Neural Decoding

Theoretical Neuroscience Ch 3 Notes (Dayan and Abbott)

Neha Peddinti

Dec 2021

Contents

1 Goal

1. Previously: Predicted neural response for a given stimulus function (Ch 2). This was encoding (determining $P[\mathbf{r}|s]$).

2. Now: Predict stimulus s from the firing rates $\mathbf{r} = (r_1, ..., r_N)$ for N neurons. This is decoding (determining $P[s|\mathbf{r}]$).

3. Methods of predicting the stimulus:

   (a) **Discrimination:** Discriminate between two different possible stimulus values by decoding the response of a single cell.

   (b) **Population decoding:** Find a parameter that predicts the value of a static stimulus from the responses of a population of neurons.

   (c) **Spike-train decoding:** Find stimulus as a function of time from the spike train it prompted the neuron to generate.

2 Probability for Stochastic Model of Stimuli

1. **Prior probability**: $P[s]$, the probability of stimulus s occurring (this probability distribution is guessed).

2. **Joint probability**: $P[\mathbf{r}, s]$, the probability of stimulus s and response $\mathbf{r}$ both occurring.

3. $P[\mathbf{r}|s]$, the conditional probability of response $\mathbf{r}$ given stimulus s.

4. **Posterior distribution of s**: $P[s|\mathbf{r}]$, the conditional probability of s given $\mathbf{r}$.

5. $P[\mathbf{r}]$, the probability of response $\mathbf{r}$ occurring, independent of the stimulus used.

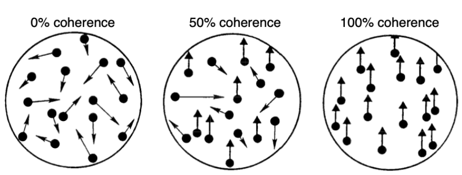
$$P[\mathbf{r}] = \sum_s P[\mathbf{r}|s]P[s] \quad P[s] = \sum_{\mathbf{r}} P[s|\mathbf{r}]P[\mathbf{r}]$$

6. Bayes theorem: Use $P[\mathbf{r}, s] = P[\mathbf{r}|s]P[s] = P[s|\mathbf{r}]P[\mathbf{r}]$ to derive the theorem

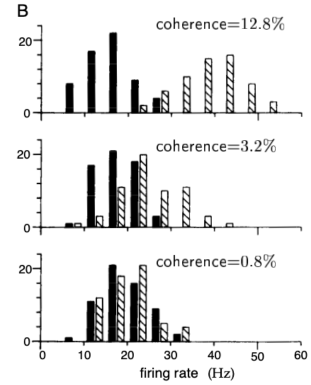
$$P[s|\mathbf{r}] = \frac{P[\mathbf{r}|s]P[s]}{P[\mathbf{r}]}$$

3 Discrimination

1. Britten et al. (1992) experiment:



Monkey was supposed to determine the overall direction of movement ("plus/+" for the neuron's preferred direction of movement, or "minus/-") when the random dot movement was shown at different coherence levels. Single neuron's response was measured.



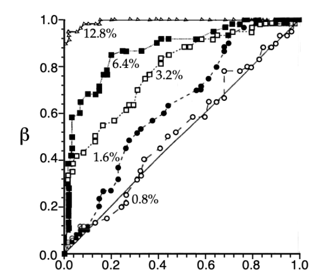
Above diagram shows the neuron's firing rates when shown stimuli in the plus (hatched histogram) and minus (solid histogram) directions (approximately Gaussian distributions).

2. ROC analysis

   (a) False alarm rate $\alpha(z)$ (size) and hit rate $\beta(z)$ (power): Choose some threshold z such that you predict a "plus" stimulus if $r \geq z$; otherwise you predict a "minus" stimulus (which works due to the separated nature of the two histograms above).

$$\alpha(z) = P[r \geq z|-] \quad \text{is the size/false alarm rate.}$$
$$\beta(z) = P[r \geq z|+] \quad \text{is the power/hit rate.}$$

   (b) Use **ROC (receiver operating characteristic) curves** to evaluate the choice of z.
   $z = 0$: r is always positive, so $\alpha = \beta = 1$.
   $z \to \infty$: r is always less than a large z value, so $\alpha = \beta = 0$.
   Vary the values of z to generate β vs α ROC curves for different coherence values. The diagram shows ROC curves for the dot experiment (Britten et al., 1992).



