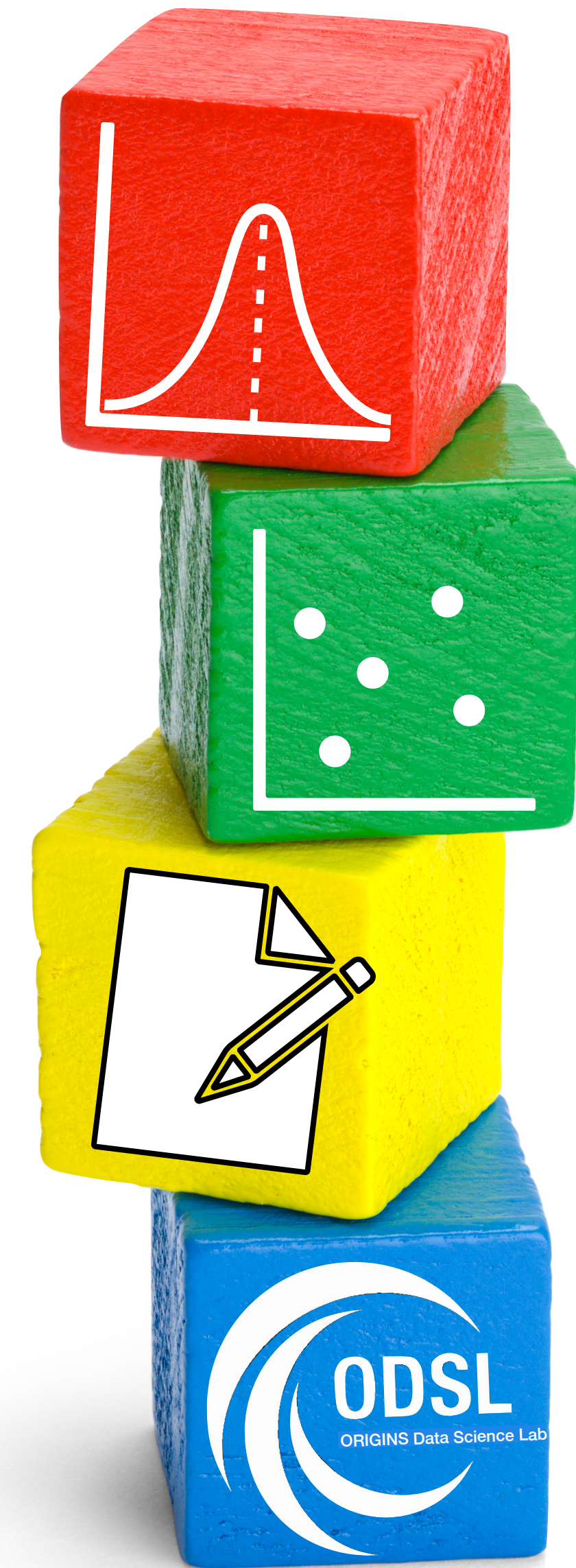


ODSL Statistics Block Course

Lecture 1: When the likelihood is tractable

Nicole Hartman
nicole.hartman@tum.de

22 Sept 2025



Welcome



Origins Data Science Lab



Led by Lukas Heinrich



Chair for Data science & Physics
Physics + CS dept, TUM
l.heinrich@tum.de

Who are we?

Nicole Hartman

Lecturer week 1
nicole.hartman@tum.de



Jonas Würzinger

Lecturer week 2
jonas.wuerzinger@tum.de



Nadine Bourriche

Tutor
nadine.bourriche@mpp.mpg.de

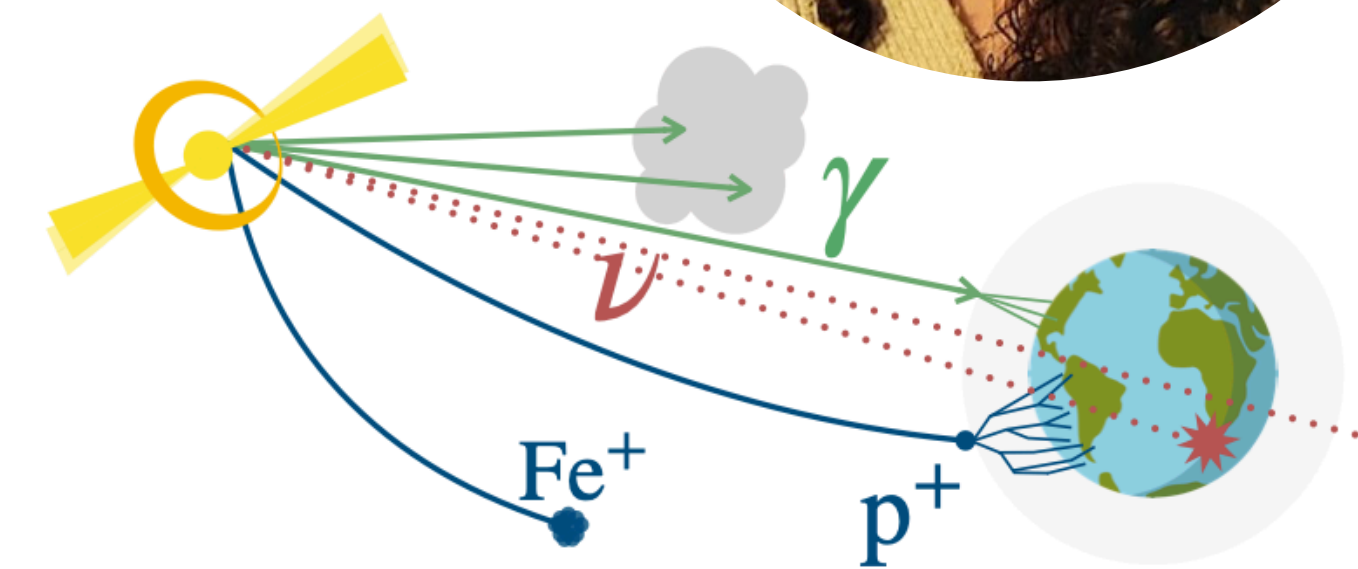
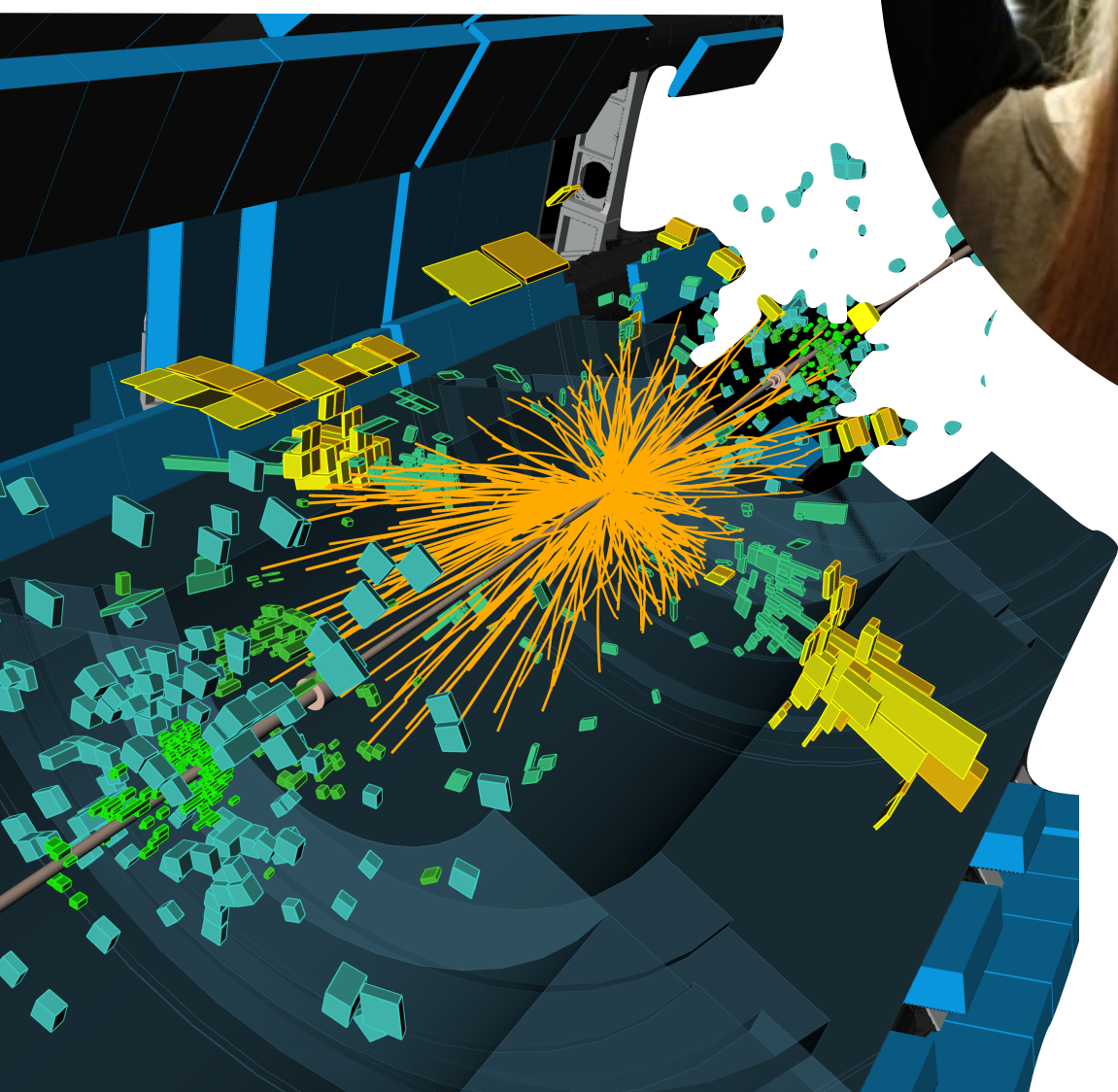


Fig courtesy of [F. Capel](#)

Our plan

Week 1: Probabilistic reasoning
Bayesian formalism

Week 2: Statistical inference
(Mostly) Frequentist formalism

Homework

If you're taking the courses for ECTS points

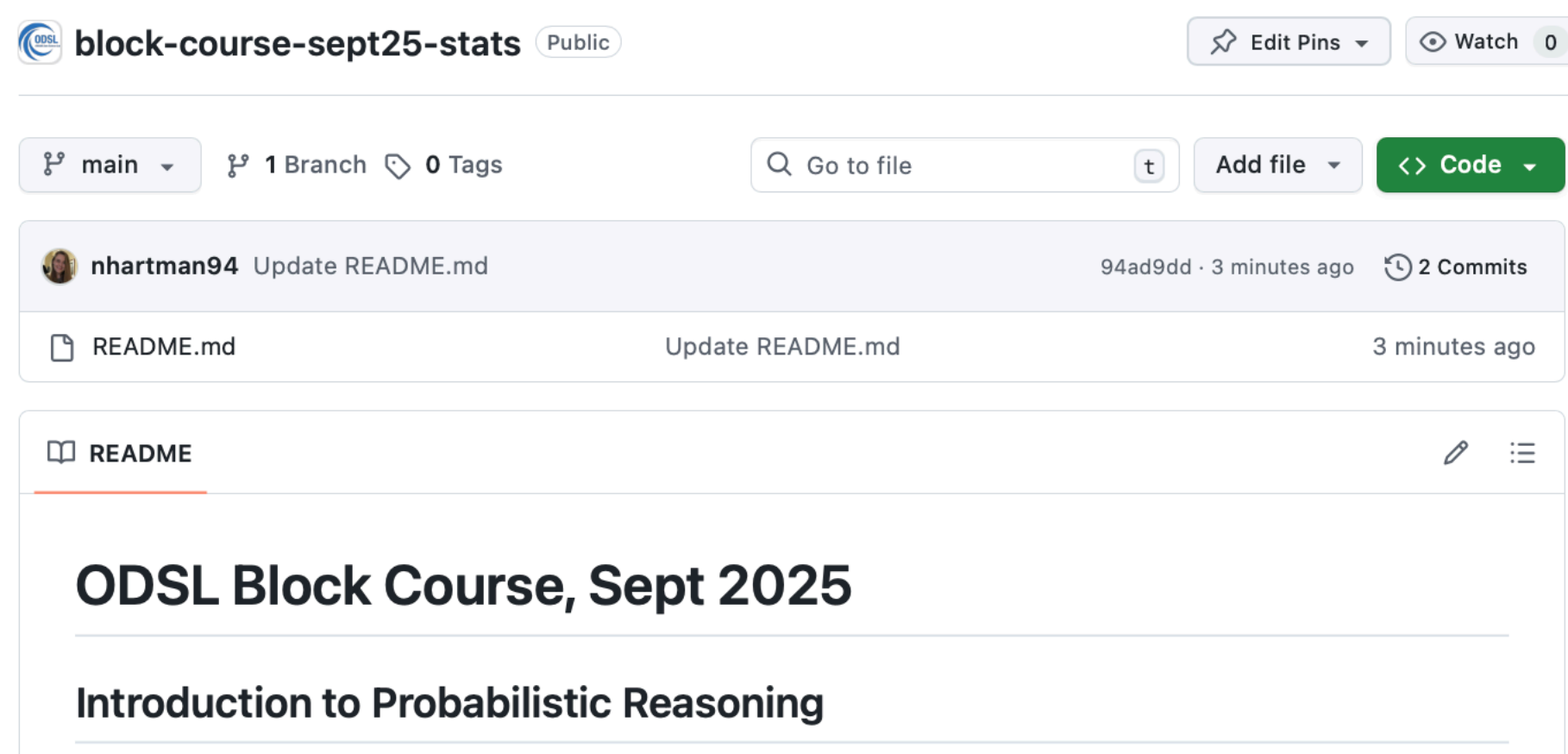
Homework due Fri, 17/10/2025

(2 weeks after end of lectures)

Please submit via email to me and Jonas:

- nicole.hartman@tum.de
- jonas.wuerzinger@tum.de

<https://github.com/odsl-team/block-course-sept25-stats/tree/main>



The screenshot shows a GitHub repository page for 'block-course-sept25-stats'. The repository is public and has 1 branch (main) and 0 tags. The commit history shows a recent commit by 'nhartman94' titled 'Update README.md' with commit hash '94ad9dd' and '2 Commits' made '3 minutes ago'. The file list shows 'README.md' with a recent update. The README content is visible, starting with 'ODSL Block Course, Sept 2025' and 'Introduction to Probabilistic Reasoning'.

