Adjust Pearson's r to Measure Arbitrary Monotone Dependence

Xinbo Ai
School of Intelligent Engineering and Automation
Beijing University of Posts and Telecommunications
Beijing 100876, China
axb@bupt.edu.cn

Abstract

Pearson's r, the most widely-used correlation coefficient, is traditionally regarded as exclusively capturing linear dependence, leading to its discouragement in contexts involving nonlinear relationships. However, recent research challenges this notion, suggesting that Pearson's r should not be ruled out *a priori* for measuring nonlinear monotone relationships. Pearson's r is essentially a scaled covariance, rooted in the renowned Cauchy-Schwarz Inequality. Our findings reveal that different scaling bounds yield coefficients with different capture ranges, and interestingly, tighter bounds actually expand these ranges. We derive a tighter inequality than Cauchy-Schwarz Inequality, leverage it to refine Pearson's r, and propose a new correlation coefficient, i.e., rearrangement correlation. This coefficient is able to capture arbitrary monotone relationships, both linear and nonlinear ones. It reverts to Pearson's r in linear scenarios. Simulation experiments and real-life investigations show that the rearrangement correlation is more accurate in measuring nonlinear monotone dependence than the three classical correlation coefficients, and other recently proposed dependence measures.

1 Introduction

Proposed in the late 19th century, Pearson's r (Pearson, 1896) has been one of the main tools for scientists and engineers to study bivariate dependence during the 20th century. It is remarkably unaffected by the passage of time (Lee Rodgers and Alan Nice Wander, 1988) and still goes strong in the 21st century (Puccetti, 2022). It has been, and probably still is, the most used measure for statistical associations, and generally accepted as *the* measure of dependence, not only in statistics, but also in most applications of natural and social sciences (Tjøstheim, Otneim, and Støve, 2022).

Despite its popularity, Pearson's r has a number of shortcomings, and the most serious issue might be that it can only capture linear dependence, as stated in classical textbooks (Wasserman, 2004) and contemporary literatures (Armstrong, 2019; Tjøstheim, Otneim, and Støve, 2022). The use of Pearson's r has been strongly discouraged for forms of associations other than linear ones (Speed, 2011).

Numerous nonlinear alternative coefficients have been proposed to address this deficiency, such as Spearman's ρ (Spearman, 1904), Kendall's τ (Kendall, 1938), Hilbert-Schmidt Independence Criterion(HSIC) (Gretton et al., 2005), distance correlation(dCor) (Székely, Rizzo, and Bakirov, 2007), Maximal Information Coefficient(MIC) (Reshef et al., 2011), and Chatterjee's ξ (Chatterjee, 2021). Their capture ranges are extending from linear dependence to monotone dependence, and then to non-monotone dependence. Without exception, all these coefficients adopt radically different approaches for nonlinear dependence, rather than following the original way of Pearson's r for a breakthrough.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

In their recent paper titled "*Myths About Linear and Monotonic Associations: Pearson's r, Spearman's ρ, and Kendall's τ*", van den Heuvel and Zhan (2022) challenged the widespread belief that Pearson's r is only a measure for linear dependence, proving this notion to be false. Their findings indicate that Pearson's r should not be ruled out *a priori* for measuring nonlinear monotone dependence. Although this potential has been recognized, the specific approach to using Pearson's r for accurate measurement of nonlinear monotone dependence remains unresolved.

Pearson's r is essentially a scaled covariance, with the renowned Cauchy-Schwarz Inequality as its mathematical foundation. We find that different scaling bounds yield coefficients with different capture ranges, and interestingly, tighter bounds actually expand these ranges. We derive a tighter inequality than Cauchy-Schwarz Inequality, leverage it to adjust Pearson's r to measure nonlinear monotone dependence. The adjusted version of Pearson's r is more accurate in measuring nonlinear monotone dependence than the three classical correlation coefficients, and other recently proposed dependence measures.

2 Methods

2.1 Definitions and notations

Consider real-valued random variables X and Y with cdf's (cumulative distribution functions) F and G respectively. We denote the covariance of X and Y as $\text{cov}(X, Y)$; the variance of X as $\text{var}(X)$, and the variance of Y as $\text{var}(Y)$. We assume that $0 < \text{var}(X) < \infty$, $0 < \text{var}(Y) < \infty$. We define $X^{\uparrow} = F^{-1}(U)$, $X^{\downarrow} = F^{-1}(1-U)$. Here, U is a random variable with the uniform distribution on $(0, 1)$, and F^{-1} is the inverse cdf or quantile function defined as $F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geqslant u\}$, $u \in (0, 1)$. Similarly, $Y^{\uparrow} = G^{-1}(U)$, $Y^{\downarrow} = G^{-1}(1-U)$.

Let $x = (x_1, \cdots, x_n)$ and $y = (y_1, \cdots, y_n)$ be samples of X and Y, each with n elements. Neither x nor y is constant. We denote the sample covariance of x and y as $s_{x,y}$; the sample variance of x as s_x^2; the sample variance of y as s_y^2. We define the increasing and decreasing rearrangement of x as $x^{\uparrow} = (x_{(1)}, x_{(2)}, \cdots, x_{(n)})$ and $x^{\downarrow} = (x_{(n)}, x_{(n-1)}, \cdots, x_{(1)})$ respectively, with $x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}$. Similarly, we define $y^{\uparrow} = (y_{(1)}, y_{(2)}, \cdots, y_{(n)})$, $y^{\downarrow} = (y_{(n)}, y_{(n-1)}, \cdots, y_{(1)})$.

Definition 1. *A subset S of $\mathbb{R}^2$ is non-decreasing (resp. non-increasing) if and only if for all (x_1, y_1), (x_2, y_2) in S, $x_1 < x_2$ implies $y_1 \leqslant y_2$ (resp. $x_1 < x_2$ implies $y_1 \geqslant y_2$). Random variables X and Y are called increasing (resp. decreasing) monotone dependent if (X, Y) lies almost surely in a non-decreasing (resp. non-increasing) subset of $\mathbb{R}^2$. Samples x and y are called increasing (resp. decreasing) monotone dependent if $\{(x, y)\}$ is a non-decreasing (resp. non-increasing) subset of $\mathbb{R}^2$.*

Definition 1 is sourced from (Mikusinski, Sherwood, and Taylor, 1991). Clearly Definition 1 is symmetrical with respect to X and Y. The monotone dependence outlined here encompasses a broader scope than definitions like the one in (Kimeldorf and Sampson, 1978), where "*each of X and Y is almost surely a monotone function of the other*". This is primarily because it doesn't necessitate a one-to-one mapping. Also, linear dependence, i.e., $P(Y = \alpha X + \beta) = 1$ at the population level or $y = ax + b$ at the sample level, is special case of monotone dependence, and we will refer to dependence which is monotone but not linear as "*nonlinear monotone dependence*".

2.2 Different bounds lead to different capture ranges

With Cauchy-Schwarz Inequality, the well-known covariance inequality can be directly derived as

$$|\text{cov}(X, Y)| \leqslant \sqrt{\text{var}(X)\,\text{var}(Y)},$$

thus the geometric mean of $\text{var}(X)$ and $\text{var}(Y)$, i.e., $\sqrt{\text{var}(X)\,\text{var}(Y)}$, mathematically provides a bound for covariance $\text{cov}(X, Y)$, which ensures that Pearson's Correlation Coefficient

$$r(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\,\text{var}(Y)}}$$

always falls into the range $[-1, +1]$. Scaled by $\sqrt{\text{var}(X)\,\text{var}(Y)}$, Pearson's r turns into a *normalized covariance*, which is dimensionless and bounded. It possesses significant advantage over the original *covariance* in the sense that its value will not be affected by the change in the units of X and Y.

2

A crucial issue that has been neglected so far is that boundedness doesn't ensure optimum. Scaling $\mathrm{cov}\,(X,Y)$ to the range $[-1, +1]$ is not the only thing that matters. In fact, different bounds can be utilized to scale covariance to be bounded coefficients, as reported in previous works (Lin, 1989; Zegers, 1986).

