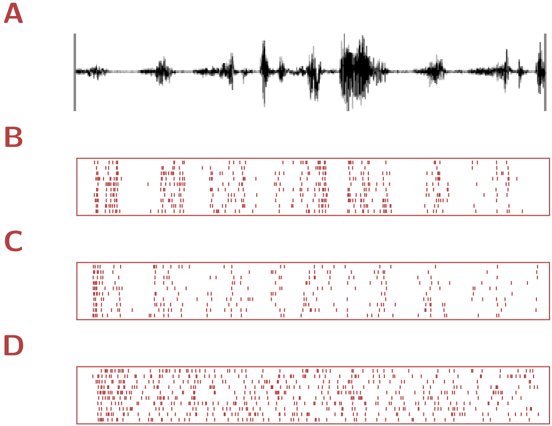
A metric space approach to the information capacity of spike trains.

Conor Houghton

Mathematical Neuroscience Laboratory
School of Mathematics
Trinity College Dublin

Berlin, 30 March 2011

Spike trains.



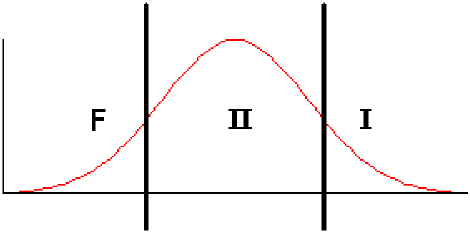
Questions about trains.

- How do the properties of spike trains change along a sensory pathway?
- Is the song rate coded or is there information in temporal features?
- Are neurons in populations redundant?

Questions about trains.

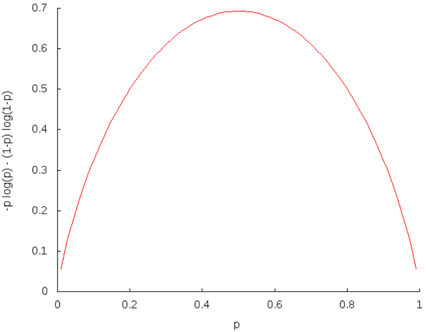
- What is the information theory of spike trains?

Information theory.



Shannon's entropy.

$$H = -\sum_{\text{events}} (\text{probability of the event}) \log (\text{probability of the event})$$



Bialek approach to information and spike trains.[1]

$$\downarrow$$

$$\underline{0,0,1,0,1,0,1,0,0,1,1,0,1,1,0,0,0,1,0,1,1,0,1,1,0,1,0,0,0,0,1,0,1,0,0,0,0,1,0,0}$$

$$\downarrow$$

$$\{\ 00101,01001,10110,00101,10110,10000,10100,00100\}$$

[1] Entropy and Information in Neural Spike Trains, Strong SP, Koberle R, de Ruyter van Steveninck RR and Bialek W (1998) Phys. Rev. Lett. 80: 197-200

Bialek approach to information - calculation.

Make a table, for example:

word	00101	01001	10110	10101	10110	10000	etc
prob.	0.011	0.022	0.052	0.011	0.054	0.098	...

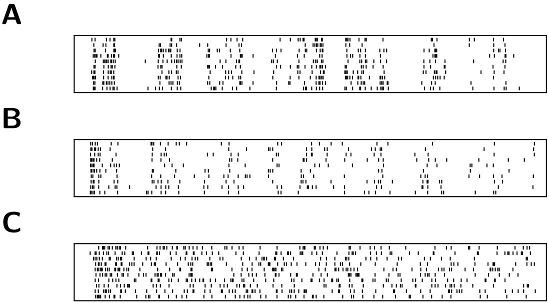
and calculate the corresponding entropy

$$H = -\sum_{\text{words}} p(\text{word}) \log p(\text{word})$$

Bialek approach to information - result.

The information is the difference between the signal entropy and the noise entropy.

$$H_s - H_\eta$$



Bialek approach to information - result.

This is the *mutual information* between the stimulus and the response.

$$\begin{aligned} H_s &= H(\text{response}) \\ H_\eta &= H(\text{response}|\text{stimulus}) \end{aligned}$$

so

$$I(\text{response}, \text{stimulus}) = H_s - H_\eta$$

Bialek approach to information and spike trains - problems.

- To take into account timing precision a small discretization lengths is needed.
- A huge number of words, most of the ones that occur are mostly zeros.
- A huge sample size needed; Bialek worked with fly, such long recording are not normally possible.
- There are also interpretational problems with any information theory approach to neuroscience, we won't deal with that here.

