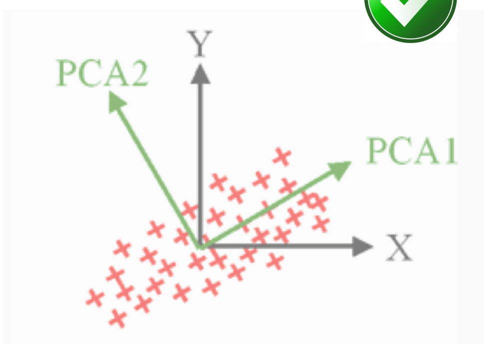
Lecture 6: Statistics

May 6th, 2015

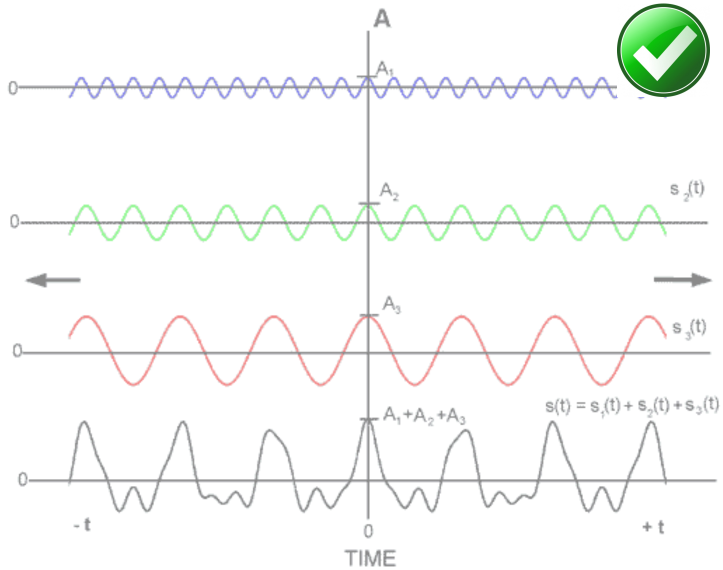
Lane McIntosh & Kiah Hardcastle

Math Tools for Neuroscience

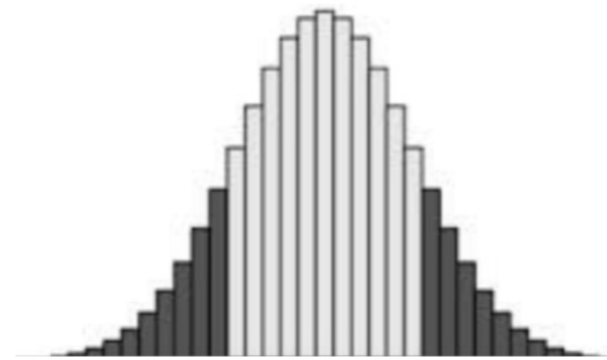
Topics we have covered this quarter



dimensionality reduction



Fourier transforms, convolutions, and filtering out noise

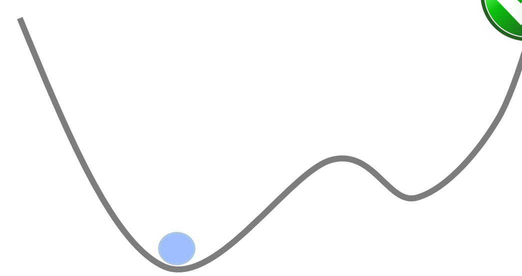


statistics, Bayesian probability, and information theory

$$C_m \frac{dV}{dt} = -g_L(V - E_L) + g_L \Delta_T e^{\frac{V-V_T}{\Delta_T}} - u$$

$$\tau_w \frac{du}{dt} = a(V - E_L) - u$$

differential equations and modeling



optimization

Goals for today

Common ways to quantitatively summarize data

Given a data set, what statistical test to use

Today’s lecture

Data statistics

- mean, variance, stdev, std. err
- common distributions
- confidence intervals

Hypothesis testing

- Parametric tests
- When to use different parametric tests
- Nonparametric tests

Example 1: Gambling

heads (0)

lose $5

tails (1)

win $5

Example 1: Bernoulli random variable

heads (0)

lose $5

tails (1)

win $5

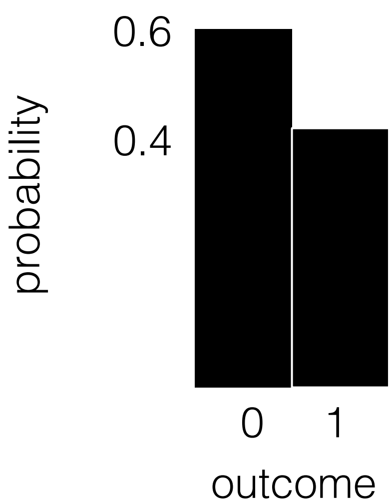
Bernoulli random variable

Example 1: Bernoulli distribution

Outcomes

0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1

9 heads, 6 tails



heads (0)

lose $5

tails (1)

win $5

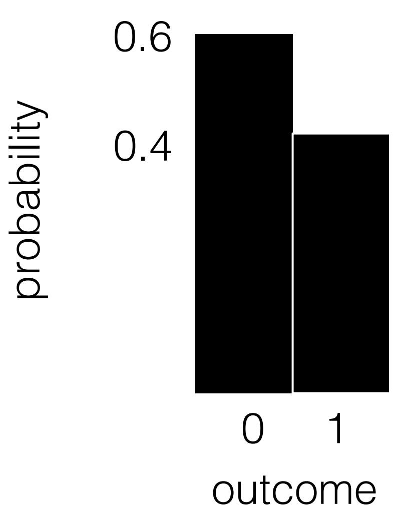
Bernoulli distribution

Example 1: Bernoulli distribution

Outcomes

0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1

9 heads, 6 tails



Mean outcome?

Average = (0+1+0+0+0+1+…+1) / 15
= 0.6*0 + 0.4*1
= 0.4

Mean winnings?

Average = (-$5+$5-$5-$5-$5+$5+…+$5) / 15
= 0.6*$-5 + 0.4*$5
= -$1

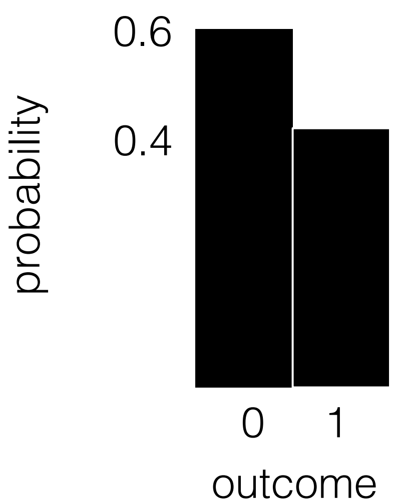
MATLAB: `mean(data)`

Example 1: Bernoulli distribution

Outcomes

0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1

9 heads, 6 tails



Mean outcome?

Average = (0+1+0+0+0+1+…+1) / 15
= 0.6*0 + 0.4*1
= 0.4

Mean winnings?

Average = (-$5+$5-$5-$5-$5+$5+…+$5) / 15
= 0.6*$-5 + 0.4*$5
= -$1

How spread out are the outcomes? (variance)

$$\sigma^2 = E[(X - \mu)^2]$$

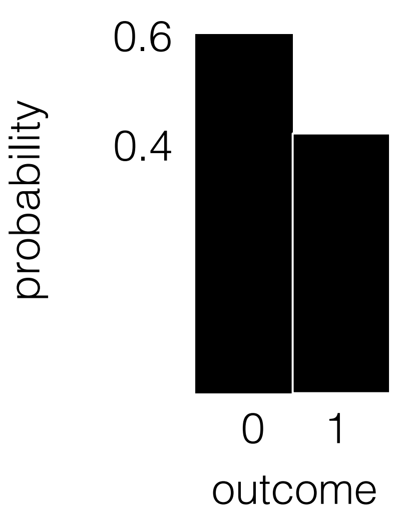
MATLAB: `var(data)`

Example 1: Bernoulli distribution

Outcomes

0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1

9 heads, 6 tails



Mean outcome?

