LP Comoments: A Prelude

Deep Mukhopadhyay

Department of Statistics
Temple University

deep@temple.edu

Nov 22, 2013

The Algorithm

(1) Input: (X, Y), sample size n.

(2) Transformation: $(X, Y) \rightarrow (F^{\mathrm{mid}}(X; X),\ F^{\mathrm{mid}}(Y; Y))$.

$$F^{\mathrm{mid}} \leftarrow \text{function}(z)\left\{ (\texttt{rank}[z] - .5)/n \right\}.$$

(3) Orthonormal Score Polynomials construction:

$$\begin{aligned} F^{\mathrm{mid}}(X; X) &\rightarrow \left\{ T_1(X; X), \ldots, T_{m_1}(X; X) \right\} := TX \\ F^{\mathrm{mid}}(Y; Y) &\rightarrow \left\{ T_1(Y; Y), \ldots, T_{m_2}(Y; Y) \right\} := TY \end{aligned}$$

$$\text{LP.poly} \leftarrow \text{function}(z, m)\left\{ \texttt{slegendre.polynomials}(F^{\mathrm{mid}}(z), m) \right\}.$$

(4) Output: Cross-Score Covariance (LP-Comoment) Operator

$$\widehat{\mathrm{LP}}_{m_1 \times m_2}(X, Y) \leftarrow \texttt{cov}(TX, TY).$$

The Algorithm

(1) Input: (X, Y), sample size n.

(2) Transformation: $(X, Y) \rightarrow (F^{\mathrm{mid}}(X; X),\ F^{\mathrm{mid}}(Y; Y))$.

$$F^{\mathrm{mid}} \leftarrow \text{function}(z)\,\{\ (\texttt{rank}[\text{z}] - .5)/\text{n}\ \}.$$

(3) Orthonormal Score Polynomials construction:

$$F^{\mathrm{mid}}(X; X) \rightarrow \{T_1(X; X), \ldots, T_{m_1}(X; X)\} := TX$$
$$F^{\mathrm{mid}}(Y; Y) \rightarrow \{T_1(Y; Y), \ldots, T_{m_2}(Y; Y)\} := TY$$

$$\text{LP.poly} \leftarrow \text{function}(z, m)\,\{\ \texttt{slegendre.polynomials}(F^{\mathrm{mid}}(z), \text{m})\ \}.$$

(4) Output: Cross-Score Covariance (LP-Comoment) Operator

$$\widehat{\mathrm{LP}}_{m_1 \times m_2}(X, Y) \leftarrow \texttt{cov}(TX, TY).$$

The Algorithm

(1) Input: (X, Y), sample size n.

(2) Transformation: $(X, Y) \;\to\; (F^{\mathrm{mid}}(X; X),\; F^{\mathrm{mid}}(Y; Y))$.

$$F^{\mathrm{mid}} \;\leftarrow\; \text{function}(z)\,\{\; (\texttt{rank}[\text{z}] \text{ - .5})/\text{n} \;\}.$$

(3) Orthonormal Score Polynomials construction:

$$F^{\mathrm{mid}}(X; X) \;\to\; \big\{T_1(X; X), \ldots, T_{m_1}(X; X)\big\} := TX$$
$$F^{\mathrm{mid}}(Y; Y) \;\to\; \big\{T_1(Y; Y), \ldots, T_{m_2}(Y; Y)\big\} := TY$$

$$\text{LP.poly} \;\leftarrow\; \text{function}(z, m)\,\{\; \texttt{slegendre.polynomials}(F^{\mathrm{mid}}(z), \text{m}) \;\}.$$

(4) Output: Cross-Score Covariance (LP-Comoment) Operator

$$\widehat{\mathrm{LP}}_{m_1 \times m_2}(X, Y) \;\leftarrow\; \texttt{cov}(TX, TY).$$

The Algorithm

(1) Input: (X, Y), sample size n.

(2) Transformation: $(X, Y) \;\rightarrow\; (F^{\mathrm{mid}}(X;X),\; F^{\mathrm{mid}}(Y;Y))$.

$$F^{\mathrm{mid}} \;\leftarrow\; \text{function}(z)\,\{\; (\texttt{rank}[\text{z}] \text{ - } .5)/\text{n} \;\}.$$

(3) Orthonormal Score Polynomials construction:

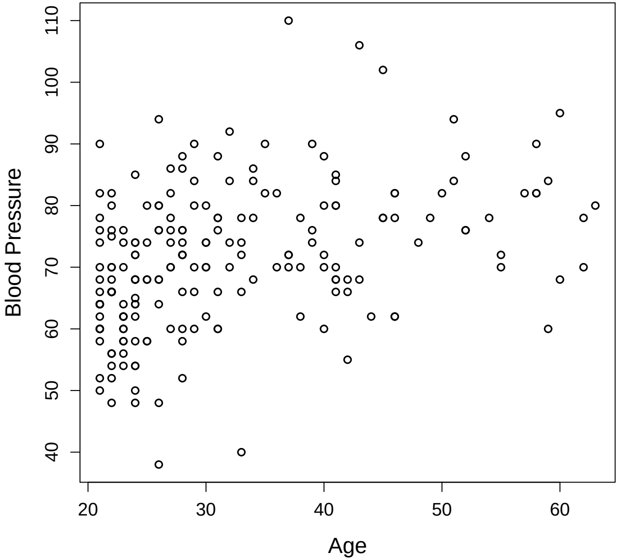
$$\begin{aligned} F^{\mathrm{mid}}(X;X) \;&\rightarrow\; \big\{T_1(X;X), \ldots, T_{m_1}(X;X)\big\} \;:=\; TX \\ F^{\mathrm{mid}}(Y;Y) \;&\rightarrow\; \big\{T_1(Y;Y), \ldots, T_{m_2}(Y;Y)\big\} \;:=\; TY \end{aligned}$$

$$\text{LP.poly} \;\leftarrow\; \text{function}(z, m)\,\{\; \texttt{slegendre.polynomials}(F^{\mathrm{mid}}(z), \text{ m}) \;\}.$$

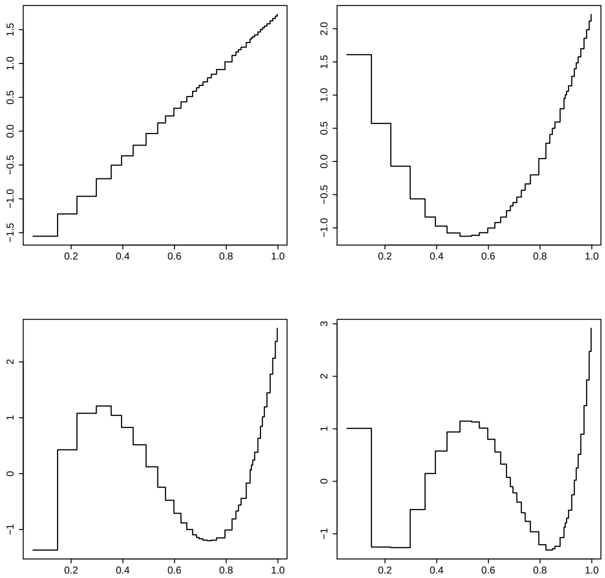
(4) Output: Cross-Score Covariance (LP-Comoment) Operator

$$\widehat{\mathrm{LP}}_{m_1 \times m_2}(X, Y) \;\leftarrow\; \texttt{cov}(TX, TY).$$