Bialek approach to information and spike trains - no noise model.

- No model of noise.
- No notion of one word being near another.

$\downarrow$

0 0 1 0 1 0 1 0 0 1 1 0 1 1 0 0 0 1 0 1 1 0 1 1 0 1 0 0 0 0 1 0 1 0 0 0 0 1 0 0

$\downarrow$

{ 00101,01001,10110,00101,10110,10000,10100,00100}

The space of spike trains?

- Should we using the discrete theory or the continuous one?
- Spike times are not discrete.

- The continuous theory assumes a continuous space, what is the space of spike trains?

Metric spaces

A metric maps pairs of points a and b, to a real number $d(a, b)$ such that

- Positive and distinguishable

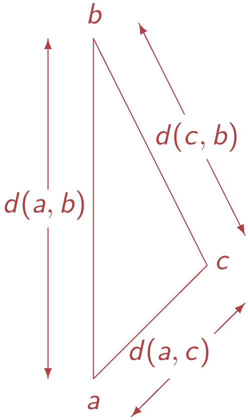
$$\begin{aligned} d(a, b) &\geq 0 \\ d(a, b) &= 0 \iff a = b, \end{aligned}$$

- Symmetric

$$d(a, b) = d(b, a).$$

- Triangle inequality

$$d(a, b) \leq d(a, c) + d(c, b).$$

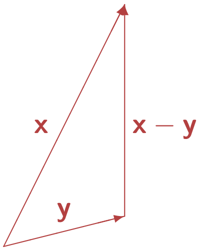


The triangle inequality

Euclidean metrics

- In $\mathbf{R}^3$ say

$$\begin{aligned}\mathbf{x} &= (x_1, x_2, x_3)\\ \mathbf{y} &= (y_1, y_2, y_3)\end{aligned}$$



- The dot product is given by $\mathbf{x}\cdot\mathbf{y} = x_1y_1 + x_2y_2 + x_3y_3$
- The dot-product of a vector with itself is *a norm*, a measure of the length of the vector $|\mathbf{x}| = \sqrt{\mathbf{x}\cdot\mathbf{x}}$.
- This norm induces a metric, called the L^2 *metric*

$$d(\mathbf{x},\mathbf{y}) = |\mathbf{x}-\mathbf{y}| = \sqrt{\sum_{i=1}^{3}(x_i - y_i)^2}.$$

Euclidean metrics on the space of functions.

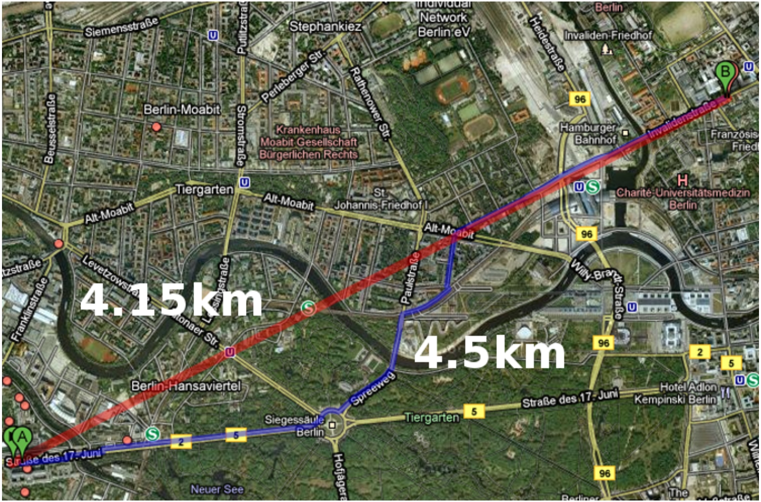
This generalizes to functions, if $f(t)$ and $g(t)$ are both real functions on the same interval, $[0, T]$ say, then the L^2-metric is

$$d(f, g) = \sqrt{\int_0^T dt(f - g)^2}.$$

Spike trains aren't a vector space.

- While it might be possible to define the addition of two spike trains by superposition, it isn't at all obvious how to define the difference.
- There is no reason to expect spike trains to be Euclidean.

A non-Euclidean metric: Metrics in towns.



‘As the crow flies’ distance versus route distance.

Metrics and spike trains.

