ON THE STATISTICAL SIGNIFICANCE OF EXCESS EVENTS - REMARKS OF CAUTION AND THE NEED FOR A STANDARD METHOD OF CALCULATION

R. Staubert
Astronomisches Institut der Universität Tübingen, Tübingen, Germany

ABSTRACT

Methods for calculating the statistical significance of excess events and the interpretation of the formally derived values are discussed. It is argued that a simple formula for a conservative estimate should generally be used in order to provide a common understanding of quoted values.

1. Introduction. Substantial nonuniformity exists in the cosmic ray literature with respect to how the statistical significance of features or excess events is being calculated (e.g. point sources, spectral lines, light curves). Consequently, there is no mutual understanding about what the confidence in some result might really be when a number of 'standard deviations' are being quoted. Some of the proposed procedures for calculation need to be taken with caution. On the other hand, there is a clear need for the adoption of a standard method to allow the reliable intercomparison of quoted results and create a common understanding of the associated confidence.

A number of methods and formulae have been proposed together with sometimes extended mathematical derivation or justification (Ref. 1-4). It has become clear however, that some of these methods need to be taken with caution. On the other hand there is a very simple formula which is being widely used by X-ray astronomers providing a common understanding.

2. Statistical Significance. An example for the statistical situation which we like to discuss is given in Figure 1.

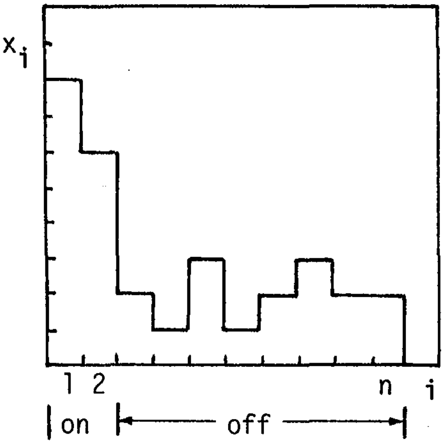

Fig. 1. Statistical example.

Numbers of events x_i are plotted versus bin number $i = 1....n$, corresponding to intervals of some physical variable (e.g. energy, phase, electric charge, time, ...). In the example given there seem to be 'excess events' in bins 1 and 2 as compared to the 'background' defined by the other bins. The excess is

$$\mathrm{ON} - \alpha\,\mathrm{OFF},$$

when ON and OFF are the integrated counts in channels 1 to 2 and in channels 3 to n, respectively and α is the ratio of the corresponding number of bins, here $2/(n-2)$.

The general questions then are:
1. Does the excess correspond to the presence of a physical signal?
2. What is the 'significance of the signal'?

It is important to distinguish between these two questions. They correspond to the assumption that one out of two alternative hypothesis is true:
- the null hypothesis H0 is, that there really is only background,
- the hypothesis H1 is, that a true signal exists in addition to background.
When a statement is made about a statistical situation, it should be clear under which hypothesis this statement holds.

The first of the two questions may be answered by giving the probability for a chance occurance of the observed excess by a statistical fluctuation (under H0). It is of course necessary to use the proper statistic (e.g. binomial statistic for small numbers of events). If a low probability for the chance occurance of 'excess events' is found, it is then usually concluded that the presence of a 'physical signal' is likely. From there on hypothesis one is advocated and all statements made should refer to H1.

Only under H1 the term significance should be used. In particular the often used formula $(ON - \alpha\, OFF)/\ \alpha \sqrt{OFF}$ is useless (as are a number of other formulae, see e.g. (4)). Also the probability which answers the first question should not be converted into a significance (as is sometimes done by using the integrated Gaussian distribution, even in cases where the Gaussian statistic does not apply).

In answering the second question then the presence of a signal is assumed (H1). The 'significance of the signal' k can be defined as the ratio of the best estimate of the signal to its uncertainty. In the case of Poissonian counting statistic for which the variance is equal to the mean a straightforward error propagation leads to the well known formula (in terms of the above defined variables):

$$\text{'significance'} \quad k = \frac{ON - \alpha\, OFF}{\sqrt{ON + \alpha^2 OFF}} \qquad [1]$$

in units of standard deviations σ (see Ref. (3,5,6,7); note that in (3) this formula is interpreted incorrectly). Formula [1] may be also derived by using the more complicated maximum likelyhood ratio (6).

