Higgs Signal Optimisation

Joe Bentley and Jake Lane

Summary

1. Kinematic variables
2. Background
3. How Higgs signals are simulated
4. How the signals are interpreted
5. How the Higgs signal is optimised
6. Problems in optimisation and improvements
7. Possible expansions of the project

Kinematic variables

4-momentum, p^μ is defined by 3 parameters for a photon:

$$E = p^t = p_T \cosh \eta \tag{1}$$
$$p^x = p_T \cos \theta \tag{2}$$
$$p^y = p_T \sin \theta \tag{3}$$
$$p^z = p_T \sinh \eta \tag{4}$$

1. p_T is the transverse momentum
2. η is the pseudo rapidity (which is the rapidity for a photon)
3. ϕ is the azimuthal angle

Kinematic Variables

Photons have the property $E = p = \sqrt{(p^x)^2 + (p^y)^2 + (p^z)^2}$ so rapidity y is the same as the pseudo rapidity, η.

$$
\begin{aligned}
\eta &\equiv \frac{1}{2}\log\left(\frac{p + p^z}{p - p^z}\right) \\
&= \frac{1}{2}\log\left(\frac{E + p^z}{E - p^z}\right) \equiv y
\end{aligned}
$$

1. Rapidity allows for a measure of how far forwards the photon travels parallel to the beam line
2. The differences $\Delta\phi^2$ and $\Delta\eta^2$ are Lorentz invariant so we can use these instead of a polar angle in the photon frame.

Background

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

Background

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

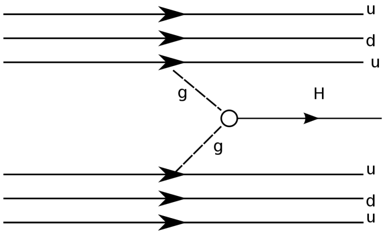

Figure : Gluons fusing into a Higgs at a proton-proton interaction, note the circle at the meeting of the 2 gluons represents the various quark (or W boson) loops that produce the Higgs

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.)

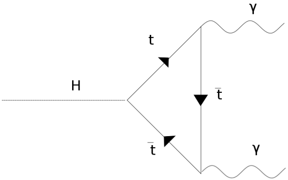

Figure : Decay of Higgs into 2 photons

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, to make the Higgs less prominent or the background more prominent
- We can apply the weightings directly when calculating our histogram

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, to make the Higgs less prominent or the background more prominent
- We can apply the weightings directly when calculating our histogram

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

Weighting

- The weighting is constructed such that the background events have weightings of 1.
- Higgs signals are weighted to produce a probability of producing a Higgs and that the Higgs decays into 2 photons and taking the ratio of simply producing 2 background photons.

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

Filter Methods

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, $\Delta\eta^2$

Physical meaning behind filtering

- E make the assumption that photons in the Higgs decay will be of high energy (or momentum)
- Transverse cuts which allow for photons to be cut that mainly decay perpendicular to the beam line.
- The $\Delta\eta$ and $\Delta\phi$ cuts allow for a Lorentz invariant angular separation between the photons. We can construct events that are 'back to back' which we expect from a Higgs decaying.

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,

$$\Sigma \equiv \frac{S}{\sqrt{S+B}} \tag{5}$$

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,

$$\Sigma \equiv \frac{S}{\sqrt{S+B}} \tag{5}$$

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.

Optimising our Filters

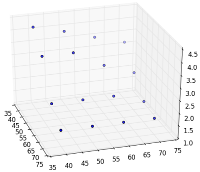

Figure : Statistical significance (height of points) for a series of different transverse momenta

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 (matplotlib) 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.

Invariant mass histogram

We calculate a histogram of invariant mass of the diphoton system ($m_{\gamma\gamma}$) using the 'numpy' (numerical python) library. Whilst this is not the primary aim (optimising a signal) it gives a visual representation of signal strength (i.e. 'seeing the Higgs').

Issues

- Since there are were so many background events, the python script was slow. Fixed by using script to generate four momentum count headers on each event. This allowed us to pre-calculate how many four momenta are in a given event, which is much faster.

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
- The program being entered around photon data would require changes to the filtering as we would have 4 momenta with non-zero invariant masses, so that background simulation is much more difficult, involving different particles.

Conclusion/Questions

Any questions?