Journal of Statistical Software
MMMMMM YYYY, Volume VV, Issue II. http://www.jstatsoft.org/

Peirce's Criterion for the Rejection of Non-Normal Outliers; Defining the Range of Applicability

Christopher Dardis
Barrow Neurological Institute, Phoenix, Arizona

Abstract

Peirce's criteria for the rejection of non-normal outliers has been with us for over 150 years. Here, I present an implementation of the method in R. A number of examples are presented and I discuss its range of applicability. Finally, I give illustrations from the early literature on the method.

Keywords: Peirce, outlier, R.

1. Introduction

Peirce's criterion for the rejection of non-normality have been with us for over 150 years. It was, in fact, the first criterion developed for the exclusion of outliers. I became interested in his methods during the course of some lab research, where it became clear that the techniques we were using were producing occasional grossly erroneous results.

I was persuaded of the merits of the technique by a paper from Ross (2003), which gave a simple but practical technique for applying the method. However given the volume of data our lab was generating, I sought to automate the method as a function in R (Team (2012)). I was also interested to see if the technique generalized to a broader range than that given in the above paper, which is limited to rejecting up to 9 outliers from 60 observations.

The original paper Peirce (1852) describes a technique for rejecting doubtful observations in the case of those arising from observing planetary motion. It is assumed that unusual departures from normality (a Gaussian distribution) of observations are the result of the observer rather the planets themselves. I believe similar assumptions are worthwhile in generalizing the findings of the natural sciences.

In brief, his technique was to generate probabilities of error occurring in the system where all N observations are retained vs. that where k are rejected. He then rejected k observations if the new system (i.e., rejecting k) is closer to normal than the old.

2. Practical application

To begin with, an illustration of the merit of his technique following, with comparisons with a number of standard existing techniques. While multiple methods are already implemented in the **outliers** package in R, the following are limited to removing only one value: `chisq.out.test, dixon.test, grubbs.test`. There is limited literature as to how legitimate it is to repeat them multiple times on a dataset.

A leading alternative for rejecting non-normal values is Chauvenet's criterion `Chauvenet.R`. There is a lack of consensus as to whether it is legitimate to repeatedly apply the function, so I have provided the option `loop=TRUE` to do so. Repeated application tends to further shrink sets, whereas Peirce's criterion does not suffer this disadvantage. I took four sample sets; that from Ross (2003), one from the National Institute of Standards and Technology Natrella (2012) and two that are already available with R. The latter two are cautionary tales: **TeachingDemos** in regard to repeated application of Chauvenet's, **compositions** sa.outliers for the perils of a set with complete separation. These sets are shown in Figure 1 and the results in Table 1. Full details in `PeirceVsChauvenet.R`.

Another approach to rejecting outliers has been proposed by bbalibor^{TM}. Their goal is to determine an average value for a series of submissions (referred to as libor, basically a nominal rate of interest). They suggest that a reasonable approach in the case of 16 submissions is to eliminate the upper and lower 4, leaving 8 and then taking the mean of these 8. (Method explained here). I compared this to using Peirce's criterion on each set of 16 observations, then averaging those, figure 2. (This is based on part of a complete set available from google docs). Although there is no reason *a priori* to assume that the submissions on which libor is based should follow a normal distribution, it can clearly be seen that Peirce's criterion gives a result almost identical to its rival, and excludes far fewer observations per application (range is 1-4 in this example, vs 8 each time). This is akin to saying that excluding larger numbers of outliers tends to have little effect on the mean value for this type of data. Alternatively, one may say that the bbalibor^{TM} method is to be preferred owing to its simplicity.

Dataset	No. observations	Method (no. removed)		
		Peirce	Chauvenet	Chauvenet (repeated)
Ross	10	2	2	8
NIST	90	11	3	13
TeachingDemos	100	7	6	17
compositions sa.outliers	300	71	0	0

Table 1: Performance of outlier detection methods.

