

Measuring Non-Linear Associations

Julian Roche

School of Mathematics and Statistics

University of Sheffield

September 2015

Abstract of Measuring Non-linear Associations

Author: Julian Roche

Date: September 2015

This study compared three different measures of association (a probabilistic measure of association called A , distance correlation ($dcor$) and Maximal Information Coefficient (MIC)) at identifying inter-variable relationships in large biological datasets. The comparisons are based on power simulations of two-way associations using a range of sample sizes, sampling distributions and additive noise levels.

The simulations found that no measure of association had greater statistical power in all cases. Using the gene expression data, $dcor$ identified all strong associations identified by A in the two-way case. The analysis suggests that A suffers from some of the same power deficiencies that have been previously identified for MIC .

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Objectives	2
1.3	Overview	2
2	Association and Non-Linearity	3
2.1	Measures of Association	3
2.2	Assessing Non-Linearity	7
3	Simulation	8
3.1	Introduction	8
3.2	Power versus Noise	10
3.3	Power versus Sample Size	13
4	Discussion and Further Work	16
A	Additional Results	20
A.1	Simulation Results	21

Chapter 1

Introduction

1.1 Motivation

Big Data is a fashionable term currently sweeping through the corporate world. While not strictly defined, it typically refers to large, complex datasets that are too big to be managed using traditional methods. Depending on the definition different statistical and computational challenges are posed. Examples of big data within pharmaceutical research include gene expression data, medical histories and clinical trial data. Often these have many more parameters than observations so $p \gg n$. The availability of these datasets combined with recent developments for identifying associations between variables offers the potential to improve biological understanding which can in turn aid in the development of treatments for different diseases.

One of the main purposes of association analysis is to produce an ordered list of variable pairs ranked by the strength of association. Typically these would form the basis of a more in-depth statistical analysis or highlight areas for further investigation or experimentation. The size of the datasets raises challenges for exploratory association mining. For example, when investigating n pairwise combinations of variables there are $\frac{1}{2} \times n \times (n - 1)$ combinations so that even for a dataset of 100 variables, there are 4095 different pairs to be analysed.

A desirable property of estimators of association is that they can identify a range of different functional relationships equally and not for example be limited to just linear relationships, as is the case with Pearson correlation. This has been highlighted as a potentially useful approach for studying gene expression where relationships may be determined by multiple distant factors¹. It is also important that different measures of association have sufficient statistical power when dealing with relatively small sample sizes. The situation becomes more complex when investigating multivariate associations or controlling for specific variables. Therefore automated, robust and efficient approaches to searching very large collections of variables are needed.

1.2 Objectives

The objective of the project is to compare three different measures of association (A , distance correlation ($dcor$) and Maximal Information Coefficient (MIC) at identifying inter-variable relationships and use these to develop efficient methods for analysing relationships in large biological datasets. The comparisons are based on simulated data and two gene expression datasets comparing lupus patients (see Section 1.5) with controls: a subset of genes deemed to be of biological interest (referred to as the training dataset) and a full Affymetrix chip (referred to as the test dataset)

1.3 Overview

Chapter 2 introduces the association measures studied and briefly reviews the concepts of dependence, association and non-linearity. Chapter 3 reports the results of a simulation study used to investigate the power of different association measures to detect dependence for a variety of functional relationships using various amounts of noise and sample sizes. Chapter 4 discusses the analysis and provides suggestions for further work. The Appendices provide ancillary material and code.

Chapter 2

Association and Non-Linearity

Background

A linear function of two random variables is one that can be written in the form:

$$y = ax + b$$

where a and b are constants. Thus a non-linear function is a relationship which cannot be written as a linear combination of the variables. Some examples of nonlinear functions are given and explored through simulation in Chapter 3.

If no relationship exists between two random variables they are said to be statistically independent. Formally, two events A and B are defined to be independent if:

$$P(A \cap B) = P(A)P(B)$$

Conversely, if two random variables X and Y are dependent then as X changes in value, the value Y will also change and vice versa. Association measures the statistical dependence between two or more random variables. In the case of two random variables this is a symmetric relationship.

2.1 Measures of Association

There are many different statistics used to measure the association between random variables. A full comparison of all of these is obviously not practical so the measures explored through simulation and on the lupus datasets are introduced here. Other measures of association which are potentially of interest include mutual information and the Heller, Heller and Gorfine (HHG) measure². By convention association is measured on a $0 - 1$ scale, 0 implying independence and 1 implying complete dependence. Correlation is also a measure of association but additionally includes the direction of the relationship between two random variables.

Generality and Equitability

Two potentially desirable properties of measures of association, introduced by Reshef et al.³, are:

- Generality: The ability to detect any departure from independence given a sufficiently large sample.
- Equitability: The ability to assign similar values of association to equally noisy relationships of different types.

The equitability property is of particular interest in large scale association studies as it offers the potential to efficiently search for pairwise relationships of different types in large datasets. It has attracted considerable interest in the context of biological studies¹ as a heuristic approach to discovering associations of interest. However, the property has also drawn criticism due to the fact that unless the measure has sufficient statistical power, inflated false discovery may reduce the utility of a given measure of association despite the equitability property⁴. Subsequent work by Kinney⁵, which established a mathematical foundation for equitability based on R^2 , suggests that R^2 equitability cannot be satisfied by any measure of association and therefore offers little practical value in exploratory analysis. However they introduced the notion of ‘self-equitability’ which is falsifiable and therefore offers a way to compare measures of association. Kinney⁵ proved that MIC is not self-equitable but that an information theoretic measure called mutual information is and hence has potential for large scale association studies.

Pearson Correlation

The Pearson correlation coefficient, ρ , measures the strength and direction of a linear association between two r.v.’s, X and Y . It is defined as

$$\rho = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$$

where $cov(X, Y)$ is the covariance between X and Y and σ_X and σ_Y are the standard deviations of X and Y respectively.

The sample correlation coefficient is estimated as follows:

$$\hat{\rho} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_X s_Y}$$

where \bar{x} and \bar{y} are the sample means and s_X and s_Y the sample standard deviations of X and Y respectively. R^2 is known as the coefficient of determination and measures the amount of observed variance explained by a function. In the case of a linear relationship between two variables, $R^2 = \rho^2$. When the form of the functional relationship between

two variables is unknown, obtaining a value of R^2 is more challenging⁶ and has led to the development of many other measures of association.

Spearman's Rank Correlation

Spearman's rank correlation is a non-parametric measure of association of paired data. For two r.v.'s X and Y , it is calculated by ordering the x 's and y 's separately and then calculating the differences d in the rank order between the two⁷. Therefore a Spearman correlation of 1 means that X and Y are in the same order. When ignoring ties it is given by the following formula:

$$\rho_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

While Spearman's rank can describe non-linear relationships better than Pearson's correlation, it has been shown to not be equitable³.

