Robust Copula Dependence for Multivariate and Conditional Dependence with Applications

Yi Li
Northeastern University
Department of Mathematics
360 Huntington Ave., Boston
USA

A. Adam Ding
Northeastern University
Department of Mathematics
360 Huntington Ave., Boston
USA
a.ding@neu.edu

Abstract: Measures of statistical dependence is of great importance for machine learning and statistical models. Recently, a new measure, the robust copula dependence (RCD) is shown to be equitable in treating dependence of linear and nonlinear relationships. The paper propose extensions of RCD to multivariate and conditional cases, which is crucial for many applications. We study the theoretical and empirical properties of the extended RCD. We successfully apply to several example applications including learning delayed time in nonlinear systems, independence testing with mixture alternatives and feature selection.

Key–Words: Multivariate dependence, robust-equitable, independence testing

1 Introduction

The measure of dependence among random variables is an important topic in statistics and machine learning. In many real word applications, the solutions of the problem are based on the usage of statistical dependence measures [de Siqueira Santos *et al.*, 2013; Jiang *et al.*, 2015]. In particular, recently there are proposals of equitable dependence measures which assigns similar values to equally noisy (linear or nonlinear) relationships [Reshef *et al.*, 2011; Murrell *et al.*, 2016; Ding *et al.*, 2017]. The equitability concept is first developed in the bivariate cases. However, applications in a wide range of studies require the dependence measures for multidimensional and/or conditional scheme. In this paper, we propose extension of a recent equitable dependence measure, the robust copula dependence (RCD), to the multivariate and conditional cases. The contributions of this paper are:

1 We summarize the two directions for extending the multivariate dependence measure and proposed the multivariate version of the robust copula dependence. The properties of the multivariate RCD and its estimator are discussed.

2 We provide the conditional extension of the robust copula dependence of which the properties are discussed.

3 We extend the concept of robust equitability to multidimension and apply RCD in the independence test with mixture alternatives.

4 We apply the conditional RCD in the feature selection problem, and proposed a conditional dependence measure based filter-type feature selection method.

5 We apply the RCD in the chaotic and delayed systems to reconstruct the delay time, which is of great importance in the study of time-delayed system.

The structure of this paper is as follows: Section 2 defines the multivariate RCD, and discuss several basic important properties based on the generalization of Renyi's axioms on bivariate case[Rényi, 1959; Wolff, 1980]. We then discuss the conditional version of RCD. In section 3, we consider the estimation of RCD as the statistical functionals of the copula density function. Its relationship with statistical test is also discussed here. The numerical examples and some applications of the multidimensional and conditional RCD are presented in section 4. Finally, we conclude this paper in section 5.

2 Methodology

Roughly speaking, the RCD is defined as the distance between the copula density of the data and the independent copula (density). The properties, such as equitability, for the bivariate RCD, and its comparison with mutual information and other dependence measures are studied in Chang *et al.* [2016]; Ding *et al.* [2017]. Here we first review the definition of RCD, then we define the multivariate and conditional extension of RCD.

2.1 The Robust Copula Dependence

The copula is a multivariate probability distribution for which the marginal probability distribution of each variable is uniform. For two random variables X and Y, let $U = F_X(X), V = F_Y(Y)$, where F_X and F_Y are the cumulative distribution functions (CDFs) of X and Y. Then U and V both follow the uniform distribution on $I = [0, 1]$. The joint density of U and V is the copula density for X and Y.

Definition 1 (Copula Distance and RCD) *Let $c(u, v)$ denote the copula density for the random vector $\vec{Z} = (X, Y)$. The* copula distance *between X and Y is*

$$CD_\alpha = \frac{1}{2}\iint_{I^2} |c(u, v) - 1|^\alpha du dv, \quad \alpha > 0. \quad (1)$$

In particular, we call CD_1 the robust copula dependence *and denote it by* $\mathsf{RCD}(X, Y)$.

2.2 Multivariate RCD and Its Properties

The above RCD measures the dependence between the two components of the bivariate vector $\vec{Z} = (X, Y)$. For a multivariate vector, there are two possible dependence that can be studied. The first type is the dependence *among* all components of a p-dimensional vector $\vec{X} = (X_1, \cdots, X_p)$. That is, we want to measure the dependence among the p random variables $X_1, \cdots, X_p$, which we call an internal-type of multivariate dependence. The second type is the dependence *between* two sub-vectors of the original p-dimensional vector, which we call an external-type of dependence. The second type of dependence is more useful in regression/classification applications, where one sub-vector represent responses and one sub-vector represent the regressors. We discuss both type of extensions of RCD in the following.

2.2.1 Internal RCD

The bivariate RCD is based on the distance between the bivariate joint copula density and the bivariate independence copula. Similarly for a p-dimensional random vector $\vec{X} = (X_1, \cdots, X_p)$, we define the internal RCD using the distance between the p-dimensional joint copula density and the p-dimensional independence copula.

