data science for (physical) scientists Ib

I: epistemological concepts and working environment

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

1 probability

frequentist vs Bayesian interpretation

basic probability arithmetic

2 statistics

distributions

moments

central limit theorem

3 data science tools

github

python

jupyter notebooks

3 data science tools

- github
- python
- jupyter notebooks
- google colab
- stackoverflow

this slide deck

http://bit.ly/dsps2019_1

0 some administrative stuff

Syllabus

http://bit.ly/dspssyllabus

Learning Outcomes

By the end of this class you should be able to formulate an appropriate analysis plan for a research question, select, gather, and prepare data for analysis, and choose and apply machine learning methods to the data.

Syllabus

http://bit.ly/dspssyllabus

Learning Outcomes

By the end of this class you should be able to formulate an appropriate analysis plan for a research question, select, gather, and prepare data for analysis, and choose and apply machine learning methods to the data.

The instructors is: Dr. **Federica Bianco** fbianco@udel.edu
office hours: Tentatively: Tuesday 1230-2PM (or by appointment).
The Class assistants are:
Grader: **Yuqi Kong** kongyq@udel.edu
Office Hours: Monday 5PM Smith Hall 220
Technical and coding questions
Physics instructor: Dr. **Alexandre David-Uraz** adu@udel.edu
Office Hours: TBD Sharp 101B physics help center
Physics and Method-Application questions

Syllabus

http://bit.ly/dspssyllabus

- 10% pre-class questions
- 10% class performance and participation
- 20% homeworks
- 25% midterm
- 35% final

quizz

http://bit.ly/dspssyllabus

- 10% pre-class questions
- 10% class performance and participation
- 20% homeworks
- 25% midterm
- 35% final

from beginning of class to 5 minutes past (be on time!)

questions on previous class material and reading assignments

participation

http://bit.ly/dspssyllabus

- 10% pre-class questions
- 10% class performance and participation
- 20% homeworks
- 25% midterm
- 35% final

ask questions

answer questions

get up and code

extra credit assignments

homework

http://bit.ly/dspssyllabus

- 10% pre-class questions
- 10% class performance and participation
- 20% homeworks
- 25% midterm
- 35% final

Homework projects must be turned in as *jupyter notebooks* by checking them into your github account in a DSPS_<firstinitialLastname> repo and the project directories HW<hw number> (unless otherwise stated).

<finitialLastname> is e.g. fBianco

homework

Please work in groups of up to 5 people on homework as a collaborative projects.

Individual notebooks must be returned for each homework. Different group members should lead different aspects of the work. A statement **must be included in the README** explaining each team member's contribution (similar to an acknowledge of contribution you would find in a *Nature* letter see, for example these contributions).

- **10% pre-class questions**
- **10% class performance and participation**
- **20% homeworks**
- **25% midterm**
- **35% final**

homework

Please work in groups of up to 5 people on homework as a collaborative projects.

Individual notebooks must be returned for each homework. Different group members should lead different aspects of the work. A statement **must be included in the README** explaining each team member's contribution (similar to an acknowledge of contribution you would find in a *Nature* letter see, for example these contributions).

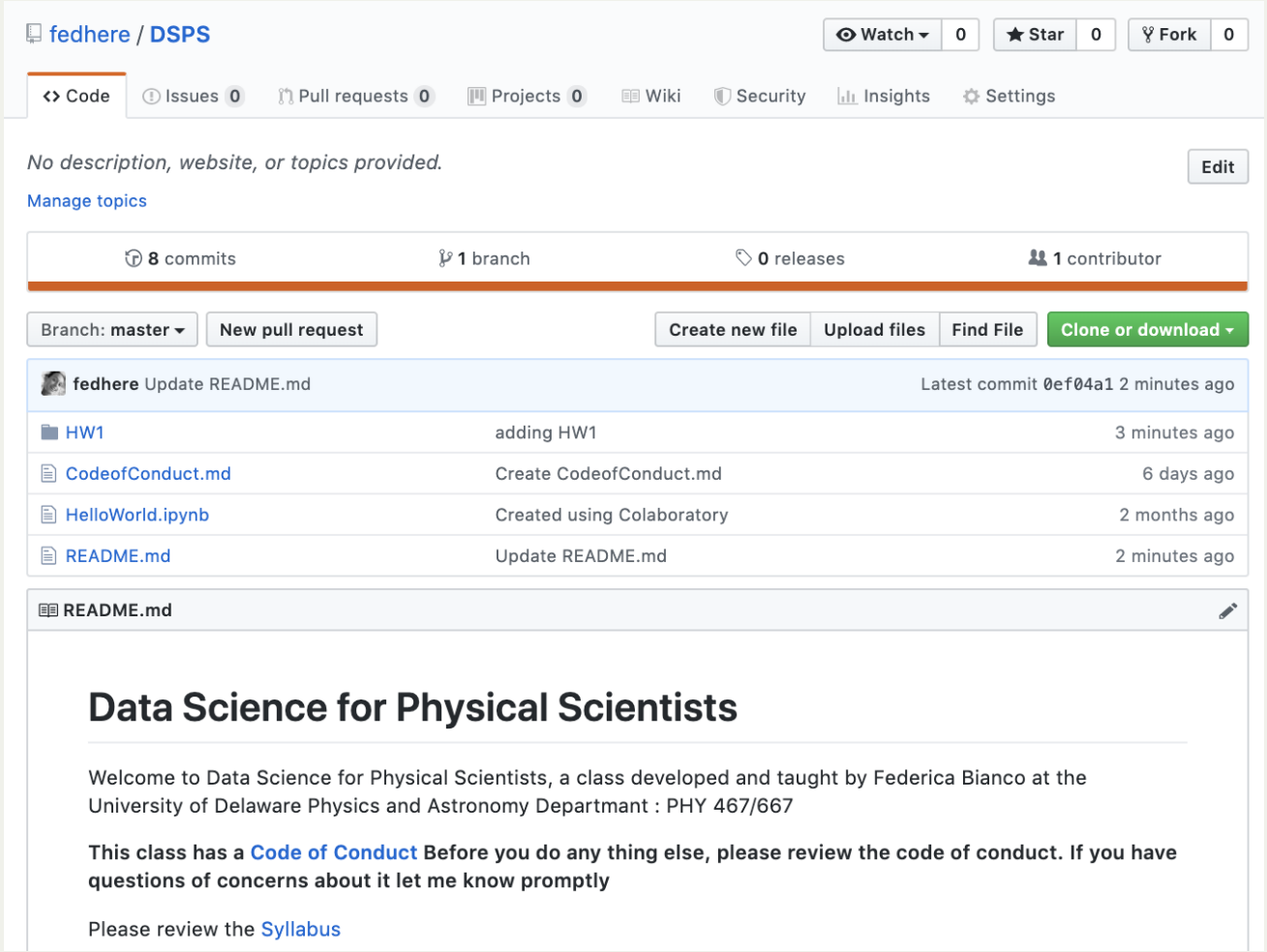
- 10% pre-class questions
- 10% class performance and participation
- 20% homeworks
- 25% midterm
- 35% final

Contributions

K.T. led the project and reduced the ALMA data. K.T. and D.I. wrote the manuscript. M.S.Y. reduced the Large Millimeter Telescope data and edited the final manuscript. Other authors contributed to the interpretation and commented on the ALMA proposal and the paper.

homework

- 10% pre-class questions
- 10% class performance and participation
- 20% homeworks
- 25% midterm
- 35% final

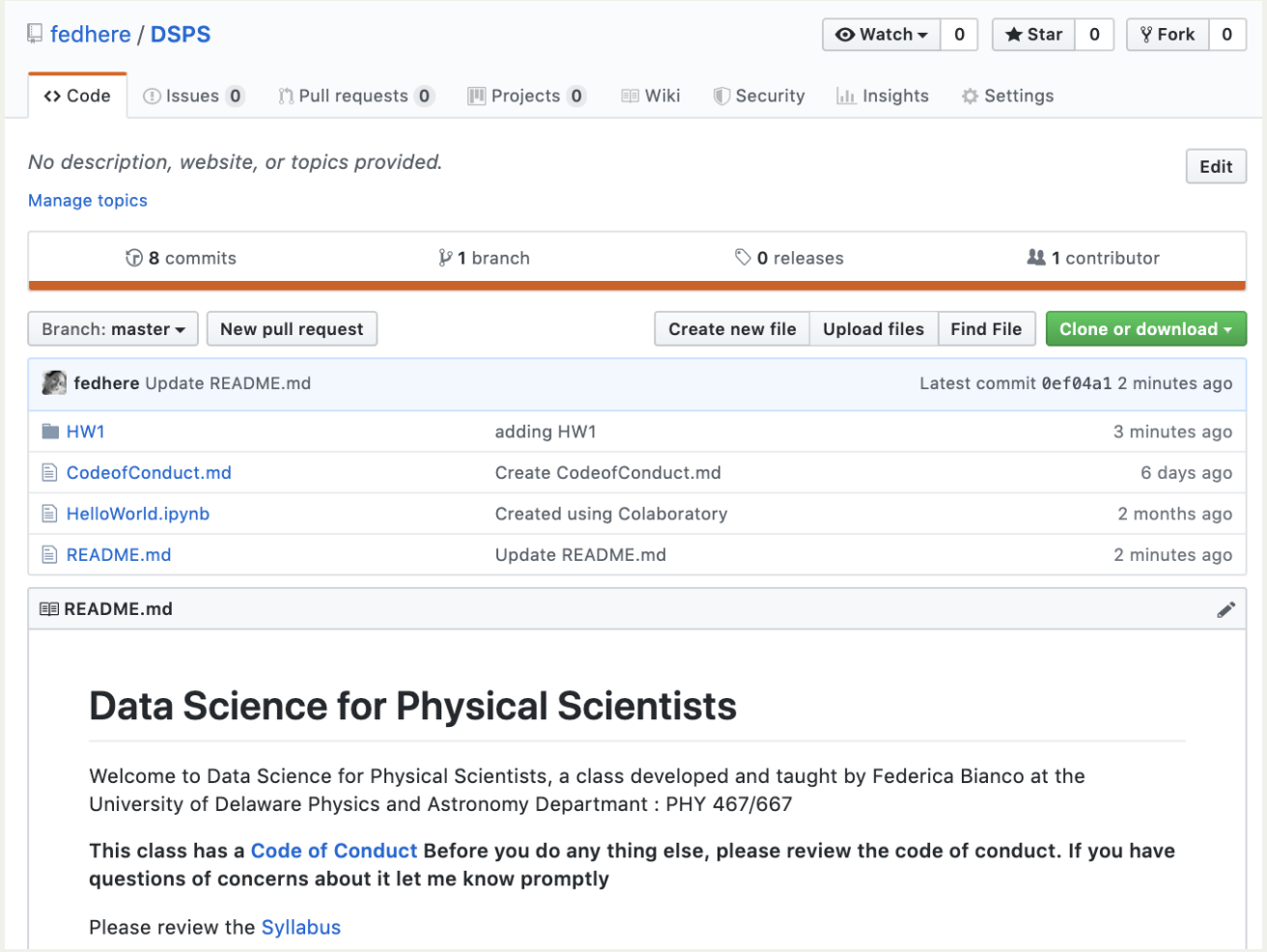


instructions will be here

https://github.com/fedhere/DSPS

homework

- 10% pre-class questions
- 10% class performance and participation
- 20% homeworks
- 25% midterm
- 35% final



instructions will be here

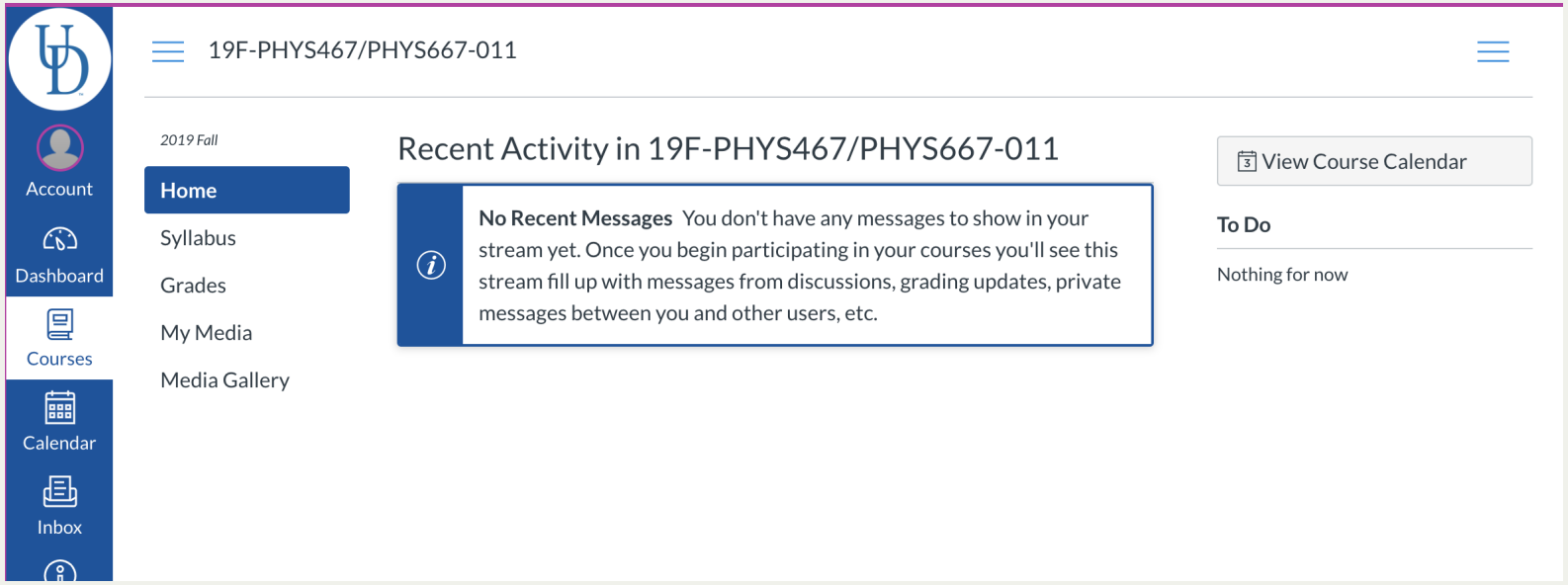
https://github.com/fedhere/DSPS

online help: http://bit.ly/2HtkQoc

please sign up asap!!

slack

homework



- 10% pre-class questions
- 10% class performance and participation
- 20% homeworks
- 25% midterm
- 35% final

of course there is also UD Canvas, which will be used to give you grades and occasionally post messages (I am still learning how to use it tho!)

