Models Day Tutorials

By Konrad Paul Kording

neuromatch academy

Your modeling day tutorial team

Konrad

Matt Laporte

with help by
Byron Galbraith

and teams: content review, tutorial optimization, copyright, day chiefs, recruit, language, technical

Who is Konrad as a scientist?

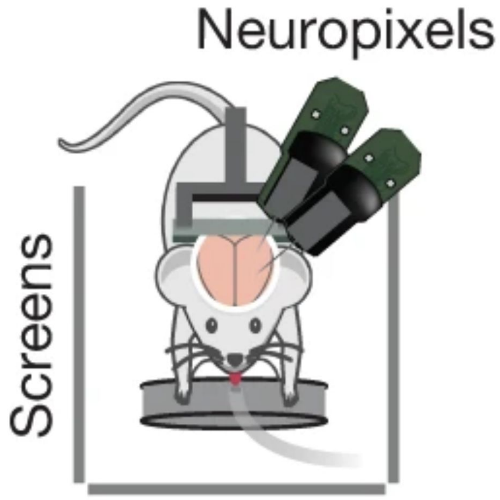
- **Everything is interesting!**
    - Bayesian (why) models.
    - Mechanistic (how) models
    - Descriptive (what) models
    - Physics, Machine Learning, etc
    - **Metascience (how does science work?)**

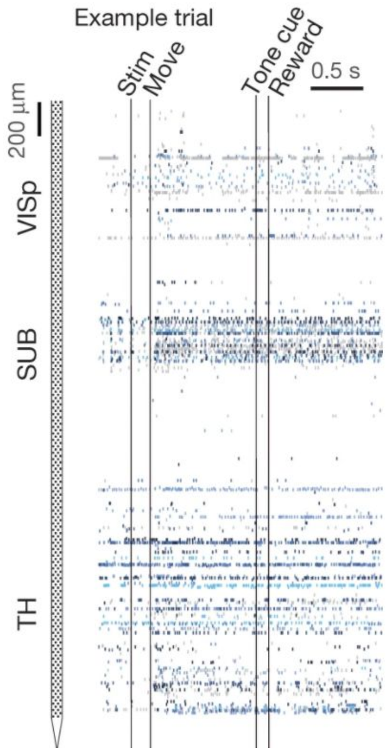
Today: three types of models

- **What models**: find an equation that describes the data
    - No promises about meaning ;)
- **How models**: how does the brain make the data
    - Mechanistic understanding
- **Why models**: why is the brain the way it is?
    - Which things matter? Are optimized? Ecological validity?

Model class I: What models

The Steinmetz data





Steinmetz, ... Harris, 2019, biorxiv

Step 1 of every data analysis

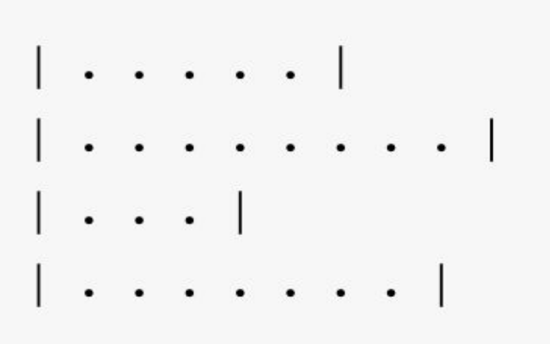
- **Really really understand the data!**
- What is the data format?
- What is in there?
- Does it make sense?

```
type(spike_times)
```

How is the data organized?

spike_times: NumPy array, size 734 (the neurons)

- Of NumPy arrays, size varying (the times of spikes)
- Which contains floats: of the exact time in seconds

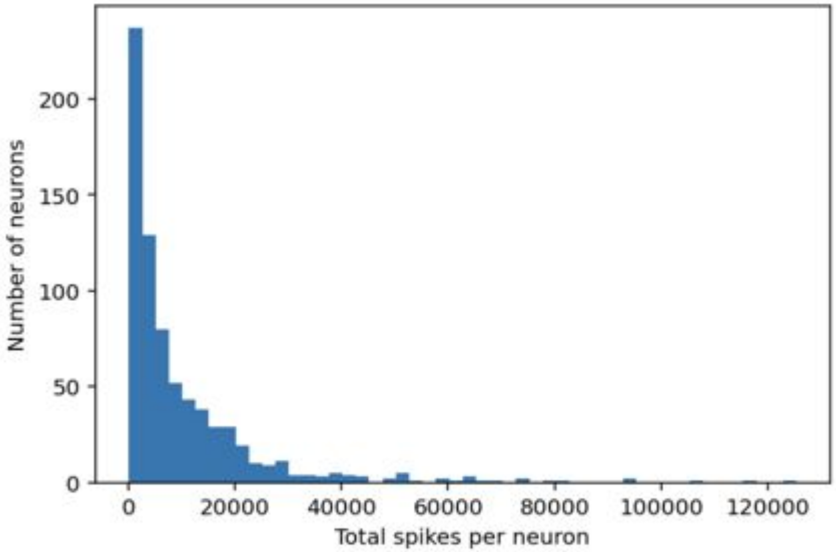


Lets understand it more deeply

- How many spikes per neuron?
- Do all neurons produce similar numbers of spikes?
- What is the distribution?

