Machine Learning for Cognitive Sciences: Principles and Applications

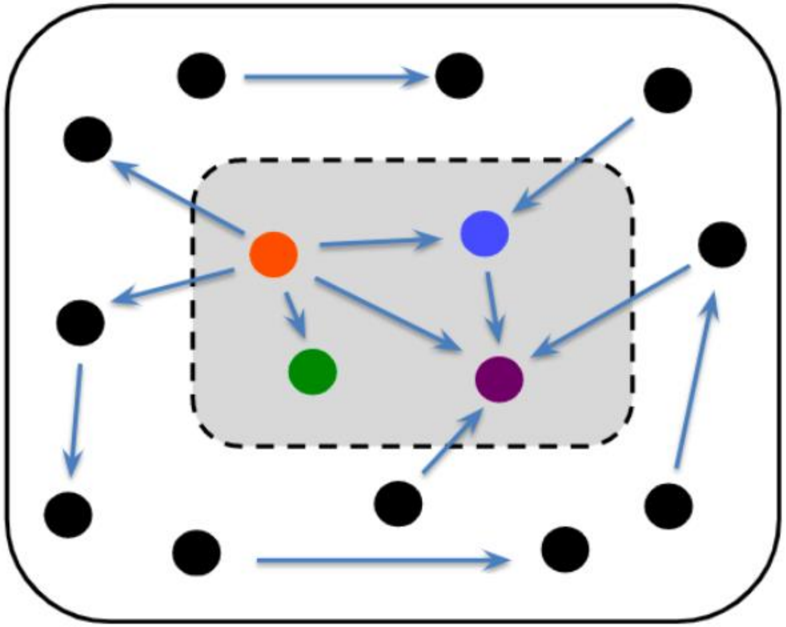
Simona Cocco (Physics Departement of ENS)
simona.cocco@phys.ens.fr CM &TD

Vito Dichio (Physics Departement of ENS)
vito.dichio@phys.ens.fr TD

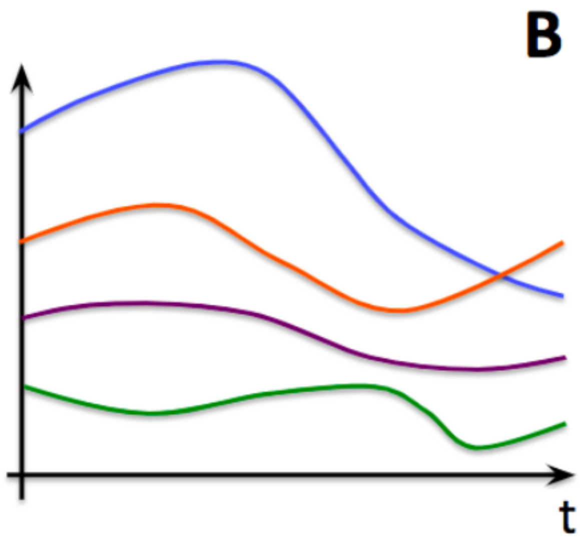
Director of the Cogmaster: Jean Baptiste Andre
jean-baptiste.andre@ens.fr

https://moodle.psl.eu/enrol/index.php?id=34555
Accès anonyme pwd:MLCS2025
Class: Wednesday salle RIBOT,
Tutorial Thursday salle Langevin

Machine Learning for Cognitive Sciences: Principles and Applications

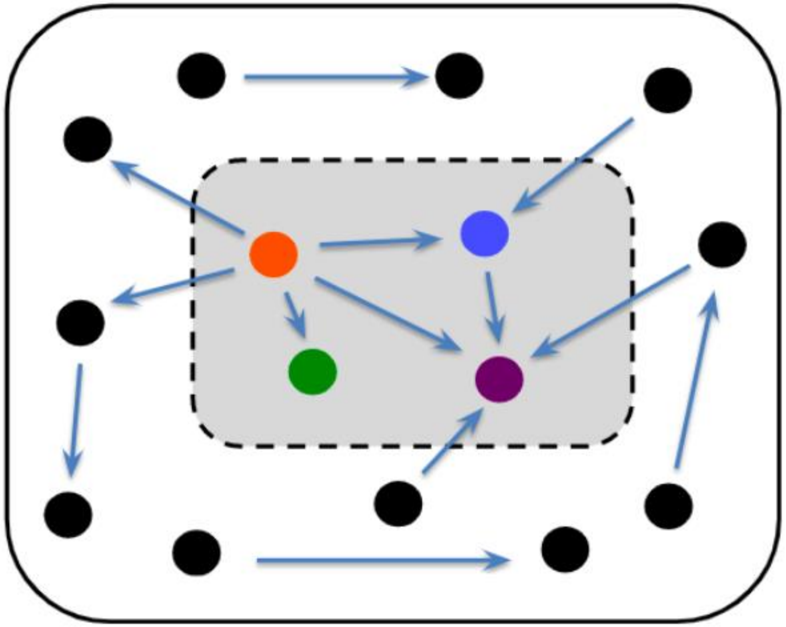


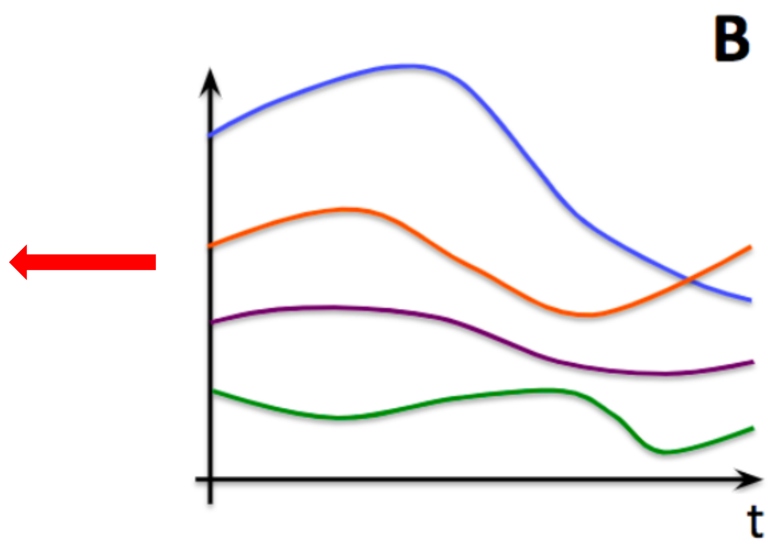
A complex system made of many particules which interact together



Observation of a limited part of the system and of the dynamics of its components

Machine Learning for Cognitive Sciences: Principles and Applications





A complex system made of many particules which interact together

Observation of a limited part of the system and of the dynamics of its components

Aim: Understand how the system works, Predict the future behavior of it.

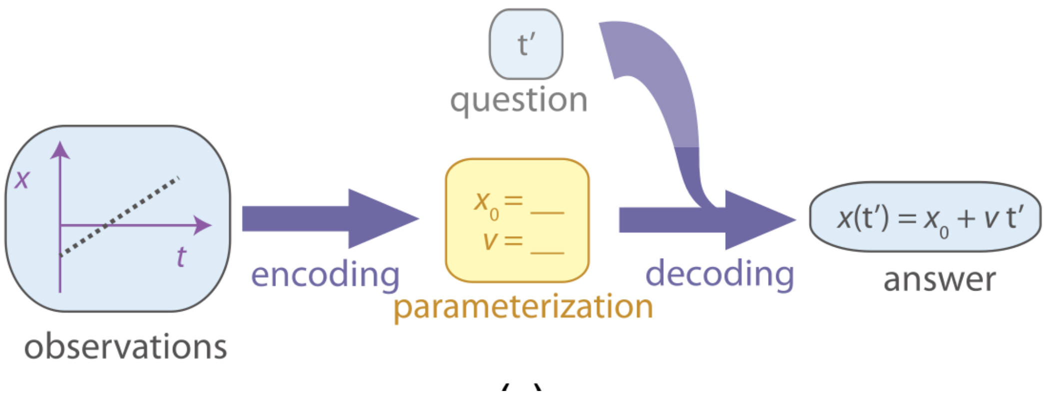
Machine Learning for Cognitive Sciences:

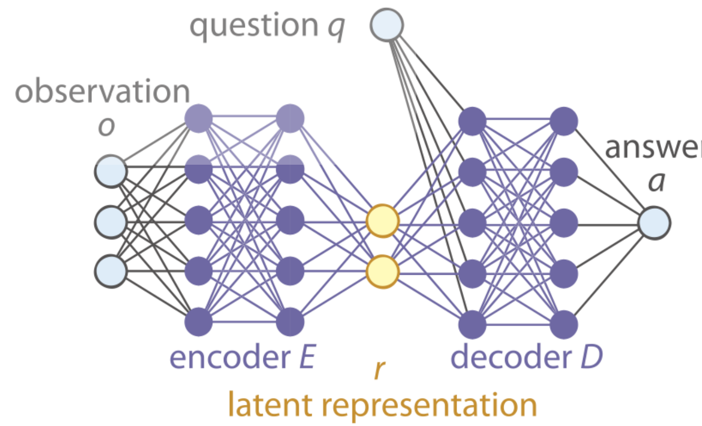
Principles and Applications

Exemples of ML utilization in Neuroscience (Neural recodings), Images (MNIST) and Language Classification (Newspaper texts), behavioral Data (trajectories of zebrafish).