A general criticism of the work of (3) and to some extent of (1) and (2) is given in (6). While it is very important, not to overestimate statistical significance, Ref. (3) does too much, leading to an underestimate.

More recently, (5) has contributed significantly to the confusion by trying to show that formula [1] is incorrect and should be replaced by another complicated formula. The main argument is that the new formula fits much better to Monte Carlo simulations than formula [1] does. The whole discussion is misleading and suffers from the fact that no distinction between H1 and H0 is made: while formula [1] refers to H1 the Monte Carlo simulations as well as the new formula refer to H0, so their distributions are necessarily different.

For the example given in Fig. 1 (with a unit of 1 for the scale of counts x_i) the two questions can be answered as follows:

1. The probability (under H0) for a chance occurance of 14 events in bins 1 and 2 whith an average rate of 6 in two bins is $\lesssim$10-3, using binomial statistic (note that Poissonian statistic gives the somewhat larger probability of 3.6 x 10-3).
2. If one feels that the probability of 10-3 is low enough to postulate the existence of a physical signal (H1), then the significance of this signal is

$$k = \frac{14 - (2/10)\ 16}{\sqrt{14 + (2/10)^2\ 16}} = 2.6 \text{ standard deviations.}$$

To put it in other words again we consider Figure 2.

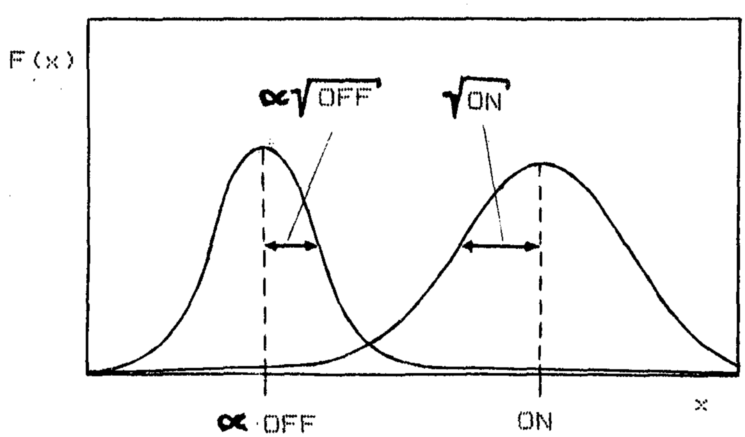

Figure 2. Representation of event number distributions.

If the signal ON $-\ \alpha$ OFF is compared to the standard deviation of the background $\alpha\sqrt{\text{OFF}}$, one gets an estimate for the chance occurance under the null hypothesis H0.
If, on the other hand, ON $-\ \alpha$ OFF is compared to both the standard deviations of the background and the signal, as is done by formula [1] under H1, one gets a different estimate. This is related to the probability that a second measurement (under identical conditions) will lead to a null result (ON $\leqslant\ \alpha$ OFF). It is this estimate that should be called 'significance of the detected signal'.

3. Final remarks

Values of significances in units of standard deviations are usually quoted when the detection of some signal is claimed. Consequently, a formula refering to H1 (existence of a signal) should be used.

Formula [1] has been widely adopted by X-ray astronomers and has as such served successfully as a standard allowing the reliable intercomparison of stated values of significance. It is up to the individual from what level of significance onward one starts to 'believe' in some reported result. Our personal view is that using formula [1] a minimum significance of 3 standard deviations (better yet 5) should be reached.

References

1. Hearn, D. 1969, Nucl.Instr.Methods 79, 200
2. O'Mongain, E. 1973, Nature 241, 376
3. Cherry, M.L. et al. 1980, Ap.J. 242, 1257
4. Li, T.P. and Ma, Y.Q. 1983, Ap.J. 272, 317
5. Li, T.P. and Ma, Y.Q. 1983, Proc.18th Int.Cosmic Ray Conf., Bangalore, XG 3-7, p.76
6. Nolan, P.L. 1982, preprint UCSD SP-82-15
7. Peterson, L. 1975, ARAA 13, 423