Estimating Neural Signal Dependence Using Kernels

Nishant Mehta

`niche@cc.gatech.edu`

Georgia Institute of Technology

Motivation

Why dependence estimation?

Consider a few fundamental problems in neural analysis

- Learning which neurons share information, conditional on function/task
    - Discovering how the brain encodes task-relevant information
- Feature selection for machine learning
    - Select features that are most relevant to some task (e.g. Brain-computer interfaces)
    - Identifying how dependence changes over time
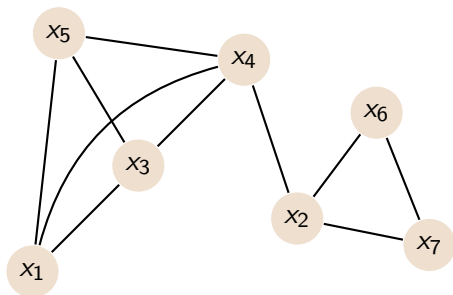
Rate coding versus temporal coding

Support for rate coding

- Complex cells tuned to specific orientations
  - Tuning function $\lambda(s)$ for stimulus s

Support for temporal coding

- Superior stimulus discriminability in the fly H1-neuron as compared to using spike counts
- Edit distance methods from Victor and Purpura
  - V1 and V2 in awake monkeys
- Firing rate difficult to code for low spike counts
  - Poor reaction time coding

Rate coding

Two sequences of spike times **x** and **y**

$\Downarrow$ (histogram with bin-width h)

Spike count over intervals

Tweak parameters:

- selection of bin-width
- where to start the first bin

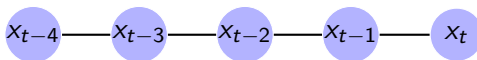# Rate coding

Intensity models:
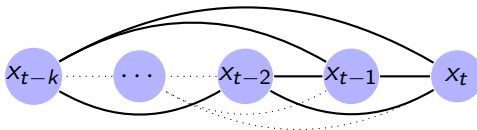
- i.i.d.

Rate coding

Intensity models:

- i.i.d.

- Markov

- Dependent on spike rates of last k intervals.

For second order dependence, consider:

$$D_2 = \mathsf{E}[(x - \mathsf{E}[x]) \otimes (y - \mathsf{E}[y])] = \mathsf{E}[(x - \mathsf{E}[x])(y - \mathsf{E}[y])^T]$$

$$\|D_2\|_F^2 = 0 \iff X \text{ and } Y \text{ uncorrelated}$$

Dependence in an RKHS

Use feature maps $\qquad x \to \phi(x) \in \mathcal{F}$

$\qquad\qquad\qquad\qquad y \to \psi(y) \in \mathcal{G}$

for RKHSs $\mathcal{F}$ and $\mathcal{G}$, with inner products $\quad k(x, x') = \langle \phi(x), \phi(x') \rangle$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad l(y, y') = \langle \psi(y), \psi(y') \rangle$

Hilbert-Schmidt Independence Criterion (HSIC) measures the cross-covariance between the kernelized x and y:

$$D_h = \mathsf{E}[(\phi(x) - \mathsf{E}[\phi(x)]) \otimes (\psi(y) - \mathsf{E}[\psi(y)])]$$

$$\|D_h\|_{\mathsf{HS}}^2 = 0 \iff X \perp\!\!\!\perp Y$$

[Gretton et al. 2005]

Dependence in an RKHS

By making use of the tensor product property $\langle x \otimes y \rangle z = x \langle y, z \rangle$, the Hilbert-Schmidt norm can be computed as

$$E_{xx'yy'}[k(x,x')l(y,y')] + E_{xx'}[k(x,x')l(y,y')]$$
$$- 2E_{xy}[E_{x'}[k(x,x')]E_{y'}[l(y,y')]]$$

The (biased) empirical estimator has the following compact form:

$$\frac{1}{T} \operatorname{tr} HKHL$$

where

K is the kernel matrix $K_{ij} = k(x_i, x_j)$

L is the kernel matrix $L_{ij} = l(y_i, y_j)$

$H = \delta_{ij} - \frac{1}{T}$

For $k(\cdot, \cdot)$ and $l(\cdot, \cdot)$ universal kernels [Steinwart 2002],
it can be shown that the cross-covariance of $\phi(x)$ and $\psi(y)$ is zero
if and only if x and y are independent random variables.

| Space for Cross-Covariance Estimation	Sensitivity
Original space	Second order dependence
Kernel space	All orders of dependence

Hilbert space distribution embeddings

Another view of this dependence measure may be more clarifying.

