Dependence Discovery from Multimodal Data via Multiscale Graph Correlation

Cencheng Shen

Joint Work with Joshua T. Vogelstein & Mauro Maggioni & Carey E. Priebe

August, 2016

Overview

1. Motivations
2. Results
   - Global Correlation Coefficient
   - Local Correlation Coefficients
   - Multiscale Graph Correlation
   - Theoretical Advantages
   - Numerical Experiments
3. Conclusion

Section 1

Motivations

Motivations

- Given multiple data sets, we would like to test whether they are independent or not.

Motivations

- Given multiple data sets, we would like to test whether they are independent or not.
- Modern data sets may be high-dimensional, nonlinear, noisy, of small sample size; and different data sets may come from disparate spaces.

Motivations

- Given multiple data sets, we would like to test whether they are independent or not.
- Modern data sets may be high-dimensional, nonlinear, noisy, of small sample size; and different data sets may come from disparate spaces.
- In particular, we are investigating the association between brain activities and various phenotypes, such as brain Connectome vs certain disease, or brain vs personality, where the brain data is usually obtained via fMRI scans for a number of brain regions at consecutive time steps.

Motivations

- Given multiple data sets, we would like to test whether they are independent or not.
- Modern data sets may be high-dimensional, nonlinear, noisy, of small sample size; and different data sets may come from disparate spaces.
- In particular, we are investigating the association between brain activities and various phenotypes, such as brain Connectome vs certain disease, or brain vs personality, where the brain data is usually obtained via fMRI scans for a number of brain regions at consecutive time steps.
- Before we try all kinds of regression / classification methods, the first task is to determine whether there exists strong dependency for given pair of data.

- **How to reliably test independence on given data?**

- **How to reliably test independence on given data?**
- We would like a test statistic that:

- **How to reliably test independence on given data?**
- We would like a test statistic that:
- 1) is consistent as the sample size increases to infinity for all dependencies.

- **How to reliably test independence on given data?**
- We would like a test statistic that:
- 1) is consistent as the sample size increases to infinity for all dependencies.
- 2) has a good finite-sample testing power, for given data of low or high dimensionality, linear or non-linearity, small sample size, with noise, etc.

- **How to reliably test independence on given data?**
- We would like a test statistic that:
- 1) is consistent as the sample size increases to infinity for all dependencies.
- 2) has a good finite-sample testing power, for given data of low or high dimensionality, linear or non-linearity, small sample size, with noise, etc.
- 3) provides insight into the dependency structure

- **How to reliably test independence on given data?**
- We would like a test statistic that:
- 1) is consistent as the sample size increases to infinity for all dependencies.
- 2) has a good finite-sample testing power, for given data of low or high dimensionality, linear or non-linearity, small sample size, with noise, etc.
- 3) provides insight into the dependency structure
- 4) does not inflate the false positive rate in the absence of dependency.

- **How to reliably test independence on given data?**
- We would like a test statistic that:
- 1) is consistent as the sample size increases to infinity for all dependencies.
- 2) has a good finite-sample testing power, for given data of low or high dimensionality, linear or non-linearity, small sample size, with noise, etc.
- 3) provides insight into the dependency structure
- 4) does not inflate the false positive rate in the absence of dependency.
- 5) easy to use and scalable on big data

- **How to reliably test independence on given data?**
- We would like a test statistic that:
- 1) is consistent as the sample size increases to infinity for all dependencies.
- 2) has a good finite-sample testing power, for given data of low or high dimensionality, linear or non-linearity, small sample size, with noise, etc.
- 3) provides insight into the dependency structure
- 4) does not inflate the false positive rate in the absence of dependency.
- 5) easy to use and scalable on big data
- To that end, we propose multiscale graph correlation in Shen et al. [2016] for testing independence, which combines distance correlation and nearest-neighbor into the same testing framework.

