Transfer entropy for population data.

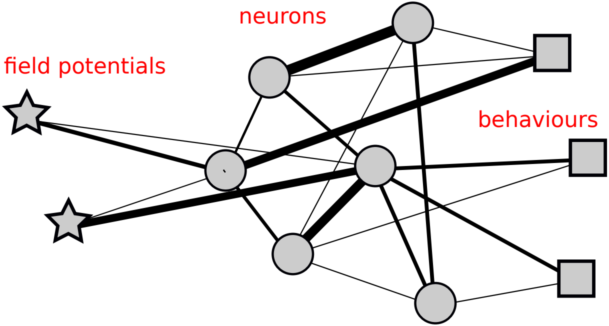
Conor Houghton

U Bristol

Edinburgh, June 2022

Thanks for the invitation

Populations - Neurons - Behaviours



What is this talk about

First we'll discuss estimating mutual information *and then we'll discuss estimating* transfer entropy.

Shannon's entropy

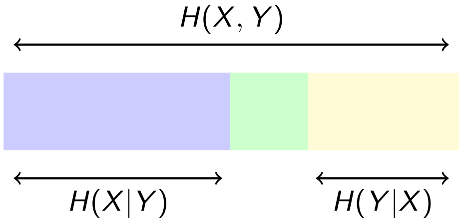
$$H(X) = -\sum_{x} p(x) \log_2 p(x)$$

Shannon's entropy

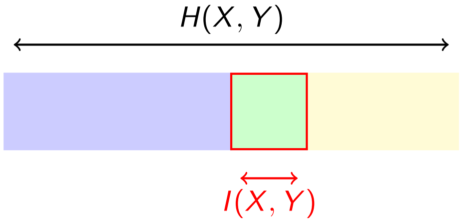
1/2	1/4	1/8	1/16	1/32	1/64	1/128	1/128
000	001	010	011	100	101	110	111
0	10	110	1110	11110	111110	1111110	1111111

$$\text{average code length} = \frac{1}{2} + \frac{1}{4}2 + \frac{1}{8}3 + \frac{1}{16}4 + \ldots = H(X) \approx 1.98 < 3$$

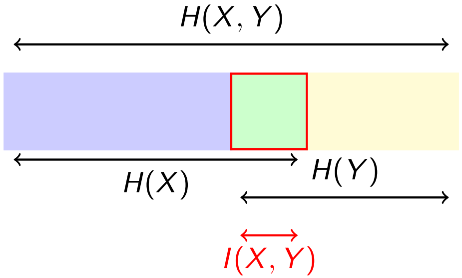
Mutual information



Mutual information



Mutual information



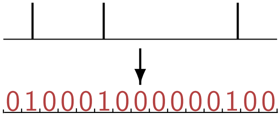
$$I(X, Y) = H(X) + H(Y) - H(X, Y)$$

Mutual information

Mutual information is the true way we measure the relationship between variables; but we ignore it because it is so hard to estimate.

Classical approach

- Discretize.



- Split into words.

$$010001000000100 \to 01000, 10000, 00100$$

Bialek, de Ruyer van Steveninck, Strong and other coworkers, late 1990s.

Classical approach

- Estimate probability of words. For example, say $w_8 = 01000$ then estimate

$$p(w_8) \approx \frac{\#\text{ occurrences of } w_8}{\#\text{ words}}$$

- Calculate

$$H(W) = -\sum_i p(w_i)\log_2 p(w_i) = -\langle \log_2 p(w_i) \rangle$$

Bialek, de Ruyer van Steveninck, Strong and other coworkers, late 1990s.

ms scale information in blow fly spike trains.

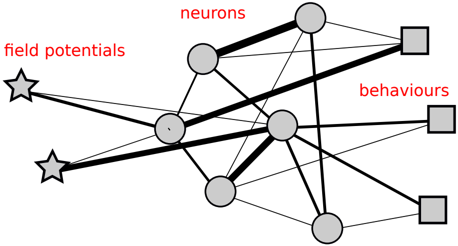
Bialek, de Ruyer van Steveninck, Strong and other coworkers, late 1990s.

Difficulties with the classical approach.

