DeepLearning.AI

Probability and Statistics for
Machine Learning and Data Science

-

Week 3: Sampling and Point Estimates

W3 Lesson 1

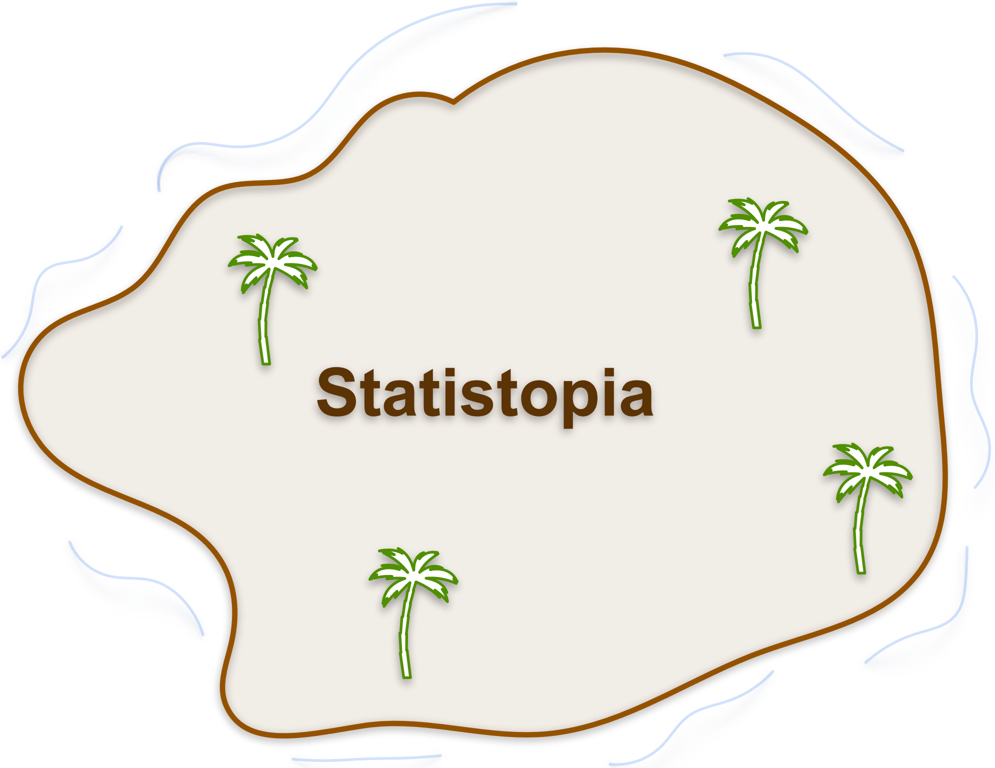
DeepLearning.AI

Sample and Population

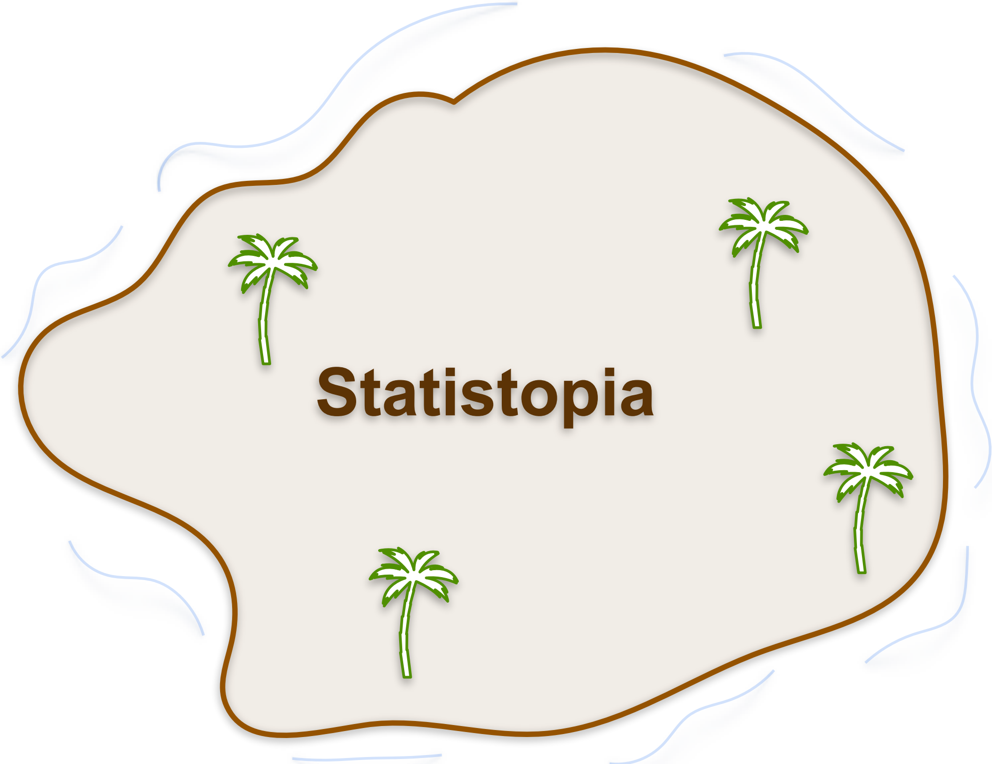
Population and Sample

Population and Sample

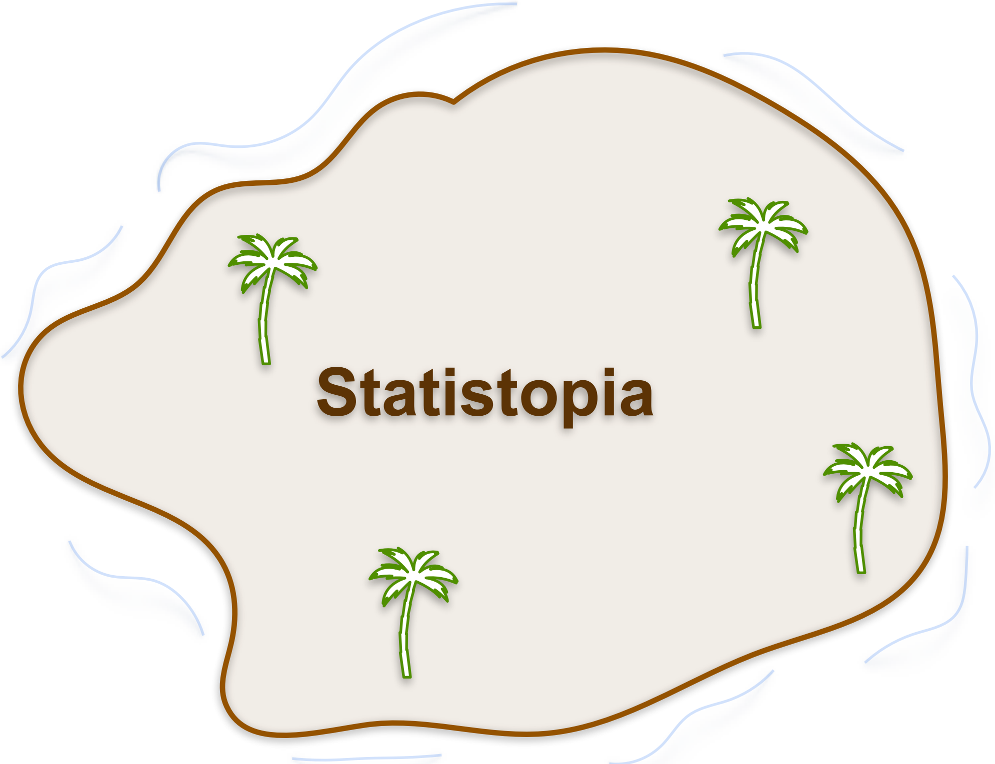
Population and Sample



Population and Sample



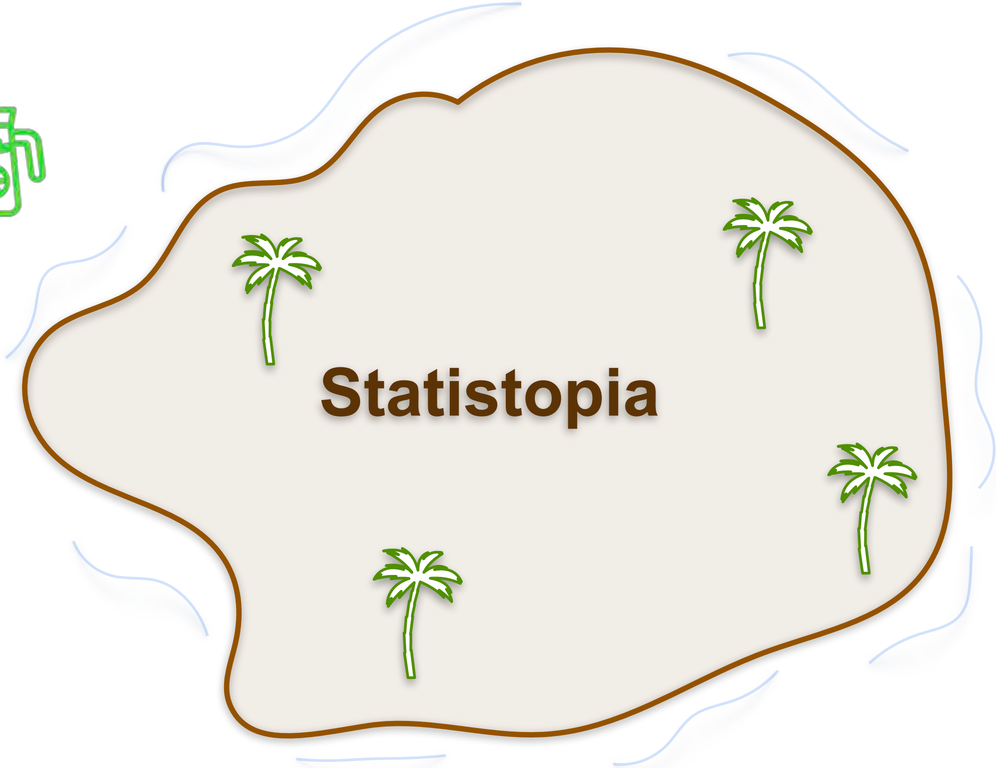
Population and Sample



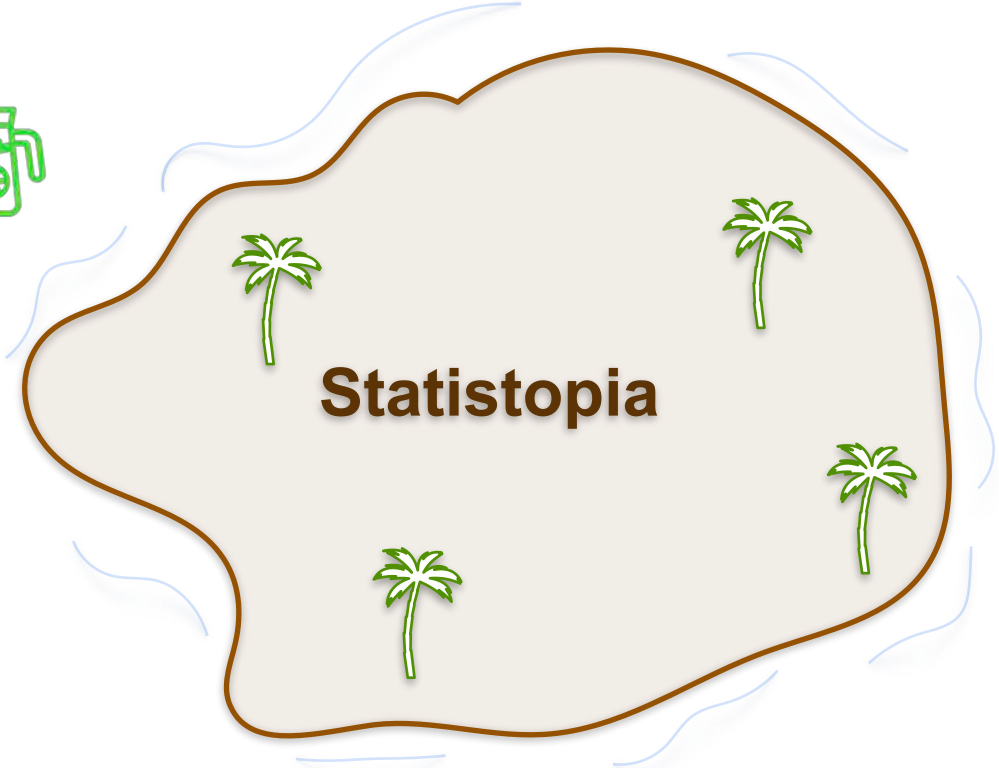
Find the **average height** of the people living on Statistopia

Population and Sample

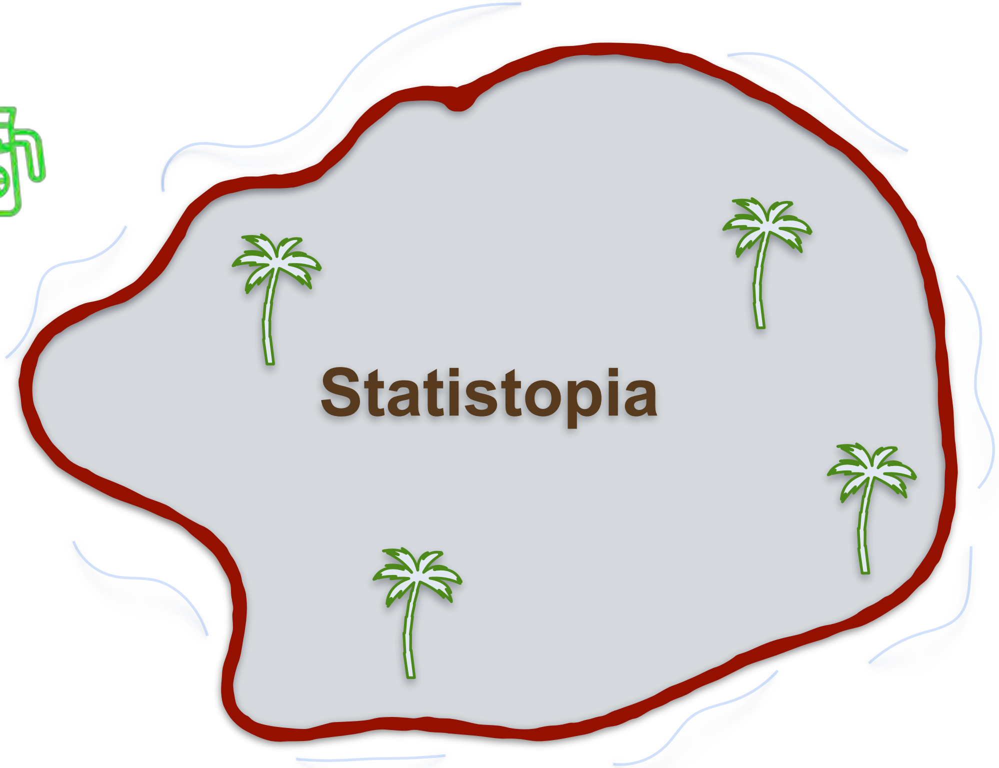
Find the **average height** of the people living on Statistopia



Population and Sample



Population and Sample

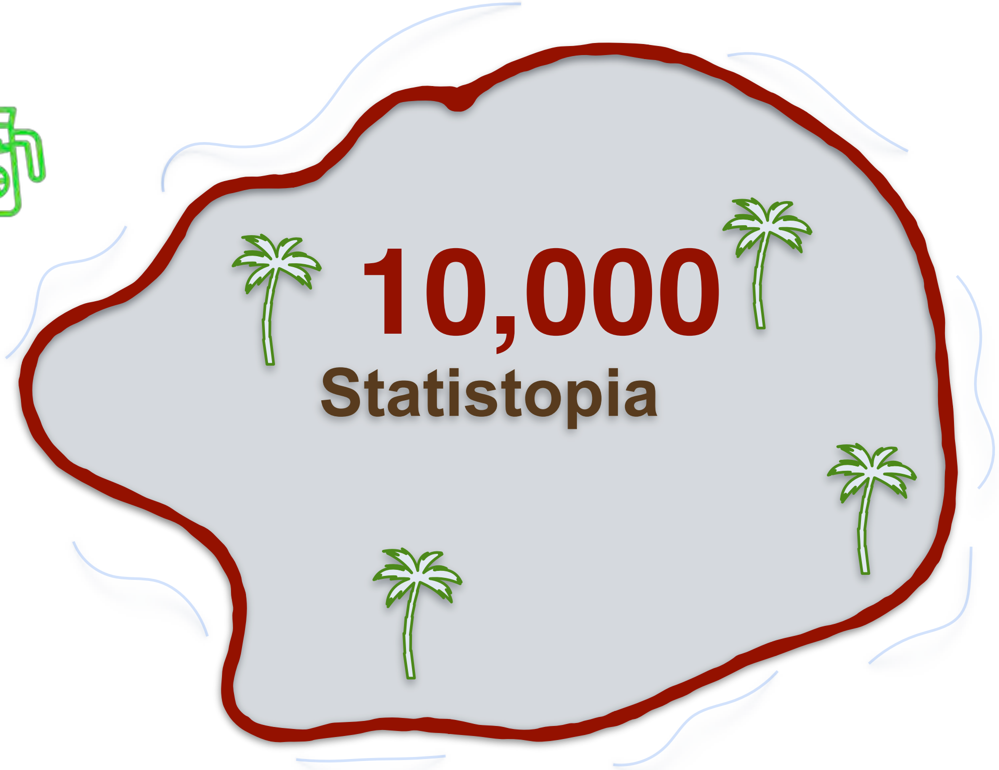


- Ask everyone on the island for their height.

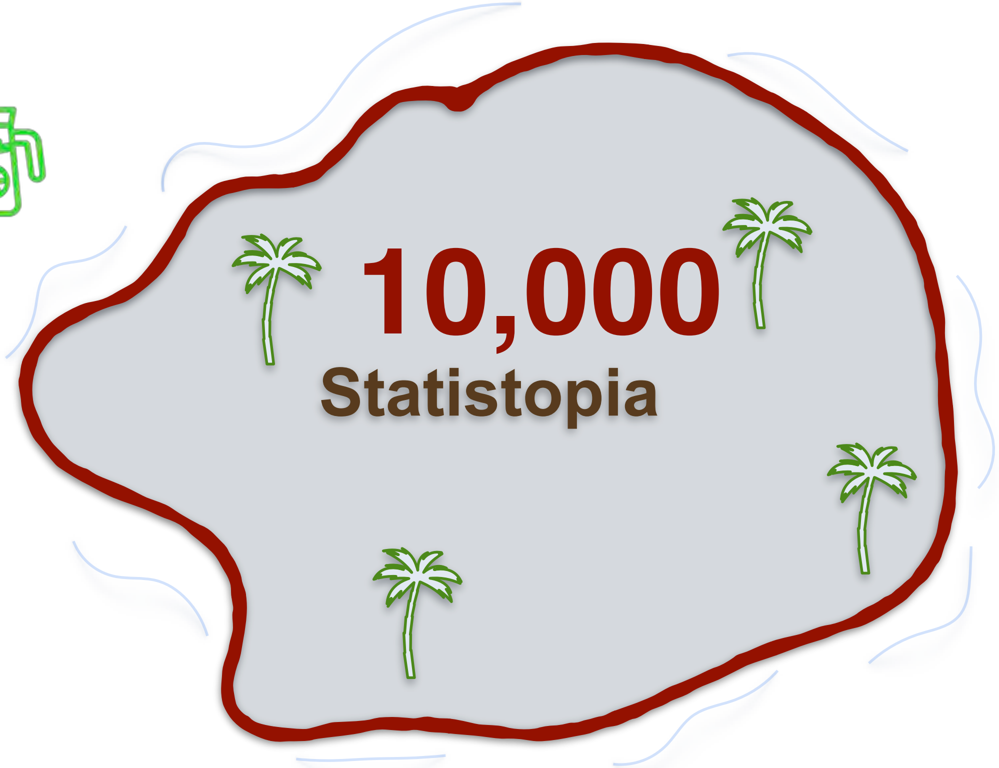
- Divide by the total number

Population and Sample

- Ask everyone on the island for their height.
- Divide by the total number

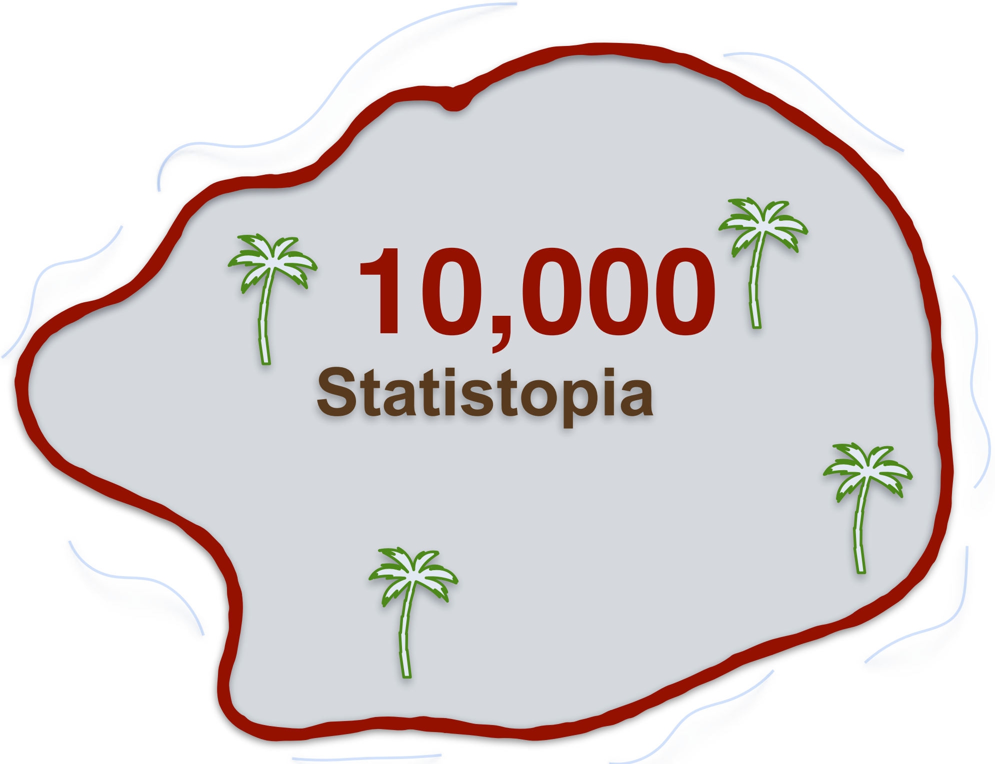


Population and Sample



- Ask everyone on the island for their height.
- Divide by the total number

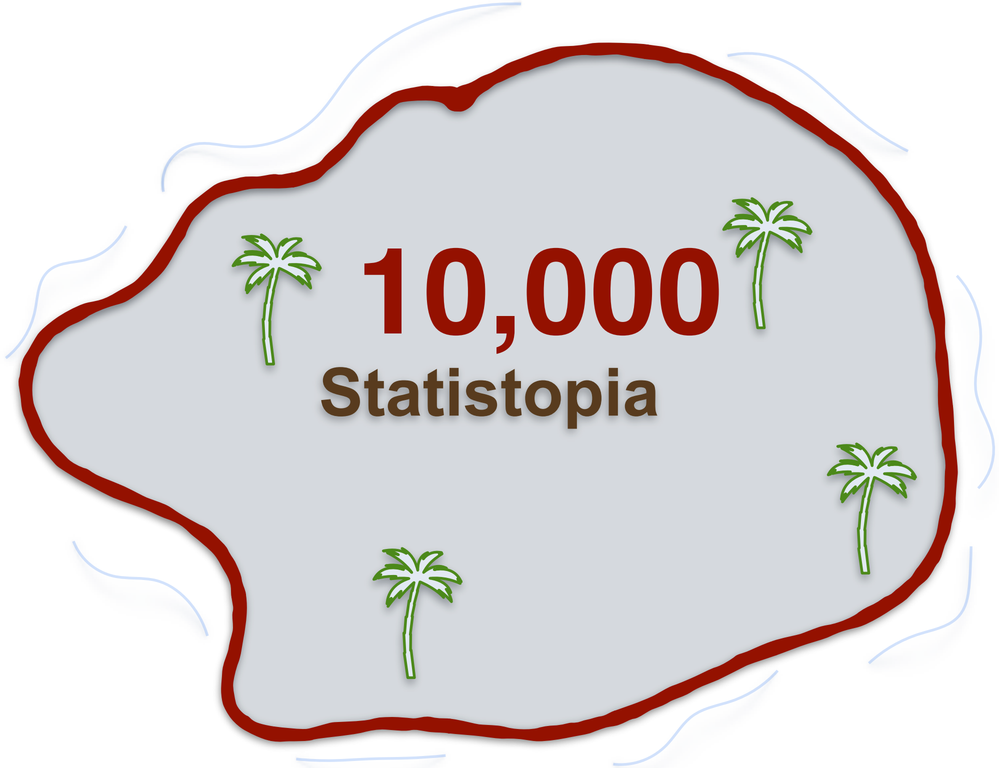
Population and Sample



- Ask everyone on the island for their height.

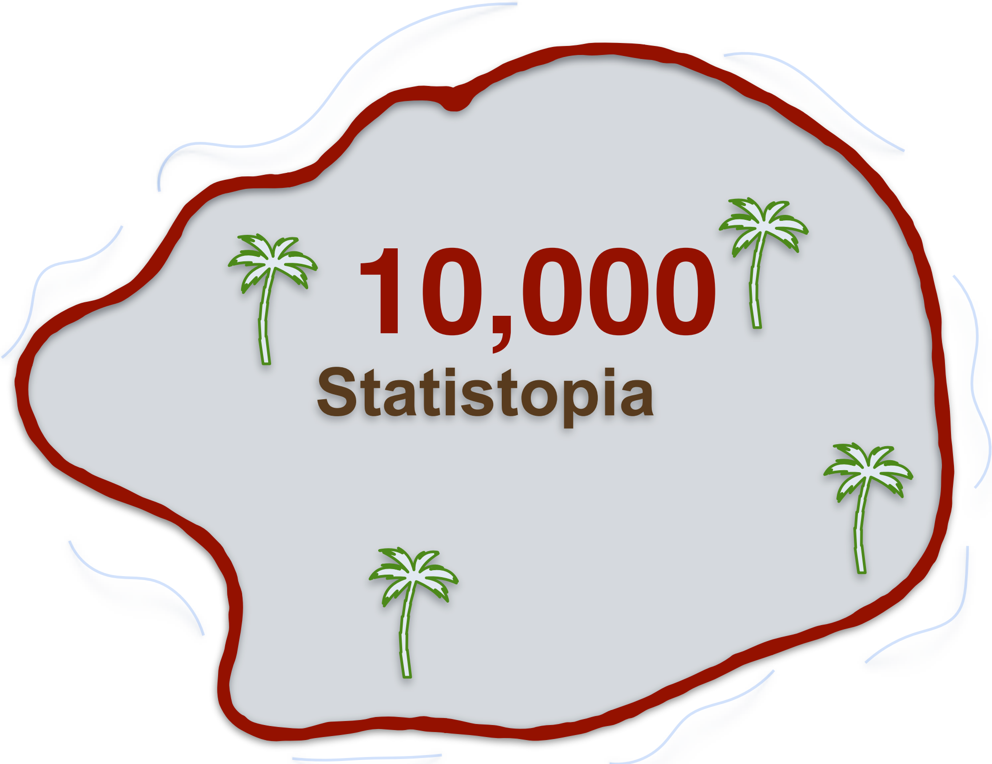
- Divide by the total number

Population and Sample



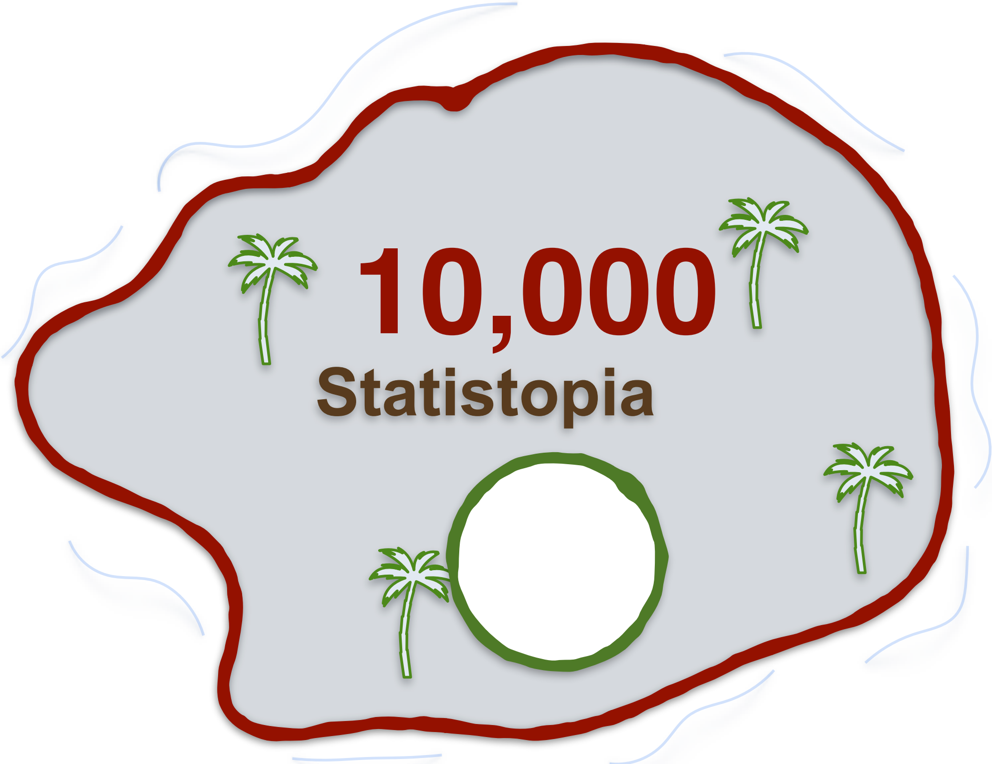
Population and Sample

- Only ask a subset of the group to estimate the average height

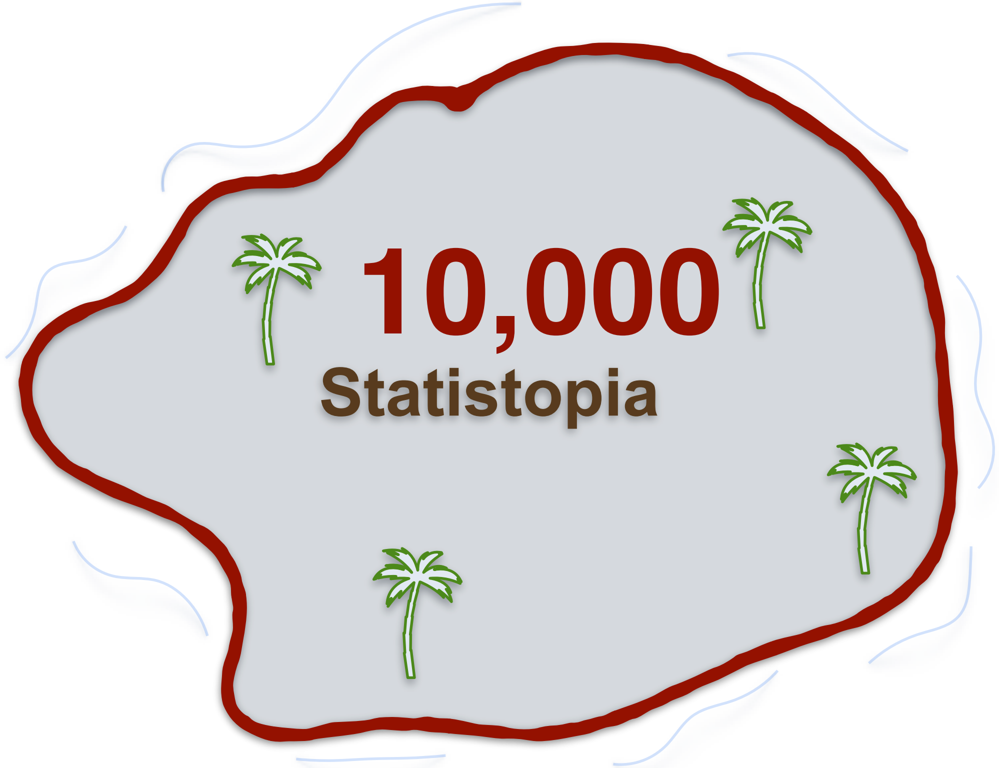


Population and Sample

- Only ask a subset of the group to estimate the average height

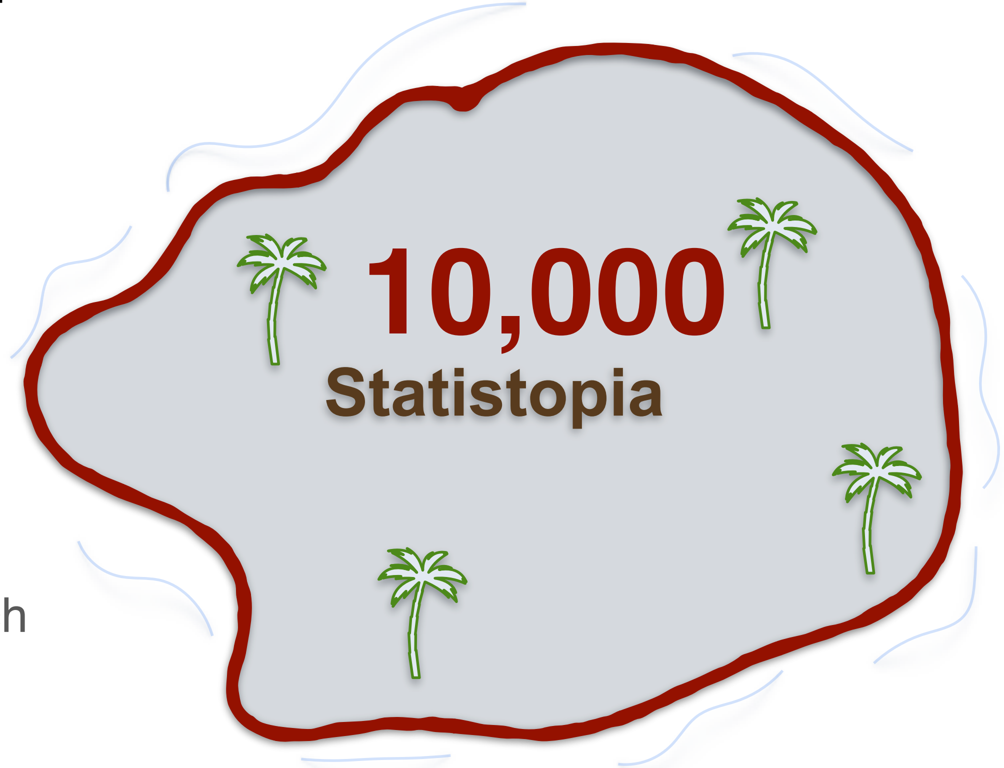


Population and Sample



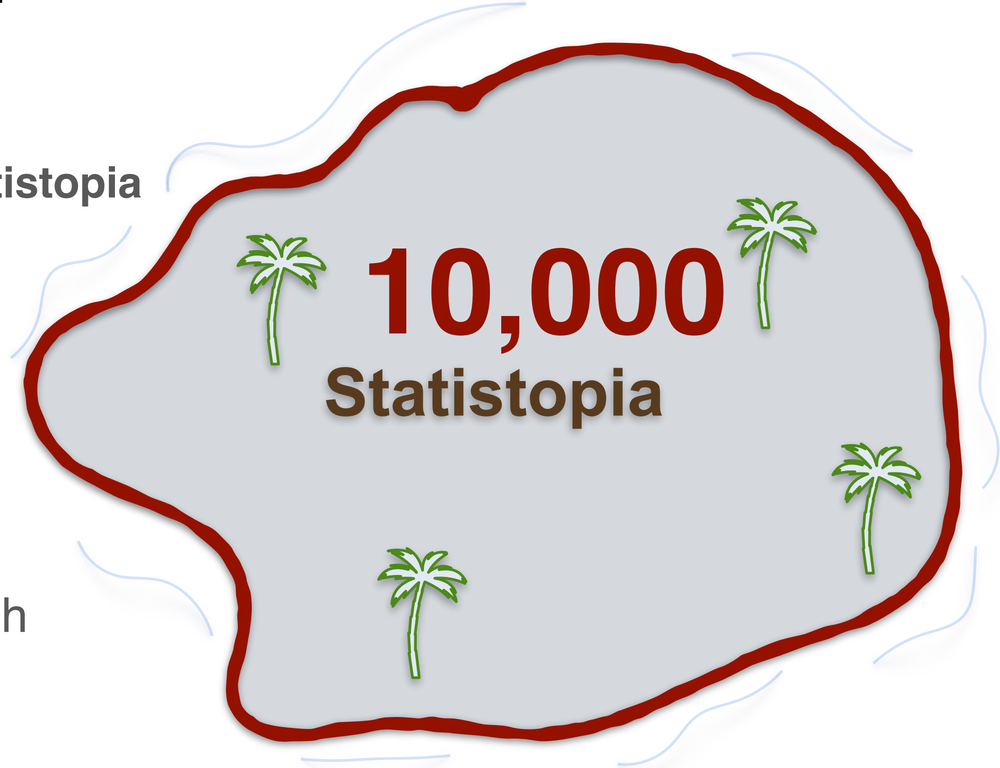
Population and Sample

Population:
the entire group of individuals or elements you want to study which share a common behaviour



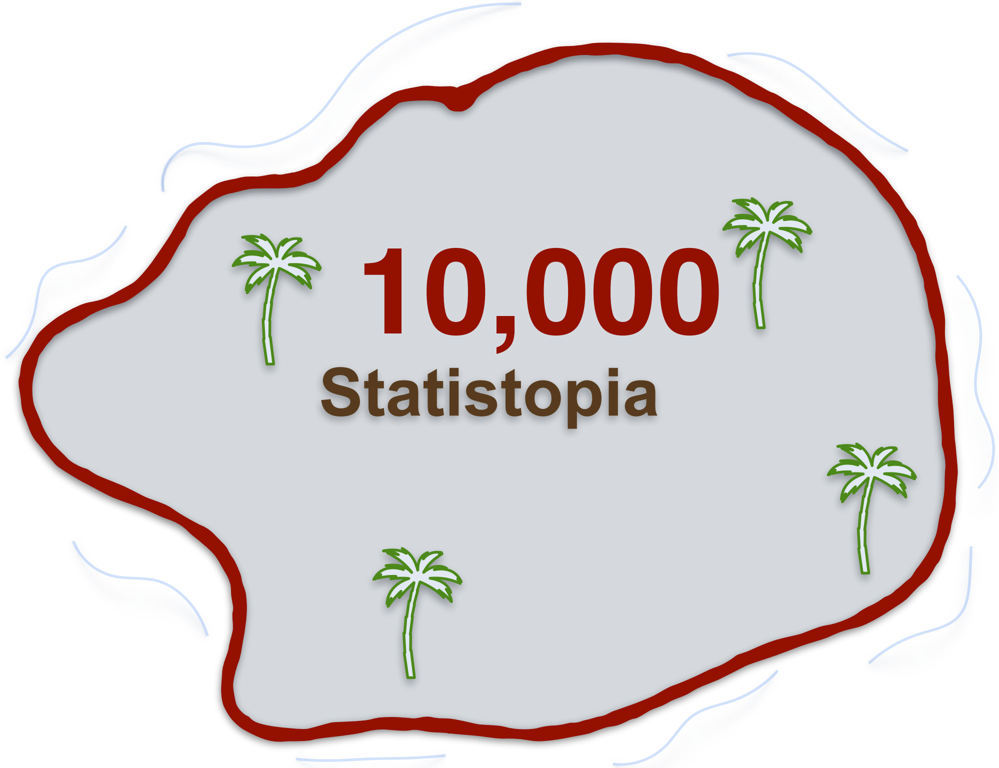
Population and Sample

The people of statistopia



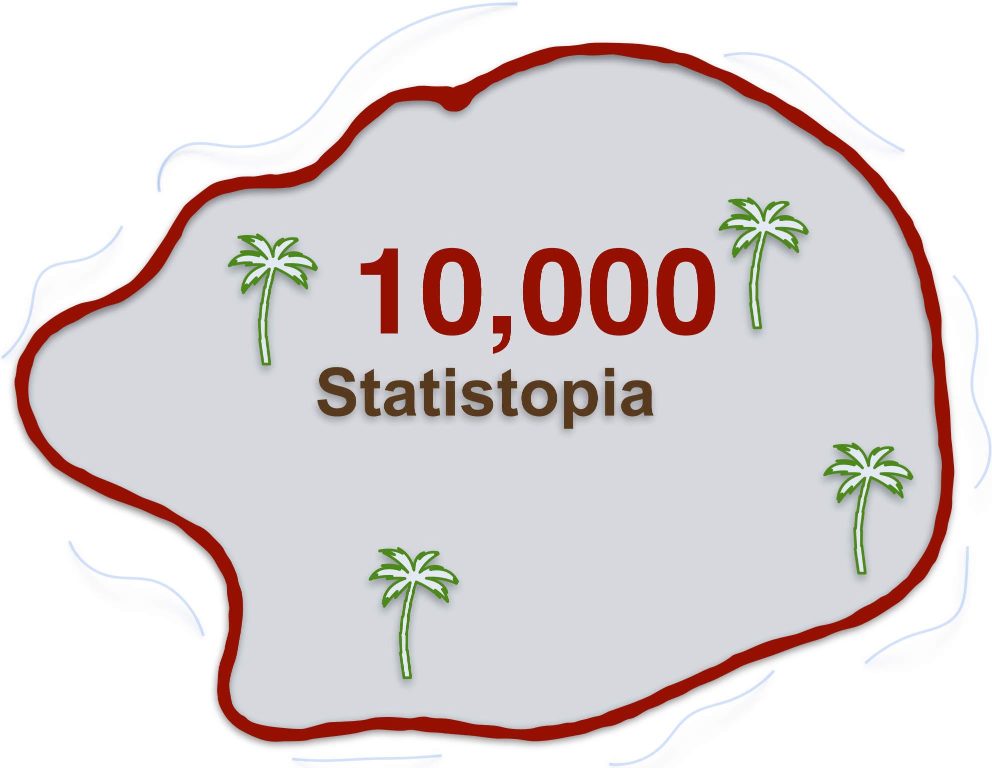
Population:
the entire group of individuals or elements you want to study which share a common behaviour

Population and Sample



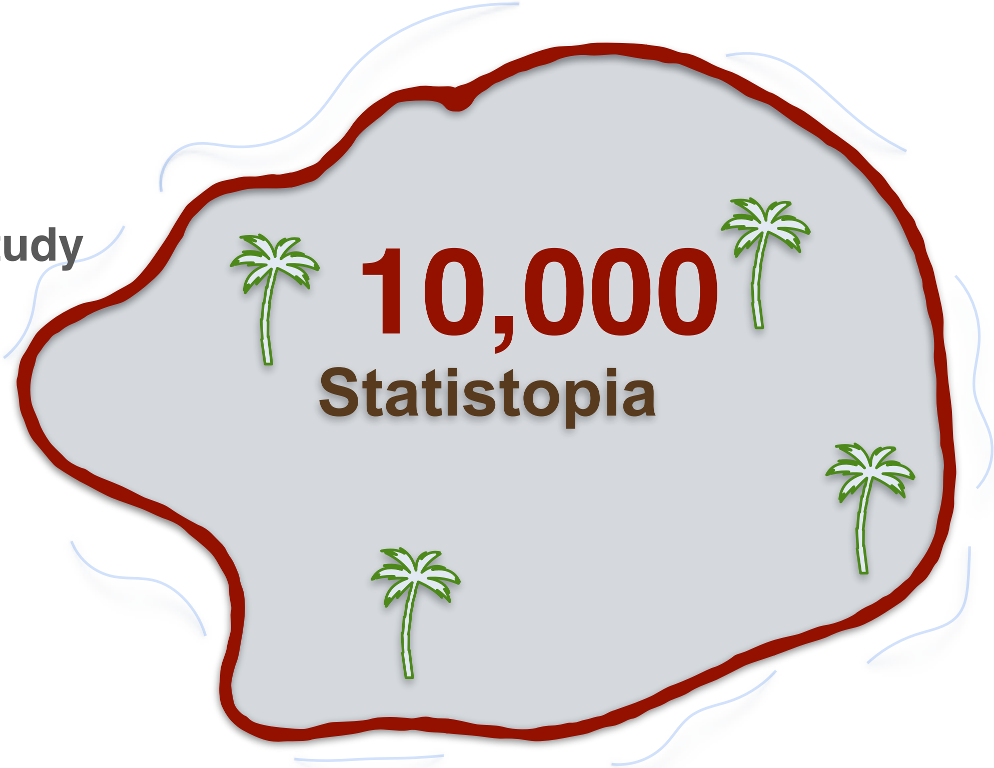
Population and Sample

Sample:
subset of the population you use to draw conclusions about the population as a whole



Population and Sample

The people you select for your study

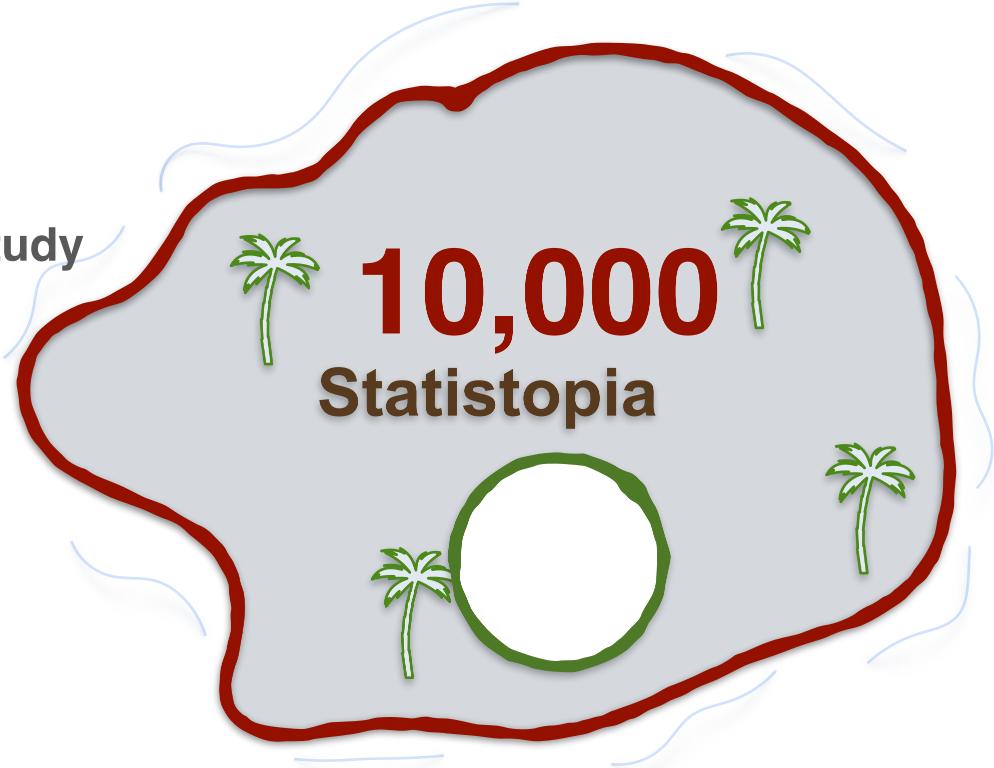


Sample:
subset of the population you use to draw conclusions about the population as a whole

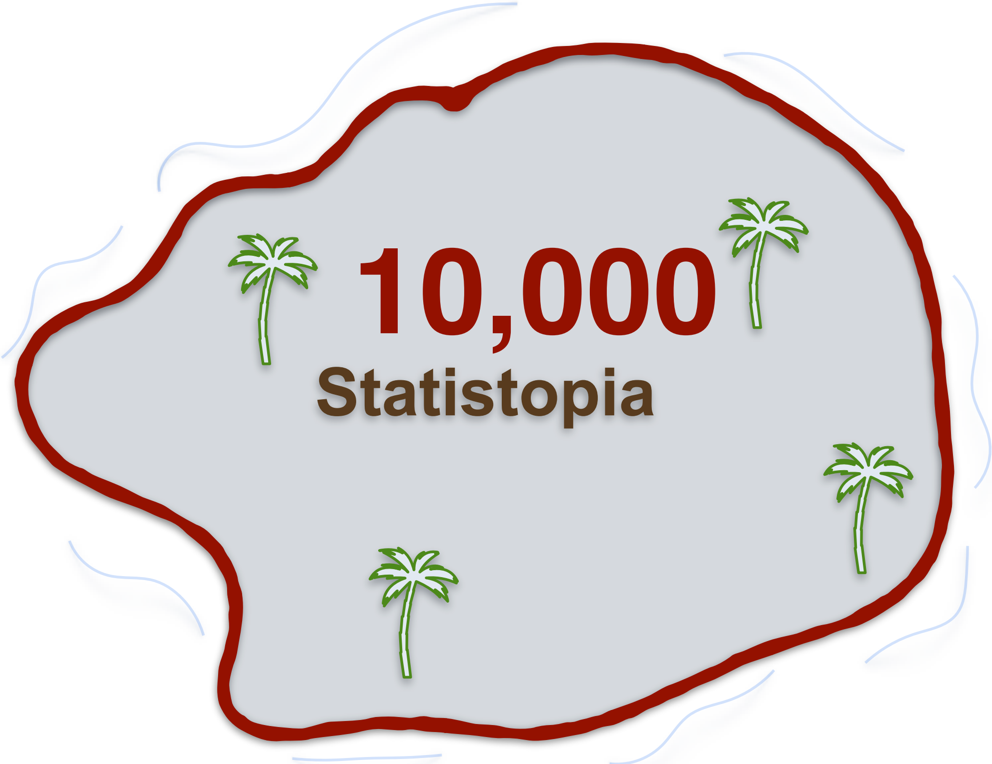
Population and Sample

The people you select for your study

Sample:
subset of the population you use to draw conclusions about the population as a whole



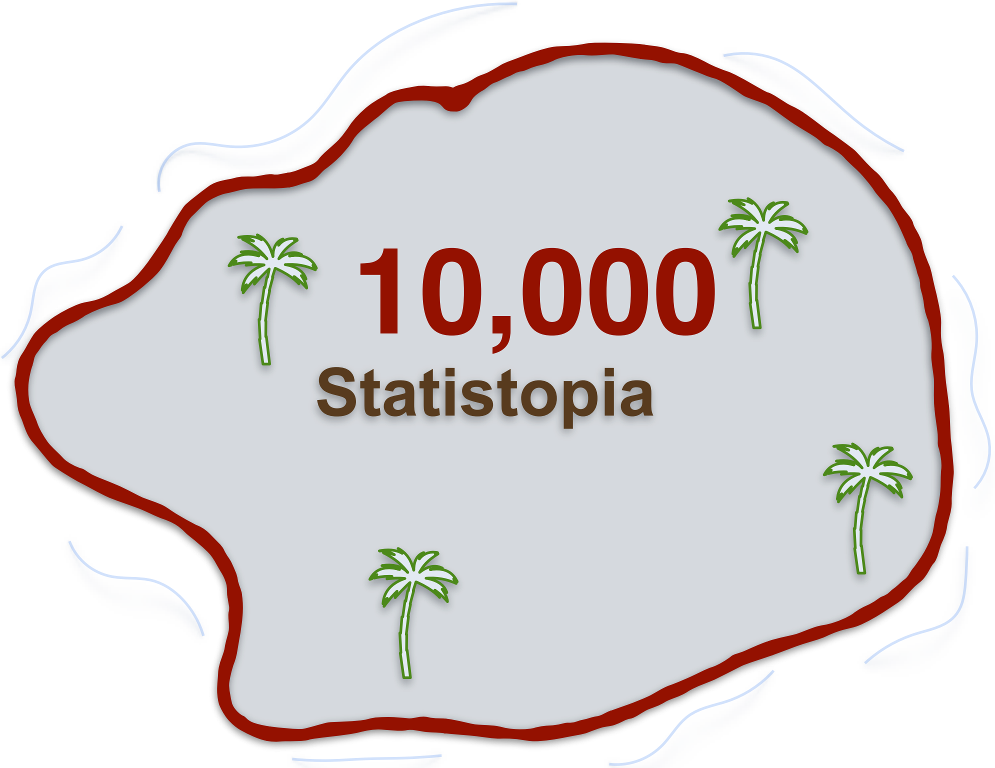
Population and Sample



Population and Sample

Population Size (N)

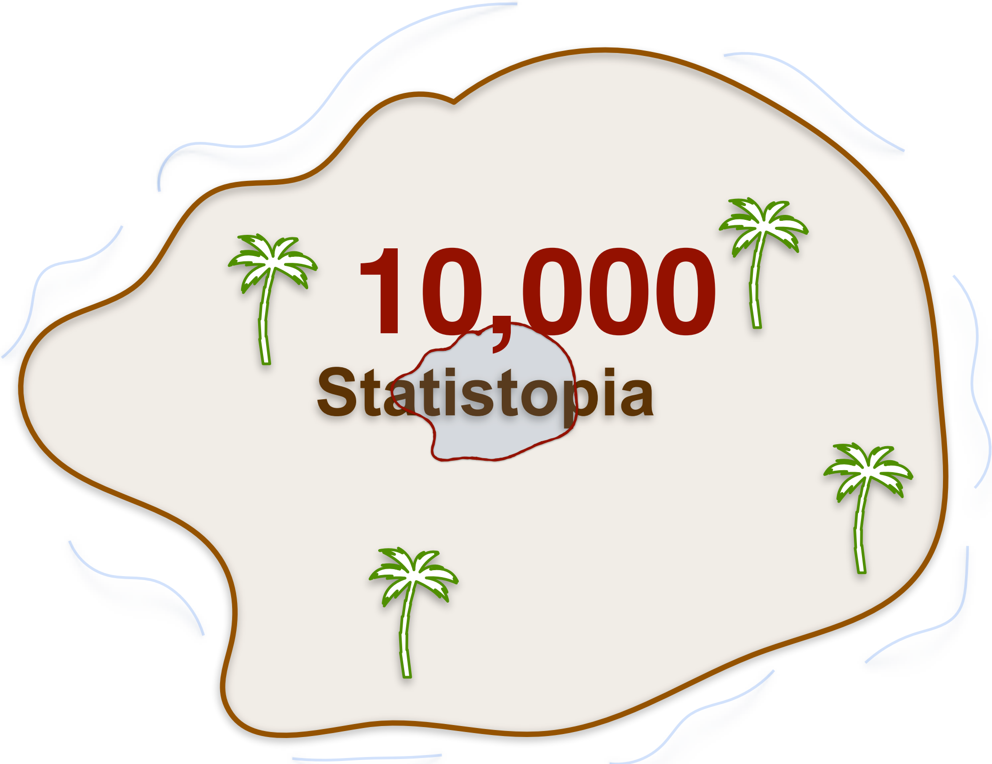
10,000



Population and Sample

Population Size (N)

10,000

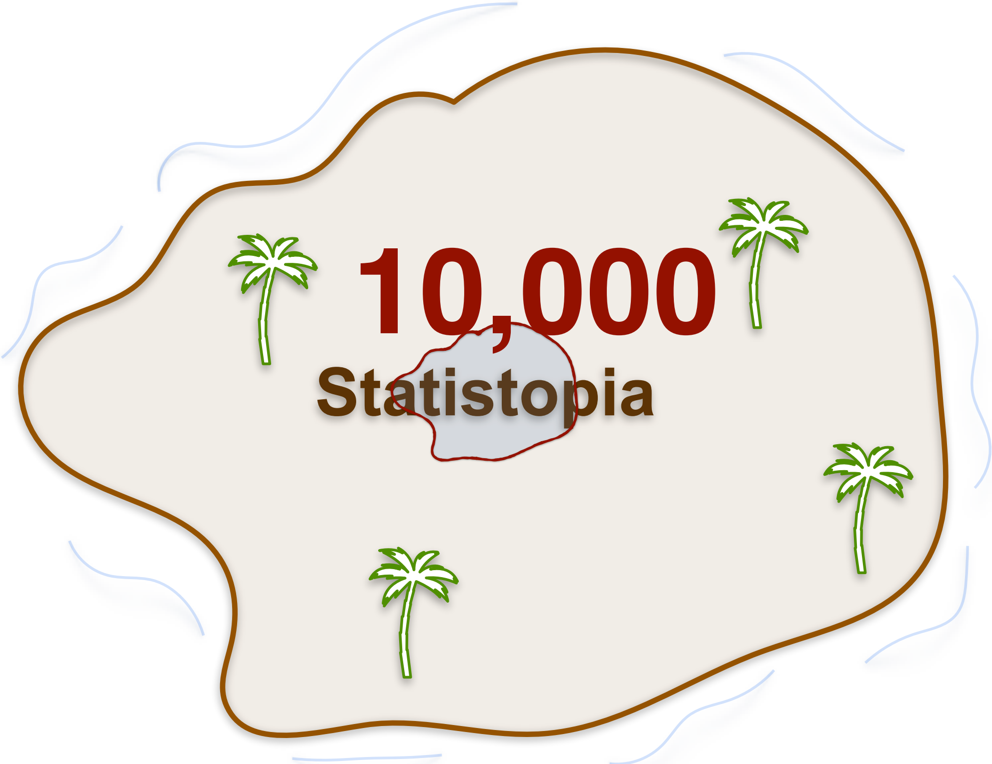


Population and Sample

Population Size (N)

10,000

Sample Size (n)



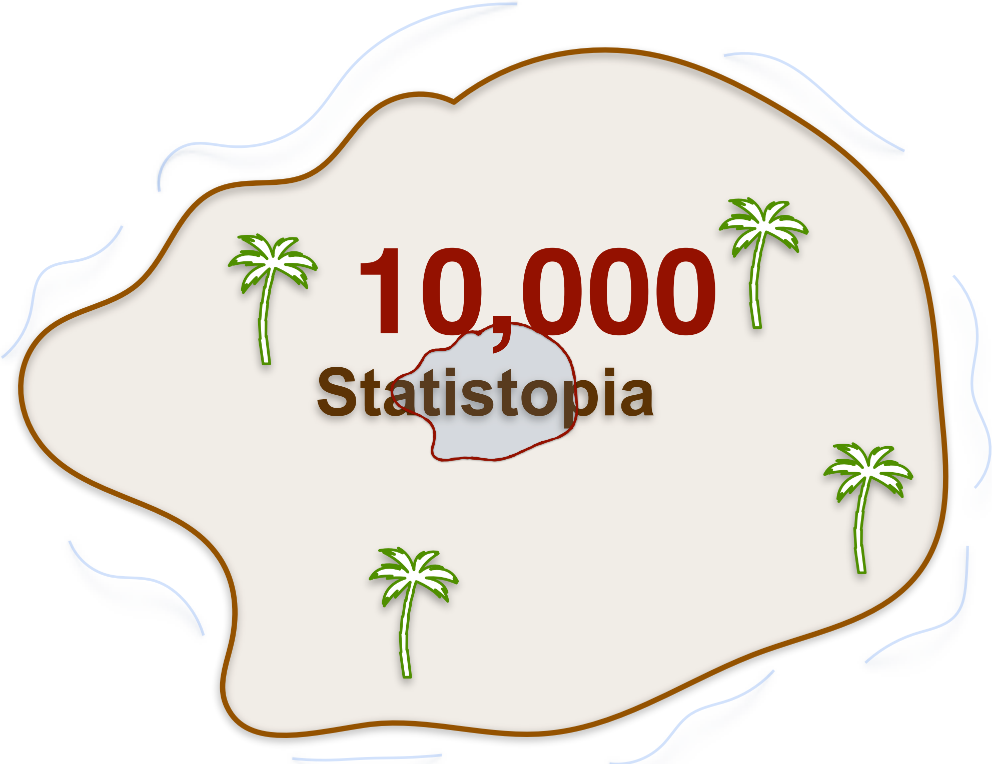
Population and Sample

Population Size (N)

10,000

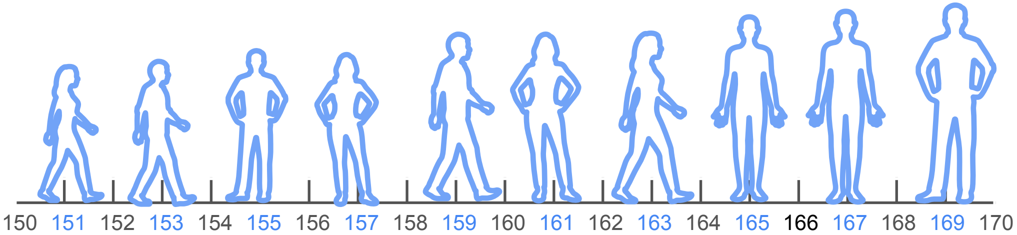
Sample Size (n)

1 - 9,999



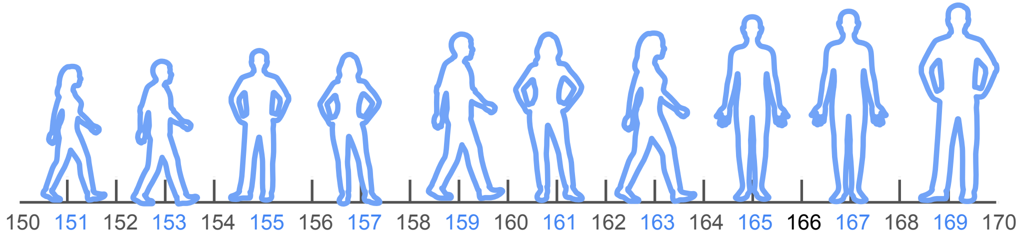
Population and Sample

Population and Sample

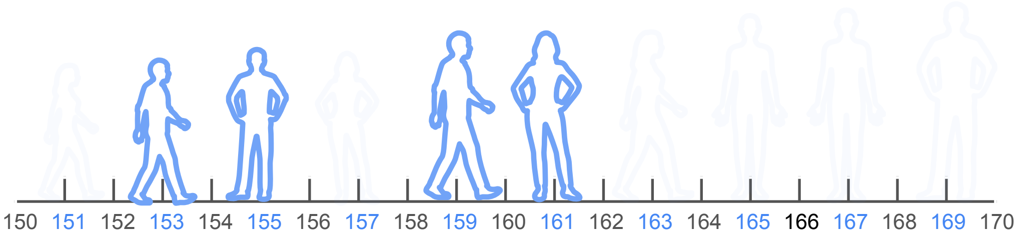


Population and Sample

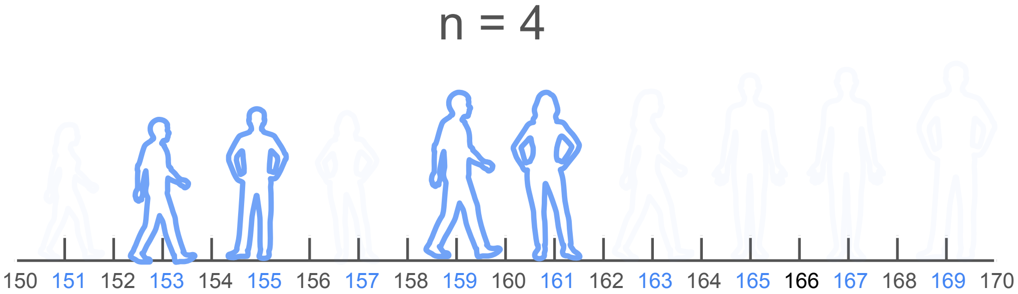
N = 10



Random Sampling

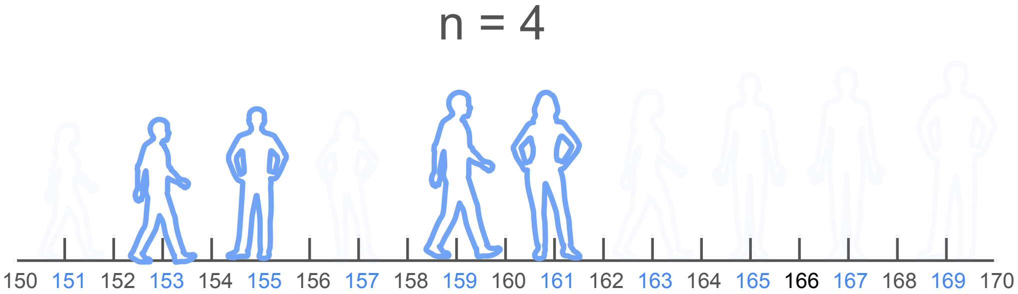


Random Sampling



Random Sampling

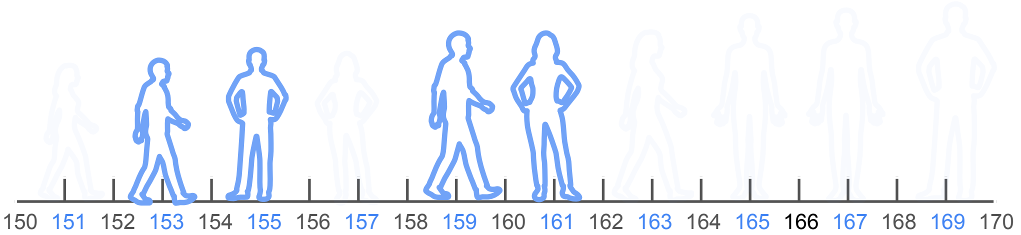
A

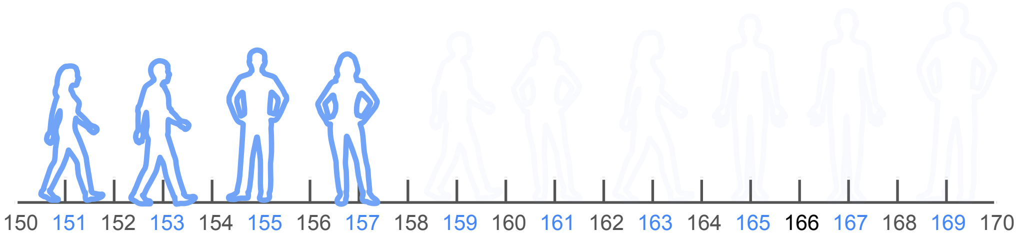


Random Sampling

n = 4

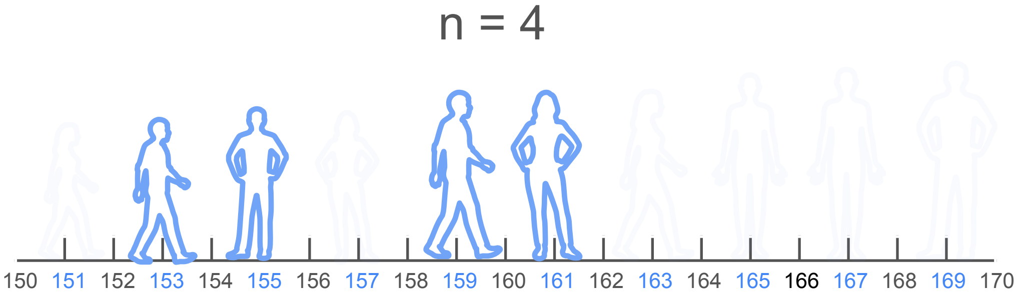
A

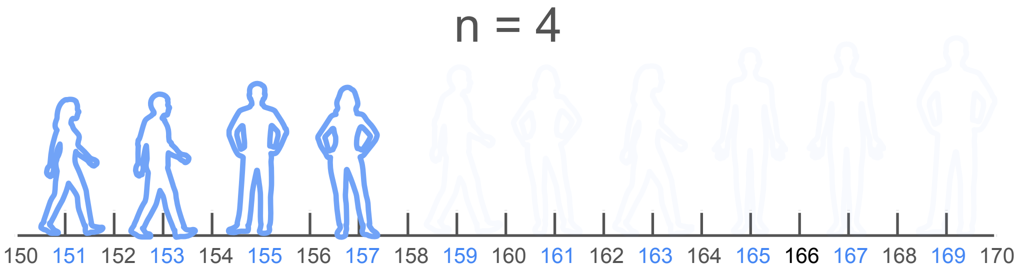




Random Sampling

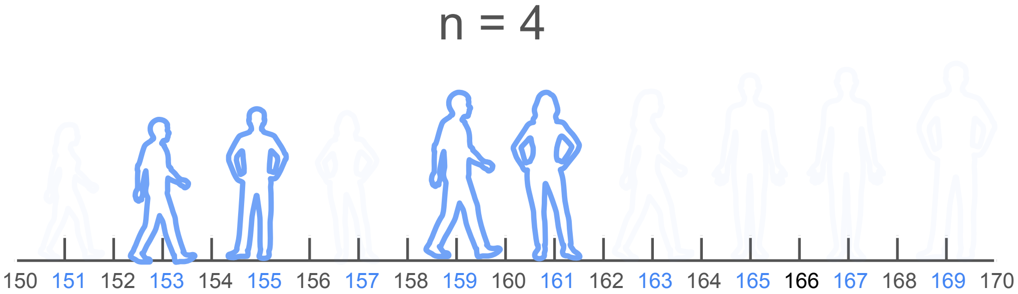
A



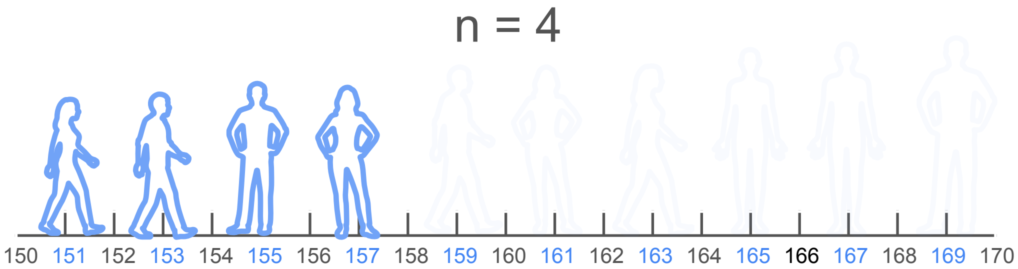


Random Sampling

A

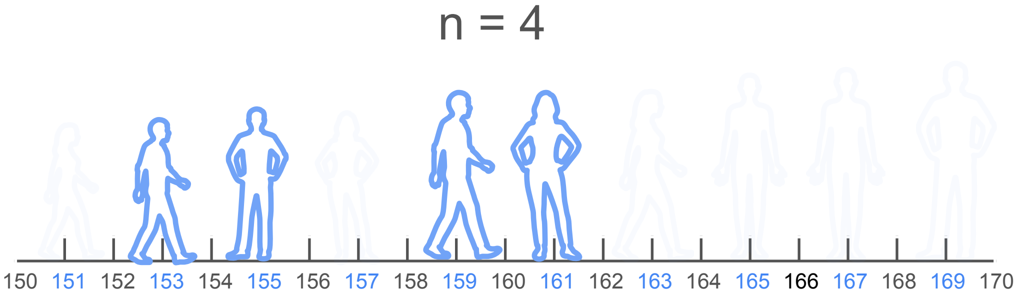


B



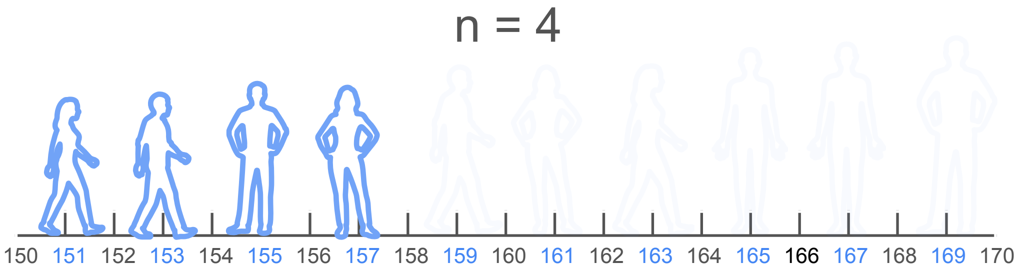
Random Sampling

A



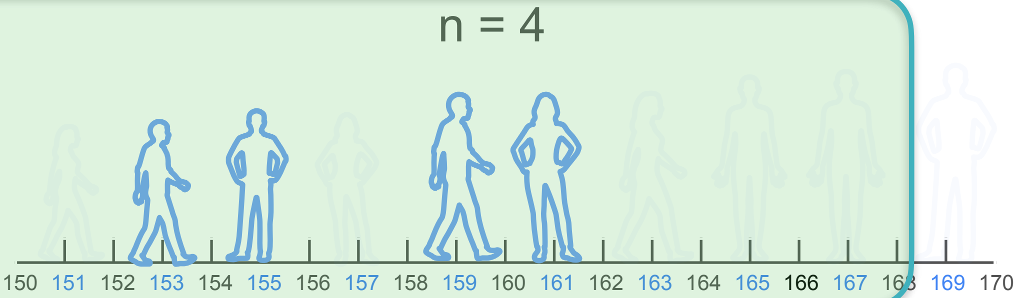
Which is the better sample to estimate the population mean height?

B



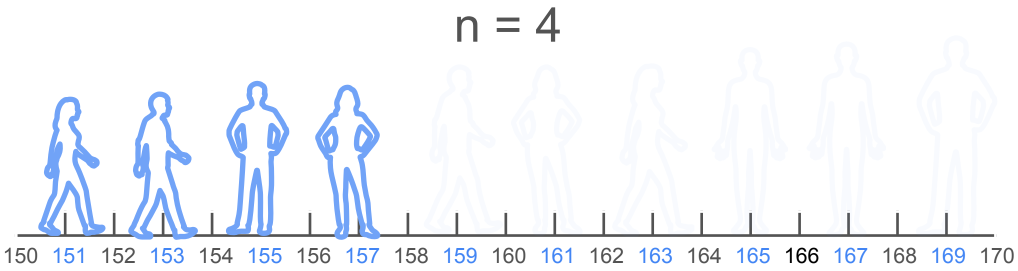
Random Sampling

A



Which is the better sample to estimate the population mean height?

B



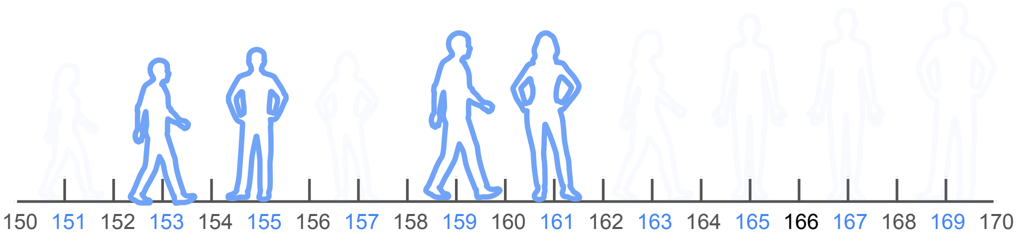
Independent Sample

Independent Sample

Example 1

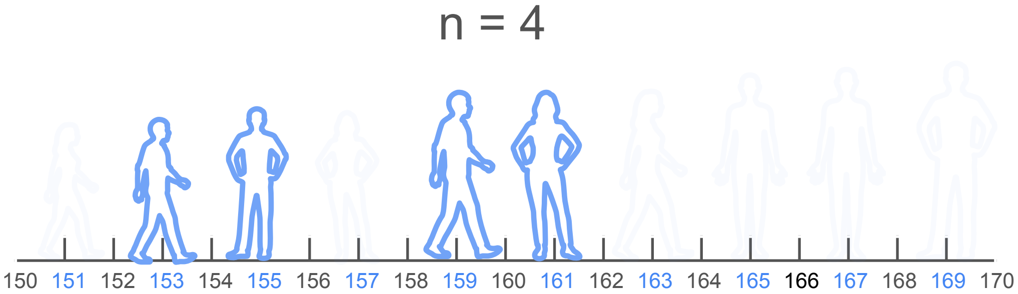
Independent Sample

Example 1



Independent Sample

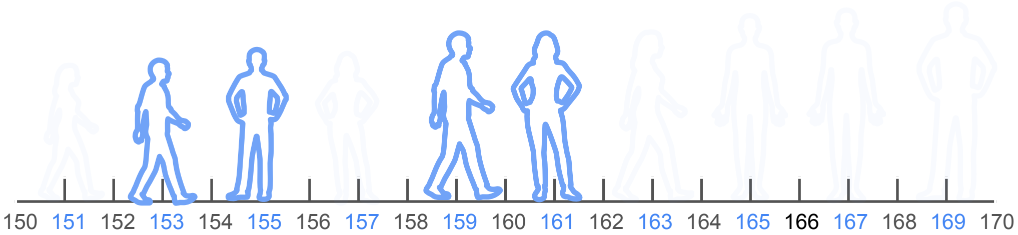
Example 1

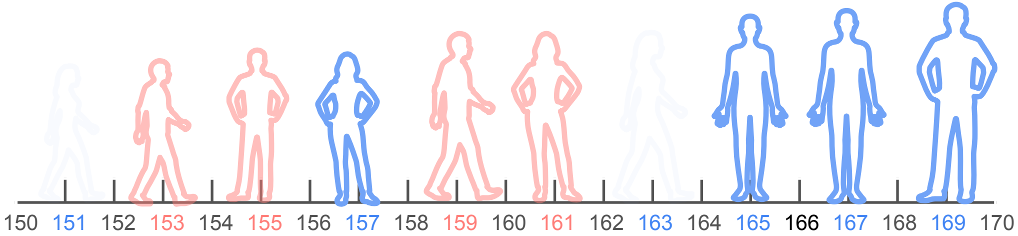


Independent Sample

Example 1

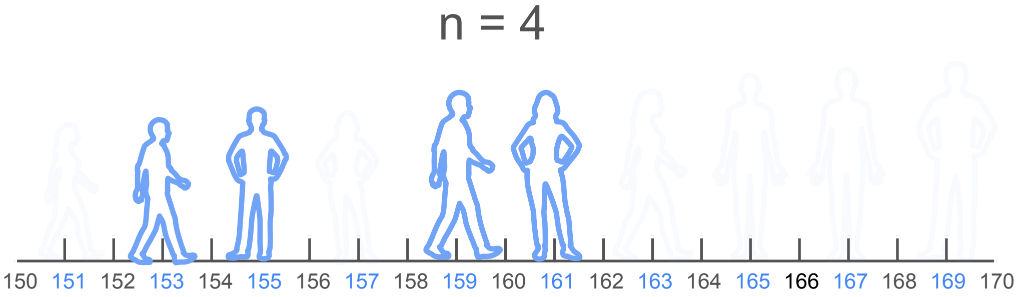
n = 4

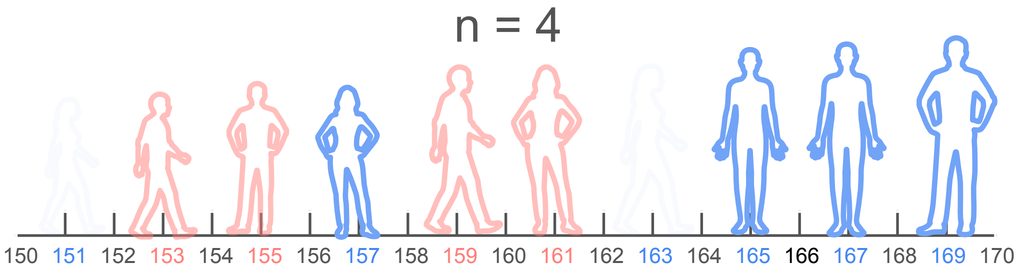




Independent Sample

Example 1

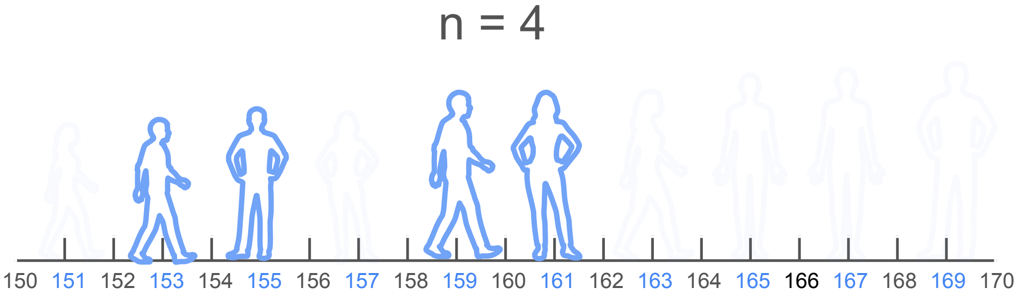


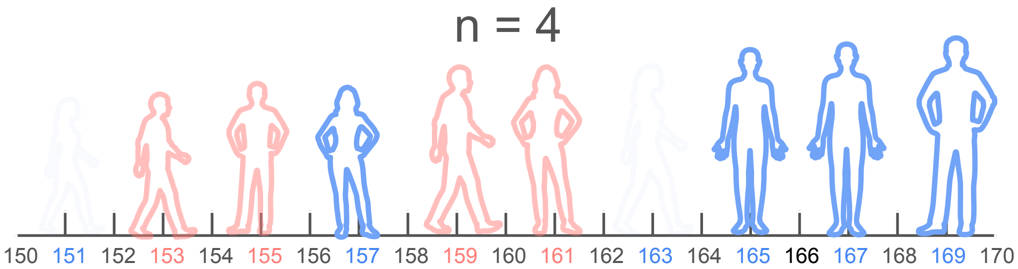


Independent Sample

Example 1

1st sample set

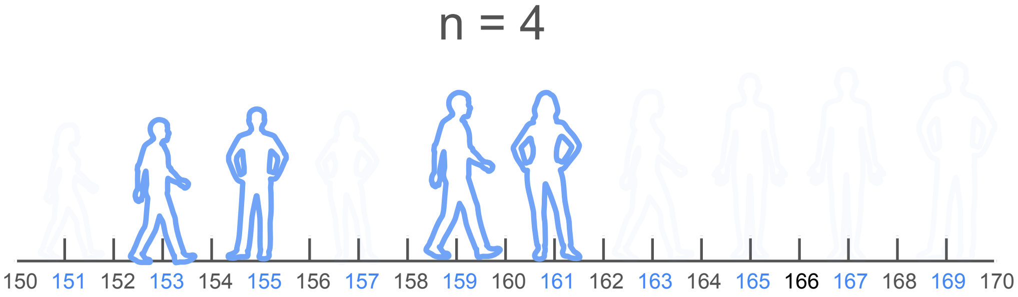




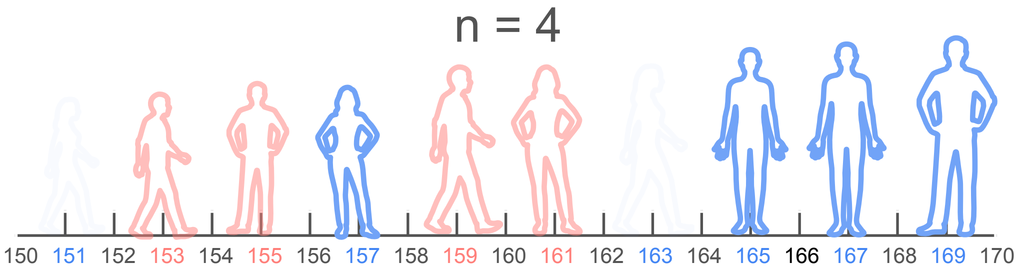
Independent Sample

Example 1

1st sample set



2nd sample set

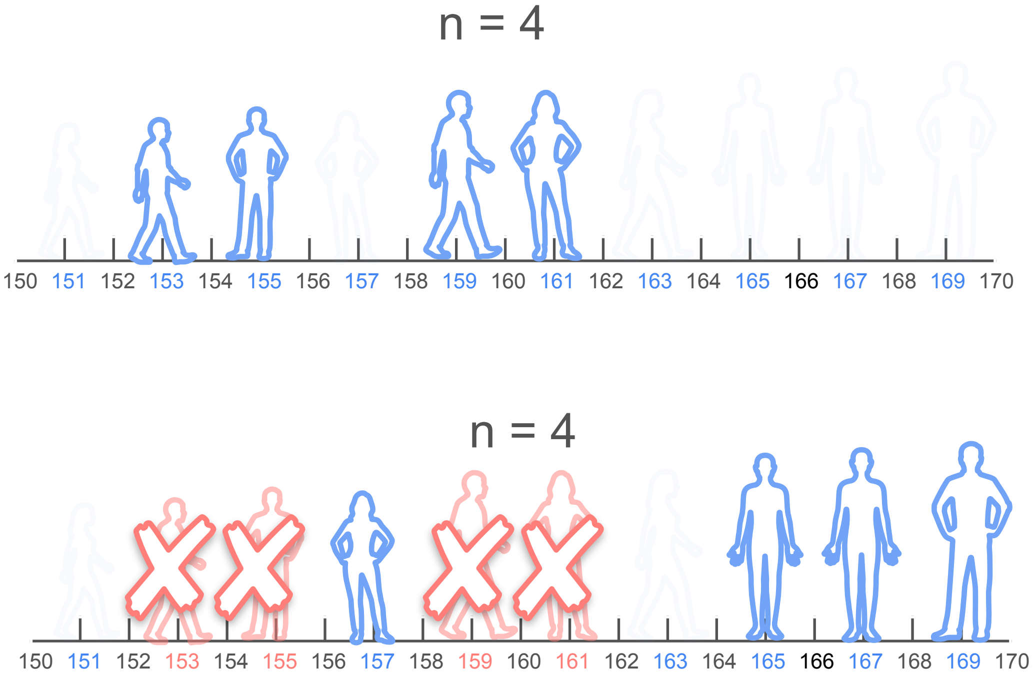


Independent Sample

Example 1

1st sample set

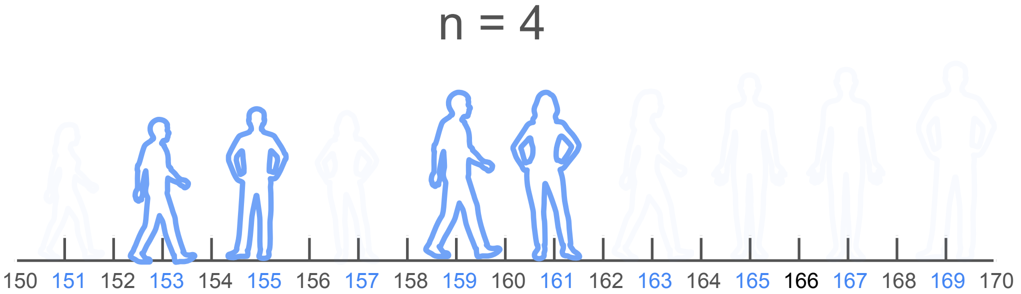
2nd sample set



Independent Sample

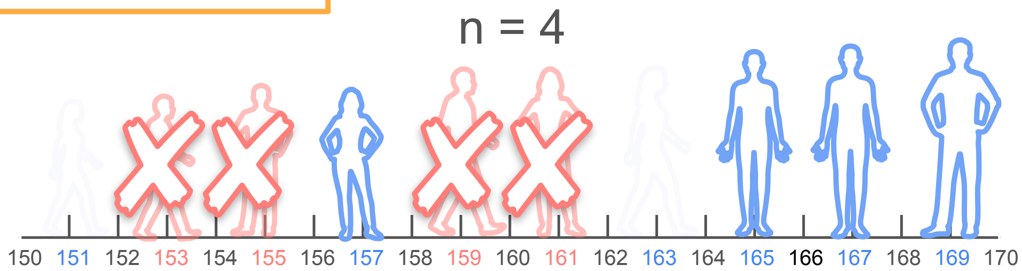
Example 1

1st sample set



Why is sample set two not a good sample?

2nd sample set

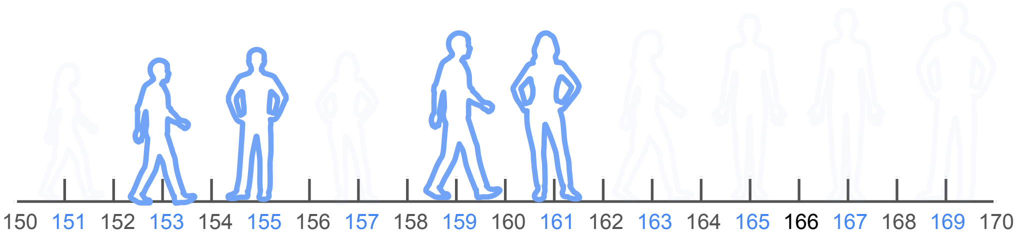


Independent Sample

Example 1

1st sample set

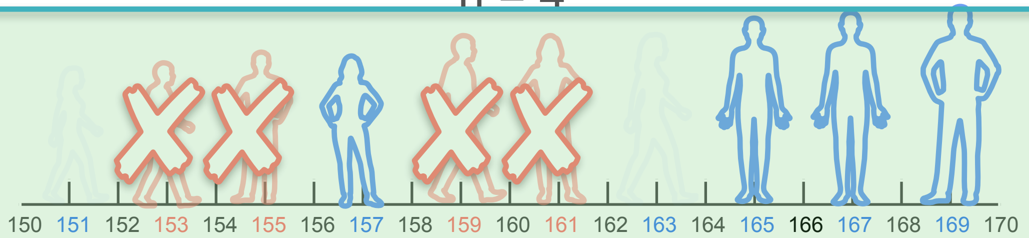
n = 4



Why is sample set two not a good sample?

2nd sample set

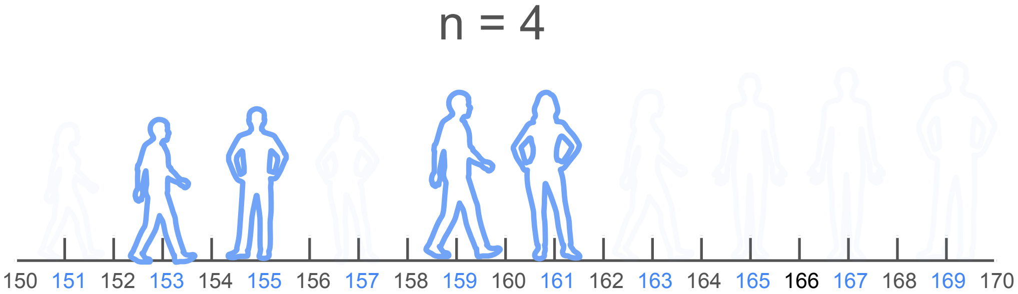
n = 4



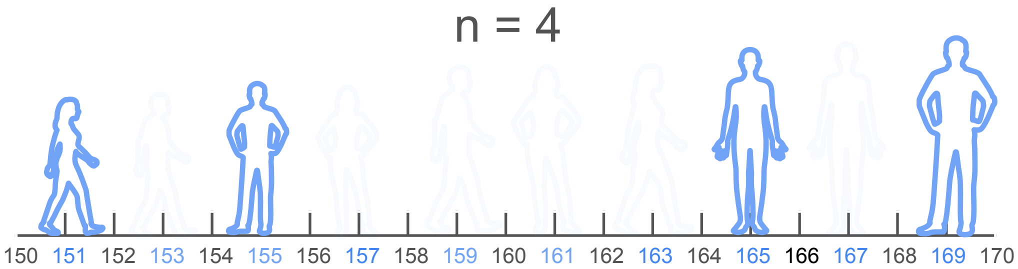
Independent Sample

Example 2

1st sample set

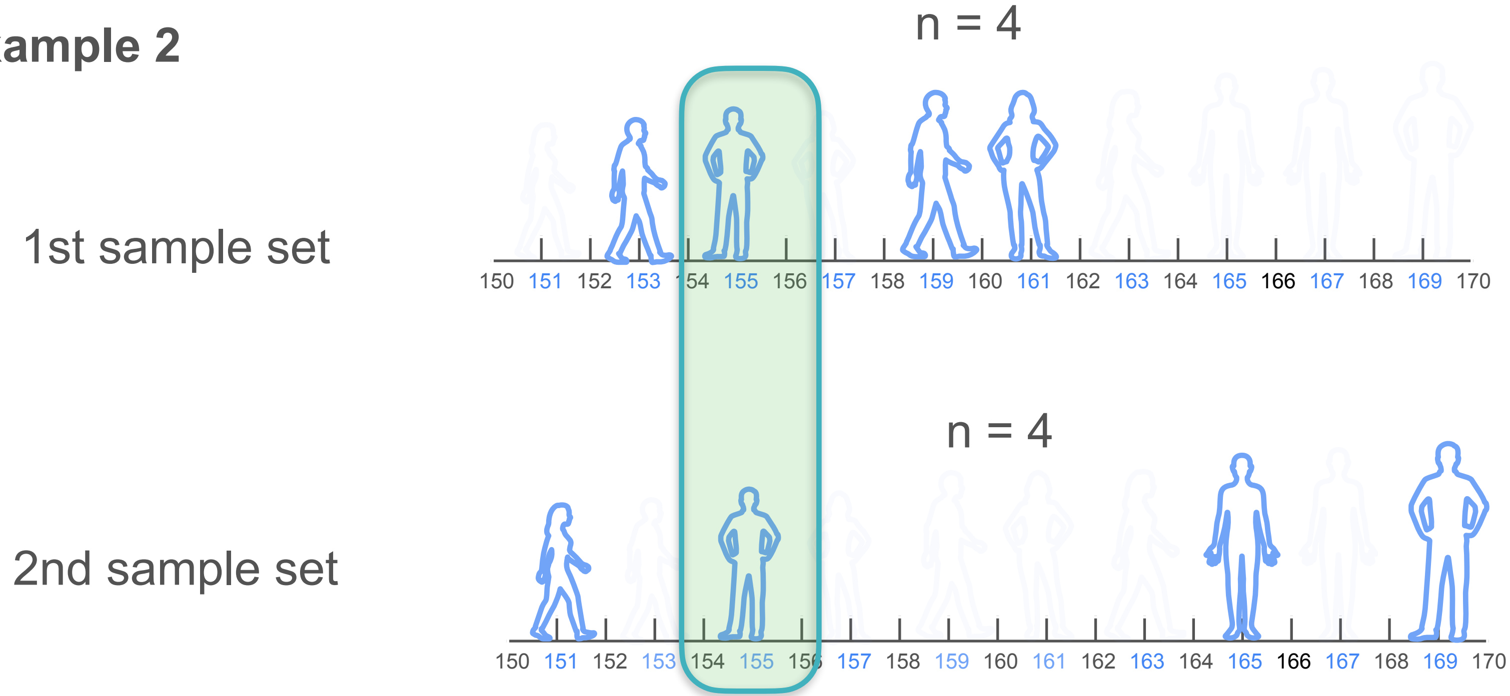


2nd sample set



Independent Sample

Example 2



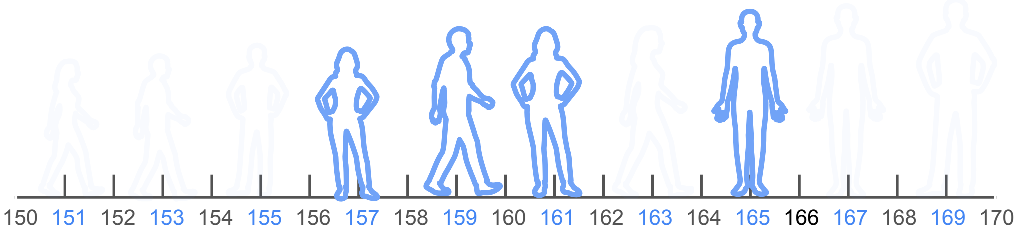
Identically Distributed Samples

Identically Distributed Samples

Example 1

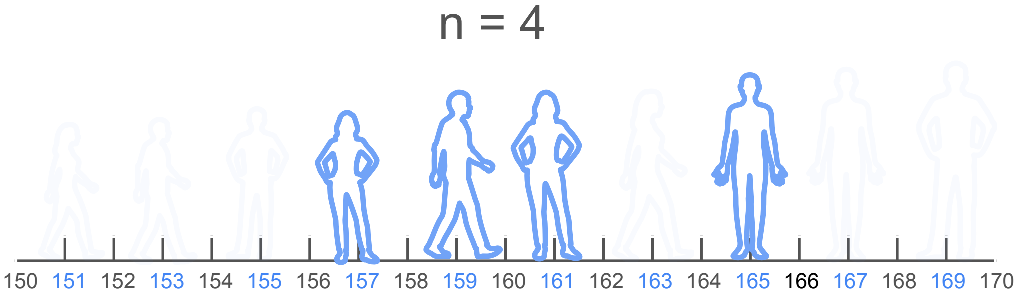
Identically Distributed Samples

Example 1



Identically Distributed Samples

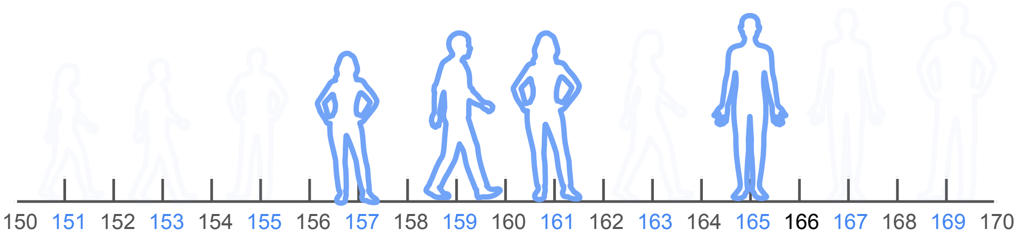
Example 1

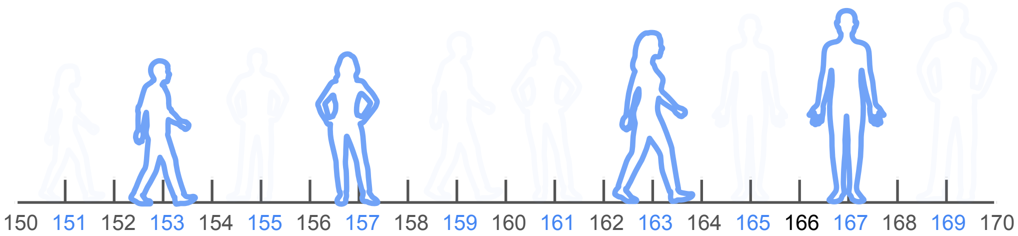


Identically Distributed Samples

Example 1

n = 4

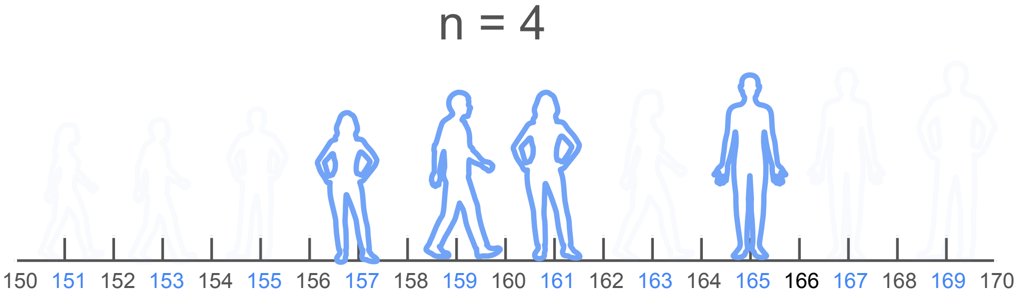


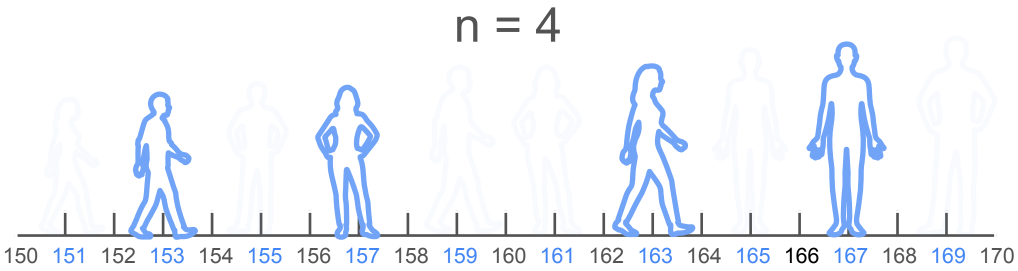


DeepLearning.AI

Identically Distributed Samples

Example 1

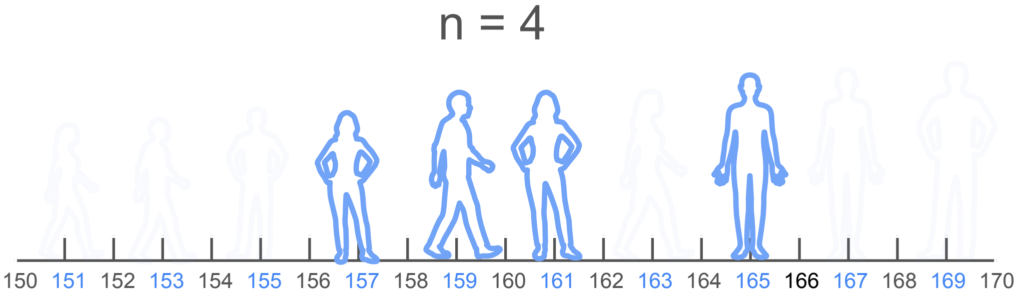


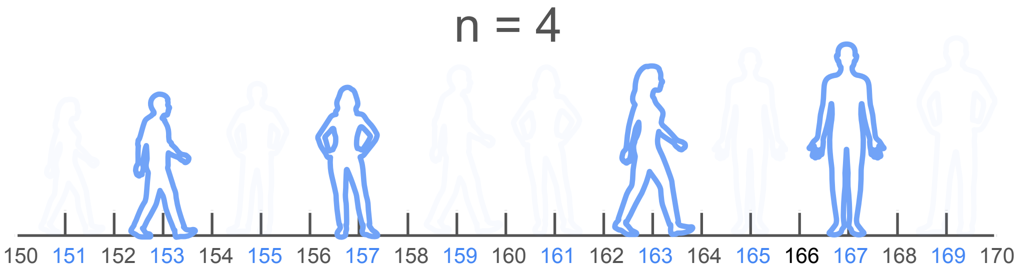


Identically Distributed Samples

Example 1

A

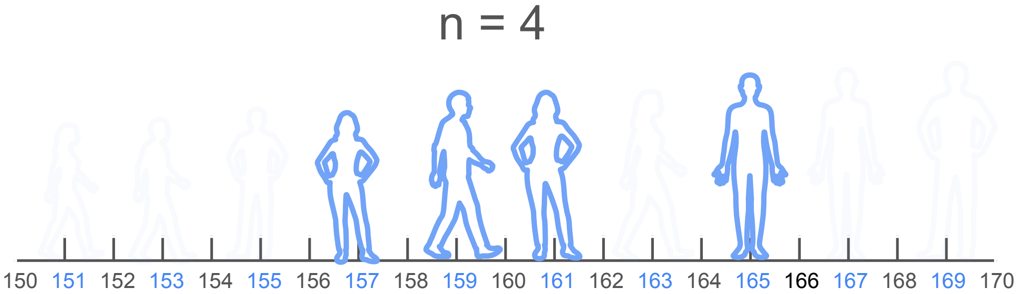




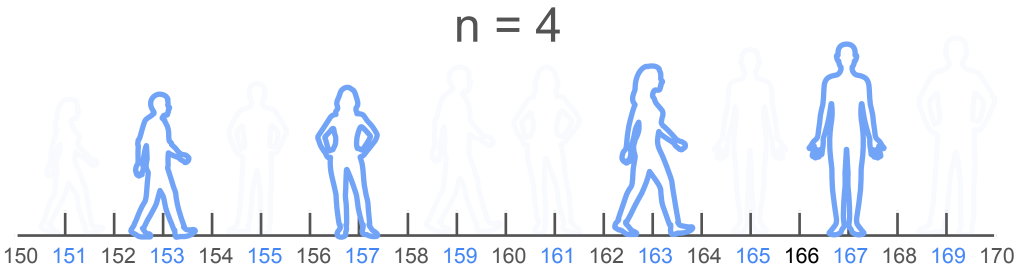
Identically Distributed Samples

Example 1

A



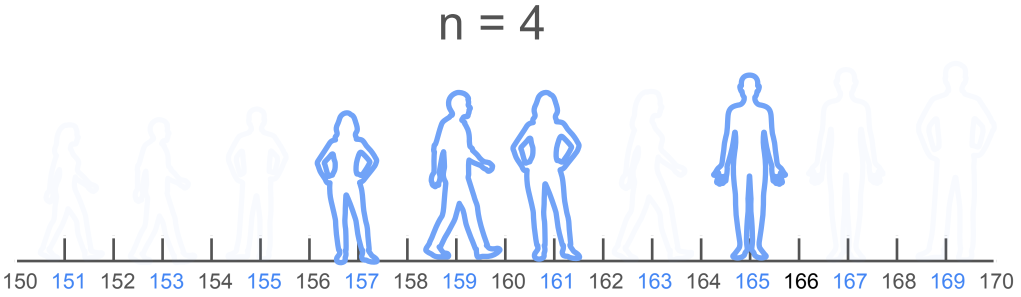
B



Identically Distributed Samples

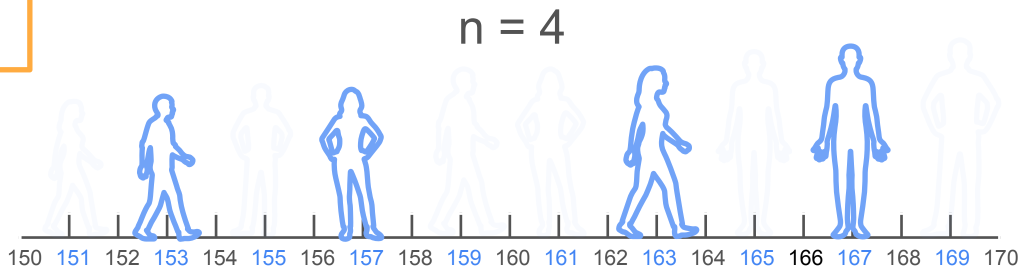
Example 1

A



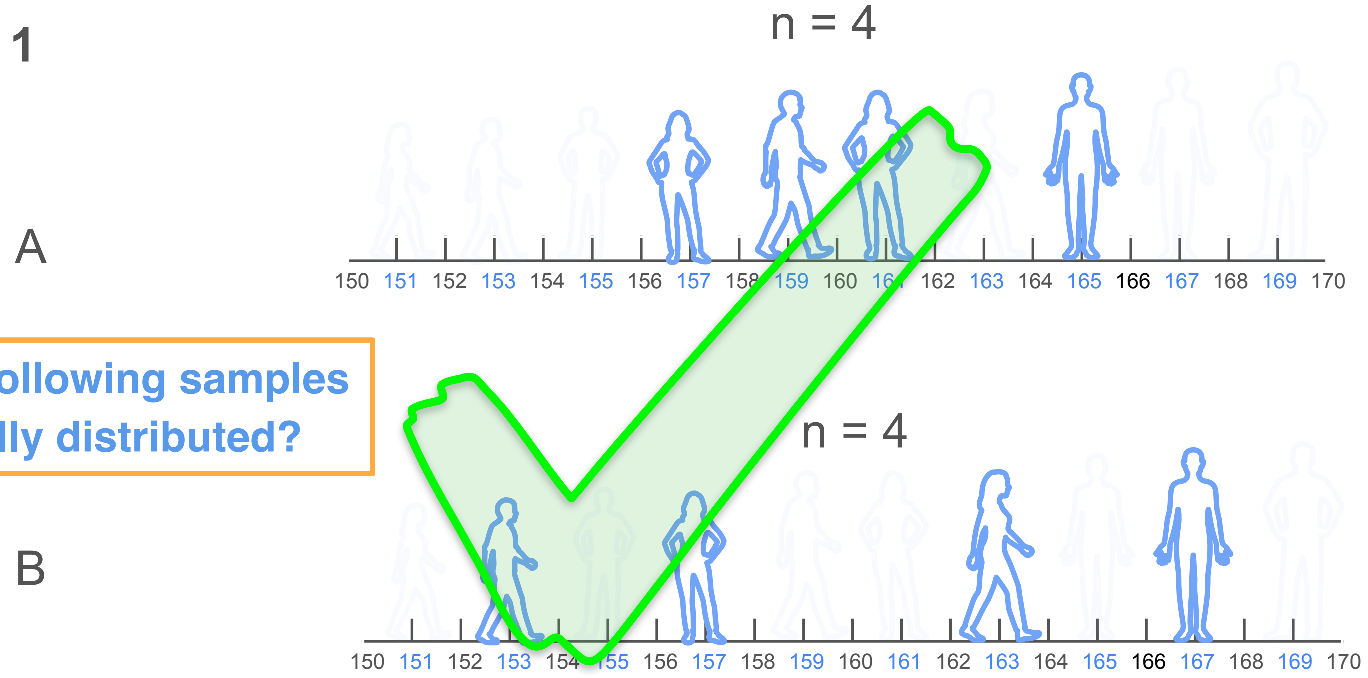
Which of the following samples are identically distributed?

B



Identically Distributed Samples

Example 1



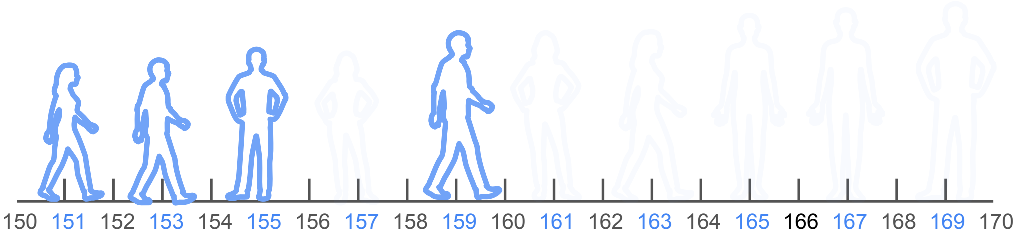
Which of the following samples are identically distributed?

Identically Distributed Samples

Example 2

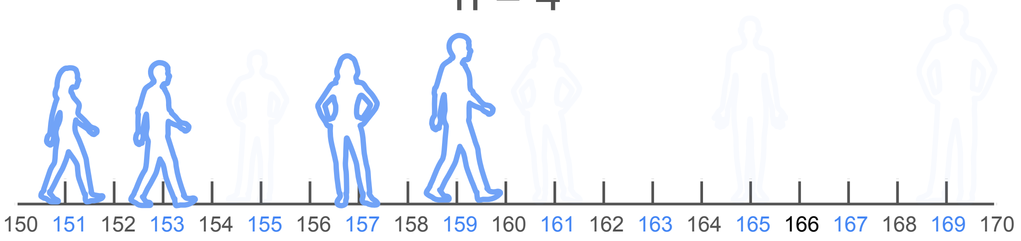
A

n = 4



B

n = 4

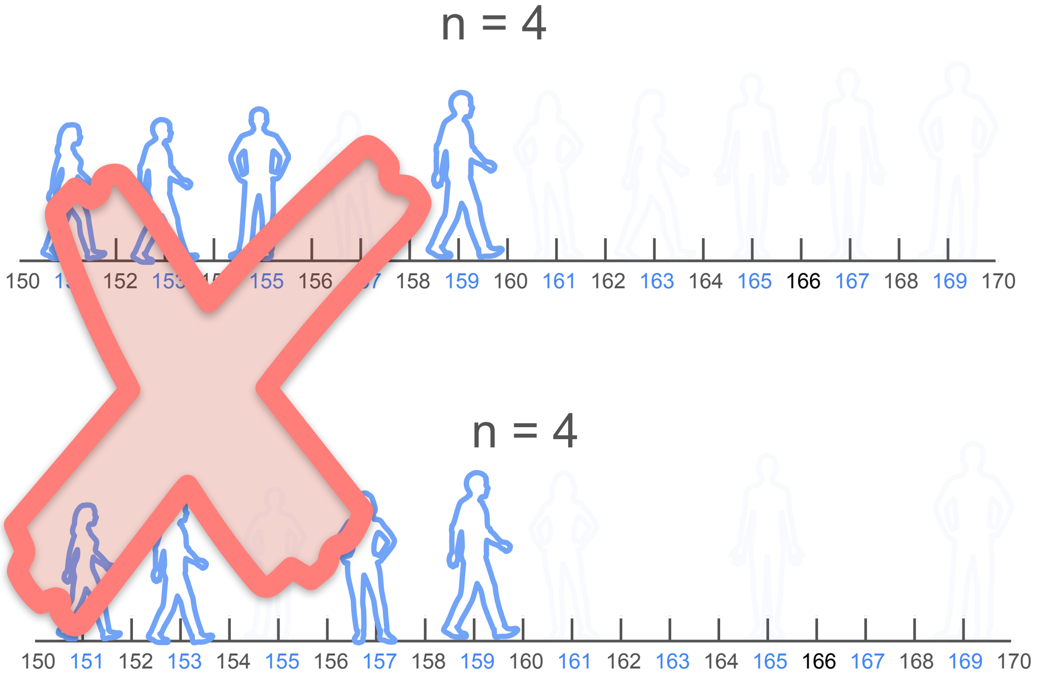


Identically Distributed Samples

Example 2

A

B

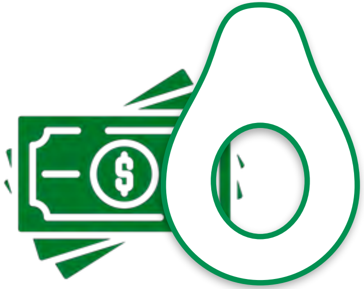


The Avocado Toast Trend

The Avocado Toast Trend

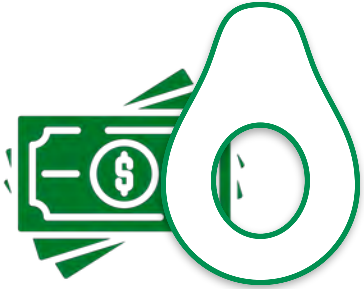
The Avocado Toast Trend

Study the price of avocados
in the United States



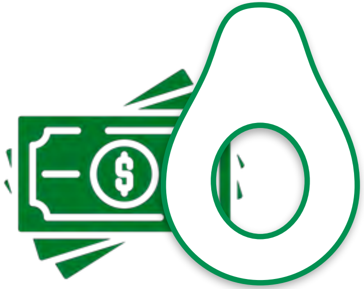
The Avocado Toast Trend

Study the price of avocados
in the United States



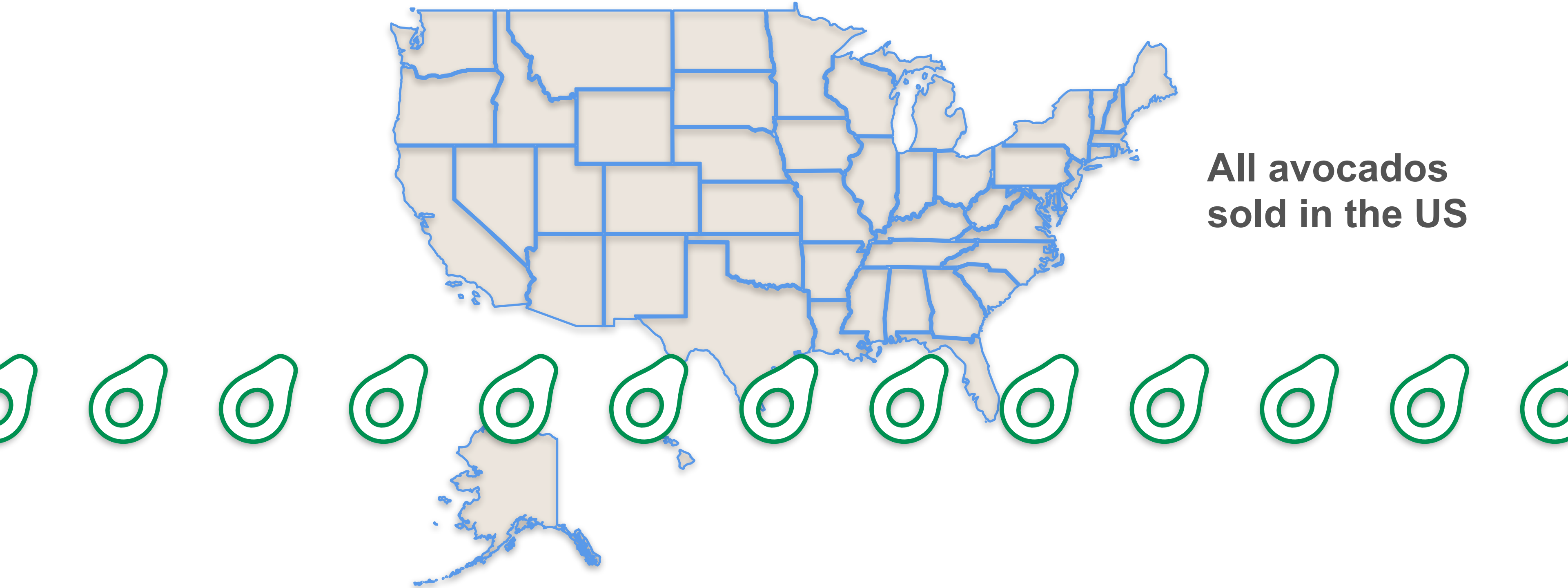
The Avocado Toast Trend

Study the price of avocados
in the United States



What is the population of your study?

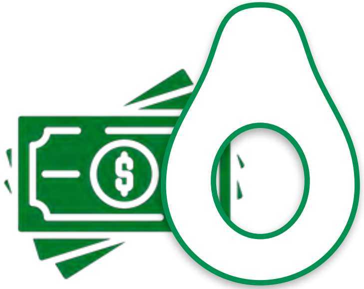
The Avocado Toast Trend



All avocados sold in the US

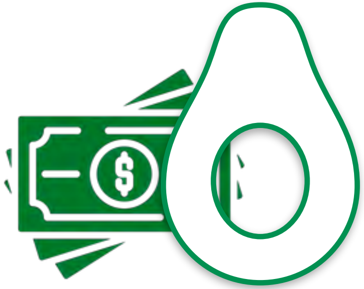
The Avocado Toast Trend

Study the price of avocados
in the United States



The Avocado Toast Trend

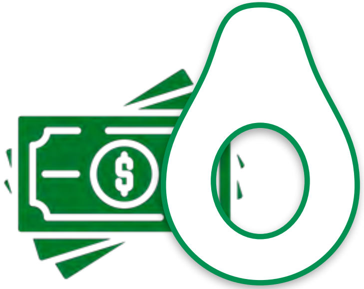
Study the price of avocados
in the United States



What is the sample of your study?

The Avocado Toast Trend

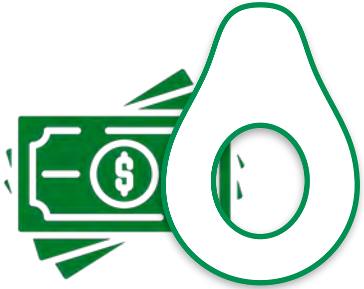
Study the price of avocados
in the United States



What is the sample of your study?

The Avocado Toast Trend

Study the price of avocados
in the United States



What is the sample of your study?

Avocados sold in the 4 stores you selected

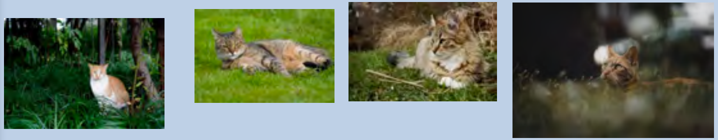
Population and Sample in Machine Learning

Population and Sample in Machine Learning

Every dataset you work with in machine learning is a sample

NOT the population

Cats



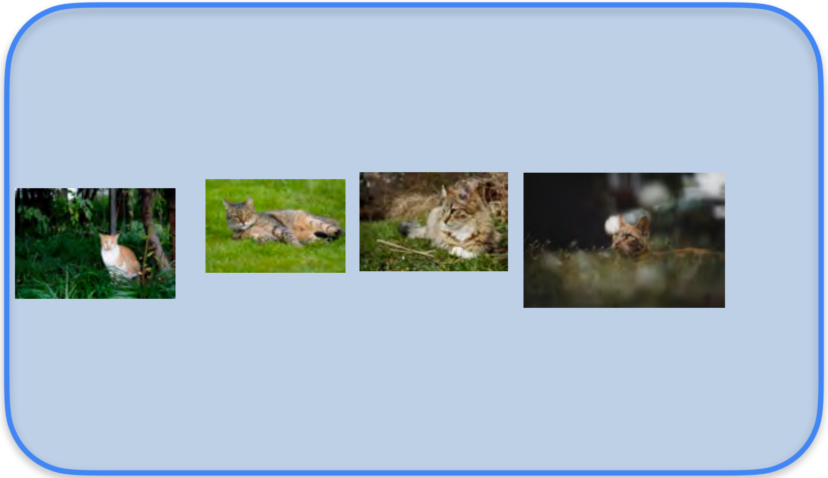
Not cats

Population and Sample in Machine Learning

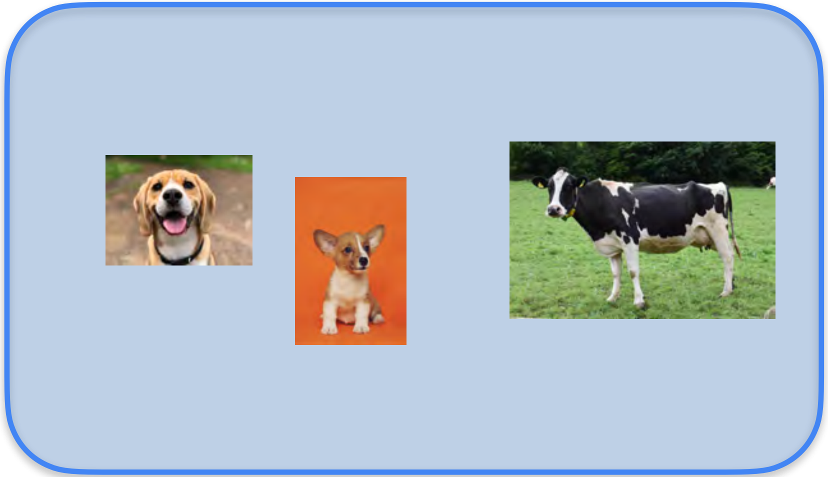
Every dataset you work with in machine learning is a sample

NOT the population

Cats



Not cats



Population and Sample in Machine Learning

Every dataset you work with in machine learning is a sample

NOT the population

Population and Sample in Machine Learning

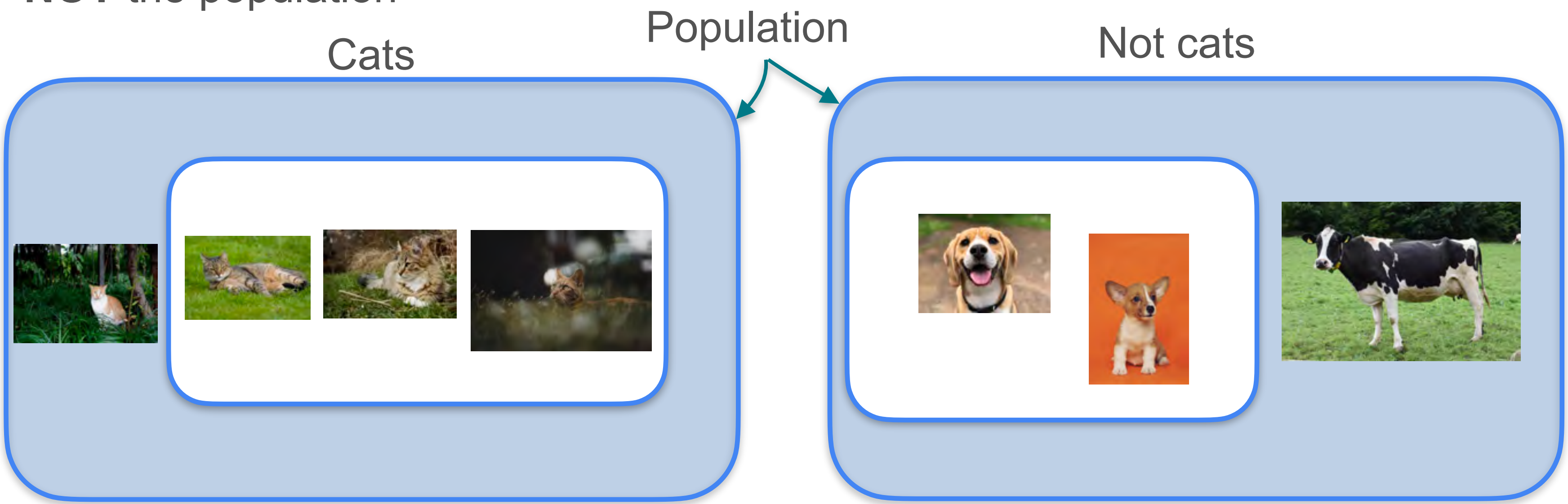
Every dataset you work with in machine learning is a sample

NOT the population

Population and Sample in Machine Learning

Every dataset you work with in machine learning is a sample

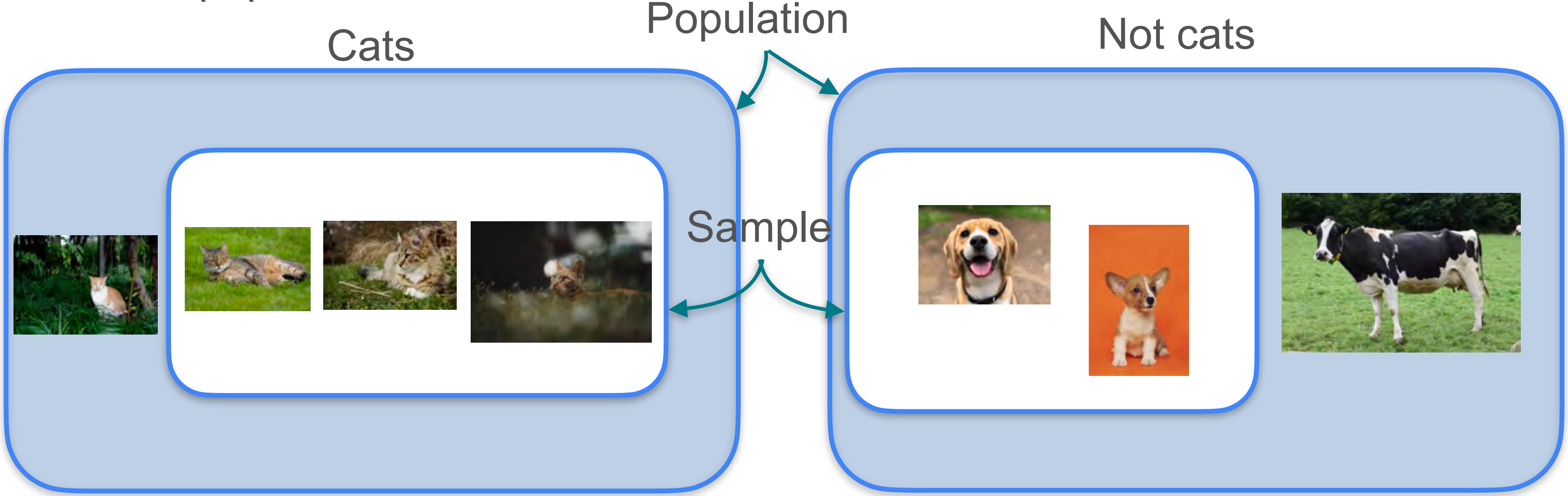
NOT the population



Population and Sample in Machine Learning

Every dataset you work with in machine learning is a sample

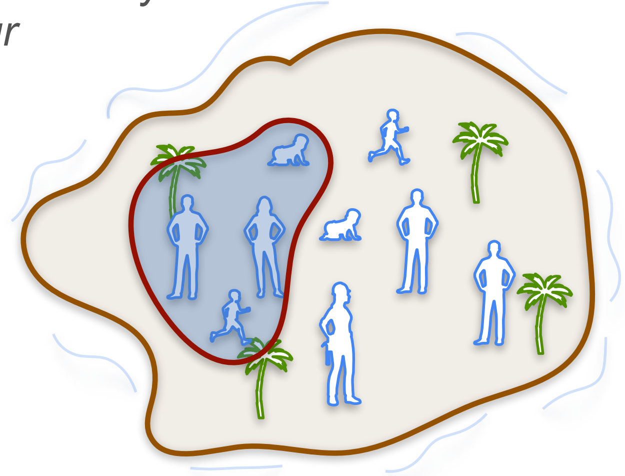
NOT the population



Recap

Population

the entire group of individuals or elements you want to study which share a common behaviour



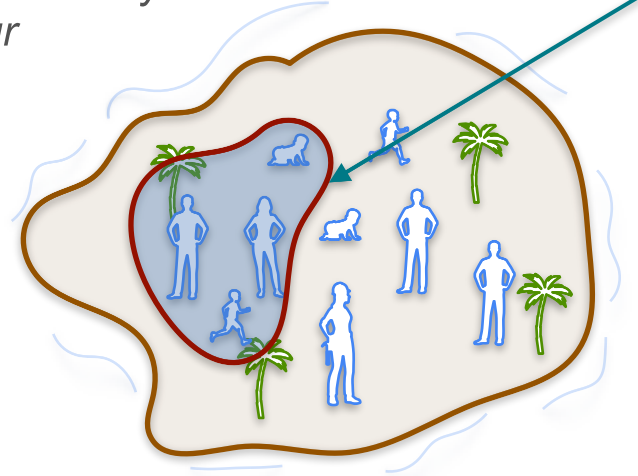
Sample

subset of the population you use to draw conclusions about the population as a whole

Recap

Population

the entire group of individuals or elements you want to study which share a common behaviour



Sample

subset of the population you use to draw conclusions about the population as a whole

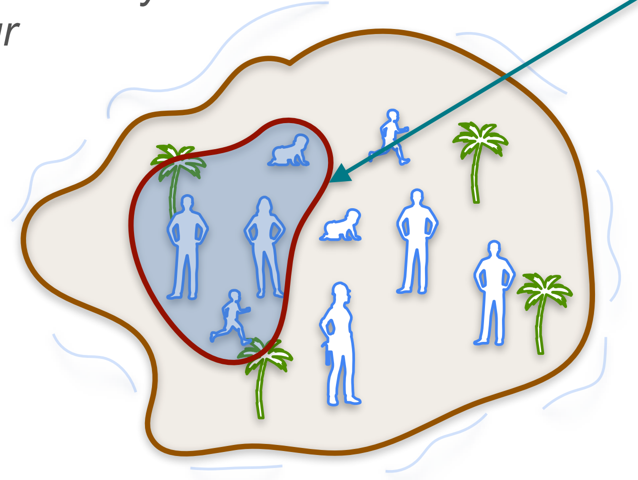
Population Size: N

Sample Size:

Recap

Population

the entire group of individuals or elements you want to study which share a common behaviour



Sample

subset of the population you use to draw conclusions about the population as a whole

Population Size: N

Sample Size: n

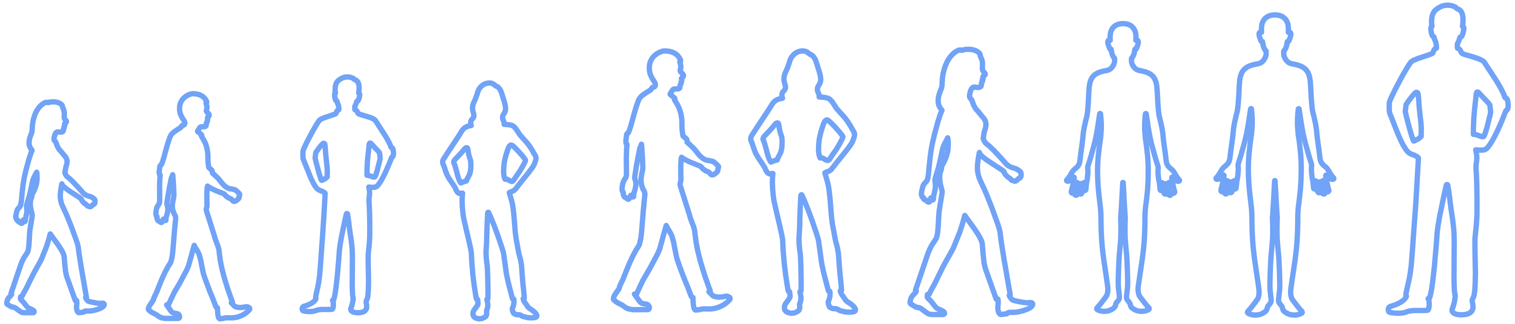
DeepLearning.AI

Sample and Population

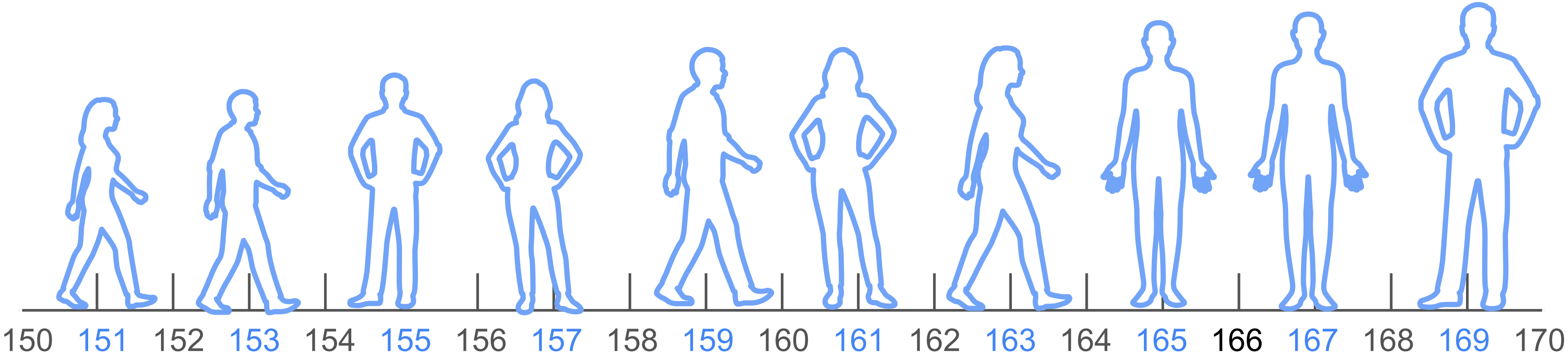
Sample Mean, Proportion, and Variance

Population and Sample Mean

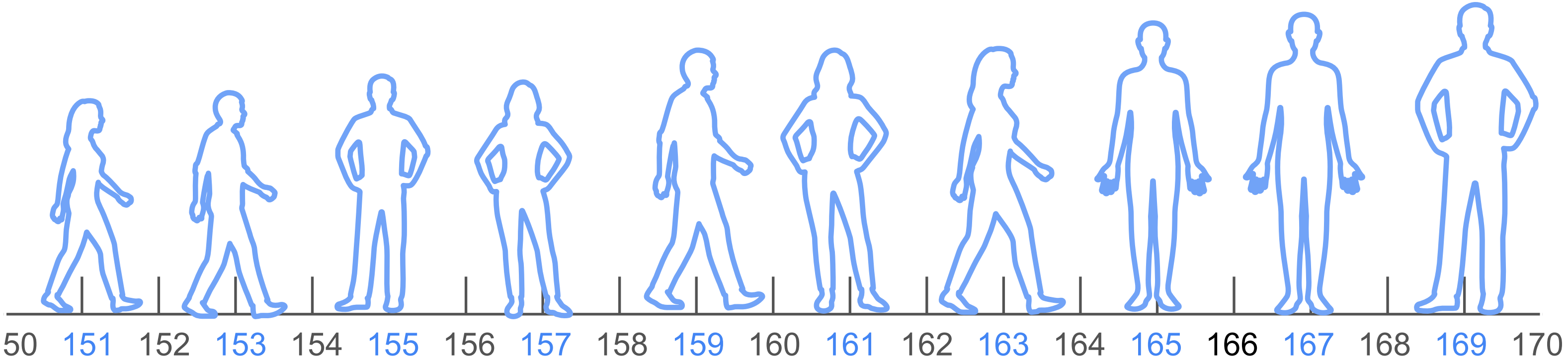
Population and Sample Mean



Population and Sample Mean

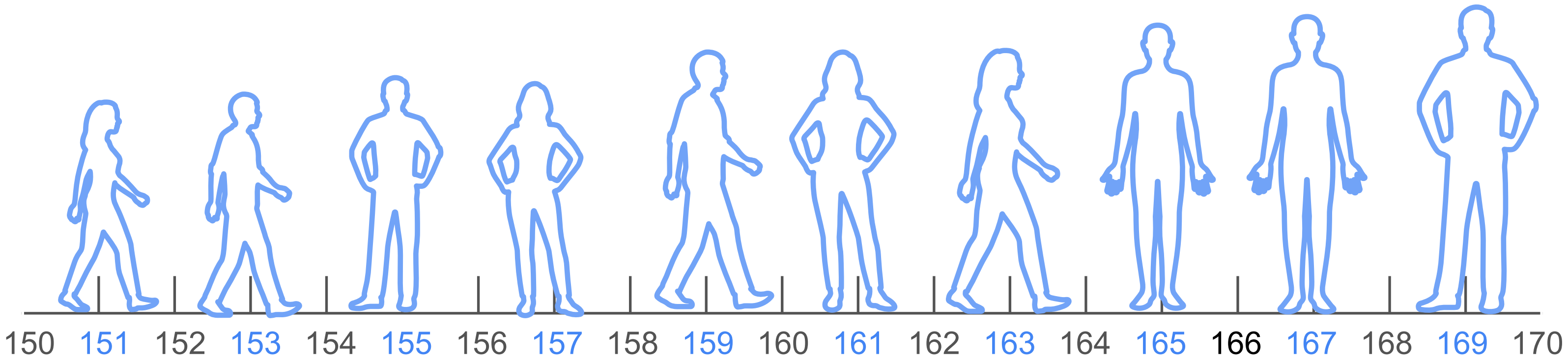


Population and Sample Mean



What is the average height in statistopia?

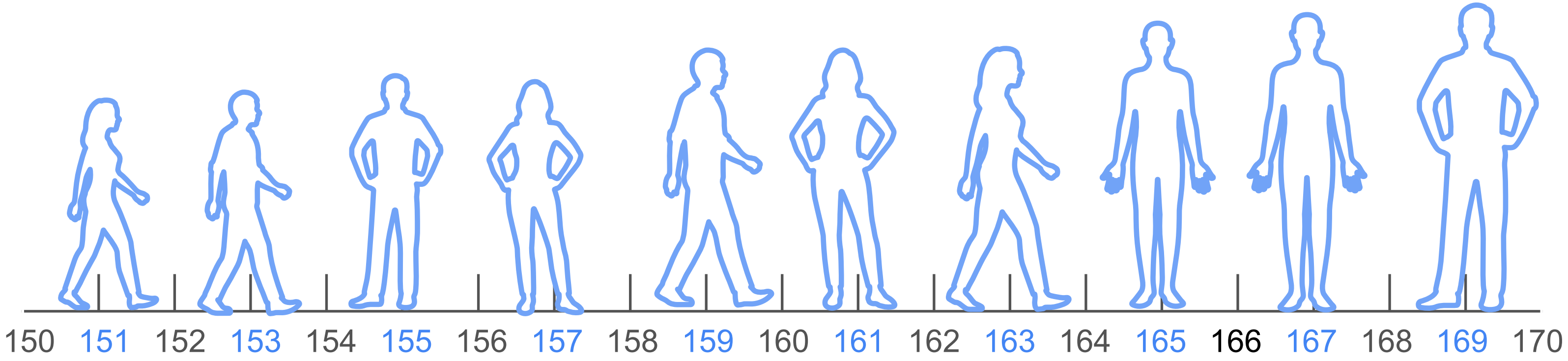
Population and Sample Mean



What is the average height in statistopia?

$$\frac{151 + 153 + 155 + 157 + 159 + 161 + 163 + 165 + 167 + 169}{10}$$

Population and Sample Mean

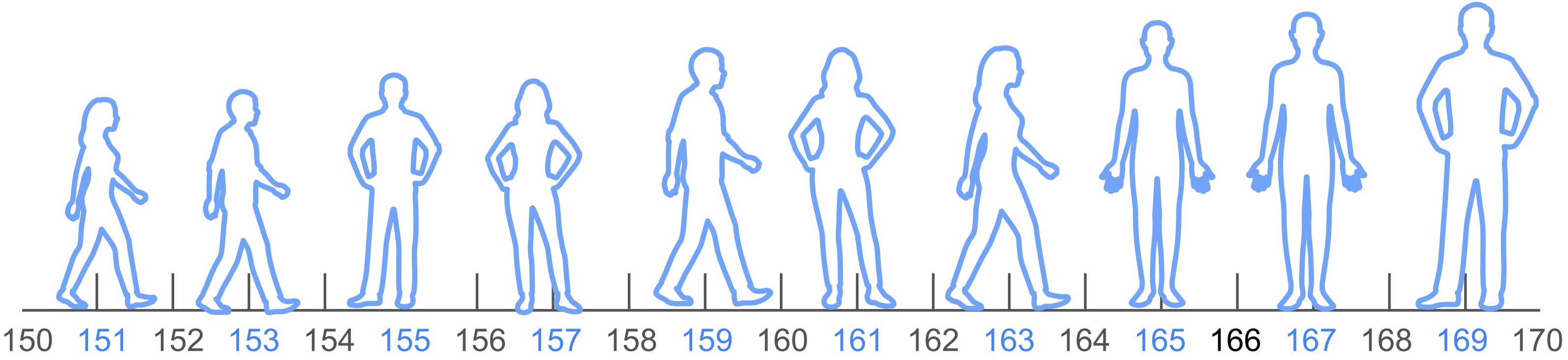


What is the average height in statistopia?

$$\frac{151 + 153 + 155 + 157 + 159 + 161 + 163 + 165 + 167 + 169}{10}$$

$$= \frac{1600}{10} = 160cm$$

Population and Sample Mean



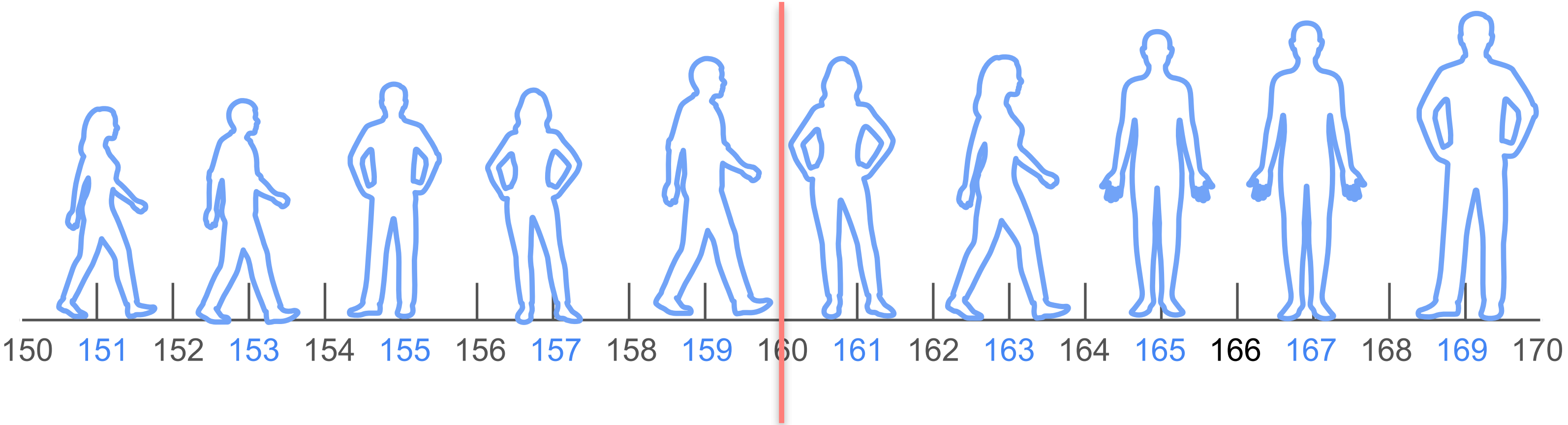
What is the average height in statistopia?

$$\frac{151 + 153 + 155 + 157 + 159 + 161 + 163 + 165 + 167 + 169}{10}$$

$$\boxed{= \frac{1600}{10} = 160cm}$$

Population mean μ

Population and Sample Mean



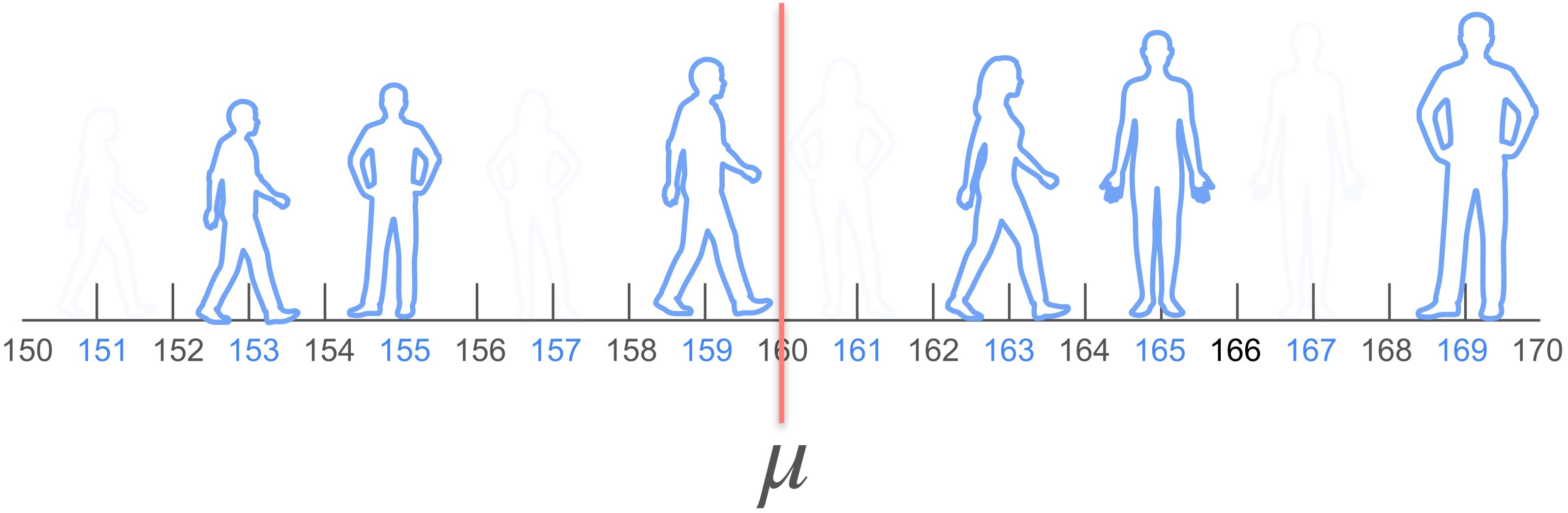
What is the average height in statistopia?

$$\frac{151 + 153 + 155 + 157 + 159 + 161 + 163 + 165 + 167 + 169}{10}$$

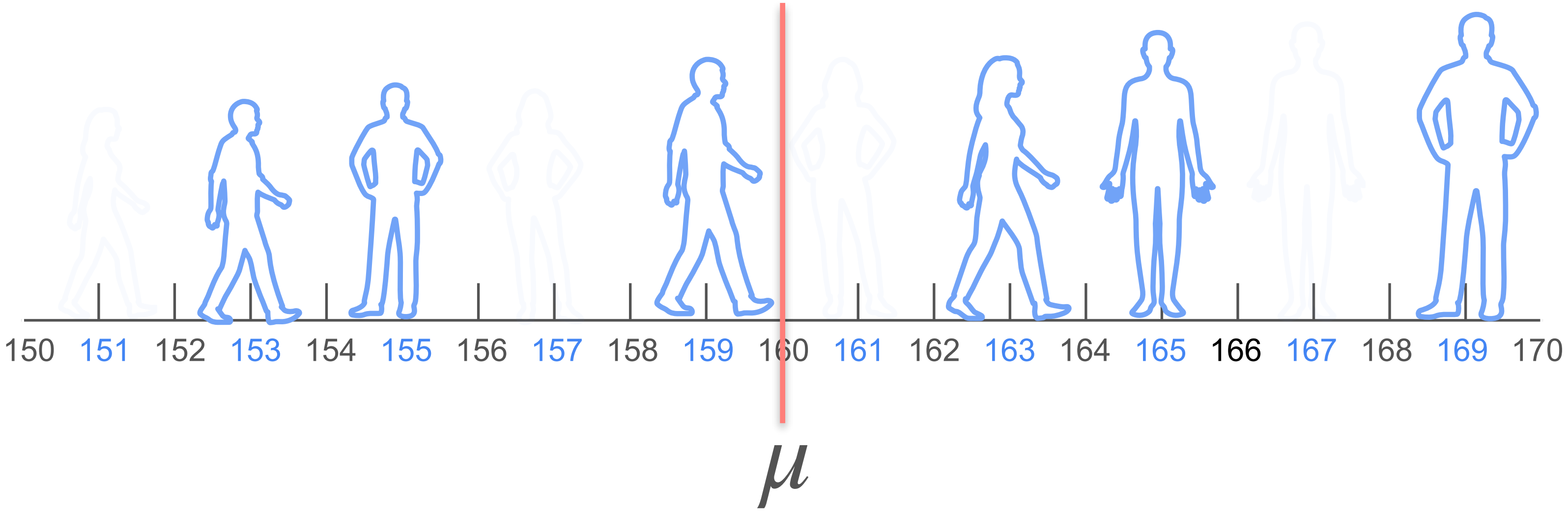
$$\boxed{= \frac{1600}{10} = 160cm}$$

Population mean μ

Population and Sample Mean

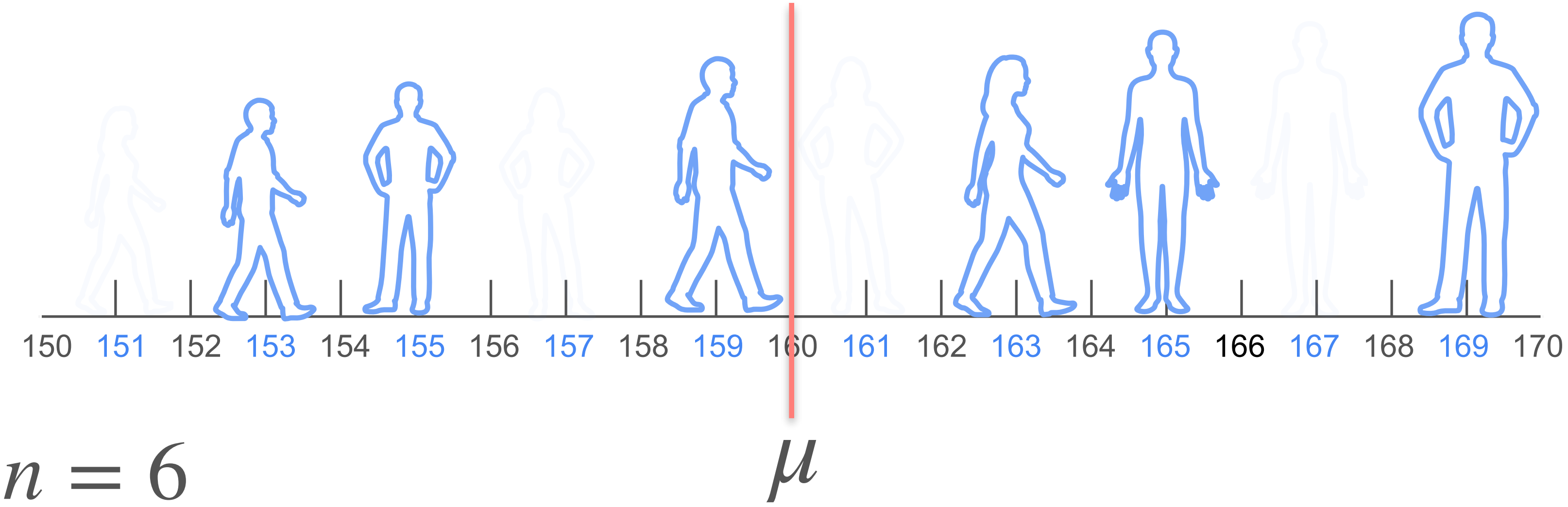


Population and Sample Mean



What is the average
height in statistopia?

Population and Sample Mean



$n = 6$

What is the average height in statistopia?

Population and Sample Mean

$$\mu$$

$$n = 6$$

What is the average height in statistopia?

$$\frac{153 + 155 + 159 + 163 + 165 + 169}{6}$$

Population and Sample Mean

μ

$n = 6$

What is the average height in statistopia?

$$\frac{153 + 155 + 159 + 163 + 165 + 169}{6} = \frac{964}{6} = 160.97$$

Population and Sample Mean

Sample mean

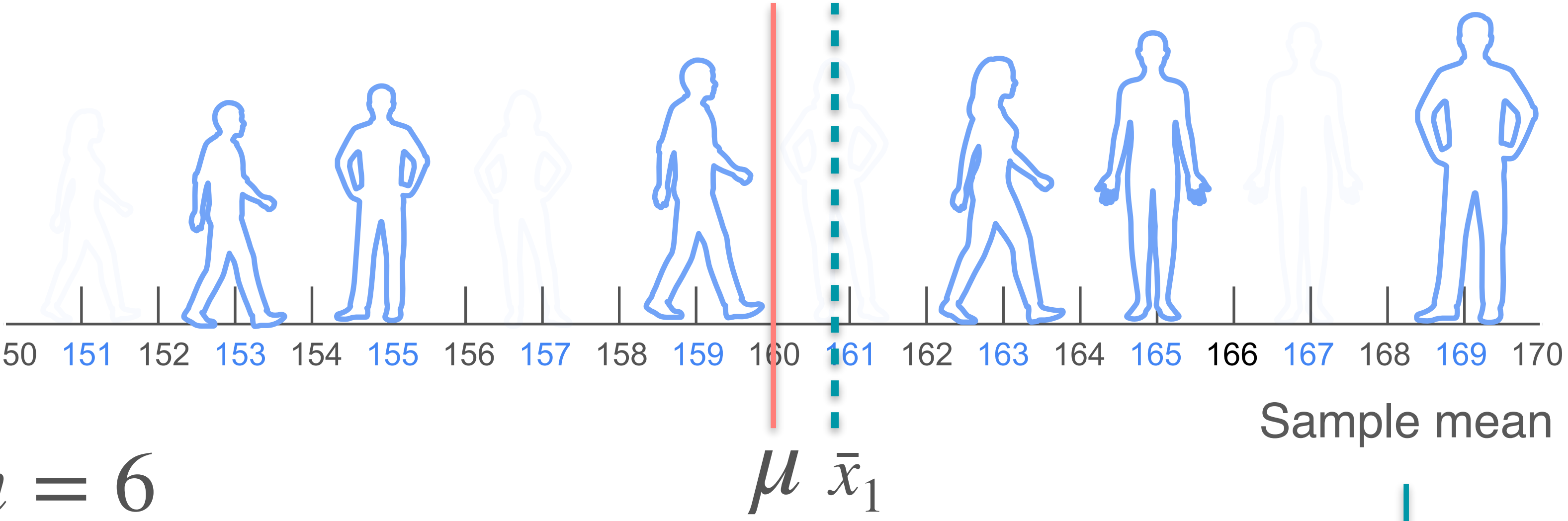
$n = 6$

$\mu \quad \bar{x}_1$

What is the average height in statistopia?

$$\frac{153 + 155 + 159 + 163 + 165 + 169}{6} = \frac{964}{6} \boxed{= 160.97} \quad \bar{x}$$

Population and Sample Mean

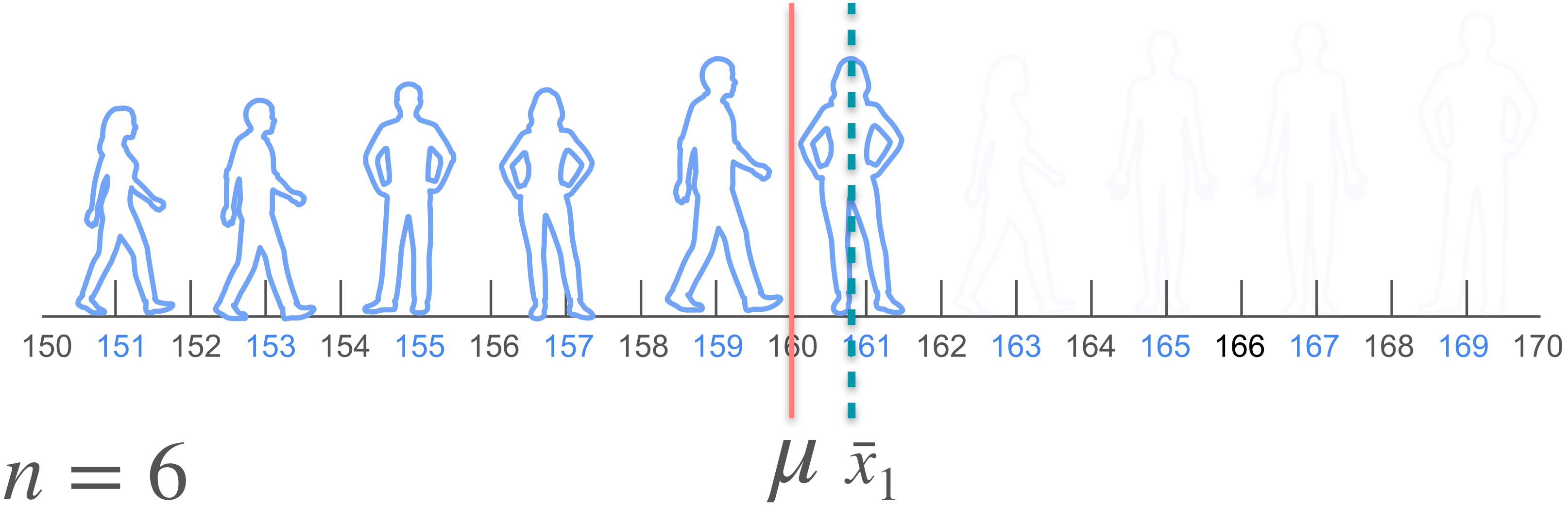


$$n = 6$$

What is the average height in statistopia?

