# Lecture 3

Spike Stimulus Analysis

# About me

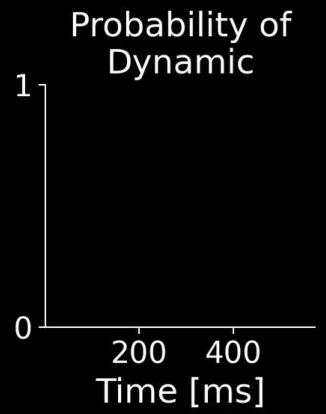
Computational Research Scientist in the lab of Dr. Loren Frank.

Postdoc with Dr. Uri Eden in the Department of Mathematics and Statistics at Boston University

Ph.D. in Computational Neuroscience from Boston University working with Drs. Daniel H. Bullock and Earl K. Miller.

Undergraduate degrees in Mathematics and Philosophy

# My Research


Animal's Position

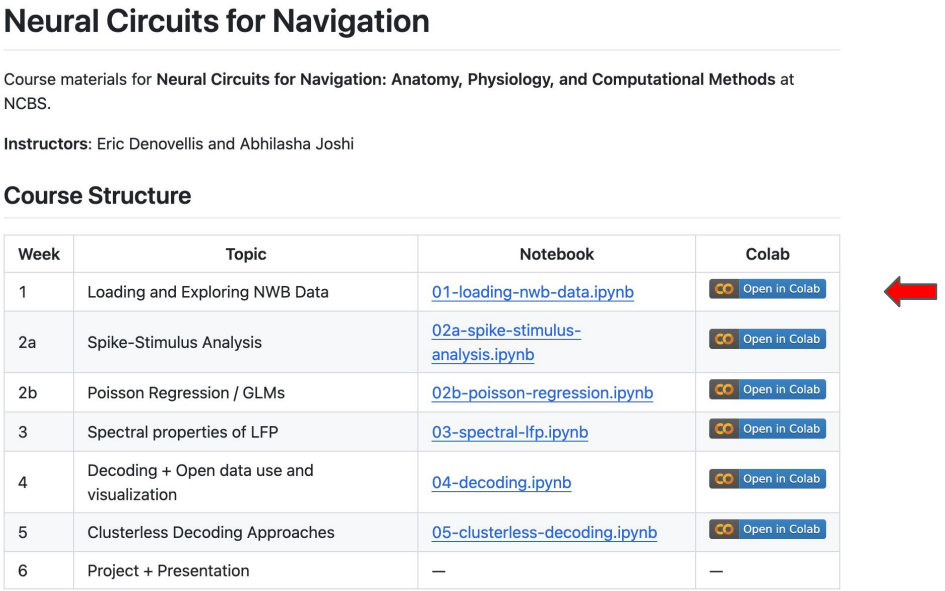
Decoded Position

Data from Alison Comrie

# Jupyter Notebooks and Google Colab

https://github.com/edeno/ncbs-neural-circuits-navigation

Click on the **Open in Colab** button



## Neural Circuits for Navigation

Course materials for **Neural Circuits for Navigation: Anatomy, Physiology, and Computational Methods** at NCBS.

**Instructors**: Eric Denovellis and Abhilasha Joshi

### Course Structure

| Week | Topic | Notebook | Colab |
|---|---|---|---|
| 1 | Loading and Exploring NWB Data | 01-loading-nwb-data.ipynb | Open in Colab |
| 2a | Spike-Stimulus Analysis | 02a-spike-stimulus-analysis.ipynb | Open in Colab |
| 2b | Poisson Regression / GLMs | 02b-poisson-regression.ipynb | Open in Colab |
| 3 | Spectral properties of LFP | 03-spectral-lfp.ipynb | Open in Colab |
| 4 | Decoding + Open data use and visualization | 04-decoding.ipynb | Open in Colab |
| 5 | Clusterless Decoding Approaches | 05-clusterless-decoding.ipynb | Open in Colab |
| 6 | Project + Presentation | — | — |

# Overall: How do we understand and characterize brain data and behavior with statistics?

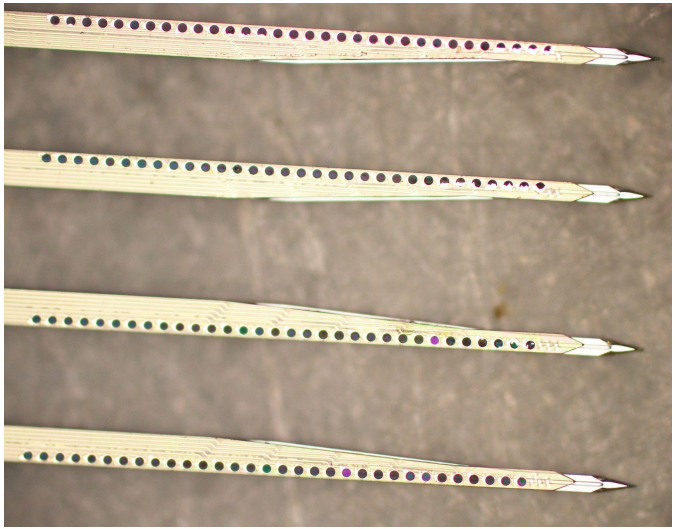
Next four weeks:

1. Relating spikes to stimuli (Encoding)
2. Basic signal processing
3. Relating Local Field Potentials (LFPs) to stimuli and analyzing oscillations
4. Decoding brain states from spikes

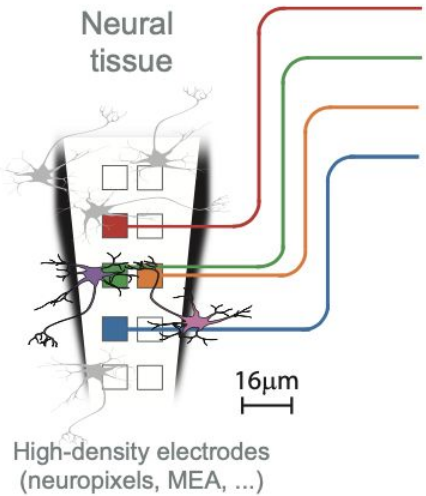
Focus on extracellular recordings (but of course there are many types of data including calcium imaging, functional neuroimaging (fMRI), EEG, etc)

