# A Statistical Analysis of the Higgs Boson

Abstract: The writing begins with basic statistics, null hypothesis, signal-free datasets and move on to more advanced statistical analysis. The final report scrutinizes the standard model H in pp collisions at $\sqrt{s} = 13$ TeV using pseudo-experiment training datasets to optimize event selection.

Shouyang Michael Wang

University of Washington

Physics 434 Capstone Project

Advisor: Dr. Miguel Morales

December 2019

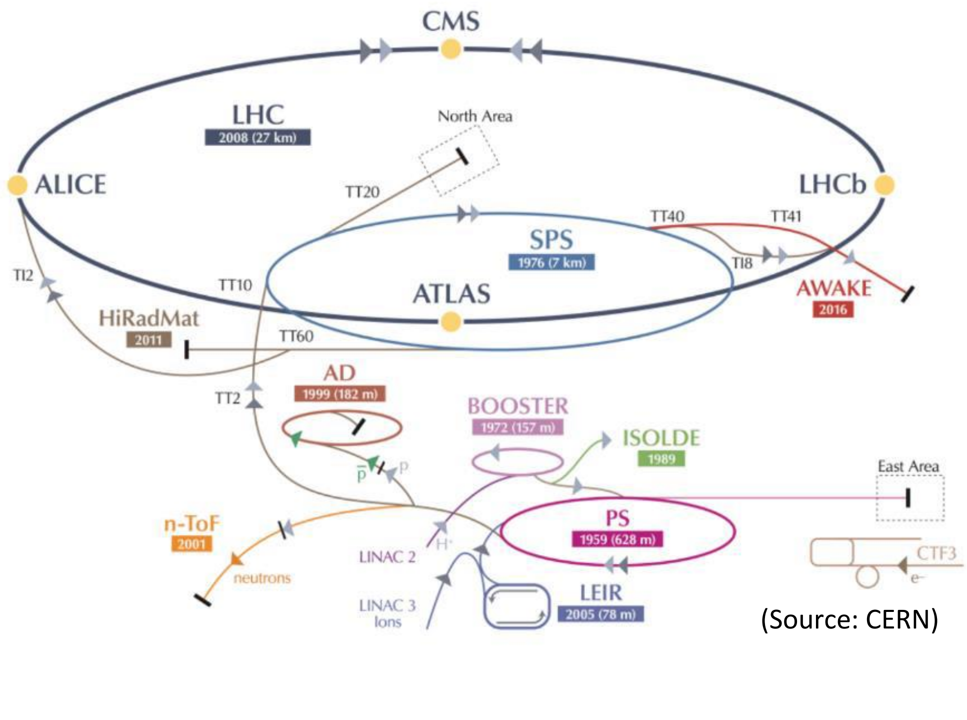
# Contents

# Background

**The Standard Model**
The Standard Model of particle physics describes all known fundamental particles that are the building blocks to everything in the universe.

The Higgs mechanism, a mechanism theorized to give rise to the masses of all the subatomic elements, is essential for the Standard Model to work mathematically - all particles that mediate forces are massless. The mechanism successfully explained how particles obtain their mass through the Higgs field and theorized the Higgs boson's existence.
Higgs boson (H) is one of the elementary particles in the Standard Model. The particle is extremely unstable, meaning that it would decay into other particles almost immediately. Higgs boson has been verified experimentally using the Large Hadron Collider at the European Organization for Nuclear Research(CERN).

**Large Hadron Collider**
The Large Hadron Collider is a circular high-energy particle accelerator and collider. It consists of a ring of "superconducting magnets," along with several accelerating structures. Inside the ring, two beams containing protons or lead ions would travel in opposite directions to collide at speed close to the speed of light. Particle detectors are placed around the collision site to record and analyze the subatomic particles that arise from the collision. The LHC is engineered to operate at a maximum collision energy of 14 TeV, meaning that each traveling beam would carry a power of 7 TeV. However, in most operations, the collision energy is controlled under 14 TeV to "optimize the delivery of particle collisions" (CERN).
Furthermore, one of the particle detectors is named ATLAS(A Toroidal LHC ApparatuS). It is used to detect a wide range of subatomic particles, including the Higgs boson.



(Source: CERN)

**Collision Energy**
In the context of proton-proton(pp) collisions, $\sqrt{s}$ indicates the total center-of-mass energy of the colliding beams.
In the laboratory frame, the center-of-mass energy is calculated as the following:

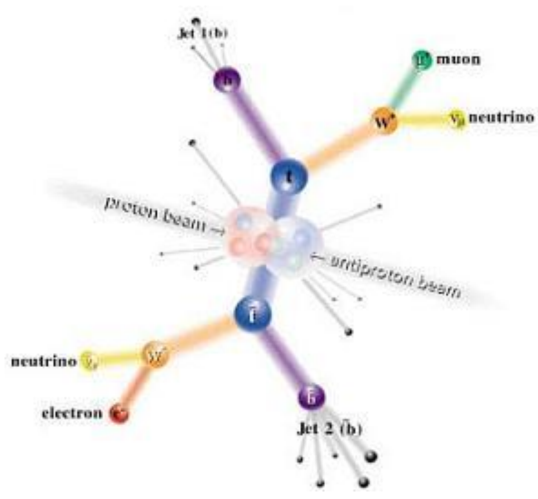
$$E_{\text{cm}} = \sqrt{s} = \sqrt{\left(\sum E_i\right)^2 - \left(\sum \vec{p}_i\right)^2}$$

We will observe and analyze an inclusive search for the standard model Higgs boson in typical pp collisions at $\sqrt{s} = 13$ TeV at the LHC using simulated data. In other words, the energy for one beam of protons is 6.5 TeV, which would result in the total energy of $\sqrt{s}$ =13 TeV when two beams collide.

**Jets**
The term refers to a large number of detectable particles flying roughly in the same direction. The process is called clustering.

**The decay of Higgs boson**
The Higgs boson is extremely unstable, and it would decay into other particles almost immediately. The standard model of particle physics theorized that the Higgs could decay into other particles like quarks. There are many "flavors" when it comes to quarks. The net electric charge of the byproducts of the decay has to be zero since the Higgs boson has a charge of zero. Quarks have electric charges, meaning that the quarks created from the decay have to emerge in pairs, such that the charge from one quark and one antiquark would cancel each other out, resulting in a net charge of zero. Very often, physicists find that the Higgs boson would decay into a pair of bottom quarks.

**Quantum Chromodynamics(QCD)**
During a proton-proton(pp) collision, almost all detected events are due to the strong interaction(QCD). Due to the Higgs boson's unstable nature, the pp collision would lead to a pair of "particle jets originating from b quarks' fragmentation. (ATLAS)" It is incredibly challenging to distinguish an actual Higgs signal from the" unwanted" b-quark pairs noise due to the QCD background.



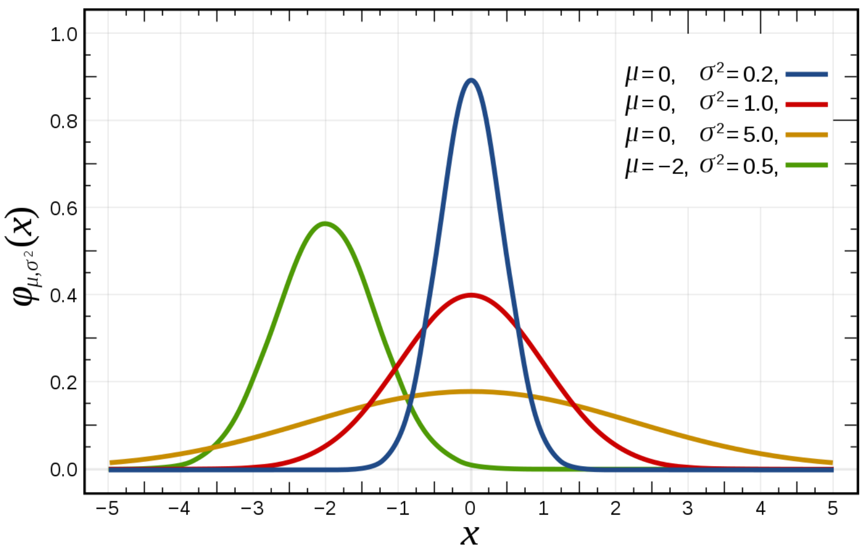

(Source: Particle Physics Theory, University of Glasgow)

# Statistics

## Normal Distribution

In physics, σ refers to a probability. The sigma implicitly refers to the standard normal distribution (a Gaussian with mean zero and a standard deviation of 1).

Integrals of the standard normal distribution give probabilities. The normal distribution is a symmetric distribution where most of the observations cluster around the center to form a bell shape. A Gaussian would remain a Gaussian after summation, multiplication, or Fourier transform.




The z-score is positive if the value lies above the mean and negative if it lies below the mean. Notice that it is possible to get a "negative value" when integrating from a negative number to another number.

For instance, if we try to integrate a function from negative infinity to a number, it's possible to get a negative area. It does not make any sense to have a negative probability. The probability from -10 to 0 should have the same probability from 0 to 10 for a Gaussian distribution since the distribution is symmetrical.

