data science for (physical) scientists 2

II probability and statistics

dr.federica bianco | fbb.space | fedhere | fedhere

this slide deck

http://slides.com/federicabianco/dsps_02

Intro survey: some (many) questions are designed to be hard! I want to probe the limit of students' preparation to better plan the course

bit.ly/3PuwZfP

How to talk about AI

How to talk about AI

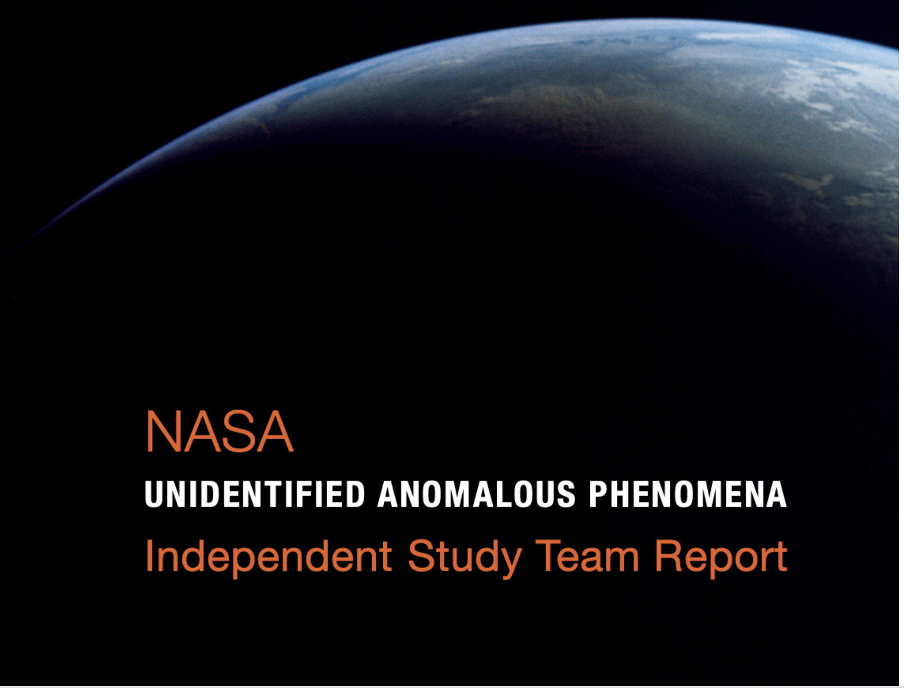
Oct 21, 2022

NASA Announces Unidentified Anomalous Phenomena Study Team Members

NASA has selected 16 individuals to participate in its independent study team on unidentified anomalous phenomena (UAP). *Observations of events in the sky that cannot be identified as aircraft or as known natural phenomena.*

The independent study team will lay the groundwork for future study on the nature of UAPs for NASA and other organizations [...] identify how data gathered by civilian government entities, commercial data, and data from other sources can potentially be analyzed to shed light on UAPs.

It will then recommend a roadmap for potential UAP data analysis by the agency going forward.



Members of the NASA Unidentified Anomalous Phenomena Independent Study Team

Chair
Dr. David Spergel
Simons Foundation

Designated Federal Official
Dr. Daniel Evans
NASA Headquarters

Panelists

Dr. Anamaria Berea
George Mason University

Dr. Federica Bianco
University of Delaware

Dr. Reggie Brothers
AE Industrial Partners

Dr. Paula Bontempi
University of Rhode Island

Dr. Jennifer Buss
Potomac Institute of Policy Studies

Dr. Nadia Drake
Science Journalist

Mr. Mike Gold
Redwire Space

Dr. David Grinspoon
Planetary Science Institute

Capt. Scott Kelly, USN, Ret.
NASA Astronaut, Ret.

Dr. Matt Mountain
Association of Universities for Research and Astronomy

Mr. Warren Randolph
Federal Aviation Administration

Dr. Walter Scott
Maxar Technologies

Dr. Joshua Semeter
Boston University

Dr. Karlin Toner
Federal Aviation Administration

Dr. Shelley Wright
University of California, San Diego

AP News

NASA says more science and less stigma are needed to understand UFOs

Reuters

NASA names chief of UFO research; panel sees no alien evidence

How to talk about

NASA Will Study UFOs Using AI and Crowdsourcing

NASA said it would begin using artificial intelligence and citizen reporting to study unidentified flying objects to develop a better...

Oct 21, 2022

NASA Announces Unidentified Anomalous Phenomena Study Team Members

NASA has selected 16 individuals to participate in its independent study team on unidentified anomalous phenomena (UAP). Observations of events in the sky that cannot be identified as aircraft or as known natural phenomena are categorized as UAPs.

The independent study team will lay the groundwork for future study on the nature of UAPs for NASA and other organizations. To do this, the team will identify how data gathered by civilian government entities, commercial data, and data from other sources can potentially be analyzed to shed light on UAPs.

It will then recommend a roadmap for potential UAP data analysis by the agency going forward.

How to talk about AI

NASA Press liaison

Nicky Fox

NASA Science Mission Directorate

Den Evans

NASA coordinator of the UAP study

David Spergle

chair of the stury group

Senator Nelson

NASA Administrator

https://www.youtube.com/embed/zjwrzHJA3PI?si=-G4UwLlYgU9ltJm4&enablejsapi=1

Reuters

NASA names chief of UFO research; panel sees no alien evidence

AP News

NASA says more science and less stigma are needed to understand UFOs https://wall.sli.do/event/qEviP9bWQyHKRZdTXuLbYJ?section=e8049dd4-c5c6-481c-83c3-23b8223708a7 https://wall.sli.do/event/ocgPy4JML3Fb1dmdPosHsR?section=c4fba1d4-ce09-4e39-b737-5f6546b443d1

0 RECAP

What is Data Science

The discipline that deals with extraction of information from data, including all phases of data driven inference from data collection through modeling and communication, and its interpretation in a domain context

What is Data Science

The field of studies that deals with extraction of information from data, including all phases of data driven inference from data collection through modeling and communication, within a domain context to enable interpretation and prediction of phenomena.

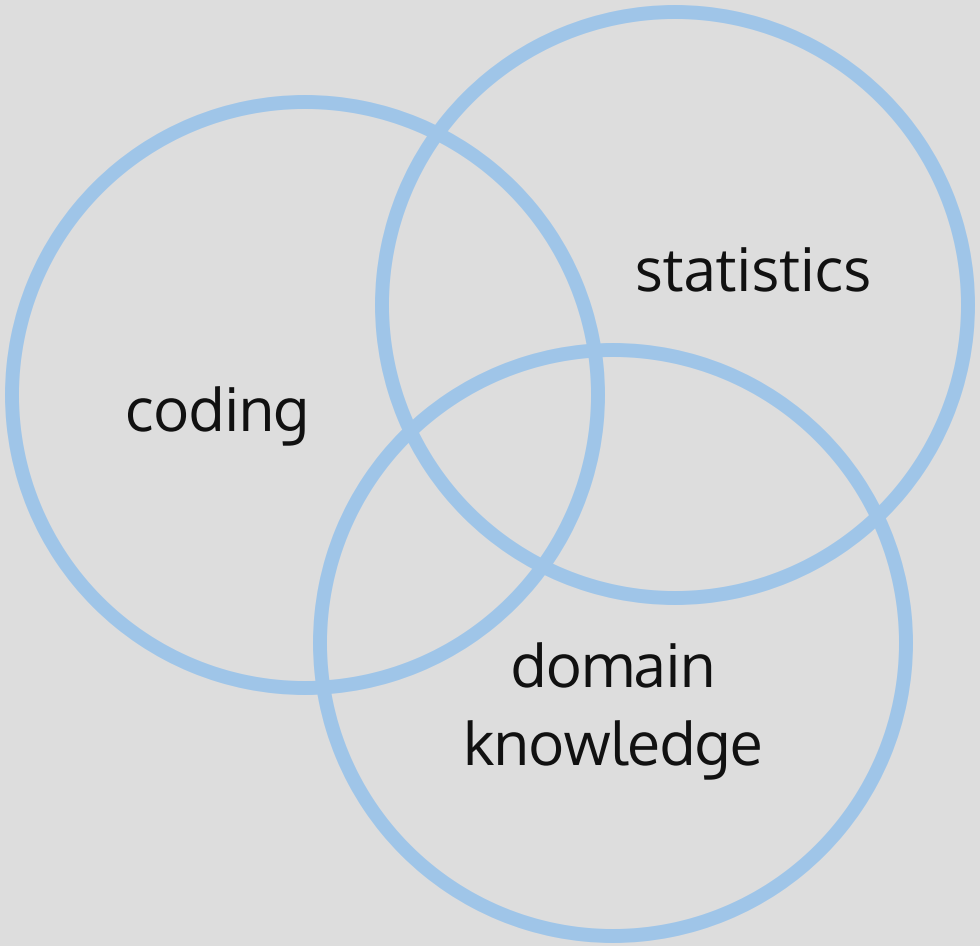
different from statistics and machine learning

different from computer science

foundations of data science for everyone

Data Science: the field of studies that deals with the extraction of information from data within a domain context to enable interpretation and prediction of phenomena.

This includes development and application of statistical tools and machine learning and AI methods

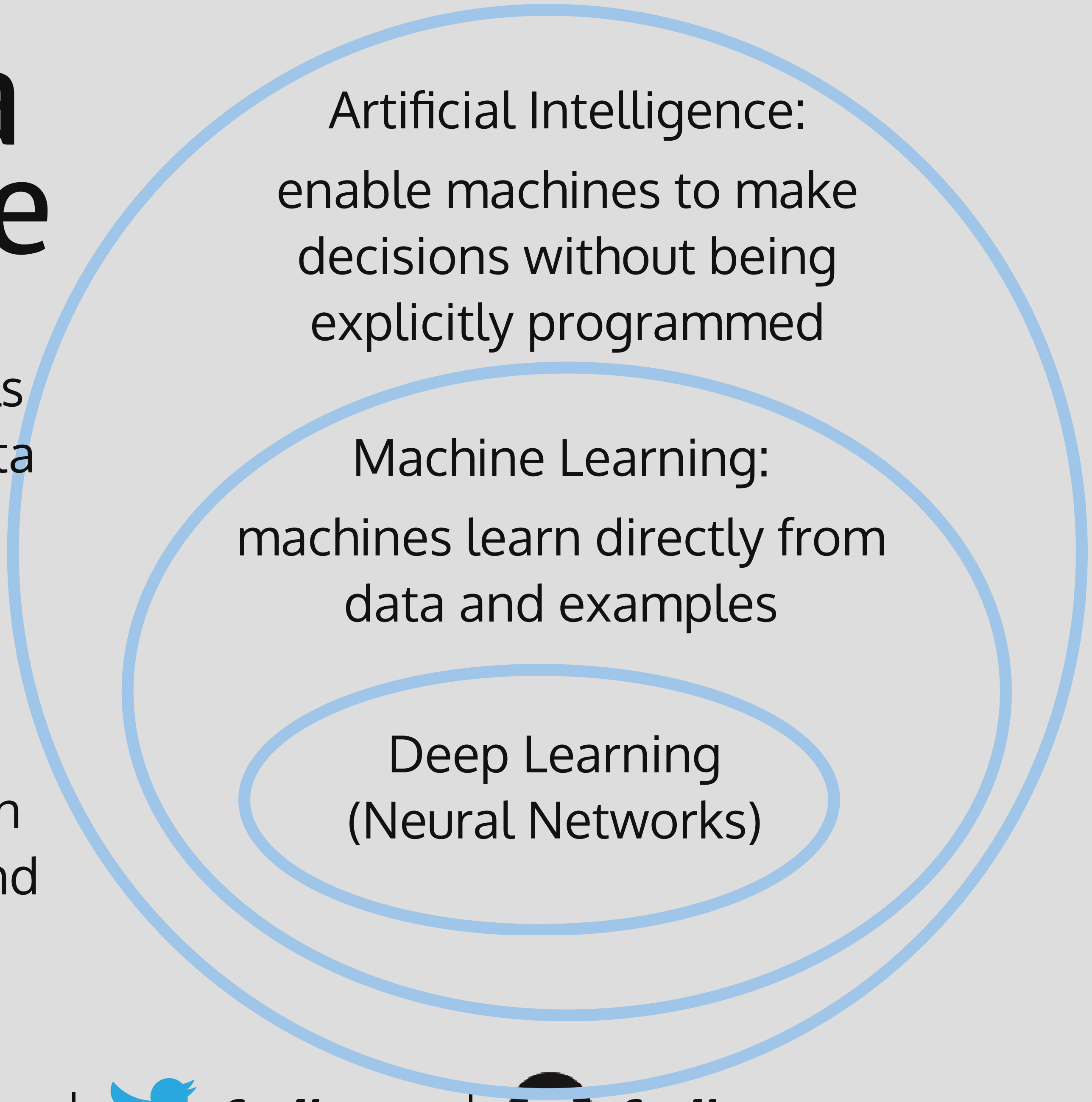


dr.federica bianco | *fbb.space* | *fedhere* | *fedhere*

foundations of data science for everyone

Data Science: the field of studies that deals with the extraction of information from data within a domain context to enable interpretation and prediction of phenomena.

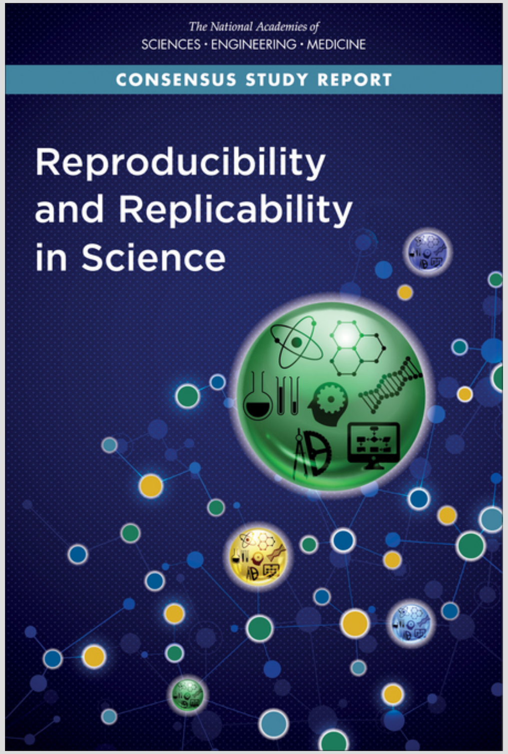
This includes development and application of statistical tools and machine learning and AI methods



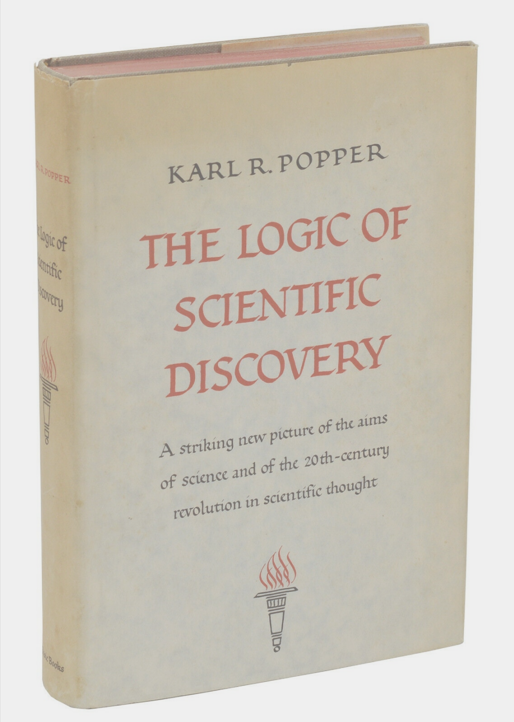
dr.federica bianco | fbb.space | fedhere | fedhere

3 General principles of "good" science

Reproducibility



Falisifiability



Parsimony

Reproducibility

Reproducible research means:

the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator. That is, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis in an attempt to yield the same results.

https://acmedsci.ac.uk/viewFile/56314e40aac61.pdf

Reproducible research in practice:

all numbers in a data analysis can be recalculated exactly (down to stochastic variables!)

using the **code** and **raw data** provided by the analyst.

- provide raw data and code to reduce it to all stages needed to get outputs
- provide code to reproduce all figures
- provide code to reproduce all number outcomes

Reproducibility

Reproducible research means:

the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator. That is, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis in an attempt to yield the same results.

https://acmedsci.ac.uk/viewFile/56314e40aac61.pdf

Reproducible research in practice:

all numbers in a data analysis can be recalculated exactly (down to stochastic variables!)

using the **code** and **raw data** provided by the analyst.

- use open data
- provide code to reproduce all figures
- provide code to reproduce all number outcomes

Reproducibility

Reproducible research means:

the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator. That is, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis in an attempt to yield the same results.

https://acmedsci.ac.uk/viewFile/56314e40aac61.pdf

Reproducible research in practice:

all numbers in a data analysis can be recalculated exactly (down to stochastic variables!)

using the **code** and **raw data** provided by the analyst.

- use open data
- code within notebooks
- provide code to reproduce all number outcomes

Reproducibility

Reproducible research means:

the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator. That is, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis in an attempt to yield the same results.

https://acmedsci.ac.uk/viewFile/56314e40aac61.pdf

Reproducible research in practice:

all numbers in a data analysis can be recalculated exactly (down to stochastic variables!)

using the **code** and **raw data** provided by the analyst.

- use open data
- code within notebooks
- set your python seeds!

```
import numpy as np
np.random.seed(302) # your favorite number goes here!
```

1 Kinds of Variables

Data Types

Categorical: These variables represent categories or groups and cannot be measured on a numerical scale.

Numerical:

... and ...

Nominal Variable: Nominal variables are categorical variables with no inherent order or ranking among categories. Examples include types of fruit or country of origin.

Ordinal Variable: Ordinal variables represent categories with a specific order or ranking but do not have equal intervals between them. An example is a survey response scale with options like "strongly disagree," "disagree," "neutral," "agree," and "strongly agree."

Data Types

Categorical:

Numerical: These variables represent quantities that can be measured on a numerical scale.

... and ...

Continuous Variable:
Continuous variables are measured on a numerical scale and can take on an infinite number of values within a given range. Examples include height, weight, or temperature.

Discrete Variable: Discrete variables also use numerical values, but they are distinct and separate, often counted in whole numbers. Examples include the number of students in a class or the number of cars in a parking lot.

Data Types

Categorical:

Numerical: These variables represent quantities that can be measured on a numerical scale.

... and ...

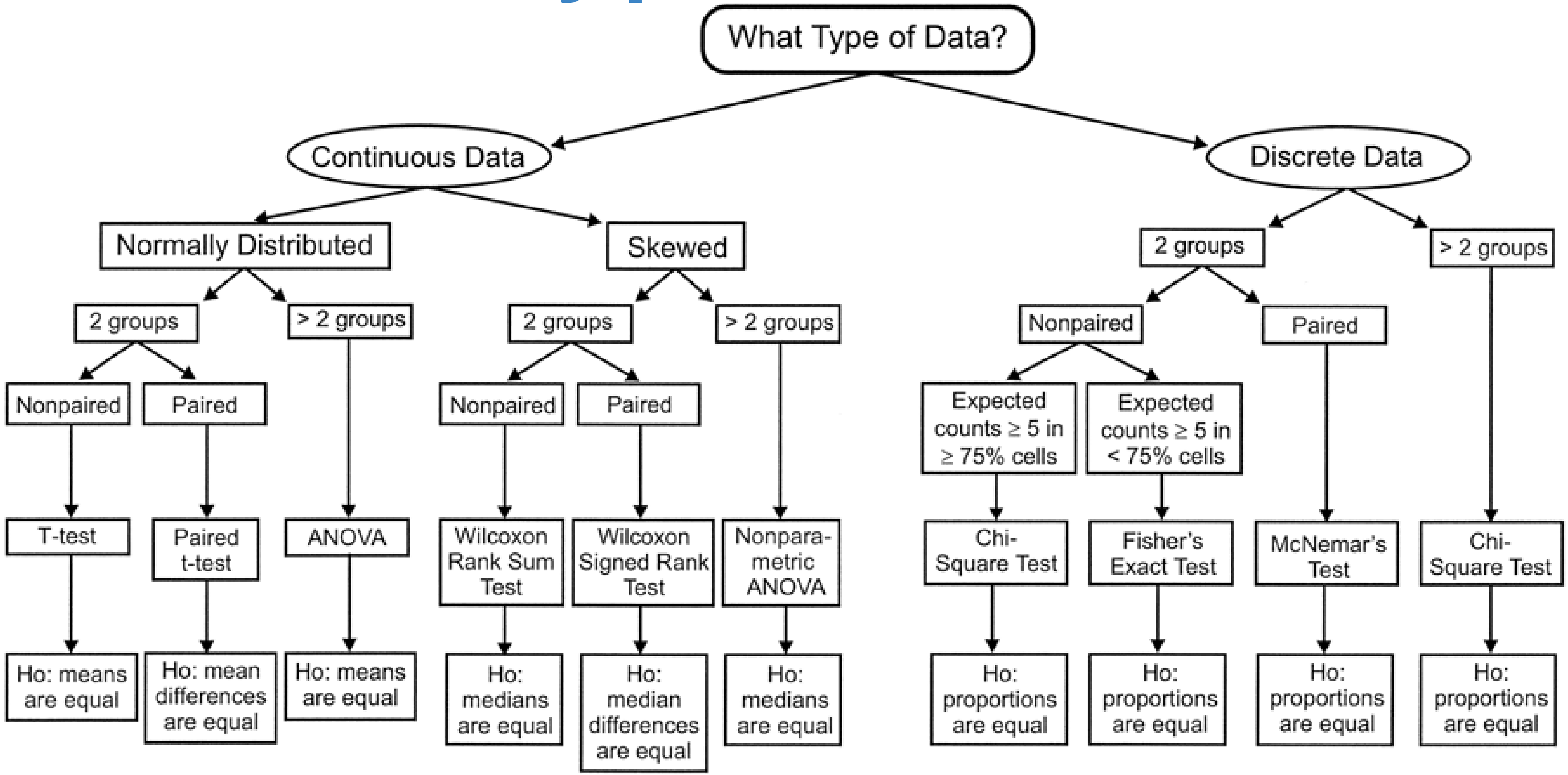
Continuous Variable: Continuous variables are measured on a numerical scale and can take on an infinite number of values within a given range. Examples include height, weight, or temperature.

Discrete Variable: Discrete variables also use numerical values, but they are distinct and separate, often counted in whole numbers. Examples include the number of students in a class or the number of cars in a parking lot.