ML as a learning and coding paradigm for real neural networks .

Machine Learning for Cognitive Sciences: Principles and Applications





Auto-Encoder with SciNet

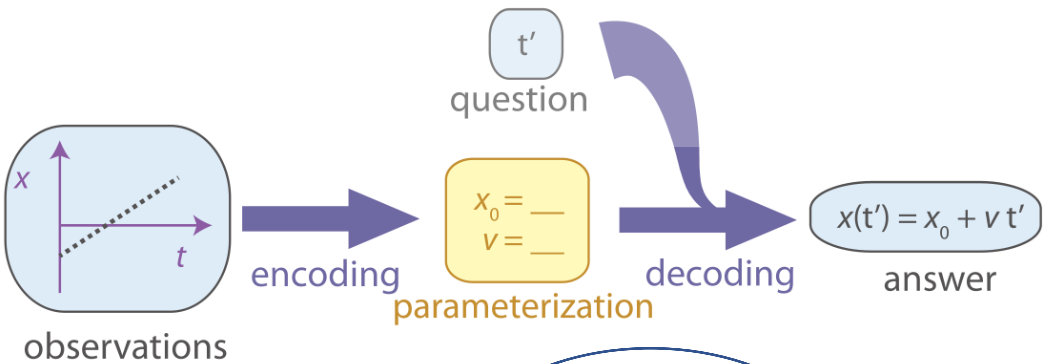
Use the network to

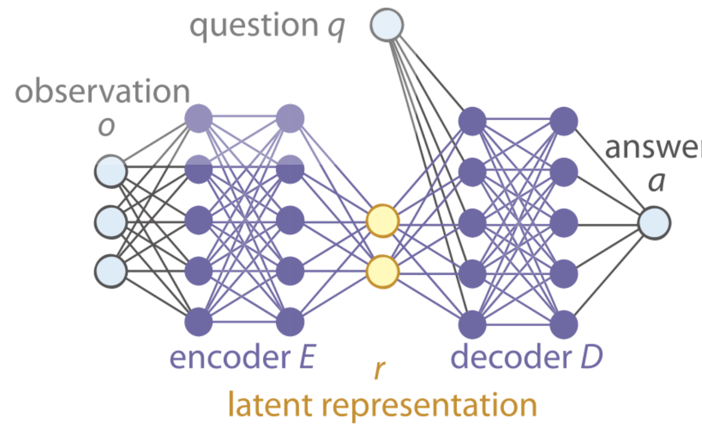
predict the behavior of the system

Interpret the parameters: how they encode the input

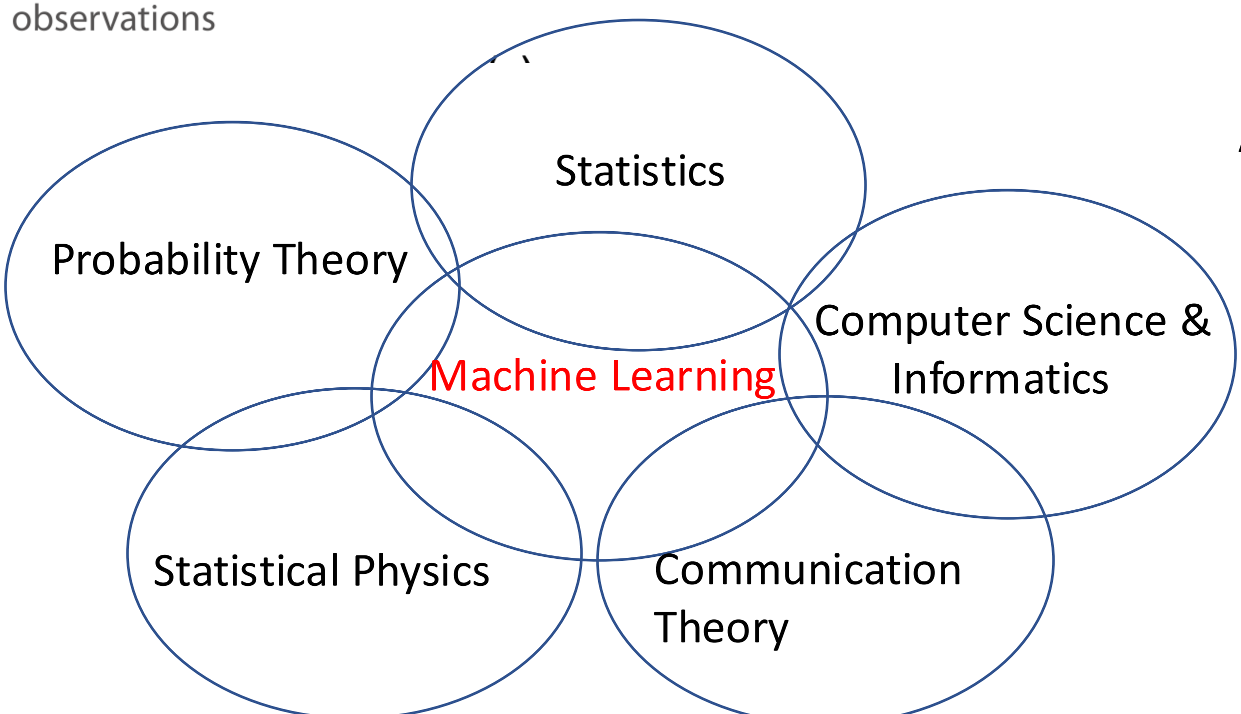
Find the best network: related to the number of data

