THE

ASTRONOMICAL JOURNAL.

No. 45.

VOL. II. CAMBRIDGE, 1852, JULY 24. NO. 21.

CRITERION FOR THE REJECTION OF DOUBTFUL OBSERVATIONS.

By BENJAMIN PEIRCE, LL. D.,

PERKINS PROFESSOR OF ASTRONOMY AND MATHEMATICS IN HARVARD UNIVERSITY.

1. In almost every true series of observations, some are found, which differ so much from the others as to indicate some abnormal source of error not contemplated in the theoretical discussions, and the introduction of which into the investigations can only serve, in the present state of science, to perplex and mislead the inquirer. Geometers have, therefore, been in the habit of rejecting those observations which appeared to them liable to unusual defects, although no exact criterion has been proposed to test and authorize such a procedure, and this delicate subject has been left to the arbitrary discretion of individual computers. The object of the present investigation is to produce an exact rule for the rejection of observations, which shall be legitimately derived from the fundamental principles of the Calculus of Probabilities.

2. *It is proposed to determine in a series of N observations, the limit of error, beyond which all observations involving so great an error may be rejected, provided there are as many as n such observations.*

The principle upon which it is proposed to solve this problem is, *that 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.*

In determining the probability of these two systems of errors, it must be carefully observed that, because observations are rejected in the second system, the corresponding observations of the first system must be regarded, not as being limited to their actual values, but only as surpassing the limit of rejection.

3. Let e be the base of the hyperbolic system of logarithms.

Let m be the number of unknown quantities.

Let N be the whole number of observations.

Let n be the number of observations proposed to be rejected.

Let $n' = N - n$ be the number of observations to be retained.

Let Δ, Δ', Δ'', &c. be the system of errors when no observation is rejected.

Let Δ_1, Δ_1', Δ_1'', &c. be the system of errors when n observations are rejected.

Let ε denote the mean error when no observation is rejected.

Let ε_1 denote the mean error when n observations are rejected.

Let $\varphi \Delta = \frac{1}{\varepsilon \sqrt{2\pi}} e^{-\frac{\Delta^2}{2\varepsilon^2}}$ denote the probability of an error Δ, in the first system, with the corresponding notation for the second system.

Let x be the ratio of the required limit of error for the rejection of n observations to the mean error ε.

Let $\psi x = \frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{1}{2}x^2}$ be the probability of an error in the first system, which exceeds the required limit.

Let y be the probability, supposed to be unknown, of such an abnormal observation that it is rejected upon account of its magnitude.

Let $y' = 1 - y$ be the probability that an observation is not of the abnormal character which involves its rejection.

Let $\Sigma . \Delta^2 = (N - m)\, \varepsilon^2$ denote the sum of the squares of all the errors Δ, Δ', &c.

Let $\Sigma . \Delta_1^2 = (n' - m)\, \varepsilon_1^2$ denote the sum of the squares of all the errors Δ_1, Δ_1', &c.

4. The probability of the first system of errors embodying the condition that n observations exceed the limit $x\varepsilon$, is

$$P = \varphi \Delta . \varphi \Delta' . \varphi \Delta'' \ldots \left(\frac{\psi x}{\varphi (x \varepsilon)}\right)^n = \frac{1}{\varepsilon^{n'} (2\pi)^{\frac{1}{2} n'}} e^{-\frac{\Sigma \Delta^2 - n x^2 \varepsilon^2}{2 \varepsilon^2}} (\psi x)^n$$

$$= \frac{1}{\varepsilon^{n'} (2\pi)^{\frac{1}{2} n'}} e^{\frac{1}{2}(-N + m + n x^2)} (\psi x)^n.$$

The probability of the second system of errors is

$$P_1 = x'^{n'} x^n, \varphi \Delta_1 . \varphi \Delta_1' \ldots = \frac{x'^{n'} x^n}{\varepsilon_1^{n'} (2\pi)^{\frac{1}{2} n'}} e^{-\frac{\Sigma \Delta_1^2}{2 \varepsilon_1^2}}$$

$$= \frac{y'^{n'} y^n}{\varepsilon_1^{n'} 2\pi^{\frac{1}{2} n'}} e^{\frac{1}{2}(-n' + m)}.$$

5. To authorize the proposed rejection of n observations we must have

$$P < P_1$$

which gives at once

$$\left(\frac{\varepsilon_1}{\varepsilon}\right)^{n'} e^{\frac{1}{2}n(x^2-1)} (\psi\, x)^n < y'^{\,n'}\, y^n.$$

