Dependence Discovery via Multiscale Generalized Correlation

Cencheng Shen

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

August, 2016

Overview

1. Motivations
2. Results
   - Multiscale Generalized Correlation
   - Theoretical Advantages
   - Numerical Simulations
   - Real Data Experiments
3. Conclusion

Section 1

Motivations

Motivations

- Given a set of paired observations, testing independence is one of the most important and fundamental tasks in statistics and data science.

Motivations

- Given a set of paired observations, testing independence is one of the most important and fundamental tasks in statistics and data science.
- Indeed, prior to embarking on a predictive machine-learning investigation, one might first check whether any dependence is detectable; if not, high-quality predictions will be unlikely.

Motivations

- Given a set of paired observations, testing independence is one of the most important and fundamental tasks in statistics and data science.
- Indeed, prior to embarking on a predictive machine-learning investigation, one might first check whether any dependence is detectable; if not, high-quality predictions will be unlikely.
- Modern data sets may be high-dimensional, nonlinear, noisy, of small sample size; and different features may come from disparate spaces.

Motivations

- Given a set of paired observations, testing independence is one of the most important and fundamental tasks in statistics and data science.
- Indeed, prior to embarking on a predictive machine-learning investigation, one might first check whether any dependence is detectable; if not, high-quality predictions will be unlikely.
- Modern data sets may be high-dimensional, nonlinear, noisy, of small sample size; and different features may come from disparate spaces.
- A particular application is to investigate the association between brain activities and various phenotypes, such as brain vs disease, or brain vs personality. Despite recent inventions of many consistent test statistics, existing methods fall short for our real data.

- We desire a test that

- We desire a test that
- 1) is consistent as the sample size increases to infinity for most dependencies.

- We desire a test that
- 1) is consistent as the sample size increases to infinity for most dependencies.
- 2) has a good finite-sample testing power, for data of low or high dimensionality, linear or non-linearity, small sample size, with noise, etc.

- We desire a test that
- 1) is consistent as the sample size increases to infinity for most dependencies.
- 2) has a good finite-sample testing power, for data of low or high dimensionality, linear or non-linearity, small sample size, with noise, etc.
- 3) does not inflate the false positive rate in the absence of dependency.

- We desire a test that
- 1) is consistent as the sample size increases to infinity for most dependencies.
- 2) has a good finite-sample testing power, for data of low or high dimensionality, linear or non-linearity, small sample size, with noise, etc.
- 3) does not inflate the false positive rate in the absence of dependency.
- 4) easy and efficient to work with for complicated data.

- We desire a test that
- 1) is consistent as the sample size increases to infinity for most dependencies.
- 2) has a good finite-sample testing power, for data of low or high dimensionality, linear or non-linearity, small sample size, with noise, etc.
- 3) does not inflate the false positive rate in the absence of dependency.
- 4) easy and efficient to work with for complicated data.
- 5) provides insights into the dependency structure.

- We desire a test that
- 1) is consistent as the sample size increases to infinity for most dependencies.
- 2) has a good finite-sample testing power, for data of low or high dimensionality, linear or non-linearity, small sample size, with noise, etc.
- 3) does not inflate the false positive rate in the absence of dependency.
- 4) easy and efficient to work with for complicated data.
- 5) provides insights into the dependency structure.
- To that end, we propose multiscale generalized correlation in [*Shen et al.(2016)*][1] for testing independence, which combines distance correlation and nearest-neighbor into the same framework.

- We desire a test that
- 1) is consistent as the sample size increases to infinity for most dependencies.
- 2) has a good finite-sample testing power, for data of low or high dimensionality, linear or non-linearity, small sample size, with noise, etc.
- 3) does not inflate the false positive rate in the absence of dependency.
- 4) easy and efficient to work with for complicated data.
- 5) provides insights into the dependency structure.
- To that end, we propose multiscale generalized correlation in [*Shen et al.(2016)*][1] for testing independence, which combines distance correlation and nearest-neighbor into the same framework.
- MGC is theoretically consistent and efficient to compute, achieves excellent testing powers in a comprehensive set of simulations, sheds light into the underlying dependency structures, and works well for our real data.

