A conditional distribution function based approach to design nonparametric tests of independence and conditional independence

Sohan Seth and José C. Príncipe

Computational NeuroEngineering Lab, University of Florida, Gainesville

Introduction

- **Definition** (X, Y) are said to be conditionally independent (CI) given Z i.e. $X \perp Y|Z$ if Y does **not** contain any additional information about X other than that contained in Z i.e.

$$P(X < x|Y = y, Z = z) = P(X < x|Z = z)\, \forall (x, y, z).$$

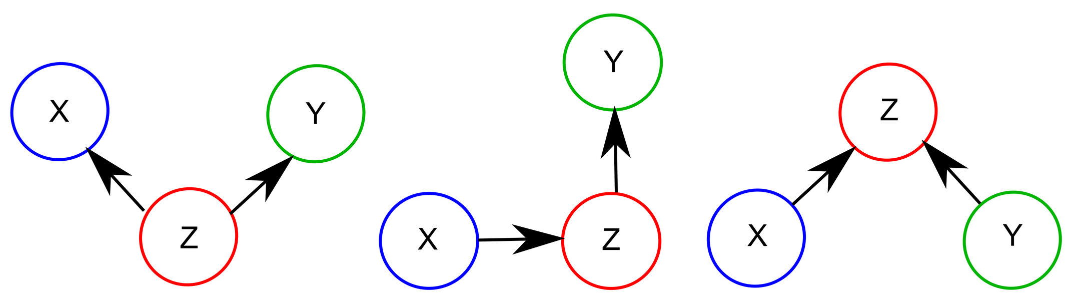


FIGURE 1: Graphical representation of CI: True, True and False

Example 1 Umbrella and Sun are conditionally independent given Rain.

- **Applications** Causal inference and feature selection.

Example 2 $\{X_t\}$ does not **Granger cause** $\{Y_t\}$ if the present value Y_t of $\{Y_t\}$ is CI of the past values $[X_{t-1}, X_{t-2}, \ldots]$ of $\{X_t\}$ given past values $[Y_{t-1}, Y_{t-2}, \ldots]$ of $\{Y_t\}$.

Example 3 Given set of features $\mathcal{S}$, m features $\mathcal{S}_m \subset \mathcal{S}$ are **redundant** if the target Y is CI of $\mathcal{S}_m$ given the rest of the features $\mathcal{S}\backslash\mathcal{S}_m$.

Measure of conditional independence (CI)

- **Measure** of CI := a statistic that attains 0 if and only CI is satisfied e.g,

$$\mathcal{M}^2_{CI} = \int (P(X < u|Y = v, Z = w) - P(X < u|Z = w))^2 \mathrm{d}F_X(u)\mathrm{d}F_{YZ}(v, w).$$

- **Estimator**

$$\hat{\mathcal{M}}^2_{CI} = \int (\hat{P}(X < u|Y = v, Z = w) - \hat{P}(X < u|Z = w))^2 \mathrm{d}\hat{F}_X(u)\mathrm{d}\hat{F}_{YZ}(v, w)$$

where $\hat{}$ denotes estimates.

