Lab 3: ATLAS Data Analysis

28-Apr-2025 (v1)

Introduction

Imagine you are a physicist working on the ATLAS (**A** **T**oroidal **L**hc **A**pparatu**S**) experiment at CERN in Geneva. At ATLAS, high energy protons are smashed together, and the byproducts are studied. One of the most important measurements to be made is the mass of the Z^0 boson.

After you have done your investigations — which you are guided through below — you will prepare a short report on your findings for grad students in your research group. More details about the report can be found at the end of the document. But given that it's for graduate students, who are very busy, your final report should be short and to the point (3 pages maximum), quantitative in nature, and contain highly polished plots, with appropriate titles, units, etc.

A table of relevant physical constants, from the 2024 edition of the Particle Data Group, is below. Good luck!

Quantity	Value
Mass of Z Boson (m_{Z^0})	91.1880 ± 0.0020 GeV/c^2
Mass of W Boson (m_W)	80.3692 ± 0.0133 GeV/c^2
Mass of Higgs Boson (m_H)	125.20 ± 0.11 GeV/c^2
Mass of Electron (m_e)	$0.51099895000 \pm 0.00000000015$ MeV/c^2
Mass of Muon (m_μ)	$105.6583755 \pm 0.0000023$ MeV/c^2
Mass of Tau (m_τ)	1776.93 ± 0.09 MeV/c^2

Introduction

At the Large Hadron Collider, at CERN, in Geneva, Switzerland, particle physicists collide beams of protons. This process breaks the protons open, and more fundamental particles are formed, interact, and decay. One of the most interesting fundamental particles to come out of proton-proton (pp) interactions is the Z^0-boson, which is the neutral carrier of the *weak* force, and is therefore responsible, along with the $W^{\pm}$-boson, for facilitating many nuclear interactions in the Universe. The photon, which you are more familiar with, is the carrier of the *electromagnetic force.*

The Z^0 is unstable and decays. About 10% of the time, it decays into a pair of charged leptons. This can be an electron (e) and and anti-electron (or positron, e^+). It can also be a muon/anti-muon pair, or a tau/anti-tau pair. Because it can do any of these, we substitute the generic letter ℓ for lepton, and this interaction is known as $Z^0 \to \ell\bar{\ell}$, where the "bar" over the ℓ indicates an anti-particle. Because charge cannot be created or destroyed, they must have opposite (or no[1]) charge. And, because matter and energy cannot be created or destroyed, if Z^0 particles are decaying to produce the leptons, then the total energy stored in the two leptons must sum to (at least) the mass of the Z^0. This means that if we can measure the energy of all double-lepton events in the detector, we should see an *excess* or a *peak* at the mass of the Z_0.

Part 1: The Invariant Mass Distribution

In the ATLAS detector (which is one of four main experiments at the LHC), it is reasonably straightforward to measure four properties of particles that come out of the proton-proton interactions. The first is the total energy E. The second is the transverse-momentum p_T, which describes the momentum the particle has in the transverse direction. The third is the pseudorapidity η, which describes the angle the particle makes with respect to the beamline. If the particle continues straight along the beamline, then $\eta \to \infty$, while if the particle is deflected out at 90^o, it has $\eta = 0$. The fourth is the azimuthal angle ϕ about the beam. That is, if you are staring down the barrel of the collider, a particle with $\eta, \phi = 0, 0$ emerges from the interaction point flying directly to the right. Where a particle with $\eta, \phi = 0, \pi/2$ emerges from the top flying straight up.

Together, these values fully define the *four momentum* of the particle: $p = (E, p_x, p_y, p_z)$ through the following mathematical relationships, where c (the speed of light) is treated in a strange fashion and set equal to 1 (which you will learn more about when you take a course that covers relativity and discuss "natural units"):

$$p_x = p_T \cos(\phi), \quad p_y = p_T \sin(\phi), \quad p_z = p_T \sinh(\eta) \tag{1}$$

The difference between the three-momentum and the energy is the particle's invariant mass:

$$M = \sqrt{E^2 - (p_x^2 + p_y^2 + p_z^2)} \tag{2}$$

If you have *two* particles, and you would like to know the total momentum of the system, you have to sum the four momenta: $p_{tot} = p_1 + p_2$.