Section 2

Results

Generalized Correlation Coefficient

Start with n pairs of observations $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ for $i = 1, \ldots, n$, where $\boldsymbol{x}$'s and $\boldsymbol{y}$'s both might be vectors of arbitrary dimensions, shapes, networks, etc.

Generalized Correlation Coefficient

Start with n pairs of observations $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ for $i = 1, \ldots, n$, where $\boldsymbol{x}$'s and $\boldsymbol{y}$'s both might be vectors of arbitrary dimensions, shapes, networks, etc.

Define a comparison function for each, i.e., $a_{ij} = \delta_x(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $b_{ij} = \delta_y(\boldsymbol{y}_i, \boldsymbol{y}_j)$.

Generalized Correlation Coefficient

Start with n pairs of observations $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ for $i = 1, \ldots, n$, where $\boldsymbol{x}$'s and $\boldsymbol{y}$'s both might be vectors of arbitrary dimensions, shapes, networks, etc.

Define a comparison function for each, i.e., $a_{ij} = \delta_x(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $b_{ij} = \delta_y(\boldsymbol{y}_i, \boldsymbol{y}_j)$.

Then $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ are the $n \times n$ interpoint comparison matrices for $X = \{\boldsymbol{x}_i\}$ and $Y = \{\boldsymbol{y}_i\}$, respectively. Assuming $\{a_{ij}\}$ and $\{b_{ij}\}$ have zero mean, a generalized correlation coefficient can then be written:

Generalized Correlation Coefficient

Start with n pairs of observations $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ for $i = 1, \ldots, n$, where $\boldsymbol{x}$'s and $\boldsymbol{y}$'s both might be vectors of arbitrary dimensions, shapes, networks, etc.

Define a comparison function for each, i.e., $a_{ij} = \delta_x(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $b_{ij} = \delta_y(\boldsymbol{y}_i, \boldsymbol{y}_j)$.

Then $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ are the $n \times n$ interpoint comparison matrices for $X = \{\boldsymbol{x}_i\}$ and $Y = \{\boldsymbol{y}_i\}$, respectively. Assuming $\{a_{ij}\}$ and $\{b_{ij}\}$ have zero mean, a generalized correlation coefficient can then be written:

$$c = \frac{1}{z}\sum_{i,j=1}^{n} a_{ij} b_{ij}, \tag{1}$$

where z is proportional to standard deviations of A and B, that is $z = n^2 \sigma_a \sigma_b$.

Examples

A few examples of c:

- It equals the Pearson's product-moment correlation coefficient by taking $a_{ij} = x_i$ and $b_{ij} = y_i$.

Examples

A few examples of c:

- It equals the Pearson's product-moment correlation coefficient by taking $a_{ij} = x_i$ and $b_{ij} = y_i$.
- Setting a_{ij} to be $rank(x_i) - rank(x_j)$ and $sign(x_i - x_j)$ respectively yields Spearman and Kendall's rank correlations.

Examples

A few examples of c:

- It equals the Pearson's product-moment correlation coefficient by taking $a_{ij} = x_i$ and $b_{ij} = y_i$.
- Setting a_{ij} to be $rank(x_i) - rank(x_j)$ and $sign(x_i - x_j)$ respectively yields Spearman and Kendall's rank correlations.
- Using $a_{ij} = |x_i - x_j|_2$ (i.e. Euclidean distance) yields the Mantel coefficient [*Mantel (1967)*][2], a popular test in biology and ecology.

Examples

A few examples of c:

- It equals the Pearson's product-moment correlation coefficient by taking $a_{ij} = x_i$ and $b_{ij} = y_i$.
- Setting a_{ij} to be $rank(x_i) - rank(x_j)$ and $sign(x_i - x_j)$ respectively yields Spearman and Kendall's rank correlations.
- Using $a_{ij} = |x_i - x_j|_2$ (i.e. Euclidean distance) yields the Mantel coefficient [*Mantel (1967)*][2], a popular test in biology and ecology.
- Using the doubly-centered distance entries for a_{ij} and b_{ij} yields the distance correlation [*Szekely et al.(2007)*][3], which is a consistent test for most dependencies.

