Asymptotic Independence of Correlation Coefficients and
Rank-preserving Scale Regeneration with Application to Testing
Hypothesis of Independence

Zhengjun Zhang†, Yongcheng Qi‡ and Xiwen Ma†

† Department of Statistics, University of Wisconsin, Madison, WI 53706, USA
‡ Department of Mathematics and Statistics, University of Minnesota Duluth, Duluth, MN 55812, USA

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 $\chi^2$ 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.

Keywords: dependence measures, distribution free test, Fisher’s $Z$-transformation tests, gamma tests, $\chi^2$ tests, order statistics, conditional expectation.

Short title: Asymptotic Independence of Correlation Coefficients.


# 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. 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. The seminal nonparametric test of independence by Hoeffding (1948), Blum, Kieffer and Rosenblatt (1960) 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, Fisher (1921) introduced the extremely useful $Z$-transformation function and test statistics, and Hotelling (1953) stated that the best present-day usage in dealing with correlation coefficients is based on Fisher (1915, 1921). Recently, Freeman and Modarres (2005) 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 (Diaconis and Efron, 1985), testing independence between two covariance stationary time series (Hong, 1996), testing the independence assumption in linear models (Christensen and Bedrick, 1997), and others such as Baek and Brock (1992), Doksum and Samarov (1995), Wilcox (1995), as well as excellent reference books by Cromwell, Labys, and Terraza (1994), and Cromwell, Hannan, Labys, and Terraza (1999), Kanji (1999), and Thode (2002), among others.

Empirical evidences have shown that the gamma test (Zhang, 2008) 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 testing 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 (Spearman, 1904), Kendall tau rank correlation coefficient (Kendall, 1938), the distance correlation coefficient (Szekely et al., 2007), the quotient correlation coefficient (Zhang, 2008), 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. Drouet-Mari and Kotz (2001) and Joe (1997) are excellent reference books related to correlations and multivariate dependencies like aforementioned ones. These dependence measures will not be considered in this paper.

Each of these correlation measures may be used 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 may be the best linear dependence measure which may be most capable of describing the data in the central region, and the quotient correlation is a nonlinear dependence measure which is also capable of describing dependence in extreme values, and most importantly they are asymp-
totically independent, we propose to use the maximum of these two sample based correlation coefficients as the association measure of the relationship between two continuous random variables.

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 $\chi^2$ 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 testing 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 testing 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 $\chi^2$ 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 $\chi^2$ distributed when the test statistic is based on the transformed data. We further propose a new testing procedure, i.e. a rank-preserving scale regeneration and conditional expectation (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 mea-
sure to a more general form and illustrate methods of data processing and variable scale
transformations. We apply the newly introduced max linear $\chi^2$ 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 $P(X \geq Y) > 0$, $P(X \leq Y) > 0$. 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 \((X, Y)\). Zhang
(2008) introduced the sample based quotient correlation coefficient:

$$q_n = \frac{\max_{i \leq n} \{Y_i/X_i\} + \max_{i \leq n} \{X_i/Y_i\} - 2}{\max_{i \leq n} \{Y_i/X_i\} \times \max_{i \leq n} \{X_i/Y_i\} - 1}.$$  \(2.1\)

It is easy to see that both $(\max_{i \leq n} \{X_i/Y_i\} - 1)$ and $(\max_{i \leq n} \{Y_i/X_i\} - 1)$ are asymptotically
nonnegative. Properties and geometric interpretation of (2.1) are illustrated in Zhang (2008).

Zhang also showed that $nq_n \xrightarrow{L} \zeta$ (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.
The following example is to show that $q_n$ has a non-zero limit.

**Example 2.1** Suppose $Z_l, l > 0$, are independent unit Fréchet random variables with distribu-
tion function $F_Z(x) = \exp(-1/x), x > 0$. Let $X = \max_{l>0} \alpha_{1l} Z_l$, $Y = \max_{l>0} \alpha_{2l} Z_l$, where
$\alpha_{li} > 0, \sum_l \alpha_{li} = 1, i = 1, 2$. Let $c_1 = \sup_{l>0} (\alpha_{1l}/\alpha_{2l}) < \infty$ and $c_2 = \sup_{l>0} (\alpha_{2l}/\alpha_{1l}) < \infty$, then $P(X/Y \leq c_1) = P(Y/X \leq c_2) = 1$ which implies that $\max_{i \leq n} \{X_i/Y_i\} \xrightarrow{a.s.} c_1$, $\max_{i \leq n} \{Y_i/X_i\} \xrightarrow{a.s.} c_2$, and $q_n \xrightarrow{a.s.} \frac{c_1 + c_2 - 2}{c_1 c_2 - 1}$, as $n \to \infty$. 

