Project 2: Discovering the Higgs boson

Key Topics: Numerical integration and maximum finding

1 Introduction

The discovery of the Higgs boson in 2012 at the Large Hadron Collider depended on optimising the analysis to produce the highest possible sensitivity to the signal from the dataset available. This project will go through a simplified version of such an analysis.

2 Searching for the Higgs

One of the ways the Higgs was discovered was through its decays to a pair of photons $H \to \gamma\gamma$. If the photons were measured perfectly, adding their energy-momentum four-vectors to find the resulting total invariant mass would give the Higgs mass. However, the experimental resolution results in the measured invariant mass having a Gaussian distribution centred on the Higgs mass with a non-zero width.

There are many other sources of photon production at the LHC and random combinations of pairs of photons not from Higgs decays will cause a large background. This background has a smooth distribution as a function of the invariant mass.

The Higgs was therefore detected by looking at the pairs of photons with an invariant mass near the Higgs mass and concluding that their number was higher than that expected from background. Any excess is quantified by putting a lower and upper mass selection cut, m_l and m_u respectively, on the photon pair invariant mass and integrating to find the expected number of photon pairs within the resulting mass range, as conceptually shown in the figure. The project is therefore about finding the optimal selection cuts to produce the best sensitivity to the Higgs signal.

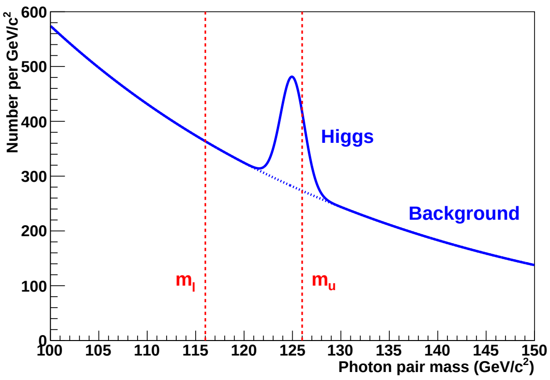

Figure 1: Conceptual shapes for the Higgs and background photon pair invariant mass distributions. The solid blue line is the total distribution of both Higgs and background pairs. The dashed blue line is the distribution for background only. Two selection cuts m_l and m_u define the selection range. The numerical values in this figure are not those given in the text as the figure is for illustrative purposes only.

3 Integration of a Gaussian

You will need to be able to integrate a Gaussian (or normal) function to determine the number of Higgs photon pairs between the selection cuts m_l and m_u. A standard Gaussian distribution with a mean of zero and a standard deviation of one is

$$G(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad \text{which gives} \qquad \int_{-\infty}^{\infty} G(x)\, dx = 1 \; . \tag{1}$$

Any other Gaussian can be derived from this function by shifting and scaling appropriately.

Write a function which integrates the above Gaussian distribution from $x = 0$ to some supplied input value of $x = a > 0$, i.e.

$$\int_0^a G(x)\, dx \, . \tag{2}$$

Treat this as a 'utility' function which could be used for any purpose and should be as accurate as possible. For your purposes, you should optimise the function accuracy over a range of a from 0 to 5.

1. Investigate several of the methods discussed in the course for doing the integral; specific integration methods, ODE methods or Monte Carlo methods could all be tried. Also, investigate if the result is more accurate when integrating the required range directly or integrating from a to ∞ and subtracting from the integral from 0 to ∞ to give the desired value. Vary any parameters of the methods to optimise the accuracy. The most accurate method and/or parameters may depend on the value of a; if so, your function should switch between them, depending on the input value.

2. Document your studies in the project report so as to state what methods you tried, the accuracy of each method, why you chose the eventual method(s) you used, and the final accuracy you obtained.

To determine the accuracy of your function, you can use a standard `erf` function or similar to get the true value.

Note: the Gaussian integration function counts as a major part of the project and should be done thoroughly. It gives you a good opportunity to show how well you have understood the concepts involved. However, be sure to allow enough time to complete the rest of the project. You could consider using a preliminary Gaussian integration function which is not yet fully optimised for accuracy to set up the code for the rest of the project and then come back to this part of the project to improve the function later.