# Overview: what do we record when we record extracellularly in the brain?

Stick electrode in brain



4 shank, 32 channel polymer probes

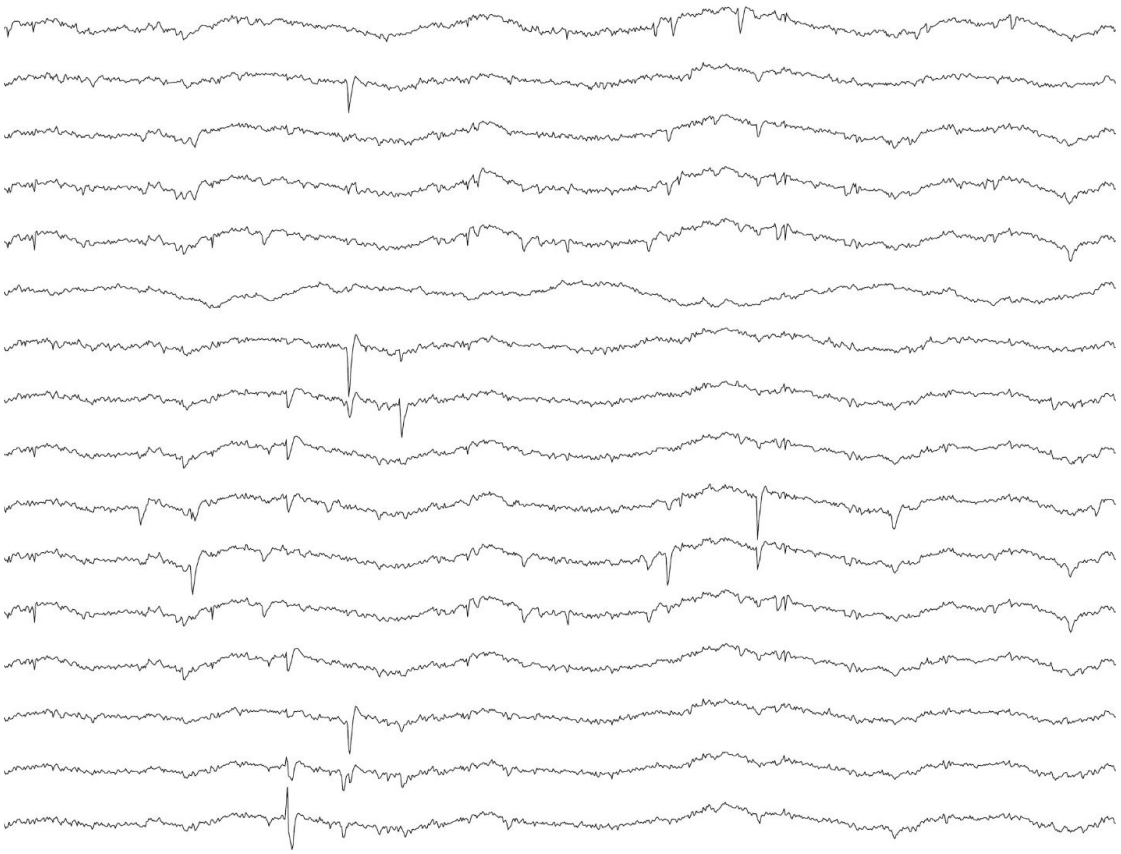



100s of neurons

Einevoll et al. 2012;
Chung et a. 2019

# Extracellular signals

16 channel trace from polymer probes

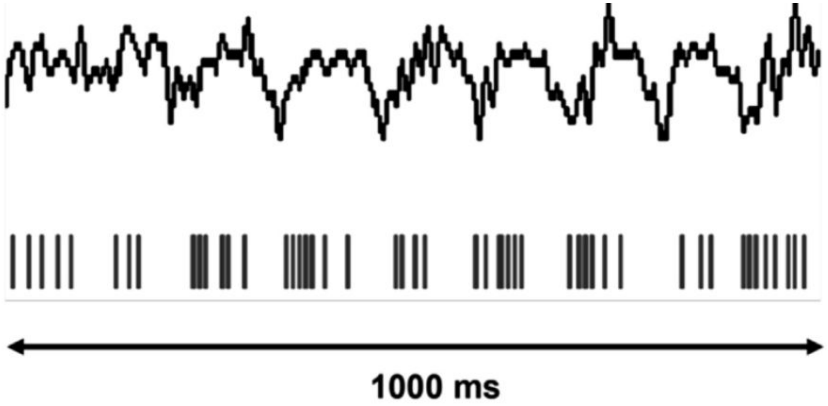


# Two Types of Data: Spikes and Local Field Potentials

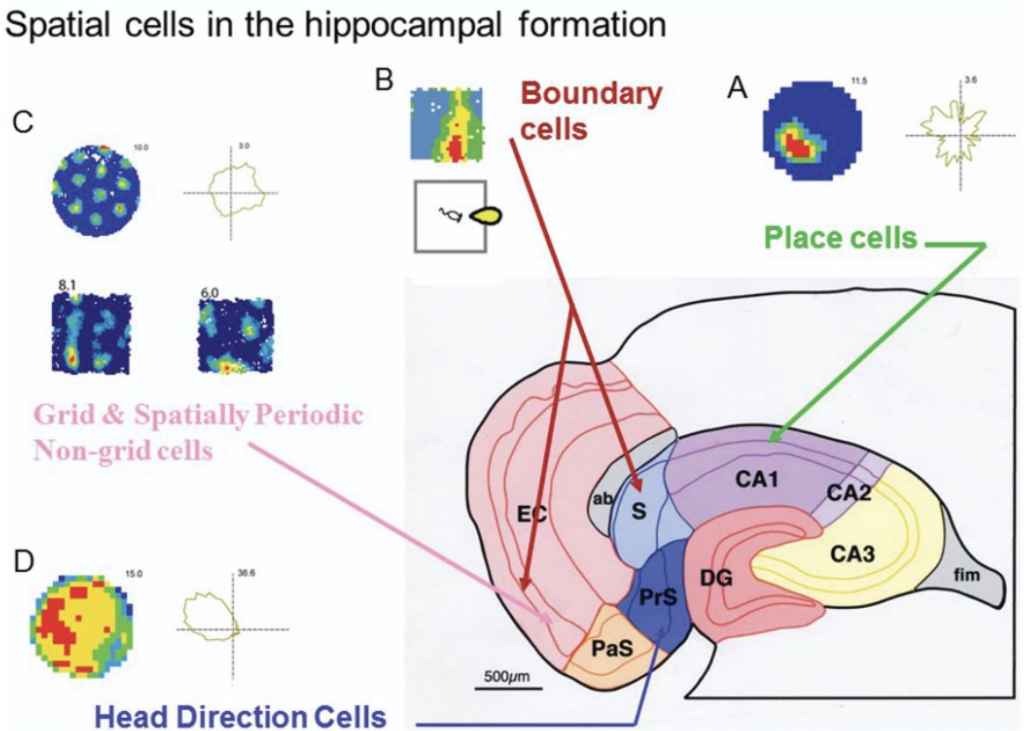
Two different types of data:

- Spikes (list of times, point process)
- Local Field Potentials (continuous)

We are going to first focus on spikes (LFPs next week)

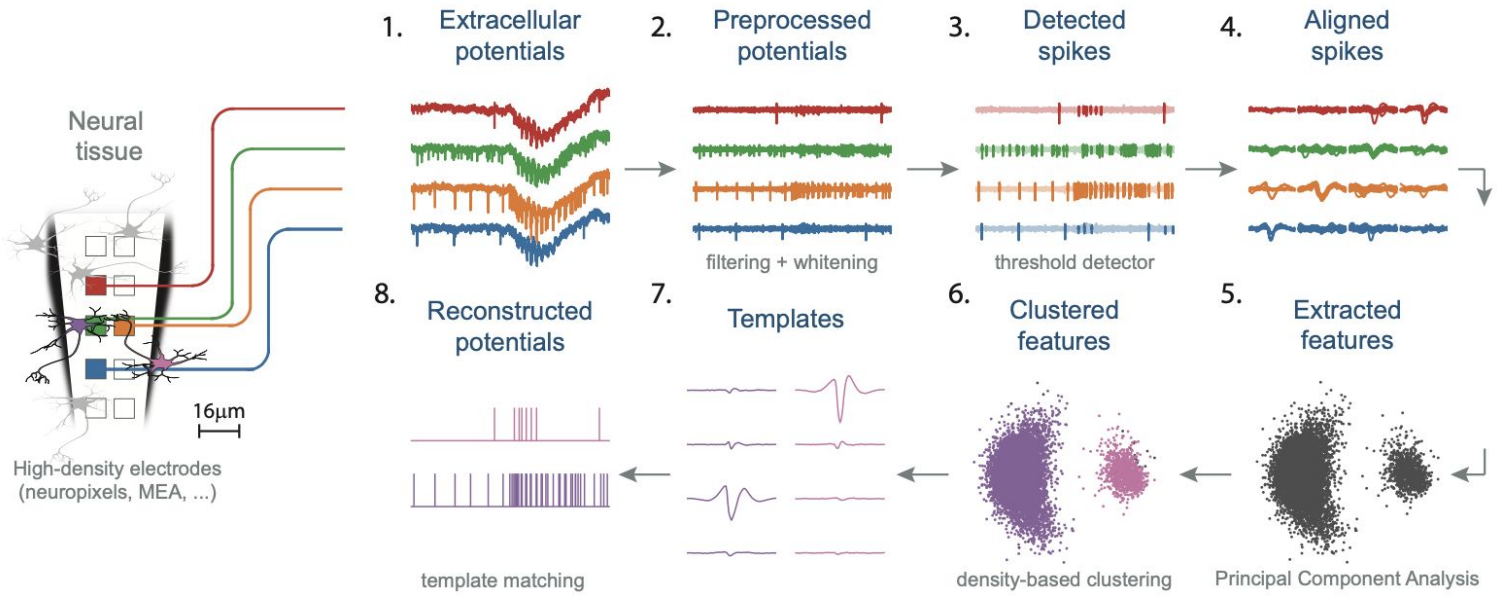


# Motivation: We want to relate this to some representation of the external world




O'Keefe Nobel lecture

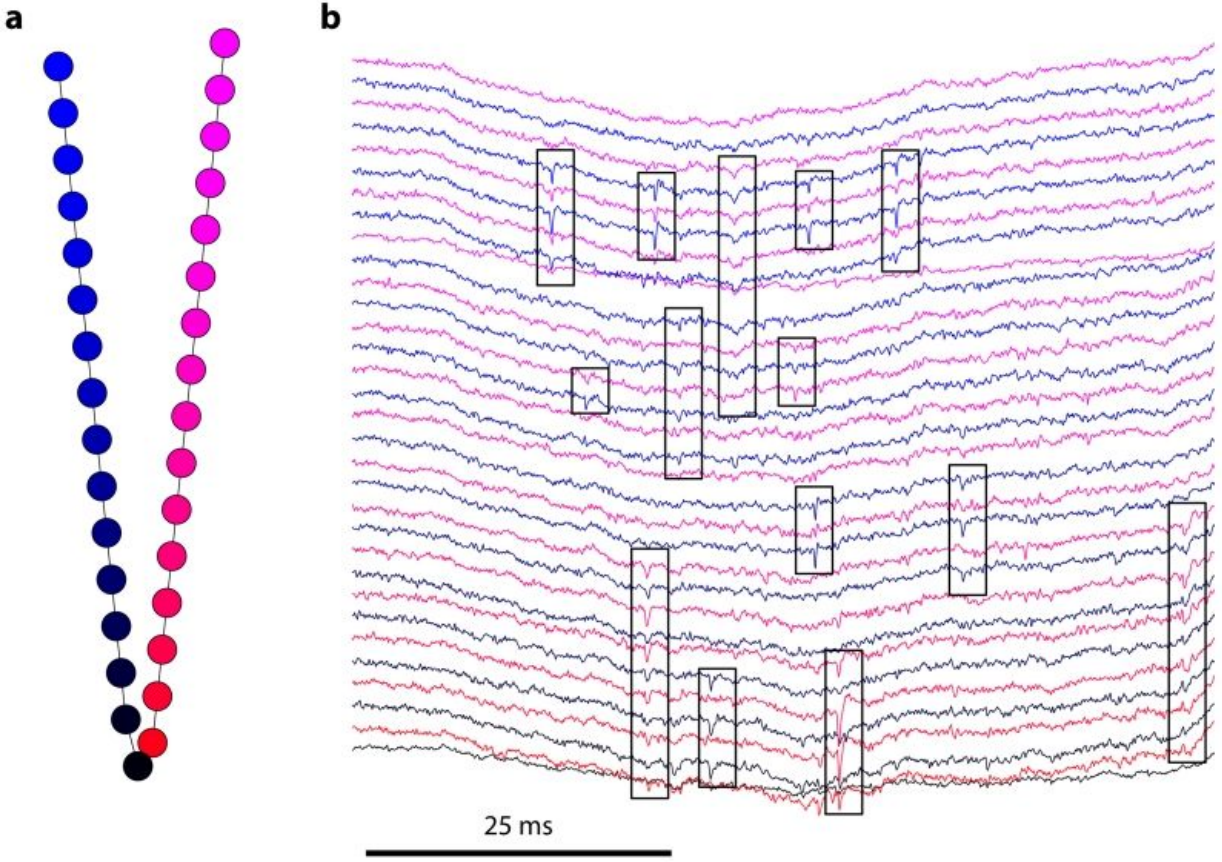
# Preprocessing: to get spikes, we must spike sort