Section 2

Results

General Correlation Coefficient

A general correlation coefficient $\mathcal{G}$ can be expressed as follows:

$$\mathcal{G} = \frac{\sum_{i,j=1}^{n}(a_{ij} - \bar{a})(b_{ij} - \bar{b})}{\sqrt{\sum_{i,j=1}^{n}(a_{ij} - \bar{a})^2 \sum_{i,j=1}^{n}(b_{ij} - \bar{b})^2}}, \tag{1}$$

where $\bar{a}$ and $\bar{b}$ denote the sample means of a_{ij} and b_{ij}.

Examples

- Suppose $X = [x_1, \cdots, x_n] \in \mathbb{R}^{d_x \times n}$ and $Y = [y_1, \cdots, y_n] \in \mathbb{R}^{d_y \times n}$ are the corresponding data sets.

Examples

- Suppose $X = [x_1, \cdots, x_n] \in \mathbb{R}^{d_x \times n}$ and $Y = [y_1, \cdots, y_n] \in \mathbb{R}^{d_y \times n}$ are the corresponding data sets.
- The Pearson's correlation coefficient can be calculated by taking $a_{ij} = x_i$ and $b_{ij} = y_i$.

Examples

- Suppose $X = [x_1, \cdots, x_n] \in \mathbb{R}^{d_x \times n}$ and $Y = [y_1, \cdots, y_n] \in \mathbb{R}^{d_y \times n}$ are the corresponding data sets.
- The Pearson's correlation coefficient can be calculated by taking $a_{ij} = x_i$ and $b_{ij} = y_i$.
- Spearman and Kendall's rank correlations set a_{ij} to be $rank(x_i) - rank(x_j)$ and $sign(x_i - x_j)$.

Examples

- Suppose $X = [x_1, \cdots, x_n] \in \mathbb{R}^{d_x \times n}$ and $Y = [y_1, \cdots, y_n] \in \mathbb{R}^{d_y \times n}$ are the corresponding data sets.
- The Pearson's correlation coefficient can be calculated by taking $a_{ij} = x_i$ and $b_{ij} = y_i$.
- Spearman and Kendall's rank correlations set a_{ij} to be $rank(x_i) - rank(x_j)$ and $sign(x_i - x_j)$.
- The Mantel coefficient Mantel [1967] considers distance matrices and takes $a_{ij} = |x_i - x_j|_2$.

Examples

- Suppose $X = [x_1, \cdots, x_n] \in \mathbb{R}^{d_x \times n}$ and $Y = [y_1, \cdots, y_n] \in \mathbb{R}^{d_y \times n}$ are the corresponding data sets.
- The Pearson's correlation coefficient can be calculated by taking $a_{ij} = x_i$ and $b_{ij} = y_i$.
- Spearman and Kendall's rank correlations set a_{ij} to be $rank(x_i) - rank(x_j)$ and $sign(x_i - x_j)$.
- The Mantel coefficient Mantel [1967] considers distance matrices and takes $a_{ij} = |x_i - x_j|_2$.
- The distance correlation (dcorr) proposed in Szekely et al. [2007] uses the doubly-centered distance entries for a_{ij} and b_{ij}.

Examples

- Suppose $X = [x_1, \cdots, x_n] \in \mathbb{R}^{d_x \times n}$ and $Y = [y_1, \cdots, y_n] \in \mathbb{R}^{d_y \times n}$ are the corresponding data sets.
- The Pearson's correlation coefficient can be calculated by taking $a_{ij} = x_i$ and $b_{ij} = y_i$.
- Spearman and Kendall's rank correlations set a_{ij} to be $rank(x_i) - rank(x_j)$ and $sign(x_i - x_j)$.
- The Mantel coefficient Mantel [1967] considers distance matrices and takes $a_{ij} = |x_i - x_j|_2$.
- The distance correlation (dcorr) proposed in Szekely et al. [2007] uses the doubly-centered distance entries for a_{ij} and b_{ij}.
- The modified distance correlation (mcorr) proposed in Szekely and Rizzo [2013] slightly modifies a_{ij}/b_{ij} of dcorr to make the resulting $\mathcal{G}$ unbiased.

