

The Data Science of Particle Physics

Lecture 1: Intro to Particle Physics & the Statistics of Discoveries

Dr. James Keaveney¹

¹james.keaveney@uct.ac.za
Room 5.05, RW James Building

June 2024



The Data Science of Particle Physics

- Week 1 - 5 sessions with me (JK)
 - All course content for week 1 is available in this git repo
[GitHub link](#)
 - Introduction to particle physics - no physics experience necessary!
 - The statistics of discoveries
 - 2-3 Open-ended data analysis assignments using Real Data from the ATLAS experiment at CERN!
- Weeks 2 and 3 with Dr. Julia Gonski (Stanford)

Who am I?



James Keaveney
RW James 5.05
james.keaveney@uct.ac.za

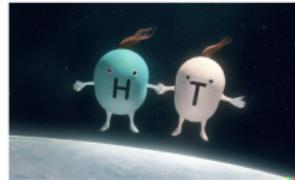


dEFT



Particle physics with ATLAS

- seeking new physics with **Higgs bosons** and **Top quarks**
- **Real-Time AI – Anomaly detection in the Trigger**

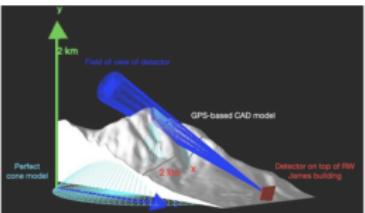


Low-cost Medical PET imaging

- Quantum Tech meets AI!

Detector development

- **ATLAS inner tracker**
- μ CT - muon tomography

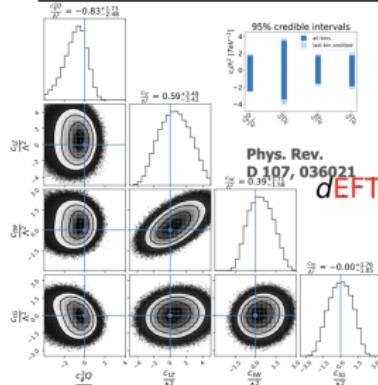


Particle phenomenology

- constraining new physics via Effective Field Theory
- **Outreach**
 - ATLAS Open Data for education



OPPENHEIMER
MEMORIAL TRUST



Read more about my research in the media... If you like :)

- OMT Article
- Engineering News Article
- UCT News Article

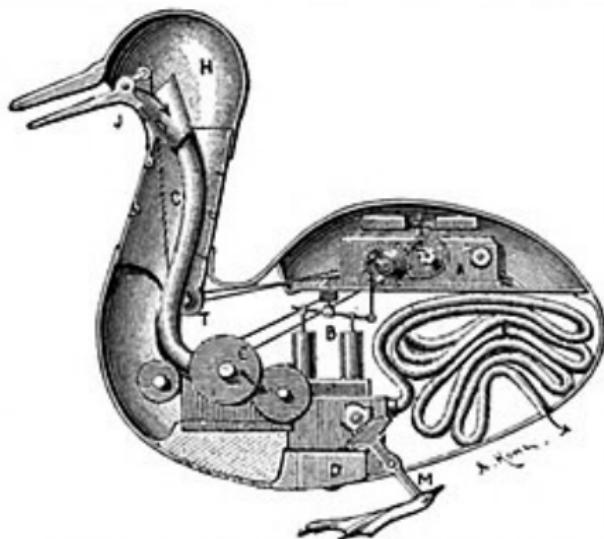
Ice-breakers (No wrong answers)

- What is particle physics?
- What is data science?

Ice-breakers 2 (No wrong answers)

- What do you hope to get out of this module?

Reductionism



INTERIOR OF VAUCANSON'S AUTOMATIC DUCK.

A, clockwork; B, pump; C, mill for grinding grain; F, intestinal tube;
J, bill; H, head; M, feet.

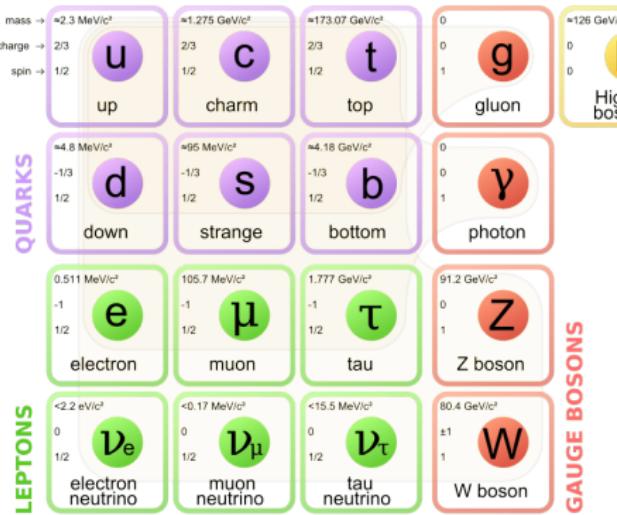
Particle Physics as *Reductionism*

- **Methodological reductionism:** the scientific attempt to provide explanations of nature in terms of ever smaller entities.
- **Particle physics** represents (**for now**) the culmination of the reductionist approach to understanding the universe

Physics at the end of the (experimental) road

- Following the reductionist approach means experimentally probing smaller and smaller distance scales ℓ
- But from Quantum Mechanics we know $\ell = \hbar c / E$
- Smaller distance scales means larger energy scales! $\ell \downarrow \equiv E \uparrow$
- **Colliding particles at the highest-ever energies allows us to probe nature at the smallest-ever distance scales**

So what do we know?



- u , d and e form all *normal* matter
- 3 fundamental forces: Electromagnetic, Weak, and Strong
 - each mediated by the exchange of elementary bosons

What do we know: fundamental fermions

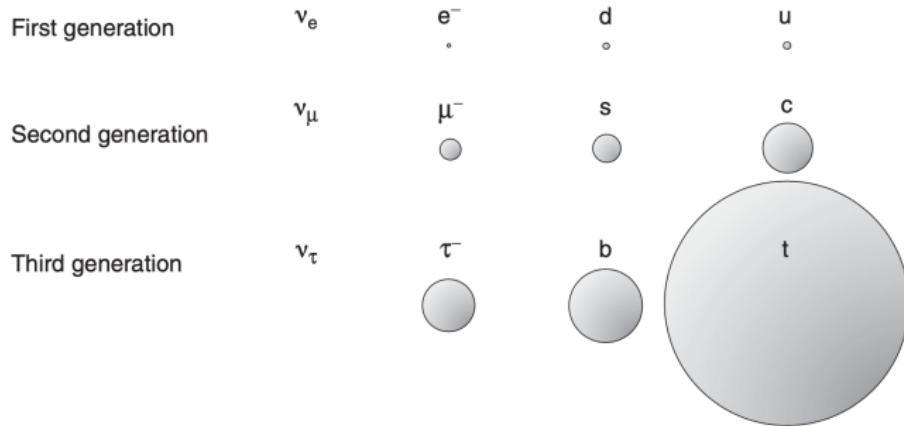


Figure: fundamental particles drawn as spheres with $V \propto m$

- Why are there three *generations*?
- Why the random masses if these are fundamental?
- $m_{top} \approx m_{Au\ atom}!! \Gamma \Gamma$

What do we know: forces through boson exchange

- fundamental particles *interact*: scatter, decay, annihilate...

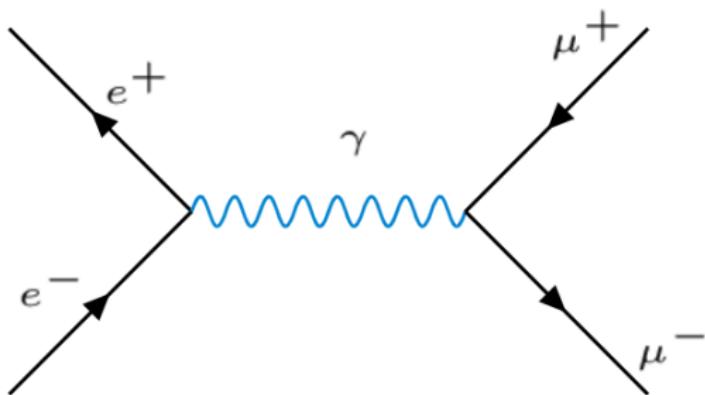


Figure: Feynman diagram for Bhabha scattering ($e^+e^- \rightarrow \mu^+\mu^-$)

- basic interactions (EM, weak, strong) understood as due to boson exchange (γ , W^\pm or Z , g)
- Nobel-prize winning discovery of the Higgs boson in 2012 completes the Standard Model, but...

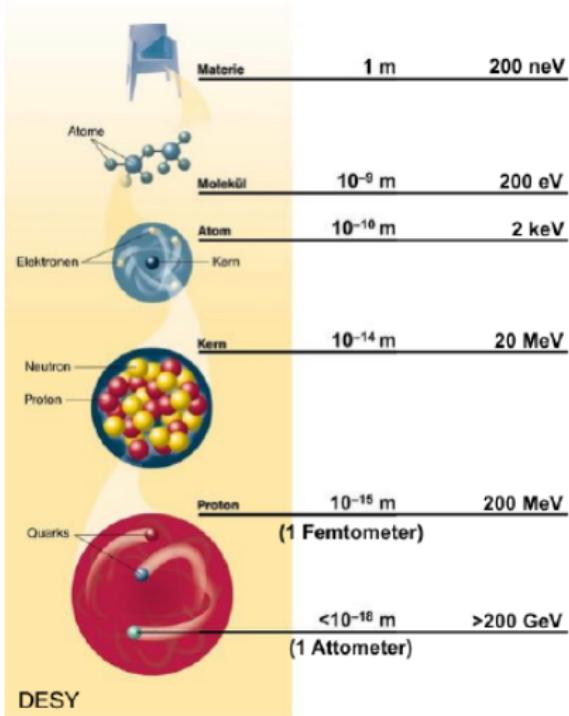
What don't we know?

