Correlation and Independence

Shuhei Mano

ISM

Detchar tetecon

June 24, 2014

Correlation

The Pearson's product-moment correlation is

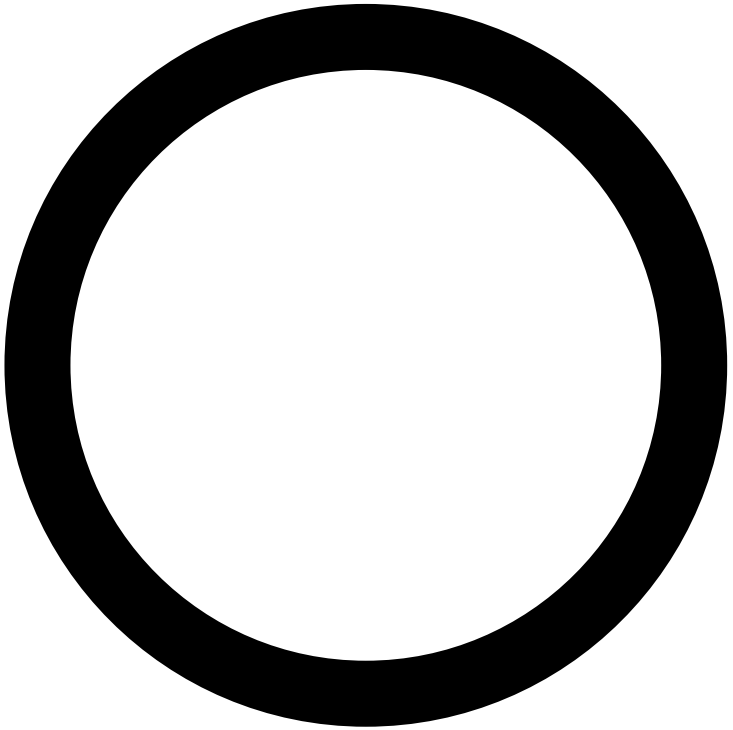
$$\rho_{X,Y} := \frac{\mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]}{\sqrt{\mathbb{E}[(X - \mathbb{E}X)^2]\mathbb{E}[(Y - \mathbb{E}Y)^2]}}.$$

- $-1 \leq \rho_{X,Y} \leq 1$ and $\rho_{X,Y} = \pm 1 \Leftrightarrow Y = cX + d$, a.s. for some const. c, d.
- $X \perp\!\!\!\perp Y \Rightarrow \rho_{X,Y} = 0$, but the reverse does not always hold.

We might want to reject $X \perp\!\!\!\perp Y$, possibly non-linear relationship.

("association" is general but ambiguous.)

Example



$$(X, Y) = (\cos(\Theta), \sin(\Theta)), \qquad \Theta \sim (0, 2\pi].$$

Y is non-linearly determined by X, but $\rho_{X,Y} = 0$.

Classics: nonparametric methods

Let R_i and S_i, $1 \leq i \leq n$ be rank of the i-th sample point in X and Y, respectively.

Spearman's rank correlation (1904):

$$\rho_{Spearman} := 1 - \frac{6 \sum_i (R_i - S_i)^2}{n(n^2 - 1)}.$$

Kendall rank correlation (1938):

$$\tau := \frac{N_1 - N_2}{n(n-1)/2},$$

where N_1 is the number of concordant pairs in X and Y, and N_2 is the number of discordant pairs.

Independence

How to test independence?

Here, I will introduce methods based on

- divergence
- characteristic function
- kernel methods

Divergence

Kullback-Leibler divergence between prob. measures P and Q:

$$D(P||Q) = \mathbb{E}_P\left[\log \frac{P}{Q}\right].$$

$D(P||Q) = 0 \Leftrightarrow P = Q$.

Let $P = P(X, Y)$ and $Q = P(X)P(Y)$, mutual information:

$$I(X, Y) := \mathbb{E}_P\left[\log \frac{P(X, Y)}{P(X)P(Y)}\right].$$

$X \perp\!\!\!\perp Y \Leftrightarrow I(X, Y) = 0$, so test $I(X, Y) = 0$.

