Checking by Differences—I

1. Introduction and Summary. When computing a table of numerical values of a mathematical function, an essential need is a check on the accuracy of the results. For this check to be fully satisfactory it must be independent, as nearly completely as possible, of the original calculations. This independence should apply to the method of computation used in the check and not only to the numerical details. Apart from this it is convenient to have a check that is as simple as possible to apply.

When the values computed form a systematic table for equally-spaced values of some associated variable—usually, but not necessarily, taken as argument—the best-known check is probably that provided by forming a table of differences. The accuracy of the results is then tested by an examination of the general run of the values of the differences of some high order, say 3rd, 5th or possibly 10th differences. It will be assumed in what follows that argument values are equidistant.

The check provided by this process of differencing is very easy to apply, and is almost always fully satisfactory in all the senses outlined above. The precise details of the process and its pitfalls do not, however, seem to have been set out fully in print. It is the purpose of this paper to consider some aspects of the process, and to discuss possible methods of detecting and finding several types of error.

In **2**, the normal difference table for equal argument-intervals is discussed. The cases of a polynomial and of a general function are both considered.

In **3**, the effect of an isolated error is exhibited, and in **4**, methods for distinguishing true errors or blunders from inevitable rounding-off errors are considered.

It is proposed to examine in a later paper some cases where there are blunders due to causes other than mistakes in function values, or where there are coupled or systematic blunders, or where the resulting effects are overlapping for other reasons. In particular, ways of distinguishing mistakes made during the formation of differences are not considered in the present paper.

2. The Normal Difference Table. 2.1. When a table of exact values of a polynomial of degree n, for equidistant values of the argument, is differenced, the values of the n-th differences are all equal, and values of higher differences are all zero. This is too well known for a numerical illustration to be needed.

If, however, the values of the polynomial tabulated are rounded off to a fixed number of decimals, the n-th differences are no longer constant, but periodic, with period depending on the degree of the polynomial and on the number of figures dropped. Higher differences also form cycles of the same period. This may be illustrated by means of the quadratic function $10x(x - 1)$. If this is tabulated to the nearest integer for interval .1 in x, the second differences run through the ten values 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, this cycle being repeated indefinitely. It may be noted that the average value is .2, in agreement with the true second difference for a table of exact values. Likewise, the third differences give a cycle 0, 0, $+1$, -1, 0, 0, 0, 0, $+1$, -1, with average zero.

2.2. Consider now the more usual case of a function that cannot be tabulated exactly. Table I gives 5-decimal values of $\log_{10} N$, for $N = 10(1)30$, with all differences to order 10. In this illustration the interval of the argument has been chosen to be small enough for checking by differences to be feasible. If too large an interval is chosen, the differences may diverge as order increases.

It will be noted that:

(i) For low orders of differences, regularity is apparent. The magnitude diminishes as order increases. This holds up to about δ^4 in the present table.

(ii) For high orders of differences, irregularity appears. This is due to the inevitable rounding-off errors, and shows first in an irregular sign-pattern. Later, as order increases, the differences increase in magnitude but more or less irregularly.

(iii) For a sequence of differences of a particular (high) order, the larger values tend to occur in groups, with signs strictly alternating and with values falling away on either side from a central maximum value, or pair of such values.