HW 1 is posted already

midterm

For the *Midterm* and the *Final* you are responsible for material in the labs, the reading, and the homework. **In preparing for the exams, use the homework as a guide to which material is essential. In the Midterm and Final YOU WILL BE EXPECTED TO WORK INDIVIDUALLY.**

- 10% pre-class questions
- 10% class performance and participation
- 20% homeworks
- 25% midterm
- 35% final

Midterm... probably in class

issues: stereotype thread - working under derass is not necessarily a required skill

advantages: interviews for jobs are often timed

final

For the *Midterm* and the *Final* you are responsible for material in the labs, the reading, and the homework. **In preparing for the exams, use the homework as a guide to which material is essential. In the Midterm and Final YOU WILL BE EXPECTED TO WORK INDIVIDUALLY.**

- 10% pre-class questions
- 10% class performance and participation
- 20% homeworks
- 25% midterm
- 35% final

Final: take home, multiple days.

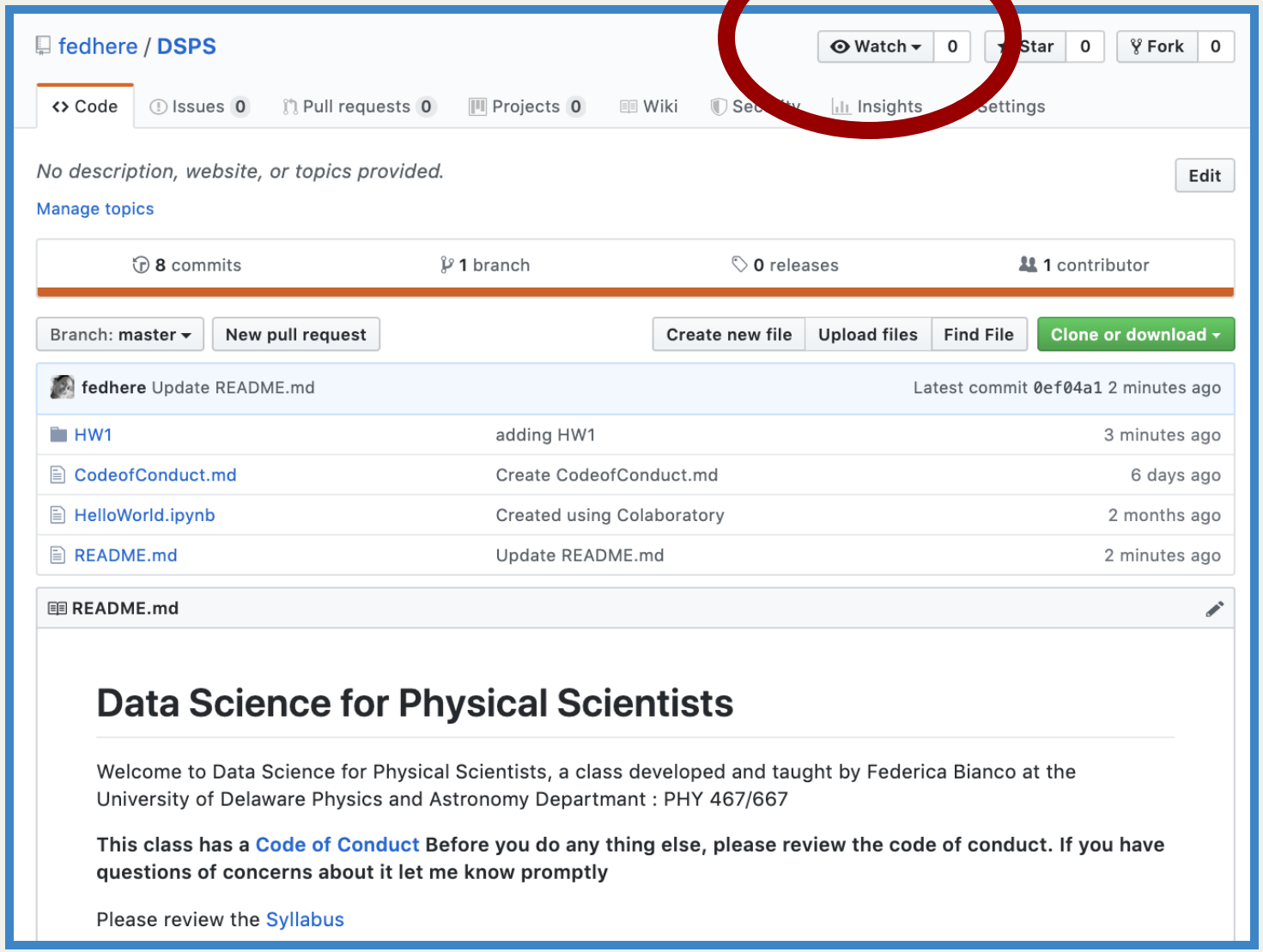
Resources

https://github.com/fedhere/DSPS

- SLIDES here
- HOMEWORK INSTRUCTIONS here
- RESOURCES here

If notebooks do not display use

https://nbviewer.jupyter.org



the scientific method

(what is science?)

PROGRAMMING

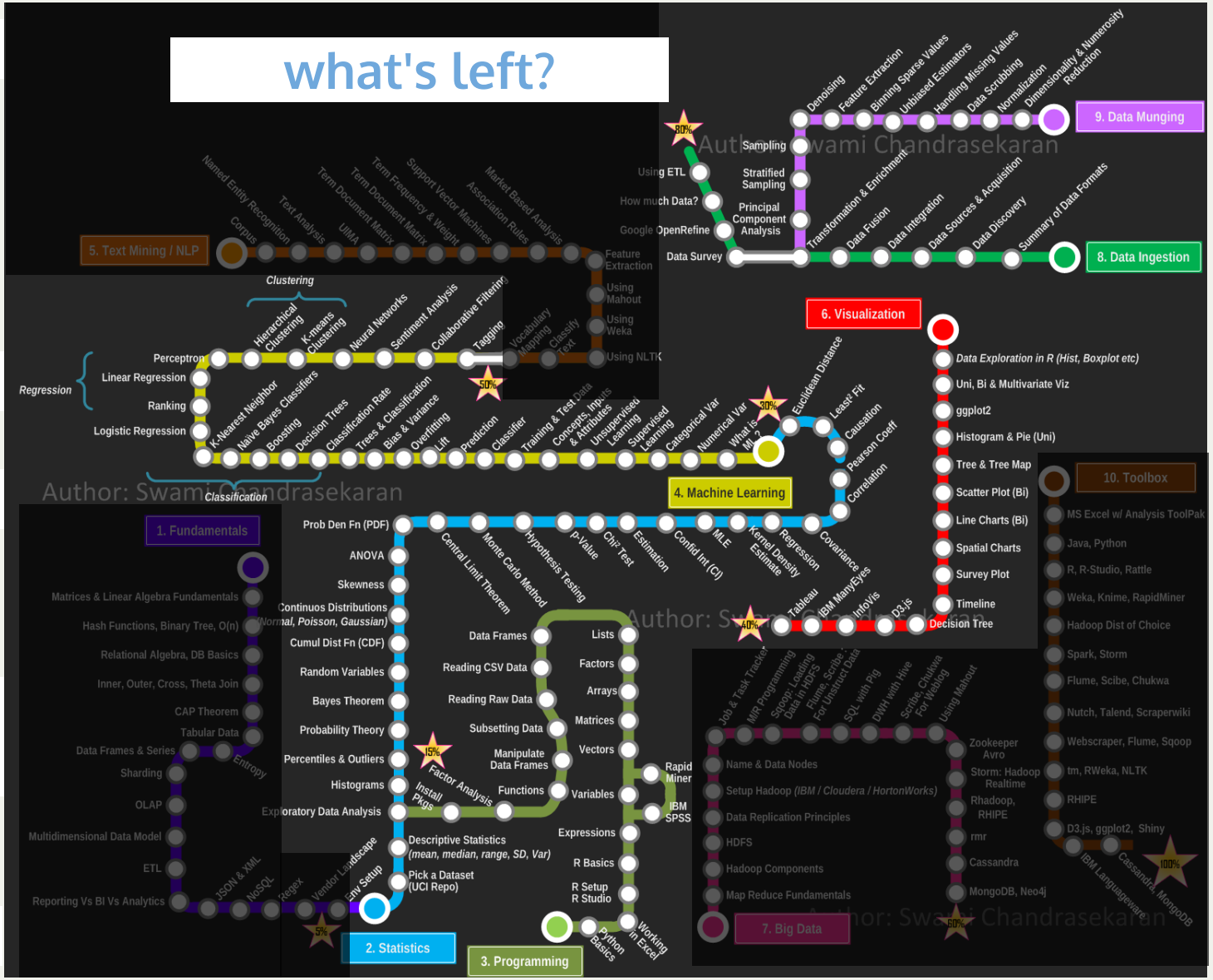
STATISTICS

DATA INGESTION

DATA MUNGING

MACHINE LEARNING

VISUALIZATION



http://nirvacana.com/thoughts/2013/07/08/becoming-a-data-scientist

python

probability distributions

p-values

uncertainties

MCMC

regression (linear, template)

classification (trees, neural networks)

clustering

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

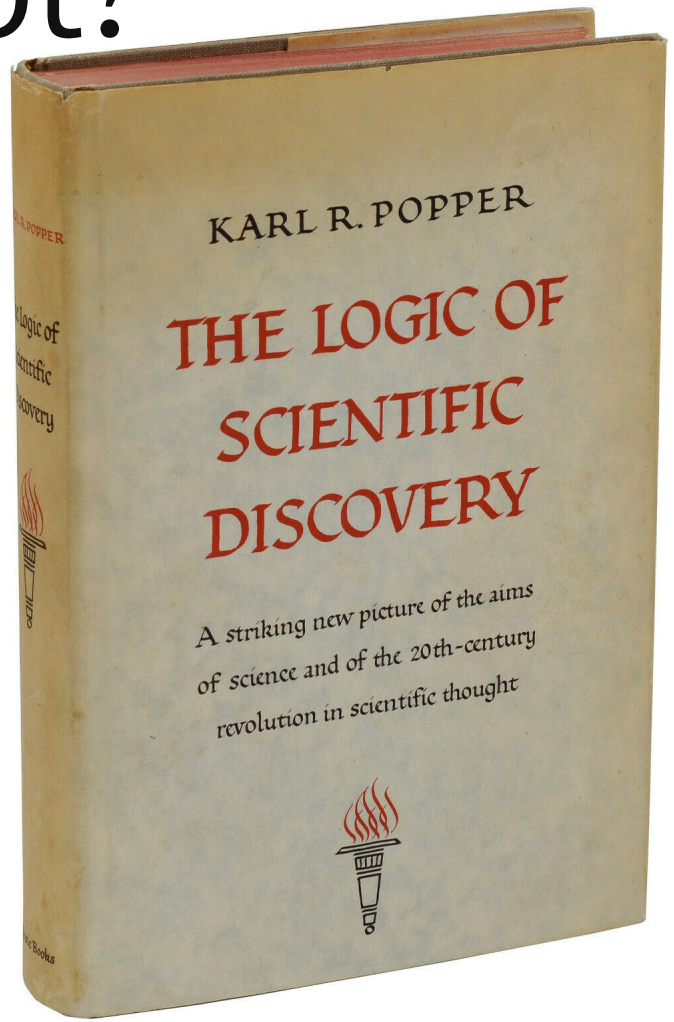
the *demarcation* problem:

what is science? what is not?

My proposal is based upon an *asymmetry* between verifiability and **falsifiability**; an asymmetry which results from the logical form of universal statements.

it is logically impossible to conclusively verify a universal proposition by reference to experience cause **not all examples could be tested**, but a single counter-instance conclusively falsifies the corresponding universal law.

—Karl Popper, *The Logic of Scientific Discovery*



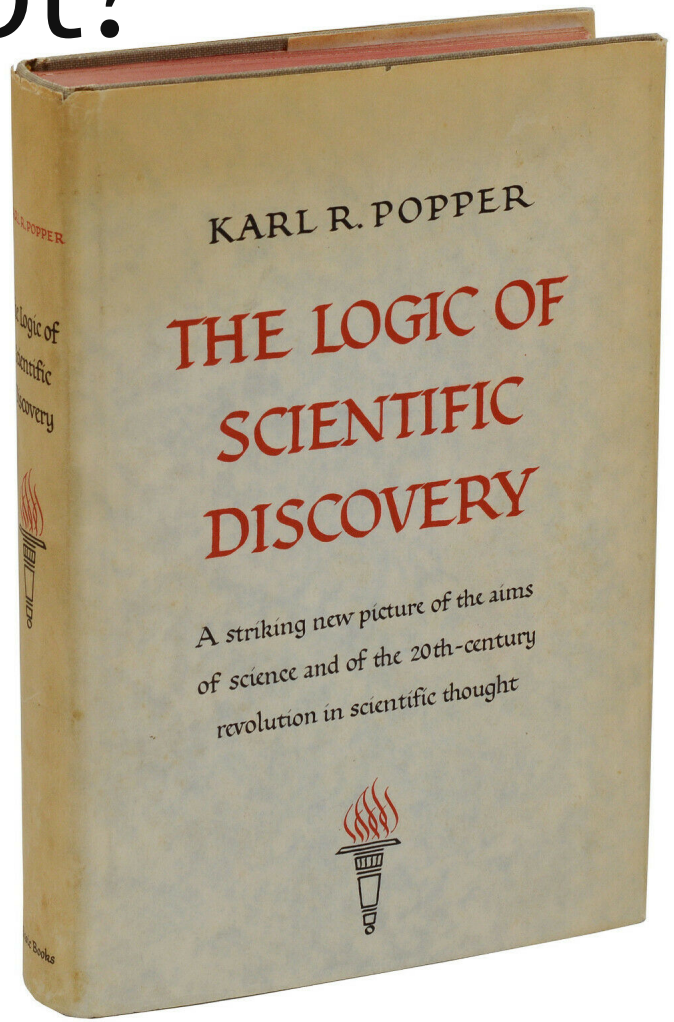
the *demarcation* problem:

what is science? what is not?

