, n\}$ of (X, Y), we repeat (4.1) N times, i.e. for each $j \in \{1, \ldots, N\}$, we generate two unit Fréchet samples of size n and denote $Z^X_{j(1)} < Z^X_{j(2)} < \cdots < Z^X_{j(n)}$, $Z^Y_{j(1)} < Z^Y_{j(2)} < \cdots < Z^Y_{j(n)}$ as the ordered samples respectively, and the rank-preserving scale regeneration marginal transformation procedure results in the following transformed data

$$X_{i,j} = Z^X_{j(\mathrm{rank}[X_i])},\ Y_{i,j} = Z^Y_{j(\mathrm{rank}[Y_i])},\ i = 1, \ldots, n,\ j = 1, \ldots, N. \quad (4.3)$$

  For each j, the correlation coefficient $r_{nj}^{\mathfrak{R}}$ and the simulation based quotient $\mathrm{q}_{nj}^{\mathfrak{R}}$ are computed using (2.7) and (4.2) based on the transformed data $(X_{i,j}, Y_{i,j})$, $i = 1, 2, \ldots, n$, $j = 1, \ldots, N$, and their corresponding marginally transformed normal values. Define p-values $\tilde{P}_{nj}^{\mathfrak{R}} = 1 - F_{\max}(n(\max\{(\mathrm{q}_{nj}^{\mathfrak{R}})^{1/2},\ |r_{nj}^{\mathfrak{R}}|,\ |D|^{1+\delta}\})^2)$.
- The decision rule is:
  - If $T_N = \frac{1}{N}\sum_{j=1}^N \tilde{P}_{nj}^{\mathfrak{R}} < \tilde{p}_\alpha$, we reject the null hypothesis of independence. The cut-off value $\tilde{p}_\alpha$ is derived from an approximation theorem presented below.

Theorem 4.3. *With the established notations, suppose $\{(X_i,\ Y_i),\ i = 1, \ldots, n\}$ is a random sample of (X, Y) with marginals being unit Fréchet, q_n and r_n are defined in (2.1) and (2.5) respectively. Then*

$$T_N \xrightarrow{a.s.} E(\tilde{P}_{n1}^{\mathfrak{R}} | \mathcal{G}) \qquad \text{as } N \to \infty$$

where $\mathcal{G}$ is the σ-algebra generated by the random vectors (X_j, Y_j), $j \geq 1$.

We want $P(T_N < \tilde{p}_\alpha) \leq \alpha$, or approximately $P\{E(\tilde{P}^{\mathfrak{R}}_{n1}|\mathcal{G}) < \tilde{p}_\alpha\} \leq \alpha$. The distribution of $E(\tilde{P}^{\mathfrak{R}}_{n1}|\mathcal{G})$ is generally unknown but is distribution-free under the null hypothesis. We obtain the cut-off value of $\tilde{p}_\alpha$ using Monte Carlo method in which we obtain 50,000 p-values. We present the following table for a set of α values which may satisfy the need of the most applications.

α	.001	.005	.01	.025	.05	.1
$\tilde{p}_\alpha$	0.0036	0.0203	0.0415	0.0956	0.1663	0.2567
$\tilde{p}_\alpha/\alpha$	3.6000	4.0600	4.1500	3.8240	3.3260	2.5670

In practice, we compare the average p-value T_N with the cut-off values $\tilde{p}_\alpha$ and then make inferences. Or equivalently, we can compare $\tilde{\tilde{p}} = T_N/(\tilde{p}_\alpha/\alpha)$ with α directly. In this case, we call $\tilde{\tilde{p}}$ a **r**ank-preserving sc**a**le regeneration and **c**onditional **e**xpectation (R.A.C.E) test p-value. We argue that this method should provide more information in testing the hypothesis of independence and make a better and accurate conclusion.

4.3. An integrated testing procedure

We now summarize our proposed test procedures into an integrated one:

- Specify the significant level α, and $\delta = 10^{-16}$;
- When marginal distributions are known,
  - Transform data to unit Fréchet scales and normal scales respectively, and compute q_n using (2.1) and r_n using (2.5);
  - If $n[C_n^\delta]^2 > \chi^2_{\max,\alpha}$, H_0 of (2.2) is rejected; otherwise it is retained.
- When marginal distributions are unknown,
  - Generate two unit Fréchet samples of size n and denote $Z^X_{(1)} < Z^X_{(2)} < \cdots < Z^X_{(n)}$, $Z^Y_{(1)} < Z^Y_{(2)} < \cdots < Z^Y_{(n)}$ as the ordered samples respectively;
  - Compute $\mathrm{q}^{\mathfrak{R}}_n$ using (4.2) and $r^{\mathfrak{R}}_n$ based on marginally transformed normal data;
  - We repeat the above two steps N times, denote $\mathrm{q}^{\mathfrak{R}}_n$ and $r^{\mathfrak{R}}_n$ as $\mathrm{q}^{\mathfrak{R}}_{nj}$ and $r^{\mathfrak{R}}_{nj}$, $j = 1, \ldots, N$ respectively, and then compute
  $$\tilde{P}^{\mathfrak{R}}_{nj} = 1 - F_{\max}(n(\max\{(\mathrm{q}^{\mathfrak{R}}_{nj})^{1/2},\ |r^{\mathfrak{R}}_{nj}|,\ |D|^{1+\delta}\})^2),\ j = 1, \ldots, N;$$
  - If $T_N = \frac{1}{N}\sum_{j=1}^N \tilde{P}^{\mathfrak{R}}_{nj} < \tilde{p}_\alpha$, we reject the null hypothesis of independence. The cut-off value $\tilde{p}_\alpha$ is derived from Theorem 4.3.

Empirical Type I error probabilities based on the above procedure and finite sample sizes are illustrated in next section. Empirical detecting powers of simulation examples will be given in Section 5.

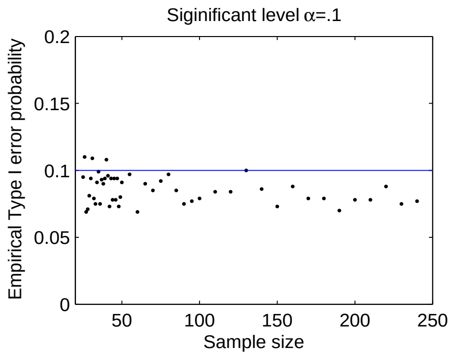

FIG 1. *Empirical Type I error probabilities. The left panel is for a test with a significant level $\alpha = .1$, and the right panel is for a test with a significant level $\alpha = .05$. Simulations are implemented in Matlab® 7.5 installed in Red Hat Enterprise Linux 5.*

4.4. Empirical Type I error probabilities

The empirical Type I errors at levels .1 and .05 are demonstrated in Figure 1. Simulation sample sizes are $n = 25, 26, \ldots, 49, 50, 55, 60, \ldots, 95, 100, 110, 120, \ldots, 230, 240$. For each fixed sample size, we run the new χ^2 test 1000 times, then compute the empirical Type I error proportions. Simulations are implemented in Matlab® 7.5 installed in Red Hat Enterprise Linux 5. One can see that the proposed χ^2 test controls Type I error probabilities within their nominal levels. The overall average empirical test level on the left panel is .0854, and the overall average empirical test level on the right panel is .0395. They are overall slightly conservative Type I errors.