Maximal Information Coefficient

Maximal Information Coefficient (MIC) is based on the concept that if two random variables, X and Y are associated, the association can be defined by drawing a grid on a scatterplot of X vs. Y . An outline of the algorithm for calculating MIC is as follows:

1. For each pair of r.v.'s X and Y find the grid between X and Y with the highest mutual information (a measure of the variables' mutual dependence).
2. Normalise the mutual information scores in order to allow fair comparison of grids (grid dimensions may differ between pairs).
3. The normalised scores form a characteristic matrix, the maximum of which is the MIC score.

The following formal definition of MIC is from Reshef et al.³: for a grid G , let I_G denote the mutual information of the probability distribution induced on the cells of G , where the probability of a cell is proportional to the number of points bounded by the cell. $m_{x,y} = \frac{\max(I_g)}{\log(\min(x,y))}$ where $m_{x,y}$ is the (x, y) -th entry of the characteristic matrix, and the maximum is taken for all x by y grids G . The MIC score is then the maximum of $m_{x,y}$ ordered pairs (x, y) such that $xy < B$, where B is a function of the sample size (the default is $B = n^{0.6}$ but the performance of MIC is sensitive to this tuning parameter). MIC is implemented in the R *minerva*⁸ package.

A

A is a probabilistic measure of association that generalises R^2 . It has a number of attractive features that make it of potential interest for association analysis, in particular the fact that it is reported⁶ to have greater power than other measures of association such as *MIC* and *dcor*, it extends to the multivariate case (e.g. how Y is associated with both X and Z) and can also measure association while controlling for covariates (i.e. partial association, the association between Y and X while controlling for Z). Additionally it is also reported to be both general and equitable⁶.

In the case of simple linear regression $y = f(x) + \epsilon$, R^2 quantifies the amount of observed systematic variation, while the residuals are assumed to be Normally distributed around $f(x)$. *A* is based on the following reformulation of R^2 :

$$R^2 = 1 - \prod_i \left(\frac{P(x_i, y_i | \text{null})}{P(x_i, y_i | \text{alt})} \right)^{\frac{2}{n}}$$

which defines R^2 in terms of the probability density of the null and alternative models. In the case of *A*, \hat{A} can be estimated if suitable models for the null and alternative hypotheses can be found. In practice this involves estimating the density at each point of the null and alternative models using kernel density estimation. The choice of null is taken to enforce independence so $P(X, Y) = P(X)P(Y)$, while the alternative is allowed to be dependent. For a detailed technical description see Murrel et al.⁶. For simplicity the remainder of this document refers to \hat{A} as *A* unless explicitly stated otherwise. It is implemented in the R *matie*⁹ package.

Distance Correlation

Distance correlation aims to measure all types of dependence between two or more random variables¹⁰. For two random vectors X and Y of size n , *dcor* is computed as follows:

1. Compute a distance matrix of all pairwise distances between samples of the X variable.
2. Compute the corresponding distance matrix for the Y variable.
3. Mean centre both distance matrices, giving two distance matrices, $A_{k,l}$ and $B_{k,l}$, where k and l are the respective row and column indices.
4. Calculate the distance covariance: $dcov(X, Y) = \sqrt{\frac{1}{n^2} \sum_{k=1, l=1}^n A_{k,l} B_{k,l}}$.
5. Scale the distance covariance to give the distance correlation:

$$dcor(X, Y) = \frac{dcov(X, Y)}{\sqrt{dcov(X, X) dcov(Y, Y)}}$$

dcor is implemented in the R *energy*¹¹ package.

2.2 Assessing Non-Linearity

The three association measures MIC , A and $dcor$ all aim to measure all types of dependence, linear and non-linear, monotonic and non-monotonic etc. between pairs of random variables equally, however none explicitly measures the non-linear association. There are a number of possible ways of trying to assess the amount of non-linear association by partitioning the observed association into linear and non-linear components. Murrel et al.⁶ implement the following different measures of non-linearity based on A and R^2 in the *matie* R package⁹. They have been generalised here so that θ refers to any association measure.

1. Residual association ($r\theta$), the association of random variables after the linear relationship has been regressed out using a standard linear model. (See also Szekely and Rizzo¹⁰.)
2. Difference, the non-linear part of the association, $\theta - R^2$. (See also Reshef et al.³.)
3. The non-linear proportion of association, $\frac{(\theta - R^2)}{\theta}$
4. The proportion of total variance that is explained by θ but not by R^2 , $\frac{(\theta - R^2)}{1 - \theta^2}$.

Chapter 3

Simulation

3.1 Introduction

Simulation was used to study a number of measures of association. The aim was to assess their utility and limitations for use in exploratory association mining in general and for the lupus dataset in particular. The analysis focused on how the statistical power of each association measures varies with noise and sample size.

Statistical power refers to the ability to identify correctly the null hypothesis when it is true. In the case of dependence the hypothesis is that two or more sets of data are independent while the alternative hypothesis is that they are dependent. Formally, for two random variables X and Y :

H_0 : X and Y are independent.

H_1 : X and Y are dependent.

A previous simulation study by Simon and Tibshirani⁴ into the power of MIC vs. $dcor$ found that in most cases $dcor$ was more powerful than MIC and that in the linear case MIC was less powerful than Pearson's Correlation Coefficient. This power deficiency led the authors to advocate the use of $dcor$ over MIC in large scale exploratory analysis as MIC might lead to too many false positives⁴.

This simulation study extends the aforementioned study by including the A score, expanding the range of functional relationships explored, studying the effect of differing sample size and also using two source distributions.

The original study used a $Uniform(0,1)$ to generate data which were then used to create dependent and independent sets of random variables X and Y . The procedure is as follows:

Let X , Y and Z be random variables and let $f(X)$ be a function of interest. For the dependant data: $X \sim U(0,1)$ and $Y = f(X)$. For the independent data: $X \sim U(0,1)$,

$Z \sim U(0, 1)$ and $Y = f(Z)$.

Their motivation for choosing $Uniform(0,1)$ distribution was primarily because the authors of *MIC* had used it in their simulation studies^{4,3}. Exploratory analysis of the lupus dataset suggested that many of the gene expressions have positively skewed distributions, so a $Beta(2,5)$ distribution was used to simulate this.

The main interest was in comparing A , MIC and $dcor$. Pearson and Spearman correlation were also included for contrast. Fifteen different functional relationships were studied. The functions are defined in Table 3.1. Normally distributed noise was added to each function so that, $y = f(x) + l \times \epsilon$, where l is a scaling parameter controlling the level of noise, ϵ , and $\epsilon \sim N(0, 1)$. In the case of the cubic, exponential, log, step, spike and sine based functions, l was scaled to compensate for the range of y . Figure 3.1 plots each of these for a sample size of 320 with minimal noise.