Examples

A few examples of c:

- It equals the Pearson's product-moment correlation coefficient by taking $a_{ij} = x_i$ and $b_{ij} = y_i$.
- Setting a_{ij} to be $rank(x_i) - rank(x_j)$ and $sign(x_i - x_j)$ respectively yields Spearman and Kendall's rank correlations.
- Using $a_{ij} = |x_i - x_j|_2$ (i.e. Euclidean distance) yields the Mantel coefficient [*Mantel (1967)*][2], a popular test in biology and ecology.
- Using the doubly-centered distance entries for a_{ij} and b_{ij} yields the distance correlation [*Szekely et al.(2007)*][3], which is a consistent test for most dependencies.
- An unbiased version of dcorr can be achieved by slightly tweaking a_{ij}/b_{ij} of dcorr, which yields the modified distance correlation [*Szekely and Rizzo (2013)*] [4].

Rank-truncated pairwise comparisons

Multiscale Generalized Correlation (MGC) combines generalized correlation coefficients with locality, i.e., calculate the correlation between two sparse matrices based on nearest-neighbor.

Rank-truncated pairwise comparisons

Multiscale Generalized Correlation ($\textsc{Mgc}$) combines generalized correlation coefficients with locality, i.e., calculate the correlation between two sparse matrices based on nearest-neighbor.

Let $R(a_{ij})$ be the "rank" of $\boldsymbol{x}_i$ relative to $\boldsymbol{x}_j$, starting from 1 to n, that is, and define $R(b_{ij})$ equivalently for the $\boldsymbol{y}$'s. For any neighborhood size k around each $\boldsymbol{x}_i$ and any neighborhood size l around each $\boldsymbol{y}_i$, we define the rank-truncated pairwise comparisons:

Rank-truncated pairwise comparisons

Multiscale Generalized Correlation (MGC) combines generalized correlation coefficients with locality, i.e., calculate the correlation between two sparse matrices based on nearest-neighbor.

Let $R(a_{ij})$ be the "rank" of $\boldsymbol{x}_i$ relative to $\boldsymbol{x}_j$, starting from 1 to n, that is, and define $R(b_{ij})$ equivalently for the $\boldsymbol{y}$'s. For any neighborhood size k around each $\boldsymbol{x}_i$ and any neighborhood size l around each $\boldsymbol{y}_i$, we define the rank-truncated pairwise comparisons:

$$\widetilde{a}_{ij}^{k} = \begin{cases} a_{ij}, & \text{if } R(a_{ij}) \leq k, \\ 0, & \text{otherwise;} \end{cases} \qquad \widetilde{b}_{ij}^{l} = \begin{cases} b_{ij}, & \text{if } R(b_{ji}) \leq l, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Rank-truncated pairwise comparisons

Multiscale Generalized Correlation ($\textsc{Mgc}$) combines generalized correlation coefficients with locality, i.e., calculate the correlation between two sparse matrices based on nearest-neighbor.

Let $R(a_{ij})$ be the "rank" of $\boldsymbol{x}_i$ relative to $\boldsymbol{x}_j$, starting from 1 to n, that is, and define $R(b_{ij})$ equivalently for the $\boldsymbol{y}$'s. For any neighborhood size k around each $\boldsymbol{x}_i$ and any neighborhood size l around each $\boldsymbol{y}_i$, we define the rank-truncated pairwise comparisons:

$$\widetilde{a}_{ij}^{k} = \begin{cases} a_{ij}, & \text{if } R(a_{ij}) \leq k, \\ 0, & \text{otherwise;} \end{cases} \qquad \widetilde{b}_{ij}^{l} = \begin{cases} b_{ij}, & \text{if } R(b_{ji}) \leq l, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Then let $a_{ij}^{k} = \widetilde{a}_{ij}^{k} - \bar{a}^{k}$, where $\bar{a}^{k}$ is the local mean, and b_{ij}^{k} similarly.