Machine Learning for Cognitive Sciences: Principles and Applications

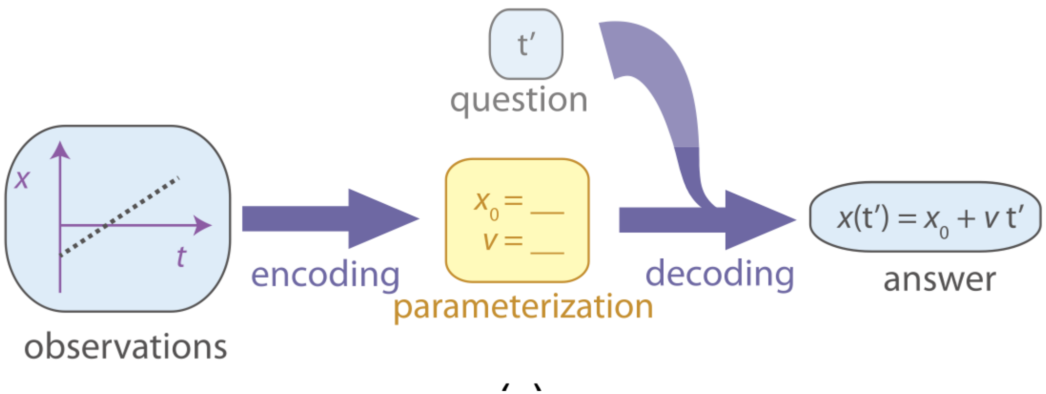


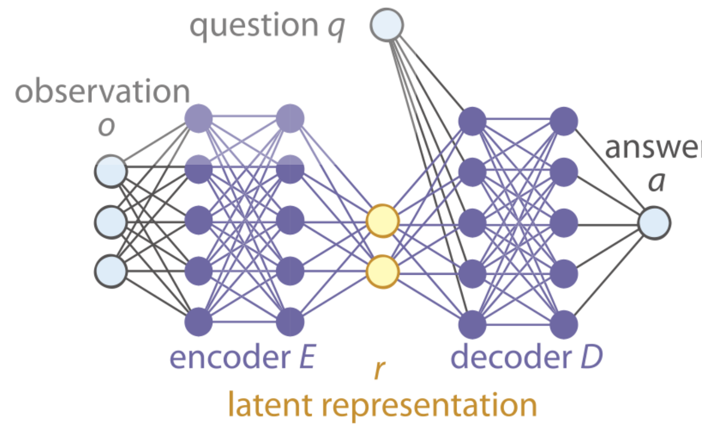


Auto-Encoder with SciNet



Machine Learning for Cognitive Sciences: Principles and Applications



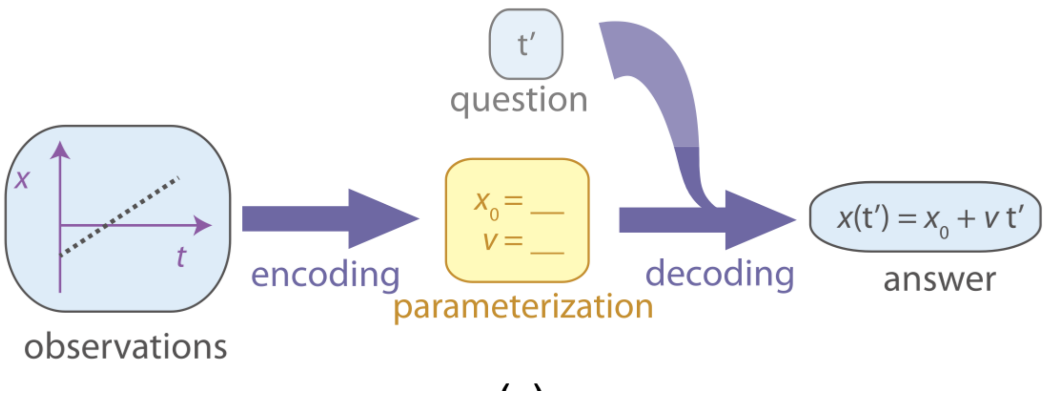


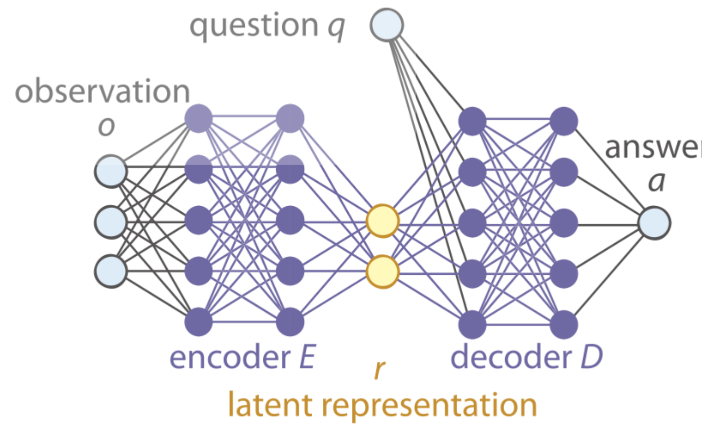
Auto-Encoder with SciNet

Aims of the Course:

- Learn **Bayesiam framework** for the Inference.
- Learn Inference Algorithm, **Programming** yourself in Python to avoid black box effect,
- Control learning: **Statistical Physics & Statistics** to understand how to adapt the inferred models to the quality/quantity of data: eg. avoid **Overfitting** (use **Regularizations**) , Compare with **Null models**, Use **Cross Validations.**

Machine Learning for Cognitive Sciences: Principles and Applications





Auto-Encoder with SciNet

Aims of the Course:

- Learn **Bayesiam framework** for the Inference.
- Learn Inference Algorithm, **Programming** yourself in Python to avoid black box effect,
- Control learning: **Statistical Physics & Statistics** to understand how to adapt the inferred models to the quality/quantity of data: eg. avoid **Overfitting** (use **Regularizations**) , Compare with **Null models**, Use **Cross Validations.**
- **Unsupervised Learning: Graphical models, Multi-Layer networks .**
- **Clustering** and **low dimensional representation of data** (eg. Principal Component Analysis)
- **Supervised Learning**: Neural Network for classification (**Perceptron algorithm**).
- **Time-dependent process**: Markov-Model and Hidden Markov-Model.

Machine Learning for Cognitive Sciences:
Principles and Applications

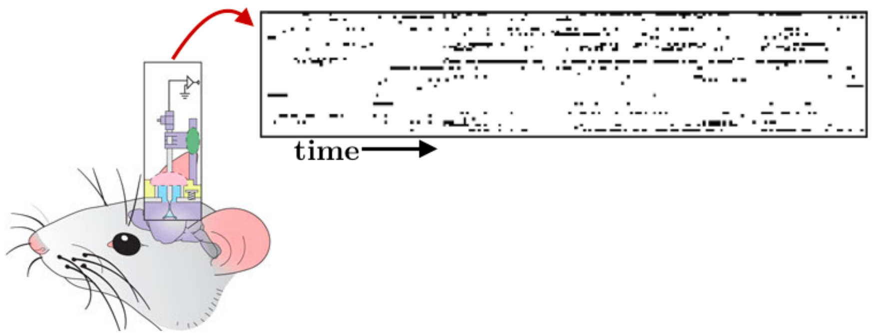
Tutorials based on programming in **Python** to **analyse** real data:

Data: Neural Recordings, Images data, behavioral recordings

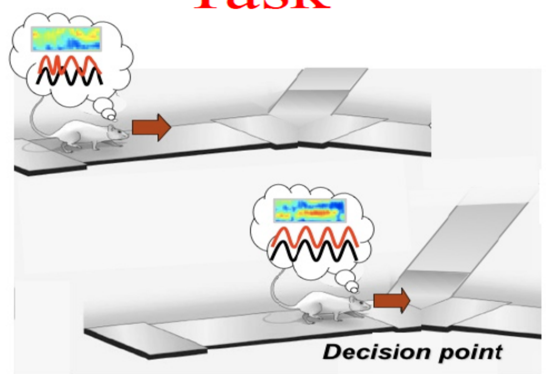
You need a laptop with jupyter notebook installed (eg. by anaconda): For tomorrow:

See:**https://moodle.psl.eu/enrol/index.php?id=34555**

Tutorial 4: Principal Component Analysis to study memory consolidation during the Sleep after a task



Sleep Pre → Task → SleepPost



Replay of rule-learning related neural patterns in the prefrontal cortex during sleep

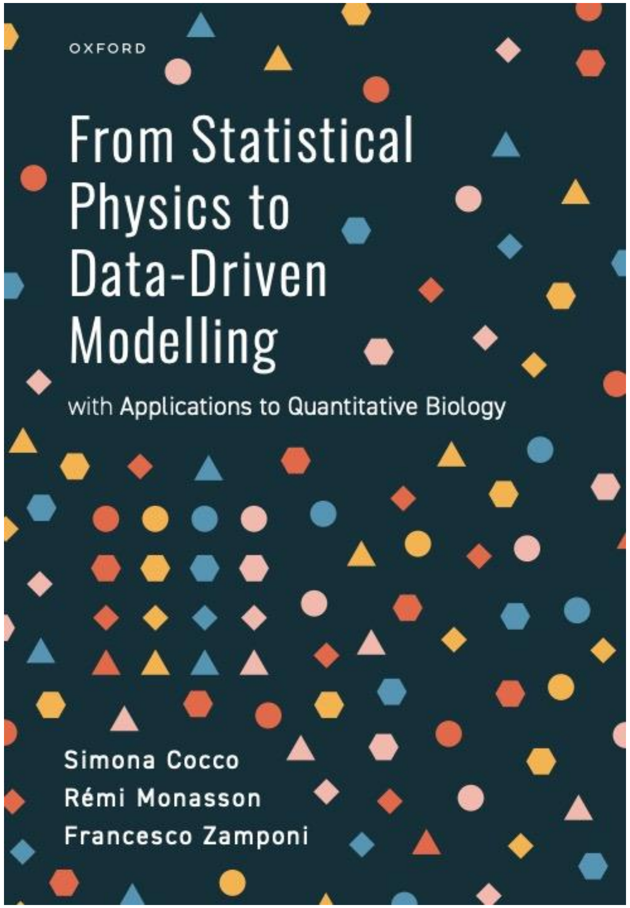
nature neuroscience (2009)

Adrien Peyrache[1], Mehdi Khamassi[1,2], Karim Benchenane[1], Sidney I Wiener[1] & Francesco P Battaglia[1,3]

How to follow the Course

Moodle contains:
https://moodle.psl.eu/enrol/index.php?id=34555

- A pdf version of the book
- Each week: The slides,
- The tutorials (data and starting notebook) ,
- The final notebook after the tutorials
- Check you have access to it.