Local Correlation

- We combine nearest-neighbor into global correlation, by defining a family of local correlation coefficients.

Local Correlation

- We combine nearest-neighbor into global correlation, by defining a family of local correlation coefficients.
- Let $rank(a_{ij})$ be the "rank" of x_i relative to x_j, that is, $rank(a_{ij}) = k$ if x_i is the k^{th} closest point (or "neighbor") to x_j; and the ranks start from $k = 1$ onwards (i.e., $rank(a_{jj}) = 1$), and the minimal rank is used for ties.

Local Correlation

- We combine nearest-neighbor into global correlation, by defining a family of local correlation coefficients.
- Let $rank(a_{ij})$ be the "rank" of x_i relative to x_j, that is, $rank(a_{ij}) = k$ if x_i is the k^{th} closest point (or "neighbor") to x_j; and the ranks start from $k = 1$ onwards (i.e., $rank(a_{jj}) = 1$), and the minimal rank is used for ties.
- Then for any general correlation coefficient $\mathcal{G}$ in the form of Equation 1, we define its *local* variants by

$$\mathcal{G}_{kl} = \frac{\sum_{i,j=1}^{n}(a_{ij}^{k} - \bar{a}^{k})(b_{ij}^{l} - \bar{b}^{l})}{\sqrt{\sum_{i,j=1}^{n}(a_{ij}^{k} - \bar{a}^{k})^2 \sum_{i,j=1}^{n}(b_{ij}^{l} - \bar{b}^{l})^2}}, \tag{2}$$

for $k = 1, \ldots, \max(rank(a_{ij}))$, $l = 1, \ldots, \max(rank(b_{ij}))$, where

$$a_{ij}^{k} = \begin{cases} a_{ij}, & \text{if } 0 < rank(a_{ij}) \leq k, \\ 0, & \text{otherwise;} \end{cases} \qquad b_{ij}^{l} = \begin{cases} b_{ij}, & \text{if } 0 < rank(b_{ij}) \\ 0, & \text{otherwise;} \end{cases} \tag{3}$$

MGC

- In the family of local correlations $\{\mathcal{G}_{kl}\}$, there always exists an optimal local correlation with respect to the independence testing power.

MGC

- In the family of local correlations $\{\mathcal{G}_{kl}\}$, there always exists an optimal local correlation with respect to the independence testing power.
- The optimal scale exists, is distribution dependent, and may not be unique.

MGC

- In the family of local correlations $\{\mathcal{G}_{kl}\}$, there always exists an optimal local correlation with respect to the independence testing power.
- The optimal scale exists, is distribution dependent, and may not be unique.
- We dub the optimal local correlation coefficient as the multiscale graph correlation, and denote it as $\mathcal{G}^*$.

MGC

- In the family of local correlations $\{\mathcal{G}_{kl}\}$, there always exists an optimal local correlation with respect to the independence testing power.
- The optimal scale exists, is distribution dependent, and may not be unique.
- We dub the optimal local correlation coefficient as the multiscale graph correlation, and denote it as $\mathcal{G}^*$.
- Although each local correlation requires $O(n^2)$ to compute, we provide a fast algorithm to compute all local correlations simultaneously in $O(n^2)$, assuming the rank information is given (note: sorting the distance matrix column-wise takes $O(n^2 \log(n))$).

MGC

- In the family of local correlations $\{\mathcal{G}_{kl}\}$, there always exists an optimal local correlation with respect to the independence testing power.
- The optimal scale exists, is distribution dependent, and may not be unique.
- We dub the optimal local correlation coefficient as the multiscale graph correlation, and denote it as $\mathcal{G}^*$.
- Although each local correlation requires $O(n^2)$ to compute, we provide a fast algorithm to compute all local correlations simultaneously in $O(n^2)$, assuming the rank information is given (note: sorting the distance matrix column-wise takes $O(n^2 \log(n))$).
- This allows the optimal scale to be efficiently determined for MGC, unlike many other applications by nearest-neighbor (say knn classification, manifold learning, etc.).