5. Simulation examples

In this section, we use the following simulation procedure:

(1) Simulate a bivariate random sample from a pre-specified joint distribution using Matlab® 7.5 installed in Red Hat Enterprise Linux 5.
(2) For each simulated univariate sequence, we use the procedure (4.1) to transform the simulated data into unit Fréchet scales.
(3) Use distribution functions to convert the transformed unit Fréchet values into normal scales.
(4) Use these transformed values to perform the new integrated tests from Section 4.3.

For comparison purpose, we also conduct Fisher's Z-transformation test, the gamma test introduced in [30].

In Figures 2, 3, 4, the blue-colored-(circled)-dotted-dashed lines, the red-colored-(squared)-dashed lines, the dark-(diamond)-solid lines, the magenta-colored-(plus)-dashed lines, the green-colored-(cross)-dotted-dashed lines are

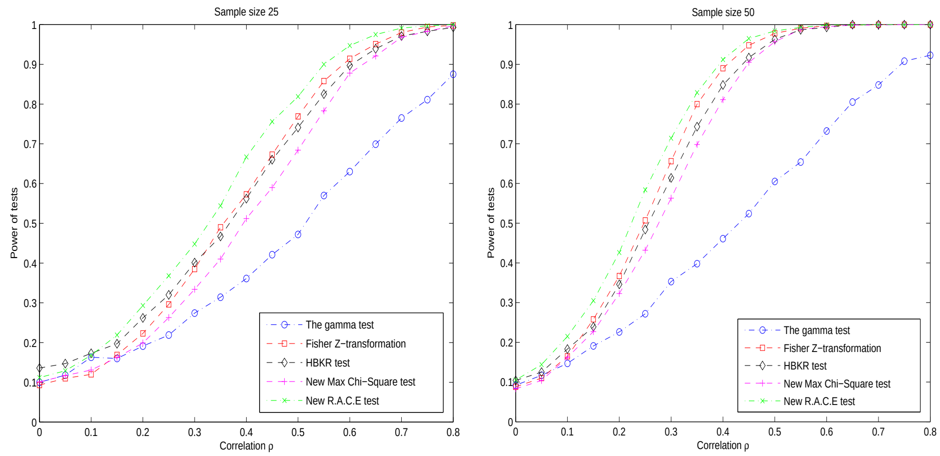

FIG 2. *Comparisons among the gamma test, Fisher's Z-transformation test (see [13, 30] for the definition of the test), the HBKR test, the new max linear χ^2 test, and the new R.A.C.E test. The left panel is for tests with sample sizes of 25. The right panel is for tests with sample size of 50.*

drawn from empirical powers calculated based on the gamma test, Fisher's Z-transformation test, HBKR test, the new max (linear) χ^2 test, the new R.A.C.E test introduced in Subsection 4.2 respectively.

Example 5.1. In this example, we simulate bivariate normal samples with correlation coefficients ranging from 0 to 0.8 with an increment 0.05 and fixed sample sizes 25 and 50 respectively. At level $\alpha = 0.10$, empirical detecting powers are calculated based on 1000 repeated samples. Performances of the five tests are shown in Figure 2.

It is clear that when data is drawn from bivariate normal, arguably Fisher's Z-transformation test should perform better than any other tests which are not constructed from the correlation coefficients. We see that the new max linear χ^2 test in this example performs as good as Fisher's Z-transformation test, the R.A.C.E test and HBKR test also perform as good as Fisher's Z-transformation test. Their corresponding power curves are complected with each other. In this example, we see that the gamma test is not as powerful as the other four tests. In other words, under the alternative hypothesis of dependence the combined test statistic (the max linear new χ^2 test statistic) boosts the power of the gamma test to as high as the most powerful test of Fisher's Z-transformation test.

Example 5.2. In this example, we simulate bivariate random samples from (X, Y), where $X = U * B + (1 + V/1000) * (1 - B)$, $Y = W * B + (1 + V/1000) * (1 - B)$, U, V, W are independent uniform random variables over [0,1], B is a Bernoulli random variable with success probability .95. Sample sizes are 25, 26, $\ldots$, 240. Empirical powers for all tests are plotted in Figure 3.

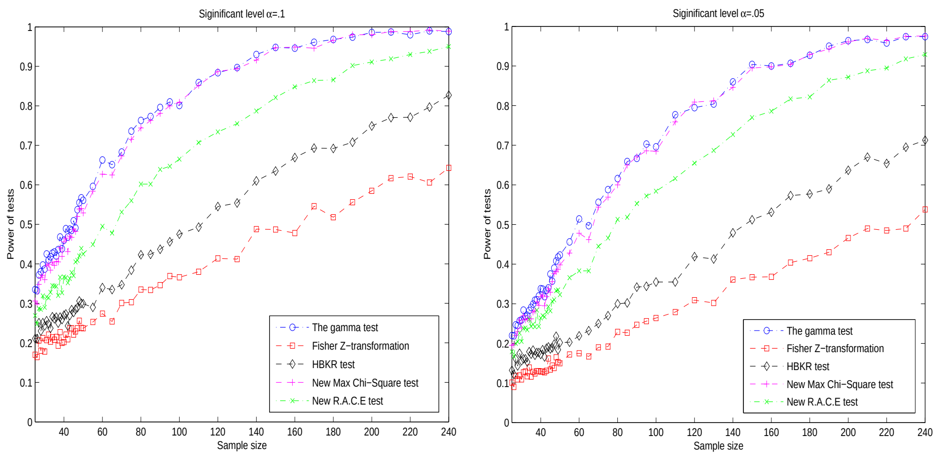

FIG 3. *Empirical powers of five tests. All tests are at level $\alpha = .1$ (left panel) and $\alpha = .05$ (right panel).*

From Figure 3, we see that the most powerful test is the gamma test in this example, and the performance of the new max linear χ^2 test is as good as the gamma test. Notice that the dependence between X and Y in this example is in extreme values, as a result the gamma test should have the best performance, see [30]. In other words, under the alternative hypothesis of dependence the combined test statistic (the new max linear χ^2 test statistic) boosts the power of Fisher's Z-transformation test statistic to as high as the most powerful test of the gamma test. Also notice that the R.A.C.E test was not as powerful as the gamma test and the new max linear χ^2 test in this example. This phenomenon suggests that when we transform data using (4.1), the original ranks of the data and how the data depends on each other affect the detecting powers. Nevertheless, it is still an acceptable approach in practice. We also make an important note here. One may concern that rank transformations kill extreme values and dependencies in data. This example shows that our proposed rank preserving simulation transformation retains dependencies in data, which is very important in practice.

