北京郵電大學
BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

NEURAL INFORMATION PROCESSING SYSTEMS

Adjust Pearson's *r* to Measure Arbitrary Monotone Dependence

Prof. Xinbo AI

Dec. 2024

Adjust Pearson's *r* to Measure Arbitrary Monotone Dependence

NEURAL INFORMATION PROCESSING SYSTEMS

Motivation

Methods

Results

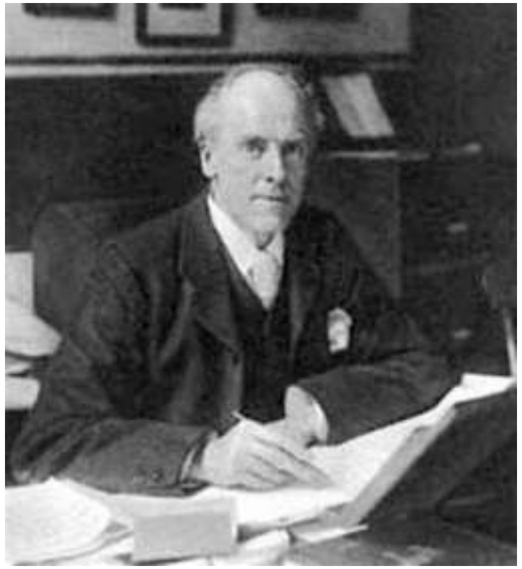
Conclusion

Motivation | Methods | Results | Conclusion

NEURAL INFORMATION PROCESSING SYSTEMS

Pearson's *r*

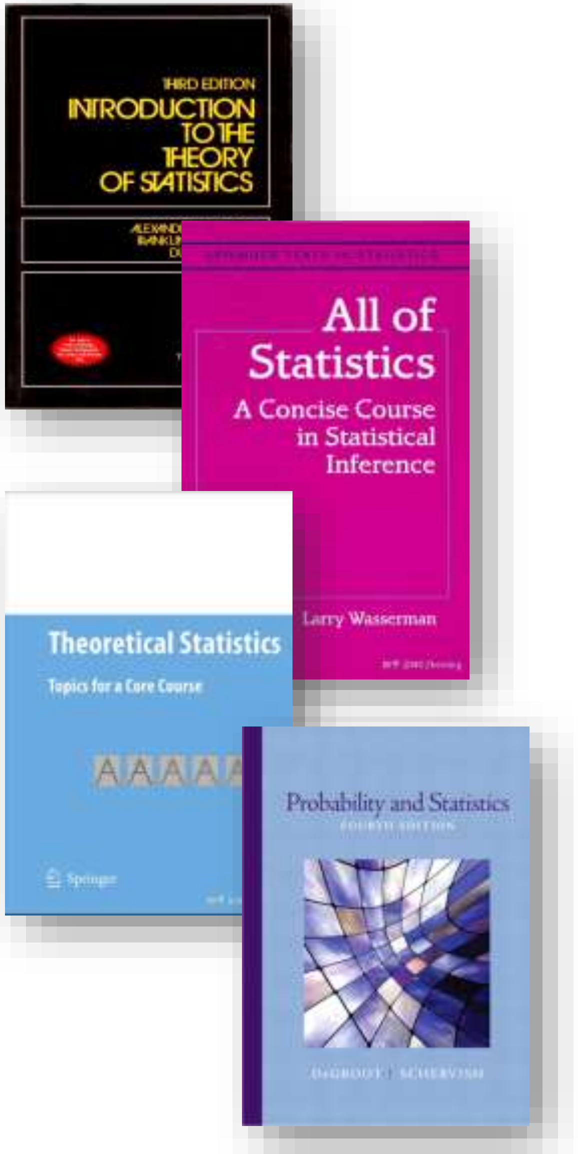
- proposed in the late 19th century (Pearson, 1896)
- has been one of the main tools for scientists and engineers to study bivariate dependence during the 20th century
- still goes strong in the 21st century (Puccetti, 2022)



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).