Definition 2 (Multivariate RCD) *Let $\vec{X} = (X_1, \cdots, X_p)$ be a p-dimensional multivariate random vector with the copula density $c(\vec{u}) = c(u_1, \cdots, u_p)$. The* multivariate robust copula distance *of $\vec{X}$ is*

$$\begin{aligned}\mathsf{RCD}(\vec{X}) &:= \frac{1}{2}\int_{\mathbb{I}} |c(\vec{u}) - 1| d\vec{u} \\ &= \frac{1}{2}\int_{\mathbb{I}} |c(u_1, \cdots, u_p) - 1| du_1 \cdots du_p.\end{aligned} \quad (2)$$

The normalization factor is of the same value $\frac{1}{2}$ in both bivariate and multivariate cases. After some basic algebra, one can show that RCD has the following two alternative forms, which can be used for estimation.

Proposition 3 (Alternative forms) *Let $\vec{X} = (X_1, \cdots, X_p)$ be a p-dimensional multivariate random vector with the copula density $c(\vec{u}) = c(u_1, \cdots, u_p)$. Then,*

$$\begin{aligned}\mathsf{RCD}(\vec{X}) &= \int_{c>1} |c(u_1, \cdots, u_p) - 1| du_1 \cdots du_p \\ &= \int_{c\leq 1} |c(u_1, \cdots, u_p) - 1| du_1 \cdots du_p.\end{aligned} \quad (3)$$

Remark. The two alternative forms in Proposition 3 could also be written as $\int [c(u_1, \cdots, u_p) - 1]_+ du_1 \cdots du_p$ or $\int [c(u_1, \cdots, u_p) - 1]_- du_1 \cdots du_p$, where f_+ and f_- denotes the positive and negative parts of f.

We can also define the RCD on the raw scale (for raw data $\vec{X}$) directly through the distance between the joint density and the independent density. The RCD definitions coincide on the raw scale and on the copula scale, as summarised in Proposition 4.

Proposition 4 (Coincide in two scales) *Let $\vec{X} = (X_1, \cdots, X_p)$ be a p-dimensional multivariate random vector with the joint density $f(x_1, \cdots, x_p)$, marginal densities $f(x_1), \cdots, f(x_p)$, and the copula density $c(u) = c(u_1, \cdots, u_p)$. Then,* $\mathsf{RCD}(\vec{X})$ *is*

$$\begin{aligned}&\frac{1}{2}\int_{\mathbb{I}} |c(u_1, \cdots, u_p) - 1| du_1 \cdots du_p \\ =&\frac{1}{2}\int_{\Omega} |f(x_1, \cdots, x_p) - f(x_1) \cdots f(x_p)| dx_1 \cdots dx_p\end{aligned} \quad (4)$$

While the definition on raw scale may be easier to understand, the copula scale definition has following advantages. From a theoretical point of view, the analysis is cleaner when we focus exclusively on the dependence structure, the copula. From a practical point of view, the copula scale definition allows us to avoid dealing with as many as $p+1$ density estimation needed in the raw scale definition.

Furthermore, RCD has some theoretical properties desirable for a general multidimensional dependence measure. Proposition 5 summarizes such properties of RCD based on the Wolff's extension of Renyi's axioms[Rényi, 1959; Wolff, 1980].

Proposition 5 (Woff's extension) *Let* $X = (X_1, \cdots, X_p)$ *be a p-dimensional multivariate random vector with copula density $c(u) = c(u_1, \cdots, u_p)$. Then,*

a. $\mathsf{RCD}(\vec{X}) \geq 0$ *and* $\mathsf{RCD}(\vec{X}) = 0$ *if and only if* $X_1, \cdots, X_p$ *are independent.*

b. $\mathsf{RCD}(\sigma(\vec{X})) = \mathsf{RCD}(\vec{X})$*, where σ is a permutation.*

c. $\mathsf{RCD}(\vec{X})$ *is invariant under strictly increasing transformations of the components of* $\vec{X}$.

d. $\mathsf{RCD}(\vec{X})$ *is invariant under strictly decreasing transformations of the components of* $\vec{X}$.

e. Let $\vec{Y} = (Y_1, \cdots, Y_s)$, $\vec{Z} = (Z_1, \cdots, Z_t)$ *and* $\vec{X} = (\vec{Y}, \vec{Z})$*, where $p = s + t$. It follows that* $\mathsf{RCD}(\vec{X}) \leq \mathsf{RCD}(\vec{Y}) + \mathsf{RCD}(\vec{Z}) + \frac{1}{2}\int |c_X - c_Y - c_Z + 1|$.

f. Let c_n be a sequence of copula density which converge almost surely to c, then $\mathsf{RCD}(c_n) \to \mathsf{RCD}(c)$.

Remark. Property (a) is equivalent to the statement that $\mathsf{RCD}(C) = 0$ if and only if $C = \Pi$, where C is the copula function for $\vec{X}$, and Π is the p-dimensional independence copula. Property (c) and (d) are the direct results of the use of copula.

