Population Coding

March 2021

Population coding

Many quantities — whether stimulus-derived or entirely internal — seem to be represented by the activity of populations of neurons with overlapping responses.

- ▶ Improved precision from combining noisy elements.
- ▶ Redundancy (e.g. against cell death).
- ▶ Computational, processing, or memory advantages?
- ▶ Representations of confidence, uncertainty or multipotent plans?

The full relationship between stimuli and population response has been difficult to study directly.

Instead, thinking about population codes has been dominated by guesses regarding (typically low-dimensional) encoded variables based on the selectivity/invariance pattern.
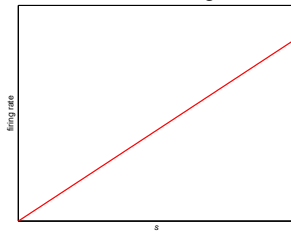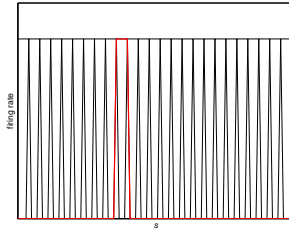
Coding a continuous variable
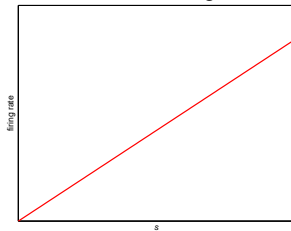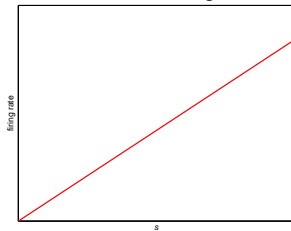

Scalar coding

Coding a continuous variable

Coding a continuous variable

Scalar coding   Labelled Line   Periodic/grid
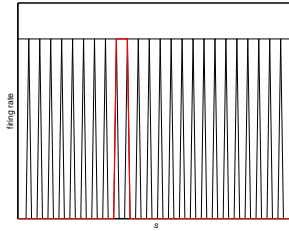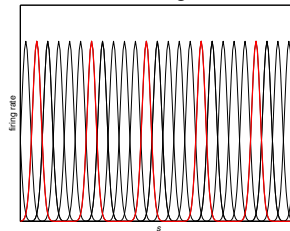
Coding a continuous variable
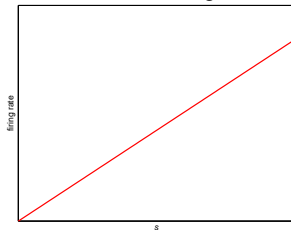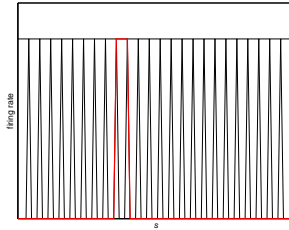
Scalar coding   Labelled Line   Periodic/grid
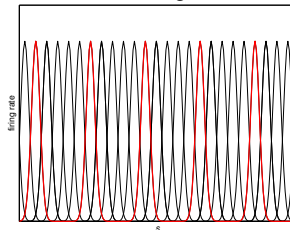
Distributed encoding
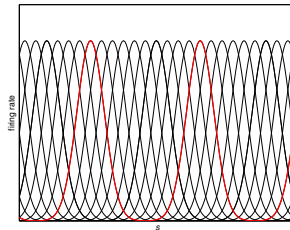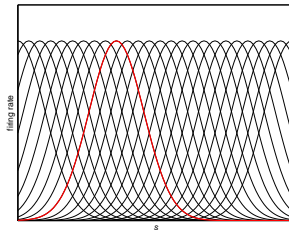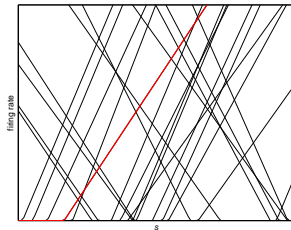
Coding a continuous variable

All of these schemes have been found in biological systems.
Issues:

1. redundancy and robustness (not scalar)
2. efficiency/resolution (not labelled line)
3. local computation (not scalar or scalar distributed)
4. multiple values (not scalar, grid)

Coding in multiple dimensions

Cartesian

Multi-D distributed
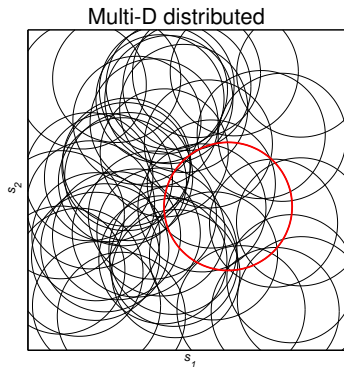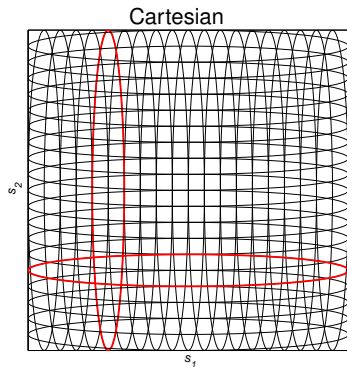
- ▶ efficient
- ▶ problems with multiple values

- ▶ represent multiple values
- ▶ may require more neurons

Theory: decoding and encoding

Theoretically tractable questions:

- ► given assumed encoding functions, how well can we (or downstream areas) decode the encoded stimulus value?
- ► what does the optimal decoder look like?
- ► what encoding schemes would be optimal, in the sense of allowing decoders to estimate stimulus values as well as possible.