Widespread belief: *Pearson's r is only for linear dependence*

- **Introduction to the theory of statistics, 1974:** Both the covariance and the correlation coefficient of random variables X and Y are measures of a **linear relationship** of X and Y
- **All of statistics: a concise course statistical inference, 2004:** If X and Y are random variables, then the covariance and correlation between X and Y measure how strong **the linear relationship** is between X and Y.
- **Theoretical statistics: topics for a core course 2010:** The covariance between two variables might be viewed as a measure of the **linear association** between the two variables
- **Probability and statistics, 2012:** The covariance and correlation are attempts to measure that dependence, but they only capture a particular type of dependence, namely **linear dependence**



Motivation | Methods | Results | Conclusion

NEURAL INFORMATION PROCESSING SYSTEMS

The widespread belief is a myth

- van den Heuvel and Zhan (2022): 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.

- This is the issue that our paper aims to address!

NEURAL INFORMATION PROCESSING SYSTEMS

Different bounds lead to different capture ranges

For covariance, we have three bounds:

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

- **1st bound: (Pearson's *r*)** $r(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}}$ → **linear:** $Y = \alpha X + \beta$
- **2nd bound: (Additivity Coefficient)** $r^{+}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\frac{1}{2}(\mathrm{var}(X)+\mathrm{var}(Y))}$ → **Additive:** $Y = \pm X + \beta$
- **3rd bound: (Concordance Coefficient)** $r^{=}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\frac{1}{2}(\mathrm{var}(X)+\mathrm{var}(Y)+|\bar{X}-\bar{Y}|^2)}$ → **Identical:** $Y = \pm X$

By 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**? **YES!**

NEURAL INFORMATION PROCESSING SYSTEMS

New inequality tighter than Cauchy-Schwarz inequality

Cauchy–Schwarz inequality

For random variables X and Y, we have

$$|EXY| \le \sqrt{EX^2 EY^2}$$

The equality holds if and only if $P(Y = \alpha X) = 1$

For samples x and y we have

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

The equality holds if and only if x and y are linearly dependent, i.e., $y = ax$ for some constant a.

Our inequality

For random variables X and Y, we have

$$|EXY| \le |EX^{\uparrow}Y^{\updownarrow}| \le \sqrt{EX^2 EY^2}$$

The equality on the left holds if and only if X and Y are monotone dependent.

For samples x and y we have

$$|\langle x, y\rangle| \le |\langle x^{\uparrow}, y^{\updownarrow}\rangle| \le \|x\|\|y\|$$

The equality on the left holds if and only if x and y are monotone dependent.

New inequality tighter than Cauchy-Schwarz inequality

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 $\operatorname{sgn}(EXY) = \operatorname{sgn}(\alpha)$.

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

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

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

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

NEURAL INFORMATION PROCESSING SYSTEMS

New inequality tighter than Cauchy-Schwarz inequality

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

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

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 $\operatorname{sgn}(\operatorname{cov}(X,Y)) = \operatorname{sgn}(\alpha)$.

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

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

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 $\operatorname{sgn}(s_{x,y}) = \operatorname{sgn}(a)$.

The proposed Rearrangement Correlation

- The Rearrangement Correlation of random variables X and Y is defined as:

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

- The Rearrangement Correlation of samples x and y is defined as:

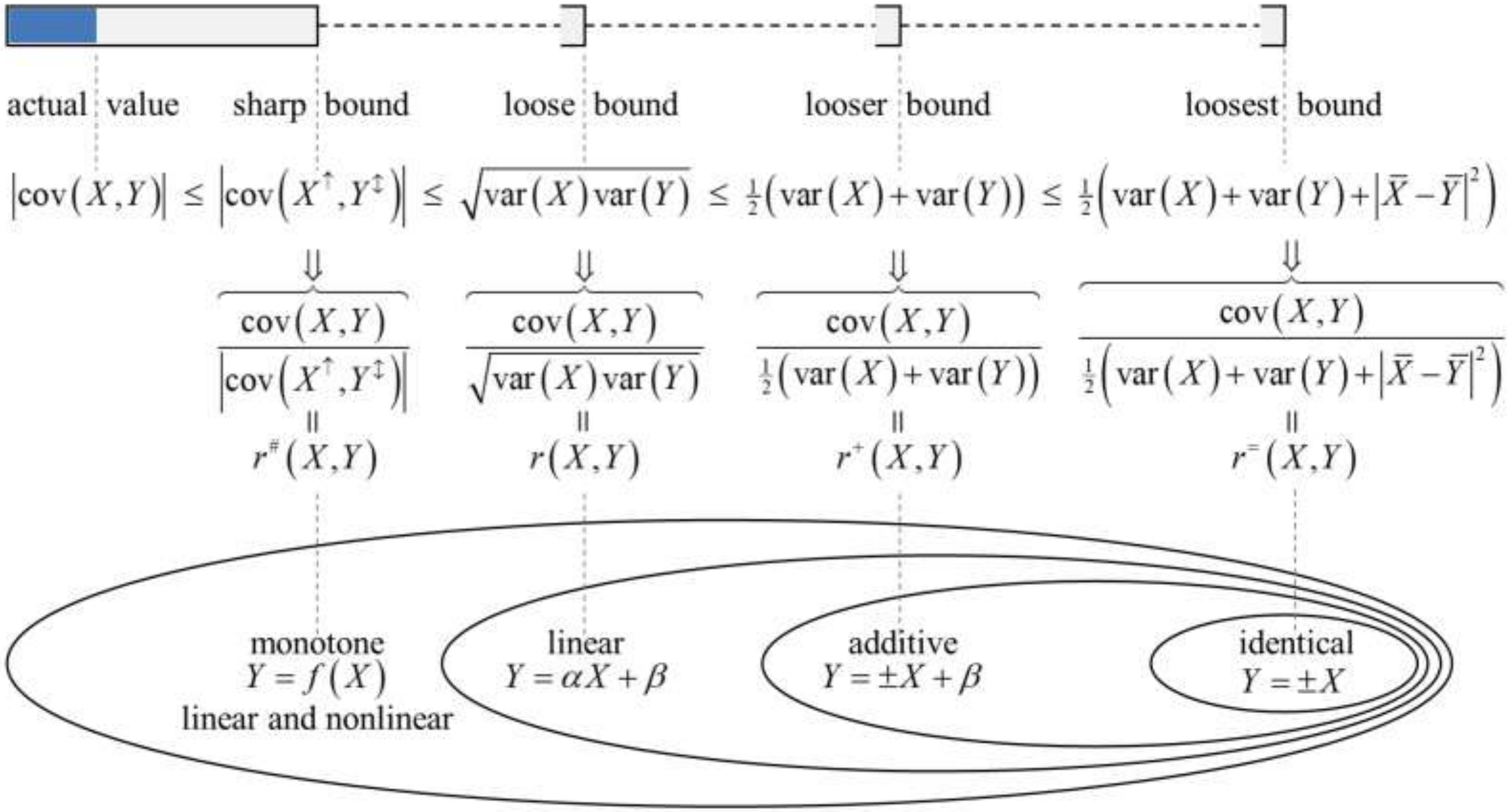
$$r^{\#}(x,y)=\frac{s_{x,y}}{\left|s_{x^{\uparrow},y^{\updownarrow}}\right|}$$

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.*

The proposed Rearrangement Correlation

- Covariance inequality series, correlation coefficients and their capture ranges