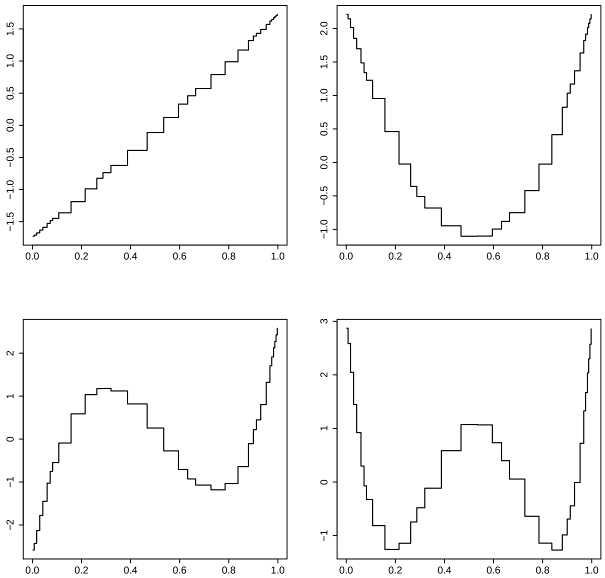
Pima Indians Diabetes Data Set



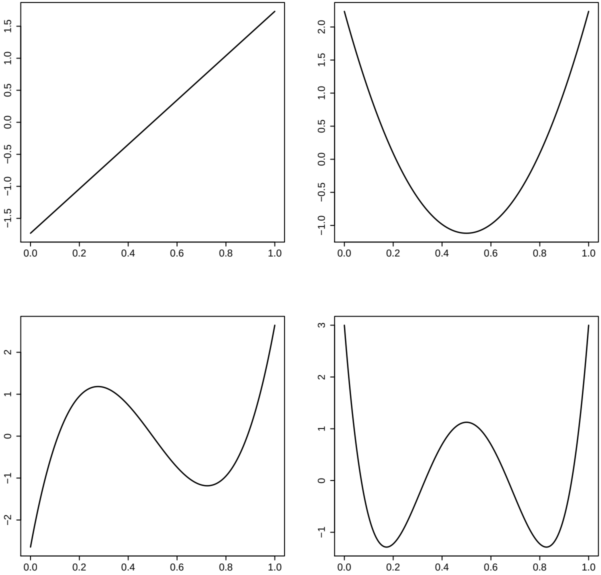
Age Score Polynomials



BP Score Polynomials



Shifted Legendre Polynomials



Shifted Orthonormal Legendre Polynomials

Orthonormal Legendre polynomials $\mathrm{Leg}_j(u)$ on interval $0 < u < 1$

$$
\begin{aligned}
\mathrm{Leg}_0(u) &= 1 \\
\mathrm{Leg}_1(u) &= \sqrt{12}(u - .5) \\
\mathrm{Leg}_1(u) &= \sqrt{5}(6u^2 - 6u + 1) \\
\mathrm{Leg}_3(u) &= \sqrt{7}(20u^3 - 30u^2 + 12u - 1) \\
\mathrm{Leg}_4(u) &= 3(70u^4 - 140u^3 + 90u^2 - 20u + 1) \\
&\vdots
\end{aligned}
$$

LP Score Comoments: $\mathrm{LP}(j, k; X, Y)$

$$\mathrm{LP}(j, k; X, Y) = \mathbb{E}\big[T_j(X;X)\,T_k(Y;Y)\big] \quad \text{for} \quad j, k > 0.$$

LP-Comoment	$T_1(Y;Y)$	$T_2(Y;Y)$	$T_3(Y;Y)$	$T_4(Y;Y)$
$T_1(X;X)$	**0.443**	-0.016	-0.011	0.033
$T_2(X;X)$	-0.009	0.065	0.080	**-0.136**
$T_3(X;X)$	-0.078	0.044	-0.088	0.008
$T_4(X;X)$	**0.110**	-0.013	-0.027	-0.019

Score Correlation Matrix

LP-Comoment: Fundamental Statistical Tool

- LP-Nonparametric Dependence Learning.
- LP-Nonparametric Copula Density Estimation and Applications.
- LP-Nonparametric Copula Shape Identification.
- LP-Modeling to Dyadic Data.
- LP-Nonparametric Copula Based Graphical Model.

$\vdots \quad \vdots$

LPD: New Class of Dependence Measures

Definition: **LP**-comoment matrix based **D**ependence measures.

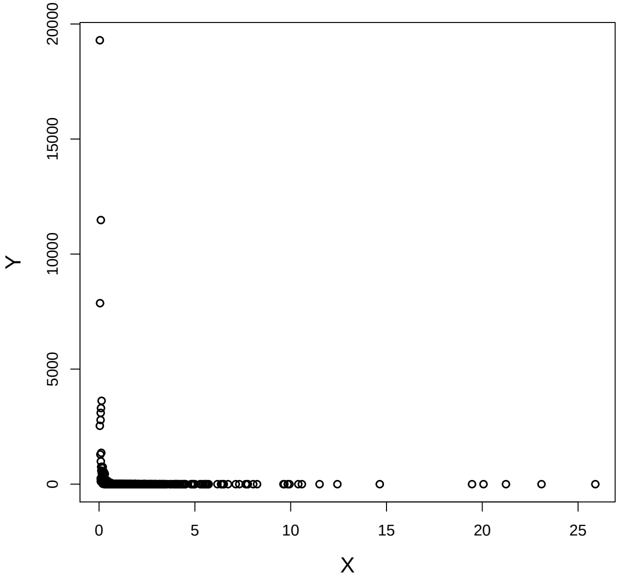
(1) LPINFOR: square Frobenius norm $\| \mathrm{LP} \|_F^2 = \sum_{j,k} \big| \mathrm{LP}[j,k] \big|^2$.

(2) LPSpec: Spectral norm $\lambda_{\max}(\mathrm{LP})$.

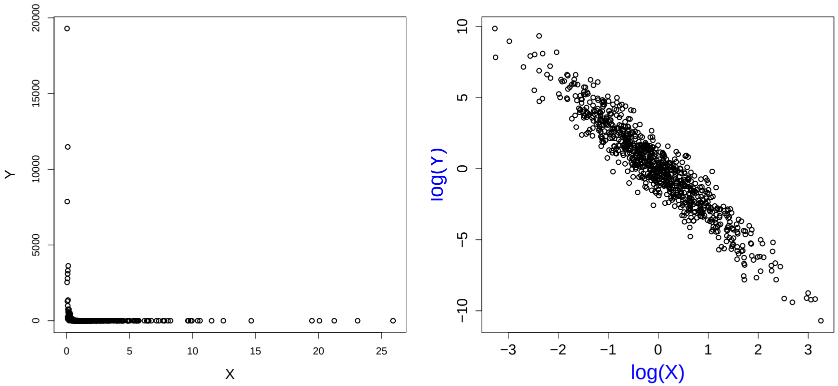
(3) LPMaximal: $\| \mathrm{LP} \|_\infty = \max_{j,k} \big| \mathrm{LP}[j,k] \big|$.