Local Correlations

We can therefore define a *local* variant of any global generalized correlation coefficient by excluding large distances:

Local Correlations

We can therefore define a *local* variant of any global generalized correlation coefficient by excluding large distances:

$$c^{kl} = \frac{1}{z_{kl}} \sum_{i,j=1}^{n} a_{ij}^{k} b_{ij}^{l}, \tag{3}$$

Local Correlations

We can therefore define a *local* variant of any global generalized correlation coefficient by excluding large distances:

$$c^{kl} = \frac{1}{z_{kl}} \sum_{i,j=1}^{n} a_{ij}^{k} b_{ij}^{l}, \tag{3}$$

where $z_{kl} = n^2 \sigma_a^k \sigma_b^l$, with σ_a^k and σ_b^l being the standard deviations for the truncated pairwise comparisons.

Local Correlations

We can therefore define a *local* variant of any global generalized correlation coefficient by excluding large distances:

$$c^{kl} = \frac{1}{z_{kl}} \sum_{i,j=1}^{n} a_{ij}^{k} b_{ij}^{l}, \tag{3}$$

where $z_{kl} = n^2 \sigma_a^k \sigma_b^l$, with σ_a^k and σ_b^l being the standard deviations for the truncated pairwise comparisons.

There are a maximum of $(n^2 - n)/2$ different local correlations, one for each possible combination of k and l. And our definition always yields symmetric local correlations, i.e. $c^{kl}(X, Y) = c^{lk}(Y, X)$, no matter A and B are symmetric or not.

MGC

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

MGC

- In the family of local correlations $\{c_{kl}\}$, 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 $\{c_{kl}\}$, 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 c^*.

MGC

- In the family of local correlations $\{c_{kl}\}$, 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 c^*.
- Although each local correlation requires $O(n^2)$ to compute, our formulation allows all local correlations to be simultaneously calculated in $O(n^2)$ as well! (Note: sorting the distance matrix column-wise takes another $O(n^2 \log(n))$).

MGC

- In the family of local correlations $\{c_{kl}\}$, 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 c^*.
- Although each local correlation requires $O(n^2)$ to compute, our formulation allows all local correlations to be simultaneously calculated in $O(n^2)$ as well! (Note: sorting the distance matrix column-wise takes another $O(n^2 \log(n))$).
- 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

The formal testing scenario is as follows: assume that $\boldsymbol{x}_i, i = 1, \ldots, n$ are identically independently distributed (i.i.d.) as $\boldsymbol{x} \sim f_x$; similarly each $\boldsymbol{y}_i$ are i.i.d. as $\boldsymbol{y} \sim f_y$.

The Testing Framework

The formal testing scenario is as follows: assume that $\boldsymbol{x}_i, i = 1, \ldots, n$ are identically independently distributed (i.i.d.) as $\boldsymbol{x} \sim f_x$; similarly each $\boldsymbol{y}_i$ are i.i.d. as $\boldsymbol{y} \sim f_y$.

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 $(\boldsymbol{x}, \boldsymbol{y})$.

The Testing Framework

The formal testing scenario is as follows: assume that $\boldsymbol{x}_i, i = 1, \ldots, n$ are identically independently distributed (i.i.d.) as $\boldsymbol{x} \sim f_x$; similarly each $\boldsymbol{y}_i$ are i.i.d. as $\boldsymbol{y} \sim f_y$.

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 $(\boldsymbol{x}, \boldsymbol{y})$.

To test on a pair of sample data, we can use the p-value of a permutation test, and reject the null when the p-value is sufficiently small.

The Testing Framework

The formal testing scenario is as follows: assume that $\boldsymbol{x}_i, i = 1, \ldots, n$ are identically independently distributed (i.i.d.) as $\boldsymbol{x} \sim f_x$; similarly each $\boldsymbol{y}_i$ are i.i.d. as $\boldsymbol{y} \sim f_y$.

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 $(\boldsymbol{x}, \boldsymbol{y})$.

To test on a pair of sample data, we can use the p-value of a permutation test, and reject the null when the p-value is sufficiently small.

The power of a test is defined as the probability that it correctly rejects the null when the null is indeed false, and has power equal to the type 1 error level when the null is true. And a test is universally consistent if its power converges to 1 as $n \to \infty$ whenever $f_{xy} \neq f_x f_y$.