In the next example, we demonstrate that the new max linear χ^2 test and the R.A.C.E test gain higher detecting power than the gamma test does. Due to the fact that the moments do not exist in this example, we note that many ordinary statistics may not be applicable to this example without performing certain variable transformation procedures such as a Box-Cox transformation or our newly proposed rank-preserving scale regeneration transformation.

Example 5.3. In this example, we simulate bivariate random samples from $(|X|, |Y|)$, where $Y = X * E + W$, and X is a standard Cauchy random variable, E and W are independent standard normal random variables. Sample sizes are 25,26,...,240. Empirical powers for all tests are plotted in Figure 4.

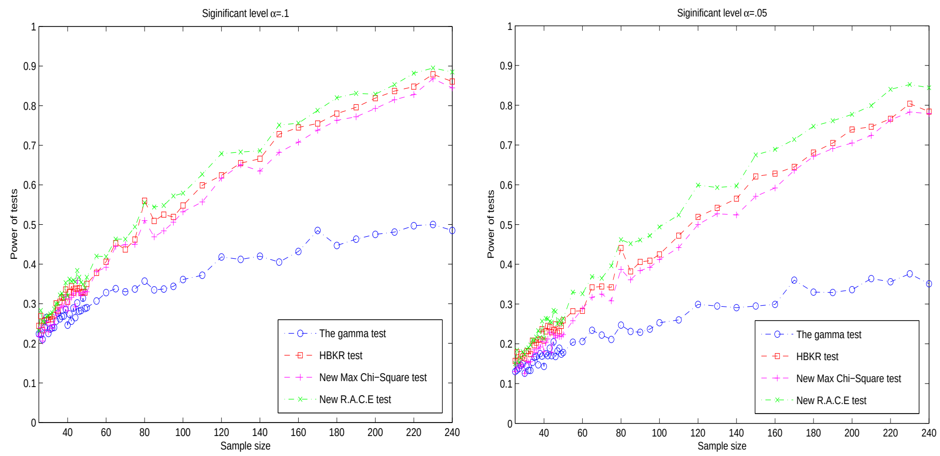

FIG 4. *Empirical powers of three tests. The significant levels in the left panel and in the right panel are $\alpha = .1$ and $\alpha = .05$ respectively.*

From Figure 4, under the alternative hypothesis of dependence the combined test statistic (the max linear χ^2 test statistic) boosts the powers of the gamma test to higher levels. Note that Fisher's Z-transformation test is not illustrated here since it is not directly applicable to original data.

These three examples clearly suggest that the new max linear χ^2 test is superior. We have extensively applied the new max linear χ^2 test to many other simulation examples and observed that the new χ^2 test outperforms existing tests. Based on these test results, it may be safe to say that the new max linear χ^2 test has a better performance in testing independence between two random variables. On the other hand, our newly proposed R.A.C.E test procedure also achieves high detecting powers in our simulated examples. In practice, the R.A.C.E test seems to be a more acceptable approach as it also suggests how the null hypothesis of independence is rejected or retained.

In the next section, we conduct a real data of tobacco markets analysis.

6. Real data analysis

We first review some basic data processing and variable transformation procedures. They are applied to our second new test procedure.

6.1. Measuring the maximum association: A practical perspective

In real data analysis, data are often transformed before computing various statistical measures. For example, in science and engineering, a log-log graph or a log-log correlation is often used in data analysis. In the data processing, there exist many transformation procedures in the literature, which satisfy different purposes. Popular transformations including Box-Cox transformation, logarithm transformation, absolute value transformation, sign transformation, etc. It is common to consider more than one transformation in a particular study. For this purpose, let $\mathbb{F}$ be a collection of candidate transformations, we extend (3.6) to be

$$C_n^* = \sup_{g_x, g_y \in \mathbb{F}} \max(\mathrm{q}_n^{1/2}(g_x, g_y),\ |r_n(g_x, g_y)|,\ |D(g_x, g_y)|^{1+\delta}), \tag{6.1}$$

where $\mathrm{q}_n(g_x, g_y) = \mathrm{q}_n^{\mathfrak{R}}$ in (4.2) with (X_i^*, Y_i^*) being obtained from

$$X_i^* = Z^X_{(\mathrm{rank}[g_x(X_i)])},\ Y_i^* = Z^Y_{(\mathrm{rank}[g_y(Y_i)])},\ i = 1, \ldots, n,$$

$\mathrm{rank}[g_x(X_i)]$ is obtained from transformed sample $\{g_x(X_1), g_x(X_2), \ldots, g_x(X_n)\}$, and $\mathrm{rank}[g_y(Y_i)]$ is defined similarly. $|r_n(g_x, g_y)|$ is calculated using (2.5) after transforming (X_i^*, Y_i^*) into normal scales by distribution transformation, so is $D(g_x, g_y)$.

We regard (6.1) as the $\mathbb{F}$ maximum association measure between two random variables. Of course, the choice of $\mathbb{F}$ is a practical issue, and one can choose transformation functions for any special purposes. In this paper, we consider the following family of six transformations which has very good performances in most of our real data analysis. (i) $X(i) = X$; (ii) $X(ii) = -X$; (iii) $X(iii) = |X|$; (iv) $X(iv) = -|X|$; (v) $X(v) = \big|\log\big(|X|/a_k\big)\big|$; (vi) $X(vi) = -\big|\log\big(|X|/a_k\big)\big|$, where a_k is the kth percentile of $|X|$.

Remark 1. When X is a positive random variable, (iii) is the same as (i), and (iv) is the same as (ii). The purpose of (ii) is to study the negative dependence between two random variables X and Y through $(X(i), Y(ii))$, or $(X(ii), Y(i))$. The purpose of (iii) is to study the dependence between values near zero in X (or respectively Y) and values in any part in Y (or respectively in X). The purpose of (v) is to study the dependence between values near the pth percentile of $|X|$ (or respectively of $|Y|$) and values in any other part in Y (or respectively in X). The rest of transformations can be interpreted accordingly. We have total 36 combinations of transformations, i.e. $(X(s),\ Y(t)), s, t, = i, ii, iii, iv, v, vi$. One can find an optimal choice of k. In our real data analysis, we will choose k from two sets of numbers $\{.05,\ .20,\ .50,\ .80,\ .95\}$ and $\{.10,\ .25,\ .30,\ .45,\ .60,\ .75\}$.

In the next subsection, we use the variable transformations (i), (ii), (v), and (vi) with the choice of k from each of the above two sets of numbers. The number of combinations of forming new bivariate transformed sample in $\{(X(s),\ Y(t)),\ s, t = i,\ ii,\ v,\ vi\}$ for each pair is 324.

Under the above variable transformations (or other $\mathbb{F}$), our refined R.A.C.E testing procedure is:

- For each transformed data $\{(X_i(s),\ Y_i(t)), i = 1, \ldots, n\}$ where $s, t \in \{i, ii, iii, iv, v, vi\}$, we repeat (4.1) N times, and hence we get N p-values. After taking the average T_N of the N p-values, we get the rank-preserving scale regeneration and conditional expectation (R.A.C.E) p-value $\tilde{\tilde{p}} = T_N/(\tilde{p}_\alpha/\alpha)$, where cut-off value $\tilde{p}_\alpha$ is given in Section 4.2. Then we keep the combination which has the smallest $\tilde{\tilde{p}}$-value.

