

MICHAEL CLARK
CENTER FOR SOCIAL RESEARCH
UNIVERSITY OF NOTRE DAME

A COMPARISON OF CORRELATION MEASURES

Contents

<i>Preface</i>	3
<i>Introduction</i>	4
<i>Pearson Correlation</i>	4
<i>Spearman's Measure</i>	5
<i>Hoeffding's D</i>	5
<i>Distance Correlation</i>	5
<i>Mutual Information and the Maximal Information Coefficient</i>	6
<i>Linear Relationships</i>	7
<i>Results</i>	7
<i>Other Relationships</i>	8
<i>Results</i>	8
<i>Less noisy</i>	8
<i>Noisier</i>	9
<i>Summary</i>	9
<i>Appendix</i>	11
<i>Mine Results</i>	11
<i>Statistics: Linear Relationships</i>	12
<i>MIC tends toward zero for independent data</i>	12
<i>Statistics: Nonlinear Relationships</i>	13
<i>Pearson and Spearman Distributions</i>	13
<i>Hoeffding's D Distributions</i>	13
<i>Less Noisy Patterns</i>	14
<i>Noisier Patterns</i>	15
<i>Additional Information</i>	16

Preface

This document provides a brief comparison of the ubiquitous Pearson product-moment correlation coefficient with other approaches measuring dependencies among variables, and attempts to summarize some recent articles regarding the newer measures. No particular statistical background is assumed besides a basic understanding of correlation. To view graphs as they are intended to be seen, make sure that the 'enhance thin lines' option is unchecked in your Acrobat Reader preferences, or just use another pdf reader.

Current version date May 2, 2013. Original April 2013. I will likely be coming back to this for further investigation and may make notable changes at that time.

Introduction

The Pearson correlation coefficient has been the workhorse for understanding bivariate relationships for over a century of statistical practice. It's easy calculation and interpretability means it is the go to measure of association in the overwhelming majority of applied practice.

Unfortunately, the Pearson r is not a useful measure of dependency *in general*. Not only does correlation not guarantee a causal relationship as Joe Blow on the street is quick to remind you, a lack of correlation does not even mean there is no relationship between two variables. For one, it is best suited to continuous, normally distributed data¹, and is easily swayed by extreme values. It is also a measure of *linear* dependency, and so will misrepresent any relationship that isn't linear, which occurs very often in practice.

Here we will examine how the Pearson r compares to other measures we might use for both linear and nonlinear relationships. In particular we will look at two measures *distance correlation* and the *maximal information coefficient*.

¹ Not that that stops people from using it for other things.

Pearson Correlation

As a reminder, the sample Pearson r is calculated as follows:

$$cov_{xy} = \sum_{i=1}^N \frac{(x_i - \bar{X})(y_i - \bar{Y})}{N-1}$$

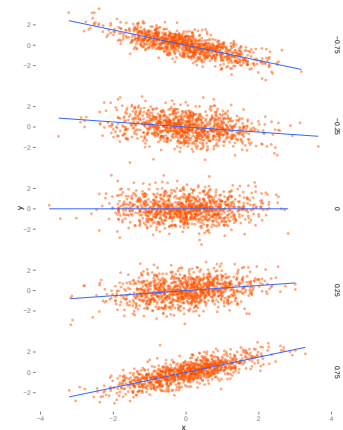
$$r_{xy} = \frac{cov_{xy}}{\sqrt{var_x var_y}}$$

In the above, we have variables X and Y for which we have N paired observations. The Pearson r is a standardized covariance, and ranges from -1, indicating a perfect negative linear relationship, and +1, indicating a perfect positive relationship. A value of zero suggests no *linear* association, but does not mean two variables are *independent*, an extremely important point to remember.

The graph to the right shows examples of different correlations with the regression line imposed. The following code will allow you to simulate your own.

```
library(MASS)
cormat = matrix(c(1, 0.25, 0.25, 1), ncol = 2) #.25 population correlation
set.seed(1234)
# empirical argument will reproduce the correlation exactly if TRUE
mydat = mvrnorm(100, mu = c(0, 0), Sigma = cormat, empirical = T)
cor(mydat)

##      [,1] [,2]
## [1,] 1.00 0.25
## [2,] 0.25 1.00
```



Spearman's Measure

More or less just for giggles, we'll also take a look at Spearman's ρ . It is essentially Pearson's r on the ranked values rather than the observed values². While it would perhaps be of use with extreme values with otherwise normally distributed data, we don't have that situation here. However since it is a common alternative and takes no more effort to produce, it is included.

