Discovering & deciphering relationships across disparate data modalities

Motivation:
↳ "Understand" the relationship between physical (brain) & mental properties
↳ Q1: are the 2 related at all?
↳ Q2: how are they related?

Stats Background:
↳ X is random variable (some measurement)
↳ F_X is distribution of X
↳ $F_X(a) = P(X \leq a)$
↳ $X \sim F_X$
↳ For 2 random variables X & Y, F_{XY} is called the joint distribution
↳ $F_{XY}(a,b) = P(X \leq a \text{ \& } Y \leq b)$
↳ $(X, Y) \sim F_{XY}$

Independence
↳ X & Y are independent if neither contains info. about the other
↳ $F_{XY} = P(X \leq a \text{ \& } Y \leq b) = P(X \leq a) \times P(Y \leq b) = F_X F_Y$

↳ $F_{XY} = F_X F_Y$

Informal Definition of Hypothesis Testing

↳ Null Hypothesis: conventional belief about a phenomenon of interest (H_0)

↳ Alternative Hypothesis: alternate belief about the same phenomenon (H_A)

↳ p-value: probability (under the null) of measurements more extreme than what was observed

Formal Definition of Independence Testing:

↳ $(X_i, Y_i) \sim F_{XY} = F_{X|Y} F_Y, \; i \in 1, \ldots, n$

↳ $H_0: F_{XY} = F_X F_Y$

↳ $H_A: F_{XY} \neq F_X F_Y$

Intuition

Intuitive Desiderata of Testing Procedure

1. Performant under any joint distribution
   ↳ low & high dimensional
   ↳ Euclidean & structured data (eg. sequence, images, networks, shapes)
   ↳ linear & nonlinear relationships

2. Reveals the "geometry" of dependence
3. Is computationally efficiency

↳ Provides a tractable algorithm that addresses the 2 motivating questions
  ↳ Q1: are the two related at all?
  ↳ Q2: how are they related?

Correlation coefficient

↳ $$r_{xy}^2 = \frac{\left(\sum_{i=1}^{n}(x-\bar{x})(y_i-\bar{y})\right)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2} \Bigg\}*$$

Mantel correlation coefficient

↳ *

↳ $$d_{xy}^2 = \frac{\left(\sum_{i,j=1}^{n}(x_i-x_j)(y_i-y_j)\right)^2}{\sum_{i,j=1}^{n}(x_i-x_j)^2 \sum_{i,j=1}^{n}(y_i-y_j)^2} \Bigg\}**$$

Generalized correlation coefficient

↳ *

↳ **

↳ $$C_{xy}^2 = \frac{\left(\sum_{i,j=1}^{n}\sigma_x(x_i,x_j)\,\sigma_y(y_i,y_j)\right)^2}{\sum_{i,j=1}^{n}\sigma_x(x_i,x_i)^2 \sum_{i,j=1}^{n}\sigma_y(y_i,y_i)^2}$$

Local distance correlation

Multiscale distance correlation

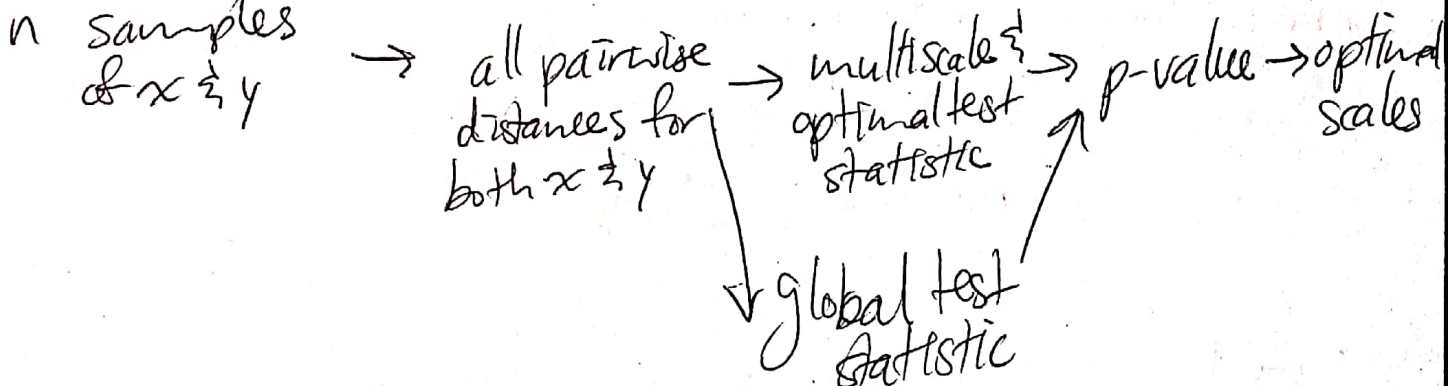
↳ compute local dcorr @ all scales

↳ find scale w/ max smoothed test statistic

↳ permutation test to determine p-value

Multiscale Generalized Correlation (MGC)

↳



Simulations

↳ 20 diff. functions

Definitions

↳ power is the probability of rejecting the null when the alternative is true

↳ $\beta_n(t)$ = power of ~~approach~~ test statistic t given n samples

↳ relative power: power of one approach minus power of another

↳ $\beta_n(mgc) - \beta_n(t)$