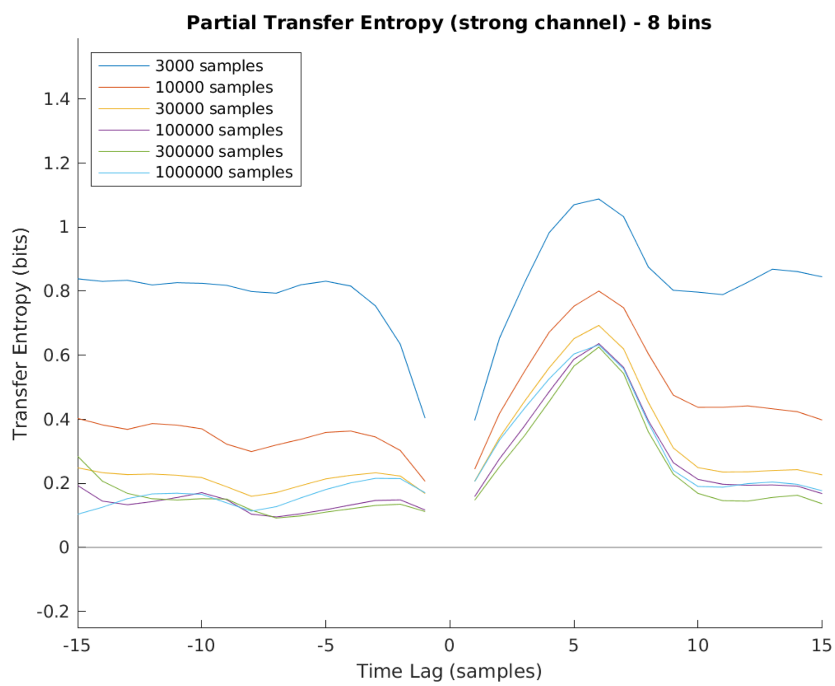
Chris's Entropy Library - User Guide

Written by Christopher Thomas – April 12, 2024.



This project is Copyright (c) 2024 by Vanderbilt University, and is released under the Creative Commons Attribution 4.0 International License.

Contents

Chapter 1

Overview

This library provides a set of functions used for calculating entropy and entropy-related measures on datasets. These were written to deal with neuroscience data (continuous brain wave signals and discrete counts of the numbers spike events seen), but the library should work with any type of data.

Measures computed are:

- **Shannon Entropy** - The amount of "surprise" associated with a given data sample, and the average "surprise" for samples in a data stream. This is the information content of the stream.
- **Conditional Entropy** - The amount learned from samples of variable Y when we already know the value of a related variable X.
- **Mutual Information** - The amount of information shared by samples of related variables Y and X. Measuring either one gets you this information.
- **Transfer Entropy** - The amount of information about the future of variable Y that you learn by knowing the past of variable X, in addition to what you already know from the past of variable Y.

These measures are discussed in detail in Section 2.

A brief description of how to use this library is given in Section **??**. For more information, see the sample code in the "`source-code`" folder. This section also provides a brief overview of considerations relating to neuroscience data analysis.

Section **??** describes the quadratic extrapolation algorithm that was used in Palmigiano 2017. This is intended to allow reliable estimates of several entropy measures with fewer data samples than would otherwise be needed (or equivalently, to improve the reliability of these measures using a fixed number of samples).

Chapter 2

Entropy

2.1 Shannon Entropy

Entropy can be thought of as measuring the information content of a stream of data samples. It is defined (per Shannon 1949) as the amount of "surprise" associated with each sample when that sample is considered to be a discrete symbol drawn from a set of possible symbols with some probability distribution. The entropy associated with one symbol is given in Equation 2.1, and the average entropy per symbol associated with a symbol stream is given in Equation 2.2. It can be seen from Equation 2.1 that symbols with lower probability result in more "surprise": rare symbols are more informative than commonly-seen symbols.