² If you have ordinal data you might also consider using a polychoric correlation, e.g. in the [psych](#) package.

```
cor(mydat, method = "spearman") #slight difference

##           [,1]      [,2]
## [1,]  1.0000 0.1919
## [2,] 0.1919 1.0000

cor(rank(mydat[, 1]), rank(mydat[, 2]))

## [1] 0.1919
```

Hoeffding's D

Hoeffding's D is another rank based approach that has been around a while³. It measures the difference between the joint ranks of (X,Y) and the product of their marginal ranks. Unlike the Pearson or Spearman measures, it can pick up on nonlinear relationships, and as such would be worth examining as well.

³ Hoeffding (1948). A non-parametric test of independence.

```
library(Hmisc)
hoeffd(mydat)$D

##           [,1]      [,2]
## [1,]  1.00000 0.01162
## [2,] 0.01162 1.00000
```

Hoeffding's D lies on the interval $[-.5,1]$ if there are no tied ranks, with larger values indicating a stronger relationship between the variables.

Distance Correlation

Distance correlation ($dCor$) is a newer measure of association (Székely et al., 2007; Székely and Rizzo, 2009) that uses the distances between observations as part of its calculation.

If we define a transformed distance matrix⁴ $\mathcal{A}$ and $\mathcal{B}$ for the X and Y variables respectively, each with elements (i,j) , then the distance covariance is defined as the square root of:

$$\mathcal{V}_{xy}^2 = \frac{1}{n^2} \sum_{i,j=1}^n \mathcal{A}_{ij} \mathcal{B}_{ij}$$

⁴ The standard matrix of euclidean distances with the row/column means subtracted and grand mean added. Elements may be squared or not.

and $dCor$ as the square root of

$$\mathcal{R}^2 = \frac{\mathcal{V}_{xy}^2}{\mathcal{V}_x \mathcal{V}_y}$$

Distance correlation satisfies $0 \leq \mathcal{R} \leq 1$, and $\mathcal{R} = 0$ only if X and Y are independent. In the bivariate normal case, $\mathcal{R} \leq |r|$ and equals one if $r \pm 1$.

Note that one can obtain a $dCor$ value for X and Y of arbitrary dimension (i.e. for whole matrices, one can obtain a multivariate estimate), and one could also incorporate a rank-based version of this metric as well.

Mutual Information and the Maximal Information Coefficient

Touted as a 'correlation for the 21st century' (Speed, 2011), the *maximal information coefficient* (MIC) is based on concepts from information theory. We can note *entropy* as a measure of uncertainty, defined for a discrete distribution with K states as:

$$\mathcal{H}(X) = - \sum_{i=k}^k p(X = k) \log_2 p(X = k)$$

A uniform distribution where each state was equally likely would have maximum entropy.

Mutual Information, a measure of how much information two variables share, then is defined as:

$$\mathcal{I}(X; Y) = \mathcal{H}(X) + \mathcal{H}(Y) - \mathcal{H}(X, Y)$$

or in terms of conditional entropy:

$$\mathcal{I}(X; Y) = \mathcal{H}(X) - \mathcal{H}(X|Y)$$

Note that $\mathcal{I}(X; Y) = \mathcal{I}(Y; X)$. Mutual information provides the amount of information one variable reveals about another between variables of any type, ranges from 0 to ∞ , does not depend on the functional form underlying the relationship. Normalized variants are possible as well.

For continuous variables, the problem becomes more difficult, but if we 'bin' or *discretize* the data, it then becomes possible. Conceptually we can think of placing a grid on a scatterplot of X and Y , and assign the continuous x (y) to the column (row) bin it belongs to. At that point we can then calculate the mutual information for X and Y .

$$\mathcal{I}(X; Y) = \sum_{X,Y} p(X, Y) \log_2 \frac{p(X, Y)}{p(X)p(Y)}$$

With $p(X, Y)$ the proportion of data falling into bin X, Y , i.e. $p(X, Y)$ is the joint distribution of X and Y .