For example, with the Mean Inequality Series (Hardy, Littlewood, and Polya, 1952), it is immediate that

$$|\mathrm{cov}\,(X,Y)| \leqslant \sqrt{\mathrm{var}\,(X)\,\mathrm{var}\,(Y)} \leqslant \frac{1}{2}\left(\mathrm{var}\,(X) + \mathrm{var}\,(Y)\right)$$

in the sense that geometric mean $\sqrt{\mathrm{var}\,(X)\,\mathrm{var}\,(Y)}$ is always less than or equal to arithmetic mean $\frac{1}{2}\left(\mathrm{var}\,(X) + \mathrm{var}\,(Y)\right)$ for nonnegative values $\mathrm{var}\,(X)$ and $\mathrm{var}\,(Y)$. Then we get a looser bound for covariance, i.e., $\frac{1}{2}\left(\mathrm{var}\,(X) + \mathrm{var}\,(Y)\right)$, with which another measure can be defined as follows(Zegers, 1986):

$$r^+\,(X,Y) = \frac{\mathrm{cov}\,(X,Y)}{\frac{1}{2}\left(\mathrm{var}\,(X) + \mathrm{var}\,(Y)\right)}$$

$r^+\,(X,Y)$ is named early as *Additivity Coefficient* (Zegers, 1986) and later as *Standardized Covariance* (Andraszewicz and Rieskamp, 2014). It is proved that the capture range of $r^+\,(X,Y)$ is limited to additive relationships, i.e., $Y = \pm X + \beta$, which are special cases of linear relationships, i.e., $Y = \alpha X + \beta$, with α being fixed to ± 1 (Zegers, 1986).

Further, we can find an even looser bound for covariance, in the sense that

$$\frac{1}{2}\left(\mathrm{var}\,(X) + \mathrm{var}\,(Y)\right) \leqslant \frac{1}{2}\left(\mathrm{var}\,(X) + \mathrm{var}\,(Y) + \left|\bar{X} - \bar{Y}\right|^2\right),$$

and define a new measure as follows:

$$r^=\,(X,Y) = \frac{\mathrm{cov}\,(X,Y)}{\frac{1}{2}\left(\mathrm{var}\,(X) + \mathrm{var}\,(Y) + \left|\bar{X} - \bar{Y}\right|^2\right)}$$

$r^=$ is named as *Concordance Correlation Coefficient* (Lin, 1989), and it is designed to measure identical relationship, i.e., $Y = \pm X$. When X and Y are both positive, it can be utilized to evaluate their agreement by measuring the variation from the $45°$ line through the origin (Lin, 1989).

As for the abovementioned three measures, r, r^+, and $r^=$, they share the same numerator, $\mathrm{cov}\,(X,Y)$, the differences lie in their denominators. These denominators serve as bounds for $\mathrm{cov}\,(X,Y)$. Different bounds lead to different capture ranges. With the bounds being looser, their capture ranges are shrinking from linear ($Y = \alpha X + \beta$) towards additive ($Y = \pm X + \beta$) and ultimately to identical ($Y = \pm X$) relationships. The looser the bound, the narrower the capture range.

Up until now, all the efforts have only led to looser bounds and measures with narrower capture ranges. *Could we possibly explore breakthroughs by approaching the problem from the opposite direction, aiming to achieve a tighter bound and consequently, devise a new measure with a broader capture range?*

The bound in Pearson's r is intrinsically provided by Cauchy-Schwarz Inequality, which is one of the most widespread and useful inequalities in mathematics. Cauchy-Schwartz Inequality is so classic and reliable that one seldom tries to improve it. Both bounds in r^+ and $r^=$ are looser than that provided by Cauchy-Schwartz Inequality. To loosen Cauchy-Schwartz Inequality might be easy while to tighten such a classic inequality might be relatively difficult. However, we find that it is not impossible to improve the tightness of Cauchy-Schwartz Inequality. In other words, there exists sharper bound for covariance, which will be depicted in the next section.

2.3 New inequality tighter than Cauchy-Schwarz Inequality

Before deriving the new inequality, we will briefly review the classic Cauchy-Schwarz inequality, which is common in textbooks. The Cauchy–Schwarz inequality states that for x and y, we have

$$|\langle x, y\rangle| \leqslant \|x\|\,\|y\|,$$

where $\langle \cdot, \cdot \rangle$ is the inner product. and $\|\cdot\|$ is the norm. The equality holds if and only if x and y are linearly dependent, i.e., $y = ax$ for some constant a.

3

After defining an inner product on the set of random variables using the expectation of their product, i.e., $\langle X, Y \rangle = EXY$, the Cauchy–Schwarz inequality becomes

$$|EXY| \leqslant \sqrt{EX^2 EY^2}.$$

Now, we will sharpen the Cauchy–Schwarz inequality. On the basis of the rearrangement theorems (Hardy, Littlewood, and Polya, 1952), we derive 6 theorems(corollaries/propositions) as follows.

Theorem 1. *For random variables X and Y, we have*

$$|EXY| \leqslant \left| EX^{\uparrow} Y^{\updownarrow} \right| \leqslant \sqrt{EX^2 EY^2}.$$

The equality on the left holds if and only if X and Y are monotone dependent, and the equality on the right holds if and only if $Y \stackrel{d}{=} \alpha X$, with $\operatorname{sgn}(EXY) = \operatorname{sgn}(\alpha)$.

Here, $\stackrel{d}{=}$ denotes equality in distribution, and $EX^{\uparrow} Y^{\updownarrow}$ is defined as:

$$EX^{\uparrow} Y^{\updownarrow} = \left\{ \begin{array}{l} EX^{\uparrow} Y^{\uparrow}, if \ \ EXY \geqslant 0 \\ EX^{\uparrow} Y^{\downarrow}, if \ \ EXY < 0 \end{array} \right.$$

For the sake of conciseness, the proofs of Theorem 1 and theorems undermentioned are all included in Appendix A.1.

Theorem 2. *For samples x and y we have*

$$|\langle x, y \rangle| \leqslant \left| \left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle \right| \leqslant \|x\| \|y\|.$$

The equality on the left holds if and only if x and y are monotone dependent, and the equality on the right holds if and only if y is arbitrary permutation of ax, with $\operatorname{sgn}(\langle x, y \rangle) = \operatorname{sgn}(a)$.

Here, $\left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle$ is defined as:

$$\left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle = \left\{ \begin{array}{l} \left\langle x^{\uparrow}, y^{\uparrow} \right\rangle, if \ \ \langle x, y \rangle \geqslant 0 \\ \left\langle x^{\uparrow}, y^{\downarrow} \right\rangle, if \ \ \langle x, y \rangle < 0 \end{array} \right.$$

Corollary 1. *For random variables X and Y, we have covariance inequality series as:*

$$\begin{aligned} |\operatorname{cov}(X, Y)| \leqslant \left| \operatorname{cov}\left(X^{\uparrow}, Y^{\updownarrow}\right) \right| &\leqslant \sqrt{\operatorname{var}(X) \operatorname{var}(Y)} \\ &\leqslant \tfrac{1}{2}\left(\operatorname{var}(X) + \operatorname{var}(Y)\right) \\ &\leqslant \tfrac{1}{2}\left(\operatorname{var}(X) + \operatorname{var}(Y) + \left|\bar{X} - \bar{Y}\right|^2\right) \end{aligned}$$

The first equality holds if and only if X and Y are monotone dependent, and the second equality holds if and only if $Y \stackrel{d}{=} \alpha X + \beta$, with $\operatorname{sgn}(\operatorname{cov}(X, Y)) = \operatorname{sgn}(\alpha)$.

Here, $\operatorname{cov}\left(X^{\uparrow}, Y^{\updownarrow}\right)$ is defined as:

$$cov\left(X^{\uparrow}, Y^{\updownarrow}\right) = \left\{ \begin{array}{ll} cov\left(X^{\uparrow}, Y^{\uparrow}\right), & if \ \ cov(X, Y) \geqslant 0 \\ cov\left(X^{\uparrow}, Y^{\downarrow}\right) & if \ \ cov(X, Y) < 0 \end{array} \right.$$

Corollary 2. *For samples x and y, we have covariance inequality series as*

$$\begin{aligned} |s_{x,y}| \leqslant \left| s_{x^{\uparrow}, y^{\updownarrow}} \right| &\leqslant \sqrt{s_x^2 s_y^2} \\ &\leqslant \frac{1}{2}\left(s_x^2 + s_y^2\right) \\ &\leqslant \frac{1}{2}\left(s_x^2 + s_y^2 + |\bar{x} - \bar{y}|^2\right) \end{aligned}$$

The first equality holds if and only if x and y are monotone dependent, and the second equality holds if and only if y is arbitrary permutation of $ax + b$, with $\operatorname{sgn}(s_{x,y}) = \operatorname{sgn}(a)$.

Here, $s_{x^{\uparrow}, y^{\updownarrow}}$ is defined as:

$$s_{x^{\uparrow}, y^{\updownarrow}} = \left\{ \begin{array}{l} s_{x^{\uparrow}, y^{\uparrow}}, if \ \ s_{x,y} \geqslant 0 \\ s_{x^{\uparrow}, y^{\downarrow}}, if \ \ s_{x,y} < 0 \end{array} \right.$$

4

2.4 The proposed Rearrangement Correlation

The inequality series in Corollary 1 and Corollary 2 provides sharper bounds for covariance at the population level and the sample level respectively. We will leverage them to define the so-called *Rearrangement Correlation*, which is the adjusted version of Pearson's r proposed here.