- If the smallest $\tilde{\tilde{p}}$ value is less than α, we reject the null hypothesis of independence.

6.2. About the data

"Did You Know? Cigarettes kill about 4.8 million people every year around the world. This is expected to rise to 10 million per year by 2020." [25]. Such situation makes it a significant public health problem, and a social economic problem as well. Here, we are interested in exploring correlation measures among seven main tobacco consumption related variables. They are cigarette tax per pack in US dollar, FY 2007 cigarette pack sales in 100 millions, FY 2007 cigarette tax revenue in 100 million US dollars, retail price per pack with all taxes in US dollar, CDC state smoking costs per pack sold in US dollar, youth smoking rate, adult smoking rate. These variables are denoted as Tax, Sales, Revenue, Price, Costs, Youth, and Adult respectively in figures and tables. Our data source is [4]. Behavior Risk Factor Surveillance Survey (BRFSS); State Cigarette Taxes: http://www.tobaccofreekids.org/research/factsheets/pdf/0099.pdf. The data sample size is 51 (50 US states plus Washington D.C.), i.e. each state has one record in Year 2007. The paired scatter plots are demonstrated in Figure 5.

From Figure 5, we can see that only cigarette tax per pack and retail price per pack with all taxes show a clear strong linear correlation (also positive dependence and co-monotone as well). All other subplots either suggest a nonlinear correlation or no correlation. Calculated kurtosis and skewness for each variable

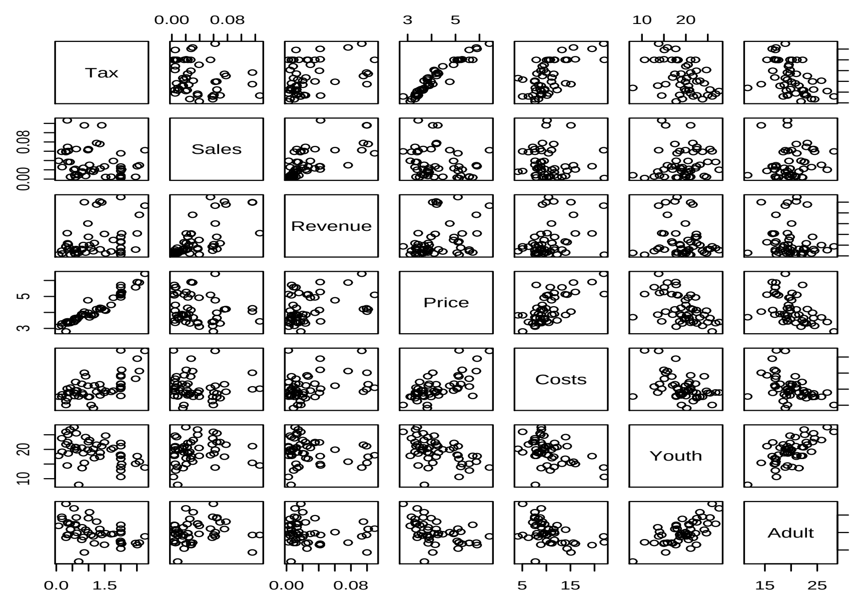

FIG 5. *Scatter plots for seven tobacco market variables.*

Table 1
This table reports p-values obtained when Fisher Z-transformation tests (upper triangle part) were applied to data drawn from Figure 5

P-value	Sales	Revenue	Price	Costs	Youth	Adult
Tax	0.0003	0.0000	0	0	0.0000	0.0000
Sales		0	0.0030	0.2654	0.0806	0.0137
Revenue			0.0000	0.0000	0.0197	0.0278
Price				0	0.0000	0.0000
Costs					0.0000	0.0000
Youth						0

show that all variables are not normally distributed; variables Sales, Revenues and Costs are showing fat tail behaviors, probably variable Adult too; variable Youth is skewed to the left, and all other variables are skewed to the right. Based on these observations, it may not be appropriate to calculate correlation coefficients and to perform Fisher's Z-transformation test since they are valid under assumptions of either a finite first moment or a finite second moment from each random variable. In other words, when using these calculated quantities, one should use them with caution.

In this analysis, one of our primary goals is to demonstrate how to use the newly combined measure, and the new χ^2 test. We still calculate correlation coefficients of variables and perform Fisher's Z-transformation test by using original data for comparison purpose.

In Table 1 (upper triangle part), we applied Fisher-Z transformation test to all pairs. We see that there are two pairs which may suggest independence or no correlation. This conclusion may be a misleading. For example, it may not be appropriate to suggest cigarette pack sales and Youth smoking rate are uncorrelated. One would argue that the higher the Youth smoking rate, the more sales. This is a clear indication that we need a more powerful test for testing independence between two random variables.

6.3. Testing hypothesis of independence via R.A.C.E test

For each combination of $(X(s),\ Y(t))$, we repeat (4.1) 100 times, and hence we get 100 p-values. For each paired random variables, we keep the combination which has the smallest $\tilde{\tilde{p}}$-value, i.e., the rank-preserving scale regeneration and conditional expectation p-value. Test results are reported in Tables 2 and 3.

In Table 2, [(s), (t)] stands for the kept combination, and two numbers in the bracket below the kept combination stand for kth percentiles used in transformations. For example, in the cell of Tax and Sales, observations in Tax are divided by its 75th sample percentile, then taking the logarithm transformation, and finally taking the absolute value transformation, while observations in Sales are divided by its 75th sample percentile, then taking the logarithm transformation, the absolute value transformation, and finally taking a negative sign transformation. In the cell of Price and Costs, observations in Price are kept as their original values. In the cell of Revenue and Youth, observations in

TABLE 2

This table reports percentiles and combinations used when the new χ^2 tests were applied to data drawn from Figure 5

	Sales	Revenue	Price	Costs	Youth	Adult
Tax	[(vi), (v)] [0.75, 0.75]	[(v), (v)] [0.2, 0.5]	[(v), (v)] [0.2, 0.05]	[(v), (v)] [0.05, 0.05]	[(v), (vi)] [0.5, 0.3]	[(vi), (v)] [0.45, 0.10]
Sales	– –	[(v), (v)] [0.95, 0.95]	[(vi), (v)] [0.8, 0.95]	[(v), (v)] [0.8, 0.05]	[(v), (vi)] [0.75, 0.10]	[(v), (vi)] [0.95, 0.20]
Revenue		– –	[(v), (v)] [0.95, 0.05]	[(vi), (vi)] [0.5, 0.05]	[(ii), (v)] [–, 0.05]	[(vi), (v)] [0.45, 0.45]
Price			– –	[(i), (v)] [–, 0.05]	[(v), (v)] [0.95, 0.05]	[(v), (v)] [0.75, 0.10]
Costs				– –	[(vi), (i)] [0.05, –]	[(vi), (v)] [0.05, 0.05]
Youth					– –	[(v), (v)] [0.05, 0.80]