Ratio Variable: has a meaningful zero point.

Interval Variable: equal intervals between values but no meaningful zero point.

When you choose a statistical test data type matters!



Source: Waning B, Montagne M: *Pharmacoepidemiology: Principles and Practice*: http://www.accesspharmacy.com

When you choose a statistical test data type matters!

Pearson's correlation

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$ *(only for continuous)*

$\bar{x}$: mean value of x

$\bar{y}$: mean value of y

n : number of datapoints

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Spearman's test

(Pearson's for ranked values)

$$\rho_{xy} = 1 - \frac{6 \sum_{i=1}^{N} (x_i - y_i)^2}{n(n^2 - 1)}$$

(continuous or ordinal)

How many dichotomous[+] (binary) variables?				
0	Both variables interval or ratio?			
	Y	Measures are linear? (No = monotonic[*])		
		Y	Pearson correlation	
		N	Spearman correlation	
	N	Both variables are ordinal?		
		Y	Kendall correlation	
		N	Both variables can be ranked?	
			Y	Kendall correlation
			N	Convert to frequency data and use Chi-square test for independence
1	Biserial Correlation Coefficient			
2	2 x 2 table?			
	Y	Phi		
	N	Cramer's V		

correlation

$\bar{x}$: mean value of x

$\bar{y}$: mean value of y

n : number of datapoints

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Pearson's correlation

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

Spearman's test

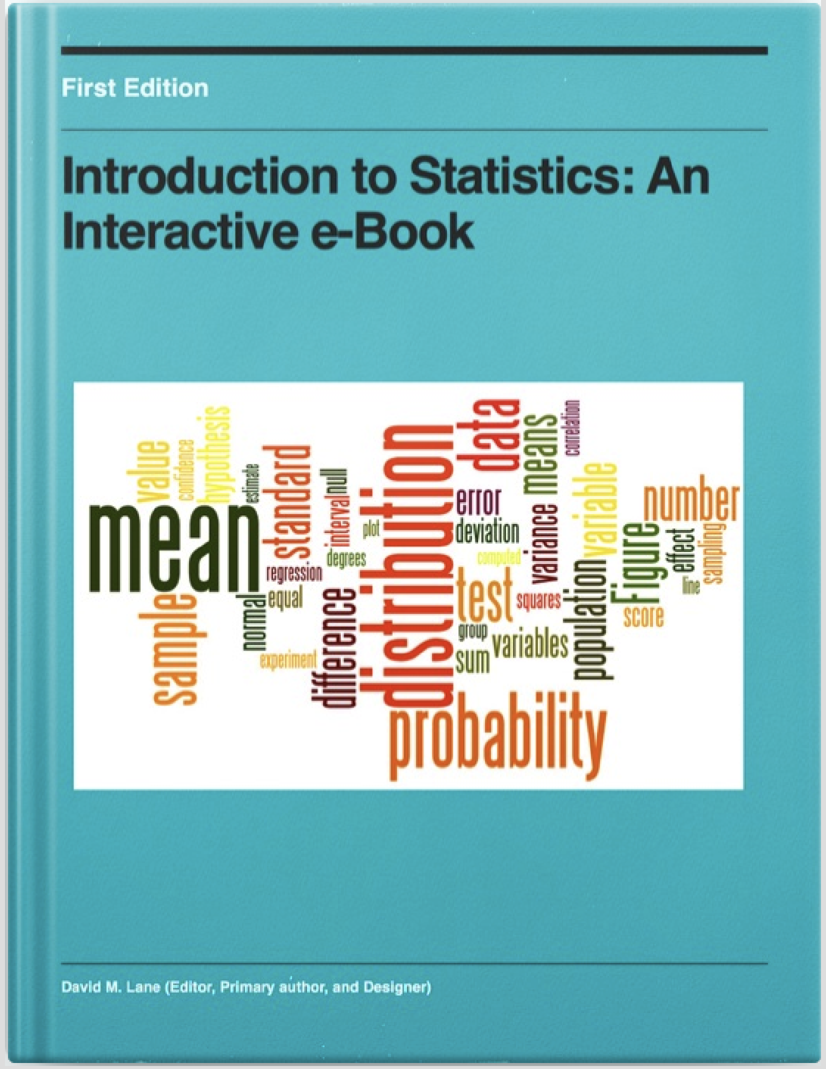
(Pearson's for ranked values)

$$\rho_{xy} = 1 - \frac{6 \sum_{i=1}^{N} (x_i - y_i)^2}{n(n^2 - 1)}$$

How many dichotomous+ (binary) variables?				
0	Both variables interval or ratio?			
0	Y	Measures are linear? (No = monotonic*)		
0	Y	Y	Pearson correlation	
0	Y	N	Spearman correlation	
0	N	Both variables are ordinal?		
0	N	Y	Kendall correlation	
0	N	N	Both variables can be ranked?	
0	N	N	Y	Kendall correlation
0	N	N	N	Convert to frequency data and use Chi-square test for independence
1	Biserial Correlation Coefficient			
2	2 x 2 table?			
2	Y	Phi		
2	N	Cramer's V		

Save the figure

2 probability and statistics



Introduction to Statistics: An Interactive e-Book

David M. Lane

Crush Course in Statistics

freee statistics book: http://onlinestatbook.com/

what are probability and statistics?

21

probability

Basic Probability *Frequentist* interpretation

fraction of times something happens

$\Updownarrow$

probability of it happening

Basic Probability Bayesian interpretation

represents a level of certainty relating to a potential outcome or idea:

if I believe the coin is unfair (tricked) then even if I get a head and a tail I will still believe I am more likely to get heads than tails

Basic Probability
Frequentist interpretation

P(E) = frequency of E

P(coin = head) = 6/11 = 0.55

fraction of times something happens

$\Updownarrow$

probability of it happening

Basic Probability
Frequentist interpretation

$$P(E) = \text{frequency of } E$$

$$P(\text{coin} = \text{head}) = 6/11 = 0.55$$

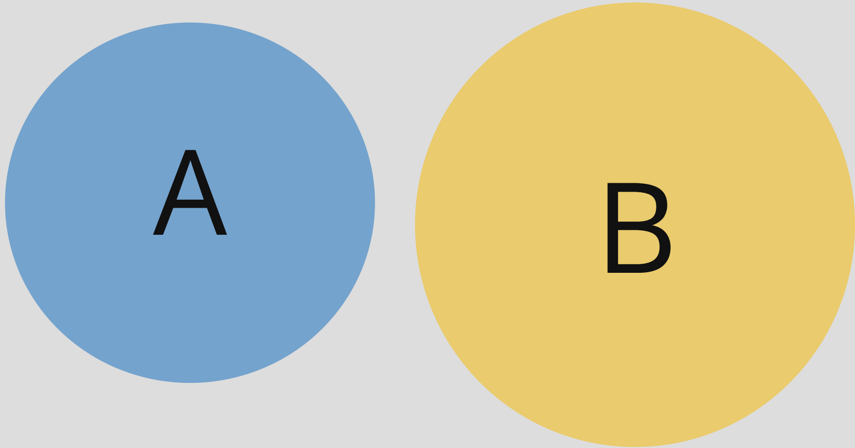
$$P(\text{coin} = \text{head}) = 51/100 = 0.51$$

fraction of times something happens

$\Updownarrow$

probability of it happening

Basic probability arithmetics



Probability Arithmetic

$$0 <= \ P(A) \ <= 1$$

$$P(A) + P(\bar{A}) = 1$$

disjoint events

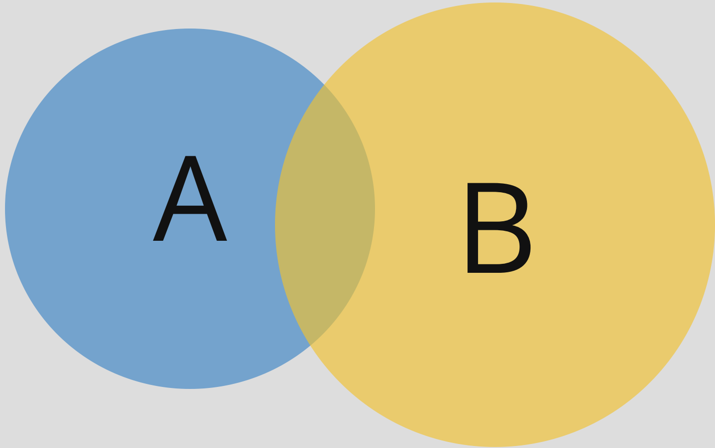
$$\text{if } P(A) \cap P(B) = 0 \text{ then :}$$

$$P(A \text{or} B) = P(A) + P(B)$$

$$P(A \text{and} B) = P(A) * P(B)$$

$$P(A|B) = P(A)$$

Basic probability arithmetics



Probability Arithmetic

in general :

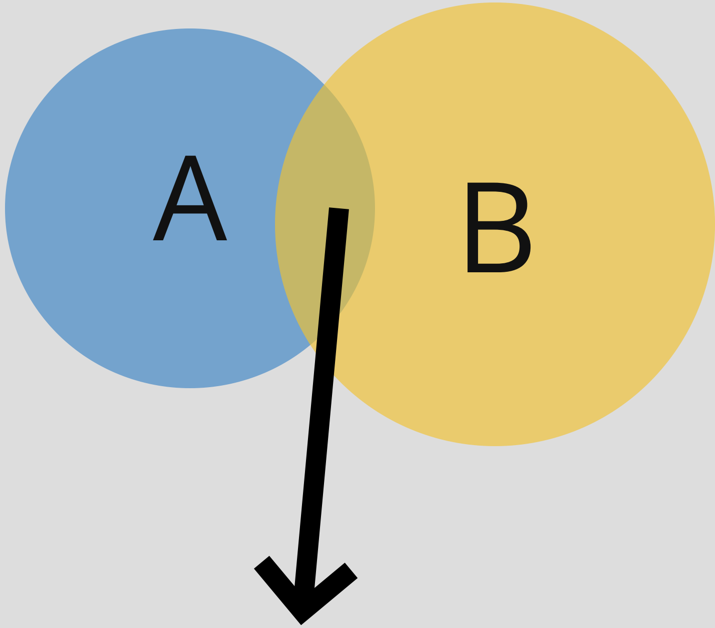
dependent events

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A|B) < P(A)$$

$$P(A \cap B) = P(A)P(B|A)$$

Basic probability arithmetics



$$P(A \cap B)$$

Probability Arithmetic

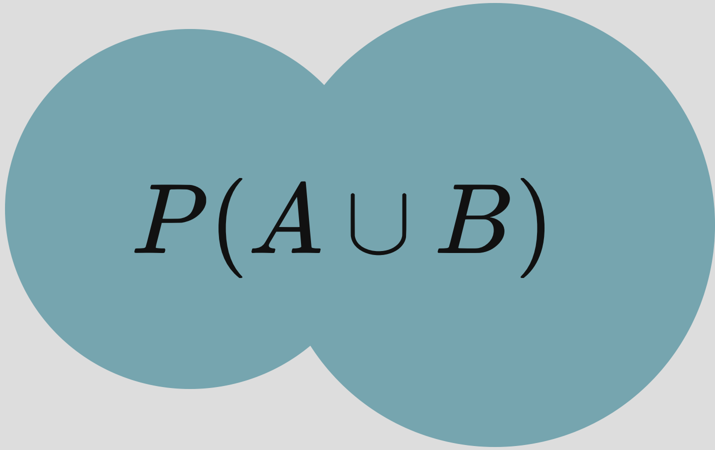
$$\text{in general :}$$

dependent events

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A|B) < P(A)$$

$$P(A \cap B) = P(A)P(B|A)$$



Basic probability arithmetics

Probability Arithmetic

in general :

dependent events

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A|B) < P(A)$$

$$P(A \cap B) = P(A)P(B|A)$$

coding time!

https://github.com/fedhere/DSPS_FBianco/blob/main/CodeExamples/Bayesian_coin_tosses_animation.ipynb

2.2 statistics

statistics

takes us from observing a limited number of samples to infer on the population

TAXONOMY

Distribution: a formula (a model)

Population: all of the elemnts of a "family"

Sample: a finite subset of the population that you observe

Phsyics Example

describe properties of the Population while the population is too large to be observed.

Statistical Mechanics:

explains the properties of the macroscopic system by statiscal knowledge of the microscopic system, even the the state of each element of the system cannot be known exactly

example: Maxwell Boltzman distribution of velocity of molecules in an ideal gas

https://upload.wikimedia.org/wikipedia/common
s/8/82/Simulation_of_gas_for_relaxation_demon
stration.gif?1567607773826

https://commons.wikimedia.org/wiki/File:Simulation_of_gas_for_relaxation_demonstration.gif

Phsyics Example

Boltzmann 1872

Entropy **2015**, *17*, 1971-2009; doi:10.3390/e17041971

OPEN ACCESS
entropy
ISSN 1099-4300
www.mdpi.com/journal/entropy

Article

Translation of Ludwig Boltzmann's Paper "On the Relationship between the Second Fundamental Theorem of the Mechanical Theory of Heat and Probability Calculations Regarding the Conditions for Thermal Equilibrium" Sitzungberichte der Kaiserlichen Akademie der Wissenschaften. Mathematisch-Naturwissen Classe. Abt. II, LXXVI 1877, pp 373-435 (Wien. Ber. 1877, 76:373-435). Reprinted in Wiss. Abhandlungen, Vol. II, reprint 42, p. 164-223, Barth, Leipzig, 1909

Kim Sharp * and Franz Matschinsky

https://www.mdpi.com/1099-4300/17/4/1971/pdf

The mechanical theory of heat assumes that the molecules of a gas are not at rest, but rather are in the liveliest motion. Hence, even though the body does not change its state, its individual molecules are always changing their states of motion, and the various molecules take up many different positions with respect to each other. The fact that we nevertheless observe completely definite laws of behaviour of warm bodies is to be attributed to the circumstance that the most random events, when they occur in the same proportions, give the same average value. For the molecules of the body are indeed so numerous, and their motion is so rapid,

Phsyics Example

Boltzmann 1872

Entropy **2015**, *17*, 1971-2009; doi:10.3390/e17041971

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Translation of Ludwig Boltzmann's Paper "On the Relationship between the Second Fundamental Theorem of the Mechanical Theory of Heat and Probability Calculations Regarding the Conditions for Thermal Equilibrium" Sitzungberichte der Kaiserlichen Akademie der Wissenschaften. Mathematisch-Naturwissen Classe. Abt. II, LXXVI 1877, pp 373-435 (Wien. Ber. 1877, 76:373-435). Reprinted in Wiss. Abhandlungen, Vol. II, reprint 42, p. 164-223, Barth, Leipzig, 1909

Kim Sharp * and Franz Matschinsky

https://www.mdpi.com/1099-4300/17/4/1971/pdf

that we can perceive nothing more than average values. One mi
compare the regularity of these average values with the amazin
constancy of the average numbers provided by statistics, which
also derived from processes each of which is determined by a com-
pletely unpredictable interaction with many other factors. Th

One must not confuse an incompletely known law, whose validity
is therefore in doubt, with a completely known law of the calculu
of probabilities; the latter, like the result of any other calculus,
a necessary consequence of definite premises, and is confirmed
insofar as these are correct, by experiment, provided sufficient

Phsyics Example

https://hal.archives-ouvertes.fr/hal-01662284/document

How statistics entered physics?

Olivier REY[1]

ABSTRACT: Now that statistics is a branch of mathematics, it is easy to imagine that its use in the field of human affairs is a by-product of modern science's way of looking at the world. Historical study contradicts such an idea: it is in the field of human affairs that quantitative statistics have developed, and it is only afterwards that it became a method for the natural sciences. Most physicists in the 19th century considered statistics all too human to have a place in the scientific study of nature. It took all Maxwell's authority and persuasion to make statistical analysis a new style of scientific thought in physics.

Such a point is of fundamental importance. Indeed, because of their initial anchoring in the human and social sphere, statistics suffered a long time from a bias among scientists and, in particular, among physicists. Making them acknowledge that statistics could be, and should be used in physics, was not a small undertaking, and that's this story I would like to sketch.

Law of large numbers

Suppose X1, X2, . . . , Xn are independent and identically-distributed variables, or i.i.d (=independent random variables with the same underlying distribution).

=>$X_{[i...j]}$ all have the same mean **μ** and standard deviation **σ**.

Let *X* be the mean of Xi $X = \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$

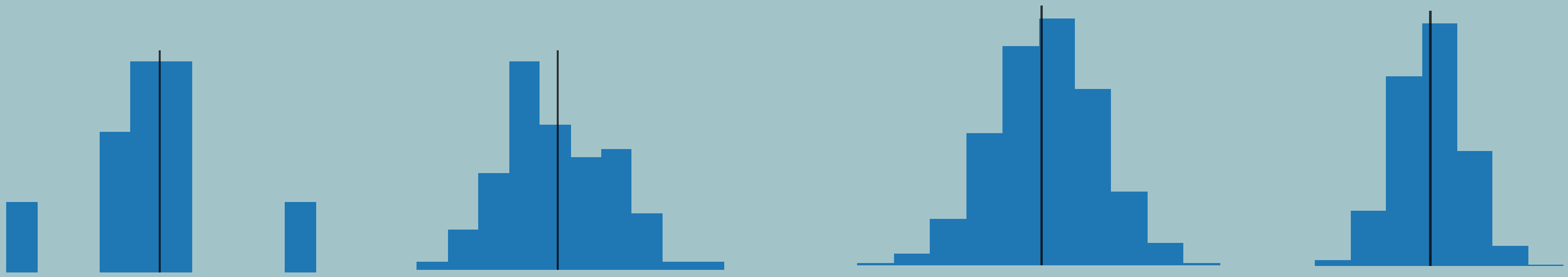
Note that *X* is itself a random variable.

As N grows, the probability that X is close to μ goes to 1.

HW

Law of Large Numbers

As the size of a _______________ tends to infinity the mean of the sample tends to the mean of the _________________

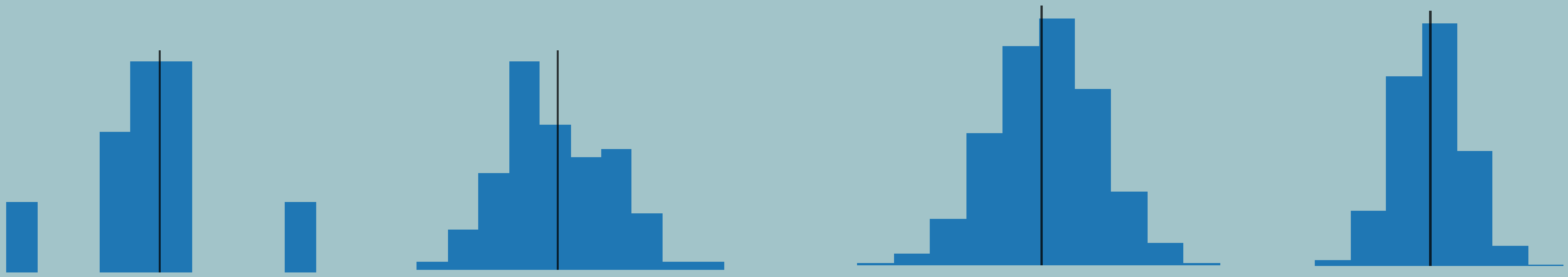


HW

Law of Large Numbers

In the limit of N -> infinity

the mean of a sample of size N approaches the mean of the population **μ**



descriptive statistics:

we summarize the proprties of a distribution

$$\mu_n = \int_{-\infty}^{\infty} (x-c)^n \, f(x) \, \mathrm{d}x.$$

3 descriptive statistics

TAXONOMY

Descriptive Statistics deals with the characterization of distributions

central tendency: mean, median, mode

spread : variance, interquantile range

Probability distributions

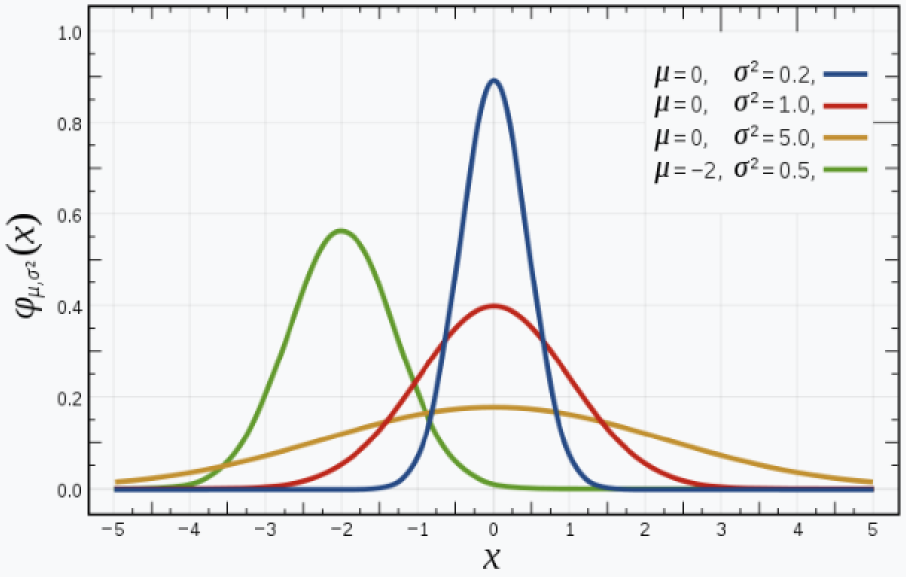
Gaussian

most common model of noise:

well behaved mathematically, symmetric, when we can we will assume our uncertainties are Gaussian distributed