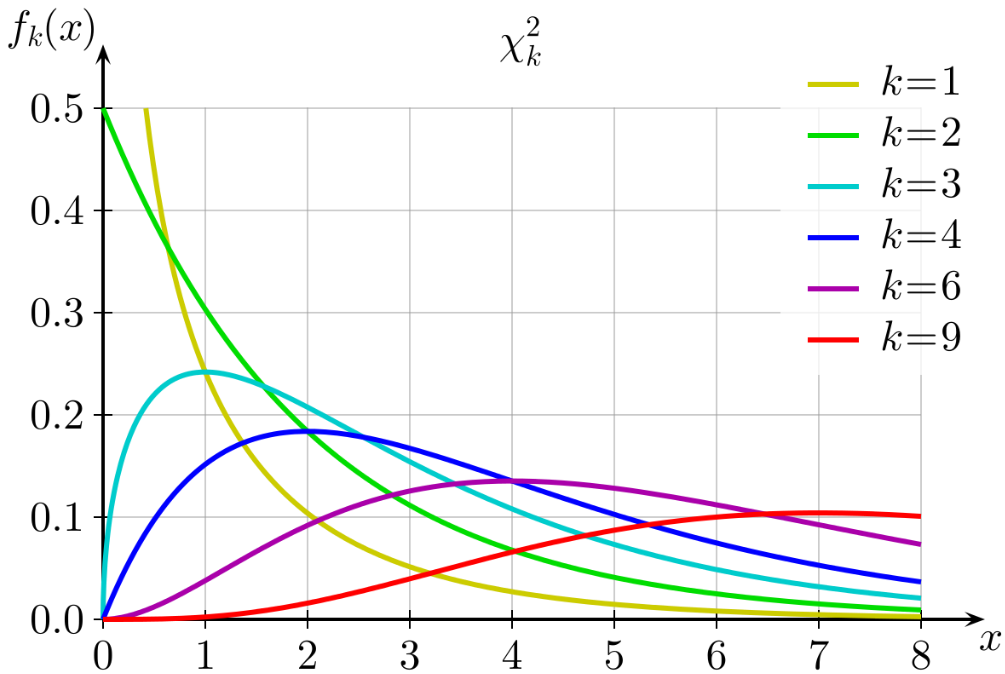
To illustrate this, we try to integrate the equation for the normal distribution using two different sets of limits.

Notice that the two equations generate the same probability. In other words, the area(probability) under the curve from -10 to 0 is the same as the area from 0 to 10.

$$\int_0^{10} \frac{\frac{e^{-x^2}}{2}}{\sqrt{2\pi}}dx = \frac{\mathrm{erf}(10)-\mathrm{erf}(0)}{4\sqrt{2}} \qquad \int_{-10}^{0} \frac{\frac{e^{-x^2}}{2}}{\sqrt{2\pi}}dx = \frac{\mathrm{erf}(0)-\mathrm{erf}(-10)}{4\sqrt{2}}$$

## χ2-distribution

Now we explore some other continuous analytic distributions. The χ2-distribution (or Chi-squared) with k degrees of freedom is the distribution of a sum of squares of k independent standard normal random variables. Now we create a χ2-distribution object with the parameter value b = 2, a special case of Gamma distribution.

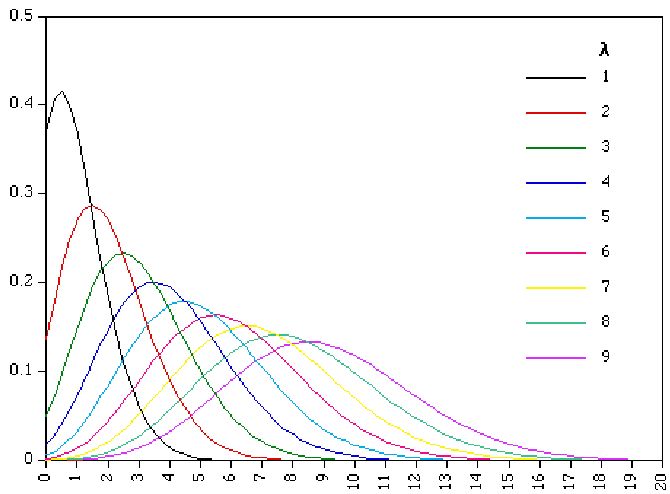



Assuming that we have signal-free data that follows the χ2-distribution, and we have a measurement for which we need to determine the σ. The test statistic is measured to be 7.1. We propose a statistical question: What is the probability that the background produces a signal that is less than the measurement we made? More specifically, given the test statistics and if we know the distribution has 3 degrees of freedom, what is the area to the left of the test statistic? First, we try to approach the question mathematically. The formula for the cumulative distribution function of the chi-square distribution is given by

$$F(x) = \frac{\gamma(\frac{k}{2}, \frac{x}{2})}{\Gamma(\frac{k}{2})} \quad \text{for } x \geq 0 \qquad \Gamma(a) = \int_0^{\infty} t^{a-1}e^{-t}dt \qquad \Gamma_x(a) = \int_0^{x} t^{a-1}e^{-t}dt$$

## Poisson distribution

Now we move on to non-continuous distributions. We will be looking at the Poisson distribution.




Notice that
1) all of the variables are discrete.
2) The graphs have a lower bond at zero(For instance, the data can't go lower than 0), but they don't have an upper limit.
3)The graph with a low mean(lambda) is highly skewed. 4) the distribution becomes more bell-like as the mean is getting large.
5) The shape isn't very symmetric like the normal distribution.
6) The mean, lambda, is a constant number throughout the experiment.

We propose a statistical question: Given that lambda = 3, what is the probability of the background producing a precisely 4? It can be calculated as the following:

```
temp = poisspdf(4,lambda)
disp('[P(x=4) = ', num2str(temp,2)])
```

Because the distributions are discrete, so are the probabilities and σ. A discrete number has countable values. Since the distribution is discrete, it makes sense to talk about probability at an exact value like P(x=4) shown above. The discrete nature is more pragmatic when it comes to experiments that expect a specified number of events to take place.

While the results are discrete, the parameters of the distributions are not. For example, the mean of a Poisson distribution can be 9.2. The mean is essentially the average of a set of numbers. It is an arithmetic number obtained from experimental data. It does not have to be a whole number. However, since the distribution is typically used to reflect the probability of a number of events occurring in a fixed interval, it does not make much sense to say that an event happened 1.5 times. An event can happen on average 1.5 times. Typically the x-axis represents a given interval of time(e.g., how many times an event happens at hours 1, 2,3,4, etc.). Just like it's either head or tail, it can't be somewhere in between.

## Central Limit Theorem

Furthermore, the sum of a Poisson with itself still has Poisson distribution. To prove this, we will use the moment generating function of the Poisson distribution. It is given by

$$M_X(t) = e^{\lambda_1(e^t-1)}$$

For a linear combination,

$$M_{X+X}(t) = M_X(t)M_X(t) = e^{\lambda_1(e^t-1)}e^{\lambda_1(e^t-1)} = e^{(2\lambda_1)(e^t-1)}$$

Hence,

$$X + X \sim \text{ Poisson } (2\lambda_1)$$

LaTex code:

```
M_{X+X}(t)=M_{X}(t) M_{X}(t)=e^{\lambda_{1}\left(e^{t}-1\right)} e^{\lambda_{1}\left(e^{t}-1\right)}=e^{\left(2 \lambda_{1}\right)\left(e^{t}-1\right)}
```

From a conceptual point of view, the Poisson distribution is the number of counts in a fixed interval of time. When we conv() or sum a Poisson distribution with itself N(in this case, five) times, it is the same as saying that we multiply the interval by N, or N times longer than the original interval. Consequently, the mean is also multiplied by N. The two methods (sum N times or multiply interval and mean by N)would get the same result. A Poisson distribution with a rescaled interval and mean is still a Poisson distribution.

A limiting form of the Poisson distribution is the Gaussian distribution. **The central limit theorem** states that when independent random variables are summed or averaged enough trials, their normalized sum or average tends toward a normal distribution. Poisson is becoming more symmetrical and bell-like as it is averaging over time. Because of the central limit theorem, Poisson tends to normal as its mean becomes sufficiently large.

# Multidimensional Gaussian

**Notes**

*I include the notes here for completeness and future reference.

Multidimensional Gaussian.

The 1-D Gaussian can be extended to a multi-dimensional Gaussian. In other words, the probability density function(pdf) of a 1D Gaussian is a special case of the pdf for multivariate Gaussian. The multivariate normal distribution has two parameters, a mean μ (in vector form , in

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^T \qquad \boldsymbol{\Sigma} = (\sigma_{ij})$$

1D, μ = constant) and a covariance matrix Σ (analogous to variance in 1D). The diagonal elements of Σ contain the variances for each variable.

The pdf of the d-dimensional Gaussian is given by

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right).$$

For a standard 1-D normal distribution(mean = 0, std = 1), the pdf has the form

$$p(x) = \frac{1}{\sqrt{2\pi}} exp(-\frac{x^2}{2})$$

For a standard 2-D (bivariate) normal distribution(correlation = 0), the pdf is

$$p(x,y) = p(x)p(y) = \frac{1}{\sqrt{2\pi}} exp(-\frac{x^2+y^2}{2})$$

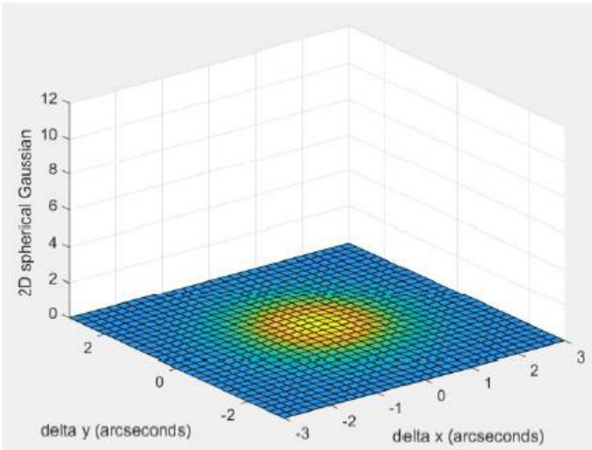
For a 2D Gaussian, assuming that the mean is zero ($\mu = [0,0]^T$), and the covariance matrix is

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

This is referred to as a spherical Gaussian because of the circular symmetry on a 2-D plot. Note that the variances are equal to 1. The determinant of a matrix is the measurement of its size.

