Equitability and the Maximal Information Coefficient

David Reshef, Yakir Reshef

CSAIL, MIT / HST, Harvard

Columbia University - May 2, 2014
Outline

Motivation

Equitability

The maximal information coefficient

Equitability in practice

Choosing a measure of dependence
A simple data exploration problem

- $D$ is a sample of size $n$ drawn from a $k$-dimensional distribution $\Pi$. 
A simple data exploration problem

- $D$ is a sample of size $n$ drawn from a $k$-dimensional distribution $\Pi$.
- Many variable pairs.
  - Can only examine a limited number.
A simple data exploration problem

- $D$ is a sample of size $n$ drawn from a $k$-dimensional distribution $\Pi$.
- Many variable pairs.
  - Can only examine a limited number.
- What are the strongest (bivariate) associations in $\Pi$?
Hypothesis testing vs. effect size

▶ How to deal with ambiguity of question?
Hypothesis testing vs. effect size

- How to deal with ambiguity of question?
- One possibility: test for independence
  - +: clear, broad objective; many good, well-characterized methods
  - -: in a dataset with 1 million non-trivial dependencies, we need a smaller list!
Hypothesis testing vs. effect size

▷ How to deal with ambiguity of question?
▷ One possibility: test for independence
  ▷ +: clear, broad objective; many good, well-characterized methods
  ▷ -: in a dataset with 1 million non-trivial dependencies, we need a smaller list!
▷ Alternatively: define ”strongest” and measure an effect size
  ▷ +: gives a list that can be narrowed down
  ▷ -: what will we be missing?
A question

Can we have our cake and eat it too?
A question

Can we have our cake and eat it too?
Equitability - definition

- $\hat{\varphi}: \mathbb{R}^{2n} \rightarrow [0, 1]$ a measure of dependence
- i.e., $\varphi = 0$ iff independence
Equitability - definition

- $\hat{\varphi} : \mathbb{R}^{2n} \rightarrow [0, 1]$ a measure of dependence
  - i.e., $\varphi = 0$ iff independence
- $Q = \{Z_\theta : \theta \in \Theta\}$ a set of standard relationships
  - Ex: $\theta$ may specify a function, a type of noise, and an amount of noise
Equitability - definition

- $\hat{\varphi} : \mathbb{R}^{2n} \rightarrow [0, 1]$ a measure of dependence
  - i.e., $\varphi = 0$ iff independence
- $Q = \{Z_\theta : \theta \in \Theta\}$ a set of standard relationships
  - Ex: $\theta$ may specify a function, a type of noise, and an amount of noise
- $\Phi : Q \rightarrow [0, 1]$ a property of interest
  - $\Phi$ captures our sense of relationship strength
Equitability - definition

- $\hat{\phi} : \mathbb{R}^{2n} \rightarrow [0, 1]$ a measure of dependence
  - i.e., $\varphi = 0$ iff independence
- $Q = \{ Z_\theta : \theta \in \Theta \}$ a set of standard relationships
  - Ex: $\theta$ may specify a function, a type of noise, and an amount of noise
- $\Phi : Q \rightarrow [0, 1]$ a property of interest
  - $\Phi$ captures our sense of relationship strength
- Equitability $\approx$ how much does $\hat{\phi}$ tell us about $\Phi$ on $Q$?
Equitability - intuition

e.g. $Q$ is the set of functional relationships of the form $(X + N_x, f(X) + N_y)$
Equitability is evaluated by analogy to estimation theory

- Treat $\hat{\phi}$ as “estimator” of $\Phi$, look at “confidence intervals” (interpretable intervals). Equitability = small intervals.
Equitability - other methods

0 iff Independence

- HSIC
- dCor
- Mutual Information
- Maximal Correlation
- RDC

Effect Size

- Pearson (on linear rels.)
- Nonparametric regression (on functional rels.)
- Smoothing splines (on functional rels.)
Equitability vs power and parameter estimation
Equitability vs power and parameter estimation
Equitability vs power and parameter estimation
Defining MIC - preliminaries

- \((X, Y)\), jointly distributed random variables on \([0, 1] \times [0, 1]\).
- For grid \(G\), let \((X, Y)|_G = (\text{col}_G(X), \text{row}_G(Y))\)
- \(G(k, \ell) = \text{all } k\text{-by}-\ell \text{ grids}\)

\[
I^*((X, Y), k, \ell) = \max_{G \in G(k, \ell)} I((X, Y)|_G)
\]
Defining MIC

The characteristic matrix of \((X, Y)\), denoted by \(M(X, Y)\), is defined by

\[
M(X, Y)_{k,\ell} = \frac{I^*((X, Y), k, \ell)}{\log \min\{k, \ell\}}
\]

for \(k, \ell > 1\).