Average = (0+1+0+0+0+1+…+1) / 15
= 0.6*0 + 0.4*1
= 0.4

Mean winnings?

Average = (-$5+$5-$5-$5-$5+$5+…+$5) / 15
= 0.6*$-5 + 0.4*$5
= -$1

How spread out are the outcomes? (variance)

$E[(X-\mu)^2]$ = [(0-.4)^2+(1-.4)^2+(0-.4)^2+…+(1-.4)^2]/ 15
= p*(1-p) (only for Bernoulli variables!)
= 0.4*0.6 = 0.24

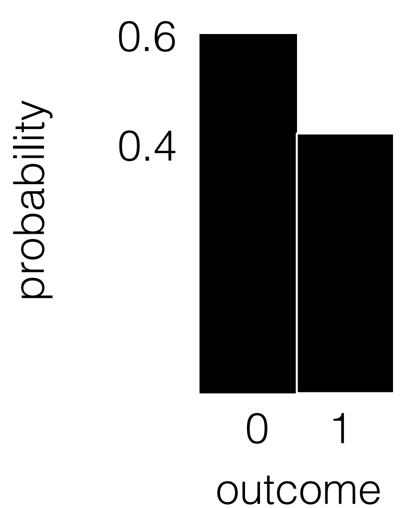
MATLAB: `var(data)`

Example 1: Bernoulli distribution

Outcomes

0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1

9 heads, 6 tails



Mean outcome?

Average = (0+1+0+0+0+1+…+1) / 15
= 0.6*0 + 0.4*1
= 0.4

Mean winnings?

Average = (-$5+$5-$5-$5-$5+$5+…+$5) / 15
= 0.6*$-5 + 0.4*$5
= -$1

How spread out are the winnings? (variance)

$E[(X-\mu)^2]$ = [(-$5-$-1)^2+($5-$-1)^2+…+($5-$-1)^2]/ 15
= $25.7143

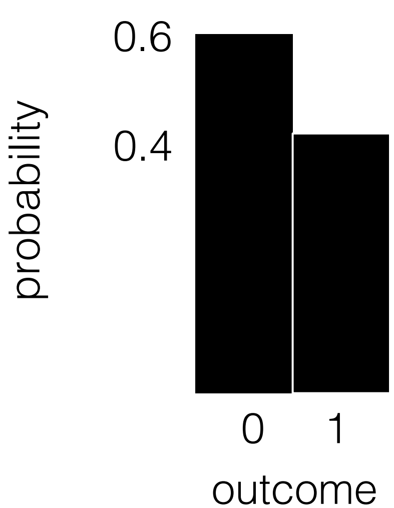
MATLAB: `var(data)`

Example 1: Bernoulli distribution

Outcomes

0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1

9 heads, 6 tails



Mean outcome?

Average = (0+1+0+0+0+1+…+1) / 15
= 0.6*0 + 0.4*1
= 0.4

Mean winnings?

Average = (-\$5+\$5-\$5-\$5-\$5+\$5+…+\$5) / 15
= 0.6*\$-5 + 0.4*\$5
= -\$1

Standard deviation of the winnings?

$$\sigma = \sqrt{E[(X-\mu)^2]}$$

= sqrt([(-\$5-\$-1)^2+(\$5-\$-1)^2+…+(\$5-\$-1)^2]/ 15)
= \$5.0709

MATLAB: `std(data)`

Example 1: Gambling

heads (0)

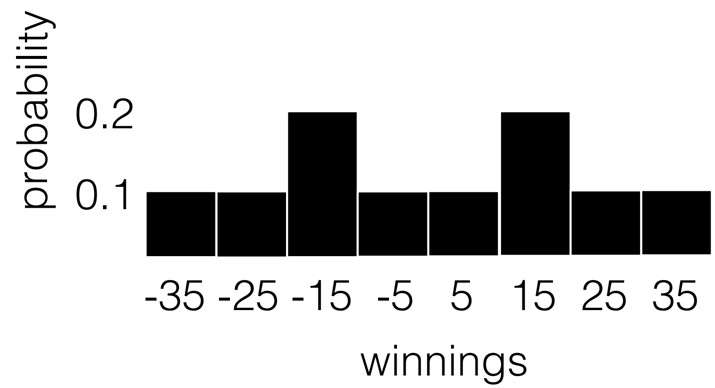
lose $5

tails (1)

win $5

Example 1: Binomial distribution

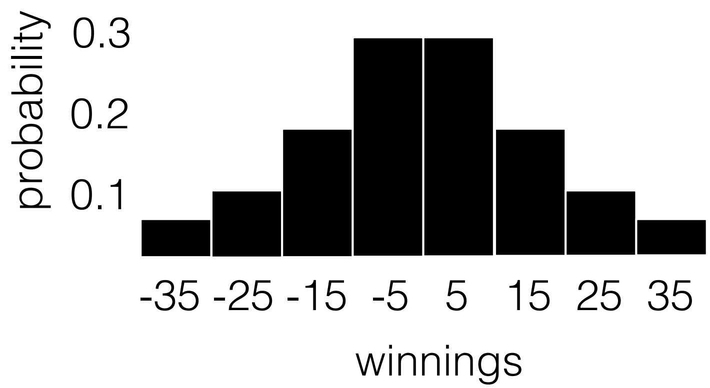
Outcomes	Winnings
0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1	-$15
0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0	$5
1, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1	$35
1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0	-$15
1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0	$25
0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1	-$5
0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0	$5
0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1	$15



Example 1: Binomial distribution

Outcomes	Winnings
0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1	-$15
0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0	$5
1, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1	$35
1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0	-$15
1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0	$25
0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1	-$5
0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0	$5
0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1	$15

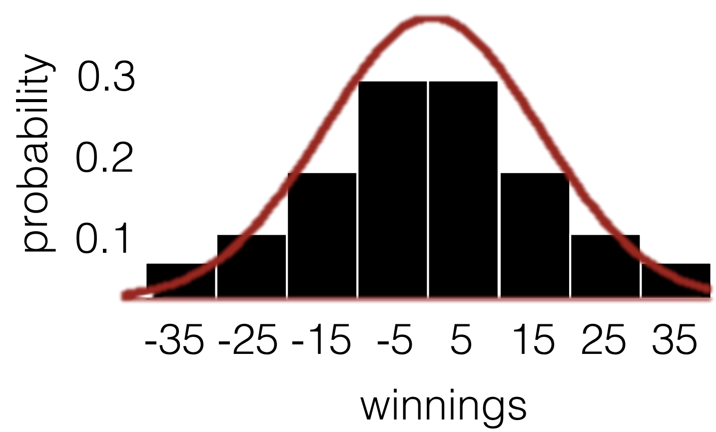
...



Central limit theorem

Central limit theorem:

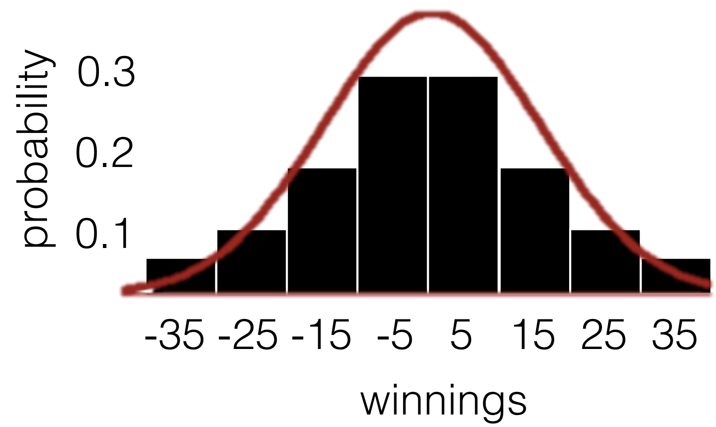
The average of many samples will tend to be normally distributed, regardless of the underlying distribution



Central limit theorem

Central limit theorem:

The average of many samples will tend to be normally distributed, regardless of the underlying distribution



Examples:

Membrane potential noise from many ion channels opening and closing randomly

Distance yeast travel from the center of the colony

Height of individuals in a population

French fry length

Central limit theorem

Central limit theorem:

The average of many samples will tend to be normally distributed, regardless of the underlying distribution

Examples:

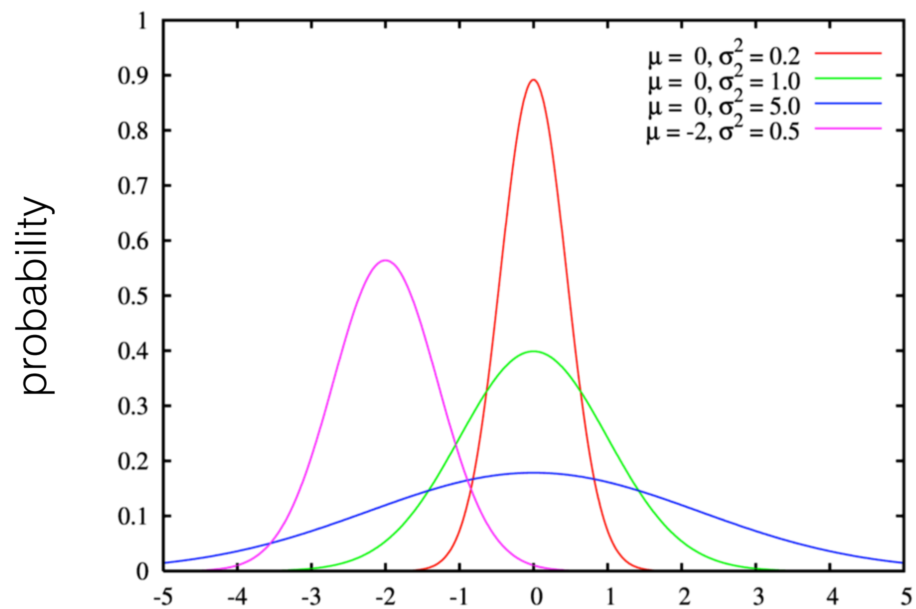
Membrane potential noise from many ion channels opening and closing randomly

Distance yeast travel from the center of the colony

Height of individuals in a population

French fry length

Normal (Gaussian) distributions with different means and variances



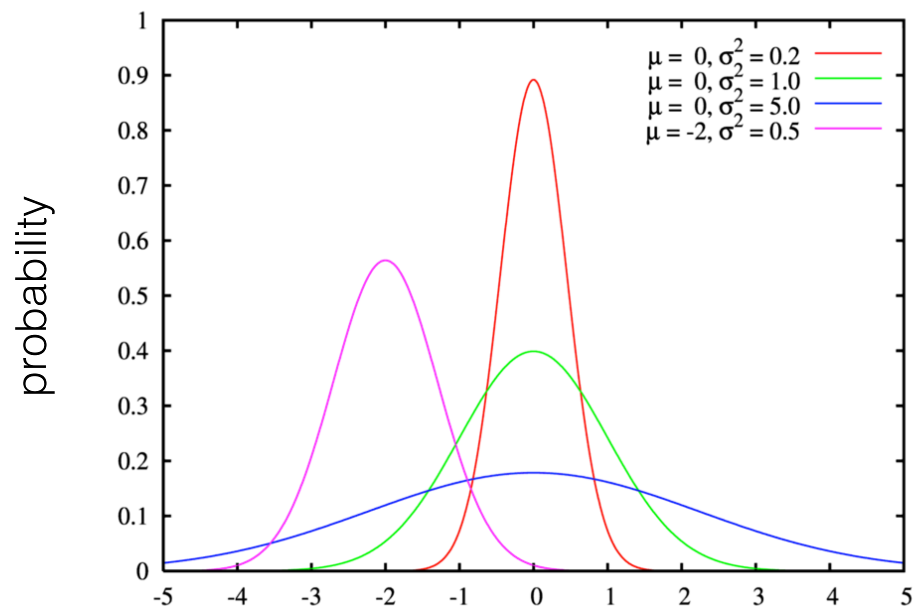
Central limit theorem

Central limit theorem:

The average of many samples will tend to be normally distributed, regardless of the underlying distribution

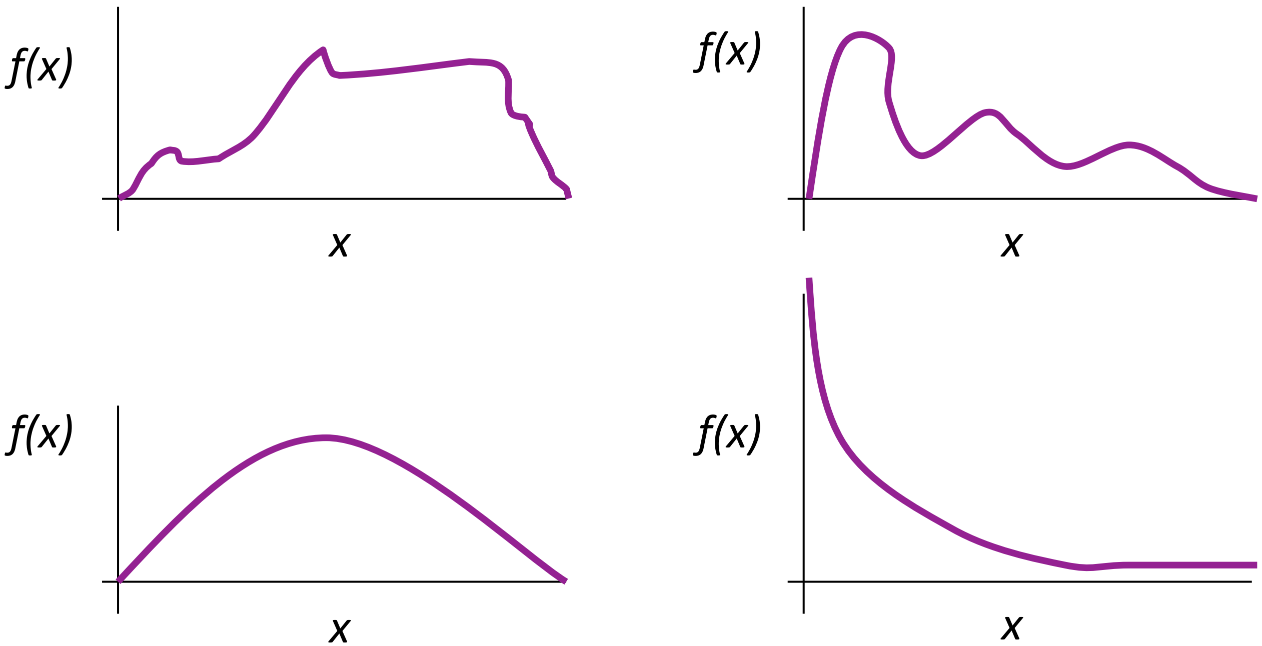
How many distributions are there?

Normal (Gaussian) distributions with different means and variances



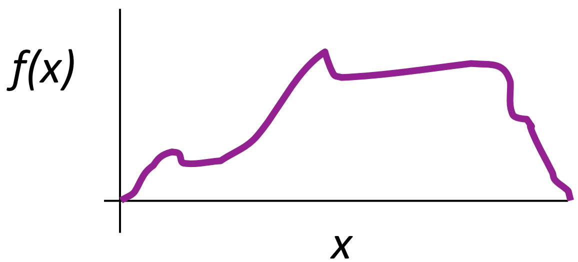
Probability distributions

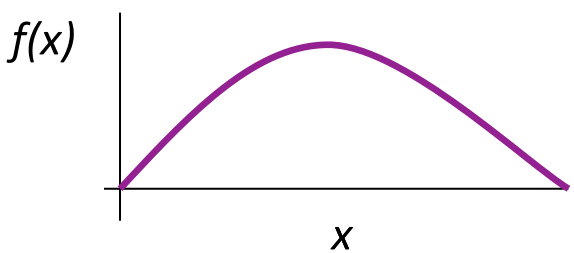
Infinitely many distributions.



