UPPER LIMITS IN THE CASE THAT ZERO EVENTS ARE OBSERVED: AN INTUITIVE SOLUTION TO THE BACKGROUND DEPENDENCE PUZZLE

P. Astone and G. Pizzella
Sezione INFN Roma 1 (“La Sapienza”), Università “Tor Vergata” and LNF Frascati. Italy

Abstract
We compare the “unified approach” for the estimation of upper limits with an approach based on the Bayes theory, in the special case that no events are observed. The “unified approach” predicts, in this case, an upper limit that decreases with the increase in the expected level of background. This seems absurd. On the other hand, the Bayesian approach leads to a result which is background independent. An explanation of the Bayesian result is presented, together with suggested reasons for the paradoxical result of the “unified approach”.

1. INTRODUCTION

The study of a new phenomenon in science often ends up in a null result. However it might be of great importance to set upper limits, as this will help our understanding by eliminating some of the theories proposed.

The determination of upper limits is presently a hotly debated issue in several fields of physics. Many papers have been devoted to this problem and different solutions have been proposed. In particular the problem has been discussed in paper [1] (“unified approach”) and, more recently, in papers [2, 3], based on the Bayes’ theory. The use of the “unified approach” (FC) to set upper limits or confidence intervals is recommended by the PDG [4]. The “unified” and the Bayesian approaches are very different, not only in the sense that they lead to different numerical results but more radically in the meaning they attribute to the quantities involved. These differences lead to intrinsic problems in any comparison of their separate results. The purpose of this letter is to try to throw some light on this contentious and important issue. We shall show that the Bayesian approach is the correct one. If our argument is accepted by the scientific community, many debates about upper limits will be clarified.

2. THE BACKGROUND DEPENDENCE PUZZLE

According to the (FC) “unified approach” the upper limit is calculated using a revised version of the classical Neyman construction for confidence intervals. This approach is usually referred to as the “unified approach to the classical statistical analysis”, and it aims to unify the treatment of upper limits and confidence intervals. On the Bayes side, according to [2, 3], the upper limit may be calculated using a function $\mathcal{R}$ that is proportional to the likelihood. This function is called the ”relative belief updating ratio” and has already been used to analyse data in papers [5, 6]. The procedure has been extensively described by G. D’ Agostini in [7].

Comparison between the two approaches is difficult for the general case. But we have noticed a special case which is easier to discuss. In this case the greater efficacy of one approach compared to the other one seems clear. This case is when the experiment gave no events, even in the presence of a background greater than zero.

When there are zero counts, the predictions obtained with the two methods are different and both are -intuitively- quite disturbing. Our intuition would, in fact, be satisfied by an upper limit that increases with the background level, and this is, in general, the case when the observation gives a number of events of the order of the background. However, when zero events are observed, the “unified approach” upper limit decreases if the background increases (a noisier experiment puts a better upper limit than a less noisy one, which seems absurd) while the Bayesian approach leads to the predictions that a constant upper limit will be found (the upper limit does not depend on the noise of the experiment). Various papers[8, 9, 10] have been devoted to the problem of solving some intrinsic difficulties with the ”unified” approach: in particular to solving the problem of ”enhancing the physical significance of frequentist confidence intervals”[8], or to imposing ”stronger classical confidence limits”[9]. In this latter article the proposed method ”gives limits that do not depend on background in the case of no observed events” (that is the Bayesian result !).

In what follows we will give an explanation for the two results.

We remind the reader that the physical quantity for which a limit must be found is the events rate (i.e. a gravitational wave burst rate) r. Here we will assume stationary working conditions. For a given hypothesis r, the number of events which can be observed in the observation time T is described by a Poisson process which has an intensity equal to the sum of that due to background and that due to signal.

In general, the main ingredients in our problem are that:

- we are practically sure about the expected rate of background events $r_b = n_b/T$ but not about the number of events that will actually be observed (which will depend on the Poissonian statistics). T is the observation time;
- we have observed a number n_c of events but, obviously, we do not know how many of these events have to be attributed to background and how many (if any) to true signals.

Under the stated assumptions, the likelihood is

$$f(n_c \,|\, r, r_b) = \frac{e^{-(r+r_b)\,T}\,((r+r_b)\,T)^{n_c}}{n_c!}\,, \tag{1}$$

We will now concentrate on the solution given by the Bayesian approach.