One of the exciting properties of the bivariate RCD is the robust equitability [Chang *et al.*, 2016; Ding *et al.*, 2017]. This concept can be generalized to multi-dimensional case. Let $\Pi(u_1, ..., u_1) = u_1 u_2 ... u_p$ be the uniform (independent) copula on the unit hyper-cube I^p, to which we add a ϵ proportion of deterministic signal C_s, where C_s is a singular copula [Nelsen, 2006]. An equitable dependence measure should give the same value ϵ regardless of which type of deterministic signal C_s is used.

Definition 6 *A dependence measure $D(\vec{X})$ is robust-equitable if and only if $D(\vec{X}) = \epsilon$ whenever $\vec{X}$ follows a distribution with copula $C = \epsilon C_s + (1 - \epsilon)\Pi$, for any singular copula C_s.*

With the above definition, we have the following property of RCD: when a deterministic signal is hidden in the background noises, RCD only depends on the noise level of the data.

Proposition 7 RCD *is robust equitable.*

2.2.2 External RCD

In some applications, like feature selection, we are not purely interested in the internal dependence measure of the multivariate vector. Rather, we care more about how strong two random vectors $\vec{X}$ and $\vec{Y}$ are related. Thus, this leads to the concept of external RCD.

Definition 8 (External RCD) *Let* $\vec{X} = (X_1, \cdots, X_p)$ *and* $\vec{Y} = (Y_1, \cdots, Y_q)$ *be p and q-dimensional multivariate random vectors with the copula density $c(\vec{u}) = c(u_1, \cdots, u_p)$ and $c(\vec{v}) = c(v_1, \cdots, v_q)$ respectively. The* multivariate robust copula distance *between $\vec{X}$ and $\vec{Y}$ is*

$$\mathsf{RCD}_*(\vec{X}, \vec{Y}) := \frac{1}{2}\int_{\mathbb{I}} |c(\vec{u}, \vec{v}) - c(\vec{u})c(\vec{v})| d\vec{u} d\vec{v}. \quad (5)$$

Remark. To distinguish from the internal RCD, we denote the external measure as RCD_* with a subscript $*$. For application such as the regression analysis, the response is one random variable Y. That can be considered as a special case, then $V = F_Y(Y)$ follows the uniform distribution with density 1 and we have

$$\mathsf{RCD}_*(\vec{X}, Y) := \frac{1}{2}\int_{\mathbb{I}} |c(\vec{u}, v) - c(\vec{u})| d\vec{u} dv. \quad (6)$$

Results similar to Proposition 3 and 4 hold for RCD_* as summarize below.

Proposition 9 *Let* $\vec{X} = (X_1, \cdots, X_p)$ *and* $\vec{Y} = (Y_1, \cdots, Y_q)$ *be p and q-dimensional multivariate random vectors with the joint density $f(x_1, \cdots, x_p, y_1, \cdots, y_q)$, marginal densities $f(x_1), \cdots, f(x_p)$, $f(y_1), \cdots, f(y_q)$, the copula densities $c(\vec{u}) = c(u_1, \cdots, u_p)$ and $c(\vec{v}) = c(v_1, \cdots, v_q)$ respectively. Then we have*

$$\begin{aligned} \mathsf{RCD}_*(X, Y) &:= \int [c(\vec{u}, \vec{v}) - c(\vec{u})c(\vec{v})]_+ d\vec{u} d\vec{v} \\ &= [c(\vec{u}, \vec{v}) - c(\vec{u})c(\vec{v})]_- d\vec{u} d\vec{v}. \end{aligned} \quad (7)$$

$$\int_{\Omega}|f(x_1,\cdots,y_q)-f(x_1,\cdots,x_p)f(y_1,\cdots,y_q)|d\vec{x}d\vec{y} = \int_{\mathbb{I}}|c(\vec{u},\vec{v})-c(\vec{u})c(\vec{v})|d\vec{u}d\vec{v} \tag{8}$$

2.3 The Conditional RCD

Conditional dependence is of great importance in many application, such as causal inference, feature selection and graphical models, etc. In this part, we extend the RCD by considering a third random vector $\vec{Z}$, and defines the conditional RCD of $\vec{X}$ and $\vec{Y}$ given $\vec{Z}$ to measure the conditional dependence.

Definition 10 (Conditional RCD) *Let $\vec{X} = (X_1,\cdots,X_p)$, $\vec{Y}=(Y_1,\cdots,Y_q)$, $\vec{Z}=(Z_1,\cdots,Z_r)$ be p, q, and r-dimensional multivariate random vectors with the copula density $c(\vec{u}) = c(u_1,\cdots,u_p)$, $c(\vec{v}) = c(v_1,\cdots,v_q)$, and $c(\vec{w}) = c(w_1,\cdots,w_r)$. The* conditional robust copula distance *of $\vec{X}$ and $\vec{Y}$ given $\vec{Z}$ is defined as:*

$$\begin{aligned}&\mathsf{RCD}(\vec{X},\vec{Y}|\vec{Z})\\ =\ &\tfrac{1}{2}\textstyle\int_{\mathbb{I}}|c(\vec{u},\vec{v},\vec{w})-c(\vec{u},\vec{w})c(\vec{v},\vec{w})|d\vec{u}d\vec{v}d\vec{w}.\end{aligned} \tag{9}$$

We have the corresponding version of Proposition 3 for the conditional RCD.