*The spike sorting problem: 1) identify the spikes in the multi-channel voltage recording and 2) assign those spikes to individual neurons*

Einevoll et al. 2012

# Spike sorting is non-trivial

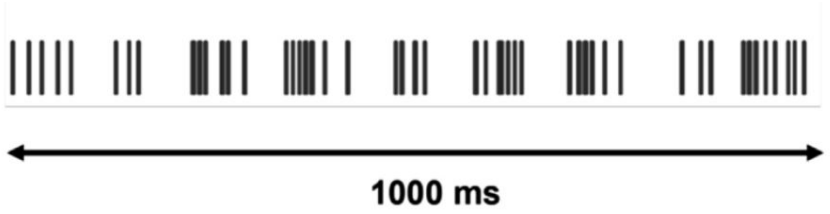


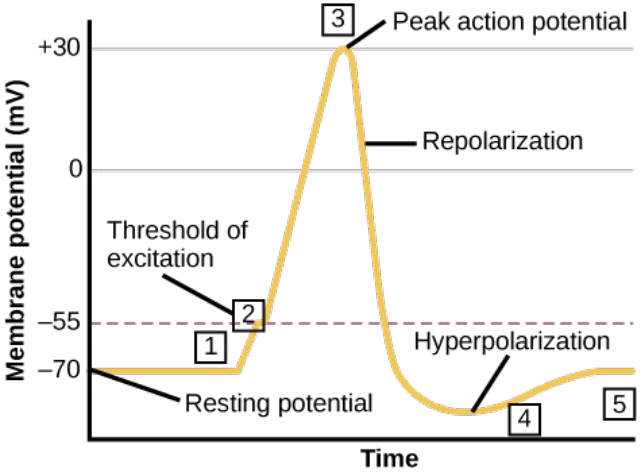

- Temporally overlapping spikes on contacts
- Simultaneous spiking from multiple neurons
- Accounting for spatio-temporal correlation given probe geometry.
- Autoregressive process (bursts)
- Electrode drift

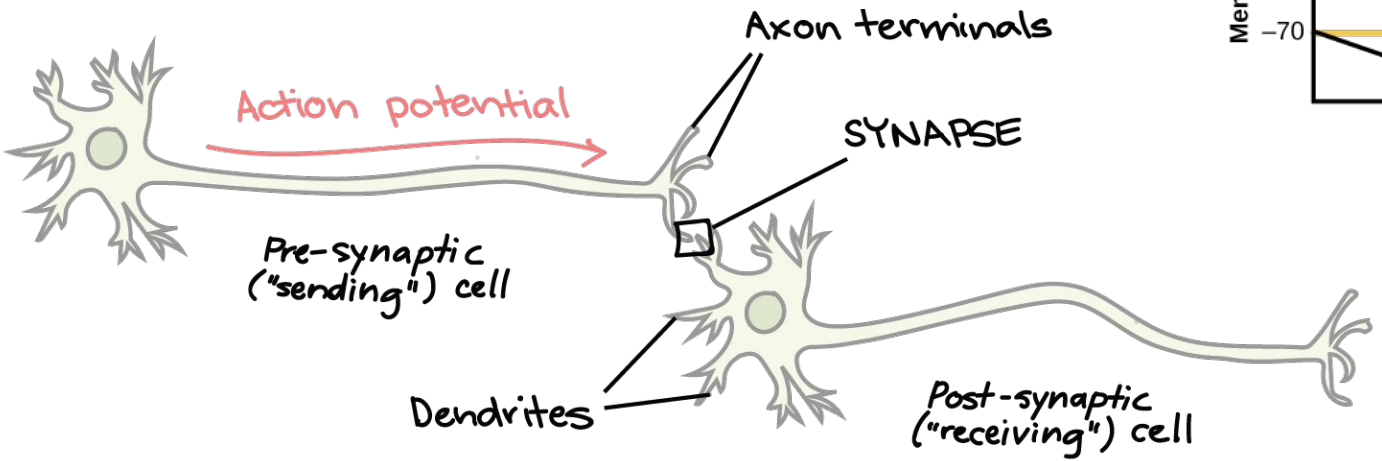
Rossant…Harris (2016) Nat. Neuro

# What does a spike represent?

Spikes have a shape, we are ignoring that and only looking at the spike time



# Summary: a spike is not a spike is not a spike

Spike times and neuron identities derived from extracellular data are created through **non-trivial preprocessing**.

We make necessary and useful simplifications but with that comes uncertainty.

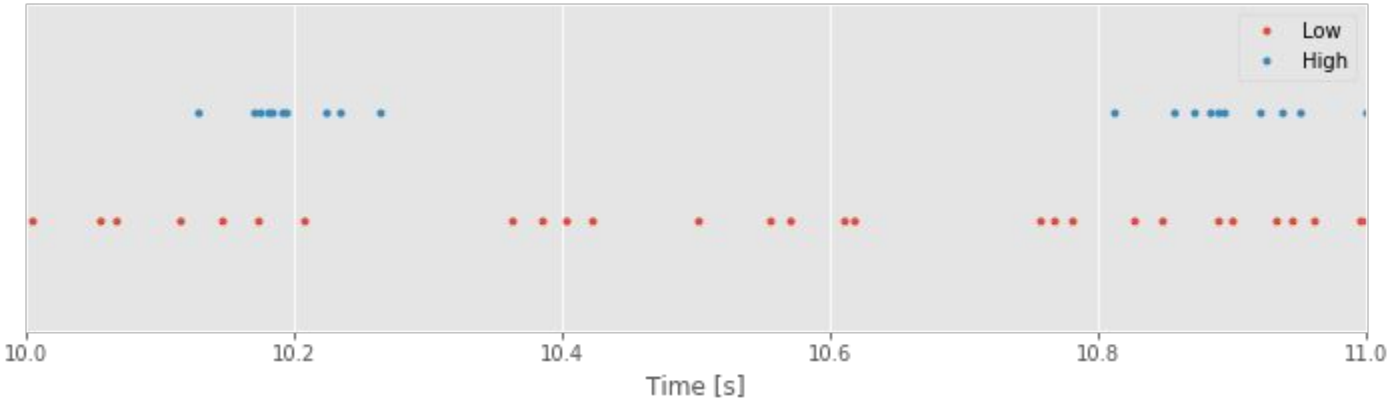
This should not stop us from analyzing spikes, but **it is important to understand that there is uncertainty and be careful with what we are analyzing**.