(4) For Diabetes Data:

- LPINFOR(X,Y) $=$ .256.
- LP-Spectral(X,Y) $=$ .463.
- LP-Maximal(X,Y) $=$ .443.



"*In practice, long tails seem more frequent than short*" Tukey (1960)



$$\widehat{\rho}(X, Y) = -\mathbf{0.021} \qquad \Rightarrow \qquad \widehat{\rho}(\log X, \log Y) = -\mathbf{.957}$$

Complete Reversal !

Model: $Y = X^{-\alpha}, \quad \alpha = 3$ and $\log X \sim \mathcal{N}(0, 1)$

Denuit and Dhaene (2003) " It is possible to have a random couple where the correlation is almost zero even though the components exhibit the *strongest kind of dependence possible* for this pair of marginals"

Theorem 1 (Characterization by Schweizer and Wolff 1981, AOS)

Any such Measure which is invariant under strictly monotonic increasing transformations has to be functional of copula density:

$$\int_{[0,1]^2} \Lambda(u,v) \; \mathrm{d}\,\mathrm{Cop}(u,v;X,Y)$$

$$\mathrm{LPINFOR}(X,Y) = \int_{[0,1]^2} \Big[\mathrm{cop}(u,v;X,Y) - 1\Big] \; \mathrm{d}\,\mathrm{Cop}(u,v;X,Y)$$

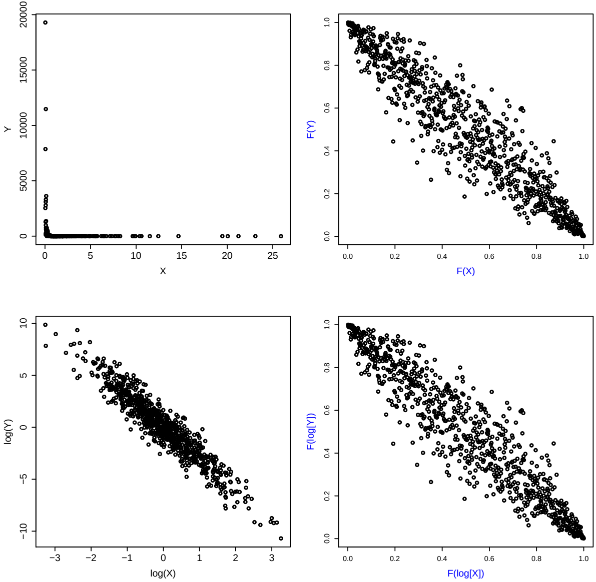
Denuit and Dhaene (2003) " It is possible to have a random couple where the correlation is almost zero even though the components exhibit the *strongest kind of dependence possible* for this pair of marginals"

Theorem 1 (Characterization by Schweizer and Wolff 1981, AOS)

Any such Measure which is invariant under strictly monotonic increasing transformations has to be functional of copula density:

$$\int_{[0,1]^2} \Lambda(u,v) \; \mathrm{d}\,\mathrm{Cop}(u,v;X,Y)$$

$$\mathrm{LPINFOR}(X,Y) = \int_{[0,1]^2} \Big[\mathrm{cop}(u,v;X,Y) - 1 \Big] \; \mathrm{d}\,\mathrm{Cop}(u,v;X,Y)$$



$$\widehat{\rho}\left[F(X;X),\ F(Y;Y)\right] = -.949$$

"We have a choice of the *distribution domain* and the *quantile domain* and that is good because we don't know in which domain nature will be most parsimonious " - Manny Parzen.

Correlation: Few Known Facts

1. We know $\rho(X, Y) = 0$. *what can we say about independence ?*

2. $\rho(X_1, Y_1) = \rho(X_2, Y_2)$. *What can we say nature of the dependence ?*

3. Let $\rho(X_1, Y_1) = .9$, $\rho(X_2, Y_2) = -.36$. *what can we say about the strength of dependence ?*

Theorem 2 (Shih and Huang (1992))

Fallacy of the reference interval $[-, 1, 1]$ *of* ρ*: Unless* $Y = aX + b$, $a \neq 0$, *the range of* ρ *is narrower than* $[-1, 1]$ *and depends on the marginal distributions.*

Example: (X, Y) bivariate standard lognormal, range of $\rho \in [-.368, 1]$.

Correlation: Few Known Facts

1. We know $\rho(X, Y) = 0$. *what can we say about independence* ?
2. $\rho(X_1, Y_1) = \rho(X_2, Y_2)$. *What can we say nature of the dependence* ?
3. Let $\rho(X_1, Y_1) = .9$, $\rho(X_2, Y_2) = -.36$. *what can we say about the strength of dependence* ?

Theorem 2 (Shih and Huang (1992))

Fallacy of the reference interval $[-,1,1]$ *of* ρ*: Unless* $Y = aX + b$, $a \neq 0$, *the range of* ρ *is narrower than* $[-1,1]$ *and depends on the marginal distributions.*

Example: (X, Y) bivariate standard lognormal, range of $\rho \in [-.368, 1]$.

Correlation: Few Known Facts

1. We know $\rho(X, Y) = 0$. *what can we say about independence ?*

2. $\rho(X_1, Y_1) = \rho(X_2, Y_2)$. *What can we say nature of the dependence ?*

3. Let $\rho(X_1, Y_1) = .9$, $\rho(X_2, Y_2) = -.36$. *what can we say about the strength of dependence ?*

Theorem 2 (Shih and Huang (1992))

Fallacy of the reference interval $[-, 1, 1]$ of ρ: Unless $Y = aX + b$, $a \neq 0$, the range of ρ is narrower than $[-1, 1]$ and depends on the marginal distributions.

Example: (X, Y) bivariate standard lognormal, range of $\rho \in [-.368, 1]$.

Desideratum

(D1.) Monotone Invariance, *Quantile domain copula based* measure.

(D2.) Range of the measure should *not* depend on marginal distributions.

(D3.) Nonlinear, Goal is to *completely* characterize (in)Dependence.

OPEN PROBLEM: (D1.) + (D2.) for Mixed X, Y.

Desideratum

(D1.) Monotone Invariance, *Quantile domain copula based* measure.

(D2.) Range of the measure should *not* depend on marginal distributions.

(D3.) Nonlinear, Goal is to *completely* characterize (in)Dependence.

OPEN PROBLEM: (D1.) + (D2.) for Mixed X, Y.

Science AAAS

Detecting Novel Associations in Large Data Sets

David N. Reshef,[1,2,3*†] Yakir A. Reshef,[2,4*†] Hilary K. Finucane,[5] Sharon R. Grossman,[2,6] Gilean McVean,[3,7] Peter J. Turnbaugh,[6] Eric S. Lander,[2,8,9] Michael Mitzenmacher,[10‡] Pardis C. Sabeti[2,6‡]

Identifying interesting relationships between pairs of variables in large data sets is increasingly important. Here, we present a measure of dependence for two-variable relationships: the maximal information coefficient (MIC). MIC captures a wide range of associations both functional and not, and for functional relationships provides a score that roughly equals the coefficient of determination (R^2) of the data relative to the regression function. MIC belongs to a larger class of maximal information-based nonparametric exploration (MINE) statistics for identifying and classifying relationships. We apply MIC and MINE to data sets in global health, gene expression, major-league baseball, and the human gut microbiota and identify known and novel relationships.