Teaching philosophy

Passive Learning



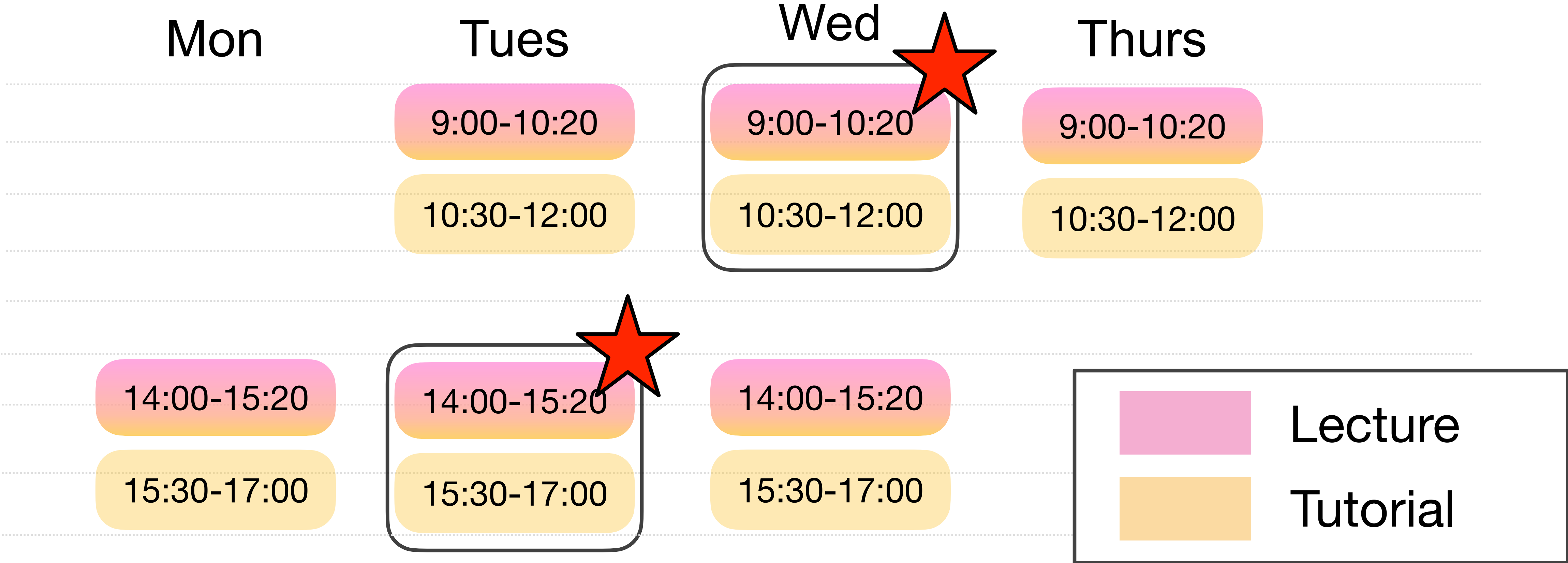
Image by [pngtree.com](https://www.pngtree.com)

Active Learning



Week 1 plan

**Two sessions in new room
(ORIGINS basement) !**

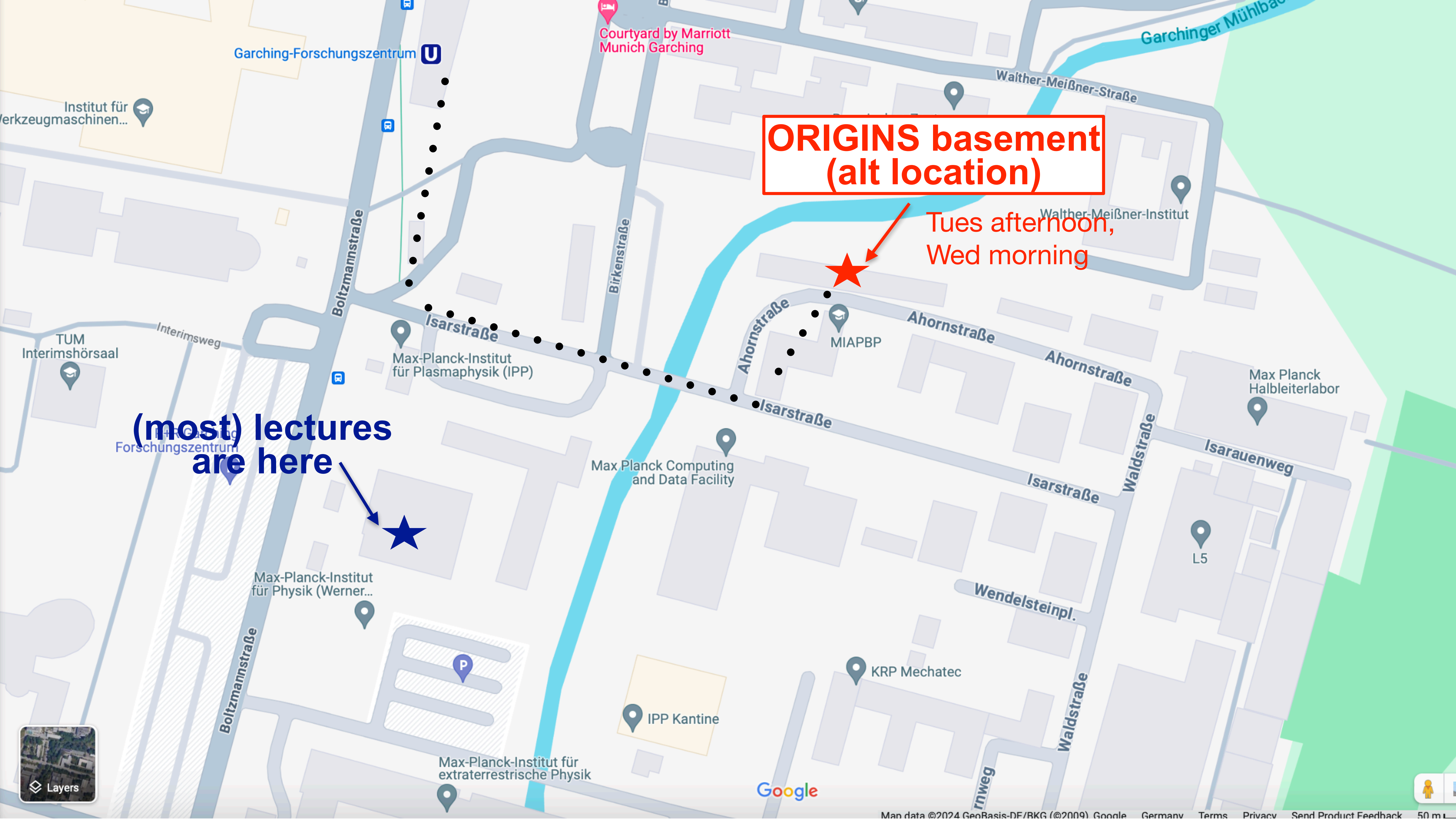


I will email the room switch reminder

**ORIGINS basement
(alt location)**

Tues afternoon,
Wed morning

**(most) lectures
are here**

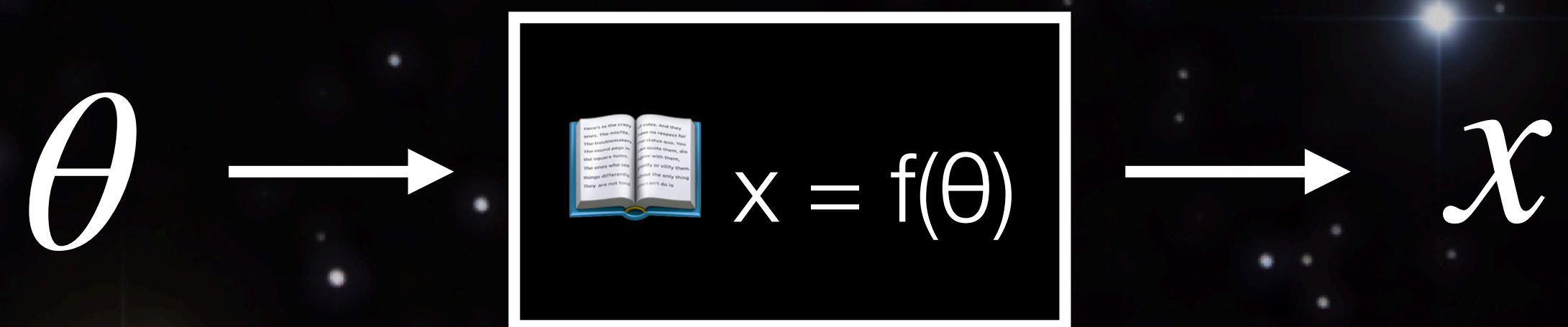


Agenda: Session 1

- Why statistics
- Prob 101
- When the likelihoods are tractable
 - Bernoulli; Binomial
 - Beta

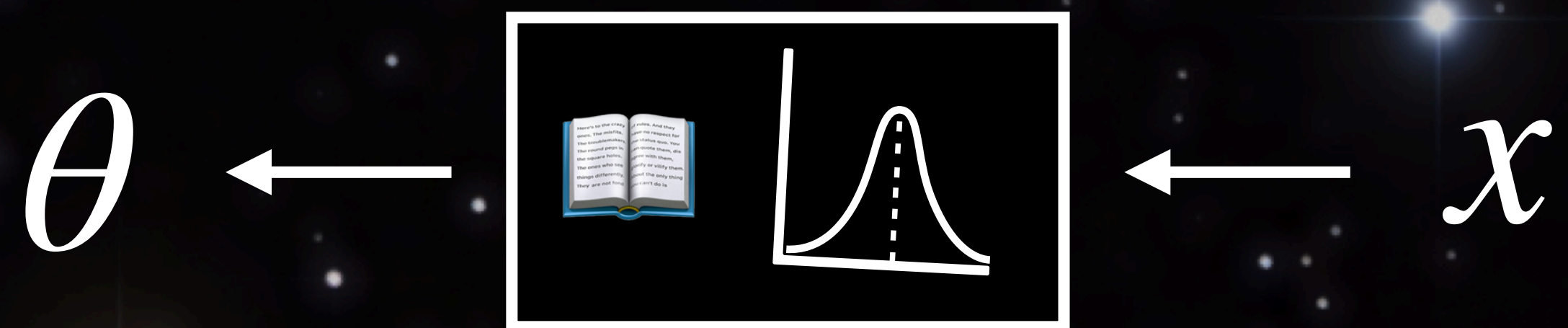
“The book nature is written in the language of mathematics.”

– Galileo

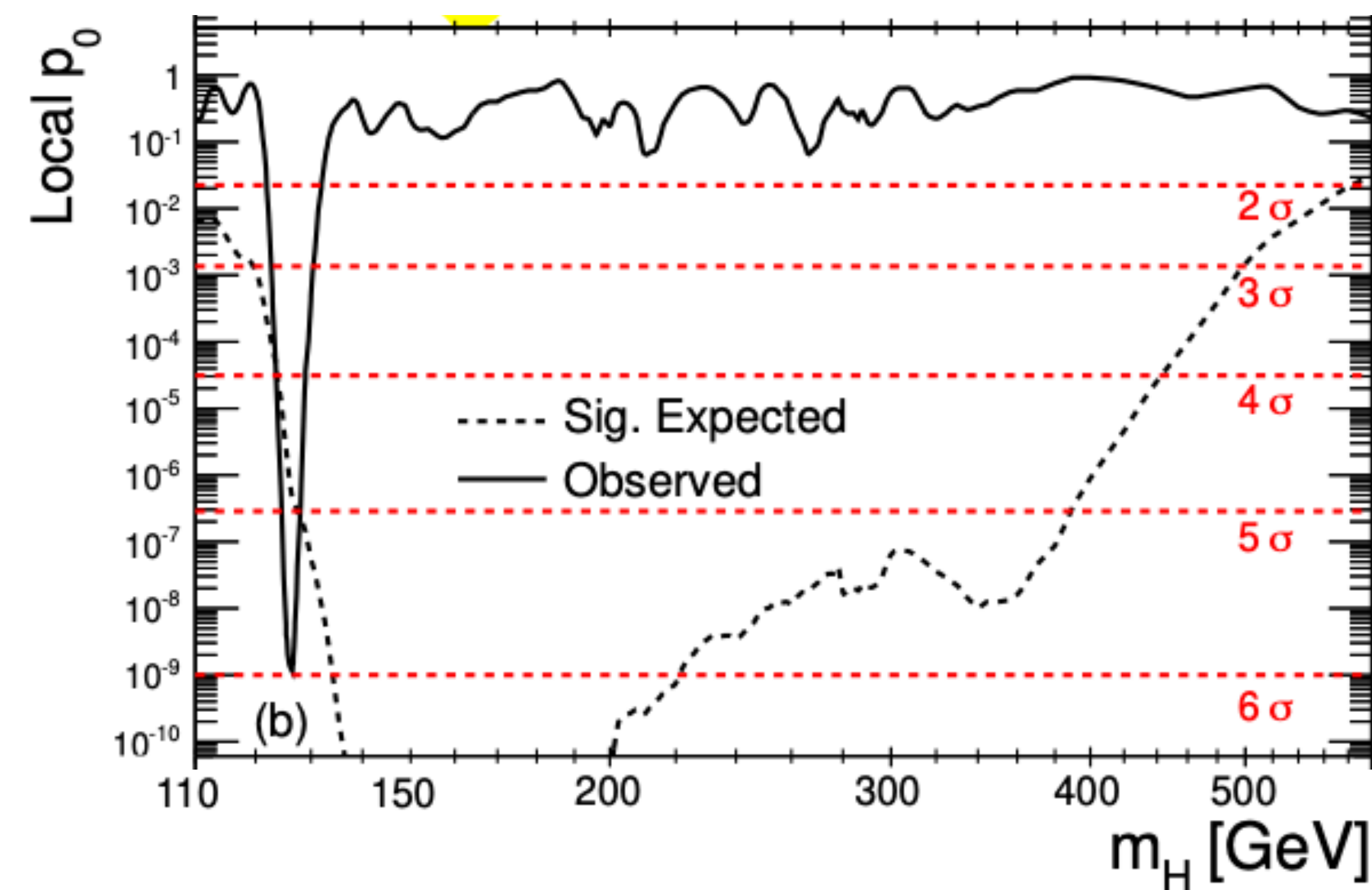
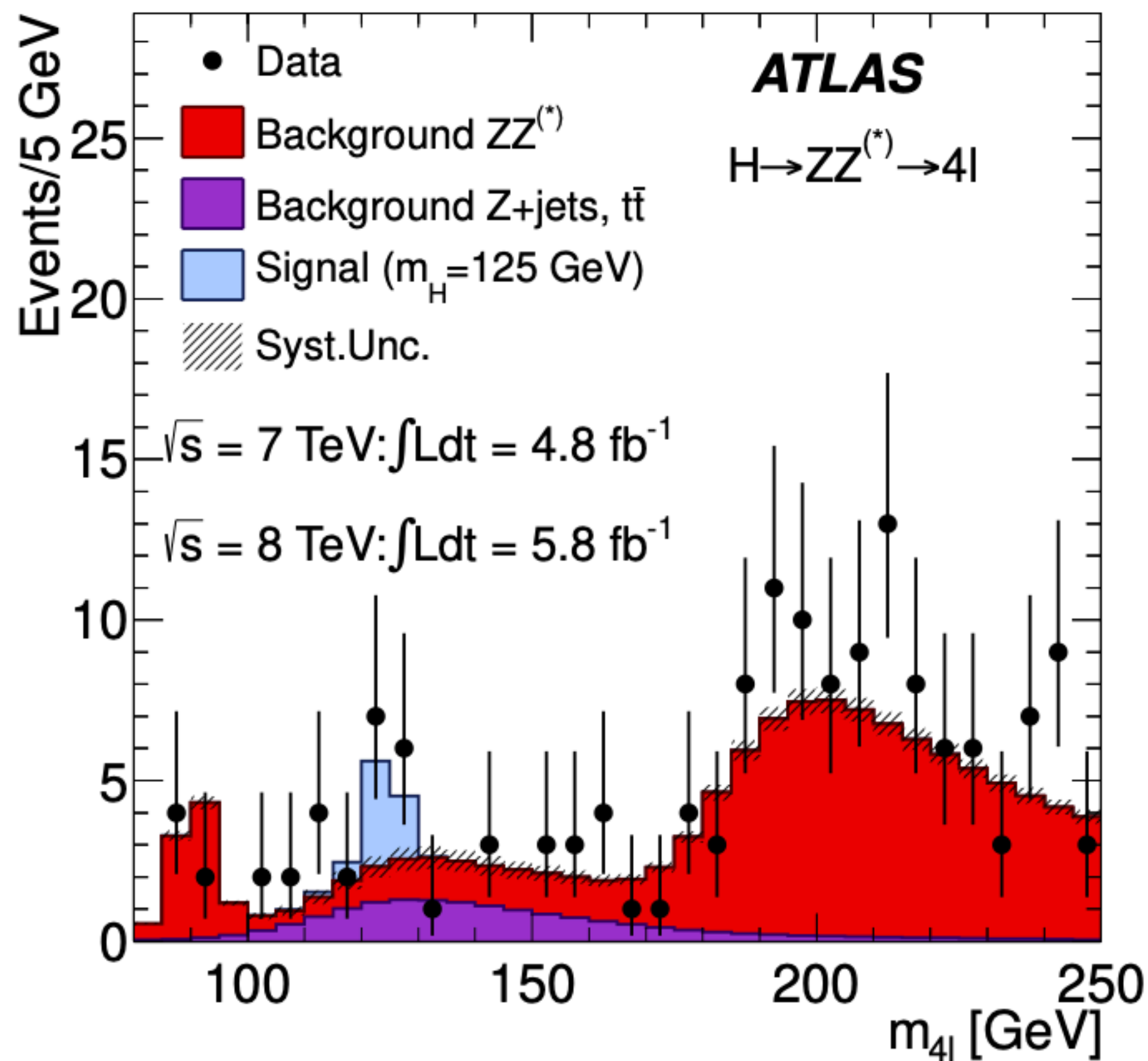


“The book of how we understand nature is in the language of Statistics.”