The maximal information coefficient (MIC) of \((X, Y)\) is defined by

\[
\text{MIC}(X, Y) = \sup M(X, Y)
\]
Elementary properties of MIC

- \( \text{MIC}(X, Y) = 0 \) iff \( X \) and \( Y \) are statistically independent.
- \( \text{MIC}(f(X), g(Y)) = \text{MIC}(X, Y) \) for monotonic functions \( f \) and \( g \).
- \( \text{MIC}(X, f(X)) = 1 \) for never-constant \( f \).
  - Holds also for superpositions of functional relationships (e.g. circle).
Estimating MIC

Thm: \( \text{MIC} = \sup \) over boundary of \( M(X, Y) \), rather than all of \( M(X, Y) \). This gives:

- A provably consistent, efficiently computable estimator of MIC.
  - Faster than the original statistic.
- An algorithm for approximating MIC of a given pdf to arbitrary precision.
Equitability in practice

The equitability of a measure of dependence can depend on:

- Model $Q = \{Z_\theta : \theta \in \Theta\}$ of *standard relationships on which can define what we mean by “noise”*
Equitability in practice

The equitability of a measure of dependence can depend on:

- Model \( Q = \{ Z_\theta : \theta \in \Theta \} \) of *standard relationships on which can define what we mean by “noise”*
  - \( \theta \) may parameterize set of relationships, type of noise, etc.
  - e.g. \( Q = \) *noisy functional relationships*
Equitability in practice

The equitability of a measure of dependence can depend on:

- Model $Q = \{ Z_\theta : \theta \in \Theta \}$ of standard relationships on which can define what we mean by “noise”
  - $\theta$ may parameterize set of relationships, type of noise, etc.
  - e.g. $Q =$ noisy functional relationships
- Choice of $\Phi : Q \rightarrow [0, 1]$, the property of interest that quantifies the noise in those relationships (e.g. $R^2$)
Equitability in practice

The equitability of a measure of dependence can depend on:

- Model \( Q = \{ Z_\theta : \theta \in \Theta \} \) of standard relationships on which can define what we mean by “noise”
  - \( \theta \) may parameterize set of relationships, type of noise, etc.
  - e.g. \( Q = \) noisy functional relationships
- Choice of \( \Phi : Q \to [0, 1] \), the property of interest that quantifies the noise in those relationships (e.g. \( R^2 \))
- Available estimator of the measure of dependence, \( \hat{\phi} \) (e.g. bias, variance)
Equitability in practice

The equitability of a measure of dependence can depend on:

- Model $Q = \{Z_\theta : \theta \in \Theta\}$ of standard relationships on which can define what we mean by “noise”
  - $\theta$ may parameterize set of relationships, type of noise, etc.
  - e.g. $Q =$ noisy functional relationships
- Choice of $\Phi : Q \rightarrow [0, 1]$, the property of interest that quantifies the noise in those relationships (e.g. $R^2$)
- Available estimator of the measure of dependence, $\hat{\phi}$ (e.g. bias, variance)

Given fixed $\Phi$, would like a $\hat{\phi}$ that works for as large a choice of $Q$ as possible.
Equitability in practice - an example to get us going

- Linear+Periodic, Low Freq (1000)
- Linear+Periodic, High Freq (1000)
- Linear+Periodic, High Freq 2 (1000)
- Linear+Periodic, Medium Freq (1000)
- Linear+Periodic, Medium Freq (500)
- Non-Fourier Freq [Low] Cosine (1000)
- Non-Fourier Freq [Low] Cosine (250)
- Cosine, High Freq (1000)
- Cosine, High Freq (500)
- Cubic (1000)
- Cubic, Y-Stretched (1000)
- L-Shaped (1000)
- Exponential [2^x] (1000)
- Exponential [10^x] (1000)
- Line (1000)
- Parabola (1000)
- Random (1000)
- Non-Fourier Freq [Low] Sine (1000)
- Sine, Low Freq (250)
- Sine, High Freq (250)
- Sine, High Freq (1000)
- Sigmoid (1000)
- Varying Freq [Medium] Cosine (1000)
- Varying Freq [Medium] Sine (1000)
- Varying Freq [Medium] Sine (500)
- Spike (1000)
- Lopsided L-Shaped (1000)
- Lopsided L-Shaped (500)
Equitability in practice

