Neural Decoding

Matthias Hennig

School of Informatics, University of Edinburgh

January 2019

⁰Acknowledgements: Mark van Rossum, Chris Williams and slides from Gatsby Liam Paninski.

Decoding brain activity



The Homunculus



Flynculus [Rieke et al., 1996]

- Stimulus discrimination, signal detection theory
- Maximum likelihood and MAP decoding
- Bounds and Fisher information
- Spike train decoding and GLMs

Understanding the neural code.

- Given spikes, what was the stimulus?
- What aspects of the stimulus does the system encode? (capacity is limited)
- What information can be extracted from spike trains:
 - By "downstream" areas? Homunculus.
 - By the experimenter? Ideal observer analysis.
- What is the coding quality?
- Design of neural prosthetic devices

Related to encoding, but encoding does not answer above questions explicitly.

- Hippocampal place cells: how is location encoded?
- Retinal ganglion cells: what information is sent to the brain? What is discarded?
- Motor cortex: how can we extract as much information as possible from a collection of M1 cells?

Probability of the stimulus, prior: P(s)Probability of a measured neural response: $P(\mathbf{r})$ Joint probability of stimulus and response: $P(\mathbf{r}, r)$ Conditional probabilities: $P(\mathbf{r}|s)$, $P(s|\mathbf{r})$ Marginal: $P(\mathbf{r}) = \sum_{s} P(\mathbf{r}|s)P(s)$ Note: $P(\mathbf{r}, s) = P(\mathbf{r}|s)P(s) = P(s|\mathbf{r})P(\mathbf{r})$

Bayes theorem:

$$P(s|\mathbf{r}) = rac{P(\mathbf{r}|s)P(s)}{P(\mathbf{r})}$$

Note that we need to know the stimulus statistics.

Example: Discrimination between two stimuli



Test subject report left or right motion (2 alternative forced choice, 2AFC). See [Dayan and Abbott, 2002], chapter 3.2.

MT neurons in this task



[Britten et al., 1992]

- Some single neurons do as well as animal!
- Possibility for averaging might be limited due to correlation?
- Population might still be better/faster? [Cohen and Newsome, 2009]



[Britten et al., 1992] Assuming rate histograms are Gaussian with equal variance σ^2 , the discriminability is

$$d' = \frac{\langle r \rangle_+ - \langle r \rangle_-}{\sigma}$$

Discriminate between response distributions $P(r_{-})$ and $P(r_{+})$. (directions + and -), and discrimination threshold *z* on firing rate:

• Hit rate:
$$\beta(z) = P(r \ge z|+)$$

• False alarm rate: $\alpha(z) = P(r \ge z|-)$

stimulus	correct	False
+	β	$1 - \beta$
-	$1 - \alpha$	α

Probability of correct answer: $(\beta(z) + 1 - \alpha(z))/2$ Can be used to find the optimal *z*.

ROC curves

Discriminate between response distributions $P(r_{-})$ and $P(r_{+})$. The Receiver Operating Characteristic (ROC) gives graphical intuition:



- Vary decision threshold and measure error rates.
- Larger area under curve means better discriminability.
- Shape relates to underlying distributions.



[Britten et al., 1992]

$$P(ext{correct}) = \int_0^1 eta dlpha$$

When responses are Gaussian:

$$P(\text{correct}) = \frac{1}{2} \operatorname{erfc} \left(\frac{\langle r \rangle_{-} - \langle r \rangle_{+}}{2\sigma} \right)$$

Population decoding





[Dayan and Abbott (2001) after Theunissen and Miller (1991)]