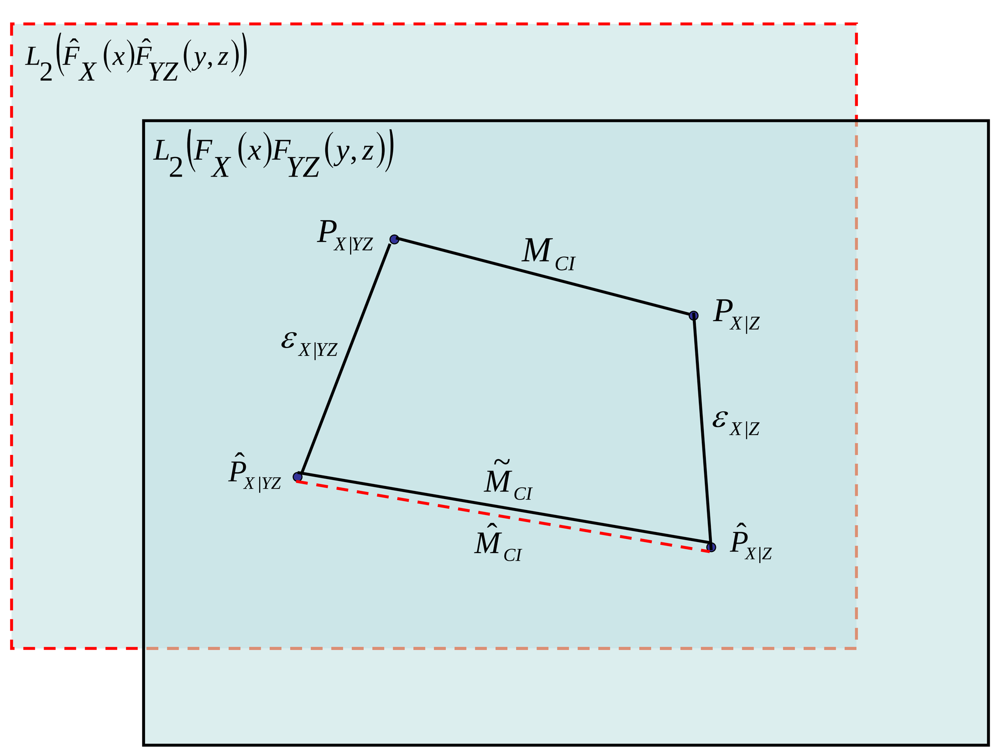


FIGURE 2: Illustration of $\mathcal{M}_{CI}$ and its estimator

$$\Rightarrow |\mathcal{M}_{CI} - \hat{\mathcal{M}}_{CI}| < \epsilon_{X|Z} + \epsilon_{X|YZ} + \underbrace{|\hat{\mathcal{M}}_{CI} - \tilde{\mathcal{M}}_{CI}|}_{\to 0 \text{ as } n\to\infty}$$

- **Problem** Find estimator $p_u(v)$ of $P(U < u|V = v)$ s.t.

$$\epsilon^2_{U|V} = \int (P(U < u|V = v) - p_u(v))^2 \mathrm{d}F_U(u)\mathrm{d}F_V(v)$$

is minimized.

- **Solution**

1 $F_U(u)$ is non-negative, therefore,

$$\text{minimize } \epsilon^2_{U|V} \Rightarrow \text{ minimize } \epsilon^2_{U|V}(u) = \int (P(U < u|V = v) - p_u(v))^2 \mathrm{d}F_V(v)\, \forall u.$$

2 $P(U < u) = \mathbf{E}\mathbb{I}(U \le u) \Rightarrow$

$$\begin{aligned}\epsilon^2_{U|V}(u) &= C - 2\int P(U < u|V = v)p_u(v)\mathrm{d}F_V(v) + \int p_u^2(v)\mathrm{d}F_V(v)\\ &= C - 2\int \mathbb{I}(u' \le u)\mathrm{d}F_{U|V}(u'|v)p_u(v)\mathrm{d}F_V(v) + \int p_u^2(v)\mathrm{d}F_V(v)\\ &= C - 2\int \mathbb{I}(u' \le u)p_u(v)\mathrm{d}F_{UV}(u', v) + \int p_u^2(v)\mathrm{d}F_V(v).\end{aligned}$$

where $C = \int P^2(U < u|V = v)\mathrm{d}F_V(v)$.

3 Assume $p_u(v) = \sum_{i=1}^m \alpha_i^u \phi_i(v)$, then,

$$\begin{aligned}\epsilon^2_{U|V}(u) &= C - 2\mathbf{E}\sum_{j=1}^m \alpha_j^u \mathbb{I}(U \le u)\phi_j(V) + \mathbf{E}\sum_{i=1}^m\sum_{j=1}^m \alpha_i^u\alpha_j^u\phi_i(V)\phi_j(V)\\ &= C - 2\mathbf{b}^\top\alpha_u + \alpha_u^\top\mathbf{A}\alpha_u\end{aligned}$$

where $[\mathbf{b}]_j = \mathbf{E}[\mathbb{I}(U \le u)\phi_j(V)]$ and $[\alpha_u]_i = \alpha_i^u$ are column vectors, and $[A]_{ij} = \mathbf{E}[\phi_i(V)\phi_j(V)]$ is a matrix.

4 Given realizations $\{(u_i, v_i)\}_{i=1}^n$, $\hat{\mathbf{b}} = \frac{1}{n}\Phi\mathbf{i}$ and $\hat{\mathbf{A}} = \frac{1}{n}\Phi\Phi^\top$, where $[\Phi]_{ij} = \phi_j(v_i)$ is a matrix and $[\mathbf{i}_u]_i = \mathbb{I}(u_i \le u)$ is a column vector.

