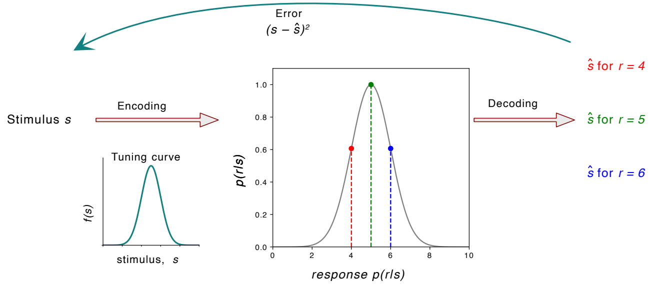
Neural Coding

5. Tuning curves (end). Fisher Information, Minimal discriminator error

Anna Levina

May 15, 2025

Reminder of our setting



Ideal observer minimizes MSE:

$$\hat{s}_{MSE}(r) = \underset{\hat{s}}{\text{argmin}}\, MSE(s) = \underset{\hat{s}}{\text{argmin}} \int p(s \mid r)(\hat{s} - s)^2 ds.$$

The minimum will be found at:

$$\hat{s}_{MSE}(r) = \int s \cdot p(s \mid r) ds = E_s[s \mid r].$$

Variance of estimator and bias

We can decompose the MSE into the variance of the estimator $\sigma^2(s) = E\left[(E[\hat{s} \mid s] - \hat{s})^2\right]$ and the bias $b(s) := E[\hat{s} \mid s] - s$. So we can write

$$MSE(s) = \sigma^2(s) + b^2(s)$$

The estimator is unbiased if

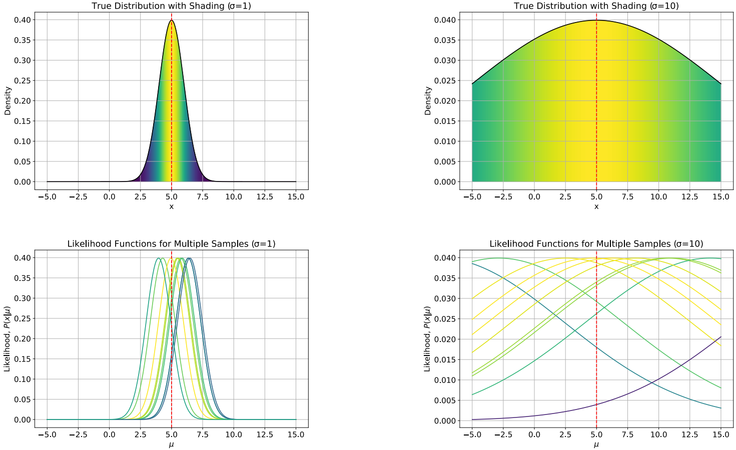
$$b(s) = E[\hat{s} \mid s] - s = 0$$

Today:

- How can we get a grip on the variance of the estimator?
- What can we do if we just need to find if it is a stimulus-1 or stimulus-2?

Fisher Information 1. Probability and likelihood

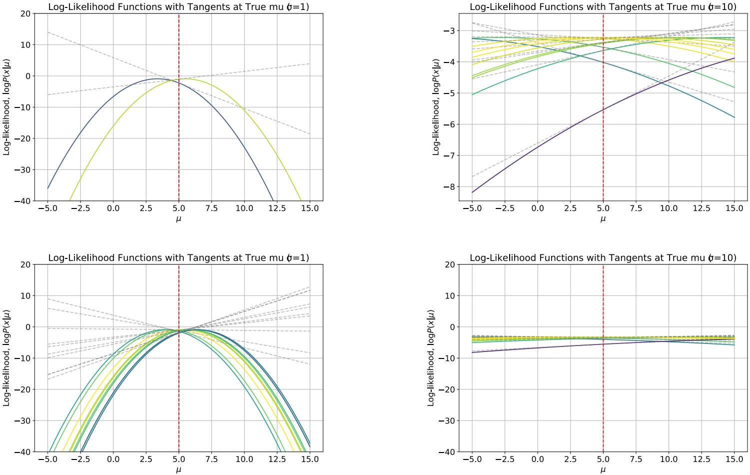
Fisher information quantifies the amount of information that an observable random variable X carries about an unknown parameter θ upon which the probability of X depends.