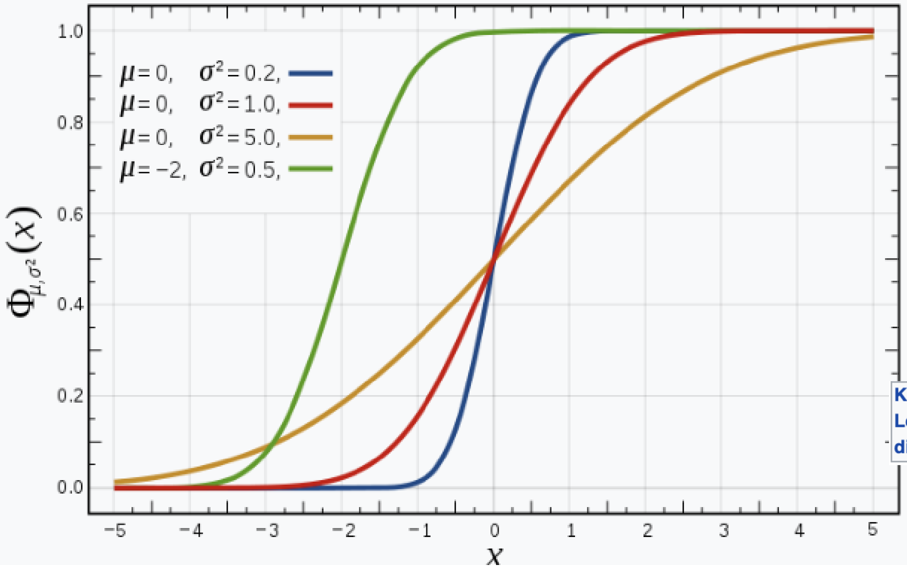
Normal Distribution

Probability density function



The red curve is the *standard normal distribution*

Cumulative distribution function



Notation	$\mathcal{N}(\mu, \sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 > 0$ = variance (squared sc
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Quantile	$\mu + \sigma\sqrt{2}\operatorname{erf}^{-1}(2F - 1)$
Mean	μ
Median	μ
Mode	μ
Variance	σ^2
Skewness	0
Ex. kurtosis	0
Entropy	$\frac{1}{2}\log(2\pi e\sigma^2)$
MGF	$\exp(\mu t + \sigma^2 t^2/2)$
CF	$\exp(i\mu t - \sigma^2 t^2/2)$
Fisher information	$\mathcal{I}(\mu, \sigma) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}$
Kullback-Leibler divergence	$D_{\mathrm{KL}}(\mathcal{N}_0 \| \mathcal{N}_1) = \frac{1}{2}\{(\sigma_0/\sigma_1)^2 + \frac{(\mu_1-\mu_0)^2}{\sigma_1^2} - 1 + 2\ln\frac{\sigma_1}{\sigma_0}\}$

Probability distributions

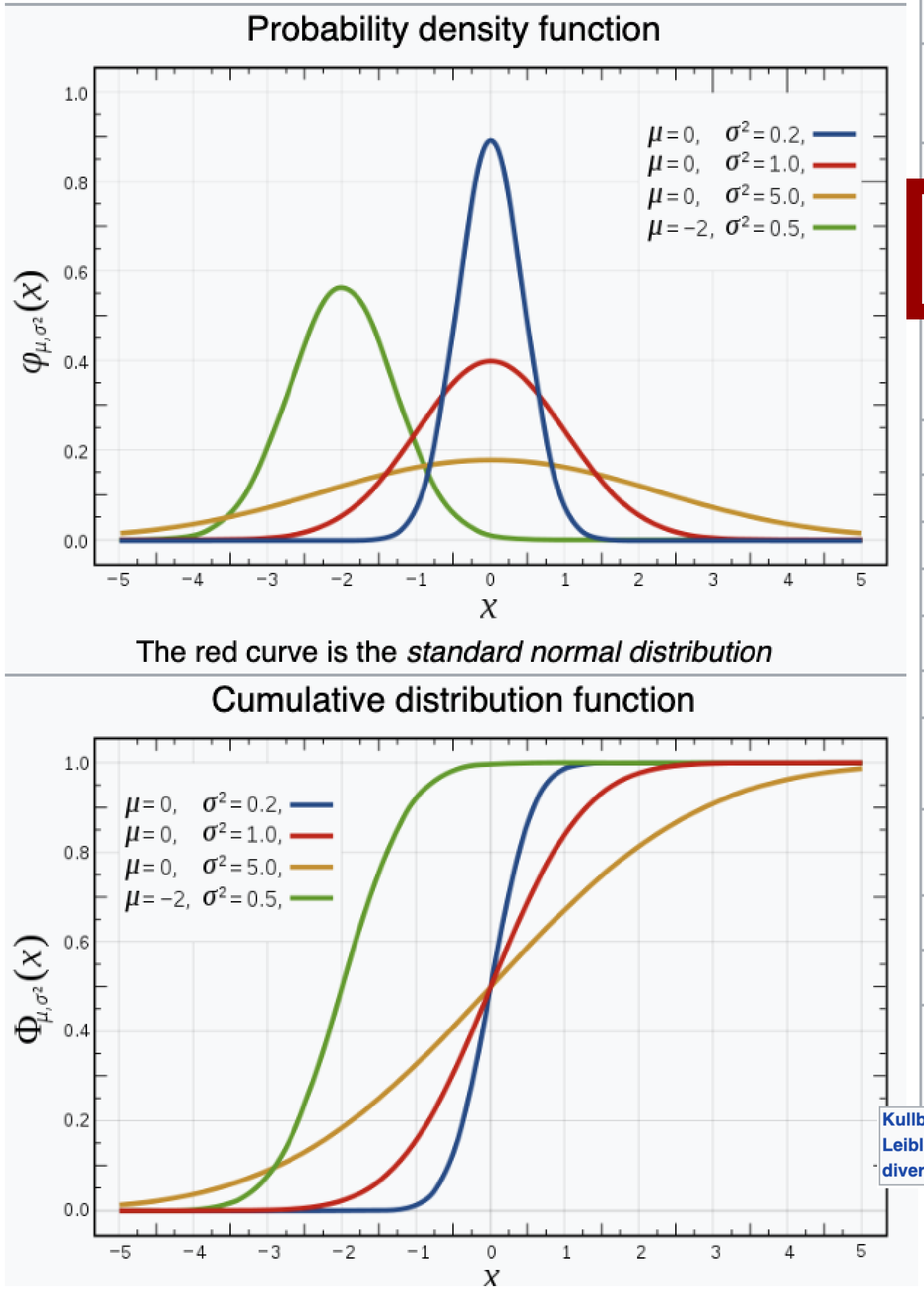
Gaussian

$$PDF: \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

most common model of noise:

well behaved mathematically, symmetric, when we can we will assume our uncertainties are Gaussian distributed

Normal Distribution



Notation	$\mathcal{N}(\mu,\sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 > 0$ = variance (squared sc
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Quantile	$\mu+\sigma\sqrt{2}\operatorname{erf}^{-1}(2F-1)$
Mean	μ
Median	μ
Mode	μ
Variance	σ^2
Skewness	0
Ex. kurtosis	0
Entropy	$\frac{1}{2}\log(2\pi e\sigma^2)$
MGF	$\exp(\mu t+\sigma^2t^2/2)$
CF	$\exp(i\mu t-\sigma^2t^2/2)$
Fisher information	$\mathcal{I}(\mu,\sigma) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}$
Kullback-Leibler divergence	$D_{KL}(\mathcal{N}_0\Vert\mathcal{N}_1) = \frac{1}{2}\{(\sigma_0/\sigma_1)^2 + \frac{(\mu_1-\mu_0)^2}{\sigma_1^2} - 1 + 2\ln\frac{\sigma_1}{\sigma_0}\}$

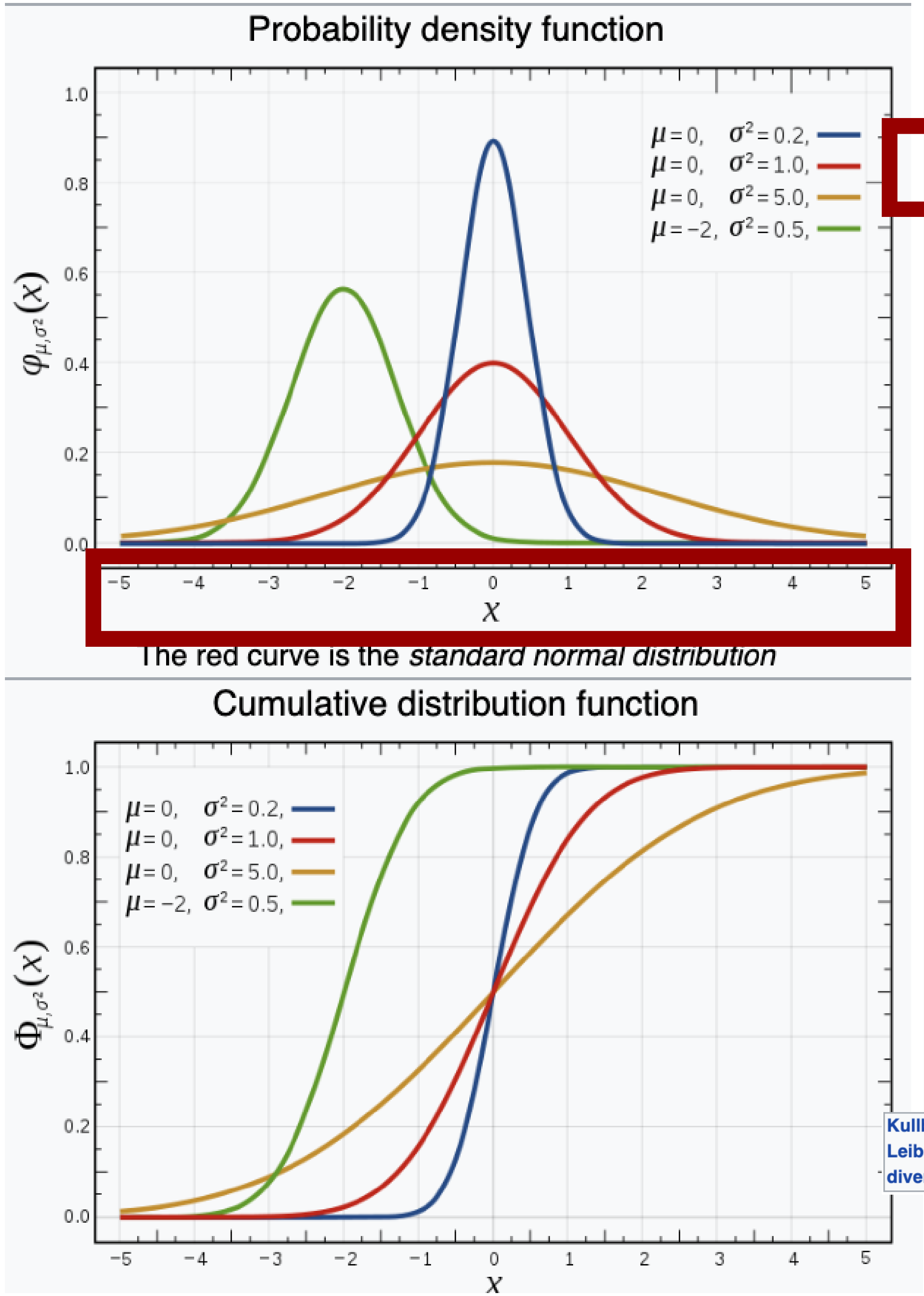
Probability distributions

Gaussian

most common model of noise:

well behaved mathematically, symmetric, when we can we will assume our uncertainties are Gaussian distributed

Normal Distribution



Notation	$\mathcal{N}(\mu,\sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 > 0$ = variance (squared sc
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Quantile	$\mu+\sigma\sqrt{2}\operatorname{erf}^{-1}(2F-1)$
Mean	μ
Median	μ
Mode	μ
Variance	σ^2
Skewness	0
Ex. kurtosis	0
Entropy	$\frac{1}{2}\log(2\pi e\sigma^2)$
MGF	$\exp(\mu t+\sigma^2 t^2/2)$
CF	$\exp(i\mu t-\sigma^2 t^2/2)$
Fisher information	$\mathcal{I}(\mu,\sigma)=\begin{pmatrix}1/\sigma^2 & 0\\ 0 & 2/\sigma^2\end{pmatrix}$
Kullback-Leibler divergence	$D_{\mathrm{KL}}(\mathcal{N}_0\|\mathcal{N}_1)=\frac{1}{2}\{(\sigma_0/\sigma_1)^2+\frac{(\mu_1-\mu_0)^2}{\sigma_1^2}-1+2\ln\frac{\sigma_1}{\sigma_0}\}$

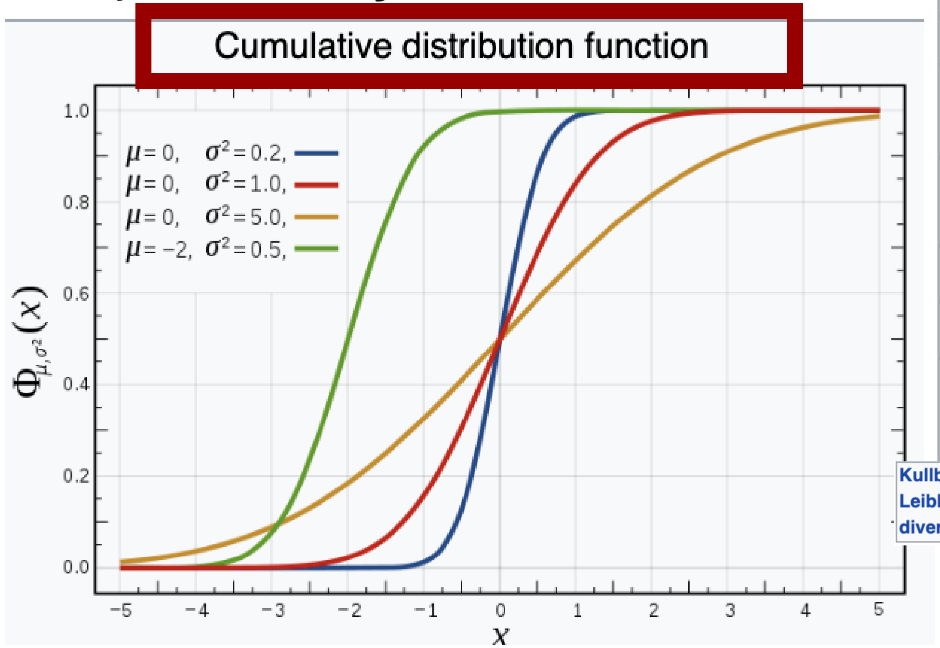
Probability distributions

Gaussian

most common model of noise:

well behaved mathematically, symmetric, when we can we will assume our uncertainties are Gaussian distributed

probability of a value > x



Notation	$\mathcal{N}(\mu,\sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 > 0$ = variance (squared sc
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Quantile	$\mu+\sigma\sqrt{2}\operatorname{erf}^{-1}(2F-1)$
Mean	μ
Median	μ
Mode	μ
Variance	σ^2
Skewness	0
Ex. kurtosis	0
Entropy	$\frac{1}{2}\log(2\pi e\sigma^2)$
MGF	$\exp(\mu t+\sigma^2 t^2/2)$
CF	$\exp(i\mu t-\sigma^2 t^2/2)$
Fisher information	$\mathcal{I}(\mu,\sigma)=\begin{pmatrix}1/\sigma^2 & 0\\ 0 & 2/\sigma^2\end{pmatrix}$
Kullback-Leibler divergence	$D_{\mathrm{KL}}(\mathcal{N}_0\Vert\mathcal{N}_1)=\frac{1}{2}\{(\sigma_0/\sigma_1)^2+\frac{(\mu_1-\mu_0)^2}{\sigma_1^2}-1+2\ln\frac{\sigma_1}{\sigma_0}\}$

descriptive statistics:

we summarize the proprties of a distribution

The moments of a distribution

$$\mu_n = \int_{-\infty}^{\infty} (x - c)^n \, f(x) \, \mathrm{d}x.$$

descriptive statistics:

we summarize the proprties of a distribution

The moments of a distribution $$\mu_n = \int_{-\infty}^{\infty} (x - c)^n \, f(x) \, \mathrm{d}x.$$

central tendency (n=1)

mean: n=1 $$\mu = \frac{1}{N} \sum_1^N x_i$$

other measures of centeral tendency:

median: 50% of the distribution is to the left,

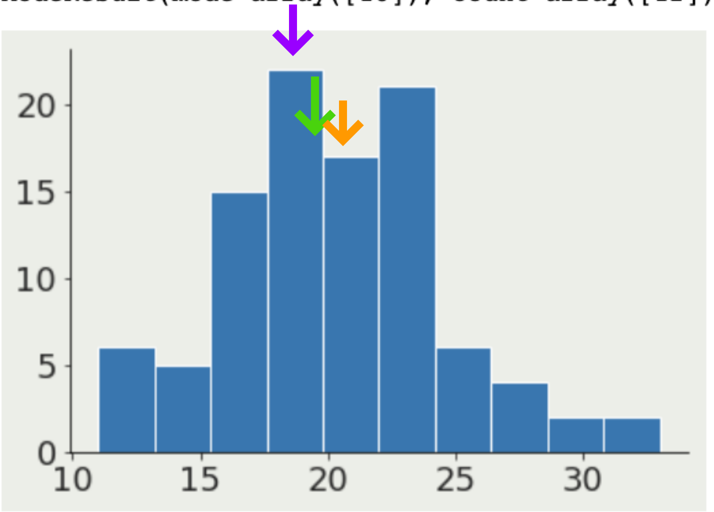
50% to the right

mode: most popular value in the distribution

```
dist = sp.stats.poisson.rvs(size=100, mu=20)
pl.hist(dist)
print(dist.mean())
print(np.median(dist))
print(sp.stats.mode(dist))
```

executed in 125ms, finished 15:01:20 2019-09-09

```
20.06
20.0
ModeResult(mode=array([18]), count=array([12]))
```



descriptive statistics:

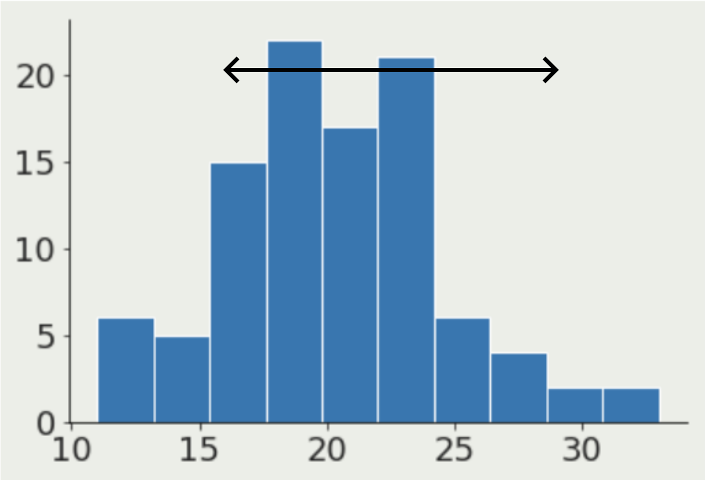
we summarize the proprties of a distribution

The moments of a distribution $$\mu_n = \int_{-\infty}^{\infty} (x-c)^n \, f(x) \, \mathrm{d}x.$$

```
dist = sp.stats.poisson.rvs(size=100, mu=20)
pl.hist(dist)
print(dist.mean())
print(np.median(dist))
print(sp.stats.mode(dist))
```

executed in 125ms, finished 15:01:20 2019-09-09

```
20.06
20.0
ModeResult(mode=array([18]), count=array([12]))
```



spread (n=2)

variance: n=2 $$\operatorname{Var}(X) = \operatorname{E}\left[(X-\mu)^2\right]$$

standard deviation $$\sigma(X) = \operatorname{E}\left[(X-\mu)\right]$$

descriptive statistics:

we summarize the proprties of a distribution

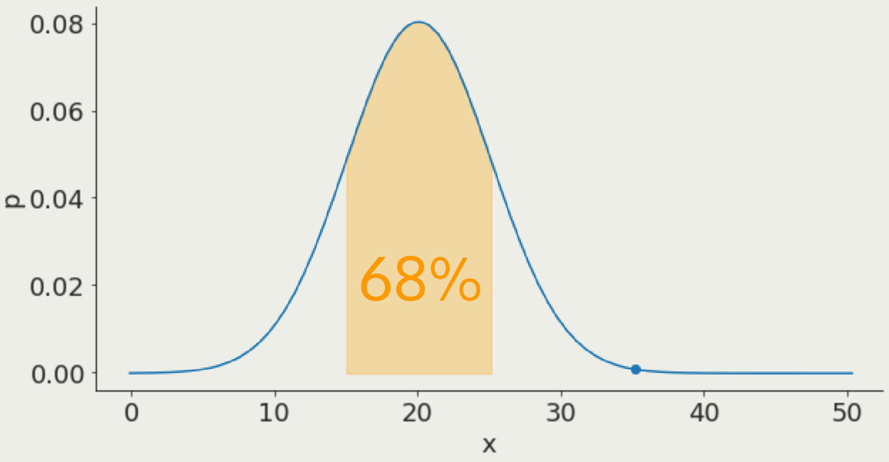
$$\mu_n = \int_{-\infty}^{\infty} (x - c)^n \, f(x) \, \mathrm{d}x.$$

variance: n=2 $$\mathrm{Var}(X) = \mathrm{E}\left[(X - \mu)^2\right]$$

standard deviation $$\sigma(X) = \mathrm{E}\left[(X - \mu)\right]$$

In a Gaussian distribution:

1σ contains 68% of the distribution



descriptive statistics:

we summarize the proprties of a distribution

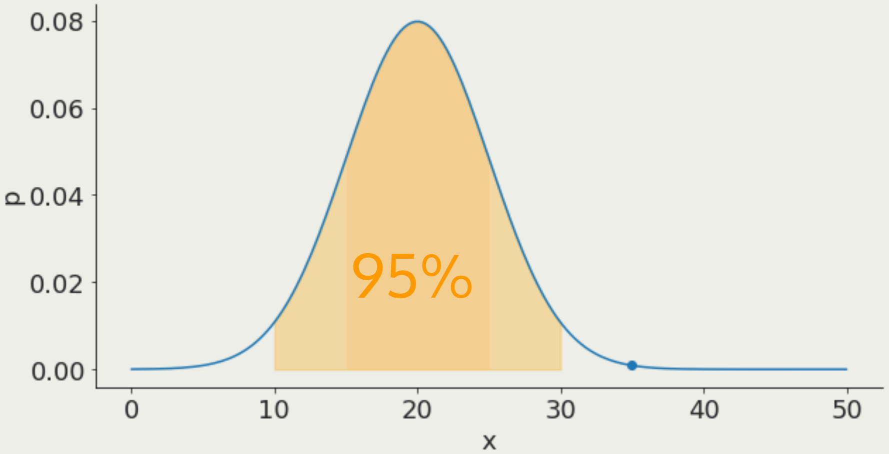
$$\mu_n = \int_{-\infty}^{\infty} (x - c)^n \, f(x) \, \mathrm{d}x.$$

variance: n=2 $$\operatorname{Var}(X) = \operatorname{E}\left[(X - \mu)^2\right]$$

standard deviation $$\sigma(X) = \operatorname{E}\left[(X - \mu)\right]$$

In a Gaussian distribution:

