Higgs Boson Discovery Report

Stefano Veroni, Bilgesu Aydın, Shuyu Jiao, Mohammed Hussain Roghey

***Abstract*—This report contains the results of the statistical analysis carried out on a randomly generated data set in Python emulating the experimental data that was collected at CERN in the discovery of the Higgs boson. The null hypothesis gave an alpha value of 8.45×10^{-8}, and is therefore rejected at a $10^{-6}\%$ significance level. This is evidence that the data simulated favours the alternative hypothesis which claims there is a signal of a particle with an alpha value of 0.284 at a mass of $124.9 \pm 0.3(\text{stat.}) \pm 0.5(\text{syst.})\text{GeV}$. Considering theoretical framework this signal most likely indicates the existence of the Higgs boson.**

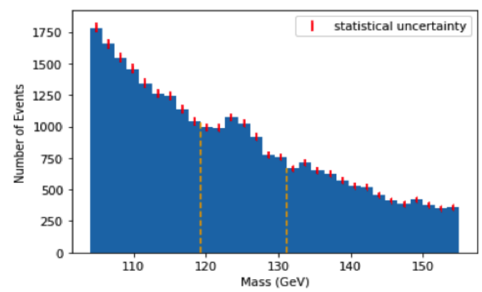

Fig. 1. The obtained data in range 104-155 GeV is plotted together with the statistical uncertainty. The dashed orange lines define the hypothetical Higgs Region

I. Introduction

By the 1970s the fundamental forces of electromagnetism and the weak interaction were unified in the electroweak force and what is known as the Standard Model, a classification of all known elementary particles and the interactions between them, was formed [1]. Over the years, experimental observations confirmed many predictions of the Standard Model. The theoretical framework which predicted the existence of the Higgs boson was established in 1964 by three independent groups [2]. The Higgs mechanism is a scalar field which interacts with particles of the Standard Model, thereby giving them mass [1]. The search for the Higgs boson, the particle associated with the Higgs field, has been one of the primary activities of the Large Hadron Collider (LHC) since its installation [2].

At the LHC beams of 10^{11} protons which pass each other at roughly 40 million times per second are collided head on which each other [3]. The are short-lived particles that are produced by these collisions are measured . The Higgs boson, which was hypothesized to be one of the products would then decay into two photons. The reconstruction of such decay was analyzed in order to investigate the existence of Higgs boson [3].

II. Data Generation and Parameterisation

A signal of approximately 10^5 events was randomly generated using Python, based on parameters [3] specifically chosen to emulate the experimentally observed data from CERN. Values considered in the range 104-155 GeV were then arranged into bins of width $dm = 1.7$ GeV, as shown in Figure 1. The background signal, which was expected to be described by an exponential function of the form $f(x) = Ae^{-\frac{x}{\lambda}}$, was parametrized considering the values below 120 GeV and above 130 GeV, as this restriction removes the region from the data, which, according to previous research [2], might be influenced by the existence of the Higgs Boson. First, reasonable estimates for the parameters were found based on some theoretical considerations. Considering the ratio between the first and the final bins in the considered range gives:

$$\frac{e^{\frac{-x_1}{\lambda}}}{e^{\frac{-x_f}{\lambda}}} = \frac{h_1}{h_f} \tag{2.1}$$

Then, solving for lambda leads to:

$$\lambda = \frac{x_f - x_1}{ln(h_1/h_f)} \tag{2.2}$$

As the number of events and the difference in height of the bins are very large, this method ensures high precision in the estimate. Then, the amplitude A of the curve was obtained by equating the area of the histograms and the area under the exponential curve:

$$dm.\sum h_i = \int_{104}^{120} Ae^{-\frac{x}{\lambda}} + \int_{130}^{155} Ae^{-\frac{x}{\lambda}} \tag{2.3}$$

Solving for A leads to:

$$A = \frac{dm}{\lambda} \frac{\sum h_i}{e^{-\frac{104}{\lambda}} + e^{-\frac{130}{\lambda}} + e^{-\frac{120}{\lambda}} + e^{-\frac{155}{\lambda}}} \tag{2.4}$$

Finally, by implementing the function `curve_fit` in the `scipy.optimize` Python package with the previously found estimates as guess values, the final values of the parameters were found. This fit gave a reduced chi squared value of 1.23 with the data in the range without the region of the hypothetical Higgs peak. Figure 2[a] displays the resulting distribution plotted on the data generated.

The alternative hypothesis taken into consideration, i.e. the existence of signal on top of the background that may be due to the Higgs Boson, can be represented by a probability distribution in which the background-only distribution, the previous exponential $f(x)$, and Higgs Boson's signal distribution, a gaussian $N(x)$, are summed together:

$$f(x) = Ae^{-\frac{x}{\lambda}} + \frac{G}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{(m_H - x)^2}{\sigma^2}} \tag{2.5}$$

In order to find the parameters describing such Gaussian, the differences between the observed data and the background distribution were examined as a function of the mass. The mass corresponding to the largest difference D was assumed to be the centre m_H of the Gaussian. Then, to calculate the width of the Gaussian consider the ratio

$$R = \frac{N(m_H - dm)}{N(m_H)} = e^{-\frac{1}{2}\frac{dm^2}{\sigma^2}} \tag{2.6}$$

which leads to the expression for the width:

$$\sigma = \sqrt{-\frac{1}{2}\frac{dm^2}{ln(R)}} \tag{2.7}$$

Finally, considering the difference D and the width of the Gaussian:

$$G = \sqrt{2\pi}\sigma D \tag{2.8}$$

Such estimates, together with the previous exponential function values, were then fitted by the `curve_fit` function. Figure 2[b] displays the resulting distribution plotted on the data generated.