5
This is a special case of multivariate maxima of moving maxima processes (or M4 processes) which have been relatively extensively studied in Smith and Weissman (1996), Zhang and Smith (2004). Theorem 2 of Deheuvels (1983) establishes that for any finite \( n \), the joint distribution of \( n \) consecutive values from a max-stable process with Fréchet marginals can be approximated by a weighted maximum \( Y_{id} = \max_{-\infty < k < \infty} \alpha_{idk} Z_k \) for suitable coefficients \( \alpha_{idk} \), where \( Z_k \)'s are independent and identically distributed (iid) unit Fréchet. Here “can be approximated” is in the sense that there is a sequence of such approximations that converges in distribution to the true distribution of the \( Y_{id} \). We refer detailed arguments to Smith and Weissman (1996) as they are beyond the scope of this paper. In addition to the theories derived in Smith and Weissman (1996), Heffernan, Tawn and Zhang (2007) and Zhang (2009) developed new theories which suggest that M4 process models are a rich family which can be used in modeling a wide range of multivariate dependencies. These results suggest that as long as two random variables are tail dependent, models in Example 2.1 can be used to approximate tail values of the two random variables, and then one uses an approximate model to conduct statistical inferences. Therefore, \( q_n \) can be used as a sample based dependence measure. For detailed discussions of \( q_n \) and its extensions to tail values above thresholds, we refer to Zhang (2008).

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 \( nq_n \). The testing hypotheses are

\[
H_0 : X \text{ and } Y \text{ are independent} \quad \text{versus} \quad H_1 : X \text{ and } Y \text{ are dependent}.
\]

We shall see, for example, in a later section that the gamma test statistic \( nq_n \) does not achieve the best testing 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 power if the dependence mainly comes from the center part of the data. Again, we note that a more efficient test statistic related to \( q_n \) for testing tail independence is the tail quotient correlation coefficient using the values above a certain threshold which can be a random threshold and can diverge to infinity almost surely. Our focus in this paper is testing independence, not 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

The following two lemmas show that the first $k$ smallest order statistics are asymptotically independent of the first $k$ largest order statistics.

**Lemma 2.1** Let $U_1, U_2, \cdots, U_n$ be $n$ independent random variables uniformly distributed over $(0, 1)$, and $U_{(1)} \leq U_{(2)} \leq \cdots \leq U_{(n)}$ be their order statistics. For any fixed integer $k \geq 1$, set $W_n = n(U_{(1)}, U_{(2)}, \cdots, U_{(k)})$ and $W'_n = n(1 - U_{(n)}, 1 - U_{(n-1)}, \cdots, 1 - U_{(n-k+1)})$. Then

\[
(W_n, W'_n) \overset{d}{\rightarrow} (W, W'),
\]

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

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

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

**Lemma 2.2** Assume that $Z_i$, $i \geq 1$ are iid random variables with a distribution function $F(x)$. Let $Z_{(1)} \leq Z_{(2)} \leq \cdots \leq Z_{(n)}$ be the order statistics of $Z_1, Z_2, \cdots, Z_n$. Assume that

\[
F(x) = x(1 + o(1)) \quad \text{and} \quad 1 - F\left(\frac{1}{x}\right) = x(1 + o(1)) \quad \text{as} \quad x \downarrow 0.
\]

Then for every fixed integer $k \geq 1$ we have

\[
\left(nZ_{(1)}, nZ_{(2)}, \cdots, nZ_{(k)}, \frac{n}{Z_{(n)}}, \frac{n}{Z_{(n-1)}}, \cdots, \frac{n}{Z_{(n-k+1)}}\right) \overset{d}{\rightarrow} (W, W'),
\]

where $W$ and $W'$ are defined as in Lemma 2.1.

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 \(j\)th smallest ratio:

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

Then for any fixed \((i,j)\), from Lemma 2.2, 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 2.1 and 2.2.

**Theorem 2.3** Suppose \(\{X_i, Y_i\}, i \geq 1\) are iid unit Fréchet distributed random variables. Then

\[
q_n(i, j) \overset{\mathcal{L}}{\rightarrow} \zeta,
\]

where \(\zeta\) 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 under some mild conditions in the following lemma that the sample mean vector of a random sample of a random vector \(\xi\) is asymptotically independent of a sample vector of componentwise maxima of a random vector \(\eta\) which is dependent of \(\xi\).

**Lemma 2.4** Let \(\{(\xi_j, \eta_j), 1 \leq j \leq n\}\) be a sequence of independent random vectors having the same distribution as that of vector \((\xi, \eta)\), where \(\xi = (\xi_1, \xi_2, \ldots, \xi_p)\) is a random vector with finite covariance matrix \(\Sigma_{\xi}\), i.e. all elements in \(\Sigma_{\xi}\) are finite, and \(\eta = (\eta_1, \eta_2, \ldots, \eta_q)\) is a random vector which belongs to the max-domain of attraction of a multivariate extreme value distribution, i.e.

\[
\lim_{n \to \infty} P^n(\eta_1 \leq a_1 x_1 + b_1, \ldots, \eta_q \leq a_q x_q + b_q) = G(x_1, \ldots, x_q) \text{ for } x \in \mathbb{R}^q \tag{2.5}
\]
where $a_{in}$ are positive normalizing constants, $x = (x_1, \ldots, x_q)$.

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

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

(2.6)
as $n \to \infty$, where $\Phi$ is the standard multivariate normal distribution function, $p$ and $q$ are fixed integer numbers.

Suppose now that $\xi = (\xi_1, \xi_2, \xi_3, \xi_4, \xi_5)$, where $\xi_3 = \xi_4^2$, $\xi_4 = \xi_5^2$, $\xi_5 = \xi_1 \xi_2$; $\eta = (\eta_1, \eta_2)$, 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 $\xi_1$ and $\xi_2$ is given by

$$r_n(\xi_1, \xi_2) = \frac{\bar{\xi}_5 - \bar{\xi}_1 \bar{\xi}_2}{\sqrt{(\xi_3 - \xi_1^2)(\xi_4 - \xi_2^2)}}$$

(2.7)

Then we have the following theorem.