The MIC (Reshef et al., 2011) can be seen as the continuous variable counterpart to mutual information. However, the above is seen as a naïve estimate, and typically will overestimate $\mathcal{I}(X; Y)$. With the MIC a search is done over various possible grids, and the MIC is the

maximum $\mathcal{I}$ found over that search, with some normalization to make the values from different grids comparable.

$$MIC(X;Y) = \max_{X,Y_{total} < B} \frac{\mathcal{I}(X;Y)}{\log_2(\min(X,Y))}$$

In the above, $\mathcal{I}$ is the naïve mutual information measure, which is divided by the lesser number of X or Y bins. X, Y_{total} is the total number of bins, B is some number, somewhat arbitrarily chosen, though [Reshef et al. \(2011\)](#) suggest a default of N^6 or N^{55} based on their experiences with various data sets.

The authors make code available on [the web](#)⁵, but just for demonstration we can use the minerva library as follows⁶.

```
library(minerva)
mine(mydat, n.cores = 3)$MIC

##           [,1]    [,2]
## [1,]  1.0000  0.2487
## [2,]  0.2487  1.0000
```

MIC is on the interval $[0, 1]$ where zero would indicate independence and a 1 would indicate a noiseless functional relationship. With MIC the goal is equitability- similar scores will be seen in relationships with similar noise levels regardless of the type of relationship. Because of this it may be particularly useful with high dimensional settings to find a smaller set of the strongest correlations. Where distance correlation might be better at detecting the *presence* of (possibly weak) dependencies, the MIC is more geared toward the assessment of strength and detecting patterns that we would pick up via visual inspection.

Linear Relationships

Just as a grounding we will start with examination of linear relationships. Random (standardized) normal data of size $N = 1000$ has been generated 1000 times for population correlations of -.8, -.6, -.4, 0, .4, .6, and .8 (i.e. for each correlation we create 1000 x,y data sets of $N = 1000$). For each data set we calculate each statistic discussed above.

Results

Distributions of each statistic are provided in the margin⁷ (squared values for Pearson and Spearman measures). Means, standard deviations, and quantiles at the 2.5% and 97.5% levels are provided in the [appendix](#). No real news to tell with the Pearson and Spearman correlations. The other measures pick up on the relationships in a manner

⁵ See the appendix regarding an issue I found with this function.

⁶ While the package availability is nice, the mine function in minerva works relatively very slow (even with the parallel option as in the example).

⁷ For this and the later similar graph, rotate the document in for a better view.

we'd hope, and assign the same values whether the original linear relation is positive or negative. Interestingly, the Hoeffding's D and MIC appear to get more variable as the values move away from zero, while for the $dCor$ gets less so. At first glance the MIC may seem to find something in nothing, in the sense it doesn't really get close to zero for the zero linear relationship. As seen later, it finds about the same value for the same pattern split into four clusters. This is a function of sample size, where MIC for independent data approaches 0 as $N \rightarrow \infty$. See the supplemental material for [Reshef et al. \(2011\)](#), particularly figure S1. I also provide an example of MICs for different sample sizes and standard deviations in the [appendix](#).

The take home message at this point is that one should feel comfortable using the other measures for standard linear relationships, as one would come to the same relative conclusions as one would with the Pearson r . Note that for perfect linear relations all statistics would be 1.0.

Other Relationships

More interesting is a comparison of these alternatives when the relationship is not linear. To this end, seven other patterns are investigated to see how these statistics compare. In the margin are examples of the patterns at two different noise levels. The patterns will be referred to as wave, trapezoid, diamond, quadratic, X, circle and cluster.