3. Methods

I sought to duplicate the table and results from Ross (2003). I was able to do so by following

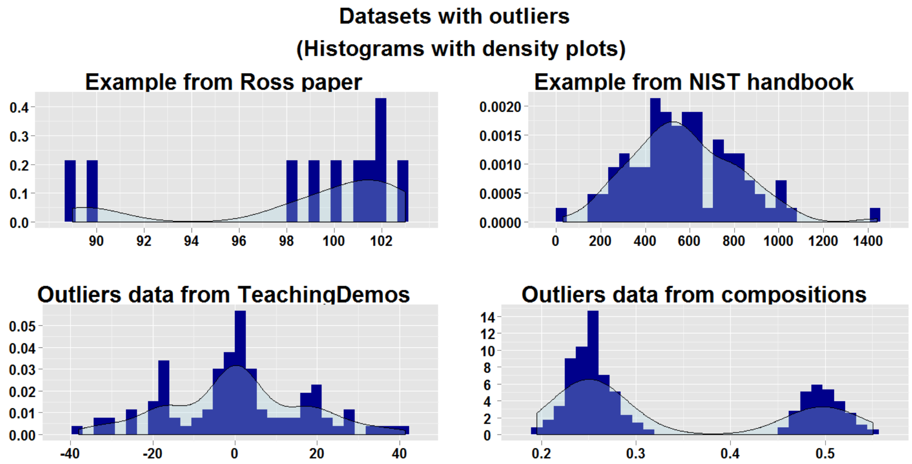

Figure 1: Illustration of datasets used to test outlier methods.

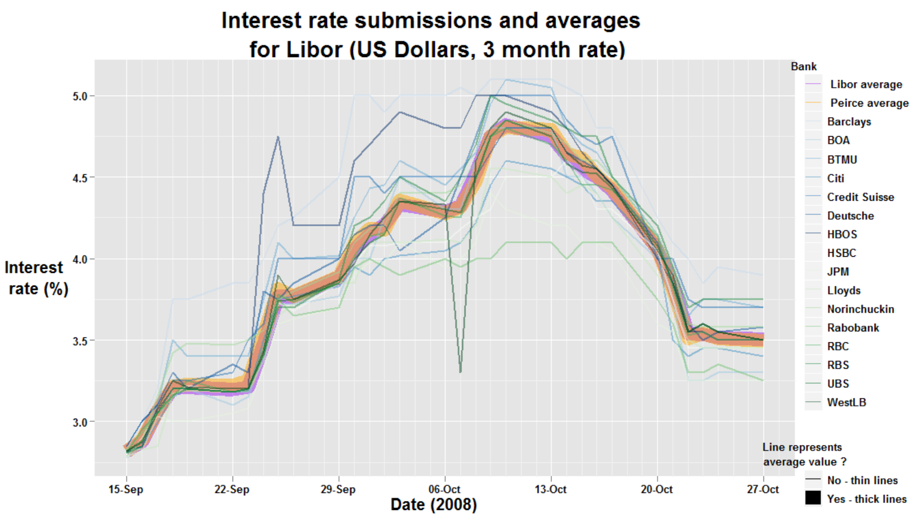

Figure 2: Libor calculated traditionally and using Peirce's criterion.

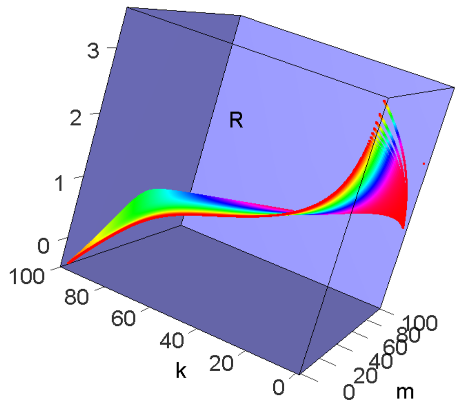

Figure 3: Values of R from % values of k and m. Sample size N=1000.

the methods in Gould (1855), (see `PeirceGould.R`). However, the upper limit for N, the number of observations in the sample, is limited by R's representation of large numbers, which is given by `.Machine$double.xmax` and is 143 on my device. A more efficient technique for achieving the same result already exists in C, and I implemented this as `Peirce.R`. Both methods rely on generating R, the ratio of the absolute error of one measurement to the sample standard deviation σ :

$$R = \frac{|x_i - \bar{x}|}{\sigma} \tag{1}$$

Where $\bar{x}$ is the sample mean. R depends on k, the number of outliers proposed to be rejected and m the number of "unknown quantities". The meaning of this latter may be destined to remain obscure; however it appears to be something akin to 'degrees of freedom', i.e., the number of *independent* processes that are giving rise to outliers in the data. Gould (1855) acknowledges that the cases of $m > 2$ are of little practical significance.

An illustration of the range of values of R for $k = 0 - 100\%$ and $m = 0 - 100\%$ of N is shown in Figure 3, with details in `PeirceLimits.R`. For values of $m = 1$ we can see R dropping below 0 at the point where $k >\approx 90\%$ i.e., it is meaningless to try to reject more than 90% of a given dataset. Additionally, for low values of k, increasing m does reduce R.

