Criterion for the rejection of doubtful observations

McIntosh, B. A.

Publisher's version / Version de l'éditeur:
https://doi.org/10.4224/21276056
Report (National Research Council of Canada. Radio and Electrical Engineering Division : ERB), 1968-12

ERB-803

UNCLASSIFIED

NATIONAL RESEARCH COUNCIL OF CANADA

RADIO AND ELECTRICAL ENGINEERING DIVISION

CRITERION FOR THE REJECTION OF DOUBTFUL OBSERVATIONS

- B. A. McINTOSH -

OTTAWA

DECEMBER 1968

ABSTRACT

It is always difficult to justify the rejection of a doubtful value from a series of measurements. Provided the total number of observations is large, a fluctuation of fairly large magnitude is tolerable in terms of its effect on the average. Objective rejection criteria, past and present, are reviewed and discussed.

CRITERION FOR THE REJECTION OF DOUBTFUL OBSERVATIONS

– B.A. McIntosh –

The exact title of this paper appears as the heading of a section in a book written by William Chauvenet [1] over 100 years ago. The so-called Chauvenet criterion is still in use by scientists and, furthermore, the basic problem indicated by the title is still a valid subject for modern statistical research. The following notes review some of the past and present approaches to the problem and add the author's own personal views. This is by no means a literature survey.

In a series of measurements of a quantity, one particular reading may have a value widely divergent from all the others. If we retain it, will it not have an undue influence on the magnitude of the average value we obtain? But do we have any objective criterion for discarding it? Before going further it is instructive to investigate quantitatively the significance of the expression 'undue influence on the mean'. Consider a set of readings X_1, X_2, X_3, ..., X_n whose mean $\overline{X}_n$ is determined in the normal way

$$\overline{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i \ .$$

The $n+1$st reading is inconsistent and we are interested in the deviation

$$x = X_{n+1} - \overline{X}_n \ .$$

The average of all $n+1$ readings is

$$\overline{X}_{n+1} = \overline{X}_n + \frac{x}{n+1} \ .$$

The effect of the deviation x on the mean is reduced by $\frac{1}{n+1}$, so that, if n is large, the *influence* of the $n+1$st reading may not be very great.

A second parameter of any error distribution is the standard deviation which we choose to define by

$$S_n^2 \equiv \frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}{n-1} \ .$$

Then adding one more reading gives

$$S_{n+1}^2 = S_n^2\left[1 - \frac{1}{n} + \frac{1}{n+1}\left(\frac{x}{S_n}\right)^2\right] .$$

Again the effect of the additional reading is reduced by $\frac{1}{n+1}$.

We note that the standard error of the mean is $S_n/\sqrt{n}$. Suppose we require that the change in the mean caused by the additional reading be less than the standard error in the mean. That is,

$$\frac{x}{n+1} < \frac{S_n}{\sqrt{n}}$$

or

$$\frac{x}{S_n} < \sqrt{n} + \frac{1}{\sqrt{n}} .$$

Although the physicist is accustomed to quote his error as $S_n/\sqrt{n}$, the statistician will more frequently quote a 5% confidence interval which is approximately $2S_n/\sqrt{n}$. Thus an additional point will not change the value of a mean beyond its normally accepted error limits if only

$$\frac{x}{S_n} \lesssim \sqrt{n} .$$

This is not a very stringent condition and would permit practically no rejection of points for large n, since for example, if $n = 100$, acceptable deviations could run as high as $10S_n$.

The foregoing discussion shows that the rejection of a seemingly inconsistent reading probably does not increase the statistical reliability of an average by very much.

Worthing and Geffner [2] state Chauvenet's criterion as follows: '... any reading of a series of n readings shall be rejected when the magnitude of its deviation from the mean of the series is such that the probability of occurrence of all deviations that large or larger does not exceed $1/2n$.' This seems to be a wholly arbitrary definition. Why is $1/2n$ a critical value? When we go back to the original book by Chauvenet we find that he credits Professor Peirce with the criterion and the derivation of it. Peirce's statement of it, as quoted by Chauvenet, seems less arbitrary, '... the proposed observations should be rejected when the probability of the system of errors obtained by retaining them is less than that of the system of errors obtained by their rejection multiplied by the probability of making so many, and no more, abnormal observations.' One calculates the probability of each of the sets of errors (with and without the divergent observations) and accepts the one with the higher probability. Chauvenet shows that an approximate criterion for the rejection of one doubtful observation results in values in agreement with Peirce's more general theory.

If $P(\epsilon)$ is the standard probability integral expressing the probability of occurrence of errors less than or equal to ϵ, then for n observations the expected number of errors greater than a critical value ϵ_c is

$$n\,[1 - P(\epsilon_c)] \;.$$

If this value is less than $\frac{1}{2}$ we will be inclined to expect no errors greater than ϵ_c; if greater than $\frac{1}{2}$, we should expect at least one error $>\epsilon_c$. Therefore the critical value ϵ_c for acceptance or rejection of an error is given by

$$n\,[1 - P(\epsilon_c)] \;=\; \frac{1}{2}$$

or

$$1 - P(\epsilon_c) \;=\; \frac{1}{2n} \;.$$

The development of modern statistical theory with its distribution functions for values obtained from finite sets of measurements provided new approaches to the problem of rejecting doubtful measurements. One of the most widely accepted is a rejection test developed by Grubbs [3] in terms of the ratio of standard deviations. Grubbs' statistic is

$$G^2 \;\equiv\; \frac{\sum_{i=1}^{n} (X_i - \overline{X}_n)^2}{\sum_{i=1}^{n+1} (X_i - \overline{X}_{n+1})^2} \;.$$