# What makes a neuron(s) spike?

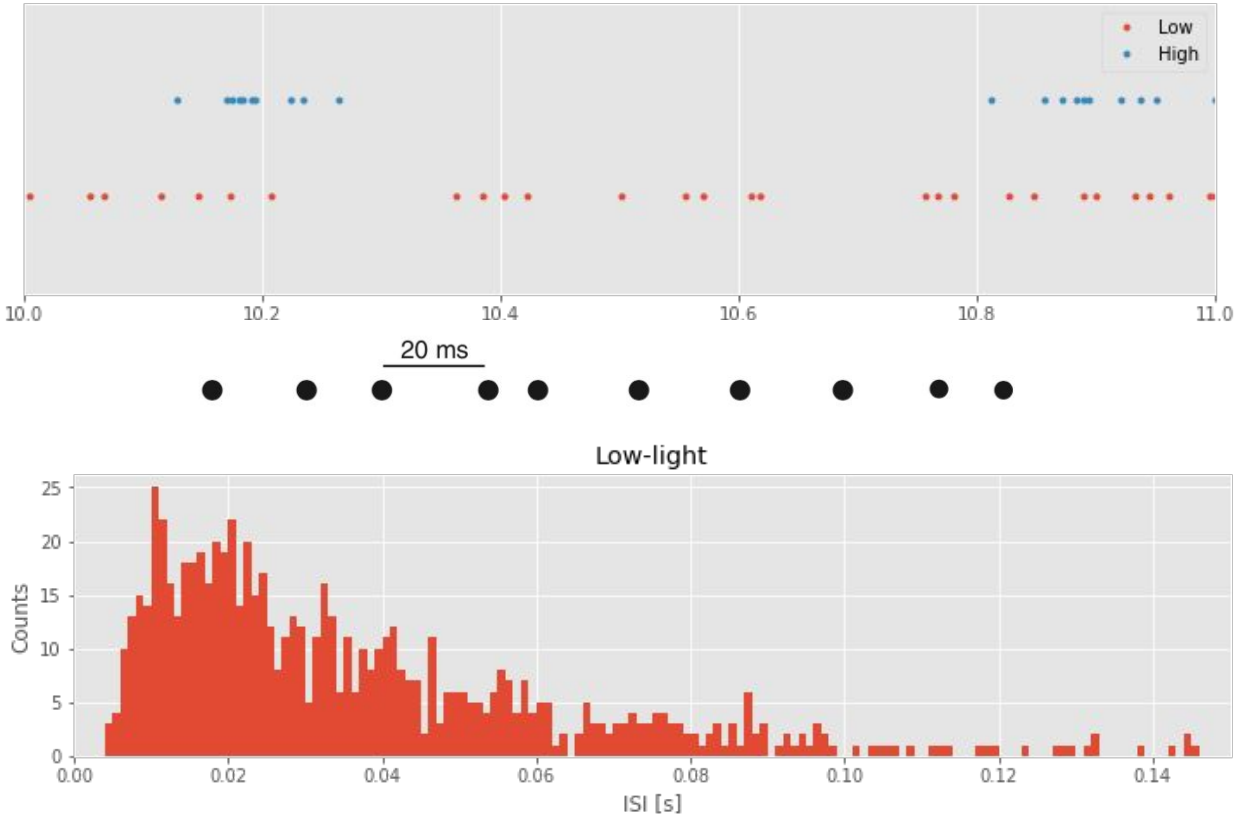
List of spike times: [0.1, 0.8, 0.9]

First step: Basic descriptive statistics and visualization

Raster plot

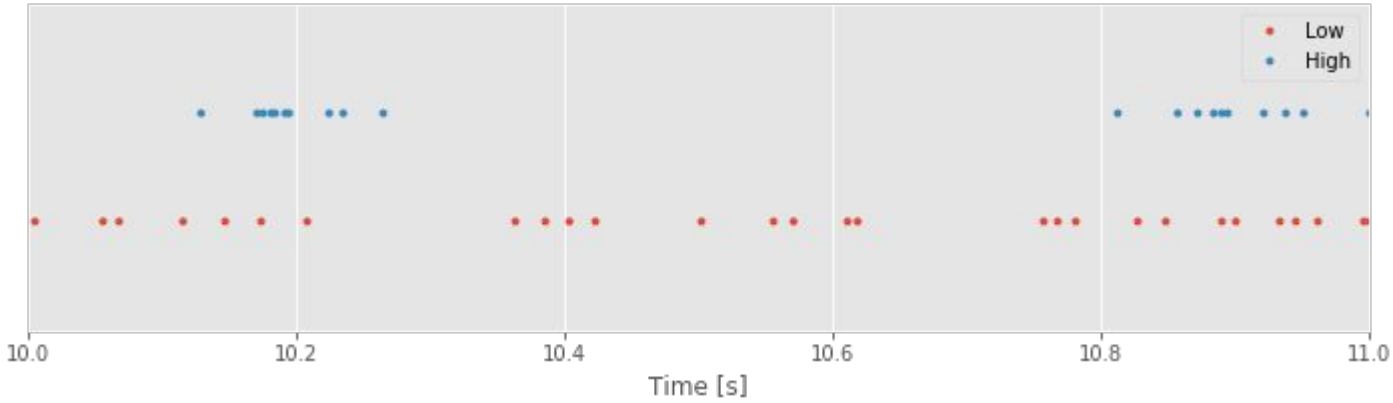


# Interspike intervals (ISIs)




$$X_i = S_i - S_{i-1}$$

# Counting spikes




Why is this a good summary of what is happening with a neuron under different conditions?

# Spike Rate

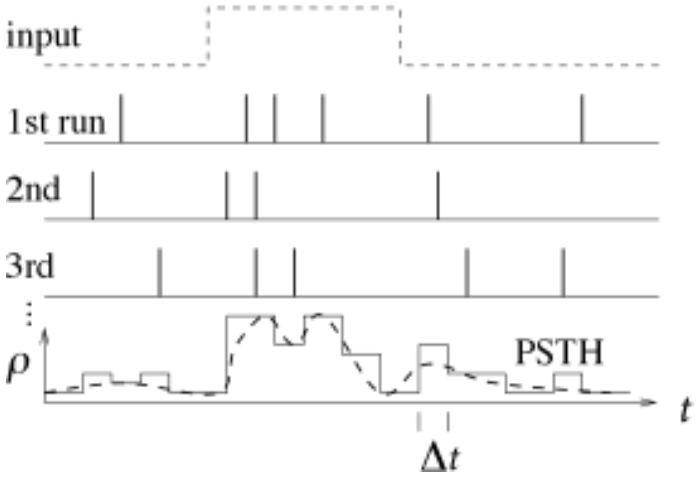
Provides a way to normalize different conditions (different length trials, different bin sizes)

$$FR = \frac{\text{number of spikes}}{\Delta t}.$$

# Peri-event time histograms

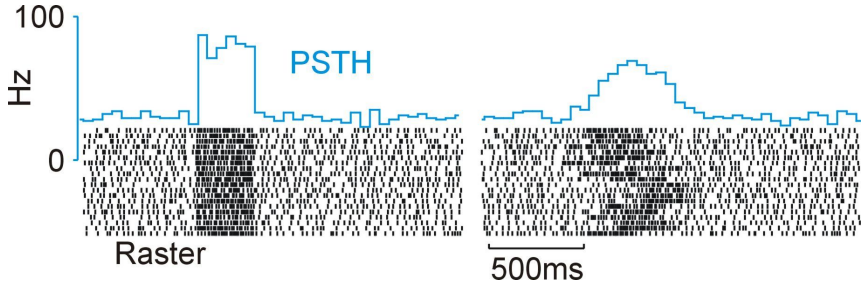
rate = average over several runs

(single neuron, repeated runs)




spike density in PSTH

$$\rho = \frac{1}{\Delta t}\frac{1}{K} n_K(t; t+\Delta t)$$



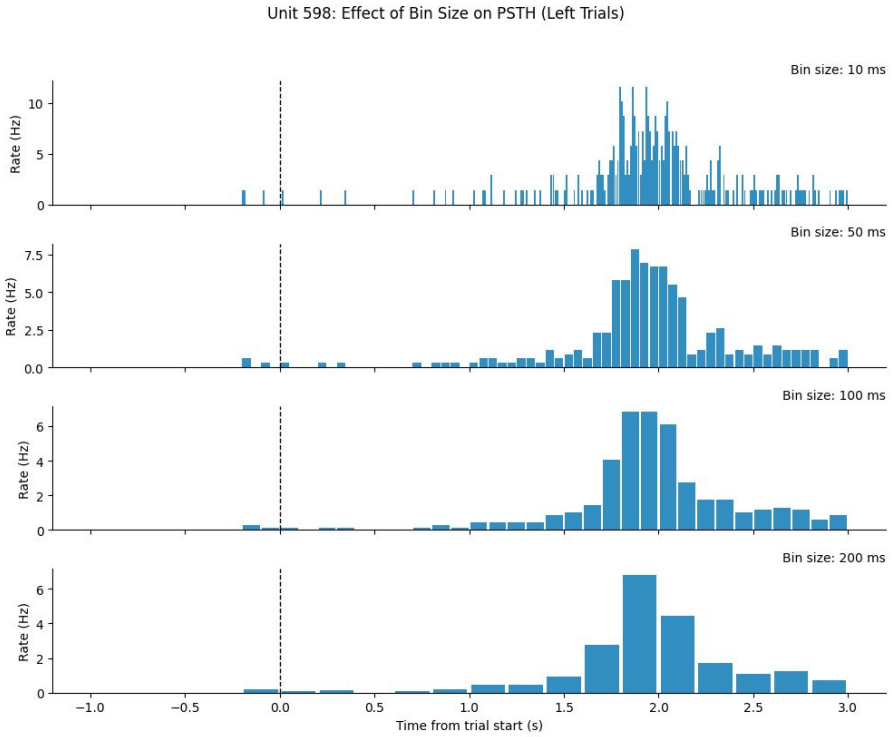

https://www.staff.ncl.ac.uk/demetris.soteropoulos/neural_variability.html

# Bin size matters

It’s a choice about how fast you believe the rate is evolving (smaller bin, evolving faster, larger bin, evolving slower).