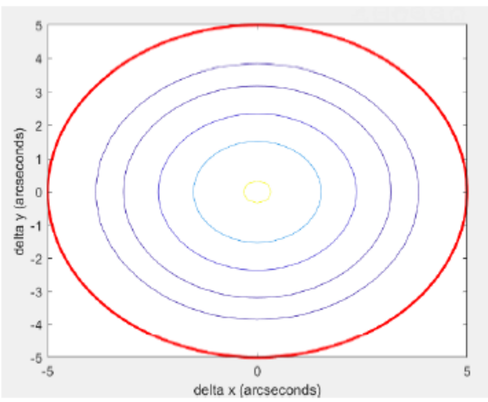
If we assume that the background is a spherical Gaussian distribution due to its circular symmetry, meaning that the distribution has a zero mean and an identity covariance matrix(analogous to variance = std = 1 in 1D).

Just for the sake of getting an actual numerical value, I created a 2D Gaussian plot with perfect data points to mimic the background (mean = 0, covariance matrix = [1 0; 0 1]).




```
%returns the cdf of the multivariate normal distribution
%with mean mu and covariance sigma
%mean
mu = [0,0];
%Create a 4-by-4 identity matrix.
%This would spit our covariance matrix
sigma = eye(2);
%grid
x1 = -3:0.2:3;
x2 = -3:0.2:3;
[X1,X2] = meshgrid(x1,x2);
X = [X1(:) X2(:)];
%evaluate pdf
y = mvnpdf(X,mu,sigma);
y = reshape(y,length(x2),length(x1));
%plot
surf(x1,x2,y)
```

```
caxis([min(y(:))-0.5*range(y(:)),max(y(:))])
axis([-3 3 -3 3 0 12])
xlabel('delta x (arcseconds)')
ylabel('delta y (arcseconds)')
zlabel('2D spherical Gaussian ')
```




```
mu = [0 0];
sigma = eye(2);
x1 = -4:.2:4;
x2 = -4:.2:4;
[X1,X2] = meshgrid(x1,x2);
X = [X1(:) X2(:)];
y = mvnpdf(X,mu,sigma);
y = reshape(y,length(x2),length(x1));
contour(x1,x2,y,[0.0001 0.001 0.01 0.05 0.15 0.25 0.35])
xlabel('delta x (arcseconds)')
ylabel('delta y (arcseconds)')
circle = viscircles([0,0],5);
```

# Higgs Bosons & QCD Parameters

## Parameteres

| Number | Variables | Symbol | Equation | Notes |
|---|---|---|---|---|
| **1** | pt | $P_T$ | $\sqrt{p_x^2 + p_y^2}$ | Transverse Momentum |
| **2** | eta | η | $-\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$ | Pseudorapidity |
| **3** | phi | ϕ | $cos^{-1}\left(\frac{x}{r}\right)$ | Azimuthal angle |
| **4** | mass | m | $E^2 = p^2 + m^2$ | Invariant Mass |
| **5** | ee2 | $e_2$ | $\sum_{i<j\in J} P_{T,i}P_{T,j}\Delta R_{ij}\frac{1}{p_{T,J}^2}$ | 2-Point ECF ratio |
| **6** | ee3 | $e_3$ | $\sum_{i<j<k\in J} P_{T,i}, P_{T,j}, P_{T,k}\Delta R_{ij}\Delta R_{ik}\Delta R_{jk} \cdot \frac{1}{P_{T,J}^3}$ | 3-Point ECF ratio |
| **7** | d2 | $D_2$ | $D_2 = \frac{e_3}{e_2^3}$ | 3 to 2-point ECF ratio |
| **8** | angularity | $a_3$ | $\frac{1}{m_J}\sum_{i\in J} E_i \cdot \sin^{-2}(\theta_i) \cdot \cos^3(\theta_i)$ | Angularity |

| | | | | |
|---|---|---|---|---|
| **9** | t1 | $\tau_1$ | See page 13 | Subjettiness |
| **10** | t2 | $\tau_2$ | | Subjettiness |
| **11** | t3 | $\tau_3$ | | Subjettiness |
| **12** | t21 | | | $\tau_2 / \tau_1$ |
| **13** | t32 | | | $\tau_3 / \tau_2$ |
| **14** | KtDeltaR | $K_t$ ΔR | $\Delta R = \sqrt{\Delta\eta^2 + \Delta\varphi^2}$ | Cluster Sequence |

# DATA EXPLORATION

The Higgs bosons are produced with large transverse momentum (pT) and decay to a bottom quark-antiquark pair. The Higgs candidates could be reconstructed as large-radius jets using Calorimeters. Due to large QCD background contamination, the direct 5-sigma observation of this Higgs channel is not accomplished yet[Phys. Rev. Lett. 120, 071802 (2018)]. We will analyze the data using MATLAB. See lab5.m for the complete code. The signal dataset is labeled as "Higgs," and the background dataset is labeled as "QCD." First, we import the data. In the data, the transverse momentum ranges from 250 GeV to 500 GeV.

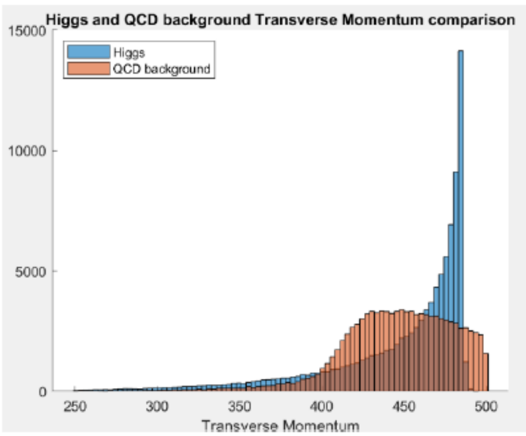
```
%import
h5disp("higgs_100000_pt_250_500.h5");
h5disp("qcd_100000_pt_250_500.h5");
higgs = h5read("higgs_100000_pt_250_500.h5",'/higgs_100000_pt_250_500');
qcd = h5read("qcd_100000_pt_250_500.h5",'/qcd_100000_pt_250_500');
```

```
Command Window
HDF5 higgs_100000_pt_250_500.h5
Group '/'
    Dataset 'higgs_100000_pt_250_500'
        Size:  14x100000
        MaxSize:  14x100000
        Datatype:   H5T_IEEE_F64LE (double)
        ChunkSize:  []
        Filters:  none
        FillValue:  0.000000
HDF5 qcd_100000_pt_250_500.h5
Group '/'
    Dataset 'qcd_100000_pt_250_500'
        Size:  14x100000
        MaxSize:  14x100000
        Datatype:   H5T_IEEE_F64LE (double)
        ChunkSize:  []
        Filters:  none
        FillValue:  0.000000
```

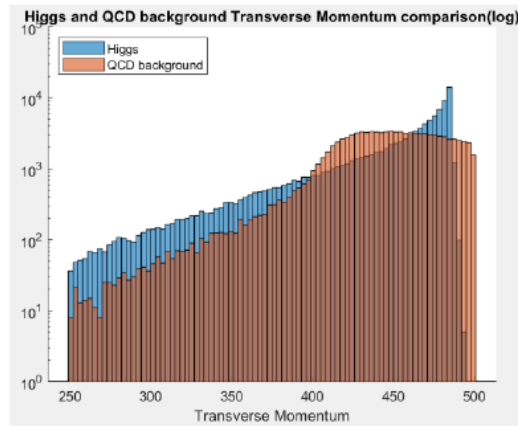
As shown above, there are two datasets. The first dataset is the Higgs signal, and the second dataset is the QCD background. Each dataset contains 14 rows and 100000 columns. 14 rows represent the 14 parameters we mentioned previously. They are 'pt', 'eta', 'phi', 'mass', 'ee2', 'ee3', 'd2', 'angularity', 't1', 't2', 't3', 't21', 't32' , and 'KtDeltaR'. For each parameter, we have recorded 100000 datapoints.

# TRANSVERSE MOMENTUM(PT) (XY PLANE)(UNIT: GEV*C-1 OR GEV)

The component of momentum in the transverse plane (in other words, perpendicular to the beamline.) The parameter is always associated with the events that happened at the vertex, which is a fairly clean sample.




```
hold on
histogram(pt);
histogram(qpt);
legend('Higgs', 'QCD background','location','northwest')
title('Higgs and QCD background Transverse Momentum comparison')
xlabel('Transverse Momentum')
hold off
```

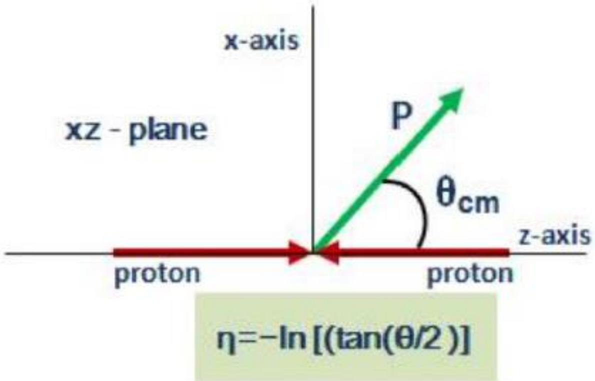


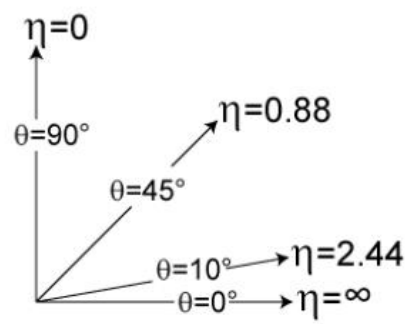

```
hold on
histogram(pt);
set(gca,'YScale','log')
histogram(qpt);
set(gca,'YScale','log')
legend('Higgs', 'QCD background','location','northwest')
title('Higgs and QCD background Transverse Momentum comparison(log)')
xlabel('Transverse Momentum')
hold off
```

Notice that the x-axis goes from 250 to 500. This agrees with the dataset we use where the transverse momentum ranges from 250 GeV to 500 GeV. One peculiar thing to notice is that there is a noticeable gap at the very end of the plot — the Higgs datapoints peak around 485 GeV.