Theorems of MGC

Theorem 1

MGC *is theoretically consistent against all dependent alternatives for which its global counterpart is.*

Theorems of MGC

Theorem 1

MGC *is theoretically consistent against all dependent alternatives for which its global counterpart is.*

Theorem 2

If $\boldsymbol{x}$ *is linearly dependent on* $\boldsymbol{y}$*, then the optimal scale for* MGC *is always the global scale.*

Theorems of MGC

Theorem 1

MGC *is theoretically consistent against all dependent alternatives for which its global counterpart is.*

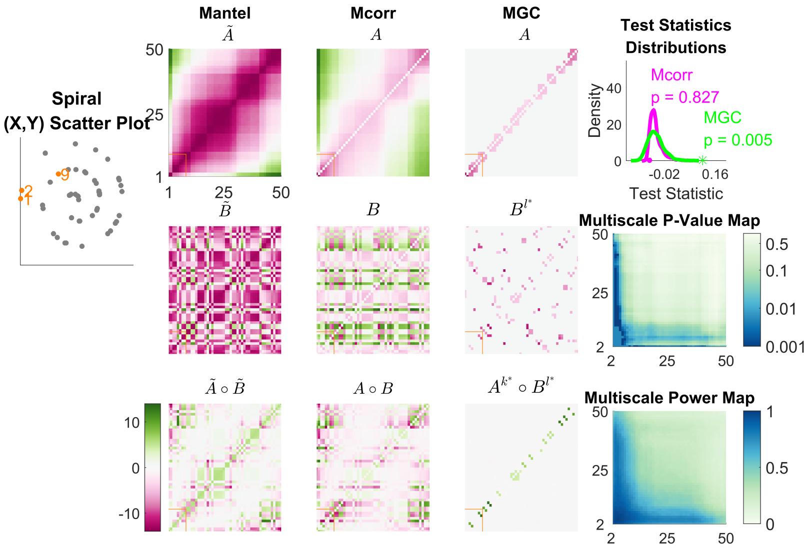
Theorem 2

If $\boldsymbol{x}$ *is linearly dependent on* $\boldsymbol{y}$, *then the optimal scale for* MGC *is always the global scale.*

Theorem 3

There exists f_{xy} *and* n *such that* MGC *is better than its global counterpart in testing power.*

Illustration of Utilizing locality



Advantage of Utilizing Locality

	Mantel	Mcorr	MGC
$\delta_x(1,2)$	-2.42	-5.21	-5.07
$\delta_y(1,2)$	-1.58	-0.91	-0.12
$\delta_x \times \delta_y$	3.82	4.74	0.61
$\delta_x(2,9)$	0.70	0.61	0.14
$\delta_y(2,9)$	-0.91	-0.28	0.12
$\delta_x \times \delta_y$	-0.63	-0.17	0.02
$\sum \delta_x \times \delta_y$	-162.14	-93.04	116.41
$\sum \delta_x \times \delta_y / \sum \delta_x^2 \sum \delta_y^2$	-0.02	-0.02	0.16

Simulation Set-Up

- We are interested in assessing the performance of our newly proposed multiscale tests in a wide variety of settings to better understand the tests, and gain insight into which to use in different settings.

Simulation Set-Up

- We are interested in assessing the performance of our newly proposed multiscale tests in a wide variety of settings to better understand the tests, and gain insight into which to use in different settings.
- In total 20 different joint distributions f_{xy} are considered, most of which are taken exactly from existing literature, including linear and nearly linear (1-5), polynomial (6-12), trigonometric (13-17), uncorrelated but nonlinearly dependent (18-19), and an independent relationship (20).

Simulation Set-Up

- We are interested in assessing the performance of our newly proposed multiscale tests in a wide variety of settings to better understand the tests, and gain insight into which to use in different settings.
- In total 20 different joint distributions f_{xy} are considered, most of which are taken exactly from existing literature, including linear and nearly linear (1-5), polynomial (6-12), trigonometric (13-17), uncorrelated but nonlinearly dependent (18-19), and an independent relationship (20).
- 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 and increasing dimensions.