From our previous definition of S_n, we have

$$G^2 \;=\; \frac{(n-1)S_n^2}{n\;S_{n+1}^2} \;.$$

From this we have the relation between the normalized deviation x/S_n and G^2.

$$\frac{x}{S_n} \;=\sqrt{\frac{n^2-1}{n}\left(\frac{1}{G^2} - 1\right)} \;.$$

Thus from the critical values which Grubbs tabulates for his test one may calculate critical values for x/S_n. A table of such values is given by Acton [4].

One should not make a direct comparison between Chauvenet's criterion and Grubbs' test. The former is an intuitive solution which has no basis in modern statistical theory of finite samples. Further, Grubbs' values are tabulated as percentage points of the integral distribution so that the particular value one obtains in actual circumstances depends on the confidence level or significance level one wishes to establish. Nevertheless, the tests are remarkably similar in form and magnitude as shown by the plots in Fig. 1. We have plotted three levels for Grubbs' test as representative. The values for Chauvenet's test have been calculated in terms of the notation used in this note.

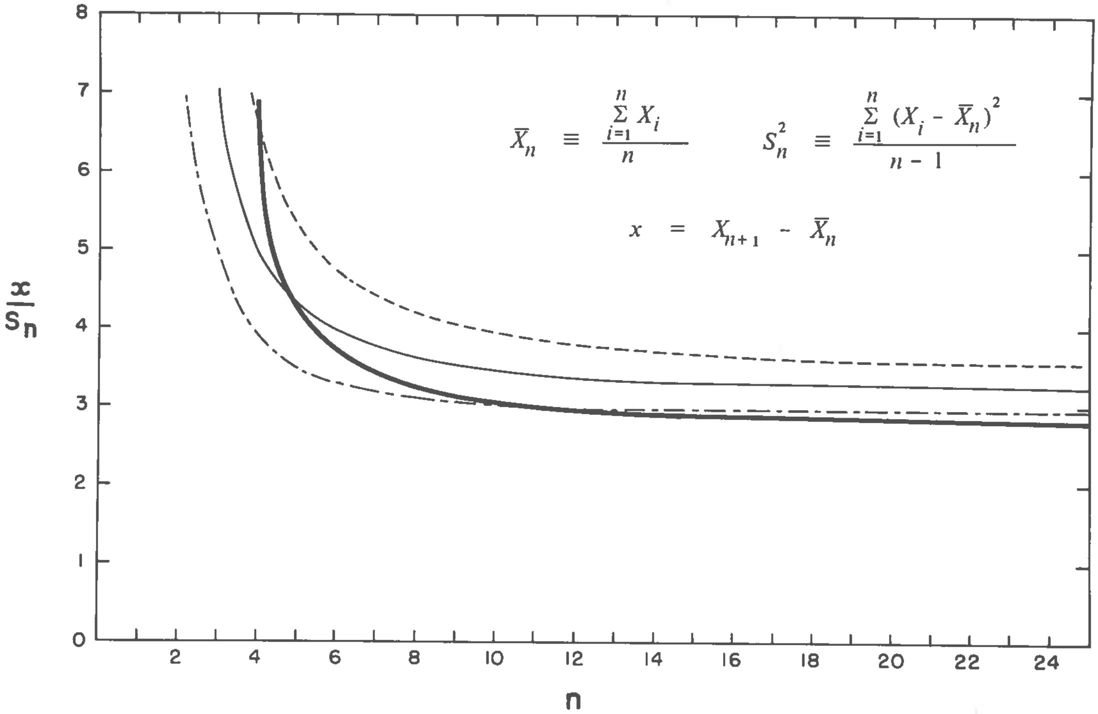

Figure 1 Critical values of normalized deviation x/S_n which, if exceeded, permit rejection of the n+1st reading. Heavy line: Chauvenet's criterion. Light lines calculated from Grubbs' test at 3 levels of significance: --- 2.5%, —— 5%, — – — 10%

It is apparent from these tests that a minimum level for rejection is about 3 times the standard deviation. Our earlier arguments showed that the effect of an additional reading decreases as n^{-1} so that provided n is large it is apparent that a large deviation can be tolerated. The situation is probably well summed up by paraphrasing some comments made by Acton [4]. The rejection tests are frequently only a 'statistical panacea to mechanize or to sanctify what is an essentially subjective operation. - - - - if there are physical explanations for the anomalous behaviour of a reading, these explanations should probably govern the eventual disposition of the data'.

References

1. Chauvenet, William. A manual of spherical and practical astronomy. Dover Publications (2 volumes), New York, 1960.

2. Worthing, A.G. and Geffner, J. Treatment of experimental data. John Wiley and Sons, Inc., New York, 1943.

3. Grubbs, F.E. Sample criteria for testing outlying observations. Annals of Mathematical Statistics, 21: 27–58; 1950.

4. Acton, F.S. Analysis of straight-line data. John Wiley and Sons, Inc., New York, 1959.

20-1-69 – BM