- we have no idea how to include gravity in the Standard Model
- we cannot reconcile our understanding of gravity with the structure and evolution of the universe without assuming $\approx 95\%$ of the universe is of unknown nature (Dark Matter/Energy), not accounted for in Standard Model
- No explanation for the masses of the Fundamental particles, especially the Higgs boson
- No sign yet of a discovery that would answer these questions...

The Large Hadron Collider

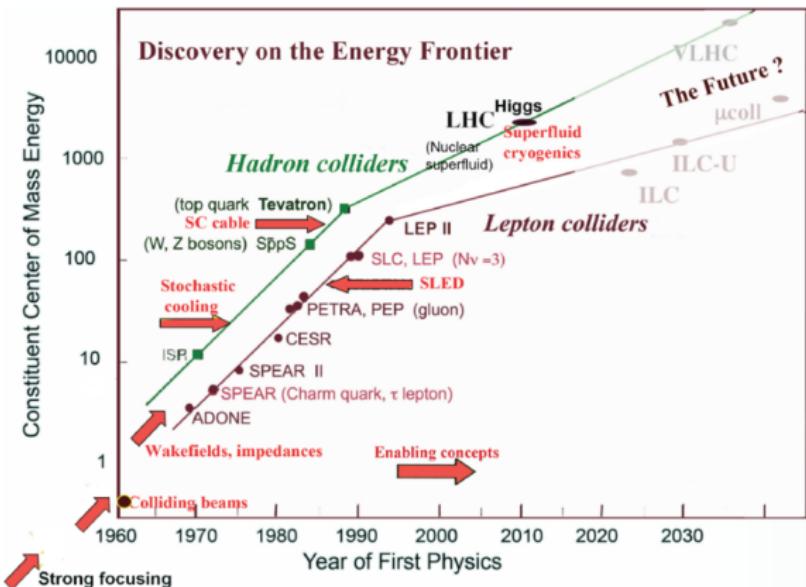


Why particles accelerators?



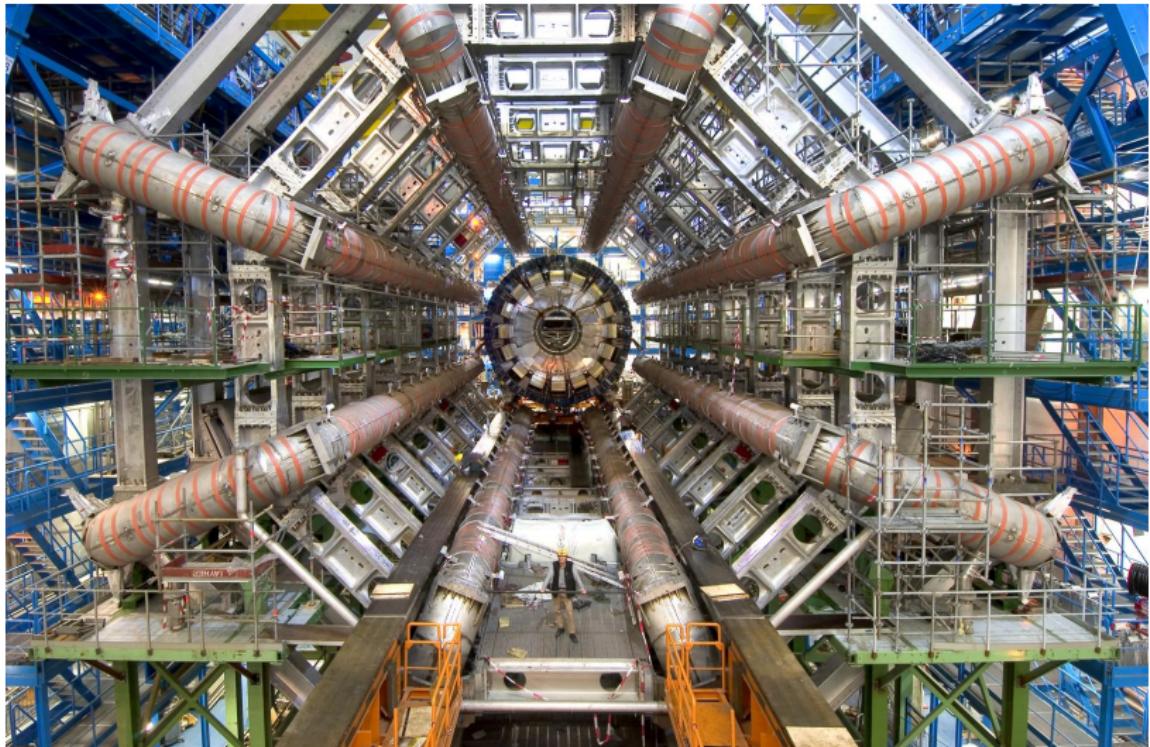
- Production of **new heavy particles** ($E = mc^2$)
 - Maximum mass of produced particle = centre-of-mass energy of elementary collision, e. g. parton-parton collision
- Resolution of **small structures**: accelerator as very powerful **microscope**
 - **De-Broglie wave length** of particle beam
$$\lambda = \frac{h}{p} = \frac{2\pi\hbar c}{pc} \rightarrow \lambda [\text{fm}] \approx \frac{1.24}{p[\text{GeV}]}$$
 - For example
 - $p = 1 \text{ GeV} \rightarrow \lambda = 1.24 \cdot 10^{-15} \text{ m}$
 - $p = 1 \text{ TeV} \rightarrow \lambda = 1.24 \cdot 10^{-18} \text{ m}$

Colliders as microscopes



- LHC has been colliding protons at 13 TeV since 2015

The ATLAS Detector

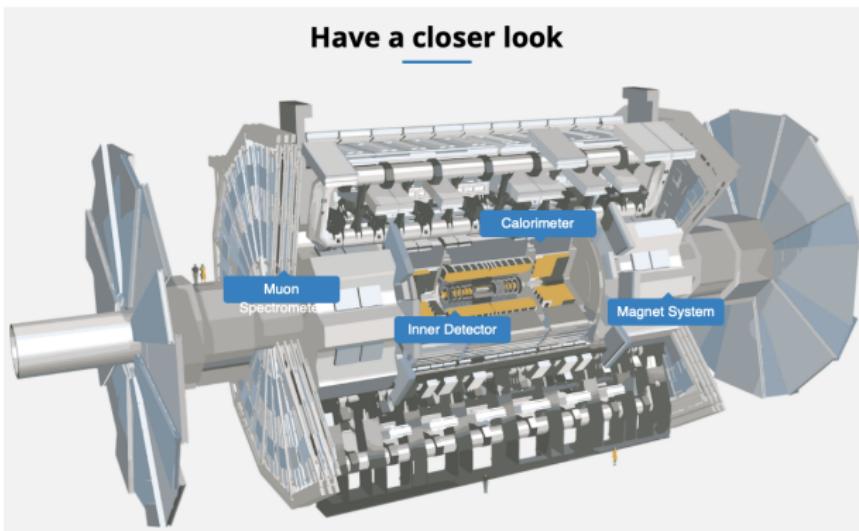
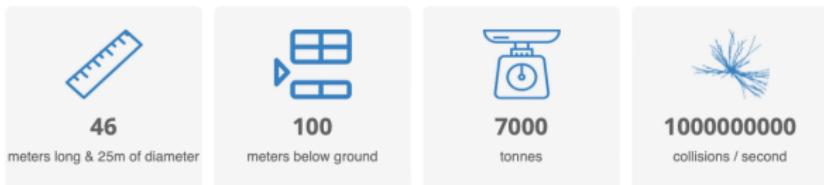