   (c) **Two-alternative force-choice task:** Given that the stimulus is presented twice (one "plus", one "minus") and generates firing rates r_1 and r_2, determine the order of the stimuli (plus-then-minus, or minus-then-plus).
   If trial 1 is "plus": $P[r_1 \geq r_2|+] = \beta(r_2)$.

Probability that r_2 is in the interval $(z, z+\Delta z)$, given that trial 2 had the "minus" stimulus (since trial 1 was "plus"): $p[z|-]\Delta z$.
Probability of correct answer in general (integrating over all values of $r_2 = z$ to get the expected value of $\beta(z)$):

$$P[\text{correct}] = \int_0^\infty dz\, p[z|-]\beta(z).$$

$P[\text{correct}]$ measures the probability that $r_1 \geq r_2$, independent of $r_2 = z$, given that trial 1 is "plus." If given that trial 1 is "minus," switch r_1 and r_2 for same result. Again the justification for this is:

$$\begin{aligned} &P[\text{correct}] = P[r_1 \geq r_2|\text{Trial 1 is } +] \\ &= P[\text{guess gives correct answer}|\text{correct answer}] \end{aligned}$$

(d) Area under the ROC curve is the probability of responding correctly in the two-alternative forced-choice test:

$$\begin{aligned} P[r \geq z|-] = \alpha(z) &= \int_z^\infty dr\, [r|-] \\ \frac{\mathrm{d}\alpha}{\mathrm{d}z} &= -p[z|-] \to dz\, p[z|-] = -d\alpha \\ P[\text{correct}] &= \int_0^\infty dz\, p[z|-]\beta(z) \\ &= \int_1^0 -d\alpha\, \beta(z) = \int_0^1 d\alpha\, \beta(z) \end{aligned}$$

(e) For Gaussian functions $p[r|\pm]$ with means $\langle r\rangle_-$ and $\langle r\rangle_+$ and same variance σ_r^2:

$$\begin{aligned} \alpha(z) &= \frac{1}{2}\text{erfc}\left(\frac{z - \langle r\rangle_-}{\sqrt{2}\sigma_r}\right) \\ \beta(z) &= \frac{1}{2}\text{erfc}\left(\frac{z - \langle r\rangle_+}{\sqrt{2}\sigma_r}\right) \\ P[\text{correct}] &= \frac{1}{2}\text{erfc}\left(\frac{\langle r\rangle_- - \langle r\rangle_+}{2\sigma_r}\right) = \frac{1}{2}\text{erfc}\left(-\frac{d'}{2}\right) \end{aligned}$$

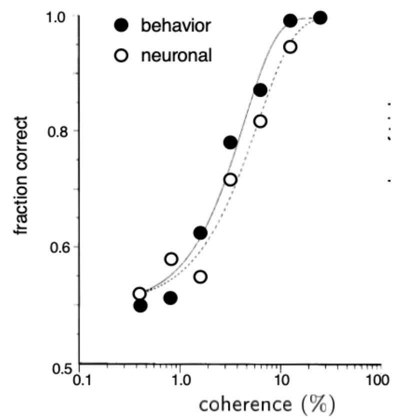
where erfc is the complementary error function

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt,$$

and d' is the discriminability

$$d' = \frac{\langle r\rangle_+ - \langle r\rangle_-}{2}.$$

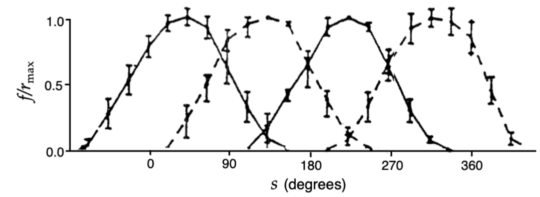
(f) Turning the previous experiment into a two-alternative forced-choice task:
Show a neuron the "plus" stimulus, record response r_1, then show it the "minus" stimulus and record r_2.
Find $P[\text{correct}]$ by calculating the area under the ROC curve for different coherence levels.
If these values match the fraction of trials that the monkeys get correct (which they somewhat do; see diagram below), then this model of Gaussian distributions and thresholds seems reasonably close to the monkey's actual neural process.



