

# ATLAS (Large Hadron Collider) Open Data Research Project

Department of Physics and Astronomy, Queen Mary University of London

## 1 Introduction

The goal of this project is to understand the ways in which particle physicists analyse data and to give it a go yourselves. All words in **bold print** are defined in the Glossary.

### 1.1 Units in Particle Physics

You will be accustomed to using SI units in your school work, however in the world of particle physics they are not very useful. The sizes of these units are based on arbitrary figures which are applicable to more everyday measurements, however they tend to be much too small (for energy) or much too large (for mass) for quantities used in particle physics. So, particle physicists use units that relate to the physical quantities they are studying. For example, instead of using Joules, they use electron volts, eV. An electron volt corresponds to the work done on one electron as it is passed through a potential difference of 1 V. Particles measured at places like CERN are travelling very fast - close to the speed of light. So we need to consider them as being **relativistic**.

- Speed of light:

$$c = 2.99 \times 10^8 \text{ m s}^{-1}$$

- Energy:

$$1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$$

- Mass:

$$m = \frac{E}{c^2}$$
$$\Rightarrow \text{Kg} = \text{JC}^{-2} = \text{eVC}^{-2}$$

For any system (a planet, a person, a particle, etc.) the total energy E, momentum p and mass m are related by the **relativistic** formula:

$$E^2 = p^2 c^2 + m^2 c^4$$

There are lots of 'c's in this equation and so physicists often set c and other physical constants to be equal to 1. These are called **natural units**. This means that we can treat energy, momentum and mass as being equivalent quantities in order to compare them. It is always possible to add values for c back in later. Using natural units here we can obtain kinematic equations which can be used to calculate particle quantities,

$$m^2 = E^2 - p^2$$

where p is momentum. So by knowing the energy and momentum associated with a particle one can calculate its **rest mass**.

#### Exercises

1. The Higgs boson has a mass of approximately 125 GeV, what is this in kg?
2. An electron has a mass of  $9.1 \times 10^{-31}$  kg, what is this in MeV?

3. In the everyday-life version of momenta  $p = mv$  we encounter a problem with massless objects like the photon, their momentum would be 0. What happens with the momentum in case of a massless particle in the special relativity version of momentum?

## 1.2 The Standard Model

The world as we know it is made up of matter. But what is matter? We know that atoms make up everything, and atoms are made up of protons, neutrons and electrons. Protons and neutrons are then made up of quarks which are **elementary particles** (also called fundamental particles).

The Elementary particles are summarised in a table called The Standard Model as shown in Figure 1 which groups them into categories called generations or families:

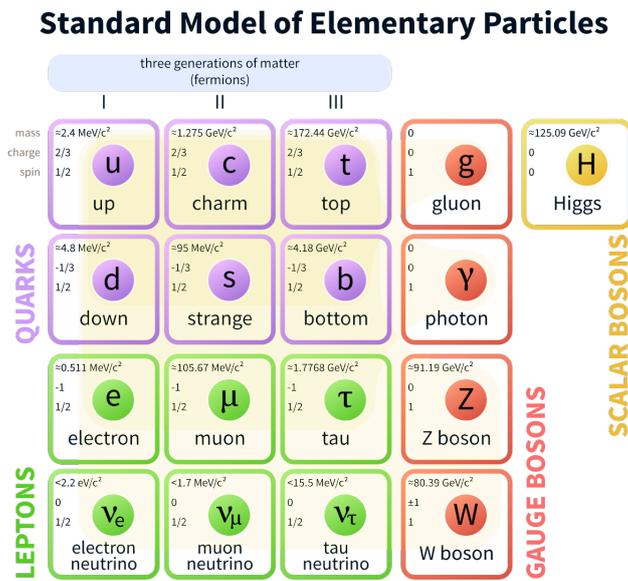


Figure 1: The Standard Model of Elementary Particles

There are 6 **flavours** of **quarks**; up, down, strange, charm, top, bottom. There are also 6 flavours of **leptons**; electron, electron neutrino, muon, muon neutrino, tau, tau neutrino. These are all **fermions**.

It also includes **force carriers** which are **bosons**. These are particles which are exchanged between matter and define the four fundamental forces: Strong, electromagnetic, weak and gravity. The most recent addition to the Standard Model is the Higgs boson which is the carrier particle for the Higgs field.