The ATLAS Detector

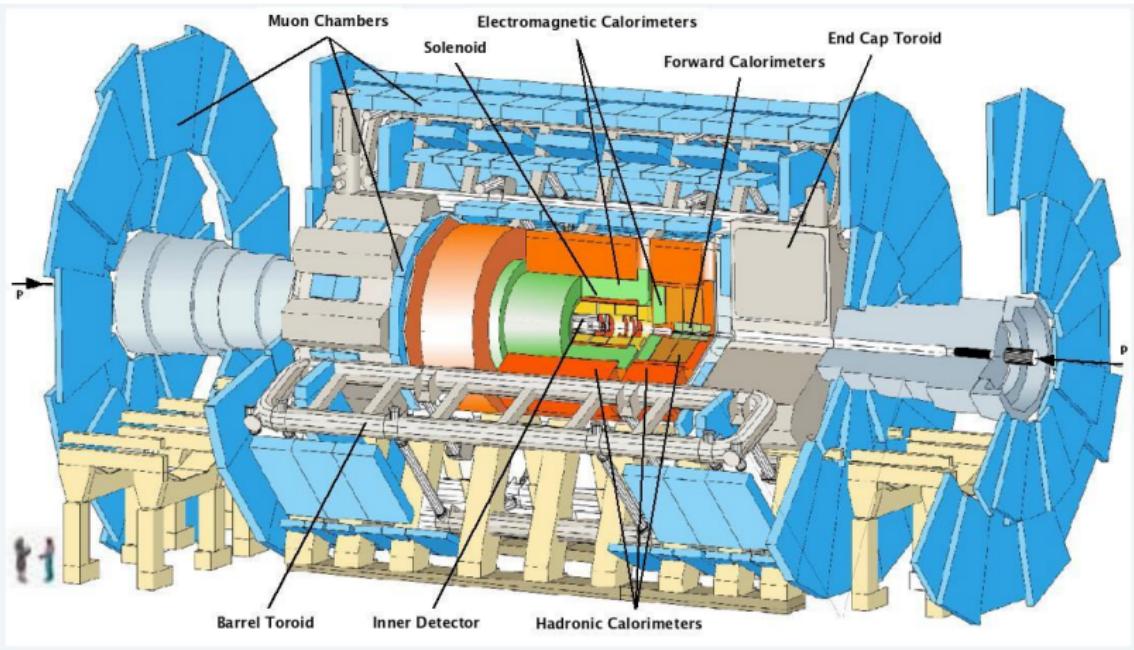


The ATLAS Detector

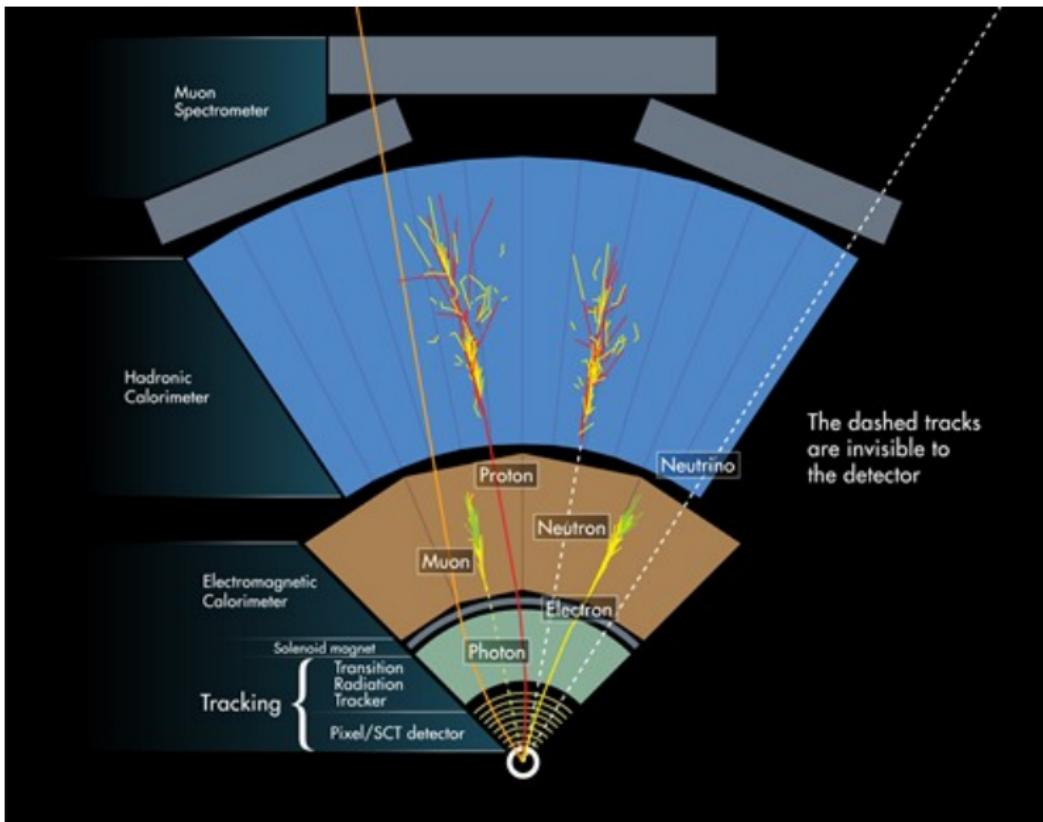
- Take a virtual tour of the ATLAS detector [here!](#)



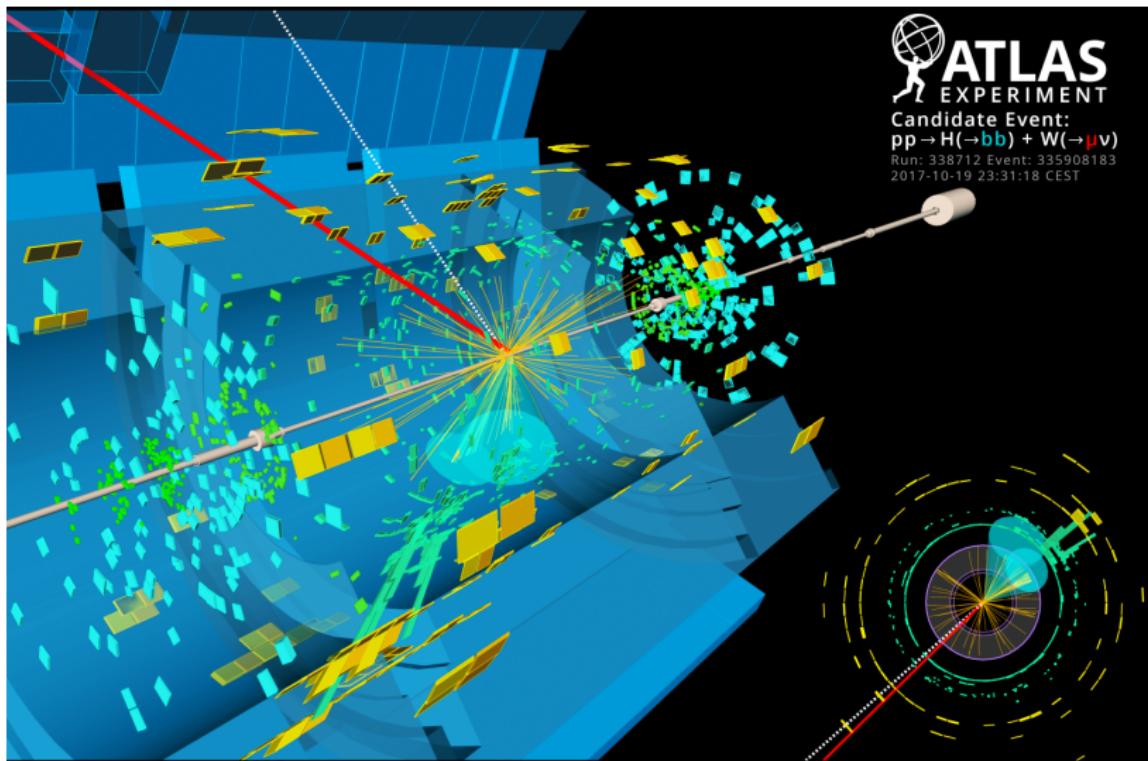
The ATLAS Detector



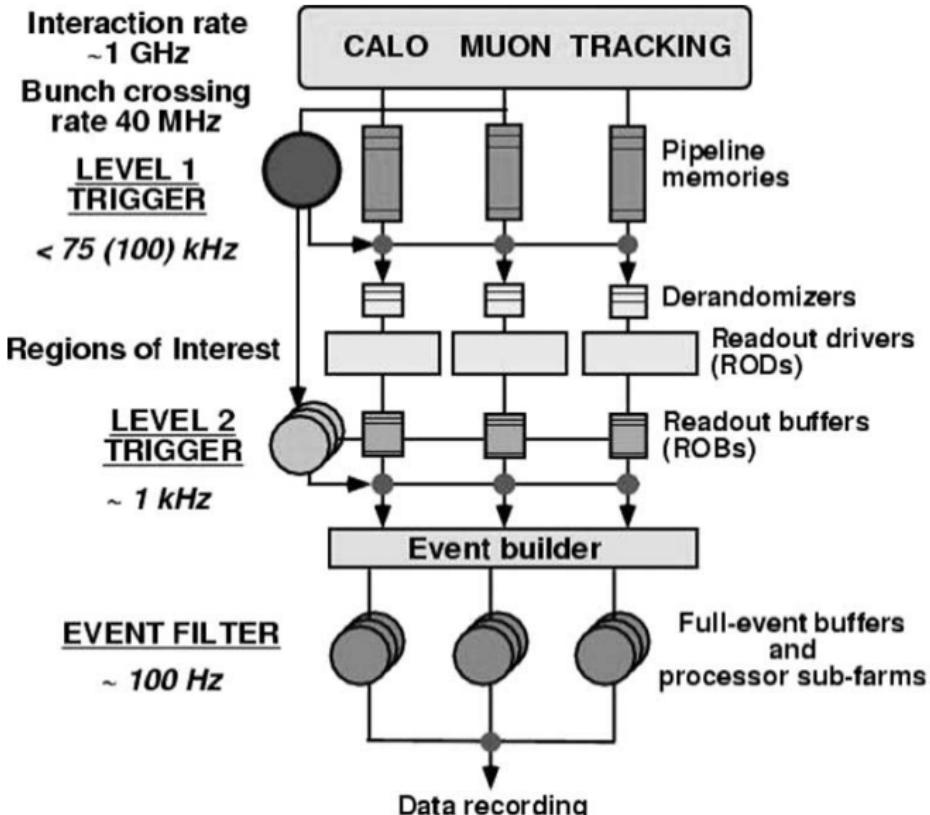
The ATLAS Detector



What happens when you collide protons?