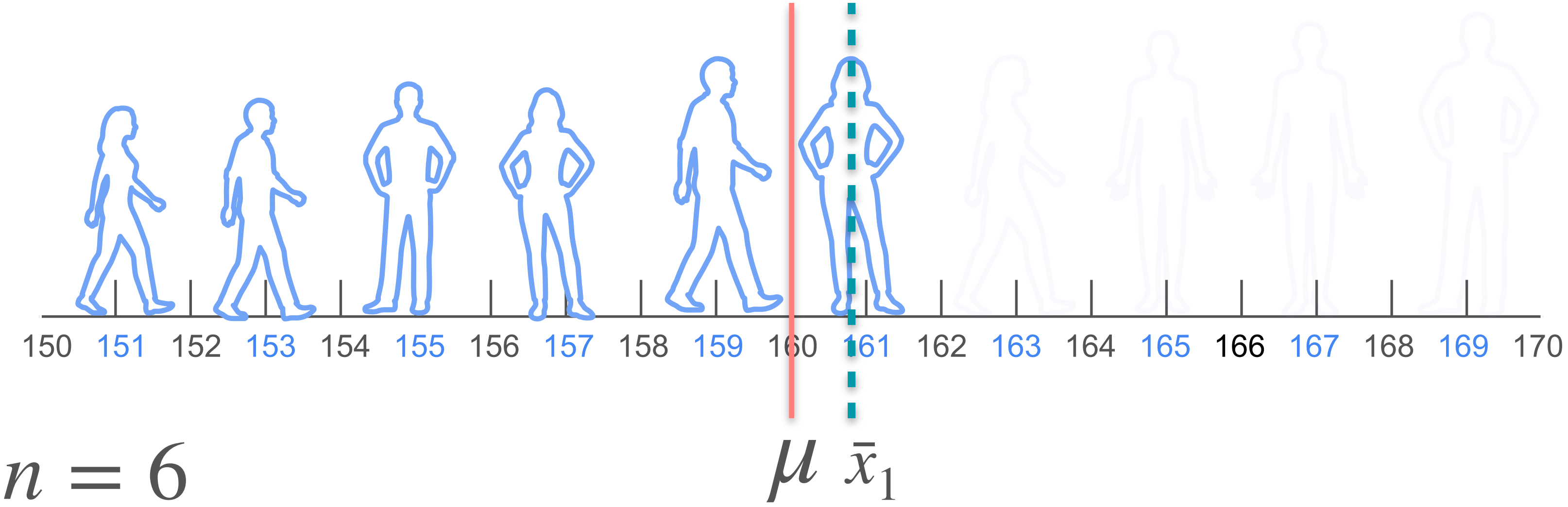
$$\frac{153 + 155 + 159 + 163 + 165 + 169}{6} = \frac{964}{6} \boxed{= 160.97} \quad \bar{x}$$

Population and Sample Mean



$n = 6$

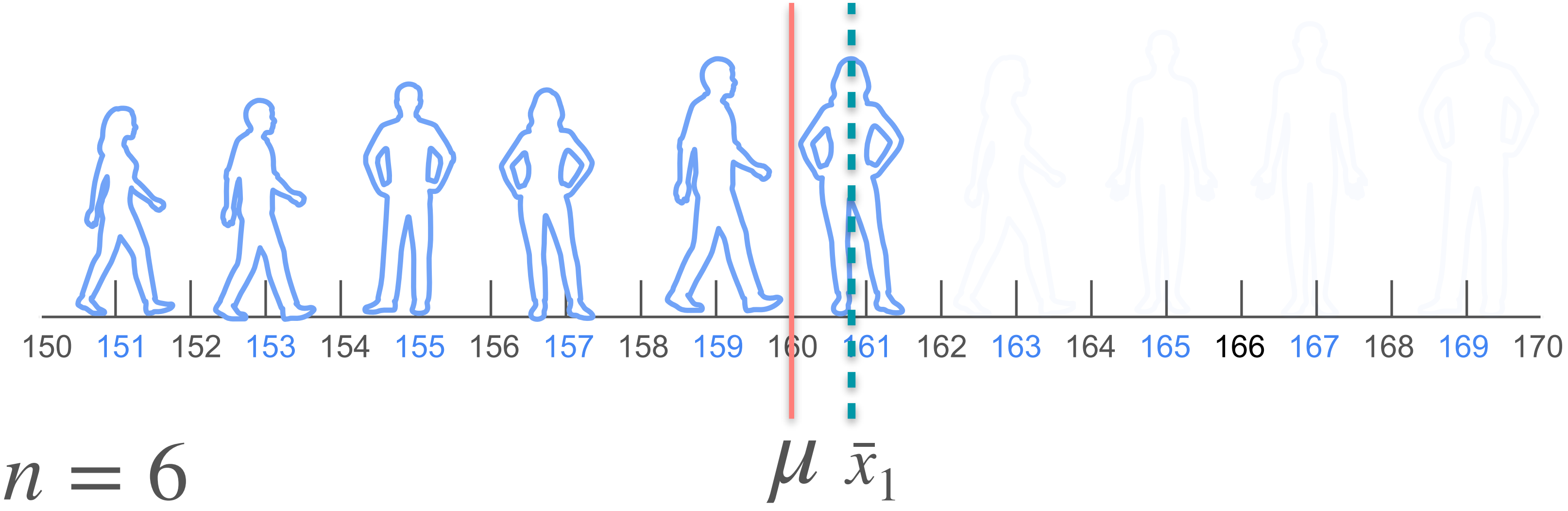
Population and Sample Mean



$$n = 6$$

What is the average height in statistopia?

Population and Sample Mean

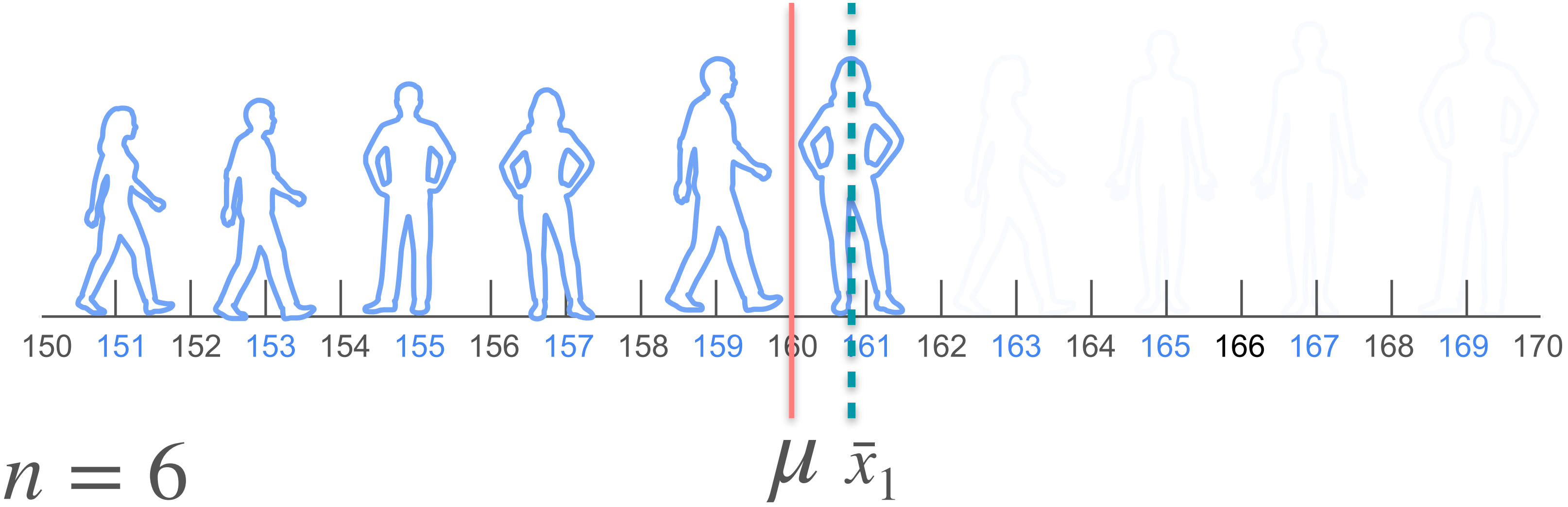


$$n = 6$$

What is the average height in statistopia?

$$\frac{151 + 153 + 155 + 157 + 159 + 161}{6}$$

Population and Sample Mean

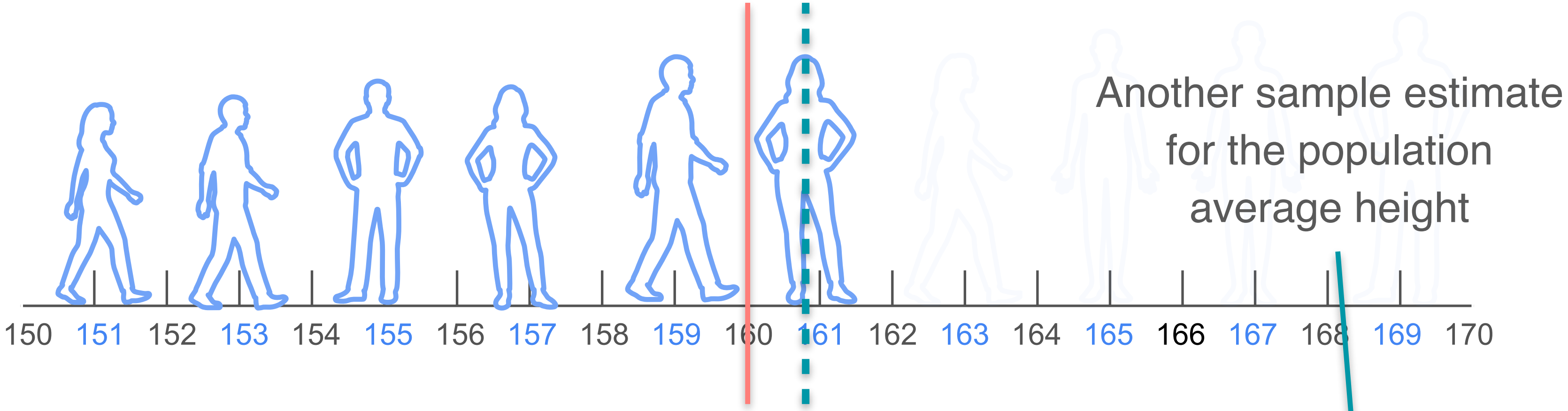


$$n = 6$$

What is the average height in statistopia?

$$\frac{151 + 153 + 155 + 157 + 159 + 161}{6} = \frac{936}{6} = 156cm$$

Population and Sample Mean

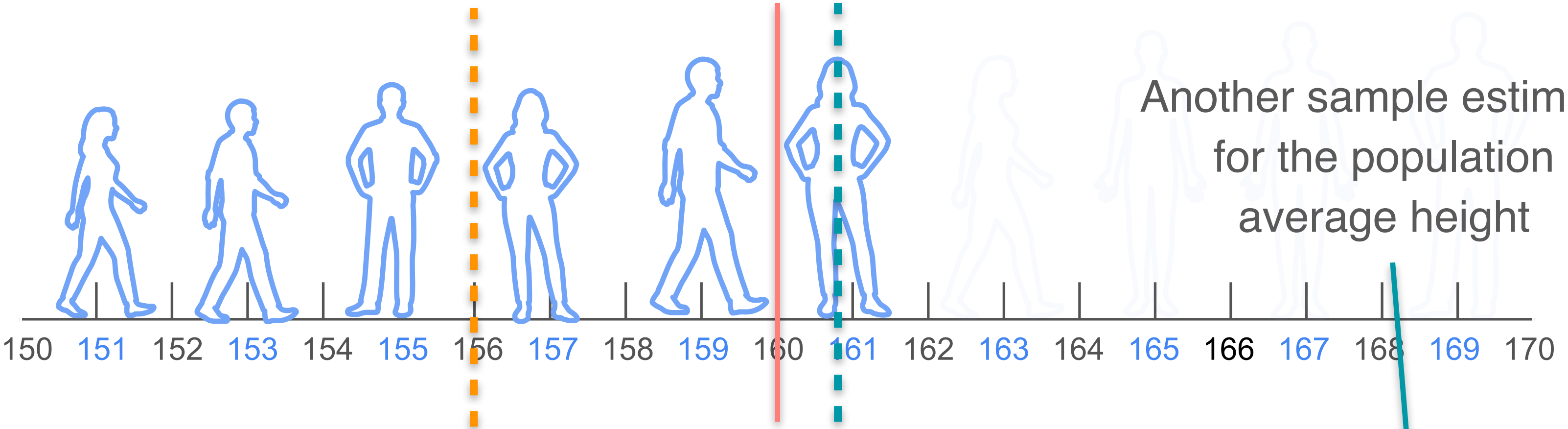


$$n = 6$$

What is the average height in statistopia?

$$\frac{151 + 153 + 155 + 157 + 159 + 161}{6} = \frac{936}{6} = \boxed{156cm}$$

Population and Sample Mean



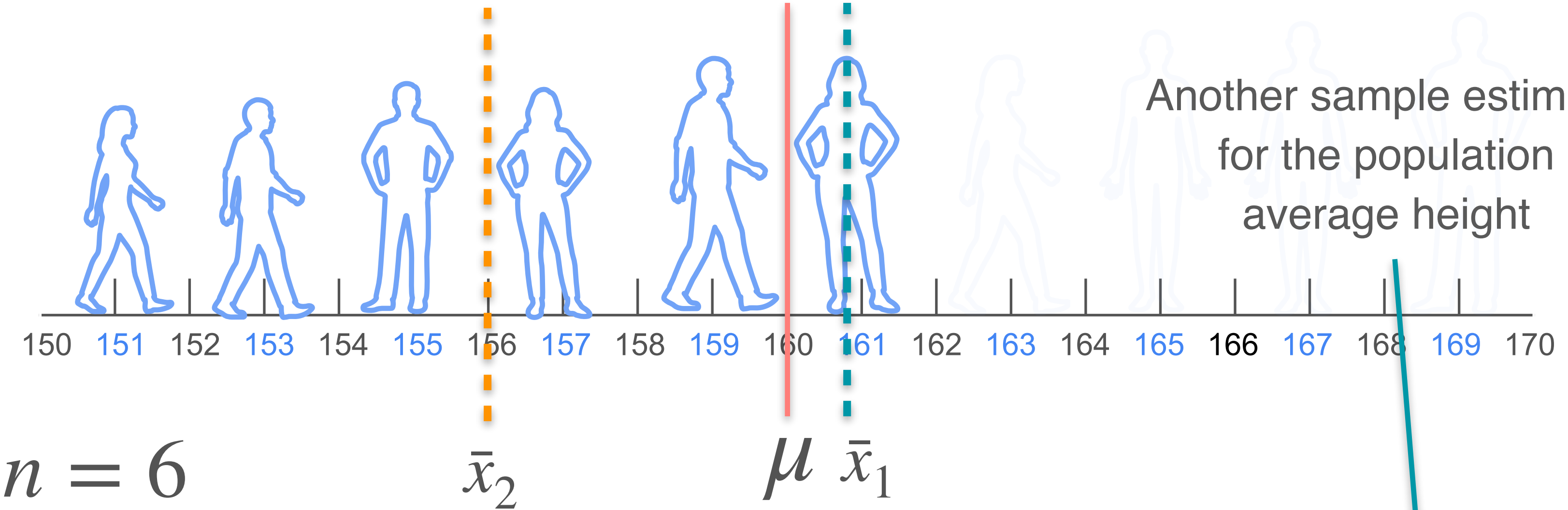
Another sample estimate for the population average height

$$n = 6$$

What is the average height in statistopia?

$$\frac{151 + 153 + 155 + 157 + 159 + 161}{6} = \frac{936}{6} = \boxed{156cm}$$

Population and Sample Mean



Another sample estimate for the population average height

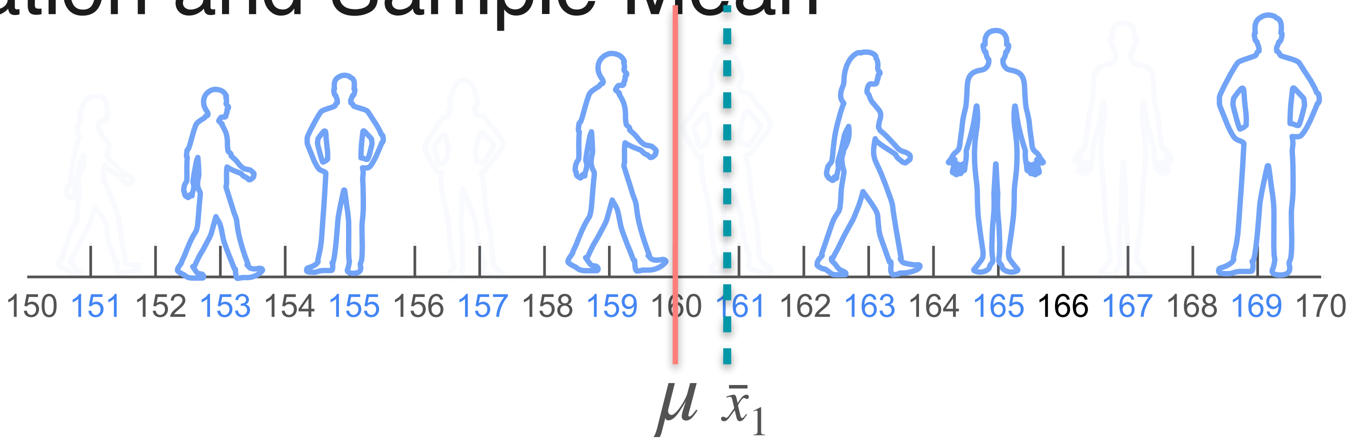
$n = 6$

What is the average height in statistopia?

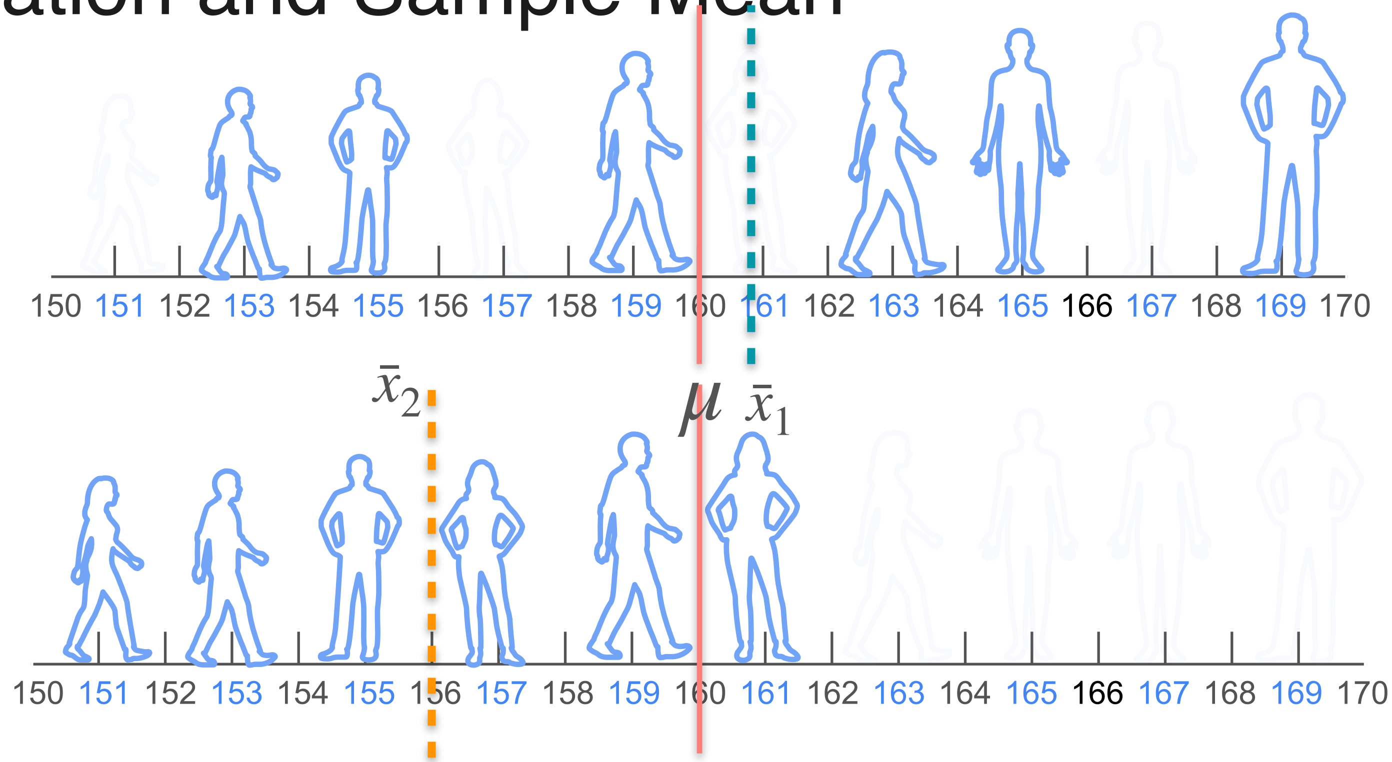
$$\frac{151 + 153 + 155 + 157 + 159 + 161}{6} = \frac{936}{6} = \boxed{156cm}$$

Population and Sample Mean

Population and Sample Mean

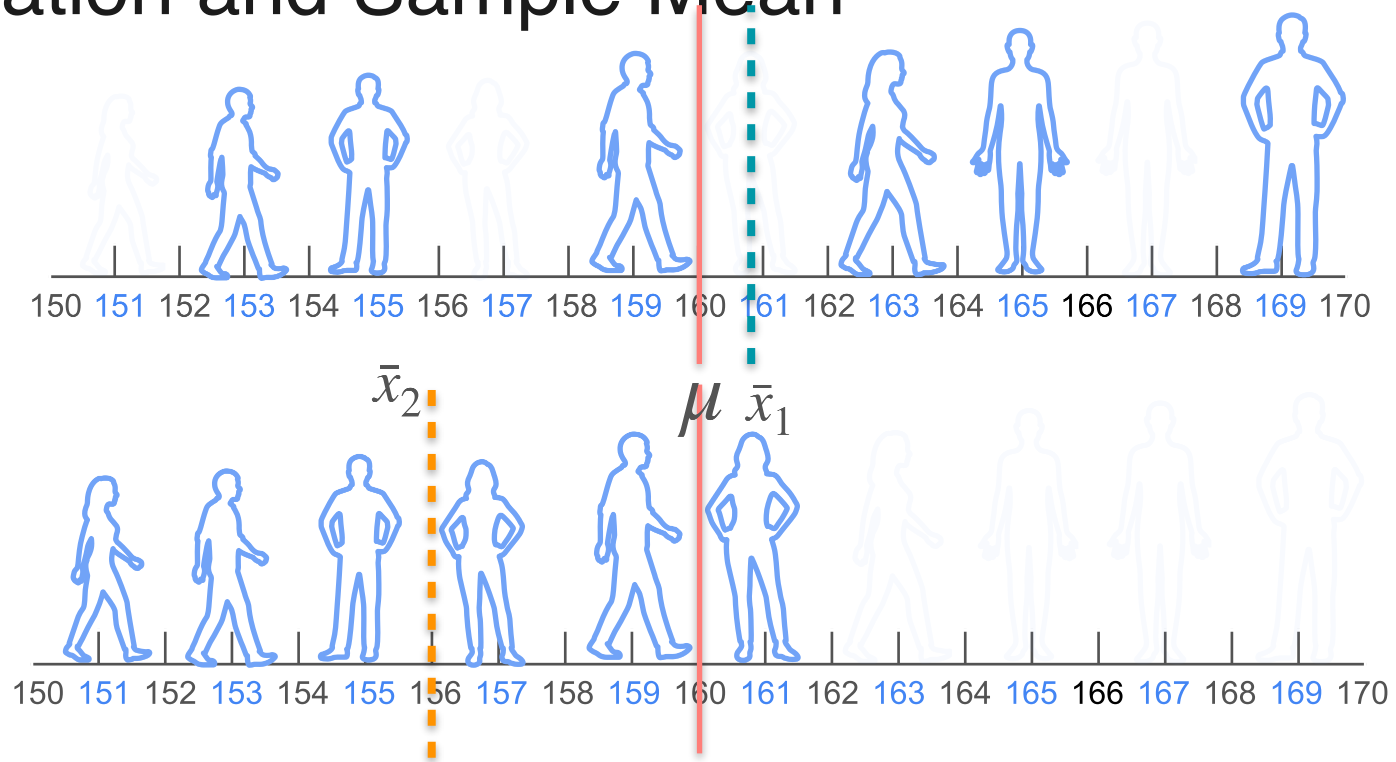


Population and Sample Mean

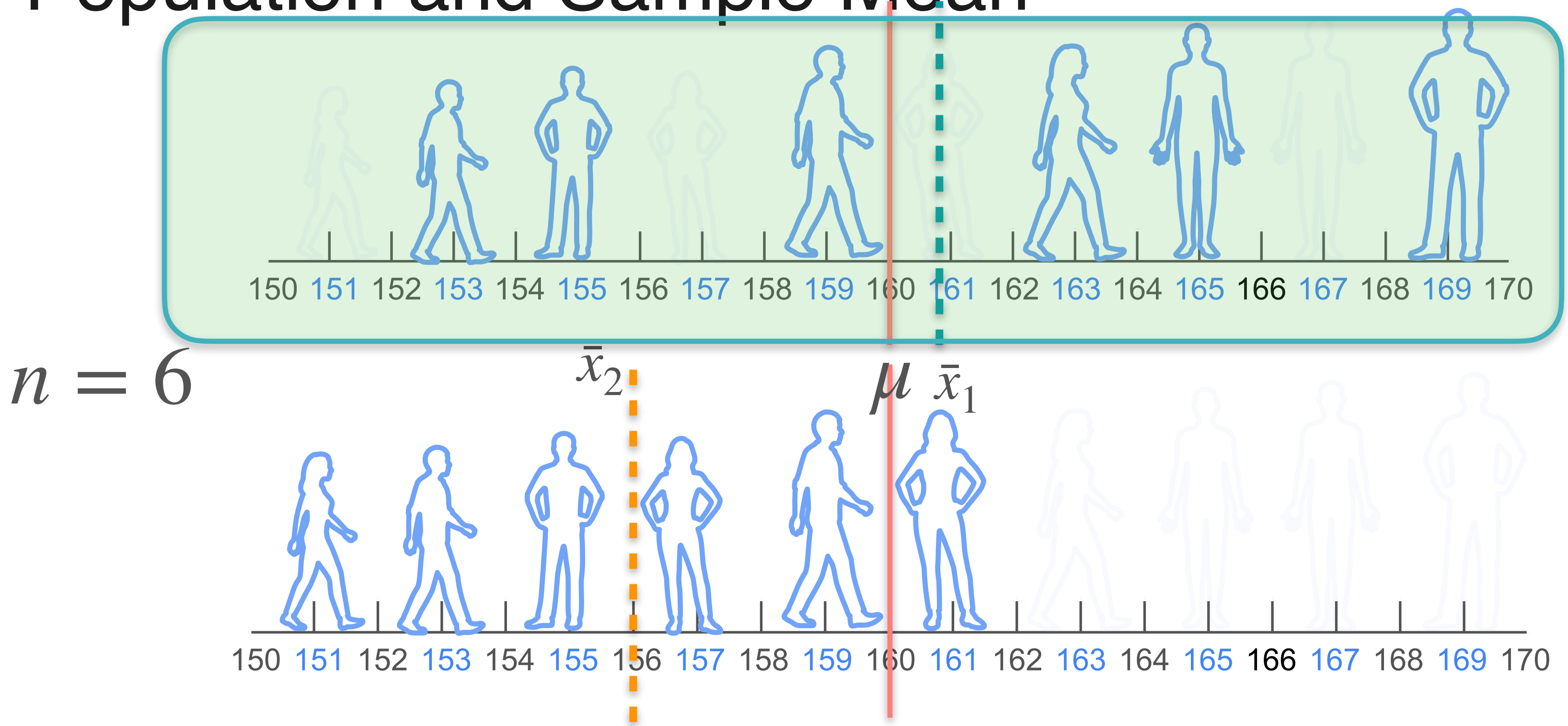


Population and Sample Mean

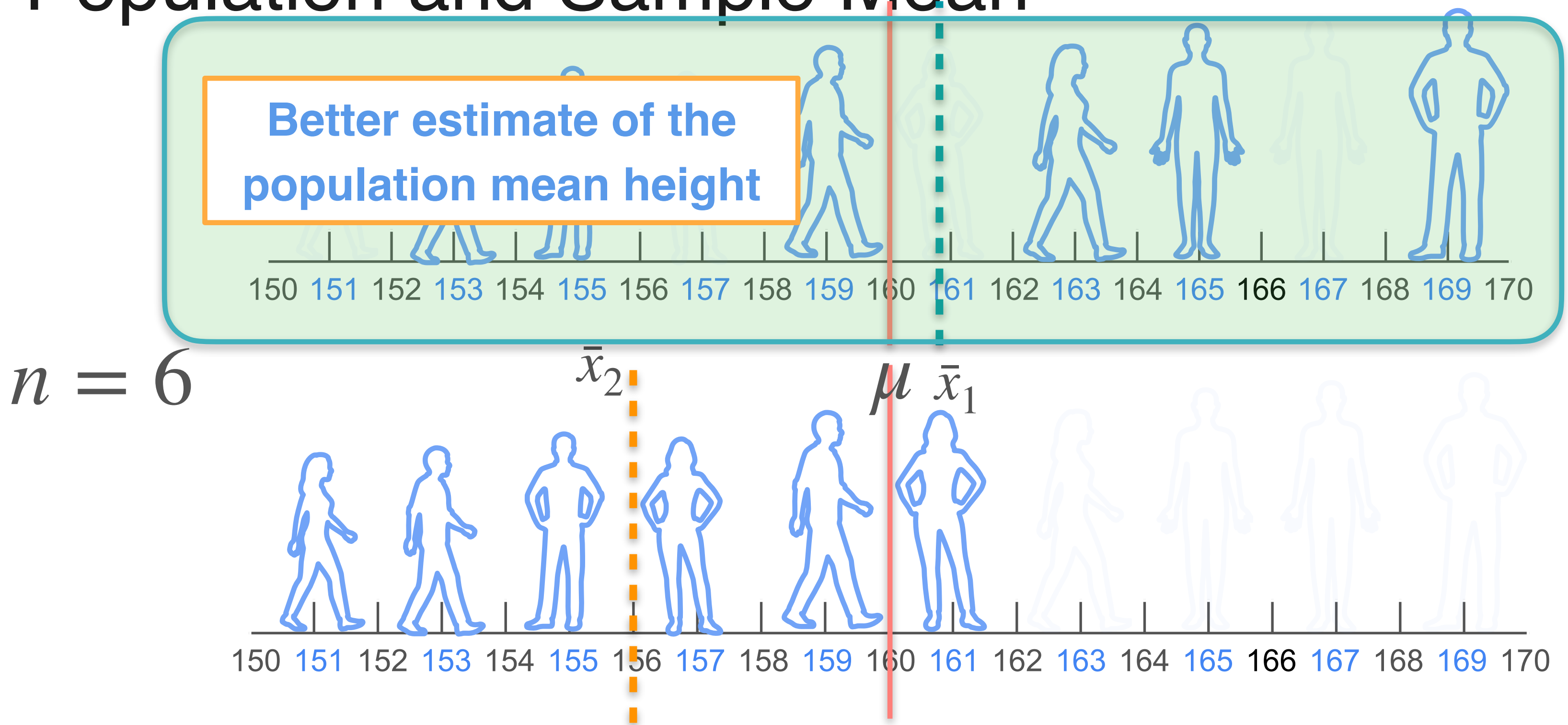
$n = 6$



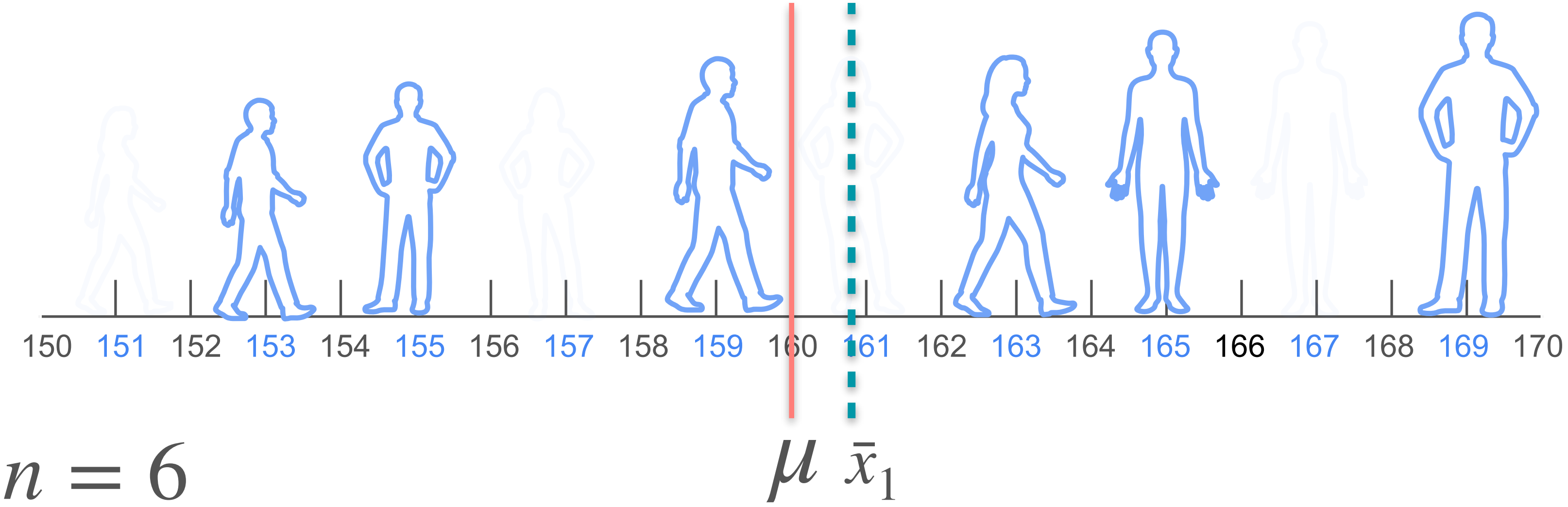
Population and Sample Mean



Population and Sample Mean

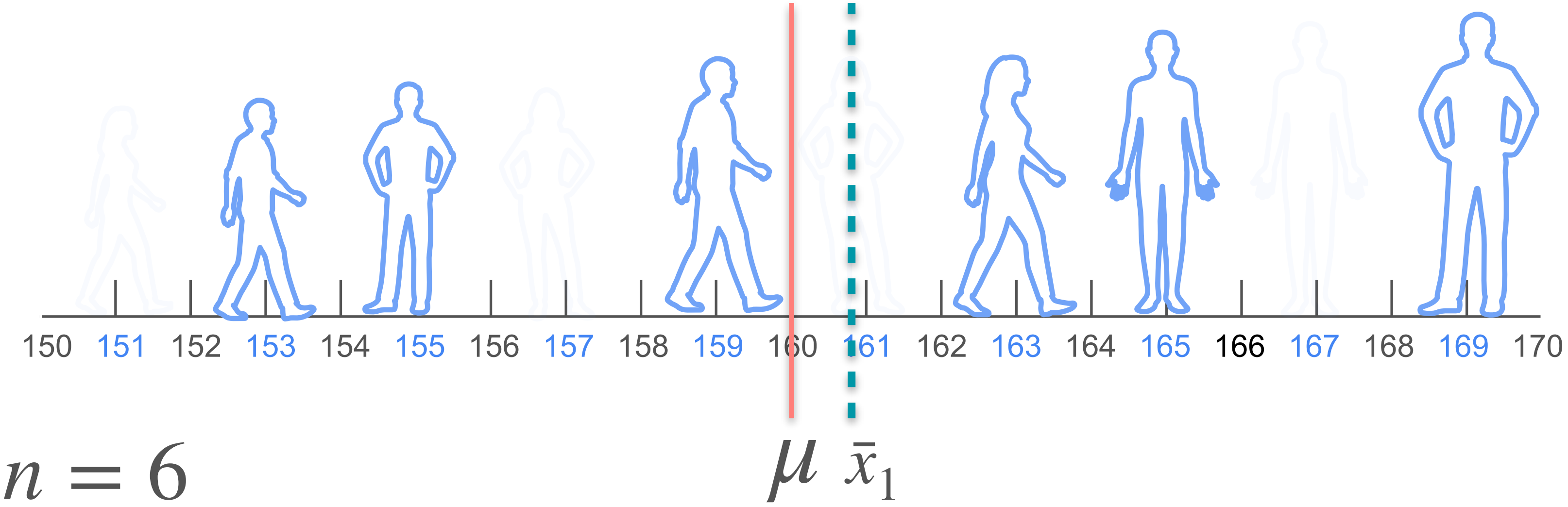


Population and Sample Mean



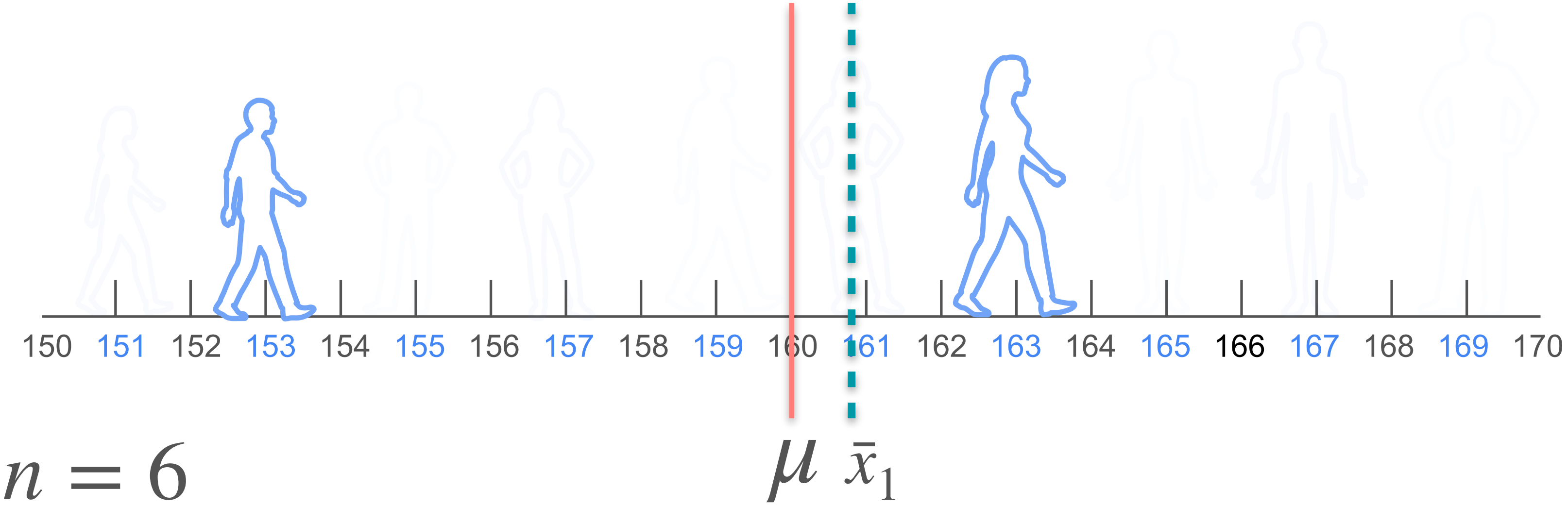
$n = 6$

Population and Sample Mean



$$n = 6$$
$$n = 2$$

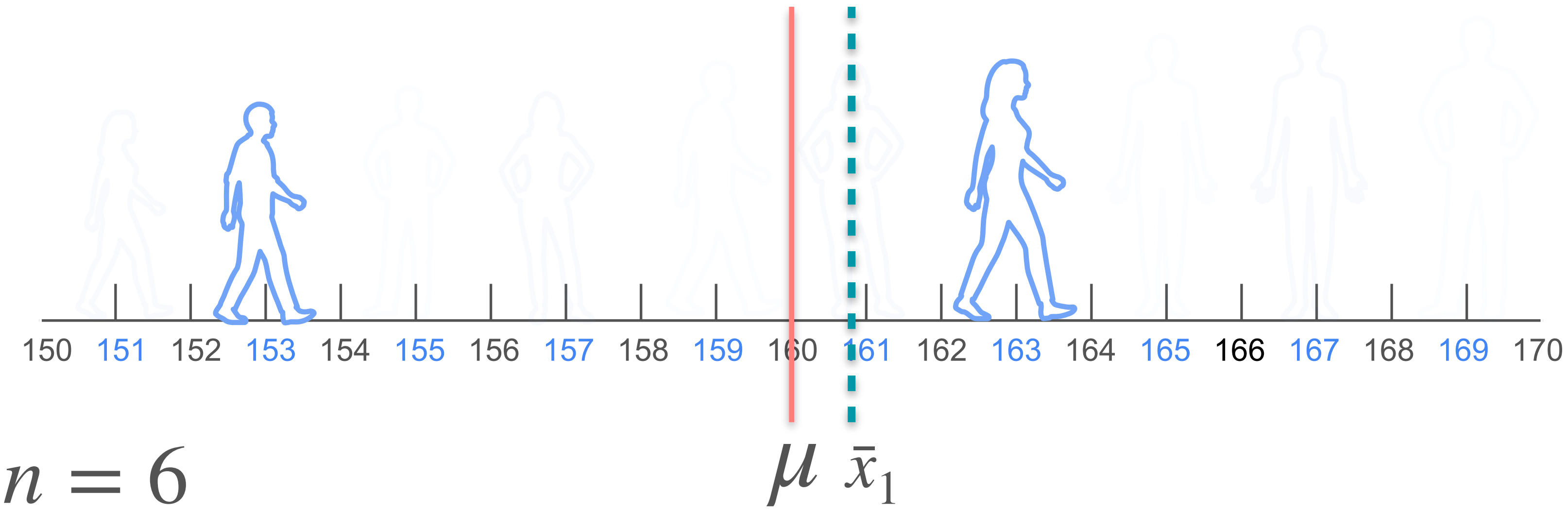
Population and Sample Mean



$n = 6$

$n = 2$

Population and Sample Mean

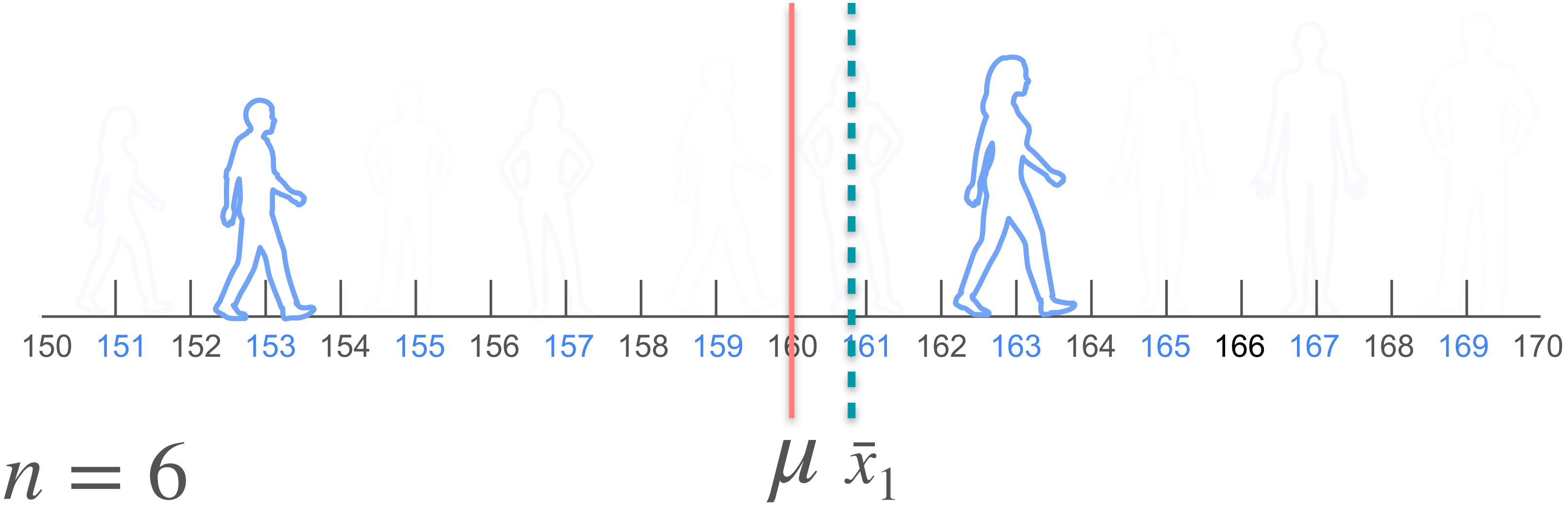


$n = 6$

$n = 2$

What is the average height in statistopia?

Population and Sample Mean



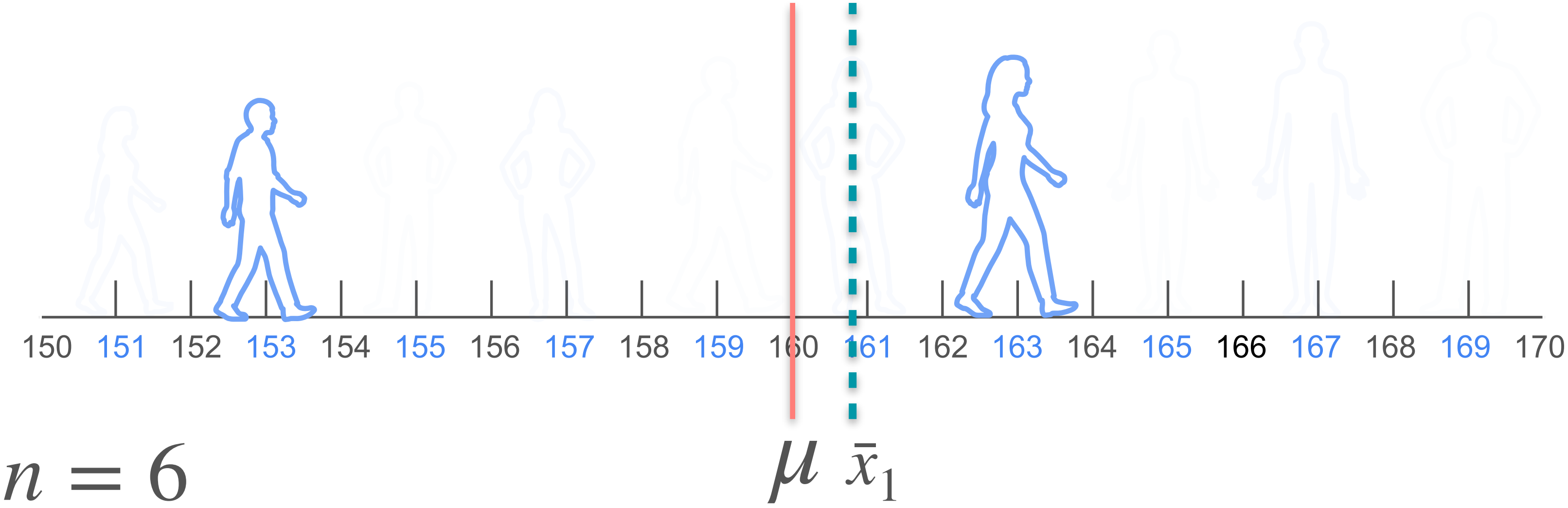
$$n = 6$$

$$n = 2$$

What is the average height in statistopia?

$$\frac{153 + 163}{2}$$

Population and Sample Mean



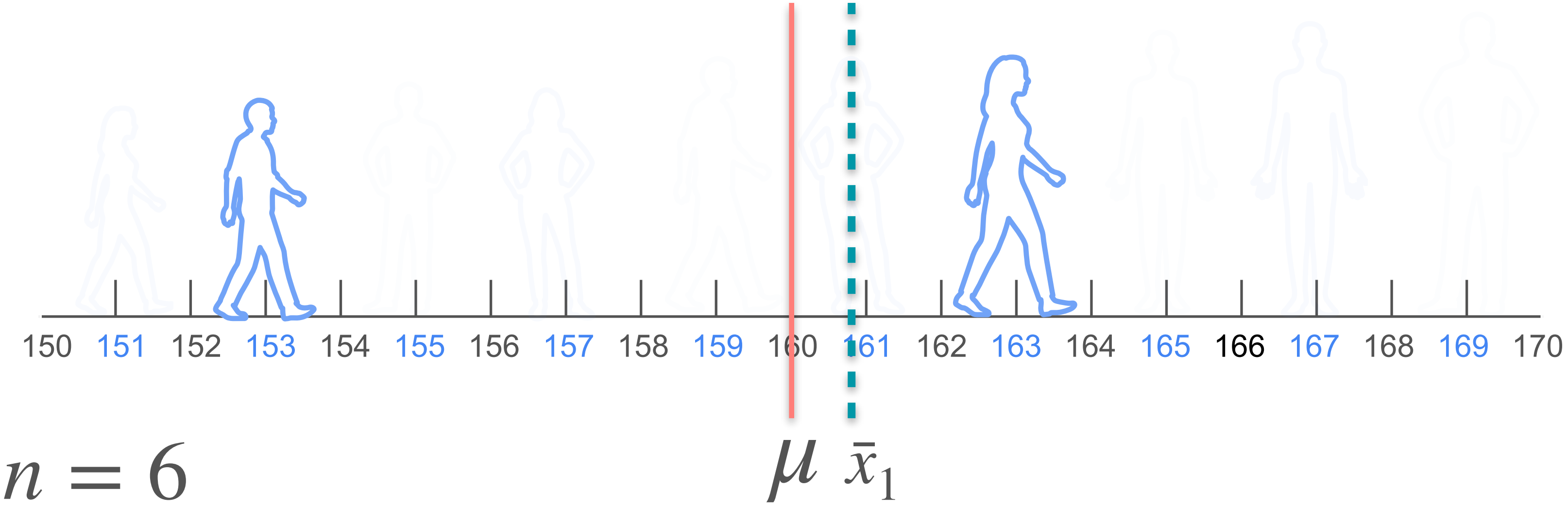
$n = 6$

$n = 2$

What is the average height in statistopia?

$$\frac{153 + 163}{2} = \frac{316}{2} = 158cm$$

Population and Sample Mean



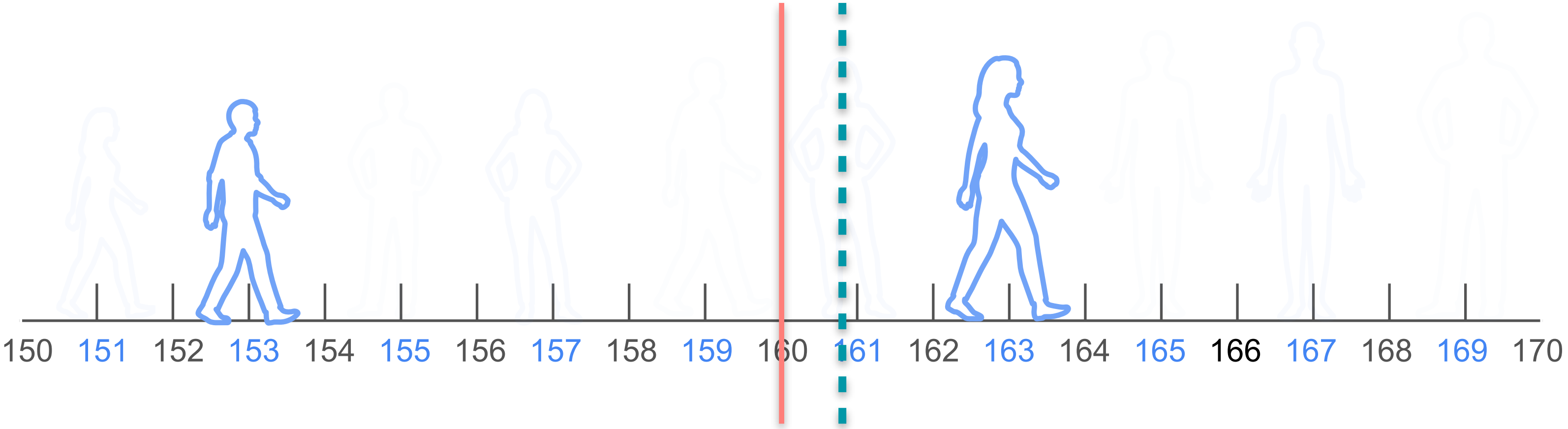
$n = 6$

$n = 2$

What is the average height in statistopia?

$$\frac{153 + 163}{2} = \frac{316}{2} = \boxed{158cm}$$

Population and Sample Mean



$\bar{x}_3$ μ $\bar{x}_1$

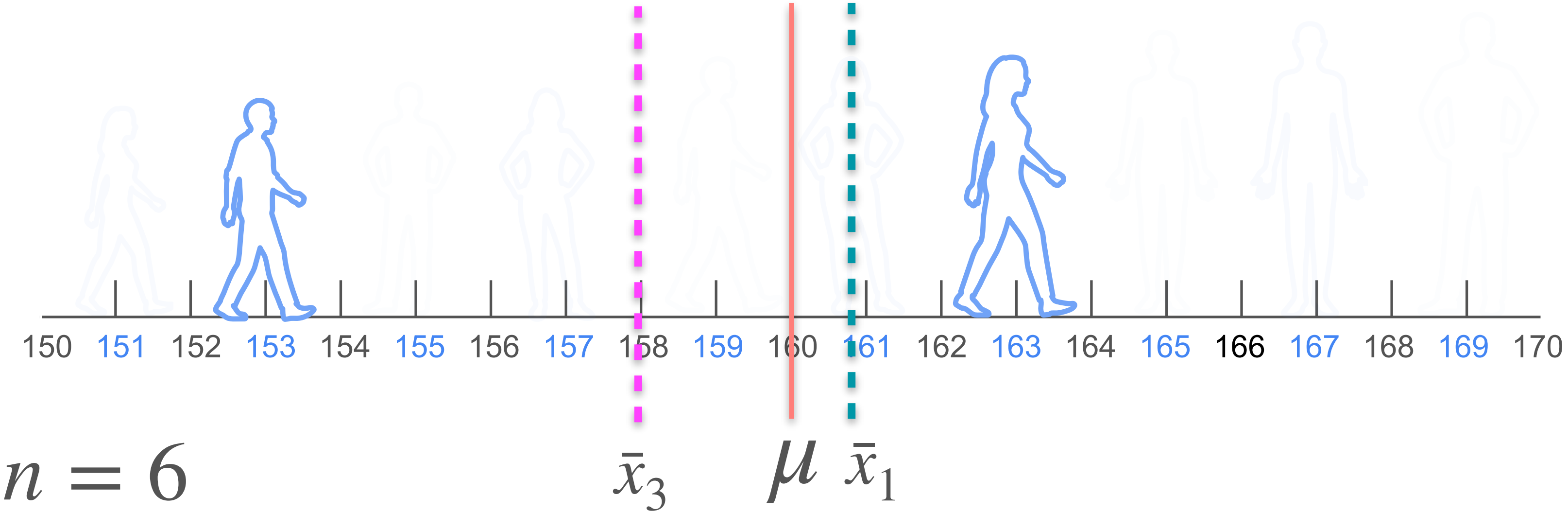
$n = 6$

$n = 2$

What is the average height in statistopia?

$$\frac{153 + 163}{2} = \frac{316}{2} = \boxed{158cm}$$

Population and Sample Mean



$n = 6$

$n = 2$

What is the average height in statistopia?

$$\frac{153 + 163}{2} = \frac{316}{2} = \boxed{158cm}$$

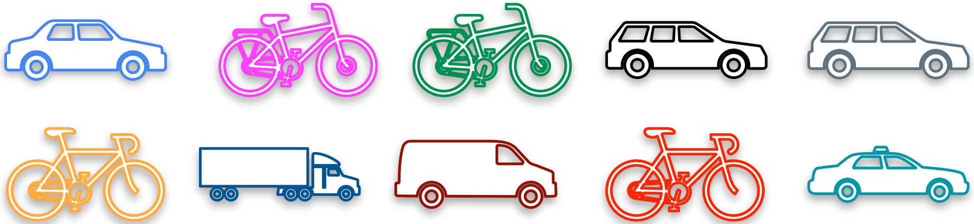
Proportion

Proportion

Population size: 10

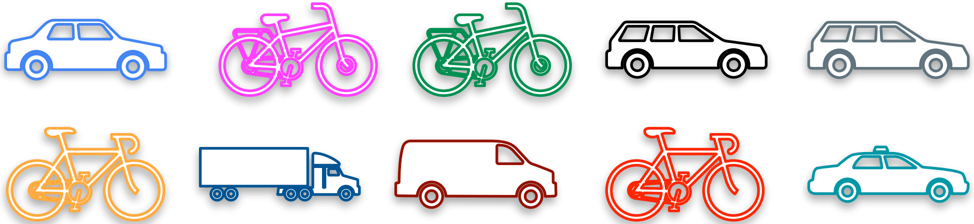
Proportion

Population size: 10



Proportion

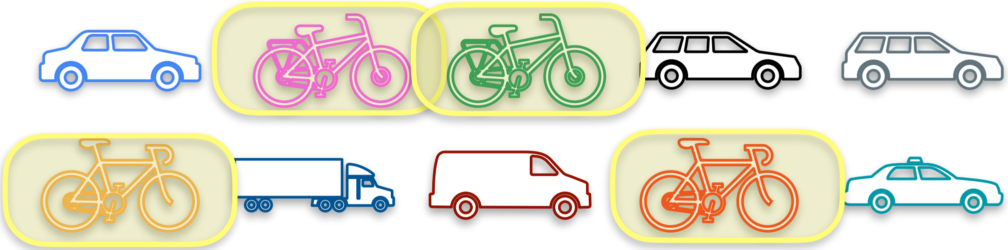
Population size: 10



What proportion of people own a bicycle?

Proportion

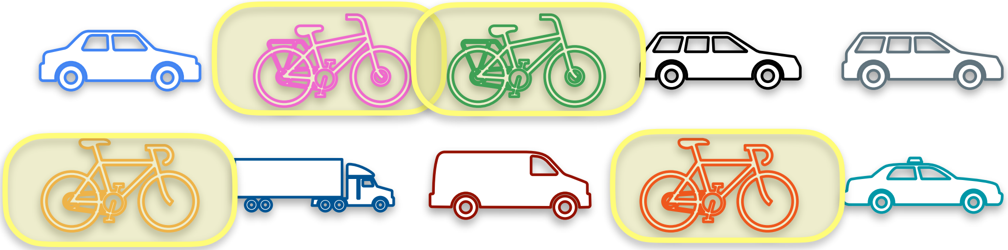
Population size: 10



What proportion of people own a bicycle?

Proportion

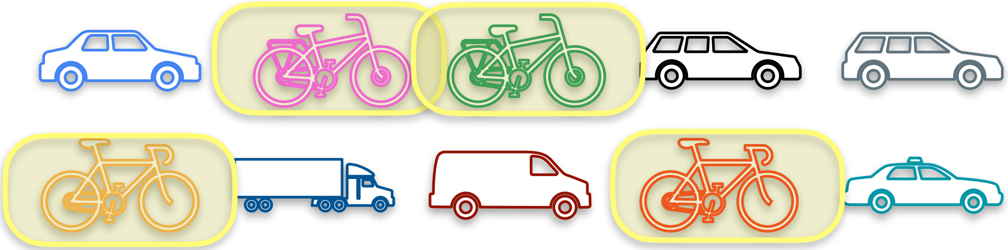
Population size: 10

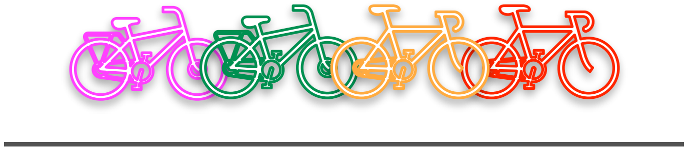


What proportion of people own a bicycle?

Proportion

Population size: 10

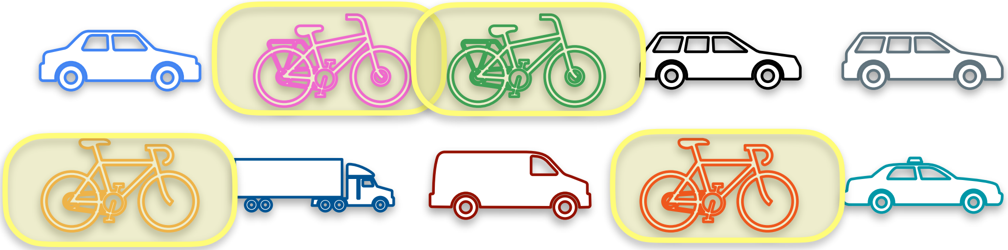


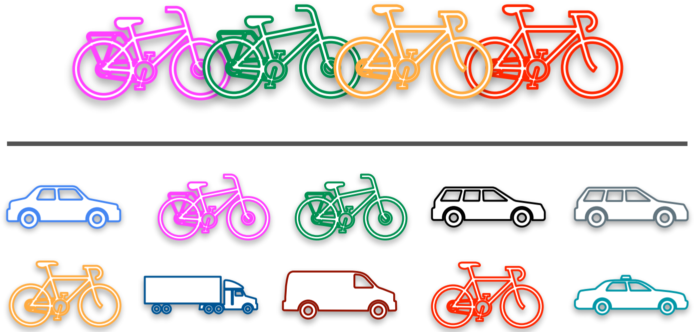


What proportion of people own a bicycle?

Proportion

Population size: 10

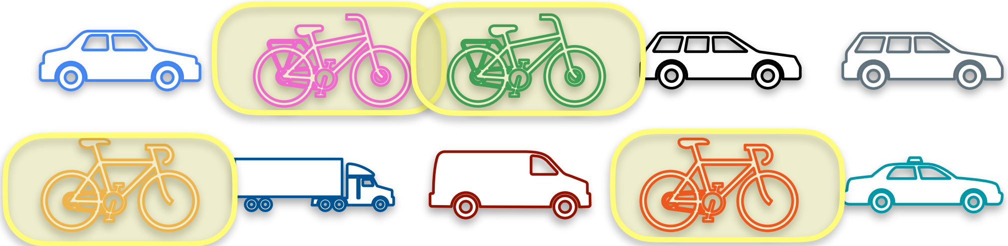




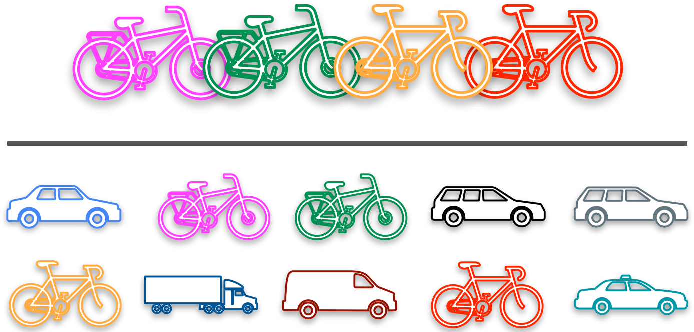
What proportion of people own a bicycle?

Proportion

Population size: 10



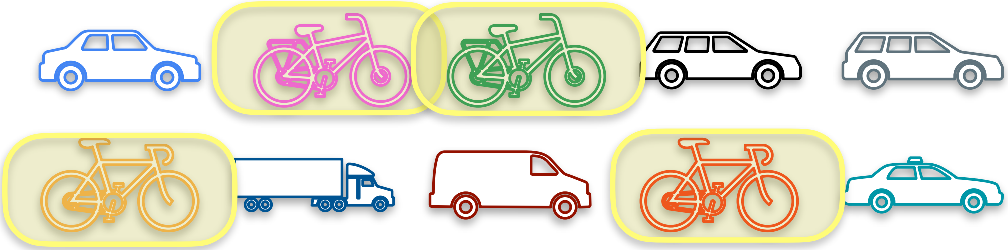
What proportion of people own a bicycle?



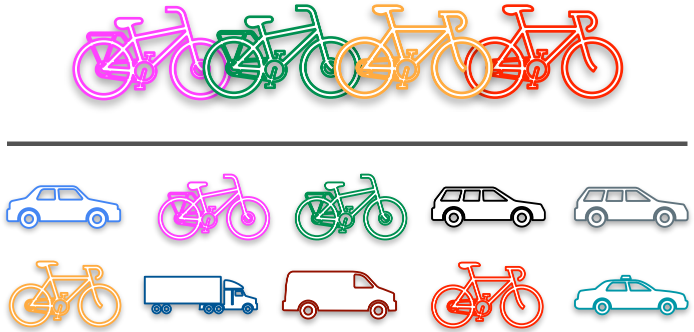
$$= \frac{4}{10} = 0.4 = 40\,\%$$

Proportion

Population size: 10



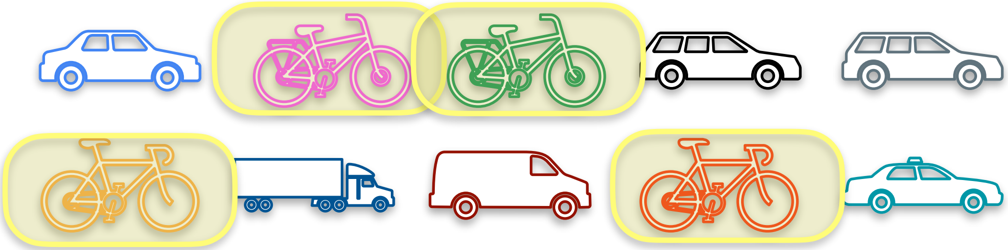
What proportion of people own a bicycle?

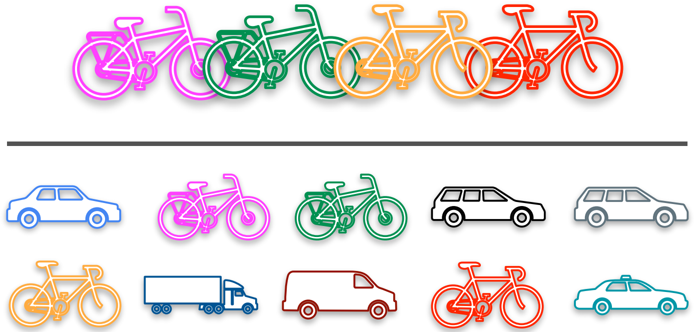


$$\text{population proportion} = \frac{4}{10} = 0.4 = 40\,\%$$

Proportion

Population size: 10





What proportion of people own a bicycle?

$$p \quad \text{population proportion} = \frac{4}{10} = 0.4 = 40\,\%$$

Proportion

$$p = \frac{\text{number of items with a given characteristic } (x)}{\text{population } (n)}$$

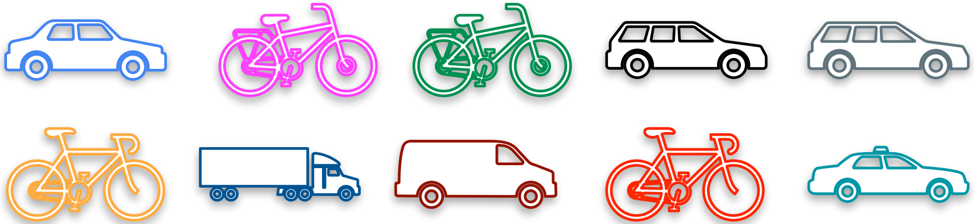
Proportion

population proportion

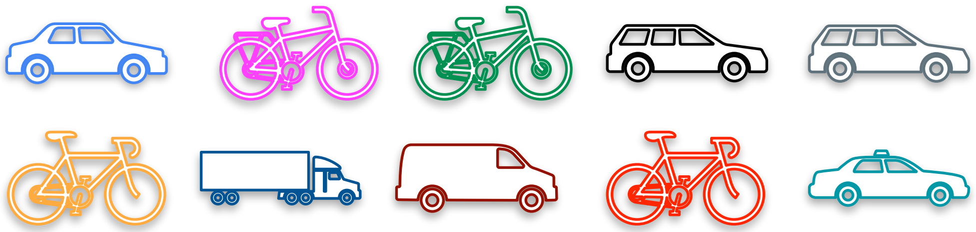
$$p = \frac{\text{number of items with a given characteristic } (x)}{\text{population } (n)}$$

Proportion

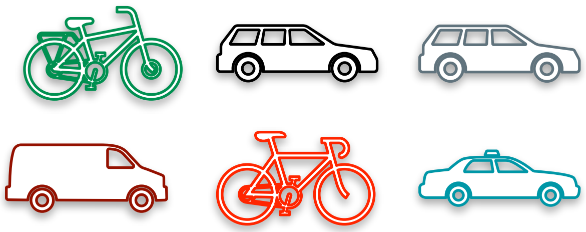
Population size: 10



Proportion

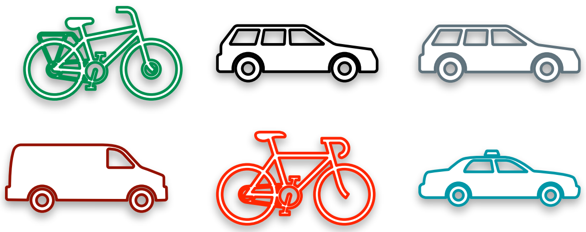


Sample Proportion



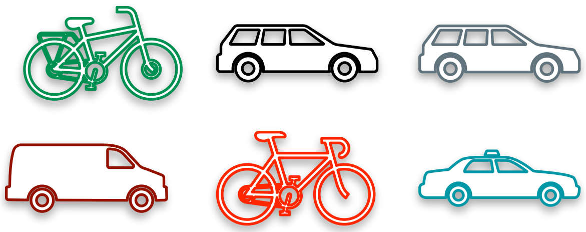
Sample Proportion

Sample size: 6



Sample Proportion

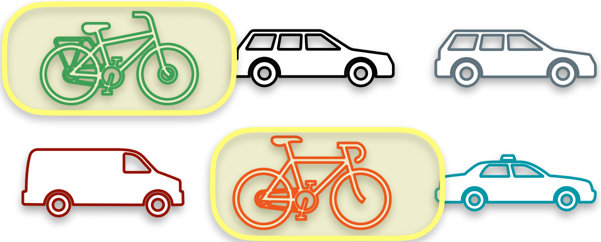
Sample size: 6



What proportion of people own a bicycle?

Sample Proportion

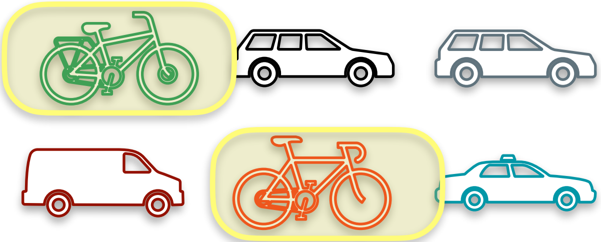
Sample size: 6



What proportion of people own a bicycle?

Sample Proportion

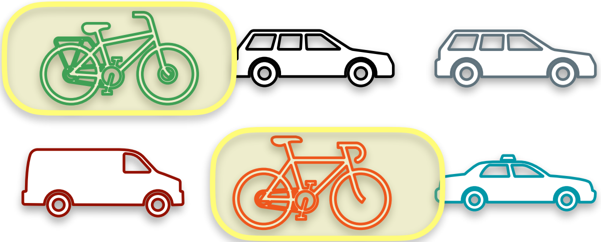
Sample size: 6



What proportion of people own a bicycle?

Sample Proportion

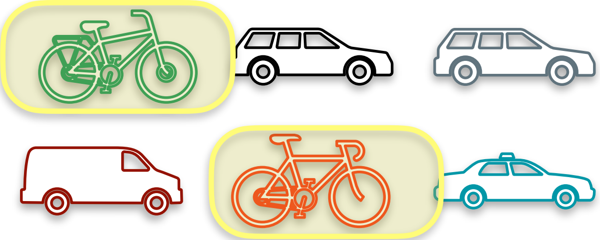
Sample size: 6



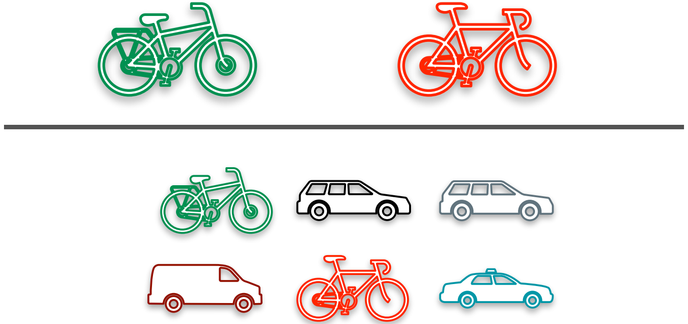
What proportion of people own a bicycle?

Sample Proportion

Sample size: 6

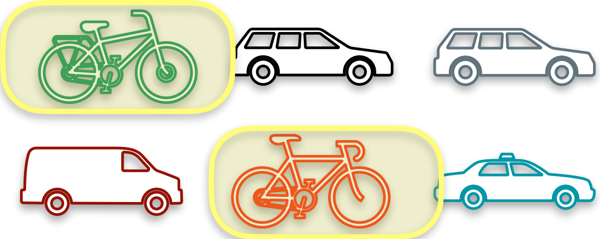


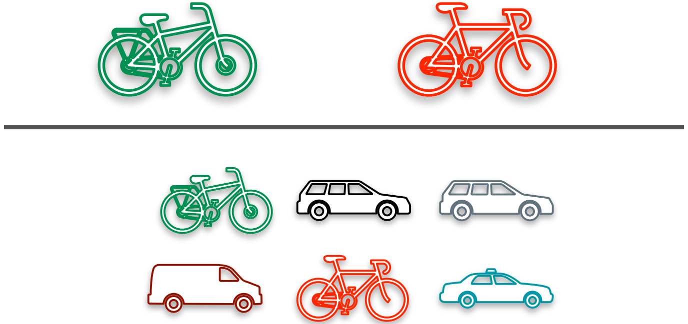
What proportion of people own a bicycle?



Sample Proportion

Sample size: 6



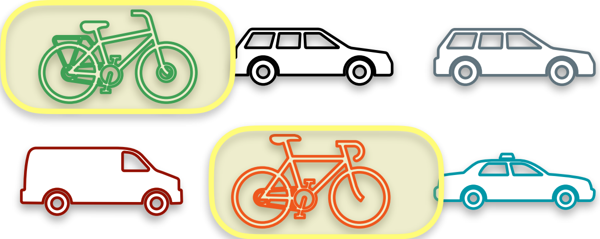


$$= \frac{2}{6} = 0.333 = 33.3\,\%$$

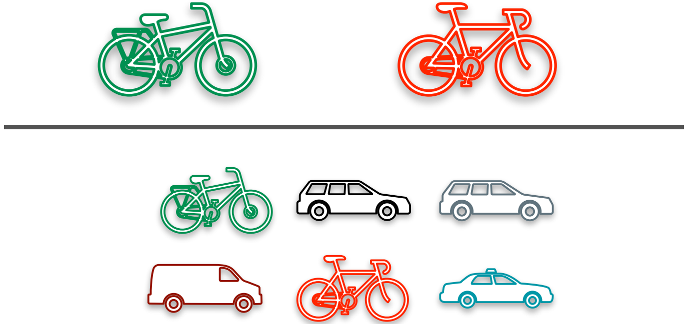
What proportion of people own a bicycle?

Sample Proportion

Sample size: 6



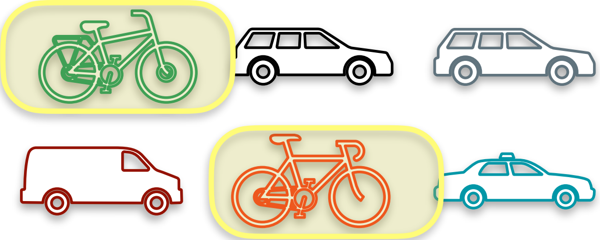
What proportion of people own a bicycle?



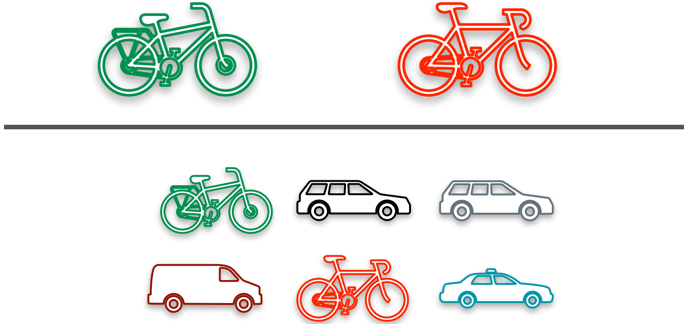
sample proportion $= \dfrac{2}{6} = 0.333 = 33.3\,\%$

Sample Proportion

Sample size: 6



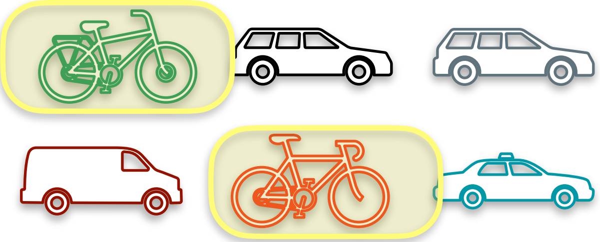
What proportion of people own a bicycle?

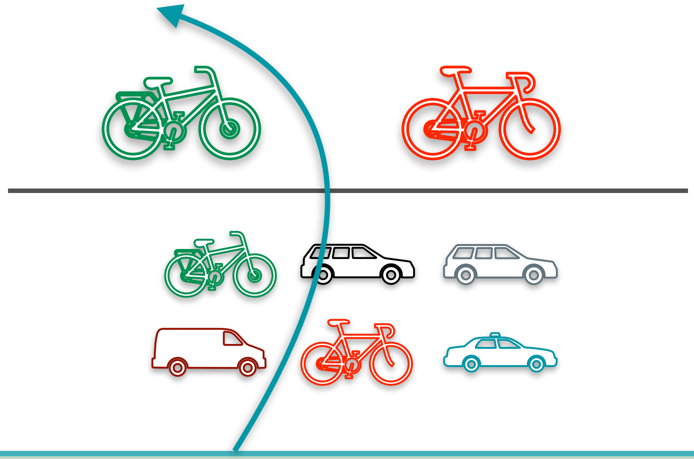


$$\hat{p} \quad \text{sample proportion} \quad = \frac{2}{6} = 0.333 = 33.3\,\%$$

Sample Proportion

Sample size: 6





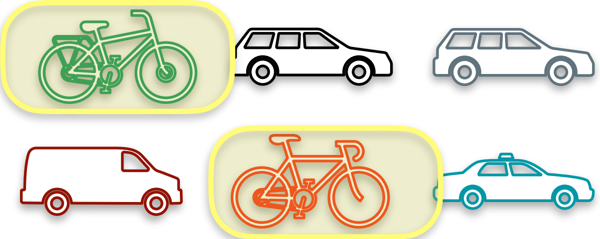
What proportion of people own a bicycle?

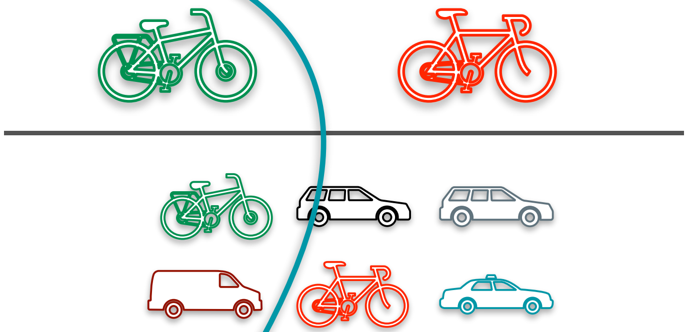
$$\hat{p} \quad \text{sample proportion} \quad = \frac{2}{6} = 0.333 = 33.3\,\%$$

Sample Proportion

Sample size: 6

estimate of the population proportion



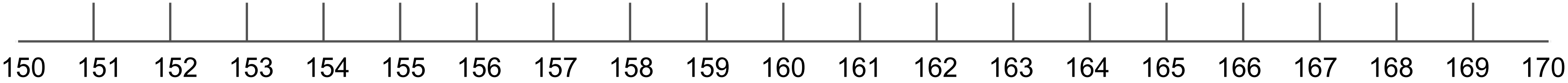


What proportion of people own a bicycle?

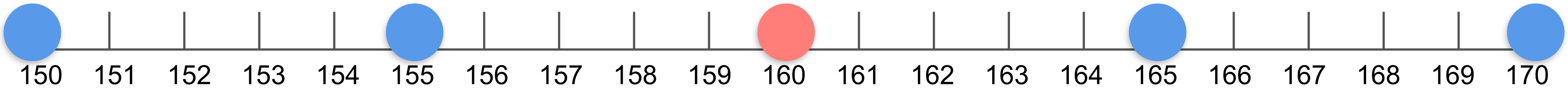
$\hat{p}$ sample proportion $= \frac{2}{6} = 0.333 = 33.3\,\%$

Sample Variance

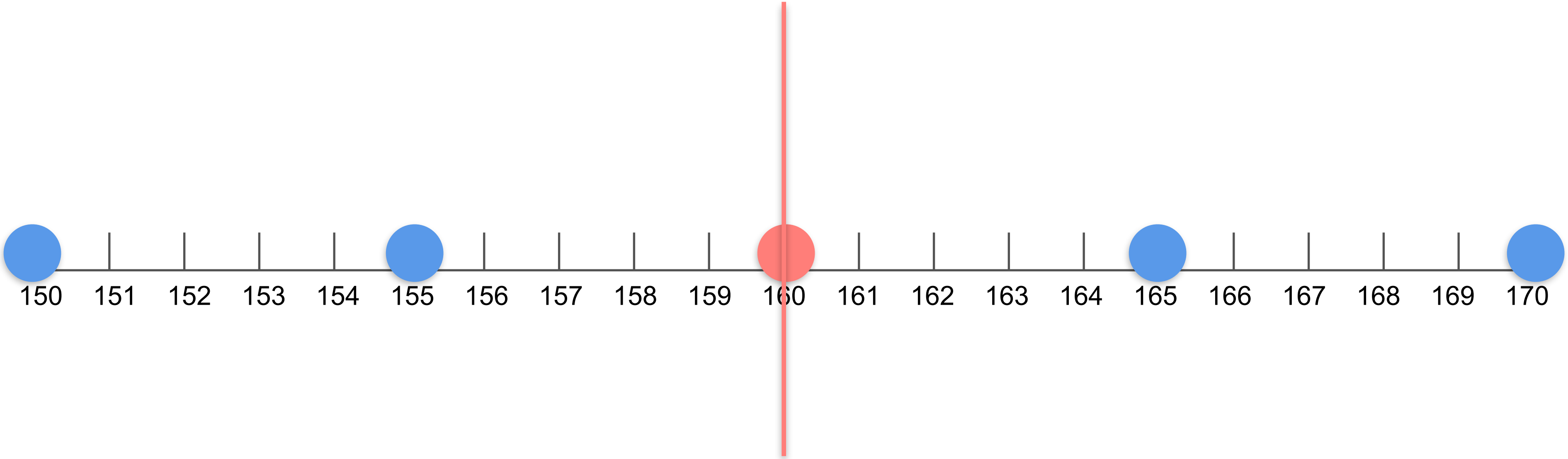
Sample Variance



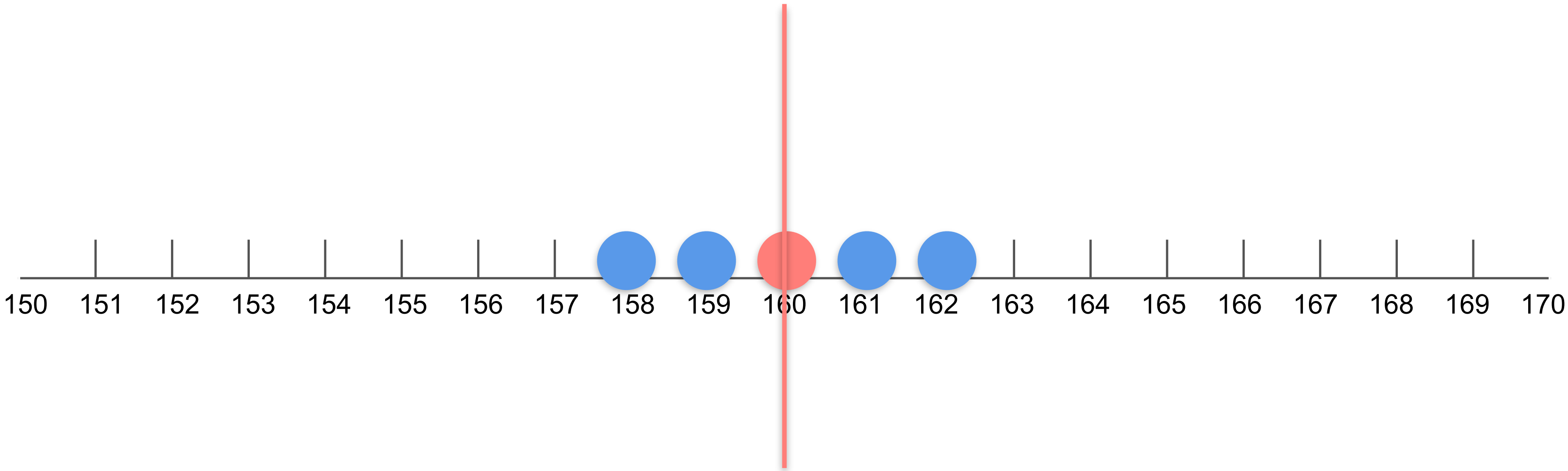
Sample Variance



Sample Variance

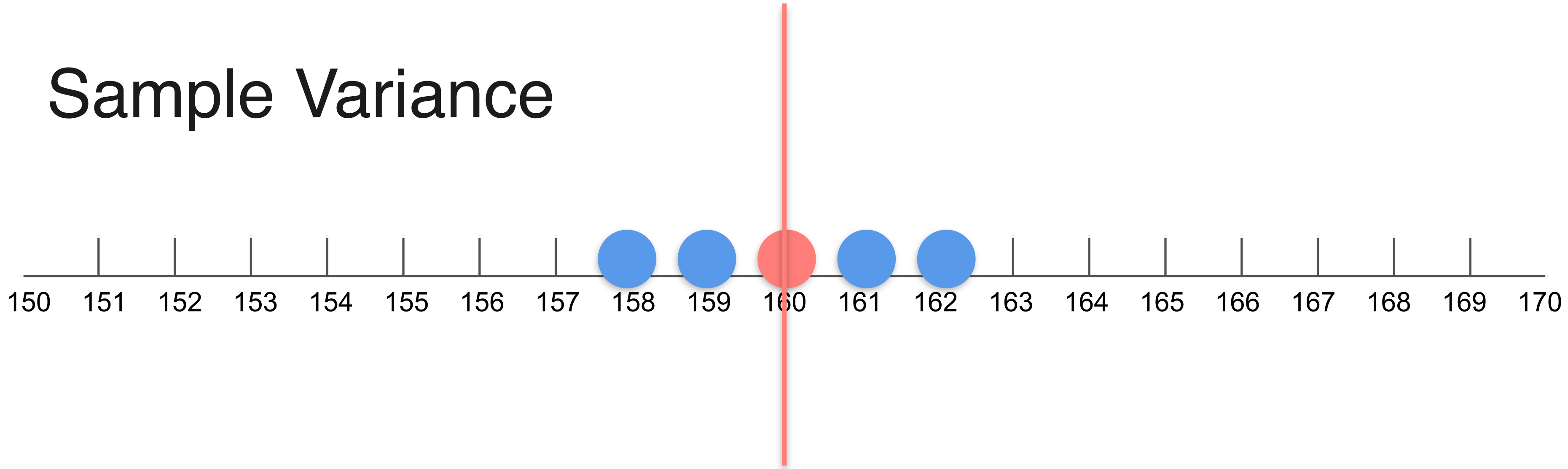


Sample Variance

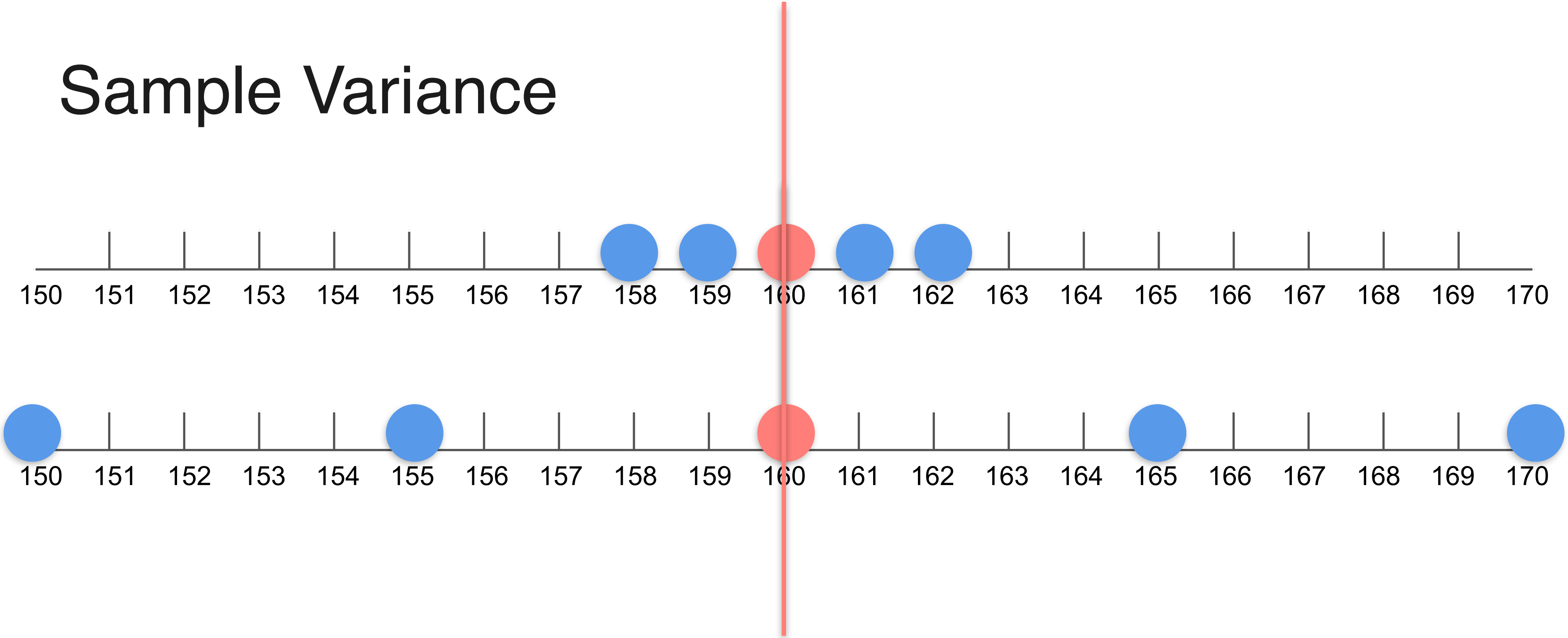


DeepLearning.AI

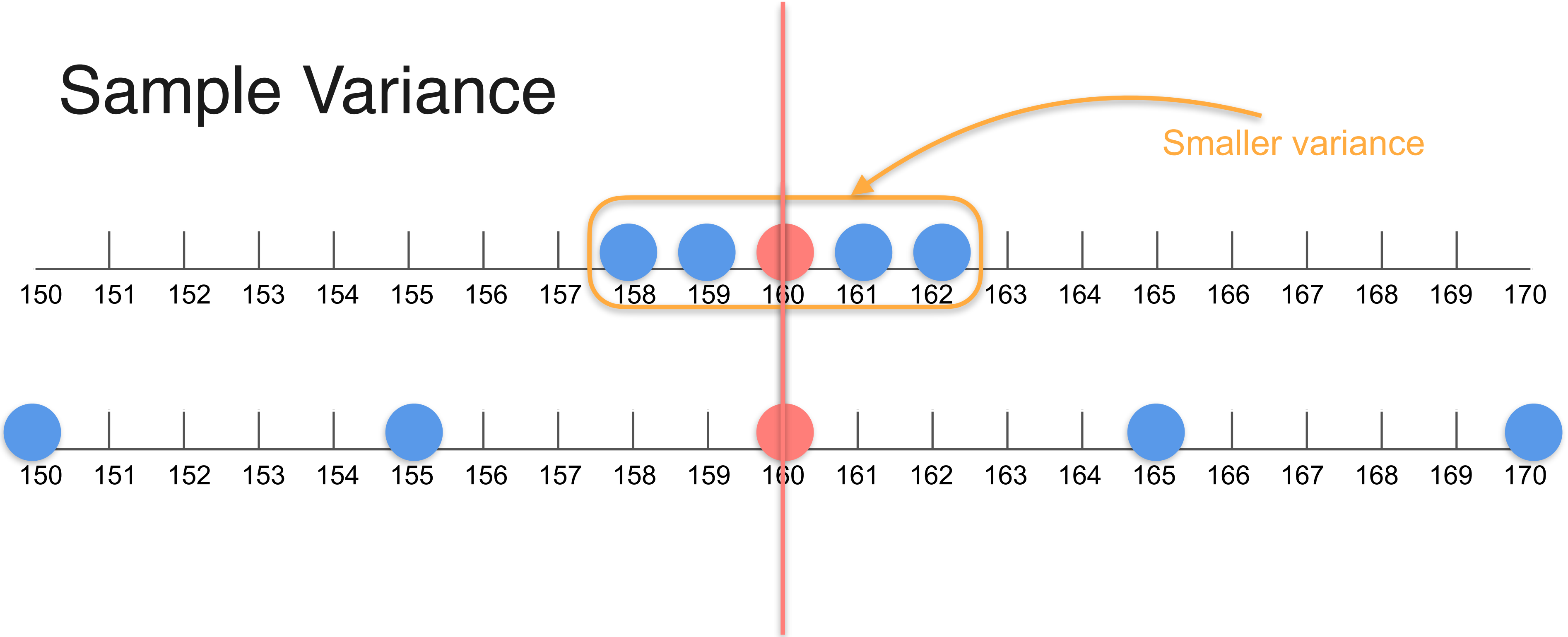
Sample Variance



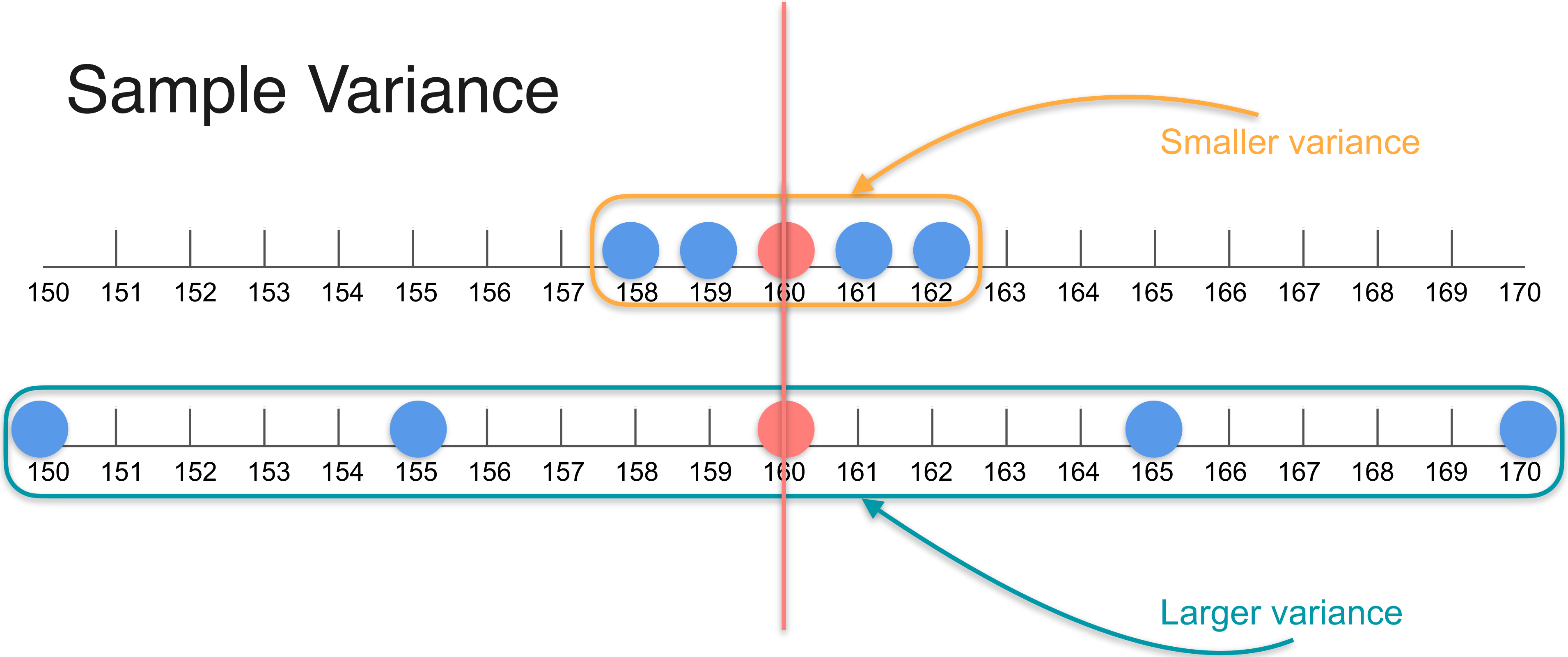
Sample Variance



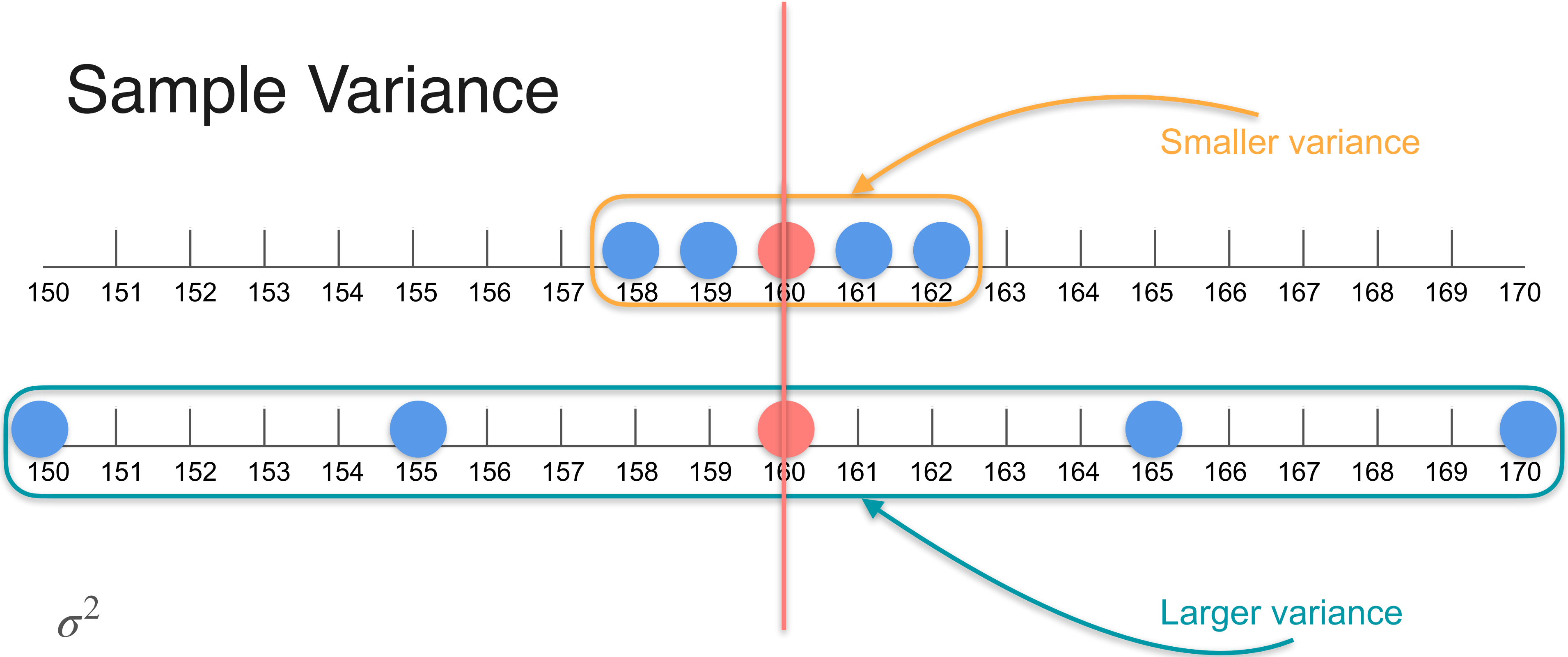
Sample Variance



Sample Variance

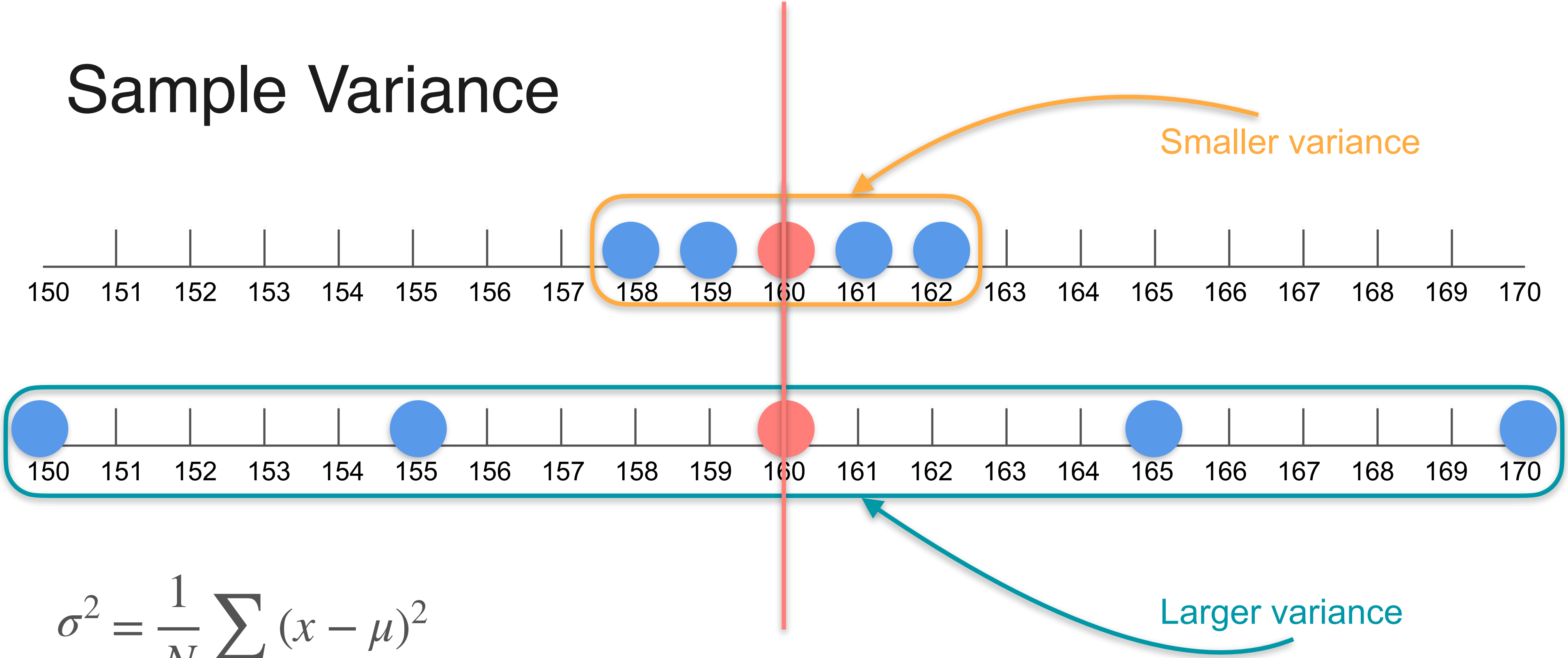


Sample Variance



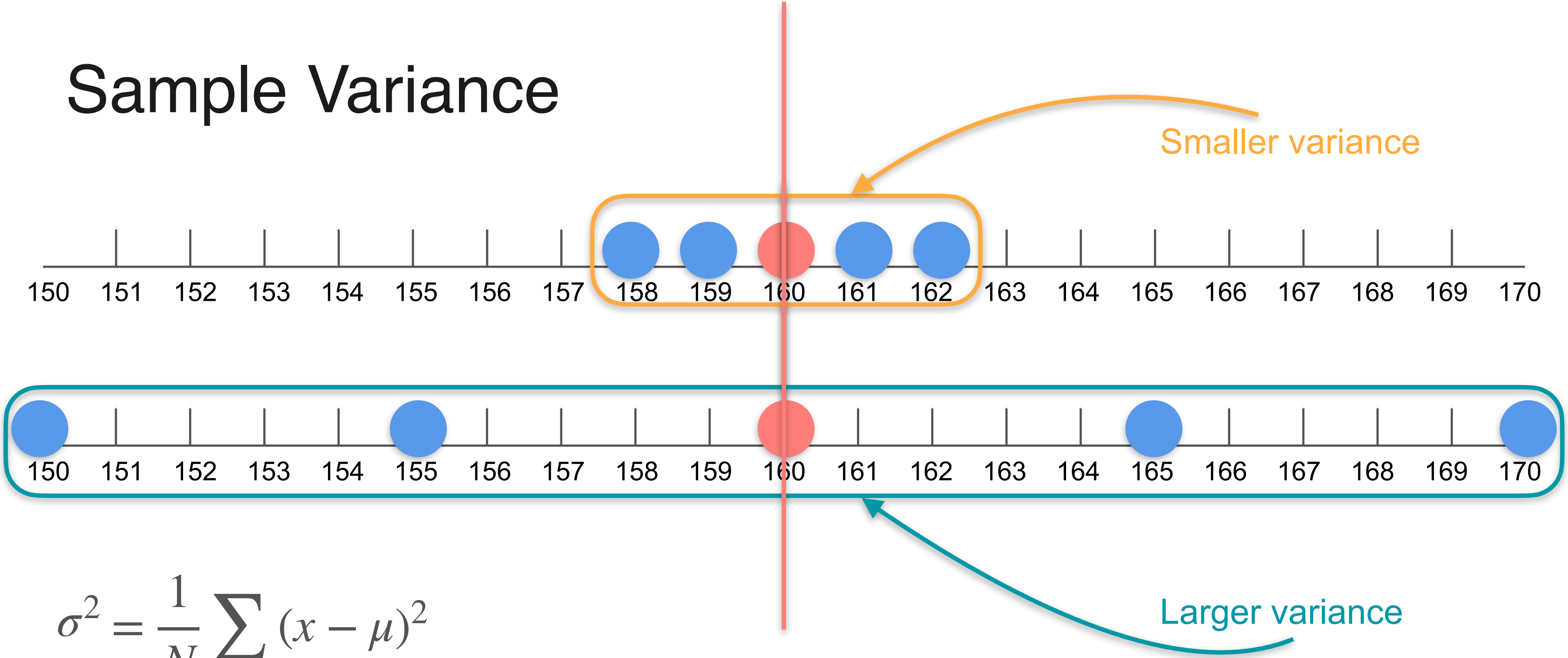
σ^2

Sample Variance



$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

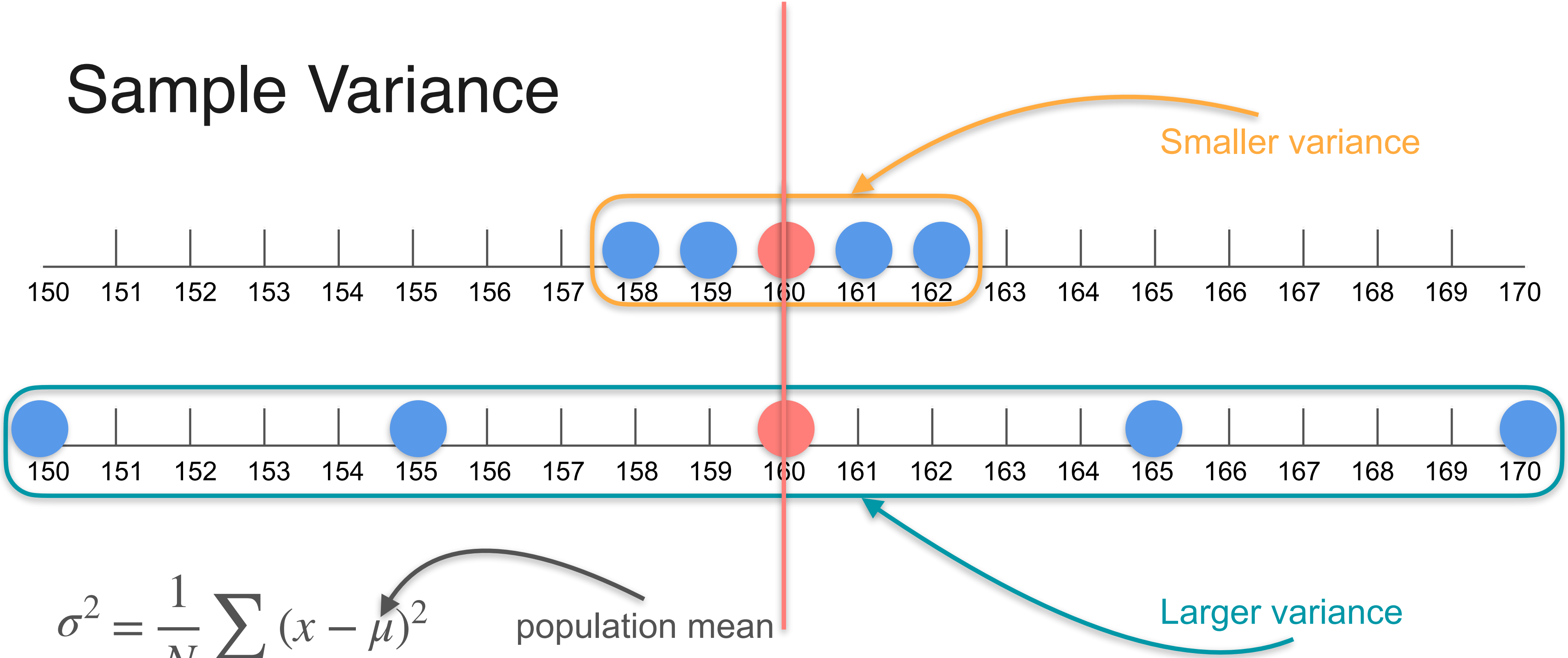
Sample Variance



$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

population size

Sample Variance

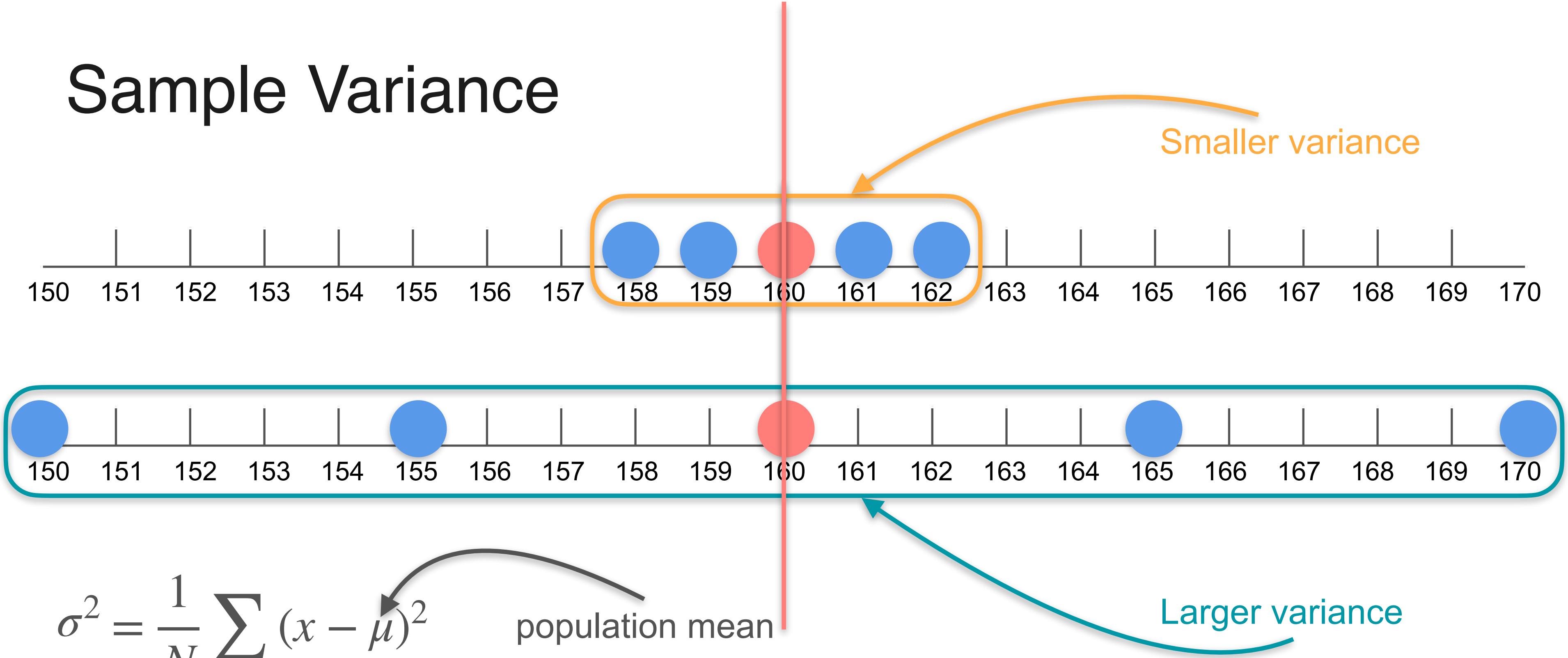


$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

population mean

population size

Sample Variance



$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

population mean

population size

How to estimate population variance with the sample?

Variance Estimation

Variance Estimation

Variance Estimation

1	2	3

μ

Variance Estimation

1	2	3

$$\mu = \frac{1+2+3}{3}$$

Variance Estimation

1	2	3

$$\mu = \frac{1+2+3}{3} = \frac{6}{3}$$

Variance Estimation

1	2	3

$$\mu = \frac{1+2+3}{3} = \frac{6}{3} = 2$$

Variance Estimation

1	2	3

$$\mu = 2$$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

Variance Estimation

1	2	3

$\mu = 2$

$x \qquad x - \mu \qquad (x-\mu)^2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

x	$x - \mu$	$(x - \mu)^2$
1		
2		
3		

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

x
1
2
3

$x - 2$

$x - \mu$ $\quad (x - \mu)^2$

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$x - 2$

x	$x - \mu$	$(x - \mu)^2$
1	-1	
2	0	
3	1	

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$x - 2$

x	$x - \mu$	$(x-\mu)^2$
1	-1	1
2	0	0
3	1	1

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$x - 2$

x	$x - \mu$	$(x - \mu)^2$
1	-1	1
2	0	0
3	1	1

$$\sum (x - \mu)^2$$

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x-\mu)^2$$

$x - 2$

x	$x-\mu$	$(x-\mu)^2$
1	-1	1
2	0	0
3	1	1

$$\sum (x-\mu)^2 = 2$$

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x-\mu)^2$$

$x - 2$

x	$x-\mu$	$(x-\mu)^2$
1	-1	1
2	0	0
3	1	1

$$\frac{\sum (x-\mu)^2}{N} \quad = 2$$

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x-\mu)^2$$

$x - 2$

x	$x-\mu$	$(x-\mu)^2$
1	-1	1
2	0	0
3	1	1

$$\frac{\sum (x-\mu)^2}{N} = \frac{2}{3}$$

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$x - 2$

x	$x-\mu$	$(x-\mu)^2$
1	-1	1
2	0	0
3	1	1

$$\frac{\sum (x-\mu)^2}{N} = \frac{2}{3}$$

σ^2

Population variance

Variance Estimation

1	2	3

$$\mu = 2$$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

Variance Estimation

$n = 2$

Samples

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples

1	1
1	2
1	3
2	1
2	2
2	3
3	1
3	2
3	3

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples

1	1
1	2
1	3
2	1
2	2
2	3
3	1
3	2
3	3

Variance Estimation

1	2	3

$\mu = 2$

$$\boxed{\sigma^2 = \frac{1}{N} \sum (x - \mu)^2}$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples

1	1
1	2
1	3
2	1
2	2
2	3
3	1
3	2
3	3

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples

1	1
1	2
1	3
2	1
2	2
2	3
3	1
3	2
3	3

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples

1	1
1	2
1	3
2	1
2	2
2	3
3	1
3	2
3	3

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples

1	1
1	2
1	3
2	1
2	2
2	3
3	1
3	2
3	3

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$
1	1	
1	2	
1	3	
2	1	
2	2	
2	3	
3	1	
3	2	
3	3	

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$
1	1	1
1	2	1,5
1	3	2
2	1	1,5
2	2	2
2	3	2,5
3	1	2
3	2	2,5
3	3	3

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$
1	1	1
1	2	1,5
1	3	2
2	1	1,5
2	2	2
2	3	2,5
3	1	2
3	2	2,5
3	3	3

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$
1	1	1
1	2	1,5
1	3	2
2	1	1,5
2	2	2
2	3	2,5
3	1	2
3	2	2,5
3	3	3

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$
1	1	1
1	2	1,5
1	3	2
2	1	1,5
2	2	2
2	3	2,5
3	1	2
3	2	2,5
3	3	3

$$Var(x) = \frac{\sum (x-\bar{x})^2}{n}$$

Variance Estimation

1	2	3

$\mu = 2$

$$\boxed{\sigma^2 = \frac{1}{N}\sum(x-\mu)^2}$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

$$Var(x) = \frac{\sum(x-\bar{x})^2}{n}$$

Samples		$\bar{x}$
1	1	1
1	2	1,5
1	3	2
2	1	1,5
2	2	2
2	3	2,5
3	1	2
3	2	2,5
3	3	3

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$	$Var(x) = \frac{\sum (x-\bar{x})^2}{n}$
1	1	1	0
1	2	1,5	0,25
1	3	2	1
2	1	1,5	0,25
2	2	2	0
2	3	2,5	0,25
3	1	2	1
3	2	2,5	0,25
3	3	3	0

Variance Estimation

1	2	3

$\mu = 2$

$$\boxed{\sigma^2 = \frac{1}{N}\sum (x-\mu)^2}$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

$$Var(x) = \frac{\sum (x-\bar{x})^2}{n}$$

Samples		$\bar{x}$	$Var(x)$
1	1	1	0
1	2	1,5	0,25
1	3	2	1
2	1	1,5	0,25
2	2	2	0
2	3	2,5	0,25
3	1	2	1
3	2	2,5	0,25
3	3	3	0

Variance Estimation

1	2	3

$\mu = 2$

$$\boxed{\sigma^2 = \frac{1}{N}\sum (x-\mu)^2}$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$	$Var(x) = \frac{\sum (x-\bar{x})^2}{n}$
1	1	1	0
1	2	1,5	0,25
1	3	2	1
2	1	1,5	0,25
2	2	2	0
2	3	2,5	0,25
3	1	2	1
3	2	2,5	0,25
3	3	3	0

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$	$Var(x) = \frac{\sum (x-\bar{x})^2}{n}$
1	1	1	0
1	2	1,5	0,25
1	3	2	1
2	1	1,5	0,25
2	2	2	0
2	3	2,5	0,25
3	1	2	1
3	2	2,5	0,25
3	3	3	0

estimated variance

$= 0.333$

$= \frac{1}{3}$

Variance Estimation

1	2	3

$\mu = 2$

$$\boxed{\sigma^2 = \frac{1}{N} \sum (x - \mu)^2}$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

$$Var(x) = \frac{\sum (x - \bar{x})^2}{n}$$

Samples		$\bar{x}$	
1	1	1	0
1	2	1,5	0,25
1	3	2	1
2	1	1,5	0,25
2	2	2	0
2	3	2,5	0,25
3	1	2	1
3	2	2,5	0,25
3	3	3	0

estimated variance

$= 0.333$

$= \frac{1}{3}$

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$	$Var(x) = \frac{\sum (x-\bar{x})^2}{n}$
1	1	1	0
1	2	1,5	0,25
1	3	2	1
2	1	1,5	0,25
2	2	2	0
2	3	2,5	0,25
3	1	2	1
3	2	2,5	0,25
3	3	3	0

estimated variance

$= 0.333$

$= \frac{1}{3}$

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$	$Var(x) = \frac{\sum (x-\bar{x})^2}{n}$
1	1	1	0
1	2	1,5	0,25
1	3	2	1
2	1	1,5	0,25
2	2	2	0
2	3	2,5	0,25
3	1	2	1
3	2	2,5	0,25
3	3	3	0

estimated variance

$= 0.333$

$= \frac{1}{3}$

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

$$Var(x) = \frac{\sum (x - \bar{x})^2}{n}$$

Samples		$\bar{x}$	$Var(x)$
1	1	1	0
1	2	1,5	0,25
1	3	2	1
2	1	1,5	0,25
2	2	2	0
2	3	2,5	0,25
3	1	2	1
3	2	2,5	0,25
3	3	3	0

estimated variance

$= 0.333$

$= \frac{1}{3}$

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$	$Var(x) = \frac{\sum (x - \bar{x})^2}{n}$
1	1	1	0
1	2	1,5	0,25
1	3	2	1
2	1	1,5	0,25
2	2	2	0
2	3	2,5	0,25
3	1	2	1
3	2	2,5	0,25
3	3	3	0

estimated variance

$$= 0.333$$

$$= \frac{1}{3}$$

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$	$Var(x) = \frac{\sum (x-\bar{x})^2}{n}$
1	1	1	
1	2	1,5	
1	3	2	
2	1	1,5	
2	2	2	
2	3	2,5	
3	1	2	
3	2	2,5	
3	3	3	

Variance Estimation

1	2	3

$\mu = 2$

$$\boxed{\sigma^2 = \frac{1}{N}\sum (x-\mu)^2}$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$	$Var(x) = \frac{\sum (x-\bar{x})^2}{n-1}$
1	1	1	
1	2	1,5	
1	3	2	
2	1	1,5	
2	2	2	
2	3	2,5	
3	1	2	
3	2	2,5	
3	3	3	

Variance Estimation

1	2	3

$\mu = 2$

$$\boxed{\sigma^2 = \frac{1}{N}\sum (x-\mu)^2}$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$	$Var(x) = \frac{\sum (x-\bar{x})^2}{n-1}$
1	1	1	
1	2	1,5	
1	3	2	
2	1	1,5	
2	2	2	
2	3	2,5	
3	1	2	
3	2	2,5	
3	3	3	

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$ Samples		$\bar{x}$	$Var(x) = \frac{\sum (x - \bar{x})^2}{n - 1}$
1	1	1	0
1	2	1,5	0,5
1	3	2	2
2	1	1,5	0,5
2	2	2	0
2	3	2,5	0,5
3	1	2	2
3	2	2,5	0,5
3	3	3	0

Variance Estimation

1 2 3

$\mu = 2$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$	$Var(x) = \frac{\sum (x - \bar{x})^2}{n - 1}$
1	1	1	0
1	2	1,5	0,5
1	3	2	2
2	1	1,5	0,5
2	2	2	0
2	3	2,5	0,5
3	1	2	2
3	2	2,5	0,5
3	3	3	0

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n=2$ Samples		$\bar{x}$	$Var(x) = \frac{\sum (x-\bar{x})^2}{n-1}$
1	1	1	0
1	2	1,5	0,5
1	3	2	2
2	1	1,5	0,5
2	2	2	0
2	3	2,5	0,5
3	1	2	2
3	2	2,5	0,5
3	3	3	0

estimated variance

$= 0.667$

$= \frac{2}{3}$

DeepLearning.AI

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$	$Var(x) = \frac{\sum (x-\bar{x})^2}{n-1}$
1	1	1	0
1	2	1,5	0,5
1	3	2	2
2	1	1,5	0,5
2	2	2	0
2	3	2,5	0,5
3	1	2	2
3	2	2,5	0,5
3	3	3	0

estimated variance

$= 0.667$

$= \frac{2}{3}$

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n = 2$

Samples		$\bar{x}$	$Var(x) = \frac{\sum (x-\bar{x})^2}{n-1}$
1	1	1	0
1	2	1,5	0,5
1	3	2	2
2	1	1,5	0,5
2	2	2	0
2	3	2,5	0,5
3	1	2	2
3	2	2,5	0,5
3	3	3	0

estimated variance

$= 0.667$

$= \frac{2}{3}$

Variance Estimation

Population Variance Formula

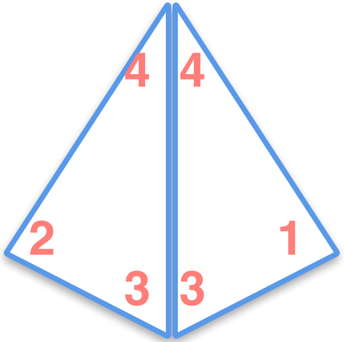
$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

Sample Variance Formula

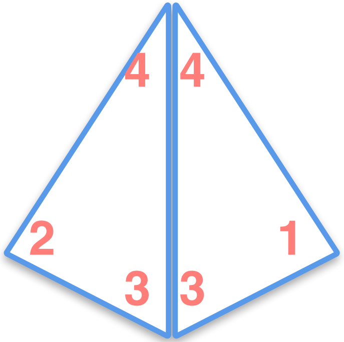
$$Var(x) = \frac{1}{n-1} \sum (x - \bar{x})^2$$

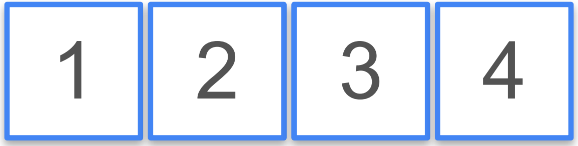
Variance Estimation

Variance Estimation

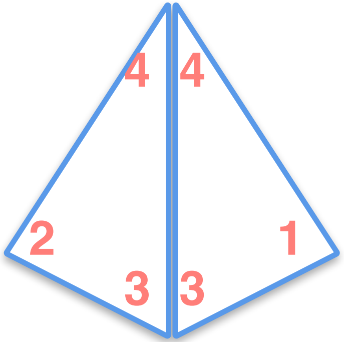


Variance Estimation





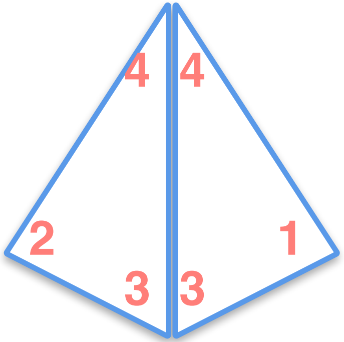
Variance Estimation



1	2	3	4

$$\mu = \frac{1+2+3+4}{4}$$

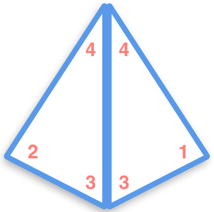
Variance Estimation



1	2	3	4

$$\mu = \frac{1+2+3+4}{4} = 2.5$$

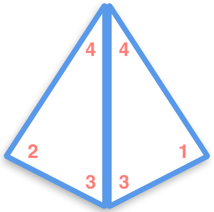
Variance Estimation



1	2	3	4

$$\mu = 2.5$$

Variance Estimation

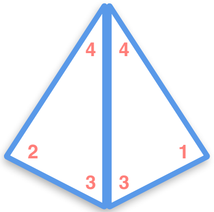


1	2	3	4

$\mu = 2.5$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

Variance Estimation



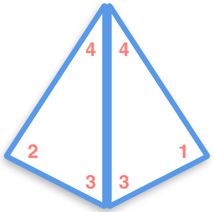
1	2	3	4

$\mu = 2.5$

x $\quad x - \mu \quad$ $(x-\mu)^2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

Variance Estimation



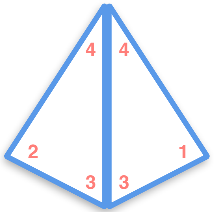
1 2 3 4

$\mu = 2.5$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

x	$x - \mu$	$(x - \mu)^2$
1		
2		
3		
4		

Variance Estimation



1 2 3 4

$\mu = 2.5$

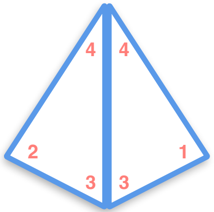
$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

x
1
2
3
4

$x - 2.5$

$x - \mu$ $(x - \mu)^2$

Variance Estimation



1 2 3 4

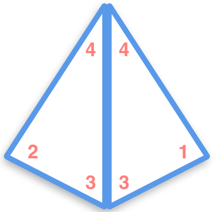
$\mu = 2.5$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$x - 2.5$

x	$x-\mu$	$(x-\mu)^2$
1	-1,5	
2	-0,5	
3	0,5	
4	1,5	

Variance Estimation



1	2	3	4

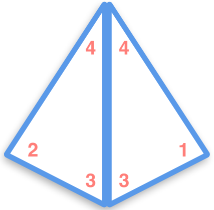
$\mu = 2.5$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$x - 2.5$

x	$x - \mu$	$(x - \mu)^2$
1	-1,5	2,25
2	-0,5	0,25
3	0,5	0,25
4	1,5	2,25

Variance Estimation



1 2 3 4

$\mu = 2.5$

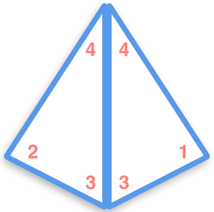
$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$x - 2.5$

x	$x-\mu$	$(x-\mu)^2$
1	-1,5	2,25
2	-0,5	0,25
3	0,5	0,25
4	1,5	2,25

$$\sum (x-\mu)^2$$

Variance Estimation



1 2 3 4

$\mu = 2.5$

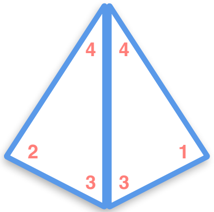
$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$x - 2.5$

x	$x-\mu$	$(x-\mu)^2$
1	-1,5	2,25
2	-0,5	0,25
3	0,5	0,25
4	1,5	2,25

$$\sum (x-\mu)^2 = 5$$

Variance Estimation



1 2 3 4

$\mu = 2.5$

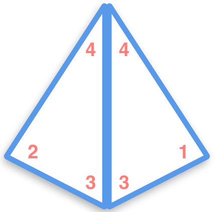
$$\sigma^2 = \frac{1}{N} \sum (x-\mu)^2$$

$x - 2.5$

x	$x-\mu$	$(x-\mu)^2$
1	-1,5	2,25
2	-0,5	0,25
3	0,5	0,25
4	1,5	2,25

$$\frac{\sum (x-\mu)^2}{N} = 5$$

Variance Estimation



1 2 3 4

$\mu = 2.5$

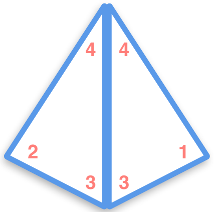
$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$x - 2.5$

x	$x - \mu$	$(x - \mu)^2$
1	-1,5	2,25
2	-0,5	0,25
3	0,5	0,25
4	1,5	2,25

$$\frac{\sum (x - \mu)^2}{N} = \frac{5}{4}$$

Variance Estimation



1 2 3 4

$\mu = 2.5$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

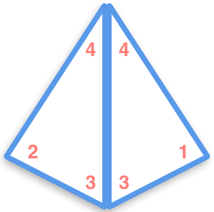
$x - 2.5$

x	$x-\mu$	$(x-\mu)^2$
1	-1,5	2,25
2	-0,5	0,25
3	0,5	0,25
4	1,5	2,25

$$\frac{\sum (x-\mu)^2}{N} = \frac{5}{4}$$

Population variance
σ^2

Variance Estimation



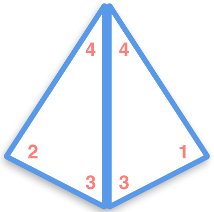
1	2	3	4

$$\mu = 2.5$$

$$\sigma^2 = \frac{1}{N}\sum (x - \mu)^2$$

$$\sigma^2 = \frac{5}{4}$$

Variance Estimation



$\boxed{1}$ $\boxed{2}$ $\boxed{3}$ $\boxed{4}$

$\mu = 2.5$

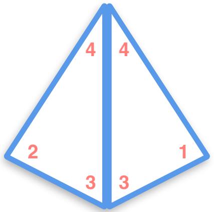
$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$$\sigma^2 = \frac{5}{4}$$

$n = 2$

Samples

Variance Estimation



1	2	3	4

$\mu = 2.5$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

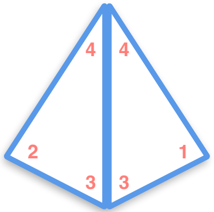
$$\sigma^2 = \frac{5}{4}$$

$n = 2$

Samples

x

Variance Estimation



1	2	3	4

$\mu = 2.5$

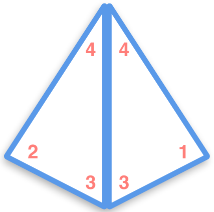
$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$$\sigma^2 = \frac{5}{4}$$

$n = 2$
Samples

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4

Variance Estimation



1	2	3	4

$\mu = 2.5$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

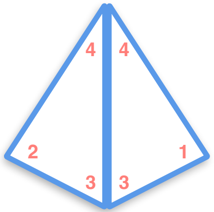
$$\sigma^2 = \frac{5}{4}$$

$n = 2$

Samples

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$																

Variance Estimation



1	2	3	4

$\mu = 2.5$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

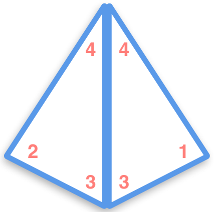
$$\sigma^2 = \frac{5}{4}$$

$n = 2$

Samples

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4

Variance Estimation



1	2	3	4

$\mu = 2.5$

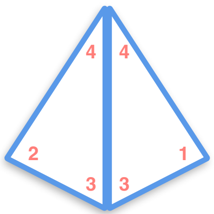
$$\boxed{\sigma^2 = \frac{1}{N} \sum (x - \mu)^2}$$

$$\sigma^2 = \frac{5}{4}$$

$n = 2$
Samples

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4

Variance Estimation



1 2 3 4

$\mu = 2.5$

$$\boxed{\sigma^2 = \frac{1}{N} \sum (x - \mu)^2}$$

$$Var(x) = \frac{\sum (x - \bar{x})^2}{n}$$

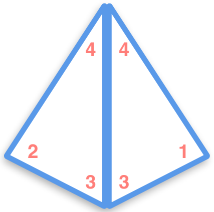
$$\sigma^2 = \frac{5}{4}$$

$n = 2$

Samples

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4

Variance Estimation



1	2	3	4

$\mu = 2.5$

$$\boxed{\sigma^2 = \frac{1}{N}\sum (x-\mu)^2}$$

$$Var(x) = \frac{\sum (x-\bar{x})^2}{n}$$

$$\sigma^2 = \frac{5}{4}$$

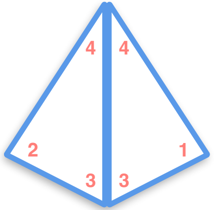
$n = 2$

Samples

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4

$var(x)$

Variance Estimation



1 2 3 4

$\mu = 2.5$

$$\boxed{\sigma^2 = \frac{1}{N}\sum (x-\mu)^2}$$

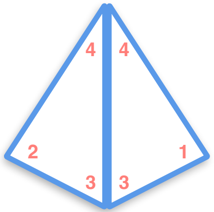
$$Var(x) = \frac{\sum (x-\bar{x})^2}{n}$$

$$\sigma^2 = \frac{5}{4}$$

$n = 2$
Samples

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4
$var(x)$	0	0,25	1	2,25	0,25	0	0,25	1	1	0,25	0	0,25	2,25	1	0,25	0

Variance Estimation



1 2 3 4

$\mu = 2.5$

$n = 2$

Samples

$$\boxed{\sigma^2 = \frac{1}{N}\sum (x-\mu)^2}$$

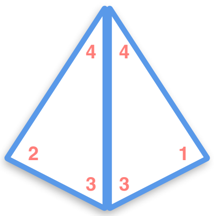
$$Var(x) = \frac{\sum (x-\bar{x})^2}{n}$$

$$\sigma^2 = \frac{5}{4}$$

$$var(x) = \frac{5}{8}$$

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4
$var(x)$	0	0,25	1	2,25	0,25	0	0,25	1	1	0,25	0	0,25	2,25	1	0,25	0

Variance Estimation



1	2	3	4

$\mu = 2.5$

$n = 2$

Samples

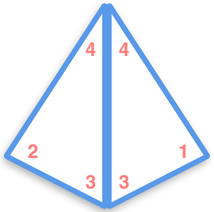
$$\boxed{\sigma^2 = \frac{1}{N}\sum (x-\mu)^2}$$

$$Var(x) = \frac{\sum (x-\bar{x})^2}{n}$$

$$\boxed{\sigma^2 = \frac{5}{4}} \qquad \boxed{var(x) = \frac{5}{8}}$$

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4
$var(x)$	0	0,25	1	2,25	0,25	0	0,25	1	1	0,25	0	0,25	2,25	1	0,25	0

Variance Estimation



1 2 3 4

$\mu = 2.5$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2 \qquad Var(x) = \frac{\sum (x - \bar{x})^2}{n}$$

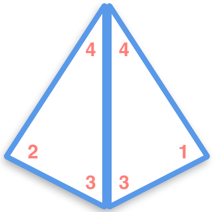
$$\sigma^2 = \frac{5}{4}$$

$n = 2$

Samples

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4

Variance Estimation



1	2	3	4

$\mu = 2.5$

$n = 2$

Samples

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

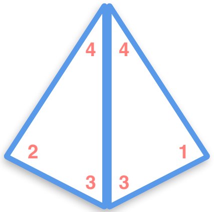
$$Var(x) = \frac{\sum (x - \bar{x})^2}{n}$$

$$\sigma^2 = \frac{5}{4}$$

$$Var(x) = \frac{\sum (x - \bar{x})^2}{n - 1}$$

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4

Variance Estimation



1	2	3	4

$\mu = 2.5$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2 \qquad Var(x) = \frac{\sum (x-\bar{x})^2}{n-1}$$

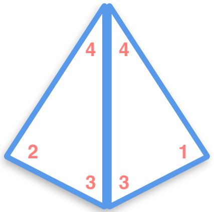
$$\sigma^2 = \frac{5}{4}$$

$n = 2$

Samples

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4

Variance Estimation



1 2 3 4

$\mu = 2.5$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2 \qquad Var(x) = \frac{\sum (x - \bar{x})^2}{n - 1}$$

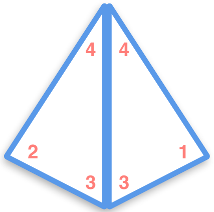
$$\sigma^2 = \frac{5}{4}$$

$n = 2$
Samples

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4

$var(x)$

Variance Estimation



1	2	3	4

$\mu = 2.5$

$$\sigma^2 = \frac{1}{N}\sum(x-\mu)^2 \qquad Var(x) = \frac{\sum(x-\bar{x})^2}{n-1}$$

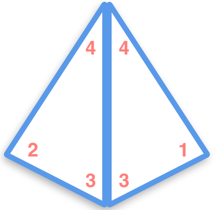
$$\sigma^2 = \frac{5}{4}$$

$n = 2$

Samples

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4
$var(x)$	0	0,5	2	4,5	0,5	0	0,5	2	2	0,5	0	0,5	4,5	2	0,5	0

Variance Estimation



1	2	3	4

$\mu = 2.5$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

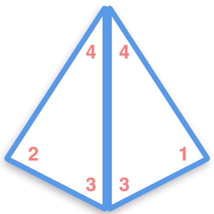
$$Var(x) = \frac{\sum (x-\bar{x})^2}{n-1}$$

$$\sigma^2 = \frac{5}{4}$$

$n = 2$
Samples

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4
$var(x)$	0	0,5	2	4,5	0,5	0	0,5	2	2	0,5	0	0,5	4,5	2	0,5	0

Variance Estimation



1	2	3	4

$\mu = 2.5$

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

$$Var(x) = \frac{\sum (x - \bar{x})^2}{n - 1}$$

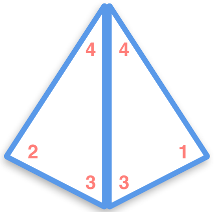
$$\sigma^2 = \frac{5}{4}$$

$$Var(x) = \frac{5}{4}$$

$n = 2$
Samples

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4
$var(x)$	0	0,5	2	4,5	0,5	0	0,5	2	2	0,5	0	0,5	4,5	2	0,5	0

Variance Estimation



1	2	3	4

$\mu = 2.5$

$n = 2$

Samples

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$Var(x) = \frac{\sum (x-\bar{x})^2}{n-1}$$

$$\boxed{\sigma^2 = \frac{5}{4}} \qquad \boxed{Var(x) = \frac{5}{4}}$$

x	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
$\bar{x}$	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4
$var(x)$	0	0,5	2	4,5	0,5	0	0,5	2	2	0,5	0	0,5	4,5	2	0,5	0

Variance Estimation

Population Variance Formula

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

Sample Variance Formula

$$Var(x) = \frac{\sum (x - \bar{x})^2}{n - 1}$$

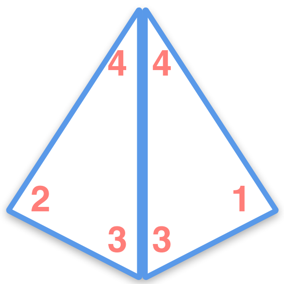
$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

Sample and Population

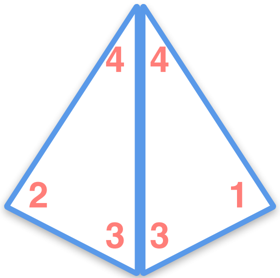
Law of Large Numbers

Law of Large Numbers

Law of Large Numbers

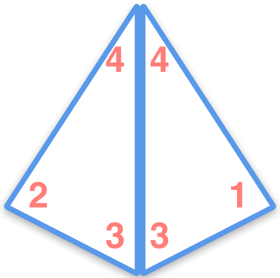


Law of Large Numbers



1 2 3 4

Law of Large Numbers



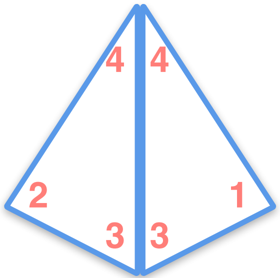
1 2 3 4

$$\mu = 2.5$$

Experiment:

Toss the 4-sided dice twice and record the average of your outcomes

Law of Large Numbers



1 2 3 4

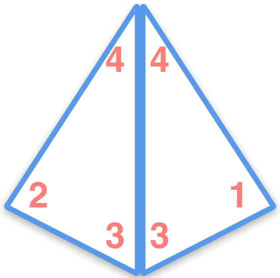
$$\mu = 2.5$$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

Experiment:

Toss the 4-sided dice twice and record the average of your outcomes

Law of Large Numbers



1 2 3 4

$$\mu = 2.5$$

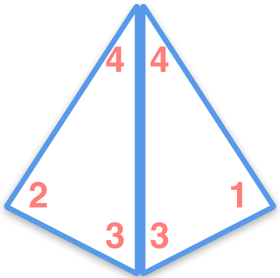
Experiment:

Toss the 4-sided dice twice and record the average of your outcomes

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

	1	2	3	4
1	1	1.5	2	2.5
2	1.5	2	2.5	3
3	2	2.5	3	3.5
4	2.5	3	3.5	4

Law of Large Numbers



1 2 3 4

$$\mu = 2.5$$

Experiment:

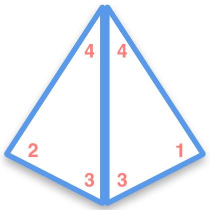
Toss the 4-sided dice twice and record the average of your outcomes

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

	1	2	3	4
1	1	1.5	2	2.5
2	1.5	2	2.5	3
3	2	2.5	3	3.5
4	2.5	3	3.5	4

	1	2	3	4
1	μ 2.5			
2				
3				
4				

Law of Large Numbers

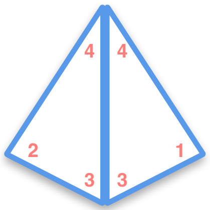


1 2 3 4

$$\mu = 2.5$$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

Law of Large Numbers



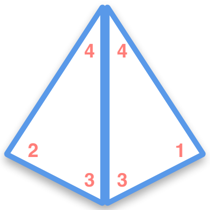
1 2 3 4

$\mu = 2.5$

1 trial

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

Law of Large Numbers



1 2 3 4

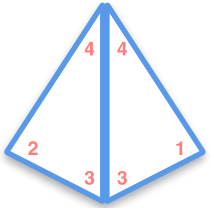
$$\mu = 2.5$$

1 trial

4,3

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

Law of Large Numbers



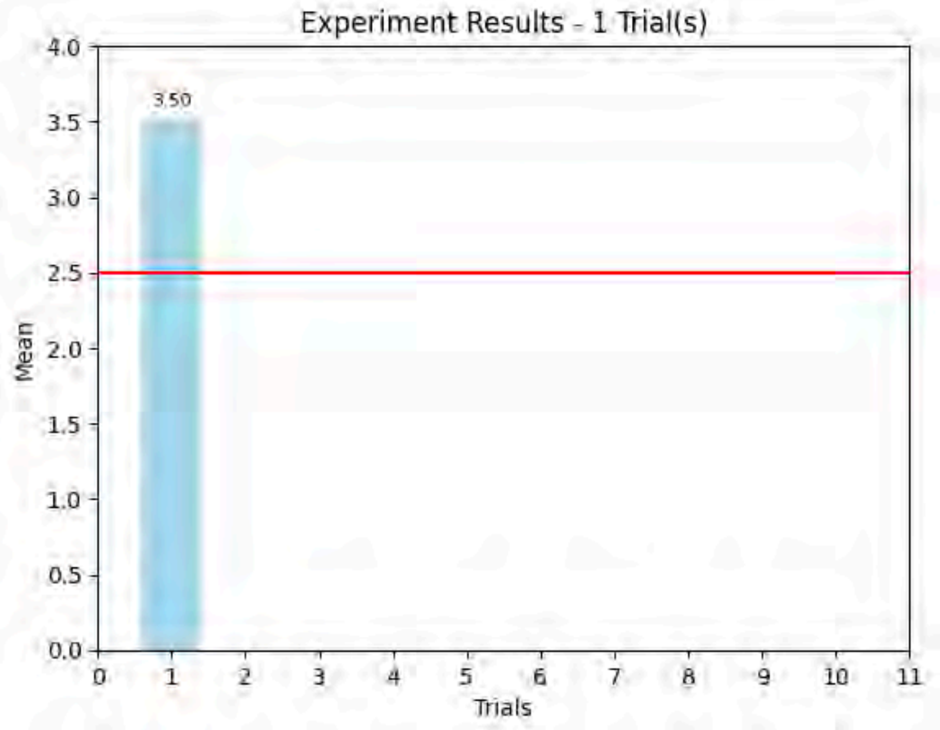
1 2 3 4

$\mu = 2.5$

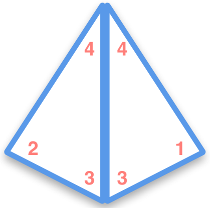
1 trial

4,3

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



Law of Large Numbers



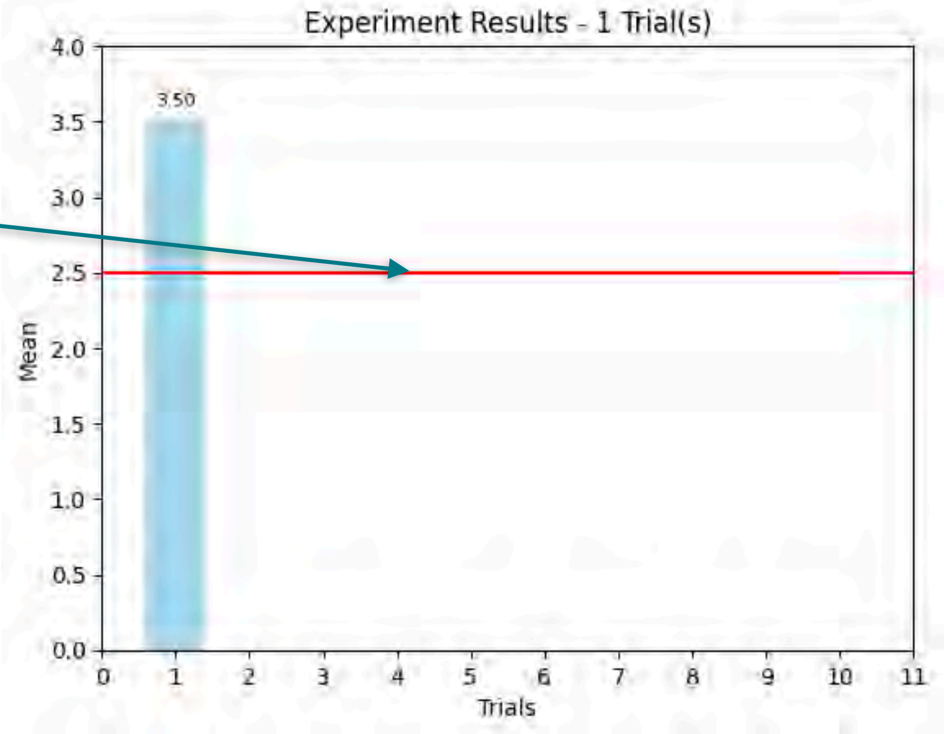
1 2 3 4

$\mu = 2.5$

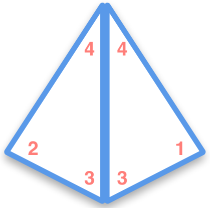
1 trial

4,3

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



Law of Large Numbers



1 2 3 4

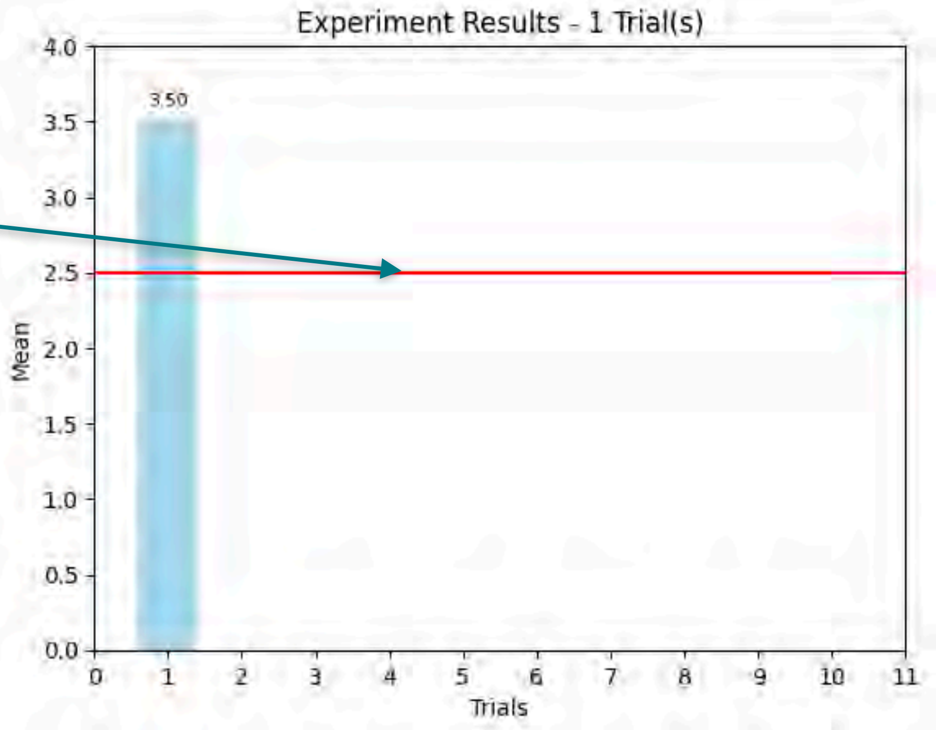
$\mu = 2.5$

1 trial

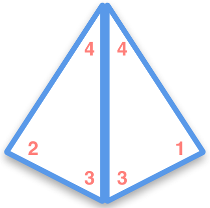
4,3

$\bar{x}_1$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



Law of Large Numbers



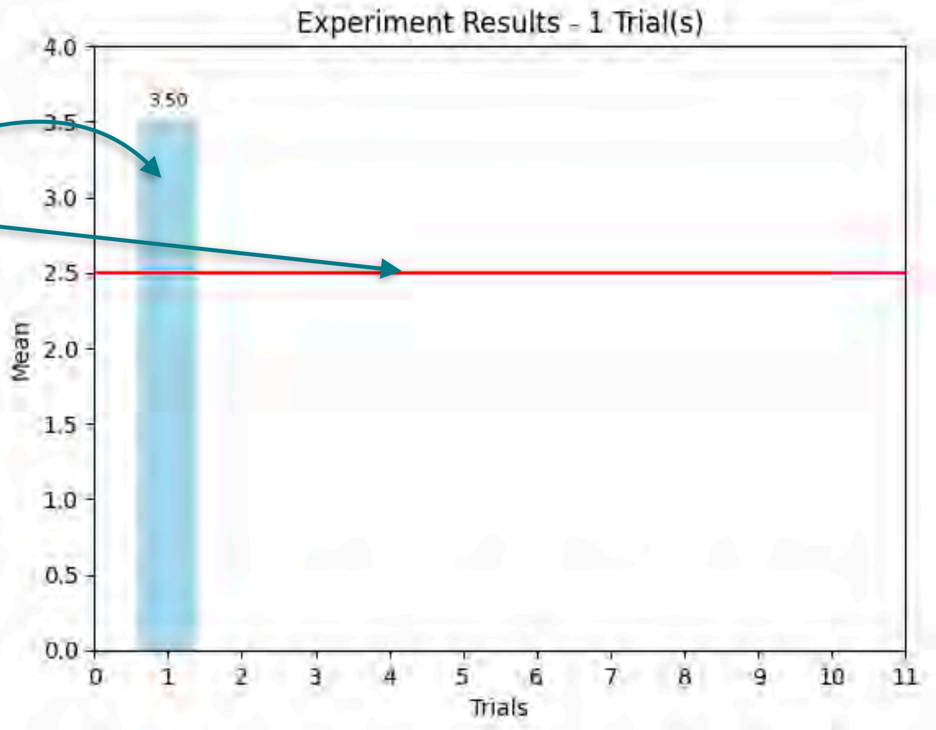
1 2 3 4

$\mu = 2.5$

1 trial

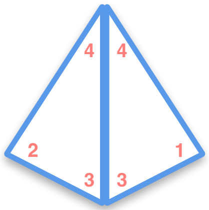
4,3

$\bar{x}_1$



	1	**2**	**3**	**4**
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

Law of Large Numbers



1 2 3 4

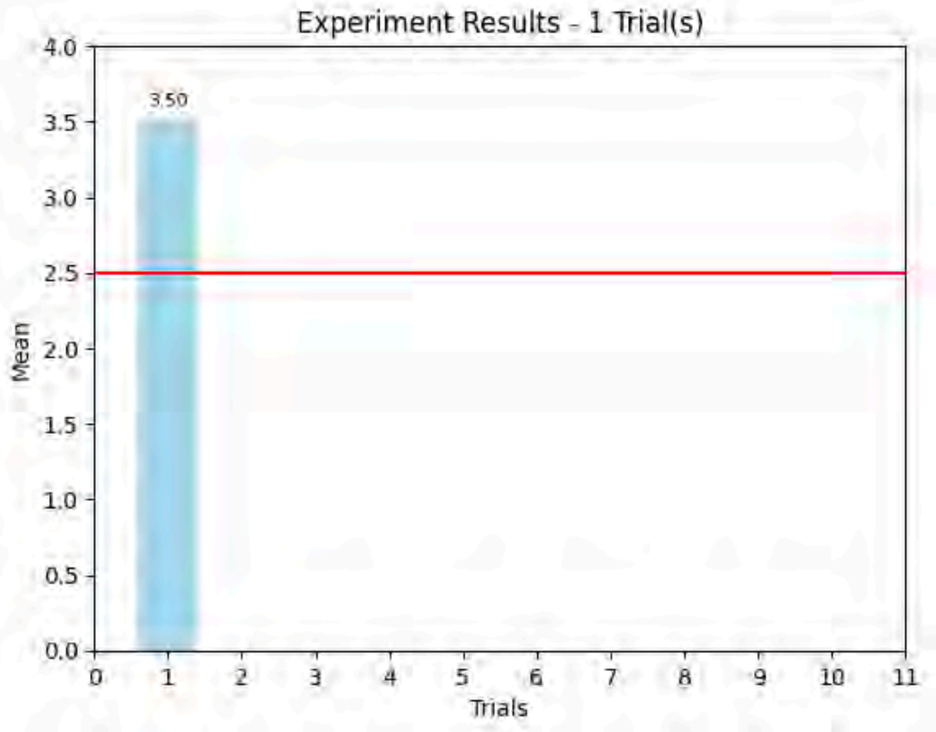
$\mu = 2.5$

1 trial

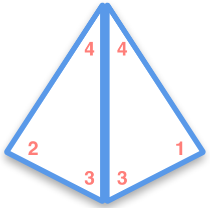
4,3

2 trials

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



Law of Large Numbers



1 2 3 4

$\mu = 2.5$

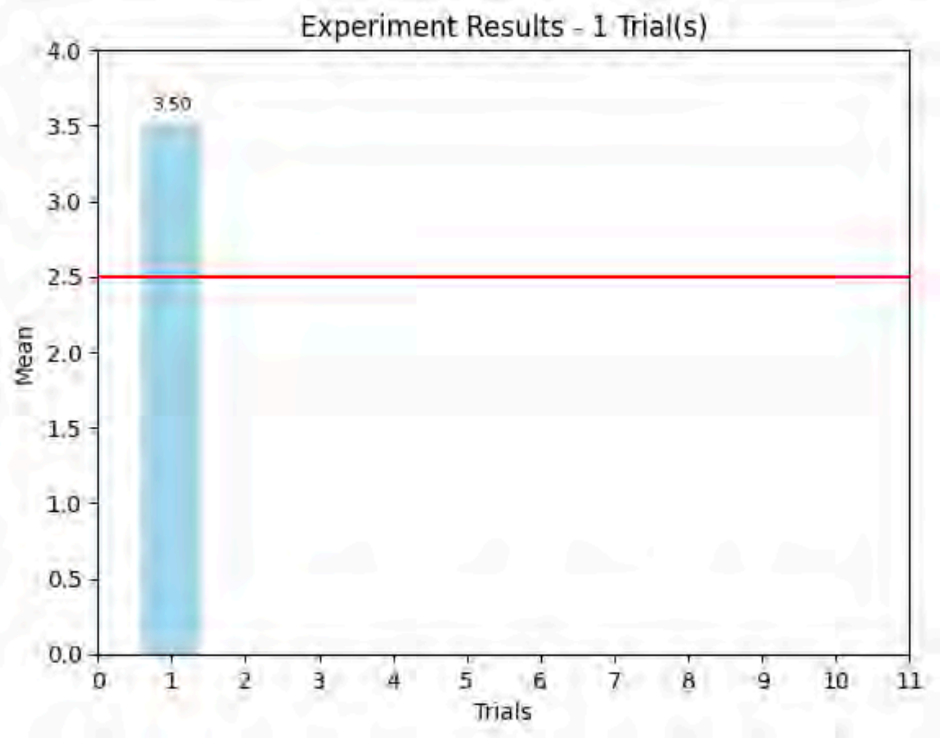
	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