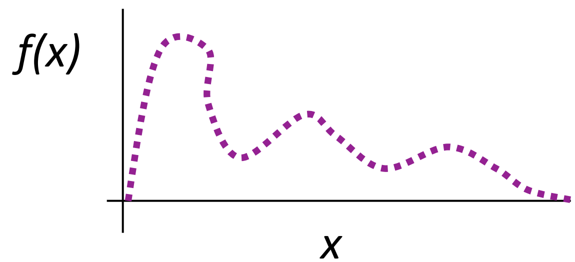
Probability distributions

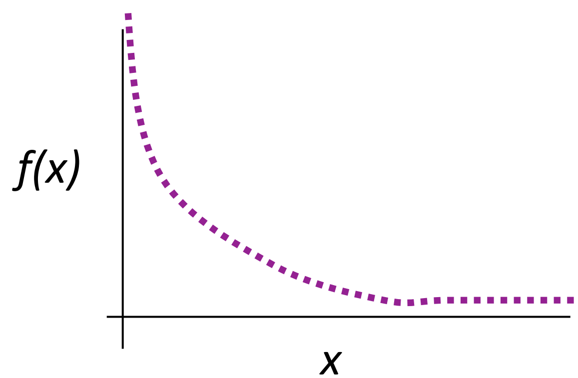
Infinitely many distributions.





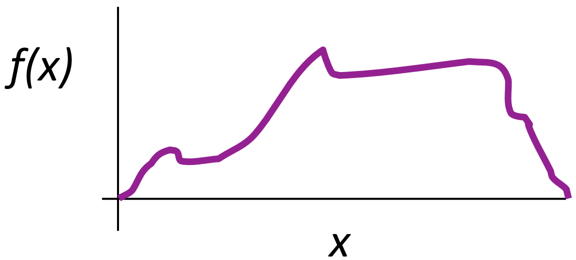
Could be continuous or discrete.

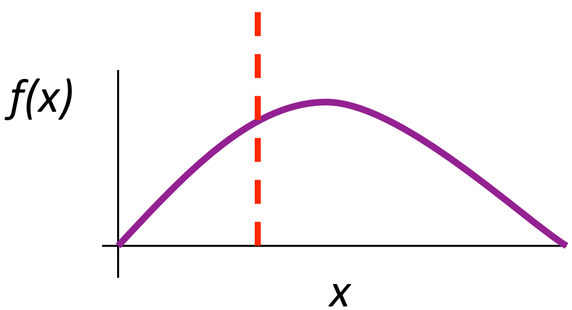




Probability distributions

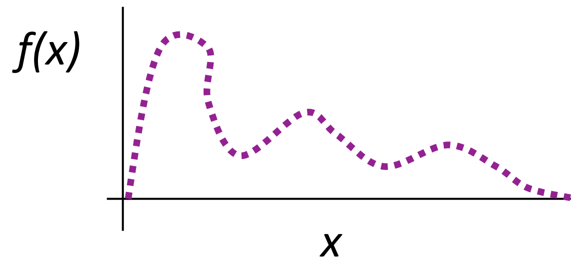
Infinitely many distributions.

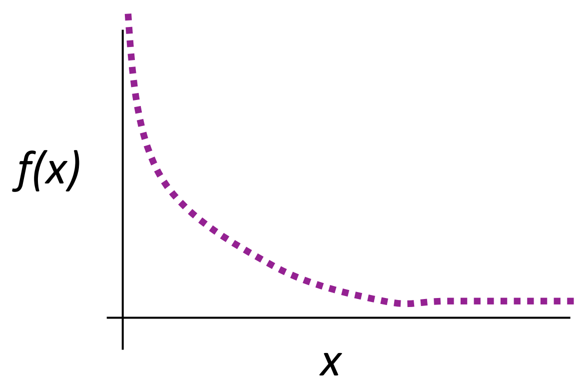




Requires sum to be 1.

Could be continuous or discrete.

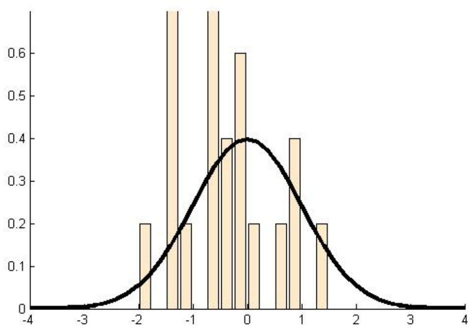




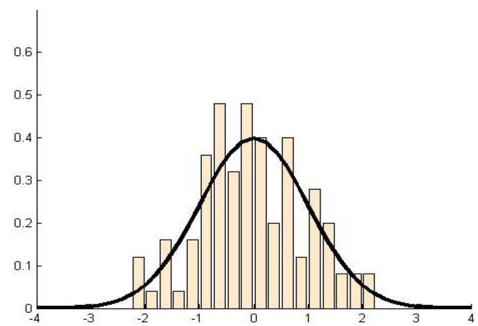
The probability of any event must be between 0 and 1.

How is your data distributed?

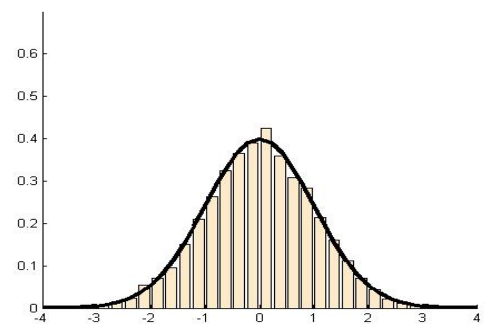
After collecting your data, how do you know its probability distribution?



n=20



n=100



n=1000

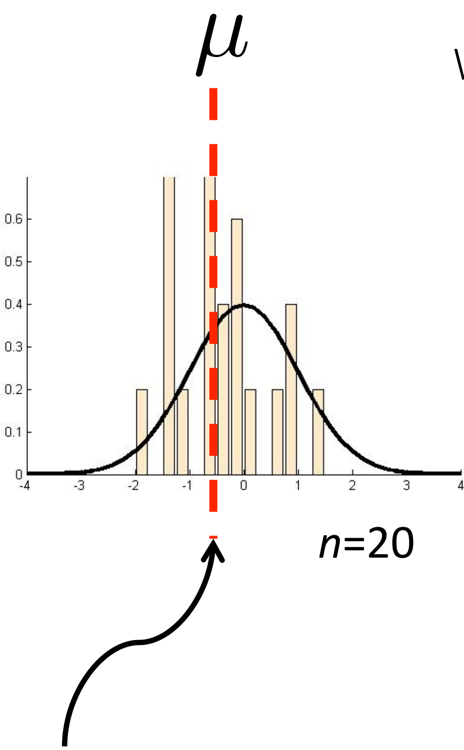
Answer: Just bin your data by making a histogram.
The more data, the better approximation.

For example:
Time for mouse to exit maze
Change in length of a pyramidal neuron
Mass/Charge for a particular chemical bond

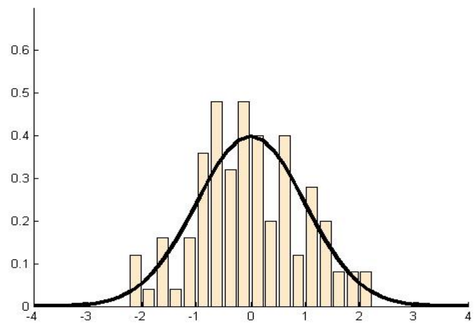
MATLAB:
```
[count, bins] = hist(data)
probs = count/sum(count)
bar(bins, probs)
```

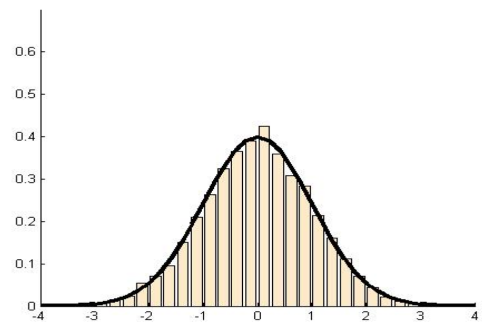
Point statistics

What is the sampling error of our mean?