The ATLAS Trigger



So what does ATLAS data look like?

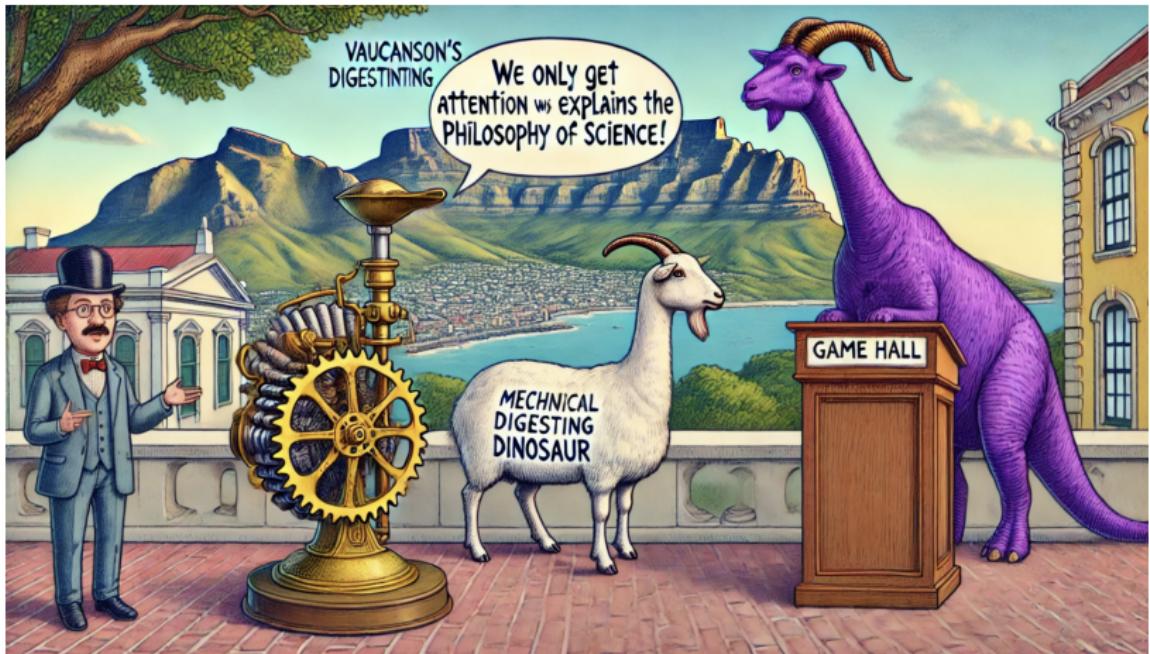
- surprisingly simple!

	df						
	✓	0.1s					Python
0	[59232.8828125 55261.05078125]	[49085.1963125 39126.15625]	[0.1477134525759094 -2.86790585178833]	[0.632411367406616 0.879536330699207]	[1 -1]	[-991.515356445312 603.2068461445312]	[0.0 0.0]
1	[12704.1875 33773.015625]	[65642.484375 33766.197953125]	[-1.462382197380606 -2.781487941749434]	[1.27910375590277 -0.018980734050273895]	[-1 1]	[-788.6085205078125 -1062.7930908203125]	[1063.2587890625 0.0]
2	[201868.75 54080.953125]	[38357.2578125 32836.46484375]	[-2.7715253829956055 0.05025695636868447]	[-2.344670295715332 -1.0837055444717407]	[-1 1]	[-320.12200927734375 670.6602172851562]	[0.0 0.0]
3	[79745.3828125 28305.0178125]	[69548.1328125 27166.60548875]	[1.5498226881027222 -0.946570873260499]	[-0.535100564020991 0.2884726822376251]	[-1 1]	[-39.91961669921875 -149.05303950578125]	[0.0 0.0]
4	[28867.724609375 98152.65625]	[27808.724609375 25207.083964375]	[-1.008341670036316 1.500999927520752]	[0.2750834822654724 -2.0356335639953613]	[-1 1]	[121.45063100585938 119.237670984375]	[0.0 0.0]
...
549252	[37704.515625 53827.45703125]	[33141.77734375 31681.8828125]	[2.15971360931965 -1.0514856576919558]	[-0.5188854932785034 1.1224994569423828]	[-1 1]	[-380.786376953125 2450.8642578125]	[0.0 1780.0516357421875]
549253	[38264.09765625 99156.9921875]	[38691.76953125 30524.439463125]	[2.4906346797943116 -0.569039462164906]	[0.2917047142982483 1.846530795097351]	[-1 1]	[188.4114379882125 -37.4415435791056]	[0.0 0.0]
549254	[39036.66640625 114896.96875]	[36656.05859375 28371.234375]	[0.891377470588684 -2.3450729846954346]	[0.3564602773185944 2.0762693630996708]	[-1 1]	[-491.8928627832031 -164.48634338378906]	[0.0 0.0]
549255	[37635.5546875 63254.0546875]	[36309.703125 34960.17578125]	[-0.868657648563385 2.35759735260098]	[0.2694103559494 -1.19911342430174]	[1 -1]	[-358.5328747558594 -358.3528747558594]	[0.0 0.0]
549256	[61195.4609375 35183.0625]	[47037.4921875 33874.4463125]	[-2.48432903888855 1.060566773223877]	[-0.7576252222061157 0.27705806493759156]	[1 -1]	[2067.1955556640625 889.0148315429688]	[0.0 0.0]

549257 rows x 7 columns

- However we are typically looking for **incredibly** rare phenomena and we collide at 40 MHz for years on end
- Hence particle physics is an inherently **statistical science**

Thought experiments



Lorentz Four-Vectors

- Four-vectors are mathematical objects that combine space and time components
- In special relativity, a four-vector is represented as (\vec{x}, t) or (ct, \vec{x})
- Four-vectors are used to describe physical quantities in a way that is consistent with the principles of special relativity
- Examples of four-vectors:
 - Four-position: $x^\mu = (ct, \vec{x})$
 - Four-momentum: $p^\mu = (E/c, \vec{p})$
 - Four-velocity: $u^\mu = \gamma(c, \vec{v})$

Lorentz Transformations and Invariance

- Lorentz transformations are a set of linear transformations that relate the coordinates of one inertial frame to another
- Four-vectors transform under Lorentz transformations in a way that preserves their inner product, known as Lorentz invariance
- Lorentz invariance ensures that the laws of physics are the same in all inertial reference frames
- The inner product of two four-vectors A^μ and B^μ is defined as:

$$A^\mu B_\mu = A^0 B_0 - \vec{A} \cdot \vec{B}$$

- Explicitly, for two four-vectors $A^\mu = (A^0, \vec{A})$ and $B^\mu = (B^0, \vec{B})$:

$$A^\mu B_\mu = A^0 B^0 - A^1 B^1 - A^2 B^2 - A^3 B^3$$

- Lorentz-invariant quantities, such as the rest mass and proper time, play a crucial role in particle physics

Four-vectors in Particle Physics

- Lorentz invariance is a fundamental principle in the Standard Model of particle physics
- Four-vectors and Lorentz-invariant quantities are used to construct kinematic variables and selection criteria in particle physics analyses
- Lorentz invariance ensures that the results of particle physics experiments are independent of the reference frame, allowing for consistent comparisons and interpretations

The Statistics of Discoveries

- Statistics is
 - peculiar, counter-intuitive, often seems easier than it is
 - elusive: (you think you understand it, you realise you don't)^N
 - **fundamental to modern experimental particle physics**
- Incorrect statistical analysis can mean the difference between a discovery and not a discovery

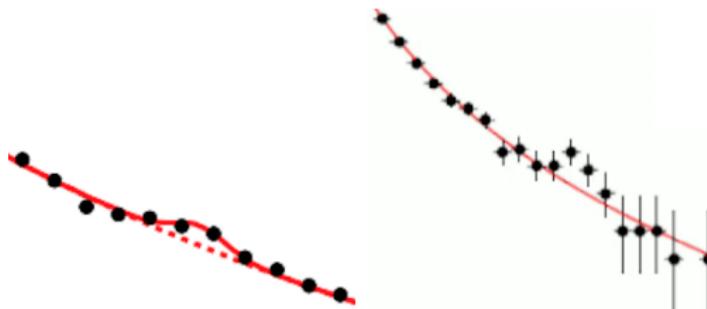


Figure: One of these *bumps* is a real discovery, the other is not...