The Testing Framework

- Given two data sets $X = [x_1, \cdots, x_n] \in \mathcal{R}^{d_X \times n}$ and $Y = [y_1, \cdots, y_n] \in \mathcal{R}^{d_Y \times n}$.

The Testing Framework

- Given two data sets $X = [x_1, \cdots, x_n] \in \mathcal{R}^{d_X \times n}$ and $Y = [y_1, \cdots, y_n] \in \mathcal{R}^{d_Y \times n}$.
- Assume that $x_i, i = 1, \ldots, n$ are identically independently distributed (i.i.d.) as $\boldsymbol{x} \sim f_x$; similarly each y_i are realizations of $\boldsymbol{y} \sim f_y$.

The Testing Framework

- Given two data sets $X = [x_1, \cdots, x_n] \in \mathcal{R}^{d_X \times n}$ and $Y = [y_1, \cdots, y_n] \in \mathcal{R}^{d_Y \times n}$.
- Assume that $x_i, i = 1, \ldots, n$ are identically independently distributed (i.i.d.) as $\boldsymbol{x} \sim f_x$; similarly each y_i are realizations of $\boldsymbol{y} \sim f_y$.
- The null and the alternative hypothesis for testing independence are

$$H_0 : f_{xy} = f_x f_y,$$
$$H_A : f_{xy} \neq f_x f_y,$$

where f_{xy} denotes the joint distribution of $(\boldsymbol{x}, \boldsymbol{y})$.

The Testing Framework

- Given two data sets $X = [x_1, \cdots, x_n] \in \mathcal{R}^{d_X \times n}$ and $Y = [y_1, \cdots, y_n] \in \mathcal{R}^{d_Y \times n}$.
- Assume that $x_i, i = 1, \ldots, n$ are identically independently distributed (i.i.d.) as $\boldsymbol{x} \sim f_x$; similarly each y_i are realizations of $\boldsymbol{y} \sim f_y$.
- The null and the alternative hypothesis for testing independence are

$$
\begin{aligned}
H_0 &: f_{xy} = f_x f_y, \\
H_A &: f_{xy} \neq f_x f_y,
\end{aligned}
$$

  where f_{xy} denotes the joint distribution of $(\boldsymbol{x}, \boldsymbol{y})$.
- For a given pair of data with unknown model, we can use the permutation test, and reject the null when the p-value is sufficiently small.

Theorems of MGC

Theorem 1

Multiscale graph correlation is consistent against all dependent alternatives of finite second moments, i.e., $\beta_\alpha(\mathcal{G}^*) \to 1$ *as* $n \to \infty$ *at any type 1 error level* α*, when distance correlation or modified distance correlation is used as the global correlation* $\mathcal{G}$*.*

Theorems of MGC

Theorem 1

Multiscale graph correlation is consistent against all dependent alternatives of finite second moments, i.e., $\beta_\alpha(\mathcal{G}^*) \rightarrow 1$ as $n \rightarrow \infty$ at any type 1 error level α, when distance correlation or modified distance correlation is used as the global correlation $\mathcal{G}$.

Theorem 2

*Suppose **x** is linearly dependent of **y**. Then for any n and α it always holds that*

$$\beta_\alpha(\mathcal{G}^*) = \beta_\alpha(\mathcal{G}). \tag{4}$$

Thus multiscale graph correlation is equivalent to the global correlation coefficient under linear dependency.

Theorems of MGC

Theorem 3

There exists f_{xy}, n and α such that

$$\beta_{\alpha}(\mathcal{G}^{*}) > \beta_{\alpha}(\mathcal{G}). \tag{5}$$

Thus multiscale graph correlation can be better than its global correlation coefficient under certain nonlinear dependency.