TABLE 3

This table reports T_N-values (upper triangle part) and their converted pseudo p-values ($\tilde{\tilde{p}} = T_N/(3.3260)$) (lower triangle part) when 100 Max linear χ^2 tests were applied to each paired data

P-value	Tax	Sales	Revenue	Price	Costs	Youth	Adult
Tax	——	0.0010	0.0194	0	0	0.0031	0.0001
Sales	0.0003	——	0	0.0091	0.0784	0.0510	0.0187
Revenue	0.0058	0	——	0.0147	0.0034	0.0678	0.1139
Price	0	0.0027	0.0044	——	0	0.0009	0
Costs	0	0.0236	0.0010	0	——	0.0055	0.0001
Youth	0.0009	0.0153	0.0204	0.0003	0.0017	——	0.0003
Adult	0.0000	0.0056	0.0342	0	0.0000	0.0001	——

Revenue are taken a negative sign transformation. The numbers in rest cells can be interpreted similarly. We note that Table 2 also tells the directions of the dependencies. For example, Tax and Sales are negatively correlated. When Tax is at the 75th (sample) percentile or higher, Sales is more likely close to the 25th (sample) percentile.

We report T_N-values in Table 3 (upper triangle part) and their converted pseudo p-values ($\tilde{\tilde{p}} = T_N/3.3260$) (lower triangle part), see the table in Section 4.2 regarding the factor number 3.3260. From the table and taking into account of all possible transformations, we see that the T_N-value between Sales and Youth is 0.0510, while the T_N-value between Revenue and Adult is 0.1139. These are two largest T_N-values in the analysis, and they are less than 0.1663.

Table 3 clearly gives more information on how a null hypothesis is rejected. From this table, it may be safe to conclude that all paired variables are dependent using the cut-off $\tilde{p}_\alpha = 0.1663$ when $\alpha = .05$. We can see that it may be problematic for conclusions based on Table 1 as some variables have heavy tails. Especially using the new tests, we found that Sales and Costs are dependent, and Revenue and Adult are also dependent.

Our analysis results can certainly be useful in guiding social economic researchers for further studying correlations among those variables and influencing their decision making. We argue that this kind of data analysis is very important in social economic studies. With powerful statistical tools, analysis results can be influential in a sequence of political decisions, and economic decisions as well. However, this task is beyond the scope of the current paper.

7. Conclusions

In general, it is not easy to derive the limiting distribution for the maxima of several measures. Combining measures with the same or similar functional properties is also not desirable as they do not give much gains in measuring variable association strengths. In our combined measures, two candidate measures behave completely different. In the study of the combined measures, we are able to not only derive the limiting distributions, but also provide normalizing constants, which are especially useful for finite sample sizes. One can see that from our simulation examples, the new max linear χ^2 test at least achieves local optimality in testing hypothesis of independence. The newly combined dependence measures (6.1) can certainly be applied to many statistical applications in clustering analysis, classifications, causality studies, just to name a few. It may be safe to say that the newly combined coefficients can be used in any study which may need to calculate correlation coefficients. In the cases where a random variable does not have a finite second moment and a correlation coefficient is not meaningful, the new measure is still meaningful. The new χ^2 test can easily be implemented in any statistical software packages. We note that the new χ^2 test performs much better than some existing tests when sample sizes are small, which may be very useful when it is used in sample size calculations.

The methodology introduced in Section 4 can be used in many applications. Lemma 4.1 and Theorem 4.3 deliver very general results which may lead to new developments of statistical research topics. As illustrated in our simulation examples and real data analysis, our refined testing procedure in Subsection 6.1 may give readers a better understanding and a clearer picture how the variables are dependent of each other and how the null hypothesis of independence is rejected.

8. Appendix

The limit results of the following two lemmas regarding the first k smallest order statistics and the first k largest order statistics will be used in the proof of Theorem 2.1.

Lemma 8.1. *Let $U_1, U_2, \ldots, U_n$ be n independent random variables uniformly distributed over $(0,1)$, and $U_{(1)} \le U_{(2)} \le \cdots \le U_{(n)}$ be their order statistics. For any fixed integer $k \ge 1$, set $W_n = n(U_{(1)}, U_{(2)}, \ldots, U_{(k)})$ and $W_n' = n(1 - U_{(n)}, 1 - U_{(n-1)}, \ldots, 1 - U_{(n-k+1)})$. Then*

$$(W_n, W_n') \xrightarrow{\mathcal{L}} (W, W'), \tag{8.1}$$

where $W = (w_1, w_2, \ldots, w_k)$ *and* $W' = (w'_1, w'_2, \ldots, w'_k)$ *are defined by*

$$w_i = \sum_{j=1}^{i} E_j \quad \text{and} \quad w'_i = \sum_{j=1}^{i} E'_j$$

for $1 \le i \le k$, *and* E_j, E'_j, $j \ge 1$ *are iid unit exponential random variables.*

Lemma 8.2. *Assume that* Z_i, $i \ge 1$ *are iid random variables with a distribution function* $F(x)$. *Let* $Z_{(1)} \le Z_{(2)} \le \cdots \le Z_{(n)}$ *be the order statistics of* $Z_1, Z_2, \ldots, Z_n$. *Assume that*

$$F(x) = x(1 + o(1)) \text{ and } 1 - F(\frac{1}{x}) = x(1 + o(1)) \text{ as } x \downarrow 0. \tag{8.2}$$

Then for every fixed integer $k \ge 1$ *we have*

$$\left(nZ_{(1)}, nZ_{(2)}, \ldots, nZ_{(k)}, \frac{n}{Z_{(n)}}, \frac{n}{Z_{(n-1)}}, \ldots, \frac{n}{Z_{(n-k+1)}}\right) \xrightarrow{\mathcal{L}} (W, W'),$$

where W *and* W' *are defined as in Lemma 8.1.*

Proof of Lemma 8.1. Since W_n and W'_n have the same joint distributions, it suffices to show that $W_n \xrightarrow{\mathcal{L}} W$ and that W_n and W'_n are asymptotically independent.

