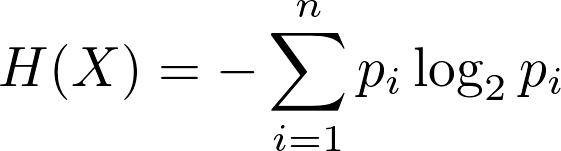
Feel free to skip through the overview of these concepts if you are familiar with them. The explanation of CaMI is further down in this document.

It is first necessary to discuss certain information theory quantities in order to infer causality:

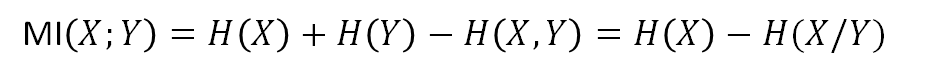
**Marginal and Joint probabilities:** marginal probabilities refer to the probability of a point in space being in a certain partition for variable X, assuming X is divided in any number of partitions, denoted as P(i), where Y is marginalised. Joint probability is the probability of a point appearing in space P(i,j) in the space of a X vs Y plot (Morin, 2016).

**Shannon Entropy:** first described in 1948, it is associated with any random variable and is interpreted as the average amount of information and the variable’s possible outcomes (Shannon, 1948). It measures the uncertainty or surprise of the given variable and is calculated as follows:



Where n is the number of points in a given time series and p is the probability of a point falling in a certain partition in the given space.

**Mutual Information (MI):** this is the measure of dependence between two random variables such as X and Y, it quantifies the transfer of information between the two variables without giving an estimate of causality (Cover & Thomas, 2006). In broader terms MI can be thought of as the correlation between two variables. It can be denoted as MI(X;Y). It is calculated as follows:



Therefore, we need Shannon entropy for X, Y and the joint entropy of X and Y. Traditionally the conditional entropy H(X/Y) is used, however, this is harder to calculate and as seen below not required for this report.

**Causal Mutual Information (CaMI):** this is a novel quantity (first described in Bianco-Martinez & Baptista, 2018) that makes the inference of causality easier, without the need of conditional probabilities to be calculated. To compute this quantity, one requires the Markov symbolic sequences partitioning of the data as described in Bianco-Martinez & Baptista, 2018 and Rubido *et al,* 2018. In practise, this reduces the dimensionality of complicated mathematical approaches by applying a simplistic symbolic encoding: first the data is separated at a given threshold (in the presented work we have chosen 0.5), where values smaller than 0.5 are encoded as 0 and values larger or equal to 0.5 are encoded as 1, this is known as order 1 partition and means that only 1 bit of information is required to present half of the data. Following orders of partitioning can be made by aggregating the 0s and 1s in pairs in the order of the time series, then sets of permutations in 3s and 4s and so on. Order of symbolic sequences is denoted with L. For a more concrete example see Borges *et al,* 2018 and Fig 1. It is important to note that the number of possible permutations grows exponentially with the order L. Therefore, for L=4, for example, the number of possible partitions is 4^2=16. CaMI is calculated by the following equation:



In practice to calculate CaMIX->Y, one simply needs to calculate the MI of symbolic sequences of length L for the variable X and symbolic sequences of length 2\*L for variable Y. This rule can be used to calculate CaMIY->X where the symbolic sequences of X and Y are 2\*L and L respectively (Bianco-Martinez & Baptista, 2018).

 **Figure 1**: Directly taken from Boges *et al,* 2018, this is an example of partitioning. As it can be seen the figure above, L denotes the order of partitioning and each consecutive symbolic encoding is taken as the number of L elements of the first order partitioning.

**Transfer Entropy (TE):** is a statistic that measures the amount and directional flow of information between 2 random events (Schreiber, 2000), in this report we will consider those to be X and Y. TE is defined as follows:



However, considering the information provided above and the works of Bianco-Martinez & Baptista, 2018 and Rubido *et al,* 2018, it can be shown that TE can also be calculated using the following equation:

Thus, removing the need to deal with conditional probabilities and making the calculations much more straightforward. It is important to note that in network systems and graph theory, it can be shown that there is a threshold, which helps infer the connections between different nodes (as described in Rubido *et al,* 2014). In practice, this means that there are 2 clusters of TE values, where the higher values are used to infer the direct connection between different nodes.

**Directionality Index (DI):** is a statistic measure of the directional flow of information from one node of a network to another. DI also measures the amount of information that flows from one node to the other. It is defined as follows:

Since

and

DI = - , then it follows that

Therefore, the DI can have negative and positive values, inferring direction of informational flow from either X to Y or Y to X. The absolute value of this statistical measure is correspondent to the amount of information flowing from one node to the other. It is possible to use DI to construct information graphs as well.

**References**

Bianco-Martinez, E., Baptista, M. S., (2018). *Space-time nature of causality*. *075509*. <https://doi.org/10.1063/1.5019917>

Borges, F. S., Lameu, E. L., Iarosz, K. C., Protachevicz, P. R., Caldas, I. L., Viana, R. L., … Baptista, M. S. (2018). Inference of topology and the nature of synapses, and the flow of information in neuronal networks. *Physical Review E*, *97*(2), 1–7. https://doi.org/10.1103/PhysRevE.97.022303

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory*. Wiley-Interscience.

Rubido, N., Martí, A. C., Bianco-Martínez, E., Grebogi, C., Baptista, M. S., & Masoller, C. (2014). Exact detection of direct links in networks of interacting dynamical units. *New Journal of Physics*, *16*. <https://doi.org/10.1088/1367-2630/16/9/093010>

Schreiber, T. (2000). Measuring Information Transfer. *Physical Review Letters*, *85*(2), 461–464. https://doi.org/10.1103/PhysRevLett.85.461

[Shannon, Claude E.](https://en.wikipedia.org/wiki/Claude_Shannon) (July 1948). ["A Mathematical Theory of Communication"](https://en.wikipedia.org/wiki/A_Mathematical_Theory_of_Communication). [*Bell System Technical Journal*](https://en.wikipedia.org/wiki/Bell_System_Technical_Journal). **27** (3): 379–423. [doi](https://en.wikipedia.org/wiki/Digital_object_identifier):[10.1002/j.1538-7305.1948.tb01338.](https://doi.org/10.1002%2Fj.1538-7305.1948.tb01338.x)