Dependence Discovery via Multiscale Generalized Correlation

Cencheng Shen

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

August, 2016

Overview

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.

Prior to embarking on a predictive machine-learning investigation, one might first check whether any dependence is detectable.

Motivations

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

Prior to embarking on a predictive machine-learning investigation, one might first check whether any dependence is detectable.

There are many real-data applications that existing methods fall short of in detecting associations.

Motivations

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

Prior to embarking on a predictive machine-learning investigation, one might first check whether any dependence is detectable.

There are many real-data applications that existing methods fall short of in detecting associations.

Modern data sets may be **high-dimensional, nonlinear, noisy, of small sample size, from disparate spaces**, etc.

We desire a test that

We desire a test that

- is consistent as the sample size increases to infinity for most dependencies.

We desire a test that

- is consistent as the sample size increases to infinity for most dependencies.
- 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

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

We desire a test that

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

We desire a test that

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

We desire a test that

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

We desire a test that

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

To that end, we propose **multiscale generalized correlation** in [*Shen et al.(2016)*][1].

We desire a test that

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

To that end, we propose **multiscale generalized correlation** in [*Shen et al.(2016)*][1].

MGC combines distance correlation and nearest-neighbor, and is able to satisfy all of the above.

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:

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

Examples

A few examples of c:

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

Examples

A few examples of c:

- The Pearson's product-moment correlation coefficient by taking $a_{ij} = x_i$ and $b_{ij} = y_i$.
- The Spearman and Kendall's rank correlations by setting a_{ij} to be $rank(x_i) - rank(x_j)$ and $sign(x_i - x_j)$ respectively.
- The Mantel coefficient [*Mantel (1967)*][2] by using $a_{ij} = |x_i - x_j|_2$ (i.e. Euclidean distance).

Examples

A few examples of c:

- The Pearson's product-moment correlation coefficient by taking $a_{ij} = x_i$ and $b_{ij} = y_i$.
- The Spearman and Kendall's rank correlations by setting a_{ij} to be $rank(x_i) - rank(x_j)$ and $sign(x_i - x_j)$ respectively.
- The Mantel coefficient [*Mantel (1967)*][2] by using $a_{ij} = |x_i - x_j|_2$ (i.e. Euclidean distance).
- The distance correlation [*Szekely et al.(2007)*][3] by using the doubly-centered distance entries for a_{ij} and b_{ij}.

A few examples of c:

- The Pearson's product-moment correlation coefficient by taking $a_{ij} = x_i$ and $b_{ij} = y_i$.
- The Spearman and Kendall's rank correlations by setting a_{ij} to be $rank(x_i) - rank(x_j)$ and $sign(x_i - x_j)$ respectively.
- The Mantel coefficient [*Mantel (1967)*][2] by using $a_{ij} = |x_i - x_j|_2$ (i.e. Euclidean distance).
- The distance correlation [*Szekely et al.(2007)*][3] by using the doubly-centered distance entries for a_{ij} and b_{ij}.
- The modified distance correlation [*Szekely and Rizzo (2013)*] [4] by slightly tweaking a_{ij}/b_{ij} of dcorr.

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 $\boldsymbol{x}_j$, and similarly $R(b_{ij})$ 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 ($\textsc{Mgc}$) 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 $\boldsymbol{x}_j$, and similarly $R(b_{ij})$ 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 $\boldsymbol{x}_j$, and similarly $R(b_{ij})$ 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 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:

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

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, and they are always symmetric, i.e. $c^{kl}(X, Y) = c^{lk}(Y, X)$, no matter A and B are symmetric or not.

MGC

In $\{c_{kl}\}$, the optimal local correlation (with respect to the independence testing power) exists, is distribution dependent, and may not be unique.

MGC

In $\{c_{kl}\}$, the optimal local correlation (with respect to the independence testing power) exists, is distribution dependent, and may not be unique.

The optimal local correlation coefficient is dubbed as the multiscale graph correlation c^*.

MGC

In $\{c_{kl}\}$, the optimal local correlation (with respect to the independence testing power) exists, is distribution dependent, and may not be unique.

The optimal local correlation coefficient is dubbed as the multiscale graph correlation c^*.

Each local correlation requires $O(n^2)$ to compute. However, once A and B are sorted, **all local correlations can be simultaneously calculated in $O(n^2)$ as well!!!**

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

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

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

The Testing Framework

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

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

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

To test on a pair of sample data, we use the p-value of a permutation test.

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 use the p-value of a permutation test.

A test is universally consistent if its power converges to 1 as $n \to \infty$ whenever $f_{xy} \neq f_x f_y$.

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

To test on a pair of sample data, we use the p-value of a permutation test.

A test is universally consistent if its power converges to 1 as $n \rightarrow \infty$ whenever $f_{xy} \neq f_x f_y$.

But we all want a test with high power in finite-sample rather than asymptotically!