Proposition 11 (Coincide in two scales) *Let $\vec{X} = (X_1,\cdots,X_p)$, $\vec{Y}=(Y_1,\cdots,Y_q)$, $\vec{Z}=(Z_1,\cdots,Z_r)$ be p, q, and r-dimensional multivariate random variables with the conditional joint density $f(\vec{x},\vec{y}|\vec{z})$, conditional marginal density $f(\vec{x}|\vec{z})$, $f(\vec{y}|\vec{z})$, and the copula density $c(\vec{u}) = c(u_1,\cdots,u_p)$, $c(\vec{v}) = c(v_1,\cdots,v_q)$, and $c(\vec{w}) = c(w_1,\cdots,w_r)$. Then,*

$$\begin{aligned}&\int_{\mathbb{I}}|c(\vec{u},\vec{v},\vec{w})-c(\vec{u},\vec{w})c(\vec{v},\vec{w})|d\vec{u}d\vec{v}d\vec{w}\\ =&\mathbb{E}_z[\int\int|f(\vec{x},\vec{y}|\vec{z})-f(\vec{x}|\vec{z})f(\vec{y}|\vec{z})|d\vec{x}d\vec{y}].\end{aligned} \tag{10}$$

3 Statistical Properties

In this section, we explore the basic property of the estimator for RCD and discuss the relationship between RCD with the independence test.

3.1 Statistical Error

We first consider the internal RCD. If we have a certain density estimator $\hat{c}$, we could plug it in the integration with respect to the empirical distribution function, which yields the following estimator.

$$\widehat{\mathsf{RCD}}(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n}|\frac{1}{\hat{c}(\vec{u}_i)}-1|. \tag{11}$$

Remark. Similarly, we could have other possible forms of the estimator based on the alternatives of RCD: $\widehat{\mathsf{RCD}}(\vec{X}) = \frac{1}{2n}\sum_{i=1}^{n}[\frac{1}{\hat{c}(\vec{u}_i)}-1]_+$, and $\widehat{\mathsf{RCD}}(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n}[\frac{1}{\hat{c}(\vec{u}_i)}-1]_-$.

We now consider the consistency result of the estimator (11). The following result is a direct consequence of the slutsky theorem [Ferguson, 1996].

Theorem 12 *Assume the p-dimensional random variable X has the copula density c bounded below by $\epsilon_0 > 0$. Let S be a compact proper subset of $\mathbb{I}^p$. For a consistent density estimator $\hat{c}$, say, the kernel density estimator, we have*

$$|\widehat{\mathsf{RCD}}(\vec{X})-\mathsf{RCD}(\vec{X})| \to 0 \tag{12}$$

as $n\to\infty$.

Similar results also hold for the estimator of external RCD_*, which is $\widehat{\mathsf{RCD}_*}(\vec{X},\vec{Y}) = \frac{1}{2n}\sum_{i=1}^{n}|1-\frac{\hat{c}(\vec{u}_i)\hat{c}(\vec{v}_i)}{\hat{c}(\vec{u}_i,\vec{v}_i)}|$, and the estimator of conditional RCD, which is $\widehat{\mathsf{RCD}}(\vec{X},\vec{Y}|\vec{Z}) = \frac{1}{2n}\sum_{i=1}^{n}|\frac{c(\vec{u}_i,\vec{w}_i)c(\vec{v}_i,\vec{w}_i)}{c(\vec{u}_i,\vec{v}_i,\vec{w}_i)}-1|$.

3.2 Link with Statistical Test

Even though the idea of equitability mainly aims at comparing the strength of hidden signal, the RCD is also closely related to the independence test. Suppose two random vectors $\vec{X}$ and $\vec{Y}$ have marginal probability measures $\mathbb{P}_x$ and $\mathbb{P}_y$, with the joint distribution $\mathbb{P}_{xy}$. Considering the following independence test between $\vec{X}$ and $\vec{Y}$

$$H_0 :(\vec{X},\vec{Y}) \sim \mathbb{P}_x\mathbb{P}_y \tag{13}$$

$$H_1 :(\vec{X},\vec{Y}) \sim \mathbb{P}_{xy} \neq \mathbb{P}_x\mathbb{P}_y. \tag{14}$$

It turns out that the RCD quantifies the difficulty of this test and provides an upper bound of the power.

