Neural Coding

4. Tuning curves. Encoding-Decoding

Anna Levina

May 7, 2025

Recap

- Define probabilities for sequences of spikes that correspond to the process with a given rate function $\lambda(t)$
- Generate these spikes with various different methods
- Decompose variability of spiking across different times in the trial to the rate variance and variability of spike generation

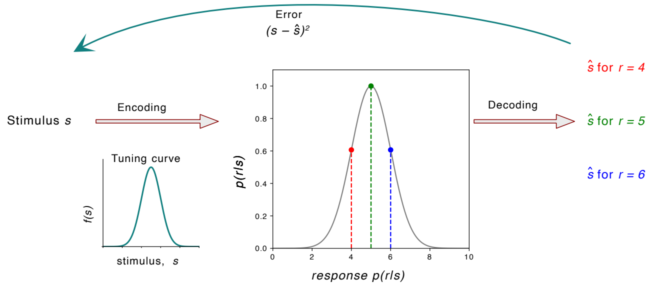
Today:

General setting: Using the rate function and generated variable number of spikes, reconstruct the presented stimulus.

Questions/tasks:

- Define how the stimulus will be translated into a response
- How to optimally reconstruct the stimulus?
- What are the favorable properties of tuning curves?

General setup



- Stimulus s, equipped with probability distribution $p(s)$
- Tuning function $f(s)$, ideal tuning
- Additional noise generates a distribution of responses for a given stimulus during the observation time T, $p(r \mid T, f(s))$
- From response r ideal observer can reconstruct the estimation of the stimulus $\hat{s}$. This is a *decoding*.
- Decoding error is $(s - \hat{s})^2$

Noise models 1. Normal noise

Normal noise

$$p(r \mid f(s)) \sim \mathcal{N}(\mu, \sigma^2)$$

- Additive: $\mu = f(s)$, $\sigma = const$
- Multiplicative: $\mu = f(s)$, $\sigma = f(s)$
- Problems: Generated firing rates might get to be negative

Noise models 2. Poisson Noise

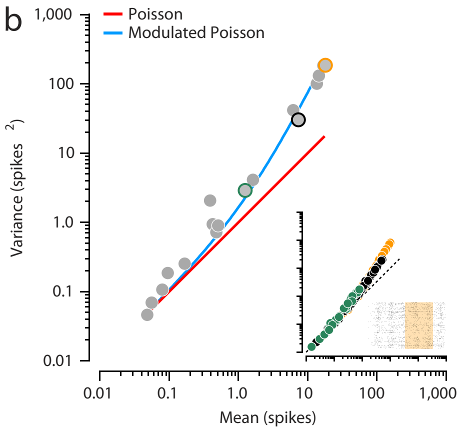
Poisson noise

$$p(r \mid f(s)) = \frac{f(s)^r}{r!} e^{-f(s)}$$

$$E(r \mid f(s)) = f(s),\ Var(r \mid f(s)) = f(s)$$

Fano-factor $F = \frac{\sigma^2}{\mu}$. For Poisson distribution $F = 1$
Definition Random variable with Fano factor larger than one is called *overdispersed*.

Poisson Noise: problem



For the real neurons, the Fano-factor is not 1 (Goris, Movshon, Simonchelli, 2014). Red: Poisson prediction, blue: real observations

Noise models 3. Negative Binomial

Response r is generated from the Negative Binomial distribution $NB(q, p)$:

$$p(r \mid q, p) = \binom{r + q - 1}{r} (1 - p)^q p^r$$

$$\xi \sim NB(q, p) \Rightarrow E\xi = \frac{pq}{1 - p},\ Var(\xi) = \frac{pq}{(1 - p)^2}$$

For neuronal response, we take $p = \frac{f(s)}{f(s) + q}$. $E(r \mid f(s)) = f(s)$ and $Var(r \mid f(s)) = \frac{f(s)}{1-p}$, thus Fano factor $F = \frac{1}{1-p} > 1$.
We can get the negative binomial response distribution from modulating the Poisson process with the Gamma-distributed gain Goris, Movshon, and Simoncelli, "Partitioning neuronal variability".

Noise models 4.Bernoulli noise

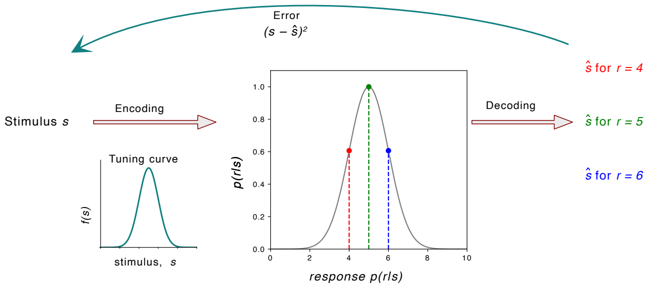
Bernoulli noise

$$p(r \mid f(s)) = f(s)^r(1 - f(s))^{1-r}$$

Requirements: $f(s) \in [0, 1]$
Reasonable for: response in a small time bin after stimulus onset

$$E(r \mid f(s)) = f(s),\ Var(r \mid f(s)) = f(s)(1 - f(s)),\ F = 1 - f(s)$$

Errors and big picture again



Error. Ideal observer

We had $\hat{s}$ an estimator of the stimulus s from the response. We define a mean squared error (MSE):

$$MSE(s) = E_r[(s - \hat{s}(r))^2 \mid s].$$

Averaging over s we get $S = E_s[E_r[(s - \hat{s}(r))^2 \mid s]]$ average risk. *Ideal observer* minimizes MSE:

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

To find the minimum, we solve the partial derivative $\frac{\partial}{\partial \hat{s}} \int \ldots ds = 0$.

$$\frac{\partial}{\partial \hat{s}} \int p(s \mid r)(\hat{s} - s)^2 ds = 2\hat{s} \int p(s \mid r) ds - 2 \int s \cdot p(s \mid r) ds$$

Hence we obtain (because $\int p(s \mid r) ds = 1$):

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

Variance of estimator and bias

For the shortening of notation, we write $\hat{s}$ for $\hat{s}(r)$

$$\begin{aligned}
\mathrm{MSE}(s) &= E[(s-\hat{s})^2 \mid s] = E[(s - E[\hat{s}] + E[\hat{s}] - \hat{s})^2 \mid s] \\
&= E[(s - E[\hat{s}])^2 \mid s] + E[(E[\hat{s}] - \hat{s})^2 \mid s] \\
&+ 2E[(E[\hat{s}] - \hat{s})(s - E[\hat{s}]) \mid s] \\
&= (s - E[\hat{s}])^2 + E[(E[\hat{s}] - \hat{s})^2 \mid s] = \sigma^2(s) + b^2(s)
\end{aligned}$$

Here $\sigma^2(s) = E\left[(E[\hat{s} \mid s] - \hat{s})^2\right]$ is a *variance of the estimator* $b(s) := E[\hat{s} \mid s] - s$ is *bias*. So we can write

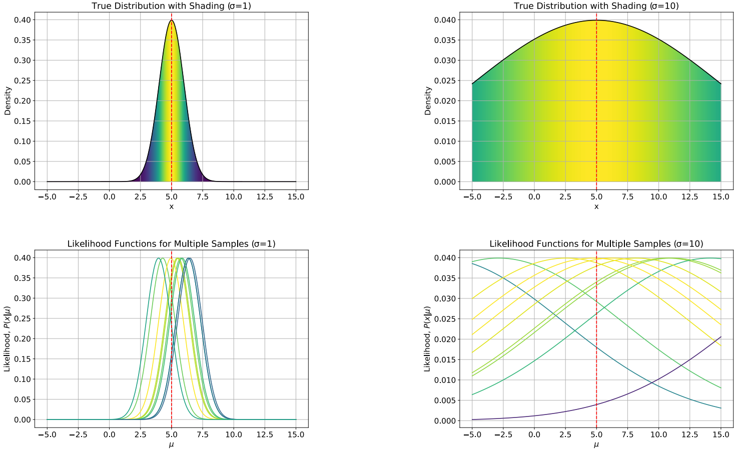
$$MSE(s) = \sigma^2(s) + b^2(s)$$

The estimator is unbiased if

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

Fisher Information 1. Probability and likelihood

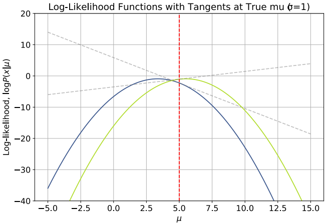
Fisher information quantifies the amount of information that an observable random variable X carries about an unknown parameter θ upon which the probability of X depends.

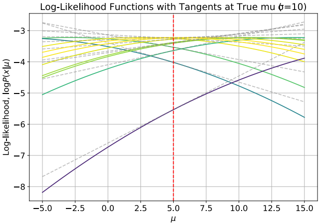


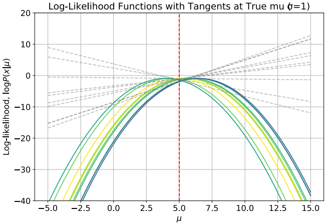
Fisher Information 2. Score function

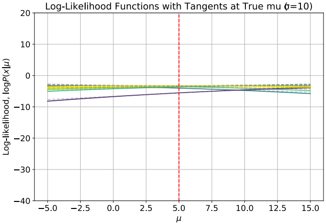
Score function $S_X(\theta)$:

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









Fisher Information 3. Finally definition

Fisher information is a variance of the score:

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

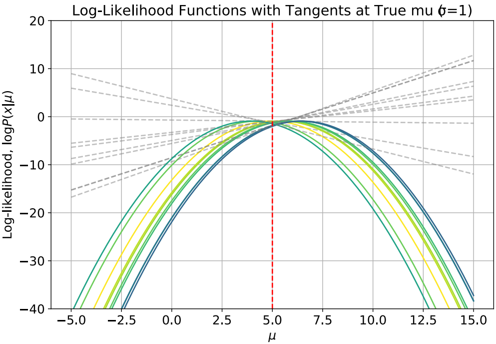
$$= \int\left(\frac{\partial}{\partial\theta}\log p(X|\theta)\right)^2 p(X|\theta)dX = -\mathrm{E}_X\left[\left(\frac{\partial^2}{\partial\theta^2}\log p(X|\theta)\right)|\theta\right]$$

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 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 a lot the properties of the estimator. 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

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