Estimate of mean

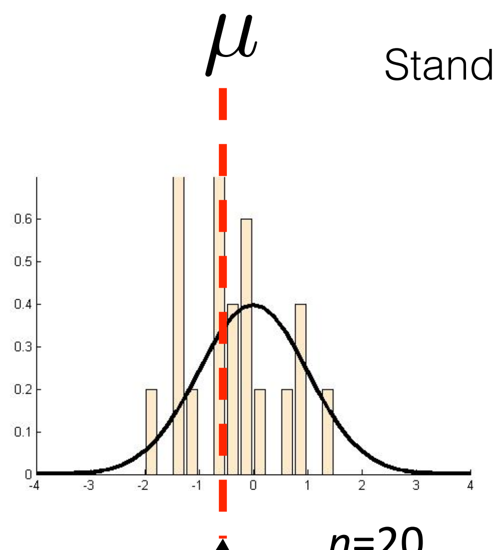




MATLAB: `sem=std(data)/sqrt(length(data))`

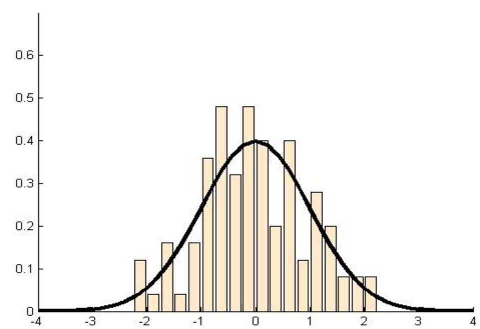
Standard error

Standard error is the sampling error of our mean

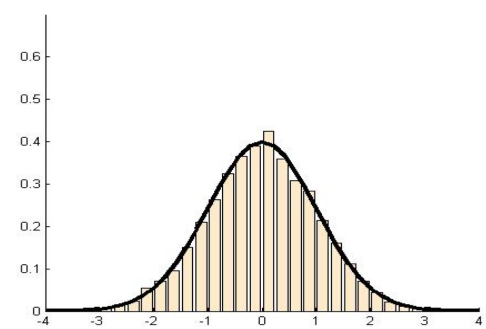


n=20

Estimate of mean



n=100



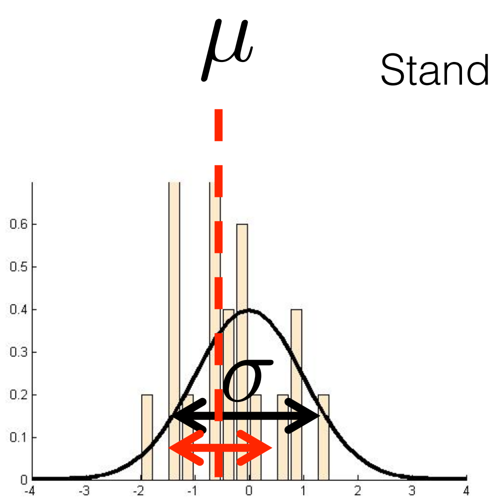
n=1000

Standard error = $\dfrac{\sigma}{\sqrt{n}}$

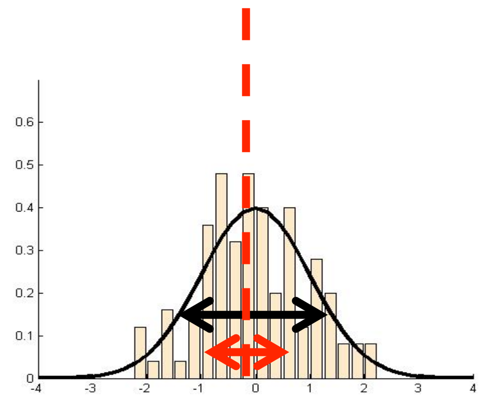
MATLAB: `sem=std(data)/sqrt(length(data))`

Standard error

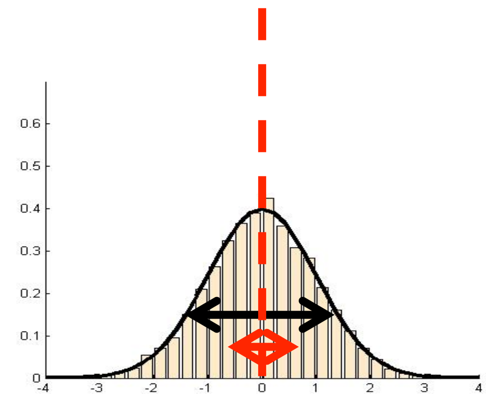
Standard error is the sampling error of our mean



n=20



n=100



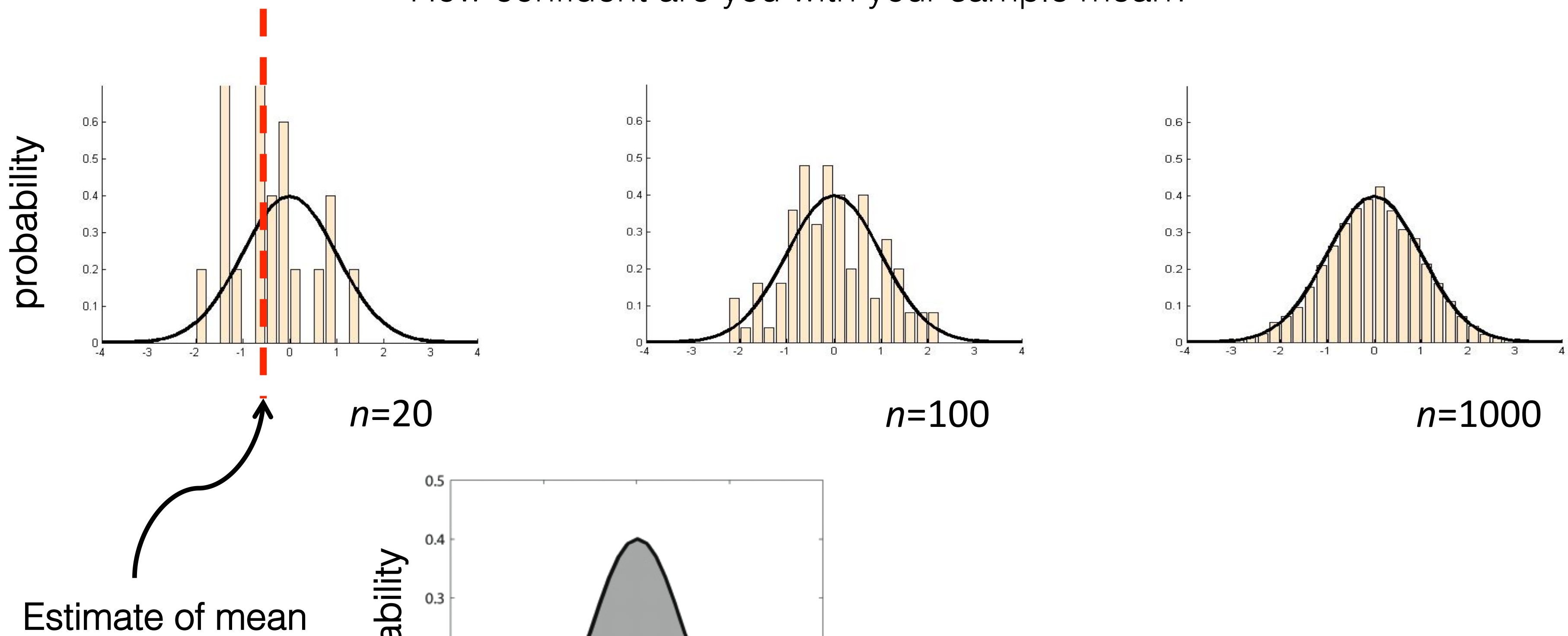
n=1000

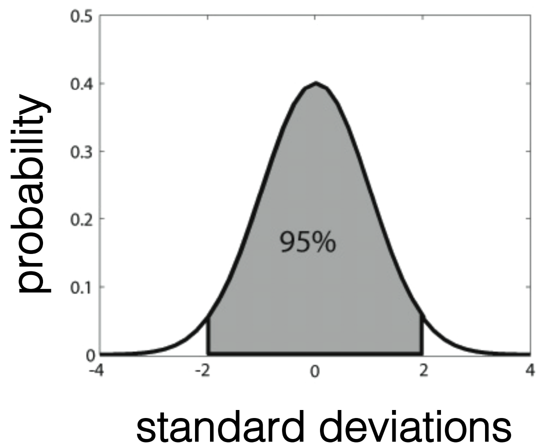
Standard error = $\dfrac{\sigma}{\sqrt{n}}$