MIC : Maximal Information Coefficient

Finding correlations in big data

A new statistical method called MIC can find diverse types of correlations in large data sets. *Nature Biotechnology* asked eight experts to weigh in on its utility.

In today's era of large data sets, statistical methods that facilitate exploratory analyses to detect patterns and generate hypotheses are critical to progress in biology. Last year, David Reshef and colleagues published a new approach to such analysis, called maximal information criteria or MIC (*Science* 334, 1518–1524, 2011). *Nature Biotechnology* solicited comments from several practitioners versed in data-intensive biological research. Their responses not only highlight the appeal of methods like MIC for biological research, but also raise some important reservations as to its widespread use and statistical power.

What is MIC?

Gustavo Stolovitzky: MIC is a quantity between 0 and 1 that measures the association between a pair of variables. If one variable deterministically dictates the value of another variable, the MIC coefficient will be 1. But if noise influences the relationship, the MIC will be smaller than 1, with larger amounts of noise resulting in a bigger deviation of MIC from 1.

Gustavo Stolovitzky, manager, Functional Genomics and Systems Biology, IBM Computational Biology Center, IBM, Yorktown Heights, New York.

Peng Qiu: When examining many potential pair-wise associations between variables measured in a large data set, MIC produces a ranked list of pairs ordered by the strength of associations.

Peng Qiu, assistant professor, Department of Bioinformatics and Computational Biology, University of Texas MD Anderson Cancer Center, Houston, Texas.

Why is this approach important?

Eran Segal: The main attraction of MIC is its ability to identify diverse types of relationships between variables without a priori favoring one relationship over the other. In contrast, commonly used statistics, such as the Pearson correlation, are either limited in the relationships they can identify (e.g., linear) or they favor certain types of relationships.

Eran Segal, associate professor, Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel.

Bill Noble: Methods for detecting nonlinear relationships are most useful in the context of exploratory knowledge discovery from large data sets, when the structure of the data itself is not yet well understood.

Bill Noble, professor, Department of Genome Sciences, Department of Computer Science and Engineering, University of Washington, Seattle, Washington.

PQ: The most appealing property of MIC is 'equitability', meaning that it is equally sensitive to different types of association relationships (linear and nonlinear).

How important is 'equitability'?

PQ: Equitability ensures that the top of the list of ranked associations is not dominated or biased by certain types of associations. As a simple comparison, if Pearson correlation is used to rank-order pair-wise associations, the top of the list will be dominated by linear relationships, and many types of nonlinear associations may receive insignificant rankings.

GS: The MIC equitability property ensures that the least noisy, associated pairs will be ranked higher regardless of the specific nature of the association, followed by pairs whose noise strength increases as we go down the list.

When might equitability be useful?

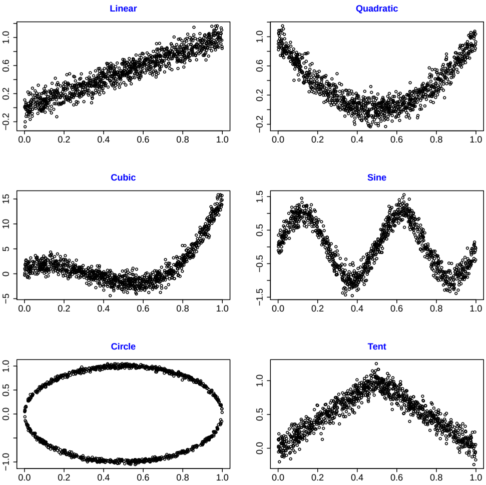
Olga Troyanskaya: The most common situation is in analyzing a collection of data sets, where many different types of pairwise relationships might occur. Another reason to use such metrics is when comparing measurements at different biological levels of control (e.g., epigenetic and transcriptional regulation) or between different biomolecules (e.g., proteins and metabolites), especially if there's a reason to suspect highly nonlinear relationships between these variables.

Olga Troyanskaya, associate professor, Department of Computer Science and the Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey.

ES: MIC might be able to identify relationships that are the result of a superposition of functions. This is particularly useful in uncovering relationships that are determined by multiple distinct factors, for example, in the regulation of gene expression.

What types of biological data would be amenable to MIC analysis?

ES: MIC could be applied to analyze the relationship between the expression of a repressor and its target gene, as the dependence of both on global cellular resources such as polymerases enforces a positive relationship, whereas a negative relationship is enforced by their specific interaction. Thus, the positive and negative relationships are superimposed to produce a more complex relationship.



- Given data : $X_1, X_2, \ldots, X_p$.
- Discover *unknown* relationship.
- Characterize Independence
- Effect Size Estimation
- **MIC**: Maximal Information Coefficient.

Intellectual Debate & Criticism: Harvard, Stanford, Berkeley, Columbia

“Power Deficiency” $+$ “Robustness”

Desideratum

(D1.) Monotone Invariance, *Quantile domain copula based* measure.

(D2.) Range of the measure should *not* depend on marginal distributions.

(D3.) Nonlinear, Goal is to *completely* characterize (in)Dependence.

(D4.) *Robust*, stability under slight departures from assumptions.

John Tukey (1962): "*That remains to us, to our willingness to take up the rocky road of real problems in preference to the smooth road of unreal assumptions, arbitrary criteria, and abstract results without real attachments. Who is for the challenge?*"

Empirical Power comparison

$$y = f(x) + \textbf{NOISE}, \quad X \sim U[0,1].$$

Setting 1. Under Gaussian error (σ), $\epsilon \sim \mathcal{N}(0, \sigma^2)$, $\sigma \uparrow 3$.

Setting 2. Proposition of outliers(η), $\eta \uparrow .4$

$$\epsilon \sim (1-\eta)\mathcal{N}(0,1) + \eta\mathcal{N}(\mu, 3), \text{ where } \mu = \pm 5 \text{ w.p } 1/2$$

Setting 3. Level of 'bad' leverage points (η), $\eta \uparrow .4$

$$\epsilon \sim (1-\eta)\mathcal{N}(0,1) + \eta\mathcal{N}(\mu, 3), \text{ where } \mu \in \{\pm 20, \pm 30\} \text{ w.p } 1/4.$$

Setting 4. Heavy-Tailed error (σ), $\epsilon \sim \text{Cauchy}(0, \text{Scale} = \sigma)$, $\sigma \uparrow 2$.

Empirical Power comparison

$$y = f(x) + \textbf{NOISE}, \quad X \sim U[0,1].$$

Setting 1. Under Gaussian error (σ), $\epsilon \sim \mathcal{N}(0, \sigma^2)$, $\sigma \uparrow 3$.