Definition 2. *The Rearrangement Correlation of random variables X and Y is defined as:*

$$r^{\#}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\left|\operatorname{cov}\left(X^{\uparrow}, Y^{\updownarrow}\right)\right|}$$

Definition 3. *The Rearrangement Correlation of samples x and y is defined as:*

$$r^{\#}(x,y) = \frac{s_{x,y}}{\left|s_{x^{\uparrow}, y^{\updownarrow}}\right|}$$

The new measure is named "*Rearrangement Correlation*" because its theoretical foundation is the rearrangement inequality, as shown in Theorem 1 and Theorem 2. We adopt the musical sharp symbol # to denote rearrangement correlation, signifying that this measure has sharp values because of its sharp bounds. Analogous to how C# is pronounced as *C-sharp*, $r^{\#}$ is pronounced as *r-sharp*.

As for the relationship between $r^{\#}(x,y)$ and $r^{\#}(X,Y)$, it is clear that $r^{\#}(x,y)$ converges to $r^{\#}(X,Y)$ when $n \to \infty$ according to their definitions.

The capture range of rearrangement correlation is no longer limited to linear dependence but monotone dependence, which is revealed by the next proposition.

Proposition 1. *For random variables X, Y, and samples x, y, the following hold:*

- $\left|r^{\#}(X,Y)\right| \leqslant 1$ *and the equality holds if and only if X and Y are monotone dependent.*

- $\left|r^{\#}(x,y)\right| \leqslant 1$ *and the equality holds if and only if x and y are monotone dependent.*

An interesting question might arise in one's mind: how can the simple adjustment, replacing $\sqrt{\operatorname{var}(X)\operatorname{var}(Y)}$ with $\left|\operatorname{cov}\left(X^{\uparrow}, Y^{\updownarrow}\right)\right|$, leads to the capture range expanding from linear dependence to (nonlinear) monotone dependence? The capture range is inherited from *covariance* itself. The capture range of covariance is limited neither to identical dependence as that of *Concordance Correlation Coefficient*, additive dependence as that of *Additivity Coefficient*, nor to linear dependence as that of Pearson's r. In fact, it can potentially detect and measure arbitrary monotone dependence, if scaled properly. In other words, Pearson's r is also measuring nonlinear monotone dependence to some extent. The adjustment is nothing more than compensating for underestimation.

The relationships among the above-mentioned *Concordance Correlation Coefficient* ($r^{=}$), *Additivity Coefficient* (r^{+}), *Pearson's r*, and the new proposed *Rearrangement Correlation* ($r^{\#}$) are depicted in Figure 1.
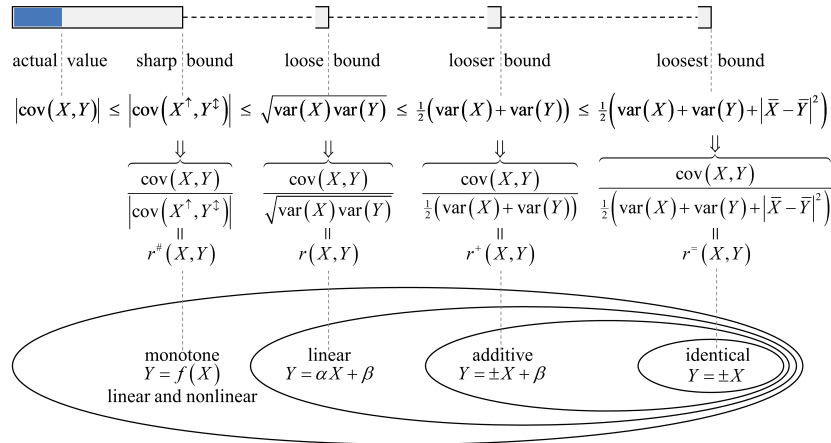
Figure 1: Covariance inequality series, correlation coefficients and their capture ranges

5

The following proposition reveals the relationship between Pearson's r and its adjusted version, i.e., Rearrangement Correlation:

Proposition 2. *For random variables X, Y, and samples x, y, the following hold:*

- $\left|r^{\#}(X,Y)\right| \geqslant |r(X,Y)|$ *and the equality holds if and only if* $Y \overset{d}{=} \alpha X + \beta$, *with* $\mathrm{sgn}\left(r(X,Y)\right) = \mathrm{sgn}(\alpha)$.

- $\left|r^{\#}(x,y)\right| \geqslant |r(x,y)|$ *and the equality holds if and only if y is arbitrary permutation of* $ax + b$, *with* $\mathrm{sgn}\left(r(x,y)\right) = \mathrm{sgn}(a)$.

Proposition 2 shows that $r^{\#}(X,Y)$ will revert to $r(X,Y)$ if and only if $Y \overset{d}{=} \alpha X + \beta$, $\mathrm{sgn}\left(r(X,Y)\right) = \mathrm{sgn}(\alpha)$, and $r^{\#}(x,y)$ to $r(x,y)$ if and only if y is arbitrary permutation of $ax + b$, with $\mathrm{sgn}\left(r(x,y)\right) = \mathrm{sgn}(a)$. It is clear that linear dependence is special case of these conditions. Thus, $r^{\#}$ reverts to r in linear scenarios.

Another question to be asked is, do we need a new monotone measure given that rank-based measures such as Spearman's ρ can already measure monotone dependence? The answer is twofold:

On the one hand, $r^{\#}$ has a higher *resolution* and is more accurate. Without exception, all measures designed for monotone dependence are utilizing the order information. However, what we utilize here is the original information, rather than the ranks. Mapping numerical values to their ranks does of course produce a certain loss of information. A small difference between two values may no longer be distinguished from a large difference. With sample size n, there are totally $\frac{n^3-n}{6}$ possible ρ values between -1 and $+1$, whatever raw values are and however correlated patterns differ. The *resolution* of Spearman's ρ might be inadequate. To take a simple example, let $x = (4,3,2,1)$ and

- $y_1 = (5,4,3,2.00)$
- $y_2 = (5,4,3,3.25)$
- $y_3 = (5,4,3,3.50)$
- $y_4 = (5,4,3,3.75)$
- $y_5 = (5,4,3,4.50)$

Obviously, y_1 and x behaves exactly in the same way, with their values getting small and small step by step. The behavior of y_2, y_3, y_4, and y_5 are becoming more and more different from that of x. However, the ρ values are all the same for y_2, y_3 and y_4. In contrast, the $r^{\#}$ values can reveal all these differences exactly.

- $r^{\#}(x,y_1) = 1.00$, $\rho(x,y_1) = 1.00$
- $r^{\#}(x,y_2) = 0.93$, $\rho(x,y_2) = 0.80$
- $r^{\#}(x,y_3) = 0.85$, $\rho(x,y_3) = 0.80$
- $r^{\#}(x,y_4) = 0.76$, $\rho(x,y_4) = 0.80$
- $r^{\#}(x,y_5) = 0.38$, $\rho(x,y_5) = 0.40$

On the other hand, $r^{\#}$ is comparable with Pearson's r, while the latter is not. For nonlinear monotone dependence, the value of Spearman's ρ might be remarkably greater than the value of Pearson's r. One may attempt to search for nonlinear relationships in data by checking whether the value of ρ far exceeds that of r. However, it might be meaningless and even impossible to compare their values directly. In cases, ρ can be either greater or less than r, and their sign can even be different. Thus the difference $|\rho| - |r|$ is confusing. On the contrary, the signs of $r^{\#}$ and r are always the same, and $\left|r^{\#}\right|$ is always greater than or equal to $|r|$. $\left|r^{\#}\right| - |r|$ equals to 0 if and only if y is arbitrary permutation of $ax + b$. Its value increases with the degree of nonlinearity.

However, Spearman's ρ can also be superior to $r^{\#}$ in the sense that the former is robust to outliers while the latter is not. Rearrangement correlation is a scaled covariance, and the limitation of being non-robust to outliers is inherited from covariance itself. In fact, concordance correlation coefficient, additivity coefficient, and Pearson's r are also scaled covariance measures, and none of them are robust to outliers.

6

To be more robust, we can also transform the raw data into their ranks before calculating $r^{\#}$. Interestingly, $r^{\#}$ becomes equivalent to Spearman's ρ when calculated on ranks. Let P and Q be the ranks of x and y respectively. Then, in the sense that $sd\left(P,Q\right) = sd\left(P^{\uparrow}, Q^{\updownarrow}\right) = \frac{n(n+1)}{12}$, $r^{\#}\left(P,Q\right) = \rho\left(P,Q\right)$. This explains why ρ can measure nonlinear monotone relationships while r only measures linear ones, despite them sharing a similar formula. The key is not just the ranking but achieving a sharp bound. Since ρ and $r^{\#}$ are equivalent when applied to ranks, and $r^{\#}$ can measure arbitrary monotone dependence (as proven in our manuscript), ρ can do the same.