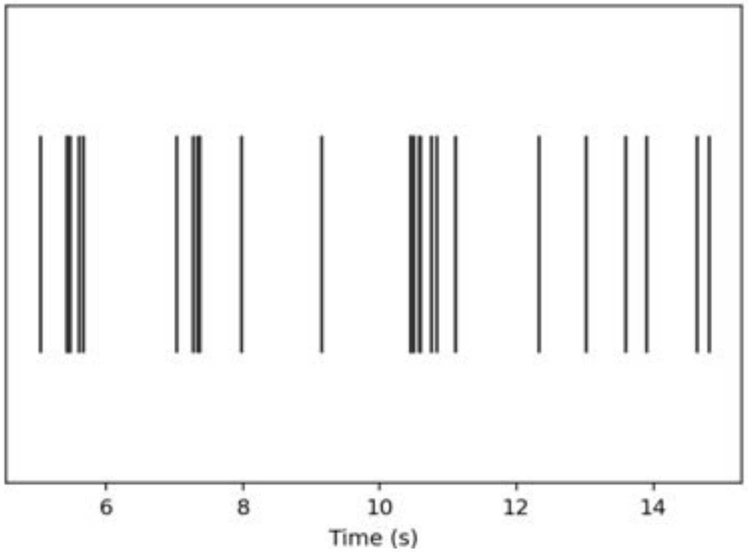
```
plt.hist
```

Spikes per neuron



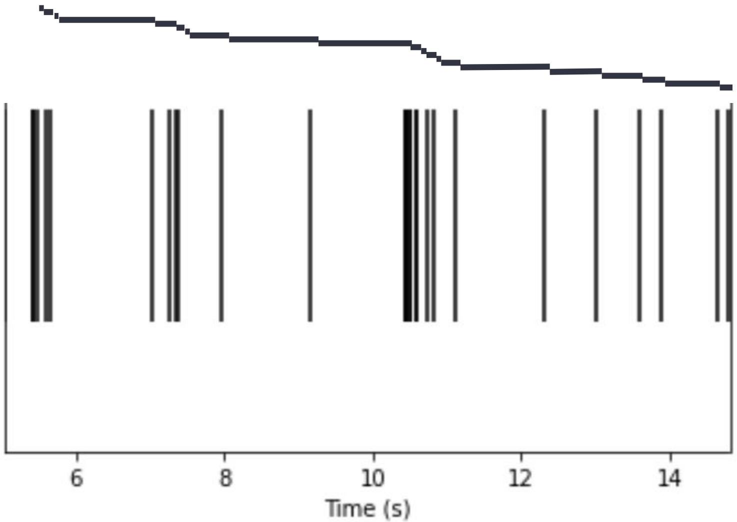
Over
2700 seconds
(~45 min)

Spikes



What is interesting here?
(there are a million things)

Inter-spike intervals



— time interval between spikes

Basic (fundamental) property of neurons

Wearing my **what** hat!

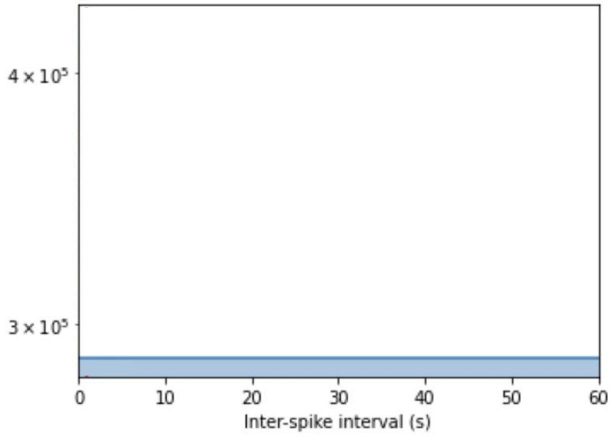
There must be a distribution of inter-spike-intervals