1 trial

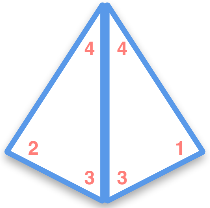
4,3

2 trials

3,4
1,3



Law of Large Numbers



1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

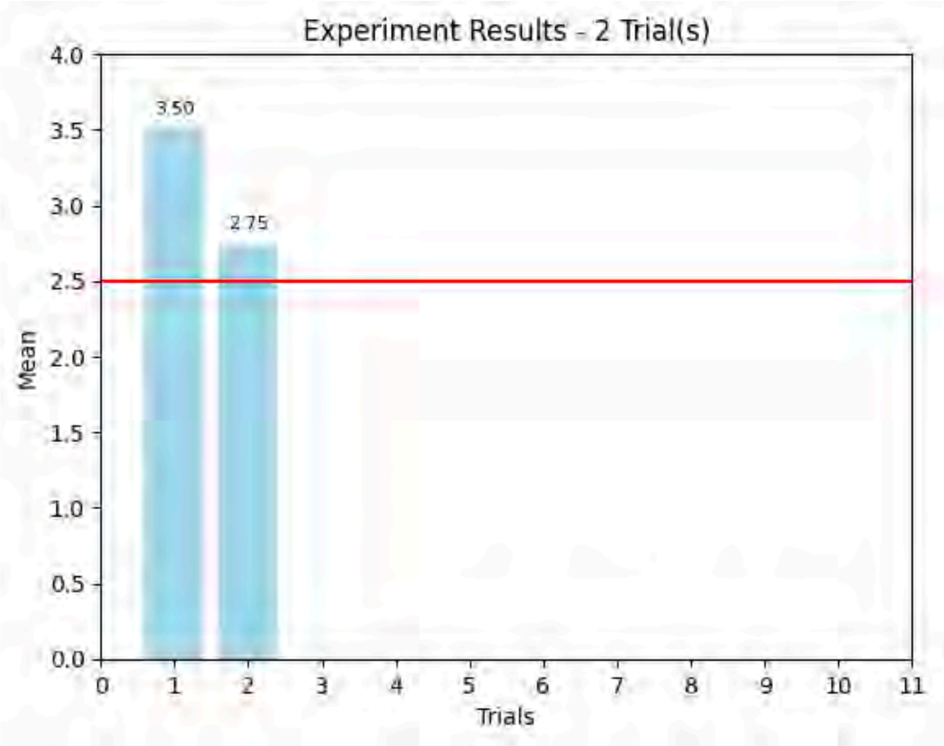
1 trial

4,3

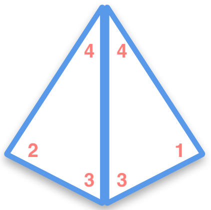
2 trials

3,4

1,3



Law of Large Numbers



1 2 3 4

$$\mu = 2.5$$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

1 trial

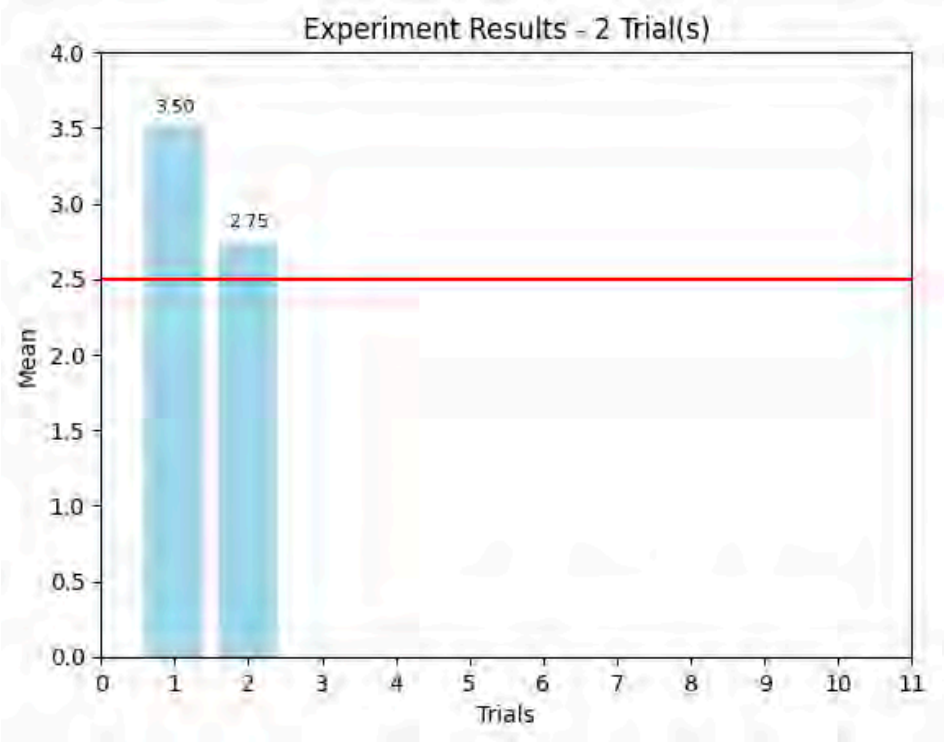
4,3

2 trials

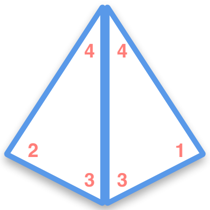
3,4

1,3

3 trials



Law of Large Numbers



1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

1 trial

4,3

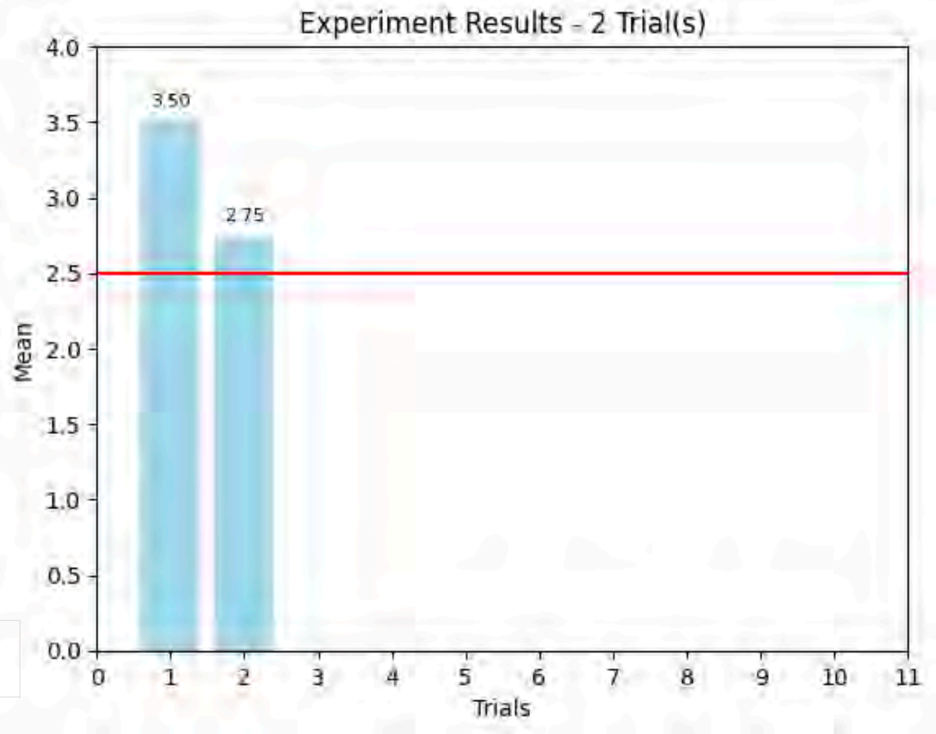
2 trials

3,4

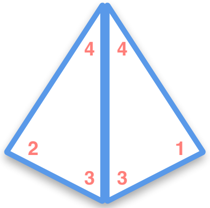
1,3

3 trials

3,1 1,4 1,1



Law of Large Numbers



1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

1 trial

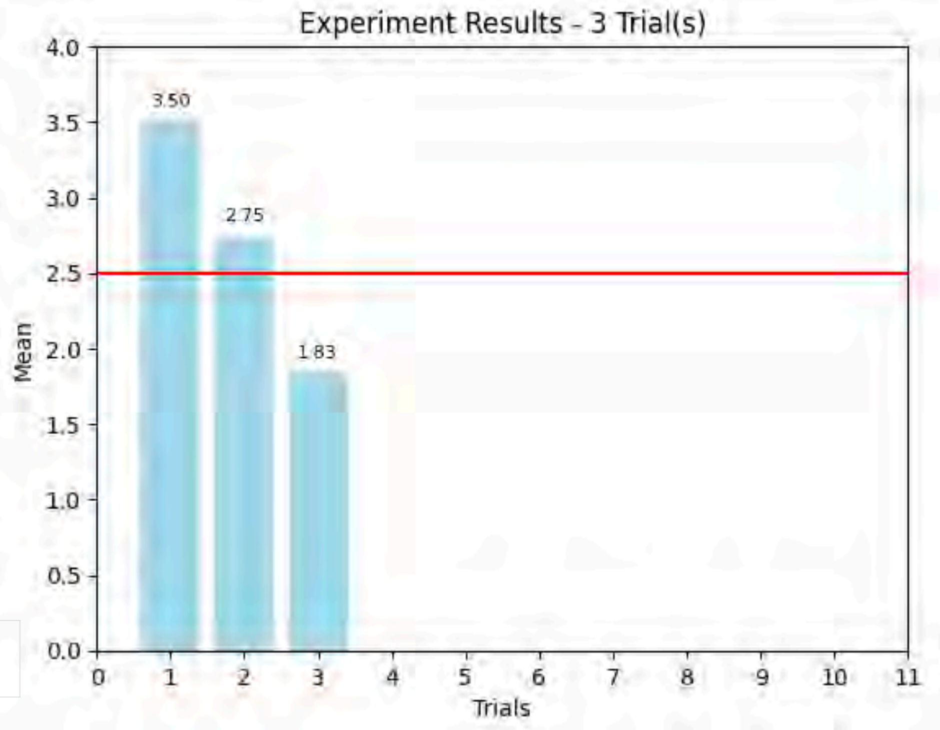
4,3

2 trials

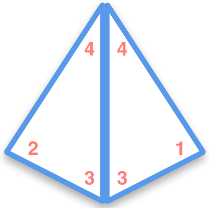
3,4
1,3

3 trials

3,1	1,4	1,1



Law of Large Numbers



1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

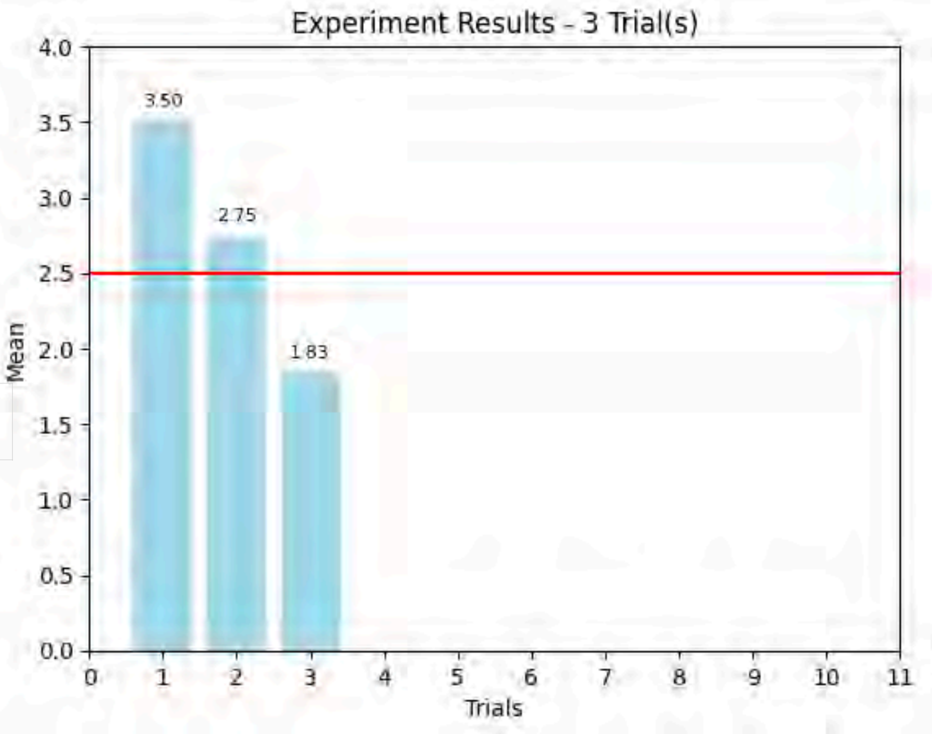
2 trials

3,4
1,3

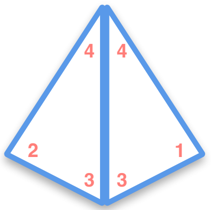
3 trials

3,1	1,4	1,1

4 trials



Law of Large Numbers



1 2 3 4

$\mu = 2.5$

1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

2 trials

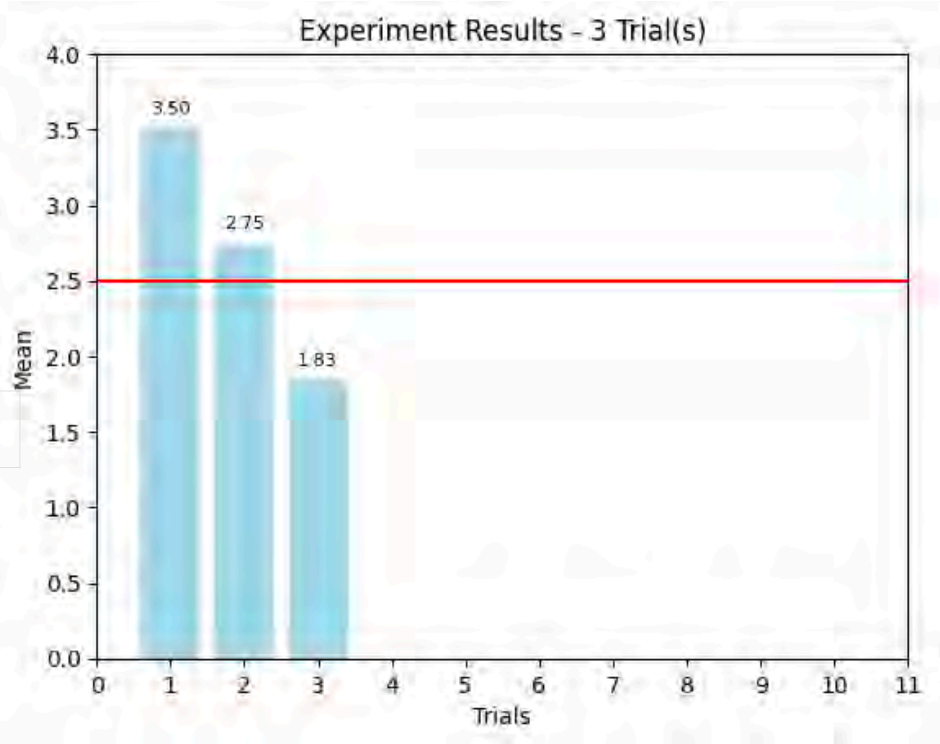
3,4
1,3

3 trials

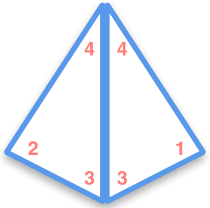
3,1	1,4	1,1

4 trials

3,1	3,1
1,2	3,2



Law of Large Numbers



1 2 3 4

$$\mu = 2.5$$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

2 trials

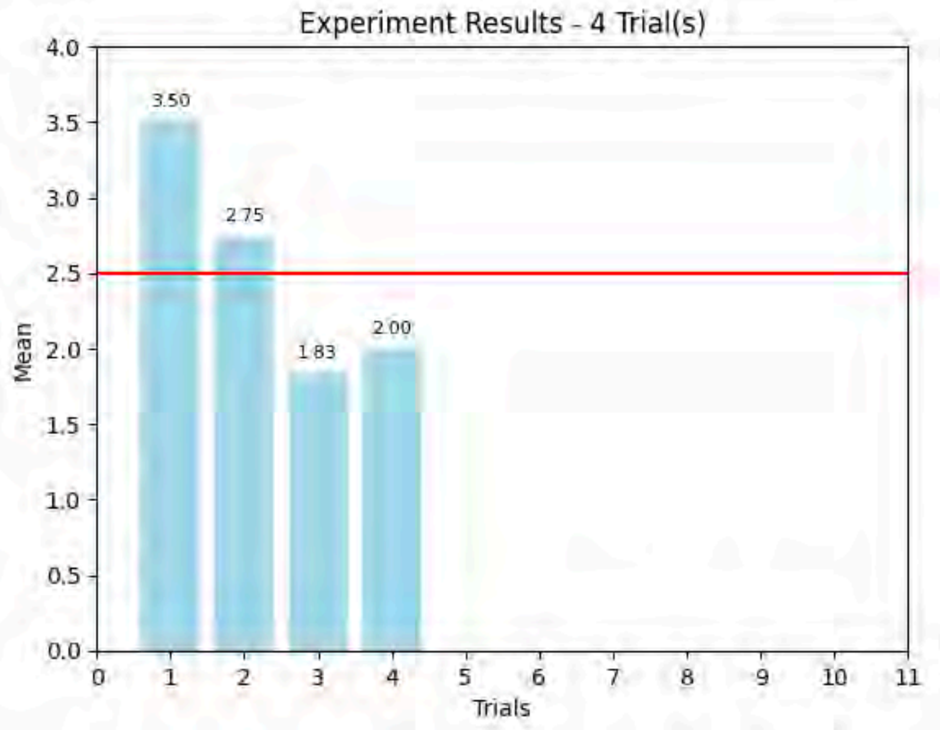
3,4
1,3

3 trials

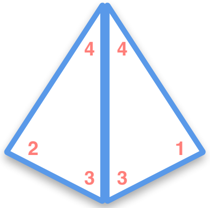
3,1	1,4	1,1

4 trials

3,1	3,1
1,2	3,2



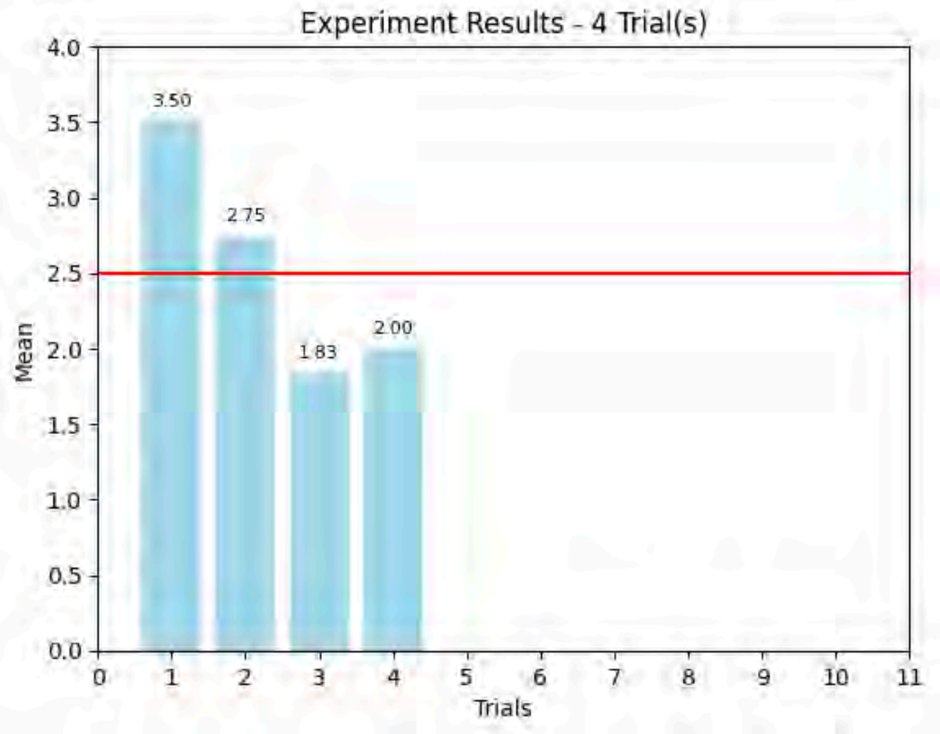
Law of Large Numbers



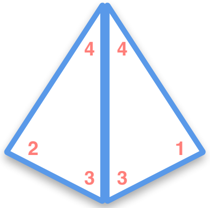
1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



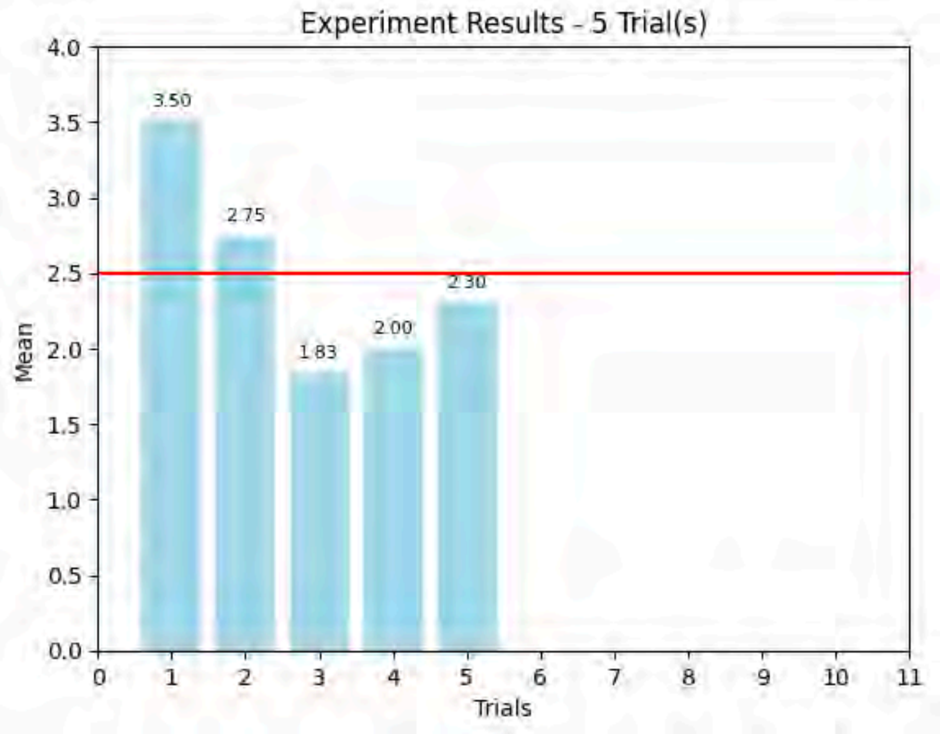
Law of Large Numbers



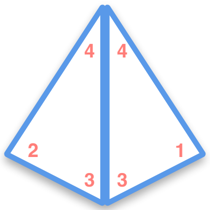
1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



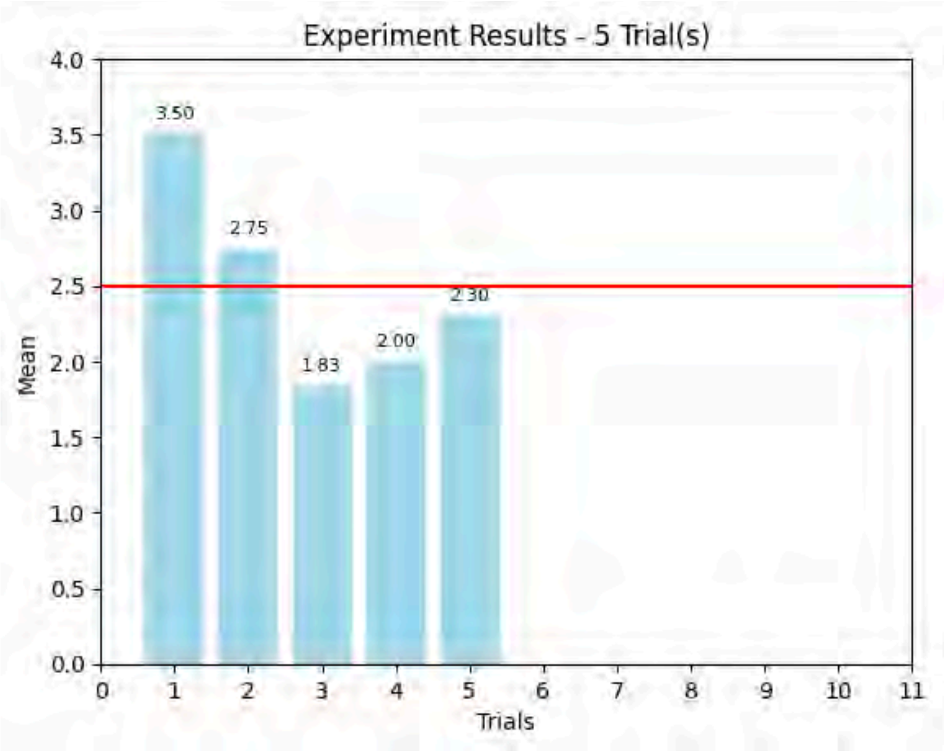
Law of Large Numbers



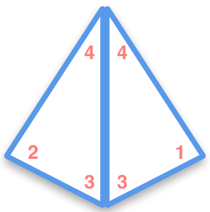
1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



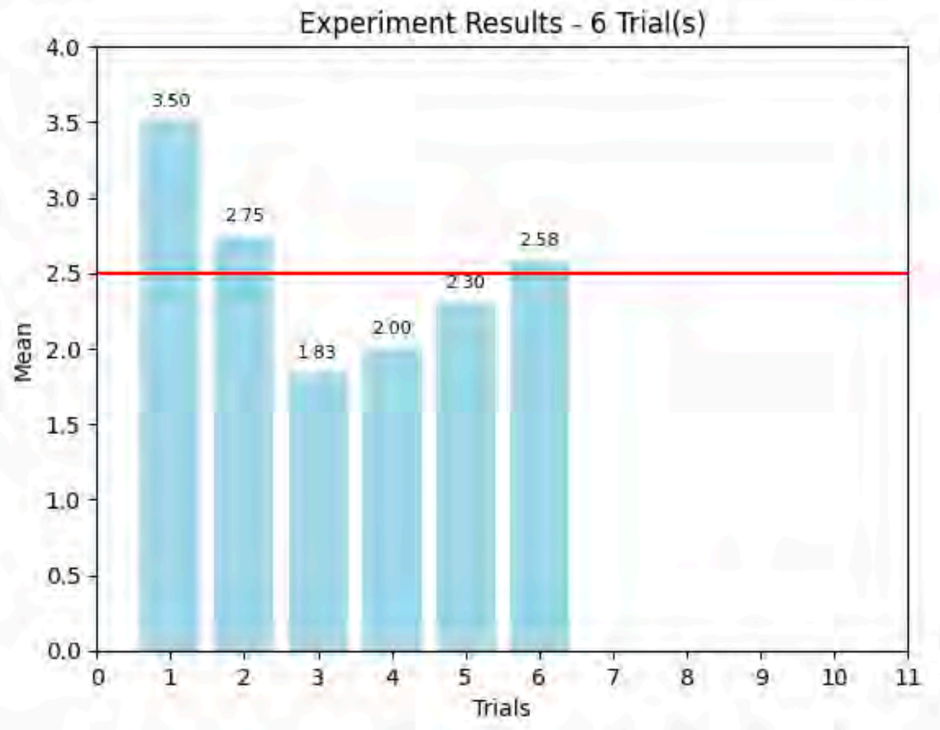
Law of Large Numbers



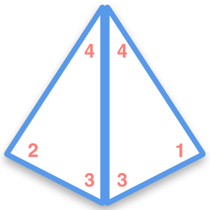
1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



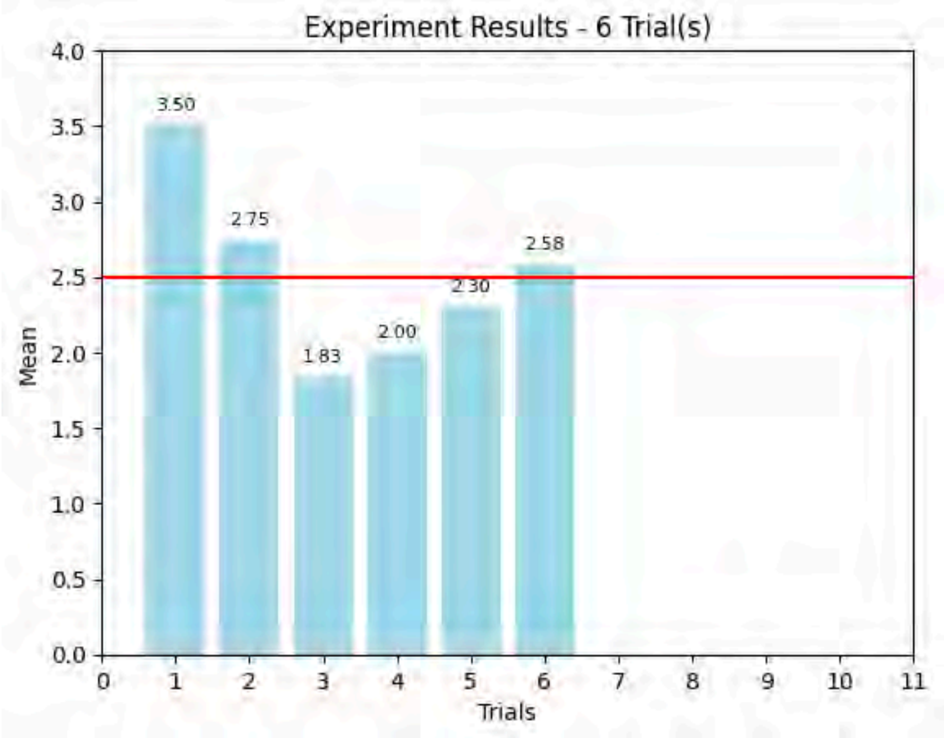
Law of Large Numbers



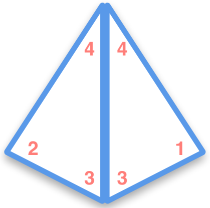
1 2 3 4

$$\mu = 2.5$$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



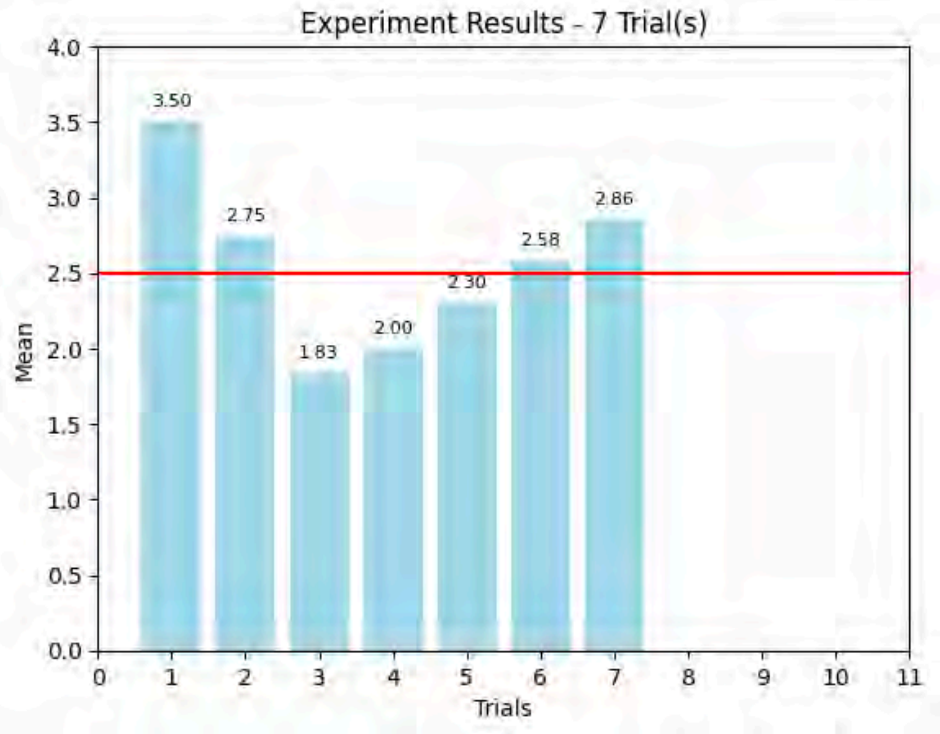
Law of Large Numbers



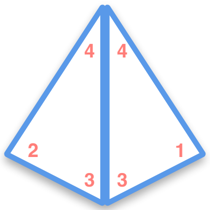
1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



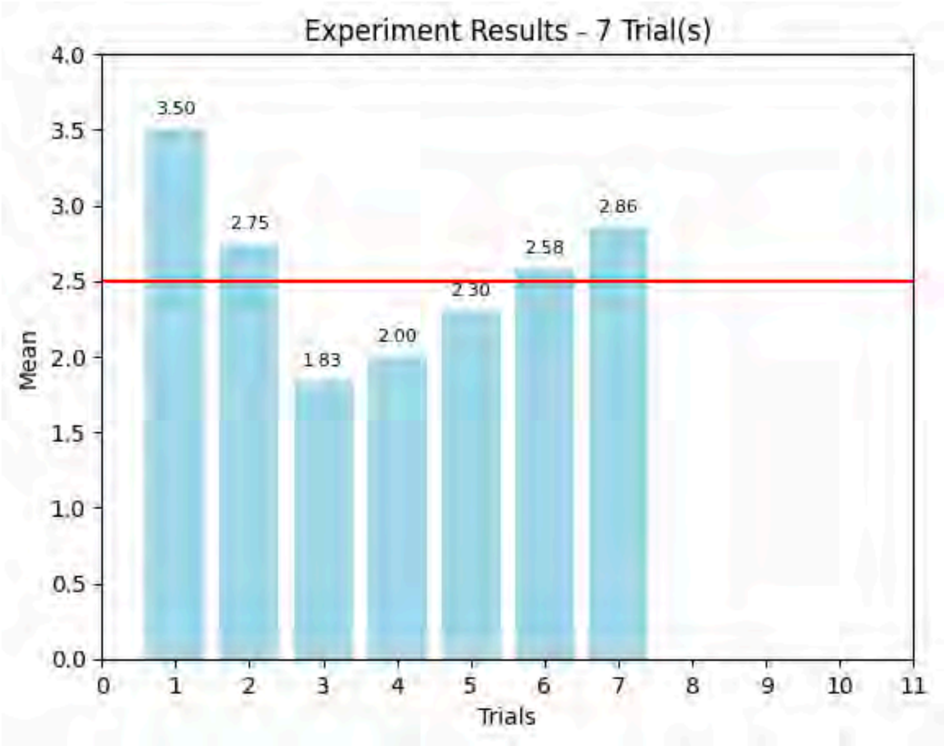
Law of Large Numbers



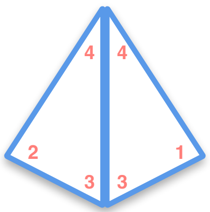
1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



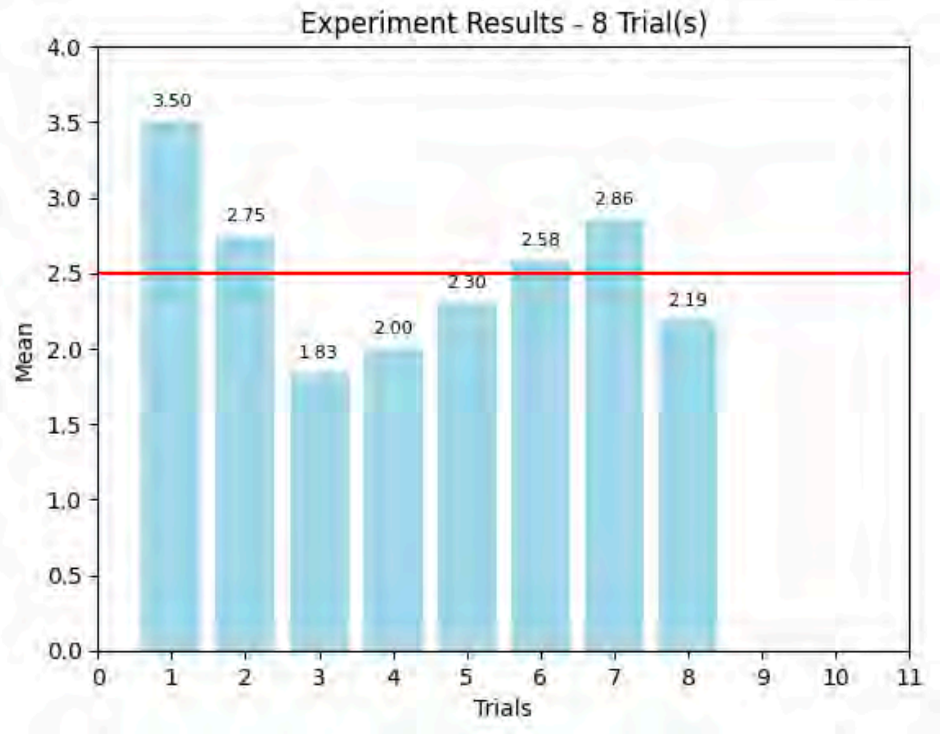
Law of Large Numbers



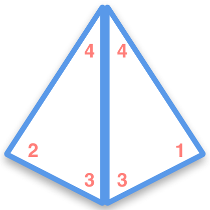
1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



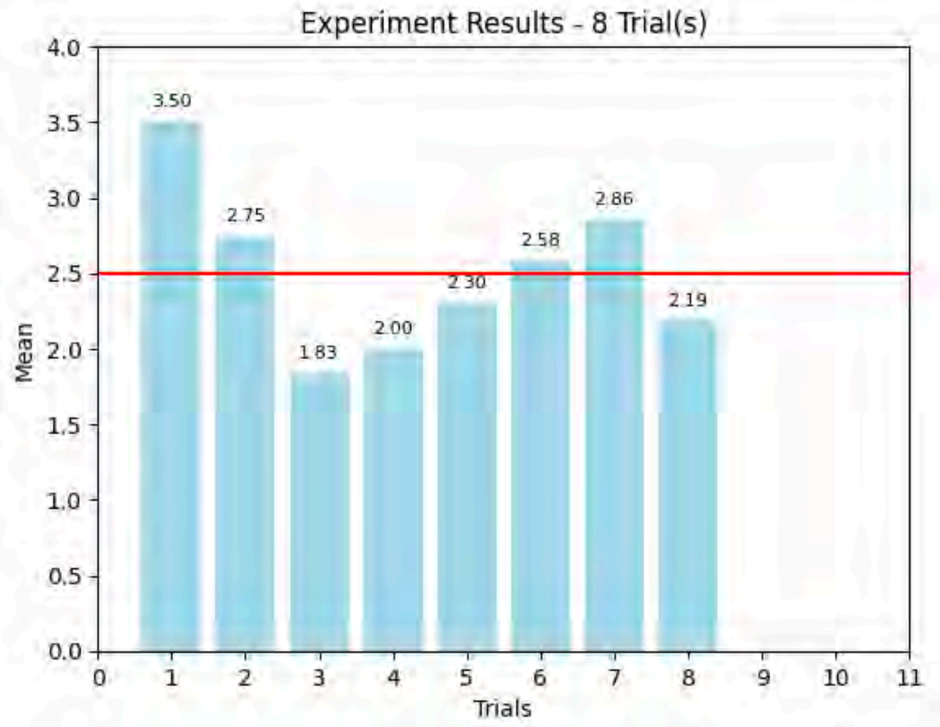
Law of Large Numbers



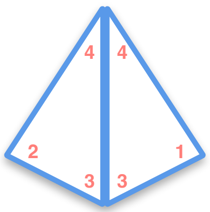
1 2 3 4

$$\mu = 2.5$$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



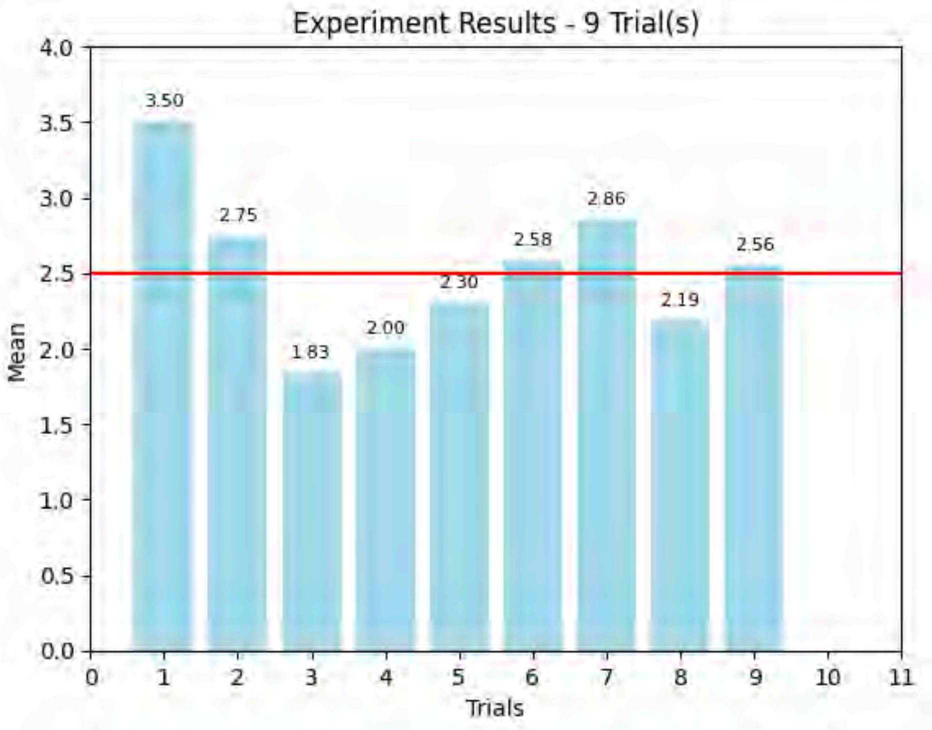
Law of Large Numbers



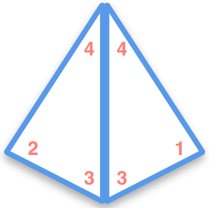
1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



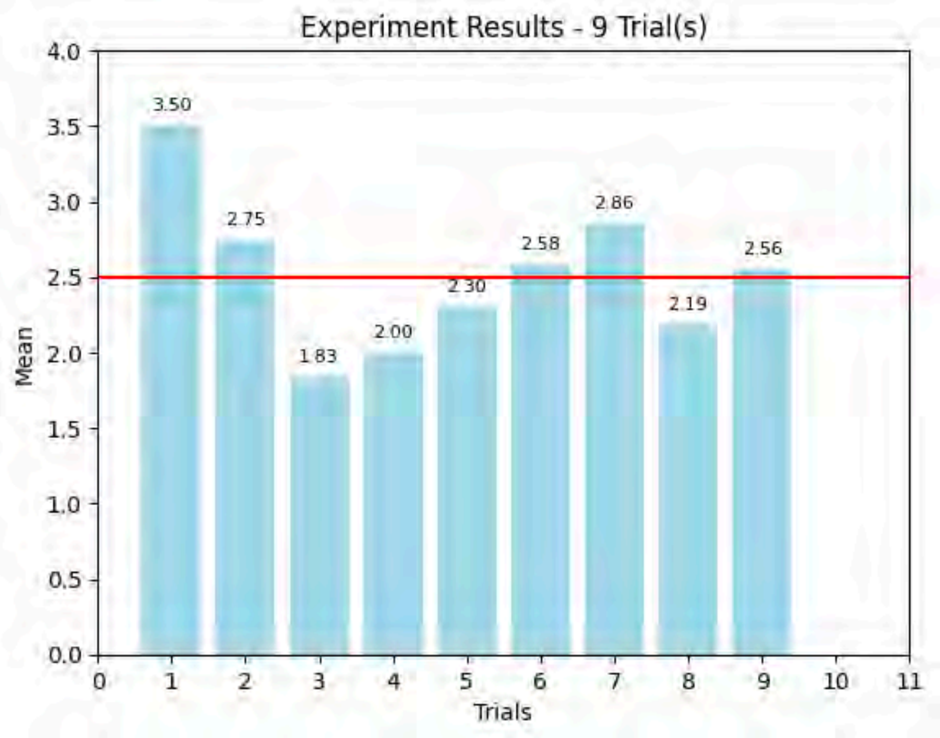
Law of Large Numbers



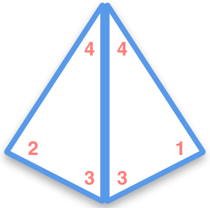
1 2 3 4

$$\mu = 2.5$$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



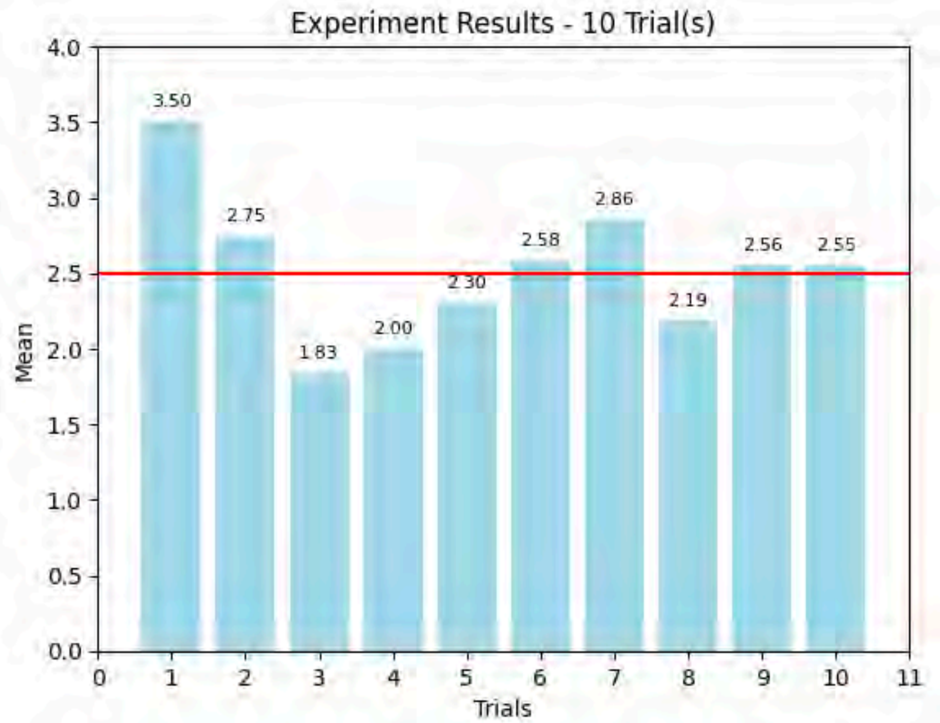
Law of Large Numbers



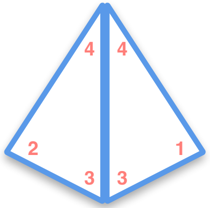
1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



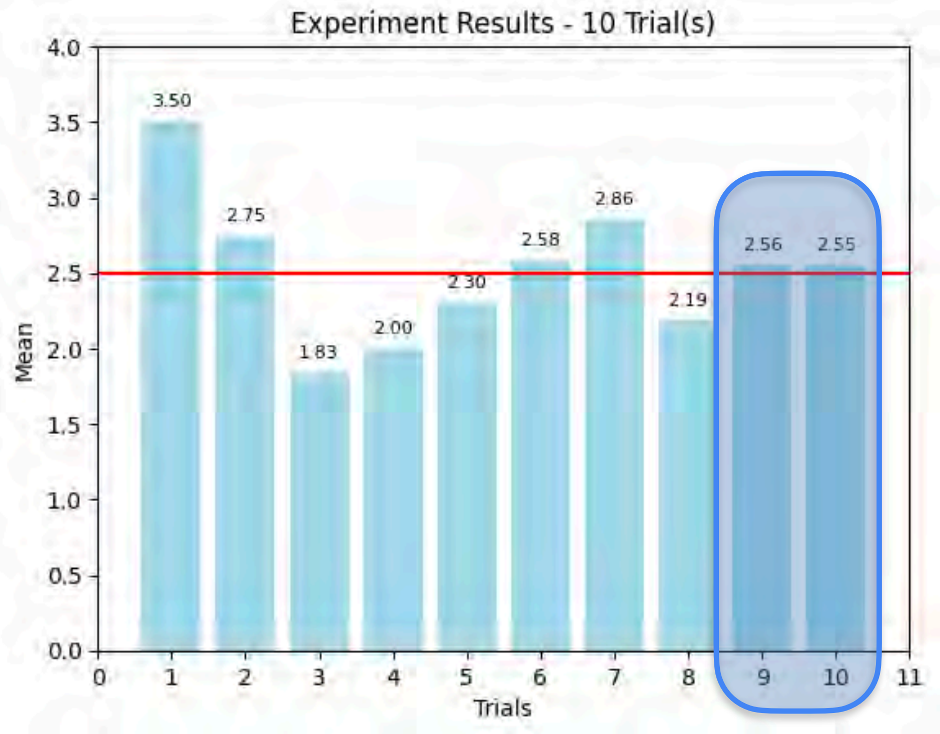
Law of Large Numbers



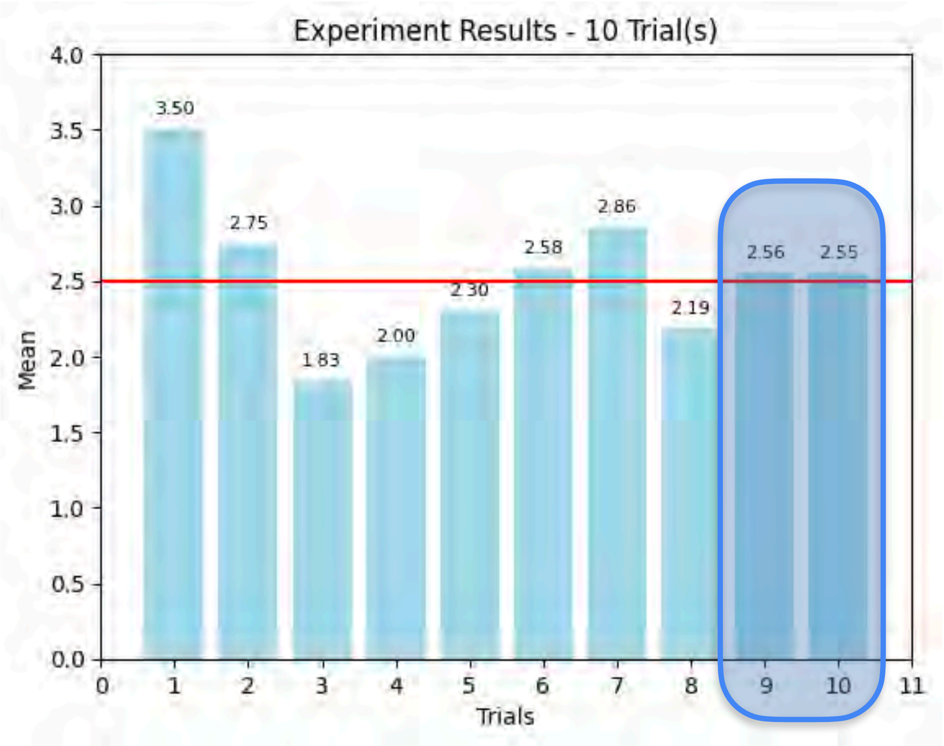
1 2 3 4

$$\mu = 2.5$$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



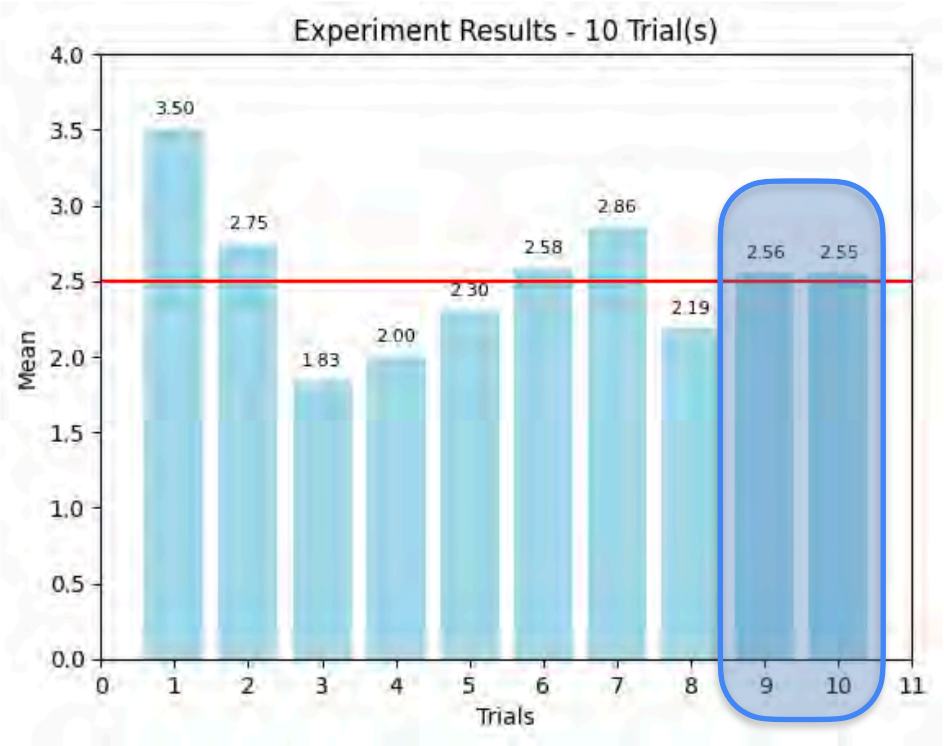
Law of Large Numbers



Law of Large Numbers

As the sample size increases, the average of the sample will tend to get closer to the average of the entire population.

Law of Large Numbers



Law of Large Numbers

Law of Large Numbers

Law of Large Numbers

Law of Large Numbers

n : number of samples

Law of Large Numbers

Law of Large Numbers

n : number of samples

X_i : some estimate X for a sample size i

Law of Large Numbers

Law of Large Numbers

n : number of samples

X_i : some estimate X for a sample size i

as $n \rightarrow \infty$

Law of Large Numbers

Law of Large Numbers

n : number of samples

X_i : some estimate X for a sample size i

as $n \rightarrow \infty$

$$\frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow \mathbb{E}[X] = \mu_X$$

Law of Large Numbers

as $n \to \infty$

Law of Large Numbers

n : number of samples

X_i : some estimate X for a sample size i

as $n \to \infty$ $$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}[X] = \mu_X$$

Law of Large Numbers

Law of Large Numbers

n : number of samples

X_i : some estimate X for a sample size i

as $n \to \infty$ $$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}[X] = \mu_X$$

as $n \to \infty$

$$\frac{X_1 + X_2 + X_3 + \ \ldots + \ X_n}{n} \longrightarrow \mathbb{E}[X] = \mu_X$$

<u>**UNDER CERTAIN CONDITIONS**</u>

Law of Large Numbers

<u>**UNDER CERTAIN CONDITIONS**</u>

Law of Large Numbers

UNDER CERTAIN CONDITIONS

- Sample is randomly drawn.

Law of Large Numbers

UNDER CERTAIN CONDITIONS

- Sample is randomly drawn.

Law of Large Numbers

UNDER CERTAIN CONDITIONS

- Sample is randomly drawn.
- Sample size must be sufficiently large.

Law of Large Numbers

UNDER CERTAIN CONDITIONS

- Sample is randomly drawn.
- Sample size must be sufficiently large.

Law of Large Numbers

UNDER CERTAIN CONDITIONS

- Sample is randomly drawn.
- Sample size must be sufficiently large.
- Independent observations.

DeepLearning.AI

Sample and Population

Central Limit Theorem

Central Limit Theorem (CLT) - Example 1

If $n = 1$

Central Limit Theorem (CLT) - Example 1

If $n = 1$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$

$\mathbf{P}(T) = 0.5$

If $n = 1$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$

$\mathbf{P}(T) = 0.5$

Random variable → X

If $n = 1$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$

$\mathbf{P}(T) = 0.5$

Random variable → X **number of heads when a coin is flipped n times**

If $n = 1$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$

$\mathbf{P}(T) = 0.5$

Random variable → X **number of heads when a coin is flipped n times**

If $n = 1$ $X = 1$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$

$\mathbf{P}(T) = 0.5$

Random variable $\rightarrow X$ **number of heads when a coin is flipped n times**

If $n = 1$ $\quad X = 1 \quad\quad X = 0$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$

$\mathbf{P}(T) = 0.5$

Random variable $\rightarrow$ X **number of heads when a coin is flipped n times**

If $n = 1$ $\quad X = 1 \quad X = 0$

Discrete Random Variable

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$

$X = 1$

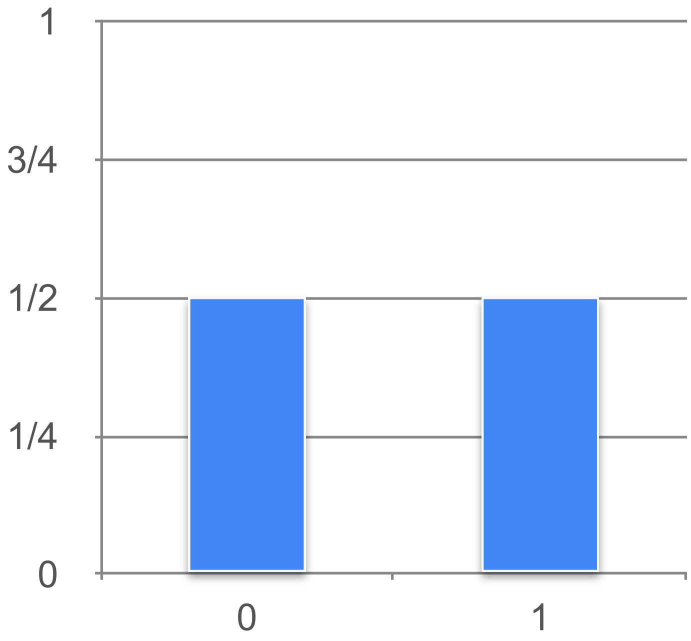
$\mathbf{P}(T) = 0.5$

$X = 0$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\mathbf{P}(T) = 0.5$

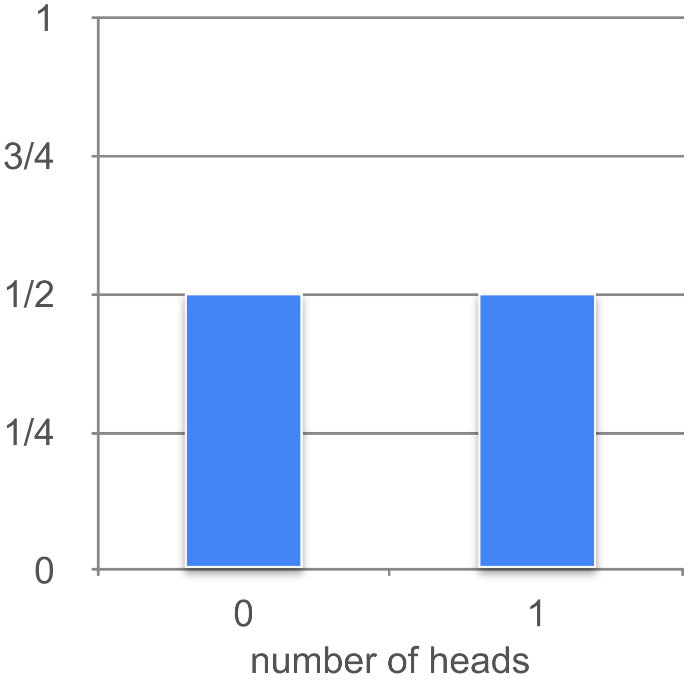
$X = 1$ $X = 0$



Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\mathbf{P}(T) = 0.5$

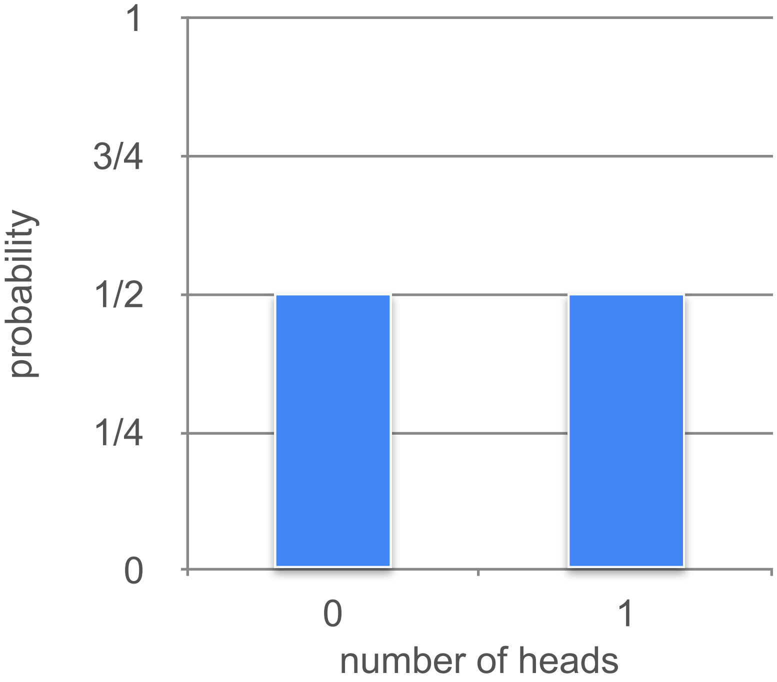
$X = 1$ $X = 0$



Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\qquad$ $\mathbf{P}(T) = 0.5$

$X = 1$ $\qquad$ $X = 0$

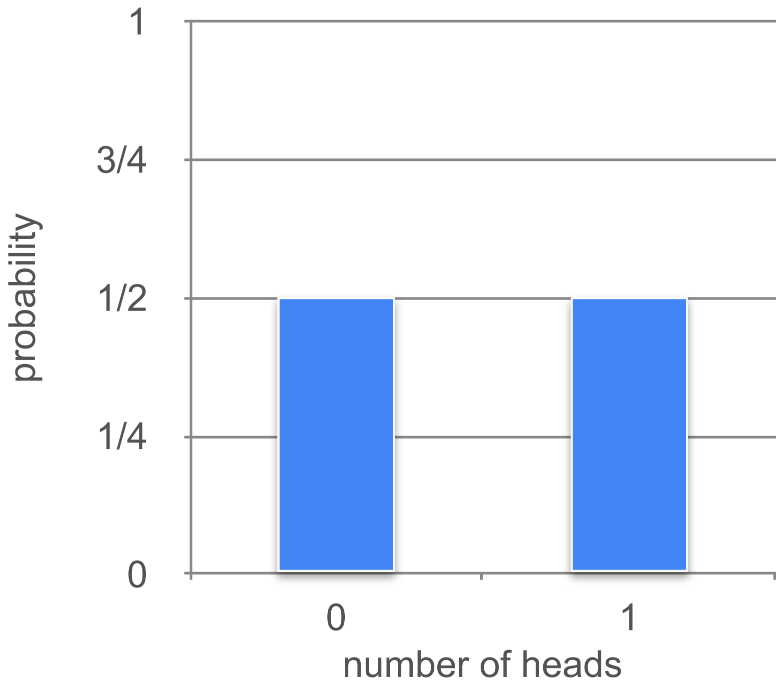


Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\quad$ $\mathbf{P}(T) = 0.5$

$X = 1$ $\quad$ $X = 0$

What can we say about the probability distribution when the number of coin flips increases?

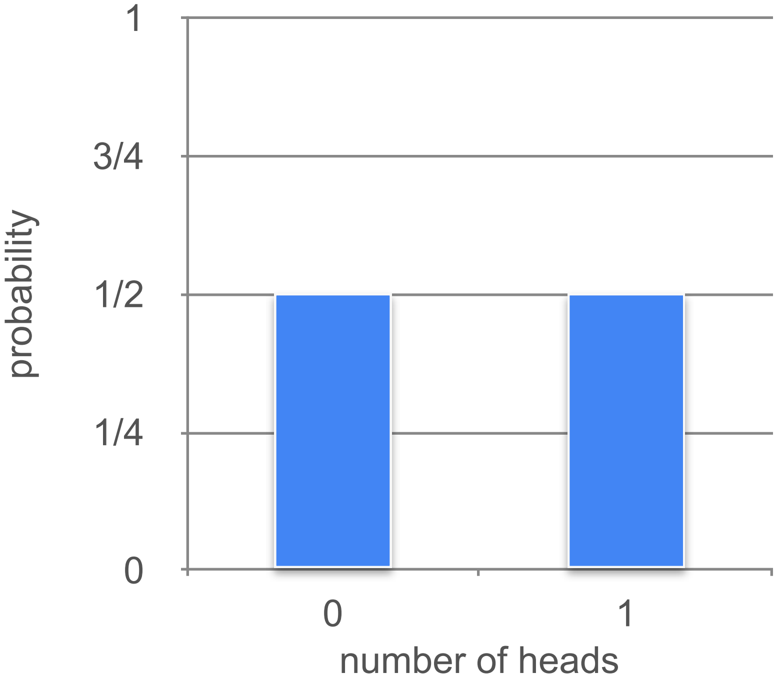


Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\quad$ $\mathbf{P}(T) = 0.5$

$X = 1$ $\quad$ $X = 0$

What can we say about the probability distribution when the number of coin flips increases?

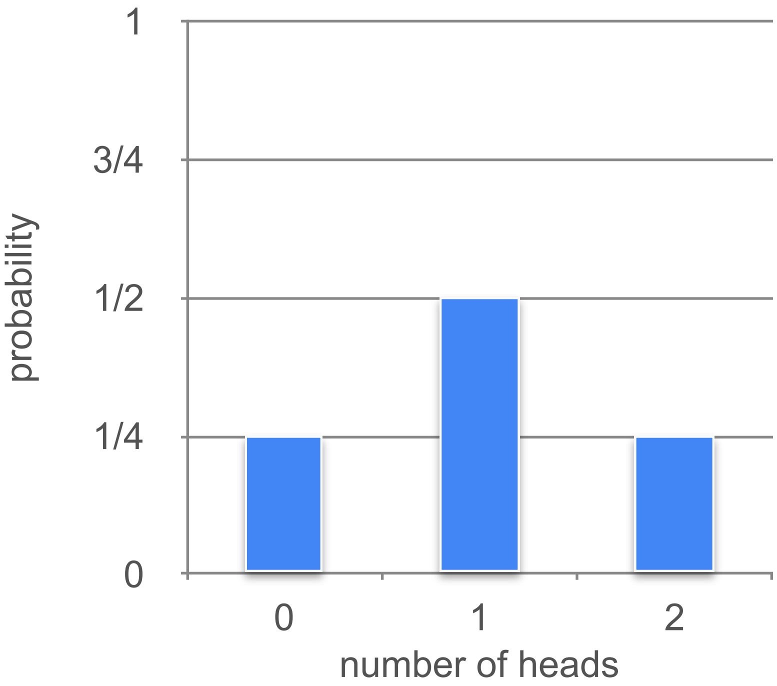


Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\mathbf{P}(T) = 0.5$

$X = 1$ $X = 0$

What can we say about the probability distribution when the number of coin flips increases?



Central Limit Theorem (CLT) - Example 1

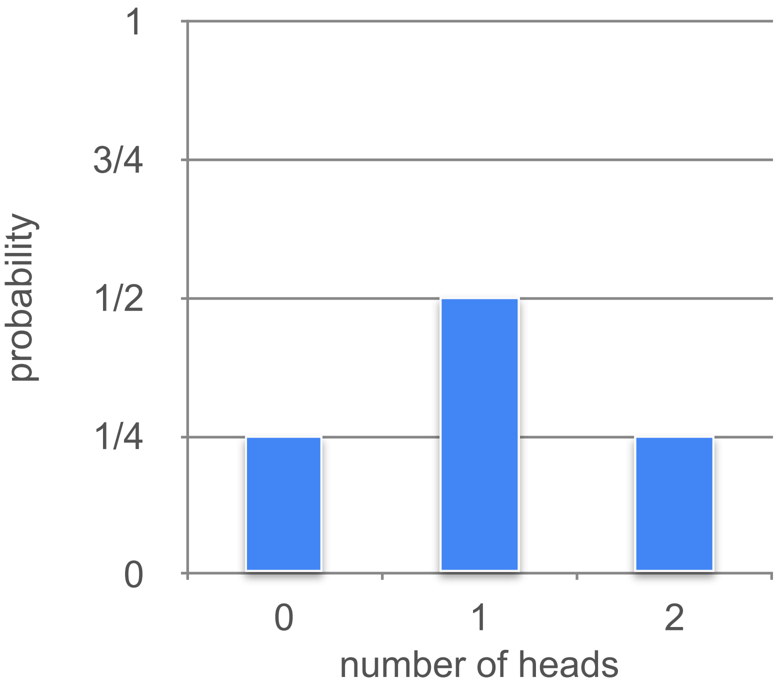
$\mathbf{P}(H) = 0.5$

$X = 1$

$\mathbf{P}(T) = 0.5$

$X = 0$

What can we say about the probability distribution when the number of coin flips increases?

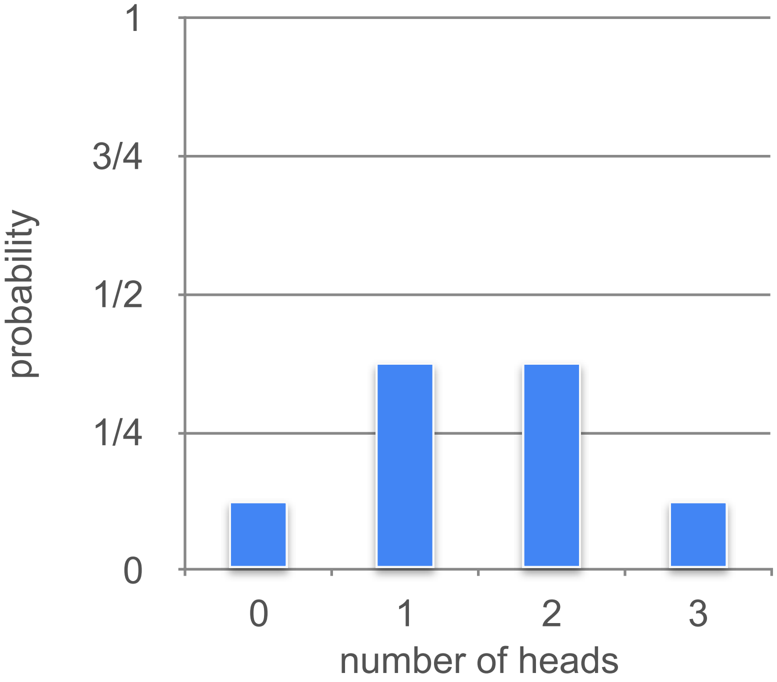


Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\quad$ $\mathbf{P}(T) = 0.5$

$X = 1$ $\quad$ $X = 0$

What can we say about the probability distribution when the number of coin flips increases?

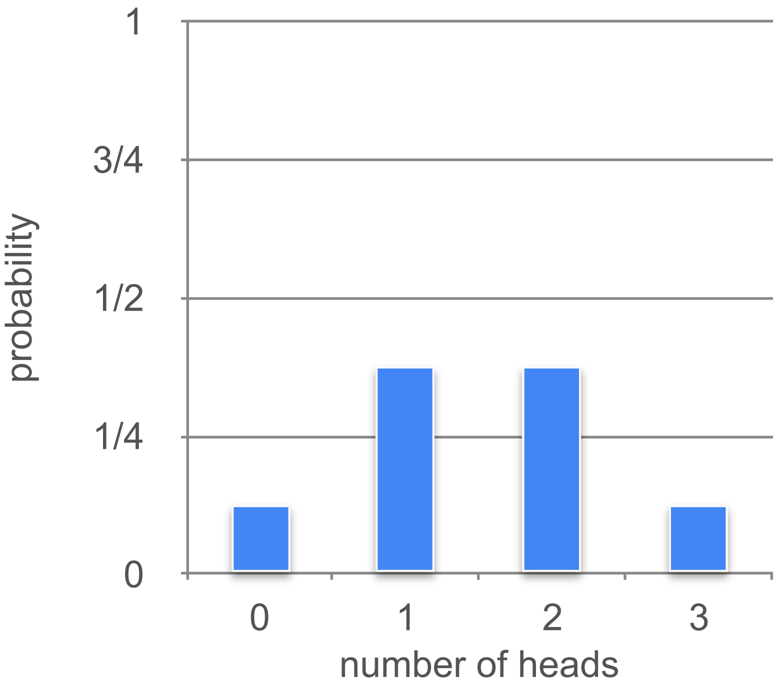


Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\mathbf{P}(T) = 0.5$

$X = 1$ $X = 0$

What can we say about the probability distribution when the number of coin flips increases?



Central Limit Theorem (CLT) - Example 1

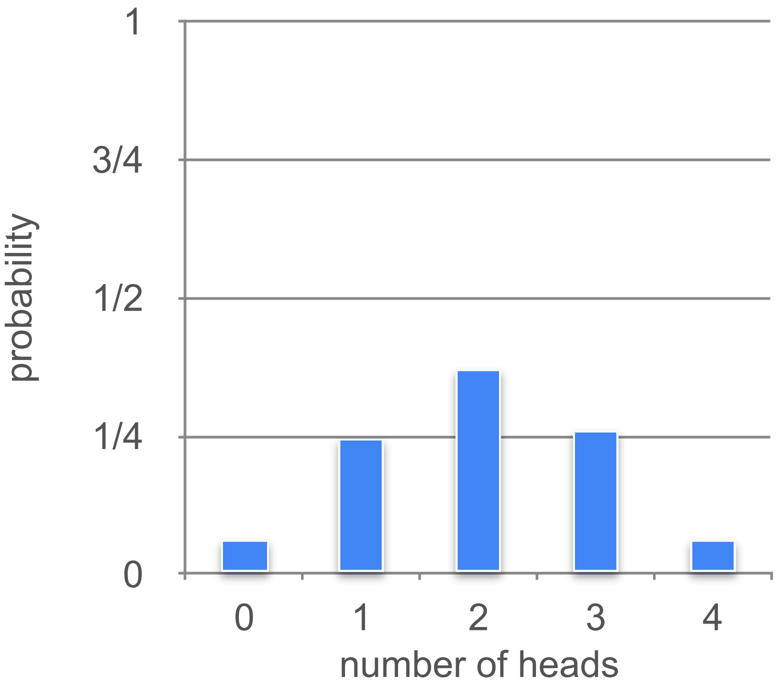
$\mathbf{P}(H) = 0.5$

$X = 1$

$\mathbf{P}(T) = 0.5$

$X = 0$

What can we say about the probability distribution when the number of coin flips increases?

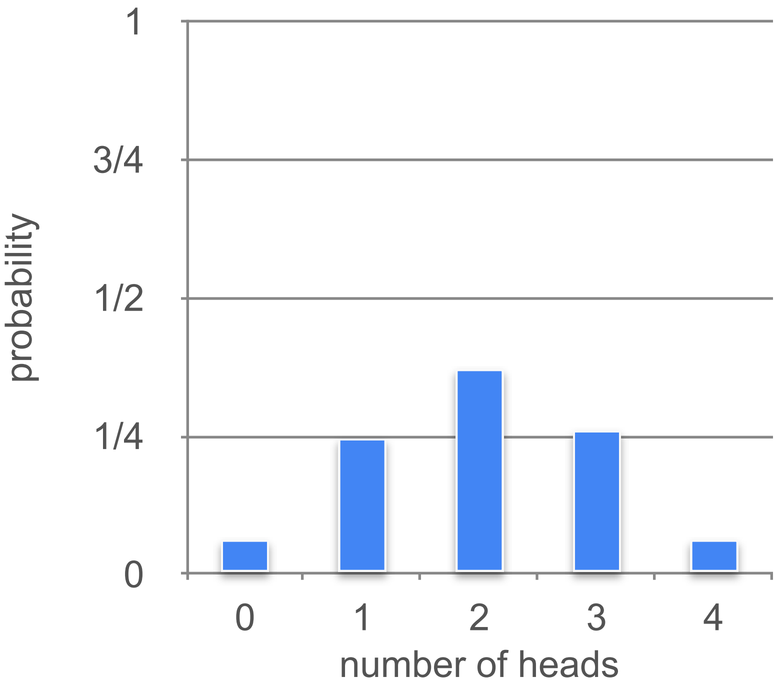


Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\mathbf{P}(T) = 0.5$

$X = 1$ $X = 0$

What can we say about the distribution when the number of coins we flip increases?



Central Limit Theorem (CLT) - Example 1

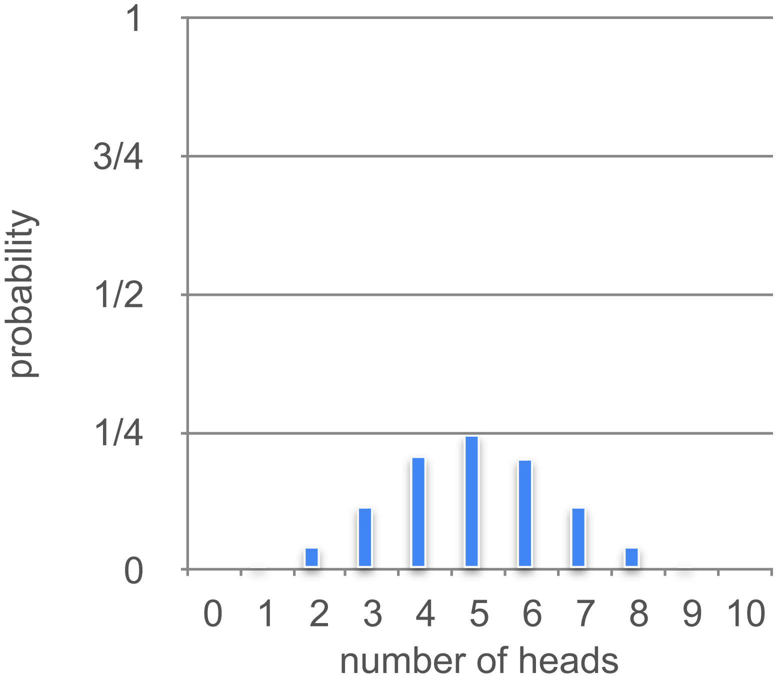
$\mathbf{P}(H) = 0.5$

$\mathbf{P}(T) = 0.5$

$X = 1$

$X = 0$

What can we say about the distribution when the number of coins we flip increases?

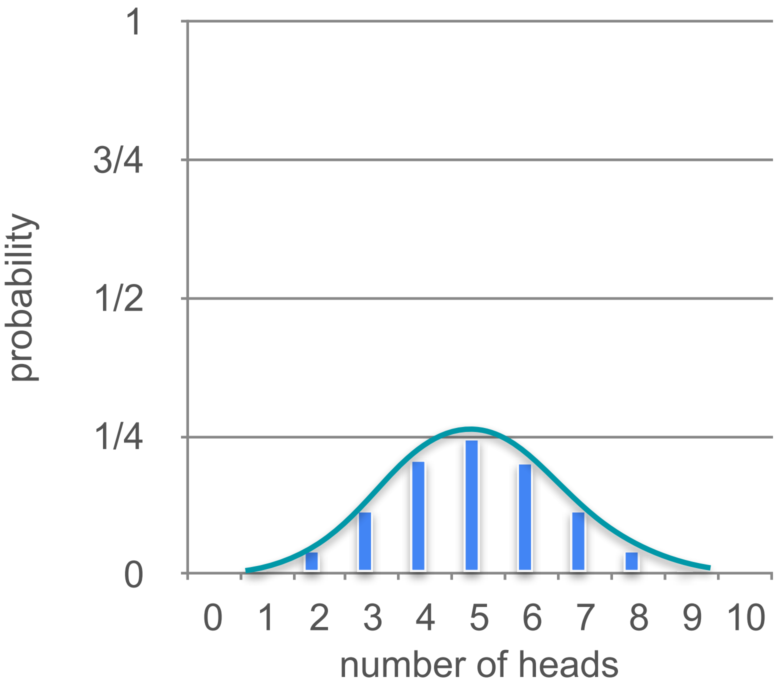


Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\mathbf{P}(T) = 0.5$

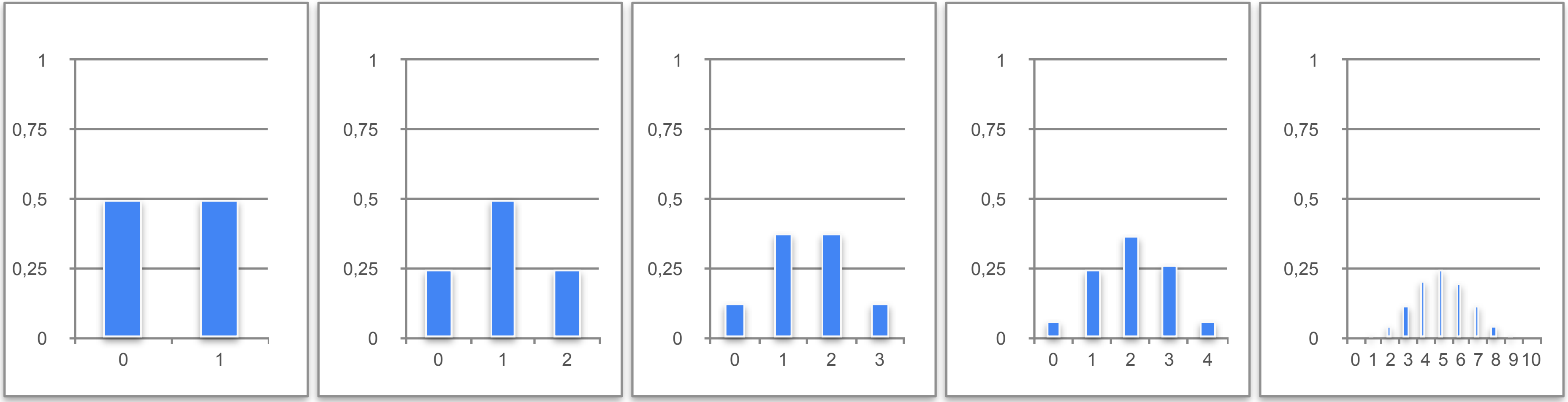
$X = 1$ $X = 0$

What can we say about the distribution when the number of coins we flip increases?

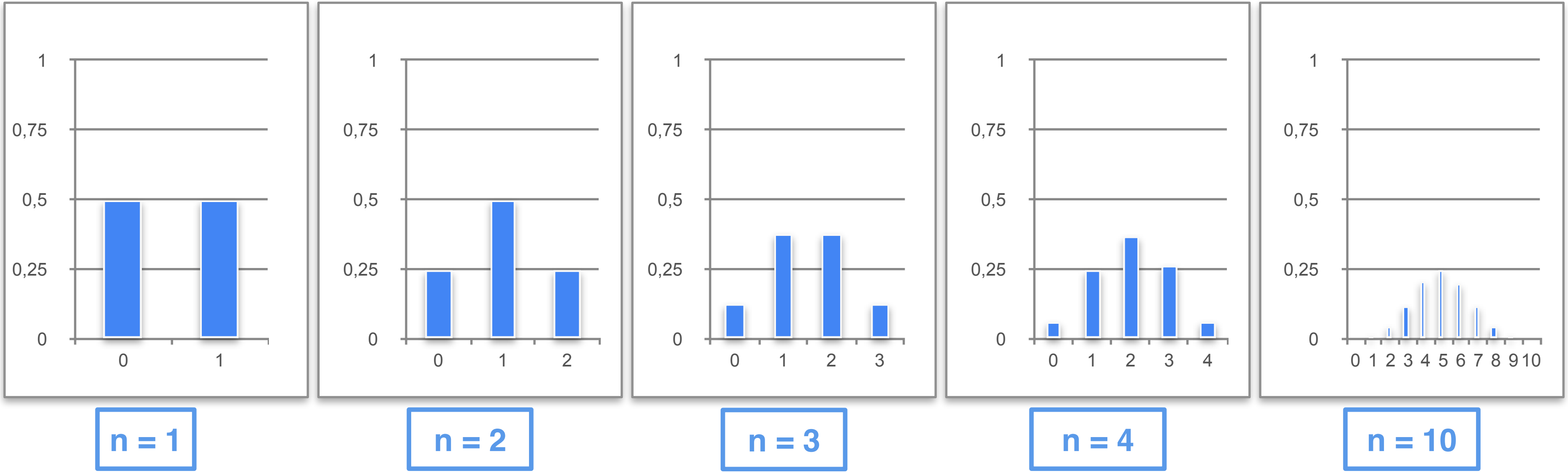


Central Limit Theorem (CLT) - Example 1

Central Limit Theorem (CLT) - Example 1

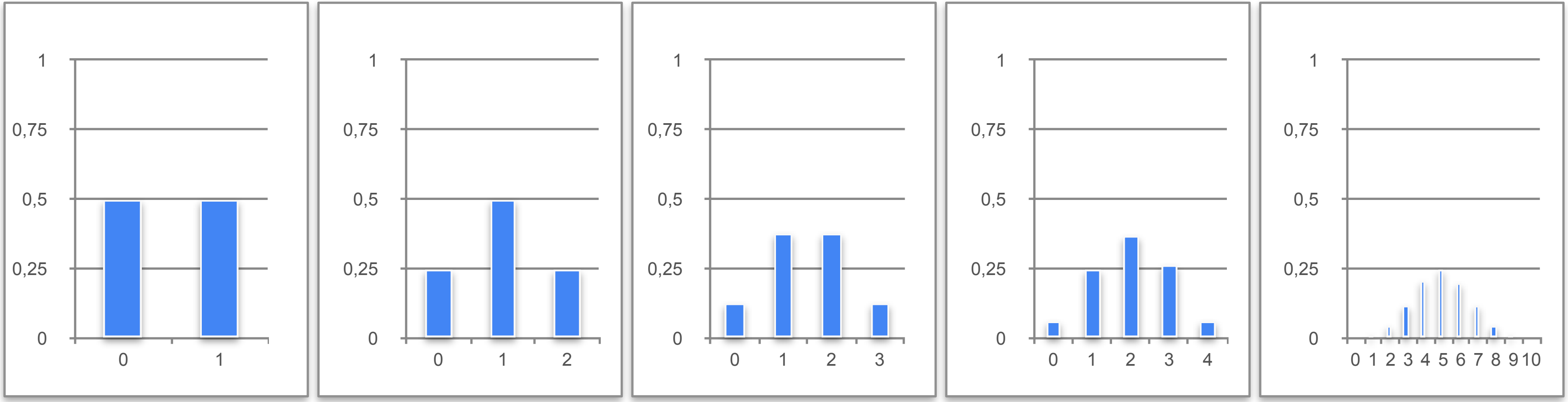


Central Limit Theorem (CLT) - Example 1



DeepLearning.AI

Central Limit Theorem (CLT) - Example 1



n = 1 | n = 2 | n = 3 | n = 4 | n = 10

As n increases, the probability distribution becomes closer to a Gaussian distribution

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$

$\mathbf{P}(T) = 0.5$

Random variable $\rightarrow X$ **number of heads when a coin is flipped n times**

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$

$\mathbf{P}(T) = 0.5$

Random variable → X **number of heads when a coin is flipped n times**

$$\mu = np$$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$

$\mathbf{P}(T) = 0.5$

Random variable $\rightarrow X$ **number of heads when a coin is flipped n times**

$$\mu = np \;\; = n\mathbf{P}(H)$$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$

$\mathbf{P}(T) = 0.5$

Random variable → X **number of heads when a coin is flipped n times**

$$\mu = np \quad = n\mathbf{P}(H)$$

$$\sigma^2 = np(1-p)$$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$

$\mathbf{P}(T) = 0.5$

Random variable → X **number of heads when a coin is flipped n times**

$$\mu = np \quad = n\mathbf{P}(H)$$

$$\sigma^2 = np(1-p) \quad = n\mathbf{P}(H)\mathbf{P}(T)$$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\mathbf{P}(T) = 0.5$

$$\mu = np \quad = n\mathbf{P}(H)$$

$$\sigma^2 = np(1-p) \quad = n\mathbf{P}(H)\mathbf{P}(T)$$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\quad$ $\mathbf{P}(T) = 0.5$

$$\mu = np \quad = n\mathbf{P}(H)$$

$$\sigma^2 = np(1-p) \quad = n\mathbf{P}(H)\mathbf{P}(T)$$

Central Limit Theorem (CLT) - Example 1

$n = 1$

$\mathbf{P}(H) = 0.5$ $\mathbf{P}(T) = 0.5$

$\mu = np \quad = n\mathbf{P}(H)$

$\sigma^2 = np(1-p) \quad = n\mathbf{P}(H)\mathbf{P}(T)$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\mathbf{P}(T) = 0.5$

$\mu = np \quad = n\mathbf{P}(H)$

$\sigma^2 = np(1 - p) \quad = n\mathbf{P}(H)\mathbf{P}(T)$

$n = 1$

$\mu = np$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\mathbf{P}(T) = 0.5$

$\mu = np \quad = n\mathbf{P}(H)$

$\sigma^2 = np(1-p) \quad = n\mathbf{P}(H)\mathbf{P}(T)$

$n = 1$

$\mu = np$

$\mu = 1 \times 0.5$

Central Limit Theorem (CLT) - Example 1

$$\mathbf{P}(H) = 0.5 \qquad \mathbf{P}(T) = 0.5$$

$$\mu = np \quad = n\mathbf{P}(H)$$

$$\sigma^2 = np(1-p) \quad = n\mathbf{P}(H)\mathbf{P}(T)$$

$$n = 1$$

$$\mu = np$$

$$\mu = 1 \times 0.5 \quad = 0.5$$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\quad$ $\mathbf{P}(T) = 0.5$

$$\mu = np \quad = n\mathbf{P}(H)$$

$$\sigma^2 = np(1-p) \quad = n\mathbf{P}(H)\mathbf{P}(T)$$

$n = 1$

$$\mu = np$$

$$\mu = 1 \times 0.5 \quad = 0.5$$

$$\sigma^2 = np(1-p)$$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5 \qquad \mathbf{P}(T) = 0.5$

$\mu = np \quad = n\mathbf{P}(H)$

$\sigma^2 = np(1-p) \quad = n\mathbf{P}(H)\mathbf{P}(T)$

$n = 1$

$\mu = np$

$\mu = 1 \times 0.5 \quad = 0.5$

$\sigma^2 = np(1-p)$

$\sigma^2 = (1 \times 0.5)(0.5)$

Central Limit Theorem (CLT) - Example 1

$\mathbf{P}(H) = 0.5$ $\quad$ $\mathbf{P}(T) = 0.5$

$$\mu = np \quad = n\mathbf{P}(H)$$

$$\sigma^2 = np(1-p) \quad = n\mathbf{P}(H)\mathbf{P}(T)$$

$n = 1$

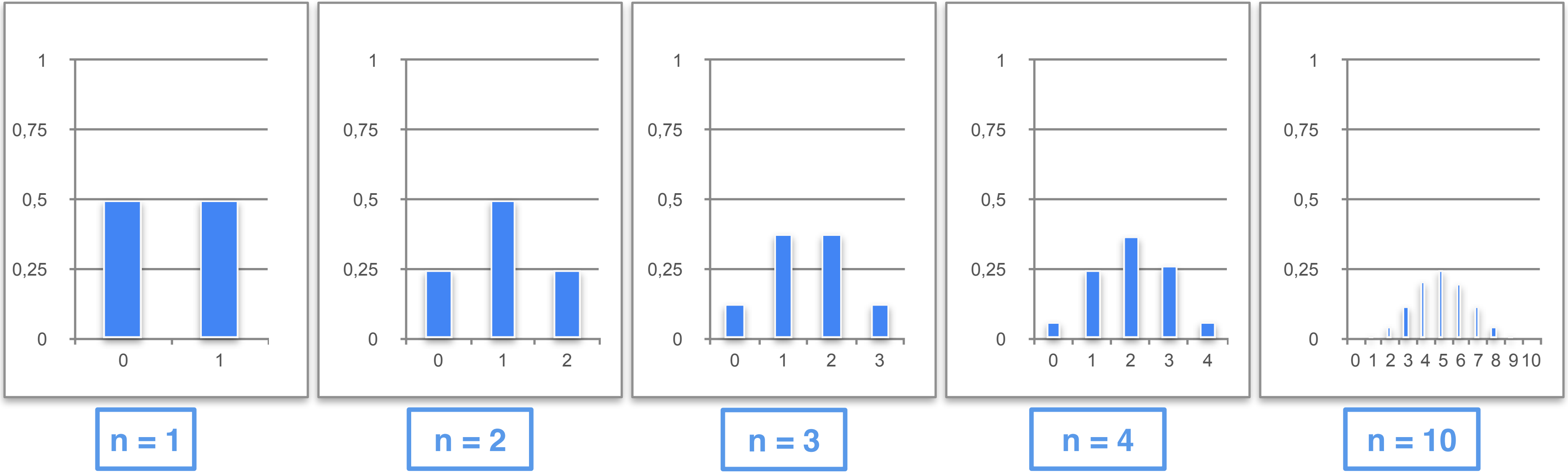
$$\mu = np$$

$$\mu = 1 \times 0.5 \quad = 0.5$$

$$\sigma^2 = np(1-p)$$

$$\sigma^2 = (1 \times 0.5)(0.5) \;= 0.25$$

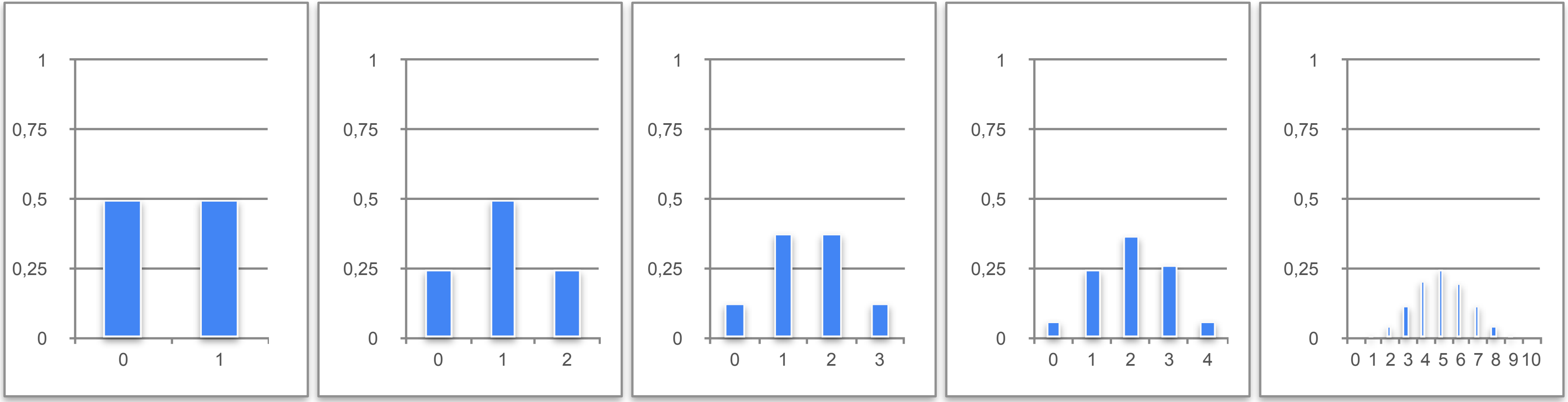
Central Limit Theorem (CLT) - Example 1



As n increases, the probability distribution becomes closer to a gaussian distribution

Central Limit Theorem (CLT) - Example 1

$$\mu = np$$

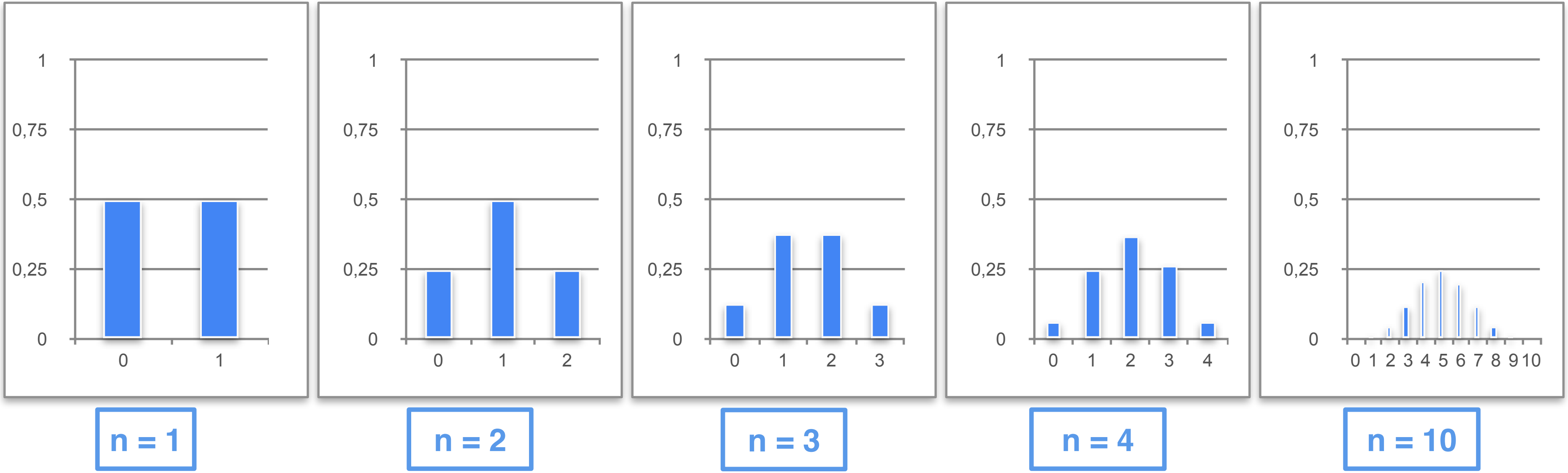


n = 1 n = 2 n = 3 n = 4 n = 10

As n increases, the probability distribution becomes closer to a gaussian distribution

Central Limit Theorem (CLT) - Example 1

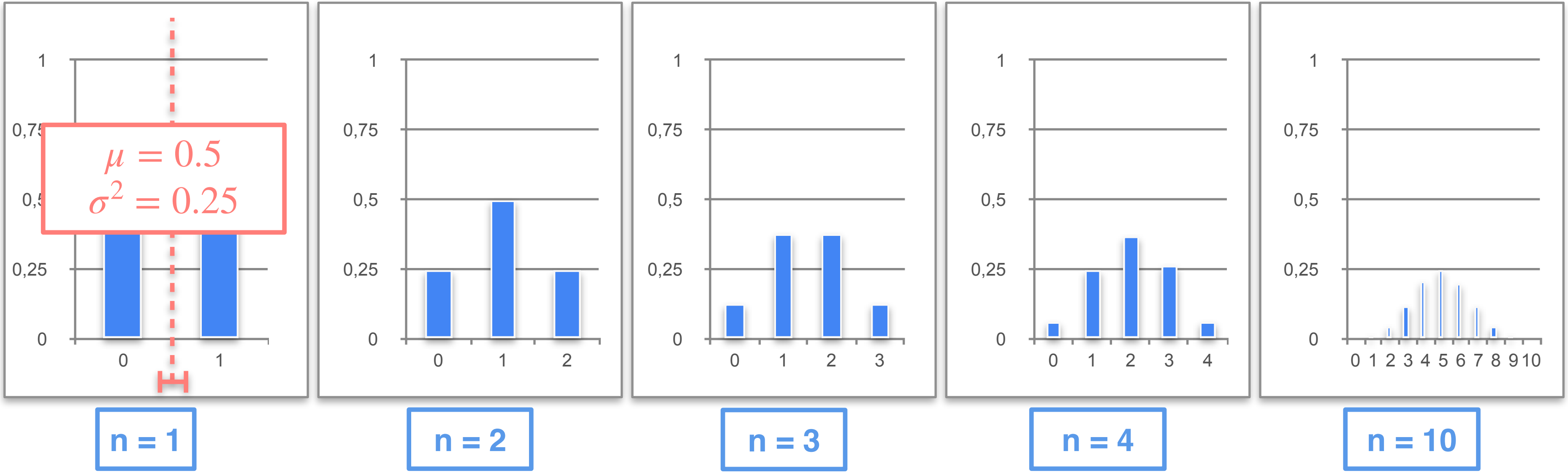
$$\mu = np \qquad \sigma^2 = np(1-p)$$



As n increases, the probability distribution becomes closer to a gaussian distribution

Central Limit Theorem (CLT) - Example 1

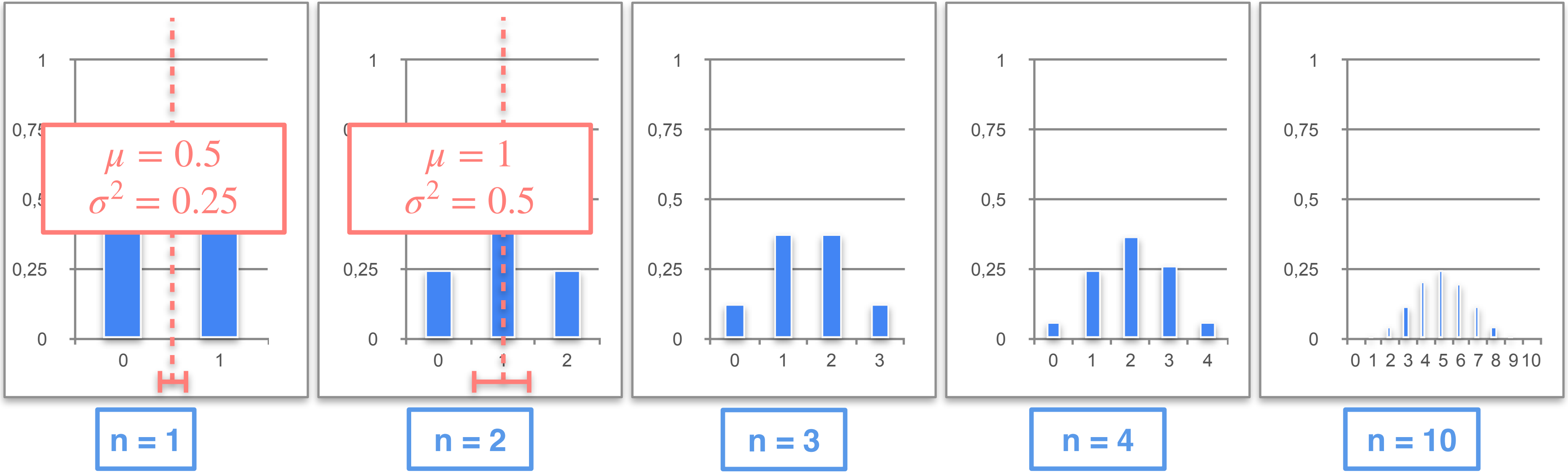
$$\mu = np \qquad \sigma^2 = np(1-p)$$



As n increases, the probability distribution becomes closer to a gaussian distribution

Central Limit Theorem (CLT) - Example 1

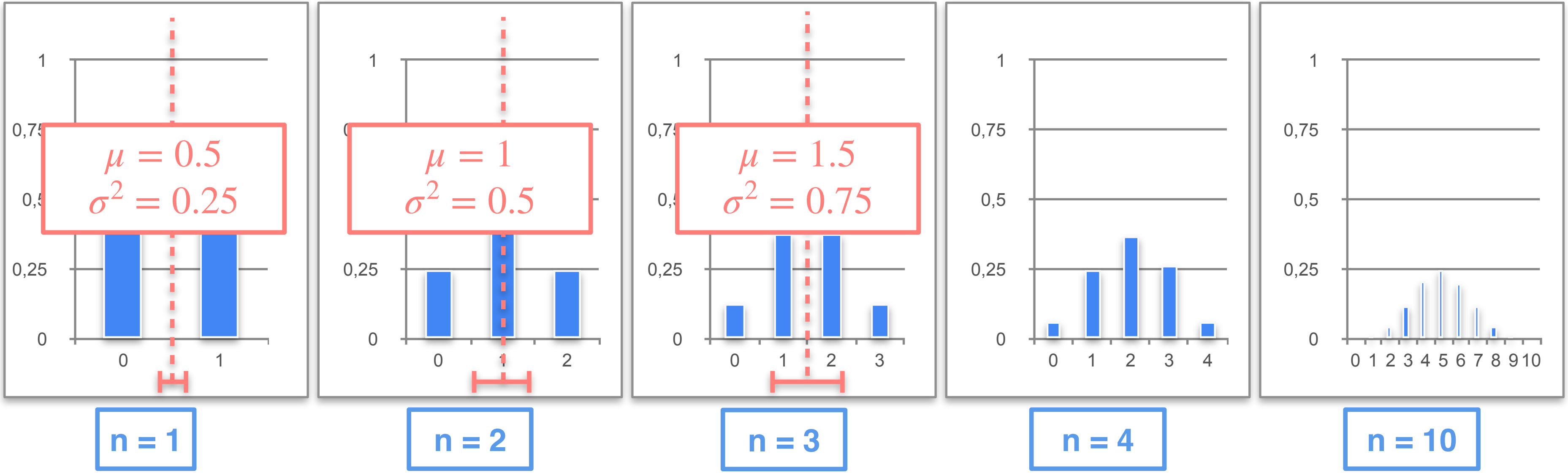
$$\mu = np \qquad \sigma^2 = np(1-p)$$



As n increases, the probability distribution becomes closer to a gaussian distribution

Central Limit Theorem (CLT) - Example 1

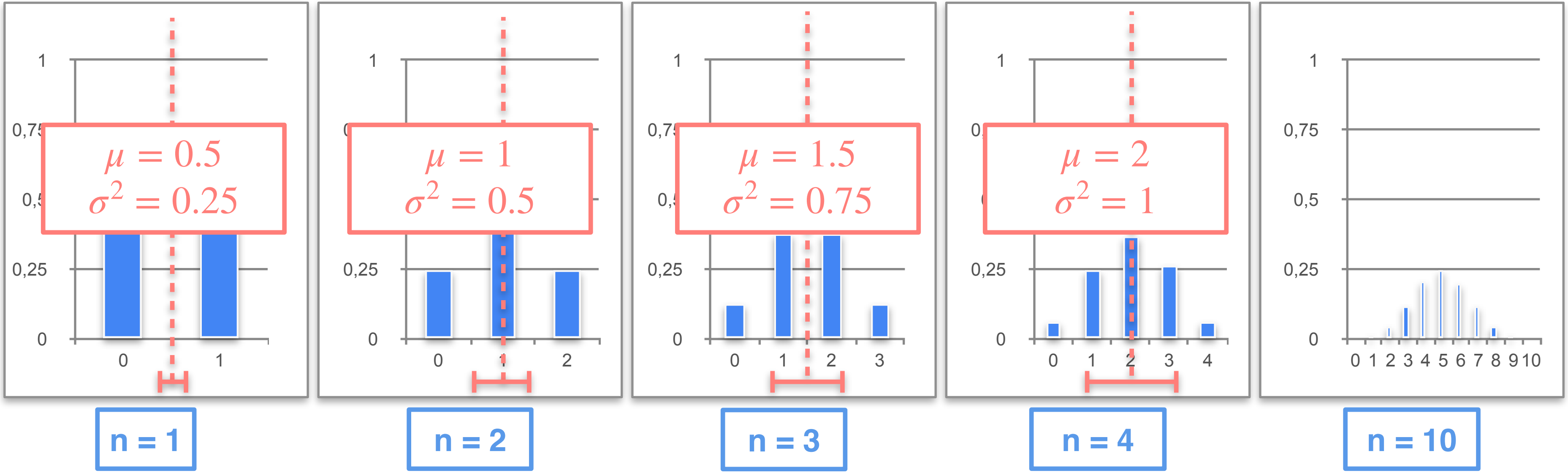
$\mu = np$ $\qquad \sigma^2 = np(1-p)$



As n increases, the probability distribution becomes closer to a gaussian distribution

Central Limit Theorem (CLT) - Example 1

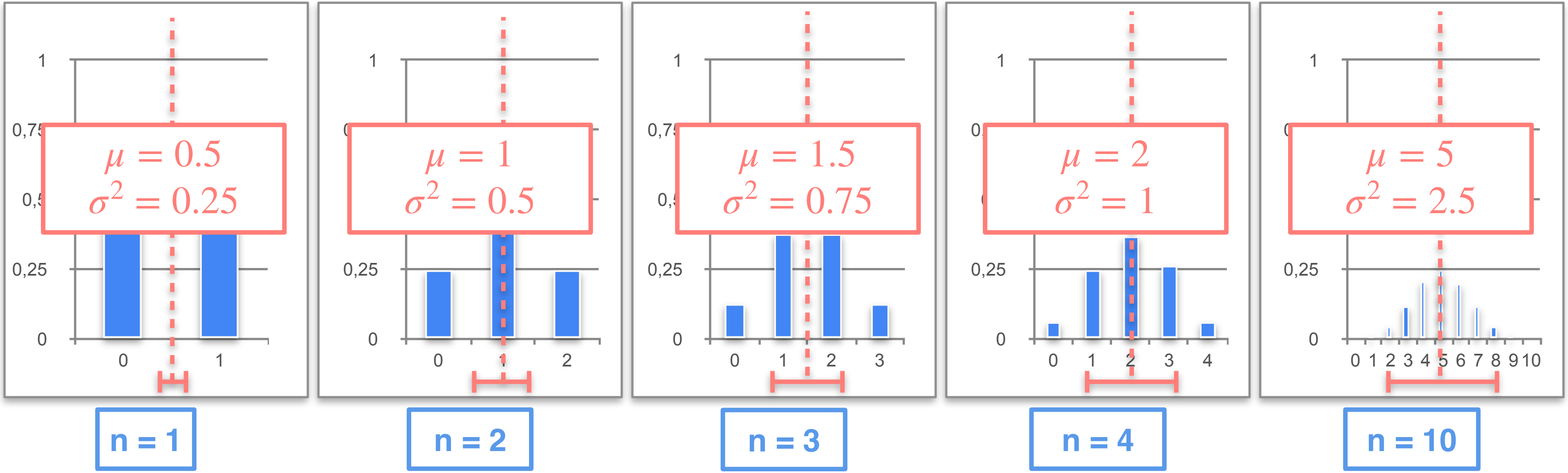
$\mu = np$ $\sigma^2 = np(1-p)$



As n increases, the probability distribution becomes closer to a gaussian distribution

Central Limit Theorem (CLT) - Example 1

$$\mu = np \qquad \sigma^2 = np(1-p)$$



As n increases, the probability distribution becomes closer to a gaussian distribution

Central Limit Theorem (CLT) - Example 1

$\mu = 0.5$
$\sigma^2 = 0.25$

$\mu = 1$
$\sigma^2 = 0.5$

$\mu = 1.5$
$\sigma^2 = 0.75$

$\mu = 2$
$\sigma^2 = 1$

$\mu = 5$
$\sigma^2 = 2.5$

Central Limit Theorem (CLT) - Example 1

$\mu = 0.5$ $\sigma^2 = 0.25$	$\mu = 1$ $\sigma^2 = 0.5$	$\mu = 1.5$ $\sigma^2 = 0.75$	$\mu = 2$ $\sigma^2 = 1$	$\mu = 5$ $\sigma^2 = 2.5$

as n become sufficiently large we will get a normal distribution with parameters

Central Limit Theorem (CLT) - Example 1

$\mu = 0.5$ $\sigma^2 = 0.25$	$\mu = 1$ $\sigma^2 = 0.5$	$\mu = 1.5$ $\sigma^2 = 0.75$	$\mu = 2$ $\sigma^2 = 1$	$\mu = 5$ $\sigma^2 = 2.5$

as n become sufficiently large we will get a normal distribution with parameters

$$\mu = np$$

Central Limit Theorem (CLT) - Example 1

$\mu = 0.5$ $\sigma^2 = 0.25$	$\mu = 1$ $\sigma^2 = 0.5$	$\mu = 1.5$ $\sigma^2 = 0.75$	$\mu = 2$ $\sigma^2 = 1$	$\mu = 5$ $\sigma^2 = 2.5$

as n become sufficiently large we will get a normal distribution with parameters

$$\mu = np$$

$$\sigma^2 = np(1-p)$$

Central Limit Theorem (CLT) - Example 1

$$\mu = np$$

$\mu = 0.5$ $\sigma^2 = 0.25$	$\mu = 1$ $\sigma^2 = 0.5$	$\mu = 1.5$ $\sigma^2 = 0.75$	$\mu = 2$ $\sigma^2 = 1$	$\mu = 5$ $\sigma^2 = 2.5$

as n become sufficiently large we will get a normal distribution with parameters

$$\mu = np$$

$$\sigma^2 = np(1 - p)$$

Central Limit Theorem (CLT) - Example 1

$\mu = np$ $\sigma^2 = np(1-p)$

$\mu = 0.5$ $\sigma^2 = 0.25$	$\mu = 1$ $\sigma^2 = 0.5$	$\mu = 1.5$ $\sigma^2 = 0.75$	$\mu = 2$ $\sigma^2 = 1$	$\mu = 5$ $\sigma^2 = 2.5$

as n become sufficiently large we will get a normal distribution with parameters

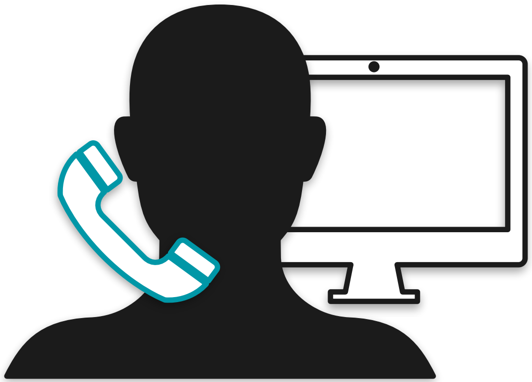
$$\mu = np$$

$$\sigma^2 = np(1-p)$$

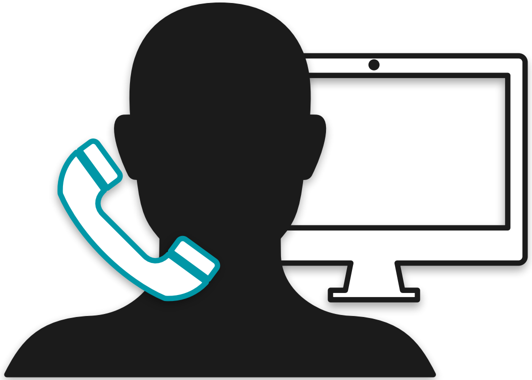
Video 4b: Central Limit Theorem

- Example 2: Continuous Random Variable (Sample mean of a sampling distribution)

Uniform Distribution: Motivation

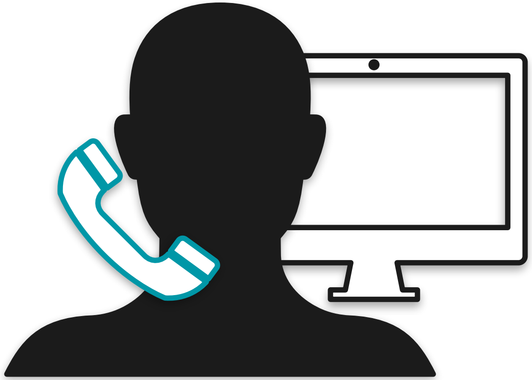


Uniform Distribution: Motivation



You’re calling a tech support line.
They can answer any time between zero and 15 minutes and if they don't answer in this time, the line is disconnected.

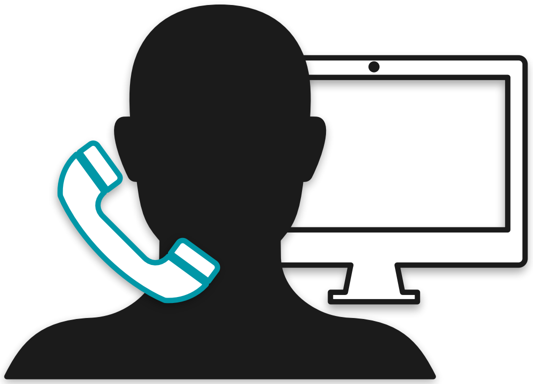
Uniform Distribution: Motivation



You're calling a tech support line.
They can answer any time between zero and 15 minutes and if they don't answer in this time, the line is disconnected.

$X =$ "Wait time for a called to be answered "

Uniform Distribution: Motivation



You’re calling a tech support line.
They can answer any time between zero and 15 minutes and if they don't answer in this time, the line is disconnected.

$X =$ ”Wait time for a called to be answered ”

$$X \sim \mathcal{U}(0,15)$$

Central Limit Theorem (CLT) - Example 2

$n = 1$ $$Y_1 = \frac{X_1}{1}$$

Central Limit Theorem (CLT) - Example 2

$n = 1$ $$Y_1 = \frac{X_1}{1}$$

$n = 2$ $$Y_2 = \frac{X_1 + X_2}{2}$$

Central Limit Theorem (CLT) - Example 2

$n = 1$ $$Y_1 = \frac{X_1}{1}$$

$n = 2$ $$Y_2 = \frac{X_1 + X_2}{2}$$

$n = 3$ $$Y_3 = \frac{X_1 + X_2 + X_3}{3}$$

$\vdots$ $\vdots$

Central Limit Theorem (CLT) - Example 2

$n = 1$ $\quad Y_1 = \frac{X_1}{1}$

$n = 2$ $\quad Y_2 = \frac{X_1 + X_2}{2}$

$n = 3$ $\quad Y_3 = \frac{X_1 + X_2 + X_3}{3}$

$\vdots \qquad \vdots$

Record the average of all n experiments

Central Limit Theorem (CLT) - Example 2

$n = 1$ $\quad Y_1 = \frac{X_1}{1}$

$n = 2$ $\quad Y_2 = \frac{X_1 + X_2}{2}$

$n = 3$ $\quad Y_3 = \frac{X_1 + X_2 + X_3}{3}$

$\vdots \qquad \vdots$

Record the average of all n experiments

$$Y_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Central Limit Theorem (CLT) - Example 2

$n = 1$ $\qquad Y_1 = \frac{X_1}{1}$

$n = 2$ $\qquad Y_2 = \frac{X_1 + X_2}{2}$

$n = 3$ $\qquad Y_3 = \frac{X_1 + X_2 + X_3}{3}$

$\vdots \qquad \vdots$

Record the average of all n experiments

$$Y_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Can we say anything about the distribution of this average?

Central Limit Theorem (CLT) - Example 2

$n = 1 \qquad Y_1 = \frac{X_1}{1}$

Central Limit Theorem (CLT) - Example 2

$n = 1 \qquad Y_1 = \frac{X_1}{1}$

What happens to the distribution of these averages as n increases?

Central Limit Theorem (CLT) - Example 2

$$n = 1 \qquad Y_1 = \frac{X_1}{1}$$

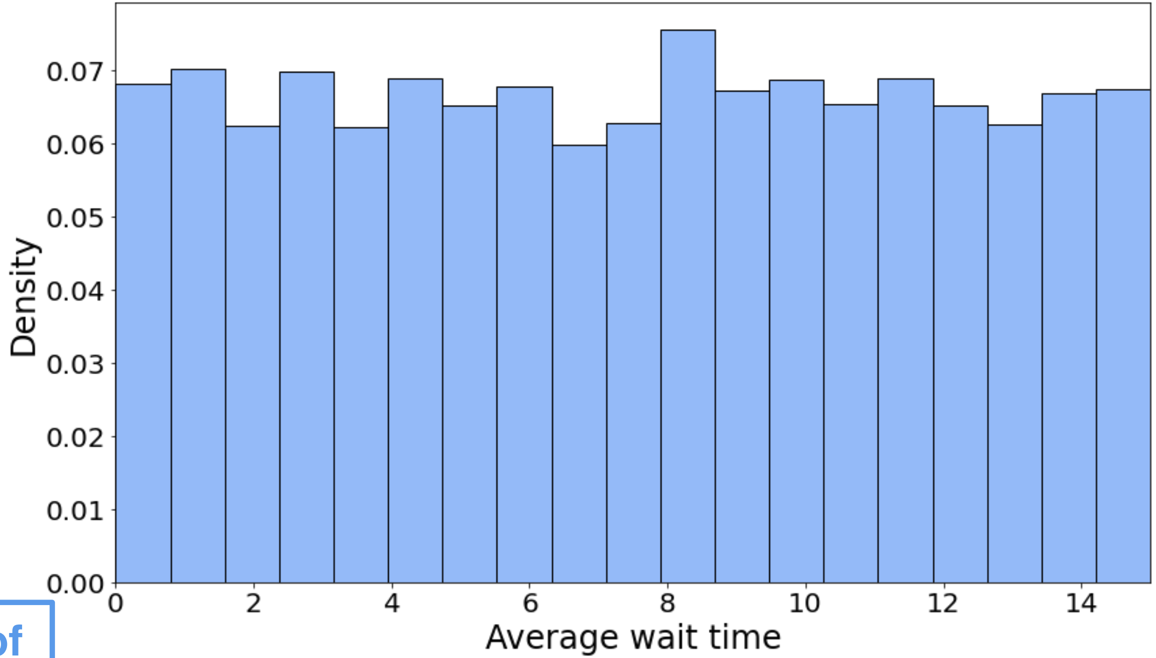
Create many samples of Y_1 so you can get a pretty histogram

What happens to the distribution of these averages as n increases?

Central Limit Theorem (CLT) - Example 2

$n = 1 \qquad Y_1 = \frac{X_1}{1}$

Create many samples of Y_1 so you can get a pretty histogram

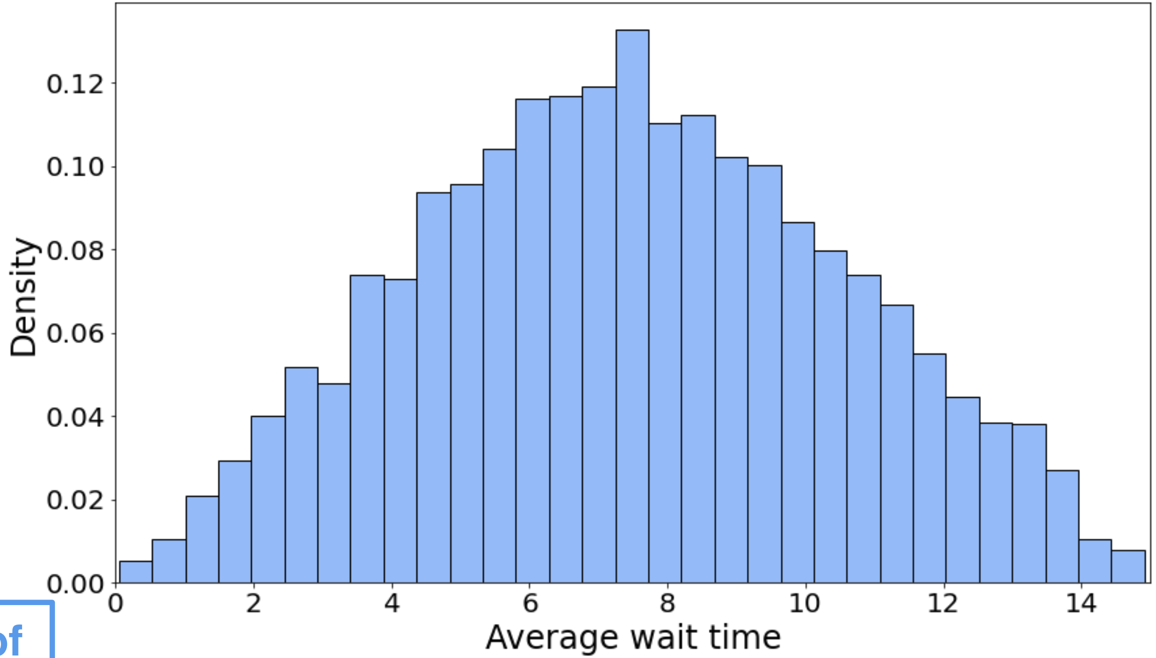


What happens to the distribution of these averages as n increases?

Central Limit Theorem (CLT) - Example 2

$n = 2$ $\qquad Y_2 = \dfrac{X_1 + X_2}{2}$

Create many samples of Y_2 so you can get a pretty histogram

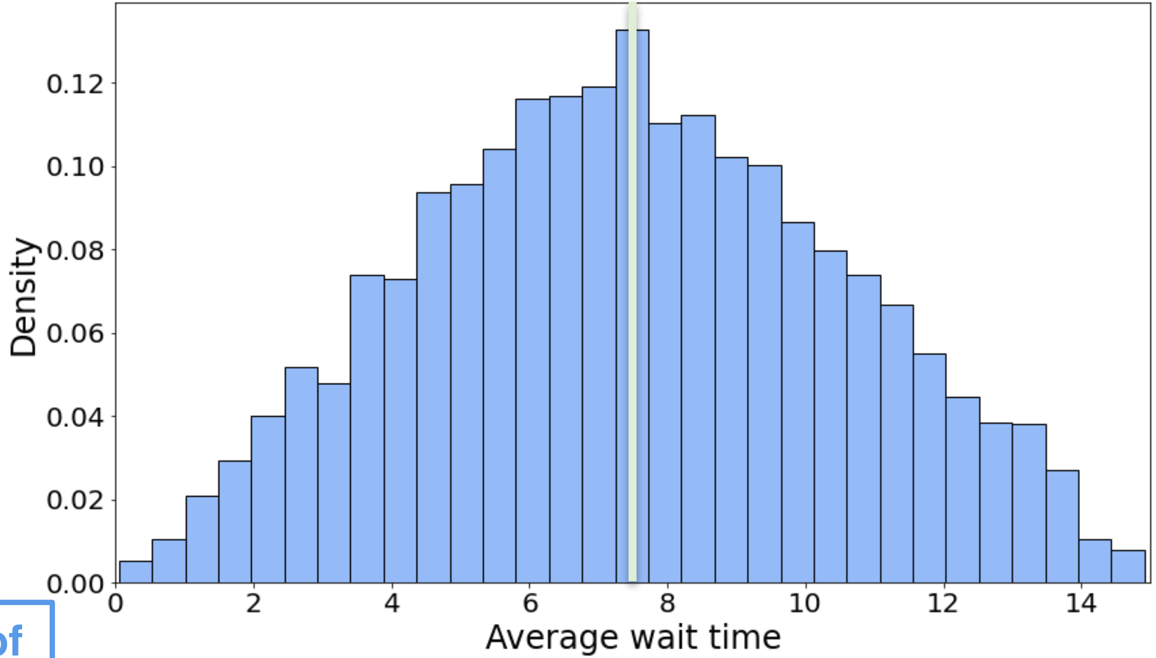


What happens to the distribution of these averages as n increases

Central Limit Theorem (CLT) - Example 2

$n = 2$ $\qquad Y_2 = \dfrac{X_1 + X_2}{2}$

Create many samples of Y_2 so you can get a pretty histogram

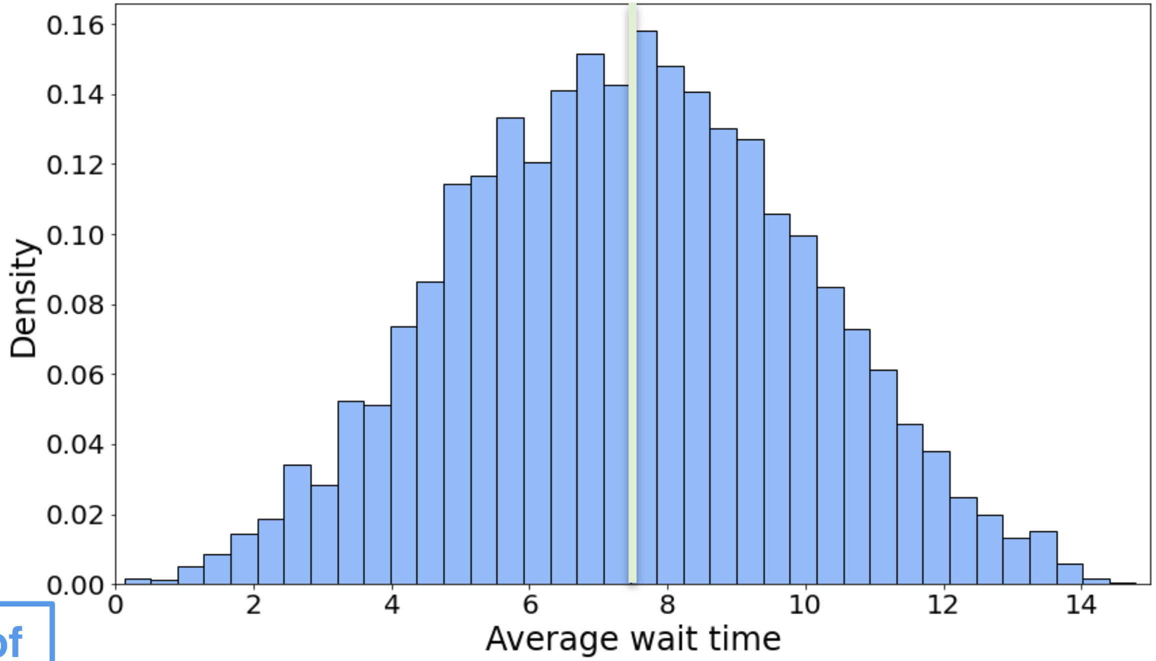


What happens to the distribution of these averages as n increases

Central Limit Theorem (CLT) - Example 2

$$n = 3 \quad Y_3 = \frac{X_1 + X_2 + X_3}{3}$$

Create many samples of Y_3 so you can get a pretty histogram

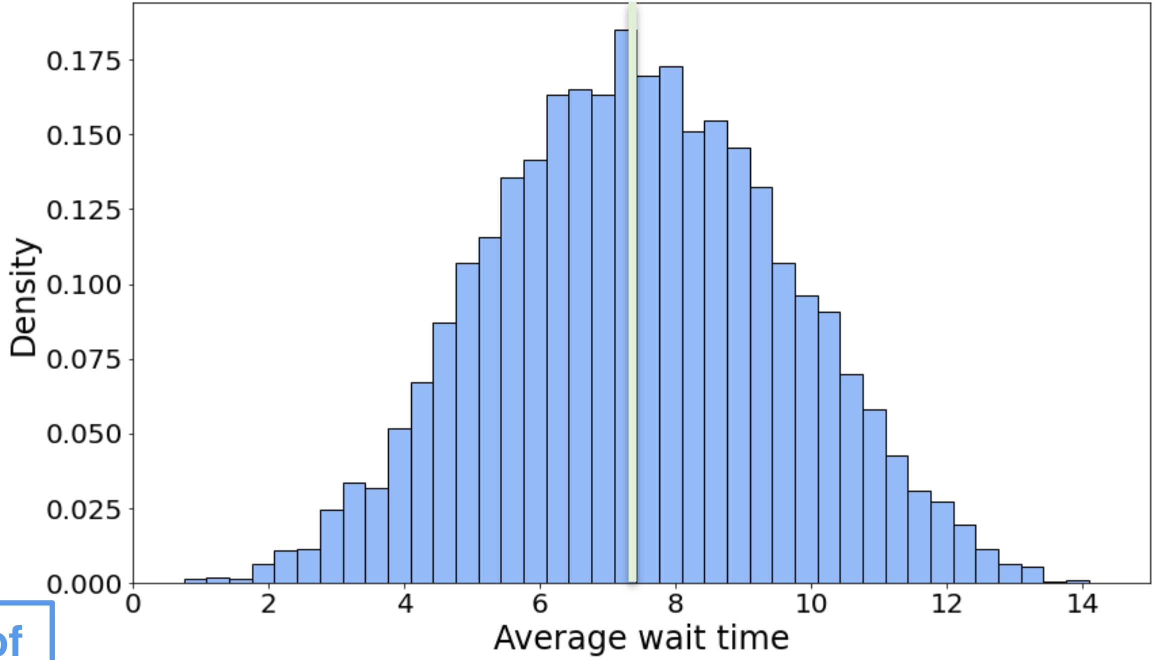


What happens to the distribution of these averages as n increases

Central Limit Theorem (CLT) - Example 2

$n = 4 \quad Y_4 = \dfrac{X_1 + X_2 + X_3 + X_4}{4}$

Create many samples of Y_4 so you can get a pretty histogram

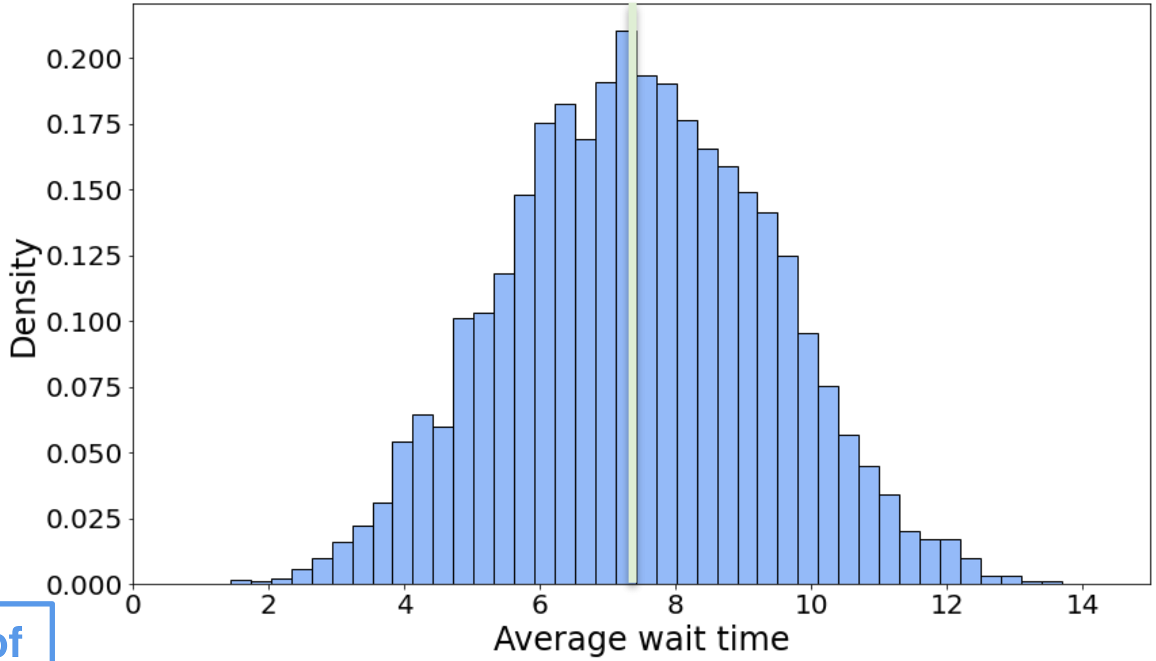


What happens to the distribution of these averages as n increases

Central Limit Theorem (CLT) - Example 2

$n = 5 \quad Y_5 = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$

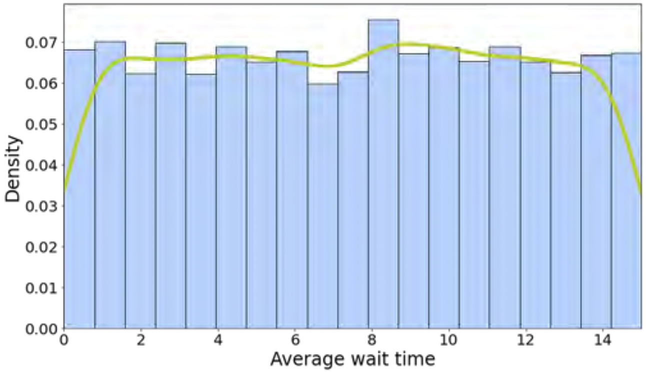
Create many samples of Y_5 so you can get a pretty histogram

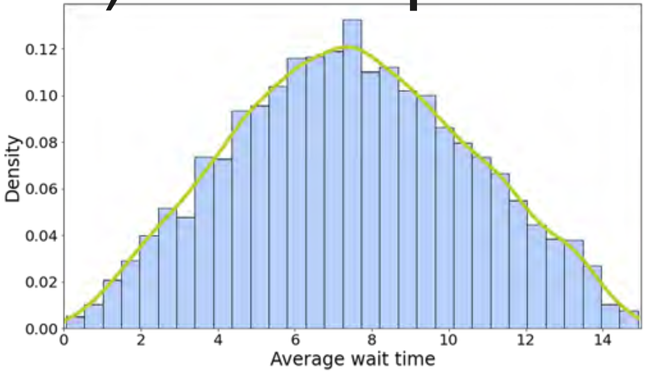


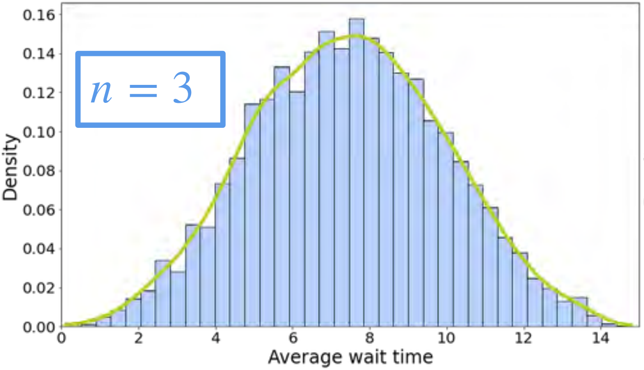
What happens to the distribution of these averages as n increases

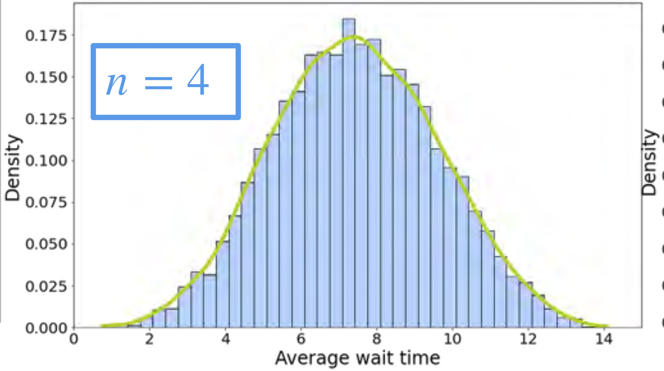
Central Limit Theorem (CLT) - Example 2

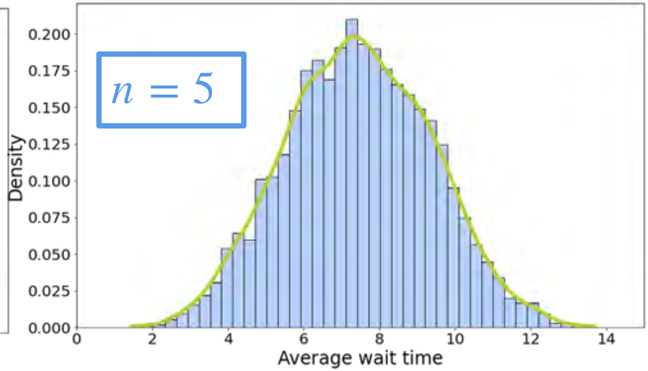
Central Limit Theorem (CLT) - Example 2











Central Limit Theorem (CLT) - Example 2

$$= \frac{1}{n^2}$$

Central Limit Theorem (CLT) - Example 2

$$\mathbb{E}[Y_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n^2}$$

Central Limit Theorem (CLT) - Example 2

$$\mathbb{E}[Y_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]$$

$$= \frac{1}{n^2}$$

Central Limit Theorem (CLT) - Example 2

$$\mathbb{E}[Y_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}n\mathbb{E}[X] = \mathbb{E}[X]$$

$$= \frac{1}{n^2}$$

Central Limit Theorem (CLT) - Example 2

$$\mathbb{E}[Y_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}n\mathbb{E}[X] = \mathbb{E}[X] = 7.5$$

$$= \frac{1}{n^2}$$

Central Limit Theorem (CLT) - Example 2

$$\mathbb{E}[Y_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}n\mathbb{E}[X] = \mathbb{E}[X] = 7.5$$

$$Var[Y_n] = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}$$

Central Limit Theorem (CLT) - Example 2

$$\mathbb{E}[Y_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}n\mathbb{E}[X] = \mathbb{E}[X] = 7.5$$

$$Var[Y_n] = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$$

Central Limit Theorem (CLT) - Example 2

$$\mathbb{E}[Y_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}n\mathbb{E}[X] = \mathbb{E}[X] = 7.5$$

$$Var[Y_n] = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$$

$$= \frac{1}{n^2} n Var(X) = \frac{Var(X)}{n}$$

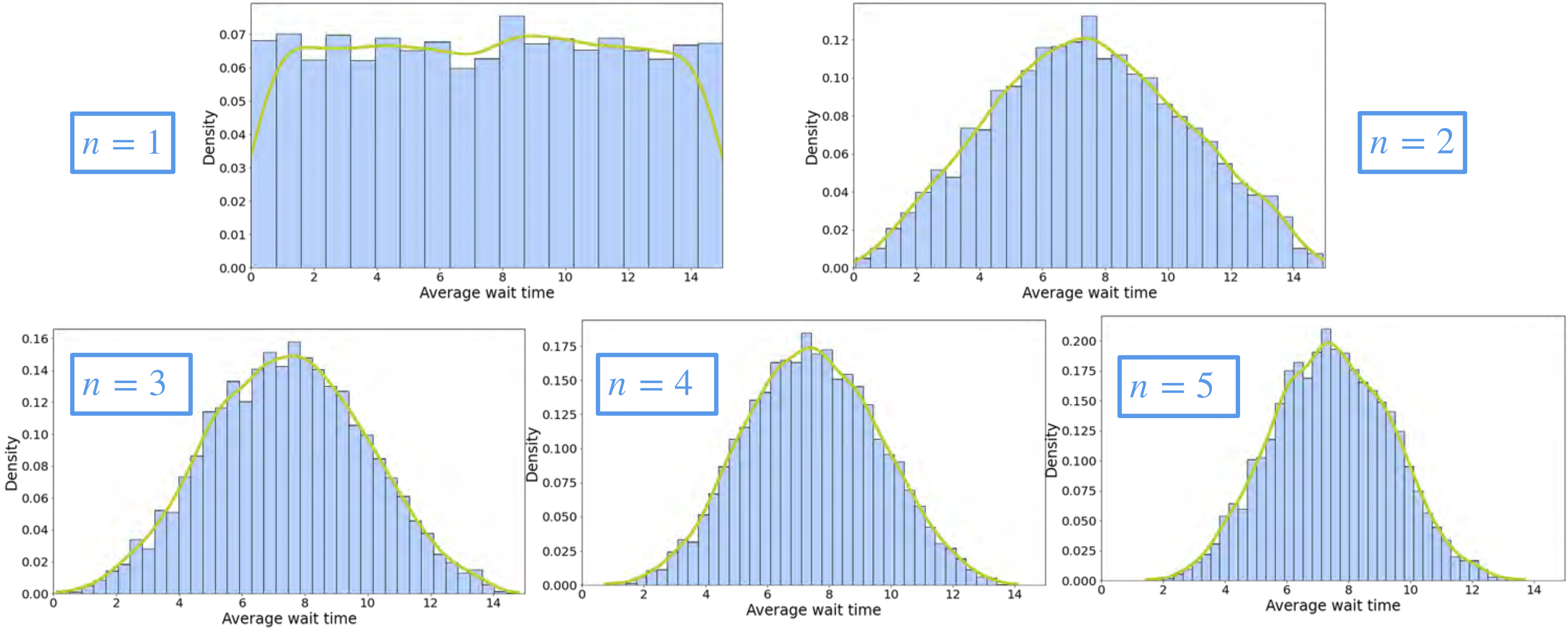
Central Limit Theorem (CLT) - Example 2

$$\mathbb{E}[Y_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n} n\mathbb{E}[X] = \mathbb{E}[X] = 7.5$$

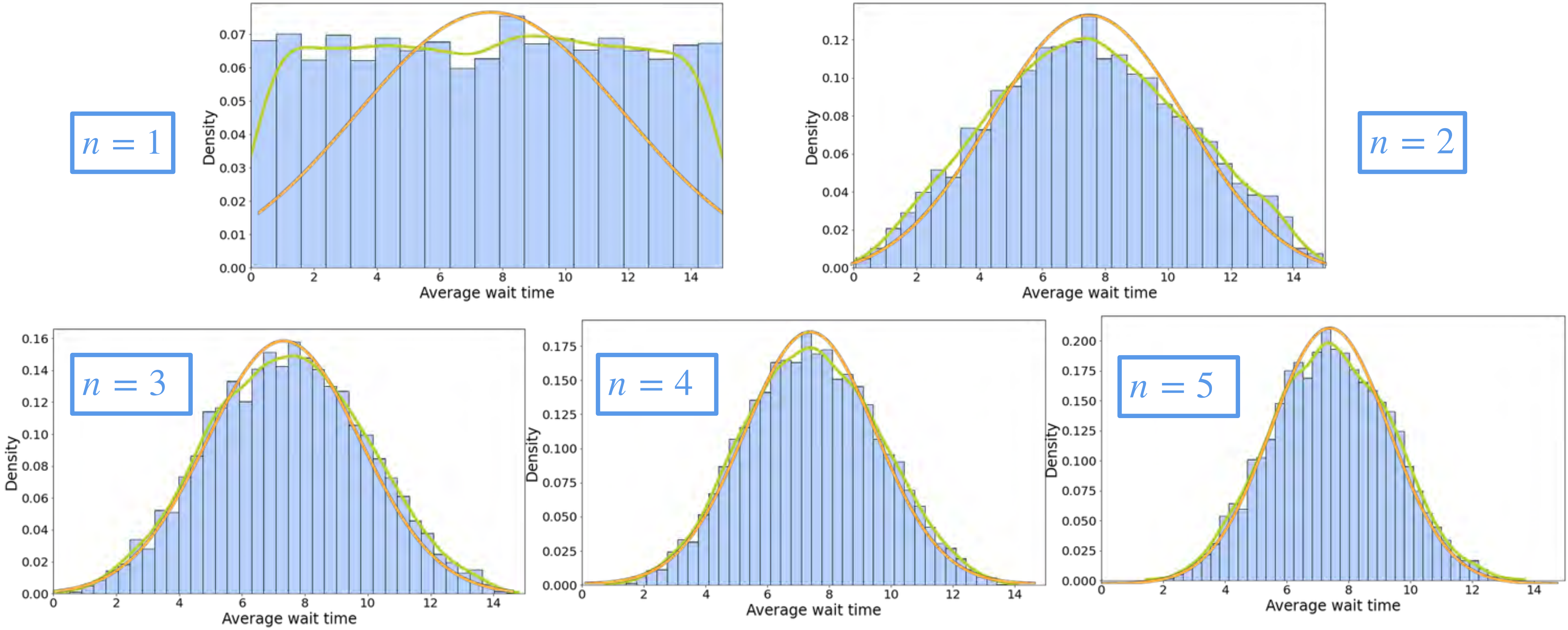
$$Var[Y_n] = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$$

$$= \frac{1}{n^2} nVar(X) = \frac{Var(X)}{n} = \frac{18.75}{n}$$

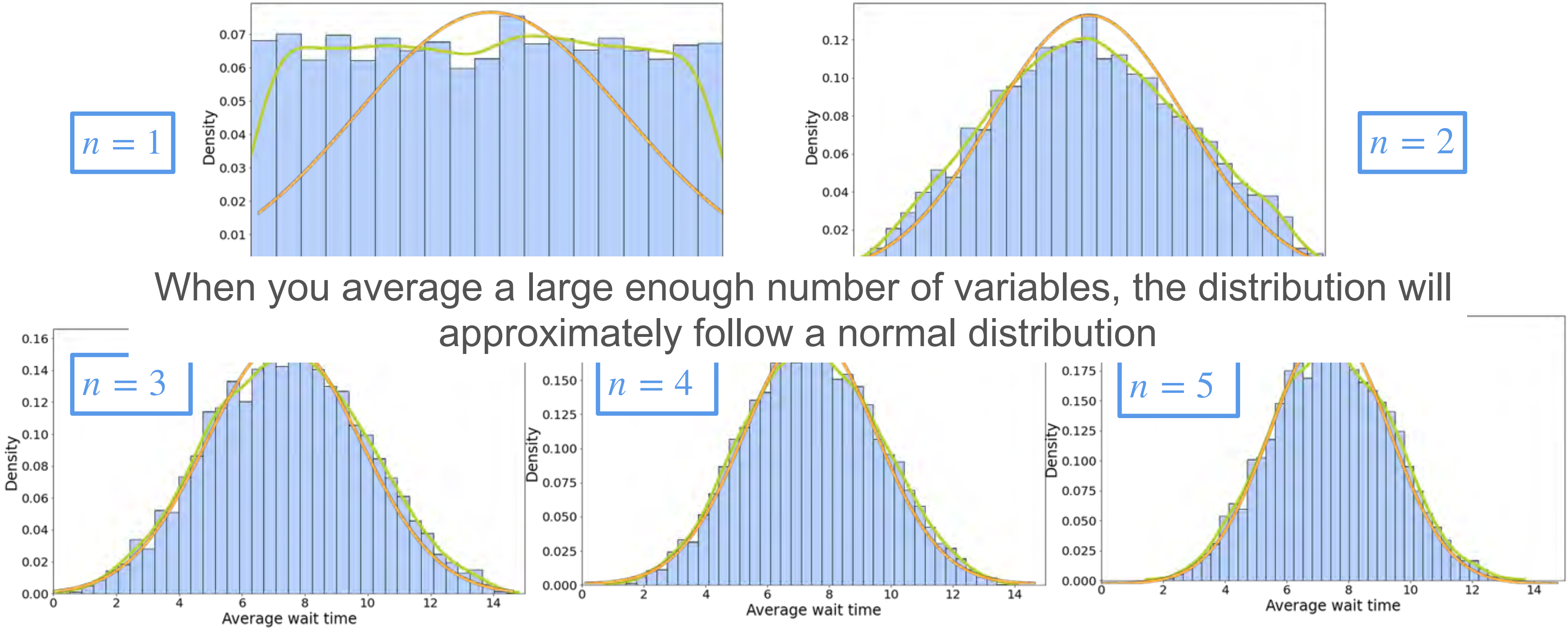
Central Limit Theorem (CLT) - Example 2



Central Limit Theorem (CLT) - Example 2



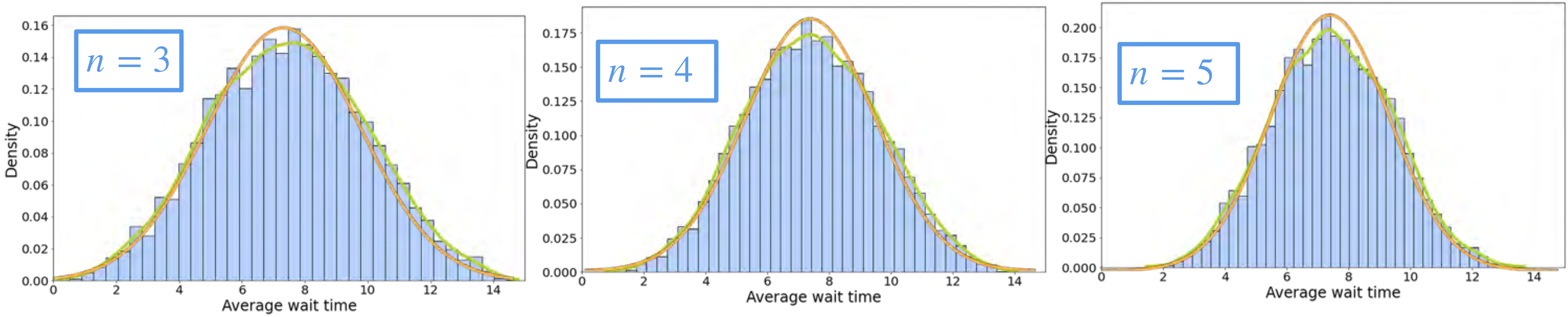
Central Limit Theorem (CLT) - Example 2



When you average a large enough number of variables, the distribution will approximately follow a normal distribution

Central Limit Theorem (CLT) - Example 2

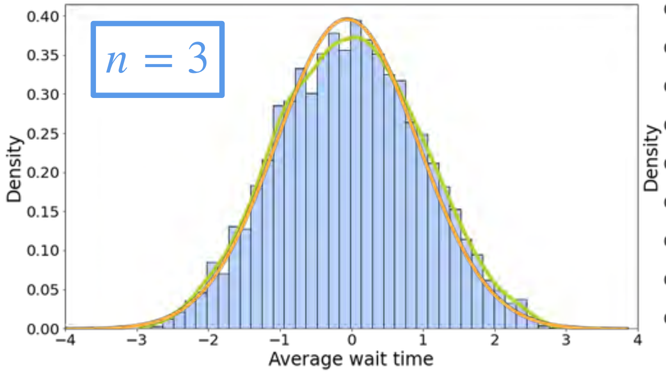
$$\frac{Y_n - 7.5}{\sqrt{18.75/n}}$$

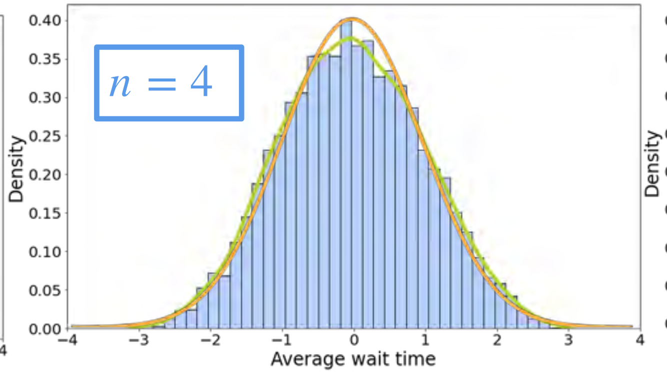


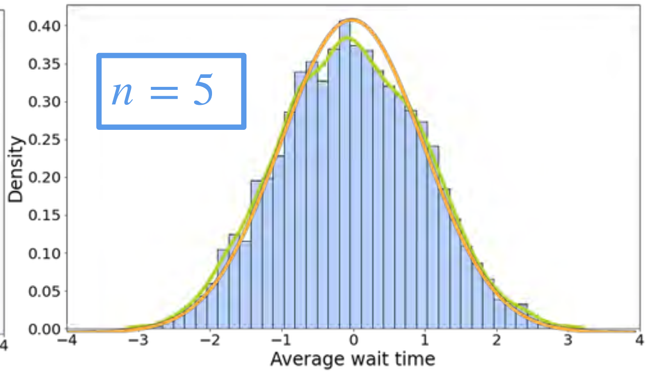
DeepLearning.AI

Central Limit Theorem (CLT) - Example 2

$$\frac{Y_n - 7.5}{\sqrt{18.75/n}} \xrightarrow{n\uparrow} \mathcal{N}(0,1)$$







Central Limit Theorem (CLT) - Example 2

Central Limit Theorem (CLT) - Example 2

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Central Limit Theorem (CLT) - Example 2

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\underbrace{\mu_{Y_n}}_{\text{mean of } Y_n} = \underbrace{\mu}_{\text{population mean}}$$

Central Limit Theorem (CLT) - Example 2

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

mean of Y_n

$$\mu_{Y_n} = \mu$$

population mean

population variance

variance of Y_n

$$\sigma_{Y_n}^2 = \frac{\sigma^2}{n}$$

Central Limit Theorem (CLT) - Formal Definition

Central Limit Theorem (CLT) - Formal Definition

As $n \to \infty$ $\quad \dfrac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]}{\sigma_X}\sqrt{n} \sim \mathcal{N}\left(0,1^2\right)$

Central Limit Theorem (CLT) - Formal Definition

As $n \to \infty$ $$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]}{\sigma_X}\sqrt{n} \sim \mathcal{N}\left(0,1^2\right)$$

As $n \to \infty$ $$\frac{1}{n}\left(\frac{\sum_{i=1}^{n} X_i - \frac{1}{n}n\mathbb{E}[X]}{\sigma_X}\right)\sqrt{n} \sim \mathcal{N}\left(0,1^2\right)$$

Central Limit Theorem (CLT) - Formal Definition

As $n \to \infty$ $$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]}{\sigma_X}\sqrt{n} \sim \mathcal{N}\left(0,1^2\right)$$

As $n \to \infty$ $$\frac{1}{n}\left(\frac{\sum_{i=1}^{n} X_i - \frac{1}{\cancel{n}}\cancel{n}\mathbb{E}[X]}{\sigma_X}\right)\sqrt{n} \sim \mathcal{N}\left(0,1^2\right)$$

Central Limit Theorem (CLT) - Formal Definition

As $n \to \infty$ $$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]}{\sigma_X}\sqrt{n} \sim \mathcal{N}\left(0,1^2\right)$$

As $n \to \infty$ $$\frac{1}{\underset{\sqrt{n}}{\cancel{n}}}\left(\frac{\sum_{i=1}^{n} X_i - \frac{1}{\cancel{n}}\cancel{n}\mathbb{E}[X]}{\sigma_X}\right)\cancel{\sqrt{n}} \sim \mathcal{N}\left(0,1^2\right)$$

Central Limit Theorem (CLT) - Formal Definition

As $n \to \infty$ $\quad \dfrac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]}{\sigma_X}\sqrt{n} \sim \mathcal{N}\left(0,1^2\right)$

As $n \to \infty$

Central Limit Theorem (CLT) - Formal Definition

As $n \to \infty$ $$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]}{\sigma_X}\sqrt{n} \sim \mathcal{N}\left(0,1^2\right)$$

As $n \to \infty$ $$\frac{\sum_{i=1}^{n} X_i - n\mathbb{E}[X]}{\sqrt{n}\sigma_X} \sim \mathcal{N}\left(0,1^2\right)$$

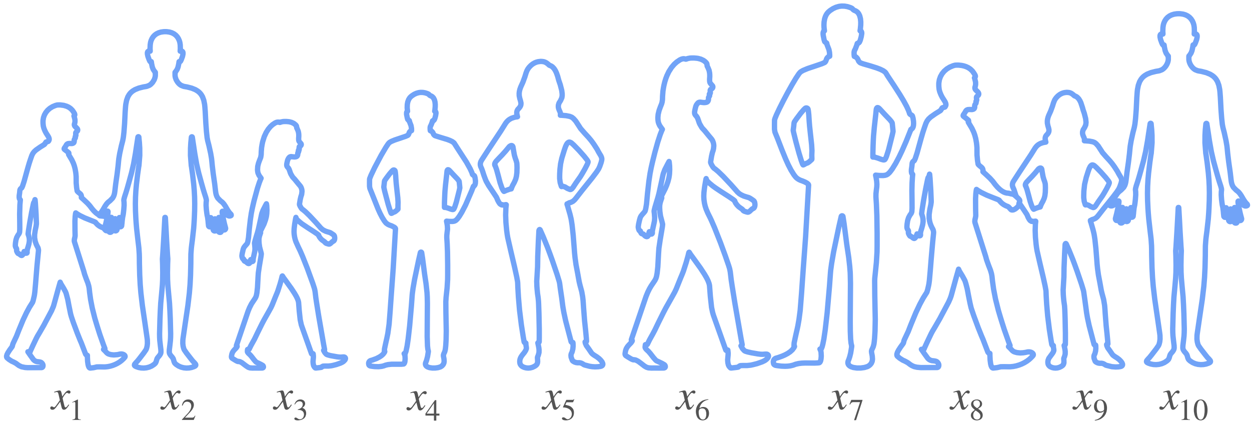
W3 Lesson 2

Point Estimation

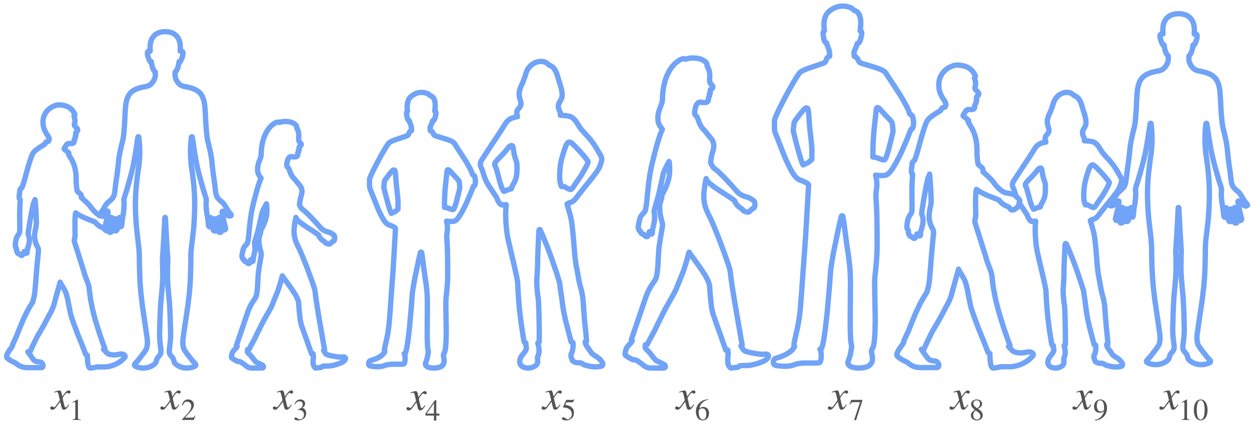
What is Point Estimation?

