A new nonparametric measure of conditional independence

Sohan Seth, IL Park and José C. Príncipe

Computational NeuroEngineering Lab (www.cnel.ufl.edu)
University of Florida (www.ufl.edu)

May 2, 2009

What is conditional independence

Example

Umbrella and Sun are conditionally independent given Rain.

Conditional independence

Definition (Conditional independence)

- X, Y, Z are three random variables.
- $f_{XY|Z}(x, y|z)$ is the joint density of (X, Y) given $Z = z$ and $f_{X|Z}(x|z)$ and $f_{Y|Z}(y|z)$ are the marginal densities of X and Y given $Z = z$.
- $X \perp Y|Z$ holds if and only if

$$f_{XY|Z}(x, y|z) = f_{X|Z}(x|z) f_{Y|Z}(y|z) \;\forall\; x, y, z \tag{1}$$

- Alternatively, $X \perp Y|Z$ holds if and only if

$$\xi(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f_{XYZ}(x, y, z) f_Z(z) - f_{XZ}(x, z) f_{YZ}(y, z) = 0 \tag{2}$$

Applications

Causality $\{X_t\}$ and $\{Y_t\}$ are two time series. Denote $Y_t^- = (Y_{t-1}, Y_{t-2}, \ldots)$ and $X_t^- = (X_{t-1}, X_{t-2}, \ldots)$.

$$X_t \perp Y_t^- | X_t^- \Leftrightarrow \{Y_t\} \not\to \{X_t\}.$$

Feature Selection $X = \{X1, \ldots, X_d\}$ are features and Y is target. $U \subset X$ is a set of redundant features. Then

$$Y \perp U | \{X \sim U\}.$$

Measure of conditional independence

Definition (measure of conditional independence)

A measure of conditional independence is a nonnegative real number that attains zero if and only if conditional independence is satisfied.

Therefore, a simple measure of conditional independence is,

L_2-norm

$$D = \int \xi^2(x, y, z) \mathrm{d}x \mathrm{d}y \mathrm{d}y \quad (3)$$

$$D = 0 \Leftrightarrow \xi(x, y, z) = 0 \Leftrightarrow X \perp Y | Z$$

Estimation

Definition (Parzen estimate)

$$\hat{f}_{XYZ}(x, y, z) = \frac{1}{n}\sum_{i=1}^{n} \kappa(x; x_i, \sigma_x)\kappa(y; y_i, \sigma_y)\kappa(z; z_i, \sigma_z)$$

where $\{(x_i, y_i, z_i)\}_{i=1}^{n}$ are n iid realizations of (X, Y, Z).

If κ is a Gaussian kernel of the form

$$\kappa(u; u_i, \sigma_u) = \frac{1}{\sqrt{2\pi}\sigma_u} \exp^{-\frac{(u-u_i)^2}{2\sigma_u^2}}$$

then D can be evaluated in closed form.

Difficulties

Estimator

$$\hat{D} = \sum_{s,t,u,v=1}^{n} \alpha_{sv}\{\beta_{sv} + \beta_{tw} - 2\beta_{sw}\}\gamma_{sv}\gamma_{sw}\gamma_{tv}\gamma_{tw}\gamma_{st}\gamma_{vw}$$

where

$$\alpha_{ij} = \kappa(x_i - x_j, \sqrt{2}\sigma_x), \beta_{ij} = \kappa(y_i - y_j, \sqrt{2}\sigma_y), \gamma_{ij} = \kappa(z_i - z_j, 2\sigma_z)$$

Difficulties

The difficulties with the current formulation is that,

1. We need to choose 3 parameters σ_x, σ_y and σ_z.
2. The complexity of evaluating the estimator is $O(n^4)$.

Approach

Theorem

A function $g(x,y,z)$ is zero almost everywhere on $\mathbb{R}^3$ if and only if

$$\int \theta_1(x-a)\theta_2(y-b)\theta_3(z-c)\, g(x,y,z)\, \mathrm{d}x\mathrm{d}y\mathrm{d}z = 0 \forall\ (a,b,c)$$

provided the Fourier transforms of $\theta_i(u) : \mathbb{R} \to \mathbb{R}$ do not vanish.

Smoothing

Smooth eq. (2) to get,

$$\zeta(\mathbf{a},\mathbf{b},\mathbf{c}) = \int \theta_1(x-a)\theta_2(y-b)\theta_3(z-c)\xi(x,y,z)\, \mathrm{d}x\mathrm{d}y\mathrm{d}z.$$

Then

$$\zeta(a,b,c) = 0 \Leftrightarrow \xi(x,y,z) = 0$$

Modified measure

New measure of conditional independence

$$D_m = \int \zeta^2(a,b,c)\,\mathrm{d}a\mathrm{d}b\mathrm{d}c \tag{4}$$

$$D_m = 0 \Leftrightarrow \zeta(x,y,z) = 0 \Leftrightarrow \xi(x,y,z) = 0 \Leftrightarrow X \perp Y|Z$$

Again, if θ is a Gaussian

$$\theta(u; u_i, \varsigma_u) = \frac{1}{\sqrt{2\pi}\varsigma_u} \exp^{-\frac{(u-u_i)^2}{2\varsigma_u^2}}$$

then D_m can be evaluated in closed form.