Theorem 13 *Let $\vec{X}$ and $\vec{Y}$ have marginal probability measures $\mathbb{P}_x$ and $\mathbb{P}_y$, $\mathbb{P}_0 = \mathbb{P}_x\mathbb{P}_y$ and $\mathbb{P}_1 = \mathbb{P}_{xy}$. For any test $\phi = \phi(\vec{X},\vec{Y}) \in \{0,1\}$ that indicates which hypothesis in (13) should be true, we have*

$$\texttt{power} := 1-\beta \le \alpha + 2\mathsf{RCD}(\vec{X},\vec{Y}), \tag{15}$$

where $\alpha = \mathbb{P}_0(\phi = 1)$ *and* $\beta = \mathbb{P}_1(\phi = 0)$ *are the Type-I and Type-II error respectively.*

Theorem 13 also shows that the total error that a test made is lower bounded by $1-2\text{RCD}(\vec{X}, \vec{Y})$. This means the test in (13) becomes easier (smaller lower bound of the total error) when RCD is large (the distributions are apart). The proof of the above theorem can be found in Appendix.

4 Numerical Examples

In this section, we consider three applications of the extended RCD. First, we use RCD for independence testing. We empirically investigate the independence testing power performance of RCD with other dependence measures (Pearson's correlation of coefficient, distance correlation [Székely *et al.*, 2009], mutual information [Blumentritt and Schmid, 2012; Joe, 1989], and some kernel based dependence measure [Gretton *et al.*, 2005; Fukumizu *et al.*, 2007; Póczos *et al.*, 2012; J Reddi and Póczos, 2013]). Second, we provide a feature selection method based on the conditional RCD. Last, we provide an application of the RCD to successfully reconstruct the time lag in nonlinear chaotic time-delayed system.

4.1 Statistical Power Analysis

In this part, we consider the power performance for the independence test under mixture alternative as defined in robust-equitability in (6). Specifically, six nonlinear functional types (circle, parabola, sine wave, cross, spiral, and Lorenz system) are considered as in Figure 1. Each signal is mixed with independent uniform noise with various mixture portion.

Within each of the six scenarios, we compare testing of independence with different dependence measures including the Pearson's correlation coefficient, distance correlation [Székely *et al.*, 2009], mutual information [Blumentritt and Schmid, 2012], RCD [Li and Ding, 2017] and HSIC [Gretton *et al.*, 2005]. The result of the power analysis is presented in Figure 2. As we can see from the plot, RCD has the best power to test the independence over all the six scenarios, while mutual information also performs well in this comparison. The other three dependence measures are unable to detect hidden signals in the mixture model.

4.2 Feature Selection

We now show how the conditional RCD could be applied in the machine learning problem of feature selection. We use it in the feature selection procedure based on dependence maximization [Song *et al.*, 2012] as following. In the first step, we find the feature that maximize the RCD with the target variable Y. Then in each iteration, the RCD, conditional on all the features that have been selected, is computed between Y and each feature which has not been selected. The feature who maximize the conditional RCD is selected in this iteration. The iterations are repeated until the desired number of features are selected.

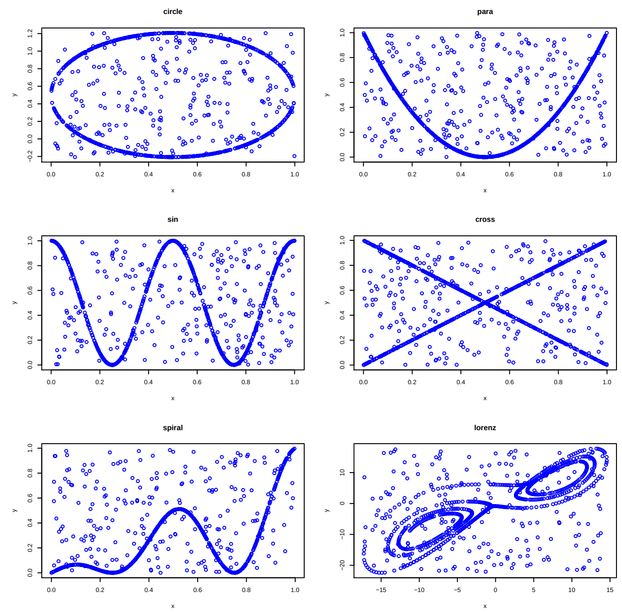

Figure 1: Six different functional types (circle, parabola, sine wave, cross, spiral, and Lorenz system) are considered for the comparison of the power analysis of the independence test. Each signal is mixed with independent uniform noise with various mixture portion.

The conditional RCD based feature selection method is tested on four different datasets from the UCI Machine Learning Repository[1] (see Figure 3). The selected features are used for the prediction of the target variables. The 5-fold cross-validated mean squared error is computed with the spline regression model. The method is compared with the forward and backward dependence maximization methods with the Hilbert-Schmidt independence criterion, which is suggested in Song *et al.* [2012]. The MSE curve in Figure 3 shows that RCD is very competitive against other dependence measures, have the best (smallest MSE) in most cases.

