Lab 5

Ondrea Robinson

November 10, 2019

In this lab we'll be looking at data from the LHC (Large Hadron Collider) in search of the Higgs boson. In this case, the data will be processed rather than raw. The way the data is gathered is by accelerating two proton beams into one another with energy 13TeV. The resulting collisions result in a string of smaller particles that then decay. Many of these particles decay very quickly into more subparticles before they reach the detectors. The detectors read the leftovers of the original collision and from these 'jets' are able to piece together the original results of the collision. Specifically we're looking for collisions that result in a bottom-anti-bottom pair of quarks, indicating the presence of the Higgs boson.

1

First let's download the data. There are two sets, one for the recorded 'signal' data and one for the background. They're labelled so that '$qcd_100000_pt_1000_1200.pkl$' is the background and the other is '$higgs_100000_pt_1000_1200.pkl$' the signal. We can use the Python pickle library to open these files using the code:

```
import pickle

# open the file of interest, and use pickle loading
infile = open ("higgs_100000_pt_1000_1200.pkl",'rb')
backgfile = open("qcd_100000_pt_1000_1200.pkl", 'rb')
signal_dict = pickle.load(infile)
backg_dict = pickle.load(backgfile)

# list all keys of the files
print(signal_dict.keys())
print(backg_dict.keys())
```

Both dictionaries have 14 indexes of data. These indexes appear like;

```
Index(['pt', 'eta', 'phi', 'mass', 'ee2', 'ee3', 'd2', 'angularity', 't1',
       't2', 't3', 't21', 't32', 'KtDeltaR'], dtype='object')
```

Each of these 14 data types tells us something about each jet/trial. Here's a table of the different data types and their meanings:

Variable	Symbol	Meaning
pt	p_T	Transverse (xy plane) momentum
eta	η	Psuedorapidity
phi	ϕ	Azimuthal angle
mass	m	Invariant mass
ee2	e_2	2-point ECF ratio
ee3	e_3	3-point ECF ratio
d2	D_2	3 ro 2 point ECF ratio
angularity	τ_{-2}	
t1		
t2		
t3		
t21		
t32		
ktDeltaR	$k\Delta R$	

Note, ECF stands for Energy Correlation Function. We can make histograms for all of these variables for each data set to see how they compare for the background and the signal. We'll use the code:

```
keys = signal_dict.keys()
titles = ['Transverse Momentum', 'Psuedorapidity', 'Azimuthal Angle',
'Invariant Mass', '2-point ECF Ratio', '3-Point ECF Ratio',
'3 to 2 point ECF Ratio', 'Angularity','t1','t2','t3','t21', 't32', 'KtDeltaR']
j=0
k=1
for i in keys:
    backg = backg_dict[i]
    signal = signal_dict[i]
    figure(k)
    plt.title(titles[j]+' (signal)')
    plt.xlabel(titles[j])
    plt.ylabel('Occurences')
    plt.hist(backg, bins=300)
    k=k+1
    figure(k)
    plt.title(titles[j]+' (background)')
    plt.xlabel(titles[j])
    plt.ylabel('Occurences')
    plt.hist(backg, bins=300)
    plt.show()
    j=j+1
    k=k+1
```

From this we have 28 figures; 14 from our signal data set and 14 from the background data set. Comapring side by side we can look at the distributions for each variable for each data set:

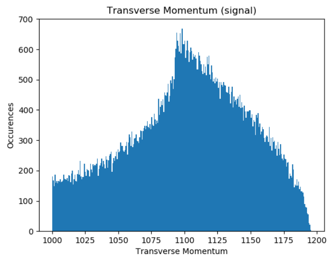

Figure 1

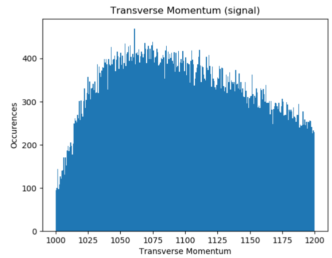

Figure 2

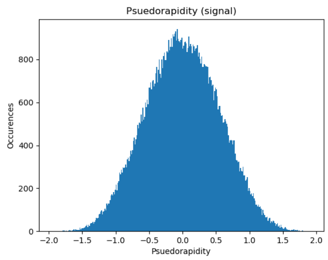

Figure 3

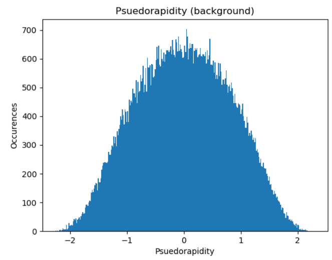

Figure 4

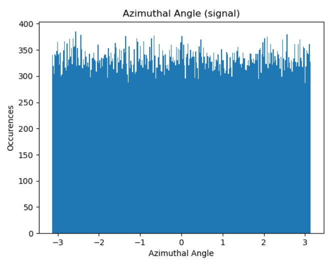

Figure 5

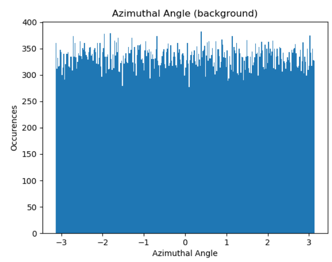

Figure 6

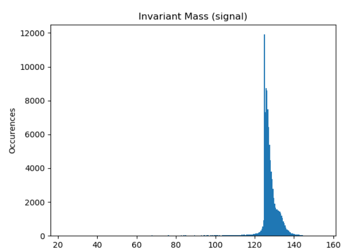

Figure 7

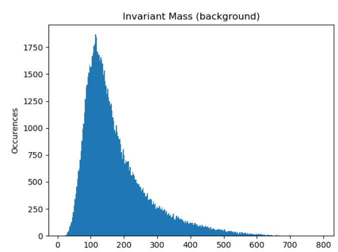

Figure 8

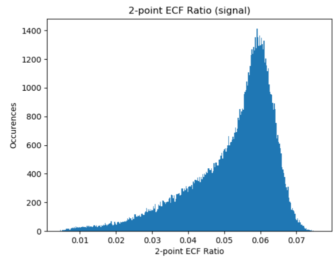

Figure 9

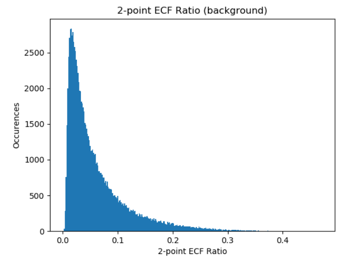

Figure 10

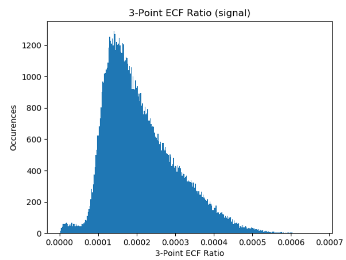

Figure 11

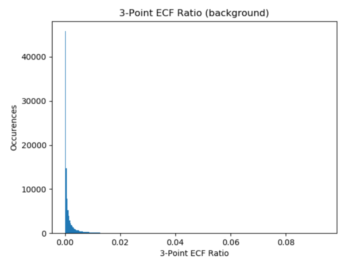

Figure 12

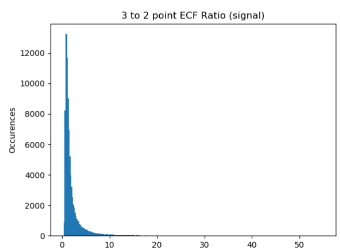

Figure 13

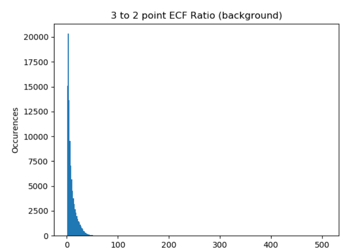

Figure 14

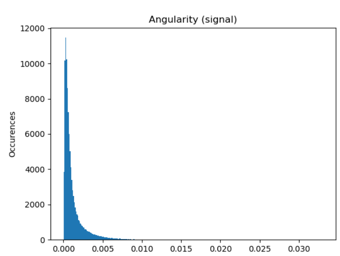

Figure 15

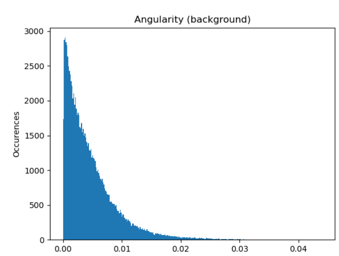