The “relative belief updating ratio” $\mathcal{R}$ is defined as:

$$\mathcal{R}(r; n_c, r_b, T) = \frac{f(n_c \,|\, r, r_b)}{f(n_c \,|\, r=0, r_b)}\,, \tag{2}$$

This function is proportional to the likelihood and it allows us to infer the probability that rT signals will be observed for given priors (using the Bayes’s theorem).

Under the hypothesis $r_b > 0$ if $n_c > 0$, $\mathcal{R}$ becomes

$$\mathcal{R}(r; n_c, r_b, T) = e^{-r\,T}\left(1 + \frac{r}{r_b}\right)^{n_c}\,. \tag{3}$$

The upper limit, or -more properly- ”standard sensitivity bound” [7], can then be calculated using the $\mathcal{R}$ function: it is the value r_{ssb} obtained when

$$\mathcal{R}(r_{ssb}; n_c; r_b; T) = 0.05 \tag{4}$$

We remark that 5% does not represent a probability, but is a useful way to put a limit independently of the priors.

Eq. 3 when no events are observed, that is, when n_c=0, becomes:

$$\mathcal{R}(r) = e^{-rT} \tag{5}$$

Thus putting $n_c = 0$ in Eq. 4 we find $r_{ssb} = 2.99$, independently of the value of the background n_b.

We will not describe the well known (FC) procedure here, but we would just observe that, according to this procedure, for $n_c = 0$ and $n_b = 0$, the upper limit is 3.09 (numerically almost identical to the Bayes' one) but it decreases as n_b increases (e.g. for $n_c = 0$ and $n_b = 15$ the upper (FC) limit at 95% CL is 1.47).

In an attempt to understand such different behaviour we will now discuss some particular cases. Suppose we have $n_c = 0$ and $n_b \neq 0$. This certainly means that the number of accidentals, whose average value can be determined with any desired accuracy, has undergone a fluctuation. The larger the n_b values, the smaller is the $a\ priori$ probability that such fluctuations will occur. Thus one could reason that it is less likely that a number n_{gw} of real signals could have been associated with a large value of n_b, since the observation gave $n_c = 0$.

According to the Bayesian approach, instead, one cannot ignore the fact that the observation $n_c = 0$ has already being made at the time the estimation of the upper limit comes to be calculated. The Bayesian approach requires that, given $n_c = 0$ and $n_b \neq 0$, one evaluates the $chance$ that a number n_{gw} of signals exists. This $chance$ of a possible signal is applied to the observation that has already been made.

Suppose that we have estimated the average background with a high degree of accuracy, for example n_b=10. In the absence of signals, the a priori probability of observing zero events, due just to a background fluctuation, is given by

$$f_n = f(n_c = 0|n_b = 10) = e^{-n_b} = 4.5 \cdot 10^{-5} \tag{6}$$

Now, suppose that we have measured zero events, that is n_c=0. In general $n_c = (n_b + n_{gw})$. It is now nonsense to ask what the probability that n_c=0 is, since the experiment has already been made and the probability is 1.

We may ask how the a priori probability would be changed if n_{gw} signals were added to the background. We get

$$f_{sn} = f(n_c = 0|n_b = 10, n_{gw}) = e^{-(n_b+n_{gw})} \tag{7}$$

It is obvious that f_{sn} can only decrease relative to f_n, since we are considering models in which signal events can only add to noise events[1].

The right answer is guaranteed if the question is well posed. Given all the previous comments, the most obvious question at this point is: **what is that signal n_{gw} which would have reduced the probability f_n by a constant factor, for example 0.05 ?**

$$f_{sn} = f_n \cdot 0.05 = e^{-n_b} \cdot e^{-n_{gw}} \tag{8}$$

Using Eqs. 6, 7 and 8 the solution is:

$$e^{-n_{gw}} = 0.05 \tag{9}$$

that is:

$$n_{gw} = 2.99 \tag{10}$$

[1]In a gravitational wave experiment signals may add up to the noise with the same phase, thus increasing the energy of the combined effect, or with a phase opposite to that of the noise, thus reducing the energy. They can in particular add up also to noise events, even if we expect this to happen with a very low probability, as we know that the events due to the signal are very "rare" compared to the events due to the noise.