Name	Function
Linear	$y = x$
Quadratic	$y = 4(x - \frac{1}{2})^2$
Cubic	$y = 128(x - \frac{1}{3})^3 - 48(x - \frac{1}{3})^3 - 12(x - \frac{1}{3})$
Fourth Root	$y = x^{\frac{1}{4}}$
Exponential	$y = e^{x^2}$
Log	$y = \ln(x)$
Sigmoid	$y = 1 + e^{10(\frac{1}{2}-x)^{-1}}$
Step	$y = \begin{cases} +1 & \text{when } x \geq 0.5 \\ -1 & \text{otherwise} \end{cases}$
Spike	$y = \begin{cases} 4 & \text{when } x \leq 0.05 \\ 1.9 - 18x & \text{when } x \leq 0.10 \\ \frac{1}{9} - \frac{x}{9} & \text{otherwise} \end{cases}$
Sine: Low	$y = \sin(4\pi x)$
Sine: High	$y = \sin(16\pi x)$
Linear + Periodic	$y = x + \sin(10\pi x)$
Varying Frequency	$y = x + \sin(5\pi \frac{1}{x})$
Circle	$y = (1 - (2x - 1)^2) \times -2z$, where $Z \sim B(1, \frac{1}{2})$
X	$y = 4(x - \frac{1}{2})^2 \times z$, where $P_Z(z) = \begin{cases} \frac{1}{2} & \text{when } z = +1 \\ \frac{1}{2} & \text{when } z = -1 \end{cases}$

Table 3.1: The fifteen functions being evaluated.

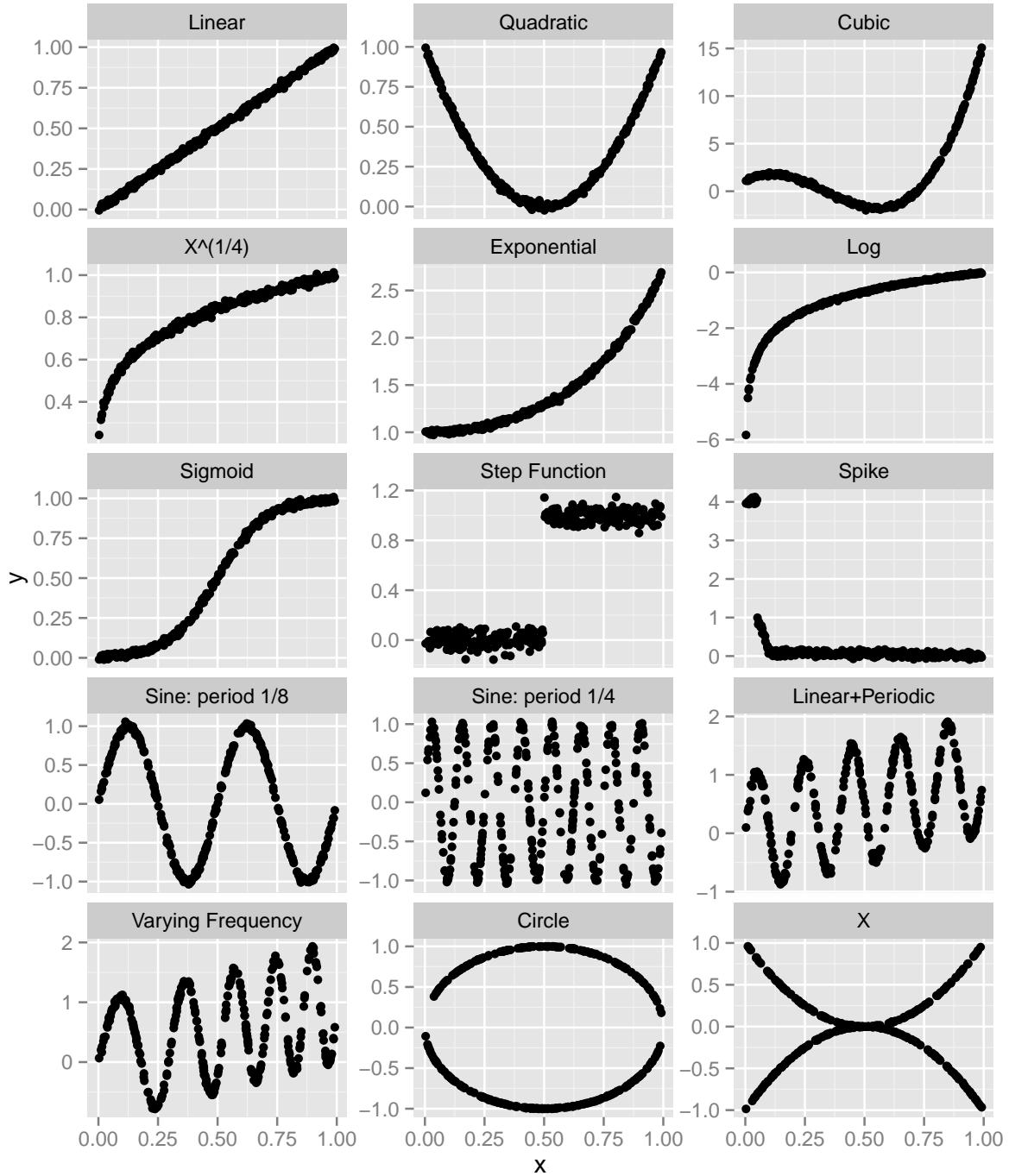


Figure 3.1: Scatter plots showing the fifteen different functional forms being evaluated.

3.2 Power versus Noise

In order for an association measure to be practically useful in association analysis it should have reasonable statistical power for noisy data. Ideally it should also be equitable, i.e. it should score equally different types of relationships with equal amounts of noise added. This simulation explored how statistical power varied for the different functional forms

for different amounts of Normally distributed noise. A sample size of 320 was used to represent a moderately size biological dataset. (Figure A.1 in Appendix A shows each functional form following the addition of noise.)

Data was generated using two distributions, *Uniform(0,1)* and *Beta(2,5)*, so for a r.v. X , $X \in \{0, 1\}$. Random noise was generated from a $N(0, 1)$ distribution and then scaled by a given noise level (0.1 to 3). The following algorithm was used in the simulation to assess if two r.v.'s X and Y were dependent:

1. For each functional form, f , and noise level, l :
2. Generate independent r.v.'s X, Y and Z : X and Z from the given distribution (so $X \in \{0, 1\}, Z \in \{0, 1\}$) and an independent r.v. $Y = f(Z) + l \times \epsilon$, where $\epsilon \sim N(0, 1)$.
3. Measure the association between X and Y using A , $dcor$, MIC , $\rho_{pearson}^2$ and $\rho_{spearman}^2$
4. Generate dependent r.v.'s X and Y : X from the given distribution and a dependent r.v. $Y = f(X) + l \times \epsilon$, where $\epsilon \sim N(0, 1)$.
5. Measure the association between X and Y using A , $dcor$, MIC , $\rho_{pearson}^2$ and $\rho_{spearman}^2$
6. Estimate the power as the number of times the association score for the dependent data exceeds the 95th percentile of the corresponding score for the independent data.
7. Repeat 500 times.