[1] It is also possible for the Z^0 to decay into two photons, or $Z^0 \to \gamma\gamma$

In the file `atlas_z_to_ll.csv`, you will find five-thousand real ATLAS data events. These have been pre-selected from the 2020 ATLAS open dataset so that the "final states" only contain the two leptons we are interested in. The first two columns are the p_T (in GeV) for the two leptons; the next two columns are the η, the next two are ϕ (radians), and the last two are the energy E (in GeV) [2].

1. Load the data into python.
2. For each lepton pair, using the formulas in Equ. 1 and Equ. 2, calculate the mass of a hypothetical particle which decayed to produce that pair. (Hint: first, calculate the vector components, then calculate the summed four-momenta, then calculate the mass.)
3. Make a histogram, with error bars, of your calculated invariant mass. You should approximate this as a Poisson counting experiment, so that the error on the number of events in the bin is equal to the square-root of the number of events in the bin. That is: $\sigma = \sqrt{N}$. Label the axes nicely, with units, etc. To make uniform grading possible, please do the histogram from 80 to 100 GeV with 41 bins. That is: `bins = np.linspace(80,100,41)`.

Part 2: Breit-Wigner Fit

You can show, using scattering theory, that the distribution of decays $\mathcal{D}$ at a reconstructed mass m follows what is known as a Breit-Wigner (to a mathematician, "Cauchy-Lorentz") peak. The distribution depends on true rest-mass of the Z^0, m_0, and on a "width" parameter Γ:

$$\mathcal{D}(m; m_0, \Gamma) = \frac{1}{\pi} \frac{\Gamma/2}{(m - m_0)^2 + (\Gamma/2)^2} \tag{3}$$

In nature, the true width parameter Γ_0, is related to the lifetime of the particle by the Heisenberg uncertainty principle. In a real detector, we can only measure the width subject to experimental uncertainties, and $\Gamma_{exp} > \Gamma_0$.

1. Code up a function that returns the decay distribution as a function of m, m_0, and Γ.
2. Fit your mass-distribution with the Breit-Wigner function. Fix the overall normalization to half the number of data points in the set. That is, you should fit $\frac{5000}{2} \times \mathcal{D}$. To make uniform grading possible, please do the fitting where the bin centers are > 87 to < 93 GeV *only*. But keep the same 41 bins from 80 to 100. If you were to make a numpy mask, it would be: `mask = (bin_centers > 87) & (bin_centers < 93)`.
3. In a second, new plot, plot the data (with error bars), and overlay your fit. Label the axes well, add a nice legend, etc. In a sub-panel, plot the residuals between the data and the fit, and draw a horizontal line at zero to indicate perfect agreement (like we did in Lecture 23). Draw two vertical dotted lines to denote the fitting range.
4. Calculate the chi-square, reduced-chi-square, and p-value of your fit *in the fitting range*.
5. Using the covariance matrix, calculate the best fit mass m_0, and its uncertainty.
6. Annotate your plot with your best fit for the mass, its uncertainty, the chi-square/NDOF, and the p-value. Use only one decimal place for all numbers.
   Label this plot clearly as **Figure 1**.

[2] One nifty utility of natural units is that momentum and energy have the same units, which makes doing math with them much easier.

Part 3: 2D Parameter Contours

As emphasized in class, this is a 2D fit, and you cannot determine Z^0 and Γ_{exp} independently. In this part, you visualize the joint probability space.

1. Perform a 2D chi-square scan of the mass-width parameter space. To make grading easier, please scan in mass from 89 to 91 GeV, and the width from 5 to 8, with 300 bins along each dimension.