Point Estimates

Point Estimates



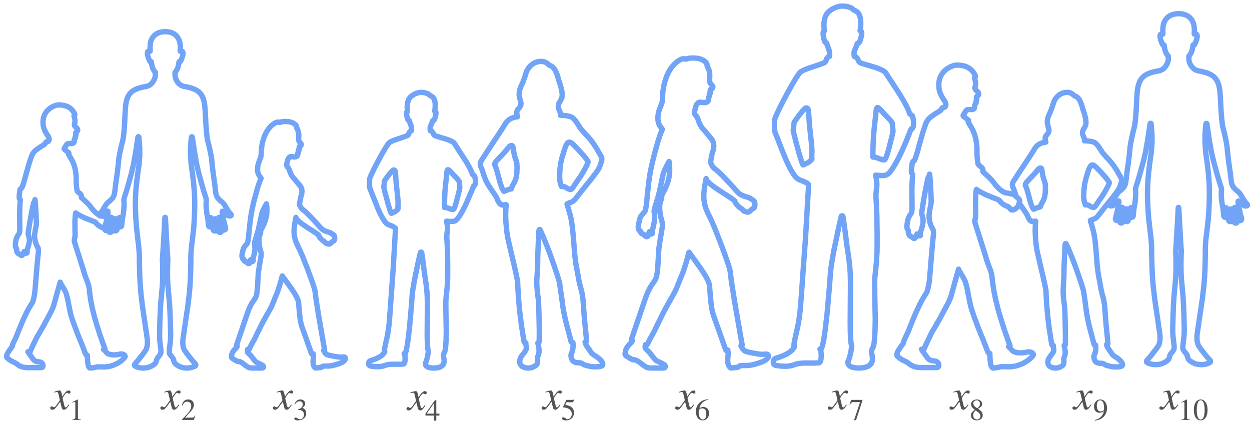
Point Estimates



Mean of the population?

$$\mu \approx \frac{1}{10}\sum_{i=1}^{10} x_i = \bar{x}$$

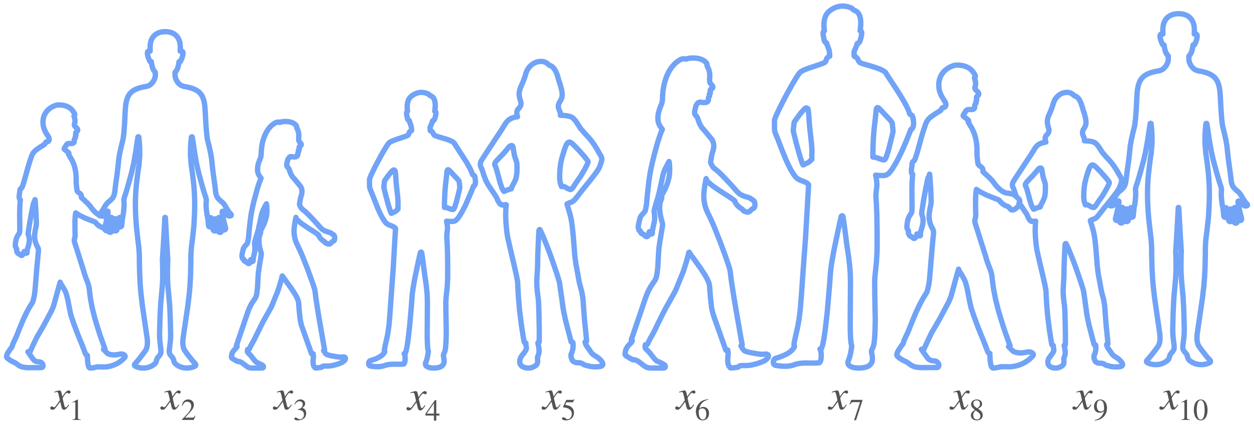
Point Estimates



Mean of the population? $\mu \approx \frac{1}{10}\sum_{i=1}^{10} x_i = \bar{x}$

Variance of the population? $\sigma^2 \approx \frac{1}{10-1}\sum_{i=1}^{10} \left(x_i - \bar{x}\right)^2 = s^2$

Point Estimates



Mean of the population? $\mu \approx \frac{1}{10}\sum_{i=1}^{10} x_i = \bar{x}$

Variance of the population? $\sigma^2 \approx \frac{1}{10-1}\sum_{i=1}^{10} (x_i - \bar{x})^2 = s^2$

$\bar{x}$ and s^2 are point estimates

Point Estimates

Point Estimates

A **point estimate** is a **single numerical value** based on **sample data** that is used to **approximate an unknown parameter** of a population or model parameter.

Point Estimates

Point Estimates

$\mathbf{P}(H)$?

Point Estimates

$\mathbf{P}(H)$?

Point Estimates

$\mathbf{P}(H)$?

10 throws
7 heads
3 tails

Point Estimates

$\mathbf{P}(H)?$

10 throws
7 heads
3 tails

$$\mathbf{P}(H) = \frac{7}{10}$$

Point Estimates

	x	y
0		
1		
2		
⋮		
50		

Point Estimates

	x	y
0		
1		
2		
⋮		
50		

Point Estimates

	x	y
0		
1		
2		
$\vdots$		
50		

$$y = \beta_0 + \beta_1 x$$

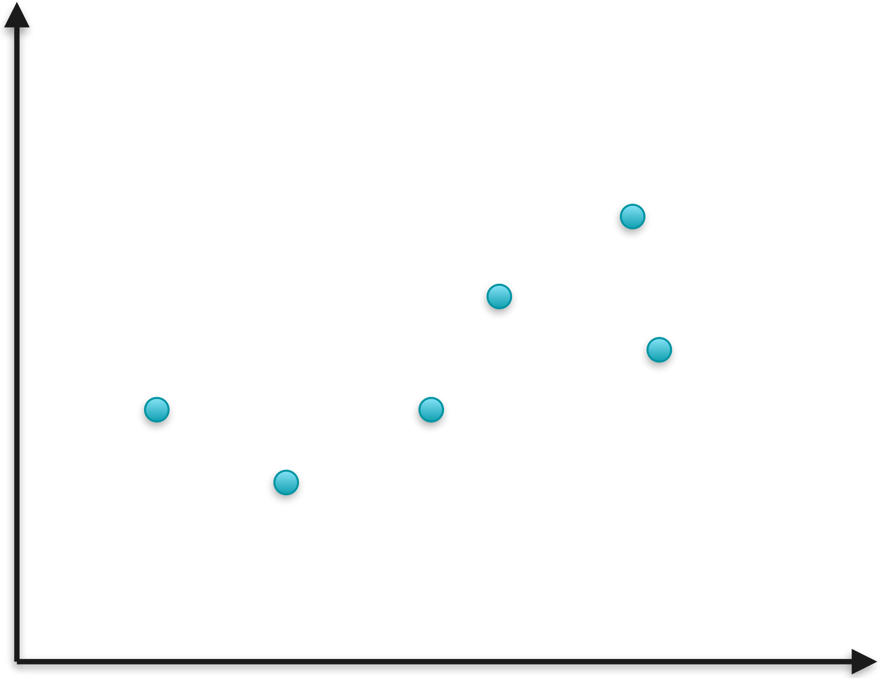
$$\beta_0, \beta_1 = ??$$

Point Estimates

	x	y
0		
1		
2		
⋮		
50		

$$y = \beta_0 + \beta_1 x$$

$$\beta_0, \beta_1 = ??$$

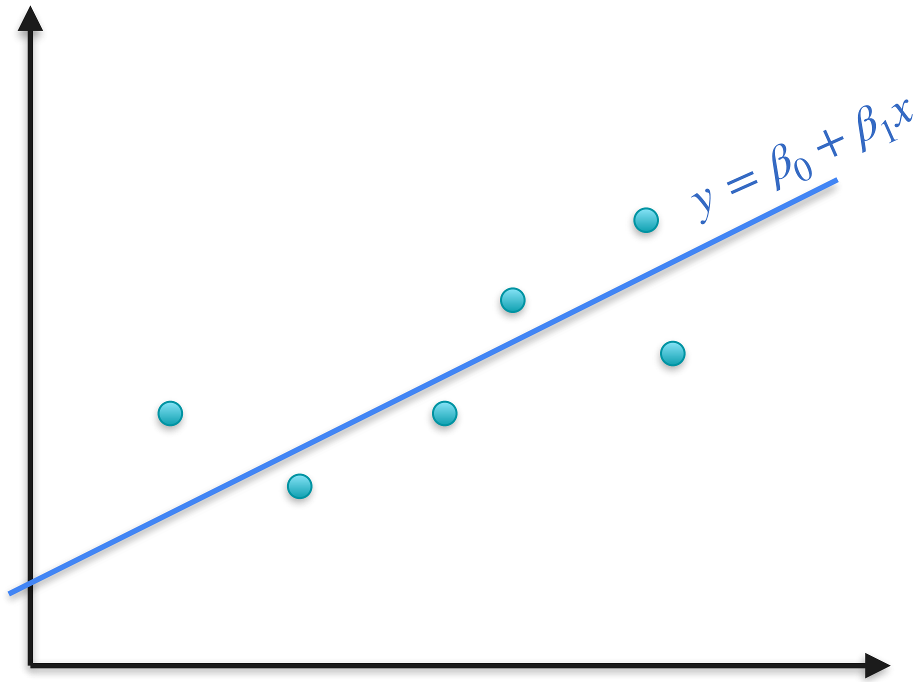


Point Estimates

	x	y
0		
1		
2		
$\vdots$		
50		

$$y = \beta_0 + \beta_1 x$$

$$\beta_0, \beta_1 = ??$$



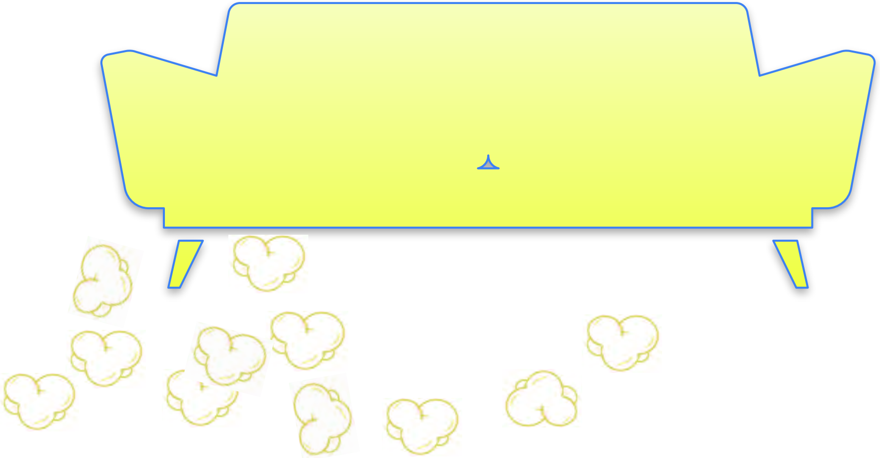
DeepLearning.AI

Point Estimation

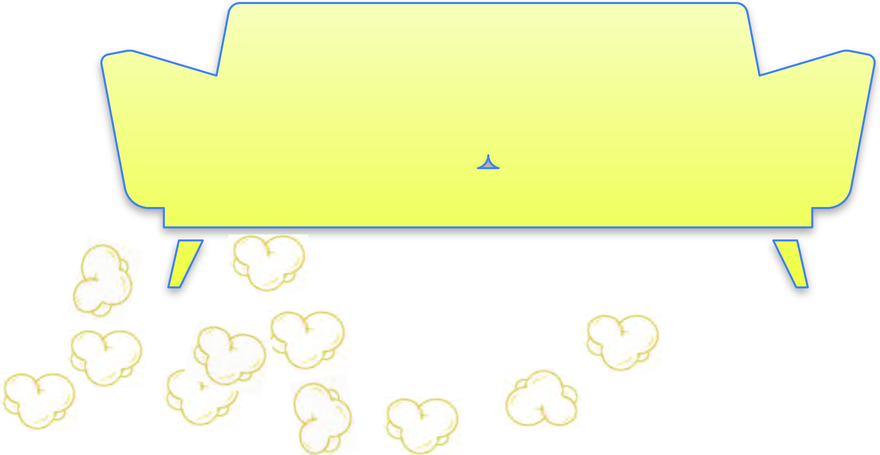
Maximum Likelihood Estimation: Motivation

There's Popcorn on the Floor. What Happened?

There's Popcorn on the Floor. What Happened?

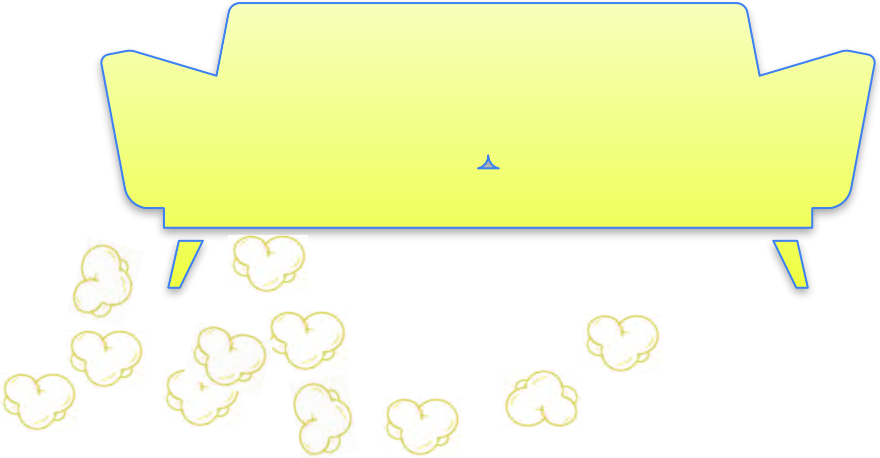


There’s Popcorn on the Floor. What Happened?

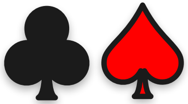


Movies

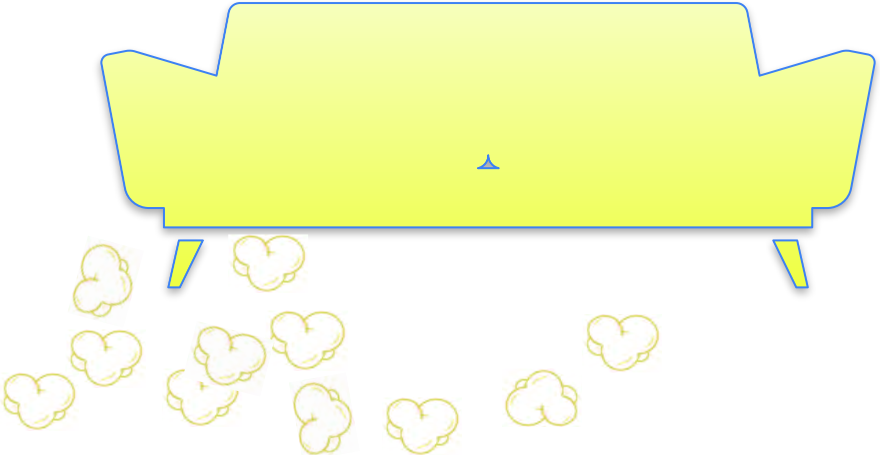
There's Popcorn on the Floor. What Happened?



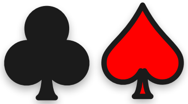
Movies 

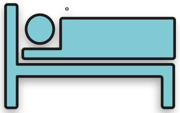
Board Games 

There's Popcorn on the Floor. What Happened?



Movies 

Board Games 

Nap 

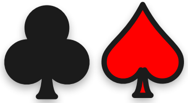
Quiz

What do you think happened?

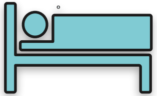
A. People were watching a movie
B. People were playing boardgames
C. People were taking a nap

There’s Popcorn on the Floor. What Happened?

Movies

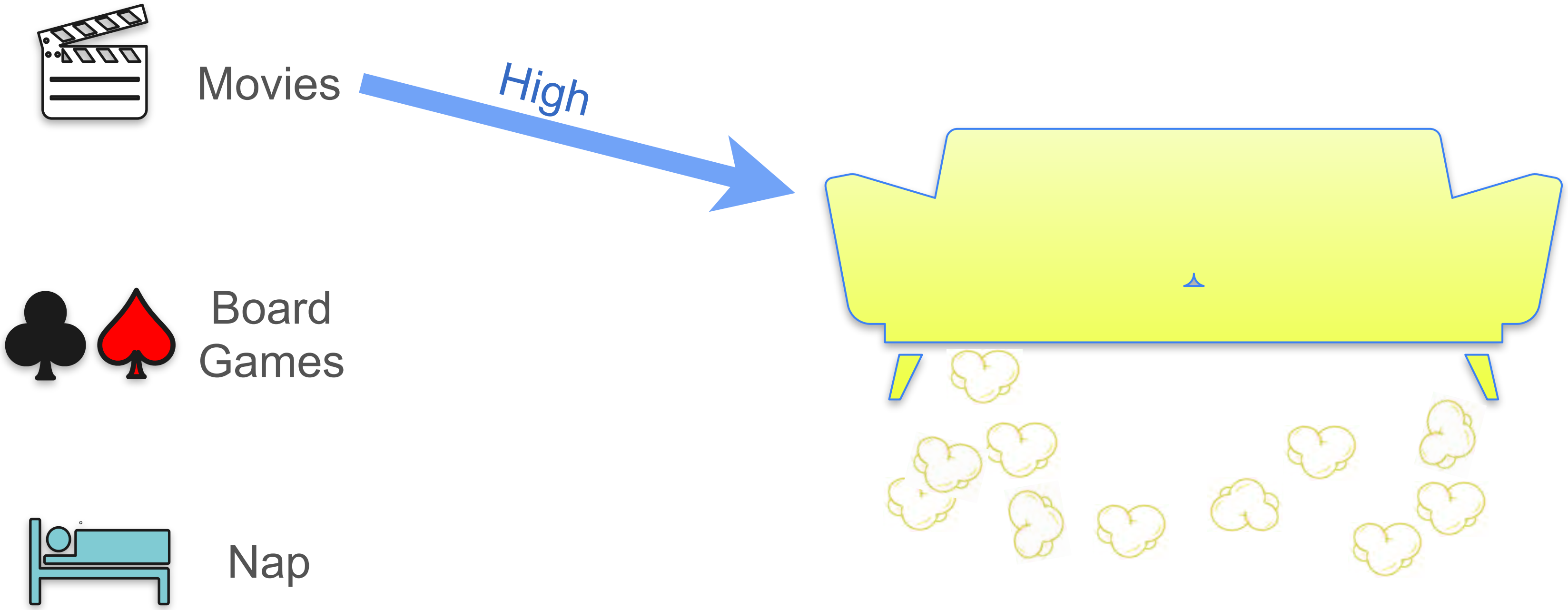


Board Games

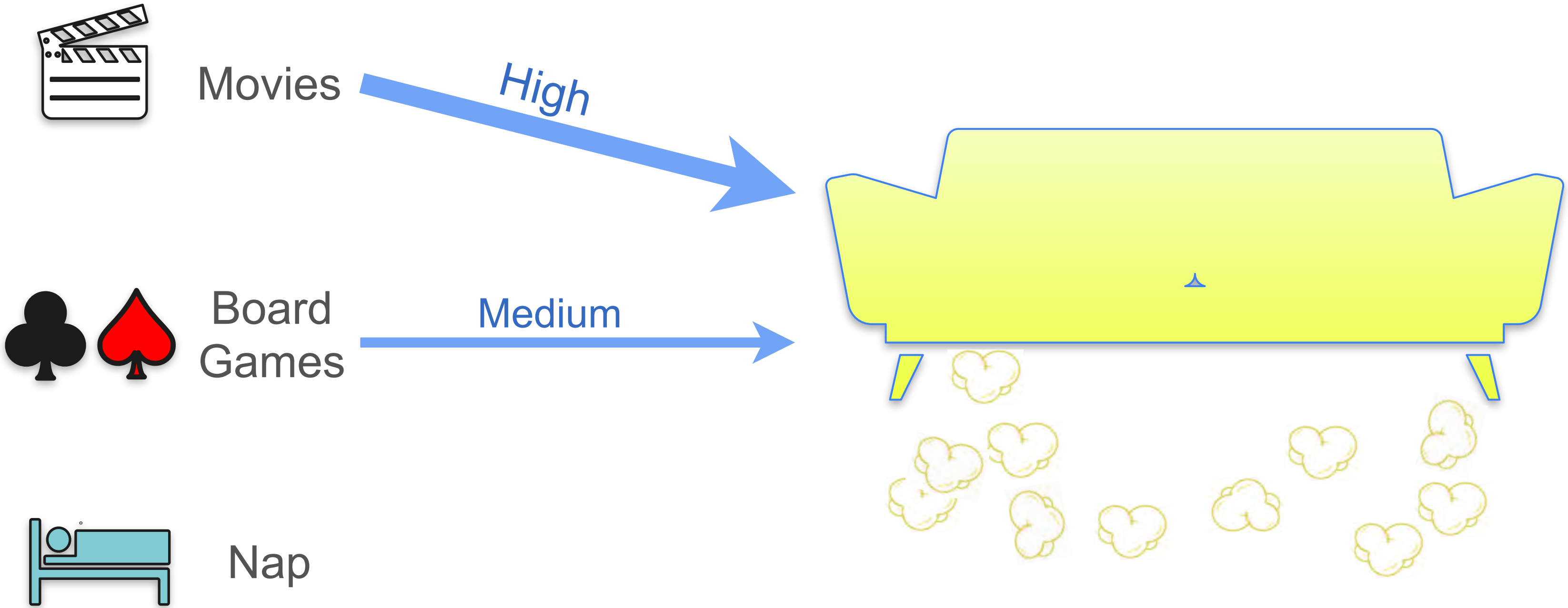


Nap

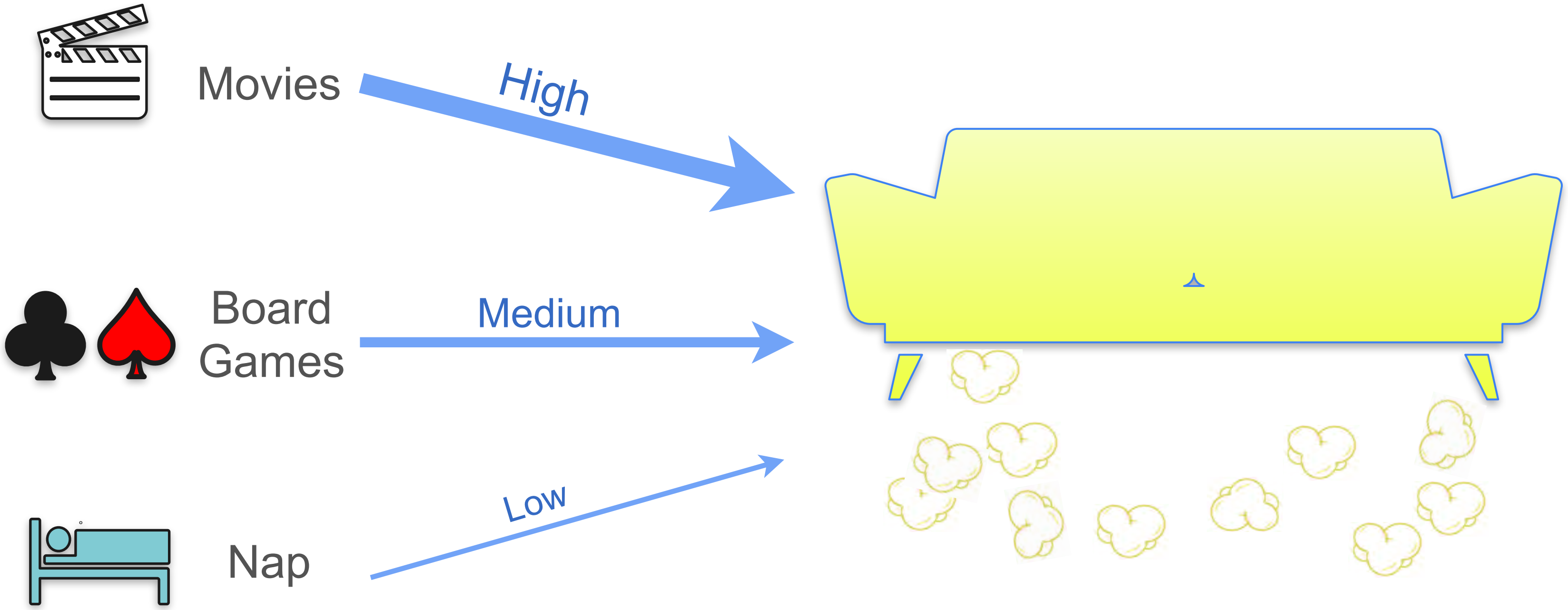
There’s Popcorn on the Floor. What Happened?



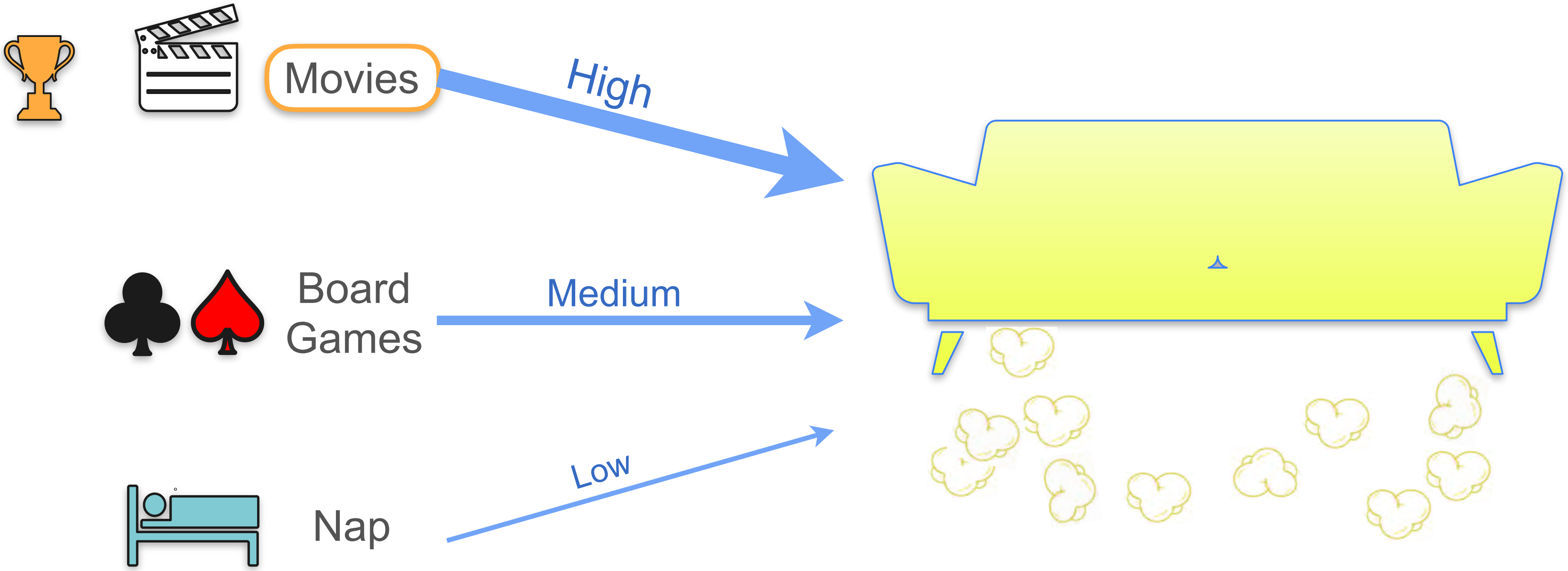
There's Popcorn on the Floor. What Happened?



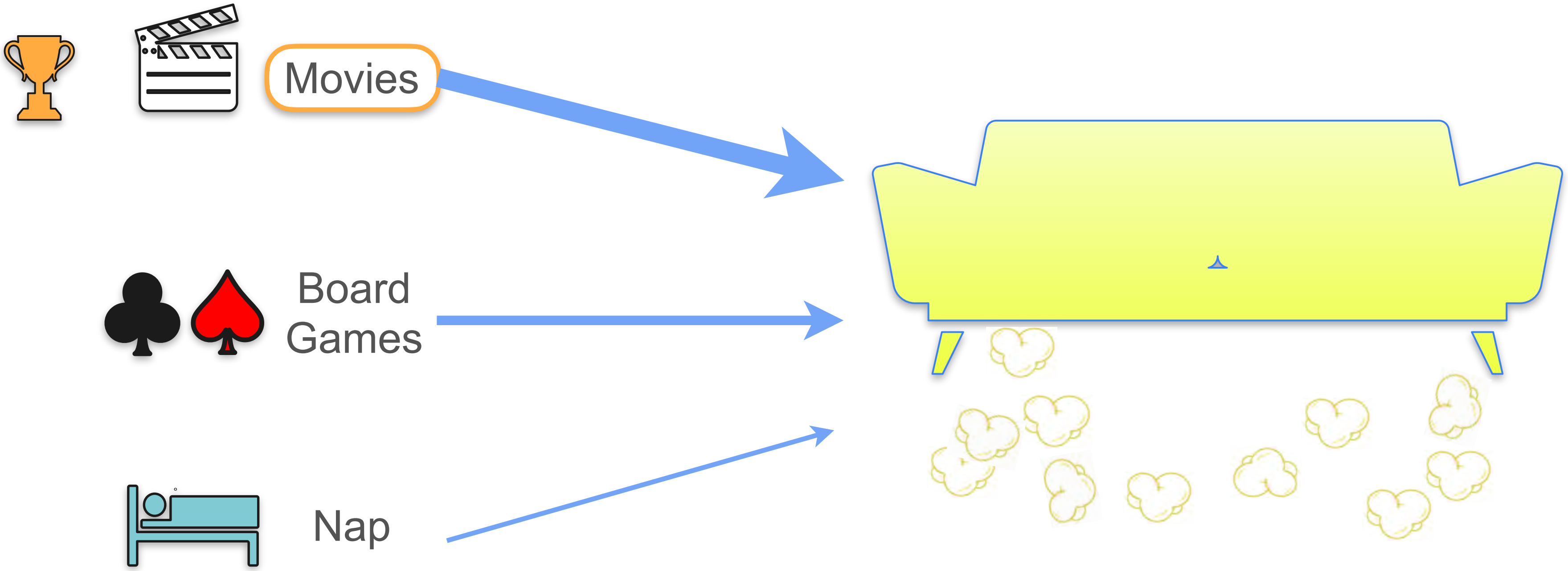
There's Popcorn on the Floor. What Happened?



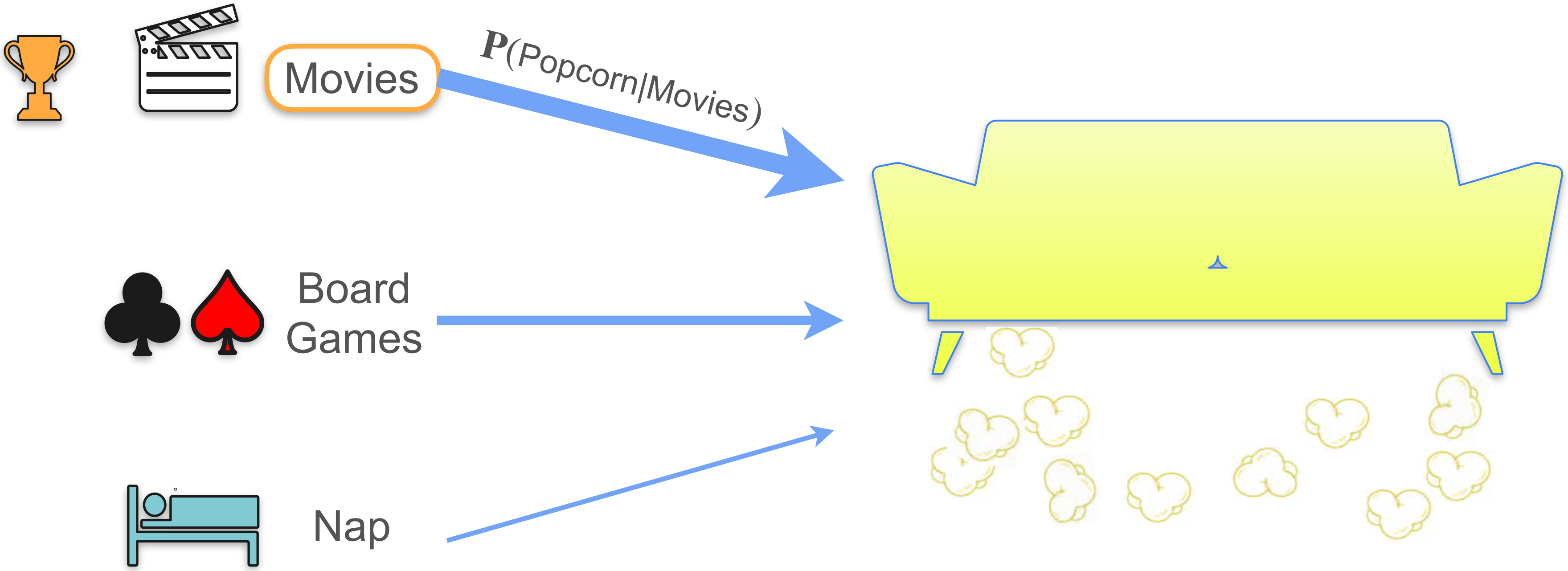
There's Popcorn on the Floor. What Happened?



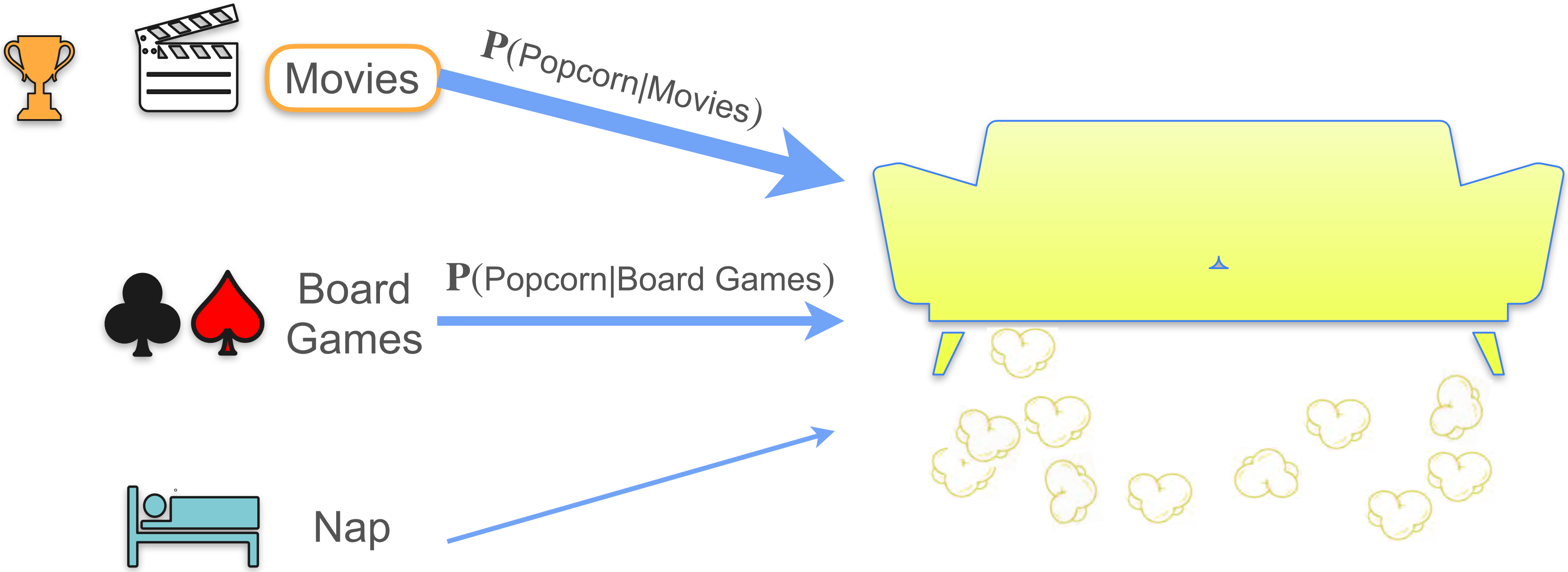
There's Popcorn on the Floor. What Happened?



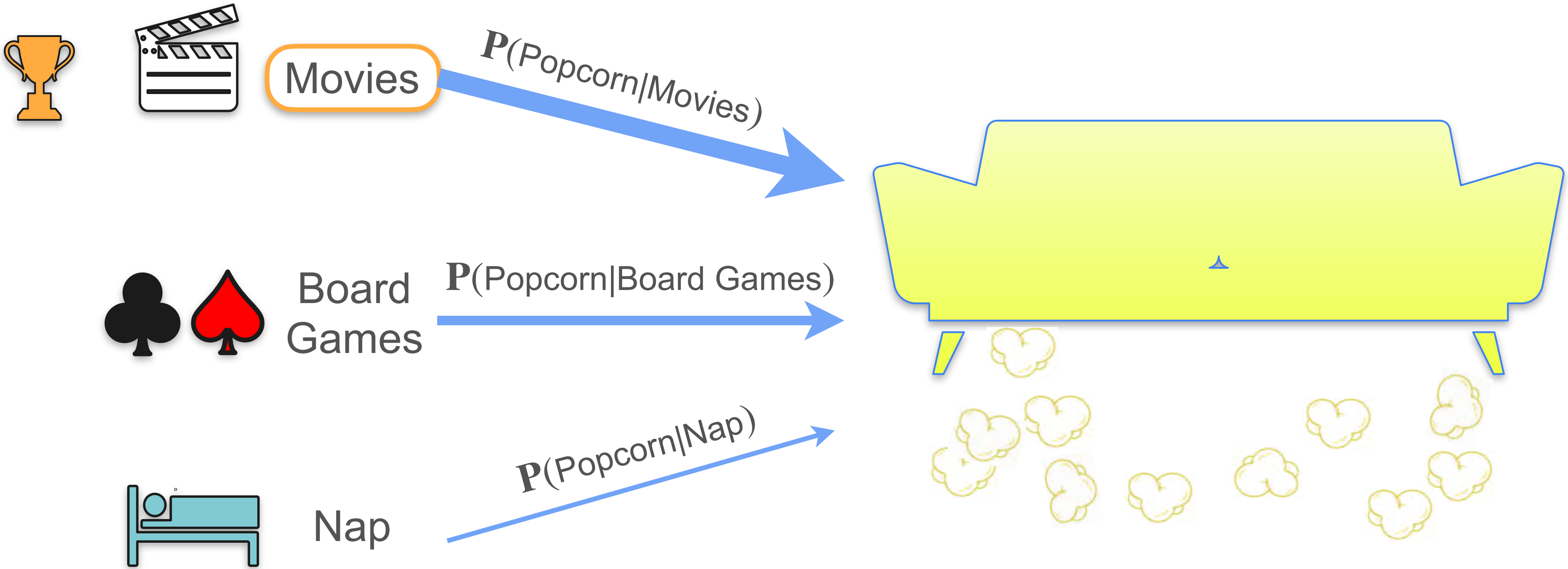
There's Popcorn on the Floor. What Happened?



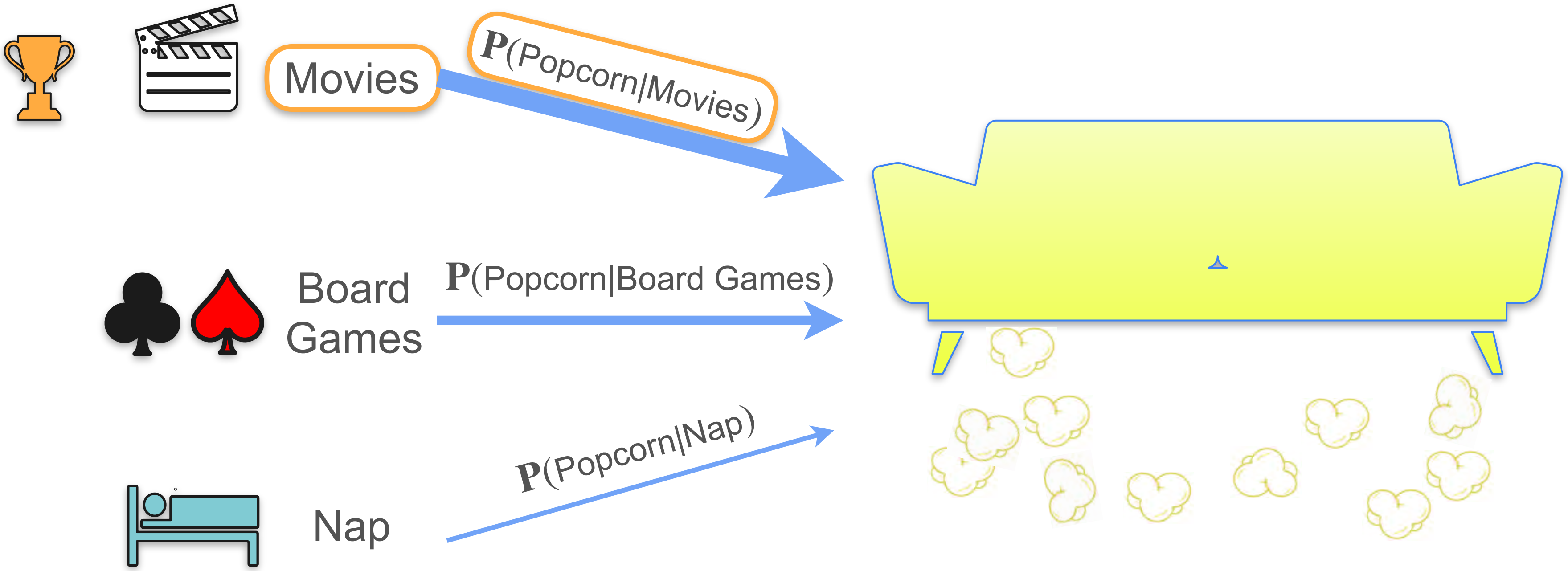
There’s Popcorn on the Floor. What Happened?



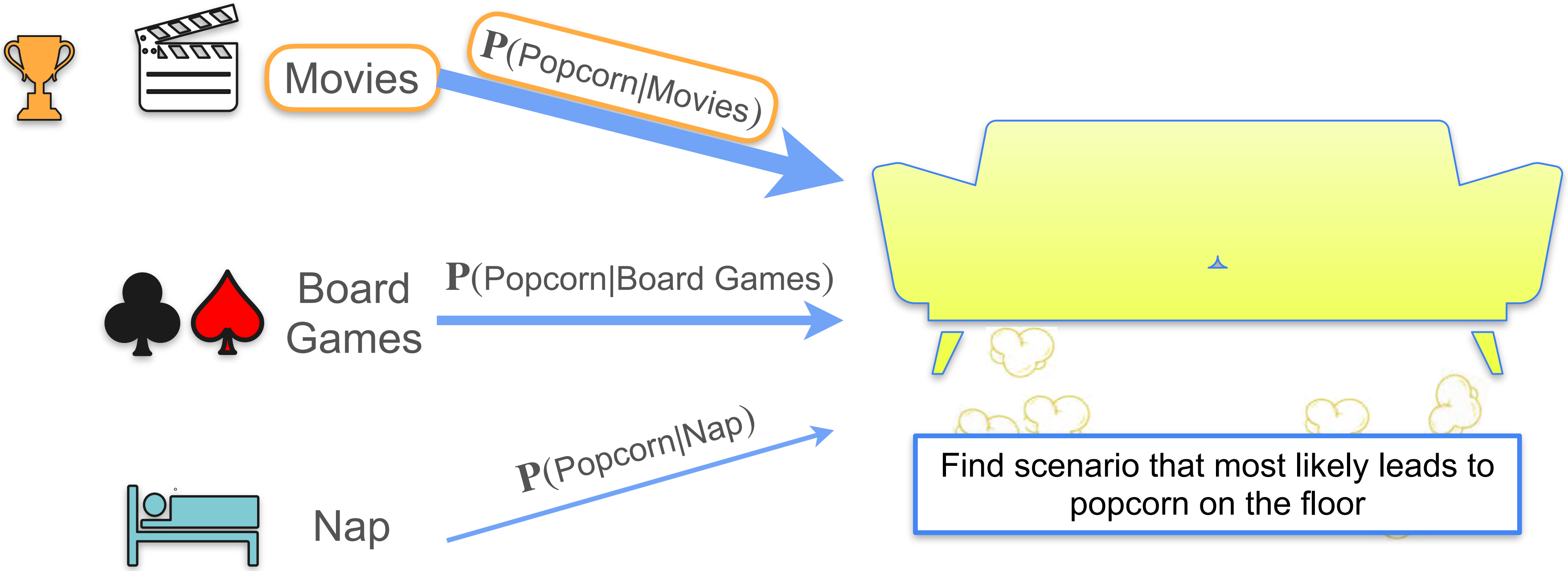
There’s Popcorn on the Floor. What Happened?



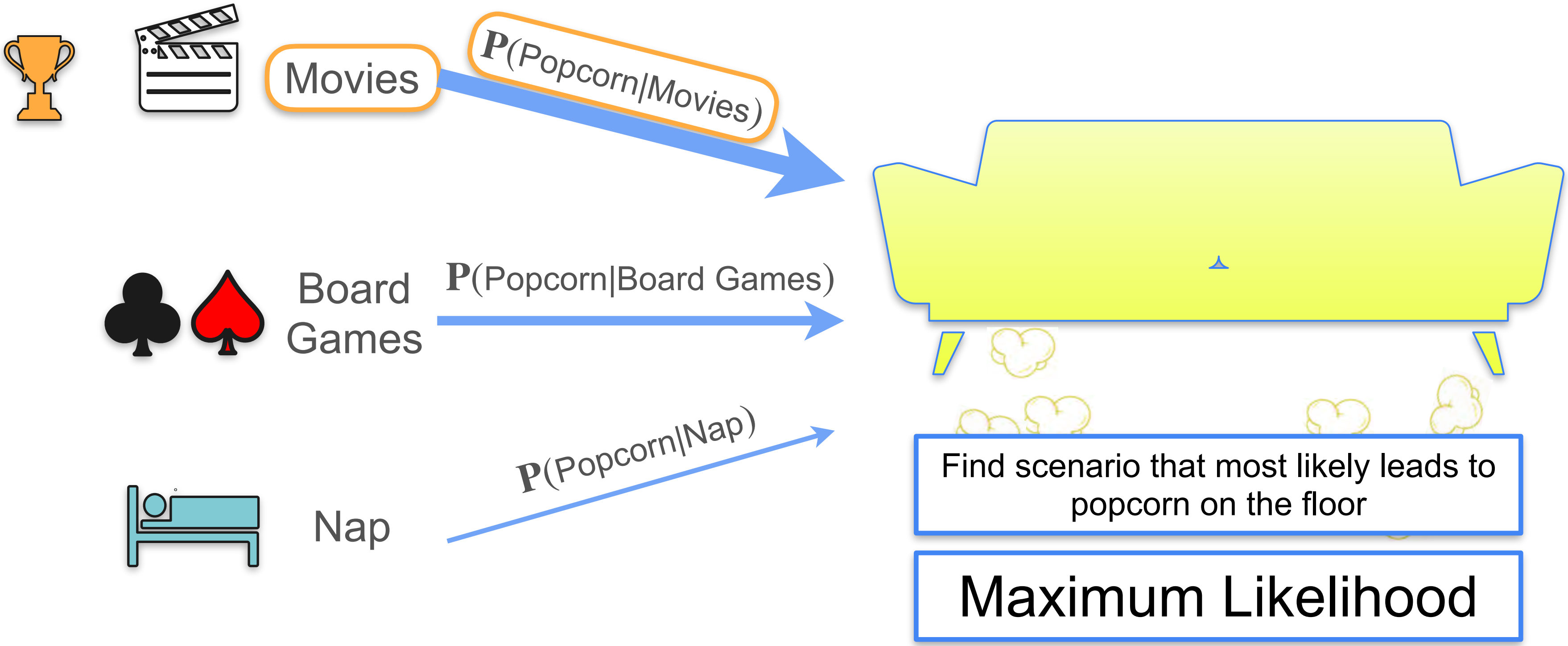
There’s Popcorn on the Floor. What Happened?



There’s Popcorn on the Floor. What Happened?



There’s Popcorn on the Floor. What Happened?



Maximum Likelihood

Maximum Likelihood

Data

Maximum Likelihood

Model 1

Data

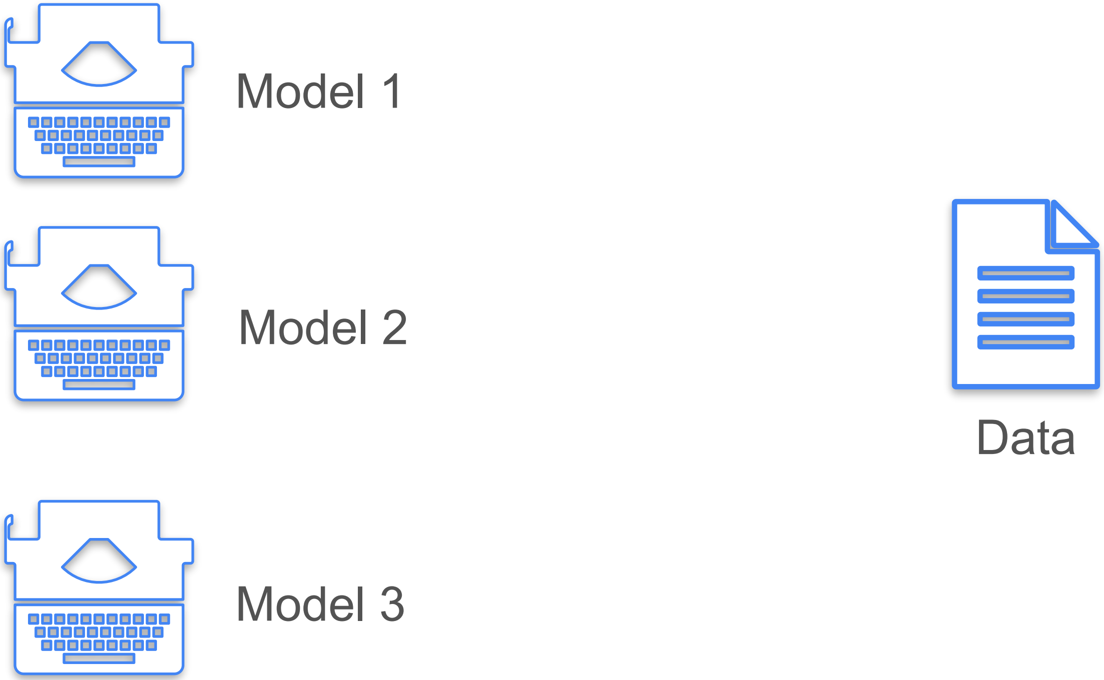
Maximum Likelihood

Model 1

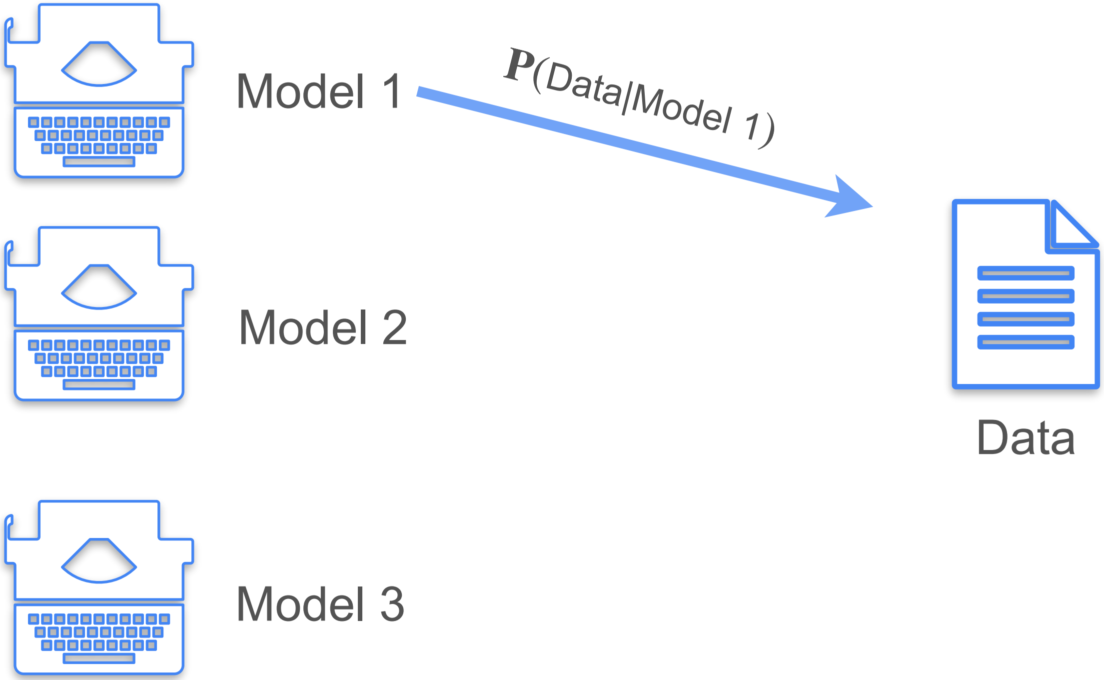
Model 2

Data

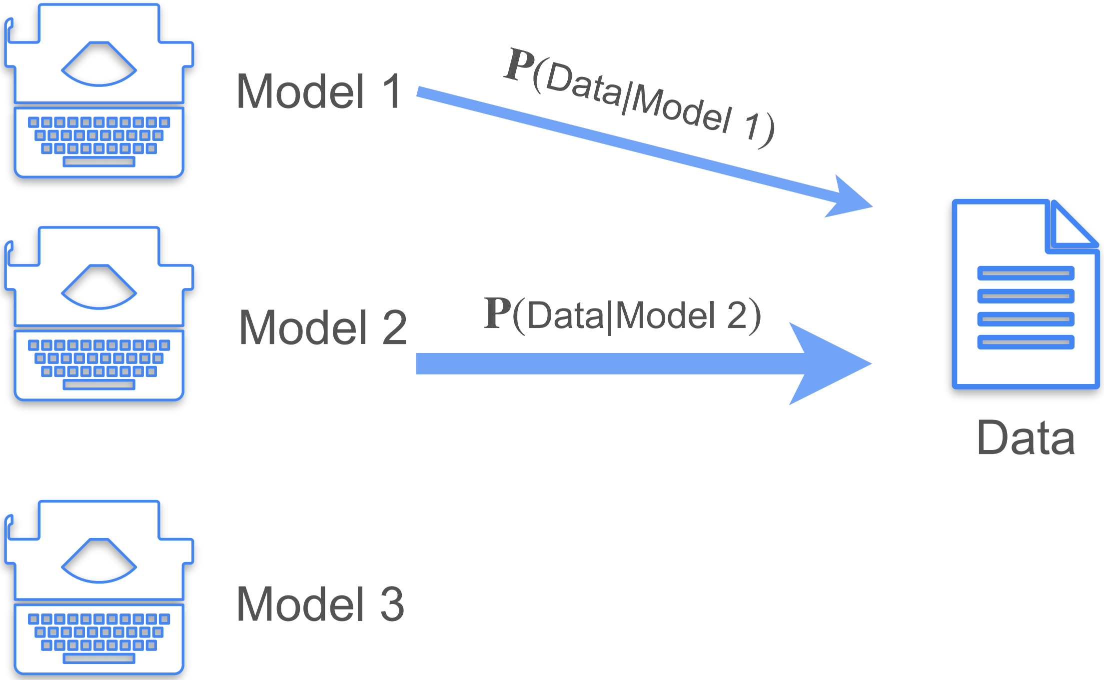
Maximum Likelihood



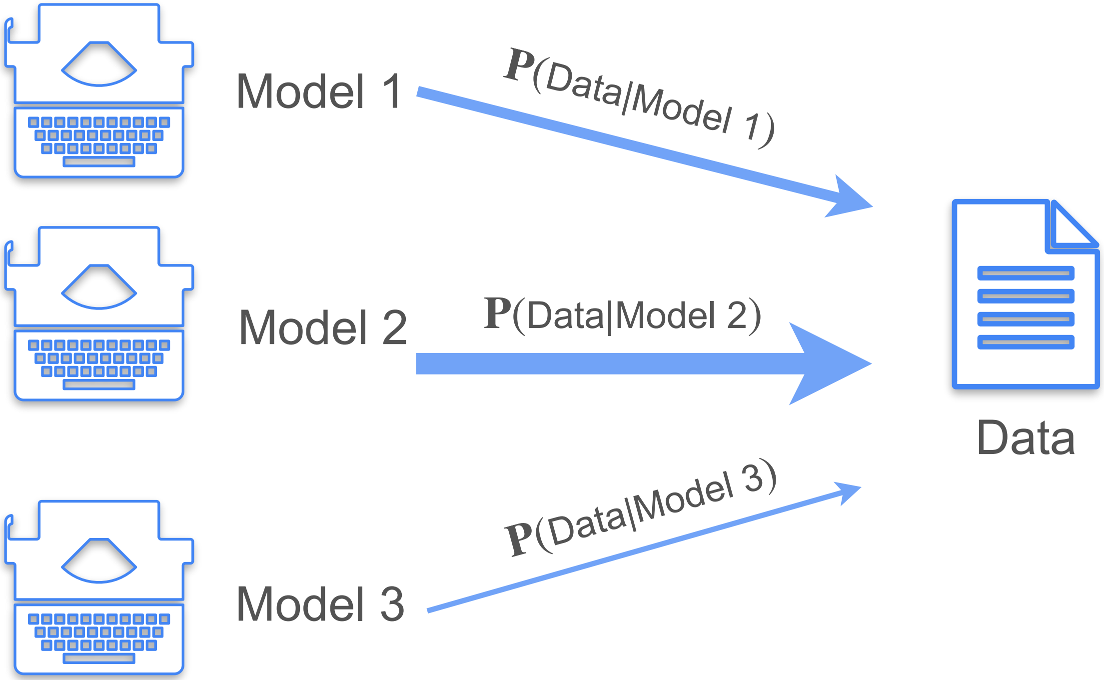
Maximum Likelihood



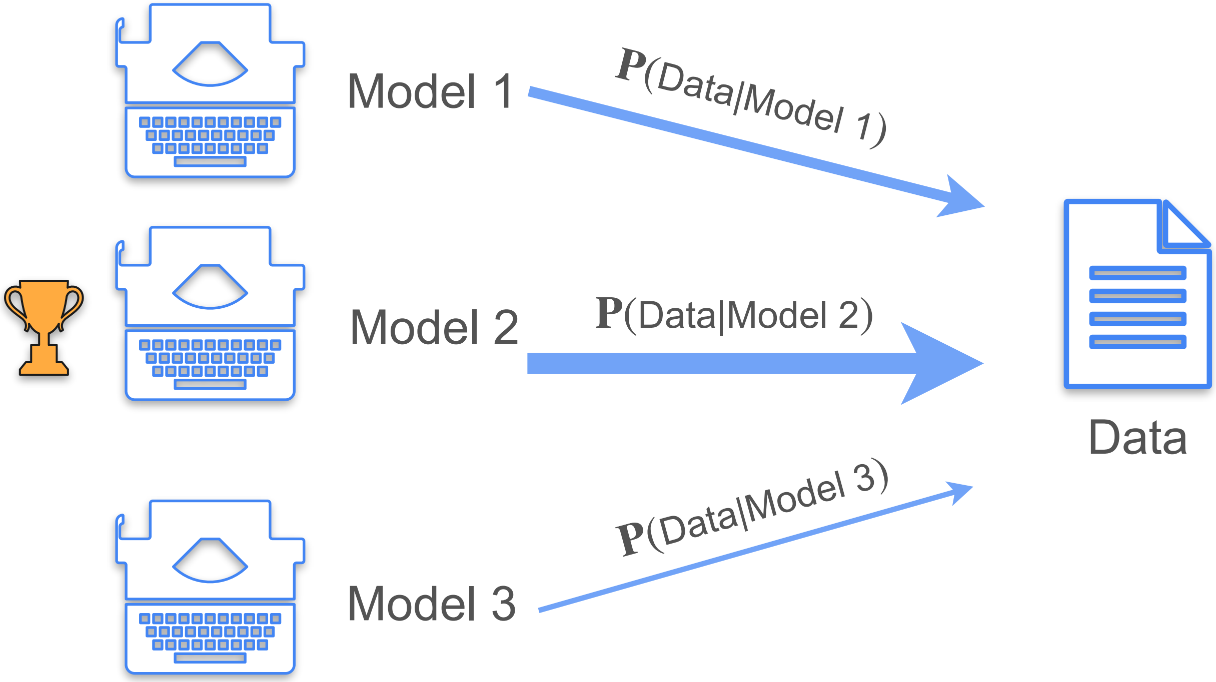
Maximum Likelihood



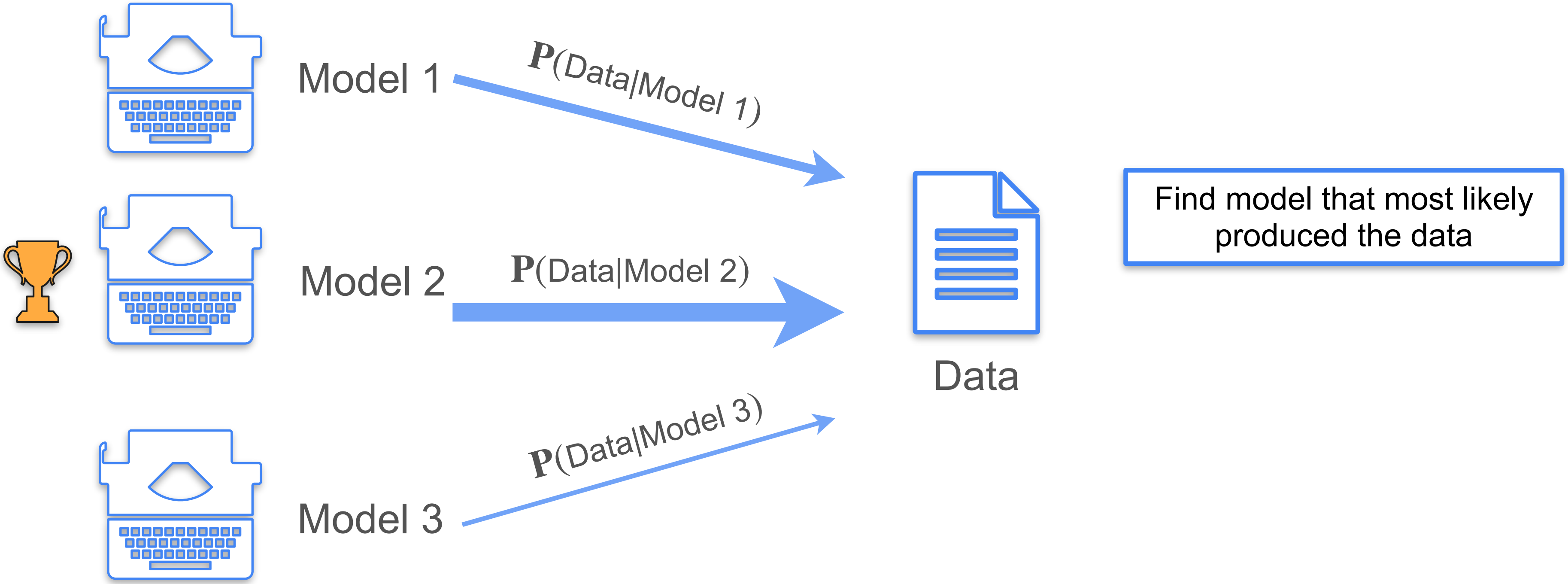
Maximum Likelihood



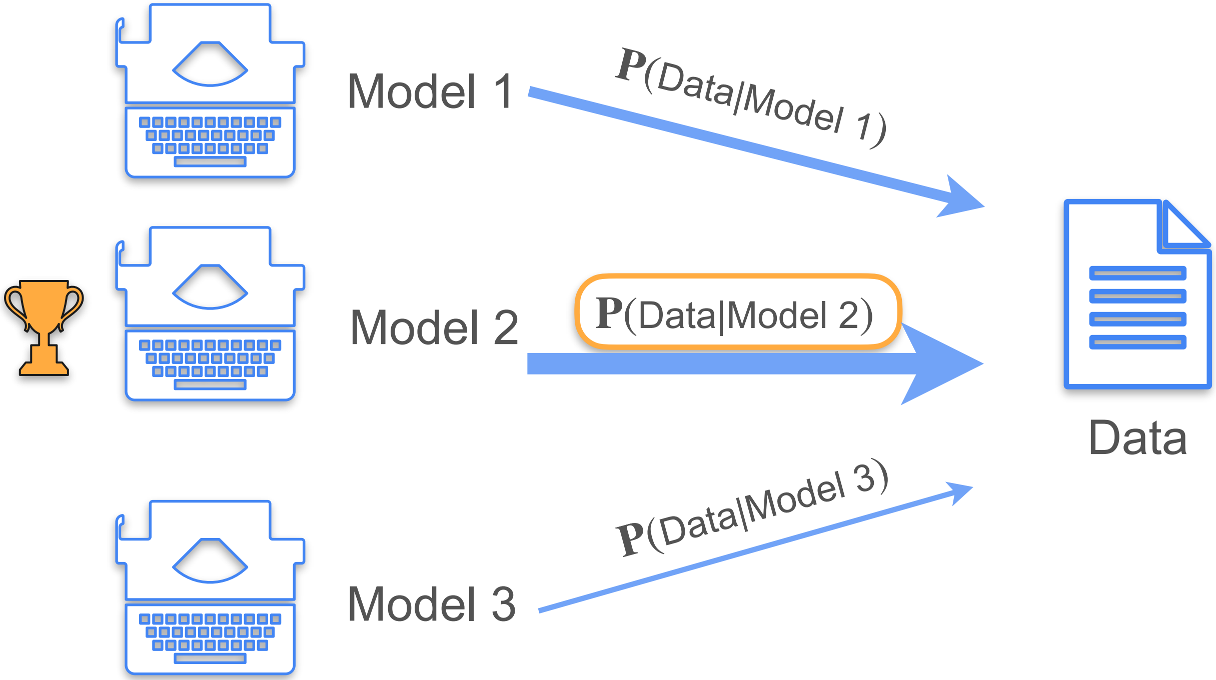
Maximum Likelihood



Maximum Likelihood

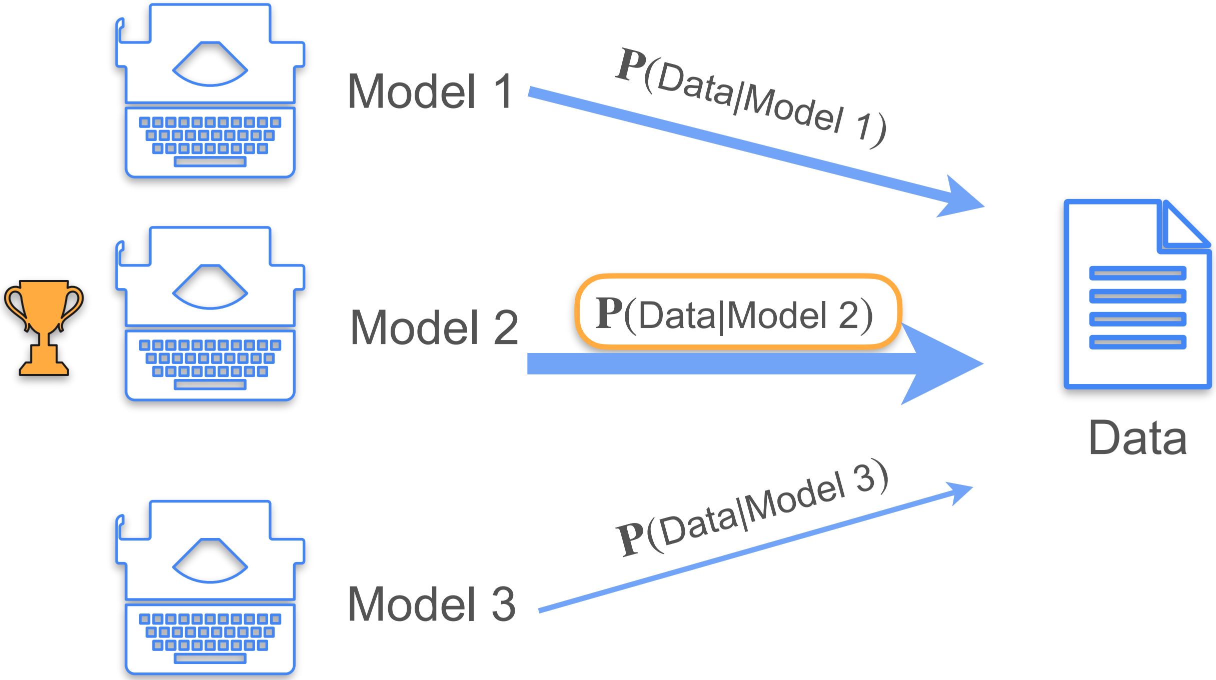


Maximum Likelihood



Find model that most likely produced the data

Maximum Likelihood

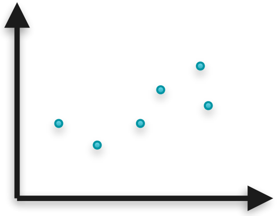


Find model that most likely produced the data

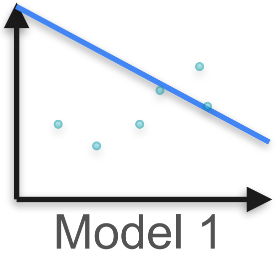
Maximize
$\mathbf{P}(\text{Data}|\text{Model})$

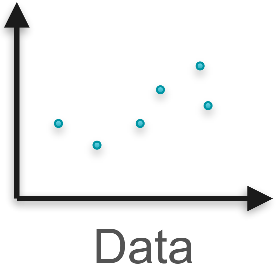
Example: Linear Regression

Example: Linear Regression

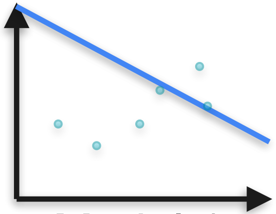

Data

Example: Linear Regression

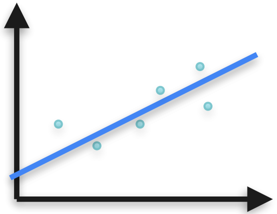




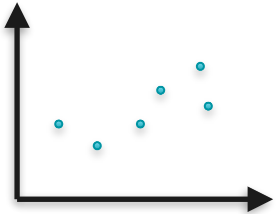
Example: Linear Regression



Model 1

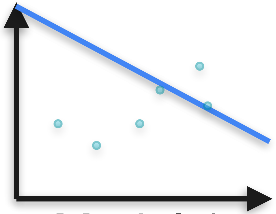


Model 2

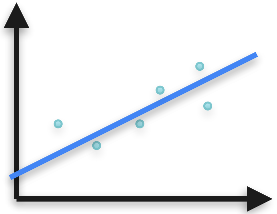


Data

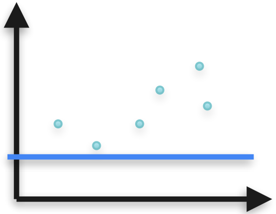
Example: Linear Regression



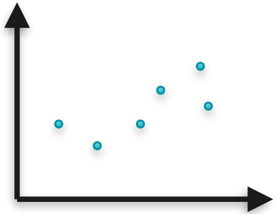
Model 1



Model 2

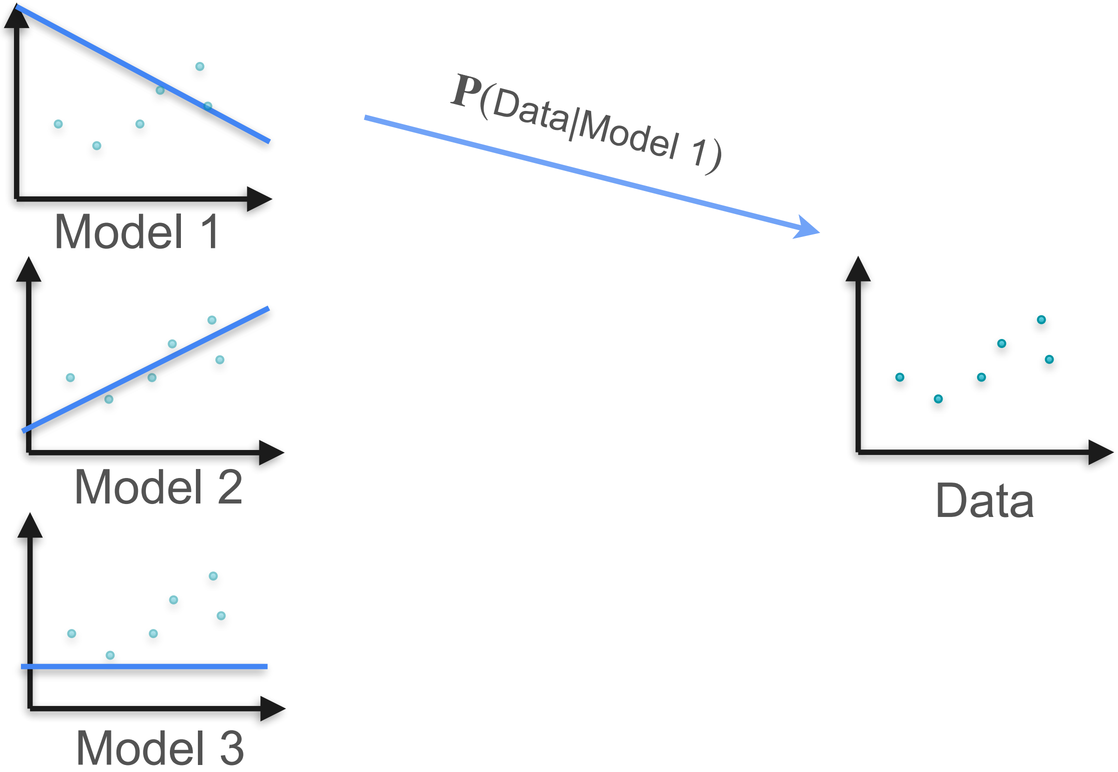


Model 3

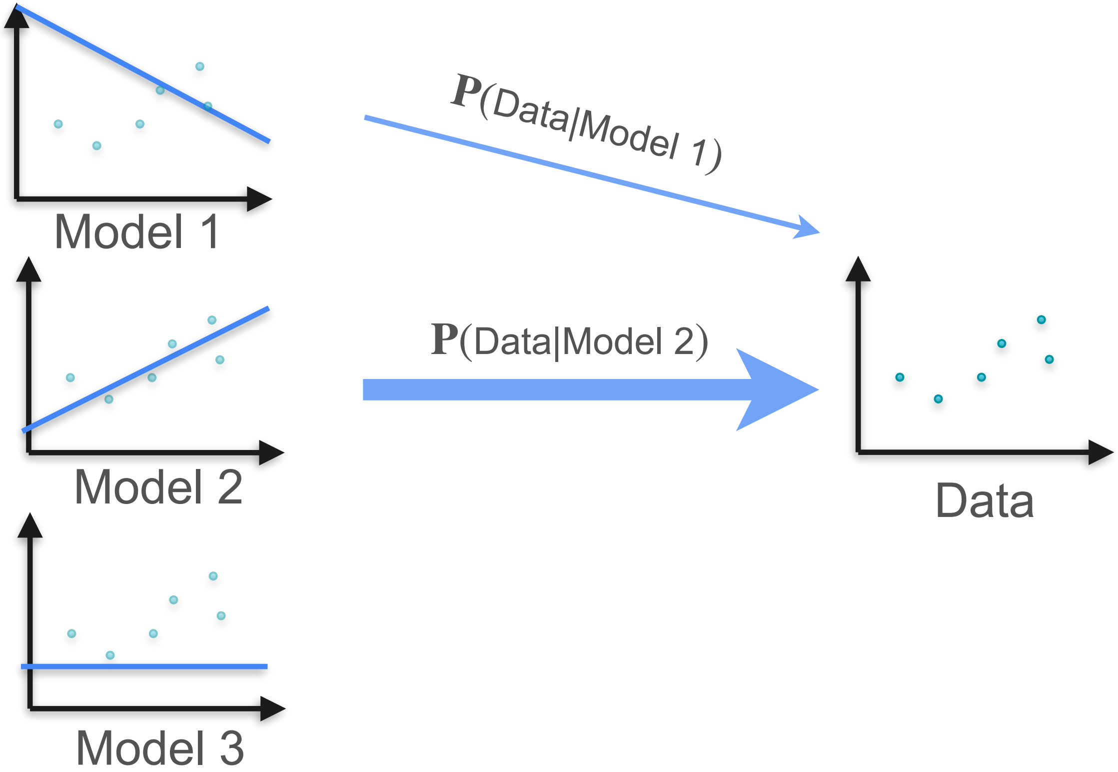


Data

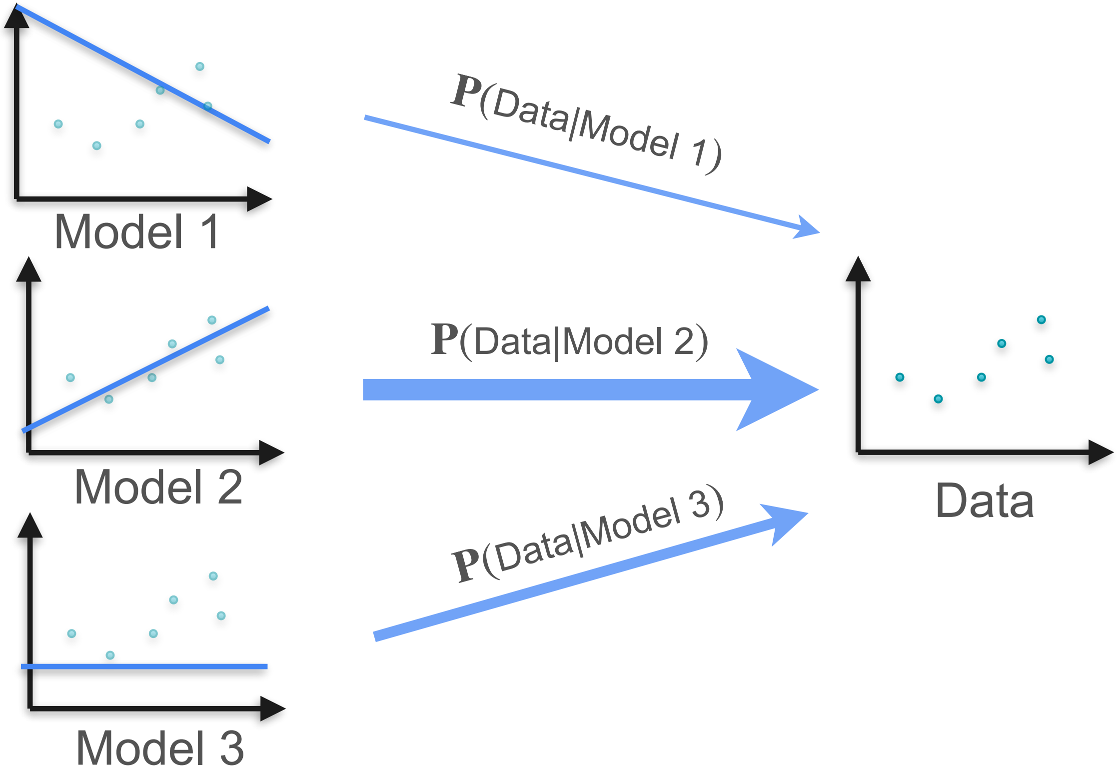
Example: Linear Regression



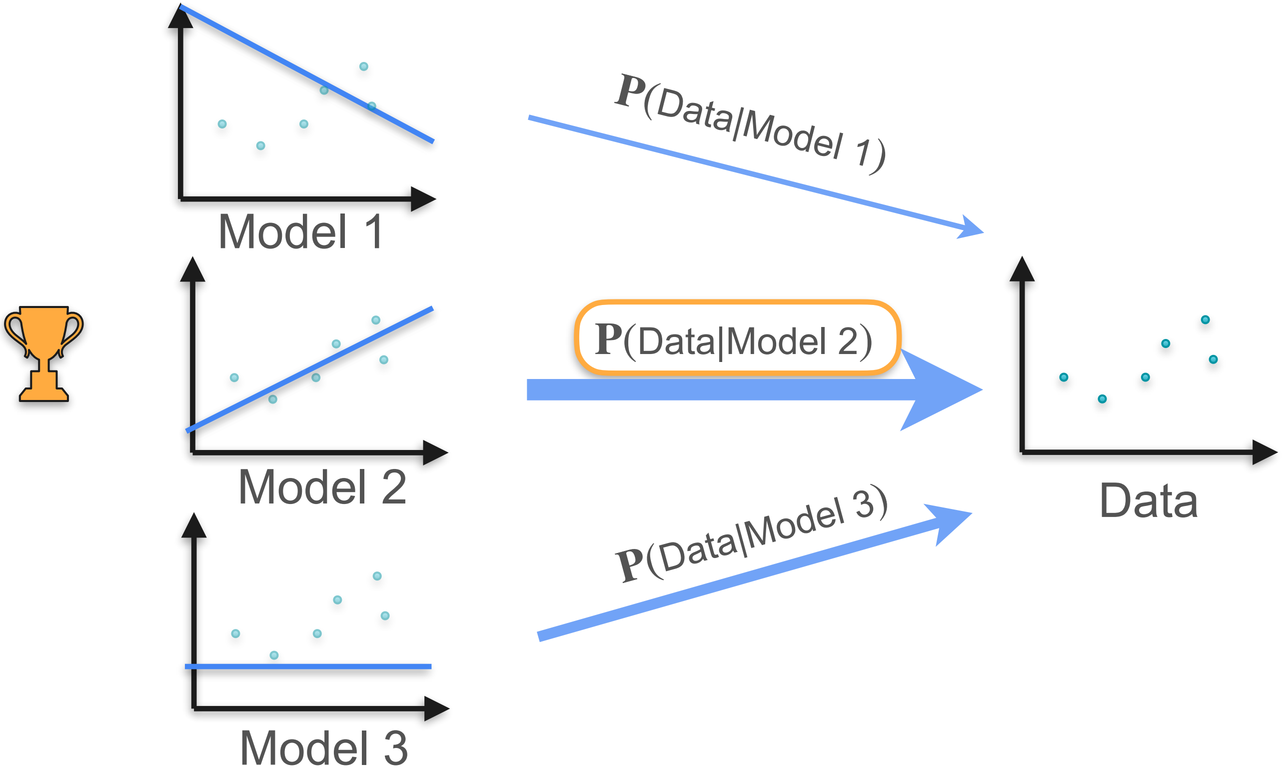
Example: Linear Regression



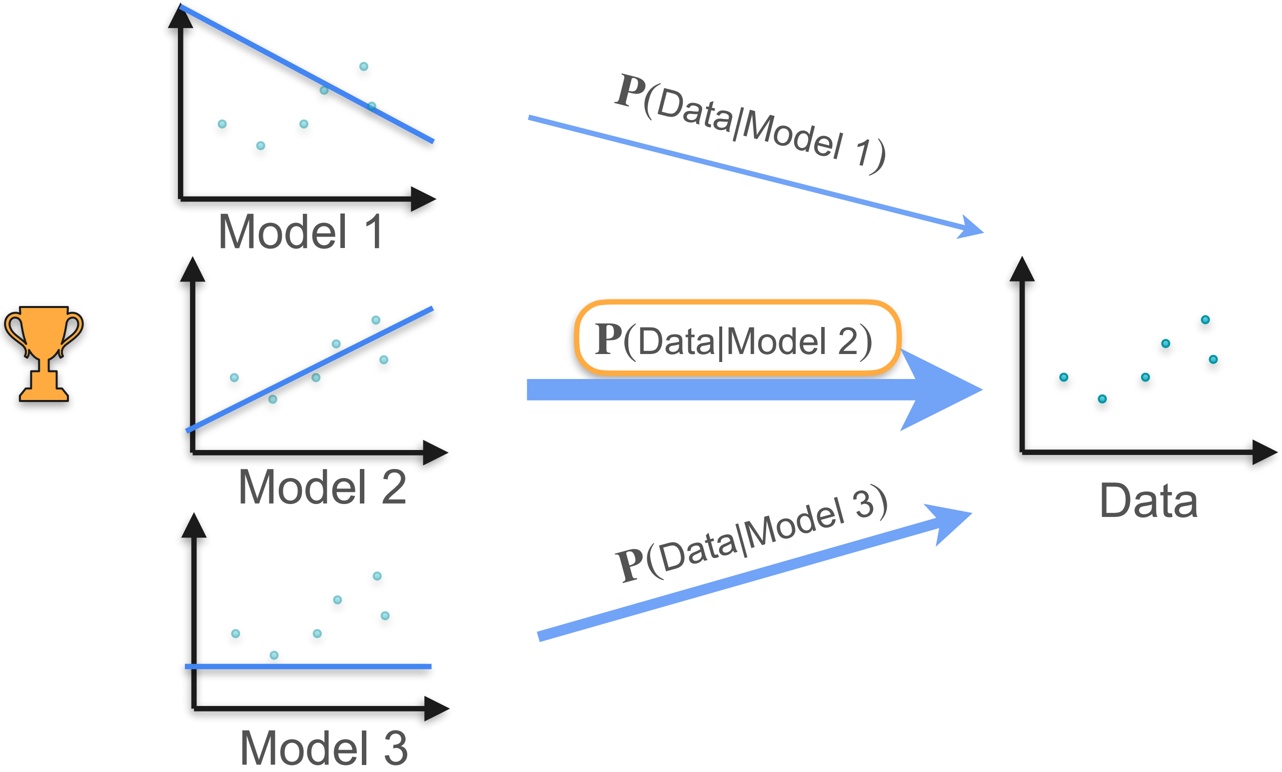
Example: Linear Regression



Example: Linear Regression



Example: Linear Regression



Find model that most likely produced the data

DeepLearning.AI

Point Estimation

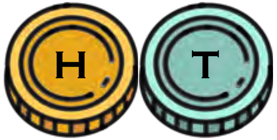
MLE: Bernoulli Example

Maximum Likelihood: Bernoulli Example

Maximum Likelihood: Bernoulli Example

Maximum Likelihood: Bernoulli Example

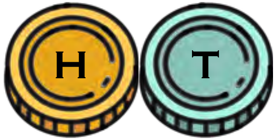
Coin 1



$\mathbf{P}(H) = 0.7$

$\mathbf{P}(T) = 0.3$

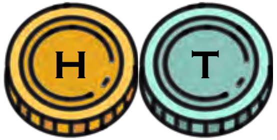
Coin 2



$\mathbf{P}(H) = 0.5$

$\mathbf{P}(T) = 0.5$

Coin 3



$\mathbf{P}(H) = 0.3$

$\mathbf{P}(T) = 0.7$

Quiz

Which of the coins is more likely?

A. Coin 1 ($P(H) = 0.7$)
B. Coin 2 ($P(H) = 0.5$)
C. Coin 3 ($P(H) = 0.3$)

Maximum Likelihood: Bernoulli Example

Maximum Likelihood: Bernoulli Example

Coin 1 0.7 0.7 0.7 0.7 0.7 0.7 0.7 0.7 0.3 0.3 = 0.0051

Maximum Likelihood: Bernoulli Example

Coin 1	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.3	0.3 = 0.0051
Coin 2	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5 = 0.0010

Maximum Likelihood: Bernoulli Example

Coin 1	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.3	0.3 = 0.0051
Coin 2	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5 = 0.0010
Coin 3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.30	0.7	0.7 = 0.00003

Maximum Likelihood: Bernoulli Example

Coin 1	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.3	0.3 = 0.0051
Coin 2	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5 = 0.0010
Coin 3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.30	0.7	0.7 = 0.00003

Maximum Likelihood: Bernoulli Example

Maximum Likelihood: Bernoulli Example

Maximum Likelihood: Bernoulli Example

Coin 1

$\mathbf{P}(H) = 0.7$

8H 2T

Maximum Likelihood: Bernoulli Example

Coin 1
$\mathbf{P}(H) = 0.7$

Coin 2
$\mathbf{P}(H) = 0.5$

8H 2T

Maximum Likelihood: Bernoulli Example

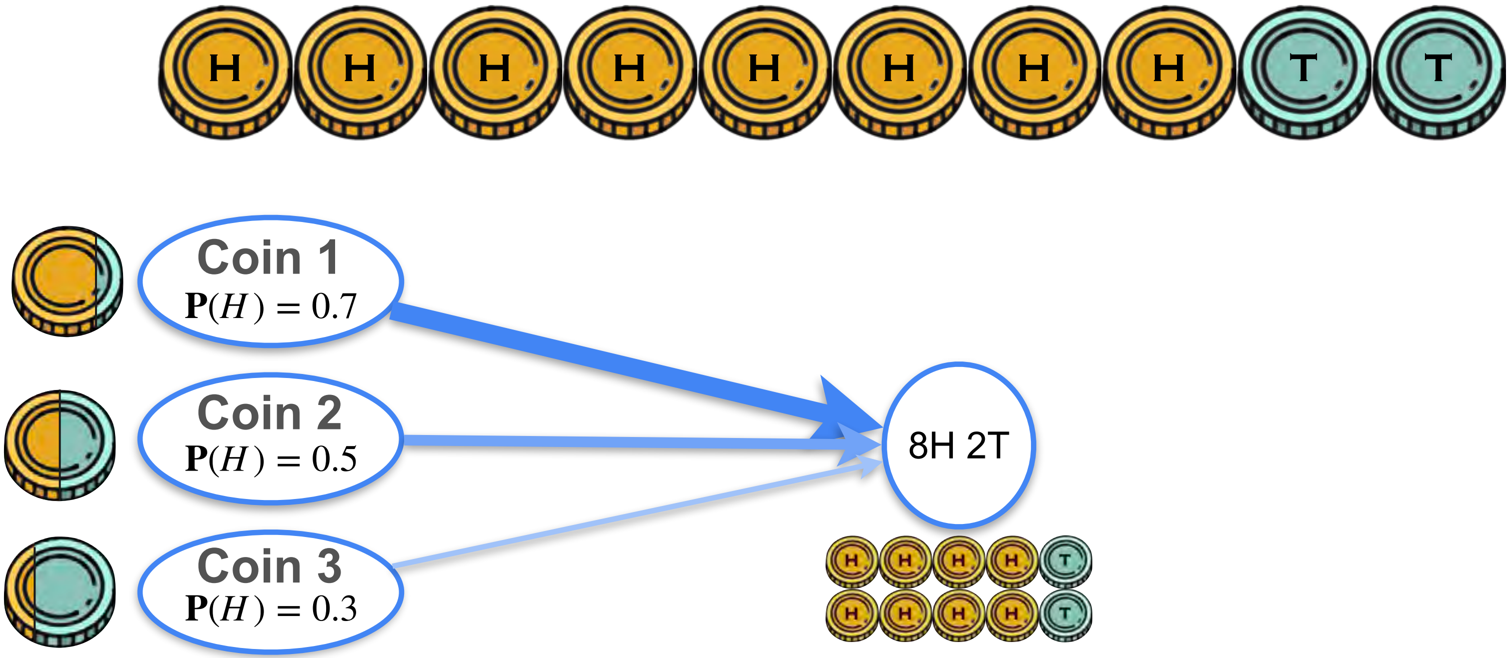
Coin 1
$\mathbf{P}(H) = 0.7$

Coin 2
$\mathbf{P}(H) = 0.5$

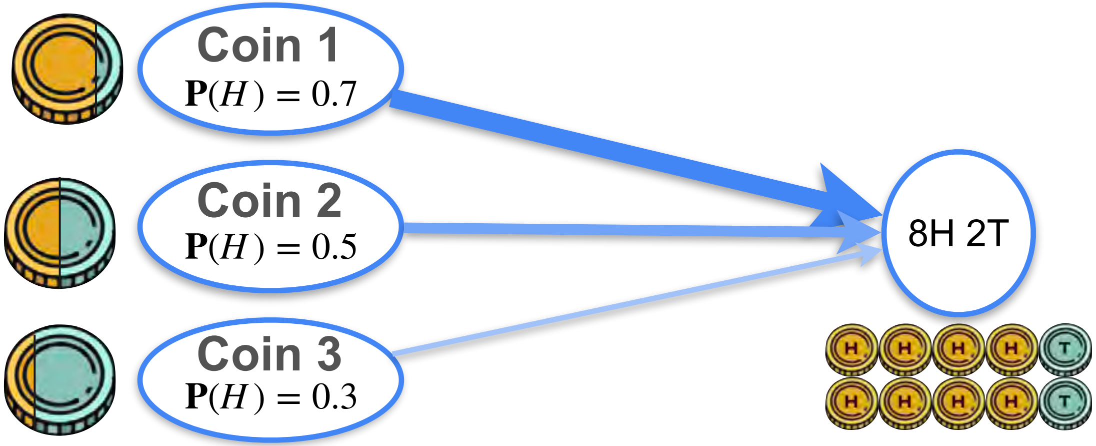
Coin 3
$\mathbf{P}(H) = 0.3$

8H 2T

Maximum Likelihood: Bernoulli Example

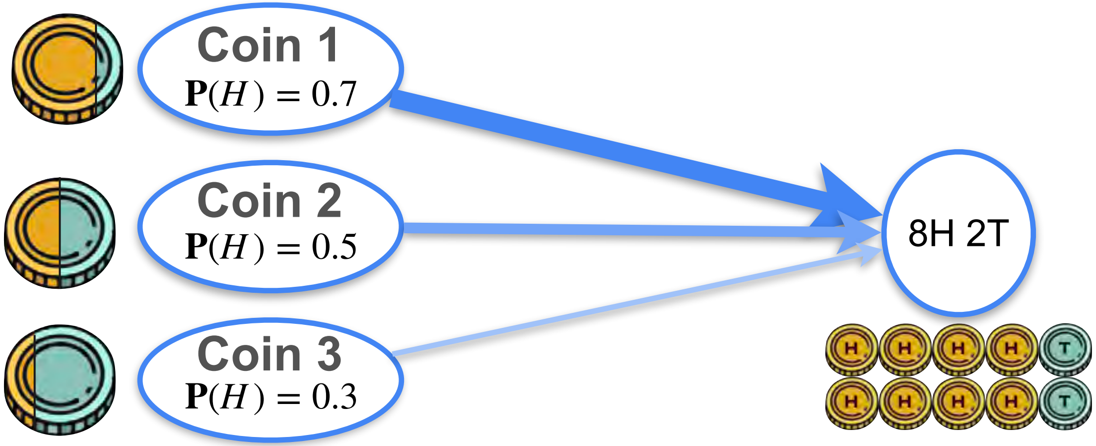


Maximum Likelihood: Bernoulli Example



Find coin that most likely produced the 8 heads and 2 tails

Maximum Likelihood: Bernoulli Example

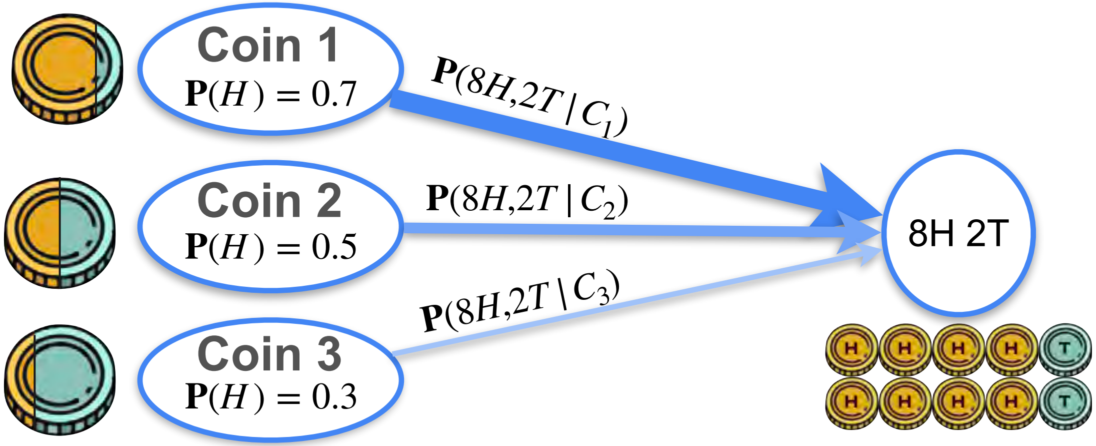


Find coin that most likely produced the 8 heads and 2 tails

Maximize

$$\mathbf{P}(8\text{H}, 2\text{T} \mid \text{Coin})$$

Maximum Likelihood: Bernoulli Example

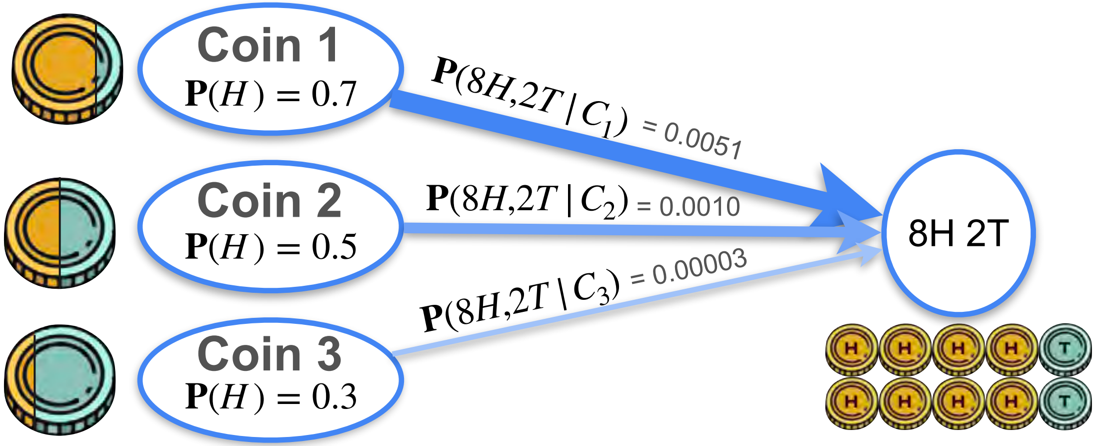


Find coin that most likely produced the 8 heads and 2 tails

Maximize

$\mathbf{P}(\text{8H, 2T} \mid \text{Coin})$

Maximum Likelihood: Bernoulli Example

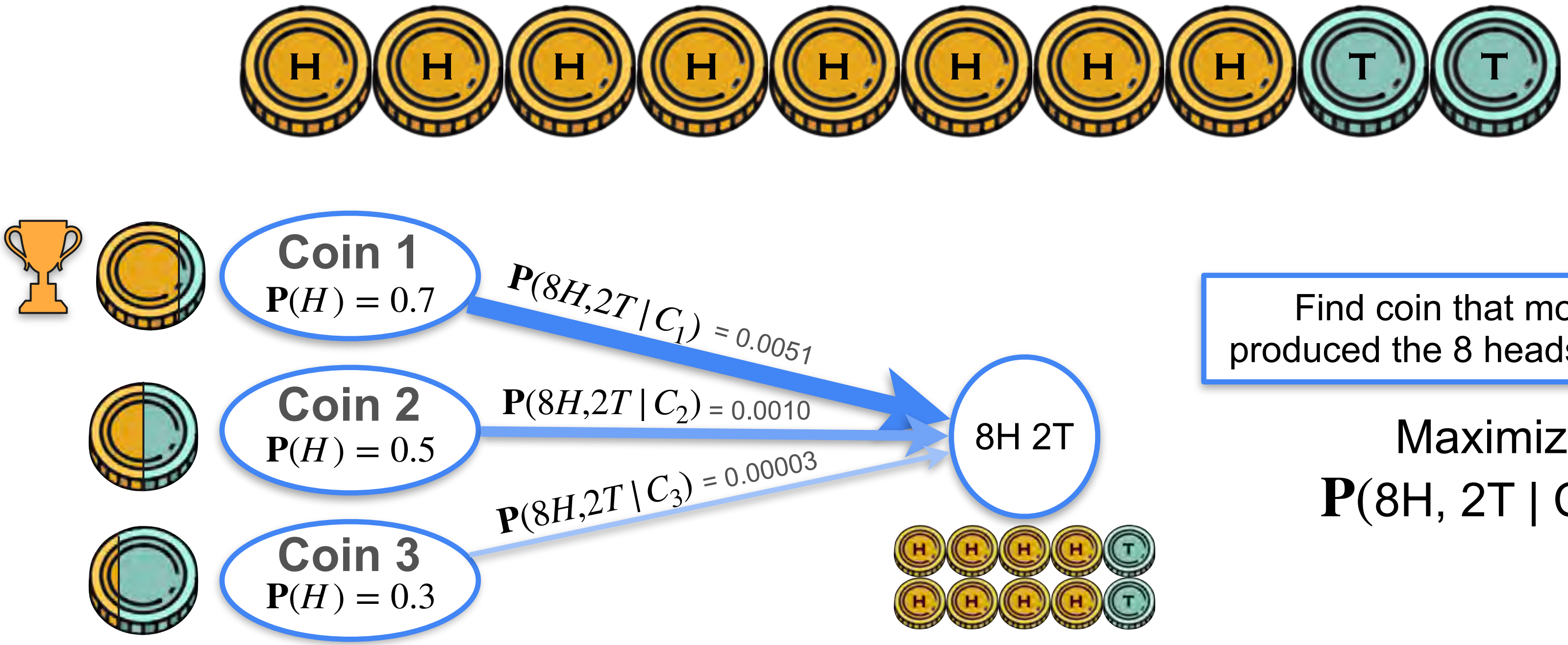


Find coin that most likely produced the 8 heads and 2 tails

Maximize

$\mathbf{P}(8\text{H}, 2\text{T} \mid \text{Coin})$

Maximum Likelihood: Bernoulli Example



Find coin that most likely produced the 8 heads and 2 tails

Maximize

$\mathbf{P}$(8H, 2T | Coin)

Maximum Likelihood: Bernoulli Example

Maximum Likelihood: Bernoulli Example

Can you do any better?

Maximum Likelihood: Bernoulli Example

$p \quad p \quad p \quad p \quad p \quad p \quad p \quad p \quad 1-p \quad 1-p$

Can you do any better?

$p = \mathbf{P}(H)$

Maximum Likelihood: Bernoulli Example

$p \quad p \quad p \quad p \quad p \quad p \quad p \quad p \quad 1-p \quad 1-p$

Can you do any better?

$p = \mathbf{P}(H)$

$p^8(1-p)^2$

Maximum Likelihood: Bernoulli Example

p p p p p p p p $1-p$ $1-p$

Can you do any better?

$p = \mathbf{P}(H)$

$p^8(1-p)^2$

You want p that maximizes the chances of seeing 8H

Maximum Likelihood: Bernoulli Example

$p = \mathbf{P}(H)$

$p^8(1-p)^2$

You want p that maximizes the chances of seeing 8H

Maximum Likelihood: Bernoulli Example

$p = \mathbf{P}(H)$ Likelihood $L(p; 8H) = p^8(1-p)^2$

You want p that maximizes the chances of seeing 8H

Maximum Likelihood: Bernoulli Example

$p = \mathbf{P}(H)$ Likelihood $L(p; 8H) = p^8(1-p)^2$

Function of p

You want p that maximizes the chances of seeing 8H

Maximum Likelihood: Bernoulli Example

$p = \mathbf{P}(H)$ Likelihood $L(p; 8H) = \; p^8(1-p)^2$

Function of p

You want p that maximizes the chances of seeing 8H

$$\log\left((p^8(1-p)^2\right)$$

Maximum Likelihood: Bernoulli Example

$p = \mathbf{P}(H)$ Likelihood $L(p; 8H) = p^8(1-p)^2$ Function of p

You want p that maximizes the chances of seeing 8H

$$\log\left((p^8(1-p)^2\right) = 8\log(p) + 2\log(1-p)$$

Maximum Likelihood: Bernoulli Example

$p = \mathbf{P}(H)$ Likelihood $L(p; 8H) = p^8(1-p)^2$

Function of p

You want p that maximizes the chances of seeing 8H

Log-likelihood $\ell(p; 8H) = \log\left((p^8(1-p)^2\right) = 8\log(p) + 2\log(1-p)$

Maximum Likelihood: Bernoulli Example

$p = \mathbf{P}(H)$ Likelihood $L(p; 8H) = p^8(1-p)^2$

Function of p

You want p that maximizes the chances of seeing 8H

Log-likelihood $\ell(p; 8H) = \log\left((p^8(1-p)^2\right) = 8\log(p) + 2\log(1-p)$

$$\frac{d}{dp}\left(8\log(p) + 2\log(1-p)\right)$$

Maximum Likelihood: Bernoulli Example

$p = \mathbf{P}(H)$ Likelihood $L(p; 8H) = p^8(1-p)^2$

Function of p

You want p that maximizes the chances of seeing 8H

Log-likelihood $\ell(p; 8H) = \log\left((p^8(1-p)^2\right) = 8\log(p) + 2\log(1-p)$

$$\frac{d}{dp}\left(8\log(p) + 2\log(1-p)\right) = \frac{8}{p} + \frac{2}{1-p}(-1)$$

Maximum Likelihood: Bernoulli Example

$p = \mathbf{P}(H)$ Likelihood $L(p; 8H) = \; p^8(1-p)^2$

Function of p

You want p that maximizes the chances of seeing 8H

Log-likelihood $\ell(p; 8H) = \log\left((p^8(1-p)^2\right) = 8\log(p) + 2\log(1-p)$

$$\frac{d}{dp}\left(8\log(p) + 2\log(1-p)\right) = \frac{8}{p} + \frac{2}{1-p}(-1) = 0 \;\Rightarrow\; \hat{p} = \frac{8}{10}$$

The General Case

Maximum Likelihood: Bernoulli Example

n coins

k heads

Maximum Likelihood: Bernoulli Example

n coins

k heads

$X_1 \quad X_2 \quad X_3 \quad \cdots \quad X_{n-1} \quad X_n$

Maximum Likelihood: Bernoulli Example

n coins
k heads

$X_1 \quad X_2 \quad X_3 \quad \cdots \quad X_{n-1} \quad X_n$

$$\mathbf{X} = (X_1, \ldots, X_n)$$

$$X_i \overset{i.i.d}{\sim} Bernoulli(p)$$

Maximum Likelihood: Bernoulli Example

n coins

k heads

$X_1 \quad X_2 \quad X_3 \qquad X_{n-1} \quad X_n$

$$\mathbf{X} = (X_1, \ldots, X_n)$$

$$X_i \overset{i.i.d}{\sim} Bernoulli(p)$$

Likelihood

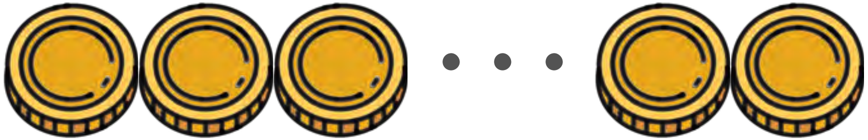
$$L(p; \mathbf{x}) = P_p(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} p_{X_i}(x_i)$$

Maximum Likelihood: Bernoulli Example

n coins

k heads

$X_1 \quad X_2 \quad X_3 \qquad \qquad X_{n-1} \quad X_n$



$\mathbf{X} = (X_1, \ldots, X_n)$

$X_i \overset{i.i.d}{\sim} Bernoulli(p)$

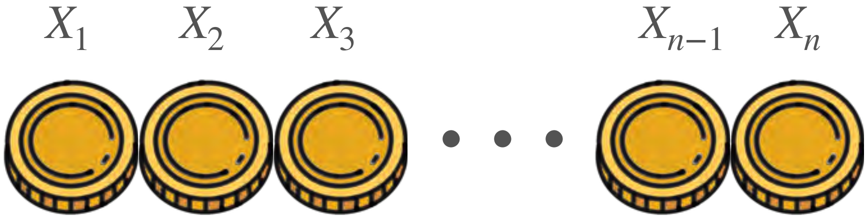
Likelihood

$$L(p; \mathbf{x}) = P_p(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} p_{X_i}(x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

Maximum Likelihood: Bernoulli Example

n coins

k heads



$\mathbf{X} = (X_1, \ldots, X_n)$

$X_i \overset{i.i.d}{\sim} Bernoulli(p)$

Likelihood

$$L(p; \mathbf{x}) = P_p(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} p_{X_i}(x_i) = \prod_{i=1}^{n} \underbrace{p^{x_i}(1-p)^{1-x_i}}$$

If $x_i = 1,\ p^{[x_i]}(1-p)^{[1-x_i]} = p$

If $x_i = 0,\ p^{[x_i]}(1-p)^{[1-x_i]} = (1-p)$

Maximum Likelihood: Bernoulli Example

n coins

k heads

$X_1 \quad X_2 \quad X_3 \qquad \cdots \qquad X_{n-1} \quad X_n$

$\mathbf{X} = (X_1, \ldots, X_n)$

$X_i \overset{i.i.d}{\sim} Bernoulli(p)$

Likelihood

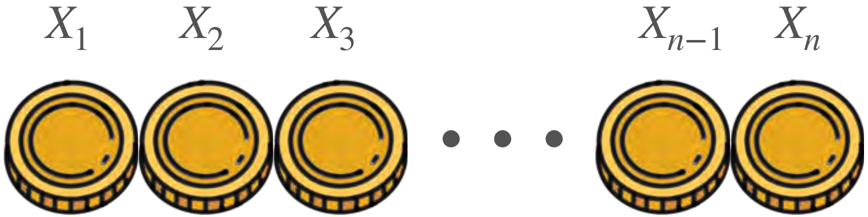
$$L(p; \mathbf{x}) = P_p(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} p_{X_i}(x_i) = \prod_{i=1}^{n} \underbrace{p^{x_i}(1-p)^{1-x_i}}$$

If $x_i = 1,\ p^{[x_i]}(1-p)^{[1-x_i]} = p$

If $x_i = 0,\ p^{[x_i]}(1-p)^{[1-x_i]} = (1-p)$

$\sum_{i=1}^{n} x_i =$ # heads $\qquad n - \sum_{i=1}^{n} x_i =$ # tails

Maximum Likelihood: Bernoulli Example



$\mathbf{X} = (X_1, \ldots, X_n)$

$X_i \overset{i.i.d}{\sim} Bernoulli(p)$

Likelihood

$$L(p; \mathbf{x}) = P_p(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} p_{X_i}(x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

Maximum Likelihood: Bernoulli Example

$$\mathbf{X} = (X_1, \ldots, X_n)$$

$$X_i \overset{i.i.d}{\sim} Bernoulli(p)$$

Likelihood

$$L(p; \mathbf{x}) = P_p(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} p_{X_i}(x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\left(\sum_{i=1}^{n} x_i\right)}(1-p)^{\left(n - \sum_{i=1}^{n} x_i\right)}$$

Maximum Likelihood: Bernoulli Example

$$\mathbf{X} = (X_1, \ldots, X_n)$$

$$X_i \overset{i.i.d}{\sim} Bernoulli(p)$$

Likelihood

$$L(p; \mathbf{x}) = P_p(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} p_{X_i}(x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\left(\sum_{i=1}^{n} x_i\right)}(1-p)^{\left(n - \sum_{i=1}^{n} x_i\right)}$$

Log-likelihood

$$\ell(p; \mathbf{x}) = \log\left((p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}\right)$$

Maximum Likelihood: Bernoulli Example

$X_1 \quad X_2 \quad X_3 \qquad\qquad X_{n-1} \quad X_n$

$\mathbf{X} = (X_1, \ldots, X_n)$

$X_i \overset{i.i.d}{\sim} Bernoulli(p)$

Likelihood

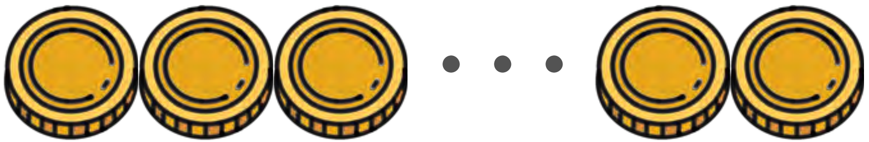
$$L(p;\mathbf{x}) = P_p(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} p_{X_i}(x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\left(\sum_{i=1}^{n} x_i\right)}(1-p)^{\left(n-\sum_{i=1}^{n} x_i\right)}$$

Log-likelihood

$$\ell(p;\mathbf{x}) = \log\left((p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}\right) = \left(\sum_{i=1}^{n} x_i\right)\log(p) + \left(n - \sum_{i=1}^{n} x_i\right)\log(1-p)$$

Maximum Likelihood: Bernoulli Example

$X_1 \quad X_2 \quad X_3 \quad \cdots \quad X_{n-1} \quad X_n$



$$\mathbf{X} = (X_1, \ldots, X_n)$$

$$X_i \overset{i.i.d}{\sim} Bernoulli(p)$$

$$\ell(p;\mathbf{x}) = \log\left(\left(p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}\right)\right) = \left(\sum_{i=1}^n x_i\right)\log(p) + \left(n - \sum_{i=1}^n x_i\right)\log(1-p)$$

Maximum Likelihood: Bernoulli Example

$$\mathbf{X} = (X_1, \ldots, X_n)$$

$$X_i \overset{i.i.d}{\sim} Bernoulli(p)$$

$$\ell(p; \mathbf{x}) = \log\left((p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} \right) = \left(\sum_{i=1}^n x_i \right) \log(p) + \left(n - \sum_{i=1}^n x_i \right) \log(1-p)$$

Find the maximum!

$$\frac{d}{dp}\ell(p; \mathbf{x}) = \frac{d}{dp}\left(\left(\sum_{i=1}^n x_i \right) \log(p) + \left(n - \sum_{i=1}^n x_i \right) \log(1-p) \right)$$

$$= \frac{\sum_{i=1}^n x_i}{p} + \frac{n - \sum_{i=1}^n x_i}{1-p}(-1) = 0$$

Maximum Likelihood: Bernoulli Example

$X_1 \quad X_2 \quad X_3 \quad \cdots \quad X_{n-1} \quad X_n$

$$\mathbf{X} = (X_1, \ldots, X_n)$$

$$X_i \overset{i.i.d}{\sim} Bernoulli(p)$$

$$\ell(p; \mathbf{x}) = \log\left(\left(p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}\right)\right) = \left(\sum_{i=1}^n x_i\right)\log(p) + \left(n - \sum_{i=1}^n x_i\right)\log(1-p)$$

Find the maximum!

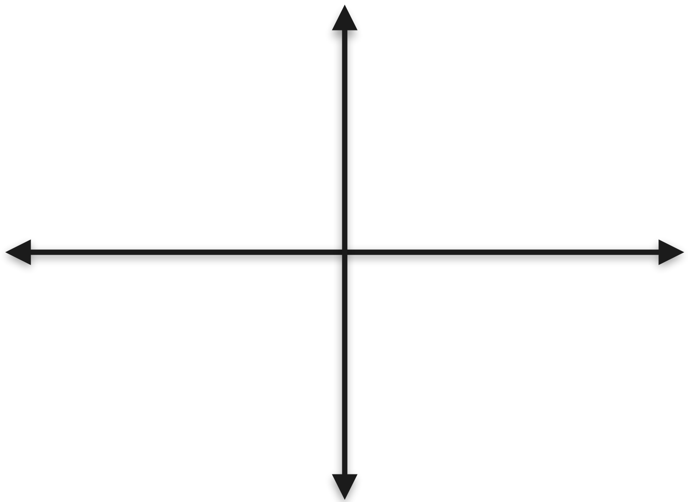
$$\frac{d}{dp}\ell(p; \mathbf{x}) = \frac{d}{dp}\left(\left(\sum_{i=1}^n x_i\right)\log(p) + \left(n - \sum_{i=1}^n x_i\right)\log(1-p)\right)$$

$$= \frac{\sum_{i=1}^n x_i}{p} + \frac{n - \sum_{i=1}^n x_i}{1-p}(-1) = 0 \quad \rightarrow \quad \hat{p} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$$

Point Estimation

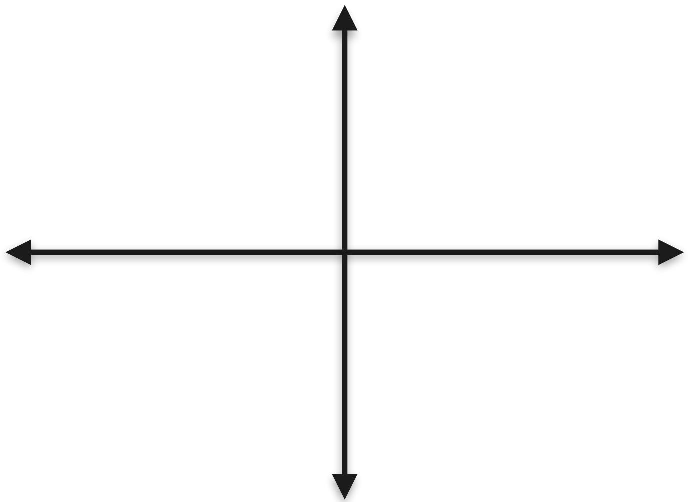
MLE: Gaussian Example

Maximum Likelihood: Gaussian Example



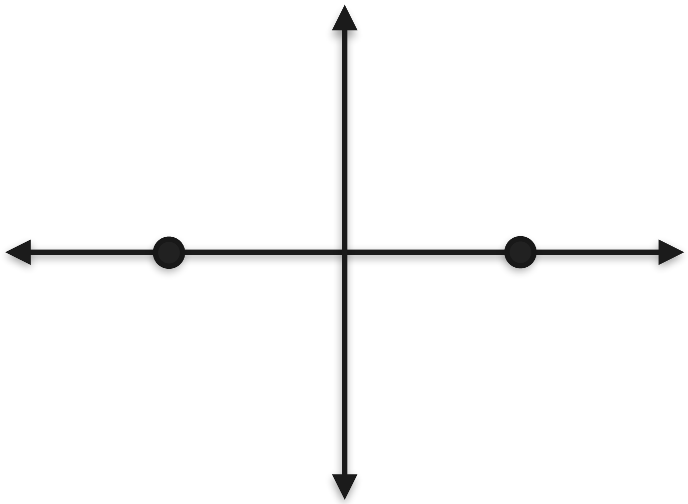
Maximum Likelihood: Gaussian Example

Observations



Maximum Likelihood: Gaussian Example

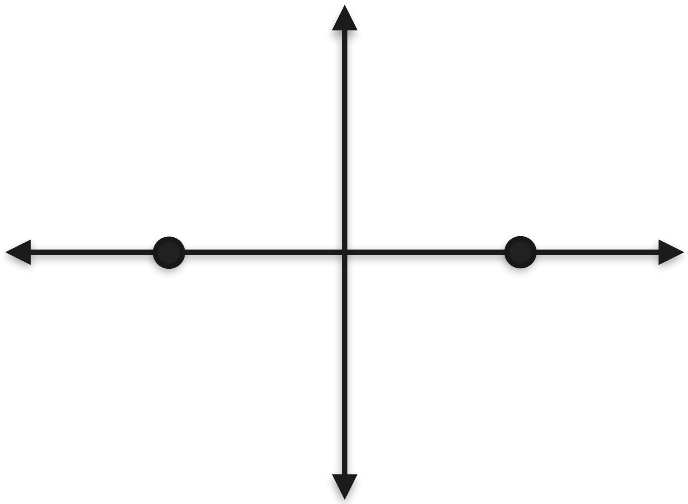
Observations



Maximum Likelihood: Gaussian Example

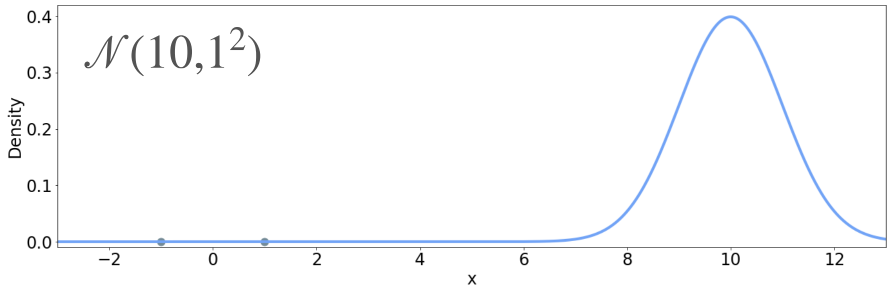
Candidates

Observations

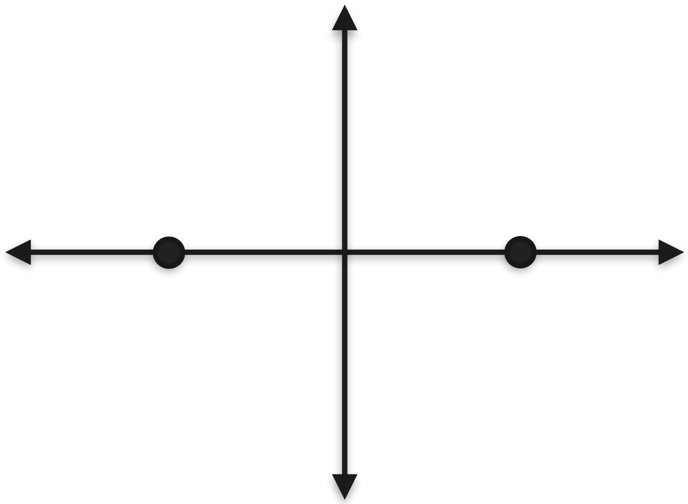


Maximum Likelihood: Gaussian Example

Candidates

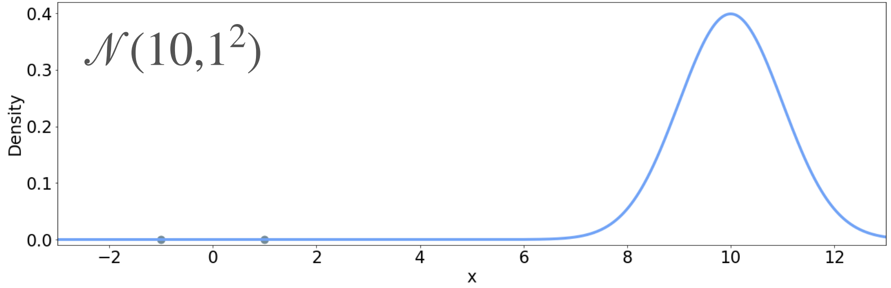


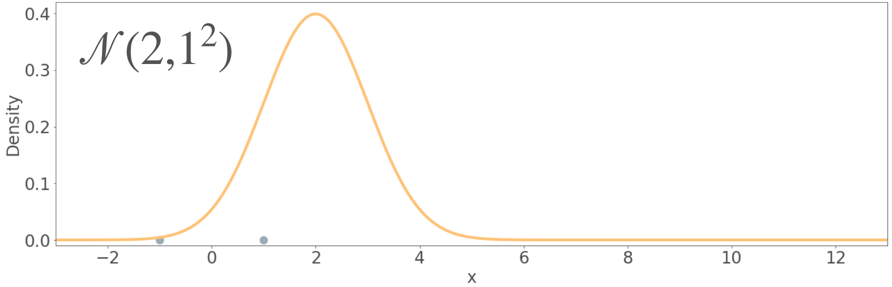
Observations



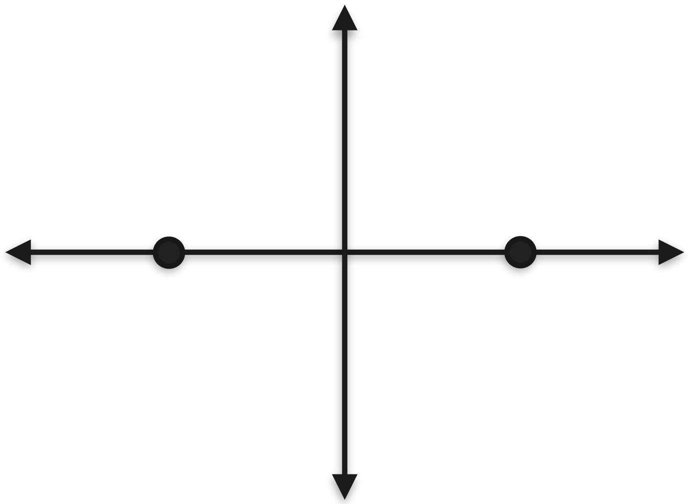
Maximum Likelihood: Gaussian Example

Candidates



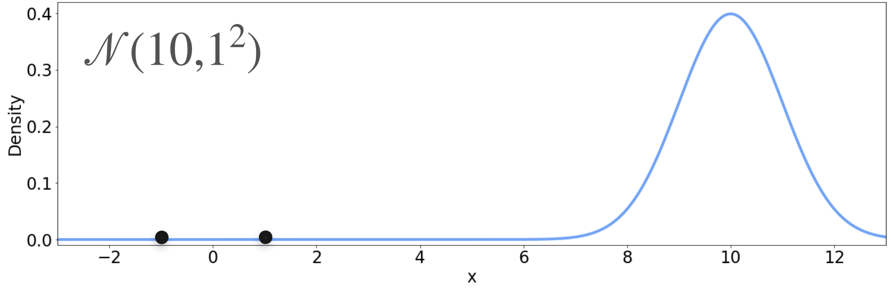


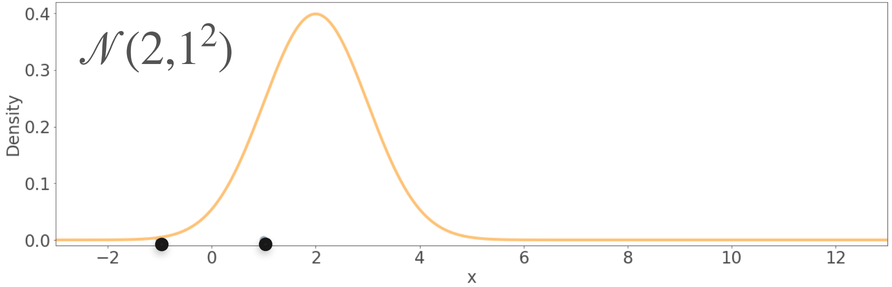
Observations



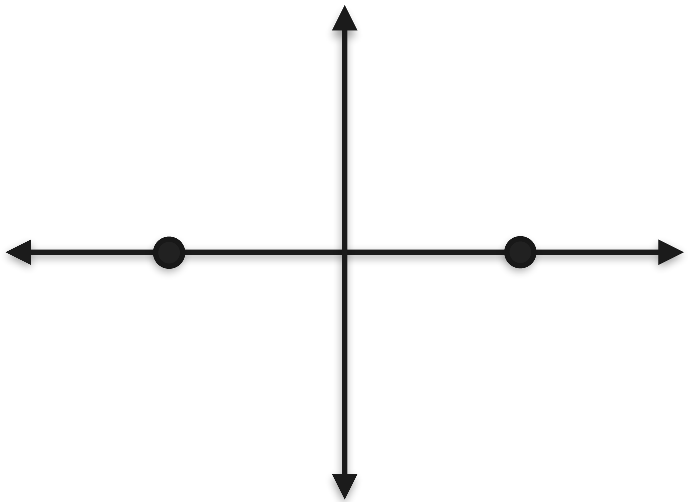
Maximum Likelihood: Gaussian Example

Candidates



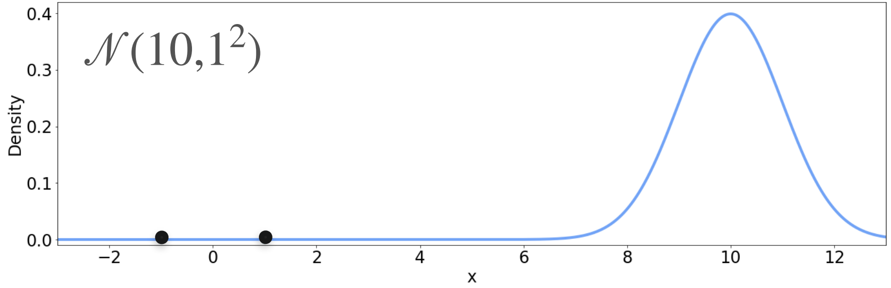


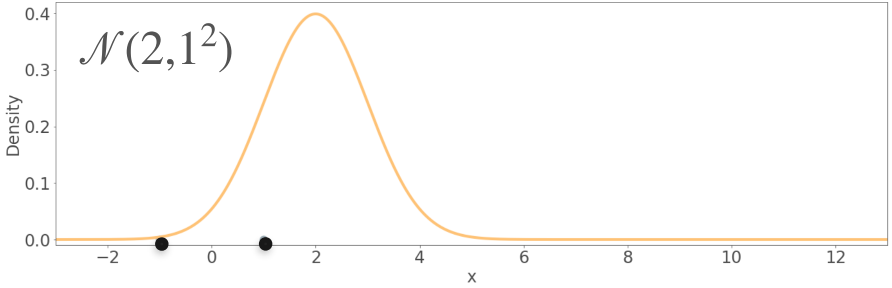
Observations



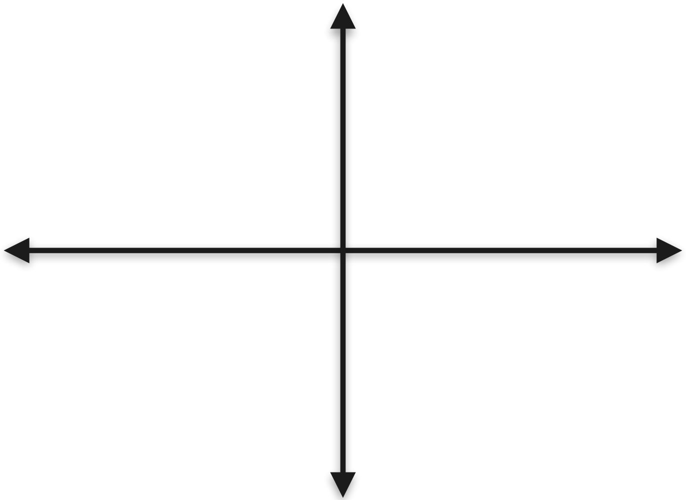
Maximum Likelihood: Gaussian Example

Candidates



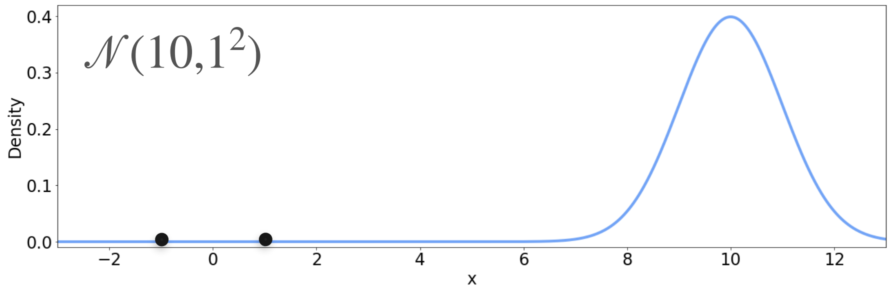


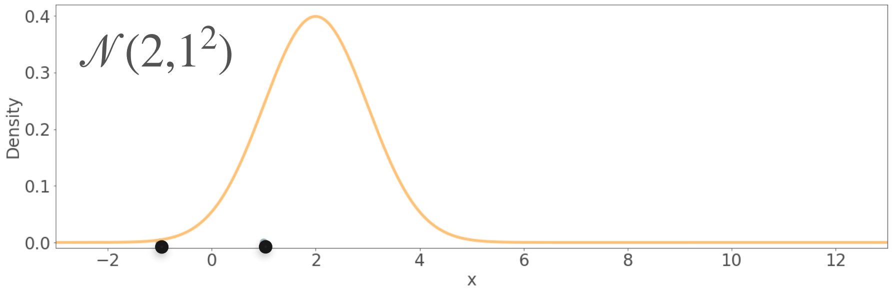
Observations



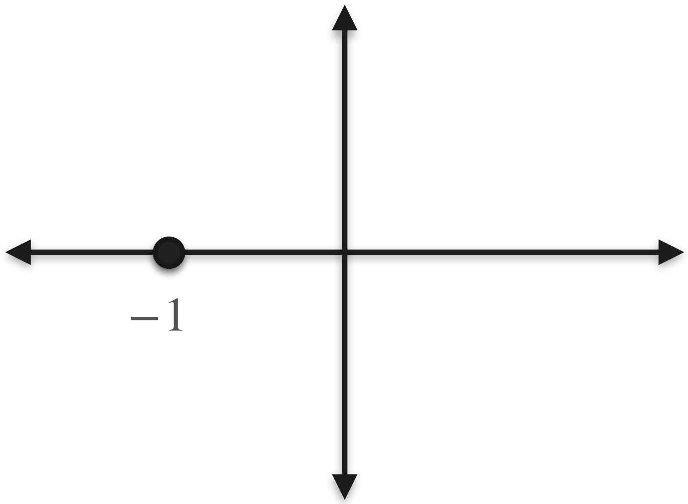
Maximum Likelihood: Gaussian Example

Candidates



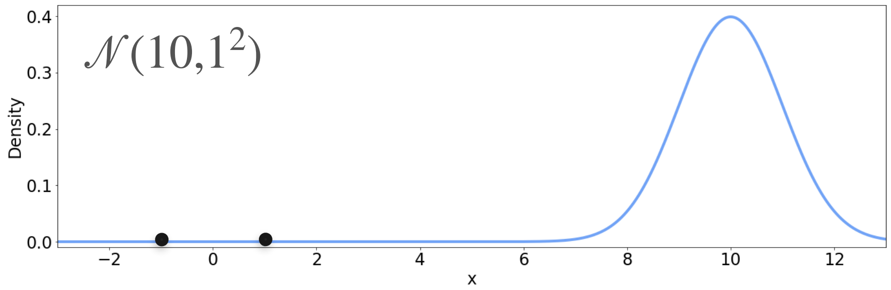


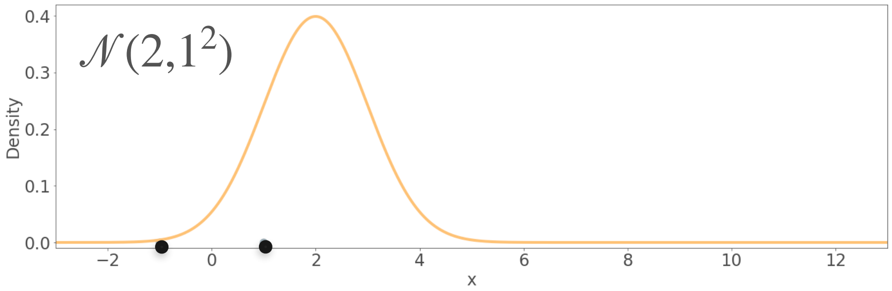
Observations



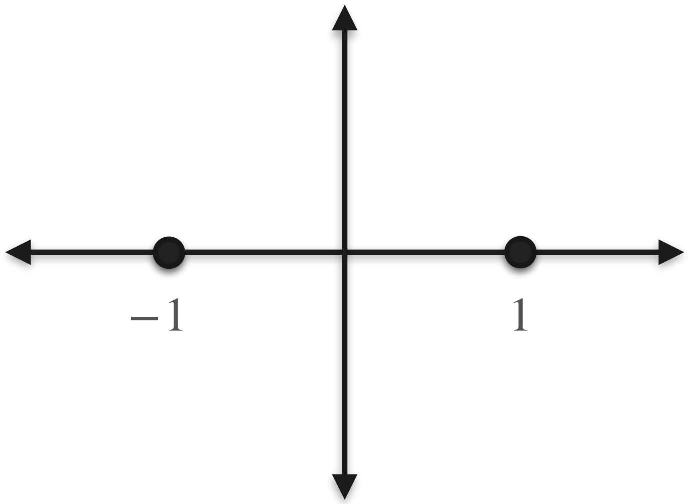
Maximum Likelihood: Gaussian Example

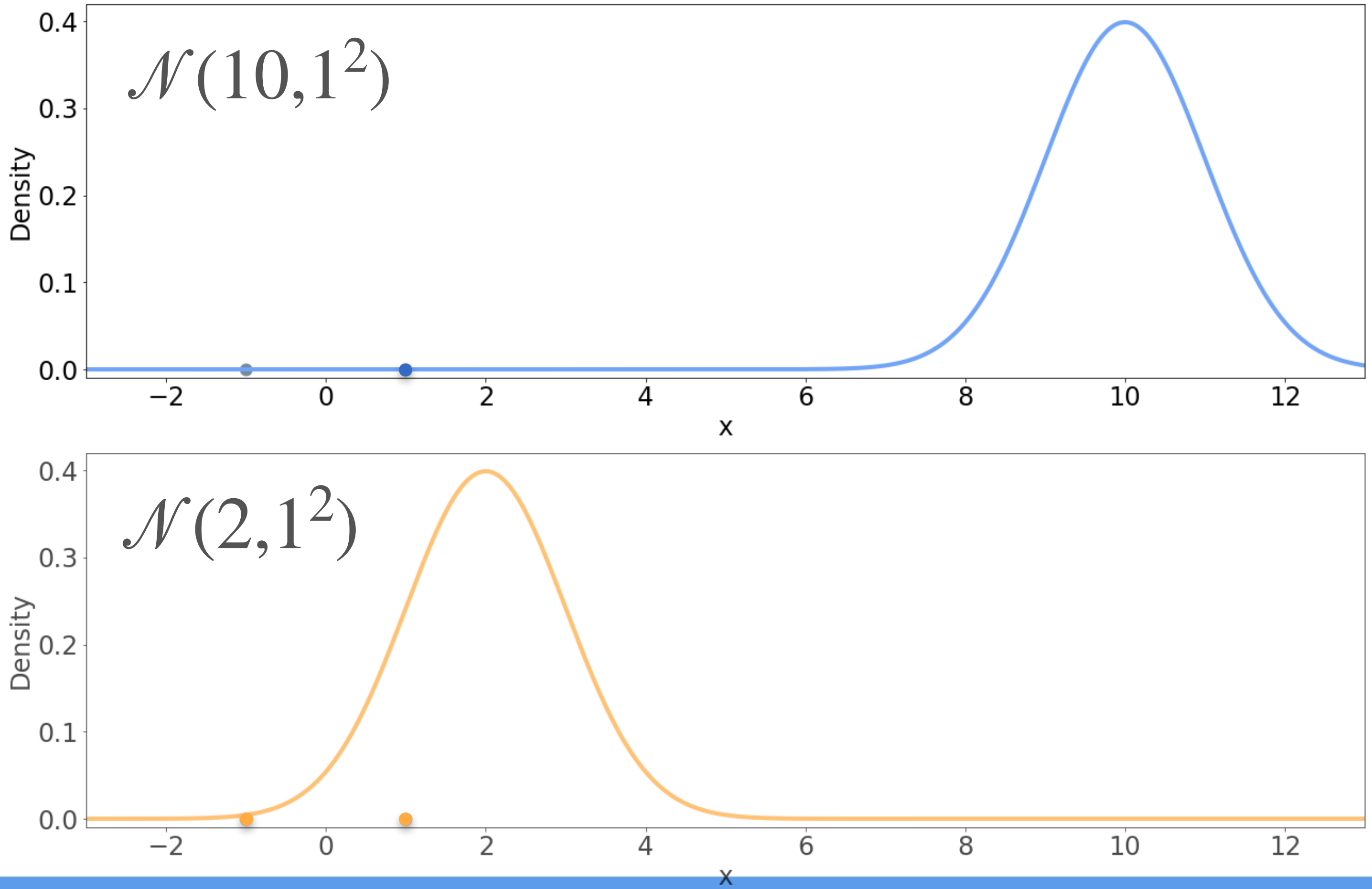
Candidates

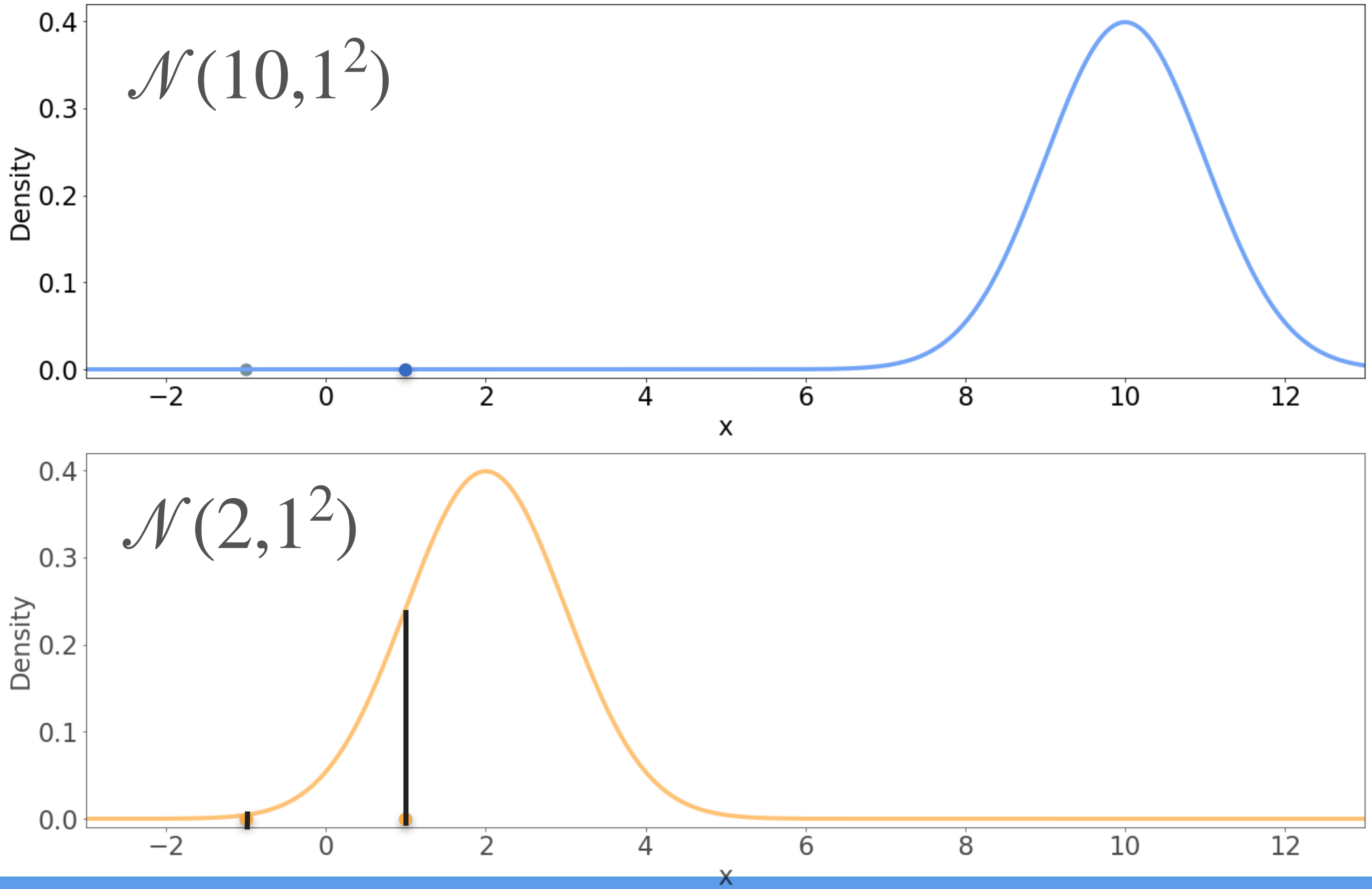




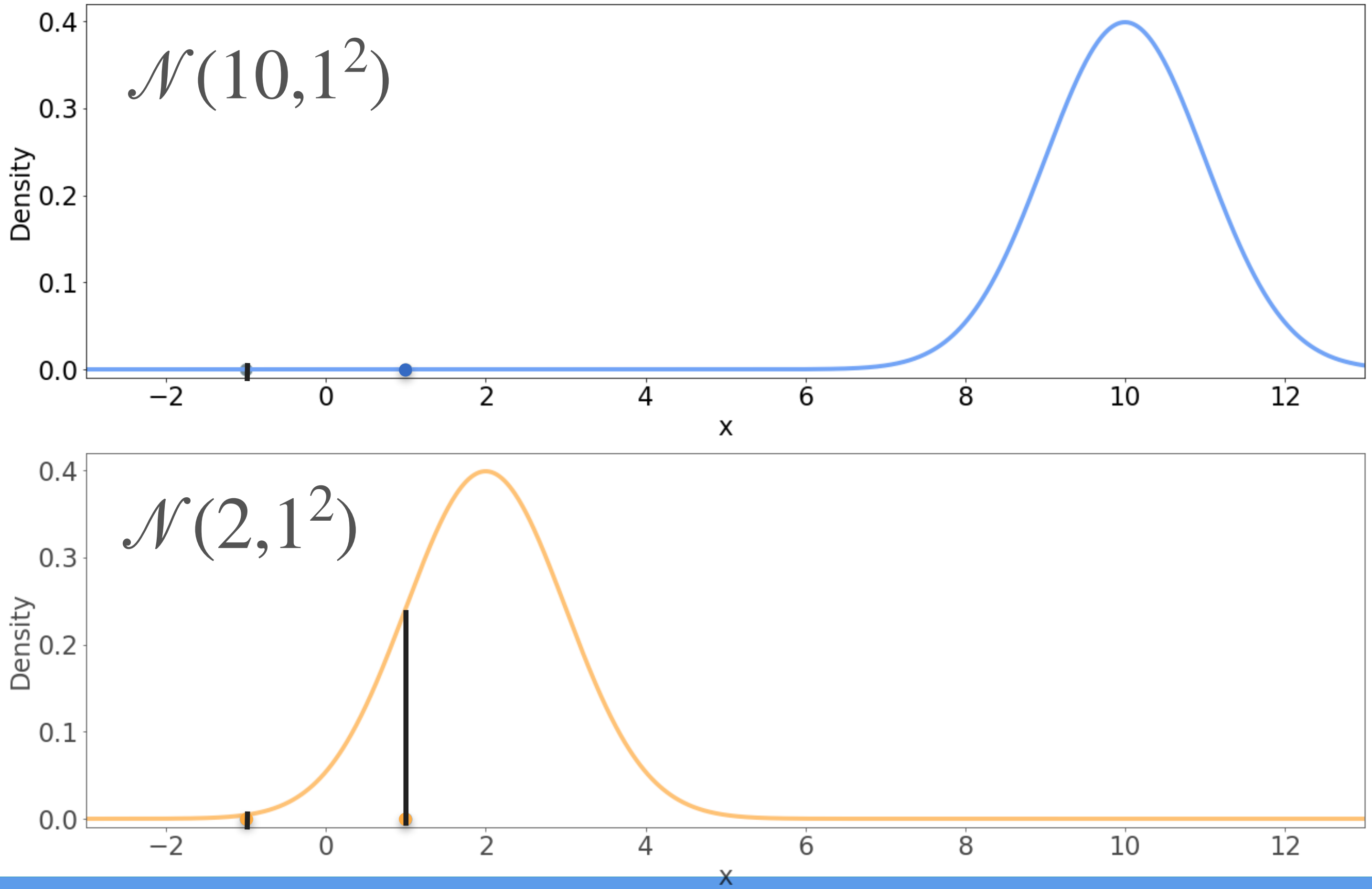
Observations



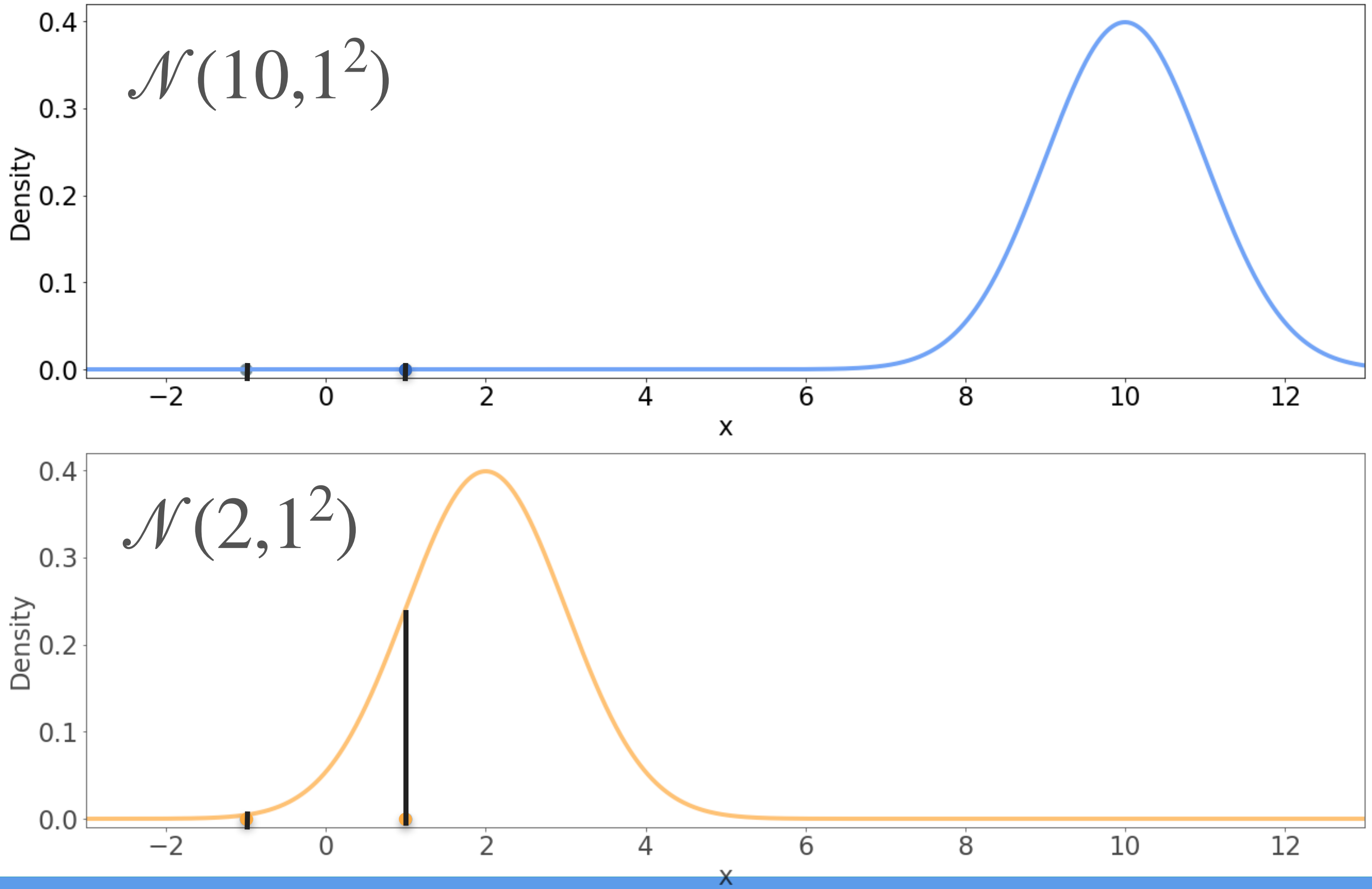

$\mathcal{N}(10,1^2)$
Density
0.4
0.3
0.2
0.1
0.0
−2
0
2
4
6
8
10
12
x
$\mathcal{N}(2,1^2)$
Density
0.4
0.3
0.2
0.1
0.0
−2
0
2
4
6
8
10
12
x