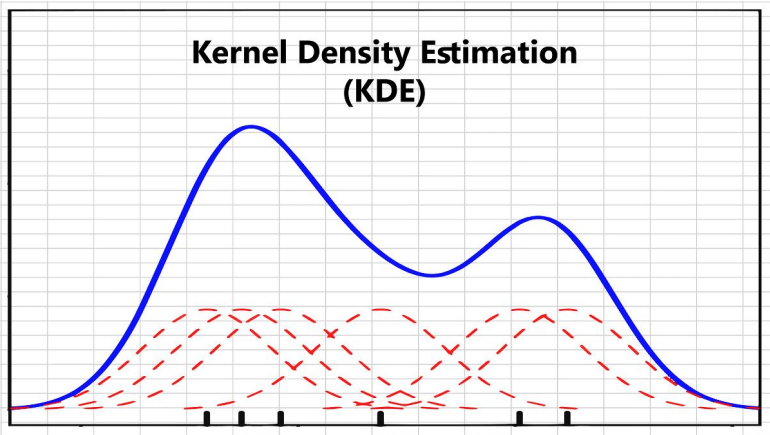
In a histogram, the prediction is a constant rate within the bin.

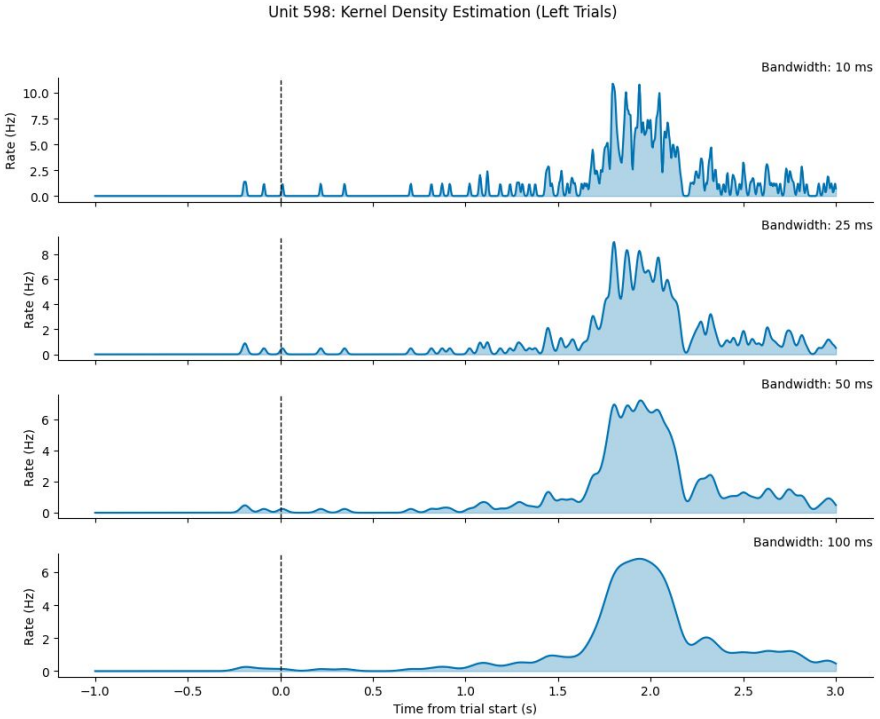
Bias-variance tradeoff



# Kernel Density Estimation

There are other “binless” methods such as kernel density estimation which give you smoother rates



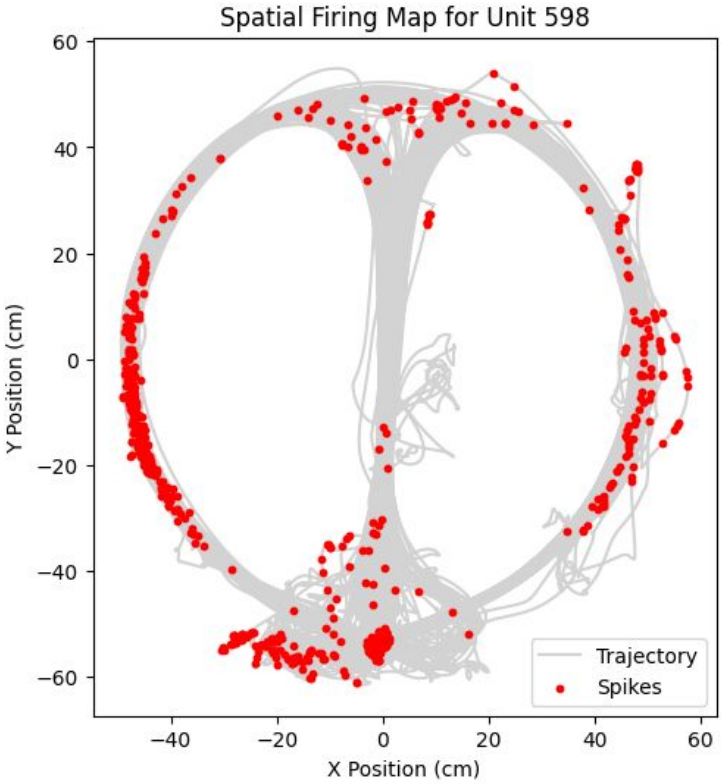



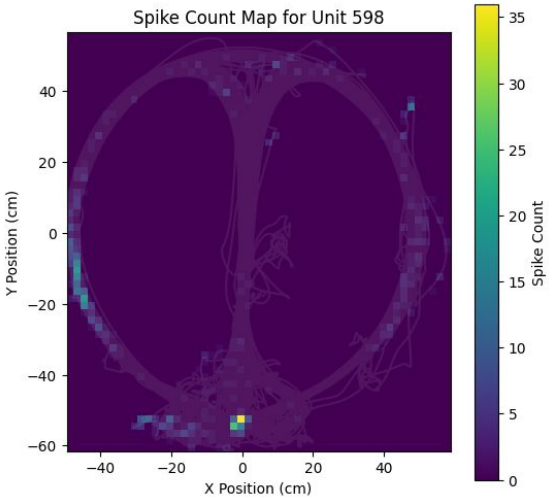
# Rate maps (e.g. place fields)

We could take a histogram over position, but this would be just counts of spikes per position. The animal could be in one position longer.

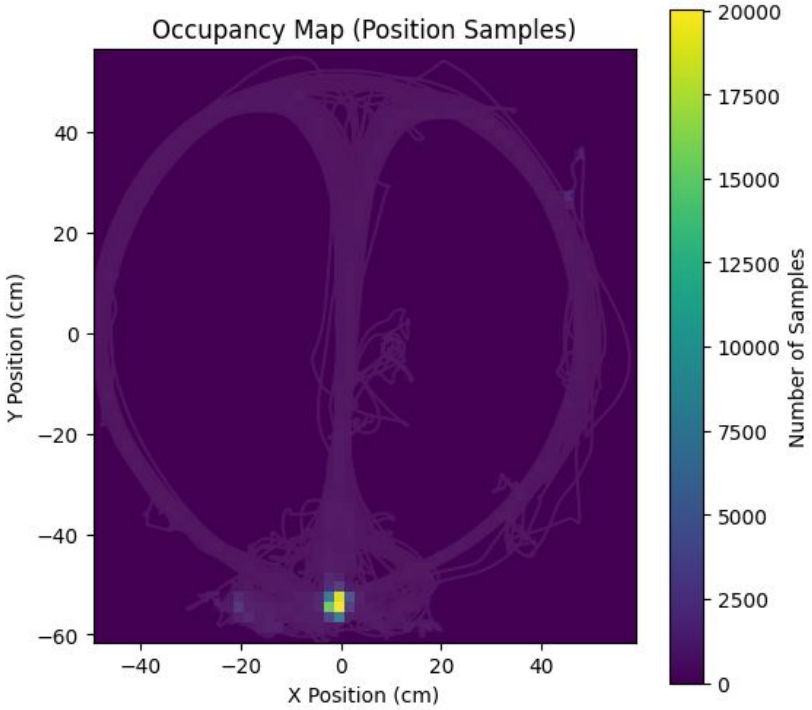
We want spike rate (spikes per second for each position bin)

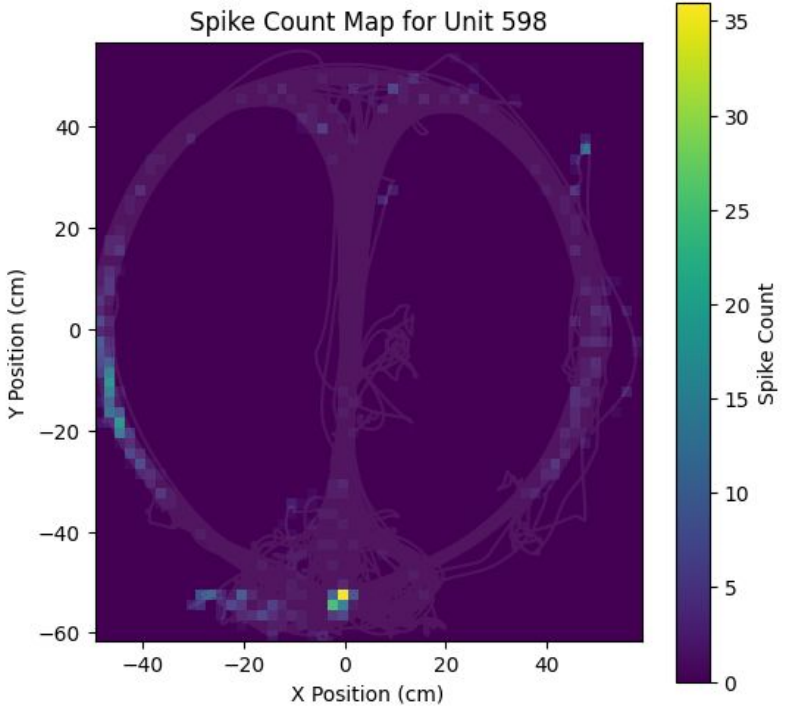
How do we do this?