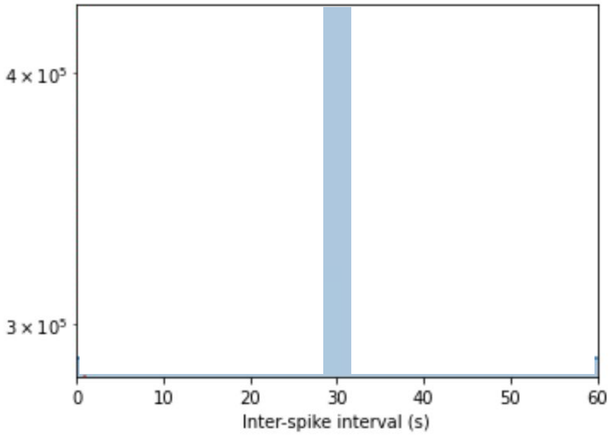
I find that interesting because it must reflect physiology

How could it look like?

However it looks like, it will help me think about brains!

How could it look like?





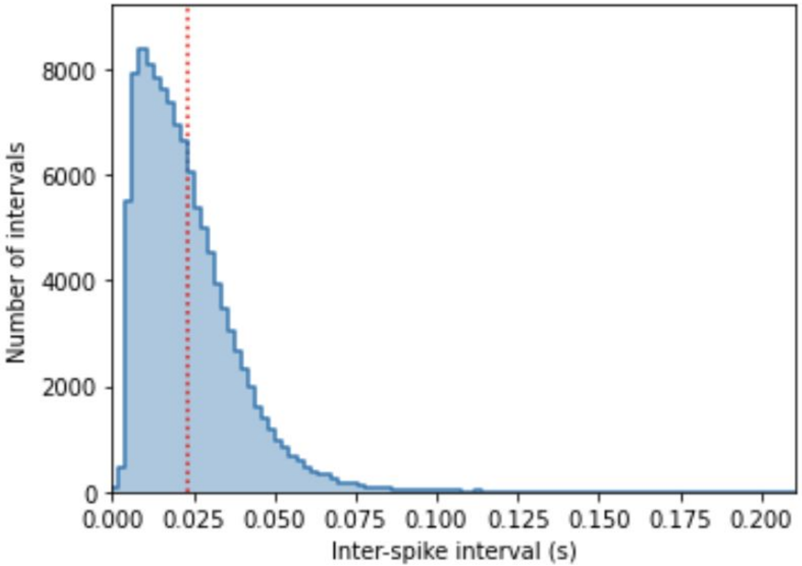
Now let us calculate the inter-spike intervals and their histogram

```
single_neuron_isis =
```

```
plt.hist
```

Empirical ISI distribution

Lets ignore
this region



What: We simply want to describe this with an equation

What is interesting here?

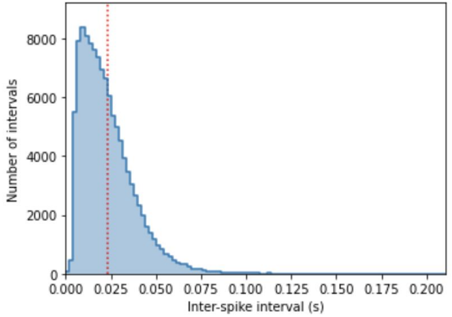
What models: we just fit a function

```
def exponential(xs, scale, rate, x0):
```

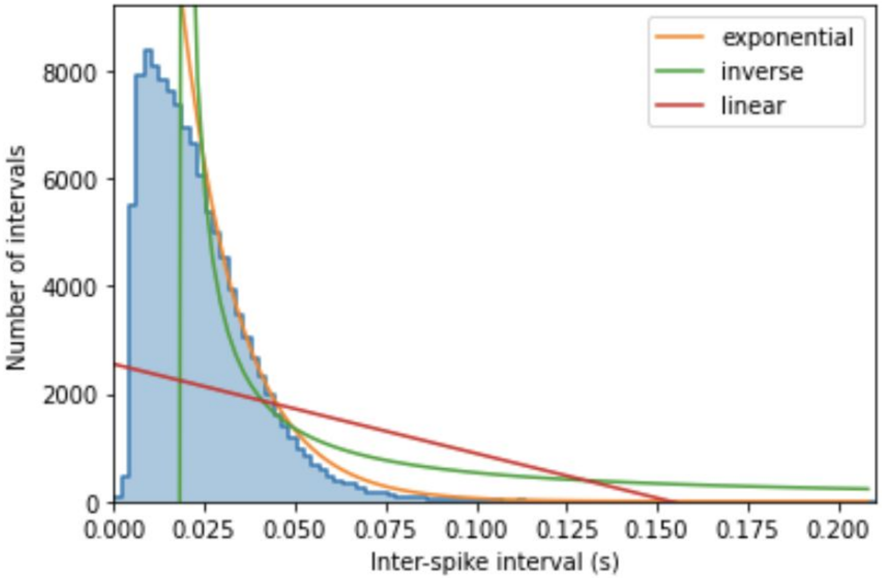
```
def inverse(xs, scale, x0):
```

```
def linear(xs, slope, y0):
```

What does it mean to be good?
Describe the data.
Nihilist!



