Higgs Boson – Digest and Discussion

By Tony O'Hagan

Introduction

On July 10th 2012, I posted the following message to the ISBA mailing list.

Dear Bayesians,

A question from Dennis Lindley prompts me to consult this list in search of answers.

We've heard a lot about the Higgs boson. The news reports say that the LHC needed convincing evidence before they would announce that a particle had been found that looks like (in the sense of having some of the right characteristics of) the elusive Higgs boson. Specifically, the news referred to a confidence interval with 5-sigma limits.

Now this appears to correspond to a frequentist significance test with an extreme significance level. Five standard deviations, assuming normality, means a p-value of around 0.0000005. A number of questions spring to mind.

1. Why such an extreme evidence requirement? We know from a Bayesian perspective that this only makes sense if (a) the existence of the Higgs boson (or some other particle sharing some of its properties) has extremely small prior probability and/or (b) the consequences of erroneously announcing its discovery are dire in the extreme. Neither seems to be the case, so why 5-sigma?

2. Rather than ad hoc justification of a p-value, it is of course better to do a proper Bayesian analysis. Are the particle physics community completely wedded to frequentist analysis? If so, has anyone tried to explain what bad science that is?

3. We know that given enough data it is nearly always possible for a significance test to reject the null hypothesis at arbitrarily low p-values, simply because the parameter will never be exactly equal to its null value. And apparently the LHC has accumulated a very large quantity of data. So could even this extreme p-value be illusory?

If anyone has any answers to these or related questions, I'd be interested to know and will be sure to pass them on to Dennis.

Regards,

Tony

Before going further, I should say that the wording of this message, including the somewhat inflammatory nature of some parts of it, was mine; I was not quoting Dennis Lindley directly. The wording was, though, quite deliberately intended to provoke discussion. In that objective it was successful – I received more than 30 substantive comments in reply. All of these were thoughtful and I learnt a great deal from them. I promised to construct a digest of the discussion. This document is that digest and a bit more – it includes some personal reflections on the issues.

In what follows, particular contributors are occasionally named wherever the context seemed to require, but on the whole I have not attempted to give credit for each piece of information or each viewpoint. A list of contributors is given at the end of this document and most of their contributions can be found on the ISBA website at http://bayesian.org/forums/news/3648.

Has anyone tried to explain?

Probably my most inflammatory comments were in question 2: Are the particle physics community completely wedded to frequentist analysis? If so, has anyone tried to explain what bad science that is?

I already knew of several serious physicists who had been propounding Bayesian ideas within the physics community, two of whom (Bill Jefferys and Giulio d'Agostini) responded to my message. In particular, d'Agostini's article http://arxiv.org/abs/1112.3620 is specifically relevant to the Higgs boson issue.

More to the point, several respondents told me about the Phystat meetings organised by Louis Lyons. The recent one in 2011 was mostly devoted to the very question of 'discovery' and the 5-sigma rule. Several eminent statisticians made presentations. In its programme (http://indico.cern.ch/conferenceOtherViews.py?view=standard&confId=107747) I recognised David Cox, Jim Berger, José Bernardo and David van Dyk. I understand that other statisticians, including Michael Goldstein, Steffen Lauritzen and Persi Diaconis, had been involved in earlier meetings. It is important to say that Phystat may not be a mainstream physics meeting, but it does demonstrate that at least a subset of physicists do engage with statisticians on questions of philosophy and methodology. A couple of weeks after posting my message, I attended the 2012 Maxent meeting (http://www.ipp.mpg.de/ippcms/eng/for/veranstaltungen/konferenzen/maxent2012/). This is a long-standing series of annual conferences which, although it has a maximum entropy focus, also takes Bayesian ideas seriously.

Broadly speaking, it seems that Bayesian methods are prominent in the astrophysics end of the physics community, but at the other end the particle physicists are, if not 'completely wedded', certainly mostly committed to frequentist methods. But it is not because nobody has attempted to convert them to Bayesian thinking, nor would it be at all fair to say that no particle physicists have listened to those arguments. Inevitably, I am sure the great majority of particle physicists do not concern themselves with the foundations of statistics, and simply follow the prevailing view; they were taught frequentist methods at university, all of their colleagues use them and the journals expect them. Nevertheless there are some who do think about such things, and in the particle physics community it seems that the overwhelming majority favour frequentist methods.