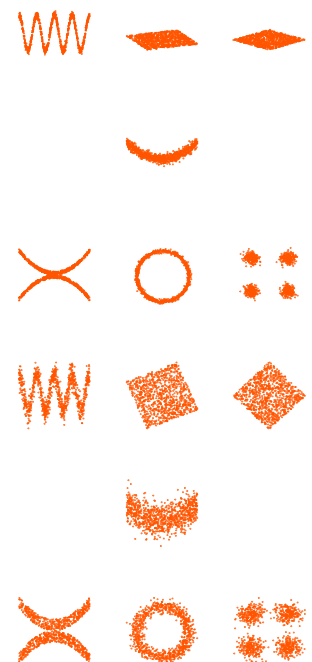
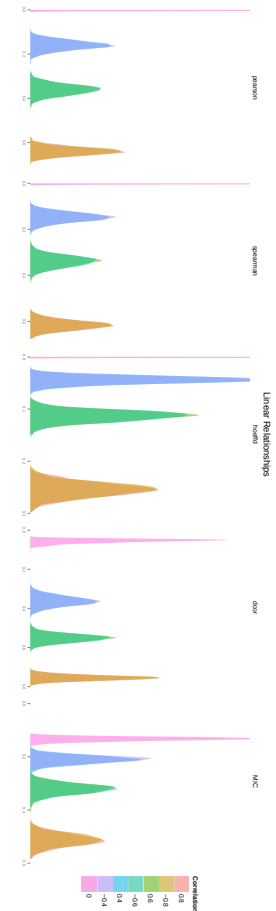
Results

Distributional results for the distance correlation and MIC can again be seen in the margin, related statistics and distributions for the those and the other statistics can be found in the [appendix](#). Neither Pearson's r nor Spearman's ρ find a relationship among any of the patterns regardless of noise level, and will no longer be a point of focus.

Less noisy

Hoeffding's D shows very little variability within its estimates for any particular pattern, and does not vary much between patterns (means range 0 to .1). In this less noisy situation Hoeffding's D does pick up on the quadratic pattern relative to the others, followed by the X and circle patterns, but in general these are fairly small values for any pattern.

The $dCor$ and MIC show notable values for most of these patterns. The $dCor$ finds the strongest relationship for the quadratic function,



followed by the X, circle, the trapezoidal and wave patterns in a group, with the cluster pattern near zero. The MIC finds a very strong relationship for the wave pattern, followed by the quadratic, about the same for the circle and X pattern, and the rest filling out toward its lower values. Just as before, the MIC won't quite reach 0 for $N = 1000$, so those last three are probably reflective of little dependence according to MIC.

Noisier

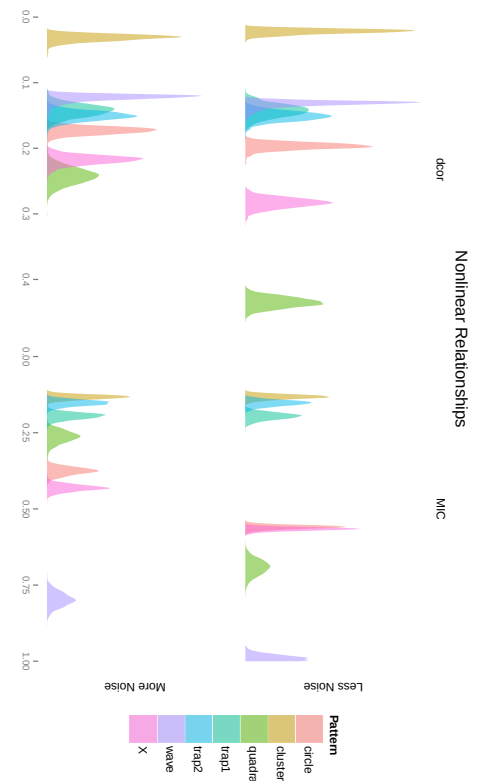
With more noise, Hoeffding's D does not pick up on the patterns well; the means now range from 0 to .02. The *dCor* maintains its previous ordering, although in general the values are smaller, or essentially the same in the case of the trapezoidal and cluster patterns for which strong dependency was not uncovered previously. The MIC still finds a strong dependency in the wave pattern, as well as the X and circle, but the drop off for the quadratic relationship is notable, and is now deemed less of a dependency than the circle and X patterns. The remaining are essentially the same as the less noisy situation. This suggests that for some relationships the MIC will produce the same value regardless of the amount of noise, and the noise level may affect the ordering of the strength one would see of different relationships.

Summary

Pearson's r and similar measures are not designed to pick up nonlinear relationships or dependency in a general sense. Other approaches such as Hoeffding's D might do a little better in some limited scenarios, and a statistical test⁸ for it might suggest dependence given a particular nonlinear situation. However it appears one would not get a good sense of the strength of that dependence, nor would various patterns of dependency be picked up. It also does not have the sort of properties the MIC is attempting to hold.