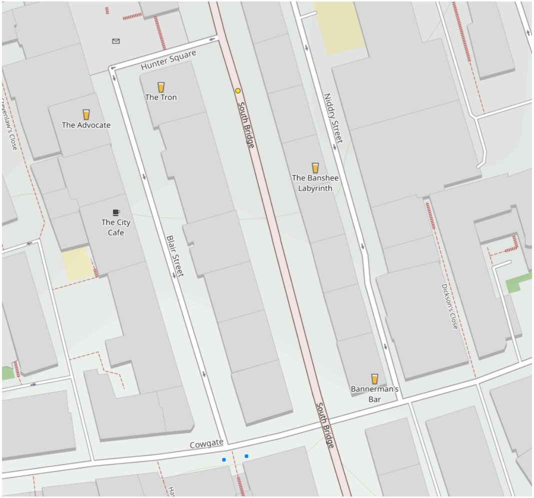
- Undersampling.
  - 100 ms words and 2 ms bins gives $2^{50} = 1125899906842624$ words.
  - Lots of clever approaches to this, for example Nemenman et al. (PRE 2004, BMC Neuroscience 2007) where a cunning prior is used for $p(w_i)$.
- Sampling bias.
  - An even distribution will never give equal counts for each word, giving different $p(w_i)$.
  - Lots of clever approaches to this too, see Panzeri et al. (J Neurophys. 2007).

Many fixes but still . . .

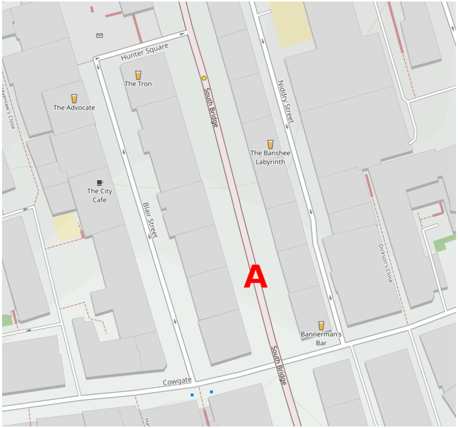
- Neuron - neuron mutual information.
- Maze - neuron mutual information.
- Mutual information between **populations**.
- Mutual information between neurons and **field potentials**.



Also ignores the proximity structure!

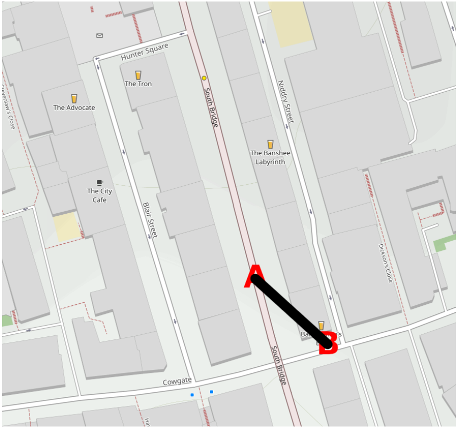


Also ignores the proximity structure!

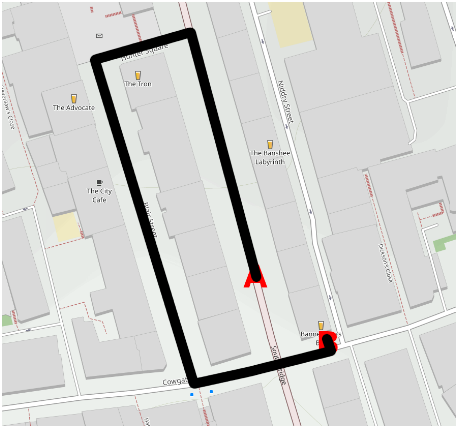


Also ignores the proximity structure!

Also ignores the proximity structure!

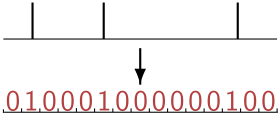


Also ignores the proximity structure!

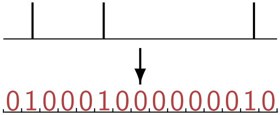


Classical approach

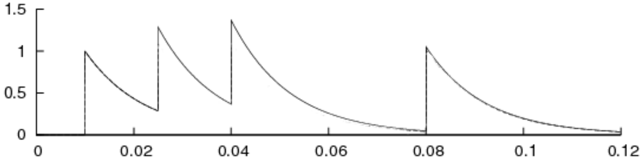
- Discretize.