3 Experiments

3.1 Performance metrics

The main purpose of proposing Rearrangement Correlation is to provide a measure of dependence strength for nonlinear monotone relationships, rather than to simply serve as a test statistic for testing independence. Thus, our performance metrics focus on strength measurement.

The basic question to be asked when measuring any attribute is how accurate is this measurement, and there should be no exception for dependence measurement. ISO 5725 uses two terms *"trueness"* and *"precision"* to describe the accuracy of a measurement method. *Trueness* refers to the closeness of agreement between the mean or median of measured results and the true or accepted reference value. *Precision* refers to the closeness of agreement between measured values (ISO, 1994). The comprehensive performance of *trueness* and *precision* can be represented as the mean absolute error (*MAE* for short), and calculated as the mean of the absolute values of the difference between the measured value and the conventional true value. On the whole, a measure with lower MAE value is better.

We evaluate the performance of different measures in a supervised way. We employ the coefficient of determination, R^2, which is defined as the proportion of variance for one variable explained by the other variable, as the ground truth of strength of dependence, which is common in practices (Reshef et al., 2011). Further, we take its square root, R, as the conventional true value of the relationship strength. A simple evidence is that Pearson's r, as the golden standard for measuring linear dependence, is equivalent to R, as long as the relationship is linear. Thus, it is reasonable to adopt R as the reference value.

3.2 Simulation procedure

We investigate the accuracy performance of $r^{\#}$, along with r^{+}, Pearson's r, Spearman's ρ, Kendall's τ and four other leading dependence measures, i.e., HSIC, dCor, MIC and Chatterjee's ξ in the following way: for each scenario $y = f\left(x\right)$, we generated 512 pairs of (x, y) from the regression model $y = f\left(x\right) + \varepsilon$, and computed the values of different measures between x and y at different R levels. In the regression model, the x sample is uniformly distributed on the unit interval $(0, 1)$, and the noise is normally distributed as $\varepsilon \sim N\left(0, \sigma\right)$, with σ controlling R to a certain level.

$$R = \sqrt{1 - \frac{\sigma^2}{\text{var}\left(Y\right)}}$$

For the sake of robustness, the computation process is repeated 10 times for each measure at each R level, and the mean value is adopted.

Simulation procedure is implemented in the *recor* R package, which is available in supplemental materials. The workflow is to call *accuracy_db*() firstly, *accuracy_results_frm_db*() secondly and *accuracy_plot_lite*() finally. To reproduce the results, just keep all the parameters as default. More details are available in the package help files.

3.3 Performance in simulated scenarios

The simulation is conducted in up to 50 types of different monotone scenarios, including all basic elementary functions, lots of composite functions and several special functions. To our knowledge, our research explores the most extensive and representative range of scenarios. Detailed descriptions of these scenarios can be found in Appendix A.3, and the results are shown in Figure 2.

7
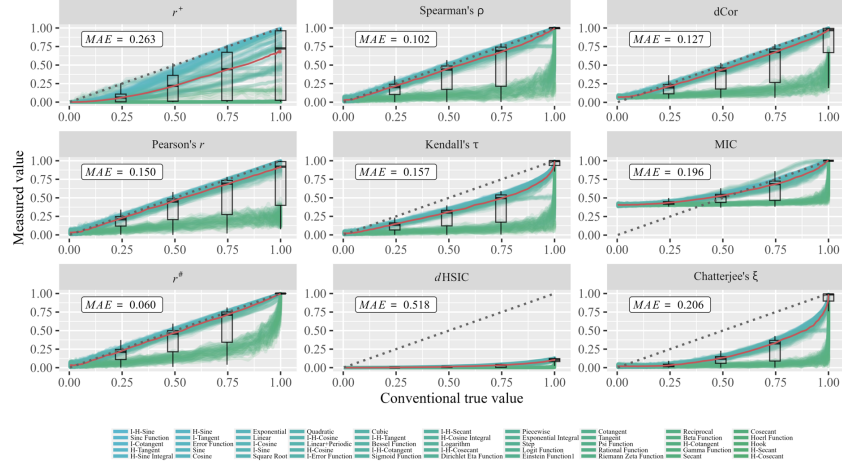
Figure 2: Performance of different measures in 50 simulated scenarios

Figure 2 shows the scatter plots of conventional true value versus measured values in different scenarios. The nine investigated measures are located in nine panels respectively. In each panel, there are totally 50 transparent green lines, representing the measured value (Y-axis) with respect to conventional true value (X-axis) for 50 scenarios. The prefixes *I-* and *H-* refer to *Inverse* and *Hyperbolic* respectively. For example, *I-H-Tangent* stands for *Inverse Hyperbolic Tangent* function. In all these panels, the dashed diagonal lines represent an ideal measure, the score of which is exactly the same as the conventional true value. Apparently, the closeness to this reference line reflects the performance of a measure for a certain scenario. The median values of each measure among scenarios at different R levels are also calculated and denoted by the non-transparent red line.

We first look at the extreme values on both sides. It is expected that a measure will score nearly zero when X and Y are randomly generated, i.e., $R \approx 0$, and score one when there is perfect monotone relationship, i.e., $R = 1$. Only Spearman's ρ, Kendall's τ and the adjusted $r^{\#}$ meet this requirement. MIC also scores $+1$ when $R = 1$, however, it tends to overestimate the strength when R is near zero, as also reported in other literature before (Chatterjee, 2021). The remaining five measures, i.e., r^{+}, r, HSIC, dCor and ξ, underestimate the strength of nonlinear relationships, and never converge to $+1$ even when R approaches to 1.

Now let's take a close look at the intermediate values. It can be seen in Figure 2 that the non-transparent red line of $r^{\#}$ is the closest one to the dashed line, which means the measured values by $r^{\#}$ possess the minimum error. To further quantify the accuracy, we add four boxplots at four representative R levels (approximately, 0.25, 0.50, 0.75, and 1.00) for each measure. $r^{\#}$ has the highest trueness in all these representative levels. As for precision, the $r^{\#}$ also outperforms all other measures except HSIC and MIC. Although HSIC and MIC possesses the best precision, they suffer from lower trueness. HSIC tends to underestimate the strength severely, and MIC is also a biased measure, tending to overestimate the strength when the signal is weak, and underestimate it when the signal is strong, as shown in Figure 2.

The overall performance in terms of MAE is ordered as $r^{\#}\,(0.060) \succ \rho\,(0.102) \succ dCor\,(0.127) \succ r\,(0.150) \succ \tau\,(0.157) \succ MIC\,(0.196) \succ \xi\,(0.206) \succ r^{+}\,(0.263) \succ HSIC\,(0.518)$. $r^{\#}$ possesses significant accuracy advantage over all other measures.

3.4 Performance in real-life scenarios

In addition to simulated scenarios, we also investigate the performance of these measures on real life scenarios provided by NIST (National Institute of Standards and Technology, 2003). There are five available monotone scenarios: Chwirut1, Hahn1, Rat43, Roszman1, and Thurber. Details for these scenarios are available in Appendix A.4.

The performance of nine measures in these five scenarios are depicted in Figure 3. Bar plots illustrate the measured values, the conventional true value verified by NIST is annotated on the top of each

8
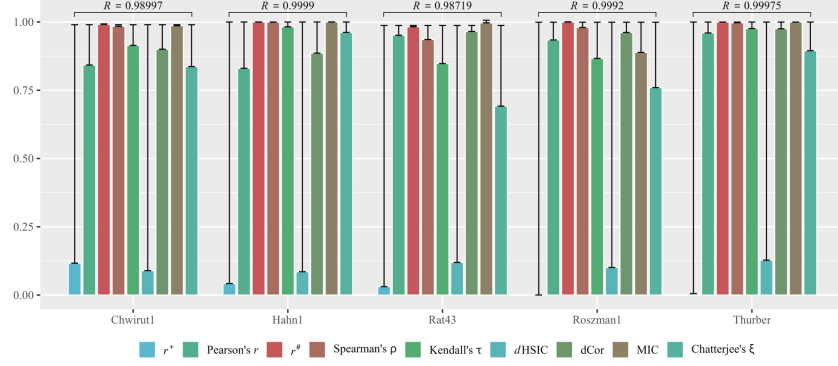
Figure 3: Performance of Different Measures in 5 Real-life Scenarios

scenario group. And the differences between the measured value and the true value are mapped as error bars.

It can be seen from Figure 3 that $r^{\#}$ possesses minimum error, or best accuracy performance among all these five scenarios, with its MAE value as 0.00141, followed by $\rho(0.0159)$, MIC(0.0249), dCor(0.0575), $\tau(0.0779)$, $r(0.0916)$, $\xi(0.166)$, HSIC(0.891) and $r^{+}(0.956)$.