Both distance correlation and MIC seem up to the challenge to find dependencies beyond the linear realm. However neither is perfect, and Kinney and Atwal (2013) note several issues with MIC in particular. They show that MIC is actually not equitable⁹, the key property Reshef and co. were claiming, nor is it 'self-equitable' either (neither is *dCor*). Kinney and Atwal also show, and we have seen here, that variable noise may not affect the MIC value for certain relationships, and they further note that one could have a MIC of 1.0 for differing noise levels.

As for distance correlation, Simon and Tibshirani (2011, see [additional information](#) for a link to their comment) show that *dCor*



⁸ The `hoeffd` function does return a p-value if interested.

⁹ The R^2 measure of equitability Reshef et al. were using doesn't appear to be a useful way to measure it either.

exhibits more statistical power than the MIC. We have also seen it will tend to zero even with smaller sample sizes, and that it preserved the ordering of dependencies found across noise levels. Furthermore, $dCor$ is straightforward to calculate and not an approximation.

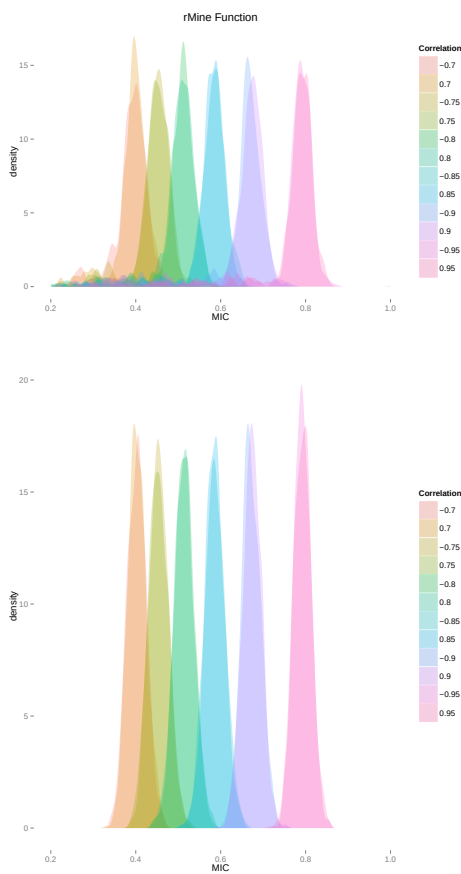
In the end we may still need to be on the lookout for a measure that is both highly interpretable and possesses all the desirable qualities we want, but we certainly have do have measures that are viable already. Kinney and Atwal (2013) suggest that the issues surrounding mutual information $\mathcal{I}$ that concern Reshef et al. were more a sample size issue, and that those difficulties vanish with very large amounts of data. For smaller sample sizes and/or to save computational costs, Kinney and Atwal suggest $dCor$ would be a viable approach.

I think the most important conclusion to draw is to try something new. Pearson's simply is not viable for understanding a great many dependencies that one will regularly come across in one's data adventures. Both $dCor$ and mutual information seem much better alternatives for picking up on a wide variety of relationships that variables might exhibit, and can be as useful as our old tools. Feel free to experiment!

Appendix

Mine Results

Note on the rMine code available at <http://www.exploredata.net/Downloads/MINE-Application>. I have come across a couple of issues. One is that it will simply not work (as the Tibshirani comment refers to it, a 'glitch'), producing a missing value. But in addition to that, I found that it also appears to produce odd values every once in a while. Consider the following graphs in which various notably high linear relationships are examined, with the rMine function and the **mine** function from the *minerva* respectively. The rMine function bounces around quite a bit, even tending toward very low values on regular occasion, although the peak of the distribution is about right. In contrast the minerva function shows more symmetric and stable distributions, even with a little bit of bump in the tails. The rMine code was beta and dated Jan. 2012, so it may have changed by the time you play with it.