The complete *R* code, based on that of Simon and Tibshirani⁴, is given in Appendix A. The results in Figure 3.2 suggest that all measures are sensitive to the level of noise and that none of the measures has greater power for all of the functional forms simulated. In particular A appears to have low power for the fourth root, step, exponential, sigmoid and linear cases. It also has lower power than $dcor$ in the cubic cases. Both A and MIC generally outperform $dcor$ for periodic functions most notably for the high frequency sine wave. Interestingly, A has the highest power in detecting the circle and 'X' shapes.

Simulation results from the skewed distribution were broadly similar to that shown in Figure 3.2 are given in Appendix A.2 (Figure A.2). For most functional forms there was a decrease in power, most notably for the step function which showed a reduction in power for all association measures. Both A and MIC also had lower power than Pearson correlation in the quadratic and cubic cases.

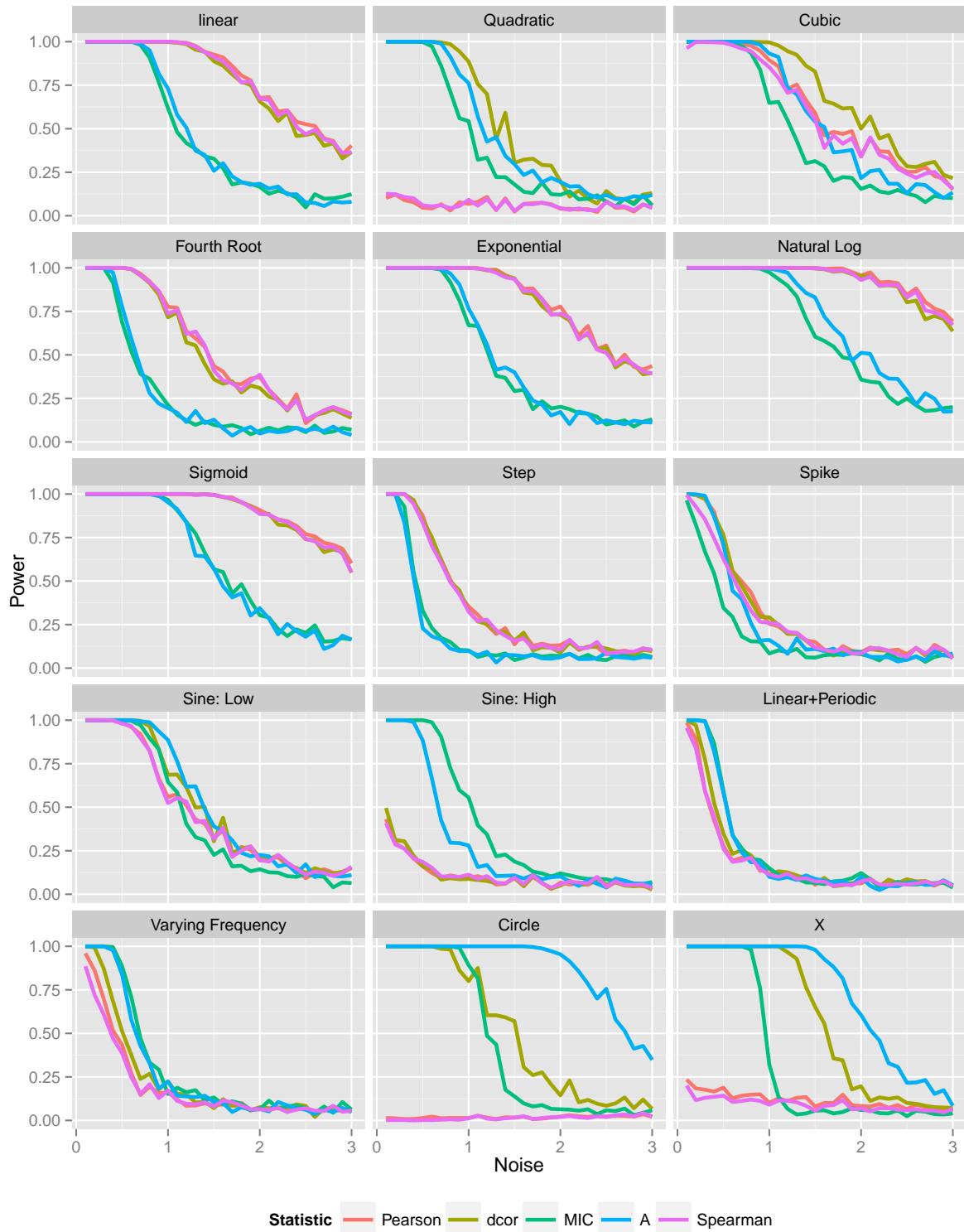


Figure 3.2: Scatter plots showing the power of each measure of association vs. noise for each functional form based on 320 observations from a $Uniform(0, 1)$ distribution.

3.3 Power versus Sample Size

Typically for biological datasets the sample size, n , is relatively small (in the case of the lupus dataset $n = 420$), while $n > 1000$ would generally considered a large dataset. Hence to be of practical use an association measure should be sufficiently powerful at detecting associations on small datasets. This was the objective of this simulation.

An adapted version of algorithm given in the previous section was used to explore how power varies with sample size in the presence of moderate noise:

1. For each functional form, f , and sample size, $n \in \{10, 20, 30, 40, 50, 75, 100, 125, 150, 200, 250, 300, 350, 400, 500, 750, 1000\}$:
2. Generate independent r.v.'s X, Y and Z : X and Z from the given distribution (so $X \in \{0, 1\}, Z \in \{0, 1\}$) and an independent r.v. $Y = f(Z) + l \times \epsilon$, where $l = 1$ and $\epsilon \sim N(0, 1)$
3. Measure the association between X and Y using A , $dcor$, MIC , $\rho_{pearson}^2$ and $\rho_{spearman}^2$
4. Generate dependent r.v.'s X and Y : X from the given distribution and a dependent r.v. $Y = f(X) + l \times \epsilon$, where $l = 1$ and $\epsilon \sim N(0, 1)$.
5. Measure the association between X and Y using A , $dcor$, MIC , $\rho_{pearson}^2$ and $\rho_{spearman}^2$
6. Estimate the power as the number of times the association score for the dependent data exceeds the 95th percentile of the corresponding score for the independent data.
7. Repeat 500 times.

