Estimation of conditional mutual information and its application as a measure of conditional dependence

Sohan Seth and José C. Príncipe
University of Florida, Gainesville, FL
{sohan,principe}@cnel.ufl.edu

Conditional independence is an important concept in many different aspects of engineering such as detecting causal relationship between two time series and extracting informative features in a classification or regression setting [2, 3, 5]. Besides several other measures, conditional mutual information (CMI) has been suggested as a measure of conditional dependence [4]. Inspired by the use of kernel based methods as a function approximator, we propose a new approach to estimate CMI. We briefly describe the proposed work below.

Given random variables (X, Y, Z), CMI between X and Y given Z is defined as

$$I(X;Y|Z) = \mathbb{E}_Z[I(X;Y)|Z] = \int f_Z(z) \left[\iint f_{XY|Z}(x,y|z) \log \frac{f_{XY|Z}(x,y|z)}{f_{X|Z}(x|z) f_{Y|Z}(y|z)} \mathrm{d}x\mathrm{d}y \right] \mathrm{d}z. \tag{1}$$

It can be easily seen that conditional mutual information can be expressed in the following form,

$$I(X;Y|Z) = \mathbb{E}_{XYZ} \left[\log \left(\frac{f_{XYZ}(X,Y,X)}{f_X(X) f_Y(Y) f_Z(Z)} \frac{f_X(X) f_Z(Z)}{f_{XZ}(X,Z)} \frac{f_Y(Y) f_Z(Z)}{f_{YZ}(Y,Z)} \right) \right].$$

Let us define

$$h(x,y,z) = \frac{f_{XYZ}(x,y,z)}{f_X(x) f_Y(y) f_Z(z)}.$$

Then

$$f_{XYZ}(x,y,z) = h(x,y,z) f_X(x) f_Y(y) f_Z(z).$$

Integrating both sides with respect to x we get,

$$f_{YZ}(y,z) = f_Y(y) f_Z(z) \mathbb{E}_X[h(X,y,z)] \Rightarrow \frac{f_{YZ}(y,z)}{f_Y(y) f_Z(z)} = \mathbb{E}_X[h(X,y,z)].$$

Similarly,

$$\frac{f_{XZ}(x,z)}{f_X(x) f_Z(z)} = \mathbb{E}_Y[h(x,Y,z)].$$

Thus,

$$I(X;Y|Z) = \mathbb{E}_{XYZ} \left[\log h(X,Y,Z) - \log \mathbb{E}_{X'}[h(X',Y,Z)] - \log \mathbb{E}_{Y'}[h(X,Y',Z)] \right] \tag{2}$$

where X' and Y' are independent copies of X and Y respectively. Therefore, an estimator of $h(x,y,z)$ directly leads to an estimator of CMI.

Inspired by the kernel based methods, we replace h by $\hat{h}(x,y,z) = \sum_{i=1}^n \alpha_i \kappa_1(x,x_i) \kappa_2(y,y_i) \kappa_3(z,z_i)$ where κ_1, κ_2 and κ_3 are positive definite kernels and minimize the following cost function solve for the coefficients,

$$J = \int \left(\frac{f_{XYZ}(x,y,z)}{f_X(x) f_Y(y) f_Z(z)} - \hat{h}(x,y,z) \right)^2 f_X(x) f_Y(y) f_Z(z) \mathrm{d}x\mathrm{d}y\mathrm{d}z + \lambda ||\alpha||_2^2$$

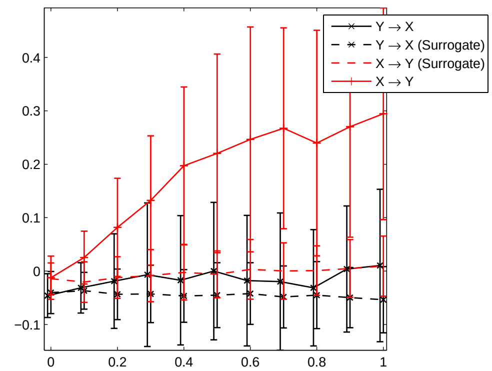

Fig. 1: Surrogate data test of causality.

where α is a vector containing the coefficients α_i's. Expanding the first term in the cost function it can be shown that

1. the actual density functions only appears as expectations i.e. we do not need to know them explicitly and,
2. the cost function is quadratic in terms of α and therefore, can be easily solved.

Replacing $\hat{h}$ in eq. (2), we get an estimator of CMI. Next, we briefly describe an experiment of detecting causality.

Consider two time series $\{X_t\}$ and $\{Y_t\}$ as follows,

$$x(t) = 3.4x(t-1)(1-x^2(t-1))e^{-x^2(t-1)} + 0.8x(t-2) + \epsilon_1$$
$$y(t) = 3.4y(t-1)(1-y^2(t-1))e^{-y^2(t-1)} + 0.5y(t-2) + cx^2(t-2) + \epsilon_2$$

where $\epsilon_1, \epsilon_2 \sim \mathcal{N}(0,1)$. Similar experiment without the noise term has been described in [1]. In this particular case $\{X_t\}$ causes $\{Y_t\}$ but not the other way around. Therefore, $X_t \perp [Y_{t-1}, Y_{t-2}] | [X_{t-1}, X_{t-2}]$ but $Y_t \not\perp [X_{t-1}, X_{t-2}] | [Y_{t-1}, Y_{t-2}]$. Figure 1 shows the surrogate data test with 100 samples. It can be clearly seen that the test supports our expectations.

References

[1] Yonghong Chen, Govindan Rangarajan, Jianfeng Feng, and Mingzhou Ding. Analyzing multiple nonlinear time series with extended granger causality. *Physics Letters A*, 324:26, 2004.

[2] Kenji Fukumizu, Francis R. Bach, and Michael I. Jordan. Dimensionality reduction for supervised learning with reproducing kernel hilbert spaces. *J. Mach. Learn. Res.*, 5:73–99, 2004.

[3] Kenji Fukumizu, Arthur Gretton, Xiaohai Sun, and Bernhard Schölkopf. Kernel measures of conditional dependence. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 489–496. MIT Press, Cambridge, MA, 2008.

[4] Katerina Hlavackova-Schindler, Milan Palus, Martin Vejmelka, and Joydeep Bhattacharya. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, 441(1):1–46, March 2007.

[5] Liangjun Su and Halbert White. A nonparametric Hellinger metric test for conditional independence. *Econometric Theory*, 24:829–864, 2008.