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.*

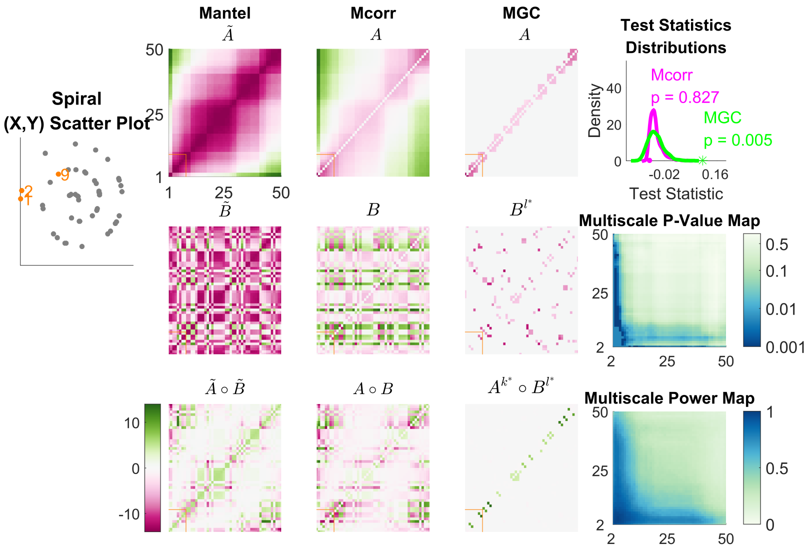
Theorem 2

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

In total 20 different joint distributions f_{xy} are considered, 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

In total 20 different joint distributions f_{xy} are considered, 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 (not shown in slides).

Simulation Set-Up

In total 20 different joint distributions f_{xy} are considered, 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 (not shown in slides).

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

In total 20 different joint distributions f_{xy} are considered, 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 (not shown in slides).

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 with single centering.

Visualizations of Simulation Settings

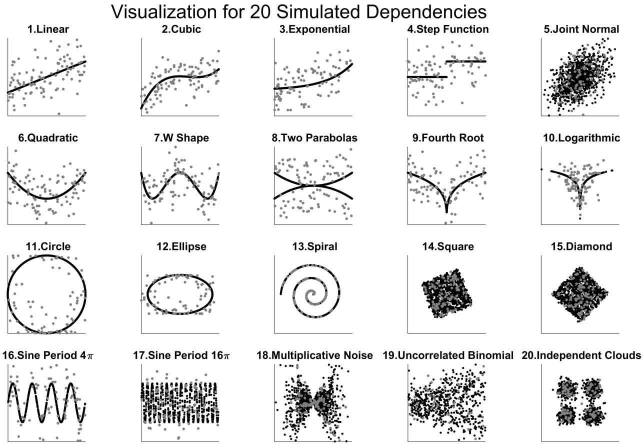


Figure: Visualization of the 20 dependencies for one-dimensional simulations.

Simulation Powers

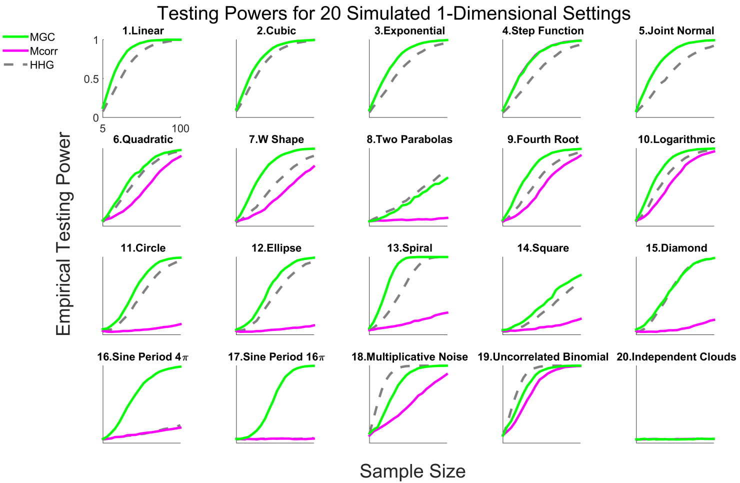


Figure: Powers of different methods for 20 different one-dimensional dependence structures.

Simulation Power Map

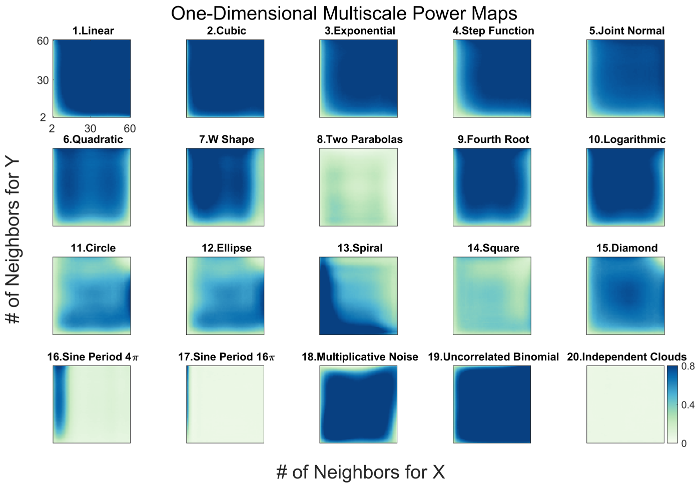


Figure: Multiscale Power Maps indicating the influence of neighborhood size on MGC testing power.