**Theorem 2.5** With the established notations, suppose $\xi$ and $\eta$ satisfy the conditions in Lemma 2.4. Then

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

(2.8)
as $n \to \infty$, where $r$ is Pearson correlation coefficient of $\xi_1$ and $\xi_2$, and $\mu_1 = E(\xi_1)$, $\mu_2 = E(\xi_2)$, $\sigma_1 = \sqrt{Var(\xi_1)}$, $\sigma_2 = \sqrt{Var(\xi_2)}$,

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

$$-rE(\frac{\xi_1 - \mu_1}{\sigma_1})(\frac{\xi_2 - \mu_2}{\sigma_2})(\frac{\xi_1 - \mu_1}{\sigma_1}^2 + (\frac{\xi_2 - \mu_2}{\sigma_2})^2),$$

and $\Phi$ is the standard normal distribution function.

Furthermore, when $\xi_1$ and $\xi_2$ are independent, and denote $X_i = \eta_{1i}$, $Y_i = \eta_{2i}$ in (2.1). Then we have

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

(2.9)

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 (2.7) 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 \( \{(X_i, Y_i), i = 1, \ldots, n\} \) 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:

\[
P(X/Y < z) = \left( \frac{z}{1+z} \right)^{1+\lambda}, \quad z > 0, \quad 0 \leq \lambda < 1,\]

\[
\text{corr} \left( \Phi^{-1}(e^{-1/X}), \Phi^{-1}(e^{-1/Y}) \right) = \rho. \]

Then \( q_n \) can be used to test the following hypotheses:

\[
H_0^\lambda : \lambda = 0 \quad \text{versus} \quad H_1^\lambda : \lambda > 0, \]

and \( r_n \) can be used to test the following hypotheses:

\[
H_0^\rho : \rho = 0 \quad \text{versus} \quad H_1^\rho : \rho \neq 0. \]
Notice that $H_0$ implies both $H_0^\lambda$ and $H_0^\rho$, but not reversely. There is no direct connection between $H_0^\lambda$ and $H_0^\rho$. It is not known yet that whether $H_0^\lambda$ 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. An improved better inference problem is to test

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

(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 $H_0$, the quotient coefficient $q_n$ is shown to be $n$ convergence (Zhang, 2008), 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:

$$C_n = \max(q_n^{1/2}, |r_n|).$$  

(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 $q_n^{1/2}$. In the case of co-monotone or positive dependent, the limit of $C_n$ can be either from $q_n^{1/2}$ 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:

$$nC_n^2 \xrightarrow{d} \max(\chi^2_1, \frac{1}{2}\chi^2_4),$$

where $\chi^2_1$ and $\chi^2_4$ are two independent $\chi^2$ 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 $\chi^2_4$ 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 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 $\chi^2$ 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 $\alpha$ percentile of the random variable $\max(\chi^2_1, \frac{1}{2}\chi^2_4)$. 

We note that the new max linear \( \chi^2 \) 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 \( \chi^2 \) 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 Wang (2010) for the derivation, and hence the max linear \( \chi^2 \) 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 \( H^1_\lambda \) with \( \lambda = \lambda_1 > 0 \), \( nq_n \) is an asymptotic power one test, and hence \( nC^2_n \) is also an asymptotic power one test under any fixed alternative hypothesis \( H^c_1 \). We now study the asymptotic powers under a class of local alternatives:

\[
H^c_0 : \lambda = \lambda_0 = 0, \rho = \rho_0 = 0 \text{ versus } H^c_{1Ln} : \rho = \rho_n = \Delta/\sqrt{n}, \lambda = \lambda_n = \Delta/\sqrt{n}, \Delta > 0.
\]

(3.7)

Note that the distribution function of random variable \( \max(\chi^2_1, \frac12 \chi^2_4) \) is given by

\[
F_{\max}(x) = 2(\Phi(\sqrt{x}) - \frac12)(1 - e^{-x} - \frac12 xe^{-x}), \quad x \geq 0,
\]

where \( \Phi \) is the standard normal distribution function.

Denote \( \sigma \) in Theorem 2.5 under \( H_0 \) and \( H^c_{0Ln} \) as \( \sigma_0 \) and \( \sigma_n \) respectively, and let \( u_\alpha = \chi^2_{\max,\alpha} \) which is the \( \alpha \) 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 \( H^c_{0Ln} \), the asymptotic power is

\[
\beta_n = P_n(nC^2_n > u_\alpha) = 1 - P_n(nq_n < u_\alpha, nr_n^2 < u_\alpha)
\]

\[
= 1 - P_n\left(nq_n < u_\alpha, \frac{\sqrt{u_\alpha}}{\sigma_0} < \sqrt{n}(\rho_n - \rho_0)\frac{\sigma_n}{\sigma_0}, \frac{\sqrt{u_\alpha}}{\sigma_0} + \frac{\sqrt{n}(\rho_n - \rho_0)}{\sigma_n} < \frac{\sqrt{u_\alpha}}{\sigma_0}\right)
\]

\[
\rightarrow 1 - \left[\Phi\left(-\frac{\Delta}{\sigma_0} + \frac{\sqrt{u_\alpha}}{\sigma_0}\right) - \Phi\left(-\frac{\Delta}{\sigma_0} - \frac{\sqrt{u_\alpha}}{\sigma_0}\right)\right](1 - e^{-u_\alpha} - \frac12 u_\alpha e^{-u_\alpha}).
\]