2. Make a filled contour plot of the $\Delta\chi^2 = \chi^2 - \chi^2_{min}$. Clip the $\Delta\chi^2$ at 35 units to make the plot easier to see. Add a colorbar. Make the plot look nice, with appropriate labels. Note that this is a $\Delta\chi^2$ map. This means the minimum value on the z-axis should be zero.

3. Draw the 1σ and 3σ confidence levels onto the plot using a solid and dashed line, respectively. Use the matplotlib `clabel` capability to label the levels appropriately. (Hint 1: If you need a refresher on how to do this, please check the matplotlib documentation page. Hint 2: You can look up the $\Delta\chi^2$ corresponding to the 1σ and 3σ levels online, or they can be seen in Lecture 22. Pay close attention to how many degrees of freedom you have.)

4. Draw a dot/cross at the best fit location from Part 2. Label this plot clearly as **Figure 2**.

Report Preparation & Code Submission

Submit your code to GitHub in a notebook with all figures included as well, along with your 3-page report. Two, and only two, figures should be embedded in your report: the Figure 1 and Figure 2 specified above. Your report will be read by your graduate students. They are in your research group, but obviously less experienced than you. As such, your report will need to give an introduction to the physics, and some basic explanations of how you did your plotting and how you made your calculations. You want your report to include:

I Introduction

II The Invariant Mass Distribution and its Fit, including Figure 1. This section should clearly state five main numbers, to 1 decimal place. Namely: the fitted mass of the Z^0 (in GeV), the fitted uncertainty on the Z^0 mass (in GeV), the chi-square, the number of degrees of freedom, and the pvalue.

III The 2D Parameter Scan, including Figure 2

IV Discussion and Future Work

In particular, Section IV should include a brief comparison of your measured Z^0 relative to the latest accepted values from the PDG. It should also include a summary of the approximations you have made in doing your calculations, and future work necessary to make the calculations more realistic. For example, your fit does not include any systematic uncertainties, or the energy resolution of the ATLAS detector.

Rubric (60 points total)

1. (10 pts) Professionalism and style
   (a) Is the text complete (has all five sections) and comprehensible? Is the report signed?
   (b) Is it written in complete sentences, free of grammatical errors and typos?
   (c) Is the tone appropriate for a technical report?
   (d) Is the length appropriate? (2 - 3 pages, including the two [and only 2!] requested plots). No last blank page.
   (e) Are the plots professional, with appropriate labels, units, etc.?
2. (5 pts) Git
   (a) The jupyterlab notebook is well documented and organized, e.g. with markdown cells
   (b) Is the first cell [1] and increments by 1 showing it's a reproducible notebook
   (c) The notebook is correctly committed, and *pushed*, to GitHub in the Lab3 folder
3. (5 pts) Introduction
   (a) Does the introduction provide a high level summary of the project?
4. (25 pts) Invariant Mass Distribution
   (a) This section should include a description of what is plotted, how it was calculated, and what model that is fitted.
   (b) Contains Figure 1: the plot of invariant mass distribution. The plot should include the data with error bars, and the fit. It should have the correct specified binning, etc.
   (c) The five numbers, reported to 1 decimal place:
      i. the fitted mass of the Z^0, in GeV
      ii. the fitted uncertainty on the Z^0 mass, in GeV
      iii. the chi-square
      iv. the number of degrees of freedom
      v. the p-value
   (d) A quantitative and insightful discussion of the results in light of the p-value, etc.
5. (10 pts) 2D Parameter Scan
   (a) This section should include a description of what is plotted and how it was calculated.
   (b) Contains Figure 2: the plot of the 2D chi-square parameter scan, with appropriate binning, etc. The plot should include the 1σ and 3σ contours, written on the plot.
   (c) The text should explain how the $\Delta\chi^2$ for each level was chosen.
6. (5 pts) Conclusions
   (a) Does the report summarize the findings.
   (b) Does the report offer meaningful discussion of assumptions and simplifications used?
   (c) Does the report offer meaningful discussion of next steps to enhance realism?
   (d) Does the report compare the fitted mass to the literature value from the PDG.