8. HOW TO REDUCE TO A MINIMUM THE MEAN ERROR OF TABLES

First published in "The Napier Tercentenary Memorial Volume", 1915, p. 345.

In the years which have passed since the memorable invention of logarithms, different systems have been devised for the arrangement and calculation of tables of logarithms, chiefly for the purpose of combining rapidity and convenience in use with a considerable degree of accuracy. Some of these systems will be discussed in the following remarks.

1. Firstly, we will consider the simple type of tables, which are intended for ordinary linear interpolation. Sometimes the first-differences are directly given, and, as a rule, tables of proportional parts are provided, while economy of space is effected by the double-entry arrangement (already used by *J. Newton* in the seventeenth century). For the sake of brevity this type of tables will in the following be referred to as "Type A". They are, even nowadays, preferred by many calculators, although other types (see below) undoubtedly surpass them as to convenience and rapidity.

Now let us consider accuracy. We may here suppose that the higher differences are unimportant, or the function nearly linear. It is here essential to give, not the maximal error, nor the probability of some special value of the error, but the mean-square of errors (which is the square of the mean-error, according to the general significance of this expression in the theory of probabilities). As unit of the error we will use the last decimal unit of the table-values. If we consider only the values directly printed, the mean-square of errors is known to be 1/12. If we suppose the interpolations distributed evenly along the whole table-interval, we find the mean-square of errors to be

$$1/18 + 1/12 = 5/36, \text{ or } 0{\cdot}1388 \ldots .$$

If, however, only the points dividing the interval into 10 equal parts are considered, we find a slightly lower value,

$$\frac{403}{3000}, \text{ or } 0{\cdot}13433 \ldots \ldots$$

It is not necessary to give here the exact law for the distribution of the errors, which is, furthermore, not very different from the "normal" law of errors; the reader interested in this question should consult *De accuratione qua possit quantitas per tabulas determinari*, by *Carolus Æmilius Mundt*, Havniæ, 1842.

2. We will now consider another well-known type, which will hereafter be referred to as "Type B". The ordinary double-entry table is here accompanied by a special auxiliary table, the separate horizontal lines of which correspond to the horizontal lines of the main table. By means of the values of the auxiliary table, which are simply 1, 2, 3 ... 9-tenths of the mean-difference of the opposite part of the main-table, all the interpolations necessary are reduced to simple additions. This convenient form of table seems to have been used not earlier than in the nineteenth century, and the oldest tables, as far as I know, are the following two: *Fünfstellige Logarithmen*, by *A. M. Nell* (1866), and *Logarithms and Antilogarithms*, issued by the Institute of Actuaries in 1877 (4 figures). The arrangement used in some parts of *Tables trigonométriques décimales*, by *Borda* and *Delambre* (1801) is, however, essentially the same. Obviously some modifications of this type of table are possible; but the arrangement described here is the most natural, because the number 10 is the basis of the system of numeration.

Considering accuracy, we find the mean-square of the error to be

$$1/6 \text{ or } 0.1666 \ldots\ldots$$

3. An increase of accuracy, without loss of convenience, has been accomplished by the appearance of *Fircifret Logaritmetabel*, by *N. E. Lomholt*, 1897 (the first edition). This author's object has been to do away with the great errors (exceeding 1·05 units), and furthermore to diminish the number of errors exceeding 0·5, as well as to diminish the "average error". He has not, as sometimes stated (for example, in the *Mathematical Encyclopædia*, both editions), reduced the average error to a minimum, and he has not thought it necessary to use a definite method, excluding entirely the personal element. Although, in my opinion at least, this is a drawback, I willingly acknowledge not only the real progress made, but also the general idea of improving the tables of Type B by means of new values both in the main table and in the auxiliary table.

4. In a Danish paper, published in *Nyt Tidsskrift for Matematik*, 1910, the present writer proposed to choose the 10 + 9 or 19 values of each horizontal line in such a manner that the sum of the square of the 100 resulting errors will be reduced to a minimum. It is hardly necessary to

say anything here in defence of this proposal[1]); but I will shortly set forth the method by which the 19 values may be found, using as an example the calculation of a single line (61) of the 4-figure table of logarithms.