We use a quadratic relationship and finite n to prove theorem 3.

Simulation Set-Up

- In total 20 different distributions of f_{xy} are considered, most of which are based on simulations from existing literature.

Simulation Set-Up

- In total 20 different distributions of f_{xy} are considered, most of which are based on simulations from existing literature.
- They consist of various linear and close to linear dependencies (e.g., exponential, joint normal), polynomial-based nonlinear relationships (e.g., quadratic, fourth-root), trigonometry-based nonlinear dependencies (e.g., circle, spiral), two uncorrelated but dependent dependencies, and an independent relationship.

Simulation Set-Up

- In total 20 different distributions of f_{xy} are considered, most of which are based on simulations from existing literature.
- They consist of various linear and close to linear dependencies (e.g., exponential, joint normal), polynomial-based nonlinear relationships (e.g., quadratic, fourth-root), trigonometry-based nonlinear dependencies (e.g., circle, spiral), two uncorrelated but dependent dependencies, and an independent relationship.
- For each distribution, we further consider two different scenarios: a 1-dimensional scenario with increasing sample size, and a high-dimensional scenario with fixed sample size but increasing dimensions.

Simulation Set-Up

- In total 20 different distributions of f_{xy} are considered, most of which are based on simulations from existing literature.
- They consist of various linear and close to linear dependencies (e.g., exponential, joint normal), polynomial-based nonlinear relationships (e.g., quadratic, fourth-root), trigonometry-based nonlinear dependencies (e.g., circle, spiral), two uncorrelated but dependent dependencies, and an independent relationship.
- For each distribution, we further consider two different scenarios: a 1-dimensional scenario with increasing sample size, and a high-dimensional scenario with fixed sample size but increasing dimensions.
- The benchmarks are distance correlation, modified distance correlation, the Mantel test, and the HHG method proposed in Heller et al. [2013].

Simulation Set-Up

- In total 20 different distributions of f_{xy} are considered, most of which are based on simulations from existing literature.
- They consist of various linear and close to linear dependencies (e.g., exponential, joint normal), polynomial-based nonlinear relationships (e.g., quadratic, fourth-root), trigonometry-based nonlinear dependencies (e.g., circle, spiral), two uncorrelated but dependent dependencies, and an independent relationship.
- For each distribution, we further consider two different scenarios: a 1-dimensional scenario with increasing sample size, and a high-dimensional scenario with fixed sample size but increasing dimensions.
- The benchmarks are distance correlation, modified distance correlation, the Mantel test, and the HHG method proposed in Heller et al. [2013].
- To better illustrate the effectiveness of distance-based local correlation, we consider three MGC implementations by dcorr / mcorr / Mantel respectively in the simulation.

Visualization of Sample Data for Each Joint Distribution

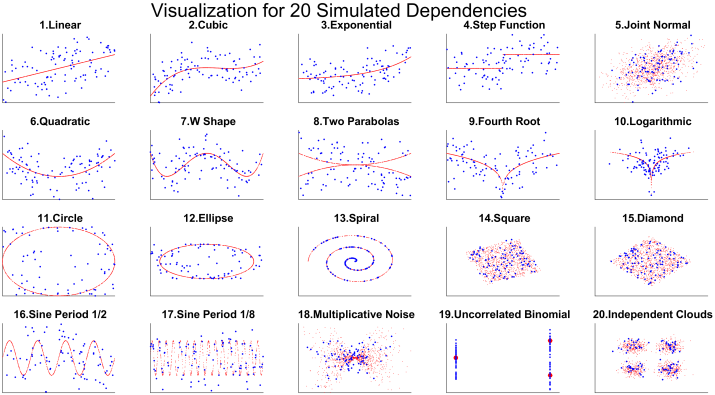


Figure: Visualization of the 20 dependencies for 1-dimensional simulations. The blue points are generated with noise (c=1) for $n = 100$ to show the actual sample data in testing, and the red points are generated with no noise for $n = 1000$ to highlight each underlying dependency.