TABLE I

N	Log N		δ^2		δ^4		δ^6		δ^8		δ^{10}
10	1.00000										
		+4139									
11	1.04139		−360								
		+3779		+57							
12	1.07918		−303		−11						
		+3476		+46		−1					
13	1.11394		−257		−12		+ 9				
		+3219		+34		+8		−20			
14	1.14613		−223		− 4		−11		+37		
		+2996		+30		−3		+17		−61	
15	1.17609		−193		− 7		+ 6		−24		+ 90
		+2803		+23		+3		− 7		+29	
16	1.20412		−170		− 4		− 1		+ 5		− 28
		+2633		+19		+2		− 2		+ 1	
17	1.23045		−151		− 2		− 3		+ 6		− 10
		+2482		+17		−1		+ 4		− 9	
18	1.25527		−134		− 3		+ 1		− 3		+ 8
		+2348		+14		0		+ 1		− 1	
19	1.27875		−120		− 3		+ 2		− 4		+ 4
		+2228		+11		+2		− 3		+ 3	
20	1.30103		−109		− 1		− 1		− 1		+ 17
		+2119		+10		+1		− 4		+20	
21	1.32222		− 99		0		− 5		+19		− 75
		+2020		+10		−4		+15		−55	
22	1.34242		− 89		− 4		+10		−36		+132
		+1931		+ 6		+6		−21		+77	
23	1.36173		− 83		+ 2		−11		+41		−153
		+1848		+ 8		−5		+20		−76	
24	1.38021		− 75		− 3		+ 9		−35		+133
		+1773		+ 5		+4		−15		+57	
25	1.39794		− 70		+ 1		− 6		+22		− 83
		+1703		+ 6		−2		+ 7		−26	
26	1.41497		− 64		− 1		+ 1		− 4		
		+1639		+ 5		−1		+ 3			
27	1.43136		− 59		− 2		+ 4				
		+1580		+ 3		+3					
28	1.44716		− 56		+ 1						
		+1524		+ 4							
29	1.46240		− 52								
		+1472									
30	1.47712										

TABLE II

N	E		δ^2		δ^4		δ^6		δ^8		δ^{10}
10	0.00										
		+.27									
11	+ .27		−.41								
		−.14		+ .77							
12	+ .12		+.35		−1.86						
		+.21		−1.10		+4.56					
13	+ .34		−.74		+2.69		−10.15				
		−.53		+1.60		−5.59		+20.42			
14	− .20		+.85		−2.90		+10.27		−37.17		
		+.32		−1.30		+4.68		−16.75		+61.07	
15	+0.13		−.45		+1.78		− 6.48		+23.90		− 90.03
		−.13		+ .47		−1.80		+ 7.16		−28.96	
16	− .00		+.02		− .03		+ .68		− 5.06		+ 27.98
		−.11		+ .44		−1.13		+ 2.10		− .98	
17	− .11		+.46		−1.16		+ 2.78		− 6.03		+ 9.99
		+.36		− .71		+1.65		− 3.93		+ 9.01	
18	+ .25		−.25		+ .49		− 1.16		+ 2.98		− 8.00
		+.11		− .22		+ .49		− .96		+ 1.01	
19	+ .36		−.47		+ .98		− 2.11		+ 3.99		− 4.00
		−.36		+ .76		−1.62		+ 3.03		− 3.00	
20	−0.00		+.29		− .64		+ .92		+ .99		− 17.00
		−.07		+ .12		− .71		+ 4.02		−20.00	
21	− .07		+.41		−1.35		+ 4.94		−19.01		+ 75.00
		+.34		−1.23		+4.23		−14.98		+55.00	
22	+ .27		−.82		+2.88		−10.05		+36.00		−132.00
		−.48		+1.65		−5.82		+21.01		−77.00	
23	− .22		+.83		−2.94		+10.96		−41.00		+153.00
		+.34		−1.29		+5.15		−19.99		+76.00	
24	+ .12		−.46		+2.21		− 9.03		+35.00		−133.00
		−.12		+ .92		−3.88		+15.01		−57.00	
25	+0.00		+.46		−1.67		+ 5.98		−22.00		+ 83.00
		+.33		− .75		+2.10		− 7.00		+26.00	
26	+ .33		−.29		+ .43		− 1.02		+ 4.00		
		+.04		− .32		+1.08		− 3.00			
27	+ .38		−.61		+1.51		− 4.01				
		−.57		+1.18		−2.93					
28	− .20		+.57		−1.43						
		−.00		− .24							
29	− .20		+.33								
		+.33									
30	+0.13										

The properties (ii) and (iii) are emphasized if a true error, or "blunder,"[1] should occur, and form the basis of the method for the detection and location of such blunders by differencing.

These observations are readily explained by noting that each value of the function is the sum of a tabular entry and a rounding-off error. This rounding-off error forces the value tabulated to be a multiple of the unit of the final decimal, and is at most half of this unit in magnitude.