5 Regularized solution; $\hat{P}(U < u|V = v) = \sum_{i=1}^n \alpha_i^{*(u)}\phi_i(v)$ where

$$\alpha_u^* = (\Phi^\top\Phi + \lambda_u^\Phi\mathbf{I})^{-1}\Phi^\top\mathbf{i}_u.$$

- **Estimator**

$$\begin{aligned}\Rightarrow \hat{\mathcal{M}}^2_{CI} &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n(\hat{P}(X < x_i|Y = y_j, Z = z_j) - \hat{P}(X < x_i|z_j))^2\\ &= \frac{1}{n^2}||(\Xi(\Xi^\top\Xi + \lambda_\Xi\mathbf{I})^{-1}\Xi^\top - \Psi(\Psi^\top\Psi + \lambda_\Psi\mathbf{I})^{-1}\Psi^\top)\mathbf{i}||_F^2.\end{aligned}$$

where $[\mathbf{i}]_{ij} = \mathbb{I}(x_i \le x_j)$ is a matrix of 0s and 1s, and Ξ and Ψ are matrix generated by basis functions ξ and ψ, chosen to be Gaussian kernel.

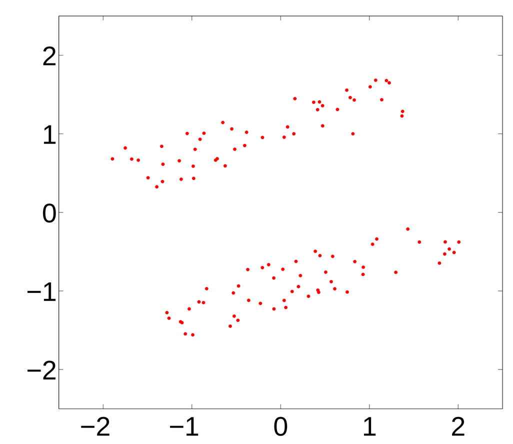
Measure of dependence

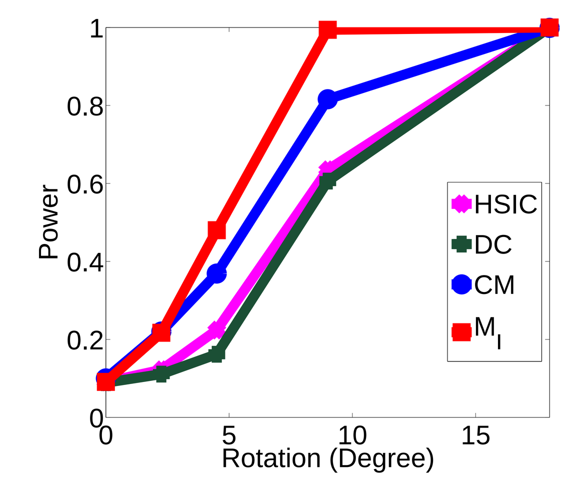
Independence is a special case of conditional independence

- **Definition** (X, Y) are said to be independent Y does **not** contain any additional information about X i.e. $P(X < x|Y = y) = P(X < x)\,\forall (x, y, z)$.
- **Applications** Independent component analysis (ICA)
- **Measure** $\mathcal{M}_I^2 = \int (P(X < u|Y = v) - P(X < u))^2 \mathrm{d}F_X(u)\mathrm{d}F_Y(v)$.
- **Estimator** $\hat{\mathcal{M}}_I^2 = \frac{1}{n^2}||(\Theta(\Theta^\top\Theta + \lambda_\Theta\mathbf{I})^{-1}\Theta^\top - n^{-1}\mathbf{J})^{-1}\mathbf{i}||_F^2$ where $\mathbf{J}$ is a matrix of ones and Θ is matrix generated by basis function θ, chosen to be Gaussian kernel.

Simulation

- **Test of dependence** Comparison among $\mathcal{M}_I$, Hilbert-Schmidt independence criterion (HSIC) [1], distance correlation (DC) [3] and Cramer-von-Mises (CM) test.