\[ Q_1 = \{(X, f(X) + N_y) : (X, f(X)) \text{ uniform on } f\} \]
Equitability in practice

\[ Q_2 = \{(X + N_x, f(X)) : (X, f(X)) \text{ uniform on } f\} \]
Equitability in practice

\[ Q_3 = \{(X + N_x, f(X) + N_y) : (X, f(X)) \text{ uniform on } f\} \]
Equitability in practice

\[ Q_4 = \{(X, f(X) + N_y) : (X, f(X)) \text{ uniform on } X\} \]
Equitability in practice

\[ Q_5 = \{(X + N_x, f(X)) : (X, f(X)) \text{ uniform on } X\} \]
Equitability in practice

Q_6 = \{(X + N_x, f(X) + N_y) : (X, f(X)) \text{ uniform on } X\}

Outline
Motivation
Equitability
The maximal information coefficient
Equitability in practice
Choosing a measure of dependence
## Equitability in practice – summary

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Model</th>
<th>Noise (Gaussian)</th>
<th>Maximal Corr. (ACE)</th>
<th>dCor</th>
<th>HSIC</th>
<th>$I_{L^2}$ (Kraskov)</th>
<th>Pearson</th>
<th>RDC</th>
<th>MIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 250$</td>
<td>Even Along $f(X)$</td>
<td>$Y$-Noise</td>
<td>0.58</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>Even Along $X$</td>
<td>$Y$-Noise</td>
<td>0.53</td>
<td>1.00</td>
<td>1.00</td>
<td>0.90</td>
<td>1.00</td>
<td>1.00</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>Even Along $f(X)$</td>
<td>$XY$-Noise</td>
<td>0.63</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>Even Along $X$</td>
<td>$XY$-Noise</td>
<td>0.65</td>
<td>1.00</td>
<td>1.00</td>
<td>0.88</td>
<td>1.00</td>
<td>1.00</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>Even Along $f(X)$</td>
<td>$X$-Noise</td>
<td>0.63</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>Even Along $X$</td>
<td>$X$-Noise</td>
<td>0.65</td>
<td>1.00</td>
<td>1.00</td>
<td>0.90</td>
<td>1.00</td>
<td>1.00</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>Worst Case</td>
<td></td>
<td>0.65</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.55</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>Even Along $f(X)$</td>
<td>$Y$-Noise</td>
<td>0.53</td>
<td>1.00</td>
<td>1.00</td>
<td>0.75</td>
<td>1.00</td>
<td>1.00</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>Even Along $X$</td>
<td>$Y$-Noise</td>
<td>0.48</td>
<td>1.00</td>
<td>1.00</td>
<td>0.48</td>
<td>1.00</td>
<td>1.00</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Even Along $f(X)$</td>
<td>$XY$-Noise</td>
<td>0.58</td>
<td>1.00</td>
<td>1.00</td>
<td>0.98</td>
<td>1.00</td>
<td>1.00</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>Even Along $X$</td>
<td>$XY$-Noise</td>
<td>0.63</td>
<td>1.00</td>
<td>1.00</td>
<td>0.85</td>
<td>1.00</td>
<td>1.00</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Even Along $f(X)$</td>
<td>$X$-Noise</td>
<td>0.58</td>
<td>1.00</td>
<td>1.00</td>
<td>0.95</td>
<td>1.00</td>
<td>1.00</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>Even Along $X$</td>
<td>$X$-Noise</td>
<td>0.63</td>
<td>1.00</td>
<td>1.00</td>
<td>0.88</td>
<td>1.00</td>
<td>1.00</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Worst Case</td>
<td></td>
<td>0.63</td>
<td>1.00</td>
<td>1.00</td>
<td>0.98</td>
<td>1.00</td>
<td>1.00</td>
<td>0.48</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>Even Along $f(X)$</td>
<td>$Y$-Noise</td>
<td>0.45</td>
<td>0.98</td>
<td>1.00</td>
<td>0.08</td>
<td>1.00</td>
<td>1.00</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Even Along $X$</td>
<td>$Y$-Noise</td>
<td>0.40</td>
<td>0.98</td>
<td>1.00</td>
<td>0.08</td>
<td>1.00</td>
<td>1.00</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Even Along $f(X)$</td>
<td>$XY$-Noise</td>
<td>0.48</td>
<td>0.98</td>
<td>1.00</td>
<td>0.33</td>
<td>1.00</td>
<td>1.00</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>Even Along $X$</td>
<td>$XY$-Noise</td>
<td>0.53</td>
<td>0.98</td>
<td>1.00</td>
<td>0.50</td>
<td>1.00</td>
<td>1.00</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>Even Along $f(X)$</td>
<td>$X$-Noise</td>
<td>0.48</td>
<td>0.98</td>
<td>1.00</td>
<td>0.38</td>
<td>1.00</td>
<td>1.00</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>Even Along $X$</td>
<td>$X$-Noise</td>
<td>0.53</td>
<td>0.98</td>
<td>1.00</td>
<td>0.58</td>
<td>1.00</td>
<td>1.00</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>Worst Case</td>
<td></td>
<td>0.53</td>
<td>0.98</td>
<td>1.00</td>
<td>0.58</td>
<td>1.00</td>
<td>1.00</td>
<td>0.38</td>
</tr>
</tbody>
</table>
Tradeoffs in choosing a measure of dependence