Remark

- c.f. Independent Component Analysis
- Power divergence:

$$\mathbb{E}_P\left\{\left[\frac{P(X, Y)}{P(X)P(Y)}\right]^{\lambda} - 1\right\}.$$

Characteristic Function

Let $\phi_{X,Y}$, ϕ_X and ϕ_Y be characteristic func. of (X, Y), $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$.

$X \perp\!\!\!\perp Y \Leftrightarrow \phi_{X,Y} = \phi_X \phi_Y$, so test $||\phi_{X,Y} - \phi_X \phi_Y||^2 = 0$, where $||\cdot||$ is a distance.

Székely et al. (2007, Ann. Stat. 35: 2769) proposal:

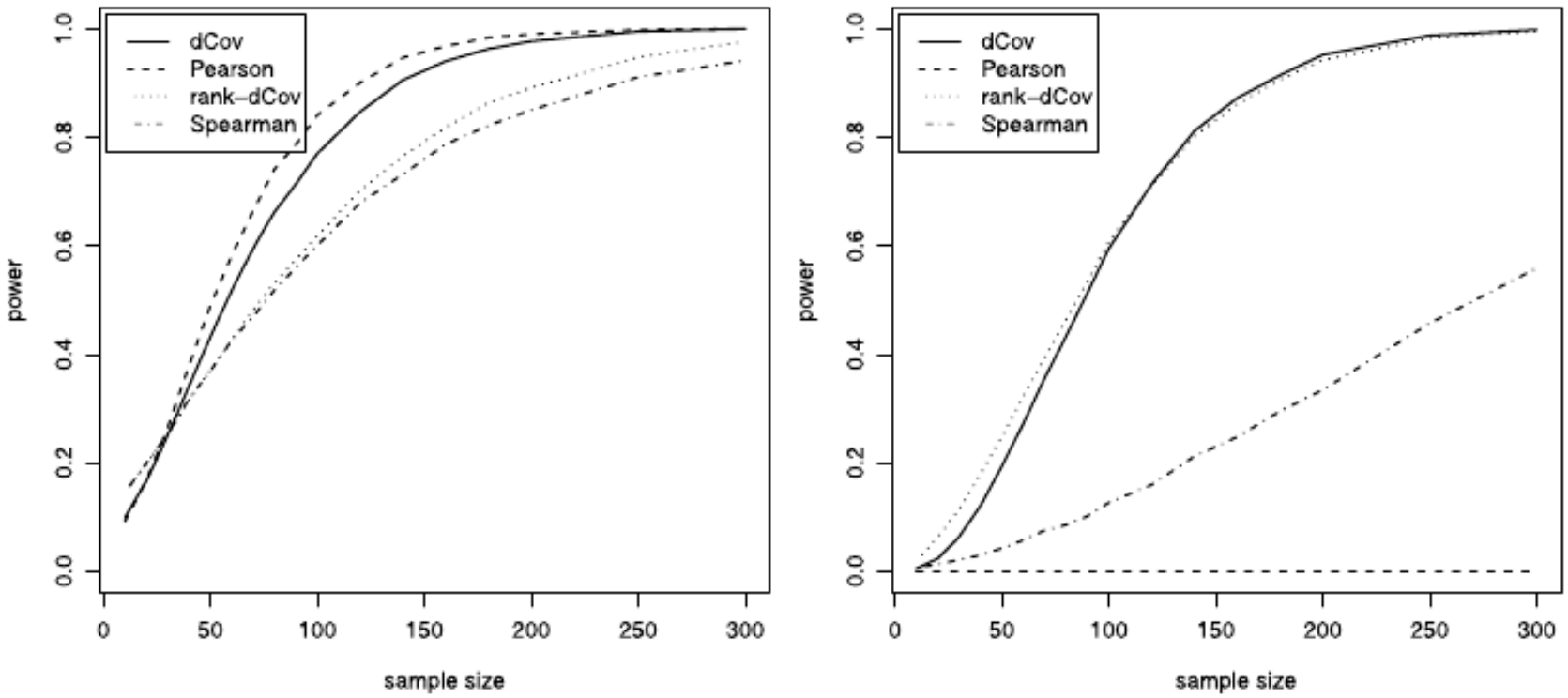
$$\mathcal{V}^2(X, Y) := ||\phi_{X,Y} - \phi_X \phi_Y||_w^2 = \int |\phi_{X,Y}(t, s) - \phi_X(t)\phi_Y(s)|^2 w(t, s) dtds,$$

