BOOTSTRAP CONTINUED

10.17.2018

OFFICE HOURS

* **Today only:** my office hours will be 2-3pm instead of 1:30-3pm

RECAP

* Bootstrap: generating fake datasets out of your real dataset

BOOTSTRAP EXAMPLE

* we have a dataset: an electrophysiological recording from a spiking neuron

* we want to estimate the mean background spike rate of the neuron

* but we have a limited sample of data!

BOOTSTRAP EXAMPLE

* we want to know the **standard error** of the **mean** firing rate

* let’s estimate it using a bootstrap!

STANDARD ERROR

* sometimes there are analytic solutions to finding the standard error

* in the neuron example our data acts like a Bernoulli random variable where we’re trying to estimate the parameter q

STANDARD ERROR

* it turns out the standard error for q in a Bernoulli random variable has an analytic form:

$$\sigma_{error} = \sqrt{\frac{q(1-q)}{n}}$$

CENTRAL LIMIT THEOREM

* you might have noticed that the distribution of bootstrap means tends to look pretty consistently like a bell curve

* even though the underlying distribution we're picking from can be super weird

CENTRAL LIMIT THEOREM

* this is because of the **central limit theorem**, which says that the sum of a large number of independent, identically-distributed (iid) random variables will tend to a Gaussian distribution

CENTRAL LIMIT THEOREM

* one consequence of the CLT is that *Gaussian distributions appear everywhere*

* this is because many interesting things are the result of complex processes where, effectively, many small independent contributions are summed

CENTRAL LIMIT THEOREM

* DEMO: let’s take a bunch of random values (from the same distribution)

* and take their average

* let’s do this many times

CENTRAL LIMIT THEOREM

* as the number of samples we take increases, the distribution of their averages converges to..

* a Gaussian distribution!

END