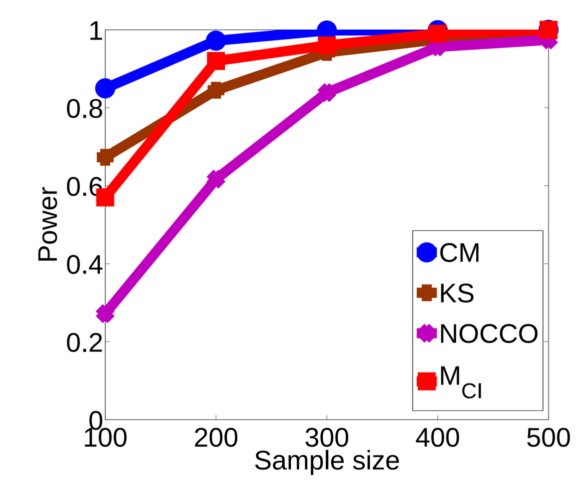


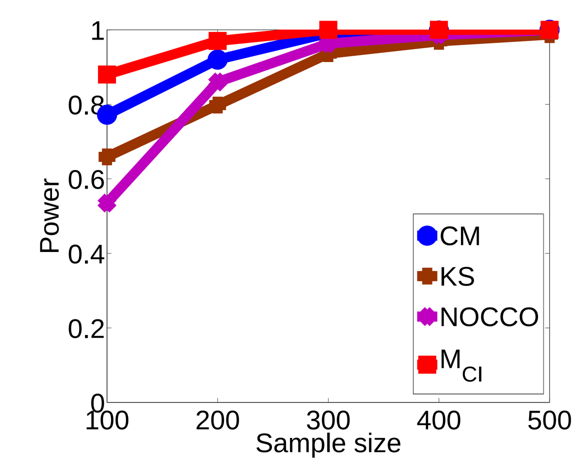


- **Test of conditional independence** Comparison among $\mathcal{M}_{CI}$, normalized cross covariance operator (NOCCO) based measure [1], CM test and Kolmogorov-Smirnov (KS) test [2]. For the following time series $Y_t \not\perp Y_{t-2}|Y_{t-1}$

$$Y_t = -0.5Y_{t-1} + 0.5Y_{t-2}\left(1 + \exp^{-0.5Y_{t-1}}\right)^{-1} + \epsilon_t, \epsilon_t \sim \mathcal{N}(0, 1)$$

$$Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + 0.125Y_{t-3} + \sqrt{0.3 + |Y_{t-3}|}\epsilon_t, \epsilon_t \sim \mathcal{N}(0, 1)$$



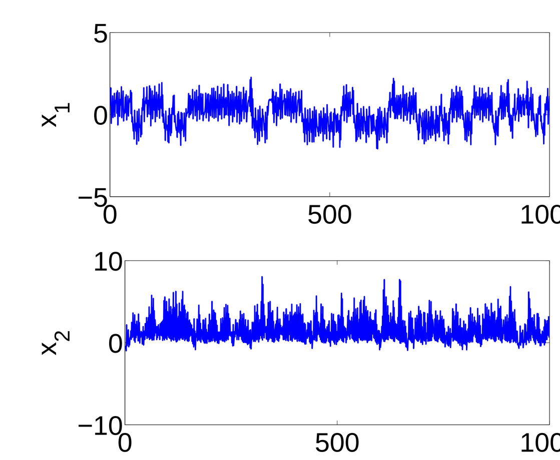


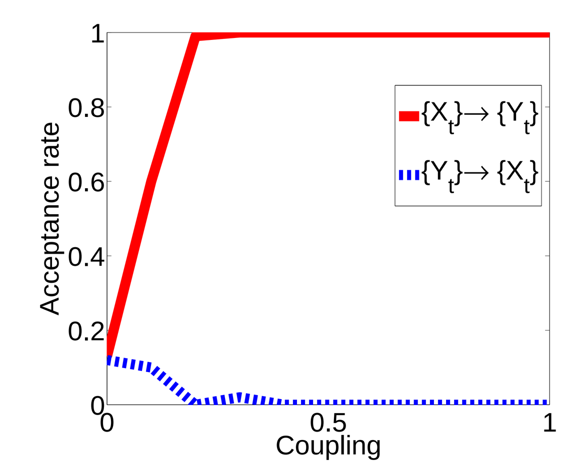
- **Test of Granger noncausality** For the following bivariate time series $\{X_t\}$ causes $\{Y_t\}$ but not the other way around.

$$x_1(t) = 3.4x_1(t-1)(1 - x_1^2(t-1))e^{-x_1^2(t-1)} + 0.8x_1(t-2) + \epsilon_1$$

$$x_2(t) = 3.4x_2(t-1)(1 - x_2^2(t-1))e^{-x_2^2(t-1)} + 0.5x_2(t-2) + cx_1^2(t-2) + \epsilon_2$$

where $\epsilon_1, \epsilon_2 \sim \mathcal{N}(0, 1)$ and $0 \le c \le 1$ is the coupling.





References

[1] K. Fukumizu, A. Gretton, X. Sun, and B. Schölkopf. Kernel measures of conditional dependence. *Advances in Neural Information Processing Systems 20*, pages 489–496, 2008.

[2] O. Linton and P. Gozalo. Conditional independence restrictions: Testing and estimation. Cowles Foundation Discussion Papers 1140, Cowles Foundation, Yale University, November 1996.

[3] G. J. Székely, M. L. Rizzo, and N. K. Bakirov. Measuring and testing dependence by correlation of distances. *Ann. Stat.*, 35(6):2769–2794, 2007.