- Discretize.



van Rossum metric



Spike trains mapped to functions and a metric on the space of functions induces a metric on the spike train space.

van Rossum (Neural Comp. 2001)

Multi-unit van Rossum metric

- There is a multi-unit easily computed version of the van Rossum metric.
- It relies on a time constant and a population parameter.

Houghton and Sen (Neural Comp. 2008) / Houghton and Kreuz (Network 2012)

The rules

We want to estimate mutual information for data on a metric space

- There is a KDE version of this, here we use a Kozachenko-Leonenko approach
- It ends up somewhat similar to the Kraskov, Stögbauer and Grassberger (KSG) estimator.

A dart board

photo from ebay (£4.20 +p.p.)

A dart board

Probability mass function

$$\text{prob(dart lands in } B) = \int_B p(\mathrm{x})\,dV$$

Estimating using the number of number of holes

$$\langle \text{number of holes in } B \rangle = \int_B p(x)dV \times (\text{total number of holes})$$

where the total volume is normalized.

Estimating the probability mass function

If the mass function varies slowly:

$$\int_B p(\mathrm{x})dV \approx p(\mathrm{x}_0) \times \operatorname{vol} B$$

so

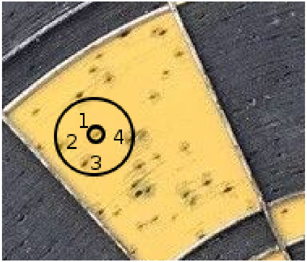
$$\text{number of holes in } B \approx p(\mathrm{x}_0) \times \operatorname{vol} B \times (\text{total number of holes})$$

Using this to find the mutual information gives a *Kozachenko-Leonenko* estimator.

Estimating using the number of number of holes

$$p(x_0) \approx \frac{\#B}{n \times \mathrm{vol}\, B}$$

where n is the total number of points and $\#B$ is the number of points in B.



so

$$p(\circ) = \frac{4}{n \mathrm{vol}\, B}$$

Problem

How do we work out the volume in the space of functions? We have no coordinates *xyz* to do

$$\operatorname{vol} B = \int_B dxdydz$$

We must respect the rules and use only the metric, well the metric and the existence of the probability density.

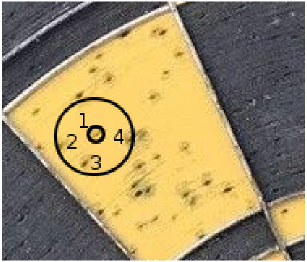
Use the mass function as a measure!

$$\mathrm{vol}\, B = \int_B p(\mathrm{x})dV$$

Volume by counting holes

$$\operatorname{vol} B \approx \frac{\text{number of holes in } B}{\text{total number of holes}}$$

Volume by counting holes



A ball with volume h/n around the circled point, where n is the total number of holes and $h = 4$.

Metric

To make a ball you need a metric; not to measure the radius since the size is being defined by the volume, but to define ‘the nearest h points’.

Oh no

$$p(\mathsf{x}_0) \approx \frac{\#B}{n \times \mathrm{vol}\, B} = \frac{h}{nh/n} = 1$$

and using this meaure gives $H(X) = 0$; in fact the differential entropy is not well-defined. However the mutual information is!

Mutual infomation

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

has two probability distributions: $p_Y(y)$ and $p_{Y|X}(y|x)$!

IDEA: use one to estimate volume, the other can then be estimated by counting!

Formula - discrete case

This is for the case where X is a discrete random variable and everything exciting is happening in Y space.

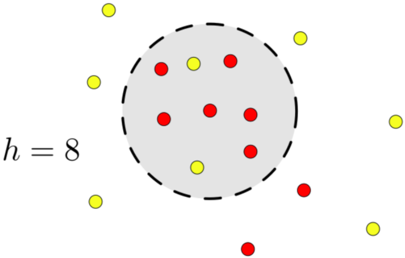
$$I(X,Y) = \frac{1}{n}\sum_{y_i} \log_2 \frac{n\#_{y_i}B}{h}$$

where$\#_y B$ are the number of points in B that correspond to the X value as y and n_s is the number of stimuli.

