Higgs Signal Optimisation

Joe Bentley and Jake Lane

Summary

1. General background on Higgs
2. How Higgs signals are simulated
3. How the signals are interpreted
4. How the Higgs signal is optimised
5. Problems in optimisation and improvements
6. Possible expansions of the project

Some Terminology

There is some terminology we will use in this, so we will define a few terms here

- Transverse momentum is the component of momentum perpendicular to the beam axis
- Azimuthal angle is the angle between the momentum and the beam axis
- Invariant mass is the same in all reference frames and defined as $m_0^2 = E^2 - |\mathbf{p}|^2$

Background

The Higgs boson is produced in many channels.

Background

The Higgs boson is produced in many channels. The most common in proton collider experiments is 'gluon gluon Fusion' (ggF)

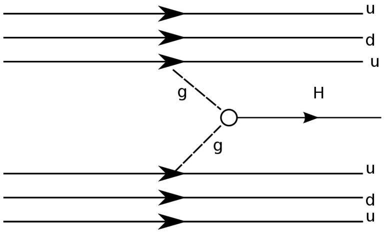

Figure: Gluons fusing into a Higgs at a proton-proton interaction

Decay

The Higgs decays in a very short period of time in many channels, the most common is 2 bottom quarks, but we investigate the decay into 2 photons (the diphoton channel.)

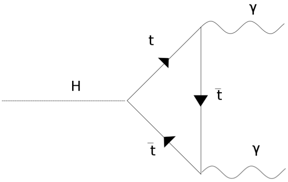

Figure: Decay of Higgs into 2 photons

This has a branching fraction of order of 10^{-3} but is much easier to detect experimentally.

Simulation

- The Higgs events and background events provided to us are simulated using PYTHIA.
- The simulation consists of a text file of the Energy and momentum (four momentum) of each photon in each collision event
- We will use one simulation of 10,000 Higgs events (which still have background in them) and one simulation of 1,000,000 background events

Parsing

- To parse the events, we wrote a script in Python 3 to parse the four momenta from the event data
- It calculates the invariant mass of each decay event and outputs this to a plaintext file for plotting in a histogram
- It can also apply a range of filtering on the events to filter out events that are not Higgs
- The script is written as a module so that we can use it from other scripts, for example we use it in the statistical significance script to apply the filtering

Weighting

In our simulated data, there are 10,000 Higgs events, and 1,000,000 background events. In an actual collider experiment this ratio would be much lower. To account for this we need to apply weightings to each dataset.

Weighting

In our simulated data, there are 10,000 Higgs events, and 1,000,000 background events. In an actual collider experiment this ratio would be much lower. To account for this we need to apply weightings to each dataset.

$$\frac{\text{num. Higgs events}}{\text{num. background events}} = \frac{\sigma(\text{Higgs produced}) \times B_f\,(H \to \gamma\gamma)}{\sigma(\text{background})}$$

By knowing the branching factor of the Higgs to two photons decay, as well as the cross sections of the background and of the Higgs production, we can use the above equation to calculate how much weighting we need to apply to the Higgs or background events in our histogram.

Filtering

- There are $\sim$ 100 Higgs events to $\sim$ 100,000,000 background events
- If we do not have good filtering then there is no chance of seeing the Higgs events amongst the background events
- We need an effective way to distinguish Higgs events from background events

Filter Methods

1. Number of particles
2. Transverse momentum
3. Energy
4. Azimuthal angle

Selection cuts

1. Number cut so that only events with 2 or more photons are counted
2. 2 Transverse momentum, p_T, cuts so that 1 photon has a larger p_T than p_{T1} and the other has a $p_T > p_{T2}$
3. 2 Energy E cuts, in a similar principle to the p_T cuts
4. Cuts for the angular difference between the 2 photons derived from the difference in azimuthal angle $\Delta\phi^2$ and the difference in pseudorapidity, $\Delta\eta^2$

Optimising our Filters

To optimise our filters so that we have the best ratio of Higgs events to background events, we need to optimise our filter methods for the highest statistical significance Σ,

$$\Sigma \equiv \frac{S}{\sqrt{S+B}} \tag{1}$$

where S is the number of filtered signal events and B is the number of filtered background events.

Optimising our Filters

To optimise our filters so that we have the best ratio of Higgs events to background events, we need to optimise our filter methods for the highest statistical significance Σ,

$$\Sigma \equiv \frac{S}{\sqrt{S+B}} \tag{1}$$

where S is the number of filtered signal events and B is the number of filtered background events.
We apply the filtering and calculate the significance for a series of different parameters, for example for different transverse momenta cuts, to see what gives us the best statistical significance.

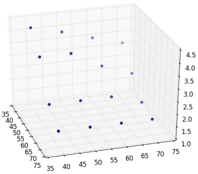

Figure: Statistical significance (height of points) for a series of different transverse momenta

This is a 3D plot generated from a script that calculates the statistical signifiance for various transverse momenta cuts. From graphs like these we can narrow down the ranges of transverse momenta that we filter until we get the highest statistical significance. We use PyPlot to generate our plots.

We can use the same method to check for example which azimuthal angles gives the highest statistical significance, by plotting a 2D scatter graph of statistical significance against scatter angle.

Seeing the Higgs

Actually observing it explanation, as in, mathematically.

Issues

- Since there are were so many background events, the python script was slow. Fixed using script to generate four momentum count headers on each event.

Further Explorations

Could possibly observe a different decay channel. Jets. Experimentation with PYTHIA.

- Since we have access to PYTHIA, the program that simulated the data, we could potentially generate new data
- Since our program is quite general, it wouldn't be too difficult to alter it for other Higgs decays

Conclusion/Questions