– Ariel Schwartzman



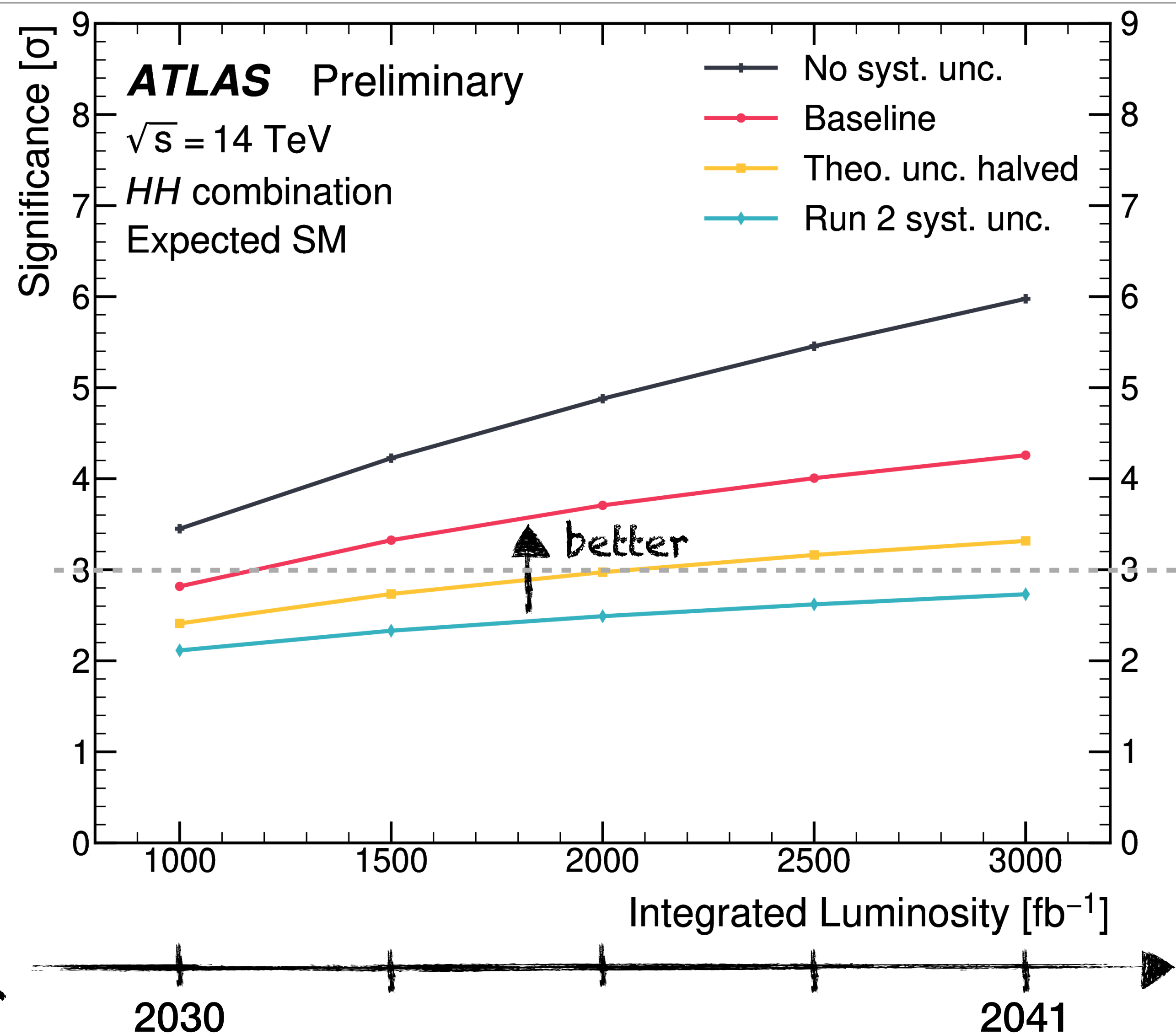
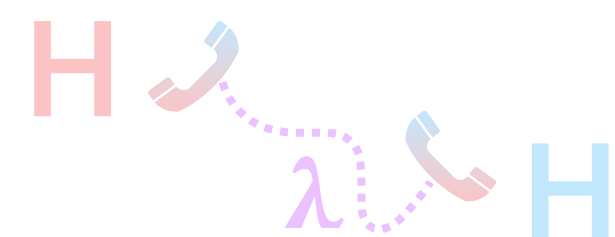
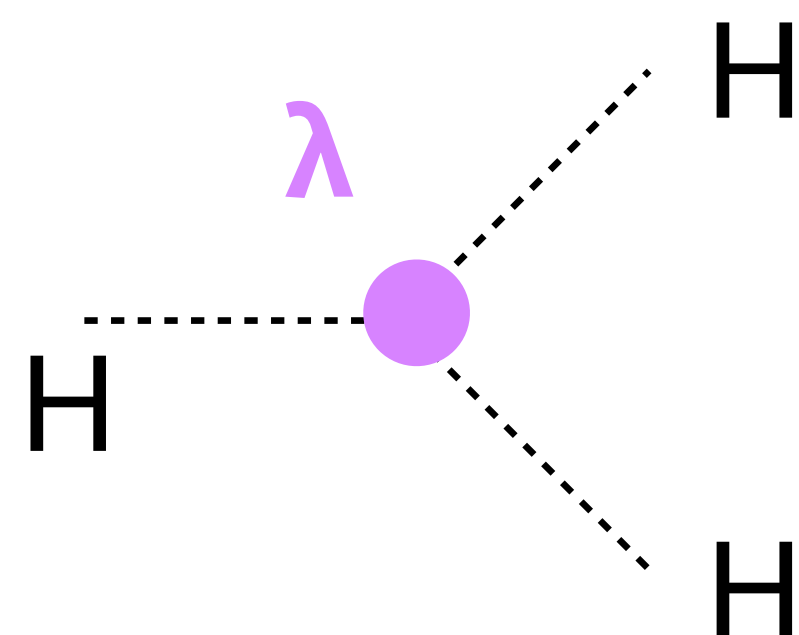
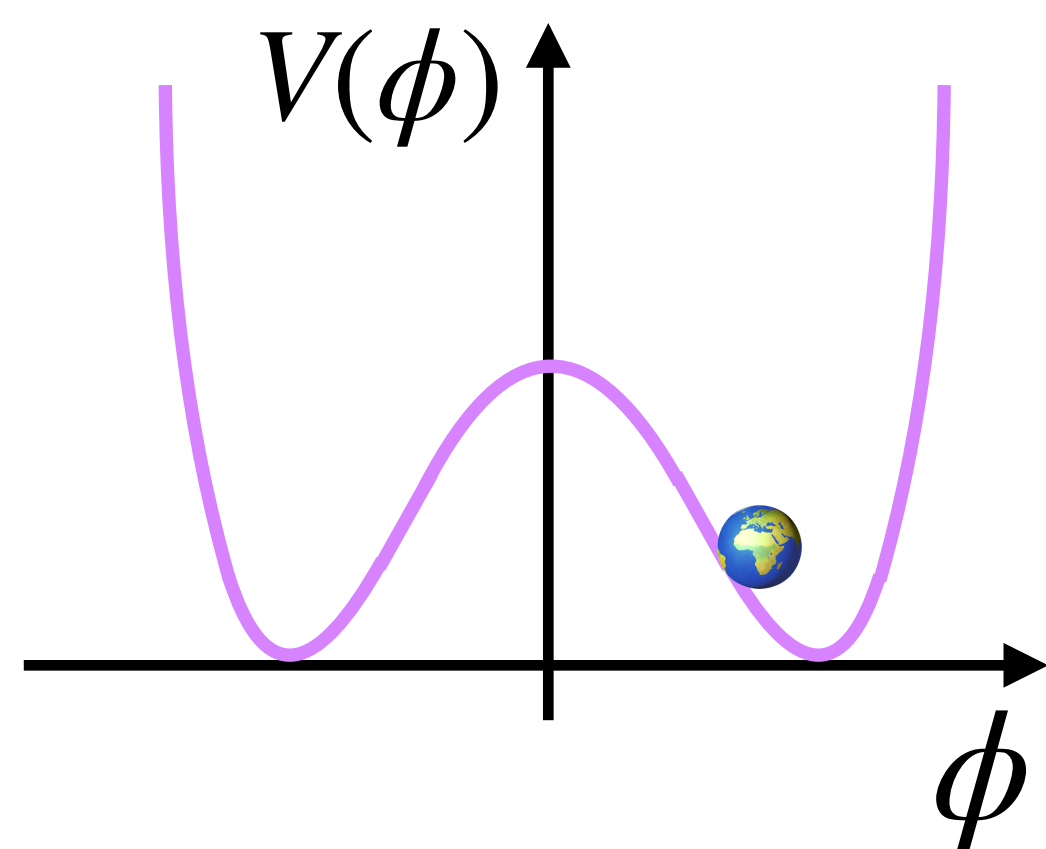
Have I found the Higgs?



Poisson probabilities covered tomorrow!

Hypothesis testing with Jonas next Tues!

When will I understand the Higgs?



Where do the highest energy particles come from?

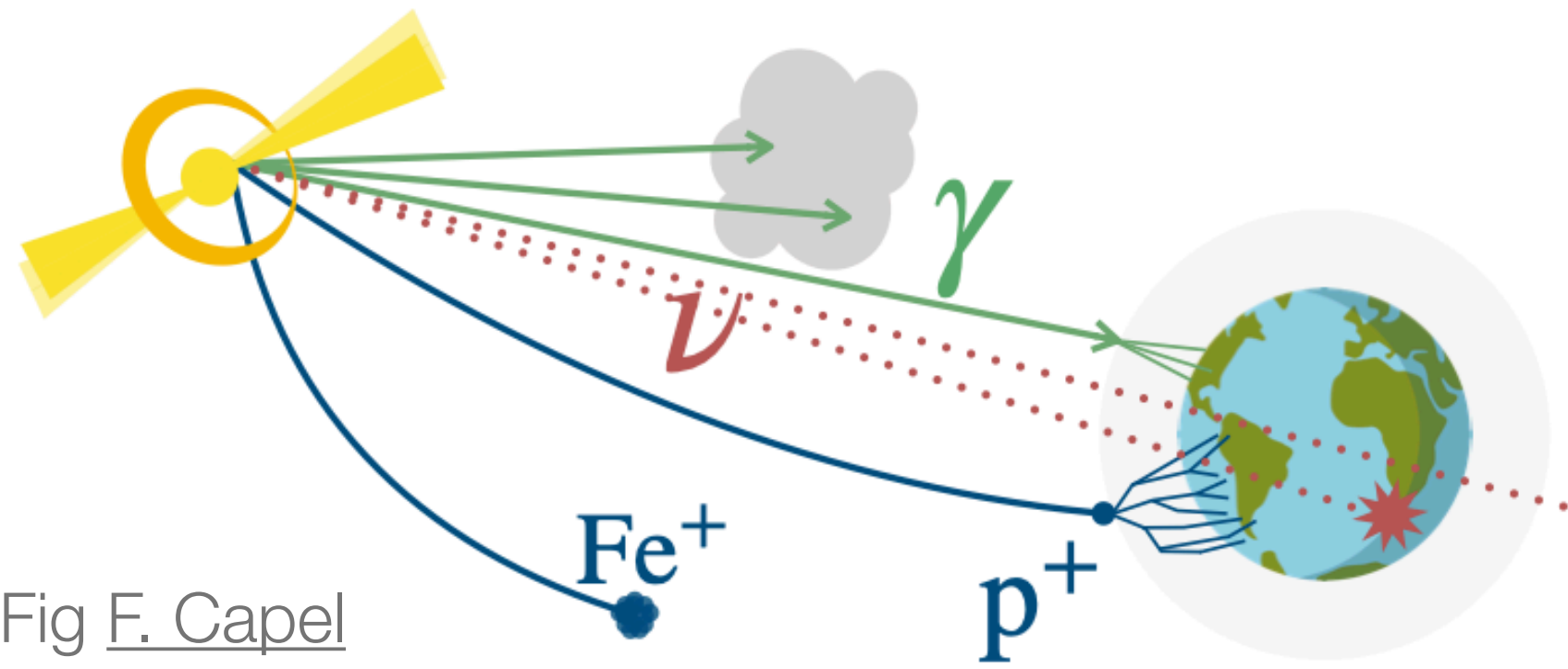
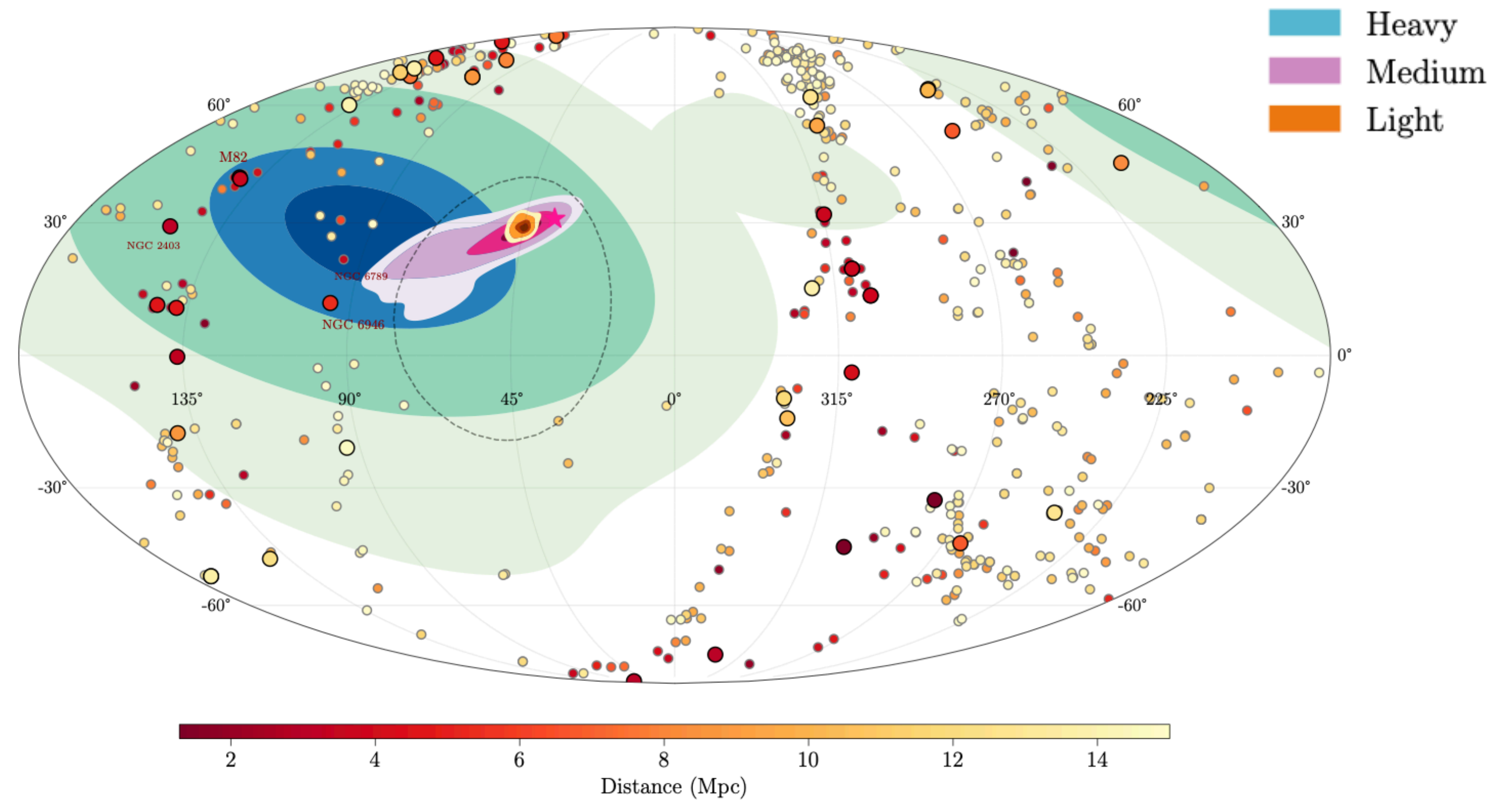
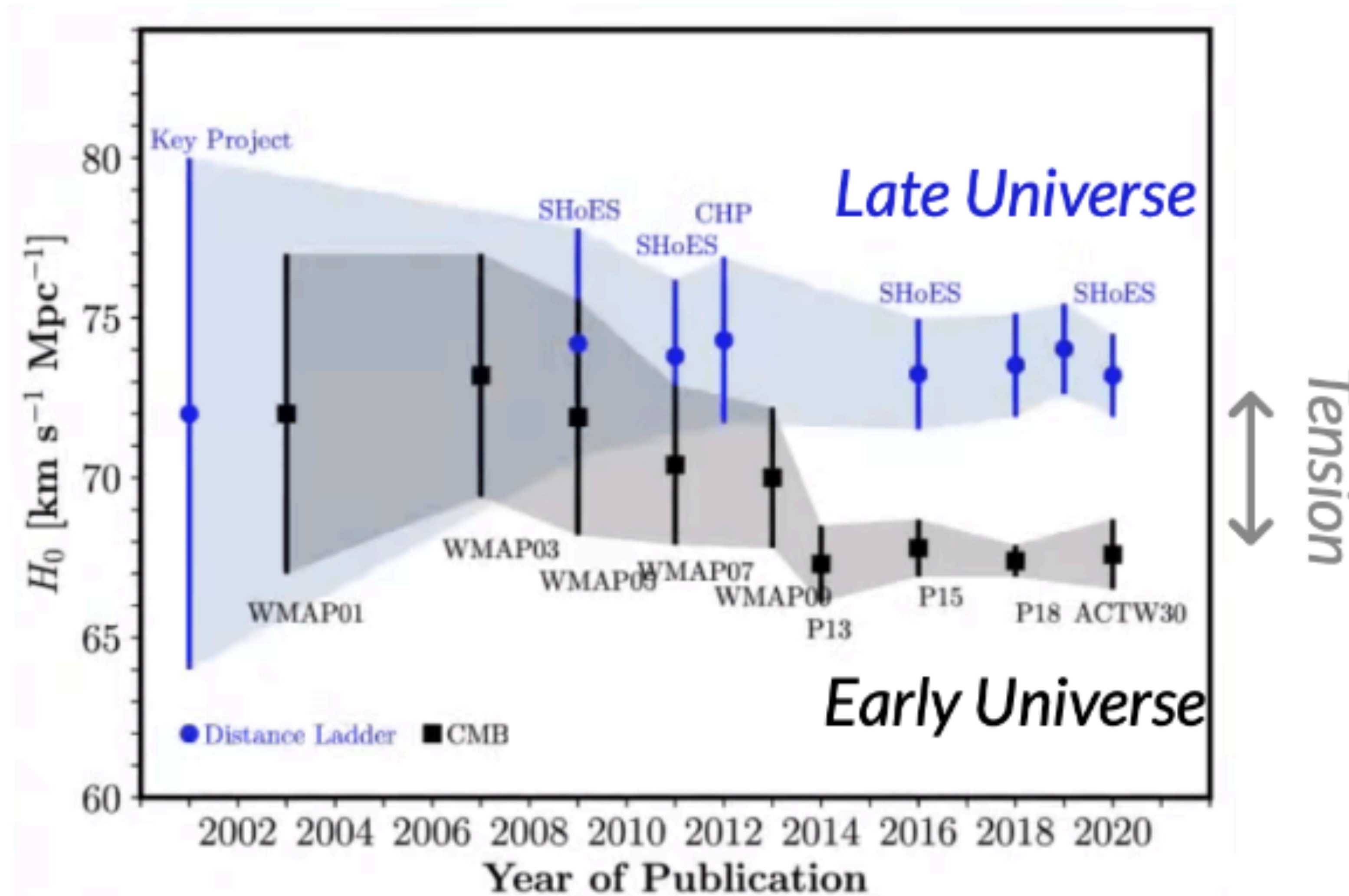