Table II shows the rounding-off errors E of Table I, with differences to the 10th. These differences were all obtained with three further figures and rounded off individually, and are within half a unit of the 7th decimal from the true value.

It will be noted that as the order of the differences increases beyond the 4th—that is, from the point in Table I where the differences begin to show obvious irregularity—the differences in Table II tend to be more and more nearly integral multiples of the 5th decimal unit, and to give values approaching zero more and more closely when added to the corresponding values in Table I. This confirms the expectation that the differences of the true values of the function tabulated continue to decrease, with increasing order of differences, beyond the order to which this decrease remains apparent in a 5-decimal table.

TABLE III

f		δ^2		δ^4	
0		0		0	
	0		0		0
0		0		0	
	0		0		$+\epsilon$
0		0		$+\epsilon$	
	0		$+\epsilon$		-5ϵ
0		$+\epsilon$		-4ϵ	
	$+\epsilon$		-3ϵ		$+10\epsilon$
ϵ		-2ϵ		$+6\epsilon$	
	$-\epsilon$		$+3\epsilon$		-10ϵ
0		$+\epsilon$		-4ϵ	
	0		$-\epsilon$		$+5\epsilon$
0		0		$+\epsilon$	
	0		0		$-\epsilon$
0		0		0	
	0		0		0
0		0		0	

3. An Isolated Error. 3.1. Consider next the effect of an isolated error or blunder ϵ. This is exhibited in Table III. This extends only to 5th differences, as it is sufficiently obvious that the coefficients are binomial coefficients with alternating signs.

3.2. The method for detecting, locating and evaluating a blunder in a table of exact values of a polynomial is now clear. A difference table is formed and the differences of some order p greater than n, the degree of the polynomial, are examined. These differences should all be zero; if, however, it is observed that some are not zero but alternate in sign and have magnitudes proportional to the binomial coefficients

$$\binom{p}{r} = \frac{p!}{r!(p-r)!}$$

then an isolated error is indicated.

If $p = 2k$ is even, there will be a pth difference that is numerically greater than the others; the blunder should be found[2] in the function value on a level with this. Denote this largest difference, with its sign attached, by D_{2k}. The value of the amount to be added to the function value in question, in order to correct it, is one of the values

$$+D_2/2,\ -D_4/6,\ +D_6/20,\ -D_8/70,\ +D_{10}/252,\ \cdots,\ (-1)^{k-1}D_{2k} \Big/ \binom{2k}{k}$$

It should be noted that these apply only when the corresponding difference D_{2k} should be zero, i.e., when $2k > n$, the degree of the polynomial concerned.

If $p = 2k + 1$ is odd, there will be two successive pth differences of equal magnitude, larger than the rest. The blunder should be found in the function value at the level half-way between these. If the upper of these differences is D_{2k+1} and the lower $-D_{2k+1}$, the correction to be added to the function value is one of the values (with $2k + 1 > n$)

$$-D_1,\ +D_3/3,\ -D_5/10,\ +D_7/35,\ -D_9/126,\ \cdots,\ (-1)^k D_{2k-1} \Big/ \binom{2k-1}{k}$$

3.3. Consider next a difference table involving a function that is not a polynomial. In this case the effect due to a true error or blunder, exceeding half a final unit, is mixed up with the effects of the rounding-off errors. Detection and location of the error involve the disentanglement of these effects.

For large blunders the method of detection is, with small modifications, the same as for a polynomial. A difference table is first formed to an order p of differences such that the *normal* vertical sequence of signs (i.e., the sequence of signs in a region free from the effects of blunders) has ceased to be regular. This means that the pth differences of the true function—which are regular—are swamped by the irregularities due to the rounding-off errors, that is, that the function differences are effectively zero.[3] Any sequence of differences having the numerically greatest difference substantially larger than the normal pth differences due to rounding-off will at once stand out, and the error indicated can be located and estimated in the same manner as in **3.2**, except that:

(i) The successive differences will be only approximately proportional to the binomial coefficients of order p.