$P(x, y) = P(x) \times P(y)$ if and only if X and Y are statistically independent.

Suppose we had some injective mapping of distributions to a Hilbert space.

Dependence test:

- In new space, measure distance between joint and product of marginal distributions
- For X and Y independent, distance asymptotically $\rightarrow 0$
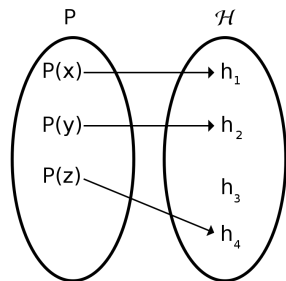
[Smola et al. 2007]

Hilbert space distribution embeddings

Suppose $X \sim p$.

For the feature map $x \rightarrow \phi(x) \in \mathcal{F}$, Smola et al. defined the expectation operator of the distribution p as

$$\mu[p] = \mathsf{E}_x[\phi(x)]$$

We similarly can embed the joint distribution $P(X, Y)$ using the mapping $(x, y) \rightarrow v((x, y), (\cdot, \cdot))$, where we factorize kernel $v((\cdot, \cdot), (\cdot, \cdot))$ as:

$$v((x, y), (x', y')) = k(x, x')l(y, y')$$

Suppose that we had an embedding of the distributions of X and Y into Hilbert spaces $\mathcal{F}$ and $\mathcal{G}$ such that we could compute the distances between the embeddings of the distributions.

Hilbert space distribution embeddings

Since an obvious measure is the difference between $P_{x,y}$ and $P_x \times P_y$, let's consider the distance between these distributions embeddings in an RKHS:

$$\|\mu[P_{xy}] - \mu[P_x \times P_y]\|^2$$

↑
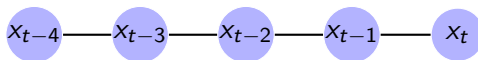Equivalent to Hilbert-Schmidt Independence Criterion from before
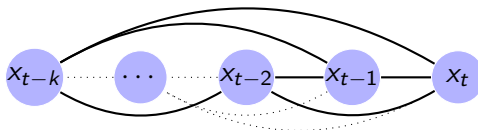
Rate coding

Intensity models:

- i.i.d.

$$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5$$

- Markov

$$x_{t-4} - x_{t-3} - x_{t-2} - x_{t-1} - x_t$$

- Dependent on spike rates of last k intervals.

$$x_{t-k} \cdots x_{t-2} - x_{t-1} - x_t$$

Dependence for graphical models

Suppose binned spike train data for bin-width h.
If we do not expect spike rate at time t to influence spike rate
beyond $t + \tau h$, then we can consider a graphical model with edges

$$(x_s, x_t), (x_s, y_t), (y_s, y_t) \text{ for } |s - t| \leq \tau.$$

Sufficient statistics decompose along the maximal cliques of this
graphical model

$$\Downarrow$$

Compute the dependence within each of the maximal cliques in
decomposition separately.

[Altun et al. 2004]

Dependence for graphical models

The graph's maximal cliques are

$$(x_i, \ldots, x_{i+\tau}, y_i, \ldots, y_{i+\tau}) \quad \text{for } 1 \leq i \leq T - \tau.$$

We kernelize the maximal cliques and consider each realization of the variables in a maximal clique as a sample.

$$\sum_{c \in \mathcal{C}} \|\mu_c[P_c(x_c, y_c)] - \mu_c[P_c(x_c)P_c(y_c)]\|_{\mathsf{HS}}^2$$

The empirical estimator is similar to the form before ($\frac{1}{T} \operatorname{tr} HKHL$), with kernel matrices now representing inner products of the clique variables in each multivariate's respective kernel space, and the sums going over all pairs of the maximal clique samples.