Let $E_{(1)} \le E_{(2)} \le \cdots \le E_{(n)}$ be the order statistics of $E_1, E_2, \ldots, E_n$. Then it is well known that

$$(U_{(1)}, U_{(2)}, \ldots, U_{(n)}) \stackrel{d}{=} \big(1 - \exp(-E_{(1)}), 1 - \exp(-E_{(2)}), \ldots, 1 - \exp(-E_{(n)})\big). \tag{8.3}$$

Further, we have from Eq (3.5.6) on page 34 of [2]:

$$(E_{(1)}, E_{(2)}, \ldots, E_{(n)}) \stackrel{d}{=} \left(\sum_{j=1}^{1} \frac{E_j}{n-j+1}, \sum_{j=1}^{2} \frac{E_j}{n-j+1}, \ldots, \sum_{j=1}^{n} \frac{E_j}{n-j+1}\right). \tag{8.4}$$

For each fixed i, since $E_{(i)} \to 0$ in probability and $nE_{(i)} \xrightarrow{\mathcal{L}} \sum_{j=1}^{i} E_j$ as n tends to infinity, we have

$$\begin{aligned} W_n &\stackrel{d}{=} n\big(1 - \exp(-E_{(1)}), 1 - \exp(-E_{(2)}), \ldots, 1 - \exp(-E_{(k)})\big) \\ &= n(E_{(1)}, E_{(2)}, \ldots, E_{(k)}) + O_p\left(\frac{1}{n}\right) \\ &\xrightarrow{\mathcal{L}} (w_1, w_2, \ldots, w_k) \quad \text{as } n \to \infty. \end{aligned}$$

Set

$$V_n = n\big(1 - \exp(-E_{(1)}), 1 - \exp(-E_{(2)}), \ldots, 1 - \exp(-E_{(k)})\big)$$

and

$$V'_n = n\big(\exp(-E_{(n)}), \exp(-E_{(n-1)}), \ldots, \exp(-E_{(n-k+1)})\big).$$

From (8.3), to show the asymptotic independence of W_n and W_n', it suffices to prove the asymptotic independence of V_n and V_n'. To this end, set

$$V_n'' = n(\exp(-(E_{(n)} - E_{(k)})), \exp(-(E_{(n-1)} - E_{(k)})), \ldots, \exp(-(E_{(n-k+1)} - E_{(k)}))).$$

It follows from (8.4) that V_n and V_n'' are independent. Since

$$V_n' = V_n'' \exp(-E_{(k)}) = V_n''(1 + o_p(1)),$$

the asymptotic independence of V_n and V_n' is proved. This completes the proof of the lemma. $\square$

Proof of Lemma 8.2. Let F^- denote the generalized inverse of F by

$$F^-(u) = \inf\{x : F(x) \ge u\} \quad \text{for } u \in (0, 1).$$

If U is a random variable uniformly distributed over $(0, 1)$, then the distribution function of $F^-(U)$ is F. Therefore, $F^-(U_i)$, $1 \le i \le n$ are iid with the distribution F and

$$\big(Z_{(1)}, Z_{(2)}, \ldots, Z_{(n)}\big) \stackrel{d}{=} \big(F^-(U_{(1)}), F^-(U_{(2)}), \ldots, F^-(U_{(n)})\big).$$

From condition (8.2) we have

$$F^-(u) = u(1 + o(1)) \text{ and } F^-(1-u) = \frac{1 + o(1)}{u} \text{ as } u \downarrow 0.$$

Since $U_{(i)} \to 0$ and $U_{(n-i+1)} \to 1$ in probability, we have

$$F^-(U_{(i)}) = U_{(i)}(1 + o_p(1)), \quad F^-(U_{(n-i+1)}) = \frac{1 + o_p(1)}{1 - U_{(n-i+1)}}$$

for $1 \le i \le k$. Thus we have from Lemma 8.1

$$\left(nZ_{(1)}, nZ_{(2)}, \ldots, nZ_{(k)}, \frac{n}{Z_{(n)}}, \frac{n}{Z_{(n-1)}}, \ldots, \frac{n}{Z_{(n-k+1)}}\right) \stackrel{d}{=} (W_n, W_n') + o_p(1)$$
$$\xrightarrow{\mathcal{L}} (W, W'),$$

which completes the proof. $\square$

Proof of Lemma 2.2. Without loss of generality assume $\boldsymbol{\mu} = (0, \ldots, 0)$. Let $\mathbf{t} = (t_1, \ldots, t_p) \in \mathbb{R}^p$. It suffices to show that $\bar{\boldsymbol{\xi}}\mathbf{t}^\tau$ and $M_{\boldsymbol{\eta}}$ are asymptotically independent for every nonzero $\mathbf{t}$. To this end, for every vector $(x_1, \ldots, x_q)$ with $G(x_1, \ldots, x_q) \in (0, 1)$, set

$$B_n = \left\{ \frac{\max_{1 \le i \le n} \eta_{ki} - b_{kn}}{a_{kn}} \le x_k, \ k = 1, \ldots, q \right\}.$$

It suffices to show that the limit of the conditional distribution of $\sqrt{n}\bar{\boldsymbol{\xi}}\mathbf{t}^\tau$ given B_n is free of $(x_1, \ldots, x_q)$. Note that $\{\tau_j := \boldsymbol{\xi}_j \mathbf{t}^\tau, \quad 1 \le j \le n\}$ are iid random variables with mean 0 and variance $\sigma_{\mathbf{t}}^2 := \mathbf{t}\Sigma_{\xi}\mathbf{t}^\tau \in (0, \infty)$.

Set $A_n = \{\eta_k \leq a_{kn}+b_{kn}x_k, \ \ k = 1,\ldots,q\}$. Then we have $P(A_n) = 1+O(\frac{1}{n})$ from (2.3). Let $\{\tau_n', \ \ 1 \leq j \leq n\}$ denote a sequence of n iid random variables distributed the same as the the conditional distribution of $\boldsymbol{\xi}\mathbf{t}^\tau$ given B_n. Then we have

$$P\big(\sqrt{n}\bar{\boldsymbol{\xi}}\mathbf{t}^\tau \leq y|B_n\big) = P\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\tau_j' \leq y\right) \tag{8.5}$$

for every $y \in \mathbb{R}$.

Let $\bar{A}_n$ denote the complement of A_n and $I(A)$ indicator function of event A. It is easy to see by using Hölder's inequality that

$$\begin{aligned}|E(\tau_1')| &= \left|\frac{1}{P(A_n)}E(\tau_1 I(A_n))\right| = \left|\frac{1}{P(A_n)}E(\tau_1 I(\bar{A}_n))\right| \\ &= \frac{1}{P(A_n)}\sqrt{E(\tau_1^2 I(\bar{A}_n))}\sqrt{P(\bar{A}_n)} = o(n^{-1/2})\end{aligned}$$

and

$$E(\tau_1')^2 = \frac{1}{P(A_n)}E(\tau_1^2 I(A_n)) = \frac{1}{P(A_n)}(\sigma_{\mathbf{t}}^2 - E(\tau_1^2 I(\bar{A}_n))) = \sigma_{\mathbf{t}}^2 + o(1).$$

Note that

$$\sqrt{n}\bar{\boldsymbol{\xi}}\mathbf{t}^\tau = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\tau_j \xrightarrow{d} N(0,\sigma_{\mathbf{t}}^2).$$

To conclude the theorem, from (8.5) we need to show that $\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\tau_j'$ has the same limiting distribution by verifying the following Lindeberg-Feller condition

$$\frac{1}{n}\sum_{j=1}^{n}E(\tau_j')^2 I(|\tau_j'| > \varepsilon\sqrt{n}\sigma_{\mathbf{t}}) \leq \frac{1}{P(A_n)}E(\tau_1^2 I(|\tau| > \varepsilon\sqrt{n}\sigma_{\mathbf{t}})) \to 0$$

for every $\varepsilon > 0$. The last step is obvious since $E(\tau_1^2) \in (0,\infty)$. This complete the proof. □