Optimality

- Biological systems (including brains) have evolved through natural selection.
  - loosely an optimisation
    - for propagation (of genes)
    - in a dynamic landscape (changing environment; competitors; predators; prey ...)
  - propagation of genes requires successful individuals
  - successful individuals require effective systems and components
- Much theory seems to understand evolved systems solns. as "optimal".
  - cost function?
  - constraints?

In a basic sensory system:

- cost – accurate, sensitive detecton
- constraints – neuron count / energy / wiring ?

Rate coding

In the rate coding context, we imagine that the firing rate of a cell r represents a single (possibly multidimensional) stimulus value s at any one time:

$$r = f(s).$$

Even if s and r are embedded in time-series we assume:

1. that coding is instantaneous (with a fixed lag),
2. that r (and therefore s) is constant over a short time Δ.

The actual number of spikes n produced in Δ is then taken to be distributed around $r\Delta$, often according to a Poisson distribution.

Tuning curves

The function $f(s)$ is known as a tuning curve.
Commonly assumed forms:

- Gaussian

$$r_0 + r_{\max} \exp\left[-\frac{1}{2\sigma^2}(x - x_{\text{pref}})^2\right]$$

- (Thresholded) Ramp

$$r_0 + \Theta(x - x_{\text{thr}})\, r_{\max}\, \rho \cdot (x - x_{\text{thr}})$$

- Cosine

$$r_0 + r_{\max} \cos(\theta - \theta_{\text{pref}})$$

- Wrapped Gaussian

$$r_0 + r_{\max} \sum_n \exp\left[-\frac{1}{2\sigma^2}(\theta - \theta_{\text{pref}} - 2\pi n)^2\right]$$

- von Mises ("circular Gaussian")

$$r_0 + r_{\max} \exp\left[\kappa \cos(\theta - \theta_{\text{pref}})\right]$$

- periodic (grid)

$$f(s) = f_1(\sin(2\pi s/\lambda))$$

Cricket cercal system
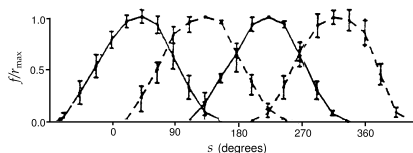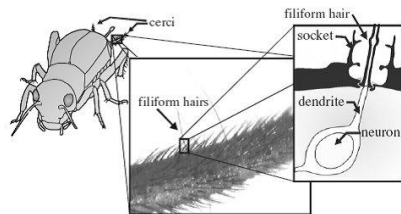
$$r_a(s) = r_a^{\max}[\cos(\theta - \theta_a)]_+ = r_a^{\max}[\mathbf{c}_a^T \mathbf{v}]_+$$

$$\mathbf{c}_1^T \mathbf{c}_2 = 0$$
$$\mathbf{c}_3 = -\mathbf{c}_1$$
$$\mathbf{c}_4 = -\mathbf{c}_2$$

So, writing $\tilde{r}_a = r_a / r_a^{\max}$:

$$\begin{pmatrix} \tilde{r}_1 - \tilde{r}_3 \\ \tilde{r}_2 - \tilde{r}_4 \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \end{pmatrix} \mathbf{v}$$

$$\mathbf{v} = (\mathbf{c}_1 \mathbf{c}_2) \begin{pmatrix} \tilde{r}_1 - \tilde{r}_3 \\ \tilde{r}_2 - \tilde{r}_4 \end{pmatrix} = \tilde{r}_1 \mathbf{c}_1 - \tilde{r}_3 \mathbf{c}_3 + \tilde{r}_2 \mathbf{c}_2 - \tilde{r}_4 \mathbf{c}_4 = \sum_a \tilde{r}_a \mathbf{c}_a$$

This is called population vector decoding.

Motor cortex (simplified)
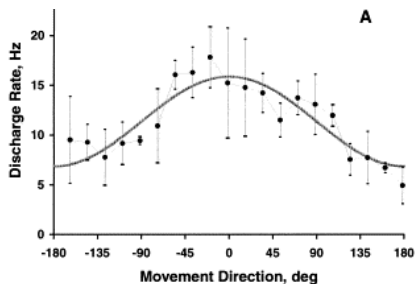
Cosine tuning, randomly distributed preferred directions.
In general, population vector decoding works for

- cosine tuning
- cartesian or dense (*tight*) directions

But:

- is it optimal?
- does it generalise? (Gaussian tuning curves)
- how accurate is it?