Cricket Cercal System: Information about wind direction is encoded by four types of neurons

$$\left(rac{f(s)}{r_{\max}}
ight) = [\cos(s - s_a)]_+$$

 Let c_a denote a unit vector in the direction of s_a, and v be a unit vector parallel to the wind velocity

$$\left(rac{f(s)}{r_{\max}}
ight) = [\mathbf{v} \cdot \mathbf{c}_a]_+$$

- Crickets are Cartesian, 4 directions 45° , 135° , -135° , -45°
- Population vector is defined as

$$\mathbf{v}_{pop} = \sum_{a=1}^{4} \left(\frac{r}{r_{\max}} \right)_{a} \mathbf{c}_{a}$$



[Dayan and Abbott (2001) after Salinas and Abbott (1994)]

- Certain neurons in M1 of the monkey can be described by cosine functions of arm movement direction (Georgopoulos et al, 1982)
- Similar to cricket cercal system, but note:
 - Non-zero offset rates r₀

$$\left(\frac{f(s)-r_0}{r_{\max}}\right)=\mathbf{v}\cdot\mathbf{c}_a$$

 Non-orthogonal: there are many thousands of M1 neurons that have arm-movement-related tuning curves



$$p(s|\mathbf{r}) = rac{p(\mathbf{r}|s)p(s)}{p(\mathbf{r})}$$

- Maximum likelihood decoding (ML): $\hat{s} = \operatorname{argmax}_{s} p(\mathbf{r}|s)$
- Maximum a posteriori (MAP): $\hat{s} = \operatorname{argmax}_{s} p(s)p(\mathbf{r}|s)$
- Note these two are equivalent if *p*(*s*) is flat.
- Bayes: mimimize loss

$$s_B = \operatorname{argmin}_{s^*} \int_s L(s, s^*) p(s|\mathbf{r}) ds$$

For squared loss L(s, s^{*}) = (s − s^{*})², optimal s^{*} is posterior mean, s_B = ∫_s p(s|r)s.

• For the cercal system, assuming indep. noise

$$p(\mathbf{r}|s) = \prod_{a} p(r_{a}|s)$$

where each $p(r_a|s)$ is modelled as a Gaussian with means and variances

- p(s) is uniform (hence MAP=ML)
- ML decoding finds a peak of the likelihood
- Bayesian method finds posterior mean
- These methods improve performance over the vector method (but not that much, due to orthogonality...)

Cricket Cercal System



[Dayan and Abbott (2001) after Salinas and Abbott (1994)]

General Consideration of Population Decoding



[Dayan and Abbott (2001)]

Gaussian tuning curves.

Poisson firing model over time *T*, count $n_a = r_a T$ spikes.

$$p(\mathbf{r}|s) = \prod_{a=1}^{N} \frac{(f_a(s)T)^{n_a}}{n_a!} \exp(-f_a(s)T)$$
$$\log p(\mathbf{r}|s) = \sum_{a=1}^{N} n_a \log f_a(s) + \dots$$

The terms in ... are independent of *s*, and we assume $\sum_{a} f_{a}(s)$ is independent of *s* (all neurons sum to the same average firing rate).

• s_{ML} is stimulus that maximizes $\log p(\mathbf{r}|s)$, determined by

$$\sum_{a=1}^{N} r_a \frac{f_a'(s_{ML})}{f_a(s_{ML})} = 0$$

• If all tuning curves are Gaussian $f_a = A \exp[-(s - s_a)^2/2\sigma_w^2]$ then

$$s_{ML} = rac{\sum_a r_a s_a}{\sum_a r_a}$$

which is simple and intuitive, known as Center of Mass (cf population vector)

• Bias and variance of an estimator sest

$$egin{array}{rcl} b_{est}(s) &=& \langle s_{est}
angle - s \ \sigma^2_{est}(s) &=& \langle (s_{est} - \langle s_{est}
angle)^2
angle \ \langle (s-s_{est})^2
angle &=& b^2_{est}(s) + \sigma^2_{est} \end{array}$$

• Thus for an unbiased estimator, MSE $\langle (s - s_{est})^2 \rangle$ is given by σ_{est}^2 , the variance of the estimator