Statistics: Linear Relationships

Means.

pop.	pearson	spearman	hoeffd	dcor	MIC
.8	0.799	0.785	0.255	0.755	0.513
.6	0.600	0.582	0.113	0.552	0.321
.4	0.399	0.383	0.044	0.363	0.211
0	-0.001	-0.001	0.000	0.055	0.133
-.4	-0.399	-0.384	0.044	0.363	0.210
-.6	-0.600	-0.582	0.113	0.552	0.320
-.8	-0.799	-0.785	0.254	0.755	0.512

Standard deviations.

pearson	spearman	hoeffd	dcor	MIC
0.011	0.013	0.014	0.013	0.024
0.020	0.022	0.011	0.021	0.021
0.027	0.028	0.007	0.026	0.016
0.031	0.031	0.000	0.010	0.007
0.027	0.027	0.007	0.025	0.014
0.020	0.022	0.011	0.021	0.020
0.012	0.013	0.014	0.013	0.023

Quantiles at the 2.5% and 97.5% levels.

pearson	spearman	hoeffd	dcor	MIC
[0.777 0.820]	[0.758 0.810]	[0.228 0.282]	[0.729 0.780]	[0.467 0.560]
[0.560 0.640]	[0.540 0.624]	[0.093 0.135]	[0.513 0.593]	[0.281 0.366]
[0.345 0.452]	[0.328 0.439]	[0.031 0.059]	[0.311 0.416]	[0.182 0.242]
[-0.062 0.060]	[-0.062 0.058]	[0.000 0.001]	[0.040 0.080]	[0.120 0.149]
[-0.449 -0.344]	[-0.436 -0.327]	[0.030 0.058]	[0.309 0.412]	[0.183 0.240]
[-0.638 -0.561]	[-0.624 -0.538]	[0.093 0.135]	[0.509 0.592]	[0.282 0.360]
[-0.821 -0.775]	[-0.810 -0.757]	[0.226 0.281]	[0.726 0.780]	[0.466 0.558]

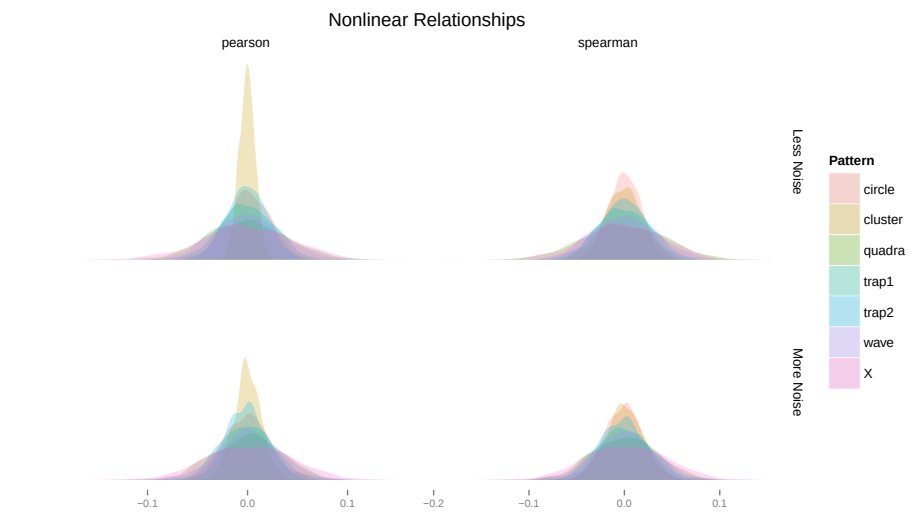
MIC tends toward zero for independent data

MIC for x and y random draws of size N and $N(0, sd)$.

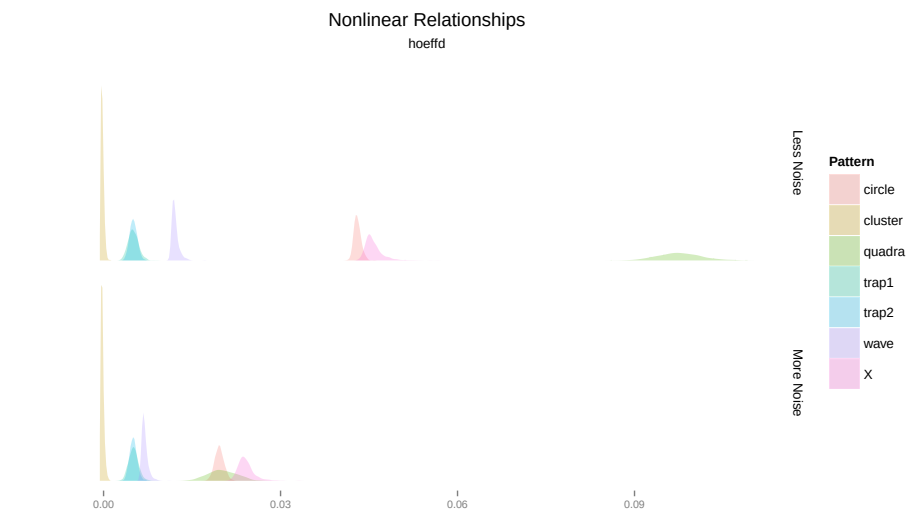
N	10	1	0.1	0.01
100	0.251	0.263	0.228	0.255
500	0.166	0.146	0.150	0.160
1000	0.144	0.137	0.133	0.144
2500	0.097	0.095	0.101	0.101
5000	0.084	0.078	0.082	0.079
10000	0.064	0.065	0.062	0.063
20000	0.050	0.046	0.046	0.045