Simulation Set-Up

- We are interested in assessing the performance of our newly proposed multiscale tests in a wide variety of settings to better understand the tests, and gain insight into which to use in different settings.
- In total 20 different joint distributions f_{xy} are considered, most of which are taken exactly from existing literature, including linear and nearly linear (1-5), polynomial (6-12), trigonometric (13-17), uncorrelated but nonlinearly dependent (18-19), and an independent relationship (20).
- 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 and increasing dimensions.
- The benchmarks are distance correlation, modified distance correlation, the Mantel test, and the HHG method proposed in [*Heller et al.(2013)*][5].

Simulation Set-Up

- We are interested in assessing the performance of our newly proposed multiscale tests in a wide variety of settings to better understand the tests, and gain insight into which to use in different settings.
- In total 20 different joint distributions f_{xy} are considered, most of which are taken exactly from existing literature, including linear and nearly linear (1-5), polynomial (6-12), trigonometric (13-17), uncorrelated but nonlinearly dependent (18-19), and an independent relationship (20).
- 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 and increasing dimensions.
- The benchmarks are distance correlation, modified distance correlation, the Mantel test, and the HHG method proposed in [*Heller et al.(2013)*][5].
- Our MGC implementation is based on modified distance correlation.

Visualizations of Simulation Settings

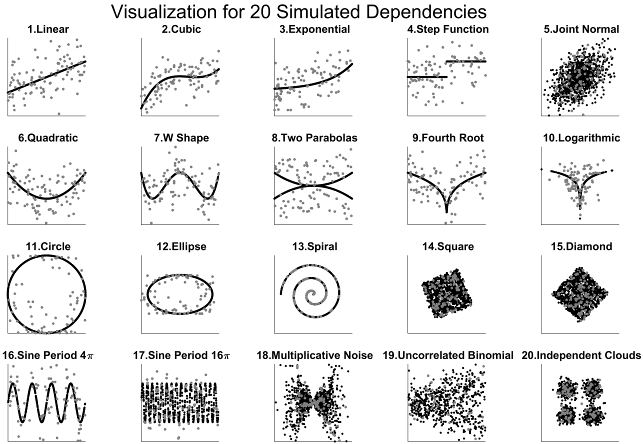


Figure: Visualization of the 20 dependencies for one-dimensional simulations.

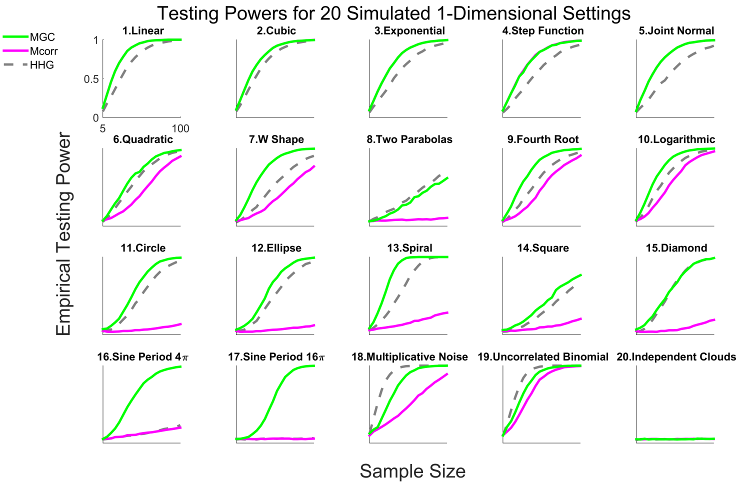


Figure: Powers of different methods for 20 different one-dimensional dependence structures.

Simulation Power Map

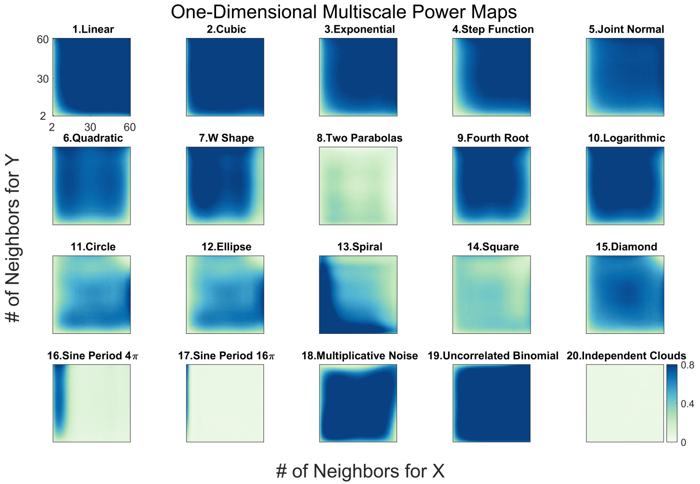


Figure: Multiscale Power Maps indicating the influence of neighborhood size on MGC testing power.