### 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 \( nC^2_n \) as a test statistic does not have testing 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:

$$C^\delta_n = \max(a_n^{1/2}, |r_n|, |D|^{1+\delta}),$$

(3.8)

where $\delta$ is a small positive number, say $\delta = 10^{-10}$, and $D$ is defined as follows (Hoeffding 1948). 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 $\Sigma''$ denotes summation over all $\alpha$ 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 Hoeffding (1948) 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 $C^\delta_n$ has the same limit distribution as $C_n$ has. And we still call $nC^\delta_n$ as the new max linear $\chi^2$ test statistic. Under the alternative hypothesis, $D$ has a positive limit, and hence $nC^\delta_n$ is an asymptotic power one test.

The use of $D$ as a nonparametric test (HBKR test) of independence has been well-known due to Hoeffding (1948), and Blum, Kieffer and Rosenblatt (1960). 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 $nC^\delta_n$ has a better performance, and the computation of $nC^\delta_n$ is very simple and fast.

The theories established in this section are valid when marginal distributions are known. When marginal distributions are not known, we introduce a very useful marginal transformation procedure next.
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 \(Z_{(1)}^X < Z_{(2)}^X < \cdots < Z_{(n)}^X\), \(Z_{(1)}^Y < Z_{(2)}^Y < \cdots < Z_{(n)}^Y\) as the ordered samples respectively, and the rank-preserving scale regeneration marginal transformation procedure results in the following transformed data

\[
X_i^* = Z_{(\text{rank}[X_i])}^X, \quad Y_i^* = Z_{(\text{rank}[Y_i])}^Y, \quad i = 1, \ldots, n.
\] (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

\[
q_{Rn} = \frac{\max_{i \leq n} \{Y_i^*/X_i^*\} + \max_{i \leq n} \{X_i^*/Y_i^*\} - 2}{\max_{i \leq n} \{Y_i^*/X_i^*\} \times \max_{i \leq n} \{X_i^*/Y_i^*\} - 1}.
\] (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 the simulation procedure here is similar to the procedure in Zhang (2008) 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), \{\((X_i, Y_i), i = 1, \ldots, n\)\} is a random sample from \((X, Y)\), and \{\((X_i', Y_i'), i = 1, \ldots, n\)\} is a different random sample from \((X, Y)\). Denote \(Z_{(1)}^X < Z_{(2)}^X < \cdots < Z_{(n)}^X\), \(Z_{(1)}^Y < Z_{(2)}^Y < \cdots < Z_{(n)}^Y\) as the ordered samples of \{\((X_i, Y_i'), i = 1, \ldots, n\)\}
respectively. Define

\[ X^*_i = Z^{X}_{\text{rank}(X_i)}, \quad Y^*_i = Z^{Y}_{\text{rank}(Y_i)}, \quad i = 1, \ldots, n. \]

Then \( \{(X^*_i, Y^*_i), \ i = 1, \ldots, n\} \) 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 \(\{(X_i, Y_i), \ i = 1, \ldots, n\}\) is a random sample from \((X, Y)\). Then \(q_n^R\) is asymptotically \(\chi^2\) distributed - i.e., \(2a_nq_n^R \xrightarrow{\mathcal{L}} \chi^2_4\), as \(a_n/n \to 1, \ n \to \infty\), where \(\chi^2_4\) is a \(\chi^2\) 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 Zhang (2008) was proved under a stronger condition than necessary. The proof of the above theorem does not need to satisfy that condition.

We find 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 \(n(\max\{(q_n^R)^{1/2}, |r_n^R|, |D|^{1+\delta}\})^2 > \chi^2_{\max, \alpha}\), \(H_0\) of (2.2) is rejected; otherwise it is retained. Here \(r_n^R\) is computed using (2.7) 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 \(\chi^2\) tests and making a convincing conclusion based on rates of rejecting the null hypothesis. We have the following remark:
Remark 4.1. Suppose \( N \) statisticians transform data to unit Fréchet scales using their own methods and models. Parameter estimators in every model are consistent. Each statistician applies the new max linear \( \chi^2 \) test to the transformed unit Fréchet scaled data and calculates a \( p \)-value. There will be \( N \) \( p \)-values. An alternative way is to implement these transformations in computer based experiment, i.e. we may use (4.1) to mimic all statisticians’ transformation procedures and calculate \( N \) \( p \)-values.

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 \( Z_{j(1)}^X < Z_{j(2)}^X < \cdots < Z_{j(n)}^X \) as the ordered samples respectively, and the rank-preserving scale regeneration marginal transformation procedure results in the following transformed data

\[
X_{i,j} = Z_{j(\text{rank}[X_i])}^X, \quad Y_{i,j} = Z_{j(\text{rank}[Y_i])}^Y, \quad i = 1, \ldots, n, \quad j = 1, \ldots, N. \tag{4.3}
\]

For each \( j \), the correlation coefficient \( r_{nj}^n \) and the simulation based quotient \( q_{nj}^n \) 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 \( \tilde{P}_{nj}^n = 1 - F_{\max}(n(\max\{(q_{nj}^n)^{1/2}, |r_{nj}^n|, |D|^{1+\delta})\}^2) \).