The Statistics of Discoveries

- Statistics is
 - peculiar, counter-intuitive, often seems easier than it is
 - elusive: (you think you understand it, you realise you don't)^N
 - **fundamental to modern experimental particle physics**
- Incorrect statistical analysis can mean the difference between a discovery and not a discovery

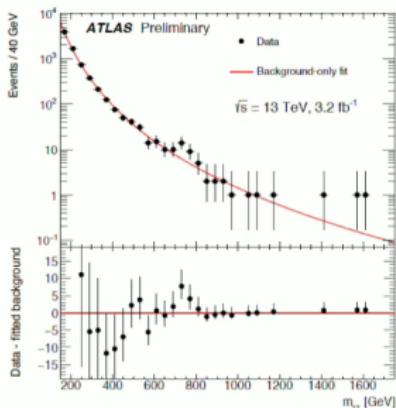
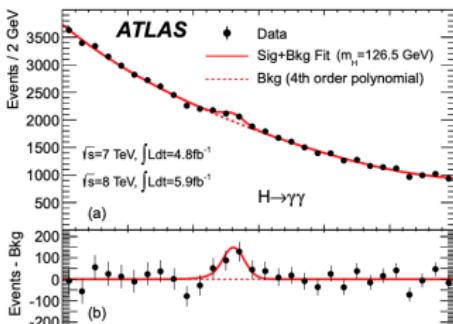
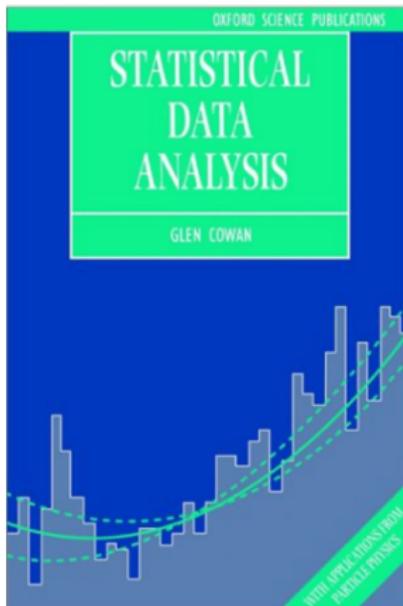


Figure: Discovery (left), Not a Discovery (right)

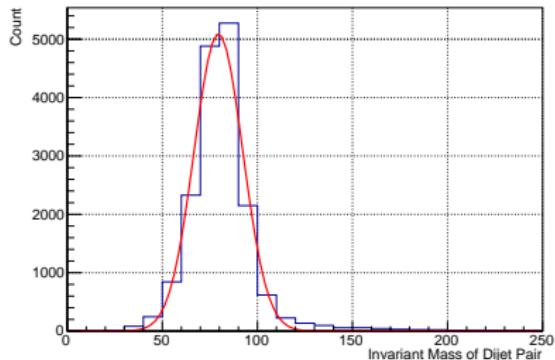
The Statistics of Discoveries

- Statistics is
 - elusive: (you think you understand it, you realise you don't) N
- often need to refer back to textbooks...



Basic concepts: random variables and probability

- Results of repeated "identical", experiments may vary.
 - Instability in apparatus/environment/experimenter
 - Fundamental QM unpredictability of the system
- A variable is **random** when it cannot be predicted with absolute certainty



Basic concepts: probability

- Statistics and Probability: two schools of thought
 - **Bayesian**: Given some data/evidence, we assign probability to some *hypothesis*, e.g. given this LHC data, how sure are we the Higgs boson exists?
 - **Frequentist**: Given some *hypothesis*, how likely is the data we observe, e.g. assuming the Higgs boson exists, how likely is the data that we observe?
- Frequentist approaches are more popular in particle physics
- I will mainly discuss frequentist ideas

Basic concepts: random variables and probability

- Frequentist Probability
 - interpreted as a **limiting frequency**
- Imagine a *repeatable* experiment repeated n times, with S the set of all possible results
- A is a subset of possible results

$$P(A) = \lim_{n \rightarrow +\infty} \frac{N_{\text{result in } A}}{n}$$

- This definition satisfies the **3 axioms of probability**:
 1. $P(A) > 0$ for all A - probabilities can't be negative
 2. $\int_S P(A) = 1$ - *something* must happen
 3. For two mutually exclusive sets A and B , ($A \cap B = \emptyset$),
 $P(A \cup B) = P(A) + P(B)$.

Ice-breaker 1

- What does the *mean* mean?

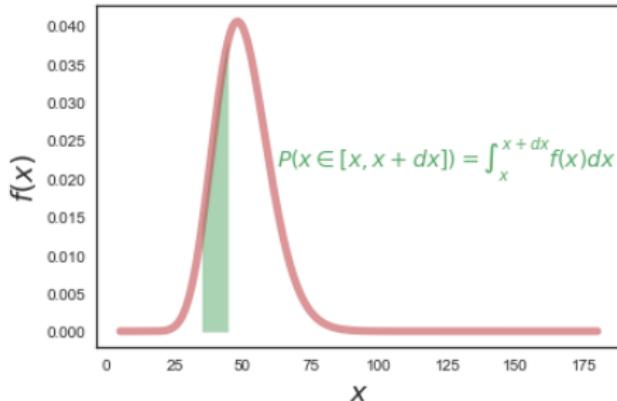
Python code / notebooks

- Code used to make the following plots (unless stated otherwise) available at [GitHub link](#)

Basic concepts: probability density functions (pdf)

- Imagine an experiment with all possible results characterised by a single continuous variable x
- S corresponds to the (1D) space of all possible results
- What is the probability of observing a result in the interval $[x, x + dx]$?
 - given by $f(x)$ (pdf)

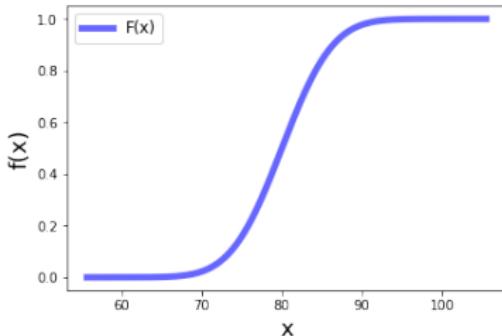
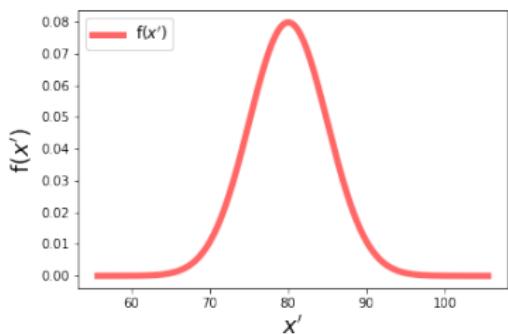
$$P(x \in [x, x + dx]) = \int_x^{x+dx} f(x) dx$$



Basic concepts: cumulative density functions (cdf)

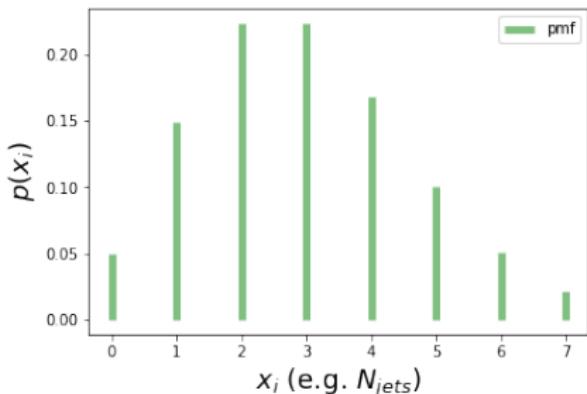
- cdf: $F(x)$
 - probability for x' to have a value $\leq x$

$$F(x) = \int_{-\infty}^x f(x')dx'$$



Basic concepts: probability mass function (pmf)

- If x can only assume discrete values (x_i), we use a *pmf* to describe its distribution
- pmf: $p(x_i) = P(x = x_i)$ where P is a probability.



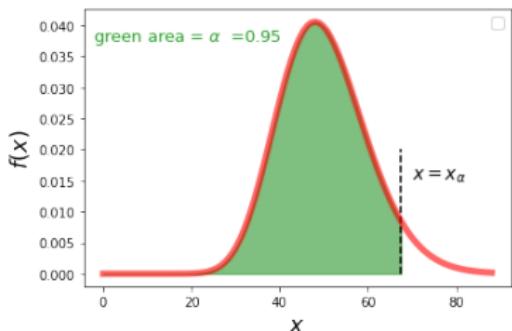
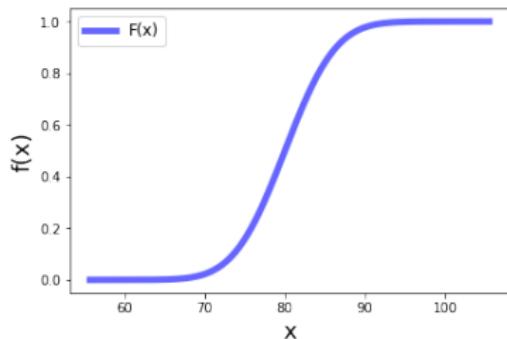
$$\sum_{x_i} p(x_i) = 1$$

- Many examples of discrete observables in particle physics!

