Causality

Granger Causality vs Transfer Entropy

Alireza Nouri

Sharif University

2024

Granger Causality

- Let X_t and Y_t be two stationary processes.

Granger Causality

- Let X_t and Y_t be two stationary processes.
- It is said that the process X_t Granger causes another process Y_t if future values of Y_t can be better predicted using the past values of X_t and Y_t rather than only past values of Y_t if:

Granger Causality

- Let X_t and Y_t be two stationary processes.
- It is said that the process X_t Granger causes another process Y_t if future values of Y_t can be better predicted using the past values of X_t and Y_t rather than only past values of Y_t if:

$$Y_t = a_0 + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \sum_{k=1}^{L} b_{2k} X_{t-k} + \varepsilon_t$$

where $\varepsilon_t \sim N(0, \sigma^2)$ and is i.i.d.

Granger Causality

- Let X_t and Y_t be two stationary processes.
- It is said that the process X_t Granger causes another process Y_t if future values of Y_t can be better predicted using the past values of X_t and Y_t rather than only past values of Y_t if:

$$Y_t = a_0 + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \sum_{k=1}^{L} b_{2k} X_{t-k} + \varepsilon_t$$

  where $\varepsilon_t \sim N(0, \sigma^2)$ and is i.i.d.
- If X_t causes Y_t, b_{2k} should not be equal to 0.

Granger Causality

- Let X_t and Y_t be two stationary processes.
- It is said that the process X_t Granger causes another process Y_t if future values of Y_t can be better predicted using the past values of X_t and Y_t rather than only past values of Y_t if:

$$Y_t = a_0 + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \sum_{k=1}^{L} b_{2k} X_{t-k} + \varepsilon_t$$

where $\varepsilon_t \sim N(0, \sigma^2)$ and is i.i.d.
- If X_t causes Y_t, b_{2k} should not be equal to 0.
- Hypothesis testing:

$$H_0 : b_{2k} = 0 \quad \text{for} \quad k \in \{1, 2, 3, \ldots, L\}$$

$$G = \frac{\frac{R_2 - R_1}{L}}{\frac{R_1}{N-2L}} \sim F_{L,N-2L}$$

1. Granger Causality
2. Transfer Entropy
3. Conclusions

Transfer Entropy

- Shannon Information (Entropy):

$$H(X) = -\int_{x\in\mathcal{X}} p(x)\log_2(p(x))\,dx$$

Transfer Entropy

- Shannon Information (Entropy):

$$H(X) = -\int_{x\in\mathcal{X}} p(x)\log_2(p(x))\,dx$$

- Mutual Information:

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

Transfer Entropy

- Shannon Information (Entropy):

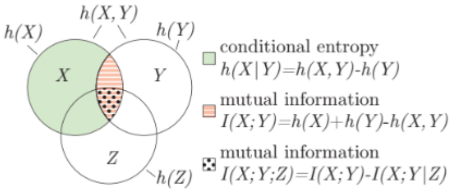
$$H(X) = -\int_{x\in\mathcal{X}} p(x)\log_2(p(x))\,dx$$

- Mutual Information:

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

- Conditional Mutual Information:

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



Transfer Entropy

- Shannon Information (Entropy):

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log_2(p(x))$$

- Mutual Information:

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

- Conditional Mutual Information:

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

- Transfer Entropy:

$$T^t_{X\to Y} \equiv I(Y_t;X_{t-1}|Y_{t-1}) \geq 0$$

$$Y_t = f(X_{t-1}) + f(Y_{t-1}) + \varepsilon_t$$

General (k, ℓ) $-$ *HistoryDefinitionofTransferEntropy*

- General Transfer Entropy:

$$T_{X\to Y}^{(K,\ell)} \equiv I(Y_t; X_{t-1}^{\ell} | Y_{t-1}^{k})$$

$$X_{t-1}^{\ell} = (X_{t-1}, X_{t-2}, \ldots, X_{t-\ell})$$

$$Y_{t-1}^{k} = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k})$$

- Multivariate Transfer Entropy:

$$T_{X\to Y|Z}^{(K,\ell,m)} \equiv I(Y_t; X_{t-1}^{\ell} | Y_{t-1}^{k}, Z_{t-1}^{m})$$

Greedy Algorithm for mTE Estimation (Five Steps)