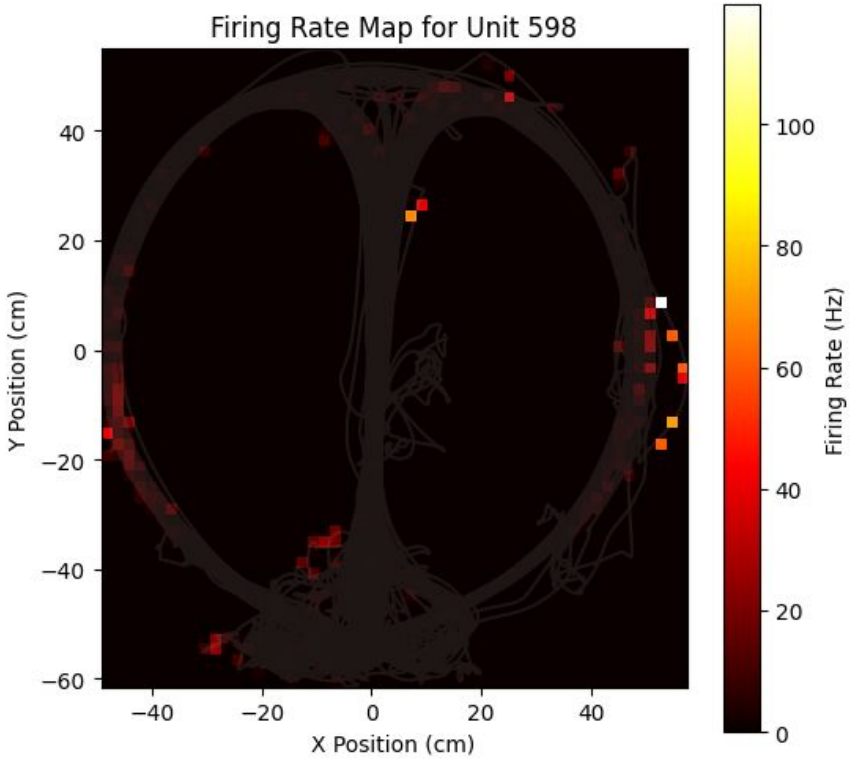
# Occupancy - How often is the animal at a given position



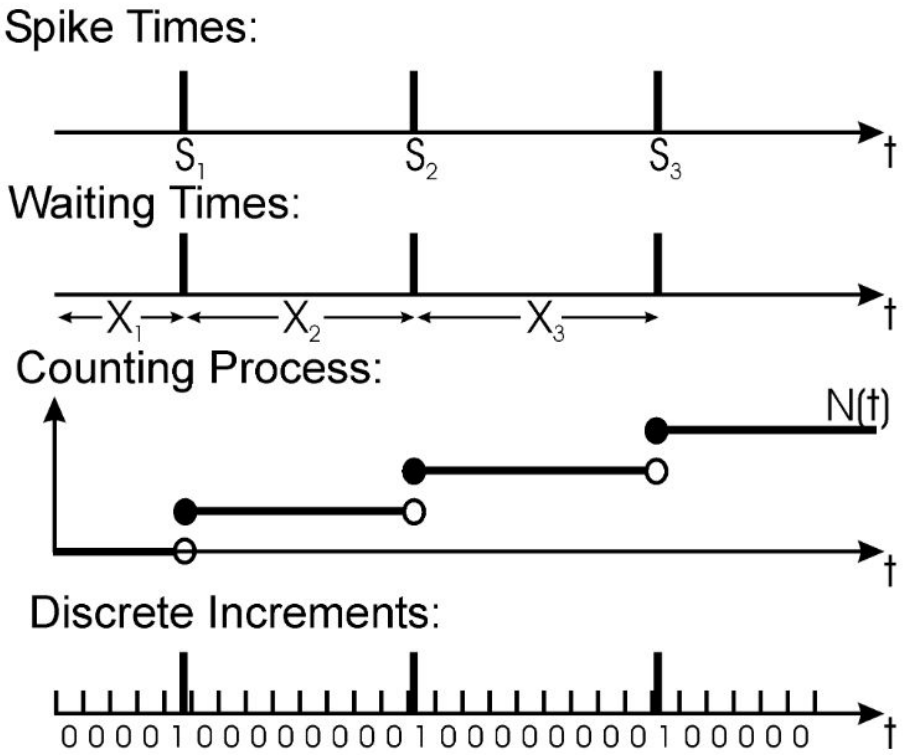



# Rate map - occupancy normalized histogram

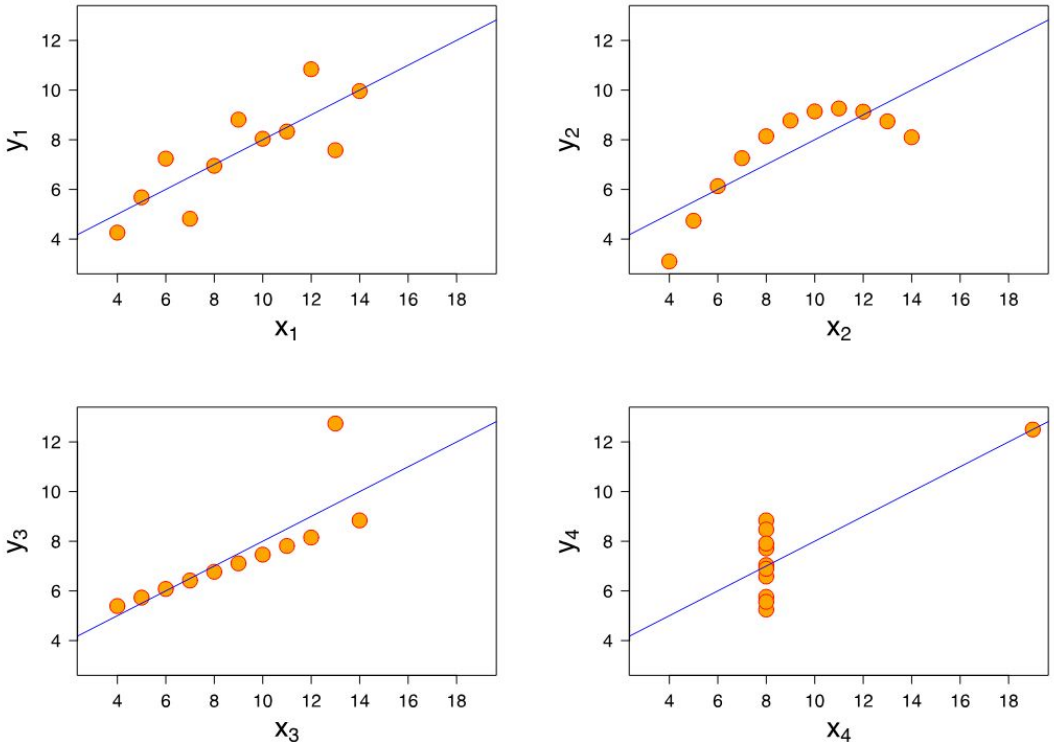
Normalize counts per position bin by how often the animal is there.



# Summary: four ways to describe a point process



# Importance of basic visualizations - Anscombe's quartet




all four of these have the same mean, variance, correlation, and regression line

So you need to plot and understand your model

# Characterizing variability spike rate

The number of spikes over a period of time (spike rate) exhibits high variability

An individual neuron presented multiple times with the same stimulus will typically produce a different number of spikes following each presentation.

But over many stimulus presentations, the ratio between the sample mean and sample variance of the spike counts over fixed intervals remains constant.