Fisher Information 2. Score function

Score function $S_X(\theta)$:

$$S_X(\theta) := \frac{\partial}{\partial\theta}\log p(X|\theta) = \frac{\frac{\partial}{\partial\theta}p(X|\theta)}{p(X|\theta)}$$



Fisher Information 3. Finally definition

Fisher information is a variance of the score:

$$J(\theta) = J(p(X|\theta)) = \mathrm{E}_X \left[S_X^2(\theta) \middle| \theta \right] = \mathrm{E}_X \left[\left(\frac{\partial}{\partial \theta} \log p(X|\theta) \right)^2 |\theta \right]$$

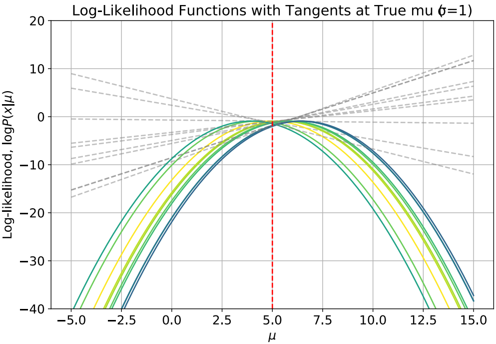
$$= \int \left(\frac{\partial}{\partial \theta} \log p(X|\theta) \right)^2 p(X|\theta) dX = -\mathrm{E}_X \left[\left(\frac{\partial^2}{\partial \theta^2} \log p(X|\theta) \right) |\theta \right]$$

The last equality is satisfied if the likelihood is doubly differentiable.

$$\frac{\partial^2}{\partial \theta^2} \log p(X|\theta) = \frac{\frac{\partial^2}{\partial \theta^2} p(X|\theta)}{p(X|\theta)} - \left(\frac{\frac{\partial}{\partial \theta} p(X|\theta)}{p(X|\theta)} \right)^2$$

$$\mathrm{E}_X \left[\frac{\frac{\partial^2}{\partial \theta^2} p(X|\theta)}{p(X|\theta)} \right] = \int_{-\infty}^{+\infty} \frac{\frac{\partial^2}{\partial \theta^2} p(X|\theta)}{p(X|\theta)} p(X|\theta) dX = \frac{\partial^2}{\partial \theta^2} \int_{-\infty}^{+\infty} p(X|\theta) dX =$$

Fisher Information 4. Why is it a variance and not mean?



$$\mathrm{E}(S_X(\theta)) = \int_{-\infty}^{+\infty} p(X|\theta)\frac{\partial}{\partial\theta}\log p(X|\theta)dX = \int_{-\infty}^{+\infty}\frac{\partial p(r|\theta)}{\partial\theta}dX = 0.$$

Hence, for the variance, we can write:

$$\mathrm{Var}(S_X(\theta)) = \mathrm{E}((S_X(\theta))^2) + \mathrm{E}((S_X(\theta)))^2 = \mathrm{E}((S_X(\theta))^2)$$

Cramér-Rao bound

For the unbiased estimator $\hat{s}(r)$ Cramér- Rao bound is:

$$\mathrm{Var}[\hat{s}(r)|s] = \mathrm{E}_r[(s - \hat{s}(r))^2|s] \geq \frac{1}{J(s)}$$

This allows us to define the efficiency of the unbiased estimator:

$$e(\hat{s}) = \frac{J(s)^{-1}}{\mathrm{Var}(\hat{s})}$$

General form: If $E[\hat{s}] = g(s)$, then $\mathrm{Var}[\hat{s}] \geq \frac{(g'(s))^2}{J_s}$. Thus, if $\hat{s}$ has bias $b(s) \neq 0$, then we can write $\hat{s} = g(s) = s + b(s)$, $g'(s) = 1 + b'(s)$. Using the general form of the bound, we get:

$$\mathrm{var}(\hat{s}) \geq \frac{[1 + b'(s)]^2}{J(s)} \Rightarrow \mathrm{E}\left((\hat{s} - s)^2\right) \geq \frac{[1 + b'(s)]^2}{J(s)} + b(s)^2$$

Asymptotically for unbiased estimator $(s - \hat{s}) \sim \mathcal{N}(0, \frac{1}{J(s)})$

Poisson noise Fischer Information, direct computation

$$
\begin{aligned}
J_s &= -\mathrm{E}_r\left[\frac{\partial^2}{\partial s^2}\log\left(\frac{f(s)^r}{r!}e^{-f(s)}\right)\right] \\
&= -\mathrm{E}_r\left[\frac{\partial^2}{\partial s^2}\left(r\log f(s)-\log(r!)-f(s)\right)\right] \\
&= -\mathrm{E}_r\left[\frac{\partial}{\partial s}\left(\frac{rf'(s)}{f(s)}-f'(s)\right)\right] = -\mathrm{E}_r\left[r\frac{f\cdot f''-(f')^2}{f^2}-f''\right] \\
&= -\frac{f\cdot f''-(f')^2-f\cdot f''}{f} = \frac{(f')^2}{f}
\end{aligned}
$$

The shape of the tuning function can influence the properties of the estimator significantly. An example of such impact on the reconstruction of direction by a linear combination of neurons Seung and Sompolinsky, "Simple models for reading neuronal population codes." The problem disappears when using a population vector.