Anyway, in principle, the presence of this fact will lead to the prediction of a signal rate that increases with the background: in fact the probability that one background event be cancelled by a signal event increases, as n_b increases. Thus, if we, at least in part, attribute the observation of n_c=0 to a cancellation of background events due to the signal the final limit on r should increase.

In the modelling we usually, as reasonable, consider this effect be negligible. If this is not the case then it must be properly modelled in the likelihood.

Now suppose another situation, n_b=20, thus $f_n = 2.1 \cdot 10^{-9}$. Repeating the previous reasoning we still get the limit 2.99.

The meaning of the Bayesian result is now clear: we do not care about the absolute value of the a priori probability of getting $n_c = 0$ in the presence of noise alone. The observation of $n_c = 0$ means that the background gave zero counts by chance. Even if the a priori probability is very small, its value has no meaning once it has happened. The fact that the single background measurement turned out to be zero, either due to a zero average background or due to the observation of a low (a priori) probability event, must not change our prediction concerning possible signals.

For $n_c = 0$ we are certain that the number of events due to the background is zero. Clearly this particular situation gives more information about the possible signals. In the case $n_c \neq 0$, instead, it is not possible to distinguish between background and signal. The mathematical aspect of this is that the Poisson formula when $n_c = 0$ reduces to the exponential term only, and thus it is possible to separate the two contributions, of the signal (unknown) and of the noise (known).

We note that the different behaviour of the limit in the unified approach is due to the non-Bayesian character of the reasoning. In such an approach an event that has already occurred is considered "improbable": given the observation of $n_c = 0$ they still consider that the probability

$$f_{sn} = f(n_c = 0|n_b, n_{gw}) = e^{-(n_b+n_{gw})} \tag{11}$$

decreases as n_b increases. As a consequence they deduce that to a larger n_b corresponds a smaller upper limit n_{gw}.

Given the previous considerations, we must now admit that our intuition to expect an upper limit that increases with increasing background, even when $n_c = 0$, was wrong. We should have expected to predict a constant signal rate, as a consequence of the observation of zero events, independently of the background level.

3. CONCLUSION

We have compared the upper limits obtained with the (FC) "unified" and with the Bayesian procedures, in the case of zero observed events.

We believe that the greater efficacy of the Bayesian approach compared to the (FC) method, demonstrated for the case $n_c = 0$, is a strong indication that the Bayesian method -natural, simple and intuitive- is the correct one. Thus we agree with the proposal in[7] that this method should be adopted by the scientific community for upper limit calculations (see, for example, [11] on upper limits in gravitational wave experiments).

References

[1] G.J. Feldman and R.D. Cousins, *Unified approach to the classical statistical analysis of small signal*, Phys.Rev.D 57, 3873 (1998)

[2] G. D' Agostini, *Contact interaction scale from Deep Inelastic scattering events- what do the data teach us?*, ZEUS note 98-079, November 1998.

[3] P. Astone and G. D'Agostini, *Inferring the intensity of Poisson processes at the limit of the detector sensitivity*, CERN-EP/99-126 and hep-ex/9909047

[4] C. Caso et al., *Review of particle physics*, *Eur. Phys. J.* **C3** (1998) 1 (http://pdg.lbl.gov/)

[5] ZEUS Collaboration, J. Breitweg et al., *Search for contact interactions in Deep-Inelastic $e^+p \to e^+X$ scattering at HERA*, DESY Report 99-058, `hep-ex/9905039`, May 1999, to be published in Eur. Phys. J. C.

[6] G. D' Agostini and G. Degrassi, *Constraints on the Higgs boson mass from direct searches and precision measurements*, internal report DFPD-99/TH/02, `hep-ph/9902226`. Feb. 1999, to be published in Eur. Phys. J. C.

