

# Comment on “Accumulating Evidence for the Associate Production of a Neutral Scalar with Mass around 151 GeV”

Andrew Fowlie\*

*Department of Physics and Institute of Theoretical Physics,  
Nanjing Normal University, Nanjing, Jiangsu 210023, China*

A recent paper [1] accumulates evidence for a new fundamental particle by combining several CMS and ATLAS searches for the Standard Model Higgs boson. The putative particle is a neutral scalar,  $S$ , with a mass of about 151 GeV. The reported significances are  $5.1\sigma$  local and  $4.8\sigma$  global. This nearly reaches the  $5\sigma$  threshold for a discovery in high-energy physics. In this brief note we cast doubt on the strength of the evidence for a new particle. After taking into account the fact that signals were fitted to six different channels, we find that the significances are only  $4.1\sigma$  local and  $3.5\sigma$  global. The code and instructions for reproducing our calculations are available at [this github repo](#).

## I. INTRODUCTION

In ref. [1] several ATLAS and CMS resonance searches are combined into six channels. In each channel, the background model predicts a smooth non-resonant spectrum parameterised by four nuisance parameters, collectively denoted by  $\theta$ . The signal model, on the other hand, predicts a Crystal-ball shaped resonance in each channel parameterised by seven parameters: a single mass parameter,  $m$ , common to each channel, and six positive signal strength parameters  $\mu_i \geq 0$  for  $i = 1 - 6$ , one for each channel.

Ref. [1] construct a test-statistic based on a profiled likelihood ratio for the background model and background plus signal model,

$$\lambda \equiv -2 \log \frac{\max_{m, \mu, \theta} \mathcal{L}(m, \mu, \theta)}{\max_{\theta} \mathcal{L}(\mu = 0, \theta)}, \quad (1)$$

where  $\mathcal{L}(m, \mu, \theta)$  is the likelihood of the observed data for the model with parameters  $m$ ,  $\mu$  and  $\theta$ ,  $\mu = 0$  corresponds to the background model, and the maximisation over the signal strengths is constrained by  $\mu_i \geq 0$ . Ref. [1] find  $\lambda = 26.01$  [2]; we haven’t verified this result, but assume that it is correct. This test-statistic is used to compute local and global significances. We discuss the interpretation of significances as evidence in ref. [3, 4]; we focus here only on their correct computation. Although the computation is complicated by the look-elsewhere effect (LEE) and the fact that the signal strength parameters must be positive, the issue with ref. [1] is simple: they computed the significances as though they fitted a single signal when in fact they fitted six independent signal strengths.

## II. LOCAL SIGNIFICANCE

To compute significances, it is common to resort to asymptotic formulae (see e.g., ref. [5]) based on Wilks’

theorem [6]. Roughly speaking, this theorem states that test-statistics of the form eq. (1) are asymptotically  $\chi_k^2$  distributed, where  $k$  is the number of parameters describing the signal. The resonance search, however, violates a required assumption (see e.g., ref. [7]), as the mass parameter isn’t identifiable under the background model, i.e., when  $\mu = 0$ , the model doesn’t depend on  $m$ .

By simply fixing the mass parameter to its best-fit value  $\hat{m}$ , however, we may compute a local  $p$ -value. In this case, the test-statistic splits into independent contributions for each channel,

$$\lambda = \sum_{i=1}^6 -2 \log \frac{\max_{\mu_i, \theta_i} \mathcal{L}_i(\hat{m}, \mu_i, \theta_i)}{\max_{\theta_i} \mathcal{L}_i(\mu_i = 0, \theta_i)}. \quad (2)$$

There is a further complication, though, as the signal strength parameters must be positive, such that the background model lies at the boundary rather than in the interior of the background plus signal model. This is addressed by Chernoff’s modification to Wilks’ theorem [8] and further explored in ref. [9]. For each channel, if the best-fit signal at a fixed mass fluctuates downwards, that channel’s contribution to the test-statistic in eq. (2) vanishes. If it fluctuates upwards, that channel’s contribution to the test-statistic is described by a chi-squared distribution with one degree of freedom as usual. This is a mixture distribution called a half chi-square distribution,  $\frac{1}{2}\chi^2$ . The probability density function is (see case 5 in ref. [9] or discussion around eq. 52 in ref. [5])

$$p(\lambda) = \frac{1}{2}\delta(\lambda) + \frac{1}{2}p_{\chi_1^2}(\lambda). \quad (3)$$

where  $p_{\chi_1^2}(\lambda)$  is the probability density function for a  $\chi^2$  distribution with one degree of freedom. The terms in this mixture correspond to downwards and upwards fluctuations and the coefficients are half because downwards and upwards fluctuations are equally likely.