Brain vs Personality

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

Brain vs Personality

[*Adelstein et al.(2011)*] [6] was able to detect dependence between certain brain regions and 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 the five-factor personality trait as quantified by the NEO Personality Inventory-Revised.

Brain vs Personality

[*Adelstein et al.(2011)*] [6] was able to detect dependence between certain brain regions and 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 the 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]. Then 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

[*Adelstein et al.(2011)*] [6] was able to detect dependence between certain brain regions and 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 the 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]. Then 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 directly use the Euclidean distance.

Permutation Test P-value map

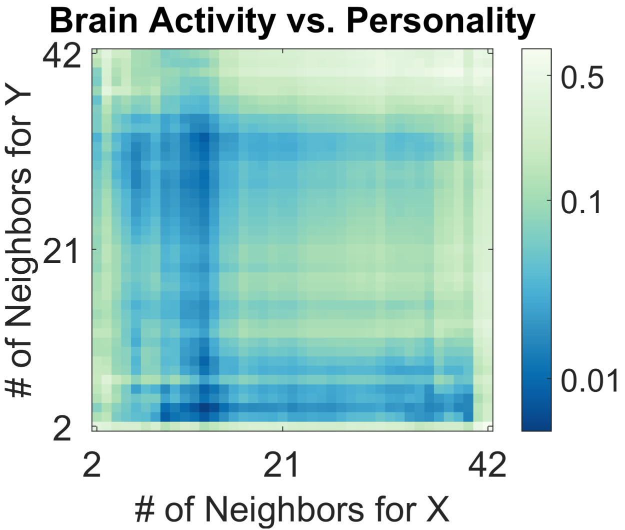


Figure: P-value map (log scale) for brain fMRI scan vs five-factor personality.

Section 3

Conclusion

Advantages of MGC

- 1) MGC is consistent against all distributions of finite second moment when based on dcorr or mcorr.

Advantages of MGC

- 1) MGC is consistent against all distributions of finite second moment when based on dcorr or mcorr.
- 2) MGC exhibits superior numerical performance in a comprehensive simulation setting, throughout various linear / nonlinear / high-dimensional / noisy dependencies.

Advantages of MGC

- 1) MGC is consistent against all distributions of finite second moment when based on dcorr or mcorr.
- 2) MGC exhibits superior numerical performance in a comprehensive simulation setting, throughout various linear / nonlinear / high-dimensional / noisy dependencies.
- 3) MGC is easy and efficient to implement for any generalized correlation.

Advantages of MGC

- 1) MGC is consistent against all distributions of finite second moment when based on dcorr or mcorr.
- 2) MGC exhibits superior numerical performance in a comprehensive simulation setting, throughout various linear / nonlinear / high-dimensional / noisy dependencies.
- 3) MGC is easy and efficient to implement for any generalized correlation.
- 4) MGC not only outputs a p-value in testing, but also provides useful information on the local scale where the dependency is the strongest, and implies the geometry of the underlying dependency.

What we do not show in slides

- MGC superiority for high-dimensional simulations.

What we do not show in slides

- MGC superiority for high-dimensional simulations.
- MGC applications for brain shape vs disease, and brain activity vs fake movie.

What we do not show in slides

- MGC superiority for high-dimensional simulations.
- MGC applications for brain shape vs disease, and brain activity vs fake movie.
- MGC implementation by single centered dcorr, which is equivalent to the original dcorr in testing but better preserves the rank information.

What we do not show in slides

- MGC superiority for high-dimensional simulations.
- MGC applications for brain shape vs disease, and brain activity vs fake movie.
- MGC implementation by single centered dcorr, which is equivalent to the original dcorr in testing but better preserves the rank information.
- MGC scale selection algorithm based on the p-value map, which heuristically selects the optimal scale by finding consecutive regions of significant p-values.

What we do not show in slides

- MGC superiority for high-dimensional simulations.
- MGC applications for brain shape vs disease, and brain activity vs fake movie.
- MGC implementation by single centered dcorr, which is equivalent to the original dcorr in testing but better preserves the rank information.
- MGC scale selection algorithm based on the p-value map, which heuristically selects the optimal scale by finding consecutive regions of significant p-values.

What we do not show in slides

- MGC superiority for high-dimensional simulations.
- MGC applications for brain shape vs disease, and brain activity vs fake movie.
- MGC implementation by single centered dcorr, which is equivalent to the original dcorr in testing but better preserves the rank information.
- MGC scale selection algorithm based on the p-value map, which heuristically selects the optimal scale by finding consecutive regions of significant p-values.

And many other details...

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.