Simulation Set-Up

- To calculate the testing power, we repeatedly generate dependent data and independent data from the given model, estimate the test statistic under the null and the alternative, and compute the critical value and power accordingly.

Simulation Set-Up

- To calculate the testing power, we repeatedly generate dependent data and independent data from the given model, estimate the test statistic under the null and the alternative, and compute the critical value and power accordingly.
- We use 10,000 Monte-Carlo replicates to generate the data for testing power computation; and 2,000 additional replicates to generate data for optimal scale estimation.

Simulation Set-Up

- To calculate the testing power, we repeatedly generate dependent data and independent data from the given model, estimate the test statistic under the null and the alternative, and compute the critical value and power accordingly.
- We use 10,000 Monte-Carlo replicates to generate the data for testing power computation; and 2,000 additional replicates to generate data for optimal scale estimation.
- For 1-dimensional simulations, each panel shows empirical testing power on the absicca, and sample size on the ordinate, with dimension fixed at 1; for high-dimensional simulations, the dimension choice is on the ordinate, with the sample size fixed at 100.

Simulation Interprations

- We will see that among the benchmarks, dcorr / mcorr can do fairly well in linear problems with mcorr being better in hd linear problems, while HHG works the best in nonlinear problems.

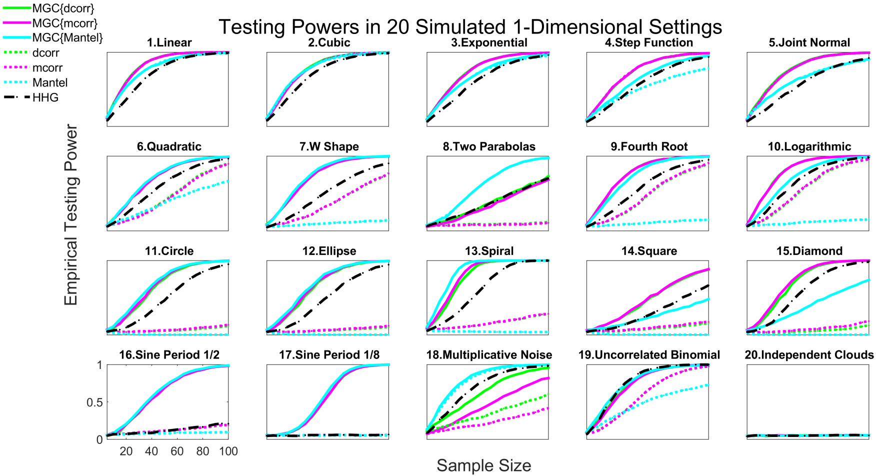
Simulation Interprations

- We will see that among the benchmarks, dcorr / mcorr can do fairly well in linear problems with mcorr being better in hd linear problems, while HHG works the best in nonlinear problems.
- Yet MGC achieves the same power as the global correlation under close to linear dependencies, and is equivalent or better than HHG under nonlinear dependencies. This makes MGC the best method throughout all sample sizes / dimensions / simulations (note: different MGC implementations vary slightly in performance, depending on the property of the global correlation).