3.5 Performance in non-monotone scenarios

It's worth noting that the aforementioned scenarios only cover monotone cases. To evaluate performance in typical non-monotone contexts, we conducted experiments in 16 scenarios, encompassing those outlined in (Reshef et al., 2011) and (Simon and Tibshirani, 2014). Details for these scenarios are available in Appendix A.5.

As anticipated, the MAE value of $r^{\#}$ reaches a significant 0.418, notably inferior to those of $\xi(0.141)$, MIC (0.157) and dCor (0.364). In essence, $r^{\#}$ struggles to accurately measure non-monotone dependence. This limitation stems from its reliance on *covariance*, which inherently fails to detect non-monotone relationships. To illustrate, consider a standard introductory text book example, i.e., $\operatorname{cov}(X, Y) = 0$ despite Y being totally dependent on X via $Y = X^2$. Attempts to tighten its bound proves futile.

However, the performance of $r^{\#}$ is also superior to those of Spearman's $\rho(0.431)$, Pearson's $r(0.437)$ and Kendall's $\tau(0.461)$. As for accuracy performance, $r^{\#}$ outperforms the three classical correlation coefficients in not only monotone, but also non-monotone scenarios.

4 Conclusion and discussion

We proposed here an adjusted version of Pearson's r, i.e., *rearrangement correlation*, which can be treated as counterpart of Pearson's r for nonlinear monotone dependence.

The basic idea of rearrangement correlation is simple and straightforward. Its mathematical foundation is a sharpened version of the famous Cauchy-Schwarz Inequality. Tighter bound leads to wider capture range. With the adjustment, the capture range of Pearson's r is extended from linear dependence to (nonlinear) monotone dependence. Simulated and real-life investigations demonstrate that the rearrangement correlation is more accurate in measuring nonlinear monotone dependence than the three classical correlation coefficients and other more recently proposed dependence measures.

We may draw the conclusion that: Pearson's r is undoubtedly the gold measure for linear dependence. Now, *it might be the gold measure also for nonlinear monotone dependence, if adjusted.*

9

References

Andraszewicz, Sandra and Jörg Rieskamp (2014). "Standardized Covariance—A Measure of Association, Similarity and Co-riskiness between Choice Options". In: *Journal of Mathematical Psychology* 61, pp. 25–37.

Armstrong, Richard A (2019). "Should Pearson's Correlation Coefficient Be Avoided?" In: *Ophthalmic and Physiological Optics* 39.5, pp. 316–327.

Chatterjee, Sourav (2021). "A New Coefficient of Correlation". In: *Journal of the American Statistical Association* 116.536, pp. 2009–2022.

Gretton, Arthur et al. (2005). "Measuring Statistical Dependence with Hilbert-Schmidt Norms". In: *Algorithmic Learning Theory*. Ed. by David Hutchison et al. Vol. 3734. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 63–77.

Hardy, G.H., J.E. Littlewood, and G. Polya (1952). *Inequalities*. London: Cambridge University Press.

ISO (1994). *Accuracy (Trueness and Precision) of Measurement Methods and Results — Part 1: General Principles and Definitions*. International Standard.

Kendall, M. G. (1938). "A New Measure of Rank Correlation". In: *Biometrika* 30.1/2, pp. 81–93.

Kimeldorf, George and Allan R. Sampson (July 1978). "Monotone Dependence". In: *The Annals of Statistics* 6.4. (Visited on 09/21/2022).

Lee Rodgers, Joseph and W. Alan Nice Wander (1988). "Thirteen Ways to Look at the Correlation Coefficient". In: *American Statistician* 42.1, pp. 59–66.

Lin, Lawrence I-Kuei (1989). "A Concordance Correlation Coefficient to Evaluate Reproducibility". In: *Biometrics* 45.1, pp. 255–268.

Mikusinski, Sherwood, and Taylor (1991). "The Fréchet Bounds Revisted". In: *Real Analysis Exchange* 17.2, p. 759.

National Institute of Standards and Technology (2003). *NIST Standard Reference Database 140 - Statistical Reference Datasets*. Tech. rep.

Pearson, Karl (1896). "Mathematical Contributions to the Theory of Evolution, III. Regression, Heredity, and Panmixia". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 187, pp. 253–318.

Puccetti, Giovanni (2022). "Measuring Linear Correlation between Random Vectors". In: *Information Sciences* 607, pp. 1328–1347.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. Manual. Vienna, Austria.

Reshef, David N. et al. (2011). "Detecting Novel Associations in Large Data Sets". In: *Science* 334.6062, pp. 1518–1524.

Simon, Noah and Robert Tibshirani (2014). *Comment on "Detecting Novel Associations In Large Data Sets" by Reshef Et Al, Science Dec 16, 2011*. arXiv: 1401.7645.

Spearman, C. (1904). "The Proof and Measurement of Association between Two Things". In: *The American Journal of Psychology* 15.1, pp. 72–101.

Speed, T. (2011). "A Correlation for the 21st Century". In: *Science* 334.6062, pp. 1502–1503.

Székely, Gábor J., Maria L. Rizzo, and Nail K. Bakirov (2007). "Measuring and Testing Dependence by Correlation of Distances". In: *Annals of Statistics* 35.6, pp. 2769–2794.

Tjøstheim, Dag, Håkon Otneim, and Bård Støve (2022). "Statistical Dependence: Beyond Pearson's ρ". In: *Statistical Science* 37.1, pp. 90–109.

van den Heuvel, Edwin and Zhuozhao Zhan (2022). "Myths About Linear and Monotonic Associations: Pearson's r, Spearman's ρ, and Kendall's τ". In: *The American Statistician* 76.1, pp. 44–52.

Wasserman, Larry (2004). *All of Statistics: A Concise Course in Statistical Inference*. New York: Springer.

Whitt, Ward (1976). "Bivariate Distributions with Given Marginals". In: *The Annals of Statistics* 4.6, pp. 1280–1289.

Zegers, Frits E. (1986). "A Family of Chance-corrected Association Coefficients for Metric Scales". In: *Psychometrika* 51.4, pp. 559–562.

10

A Appendix

A.1 Proofs of theorems, corollaries and propositions

Theorem 1. *For random variables X and Y, we have*

$$|EXY| \leqslant \left|EX^{\uparrow}Y^{\updownarrow}\right| \leqslant \sqrt{EX^2 EY^2}.$$

The equality on the left holds if and only if X and Y are monotone dependent, and the equality on the right holds if and only if $Y \overset{d}{=} \alpha X$, with $\mathrm{sgn}\,(EXY) = \mathrm{sgn}\,(\alpha)$.

Here, $\overset{d}{=}$ denotes equality in distribution, and $EX^{\uparrow}Y^{\updownarrow}$ is defined as:

$$EX^{\uparrow}Y^{\updownarrow} = \begin{cases} EX^{\uparrow}Y^{\uparrow}, if\ EXY \geqslant 0 \\ EX^{\uparrow}Y^{\downarrow}, if\ EXY < 0 \end{cases}$$

Proof. The proof will be completed in two parts.

- The proof of $|EXY| \leqslant \left|EX^{\uparrow}Y^{\updownarrow}\right|$ is mainly based on the rearrangement theorem for functions, i.e., Theorem 378 on page 278 of (Hardy, Littlewood, and Polya, 1952): Let $f^{\uparrow}$, $g^{\uparrow}$ denote increasing rearrangements and $f^{\downarrow}$, $g^{\downarrow}$ decreasing rearrangements of f and g on $[0,1]$ as defined on page 276 of (Hardy, Littlewood, and Polya, 1952). Then we have

$$\int_0^1 f^{\uparrow}(u)\, g^{\downarrow}(u) du \leqslant \int_0^1 f(u)\, g(u) du \leqslant \int_0^1 f^{\uparrow}(u)\, g^{\uparrow}(u) du.$$

  Let $\prod(F, G)$ be the set of all joint cdf's on $\mathbb{R}^2$ having F and G as marginal cdf's. For arbitrary cdf $H \in \prod$ there exists $(X, Y): [0,1] \to \mathbb{R}^2$ such that $[X(U), Y(U)]$ has cdf H. We can let $f(u) = X(u)$ and $g(u) = Y(u)$ so that $EXY = \int_0^1 f(u)\, g(u) du$. The increasing and decreasing rearrangements of f and g are just $f^{\uparrow}(u) = F^{-1}(u)$, $f^{\downarrow}(u) = F^{-1}(1-u)$, $g^{\uparrow}(u) = G^{-1}(u)$, and $g^{\downarrow}(u) = G^{-1}(1-u)$, as stated in (Whitt, 1976). Thus, we have
  $$EX^{\uparrow}Y^{\downarrow} \leqslant EXY \leqslant EX^{\uparrow}Y^{\uparrow}.$$

  The right-hand (*resp.* left-hand) equality holds if and only if $(X, Y) \overset{d}{=} \left(F^{-1}(U), G^{-1}(U)\right)$ (*resp.* $(X, Y) \overset{d}{=} \left(F^{-1}(U), G^{-1}(1-U)\right)$). The equality conditions can be equivalently expressed as X and Y are increasing (*resp.* decreasing) monotone dependent (Mikusinski, Sherwood, and Taylor, 1991).