- The framework for continuous version of information theory is a manifold, but perhaps that isn't needed, perhaps it can be rephrased in terms of metric spaces.

The van Rossum metric.

- A spike train is a list of spike times.

$$\mathbf{u} = \{u_1, u_2, \cdots, u_m\}$$

- Map spike trains to functions of t

$$\mathbf{u} \mapsto f(t; \mathbf{u}) = \sum_{i=1}^{m} h(t - u_i)$$

- $h(t)$ is a kernel, here, it is a causal exponential function

$$h(t) = \begin{cases} \exp(-t/\delta_T) & t > 0 \\ 0 & t \leq 0 \end{cases}$$

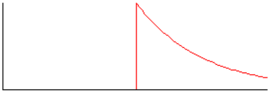
- Now

$$d(\mathbf{u}, \mathbf{v}) = \sqrt{\int dt [f(t; \mathbf{u}) - f(t; \mathbf{v})]^2}.$$

The van Rossum metric.

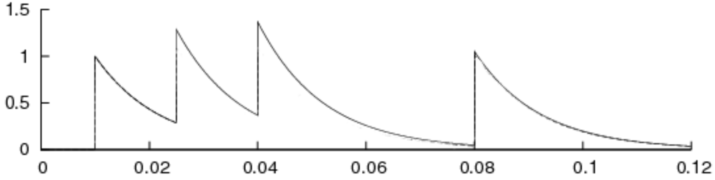
Two steps

- Maps from spike trains to functions using a filter.



- Use the metric on the space of functions.

The van Rossum metric.



Comparing metrics

The basic idea is to use the candidate metric to cluster a set of spike trains, and to compare this clustering with a "gold standard", namely, clustering the spike trains according to the stimuli that elicited them.

The scheme we use is a jack-knife calculation of a confusion matrix. The transmitted information $\tilde{h}$ is used to score clustering with one, the highest, corresponding to perfect clustering.

How would information theory work on the metric space of spike trains?

Imagine . . .

- First lets ask how it would look if there were coordinates for spike trains.
- Imagine there is a space of spike trains with coordinates and all that.

- Imagine there is a coordinate for each length L piece of spike train.

Imagine further . . .

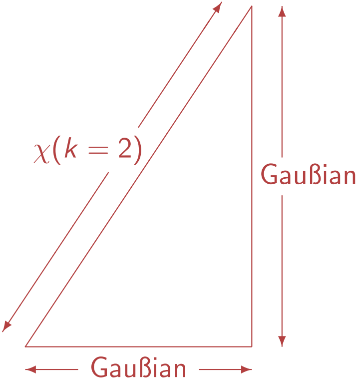
- Imagine that each variable has independent additive Gaussian noise.

$$X_i = Y_i + \eta$$

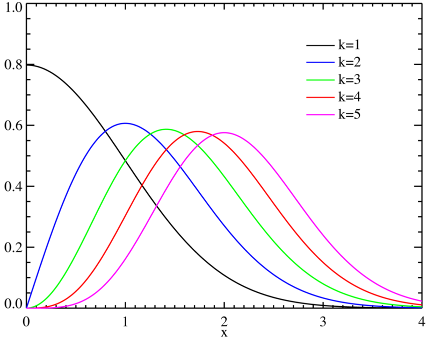
η

The χ-distribution.

- The distance between two such vectors satisfies a χ-distribution: $\mathbf{X} = (X_1, X_2, \ldots, X_k)$, $\mathbf{X}' = (X_1', X_2', \ldots, X_k')$ has $|\mathbf{X} - \mathbf{X}'| \sim \chi(\sigma, k)$.



The χ-distribution.



Idea!

- Turn this around![2]

[2] A metric space approach to the information channel capacity of spike trains Gillespie, JB and Houghton, CJ (2011) J. Comput. Neurosci. 30(1).

Idea!

- Turn this around![2]
  - Propose this as the distribution of distances.
  - Calculate k from the distribution and use this to work out L.

$$k = \frac{2\langle\zeta^2\rangle^2}{\langle\zeta^4\rangle - \langle\zeta^2\rangle^2}.$$

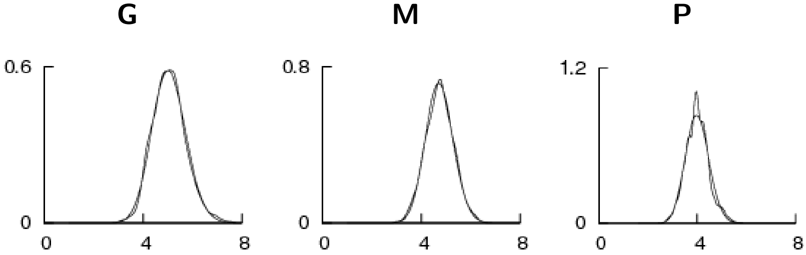
  - Use the noise model to calculate information.