2σ contains 95% of the distribution



descriptive statistics:

we summarize the proprties of a distribution

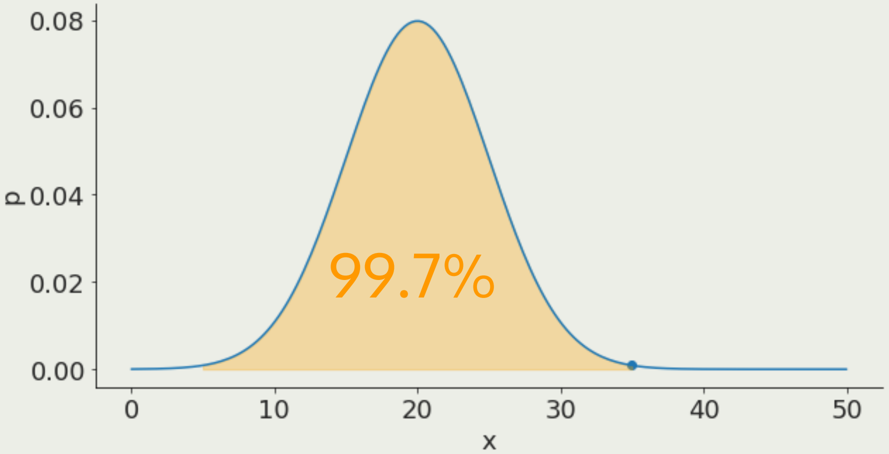
$$\mu_n = \int_{-\infty}^{\infty} (x-c)^n \, f(x) \, \mathrm{d}x.$$

variance: n=2 $$\operatorname{Var}(X) = \operatorname{E}\left[(X-\mu)^2\right]$$

standard deviation $$\sigma(X) = \operatorname{E}\left[(X-\mu)\right]$$

In a Gaussian distribution:

3σ contains 99.7% of the distribution



Memorize the following:

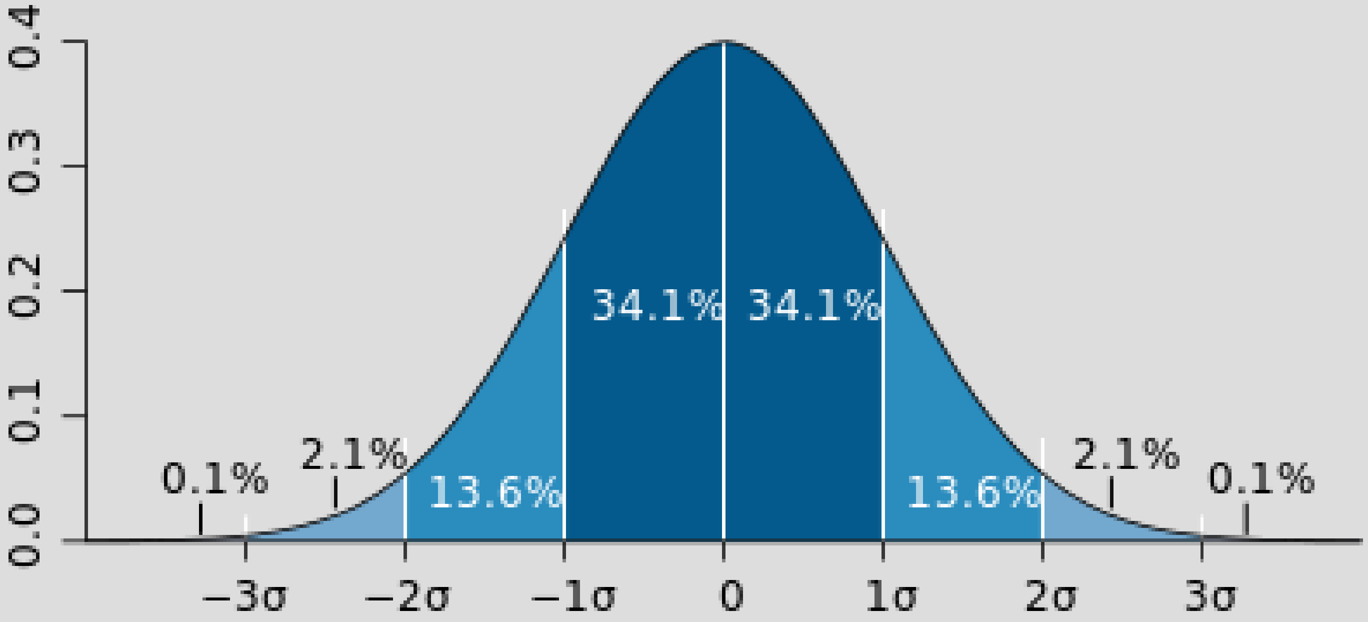
$1\sigma = 68\%$

$2\sigma = 95\%$

$3\sigma = 99.7\%$

$5\sigma = 99.999971428$

$= 1$ in 3.5 million



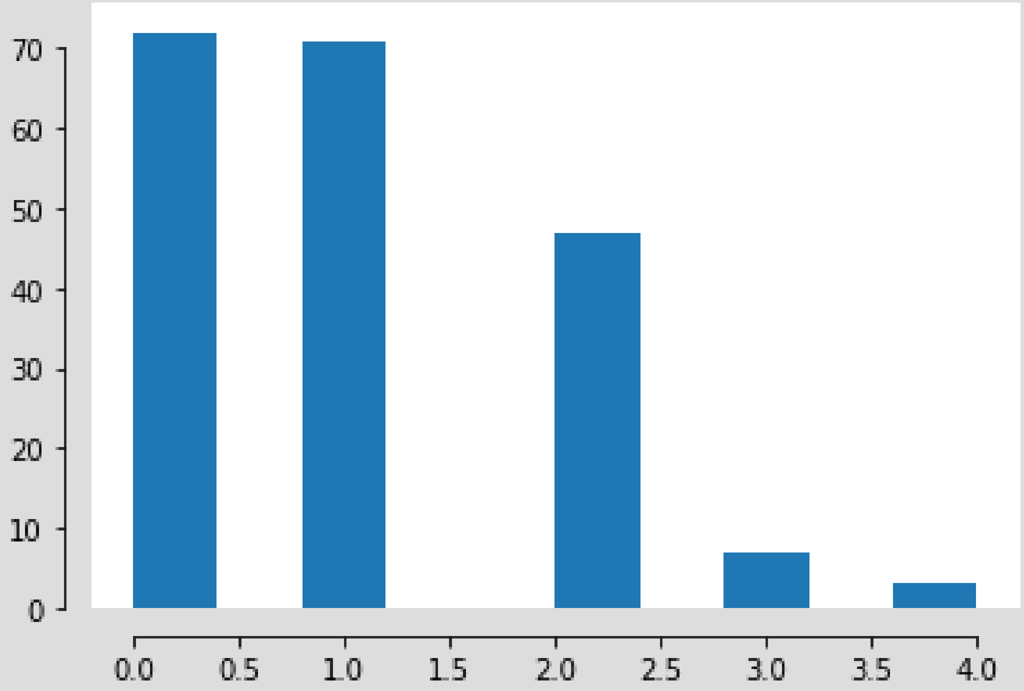
p-value statistics

When we set a confidence value or interval on inferred quantities we imply that we had 1 in x chances of getting that result (technically "a result as extreme as that" ... we will see this in more detail

distributions

parameters (*lambda*=1)

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



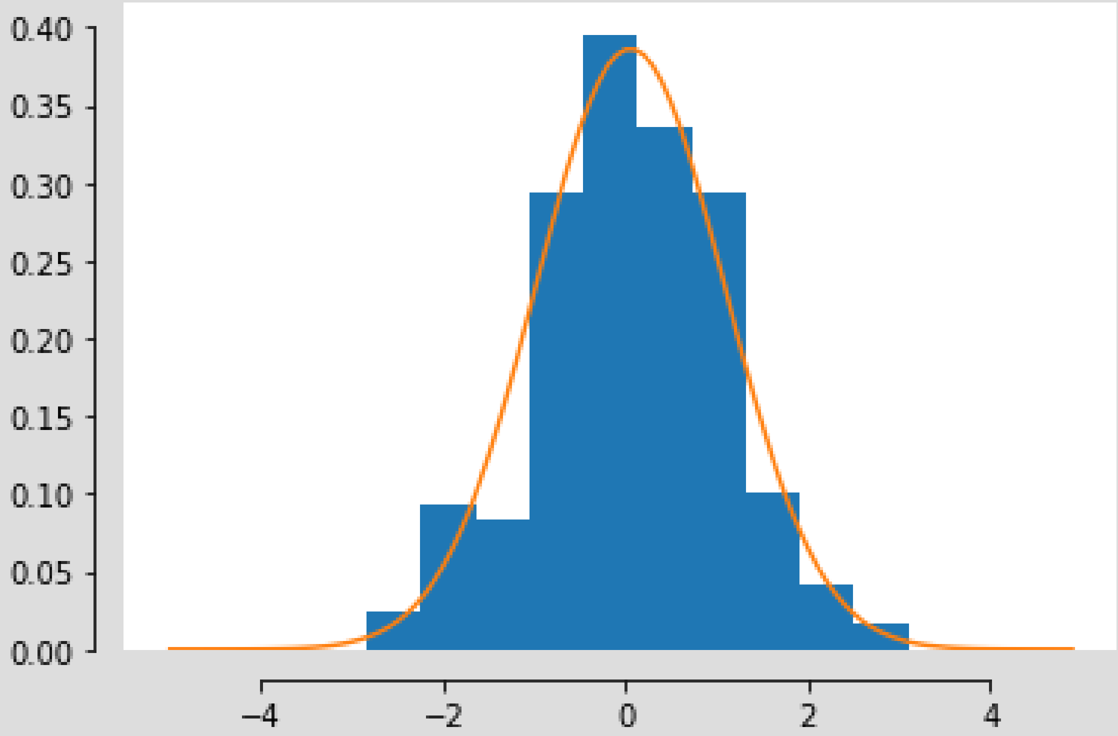
Poisson

discrete support $(1, +\inf]$

support

parameters (-0.1, 0.9)

$$N(r|\mu, \sigma) \sim \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(r-\mu)^2}{2\sigma^2}}$$



normal or Gaussian

continuous support $[-\inf, +\inf]$

Probability distributions

Binomial

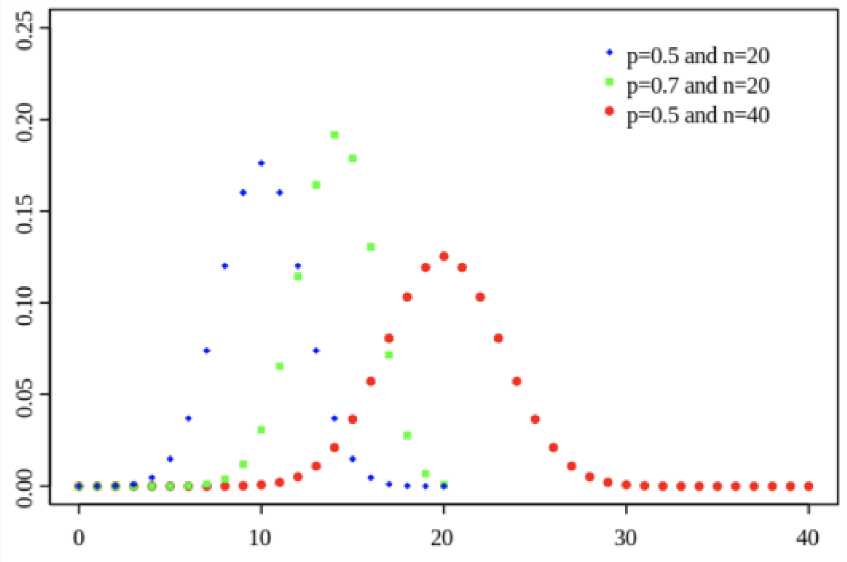
Coin toss:

fair coin: p=0.5 n=1

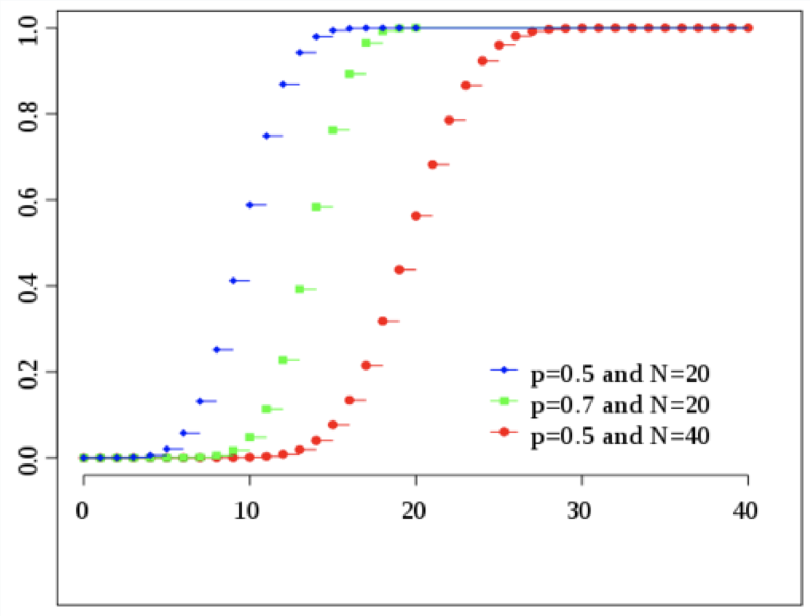
Vegas coin: p≠0.5 n=1

Binomial distribution

Probability mass function



Cumulative distribution function



Notation	$B(n,p)$
Parameters	$n \in \{0,1,2,\ldots\}$ – number of trials $p \in [0,1]$ – success probability for each trial
Support	$k \in \{0,1,\ldots,n\}$ – number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	$np(1-p)$
Skewness	$\frac{1-2p}{\sqrt{np(1-p)}}$
Ex. kurtosis	$\frac{1-6p(1-p)}{np(1-p)}$
Entropy	$\frac{1}{2}\log_2(2\pi e np(1-p)) + O\left(\frac{1}{n}\right)$ in shannons. For nats, use the natural log in the log.
MGF	$(1-p+pe^t)^n$
CF	$(1-p+pe^{it})^n$
PGF	$G(z) = [(1-p)+pz]^n$
Fisher information	$g_n(p) = \frac{n}{p(1-p)}$ (for fixed n)

Probability distributions

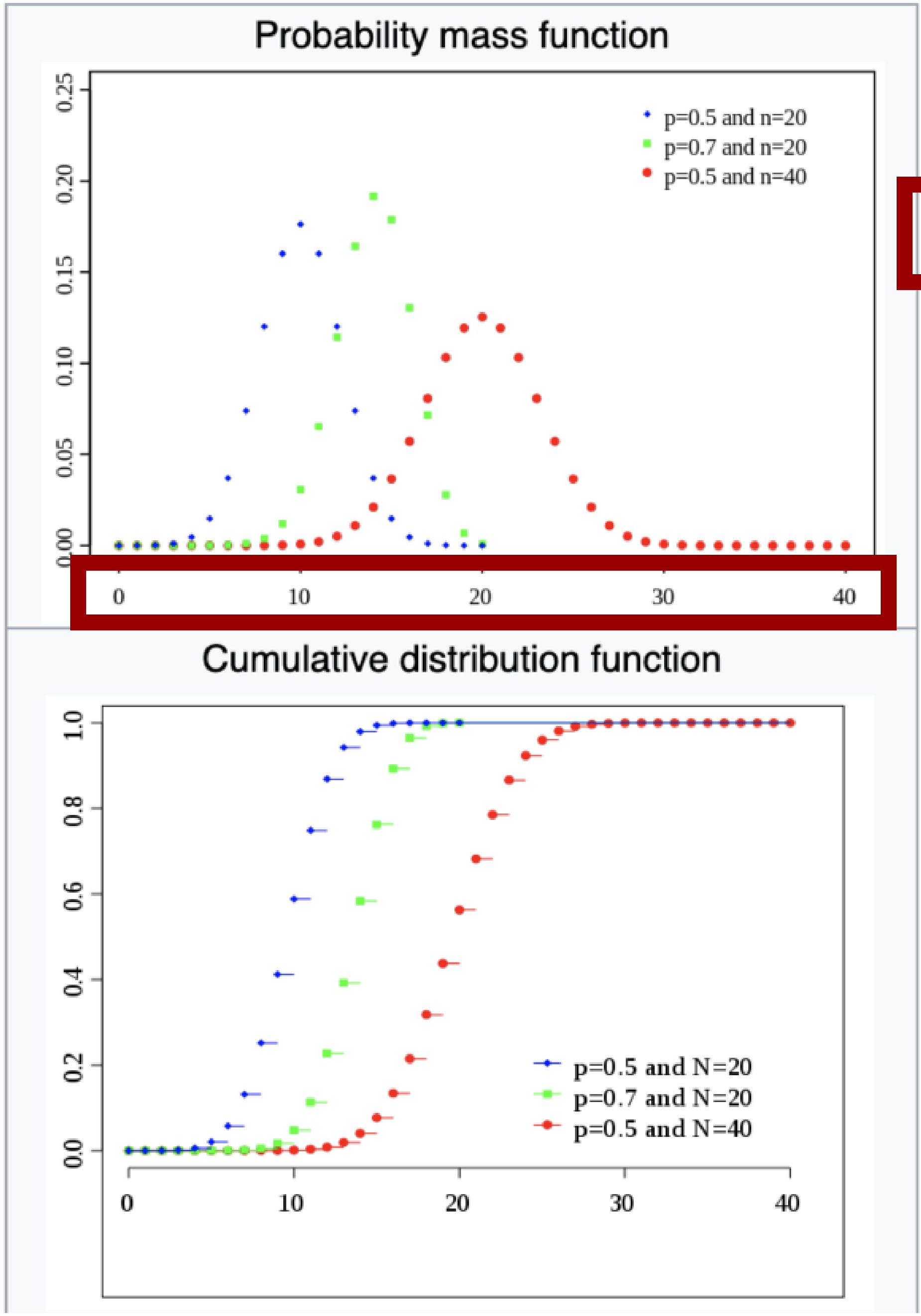
Binomial

Coin toss:

fair coin: p=0.5 n=1

Vegas coin: p≠0.5 n=1

Binomial distribution



Notation	$B(n, p)$
Parameters	$n \in \{0, 1, 2, \ldots\}$ – number of trials $p \in [0, 1]$ – success probability for each trial
Support	$k \in \{0, 1, \ldots, n\}$ – number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	$np(1-p)$
Skewness	$\frac{1-2p}{\sqrt{np(1-p)}}$
Ex. kurtosis	$\frac{1-6p(1-p)}{np(1-p)}$
Entropy	$\frac{1}{2}\log_2(2\pi e np(1-p)) + O\left(\frac{1}{n}\right)$ in shannons. For nats, use the natural log in the log.
MGF	$(1-p+pe^t)^n$
CF	$(1-p+pe^{it})^n$
PGF	$G(z) = [(1-p)+pz]^n$
Fisher information	$g_n(p) = \frac{n}{p(1-p)}$ (for fixed n)

Probability distributions

Binomial

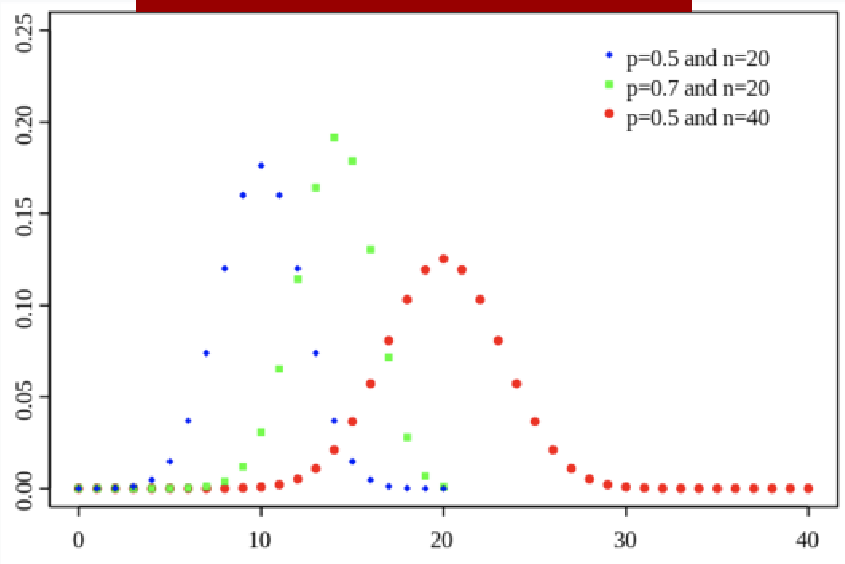
Coin toss:

fair coin: p=0.5 n=1

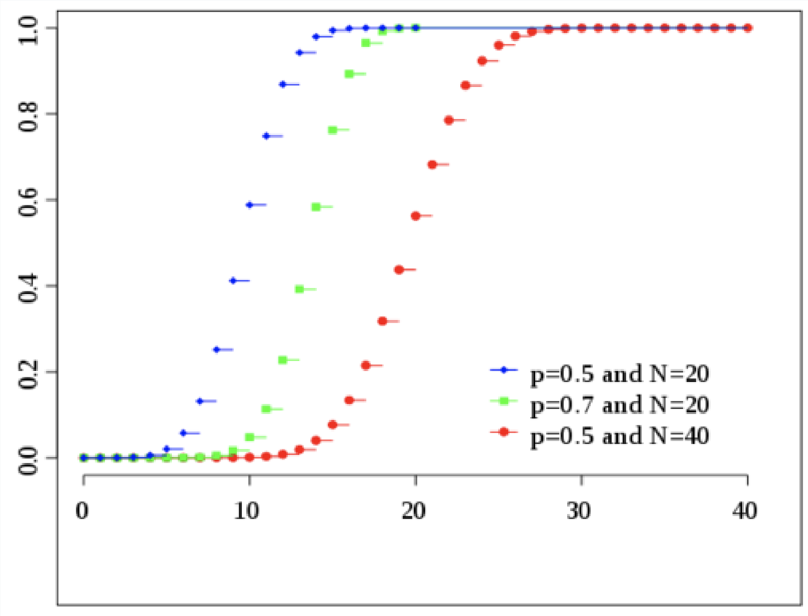
Vegas coin: p≠0.5 n=1

Binomial distribution

Probability mass function



Cumulative distribution function



Notation	$B(n,p)$
Parameters	$n \in \{0,1,2,\dots\}$ – number of trials $p \in [0,1]$ – success probability for each trial
Support	$k \in \{0,1,\dots,n\}$ – number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	$np(1-p)$
Skewness	$\frac{1-2p}{\sqrt{np(1-p)}}$
Ex. kurtosis	$\frac{1-6p(1-p)}{np(1-p)}$
Entropy	$\frac{1}{2}\log_2(2\pi enp(1-p)) + O\left(\frac{1}{n}\right)$ in shannons. For nats, use the natural log in the log.
MGF	$(1-p+pe^t)^n$
CF	$(1-p+pe^{it})^n$
PGF	$G(z) = [(1-p)+pz]^n$
Fisher information	$g_n(p) = \frac{n}{p(1-p)}$ (for fixed n)