Fig F. Capel



We'll cover SBI and the ABC method on Thursday!

Is our understanding of the universe consistent?

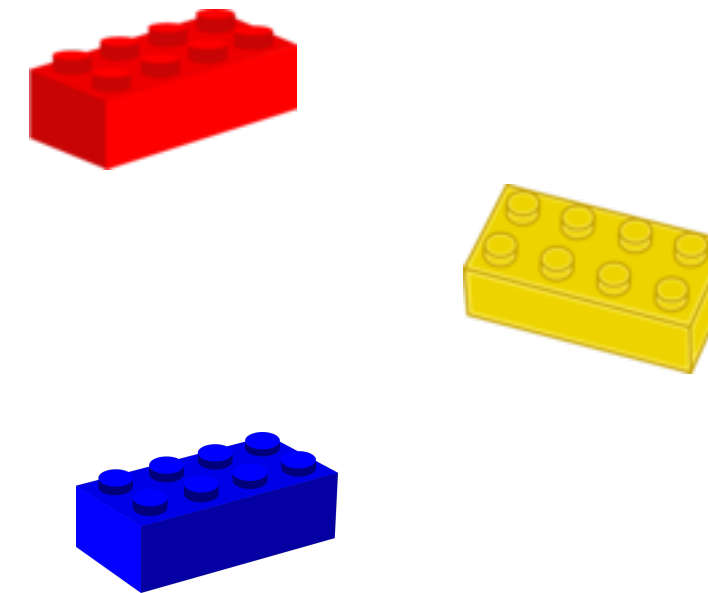


Machine Learning

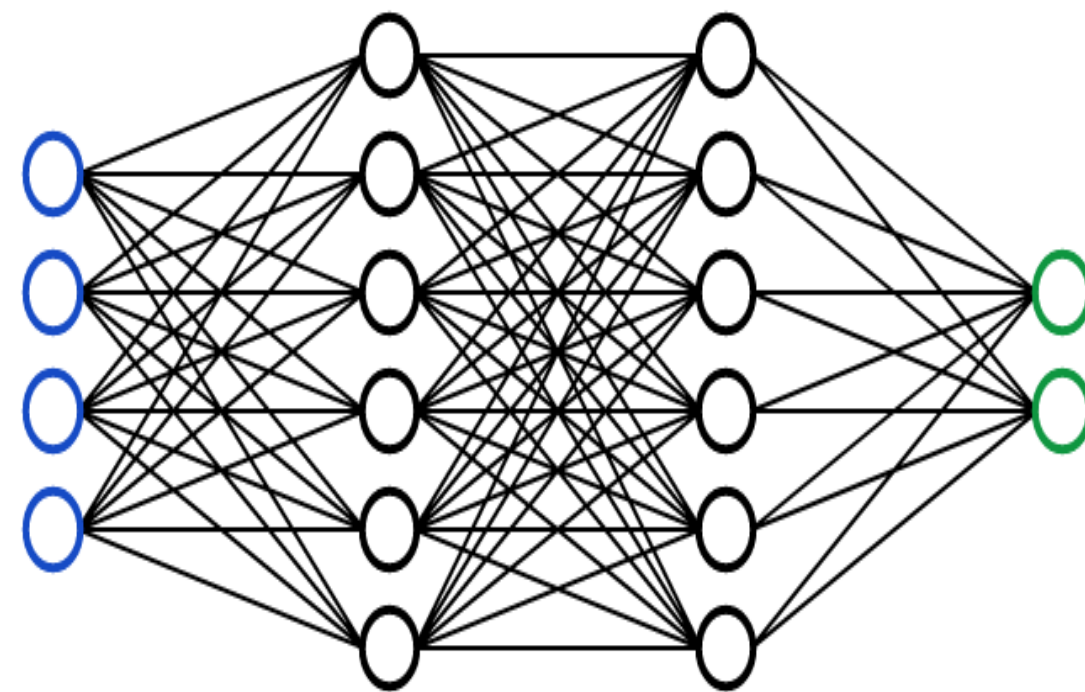
🔑 Probabilistic description data

$$p(x | \theta)$$

data



Benefits from big data



Drives the advancements of deep NNs



More on Thurs how ML can modernize your inference pipelines! (SBI)

Who are you?



[PollEv.com /nicolehartman968](https://www.pollEv.com/nicolehartman968)

Course syllabus



Week 1:

Analytical probabilities (Mon & Tues)

Tractable inference (Tues)

- Conjugate priors
- Model comparison

When inference isn't tractable (Wed & thurs)

- Hierarchical graphical models
- Monte Carlo Sampling
- Markov Chain Monte Carlo
- Variational Inference
- Intro to SBI

Week 2:

Bayesian vs. Frequentist

Parameter estimation

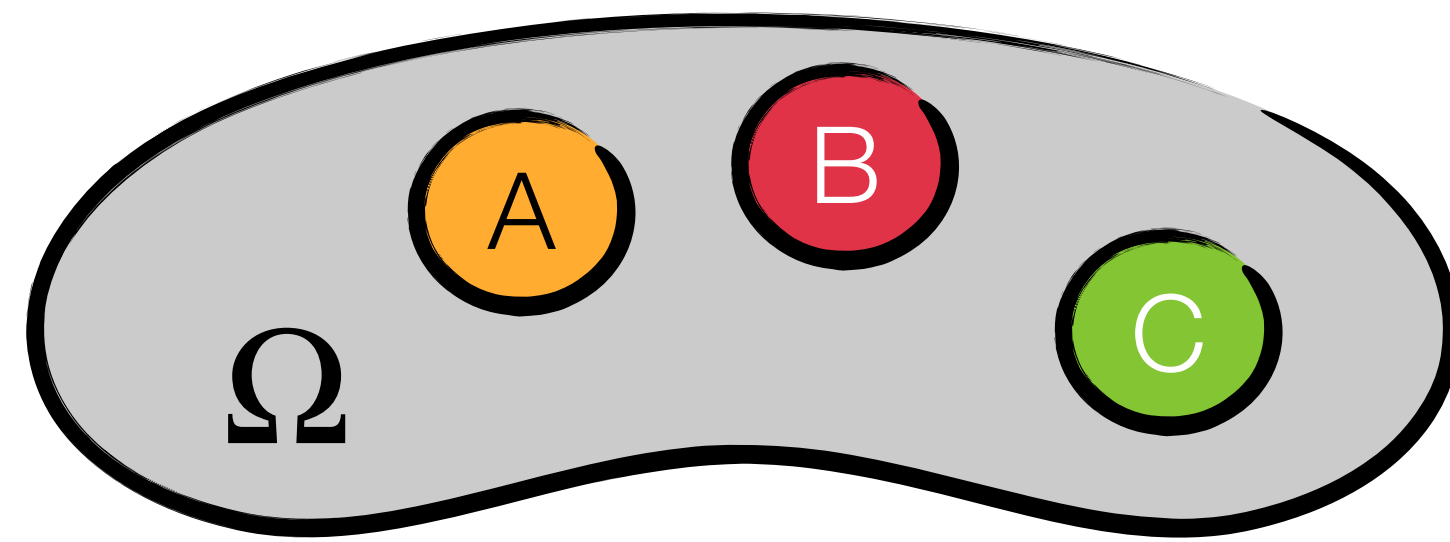
- MLE
- Confidence intervals

Hypothesis testing

Agenda: Session 1

- Why statistics
- Prob 101
- When the likelihoods are tractable
 - Bernoulli; Binomial
 - Beta

Sample space: Ω




Defines the set of possible outcomes

Examples:

Flipping a coin:

elementary event


$$\Omega = \left\{ \text{heads}, \text{tails} \right\}$$

Rolling a dice:

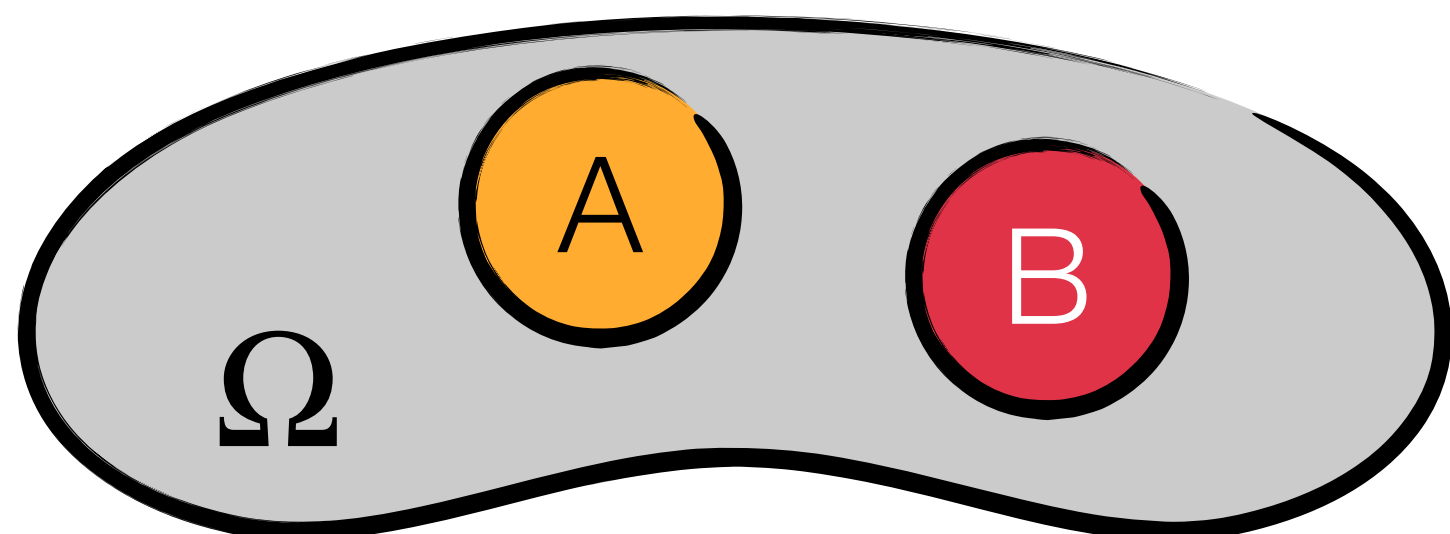
$$\Omega = \left\{ 1, 2, 3, 4, 5, 6 \right\}$$

What does a probability need to satisfy?