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.

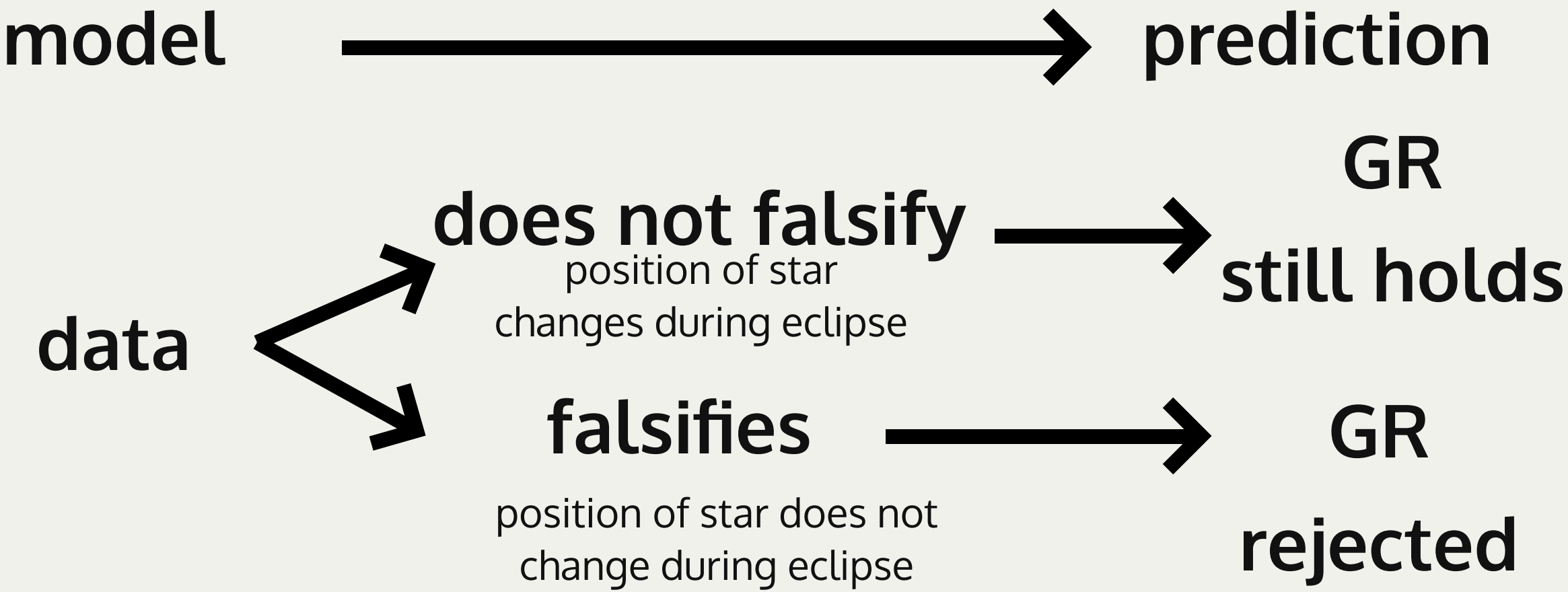
—Karl Popper, *The Logic of Scientific Discovery*

a scientific theory must be *falsifiable*



the *demarcation* problem

the *demarcation* problem



the *demarcation* problem

is phsychology a science?

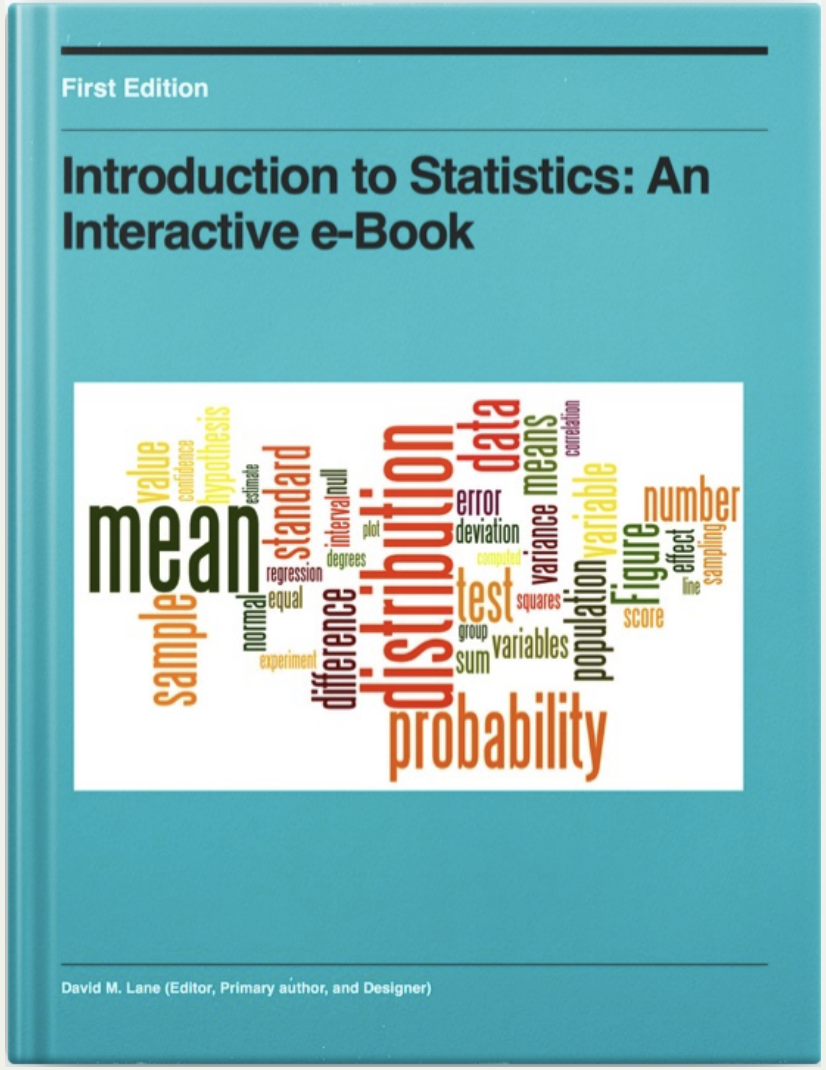
DISCUSS!

the *demarcation* problem

things can get more complicated though:

most scientific theories are actually based largely on *probabilistic induction* and modern *inductive inference* (Solomonoff, frequentist vs Bayesian methods...)

1 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?

1 probability

Basic Probability
Frequentist interpretation

fraction of times something happens

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

$\wedge$
$=$
$\vee$

probability of it happening

Basic Probability
Frequentist interpretation

P(E) = frequency of E

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

P(coin = head) = 51/100 = 0.51

fraction of times something happens

$\wedge$
$=$
$\vee$

probability of it happening

Basic probability arithmetics

Probability Arithmetic $0 <= P(A) <= 1$

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

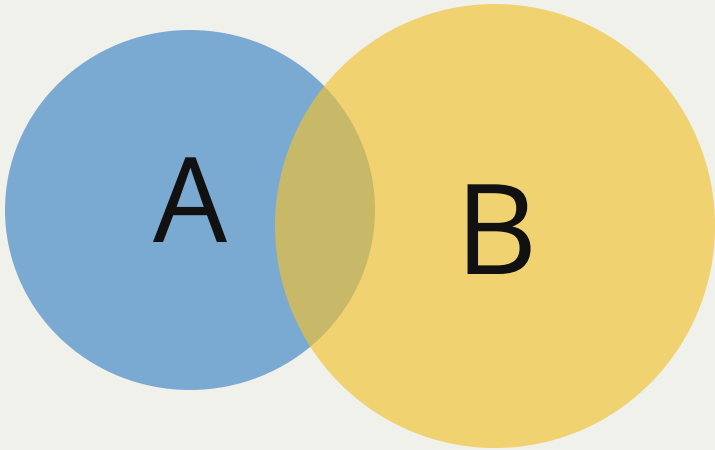
disjoint events $if\, P(A) \cap P(B) = 0 \text{ then :}$

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

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

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

Basic probability arithmetics



Probability Arithmetic

$$\text{in general :}$$

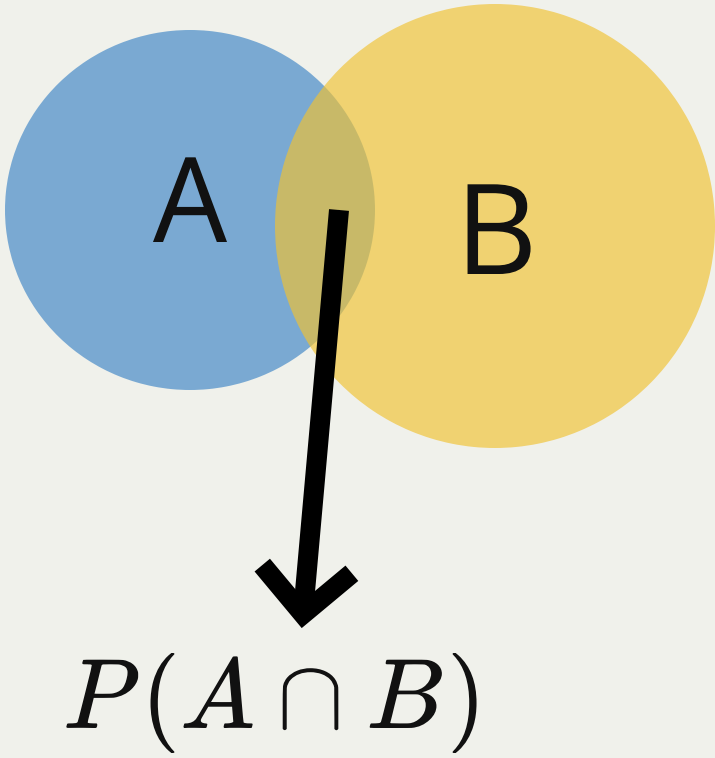
dependent events

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

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

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

Basic probability arithmetics



Probability Arithmetic

$$\text{in general :}$$

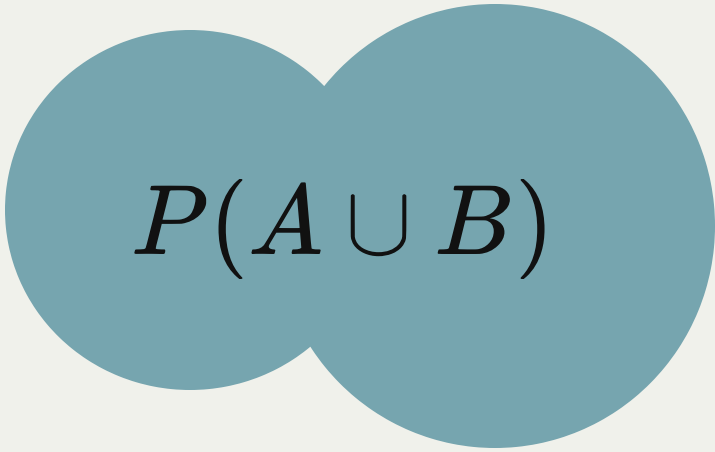
dependent events

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

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

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

Basic probability arithmetics



Probability Arithmetic

$$\text{in general :}$$

dependent events

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

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

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

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

HW

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/commons/8/82/Simulation_of_gas_for_relaxation_demonstration.gif?1567607773826