# Fano Factor

The ratio between the sample mean and sample variance of the spike counts

$$F_n = \frac{\mathbb{V}[y_{t,n}]}{\mathbb{E}[y_{t,n}]}$$

Dimensionless - easy to compare between in different conditions (e.g. different time windows, different stimulus intensities) and different rates (variability can be higher at higher rates)

Overdispersion: Fano Factor > 1

Underdispersion: Fano Factor < 1

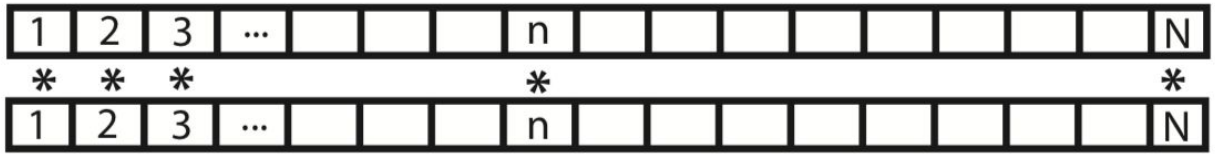
Poisson: Fano Factor = 1

Autocorrelation: how is the current spiking related to past spiking

$$\rho_{xx}[L] = \frac{\sum_{i=1}^{N-L}(x_i - \bar{x})(x_{i+L} - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

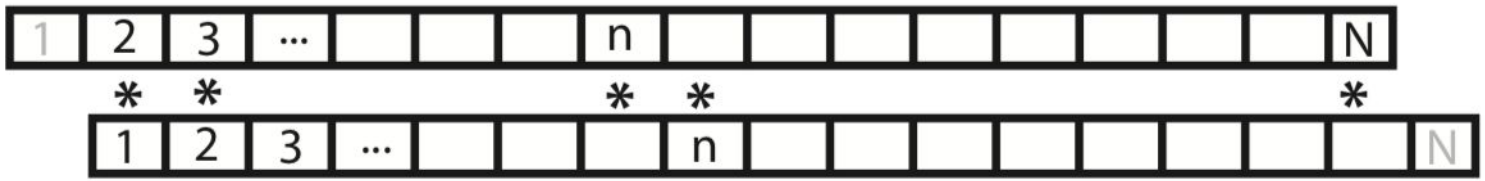
$$\rho_{xx}[L] = \frac{\sum_{i=1}^{N-L}(x_i - \bar{x})(x_{i+L} - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

L=0




sum $r_{xx}[0]$

L=1




sum $r_{xx}[1]$

# slide and inner product

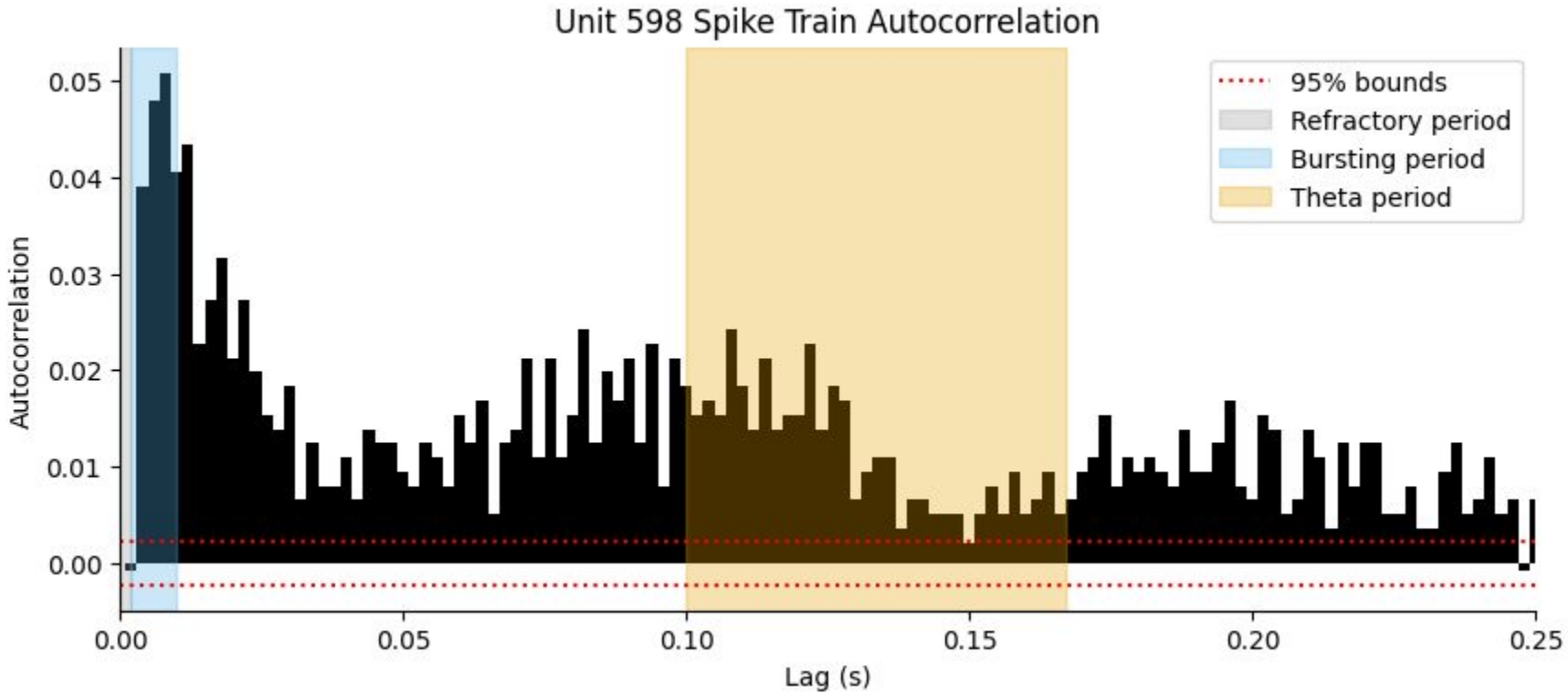
inner product (multiply & sum) measures similarity (how orthogonal vectors are, inner product =0 -> vectors orthogonal, inner product = 1 -> vectors parallel)

We measure similarity at each lag. We look at the original time series versus a lagged copy (the same time timeseries shifted backwards L time steps)

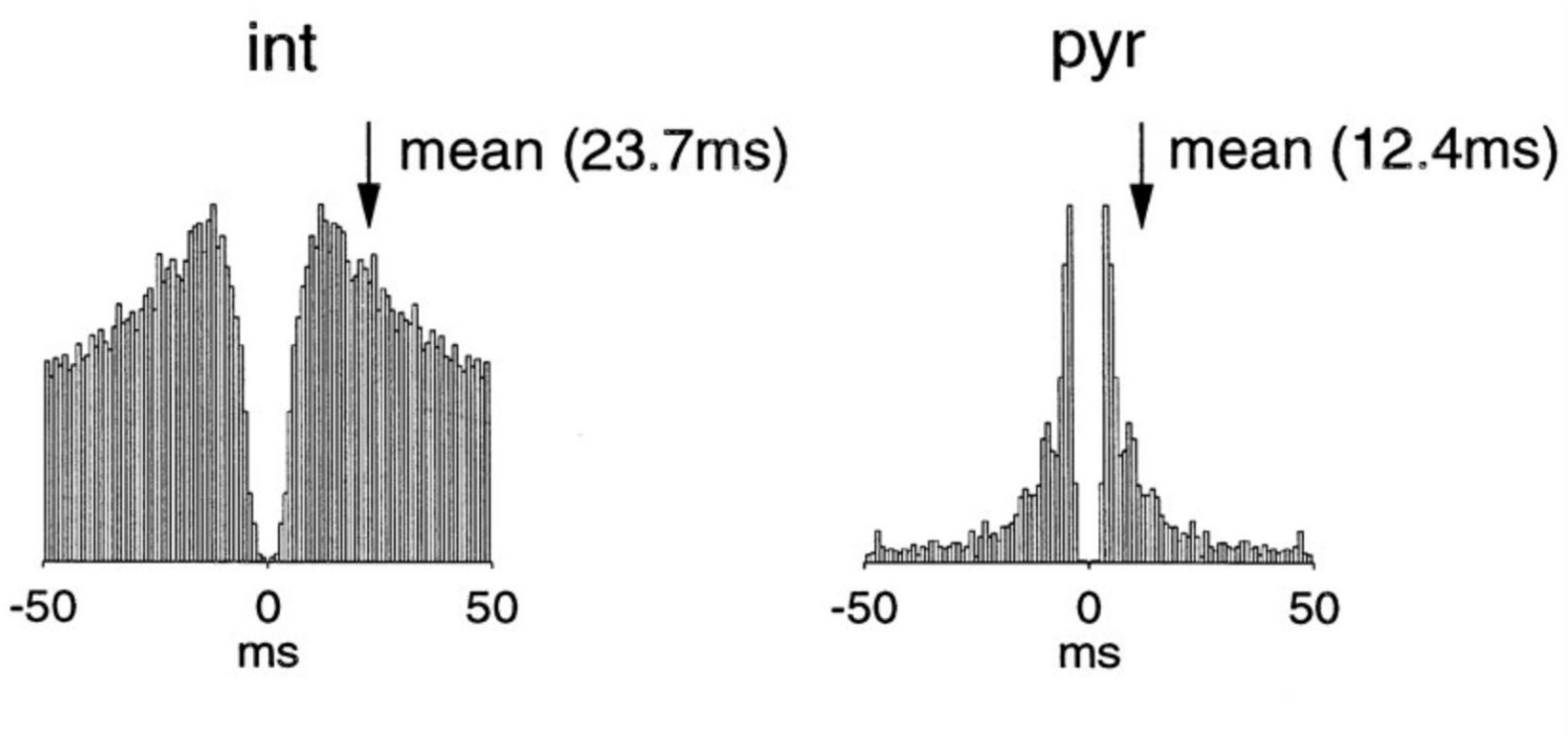
Remember that we have to subtract the mean (we are looking at residuals, similarity of fluctuations)

$$\rho_{xx}[L] = \frac{\sum_{i=1}^{N-L}(x_i - \bar{x})(x_{i+L} - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

# Hippocampal example



# Hippocampal example




Csicsvari et al. 1998

# Poisson model: Statistical Model of a Neuron

Simple model (not modeling complex biophysics like Hodgkin-Huxley)

Well suited for modeling expected number of counts (number of spikes is are discrete non-negative integers 0,1,2…)

Neuron firing is stochastic (spike times are variable across conditions)