- The decision rule is:

  - If \( T_N = \frac{1}{N} \sum_{j=1}^{N} \tilde{P}_{nj}^n < \tilde{p}_\alpha \), 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 (2.7) respectively. Then

\[
T_N \xrightarrow{a.s.} E(\tilde{P}_{n1}^n | \mathcal{G}) \quad \text{as } N \to \infty
\]

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

We want \( P(T_N < \tilde{p}_\alpha) \leq \alpha \), or approximately \( P\{E(\tilde{P}_{n1}^n | \mathcal{G}) < \tilde{p}_\alpha\} \leq \alpha \). The distribution of \( E(\tilde{P}_{n1}^n | \mathcal{G}) \) 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 $\alpha$ values which may satisfy the need of the most applications.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>.001</th>
<th>.005</th>
<th>.01</th>
<th>.025</th>
<th>.05</th>
<th>.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{p}_\alpha$</td>
<td>0.0036</td>
<td>0.0203</td>
<td>0.0415</td>
<td>0.0956</td>
<td>0.1663</td>
<td>0.2567</td>
</tr>
<tr>
<td>$\tilde{p}_\alpha / \alpha$</td>
<td>3.6000</td>
<td>4.0600</td>
<td>4.1500</td>
<td>3.8240</td>
<td>3.3260</td>
<td>2.5670</td>
</tr>
</tbody>
</table>

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 $\tilde{p} = T_N / (\tilde{p}_\alpha / \alpha)$ with $\alpha$ directly. In this case, we call $\tilde{p}$ a rank-preserving scale regeneration and conditional expectation (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 $\alpha$, 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 (2.7);
  - If $n[C_n^2] > \chi_{\max, \alpha}^2$, $H_0$ 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 $q_n^{\text{Fr}}$ using (4.2) and $r_n^{\text{Fr}}$ based on marginally transformed normal data;
  - We repeat the above two steps $N$ times, denote $q_n^{\text{Fr}}$ and $r_n^{\text{Fr}}$ as $q_{nj}^{\text{Fr}}$ and $r_{nj}^{\text{Fr}}$, $j = 1, \ldots, N$ respectively, and then compute
    $$\tilde{P}_{nj}^{\text{Fr}} = 1 - F_{\max}(n(\max\{(q_{nj}^{\text{Fr}})^{1/2}, |r_{nj}^{\text{Fr}}|, |D|^{1+\delta})\}), \quad j = 1, \ldots, N;$$
  - If $T_N = \frac{1}{N} \sum_{j=1}^{N} \tilde{P}_{nj}^{\text{Fr}} < \tilde{p}_\alpha$, 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. Testing powers of simulation examples will be given in Section 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 \( \chi^2 \) 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 \( \chi^2 \) 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.

![Figure 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.](image)

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.

As discussed in the introduction section, one may think that high testing powers can be achieved via combining more dependence measures into one maximal strength association measure and using a permutation test. It is generally true that the combined measure performs better than any of its component measures in achieving high testing powers. When compared to different combinations, we find that a combined measure with a larger number of component measures does not necessarily performs better than a different combined measure with a smaller number of component measures. It really depends on which measures are used in the combination. There will be many combinations of dependence measures introduced in the literature, and it is not feasible to study all combinations. Ideally, one prefers a combined measure with the number of component measures being as small as possible. Simulation examples in this section aim to give evidences that our proposed combination works as good as the best testing procedure in each case.

For comparison purpose, we also conduct Fisher’s $Z$-transformation test, the gamma test introduced in Zhang (2008).

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 drawn from empirical powers calculated based on the gamma test, Fisher’s $Z$-transformation test, HBKR test, the new max (linear) $\chi^2$ 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 testing 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
Figure 2: Comparisons among the gamma test, Fisher’s Z-transformation test (see Fisher (1921), or Zhang (2008) for the definition of the test), the HBKR test, the new max linear $\chi^2$ 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.

coefficients. We see that the new max linear $\chi^2$ 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 completed 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 $\chi^2$ 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, …, 240. Empirical powers for all tests are plotted in Figure 3.

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 $\chi^2$ 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 Zhang (2008). In other words, under the alternative hypothesis of dependence the combined test statistic (the new max linear $\chi^2$ 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 $\chi^2$ 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 testing 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 plus simulation transformation retains dependencies in data, which is very important in practice.

In the next example, we demonstrate that the new max linear $\chi^2$ test and the R.A.C.E test gain higher testing 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 $\left( |X|, |Y| \right)$, where $Y = X \ast E + W$, 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.

From Figure 4, under the alternative hypothesis of dependence the combined test statistic (the max linear $\chi^2$ 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 $\chi^2$ test is superior. We have extensively applied the new max linear $\chi^2$ test to many other simulation examples and observed that the new $\chi^2$ test outperforms existing tests. Based on these test results, it may be safe to say that the new max linear $\chi^2$ 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 testing 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 $\mathcal{F}$ be a collection of candidate transformations, we extend (3.6) to be

$$C_n^* = \sup_{g_x, g_y \in \mathcal{F}} \max(q_n^{1/2}(g_x, g_y), |r_n(g_x, g_y)|, |D(g_x, g_y)|^{1+\delta}),$$