### Exercise

4. A mystery particle is found to have energy,  $E = 230 \text{ GeV}$  and momentum,  $p = 1.5 \times 10^{11} \text{ eV c}^{-1}$ . Find the mass of the particle, and then see if you can identify it using the standard model.

## 1.3 The Higgs Boson

The Higgs boson is the most recent addition to the standard model. It was theorised by François Englert and Peter Higgs in 1964, and they were awarded the 2013 Nobel Prize for it. Much work has been produced on it since in refining detection of it and physicists are gradually gaining a deeper insight into its mechanisms. It is colloquially known as the 'god particle' due to its role in giving particles their mass.

The mass of the Higgs was not part of the theory and so was unknown until it was first detected and measured. Since then, measurements have become far more accurate.

As a massless particle travels through a Higgs field it can become surrounded by a cloud of virtual Higgs particles causing it to travel slower through the field and gain mass. Only certain particles interact with the Higgs field and to different extents thus giving the range of masses existent in different particles.

A good analogy for the Higgs mechanism is depicted in Figure 2 and works as follows:

There is a pub full of physicists and Einstein walks in. He wants to get to the bar at the other side of the room, but as he walks through the crowd, people want to talk to him and gather around him slowing him down. He becomes a more massive entity as he walks through and people gather around him. Not just anyone would have this effect on the crowd however, if a normal, uninteresting person were to walk to the bar, they would have no problems at all and go straight through without gaining any followers. A photon works like this.

This explained the unsolved problems associated with **Quantum Field Theory**. It dictated that all particles must be massless which was nonsensical as you couldn't then just add mass on to particles. The Higgs mechanism provided an answer for this.

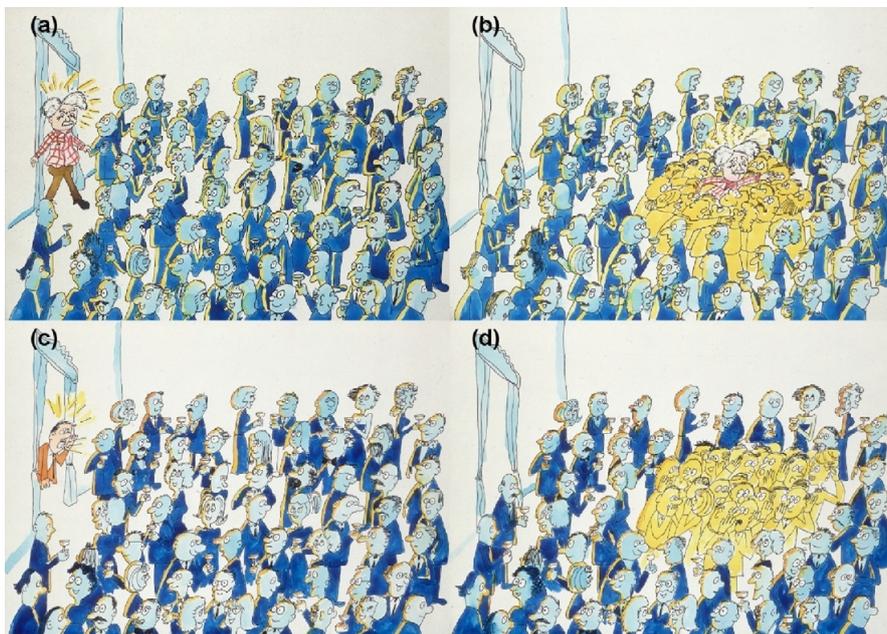


Figure 2: a,b) As Einstein makes his way through the crowd he attracts the attention of lots of people and becomes more massive. This corresponds to a particle gaining mass as it travels through a Higgs field. c,d) A message is passed through the crowd with people gathering together to share it. This is akin to the Higgs boson itself.

The Higgs boson itself can be compared to a generic clumping of people in the pub, think of a message be passed through the room from group to group (see Figure 2). It can be produced through a variety of interactions, the most common at the LHC is gluon-gluon fusion.

The Higgs can **decay** into various pairs of fermions or bosons and depending on the mass of the Higgs particle, different decays have higher and lower probabilities of occurring. These decay products are then detected and so signify the presence of a Higgs in an interaction. Each decay is characterised by a **branching ratio**. Essentially this gives the probability of a certain decay occurring over the other types. Figure 3 shows a graph of the different decays arranged by their branching ratios. The decays with the highest branching ratios aren't always the easiest to detect however.