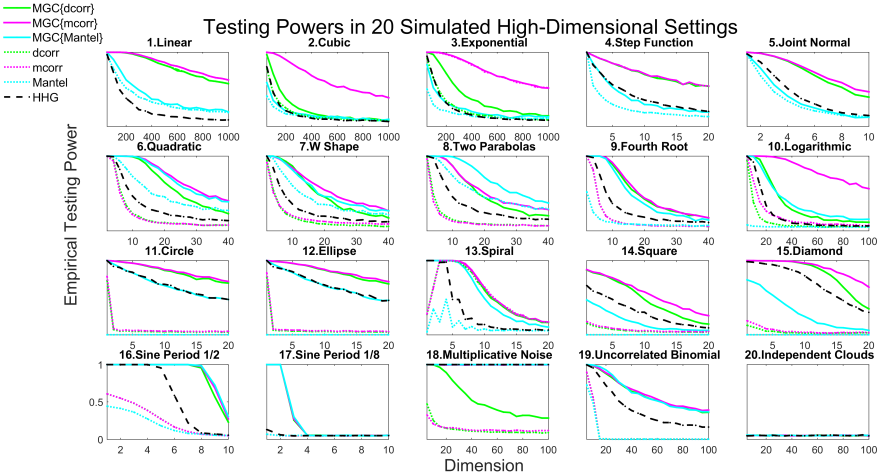
Simulation Interprations

- We will see that among the benchmarks, dcorr / mcorr can do fairly well in linear problems with mcorr being better in hd linear problems, while HHG works the best in nonlinear problems.
- Yet MGC achieves the same power as the global correlation under close to linear dependencies, and is equivalent or better than HHG under nonlinear dependencies. This makes MGC the best method throughout all sample sizes / dimensions / simulations (note: different MGC implementations vary slightly in performance, depending on the property of the global correlation).
- We also show the MGC power heatmap with respect to neighborhood choices, where we can observe that: linear dependency always favors the largest scale (i.e., $\mathcal{G}^* = \mathcal{G}$), while nonlinear dependency always favors a smaller scale such that ($\mathcal{G}^* > \mathcal{G}$); and similar dependency structure usualy yield similar optimal scales.

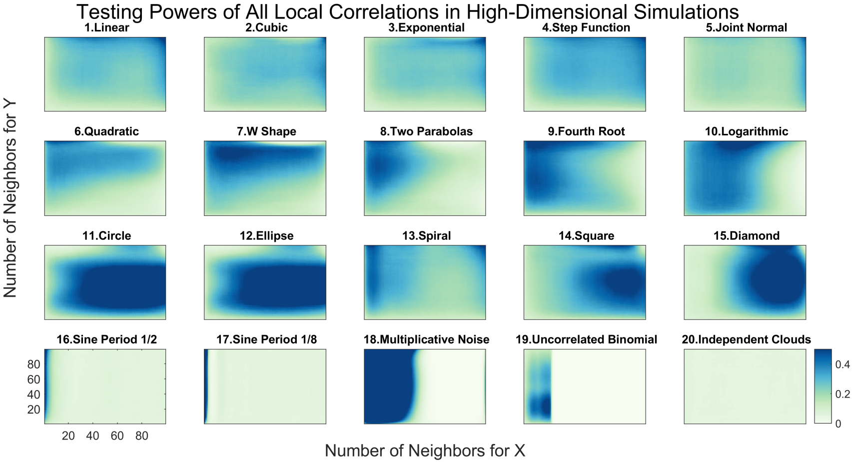
1-Dimensional Testing Powers



High-Dimensional Testing Powers



High-Dimensional Local Mcorr Power Heatmap



Section 3

Conclusion

Advantages of MGC

- 1) Easy to implement for any global correlation / any data with a suitable distance metric, and comparable in running time to existing methods.

Advantages of MGC

- 1) Easy to implement for any global correlation / any data with a suitable distance metric, and comparable in running time to existing methods.
- 2) When MGC is based on dcorr/mcorr, MGC is consistent for testing independence.

Advantages of MGC

- 1) Easy to implement for any global correlation / any data with a suitable distance metric, and comparable in running time to existing methods.
- 2) When MGC is based on dcorr/mcorr, MGC is consistent for testing independence.
- 3) MGC is equivalent to the respective global correlation under linear dependence, but can perform better under nonlinear dependence.