MATLAB: `sem=std(data)/sqrt(length(data))`

Confidence intervals

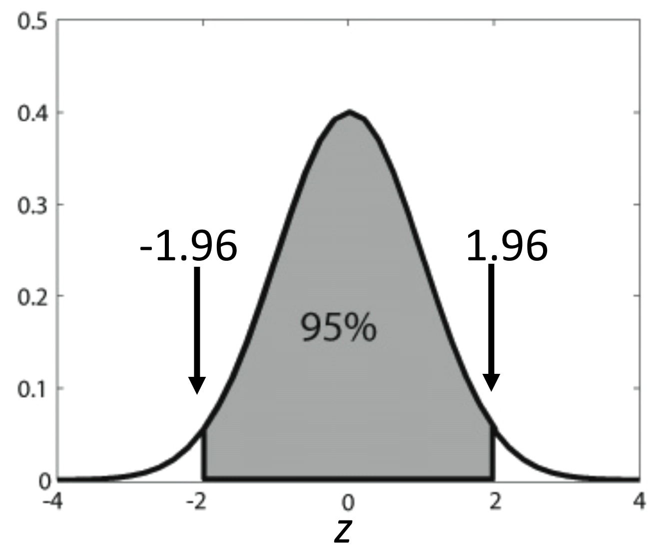
How confident are you with your sample mean?





Background: The standard normal distribution (Z)

The standard normal distribution has mean 0 and standard deviation 1.



P(-1.96 < Z < 1.96) = 0.95

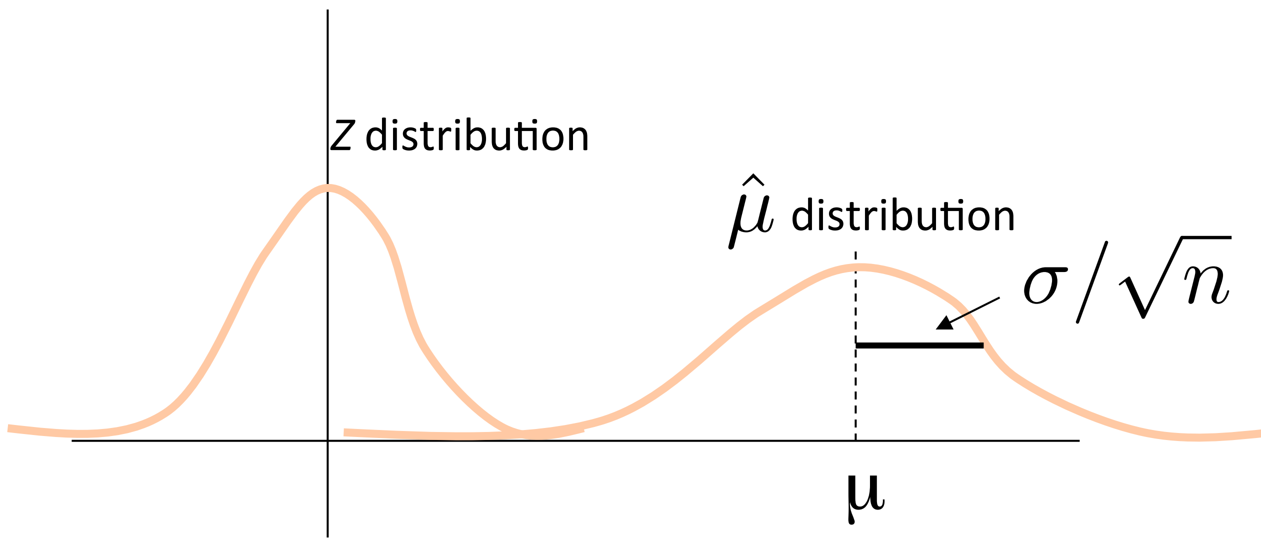
– i.e., 95% of the time, a sample from the Z distribution is between -1.96 and +1.96

Calculating a confidence interval

You measured the sciatic nerves of 25 mice: x_1, x_2, .. x_{25}

Your point estimate of the sciatic nerve length is: $\hat{\mu} = \sum_i \frac{x_i}{25} = 1.5$

$$Z = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}}$$



Calculating a confidence interval

$$Z = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}}$$

P(-1.96 < Z < 1.96) = 0.95

$$P(-1.96 < \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} < 1.96) = 0.95$$

Solve for μ:

$$P\left(\hat{\mu} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \hat{\mu} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Calculating a confidence interval

lower bound

upper bound

$$P\left(\hat{\mu} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \hat{\mu} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$\hat{\mu} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

what is the 95% confidence interval for our problem?

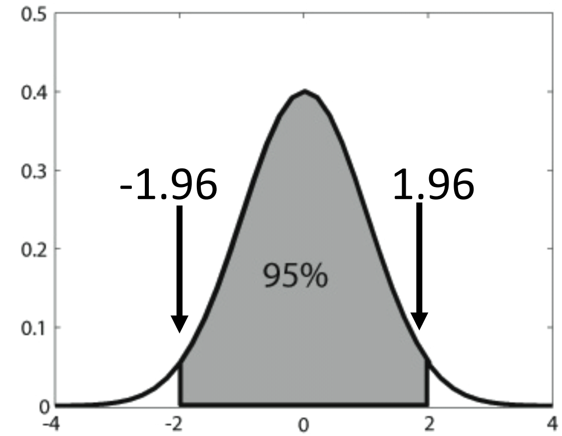
n = 25 mice

$\hat{\mu}$ = 1.5 microns

σ = 1.3 microns

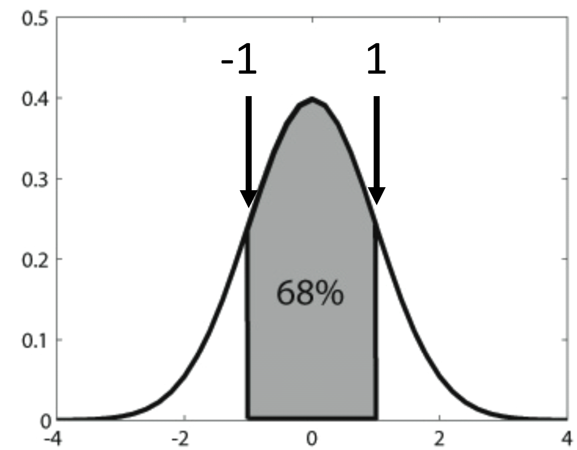
CIs are related to the standard error of the mean (SEM)

For a 95% CI, $\hat{\mu} \pm 1.96 \frac{\sigma}{\sqrt{n}}$

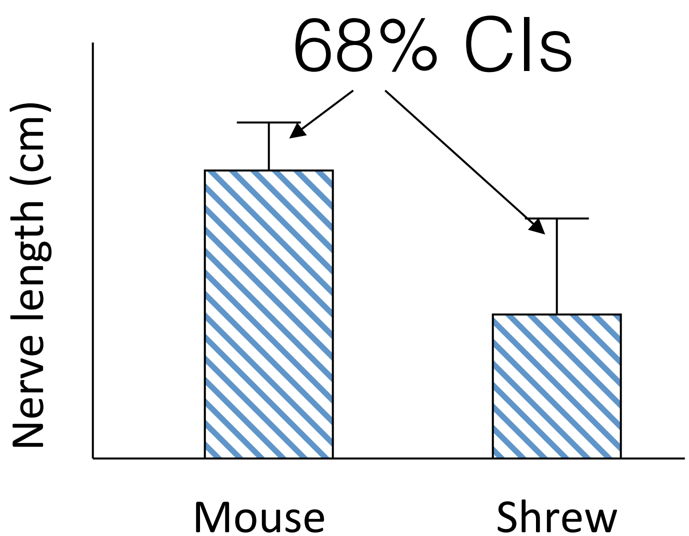


For a 68% CI, $\hat{\mu} \pm \frac{\sigma}{\sqrt{n}}$

SEM



CIs are related to the standard error of the mean (SEM)