The significance of the Higgs within the standard model led to a 40 year hunt for it. It had become a very important part of particle physics, but was yet to be observed. It would be the final piece of the standard model. This hunt motivated the construction of the Large Hadron Collider at CERN, designed specifically to be able to try to find it.

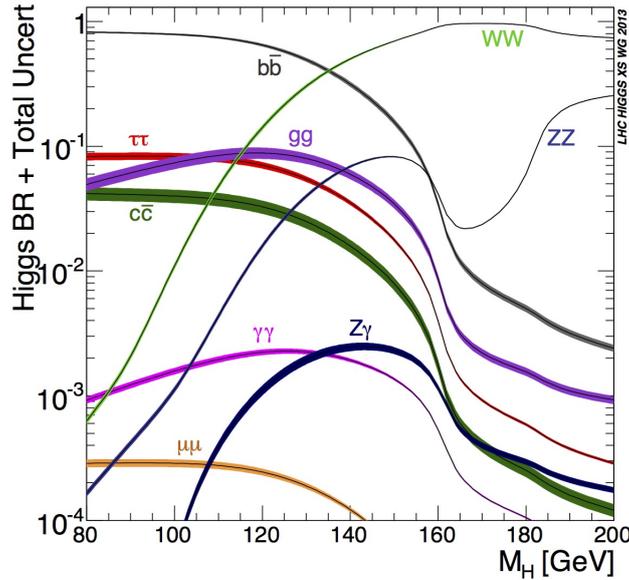


Figure 3: Graph summarising branching ratios of various Higgs decays.

## 2 Particle Detection

The Large Hadron Collider (LHC) at CERN has four main particle detectors: ALICE, ATLAS, CMS and LHCb. These detectors are showered with the particles produced when the two beams of protons circulating around the LHC collide. Each layer of a detector has a specific function, but the main components are semiconductor detectors which measure charged particles and calorimeters which measure particle energy. The semiconductor detectors use materials such as silicon to create diodes - components that conduct electrical current in only one direction. Charged particles passing through the large number of strips of silicon placed around the proton beam collision point create a current that can be tracked and measured.

**Calorimeters** are calibrated to be either electromagnetic or hadronic. The former will detect particles such as electrons and photons, while the latter will pick up protons and neutrons. Particles entering a calorimeter are absorbed and a particle shower is created by cascades of interactions. It is the energy deposited into the calorimeter from these interactions that is measured. Stacking calorimeters allows physicists to build a complete picture of the direction in which the particles travelled as well as the energy deposited, or determine the shape of the particle shower produced.

ATLAS and CMS are also designed to detect muons, particles much like electrons but 200 times more massive, which pass right through the other detection equipment. A muon spectrometer surrounds the calorimeter, and functions in a similar way to the inner silicon detector.

However, there are often particles which are impossible to detect on their own due to short lifetimes, neutral charges etc. However it is possible to infer their existence by detecting their subsequent interactions and decays and extrapolating information on their position, mass or energy from these and tracking back their paths. It is almost like tracing its footsteps back through time. In order to do this it is necessary to build up a relatively accurate map of those particles that can be detected. These interactions are discussed in terms of their **vertices**, with the **primary vertex** being the initial proton collision with the highest **transverse momentum**.

### 2.1 ATLAS

ATLAS is one of two general-purpose detectors at the LHC. It investigates a wide range of physics, from the discovery of the Higgs boson to the search for extra dimensions and particles that could make up **dark matter**. Beams of particles from the LHC collide at the centre of the ATLAS detector making collision debris in the form of new particles, which fly out from the collision point in all directions. Six

different detecting subsystems arranged in layers around the collision point record the paths, momentum, and energy of the particles, allowing them to be individually identified. Figures 4 and 5 show schematics of the detection system at ATLAS.

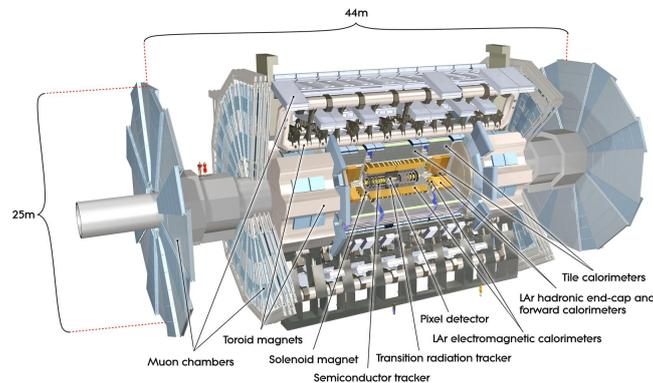


Figure 4: A schematic of the ATLAS detector. This diagram shows it to scale with two people standing on it for size comparison.