Exponential fits well

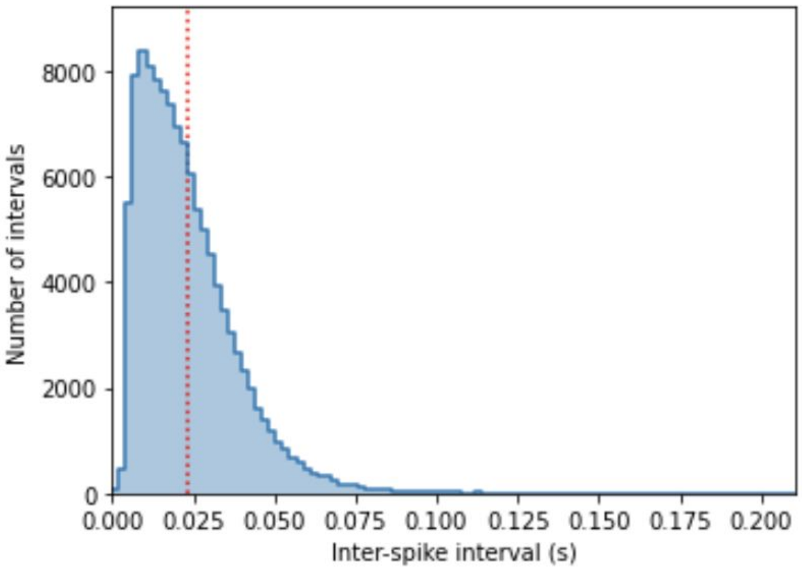


What we just learned

- We take the data and choose something to look at
    - Based on intuition of what matters (we chose ISI histogram)
- Then we try and describe what we see
- With an equation (e.g. exponential function)
- **What models:** We just want to describe what we see

Model class II: How models

Explain: Empirical ISI distribution



How: We want to describe this with an equation that we believe approximates the mechanism

Wearing my how hat!

Well, I know that ISI histograms look exponential
There must be a mechanism for that!
Hey, I know that neurons have a capacitance, so they should integrate signals.
Does that explain their ISI histogram?

The linear integrate and fire neuron

$$\Delta V_t = \alpha I_t$$

- If $V_t > 1$

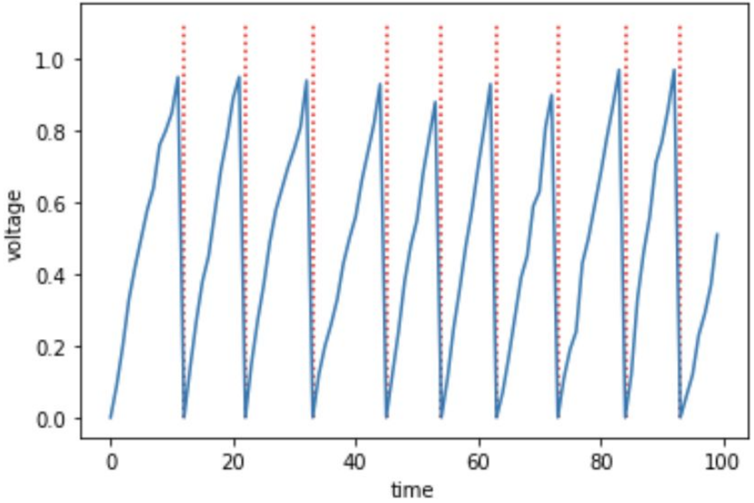
    Spike

    $V_t = 0$

Poisson input

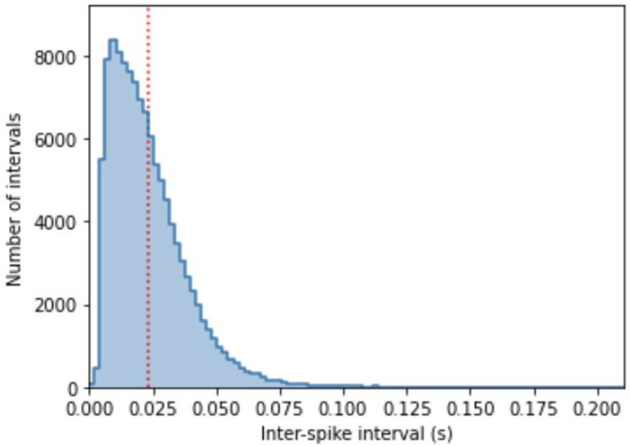
```
v = np.zeros(n_steps)
spike_times = []
for i in range(1, n_steps):
    dv = ...
    v[i] = v[i-1] + dv
    if v[i] > 1:
        spike_times.append(i)
        v[i] = 0
```