Conor Houghton, Royal Society Open Science, 2 (2015) 140391.

Formula - discrete case

$$I(X,Y) = \frac{1}{n}\sum_{y_i} \log_2 \frac{n_s \#_{y_i} B}{h}$$



h

There are two approximations:

$$\int_B p(x)dV \approx \#B \times \text{vol}\, B$$

and

$$\int_B p(x)dV \approx \text{V} \times p(x_0)$$

The first approximation gets better if the volume is bigger, the second gets worse; the correct choice of h is a compromize between these two. There is actually a successful approach to picking h that seems to work, based on the bias, which can be calculated analytically.

Two continuous variable

This also works for the case where X and Y are both continuous; as for example, when comparing neuronal populations!

Two continuous variable

In this case we use exploit the fact that there are two probability distributions on the joint space (X, Y).

The joint distribution:

$$p_{X,Y}(x, y)$$

and the marginalized distribution

$$p_X(x)p_Y(y)$$

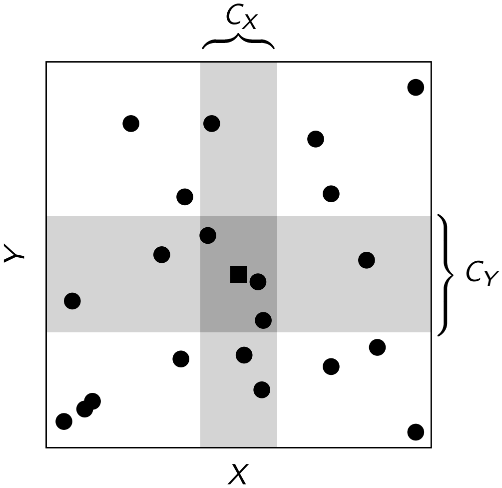
Two continuous variables

$$I(X,Y)=\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{n\#[C(x_i,y_i)]}{h^2}$$

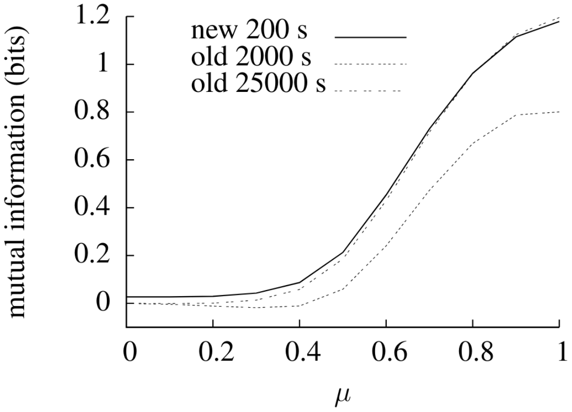
with $C(x_i,y_i)=C_X(x_i,y_i)\cup C_Y(x_i,y_i)$

Two continuous variables

$$I(X,Y) = \frac{1}{n}\sum_{i=1}^{n} \log_2 \frac{n\#[C(x_i,y_i)]}{h^2}$$



Two continuous variables



Conor Houghton, Neural Computation (2019) 31:330-343

Transfer entropy

Work with Jake Witter *. . . with a paper in preparation.*

What about transfer entropy?

The transfer entropy is a measure of causality!

What about transfer entropy?

Isn't that Granger causality? Transfer entropy reduces to Granger causality for vector auto-regressive processes!

Lionel Barnett, Adam B. Barrett, and Anil K. Seth Phys. Rev. Lett. 103, 238701

Transfer entropy

$$T(X \to Y) = I[X(\text{past}), Y(\text{now}) | Y(\text{past})]$$

Transfer entropy

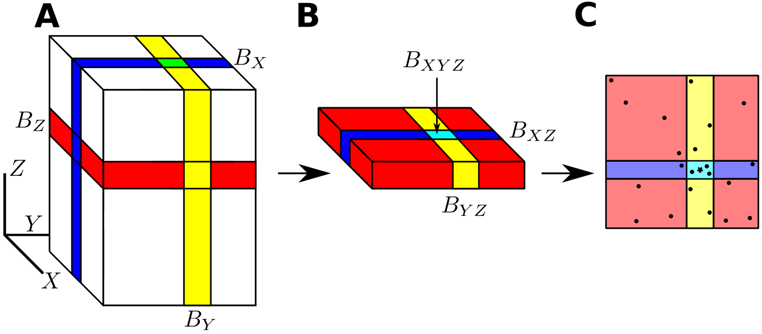
Transfer entropy is a sort of conditional mutual information.

$$I(X, Y|Z)$$

and this suffers even more acutely from sampling problems.