1. Initialise Z as an empty set and consider the candidate sets for Y's past (CY) and X's past (CX).

Greedy Algorithm for mTE Estimation (Five Steps)

1. Initialise Z as an empty set and consider the candidate sets for Y's past (CY) and X's past (CX).
2. Select variables from CY:

Greedy Algorithm for mTE Estimation (Five Steps)

1. Initialise Z as an empty set and consider the candidate sets for Y's past (CY) and X's past (CX).
2. Select variables from CY:
   - For each candidate, $c \in CY$, estimate the information contribution to c as $I(Y_n; c|Z)$
   - Find the candidate with maximum information contribution, c^*, and perform a significance test; if significant, add c^* to Z and remove it from CY. We use the maximum statistic to test for significance.
   - Terminate if $I(Y_n; c^*|Z)$ is not significant or CY is empty.

Greedy Algorithm for mTE Estimation (Five Steps)

1. Initialise Z as an empty set and consider the candidate sets for Y's past (CY) and X's past (CX).
2. Select variables from CY:
   - For each candidate, $c \in CY$, estimate the information contribution to c as $I(Y_n; c|Z)$
   - Find the candidate with maximum information contribution, c^*, and perform a significance test; if significant, add c^* to Z and remove it from CY. We use the maximum statistic to test for significance.
   - Terminate if $I(Y_n; c^*|Z)$ is not significant or CY is empty.
3. Select variables from X's past (i.e., from CX):

Greedy Algorithm for mTE Estimation (Five Steps)

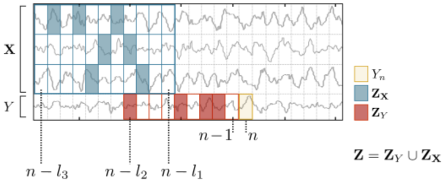
1. Initialise Z as an empty set and consider the candidate sets for Y's past (CY) and X's past (CX).
2. Select variables from CY:
   - For each candidate, $c \in CY$, estimate the information contribution to c as $I(Y_n; c|Z)$
   - Find the candidate with maximum information contribution, c^*, and perform a significance test; if significant, add c^* to Z and remove it from CY. We use the maximum statistic to test for significance.
   - Terminate if $I(Y_n; c^*|Z)$ is not significant or CY is empty.
3. Select variables from X's past (i.e., from CX):
   - Iteratively test candidates in CX following the procedure described in step 2.

Greedy Algorithm for mTE Estimation (Five Steps)

1. Initialise Z as an empty set and consider the candidate sets for Y's past (CY) and X's past (CX).
2. Select variables from CY:
   - For each candidate, $c \in CY$, estimate the information contribution to c as $I(Y_n; c|Z)$
   - Find the candidate with maximum information contribution, c^*, and perform a significance test; if significant, add c^* to Z and remove it from CY. We use the maximum statistic to test for significance.
   - Terminate if $I(Y_n; c^*|Z)$ is not significant or CY is empty.
3. Select variables from X's past (i.e., from CX):
   - Iteratively test candidates in CX following the procedure described in step 2.
4. Prune Z: Test and remove redundant variables in Z using the minimum statistic.

Greedy Algorithm for mTE Estimation (Five Steps)

5. Perform statistical tests on the final set Z:
   - Perform an omnibus test to test the collective information transfer from all the relevant (informative) source variables to the target: $I(Y_n; Z_X | Z_Y)$.
   - If the omnibus test is significant, perform a sequential maximum statistic on each selected variable $z \in Z$ to obtain each variable's final information contribution and p-value.



Conclusion

- Granger causality is particularly well-suited for analyzing linear time series data.
- Transfer entropy, being a non-parametric method, is versatile and effective for both linear and non-linear time series analysis.

References

- Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross-spectral Methods". Econometrica. 37 (3): 424–438. doi:10.2307/1912791. JSTOR 1912791.
- Bossomaier, T., Barnett, L., Harré, M., & Lizier, J. T. (2016). An introduction to transfer entropy. Springer International Publishing
- Zhu, J.; Bellanger, J.-J.; Shu, H.; Le Bouquin Jeannès, R. Contribution to Transfer Entropy Estimation via the k-Nearest-Neighbors Approach. Entropy 2015, 17, 4173-4201. https://doi.org/10.3390/e17064173

Thanks!