Probability distributions

Binomial

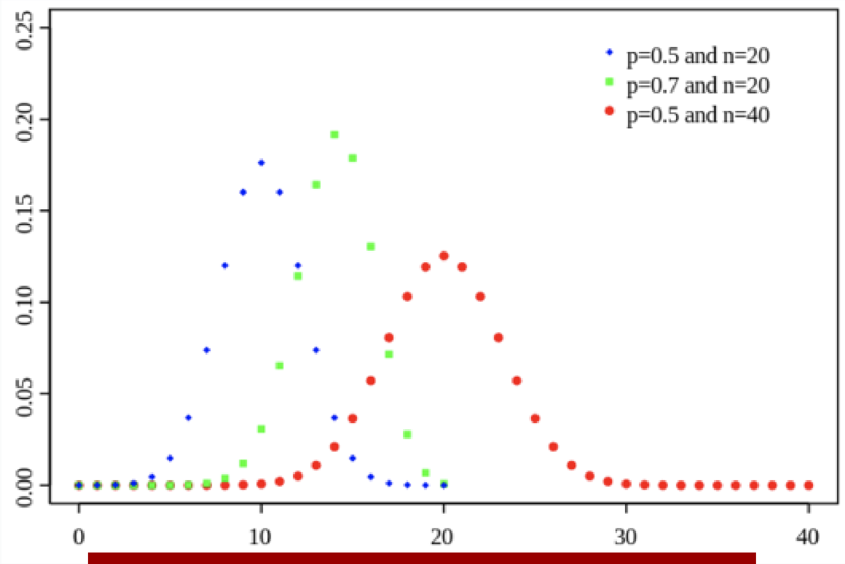
Coin toss:

fair coin: p=0.5 n=1

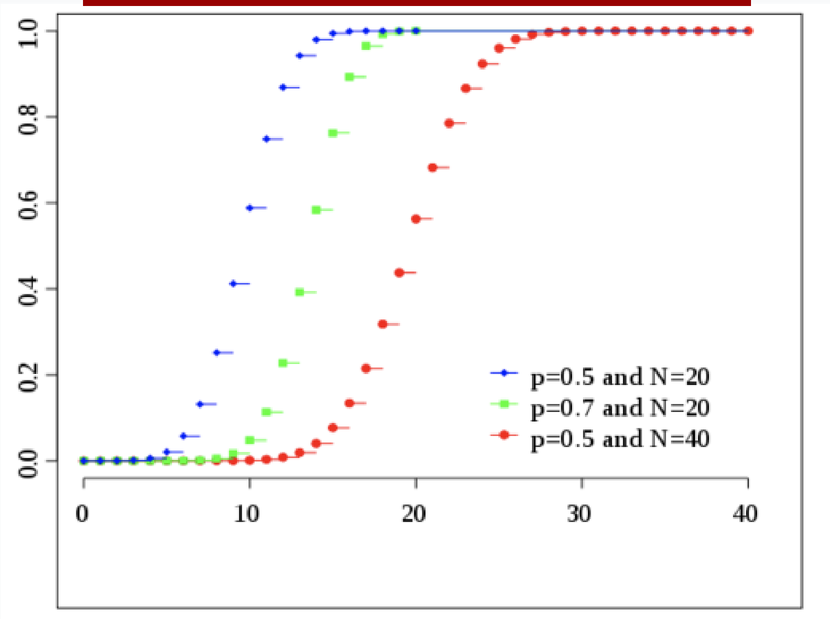
Vegas coin: p≠0.5 n=1

Binomial distribution

Probability mass function



Cumulative distribution function



Notation	$B(n,p)$
Parameters	$n \in \{0,1,2,\ldots\}$ – number of trials $p \in [0,1]$ – success probability for each trial
Support	$k \in \{0,1,\ldots,n\}$ – number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	$np(1-p)$
Skewness	$\frac{1-2p}{\sqrt{np(1-p)}}$
Ex. kurtosis	$\frac{1-6p(1-p)}{np(1-p)}$
Entropy	$\frac{1}{2}\log_2(2\pi enp(1-p)) + O\left(\frac{1}{n}\right)$ in shannons. For nats, use the natural log in the log.
MGF	$(1-p+pe^t)^n$
CF	$(1-p+pe^{it})^n$
PGF	$G(z) = [(1-p)+pz]^n$
Fisher information	$g_n(p) = \frac{n}{p(1-p)}$ (for fixed n)

Probability distributions

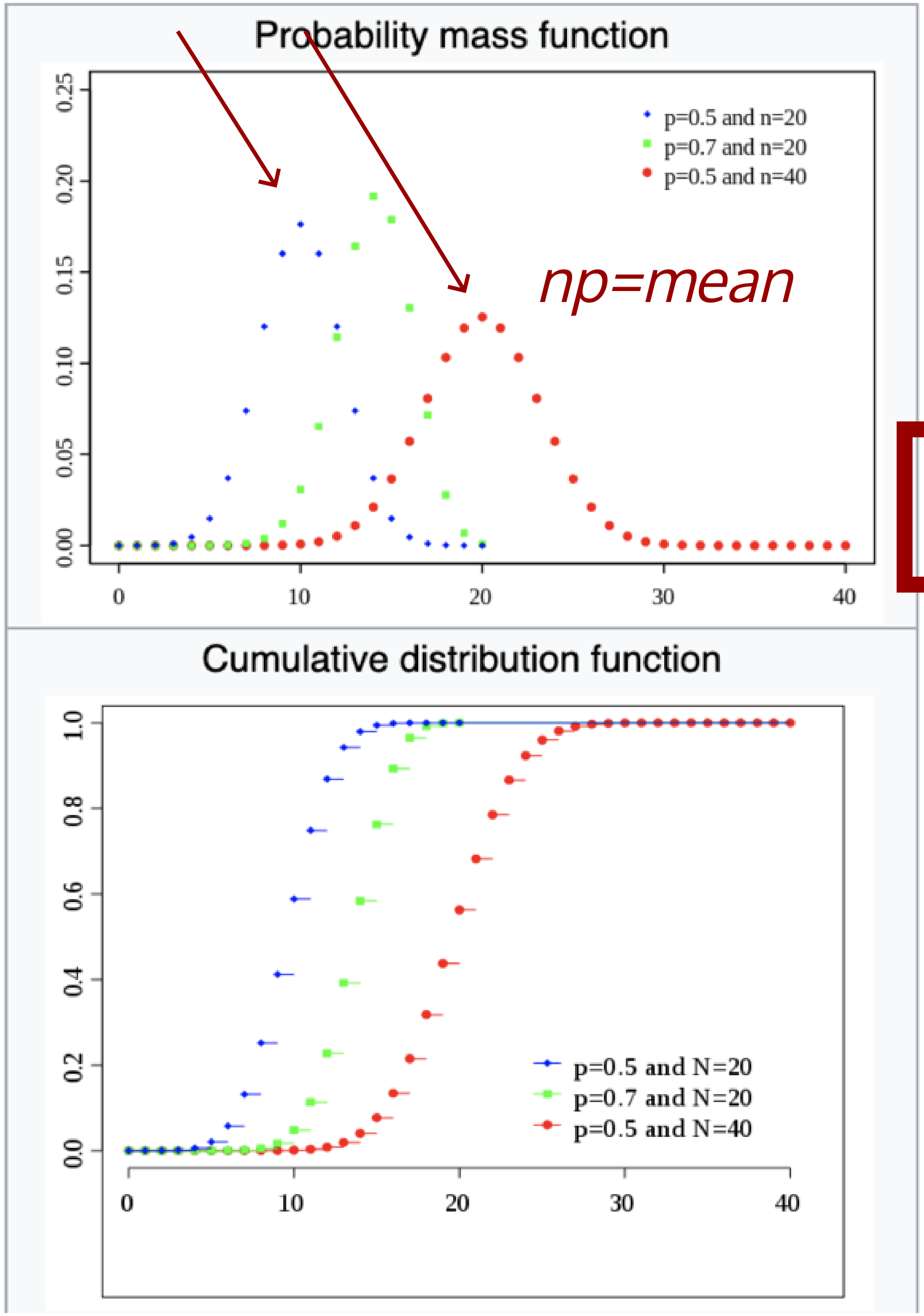
Binomial

Coin toss:

fair coin: p=0.5 n=1

Vegas coin: p≠0.5 n=1

Binomial distribution



central tendency

Notation	$B(n,p)$
Parameters	$n \in \{0,1,2,\ldots\}$ – number of trials $p \in [0,1]$ – success probability for each trial
Support	$k \in \{0,1,\ldots,n\}$ – number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	$np(1-p)$
Skewness	$\frac{1-2p}{\sqrt{np(1-p)}}$
Ex. kurtosis	$\frac{1-6p(1-p)}{np(1-p)}$
Entropy	$\frac{1}{2}\log_2(2\pi enp(1-p)) + O\left(\frac{1}{n}\right)$ in shannons. For nats, use the natural log in the log.
MGF	$(1-p+pe^t)^n$
CF	$(1-p+pe^{it})^n$
PGF	$G(z) = [(1-p)+pz]^n$
Fisher information	$g_n(p) = \frac{n}{p(1-p)}$ (for fixed n)

Probability distributions

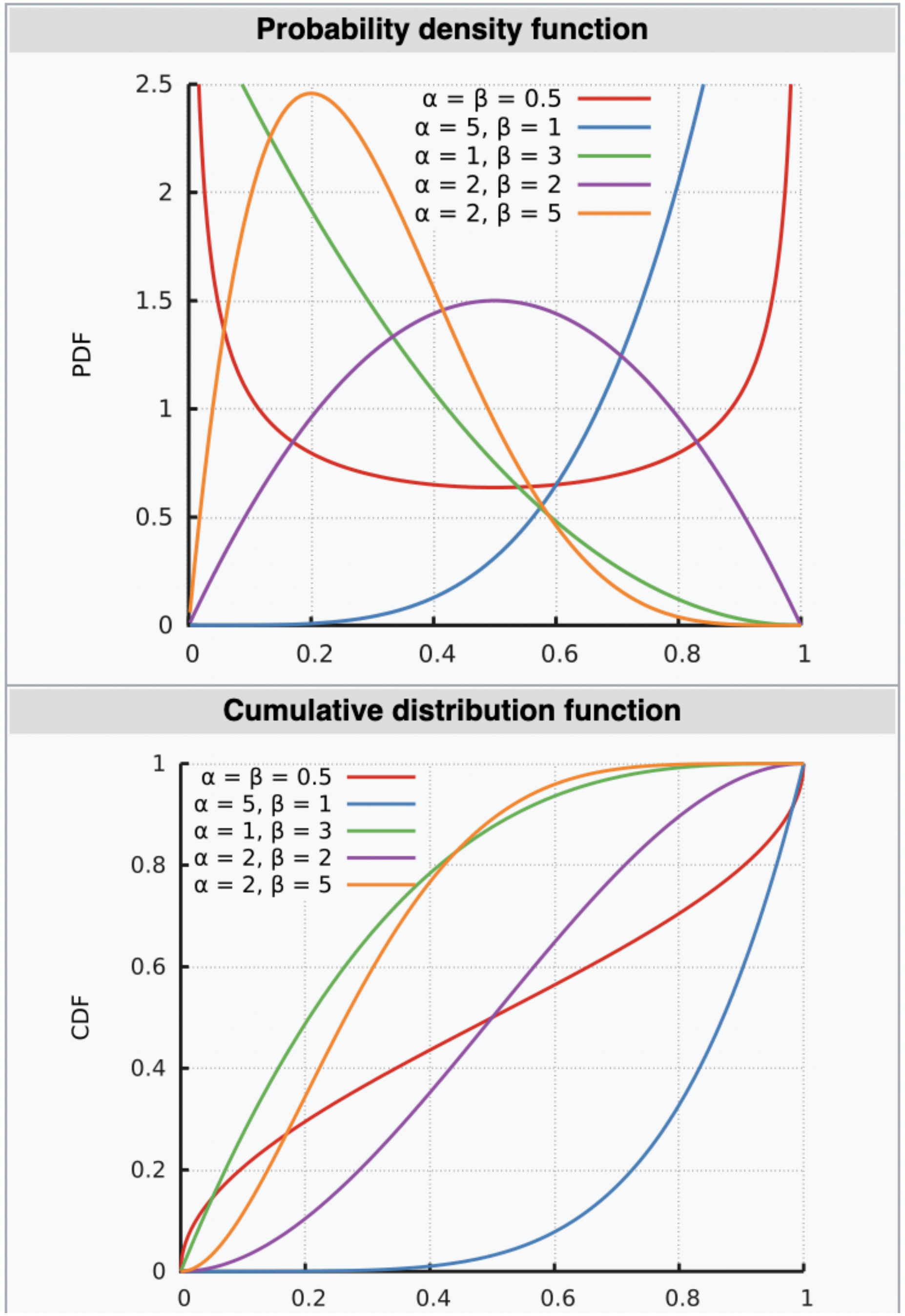
Beta

Bayesian Conjugate Prior of Binomial

a: number of success (head)

b: number of failures (tail)

Beta



Parameters	$\alpha > 0$ shape (real) $\beta > 0$ shape (real)
Support	$x \in [0,1]$ or $x \in (0,1)$
PDF	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$ where $\mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and Γ is the Gamma function.
CDF	$I_x(\alpha,\beta)$ (the regularized incomplete beta function)
Mean	$\mathrm{E}[X] = \frac{\alpha}{\alpha+\beta}$ $\mathrm{E}[\ln X] = \psi(\alpha) - \psi(\alpha+\beta)$ $\mathrm{E}[X \ln X] = \frac{\alpha}{\alpha+\beta}[\psi(\alpha+1) - \psi(\alpha+\beta+1)]$ (see section: Geometric mean) where ψ is the digamma function
Median	$I^{[-1]}_{\frac{1}{2}}(\alpha,\beta)$ (in general) $\approx \frac{\alpha-\frac{1}{3}}{\alpha+\beta-\frac{2}{3}}$ for $\alpha,\beta > 1$
Mode	$\frac{\alpha-1}{\alpha+\beta-2}$ for $\alpha, \beta > 1$ any value in $(0,1)$ for $\alpha, \beta = 1$ {0, 1} (bimodal) for $\alpha, \beta < 1$ 0 for $\alpha \le 1, \beta > 1$ 1 for $\alpha > 1, \beta \le 1$
Variance	$\mathrm{var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ $\mathrm{var}[\ln X] = \psi_1(\alpha) - \psi_1(\alpha+\beta)$ (see trigamma function and see section: Geometric variance)
Skewness	$\frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}}$

coding time!

https://github.com/fedhere/DSPS_FBianco/blob/main/CodeExamples/stats/coin_tosses.ipynb

coding time!

https://github.com/fedhere/DSPS_FBianco/blob/main/labs/Instructions_Bayesian_posteriors.ipynb

Alex Ji
@alexanderpji

"Everyone should use Bayesian statistics, it's so intuitive and natural!"

Probability distributions

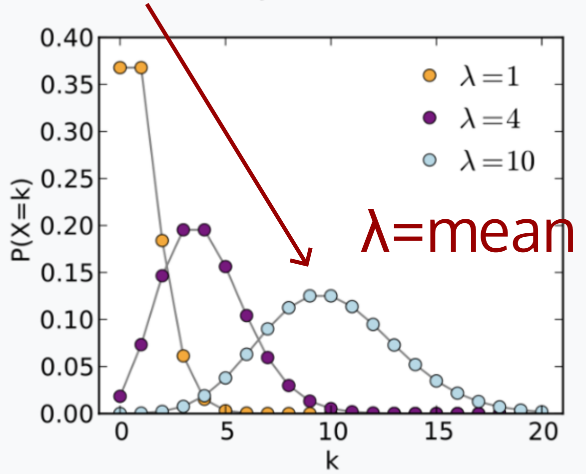
Poisson

Shut noise/count noise

The innate noise in natural steady state processes (star flux, rain drops...)

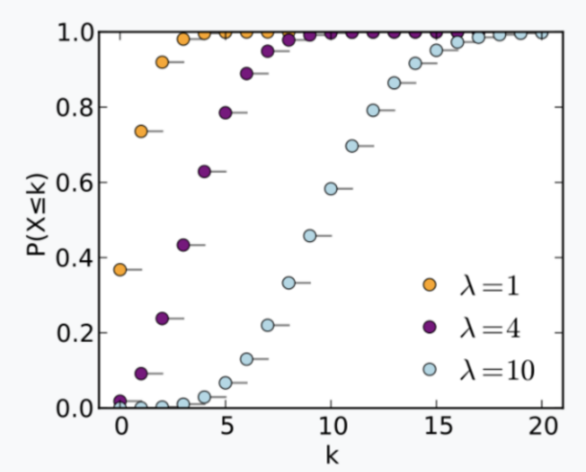
Poisson

Probability mass function



The horizontal axis is the index *k*, the number of occurrences. *λ* is the expected number of occurrences, which need not be an integer. The vertical axis is the probability of *k* occurrences given *λ*. The function is defined only at integer values of *k*. The connecting lines are only guides for the eye.

Cumulative distribution function



The horizontal axis is the index *k*, the number of occurrences. The CDF is discontinuous at the integers of *k* and flat everywhere else because a variable that is Poisson distributed takes on only integer values.

Notation	$\mathrm{Pois}(\lambda)$
Parameters	$\lambda > 0$, (real) — rate
Support	$k \in \{0, 1, 2, \ldots\}$
pmf	$\frac{\lambda^k e^{-\lambda}}{k!}$
CDF	$\frac{\Gamma(\lfloor k+1 \rfloor, \lambda)}{\lfloor k \rfloor!}$, or $e^{-\lambda} \sum_{i=0}^{\lfloor k \rfloor} \frac{\lambda^i}{i!}$, or $Q(\lfloor k+1 \rfloor, \lambda)$ (for $k \geq 0$, where $\Gamma(x, y)$ is the upper incomplete gamma function, $\lfloor k \rfloor$ is the floor function, and Q is the regularized gamma function)
Mean	λ
Median	$\approx \lfloor \lambda + 1/3 - 0.02/\lambda \rfloor$
Mode	$\lceil \lambda \rceil - 1, \lfloor \lambda \rfloor$
Variance	λ
Skewness	$\lambda^{-1/2}$
Ex. kurtosis	λ^{-1}
Entropy	$\lambda[1 - \log(\lambda)] + e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k \log(k!)}{k!}$ (for large λ) $\frac{1}{2}\log(2\pi e \lambda) - \frac{1}{12\lambda} - \frac{1}{24\lambda^2} - \frac{19}{360\lambda^3} + O\left(\frac{1}{\lambda^4}\right)$
MGF	$\exp(\lambda(e^t - 1))$
CF	$\exp(\lambda(e^{it} - 1))$
PGF	$\exp(\lambda(z - 1))$
Fisher information	$\frac{1}{\lambda}$

Probability distributions

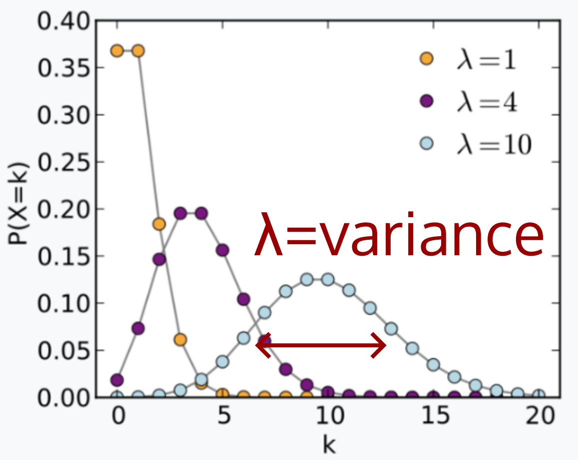
Poisson

Shut noise/count noise

The innate noise in natural steady state processes (star flux, rain drops...)

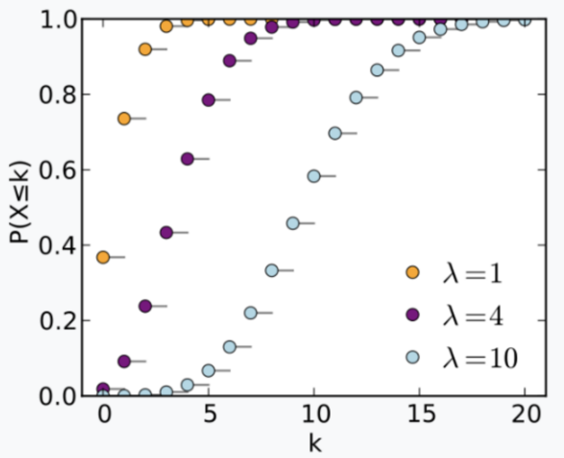
Poisson

Probability mass function



The horizontal axis is the index *k*, the number of occurrences. *λ* is the expected number of occurrences, which need not be an integer. The vertical axis is the probability of *k* occurrences given *λ*. The function is defined only at integer values of *k*. The connecting lines are only guides for the eye.

Cumulative distribution function



The horizontal axis is the index *k*, the number of occurrences. The CDF is discontinuous at the integers of *k* and flat everywhere else because a variable that is Poisson distributed takes on only integer values.