  If $EXY \geqslant 0$, we have $EX^{\uparrow}Y^{\updownarrow} = EX^{\uparrow}Y^{\uparrow} \geqslant EXY \geqslant 0$, which implies that $\left|EX^{\uparrow}Y^{\updownarrow}\right| = EX^{\uparrow}Y^{\uparrow} \geqslant EXY = |EXY|$, and the equality holds if and only if X and Y are increasing monotone dependent. If $EXY < 0$, we have $EX^{\uparrow}Y^{\updownarrow} = EX^{\uparrow}Y^{\downarrow} \leqslant EXY < 0$, which implies that $\left|EX^{\uparrow}Y^{\updownarrow}\right| = -EX^{\uparrow}Y^{\downarrow} \geqslant -EXY = |EXY|$, and the equality holds if and only if X and Y are decreasing monotone dependent. Either way, we have $\left|EX^{\uparrow}Y^{\updownarrow}\right| \geqslant |EXY|$, and the equality holds if and only if X and Y are monotone dependent.

- The proof of $\left|EX^{\uparrow}Y^{\updownarrow}\right| \leq \sqrt{EX^2 EY^2}$ is mainly based on Cauchy-Schwarz Inequality: If $EXY \geqslant 0$, $\left|EX^{\uparrow}Y^{\updownarrow}\right| = \left|EX^{\uparrow}Y^{\uparrow}\right| \leqslant \sqrt{E(X^{\uparrow})^2 E(Y^{\uparrow})^2} = \sqrt{EX^2 EY^2}$ in the sense that $X^{\uparrow} \overset{d}{=} X$, and $Y^{\uparrow} \overset{d}{=} Y$. And the equality holds if and only if $Y^{\uparrow} = \alpha X^{\uparrow}$, equivalently, $Y \overset{d}{=} \alpha X$, for some constant $\alpha \geq 0$. Similarly, If $E(XY) < 0$, $\left|EX^{\uparrow}Y^{\updownarrow}\right| = \left|EX^{\uparrow}Y^{\downarrow}\right| \leqslant \sqrt{E(X^{\uparrow})^2 E(Y^{\downarrow})^2} = \sqrt{EX^2 EY^2}$ in the sense that $X^{\uparrow} \overset{d}{=} X$, and $Y^{\downarrow} \overset{d}{=} Y$. And the equality holds if and only if $Y^{\downarrow} = \alpha X^{\uparrow}$, equivalently, $Y \overset{d}{=} \alpha X$, for some constant $\alpha < 0$. Either way, we have $\left|EX^{\uparrow}Y^{\updownarrow}\right| \leqslant \sqrt{EX^2 EY^2}$ and the equality holds if and only if $Y \overset{d}{=} \alpha X$, with $\mathrm{sgn}\,(EXY) = \mathrm{sgn}\,(\alpha)$.

11

□

Theorem 2. *For samples x and y we have*

$$|\langle x, y \rangle| \leqslant \left| \left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle \right| \leqslant \|x\| \, \|y\| \, .$$

The equality on the left holds if and only if x and y are monotone dependent, and the equality on the right holds if and only if y is arbitrary permutation of ax, with $\operatorname{sgn}(\langle x, y \rangle) = \operatorname{sgn}(a)$.

Here, $\left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle$ is defined as:

$$\left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle = \left\{ \begin{array}{ll} \left\langle x^{\uparrow}, y^{\uparrow} \right\rangle, if & \langle x, y \rangle \geqslant 0 \\ \left\langle x^{\uparrow}, y^{\downarrow} \right\rangle, if & \langle x, y \rangle < 0 \end{array} \right.$$

Proof. The proof will also be completed in two parts.

- The proof of $|\langle x, y \rangle| \leqslant \left| \left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle \right|$ is mainly based on another rearrangement theorem for finite sets, i.e., Theorem 368 on page 261 of (Hardy, Littlewood, and Polya, 1952): With $x^{\uparrow} = \left(x_{(1)}, x_{(2)}, \cdots, x_{(n)} \right)$, $y^{\uparrow} = \left(y_{(1)}, y_{(2)}, \cdots, y_{(n)} \right)$, and $y^{\downarrow} = \left(y_{(n)}, y_{(n-1)}, \cdots, y_{(1)} \right)$, we have

$$\sum_{i=1}^{n} x_{(i)} y_{(n-i+1)} \leqslant \sum_{i=1}^{n} x_i y_i \leqslant \sum_{i=1}^{n} x_{(i)} y_{(i)}.$$

  That is,

$$\left\langle x^{\uparrow}, y^{\downarrow} \right\rangle \leqslant \langle x, y \rangle \leqslant \left\langle x^{\uparrow}, y^{\uparrow} \right\rangle,$$

  and the right-hand (*resp.* left-hand) equality holds if and only if x and y are similarly(*resp.* oppositely) ordered. According to the definitions of "*similarly(resp. oppositely) ordered*" on page 43 in (Hardy, Littlewood, and Polya, 1952), the equality conditions can be equivalently expressed as *x and y are increasing(resp. decreasing) monotone dependent*.

  If $\langle x, y \rangle \geqslant 0$, we have $\left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle = \left\langle x^{\uparrow}, y^{\uparrow} \right\rangle \geqslant \langle x, y \rangle \geqslant 0$, which implies $\left| \left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle \right| = \left\langle x^{\uparrow}, y^{\uparrow} \right\rangle \geqslant \langle x, y \rangle = |\langle x, y \rangle|$, and the equality holds if and only if x and y are increasing monotone dependent. If $\langle x, y \rangle < 0$, we have $\left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle = \left\langle x^{\uparrow}, y^{\downarrow} \right\rangle \leqslant \langle x, y \rangle < 0$, which implies $\left| \left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle \right| = - \left\langle x^{\uparrow}, y^{\downarrow} \right\rangle \geqslant - \langle x, y \rangle = |\langle x, y \rangle|$, and the equality holds if and only if x and y are decreasing monotone dependent. Either way, we have $\left| \left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle \right| \geqslant |\langle x, y \rangle|$ and the equality holds if and only if x and y are monotone dependent.

- The proof of $\left| \left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle \right| \leqslant \|x\| \, \|y\|$ is mainly based on Cauchy-Schwarz Inequality: in the sense that norm $\|\cdot\|$ is permutation invariant, we have $\left\| x^{\uparrow} \right\| = \|x\|$ and $\left\| y^{\uparrow} \right\| = \left\| y^{\downarrow} \right\| = \|y\|$. If $\langle x, y \rangle \geqslant 0$, we have $\left| \left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle \right| = \left| \left\langle x^{\uparrow}, y^{\uparrow} \right\rangle \right| \leqslant \left\| x^{\uparrow} \right\| \left\| y^{\uparrow} \right\| = \|x\| \, \|y\|$, and the equality holds if and only if $y^{\uparrow} = ax^{\uparrow}$, or equivalently, y is arbitrary permutation of ax for some constant $a \geq 0$. If $\langle x, y \rangle < 0$, we have $\left| \left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle \right| = \left| \left\langle x^{\uparrow}, y^{\downarrow} \right\rangle \right| \leqslant \left\| x^{\uparrow} \right\| \left\| y^{\downarrow} \right\| = \|x\| \, \|y\|$, and the equality holds if and only if $y^{\downarrow} = ax^{\uparrow}$, or equivalently, y is arbitrary permutation of ax for some constant $a < 0$. Either way, we have $\left| \left\langle x^{\uparrow}, y^{\updownarrow} \right\rangle \right| \leqslant \|x\| \, \|y\|$, and the equality holds if and only if y is arbitrary permutation of ax, , with $\operatorname{sgn}(\langle x, y \rangle) = \operatorname{sgn}(a)$.

□

Corollary 1. *For random variables X and Y, we have covariance inequality series as:*

$$\begin{aligned} |\operatorname{cov}(X, Y)| \leqslant \left| \operatorname{cov}\left(X^{\uparrow}, Y^{\updownarrow} \right) \right| &\leqslant \sqrt{\operatorname{var}(X) \operatorname{var}(Y)} \\ &\leqslant \tfrac{1}{2} \left(\operatorname{var}(X) + \operatorname{var}(Y) \right) \\ &\leqslant \tfrac{1}{2} \left(\operatorname{var}(X) + \operatorname{var}(Y) + \left| \bar{X} - \bar{Y} \right|^2 \right) \end{aligned}$$

The first equality holds if and only if X and Y are monotone dependent, and the second equality holds if and only if $Y \overset{d}{=} \alpha X + \beta$, with $\operatorname{sgn}(\operatorname{cov}(X, Y)) = \operatorname{sgn}(\alpha)$.