As an aside, I sought to generate the values in Table III in Gould (1855), giving values of $NlogQ$ for N observations and k proposed rejections. I followed his equation (B.) which is:

$$Q^k = \frac{k^k (N-k)^N - k}{N^k} \tag{2}$$

whence

$$N \log_{10} Q = N \log_{10} \sqrt[k]{\frac{k^k (N-k)^N - k}{N^k}} \tag{3}$$

However in faithfully replicating the table, I found additional adjustments were necessary such that

$$Nlog_{10}Q \rightarrow 10 + Nlog_{10}Q \tag{4}$$

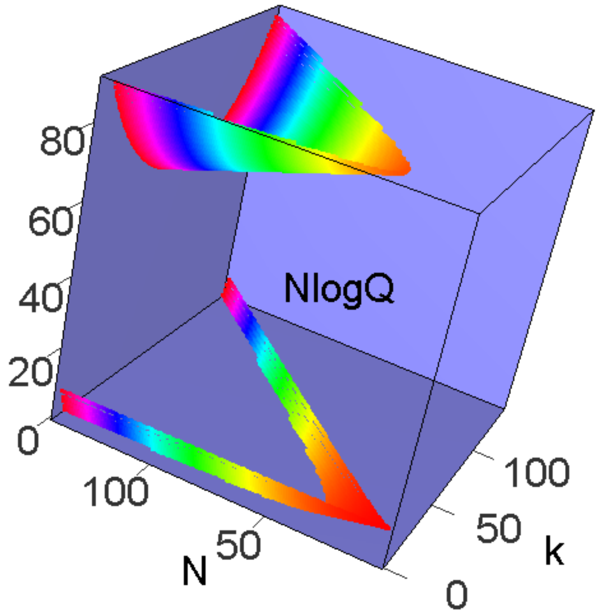

Figure 4: Values of $NlogQ$ for N and k, per Gould paper.

and

$$Nlog_{10}Q < 0.05 \quad \Rightarrow \quad Nlog_{10}Q \to 100 + Nlog_{10}Q \tag{5}$$

The reason for this is not entirely self-evident. An illustration of the function is shown in Figure 4, which may be manipulated, if desired, with `NlogQLimits.R`:

4. Original paper

Peirce himself appears to have been aware of the difficulties in interpreting his orininal paper.

> I perceive that the theory of my criterion has been frequently misunderstood. I presume this to be due in a great degree to the conciseness of the argument with which it was published.
>
> Peirce (1877)

However this did not prevent the methods becoming widely adopted in his own time, largely due to the clarity of Gould's implementation. I had some difficulty in replicating all of the results from the original paper Peirce (1852). However a number of formulas are of interest.

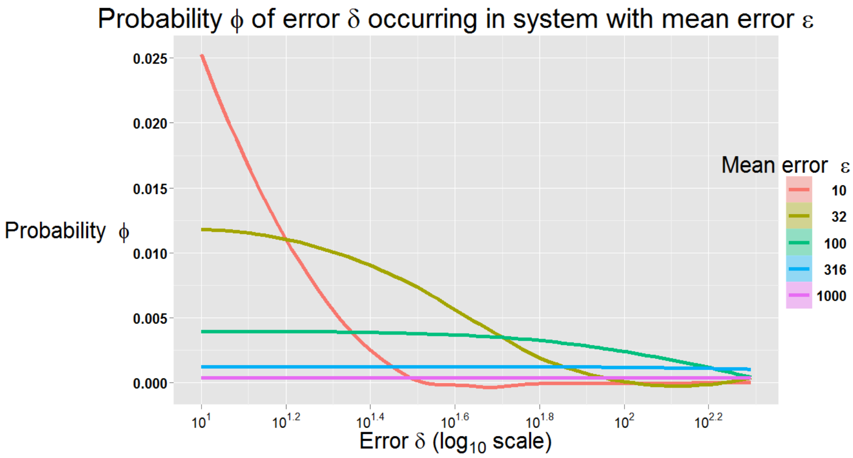

Figure 5: Probability of an error occurring, varying by mean error of system.