III. Hypothesis Testing

As the background-only hypothesis was tested using the chi squared test on the whole range 104-155 GeV, the reduced chi squared value obtained was 3.07, implying a p-value of 8.45×10^{-8}. Oppositely, the alternative hypothesis's chi squared test, applied to the whole range, returned a reduced chi squared 1.134 and a p-value 0.284. The mass estimate for Higgs Boson was $m_H = 124.85 \pm 0.32$ GeV.

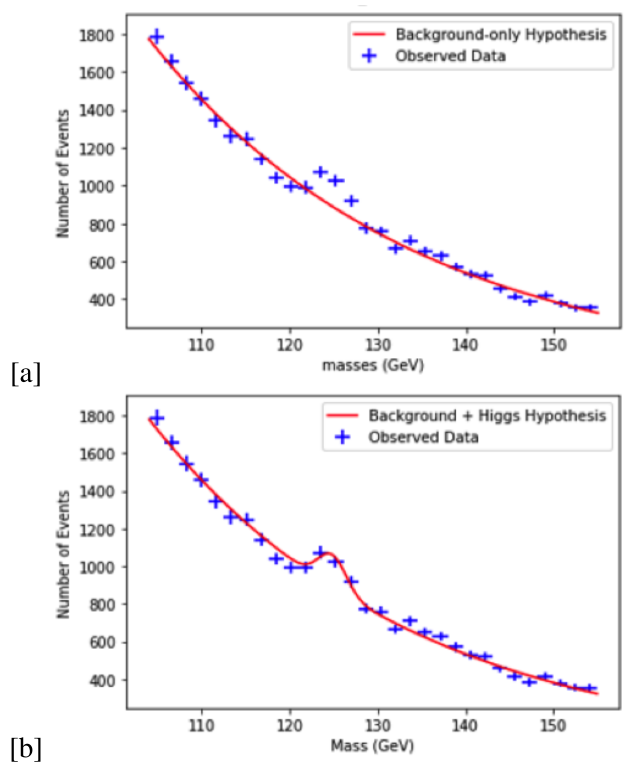

Fig. 2. Plots comparing the obtained fit with the data for [a] the background-only hypothesis and [b] the Higgs hypothesis.

IV. Results and Analysis

The statistical uncertainty in the data was taken as the square root of the bin height at every bin, as the number of events in each bin can be approximated by a Poisson distribution and the standard deviation for this can be obtained by square rooting the mean, that is, the height of the bin.

We reject the background only hypothesis at a 10^{-7} significance level since the infinitesimal p-value 8.45×10^{-8} is much smaller than 3×10^{-6} [4]. Such probability value and the corresponding chi square, in fact, cannot be explained by random fluctuations (see Appendix). Instead, the alternative hypothesis proved to be a much better fit to the data. The much higher p-value indicates high evidence of the existence of a new particle, likely the Higgs Boson, with a mass $m_H = 124.9 \pm 0.3(stat.) \pm 0.5(syst.)$ GeV. The 0.5 GeV uncertainty is a consequence of the impassable systematic error in the experiment [2].

V. Conclusion

By performing the chi squared test on both the null and alternative hypotheses, it was found that the former describes the observed data poorly, with a p-value equal to $8.45.10^{-8}$, whereas the alternative proved to be a much better fit, with a p-value of 0.284. There is sufficient data to conclude that there is a signal of a particle with mass $m_H = 124.9 \pm 0.3(stat.) \pm 0.5(syst.)$ GeV, in accord with the theoretical framework of the Higgs boson.

VI. References

[1] "The higgs boson." [Online]. Available: https://home.cern/science/physics/higgs-boson

[2] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," 2012. [Online]. Available: www.elsevier.com/locate/physletb

[3] M. Richards, "PHYS40005: Statistics of Measurement Group Computational Assignment 2020 SToM Computing Problem-Discovering the Higgs Boson," Tech. Rep., 2020.

[4] P. K. Sinervo, "Signal Significance in Particle Physics," pp. 18–22, 2002.

Appendix

In this section, the chi square probability density function (PDF) is briefly treated, aiming to show the reader that the chi squared obtained as we tested the null hypothesis cannot be satisfactorily explained by random fluctuations. To do so, 10000 random datasets were generated and fitted. The fitting function was the same used to generate the data, therefore we would expect perfect matching, i.e. the chi squared distribution should peak around the number of degrees of freedom: in this case it was 28. Figure 4 compares the 10000 chi squared values obtained with the expected chi squared distribution and the chi squared obtained as we tested the null hypothesis (≈ 85.9). It can be seen that the value of the chi squared PDF at that point is infinitesimal, and none of the 10000 chi squared obtained is not even near such value.

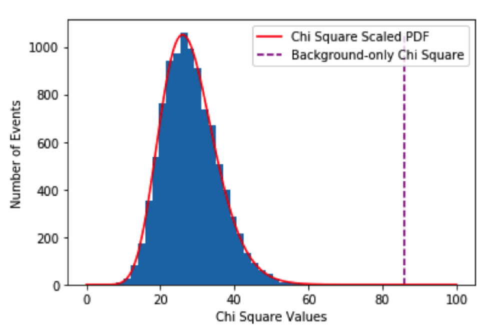

Fig. 3. The obtained data in range 104-155 GeV is plotted together with the statistical uncertainty. The dashed orange lines define the hypothetical Higgs Region