12

Here, $\mathrm{cov}\left(X^{\uparrow}, Y^{\updownarrow}\right)$ *is defined as:*

$$cov\left(X^{\uparrow}, Y^{\updownarrow}\right) := \left\{ \begin{array}{ll} cov\left(X^{\uparrow}, Y^{\uparrow}\right), & if \ \ cov\left(X, Y\right) \geqslant 0 \\ cov\left(X^{\uparrow}, Y^{\downarrow}\right) & if \ \ cov\left(X, Y\right) < 0 \end{array} \right.$$

Proof. $\left|\mathrm{cov}\left(X, Y\right)\right| \leqslant \left|\mathrm{cov}\left(X^{\uparrow}, Y^{\updownarrow}\right)\right| \leqslant \sqrt{\mathrm{var}\left(X\right)\mathrm{var}\left(Y\right)}$, and the equality conditions are immediate from Theorem 1. The remaining parts of the inequality series are obvious. □

Corollary 2. *For samples x and y, we have covariance inequality series as*

$$|s_{x,y}| \leqslant \left|s_{x^{\uparrow}, y^{\updownarrow}}\right| \leqslant \sqrt{s_x^2 s_y^2}$$
$$\leqslant \frac{1}{2}\left(s_x^2 + s_y^2\right)$$
$$\leqslant \frac{1}{2}\left(s_x^2 + s_y^2 + |\bar{x} - \bar{y}|^2\right)$$

The first equality holds if and only if x and y are monotone dependent, and the second equality holds if and only if y is arbitrary permutation of $ax + b$, with $\mathrm{sgn}\left(s_{x,y}\right) = \mathrm{sgn}\left(a\right)$.

Here, $s_{x^{\uparrow}, y^{\updownarrow}}$ is defined as:

$$s_{x^{\uparrow}, y^{\updownarrow}} = \left\{ \begin{array}{ll} s_{x^{\uparrow}, y^{\uparrow}}, if \ \ s_{x,y} \geqslant 0 \\ s_{x^{\uparrow}, y^{\downarrow}}, if \ \ s_{x,y} < 0 \end{array} \right.$$

Proof. $|s_{x,y}| \leqslant \left|s_{x^{\uparrow}, y^{\updownarrow}}\right| \leqslant \sqrt{s_x^2 s_y^2}$, and the equality conditions are immediate from Theorem 2. The remaining parts of the inequality series are obvious. □

Proposition 1. *For random variables X, Y, and samples x, y, the following hold:*

- $\left|r^{\#}\left(X, Y\right)\right| \leqslant 1$ *and the equality holds if and only if X and Y are monotone dependent.*

- $\left|r^{\#}\left(x, y\right)\right| \leqslant 1$ *and the equality holds if and only if x and y are monotone dependent.*

Proof. The proposition is immediate from Corollary 1 and Corollary 2. □

Proposition 2. *For random variables X, Y, and samples x, y, the following hold:*

- $\left|r^{\#}\left(X, Y\right)\right| \geqslant \left|r\left(X, Y\right)\right|$ *and the equality holds if and only if $Y \overset{d}{=} \alpha X + \beta$, with $\mathrm{sgn}\left(r\left(X, Y\right)\right) = \mathrm{sgn}\left(\alpha\right)$.*

- $\left|r^{\#}\left(x, y\right)\right| \geqslant \left|r\left(x, y\right)\right|$ *and the equality holds if and only if y is arbitrary permutation of $ax + b$, with $\mathrm{sgn}\left(r\left(x, y\right)\right) = \mathrm{sgn}\left(a\right)$.*

Proof. The proposition is immediate from Corollary 1 and Corollary 2. □

A.2 Experiments settings

All the experiments are implemented with the R language (R Core Team, 2024), along with several add-on packages. The following are lists of packages and functions for the implementation of the nine measures involved in our study:

- r^{+}, recor::loose_pearson()
- r, stats::cor()
- $r^{\#}$, recor::sharp_pearson()
- ρ, stats::cor(), with the argument *method* set as "spearman"
- τ, stats::cor(), with the argument *method* set as "kendall"
- HSIC, dHSIC::dhsic()
- dCor, energy::dcor()

13

- MIC, minerva::mine_stat()
- ξ, XICOR::calculateXI()

For convenience, we developed an R package *recor*, which encapsulates all these measures as *cor_XXX()* functions. The *recor* package is available as *recor_1.0.2.tar.gz* in supplemental materials. For a latest version, please visit https://github.com/byaxb/recor.

Hardware environment configuration for this study was: DELL OptiPlex 7070 Tower, equipped with 8-core CPU Core i7-9700 @ 3.00GHz, 24G DDR4 2666MHz RAM. Under this configuration, it took about 5 days to complete all the experiments.

A.3 Simulated scenarios

We carry out our experiments on 50 simulated scenarios, including all basic elementary functions, lots of composite functions and several typical special functions.

1. Linear: $y = 2x + 1, x \in [0, 1]$
2. Quadratic [asymmetry]: $y = x^2, x \in [0, 1]$
3. Square Root: $y = \sqrt{x}, x \in [0, 1]$
4. Cubic: $y = x^3, x \in [0, 1]$
5. Reciprocal: $y = \frac{1}{x}, x \in [0, 1]$
6. Exponential: $y = e^x$, with $x \in [0, 1]$
7. Logarithm: $y = \ln x, x \in [0, 1]$
8. Sine [quarter period]: $y = \sin(x), x \in \left[0, \frac{\pi}{2}\right]$
9. Cosine [quarter period]: $y = \cos(x), x \in \left[0, \frac{\pi}{2}\right]$
10. Tangent [half period]: $y = \tan(x), x \in \left[0, \frac{\pi}{2}\right]$
11. Cotangent [half period]: $y = \cot(x), x \in \left[0, \frac{\pi}{2}\right]$
12. Inverse Sine: $y = \arcsin(x), x \in [0, 1]$
13. Inverse Cosine: $y = \arccos(x), x \in [0, 1]$
14. Inverse Tangent: $y = \arctan(x), x \in [0, 1]$
15. Inverse Cotangent: $y = \text{arccot}(x), x \in [0, 1]$
16. Secant [quarter period]: $y = \sec(x), x \in \left[0, \frac{\pi}{2}\right]$
17. Cosecant [quarter period]: $y = \csc(x), x \in \left[0, \frac{\pi}{2}\right]$
18. Hyperbolic Sine: $y = \sinh x = \frac{e^x - e^{-x}}{2}, x \in [0, 1]$
19. Hyperbolic Cosine: $y = \cosh x = \frac{e^x + e^{-x}}{2}, x \in [0, 1]$
20. Hyperbolic Tangent: $y = \tanh x = \frac{e^{2x} - 1}{e^{2x} + 1}, x \in [0, 1]$
21. Hyperbolic Cotangent: $y = \coth x = \frac{e^{2x} + 1}{e^{2x} - 1}, x \in [0, 1]$
22. Hyperbolic Secant: $y = \text{sech}(x) = \frac{2}{e^x + e^{-x}}, x \in [0, 100]$
23. Hyperbolic Cosecant: $y = csch(x) = \frac{2}{e^x - e^{-x}}, x \in [0, 100]$
24. Inverse Hyperbolic Sine: $y = arcsinh(x) = \ln\left(x + \sqrt{x^2 + 1}\right), x \in [0, 1]$
25. Inverse Hyperbolic Cosine: $y = arccosh(x) = \ln\left(x + \sqrt{x^2 - 1}\right), x \in [1, 2]$
26. Inverse Hyperbolic Tangent: $y = arctanh(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right), x \in [0, 1]$
27. Inverse Hyperbolic Cotangent: $y = arccoth(x) = \frac{1}{2}\ln\left(\frac{x+1}{x-1}\right), x \in [1, 2]$
28. Inverse Hyperbolic Secant: $y = arcsech(x) = \ln\left(\frac{1}{x} + \sqrt{\frac{1}{x^2} - 1}\right), x \in [0, 1]$

14

29. Inverse Hyperbolic Cosecant: $y = arccsch\,(x) = \ln\left(\frac{1}{x} + \sqrt{\frac{1}{x^2} + 1}\right), x \in [0, 1]$