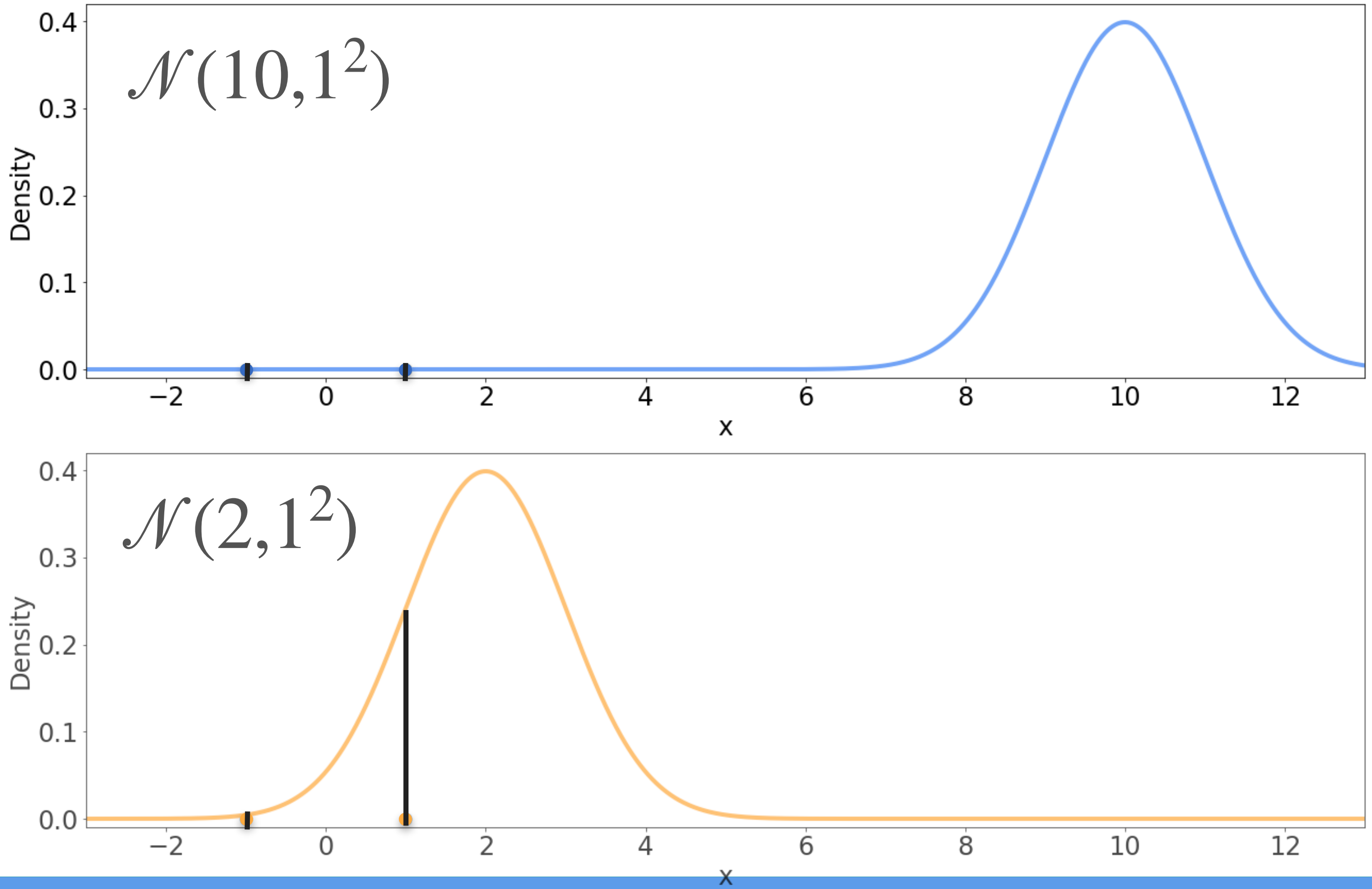


$\mathcal{N}(10,1^2)$
Density
0.4
0.3
0.2
0.1
0.0
−2
0
2
4
6
8
10
12
x
$\mathcal{N}(2,1^2)$
Density
0.4
0.3
0.2
0.1
0.0
−2
0
2
4
6
8
10
12
x

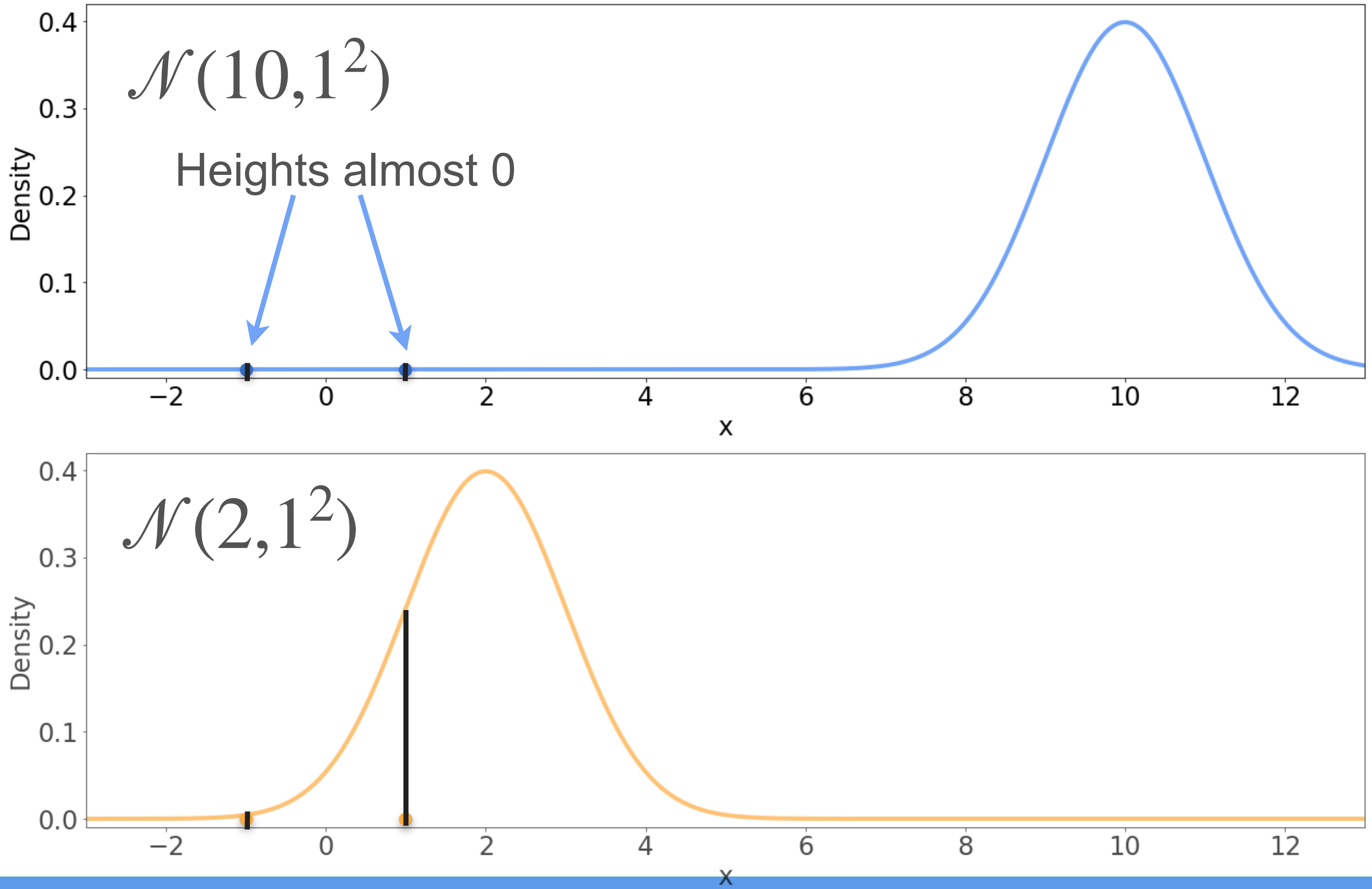


$\mathcal{N}(10,1^2)$
Density
0.4
0.3
0.2
0.1
0.0
−2
0
2
4
6
8
10
12
x
$\mathcal{N}(2,1^2)$
Density
0.4
0.3
0.2
0.1
0.0
−2
0
2
4
6
8
10
12
x

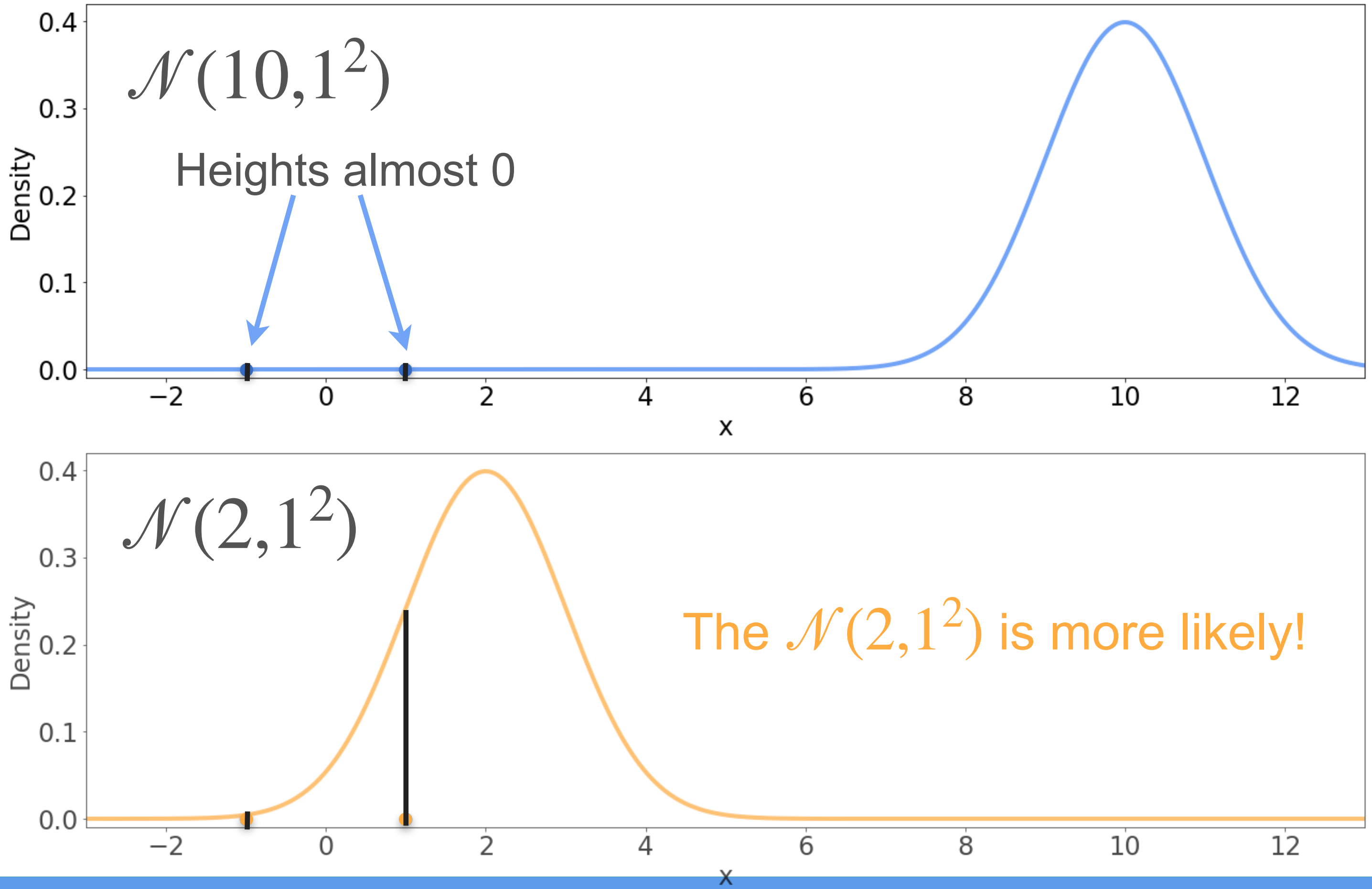

$\mathcal{N}(10,1^2)$
Density
0.4
0.3
0.2
0.1
0.0
−2
0
2
4
6
8
10
12
x
$\mathcal{N}(2,1^2)$
Density
0.4
0.3
0.2
0.1
0.0
−2
0
2
4
6
8
10
12
x

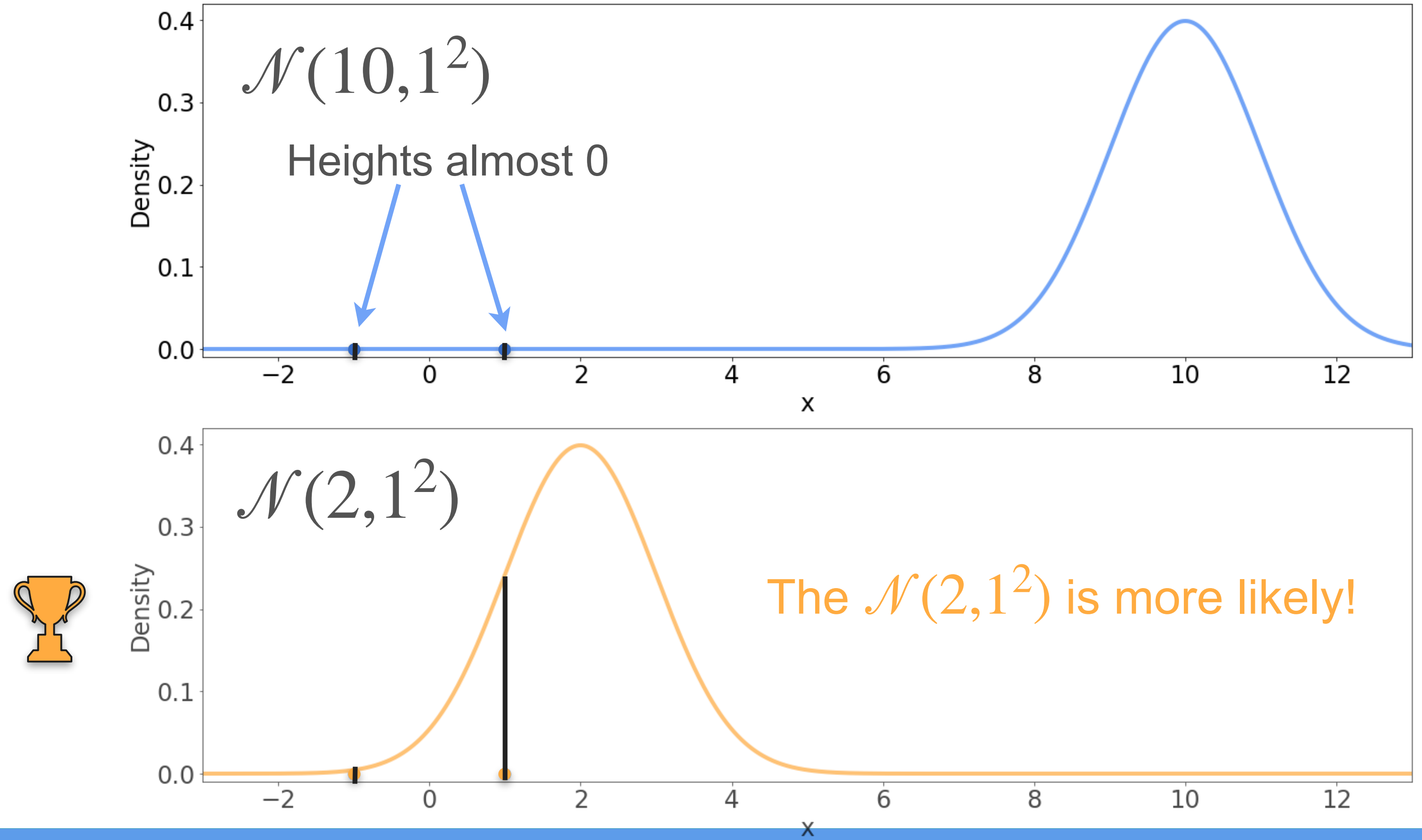


$\mathcal{N}(10,1^2)$
Density
0.4
0.3
0.2
0.1
0.0
−2
0
2
4
6
8
10
12
x
$\mathcal{N}(2,1^2)$
Density
0.4
0.3
0.2
0.1
0.0
−2
0
2
4
6
8
10
12
x



$\mathcal{N}(10,1^2)$
Heights almost 0
Density
0.4
0.3
0.2
0.1
0.0
−2
0
2
4
6
8
10
12
x
$\mathcal{N}(2,1^2)$
Density
0.4
0.3
0.2
0.1
0.0
−2
0
2
4
6
8
10
12
x

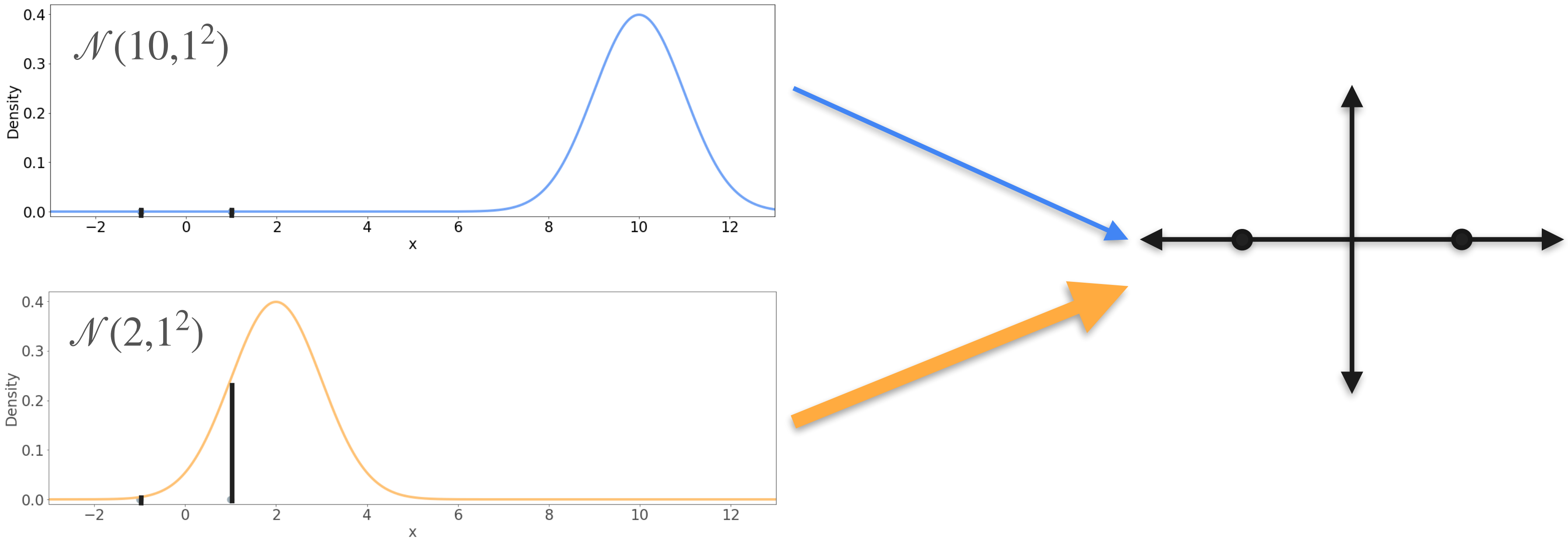

$\mathcal{N}(10,1^2)$
Heights almost 0
Density
0.4
0.3
0.2
0.1
0.0
−2
0
2
4
6
8
10
12
x
$\mathcal{N}(2,1^2)$
The $\mathcal{N}(2,1^2)$ is more likely!
Density
0.4
0.3
0.2
0.1
0.0
−2
0
2
4
6
8
10
12
x


$\mathcal{N}(10,1^2)$
Heights almost 0
Density
x
$\mathcal{N}(2,1^2)$
The $\mathcal{N}(2,1^2)$ is more likely!
Density
x

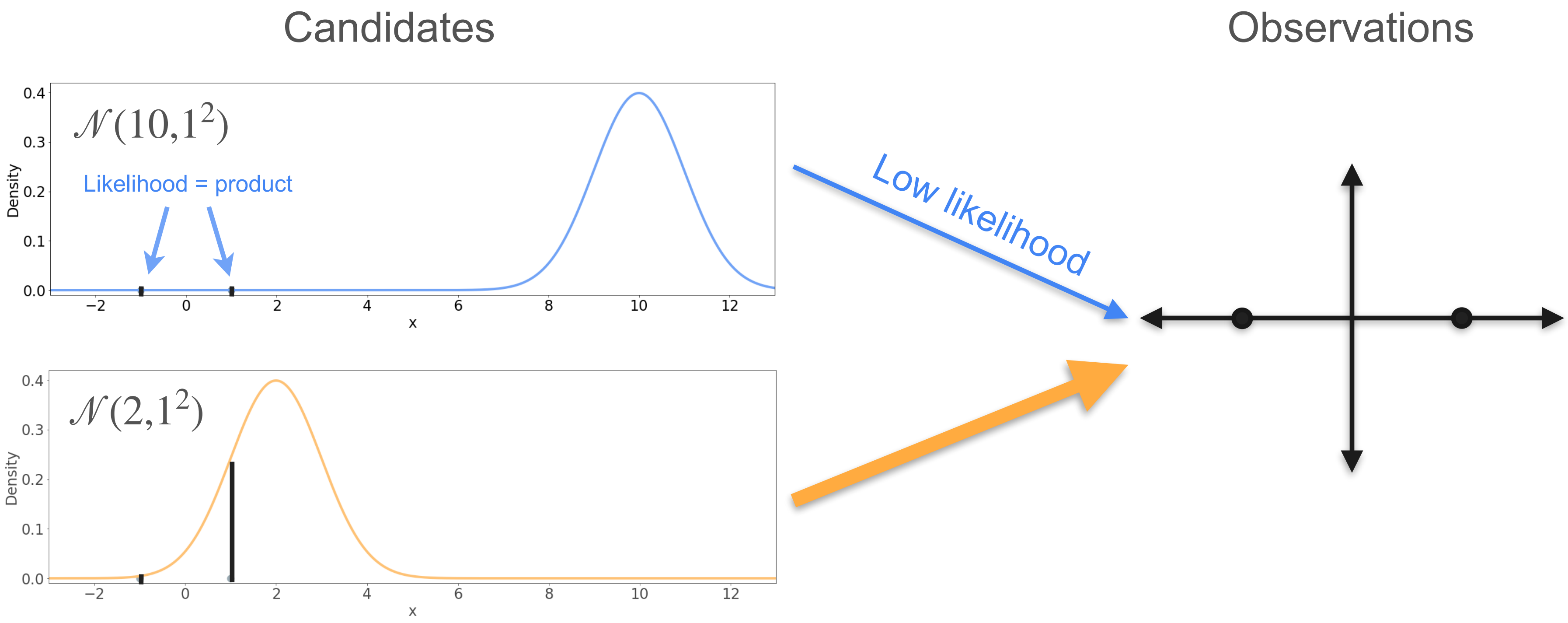
Maximum Likelihood: Gaussian Example

Candidates

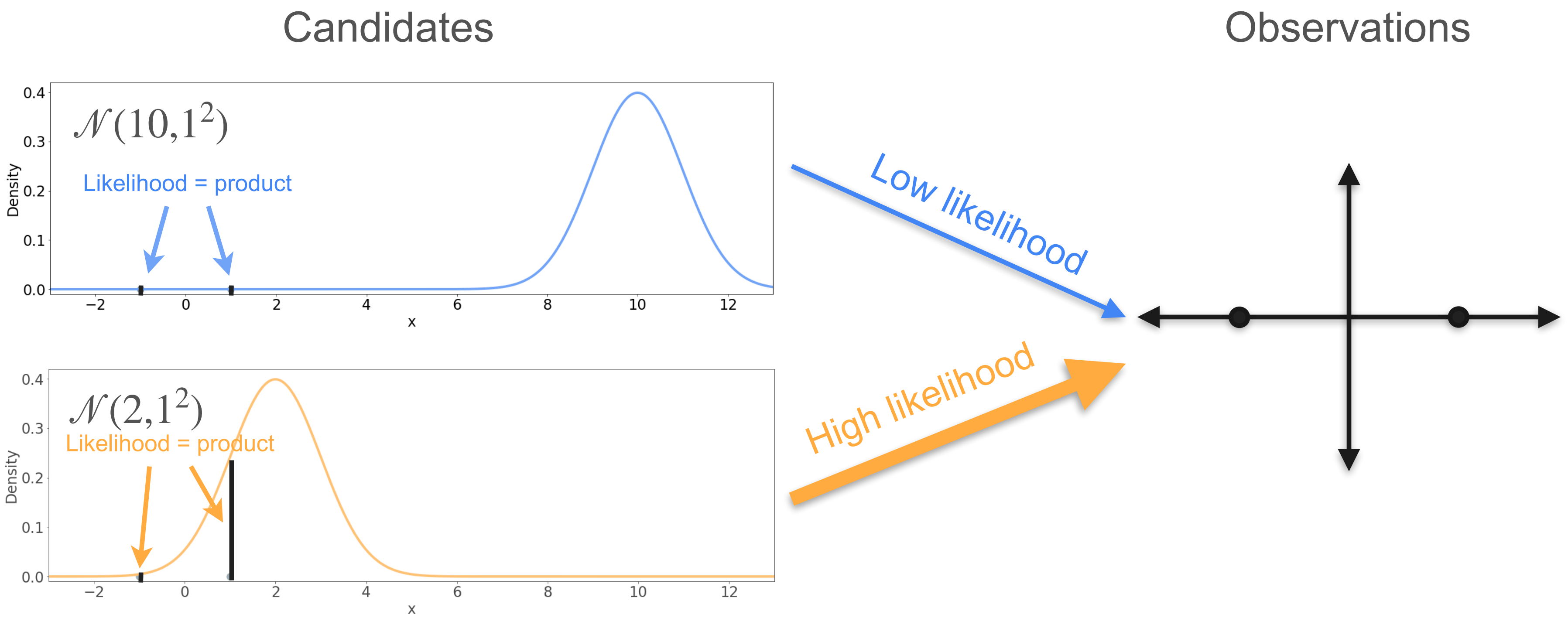
Observations



Maximum Likelihood: Gaussian Example



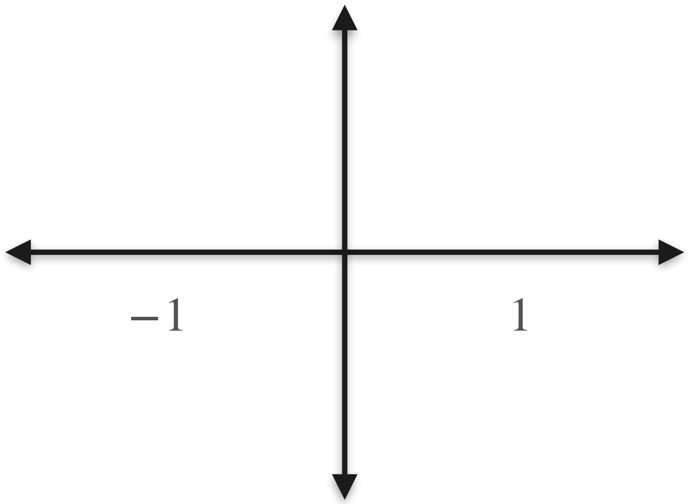
Maximum Likelihood: Gaussian Example



Gaussians With Three Different Means

Candidates

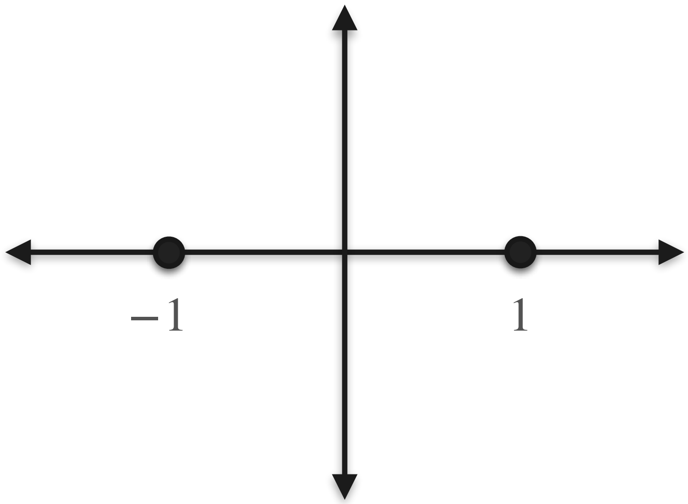
Observations



Gaussians With Three Different Means

Candidates

Observations

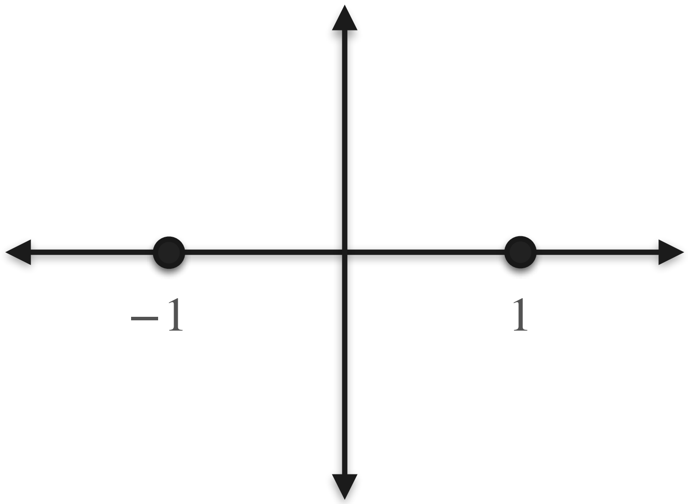


Gaussians With Three Different Means

Candidates

$\mathcal{N}(-1, 1^2)$

Observations



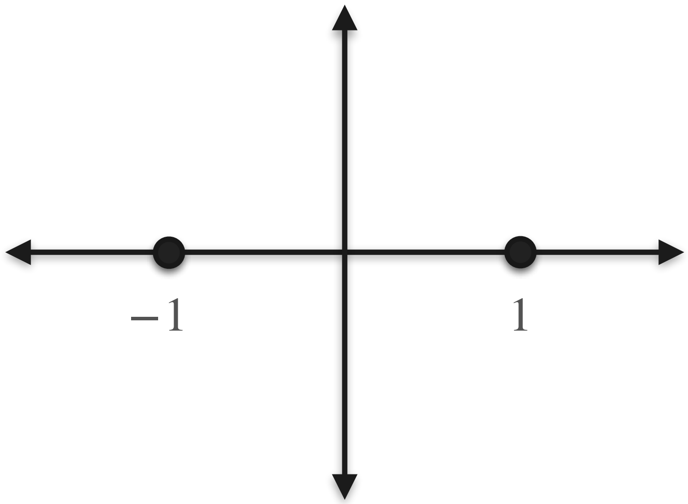
Gaussians With Three Different Means

Candidates

$\mathcal{N}(-1,1^2)$

$\mathcal{N}(0,1^2)$

Observations



Gaussians With Three Different Means

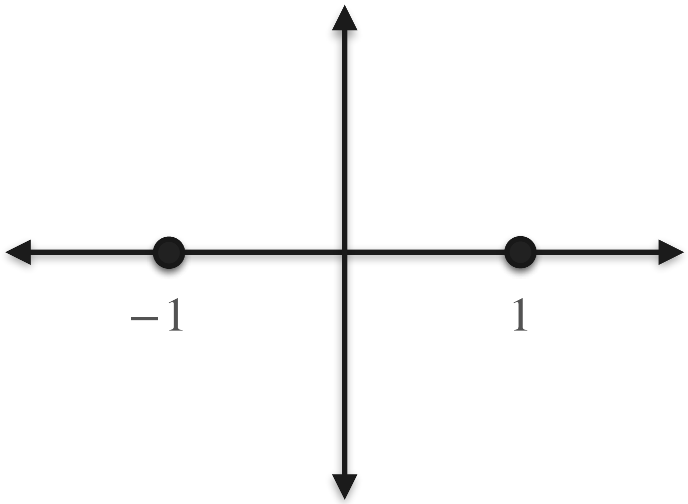
Candidates

$\mathcal{N}(-1, 1^2)$

$\mathcal{N}(0, 1^2)$

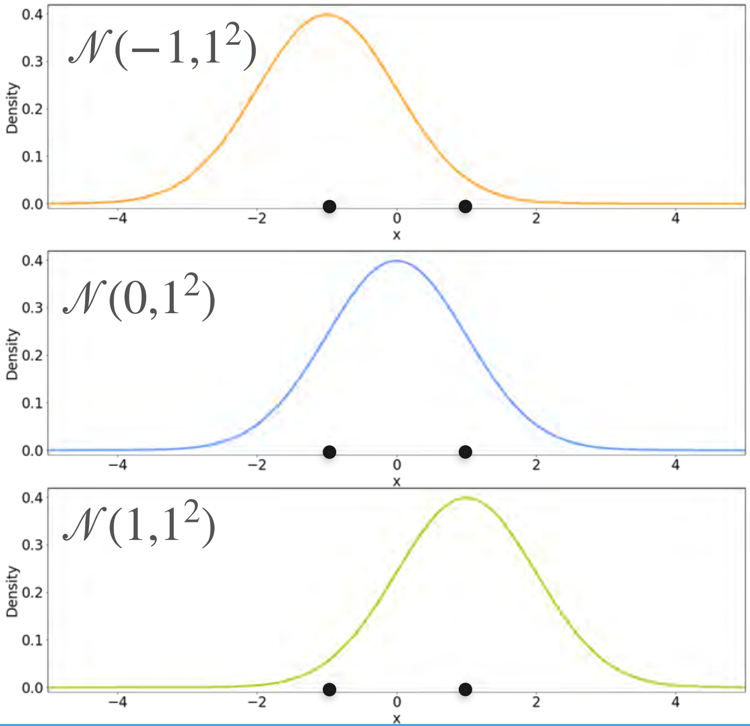
$\mathcal{N}(1, 1^2)$

Observations

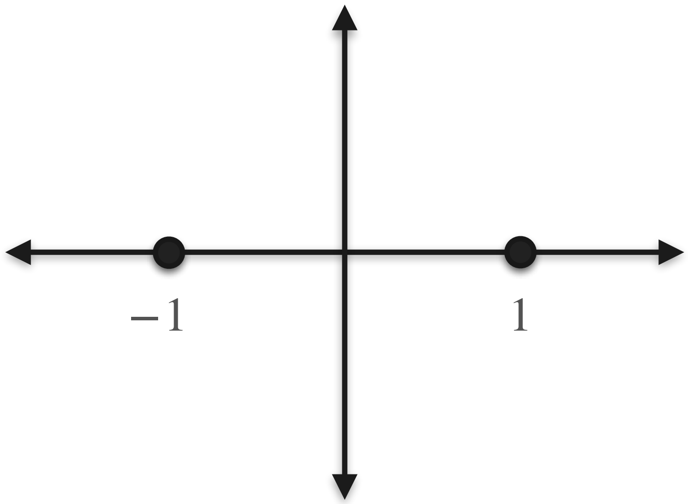


Gaussians With Three Different Means

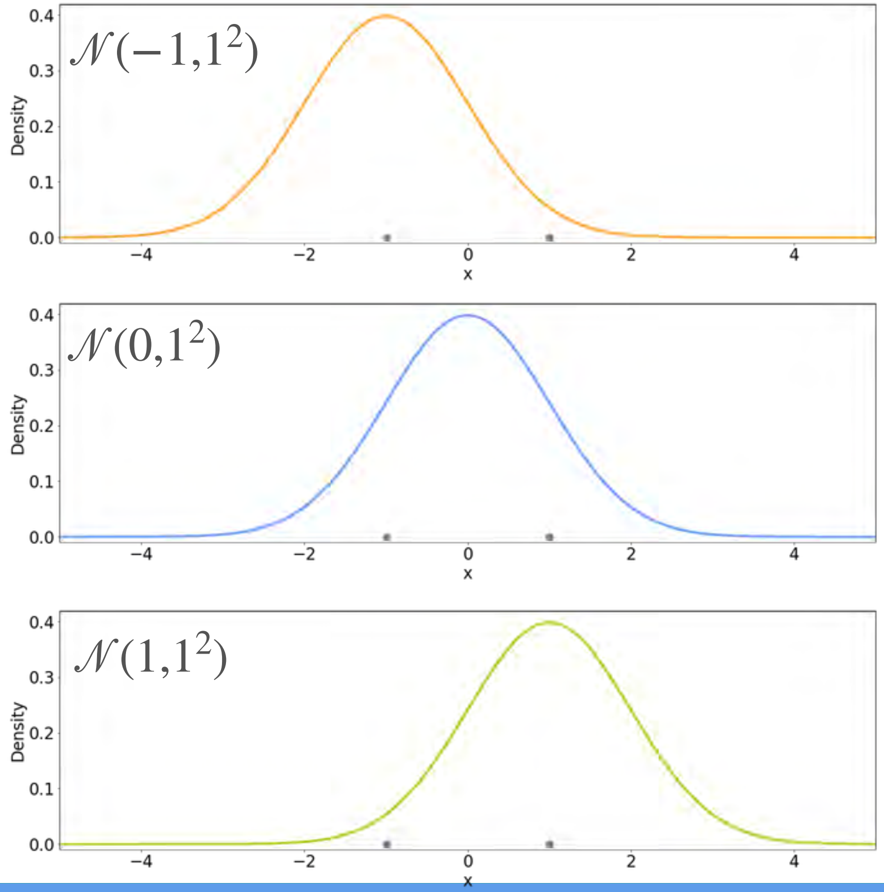
Candidates



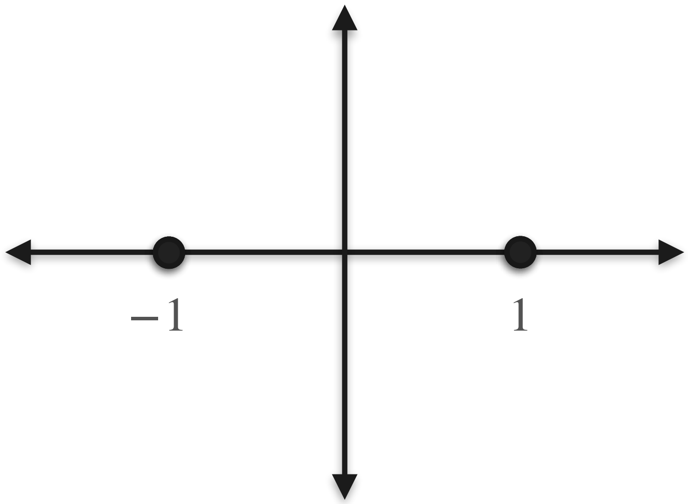
Observations



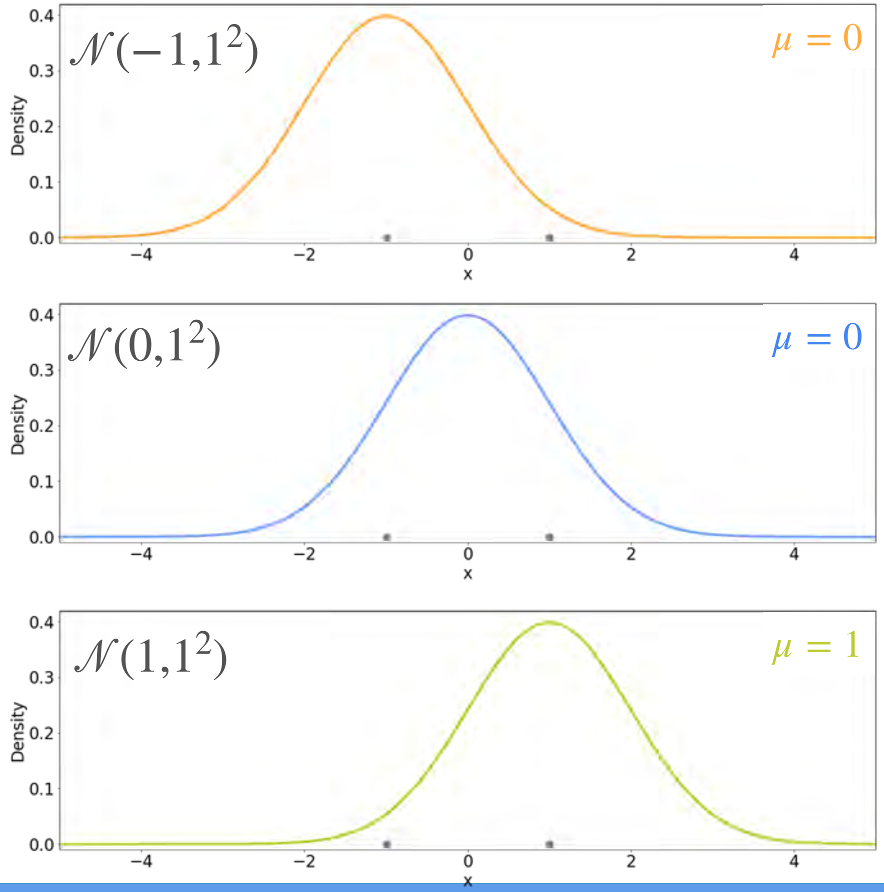
Candidates



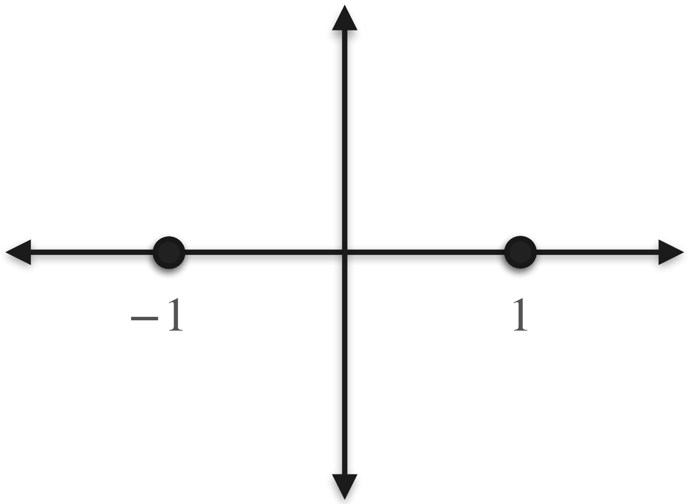
Observations



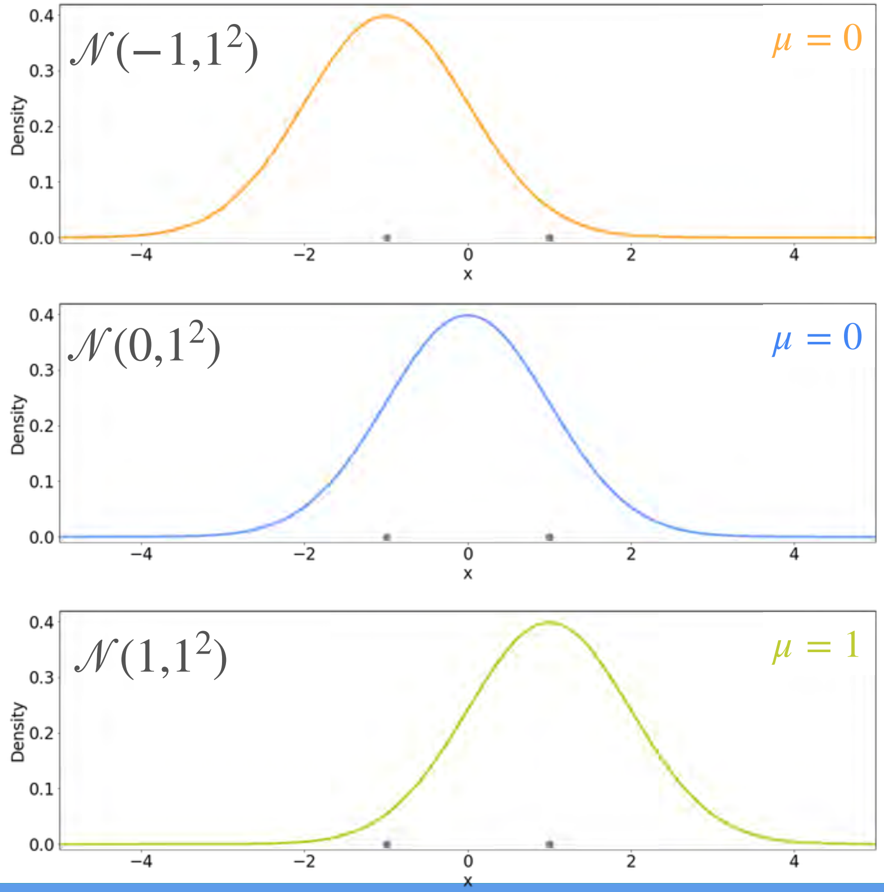
Candidates



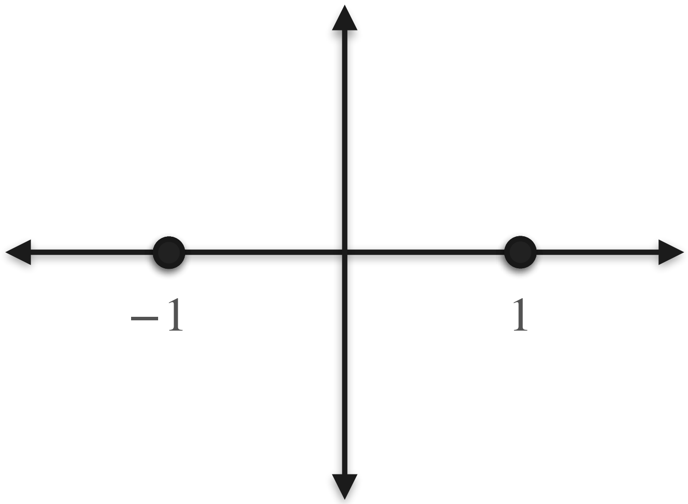
Observations



Candidates

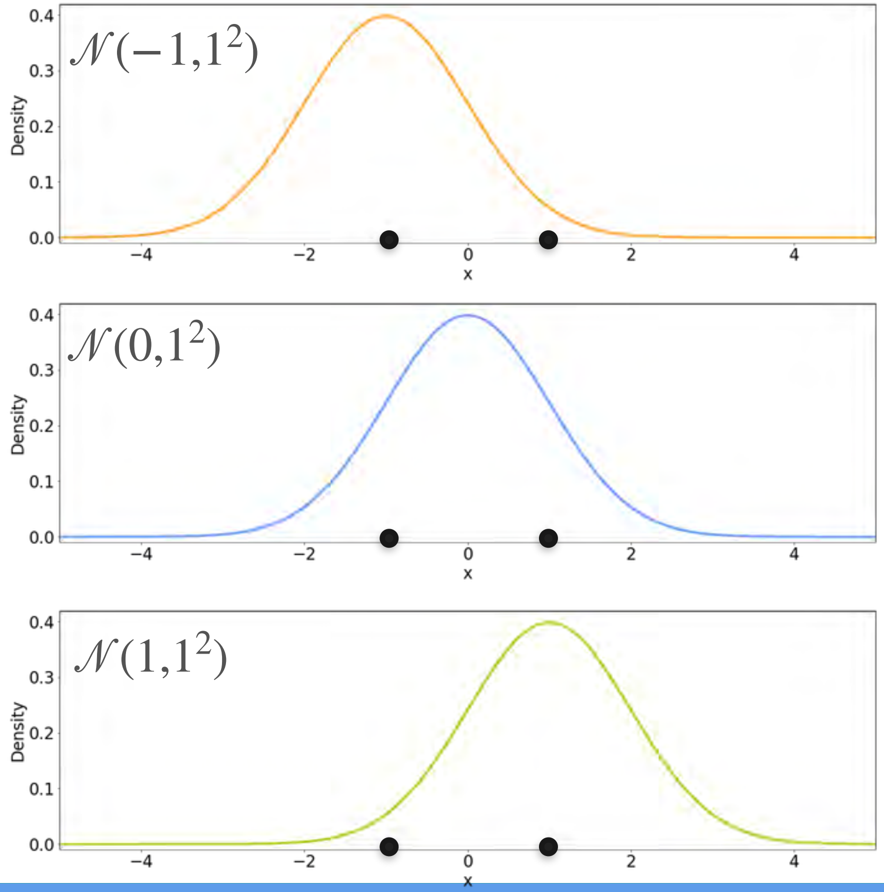


Observations



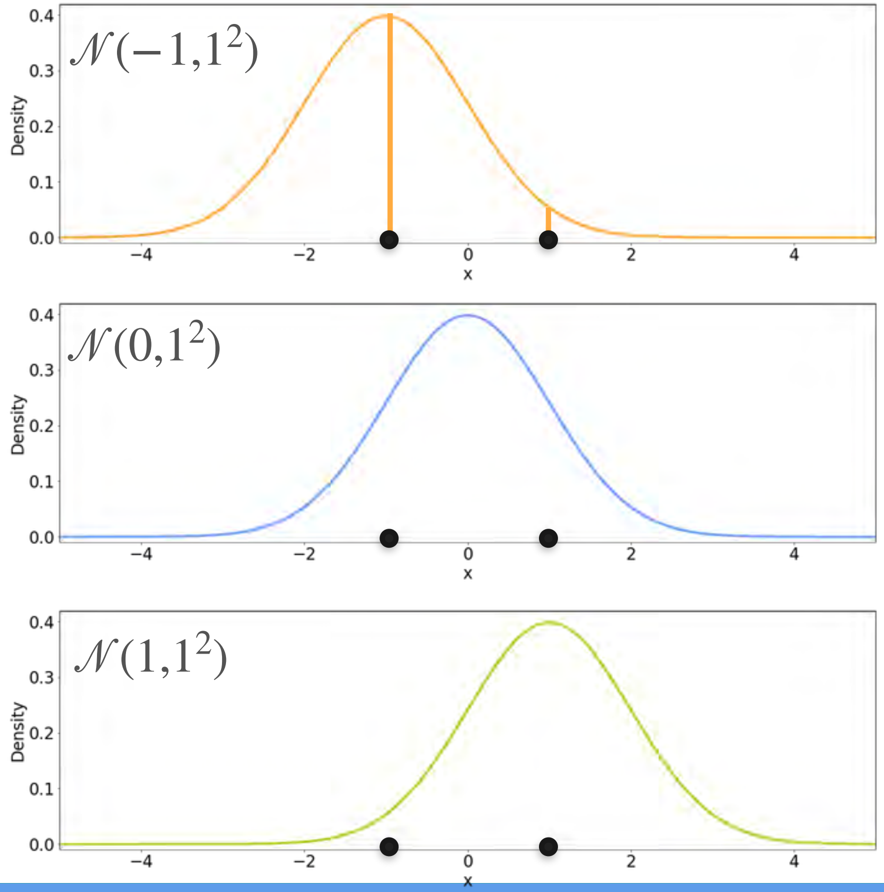
Candidates

Observations



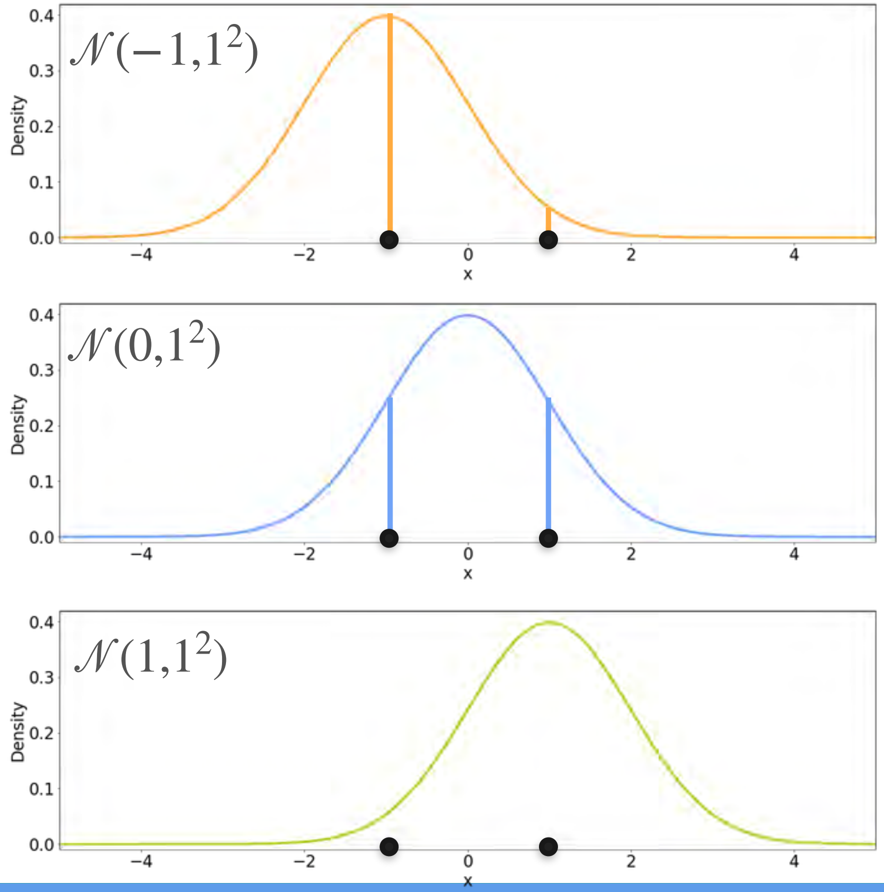
Candidates

Observations



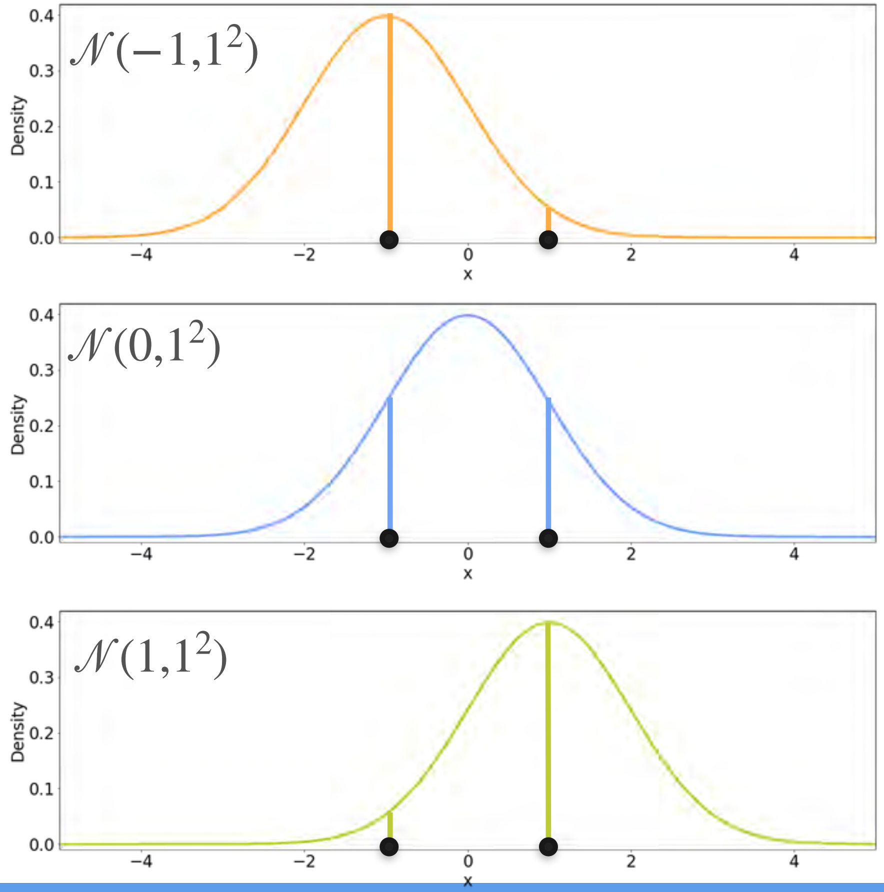
Candidates

Observations



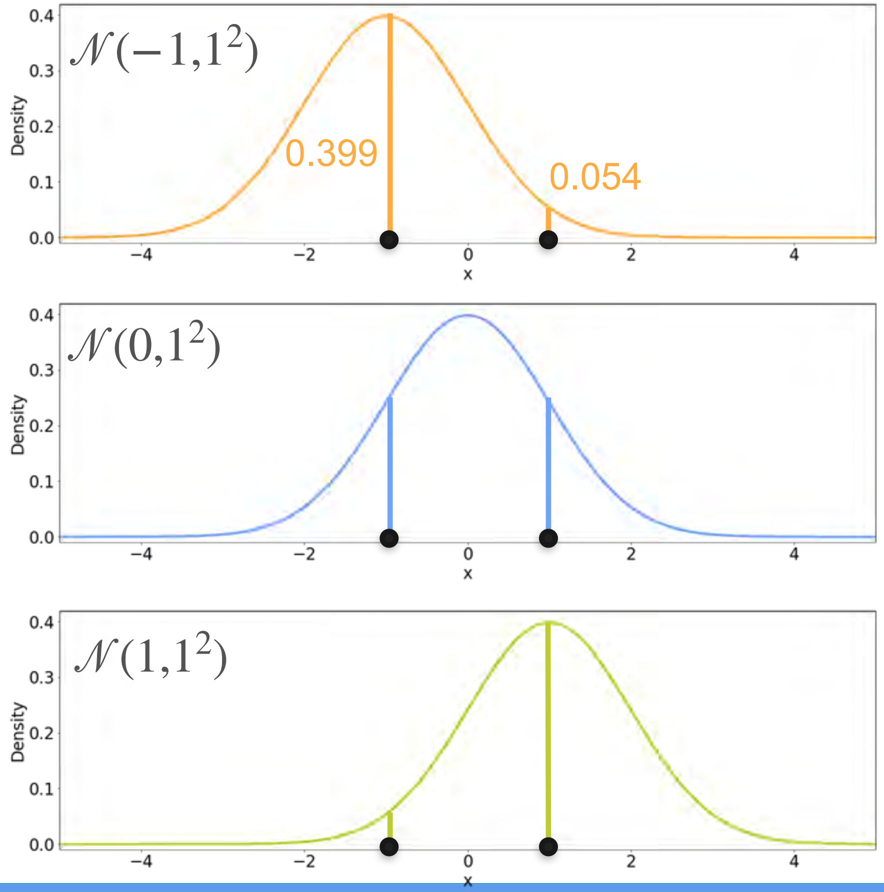
Candidates

Observations



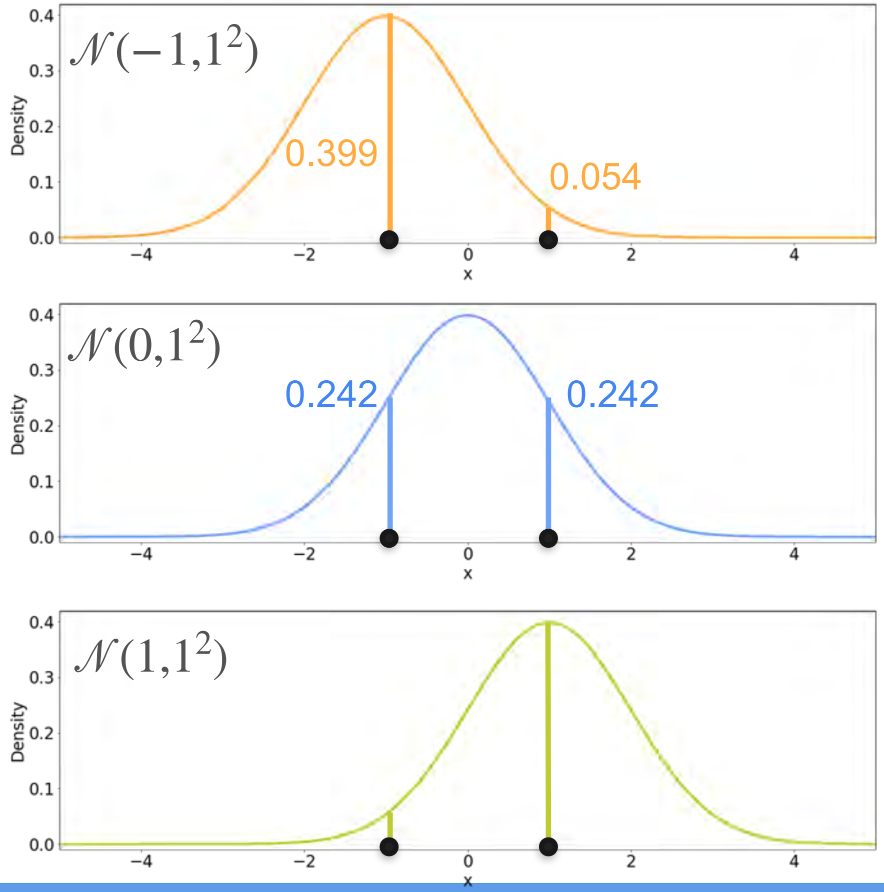
Candidates

Observations



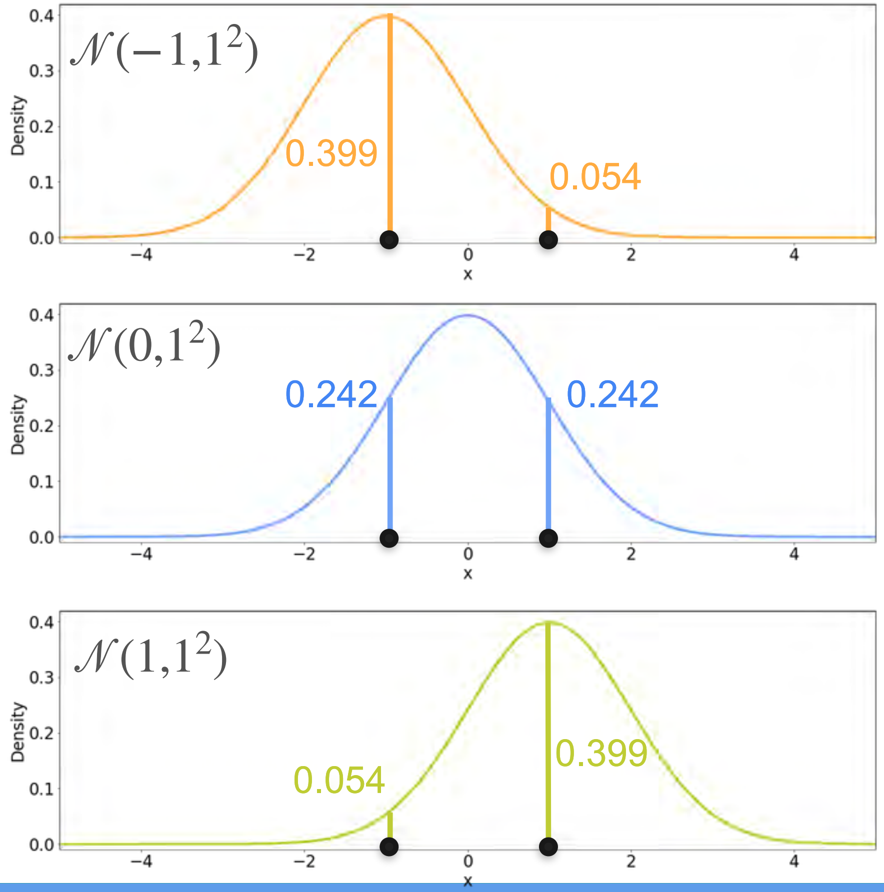
Candidates

Observations

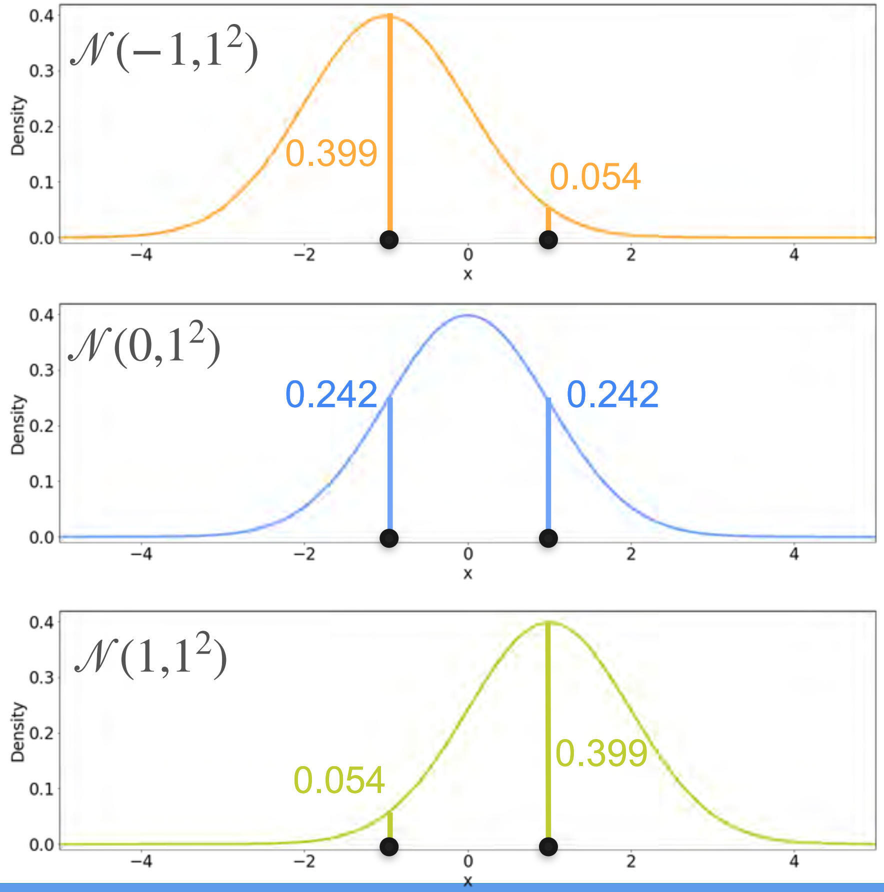


Candidates

Observations



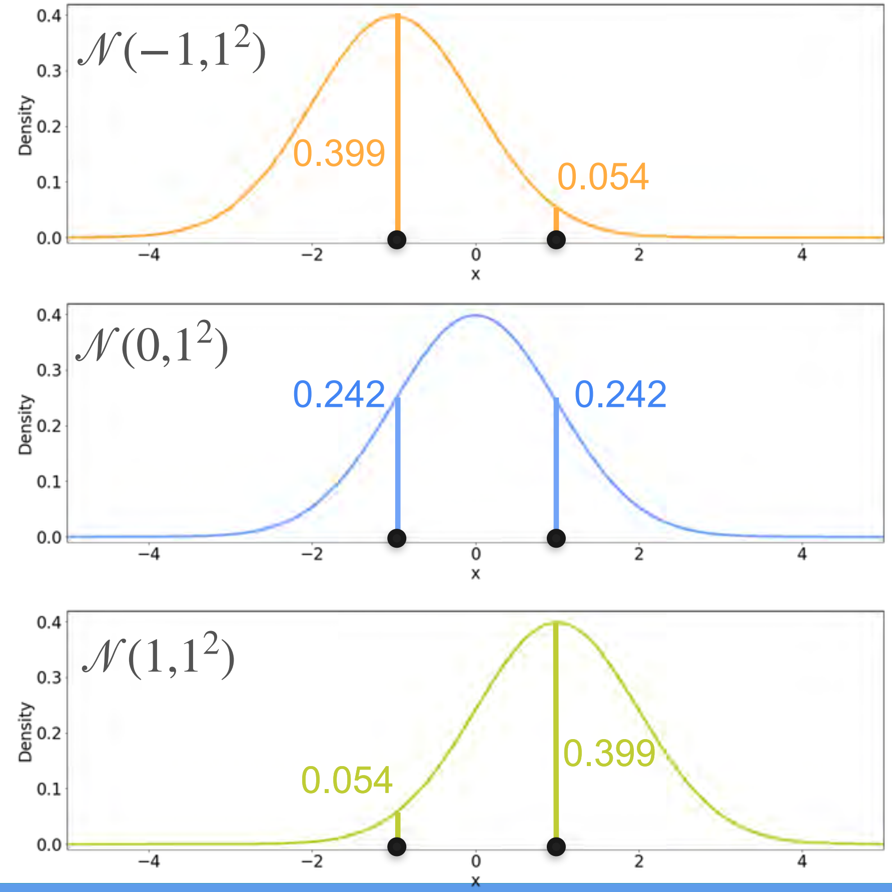
Candidates



Observations

$0.399 \cdot 0.054 = 0.022$

Candidates



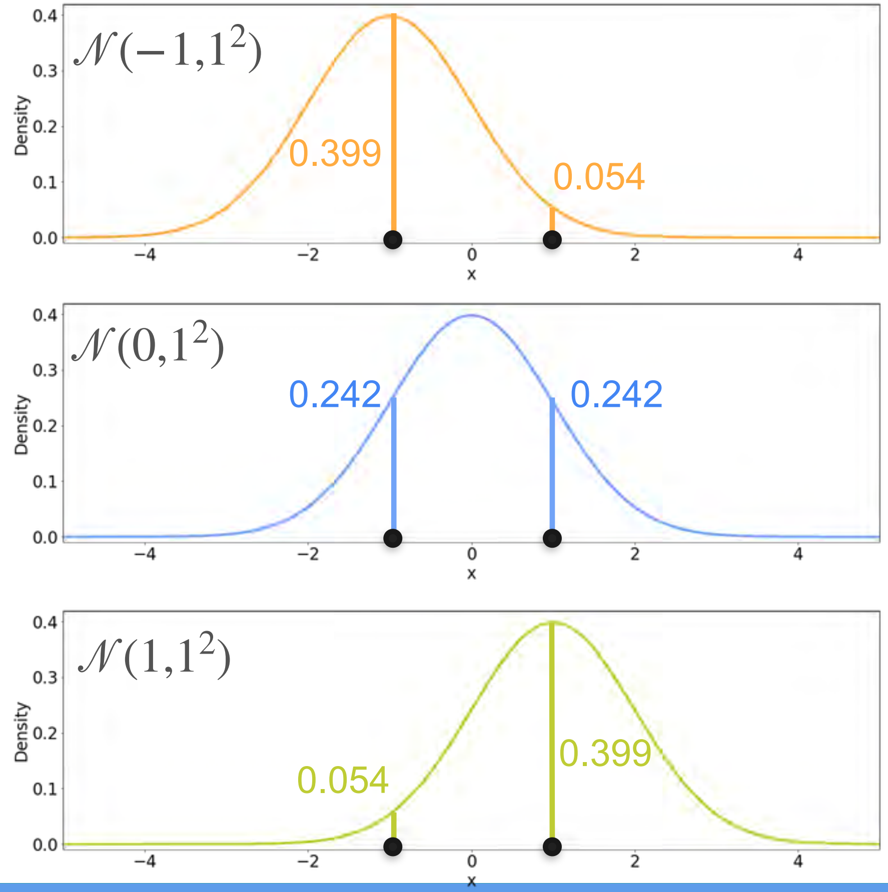
Observations

$0.399 \cdot 0.054 = 0.022$

$0.242 \cdot 0.242 = 0.059$

Candidates

Observations



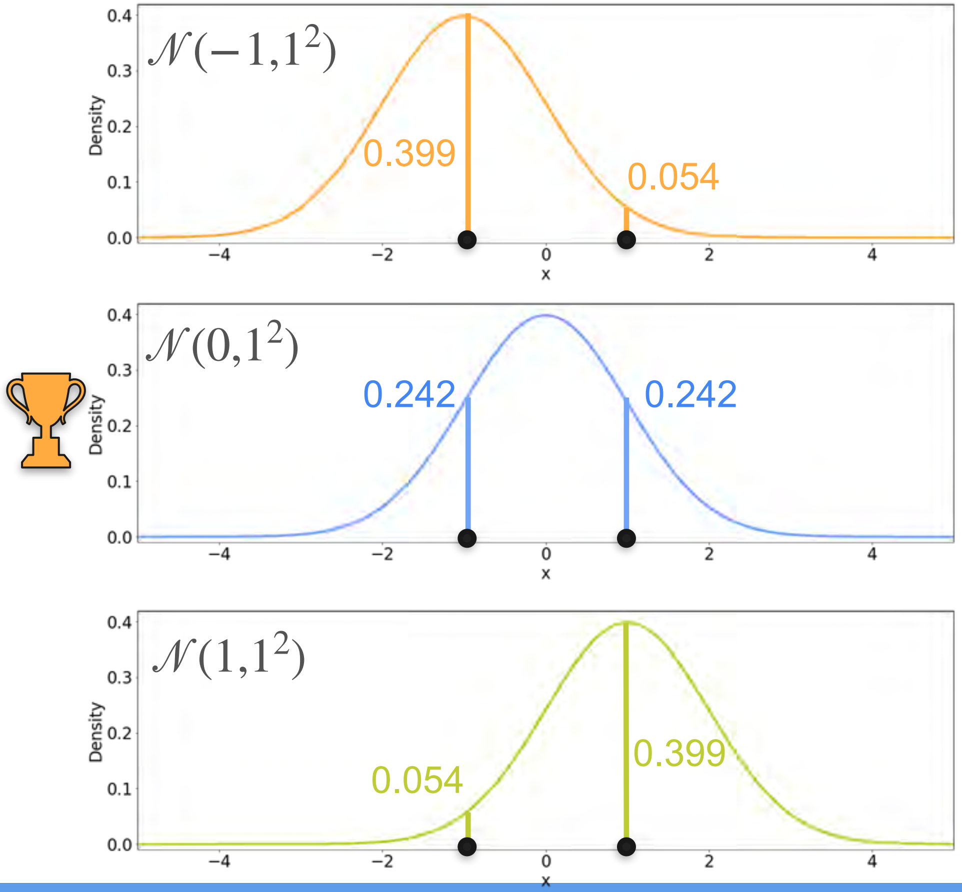
$0.399 \cdot 0.054 = 0.022$

$0.242 \cdot 0.242 = 0.059$

$0.054 \cdot 0.399 = 0.022$

Candidates

Observations

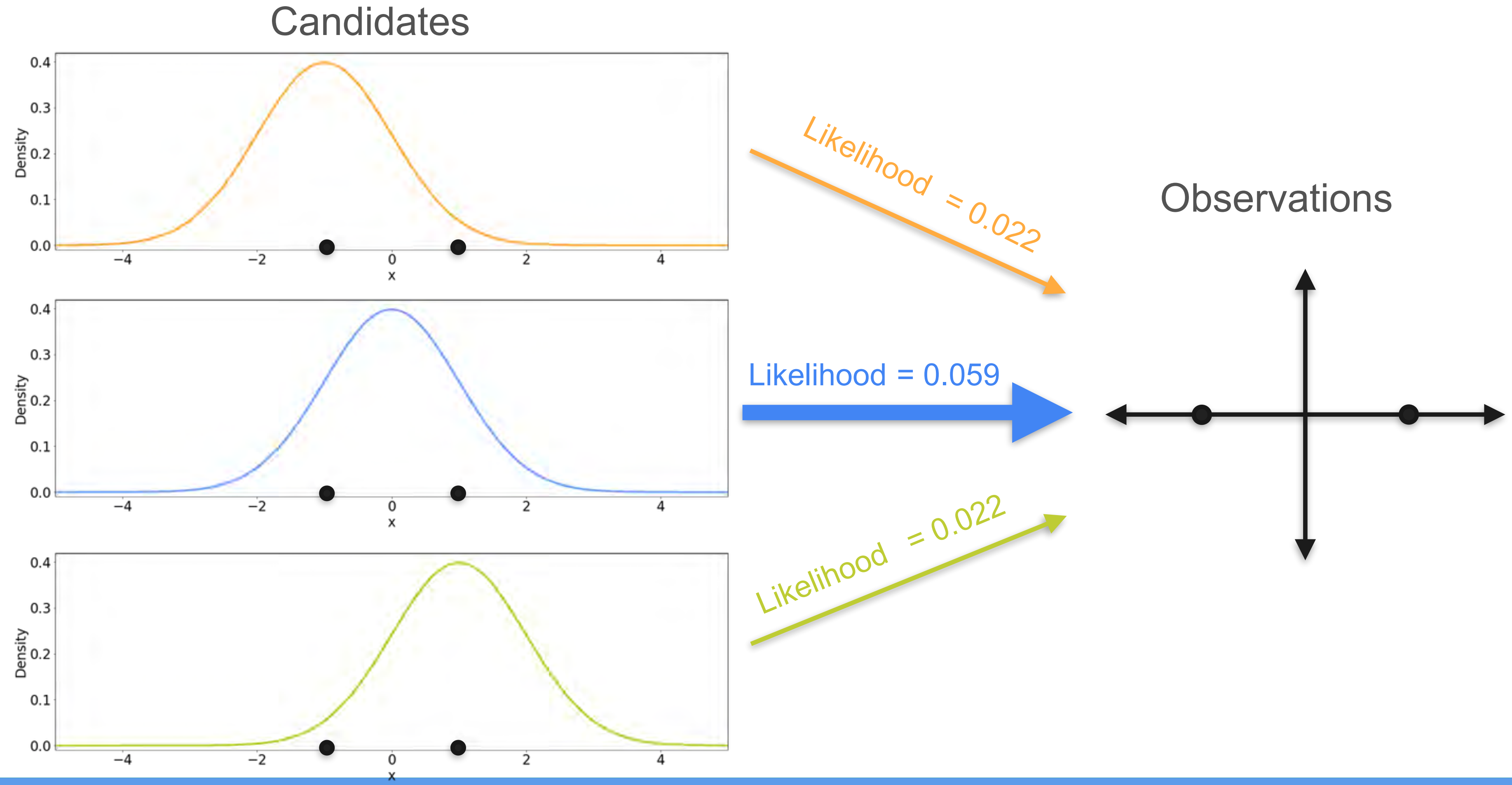


$0.399 \cdot 0.054 = 0.022$

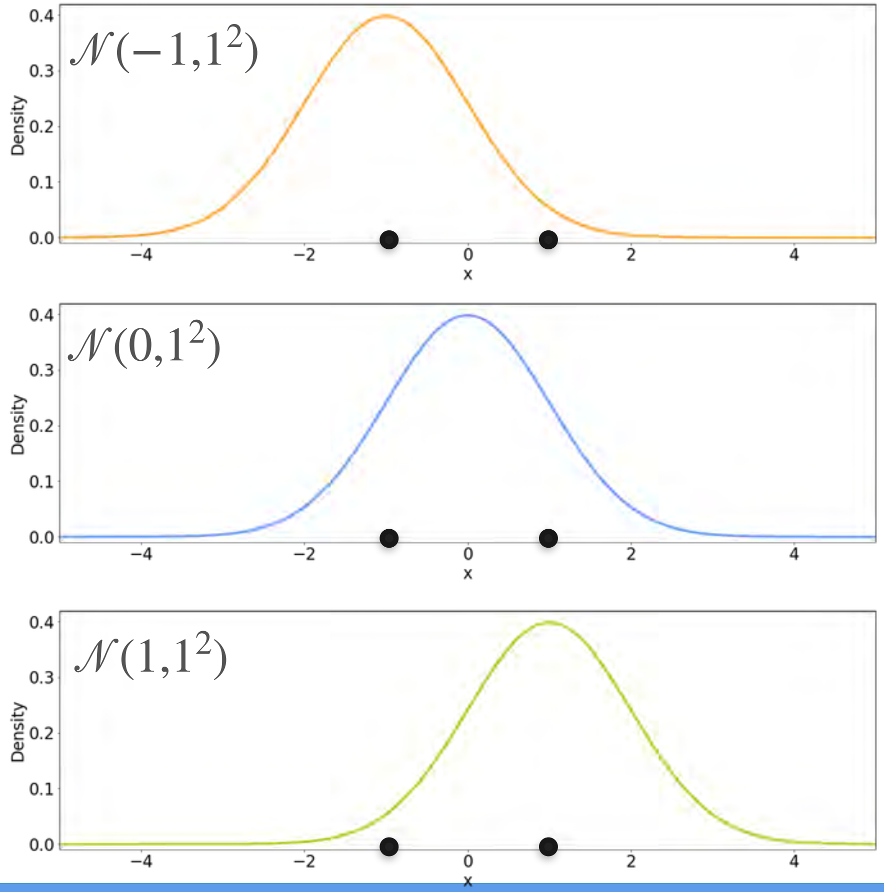
The $\mathcal{N}(0,1)$ is more likely!

$0.242 \cdot 0.242 = 0.059$

$0.054 \cdot 0.399 = 0.022$

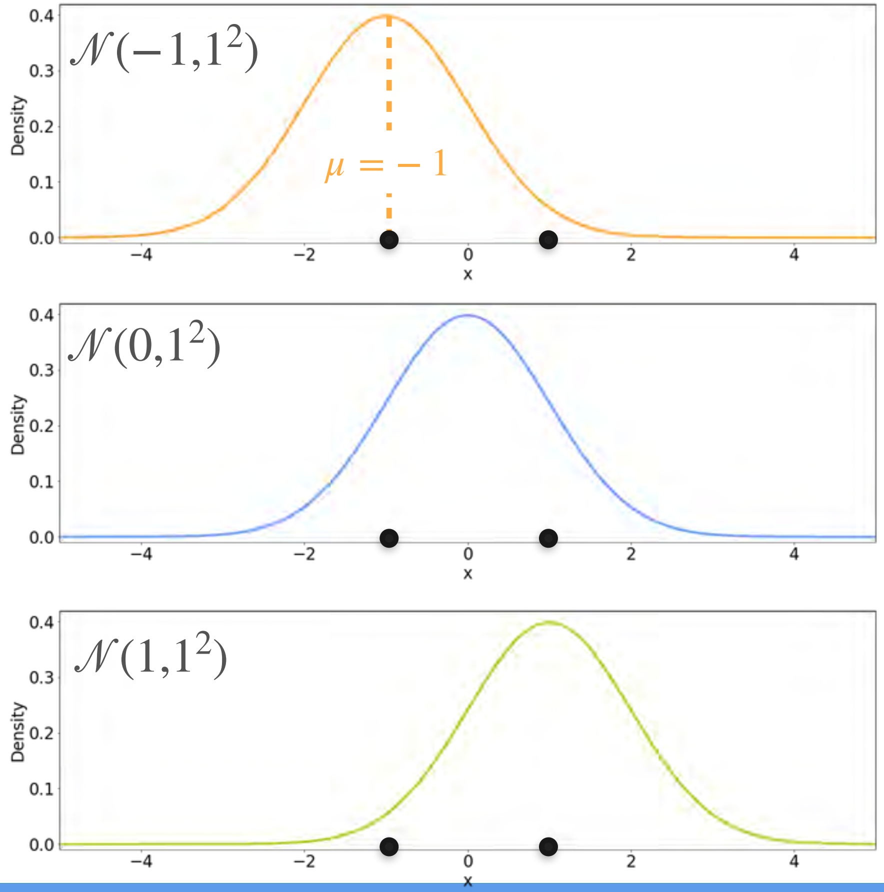

Candidates
Density
0.4
0.3
0.2
0.1
0.0
−4
−2
0
2
4
x
Likelihood = 0.022
Likelihood = 0.059
Likelihood = 0.022
Observations

Candidates



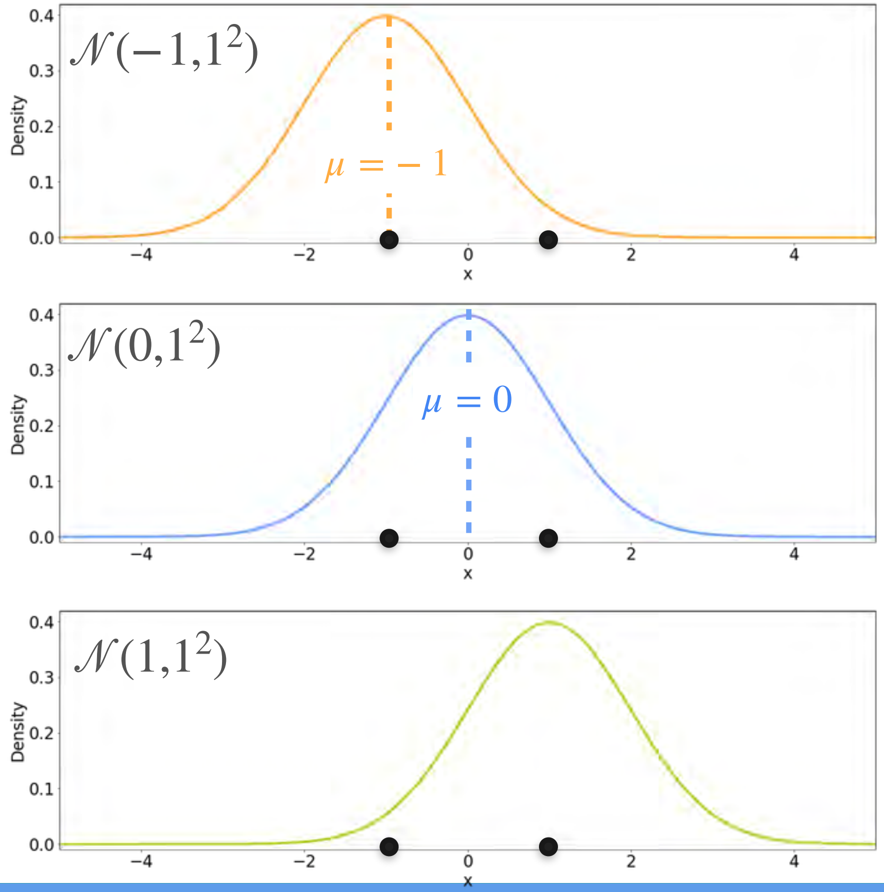
$\mathcal{N}(-1,1^2)$
$\mathcal{N}(0,1^2)$
$\mathcal{N}(1,1^2)$
Density
0.4
0.3
0.2
0.1
0.0
−4
−2
0
2
4
x

Candidates



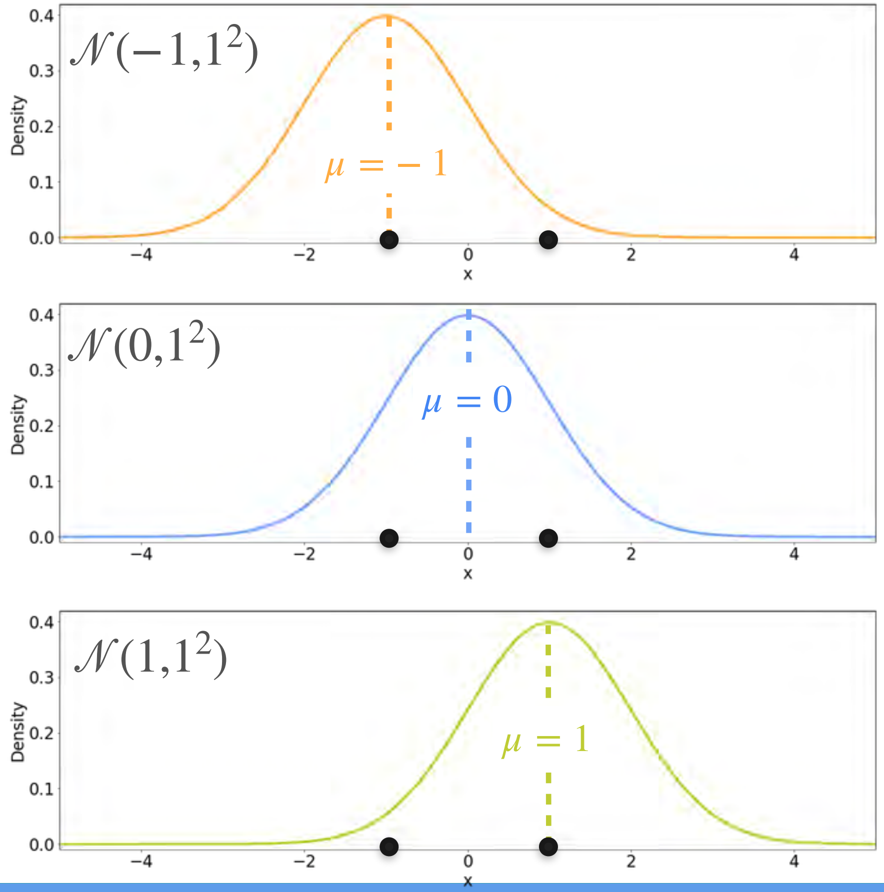
$\mathcal{N}(-1,1^2)$
$\mu = -1$
$\mathcal{N}(0,1^2)$
$\mathcal{N}(1,1^2)$
Density
x
0.4
0.3
0.2
0.1
0.0
−4
−2
0
2
4

Candidates



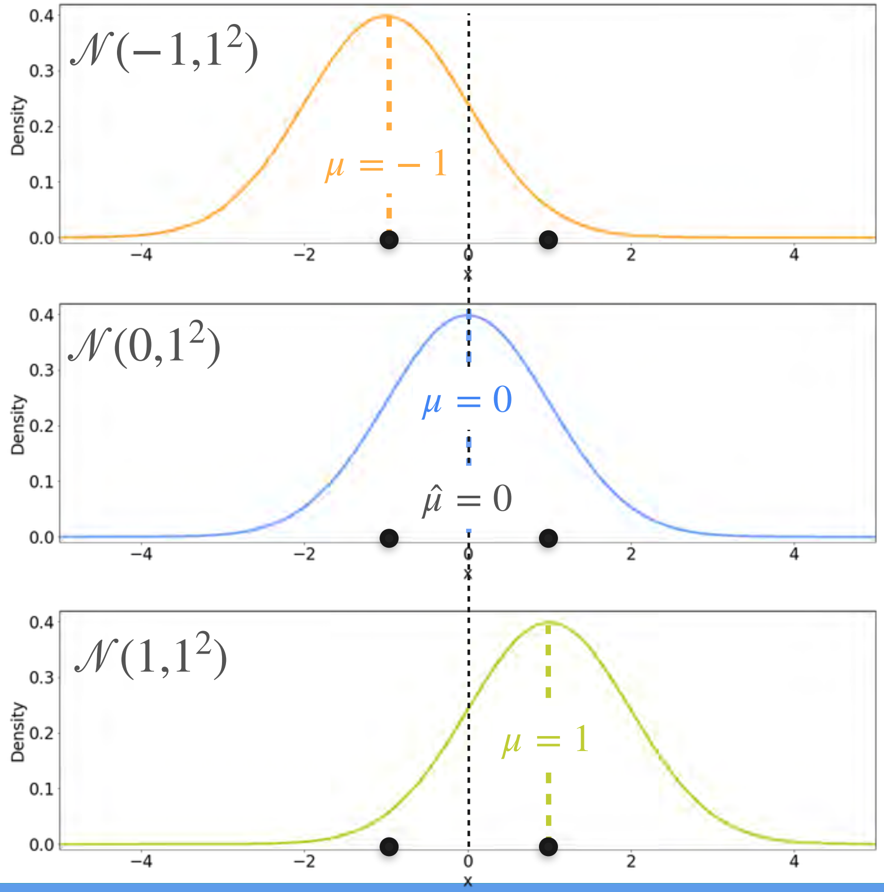
$\mathcal{N}(-1,1^2)$
$\mu = -1$
$\mathcal{N}(0,1^2)$
$\mu = 0$
$\mathcal{N}(1,1^2)$
Density
x
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4

Candidates



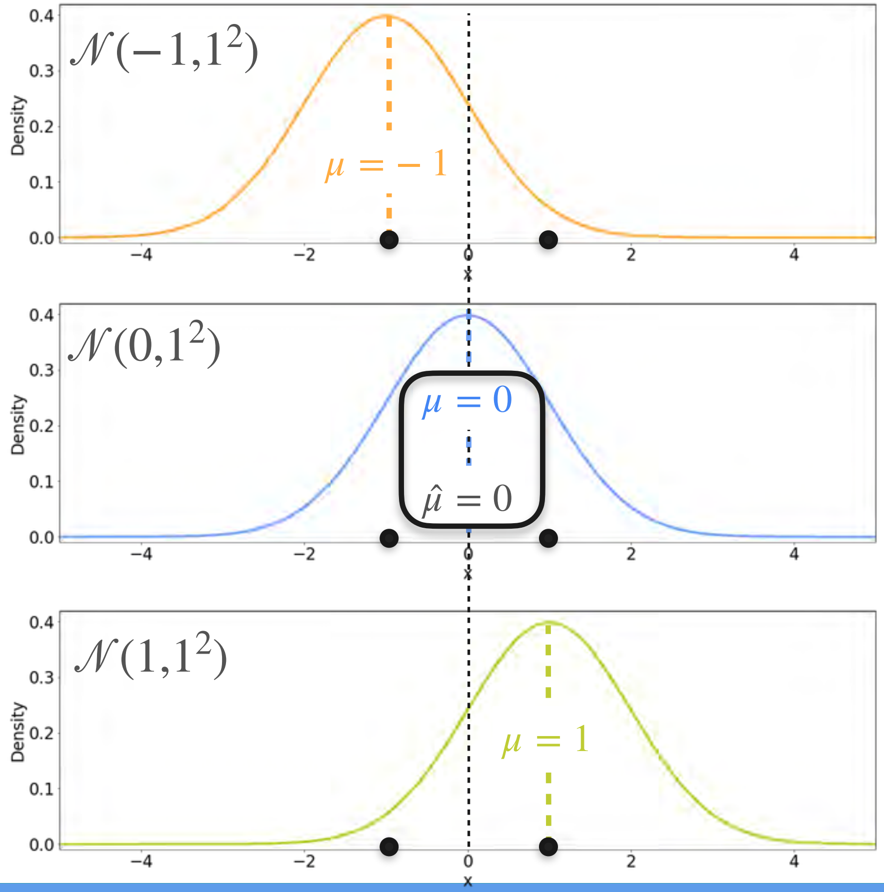
$\mathcal{N}(-1,1^2)$
$\mu = -1$
Density
x
$\mathcal{N}(0,1^2)$
$\mu = 0$
Density
x
$\mathcal{N}(1,1^2)$
$\mu = 1$
Density
x

Candidates

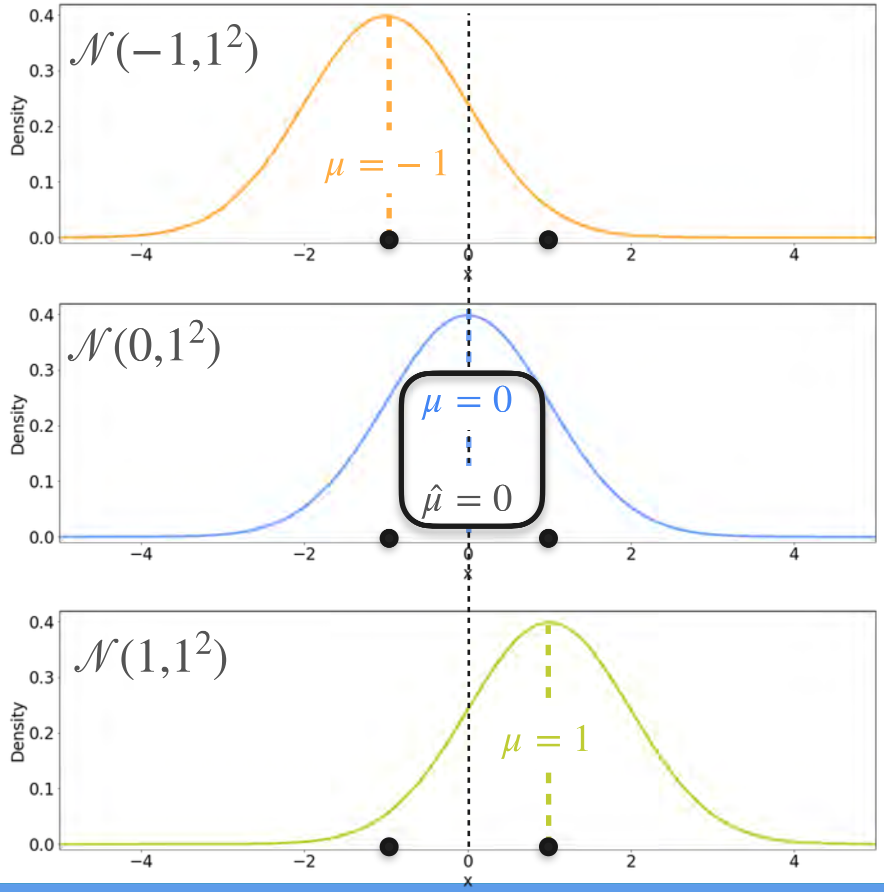


$\mathcal{N}(-1,1^2)$
$\mu = -1$
$\mathcal{N}(0,1^2)$
$\mu = 0$
$\hat{\mu} = 0$
$\mathcal{N}(1,1^2)$
$\mu = 1$
Density
x
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4

Candidates


$\mathcal{N}(-1,1^2)$
$\mu = -1$
$\mathcal{N}(0,1^2)$
$\mu = 0$
$\hat{\mu} = 0$
$\mathcal{N}(1,1^2)$
$\mu = 1$
Density
x

Candidates

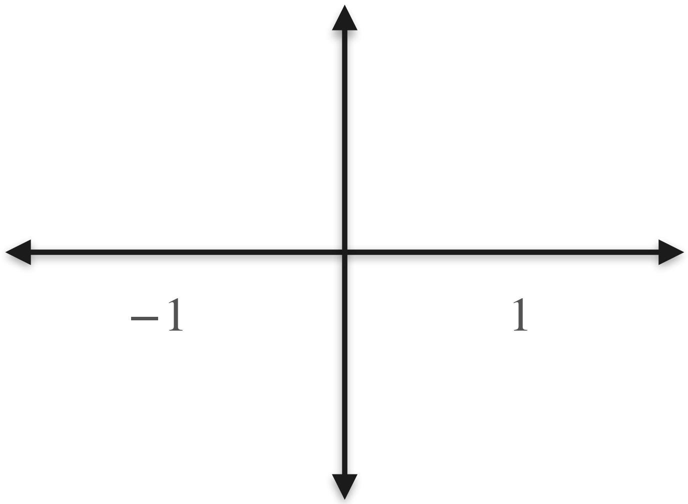


The best distribution is the one where the **mean** of the distribution is the **mean** of the sample

Gaussians With Three Different Variance

Candidates

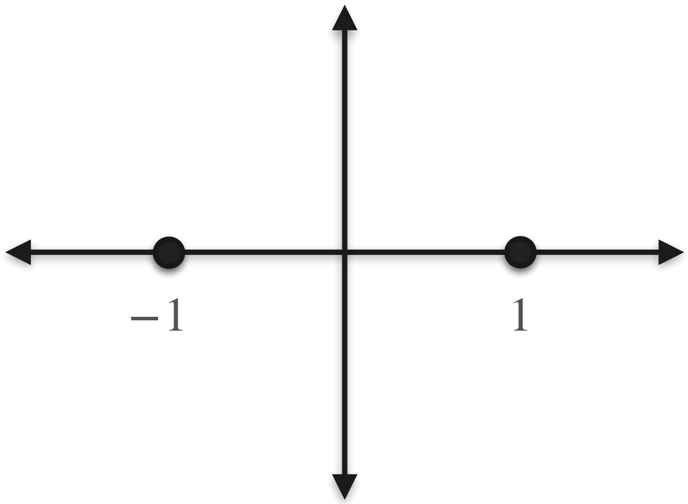
Observations



Gaussians With Three Different Variance

Candidates

Observations

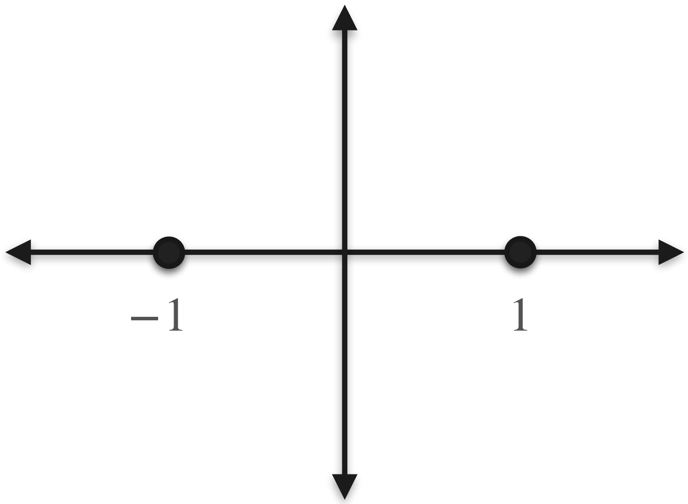


Gaussians With Three Different Variance

Candidates

$\mathcal{N}(0, 0.5^2)$

Observations



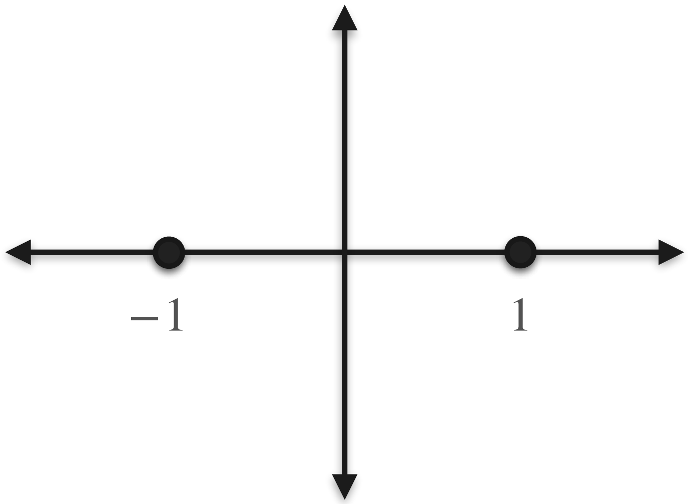
Gaussians With Three Different Variance

Candidates

$\mathcal{N}(0, 0.5^2)$

$\mathcal{N}(0, 1^2)$

Observations



Gaussians With Three Different Variance

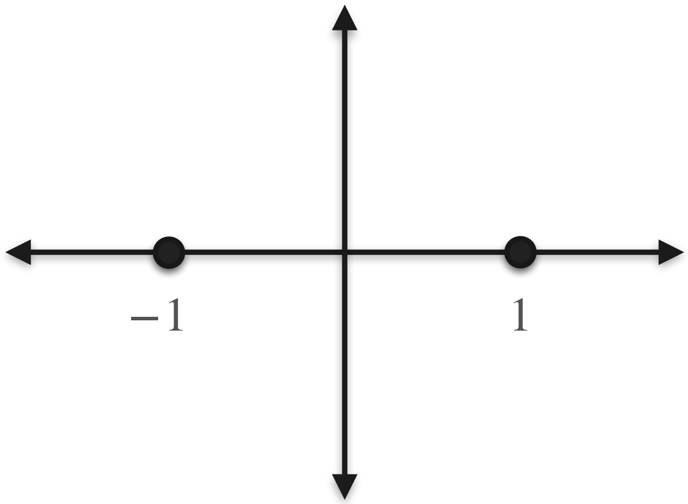
Candidates

$\mathcal{N}(0, 0.5^2)$

$\mathcal{N}(0, 1^2)$

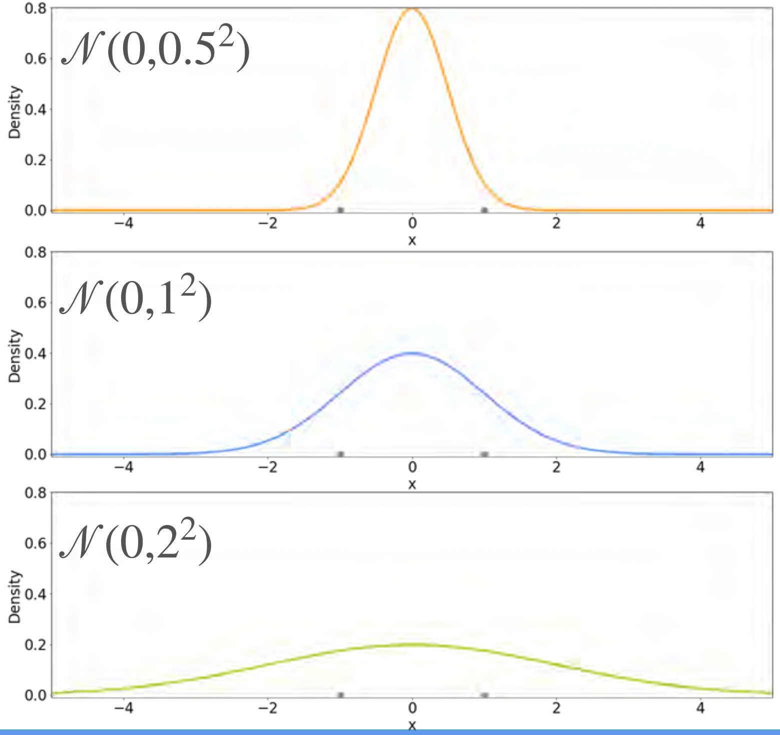
$\mathcal{N}(0, 2^2)$

Observations

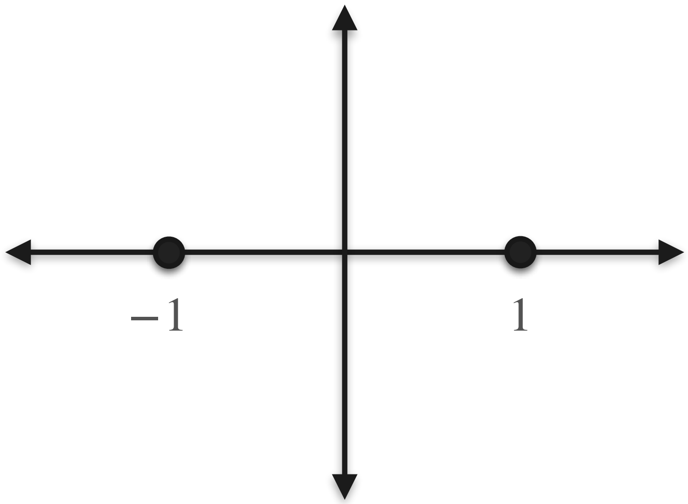


Gaussians With Three Different Variance

Candidates

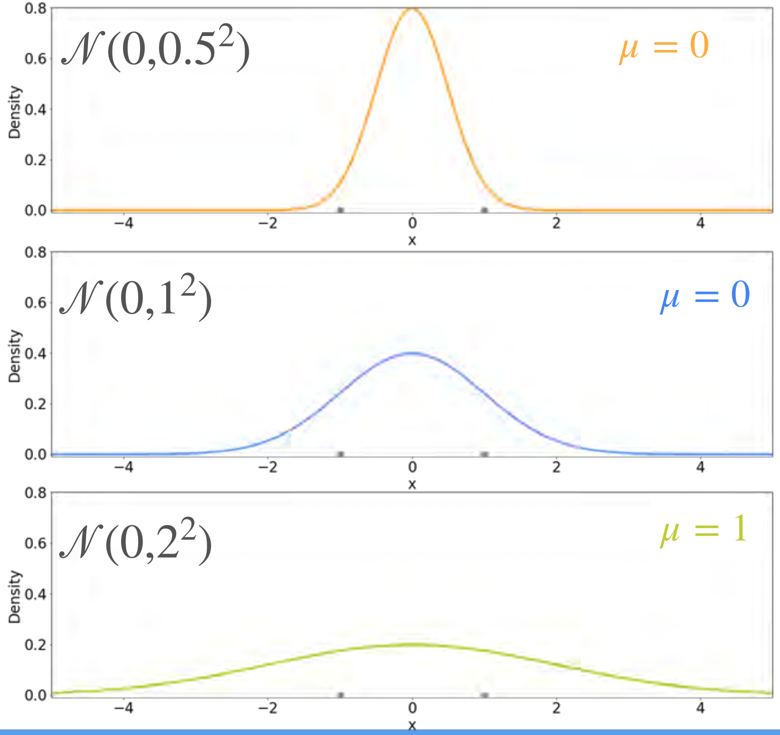


Observations

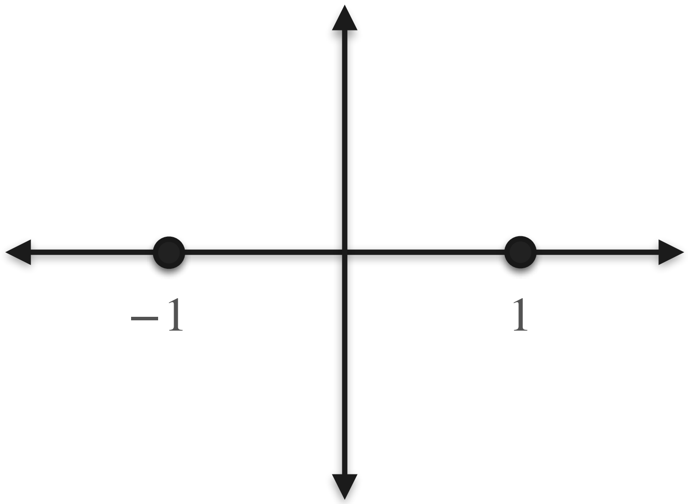


Gaussians With Three Different Variance

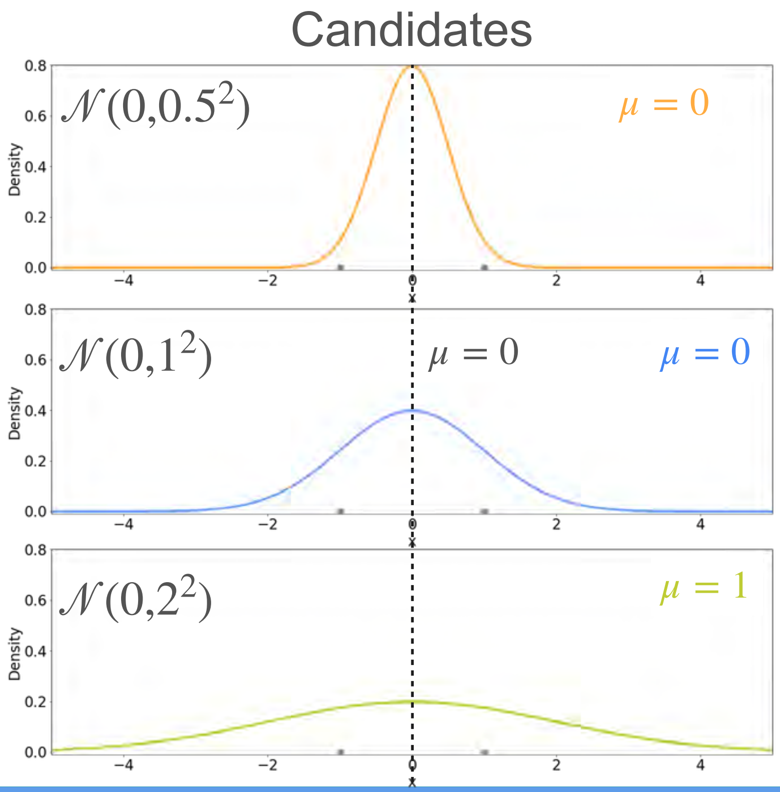
Candidates

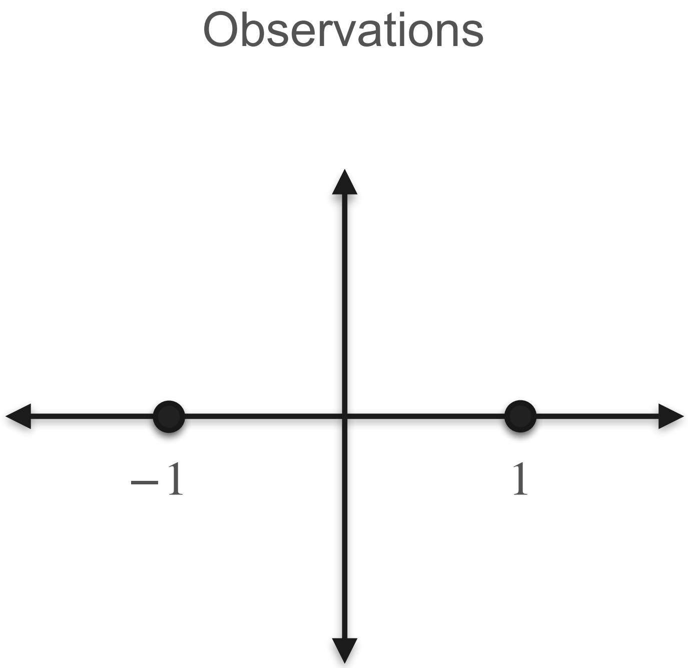


Observations



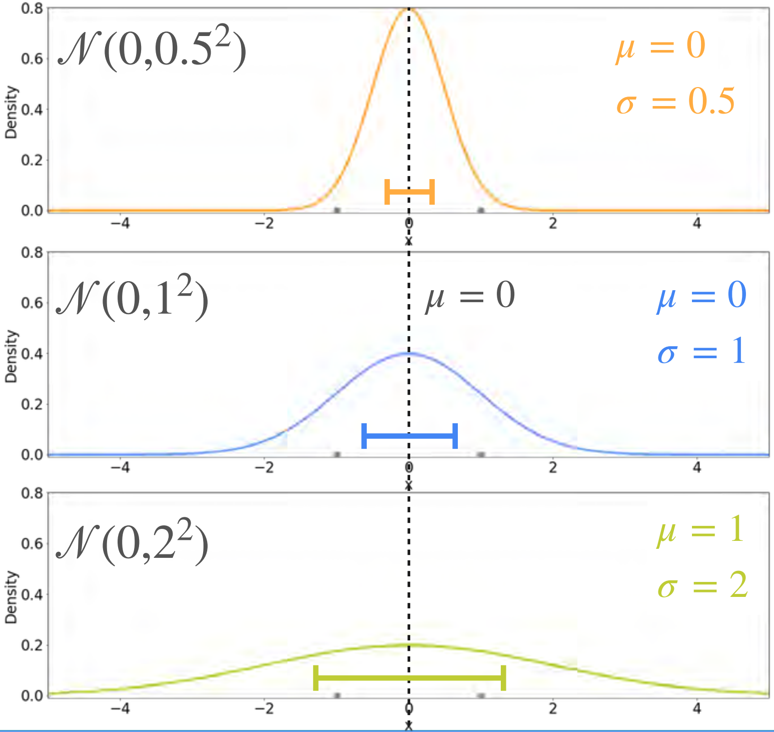
Gaussians With Three Different Variance



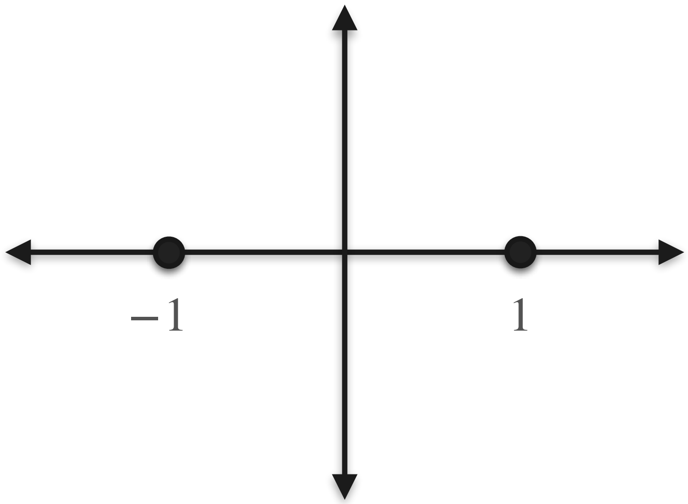


Gaussians With Three Different Variance

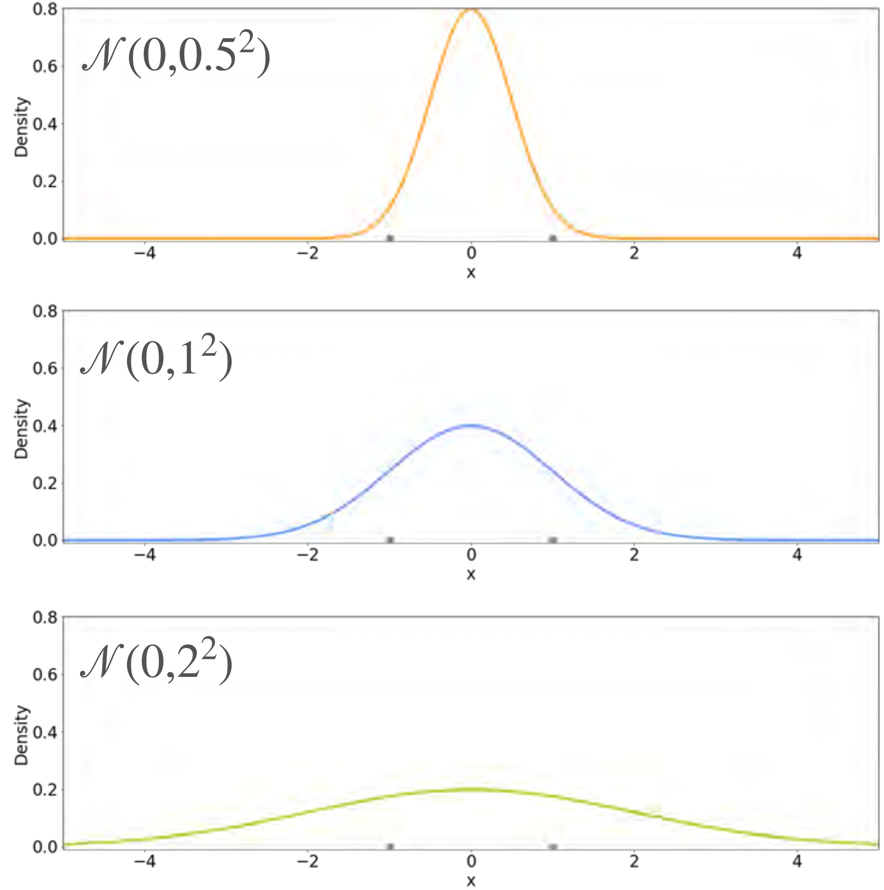
Candidates



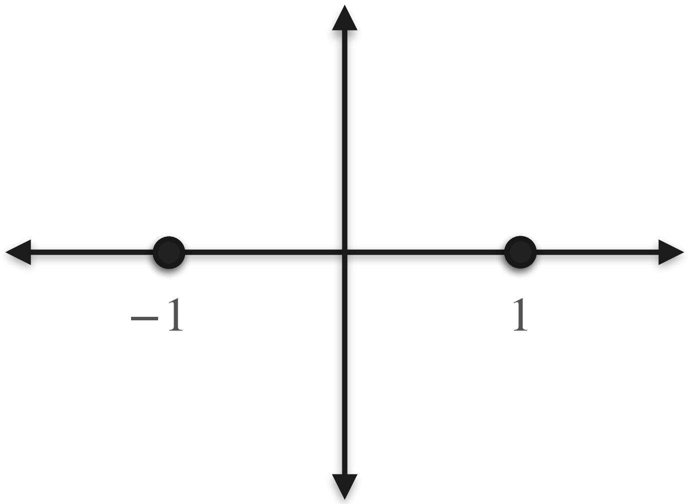
Observations



Candidates

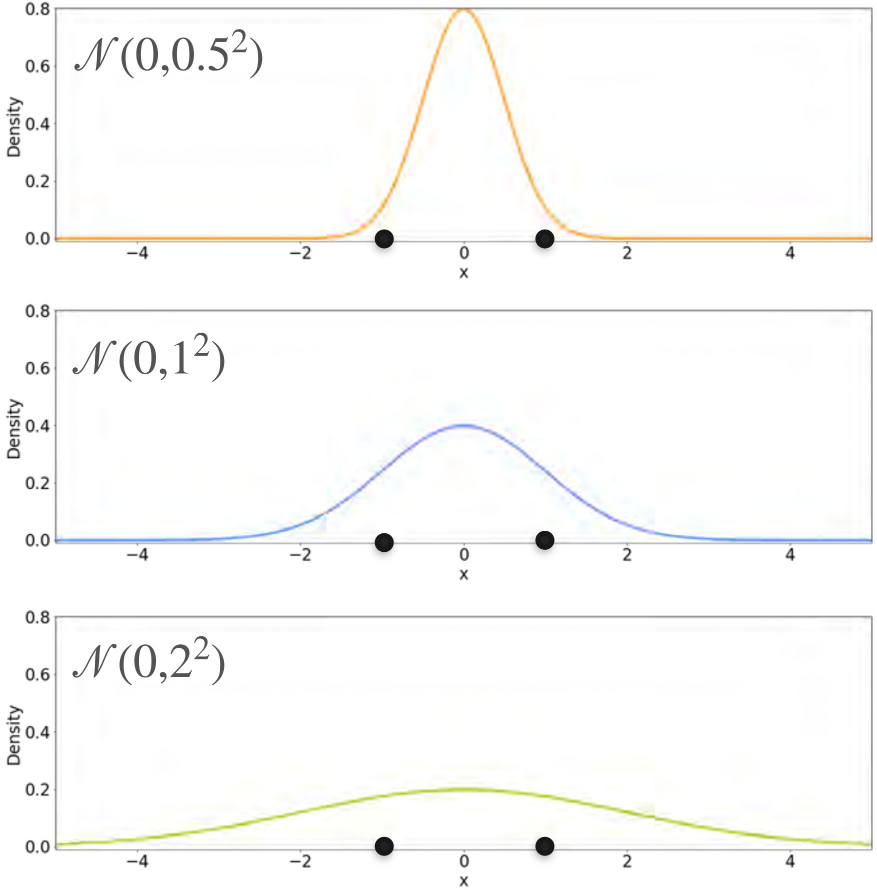


Observations

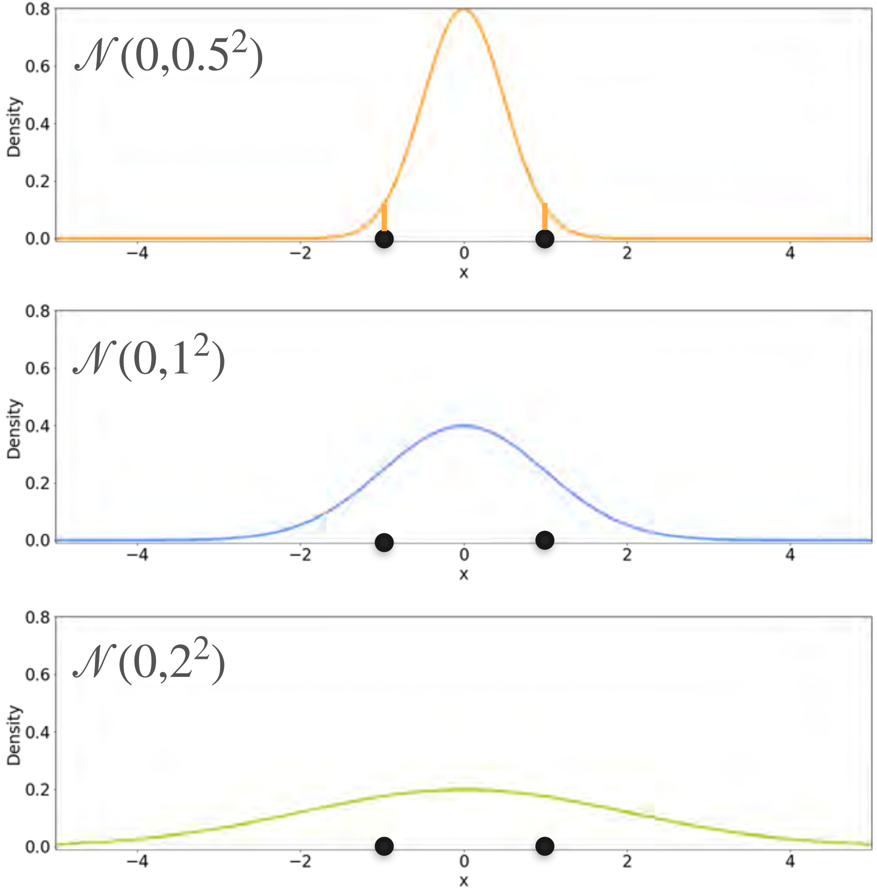


Candidates

Observations



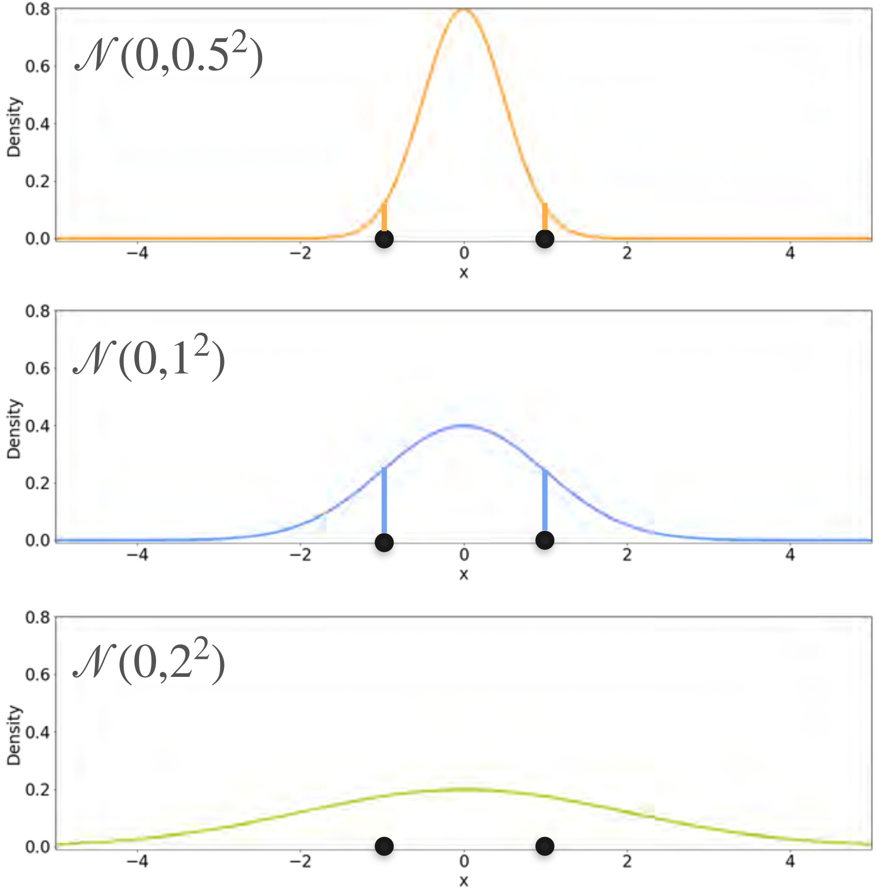
Candidates



Observations

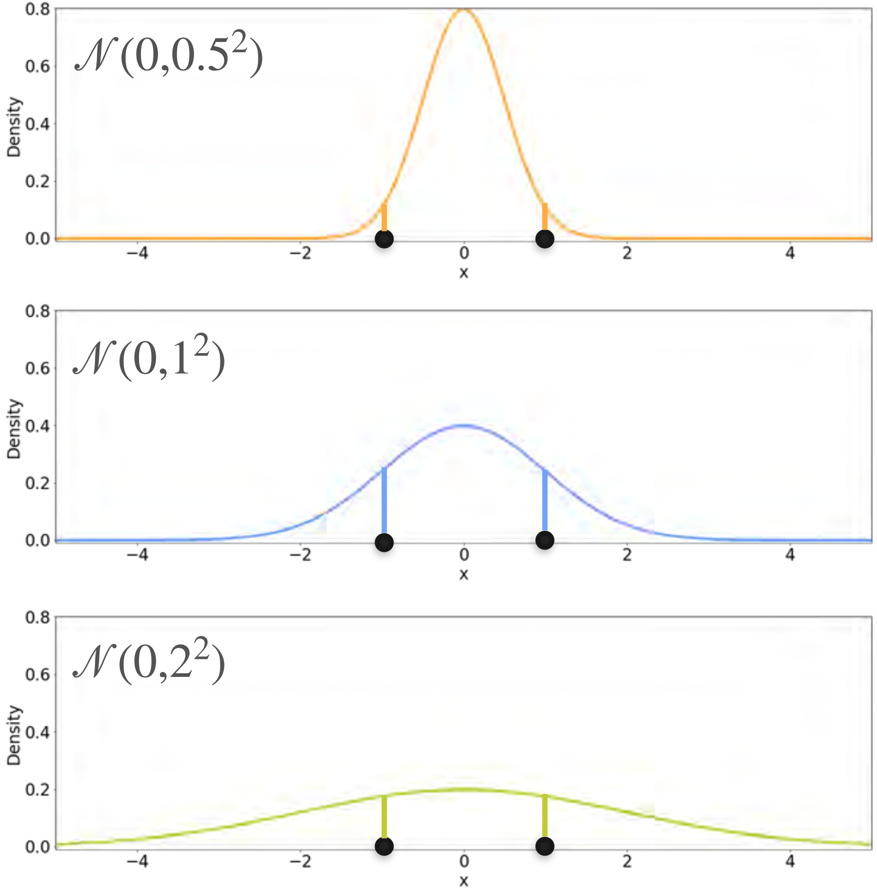
Candidates

Observations

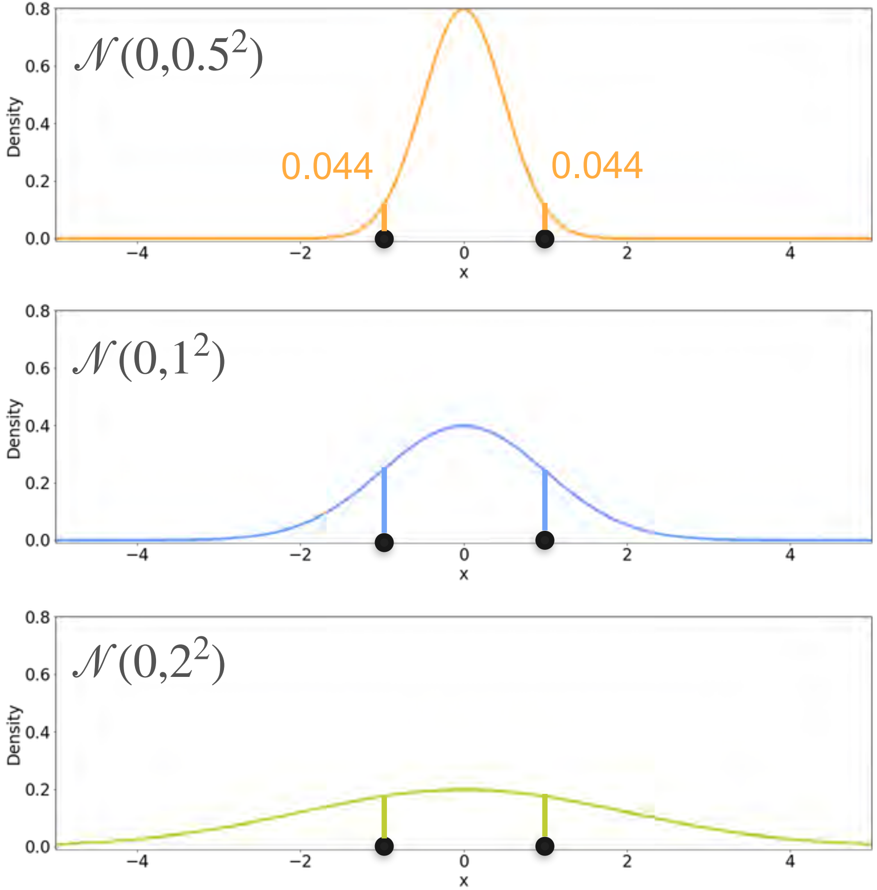


Candidates

Observations



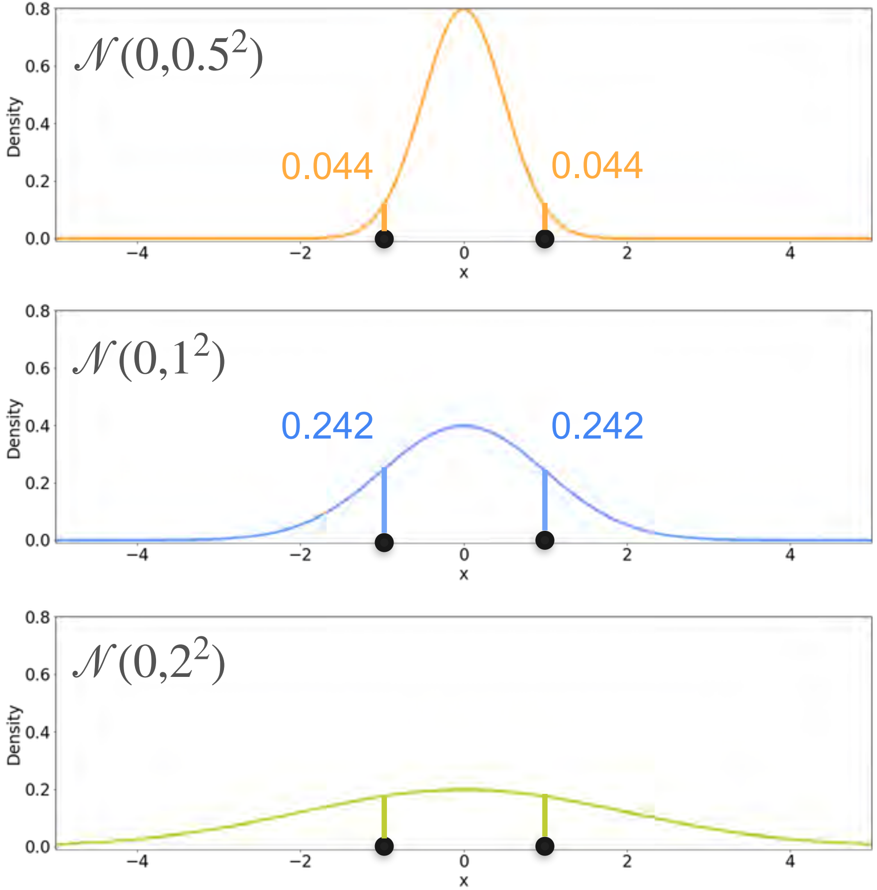
Candidates



Observations

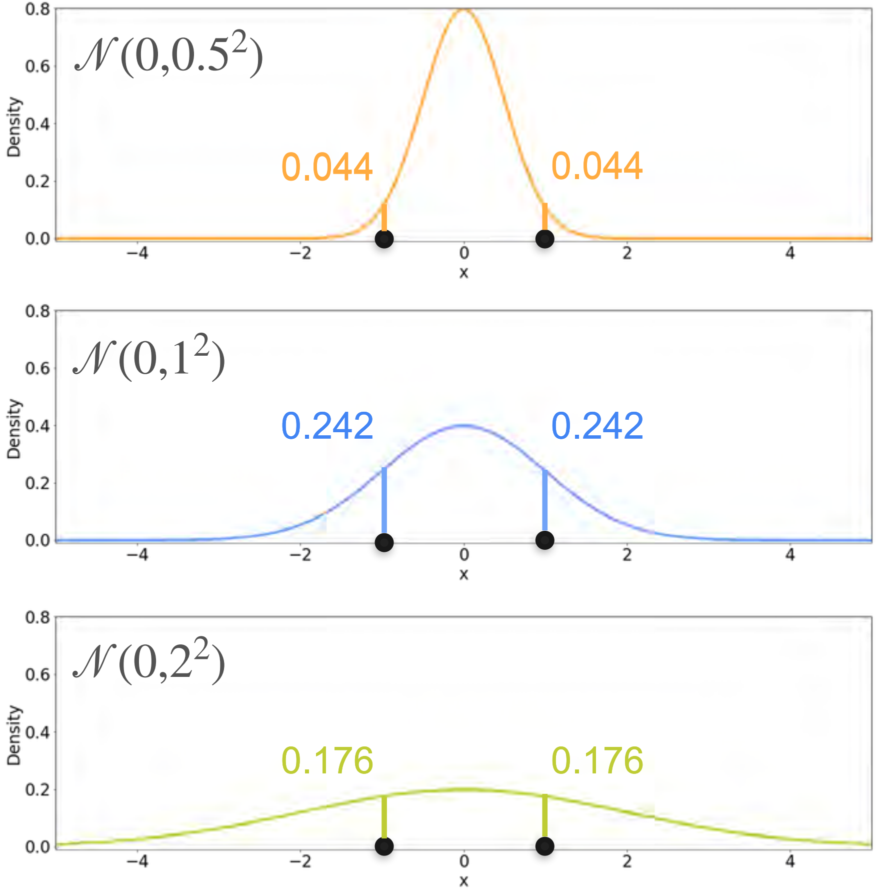
Candidates

Observations

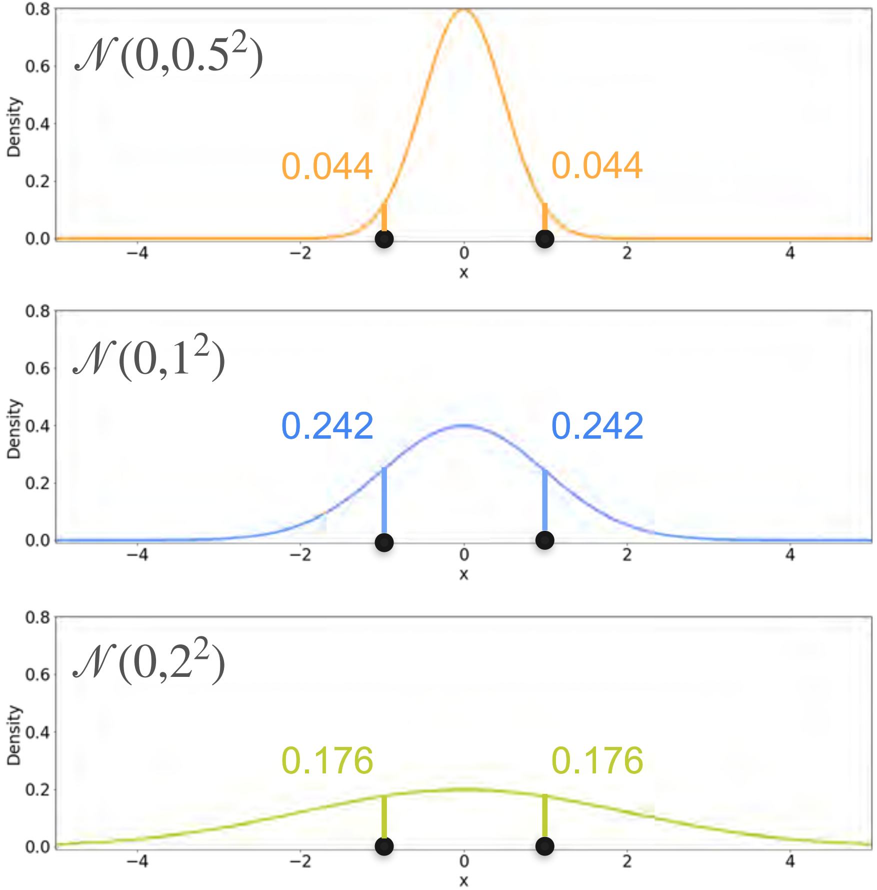


Candidates

Observations



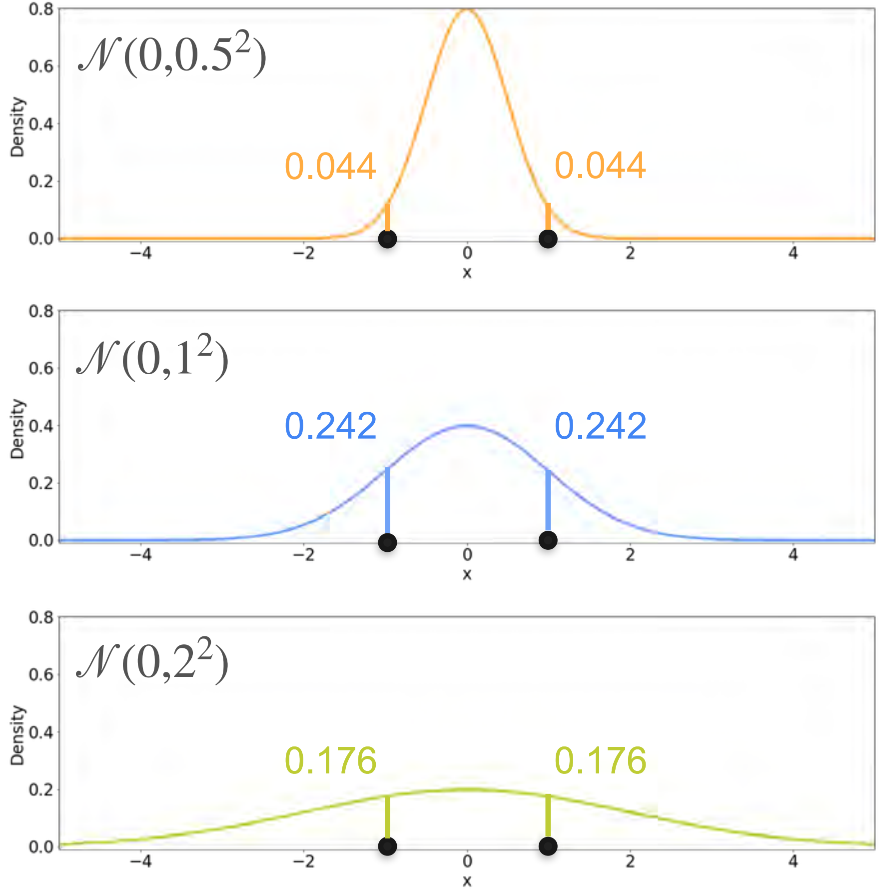
Candidates



Observations

$0.044 \cdot 0.044 = 0.002$

Candidates

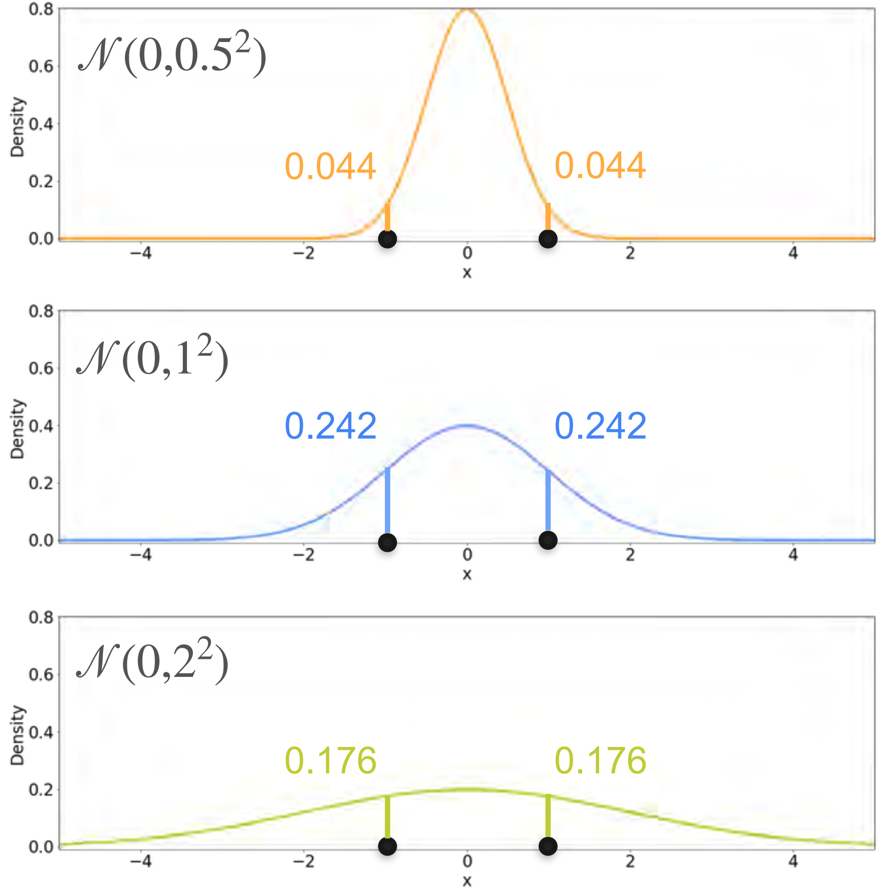


Observations

$0.044 \cdot 0.044 = 0.002$

$0.242 \cdot 0.242 = 0.059$

Candidates



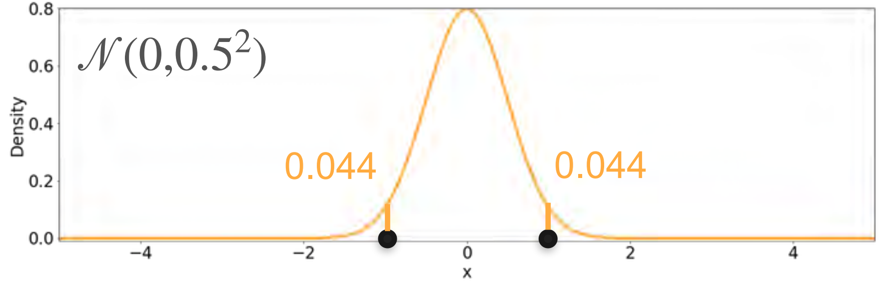
Observations

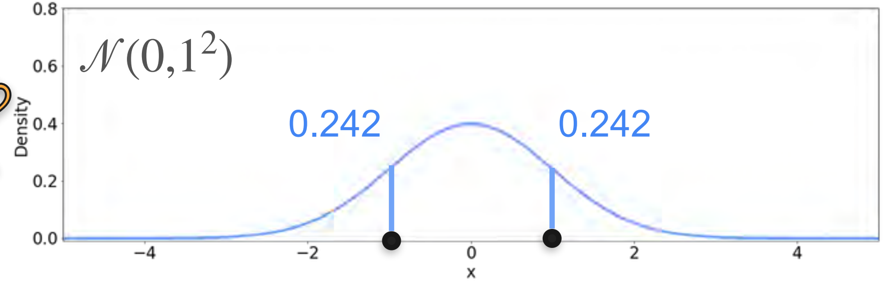
$0.044 \cdot 0.044 = 0.002$

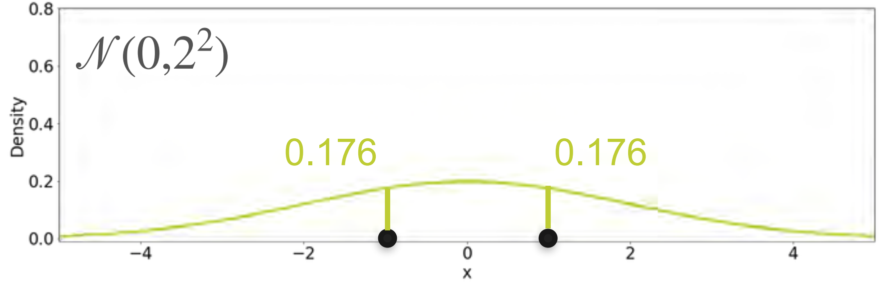
$0.242 \cdot 0.242 = 0.059$

$0.176 \cdot 0.176 = 0.031$

Candidates







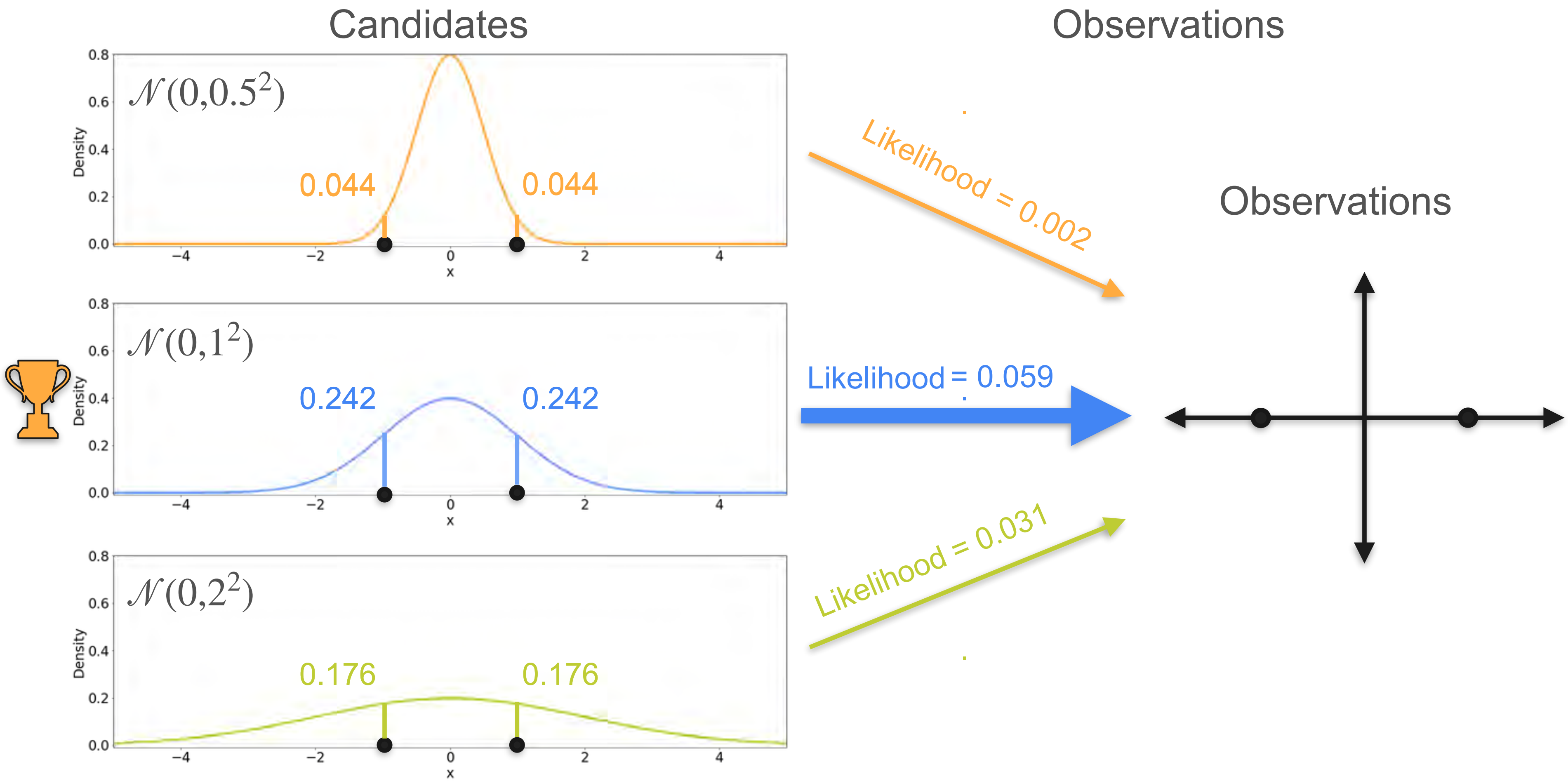
Observations

$0.044 \cdot 0.044 = 0.002$

The $\mathcal{N}(0,1^2)$ is more likely!