6. The value of y must be determined by the condition that P_1 is a maximum, and therefore that $y'^{\,n'}\, y^n$ is a maximum. The differential of the logarithm of this equation gives then for the maximum

$$\frac{y'}{n'} = \frac{y}{n}$$

$$y = \frac{n}{N}\,, \quad y' = \frac{n'}{N}$$

$$y'^{\,n'}\, y^n = \frac{n^m\, n'^{n'}}{N^n}$$

which may be denoted by Q^N; and *the required condition of rejection becomes*, by extracting the Nth root,

$$\left(\frac{\varepsilon_1}{\varepsilon}\right)^{y'} \left(e^{\frac{1}{2}(x^2-1)}\, \psi\, x\right)^{y} < Q \qquad (A.)$$

in which

$$Q = y'^{\,y'}\, y^y = y^y\,(1-y)^{1-y}. \qquad (B.)$$

7. The relation of ε' to ε must depend upon the nature of the equations which correspond to the rejected observations. But in the form in which the problem is here presented, the attention is directed to no special case; and it is sufficient for a first approximation to suppose that the excess of $\Sigma\Delta^2$ over $\Sigma\Delta_1^2$ is only equal to the sum of the squares of the rejected observations, which gives the equation

$$(1-m')\,\varepsilon^2 - y\,x^2\,\varepsilon^2 = (y'-m')\,\varepsilon_1^2$$

in which $\quad m' = \frac{m}{N}.$

Hence,

$$\frac{\varepsilon_1}{\varepsilon} = \left(\frac{1-m'-y\,x^2}{y'-m'}\right). \qquad (C.)$$

The combination of (A), (B), and (C) for various values of m' and y' gives a set of values of x, which will answer for most cases. Table I. contains such values for different values of m' and y.

8. When the attention is directed to the rejection of special observations, the values of ε' and ε can be determined by direct investigation, for which the usual forms of computation in the method of least squares afford various facilities; and the application of the criterion involved in (A) is simplified by means of Tables II. and III. Table II. contains Brigg. log. Q for different values of y, and Table III. contains Brigg. log. R for values of x, in which

$$R = e^{\frac{1}{2}(x^2-1)}\, \psi\, x. \qquad (D.)$$

Previously to the publication of these tables, computers will easily determine $\psi\, x$ from Bessel's table of $\int e^{-t^2}$, published in the *Fundamenta Astronomiæ*, or more conveniently from the tables given by Encke in his memoir, *Ueber die Methode der Kleinsten Quadrate*, in the *Berlin Jahrbuch* for 1834.

The slow rate at which R changes will be found greatly to facilitate the determination of the criterion, as will be seen in the following examples: —

Example 1. To determine the limit of rejection of a single observation in the case of thirty observations of the vertical semidiameter of *Venus*, made by Professor Coffin and Lieutenant Page with the mural-circle, at Washington, in the year 1846.

In the reduction of these observations, I have assumed two unknown quantities, of which one is a constant and the other is proportional to the horizontal parallax. I have then found for the residual errors

—0″.63	+0″.27	+0″.60	+0″.30	—0″.10	—0″.19
+0.45	—0.43	—0.17	—0.05	+0.16	—0.52
+0.34	+0.08	—0.23	—0.45	—0.47	—0.51
—0.26	—0.11	+0.83	+0.11	+0.04	—0.49
—0.28	+0.13	+0.69	+0.45	—0.12	+0.50

and I also find $\qquad \log.\ \varepsilon^2 = 9.2368.$

We have then, in this case,

$$N = 30, \qquad n = 1, \qquad n' = 29, \qquad m = 2$$

whence

$$\frac{\varepsilon'}{\varepsilon} = \left(\frac{28-x^2}{27}\right)^{\frac{1}{2}}, \quad Q^N = \frac{29^{29}}{30^{30}}$$

$$-N \log.\ Q = 1.9040.$$

For a first approximation let $\log.\ R = 9.3$

which gives $\log.\ R - N\log.\ Q = 1.2 < \left(\frac{\varepsilon}{\varepsilon'}\right)^{29}$

$$\begin{aligned} \log.\ \varepsilon^2 - \log.\ \varepsilon'^2 &> \ .08 \\ \log.\ (28-x^2) &< 1.35 \\ 28 - x^2 &< 22.4 \\ x^2 &> 5.6 \\ x &> 2.4. \end{aligned}$$

If this value of the limit is substituted in the next approximation, we have $\qquad \log.\ R = 9.248$

whence

$$\begin{aligned} \log.\ \varepsilon^2 - \log.\ \varepsilon'^2 &> .0795 \\ x^2 &> 5.51 \\ x\,\varepsilon &> 0''.97 \end{aligned}$$

which is so much greater than either of the above errors of observation, that neither of them is to be rejected.

Example 2. To determine the limit of rejection of one or two observations in the case of fifteen observations of the vertical semidiameter of *Venus*, made by Lieutenant Herndon, with the meridian-circle at Washington, in the year 1846.