Estimator

Estimator

$$\hat{D}_m = \sum_{s,t,u,v=1}^{n} \alpha_{sv}\{\beta_{sv} + \beta_{tw} - 2\beta_{sw}\}\gamma_{sv}\gamma_{sw}\gamma_{tv}\gamma_{tw}\delta_{st}\delta_{vw}$$

where

$$\alpha_{ij} = \kappa(x_i - x_j, \sqrt{2(\sigma_x^2 + \varsigma_x^2)}), \beta_{ij} = \kappa(y_i - y_j, \sqrt{2(\sigma_y^2 + \varsigma_y^2)})$$

$$\gamma_{ij} = \kappa(z_i - z_j, 2\sqrt{(\sigma_z^2 + 2\varsigma_z^2)}), \delta_{ij} = \kappa(z_i - z_j, 2\sigma_z\sqrt{\frac{\sigma_z^2 + 2\varsigma_z^2}{\sigma_z^2 + 4\varsigma_z^2}})$$

The proposed approach gives rise to a similar estimator.

Advantages

1. The new measure is robust to the choice of σ_x and σ_y as $\sqrt{2}\sigma_x$ and $\sqrt{2}\sigma_y$ are replaced by $\sqrt{2(\sigma_x^2 + \varsigma_x^2)}$ and $\sqrt{2(\sigma_y^2 + \varsigma_y^2)}$. However, the same is not true for σ_z.
2. The new estimator can be evaluated in $O(n^3)$ by assuming $\gamma_{tv} \approx 1$ which is true for sufficiently large ς_z.
3. Further improvement of computation is possible by noting that the matrix $[\mathbf{K}]_{ij} = \kappa(u_i - u_j)$ is positive definite with a fast decaying eigenstructure. If the rank of this matrix is much lower than n then the computation required is linear in terms of n.

Simulation

Figure: The picture depicts the surrogate data set. The f_i (red) and f_d (green) are the density functions of the estimator under conditional independence and conditional dependence, respectively. green region is the size of the test and the red + green region is the power of the test.

For the following simulations, we choose,

$$\sigma_x^2 + \varsigma_x^2 = \sigma_y^2 + \varsigma_y^2 = 1, \sigma_z^2 = 0.5n^{-\frac{1}{d+4}} \text{ and } \varsigma_z^2 = 1$$

where d is the dimension of (X, Y, Z). σ_z can also be chosen via crossvalidation.

Test of conditional independence

For the following datasets, blue implies $Y_t \perp Y_{t-2} | Y_{t-1}$ and brown implies $Y_t \not\perp Y_{t-2} | Y_{t-1}$,

1. $Y_t = 0.3 Y_{t-1} + \epsilon_t$
2. $Y_t = (-0.5 Y_{t-1} + \epsilon_t) I_{Y_{t-1} \leq 1} + (0.4 Y_{t-1} + \epsilon_t) I_{Y_{t-1} > 1} + \epsilon_t$
3. $Y_t = 0.8 |Y_{t-1}|^{0.5} + \epsilon_t$
4. $Y_t = 0.6 \Phi(Y_{t-1}) Y_{t-1} + \epsilon_t$
5. $Y_t = -0.5 Y_{t-1} + 0.5 Y_{t-2} \left(1 + \exp^{-0.5 Y_{t-1}}\right)^{-1} + \epsilon_t$
6. $Y_t = \exp^{-Y_{t-1}^2} + |0.1 Y_{t-2} (16 - Y_{t-2})| \tau_t$
7. $Y_t = 0.5 Y_{t-1} + 0.25 Y_{t-2} + 0.125 Y_{t-3} + \sqrt{0.3 + |Y_{t-3}|} \epsilon_t$
8. $Y_t = \sqrt{0.01 + 0.8 Y_{t-1} + 0.64 Y_{t-2}^2 + 0.512 Y_{t-3}^2} \epsilon_t$

where $\epsilon_t \sim \mathcal{N}(0, 1)$ and $\tau = \sum_{i=1}^{n} \tau_i$ where $\tau_i \sim \mathcal{U}(-1/7, 1/7)$. We choose $n = 200$.

Result

Table: Power of the test. The size of the test is 0.05. Therefore, datasets $1-4$ should have a value close to 0.05 and datasets $5-8$ should have a value close to 1.

1	2	3	4	5	6	7	8
0.09	0.05	0.05	0.04	0.52	0.21	0.64	0.42

Test of causality

For the following dataset, $\{X_t\} \to \{Y_t\}$ but $\{Y_t\} \not\to \{X_t\}$.

$$x(t) = 3.4x(t-1)(1 - x^2(t-1))e^{-x^2(t-1)} + 0.8x(t-2) + 0.1\epsilon_t$$

$$y(t) = 3.4y(t-1)(1 - y^2(t-1))e^{-y^2(t-1)} + 0.5y(t-2) + cx^2(t-2) + 0.1\epsilon_t$$

where $0 \leq c \leq 1$ is the coupling strength and $\epsilon \sim \mathcal{N}(0,1)$. Therefore, $X_t \perp [Y_{t-1}, Y_{t-2}] \mid [X_{t-1}, X_{t-2}]$ but $Y_t \not\perp [X_{t-1}, X_{t-2}] \mid [Y_{t-1}, Y_{t-2}]$. We choose $n = 100$.

Result

Figure: Result of the surrogate data test. The test strongly suggests that $X \to Y$ but $Y \to X$.

Result II

Table: Power of the test. The size of the test is 0.05. Therefore, $Y \to X$ should have a value close to 0.05 and $X \to Y$ should have a value close to 1 except at $\gamma = 0$.

	0	0.1	0.2	0.3	0.4	
$X \to Y$	0.01	0.02	0.19	0.53	0.95	
$Y \to X$	0.01	0.01	0.03	0.02	0.05	
	0.5	0.6	0.7	0.8	0.9	1
	0.97	0.97	0.99	0.99	0.99	0.99
	0.03	0.01	0.02	0.03	0.01	0.02

Summary

We introduce a novel nonparametric measure of conditional independence which is

1. easy to compute, and
2. robust to the selection of parameters.

Thank you

Questions?