HW

Phsyics Example

Boltzmann 1872

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,

HW

Phsyics Example

Boltzmann 1872

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

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

Probability distributions

Binomial

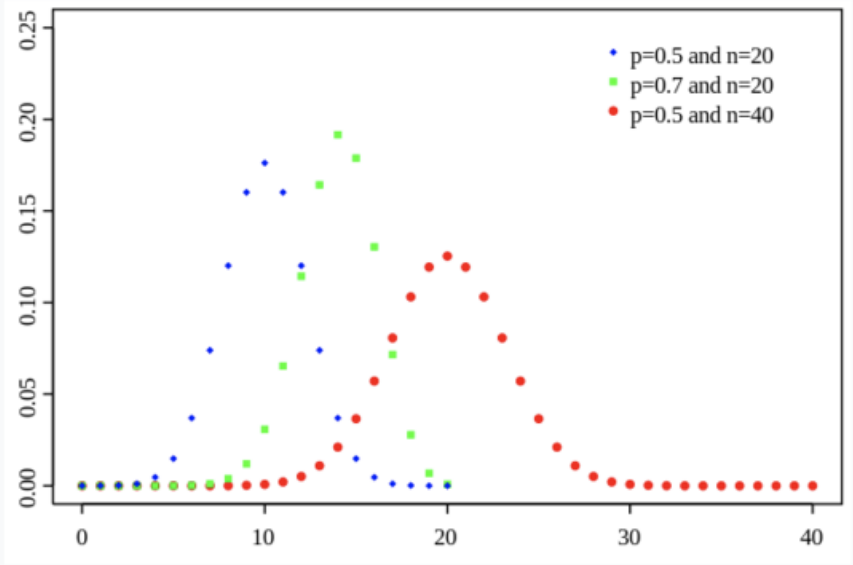
Coin toss:

fair coin: p=0.5 n=1

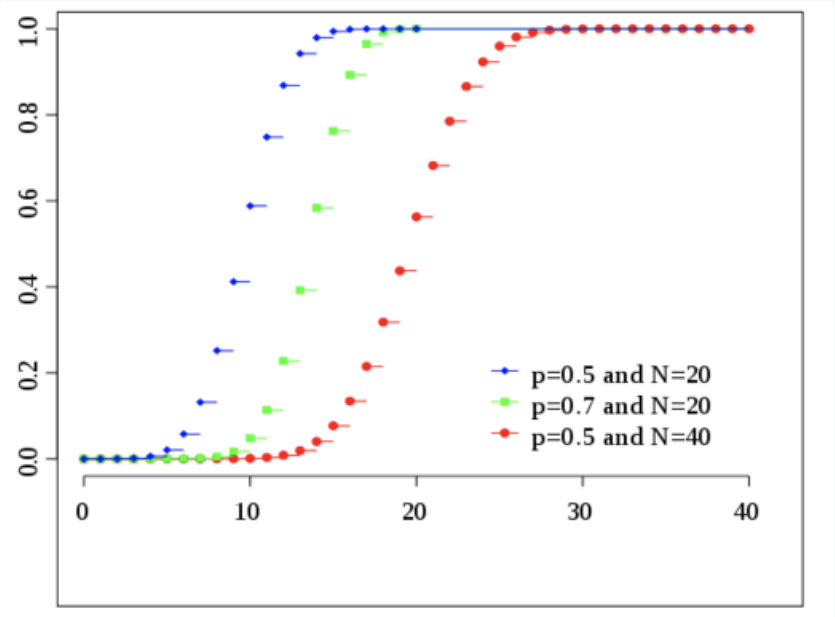
Vegas coin: p≠0.5 n=1

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

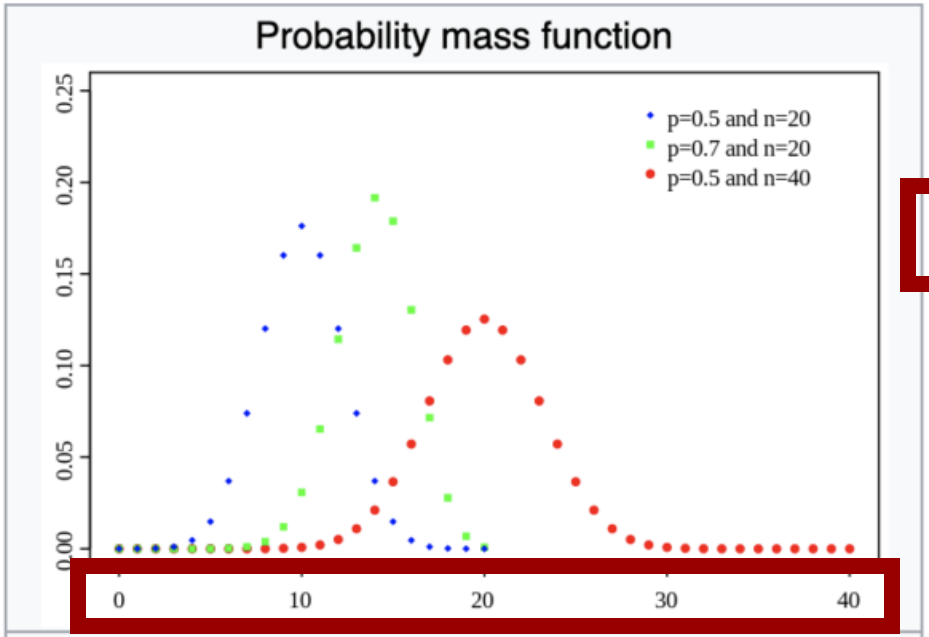
Binomial

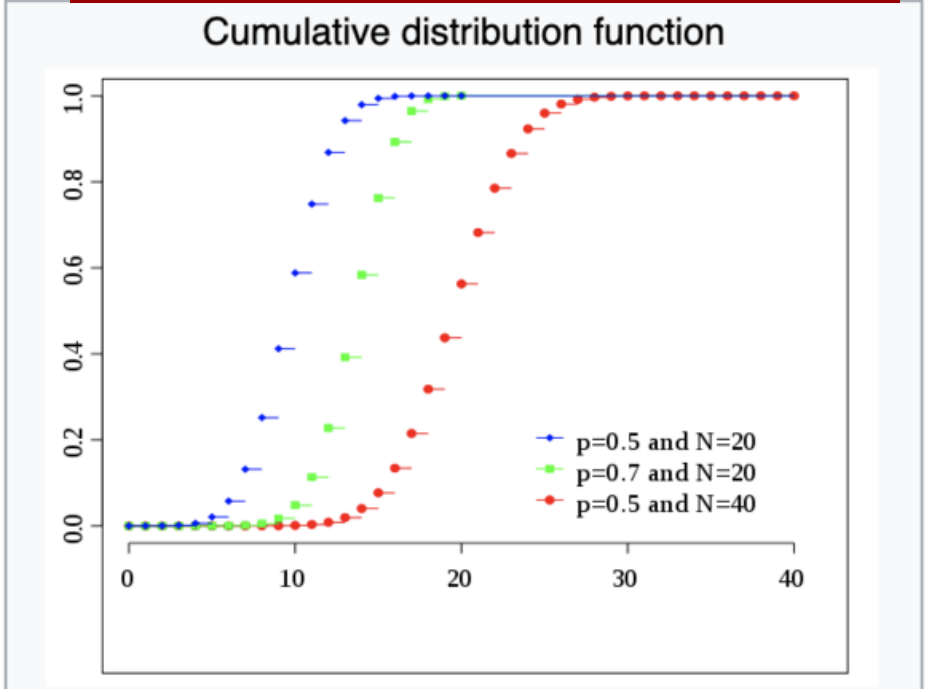
Coin toss:

fair coin: p=0.5 n=1

Vegas coin: p≠0.5 n=1

Binomial distribution





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

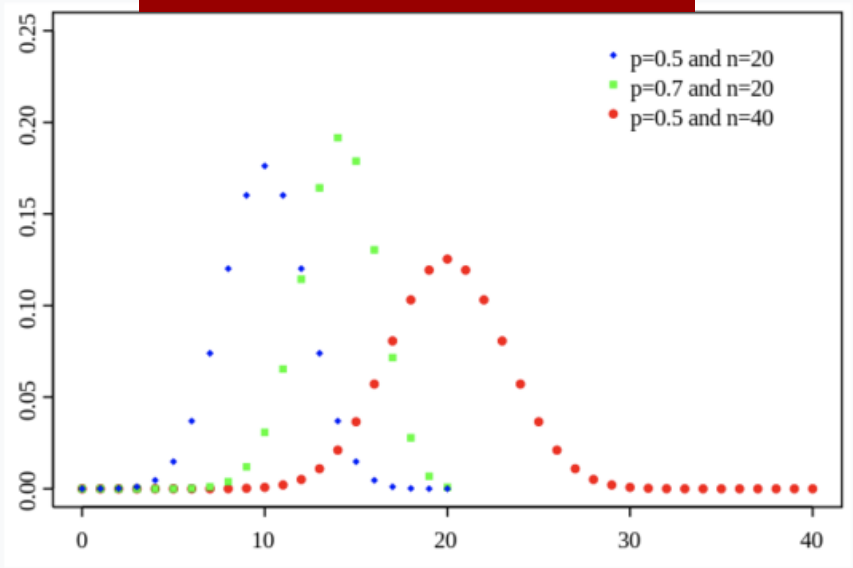
Coin toss:

fair coin: p=0.5 n=1

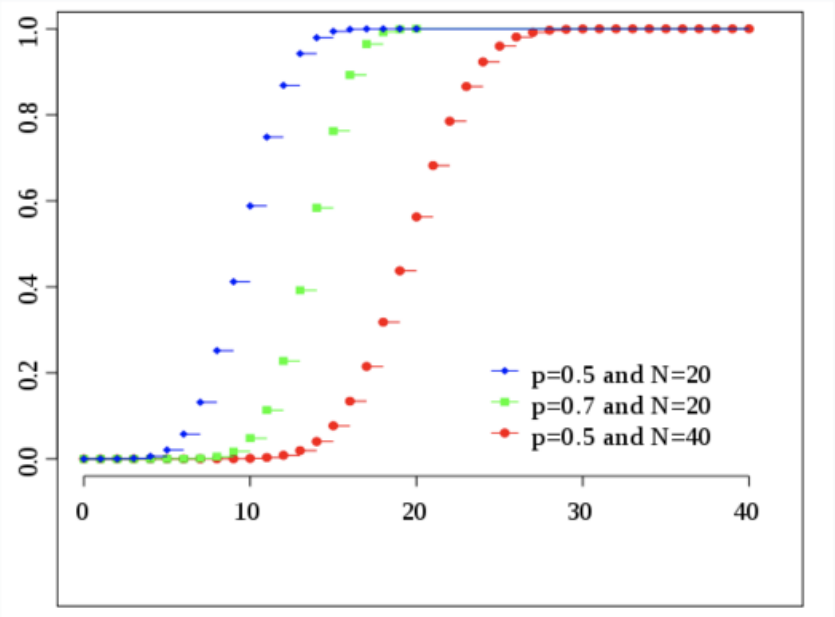
Vegas coin: p≠0.5 n=1

Binomial distribution

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

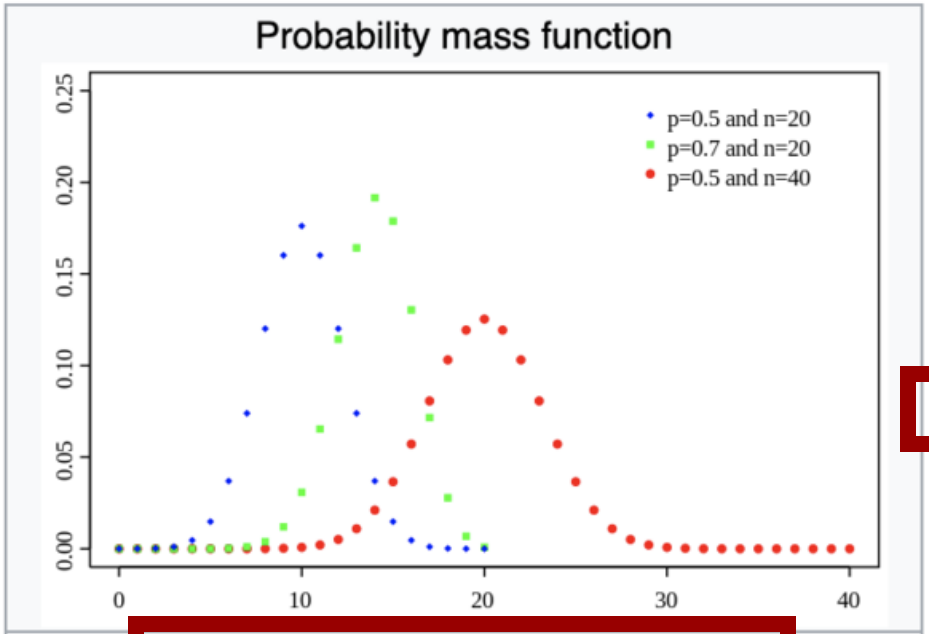
Binomial

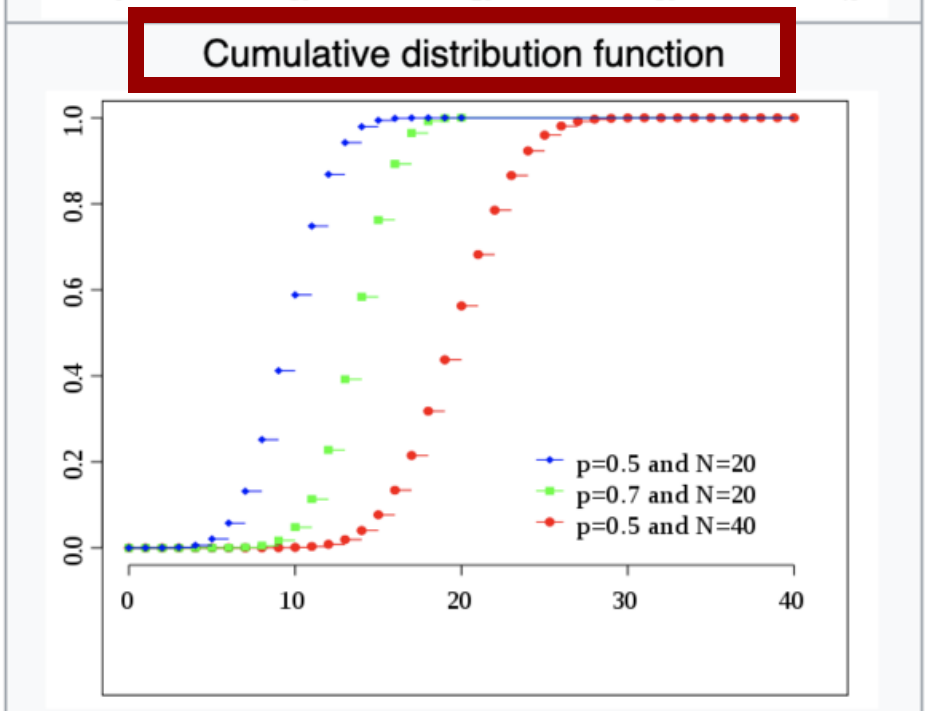
Coin toss:

fair coin: p=0.5 n=1

Vegas coin: p≠0.5 n=1

Binomial distribution





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

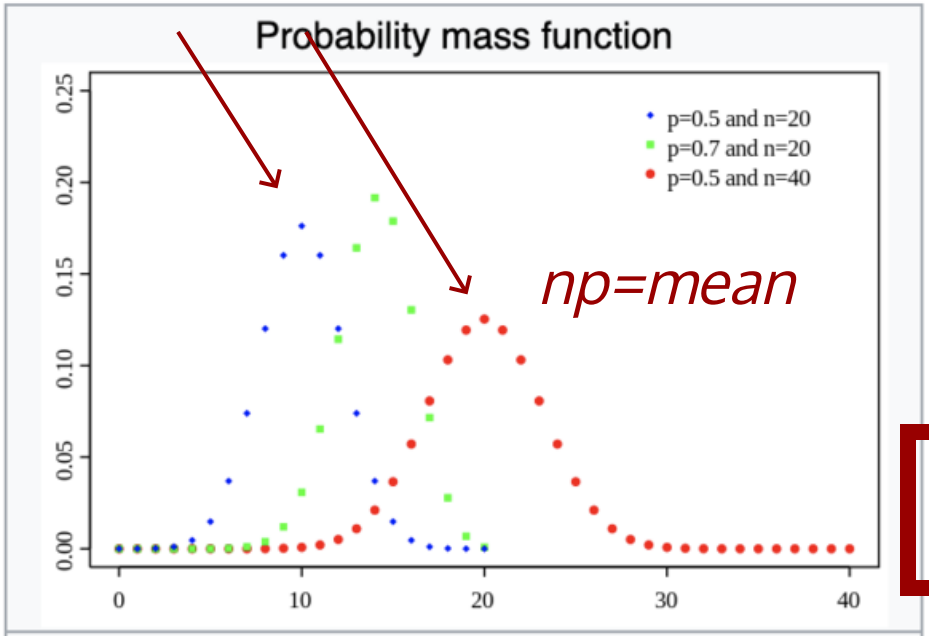
Binomial

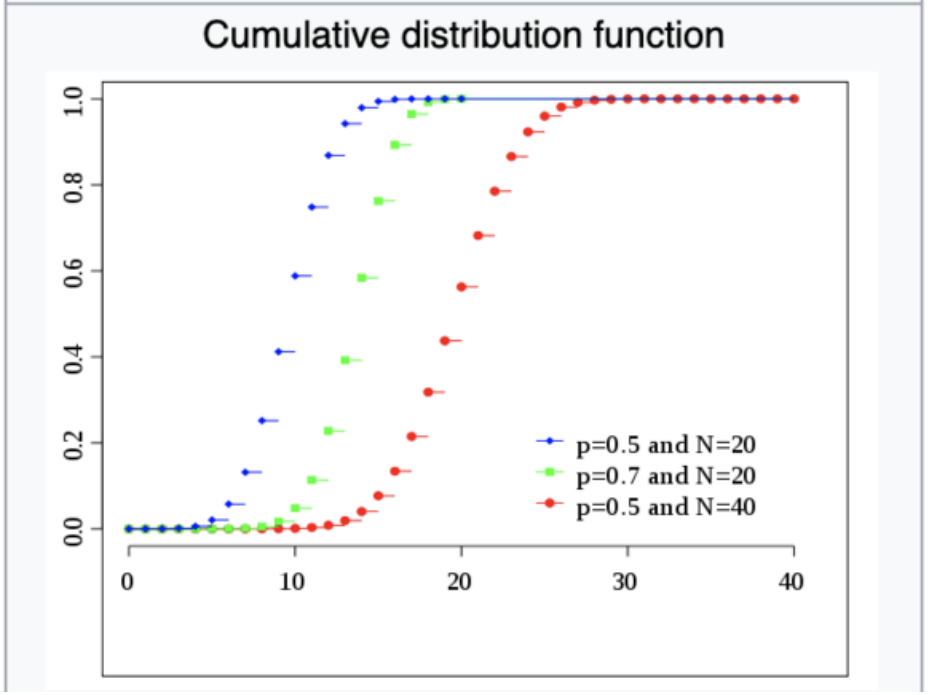
Coin toss:

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

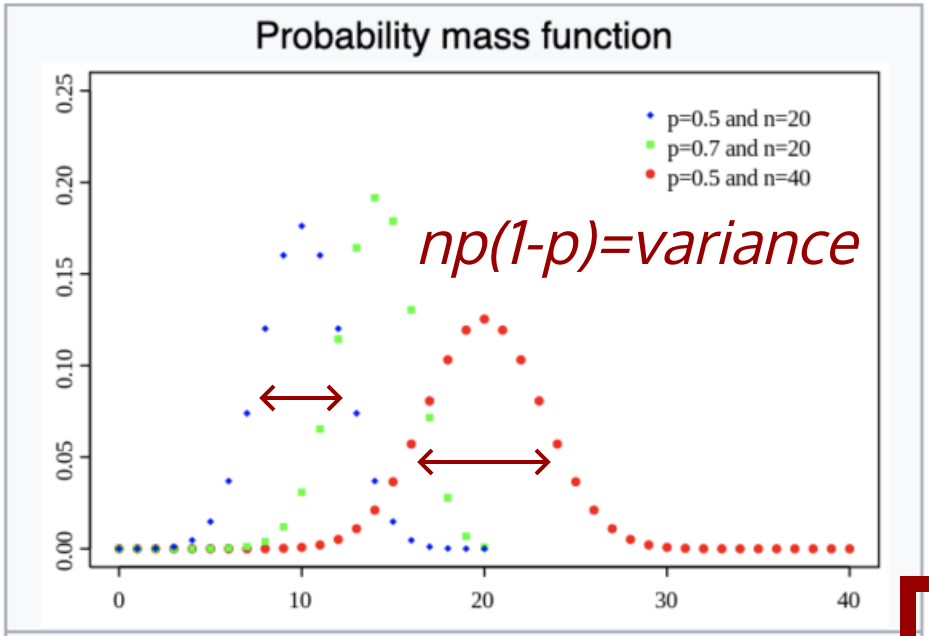
Binomial

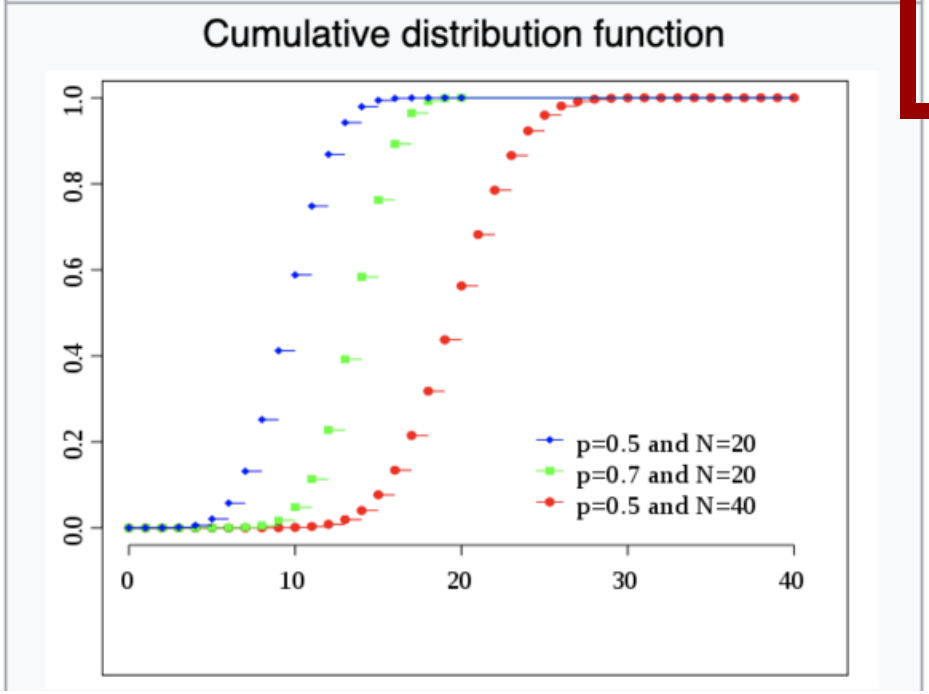
Coin toss:

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	$\frac{1}{2}\log_2(2\pi enp(1-p)) + O\left(\frac{1}{n}\right)$ 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

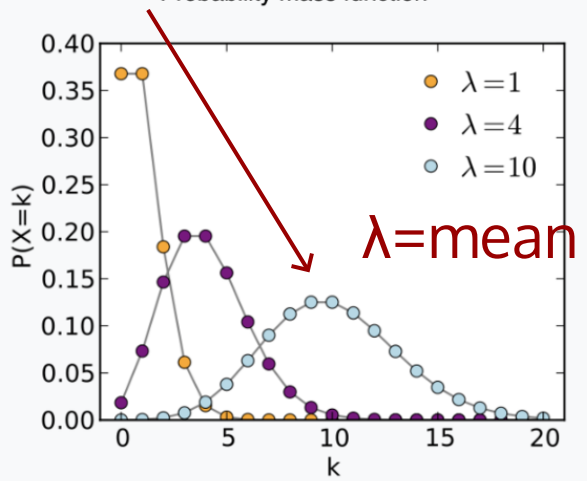
Poisson

Shut noise/count noise

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

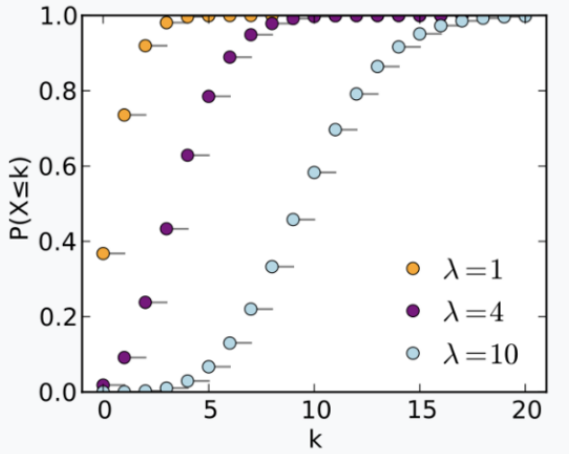
Poisson

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

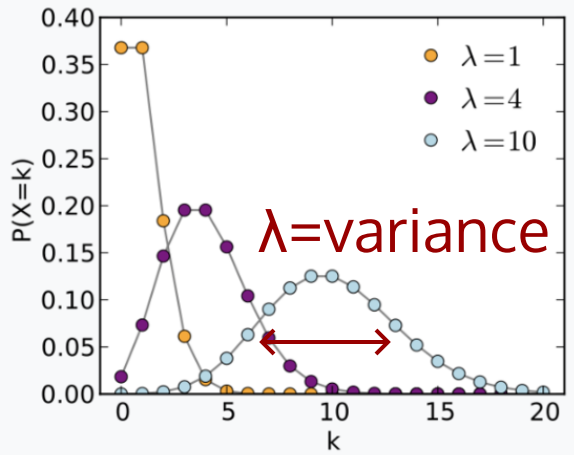
Poisson

Shut noise/count noise

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

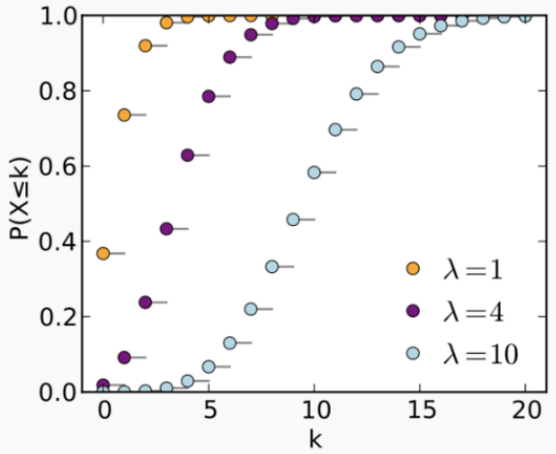
Poisson

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

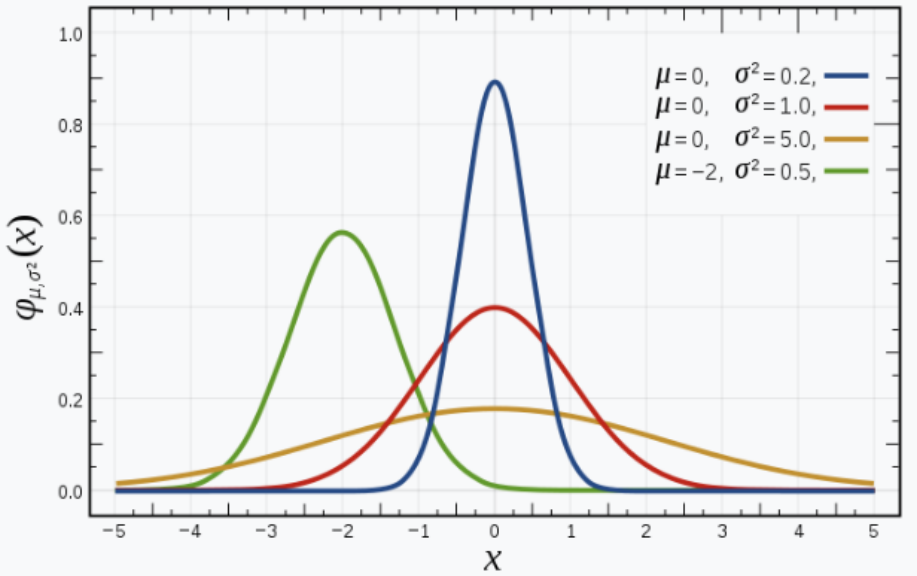
Gaussian

most common noise:

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

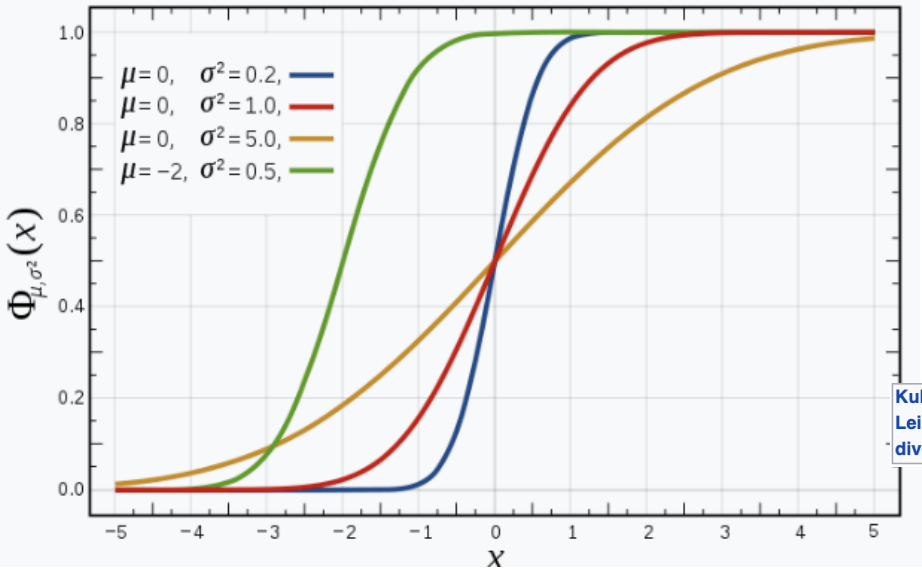
Normal Distribution

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	$D_{\mathrm{KL}}(\mathcal{N}_0 \| \mathcal{N}_1) = \frac{1}{2}\{(\sigma_0/\sigma_1)^2 + \frac{(\mu_1-\mu_0)^2}{\sigma_1^2} - 1 + 2\ln\frac{\sigma_1}{\sigma_0}\}$

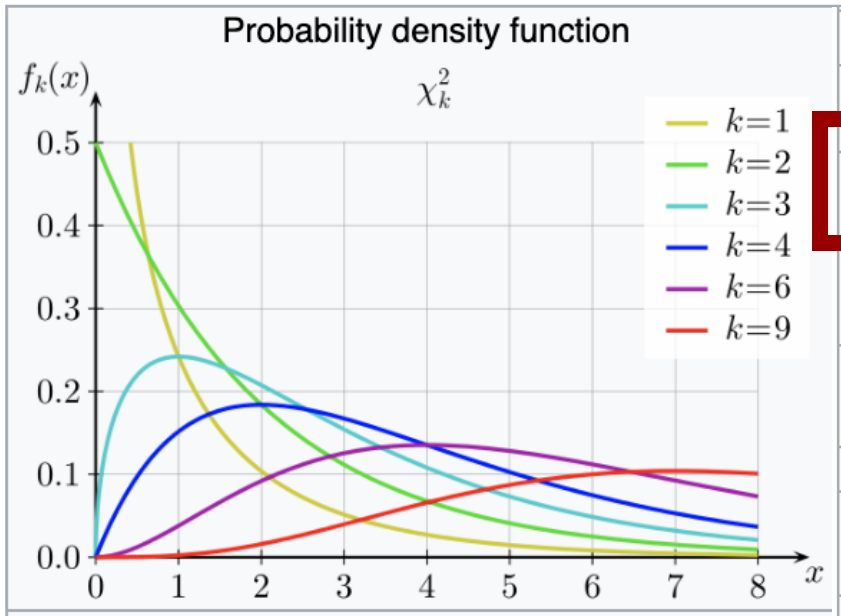
Probability distributions

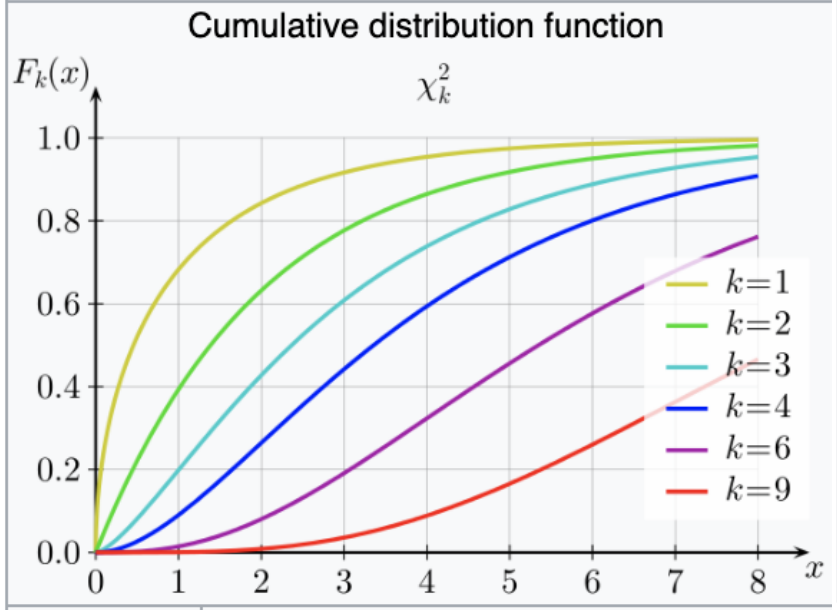
Chi-square (χ^2)

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 $0 < t < \sqrt{e}$

coding time!

https://colab.research.google.com/

https://github.com/fedhere/DSPS/blob/master/statistics/distributionParametersDemo.ipynb

HW

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

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

3

the scientific method in a probabilistic context

p(physics | data)

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

Bayesian Inference

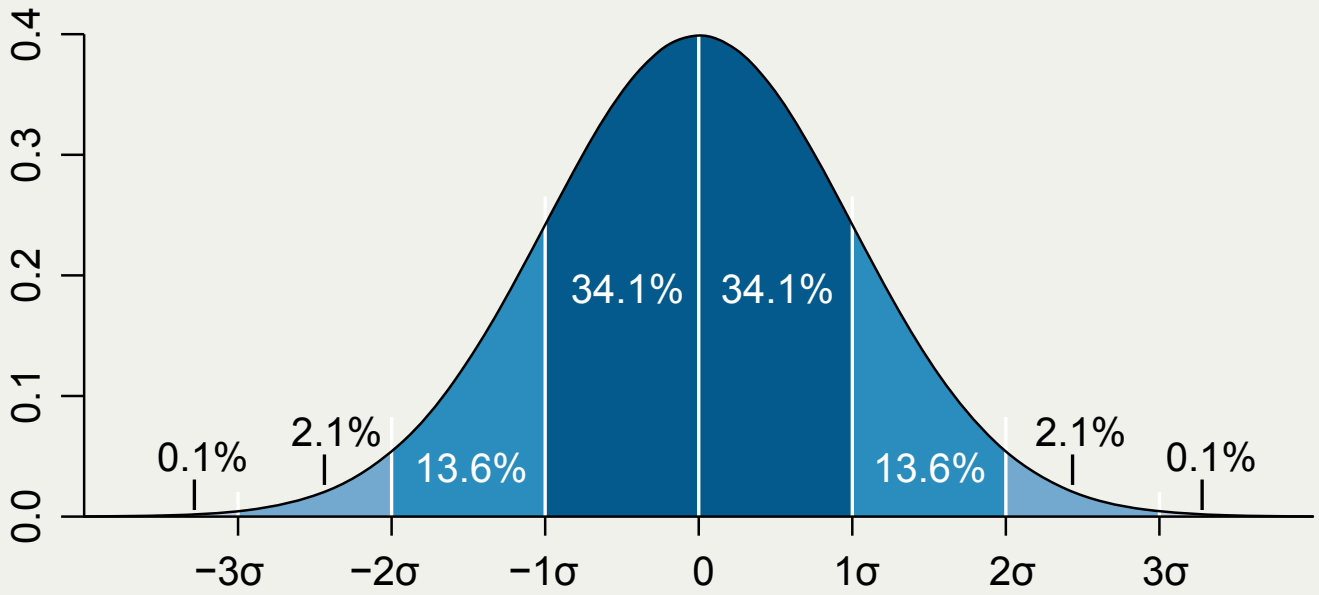
Forward Modeling

Frequentist approach
(NHRT)

p(physics | data)