A toy example

- Compared to Spearman's ρ, $r^{\#}$ has a higher resolution and is more accurate

$$x=(4,3,2,1)$$

$$y_1=(5,4,3,2.00)\rightarrow r^{\#}(x,y_1)=1.00, \rho(x,y_1)=1.00$$

$$y_2=(5,4,3,3.25)\rightarrow r^{\#}(x,y_2)=0.93, \rho(x,y_2)=0.80$$

$$y_3=(5,4,3,3.50)\rightarrow r^{\#}(x,y_3)=0.85, \rho(x,y_3)=0.80$$

$$y_4=(5,4,3,3.75)\rightarrow r^{\#}(x,y_4)=0.76, \rho(x,y_4)=0.80$$

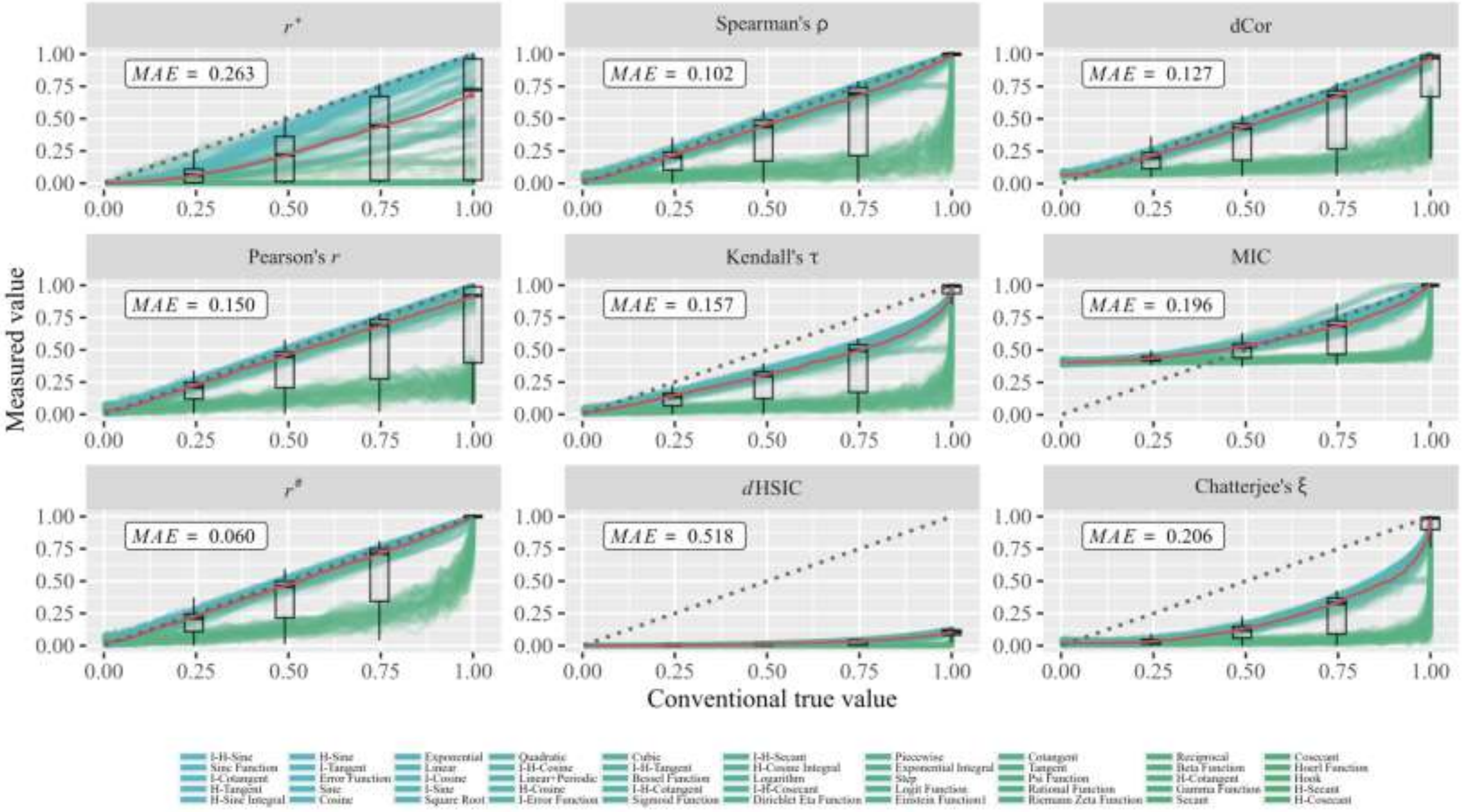
$$y_5=(5,4,3,4.50)\rightarrow r^{\#}(x,y_5)=0.38, \rho(x,y_5)=0.40$$

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.