Advantages of MGC

- 1) Easy to implement for any global correlation / any data with a suitable distance metric, and comparable in running time to existing methods.
- 2) When MGC is based on dcorr/mcorr, MGC is consistent for testing independence.
- 3) MGC is equivalent to the respective global correlation under linear dependence, but can perform better under nonlinear dependence.
- 4) Superior numerical performance in a comprehensive simulation setting, including various linear/nonlinear/high-dimensional/noisy dependency.

Advantages of MGC

- 1) Easy to implement for any global correlation / any data with a suitable distance metric, and comparable in running time to existing methods.
- 2) When MGC is based on dcorr/mcorr, MGC is consistent for testing independence.
- 3) MGC is equivalent to the respective global correlation under linear dependence, but can perform better under nonlinear dependence.
- 4) Superior numerical performance in a comprehensive simulation setting, including various linear/nonlinear/high-dimensional/noisy dependency.
- 5) Provides information on the local scale where the dependency is the strongest, and implies the linearity/nonlinearity of the underlying dependency.

Advantages of MGC

- 1) Easy to implement for any global correlation / any data with a suitable distance metric, and comparable in running time to existing methods.
- 2) When MGC is based on dcorr/mcorr, MGC is consistent for testing independence.
- 3) MGC is equivalent to the respective global correlation under linear dependence, but can perform better under nonlinear dependence.
- 4) Superior numerical performance in a comprehensive simulation setting, including various linear/nonlinear/high-dimensional/noisy dependency.
- 5) Provides information on the local scale where the dependency is the strongest, and implies the linearity/nonlinearity of the underlying dependency.
- 6) Does not inflate the false positive rate.

Advantages of MGC

- 1) Easy to implement for any global correlation / any data with a suitable distance metric, and comparable in running time to existing methods.
- 2) When MGC is based on dcorr/mcorr, MGC is consistent for testing independence.
- 3) MGC is equivalent to the respective global correlation under linear dependence, but can perform better under nonlinear dependence.
- 4) Superior numerical performance in a comprehensive simulation setting, including various linear/nonlinear/high-dimensional/noisy dependency.
- 5) Provides information on the local scale where the dependency is the strongest, and implies the linearity/nonlinearity of the underlying dependency.
- 6) Does not inflate the false positive rate.
- 7) MGC is a better scalable method, for big data where distance-based testing method is often performed on sub-samples.

What We Do Not Show In Slides

- MGC is robust against outliers;

What We Do Not Show In Slides

- MGC is robust against outliers;
- In case of one pair of given data of unknown model, how to choose the optimal scale heuristically;

What We Do Not Show In Slides

- MGC is robust against outliers;
- In case of one pair of given data of unknown model, how to choose the optimal scale heuristically;
- Real data experiments where MGC is used to detect local signals between brain data vs phenotypes.

In The Future

- Testing dependence on networks.

In The Future

- Testing dependence on networks.
- Better optimal scale estimation for unknown model.

In The Future

- Testing dependence on networks.
- Better optimal scale estimation for unknown model.
- Investigate alternative form of local correlations.

In The Future

- Testing dependence on networks.
- Better optimal scale estimation for unknown model.
- Investigate alternative form of local correlations.
- Local correlation is potentially useful for nonlinear embedding, variable selection, etc.

Section 4

References

R. Heller, Y. Heller, and M. Gorfine. A consistent multivariate test of association based on ranks of distances. *Biometrika*, 100(2):503–510, 2013.

N. Mantel. The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2):209–220, 1967.

C. Shen, C. E. Priebe, M. Maggioni, and J. T. Vogelstein. Dependence discovery from multimodal data via multiscale graph correlation. *Submitted*, 2016.

G. Szekely and M. Rizzo. The distance correlation t-test of independence in high dimension. *Journal of Multivariate Analysis*, 117:193–213, 2013.

G. Szekely, M. Rizzo, and N. Bakirov. Measuring and testing independence by correlation of distances. *Annals of Statistics*, 35(6):2769–2794, 2007.