Let us, for the present at least, accept the values of a 7-figure table as exact. As a starting-point we shall use a preliminary set of 19 values taken from a table of Type B. From these we derive 100 preliminary values of logarithms, and we write them down in form of a square. The 100 values are subtracted from the 100 corresponding "exact" values, and the 100 differences (or "errors") are written down in a similar manner. As some of the differences are negative, it is convenient to use either the number 9 or, still better (in accordance with the late Professor *T. N. Thiele*'s suggestion), the letter ν as representing a negative unit prefixed to a number, the following decimal parts being positive. By adding the vertical and the horizontal arrays we can obtain the sums V and H. These sums we arrange according to magnitude in two parallel columns, the former to the left, with the 10 terms decreasing from the top downwards, the latter to the right, with the 10 terms increasing from the top downwards. (Practically we always find that the difference between two arbitrary terms of each column will be less than 10, numerically; if not, we use another preliminary set of values satisfying this condition). We further extend the two columns downward by the repetition of some of the terms at the top, having previously subtracted 10 from each of the values V and added 10 to each of the values H.

If now we wish to alter the preliminary table in order to obtain the final table, we may represent the resulting changes in the 100 values of logarithms by following a certain number (m) of the vertical arrays and a certain number (n) of the horizontal arrays, adding 1 to the values of

[1]) In: *Nyt Tidsskrift for Matematik B*, vol. 22, 1911, p. 10, Erlang states the following reasons why the method of least squares should be used: "In practice, the errors contained in any particular table do not manifest themselves immediately or one by one. The numbers corresponding to the logarithms to be looked up in the table — usually the results of measurements — contain errors already; several logarithms must be added to, or subtracted from, other logarithms in order to find the logarithm of the number that is to be looked up in the table of antilogarithms. The detrimental effect of the resultant error is some function of the magnitude of the error. This function cannot, as a rule, be determined; but then, this is fortunately not necessary. The probability that the resultant error exceeds a certain quantity x depends only on the ratio of x to the resultant mean error and decreases as the latter decreases; but this resultant mean error can be expressed in a well-known and very simple manner by the mean errors corresponding to the different sources of errors. These mean errors should therefore be minimum. It is assumed in the above that the resultant errors satisfy the typical, or exponential, law of errors; that this is actually the case can be proved, however, even without knowledge of the particular laws of errors applying to the different error sources in question if only the number of these sources is fairly great."

Log 6100 ... 6199.

	(+)	(+)									
	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917	
(—) 1	4	1	9	6	3	0	7	4	1	8	
1	4	1	9	6	3	0	7	4	1	8	
2	5	2	0	7	4	1	8	5	2	9	
(—) 3	6	3	1	8	5	2	9	6	3	0	
(—) 4	7	4	2	9	6	3	0	7	4	1	
4	7	4	2	9	6	3	0	7	4	1	
5	8	5	3	0	7	4	1	8	5	2	
(—) 6	9	6	4	1	8	5	2	9	6	3	
6	9	6	4	1	8	5	2	9	6	3	
	298	412	ν514	ν605	ν684	ν751	ν807	ν852	ν885	ν906	ν8714
	010	123	ν224	ν313	ν391	ν457	ν512	ν555	ν587	ν608	ν5780 (+)
	722	833	ν933	021	098	163	217	259	290	309	2845
	434	544	ν643	ν730	ν805	ν869	ν922	ν963	ν992	011	ν9913
	145	254	ν352	ν438	ν512	ν575	ν626	ν666	ν695	ν712	ν6975 (+)
	ν857	ν965	ν061	ν146	ν219	ν281	ν331	ν370	ν397	ν413	ν4040 (+)
	568	675	ν770	ν854	ν926	ν986	035	073	099	114	1100
	279	385	ν479	ν561	ν632	ν692	ν739	ν776	ν801	ν815	ν8159
	ν990	095	ν188	ν269	ν339	ν397	ν444	ν479	ν503	ν516	ν5220 (+)
	701	805	ν896	ν976	045	102	148	182	205	216	2276
	3004	4091	ν5060	ν5913	ν6651	ν7273	ν7781	ν8175	ν8454	ν8620	
	(—)	(—)									