with $w(t, s) \propto t^{-1-p} s^{-1-q}$.

Distance correlation:

$$R^2(X, Y) := \frac{\mathcal{V}^2(X, Y)}{\sqrt{\mathcal{V}^2(X)\mathcal{V}^2(Y)}}.$$

Experiments



$$f(x, y) = [(1 + \theta x)(1 + \theta y)] \exp(-x - y - \theta xy),\ x, y > 0,\ 0 \leq \theta \leq 1.$$

Székely & Rizzo (2009) Ann. Appl. Stat. 3: 1236.

Kernel Methods

Reproducing Kernel Hilbert Space: A positive definite kernel k uniquely determine a Hilbert space $\mathcal{H}_k$ with

- $k(\cdot, x) \in \mathcal{H}_k, \forall x \in \mathcal{X}$.
- Reproducing property:

$$\langle f(\cdot), k(\cdot, x)\rangle_{\mathcal{H}_k} = f(x), \qquad \forall x \in \mathcal{X}, \forall f \in \mathcal{H}_k.$$

Remark

- Let $x \mapsto \Phi(x) = k(\cdot, x)$. Inner product in $\mathcal{H}_k$ is a Grammian:

$$\langle \Phi(x_i), \Phi(x_j)\rangle_{\mathcal{H}_k} = k(x_i, x_j),$$

  so we do not know explicit form of mapping, possibly non-linear.
- c.f. Support Vector Machine

Cross-Covariance Operator

Expectation operator:

$$m_X := \mathbb{E}[\Phi(X)] = \mathbb{E}[k(\cdot, X)].$$

- $\langle m_X, f \rangle = \mathbb{E}[f(X)]$. (c.f. characteristic func.).
- Estimator: $\hat{m}_X^{(n)} = \frac{1}{n} \sum_{i=1}^{n} k(\cdot, X_i)$.

Cross-Covariance operator:

- $\forall f \in \mathcal{H}_X, \forall g \in \mathcal{H}_Y$,

$$\langle \Sigma_{Y,X} f, g \rangle = \mathbb{E}[f(X)g(Y)] - \mathbb{E}[f(X)]\mathbb{E}[g(Y)],$$

- $X \perp\!\!\!\perp Y \Leftrightarrow \Sigma_{Y,X} = O$, so test $\|\hat{\Sigma}_{Y,X}\|^2_{Hilbert-Schmidt} = 0$ (Gretton et al. 2005 16th ALT).

Randomized Dependence Coefficient

Hirschfeld-Gebelein-Rényi maximum correlation coefficient:

$$\rho_{X,Y}^{HGR} = \sup_{f,g} \rho_{f(X),g(Y)}$$

with $X \perp\!\!\!\perp Y \Leftrightarrow \rho_{X,Y}^{HGR} = 0$. DRC is an "approximation" of $\rho_{X,Y}^{HGR}$ (Lopez-Paz et al. NIPS 2013):