Brain vs Personality

Our first real data experiment investigates whether there is any dependency between the brain activity and personality.

Brain vs Personality

Our first real data experiment investigates whether there is any dependency between the brain activity and personality.

[*Adelstein et al.(2011)*] [6] was able to detect dependence between certain regions and dimensions of personality, but lacked the tools to test for dependence of the whole brain activity against all five dimensions of personality.

Brain vs Personality

Our first real data experiment investigates whether there is any dependency between the brain activity and personality.

[*Adelstein et al.(2011)*] [6] was able to detect dependence between certain regions and dimensions of personality, but lacked the tools to test for dependence of the whole brain activity against all five dimensions of personality.

In this dataset, we have $n = 42$ subjects, for each we obtained 197 time-steps of resting-state functional MRI activity, as well as her five-factor personality trait as quantified by the NEO Personality Inventory-Revised.

Brain vs Personality

Our first real data experiment investigates whether there is any dependency between the brain activity and personality.

[*Adelstein et al.(2011)*] [6] was able to detect dependence between certain regions and dimensions of personality, but lacked the tools to test for dependence of the whole brain activity against all five dimensions of personality.

In this dataset, we have $n = 42$ subjects, for each we obtained 197 time-steps of resting-state functional MRI activity, as well as her five-factor personality trait as quantified by the NEO Personality Inventory-Revised.

The raw brain activity was processed using CPAC [*Craddock et al.(2015)*][7] resulting in 197 brain regions. We ran a spectral analysis on each region, bandpassed and normalized it, and then calculated the Kullback-Leibler divergence across regions and the normalized Hellinger distance between each subject.

Brain vs Personality

Our first real data experiment investigates whether there is any dependency between the brain activity and personality.

[*Adelstein et al.(2011)*] [6] was able to detect dependence between certain regions and dimensions of personality, but lacked the tools to test for dependence of the whole brain activity against all five dimensions of personality.

In this dataset, we have $n = 42$ subjects, for each we obtained 197 time-steps of resting-state functional MRI activity, as well as her five-factor personality trait as quantified by the NEO Personality Inventory-Revised.

The raw brain activity was processed using CPAC [*Craddock et al.(2015)*][7] resulting in 197 brain regions. We ran a spectral analysis on each region, bandpassed and normalized it, and then calculated the Kullback-Leibler divergence across regions and the normalized Hellinger distance between each subject.

For the five-factor personality data, we use the Euclidean distance.

Permutation Test P-value map

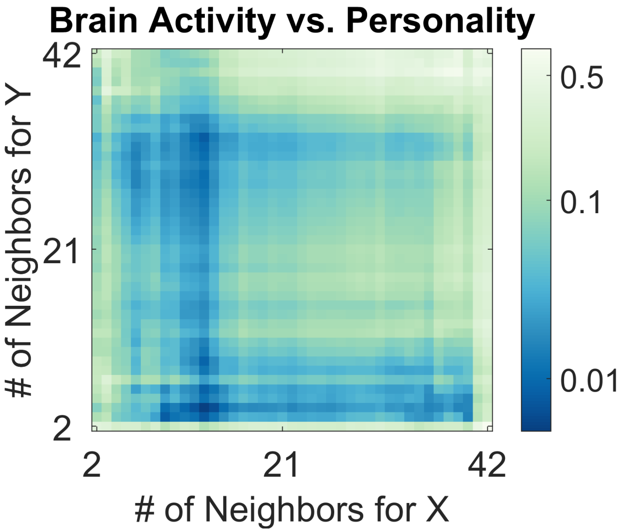


Figure: P-value map (log scale) for brain fMRI scan vs five-factor personality.

Section 3

Conclusion

Advantages of MGC

- 1) MGC is easy and efficient to implement for any generalized correlation.