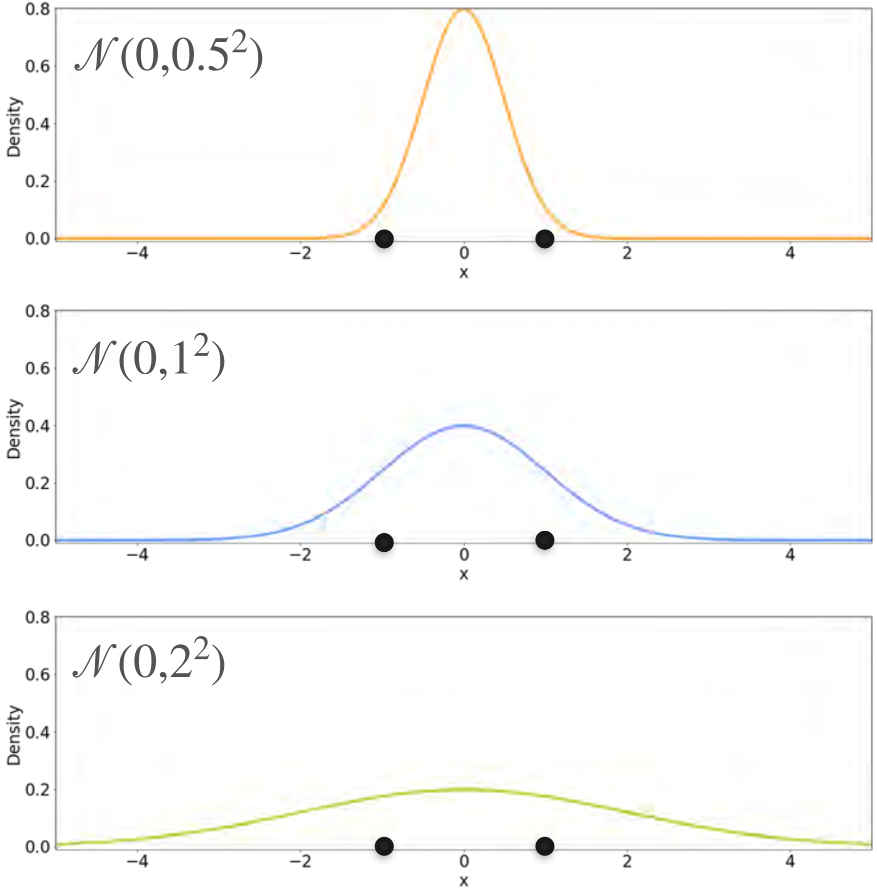
$0.242 \cdot 0.242 = 0.059$

$0.176 \cdot 0.176 = 0.031$

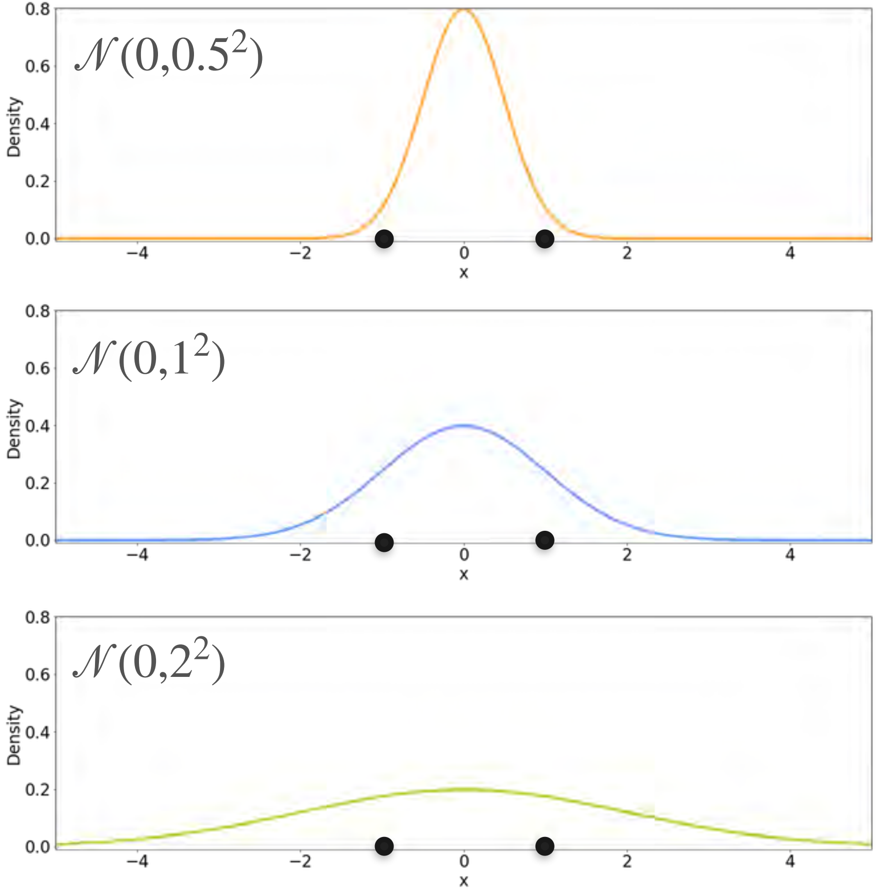

Candidates
Observations
$\mathcal{N}(0,0.5^2)$
0.044
0.044
Likelihood = 0.002
Observations
$\mathcal{N}(0,1^2)$
0.242
0.242
Likelihood = 0.059
$\mathcal{N}(0,2^2)$
0.176
0.176
Likelihood = 0.031
Density
x

Candidates

Observations



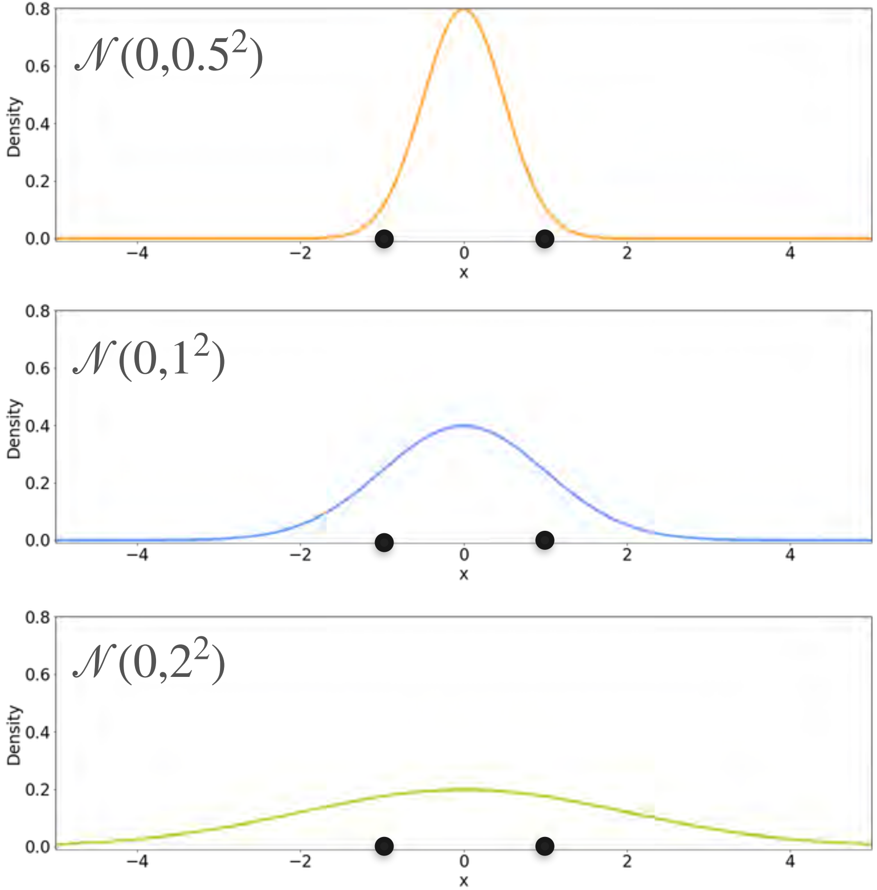
Candidates



Observations

Variance of the observations

Candidates

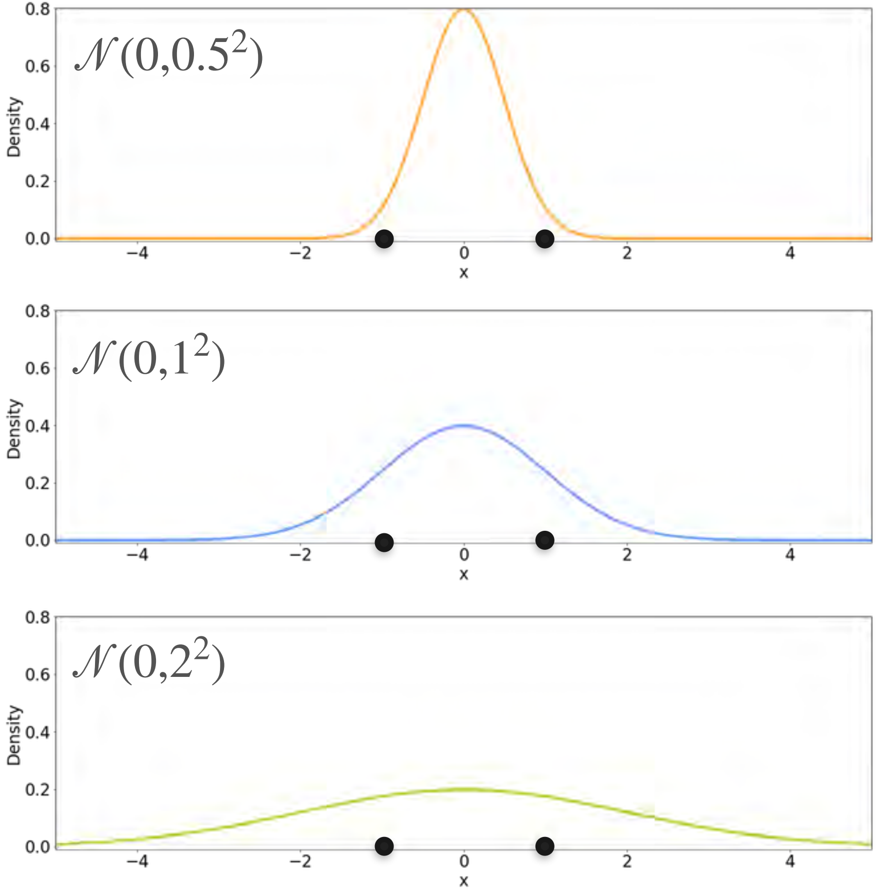


Observations

Variance of the observations

$$\widehat{\sigma^2} = \frac{1}{2}\left((0-1)^2 + (0+1)^2\right)$$

Candidates

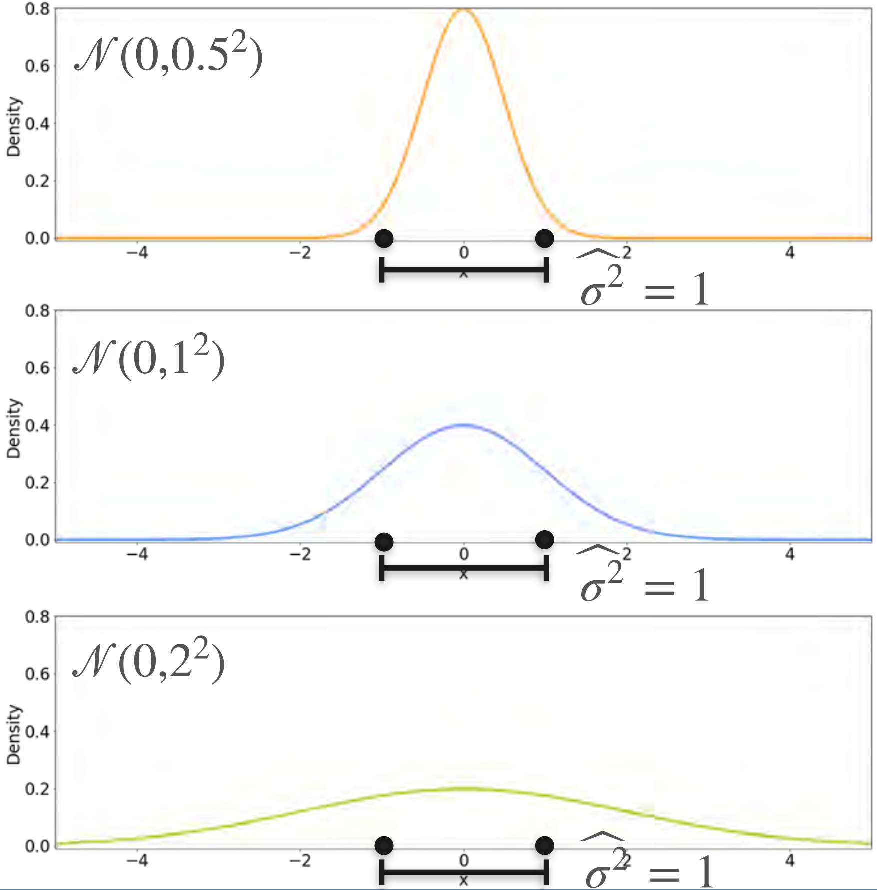


Observations

Variance of the observations

$$\widehat{\sigma^2} = \frac{1}{2}\left((0-1)^2 + (0+1)^2\right) = 1$$

Candidates

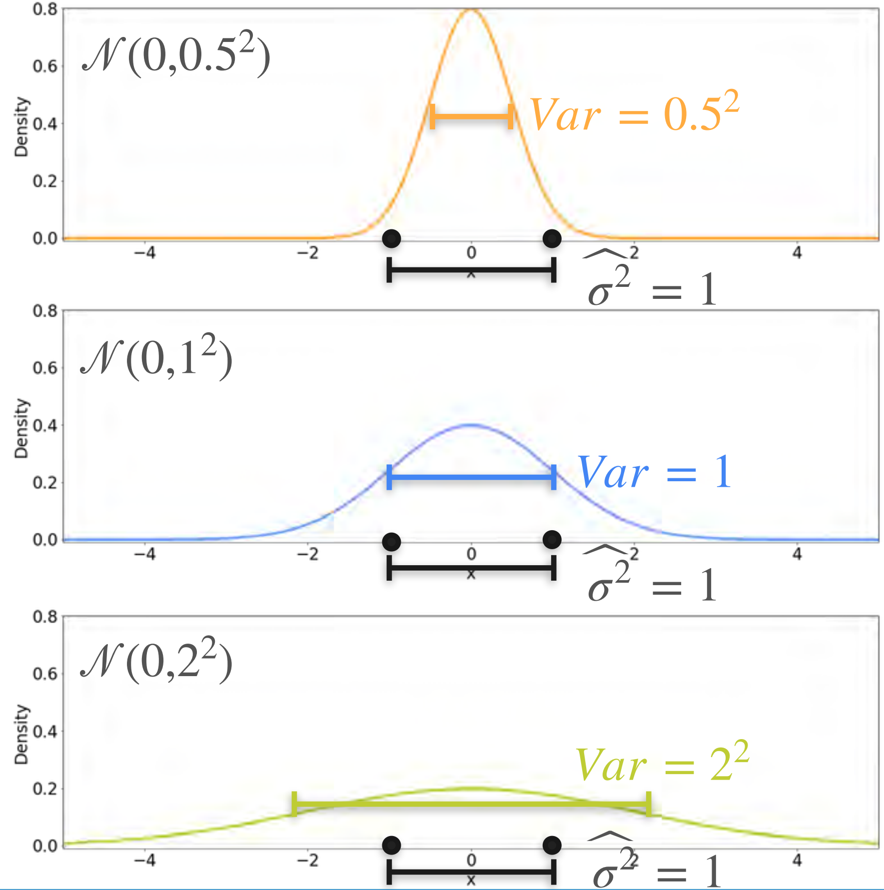


Observations

Variance of the observations

$$\widehat{\sigma^2} = \frac{1}{2}\left((0-1)^2 + (0+1)^2\right) = 1$$

Candidates

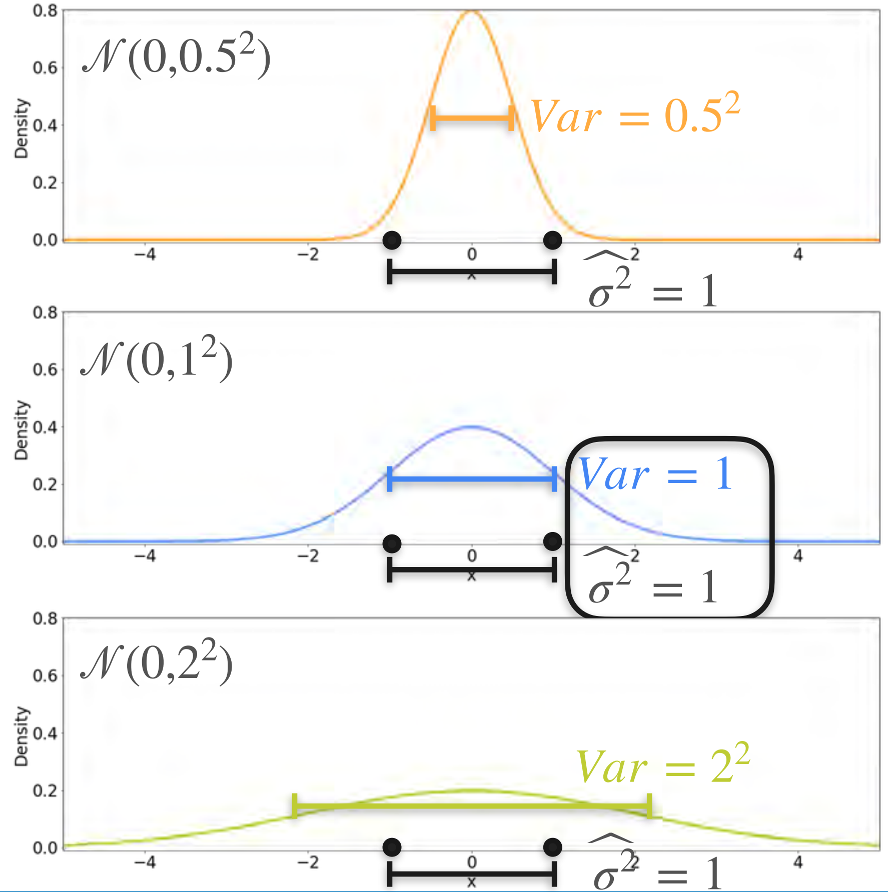


Observations

Variance of the observations

$$\widehat{\sigma^2} = \frac{1}{2}\left((0-1)^2 + (0+1)^2\right) = 1$$

Candidates

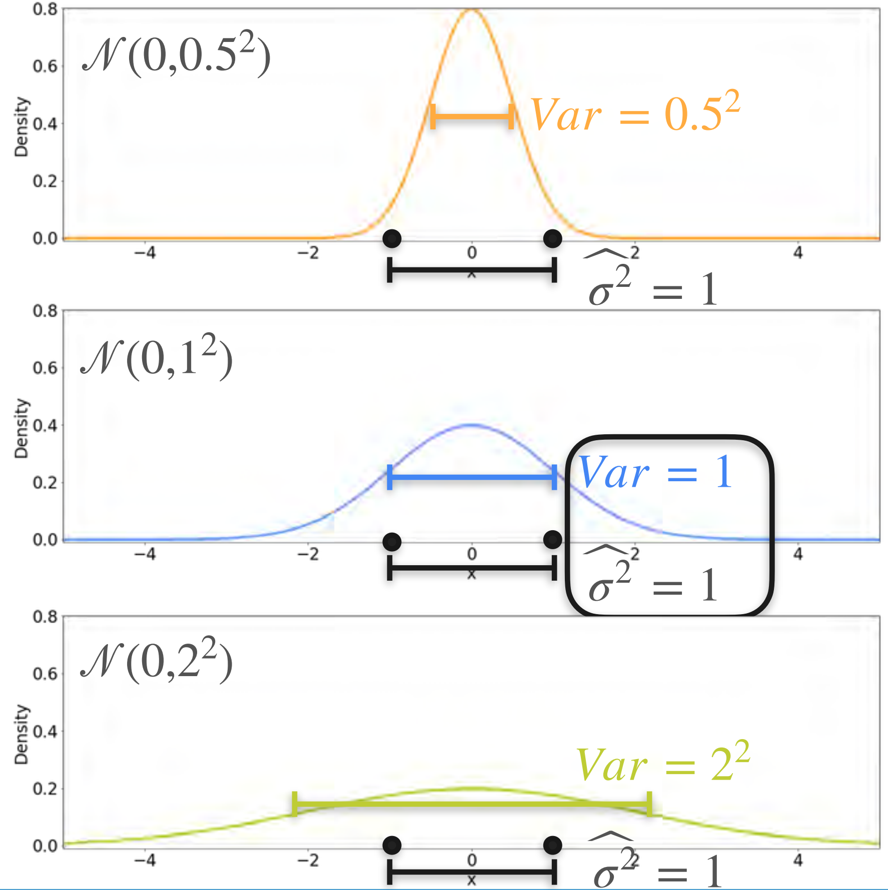


Observations

Variance of the observations

$$\widehat{\sigma^2} = \frac{1}{2}\left((0-1)^2 + (0+1)^2\right) = 1$$

Candidates



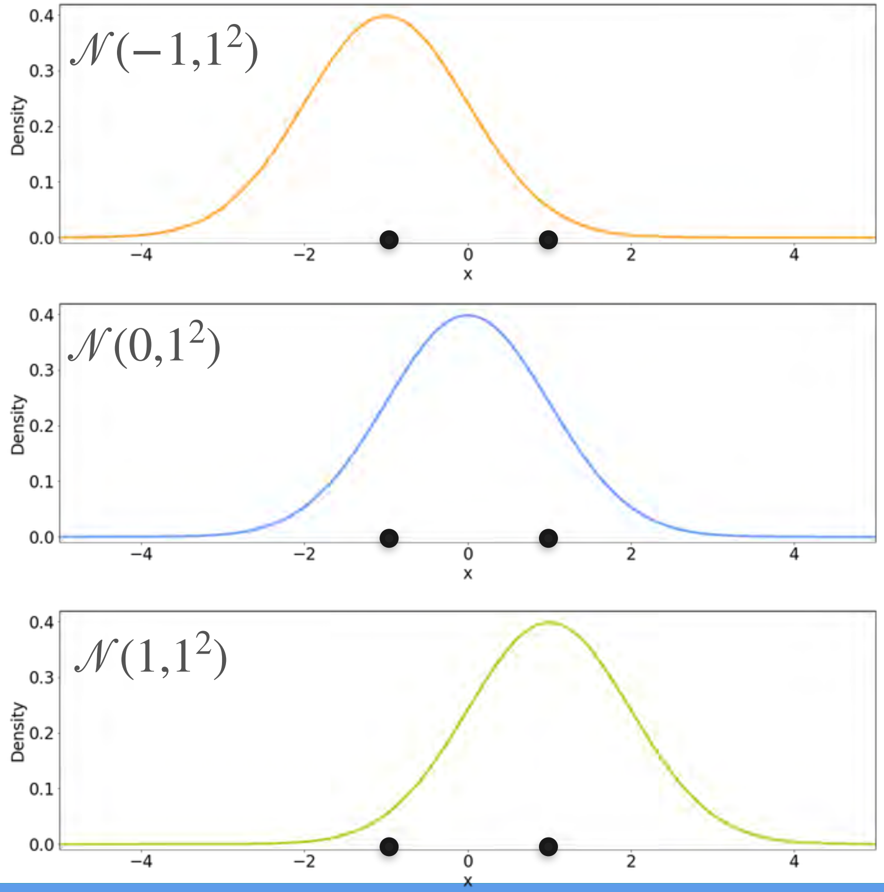
Observations

Variance of the observations

$$\widehat{\sigma^2} = \frac{1}{2}\left((0-1)^2 + (0+1)^2\right) = 1$$

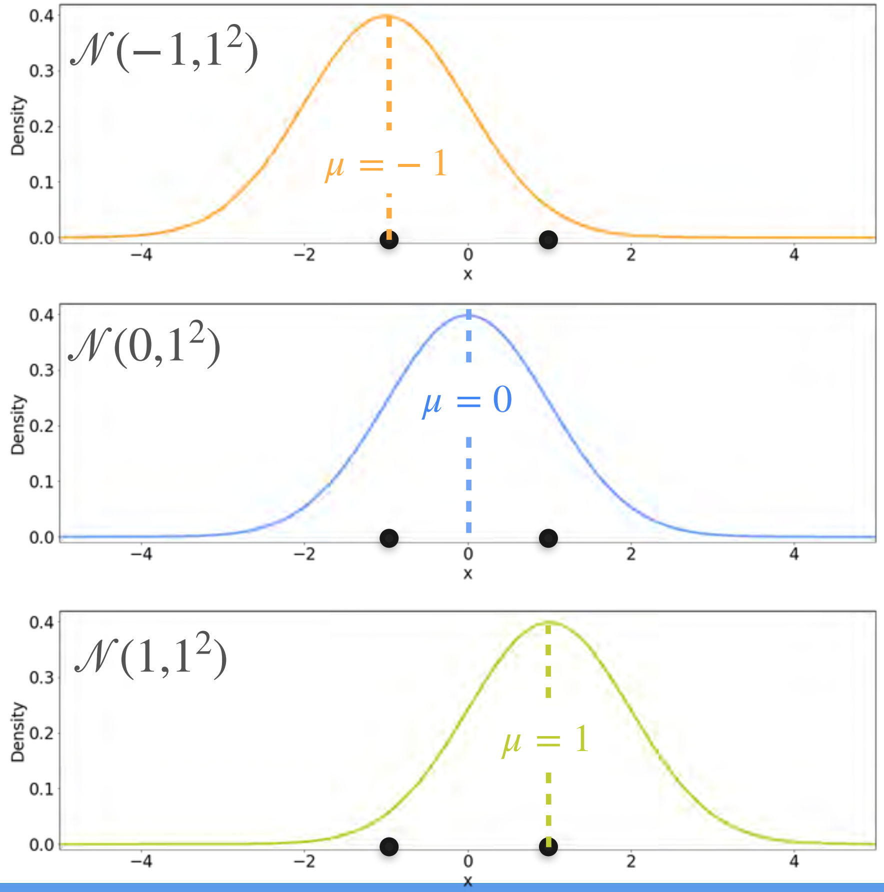
The best distribution is the one where the **variance** of the distribution is the **variance** of the sample

Candidates



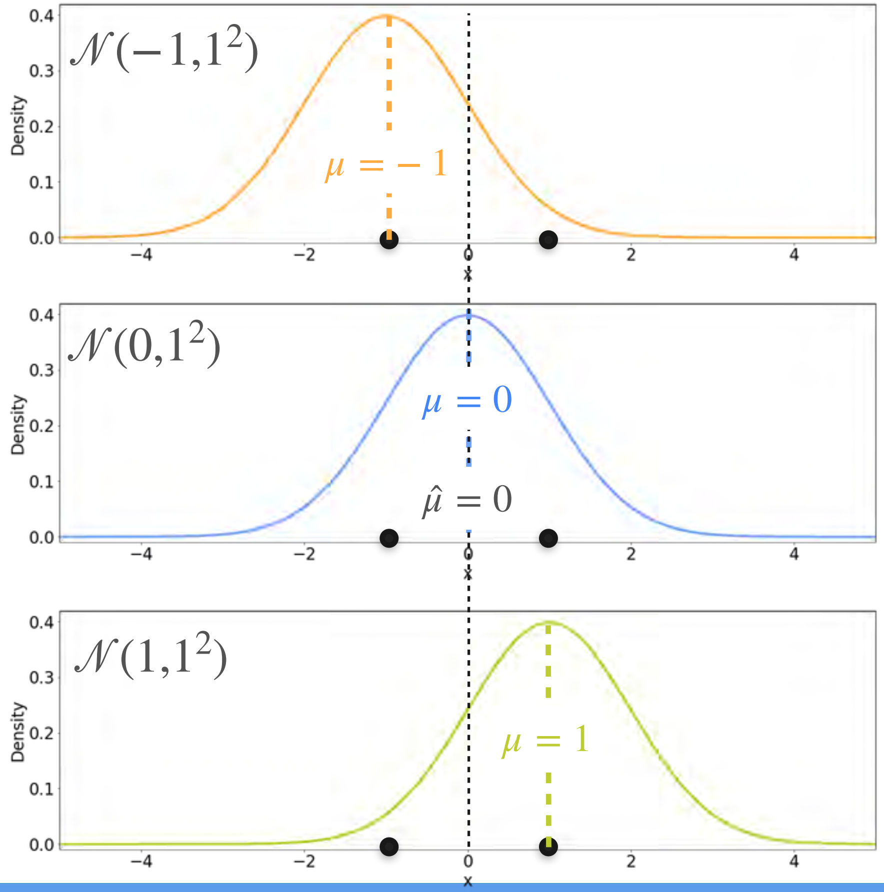
$\mathcal{N}(-1,1^2)$
$\mathcal{N}(0,1^2)$
$\mathcal{N}(1,1^2)$
Density
x
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4

Candidates



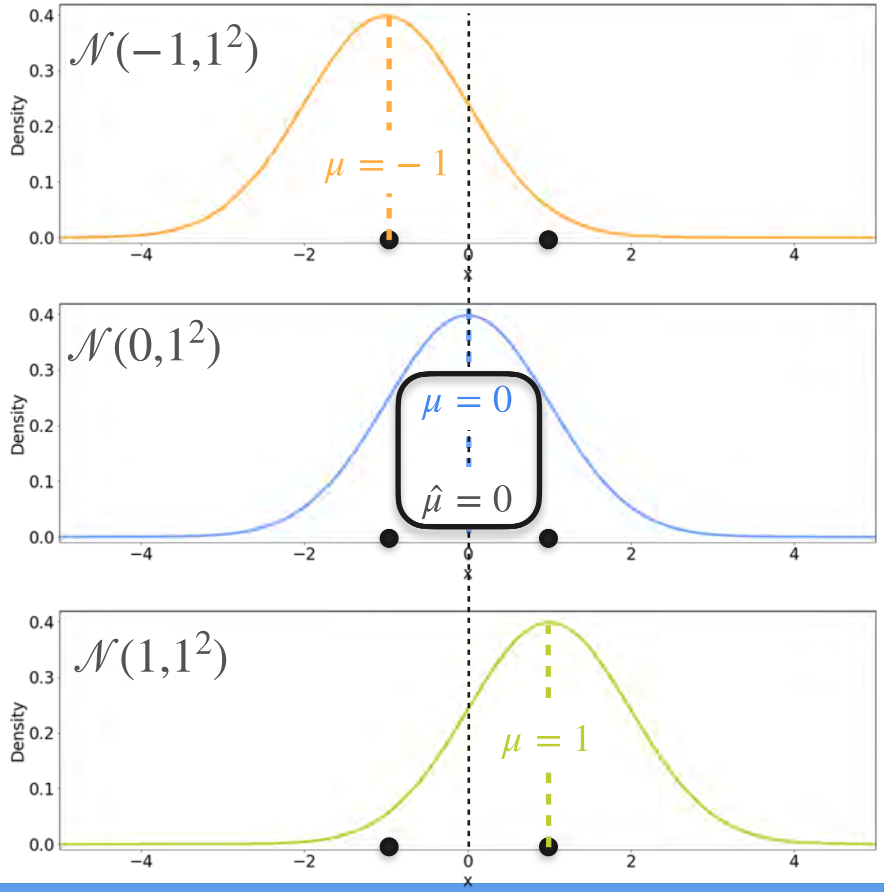
$\mathcal{N}(-1,1^2)$
$\mu = -1$
$\mathcal{N}(0,1^2)$
$\mu = 0$
$\mathcal{N}(1,1^2)$
$\mu = 1$
Density
x
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4

Candidates



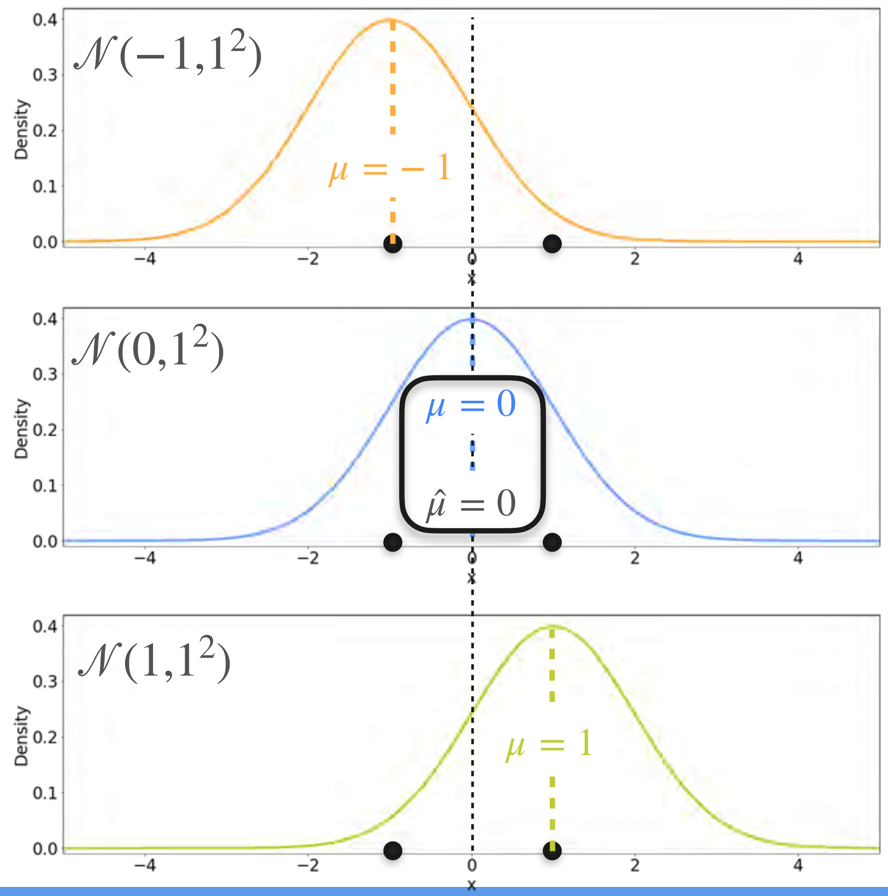
$\mathcal{N}(-1,1^2)$
$\mu = -1$
$\mathcal{N}(0,1^2)$
$\mu = 0$
$\hat{\mu} = 0$
$\mathcal{N}(1,1^2)$
$\mu = 1$
Density
x

Candidates



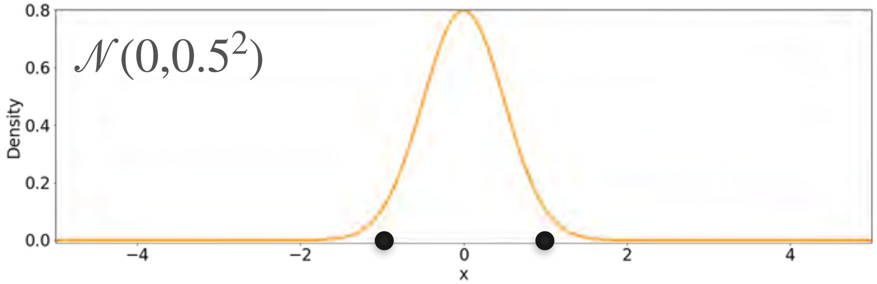
$\mathcal{N}(-1,1^2)$
$\mu = -1$
$\mathcal{N}(0,1^2)$
$\mu = 0$
$\hat{\mu} = 0$
$\mathcal{N}(1,1^2)$
$\mu = 1$
Density
x

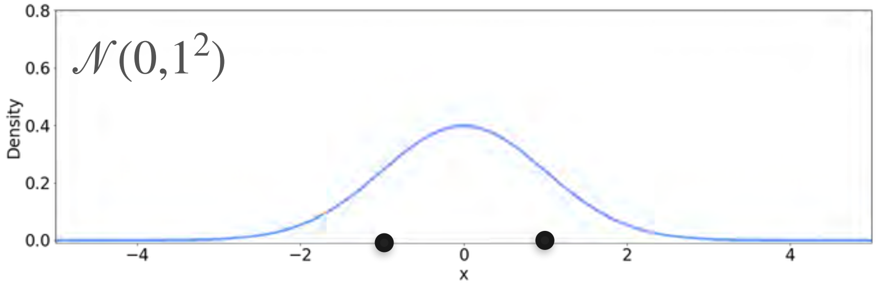
Candidates

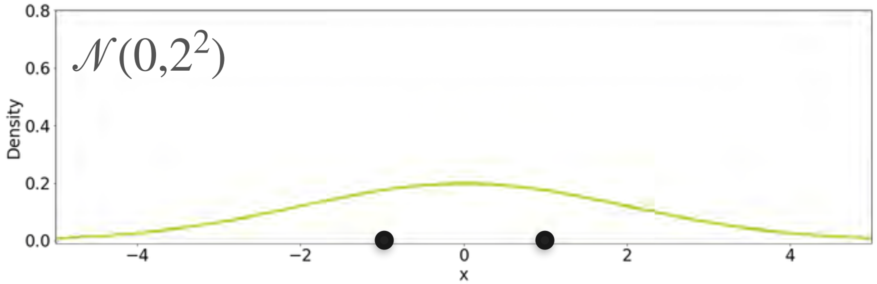


The best distribution is the one where the **mean** of the distribution is the **mean** of the sample

Candidates



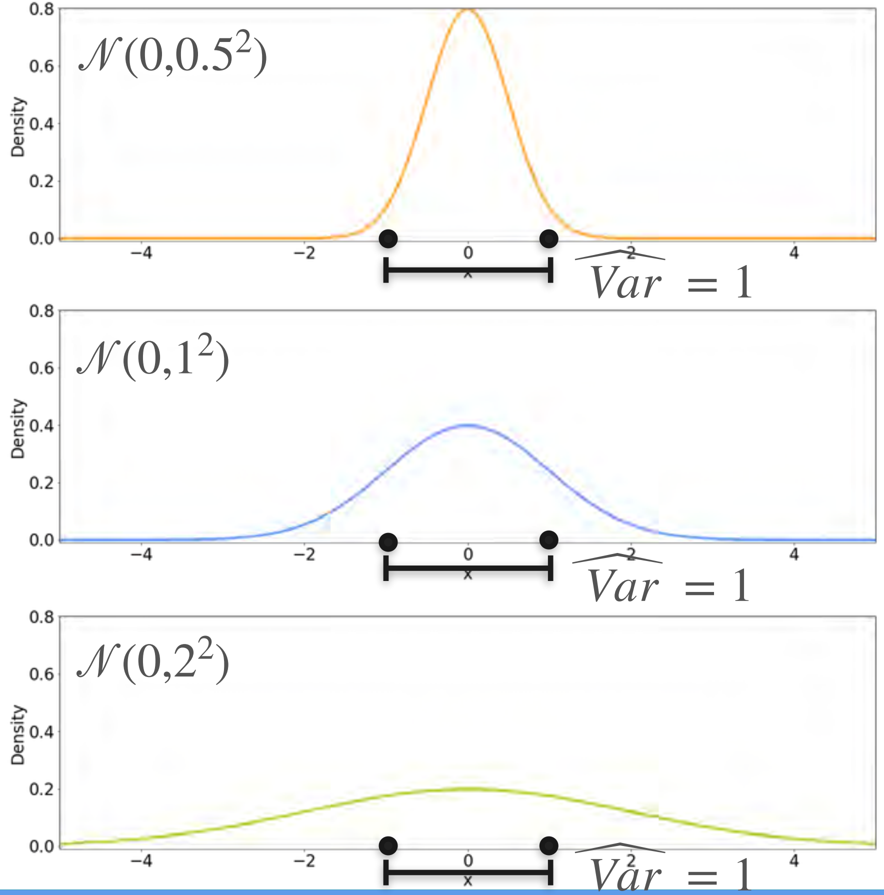




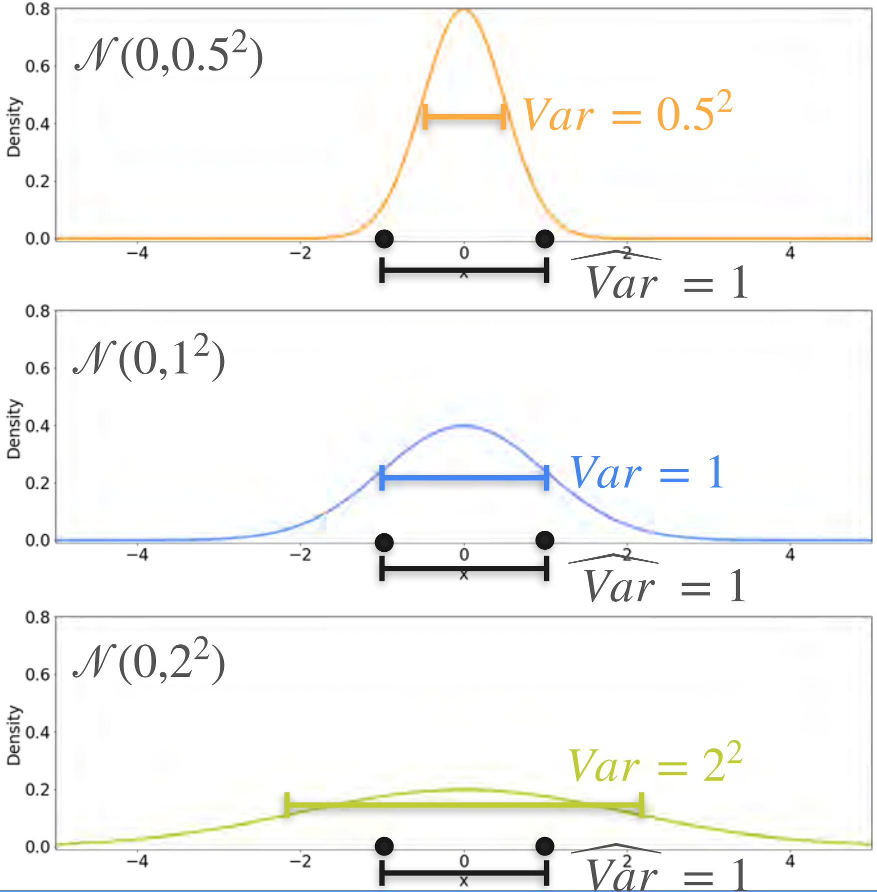
Observations

Candidates

Observations



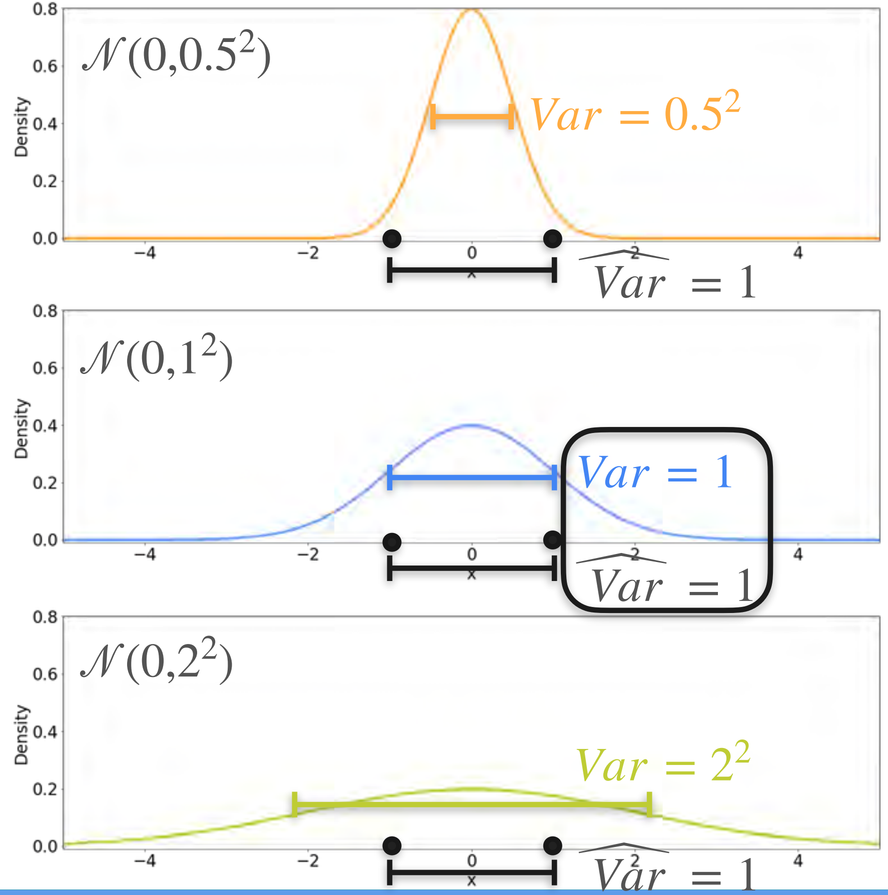
Candidates

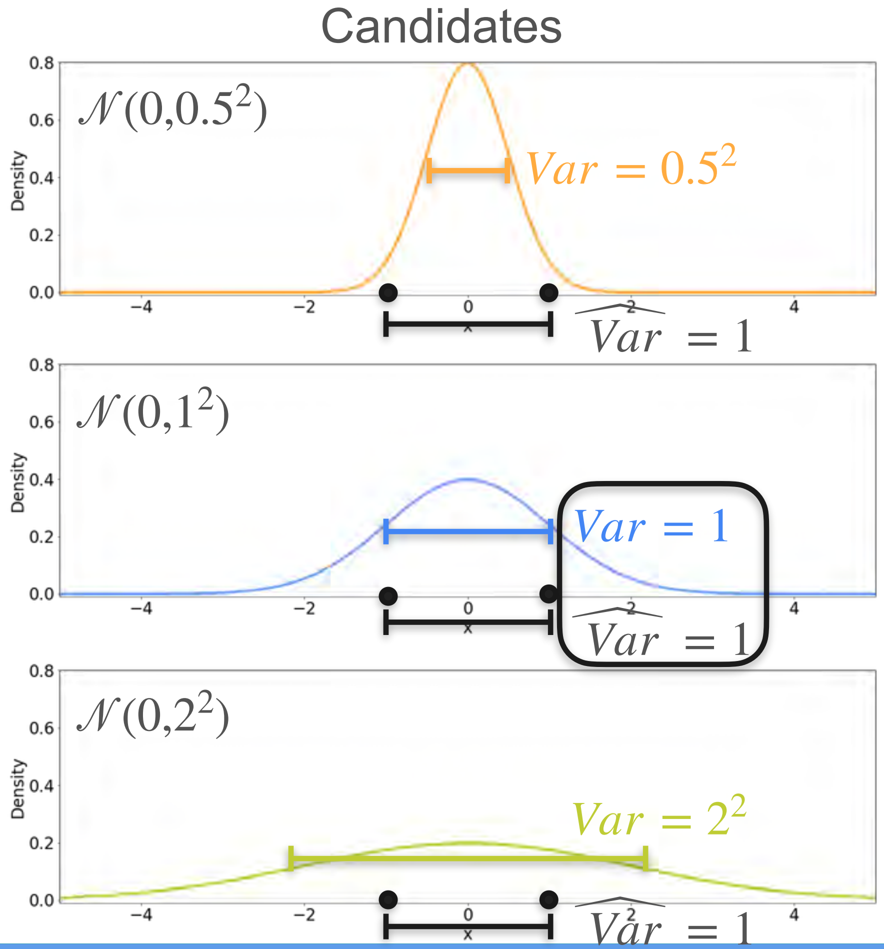


Observations

Candidates

Observations

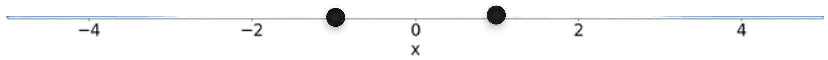




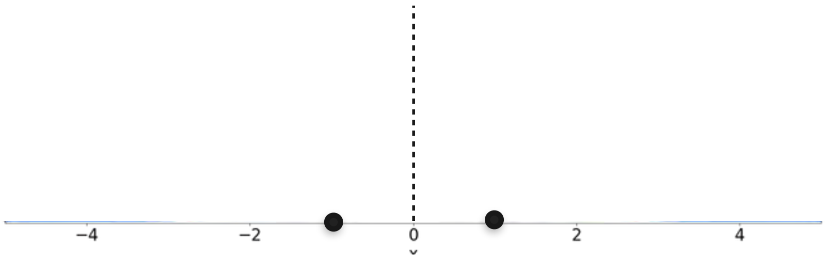
Observations

The best distribution is the one where the **variance** of the distribution is the **variance** of the sample

Best Gaussian Fit

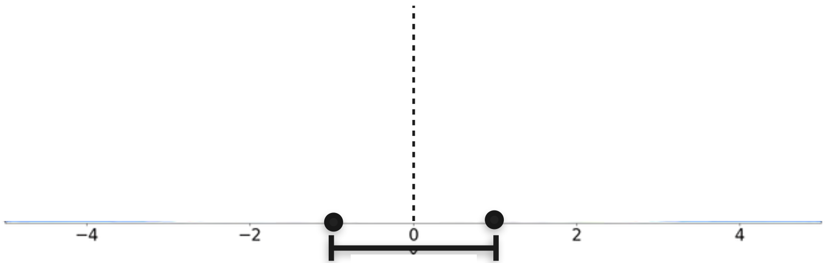


Best Gaussian Fit



$\hat{\mu} = 0$

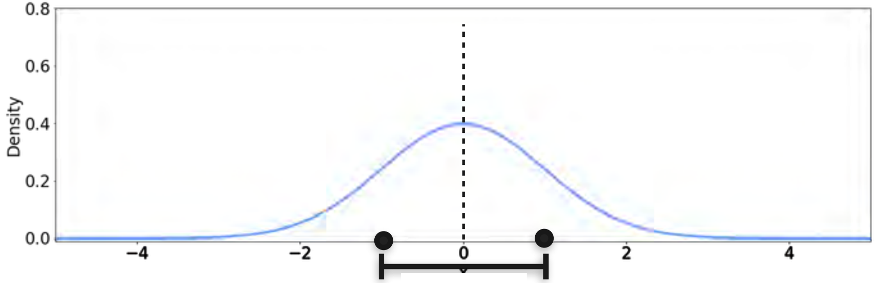
Best Gaussian Fit



$$\hat{\mu} = 0$$

$$\widehat{\sigma^2} = 1$$

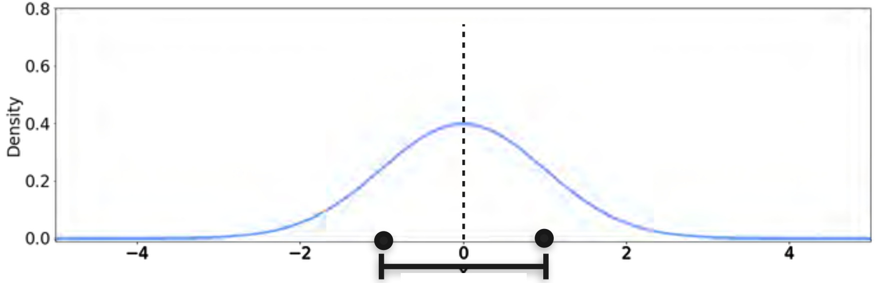
Best Gaussian Fit



$$\hat{\mu} = 0$$

$$\widehat{\sigma^2} = 1$$

Best Gaussian Fit

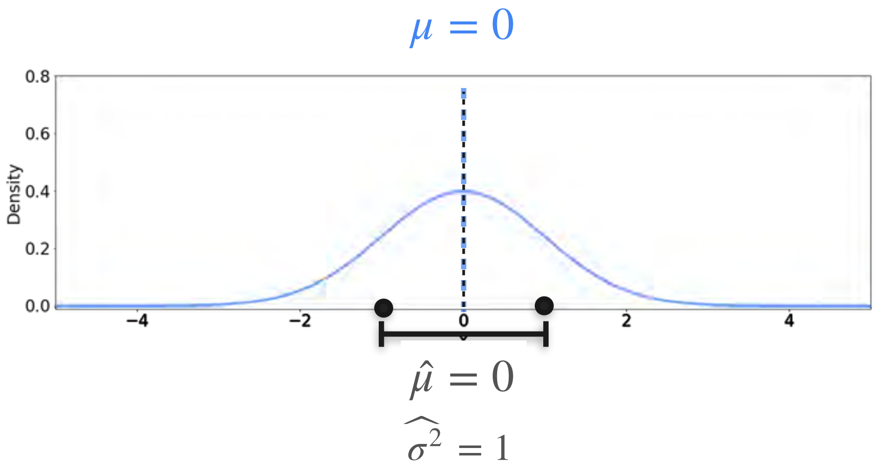


$$\hat{\mu} = 0$$

$$\widehat{\sigma^2} = 1$$

The best distribution is the one where the **mean** of the distribution is the **mean** of the sample

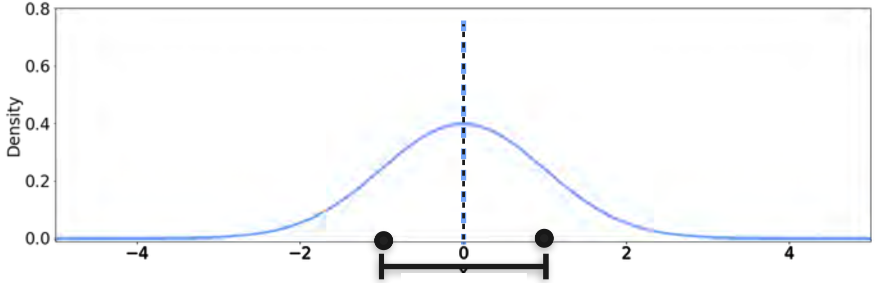
Best Gaussian Fit



The best distribution is the one where the **mean** of the distribution is the **mean** of the sample

Best Gaussian Fit

$\mu = 0$



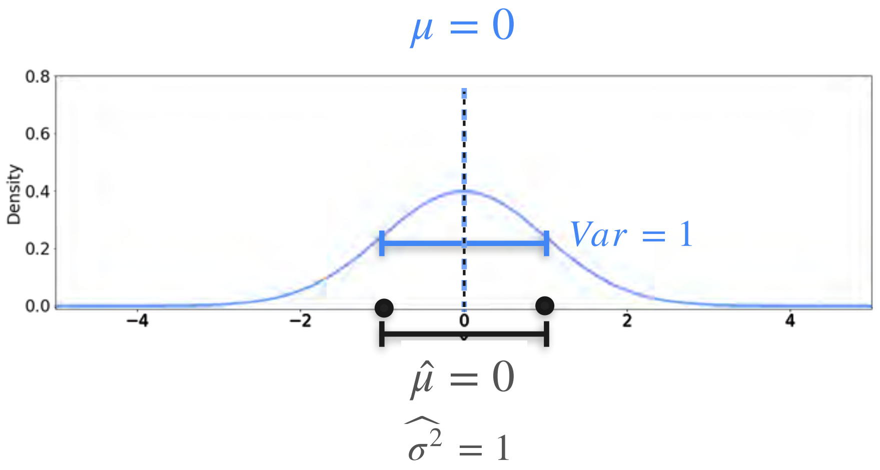
$\hat{\mu} = 0$

$\widehat{\sigma^2} = 1$

The best distribution is the one where the **mean** of the distribution is the **mean** of the sample

The best distribution is the one where the **variance** of the distribution is the **variance** of the sample

Best Gaussian Fit



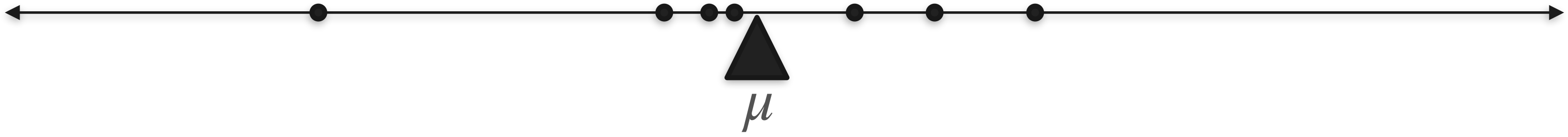
The best distribution is the one where the **mean** of the distribution is the **mean** of the sample

The best distribution is the one where the **variance** of the distribution is the **variance** of the sample

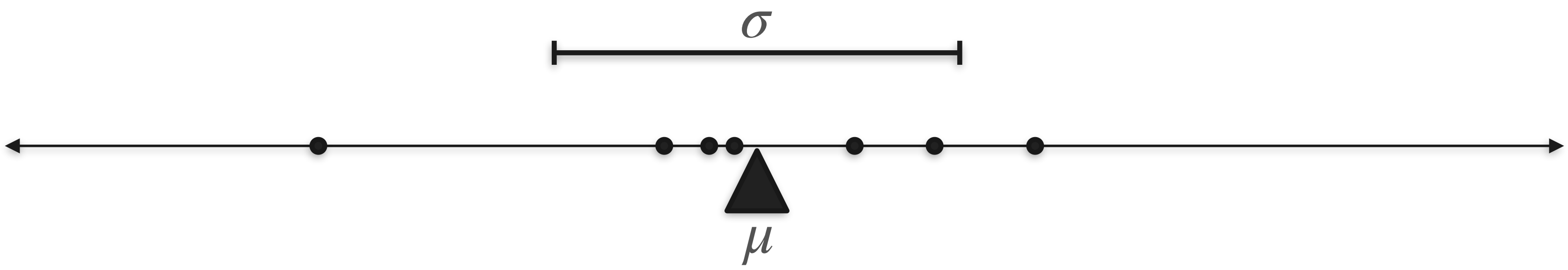
Best Gaussian Fit

Best Gaussian Fit

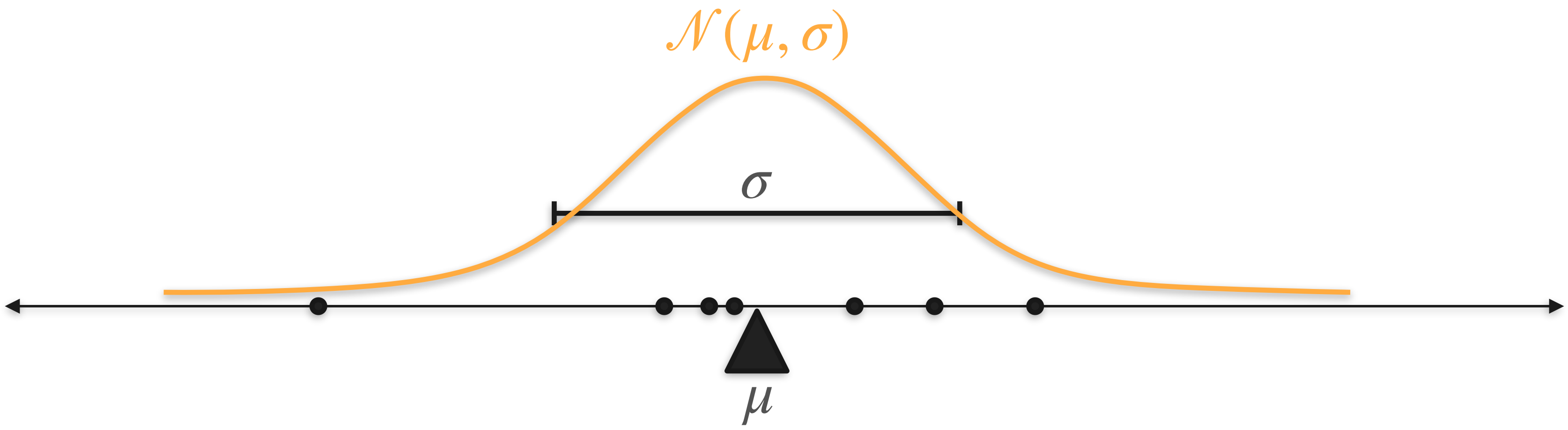
Best Gaussian Fit



Best Gaussian Fit



Best Gaussian Fit



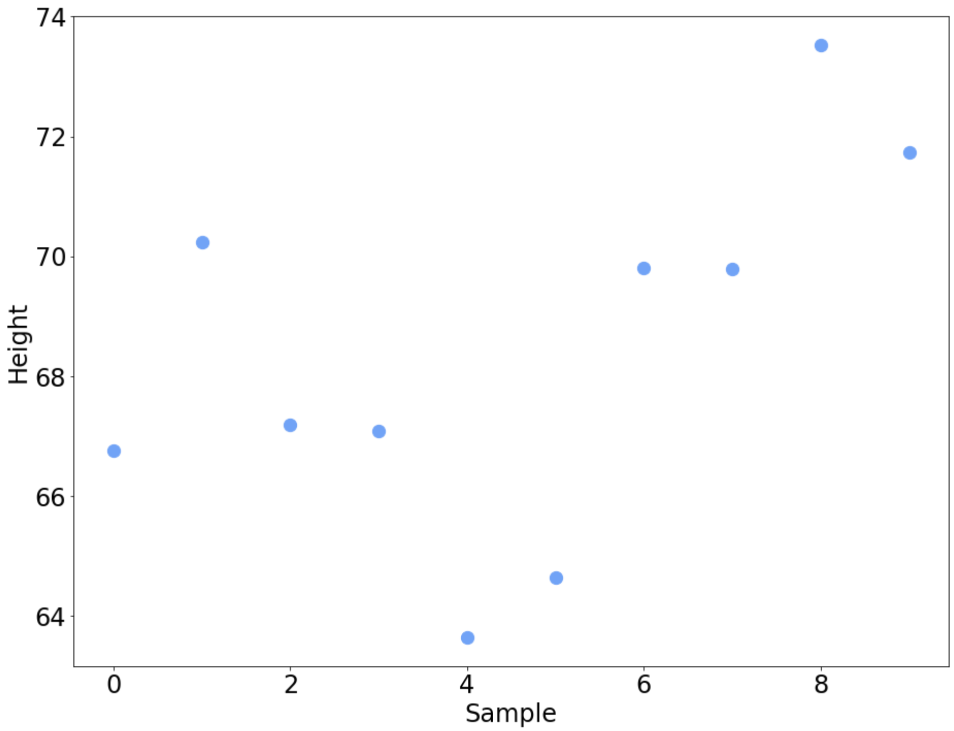
Maximum Likelihood: Gaussian Example

Maximum Likelihood: Gaussian Example

X = "Height of an 18 year old"

You measure 10 people
(i.e. sample size =10)

$$\mathbf{X} = (X_1, X_2, \ldots, X_{10})$$



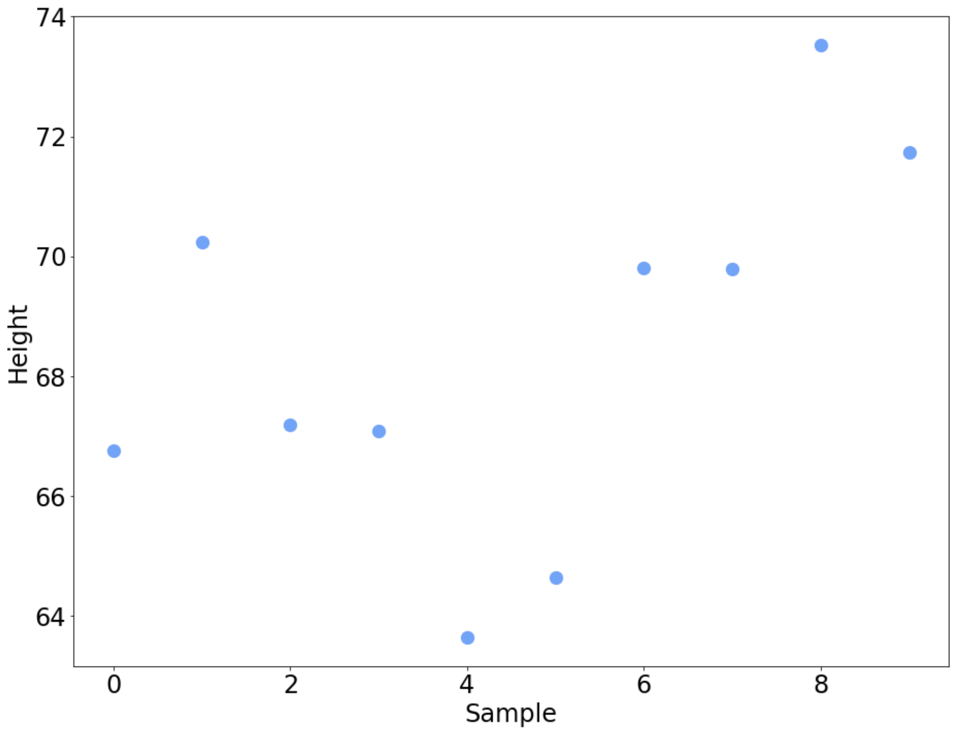
Maximum Likelihood: Gaussian Example

X = "Height of an 18 year old"

You measure 10 people
(i.e. sample size =10)

$$\mathbf{X} = (X_1, X_2, \ldots, X_{10})$$

$$X_1, X_2, \ldots, X_{10} \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma)$$



Maximum Likelihood: Gaussian Example

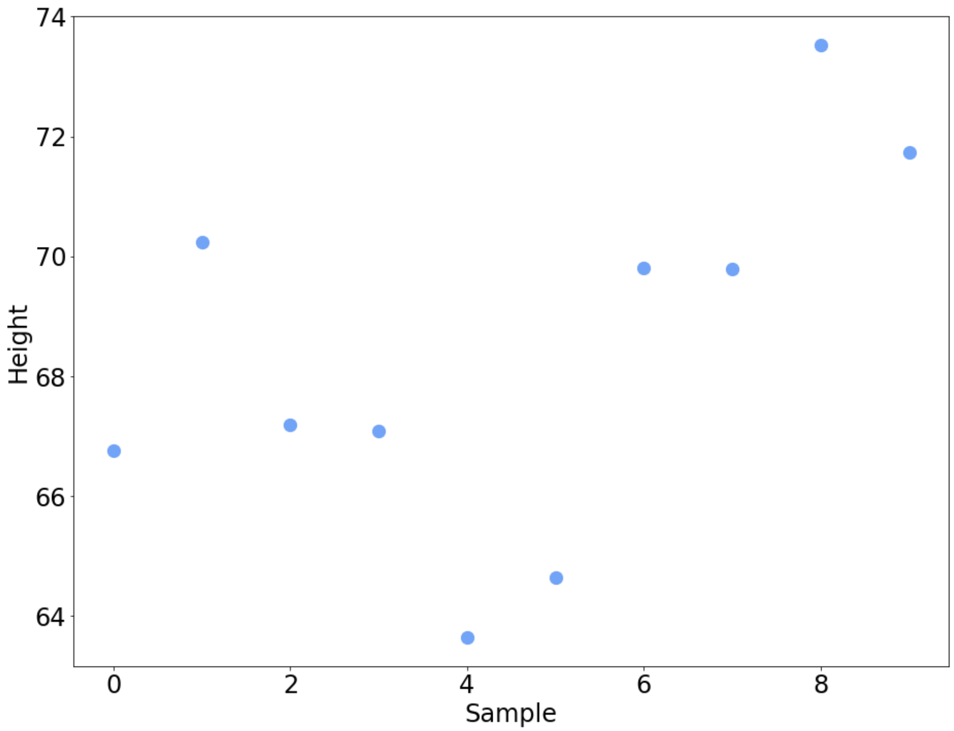
X = "Height of an 18 year old"

You measure 10 people
(i.e. sample size =10)

$$\mathbf{X} = (X_1, X_2, \ldots, X_{10})$$

$$X_1, X_2, \ldots, X_{10} \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma)$$

$$\mu ?$$



Maximum Likelihood: Gaussian Example

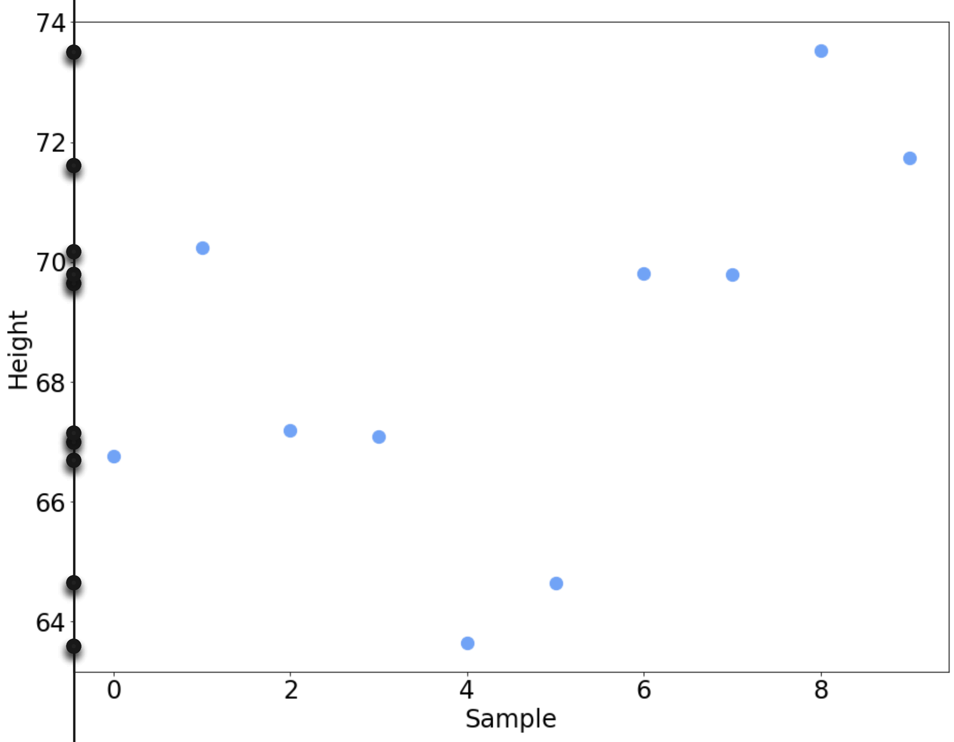
X = "Height of an 18 year old"

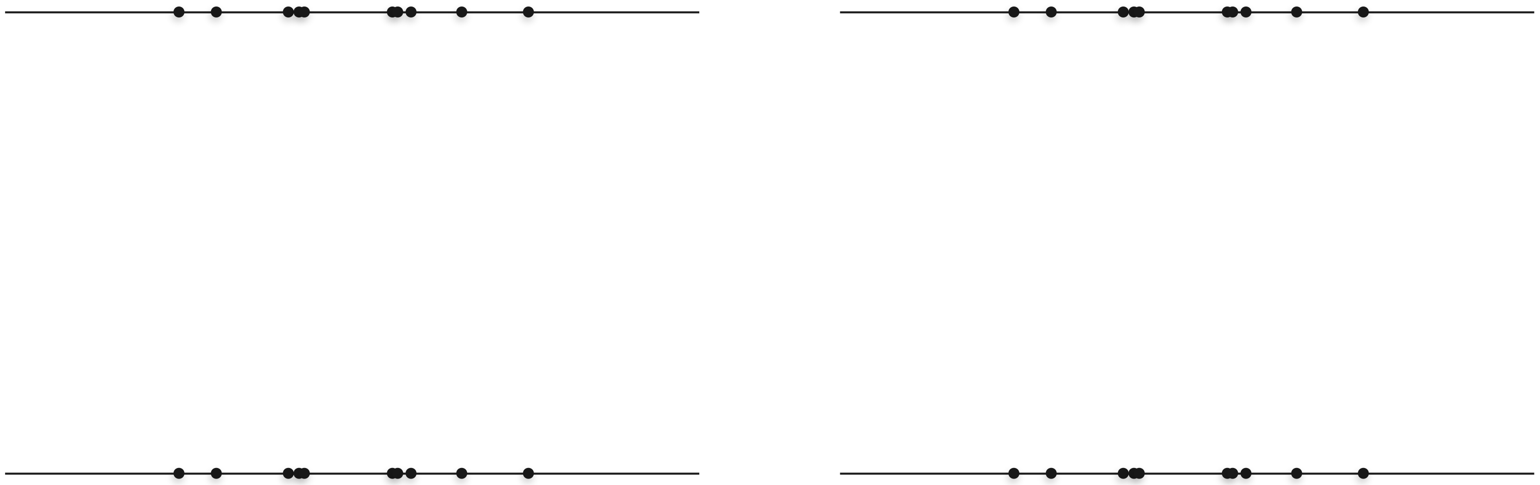
You measure 10 people
(i.e. sample size =10)

$$\mathbf{X} = (X_1, X_2, \ldots, X_{10})$$

$$X_1, X_2, \ldots, X_{10} \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma)$$

$$\mu ?$$





$\mu = 64$

$\mu = 68$

$\mu = 66$

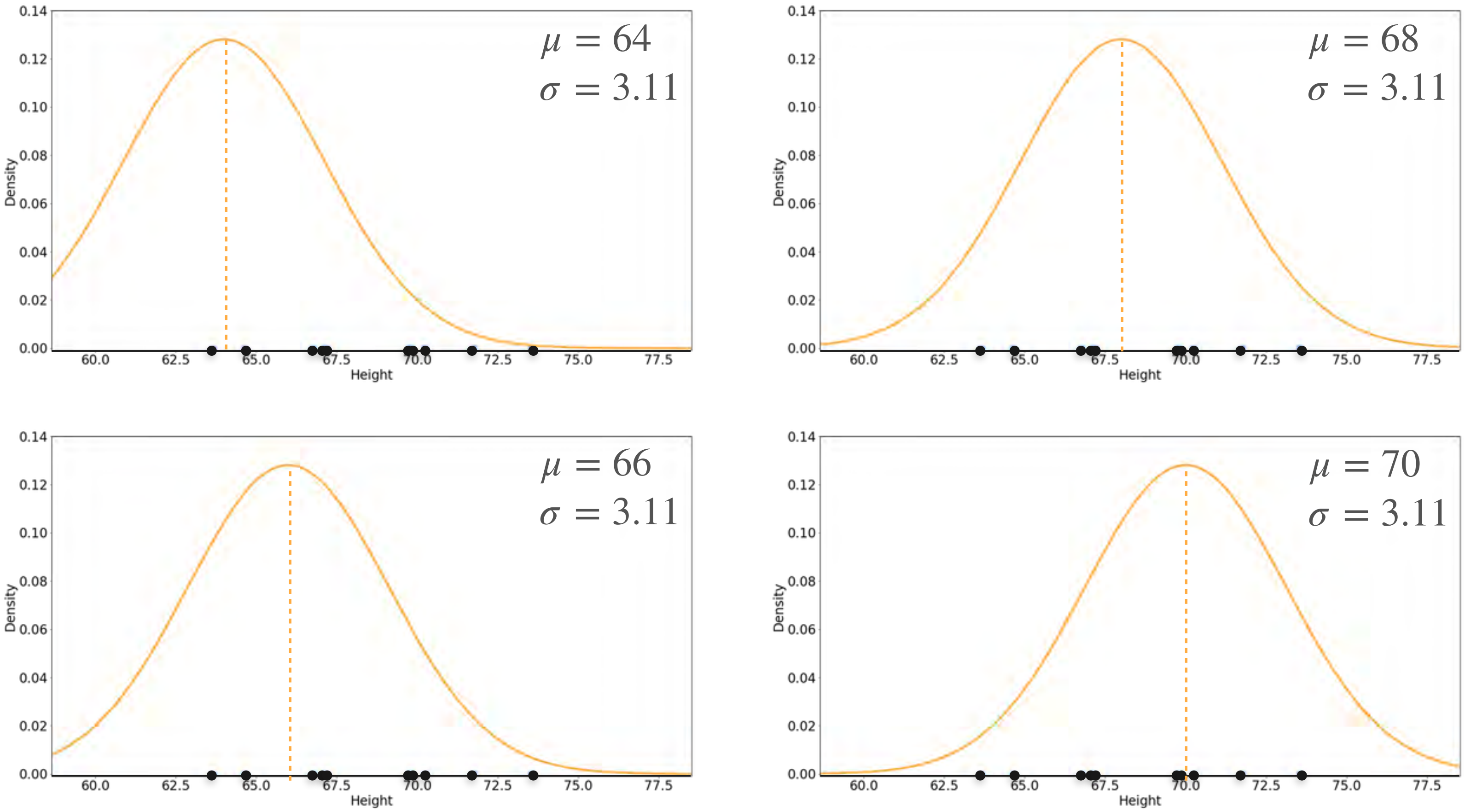
$\mu = 70$

$\mu = 64$
$\sigma = 3.11$

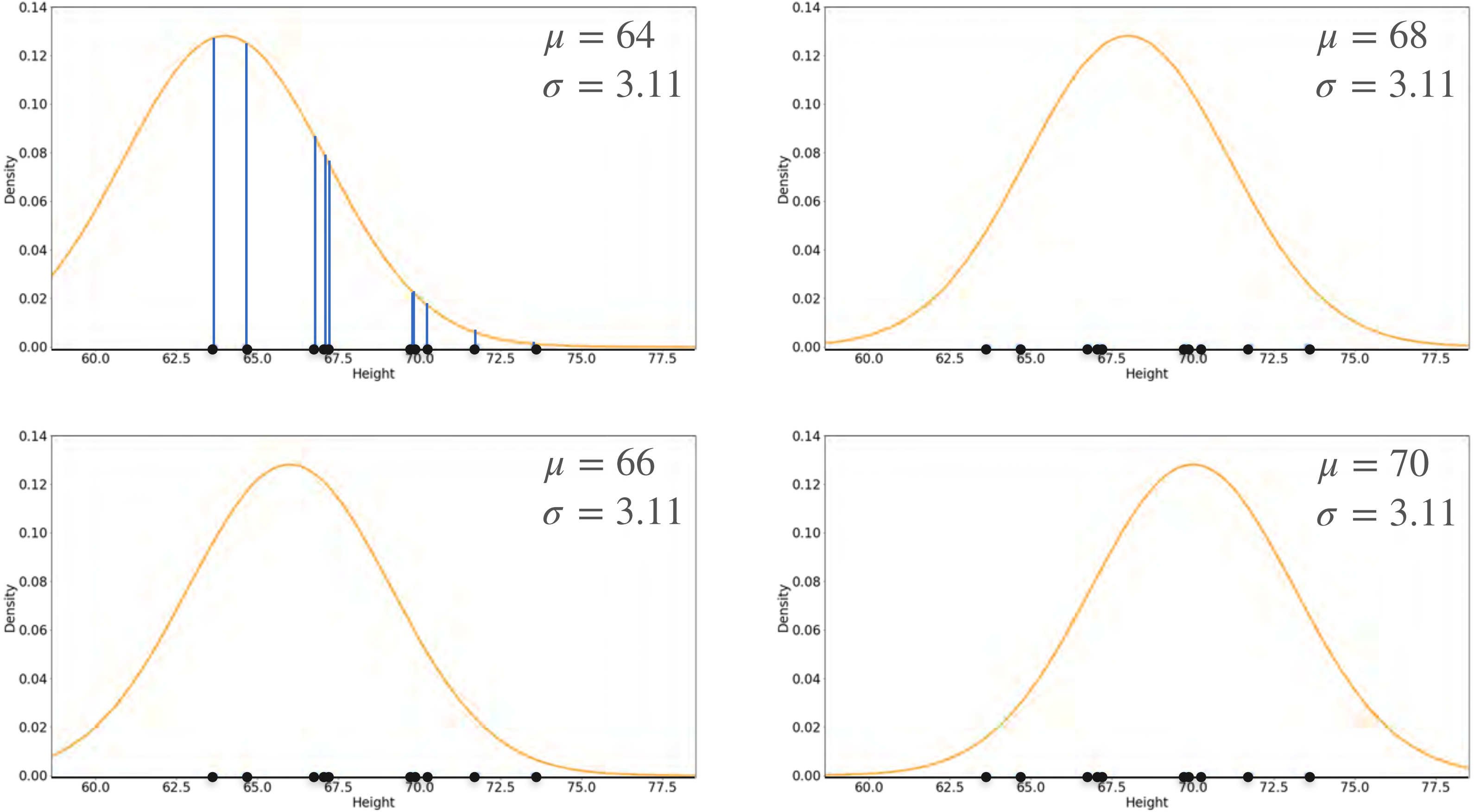
$\mu = 68$
$\sigma = 3.11$

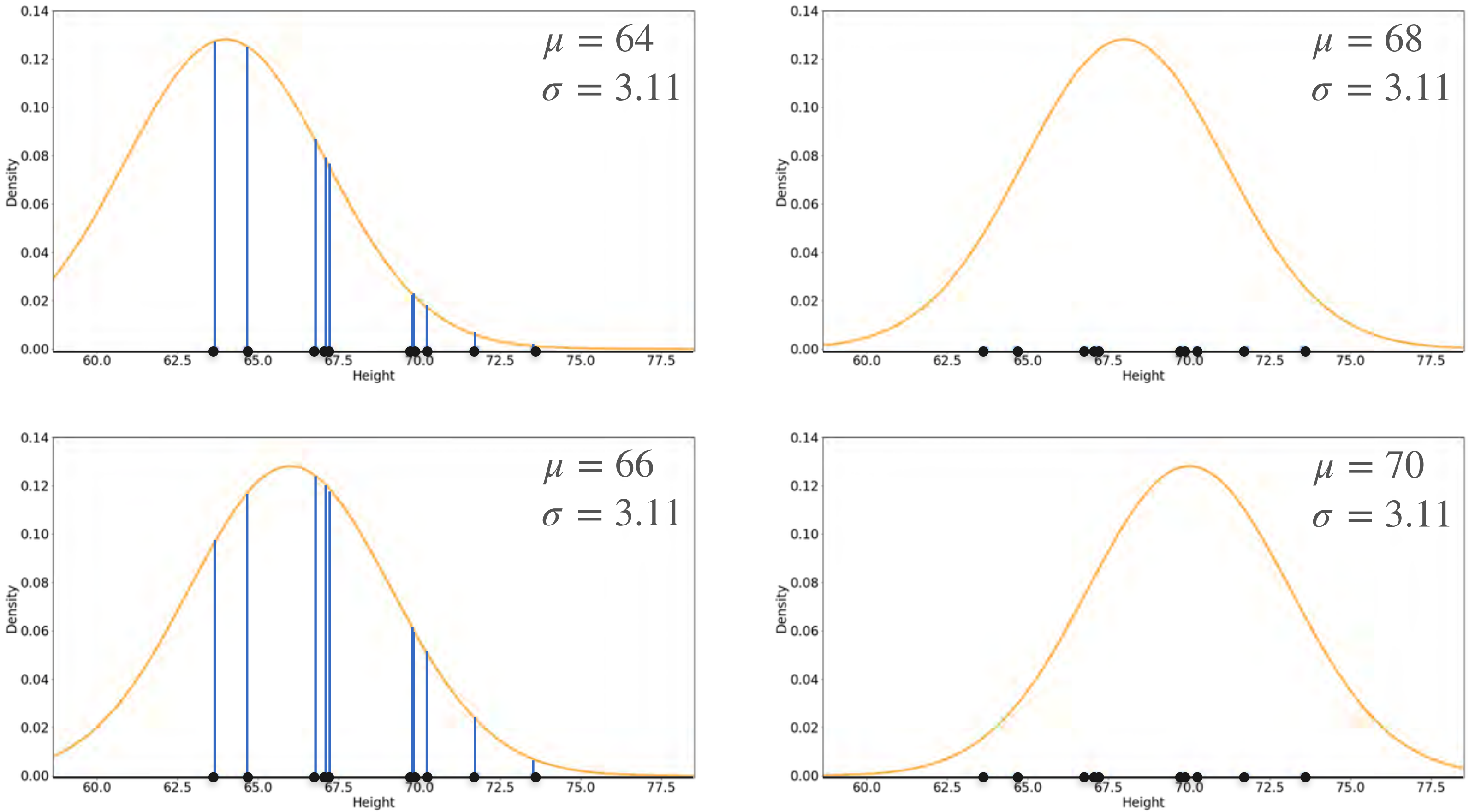
$\mu = 66$
$\sigma = 3.11$

$\mu = 70$
$\sigma = 3.11$

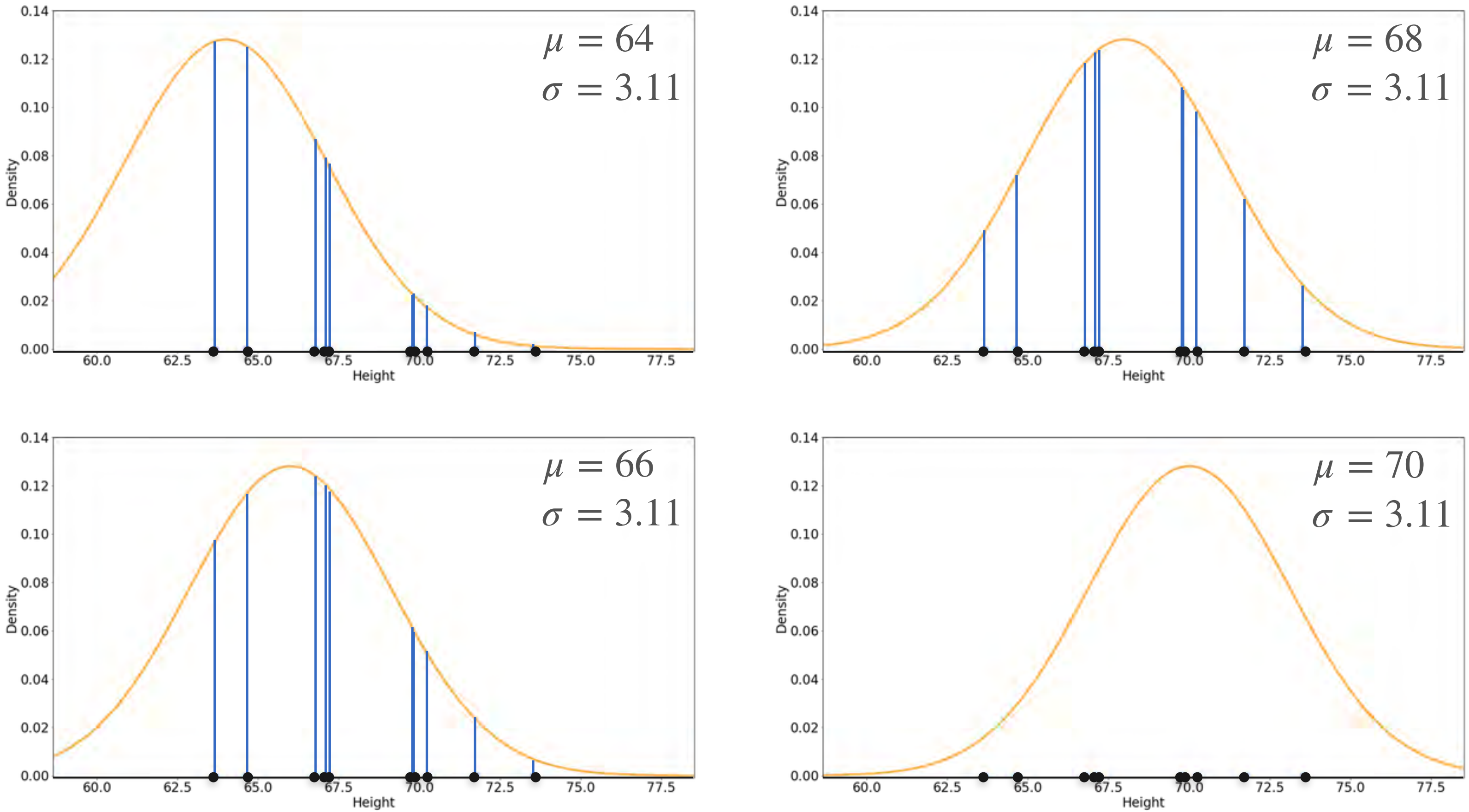


$\mu = 64$
$\sigma = 3.11$
$\mu = 68$
$\sigma = 3.11$
$\mu = 66$
$\sigma = 3.11$
$\mu = 70$
$\sigma = 3.11$
Density
Height
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
60.0
62.5
65.0
67.5
70.0
72.5
75.0
77.5

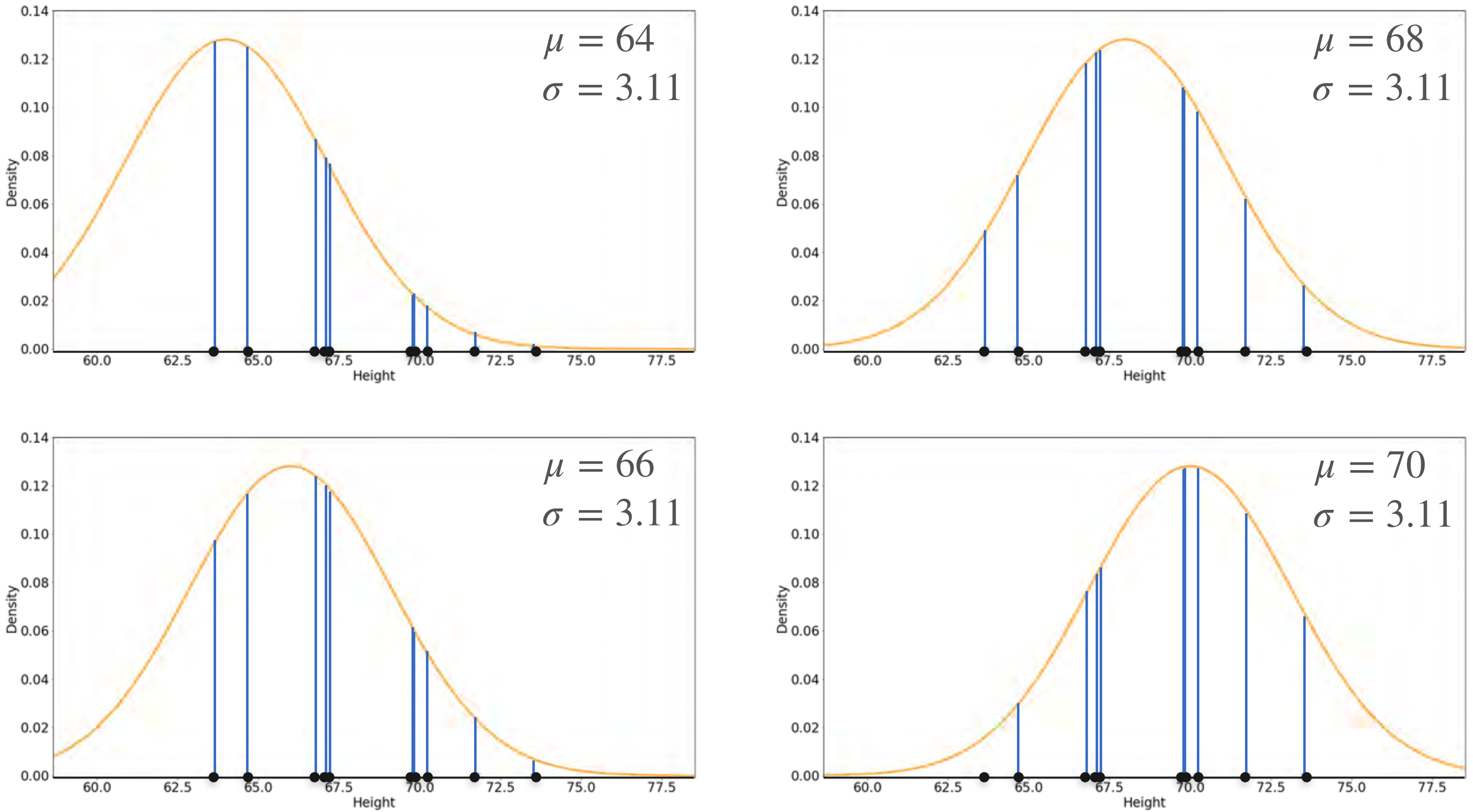

$\mu = 64$
$\sigma = 3.11$
$\mu = 68$
$\sigma = 3.11$
$\mu = 66$
$\sigma = 3.11$
$\mu = 70$
$\sigma = 3.11$
Density
Height
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
60.0
62.5
65.0
67.5
70.0
72.5
75.0
77.5



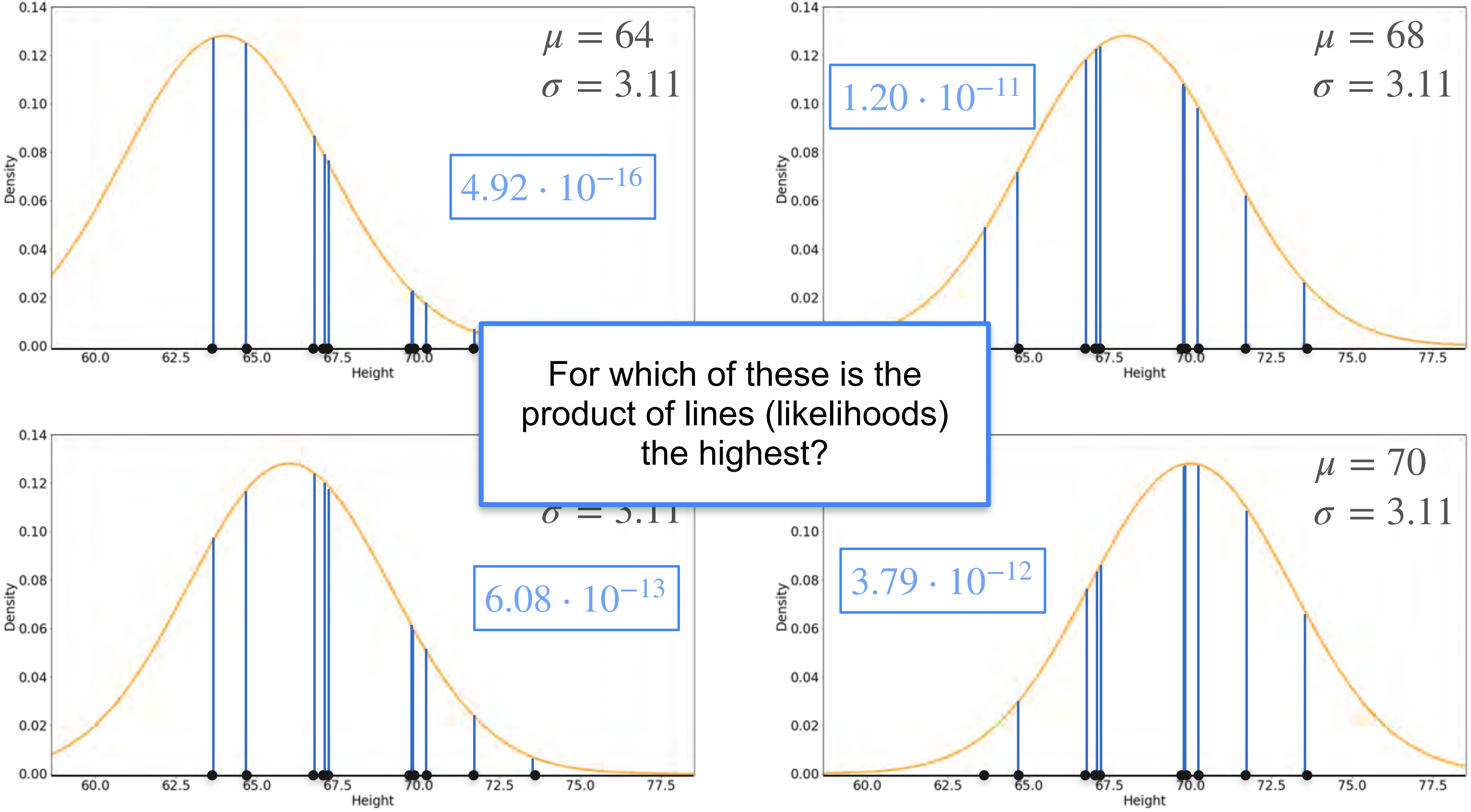
$\mu = 64$
$\sigma = 3.11$
$\mu = 68$
$\sigma = 3.11$
$\mu = 66$
$\sigma = 3.11$
$\mu = 70$
$\sigma = 3.11$
Density
Height
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
60.0
62.5
65.0
67.5
70.0
72.5
75.0
77.5



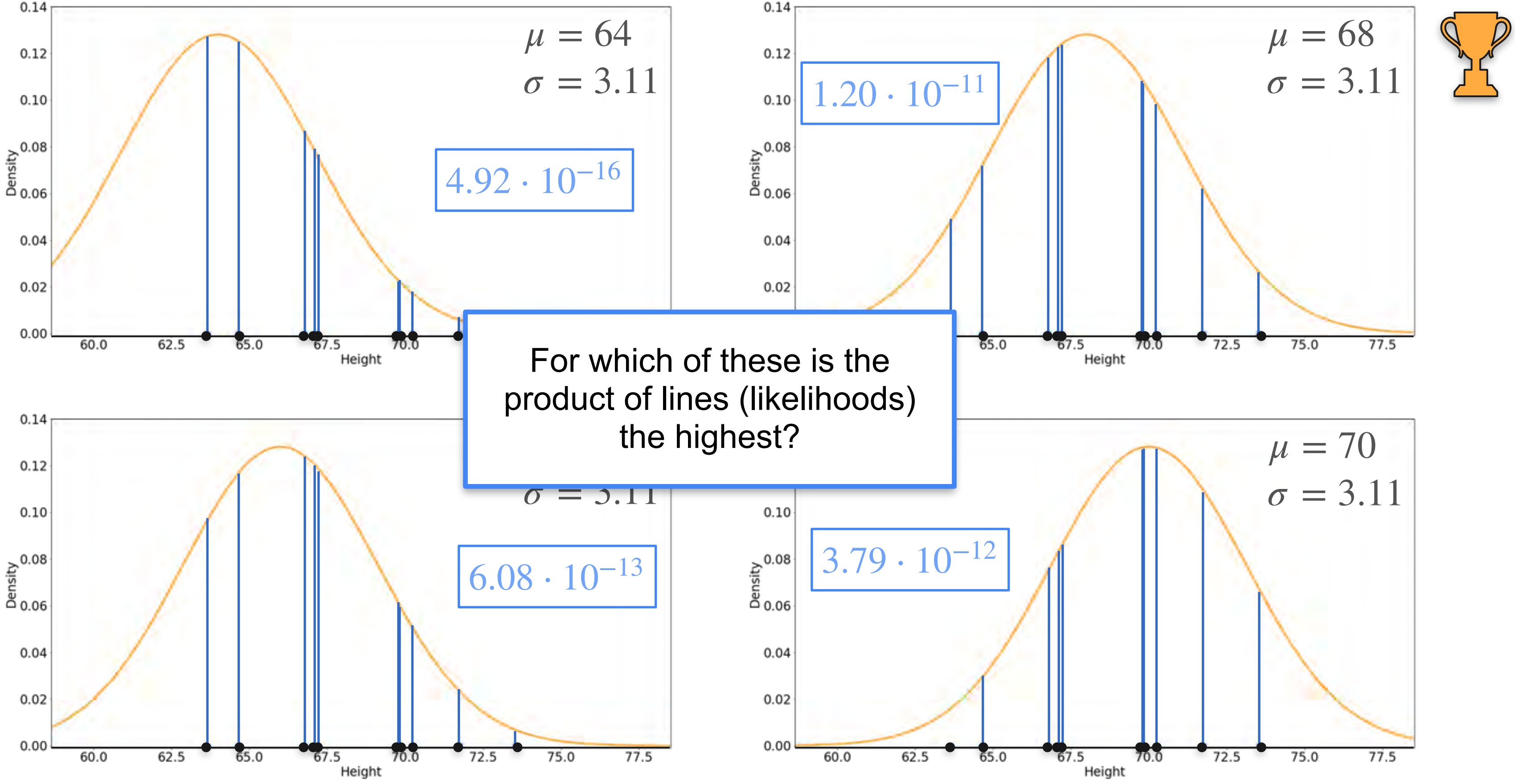
$\mu = 64$
$\sigma = 3.11$
$\mu = 68$
$\sigma = 3.11$
$\mu = 66$
$\sigma = 3.11$
$\mu = 70$
$\sigma = 3.11$
Density
Height
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
60.0
62.5
65.0
67.5
70.0
72.5
75.0
77.5



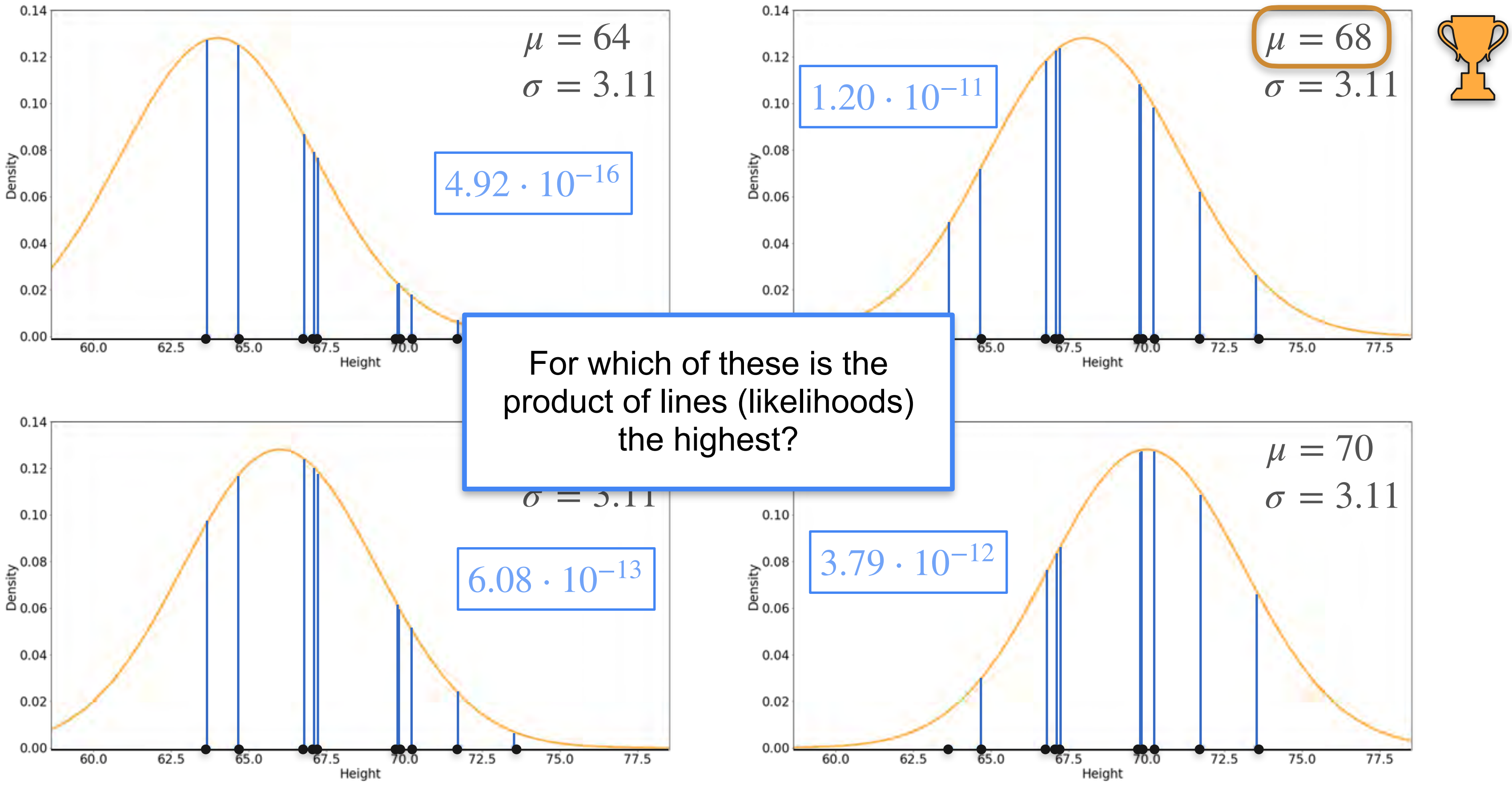
$\mu = 64$
$\sigma = 3.11$
$\mu = 68$
$\sigma = 3.11$
$\mu = 66$
$\sigma = 3.11$
$\mu = 70$
$\sigma = 3.11$
Density
Height
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
60.0
62.5
65.0
67.5
70.0
72.5
75.0
77.5



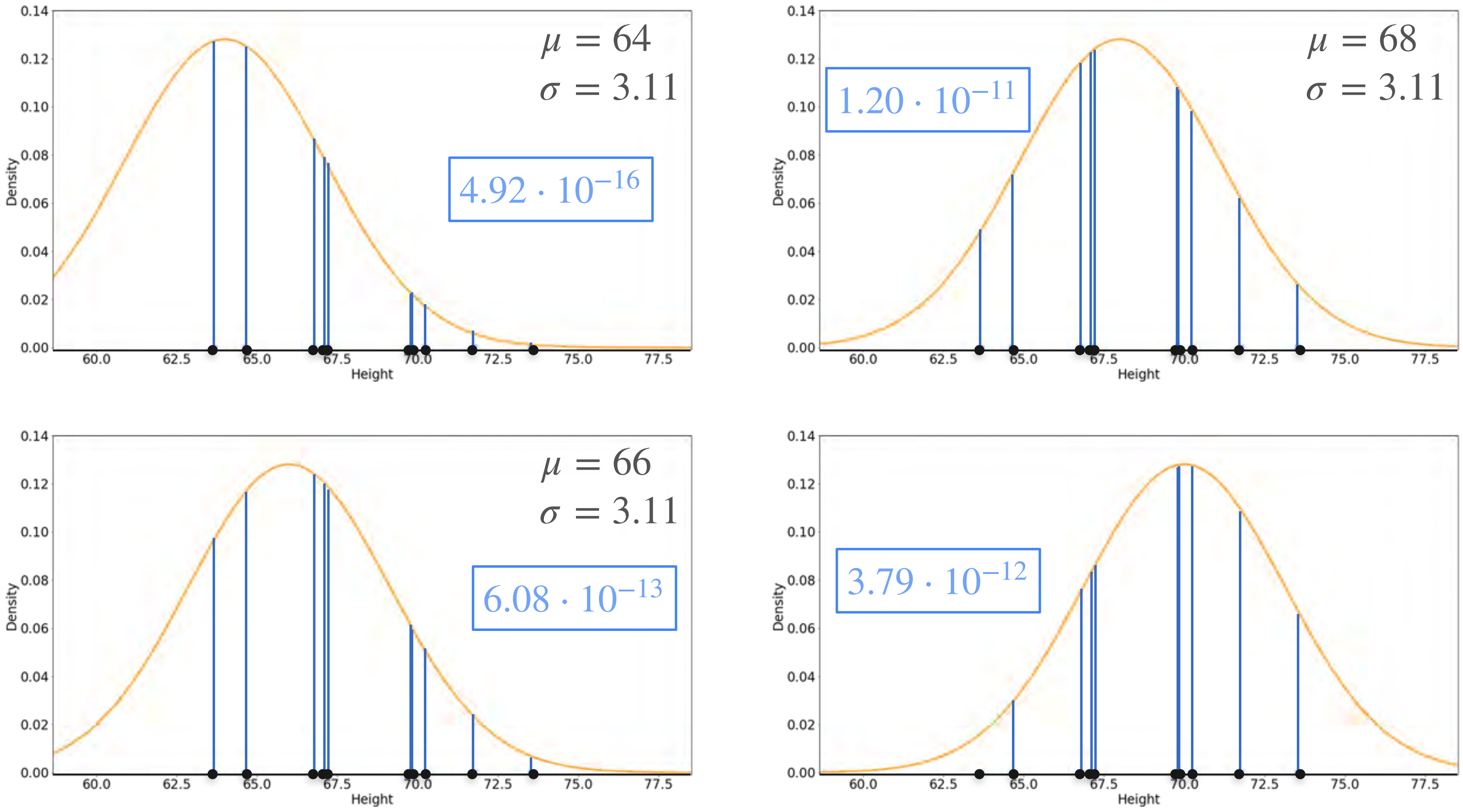
For which of these is the product of lines (likelihoods) the highest?



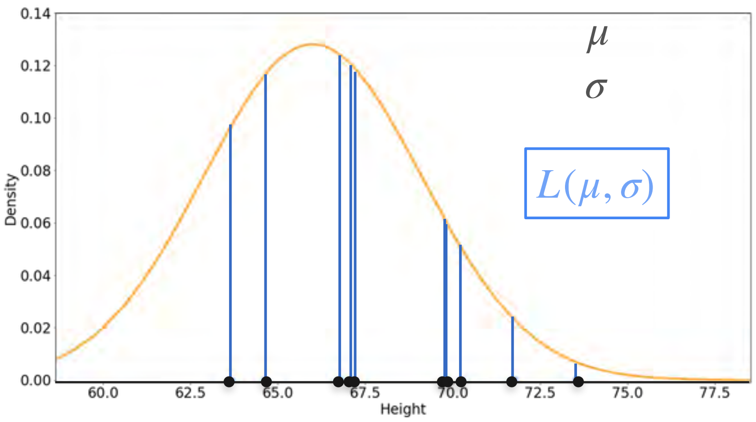
For which of these is the product of lines (likelihoods) the highest?



$\mu = 64$
$\sigma = 3.11$
$4.92 \cdot 10^{-16}$
$1.20 \cdot 10^{-11}$
$\mu = 68$
$\sigma = 3.11$
$6.08 \cdot 10^{-13}$
$\mu = 70$
$\sigma = 3.11$
$3.79 \cdot 10^{-12}$
Density
Height
For which of these is the product of lines (likelihoods) the highest?

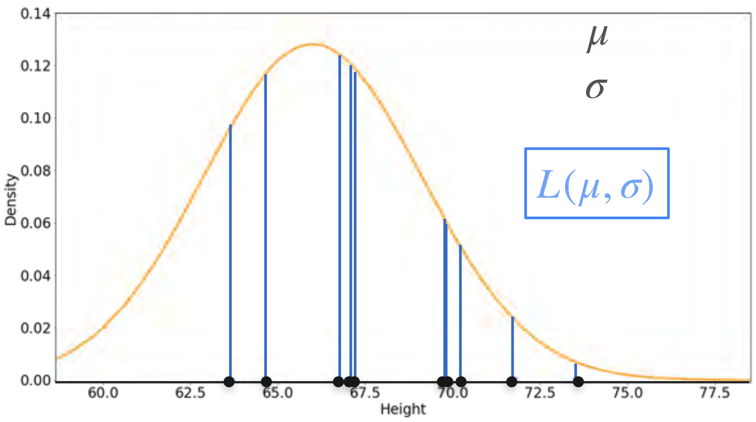


$\mu = 64$
$\sigma = 3.11$
$4.92 \cdot 10^{-16}$
$1.20 \cdot 10^{-11}$
$\mu = 68$
$\sigma = 3.11$
$\mu = 66$
$\sigma = 3.11$
$6.08 \cdot 10^{-13}$
$3.79 \cdot 10^{-12}$
Density
Height
0.00 0.02 0.04 0.06 0.08 0.10 0.12 0.14
60.0 62.5 65.0 67.5 70.0 72.5 75.0 77.5



$$\frac{d}{d\mu}\ell(\mu;\mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i - \mu)^2}{\sigma^2}\right]$$

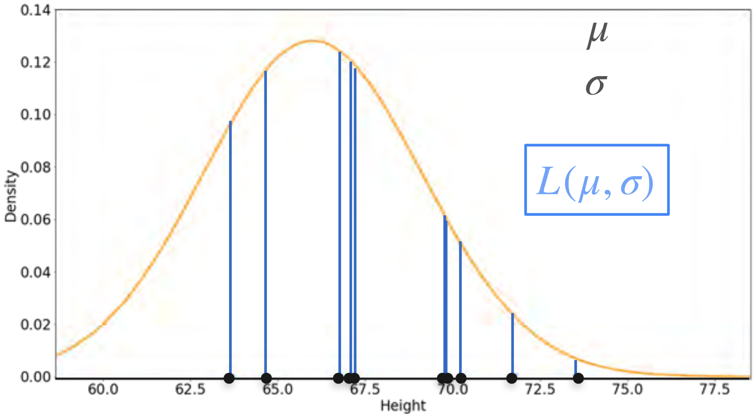
Likelihood $L(\mu; \underline{x}) = \prod_{i=1}^{10} f_\mu(x_i)$



$$\frac{d}{d\mu}\ell(\mu; \mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i - \mu)^2}{\sigma^2}\right]$$

Likelihood $L(\mu;\underline{x}) = \prod_{i=1}^{10} f_\mu(x_i)$

$$= \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right)$$



μ

σ

$\boxed{L(\mu,\sigma)}$

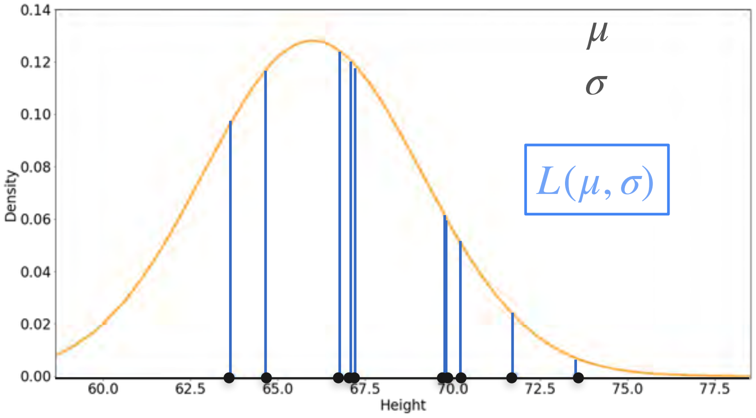
$$\frac{d}{d\mu}\ell(\mu;\mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right]$$

Likelihood $L(\mu; \underline{x}) = \prod_{i=1}^{10} f_\mu(x_i)$

Observations

x_i are fixed!

$$= \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\frac{(\boxed{x_i}-\mu)^2}{\sigma^2}\right)$$



μ

σ

$L(\mu, \sigma)$

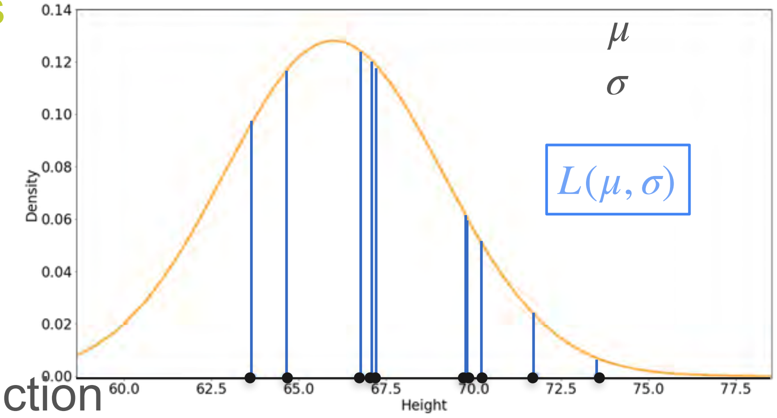
$$\frac{d}{d\mu}\ell(\mu; \mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right]$$

Likelihood $L(\mu; \underline{x}) = \prod_{i=1}^{10} f_\mu(x_i)$

$$= \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\frac{(\boxed{x_i}-\boxed{\mu})^2}{\sigma^2}\right)$$

Observations x_i are fixed!

It's a function of μ

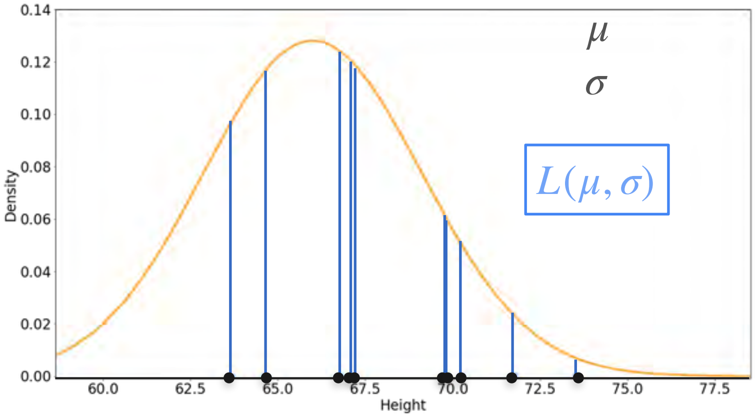


$$\frac{d}{d\mu}\ell(\mu; \mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right]$$

Likelihood $L(\mu; \underline{x}) = \prod_{i=1}^{10} f_\mu(x_i)$

$$= \boxed{\prod_{i=1}^{10} \frac{1}{\sqrt{2\pi}\,\sigma}} \exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}$$

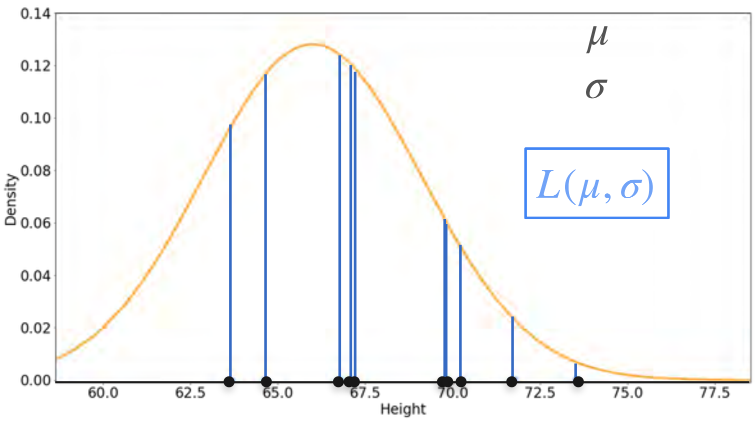


$$\frac{d}{d\mu}\ell(\mu; \mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right]$$

Likelihood $L(\mu; \underline{x}) = \prod_{i=1}^{10} f_\mu(x_i)$

$$= \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi}\,\sigma} \boxed{\exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right)}$$

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10} \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right)$$

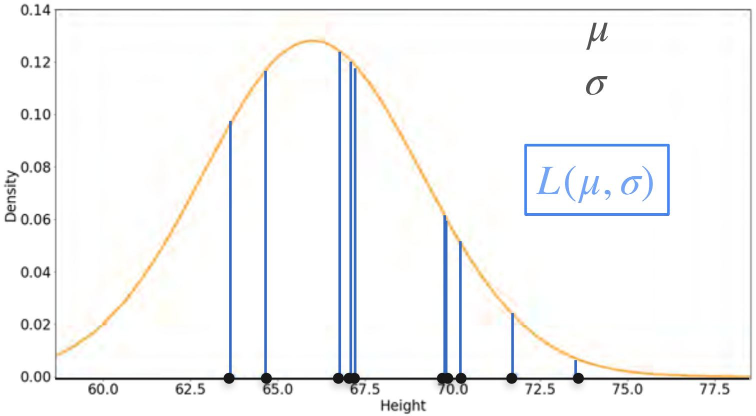


$$\frac{d}{d\mu}\ell(\mu; \mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right]$$

Likelihood $L(\mu; \underline{x}) = \prod_{i=1}^{10} f_\mu(x_i)$

$$= \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\frac{(x_i - \mu)^2}{\sigma^2}\right)$$

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10} \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{10}(x_i - \mu)^2}{\sigma^2}\right)$$

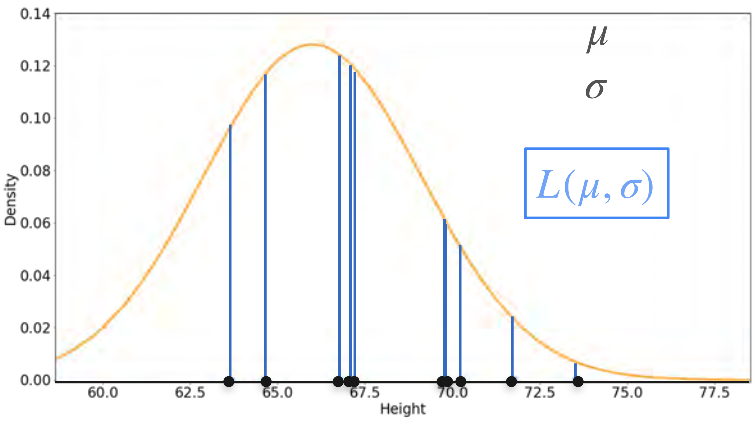


$$\frac{d}{d\mu}\ell(\mu; \mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i - \mu)^2}{\sigma^2}\right]$$

Likelihood $L(\mu; \underline{x}) = \prod_{i=1}^{10} f_\mu(x_i)$

$$= \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10} \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right)$$



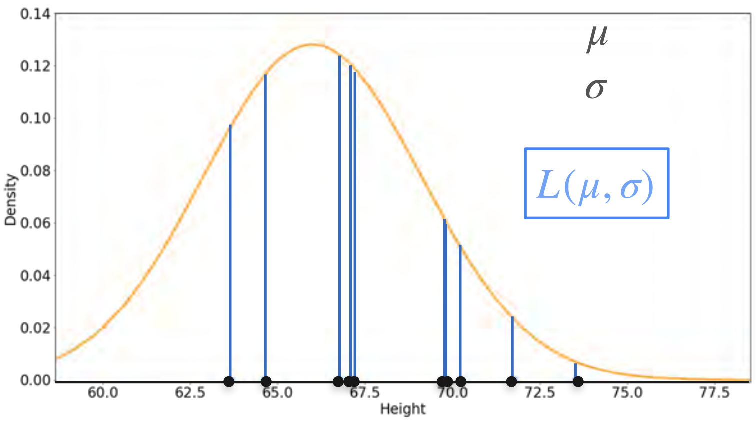
$$\ell(\mu; \mathbf{x}) = \log\left(L(\mu; \mathbf{x})\right)$$

$$\frac{d}{d\mu}\ell(\mu; \mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right]$$

$$\text{Likelihood } L(\mu;\underline{x}) = \prod_{i=1}^{10} f_\mu(x_i)$$

$$= \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10} \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right)$$



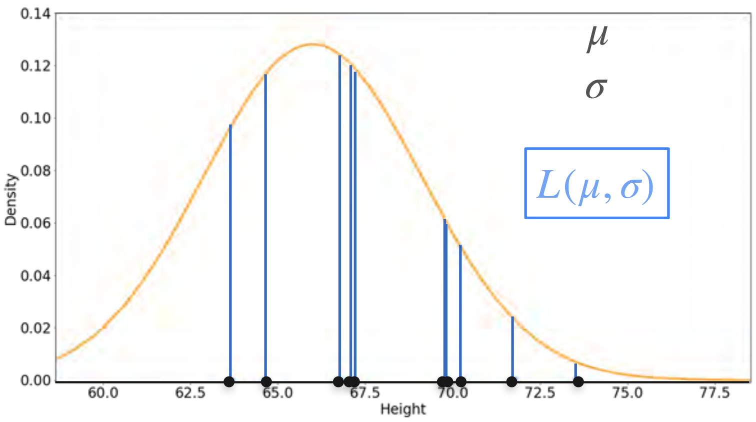
$$\ell(\mu;\mathbf{x}) = \log\left(L(\mu;\mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}$$

$$\frac{d}{d\mu}\ell(\mu;\mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right]$$

$$\text{Likelihood } L(\mu; \underline{x}) = \prod_{i=1}^{10} f_\mu(x_i)$$

$$= \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10} \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right)$$



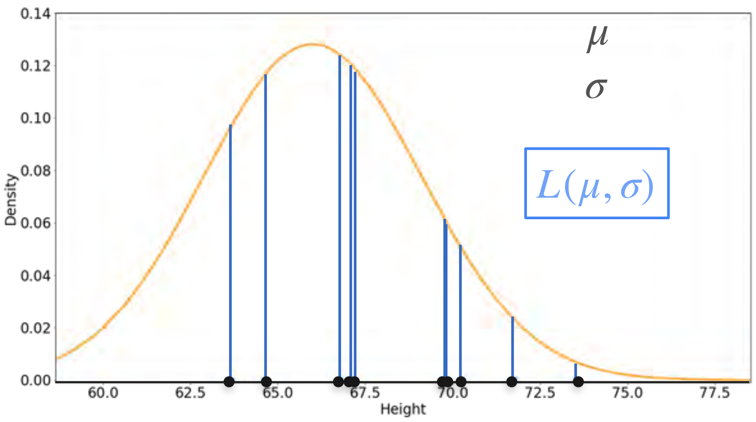
$$\text{Log-Likelihood } \ell(\mu; \mathbf{x}) = \log\left(L(\mu; \mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}$$

$$\frac{d}{d\mu}\ell(\mu; \mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right]$$

Likelihood $L(\mu; \underline{x}) = \prod_{i=1}^{10} f_{\mu}(x_i)$

$$= \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10} \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right)$$



Log-Likelihood $\ell(\mu; \mathbf{x}) = \log\left(L(\mu; \mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}$

Derivative $\frac{d}{d\mu}\ell(\mu; \mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right]$

Log-Likelihood $\ell(\mu; \mathbf{x}) = \log\left(L(\mu; \mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i - \mu)^2}{\sigma^2}$

Derivative $\frac{d}{d\mu}\ell(\mu; \mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i - \mu)^2}{\sigma^2}\right]$

Log-Likelihood $\ell(\mu; \mathbf{x}) = \log\left(L(\mu; \mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i - \mu)^2}{\sigma^2}$

Derivative $\frac{d}{d\mu}\ell(\mu; \mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i - \mu)^2}{\sigma^2}\right] = 0$

Log-Likelihood $\ell(\mu;\mathbf{x}) = \log\left(L(\mu;\mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}$

Derivative $\frac{d}{d\mu}\ell(\mu;\mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right] = 0$

$$= \quad 0$$

Log-Likelihood $\ell(\mu;\mathbf{x}) = \log\left(L(\mu;\mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}$

Derivative $\frac{d}{d\mu}\ell(\mu;\mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right] = 0$

$$= \quad 0 \quad -\frac{1}{2}\frac{1}{\sigma^2}\sum_{i=1}^{10}2(x_i-\mu)(-1)$$

Log-Likelihood $\ell(\mu;\mathbf{x}) = \log\left(L(\mu;\mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}$

Derivative $\frac{d}{d\mu}\ell(\mu;\mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right] = 0$

$$= \quad 0 \quad \boxed{-\frac{1}{2}\frac{1}{\sigma^2}\sum_{i=1}^{10} 2(x_i-\mu)(-1)}$$

This needs to be zero

Log-Likelihood $\ell(\mu; \mathbf{x}) = \log\left(L(\mu; \mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i - \mu)^2}{\sigma^2}$

Derivative $\frac{d}{d\mu}\ell(\mu; \mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i - \mu)^2}{\sigma^2}\right] = 0$

$$= \quad 0 \quad \boxed{-\cancel{\frac{1}{2}}\frac{1}{\sigma^2}\sum_{i=1}^{10} 2(x_i - \mu)(-1)}$$

This needs to be zero

Log-Likelihood $\ell(\mu;\mathbf{x}) = \log\left(L(\mu;\mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}$

Derivative $\frac{d}{d\mu}\ell(\mu;\mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right] = 0$

$$= \quad 0 \quad \boxed{-\cancel{\frac{1}{2}\frac{1}{\sigma^2}}\sum_{i=1}^{10}\cancel{2}(x_i-\mu)(\cancel{-1})}$$

This needs to be zero

Log-Likelihood $\ell(\mu;\mathbf{x}) = \log\left(L(\mu;\mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}$

Derivative $\frac{d}{d\mu}\ell(\mu;\mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right] = 0$

$$= \quad 0 \quad \boxed{-\cancel{\frac{1}{2}\frac{1}{\sigma^2}}\sum_{i=1}^{10}\cancel{2}(x_i-\mu)(\cancel{-1})}$$

This needs to be zero

$$\sum_{i=1}^{10}(x_i-\mu) = 0$$

Log-Likelihood $\ell(\mu;\mathbf{x}) = \log\left(L(\mu;\mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}$

Derivative $\frac{d}{d\mu}\ell(\mu;\mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma^2}\right] = 0$

$$= \quad 0 \quad \boxed{\cancel{-\frac{1}{2}\frac{1}{\sigma^2}}\sum_{i=1}^{10}\cancel{2}(x_i-\mu)(\cancel{-}1)}$$

This needs to be zero

$$\sum_{i=1}^{10}(x_i-\mu) = 0$$

$$\left(\sum_{i=1}^{10}x_i\right) - 10\mu = 0$$

Log-Likelihood $\ell(\mu; \mathbf{x}) = \log\left(L(\mu; \mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i - \mu)^2}{\sigma^2}$

Derivative $\frac{d}{d\mu}\ell(\mu; \mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{10}\right) - \frac{1}{2}\frac{\sum_{i=1}^{10}(x_i - \mu)^2}{\sigma^2}\right] = 0$

$$= \quad 0 \quad \boxed{-\cancel{\frac{1}{2}\frac{1}{\sigma^2}} \sum_{i=1}^{10} \cancel{2}(x_i - \mu)(\cancel{-1})}$$

This needs to be zero

$$\sum_{i=1}^{10}(x_i - \mu) = 0$$

$$\left(\sum_{i=1}^{10} x_i\right) - 10\mu = 0 \quad \Rightarrow \quad \mu = \frac{1}{10}\sum_{i=1}^{10} x_i$$

Log-Likelihood $\ell(\mu; \mathbf{x}) = \log\left(L(\mu;\mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\right) - \frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}$

Derivative $\frac{d}{d\mu}\ell(\mu;\mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\right) - \frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}\right] = 0$

$$= \quad 0 \quad \boxed{-\frac{\cancel{1}\;1}{\cancel{2}\;\sigma^2}\sum_{i=1}^{n}\cancel{2}(x_i-\mu)(\cancel{-}1)}$$

This needs to be zero

$$\sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\left(\sum_{i=1}^{n}x_i\right) - n\,\mu = 0 \quad \Rightarrow \quad \mu = \frac{1}{n}\sum_{i=1}^{n}x_i$$

Log-Likelihood $\ell(\mu; \mathbf{x}) = \log\left(L(\mu; \mathbf{x})\right) = \log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\right) - \frac{1}{2}\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{\sigma^2}$

Derivative $\frac{d}{d\mu}\ell(\mu; \mathbf{x}) = \frac{d}{d\mu}\left[\log\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\right) - \frac{1}{2}\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{\sigma^2}\right] = 0$

$$= \quad 0 \quad \boxed{\cancel{-\frac{1}{2}\frac{1}{\sigma^2}}\sum_{i=1}^{n}\cancel{2}(x_i - \mu)(\cancel{-1})}$$

This needs to be zero

$$\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\left(\sum_{i=1}^{n} x_i\right) - n\,\mu = 0 \quad \Rightarrow \quad \mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The best distribution is the one where the **mean** of the distribution is the **mean** of the sample

Maximum Likelihood: Gaussian Example

Maximum Likelihood: Gaussian Example

Heights	66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

Maximum Likelihood: Gaussian Example

Heights	66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

$$\hat{\mu} = \frac{66.75 + 70.24 + 67.19 + 67.09 + 63.65 + 64.64 + 69.81 + 69.79 + 73.52 + 71.74}{10}$$

Maximum Likelihood: Gaussian Example

Heights	66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

$$\hat{\mu} = \frac{66.75 + 70.24 + 67.19 + 67.09 + 63.65 + 64.64 + 69.81 + 69.79 + 73.52 + 71.74}{10}$$

$$= 68.442$$

Maximum Likelihood: Gaussian Example

Heights	66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

Maximum Likelihood: Gaussian Example

What about the variance?

Heights	66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

Maximum Likelihood: Gaussian Example

What about the variance?

Heights	66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

If you repeat the process, assuming unknown variance, you'll get that

Maximum Likelihood: Gaussian Example

What about the variance?

Heights	66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

If you repeat the process, assuming unknown variance, you'll get that

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{10} (x_i - \overline{x})^2}{10}$$

Maximum Likelihood: Gaussian Example

What about the variance?

Heights	66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

If you repeat the process, assuming unknown variance, you'll get that

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{10} (x_i - \bar{x})^2}{10} = \frac{1}{10}\Big((66.75 - 68.442)^2 + (70.24 - 68.442)^2 + (67.19 - 68.442)^2 +$$

$$(67.09 - 68.442)^2 + (63.65 - 68.442)^2 + (64.64 - 68.442)^2 +$$

$$(69.81 - 68.442)^2 + (69.79 - 68.442)^2 + (73.52 - 68.442)^2 +$$

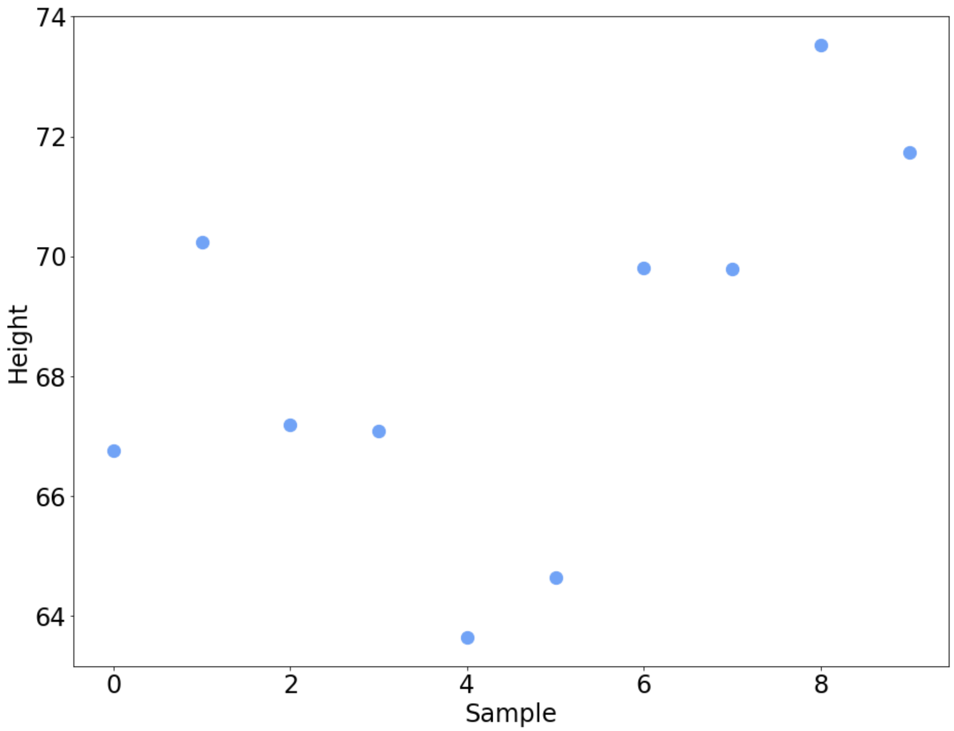
$$(71.74 - 68.442)^2\Big)$$

$$= 8.72$$

Maximum Likelihood: Gaussian Example

X = "Height of an 18 year old"

$$\mathcal{N}(\mu, \sigma)$$

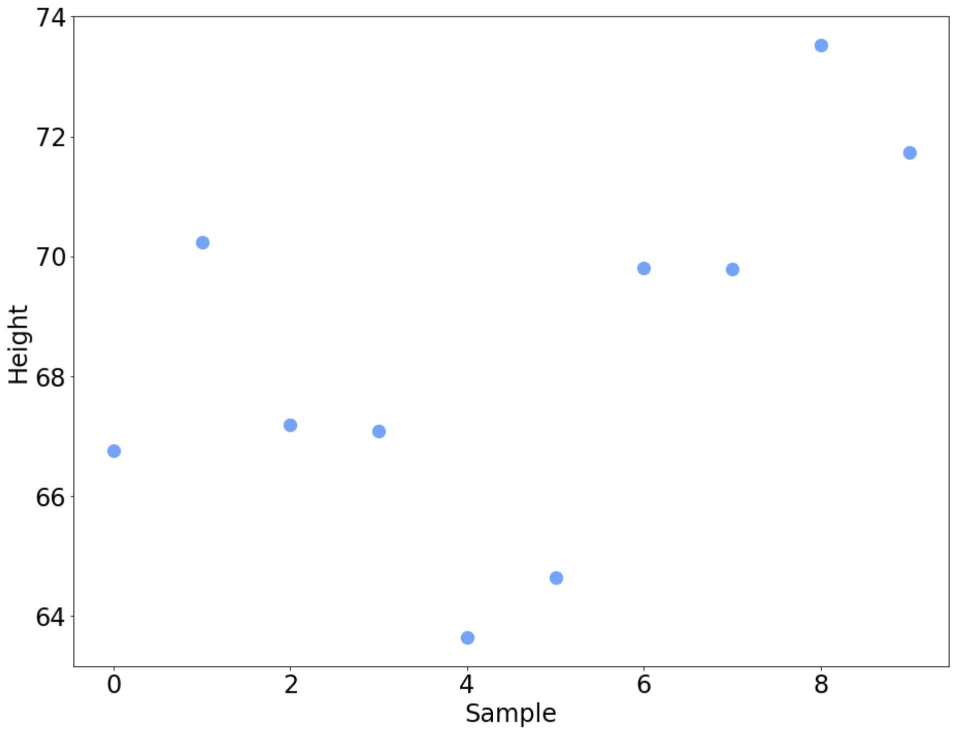


Maximum Likelihood: Gaussian Example

X = "Height of an 18 year old"

$\mathcal{N}(\mu, \sigma)$

$\hat{\mu} = 68.442$



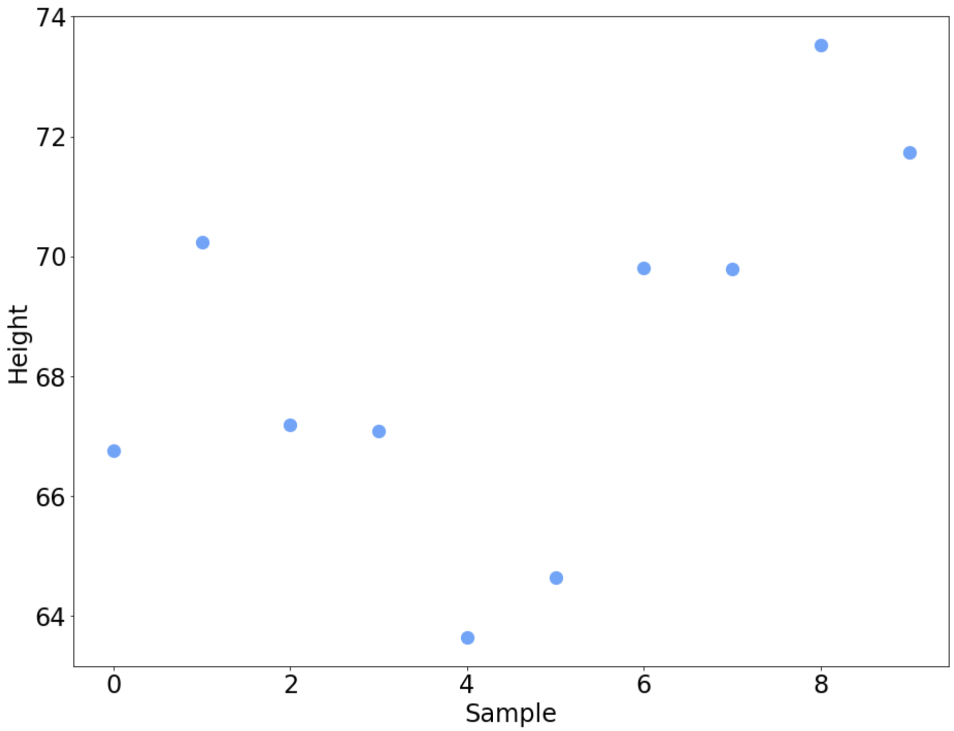
Maximum Likelihood: Gaussian Example

X = "Height of an 18 year old"

$$\mathcal{N}(\mu, \sigma)$$

$$\hat{\mu} = 68.442$$

$$\hat{\sigma} = 8.72$$



DeepLearning.AI

Point Estimation

MLE: Linear Regression

Maximum Likelihood

Maximum Likelihood

Data

Maximum Likelihood

Model 1

Data

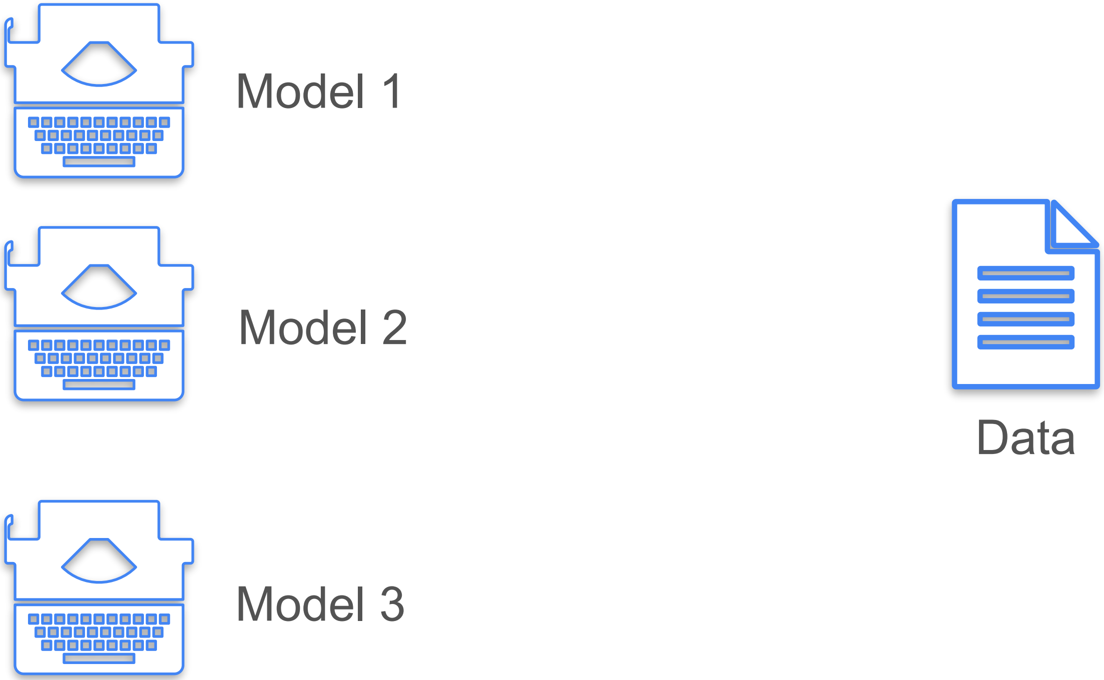
Maximum Likelihood

Model 1

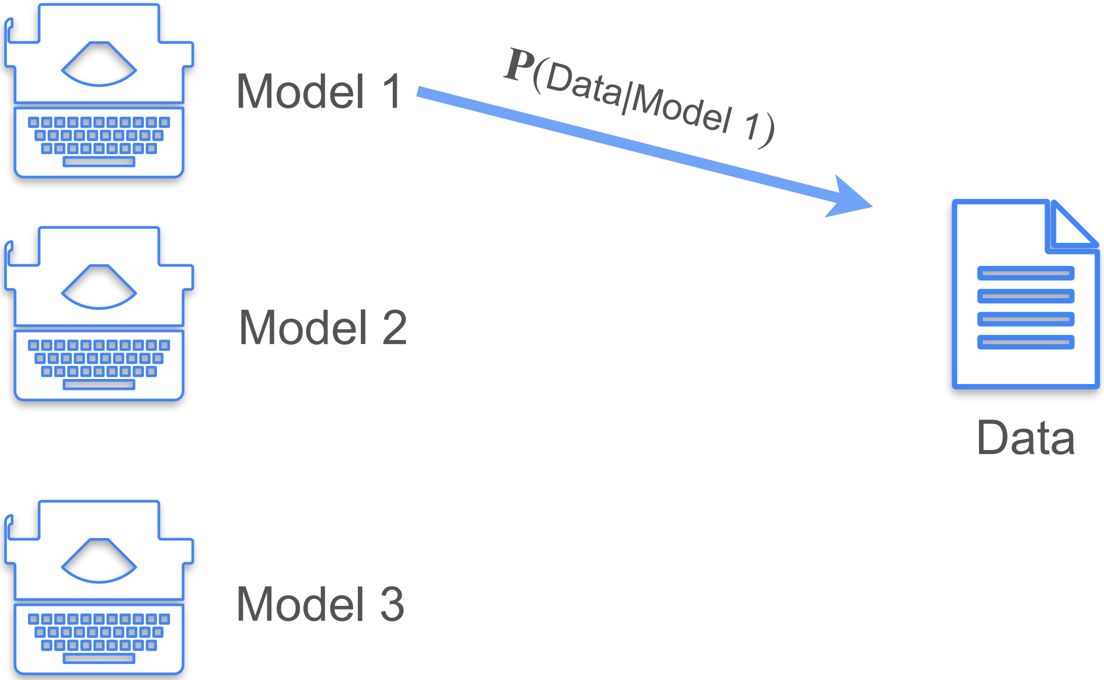
Model 2

Data

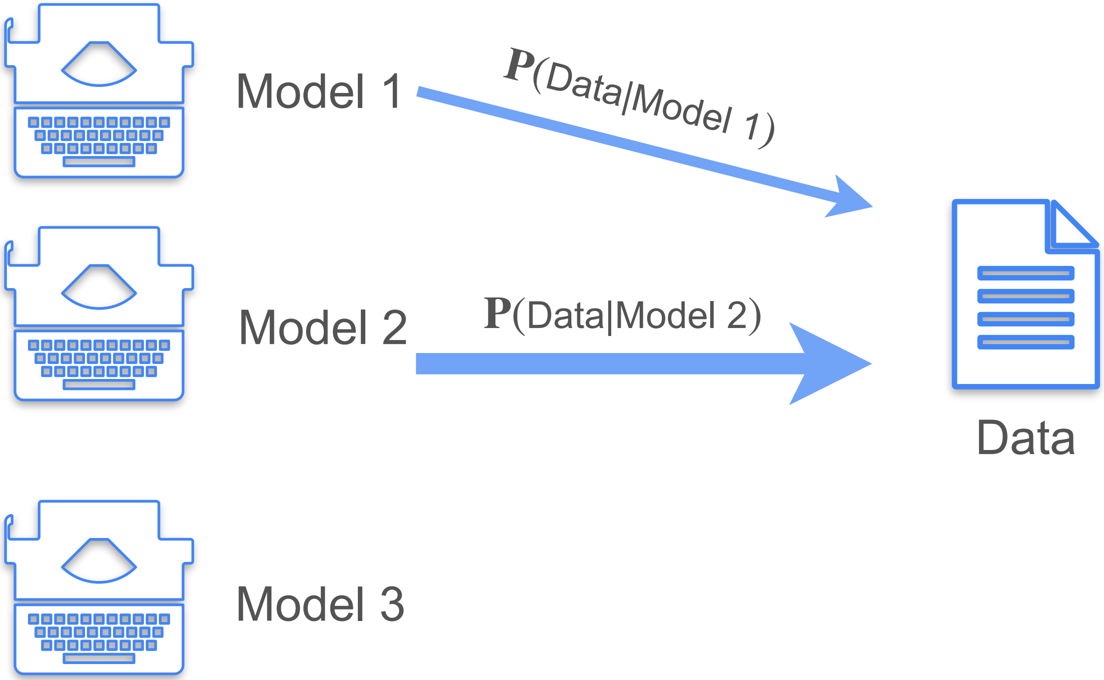
Maximum Likelihood



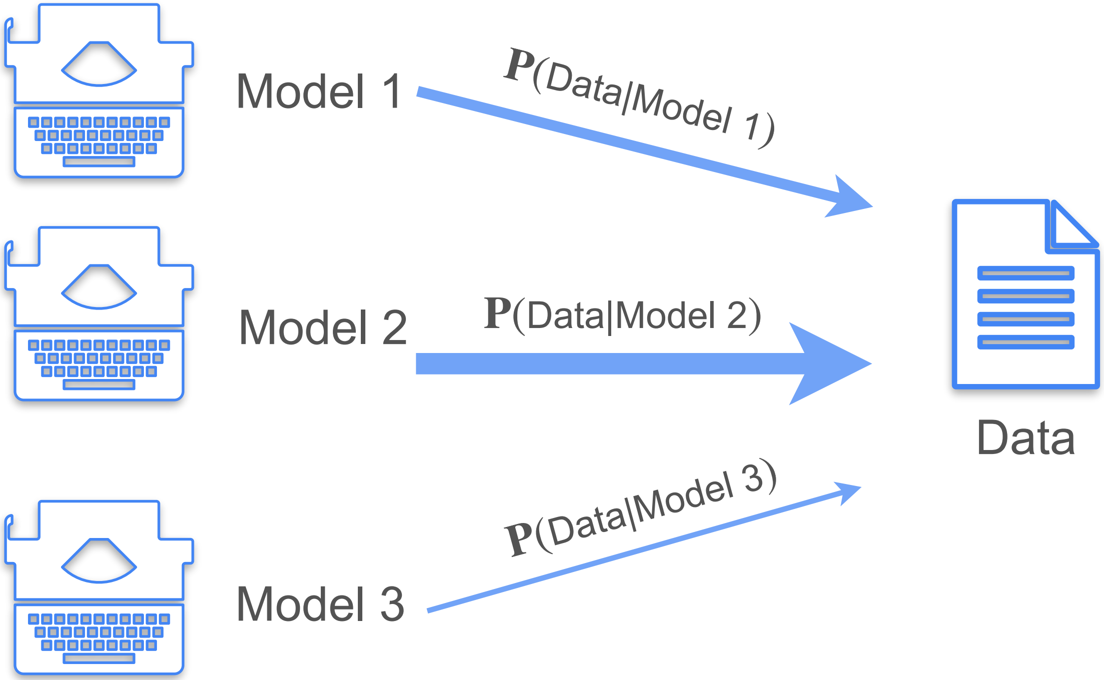
Maximum Likelihood



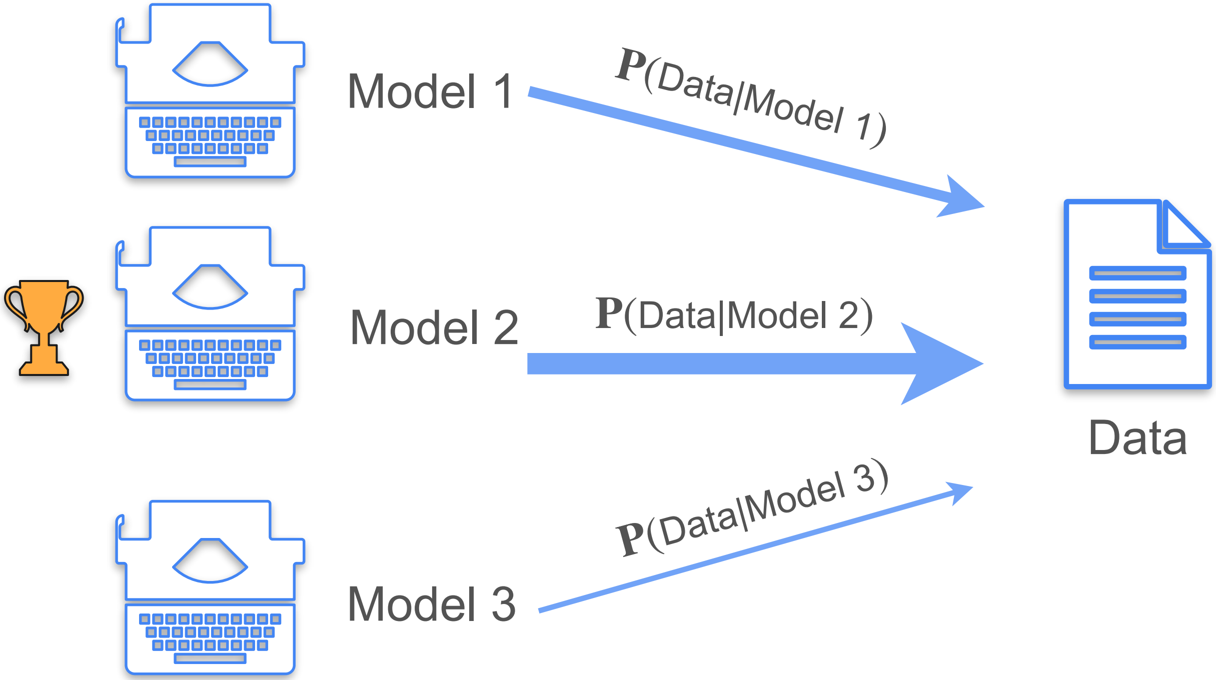
Maximum Likelihood



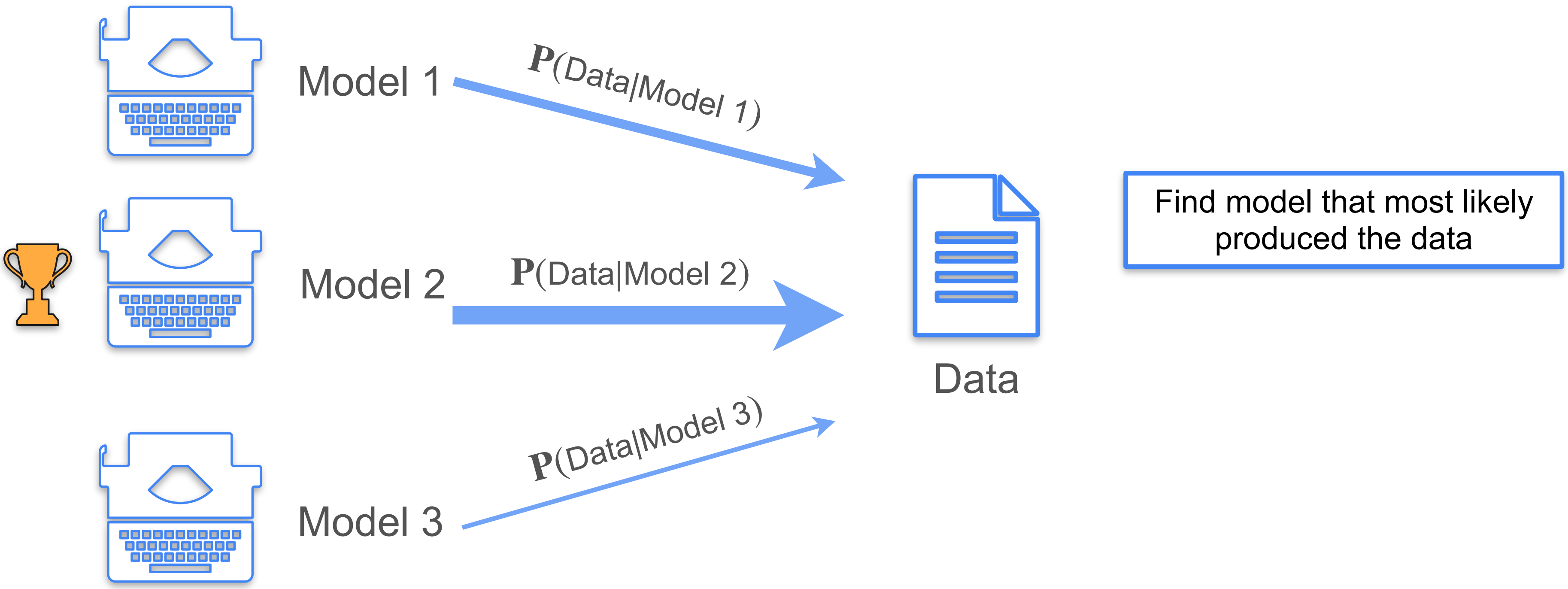
Maximum Likelihood



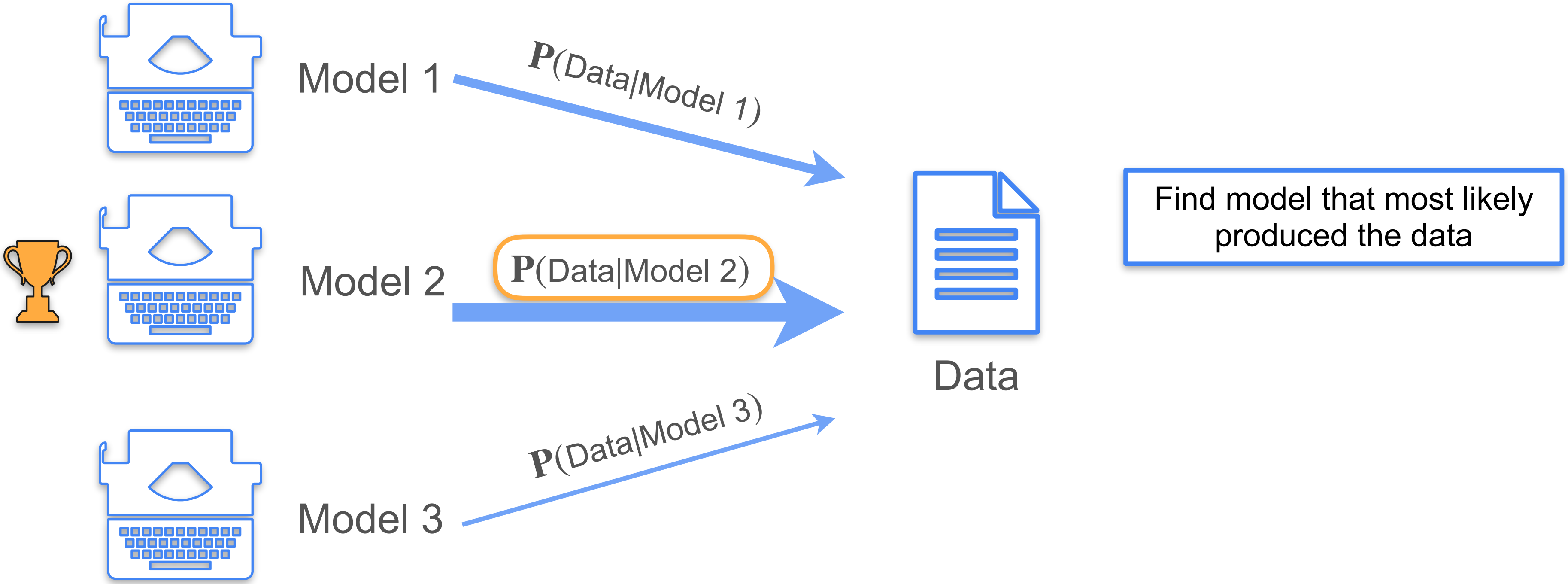
Maximum Likelihood



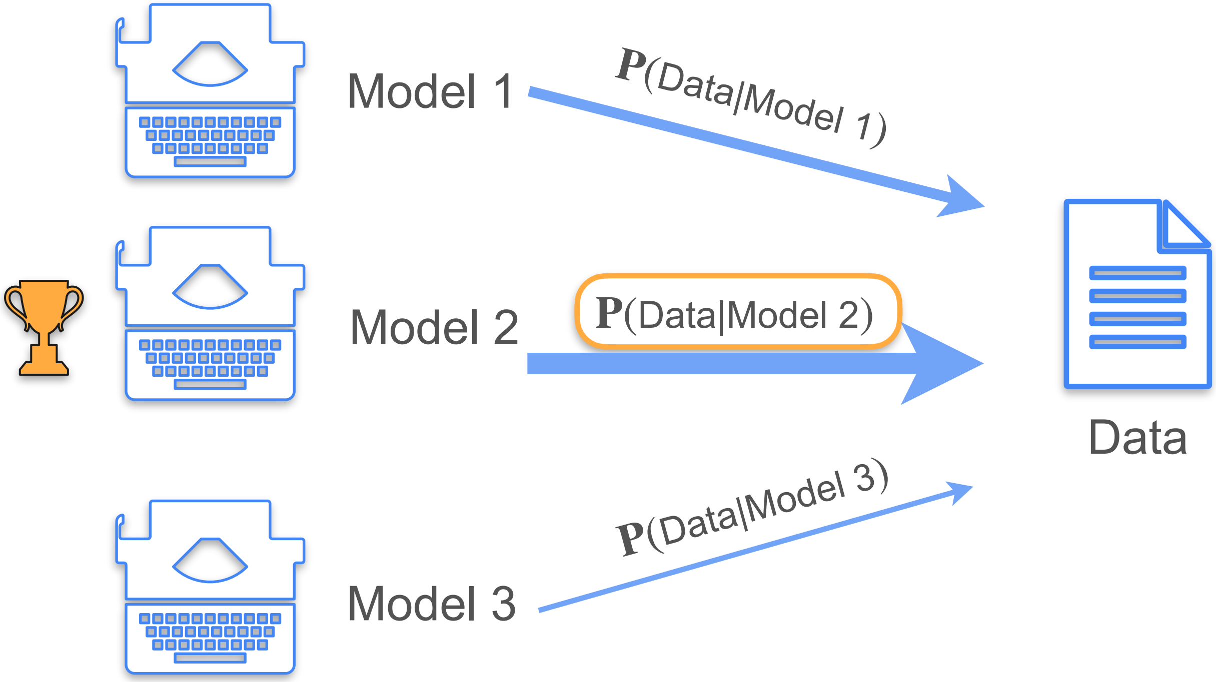
Maximum Likelihood



Maximum Likelihood



Maximum Likelihood

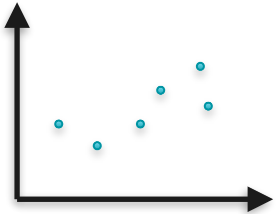


Find model that most likely produced the data

Maximize
$\mathbf{P}(\text{Data}|\text{Model})$

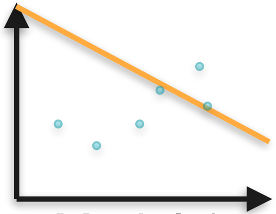
Example: Linear Regression

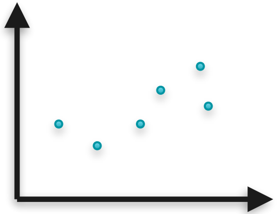
Example: Linear Regression



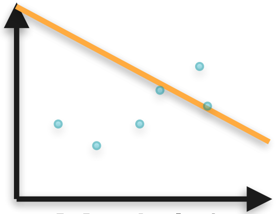
Data

Example: Linear Regression

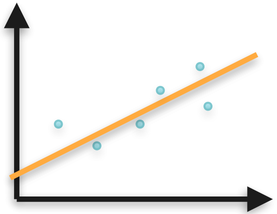

Model 1


Data

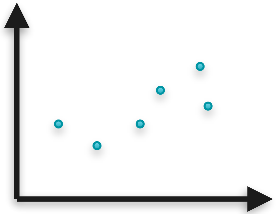
Example: Linear Regression



Model 1

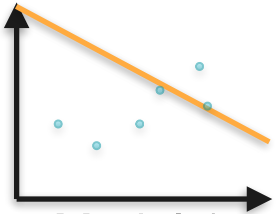


Model 2

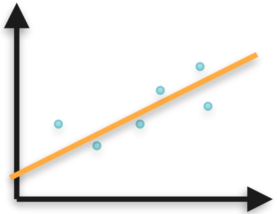


Data

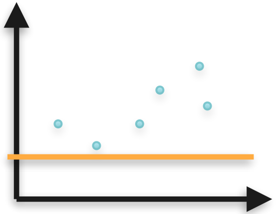
Example: Linear Regression



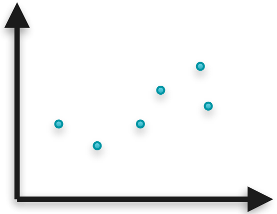
Model 1



Model 2

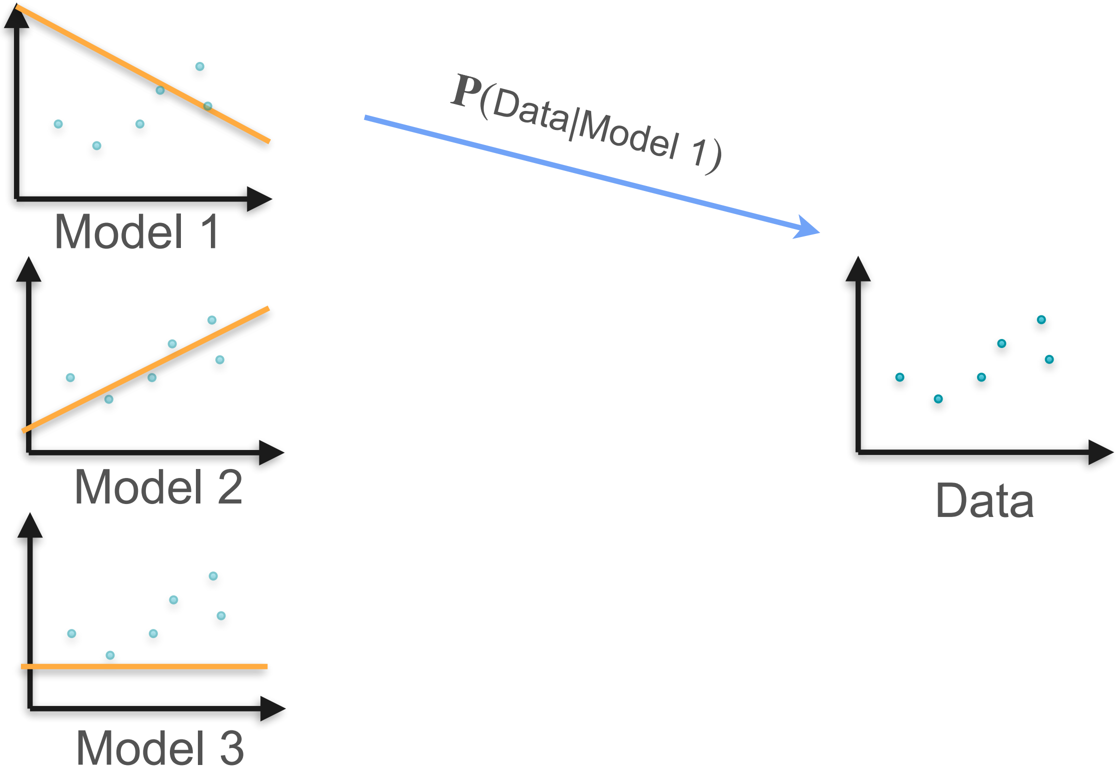


Model 3

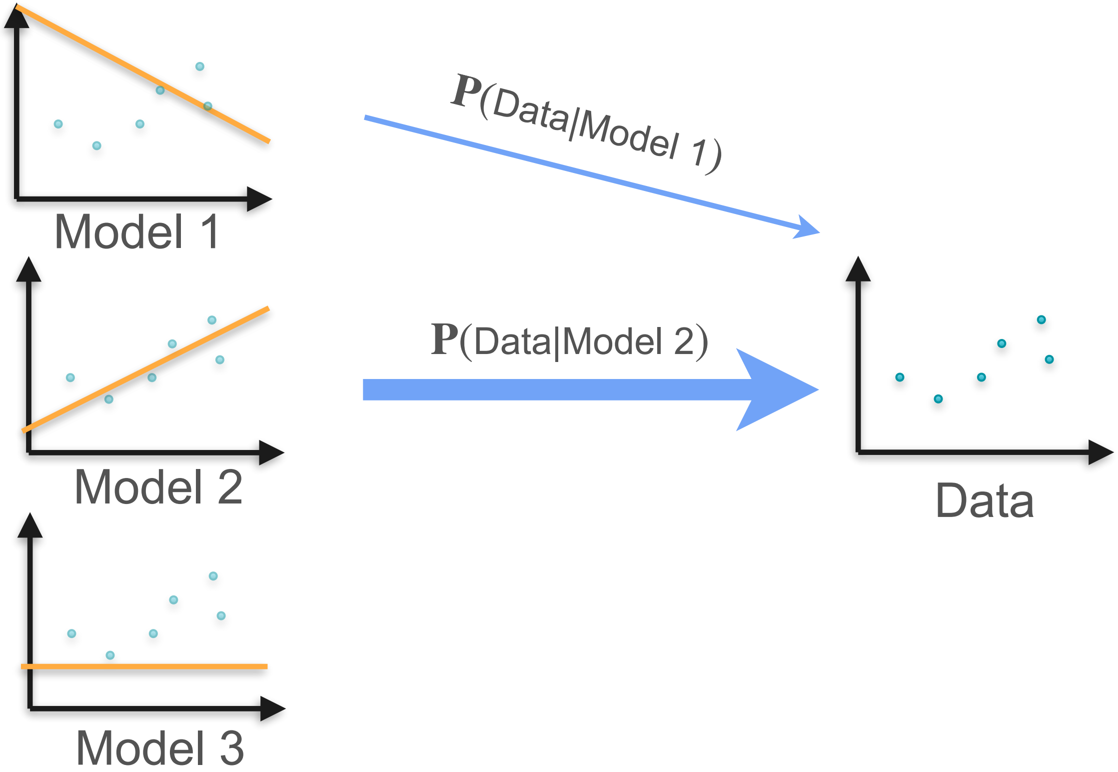


Data

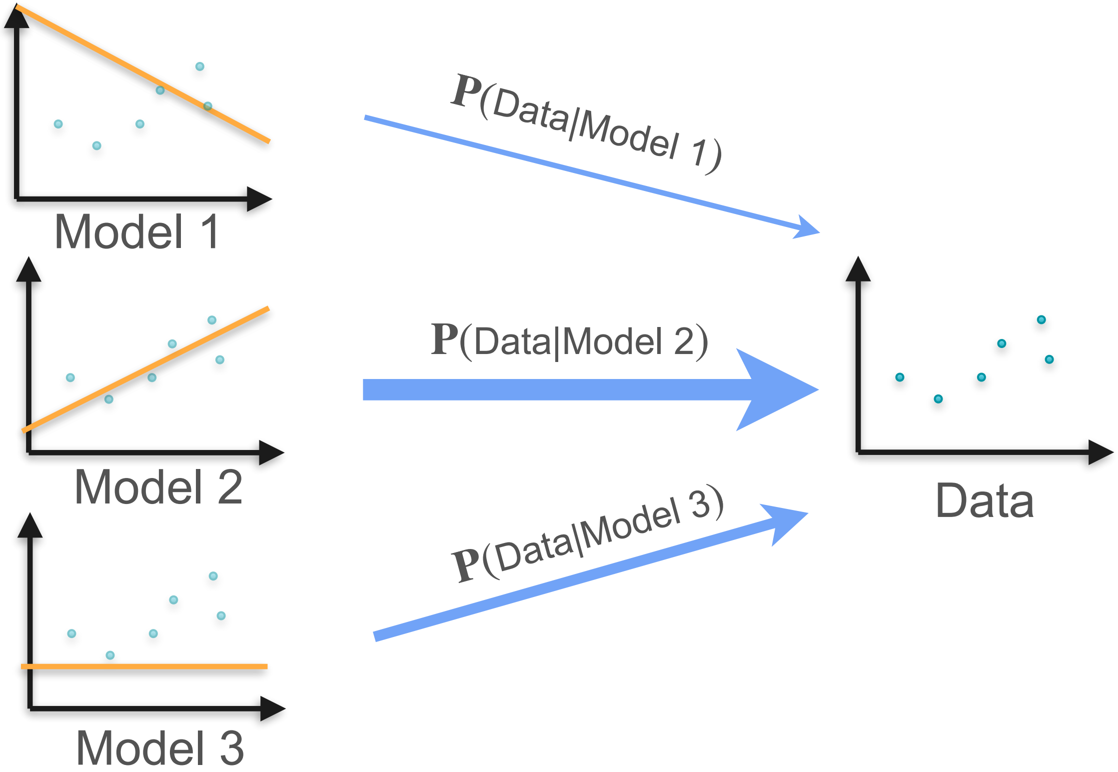
Example: Linear Regression



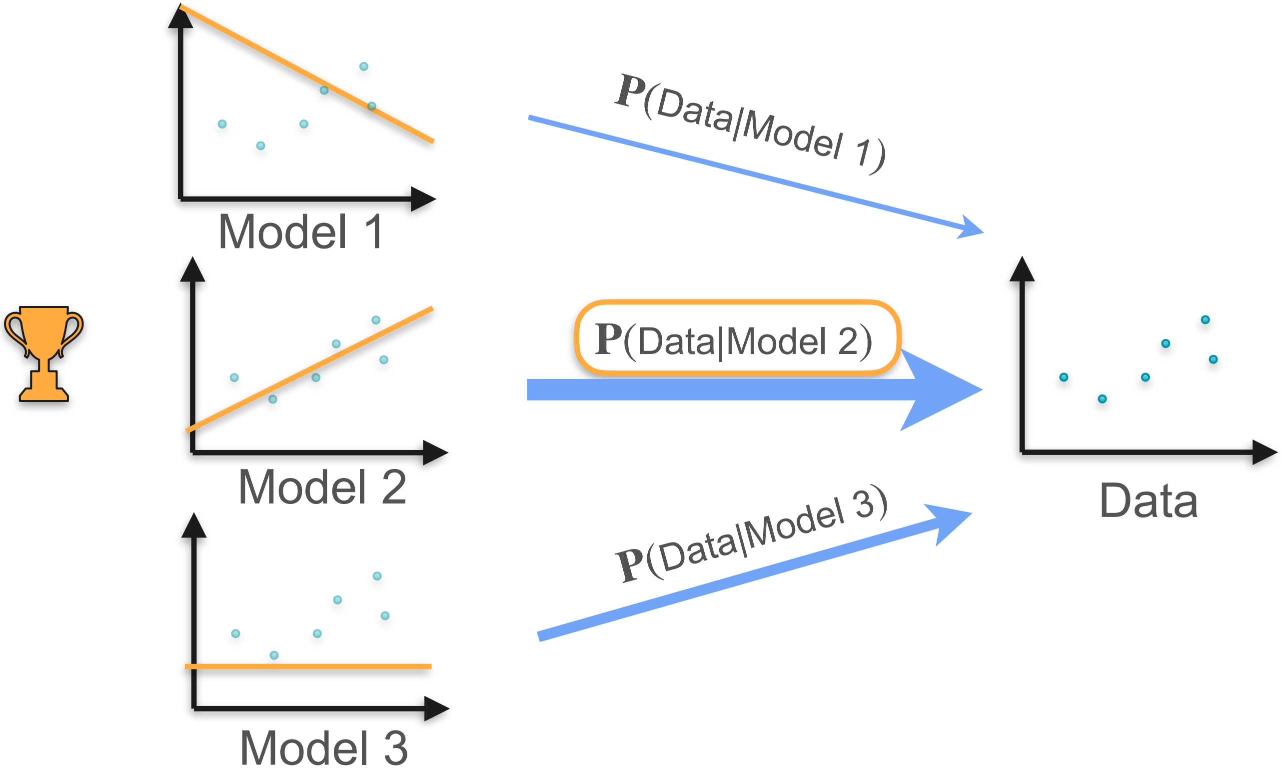
Example: Linear Regression



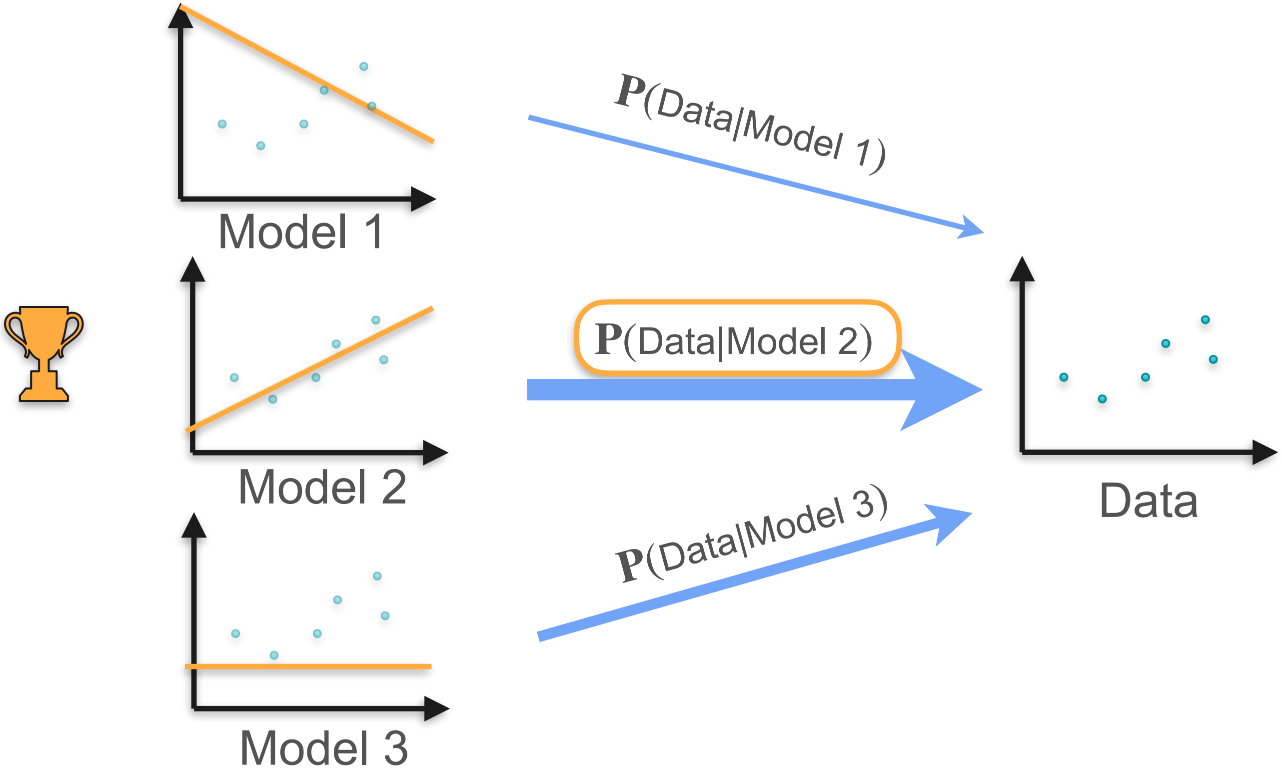
Example: Linear Regression



Example: Linear Regression

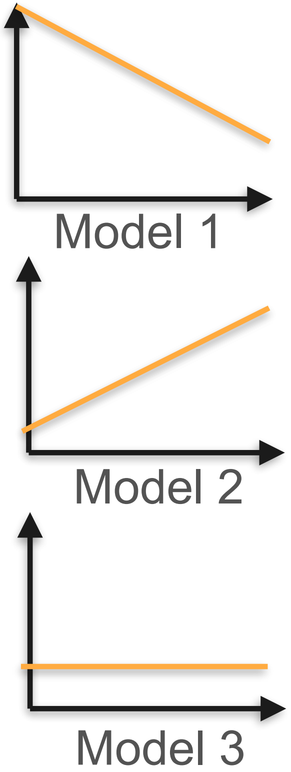


Example: Linear Regression

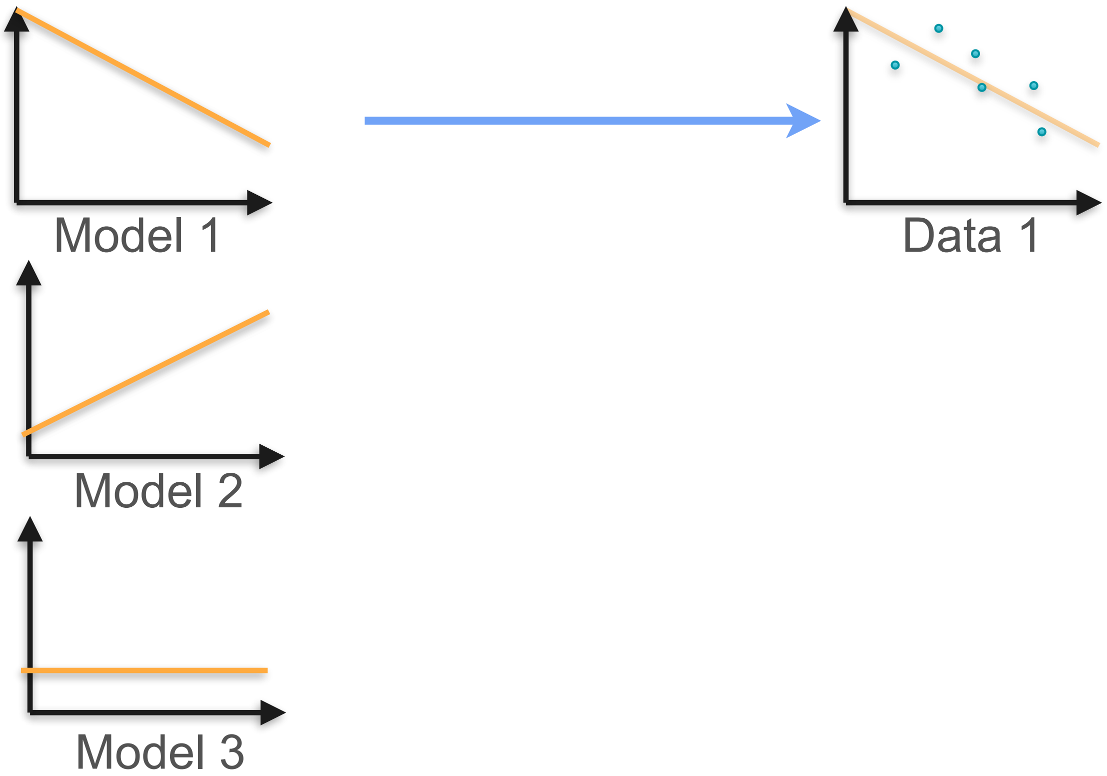


Find model that most likely produced the data

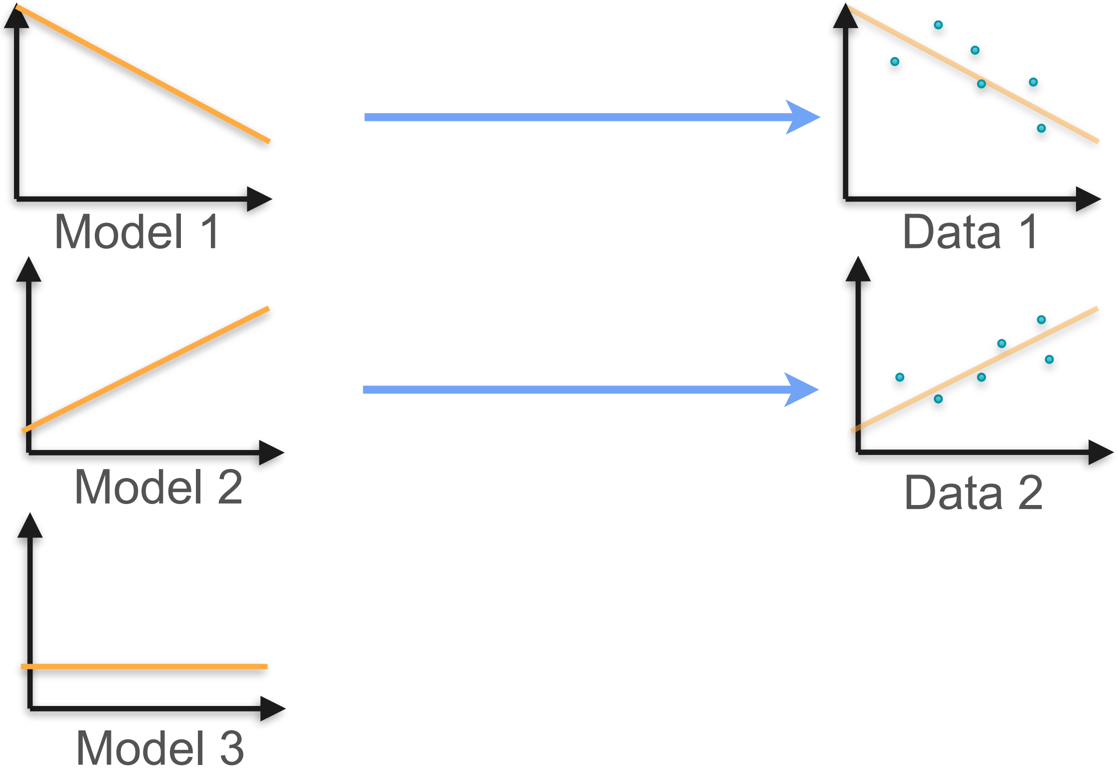
How Exactly Does a Line Produce Points?