Fisher information

 Fisher information is a measure of the curvature of the log likelihood near its peak

$$I_{F}(s) = \left\langle -\frac{\partial^{2} \log p(\mathbf{r}|s)}{\partial s^{2}} \right\rangle_{s} = -\int d\mathbf{r} p(\mathbf{r}|s) \frac{\partial^{2} \log p(\mathbf{r}|s)}{\partial s^{2}}$$

(the average is over trials measuring r while s is fixed)

 Cramér-Rao bound says that for any estimator [Cover and Thomas, 1991]

$$\sigma_{est}^2 \ge rac{(1+b_{est}'(s))^2}{I_F(s)}$$

- efficient estimator if $\sigma_{est}^2 = \frac{(1+b'_{est}(s))^2}{I_F(s)}$. In the bias-free case an efficient estimator $\sigma_{est}^2 = 1/I_F(s)$.
- ML decoder is typically efficient when $N \to \infty$.

- In homogeneous systems I_F indep. of s.
- More generally Fisher matrix $(I_F)_{ij}(\mathbf{s}) = \left\langle -\frac{\partial^2 \log p(\mathbf{r}|s)}{\partial s_i \partial s_i} \right\rangle_c$.
- Taylor expansion of Kullback-Leibler $D_{KL}(P(\mathbf{s}), P(\mathbf{s} + \delta \mathbf{s})) \approx \sum_{ij} \delta s_i \delta s_j (I_F)_{ij}$
- Not a Shannon information measure (not in bits), but related to Shannon information in special cases,e.g.
 [Brunel and Nadal, 1998, Yarrow et al., 2012].

Fisher information for a population

For independent Poisson spikers

$$I_{F}(s) = \left\langle -\frac{\partial^{2} \log p(\mathbf{r}|s)}{\partial s^{2}} \right\rangle = T \sum_{a} \langle r_{a} \rangle \left(\left(\frac{f_{a}'(s)}{f_{a}(s)} \right)^{2} - \frac{f_{a}''(s)}{f_{a}(s)} \right)$$

For dense, symmetric tuning curves, the second term sums to zero. Using $f_a(s) = \langle r_a \rangle$ we obtain

$$I_F(s) = T \sum_a rac{(f_a'(s))^2}{f_a(s)}$$

For dense $f_a(s) = Ae^{-(s-s0+a.ds)^2/2\sigma_w^2}$ with density $\rho = 1/ds$, sum becomes integral

$$I_F = \sqrt{2\pi} T A \rho / \sigma_w$$

For Gaussian tuning curves



[Dayan and Abbott (2001)]

- Note that Fisher information vanishes at peak as $f'_a(s) = 0$ there.
- Can be used to create optimal tuning curves, [Dayan and Abbott, 2002] chapter 3.3.
- Discriminability $d' = \Delta s \sqrt{I_F(s)}$ for a small ΔF .

FI predicts human performance



[Dayan and Abbott (2001)]

- Orientation discrimination for stimuli of different size (different N)
- Solid line: estimated minimum standard deviation at the Cramer Rao bound
- Triangles: human standard deviation as function of stimulus size (expressed in *N*)

Slope as strategy

Fig. 4. Prediction for other sensory systems (olfaction). Color map, schematic odor trail; gray line, path of an organism that followed the trail's peak concentration. This strategy is typically assumed for odor-trail following (3). Black line, path of the same organism when using a strategy similar to that of our bats, that is, following the maximum slope of the odorant concentration (17). The movement jitter in this case is smaller, making the tracking smoother and therefore faster.