4.3 Recovery of Delayed Time

Time-delayed systems, which arise from a wide range of natural phenomenons, have been received a lot

[1] `http://archive.ics.uci.edu/ml/index.php`

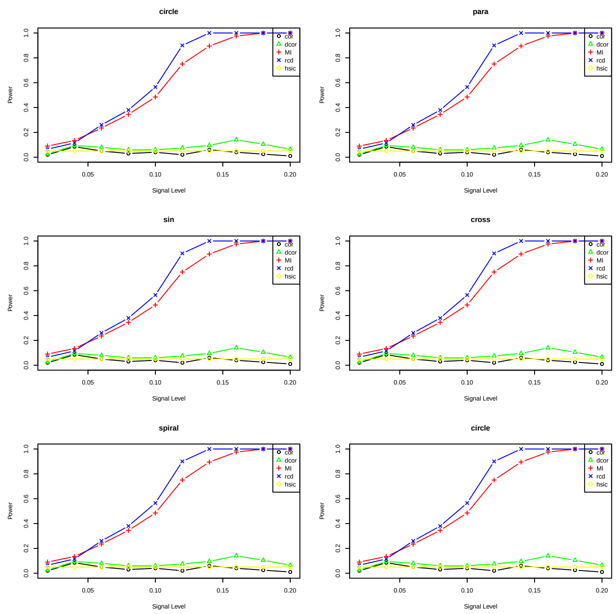

Figure 2: Power curves for the independence test with mixture alternatives of five dependence measures (Pearson's correlation coefficient, distance correlation, mutual information, RCD and HSIC). Results are based on 200 replications in the null and alternative cases. The sample size is $n = 300$, and ten signal levels (portions) between 0 and 0.4. The distance correlation is computed with with `energy` package in R. The mutual information is computed with the `parmigene` package in R.

of attention [Bezruchko *et al.*, 2001; Prokhorov and Ponomarenko, 2005; Prokhorov *et al.*, 2013]. Specifically, the reconstruction of the delayed time of the delay-induced dynamics is one of the central issues in the study of chaotic time series. In this part, we apply our proposed dependence measure RCD to recover the delayed time from delayed series. We consider the following first-order delayed differential equation with a single delay:

$$\epsilon x'(t) = -x(t) + f(x(t-\tau)), \tag{16}$$

where τ is the delayed time, ϵ is a model parameter, and f is a (nonlinear) function. The delayed time τ is prior unknown to us, and our task is to recovery τ from the data generated from the system described in (16). We note that the first order derivative can be discretized with $x'(t) \approx \frac{x(t+\Delta t)-x(t)}{\Delta t}$ where Δt is the sampling time unit. In this way, the discretized delayed time is $d = \frac{\tau}{\Delta t}$.

The central idea of the application is to vary the trial delayed time $\hat{\tau}$, and calculate the dependence between the series of $x(t)$ and $x(t-\hat{\tau})$. If $\hat{\tau}$ hits the true delayed time, the dependence structure will exhibit a large spike, compared with other false delayed time values, as the true system is driven by the current state and the delayed state with the true delayed time.

Procedure 1 Conditional RCD for feature selection

Input: The full set of features $\mathcal{S} = \{X_1, \cdots, X_p\}$ and the target variable Y

Output: The ordered set of features $\mathcal{S}^*$

$\mathcal{S}^* \leftarrow \emptyset$
$I \leftarrow \underset{j}{\operatorname{argmax}}\, \mathrm{RCD}(X_j, Y)$
$\mathcal{S} \leftarrow \mathcal{S} \backslash \{X_j\}$
$\mathcal{S}^* \leftarrow (X_j)$
while $\mathcal{S} \neq \emptyset$ **do**
    $I \leftarrow \underset{j}{\operatorname{argmax}}\, \mathrm{RCD}(\mathcal{S}\backslash\{X_j\}, Y | \mathcal{S}^*)$
    $\mathcal{S} \leftarrow \mathcal{S} \backslash \{X_j\}$
    $\mathcal{S}^* \leftarrow (\mathcal{S}^*, X_j)$
end while

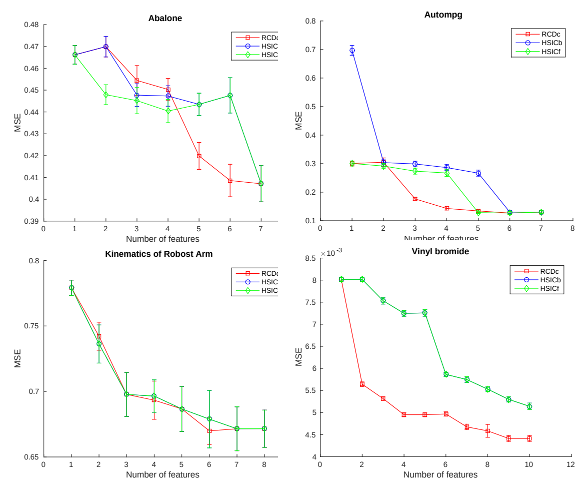

Figure 3: Feature selection results on four different datasets based on the backward/forward dependence maximization feature selection in Song *et al.* [2012] and the proposed method with conditional RCD.