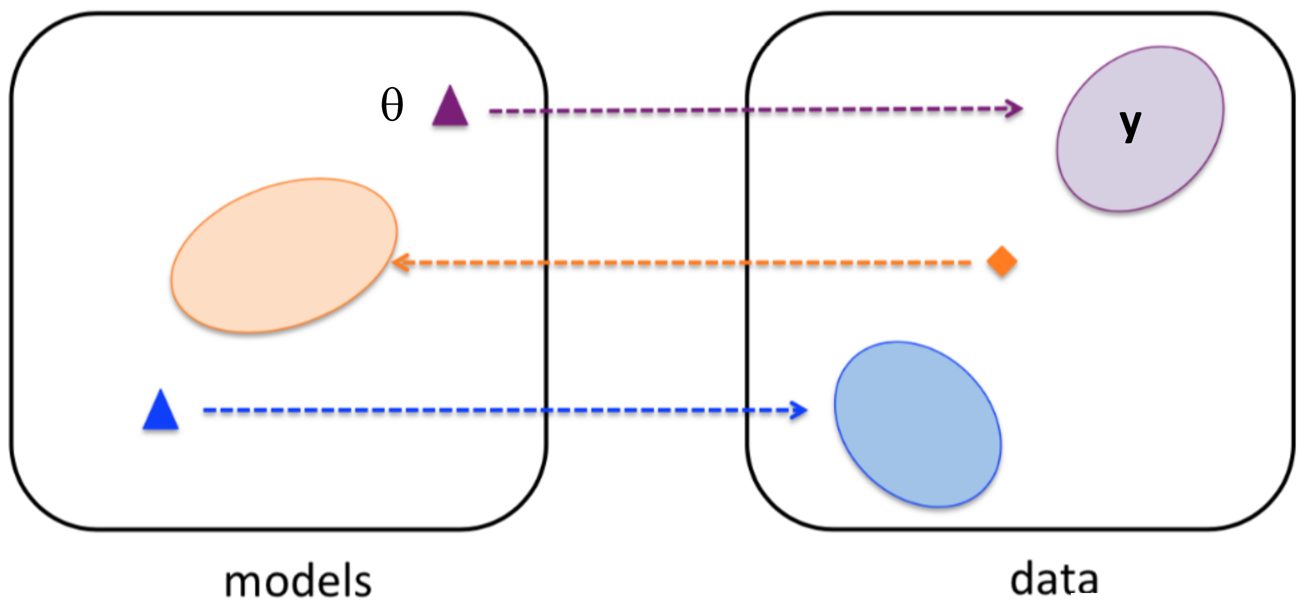


Class each Wednesday 8h30-10h30
Tutorials each Thursday 8h30-10h30

- Install Jupyter Notebook (Anaconda) Phython 3
- Revise lectures before the tutorials
- Attendance is mandatory.
- Final scores: One homework before christmas,

  \+ final examination : python program based on the tutorial in class
  and a session of theoretical questions

Bayesian Inference



- Each model defines a distribution of possible data

- Each data defines a distribution of possible models

Probabilities and conditional probabilities

Dice : faces y = 1,2,3,4,5,6 (here, y = 3)

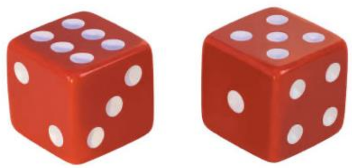
Notation: Probability (face = y) $\equiv p(y)$

Properties : $p(y) \geq 0, \sum_{y=1,..6} p(y) = 1$

Unbiased dice : each face is equally likely through a draw

$$p(1) = p(2) = \ldots = p(6) = \frac{1}{6}$$

Probabilities and conditional probabilities



Two dices : faces y_1 and y_2 (here, y_1 =6, y_2 = 5)

Joint probability (1st dice = y_1 & 2nd dice = y_2) $\equiv p(y_1, y_2)$

Two important considerations:

- Care about one event only -> definition of **marginal probability**

$$p_1(y_1) = \sum_{y_2} p(y_1, y_2) \ , \ p_2(y_2) = \sum_{y_1} p(y_1, y_2)$$

- Ask whether events are **independent,** i.e. are the two dices correlated or not?

$$p(y_1, y_2) = p_1(y_1) \times p_2(y_2) \ ?$$

An example of dependent events

Box A

20 plain cookies
\+
20 chocolate cookies

Box B

10 plain cookies
\+
30 chocolate cookies

One picks up uniformly at random one box and one cookie out of this box

Box: $x_1 = A$ or B □ $$p_1(A) = p_1(B) = \frac{1}{2}$$

Cookie type: $x_2 = plain$ or $chocolate$ □ $$p_2(plain) = \frac{3}{8}$$

$$p_2(chocolate) = \frac{5}{8}$$

Conditional Probability

but $$p_2(chocolate|A) = \frac{1}{2} < p_2(chocolate) = \frac{5}{8} < p_2(chocolate|B) = \frac{3}{4}$$

An example of dependent events

Box *A*

20 plain cookies
+
20 chocolate cookies

Box *B*

10 plain cookies
+
30 chocolate cookies

One picks up uniformly at random one box and one cookie out of this box

Box*:* θ = *A* or *B* □ $p_1(A) = p_1(B) = \frac{1}{2}$

Cookie type: *y* = *plain* or *chocolate* □ $p_2(plain) = \frac{3}{8}$

$$p_2(chocolate) = \frac{5}{8}$$

- When the events are **dependent:**

$$p(y \;\; , \theta \;\;) = p_1(\theta) \times p_2(y \;\; | \theta \;)$$

Characterization of dependent events

Definition of **conditional probability:**

$$p_2(y|\theta) \equiv \frac{p(y,\theta)}{p_1(\theta)}$$

<u>Exercise 1:</u> *check that marginal probability over the boxes give the probabilities of Chocolate of plain cookies computed in previous slide .*

<u>Exercise 2:</u> *check that conditional probability is normalized.*

Box A	Box B
20 plain cookies + 20 chocolate cookies	10 plain cookies + 30 chocolate cookies

$p_2(y = \text{plain} \mid \theta = A) = 1/2$,
$p_2(y = \text{chocolate} \mid \theta = A) = 1/2$

$p_2(y = \text{plain} \mid \theta = B) = 1/4$,
$p_2(y = \text{chocolate} \mid \theta = B) = 3/4$

Characterization of dependent events

Definition of **conditional probability:** $$p_1(\theta|y) \equiv \frac{p(y,\theta)}{p_2(y)} \equiv \frac{p_2(y|\theta) \times p_1(\theta)}{p_2(y)}$$

Probability that the Box is A or B given that I have drawn a plain cookie

Box A	Box B
20 plain cookies + 20 chocolate cookies	10 plain cookies + 30 chocolate cookies

$p_2(\theta=A \mid \text{plain}) = ?$

$p_2(\theta=B \mid \text{plain}) = ?$

Characterization of dependent events

1/2

Definition of **conditional probability**:

$$p_1(\theta|y) \equiv \frac{p_2(y|\theta) \times p_1(\theta)}{p_2(y)}$$

3/8

Probability that the Box is A or B given that I have drawn a plain cookie

Box A

20 plain cookies
+
20 chocolate cookies

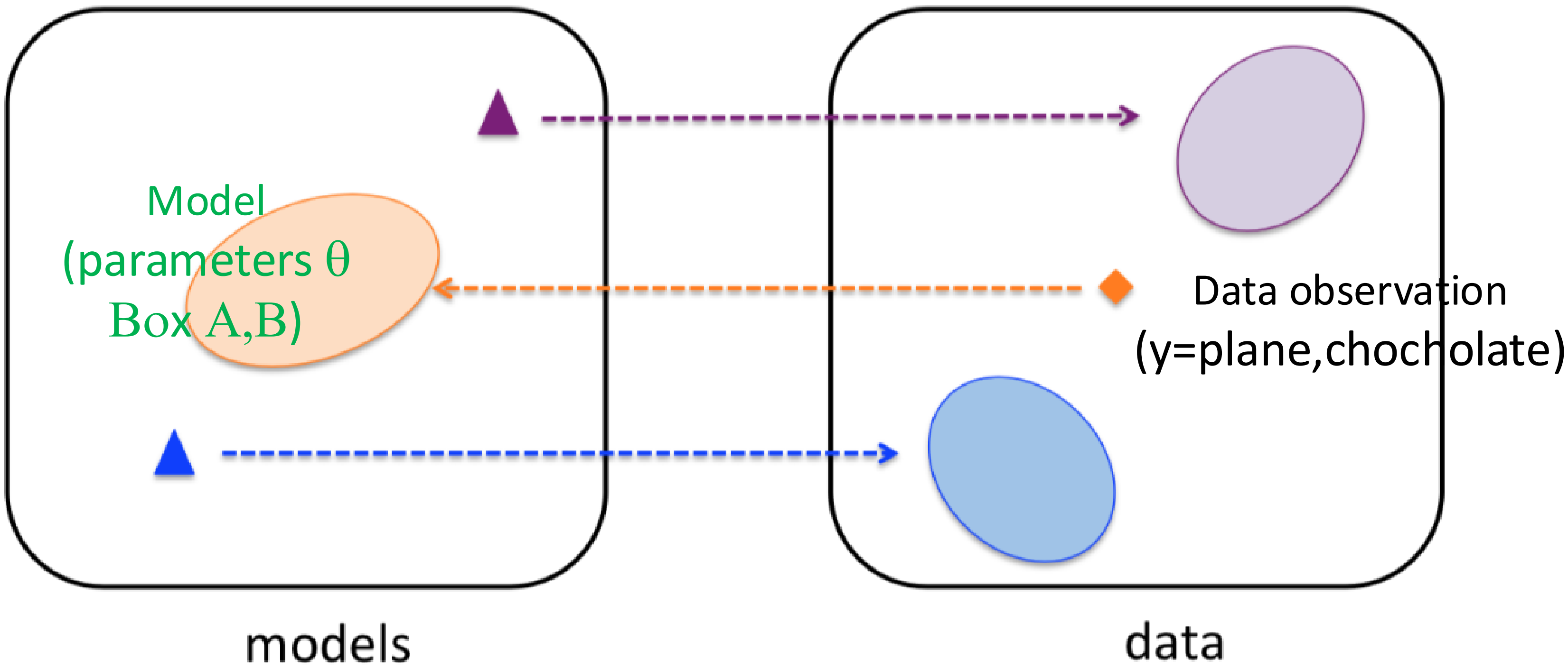
Box B

10 plain cookies
+
30 chocolate cookies

$$p_2(\theta = A \mid \text{plain}) = \frac{{}^1/_2 \times {}^1/_2}{{}^3/_8} = \frac{2}{3}$$

$$p_2(\theta = B \mid \text{plain}) = \frac{{}^1/_4 \times {}^1/_2}{{}^3/_8} = \frac{1}{3}$$