Bayesian decoding

Take $n_a \sim \text{Poisson}[f_a(s)\Delta]$, independently for different cells.
Then

$$P(\mathbf{n}|s) = \prod_a \frac{e^{-f_a(s)\Delta}(f_a(s)\Delta)^{n_a}}{n_a!}$$

and

$$\log P(s|\mathbf{n}) = -\sum_a f_a(s)\Delta + n_a log(f_a(s)\Delta) - \log n_a! + \log P(s)$$

Assume $\sum_a f_a(s)$ is independent of s for a homogeneous population, and prior is flat.

$$\frac{d}{ds}\log P(s|\mathbf{n}) = \frac{d}{ds}\sum_a n_a log(f_a(s)\Delta)$$

$$= \sum_a \frac{n_a}{f_a(s)\Delta} f_a'(s)\Delta$$

Bayesian decoding

Now, consider $f_a(s) = e^{-(s-s_a)^2/2\sigma^2}$, so $f_a'(s) = -(s-s_a)/\sigma^2 e^{-(s-s_a)^2/2\sigma^2}$
and set the derivative to 0:

$$\sum_a n_a(s-s_a)/\sigma^2 = 0$$

$$\hat{s}_{\text{MAP}} = \frac{\sum_a n_a s_a}{\sum_a n_a}$$

So the MAP estimate is a population average of preferred directions. Not exactly a population vector.

Measuring the potential quality of a representation

Now consider a (one dimensional) stimulus that takes on continuous values (e.g. angle).

- contrast
- orientation
- motion direction
- movement speed

Suppose a neuron fires n spikes in response to stimulus s according to some distribution

$$P(n|f(s)\Delta)$$

Given an observation of n, how well can we estimate s?

Continuous estimation

Useful to consider a limit given $N \to \infty$ measurements n_i all generated by the same stimulus s^*.

Continuous estimation

Useful to consider a limit given $N \to \infty$ measurements n_i all generated by the same stimulus s^*.

The posterior over s is

$$\log P(s|\{n_i\}) = \sum_i \log P(n_i|s) + \log P(s) - \log Z(\{n_i\})$$

Continuous estimation

Useful to consider a limit given $N \to \infty$ measurements n_i all generated by the same stimulus s^*.

The posterior over s is

$$\log P(s|\{n_i\}) = \sum_i \log P(n_i|s) + \log P(s) - \log Z(\{n_i\})$$

Taking $N \to \infty$ we have

$$\frac{1}{N} \log P(s|\{n_i\}) \to \langle \log P(n|s) \rangle_{n|s^*} + 0 - \log Z(s^*)$$

Continuous estimation

Useful to consider a limit given $N \to \infty$ measurements n_i all generated by the same stimulus s^*.

The posterior over s is

$$\log P(s|\{n_i\}) = \sum_i \log P(n_i|s) + \log P(s) - \log Z(\{n_i\})$$

Taking $N \to \infty$ we have

$$\frac{1}{N} \log P(s|\{n_i\}) \to \langle \log P(n|s) \rangle_{n|s*} + 0 - \log Z(s^*)$$

and so

$$P(s|\{n_i\}) \to e^{N\langle \log P(n|s) \rangle_{n|s*}} / Z$$

$$= e^{-N\mathbf{KL}\left[P(n|s^*)\middle\|P(n|s)\right]} / Z$$

(Note: Z is being redefined as we go, but never depends on s)

Continuous estimation

Now, Taylor expand the KL divergence in s around s^*:

$$\mathbf{KL}[P(n|s^*)\|P(n|s)]$$

$$= -\langle \log P(n|s)\rangle_{n|s^*} + \langle \log P(n|s^*)\rangle_{n|s^*}$$

Continuous estimation

Now, Taylor expand the KL divergence in s around s^*:

$$\mathbf{KL}[P(n|s^*)\|P(n|s)]$$

$$= -\langle \log P(n|s) \rangle_{n|s^*} + \langle \log P(n|s^*) \rangle_{n|s^*}$$

$$= -\langle \log P(n|s^*) \rangle_{n|s^*} - (s - s^*)\left\langle \frac{d \log P(n|s)}{ds}\bigg|_{s^*} \right\rangle_{s^*} - \frac{1}{2}(s - s^*)^2 \left\langle \frac{d^2 \log P(n|s)}{ds^2}\bigg|_{s^*} \right\rangle_{s^*}$$

$$+ \langle \log P(n|s^*) \rangle_{n|s^*}$$

Continuous estimation

Now, Taylor expand the KL divergence in s around s^*:

$$\mathbf{KL}[P(n|s^*)\|P(n|s)]$$

$$= -\langle \log P(n|s)\rangle_{n|s^*} + \langle \log P(n|s^*)\rangle_{n|s^*}$$

$$= -\langle \log P(n|s^*)\rangle_{n|s^*} - (s-s^*)\left\langle \frac{d \log P(n|s)}{ds}\Big|_{s^*}\right\rangle_{s^*} - \frac{1}{2}(s-s^*)^2\left\langle \frac{d^2 \log P(n|s)}{ds^2}\Big|_{s^*}\right\rangle_{s^*}$$

$$\quad + \langle \log P(n|s^*)\rangle_{n|s^*}$$

$$= -\frac{1}{2}(s-s^*)^2\left\langle \frac{d^2 \log P(n|s)}{ds^2}\Big|_{s^*}\right\rangle_{s^*} + \dots$$