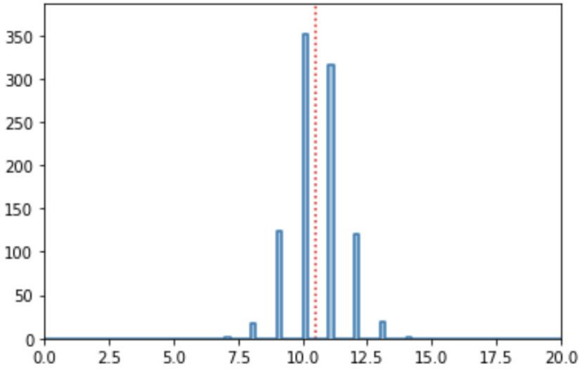
The linear integrate and fire neuron with excitatory drive



Very predictable
(like clockwork)

Exp(-dt) goes in, regular comes out!





Wearing my how hat!

Dang. My model must be wrong!

A network that is primarily excitatory becomes regular!

Well, but I know there is inhibition too!

What if the two are the same strength?

Balanced Excitation and Inhibition!

- Make it leaky integrate and fire

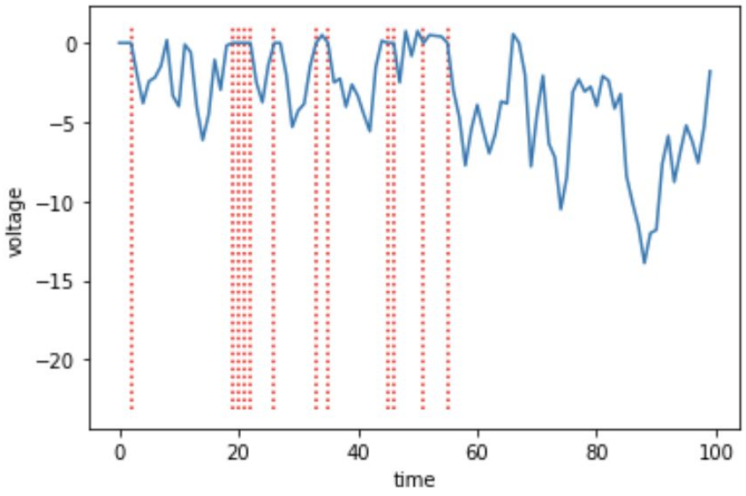
$$\Delta V_t = -\alpha V_t + I$$

- Balanced excitation and inhibition

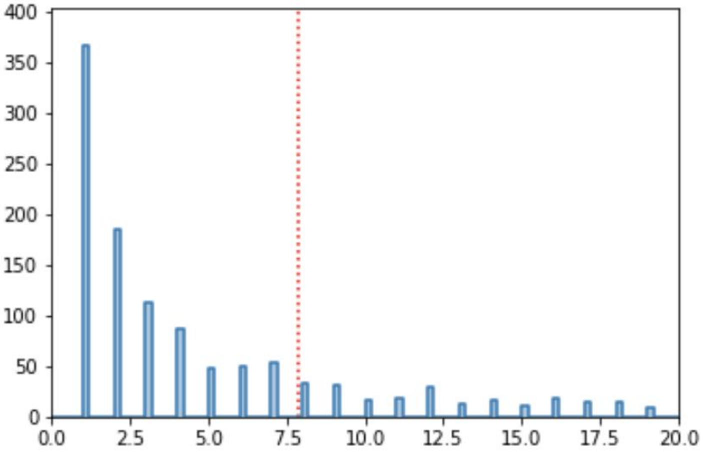
$$I = \alpha\left(Poisson(\lambda) - Poisson(\lambda)\right)$$