Bayesian Inference



Likelihood of data
given the model

A priori distribution of
the model parameters

$$\mathrm{p}(\theta|y) = \frac{p(y|\theta)\times p(\theta)}{p(y)}$$

Posterior of model
parameter given the data

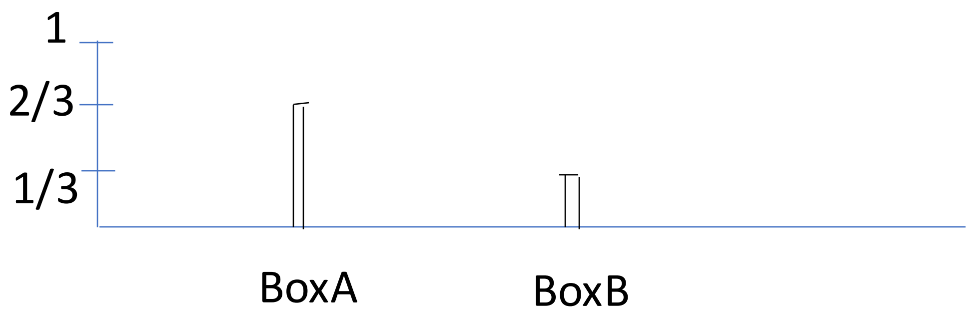
Probability of that data:
Evidence

An Illustration: Model (Box) Inference from the Bayes formula

Suppose you have drawn a plane cookie

Posterior of the model parameters given the data

$\mathrm{p}(\theta|plane)$



- Most Probable model: Box A $\quad p_2(\text{plane})=1/2$, $p_2(\text{pchocolate})=1/2$

From the model you can predict the probability of having a chocolate cookie if you draw again from the same box

- Probability that the prediction is wrong: 1/3

How you can increase the confidence in your model?
->lower the probability of a wrong prediction

Laplace and the birth rate of boys & girls

Historical example: « proof » by Laplace that the female and male birth rates are different

Data: Nbs of girls born in Paris from 1745 to 1770 : 245,945

... boys ... : 251,527

y = nb. of female births, M = total number of births

Laplace and the birth rate of boys & girls

Historical example: « proof » by Laplace that the female and male birth rates are different

Data: Nbs of girls born in Paris from 1745 to 1770 : 245,945

... boys ... : 251,527

y = nb. of female births, M = total number of births

Inference: θ= probability that a newborn baby is a girl

- Prior distribution: uniform over θ in $[0;1]$

INTUITION: $\theta\sim$

Laplace and the birth rate of boys & girls

Historical example: « proof » by Laplace that the female and male birth rates are different

Data: Nbs of girls born in Paris from 1745 to 1770 : 241,945

… boys … : 251,527

y = nb. of female births, M = total number of births

Inference: θ= probability that a newborn baby is a girl

- Prior distribution: uniform over θ in [0;1]

INTUITION: $\theta \sim \frac{241{,}945}{493{,}472} = 0.4903$

Laplace and the birth rate of boys & girls

Historical example: « proof » by Laplace that the female and male birth rates are different

Data: Nbs of girls born in Paris from 1745 to 1770 : 241,945

... boys ... : 251,527

σ = nb. of female births, n = total number of births

Inference: θ = probability that a newborn baby is a girl

- Prior distribution: uniform over θ in [0;1]
- Likelihood: $\mathrm{p}(y|\vartheta) = \binom{M}{y}\ \vartheta^{y}(1-\vartheta)^{M-y}$

Binomial Distribution.

Laplace and the birth rate of boys & girls

Historical example: « proof » by Laplace that the female and male birth rates are different

Data: Nbs of girls born in Paris from 1745 to 1770 : 241,945

... boys ... : 251,527

σ = nb. of female births, n = total number of births

Inference: θ = probability that a newborn baby is a girl

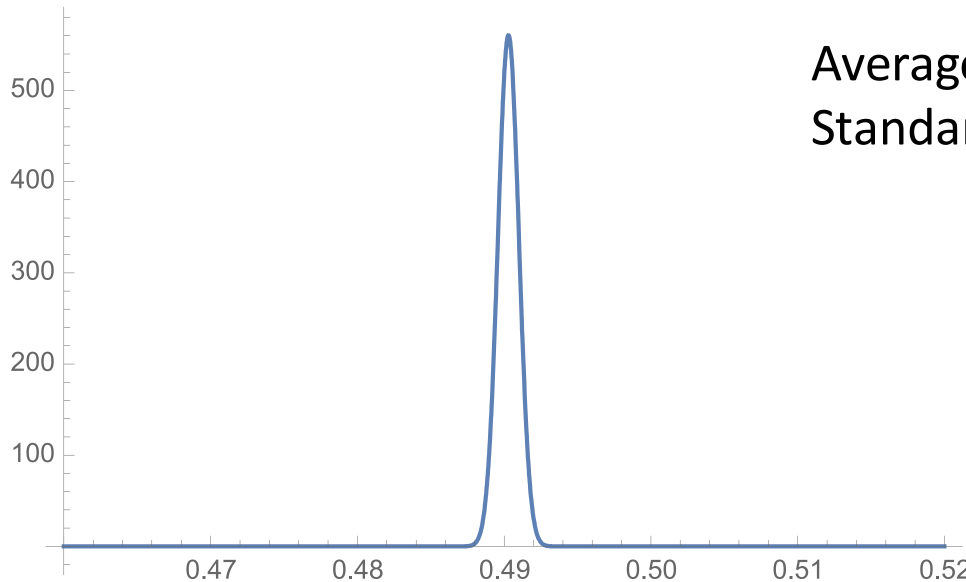
- Prior distribution: uniform over θ in [0;1]
- Likelihood: $p(y|\theta) = \binom{M}{y}\ \vartheta^{y}(1-\vartheta)^{M-y}$
- Bayes: $$p(\theta|y) = \frac{p(y|\theta)\times p(\theta)}{p(y)}$$

Uniform in interval [0,1]

Cst $\int_0^1 d\theta\ \theta^{y}\,(1-\theta)^{M-y}$

Laplace and the birth rate of boys & girls

Posterior distribution:



Average θ = 0.490291
Standard deviation θ = 0.007117

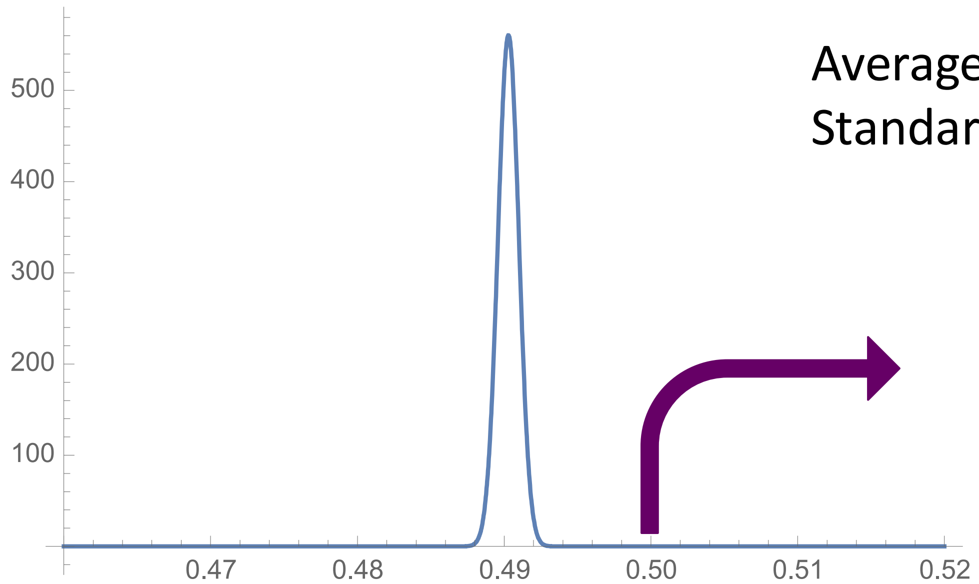
Uniform in interval [0,1]