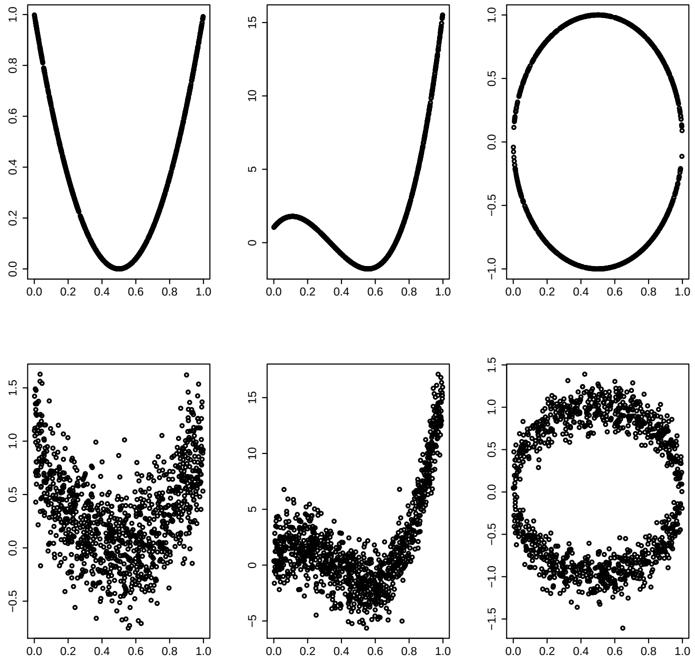
Setting 2. Proposition of outliers(η), $\eta \uparrow .4$

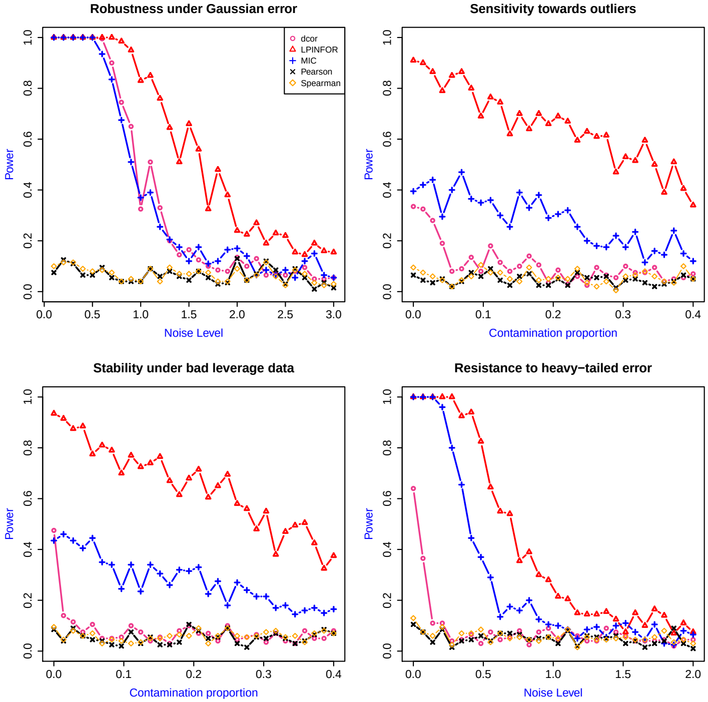
$$\epsilon \sim (1-\eta)\mathcal{N}(0,1) + \eta\mathcal{N}(\mu, 3), \text{ where } \mu = \pm 5 \text{ w.p } 1/2$$

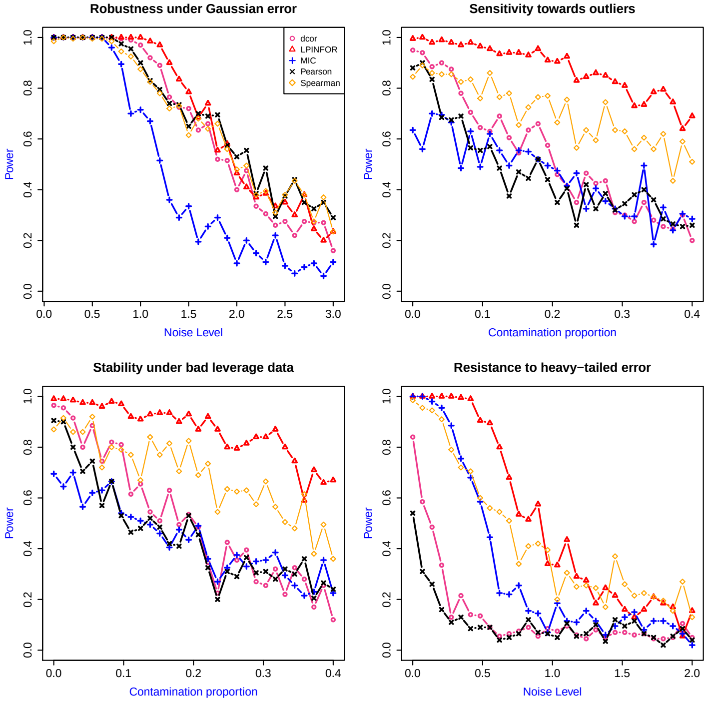
Setting 3. Level of 'bad' leverage points (η), $\eta \uparrow .4$

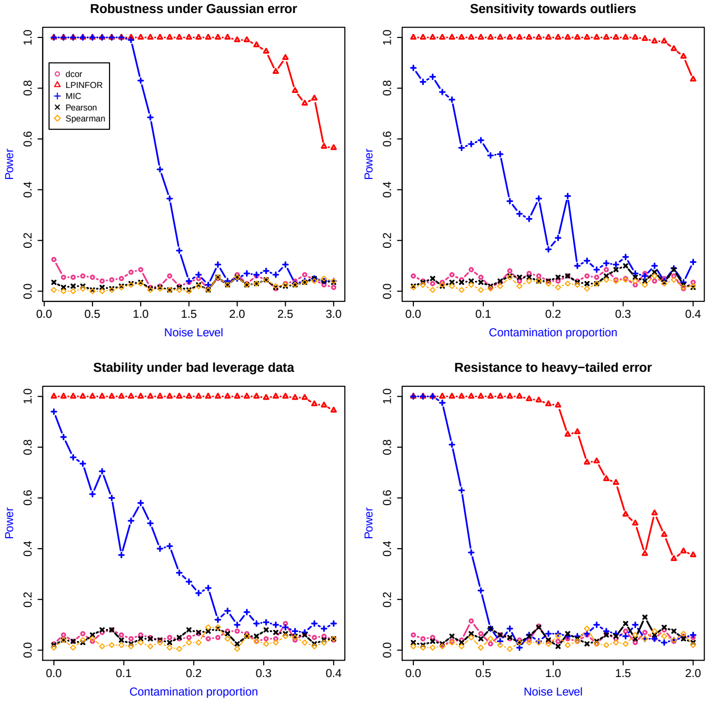
$$\epsilon \sim (1-\eta)\mathcal{N}(0,1) + \eta\mathcal{N}(\mu, 3), \text{ where } \mu \in \{\pm 20, \pm 30\} \text{ w.p } 1/4.$$

Setting 4. Heavy-Tailed error (σ), $\epsilon \sim \text{Cauchy}(0, \text{Scale} = \sigma)$, $\sigma \uparrow 2$.