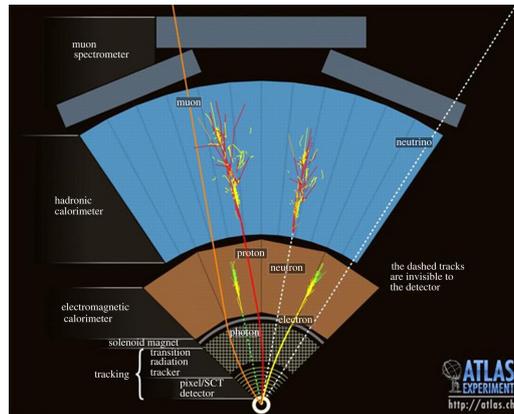


Figure 5: A cross section of the detector showing where each particle would be detected.

A large magnet system bends the paths of charged particles so that their momenta can be measured. This works because particles with more momentum bend less than those with greater momentum as they are in the magnetic field for less time. The particles bend according to the following equation.

$$\vec{F} = Q\vec{v} \times \vec{B}$$

$$|F| = |Q| |B| |v| \sin \theta$$

Where  $\vec{F}$  is the force acting on the particle,  $Q$  is the charge of the particle,  $\vec{v}$  is the velocity of the particle,  $\vec{B}$  is the strength of the magnetic field and  $\theta$  is the angle between the particle's velocity and the magnetic field. The momentum is then calculated by measuring a particle's **gyroradius**, this can then be related to the momentum by the following equation

$$r_g = \frac{p_{\perp}}{|Q|B}$$

where  $p_{\perp}$  is the component of momentum perpendicular to the direction of the magnetic field.

### Exercises

5. What is the force on a proton travelling at 99% the speed of light at  $45^\circ$  to a 4 T magnetic field?
  
6. Calculate the gyroradius of a 1 GeV electron in a 2 T magnetic field if its velocity is:
  - (a) Directed along the magnetic field.
  
  - (b) Perpendicular to the magnetic field.
  
  - (c) At  $30^\circ$  to the magnetic field.

The interactions in the ATLAS detectors create an enormous flow of data. To digest the data, ATLAS uses an advanced **trigger system** to tell the detector which events to record and which to ignore. Complex data-acquisition and computing systems are then used to analyse the collision events recorded. At 46 m long, 25 m high and 25 m wide, the 7000-tonne ATLAS detector is the largest volume particle detector ever constructed. It sits in a cavern 100 m below ground near the main CERN site, close to the village of Meyrin in Switzerland. More than 3000 scientists from 174 institutes in 38 countries work on the ATLAS experiment.

### 3 Histogram Analyser

Here you can take on the role of a real particle physicist and play with some real data from ATLAS using the Histogram analyser. The histograms give data from W boson pair production from Higgs decay - the signature process for identifying Higgs, and three background processes, WW,  $t\bar{t}$  and Z (these are separate processes which have similar signature products to the desired Higgs process).

The signature process:  $H \rightarrow W^+W^-$  goes on to decay to pairs of particles (quark-antiquark,  $q\bar{q}$ , lepton-antineutrino,  $\ell^-\bar{\nu}$ , antilepton-neutrino,  $\ell^+\nu$ ) and is one of the most likely decays for a Higgs. In this case we are looking for the following process:

$$H \rightarrow W^+W^- \rightarrow \ell^-\ell^+\nu\bar{\nu}$$

where the leptons are electrons or muons here. It is identified by an isolated lepton with high **transverse momentum** ( $p_T$ ). The *Get Started* book then goes into more detail about the background decays and how they are identified. The *Separate Signals* tab describes how you can analyse the signal data to help you know which parts of each graph to cut to increase signal significance and summarises characteristics of each process.

#### 3.1 Significance

Significance is introduced to quantify the probability of a statistical fluctuation expected to observe a certain events or more when you know how many background events you expect. It therefore measures the likelihood that the number of events you observed were just a statistical fluke or whether some other process may be present. The idea is to get as high a **significance** as possible for the H to WW decay, thereby minimising contributions from background decays.