- Bayes: $$\mathrm{p}(\theta|y) = \frac{p(y|\theta)\times p(\theta)}{p(y)}$$

Cst $\int_0^1 d\theta\ \theta^y\,(1-\theta)^{M-y}$

Laplace and the birth rate of boys & girls

Posterior distribution:



Average θ = 0.490291
Standard deviation θ = 0.007117

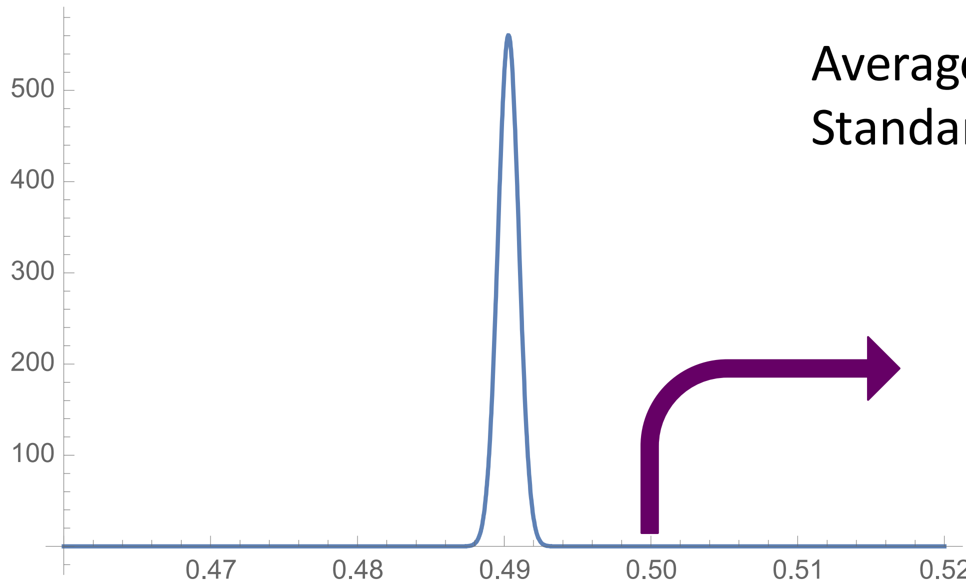
Probability that θ is equal or larger than 0.5 =

$$\int_{0.5}^{1} d\theta \ p(\theta|y) \approx 10^{-42}$$

Extremely unlikely!

Laplace and the birth rate of boys & girls

Posterior distribution:



Average θ = 0.490291
Standard deviation θ = 0.007117

Probability that θ exceeds 0.5 = $\int_{0.5}^{1} d\theta \;\; p(\theta|y) \;\approx 10^{-42}$ Extremely unlikely!

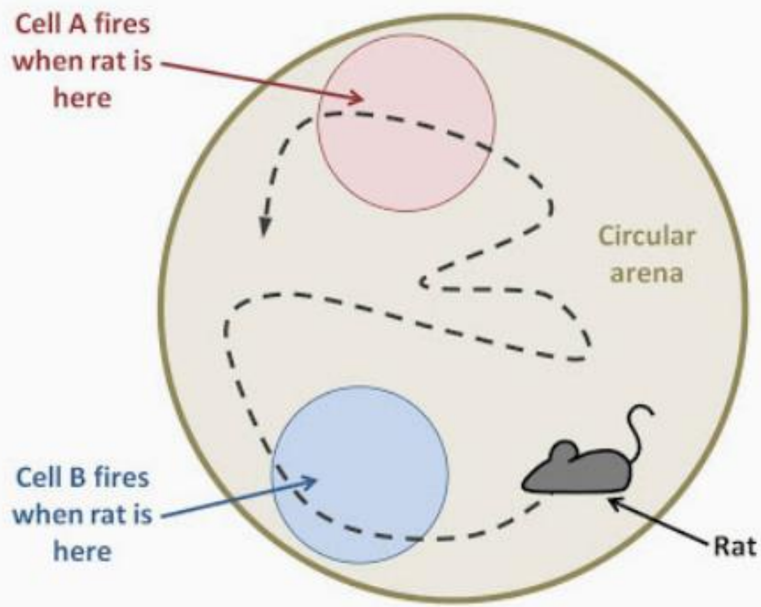
- In the tutorial tomorrow you will compute the posterior distribution, the most likely value, and average value for the spiking rate of a neuron from a neural recording (Poissonian Distribution)

Application:

Decoding of position from recorded activity in the Hippocampus (HPC)

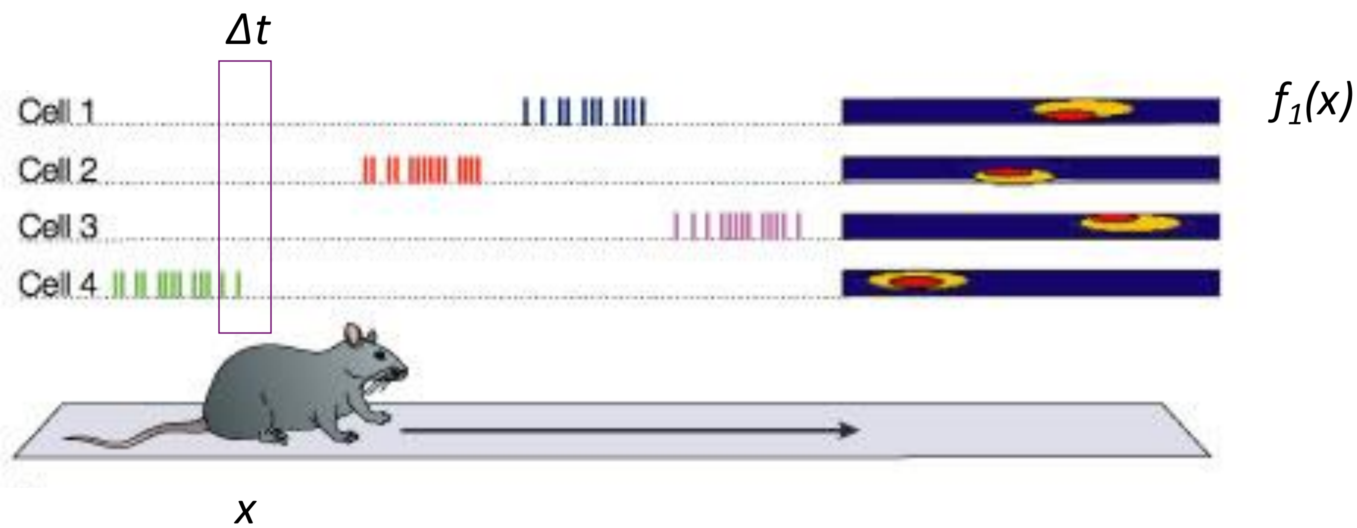
Decoding of position from HPC activity

Place cells in the hippocampus regions CA1 and CA3 present spatially-located firing fields.



- Place fields are retrieved when the animal is placed in the same environment after days
- Stable in dark and against limited changes of environment

Decoding of position from HPC activity



- Assume place cells are Poisson :

Probability of y spikes for cell i:

$$p_i(y_i|x) = e^{-f_i(x)\Delta t} \frac{(f_i(x)\ \Delta t)^{y_i}}{y_i!}$$

- First one infers the spiking rates of the cells at each position $f_i(x)$
- Likeliwood for the population activity:

Assume place cell activities are independent (conditional to position x):

$$\mathrm{P}(\{y_i\}|x) = \prod_i p_i(y_i|x)$$

Decoding of position from HPC activity

- Prior over trajectory: $P_{prior}(x_t) =$ uniform over the environment

- Find most likely position:

MAP=ML decoding $x_t = \text{argmax}_x \left[P(\{y_i\}|x) \times P_{prior}(x_t) \right]$

Decoding of position from HPC activity

- Prior over trajectory: $P_{prior}(x_t) =$ uniform over the environment

- Find most likely position:

MAP=ML decoding $x_t = \text{argmax}_x \left[P(\{y_i\}|x) \times P_{prior}(x_t) \right]$

Example: CA1 recording in freeling moving rats, 60 cm squared box, 34 cells
Place field: 400 squared bins per cell (discretized x – in 2D)
(data from K. Jezek)