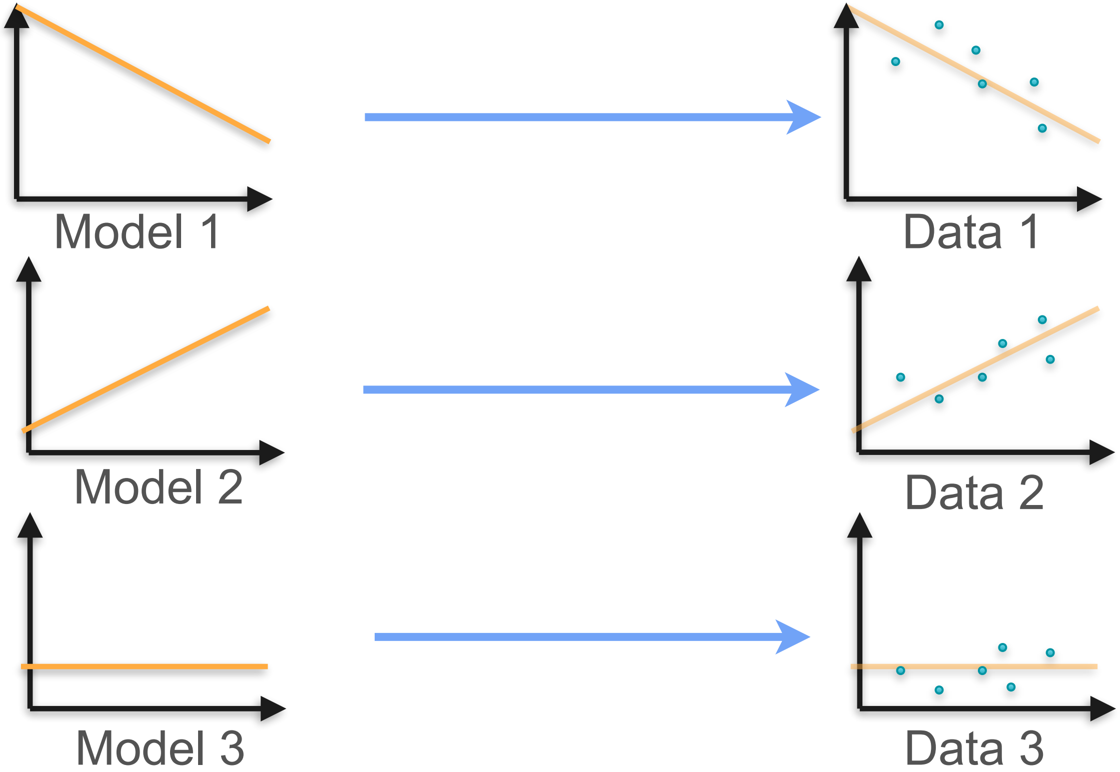
How Exactly Does a Line Produce Points?



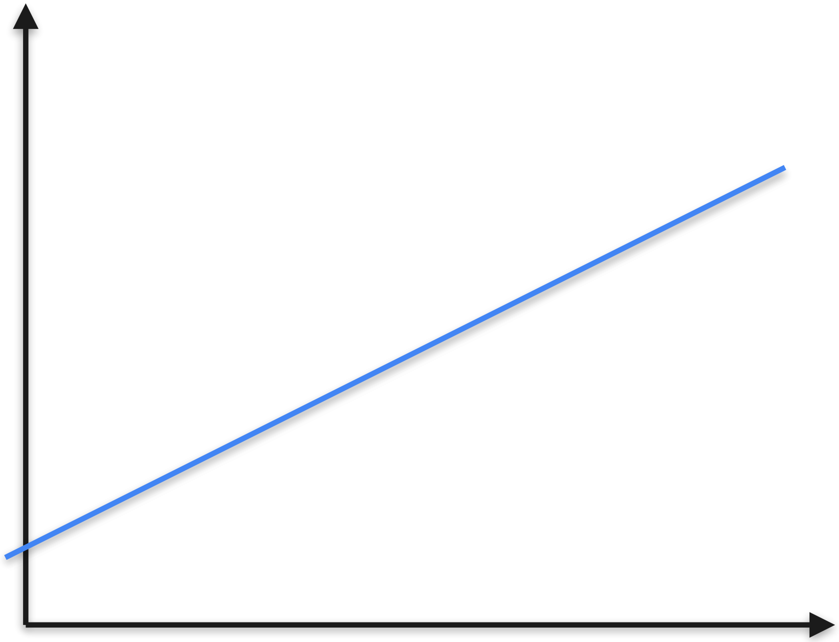
How Exactly Does a Line Produce Points?



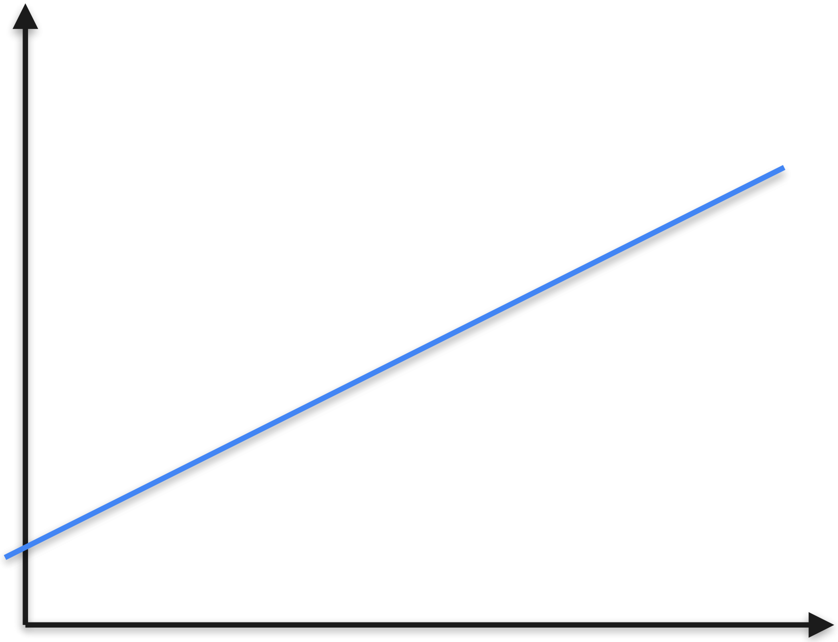
How Exactly Does a Line Produce Points?



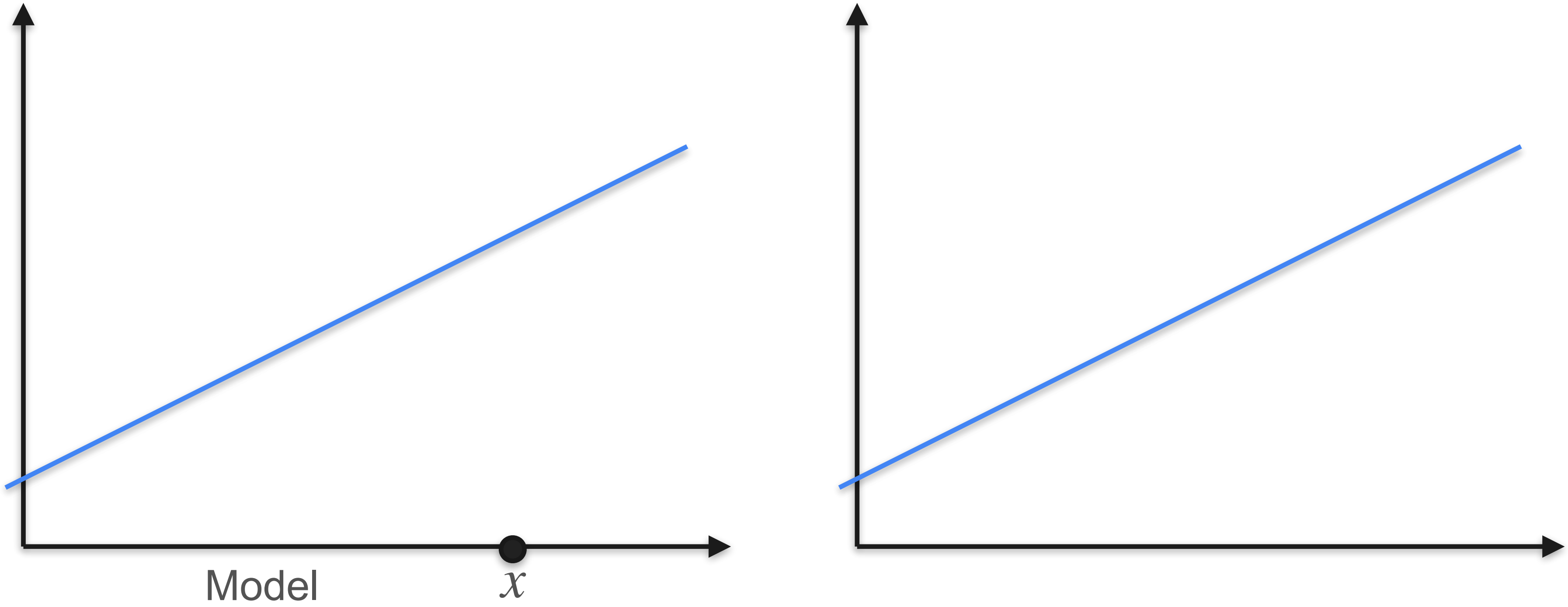
How Exactly Does a Line Produce Points?



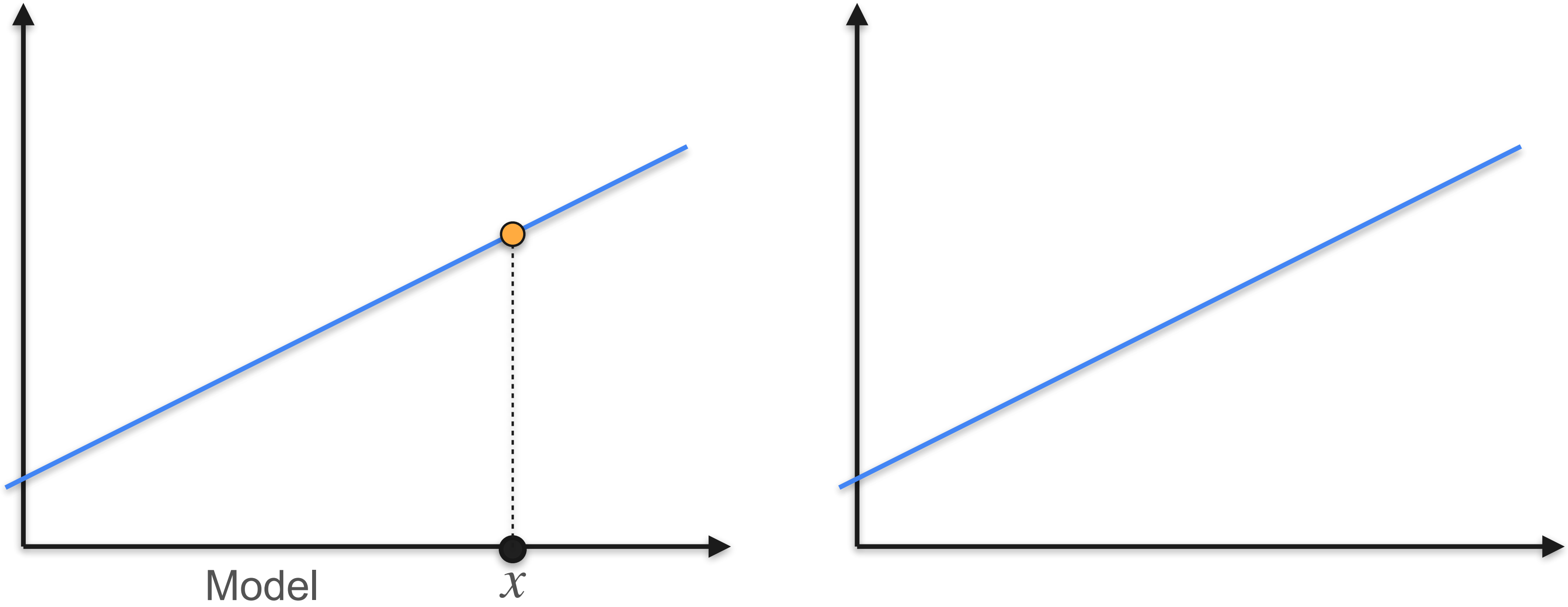
Model



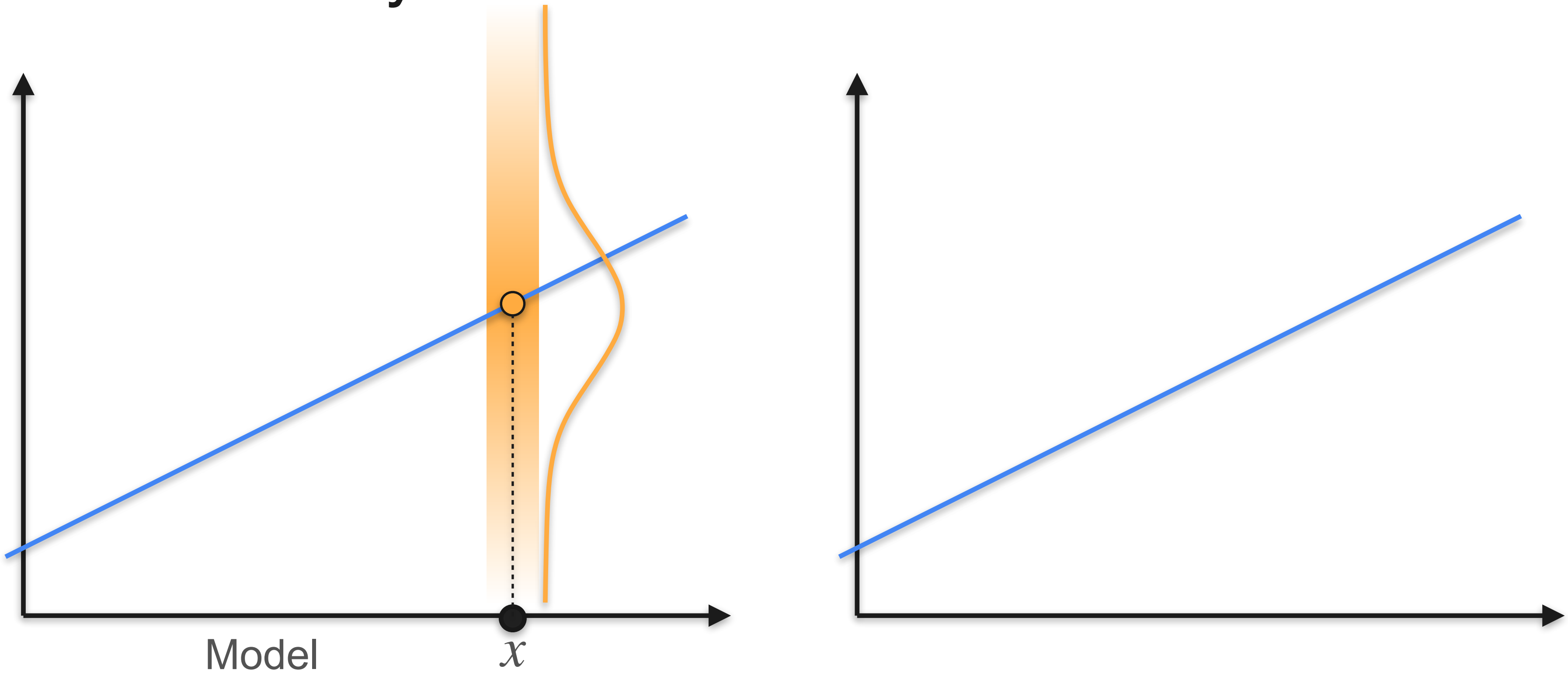
How Exactly Does a Line Produce Points?



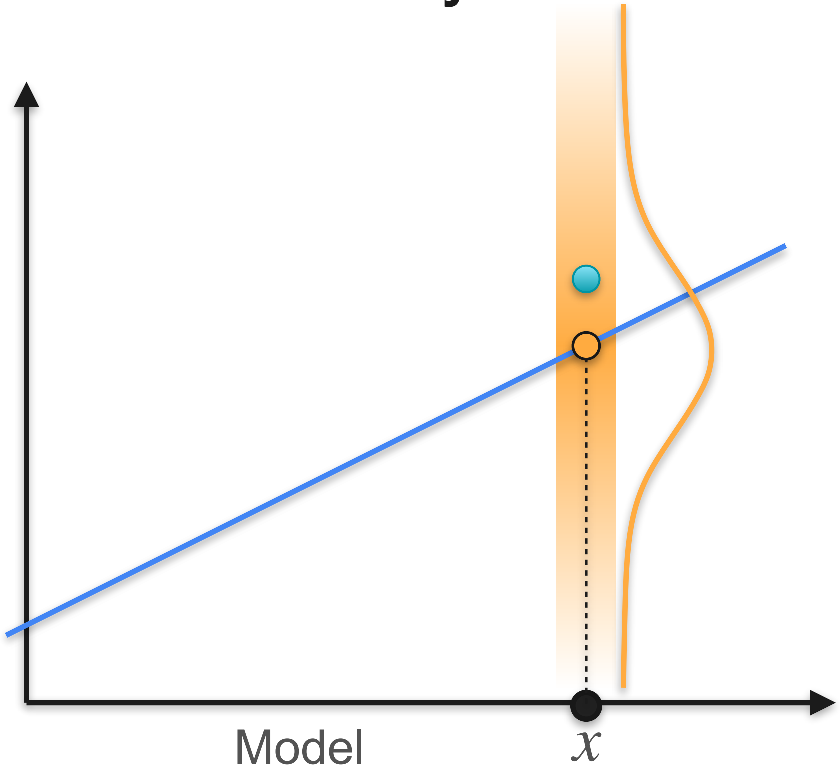
How Exactly Does a Line Produce Points?

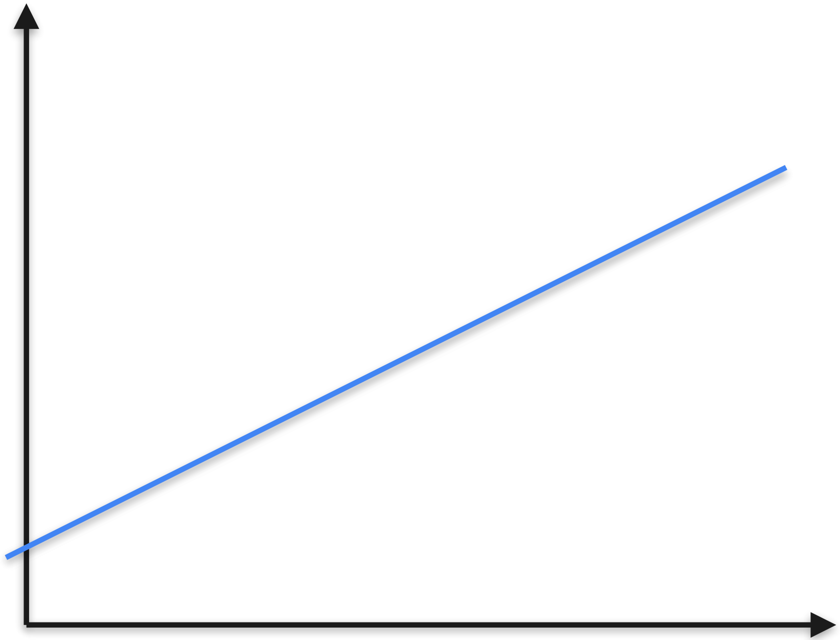


How Exactly Does a Line Produce Points?

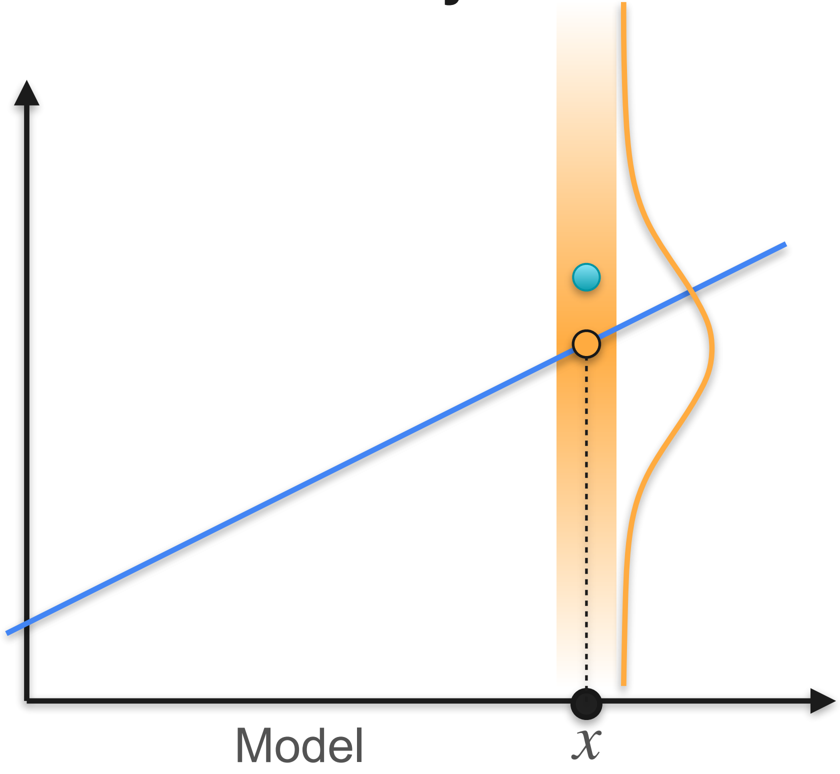


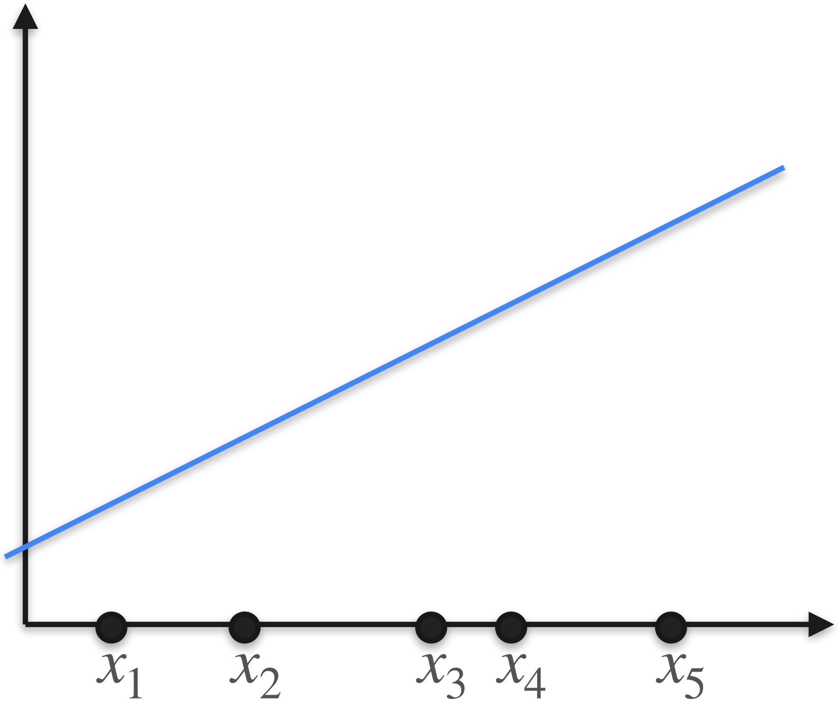
How Exactly Does a Line Produce Points?





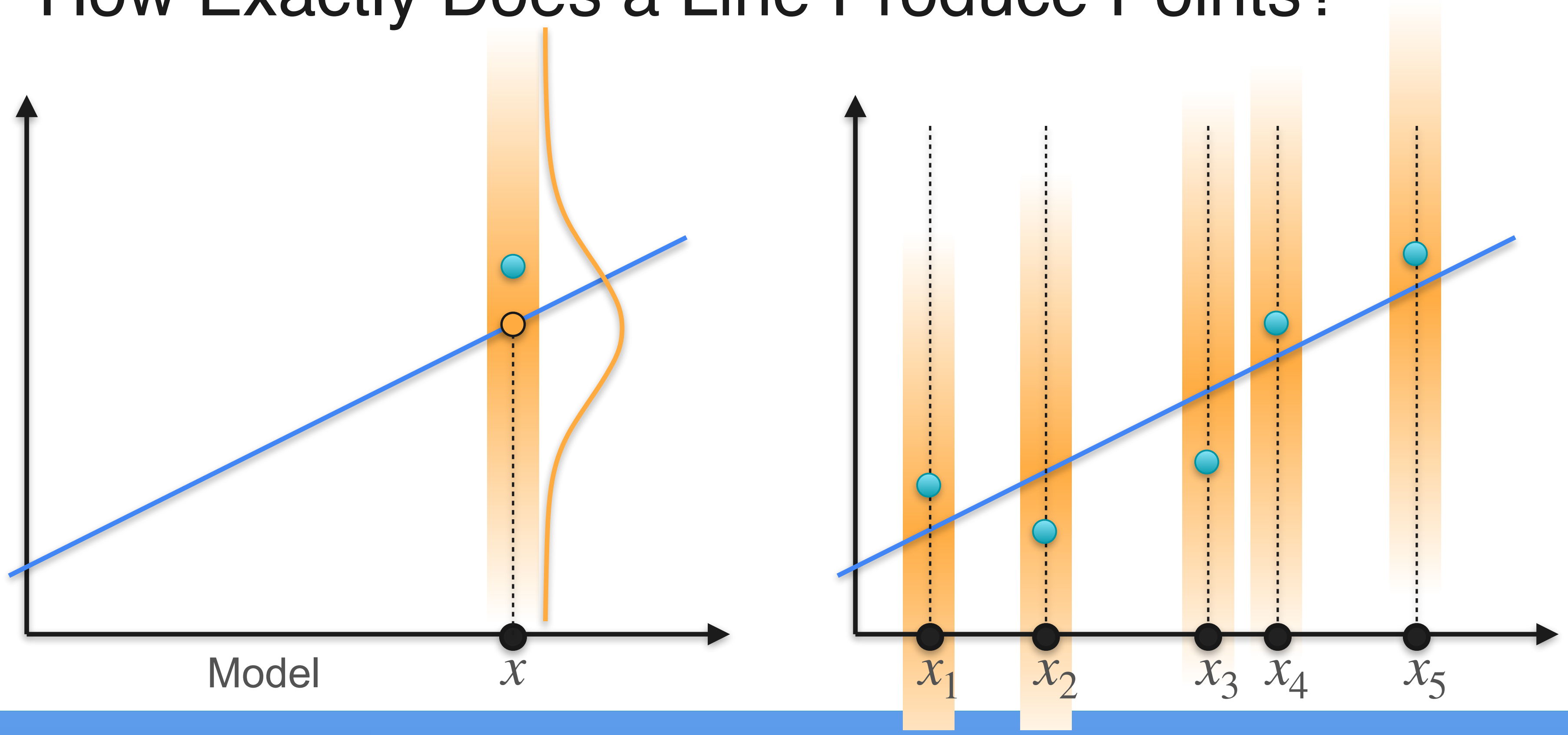
How Exactly Does a Line Produce Points?



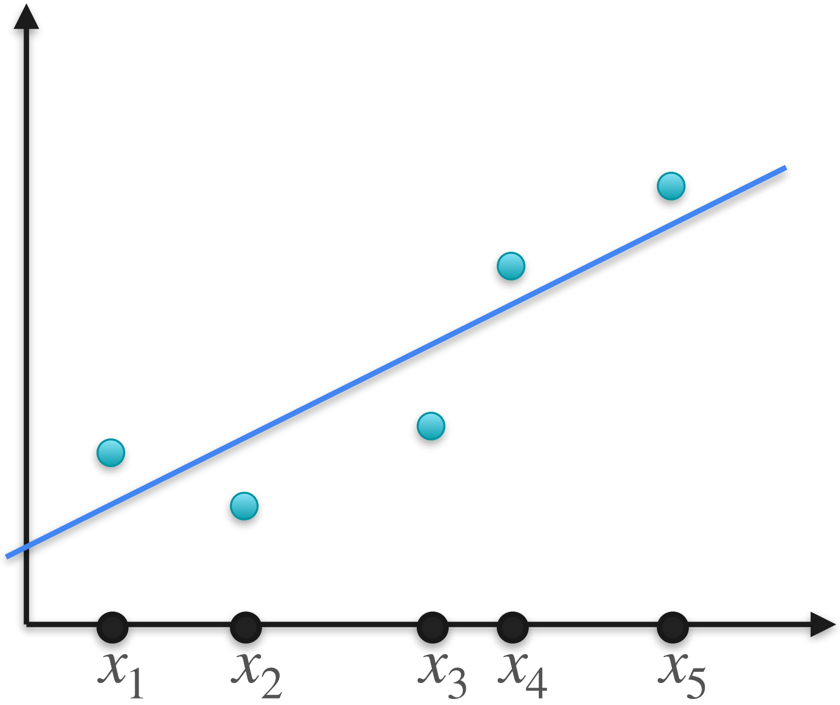


DeepLearning.AI

How Exactly Does a Line Produce Points?

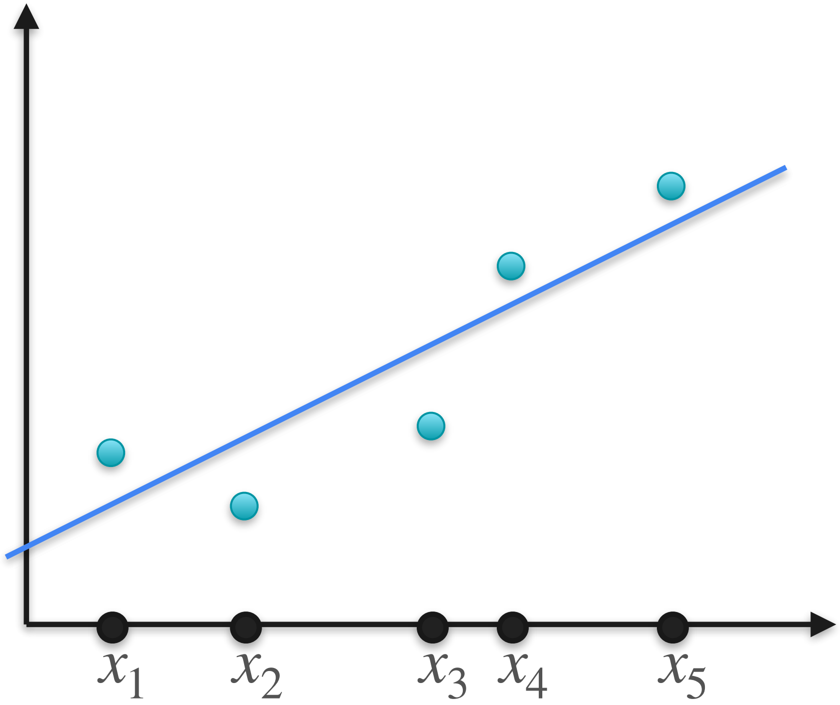


Linear Regression



Linear Regression

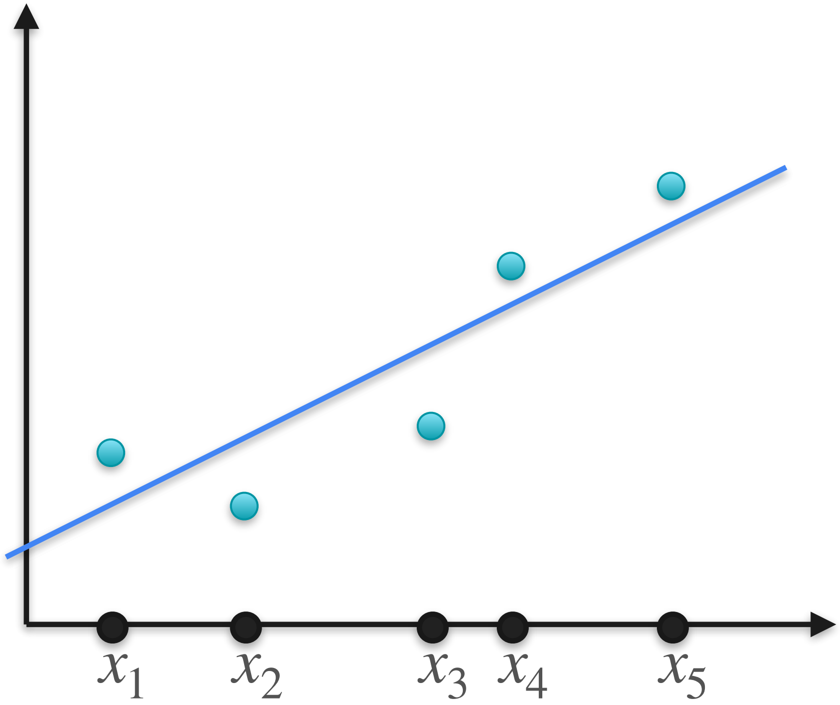
Line that best produced the points



Linear Regression

Line that best produced the points

Line that best fits the data (linear regression)

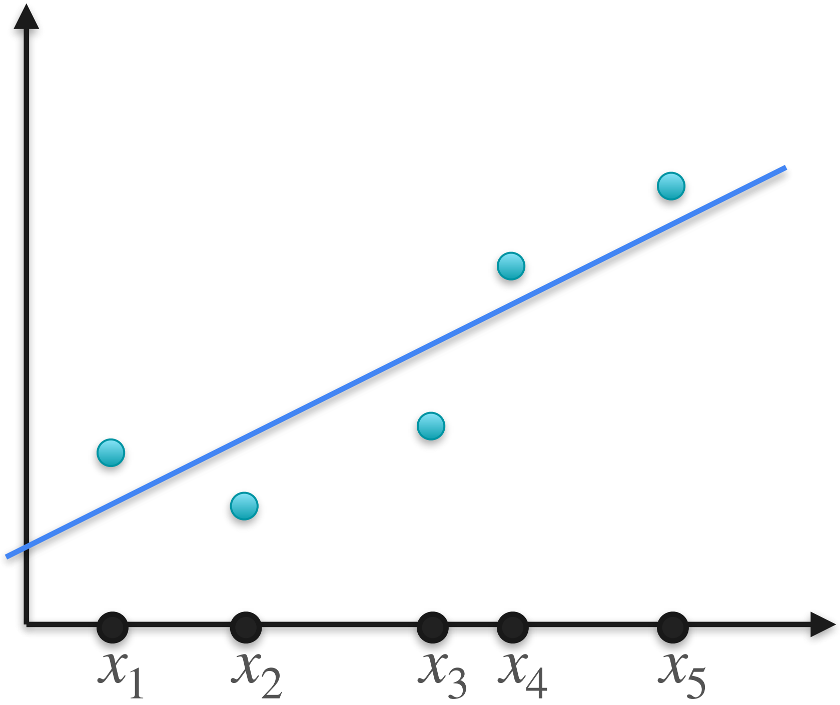


Linear Regression

Line that best produced the points

↕

Line that best fits the data (linear regression)

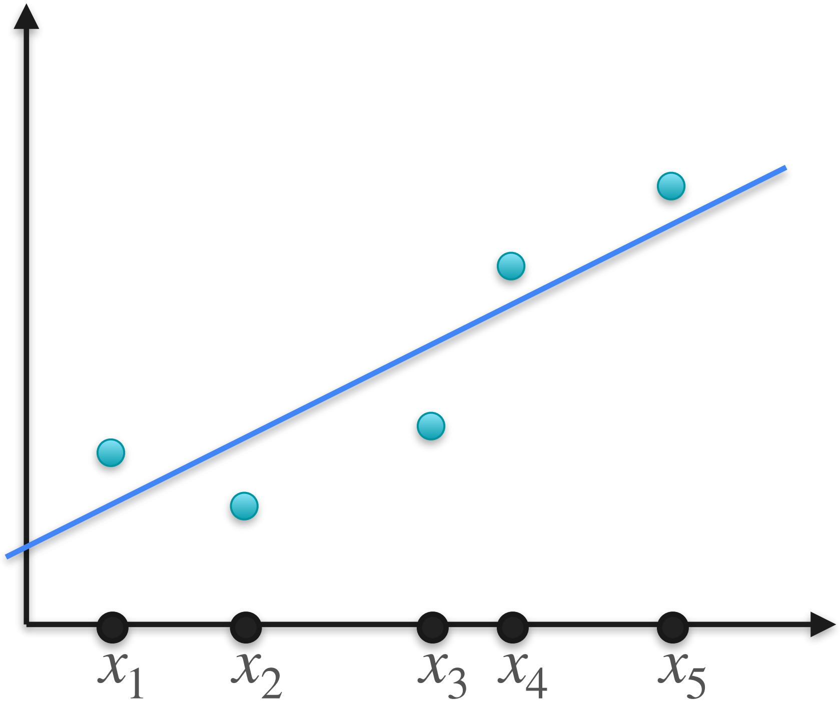


Linear Regression

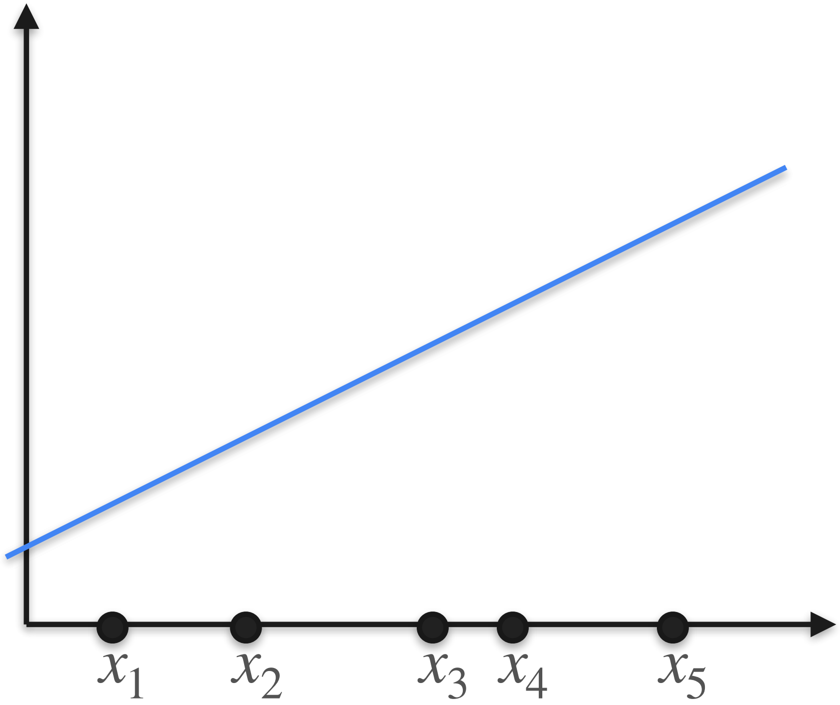
Line that best produced the points

How?

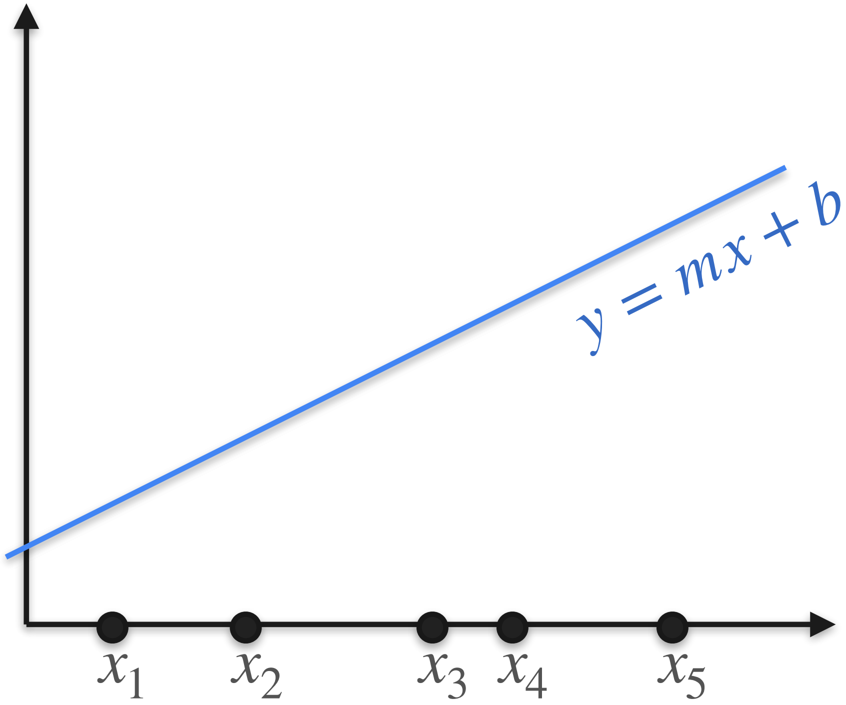
Line that best fits the data (linear regression)



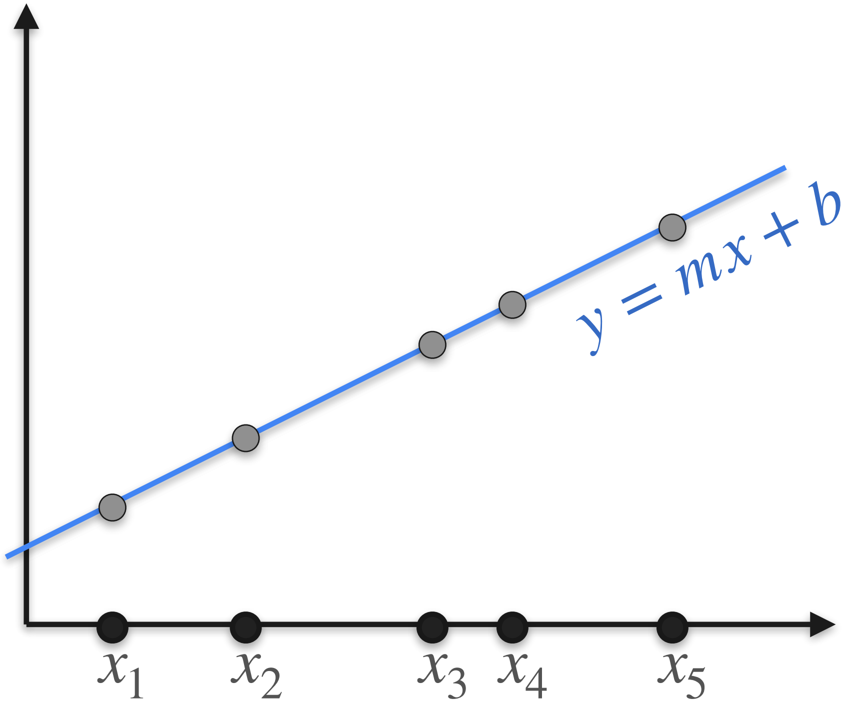
Linear Regression and Likelihood



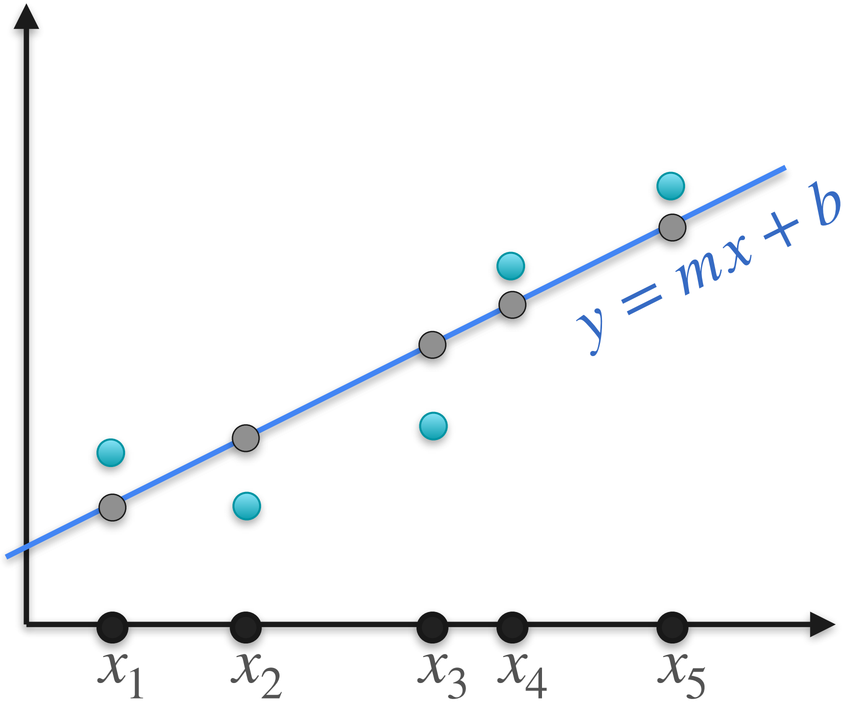
Linear Regression and Likelihood



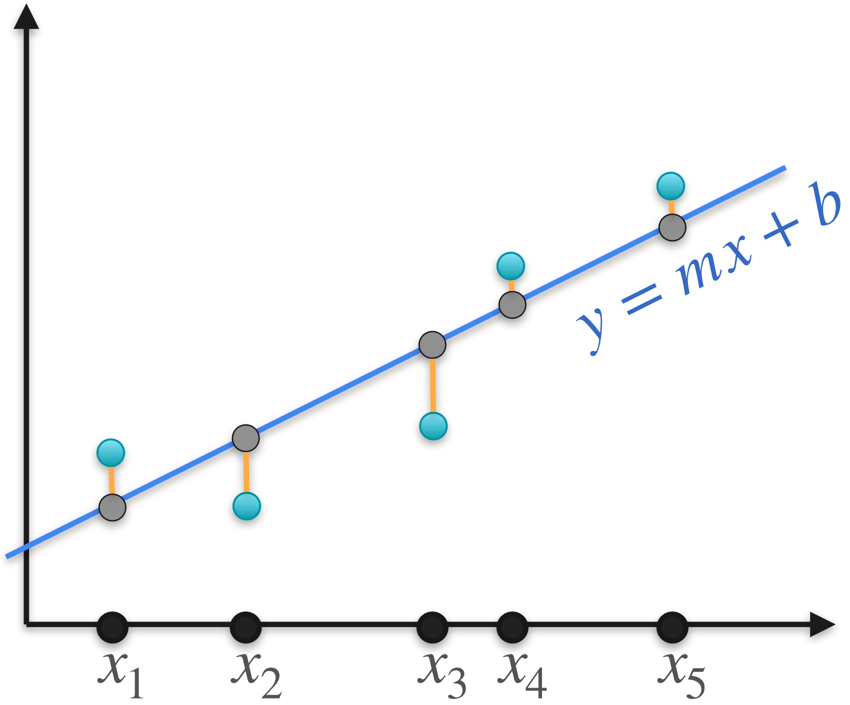
Linear Regression and Likelihood



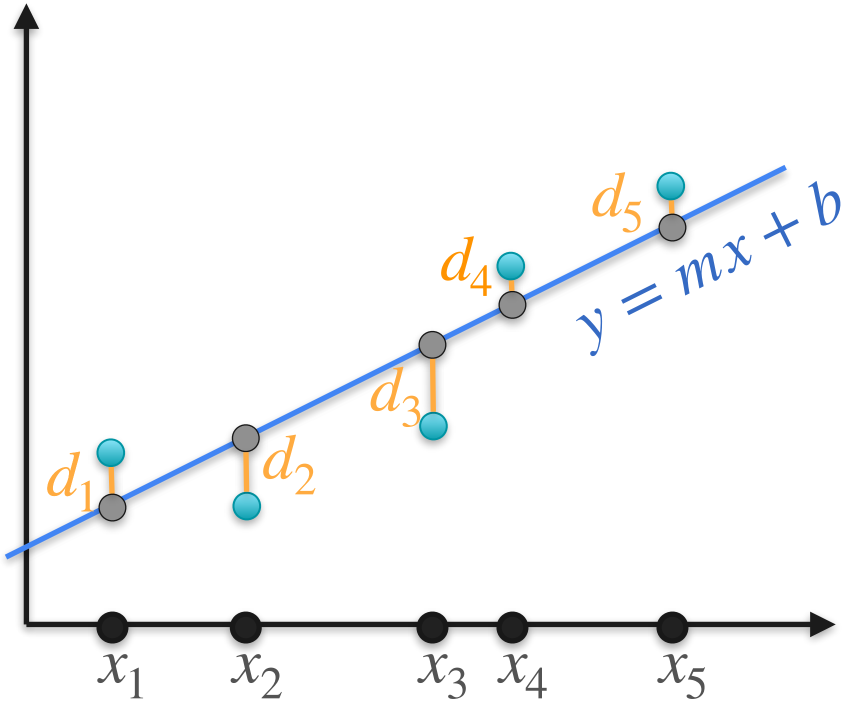
Linear Regression and Likelihood



Linear Regression and Likelihood

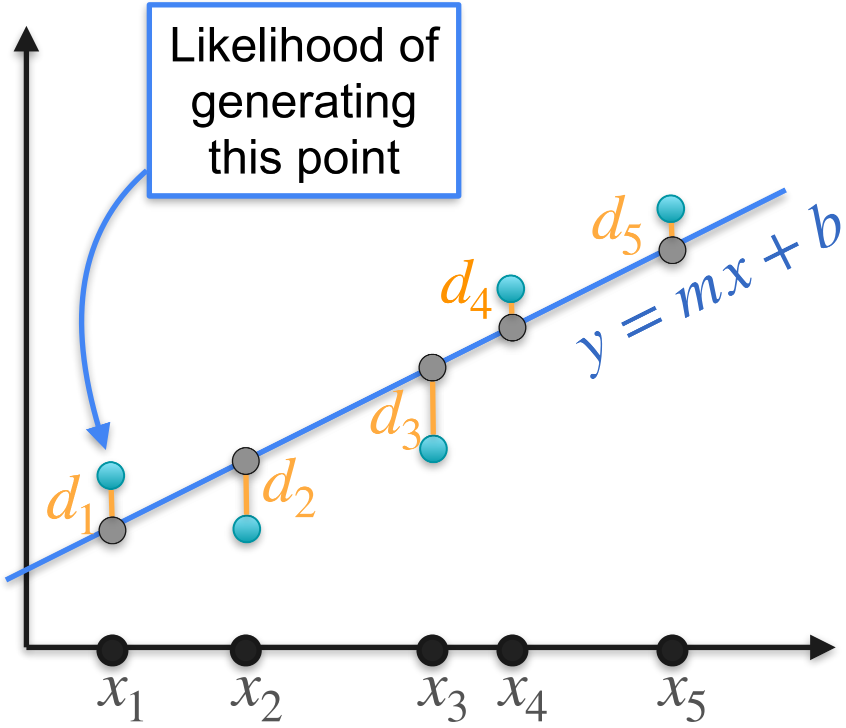


Linear Regression and Likelihood



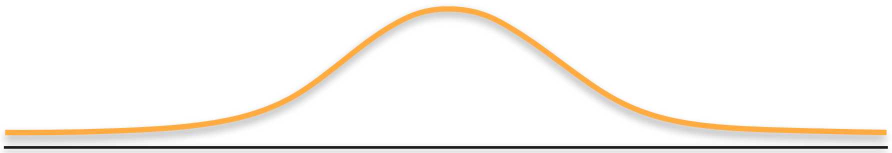
Linear Regression and Likelihood

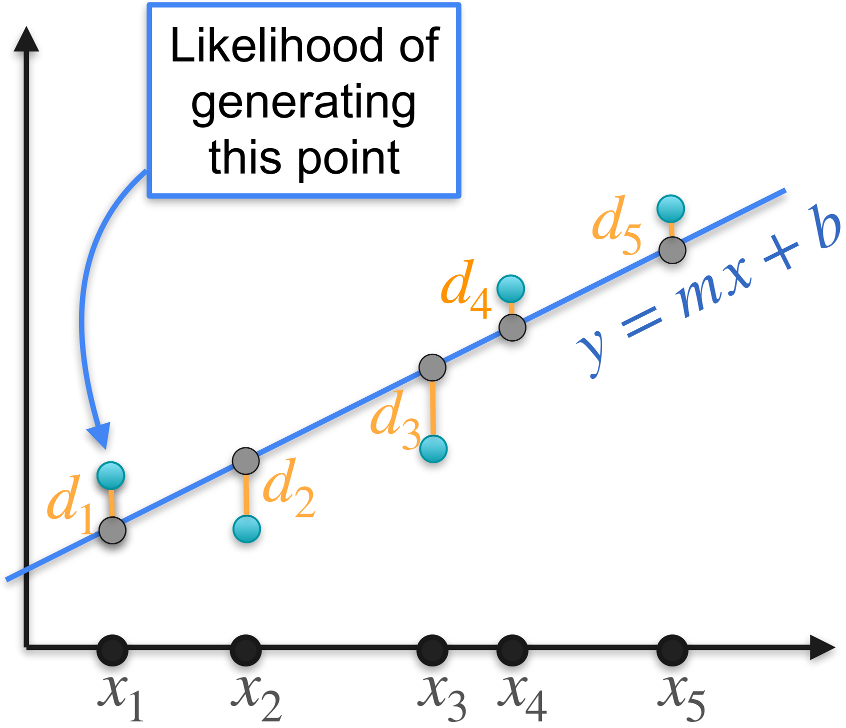
Likelihood:



Linear Regression and Likelihood

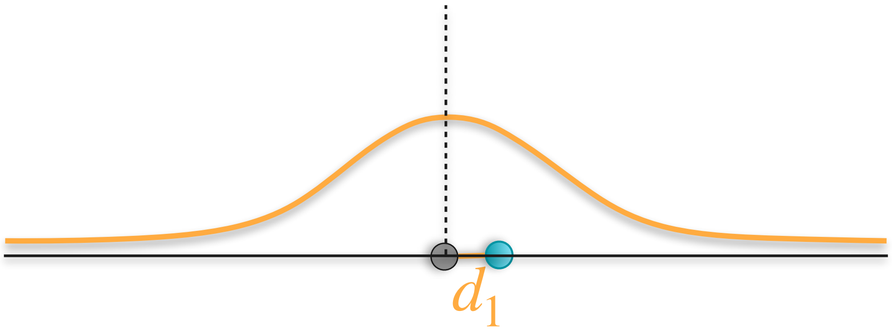
Likelihood:

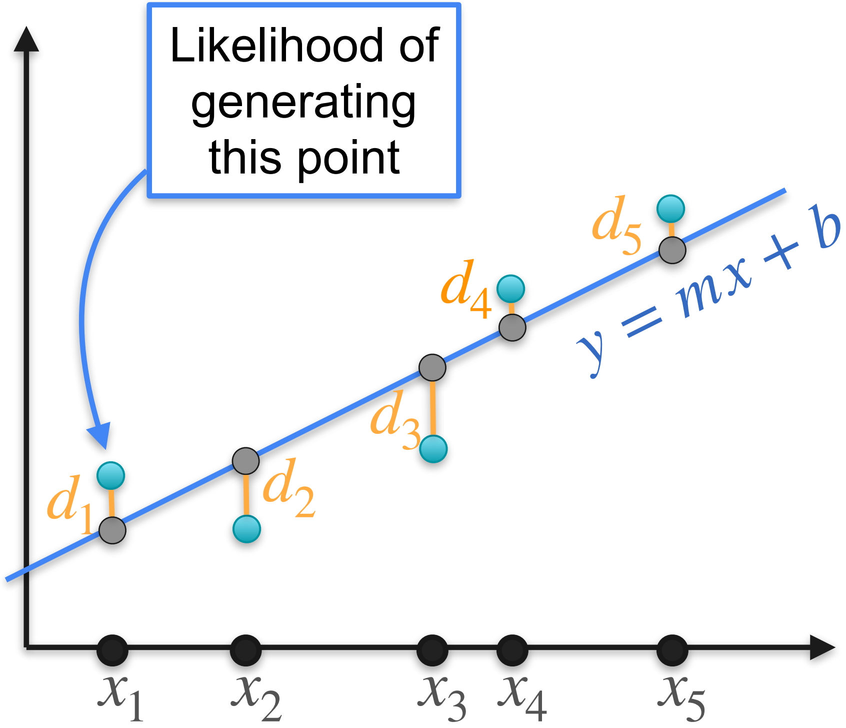




Linear Regression and Likelihood

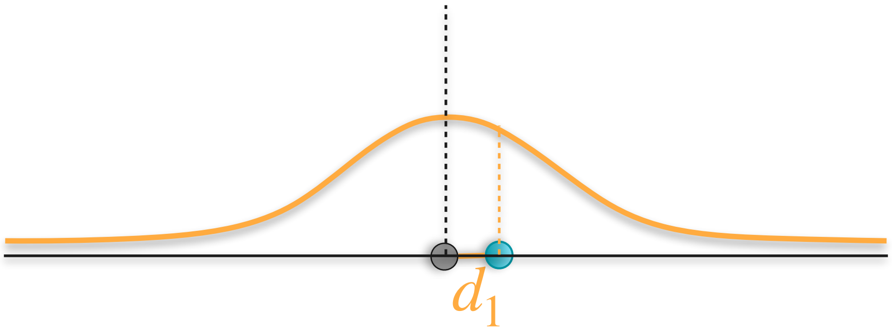
Likelihood:

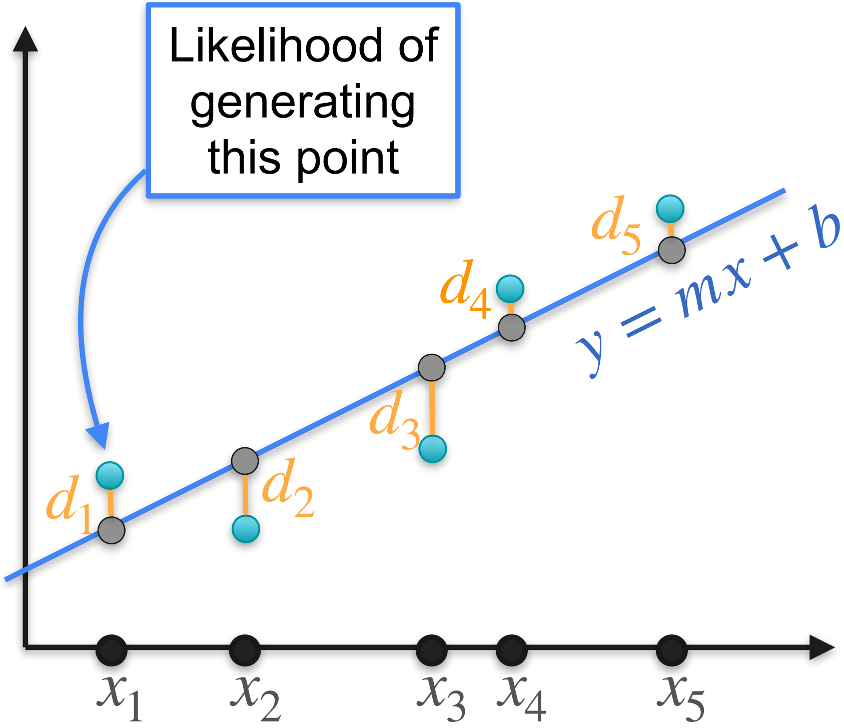




Linear Regression and Likelihood

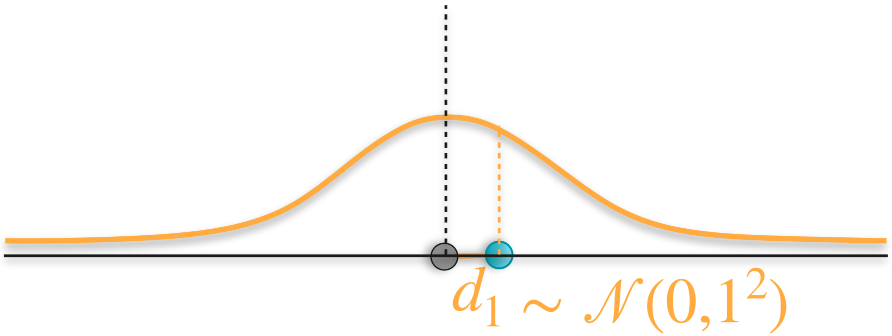
Likelihood:

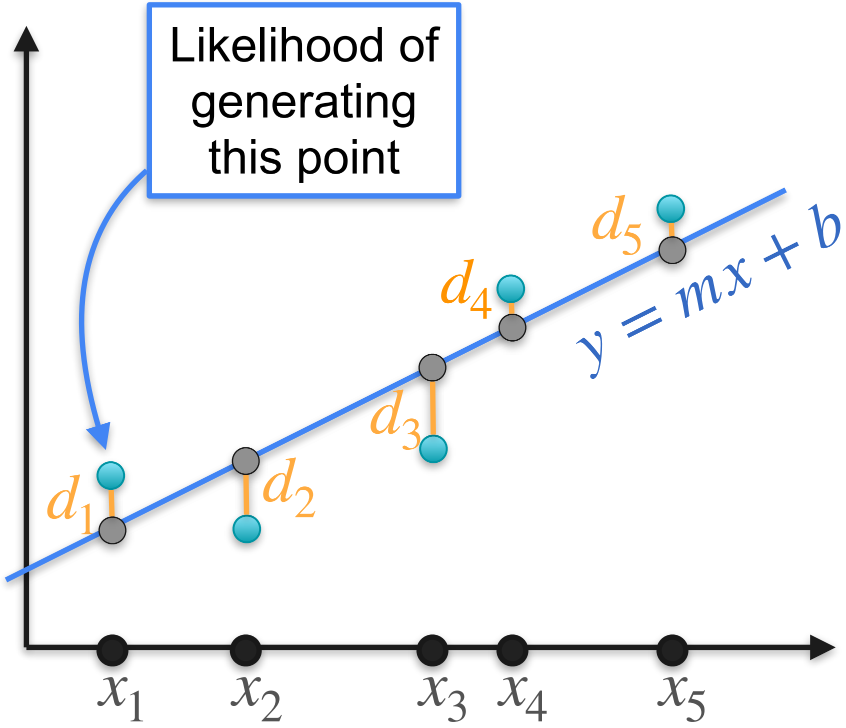




Linear Regression and Likelihood

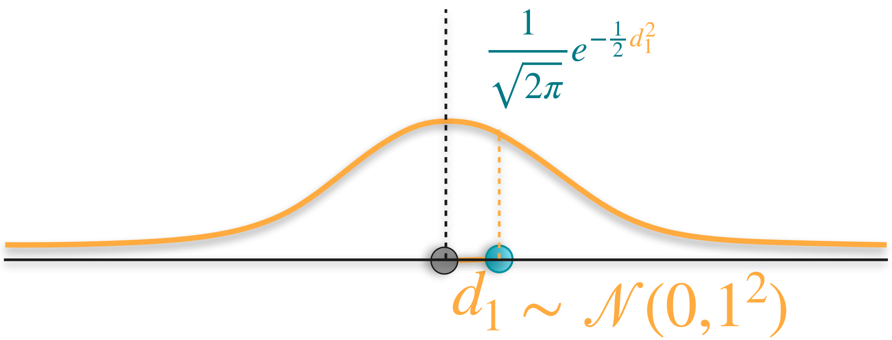
Likelihood:

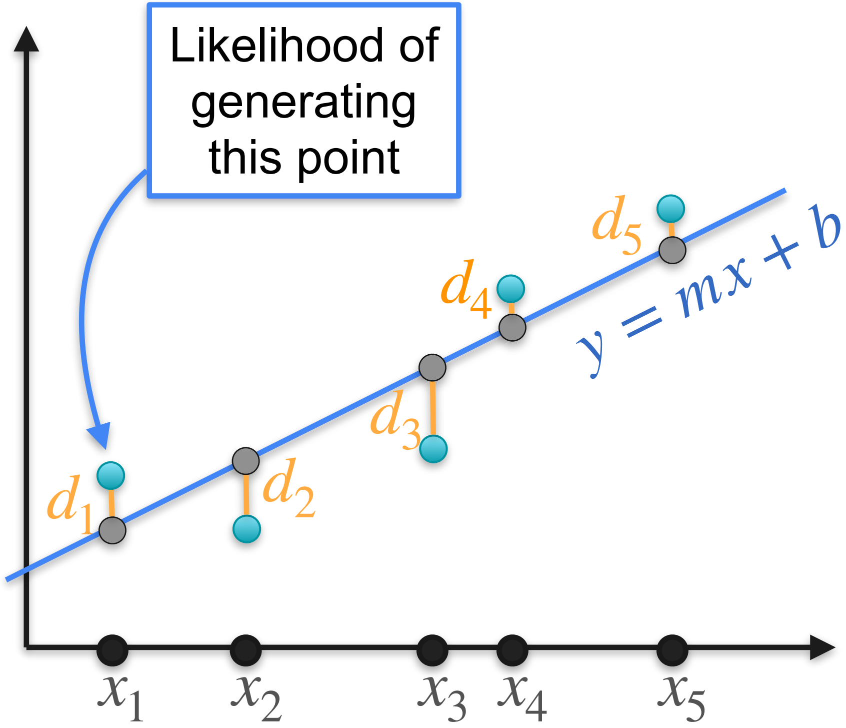




Linear Regression and Likelihood

Likelihood:

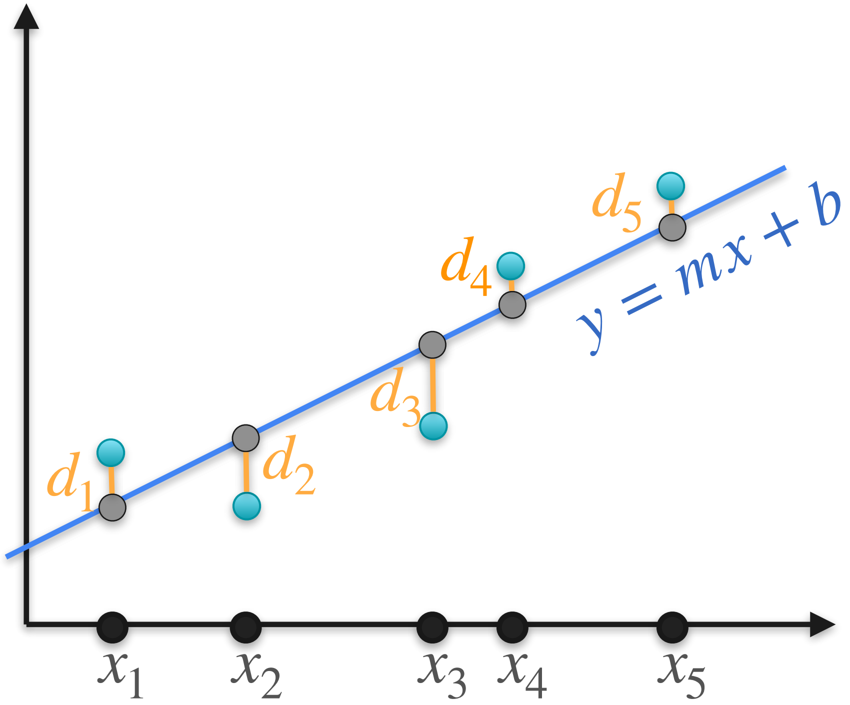




Linear Regression and Likelihood

Likelihood:

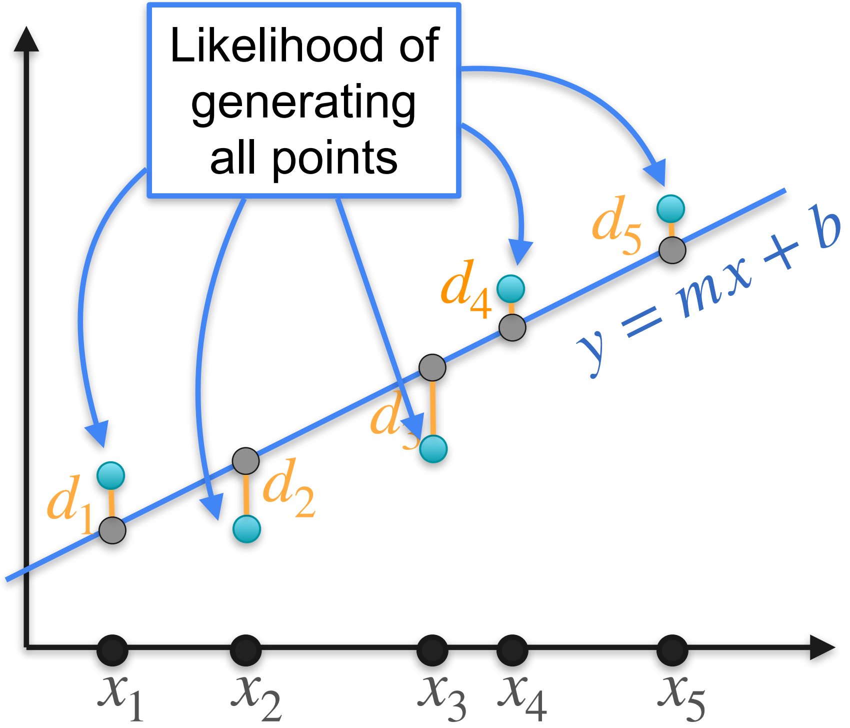
$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} d_1^2}$$



Linear Regression and Likelihood

Likelihood:

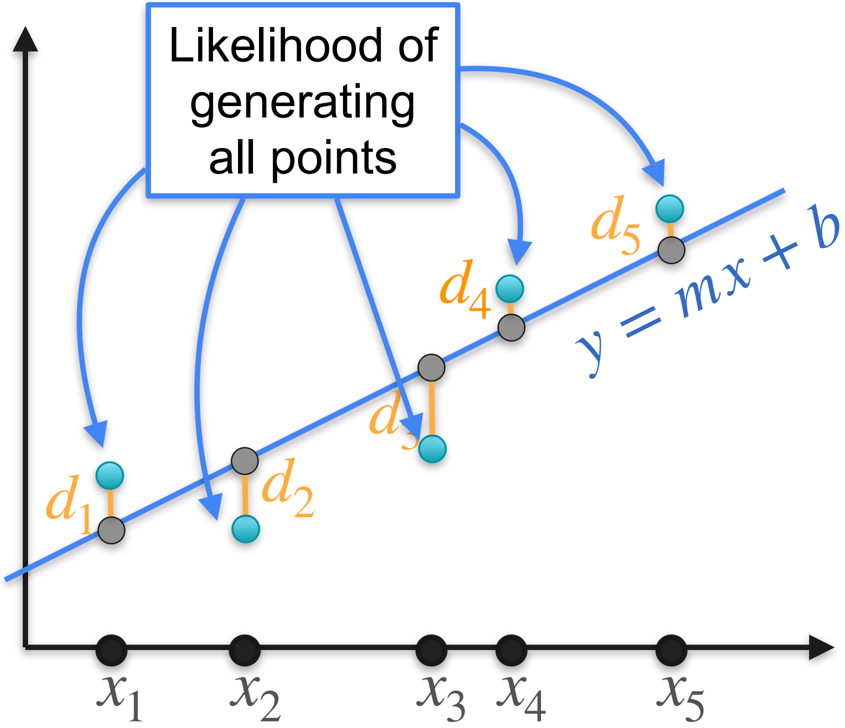
$$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_1^2}$$



Linear Regression and Likelihood

Likelihood:

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_1^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_2^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_3^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_4^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_5^2}$$

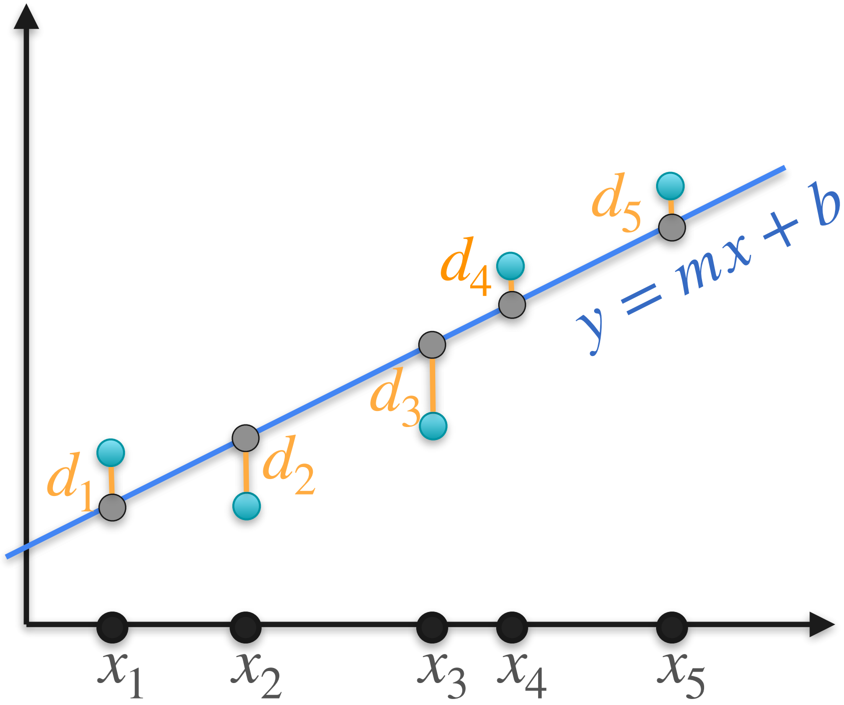


Linear Regression and Likelihood

Likelihood:

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_1^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_2^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_3^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_4^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_5^2}$$

Maximize

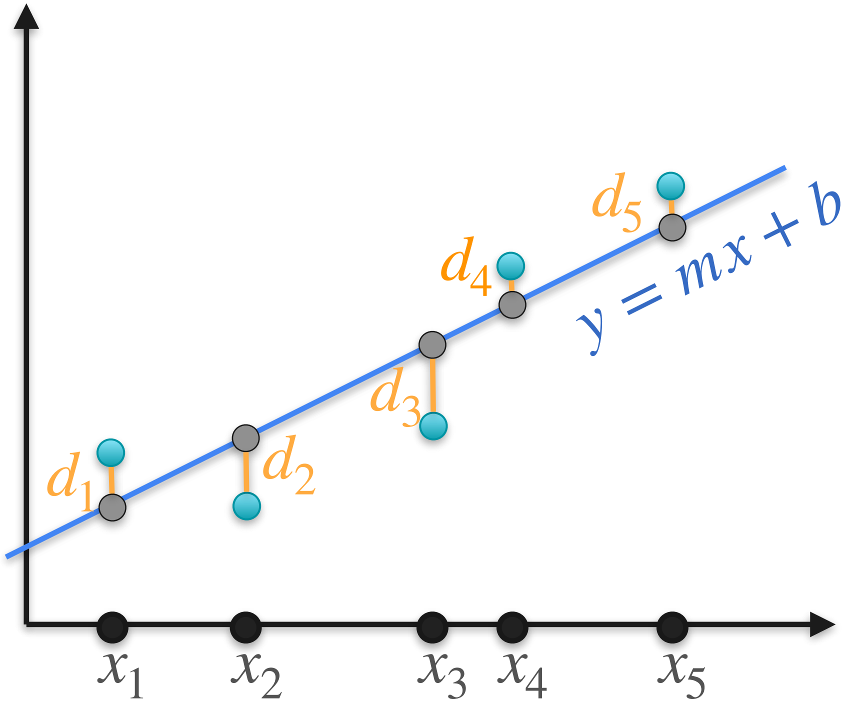


Linear Regression and Likelihood

Likelihood:

$$\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2}$$

Maximize



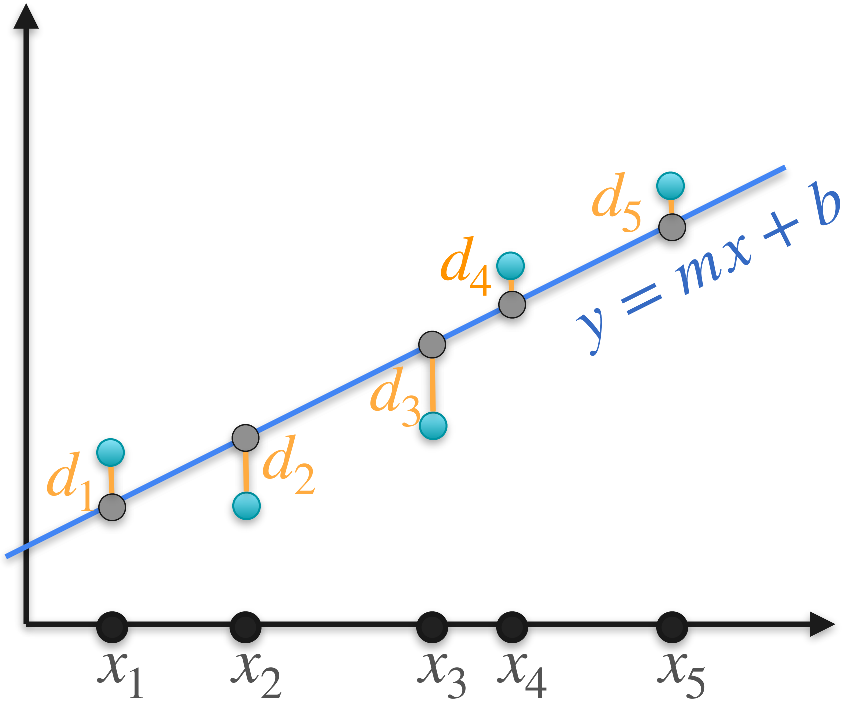
Linear Regression and Likelihood

Likelihood:

$$\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2}$$

Maximize

$$e^{-\frac{1}{2}d_1^2} \cdot e^{-\frac{1}{2}d_2^2} \cdot e^{-\frac{1}{2}d_3^2} \cdot e^{-\frac{1}{2}d_4^2} \cdot e^{-\frac{1}{2}d_5^2}$$



Linear Regression and Likelihood

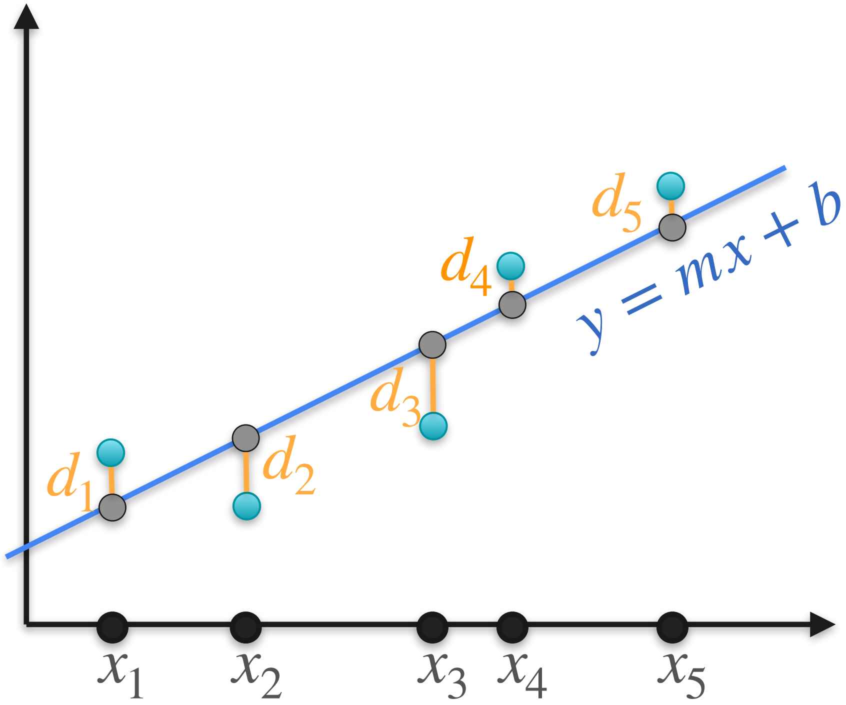
Likelihood:

$$\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2}$$

Maximize

$$e^{-\frac{1}{2}d_1^2}\cdot e^{-\frac{1}{2}d_2^2}\cdot e^{-\frac{1}{2}d_3^2}\cdot e^{-\frac{1}{2}d_4^2}\cdot e^{-\frac{1}{2}d_5^2}$$

$$e^{-\frac{1}{2}\left(d_1^2+d_2^2+d_3^2+d_4^2+d_5^2\right)}$$



Linear Regression and Likelihood

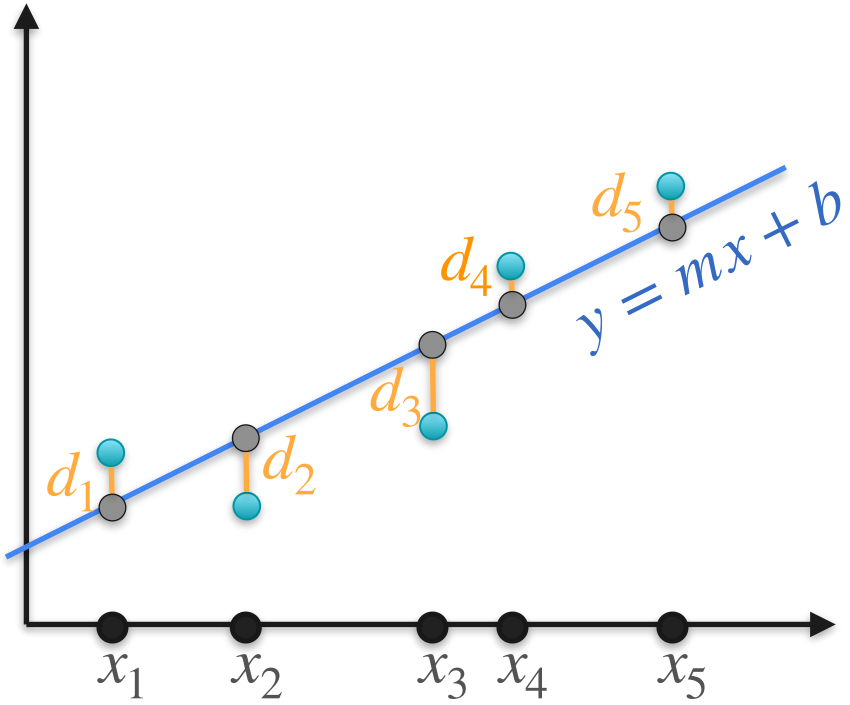
Likelihood:

$$\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2}$$

Maximize

$$e^{-\frac{1}{2}d_1^2} \cdot e^{-\frac{1}{2}d_2^2} \cdot e^{-\frac{1}{2}d_3^2} \cdot e^{-\frac{1}{2}d_4^2} \cdot e^{-\frac{1}{2}d_5^2}$$

$$\cancel{e^{\boxed{-}\frac{1}{2}}}(d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2)$$



Linear Regression and Likelihood

Likelihood:

$$\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2}$$

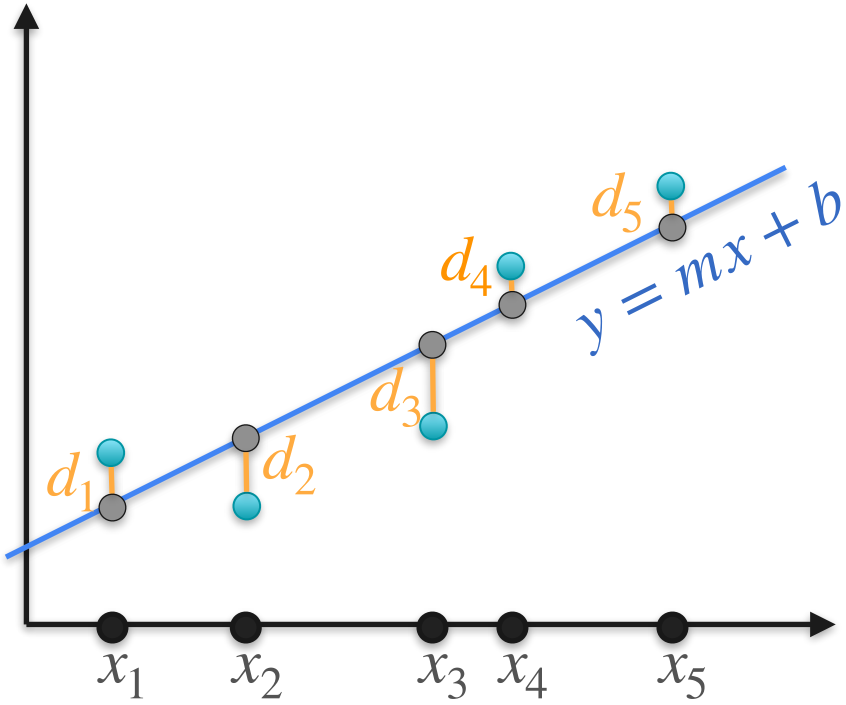
Maximize

$$e^{-\frac{1}{2}d_1^2}\cdot e^{-\frac{1}{2}d_2^2}\cdot e^{-\frac{1}{2}d_3^2}\cdot e^{-\frac{1}{2}d_4^2}\cdot e^{-\frac{1}{2}d_5^2}$$

$$\cancel{e^{\boxed{-}\frac{1}{2}}}\left(d_1^2+d_2^2+d_3^2+d_4^2+d_5^2\right)$$

Minimize

$$d_1^2+d_2^2+d_3^2+d_4^2+d_5^2$$



Linear Regression and Likelihood

Likelihood:

$$\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2}$$

Maximize

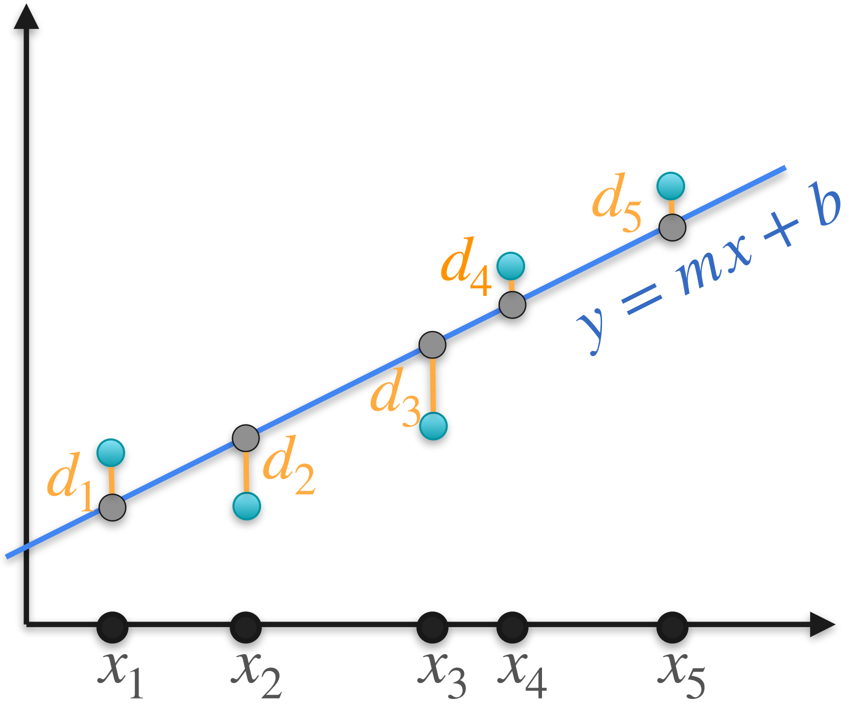
$$e^{-\frac{1}{2}d_1^2} \cdot e^{-\frac{1}{2}d_2^2} \cdot e^{-\frac{1}{2}d_3^2} \cdot e^{-\frac{1}{2}d_4^2} \cdot e^{-\frac{1}{2}d_5^2}$$

$$\cancel{e^{\boxed{-}\frac{1}{2}}}(d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2)$$

Minimize

$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$$

Least squares error!



Linear Regression and Likelihood

Likelihood:

$$\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2}$$

Maximize

$$e^{-\frac{1}{2}d_1^2} \cdot e^{-\frac{1}{2}d_2^2} \cdot e^{-\frac{1}{2}d_3^2} \cdot e^{-\frac{1}{2}d_4^2} \cdot e^{-\frac{1}{2}d_5^2}$$

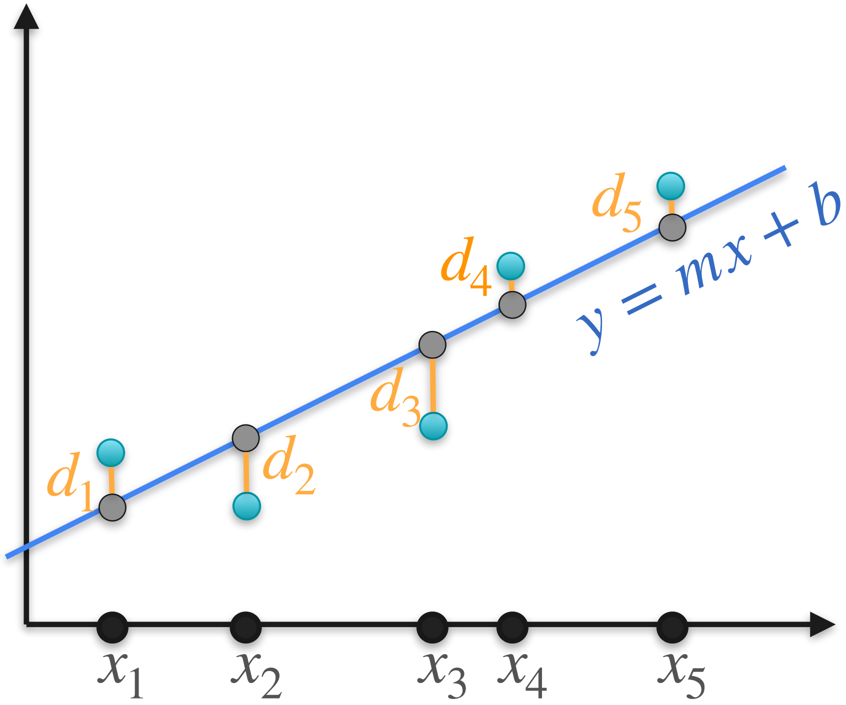
$$\cancel{e^{\boxed{-}\frac{1}{2}}}(d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2)$$

Minimize

$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$$

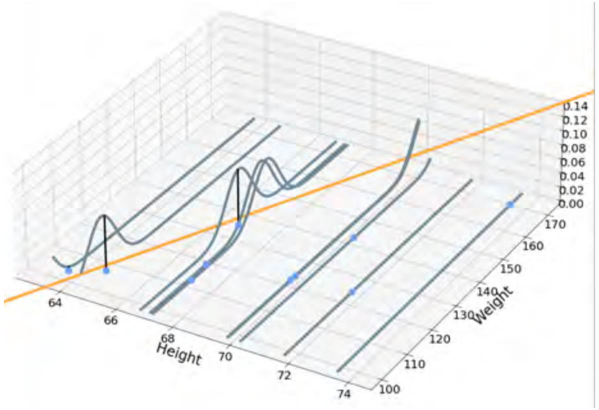
Least squares error!

Linear regression!

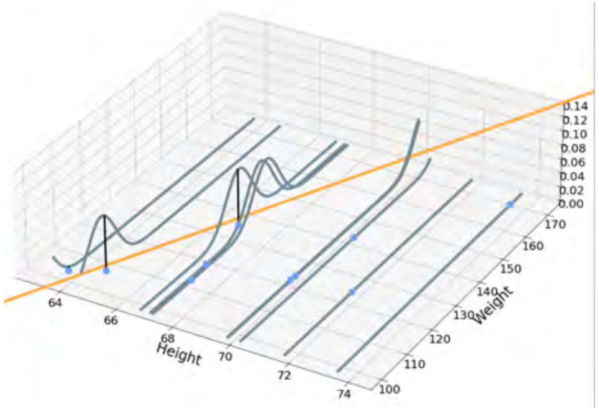


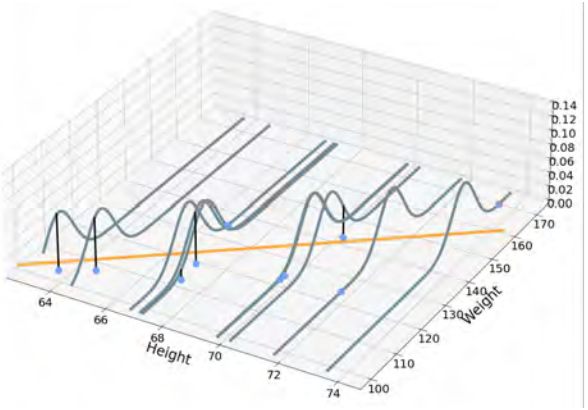
Picking the Right Model

Picking the Right Model

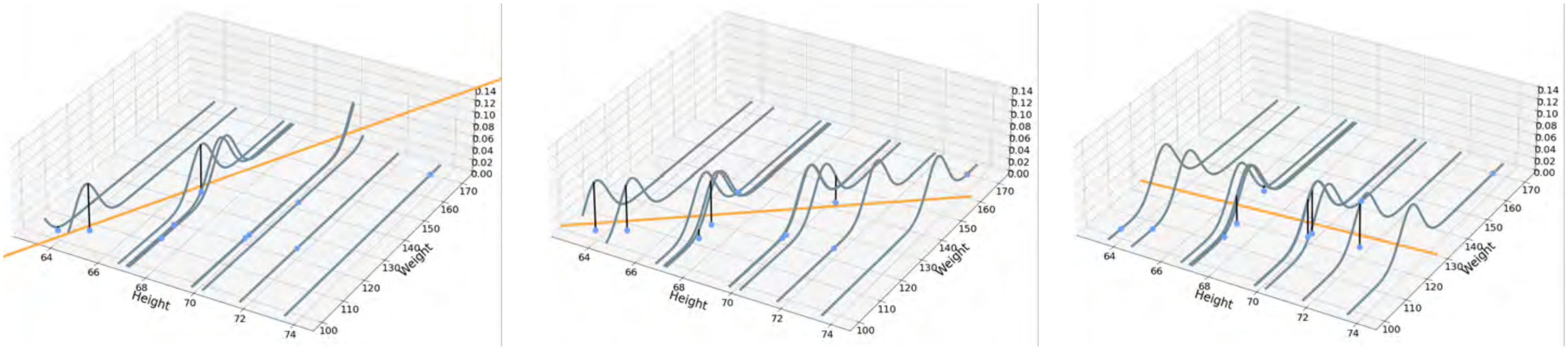


Picking the Right Model

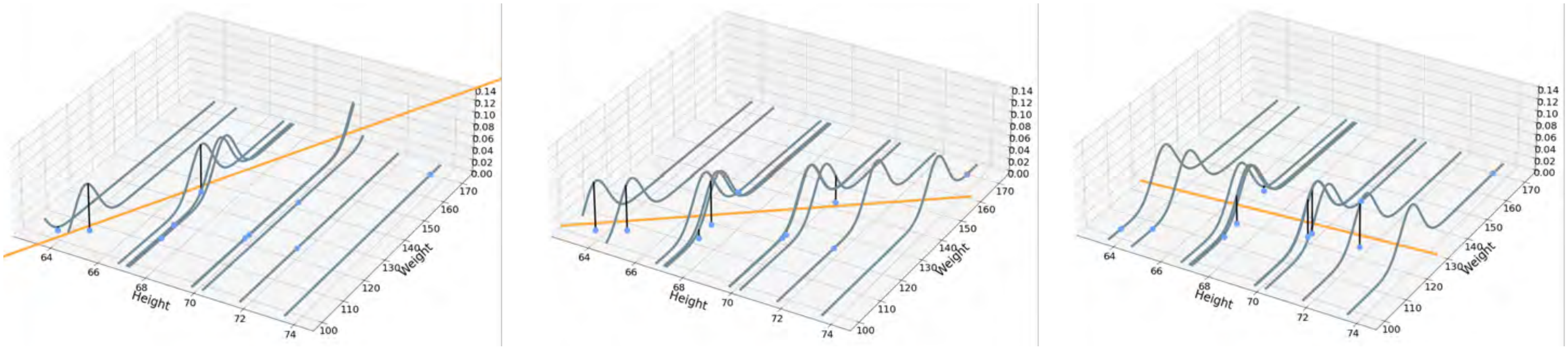




Picking the Right Model

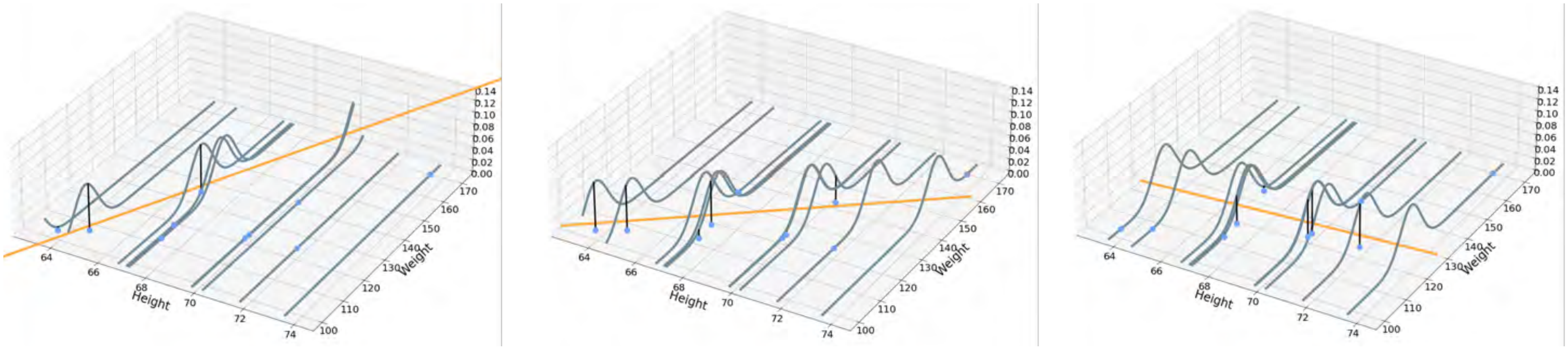


Picking the Right Model



Model 1:
Likelihood $= 4.91 \cdot 10^{-260}$

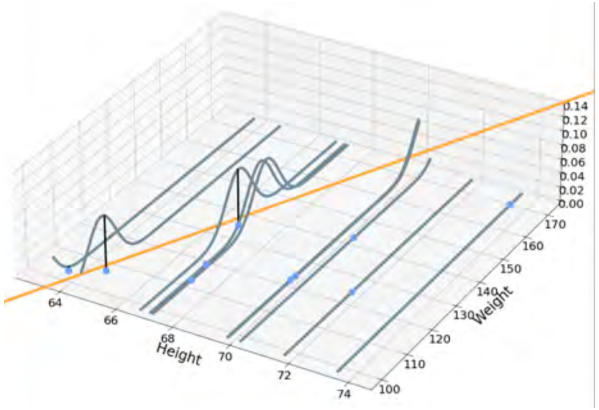
Picking the Right Model



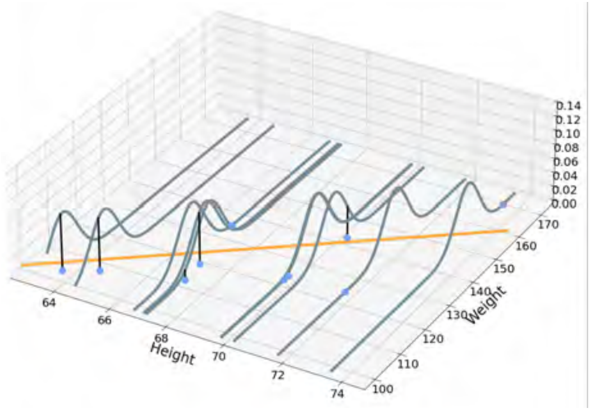
Model 1:
Likelihood $= 4.91 \cdot 10^{-260}$

Model 2:
Likelihood $= 8.16 \cdot 10^{-28}$

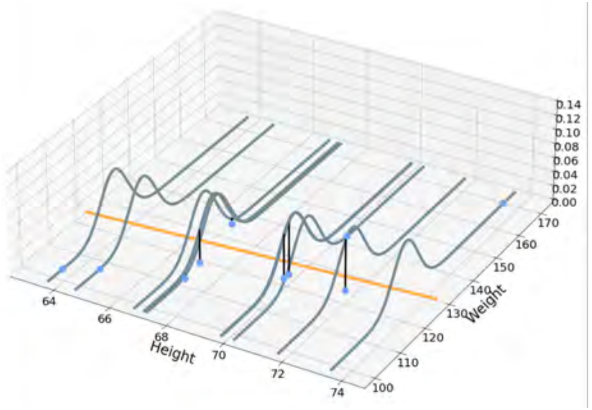
Picking the Right Model



Model 1:
Likelihood $= 4.91 \cdot 10^{-260}$

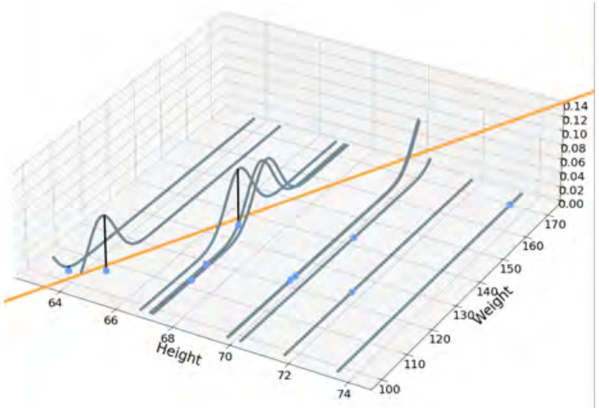


Model 2:
Likelihood $= 8.16 \cdot 10^{-28}$



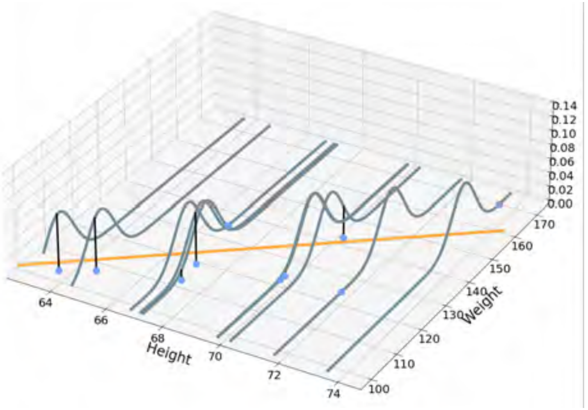
Model 3:
Likelihood $= 3.48 \cdot 10^{-49}$

Picking the Right Model



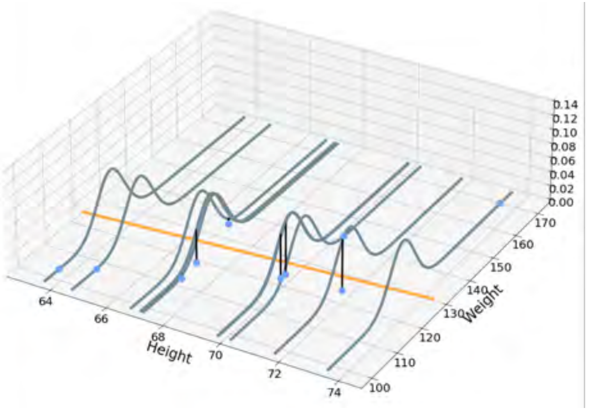
Model 1:

Likelihood $= 4.91 \cdot 10^{-260}$



Model 2:

Likelihood $= 8.16 \cdot 10^{-28}$



Model 3:

Likelihood $= 3.48 \cdot 10^{-49}$

DeepLearning.AI

Point Estimation

Frequentist vs Bayesian Statistics

Frequentist Vs. Bayesian Statistics

Frequentist Vs. Bayesian Statistics

Frequentists	Bayesians

Frequentist Vs. Bayesian Statistics

Frequentists

- Probabilities represent long term frequency of events

Bayesians

Frequentist Vs. Bayesian Statistics

Frequentists

- Probabilities represent long term frequency of events

Bayesians

- Probabilities represent the degree of belief (or certainty)

Frequentist Vs. Bayesian Statistics

Frequentists

- Probabilities represent long term frequency of events

- Concept of Likelihood

Bayesians

- Probabilities represent the degree of belief (or certainty)

Frequentist Vs. Bayesian Statistics

Frequentists

- Probabilities represent long term frequency of events

- Concept of Likelihood

Bayesians

- Probabilities represent the degree of belief (or certainty)

- Concept of Prior

Frequentist Vs. Bayesian Statistics

Frequentists

- Probabilities represent long term frequency of events
- Concept of Likelihood

Bayesians

- Probabilities represent the degree of belief (or certainty)
- Concept of Prior
- Goal: update prior belief based on observations

Frequentist Vs. Bayesian Statistics

Frequentists

- Probabilities represent long term frequency of events
- Concept of Likelihood
- Goal: Find the model that most likely generated the observed data

Bayesians

- Probabilities represent the degree of belief (or certainty)
- Concept of Prior
- Goal: update prior belief based on observations

Frequentist Approach

Frequentist Approach

Frequentist Approach

$p =$ probability of heads

Frequentist Approach

p = probability of heads

$$p = 0.8$$

Frequentist Approach

p = probability of heads

$$p = 0.8$$

Maximum likelihood

Bayesian Approach

$p =$ probability of heads

Bayesian Approach

$p =$ probability of heads

p could be anything between 0 and 1

Bayesian Approach

$p =$ probability of heads

p could be anything between 0 and 1

p could still be anything between 0 and 1, but it may be closer to 1

Bayesian Approach

$p =$ probability of heads

p could be anything between 0 and 1

p could still be anything between 0 and 1, but it may be closer to 1

p could still be anything between 0 and 1, but it's probably around 0.5

Bayesian Approach

$p =$ probability of heads

p could be anything between 0 and 1

p could still be anything between 0 and 1, but it may be closer to 1

p could still be anything between 0 and 1, but it's probably around 0.5

p could still be anything between 0 and 1, but it's more likely closer to 0.66

Bayesian Approach

$p =$ probability of heads

p could be anything between 0 and 1

p could still be anything between 0 and 1, but it may be closer to 1

p could still be anything between 0 and 1, but it's probably around 0.5

p could still be anything between 0 and 1, but it's more likely closer to 0.66

…

Bayesian Approach

$p =$ probability of heads

p could be anything between 0 and 1

p could still be anything between 0 and 1, but it may be closer to 1

p could still be anything between 0 and 1, but it's probably around 0.5

p could still be anything between 0 and 1, but it's more likely closer to 0.66

…

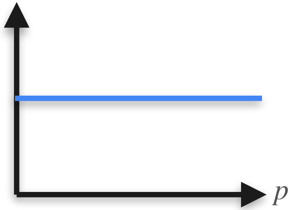
p could still be anything between 0 and 1, but it's very probably close to 0.8

Bayesian Approach

$p =$ probability of heads

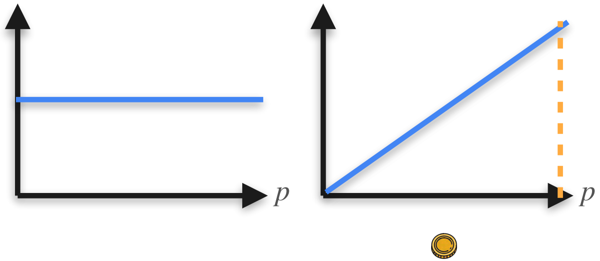
Bayesian Approach

$p =$ probability of heads



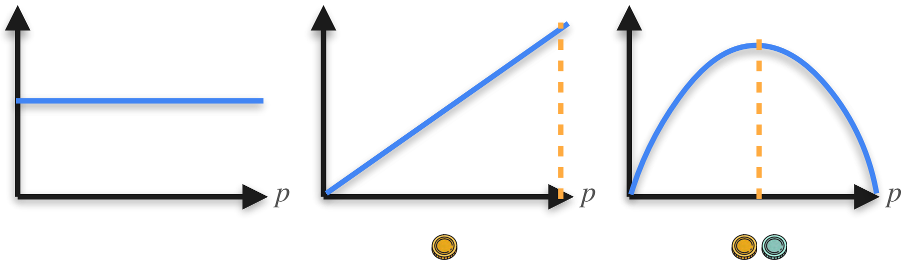
Bayesian Approach

$p =$ probability of heads



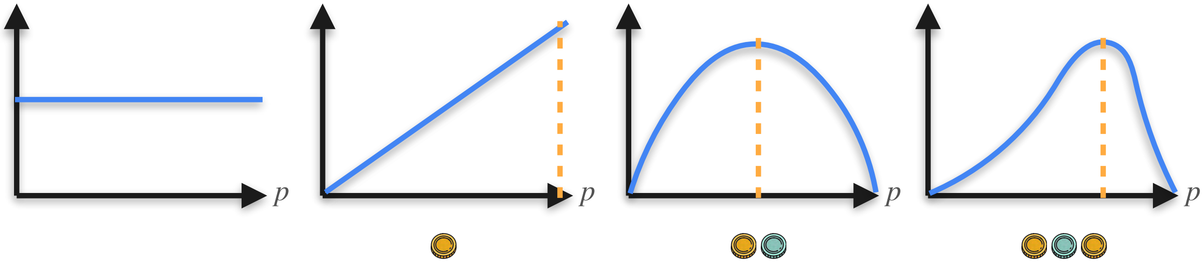
Bayesian Approach

$p =$ probability of heads



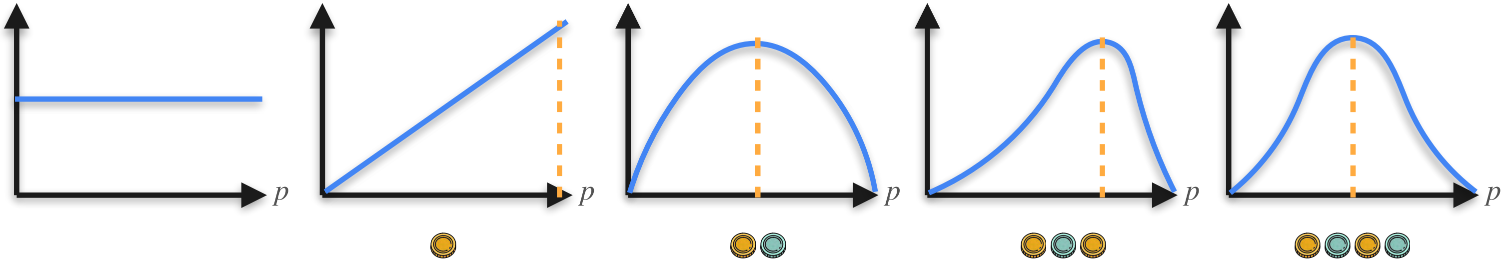
Bayesian Approach

$p =$ probability of heads



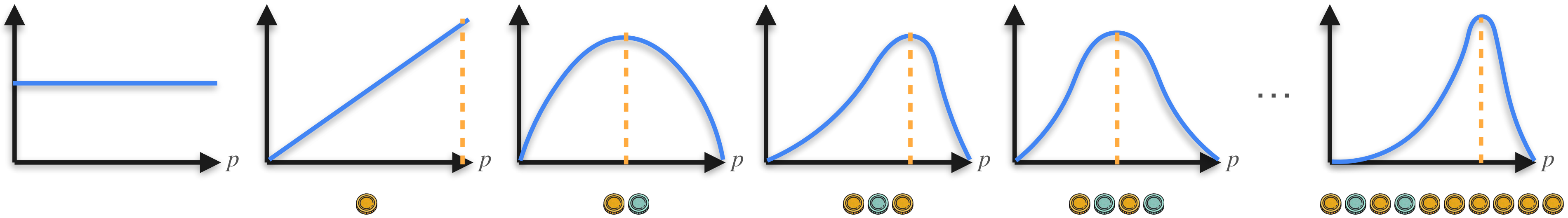
Bayesian Approach

$p =$ probability of heads



Bayesian Approach

$p =$ probability of heads



Bayesian Statistics

Bayesian Statistics

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

Bayesian Statistics

Remember Bayes' theorem? $\boxed{\mathbf{P}(B \mid A)} = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

Posterior

Bayesian Statistics

Remember Bayes' theorem? $\boxed{\mathbf{P}(B \mid A)} = \dfrac{\mathbf{P}(A \mid B)\boxed{\mathbf{P}(B)}}{\mathbf{P}(A)}$ Prior

Posterior

Bayesian Statistics

Remember Bayes' theorem? $\boxed{\mathbf{P}(B \mid A)} = \dfrac{\mathbf{P}(A \mid B)\boxed{\mathbf{P}(B)}}{\boxed{\mathbf{P}(A)}}$ Prior

Posterior

Normalizing constant

Bayesian Statistics

The parameters (or model) you want to estimate

Remember Bayes' theorem? $\boxed{\mathbf{P}(B \mid A)} = \dfrac{\mathbf{P}(A \mid B)\boxed{\mathbf{P}(B)}}{\boxed{\mathbf{P}(A)}}$

Prior

Posterior

Normalizing constant

Bayesian Statistics

The parameters (or model) you want to estimate

Your sample vector

Remember Bayes' theorem? $\boxed{\mathbf{P}(B \mid A)} = \dfrac{\mathbf{P}(A \mid B)\boxed{\mathbf{P}(B)}}{\boxed{\mathbf{P}(A)}}$ Prior

Posterior

Normalizing constant

Bayesian Statistics

Remember Bayes' theorem?

$$\boxed{\mathbf{P}(B \mid A)} = \frac{\boxed{\mathbf{P}(A \mid B)}\,\boxed{\mathbf{P}(B)}}{\boxed{\mathbf{P}(A)}}$$

B: The parameters (or model) you want to estimate

A: Your sample vector

$\mathbf{P}(B \mid A)$: Posterior

$\mathbf{P}(A \mid B)$: Probability of samples given model B

$\mathbf{P}(B)$: Prior

$\mathbf{P}(A)$: Normalizing constant

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

$$\Theta = \mathbf{P}(H)$$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

Θ is a random variable $\longrightarrow$ $\Theta = \mathbf{P}(H)$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

Θ is a random variable $\longrightarrow$ $\Theta = \mathbf{P}(H)$

$X_i = 1$ if H, 0 if T

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

Θ is a random variable $\longrightarrow$ $\Theta = \mathbf{P}(H)$

$X_i = 1$ if H, 0 if T

Observations: $\sum_{i=1}^{10} X_i = 8$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

Θ is a random variable $\Theta = \mathbf{P}(H)$ $X_i = 1$ if $H, 0$ if T

Observations: $\sum_{i=1}^{10} X_i = 8$

Suppose that Θ takes on the particular value θ

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

Θ is a random variable $\longrightarrow$ $\Theta = \mathbf{P}(H)$ $\quad\quad X_i = 1$ if H, 0 if T

Observations: $\sum_{i=1}^{10} X_i = 8$

Suppose that Θ takes on the particular value θ

$\mathbf{P}(A \mid B)$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

Θ is a random variable $\longrightarrow \Theta = \mathbf{P}(H)$ $\qquad X_i = 1$ if H, 0 if T

Observations: $\sum_{i=1}^{10} X_i = 8 \longleftarrow A$

Suppose that Θ takes on the particular value θ

$\mathbf{P}(A \mid B)$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

Θ is a random variable $\longrightarrow \Theta = \mathbf{P}(H)$ $\qquad X_i = 1$ if H, 0 if T

Observations: $\displaystyle\sum_{i=1}^{10} X_i = 8 \longleftarrow A$

Suppose that $\boxed{\Theta \text{ takes on the particular value } \theta} \longleftarrow B$

$\mathbf{P}(A \mid B)$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

Θ is a random variable $\longrightarrow$ $\Theta = \mathbf{P}(H)$ $\qquad X_i = 1$ if H, 0 if T

Observations: $\sum_{i=1}^{10} X_i = 8 \longleftarrow A$

Suppose that $\boxed{\Theta \text{ takes on the particular value } \theta} \longleftarrow B$

$$\mathbf{P}(A \mid B) \longrightarrow \mathbf{P}\left(\sum_{i=1}^{10} X_i = 8 \mid \Theta = \theta \right) =$$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

$\theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad 1-\theta \quad 1-\theta$

Θ is a random variable $\longrightarrow \Theta = \mathbf{P}(H)$ $\qquad X_i = 1$ if H, 0 if T

Observations: $\displaystyle\sum_{i=1}^{10} X_i = 8 \longleftarrow A$

Suppose that $\boxed{\Theta \text{ takes on the particular value } \theta} \longleftarrow B$

$$\mathbf{P}(A \mid B) \longrightarrow \mathbf{P}\left(\sum_{i=1}^{10} X_i = 8 \mid \Theta = \theta\right) = \theta^8(1-\theta)^2$$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

$\theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad 1-\theta \quad 1-\theta$

Θ is a random variable $\longrightarrow$ $\Theta = \mathbf{P}(H)$ $\qquad X_i = 1$ if H, 0 if T

Observations: $\displaystyle\sum_{i=1}^{10} X_i = 8 \longleftarrow A$

Suppose that $\boxed{\Theta \text{ takes on the particular value } \theta} \longleftarrow B$

$$\mathbf{P}(A \mid B) \longrightarrow \mathbf{P}\left(\sum_{i=1}^{10} X_i = 8 \mid \Theta = \theta\right) = \theta^8(1-\theta)^2 \qquad X_i \mid \Theta = \theta \sim Bernoulli(\theta)$$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

$\theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad 1-\theta \quad 1-\theta$

$X_i \mid \Theta = \theta \sim Bernoulli(\theta)$

How do you choose the prior?

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

$\theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad 1-\theta \quad 1-\theta$

$$X_i \mid \Theta = \theta \sim Bernoulli(\theta)$$

How do you choose the prior?

If before tossing the coins you have no idea whether the coin is fair or not, then your prior belief about Θ should not favor any value.

You can model this with a uniform distribution

$$\Theta \sim Uniform(0,1) \qquad f_\Theta(\theta) = 1, \quad 0 \le \theta \le 1$$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

$\theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad 1-\theta \quad 1-\theta$

$X_i \mid \Theta = \theta \sim Bernoulli(\theta)$

How do you choose the prior?

If before tossing the coins you have no idea whether the coin is fair or not, then your prior belief about Θ should not favor any value.

You can model this with a uniform distribution

$\Theta \sim Uniform(0,1)$ $\quad \boxed{f_\Theta(\theta) = 1, \quad 0 \le \theta \le 1} \longleftarrow$ Prior $(\mathbf{P}(B))$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

$\theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad 1-\theta \quad 1-\theta$

$X_i \mid \Theta = \theta \sim Bernoulli(\theta) \longleftarrow (\mathbf{P}(A \mid B))$

$\Theta \sim Uniform(0,1) \longleftarrow$ Prior $(\mathbf{P}(B))$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

$\theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad 1-\theta \quad 1-\theta$

$X_i \mid \Theta = \theta \sim Bernoulli(\theta) \leftarrow (\mathbf{P}(A \mid B))$ $\qquad \Theta \sim Uniform(0,1) \leftarrow$ Prior $(\mathbf{P}(B))$

How do you get the posterior? Bayes!

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

$\theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad 1-\theta \quad 1-\theta$

$X_i \mid \Theta = \theta \sim Bernoulli(\theta) \longleftarrow (\mathbf{P}(A \mid B))$

$\Theta \sim Uniform(0,1) \longleftarrow$ Prior $(\mathbf{P}(B))$

How do you get the posterior? Bayes!

$$f_{\Theta \mid \mathbf{X}=\mathbf{x}}(\theta) = \frac{p_{\mathbf{X} \mid \Theta=\theta}(\mathbf{x}) f_\Theta(\theta)}{p_{\mathbf{X}}(\mathbf{x})}$$

Bayesian Statistics: Bernoulli Example

Remember Bayes' theorem? $\mathbf{P}(B \mid A) = \dfrac{\mathbf{P}(A \mid B)\mathbf{P}(B)}{\mathbf{P}(A)}$

$\theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad \theta \quad 1-\theta \quad 1-\theta$

$X_i \mid \Theta = \theta \sim Bernoulli(\theta) \longleftarrow (\mathbf{P}(A \mid B))$ $\qquad \Theta \sim Uniform(0,1) \longleftarrow$ Prior $(\mathbf{P}(B))$

How do you get the posterior? Bayes!

$$\boxed{f_{\Theta \mid \mathbf{X}=\mathbf{x}}(\theta)} = \frac{p_{\mathbf{X} \mid \Theta=\theta}(\mathbf{x}) f_\Theta(\theta)}{\boxed{p_{\mathbf{X}}(\mathbf{x})}}$$

$\longleftarrow$ Normalizing constant

$\uparrow$

Posterior $(\mathbf{P}(B \mid A))$

Bayesian Statistics: MAP Estimator

$$f_{\Theta|\mathbf{X}=\mathbf{x}}(\theta) \propto p_{\mathbf{X}|\Theta=\theta}(\mathbf{x}) f_{\Theta}(\theta)$$

Bayesian Statistics: MAP Estimator

$$f_{\Theta|\mathbf{X}=\mathbf{x}}(\theta) \propto \underbrace{\boxed{p_{\mathbf{X}|\Theta=\theta}(\mathbf{x})}}_{\mathbf{P}(\text{data}\,|\,\text{model})} \underbrace{\boxed{f_\Theta(\theta)}}_{\mathbf{P}(\text{model})}$$

Bayesian Statistics: MAP Estimator

$$f_{\Theta|\mathbf{X}=\mathbf{x}}(\theta) \propto \boxed{p_{\mathbf{X}|\Theta=\theta}(\mathbf{x})}\boxed{f_{\Theta}(\theta)}$$

Does this look familiar? $\mathbf{P}$(data | model) $\mathbf{P}$(model)

Bayesian Statistics: MAP Estimator

$$f_{\Theta|\mathbf{X}=\mathbf{x}}(\theta) \propto \underbrace{\boxed{p_{\mathbf{X}|\Theta=\theta}(\mathbf{x})}}_{\mathbf{P}(\text{data} \mid \text{model})} \underbrace{\boxed{f_{\Theta}(\theta)}}_{\mathbf{P}(\text{model})}$$

Does this look familiar?

Point estimators?