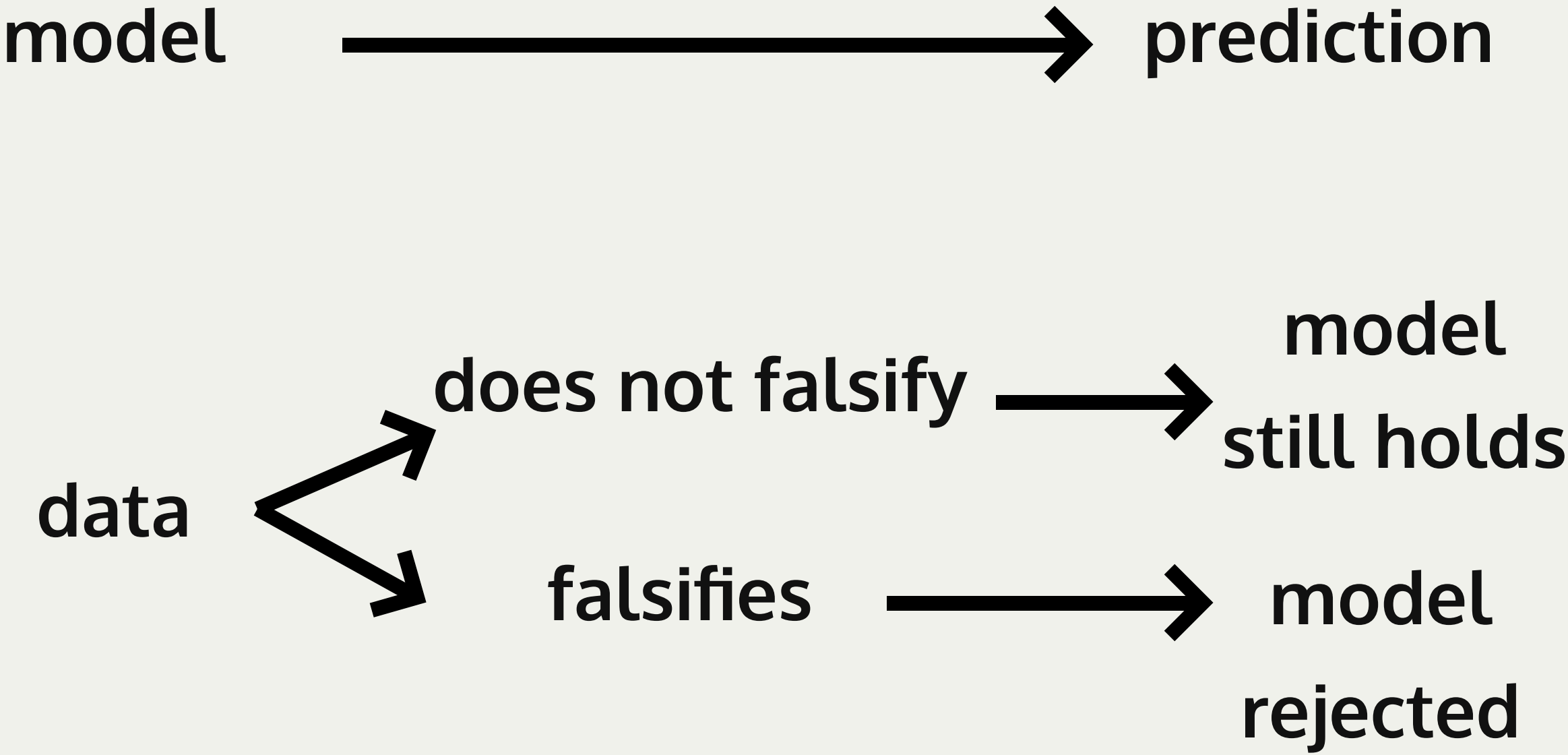
https://speakerdeck.com/dfm/emcee-odi

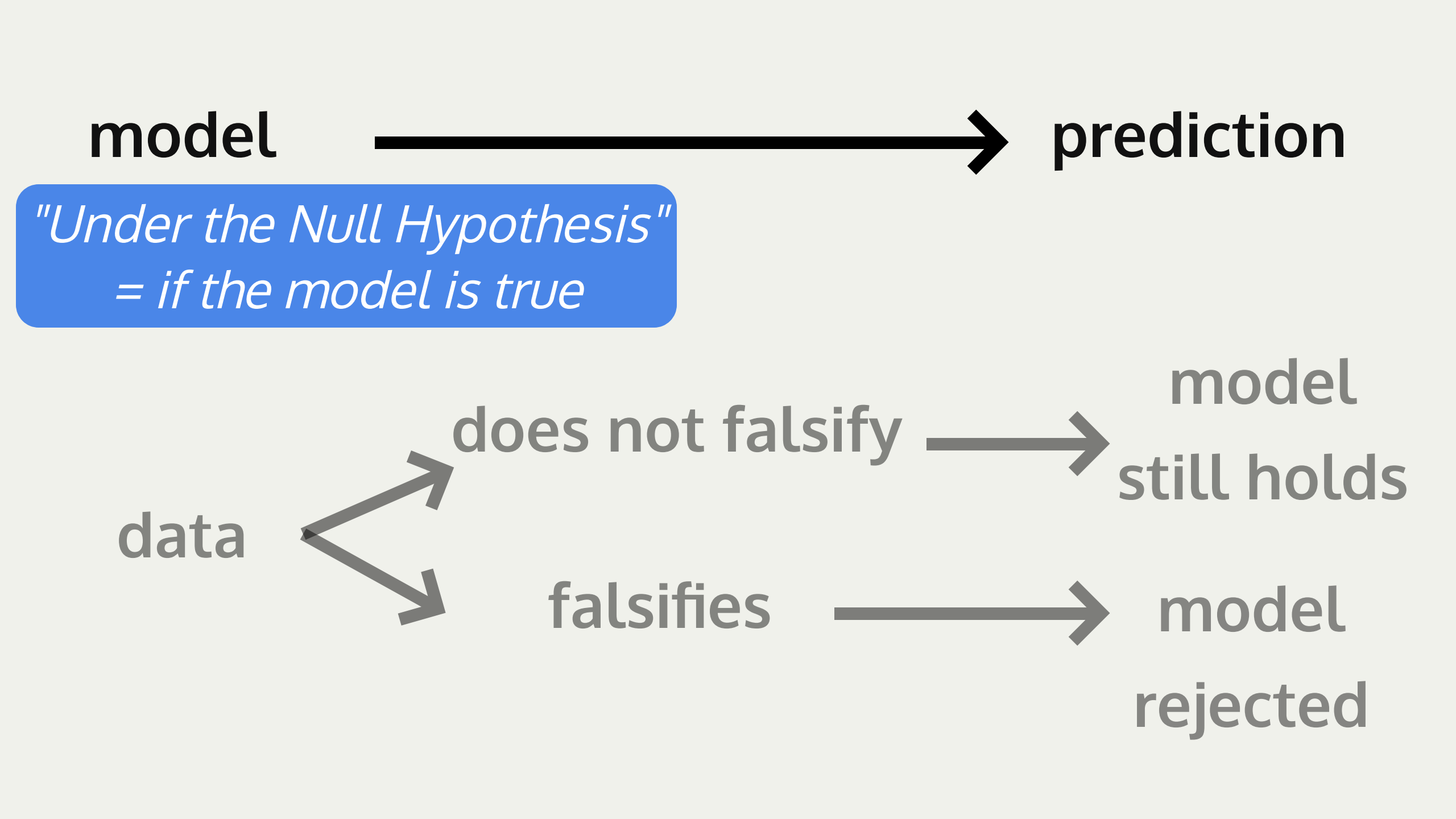
Null
Hypothesis
Rejection
Testing

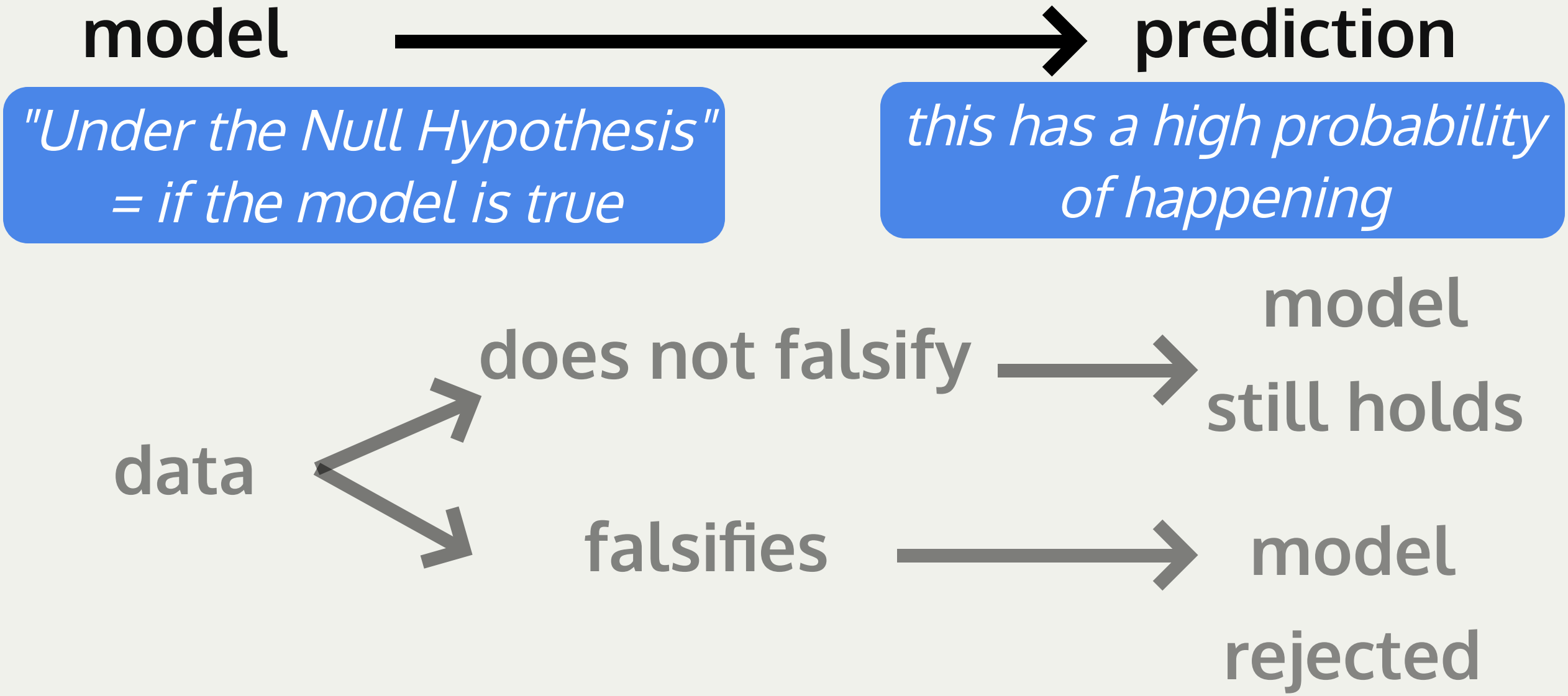


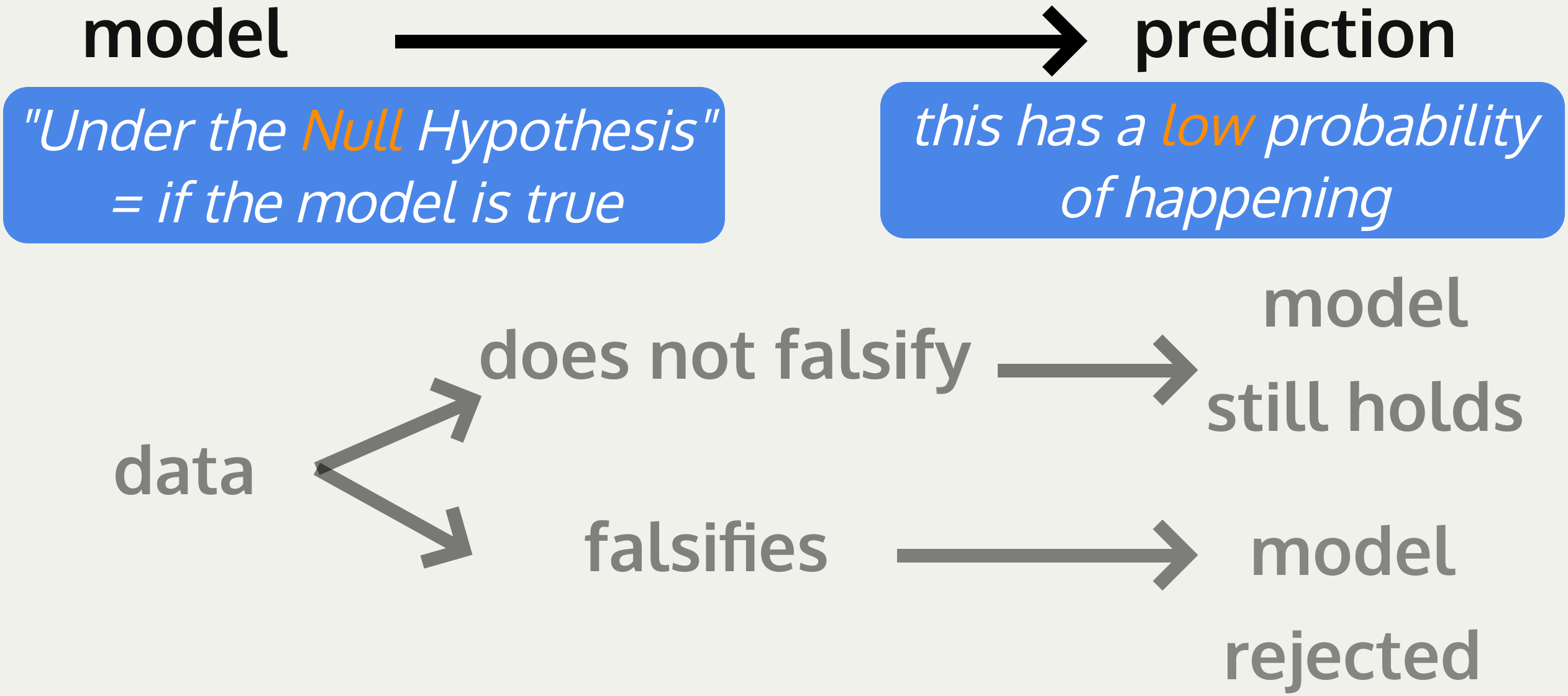
p(physics | data)

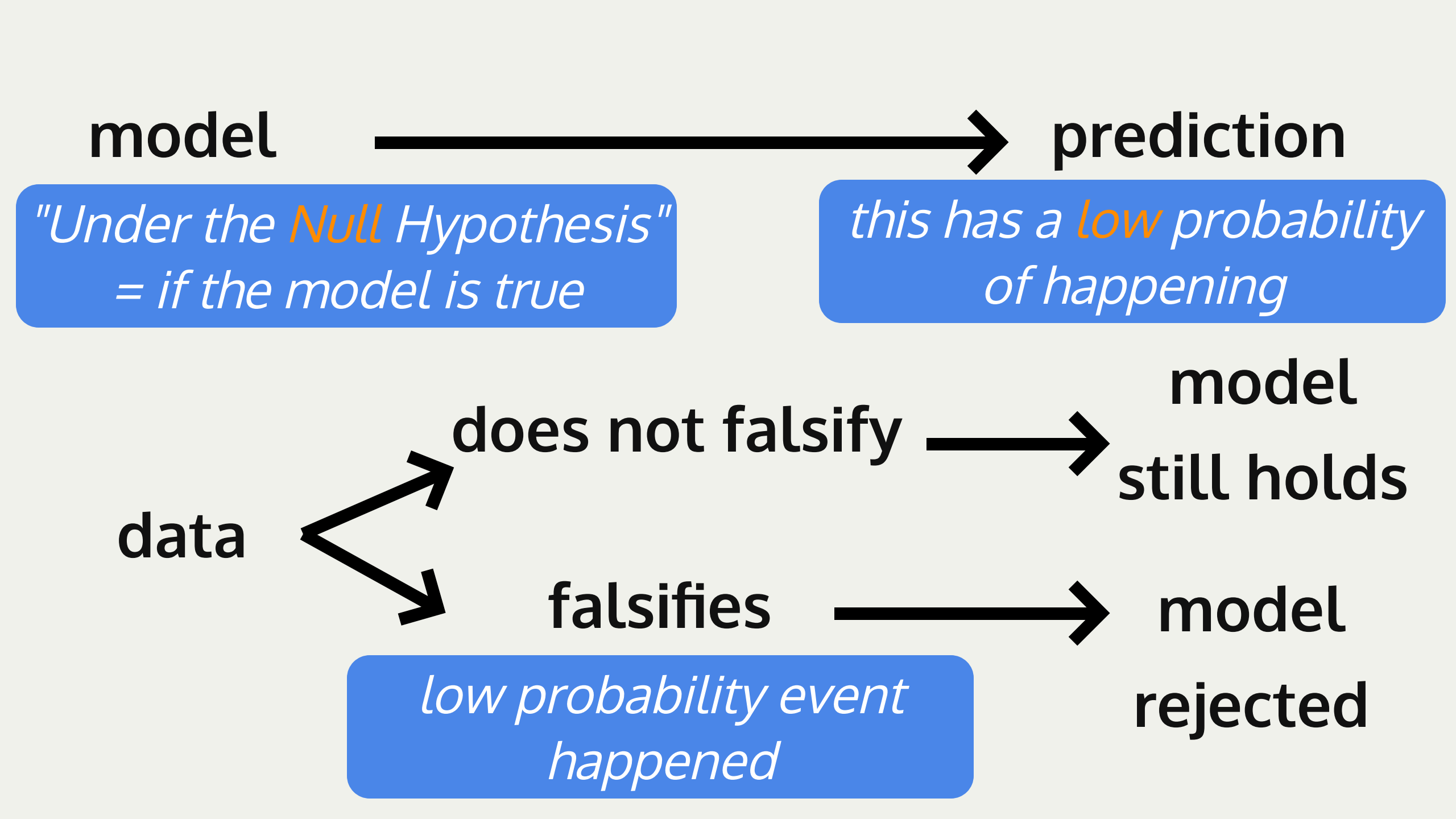
https://speakerdeck.com/dfm/emcee-odi

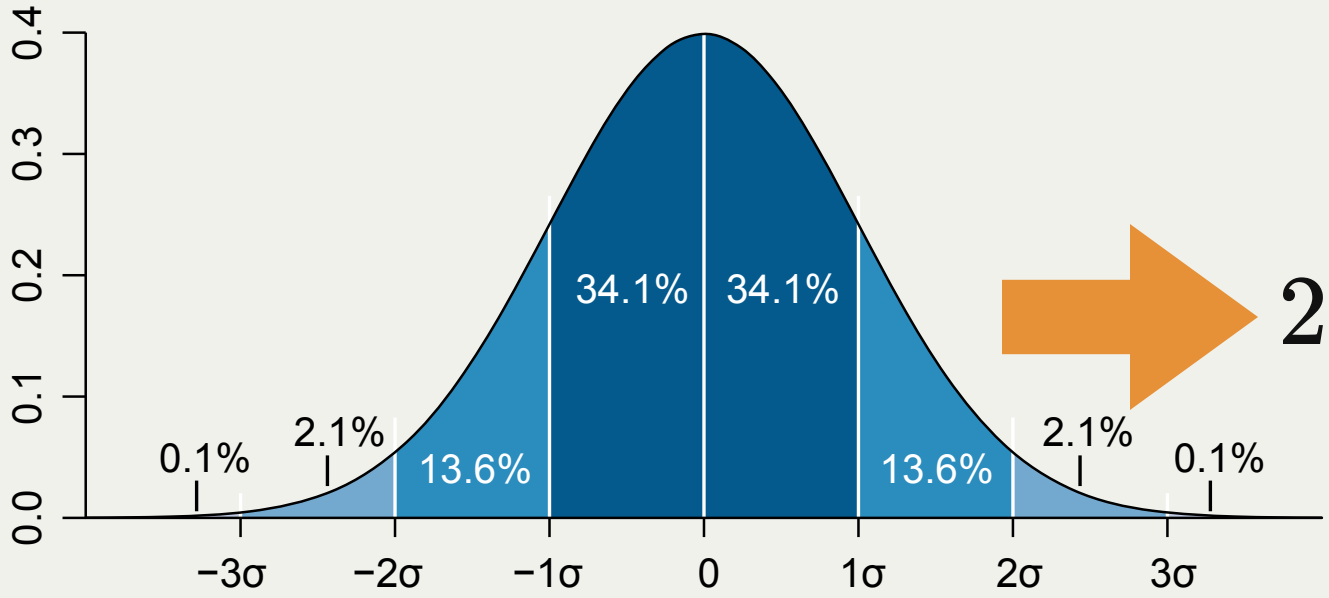

model
prediction
data
does not falsify
model
still holds
falsifies
model
rejected


model
prediction
"Under the Null Hypothesis"
= if the model is true
data
does not falsify
model
still holds
falsifies
model
rejected


model
prediction
"Under the Null 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 true
this has a low probability
of happening
data
does not falsify
model
still holds
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