As an example, we consider the Mackey-Glass equation under this type. The system is defined as follow with a nonlinear function f:

$$x'(t) = -bx(t) + \frac{ax(t-\tau)}{1 + x^c(t-\tau)}, \tag{17}$$

where a, b, c are three parameters. As we can see from Figure 4, the left plot is an example of the sampled series of $x(t)$. The first 20000 points of the series are not taken into account to exclude a transient process. In the right plot of Figure 4, the dependence measure is calculated between $x(t)$ and $x(t-\hat{\tau})$ with different choice of the trial delayed time. Result shows that

RCD could successfully detect the relatively large dependence with the correct delayed time.

Similar settings is applicable with coupled systems. For example, let $x(t)$ and $y(t)$ are coupled with the above Mackey-Glass equation with different initial conditions:

$$\begin{cases} x'(t) &= -by(t) + \frac{ay(t-\tau)}{1+y^c(t-\tau)}, \\ y'(t) &= -bx(t) + \frac{ax(t-\tau)}{1+x^c(t-\tau)}. \end{cases}$$

The delayed time could be reconstructed in the similar manner by using the multivariate dependence between the vector $(x(t), y(t))$ and the vector $(x(t-\tau), y(t-\tau))$. The results of the simulation is presented in Figure 5. The left panel is an example of the sampled data for $x(t)$, while similar pattern holds for $y(t)$. After applying the multivariate dependence between $(x(t), y(t))$ and $(x(t-\tau), y(t-\tau))$, the result for various trial delayed time is in the right plot of Figure 5. It can be seen from this plot that the RCD

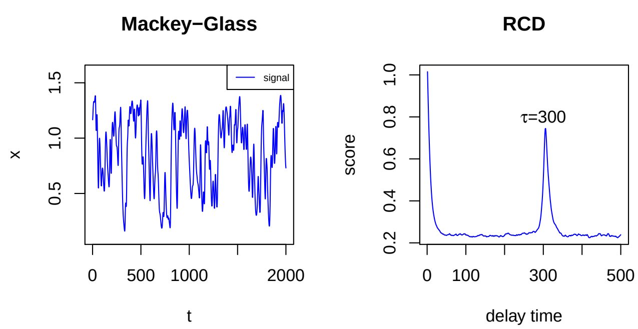

Figure 4: An example of the Mackey-Glass delayed system. The parameters are $a = 0.2, b = 0.1, c = 10, \tau = 300, \Delta t = 1$. The number of points $N = 10000$ with a 20000 washout period. The initial value is chosen with $x(0) = 0.5$. The left plot is an example of the sampled series from the system. The right plot shows the spike of the dependence at the true delayed time 300.

5 Conclusion and Future Work

In this paper, we generalize the robust copula dependence to multivariate and conditional versions. We also discuss the robust equitability for multivariate RCD, and the application of RCD in independence test, feature selection, time-delayed system analysis, etc. Results show that RCD could be used as a reliable tool for detecting strong linear and nonlinear relationships in noisy data sets. It can also be a powerful tool for testing independence and detecting (nonlinear) signals. One open problem arising out of our

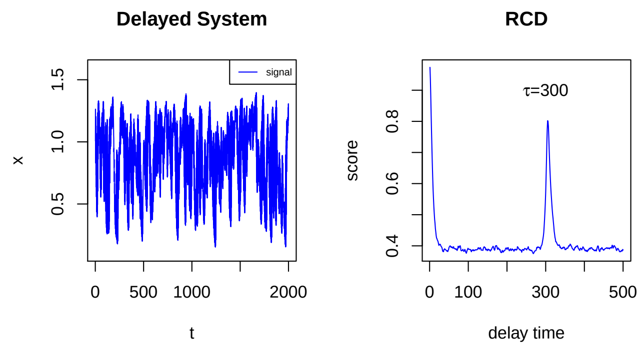

Figure 5: An example of the coupled delayed system with the application of multivariate dependence measures. The parameters are $a = 0.2, b = 0.1, c = 10, \tau = 300, \Delta t = 1$. The number of points $N = 10000$ with a 20000 washout period. The initial value is chosen with $x(0) = 0.4$ and $y(0) = 0.6$. The left plot is an example of the sampled series from the system. The right plot shows the spike of the dependence at the true delayed time 300.

work is how to apply the conditional RCD in causal inference. In addition, the application of (multivariate) RCD in other complex systems, such as the climatology or biology is also of great interests.

Acknowledgments

We would like to acknowledge support of this project from the NSF grant CCF-1442728.