Extension to fMRI data

- Measuring dependence in spike rate data
- Similar setting: Measuring dependence among fMRI voxels
    Explored in Gretton et al. (2006) for Macaque monkey visual cortex
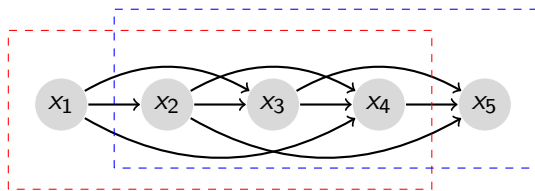
Extension to fMRI data

- Measuring dependence in spike rate data
- Similar setting: Measuring dependence among fMRI voxels
    Explored in Gretton et al. (2006) for Macaque monkey visual cortex

Non-iid

- Temporal dependence in neural activation
- HRF convolution-induced latency

Vector space embeddings of spike trains

Suppose that we have two sequences of spike times:

$$\mathbf{x} = x(1), x(2), \ldots, x(m_x)$$
$$\mathbf{y} = y(1), y(2), \ldots, y(m_y)$$

We use $x(i)$ to refer both to the ith spike and its corresponding time of occurrence.

Spike $y(j)$ directly precedes $x(i)$ if $j = \max_{j'}\{j' : y(j') < x(i)\}$.

If $y(j)$ directly precedes $x(i)$,
then we refer to $y(j - k)$ as y's k^{th} preceding spike of $x(i)$.

Spike history space

Let $\mathbf{x}^{(i)}$ be a realization of a random variable $X = (x_1, x_2, \ldots, x_\tau)$, where x_j is time difference between $x(i)$ and x's j^{th} preceding spike of $x(i)$.

Note that x's k^{th} preceding spike of $x(i)$ is simply $x(i - k)$.

Similarly, let $\mathbf{y}^{\mathbf{i}}$ be a realization of a R.V. $Y = (y_1, y_2, \ldots, y_\tau)$, where y_j is the time difference between $x(i)$ and y's j^{th} preceding spike of $x(i)$.

Probabilistic model

Consider the joint densities $P(\mathbf{x^i})$ and $P(\mathbf{y^i})$.

If underlying processes generating spike times $\mathbf{x}$ and $\mathbf{y}$ independent, then for reference spike $x(i)$:

Time since each of x's last τ spikes preceding $x(i)$

$$\perp\!\!\!\perp$$

Time since each of y's last τ spikes preceding $x(i)$

More generally, we have $X \perp\!\!\!\perp Y$.

Samples from the spike history space

Similar to structured HSIC method, we use τ to impose a finite dependence history to bound the dimensionality of our space. We have joint samples $(x_{t_0}, y_{t_0}), (x_{t_0+1}, y_{t_0+1}), \ldots, (x_{m_x}, y_{m_x})$. For computing the independence test, we consider

$$\|\mu[P(X, Y)] - \mu[P(X)P(Y)]\|_{\mathsf{HS}}^2$$

with empirical estimator $\frac{1}{T} \operatorname{tr} HKLH$ for summation over all T existing pairs (x_t, y_t).

Samples from the spike history space

Similar to structured HSIC method, we use τ to impose a finite dependence history to bound the dimensionality of our space.
We have joint samples $(x_{t_0}, y_{t_0}), (x_{t_0+1}, y_{t_0+1}), \ldots, (x_{m_x}, y_{m_x})$.
For computing the independence test, we consider

$$\|\mu[P(X, Y)] - \mu[P(X)P(Y)]\|_{\mathsf{HS}}^2$$

with empirical estimator $\frac{1}{T} \operatorname{tr} HKLH$ for summation over all T existing pairs (x_t, y_t).
The above samples are derived from referencing to spikes in spike train **x**.
Also, need to repeat the above for alternate X and Y which are referenced to spikes in spike train **y**.

Sign language recognition

- Sign language recognition using fMRI scanning of motor cortex
- Feature space, $64 \times 64 \times 3$ voxels $> 10^4$.
- Hidden Markov models work fairly well for language recognition tasks
  - Need lower dimensionality!

Very roughly, backward elimination where we retain the features whose removal significantly reduces the dependence between the active set and the labels.

Decomposing signs into smaller motor units provides a method for doing feature selection for individual signs.

Future work and open problems

- Experiments!
  - Spike sequence data (synthetic and actual)
    - Spike history space
    - Firing rate encoding
  - fMRI data
    - Voxel-stimulus dependence estimation
    - Conditional on task, voxel-voxel dependence estimation

Future work and open problems

- Experiments!
    - Spike sequence data (synthetic and actual)
        - Spike history space
        - Firing rate encoding
    - fMRI data
        - Voxel-stimulus dependence estimation
        - Conditional on task, voxel-voxel dependence estimation

Open

1. Extending HSIC to dependency among 3 or more neurons.
2. Improvements on spike history space?

Questions?

Answers?