Again the algorithm was run with a $Uniform(0,1)$ distribution and then with a $Beta(2,5)$ distribution to see if a skewed source distribution affected the results. The results for the $Uniform$ case are given in Figure 3.3. The results show that A , MIC and $dcor$ only converged in certain cases but that $dcor$ converged more quickly (i.e. is more powerful on smaller sample sizes) in the linear, quadratic, cubic and exponential cases. The power of $dcor$ was again noticeably low in the high frequency sine case. No measures converged for the step, spike, linear+periodic or varying frequency functions function. A and MIC both failed to converge for $X^{\frac{1}{4}}$.

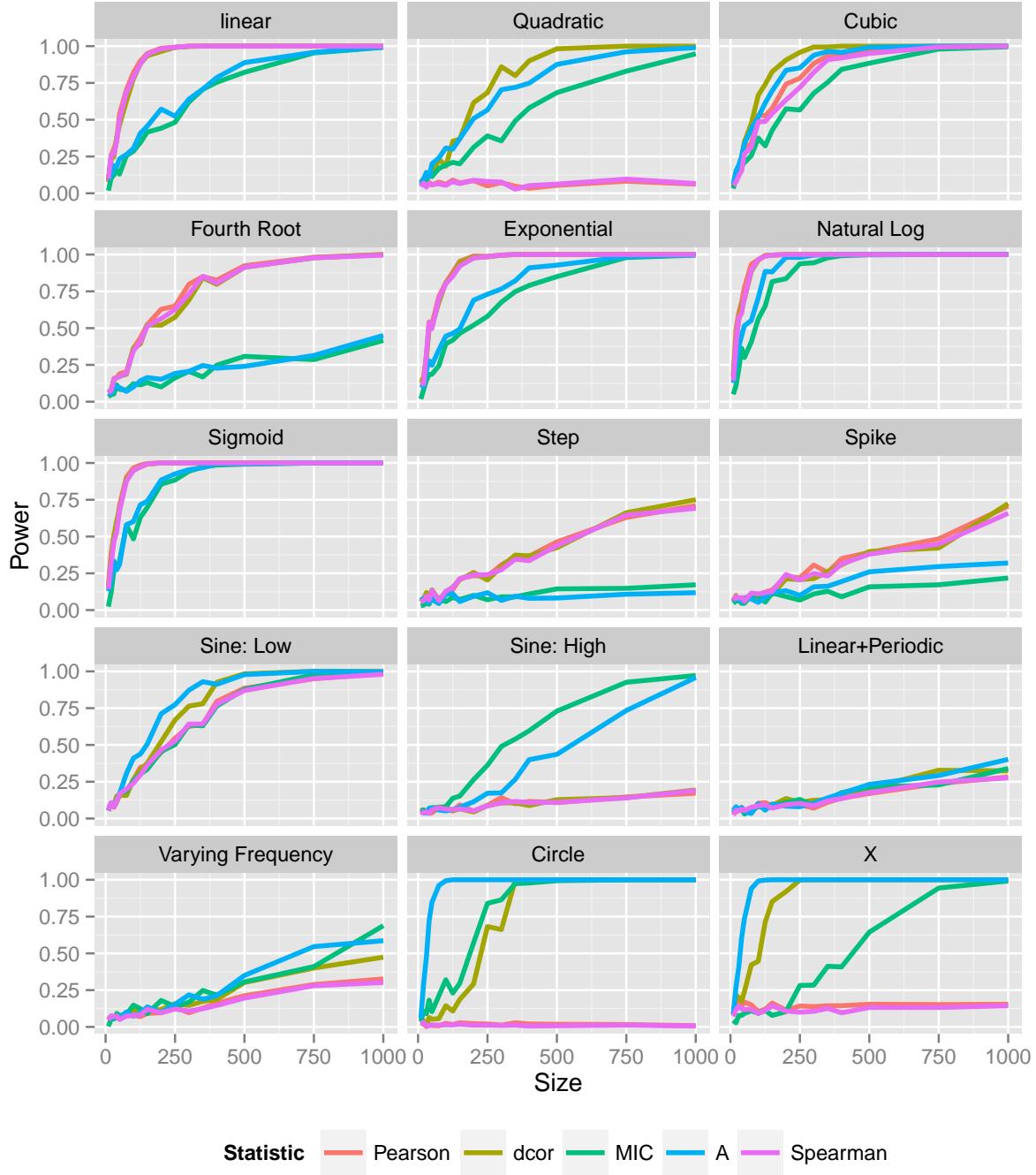


Figure 3.3: Scatter plots showing the power of each measure of association vs. sample size using a $Uniform(0, 1)$ distribution.

In the case of the skewed distribution, the performance of all measures generally declined, particularly for A and MIC in the quadratic and cubic cases. The power of A , MIC and $dcor$ also reduced for the exponential function. The results are shown in Figure A.3 in Appendix A.

Simulations were also run using the maximum level of noise and using a skewed distri-

bution. The results are shown in Figure A.4 in Appendix A.2 and show that none of the association measures were powerful at detecting dependence on very noisy relationships even when the sample size was large ($n = 1000$).

Chapter 4

Discussion and Further Work

The simulation studies focused on how the power of the different measures of association perform in the presence of noise and with different sample sizes. This is of particular relevance to biological datasets which can contain a large amount of technical variability which is not of interest and typically have small sample sizes but many dimensions.

Unsurprisingly the simulations showed that in general power decreases in the presence of increasing noise while power increases with increased sample size even in the presence of noise. Of greater importance is that the analysis suggested that A suffers from some of the same power deficiencies as MIC for certain important cases such as linear and cubic but both surpass $dcor$ where high frequency periodic variation exists. This appears to confirm the reservations regarding the equitability heuristic expressed by Kinney and Atwal⁵ and Simon and Tibshirani⁴. The size simulations questions the practical usefulness of the generality heuristic.

No one measure exceeded all of the others in all of the cases, however on balance the simulation study suggests that $dcor$ should perform better than either MIC or A on the basis of capacity to handle noise and deal with small sample sizes. From a practical perspective if the form of a relationships are unknown then $dcor$ would appear to be the best choice of association measure, possibly in conjunction with Pearson Correlation (to assess non-linearity) and either A or MIC if periodic effects are suspected or known a priori.

There are many other simulation based analyses that could be performed. The noise and size studies both held either size or noise constant, an obvious extension would be to allow both noise and sample size to vary simultaneously to get a more complete power profile. Further extensions would be to expand the range of functional forms and source distributions and to include three way and partial associations where possible. Additionally it would be of interest to explore relative performance of HHG^2 which has been identified as being of potential use for exploratory association mining.

The goal of the analysis was to compare the relative of ability different measures of association to estimate extant non-linear relationships.