Basic concepts: quantiles

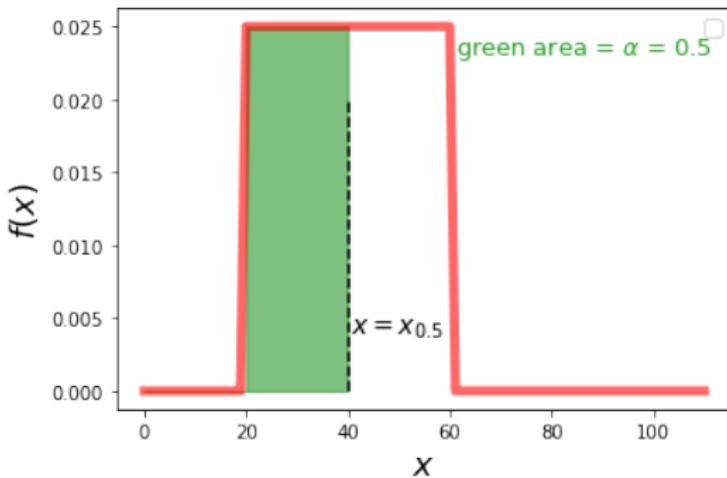
- also known as α points
- the quantile x_α is the value of x such that $F(x_\alpha) = \alpha$
- simply the inverse of the cdf

$$x_\alpha = F^{-1}(\alpha)$$



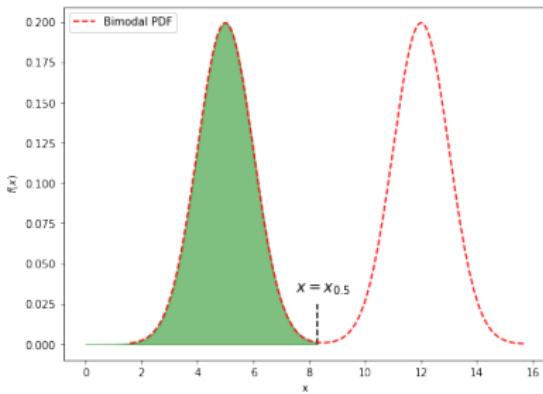
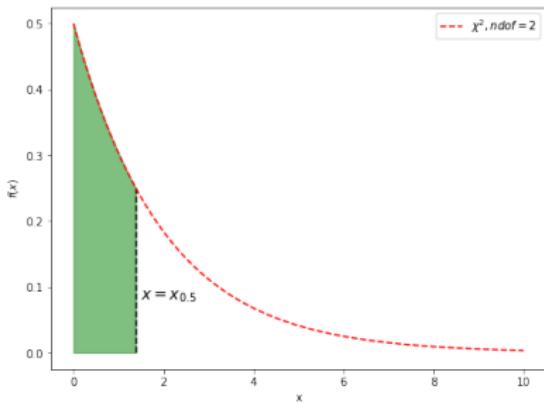
Basic concepts: median

- $x_{0.5}$ is a special case known as the **median**
- median often interpreted as the *typical location of x*
- when can this interpretation break down?



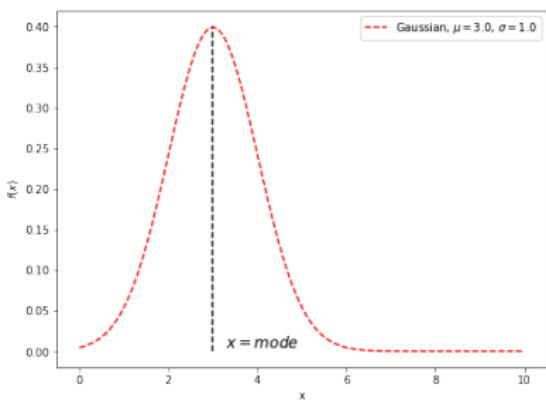
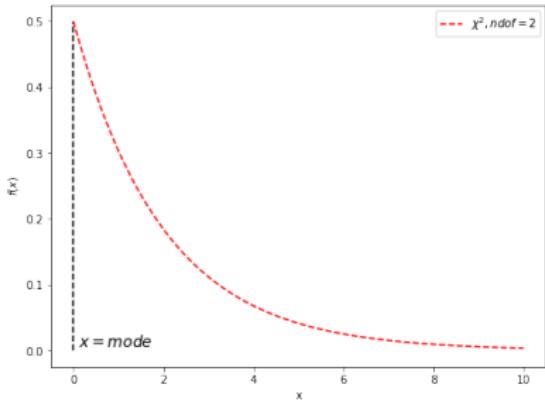
Basic concepts: median

- The median often interpreted as the *typical location of x*
- when can this interpretation break down?



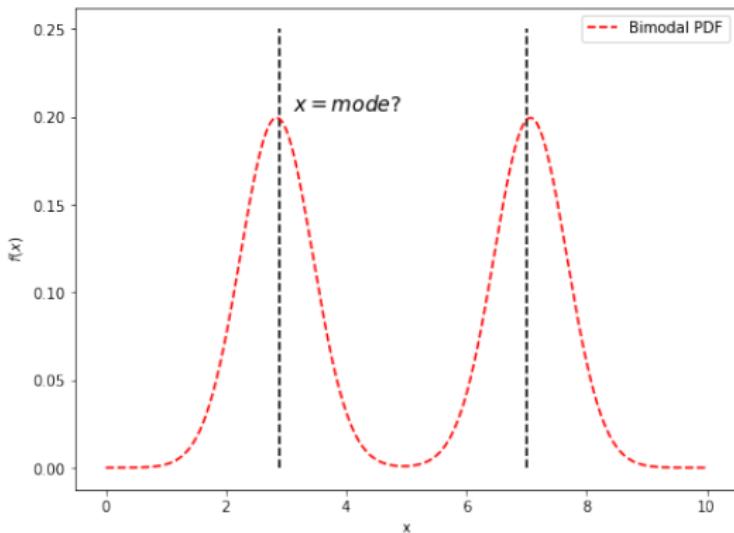
Basic concepts: mode

- The **mode** is the value of x for which $pdf(x)$ is maximal
 - The *typical location of the variable* is often better captured by the mode



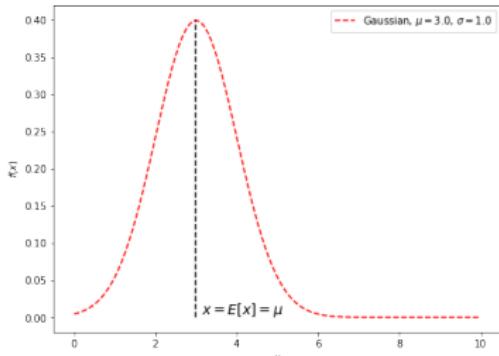
Basic concepts: mode

- mode is the value of x for which $pdf(x)$ is maximal
- when can this breakdown?



Basic concepts: expectation value

- The **expectation value** $E[x]$ of a variable x distributed according to $f(x)$ is often referred to as the **mean** μ .
- $E[x]$ is **not** a function of x , rather depends on form of $f(x)$.

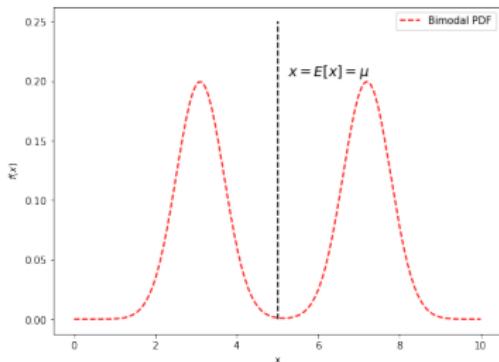


$$E[x] = \int_{-\infty}^{\infty} x.f(x)dx = \mu$$

- If the $f(x)$ is *concentrated in one region*, $E[x]$ represents a measure of where values of x are likely to be observed.
- When can this interpretation break down?**

Basic concepts: expectation value

- What if $f(x)$ is *multimodal*?, e.g, two gaussian peaks



$$E[x] = \int_{-\infty}^{\infty} x.f(x)dx = \mu$$

- x is never equal to μ !

Basic concepts: variance

- Functions of x also have expectation values
 - e.g. the expectation value of the squared difference between x and μ .
- $E[(x - \mu)^2]$ is called the **variance** V
 - V measures how *spread out* $f(x)$ is
 - Note $E[(x - \mu)^2] = E[x^2] - \mu^2$
- usually use the **standard deviation** σ instead
 - $\sigma = \sqrt{V}$

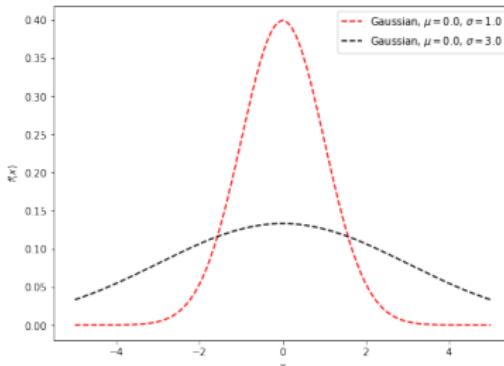


Figure: The two pdfs have the same μ but different σ

Basic concepts: variance

- Functions of x also have expectation values
 - e.g. the expectation value of the squared difference between x and μ .
- $E[(x - \mu)^2]$ is called the **variance** V
 - V measures how *spread out* $f(x)$ is
 - Note $E[(x - \mu)^2] = E[x^2] - \mu^2$
- usually use the **standard deviation** σ instead
 - $\sigma = \sqrt{V}$

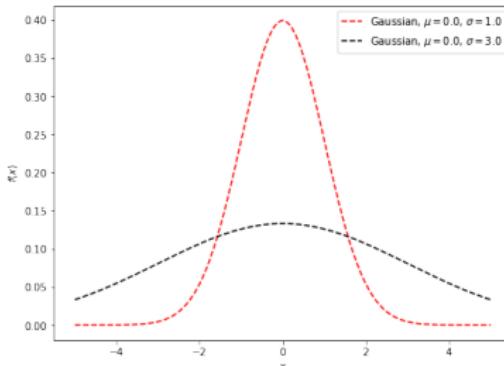
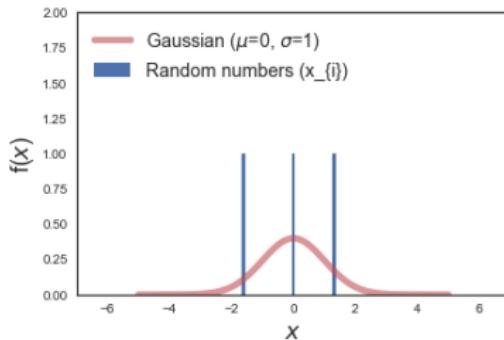


Figure: The two pdfs have the same μ but different σ

Basic concepts: *random numbers*

- We have been talking about abstract notions of probability
 - but what about real data?
 - imagine some data x_i : n observations of some quantity x
 - what then is the μ and σ of x_i ?