In large counting experiments we use Poisson statistics, which tell us that the standard deviation of the number of events should go as  $\sqrt{N_k}$  where  $N_k$  is the number of events of type  $k$ .

#### Exercises

7. Flip a coin  $M$  times (e.g. 10) and count the number of heads. Repeat this a few times (e.g. 5) and calculate the mean and standard deviation of the number of heads in each try. Is the mean number of heads consistent with the expected value  $pM$ , where  $p = 0.5$  is the probability of landing heads, and the standard deviation around  $\sqrt{pM}$ ?

8. Retry the experiment this time with a dice, i.e. changing  $p$ , and compare again with the theoretical result.

From this fact we can calculate the significance (in units of standard deviations or  $\sigma$ ) by using the approximate formula

$$z = \frac{\text{Number of signal events}}{\sqrt{\text{Number of background events}}} \quad (1)$$

In Particle Physics, for there to be considered evidence of a new result requires at least  $3\sigma$  and for a discovery they have to be at least  $5\sigma$ , corresponding to a 1 in 3.5 million chance of simply being down to statistical fluctuations. By making **cuts** on the different histograms it is possible to filter the total data for each background decay and possibly increase the significance of the signature decay.

## 3.2 Find the Higgs

Now use the *Find the Higgs* tab to isolate some Higgs processes. As you're going note down the cuts you have made and the effects they have and take screenshots of the graphs as you go. Try not to just use trial and error, but instead work out exactly which cuts you need to make and why these are making a difference to the significance. As explained in the *Get Started* book, you can separate the signals in order to see what characteristics each signal has on each graph.

Compare the results you obtain from the real and simulated data and see if there is any difference and if so, why that might be.

## 4 Research

Now that you've gained some first-hand experience with using ATLAS data, it is up to you to continue this work with your own independently motivated research. It is imperative that you perform a literature review before continuing, reading up on and around the underlying particle physics. Here we merely suggest some avenues you may wish to explore in your own research:

- Why do your selected cuts optimise a search for the Higgs boson?
- Can you estimate the amount of data you might need for a statistically significant Higgs boson result based on the data so far?
- Can you use the data to measure various unstable particles' masses and/or lifetimes?
- Can you explain any differences between the data and Monte Carlo simulations?

You may use one of these topics as inspiration or come up with one of your own to research. Be sure to read around your chosen topic and undertake an experiment using the data and tools provided that will address your research question.

**If you're still unsure what to do, please tell your teacher to get in touch with us so we can visit and provide guidance and assistance. If at any point if you find yourself unable to continue or completely unsure about something, ask your teacher to get in touch with us so that we can help you.**

## 5 Tools for projects

The recommended project is ttZ because it has the clearest webpages and easiest to use code. However there are some further options on the Higgs that are further down.

## 5.1 ttZ project

Key links for ttZ project:

1. General rare to processes Introduction: <http://opendata.atlas.cern/release/2020/documentation/visualization/rare-top-processes.html>
2. ttZ analysis introduction: <http://opendata.atlas.cern/release/2020/documentation/visualization/ttZanalyses.html>
3. Histogram analyser app: <http://opendata.atlas.cern/release/2020/documentation/visualization/histogram-analyser-ttZ.html>
4. Online strategy for analyser: [http://opendata.atlas.cern/release/2020/documentation/visualization/separate\\_signals\\_ttZ.html](http://opendata.atlas.cern/release/2020/documentation/visualization/separate_signals_ttZ.html)
5. More details/examples: [http://opendata.atlas.cern/release/2020/documentation/visualization/find\\_ttZ.html](http://opendata.atlas.cern/release/2020/documentation/visualization/find_ttZ.html)
6. Jupyter notebook for more advanced work: [https://github.com/atlas-outreach-data-tools/notebooks-collection-opendata/blob/master/13-TeV-examples/uproot\\_python/ttZ\\_ML\\_from\\_csv.ipynb](https://github.com/atlas-outreach-data-tools/notebooks-collection-opendata/blob/master/13-TeV-examples/uproot_python/ttZ_ML_from_csv.ipynb)
7. Input csv file for advanced work: [http://opendata.atlas.cern/release/2020/documentation/visualization/CrossFilter/13TeV\\_ttZ.csv](http://opendata.atlas.cern/release/2020/documentation/visualization/CrossFilter/13TeV_ttZ.csv)