The test-statistic for  $n$  channels as in eq. (2) for  $n = 6$  is thus the sum of  $n$  half-chi-squared variates. We denote this as a  $\frac{1}{2}\chi_n^2$  distribution. The density is (see case 9 with

\* [andrew.j.fowlie@NJNU.edu.cn](mailto:andrew.j.fowlie@NJNU.edu.cn)

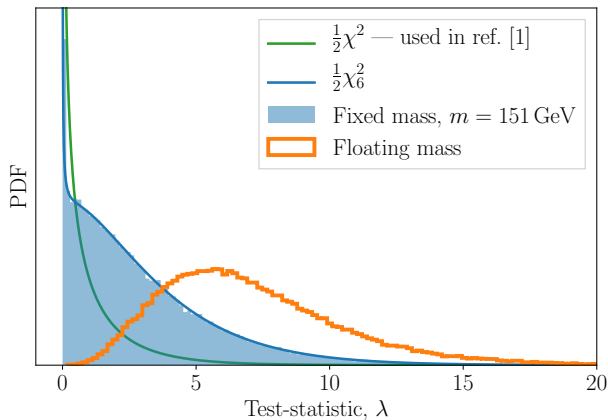


Figure 1. Distribution of test-statistics from MC simulations. The  $\frac{1}{2}\chi^2$  and  $\frac{1}{2}\chi_6^2$  distributions are shown by smooth lines for comparison.

$s = 0$  in ref. [9])

$$p(\lambda) = \left(\frac{1}{2}\right)^n \sum_{i=0}^n \binom{n}{i} p_{\chi_i^2}(\lambda), \quad (4)$$

where  $p_{\chi_0^2}(\lambda) = \delta(\lambda)$ . The coefficients are just the binomial coefficients describing the chances of  $i$  upwards fluctuations at the fixed mass in  $n$  channels. When  $i$  of them fluctuate upwards, the test-statistic is a sum of  $i$  chi-squared variates which follows a  $\chi_i^2$  distribution. When  $n = 1$ , this reduces to eq. (3).

In the  $n = 1$  case, the density in eq. (3) leads to the simple result for the local significance (eq. 52 in ref. [5]),

$$Z_{\text{Local}} = \sqrt{\lambda}. \quad (5)$$

This does not apply when  $n \neq 1$  but was used [2] in ref. [1] to calculate  $Z_{\text{Local}} = 5.1\sigma$ . In contrast, using eq. (4), we obtain  $Z_{\text{Local}} = 4.1\sigma$  when combining  $n = 6$  channels as in ref. [1]. We validate this result through Monte Carlo (MC) simulations described in section III. Thus to avoid overstating the significance and reaching potentially faulty conclusions, we must take into account that we searched six channels.

### III. GLOBAL SIGNIFICANCE

We previously considered a local significance based on a fixed mass. We should compute a global significance by taking into account every test that was performed and every test that would have been performed were the data different (see problems 1 and 2 in ref. [10] for a pedagogical if partisan discussion). Unfortunately, we don't know the analysis plan of the authors of ref. [1] and thus there are imponderable look-elsewhere effects associated with their choices of statistical tests and datasets. Even if the authors of ref. [1] began with a particular theory of a new scalar boson in mind, many analysis choices could

have been influenced by the observed data. We are not, however, suggesting conscious hacking of any sort and direct readers to the nuanced discussion in ref. [11, 12].

We don't further discuss these issues and instead compute the global significance following the scope of the LEE considered in ref. [1] by taking into account the LEE for the mass range that was searched. In this case, we consider the test-statistic as a random field over the mass range and consider its most extreme fluctuation. Ref. [1] accounted for this LEE by multiplying the local  $p$ -value by a trial factor of 5 [2]. This was an estimate motivated by the 140 GeV – 155 GeV range and the 1.5 GeV and 14 GeV resolutions in the six channels. We instead compute it.

There are rules of thumb for the trial factor based on the range,  $\Delta m$ , and resolution,  $\sigma_m$ , e.g.,

$$N \approx \frac{1}{3} \frac{\Delta m}{\sigma_m} Z_{\text{Local}}. \quad (6)$$

This rule yields a trial factor of about 14, but it was found and tested only for the case of  $n = 1$  channels [13]. Thus rather than using a rule of thumb, we compute it through MC simulations and the Gross-Vitells method (see e.g., ref. [14–16]).

The Gross-Vitells method is a sophisticated treatment based on the expected numbers of up-crossings of random fields above specified thresholds. To apply it, we required the so-called Euler characteristic densities for the  $\frac{1}{2}\chi_6^2$  random field and an estimate of the expected number of up-crossings at a specified level. From theorem 1 and remark 2 in ref. [17], the Euler characteristic densities for the mixture  $\frac{1}{2}\chi_6^2$  are just the same mixtures of the known densities for  $\chi_n^2$  random fields (see theorem 15.10.1 in ref. [14]). To estimate the expected number of up-crossings, we perform simulations using a toy treatment of a resonance problem with five channels with resolution 1.5 GeV and one channel with resolution 14 GeV in a window 140 GeV – 155 GeV. In this toy treatment, we fit the mass and signal strengths of a Crystal ball function on top of a fixed background. The resulting test-statistic should obey the same asymptotic distribution as that in the full treatment. We ultimately find a global significance of about  $3.5\sigma$ , which corresponds to a trial factor of about 12.