Continuous estimation

Now, Taylor expand the KL divergence in s around s^*:

$$\mathbf{KL}[P(n|s^*)\|P(n|s)]$$

$$= -\langle \log P(n|s) \rangle_{n|s^*} + \langle \log P(n|s^*) \rangle_{n|s^*}$$

$$= -\langle \log P(n|s^*) \rangle_{n|s^*} - (s - s^*)\left\langle \frac{d \log P(n|s)}{ds}\bigg|_{s^*} \right\rangle_{s^*} - \frac{1}{2}(s - s^*)^2 \left\langle \frac{d^2 \log P(n|s)}{ds^2}\bigg|_{s^*} \right\rangle_{s^*}$$

$$+ \langle \log P(n|s^*) \rangle_{n|s^*}$$

$$= -\frac{1}{2}(s - s^*)^2 \left\langle \frac{d^2 \log P(n|s)}{ds^2}\bigg|_{s^*} \right\rangle_{s^*} + \dots$$

$$= \frac{1}{2}(s - s^*)^2 J(s^*) + \dots$$

Continuous estimation

Now, Taylor expand the KL divergence in s around s^*:

$$\textbf{KL}[P(n|s^*)\|P(n|s)]$$

$$= -\langle \log P(n|s)\rangle_{n|s^*} + \langle \log P(n|s^*)\rangle_{n|s^*}$$

$$= -\langle \log P(n|s^*)\rangle_{n|s^*} - (s - s^*)\left\langle \frac{d \log P(n|s)}{ds}\Big|_{s^*}\right\rangle_{s^*} - \frac{1}{2}(s - s^*)^2\left\langle \frac{d^2 \log P(n|s)}{ds^2}\Big|_{s^*}\right\rangle_{s^*}$$

$$\quad + \langle \log P(n|s^*)\rangle_{n|s^*}$$

$$= -\frac{1}{2}(s - s^*)^2\left\langle \frac{d^2 \log P(n|s)}{ds^2}\Big|_{s^*}\right\rangle_{s^*} + \ldots$$

$$= \frac{1}{2}(s - s^*)^2 J(s^*) + \ldots$$

So in asymptopia, the posterior $\rightarrow \mathcal{N}(s^*, 1/J(s^*))$.
$J(s^*)$ is called the *Fisher Information*.

Continuous estimation

Now, Taylor expand the KL divergence in s around s^*:

$$\mathbf{KL}[P(n|s^*)\|P(n|s)]$$

$$= -\langle \log P(n|s) \rangle_{n|s^*} + \langle \log P(n|s^*) \rangle_{n|s^*}$$

$$= -\langle \log P(n|s^*) \rangle_{n|s^*} - (s - s^*) \left\langle \frac{d \log P(n|s)}{ds} \Big|_{s^*} \right\rangle_{s^*} - \frac{1}{2}(s - s^*)^2 \left\langle \frac{d^2 \log P(n|s)}{ds^2} \Big|_{s^*} \right\rangle_{s^*}$$

$$\quad + \langle \log P(n|s^*) \rangle_{n|s^*}$$

$$= -\frac{1}{2}(s - s^*)^2 \left\langle \frac{d^2 \log P(n|s)}{ds^2} \Big|_{s^*} \right\rangle_{s^*} + \dots$$

$$= \frac{1}{2}(s - s^*)^2 J(s^*) + \dots$$

So in asymptopia, the posterior $\rightarrow \mathcal{N}(s^*, 1/J(s^*))$.
$J(s^*)$ is called the *Fisher Information*.

$$J(s^*) = -\left\langle \frac{d^2 \log P(n|s)}{ds^2} \Big|_{s^*} \right\rangle_{s^*} = \left\langle \left(\frac{d \log P(n|s)}{ds} \Big|_{s^*} \right)^2 \right\rangle_{s^*}$$

(You will show that these are identical in the homework.)

Cramér-Rao bound

The Fisher Information is important even outside the large data limit due to a deeper result that is due to Cramér and Rao.

This states that for any N, any *unbiased* estimator $\hat{s}(\{n_i\})$ of s will have the property that

$$\left\langle (\hat{s}(\{n_i\}) - s^*)^2 \right\rangle_{n_i|s^*} \geq \frac{1}{J(s^*)}.$$

Thus, Fisher Information gives a lower bound on the variance of any unbiased estimator. This is called the Cramér-Rao bound.