(6.1)

where $q_n(g_x, g_y) = q_n^R$ in (4.2) with $(X^*_i, Y^*_i)$ being obtained from

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

$\text{rank}[g_x(X_i)]$ is obtained from transformed sample $\{g_x(X_1), g_x(X_2), \ldots, g_x(X_n)\}$, and $\text{rank}[g_y(Y_i)]$ is defined similarly. $|r_n(g_x, g_y)|$ is calculated using (2.7) 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 $\mathcal{F}$ maximum association measure between two random variables. Of course, the choice of $\mathcal{F}$ 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) $X(v) = |\log(|X|/a_k)|$; (vi) $X(vi) = -|\log(|X|/a_k)|$, where $a_k$ is the $k$th 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 $p$th 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 \( F \)), our refined R.A.C.E testing procedure is:

- For each transformed data \( \{(X_i(s), Y_i(t)), i = 1, ..., n\} \) 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 \( \tilde{p} = T_N/(\tilde{p}_\alpha/\alpha) \), where cut-off value \( \tilde{p}_\alpha \) is given in Section 4.2. Then we keep the combination which has the smallest \( \tilde{p} \)-value.

- If the smallest \( \tilde{p} \) value is less than \( \alpha \), 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.” Special Report (2008). 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 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. 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 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 $\chi^2$ 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.
Table 1: This table reports $p$-values obtained when Fisher $Z$-transformation tests (upper triangle part) were applied to data drawn from Figure 5.

<table>
<thead>
<tr>
<th>P-value</th>
<th>Sales</th>
<th>Revenue</th>
<th>Price</th>
<th>Costs</th>
<th>Youth</th>
<th>Adult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tax</td>
<td>0.0003</td>
<td>0.0000</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Sales</td>
<td>0</td>
<td>0.0030</td>
<td>0.2654</td>
<td>0.0806</td>
<td>0.0137</td>
<td></td>
</tr>
<tr>
<td>Revenue</td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0197</td>
<td>0.0278</td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td></td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costs</td>
<td></td>
<td></td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>Youth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

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 $\tilde{p}$-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 $k$th 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 Revenue are taken a negative sign transformation. The numbers in rest cells can be interpreted similarly. We note 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 ($\tilde{p} = T_N/3.3260$) (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,
Table 2: This table reports percentiles and combinations used when the new $\chi^2$ tests were applied to data drawn from Figure 5.

<table>
<thead>
<tr>
<th></th>
<th>Sales</th>
<th>Revenue</th>
<th>Price</th>
<th>Costs</th>
<th>Youth</th>
<th>Adult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tax</td>
<td>[(vi), (v)]</td>
<td>[(v), (v)]</td>
<td>[(v), (v)]</td>
<td>[(v), (vi)]</td>
<td>[(vi), (v)]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.75, 0.75]</td>
<td>[0.2, 0.5]</td>
<td>[0.2, 0.05]</td>
<td>[0.05, 0.05]</td>
<td>[0.5, 0.3]</td>
<td>[0.45, 0.10]</td>
</tr>
<tr>
<td>Sales</td>
<td>–</td>
<td>[(v), (v)]</td>
<td>[(vi), (v)]</td>
<td>[(v), (v)]</td>
<td>[(v), (vi)]</td>
<td>[(v), (vi)]</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>[0.95, 0.95]</td>
<td>[0.8, 0.95]</td>
<td>[0.8, 0.05]</td>
<td>[0.75, 0.10]</td>
<td>[0.95, 0.20]</td>
</tr>
<tr>
<td>Revenue</td>
<td>–</td>
<td>–</td>
<td>[(v), (v)]</td>
<td>[(vi), (v)]</td>
<td>[(ii), (v)]</td>
<td>[(vi), (v)]</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>[0.95, 0.05]</td>
<td>[0.5, 0.05]</td>
<td>[, 0.05]</td>
<td>[0.45, 0.45]</td>
</tr>
<tr>
<td>Price</td>
<td>–</td>
<td>–</td>
<td>[(i), (v)]</td>
<td>[(v), (v)]</td>
<td>[(v), (v)]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>[, 0.05]</td>
<td>[0.95, 0.05]</td>
<td>[0.75, 0.10]</td>
<td></td>
</tr>
<tr>
<td>Costs</td>
<td>–</td>
<td>–</td>
<td>[(vi), (i)]</td>
<td>[(vi), (v)]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>[0.05, –]</td>
<td>[0.05, 0.05]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth</td>
<td>–</td>
<td>–</td>
<td>[(v), (v)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>[0.05, 0.80]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 the 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.
Table 3: This table reports $T_N$-values (upper triangle part) and their converted pseudo p-values ($\tilde{p} = T_N/(3.3260)$) (lower triangle part) when 100 Max linear $\chi^2$ tests were applied to each paired data.

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

7 Conclusions

In general, it is not easy to derive the limiting distribution for the maxima of several measures. Also combining measures with the same or similar functional properties is 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 $\chi^2$ 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 $\chi^2$ test can easily be implemented in any statistical software packages. We note that the new $\chi^2$ 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

Proof of Lemma 2.1. Since \( W_n \) and \( W'_n \) have the same joint distributions, it suffices to show that \( W_n \overset{L}{\to} W \) and that \( W_n \) and \( W'_n \) are asymptotically independent.