Robustness under Gaussian error
Sensitivity towards outliers
Stability under bad leverage data
Resistance to heavy−tailed error
dcor
LPINFOR
MIC
Pearson
Spearman
Power
Noise Level
Contamination proportion
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.5
2.0
2.5
3.0
0.1
0.3


Robustness under Gaussian error
Sensitivity towards outliers
Stability under bad leverage data
Resistance to heavy−tailed error
dcor
LPINFOR
MIC
Pearson
Spearman
Power
Noise Level
Contamination proportion


Robustness under Gaussian error
Sensitivity towards outliers
Stability under bad leverage data
Resistance to heavy−tailed error
dcor
LPINFOR
MIC
Pearson
Spearman
Power
Noise Level
Contamination proportion
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.5
2.0
2.5
3.0
0.1
0.3

Computational Complexity

(D5.) Computationally fast, scalable to large data sets.

$(X, Y) \sim Uniform[0, 1]^2$, *samples size* n*, averaged over* 100 *runs (in secs).*

Methods	Size of the data sets			
	$n = 1000$	$n = 2500$	$n = 5000$	$n = 10,000$
LPD$^+$	0.002 (.005)	0.005 (.007)	0.011 (.007)	**0.018** (.007)
Dcor*	0.371 (.013)	1.773 (.450)	7.756 (.810)	**44.960** (12.64)
MIC†	1.357 (.035)	5.584 (.110)	19.052 (.645)	**65.170** (2.54)

$+$ Written in R.
$*$ Written in C.
$\dagger$ Written in Java.

What Happens for Bivariate Normal ?

Theorem 3

For (X, Y) bivariate normal LPINFOR *is an increasing function of $|\rho|$:*

$$\text{LPINFOR}(X, Y) = \frac{\rho^2}{1-\rho^2}.$$

Trick: Switch basis to compute the integral representation of LPINFOR.

Lemma 1 (Eagleson, 1964 and Koziol, 1979)

Let $H_j(x)$ denotes the jth orthonormalized Hermite polynomial, then :

$$\int_{[0,1]^2} \frac{\phi[\Phi^{-1}(u;X), \Phi^{-1}(v;Y);\rho]}{\phi[\Phi^{-1}(u;X)]\,\phi[\Phi^{-1}(v;Y)]}\, H_j[\Phi^{-1}(u)]\, H_k[\Phi^{-1}(v)]\, \mathrm{d}u\, \mathrm{d}v = \delta_{jk}\rho^j.$$

Rényi Axioms (1959)

1. LPD(X, Y) invariant under Strictly Monotone Transformations.

2. LPD$(X, Y) = 0$ if and only if X and Y are independent.

3. LPD$(X, Y) = \text{LPD}(Y, X)$.

4. For $(X, Y; \rho)$ Bivariate Normal LPD(X, Y) is $\Psi(|\rho|)$, $\Psi \uparrow$.

5. LPD(X, Y) is *defined for any* X, Y (any combination of discrete and continuous).

Desideratum: (D6.) Satisfies Rényi Axioms.

Discrete Marginals: 2 by 2 Contingency Table

Is there any relationship between taking aspirin and risk of heart attack ?

Drug	**Male Heart Attack**	
	Yes	No
Aspirin	104	10,933
Placebo	189	10,845

Theorem 4 (Connection with Pearson ϕ-coefficient.)

For (X, Y) forming a 2×2 contingency table we have,

$$\mathrm{LPINFOR}(X, Y) = \phi^2(X, Y).$$

Discrete Marginals: 2 by 2 Contingency Table

Is there any relationship between taking aspirin and risk of heart attack ?

Drug	**Male Heart Attack**	
	Yes	No
Aspirin	104	10,933
Placebo	189	10,845

Theorem 4 (Connection with Pearson ϕ-coefficient.)

For (X, Y) forming a 2×2 contingency table we have,

$$\text{LPINFOR}(X, Y) = \phi^2(X, Y).$$

Discrete Marginals: R by C Contingency Table

Is there any association between Hair color and Eye color ?

Eye Color	**Hair Color**				
	Fair	Red	Medium	Dark	Black
Blue	326	38	241	110	3
Light	688	116	584	188	4
Medium	343	84	909	412	26
Dark	98	48	403	681	85

Theorem 5 (Connection with Chi-squared Measure)

For (X, Y) forming a $R \times C$ contingency table we have,

$$\mathrm{LPINFOR}(X, Y) = \mathrm{CHIDIV}(X, Y).$$

Large Sparse Matrix ?

OPEN PROBLEM: Tackle $p \gg n$ case, Modified smooth-χ^2, Classical approaches fails miserably.

"Despite the 35 years that have passed since the emergence of the loglinear model literature, the interest in succinct parametric models for large sparse contingency tables remains" - Fienberg, 2003

Mixed Marginals

OPEN PROBLEM: To Date No Copula based Method exists.

Example: Y discrete and X continuous. LP-Compression.

Theorem 6 (Two Sample Case)

Nonparametric Wilcoxon rank statistics has the following LP-Comoment based representation

$$\mathrm{Wilcoxon(X, Y)} \ = \ \mathsf{LP}[1, 1; \mathrm{X}, \mathrm{Y}].$$

High order Wilcoxon statistics are LP comoments of high order

$$\mathsf{LP}(k, 1; X, Y) \ = \ \mathbb{E}[T_k(X; X) T_1(Y; Y)]$$

Mixed Marginals

OPEN PROBLEM: To Date No Copula based Method exists.

Example: Y discrete and X continuous. LP-Compression.

Theorem 6 (Two Sample Case)

Nonparametric Wilcoxon rank statistics has the following LP-Comoment based representation

$$\mathrm{Wilcoxon}(\mathrm{X},\mathrm{Y}) \;=\; \mathsf{LP}[1,1;\mathrm{X},\mathrm{Y}].$$

High order Wilcoxon statistics are LP comoments of high order

$$\mathsf{LP}(k,1;X,Y) \;=\; \mathbb{E}[T_k(X;X)T_1(Y;Y)]$$

Desideratum

(D1.) Monotone Invariance, *Quantile domain copula based* measure.

(D2.) Range of the measure should *not* depend on marginal distributions.

(D3.) Nonlinear, Goal is to *completely* characterize (in)Dependence.

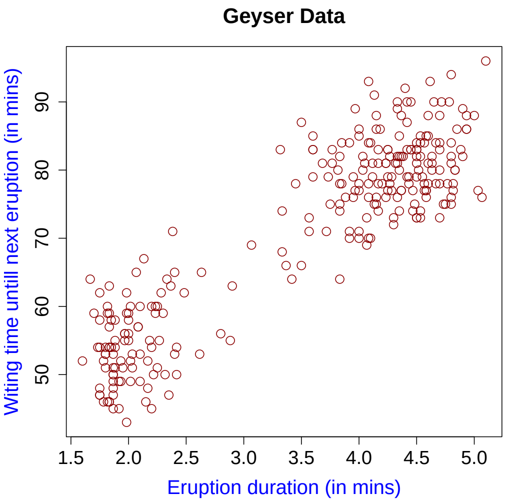
(D4.) *Robust*, stability under slight departures from assumptions.