[2] A metric space approach to the information channel capacity of spike trains Gillespie, JB and Houghton, CJ (2011) J. Comput. Neurosci. 30(1).

Idea!

- k is a sort of noise dimension or effective dimension.

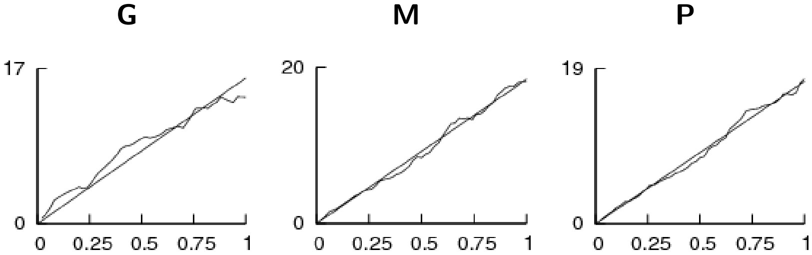
χ-distribution.



- Tested using the Anderson-Darling test.

k as a function of spike train length.

- k should increase linearly with sample length.



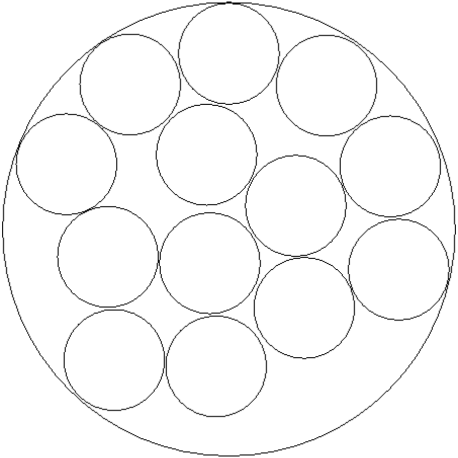
Channel capacity.

The channel capacity for a single Gaußian variable X is

$$C = \frac{1}{2}\log_2\left(1+\frac{\nu^2}{\sigma^2}\right) \text{ bits per time unit}$$

where σ^2 is the signal variance, usually taken to be the bound by the power constraint and ν^2 is the noise variance.

Information theory - this works.



- Model the spike train as a Gaußian channel but re-express the calculations in terms of distance based quantities!

Information theory - this works.

For example,

- If X and X' are iid Gaussian variables with variance σ^2 their difference is Gaussian with variance $\sigma_d^2 = 2\sigma^2$.

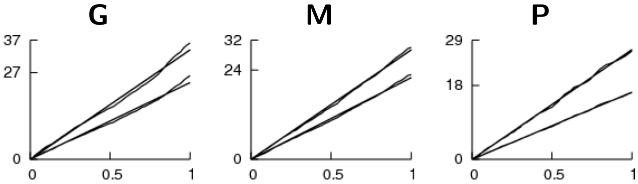
Channel capacity.

$$C = \frac{1}{2}\log_2\left(\frac{\xi_d^2}{\sigma_d^2}\right) \text{ bits per } L.$$

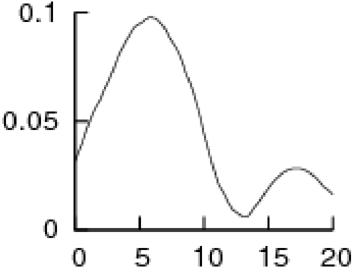
where ξ_d^2 is the signal variance and σ_d^2 the noise variance and $L = (\text{sample length})/k$.

Variances.

- ξ_d^2 and σ_d^2 are calculated by least squares fit.



Channel capacity of the cells we looked at.



Information theory on the metric space.

- The noise model fits the data we have.
- Seems to be the natural arena for information theory calculations.
- The channel capacity theory is about encoding discrete information in a continuous signal.
  - What we actually need is distortion theory.
- A multi-neuron version is needed for populations.
- Most of all, need to apply to more data.

More general conclusions.

- Information theory - what's the story with that?
- So, what is the space of spike trains?