$$H(x_k) = \log_2 \left[\frac{1}{P(x_k)} \right] = -\log_2 [P(x_k)] \tag{2.1}$$

$$H(X) = -\sum_k P(x_k) \log_2 [P(x_k)] \tag{2.2}$$

For discrete-valued data, or data such as text that is readily interpreted as symbols, entropy may be calculated on the raw data values. A histogram of observed symbols is made, and this is used as the probability distribution. For continuous-valued data, histogram bins are typically defined and the bin labels are interpreted as symbols. For data consisting of event arrival times, time bins are typically defined and the number of events within each bin are counted. These counts may be taken as individual symbols or several adjacent time bins may be considered to form a word, which is used as a symbol for entropy calculations. These approaches are discussed in more detail in Section **??**.

2.2 Conditional Entropy

Conditional entropy can be thought of as the amount of additional information learned from samples of variable Y when we already know the value of a related variable X. This is illustrated in Figure **??**. It is defined by Equation 2.3:

$$H(Y|X) = -\sum_{j,k} P(x_j, y_k) \log_2 \left[\frac{P(x_j, y_k)}{P(x_j)} \right] \tag{2.3}$$

For the case of independent variables where $P(x_j, y_k) = P(x_j)P(y_k)$, this reduces to Shannon entropy.

FIXME: Figure for conditional entropy.

This is generalized to several X variables per Equation 2.4; this case is illustrated in Figure **??**.

$$H(Y|X_A, X_B) = -\sum_{j,k,m} P(x_{aj}, x_{bk}, y_m) \log_2 \left[\frac{P(x_{aj}, x_{bk}, y_m)}{P(x_{aj}, x_{bk})} \right] \tag{2.4}$$

FIXME: Figure for conditional entropy with multiple X.

2.3 Mutual Information

Mutual information can be thought of as the amount of information shared between related variables X and Y. Sampling either variable X *or* variable Y reveals the shared information. This is illustrated in Figure **??**. Mutual information is defined by Equation 2.5:

$$I(X,Y) = \sum_{j,k} P(x_j, y_k) \log_2 \left[\frac{P(x_j, y_k)}{P(x_j)P(y_k)} \right] \tag{2.5}$$

For the case of independent variables where $P(x_j, y_k) = P(x_j)P(y_k)$, this is equal to zero (due to the $log_2(1)$ term).

FIXME: Figure for mutual information.

This is generalized to more than two variables per Equation 2.6; this case is illustrated in Figure **??**.

$$I(X,Y,Z) = \sum_{j,k,m} P(x_j, y_k, z_m) \log_2 \left[\frac{P(x_j, y_k, z_m)}{P(x_j)P(y_k)P(z_m)} \right] \tag{2.6}$$

FIXME: Figure for mutual information with three variables.

2.4 Transfer Entropy

Transfer entropy can be thought of as the amount of information about the future of variable Y that you learn by knowing the past of variable X, in addition to what you already know from the past of variable Y. This is illustrated in Figure **??**. Transfer entropy is defined by Equation 2.7:

$$TE_{X \to Y} = H(Y|Y_{past}) - H(Y|Y_{past}, X_{past}) \tag{2.7}$$

The Y_{past} and X_{past} terms are usually approximated by taking the past value at some time $t - \delta t$ as a proxy for the entire past history:

$$\begin{aligned} \hat{Y}_{past} &= Y(t - \delta t) \\ \hat{X}_{past} &= X(t - \delta t) \end{aligned} \tag{2.8}$$

FIXME: Figure for transfer entropy.

Partial transfer entropy is used to disentangle the contributions of multiple X variables to some Y variable. This is illustrated in Figure 2.9. Partial transfer entropy is defined by Equation 2.9:

$$pTE_{X_A \to Y} = H(Y_A|Y_{past}, X_{Bpast}) - H(Y|Y_{past}, X_{Apast}, X_{Bpast}) \tag{2.9}$$

FIXME: Figure for partial transfer entropy.

2.5 References

- C. E. Shannon, *A Mathematical Theory of Communication*, The Bell System Technical Journal, v 27, pp 379–423,623-656, July, October, 1948