3. Neyman-Pearson lemma and likelihood ratio. **Come back to this!!, and Exercise 1.**
   (a) Why the data of Britten et al. allowed the firing-rate threshold test (instead of the likelihood ratio test) to be optimal; using logs.
   (b) Show that the slope of the ROC curve is equal to the likelihood ratio (and what this means intuitively).
   (c) Introducing a loss function and how to minimize expected loss.
   (d) Finding $P[+|r]$.
   (e) Likelihood ratio test for small Δs between the two stimulus values between which you want to discriminate.
4. For populations of neurons, similar techniques apply by terms such as $p[r|s]$ are replaced with $p[\mathbf{r}|s]$.

4 Population Decoding

1. Vector method of decoding
   (a) Cricket cercal system problem (Theunissen and Miller, 1991):
   4 neurons shown to have perpendicular preferred wind directions s_a (as shown in the tuning curve graphs below). Use firing rates of the population of neurons to predict wind direction.



   (b) Tuning curve for interneuron a with preferred wind direction angle s_a (which has the vector form $\mathbf{c_a}$):

$$\left(\frac{f(s)}{r_{\max}}\right)_a = [(\cos(s - s_a))]_+ = [\mathbf{v}\cdot\mathbf{c_a}]_+,$$

   . which gives the predicted firing rate of neuron a when the wind direction is $\mathbf{v}$ (or has angle s).

(c) Population vector to estimate wind direction (derived from Cartesian coordinates):

$$\mathbf{v}_{\text{pop}} = \sigma_{a=1}^{4} \left(\frac{r}{r_{\max}} \right) \mathbf{c_a},$$

. which essentially projects the predicted wind direction $\mathbf{v}$ (based on firing rate r using the tuning curve above) onto each neuron's preferred direction $\mathbf{c_a}$, and adds the projections together.

(d) Including offset firing rate r_0 in some types of neurons (such as primary motor cortex, M1, neurons in a monkey):

$$\left(\frac{\langle r \rangle - r_0}{r_{\max}} \right)_a = \left(\frac{f(s) - r_0}{r_{\max}} \right)_a = \mathbf{v} \cdot \mathbf{c_a}.$$

(e) Population vector for a large number of neurons N (more than 4, so directions definitely don't form Cartesian coordinate axes in this case, but $\mathbf{v}_{\text{pop}}$ is still parallel to $\mathbf{v}$ for large N):

$$\mathbf{v}_{\text{pop}} = \sum_{a=1}^{N} \left(\frac{r - r_0}{r_{\max}} \right)_a \mathbf{c_a}$$

Averaging over several trials:

$$\langle \mathbf{v}_{\text{pop}} \rangle = \sum_{a=1}^{N} (\mathbf{v} \cdot \mathbf{c_a}) \mathbf{c_a}.$$

2. Bayesian inference: minimizing a loss function.

(a) Loss function: $L(s, s_{\text{bayes}})$, measures cost of reporting s_{bayes} when the actual stimulus is s.
To minimize loss, minimize: $\int dsL(s, s_{\text{bayes}})p[s|\mathbf{r}]$.

(b) Stimulus prediction: s_{bayes}

$$L(s, s_{\text{bayes}}) = (s - s_{\text{bayes}})^2 \to s_{\text{bayes}} = \text{mean of } p[s|\mathbf{r}]$$
$$L(s, s_{\text{bayes}}) = |s - s_{\text{bayes}}| \to s_{\text{bayes}} = \text{median}$$

3. Maximum a posteriori (MAP) inference: maximize conditional probability density of the stimulus.

4. Maximum likelihood (ML) inference: MAP inference when $p[s]$ is independent of s.
Using the procudure outlined in the Summary section, the stimulus estimate S_{ML} ends up as

$$S_{ML} = \frac{\sum r_a s_a / \sigma_a^2}{\sum r_a \sigma_a^2},$$

where r_a is the measured firing rate of neuron a, and s_a is the neuron's preferred stimulus value (the stimulus corresponding with the max value of the neuron's tuning curve).