Association analyses are essentially exploratory in nature and provide no evidence of causation. There are numerous sources of technical and biological variation which might be responsible for observed associations. In reality the results of an association analysis would serve as the basis for a more refined follow on study or, if the goal is to establish causality, a suitably designed controlled experiment.

Of the measures of association studied, on balance, *dcor* had better performance than either *A* or *MIC* on the lupus data in terms of associations identified. The simulations suggest that *dcor* generally has greater power with small sample sizes and noisy datasets but that *MIC* and *A* have greater power for certain functional types. Knowing the relative strengths and weaknesses of each measure is a critical aspect to choosing which association to use. The choice of which to use might depend on the types of functional relationships suspected to exist in the data or to the extent of the computational resources available. A pragmatic option is to combine a number of measures such as *dcor* and *A* and use the knowledge of the measures' properties to critically evaluate the results obtained.

There are many other measures of association that could be compared, most notably mutual information and the Heller-Heller-Gorfine² measure. Another extension would be to investigate partial associations, i.e. measuring the strength of association while one variable is controlled for. There also a number of alternative methods that can be used to assess inter-variable association and whether or not it can be best described linearly or non-linearly. Two Bayesian approaches to this are an R package called *spikeSlabGAM*¹² and another approach presented by Ni et al.¹³ which would make an interesting comparison in a future study.

A further extension would be to study the sensitivity of both *A* and *dcor* to their respective tuning parameters. The bootstrap estimates of both *MIC* and *A* appeared to converge more closely with *dcor* than their point estimates. This could also be studied through simulation however even if the accuracy improves, the added computational overhead might render it impractical.

Bibliography

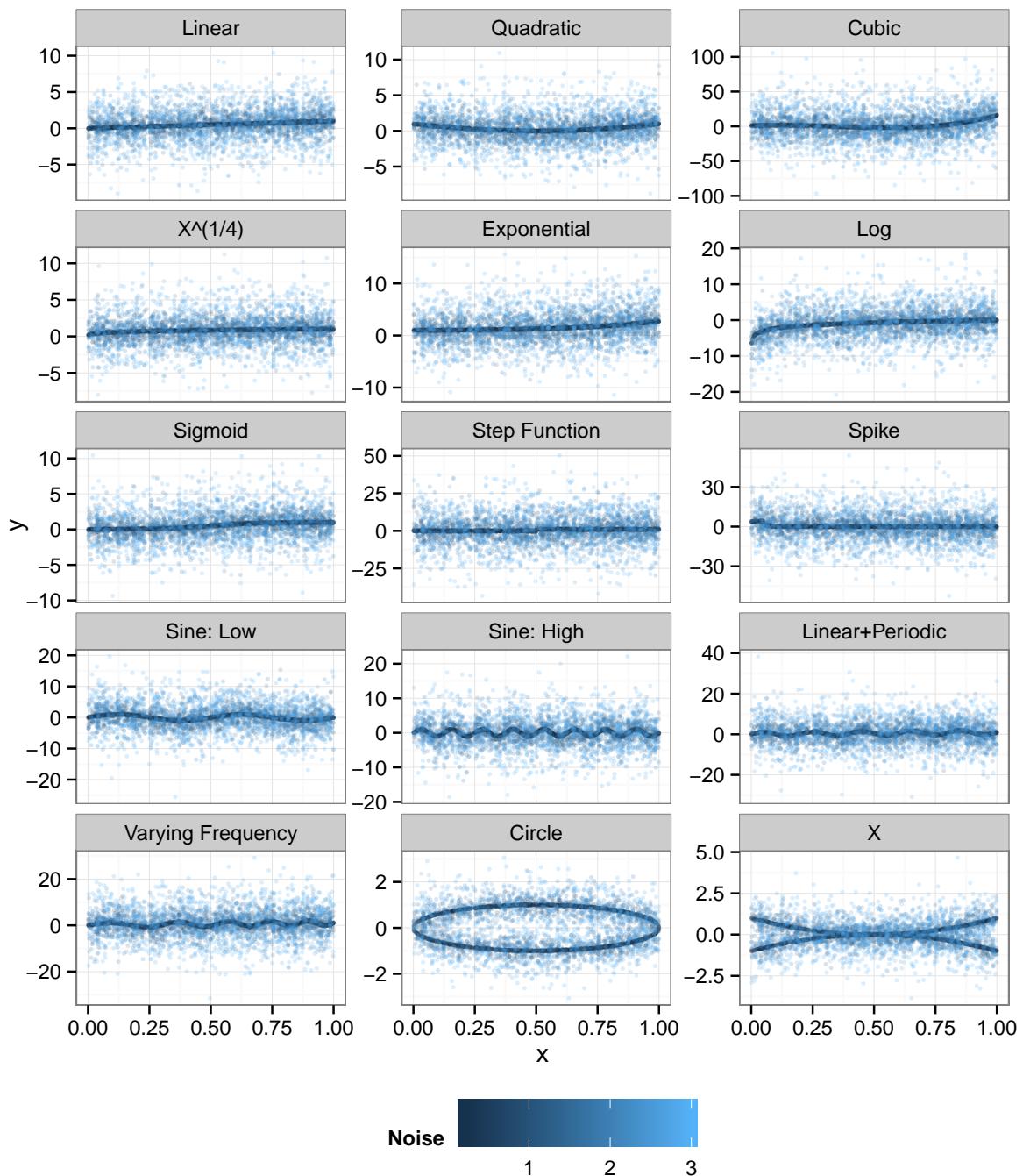
- [1] Finding correlations in big data. *Nature Biotechnology*, 30(4):334–335, April 2012.
- [2] Ruth Heller, Yair Heller, and Malka Gorfine. A consistent multivariate test of association based on ranks of distances. *Biometrika*, 100(2):503–510, 2013. doi: 10.1093/biomet/ass070. URL <http://biomet.oxfordjournals.org/content/100/2/503.abstract>.
- [3] D. Reshef, Y. Reshef, and H. Finucane. Detecting novel associations in large data sets. *Science*, 334(6062):1518–1524, December 2011.
- [4] N Simon and R Tibshirani. Comment on detecting novel associations in large data sets by reshef et al, December 2011. URL <http://statweb.stanford.edu/~tibs/reshef/comment.pdf>.
- [5] Justin B. Kinney and Gurinder S. Atwal. Reply to reshef et al.: Falsifiability or bust. *Proceedings of the National Academy of Sciences*, 111(33):E3364, 2014. doi: 10.1073/pnas.1410317111. URL <http://www.pnas.org/content/111/33/E3364.short>.
- [6] B. Murrel, D. Murrel, and H. Murrel. Discovering general multidimensional associations, March 2013. URL <http://arxiv.org/pdf/1303.1828v1.pdf>.
- [7] Irwin and Marylees Miller. *John E. Freund’s mathematical statistics with applications*. Pearson Prentice Hall, Upper Saddle River, NJ, 7th edition, 2004.
- [8] Roberto Visintainer [aut] Davide Albanese [aut] Samantha Riccadonna [ctb] Giuseppe Jurman [ctb] Cesare Furlanello [ctb] Michele Filosi [aut, cre]. *minerva: Maximal Information-Based Nonparametric Exploration R package for Variable Analysis*, 2014. URL <http://CRAN.R-project.org/package=minerva>.
- [9] Ben Murrell, Dan Murrell, and Hugh Murrell. *Measuring Association and Testing Independence Efficiently*, 2015. URL <http://CRAN.R-project.org/package=matie>.
- [10] G. Szekely and M. L Rizzo. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272, August 2013.