[7] G. D'Agostini, *Confidence limits: what is the problem? is there the solution?*, Proceedings of the Workshop on "Confidence Limits", CERN, 17-18/1/2000.
http://www.cern.ch/CERN/Divisions/EP/Events/CLW/

[8] C. Giunti, *Enhancing the physical significance of frequentist confidence intervals*, Proceedings of the Workshop on "Confidence Limits", CERN, 17-18/1/2000.
http://www.cern.ch/CERN/Divisions/EP/Events/CLW/

[9] G. Punzi, *Stronger classical confidence limits*, Proceedings of the Workshop on "Confidence Limits", CERN, 17-18/1/2000.
http://www.cern.ch/CERN/Divisions/EP/Events/CLW/ and hep-ex/9912048

[10] M. Woodroofe, *On the problem of low counts in a signal plus noise model*, Proceedings of the Workshop on "Confidence Limits", CERN, 17-18/1/2000.
http://www.cern.ch/CERN/Divisions/EP/Events/CLW/

[11] P. Astone and G. Pizzella, *"On upper limits for gravitational radiation"*, gr-qc/0001035. 12 Jan 2000

Discussion after talk of Pia Astone. Chairman Peter Igo-Kemenes.

Don Groom

If you expect 10 background events and observe zero, wouldn't you conclude simply that something was broken?

P. Astone

Clearly the numbers I have used in the example are very high, but an event of very low probability might have occurred. I simply wanted to force you to understand my point, even with a rather extreme situation. In any case it is very easy to get zero events, when the estimated background is, for example, two or three, and the meaning of my example is still the same. I work on gravitational wave detection, and we are now analysing the data of five detectors in coincidence.It really will not surprise me if, in the presence of a background different from zero, we will measure zero coincidence events, as happened many times in the past.

As a general comment, I want to say that I am assuming here that your estimation of the background has been well done. It is obvious that we have to control and check the behaviour of our detector and the procedure to estimate the background!

Alex Read

You said twice something about the expectation of an upper limit that increases with the average background. Don't you mean, decreases ?

P. Astone

It depends on the assumed working condition. Assume that the observation n_c is roughly equal to the background n_b (and this is the situation that usually happens), and that n_c and n_b are both much higher than the signal. In this case the upper limit increases as $\sqrt{n_b}$, in the Bayesian approach. In this case we have assumed that n_c increases as the background increases. In contrast, if n_c is fixed, and the background increases, then I agree with you, the upper limit decreases.

A. Read

Yes, in fact the second case was the thing I was thinking about because you concentrated most of your talk about zero counts and if you had zero counts and after the experiment you increase the background, the limit improves.

C. Giunti

I don't know if I understood well, but you said, for example, if you measure zero events, you want that the upper limit in the case of 10 or 20 expected background events is the same. This is what happens also in the unified approach. If you increase the background the upper limit remains constant, practically. It goes down only in the beginning. For example, if you measure zero events, the upper limit decreases from one to two expected background events, but above that it remains practically constant.

P. Astone

It increases, even if the increase is not very high. In the first transparency I gave a numerical result. For example, if the background is 15, the upper limit decreases from 3 to 1.47. Numerically it is really not so important, but I said that I am not interested in comparing the numbers. What is important is not to get a limit of 3 or 2 as we are interested in the order of magnitude of things, so from my point of view, 3 or 2 is the same. But I tried to understand the meaning of these different behaviours and the reasoning that lead to the two results. What I found out is that the difference is due to the non-Bayesian character of the Feldman-Cousins reasoning: they consider still improbable, an event that has already occurred. On the contrary, using the Bayesian approach, the absolute value of the a priori noise probability is not important, but what is important is how it is rescaled once you suppose that the signals do exist.

Gunter Zech

I think this likelihood ratio is a rather rational approach, and it is neither Bayesian nor classical, but once you cut at a certain value this is somehow arbitrary because, I mean, if you have a long tail or not in the in the likelihood, and cut at 5%, it makes a difference. So all these kinds of approaches have their problems. In the Poisson case it would be interesting to see if you get a different limit, if you integrate the Bayesian way or if you cut at a corresponding value of the likelihood. Is there a difference?

G. d'Agostini

Perhaps I can comment. First of all, in the cases we are interested in, the likelihood is really steep on the right side. Usually that is the case. As I always say, if it is not the case, publish the likelihood, our $\mathcal{R}$ and so on (for example the log-likelihood). We have done an exercise to see what happens if you try to integrate the likelihood in the sense that you assume a flat prior, and the order of magnitude is the same. We have a table of comparison in our paper. We also give a justification of this uniform prior. It is the prior that gives the same results - really they are almost identical results - that you would get from a prior which reflects the positive attitude of experimentalists, who are not losing time and money, but they do research because they hope to see something. If you plug in these kinds of priors, you get essentially the same answer as a uniform prior.