A trick to compute Fisher information for tuning curves

In the reference materials (like Wikipedia), you will find Fisher information for the simple distribution depending on a parameter:

$$I(\theta) = -\int \left[\frac{\partial^2}{\partial \theta^2} \log p(x|\theta) \right] p(x|\theta) dx = E\left(\left\{ \frac{\partial}{\partial \theta} \log p(x|\theta) \right\}^2 \right)$$

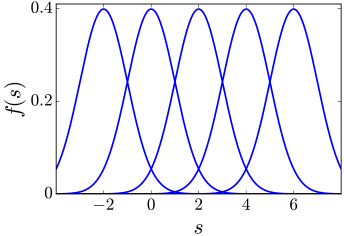
But in our case, $\theta = f(s)$, so the derivatives are more complicated. But the following theorem can save us:

Theorem

Let X be a random variable with density function $p(x|\theta)$ and $I_0(\theta)$ be the Fisher information of X. Suppose now the parameter θ is replaced by a new parameter μ, where $\theta = \phi(\mu)$, and ϕ is a differentiable function. Let $I_1(\mu)$ denote the Fisher information of X when the parameter is μ. Then

$$I_1(\mu) = \left[\phi'(\mu)\right]^2 I_0[\phi(\mu)]$$

Sharpen or broaden the tuning curves? *



Let the stimulus be D-dimensional, and the preferred stimulus be uniformly distributed across neurons with density ρ_s. Tuning curves are Gaussian.

Tuning function $f(x) = F\phi\left(\frac{|x-c|^2}{\sigma^2}\right)$

The total tuning curve is a multiplication of tuning in each direction, std $= \sigma$. Then Fisher information

$$J = \frac{(2\pi)^{D/2}\rho_s\sigma^{D-2}r_{max}T}{D}$$

So if $D = 1$, sharper tuning is good (increasing Fisher information), $D > 2$ sharpening decreases Fisher information Dayan and Abbott ("Theoretical neuroscience: computational and mathematical modeling of neural systems") and Zhang and Sejnowski ("Neuronal Tuning: To Sharpen or Broaden?")

Stimuli discrimination setting

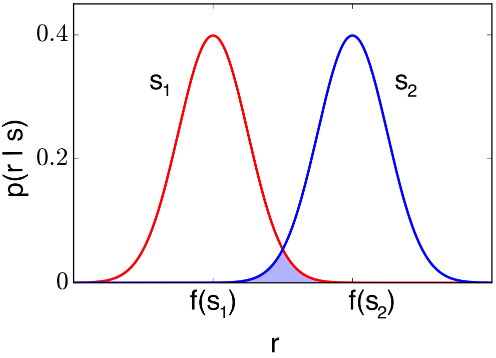
- Let there be two possible stimuli s_1 and s_2
- We want to find out from the response r which stimulus was presented.
- We suppose that the prior information gives us probabilities of the individual stimuli:

$$p(s = s_1) = \lambda, \;\; p(s = s_2) = 1 - \lambda$$

Then *Maximal a posteriori* (MAP) estimator would be:

$$\hat{s}_{MAP} = \begin{cases} s_1 \text{ if } p(s_1|r) > p(s_2|r) \\ s_2 \text{ otherwise} \end{cases}$$

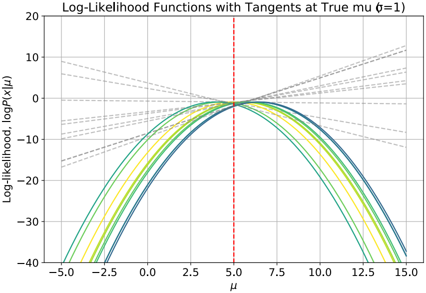
Minimal discriminator error



$$\mathrm{MDE} = \int \min\left\{\lambda p\left(\mathsf{r}|s_1\right), (1-\lambda) p\left(\mathsf{r}|s_2\right)\right\} d\mathsf{r}$$

Thus, the MDE is the blue-shaded area.

Summary



We learned:

- How to define and understand Fisher information
- What could be tweaked to improve coding by changing the tuning
- How to formulate the discrimination problem and compute the minimal discriminator error

Next week

The lecture and the tutorial are switched again!

- There is a CRC 1233 "Robust Vision" and Bernstein Lecture by Marion Silies: "From heterogeneous wiring to degenerative function in motion-detection circuits", at 16:15 in the HNO room (same room). Lecture is obligatory for GTC students.
- The tutorial shifting will be discussed with Tim (maybe doubling the tutorim for the week after, without going over-time)

Topics for the next two lectures

- How to deal with correlations between neurons
- What if the coding is not done by averaging neurons but by looking at population activity?