Proof of Theorem 2.3. Without loss of generality we assume $\mu_1 = \mu_2 = 0$, and $\sigma_1 = \sigma_2 = 1$. Then it follows from the central limit theorem that

$$\sqrt{n}(\bar{\xi}_1,\bar{\xi}_2) \xrightarrow{d} N(0,\Sigma_1) \tag{8.6}$$

and

$$\sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n}\xi_{1k}\xi_{2k} - r, \frac{1}{n}\sum_{k=1}^{n}\xi_{1k}^2 - 1, \frac{1}{n}\sum_{k=1}^{n}\xi_{2k}^2 - 1\right) \xrightarrow{d} N(0,\Sigma_2), \tag{8.7}$$

where

$$\Sigma_1 = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$$

and

$$\Sigma_2 = \begin{pmatrix} E(\xi_1^2\xi_2^2) - r^2 & E(\xi_1^3\xi_2) - r & E(\xi_1\xi_2^3) - r \\ E(\xi_1^3\xi_2) - r & E(\xi_1^4) - 1 & E(\xi_1^2\xi_2^2) - 1 \\ E(\xi_1\xi_2^3) - r & E(\xi_1^2\xi_2^2) - 1 & E(\xi_2^4) - 1 \end{pmatrix}.$$

It follows from (8.6), (8.7) and Taylor's expansion that

$$r_n - r = \frac{1}{n}\sum_{k=1}^{n}\xi_{1k}\xi_{2k} - r - \frac{r}{2n}\sum_{k=1}^{n}(\xi_{1k}^2 - 1) - \frac{r}{2n}\sum_{k=1}^{n}(\xi_{2k}^2 - 1) + o_p(n^{-1/2}), \quad (8.8)$$

which together with (8.7) yields

$$\sqrt{n}(r_n - r) \xrightarrow{\mathcal{L}} N(0, \sigma^2).$$

The rest of the proof follows from Lemma 2.2 directly. □

The following result is closely related to Theorem 3.1 in [30]. It shows the properties of a random sample of (X, Y) with unit Fréchet margins.

Proposition 8.3. *If X and Y are independent and have unit Fréchet margins, i.e. $F_X(x) = F_Y(x) = e^{-1/x},\ x > 0$, and $(X_i,\ Y_i),\ i = 1, \ldots, n$ is a random sample from (X, Y), then*

$$2a_n \mathrm{q}_n \xrightarrow{\mathcal{L}} \chi_4^2, \quad (8.9)$$

where $a_n/n \to 1$ as $n \to \infty$, and χ_4^2 is a χ^2 random variable with 4 degrees of freedom.

Proof of Lemma 4.1. Since $\{X_1, \ldots, X_n\}$ is a random sample from a continuous random variable, we have $P\{\text{rank}[X_i] = \text{rank}[X_j],\ \text{for } i \neq j\} = 0$, which implies that $(\text{rank}[X_1], \ldots, \text{rank}[X_n])$ takes any permutation of $\{1, \ldots, n\}$ with the same probability $1/n!$. Similarly, $(\text{rank}[Y_1], \ldots, \text{rank}[Y_n])$ has the same distribution.

Note that $(\text{rank}[X_1], \ldots, \text{rank}[X_n])$ and $(\text{rank}[Y_1], \ldots, \text{rank}[Y_n])$ are independent, therefore $(Z^X_{(\text{rank}[X_1])}, \ldots, Z^X_{(\text{rank}[X_n])})$ and $(Z^Y_{(\text{rank}[Y_1])}, \ldots, Z^Y_{(\text{rank}[Y_n])})$ are independent.

We now derive the joint distribution of $(Z^X_{(\text{rank}[X_1])}, \ldots, Z^X_{(\text{rank}[X_n])})$. We have

$$\begin{aligned}
&P\{Z^X_{(\text{rank}[X_1])} \le x_1, \ldots, Z^X_{(\text{rank}[X_n])} \le x_n\} \\
&= \sum_{(k_1,\ldots k_n)\in\mathfrak{T}} P\{Z^X_{(k_1)} \le x_1, \ldots, Z^X_{(k_n)} \le x_n,\ \text{rank}[X_1] = k_1, \ldots, \text{rank}[X_n] = k_n\} \\
&= \sum_{(k_1,\ldots k_n)\in\mathfrak{T}} P\{Z^X_{(k_1)} \le x_1, \ldots, Z^X_{(k_n)} \le x_n\} \\
&\qquad\qquad \times P\{\text{rank}[X_1] = k_1, \ldots, \text{rank}[X_n] = k_n\} \\
&= \frac{1}{n!}\sum_{(k_1,\ldots k_n)\in\mathfrak{T}} P\{Z^X_{(k_1)} \le x_1, \ldots, Z^X_{(k_n)} \le x_n\}
\end{aligned}$$

where $\mathfrak{T}$ is the collection of permutations of $\{1,\ldots,n\}$. For any $(k_1,\ldots,k_n)\in\mathfrak{T}$, the joint density function of $(Z^X_{(k_1)},\ldots,Z^X_{(k_n)})$ is

$$f_{Z^X_{(k_1)},\ldots,Z^X_{(k_n)}}(t_1,\ldots,t_n) = \begin{cases} n!f_Z(t_1)f_Z(t_2)\cdots f_Z(t_n), & \text{if rank}[t_i]=k_i,\ i=1,\ldots,n;\\ 0, & \text{otherwise.}\end{cases}$$

Now define

$$\mathcal{D}=\{(t_1,\ldots,t_n)\mid 0<t_i\le x_i,\ i=1,\ldots,n\},$$
$$\mathcal{A}(k_1,\ldots,k_n)=\{(t_1,\ldots,t_n)\mid \text{rank}[t_i]=k_i,\ 0<t_i\le x_i, i=1,\ldots,n\}.$$

Then we have

$$\begin{aligned}\sum_{(k_1,\ldots k_n)\in\mathfrak{T}} & P\{Z^X_{(k_1)}\le x_1,\ldots,Z^X_{(k_n)}\le x_n\}\\ &= \sum_{(k_1,\ldots k_n)\in\mathfrak{T}}\int_{\mathcal{A}(k_1,\ldots,k_n)} n!f_Z(t_1)f_Z(t_2)\cdots f_Z(t_n)dt_1\ldots dt_n\\ &= n!\int_{\mathcal{D}} f_Z(t_1)f_Z(t_2)\cdots f_Z(t_n)dt_1\ldots dt_n = n!\prod_{i=1}^n P\{Z^X_i\le x_i\}.\end{aligned}$$

