Discovering the Higgs boson using a simulated data set

D. Bai, A. Li, H. Niu, M. Yan, Q. Zhao

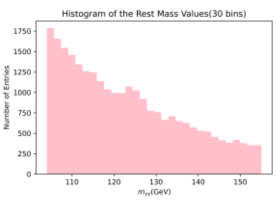
Abstract— The data generated were tested under two hypotheses over 104-155 GeV mass range. The background only hypothesis gave a χ^2 value of 677 while the background and signal hypothesis gave a χ^2 value of 1.32 and a corresponding p-value of 0.120, showing the existence of the Higgs boson. The χ^2 value is then plotted against the mass range to confirm that it has a rest mass of around 125 GeV

I. INTRODUCTION

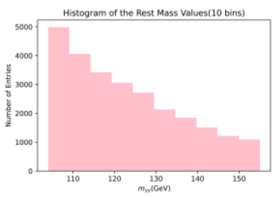
The Standard Model predicted the existence of Higgs boson, however, the proof of this took physicists 40 years. One of the difficulties of observing the Higgs boson is that its lifetime is too small. Instead of observing Higgs boson, the better method is to observe the products of the decay of Higgs boson. One of the decay modes of Higgs boson is two photons. Due to the Conservation of Mass-Energy, we can use the energy of the final products which are two photons to derive the mass of Higgs boson. The mass distribution of photons can be analysed using different statistical tests including Chi-square and hypothesis testing. If the small bump in the middle of the graph corresponding to the background and signals that has a P value larger than the critical value, then the existence of Higgs boson can be proved.[1][2]

II. DATA GENERATION AND PARAMETERISATION

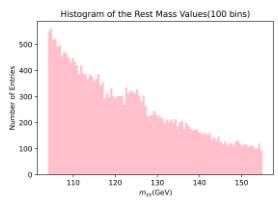
The *generate_data(n=400)* function creates a total of 1000400 data - 400 of which corresponds to the signals in real-life. The remaining data are background



(a)



(b)



(c)

Fig. 1: the histograms of the rest mass values for 30, 100, and 10 bins

entries. The data generated will be different each time but is fixed by the random seed.

As shown in Fig. 1(a), the shape of the histogram is similar to the results published by the CMS experiment and the bump at around 125 GeV can be observed. For Fig. 1(b), the bump becomes more pronounced. As the number of bins decreases to 10, the bump can no longer be observed.

The background distribution is given by:

$$B(x) = Ae^{-x/\lambda} \quad (1)$$

Where λ can be estimated by[3]:

$$\lambda \approx \hat{a} = \frac{\sum_i x_i}{N} \quad (2)$$

Where $\hat{a}$ is the maximum likelihood estimation for λ and N is the amount of data. When there is an uppercut at 120 MeV the estimation for λ is 27.75 $\pm$ 0.03 and when the signals are included, the estimation becomes 30.02 $\pm$ 0.03. The uppercut value will be used because only the background should be considered. A from eq. 1 can be estimated by equating the area of the histogram to the area of the background distribution from 0 to 120 GeV. Fig. 2 shows that when only the background is considered, the curve does not fit nicely on the histogram.

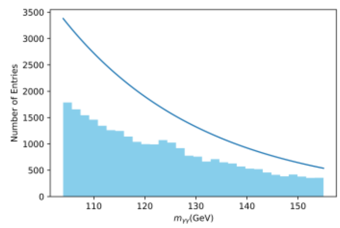


Fig. 2: The background expectation curve extrapolated over the 104-155 GeV mass range

III. HYPOTHESIS TESTING

Both the background only and the background and signal hypotheses were tested by calculating a reduced χ^2 statistics, using the *get_B_chi* function and the *get_SB_chi* function. Their respective p-values are then calculated using the *chi2* function and compared to $\alpha = 0.05$.

By using the *get_B_chi* function on the parameters found using the estimation method, a reduced χ^2 statistics of 159.6 was calculated. The reduced χ^2 statistics can be improved by using a 2D scan that finds the minimum reduced χ^2 statistics for selected ranges of A and λ, and the values of A and λ obtained can be used as the initial guess in the *curve_fit* function. After fitting, a reduced χ^2 statistics of 0.96 was obtained when A = 133655 and ⅄ = 29.9316. Moreover, due to random fluctuations, the reduced χ^2 statistics varies. This variation can be investigated by plotting a histogram of the reduced χ^2 statistics over many simulations. Fig. 3 shows that this distribution is approximately gaussian, which is in accord with the Central Limit Theorem.

The background and signal hypothesis is tested using the *get_SB_chi* function where the signal follows a gaussian distribution with a signal amplitude of 700, a μ of 125 GeV and a σ of 1.25 GeV.