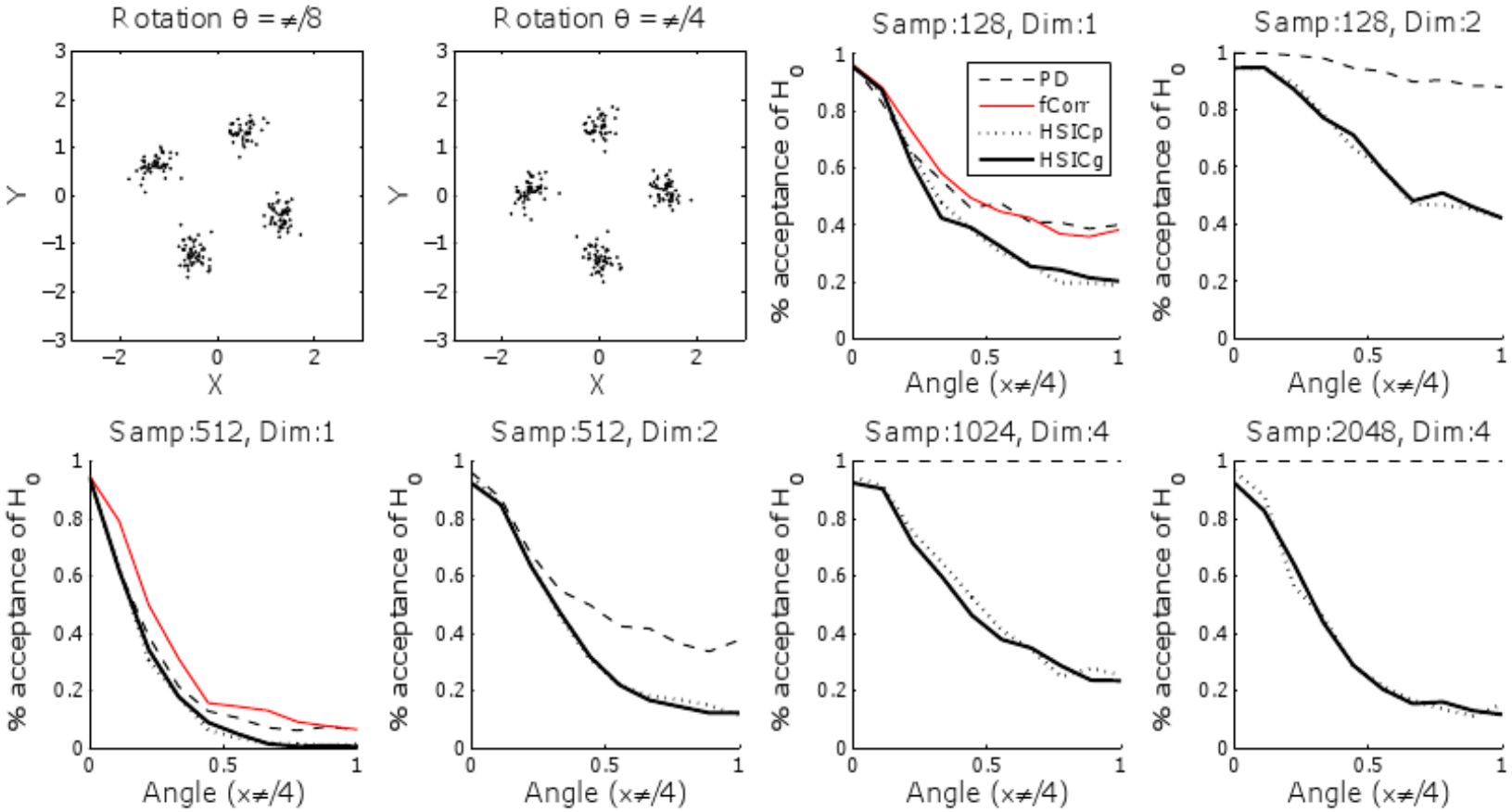
$$\max_{a,b} \rho_{a'\Phi(F_X(X)),b'\Phi(F_Y(Y))},$$

(c.f. canonical correlation), with

$$(\Phi(X))_{i,j} = \Phi(w_i' x_j + b_i), \qquad 1 \leq i \leq k,\ 1 \leq j \leq n,$$

