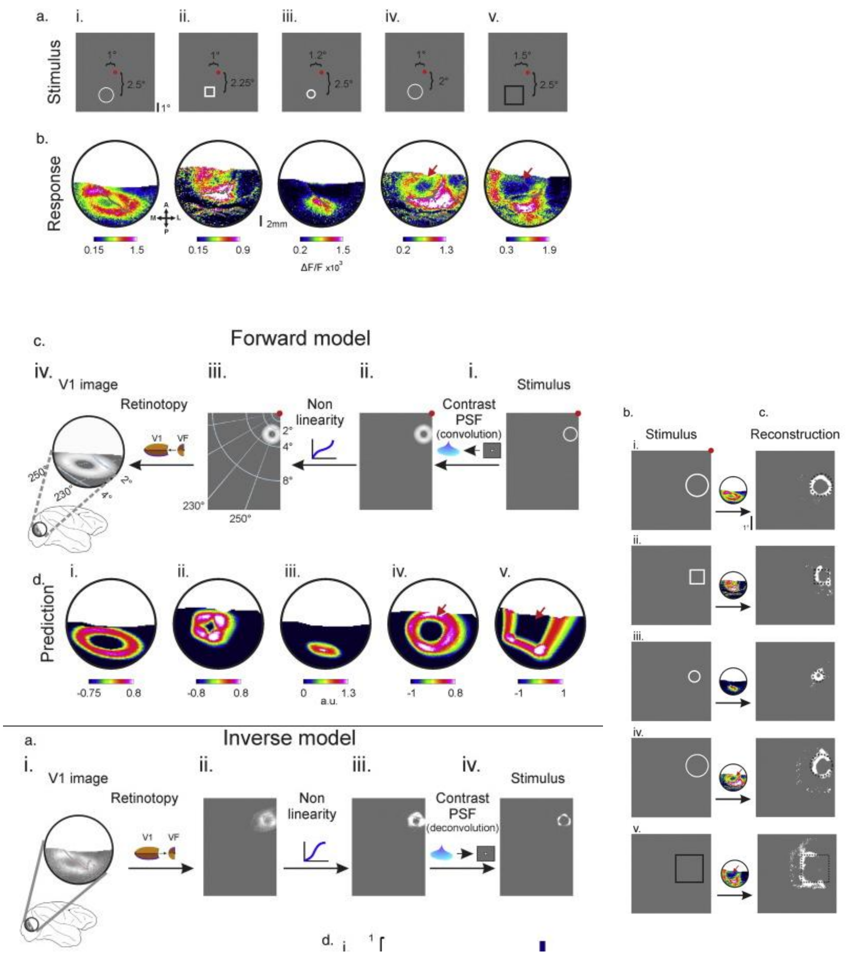
Written by Yarden Nativ, mail: nativyardenac@gmail.com

## 6th Recitation 4.5.23

## Population coding, MAP, MLE

# Neural decoding

Neural decoding is commonly considered as “mind reading” although practically this is a very important step used in a wide range of studies. For example, studies aiming to construct a prothesis replacing neuronal functions, for example visual prothesis to assist blind people, are interested in general decoding of the neural activity back to the stimulus given. For example (adopted from Zurawel et al. 2016):




Decoding involves many kinds of transformations, in which most of them are non-linear. In this course we will focus on the linear transformations:

Written by Yarden Nativ, mail: nativyardenac@gmail.com

## Examples from past exams: population vector

**2010 exam:** The Green Grass Goblin has four neurons for identifying Elf movements with different optimal directions.

The neurons fired at the following rates for an Elf approaching from direction X:

| Optimal direction | 0 | 90 | 180 | 270 |
|---|---|---|---|---|
| Firing rate for X | 20 | 10 | 4 | 8 |

All the neurons have a max FR of 30 spikes/s and baseline FR is 0 sp/sec.

Calculate the movement direction of the Elf using a population vector.

$$\bar{v}_{population} = \sum_{i=1}^{N} \frac{r_i - r_0}{r_{max}} \cdot \bar{v}_i$$

**Solution:**

The given sum is based on normalized firing rate to baseline. Each neuron can have its own $r_0$ and $r_{max}$, and in such a case we would assign the corresponding value of the neuron. Nevertheless, for this question $r_0, r_{max}$ will be the same for all neurons.