Specifically in regard to the Higgs boson, I read that "*There is a bureaucracy that determines in advance the procedures (statistical and otherwise) that will be implemented. Like most bureaucracies, this one is rather conservative and is unlikely to be far out-of-step with the general consensus of its scientific community, a community that is much larger than group that participates in Phystat workshops. So far, the official statistical methods have not been Bayesian*." Whether this bureaucracy is drawn from those who are not concerned about the foundations of statistics and just follow precedent, or whether some of them have given serious thought to the issue is not mentioned.

Particle physicists are exploring the fundamental nature of matter and generally regard science as uncompromisingly objective. I believe that this is misguided, but the claimed objectivity of science is deeply ingrained. Regardless of this particular point, one physicist remarked that "*most find personal Bayesianism decidedly uncongenial as a way of proceeding with a publication with* *hundreds (actually thousands) of signatories.*" For an announcement as momentous as the discovery of the Higgs boson, we are not concerned with the personal judgements of individual scientists. Whilst it may be feasible (and from a Bayesian perspective, the right thing to do) to evaluate the individual physicists' posterior beliefs and assess their degree of consensus, such consensus would surely require very compelling evidence.

It remains to consider just how compelling the evidence is, and in this regard the meaning of 5-sigma is obviously a key factor.

The analysis

Before discussing the nature of 5-sigma, it is important to begin by giving some flavour of the complexity of the computations that were carried out before the announcement was made by the physicists. As might be imagined, this was not just a *t* test for whether a population mean was zero or positive.

Several contributors provided insights into what was actually done. I am sure that to give a full description would require many pages, but here are some of the aspects that I have gleaned. My apologies to everyone if any of this is wrong!

The standard model

The most accepted body of theory in particle physics is the Standard Model (SM). The SM is a complex theory with many ramifications, which in particular predicts the existence, role and properties of the Higgs boson. The official symbol for the Higgs boson is H^0, but at the risk of irritating physicists I will simply call it H.

The statistical analysis which ultimately leads to a 5-sigma result is based on comparing two hypotheses. The null hypothesis is the SM without H, while the alternative hypothesis is the SM with H. One respondent questioned what the null hypothesis could mean if the SM predicts H. I presume that what is meant is that the SM is true *except* in its prediction of H, but there is still a potential problem with this. If the SM is wrong about H, just what bits of it must be changed so that H is no longer a consequence? Probably there are many candidates for such changes. Without explicitly stating what changes are to be applied, the null hypothesis is not properly defined and it is not possible to compute the likelihood of the data under the null.

I think the argument must be that whatever analysis is done under the null hypothesis must use aspects and implications of the SM that are believed to hold notwithstanding the existence or non-existence of H. These other aspects of the SM are not questioned, presumably because they have been confirmed by abundant evidence elsewhere. If H does not exist, the SM will need to be changed in subtle ways that do not affect all that established body of evidence. I wonder to what extend this argument is explicit in the reported analysis?