To check the asymptotic results for the local and global significances, we perform MC simulations of the test-statistic. We perform 100 000 pseudo-experiments with the toy treatment of the problem described above, computing the test-statistics eqs. (1) and (2) in each case. By simply counting the number of simulations in which the test-statistics exceeded that observed, we again find  $4.1\sigma$  local and  $3.5\sigma$  global significances, validating our previous results.<sup>1</sup> The distribution of the test-statistic and

<sup>1</sup> Note though that this does not validate that the asymptotic limit was valid in the real problem, as our toy treatment used large numbers of expected events in all channels.

the test-statistic at a fixed mass of 151 GeV are shown in fig. 1. We see that the simulations in the latter case closely matches a  $\frac{1}{2}\chi_6^2$  distribution, as expected.

#### IV. SUMMARY

Taking into account the fact that in ref. [1] six signals were fitted rather than one and properly computing the LEE, we find the significances for a new neutral scalar are only  $4.1\sigma$  local and  $3.5\sigma$  global. Thus while this

anomaly may remain intriguing, it isn't as significant as first thought.

#### ACKNOWLEDGEMENTS

I was supported by an NSFC Research Fund for International Young Scientists grant 11950410509. I thank the authors of ref. [1] for time spent clarifying details of their paper.

- 
- [1] A. Crivellin, Y. Fang, O. Fischer, A. Kumar, M. Kumar, E. Malwa, B. Mellado, N. Rapheeha, X. Ruan, and Q. Sha, “Accumulating Evidence for the Associate Production of a Neutral Scalar with Mass around 151 GeV,” [arXiv:2109.02650 \[hep-ph\]](#).
- [2] Private correspondence with the authors of ref. [1].
- [3] A. Fowlie, “Bayesian and frequentist approaches to resonance searches,” *JINST* **14** (2019) P10031, [arXiv:1902.03243 \[hep-ph\]](#).
- [4] A. Fowlie, “Comment on “reproducibility and replication of experimental particle physics results,”” *Harvard Data Science Review* (2021) , [arXiv:2105.03082 \[physics.data-an\]](#).
- [5] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” *Eur. Phys. J. C* **71** (2011) 1554, [arXiv:1007.1727 \[physics.data-an\]](#). [Erratum: *Eur. Phys. J.C* **73**, 2501 (2013). We refer to equation numbers in the published version.].
- [6] S. S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses,” *Annals Math. Statist.* **9** no. 1, (1938) 60–62.
- [7] S. Algeri, J. Aalbers, K. Dundas Morã, and J. Conrad, “Searching for new phenomena with profile likelihood ratio tests,” *Nature Rev. Phys.* **2** no. 5, (2020) 245–252, [arXiv:1911.10237 \[physics.data-an\]](#).
- [8] H. Chernoff, “On the Distribution of the Likelihood Ratio,” *Annals Math. Statist.* **25** no. 3, (1954) 573 – 578.
- [9] S. G. Self and K.-Y. Liang, “Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions,” *J. Am. Stat. Assoc.* **82** no. 398, (1987) 605–610.
- [10] E.-J. Wagenmakers, “A practical solution to the pervasive problems of  $p$  values,” *Psychonomic Bulletin & Review* **14** no. 5, (Oct., 2007) 779–804.
- [11] A. Gelman and E. Loken, “The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time,” (2013) .
- [12] A. Gelman and E. Loken, “The statistical crisis in science: data-dependent analysis – a “garden of forking paths” – explains why many statistically significant comparisons don’t hold up,” *American Scientist* **102** no. 6, (2014) 460–466.
- [13] E. Gross and O. Vitells, “Look Elsewhere Effect,” 2010. Talk at “Statistical issues relevant to significance of discovery claims”.
- [14] R. Adler and J. Taylor, *Random Fields and Geometry*. Springer New York, first ed., 2007.
- [15] E. Gross and O. Vitells, “Trial factors for the look elsewhere effect in high energy physics,” *Eur. Phys. J. C* **70** (2010) 525–530, [arXiv:1005.1891 \[physics.data-an\]](#).
- [16] S. Algeri and D. A. van Dyk, “Testing one hypothesis multiple times: The multidimensional case,” *J. Comput. Graph. Stat.* **29** no. 2, (2020) 358–371, [arXiv:1803.03858 \[stat.ME\]](#).
- [17] J. E. Taylor and K. J. Worsley, “Detecting sparse cone alternatives for gaussian random fields, with an application to fMRI,” *Statistica Sinica* (2013) 1629–1656, [arXiv:1207.3840 \[math.ST\]](#).