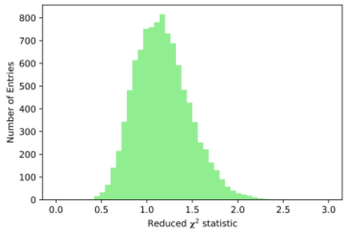


Fig. 3: The reduced χ^2 statistics for 10,000 simulations

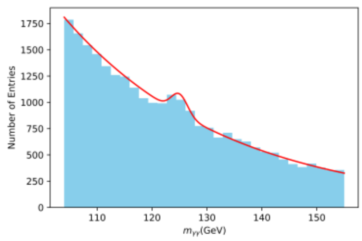


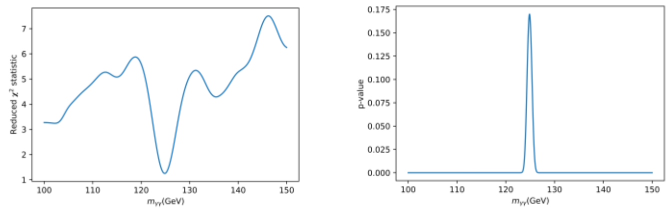
Fig.4: The background and signal curve extrapolated over the 104-155 GeV mass range

IV. Results and Analysis

As shown in Fig.3, when the χ^2 values are calculated for the background only scenario (0-120 GeV), they are mostly below 2.0. When the reduced χ^2 statistics for the background only hypothesis is calculated over the 104-155 GeV range, which includes the signal, its value becomes 677 and at $N_{dof} = 28$, the p-value is approximately 0, hence we reject the background only hypothesis.

For the background and signal hypothesis, by performing another 2D scan over the full mass region which includes the signal, the new fitted values for A and ⅄ are 59474 and 29.778 respectively. The signal and background χ^2 statistics is 1.32, which is considered a good fit. Fig. 4 shows the fitted curve over the 104-155 GeV mass range and a bump at approximately 125 GeV can be observed. At $N_{dof} = 28$, the p-value for the χ^2 statistics is 0.120 which is larger than $\alpha = 0.05$. Therefore, the p-value is not in the critical region and the background and signal hypothesis is not rejected.

Furthermore, assuming that the signal mass position is unknown, the reduced χ^2 statistics can be calculated for a range of masses under the signal and background hypothesis. Fig.5 (a) shows that the χ^2 value is at a local minimum when the rest mass is around 125 GeV, and the spike in p-value in Fig.5 (b) also corresponds to 125 GeV. These suggest the existence of the Higgs boson.



(a) (b)

Fig. 5 (a): the reduced χ^2 statistics for mass range 100-150 GeV. (b): the p-value for mass range 100-150 GeV

V. Conclusion

The data generated were tested using both hypotheses. For the background only hypothesis the χ^2 value obtained in the 104-155 mass range is 677 which is rejected. For the background and signal hypothesis the χ^2 value obtained in the 104-155 mass range is 1.32 and the corresponding p-value is 0.120 which is larger than $\alpha = 0.05$. Moreover, the plot of the reduced χ^2 statistics against rest mass suggests that χ^2 is minimum when the rest mass is approximately 125 GeV confirms the rest mass of the Higgs Boson.

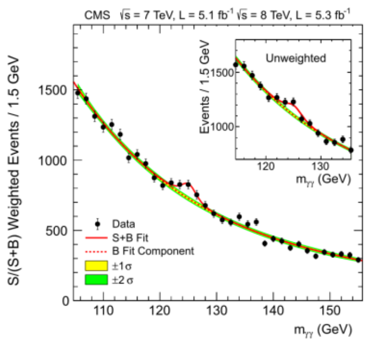
VI. References

[1] S. Chatrchyan etc.Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,Physics Letters B,Volume 716, Issue 1,2012,Pages 30-61, ISSN 0370-2693, https://doi.org/10.1016/j.physletb.2012.08.021.

[2] Carmi, D., Falkowski, A., Kuflik, E. et al. Higgs after the discovery: a status report. J. High Energ. Phys. 2012, 196 (2012). https://doi.org/10.1007/JHEP10(2012)196

[3] Dr Mark Richards, 2022, Year 1 Statistics of Measurement

Appendix A



Appendix A: The diphoton invariant mass distribution with each event weighted by the S/(S + B) value of its category. The lines represent the fitted background and signal, and the coloured bands represent the ±1 and ±2 standard deviation uncertainties in the background estimate. The inset shows the central part of the unweighted invariant mass distribution. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this Letter.) Reference: The CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Physics Letters B, 716-1