5 Spike-Train Decoding

1. **Goal:** Estimate stimulus at time $t - \tau_0$ from the n spikes that occurred until time t (where τ_0 is the **prediction delay**).
Note: This assumes that future stimulus values are correlated with previous stimulus values.
Benefit of higher τ_0: more accurate prediction estimate at $t - \tau_0$ since more spikes that occurred after (and were the result of) the stimulus are taken into account.
Downside of higher τ_0: less relevant result that cannot be used to predict immediate/future responses.

2. Stimulus estimate: uses linear kernel K.

$$s_{\text{est}}(t - \tau_0) = \sum_{i=1}^{n} K(t - t_i) - \langle r \rangle \int_{-\infty}^{\infty} d\tau \, K(\tau),$$

where $\langle r \rangle = \langle n \rangle / T$ is the average firing rate over the trial, and that last term is used to ensure that the time average of s_{est} is 0.
K is a function of τ, which counts backward before time t, so the last term is a time average of the quantity in the first term, so that subtracting it gives a zero time-average for s_{est}.
Using $\rho(t) = \sum \delta(t - t_i)$:

$$s_{\text{est}}(t - \tau_0) = \int_{-\infty}^{\infty} d\tau \, (\rho(t - \tau) - \langle r \rangle) \, K(\tau)$$

Minimize squared difference between estimated and actual stimulus for optimal K (similar to Ch 2): Take the functional derivative of both sides of the equation, set it equal to zero, rearrange and write in terms of $Q_{rs}(-\tau)$ and $Q_{ss}(\tau - \tau')$.

$$\int_{-\infty}^{\infty} d\tau' \, Q_{\rho\rho}(\tau - \tau') \, K(\tau') = Q_{rs}(\tau - \tau_0),$$

where $Q_{\rho\rho}$ is the spike-train autocorrelation function,

$$Q_{\rho\rho}(\tau - \tau') = \frac{1}{T} \int_0^T dt \langle (\rho(t - \tau) - \langle r \rangle)(\rho(t - \tau') - \langle r \rangle) \rangle,$$

and Q_{rs} is the correlation of the firing rate and the stimulus,

$$Q_{rs}(\tau - \tau_0) = \frac{1}{T} \left\langle \sum_{i=1}^{n} s(t_i + \tau - \tau_0) \right\rangle = \langle r \rangle C(\tau_0 - \tau),$$

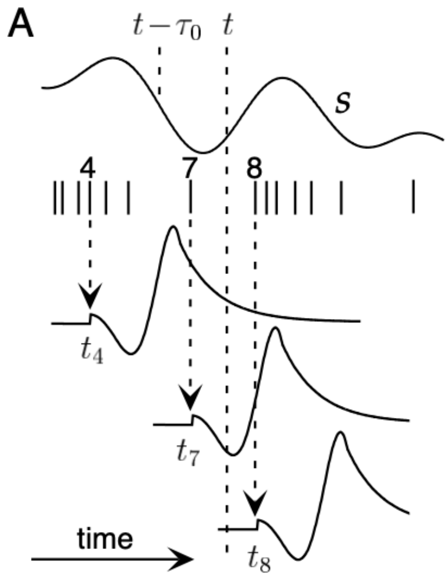
related to the spike-triggered average C.
Assume uncorrelated spike train (generally occurs at low firing rates), so that:

$$Q_{\rho\rho}(\tau) = \langle r \rangle \delta(\tau),$$

so the previous equation for K simplifies to the spike-triggered average $\tau - \tau_0$ before a spike:

$$K(\tau) = \frac{1}{\langle r \rangle} Q_{rs}(\tau - \tau_0) = C(\tau_0 - \tau)$$

Stimulus estimate obtained from summing the kernels K just adds the spike-triggered average of the neuron (which are all functions), but the STA function is translated to match all neuron spikes. This is shown in the diagram below, where the functions at t_4, t_7, and t_8 (theoretically along with the STAs translated to other times), are added up the estimate the stimulus function at the top of the figure.