Common probability distributions

Commonly arise in nature

Finding what distribution matches your data helps characterize the mechanism underlying your data

Identifying your distribution is import for choosing a statistical test when you are testing a hypothesis

Common discrete probability distributions

Binomial

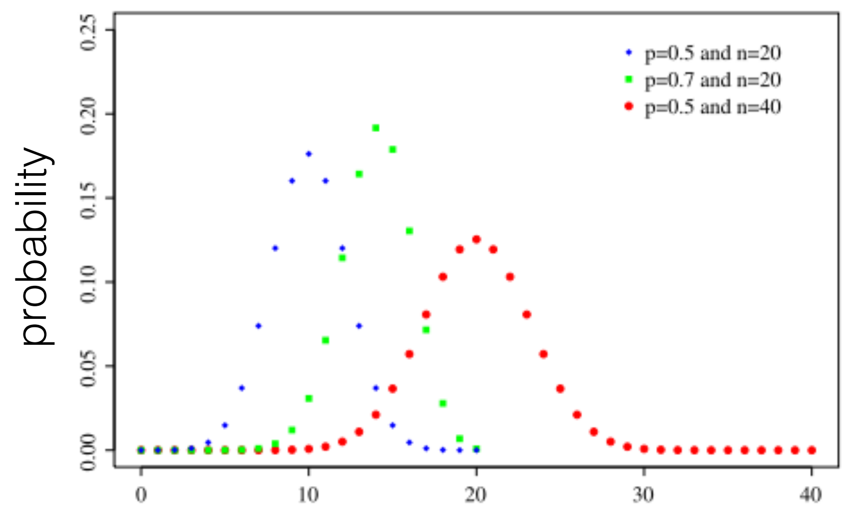
Probability of a certain number of “successes”

Examples:

Probability of getting a certain number of heads, when a coin is tossed 50 times.

Probability of a certain number of children being born with a homozygous mutation when both parents are heterozygous for the mutation and there are 3 children.

Probability that an animal chooses one of two choices a certain number of times, given 20 trials.



Poisson

Probability of a certain number of events happening in a fixed period

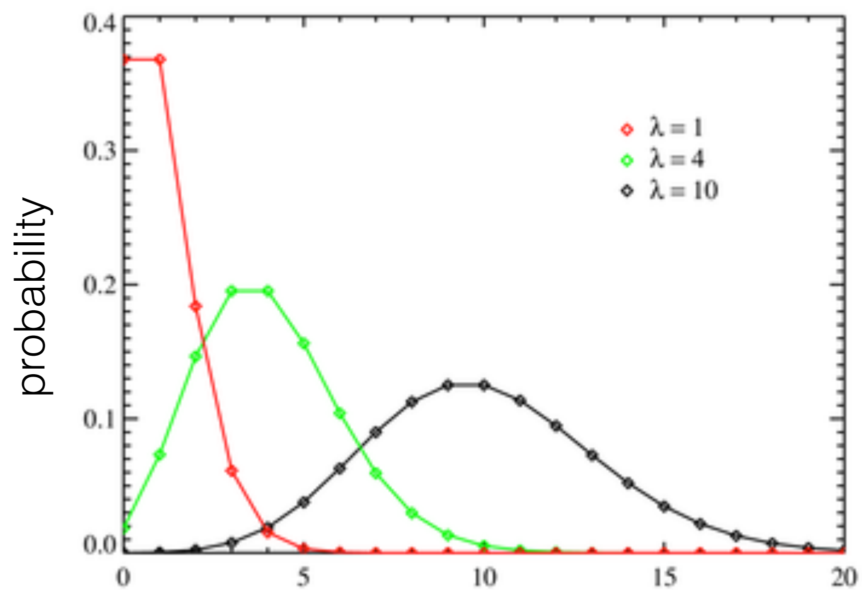
Poisson is the limit of the binomial distribution for large *n* and small *p*

Examples:

Spike rates

Number of Marguerite shuttles that arrive at a bus stop

Number of particles that hit a detector



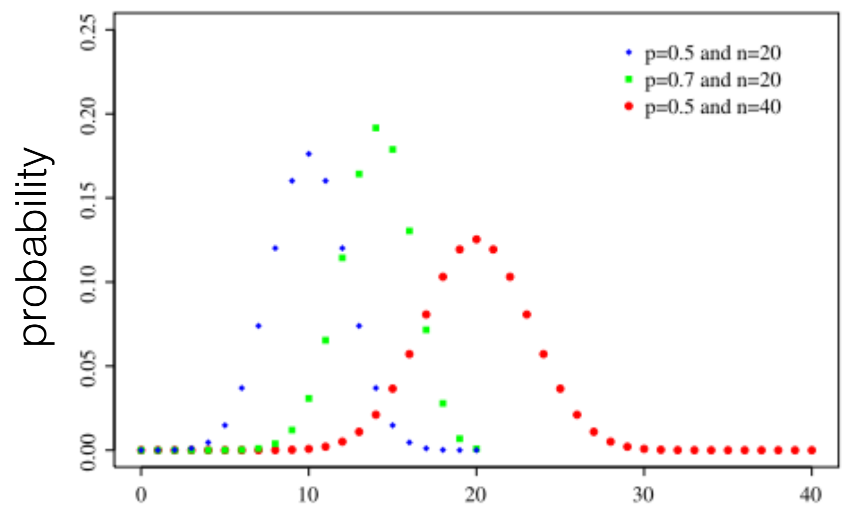
Common discrete probability distributions

Binomial

Probability of a certain number of "successes"