Statistics: Nonlinear Relationships

Pearson and Spearman Distributions



Hoeffding's D Distributions



Less Noisy Patterns

Means.

pattern	pearson	spearman	hoeffd	dcor	MIC
<i>wave</i>	0.000	0.000	0.012	0.132	0.988
<i>trapezoid</i>	-0.001	-0.001	0.005	0.143	0.196
<i>diamond</i>	-0.001	-0.001	0.005	0.151	0.153
<i>quadra</i>	0.000	-0.001	0.098	0.435	0.692
<i>X</i>	-0.001	-0.001	0.046	0.283	0.566
<i>circle</i>	0.000	0.000	0.043	0.197	0.560
<i>cluster</i>	-0.001	-0.002	0.000	0.022	0.133

Standard deviations.

pattern	pearson	spearman	hoeffd	dcor	MIC
<i>wave</i>	0.033	0.034	0.001	0.006	0.010
<i>trapezoid</i>	0.027	0.029	0.001	0.010	0.012
<i>diamond</i>	0.020	0.024	0.001	0.007	0.010
<i>quadra</i>	0.039	0.042	0.004	0.008	0.026
<i>X</i>	0.044	0.042	0.002	0.008	0.006
<i>circle</i>	0.022	0.018	0.001	0.005	0.007
<i>cluster</i>	0.008	0.021	0.000	0.004	0.008

Quantiles at the 2.5% and 97.5% levels.

pattern	pearson	spearman	hoeffd	dcor	MIC
<i>wave</i>	[-0.062 0.064]	[-0.062 0.068]	[0.011 0.014]	[0.125 0.148]	[0.965 1.000]
<i>trapezoid</i>	[-0.053 0.048]	[-0.057 0.053]	[0.003 0.007]	[0.123 0.163]	[0.174 0.221]
<i>diamond</i>	[-0.040 0.037]	[-0.048 0.047]	[0.004 0.006]	[0.137 0.166]	[0.135 0.174]
<i>quadra</i>	[-0.074 0.076]	[-0.083 0.083]	[0.091 0.105]	[0.420 0.451]	[0.642 0.746]
<i>X</i>	[-0.093 0.084]	[-0.085 0.077]	[0.044 0.050]	[0.269 0.299]	[0.556 0.578]
<i>circle</i>	[-0.043 0.042]	[-0.034 0.034]	[0.042 0.044]	[0.188 0.208]	[0.548 0.573]
<i>cluster</i>	[-0.017 0.014]	[-0.045 0.039]	[-0.001 0.000]	[0.016 0.030]	[0.119 0.149]

Noisier Patterns

Means.

pattern	pearson	spearman	hoeffd	dcor	MIC
<i>wave</i>	0.001	0.001	0.007	0.123	0.797
<i>trapezoid</i>	-0.001	0.000	0.005	0.143	0.196
<i>diamond</i>	0.000	-0.001	0.005	0.151	0.153
<i>quadra</i>	0.000	0.000	0.020	0.242	0.264
<i>X</i>	0.001	0.001	0.024	0.218	0.431
<i>circle</i>	0.000	0.000	0.020	0.173	0.376
<i>cluster</i>	0.000	0.000	0.000	0.032	0.133

Standard deviations.