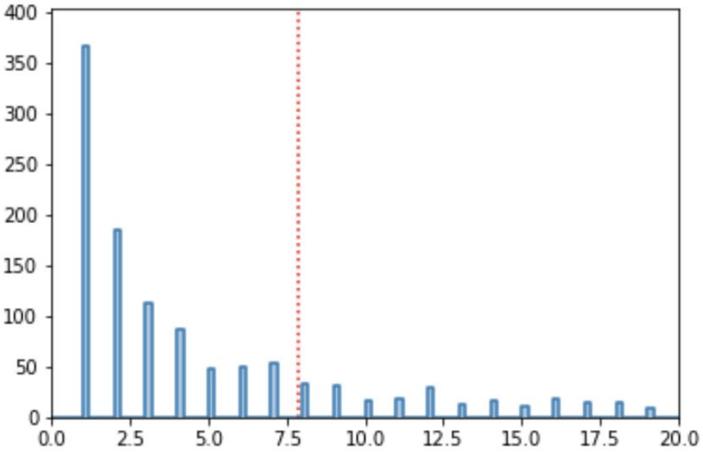
```
for i in range(1, n_steps):
    dv = ...
    v[i] = v[i-1] + dv
    if v[i] > 1:
        spike_times.append(i)
        v[i] = 0
```

Results of balanced inputs



What goes in is what goes out!





Intuition

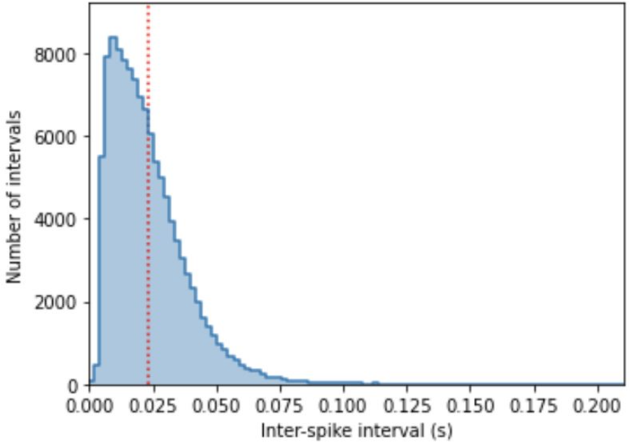
- Over period of integration window
    - Certain probability of hitting threshold (more noise = higher p)
- All integration windows are independent
- Exponential because fixed probability of ending.

What we just learned

- In how models we start with an idea of how biology solves a problem
- We hypothesize a mechanism
- We then simulate this mechanism
- And check how well it coincides with our measurements
- **How models:** We want to describe what we measured in terms of a mechanism we believe in

Model class III: Why models

Explain: Empirical ISI distribution



Why: We want to describe this with an equation that comes from optimizing a problem we believe in

Wearing my why hat!

I know the ISI histogram looks like an exponential

My **how** friends feel they understand how that comes about.

But **why** is that good? Evolution could have come up with different distributions (e.g. by not making excitation and inhibition be balanced)

An optimization problem

- **Objective:** Maximize Information transmitted by ISIs
- How would optimal ISI distribution look like?