NEURAL INFORMATION PROCESSING SYSTEMS

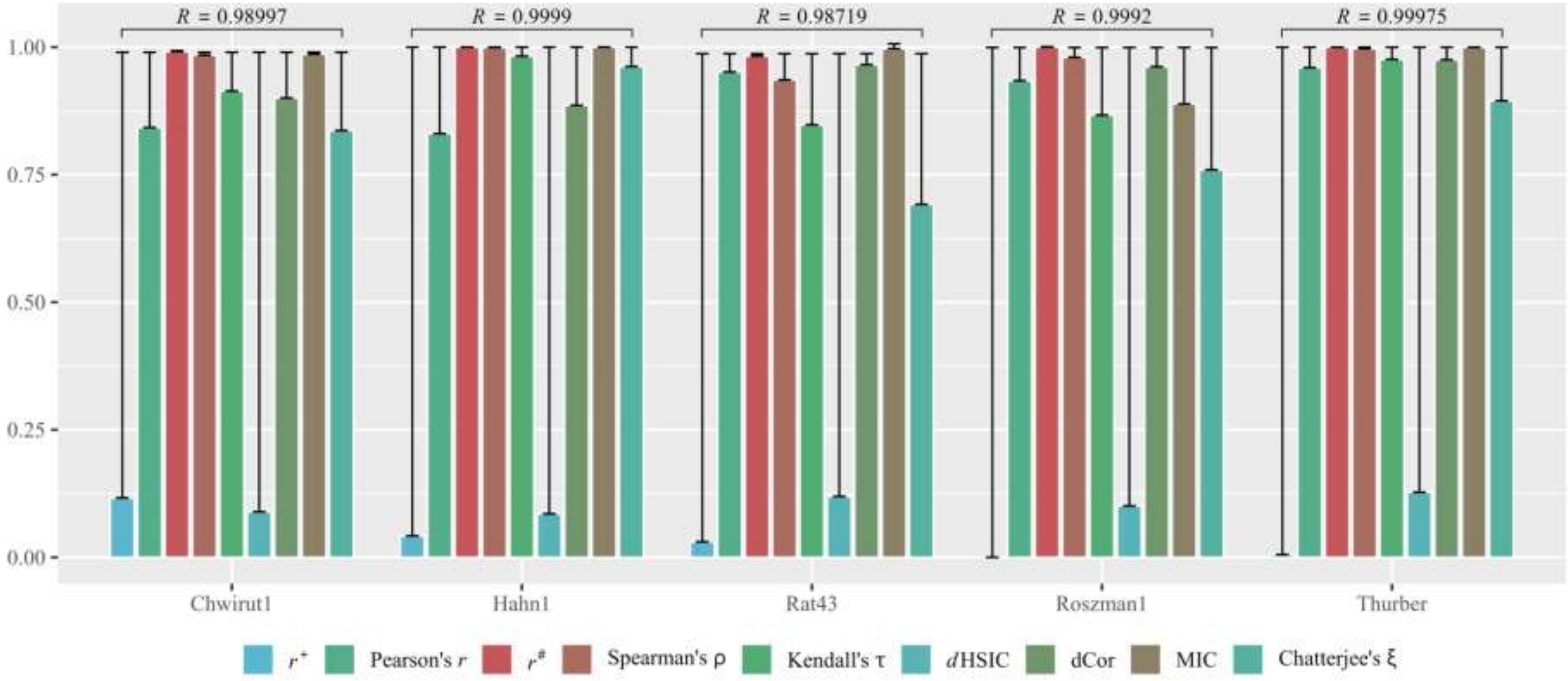
Performance of different measures

- in 50 simulated scenarios



Performance of different measures

- in 5 Real-life Scenarios



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*.

北京郵電大學
BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

NEURAL INFORMATION PROCESSING SYSTEMS

Thank you