Notation	$\mathrm{Pois}(\lambda)$
Parameters	$\lambda > 0$, (real) — rate
Support	$k \in \{0, 1, 2, \ldots\}$
pmf	$\frac{\lambda^k e^{-\lambda}}{k!}$
CDF	$\frac{\Gamma(\lfloor k+1 \rfloor, \lambda)}{\lfloor k \rfloor !}$, or $e^{-\lambda} \sum_{i=0}^{\lfloor k \rfloor} \frac{\lambda^i}{i!}$, or $Q(\lfloor k+1 \rfloor, \lambda)$ (for $k \geq 0$, where $\Gamma(x, y)$ is the upper incomplete gamma function, $\lfloor k \rfloor$ is the floor function, and Q is the regularized gamma function)
Mean	λ
Median	$\approx \lfloor \lambda + 1/3 - 0.02/\lambda \rfloor$
Mode	$\lceil \lambda \rceil - 1, \lfloor \lambda \rfloor$
Variance	λ
Skewness	$\lambda^{-1/2}$
Ex. kurtosis	λ^{-1}
Entropy	$\lambda[1 - \log(\lambda)] + e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k \log(k!)}{k!}$ (for large λ) $\frac{1}{2}\log(2\pi e \lambda) - \frac{1}{12\lambda} - \frac{1}{24\lambda^2} - \frac{19}{360\lambda^3} + O\left(\frac{1}{\lambda^4}\right)$
MGF	$\exp(\lambda(e^t - 1))$
CF	$\exp(\lambda(e^{it} - 1))$
PGF	$\exp(\lambda(z - 1))$
Fisher information	$\frac{1}{\lambda}$

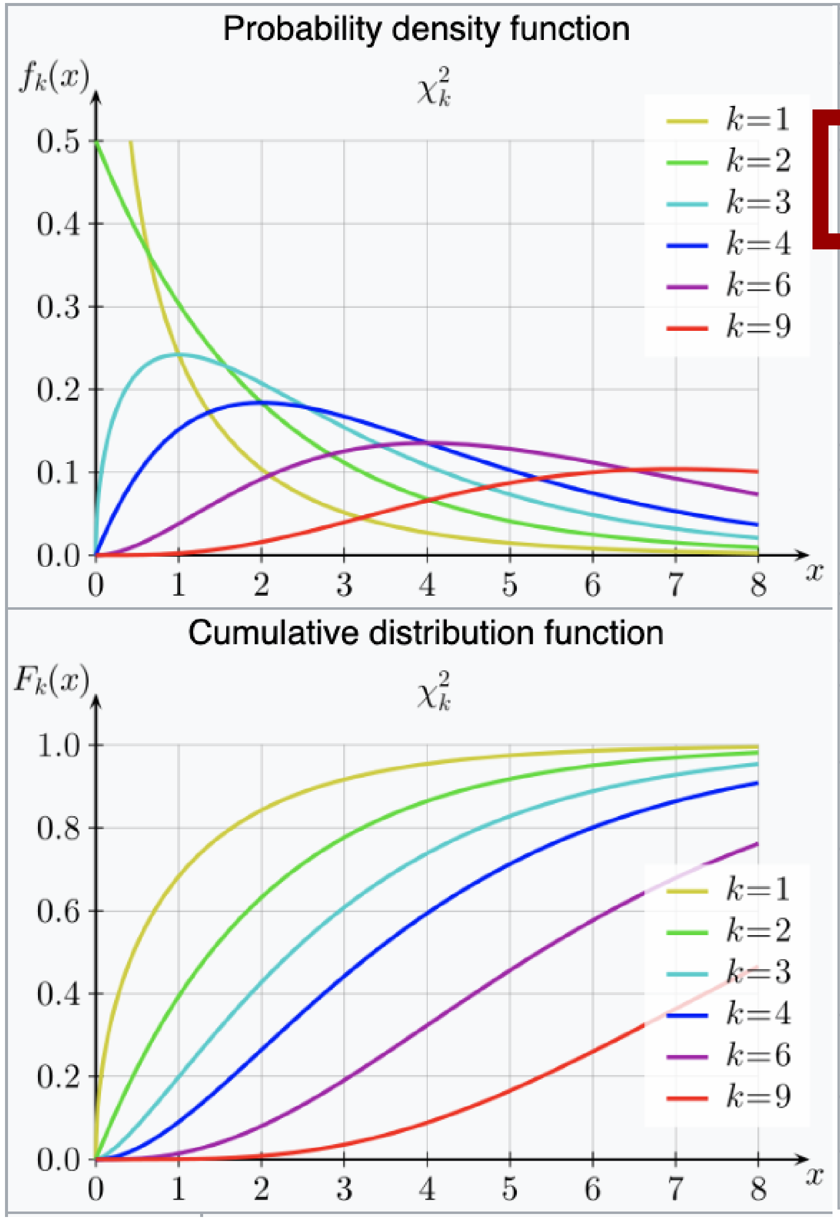
Probability distributions

Chi-square (χ2)

turns out its extremely common

many pivotal quantities follow this distribution and thus many tests are based on this

chi-square



Notation	$\chi^2(k)$ or χ^2_k
Parameters	$k \in \mathbb{N}_{>0}$ (known as "degrees of freedom")
Support	$x \in (0, +\infty)$ if $k = 1$, otherwise $x \in [0, +\infty)$
PDF	$\frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$
CDF	$\frac{1}{\Gamma(k/2)} \gamma\left(\frac{k}{2}, \frac{x}{2}\right)$
Mean	k
Median	$\approx k\left(1 - \frac{2}{9k}\right)^3$
Mode	$\max(k-2, 0)$
Variance	$2k$
Skewness	$\sqrt{8/k}$
Ex. kurtosis	$\frac{12}{k}$
Entropy	$\frac{k}{2} + \ln(2\Gamma(\frac{k}{2})) + (1 - \frac{k}{2})\psi(\frac{k}{2})$
MGF	$(1-2t)^{-k/2}$ for $t < \frac{1}{2}$
CF	$(1-2it)^{-k/2}$ [1]
PGF	$(1 - 2\ln t)^{-k/2}$ for $0 < t < \sqrt{e}$

Basic Probability *Frequentist* interpretation

fraction of times something happens

$\Updownarrow$

probability of it happening

Basic Probability Bayesian interpretation

represents a level of certainty relating to a potential outcome or idea:

if I believe the coin is unfair (tricked) then even if I get a head and a tail I will still believe I am more likely to get heads than tails

$$P(A|B)P(B) = P(B|A)P(A)$$

4 Bayes theorem

Bayes theorem

$$P(A|B)P(B) = P(B|A)P(A)$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes theorem

$$P(\text{model}|\text{data})P(\text{data}) = P(\text{data}|\text{model})P(\text{model})$$

$$P(\text{model}|\text{data}) = \frac{P(\text{data}|\text{model})P(\text{model})}{P(\text{data})}$$

Bayes theorem

$$\cancel{P(\theta|\mathrm{D}) = P(\mathrm{D}|\theta)}$$

θ model parameters (fair)

D data (number of tails)

implicity in the frequentist approach

What is the probability that the coin is fair?
its the same as the probability of getting
that many tails if I assume the coin is fair

Bayes theorem

$$\underbrace{P(\theta|\mathrm{D})}_{\text{posterior}} = \frac{\overbrace{P(\mathrm{D}|\theta)}^{\text{likelihood}}\ \overbrace{P(\theta)}^{\text{prior}}}{\underbrace{P(\mathrm{D})}_{\text{evidence}}}$$

θ model parameters

D data

E.g. .... not likely to be Aliens after all...

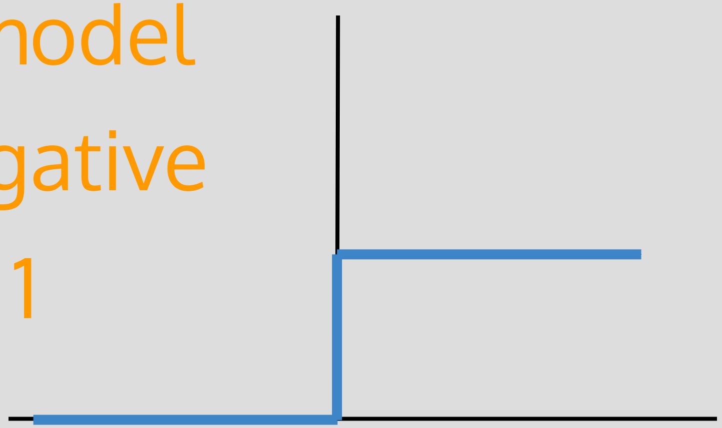
Bayes theorem

$$P(\theta|\mathrm{D}) = \frac{P(\mathrm{D}|\theta)P(\theta)}{P(\mathrm{D})}$$

prior: constraints on the model

e.g. flux is never negative

P(f<0) = 0 P(f>=0) = 1



θ model parameters

D data

Bayes theorem

$$P(\theta|\mathrm{D}) = \frac{P(\mathrm{D}|\theta)P(\theta)}{P(\mathrm{D})}$$

θ model parameters

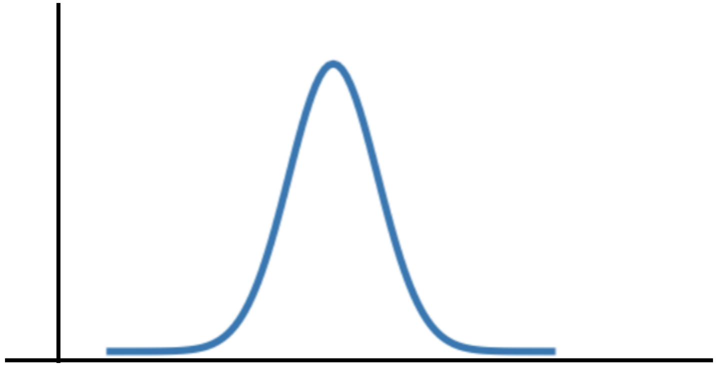
D data

prior: constraints on the model

people's weight <1000lb

& people's weight >0lb

P(w) ~ $\mathcal{N}$(105lb, 90lb)



Bayes theorem

$$P(\theta|\mathrm{D}) = \frac{P(\mathrm{D}|\theta)P(\theta)}{P(\mathrm{D})}$$

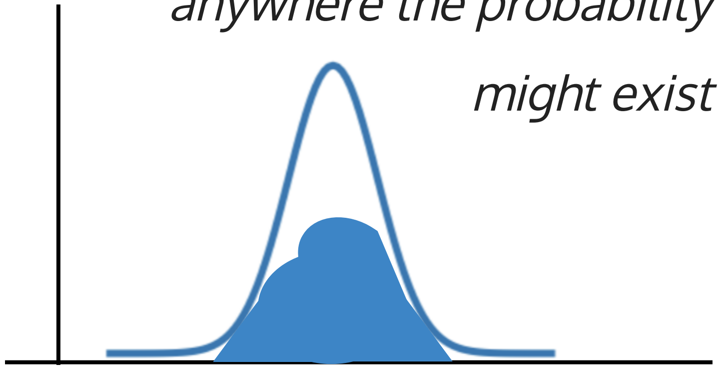
prior: constraints on the model

people's weight <1000lb

& people's weight >0lb

P(w) ~ $\mathcal{N}$(105lb, 90lb)

the prior should not be 0 anywhere the probability might exist



θ model parameters

D data

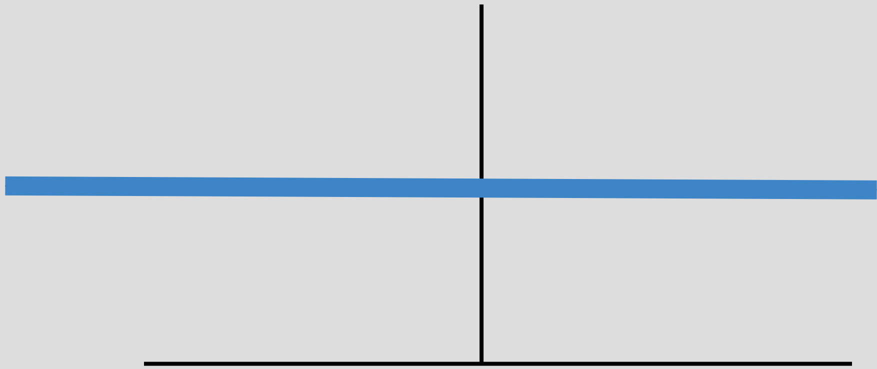
Bayes theorem

$$P(\theta|\mathrm{D}) = \frac{P(\mathrm{D}|\theta)P(\theta)}{P(\mathrm{D})}$$

prior: "uninformative prior"

θ model parameters

D data



Bayes theorem

$$P(\theta|\mathrm{D}) = \frac{P(\mathrm{D}|\theta)P(\theta)}{P(\mathrm{D})}$$

evidence

????

θ model parameters

D data

it does not matter if I want to use this for model comparison

Bayes theorem

$$P(\theta_1|\mathrm{D}) = \frac{P(\mathrm{D}|\theta_1)P(\theta_1)}{\cancel{P(\mathrm{D})}} \qquad P(\theta_2|\mathrm{D}) = \frac{P(\mathrm{D}|\theta_2)P(\theta_2)}{\cancel{P(\mathrm{D})}}$$

$$P(\theta|\mathrm{D}) \propto P(\mathrm{D}|\theta)P(\theta)$$

which has the highest posterior probability?

θ model parameters

D data

it does not matter if I want to use this for model comparison

Bayes theorem

$$P(\theta|\mathrm{D}) = \frac{P(\mathrm{D}|\theta)P(\theta)}{P(\mathrm{D})}$$

posterior: joint probability distributin of a parameter set (θ, e.g. (m, b)) condition upon some data D and a model hypothesys f

prior: "intellectual" knowledge about the model parameters condition on a model hypothesys f. *This should come from domain knowledge or knowledge of data that is not the dataset under examination*

evidence: marginal likelihood of data under the model

Note: in practive all of these quantities are conditioned on the shape of the model:

$$P(D|f) = \int_{-\infty}^{\infty} P(D|\theta, f)P(\theta|f)d\theta$$

this is a model fitting, not a model selection methodology...

coding time!

https://github.com/fedhere/DSPS_FBianco/blob/main/CodeExamples/stats/coin_tosses_animation.ipynb

lab time!

https://github.com/fedhere/DSPS_FBianco/blob/main/labs/Instructions_Bayesian_posteriors.ipynb

http://bit.ly/3Rpl9oy

Central Limit Theorem

Laplace (1700s) but also: Poisson, Bessel, Dirichlet, Cauchy, Ellis

Let $x_1 \ldots x_N$ be an N-elements sample from a population whose distribution has

mean **μ** and standard deviation **σ**

In the limit of N -> infinity

the sample mean $\bar{x}$ approaches a Normal (Gaussian) distribution with mean μ and standard deviation σ

regardless of the distribution of X

$$\bar{x} \sim N\left(\mu, \sigma/\sqrt{N}\right)$$

Central Limit Theorem

CAVEATS:

For the theorem to hold the sample size has to be large enough (N~30 is usually considered sufficient)

The underlying distribution should be unbound (otherwise the tails of the Gaussian would be truncated)

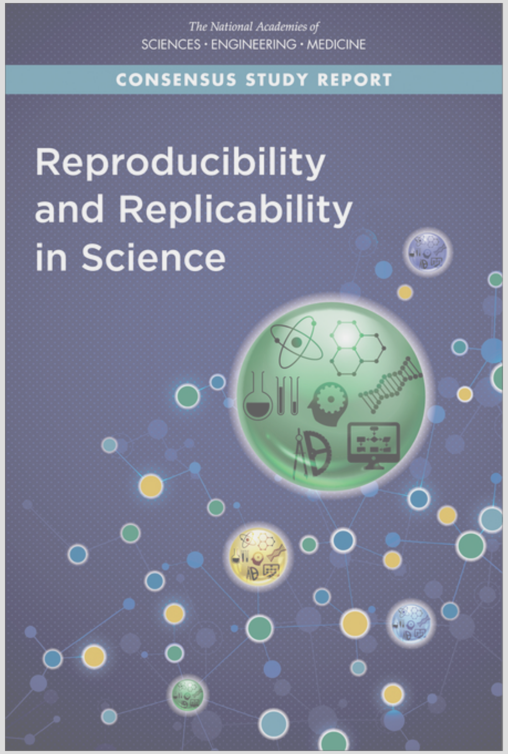
$$\bar{x} \ \sim \ N\left(\mu, \sigma/\sqrt{N}\right)$$

5

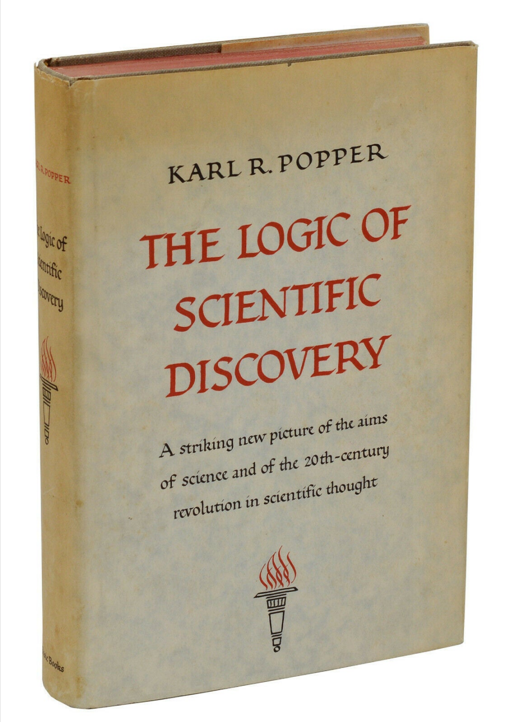
the principle of Falsifiability

3 General principles of "good" science

Reproducibility



Falisifiability



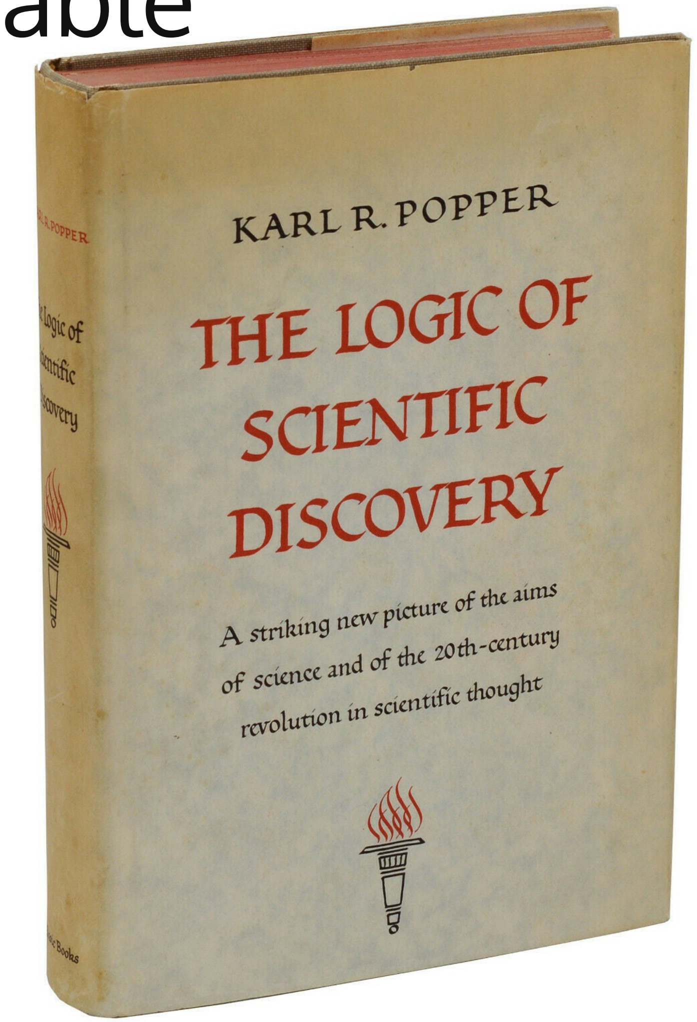
Parsimony

the *demarcation* problem:

science hypotheses needs to be falsifiable

My proposal is based upon an *asymmetry* between verifiability and **falsifiability**; an asymmetry which results from the logical form of universal statements. For these are never derivable from singular statements, but can be contradicted by singular statements.

—Karl Popper, *The Logic of Scientific Discovery*



the *demarcation* problem:

science hypotheses needs to be falsifiable

My proposal is based upon an *asymmetry* between verifiability and **falsifiability**; an asymmetry which results from the logical form of universal statements. For these are never derivable from singular statements, but can be contradicted by singular statements.

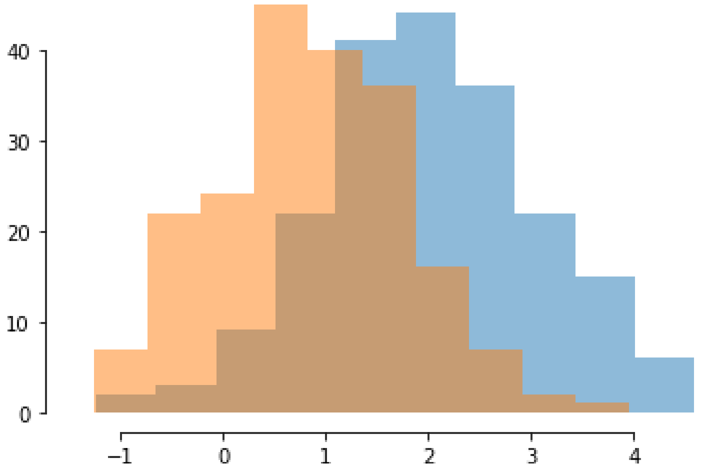
—Karl Popper, *The Logic of Scientific Discovery*

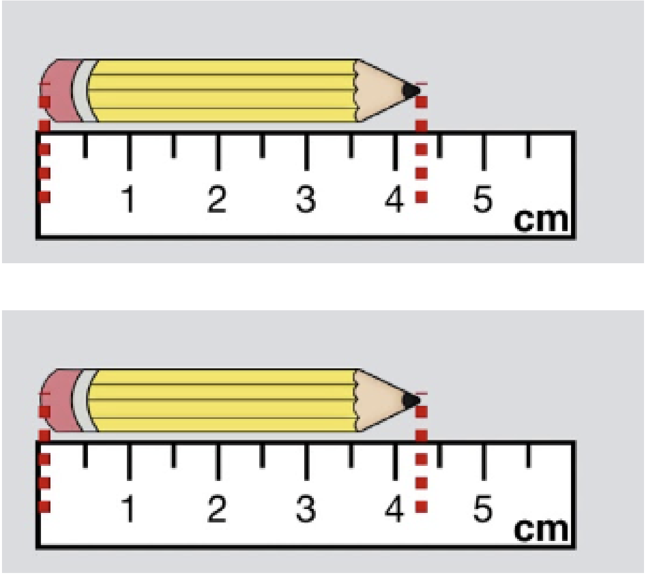
I need to see only 1 black swan to tell that the statement that all swans are white is not true. But even if I dont see a black one it does not mean all swans are white

the *demarcation* problem:

science hypotheses needs to be falsifiable

But what happens when I have distributions of measurements?