The same method of reduction used in this example as in the preceding gave for the residual errors

—0″.30	+0″.48	+0″.63	—0″.22	+0″.18
—0.44	—0.24	—0.13	—0.05	+0.39
+1.01	+0.06	—1.40	+0.20	+0.10

$$\log.\ \varepsilon^2 = 9.53050.$$

For the criterion of rejecting one observation, we have

$$N = 15, \qquad n = 1, \qquad n' = 14, \qquad m = 2$$

whence

$$\frac{\varepsilon'}{\varepsilon} = \left(\frac{13-x^2}{12}\right)^{\frac{1}{2}}, \quad Q^N = \frac{14^{14}}{15^{15}}$$

$$-N \log.\ Q = 1.5955.$$

For a first approximation let log. $R = 9.3$

which gives log. $R - N$ log. $Q = 0.90 < \left(\frac{\varepsilon}{\varepsilon'}\right)^{14}$

$$\begin{aligned} \log. \varepsilon^2 - \log. \varepsilon'^2 &> 0.13 \\ \log. (13 - x^2) &< 0.95 \\ 13 - x^2 &< 8.9 \\ x^2 &> 4.1 \\ x &> 2.0. \end{aligned}$$

If this value of the limit is substituted in the next approximation, we have

$$\log. R = 9.309$$

whence

$$\begin{aligned} \log. \varepsilon^2 - \log. \varepsilon'^2 &> 0.1292 \\ x^2 &> 4.083 \\ x\,\varepsilon &> 1''.18 \end{aligned}$$

which shows that the error 1″.40 should be rejected.

For the criterion for rejecting two observations, we have

$$N = 15, \qquad m = 2, \qquad n' = 13, \qquad m = 2$$

whence $\quad \frac{\varepsilon}{\varepsilon'} = \left(\frac{13 - 2\,x^2}{11}\right)^{\frac{1}{2}}, \quad Q^N = \frac{2^2 \,.\, 13^{13}}{15^{15}}$

$$-\frac{N}{n} \log. Q = 1.2790.$$

For a first approximation let log. $R = 9.3$

which gives log. $R - \frac{N}{n}$ log. $Q = 0.58 < \left(\frac{\varepsilon}{\varepsilon'}\right)^{6\frac{1}{2}}$

$$\begin{aligned} \log. \varepsilon^2 - \log. \varepsilon'^2 &> 0.18 \\ \log. (13 - 2\,x^2) &< 0.86 \\ 13 - 2\,x^2 &< 7.24 \\ x^2 &> 2.88 \\ x &> 1.7. \end{aligned}$$

If this value of the limit is adopted as a second approximation, we have $\quad \log. R = 9.360$

whence

$$\begin{aligned} \log. \varepsilon^2 - \log. \varepsilon'^2 &> 0.1882 \\ x^2 &> 2.930 \\ x\,\varepsilon &> 0''.997 \end{aligned}$$

which shows that the two observations corresponding to errors +1″.01 and —1″.40 should be rejected.

FIRST PART OF TABLE III. LOG. *R*.

Argument x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.0	9.5015	9.4992	9.4970	9.4947	9.4924	9.4902	9.4880	9.4857	9.4835	9.4813
1	.4791	.4769	.4747	.4725	.4704	.4682	.4661	.4639	.4618	.4597
2	.4575	.4554	.4533	.4512	.4491	.4471	.4450	.4429	.4409	.4388
3	.4367	.4347	.4327	.4307	.4286	.4266	.4246	.4226	.4206	.4187
4	.4167	.4147	.4128	.4108	.4088	.4069	.4050	.4030	.4011	.3992
5	.3973	.3954	.3935	.3916	.3897	.3878	.3860	.3841	.3822	.3804
6	.3786	.3767	.3749	.3731	.3713	.3694	.3676	.3658	.3640	.3622
7	.3604	.3587	.3569	.3551	.3533	.3516	.3498	.3481	.3464	.3446
8	.3429	.3412	.3395	.3378	.3360	.3343	.3327	.3310	.3293	.3276
9	.3259	.3243	.3226	.3209	.3193	.3176	.3160	.3144	.3127	.3111
2.0	.3095	.3079	.3063	.3046	.3030	.3014	.2999	.2983	.2967	.2951
1	.2935	.2920	.2904	.2888	.2873	.2857	.2842	.2827	.2811	.2796
2	.2780	.2765	.2750	.2735	.2720	.2705	.2690	.2675	.2660	.2645
3	.2630	.2615	.2600	.2586	.2571	.2556	.2542	.2527	.2512	.2498
4	.2483	.2469	.2455	.2440	.2426	.2412	.2398	.2383	.2369	.2355
5	.2341	.2327	.2313	.2299	.2285	.2272	.2258	.2244	.2230	.2217
6	.2203	.2189	.2176	.2162	.2149	.2135	.2122	.2108	.2105	.2082
7	.2068	.2055	.2042	.2029	.2016	.2002	.1989	.1976	.1963	.1950
8	.1937	.1924	.1911	.1898	.1886	.1873	.1860	.1847	.1835	.1822
9	.1810	9.1797	9.1784	9.1772	9.1759	9.1747	9.1735	9.1722	9.1710	9.1697
3.0	9.1685									

FROM A LETTER OF PROFESSOR ARGELANDER TO THE EDITOR.

Bonn, 1852, *June* 25.

Herewith I send you an ephemeris of *Thetis* for July, — which will, I trust, represent the course of this planet with considerable accuracy. One of my pupils, Mr. Schoenfeld, has computed it from elements of his own, which I also communicate. After the moonshine *Thetis* will be too faint for observation by our telescopes here, and if we are to have any observations at all, they must be made at observatories provided with the great refractors, of which you have so many in America.

I have just sent you, by the way of Leipsic, the first part of