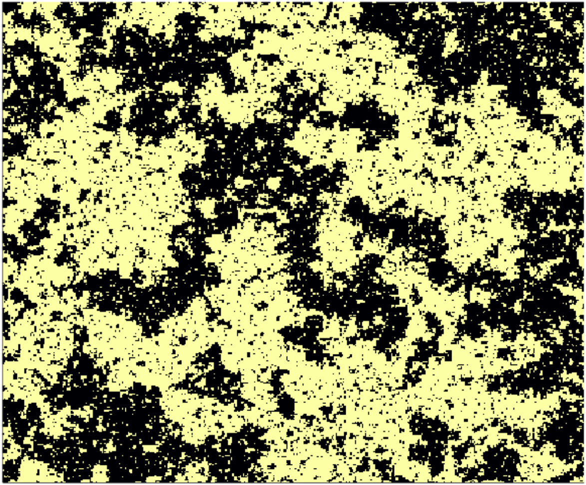
Conditional mutual information

The metric Kozachenko-Leonenko estimator can be extended to this case; it involves three-way intersections of the nearest-neighbour sets.

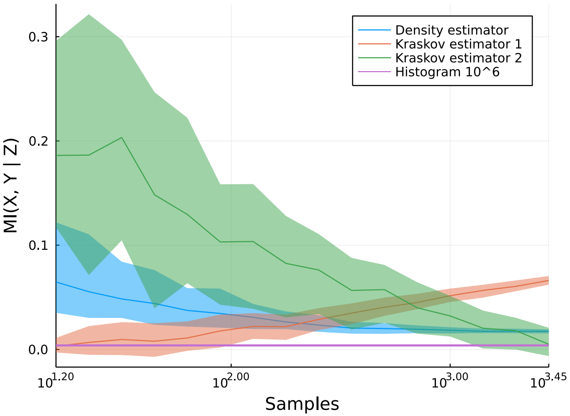


$$I(X, Y|Z) \approx \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{h_{xyz}(i)h}{h_{xz}(i)h_{yz}(i)}\right)$$

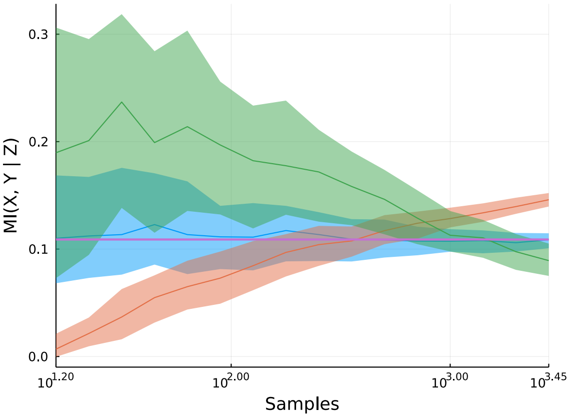
Ising model



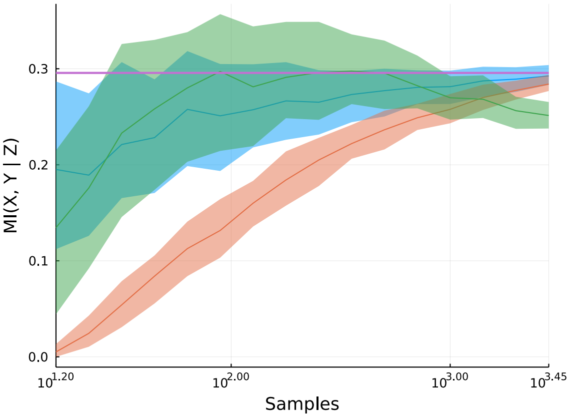
Transfer Entropy



Transfer Entropy



Transfer Entropy



The End

THANK YOU!