From paper on bat echo location [Yovel et al., 2010])

- Dayan and Abbott §3.4
- Estimate the stimulus from spike times t_i to minimize e.g. $\langle s(t) s_{est}(t) \rangle^2$
- First order reconstruction:

$$m{s}_{est}(t- au_0) = \sum_{t_i} m{K}(t-t_i) - \langle r
angle \int m{d} au m{K}(au)$$

- The second term ensures that $\langle s_{est}(t) \rangle = 0$
- Delay τ₀ can be included to make decoding easier: predict stimulus at time t - τ₀ based on spikes up to time t

- Organism faces causal (on-line) decoding problem.
- Prediction of the current/future stimulus requires temporal correlation of the stimulus. Example: in head-direction system neural code correlates best with future direction.
- Requires $K(t t_i) = 0$ for $t \le t_i$.

$$m{s}_{est}(t- au_0) = \sum_{t_i} m{K}(t-t_i) - \langle r
angle \int m{d} au m{K}(au)$$

• Delay τ_0 buys extra time

Causal decoding



Delay $\tau_0 = 160$ ms. (B: shifted/causal kernel, C: non-causal kernel) At time *t* estimate $s(t - \tau 0)$:

Spikes 1..4: contribute because stimulus is correlated (right tail of K) Spikes 5..7: contribute because of τ_0

Spikes 8, 9,... : have not occurred yet. Stimulus uncorrelated: Kernel from STA [Dayan and Abbott (2001)]

Decoding from a GLM



- H1 neuron of the fly
- Solid line is reconstruction using acausal filter

 Note, reconstruction quality will depend on stimulus



[Dayan and Abbott (2001) after Rieke et al (1997)]

According to Bayes theorem

$$p(s|r) = p(r|s)p(s)/p(r)$$

$$\log p(s|r) = \log p(r|s) + \log p(s) + c$$

Tractable as long as rhs. is concave, which excludes heavy tailed p(s). Requires to compute:

$$\hat{s} = \operatorname{argmax}_{s} \log p(s|r)$$

This is numerically hard, see [Pillow et al., 2011] for methods.

MAP decoding simulation



[Pillow et al., 2011]

- Stimulus reconstruction similar to encoding problem. But
 - Response is given, can not be choosen to be white
 - Imposing causality adds realism but reduces quality
- The reconstruction problem can be ill-posed. It is not always possible to reconstruct stimulus (cf dictionary). For instance: complex cell.

Still, the cell provides information about the stimulus. Could try to read the code, rather than reconstruct the stimulus (e.g. ideal observer)

[Hung et al., 2005]

- $\bullet\,$ Recording from \sim 300 sites in the Inferior Temporal (IT) cortex
- Present images of 77 stimuli (of different objects) at various locations and scales in the visual field.
- Task is to categorize objects into 8 classes, or identify all 77 objects
- Predictions based on one-vs-rest linear SVM classifiers, using data in 50 ms bins from 100 ms to 300 ms after stimulus onset

Fig. 1. Accurate readout of object category and identity from IT population activity. (A) Example of multi-unit spiking responses of 3 independently recorded sites to 5 of the 77 objects. Rasters show spikes in the 200 ms after stimulus onset for 10 repetitions



(black bars indicate object presentation). (B) Performance of a linear classifier over the entire object set on test data (not used for training) as a function of the number of sites for reading out object category (red, chance = 1.25%) or identity (blue, chance = 1.3%). The input from each site was the spike count in consecutive 50-ms bins from 100 to 300 ms







[Hung et al., 2005]

What does this tell us?

- Performance of classifiers can provide a lower bound on the information available in the population activity.
- Assuming independence, correlations are ignored. Correlation could limit or enhance information.
- Distributed representation
- Linear classifier can plausibly be implemented in neural hardware

- Dayan and Abbott §3.3
- Population encoding uses a large number of neurons to represent information
- Advantage 1: reduction of uncertainty due to neuronal variability (Improves reaction time).
- Advantage 2: Ability to represent a number of different stimulus attributes simultaneously (e.g. in V1 location and orientation).