Could be a distribution of measurements

Could be an intrinsically stochastic phenomenon

Beyond any reasonable doubt

same concept guides prosecutorial justice

guilty beyond reasonable doubt

in a probabilistic sense, all hypotheses we make are possible

We will reject a hypothesis if its probability is lower than a predefined threshold

the *demarcation* problem in *Bayesian* context

The probability that a belief is true given **new evidence** equals the probability that the belief is true **regardless of that evidence** times the **probability that the evidence is true given that the belief is true** divided by the **probability that the evidence is true regardless** of whether the belief is true.

- *Thomas Bayes Essay towards solving a Problem in the Doctrine of Chances (1763)*

$$p(M|D) = \frac{P(M)\,P(D|M)}{P(D)}$$

7

the scientific method in a probabilistic context FIRST PASS

$$p(\text{physics} \mid \text{data})$$

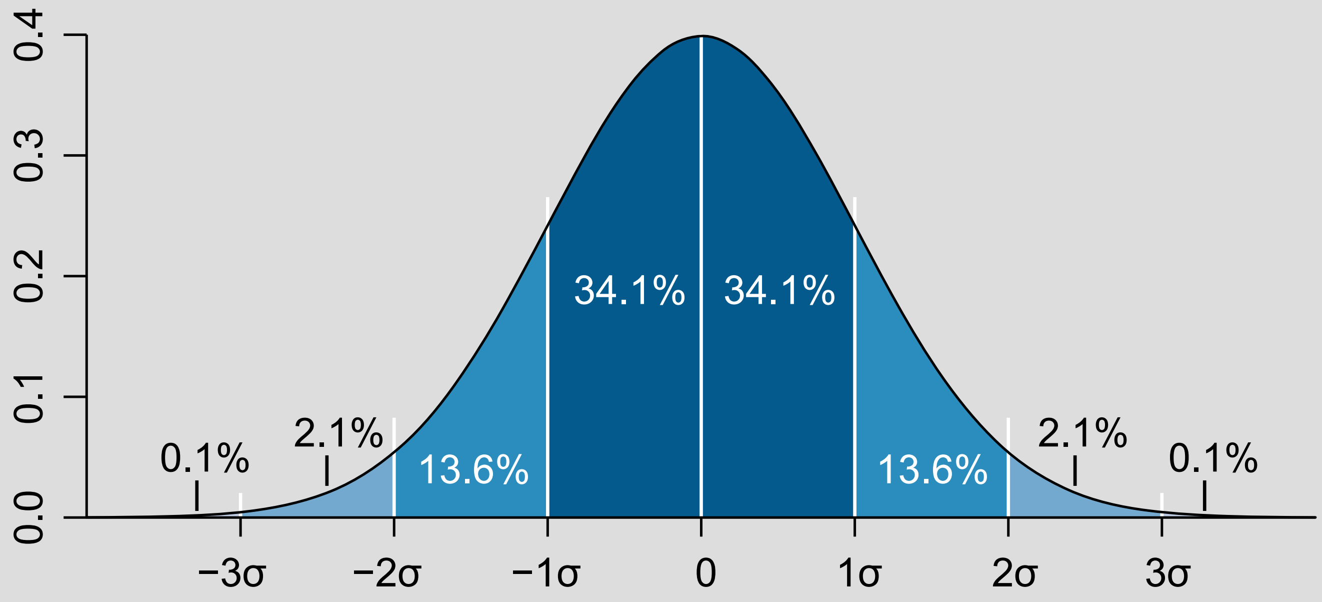
Bayesian Inference

Forward Modeling

Frequentist approach
(NHRT)

p(physics | data)

Null
Hypothesis
Rejection
Testing



p(physics | data)

https://speakerdeck.com/dfm/emcee-odi

model → prediction

data → does not falsify → model still holds

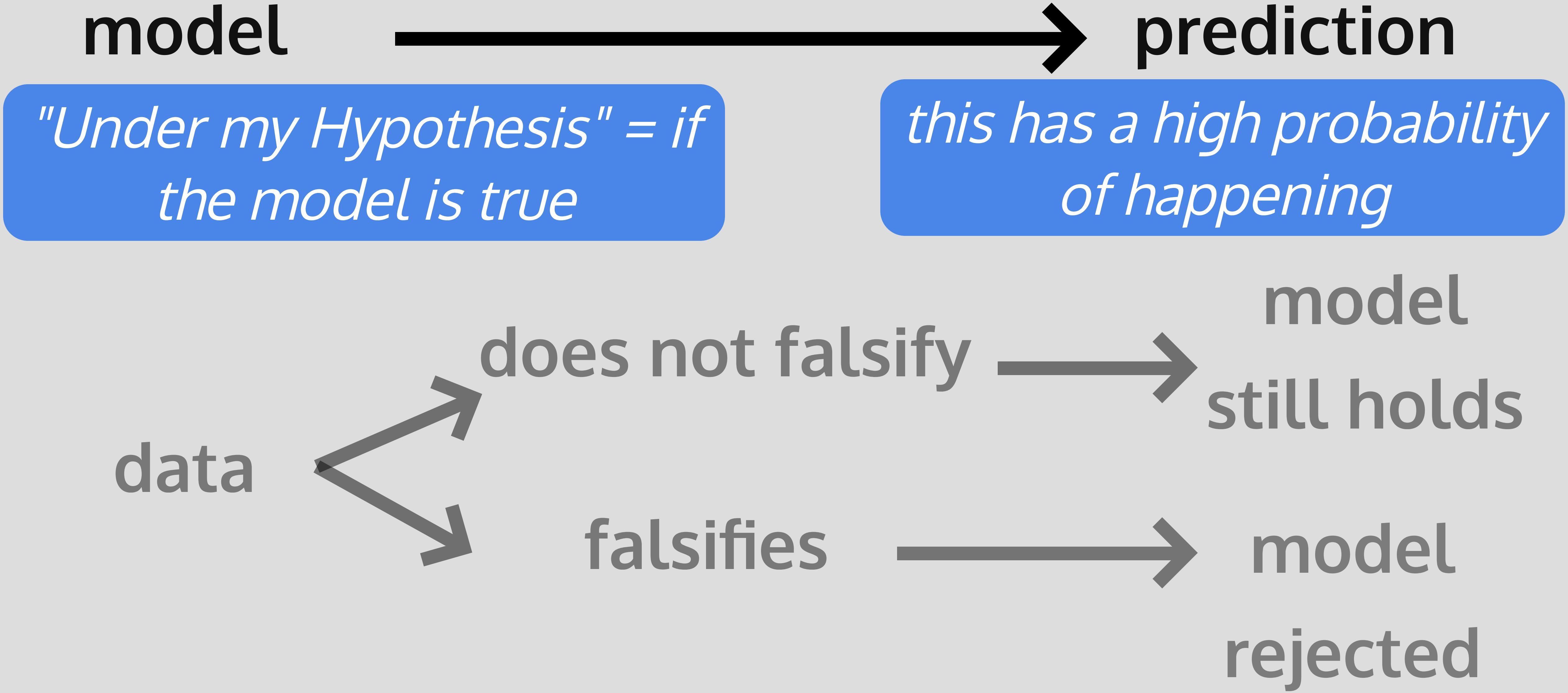
data → falsifies → model rejected

model → **prediction**

"Under my Hypothesis" = if the model is true

data → does not falsify → model still holds

data → falsifies → model rejected


model
prediction
"Under my Hypothesis" = if the model is true
this has a high probability of happening
data
does not falsify
model still holds
falsifies
model rejected

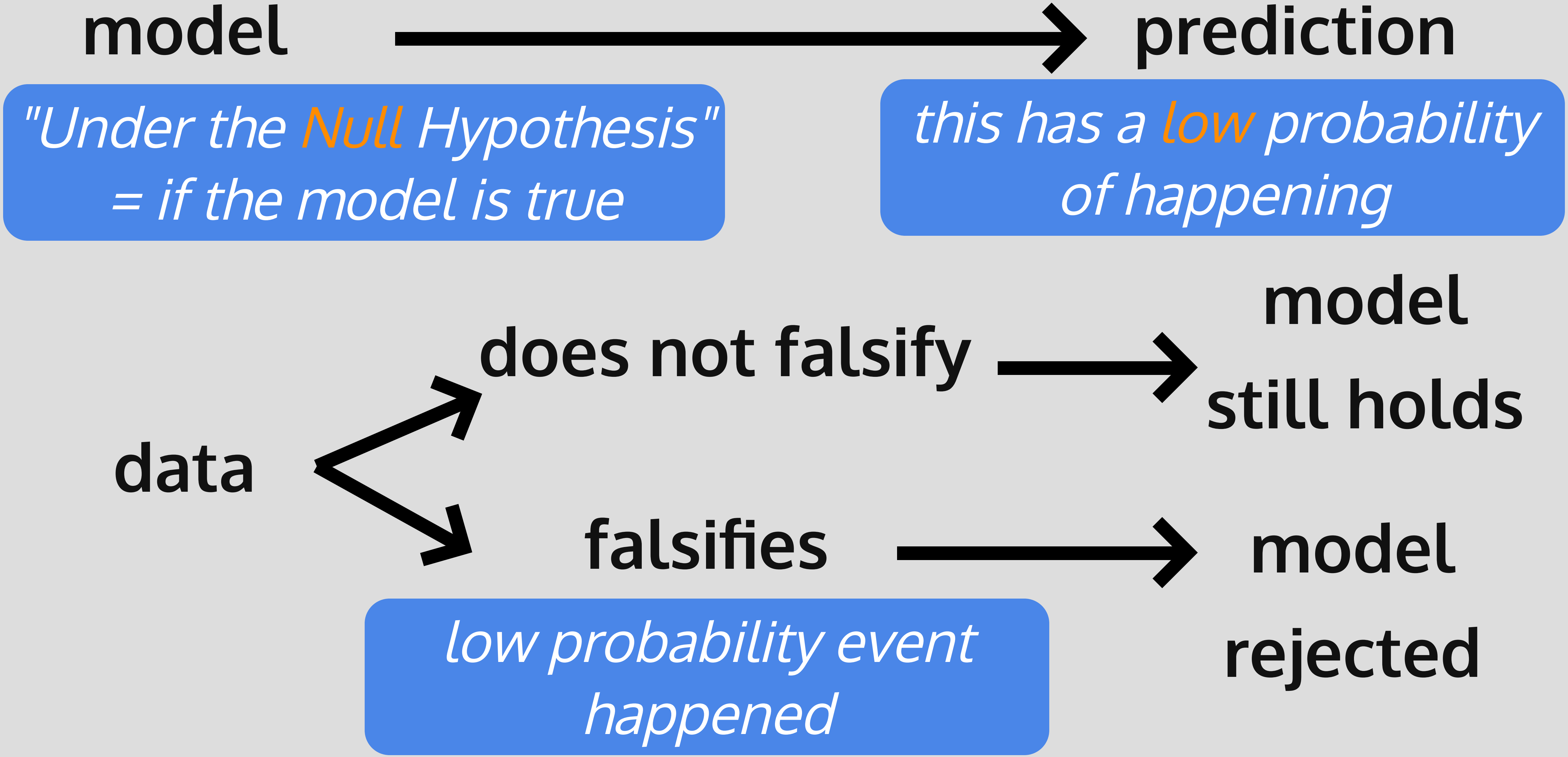
model → prediction

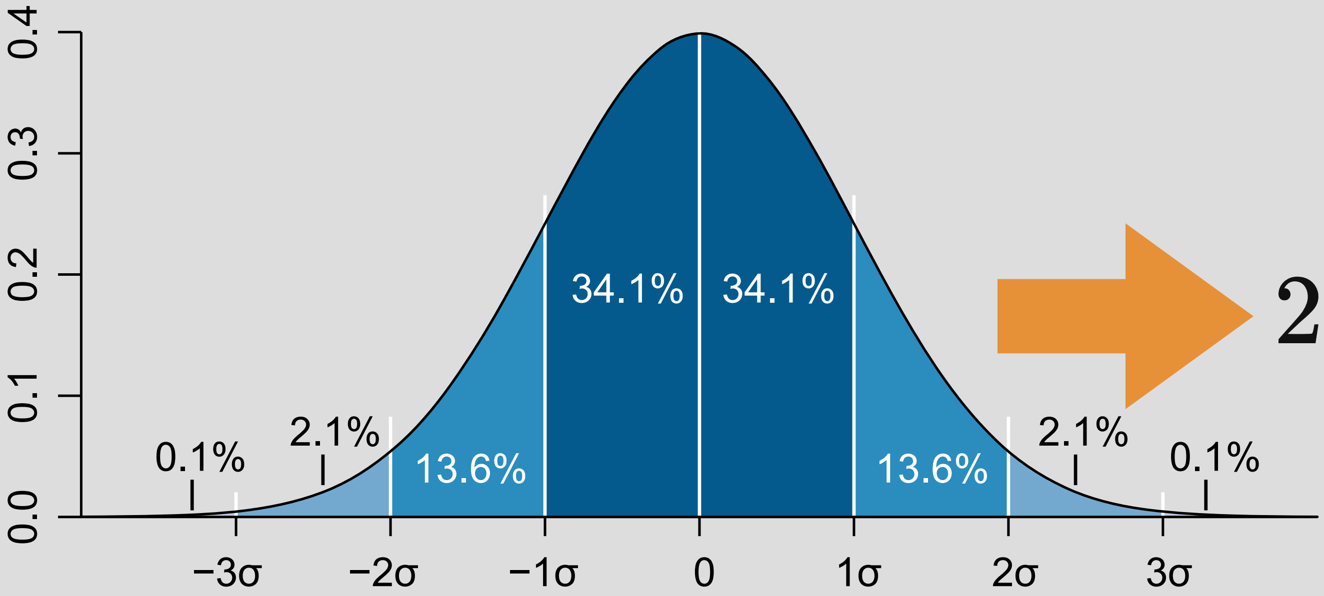
"Under the Null Hypothesis"
= if the model is false

this has a low probability
of happening

data → does not falsify → model still holds

data → falsifies → model rejected


model
prediction
"Under the Null Hypothesis"
= if the model is true
this has a low probability
of happening
data
does not falsify
model
still holds
falsifies
model
rejected
low probability event
happened



2σ

rejected at 95%

0.05 p-value

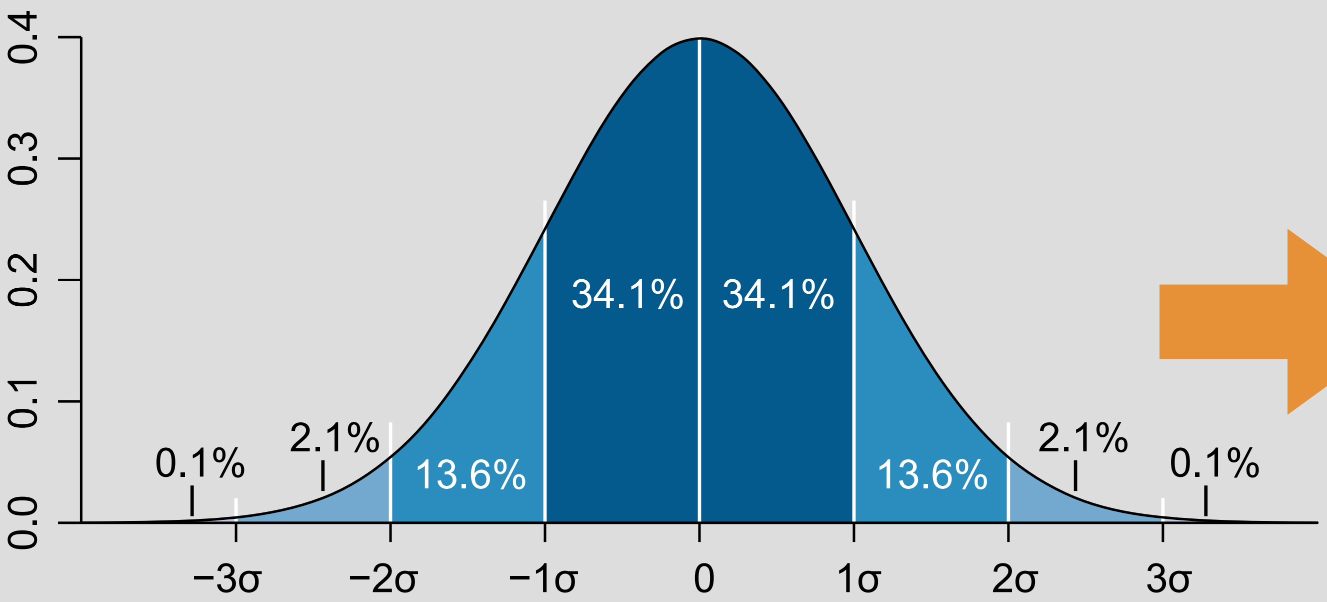
5% confidence

data

does not falsify → model still holds

falsifies → model rejected

low probability event happened



rejected at 99.7%

3σ 0.003 p-value

0.3% confidence

data

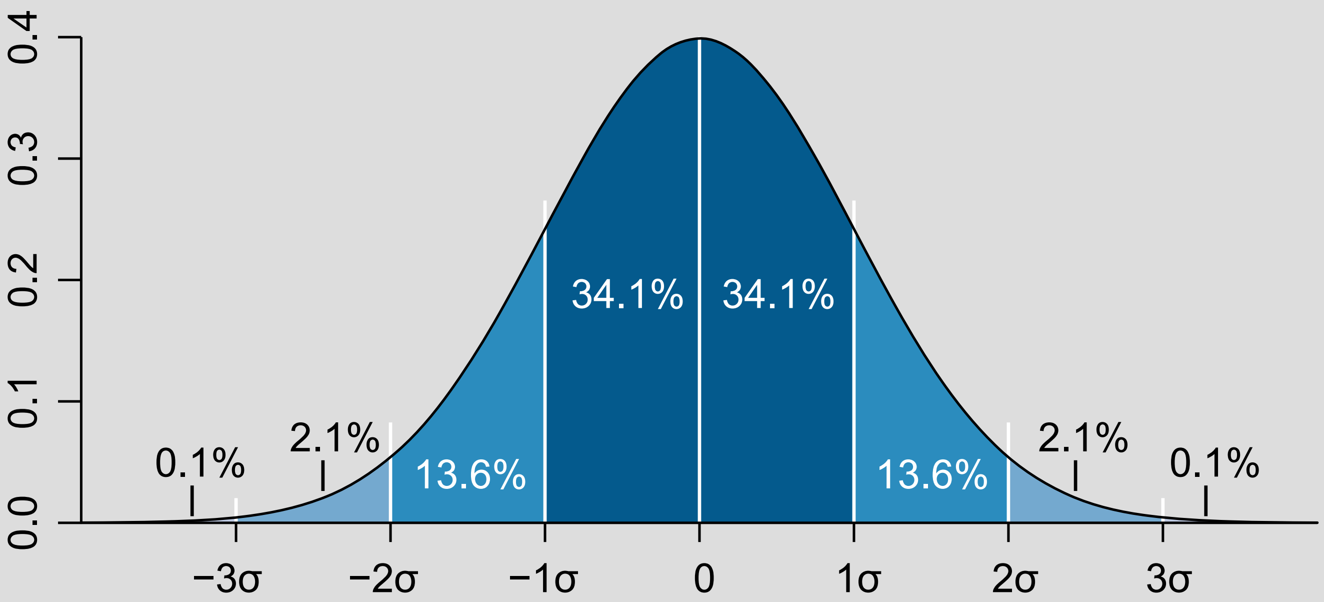
does not falsify

model still holds

falsifies

model rejected

low probability event happened



rejected at 99.99...%

3-e7 p-value

3-e5% confidence

5σ

model still holds

does not falsify

data

falsifies

model rejected

low probability event happened

Null
Hypothesis
Rejection
Testing

1

formulate your prediction

Null Hypothesis

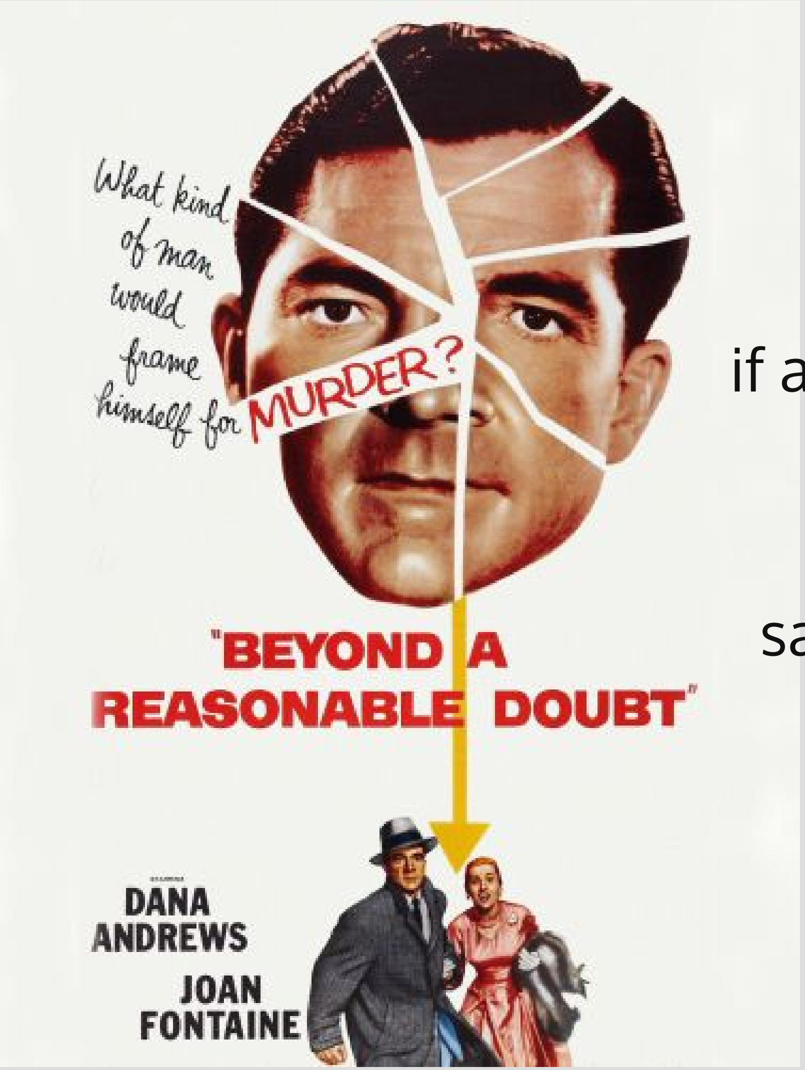
Null
Hypothesis
Rejection
Testing

2

identify all alternative outcomes

Alternative Hypothesis

Null
Hypothesis
Rejection
Testing



if all alternatives to our model are ruled out,
then our model must hold

same concept guides prosecutorial justice

guilty beyond reasonable doubt

2

identify all alternative outcomes

Alternative Hypothesis

Null
Hypothesis
Rejection
Testing

3

set confidence threshold

2σ	confidence level
0.05	p-value
95%	α threshold

Null

Hypothesis

Rejection

Testing

pivotal quantities

quantities that under the Null Hypotheis follow a known distribution

4

$$p(\text{pivotal quantity}|NH) \sim p(NH|D)$$

Null
Hypothesis
Rejection
Testing

pivotal quantities

quantities that under the Null Hypotheis follow a known distribution

also called "statistics"

e.g.: *χ2 statistics:* difference between prediction and reality squared

Z statistics: difference between means

K-S statistics: maximum distance of cumulative distributions.