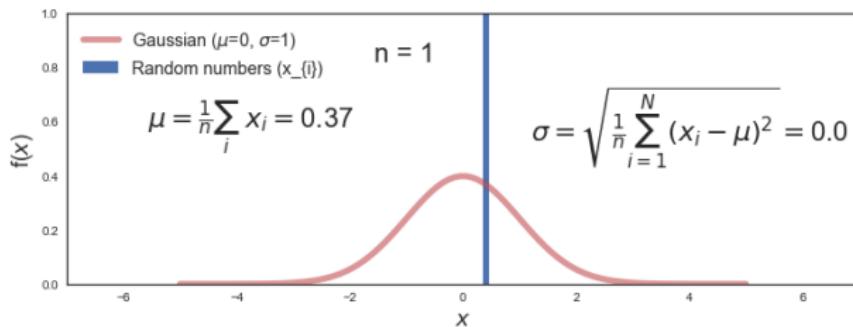


$$\mu = \frac{1}{n} \sum_{i=1}^n x_i, \quad \sigma = \sqrt{\frac{1}{n} \sum_{i=1}^N (x_i - \mu)^2}$$

- Let's think about how these definitions correspond to the defns. for pdfs

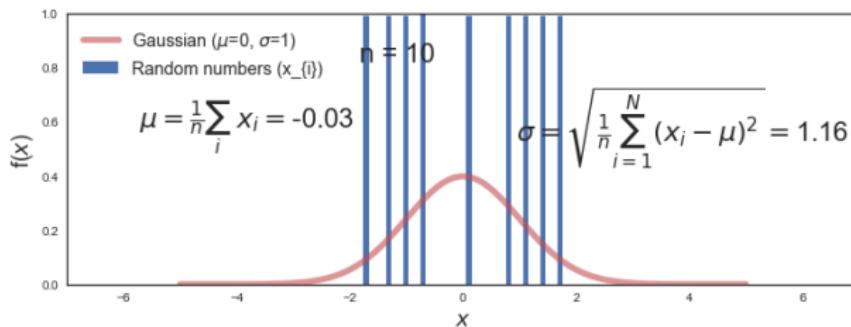
Basic concepts: *random* numbers

- *Random* numbers are useful in simulating data that is governed by a pdf
- Software tools can generate random numbers that are governed by any pdf...



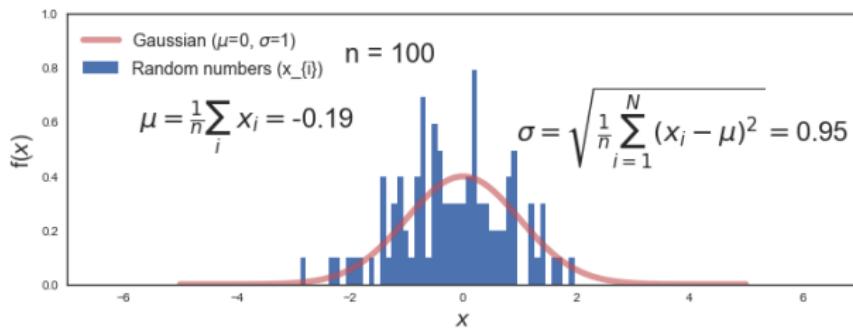
Basic concepts: *random* numbers

- *Random* numbers are useful in simulating data that is governed by a pdf



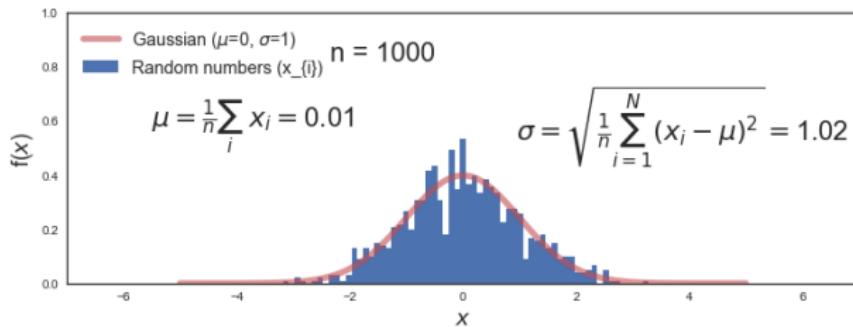
Basic concepts: *random* numbers

- *Random* numbers are useful in simulating data that is governed by a pdf



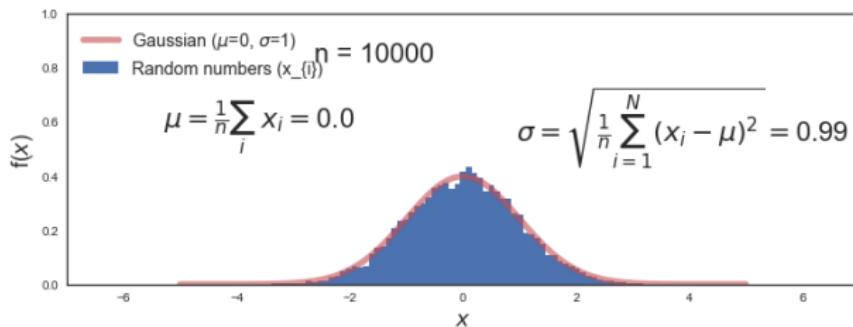
Basic concepts: *random numbers*

- *Random numbers* are useful in simulating data that is governed by a pdf



Basic concepts: *random* numbers

- *Random* numbers are useful in simulating data that is governed by a pdf

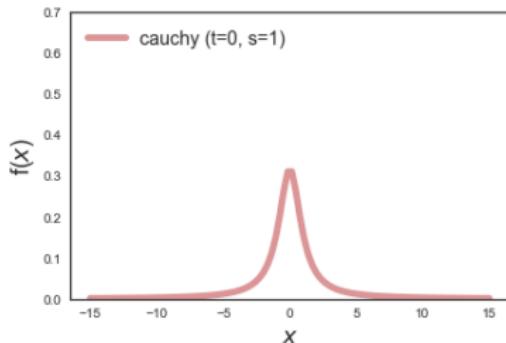


Basic concepts: mean & standard deviation limitations

- When the pdf has *fat tails*, μ and σ stop being useful
 - e.g. the Cauchy pdf

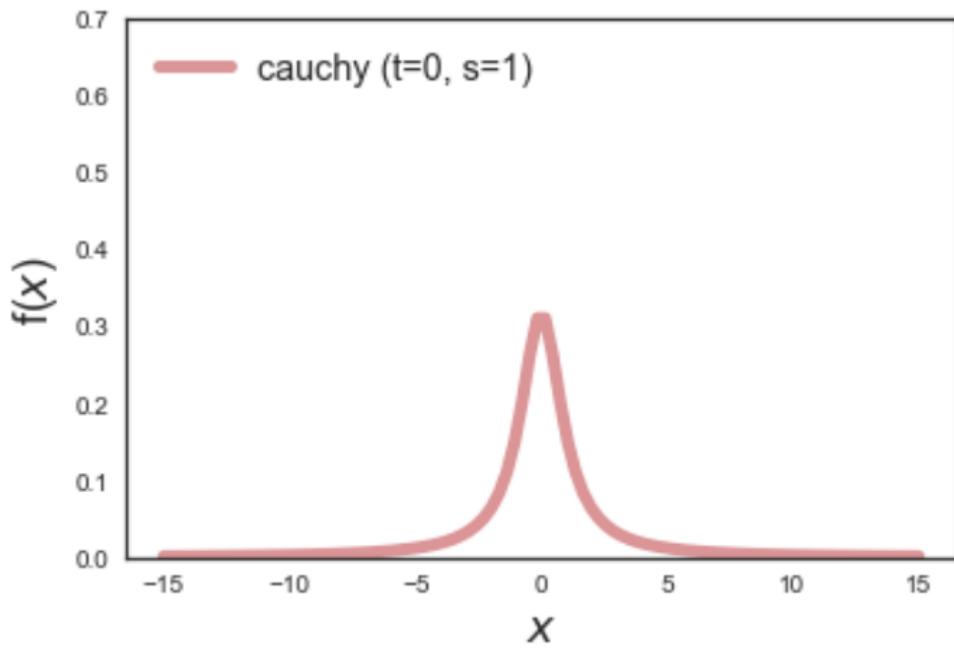
$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma} \left[\frac{\gamma^2}{(x - x_0)^2 + \gamma^2} \right],$$

- This pdf comes up a lot in physics
- $E[x]$ is undefined!
- $E[(x - \mu)^2]$ is undefined!