- **Computational budget**: how fast must analysis be?
Tradeoffs in choosing a measure of dependence

- **Computational budget**: how fast must analysis be?
- **Sample size**: can you afford to care about anything more than power (e.g. effect size)?
Tradeoffs in choosing a measure of dependence

- **Computational budget**: how fast must analysis be?
- **Sample size**: can you afford to care about anything more than power (e.g. effect size)?
- **Anticipated relationship strength**: searching for weak dependencies only, or stronger ones too?
Tradeoffs in choosing a measure of dependence

- **Computational budget**: how fast must analysis be?
- **Sample size**: can you afford to care about anything more than power (e.g. effect size)?
- **Anticipated relationship strength**: searching for weak dependencies only, or stronger ones too?
- **Heterogeneity of relationships**: looking for relationships of just one kind, or many kinds?
Tradeoffs in choosing a measure of dependence

- **Computational budget**: how fast must analysis be?
- **Sample size**: can you afford to care about anything more than power (e.g. effect size)?
- **Anticipated relationship strength**: searching for weak dependencies only, or stronger ones too?
- **Heterogeneity of relationships**: looking for relationships of just one kind, or many kinds?
- **Dimensionality**: can you manually follow up on all non-trivial relationships?
Tradeoffs in choosing a measure of dependence

- **Computational budget**: how fast must analysis be?
- **Sample size**: can you afford to care about anything more than power (e.g. effect size)?
- **Anticipated relationship strength**: searching for weak dependencies only, or stronger ones too?
- **Heterogeneity of relationships**: looking for relationships of just one kind, or many kinds?
- **Dimensionality**: can you manually follow up on all non-trivial relationships?
- **Noise type**: What kind of noise do you have, and how well can you characterize it?
When is equitability useful

**Anticipated relationship strength**

- Only low signal
- Mixed / High
- No (focus on maximizing power)
When is equitability useful

- **Anticipated relationship strength**
  - Only low signal
  - Mixed / High

- **Heterogeneity of relationships**
  - Low
  - Medium / High

- No (focus on maximizing power)
- No (use parametric methods)
When is equitability useful

- **Anticipated relationship strength**
  - Only low signal
  - No (focus on maximizing power)
  - Mixed / High

- **Heterogeneity of relationships**
  - Low
  - No (use parametric methods)
  - Medium / High

- **Dimensionality of data**
  - Low
  - No (manually examine all results?)
  - Medium / High
When is equitability useful

- **Anticipated relationship strength**
  - Only low signal
  - Mixed / High
    - No (focus on maximizing power)

- **Heterogeneity of relationships**
  - Low
  - Medium / High
    - No (use parametric methods)

- **Dimensionality of data**
  - Low
  - Medium / High
    - No (manually examine all results?)