Kolmogorov's axioms

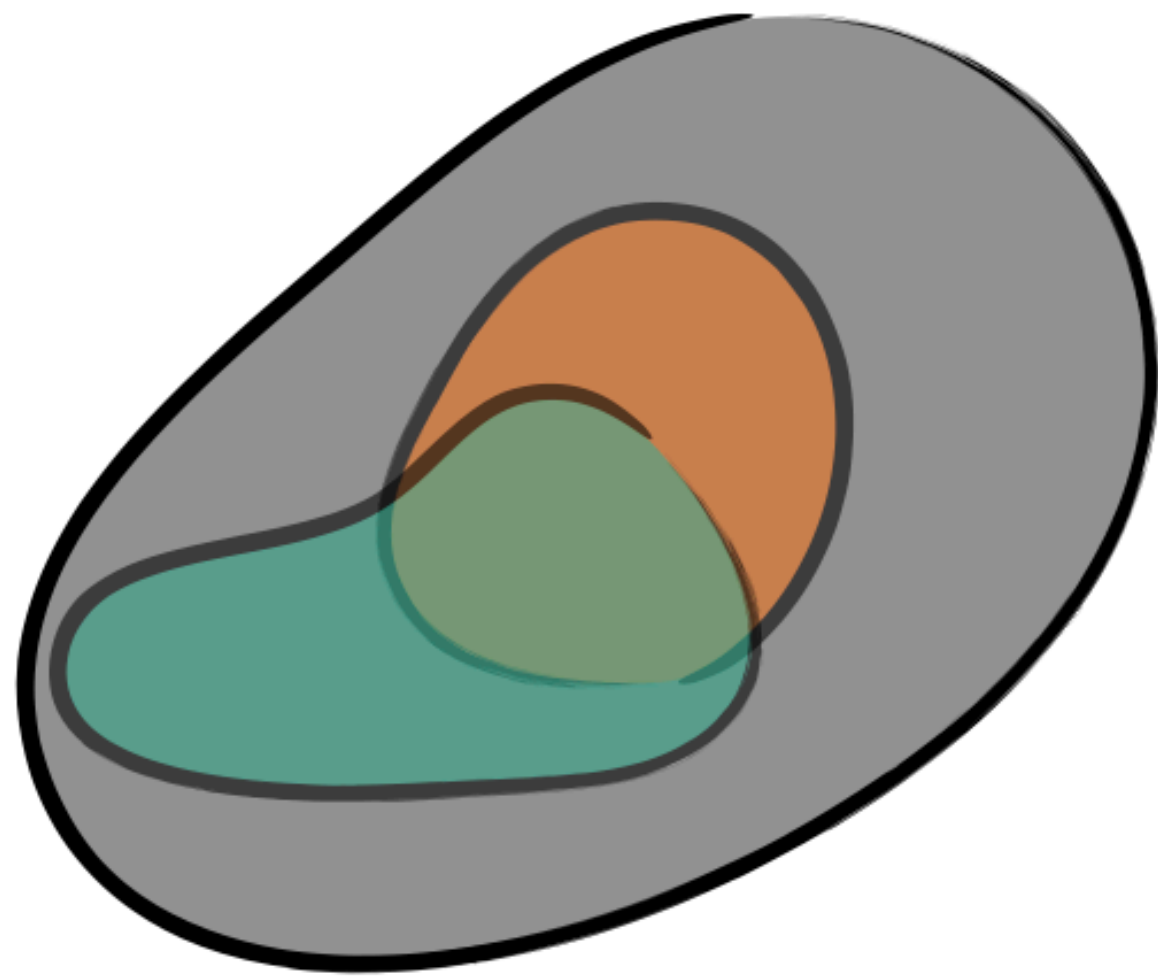
For state space Ω with a set of **events** $\{A, B, C, \dots\}$, a probability distribution p must satisfy:

1. Non-negativity $p(A) > 0$
2. Sum to unity $\sum_{s \in \Omega} p(s) = 1$
3. Disjoint events sum to 1 $p(A \cup B) = p(A) + p(B)$



The joint

Since events are sets of elementary events, we can use standard set operations and think about the effects on probability

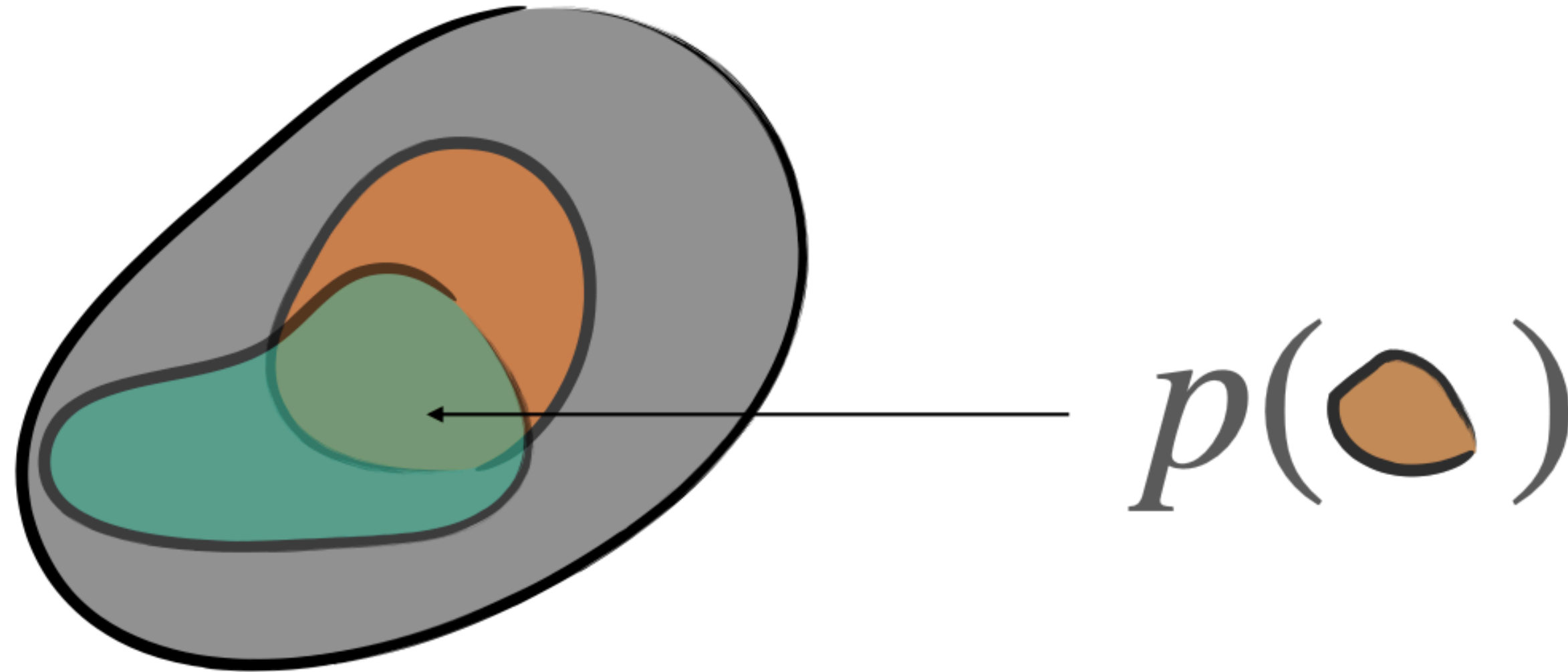


$$p(\text{teal}) + p(\text{orange}) - p(\text{brown})$$

$$p(A \text{ or } B) = p(A \cup B) = p(A) + p(B) - p(A \text{ and } B)$$

The joint

Since events are sets of elementary events, we can use standard set operations and think about the effects on probability



$$p(A \text{ and } B) = p(A \cap B)$$

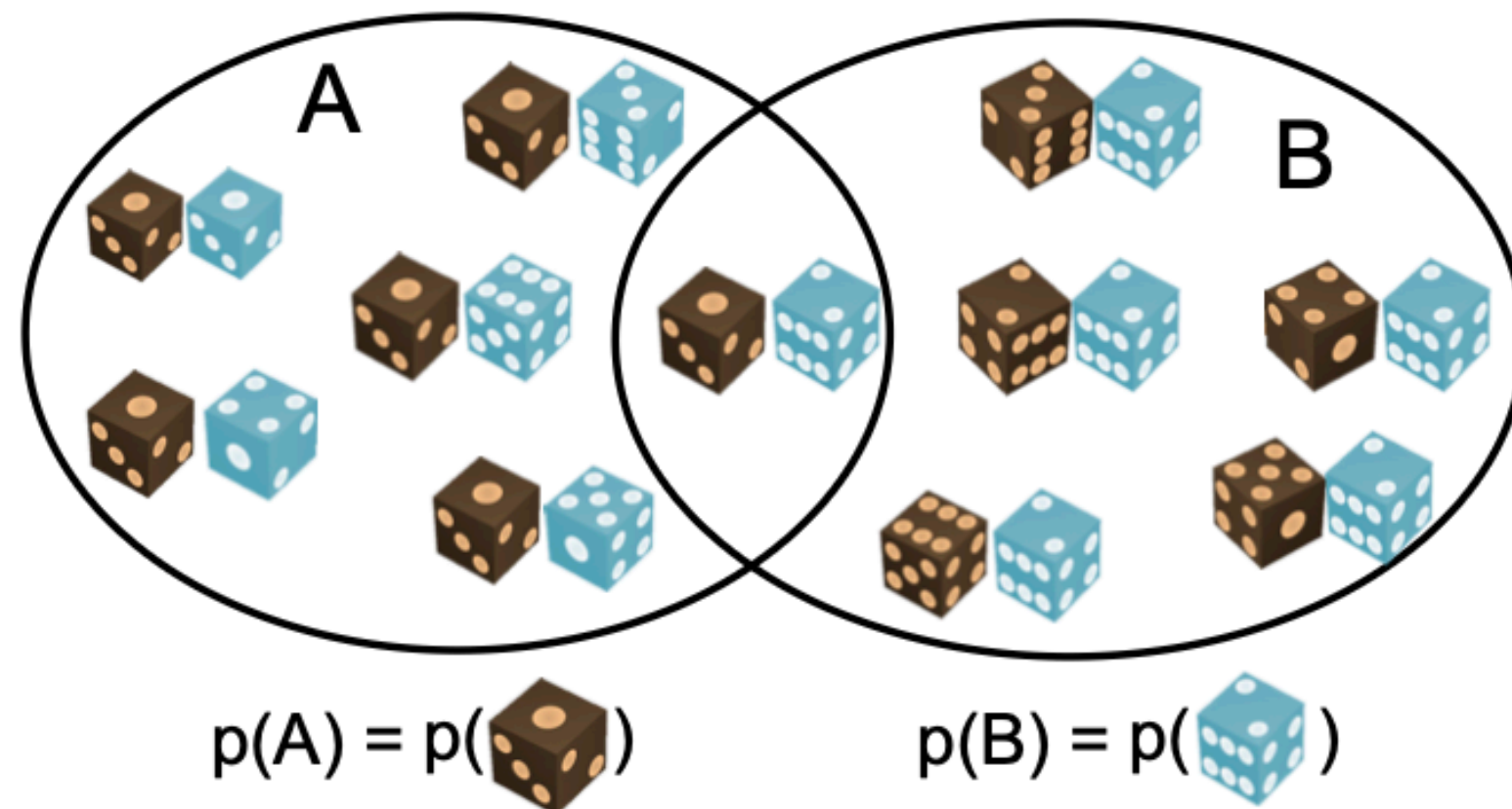
Independence of events

Two events A and B are **independent** if

$$p(A \cap B) = p(A)p(B)$$

$\Omega = \{ \text{die}_1, \text{die}_2, \dots, \text{die}_6 \}$ 2 die tosses
 $6^2 = 36$ outcomes

Are these two events independent?



Yes!

$$p(A \cap B) = p(\text{die } 1, \text{die } 2)$$

$$p(A)p(B) = p(\text{die } 1) p(\text{die } 2)$$

Random variables

Translates elementary events into real numbers

Examples:

Flipping a coin:

$$\Omega = \left\{ \text{heads}, \text{tails} \right\} \longrightarrow \{ 1, 0 \}$$

Rolling a dice:

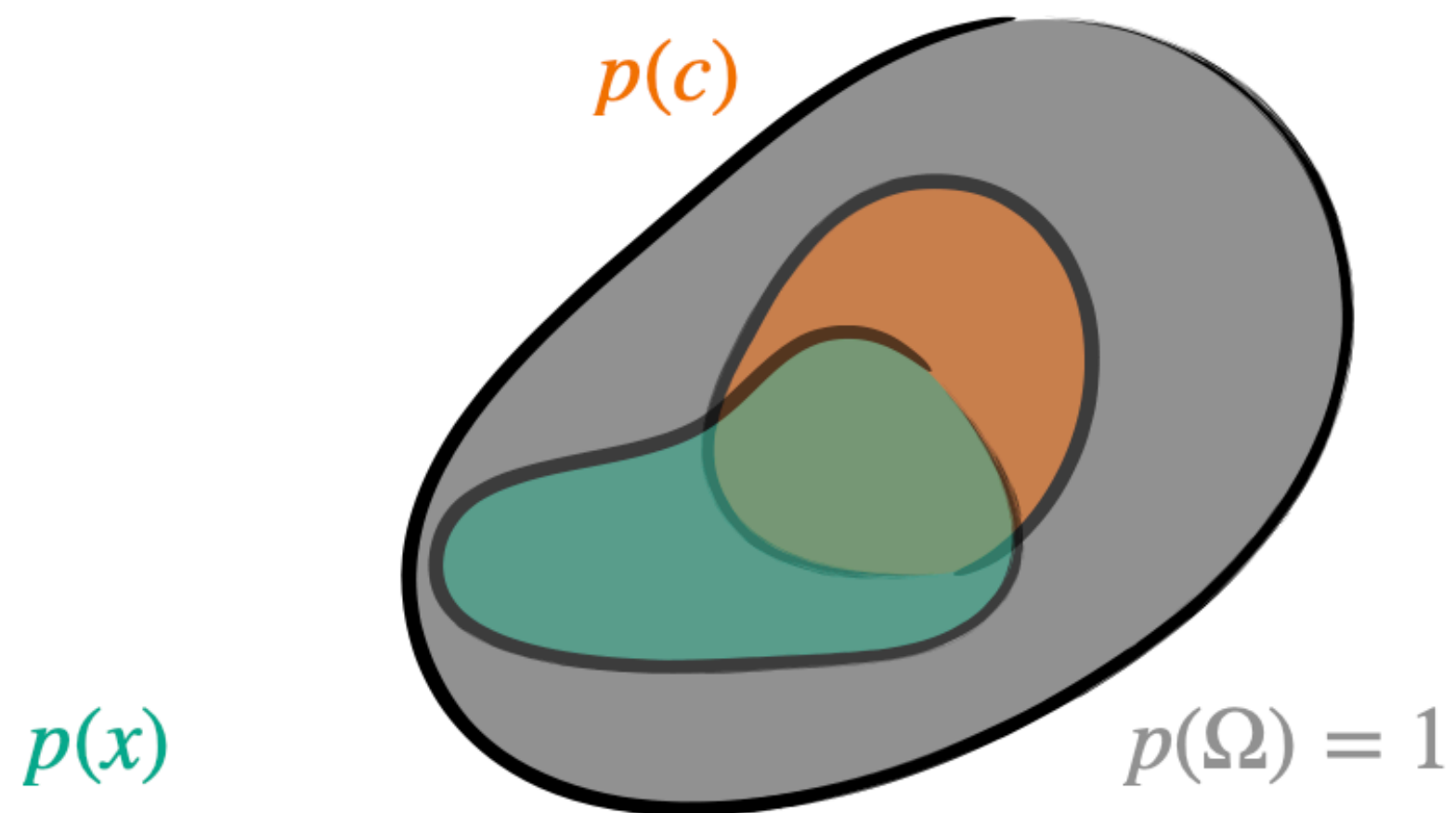
$$\Omega = \left\{ \text{1 dot}, \text{2 dots}, \text{3 dots}, \text{4 dots}, \text{5 dots}, \text{6 dots} \right\}$$
$$\{ 1, 2, 3, 4, 5, 6 \}$$

