Mutual information for functions, maybe even ERPs.

Conor Houghton

CS, U Bristol

Frankfurt, May 2018

Shannon's Entropy 1

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

Shannon's Entropy 2

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 1

$$I(X,Y)=\sum_{x,y} p_{X,Y}(x,y)\log_2\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}$$

Mutual Information 2

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

Mutual Information 3

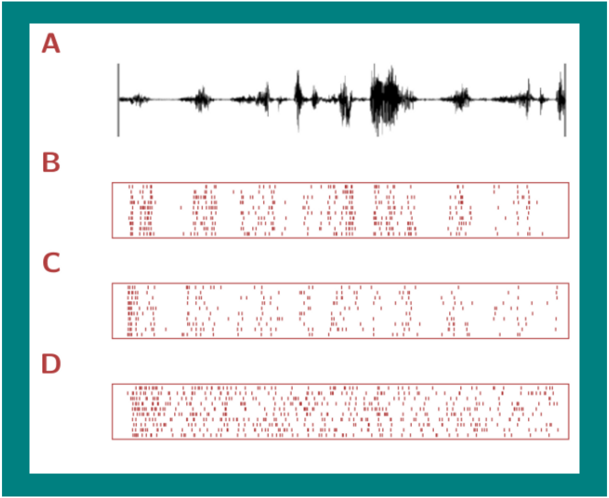
$$I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

or, for example

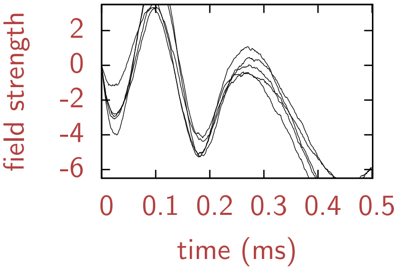
$$H(Y) = H(Y|X) + I(X,Y)$$

info in Y=(info remaining in Y if you know X)+(mutual info)

Spike trains



Functions



A dart board 1

photo from ebay (£4.20 +p.p.)

A dart board 2

Probability mass function

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

Estimating using the number of number of holes 1

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

where the total volume is normalized.

Estimating the probability mass function

If the mass function varies slowly:

$$\int_B p(\mathbf{x})dV \approx p(\mathbf{x}_0) \times \text{vol}\, B$$

so

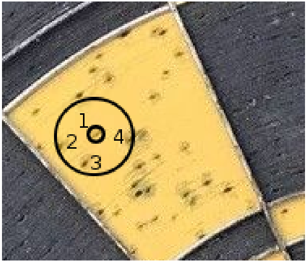
$$\text{number of holes in } B \approx p(\mathbf{x}_0) \times \text{vol}\, B \times (\text{total number of holes})$$

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

Estimating using the number of number of holes 3

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

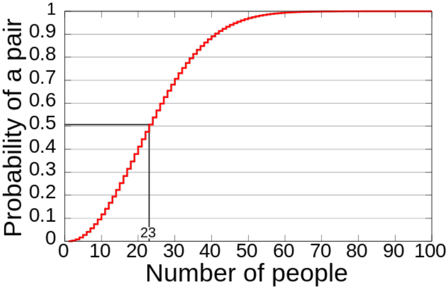
where n is the total number of points and $\#B$ is the number of points in B.



so

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

Kozachenko-Leonenko estimators are very good



graph from wikipedia article on the birthday paradox

Problem

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

$$\text{vol}\, B = \int_B dxdydz$$

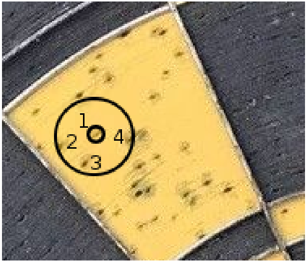
Use the mass function as a measure!

$$\text{vol } B = \int_B p(\mathbf{x}) dV$$

Volume by counting holes 1

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

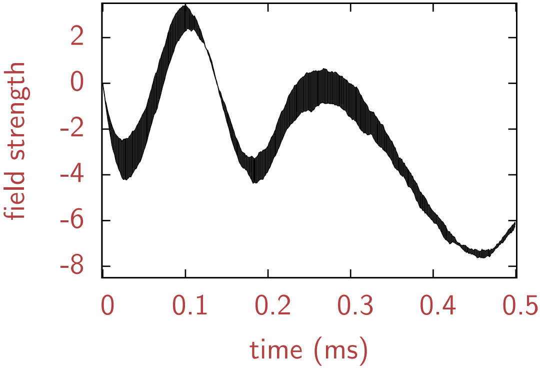
Volume by counting holes 2



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

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

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 1

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

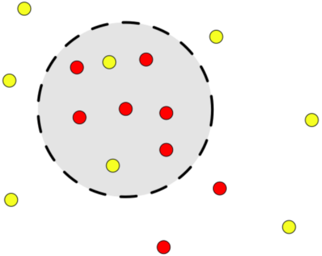
$$I(X,Y) = \frac{1}{n}\sum_{y_i} \log_2 \frac{n\#_{y_i}B}{h}$$

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.