mean = np

variance = $np(1-p)$



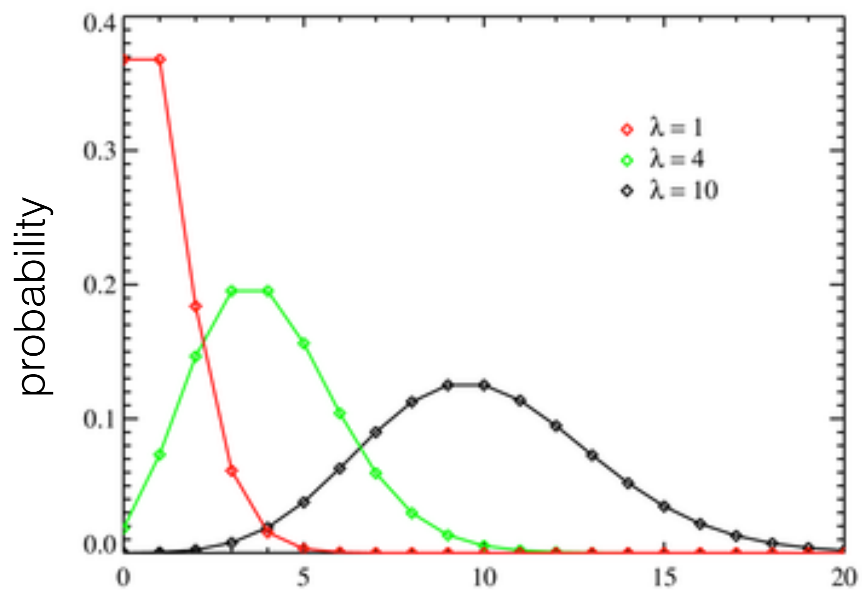
Poisson

Probability of a certain number of events happening in a fixed period

Poisson is the limit of the binomial distribution for large *n* and small *p*

mean = np

variance = np



Common continuous probability distributions

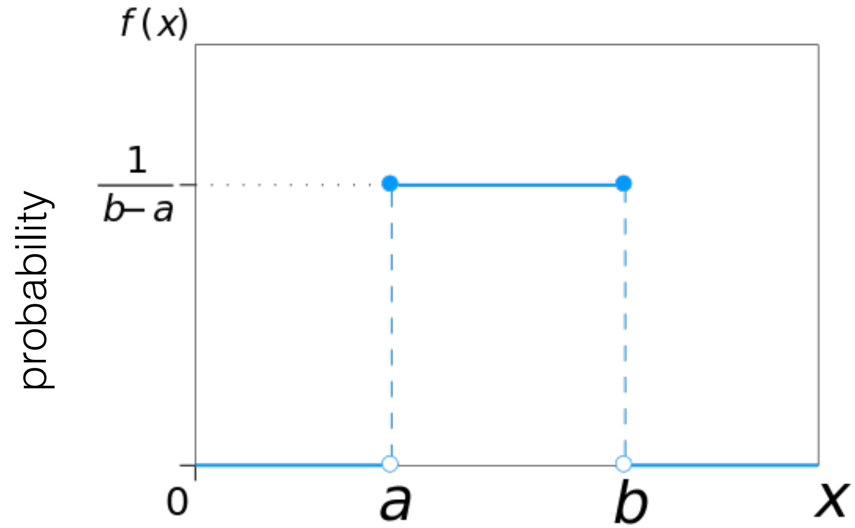
Uniform

Equal probability everywhere an event is possible

Examples:

Probability of a meteor hitting any given spot on Earth

Probability of an ion channel being located on a particular part of the membrane



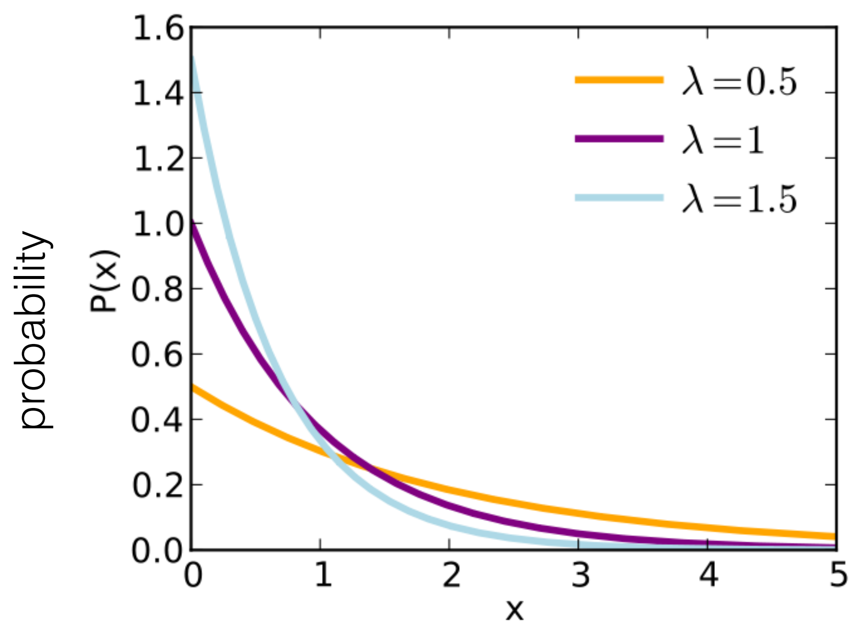
Exponential

Probability of the time between events in a Poisson process

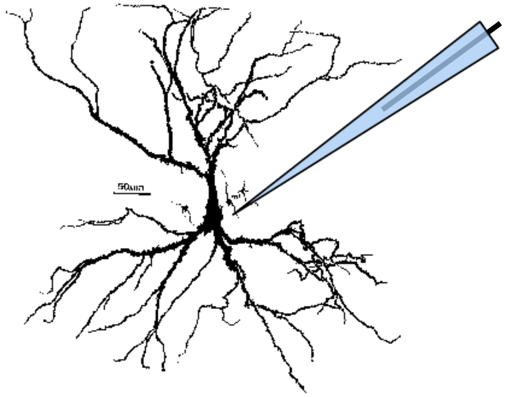
Examples:

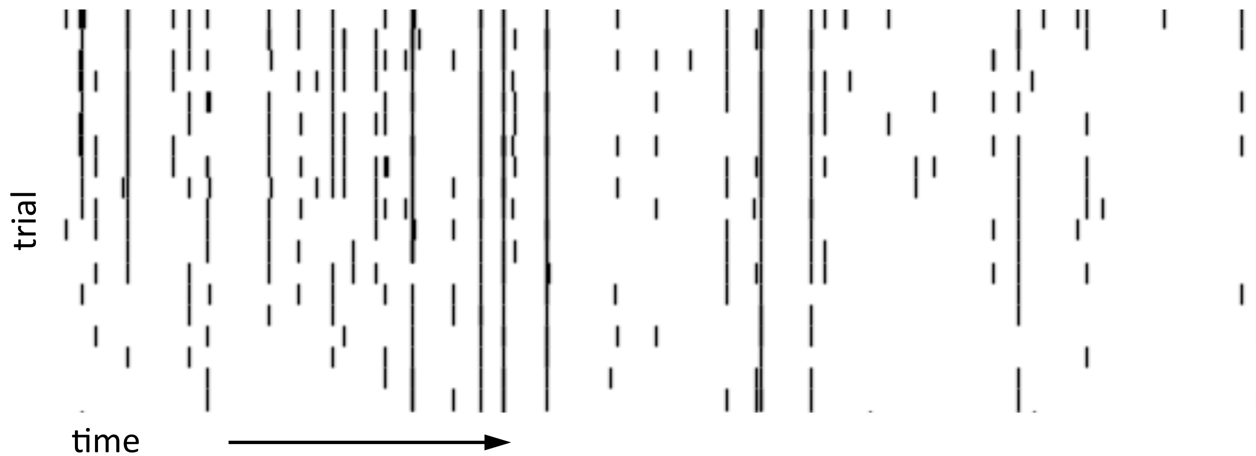
Time between buses

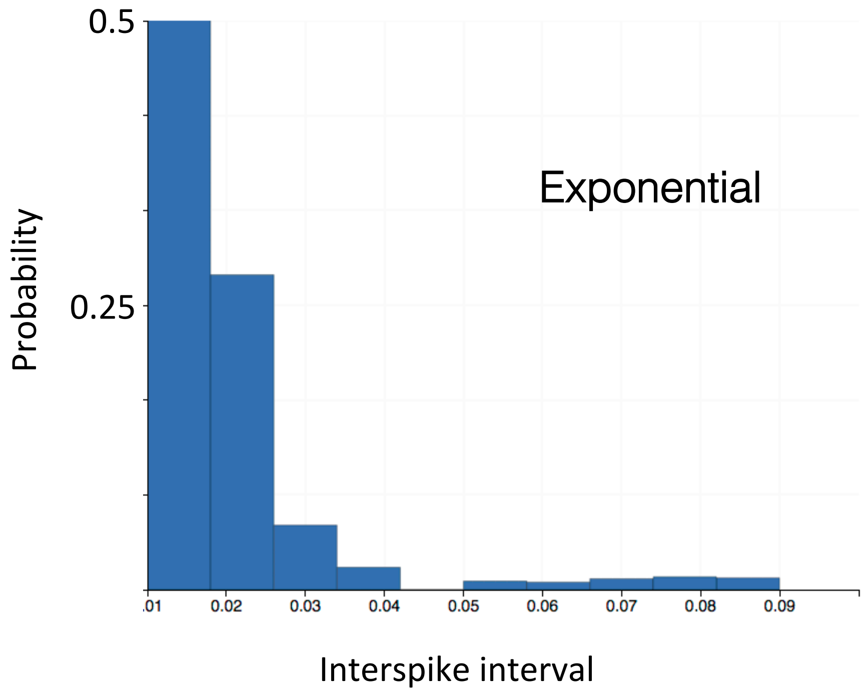
Time between spikes



Example 2: Interspike intervals







Today’s lecture

Data statistics

- mean, variance, stdev, std. err
- common distributions
- confidence intervals

Hypothesis testing

- Parametric tests
- When to use different parametric tests
- Nonparametric tests