- **Sample size**
  - Very low
  - Medium
  - Very high
    - No (focus on maximizing power)

- **MIC**
  - Med.
When is equitability useful

- Anticipated relationship strength
  - Only low signal
  - Mixed / High
- Heterogeneity of relationships
  - Low
  - Medium / High
- Dimensionality of data
  - Low
  - Medium / High
- Sample size
  - Very low
  - Med.
  - Very high
- Computational budget
  - MIC
  - Low
  - Medium / High
When is equitability useful

- **Anticipated relationship strength**
  - Only low signal
  - Mixed / High
    - No
      - (focus on maximizing power)

- **Heterogeneity of relationships**
  - Low
    - No
      - (use parametric methods)
  - Medium / High

- **Dimensionality of data**
  - Low
    - No
      - (manually examine all results?)
  - Medium / High
    - No
      - (focus on maximizing power)

- **Sample size**
  - Very low
    - No
      - (focus on maximizing power)
  - Med.
    - No
      - (focus on maximizing power)
  - Very high
    - No
      - (focus on maximizing power)

- **Computational budget**
  - Low
    - No
      - (focus on maximizing power)
  - Medium / High

- **Noise type**
  - Well-characterized
    - Y-noise
    - No
    - MIC
  - Unknown / not-characterized
    - MIC
When is equitability useful

- **Anticipated relationship strength**
  - Low
  - Mixed / High
  - (focus on maximizing power)

- **Heterogeneity of relationships**
  - Low
  - Medium / High
  - (use parametric methods)

- **Dimensionality of data**
  - Low
  - Medium / High
  - (manually examine all results?)

- **Sample size**
  - Very low
  - Med.
  - MIC
  - Very high
  - Computational budget
    - Low
    - Medium / High
    - Noise type
      - Well-characterized
      - Unknown / not-characterized
      - Y-noise
      - MIC

- **Equitability**

David Reshef, Yakir Reshef: Equitability and the Maximal Information Coefficient

CSAIL, MIT / HST, Harvard
Acknowledgments

- Michael Mitzenmacher
- Pardis Sabeti
- Hilary Finucane
- Shari Grossman
- Peter Turnbaugh
- Gil McVean
- Eric Lander
Kinney and Atwal’s critique
Kinney and Atwal’s critique
Kinney and Atwal’s critique
Equitability vs power

- Measures of dependence traditionally focus on maximizing power at rejecting statistical independence (i.e. $\Phi = 0$)
Equitability vs power

- Measures of dependence traditionally focus on maximizing power at rejecting \textit{statistical independence} (i.e. $\Phi = 0$)
- A statistic with good worst-case interpretability has good power at rejecting any null hypothesis of the form $\Phi \leq \Phi_0$
  - This allows for relationship ranking by $\Phi$. 
Equitability vs power

- Measures of dependence traditionally focus on maximizing power at rejecting \textit{statistical independence} (i.e. $\Phi = 0$).
- A statistic with good worst-case interpretability has good power at rejecting any null hypothesis of the form $\Phi \leq \Phi_0$.
  - This allows for relationship ranking by $\Phi$.
  - If we want strongest relationships, we “don’t care” about relationships with a low value of $\Phi$. 

David Reshef, Yakir Reshef: Equitability and the Maximal Information Coefficient

CSAIL, MIT / HST, Harvard
### Equitability vs parameter estimation

<table>
<thead>
<tr>
<th></th>
<th>Estimating $\theta$ (confidence)</th>
<th>Estimating $\Phi$ (interpretability)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td>Only one distribution for each value of the estimand</td>
<td>Multiple distributions for each value of the estimand</td>
</tr>
<tr>
<td><strong>Error</strong></td>
<td>Confidence intervals wide due to finite sample effects</td>
<td>Interpretable intervals wide due to finite sample effects, as well as infinite sample relationship between $\varphi$ and $\Phi$</td>
</tr>
<tr>
<td><strong>Tests</strong></td>
<td>Small confidence intervals give power at rejecting $H_0 : \theta &lt; \theta_0$</td>
<td>Small interpretable intervals give power at rejecting $H_0 : \Phi &lt; \theta_0$</td>
</tr>
</tbody>
</table>
MIC and mutual information

- Without normalization, we get $\sup_G I((X, Y)|G) = I(X, Y)$
- Normalization “groups together” grids of similar complexity, subjects them to same regularization.
- This makes MIC a continuous function of the pdf of $(X, Y)$.
  - Normalization used is the “minimal” one necessary to achieve this.
  - In particular, mutual information not continuous.
Equitability on more complicated relationships

Increasing Noise

- Line
- Two Lines
- Line & Parabola
- Ellipse
Equitability on more complicated relationships
Equitability on more complicated relationships