- [11] Maria L. Rizzo and Gabor J. Szekely. *energy: E-statistics (energy statistics)*, 2014. URL <http://CRAN.R-project.org/package=energy>.
- [12] Fabian Scheipl. spikeSlabGAM: Bayesian variable selection, model choice and regularization for generalized additive mixed models in R. *Journal of Statistical Software*, 43(14):1–24, 2011. URL <http://www.jstatsoft.org/v43/i14/>.
- [13] Yang Ni, Francesco C. Stingo, and Veerabhadran Baladandayuthapani. Bayesian nonlinear model selection for gene regulatory networks. *Biometrics*, pages n/a–n/a, 2015. ISSN 1541-0420. doi: 10.1111/biom.12309. URL <http://dx.doi.org/10.1111/biom.12309>.

Appendix A

Additional Results

A.1 Simulation Results



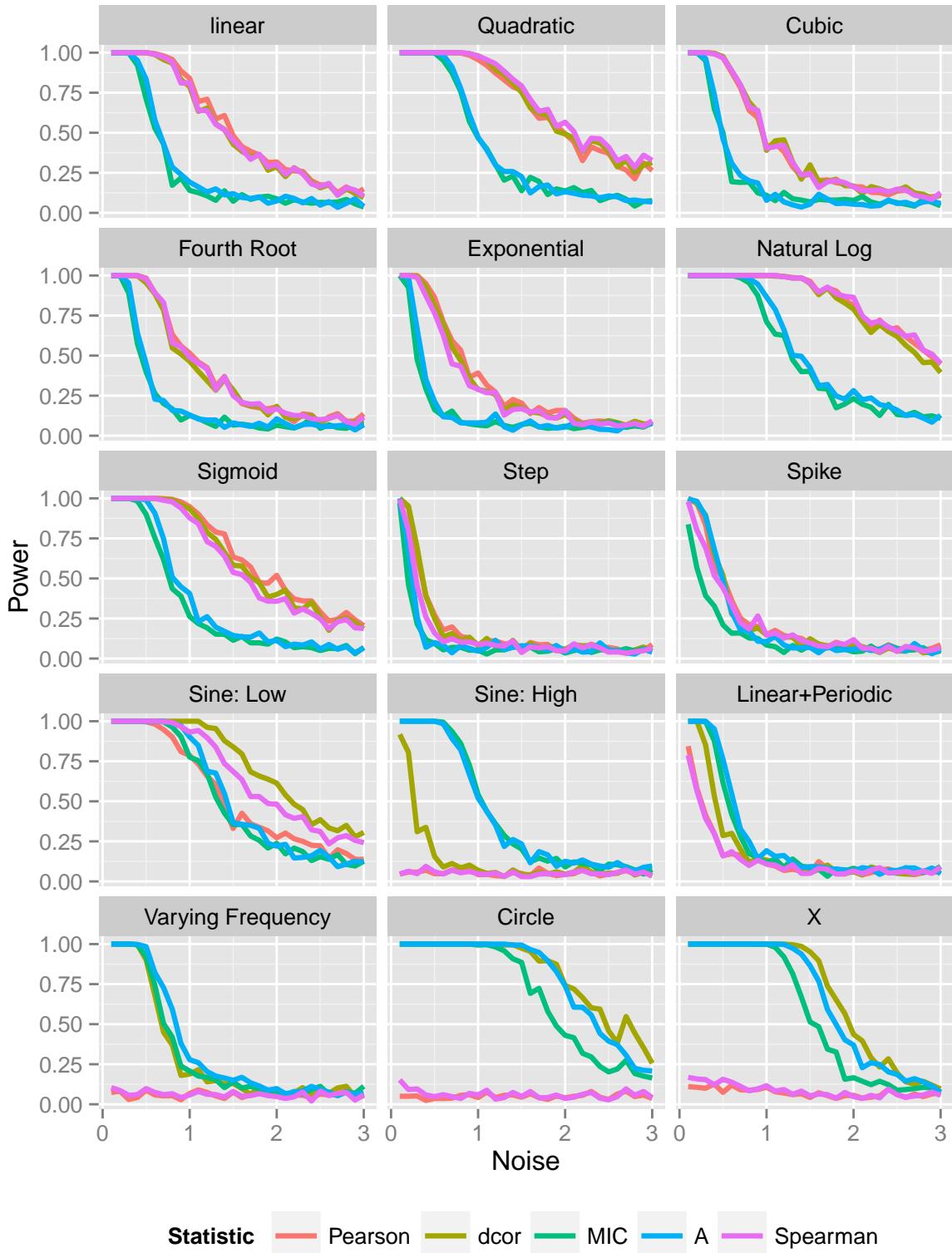


Figure A.2: Scatter plots showing the power of each measure of association vs. noise for each functional form based on 320 observations from a $Beta(2, 5)$ distribution. Note the reduction in power for the step function.