4 Expected Higgs signal significance based on average values

To avoid any biases from the actual data sample obtained, the analysis selection cuts must be decided based on the average expected values *before* looking at the data. This is called a 'blind' analysis and is the gold-standard for scientific work.

A given set of selection cuts m_l and m_u will on average result in a number of background photon pairs N_B, given by integrating the background distribution over the selection range. This average number of selected background photon pairs can be parameterised by

$$N_B = \int_{m_l}^{m_u} B(m)\, dm \, , \qquad \text{where} \qquad B(m) = Ae^{-(m-m_H)/k} \tag{3}$$

with $A = 1500\,(\mathrm{GeV}/c^2)^{-1}$ and $k = 20\,\mathrm{GeV}/c^2$. This integral can be calculated analytically. If there is no Higgs boson, the actual number which will be observed in real data will be a random value from a Poisson distribution with a mean of N_B and so will have a standard deviation of $\sqrt{N_B}$.

Hence, to discover the Higgs, we need to observe a number of photon pairs greater than N_B which is not compatible with a Poisson statistical fluctuation around the average value. If the Higgs exists and produces an average of N_H photon pairs within the selection range, the average significance S of the Higgs excess will be

$$S = \frac{N_H}{\sqrt{N_B}} \, . \tag{4}$$

To claim a discovery, a value of $S \geq 5$ (often called a "five-sigma signal") is considered to be necessary.

The Higgs mass is $m_H = 125.1\,\mathrm{GeV}/c^2$ and the experimental resolution gives a Gaussian mass distribution centred on this value with an RMS of $1.4\,\mathrm{GeV}/c^2$. Assume theory predicts an average number of 470 Higgs decays to a pair of photons in total. This means the average number of selected photon pairs from the Higgs, i.e. N_H, is the integral of 470 times a Gaussian distribution integrated over the mass selection range

$$N_H = 470 \int_{m_l}^{m_u} H(m)\, dm \, , \tag{5}$$

where $H(m)$ is an appropriately shifted and scaled Gaussian distribution.

1. Use your Gaussian integration function within another function which returns the significance as a function of the low and high mass selection cuts. Study the dependence of the significance on the mass cuts and present your findings in your report.

2. Write code to find the optimal low and high mass selection cuts simultaneously by maximising the significance. This can use any of the 2D maximum-finding methods discussed in the notes. Find the optimal mass selection cut values to an accuracy of 1 keV/c^2 (note, not GeV/c^2). Comment on your results.

5 Observed Higgs signal significance

When the LHC experiments took real data, the actual number N of photon pairs observed in the chosen mass selection range was compared with the expected average background to give an observed excess of photon pairs $N - N_B$.

Assuming the Higgs exists, statistical fluctuations mean this observed excess will in general not be equal to the expected average N_H, but could lower or higher. Estimate the probability that an actual measurement will have a five-sigma signal.

6 Checks on analysis inputs

The inputs to your calculation are not known perfectly, so these will cause further errors on the number of excess events. Assume the following:

1. The Higgs mass is known to ± 0.2 GeV/c^2.

2. Some photons interact with the detector material and lose energy before being measured. If one of the photons in the pair does this, the average invariant mass is shifted down to 124.5 GeV/c^2 and the resolution is also degraded to 2.6 GeV/c^2. (If both photons in the pair interact, the mass is so smeared out it becomes part of the background.) The number of photons affected is uncertain and could be between 0 and 4%.

3. The theory predicting the number of Higgs bosons expected to be created at the LHC has a $\pm 3\%$ error.

Each of these uncertainties could cause a change to the average number of Higgs photon pairs within the selection range.

With your selection cuts fixed to the optimal range, study the changes to the average value of N_H as a function of the values of the Higgs mass, photon interaction fraction and number created, for each effect separately. Discuss which of the resulting uncertainties would not be negligible compared with the statistical error from the Poisson distribution.

Assuming these uncertainties are all uncorrelated, then by combining them with the statistical error, re-estimate the probability that the actual measurement will have a five-sigma signal.