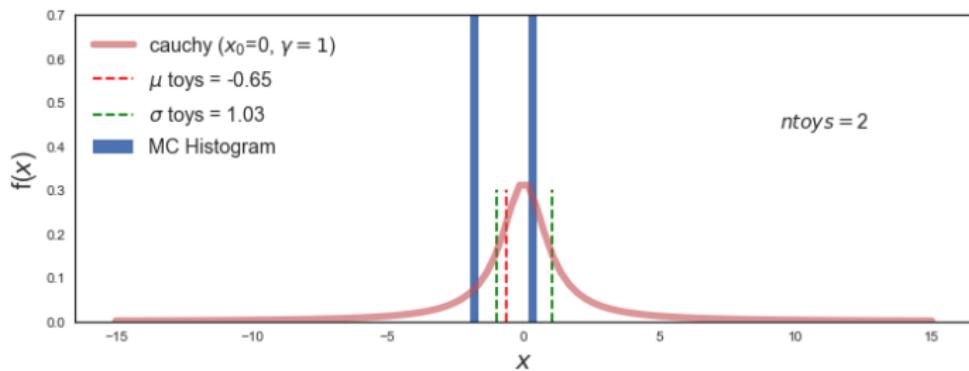
Basic concepts: mean & standard deviation limitations

- $E[x]$ is undefined!
- $E[(x - \mu)^2]$ is undefined!
- Taking the μ and σ of random numbers distributed according to a Cauchy does not work



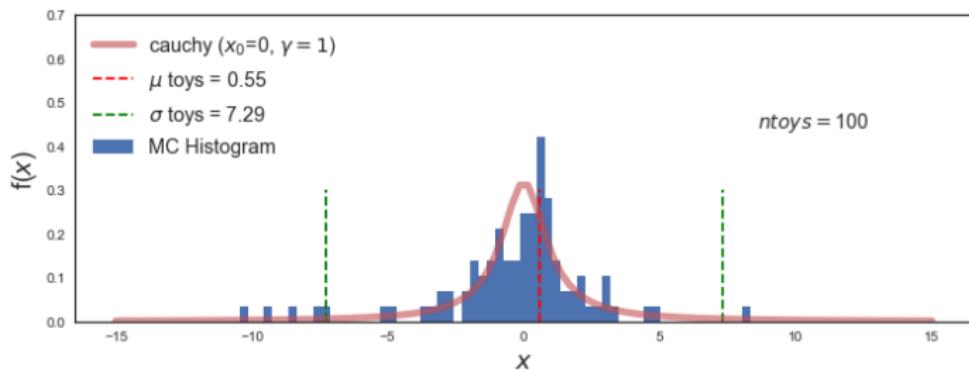
Basic concepts: mean & standard deviation limitations

- $E[x]$ is undefined!
- $E[(x - \mu)^2]$ is undefined!
- Taking the μ and σ of random numbers distributed according to a Cauchy does not work



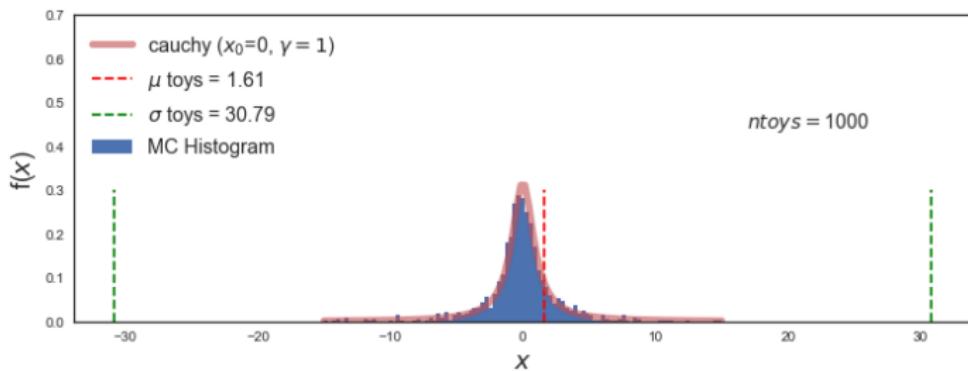
Basic concepts: mean & standard deviation limitations

- $E[x]$ is undefined!
- $E[(x - \mu)^2]$ is undefined!
- Taking the μ and σ of random numbers distributed according to a Cauchy does not work



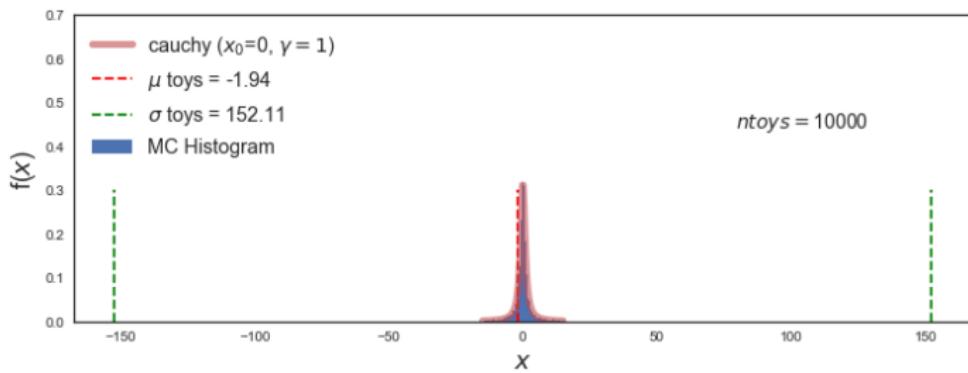
Basic concepts: mean & standard deviation limitations

- $E[x]$ is undefined!
- $E[(x - \mu)^2]$ is undefined!
- Taking the μ and σ of random numbers distributed according to a Cauchy does not work



Basic concepts: mean & standard deviation limitations

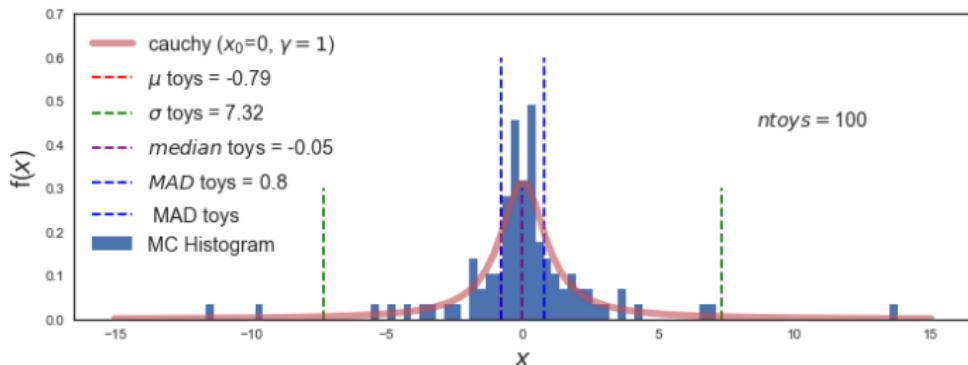
- $E[x]$ is undefined!
- $E[(x - \mu)^2]$ is undefined!
- Taking the μ and σ of random numbers distributed according to a Cauchy does not work



Basic concepts: alternatives: median and MAD

- If you suspect your data has fat tails, it's better to avoid the μ and σ
- Instead of μ how about the median?
- Instead of σ how about something MAD? (Mean Absolute Deviation)

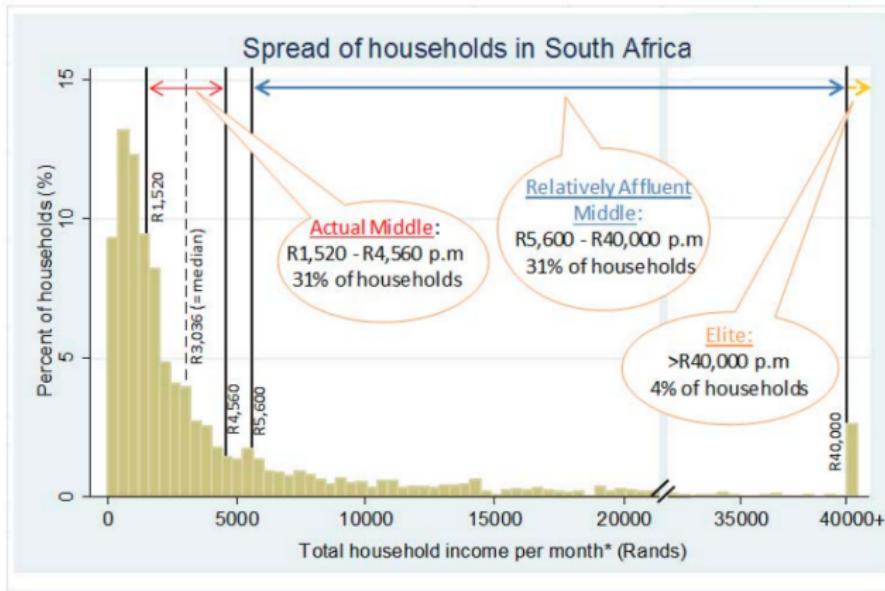
$$MAD = \frac{1}{n} \sum_{i=1}^n |x_i - \mu(x)|$$



Basic concepts: alternatives: mode

- When does the median fail?

Figure 1: The spread of households within the income distribution in South Africa, 2008



Source: NIDS 2008, own estimates

Figure: Source: Who are the middle class in South Africa? Does it matter for policy? Visagie 2013