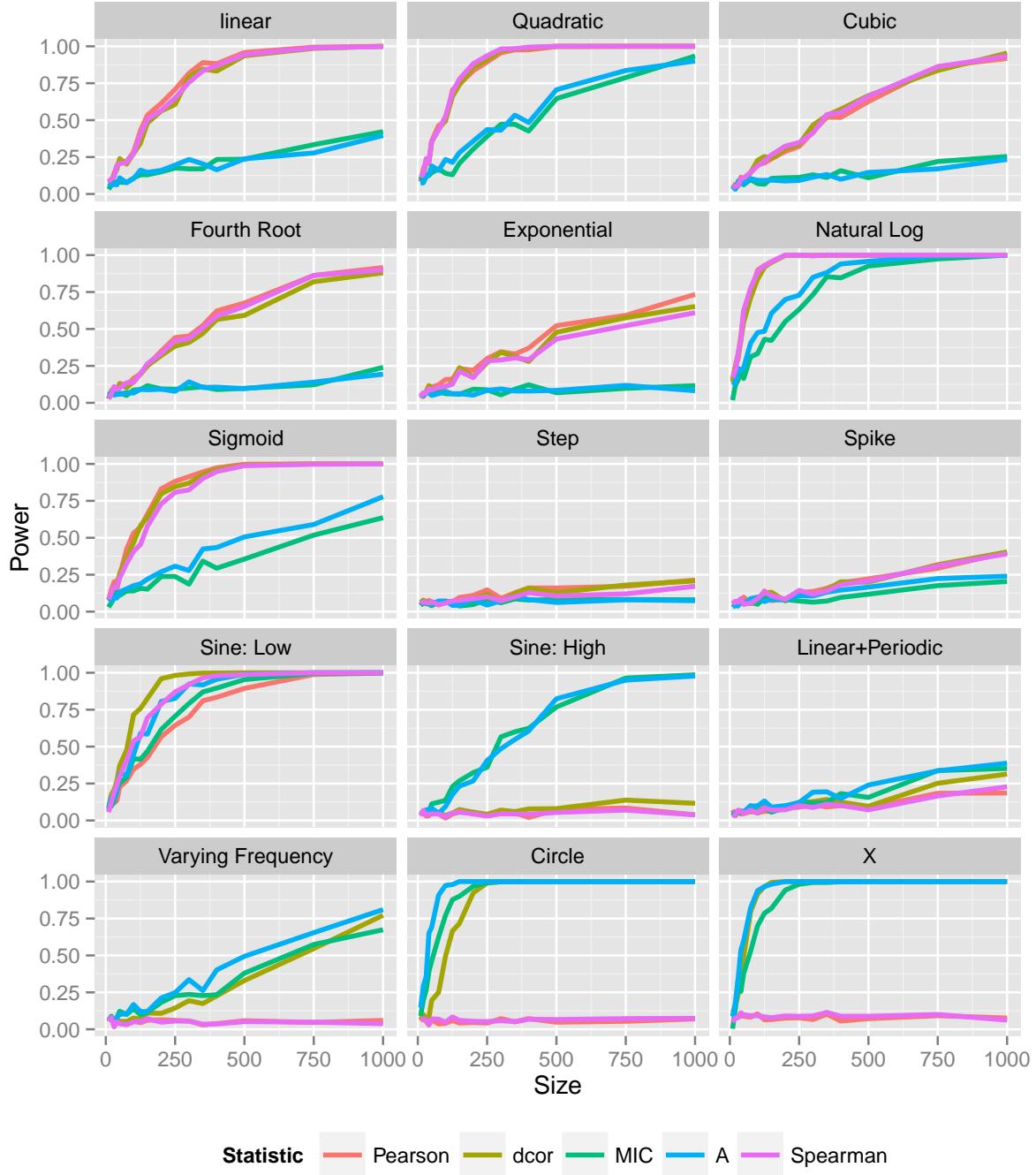


Figure A.3: Scatter plots showing the power of each measure of association vs. size for each functional form based on a $Beta(2, 5)$ distribution. For the exponential function the power reduced for *dcor*, *A* and *MIC*.

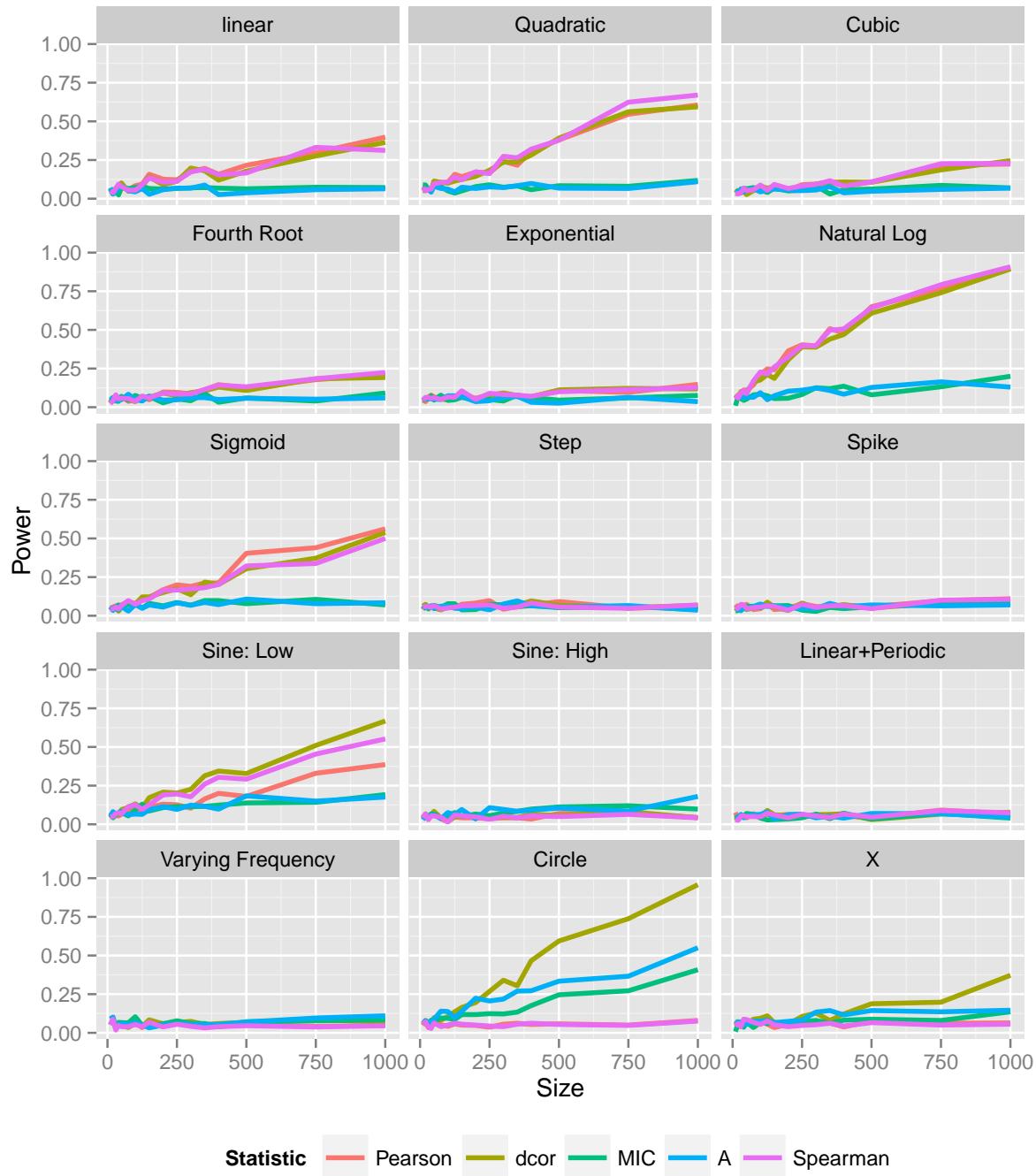


Figure A.4: Scatter plots showing the power of each measure of association vs. sample size for each functional forms. None of the measures converged.