(D5.) Computationally fast, scalable to large data sets.

(D6.) Satisfy Rényi Axioms.

(D7.) Simultaneously *extend and integrate* classical and novel statistical methods, compute them systematically under a unified framework.

Omnibus to Directional Measure. Linear Correlation ?



[Terence Speed, IMS Bulletin **15**, March 2012]

$$\mathrm{LP}\left[X, Y\right] = \begin{bmatrix} \mathbf{0.781} & \mathbf{-0.190} & -0.128 & \mathbf{0.208} \\ -0.181 & \mathbf{0.291} & 0.037 & -0.039 \\ -0.136 & 0.052 & 0.169 & -0.018 \\ \mathbf{0.189} & -0.095 & 0.042 & 0.108 \end{bmatrix}$$

Linearity Coefficient: $\widehat{\mathrm{LP}}[1,1;X,Y]^2/\,\mathrm{LPINFOR} = .65.$

How to Test ? Behaviour of $\widehat{\mathrm{LP}}[j,k;X,Y]$

Theorem 7 (Asymptotic Normality)

Under the H_0, the sample LP-comoments (by expressing them as functional of copula process) are i.i.d and

$$\sqrt{n}\,\widehat{\mathrm{LP}}[j,k;X,Y] \sim \mathcal{N}(0,1).$$

$$\mathrm{LP}\left[X,Y\right] = \begin{bmatrix} \mathbf{0.781} & -\mathbf{0.190} & -0.128 & \mathbf{0.208} \\ -0.181 & \mathbf{0.291} & 0.037 & -0.039 \\ -0.136 & 0.052 & 0.169 & -0.018 \\ \mathbf{0.189} & -0.095 & 0.042 & 0.108 \end{bmatrix}$$

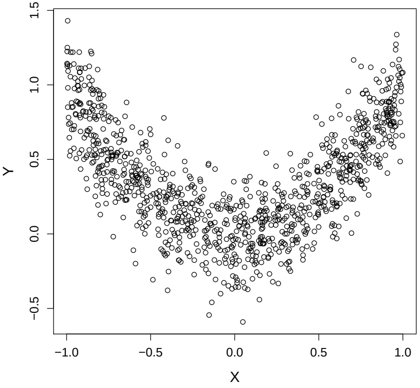
Linearity Coefficient: $\widehat{\mathrm{LP}}[1,1;X,Y]^2/\,\mathrm{LPINFOR} = .65.$

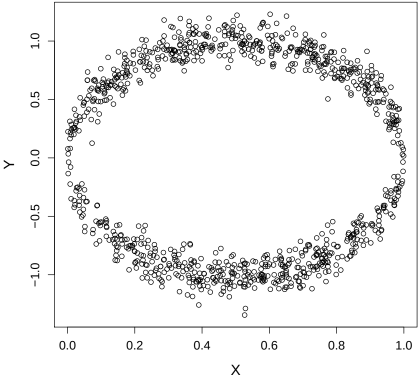
How to Test ? Behaviour of $\widehat{\mathrm{LP}}[j,k;X,Y]$

Theorem 7 (Asymptotic Normality)

Under the H_0, the sample LP-comoments (by expressing them as functional of copula process) are i.i.d and

$$\sqrt{n}\,\widehat{\mathrm{LP}}[j,k;X,Y] \sim \mathcal{N}(0,1).$$





$$\begin{bmatrix} 0.005 & 0.010 & -0.042 & -0.100 \\ \mathbf{0.831} & 0.201 & -0.023 & -0.020 \\ 0.009 & -0.034 & -0.038 & -0.019 \\ 0.144 & 0.365 & 0.148 & 0.035 \end{bmatrix} \qquad \begin{bmatrix} -0.011 & -0.062 & -0.007 & 0.014 \\ 0.024 & \mathbf{-0.767} & -0.033 & 0.121 \\ 0.020 & 0.077 & 0.018 & 0.032 \\ -0.033 & 0.216 & 0.043 & 0.253 \end{bmatrix}$$

LP-Optimal Transformation

- Nonparametrically identifies the optimal nonlinear transformations

$$\rho\big(\Psi^*(Y),\, \phi^*(X)\big) \;=\; \max_{\Psi,\phi} \rho\big(\Psi(Y),\, \phi(X)\big).$$

- Optimal Transformation (Breiman and Friedman, 1985).
- Contrast the Simplicity of ACE-Algorithm and LP-Algorithm.
- Does Optimal Transformation always exist ?

Theorem 8 (Existence)

The optimal LP-Transformation exists if the copula density is square-integrable,

$$\iint_{[0,1]^2} \mathrm{cop}^2(u, v; X, Y)\, \mathrm{d}u\, \mathrm{d}v \;<\; \infty.$$

LP-Optimal Transformation

- Nonparametrically identifies the optimal nonlinear transformations

$$\rho\big(\Psi^*(Y),\, \phi^*(X)\big) \;=\; \max_{\Psi,\phi} \rho\big(\Psi(Y),\, \phi(X)\big).$$

- Optimal Transformation (Breiman and Friedman, 1985).

- Contrast the Simplicity of ACE-Algorithm and LP-Algorithm.

- Does Optimal Transformation always exist ?

Theorem 8 (Existence)

The optimal LP-Transformation exists if the copula density is square-integrable,

$$\iint_{[0,1]^2} \mathrm{cop}^2(u, v; X, Y)\, \mathrm{d}u\, \mathrm{d}v \;<\; \infty.$$

LP-Smooth Regression for Mixed Covariates

Can we produce Nonparametric Smoother to model the dependence ?

Simultaneously: Detecting $\rightarrow$ Exploring $\rightarrow$ Modeling

Theorem 9 (LP-Regression Expansion)

Nonparametric nonlinear regression is equivalent to conditional expectation $\mathbb{E}[Y|X = x = Q(u;X)]$*, which can be approximated by a linear combination of orthonormal score functions* $T_j(X;X)$ *satisfying:*

$$\mathbb{E}\big[Y \mid X = x\big] - \mathbb{E}[Y] = \sum_j T_j(x;X)\ \mathrm{LP}(j,0;X,Y).$$

LP-Smooth Regression for Mixed Covariates

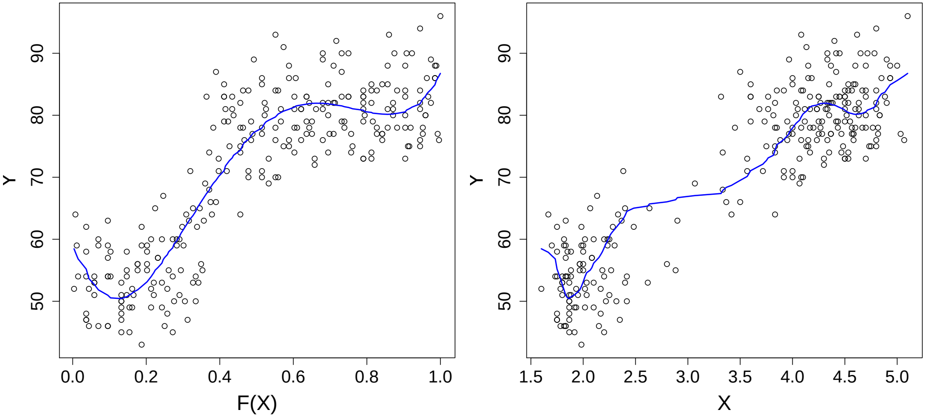
Can we produce Nonparametric Smoother to model the dependence ?