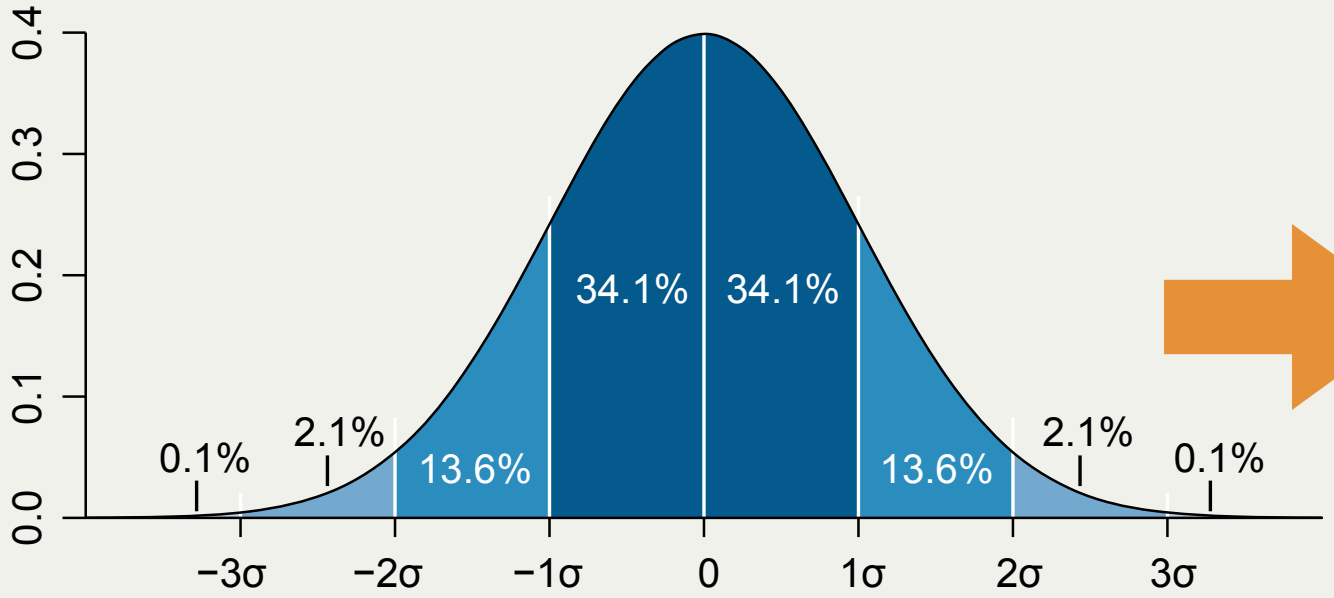
rejected at 95%
0.05 p-value
5% confidence

data

does not falsify → **model still holds**

falsifies → **model rejected**

low probability event happened



rejected at 99.7%

3σ 0.003 p-value

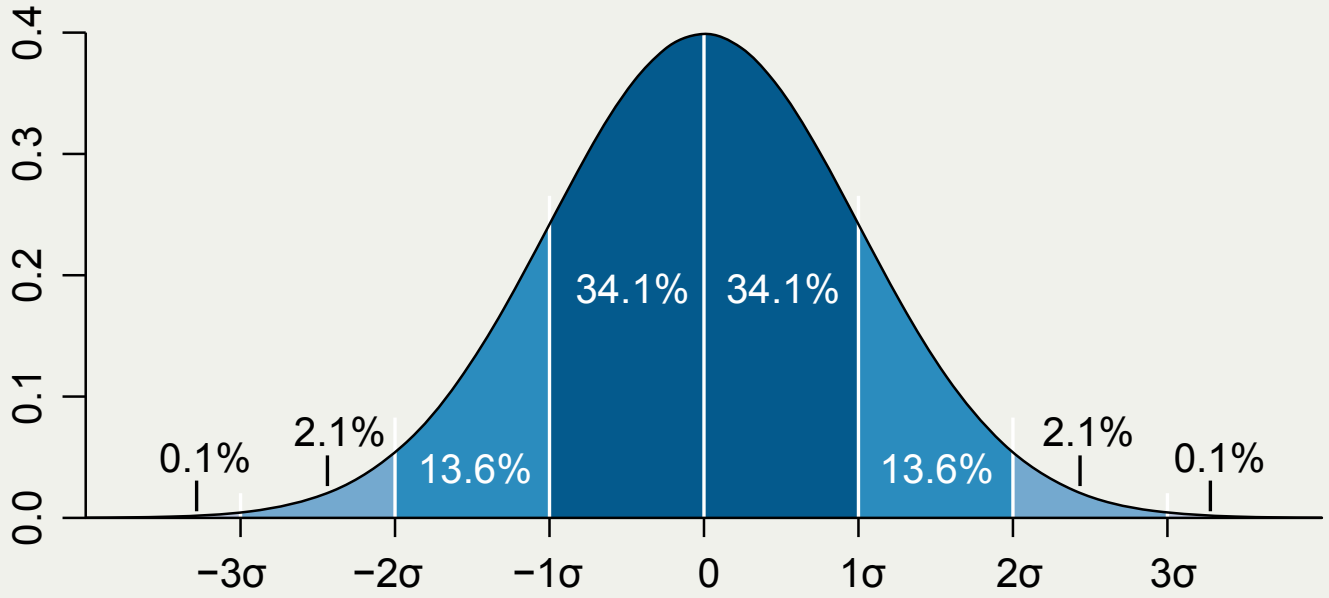
0.3% confidence

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

data

does not falsify → **model still holds**

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

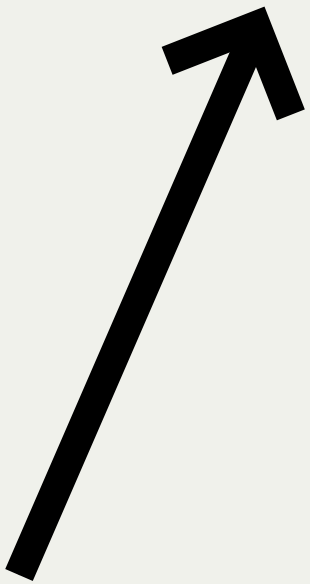
Null
Hypothesis
Rejection
Testing

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

prediction is unlikely

Alternative rejected

Null holds



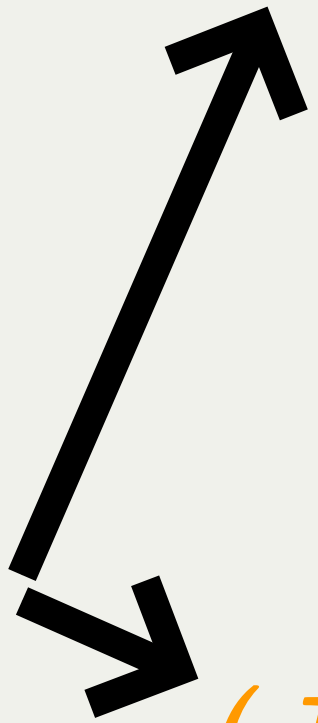
4

test data against
alternative outcomes

Null
Hypothesis
Rejection
Testing

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

prediction is unlikely
Alternative rejected
Null holds



4

test data against
alternative outcomes

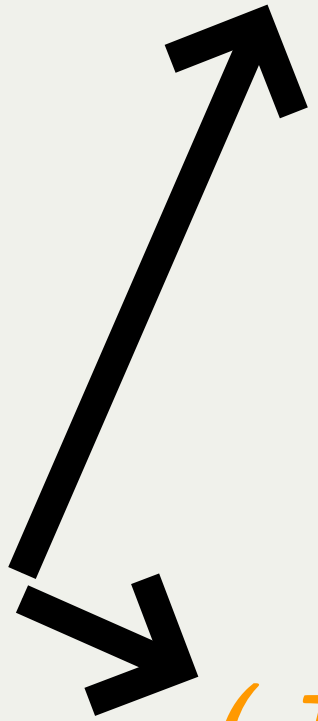
$$p(NH|D) \geq \alpha$$

prediction is likely
Alternative holds
Null rejected

Null
Hypothesis
Rejection
Testing

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

prediction is unlikely
Alternative rejected
Null holds



4

test data against
alternative outcomes

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

prediction is likely
Alternative holds
Null rejected

formulate the Null as the comprehensive opposite of your theory

model → **prediction**

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

this has a low probability of happening

data → **does not falsify alternative** → **model rejected**

data → **falsifies alternative** → **model still holds**

low probability event happened

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.

- *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 **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

central limit theorem

null hypothesis rejection testing setup

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

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 ($p < .05$)

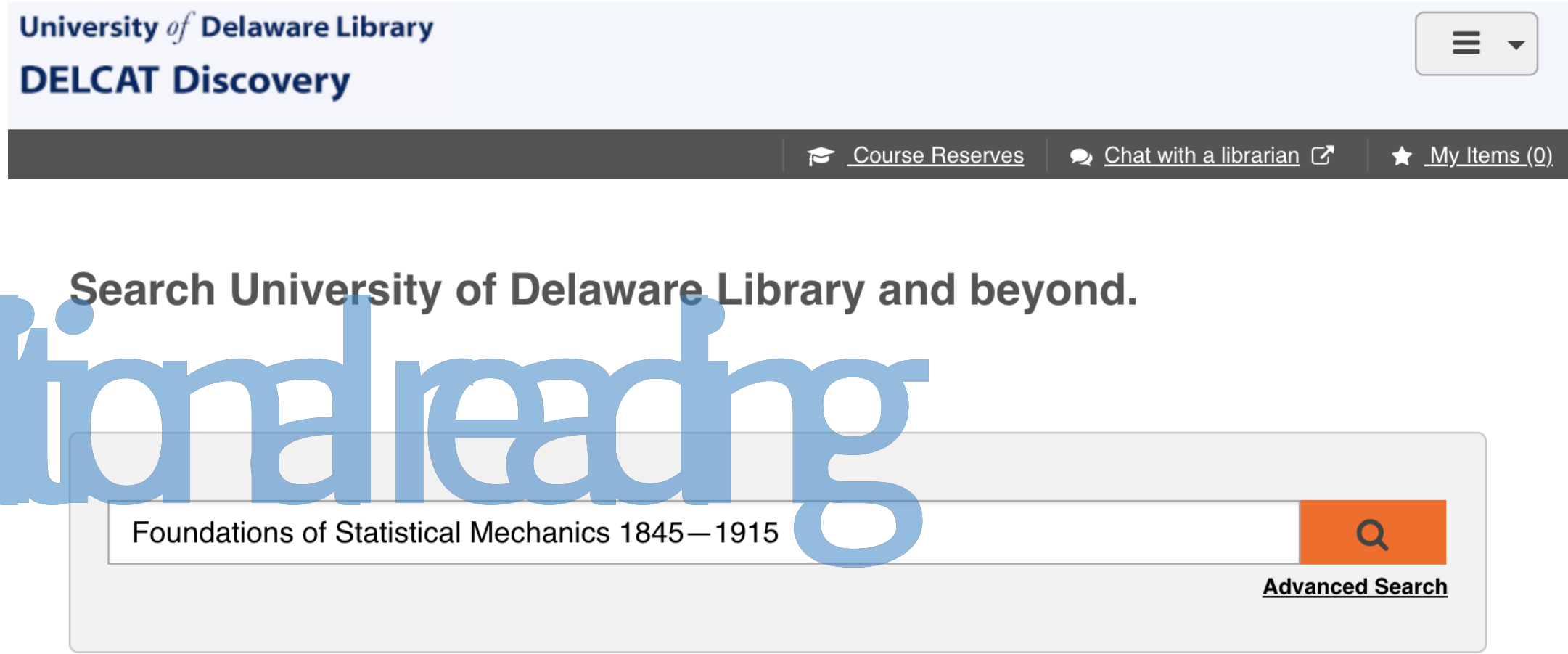
Jacob Cohen

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



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