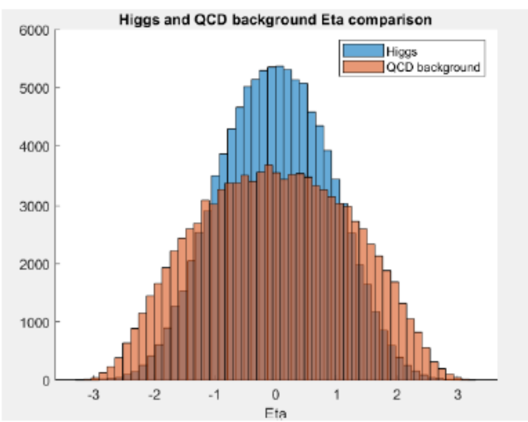
# PSEUDORAPIDITY ETA (η)

Eta describes the angle of a particle relative to the beam axis(z-axis). θ and η are convertible.

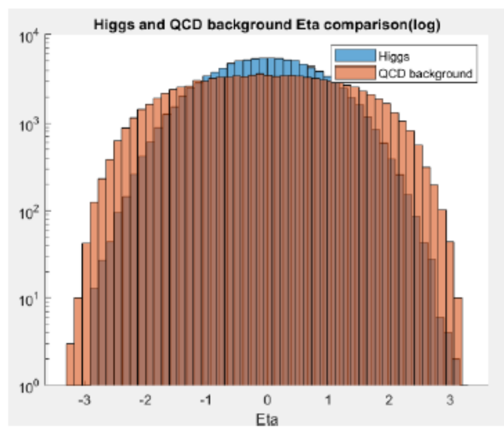



(https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.momentum)
(https://en.wikipedia.org/wiki/Pseudorapidity#/media/File:Pseudorapidity2.png)




```
hold on
histogram(eta,b);
histogram(qeta,b);
legend('Higgs', 'QCD background')
title('Higgs and QCD background Eta comparison')
xlabel('Eta')
hold off
```

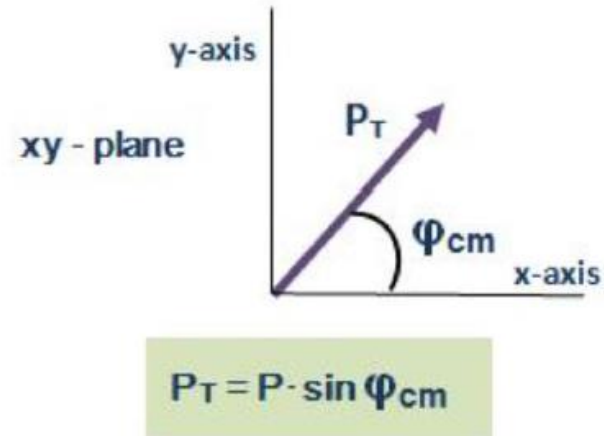



```
hold on
histogram(eta,b);
set(gca,'YScale','log')
histogram(qeta,b);
set(gca,'YScale','log')
legend('Higgs', 'QCD background')
title('Higgs and QCD background Eta comparison(log)')
xlabel('Eta')
hold off
```

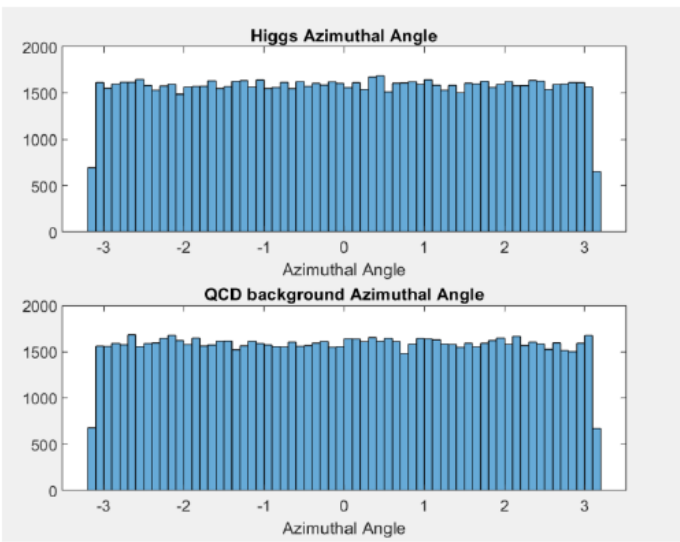
The plot goes from -pi to pi, and it is centered around zero.

## PHI (Φ)

Phi describes the azimuthal angle (angle from the x-axis).




(https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.momentum)




```
subplot(2,1,1)
histogram(phi);
title('Higgs Azimuthal Angle')
xlabel('Azimuthal Angle')
subplot(2,1,2)
histogram(qphi);
title('QCD background Azimuthal Angle')
xlabel('Azimuthal Angle')
hold off
```

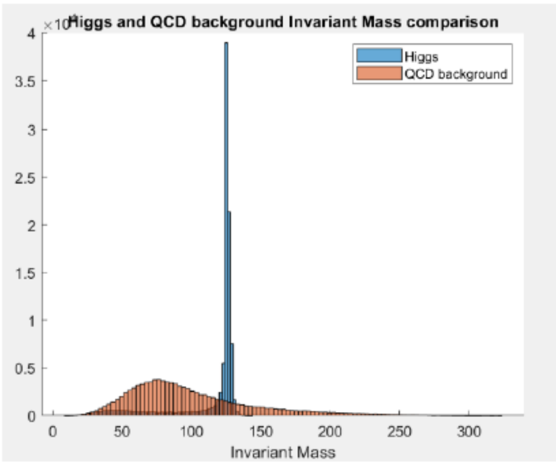
The plot ranges from -pi to pi, and it is centered around zero.

Equation:

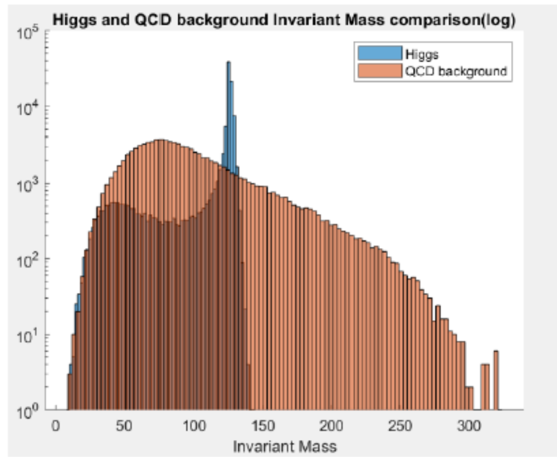
$$\phi = cos^{-1}\left(\frac{x}{r}\right)$$

# INVARIANT MASS (M)(UNIT: GEV*C-2 OR GEV)

When a Higgs boson particle decays into other particles, its mass before the decay can be calculated “from the energies and momenta of the decay products”(ATLAS). The mass is an invariant quantity, which is the same for all observers in all reference frames.
We will plot the Higgs and QCD background mass and compare the two plots.




```
hold on
histogram(mass);
histogram(qmass);
legend('Higgs', 'QCD background')
title('Higgs and QCD background Invariant Mass comparison')
xlabel('Invariant Mass')
hold off
```



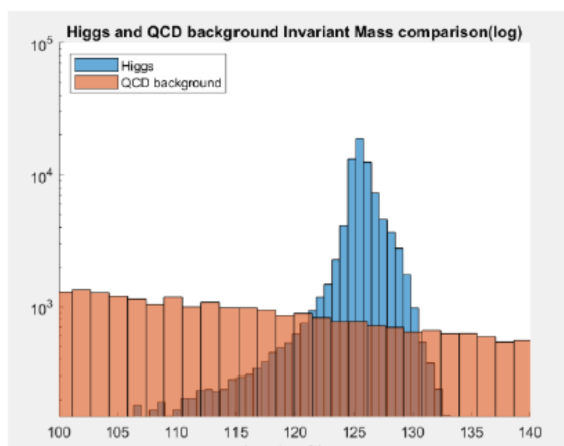

```
hold on
histogram(mass);
set(gca,'YScale','log')
histogram(qmass);
set(gca,'YScale','log')
legend('Higgs', 'QCD background')
title('Higgs and QCD background Invariant Mass comparison(log)')
xlabel('Invariant Mass')
hold off
```

Here is a zoomed-in version.