Conditional probability

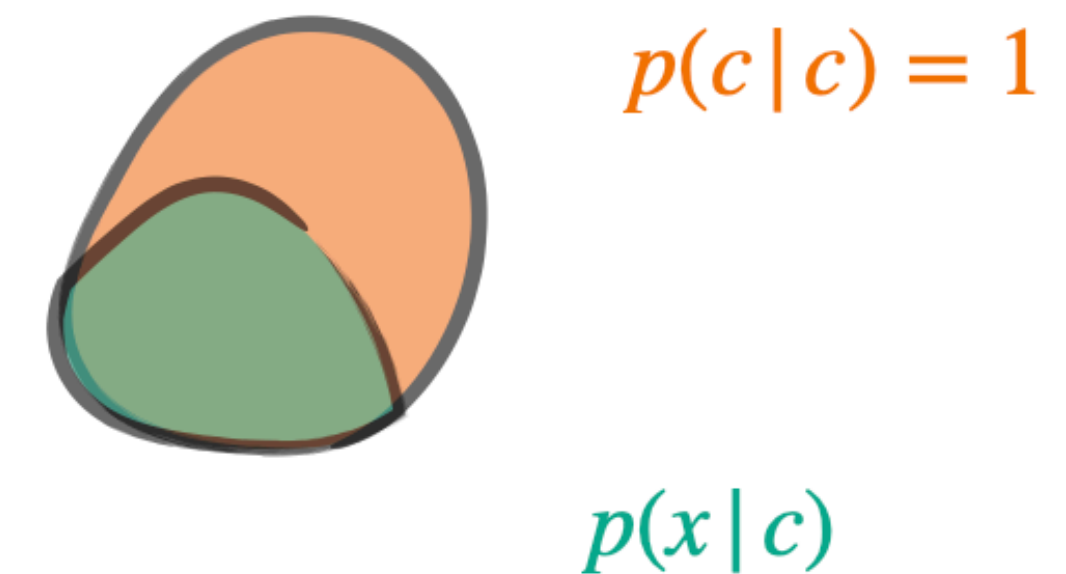
We can define the conditional probability as the probability of **event x** if we restrict our world to events where a **condition c** is true

$$p(c) \rightarrow \frac{p(c)}{p(c)} = 1$$

$$p(x) \rightarrow p(x|c) := \frac{p(x \text{ and } c)}{p(c)}$$



restrict world



Chain rule of probability

$$p(A, B) = p(A | B)p(B)$$

Marginalization

$$p(A) = \sum_{B \in \Omega_B} p(A, B) = \sum_{B \in \Omega_B} p(A | B)p(B)$$

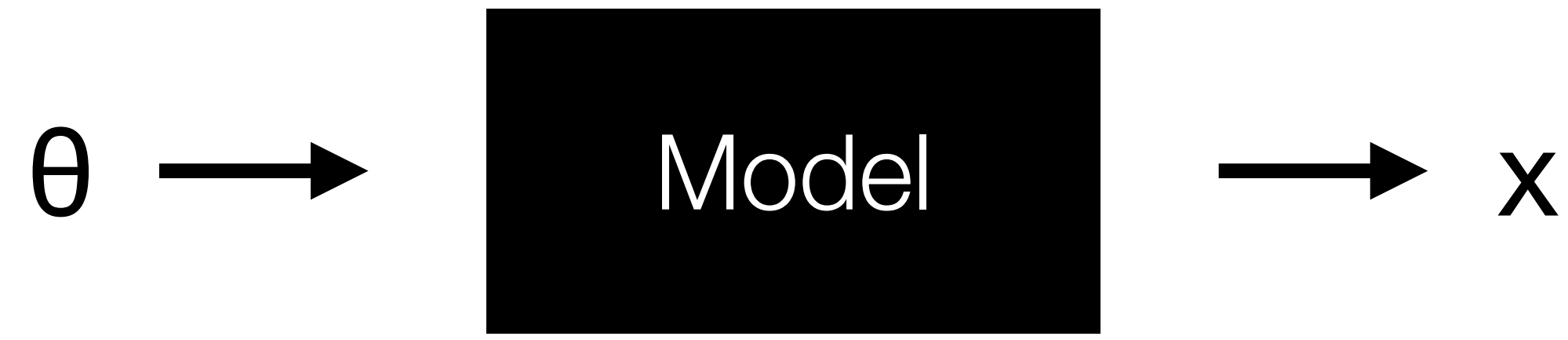
Ways to expand the joint

Bayes theorem

$$\begin{aligned} p(A, B) &= p(A | B) p(B) \\ &= p(B | A) p(A) \end{aligned}$$

Can reshuffle terms to get...

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



Bayes theorem

Likelihood Prior

$p(\theta | x) = \frac{p(x | \theta) p(\theta)}{p(x)}$

Posterior Evidence





Example: testing for rare disease

Given a positive test, how likely is it that I have the disease?

A : Test is positive

\bar{A} : Test is negative

B : Have disease

\bar{B} : Don't have disease

Test specifications:

- True positive rate, $p(A | B) = 0.97$
- False positive rate, $p(A | \bar{B}) = 0.01$

Rare disease: $p(B) = 0.005$



Soln: Testing for rare disease

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

$$p(A) = \sum_{B_*} p(A, B_*) = p(A, B) + p(A, \bar{B})$$

Kolmogrov's 3rd axiom

$$= p(A | B)p(B) + p(A | \bar{B})p(\bar{B})$$

Chain rule of prob

$$p(B | A) = \frac{0.97 \cdot 0.005}{0.97 \cdot 0.005 + 0.01 \cdot 0.995} = 33\%$$

Follow-up

What if the disease isn't so rare (e.g, during covid outbreak 🤒).

$$p(B) = 0.4$$

Sick society

$$p(B|A) = \frac{0.97 \cdot 0.4}{0.97 \cdot 0.4 + 0.01 \cdot 0.6} = 98\%$$

Healthy society

$$p(B|A) = \frac{0.97 \cdot 0.005}{0.97 \cdot 0.005 + 0.01 \cdot 0.995} = 33\%$$

Priors can have a big impact on the results!

Discussing this more tmr

Expectation values

Discrete distributions

$$\mathbb{E}_p[f(x)] = \sum_i p_i f(x_i)$$

Continuous distributions

$$\mathbb{E}_p[f(x)] = \int dx p(x) f(x)$$

Moments of the distribution

First moment (mean)

$$\mu = \mathbb{E}_p[x] = \int xp(x)dx$$

Second moment

$$\mathbb{E}_p [x^2] = \int x^2p(x)dx$$

Variance

$$\sigma^2 = \mathbb{E}_p [(x - \mu)^2] = \mathbb{E}_p [x^2] - \mathbb{E}_p[x]^2$$



Prove in homework

Agenda: Session 1

- Why statistics
- Course org & teaching philosophy
- Prob 101
- When the likelihoods are tractable
 - **Discrete:** Bernoulli; Binomial
 - **Continuous:** Gamma (exponential, χ^2); Beta

Bernoulli

Two classes: $\Omega = \{0,1\}$

Ex:



p = probability of “success”, and one trial.

$$p(k) = \begin{cases} p, & k = 1 \\ 1 - p, & k = 0 \end{cases} = \text{Can we make this formula more concise?}$$

A) $p^k \cdot (1 - p)^{1-k}$

B) $p^k + (1 - p)^{1-k}$

C) $p^k - (1 - p)^{1-k}$

D) $k \cdot p - (1 - k)(1 - p)$

[PollEv.com/nicolehartman968](https://www.pollEv.com/nicolehartman968)



Bernoulli

Two classes: $\Omega = \{0,1\}$

Ex:

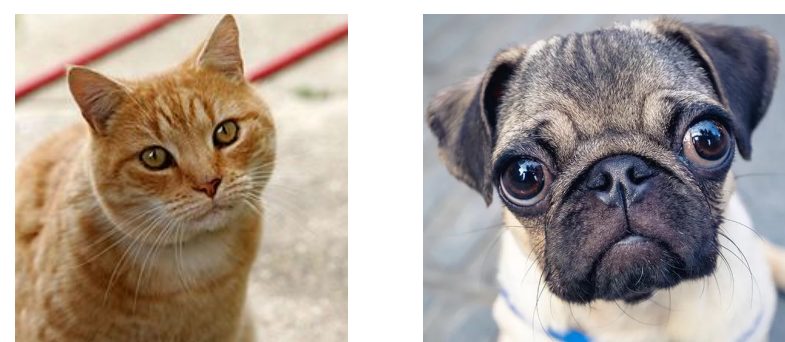


p = probability of “success”, and one trial.

$$p(k) = \begin{cases} p, & k = 1 \\ 1 - p, & k = 0 \end{cases} = p^k \cdot (1 - p)^{1-k}$$

Fun fact! In ML the Bernoulli distribution is used for the 2 class classification, “binary cross entropy”

$x = [\text{cat}, \text{dog}]$



Your turn

For the Bernoulli distribution, what's the mean and variance of the distribution?

PollEv.com/nicolehartman968



Your turn

For the Bernoulli distribution, what's the mean and variance of the distribution?

$$\mu = \mathbb{E}_p[k] = \sum_{x \in \{1,0\}} k p(k)$$

$$= 1 \cdot p + 0 \cdot (1 - p) = p$$

To find the variance, first solve for the second moment...

$$\mathbb{E}_p[k^2] = \sum_{k \in \{1,0\}} k^2 p(k) = 1^2 \cdot p + 0^2 \cdot (1 - p) = p$$

Putting the pieces together

$$\sigma^2 = \mathbb{E}_p[k^2] - \mathbb{E}_p[k]^2 = p - p^2 = p(1 - p)$$

Binomial

Two outcomes (**success** / **failure**), p = probability of success, **n trials**, and x is the # of successes. $\Omega = \{0, 1, \dots, n\}$

Ex:



Binomial “event” x is the sum of n independent Bernoulli trials ($x_1, x_2, \dots, x_n: x_i \in \{0,1\}$)

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Recall: for independent events x and y , probabilities multiply.



Combinatorial factor

x successes

$n - x$ failures

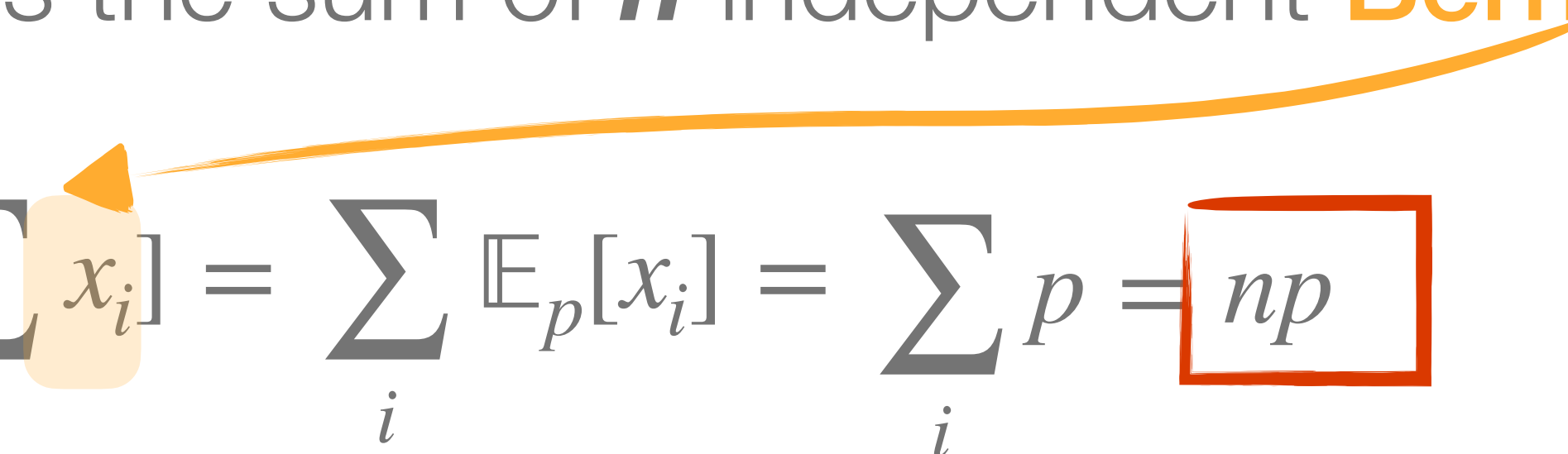
E.g, $n = 4, x = 2$

$$\binom{n}{x} = \binom{4}{2} = \frac{4!}{(4-2)!2!} = \frac{4 \cdot 3}{2} = 6$$

Your turn : mean

For the Binomial distribution, what's the mean of the distribution?

Binomial “event” x is the sum of n independent **Bernoulli** trials ($x_1, x_2, \dots, x_n: x_i \in \{0,1\}$)

$$\mu = \mathbb{E}_p[x] = \mathbb{E}_p\left[\sum_i x_i\right] = \sum_i \mathbb{E}_p[x_i] = \sum_i p = np$$




Your turn : Binomial Variance

For the Binomial distribution, what's the mean and variance of the distribution?

$$\mathbb{E}_p[k^2] = \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n k^2 \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^n k \frac{n!}{(n-k)!(k-1)!} p^k (1-p)^{n-k}$$

Issue: can't cancel this k ... but
could cancel factor $k-1$

Trick: $\mathbb{E}[k(k-1)] = \mathbb{E}[k^2] - \mathbb{E}[k]$



Your turn : Binomial Variance

Trick: $\mathbb{E}[k(k-1)] = \mathbb{E}[k^2] - \mathbb{E}[k]$

$$\mathbb{E}_p[k(k-1)] = \sum_{k=0}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=2}^n \frac{n!}{(n-k)!(k-2)!} p^k (1-p)^{n-k}$$

Terms in the sum are 0 for
 $k = 0$ and $k = 1$

Let $k' = k - 2$

$$= \sum_{k=0}^{n-2} \frac{n!}{(n-k-2)!k!} p^{k+2} (1-p)^{n-k-2} = n(n-1)p^2 \sum_{k=0}^{n-2} \frac{(n-2)!}{(n-2-k)!k!} p^k (1-p)^{(n-2)-k}$$

$$= n(n-1)p^2$$

$$= 1 \quad \text{Kolmogorov's 2nd axiom}$$

Putting the pieces together

$$\begin{aligned} \sigma^2 &= \mathbb{E}_p[k(k-1)] + \mathbb{E}_p[k] - \mathbb{E}_p[k]^2 = \cancel{n(n-1)p^2} + np - \cancel{(np)^2} \\ &= -np^2 + np = \boxed{np(1-p)} \end{aligned}$$

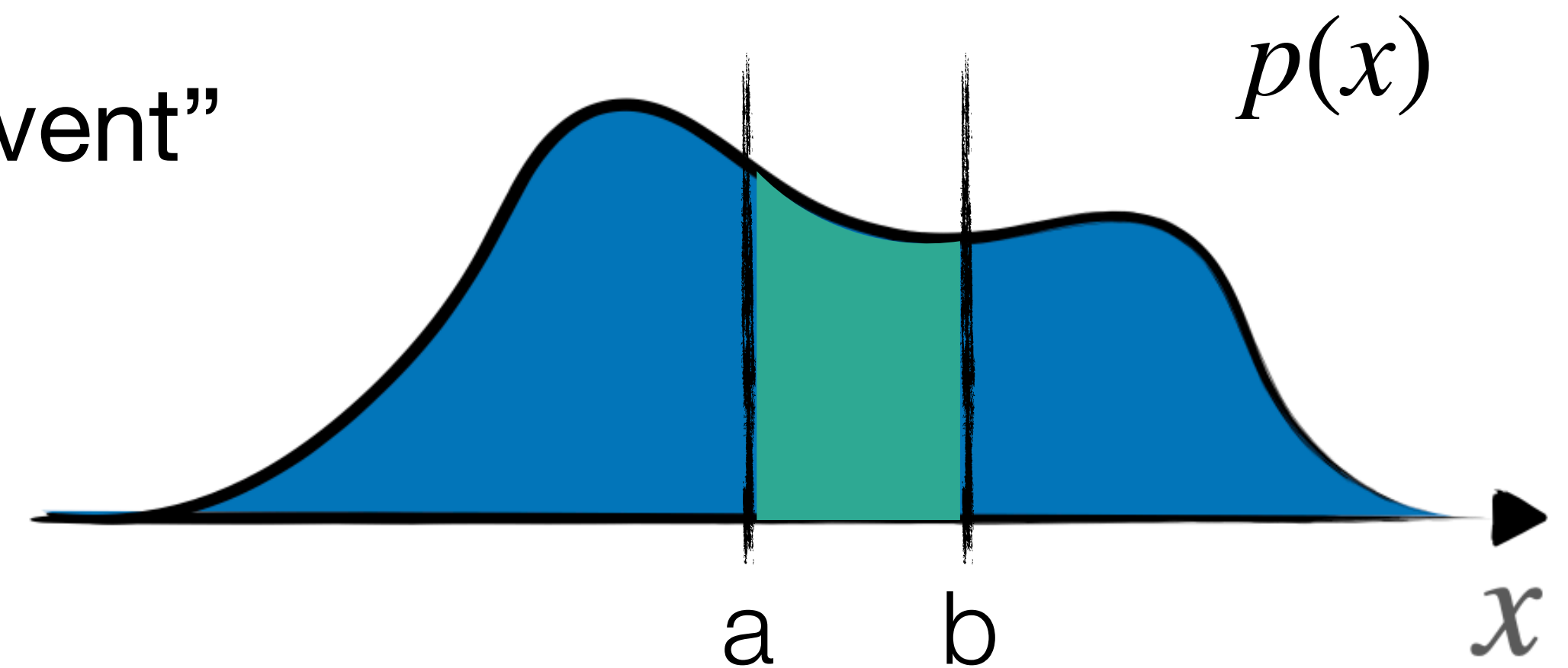
Continuous probability distributions

Probability density function, $f(x)$:

↪ continuous analog of an “elementary event”

$$p(x \in [a, b]) = \int_a^b dx p(x)$$

Discrete probability Probability density



The fundamental theorem of calculus for probability distributions

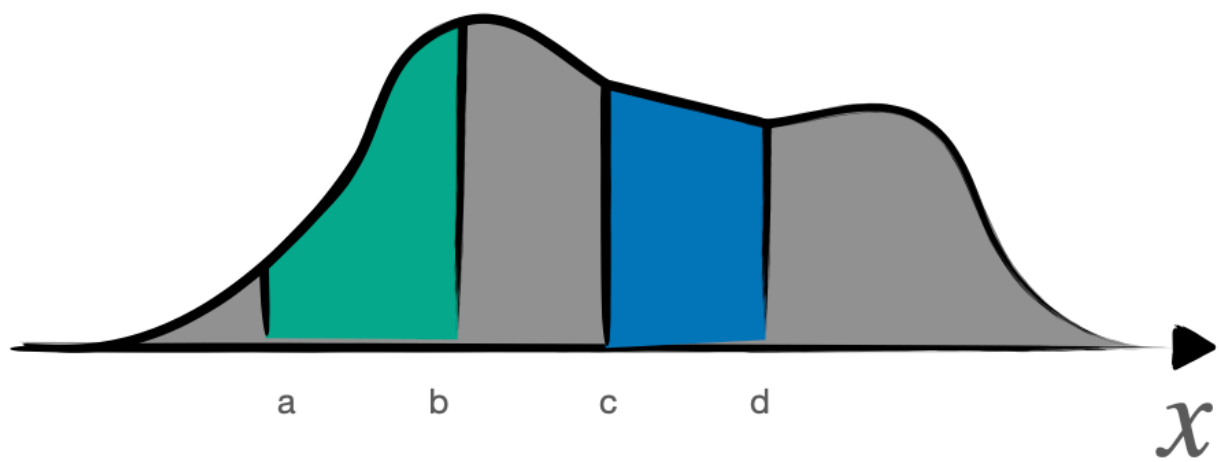
Continuous probability distributions

Kolmogorov's axioms:

1) Positivity: $f(x) > 0$

2) Integratability: $\int_{\Omega} dx p(x) = 1$

3) Sum probs of disjoint events



$$\begin{aligned} p(x \in [a, b] \text{ or } x \in [c, d]) &= \int_a^b dx p(x) + \int_c^d dx p(x) \\ &= p(x \in [a, b]) + p(x \in [c, d]) \end{aligned}$$

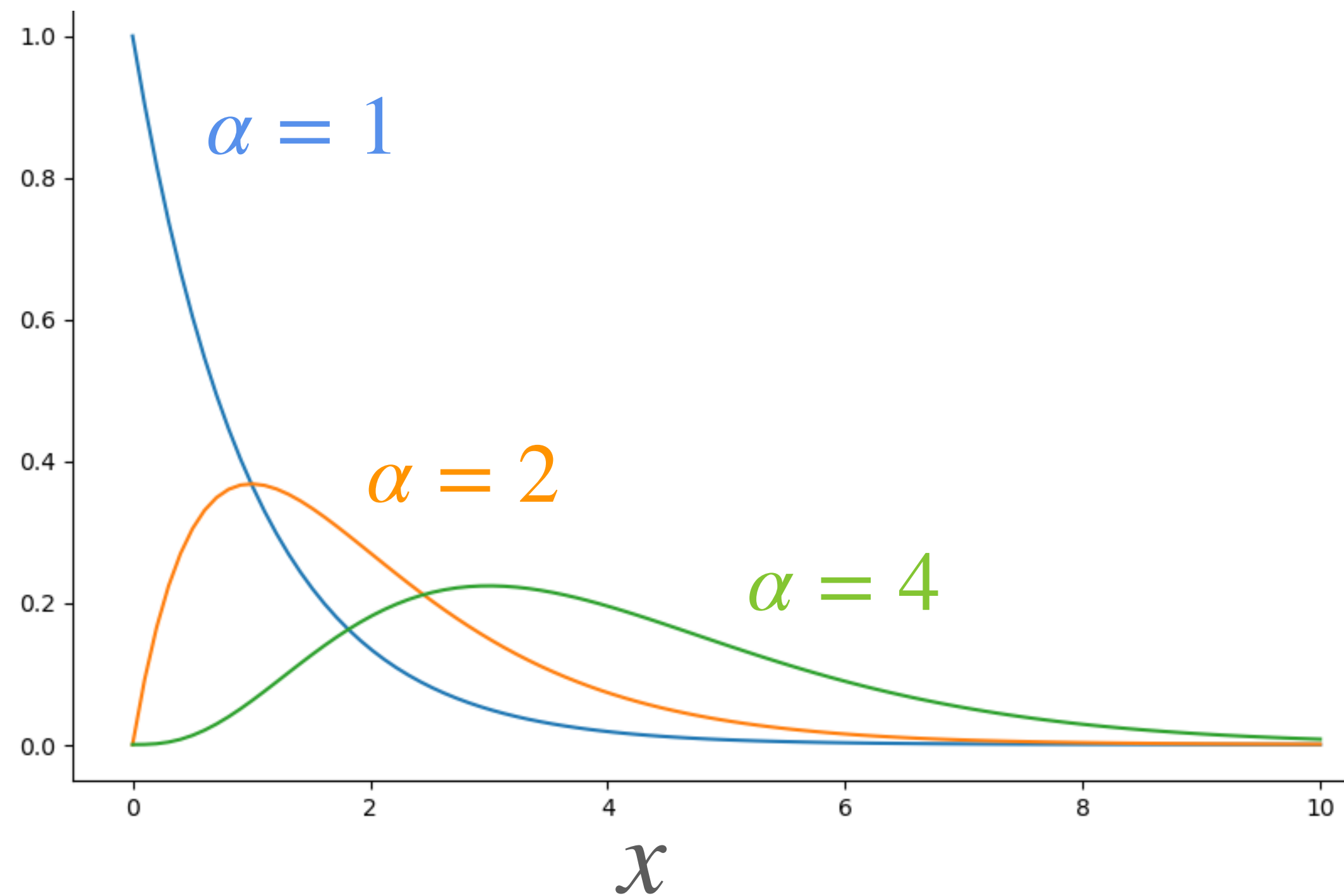
Gamma distribution

$$\Gamma(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$$

$$x \in [0, \infty) \quad \text{and} \quad \alpha > 0, \beta > 0$$

$$\Gamma_{\alpha, \beta}(x)$$

Gamma density
functions for $\beta = 1$



α : shape parameter

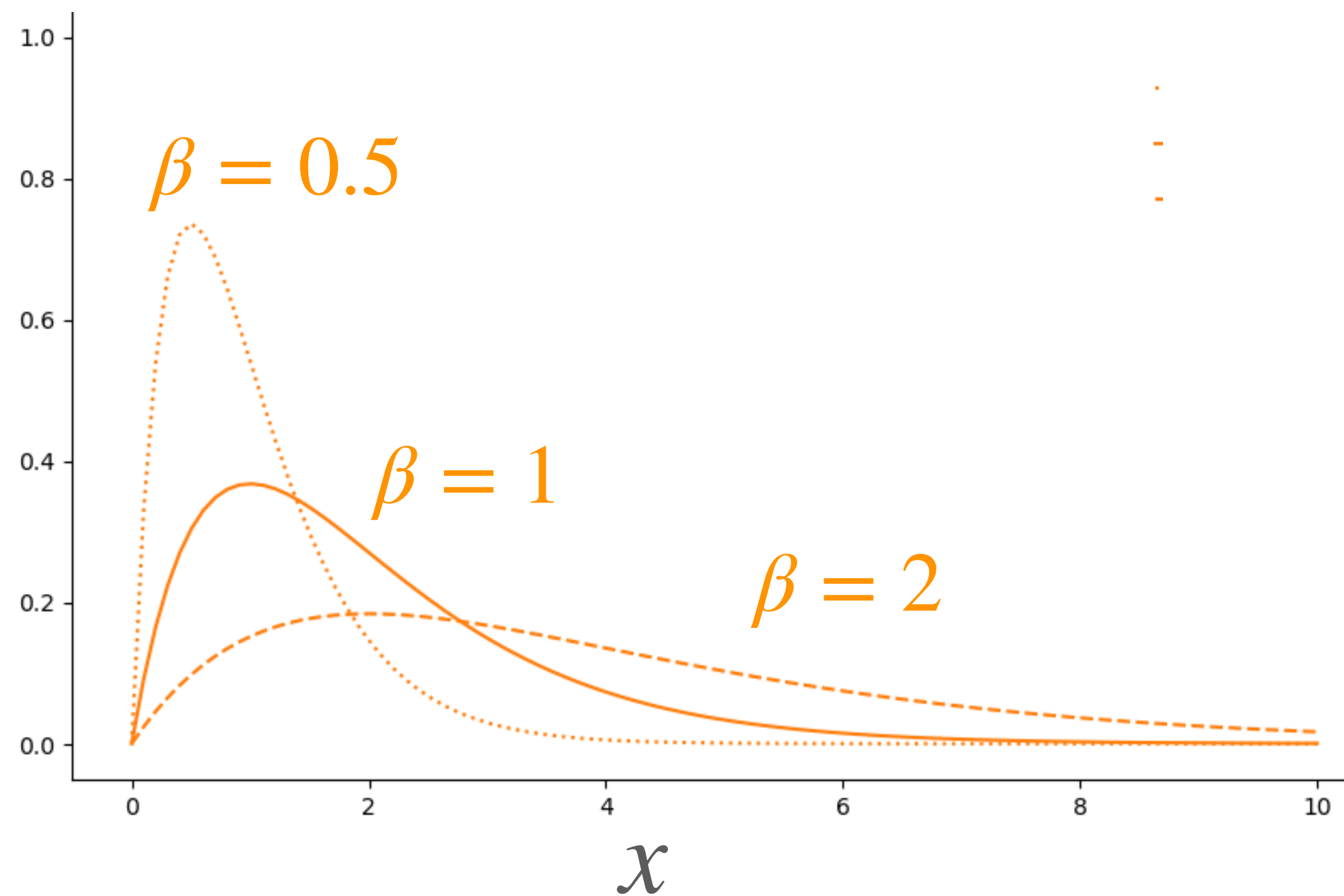
Gamma distribution

$$\Gamma(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$$

$$x \in [0, \infty) \quad \text{and} \quad \alpha > 0, \beta > 0$$

$$\Gamma_{\alpha, \beta}(x)$$

Gamma density functions for $\alpha = 2$



α : shape parameter

β : scale parameter

Gamma distribution

$$\Gamma(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad x \in [0, \infty) \quad \text{and } \alpha > 0, \beta > 0$$

Gamma function:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

Integral can be recursively calculated:

$$\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1)$$

$$\Gamma(1) = 1$$

For integer α : $\Gamma(\alpha) = (\alpha - 1)!$

Gamma distribution

$$\Gamma(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad x \in [0, \infty) \quad \text{and } \alpha > 0, \beta > 0$$

$$\mu = \mathbb{E}_{x \sim \Gamma}[x] = \alpha\beta$$

$$\sigma^2 = \mathbb{E}_{\Gamma}[x^2] - \mathbb{E}_{\Gamma}[x]^2 = \alpha\beta^2$$

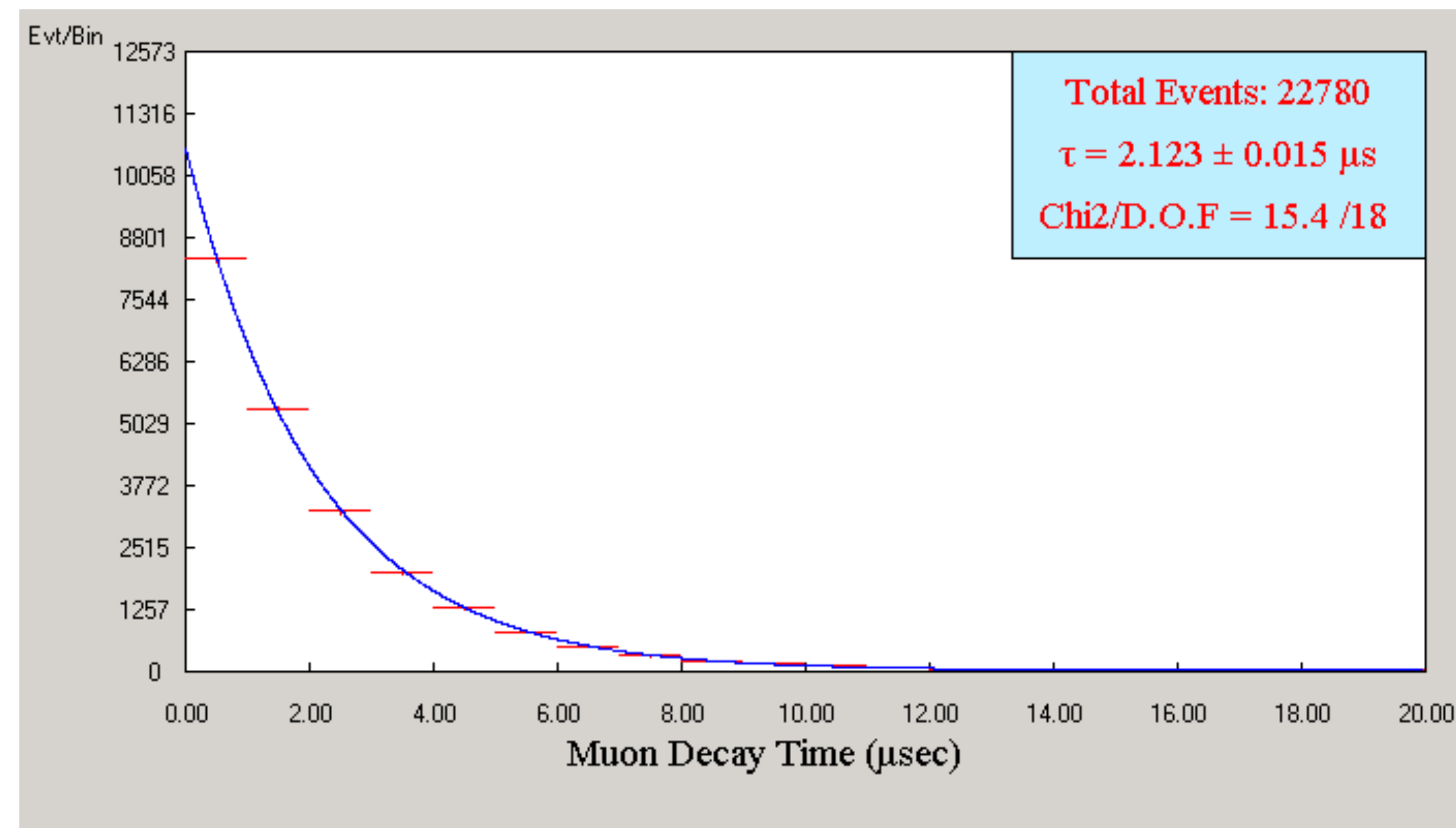


Will prove in the homeworks

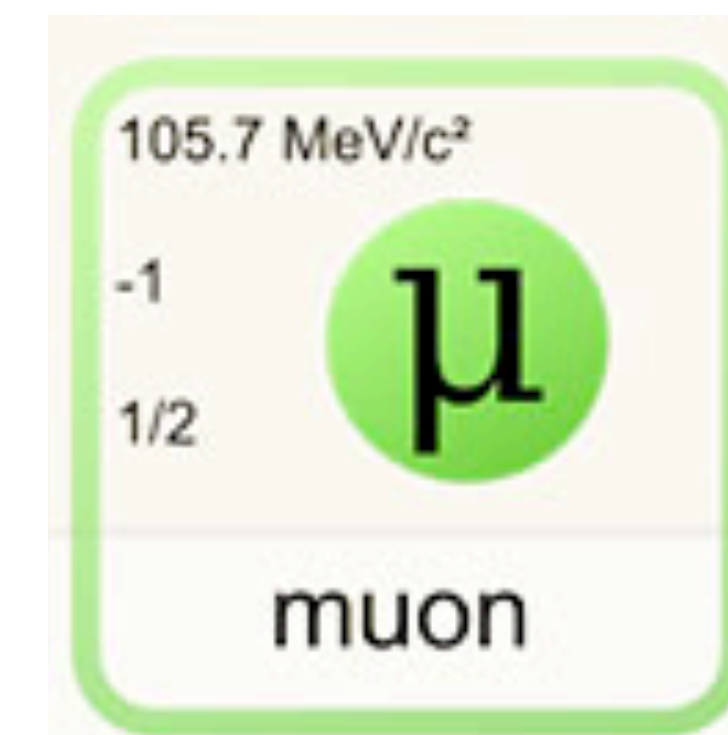
Exponential distribution

Special case of the Gamma dist. when $\alpha = 0$

$$\Gamma(x; \alpha, \lambda) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \xrightarrow{\alpha = 0} \text{Exp}(x; \lambda) = \frac{1}{\beta} e^{-x/\beta}$$



← Example: measure the muon lifetime by fitting this exponent, β .