[For estimators with *bias* $b(s^*) = \langle \hat{s}(\{n_i\}) - s^* \rangle$ the bound is:
$\left\langle (\hat{s}(\{n_i\}) - s^*)^2 \right\rangle_{n_i|s^*} \geq \frac{(1+b'(s^*))^2}{J(s^*)} + b^2(s^*)$]

The Fisher Information will be our primary tool to quantify the performance of a population code.

Fisher Info and tuning curves

$n = r\Delta + \text{noise}; \quad r = f(s) \Rightarrow$

$$J(s^*) = \left\langle \left(\frac{d}{ds}\Big|_{s^*} \log P(n|s) \right)^2 \right\rangle_{s^*}$$

Fisher Info and tuning curves

$n = r\Delta + \text{noise}; \quad r = f(s) \Rightarrow$

$$J(s^*) = \left\langle \left(\frac{d}{ds}\Big|_{s^*} \log P(n|s) \right)^2 \right\rangle_{s^*}$$

$$= \left\langle \left(\frac{d}{dr\Delta}\Big|_{f(s^*)} \log P(n|r\Delta)\Delta f'(s^*) \right)^2 \right\rangle_{s^*}$$

Fisher Info and tuning curves

$n = r\Delta + \text{noise}; \quad r = f(s) \Rightarrow$

$$J(s^*) = \left\langle \left(\frac{d}{ds}\Big|_{s^*} \log P(n|s) \right)^2 \right\rangle_{s^*}$$

$$= \left\langle \left(\frac{d}{dr\Delta}\Big|_{f(s^*)} \log P(n|r\Delta) \Delta f'(s^*) \right)^2 \right\rangle_{s^*}$$

$$= J_{\text{noise}}(r\Delta) \Delta^2 f'(s^*)^2$$

Fisher Info and tuning curves

$n = r\Delta + \text{noise}; \quad r = f(s) \Rightarrow$

$$J(s^*) = \left\langle \left(\frac{d}{ds}\Big|_{s^*} \log P(n|s) \right)^2 \right\rangle_{s^*}$$

$$= \left\langle \left(\frac{d}{dr\Delta}\Big|_{f(s^*)} \log P(n|r\Delta)\Delta f'(s^*) \right)^2 \right\rangle_{s^*}$$

$$= J_{\text{noise}}(r\Delta)\Delta^2 f'(s^*)^2$$

Fisher info for Poisson neurons

For Poisson neurons

$$P(n|r\Delta) = \frac{e^{-r\Delta}}{(r\Delta)^n}n!$$

so

$$J_{\text{noise}}[r\Delta] = \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} \log P(n|r\Delta) \right)^2 \right\rangle_{s^*}$$

Fisher info for Poisson neurons

For Poisson neurons

$$P(n|r\Delta) = \frac{e^{-r\Delta}}{(r\Delta)^n} n!$$

so

$$J_{\text{noise}}[r\Delta] = \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} \log P(n|r\Delta) \right)^2 \right\rangle_{s^*}$$

$$= \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} - r\Delta + n\log r\Delta - \log n! \right)^2 \right\rangle_{s^*}$$

Fisher info for Poisson neurons

For Poisson neurons

$$P(n|r\Delta) = \frac{e^{-r\Delta}}{(r\Delta)^n} n!$$

so

$$J_{\text{noise}}[r\Delta] = \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} \log P(n|r\Delta) \right)^2 \right\rangle_{s^*}$$

$$= \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} - r\Delta + n \log r\Delta - \log n! \right)^2 \right\rangle_{s^*}$$

$$= \left\langle \left(-1 + n/r^*\Delta \right)^2 \right\rangle_{s^*}$$

Fisher info for Poisson neurons

For Poisson neurons

$$P(n|r\Delta) = \frac{e^{-r\Delta}}{(r\Delta)^n} n!$$

so

$$
\begin{aligned}
J_{\text{noise}}[r\Delta] &= \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} \log P(n|r\Delta) \right)^2 \right\rangle_{s^*} \\
&= \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} - r\Delta + n\log r\Delta - \log n! \right)^2 \right\rangle_{s^*} \\
&= \left\langle \left(-1 + n/r^*\Delta \right)^2 \right\rangle_{s^*} \\
&= \left\langle \frac{(n - r^*\Delta)^2}{(r^*\Delta)^2} \right\rangle_{s^*}
\end{aligned}
$$

Fisher info for Poisson neurons

For Poisson neurons

$$P(n|r\Delta) = \frac{e^{-r\Delta}}{(r\Delta)^n} n!$$

so

$$
\begin{aligned}
J_{\text{noise}}[r\Delta] &= \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} \log P(n|r\Delta) \right)^2 \right\rangle_{s^*} \\
&= \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} - r\Delta + n\log r\Delta - \log n! \right)^2 \right\rangle_{s^*} \\
&= \left\langle \left(-1 + n/r^*\Delta \right)^2 \right\rangle_{s^*} \\
&= \left\langle \frac{(n - r^*\Delta)^2}{(r^*\Delta)^2} \right\rangle_{s^*} \\
&= \frac{r^*\Delta}{(r^*\Delta)^2}
\end{aligned}
$$