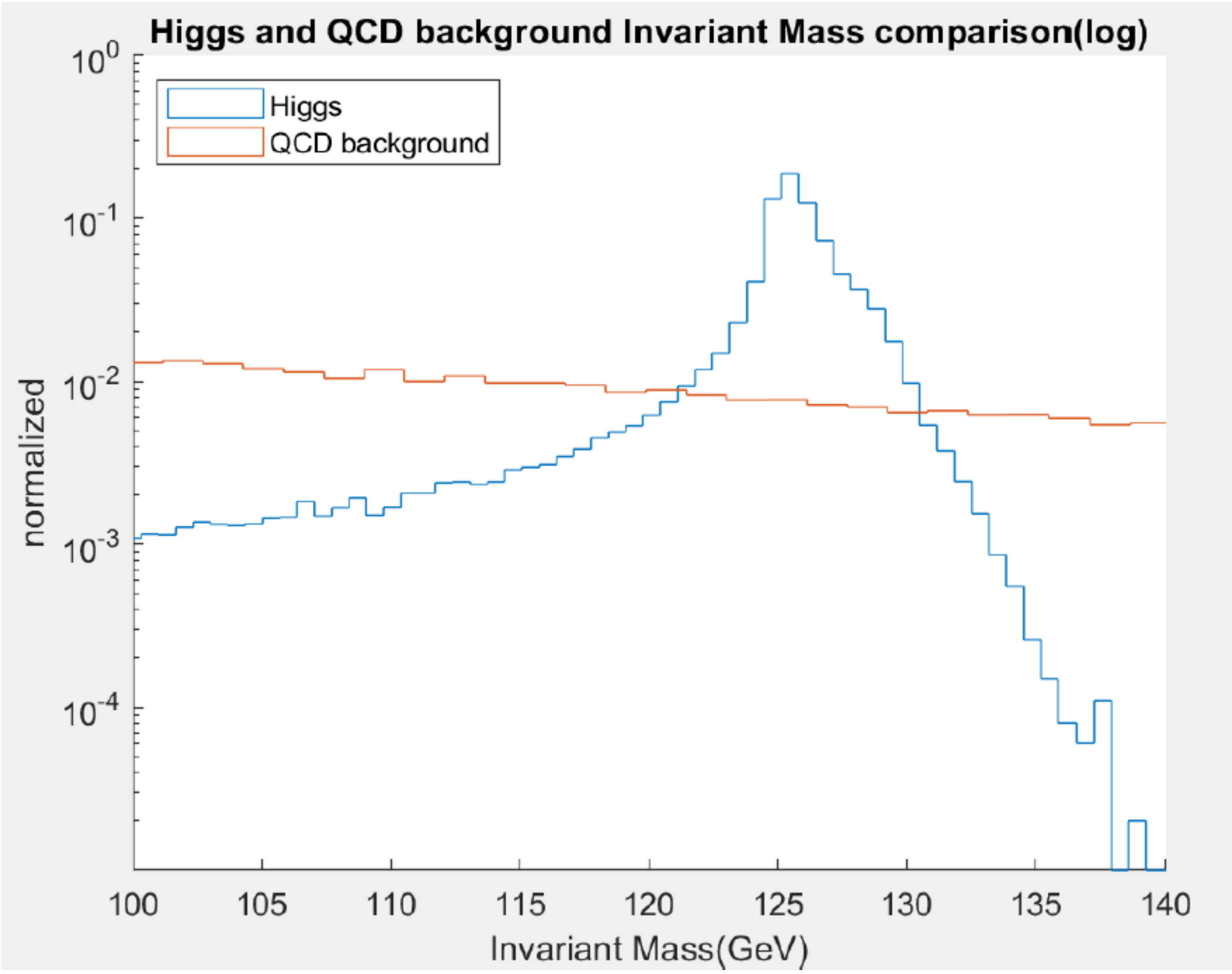
The most noticeable characteristic of the Higgs plot is that it peaks around 125. The invariant mass(m) is calculated using $m^2 = E^2 – p^2$, where E is the energy, and p is the momentum. This is the reduced form of the energy-momentum equation where c = 1. To get the invariant mass, we measure the momenta of the outgoing protons. The peak gives us information about the mass of the emerged Higgs boson. Thus, we can conclude that the mass of the Higgs boson is approximately 125 (GeV). The result agrees with the mass of the Higgs boson particle found in 2012 at CERN.

Equation:

$$M = \sqrt{\left(\sum E_i\right)^2 - \left(\sum \vec{p}_i\right)^2}$$



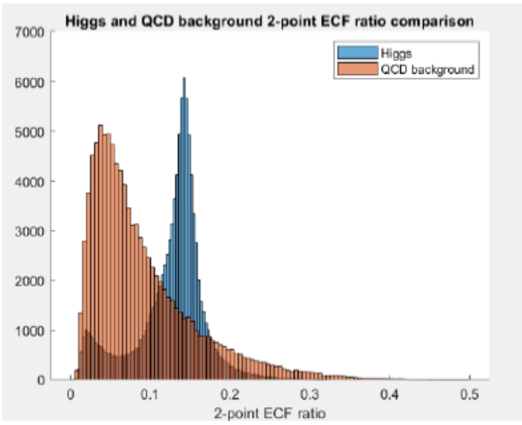

```
hold on
histogram(mass,b2);
set(gca,'YScale','log')
histogram(qmass,b2);
xlim([100,140]);
ylim([150, 1e5]);
set(gca,'YScale','log')
legend('Higgs', 'QCD background','location','northwest')
title('Higgs and QCD background Invariant Mass comparison(log)')
xlabel('Invariant Mass')
hold off
```



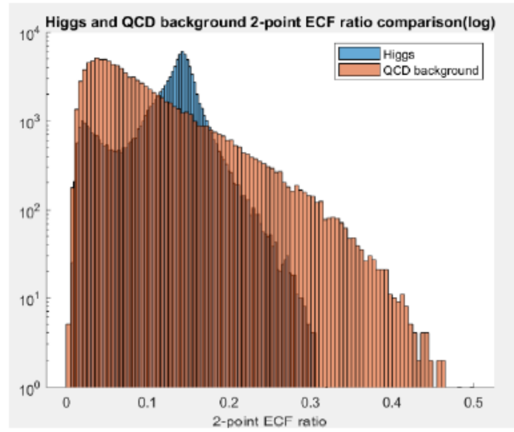

The difference between the Higgs and QCD data is fairly noticeable. Here, we are looking at the “irregularities” of the plot, which would give us information about the signal.

# EE2(E2) AND EE3(E3): ENERGY CORRELATION FUNCTIONS

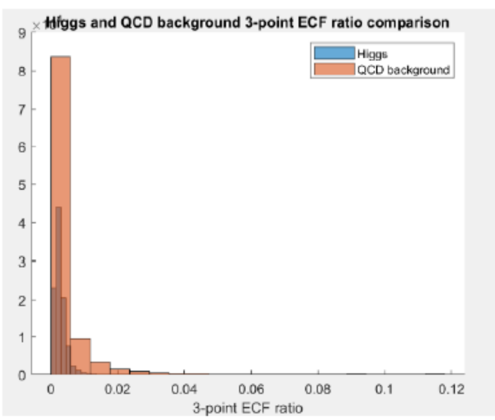
2-Point ECF ratio and 3-Point ECF ratio.



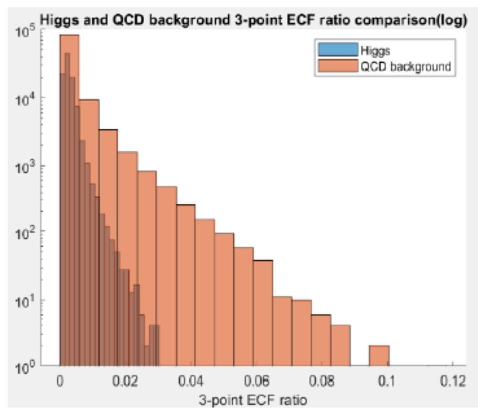

```
hold on
histogram(eetwo);
histogram(qeetwo);
legend('Higgs', 'QCD background')
title('Higgs and QCD background 2-point ECF ratio comparison')
xlabel('2-point ECF ratio')
hold off
```




```
hold on
histogram(eetwo);
set(gca,'YScale','log')
histogram(qeetwo);
set(gca,'YScale','log')
legend('Higgs', 'QCD background')
title('Higgs and QCD background 2-point ECF ratio comparison(log)')
xlabel('2-point ECF ratio')
hold off
```




```
hold on
histogram(eethree,b1);
histogram(qeethree,b1);
legend('Higgs', 'QCD background')
title('Higgs and QCD background 3-point ECF ratio comparison')
xlabel('3-point ECF ratio')
hold off
```




```
hold on
histogram(eethree,b1);
set(gca,'YScale','log')
histogram(qeethree,b1);
set(gca,'YScale','log')
legend('Higgs', 'QCD background')
title('Higgs and QCD background 3-point ECF ratio comparison(log)')
xlabel('3-point ECF ratio')
hold off
```

ee2:

$$\sum_{i<j\in J} P_{T,i}P_{T,j}\Delta R_{ij}\frac{1}{p_{T,J}^{2}}$$

LaTex:

\sum_{i<j \in J} P_{T, i} P_{T,j} \Delta R_{i j} \frac{1}{p_{T,J}^{2}}

ee3:

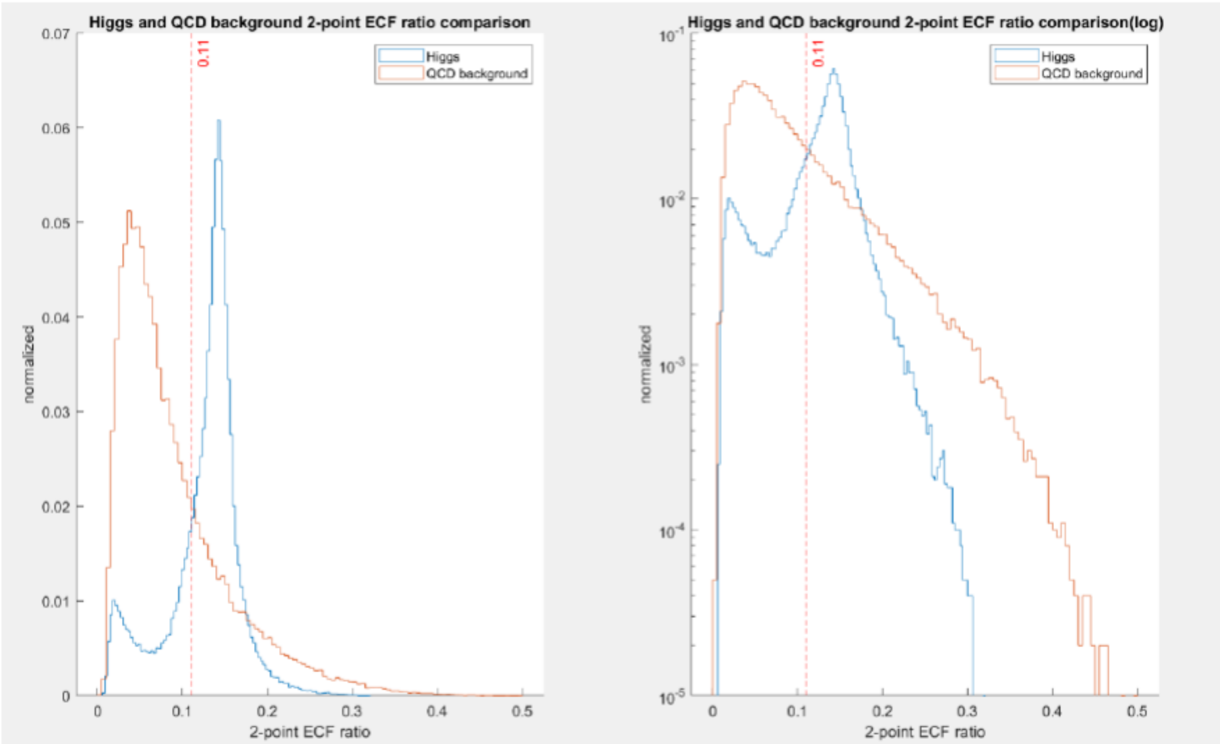
$$\sum_{i<j<k\in J} P_{T,i}, P_{T,j}, P_{T,k}\Delta R_{ij}\Delta R_{ik}\Delta R_{jk}\cdot\frac{1}{P_{T,J}^{3}}$$

LaTex:

\sum_{i<j<k \in J} P_{T,i}, P_{T,j}, P_{T,k} \Delta R_{i j} \Delta R_{i k} \Delta R_{j k} \frac{1}{P_{T,J}^3}

Note: for ee3, the QCD background "covers" the Higgs data completely. It is hard to differentiate the Higgs signal from the QCD background based on the parameter.

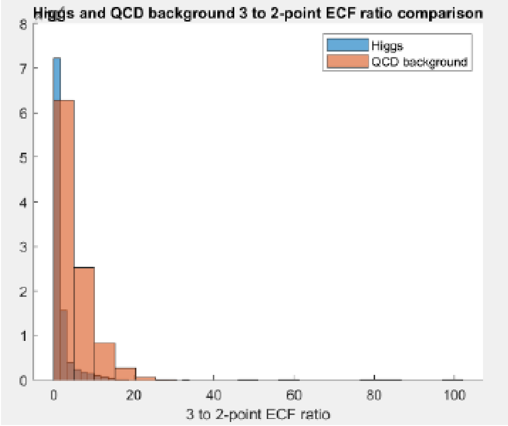
More on the Energy Correlation Functions
Energy Correlations Functions can be used as an inquiry for the jet substructure, meaning that we can use the functions to identify the N-prong substructure without a subject finding procedure.
We can use ECFs to discriminate between QCD jets and Higgs Bosons.

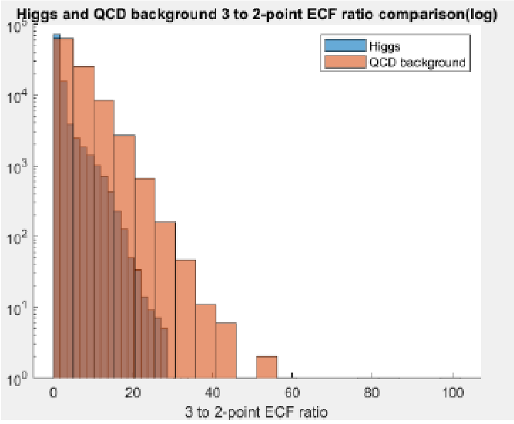



A noticeable difference between the signal and background occurs when 2-point ECF is greater than 0.11.

# D2(D2) JET SHAPE FUNCTION, USED IN TAGGING

3 to 2 - Point ECF ratio




```
hold on
histogram(dtwo,b1);
histogram(qdtwo,b1);
legend('Higgs', 'QCD background')
title('Higgs and QCD background 3 to 2-point ECF ratio comparison')
xlabel('3 to 2-point ECF ratio')
hold off
```

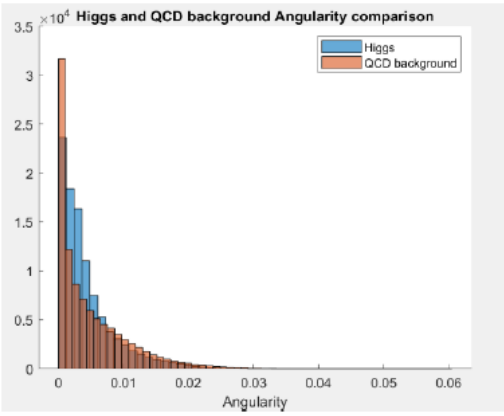
```
hold on
histogram(dtwo,b1);
set(gca,'YScale','log')
histogram(qdtwo,b1);
set(gca,'YScale','log')
legend('Higgs', 'QCD background')
title('Higgs and QCD background 3 to 2-point ECF ratio comparison(log)')
xlabel('3 to 2-point ECF ratio')
hold off
```

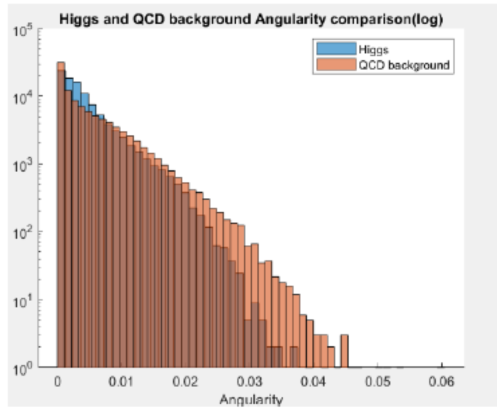
This parameter is calculated using the following equation:

$$D_2 = \frac{e_3}{e_2^3}$$

Note: The QCD background covers the Higgs data. The two datasets share similar distributions.

# ANGULARITY




```
hold on
histogram(angularity,b);
histogram(qangularity,b);
legend('Higgs', 'QCD background')
title('Higgs and QCD background Angularity comparison')
xlabel('Angularity')
hold off
```

```
hold on
histogram(angularity,b);
set(gca,'YScale','log')
histogram(qangularity,b);
set(gca,'YScale','log')
legend('Higgs', 'QCD background')
title('Higgs and QCD background Angularity comparison(log)')
xlabel('Angularity')
hold off
```

The parameter is calculated using the following equation:

$$\frac{1}{m_{J}} \sum_{i \in J} E_{i} \cdot \sin ^{-2}\left(\theta_{i}\right) \cdot \cos ^{3}\left(\theta_{i}\right)$$

Latex:

\frac{1}{m_{J}} \sum_{i \in J} E_{i} \cdot \sin ^{-2}\left(\theta_{i}\right) \cdot \cos ^{3}\left(\theta_{i}\right)

Notes:
"Angularities are a family of observables that are sensitive to the degree of symmetry in the energy flow inside a jet…..The measurement is aimed primarily at testing QCD, which makes predictions for the shape of the angularity distribution in jets where the small-angle approximation is valid…….Angularity is largely uncorrelated with all the other observables." (source: https://arxiv.org/pdf/1206.5369.pdf)
The Higgs and QCD data follow similar distributions.

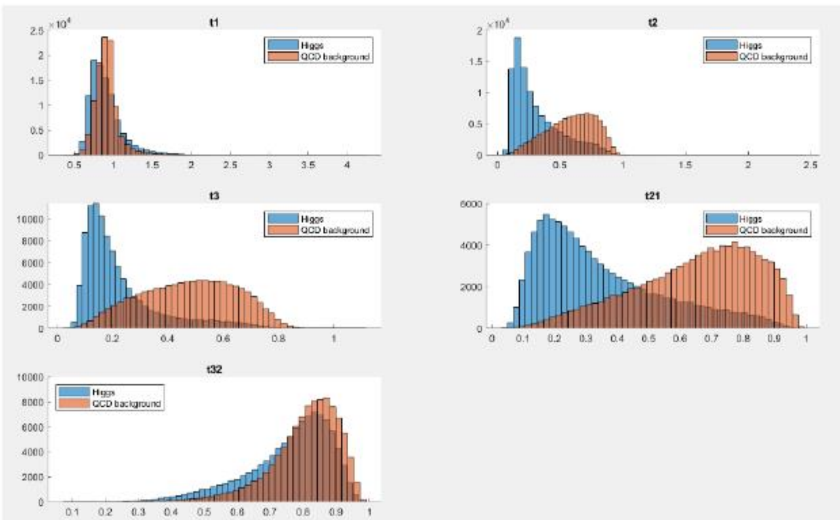
# N-SUBJETTINESS($\tau$N); JET SHAPE FUNCTION, USED IN TAGGING

"The variable is calculated by clustering the constituents of the jet with the $k_t$ algorithm and requiring that exactly N subjets be found.