Average error (Dt = 120 ms): 12 cm …

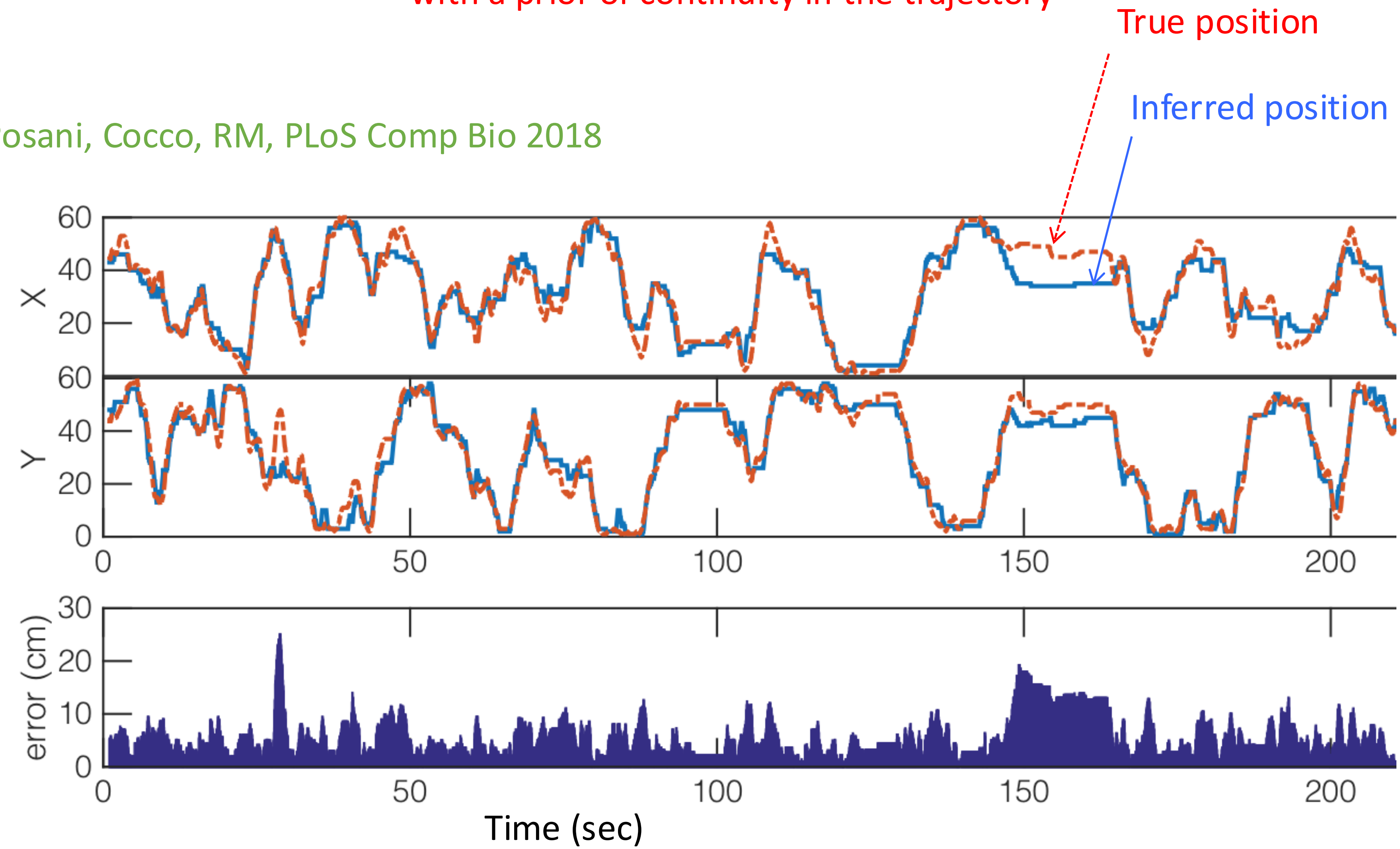
- ➢ Some regions in space are barely covered by place fields …
- ➢ Synthetic data study: error on positional decoding decreases as (nb. cells)$^{-1/2}$

How to improve precision?

Decoding of position from HPC activity

with a prior of continuity in the trajectory

Posani, Cocco, RM, PLoS Comp Bio 2018



Average error (Dt = 120 ms): 5.5 cm (compared to 12 cm)

Neuron

Hippocampal Reactivation of Random Trajectories Resembling Brownian Diffusion

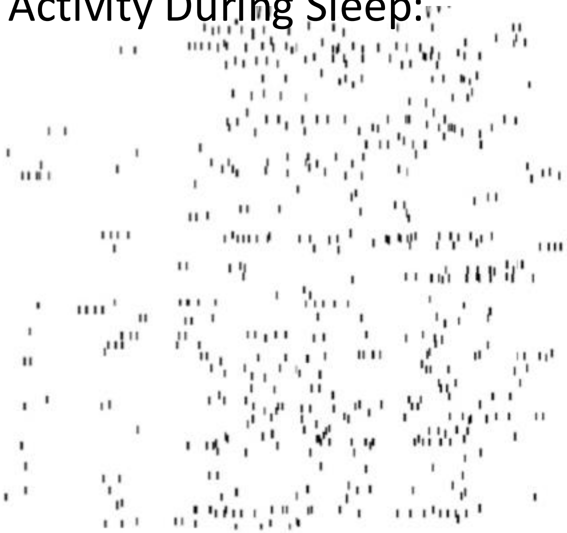
In Brief

Stella et al. examine the dynamical properties of reactivated spatial trajectories in the hippocampus following non-stereotypical exploration and find that reactivated trajectories are governed by a Brownian diffusion process that occur at varying lengths and timescales, without directly reflecting behavior.

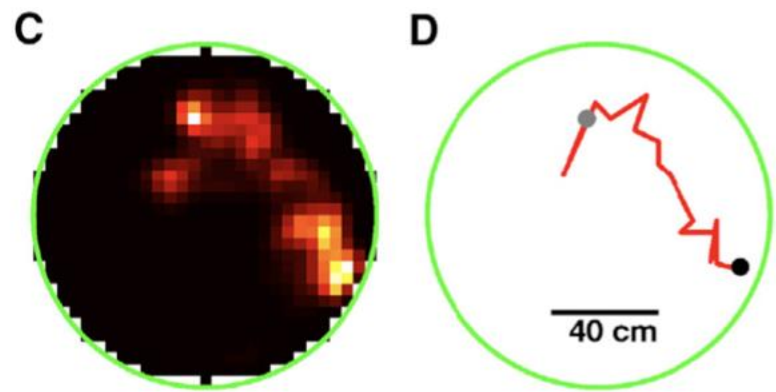
Authors

Federico Stella, Peter Baracskay, Joseph O'Neill, Jozsef Csicsvari

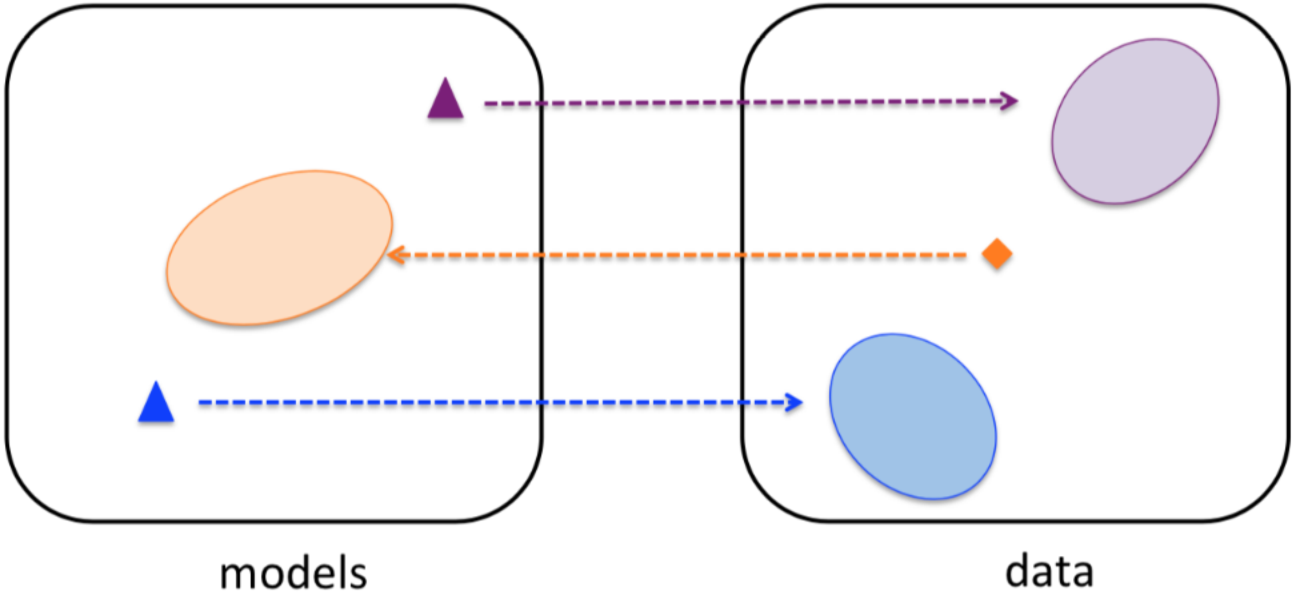
Activity During Sleep:



Decoding model form recordings during awake exploration



Take home Message

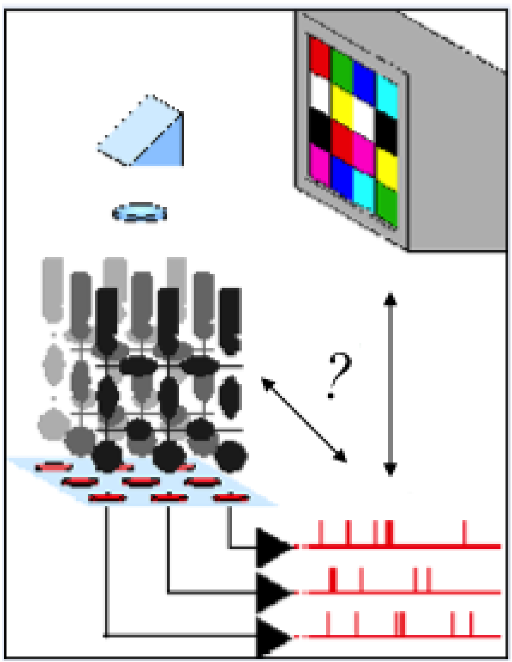


- Use Bayes law to derive the posterior distribution of parameter value given the observable.

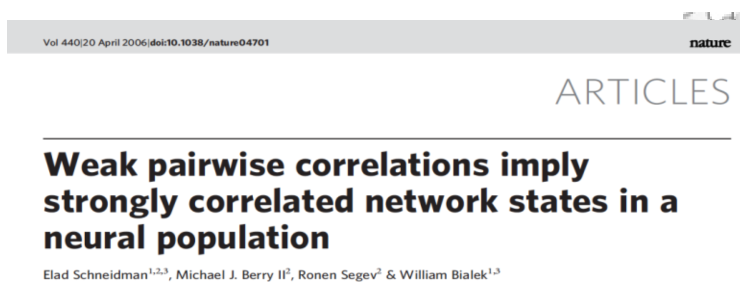
Key concepts: Posterior probability of a model given the data, Likelihood of data given a model, Prior probability of model parameters,

Tutorial: A Poisson process to model the spiking activity of a neuron and inference of the spiking frequency

In vitro recording on retina



tutorial:
40 ganglions, salamandra retina
natural movies. Recordings last for about 1 hour



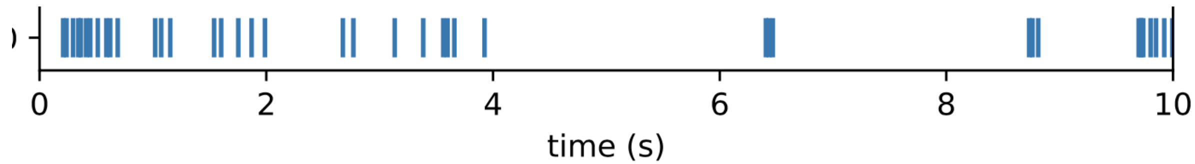
Vol 440|20 April 2006|doi:10.1038/nature04701 nature

ARTICLES

Weak pairwise correlations imply strongly correlated network states in a neural population

Elad Schneidman[1,2,3], Michael J. Berry II[2], Ronen Segev[2] & William Bialek[1,3]

Tutorial
Modeling Spike trains with a Poissonian process



$$p(y|\theta) = \frac{\theta^y e^{-\theta}}{y!}$$

Θ=f Δt

- Plot the Posterior distribution, Infer from a real spike train $\theta^* \ \langle\theta\rangle \ \sigma_\theta^2$
- For different recording times T .
- Is the Poisson process a good model for the data?

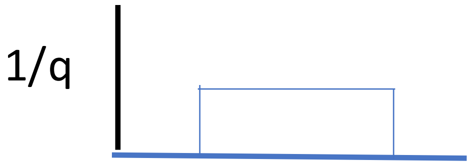
Before the tutorial of tomorrow:

- Install Jupyter Notebook (Anaconda) Phython 3
- Study slides of the class today, jupyther notebook of the class
- Book's Chapter 1

Supplementary Slides

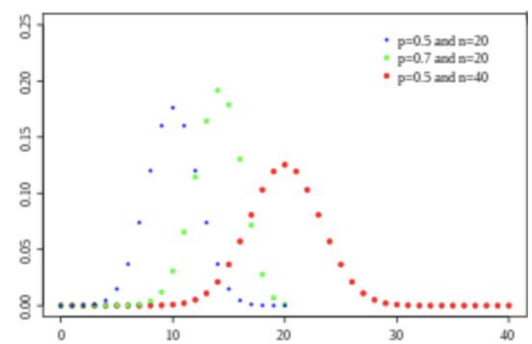
5 Important Probability Distributions For Random variables

- **Uniform Distribution**



$$y \in \{a_1, a_2, \cdots, a_q\} \qquad p(y) = 1/q$$

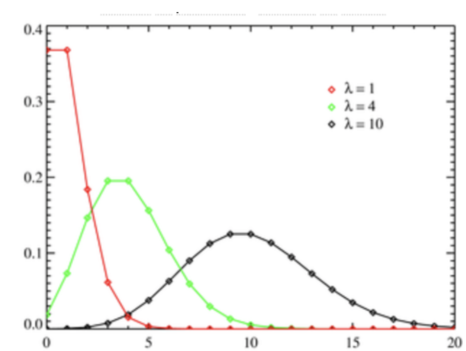
- **Binomialmial Distribution**



$$p(y|\theta) = \binom{M}{y} \theta^y (1-\theta)^{M-y}$$

$$y \in \{0,1,2,M\} \quad \in]\mathbb{N}$$

- **Poisson Distribution**

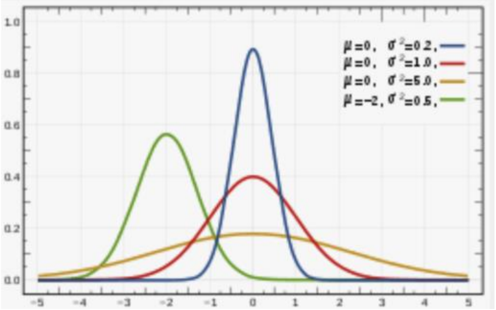


$$p(y|\lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \qquad y \in \{0, 1, 2, 3 \ldots \infty\} \in]\mathbb{N}$$

Note, the probability is always [0,1] and normalized to 1 !!

5 Important Probability Distributions For Random variables

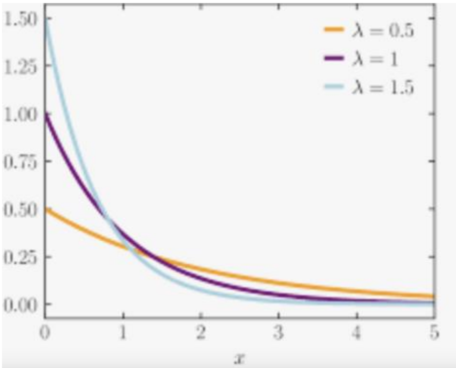
- **Normal/Gaussian Distribution**



$y \in \mathbb{R}$

$$p(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{(y-\mu)}{\sigma}\right)^2}$$

- **Exponential Distribution**



$y \in \mathbb{R}$, positives $\quad p(y|\lambda) = \lambda\, e^{-\lambda\, y}$

Note, the probability is always [0,1] and normalized to 1 !!

Bayes Theorem in Inference

Suppose we have an ensemble of **M** data points $\boldsymbol{y}_i \in \mathbb{R}^L$

Generated with a model with **D** unknown parameters $\boldsymbol{\theta} \in \mathbb{R}^D$

Bayes' rule:

Posterior Likelihood Prior

$$p(\boldsymbol{\theta}|Y) = \frac{p(Y|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(Y)} \ .$$

Evidence

The Evidence ensures the normalization of the Posterior

$$p(Y) = \int \mathrm{d}\boldsymbol{\theta}\, p(Y|\boldsymbol{\theta})p(\boldsymbol{\theta}) \ .$$

Technical Tricks:

Use the property of the Beta Distribution

$$\mathrm{Beta}(\theta;\alpha,\beta)=\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$$

$$B(\alpha,\beta)=\int_0^1 \mathrm{d}\theta\,\theta^{\alpha-1}(1-\theta)^{\beta-1}=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

In our case: $\alpha=y+1 \quad \beta=M-y+1,$

For α integer

$$\Gamma(\alpha)==\int_0^\infty \mathrm{d}\theta\,\theta^{\alpha-1}e^{-\theta}=(\alpha-1)!$$