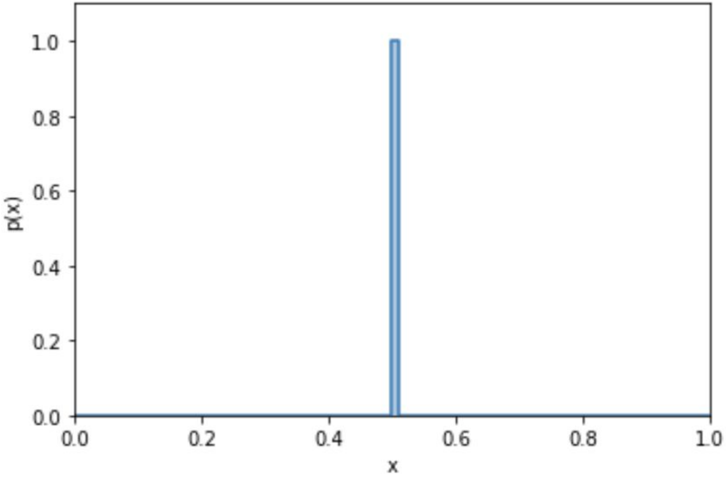
Entropy ~ Information

$$H_b(X) = -\sum_{x \in X} p(x) \log_b p(x)$$

$b = 2$ for units of *bits*

$b = e$ for *nats*

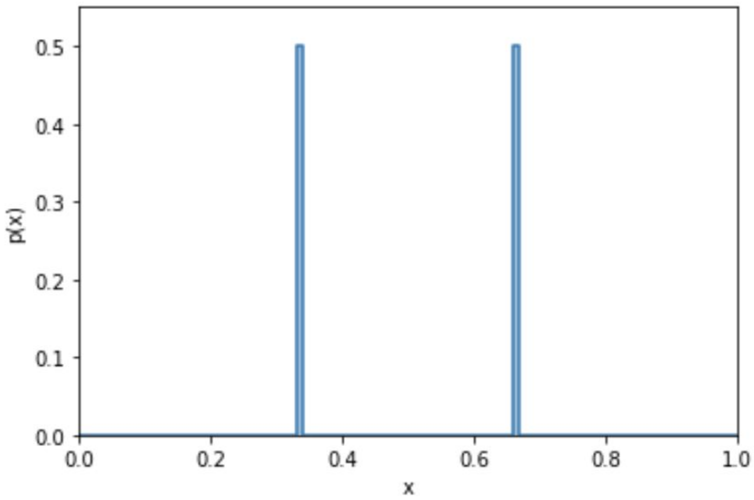
Unimodal distribution



$$H_b(X) = -\sum_{x \in X} p(x) \log_b p(x)$$

0 bits

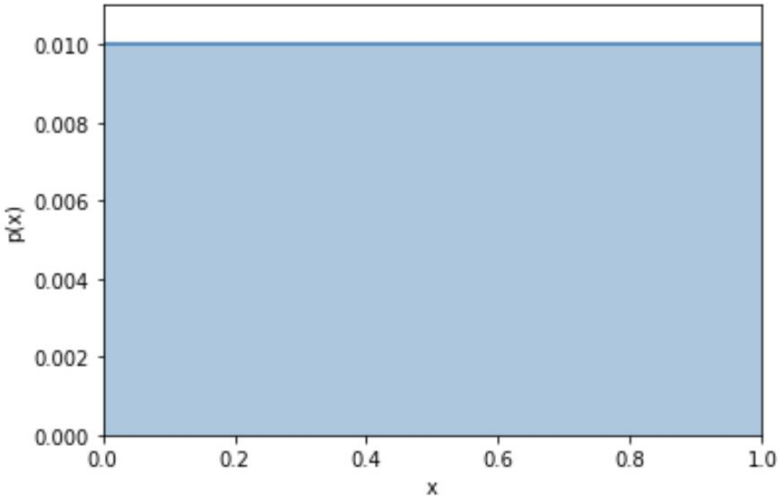
Bimodal distribution



$$H_b(X) = -\sum_{x \in X} p(x) \log_b p(x)$$

1 bit

Uniform



$$H_b(X) = -\sum_{x \in X} p(x) \log_b p(x)$$

Most bits

```
entropy(pmf)
```

Again wearing my **why** hat

It looks exponential not uniform

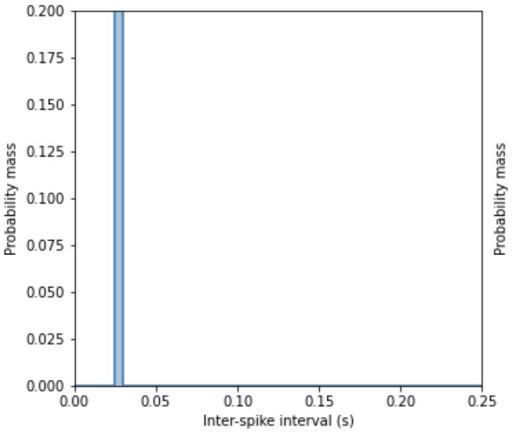
But hey, uniform kinda does not make sense: uniform until which value?

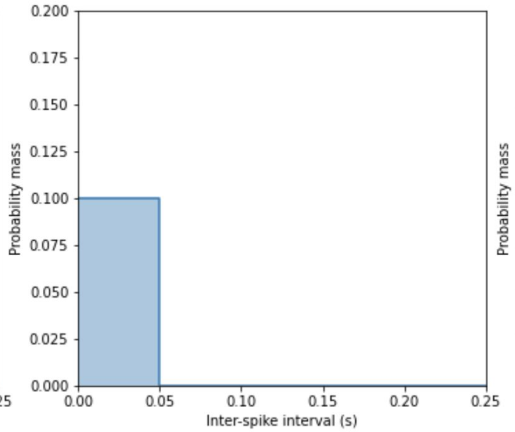
Our model must miss something!

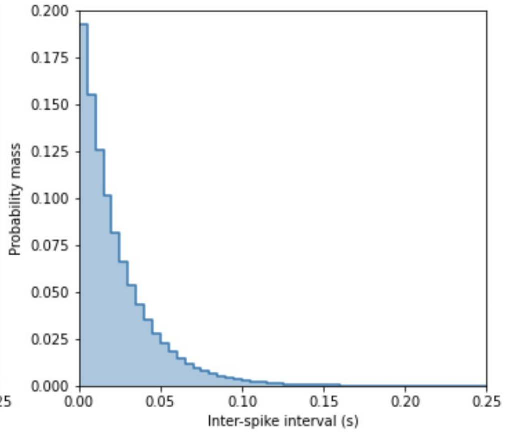
An optimization problem

- **Constraint:** Each neuron can only produce a certain number of spikes (they cost energy)
- **Objective:** Maximize Entropy of ISIs given fixed number of spikes
- How would optimal ISI distribution now look like?