To calculate the vector of each neuron, we can write the directions using equivalent unit vectors:

$$0° = [1\ 0] \quad 90° = [0\ 1] \quad 180° = [-1\ 0] \quad 270° = [0 - 1]$$

Therefore, the population vector is given by:

$$\left(\frac{20}{30}, \frac{10}{30}, \frac{4}{30}, \frac{8}{30}\right) \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix} = (0.5333, 0.067)\ \ (x, y)$$

And the real direction of the elf is:

$$\alpha = \arctan\left(\frac{y}{x}\right) = 7.1°$$

## Class discussion:

- What assumptions we must have for population decoding analysis?
- How useful are population vectors for studying neuronal behavior?

**Estimators:**

Given a model function $f(x)$ and a given set of observations $X = \{x_1, x_2, \ldots, x_n\}$, we would like to estimate the model parameters. For example, for a polynomial model $f(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots$ estimating the parameters is finding the values (in numbers) of $\{\theta_0, \theta_1, \theta_2, \ldots\}$. At this level, we are not interested in evaluating the model itself in relation to other possible models. We use it as a ground truth and try to find the parameters which explain in the best way the results.

Hereinafter are the three estimators we will use:

- Maximum likelihood estimator (MLE):

$$\widehat{\Theta} = \arg\max P(x|\Theta)$$

- Maximum a posterior estimator (MAP):
  Same as MLE without the a posterior knowledge (non-Naïve)

$$\widehat{\Theta} = \arg\max P(\Theta|\mathrm{x})$$

$$P(\Theta|x) = \frac{p(x|\Theta) \cdot p(\Theta)}{p(x)}$$

$$\widehat{\Theta} = \arg\max P(x|\Theta) \cdot p(\Theta)$$

- Bayesian approach:
  Given the cost function $C(\epsilon = \Theta - \widehat{\Theta})$ and $x$:

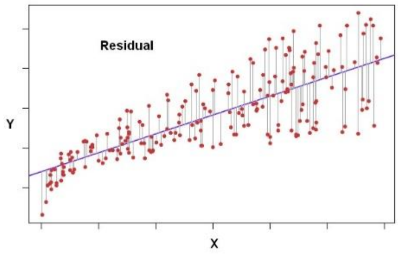
$$\widehat{\Theta} = \arg\min_{\widehat{\Theta}} \int_{\Theta} C\left(\Theta, \widehat{\Theta}\right) \cdot p(\Theta|x) d\Theta$$

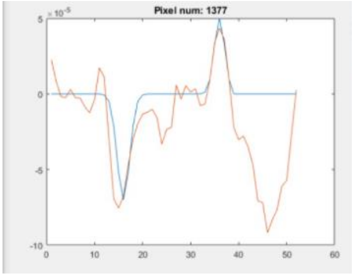
**Optimizing- computationally and analytically**

If we want to evaluate a model computationally, we use an error function which describes for a given set of parameters how good is our model. Commonly, root mean square error (RMSE, aka residuals) is a useful tool for this evaluation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Model_i - Actual_i)^2}{N}} \qquad N - number\ of\ measurements$$

**Two examples:**




If we want to optimize it analytically, we will use the derivative of the given function and find its maximum.

## Examples from past exams:

**2005 exam:** During 5 trials the recorded rates of the neuron were: 70, 62, 95, 59, 65. Assuming that the recorded rate is the result of a Poisson distribution, find the MLE for the parameter λ.

**Solution:**

We will rewrite the question data for estimation $\bar{X} = [59,62,65,70,95]$ from a given Poisson distribution $\frac{\lambda^x e^{-\lambda}}{x_i!}$.

We will use the estimator MLE as defined:

$$\hat{\lambda} = \arg\max_{\lambda} P(\bar{X}|\lambda) \qquad /MLE$$

The function for estimation is:

$$f(\bar{X}|\lambda) = \prod_i \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

Using logarithmic function it can be solved by:

$$\ln\left(f(\bar{X}|\lambda)\right) = \sum_i \ln\left(\frac{\lambda^{x_i} e^{-\lambda}}{x_i!}\right)$$

$$F = \sum_i \ln\left(\lambda^{x_i} e^{-\lambda}\right) = \sum_i x_i \ln(\lambda) - \lambda \cdot N$$

To find the maximum, we will use the derivative:

$$\frac{dF}{d\lambda} = \frac{\sum_i x_i}{\lambda} - N = 0 \rightarrow \lambda = \frac{1}{N}\sum_i x_i$$

Therefore:

$$\hat{\lambda} = E(x_i)$$

And in our case:

$$\hat{\lambda} = 70.2$$

**2007 exam:** The time difference (in weeks) between occurrences of amnesia in Foergetis Homeworkis patients may generally be estimated by an exponential distribution:

$$p(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

The parameter λ describing the distribution is different for each patient. The five observations of amnesia in patient X are spaced by 2, 1, 5, 1, 3 weeks.