(These functions with corresponding plots are included in `Pierce1852.R`). His expression for the probability of certain error δ in a system with mean error ϵ is given by:

$$\phi(\delta) = \frac{1}{\epsilon\sqrt{2\pi}} e^{-\frac{\delta^2}{2\epsilon^2}} \tag{6}$$

This is illustrated for a number of sample ranges of interest in Figure 5:

His next formula gives the probability ψ of an error in a system which exceeds the required limit, $x\epsilon$:

$$\psi(x) = \frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{1}{2}x^2} \tag{7}$$

The reader may recognize ψ as closely related to the complementary error function, $erfc$, which is already implemented in R as `NORMT3::erfc`:

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2}\, dt \tag{8}$$

A comparison of both is shown in Figure 6.

His next equation for the probability of k observations exceeding the required limit $x\epsilon$ I took to be:

$$P = \left(\frac{\psi(x)}{\phi(x\epsilon)}\right)^k \tag{9}$$

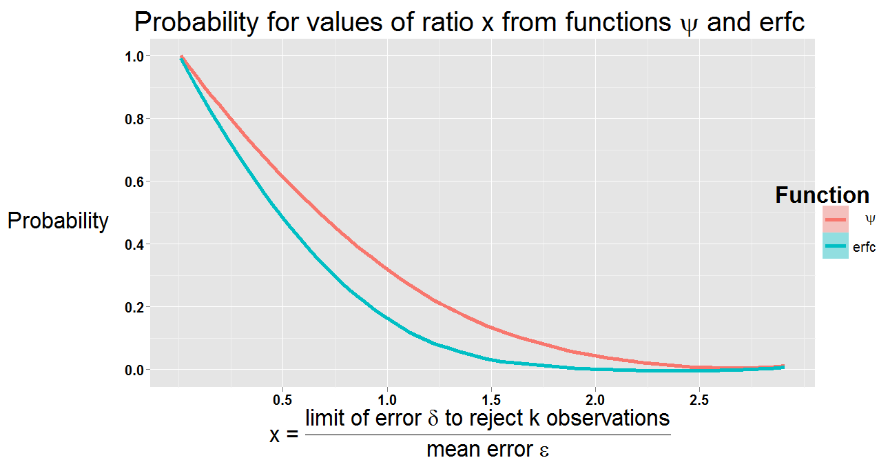

Figure 6: Comparison of Peirce's ψ with $erfc$.
ψ shows probability varying by ratio of limit of acceptable error to mean error.

whereby he derives:

$$P = \frac{1}{\epsilon^{N-k} 2\pi^{\frac{N^k}{2}}} e^{\frac{-N+m+kx^2}{2}} (\psi x)^k \tag{10}$$

However substituting arbitrary values in both lead to values of > 1. Again, the reason for this is not entirely self-evident to me.

5. Conclusion

Peirce deserves credit as the first to suggest a method of excluding outliers. Given the prevalence of normal distributions in routine observations, a revival of his methods may be timely. I hope these illustrations will clarify the range over which his methods may be applied.

6. Acknowledgements

Knud Thomsen for sharing the method in C. Kevin Mullin for converting the method to R.

References

Gould BA (1855). "On Peirce's Criterion for the Rejection of Doubtful Observations, with Tables for Facilitating its Application." *Astronomical Journal*, **4**(83), 81–87. doi:10.1086/100480. URL http://adsabs.harvard.edu/abs/1855AJ......4...81G.

Natrella M (2012). "NIST/SEMATECH e-Handbook of Statistical Methods." URL http://www.itl.nist.gov/div898/handbook/index.htm.

Peirce B (1852). "Criterion for the Rejection of Doubtful Observations." *The Astronomical Journal*, **2**(45), 161–163. URL http://articles.adsabs.harvard.edu/

cgi-bin/nph-iarticle_query?1852AJ......2..161P;data_type=PDF_HIGHhttp://adsabs.harvard.edu/full/1852AJ......2..161P.

Peirce B (1877). "On Peirce's Criterion." *Proceedings of the American Academy of Arts and Sciences*, **13**, 348–351. URL http://www.jstor.org/stable/25138498.

Ross S (2003). "Peirce's Criterion for the Elimination of Suspect Experimental Data." *Journal of Engineering Technology*, **20**(2), 1–12. URL http://classes.engineering.wustl.edu/2009/fall/che473/handouts/OutlierRejection.pdf.

Team RDC (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Affiliation:

Christopher Dardis
Department of Neurology
Barrow Neurological Institute
350 W. Thomas Road
Phoenix, AZ 85013
E-mail: christopherdardis@gmail.com
URL: https://christopherdardis.wordpress.com/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume VV, Issue II *Submitted:* yyyy-mm-dd
MMMMMM YYYY *Accepted:* yyyy-mm-dd