Let \( E(1) \leq E(2) \leq \cdots \leq E(n) \) be the order statistics of \( E_1, E_2, \cdots, E_n \). Then it is well known that

\[
(U_1, U_2, \cdots, U_n) \overset{d}{=} (1 - \exp(-E(1)), 1 - \exp(-E(2)), \cdots, 1 - \exp(-E(n))).
\]  

(8.1)

Further, we have from Eq (3.5.6) on page 34 of Balakrishnan and Cohen (1991):

\[
(E_1, E_2, \cdots, E_n) \overset{d}{=} \left(\sum_{j=1}^{1} \frac{E_j}{n - j + 1}, \sum_{j=1}^{2} \frac{E_j}{n - j + 1}, \cdots, \sum_{j=1}^{n} \frac{E_j}{n - j + 1}\right).
\]  

(8.2)

For each fixed \( i \), since \( E(i) \to 0 \) in probability and \( nE(i) \overset{L}{\to} \sum_{j=1}^{i} E_j \) as \( n \) tends to infinity, we have

\[
W_n \overset{d}{=} n\left(1 - \exp(-E(1)), 1 - \exp(-E(2)), \cdots, 1 - \exp(-E(k)) \right)
= n(E_1, E_2, \cdots, E_k) + O_p\left(\frac{1}{n}\right)
\overset{L}{\to} (w_1, w_2, \cdots, w_k) \quad \text{as} \quad n \to \infty.
\]

Set

\[
V_n = n\left(1 - \exp(-E(1)), 1 - \exp(-E(2)), \cdots, 1 - \exp(-E(k)) \right)
\]

and

\[
V'_n = n\left(\exp(-E(n)), \exp(-E(n-1)), \cdots, \exp(-E(n-k+1)) \right).
\]

From (8.1), 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

\[
V''_n = n\left(\exp(-(E(n) - E(k))), \exp(-(E(n-1) - E(k))), \cdots, \exp(-(E(n-k+1) - E(k))) \right).
\]
It follows from (8.2) 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 2.2.** Let $F^-$ denote the generalized inverse of $F$ by

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

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

$$(Z(1), Z(2), \cdots, Z(n)) \overset{d}{=} (F^-(U(1)), F^-(U(2)), \cdots, F^-(U(n))).$$

From condition (2.4) we have

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

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 $1 \leq i \leq k$. Thus we have from Lemma 2.1

$$\left( nZ(1), nZ(2), \cdots, nZ(k), \frac{n}{Z(n)}, \cdots, \frac{n}{Z(n-k+1)} \right) \overset{d}{=} (W_n, W'_n) + o_p(1) \overset{L}{\to} (W, W'),$$

which completes the proof. \[ \square \]

**Proof of Lemma 2.4.** Without loss of generality assume $\mathbf{\mu} = (0, \cdots, 0)$. Let $t = (t_1, \cdots, t_p) \in \mathbb{R}^p$. It suffices to show that $\tilde{\xi}t^\tau$ and $M_\eta$ are asymptotically independent for every nonzero $t$. To this end, for every vector $(x_1, \cdots, x_q)$ with $G(x_1, \cdots, x_q) \in (0, 1)$, set

$$B_n = \{ \max_{1 \leq i \leq q} \eta_{ki} - b_{kn} \leq x_k, \quad k = 1, \ldots, q \}.$$ 

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

31
Set $A_n = \{ \eta_k \leq a_{kn} + b_{kn} x_k, \ k = 1, \ldots, q \}$. Then we have $P(A_n) = 1 + O\left(\frac{1}{n}\right)$ from (2.5).

Let $\{\tau_n', \ 1 \leq j \leq n\}$ denote a sequence of $n$ iid random variables distributed the same as the conditional distribution of $\xi t^\tau$ given $B_n$. Then we have

\[
P\left(\frac{\sqrt{n} \xi t^\tau}{\|B_n\|} \leq y\right) = P\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \tau_j' \leq y\right)
\]

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

\[
|E(\tau_1')| = \left| \frac{1}{P(A_n)} E(\tau_1 I(A_n)) \right| = \left| \frac{1}{P(A_n)} E(\tau_1 I(\bar{A}_n)) \right| = \frac{1}{P(A_n)} \sqrt{E(\tau_1^2 I(\bar{A}_n))} \sqrt{P(\bar{A}_n)} = o(n^{-1/2})
\]

and

\[
E(\tau_1')^2 = \frac{1}{P(A_n)} E(\tau_1^2 I(A_n)) = \frac{1}{P(A_n)} (\sigma_t^2 - E(\tau_1^2 I(\bar{A}_n))) = \sigma_t^2 + o(1).
\]

Note that

\[
\sqrt{n} \xi t^\tau = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \tau_j^' \xrightarrow{d} N(0, \sigma_t^2).
\]

To conclude the theorem, from (8.3) 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

\[
\frac{1}{n} \sum_{j=1}^{n} E(\tau_j'^2 I(|\tau_j'| > \epsilon \sqrt{n} \sigma_t)) \leq \frac{1}{P(A_n)} E(\tau_1^2 I(|\tau| > \epsilon \sqrt{n} \sigma_t)) \to 0
\]

for every $\epsilon > 0$. The last step is obvious since $E(\tau^2_1) \in (0, \infty)$. This complete the proof. \(\Box\)

**Proof of Theorem 2.5** 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

\[
\sqrt{n}(\xi_1, \xi_2) \xrightarrow{d} N(0, \Sigma_1)
\]

and

\[
\sqrt{n}(\frac{1}{n} \sum_{k=1}^{n} \xi_{1k} \xi_{2k} - r, \frac{1}{n} \sum_{k=1}^{n} \xi_{1k}^2 - 1, \frac{1}{n} \sum_{k=1}^{n} \xi_{2k}^2 - 1) \xrightarrow{d} N(0, \Sigma_2),
\]

where

\[
\Sigma_1 = \begin{pmatrix}
1 & r \\
r & 1
\end{pmatrix}
\]