Fisher info for Poisson neurons

For Poisson neurons

$$P(n|r\Delta) = \frac{e^{-r\Delta}}{(r\Delta)^n} n!$$

so

$$
\begin{aligned}
J_{\text{noise}}[r\Delta] &= \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} \log P(n|r\Delta) \right)^2 \right\rangle_{s^*} \\
&= \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} - r\Delta + n\log r\Delta - \log n! \right)^2 \right\rangle_{s^*} \\
&= \left\langle \left(-1 + n/r^*\Delta \right)^2 \right\rangle_{s^*} \\
&= \left\langle \frac{(n - r^*\Delta)^2}{(r^*\Delta)^2} \right\rangle_{s^*} \\
&= \frac{r^*\Delta}{(r^*\Delta)^2} = \frac{1}{r^*\Delta}
\end{aligned}
$$

Fisher info for Poisson neurons

For Poisson neurons

$$P(n|r\Delta) = \frac{e^{-r\Delta}}{(r\Delta)^n} n!$$

so

$$
\begin{aligned}
J_{\text{noise}}[r\Delta] &= \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} \log P(n|r\Delta) \right)^2 \right\rangle_{s^*} \\
&= \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} - r\Delta + n\log r\Delta - \log n! \right)^2 \right\rangle_{s^*} \\
&= \left\langle \left(-1 + n/r^*\Delta \right)^2 \right\rangle_{s^*} \\
&= \left\langle \frac{(n - r^*\Delta)^2}{(r^*\Delta)^2} \right\rangle_{s^*} \\
&= \frac{r^*\Delta}{(r^*\Delta)^2} = \frac{1}{r^*\Delta} \qquad \text{[not surprising! } \widehat{r^*\Delta} = n \text{ and } \mathbb{V}[n] = r^*\Delta\text{]}
\end{aligned}
$$

Fisher info for Poisson neurons

For Poisson neurons

$$P(n|r\Delta) = \frac{e^{-r\Delta}}{(r\Delta)^n} n!$$

so

$$
\begin{aligned}
J_{\text{noise}}[r\Delta] &= \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} \log P(n|r\Delta) \right)^2 \right\rangle_{s^*} \\
&= \left\langle \left(\frac{d}{dr\Delta}\Big|_{r^*\Delta} - r\Delta + n\log r\Delta - \log n! \right)^2 \right\rangle_{s^*} \\
&= \left\langle \left(-1 + n/r^*\Delta \right)^2 \right\rangle_{s^*} \\
&= \left\langle \frac{(n - r^*\Delta)^2}{(r^*\Delta)^2} \right\rangle_{s^*} \\
&= \frac{r^*\Delta}{(r^*\Delta)^2} = \frac{1}{r^*\Delta} \qquad [\text{not surprising! } \widehat{r^*\Delta} = n \text{ and } \mathbb{V}[n] = r^*\Delta]
\end{aligned}
$$

and, referred back to the stimulus value:

$$J[s^*] = f'(s^*)^2 \Delta / f(s^*)$$

Population Fisher Info – independent noise

Fisher Informations for independent random variates add:

$$
\begin{aligned}
J_{\mathbf{n}}(s) &= \left\langle -\frac{d^2}{ds^2} \log P(\mathbf{n}|s) \right\rangle \\
&= \left\langle -\frac{d^2}{ds^2} \sum_a \log P(n_a|s) \right\rangle \\
&= \sum_a \left\langle -\frac{d^2}{ds^2} \log P(n_a|s) \right\rangle = \sum_a J_{n_a}(s). \\
&= \Delta \sum_a \frac{f_a'(s)^2}{f_a(s)} \qquad \text{[for Poisson cells]}
\end{aligned}
$$

