Causality via Transfer Entropy

Àlex Serés

4/1/2016

Granger Causality

Based on the two following principles:

- The cause happens prior to its effects (there is a **lag**).
- The cause has unique information about about the future values of its effect.

Then, given two jointly distributed, stationary, multivariate stochastic processes, X_t and Y_t:

X_t is said to Granger cause Y_t, if Y_t can be better predicted using the histories of both X_t and Y_t than it can by using the history of Y_t alone

Granger Causality

To measure it:

$$Y_t = \alpha_t + Y_t^{(p)} \cdot A + \epsilon_t$$

$$Y_t = \alpha'_t + (Y_t^{(p)} \oplus X_t^{(q)}) \cdot A' + \epsilon'_t$$

$$\mathcal{F}_{X \to Y} \equiv ln \frac{var(\epsilon_t)}{var(\epsilon'_t)}$$

Note that the linear version of causality is not the only one.

Entropy

Entropy is a measure of the disorder of a system. In Information theory, (Shanon) entropy is defined as:

$$H(X) = -\sum_{x \in \chi} p(x) \log p(x) = -E_p(\log p(X))$$

Given two random variables, the join entropy is:
$H(X, Y) = -\sum_{x \in \chi} \sum_{y \in \iota} p(x, y) \log p(x, y) = -E_p(\log p(X, Y))$

Conditional Entropy

$$H(X|Y) = -\sum_{x \in \chi} p(x) \sum_{y \in \iota} p(y|x) \log p(y|x) = -E_p(\log p(Y|X))$$

Chain rule for entropy:

$$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$$

$$H(X_1, X_2, ...X_n) = \sum_{i=1}^{n} H(X_i | X_{i-1}, X_{i-2}, ...X_1)$$

Mutual Information (MI)

Given two random variables, it measures the deviation in entropy from the system where both variables are independent.

$$H_1(X, Y) = -\sum_{x \in \chi} \sum_{y \in \iota} p(x, y) \log p(x, y)$$

$$H_2(X, Y) = -\sum_{x \in \chi} \sum_{y \in \iota} p(x, y) \log p(x) p(y)$$

$$I(X; Y) = H_1 - H2 =$$

$$= -\sum_{x \in \chi} \sum_{y \in \iota} p(x, y) \log \frac{p(x, y)}{p(x) p(y)} = -E_p(\log \frac{p(X, Y)}{p(X) p(Y)})$$

Mutual Information Properties

$$I(X;X) = H(X)$$

$$I(X;Y) = I(Y;X)$$

$$I(X;Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y)$$

Chain rule for mutual information:

$$I(X_1, ...X_n; Y) = \sum_{i=1}^{n} I(X_i; Y|X_{i-1}, X_{i-2}, ...X_1)$$

Transfer entropy (TE)

Non-parametric statistic measuring the amount of directed (time-asymmetric) transfer of information between two random processes. The transfer entropy of X to Y conditioned to Z can be written as:

$$TE_{X \longrightarrow Y|Z} = H(Y_t|Y^- \oplus Z^-) - H(Y_t|X^- \oplus Y^- \oplus Z^-)$$

And by MI properties:

$$TE_{X \longrightarrow Y|Z} = I(Y_t; X^-|Y^- \oplus Z^-)$$

TE vs G-Causality

- For Gaussian variables, TE and Granger Causality are entirely equivalent[1].
- It is expected that at least TE is bounded inferiorly by Granger causality

State Space Representation

A state space model for a time series, Y_t consits on two equations:

- The observation equation:

$$Y_t = G_t \boldsymbol{X_t} + W_t \qquad W_t \sim WN(0, R_t)$$

- The observation equation:

$$\boldsymbol{X_{t+1}} = F_t \boldsymbol{X_t} + V_t \qquad V_t \sim WN(0, Q_t)$$

With $E(W_t V_s') = 0$ for all t and s.