Formula 2

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

$h = 8$



h

There are two approximations:

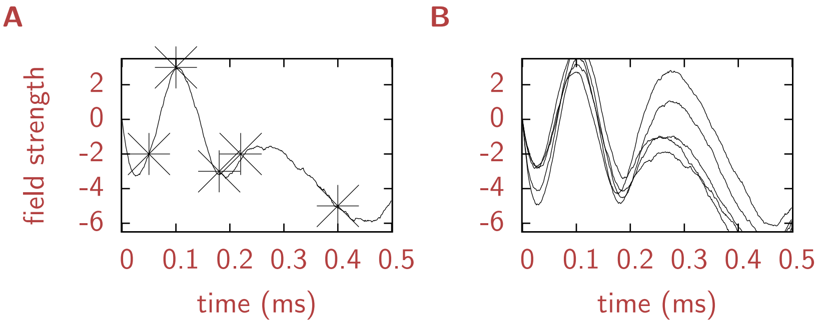
$$\int_B p(x)dV \approx \#B \times \text{vol}\, B$$

and

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

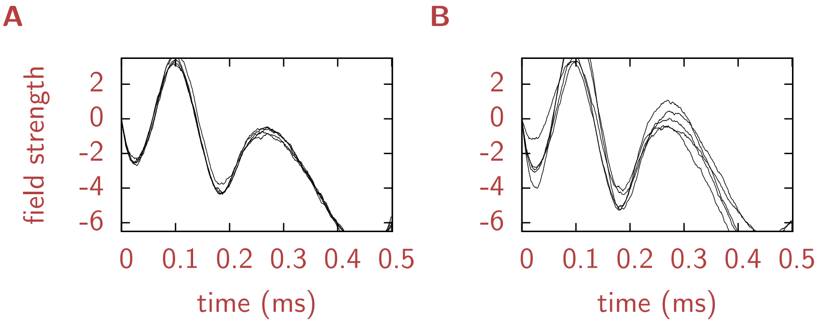
The first approximation gets better if the volume is bigger, the second gets worse; the correct choice of h is a compromize between these two. There is actually a clever approach to picking h that seems to work, based on the bias.

Pretend ERPs 1



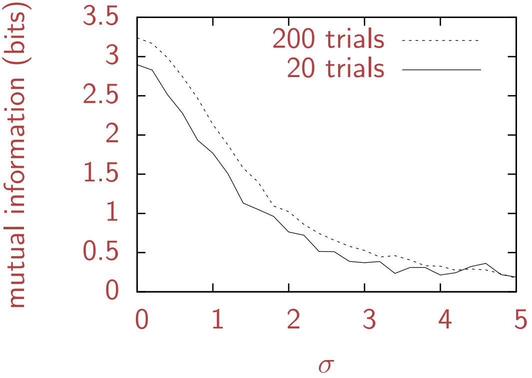
Fictive event related potentials. The ‘stimuli’ are random 5-vectors of landmarks; an ERP is produces by perturbing the landmarks, interpolating with splines and adding noise. **B** shows responses to different stimuli.

Pretend ERPs 2



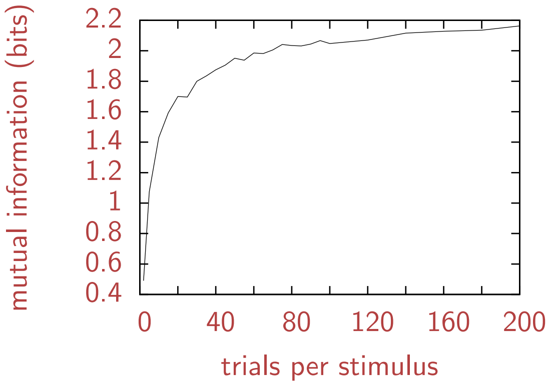
Multiple trials to the same stimulus. The amount the landmark points are moved is determined by σ; for **A** $\sigma = 0.0$, for **B** $\sigma = 2$.

Results 1



Estimated mutual information.

Results 2



Estimated mutual information. Here $\sigma = 1$.

The End

THANK YOU!