4

Null
Hypothesis
Rejection
Testing

pivotal quantities

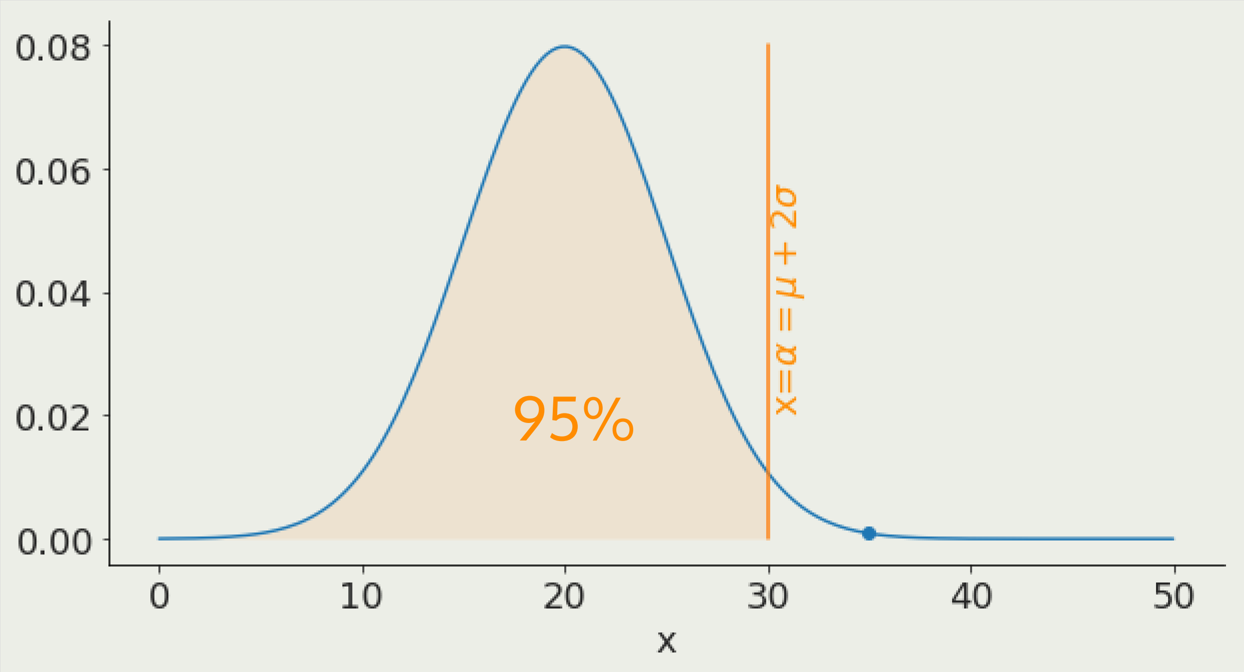
5

calculate it!

Null
Hypothesis
Rejection
Testing

what is α?

α is the x value corresponding to a chosen threshold



6

test data against
alternative outcomes

Null

Hypothesis

Rejection

Testing

$$p(NH|D) < \alpha$$

prediction is unlikely

Null rejected

Alternative holds

☺

6

test data against

alternative outcomes

Null

Hypothesis

Rejection

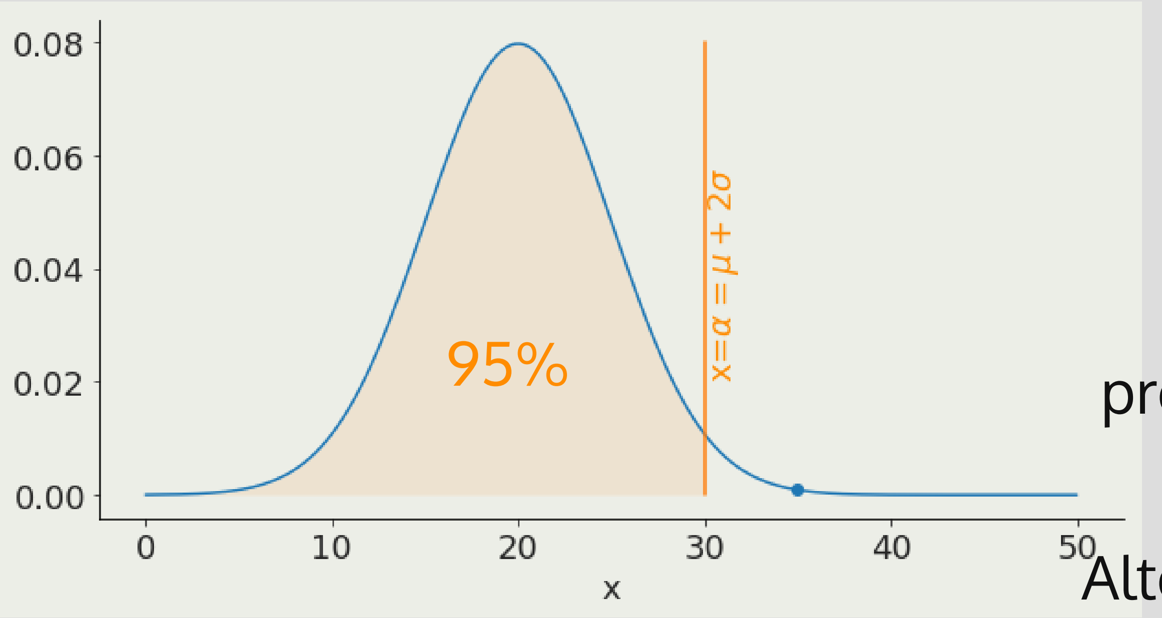
Testing

$$p(NH|D) < \alpha$$

prediction is unlikely

Null rejected

Alternative holds



prediction is likely

Null holds

Alternative rejected

6

test data against
alternative outcomes

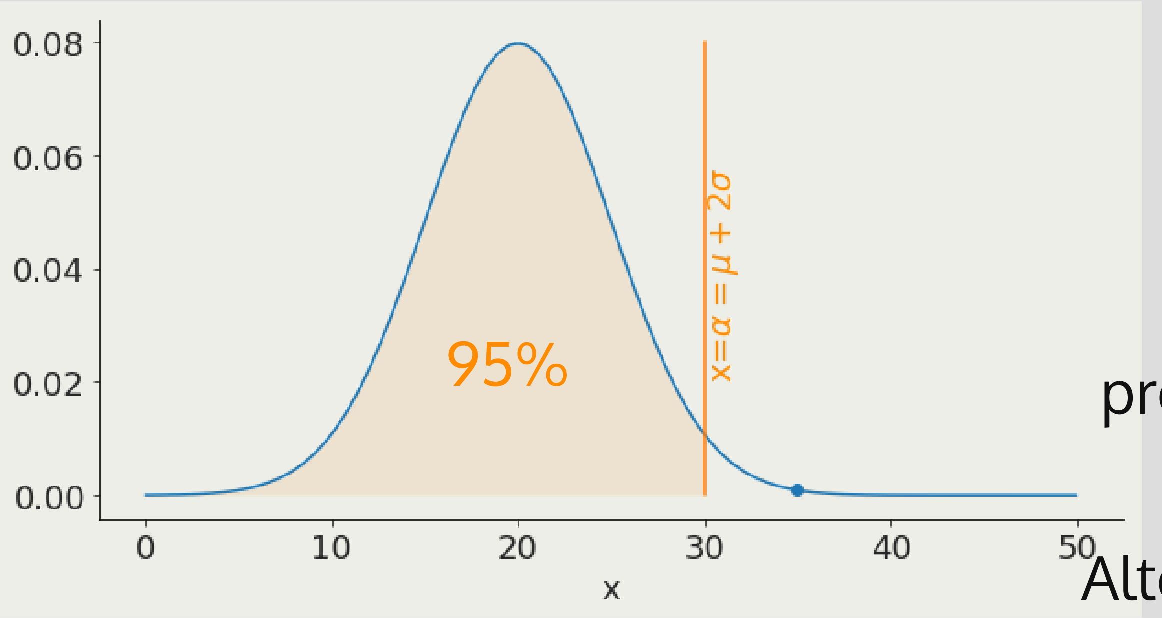
$$p(NH|D) \geq \alpha$$

Null
Hypothesis
Rejection
Testing

6

$$p(NH|D) < \alpha$$

prediction is unlikely
Null rejected
Alternative holds



prediction is likely
Null holds
Alternative rejected

test data against
alternative outcomes

$$p(NH|D) \geq \alpha$$

Null
Hypothesis
Rejection
Testing

$$p(NH|D) < \alpha$$

prediction is unlikely
Alternative rejected
Null holds

6

test data against
alternative outcomes

$$p(NH|D) \geq \alpha$$

prediction is likely
Alternative holds
Null rejected

K-S test

Kolmogorof-Smirnoff :

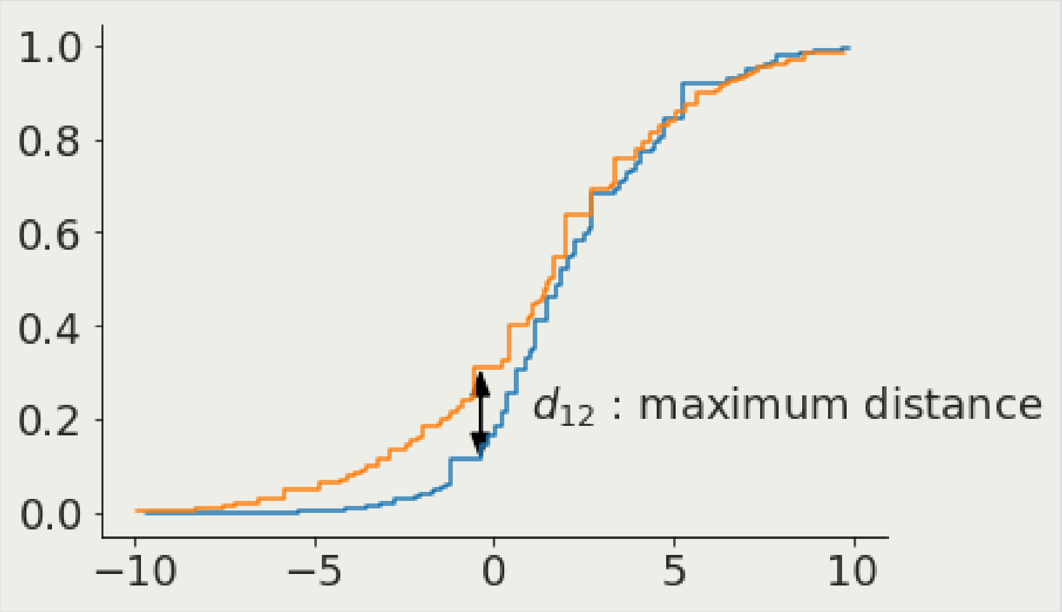
do two samples come from the same parent distribution?

pivotal quantity

$$d_{12} \equiv max_x \left|C_1(x) - C_2(x)\right|$$

Cumulative distribution 1

Cumulative distribution 2



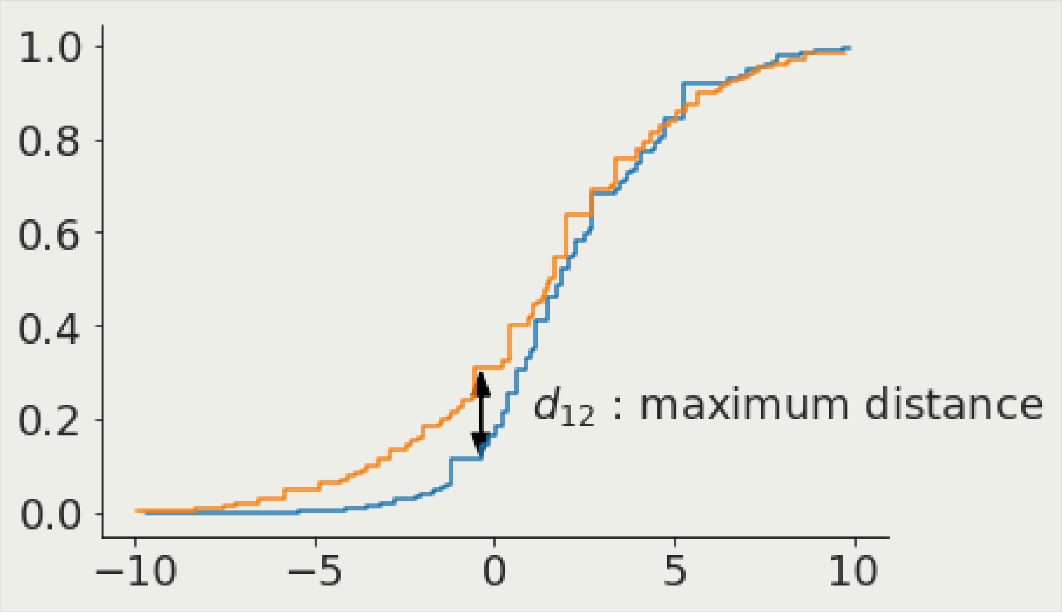
$$P(d > observed) = 2\sum_{j=1}^{\infty}(-1)^{j-1}e^{-2j^2x^2}\sqrt{\frac{N_1N_2}{N_1+N_2}D}$$

K-S test

Kolmogorof-Smirnoff :

do two samples come from the same parent distribution?

pivotal quantity



$$d_{12} \equiv max_x \left|C_1(x) - C_2(x)\right|$$

Cumulative distribution 1

Cumulative distribution 2

$$P(d > observed) =$$

```
sp.stats.ks_2samp(x, y)
```

executed in 7ms, finished 14:45:10 2019-09-09

```
Ks_2sampResult(statistic=0.4, pvalue=0.3128526760169558)
```

formulate the Null as the comprehensive opposite of your theory

model → prediction

"Under the Null Hypothesis" = if the proposed model is false

this has a low probability of happening

data → does not falsify alternative → everything but model is rejected

data → falsifies alternative → model holds

low probability event happened

Key Slide

1 formulate your prediction (NH)

2 identify all alternative outcomes (AH)

3 set confidence threshold (*p*-value)

4 find a measurable quantity which under the Null has a known distribution (pivotal quantity)

5 calculate the pivotal quantity

6 calculate probability of value obtained for the pivotal quantity under the Null

if probability $<$ *p*-value : reject Null

Key Slide

the *demarcation* problem in *Bayesian* context

The probability that a belief is true given **new evidence** equals the probability that the belief is true **regardless of that evidence** times the **probability that the evidence is true given that the belief is true** divided by the **probability that the evidence is true regardless** of whether the belief is true.

- *Thomas Bayes* *Essay towards solving a Problem in the Doctrine of Chances* (1763)

the *demarcation* problem in *Bayesian* context

The probability that a belief is true given **new evidence** equals the probability that the belief is true **regardless of that evidence** times the **probability that the evidence is true given that the belief is true** divided by the **probability that the evidence is true regardless** of whether the belief is true.

- *Thomas Bayes* *Essay towards solving a Problem in the Doctrine of Chances* (1763)

$$p(M|D) =$$

the *demarcation* problem in *Bayesian* context

The probability that a belief is true given **new evidence** equals the probability that the belief is true **regardless of that evidence** times the **probability that the evidence is true given that the belief is true** divided by the **probability that the evidence is true regardless** of whether the belief is true.

1

- *Thomas Bayes* *Essay towards solving a Problem in the Doctrine of Chances* (1763)

$$p(M|D) = P(M)...$$

"prior"

the *demarcation* problem in *Bayesian* context

The probability that a belief is true given **new evidence** equals the probability that the belief is true **regardless of that evidence** times the **probability that the evidence is true given that the belief is true** divided by the **probability that the evidence is true regardless** of whether the belief is true.

- *Thomas Bayes* *Essay towards solving a Problem in the Doctrine of Chances* (1763)

$$p(M|D) = P(M)\,P(D|M)...$$

the *demarcation* problem in *Bayesian* context

The probability that a belief is true given **new evidence** equals the probability that the belief is true **regardless of that evidence** times the **probability that the evidence is true given that the belief is true** divided by the **probability that the evidence is true regardless** of whether the belief is true.

1

- *Thomas Bayes* *Essay towards solving a Problem in the Doctrine of Chances* (1763)

$$p(M|D) = \frac{P(M)\,P(D|M)}{P(D)}$$

"evidence"

the *demarcation* problem in *Bayesian* context

The probability that a belief is true given[1] **new evidence** equals the probability that the belief is true **regardless of that evidence** times the **probability that the evidence is true given that the belief is true** divided by the **probability that the evidence is true regardless** of whether the belief is true.

\- *Thomas Bayes Essay towards solving a Problem in the Doctrine of Chances (1763)*

$$p(M|D) = \frac{P(M)\,P(D|M)}{P(D)}$$

key concepts

interpretation of probability

distributions

law of large numbers

central limit theorem

homework

- HW1 : explore the Maxwell Boltzmann distribution
- HW2: graphic demonstration of the Central Limit Theorem

Jacob Cohen, 1994

The earth is round (p=0.05)

http://fbb.space/dsps/Cohen1994_TheEarthIsRound_AmPsych.pdf

reading

the original link:
http://psycnet.apa.org/fulltext/1995-12080-001.html
(this link nees access to science magazine, but ou can use the link above which is the same file)

The Earth Is Round ($p < .05$)

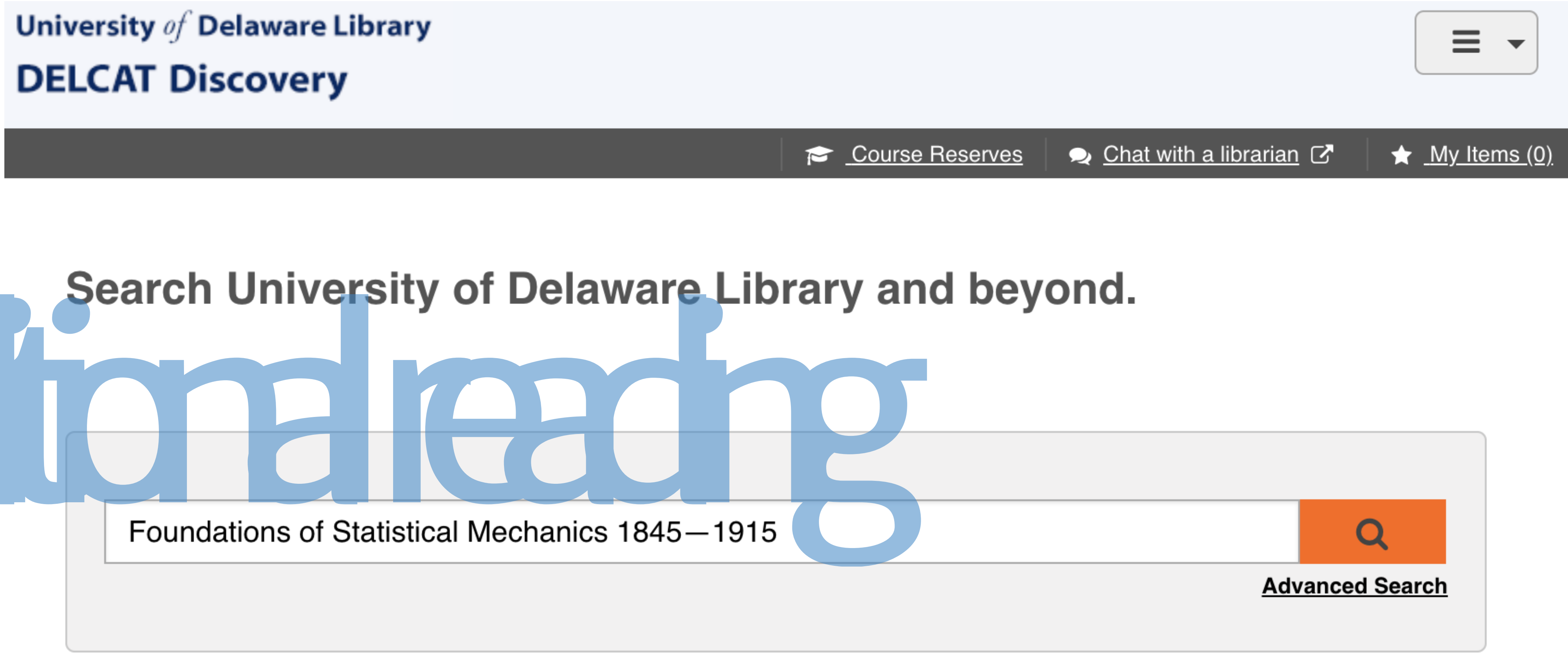
Jacob Cohen

After 4 decades of severe criticism, the ritual of null hypothesis significance testing—mechanical dichotomous decisions around a sacred .05 criterion—still persists. This article reviews the problems with this practice, including its near-universal misinterpretation of p *as the probability that* H_0 *is false, the misinterpretation that its complement is the probability of successful replication, and the mistaken assumption that if one rejects* H_0 *one thereby affirms the theory that led to the test. Exploratory data analysis and the use of graphic methods, a steady improvement in and a movement toward standardization in measurement, an emphasis on estimating effect sizes using confidence intervals, and the informed use of available statistical methods is suggested. For generalization, psychologists must finally rely, as has been done in all the older sciences, on replication.*

Foundations of Statistical Mechanics 1845—1915

Stephen G. Brush

Archive for History of Exact Sciences Vol. 4, No. 3 (4.10.1967), pp. 145-183



additional reading

Sarah Boslaugh, Dr. Paul Andrew Watters, 2008

Statistics in a Nutshell (Chapters 3,4,5)

https://books.google.com/books/about/Statistics_in_a_Nutshell.html?id=ZnhgO65Pyl4C

David M. Lane et al.

Introduction to Statistics (XVIII)

http://onlinestatbook.com/Online_Statistics_Education.epub

http://onlinestatbook.com/2/index.html

resources