Population Fisher Info – correlated (Gaussian) noise

$$\mathbf{r} \sim \mathcal{N}\left(\boldsymbol{\mu}(s), \Sigma(s)\right)$$

$$
\begin{aligned}
J(s) &= -\left\langle \frac{\partial^2}{\partial s^2} \log p(\mathbf{r}|s) \right\rangle \\
&= \frac{1}{2}\left\langle -\frac{\partial^2}{\partial s^2}\log\left|\Sigma^{-1}\right| + \frac{\partial^2}{\partial s^2}\mathsf{Tr}\left[\Sigma^{-1}(\mathbf{r}-\boldsymbol{\mu})(\mathbf{r}-\boldsymbol{\mu})^{\mathsf{T}}\right] \right\rangle \\
&= \frac{1}{2}\left\langle -\frac{\partial}{\partial s}\mathsf{Tr}\left[\Sigma(\Sigma^{-1})'\right] + \frac{\partial}{\partial s}\mathsf{Tr}\left[(\Sigma^{-1})'(\mathbf{r}-\boldsymbol{\mu})(\mathbf{r}-\boldsymbol{\mu})^{\mathsf{T}}\right] - 2\frac{\partial}{\partial s}\mathsf{Tr}\left[\Sigma^{-1}\boldsymbol{\mu}'(\mathbf{r}-\boldsymbol{\mu})\right] \right\rangle \\
&= \frac{1}{2}\left\langle -\mathsf{Tr}\left[\Sigma'(\Sigma^{-1})'\right] - \mathsf{Tr}\left[\Sigma(\Sigma^{-1})''\right] \right. \\
&\qquad + \mathsf{Tr}\left[(\Sigma^{-1})''(\mathbf{r}-\boldsymbol{\mu})(\mathbf{r}-\boldsymbol{\mu})^{\mathsf{T}}\right] - 2\mathsf{Tr}\left[(\Sigma^{-1})'\boldsymbol{\mu}'(\mathbf{r}-\boldsymbol{\mu})^{\mathsf{T}}\right] \\
&\qquad \left. -2\mathsf{Tr}\left[(\Sigma^{-1})'\boldsymbol{\mu}'(\mathbf{r}-\boldsymbol{\mu})\right] - 2\mathsf{Tr}\left[\Sigma^{-1}\boldsymbol{\mu}''(\mathbf{r}-\boldsymbol{\mu})\right] - 2\mathsf{Tr}\left[\Sigma^{-1}\boldsymbol{\mu}'\boldsymbol{\mu}'^{\mathsf{T}}\right] \right\rangle \\
&= \frac{1}{2}\left(-\mathsf{Tr}\left[\Sigma'(\Sigma^{-1})'\right] - \mathsf{Tr}\left[\Sigma(\Sigma^{-1})''\right] + \mathsf{Tr}\left[(\Sigma^{-1})''\Sigma\right] - 2\mathsf{Tr}\left[\Sigma^{-1}\boldsymbol{\mu}'\boldsymbol{\mu}'^{\mathsf{T}}\right] \right) \\
&= \frac{1}{2}\mathsf{Tr}\left[\Sigma'\Sigma^{-1}\Sigma'\Sigma^{-1}\right] + \boldsymbol{\mu}'^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu}'
\end{aligned}
$$

Gaussian FI

$$J(s) = \frac{1}{2}\text{Tr}\left[\mathbf{\Sigma}'\mathbf{\Sigma}^{-1}\mathbf{\Sigma}'\mathbf{\Sigma}^{-1}\right] + \boldsymbol{\mu}'^{\mathsf{T}}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}'$$

▶ Most focus on second term.

▶ "Linear" Fisher information – gives performance of a locally-optimal linear decoder.

▶ Eigendecomposition $\mathbf{\Sigma} = \sum_i \lambda_i \mathbf{v}_i \mathbf{v}_i^{\mathsf{T}}$

$$\Rightarrow J_1(s) = \sum_i \lambda_i^{-1}(\mathbf{v}_i^{\mathsf{T}}\boldsymbol{\mu}')^2.$$

▶ Large eigenvalue along $\boldsymbol{\mu}'$ gives low FI – sometimes called 'differential correlation' or 'information-limiting' correlation.

Optimal tuning properties

Consider a population of cells that codes the value of a D dimensional stimulus, $\mathbf{s}$. Let the ath cell emit r spikes in an interval τ with probability distribution that is conditionally independent of the other cells (given $\mathbf{s}$) and has the form

$$P_a(r \mid \mathbf{s}, \tau) = S(r, f^a(\mathbf{s}), \tau).$$

Also let the tuning curve of the ath cell, $f^a(\mathbf{s})$, be circularly symmetric:

$$f^a(\mathbf{s}) = F \cdot \phi\left((\xi^a)^2\right); \qquad (\xi^a)^2 = \sum_i^D (\xi_i^a)^2; \qquad \xi_i^a = \frac{s_i - c_i^a}{\sigma},$$

where F is a maximal rate and the function ϕ is monotically decreasing. The parameters $\mathbf{c}^a$ and σ give the centre of the ath tuning curve and the (common) width.

Optimal tuning properties

Now, the (ij)th term in the FI matrix for the ath cell is (by definition)

$$J_{ij}^a(\mathbf{s}) = \mathsf{E}\left[\frac{\partial}{\partial s_i}\log \mathsf{P}^a(r \mid \mathbf{s}, \tau)\frac{\partial}{\partial s_j}\log \mathsf{P}^a(r \mid \mathbf{s}, \tau)\right]$$

Applying the chain rule repeatedly, we find that

$$\frac{\partial}{\partial s_i}\log \mathsf{P}^a(r \mid \mathbf{s}, \tau) = \frac{1}{S(r, f^a(\mathbf{s}), \tau)}\frac{\partial}{\partial s_i}S(r, f^a(\mathbf{s}), \tau)$$

$$= \frac{S^{(2)}(r, f^a(\mathbf{s}), \tau)}{S(r, f^a(\mathbf{s}), \tau)}\frac{\partial}{\partial s_i}f^a(\mathbf{s})$$