30. Hook: $y = ax + \frac{b}{x}, a = 1, b = 1, x \in [0, 1]$

31. Rational: $y = \frac{x+1}{x-1}, x \in [0, 1]$

32. Hoerl: $y = x^a e^x, a = -1, x \in [0, 1]$

33. Sigmoid: $y = \frac{1}{1+e^{-x}}, x \in [-0.5, 0.5]$

34. Logit: $y = \ln\frac{x}{1-x}, x \in [0, 1]$

35. Step: $y = \begin{cases} 0, if\ 0 \leqslant x < \frac{1}{2} \\ 1, if\ \frac{1}{2} \leqslant x \leqslant 1 \end{cases}$

36. Piecewise [Sigmoid]: $y = \begin{cases} 0, & if\ 0 \leqslant x \leqslant \frac{49}{100} \\ 50\left(x - \frac{1}{2}\right) + \frac{1}{2}, & if\ \frac{49}{100} < x < \frac{51}{100} \\ 1, & if\ \frac{51}{100} \leqslant x \leqslant 1 \end{cases}$

37. Linear + Periodic, High Freq: $y = \frac{1}{10}\sin(10.6(2x - 1)) + \frac{11}{10}(2x - 1), x \in [0, 1]$

38. Sinc Function: $S_{k,h}\,(x) = \frac{\sin(\pi(x-kh)/h)}{\pi(x-kh)/h}, k = 0, h = 1, x \in [0, 1]$

39. Einstein Function: $\mathrm{Einstein}_1\,(x) = \frac{x^2 e^x}{(e^x - 1)^2}, x \in [0, 1]$

40. Exponential Integral: $E_1\,(x) = \int_x^\infty \frac{e^{-t}}{t}dt, x \in [0, 1]$

41. Hyperbolic Sine Integral: $Shi\,(x) = \int_0^x \frac{\sinh t}{t}dt, x \in [0, 1]$

42. Hyperbolic Cosine Integral: $Chi\,(x) = \gamma + \ln x + \int_0^x \frac{\cosh t - 1}{t}dt, x \in [0, 1]$. Here γ is Euler's Constant

43. Error Function: $erf\,(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt, x \in [0, 1]$

44. Inverse Error Function: $inverf\,(x) = t + \frac{1}{3}t^3 + \frac{7}{30}t^5 + \cdots, t = \frac{1}{2}\sqrt{\pi}x, x \in [0, 1]$

45. Gamma Function: $\Gamma\,(x) = \int_0^\infty t^{x-1}e^{-t}dt, x \in [0, 1]$

46. Psi Function: $\psi\,(x) = \frac{d^{k+1}}{dx^{k+1}}\ln\Gamma\,(x) = \frac{\Gamma'(x)}{\Gamma(x)}, x \in [0, 1], k = 0$

47. Riemann Zeta Function: $\zeta\,(x) = \sum_{n=1}^{\infty}\frac{1}{n^x}, x \in [0, 1]$

48. Bessel Function: $Y_v\,(x) = \frac{J_v(x)\cos(v\pi) - J_{-v}(x)}{\sin(v\pi)}, J_v\,(x) = \left(\frac{1}{2}x\right)^v \sum_{k=0}^{\infty}(-1)^k \frac{\left(\frac{1}{4}x^2\right)^k}{k!\Gamma(v+k+1)}, v = 0, x \in [0, 1]$

49. Beta Function: $B\,(x, w) = \frac{\Gamma(x)\Gamma(w)}{\Gamma(x+w)}, w = 1, x \in [0, 1]$

50. Dirichlet Eta Function: $\eta\,(x) = \sum_{n=1}^{\infty}\frac{-1^{n-1}}{n^x}, x \in [0, 1]$

A.4 Real-life scenarios

All the five real life scenarios are provided by NIST (National Institute of Standards and Technology, 2003) as follows:

- Chwirut1: ultrasonic calibration, with Y as ultrasonic response, and X as metal distance.
- Hahn1: thermal expansion of copper, with Y as the coefficient of thermal expansion, and X as temperature in degrees kelvin.
- Rat43: sigmoid growth, with Y as dry weight of onion bulbs and tops, and X as growing time.
- Roszman1: quantum defects in iodine atoms, with Y as the number of quantum defects, and X as the excited energy state.

15

- Thurber: semiconductor electron mobility, with Y as a measure of electron mobility, and X as the natural log of the density.

Data and details about these scenarios are available publicly at:
https://www.itl.nist.gov/div898/strd/nls/nls_main.shtml

A.5 Non-monotone scenarios

We conducted our experiments on 16 non-monotone scenarios, comprehensively covering all the scenarios from (Reshef et al., 2011) and (Simon and Tibshirani, 2014).

1. Quadratic [symmetry]: $y = 4x^2, x \in \left[-\frac{1}{2}, \frac{1}{2}\right]$

2. Cubic 2: $y = 128\left(x - \frac{1}{3}\right)^3 - 48\left(x - \frac{1}{3}\right)^2 - 12\left(x - \frac{1}{3}\right), x \in [0, 1]$

3. Sine, High Freq: $y = \sin(16\pi x), x \in [0, 1]$

4. Cosine [High Freq]: $y = \cos(14\pi x), x \in [0, 1]$

5. Lopsided L-shaped: $y = \begin{cases} 200x, \ if \ 0 \leqslant x < \frac{1}{200} \\ -198x + \frac{199}{100}, if \ \frac{1}{200} \leqslant x < \frac{1}{100} \\ -\frac{x}{99} + \frac{1}{99}, \ if \ \frac{1}{100} \leqslant x \leqslant 1 \end{cases}$

6. Circle: $y = \sqrt{1 - (2x - 1)^2}, x \in [0, 1]$

7. Linear + Periodic, Medium Freq: $y = \sin(10\pi x) + x, x \in [0, 1]$

8. Cubic 3: $y = 4x^3 + x^2 - 4x, x \in [-1.1, 1.3]$

9. Cubic, Y-stretched: $y = 41\left(4x^3 + x^2 - 4x\right), x \in [-1.1, 1.3]$

10. Sine [Two periods]: $y = \sin(4\pi x), x \in [0, 1]$

11. Sine [Low Freq]: $y = \sin(8\pi x), x \in [0, 1]$

12. Sine, Non-Fourier Freq [Low]: $y = \sin(9\pi x), x \in [0, 1]$

13. Cosine, Non-Fourier Freq [Low]: $y = \cos(7\pi x), x \in [0, 1]$

14. Sine, Varying Freq [Medium]: $y = \sin(6\pi x(1 + x)), x \in [0, 1]$

15. Cosine, Varying Freq [Medium]: $y = \cos(5\pi x(1 + x)), x \in [0, 1]$

16. Linear + Periodic, High Freq 2: $y = \frac{1}{5}\sin(10.6(2x - 1)) + \frac{11}{10}(2x - 1), x \in [0, 1]$

16

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The contributions are summarized in *Abstract* and the last paragraph of *Introduction*.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The proposed $r^{\#}$ measures linear and nonlinear monotone relationships accurately. However, it may fail to measure non-monotone dependence. The limitations are discussed in *3.5 Performance in non-monotone scenarios*.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

17

Answer: [Yes]

Justification: Full set of assumptions and complete and correct proofs are provided in *2.3 New inequality tighter than Cauchy-Schwarz Inequality*, *2.4 The proposed Rearrangement Correlation* and *A.1 Proofs of theorems, corollaries and propositions*.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We fully disclosed all the information needed to reproduce the experimental results, as depicted in *3.2 Simulation procedure*.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
    (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
    (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
    (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
    (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

18

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The data and code are available as an attached zip file, *Code and data.zip*. We developed an R package *recor*, which implemented all the experiments. The *recor* package is available as *recor_1.0.2.tar.gz*, included in the zip file. Sufficient instructions to faithfully reproduce the experiments are available in the package help files.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The experimental settings are provided in *3 Experiments*, and *A.2 Experiments Settings*.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We evaluate the performance of proposed method and others according to ISO 5725. Please see *3 Experiments* for details.

   Guidelines:

   - The answer NA means that the paper does not include experiments.

19

- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: Details about the hardware configuration and the time of execution are shown in section "*A.2 Experiments Settings*".

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

Answer: [Yes]

Justification: We have thoroughly reviewed the NeurIPS Code of Ethics. And we confirm that our research fully complies with all of its provisions.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

20

Justification: What we proposed here is an adjusted version of Pearson's r. This study simply provides theoretical results for measuring dependence and does not involve societal impacts.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: Neither the proposed method nor the released data has risk for misuse. The applicability range of this method has undergone comprehensive discussion in *3.5 Performance in non-monotone scenarios*.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All code and data used in our paper are properly credited and explicitly mentioned, as shown in *A.2 Experiments Settings*.

Guidelines:

21

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: We developed an R package to reproduce the experiments. And all the functions in this package are well documented, which can be accessed by the command "?function_name" after installation. For more details, see *A.2 Experiments Settings*.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: Our research does not involve crowdsourcing or research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

22

Answer: [NA]

Justification: Our research does not involve crowdsourcing or research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

23