Population codes and correlations: Retina

Fit coupled GLM-model (see encoding) to retina data



[Pillow et al., 2008]

Population codes and correlations: Retina



Bayesian decoding

[Pillow et al., 2008]

[Brown et al., 1998]

- Encoding in place cells: modelled as inhomogeneous Poisson processes with Gaussian receptive fields (incl. theta oscillations), assuming independence
- Encoding as path: random walk
- Bayesian filter decoding: compute posterior and then update estimate in next time step using information from spikes (combine prior and new information).
- Non-Bayesian decoding: compute estimate in time steps given spikes in time window (only current information).



[Brown et al., 1998]

Bayes, maximum likelihood, linear kernel and correlation based.

[Shpigelman et al., 2005]

- Rhesus monkey, 43 electrodes in M1
- Monkey controls cursors on a screen using two manipulanda to perform a centre-out reaching task
- Predict hand velocity based on 10 time bins, each of length 100 ms in all 43 neurons.
- Can use linear regression, polynomial regression, Gaussian kernel (support vector regression), spikernel (allows time warping)
- More sophisticated methods outperform linear regression, but linear is already decent

State-of-the-art w. Kalman filters [Gilja et al., 2012]



[[]Shpigelman et al., 2005]

Decoding from your brain



Time [2 s / image]

[Miyawaki et al., 2008]

fMRI, linear decoder from voxel activations.

- Discrimination between stimuli (or actions) using just recorded spikes is possible and tractable.
- Full reconstruction of a stimulus is hard, especially when stimulus dimensionality is high.
- But also unclear to what extent this is this even possible.
- Decoding can tell us how much information neural activity carries about the outside world, it provides an upper bound.
- BMI applications are on the horizon.

References I



Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci*, 12:4745–4765.



Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C., and Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, 18(18):7411–7425.



Brunel, N. and Nadal, J.-P. (1998).

Mutual information, Fisher information, and population coding. *Neural Comp.*, 10:1731–1757.



Cohen, M. R. and Newsome, W. T. (2009).

Estimates of the contribution of single neurons to perception depend on timescale and noise correlation.

J Neurosci, 29(20):6635–6648.



Cover, T. M. and Thomas, J. A. (1991). *Elements of information theory*. Wiley, New York.



Dayan, P. and Abbott, L. F. (2002). *Theoretical Neuroscience*. MIT press, Cambridge, MA.

References II

Gilja, V., Nuyujukian, P., Chestek, C. A., Cunningham, J. P., Yu, B. M., Fan, J. M., Churchland, M. M., Kaufman, M. T., Kao, J. C., Ryu, S. I., and Shenoy, K. V. (2012). A high-performance neural prosthesis enabled by control algorithm design. *Nat Neurosci*, 15(12):1752–1757.



Hung, C. P., Kreiman, G., Poggio, T., and DiCarlo, J. J. (2005). Fast Readout of Object Identity from Macaque Inferior Temporal Cortex. *Science*, 310:863–866.



Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M.-a., Morito, Y., Tanabe, H. C., Sadato, N., and Kamitani, Y. (2008).

Visual image reconstruction from human brain activity using a combination of multiscale local image decoders.

Neuron, 60(5):915-929.



Pillow, J. W., Ahmadian, Y., and Paninski, L. (2011).

Model-based decoding, information estimation, and change-point detection techniques for multineuron spike trains.

Neural computation, 23(1):1–45.



Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli, E. P. (2008).

Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999.

	_	ς.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1996). *Spikes: Exploring the neural code.* MIT Press, Cambridge.



Shpigelman, L., Singer, Y., Paz, R., and Vaadia, E. (2005).

Spikernels: Predicting Arm Movements by Embedding Population Spike Rate Patterns in Inner-Product Spaces.

Neural Comput, 17:671-690.



Yarrow, S., Challis, E., and Seri?s, P. (2012). Fisher and Shannon information in finite neural populations. *Neural Comput*, 24(7):1740–1780.



Yovel, Y., Falk, B., Moss, C. F., and Ulanovsky, N. (2010). Optimal localization by pointing off axis. *Science*, 327(5966):701–704.