pattern	pearson	spearman	hoeffd	dcor	MIC
<i>wave</i>	0.032	0.032	0.001	0.006	0.024
<i>trapezoid</i>	0.027	0.029	0.001	0.010	0.011
<i>diamond</i>	0.020	0.024	0.001	0.007	0.010
<i>quadra</i>	0.033	0.034	0.002	0.013	0.021
<i>X</i>	0.043	0.042	0.001	0.007	0.011
<i>circle</i>	0.023	0.020	0.001	0.006	0.013
<i>cluster</i>	0.014	0.021	0.000	0.005	0.008

Quantiles at the 2.5% and 97.5% levels.

pattern	pearson	spearman	hoeffd	dcor	MIC
<i>wave</i>	[-0.066 0.062]	[-0.063 0.063]	[0.006 0.008]	[0.116 0.139]	[0.750 0.843]
<i>trapezoid</i>	[-0.049 0.053]	[-0.052 0.054]	[0.004 0.007]	[0.124 0.162]	[0.176 0.219]
<i>diamond</i>	[-0.038 0.039]	[-0.048 0.049]	[0.004 0.006]	[0.137 0.166]	[0.135 0.173]
<i>quadra</i>	[-0.061 0.063]	[-0.065 0.066]	[0.016 0.025]	[0.218 0.268]	[0.229 0.307]
<i>X</i>	[-0.080 0.084]	[-0.078 0.080]	[0.022 0.028]	[0.204 0.235]	[0.409 0.454]
<i>circle</i>	[-0.043 0.048]	[-0.038 0.042]	[0.018 0.021]	[0.163 0.185]	[0.352 0.403]
<i>cluster</i>	[-0.028 0.027]	[-0.040 0.041]	[-0.001 0.000]	[0.023 0.043]	[0.120 0.148]

References

- Fujita, A., Sato, J. R., Demasi, M. A. A., Sogayar, M. C., Ferreira, C. E., and Miyano, S. (2009). Comparing Pearson, Spearman and Hoeffding's D measure for gene expression association analysis. *Journal of bioinformatics and computational biology*, 7(4):663–684. PMID: 19634197.
- Heller, R., Heller, Y., and Gorfine, M. (2012). A consistent multivariate test of association based on ranks of distances. *arXiv:1201.3522*. Biometrika 2012.
- Kinney, J. B. and Atwal, G. S. (2013). Equitability, mutual information, and the maximal information coefficient. *arXiv:1301.7745*.
- Reshef, D., Reshef, Y., Mitzenmacher, M., and Sabeti, P. (2013). Equitability analysis of the maximal information coefficient, with comparisons. *arXiv:1301.6314*.
- Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., and Sabeti, P. C. (2011). Detecting novel associations in large data sets. *Science*, 334(6062):1518–1524.
- Simon, N. and Tibshirani, R. (2011). Comment on "Detecting Novel Associations in Large Data Sets" by Reshef et al. pages 1–3.
- Székely, G. J. and Rizzo, M. L. (2009). Brownian distance covariance. *The Annals of Applied Statistics*, 3(4):1236–1265. Mathematical Reviews number (MathSciNet): MR2752127.
- Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794. Mathematical Reviews number (MathSciNet): MR2382665; Zentralblatt MATH identifier: 1129.62059.

Additional Information

Some R packages to calculate mutual information (package names linked):

[minet](#): implements various algorithms for inferring mutual information networks from data¹⁰

[infotheo](#): for continuous variables requires discretization via [discretize](#) function

[mpmi](#): Fast calculation mutual information for comparisons between all types of variables including continuous vs continuous, continuous vs discrete and discrete vs discrete.

R package to calculate distance correlation (package names linked):

[energy](#): E-statistics (energy) tests and statistics for comparing distributions. From Székely and Rizzo, authors of the papers cited.

R package for the [HHG](#) measure of association/test; provided by the cited author. I may come back to this in a future version of this paper.

Commentary for [Reshef et al. \(2011\)](#) at article [website](#).

Some discussion at Andrew Gelman's blog [\(1\)](#) [\(2\)](#)

Simon and Tibshirani's [comment](#).

The code I used for this paper was a modification of that available both at [Tibshirani's website](#), and code available for the patterns seen at the [Wikipedia entry](#) for the Pearson r .

¹⁰ Via the Bioconductor repository.