A further document (chapter from M Owen thesis) has been provided to teachers). Once you have got a good score on the histogram analyser, roughly matching the thesis document, make sure you understand what you have done. Then consider the following next steps (which get bigger as the list goes on):

1. Write your own code based on the jupyter notebook to duplicate the histogram analyser, including getting the same value for significance.
2. Change the code to get a better selection if you can
3. Make 2d plots, looking for correlations in pairs of variables
4. Consider modifying your code to make more complicated selections (e.g. based on linear combinations of variables)
5. Apply a machine learning algorithm to make improved selections

## 5.2 Higgs project

WARNING: the resources here are more challenging to use than the ttZ project.

1. Histogram analyser: <http://opendata.atlas.cern/visualisations/analyser-js.php>
2. Starting python code for more advanced selections on the histogram analyser data: <https://github.com/oliverh100/HiggsAnalysis>
  - Credit: Highgate School partnership
  - Principal author: Oliver Herskovits
  - Corresponding author: Anson Cheung
3. ATLAS Open Data Jupyter Notebooks using latest dataset:
  - Launch link: <http://opendata.atlas.cern/release/2020/documentation/notebooks/intro.html>
  - Manual: <http://opendata.atlas.cern/release/2020/documentation/notebooks/analysis-examples.html>
  - See also links on the top bar!

## 6 Glossary

- **Boson** - The particles associated with all the fundamental interactions (forces) and composite particles with even numbers of fermion constituents (quarks) are bosons.
- **Branching ratio** - The fraction of particles which decay by an individual decay mode with respect to the total number of particles which decay.
- **Calorimeter** - An experimental apparatus that measures the energy of particles. Most particles enter the calorimeter and initiate a particle shower and the particles' energy is deposited in the calorimeter, collected, and measured.
- **Dark matter** - Matter that exists but is not visible to us because it emits no observable radiation.
- **Decay** - A process in which a particle disappears and in its place different particles appear. The sum of the masses of the produced particles is always less than the mass of the original particle
- **Elementary particles**- A particle with no internal substructure, i.e. it cannot be broken down into any further parts as far as we know.
- **Fermion** - Fermions obey a rule called the Pauli Exclusion Principle, which states that no two fermions can exist in the same state at the same time. All particles are either fermions or bosons. All quarks and leptons are fermions.
- **Flavour**- The name used for the different quarks types (up, down, strange, charm, bottom, top) and for the different lepton types (electron, muon, tau). For each charged lepton flavour there is a corresponding neutrino flavour. In other words, flavour is the quantum number that distinguishes the different quark/lepton types. Each flavour of quark and lepton has a different mass.
- **Gyroradius** - The radius of curvature of a charged particle in the presence of a uniform magnetic field.
- **Mass**- The rest mass ( $m$ ) of a particle is the mass defined by the energy of the isolated (free) particle at rest, divided by the speed of light squared. When particle physicists use the word 'mass' they normally mean the 'rest mass' ( $m$ ) of the object in question.
- **Primary vertex** - The inelastic collision of two protons with the highest overall transverse momentum.
- **Quantum Field Theory** - A field theory that incorporates quantum mechanics and the principles of the theory of relativity.
- **Relativistic** - At speeds close to the speed of light.
- **Transverse momentum** - The component of momentum perpendicular to the beam line.
- **Trigger system** - This selects events with distinguishing characteristics that make them interesting for physics analyses.
- **Vertex** - The point of collision of two particles - the source of the interaction.
- **Virtual machine** - A virtual machine will transform your computer into an analysis machine. Your physical computer will be the "host", while the virtual machine will be a "guest". Most of the guest code runs unmodified, directly on the host computer, and the guest operating system "thinks" it's running on a real machine.

## 7 Useful links

- Video explaining the Higgs mechanism - <https://www.youtube.com/watch?v=joTKd5j3mzk>
- Video on significance in Particle Physics - <https://www.youtube.com/watch?v=Gr1rIikTYac>
- Paper on significance in Particle Physics - <https://arxiv.org/pdf/hep-ex/0208005.pdf>
- Statistics - [http://www.phys.ufl.edu/~korytov/phz6355/note\\_A13\\_statistics.pdf](http://www.phys.ufl.edu/~korytov/phz6355/note_A13_statistics.pdf)