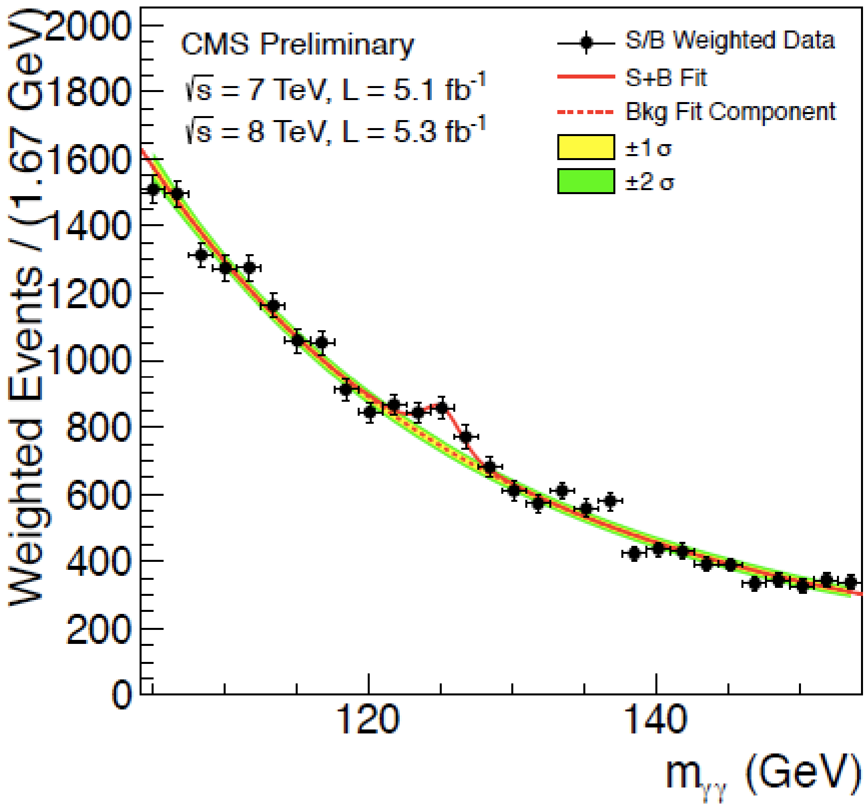
Scale of the work

Huge quantities of data are generated. The Large Hadron Collider (LHC) is "*a machine producing 40 million events a second and recording the outcomes on 150 million detectors – they collect … truly enormous amounts of data (all but the data on about 100 of those 40 million are discarded in real time by server farms which declare most events to be boring).*" A similar comment is that "*although* *the LHC has accumulated enormous amounts of data, the Higgs search is like looking for a needle in a haystack. The final samples of events that are used to look for the Higgs contain only tens to thousands of events*." However, another contributor complained that it was more like looking for a needle in a needle factory, since the data are so full of features and anomalies that it is easy to find something that looks interesting.

One thing is sure, that the scale of the work in terms of people, time and money has also been huge. Although computers screen out most of the data, hundreds of people have been involved in processing the data that eventually led to the announcement.

The model

After what is evidently a lot of sophisticated pre-processing and modelling, the data eventually end up as something like the graph below (which I am assured is in the public domain).



I have not had a detailed explanation but the graph seems to plot counts of 'events' at different levels of energy (or mass), measured in GeV on the X axis. The SM without H predicts a smooth curve of general form

$f(x) = \exp(a_0 + a_1x + a_2x^2)$

which is called the background. The SM with H predicts that a 'bump' is added to this background whose shape is a Gaussian. The model has 6 parameters, a_0, a_1, a_2, plus the location, width and amplitude of the bump. If we just assume Poisson distributions then there are lots of significant deviations from the fitted curve, but the bump around 125 GeV is the one that is so significant it hits the 5-sigma criterion.

Actually, the community insists on replication, and there have apparently been two such experiments in which the H has been 'found' at the 5-sigma level. Furthermore, the two experiments were performed at two different particle beam energies and involved two different kinds of event.

H is predicted by the SM to be extremely short lived and to decay almost immediately. It can decay into a pair of photons or a pair of Z bosons (or other things, I think), and it is these two forms of decay that have been seen. The above graph relates to the two-photons decay. I'm not sure exactly how, but this is the essence of the analysis as I understand it. Of course, there are many questions that arise in a statistician's mind on seeing something like this, and the following discussion is around those to the extent that I can follow it.

Why 5 sigma?

I was given many answers to this question. One was the point that I had raised in my message – the consequences of getting it wrong are dire.

The credibility of the scientists is at stake. As one respondent put it: "*Given that the search for the Higgs took some 45 years, tens of thousands of scientists and engineers, billions of dollars, not to mention numerous divorces, huge amounts of sleep deprivation, tens of thousands of bad airline meals, etc., etc., we want to be sure as is humanly possible that this is real*." The search for H is of huge significance in particle physics and a loss of credibility could render all that effort wasted if it meant withdrawal of further funding.