7·095
ν82·015

25·080 — 22 = 3·080

2·344
ν78·801

23·543 — 25 = ν8·543

97·398
ν79·901

17.497 — 19 = ν8.497

p	V	H	q
ν5			5
	4091	ν4040	
ν6			4
	3004	ν5220	
ν7			3
	ν8620	ν5780	
ν8			2
	ν8454	ν6975	
ν9			1
	ν8175	ν8159	
0			0
	ν7781	ν8714	
1			ν9
	ν7273	ν9913	
2			ν8
	ν6651	1100	
3			ν7
	ν5913	2276	
4			ν6
	ν5060	2845	
5			ν5
	ν4091	4040	
6			ν4
	ν3004	5220	
			ν3
		5780	
			ν2
		6975	

the former and subtracting 1 from the values of the latter. In the corresponding arrays of the set of 100 errors corresponding changes will take place, subtractions in the vertical and additions in the horizontal arrays. The resulting improvement (the diminution of the sum of the squares of errors) we can easily find by the formula:

$$I = 2\,[V] - 2\,[H] - 10\,m - 10\,n + 2\,mn,$$

[V] being the sum of the m quantities V, and [H] the sum of the n quantities H. It is now at once obvious that in order to find these quantities in the columns, we must proceed from the top of each column downwards to a certain point, without skipping over any terms. Thus the only problem left is to find the points where we are to stop, or, in other words, the numbers m and n. But if we are to stop between V_m and V_{m+1}, and between H_n and H_{n+1}, the following conditions —

$$V_m > 5 - n > V_{m+1}$$

$$H_n < m - 5 < H_{n+1}$$

— must be satisfied, as, otherwise, it would be better to take one step more or less, down the right or the left column. If we use the numbers

$$p = m - 5$$

$$q = 5 - n$$

instead of m and n for the numeration of the successive intervals of the columns H and V (see example), the two conditions and also the expression of I take the simple forms:

$$V_m > q > V_{m+1}$$

$$H_n < p < H_{n+1}$$

$$I/2 = [V] - [H] - (pq + 25).$$

It is now easy to find out the only pairs of values of p and q compatible with the two conditions (giving what may be called the relative maxima of I). The result may be marked by limiting lines, as shown in the example. For the final choice we must, by means of the formula, calculate the corresponding values of I, ordinarily very few in number, taking the values $pq + 25$ from a small table with two entries. The greatest value of I/2 is chosen. The signs (+) and (—) indicate the resulting deviations from the preliminary table.

It happens in a few cases that the number of decimal places (here 7) used in the calculation turns out to be insufficient, but it is easy to give a rule covering these cases.

If this sort of calculation is to be undertaken on a large scale it is best to try to get rid of part of the work, for example, by finding the sum of 10 function-values with equidistant arguments, without actually undertaking the addition. For special functions, such as antilogarithms, the way to proceed is obvious. Furthermore, the function considered in most cases is so nearly linear that we can find the mean of the 10 values from the value corresponding to the mean-argument, applying, if necessary, a small and easily determinable correction.

A set of 4-figure tables[1]) of the type described, has been calculated by *H. C. Nybølle*, Mathematical Assistant at the Danish Statistical Department, and myself. Another collection also containing 5-figure tables is at present being elaborated.

I would only like to mention that similar principles might possibly find application for the construction of tables of some simple and practically important functions of complex variables.

5. Concerning the mean-square of errors in tables of Type B_1, I have tried unsuccessfully to find the exact value as in the cases A and B (see above), especially for the purpose of finding the difference in this respect between B and B_1. The solution of this problem is theoretically possible, under the same supposition as to the nature of the function as above, and the integrations necessary are very simple; but the number of cases to be considered is very great. Some indications may, however, be had from the experiences available; thus we find, that the improvement I (for a set of 100 values) is, on the average, about 2 or 3 units; sometimes, although seldom, it will be as great as 50 (about), sometimes 0. We might also consider the case in which the second differences of the function are considerable (although one can, of course, get rid of this case by altering the interval). In this case the mean-square of errors produced by the aforesaid cause will obviously be about four times greater for Type B than for Type B_1.

It seems probable that Types B or B_1 will be much used in the future for the construction of tables of different functions, and if stress is laid on the greatest accuracy compatible with the arrangement, space, and number of figures chosen, Type B_1 should be preferred.

---

[1]) *A. K. Erlang, Fircifrede Logaritmetavler,* G. E. C. Gad, København. (1910-11), three editions, A, B, C, the last being the most complete.