32
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 (8.4), (8.5) and Taylor’s expansion that
\[
r_n - r = \frac{1}{n} \sum_{k=1}^{n} \xi_{1k} \xi_{2k} - r - \frac{r}{2n} \sum_{k=1}^{n} (\xi_{1k}^2 - 1) - \frac{r}{2n} \sum_{k=1}^{n} (\xi_{2k}^2 - 1) + o_p(n^{-1/2}),
\]
which together with (8.5) yields
\[
\sqrt{n}(r_n - r) \xrightarrow{\cal L} N(0, \sigma^2).
\]
The rest of the proof follows from Lemma 2.4 directly.

The following result is closely related to Theorem 3.1 in Zhang (2008). It shows the properties of a random sample of \((X, Y)\) with unit Fréchet margins.

**Proposition 8.1** 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
\[
2a_n q_n \xrightarrow{\cal L} \chi_4^2,
\]
where \(a_n/n \to 1\) as \(n \to \infty\), and \(\chi_4^2\) is a \(\chi^2\) 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
\((Z_{(\text{rank}[X_1])}^X, \ldots, Z_{(\text{rank}[X_n])}^X)\) and \((Z_{(\text{rank}[Y_1])}^Y, \ldots, Z_{(\text{rank}[Y_n])}^Y)\) are independent.

We now derive the joint distribution of \((Z_{(\text{rank}[X_1])}^X, \ldots, Z_{(\text{rank}[X_n])}^X)\). We have
\[
P\{Z_{(\text{rank}[X_1])}^X \leq x_1, \ldots, Z_{(\text{rank}[X_n])}^X \leq x_n\}
= \sum_{(k_1, \ldots, k_n) \in \mathcal{I}} P\{Z_{(k_1)}^X \leq x_1, \ldots, Z_{(k_n)}^X \leq x_n, \text{rank}[X_1] = k_1, \ldots, \text{rank}[X_n] = k_n\}
= \sum_{(k_1, \ldots, k_n) \in \mathcal{I}} P\{Z_{(k_1)}^X \leq x_1, \ldots, Z_{(k_n)}^X \leq x_n\} \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 \mathcal{I}} P\{Z_{(k_1)}^X \leq x_1, \ldots, Z_{(k_n)}^X \leq x_n\}
where $\mathcal{T}$ is the collection of permutations of \{1, ..., $n$\}. For any $(k_1, ..., k_n) \in \mathcal{T}$, the joint density function of $(Z_{(k_1)}^X, ..., Z_{(k_n)}^X)$ is

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

Now define

$$\mathcal{D} = \{(t_1, ..., t_n) \mid 0 < t_i \leq x_i, \ i = 1, ..., n\},$$

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

Then we have

$$\sum_{(k_1, ..., k_n) \in \mathcal{T}} P\{Z_{(k_1)}^X \leq x_1, ..., Z_{(k_n)}^X \leq x_n\}$$

$$= \sum_{(k_1, ..., k_n) \in \mathcal{T}} \int_{\mathcal{A}(k_1, ..., k_n)} n! f_Z(t_1) f_Z(t_2) \cdots f_Z(t_n) dt_1 ... dt_n$$

$$= n! \int_{\mathcal{D}} f_Z(t_1) f_Z(t_2) \cdots f_Z(t_n) dt_1 ... dt_n = n! \prod_{i=1}^nP\{Z_i^X \leq x_i\}.$$

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

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

which results in a zero probability event. Therefore, we obtain the joint probability of

$$(Z_{\text{rank}[X_1]}^X, ..., Z_{\text{rank}[X_n]}^X, Z_{\text{rank}[Y_1]}^Y, ..., Z_{\text{rank}[Y_n]}^Y)$$

from

$$P\{Z_{\text{rank}[X_1]}^X \leq x_1, ..., Z_{\text{rank}[X_n]}^X \leq x_n, Z_{\text{rank}[Y_1]}^Y \leq y_1, ..., Z_{\text{rank}[Y_n]}^Y \leq y_n\}$$

$$= P\{Z_1^X \leq x_1\} P\{Z_2^X \leq x_2\} \cdots P\{Z_n^X \leq x_n\} P\{Z_1^Y \leq y_1\} P\{Z_2^Y \leq y_2\} \cdots P\{Z_n^Y \leq y_n\}$$

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

**Proof of Theorem 4.2.** By Lemma 4.1, we have $\{Z_{\text{rank}[X_1]}^X, ..., Z_{\text{rank}[X_n]}^X\}$ and $\{Z_{\text{rank}[Y_1]}^Y, ..., Z_{\text{rank}[Y_n]}^Y\}$ are two independent samples from two independent unit Fréchet random variables. Then we complete the proof by applying Proposition 8.1 \hfill \Box

**Proof of Theorem 4.3.** For any fixed $n$, conditional on $\mathcal{G}$, $\tilde{P}_{n_1}^{\mathcal{G}_j}$, $j \geq 1$ are bounded iid random variables with a conditional expectation $E(\tilde{P}_{n_1}^{\mathcal{G}_j})$. Therefore, the theorem follows directly from the strong law of large numbers. \hfill \Box
References


