TEAM REFERENCE NO.DEF2

Statistics of the Measurement Project: Discovery of the Higgs Boson

Viraj Patel, Eleanor Borrill, Carla Mercier

***Abstract*— A significant number of events compared to the expected background was observed around a mass of 125 GeV. Our α value, also called p-value, for the background only hypothesis was 6.2e-8 against 0.0374 for the background plus signal hypothesis. As this excess signal is more than 5σ away from the background signal, we can confidently assert the existence of the Higgs Boson.**

I. INTRODUCTION

The Standard Model, a theory describing all interactions and elementary particles, has successfully predicted the interactions of high energy particles. The Standard model describes matter interactions through the exchange of force carriers[1], the photon, the W and Z bosons and the gluons. As the predicted symmetry between the electromagnetic and weak force interactions is broken, the elementary particles acquire mass, through the existence of a scalar particle: the Higgs Boson.[2]The research of the Higgs Boson is one of the main highlights of the Large Hadron Collider. The standard model Higgs boson is an unstable particle with immeasurably short lifetime, so it can be detected only through its decay products.

By analysing the decay product of a head on collision of two highly energetic proton beams at the Large Hadron Collider and looking at the mass distribution, one can observe the signal of the Higgs Boson at a high level of confidence. Simulating experimental data, we parameterized and fitted the signal distribution giving a smooth background shape, while the Higgs Boson gives a narrow Gaussian peak in mass around the excess signal of 125GeV. We used a chi squared method to test the hypothesis of a background signal only against the existence of the particle. The background hypothesis was rejected due to a poor goodness of fit. The statistical markers and uncertainties of the excess signal helped us prove the existence of the Higgs Boson Particle.

II. DATA GENERATION AND PARAMETERIZATION

By simulating a data set of 400 events corresponding to the number of events at the time of the discovery [3],we generated a set containing the corresponding rest mass values, and thus, we created a histogram corresponding to the rest mass distribution. To be able to analyze the signal, one can firstly parameterize the background signal, using an exponential fit of the form:

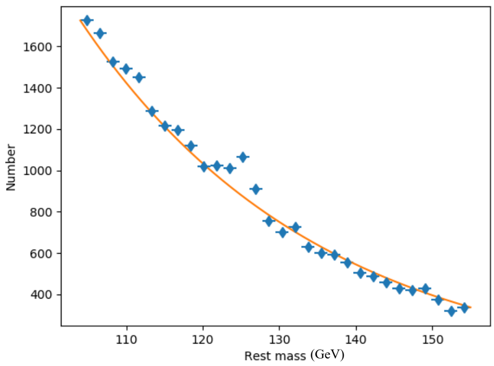
$$B(x) = ae^{-\frac{x}{\lambda}} \quad [1]$$

Where a is the normalization factor, and λ is the gradient of the exponential decay.

The maximum likelihood method was the first method used to estimate these two background parameters. With this method we obtained a=1964.0 and λ=126.1. However, the maximum likelihood method can only be applied to a normalized probability density function, which is not the case in our data, hence giving poor fitting parameters, as discussed later.

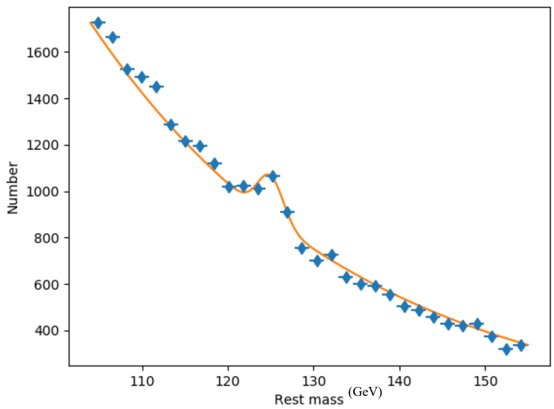
Therefore, we took the value of the first and last points of the histogram to have a first estimate of the parameters and then we solved the corresponding simultaneous equations. Using a 2D chi square loop to optimize the fit, we found the final values for the background parameters, by assuming the estimated values were on the fitting line. We found as final values a=4820.6 and λ = 31.2.

Using the final parameters, we fitted the background signal giving the expected smooth background shape.



Graph 1: Histogram of the simulated positive rest mass data sets with an exponential distribution fitting for the background signal. The vertical uncertainties represent the associated standard deviation of the stimulated set.

Additionally, the excess signal could be parametrized as Gaussian Signal with parameters- A=700, μ = 125 GeV, σ = 1.5 GeV.



Graph 4: Fitting of the signal using a Gaussian distribution, the goodness of the fit is discussed later

In order to confirm that the peak was indeed the Higgs Boson, we generalised our fitting algorithm to find the optimized parameters of the signal. We used the same estimation method for A and lambda as the background only hypothesis. To estimate the mean, one can find the location of the local maximum where the prominence is maximum. We assumed that the signal meets the background at 2σ away from the mean, so by finding the difference between the mean and the point where the background met the signal and dividing by two, we found an estimate for σ. Finally, to estimate the signal amplitude, one can find the height of the Gaussian peak using the formula: $A = (G(\mu) - B(\mu))\sigma\sqrt{2\pi}$ [2]

Where G is the function for background plus signal hypothesis and μ is the mean

Hence, we found for the Gaussian parameters: A=818.7, σ=1.7, and μ=125.25, and for the background signal, λ=31.2 and a= 48270.6

III. Hypothesis Testing

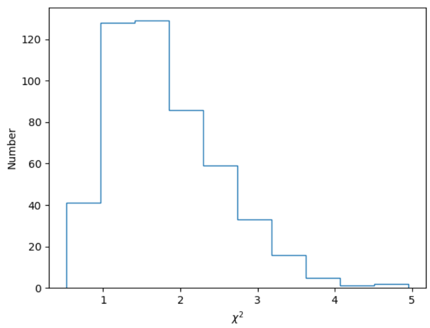
By looking at the goodness of the resulting fit, we then tested two hypotheses to see which one was more consistent with the experimental data.