Simultaneously: Detecting $\rightarrow$ Exploring $\rightarrow$ Modeling

Theorem 9 (LP-Regression Expansion)

Nonparametric nonlinear regression is equivalent to conditional expectation $\mathbb{E}[Y|X = x = Q(u;X)]$, *which can be approximated by a linear combination of orthonormal score functions* $T_j(X;X)$ *satisfying:*

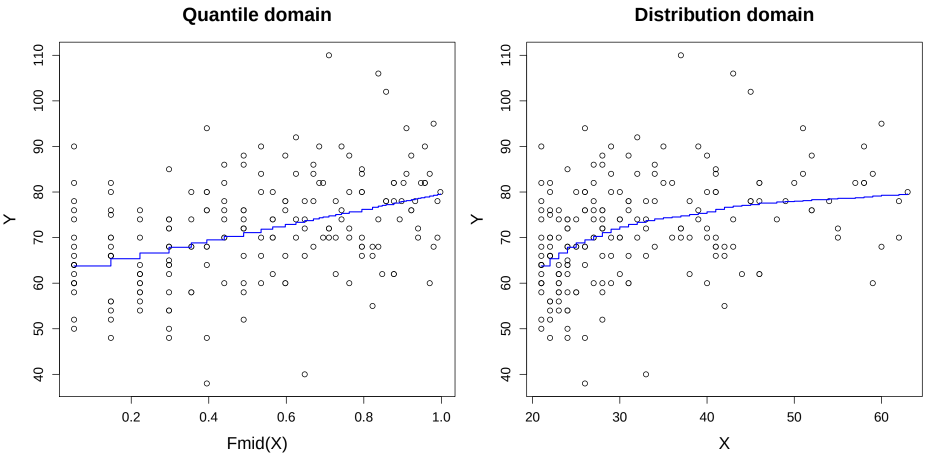
$$\mathbb{E}\big[Y \mid X = x\big] - \mathbb{E}[Y] = \sum_j T_j(x;X)\ \mathrm{LP}(j,0;X,Y).$$



Theorem 10

With probability 1, $\mathbb{E}[Y|X] = \mathbb{E}[Y|F(X;X)]$

X Continuous



Theorem 11

With probability 1, $\mathbb{E}[Y|X] = \mathbb{E}[Y|F(X;X)] = \mathbb{E}[Y|F^{\mathrm{mid}}(X;X)]$.

X Discrete

LP-Comoment based Nonparametric Dependence Modeling

(D1.) Monotone Invariance, *Quantile domain copula based* measure.

(D2.) Range of the measure should *not* depend on marginal distributions.

(D3.) Nonlinear, Goal is to *completely* characterize (in)Dependence.

(D4.) *Robust*, stability under slight departures from assumptions.

(D5.) Computationally fast, scalable to large data sets.

(D6.) Satisfy Rényi Axioms.

(D7.) Simultaneously *extend and integrate* classical and novel statistical methods, compute them systematically under a unified framework.

(D8.) Insight into the *nature of dependence*, Directional measure that yields smooth Regression automatically.

LP Mixed Data Science: Bridge the "Disconnect"

Theory: *How can we create a framework for awareness of many cultures of Statistical Science based on Mixed Data that could facilitate cross-fertilization between our profession and the Big Data Science ?*

Practice: *How to develop a Systematic Data Modeling Strategy ? How to design Flexible and Reusable algorithms based on General Theory that can be Adapted to solve specific Practical Problems ?*

Teaching: *How to create a concrete educational plan for broadly preparing a data science workforce of trained Next-Generation Statisticians which is equally applicable for introductory and advance level ?*

“To earn more, learn more, and believe that learning a lot (answering all related questions) is easier than learning little (answering only the questions asked)” — Manny Parzen

——— Thanks.

Multivariate LP-Coherence Matrix

1. Dependence Between Random Vectors: $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^{n \times q}$.

2. Define: $\mathrm{COH}(X, Y) \; = \; \Sigma_{XX}^{-1} \, \Sigma_{XY} \, \Sigma_{YY}^{-1} \, \Sigma_{YX}$.

3. LP-Coherence Dependence Measure: $\mathrm{Trace}\big[\, \mathrm{COH}(TX, \; TY) \,\big]$.

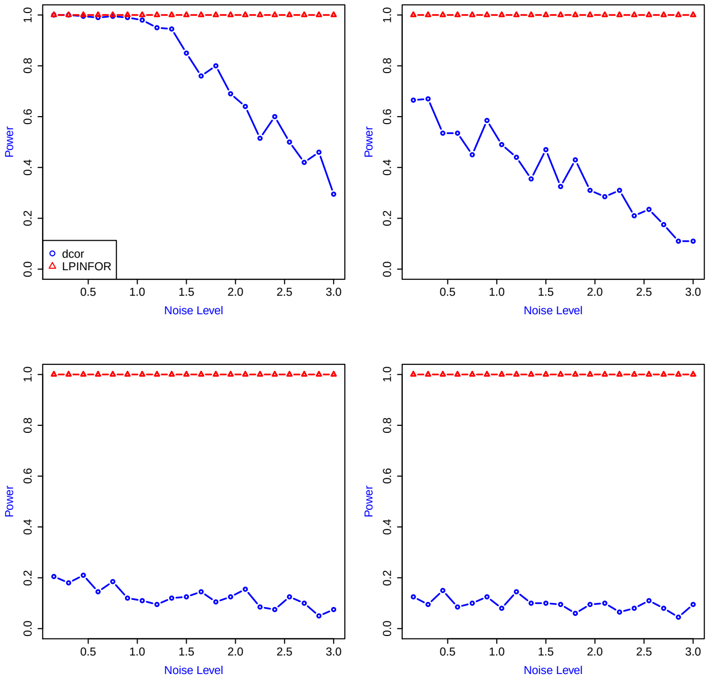
4. Unifying Univariate and Multivariate. For X, Y Univariate Verify that:

$$\mathrm{LPINFOR}(X, Y) \;\; = \;\; \mathrm{Trace}\big[\, \mathrm{COH}(TX, \; TY) \,\big].$$

5. Works for Mixed Multivariate Random Variables. Nonlinear, Robust.

6. Empirical Power Investigation: $p = q = \big\{5, 10, 25, 50\big\}$. Székely 2007

$$Y_k \; = \; \log(X_k^2) + \epsilon_k, \;\; \epsilon_k \sim \mathcal{N}(0, \sigma^2), \;\; \sigma \uparrow 3.$$


Power
Noise Level
dcor
LPINFOR
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.0
1.5
2.0
2.5
3.0