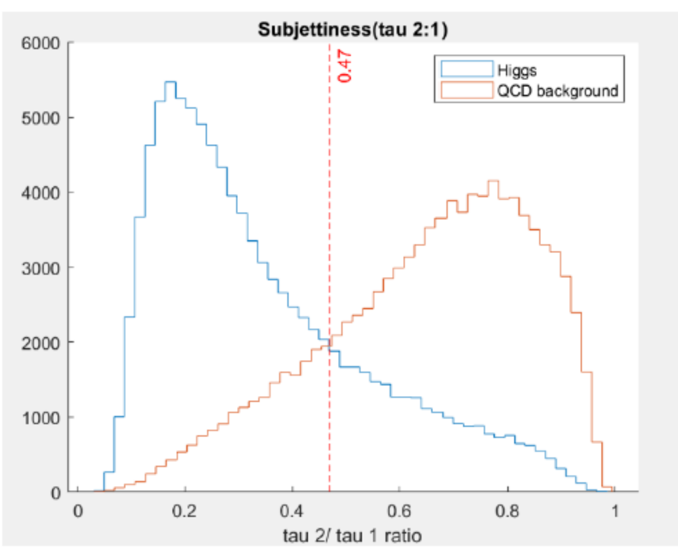
$$\tau_N = \frac{1}{d_0} \sum_k p_{\mathrm{T}k} \times \min(\delta R_{1k}, \delta R_{2k}, \ldots, \delta R_{Nk}) \; , \quad \text{with} \quad d_0 \equiv \sum_k p_{\mathrm{T}k} \times R$$

where R is the jet radius parameter in the jet algorithm, $p_{Tk}$ is the $p_T$ of constituent k, and $\delta R_{ik}$ is the distance from subjet i to constituent k. "(DSpace@MIT; http://hdl.handle.net/1721.1/84505)

- t1, t2, t3
  t21 = t2 / t1
  t32 = t3 / t2



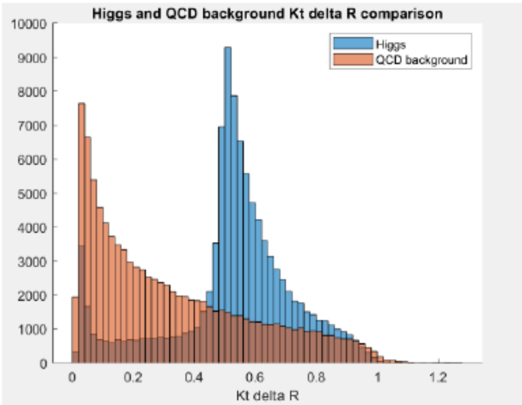

```
subplot(3,2,1)
hold on
histogram(tone,b);
histogram(qtone,b);
legend('Higgs', 'QCD background')
title('t1')
hold off
subplot(3,2,2)
hold on
histogram(ttwo,b);
histogram(qttwo,b);
legend('Higgs', 'QCD background')
title('t2')
hold off
subplot(3,2,3)
hold on
histogram(tthree,b);
histogram(qtthree,b);
legend('Higgs', 'QCD background')
title('t3')
hold off
subplot(3,2,4)
hold on
histogram(ttwoone,b);
histogram(qttwoone,b);
legend('Higgs', 'QCD background')
title('t21')
hold off
subplot(3,2,5)
hold on
histogram(tthreetwo,b);
histogram(qtthreetwo,b);
legend('Higgs', 'QCD background','location','northwest')
title('t32')
hold off
```



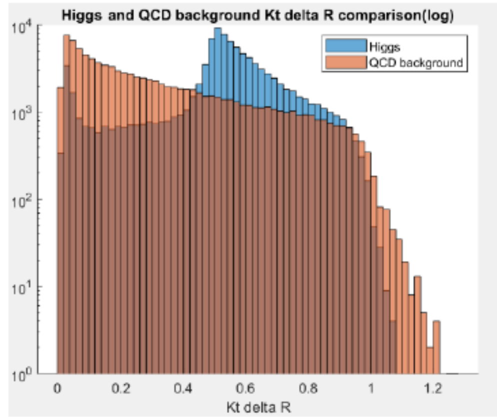

A tau21 ratio that is less than 0.47 would indicate that the data is more likely to be a Higgs signal rather than the QCD background.

More on Subjettiness.
The variable “describes to what degree the substructure of a given jet J is compatible with being composed of N or fewer subjects.”
Tau 21, the ratio of the N-subjecttiness functions associated with the standard subject axes, is used to generate the dimensionless variables that have shown to be particularly useful in identifying two-body structures within jets. (ATLAS Collaboration)

# KTDELTAR(KT ΔR) – KT CLUSTERING



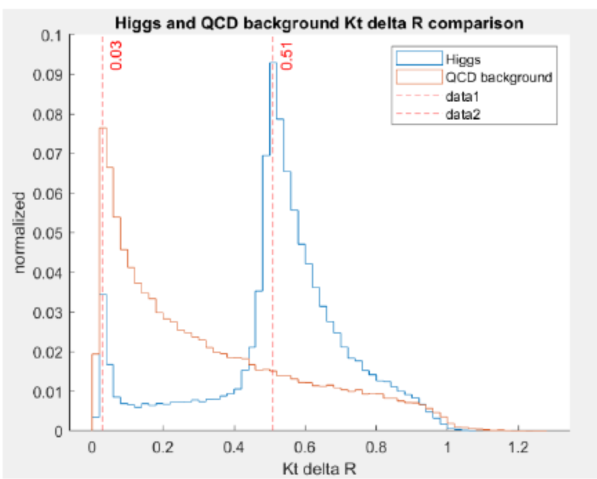

```
hold on
histogram(ktdeltar);
histogram(qktdeltar);
legend('Higgs', 'QCD background')
title('Higgs and QCD background Kt delta R comparison')
xlabel('Kt delta R')
hold off
```




```
hold on
histogram(ktdeltar);
set(gca,'YScale','log')
histogram(qktdeltar);
set(gca,'YScale','log')
legend('Higgs', 'QCD background')
title('Higgs and QCD background Kt delta R comparison(log)')
xlabel('Kt delta R')
hold off
```

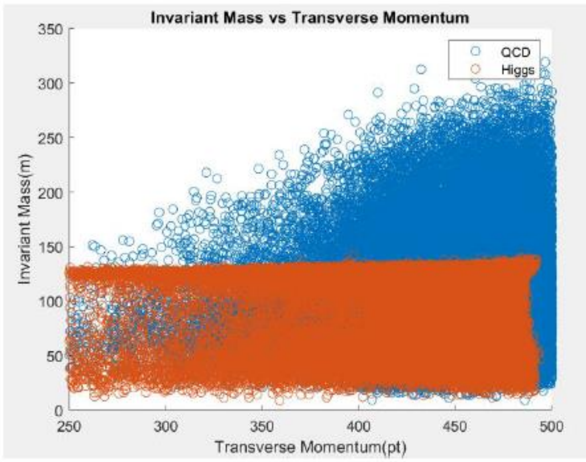
$K_t\Delta R$: ΔR of two subjects within the large-R jet

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\varphi^2}$$

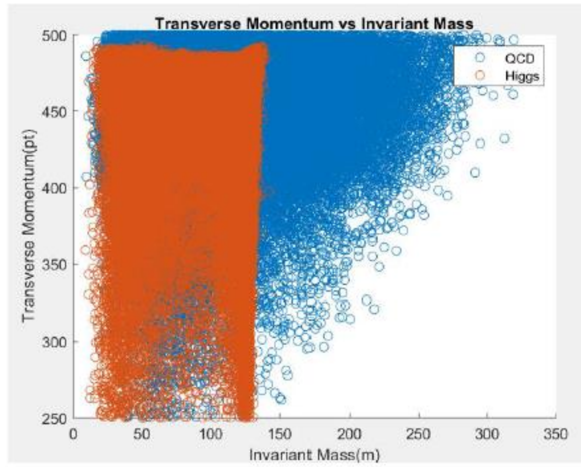



```
hold on
histogram(ktdeltar,'DisplayStyle', 'stairs','Normalization','probability');
histogram(qktdeltar,'DisplayStyle', 'stairs','Normalization','probability');
legend('Higgs', 'QCD background')
xline(0.03,'--r','0.03');
xline(0.51,'--r','0.51');
title('Higgs and QCD background Kt delta R comparison')
xlabel('Kt delta R')
ylabel('normalized')
hold off
```