(where $S^{(2)}$ indicates differentiation with respect to the second argument)

$$= \frac{S^{(2)}(r, f^a(\mathbf{s}), \tau)}{S(r, f^a(\mathbf{s}), \tau)}F\phi'\left((\xi^a)^2\right)\frac{\partial}{\partial s_i}\sum_i^D(\xi_i^a)^2$$

$$= \frac{S^{(2)}(r, f^a(\mathbf{s}), \tau)}{S(r, f^a(\mathbf{s}), \tau)}F\phi'\left((\xi^a)^2\right)\frac{2(s_i - c_i^a)}{(\sigma_i^a)^2}$$

Optimal tuning properties

So,

$$J_{ij}^a(\mathbf{s}) = \mathsf{E}\left[\left(\frac{S^{(2)}(r, f^a(\mathbf{s}), \tau)}{S(r, f^a(\mathbf{s}), \tau)}\right)^2\right] 4F^2 \left(\phi'\left((\xi^a)^2\right)\right)^2 \frac{(s_i - c_i^a)(s_j - c_j^a)}{\sigma^4}$$

$$= A_\phi\left((\xi^a)^2, F, \tau\right) \frac{(s_i - c_i^a)(s_j - c_j^a)}{\sigma^4}$$

where the function A_ϕ does not depend explicitly on σ.

Optimal tuning properties

We assumed neurons were independent $\Rightarrow$ Fisher information adds. Approximate by integral over the tuning curve centres, assuming uniform density η of neurons.

$$
\begin{aligned}
J_{ij}(\mathbf{s}) &= \sum_a J_{ij}^a(\mathbf{s}) \\
&\approx \int_{-\infty}^{+\infty} dc_1^a \cdots \int_{-\infty}^{+\infty} dc_D^a \, \eta J_{ij}^a(\mathbf{s}) \\
&= \int_{-\infty}^{+\infty} dc_1^a \cdots \int_{-\infty}^{+\infty} dc_D^a \, \eta A_\phi \left((\xi^a)^2, F, \tau \right) \frac{(s_i - c_i^a)(s_j - c_j^a)}{\sigma^4}
\end{aligned}
$$

Change variables: $c_i^a \to \xi_i^a$

$$
\begin{aligned}
&= \int_{-\infty}^{+\infty} \sigma d\xi_1^a \cdots \int_{-\infty}^{+\infty} \sigma d\xi_D^a \, \eta A_\phi \left((\xi^a)^2, F, \tau \right) \frac{\xi_i^a \xi_j^a}{\sigma^2} \\
&= \frac{\sigma^D}{\sigma^2} \eta \int_{-\infty}^{+\infty} d\xi_1^a \cdots \int_{-\infty}^{+\infty} d\xi_D^a \, A_\phi \left((\xi^a)^2, F, \tau \right) \xi_i^a \xi_j^a
\end{aligned}
$$

Now, if $i \neq j$, integral is odd in both ξ_i^a and ξ_j^a, and thus vanishes. If $i = j$, then the integral has some value $D \cdot K_\phi(F, \tau, D)$, independent of σ. Thus,

$$
J_{ii} = \sigma^{D-2} \eta D K_\phi(F, \tau, D)
$$

and the total Fisher information is proportional to σ^{D-2}.

Optimal tuning properties

Thus optimal tuning width depends on the stimulus dimension through the interplay of two effects:

$$\text{slope:} \qquad f'(s) \propto \sigma^{-1} \Rightarrow J_a(s) \propto \sigma^{-2} \qquad \textit{per cell}$$

$$\text{number of cells:} \qquad N(s) \propto \sigma^{D} \Rightarrow J(s) \propto \sigma^{D-2} \qquad \textit{population}$$

- $D = 1$
  - $\Rightarrow \sigma \rightarrow 0$ (although a lower limit is encountered when the tuning width falls below the inter-cell spacing)
- $D = 2$
  - $\Rightarrow J$ independent of σ.
- $D > 2$
  - $\Rightarrow \sigma \rightarrow \infty$ (actual limit set by valid stimuli).

Optimal tuning properties

Thus optimal tuning width depends on the stimulus dimension through the interplay of two effects:

$$\text{slope:} \qquad f'(s) \propto \sigma^{-1} \Rightarrow J_a(s) \propto \sigma^{-2} \qquad \textit{per cell}$$

$$\text{number of cells:} \qquad N(s) \propto \sigma^D \Rightarrow J(s) \propto \sigma^{D-2} \qquad \textit{population}$$

- $D = 1$
  - $\Rightarrow \sigma \to 0$ (although a lower limit is encountered when the tuning width falls below the inter-cell spacing)
- $D = 2$
  - $\Rightarrow J$ independent of σ.
- $D > 2$
  - $\Rightarrow \sigma \to \infty$ (actual limit set by valid stimuli).

- If circular symmetry is relaxed to allow different scales in each dimension for different cells then solution is a Cartesian code (narrow in one dimension, wide in others).
- Single-bump constraint is essential to analysis. Fisher information cannot address ambiguity between bumps.
- Single coded value – analysing multiple values or distributions is more complex.