Advantages of MGC

- 1) MGC is easy and efficient to implement for any generalized correlation.
- 2) MGC is consistent for testing independence when based on dcorr or mcorr.

Advantages of MGC

- 1) MGC is easy and efficient to implement for any generalized correlation.
- 2) MGC is consistent for testing independence when based on dcorr or mcorr.
- 3) MGC is equivalent to the respective global correlation under linear dependence, but can perform better under nonlinear dependence.

Advantages of MGC

- 1) MGC is easy and efficient to implement for any generalized correlation.
- 2) MGC is consistent for testing independence when based on dcorr or mcorr.
- 3) MGC is equivalent to the respective global correlation under linear dependence, but can perform better under nonlinear dependence.
- 4) MGC exhibits superior numerical performance in a comprehensive simulation setting, including various linear/nonlinear/high-dimensional/noisy dependency.

Advantages of MGC

- 1) MGC is easy and efficient to implement for any generalized correlation.
- 2) MGC is consistent for testing independence when based on dcorr or mcorr.
- 3) MGC is equivalent to the respective global correlation under linear dependence, but can perform better under nonlinear dependence.
- 4) MGC exhibits superior numerical performance in a comprehensive simulation setting, including various linear/nonlinear/high-dimensional/noisy dependency.
- 5) MGC not only yields a p-value in real data testing, but also provides useful information on the local scale where the dependency is the strongest, which implies the geometry of the underlying dependency.

What we do not show in slides

- For high-dimensional simulations MGC is superior for almost all simulations, with more

What we do not show in slides

- For high-dimensional simulations MGC is superior for almost all simulations, with more
- We applied MGC to detect dependency between brain shape vs disease; and another experiment to show MGC does not inflate false positive error in brain activity vs fake movie.

What we do not show in slides

- For high-dimensional simulations MGC is superior for almost all simulations, with more
- We applied MGC to detect dependency between brain shape vs disease; and another experiment to show MGC does not inflate false positive error in brain activity vs fake movie.
- Our MGC implementation is based on an equivalent but different mcorr version. In particular, we used single-centering rather than double centering, which is equivalent for the global correlation in testing power and p-value, but preserves the rank information better for local correlations.

What we do not show in slides

- For high-dimensional simulations MGC is superior for almost all simulations, with more
- We applied MGC to detect dependency between brain shape vs disease; and another experiment to show MGC does not inflate false positive error in brain activity vs fake movie.
- Our MGC implementation is based on an equivalent but different mcorr version. In particular, we used single-centering rather than double centering, which is equivalent for the global correlation in testing power and p-value, but preserves the rank information better for local correlations.
- When applying MGC to sample data without known model or training data, we provide a heuristic algorithm to accurately select the optimal scale by finding consecutive regions of significant p-values.

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

N. Mantel, "The detection of disease clustering and a generalized regression approach," *Cancer Research*, vol. 27, no. 2, pp. 209–220, 1967.

G. Szekely, M. Rizzo, and N. Bakirov, "Measuring and testing independence by correlation of distances," *Annals of Statistics*, vol. 35, no. 6, pp. 2769–2794, 2007.

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

R. Heller, Y. Heller, and M. Gorfine, "A consistent multivariate test of association based on ranks of distances," *Biometrika*, vol. 100, no. 2, pp. 503–510, 2013.

J. Adelstein, Z. Shehzad, M. Mennes, C. DeYoung, X. Zuo, C. Kelly, D. Margulies, A. Bloomfield, J. Gray, F. Castellanos, and M. Milham, "Personality is reflected in the brain's intrinsic functional architecture," *PLoS ONE*, vol. 6, no. 11, p. e27633, 2011.

C. Craddock, S. Sikka, B. Cheung, R. Khanuja, S. S. Ghosh, C. Yan, Q. Li, D. Lurie, J. Vogelstein, R. Burns, S. Colcombe, M. Mennes, C. Kelly, A. Di Martino, F. X. Castellanos, and M. Milham, "Towards Automated Analysis of Connectomes: The Configurable Pipeline for the Analysis of Connectomes (C-PAC)," *Frontiers in Neuroinformatics*, vol. 42, 2015.