**The goal is to find a rate parameter that explains the expected number of spikes.**

# Poisson distribution

**The goal is to find a rate parameter that explains the expected number of spikes.**

Probability of k spikes for a given time bin:

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

lambda = rate (expected number of events)

P(k) = probability of k spikes

**What is the probability of 0 spikes?**

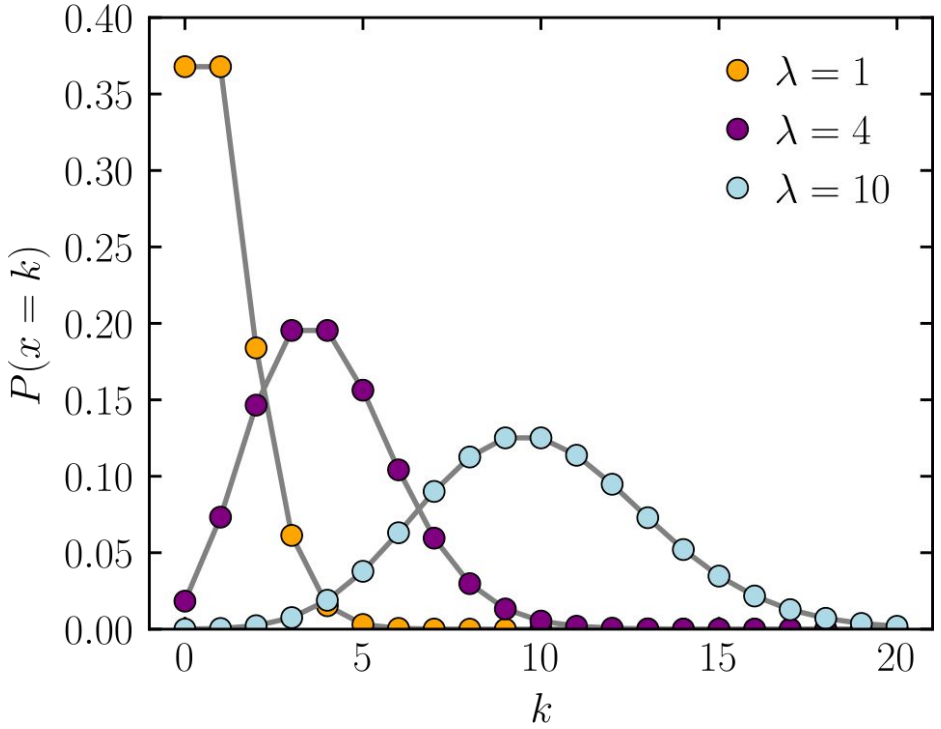
# Poisson

Low rate (lambda = 1): most probable one spike

Medium rate (lambda = 4):

most probable 4 spikes but long tail

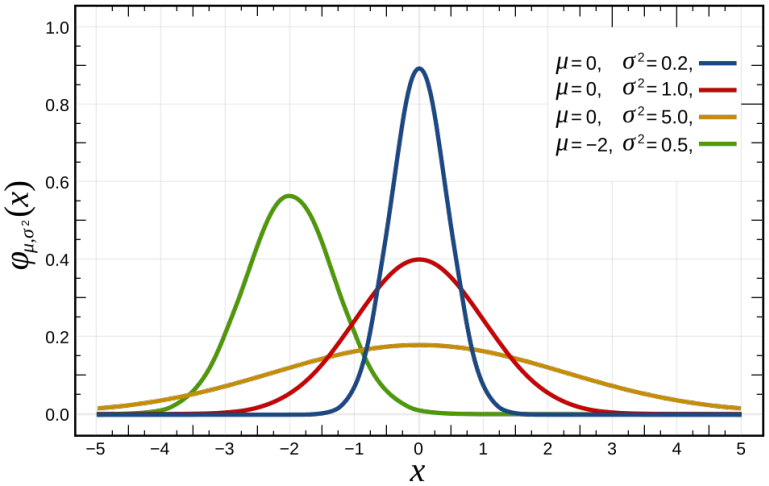
High rate (lambda = 10): Gaussian-like



# Why this is better than using a Gaussian

With a Gaussian, you can go below zero but spike rate is non-negative

With a Gaussian, mean and variance are linked: higher rate → higher variability. Poisson captures this with one parameter; Gaussian can't.

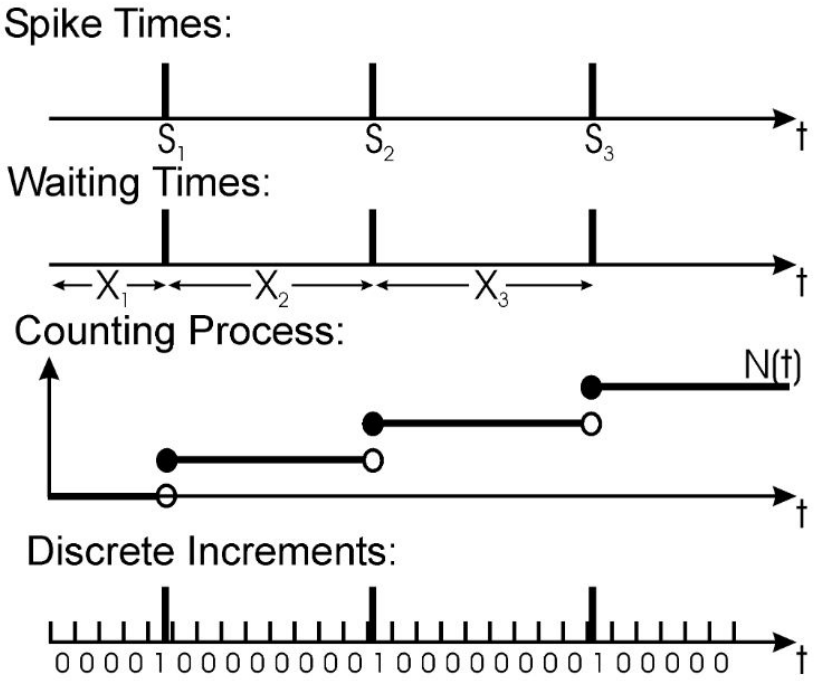


# Poisson

Number of spikes in any time bin is independent of all previous (and future) spiking (memoryless).

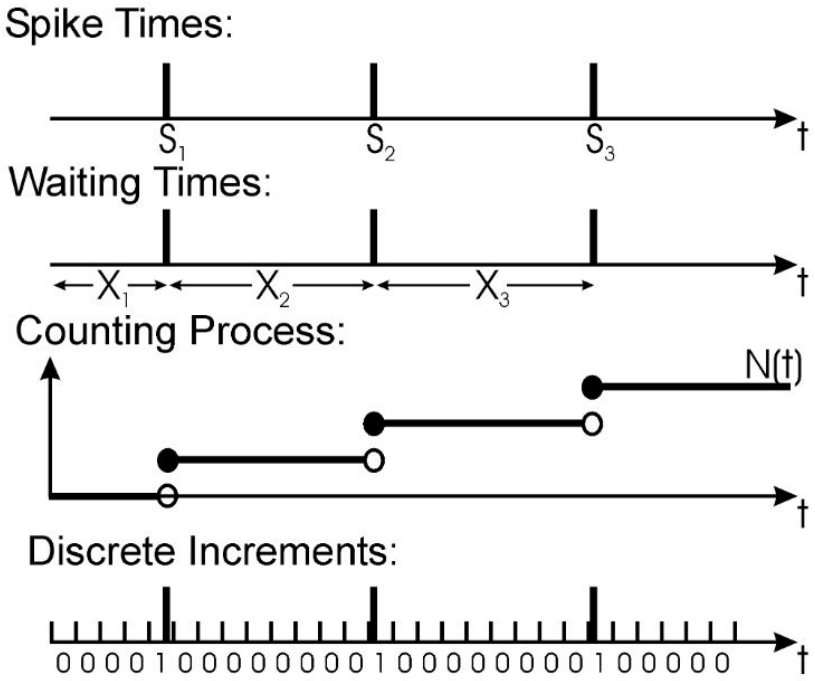
Variance scales with the mean

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

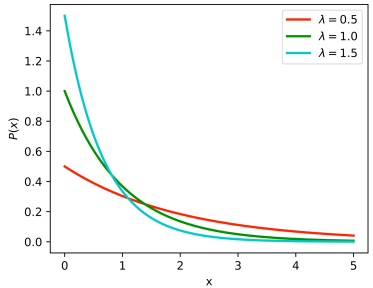


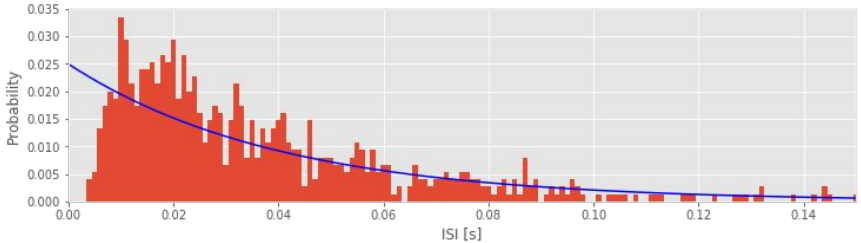
# Exponential waiting times (ISIs)

$$f(x) = \lambda \exp(-\lambda x)$$







# Does our model fit well?