Figure 16

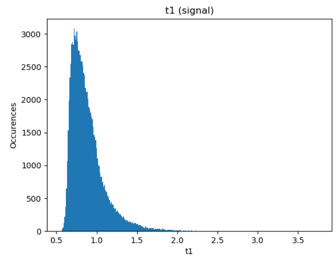

Figure 17

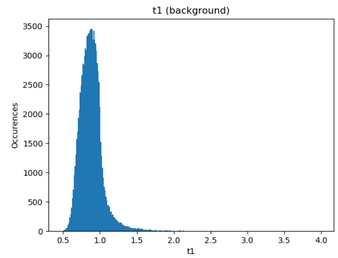

Figure 18

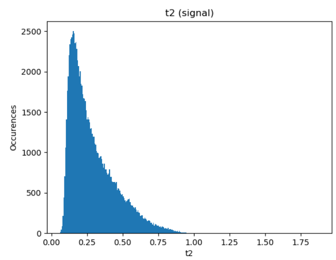

Figure 19

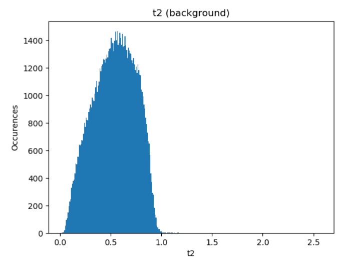

Figure 20

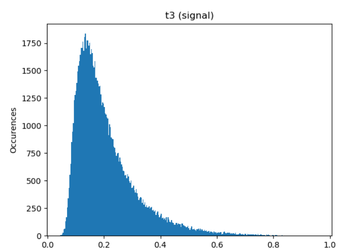

Figure 21

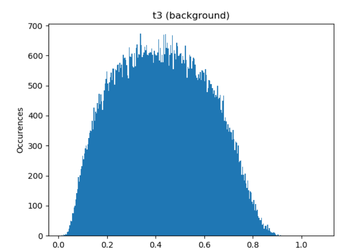

Figure 22

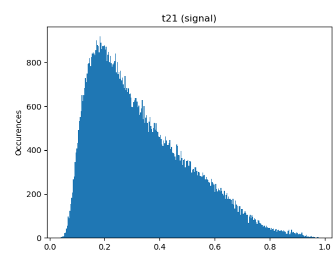

Figure 23

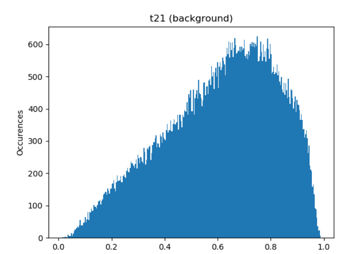

Figure 24

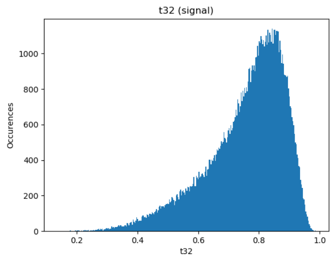

Figure 25

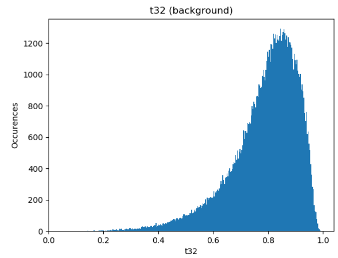

Figure 26

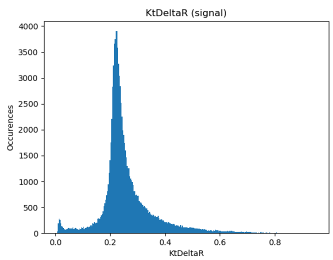

Figure 27

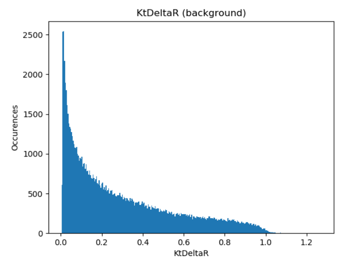

Figure 28

One thing we're able to see from this data is that the ee2 signature for the signal is completely buried in the background and so it would be better to use the ee3.