Lab 5

Written by: Lexie Peterson

Lab Partner: Kun Lee

The LHC, or the Large Hadron Collider, is a particle accelerator that beams two high-energy particles traveling near the speed of light. These two particles, traveling in the opposite direction, are guided by superconducting electromagnets until they collide at one of the detectors in the LHC. ATLAS, A Toroidal LHC ApparatuS, is a detector that has wide range of purposes that includes, most importantly for this lab, the search for Higgs boson. This particle is associated with Higgs field, a theoretical field that gives mass to interacting particles as the field spontaneously grew after the big bang.

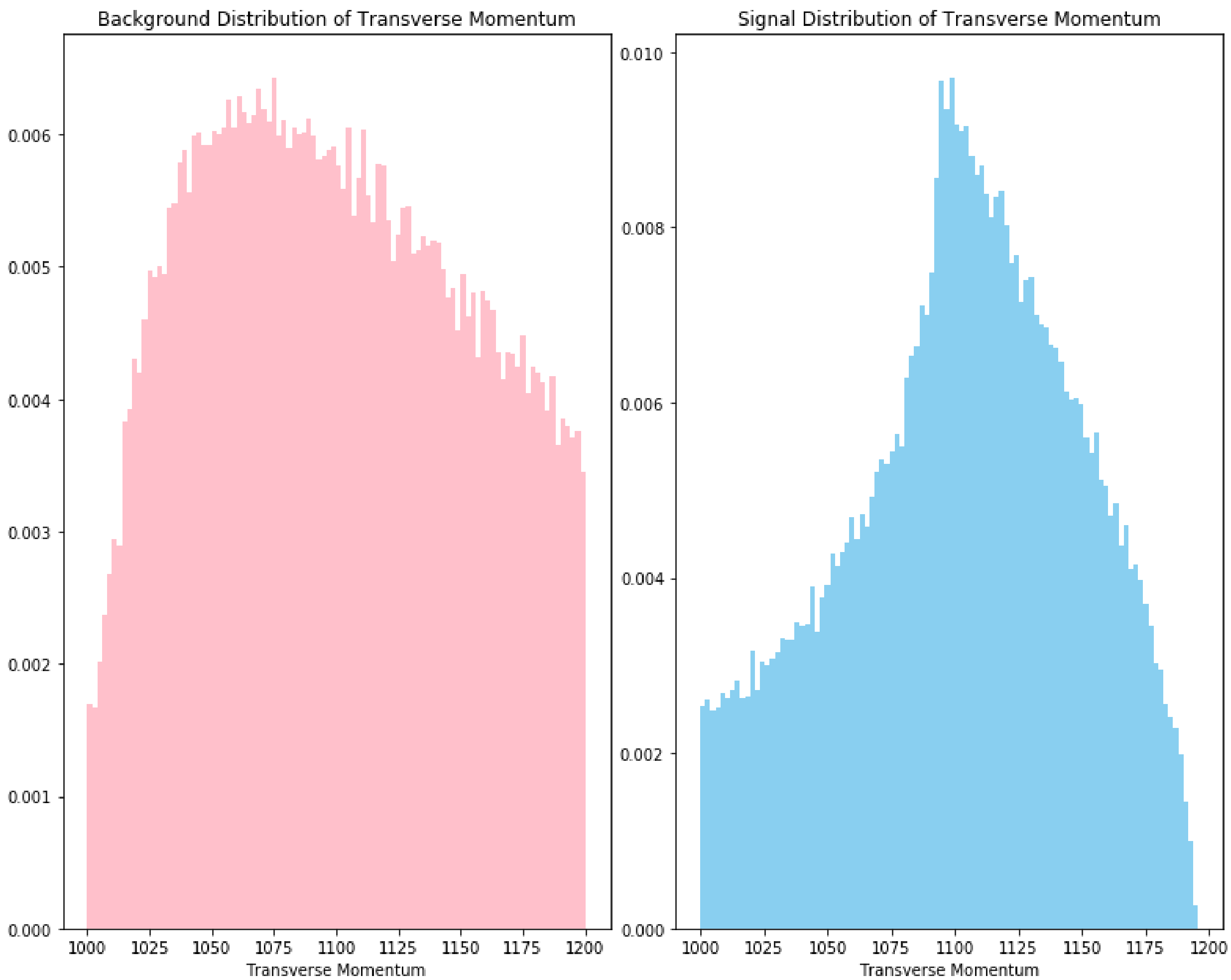
Over the next few weeks I will be studying a data set derived from the ATLAS detector. This data looks at the results of proton-proton collisions at a $\sqrt{s} = 13$ TeV. The purpose of these collisions is to help create a standard model Higgs boson. The Higgs bosons are produced when the colliding protons have large transverse momentum, and they decay into a bottom quark-antiquark pair. There are two sets of data; one is the background caused by QCD (quantum chromodynamics) contamination caused by strong interactions between quarks and gluons, the other is the is the Higgs training signal data set. There has not been a significant discovery of a Higgs yet, so this training set is necessary for determining discovery sensitivities. Each data set has 14 variables to consider, which some important ones are listed in the table below.