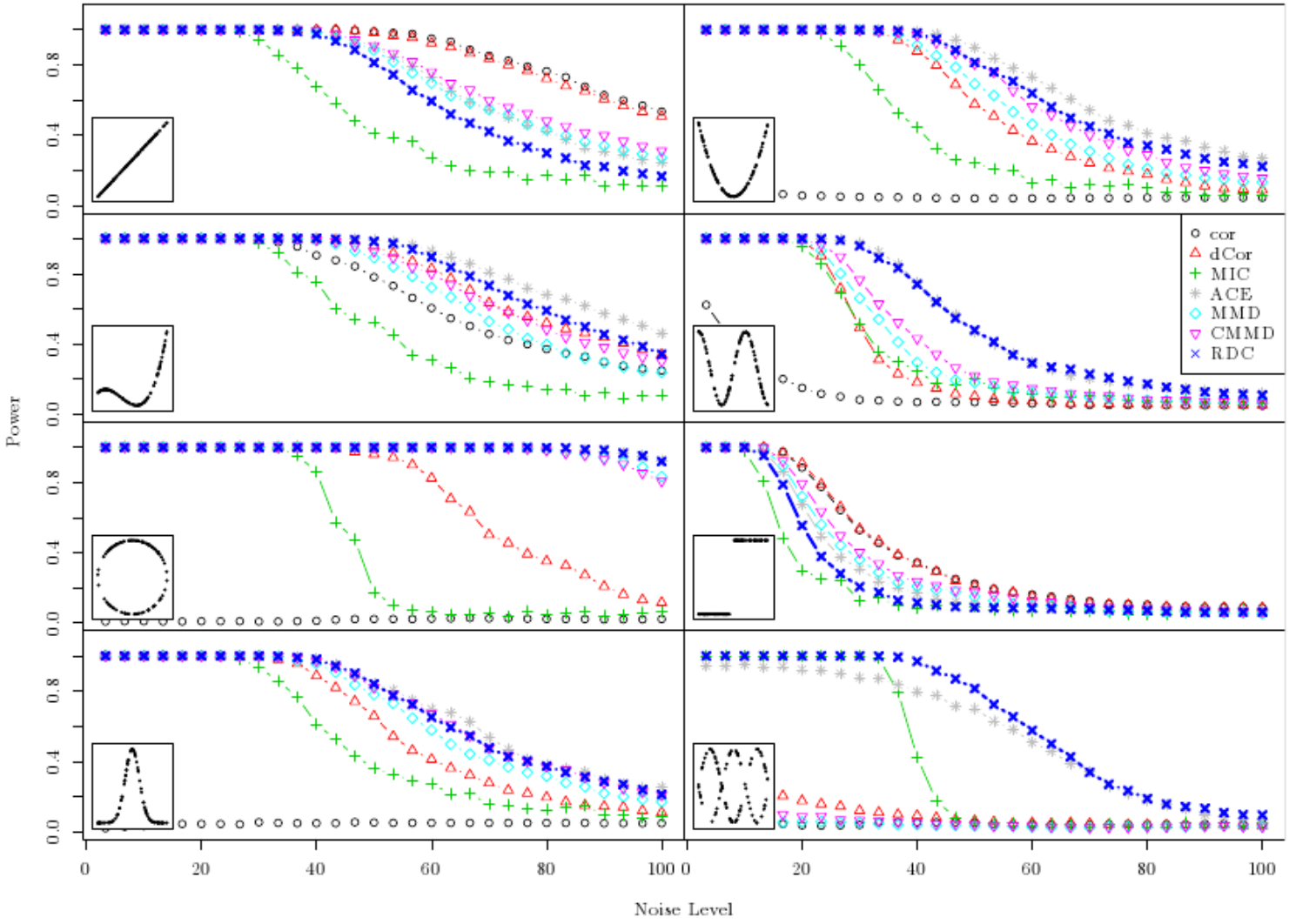
where w_i and b_i are random variables.

Experiments



Gretton et al. 2008, Advances in NIPS 20, 585.

Experiments



Lopez-Paz et al. NIPS 2013.