Announcements at 3-sigma or more in the past have proved wrong. The community has responded by raising the bar to its present level of 5-sigma. Even here, I was told of "*a nice example in the literature of a discovery of the 'pentaquark' whose existence was detected at the 5.6 sigma level, according to the paper. The authors had later to retract the claim even though the area beyond 5.6 is about 10^{-8}. The analysis was all very much derided by the physicists I have been talking to – lots of things which the physicists regard as rookie mistakes. But it shows, I think, one source of the extraordinarily conservative P-value rules they have developed*. "

Another example is the recent experiment at CERN which claimed that neutrinos could travel faster than the speed of light, which turned out to be caused by a fault in a cable.

That leads into the next reason for 5-sigma. So much can go wrong in the analysis that it makes sense to guard against false positives caused by errors in underlying assumptions, pre-processing, experimental controls, etc.

The other major reason is associated with multiple testing. "*They then carry out searches for many objects whose existence is somewhere from quite unlikely to very likely. The Higgs was in the latter category I suppose. The fact that they want to avoid getting all stirred up over unimportant coincidences in 10^15 events motivates them to apply this very ad hoc multiple comparisons control. But more importantly when they search for something like the Higgs they look at a sequence of possible masses m -- say running from 50 GeV to 600 Gev at steps of less than 1 GeV (they report* *masses in units of energy quoted in GeV usually). As they search they carry out tests for the existence of Higgs at each mass in the sequence. They plot P-values against the mass and look for dips to very low levels. The procedure is a test at each mass m on a Poisson count (or set of Poisson counts) of the null hypothesis that there is no such thing as a Higgs boson with mass m. That null is translated to a null hypothesis about the set of Poisson means*."

A similar issue is what they call the Look Elsewhere Effect (LEE). In the words of one respondent: "*The 'Look Elsewhere Effect'. We are worried about the chance of a statistical fluctuation mimicking our observation, not only at the given mass of 125 GeV but anywhere in the spectrum. The quoted p-values are 'local' i.e. the chance of a fluctuation at the observed mass. Unfortunately the LEE correction factor is not very precisely defined, because of ambiguities about what is meant by 'elsewhere'*."

In a report at http://www.technologyreview.com/view/428428/higgs-boson-may-be-an-imposter-say-particle/ it is claimed that the observed effect may indeed be caused by something else – "*a generic Higgs doublet or a triplet imposter*. "

Yet another kind of multiple testing is the 'Keep Looking Bias' (http://www.science20.com/quantum_diaries_survivor/blog/keeplooking_bias). If the physicists find a result that reaches 3 sigma, then they keep looking and keep testing until eventually they reach 5 sigma. They are 'sampling to a foregone conclusion', hoping eventually to reach the required standard of evidence. Of course, they may never reach it and the funding agencies may give up waiting, so it is not really a foregone conclusion, but nevertheless this kind of multiple testing is something that in frequentist statistics causes all kinds of heart-searching and complicated analysis in clinical trials.

Discussion

So here are some of my own views on this.

There are good reasons for being cautious and demanding a very high standard of evidence before announcing something as momentous as H. It is acknowledged by those who use it that the 5-sigma standard is a fudge, though. They would surely be willing to make such an announcement if they were, for instance, 99.99% certain of H's existence, as long as that 99.99% were rigorously justified. 5-sigma is used because they don't feel able to quantify the probability of H rigorously. So they use the best statistical analysis that they know how to do, but because they also know there are numerous factors not taken into account by this analysis – the multiple testing, the likelihood of unrecognised or unquantified deficiencies in the data, experiment or statistics, and the possibility of other explanations – they ask for what on the face of it is an absurdly high level of significance from that analysis.

To do better would be a very demanding task. First, to be able to give a probability for H's existence requires a Bayesian analysis, which means embracing a prior probability, and also prior distributions on the parameters of the background and the signal. A range of choices for those would have to be considered in order to ensure that the certainty of H's existence at the end is uncontroversial. A benefit of a Bayesian analysis is that it would avoid the well-documented problems with p-values as measures of the strength of evidence, but a Bayesian analysis would not be easy to do.