Variable	Symbol	Meaning	Equation	Comment
pt	p_t	Transverse Momentum	$\sqrt{(p_x^2 + p_y^2)}$	This momentum is perpendicular to beam line of the particles.
eta		Pseudorapidity	$-\ln\frac{\theta}{2}$	Angle of the particle relative to the beam axis
phi		Azimuthal Angle	$\cos^{-1}\frac{x}{r}$	Angle above the z-axis
mass	m	Invariant Mass	$E^2 = p^2 + m^2$	Portion of the total mass of an object that is independent of the overall motion of the system.
ee2	e_2	2-Point ECF Ratio		Energy correlation functions
ee3	e_3	3-Point ECF Ratio		Energy correlation functions
d2	D_2	3-to-2-Point ECF Ratio		Energy correlation functions

The rest of the variables not included in the table above are t_1, t_2, t_3, t_{21}, t_{32}, and ktDetlaR. The series of t's are variables describing the jets of the resulting from the collision. Additionally, some of the above variables refer to x, y, and z axes. These three axes refer to the ATLAS set up; z is the direction of the beam, while x and y are on the plane of the circular detector.

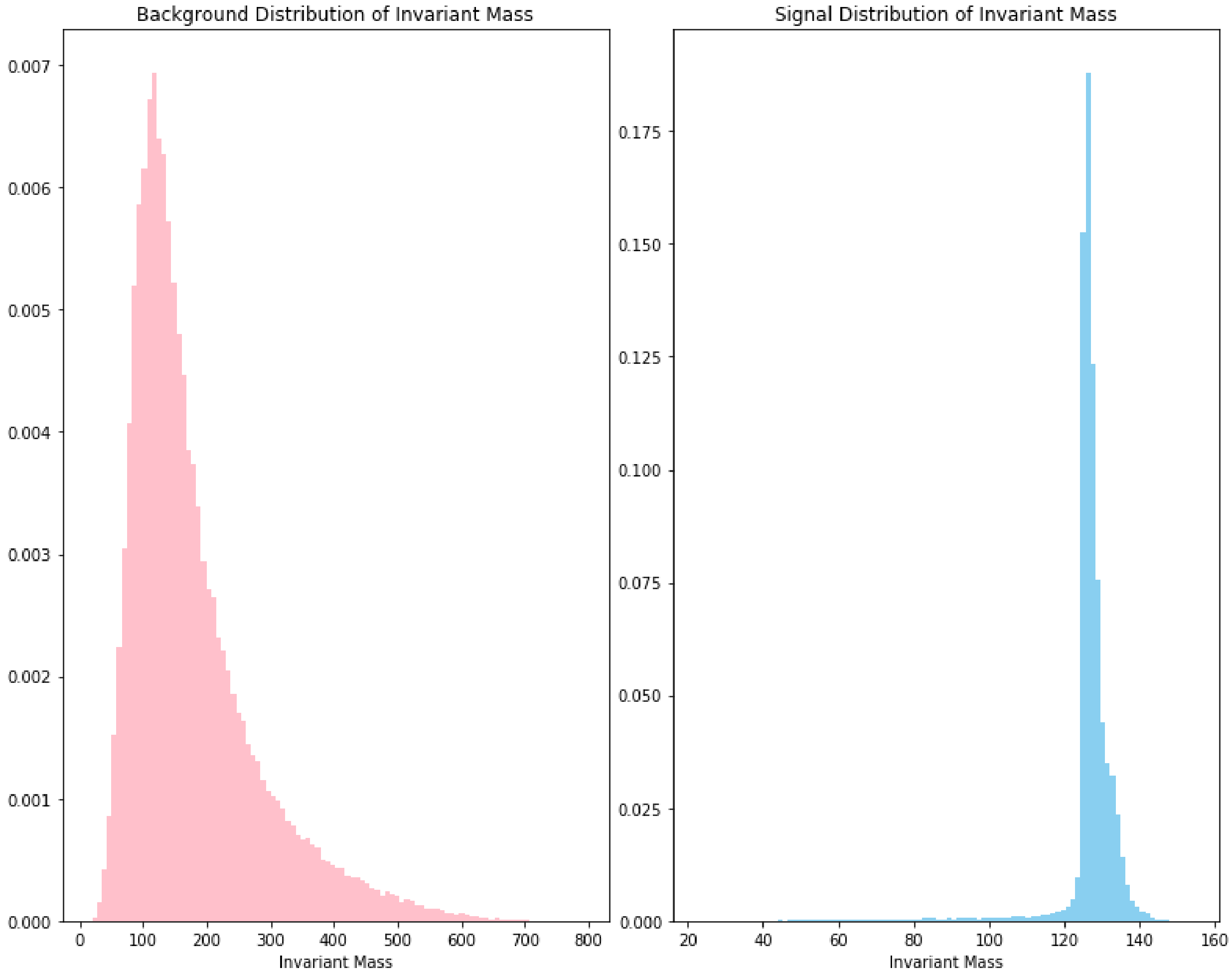
I have begun a rudimentary exploration of the two data sets. This means I have just plotted some basic histograms to make distributions of the various variables and find any points of interest. Going into detail of every variable would make for a tedious and long paper, so I will only go through plots I found most interesting.