Bayesian Statistics: MAP Estimator

$$f_{\Theta|\mathbf{X}=\mathbf{x}}(\theta) \propto \boxed{p_{\mathbf{X}|\Theta=\theta}(\mathbf{x})}\,\boxed{f_{\Theta}(\theta)}$$

Does this look familiar? **P**(data | model) **P**(model)

Point estimators?

$$\hat{\theta} = \arg\max_{\theta} f_{\Theta|\mathbf{X}=\mathbf{x}}(\theta) = \arg\max_{\theta} p_{\mathbf{X}|\Theta=\theta}(\mathbf{x}) f_{\Theta}(\theta)$$

Bayesian Statistics: MAP Estimator

$$f_{\Theta|\mathbf{X}=\mathbf{x}}(\theta) \propto \boxed{p_{\mathbf{X}|\Theta=\theta}(\mathbf{x})}\boxed{f_{\Theta}(\theta)}$$

Does this look familiar? $\mathbf{P}$(data | model) $\mathbf{P}$(model)

Point estimators?

$$\hat{\theta} = \arg\max_{\theta} f_{\Theta|\mathbf{X}=\mathbf{x}}(\theta) = \arg\max_{\theta} p_{\mathbf{X}|\Theta=\theta}(\mathbf{x}) f_{\Theta}(\theta)$$

This is called the Maximum a Posteriori (MAP) estimator

Bayesian Statistics: MAP Estimator

$$f_{\Theta|\mathbf{X}=\mathbf{x}}(\theta) \propto \boxed{p_{\mathbf{X}|\Theta=\theta}(\mathbf{x})}\boxed{f_{\Theta}(\theta)}$$

Does this look familiar? $\mathbf{P}$(data | model) $\mathbf{P}$(model)

Point estimators?

$$\hat{\theta} = \arg\max_{\theta} f_{\Theta|\mathbf{X}=\mathbf{x}}(\theta) = \arg\max_{\theta} p_{\mathbf{X}|\Theta=\theta}(\mathbf{x}) f_{\Theta}(\theta)$$

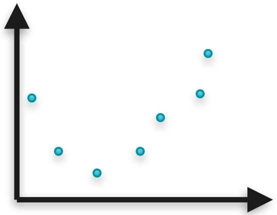
This is called the Maximum a Posteriori (MAP) estimator

Notice that if the prior is non informative, then MAP = MLE

Point Estimation

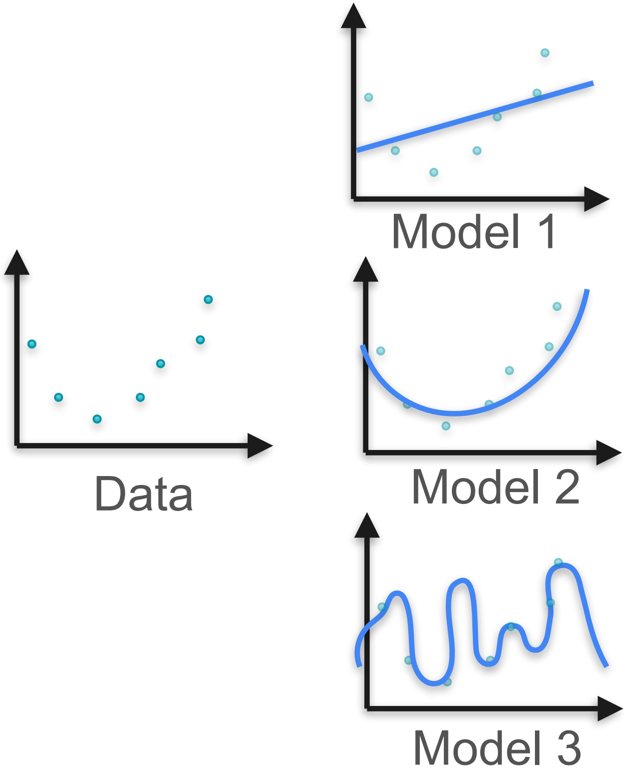
Regularization

Example: Polynomial Regression

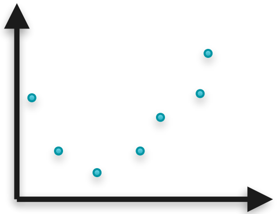


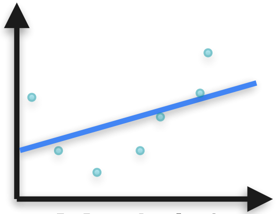
Data

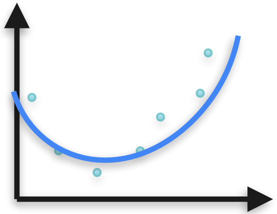
Example: Polynomial Regression

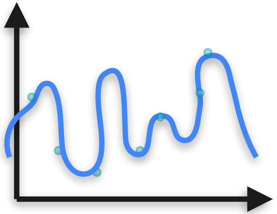


Example: Polynomial Regression



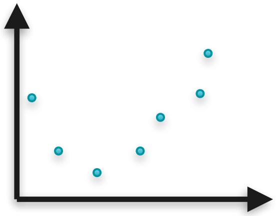


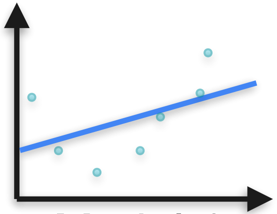


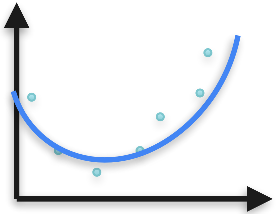


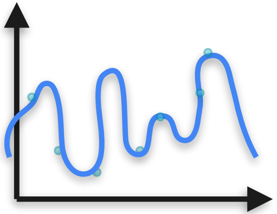
Loss

Example: Polynomial Regression





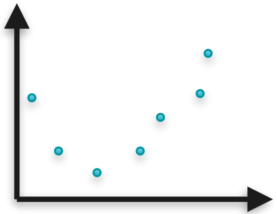




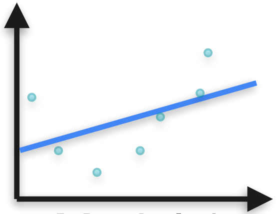
Loss

10

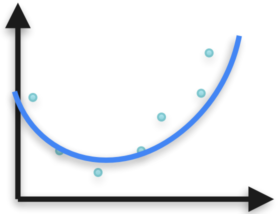
Example: Polynomial Regression



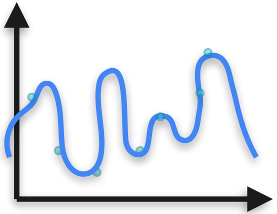
Data



Model 1



Model 2



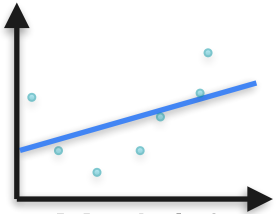
Model 3

Loss

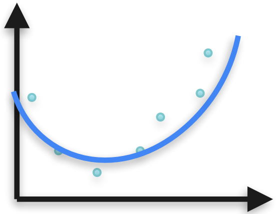
10

2

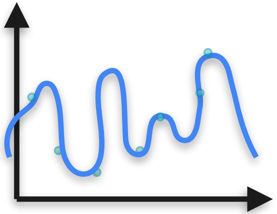
Example: Polynomial Regression



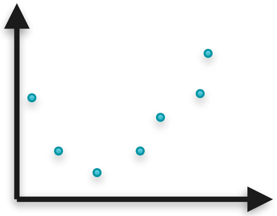
Model 1



Model 2



Model 3



Data

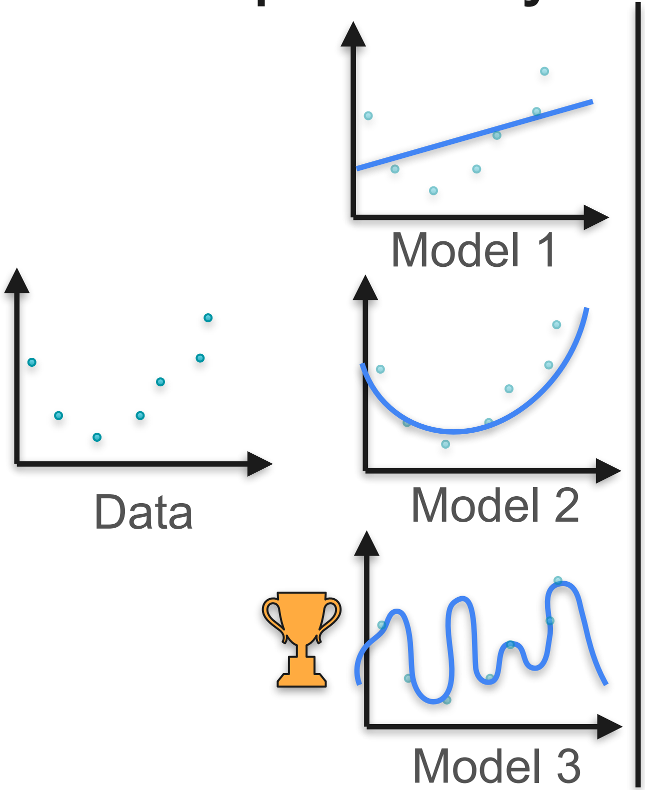
Loss

10

2

0.1

Example: Polynomial Regression



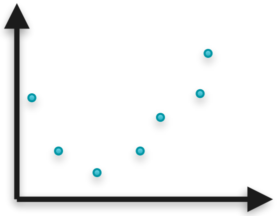
Loss

10

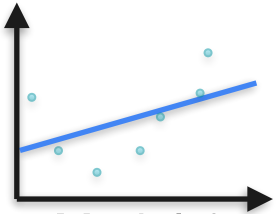
2

0.1

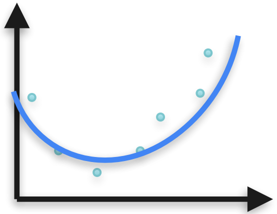
Example: Polynomial Regression



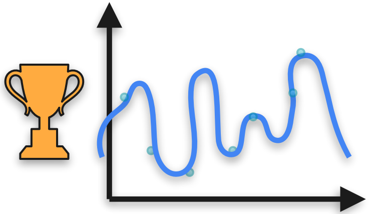
Data



Model 1



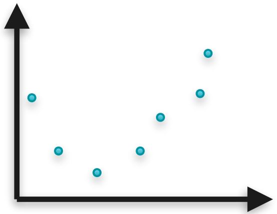
Model 2



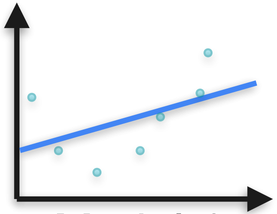
Model 3

Loss	Equation
10	
2	
0.1	

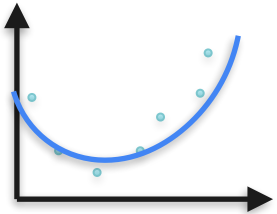
Example: Polynomial Regression



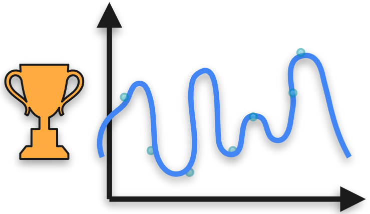
Data



Model 1



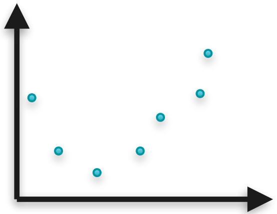
Model 2



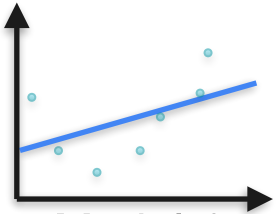
Model 3

	Loss	Equation
Model 1	10	$y = 4x + 3$
Model 2	2	
Model 3	0.1	

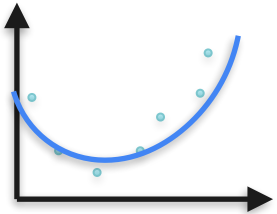
Example: Polynomial Regression



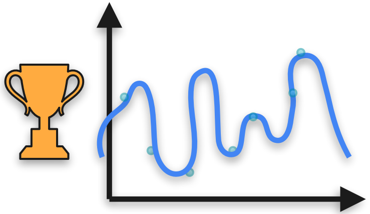
Data



Model 1



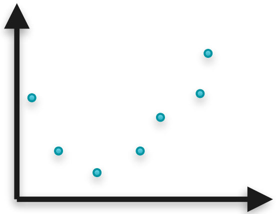
Model 2



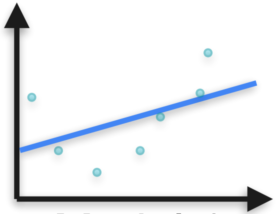
Model 3

	Loss	Equation
Model 1	10	$y = 4x + 3$
Model 2	2	$y = 2x^2 - 4x + 5$
Model 3	0.1	

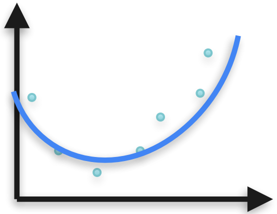
Example: Polynomial Regression



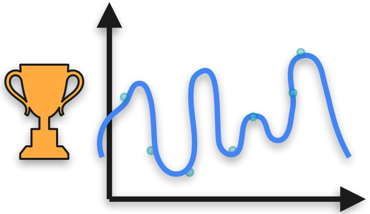
Data



Model 1



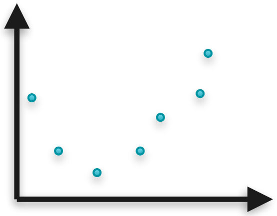
Model 2



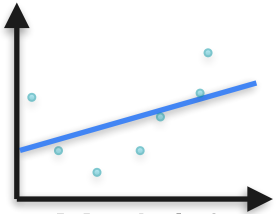
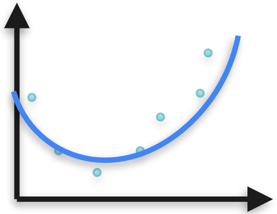
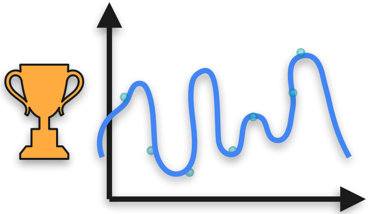
Model 3

	Loss	Equation
Model 1	10	$y = 4x + 3$
Model 2	2	$y = 2x^2 - 4x + 5$
Model 3	0.1	$y = 4x^{10} - 9x^8 - 2x^6$ $+3x^5 - 6x^4 - 10x + 4$

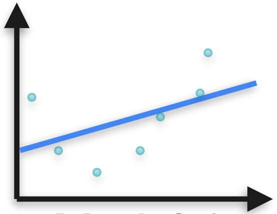
Example: Polynomial Regression



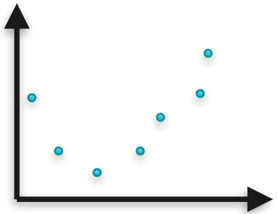
Data

	Loss	Equation	Penalty
 Model 1	10	$y = 4x + 3$	
 Model 2	2	$y = 2x^2 - 4x + 5$	
 Model 3	0.1	$y = 4x^{10} - 9x^8 - 2x^6$ $+3x^5 - 6x^4 - 10x + 4$	

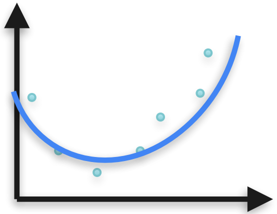
Example: Polynomial Regression



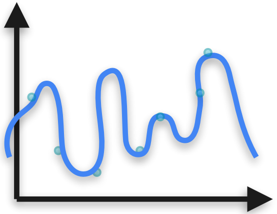
Model 1



Data



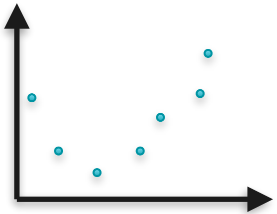
Model 2

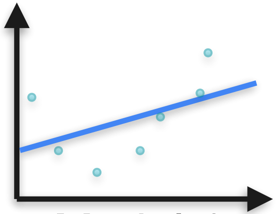


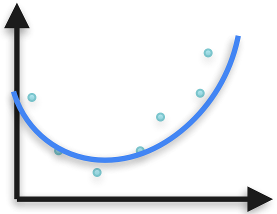
Model 3

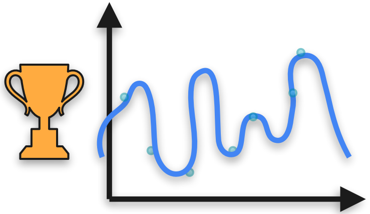
Loss	Equation	Penalty
10	$y = \boxed{4x} + 3$	$L_2 = 4^2$
2	$y = 2x^2 - 4x + 5$	
0.1	$y = 4x^{10} - 9x^8 - 2x^6 + 3x^5 - 6x^4 - 10x + 4$	

Example: Polynomial Regression



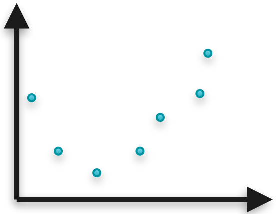




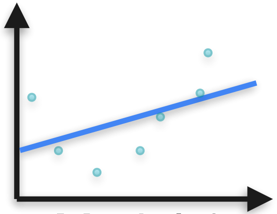
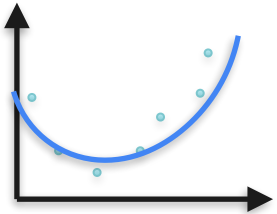
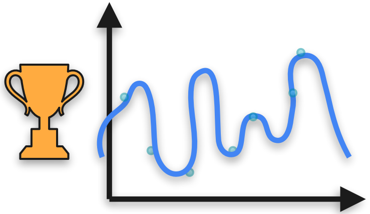


	Loss	Equation	Penalty
Model 1	10	$y = \boxed{4}x + 3$	$L_2 = 4^2 = 16$
Model 2	2	$y = 2x^2 - 4x + 5$	
Model 3	0.1	$y = 4x^{10} - 9x^8 - 2x^6 + 3x^5 - 6x^4 - 10x + 4$	

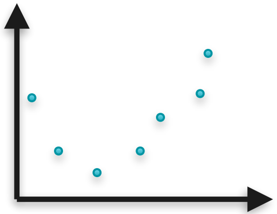
Example: Polynomial Regression



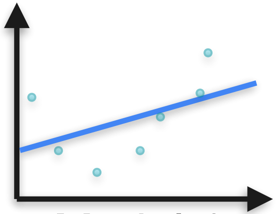
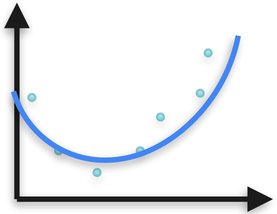
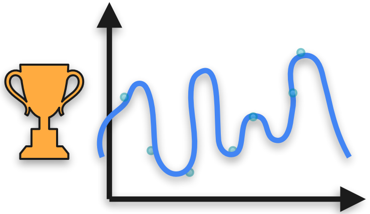
Data

	Loss	Equation	Penalty
 Model 1	10	$y = \boxed{4}x + 3$	$L_2 = 4^2 = 16$
 Model 2	2	$y = \boxed{2}x^2 - \boxed{4}x + 5$	$L_2 = 2^2 + (-4)^2$
 Model 3	0.1	$y = 4x^{10} - 9x^8 - 2x^6 + 3x^5 - 6x^4 - 10x + 4$	

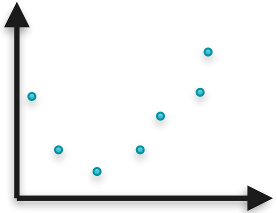
Example: Polynomial Regression



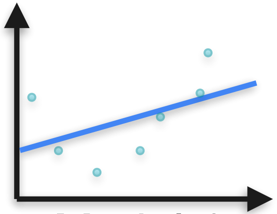
Data

	Loss	Equation	Penalty
 Model 1	10	$y = \boxed{4}x + 3$	$L_2 = 4^2$ $= 16$
 Model 2	2	$y = \boxed{2}x^2 \boxed{-4}x + 5$	$L_2 = 2^2 + (-4)^2$ $= 20$
 Model 3	0.1	$y = 4x^{10} - 9x^8 - 2x^6$ $+3x^5 - 6x^4 - 10x + 4$	

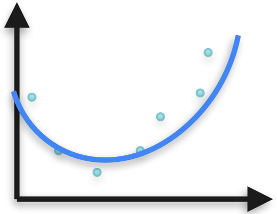
Example: Polynomial Regression



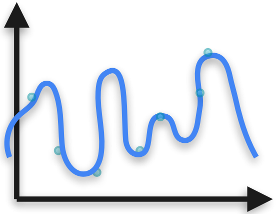
Data



Model 1



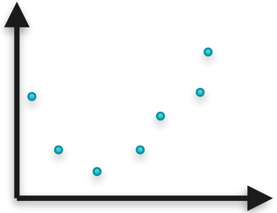
Model 2



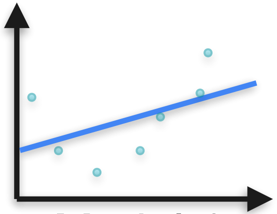
Model 3

	Loss	Equation	Penalty
Model 1	10	$y = \boxed{4}x + 3$	$L_2 = 4^2 = 16$
Model 2	2	$y = \boxed{2}x^2 \boxed{-4}x + 5$	$L_2 = 2^2 + (-4)^2 = 20$
Model 3	0.1	$y = \boxed{4}x^{10} \boxed{-9}x^8 \boxed{-2}x^6 \boxed{+3}x^5 \boxed{-6}x^4 \boxed{-10}x + 4$	$L_2 = 4^2 + (-9)^2 + (-2)^2 + 3^2 + (-6)^2 + (-10)^2$

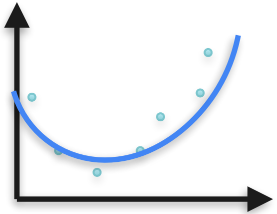
Example: Polynomial Regression



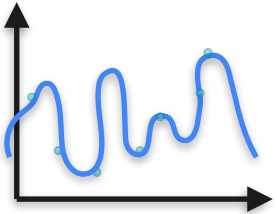
Data



Model 1



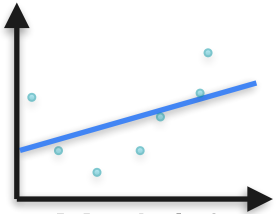
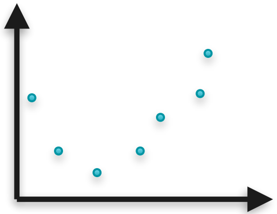
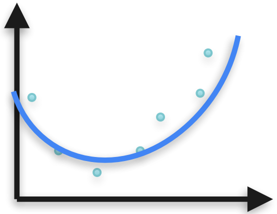
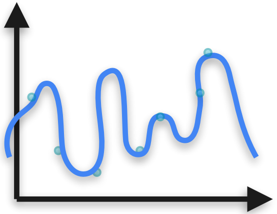
Model 2



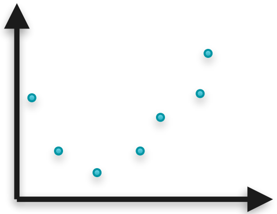
Model 3

	Loss	Equation	Penalty
Model 1	10	$y = \boxed{4}x + 3$	$L_2 = 4^2$ $= 16$
Model 2	2	$y = \boxed{2}x^2 \boxed{-4}x + 5$	$L_2 = 2^2 + (-4)^2$ $= 20$
Model 3	0.1	$y = \boxed{4}x^{10} \boxed{-9}x^8 \boxed{-2}x^6$ $\boxed{+3}x^5 \boxed{-6}x^4 \boxed{-10}x + 4$	$L_2 = 4^2 + (-9)^2 + (-2)^2$ $+3^2 + (-6)^2 + (-10)^2$ $= 246$

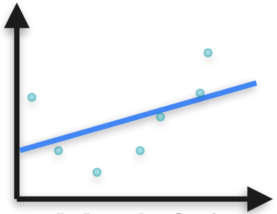
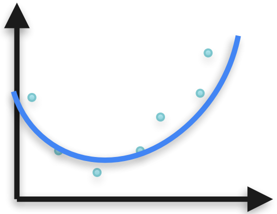
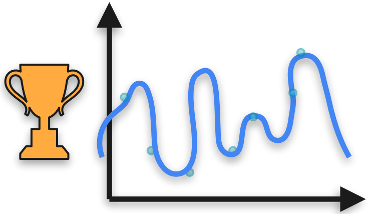
Example: Polynomial Regression

	Loss	Equation	Penalty	New loss
 Model 1	10	$y = \boxed{4}x + 3$	$L_2 = 4^2 = 16$	
 Data  Model 2	2	$y = \boxed{2}x^2 \boxed{-4}x + 5$	$L_2 = 2^2 + (-4)^2 = 20$	
  Model 3	0.1	$y = \boxed{4}x^{10} \boxed{-9}x^8 \boxed{-2}x^6 \boxed{+3}x^5 \boxed{-6}x^4 \boxed{-10}x + 4$	$L_2 = 4^2 + (-9)^2 + (-2)^2 + 3^2 + (-6)^2 + (-10)^2 = 246$	

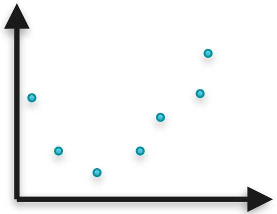
Example: Polynomial Regression



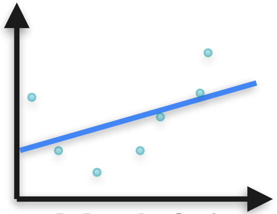
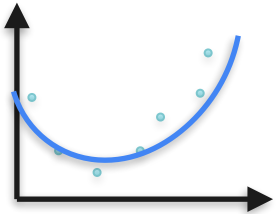
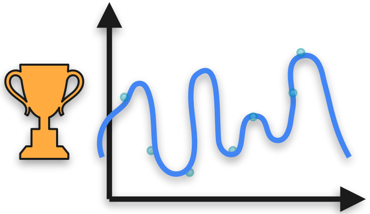
Data

	Loss	Equation	Penalty	New loss
 Model 1	$\boxed{10}$	$y = \boxed{4}x + 3$	$L_2 = 4^2 = \boxed{16}$	26
 Model 2	2	$y = \boxed{2}x^2 \boxed{-4}x + 5$	$L_2 = 2^2 + (-4)^2 = 20$	
 Model 3	0.1	$y = \boxed{4}x^{10} \boxed{-9}x^8 \boxed{-2}x^6 \boxed{+3}x^5 \boxed{-6}x^4 \boxed{-10}x + 4$	$L_2 = 4^2 + (-9)^2 + (-2)^2 + 3^2 + (-6)^2 + (-10)^2 = 246$	

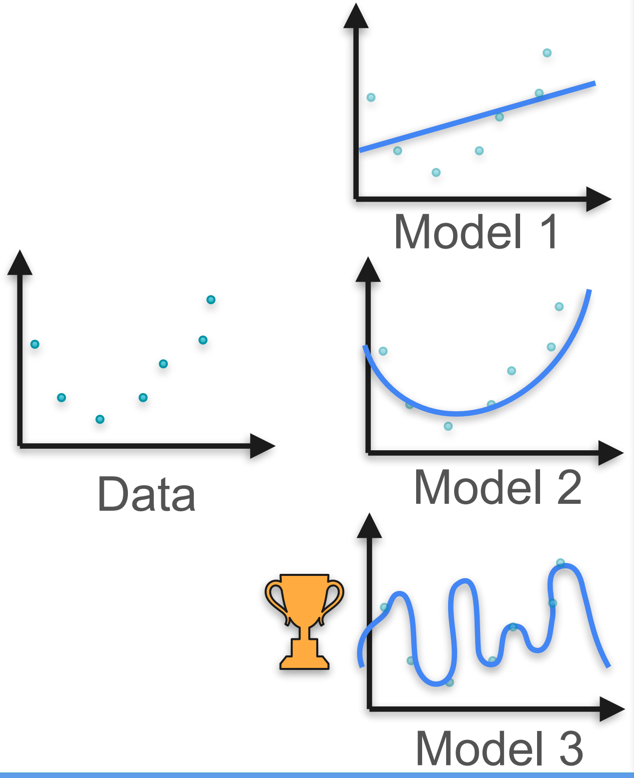
Example: Polynomial Regression



Data

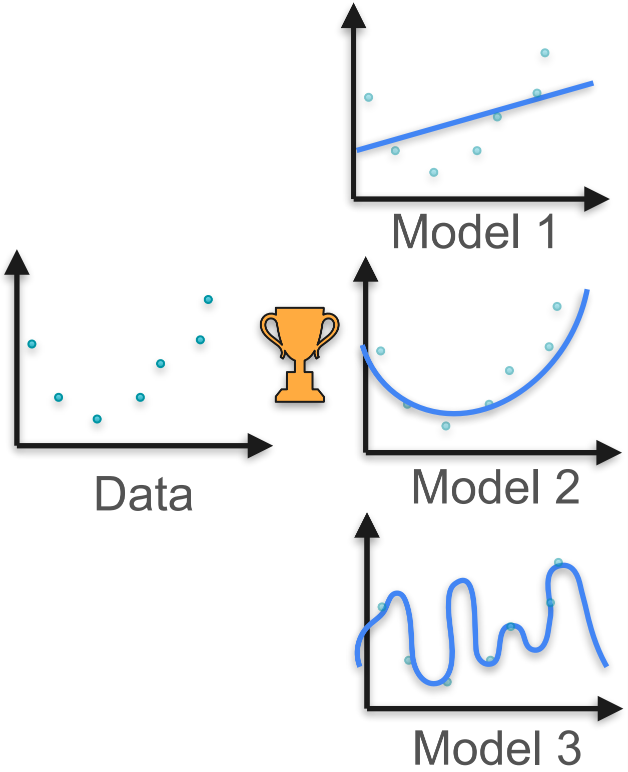
	Loss	Equation	Penalty	New loss
 Model 1	$\boxed{10}$	$y = \boxed{4}x + 3$	$L_2 = 4^2 = \boxed{16}$	26
 Model 2	$\boxed{2}$	$y = \boxed{2}x^2 \boxed{-4}x + 5$	$L_2 = 2^2 + (-4)^2 = \boxed{20}$	22
 Model 3	0.1	$y = \boxed{4}x^{10} \boxed{-9}x^8 \boxed{-2}x^6 \boxed{+3}x^5 \boxed{-6}x^4 \boxed{-10}x + 4$	$L_2 = 4^2 + (-9)^2 + (-2)^2 + 3^2 + (-6)^2 + (-10)^2 = 246$	

Example: Polynomial Regression



	Loss	Equation	Penalty	New loss
Model 1	$\boxed{10}$	$y = \boxed{4}x + 3$	$L_2 = 4^2 = \boxed{16}$	26
Model 2	$\boxed{2}$	$y = \boxed{2}x^2 \boxed{-4}x + 5$	$L_2 = 2^2 + (-4)^2 = \boxed{20}$	22
Model 3	$\boxed{0.1}$	$y = \boxed{4}x^{10} \boxed{-9}x^8 \boxed{-2}x^6 \boxed{+3}x^5 \boxed{-6}x^4 \boxed{-10}x + 4$	$L_2 = 4^2 + (-9)^2 + (-2)^2 + 3^2 + (-6)^2 + (-10)^2 = \boxed{246}$	246.1

Example: Polynomial Regression



	Loss	Equation	Penalty	New loss
Model 1	$\boxed{10}$	$y = \boxed{4}x + 3$	$L_2 = 4^2 = \boxed{16}$	26
Model 2	$\boxed{2}$	$y = \boxed{2}x^2 \boxed{-4}x + 5$	$L_2 = 2^2 + (-4)^2 = \boxed{20}$	22
Model 3	$\boxed{0.1}$	$y = \boxed{4}x^{10} \boxed{-9}x^8 \boxed{-2}x^6 \boxed{+3}x^5 \boxed{-6}x^4 \boxed{-10}x + 4$	$L_2 = 4^2 + (-9)^2 + (-2)^2 + 3^2 + (-6)^2 + (-10)^2 = \boxed{246}$	246.1

Regularization Term

Regularization Term

Model: $y = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$

Regularization Term

Model: $y = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$

Log-loss: $\ell\ell$

Regularization Term

Model: $y = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$

Log-loss: $\ell\ell$

L2 Regularization Error: $a_n^2 + a_{n-1}^2 + \cdots + a_1^2$

Regularization Term

Model: $y = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$

Log-loss: $\ell\ell$

L2 Regularization Error: $a_n^2 + a_{n-1}^2 + \cdots + a_1^2$

Regularization parameter: λ

Regularization Term

Model: $y = \boxed{a_n} x^n + \boxed{a_{n-1}} x^{n-1} + \cdots + \boxed{a_1} x + a_0$

Log-loss: $\ell\ell$

L2 Regularization Error: $a_n^2 + a_{n-1}^2 + \cdots + a_1^2$

Regularization parameter: λ

Regularization Term

Model: $y = \boxed{a_n} x^n + \boxed{a_{n-1}} x^{n-1} + \cdots + \boxed{a_1} x + a_0$

Log-loss: $\ell\ell$

L2 Regularization Error: $a_n^2 + a_{n-1}^2 + \cdots + a_1^2$

Regularization parameter: λ

Regularized error:

Regularization Term

Model: $y = \boxed{a_n} x^n + \boxed{a_{n-1}} x^{n-1} + \cdots + \boxed{a_1} x + a_0$

Log-loss: $\ell\ell$

L2 Regularization Error: $a_n^2 + a_{n-1}^2 + \cdots + a_1^2$

Regularization parameter: λ

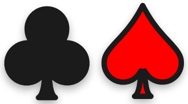
Regularized error: $\ell\ell + \lambda\,(a_n^2 + a_{n-1}^2 + \cdots + a_1^2)$

Point Estimation

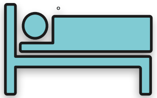
Back to Bayesics

There’s Popcorn on the Floor. What Happened?

Movies

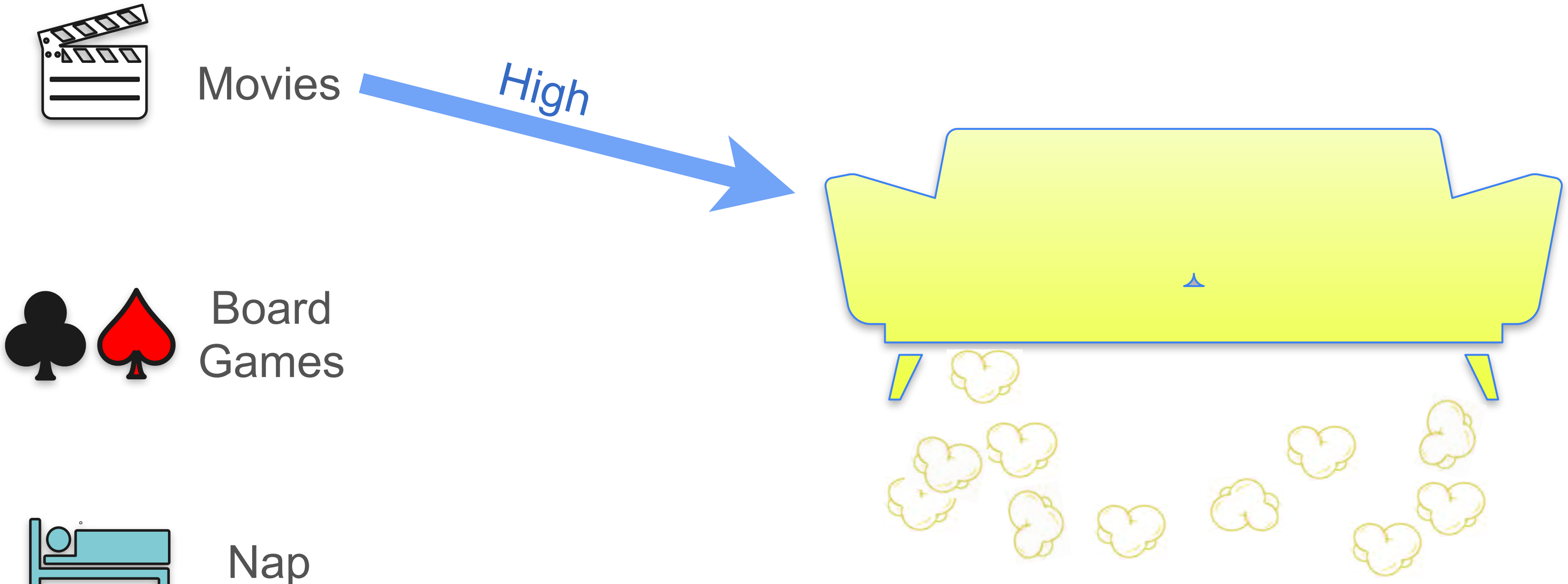


Board Games

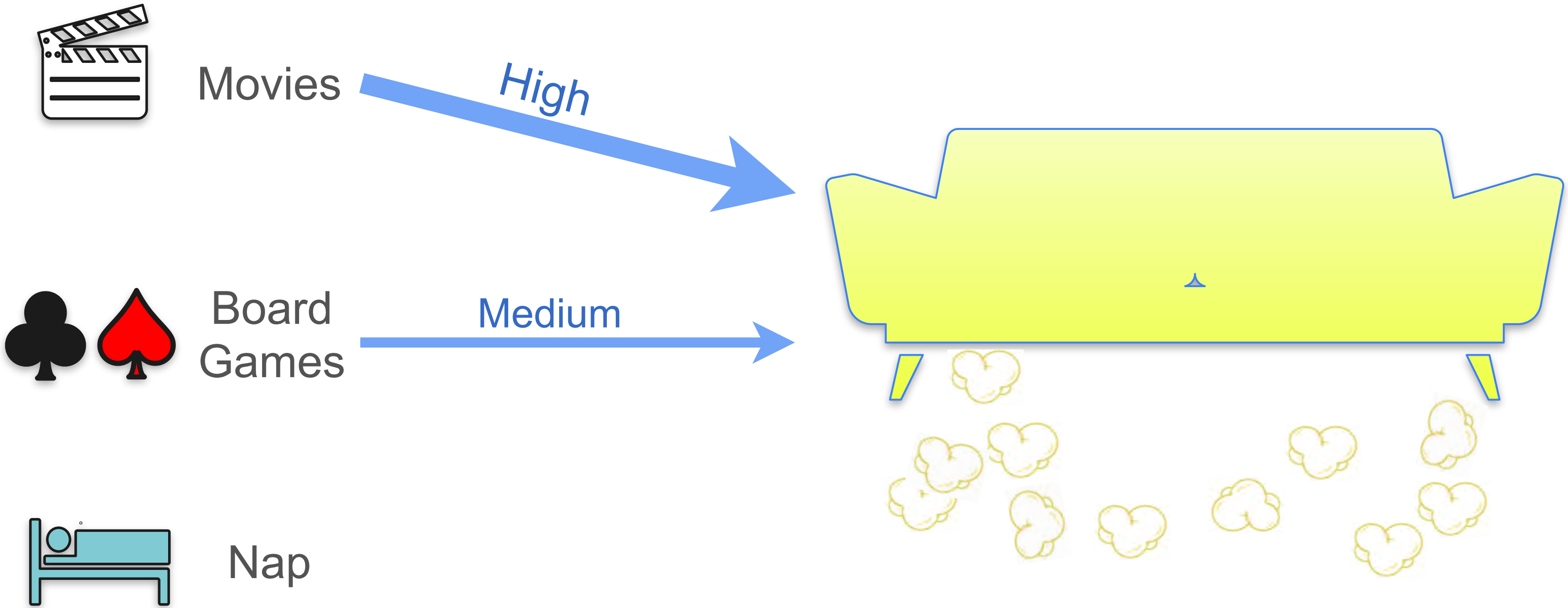


Nap

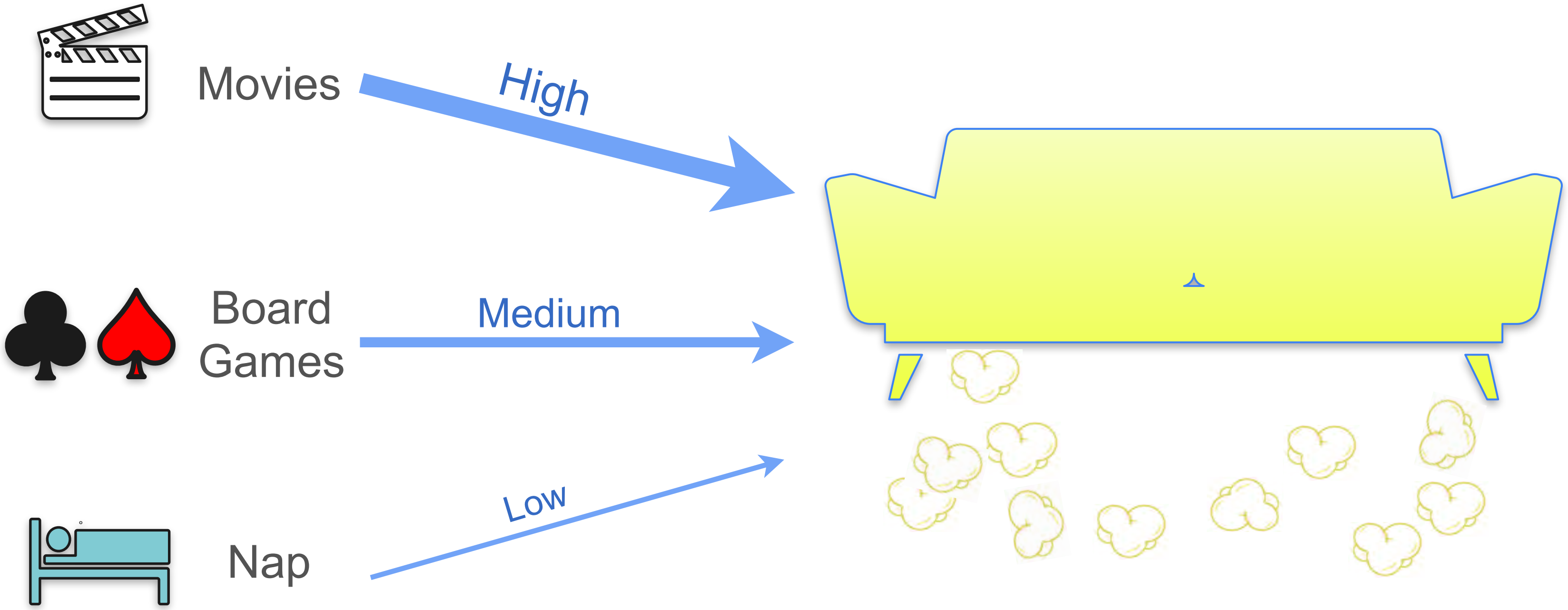
There’s Popcorn on the Floor. What Happened?



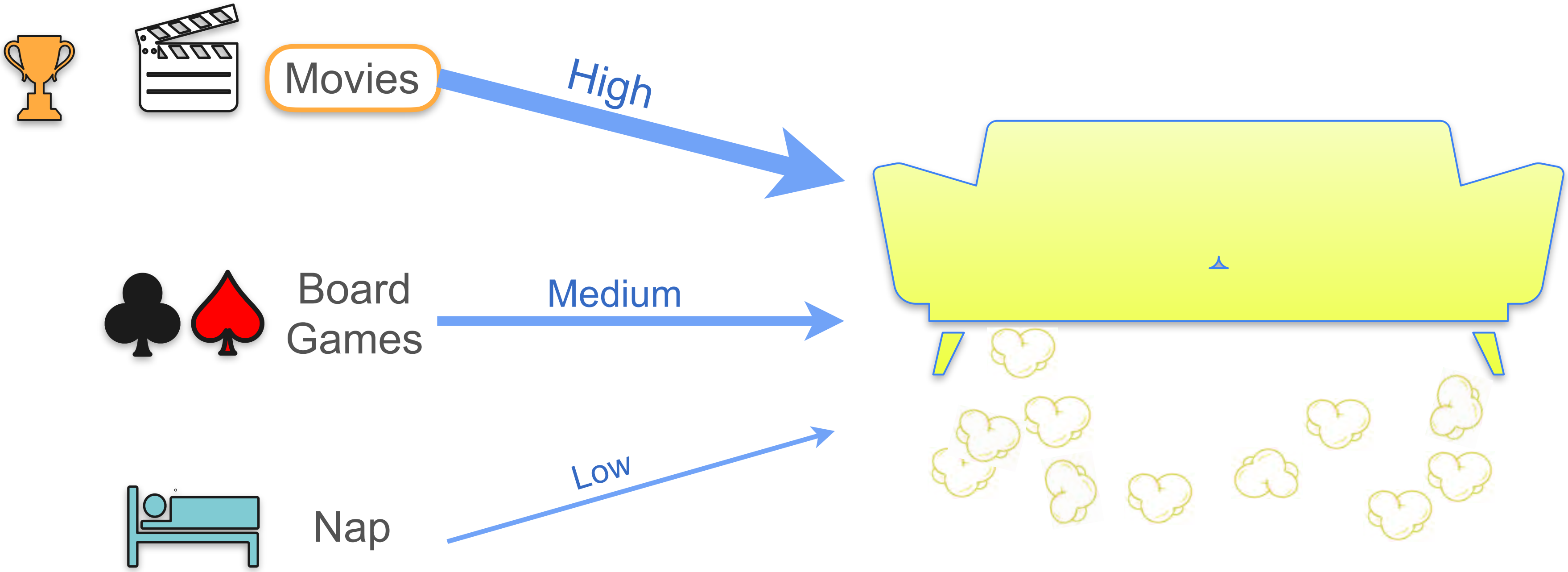
There's Popcorn on the Floor. What Happened?



There's Popcorn on the Floor. What Happened?



There's Popcorn on the Floor. What Happened?

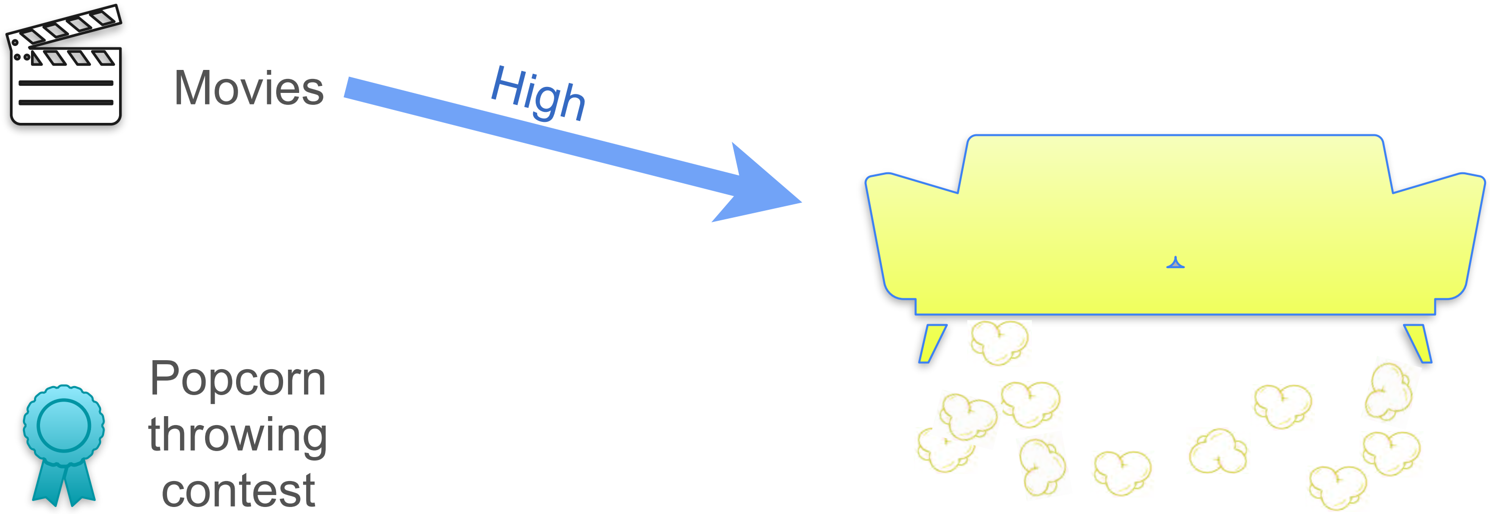


There’s Popcorn on the Floor. What Happened?

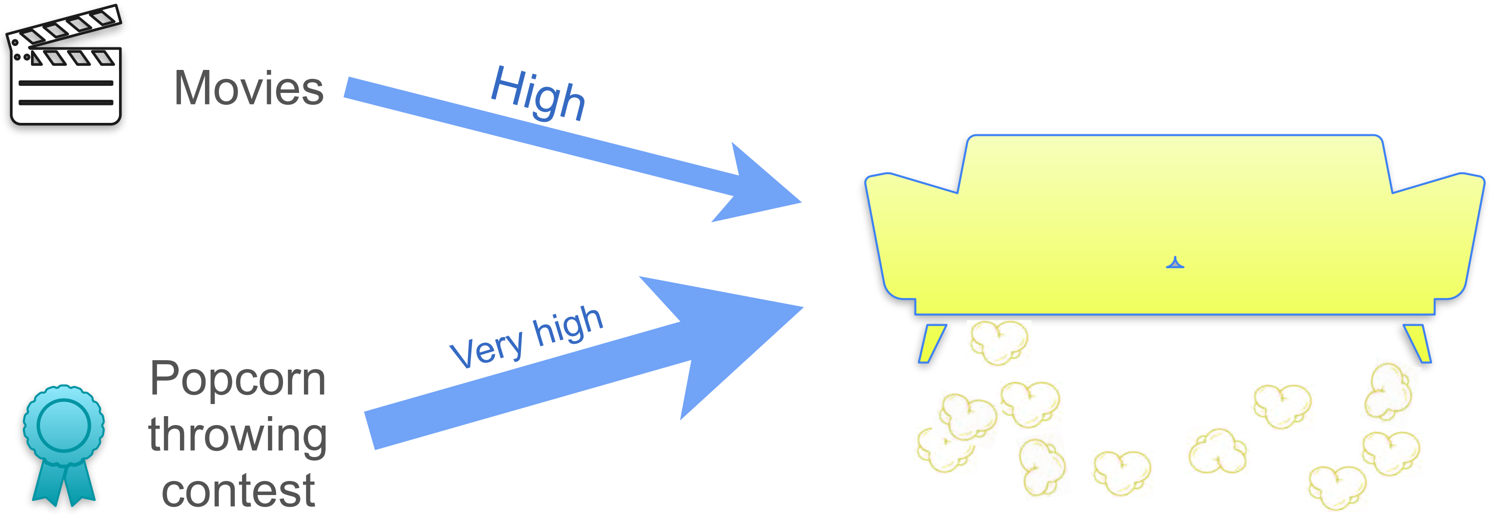
Movies

Popcorn throwing contest

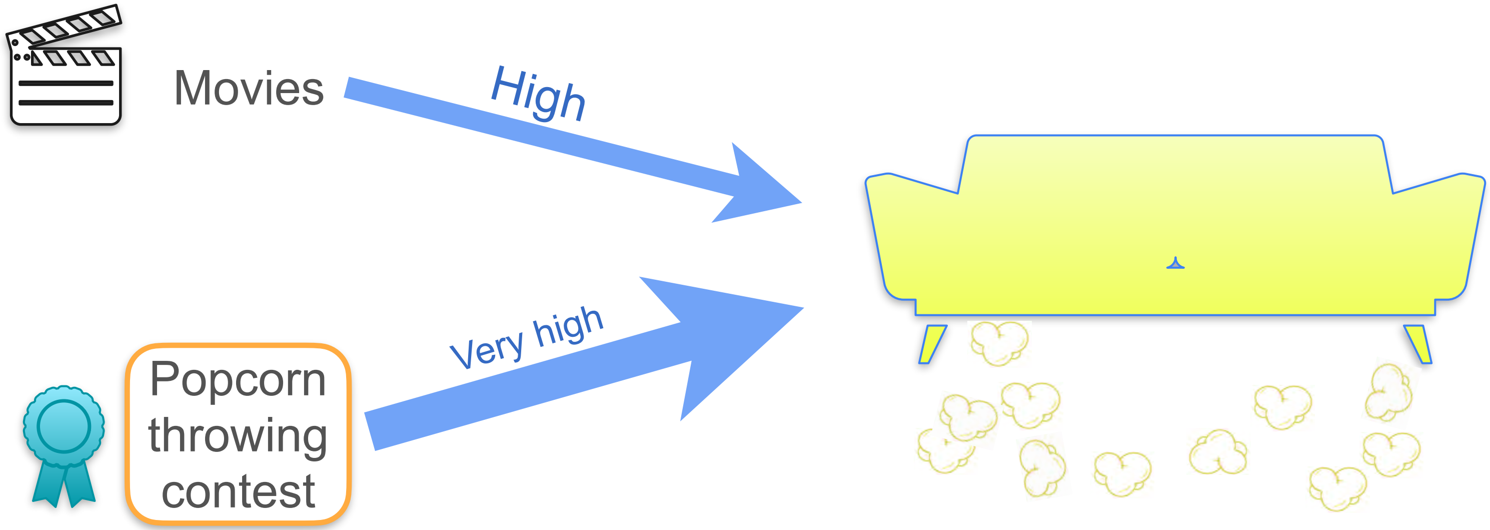
There’s Popcorn on the Floor. What Happened?



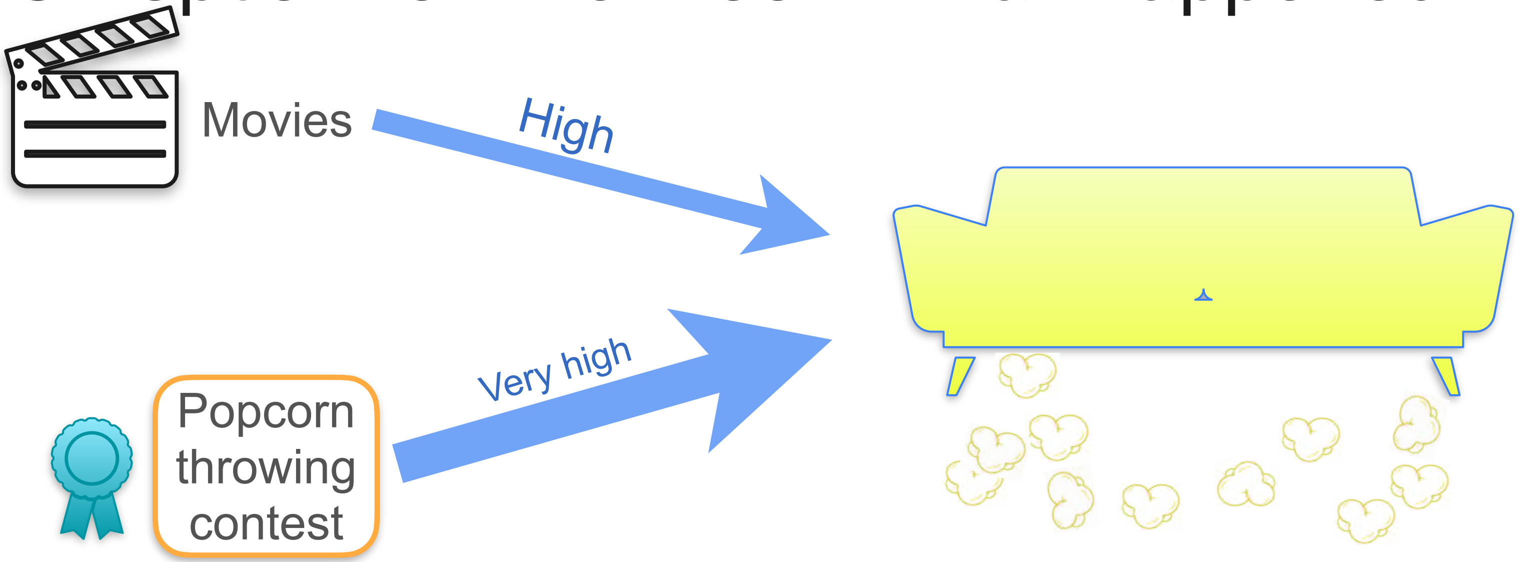
There’s Popcorn on the Floor. What Happened?



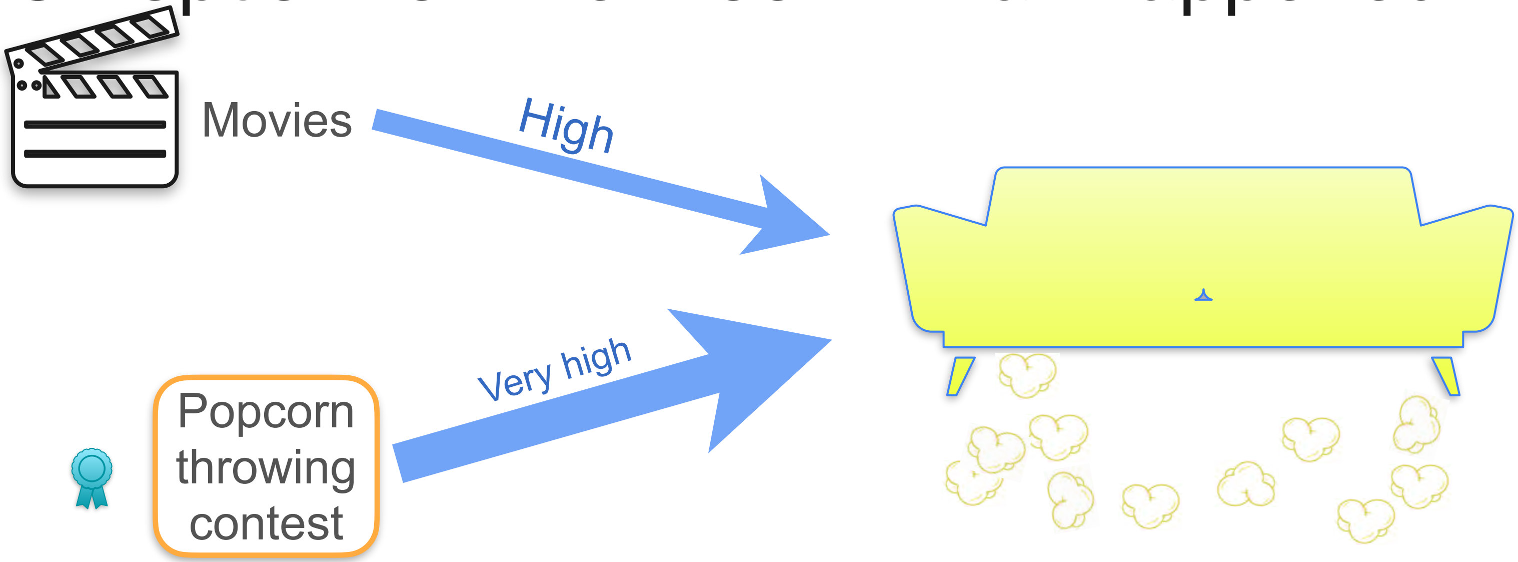
There’s Popcorn on the Floor. What Happened?



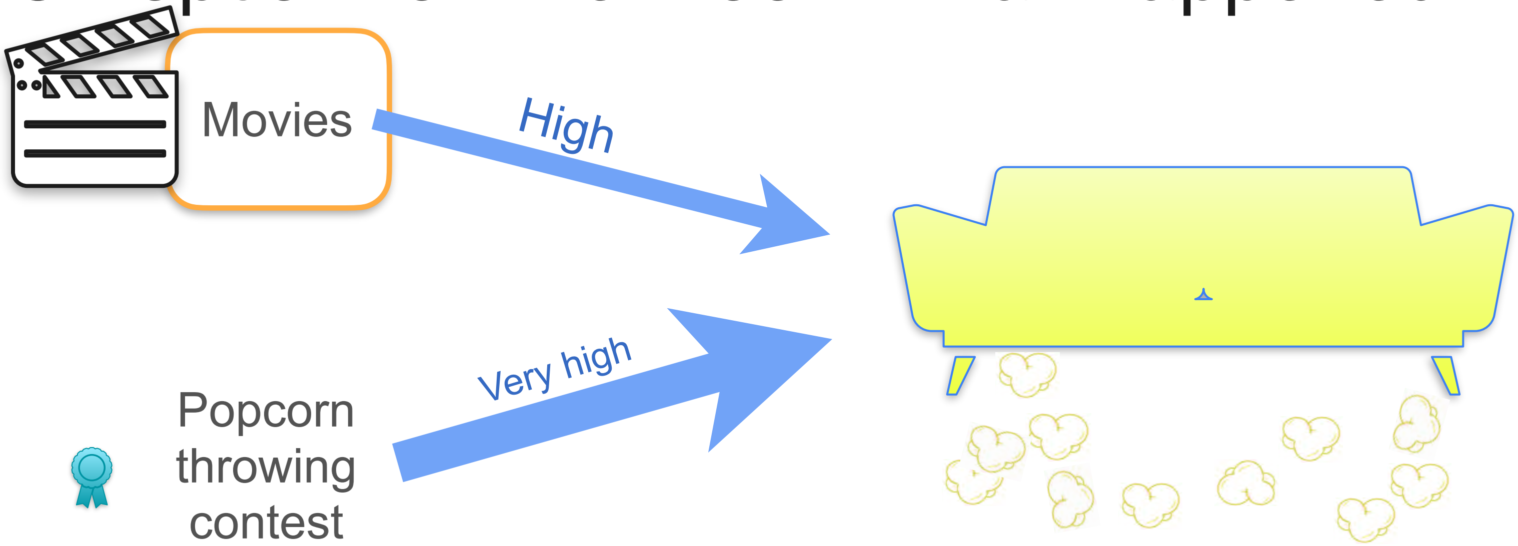
There's Popcorn on the Floor. What Happened?



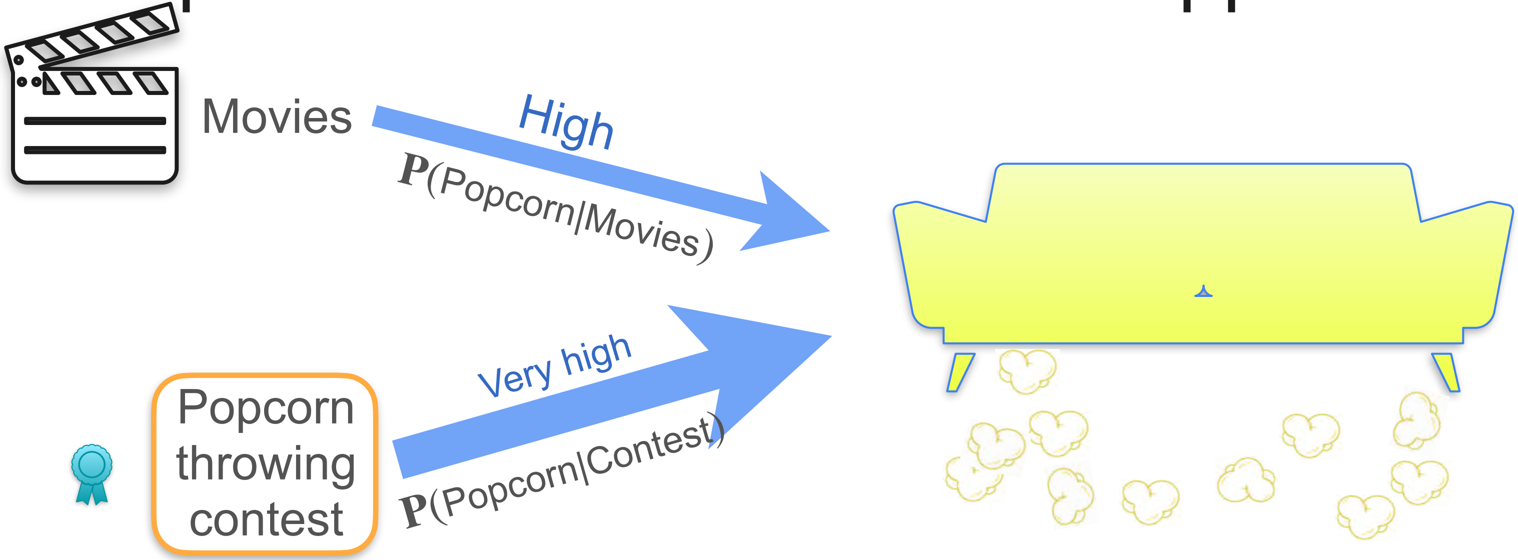
There's Popcorn on the Floor. What Happened?



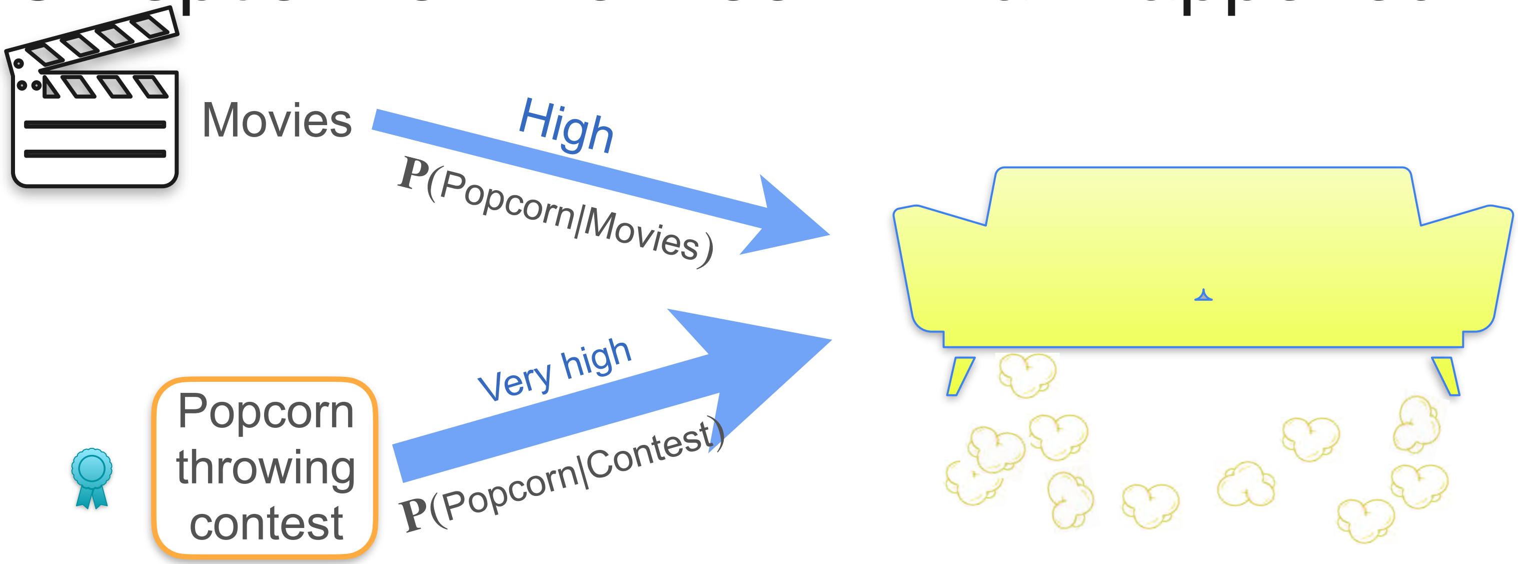
There's Popcorn on the Floor. What Happened?



There's Popcorn on the Floor. What Happened?

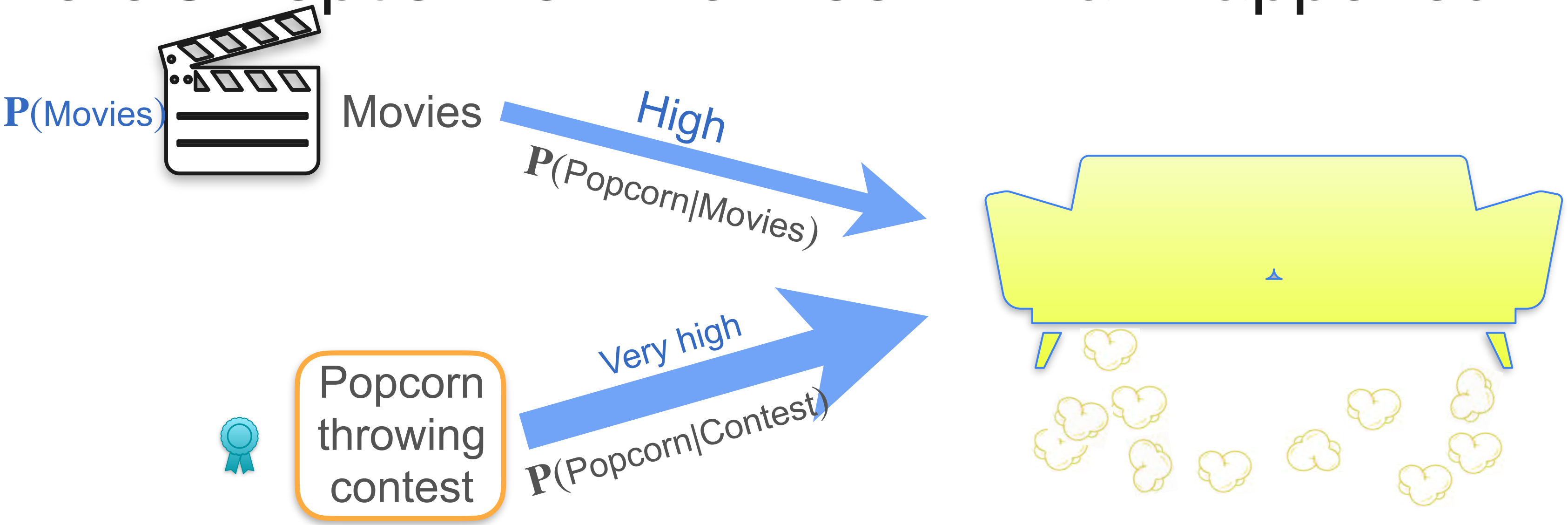


There's Popcorn on the Floor. What Happened?



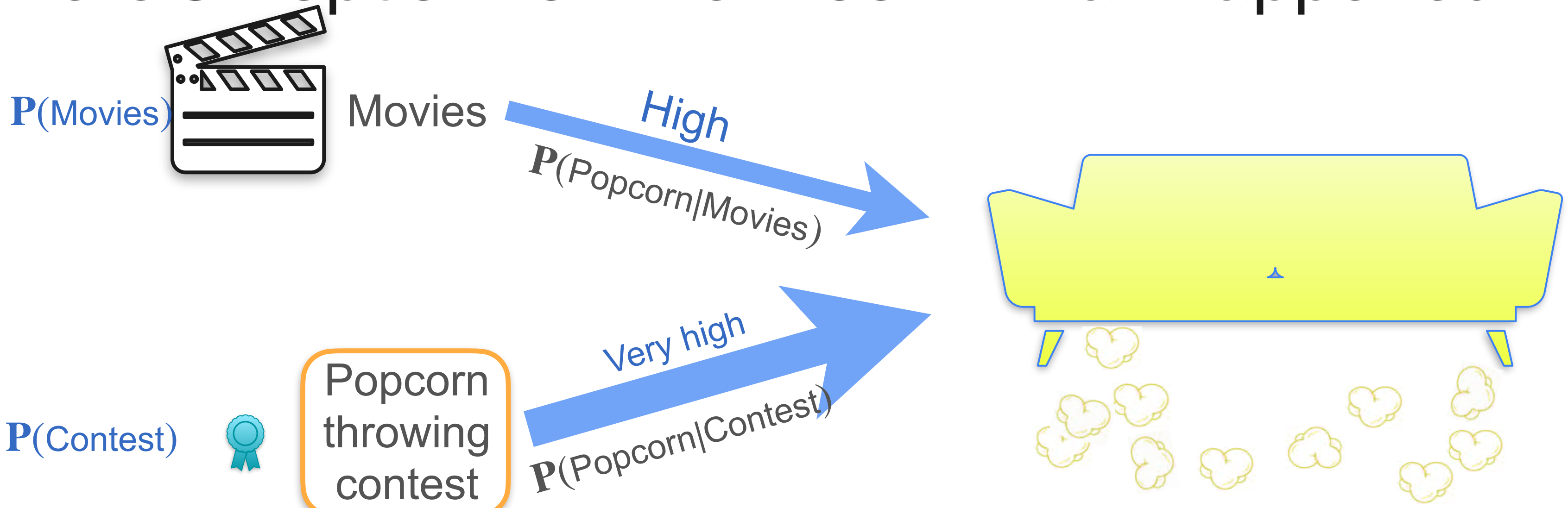
P(Popcorn|Movies) < P(Popcorn|Contest)

There's Popcorn on the Floor. What Happened?



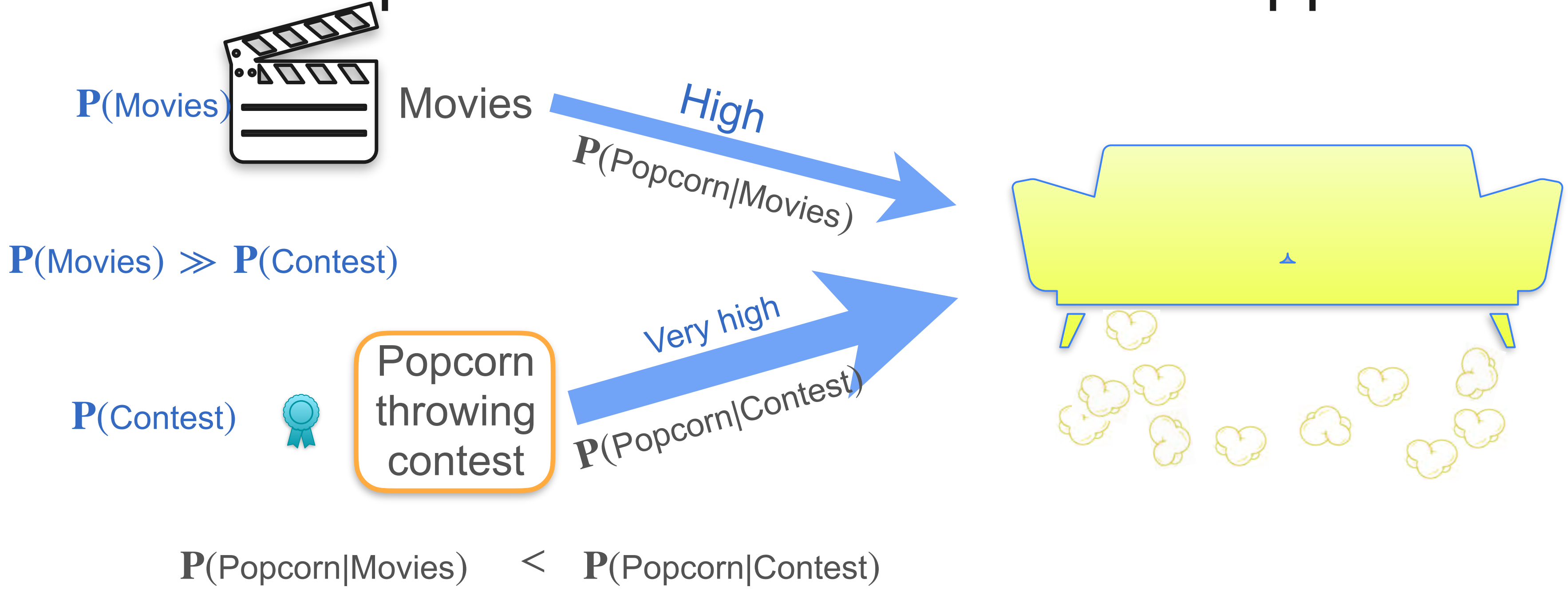
$$P(\text{Popcorn}|\text{Movies}) < P(\text{Popcorn}|\text{Contest})$$

There's Popcorn on the Floor. What Happened?

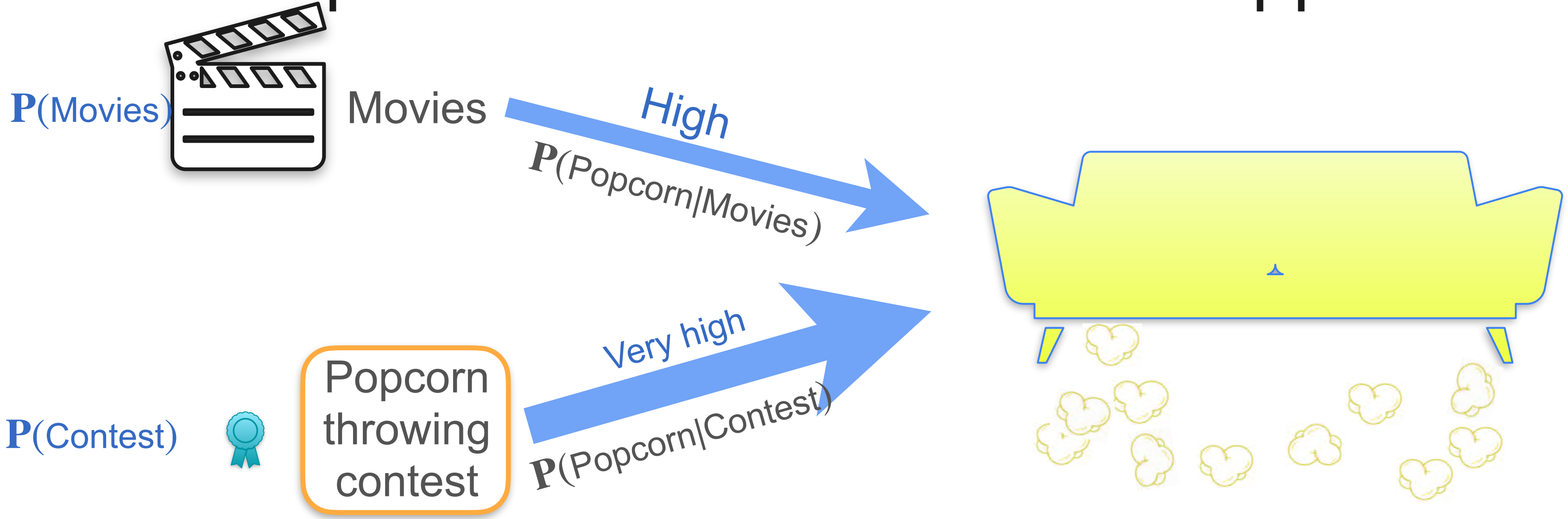


$$\mathbf{P}(\text{Popcorn}|\text{Movies}) < \mathbf{P}(\text{Popcorn}|\text{Contest})$$

There's Popcorn on the Floor. What Happened?

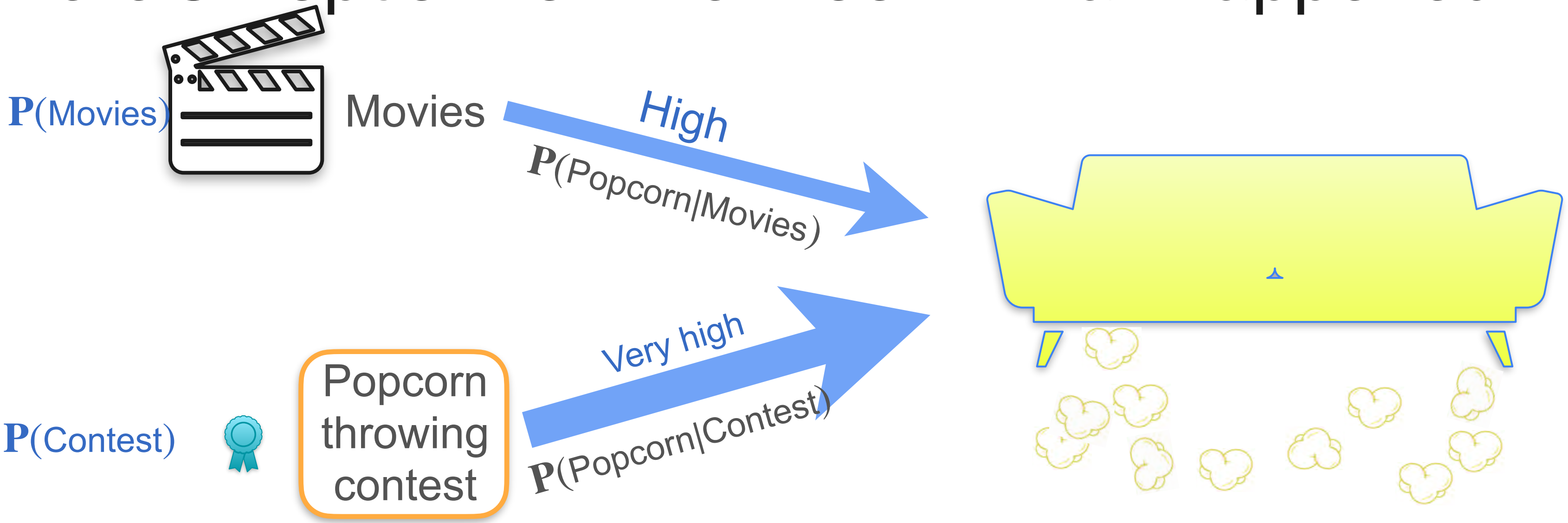


There's Popcorn on the Floor. What Happened?



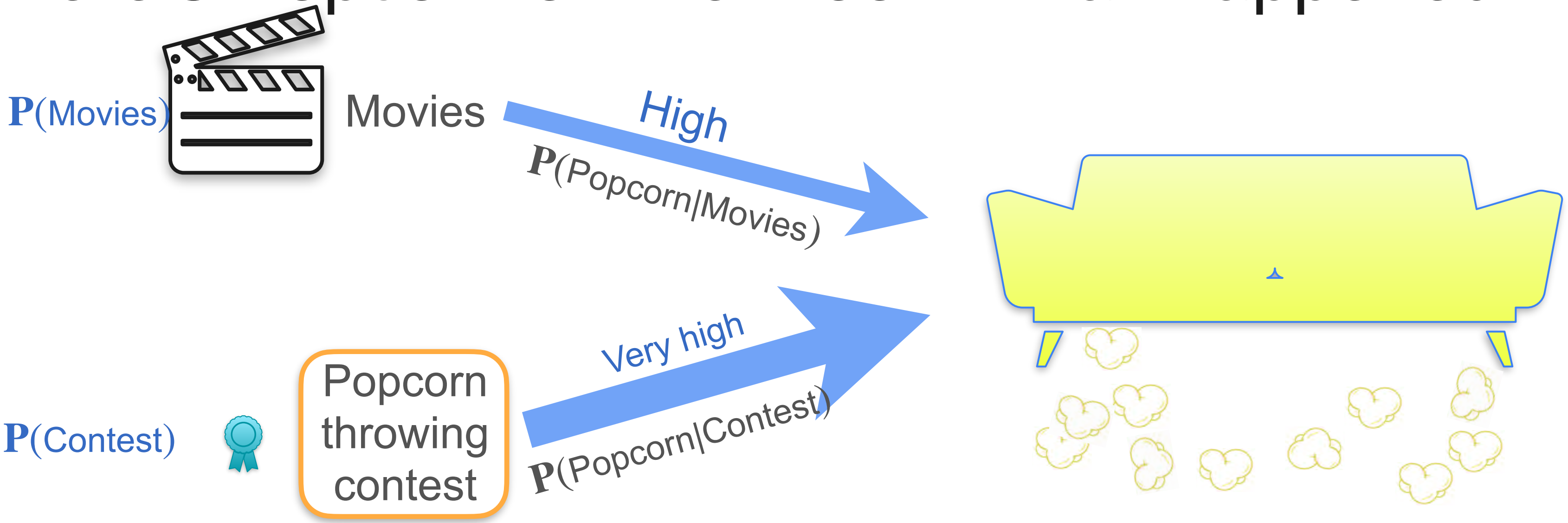
$$\mathbf{P}(\text{Popcorn}|\text{Movies})\ \mathbf{P}(\text{Movies}) < \mathbf{P}(\text{Popcorn}|\text{Contest})\ \mathbf{P}(\text{Contest})$$

There's Popcorn on the Floor. What Happened?



$$\mathbf{P}(\text{Popcorn}|\text{Movies})\ \mathbf{P}(\text{Movies}) > \mathbf{P}(\text{Popcorn}|\text{Contest})\ \mathbf{P}(\text{Contest})$$

There's Popcorn on the Floor. What Happened?



$\mathbf{P}(\text{Popcorn}|\text{Movies})\ \mathbf{P}(\text{Movies}) > \mathbf{P}(\text{Popcorn}|\text{Contest})\ \mathbf{P}(\text{Contest})$

$$\boxed{\mathbf{P}(A \mid B)\mathbf{P}(B) = \mathbf{P}(A \cap B)}$$

There's Popcorn on the Floor. What Happened?

$\mathbf{P}$(Movies)

Movies

High

$\mathbf{P}$(Popcorn|Movies)

$\mathbf{P}$(Contest)

Popcorn throwing contest

Very high

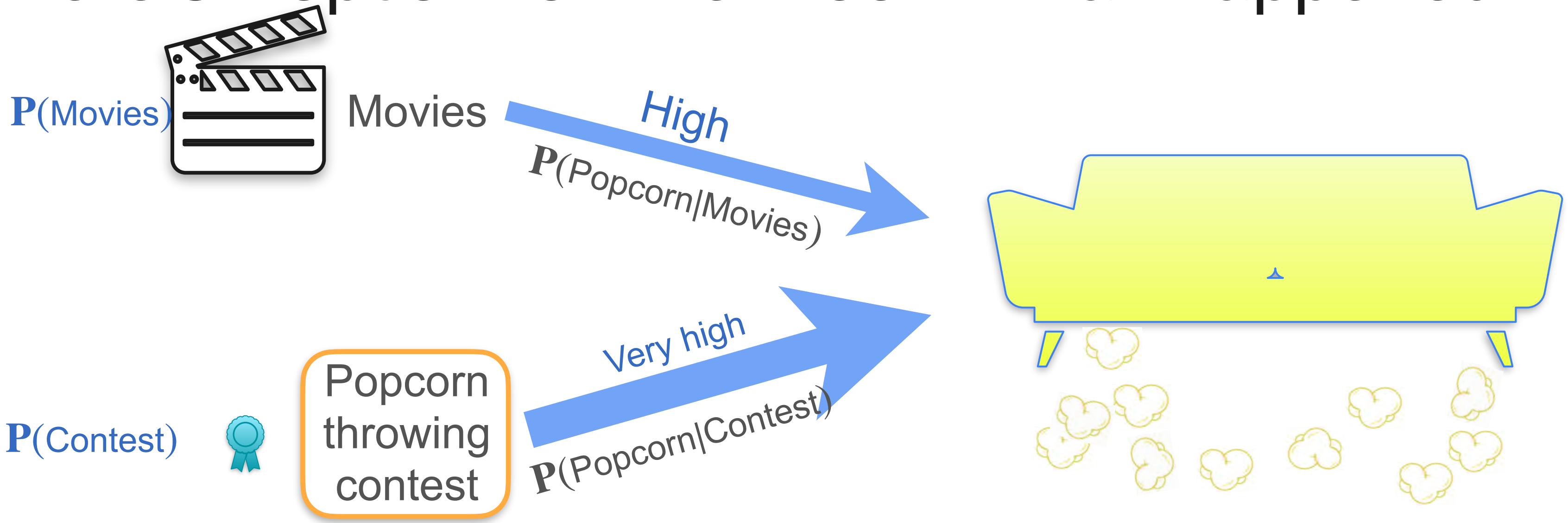
$\mathbf{P}$(Popcorn|Contest)

$\mathbf{P}$(Popcorn|Movies) $\mathbf{P}$(Movies) $>$ $\mathbf{P}$(Popcorn|Contest) $\mathbf{P}$(Contest)

$$\boxed{\mathbf{P}(A\,|\,B)\mathbf{P}(B) = \mathbf{P}(A \cap B)}$$

$\mathbf{P}$(Popcorn ∩ Movie)

There's Popcorn on the Floor. What Happened?

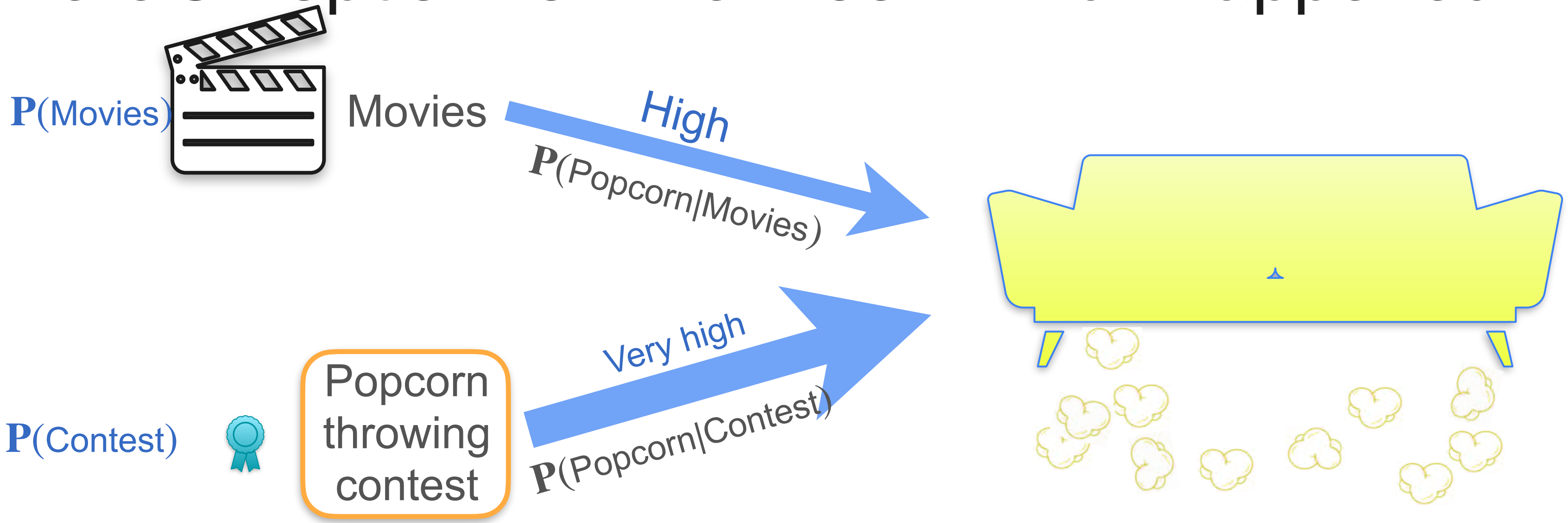


$\mathbf{P}(\text{Popcorn}|\text{Movies})\ \mathbf{P}(\text{Movies}) > \mathbf{P}(\text{Popcorn}|\text{Contest})\ \mathbf{P}(\text{Contest})$

$\mathbf{P}(\text{Popcorn} \cap \text{Movie})$ $\qquad$ $\mathbf{P}(\text{Popcorn} \cap \text{Contest})$

$$\boxed{\mathbf{P}(A \mid B)\mathbf{P}(B) = \mathbf{P}(A \cap B)}$$

There's Popcorn on the Floor. What Happened?



$$\mathbf{P}(\text{Popcorn}|\text{Movies})\ \mathbf{P}(\text{Movies}) > \mathbf{P}(\text{Popcorn}|\text{Contest})\ \mathbf{P}(\text{Contest})$$

$$\boxed{\mathbf{P}(A\,|\,B)\mathbf{P}(B) = \mathbf{P}(A \cap B)}$$

$$\boxed{\mathbf{P}(\text{Popcorn} \cap \text{Movie}) > \mathbf{P}(\text{Popcorn} \cap \text{Contest})}$$

What Does This Have To Do With Regularization?

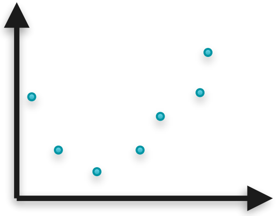
DeepLearning.AI

Point Estimation

Bayes Theorem and Regularization

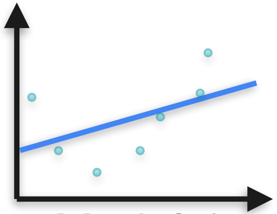
Example: Polynomial Regression

Example: Polynomial Regression

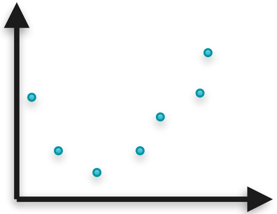


Data

Example: Polynomial Regression

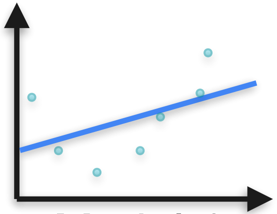


Model 1

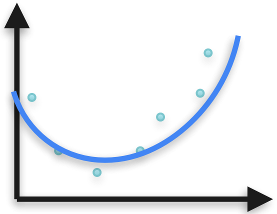


Data

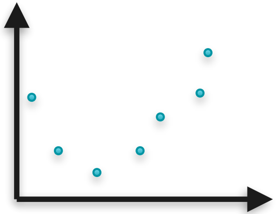
Example: Polynomial Regression



Model 1



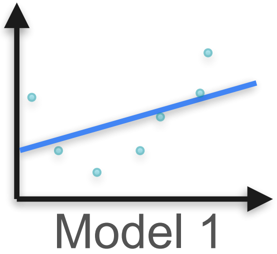
Model 2

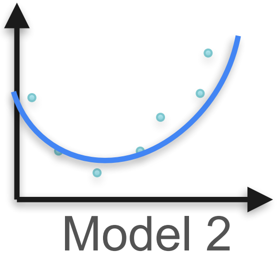


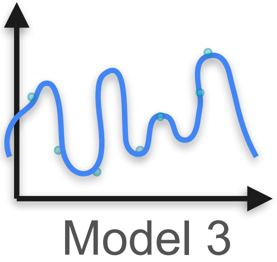
Data

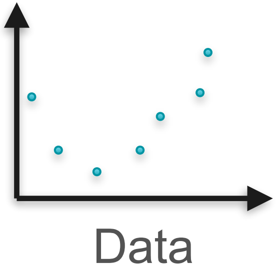
DeepLearning.AI

Example: Polynomial Regression

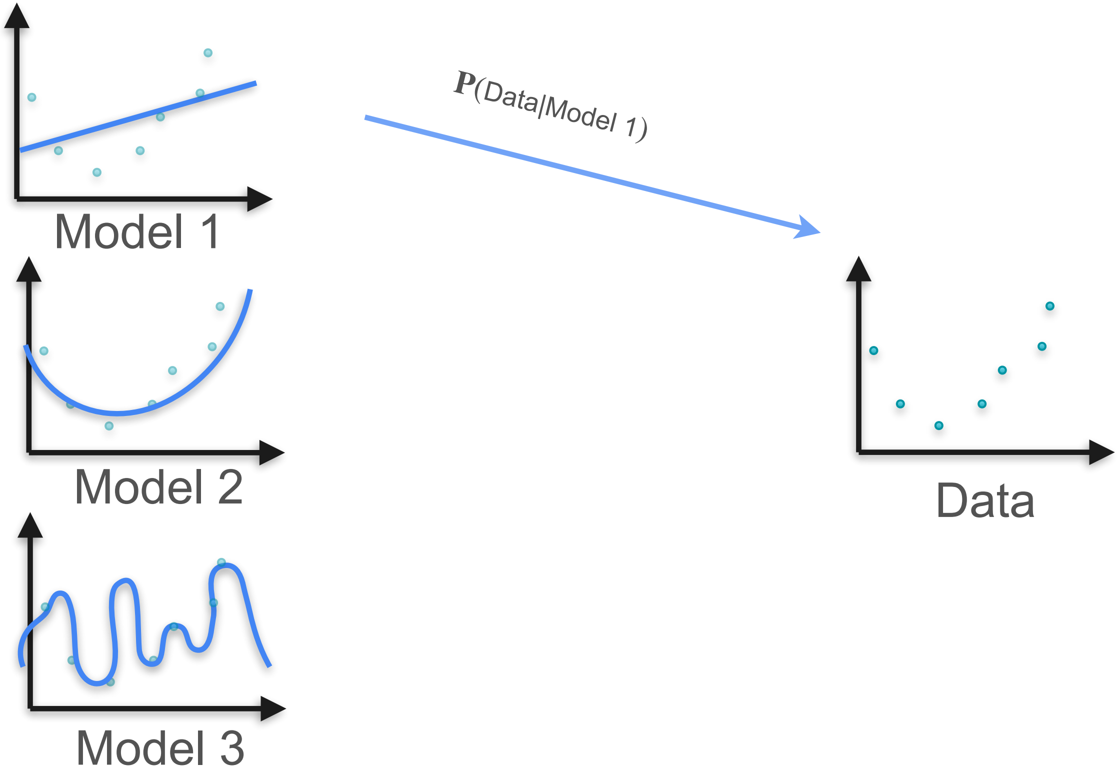




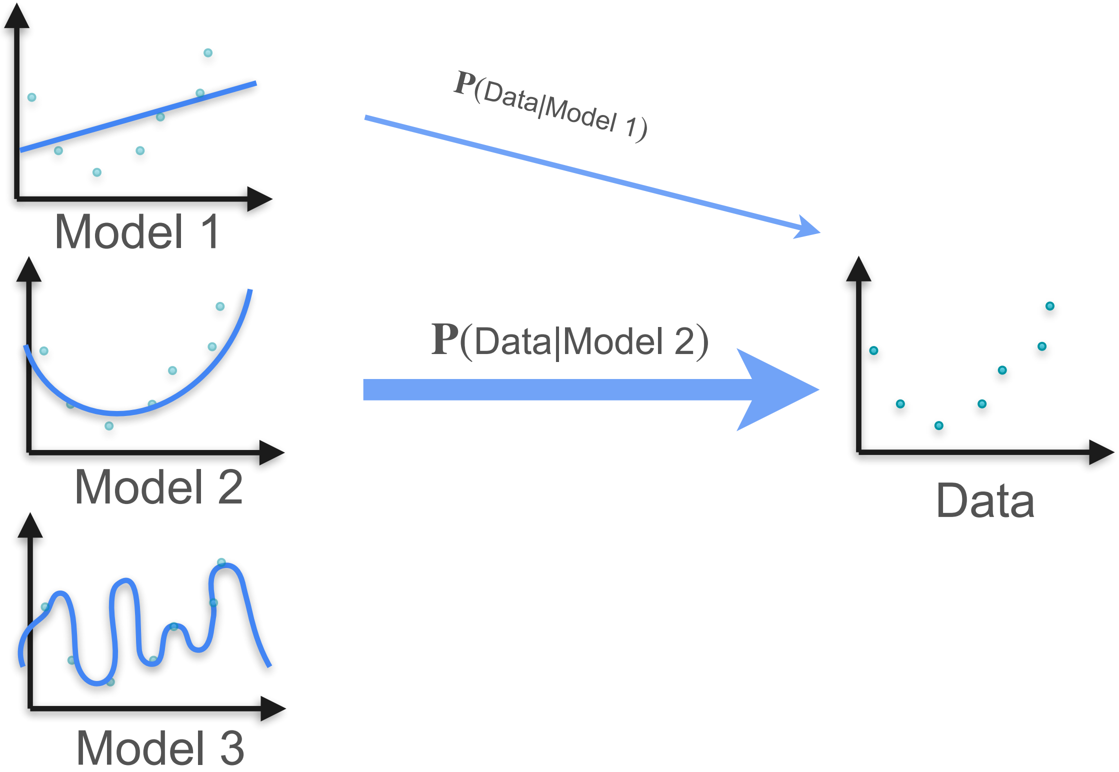




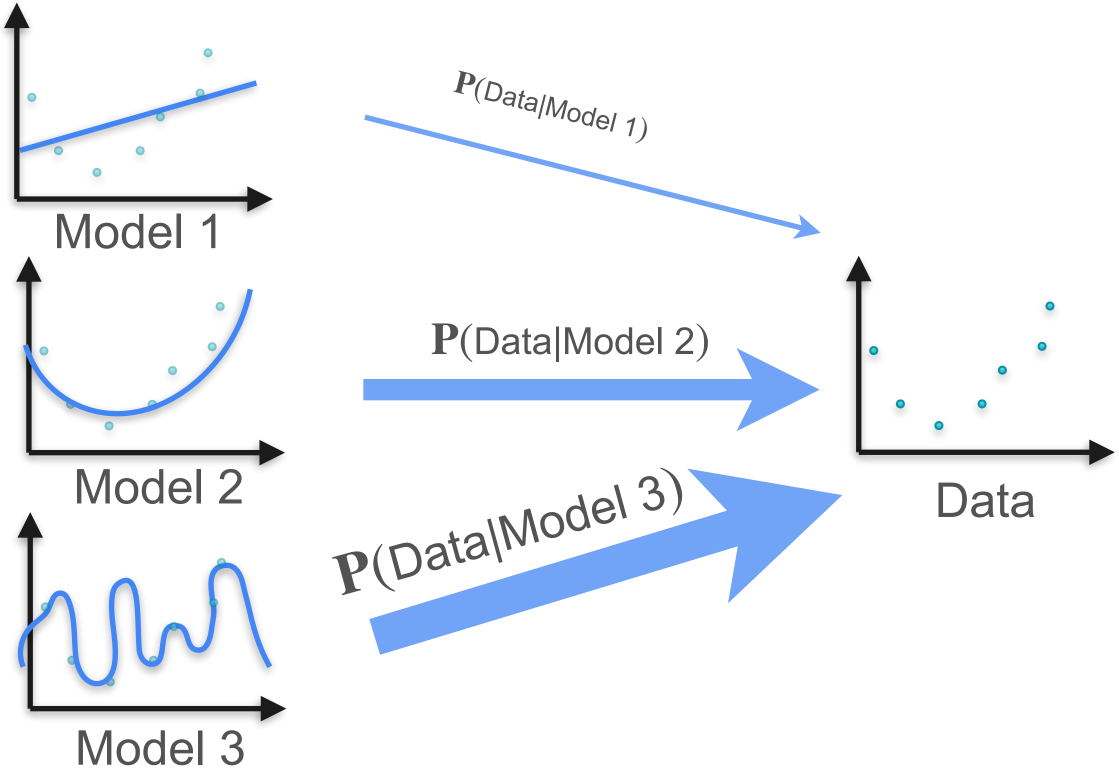
Example: Polynomial Regression



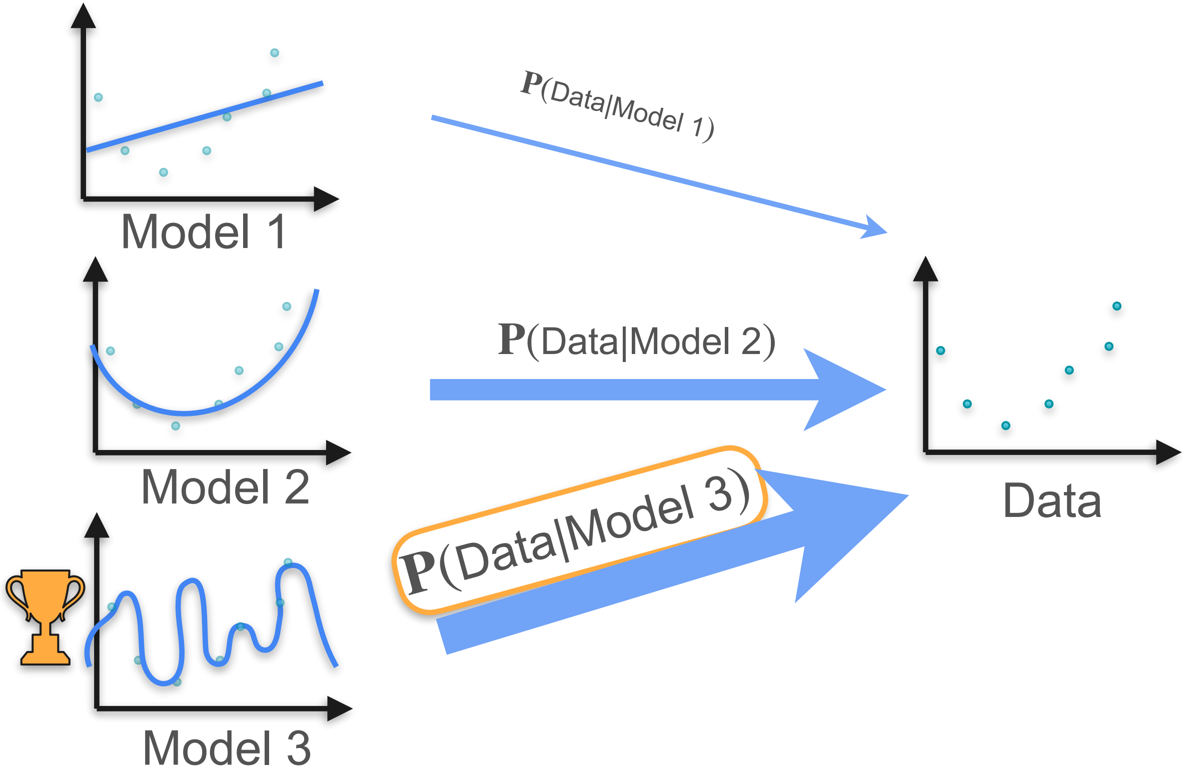
Example: Polynomial Regression



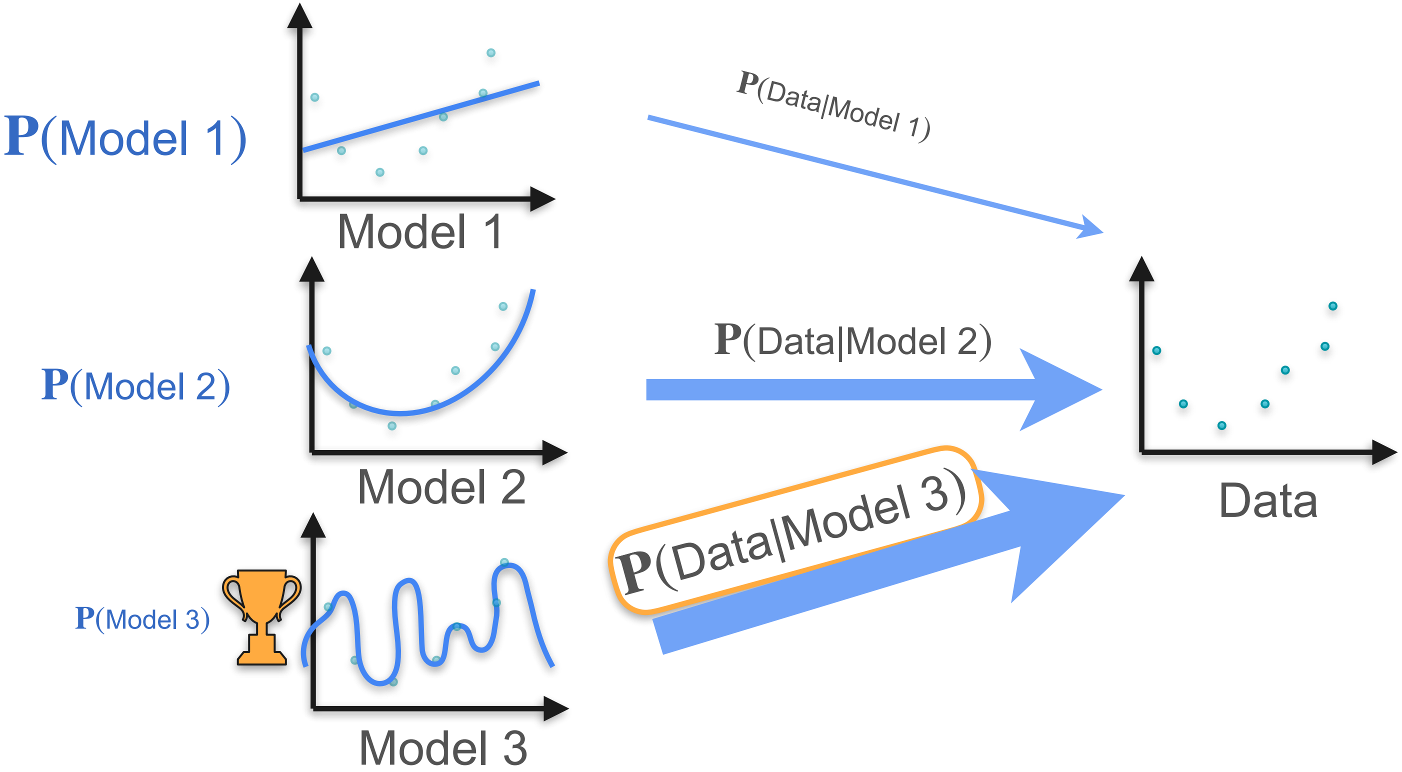
Example: Polynomial Regression



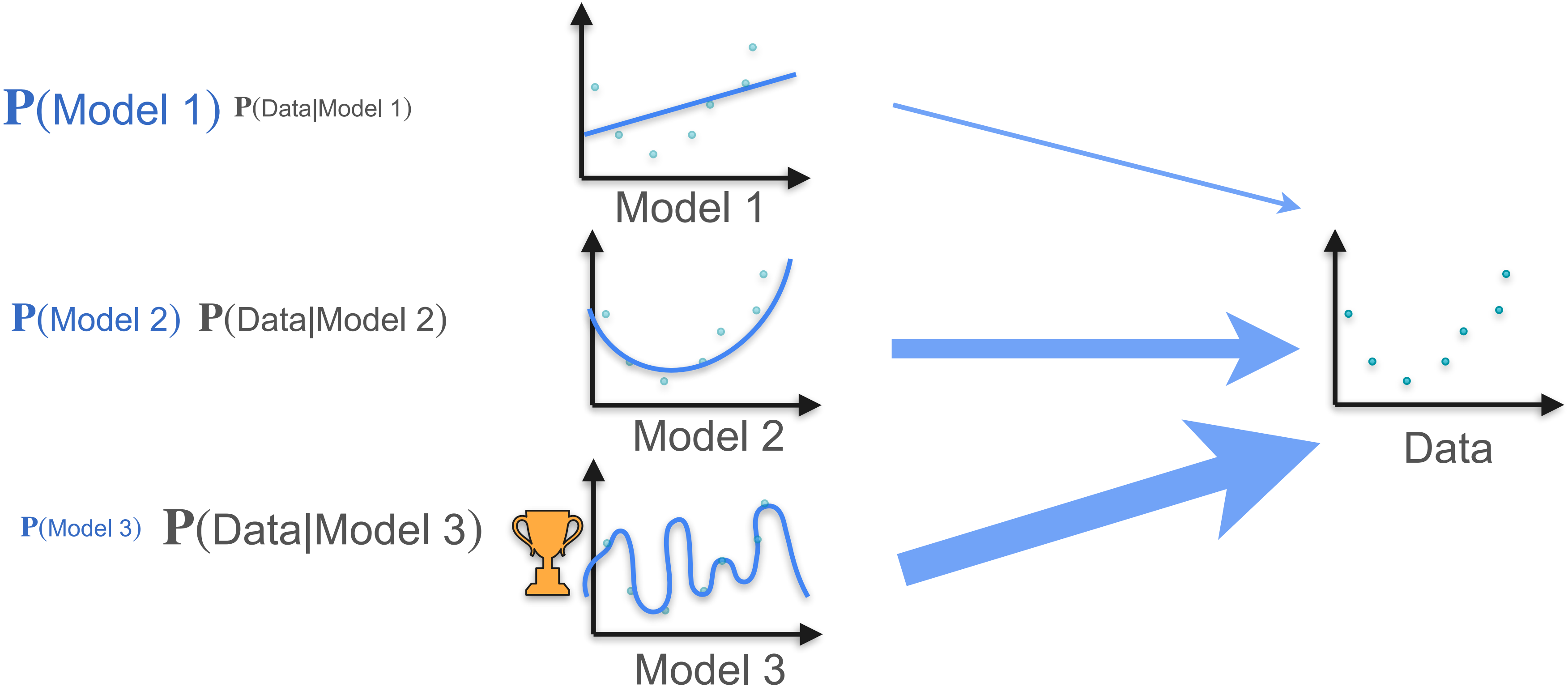
Example: Polynomial Regression



Example: Polynomial Regression



Example: Polynomial Regression

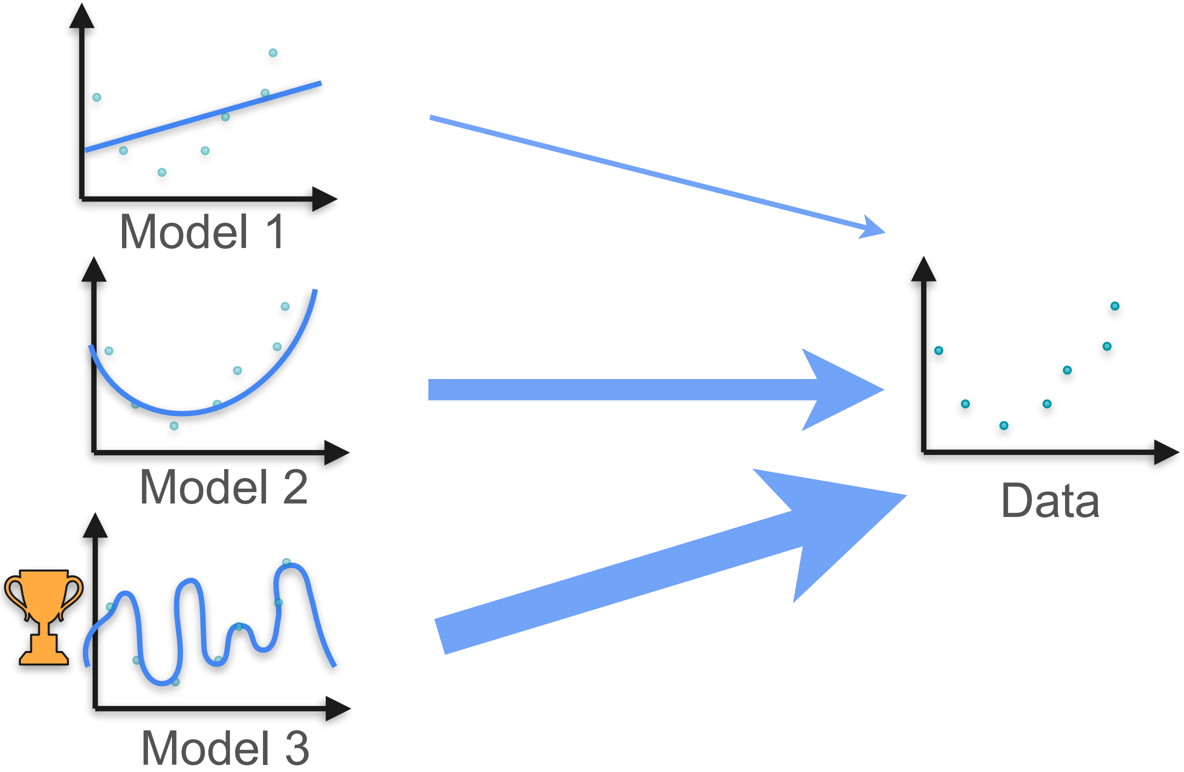


Example: Polynomial Regression

$\mathbf{P}$(Model 1) $\mathbf{P}$(Data|Model 1)

$\mathbf{P}$(Model 2) $\mathbf{P}$(Data|Model 2)

$\mathbf{P}$(Model 3) $\mathbf{P}$(Data|Model 3)

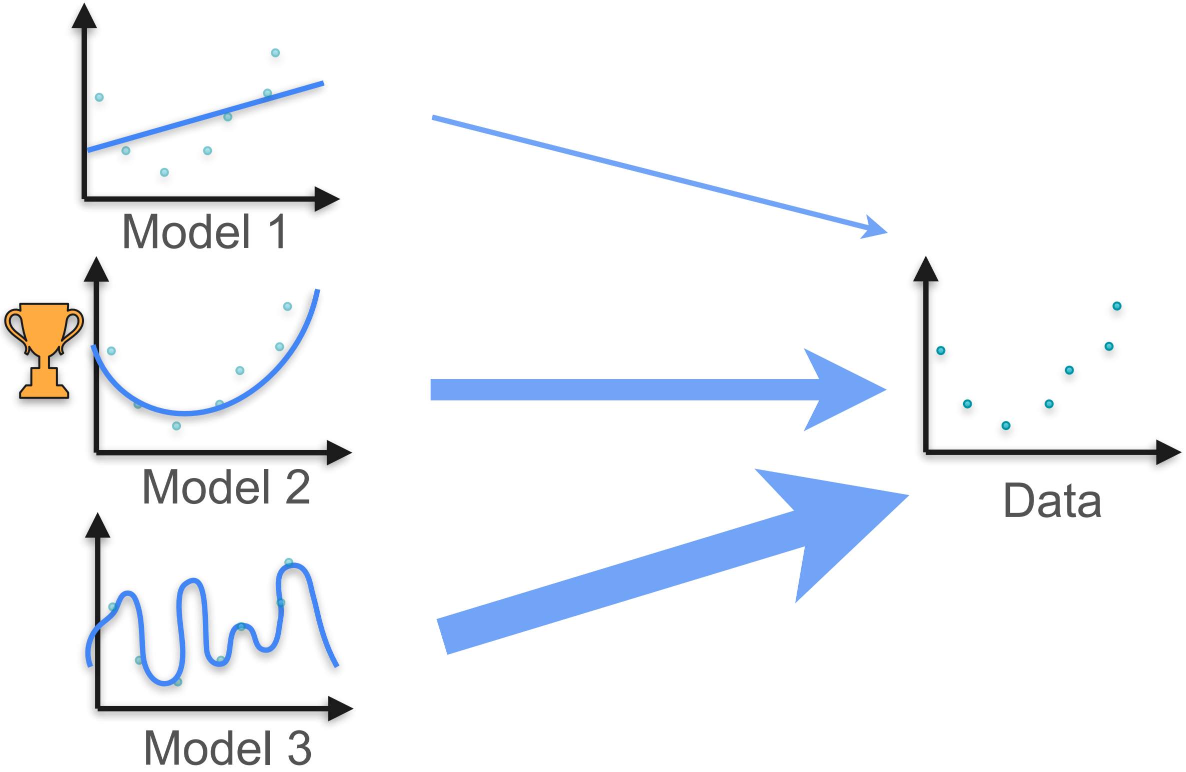


Example: Polynomial Regression

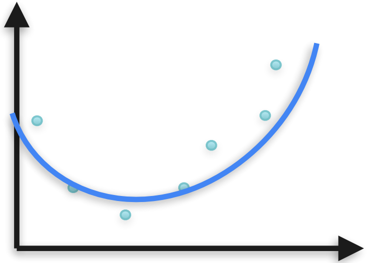
P(Model 1) P(Data|Model 1)

P(Model 2) P(Data|Model 2)

P(Model 3) P(Data|Model 3)



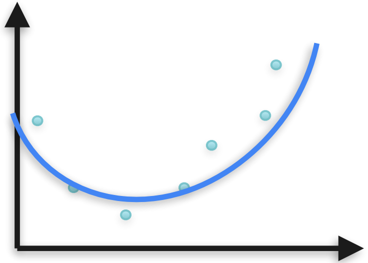
Maximum likelihood



Model

Polynomial regression

Maximum likelihood

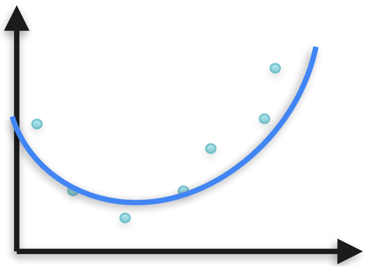


Model

Polynomial regression

Log-loss

Maximum likelihood



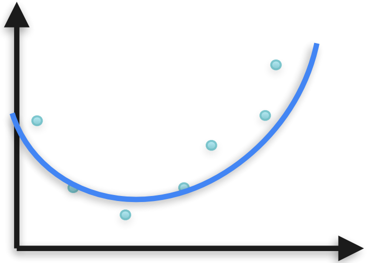
Model

Polynomial regression

$\mathbf{P}$(Data|Model)

Log-loss

Maximum likelihood with Bayes

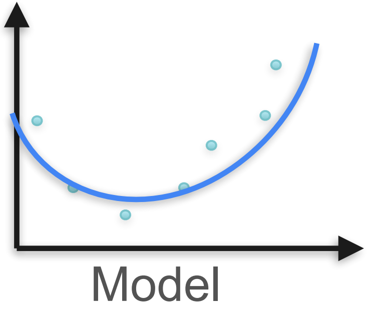


Polynomial regression

$\mathbf{P}(\text{Data}|\text{Model})$

Log-loss

Maximum likelihood with Bayes



Polynomial regression

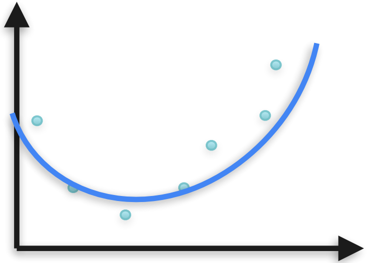
$\mathbf{P}(\text{Data}|\text{Model})$

$\cdot$

$\mathbf{P}(\text{Model})$

Log-loss

Maximum likelihood with Bayes



Polynomial regression with regularization

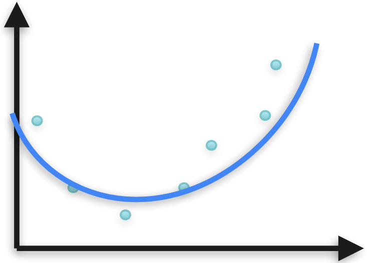
$\mathbf{P}$(Data|Model)

•

$\mathbf{P}$(Model)

Log-loss

Maximum likelihood with Bayes



Polynomial regression with regularization

$\mathbf{P}(\text{Data}|\text{Model})$

$\cdot$

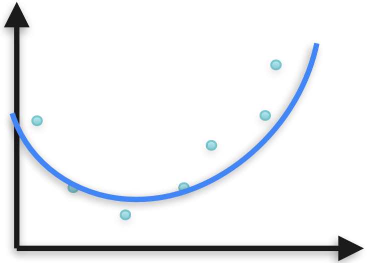
$\mathbf{P}(\text{Model})$

Log-loss

$+$

Regularization term

Maximum likelihood with Bayes



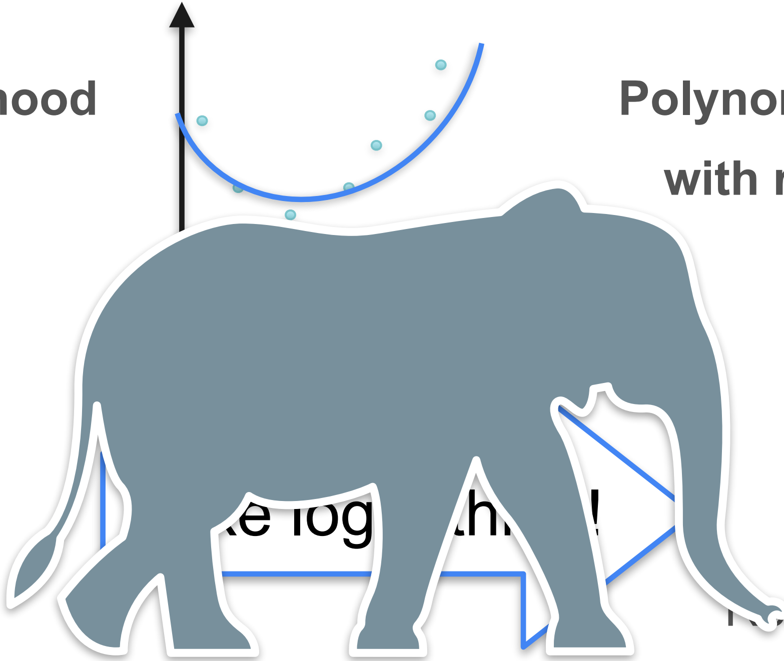
Polynomial regression with regularization

$\mathbf{P}$(Data|Model) $\cdot$ $\mathbf{P}$(Model)

Take logarithms!

Log-loss + Regularization term

Maximum likelihood with Bayes



Polynomial regression with regularization

$\mathbf{P}(\text{Data}|\text{Model})$

$\cdot$

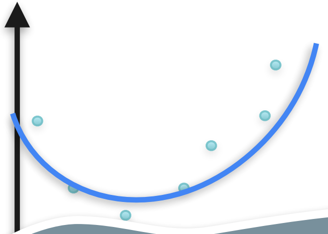
$\mathbf{P}(\text{Model})$

Log-loss

$+$

Regularization term

Maximum likelihood with Bayes



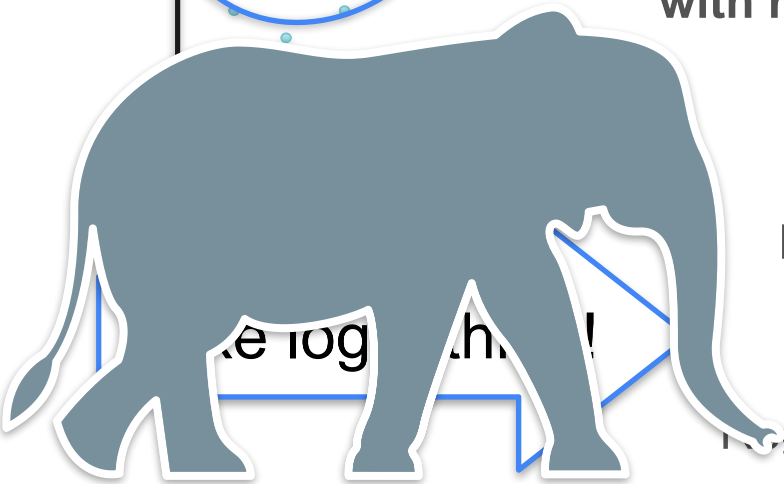
Polynomial regression with regularization

$\mathbf{P}$(Data|Model)

•

$\mathbf{P}$(Model)

?



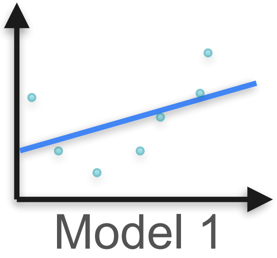
Log-loss

+

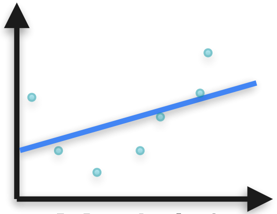
gularization term

What Is the Probability of a Model?

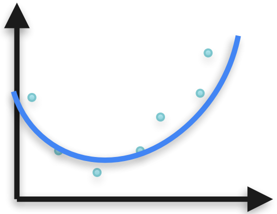
What Is the Probability of a Model?



What Is the Probability of a Model?

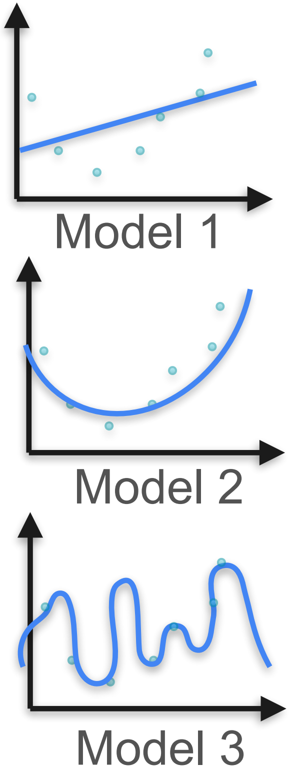


Model 1



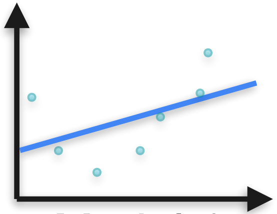
Model 2

What Is the Probability of a Model?

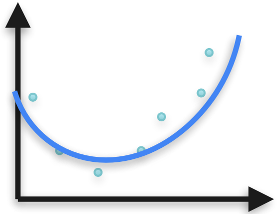


What Is the Probability of a Model?

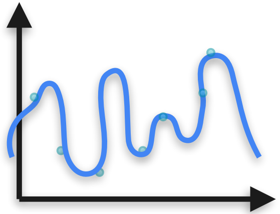
$\mathbf{P}(\text{Model 1})$


Model 1

$\mathbf{P}(\text{Model 2})$


Model 2

$\mathbf{P}(\text{Model 3})$


Model 3

What Is the Probability of a Model?

$\mathbf{P}(\text{Model 1})$

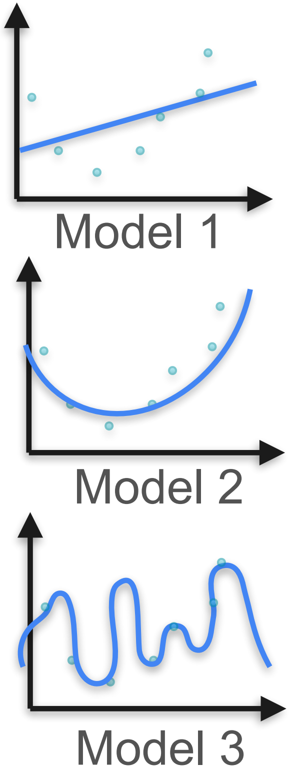
$$a_1 x + b$$

$\mathbf{P}(\text{Model 2})$

$$a_1 x^2 + a_2 x + b$$

$\mathbf{P}(\text{Model 3})$

$$a_1 x^{10} + a_2 x^9 + \cdots + a_{10} x + b$$



What Is the Probability of a Model?

$\mathbf{P}(\text{Model 1})$

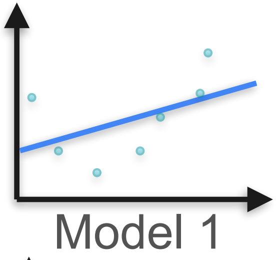
$\boxed{a_1}x + b$

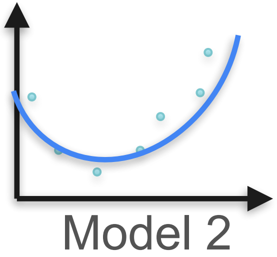
$\mathbf{P}(\text{Model 2})$

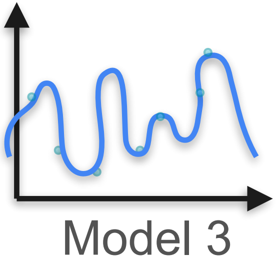
$a_1x^2 + a_2x + b$

$\mathbf{P}(\text{Model 3})$

$a_1x^{10} + a_2x^9 + \cdots + a_{10}x + b$







What Is the Probability of a Model?

$\mathbf{P}(\text{Model 1})$

$\boxed{a_1}x + b$

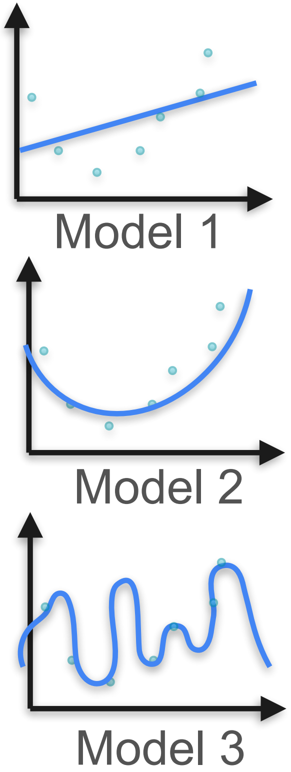
$a_i \sim \mathcal{N}(0,1)$

$\mathbf{P}(\text{Model 2})$

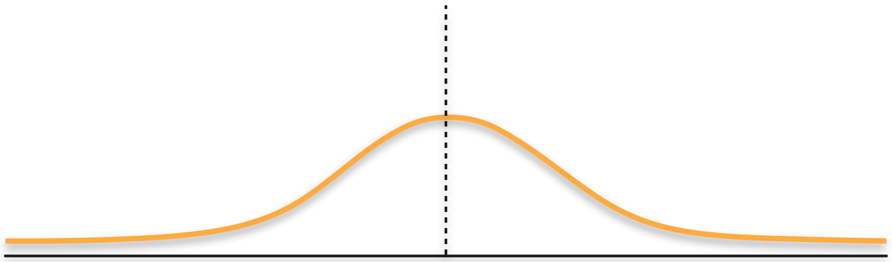
$a_1x^2 + a_2x + b$

$\mathbf{P}(\text{Model 3})$

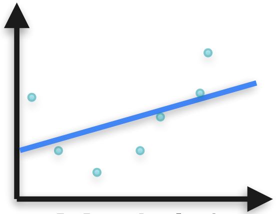
$a_1x^{10} + a_2x^9 + \cdots + a_{10}x + b$



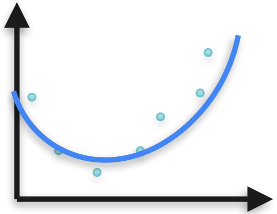
What Is the Probability of a Model?



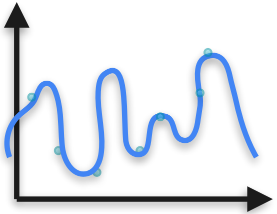
$$a_i \sim \mathcal{N}(0,1)$$



Model 1



Model 2



Model 3

$\mathbf{P}(\text{Model 1})$

$$\boxed{a_1}x + b$$

$\mathbf{P}(\text{Model 2})$

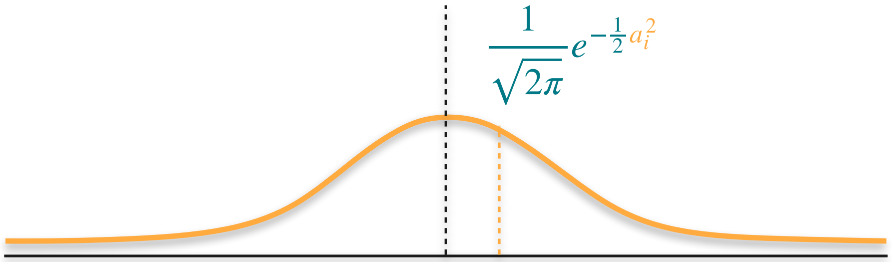
$$a_1x^2 + a_2x + b$$

$\mathbf{P}(\text{Model 3})$

$$a_1x^{10} + a_2x^9 + \cdots + a_{10}x + b$$

What Is the Probability of a Model?

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_i^2}$$



$$a_i \sim \mathcal{N}(0,1)$$

P(Model 1)

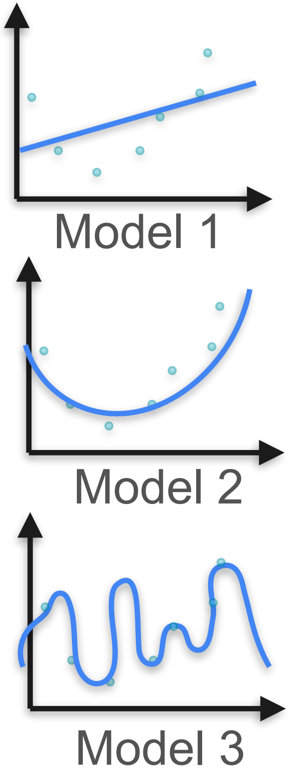
$$\boxed{a_1}x + b$$

P(Model 2)

$$a_1x^2 + a_2x + b$$

P(Model 3)

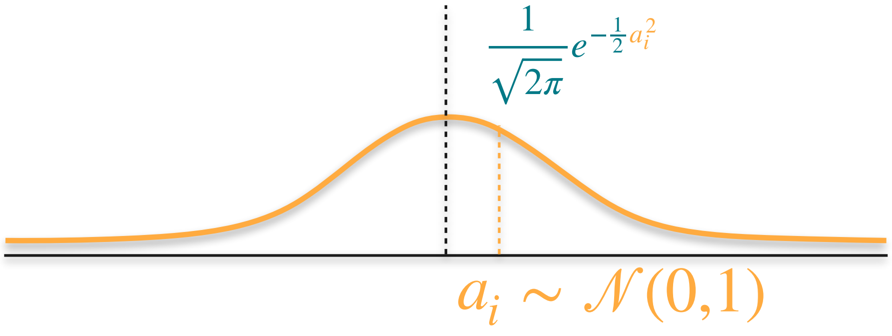
$$a_1x^{10} + a_2x^9 + \cdots + a_{10}x + b$$

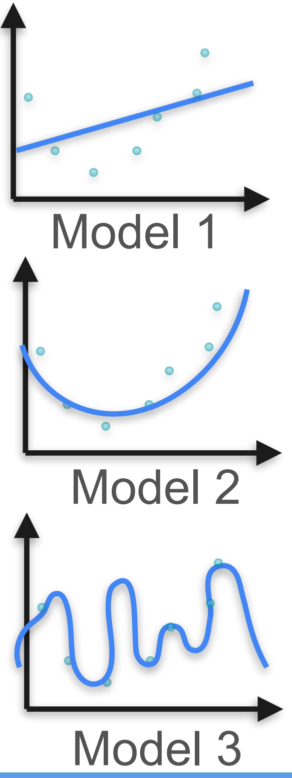


What Is the Probability of a Model?

$$\mathbf{P}(\text{Model 1}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_1^2}$$

$$\boxed{a_1}x + b$$





$\mathbf{P}(\text{Model 2})$

$$a_1x^2 + a_2x + b$$

$\mathbf{P}(\text{Model 3})$

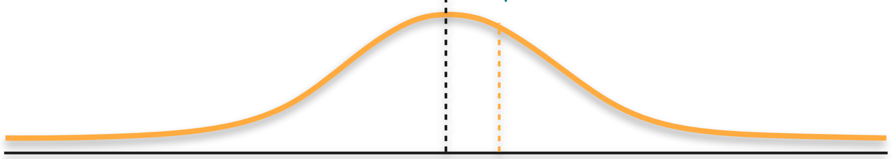
$$a_1x^{10} + a_2x^9 + \cdots + a_{10}x + b$$

What Is the Probability of a Model?

$$\mathbf{P}(\text{Model 1}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_1^2}$$

$$\boxed{a_1}x + b$$

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_i^2}$$



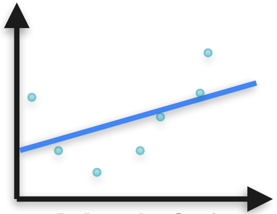
$$a_i \sim \mathcal{N}(0,1)$$

$$\mathbf{P}(\text{Model 2})$$

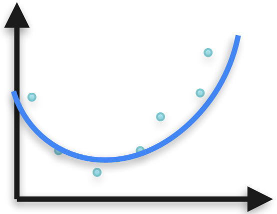
$$\boxed{a_1}x^2 + a_2x + b$$

$$\mathbf{P}(\text{Model 3})$$

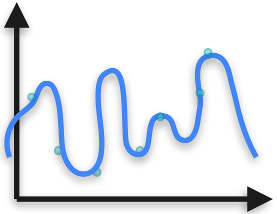
$$a_1x^{10} + a_2x^9 + \cdots + a_{10}x + b$$



Model 1



Model 2



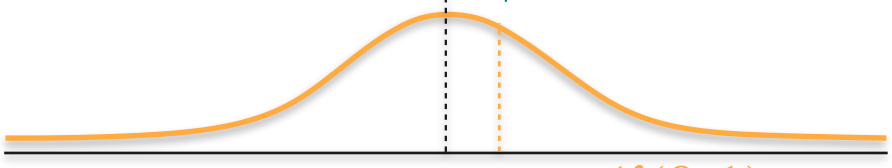
Model 3

What Is the Probability of a Model?

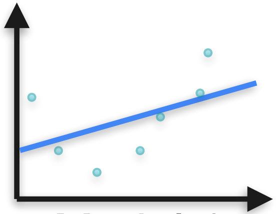
$$\mathbf{P}(\text{Model 1}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_1^2}$$

$$\boxed{a_1}x + b$$

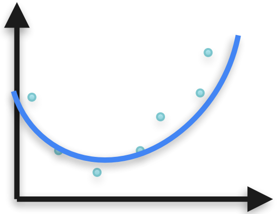
$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_i^2}$$



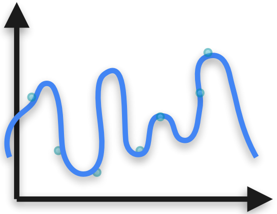
$$a_i \sim \mathcal{N}(0,1)$$



Model 1



Model 2



Model 3

$\mathbf{P}(\text{Model 2})$

$$\boxed{a_1}x^2 + \boxed{a_2}x + b$$

$\mathbf{P}(\text{Model 3})$

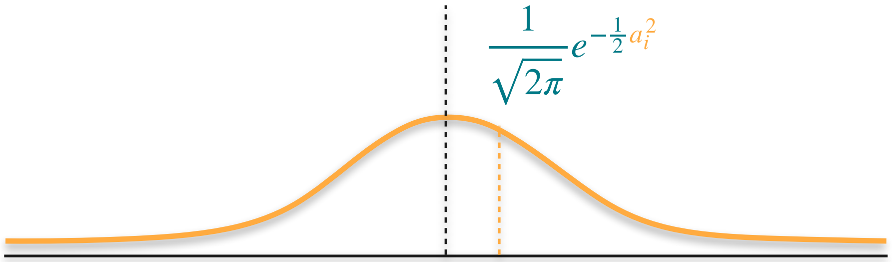
$$a_1x^{10} + a_2x^9 + \cdots + a_{10}x + b$$

What Is the Probability of a Model?

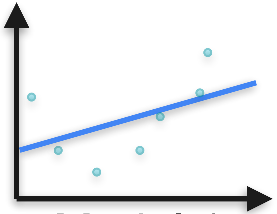
$$\mathbf{P}(\text{Model 1}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_1^2}$$

$$\boxed{a_1}x + b$$

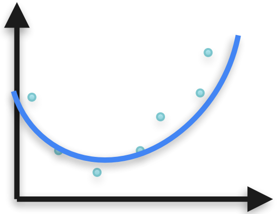
$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_i^2}$$



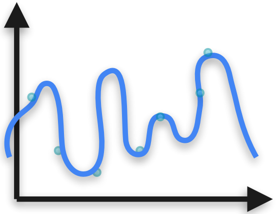
$$a_i \sim \mathcal{N}(0,1)$$



Model 1



Model 2



Model 3

$\mathbf{P}(\text{Model 2})$

$$\boxed{a_1}x^2 + \boxed{a_2}x + b = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_1^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_2^2}$$

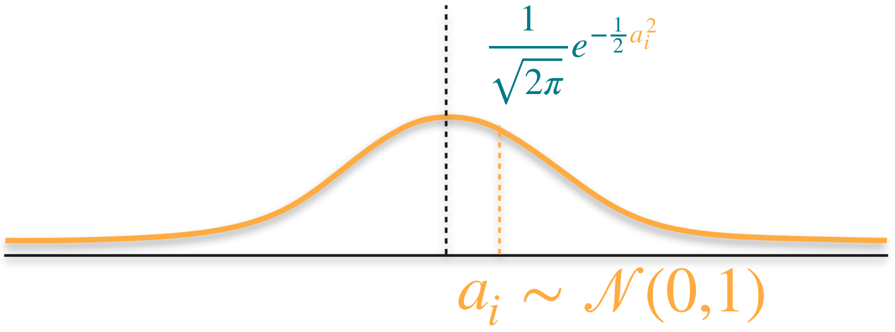
$\mathbf{P}(\text{Model 3})$

$$a_1 x^{10} + a_2 x^9 + \cdots + a_{10} x + b$$

What Is the Probability of a Model?

$$\mathbf{P}(\text{Model 1}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_1^2}$$

$$\boxed{a_1} x + b$$

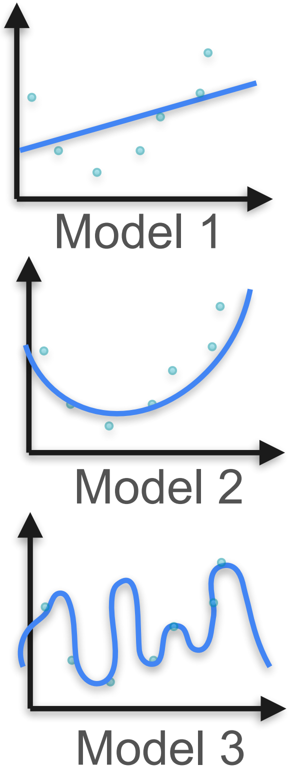


$\mathbf{P}(\text{Model 2})$

$$\boxed{a_1} x^2 + \boxed{a_2} x + b = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_1^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_2^2}$$

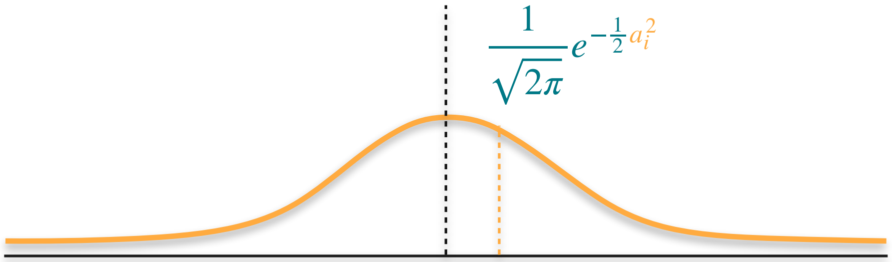
$\mathbf{P}(\text{Model 3})$

$$\boxed{a_1} x^{10} + \boxed{a_2} x^9 + \cdots + \boxed{a_{10}} x + b$$

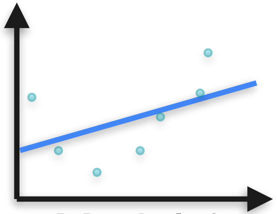


What Is the Probability of a Model?

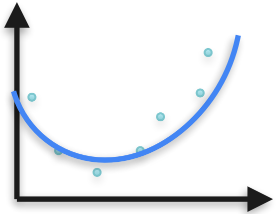
$$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}a_i^2}$$



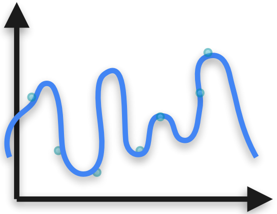
$$a_i \sim \mathcal{N}(0,1)$$



Model 1



Model 2



Model 3

$$\mathbf{P}(\text{Model 1}) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}a_1^2}$$

$$\boxed{a_1}x + b$$

$\mathbf{P}$(Model 2)

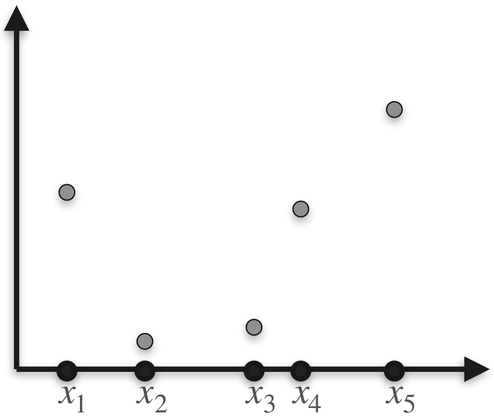
$$\boxed{a_1}x^2 + \boxed{a_2}x + b = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}a_1^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}a_2^2}$$

$\mathbf{P}$(Model 3)

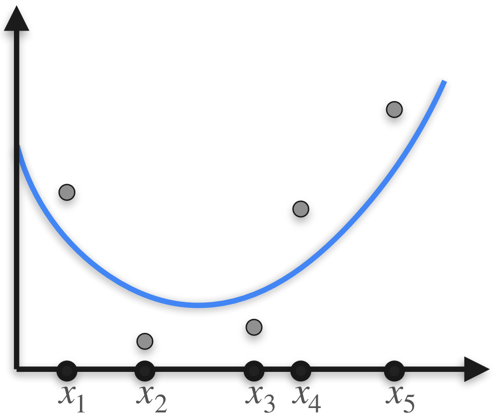
$$\boxed{a_1}x^{10} + \boxed{a_2}x^9 + \cdots + \boxed{a_{10}}x + b = \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}a_i^2}$$

Bayes and Regularization

Bayes and Regularization

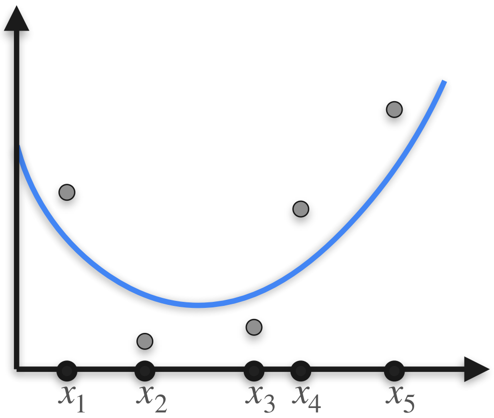


Bayes and Regularization



Bayes and Regularization

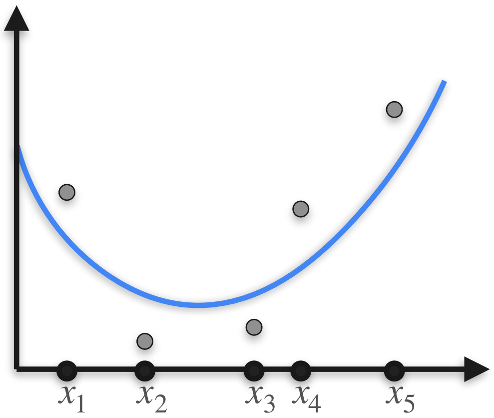
$\mathbf{P}(\text{Data}|\text{Model})$



Bayes and Regularization

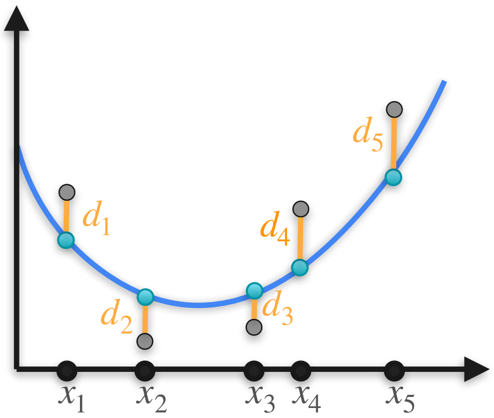
$\mathbf{P}$(Data|Model)

$\mathbf{P}$(Model)



Bayes and Regularization

$\mathbf{P}(\text{Data}|\text{Model})$ $\mathbf{P}(\text{Model})$

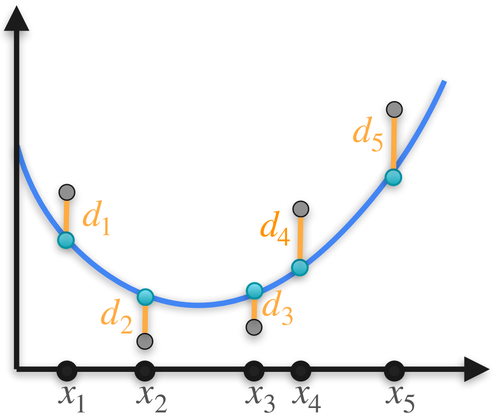


Bayes and Regularization

P(Data|Model)

P(Model)

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_1^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_2^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_3^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_4^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_5^2}$$

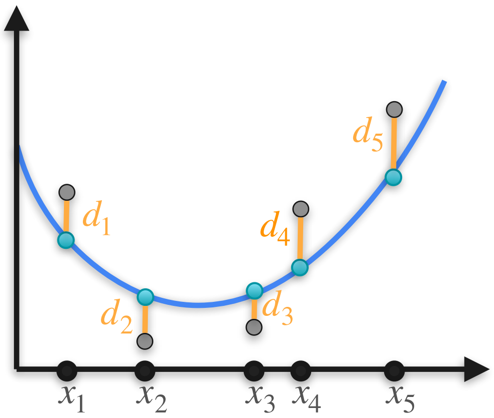


Bayes and Regularization

$\mathbf{P}(\text{Data}|\text{Model})$

$\mathbf{P}(\text{Model})$

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_1^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_2^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_3^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_4^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_5^2}$$



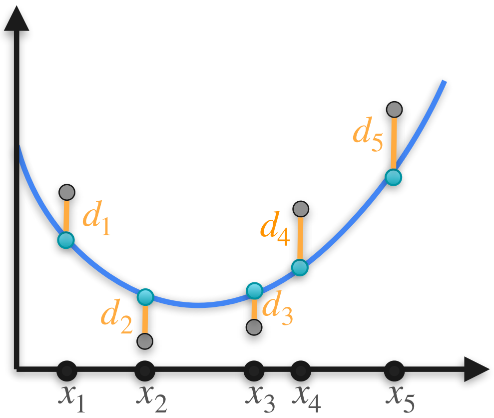
$$a_1x^2 + a_2x + b$$

Bayes and Regularization

$\mathbf{P}$(Data|Model)

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_1^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_2^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_3^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_4^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_5^2}$$

$\mathbf{P}$(Model)

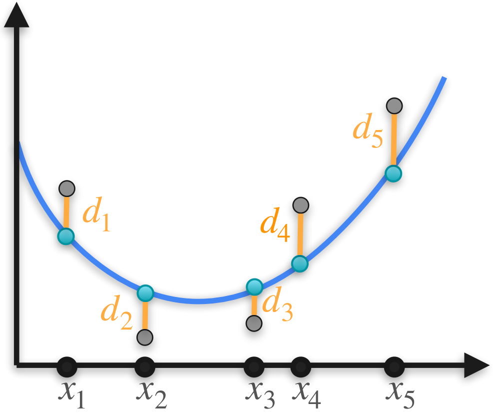


$$\boxed{a_1}x^2 + \boxed{a_2}x + b$$

Bayes and Regularization

P(Data|Model) **P**(Model)

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_1^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_2^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_3^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_4^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_5^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}a_1^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}a_2^2}$$

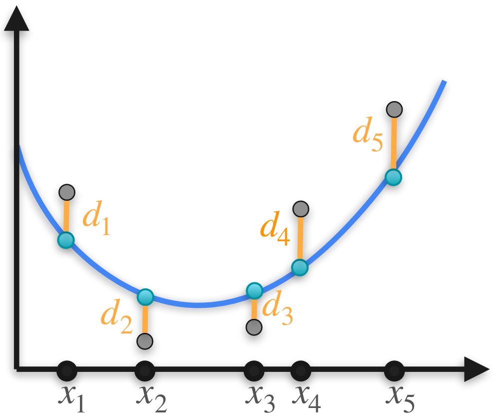


$$\boxed{a_1}x^2 + \boxed{a_2}x + b$$

Bayes and Regularization

P(Data|Model) **P**(Model)

$$\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}a_1^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}a_2^2}$$

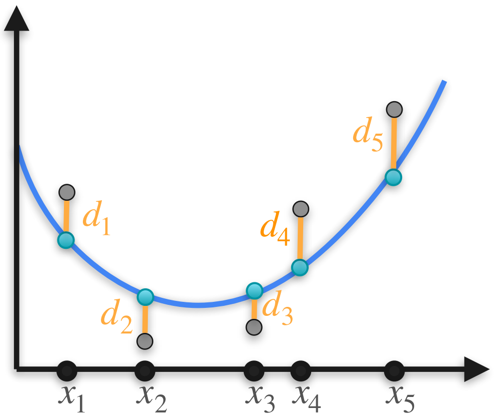


$$\boxed{a_1}x^2+\boxed{a_2}x+b$$

Bayes and Regularization

P(Data|Model) **P**(Model)

$$\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}a_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}a_2^2}$$



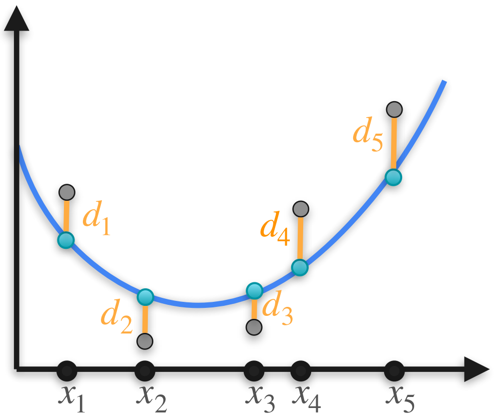
$$\boxed{a_1}x^2 + \boxed{a_2}x + b$$

Bayes and Regularization

$\mathbf{P}$(Data|Model) $\mathbf{P}$(Model)

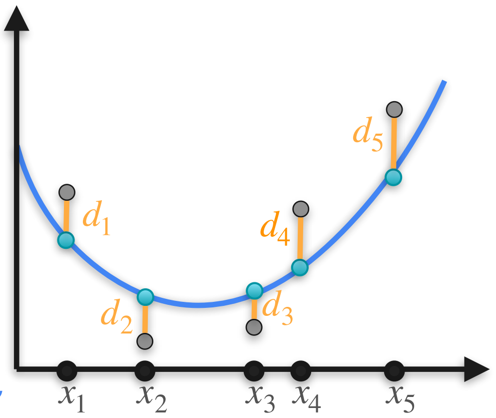
$$\underbrace{\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2}}_{\boxed{\log}} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}a_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}a_2^2}$$

$$-\frac{1}{2}(d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2)$$



$$\boxed{a_1}x^2 + \boxed{a_2}x + b$$

Bayes and Regularization



$$\boxed{a_1}x^2 + \boxed{a_2}x + b$$

P(Data|Model)

$$\underbrace{\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2}}_{\log}$$

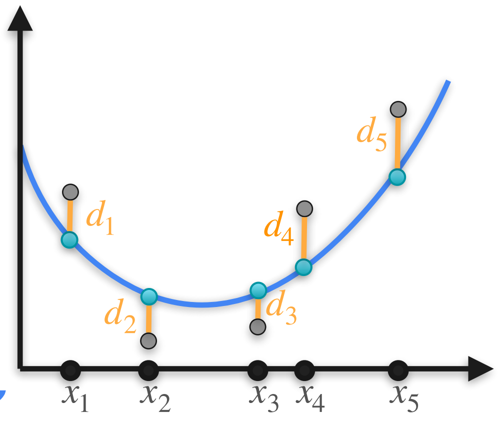
$$-\frac{1}{2}(d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2)$$

P(Model)

$$\cdot\ \underbrace{\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}a_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}a_2^2}}_{\log}$$

$$-\frac{1}{2}(a_1^2 + a_2^2)$$

Bayes and Regularization



$$\boxed{a_1}x^2 + \boxed{a_2}x + b$$

$\mathbf{P}$(Data|Model) $\qquad\qquad$ $\mathbf{P}$(Model)

$$\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}a_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}a_2^2}$$

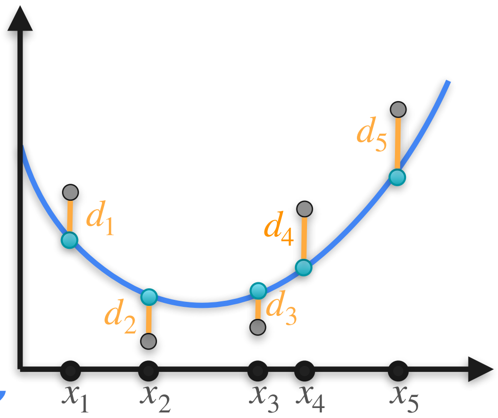
log $\qquad$ log $\qquad$ log

$$-\frac{1}{2}(d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2) \quad + \quad -\frac{1}{2}(a_1^2 + a_2^2)$$

Bayes and Regularization

$\mathbf{P}$(Data|Model) $\mathbf{P}$(Model)

$$\underbrace{\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2}}_{\boxed{\log}}\cdot\underbrace{\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}a_1^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}a_2^2}}_{\boxed{\log}}$$



$$\boxed{a_1}x^2+\boxed{a_2}x+b$$

Maximize

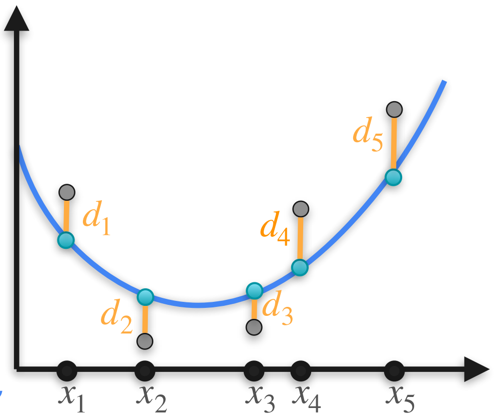
$\boxed{\log}$

$$-\frac{1}{2}(d_1^2+d_2^2+d_3^2+d_4^2+d_5^2) \quad + \quad -\frac{1}{2}(a_1^2+a_2^2)$$

Bayes and Regularization

$\mathbf{P}$(Data|Model)　　　　　　　　$\mathbf{P}$(Model)

$$\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}a_1^2} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}a_2^2}$$



$$\boxed{a_1}x^2 + \boxed{a_2}x + b$$

log　　　　log　　　　log

Maximize

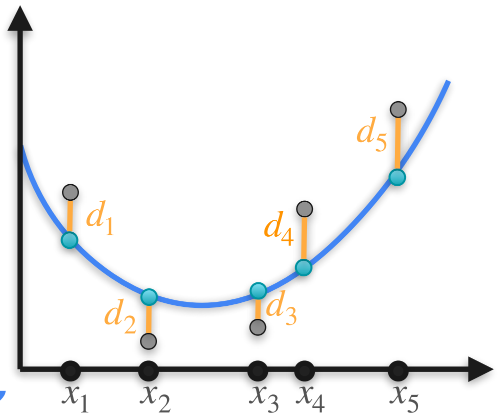
$$-\frac{1}{2}(d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2) \quad + \quad -\frac{1}{2}(a_1^2 + a_2^2)$$

Minimize

Bayes and Regularization

$\mathbf{P}$(Data|Model) $\qquad\qquad$ $\mathbf{P}$(Model)

$$\underbrace{\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_1^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_2^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_3^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_4^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}d_5^2}}_{\log}\ \cdot\ \underbrace{\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}a_1^2}\cdot\cancel{\frac{1}{\sqrt{2\pi}}}e^{-\frac{1}{2}a_2^2}}_{\log}$$



$$\boxed{a_1}x^2 + \boxed{a_2}x + b$$

log

Maximize

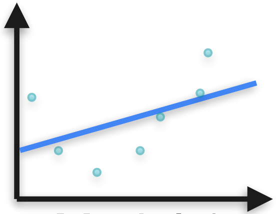
$$-\frac{1}{2}(d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2) \quad + \quad -\frac{1}{2}(a_1^2 + a_2^2)$$

Minimize

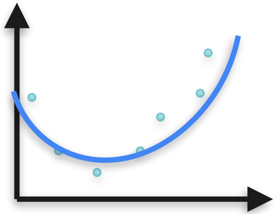
$$\boxed{d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 \quad + \quad a_1^2 + a_2^2}$$

Regularization

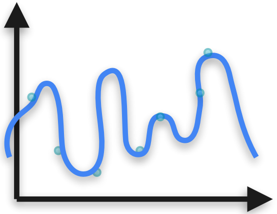
Regularization



Model 1



Model 2



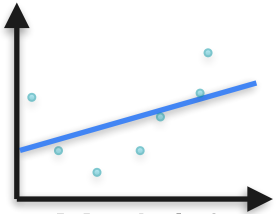
Model 3

Regularization

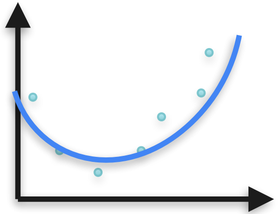
$\mathbf{P}(\text{Model 1})$

$\mathbf{P}(\text{Model 2})$

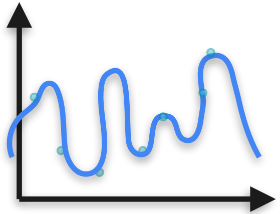
$\mathbf{P}(\text{Model 3})$



Model 1



Model 2



Model 3

Regularization

$\mathbf{P}(\text{Model 1})$

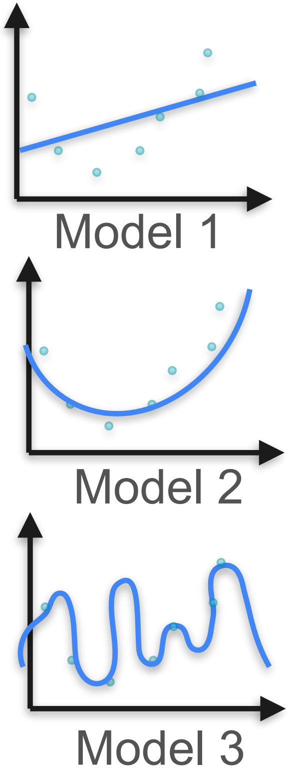
Minimize x_1^2

$\mathbf{P}(\text{Model 2})$

Minimize $x_1^2 + x_2^2$

$\mathbf{P}(\text{Model 3})$

Minimize $x_1^2 + \cdots + x_{10}^2$



Regularization

$\mathbf{P}$(Model 1)

Minimize x_1^2

$\mathbf{P}$(Data|Model 1)

$\mathbf{P}$(Model 2)

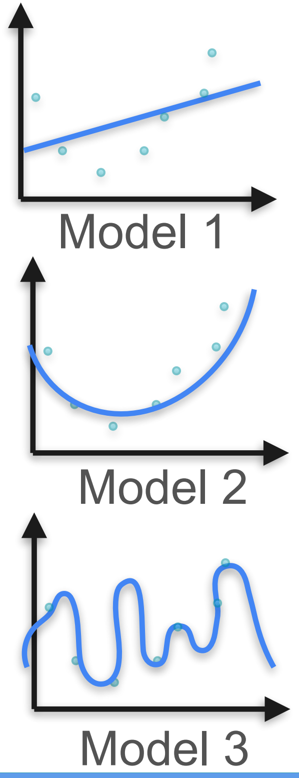
Minimize $x_1^2 + x_2^2$

$\mathbf{P}$(Data|Model 2)

$\mathbf{P}$(Model 3)

Minimize $x_1^2 + \cdots + x_{10}^2$

$\mathbf{P}$(Data|Model 3)



Regularization

$\mathbf{P}$(Model 1)

Minimize x_1^2

$\mathbf{P}$(Data|Model 1)

$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$

$\mathbf{P}$(Model 2)

Minimize $x_1^2 + x_2^2$

$\mathbf{P}$(Data|Model 2)

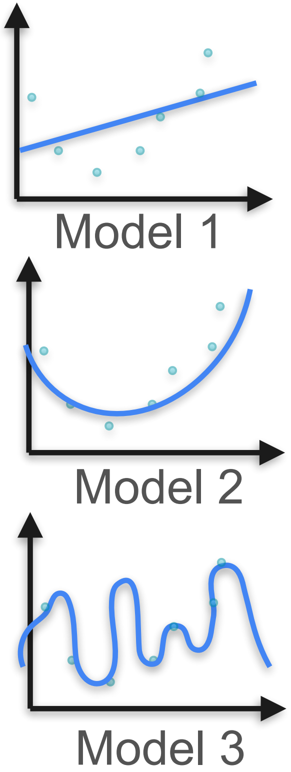
$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$

$\mathbf{P}$(Model 3)

Minimize $x_1^2 + \cdots + x_{10}^2$

$\mathbf{P}$(Data|Model 3)

$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$



Regularization

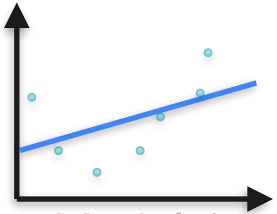
New Loss

$\mathbf{P}(\text{Model 1})$

Minimize x_1^2

$\mathbf{P}(\text{Data}|\text{Model 1})$

$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$



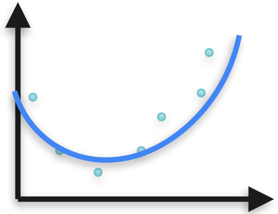
Model 1

$\mathbf{P}(\text{Model 2})$

Minimize $x_1^2 + x_2^2$

$\mathbf{P}(\text{Data}|\text{Model 2})$

$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$



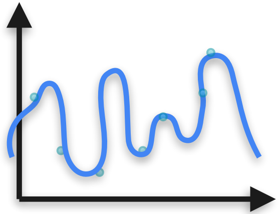
Model 2

$\mathbf{P}(\text{Model 3})$

Minimize $x_1^2 + \cdots + x_{10}^2$

$\mathbf{P}(\text{Data}|\text{Model 3})$

$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$



Model 3

Regularization

New Loss

$\mathbf{P}$(Model 1) $\qquad$ $\mathbf{P}$(Data|Model 1)

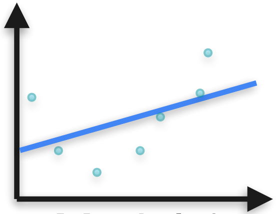
Minimize $x_1^2 + d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$

$\mathbf{P}$(Model 2) $\qquad$ $\mathbf{P}$(Data|Model 2)

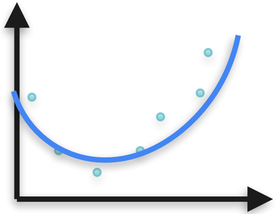
Minimize $x_1^2 + x_2^2 + d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$

$\mathbf{P}$(Model 3) $\qquad$ $\mathbf{P}$(Data|Model 3)

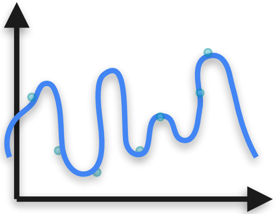
Minimize $x_1^2 + \cdots + x_{10}^2 + d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$



Model 1



Model 2



Model 3

DeepLearning.AI

Point Estimation

Conclusion