# Higgs Sensitivity

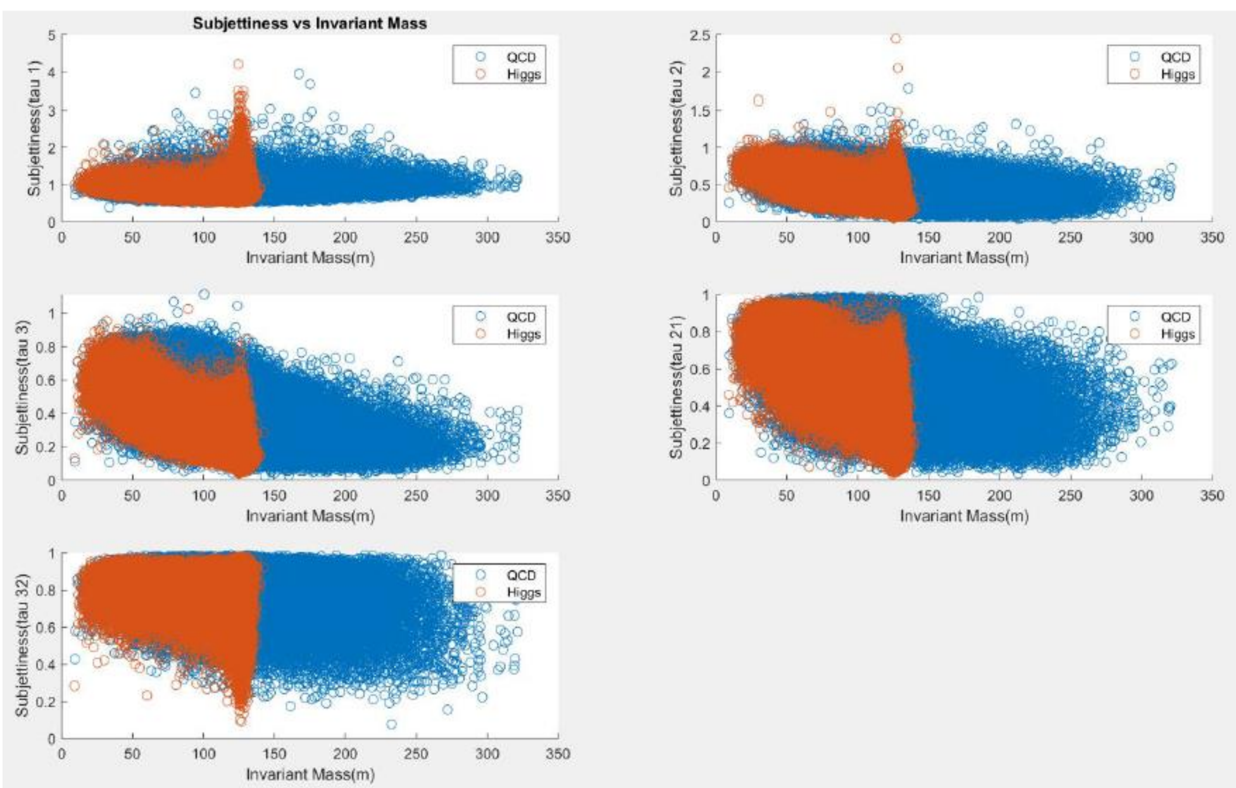


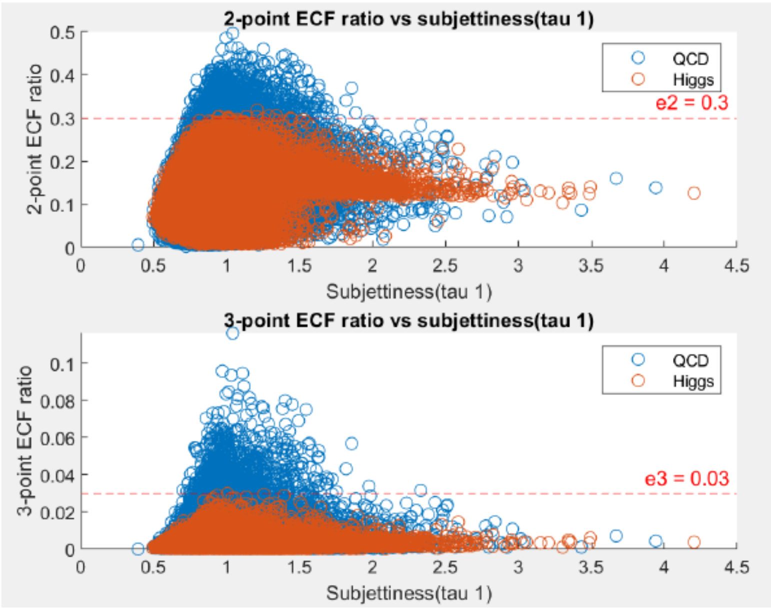

```
hold on
scatter(qpt,qmass)
scatter(pt,mass);
legend('QCD', 'Higgs')
title('Invariant Mass vs Transverse Momentum')
xlabel('Transverse Momentum(pt)')
ylabel('Invariant Mass(m)')
hold off
```



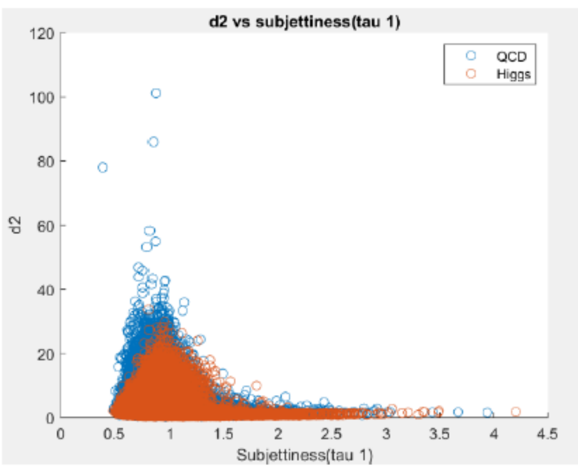

```
hold on
scatter(qmass,qpt)
scatter(mass,pt);
legend('QCD', 'Higgs')
title('Transverse Momentum vs Invariant Mass')
ylabel('Transverse Momentum(pt)')
xlabel('Invariant Mass(m)')
hold off
```

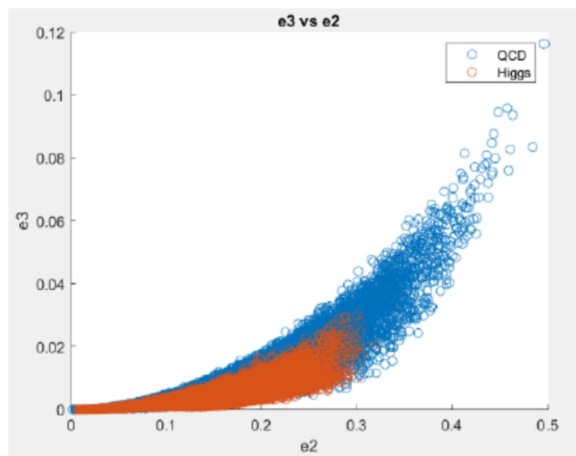
The red part is the Higgs signal. It seems that the signal has a threshold of around 125 GeV. Any invariant mass value significantly greater than this number should be considered a part of the background.






```
subplot(2,1,1)
hold on
scatter(qtone, qeetwo)
scatter(tone, eetwo)
title('2-point ECF ratio vs subjettiness(tau 1)')
yline(0.3,'--r','e2 = 0.3');
legend('QCD', 'Higgs')
ylabel('2-point ECF ratio')
xlabel('Subjettiness(tau 1)')
hold off
subplot(2,1,2)
hold on
scatter(qtone, qeethree)
scatter(tone, eethree)
title('3-point ECF ratio vs subjettiness(tau 1)')
yline(0.03,'--r','e3 = 0.03');
legend('QCD', 'Higgs')
ylabel('3-point ECF ratio')
xlabel('Subjettiness(tau 1)')
hold off
```



# Conclusions

1. The most noticeable and practical result we can obtain from analyzing the datasets is that the potential Higgs boson signal has an invariant mass of approximately 125 GeV.

2. It is observed that both the energy correlation function(ECF) and the N-subjettiness ($\tau$) can be used as a probe for the jet substructure. One parameter could be better than the other to be used depending on the situation.

Energy Correlation functions: In the d2 parameter, the background and the signal share the same distribution(Poisson-like). It seems like a good point where we apply a statistical test to test the significance of a given candidate signal.

N-subjettiness: $\tau21$, the ratio, showed a noticeable separation between the background and the Higgs signal.

3. It seems that it would be an effective way to partially separate the Higgs data from the QCD background to increase the signal detection accuracy and reduce the background noise by applying some constraints or thresholds depending on the parameters.

4. The ECF and the Subjettiness are the two most crucial parameters to discriminate particles. Further detailed statistical tests and analysis should be done based on the two parameters by truncating the datapoints or applying constraints.

# Bibliography

"The Large Hadron Collider | CERN." *Home.Cern*, 2019, home.cern/science/accelerators/large-hadron-collider.

Sirunyan, A. M., et al. "Inclusive Search for a Highly Boosted Higgs Boson Decaying to a Bottom Quark-Antiquark Pair." *Physical Review Letters*, vol. 120, no. 7, 14 Feb. 2018,

arxiv.org/abs/1709.05543, 10.1103/physrevlett.120.071802. Accessed 6 Nov. 2019.
"Higgs Boson Observed Decaying to b Quarks – at Last! | ATLAS Experiment at CERN."

*ATLAS Experiment at CERN*, 9 July 2018, atlas.cern/updates/physics-briefing/higgs-observed-decaying-b-quarks. Accessed 7 Nov. 2019.

"Mass / Invariant Mass | ATLAS Experiment at CERN." *ATLAS Experiment at CERN*, 25 Jan. 2018, atlas.cern/glossary/mass-invariant-mass. Accessed 9 Nov. 2019.

Aad, G, et al. "Performance of Jet Substructure Techniques for Large-R Jets in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV Using the ATLAS Detector." *Mit.Edu*, July 2013, dspace.mit.edu/handle/1721.1/84505, 1029-8479. Accessed 10 Nov. 2019.

Aad, G., et al. "Identification of Boosted, Hadronically Decaying W Bosons and Comparisons with ATLAS Data Taken at $$\sqrt{s} = 8$$ s = 8 TeV." *The European Physical Journal C*, vol. 76, no. 3, Mar. 2016, 10.1140/epjc/s10052-016-3978-z. Accessed 9 Nov. 2019.