Starting with the first variable, transverse momentum, I made a histogram of both the QCD data and the Higgs data. This can be seen in the image below. The plot on the left is that of



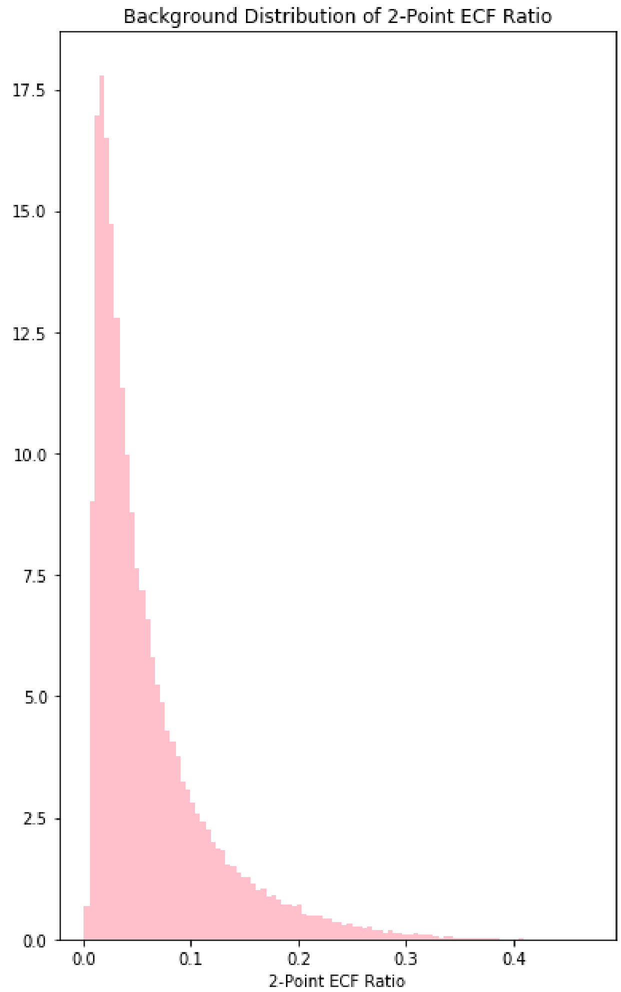
the QCD data, hence the "background distribution," while the plot on the right is that of the Higgs data, hence the "signal distribution." All images from here on will be formatted this way, the background QCD on the left, while the Higgs signal distribution on the right. The first thing I noticed about these plots is the range of the transverse momentum. The data only has transverse momentum of 1000 to 1200, which makes sense as this is the range defined by my training data set. This is important to note though, as I would not want anyone thinking it was due to some physical factor. The next, and probably most glaring, feature is the shape of the right distribution. It looks like a shark fin, and I have yet to come across such a distribution. The right distribution, on the other hand, looks like a log-normal distribution with a high sigma parameter. The mean of the Higgs plot seems to be around 1100, while the QCD mean around 1050. This may be important when determine significance later down the road.