The second equality holds because $\bigcup_{(k_1,\ldots k_n)\in\mathfrak{T}}\mathcal{A}(k_1,\ldots,k_n)=\mathcal{D}$, and for any two permutations $(k_1,\ldots,k_n)$ and $(k'_1,\ldots,k'_n)$ we have

$$\mathcal{A}(k_1,\ldots,k_n)\cap\mathcal{A}(k'_1,\ldots,k'_n)\subset\{(t_1,\ldots,t_n)\mid t_i=t_j,\ \text{for some } i,j\}$$

which results in a zero probability event. Therefore, we obtain the joint probability of $(Z^X_{(\text{rank}[X_1])},\ldots,Z^X_{(\text{rank}[X_n])},Z^Y_{(\text{rank}[Y_1])},\ldots,Z^Y_{(\text{rank}[Y_n])})$ from

$$\begin{aligned}P\{Z^X_{(\text{rank}[X_1])}&\le x_1,\ldots,Z^X_{(\text{rank}[X_n])}\le x_n,Z^Y_{(\text{rank}[Y_1])}\le y_1,\ldots,Z^Y_{(\text{rank}[Y_n])}\le y_n\}\\ &=P\{Z^X_1\le x_1\}P\{Z^X_2\le x_2\}\cdots P\{Z^X_n\le x_n\}P\{Z^Y_1\le y_1\}P\{Z^Y_2\le y_2\}\\ &\quad\cdots P\{Z^Y_n\le y_n\}\end{aligned}$$

which implies that $\{Z^X_{(\text{rank}[X_1])},\ldots,Z^X_{(\text{rank}[X_n])}\}$ and $\{Z^Y_{(\text{rank}[Y_1])},\ldots,Z^Y_{(\text{rank}[Y_n])}\}$ are two independent samples from (X,Y), and hence the proof is completed. □

Proof of Theorem 4.2. By Lemma 4.1, $\{Z^X_{(\text{rank}[X_1])},\ldots,Z^X_{(\text{rank}[X_n])}\}$ and $\{Z^Y_{(\text{rank}[Y_1])},\ldots,Z^Y_{(\text{rank}[Y_n])}\}$ are two independent samples from two independent unit Fréchet random variables. Then we complete the proof by applying Proposition 8.3. □

Proof of Theorem 4.3. For any fixed n, conditional on $\mathcal{G}$, $\tilde{P}^{\mathfrak{R}}_{nj}$, $j\ge 1$ are bounded iid random variables with a conditional expectation $E(\tilde{P}^{\mathfrak{R}}_{n1}|\mathcal{G})$. Therefore, the theorem follows directly from the strong law of large numbers. □

Acknowledgments

The authors are grateful to an associate editor and a referee for their valuable comments. Zhang's research was supported by NSF Grant DMS-0804575, and Qi's research was supported by NSF Grant DMS-1005345.

References

[1] Baek, E. G., and Brock, W. A. (1992). A nonparametric test for independence of a multivariate time series. *Statistica Sinica*, **2**, 137–156. MR1152301

[2] Balakrishnan, N. and Cohen, A. C. (1991). Order Statistics & Inference: Estimation Methods. Academic Press. MR1084812

[3] Blum, J.R., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.* **32**, 485–498. MR0125690

[4] Centers for Disease Control and Prevention (CDC). (2007). Behavior Risk Factor Surveillance Survey (BRFSS); State Cigarette Taxes: http://www.tobaccofreekids.org/research/factsheets/pdf/0099.pdf

[5] Christensen R. and Bedrick, E. J. (1997). Testing the independence assumption in linear models. *Journal of the American Statistical Association*, **92**, 1006–1016. MR1482131

[6] Cromwell, J. B., Labys, W. C., and Terraza, M. (1994). *Univariate tests for time series models*. Sage Publications.

[7] Cromwell, J. B., Hannan, M. J., Labys, W. C., and Terraza, M. (1999). *Multivariate tests for time series models*. Sage Publications.

[8] Deheuvels, P. (1983). Point processes and multivariate extreme values. *Jour. of Multi. Anal.* **13**, pp. 257–272. MR0705550

[9] Diaconis, P. and Efron, B. (1985). Testing for independence in a two-way table. *Annals of Statistics*, **13**, 845–874. MR0803747

[10] Doksum, K. and Samarov, A. (1995). Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *The Annals of Statistics*, **23**, 1443–1473 MR1370291

[11] Drouet-Mari, D. and Kotz, S. (2001). *Correaltion and dependence*. Imperial College Press.

[12] Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, **10**, 507–521.

[13] Fisher, R. A. (1921). On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample. *Metron*, **1**, 3–32.

[14] Freeman, J. and Modarres, R. (2005). Efficiency of test for independence after Box-Cox transformation. *Journal of Multivariate Analysis*, **95**, 107–118. MR2164125

[15] Heffernan, J. E., Tawn, J. A., and Zhang, Z. (2007). Asymptotically (in)dependent multivariate maxima of moving maxima processes. *Extremes*, **10**, 57–82. MR2397551

[16] Hoeffding, W. (1948). A nonparametric test of independence. *Ann. Math. Statist.* **19** (1948), 546–557. MR0029139
[17] Hong, Y. (1996). Testing for independence between two covariance stationary time series. *Biometrika*. **83**, 615–625. MR1423878
[18] Hotelling, H. (1953). New light on the correlation coefficient and its transform. *Journal of the Royal Statistical Society*, Series B **15**, 193–232. MR0060794
[19] Ikeda, S. and Matsunawa, T. (1970). On asymptotic independence of order statistics. *Ann. Inst. Statist. Math.*, **22**, 435–449. MR0282481
[20] Joe, H. (1997). *M*ultivariate models and dependence concepts, Chapman & Hall. MR1462613
[21] Kanji, G. K. (1999). 100 Statistical Tests. Sage Publications.
[22] Kendall, M. (1938). A new measure of rank correlation, *Biometrika*, **30**, 81–89.
[23] Smith, R. L. and Weissman, I. (1996). Characterization and estimation of the multivariate extremal index. Unpublished manuscript, University of North Carolina. http://www.stat.unc.edu/postscript/rs/extremal.pdf
[24] Spearman, C. (1904). The proof and measurement of association between two things. *Amer. J. Psychol.*, **15**, 72–101.
[25] Special Report: (2008). Higher cigarette taxes: reduce smoking, save lives, save money. http://www.tobaccofreekids.org/reports/prices/
[26] Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Annal. Statist.* **35**, 2769–2794. MR2382665
[27] Thode, H. C. Jr. (2002). *Testing for Normality*. Marcel Dekker. MR1989476
[28] Wang, Y. (2010). Quotient correlation of dependent random variables. Manuscript. Department of Statistics, University of Michigan.
[29] Wilcox, R. R. (1995). Testing the hypothesis of independence between two sets of variates. *Multivariate Behavioral Research*. 30, No. 2, Pages 213–225.
[30] Zhang, Z. (2008). Quotient correlation: A sample based alternative to Pearson's correlation. *Annals of Statistics*. **36**(2), 1007–1030. MR2396823
[31] Zhang, Z. (2009). On approximating max-stable processes and constructing extremal copula functions. *Statistical Inference for Stochastic Processes*, **12**, 89–114. MR2486118
[32] Zhang, Z. and Smith, R.L. (2004). The behavior of multivariate maxima of moving maxima processes. *Journal of Applied Probability*, **41**, 1113–1123. MR2122805