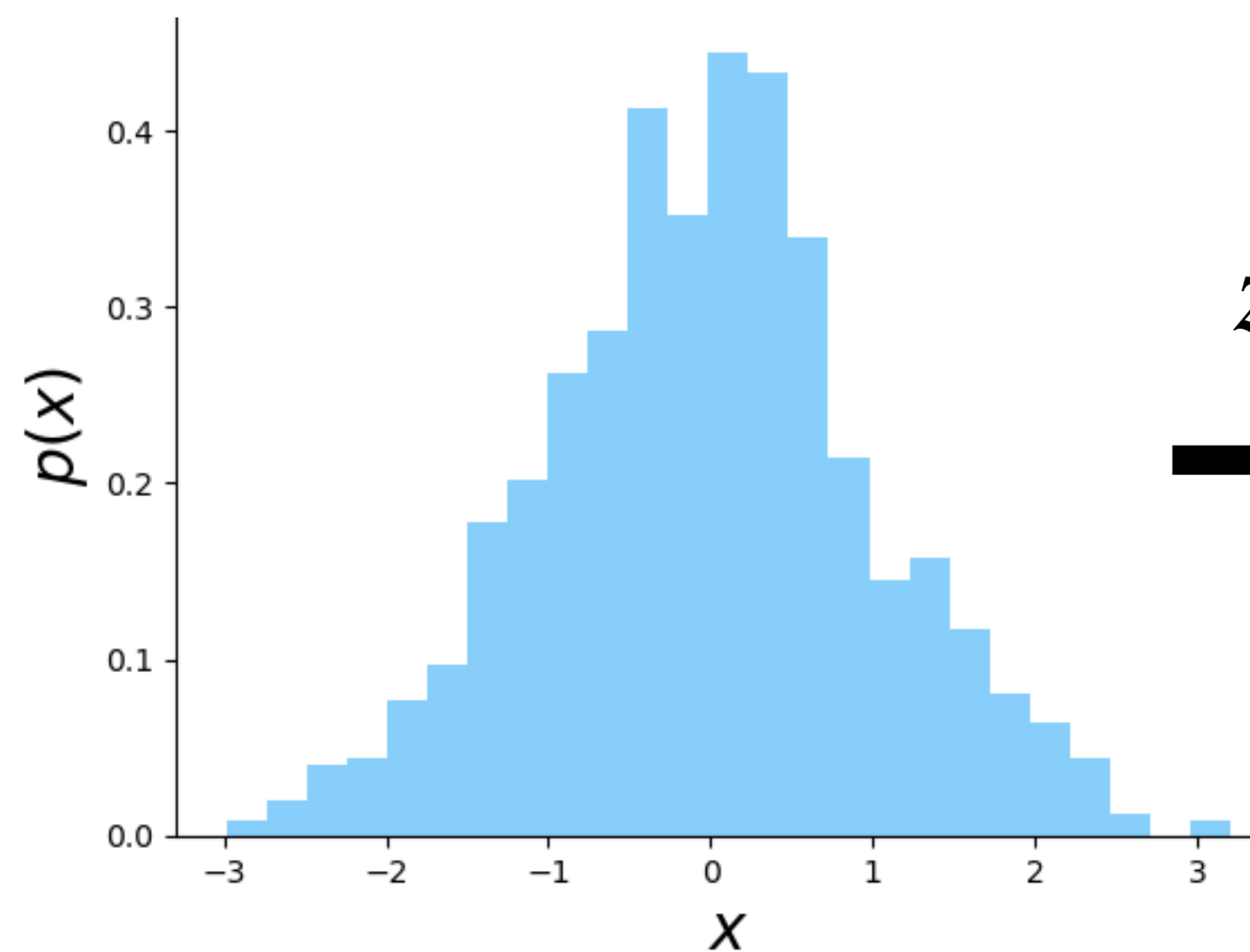
Chi2 distribution

Special case of the Gamma dist. when $\alpha = \nu/2, \beta = 2$:

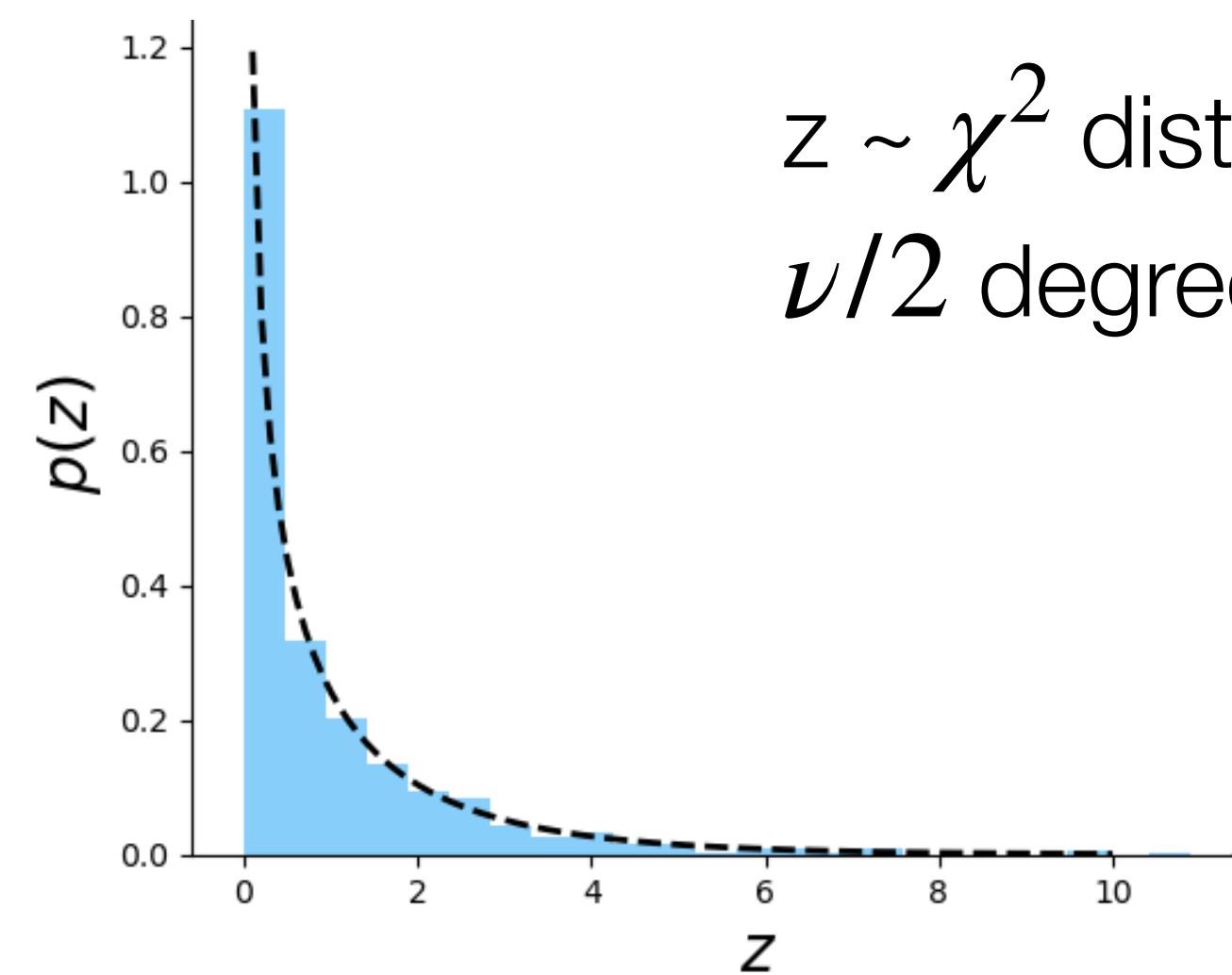
$$\Gamma(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$$

$$\frac{\alpha = \nu/2}{\beta = 2} \rightarrow \chi^2(x; \nu) = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}$$

Ex: MSE on a Gaussian



$$z = x^2$$



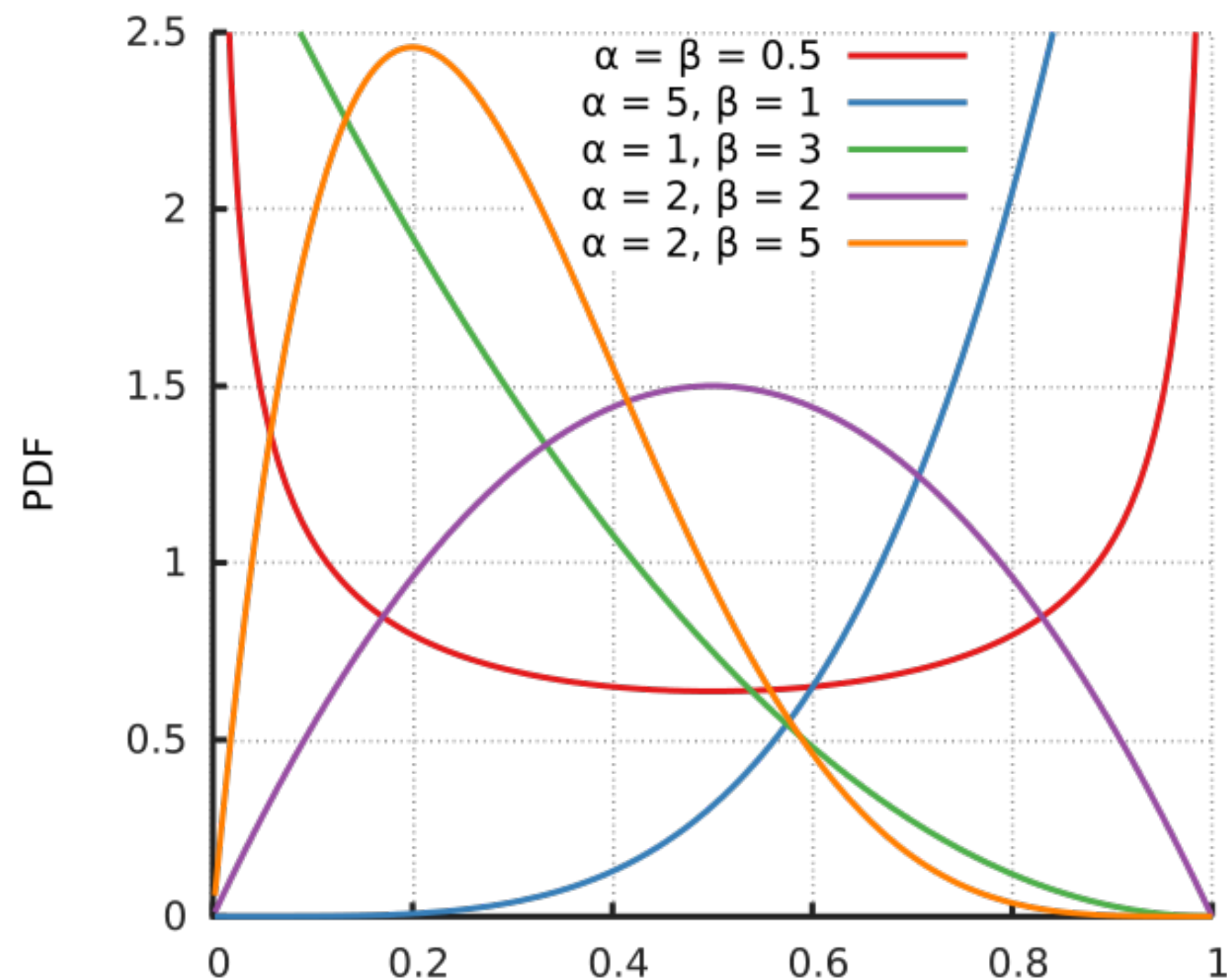
$z \sim \chi^2$ distribution with $\nu/2$ degrees of freedom

More in Jonas's lectures **next week!**

Beta distribution

$$\text{Beta}(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad x \in [0, 1]$$

$$\alpha, \beta > 0$$



https://en.wikipedia.org/wiki/Beta_distribution

Beta distribution

$$\text{Beta}(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad x \in [0, 1]$$

Beta function:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

if $\alpha, \beta \in \mathbb{N}$, $B(\alpha, \beta) = \frac{(\alpha - 1)! (\beta - 1)!}{(\alpha + \beta - 1)!}$

$$= \frac{(\alpha + \beta)}{\alpha \cdot \beta} \cdot \frac{\alpha! \beta!}{(\alpha + \beta)!}$$
$$= \frac{(\alpha + \beta)}{\alpha \cdot \beta} \cdot \frac{1}{\binom{\alpha + \beta}{\alpha}}$$

Combinatorial factor

from binomial dist

👁️ Will explore connection b/w beta and binomial tomorrow.

Beta distribution

$$\text{Beta}(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad x \in [0, 1]$$

$$\mathbb{E}[x] = \frac{\alpha}{\alpha + \beta}$$

$$\text{Var}(x) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Will prove in the
tutorial today 

Feedback for me

How was the material today?

- A) Too fast
- B) Too slow
- C) OK

PollEv.com/nicolehartman968



Summary

Probability distributions:

- Kolmogorov's axioms
- Chain rule of probability:
 - $p(A, B) = p(A | B)p(B)$
- Bayes Rule

$$p(\theta | x) = \frac{\overset{\text{Likelihood}}{p(x | \theta)} \overset{\text{Prior}}{p(\theta)}}{\underset{\text{Evidence}}{p(x)}}$$

↻ Posterior ↻ Evidence



Probability Distributions

	Name	Formula	Domain	Mean	Variance	Notes
Discrete: k	Bernoulli (p)	$p^k(1-p)^{1-k}$	$\{0,1\}$	p	$p(1-p)$	Used for ML: binary classification
	Binomial ($p;n$)	$\binom{n}{k} p^k(1-p)^{n-k}$	$\{0,1,\dots,n\}$	np	$np(1-p)$	Bernoulli with n trials
	Categorical					
	Multinomial					
	Poisson (λ)					
Continuous: x	Gamma (α,β)	$\frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$	$\mathbb{R}_{\geq 0}$	$\alpha\beta$	$\alpha\beta^2$	$\Gamma(\alpha) = \int_0^\infty dx x^{\alpha-1} e^{-x}$: Gamma function $\alpha, \beta > 0$ $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, for integer $\alpha = n$: $\Gamma(n + 1) = n!$
	Exponential (β)	$\frac{1}{\beta} e^{-x/\beta}$	$\mathbb{R}_{\geq 0}$	β	β^2	Special case of Gamma dist for $\alpha = 1$
	Chi2 (ν)	$\frac{1}{2^{\nu/2} \Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}$	$\mathbb{R}_{\geq 0}$	ν	2ν	Special case of Gamma dist for $\alpha = \nu/2, \beta = 2$
	Beta (α,β)	$\frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$[0,1]$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$B(\alpha,\beta) = \int_0^1 dx x^{\alpha-1} (1-x)^{\beta-1} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$: Beta function
	Dirichlet					
	Gaussian (μ,Σ)					
	Delta (μ)					