Now, compare one peak, uniform, and exponential



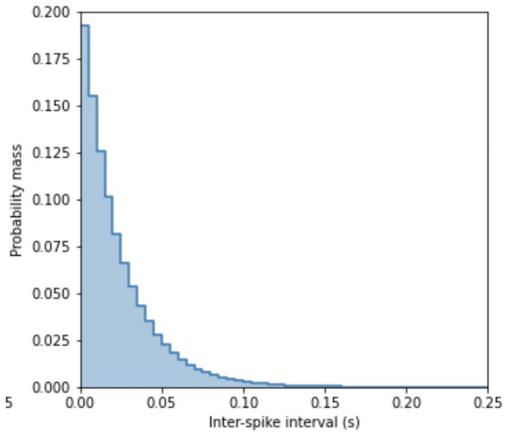




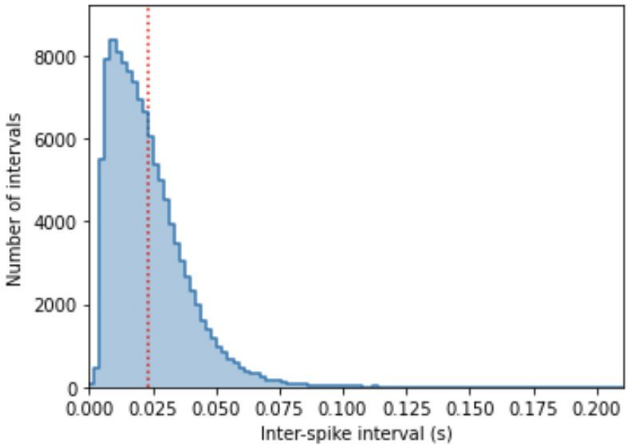
Now, calculate the entropy of each

```
entropy(ps)
```

The exponential has most entropy



What about real neurons? ISI histogram into probabilities



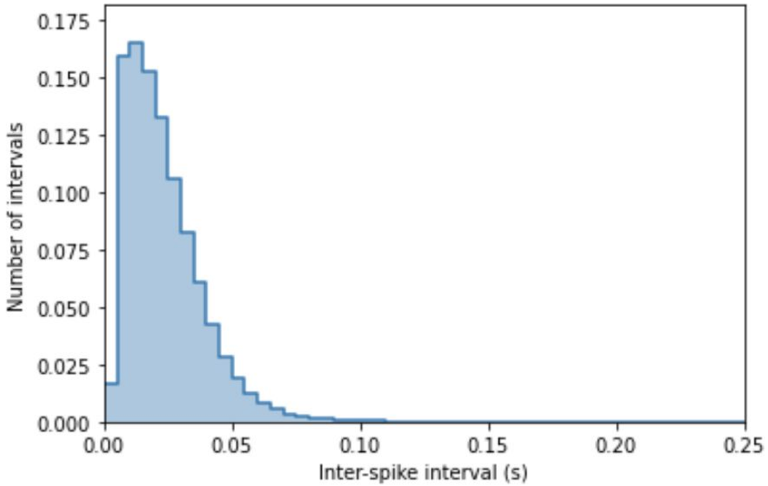
Roughly:

$$p(i) = \frac{n_i}{\sum_i n_i}$$

Now, you convert the histogram into probabilities

```
def pmf_from_counts(counts)
```

Now we have a probability distribution



Now, calculate entropy for real data

```
h = entropy(ps)
```

Analytically

- Among all distributions with fixed mean on positive numbers
- Exponential is the highest entropy one.

- Intuition: If I make a big ISI bigger, I force other ISIs to be smaller. And hence, it is better to have more smaller than bigger values.

What we just learned

- In why models we start with an idea of what biology should optimize
- We hypothesize constraints
- Then we solve for the best solution, given the constraints (here exp)
- And check how well it coincides with our measurements (here ~exp)
- **Why models:** We want to understand which behavior would be optimal (for the real problem in the real world)

What we learned overall in today's tutorials

- For all models we want to describe something we feel is important (here ISI distribution)
- In **What** models we just want to describe the data
- In **How** models we want to describe it in terms of mechanisms
- In **Why** models we want to describe it in terms of optimal solutions to (constrained) problems in the real world that the brain needs to solve