Takens' Embedding

By Takens' theorem [4] we know that we can reconstruct a chaotic dynamical system from a sequence of univariate observations of its state, using the embedding:

$$\boldsymbol{X_t} = (Y_t, Y_{t-\tau}, ... Y_{t-(m-1)\tau})$$

Embedding parameters

We must estimate m (the embedding dimension) and τ (the embedding delay) for each variable of X_t [3], as the observation function can be considered locally linear[3].

- The simplest reasonable estimate of an optimal delay is the first zero of the autocorrelation function.
- An estimate of m can be obtained using the method of false neighbors[2].

Estimating the dimension

To choose m [5] we increase 1 by 1 the dimension and check for possible false neighbors. When none are found we stick with the last dimension size.

The rth neighbour of a m-dimensional point $\boldsymbol{X_t}$ is considered false if:

- $\frac{d_{m+1}(\boldsymbol{X_t},\boldsymbol{X_t^{(r)}})-d_m(\boldsymbol{X_t},\boldsymbol{X_t^{(r)}})}{d_m(\boldsymbol{X_t},\boldsymbol{X_t^{(r)}})} > R_{tol}$

  Or

- $\frac{d_{m+1}(\boldsymbol{X_t},\boldsymbol{X_t^{(r)}})}{Var(Y_t)} > A_{tol}$

TE with Optimal Self Prediction

As it is reasoned in [2], from the Granger-Wiener principles that define causality, and also asking for optimal self prediction, TE is calculated as:

$$TE_{X \longrightarrow_u Y} = I(Y_t; \boldsymbol{X_{t-u}} | \boldsymbol{Y_{t-1}}) =$$

$$= I(Y_t, \boldsymbol{Y_{t-1}}; \boldsymbol{X_{t-u}}) - I(Y_t; \boldsymbol{Y_{t-1}})$$

It is not necessary to know the true lag δ, as TE is maximal for $u = \delta$.

Estimating MI

Mutual information can be estimated from the average distance to the k-nearest neighbor[6]. Lets consider $Z = (X, Y)$, $z_i = (x_i, y_i)$, with the norm $||z|| = ||(||x||_X, ||y||_Y)||_\infty$. Let $\epsilon(i)/2$ be the distance from z_i to its kth neighbor, and $\epsilon(i)_X/2$ the distance in X of their projection, $\epsilon(i) = max\{\epsilon_X(i), \epsilon_Y(i)\}$. Then $n_x(i)$ is the number of points x_j whose distance from x_i is strictly less than $\epsilon(i)/2$. Then:

$$I(X;Y) \simeq \psi(k) - < \psi(n_x + 1) + \psi(n_y + 1) > + \psi(N)$$

Estimating TE

Thus TE can be estimated with[1]:

$$TE_{X \longrightarrow_u Y} \simeq \psi(k) - < \psi(n_{\mathbf{Y_{t-1}}} + 1) - \psi(n_{Y_t, \mathbf{Y_{t-1}}} + 1) - \psi(n_{\mathbf{X_{t-1}}, \mathbf{Y_{t-1}}} + 1) >$$

Bibliography

[1] Barnett L, Barrett AB, Seth AK (2009) Granger causality and transfer entropy are equivalent for Gaussian variables. Phys Rev Lett 103: 238701.

[2]Wibral M, Pampu N, Priesemann V, Siebenhühner F, Seiwert H, Lindner M, et al. (2013) Measuring Information-Transfer Delays. PLoS ONE 8(2): e55809.

[3]Matthew B. Kennel, Reggie Brown, and Henry D. I. Abarbanel Phys. Rev. A 45, 3403 – Published 1 March 1992

[4]Takens F (1981) Dynamical Systems and Turbulence, Warwick 1980, Springer, volume 898 of Lecture Notes in Mathematics , chapter Detecting Strange Attractors in Turbulence. 366–381.

[5]Mario Ragwitz and Holger Kantz Phys. Rev. E 65, 056201 – Published 15 April 2002

[6] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger Phys. Rev. E 69, 066138 – Published 23 June 2004; Erratum Phys. Rev. E 83, 019903 (2011)