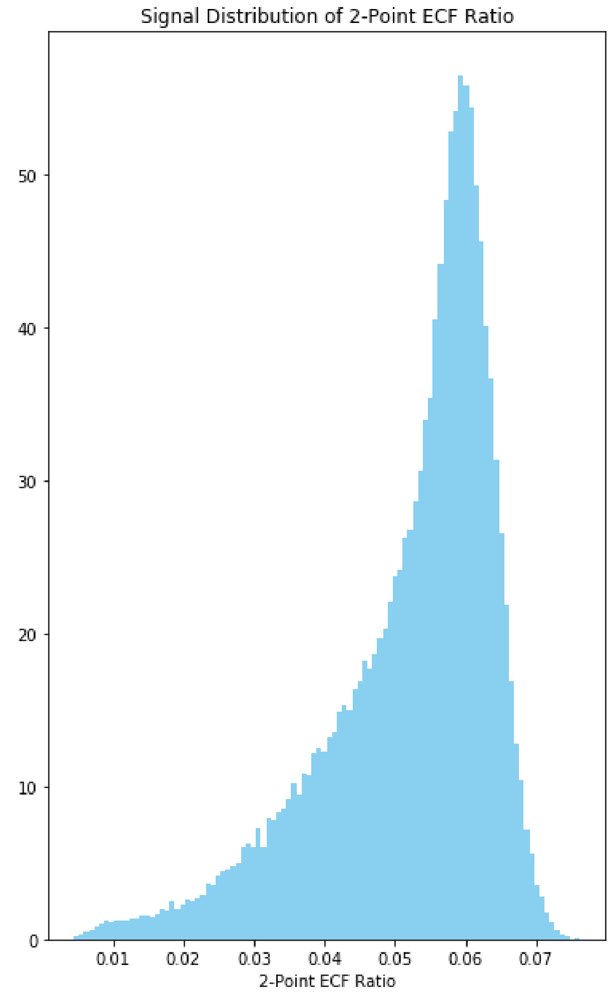
The next important variable is the invariant mass, which, remember, is the mass of an object that is not affected by the motion of the said object. The plot of the QCD and Higgs invariant mass distributions is below. The QCD distibution appears to be a Reyleigh distribution



in shape. I could say the same of the Higgs distribution. This distribution also looks like a Rayleigh distribution, just much skinner. Unfortunately, the mean of both distributions look to be around the same location at about 130 GeV (I am assuming these are the units, I have yet to find written confirmation of this). This similarity in mean location might make it hard for us to find our signal, and find it with significance using just this variable, as it blends in with the background distribution. This leads me to believe I will need to make a combination of variables in the future to get a significant signal. However, the frequencies of the two peaks are different (QCD is around 0.007 while Higgs is around 0.18), so that might be a helpful distinguishing factor between our distribution and signal. I am unsure how it could be helpful, but I think it is important to keep it in the back of my mind for now in case of sudden epiphanies.

One last variable I wish to touch upon is the 2-Point ECF ratio. I find this ECF ratio interesting due to the signifcant differences in the distributions of the QCD and the Higgs.





Both distributions have a similar mean around 0.6 (a similar tend to the other ECF ratios), but their traillling tails are flip (unlike the other ECF ratio distributions). The QCD distribution trailing tail is on the right while the Higgs is on the left. I am unsure of the importance of this, but due to it being different to the tend I saw with the other ECF ratios, I thought I should make note of it.