3. If spike train *is* correlated ($Q_{\rho\rho}$ isn't that delta function above), you can find $K(\tau)$ directly by from the equation we already had (where K could be found implicitly from the integral): Take the Fourier transform of both sides, rearrange, and take the inverse Fourier transform:

$$K(\tau) = \frac{1}{2\pi}\int d\omega\, \hat{K}(\omega)\exp(-i\omega\tau)$$
$$= \frac{\hat{Q}(\omega)\exp(i\omega_0)}{\hat{Q_{\rho\rho}}(\omega)}$$

Note that all solutions above are **acausal** because they include non-zero K values for times $t-\tau$ that occur *after* t as well.

4. Causality constraints:
Causal kernel: $K(t-t_i) = 0$ for $t - t_i \leq 0$, which means that only events that occurred before time t ($t_i < t$) are weighted. Ex: neural response model that only takes into account stimulus values that occurred before time t to determine the response at time t.
Replace the kernel with $\Theta(\tau)K(\tau)$, where

$$\Theta(\tau) = \begin{cases} 1, & \tau > 0, \\ 0, & \tau < 0. \end{cases}$$

6 Summary

1. Discrimination (single neuron)
   (a) Generate histograms of the neuron's firing rate when presented with the different stimuli you want to discriminate between.
   (b) Now you can use the measured firing rate r to determine which distribution it's most likely to be from and predict that the stimulus presented was the one that corresponded with that distribution:
   Set an optimal threshold to maximize the probability of a correct choice.
   Modify the problem to be a two-alternative force-choice task and use ROC analysis.
   Use the Neyman-Pearson lemma and compare r to the likelihood ratio for the optimal result. It can be shown that this is equivalent to using a Bayesian approach where you try to minimize a loss function.

2. Population decoding (multiple neurons)
   (a) Vector method: A group of neurons might act like coordinate axes for different stimulus values, so that a stimulus is projected onto these axes to determine the firing rates of the different neurons, and therefore the rates can be used to backtrack (add up the component vectors) and determine the initial stimulus.
   (b) Bayesian approach: Minimize a loss function.
   If the loss function is the squared difference between the estimate and the true value, the predicted stimulus ends up being the mean stimulus, and if the loss function is the absolute value of this difference, the predicted s ends up being the median of the stimulus distribution. Other loss functions give other s prediction functions.
   (c) MAP inference method: Choose the stimulus value that maximizes the conditional probability of the stimulus, given response $\mathbf{r}$.
   Find the conditional probability of s evoking a certain firing rate $r_a = n_a/T$ using a homogeneous Poisson model, assume that each neuron fires independently, and take a derivative (setting it equal to 0) to find the maximum value.
   (d) ML inference: MAP inference, but when the prior (stimulus probability density) is independent of the stimulus s.
   (e) Cramér-Rao bound: limits the variance of any estimate s_{est}, which limits the accuracy with which a decoding algorithm can estimate the encoded quantity.

$$\sigma_{\text{est}}^2 \geq \frac{(1 + b'_{\text{est}}(s))^2}{I_F(s)},$$

   where $b'_{\text{est}}(s)$ is the derivative of the bias $b_{\text{est}}(s)$, defined so that $s = \langle s_{\text{est}}\rangle - b_{\text{est}}(s)$, and $I_F(s)$ is the **Fisher information** of the firing-rate distribution, which has an equation associated with it.

3. Spike-train decoding (single neurons)
   (a) Find the STA of the neuron, which can be shown to be equal to the optimal kernel K that you take a linear sum of to estimate $s(t)$.
   (b) Estimate the time-varying stimulus function by adding up the STA functions translated to correspond with various spike times prior to t.