In the hypothesis of a background only signal, we calculated the reduced chi square and the chi square value to examine the goodness of the fit in the mass range region of [104,155] GeV.

	Chi square λ^2	*Reduced Chi square*	*α Value*
Background only Hypothesis	86.8	3.1	6.2e-8

Table 2: The reduced chi square value corresponds to the chi square value divided by the degree of freedom. When parsing the chi square value in the α value function, one needs to multiply it by the degree of freedom.

We can clearly see that the α value, also called the significance level is relatively small. By definition, the α value is the probability of rejecting the null hypothesis when the null hypothesis is true[4]. Therefore, we can say that the hypothesis is false at the 6.2e-6% level. Even if the probability of making a Type 1 error is extremely small, one has to take into account the random fluctuations between each simulation. To account for this uncertainty, we repeated the simulation 500 times and formed a distribution of the associated reduced chi square value.



Graph 2: Histogram distribution of the reduced chi square value of the 500 simulations in the hypothesis of a background only signal

The distribution looks as expected because we can observe a Gaussian distribution after its peak which is characteristic of the order of 20 degrees of freedom[5]. In this graph, we can find the value of the background only reduced chi square, however the low number of occurrences highlights the poorness of this fit. We then tested the background plus signal hypothesis:

	Chi square λ^2	*Reduced Chi square*	*α Value*
Background+signal Hypothesis	13.95	0.558	0.0374

Table 3: Statistical parameters found for the background plus signal hypothesis.

IV. Results and Analysis

The true values with which we compared our estimates were a mean of 125 GeV, a standard deviation of 1.7 and a signal amplitude of 700 [2] .The analysis of our estimates is therefore dependent on prior credibility of the theory we are testing.

We used a 5σ marker to test the likelihood of our hypotheses, corresponding to a p-value of 3e-7 [6] .This is our threshold value for evidence of this particle. In other words, the probability that our background-only hypothesis results in the observed signal is less than 1 in about 3.5 million [6].

Our calculated α value for the background-only hypothesis was 6.2e-8. As this result is lower than the 5σ significance level, we can confidently reject this hypothesis because the α value implies that it is not a good fit for the data. Our calculated α value for the background and Higgs signal hypothesis was 0.0374. This value is greater than our threshold α value so we can infer that this hypothesis is likely to be true. Therefore, this signal is a statistically significant marker of a new particle at the Higgs energy. Also, from this hypothesis, we calculated a mean of 125.25 GeV, standard deviation of 1.7 (equal to the uncertainty of our results) and signal amplitude of 818.7. Comparing these values with the true values above, with a percentage error of 0.2% for the mean estimate, and 17% for the amplitude, and a negligible error for the standard deviation value, we can conclude that this hypothesis is a good fit for the data as it has yielded highly accurate results.

V. Conclusion

Our results from a chi-squared test corresponded to a p-value lower than the threshold value for the background-only hypothesis; causing us to reject this hypothesis at the 5σ significance level. We inferred that the background and Higgs signal hypothesis is likely to be true as the α value was higher than the threshold value and the calculated values of the

Higgs mass and standard deviation were very accurate. A main restrain to this project was the processing speed of our computer which restricted us in the amount of data we could process, and a potential source of uncertainty was the number of significant figures produced by the Python programming language. Had we used a language built for data analysis for particle physics like ROOT, our uncertainty would be reduced.

VI. References

[1] Ecker, Gerhard., and SpringerLink. 2019. *Particles, Fields, Quanta From Quantum Mechanics to the Standard Model of Particle Physics* Web.
[2] Aad, G., Abajyan, T., Abbott, B., Abdallah, J. and Abdel Khalek *et al*, 2012. *Observation Of A New Particle In The Search For The Standard Model Higgs Boson With The ATLAS Detector At The LHC.* ATLAS Collaboration.
[3]Ali, A., 2013. *Higgs Particle(S) Higgs Particle(S): Physics Issues And Experimental Searches In High-Energy Collisions.* Springer-Verlag New York, p.396.
[4] Tenny, S. and Abdelgawad, I., 2017. *Statistical Significance*. Europe PubMed Central.
[5] Courses.atlas.illinois.edu. 2016. *P Value Calculator, Chi-Square.* [online] Available at: <http://courses.atlas.illinois.edu/spring2016/STAT/STAT200/pchisq.html> [Accessed 16 June 2020].
[6] Lamb, E. 2012. *5 Sigma What's That?* Scientific American.
[7] Verkerke, W., 2020. *Discovering The Higgs – Finding The Needle In The Haystack.* [online] Indico.cern.ch. Available at: <https://indico.cern.ch/event/327060/contributions/1713139/attachments/635288/874524/cern-openlab-higgsstat-v9.pdf> [Accessed 16 June 2020].