The hardest part is to quantify all the things that could go wrong, which we can consider generically as model errors. It is so often said that 'all models are wrong' that it is something of a cliché, but very little consideration has been given to the implications. I have recently been working on this issue and, I am pleased to say, so have others. In particular I have contributed a discussion for a paper written by Stephen Walker to be published in the *Journal of Statistical Planning and Inference* on this subject.

Having said that to do better than the 5-sigma fudge would be a very demanding task, I see no reason why it should not be attempted. The discovery of H would have been a great opportunity. Of the hundreds or thousands of people working on that project, including no doubt in some way all of the top particle physicists and many other leading figures in physics, engineering, etc., how many were statisticians? And how many of those were top-flight statistics researchers?

I wrote a follow-up to my original message when it was clear how many people had taken the trouble already to respond, including physicists. In it, I made it clear that I was delighted that those people had risen to my rather crude insinuations. I wrote:

Incidentally, I am delighted to have received a number of responses from physicists. Some of them take me to task for the provocative style of my original message. I know how important it is to respect the skills and expertise of people in other professions, but I also did want to generate discussion - in which objective I seem to have been successful. I am genuinely grateful to all those who have taken the time to give me the benefit of their knowledge and wisdom.

Respect for the skills of others cuts both ways. I may be maligning the physicists again, but my experience is that in most fields where statistics is needed it is generally done by non-statisticians. By that I mean people who are not primarily trained as statisticians and whose employment is not mainly as statisticians. I do not wish to suggest that a physicist, or an ecologist or whatever, who has taken an interest in statistics and studied statistical methods in order to apply them in their discipline is necessarily incompetent as a statistician. What bothers me is the commonly prevailing notion that a fully trained, specialist, professional statistician who has spent a lifetime honing their skills in their chosen field has nothing to teach them.

The discovery of H is big. The people involved would not dream of failing to use the very best possible physicists, not to mention the very best possible engineers to build their equipment, the very best possible computers, etc. But I'm willing to bet they didn't consider it necessary to seek the advice, throughout every stage of working with their data, of the very top professional statisticians. Yet I would be very surprised if those top statisticians would not have welcomed the opportunity to work on such a high-profile, exciting and challenging project, in the same way that top physicists would jump at the chance to be involved.

It is simply not possible for us to do this now from outside the project. First, we do not have access to the data, and I'm sure will not be allowed access to it. Second, it's a big job and should have had man-years of high-level statistical expertise devoted to it, like the very many man-years of effort of other top scientists.

In short, from the perspective of statistics, H looks to me like an opportunity missed. And so particle physicists will continue to use their 5-sigma rule without any real understanding of whether it does protect them (or over-protect them) against the feared slings and arrows of outrageous fortune.

List of contributors

Simon Jackman, Bill Jefferys, Murray Clayton, Rob Weiss, David Kaye, Jim Linnemann, Frank Lad, David van Dyk, Luis Manuel Sarro Baro, Richard Lockhart, David Cox, Peter Craig, Eric-Jan Wagenmakers, Ashley Ford, Jeff Morris, Giulio d'Agostini, Louis Lyons, David Spiegelhalter (see his blog at http://understandinguncertainty.org/explaining-5-sigma-higgs-how-well-did-they-do), Larry Wasserman, Prasanta Bandyopadhyay, Harrison Prosper, Deborah Mayo (who also has an interesting blog at http://errorstatistics.com/2012/07/11/is-particle-physics-bad-science/), Jørgen Hilden, Peter Thall, John Pratt, Herman Rubin, Peter Diggle, Tom Leonard, Nozer Singpurwalla, Udo von Toussaint, Valen Johnson.

My apologies to people in this list whose comments I have not explicitly included in the above digest. In the interests of keeping it reasonably short and to the point, and also in the interests of not using up an inordinate amount of my own time, I have been somewhat selective. If there are strands that I have ignored which you feel should be aired more openly, then I invite you to start your own discussion thread!