A Proof of Theorem 13

Proof: Let p_0 and p_1 be the density of $\mathbb{P}_0$ and $\mathbb{P}_1$ respectively. For any test ϕ, define the rejection region $R = \{\phi = 1\}$ and let $R^* = \{p_1 \geq p_0\}$. We have

$$\begin{aligned} \alpha + \beta &= 1 + \mathbb{P}_0(R) - \mathbb{P}_1(R) \\ &= 1 + \int_R (p_0 - p_1) \\ &= 1 + \int_{R \cap R^*} (p_0 - p_1) + \int_{R \cap (R^*)^c} (p_0 - p_1) \\ &= 1 - \int_{R \cap R^*} |p_0 - p_1| + \int_{R \cap (R^*)^c} |p_0 - p_1| \\ &= 1 + \int |p_0 - p_1| (1_{R \cap (R^*)^c} - 1_{R \cap R^*}) \\ &\geq 1 - 2\text{RCD}(\vec{X}, \vec{Y}) \end{aligned}$$

The last inequality will become an equality when $R = R^*$.

□

References:

B. P. Bezruchko, A. S. Karavaev, V. I. Ponomarenko, and M. D. Prokhorov. Reconstruction of time-delay systems from chaotic time series. *Phys. Rev. E*, 64:056216, Oct 2001.

Thomas Blumentritt and Friedrich Schmid. Mutual information as a measure of multivariate association: analytical properties and statistical estimation. *Journal of Statistical Computation and Simulation*, 82(9):1257–1274, 2012.

Yale Chang, Yi Li, Adam Ding, and Jennifer Dy. A robust-equitable copula dependence measure for feature selection. In *Artificial Intelligence and Statistics*, pages 84–92, 2016.

Suzana de Siqueira Santos, Daniel Yasumasa Takahashi, Asuka Nakata, and Andre Fujita. A comparative study of statistical methods used to identify dependencies between gene expression signals. *Briefings in Bioinformatics*, 2013.

A Adam Ding, Jennifer G Dy, Yi Li, and Yale Chang. A robust-equitable measure for feature ranking and selection. *The Journal of Machine Learning Research*, 18(1):2394–2439, 2017.

T.S. Ferguson. *A Course in Large Sample Theory*. Chapman & Hall Texts in Statistical Science Series. Taylor & Francis, 1996.

Kenji Fukumizu, Arthur Gretton, Xiaohai Sun, and Bernhard Schölkopf. Kernel measures of conditional dependence. In *NIPS*, volume 20, pages 489–496, 2007.

Arthur Gretton, Ralf Herbrich, Alexander Smola, Olivier Bousquet, and Bernhard Schölkopf. Kernel methods for measuring independence. *J. Mach. Learn. Res.*, 6:2075–2129, December 2005.

Sashank J Reddi and Barnabás Póczos. Scale invariant conditional dependence measures. In *Proceedings of The 30th International Conference on Machine Learning*, pages 1355–1363, 2013.

Bo Jiang, Chao Ye, and Jun S. Liu. Nonparametric k-sample tests via dynamic slicing. *Journal of the American Statistical Association*, 110(510):642–653, 2015.

Harry Joe. Relative entropy measures of multivariate dependence. *Journal of the American Statistical Association*, 84(405):157–164, 1989.

Yi Li and A. Adam Ding. An r package for estimating robust copula dependence. *International Journal of Pure Mathematics*, 4:17–21, 2017.

Ben Murrell, Daniel Murrell, and Hugh Murrell. Discovering general multidimensional associations. *PloS one*, 11(3):e0151551, 2016.

R. B. Nelsen. *An introduction to copulas (Springer series in statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

Barnabás Póczos, Zoubin Ghahramani, and Jeff G. Schneider. Copula-based kernel dependency measures. In *International Conference on Machine Learning*, 2012.

M. D. Prokhorov and V. I. Ponomarenko. Estimation of coupling between time-delay systems from time series. *Phys. Rev. E*, 72:016210, Jul 2005.

M.D. Prokhorov, V.I. Ponomarenko, and V.S. Khorev. Recovery of delay time from time series based on the nearest neighbor method. *Physics Letters A*, 377(43):3106 – 3111, 2013.

A. Rényi. On measures of dependence. *Acta Mathematica Academiae Scientiarum Hungarica*, 10(3-4):441–451, 1959.

David N Reshef, Yakir A Reshef, Hilary K Finucane, Sharon R Grossman, Gilean McVean, Peter J Turnbaugh, Eric S Lander, Michael Mitzenmacher, and Pardis C Sabeti. Detecting novel associations in large data sets. *science*, 334(6062):1518–1524, 2011.

Le Song, Alex Smola, Arthur Gretton, Justin Bedo, and Karsten Borgwardt. Feature selection via dependence maximization. *J. Mach. Learn. Res.*, 13(1):1393–1434, May 2012.

Gábor J Székely, Maria L Rizzo, et al. Brownian distance covariance. *The annals of applied statistics*, 3(4):1236–1265, 2009.

Edward F. Wolff. N-dimensional measures of dependence. *Stochastica*, 4(3):175–188, 1980.