Assuming the prior distribution $P_0(\lambda) = \frac{1}{2}\lambda \quad (0 \leq \lambda \leq 2)$ known for the general patient population, what is the maximum a-posteriori (MAP) estimator?

**Solution:**

As a first step we should note the model and its parameters:

$$f(x) = \lambda e^{-\lambda x} \rightarrow \lambda \; is \; the \; parameter \; to \; estimate$$

We will re-write the data in the question using the notations for estimators:

Given $p(\lambda) = 0.5\lambda$, $\bar{X} = [1,1,2,3,5]$ than MAP is given by:

Map estimator: $\hat{\lambda} = \arg\max_{\lambda} P(\lambda|\bar{X})$ → $\hat{\lambda} = \arg\max P(x|\lambda) \cdot p(\lambda)$

We will assume the variables of $x_i$ are independent random variables:

$$F = \prod_i P(x_i|\lambda) \cdot P(\lambda) = \lambda e^{-\lambda \cdot 1} \cdot \lambda e^{-\lambda \cdot 1} \cdot \lambda e^{-\lambda \cdot 2} \cdot \lambda e^{-\lambda \cdot 3} \cdot \lambda e^{-\lambda \cdot 5} \cdot 0.5\lambda$$

$$F = \lambda^5 e^{-12\lambda} \cdot \frac{\lambda}{2} = \frac{1}{2}\lambda^6 e^{-12\lambda}$$

To find the max $\lambda$, we will use a derivative:

$$\frac{dF}{d\lambda} = \frac{1}{2}\lambda^5 e^{-12\lambda}(6 - 12\lambda) \rightarrow \begin{matrix} \lambda_1 = 0 \\ \lambda_2 = \frac{1}{2} \end{matrix}$$

To decide which one is the maximum, we will check the value of $P(\lambda_1|\bar{X})$, using Bayes $P(\lambda|\bar{X}) = \frac{P(\bar{X}|\lambda)\cdot P(\lambda)}{P(\bar{X})}$

$$\lambda_1 = 0 \quad P(\lambda_1|\bar{X}) = 0$$

$$\lambda_2 = \frac{1}{2} \quad P(\lambda_2|\bar{X}) > 0$$

Therefore $\lambda_2 = \frac{1}{2}$ is the point for maximum value, and we can conclude $\hat{\lambda} = \frac{1}{2}$.

**2006 exam:** The time difference (in hours) between occurrences of hiccups in Hiccupitis Neuralitis Syndrome patients may be estimated by the following uniform distribution:

$$p(x) = \begin{cases} \frac{1}{n} & 0 \le x \le n \\ 0 & x > n \end{cases}$$

The parameter n describing the distribution is different for each patient. The five observations of hiccups in patient X are spaced by 2, 3, 7, 2, 1 hours.

- Find the maximum likelihood (ML) estimator for n in patient X and explain the results.
- Assuming the priors $P_0(7)$=0.1 & $P_0(8)$=0.9 known for the general patient population. Is the maximum aposteriori (MAP) estimator different from the ML estimator calculated in section (a)?

**Solution:**

Rewriting the data: $\bar{X} = [1,2,2,3,7], P(x_i) = \frac{1}{n}$

The MLE is:

$$\hat{n} = \arg\max_n P(\bar{X}|n) = \prod_i P(x_i|n) = \left(\frac{1}{n}\right)^N$$

We can see that the smaller $n$ is, the higher is the value for $n$. We don't know what is the real $n$ of the distribution, but we know that $n \geq 7$, the highest value from the observations for $x$. Therefore, maximizing the range will give us $n = 7$.

The MAP estimator can use the given probabilities to evaluate that there are more patients for which the smallest $n$ is $n = 8$ than those with $n = 7$. So for the MLE:

$$\hat{n} = \arg\max_n P(n|\bar{X})$$

Using Bayes:

$$P(n|\bar{X}) = \frac{P(\bar{X}|n) \cdot P(n)}{P(\bar{X})}$$

Therefore:

$$F = P(\bar{X}|n) \cdot P(n) = \prod_i P(x_i|n) \cdot P(n)$$

$$F = \left(\frac{1}{n_0}\right)^N \cdot P(n = n_0)$$

$$F(n = 7) = \left(\frac{1}{7}\right)^5 \cdot 0.1 = 6 \cdot 10^{-6}$$

$$F(n = 8) = \left(\frac{1}{8}\right)^5 \cdot 0.9 = 2.7 \cdot 10^{-5}$$

We got for $n = 8$ higher probability, therefore using the MAP estimator we get a different estimator, $\hat{n} = 8$