(ii) Instead of estimating the correction from a single value D_p of the pth difference, it is better to add the numerical values of a sequence of differences, centered about the largest, and to divide by the sum of the corresponding binomial coefficients. It is also sometimes useful to repeat and verify the estimate with differences of a higher order, where true function differences will usually be smaller.

Table IV illustrates these various points.

An error is apparent in δ^2, but, away from the neighborhood of the error, the signs are regular in δ^3 and even in δ^4, as may be seen in Table I. In δ^5 the large differences are in the ratios -10, $+10$, -5. Thus, approximately, the correction C is given by $(10 + 10 + 5)C = 180 + 182 + 89 = 451$. Hence $25C = 451$ and $C = 18$. Hence, log 19 should read 1.27875, agreeing with Table I.

Use of a single value of δ^7 gives

$$(10 + 2\cdot 10 + 5)C = 180 + 2\cdot 182 + 89 = 633$$

giving, again, $C = 18$.

Choice of suitable difference to which to apply the process is determined by the equality of results from differences of two successive orders; this equality is taken to indicate that the variation of the result with increasing order of difference has ceased.

A process for filling in the gap that appears, at first sight, more satisfactory is to use LAGRANGE's interpolation formula based on tabular values but omitting, of course, the value needing correction. The gap should be as near the middle of the run of values as possible. If, however, p points are used, Lagrange's formula assumes that the p-th difference is zero; that is, the result will be precisely that obtained by equating to zero the appropriate p-th difference in order to determine the error.

TABLE IV

N	log N		δ^2		δ^4	
16	1.20412					
		+2633				
17	1.23045		−151			
		+2482		− 1		
18	1.25527		−152		+ 69	
		+2330		+68		−180
19	1.27857		− 84		−111	
		+2246		−43		+182
20	1.30103		−127		+ 71	
		+2119		+28		− 89
21	1.32222		− 99		− 18	
		+2020		+10		+ 14
22	1.34242		− 89		− 4	
		+1931		+ 6		
23	1.36173		− 83			
		+1848				
24	1.38021					

4. Disentanglement of Blunders from Rounding-off Errors. 4.1. It is desirable to know a lower limit to the size of blunder that can be detected with certainty, and an upper limit to the size of those blunders for which it is almost useless to attempt detection by differencing. Blunders intermediate in size may or may not be detected, depending on the run of neighboring rounding-off errors; it is useful to have an estimate of the probability of detecting such an intermediate blunder according to its size. A more immediate problem, however, is the determination of the probability that a difference of given order and given size is due solely to the effect of rounding-off errors. The complementary probability gives the likelihood of a blunder. These limits and probabilities depend, of course, on the order of the differences examined.

In practice the procedure is as follows: All the differences of a particular, sufficiently high, order are examined. Those numerically exceeding the limit L that indicates a blunder with certainty (there may be several such large differences in succession, of alternating signs, due to a single blunder) are noted and examined carefully in order to locate the blunder, which must then be removed. When all such blunders have been eliminated there remains a run of differences, none numerically larger than L, but which may have some entries, larger than the majority, that could arise from an unlikely combination of rounding-off errors, but have a good chance of being due to a small end-figure blunder. The nearer such a difference is to L in magnitude, the more nearly does the probability that it is due to a blunder approach unity. The problem, then, is to choose K such that all differences numerically greater than K should be examined, while all those numerically not greater than K may be accepted as satisfactory, being almost certainly due to rounding-off errors.

4.2. It is easy to determine the limit L above which a blunder is certainly indicated. The sequence of rounding-off errors giving rise to the greatest possible effect in the differences is $\cdots + \frac{1}{2}, -\frac{1}{2}, +\frac{1}{2}, -\frac{1}{2}, \cdots$ extending indefinitely in both directions. This leads to the sequence $\cdots, +2^{p-1}, -2^{p-1}, +2^{p-1}, -2^{p-1}, \cdots$ in the pth differences. The maximum rounding-off effect L in the pth difference is thus 2^{p-1} in magnitude.

On the other hand, if a blunder ϵ (assumed positive and not too small) is made in a tabular value, the case most unfavorable for detection comes from the sequence

$$(S) \qquad \cdots + \tfrac{1}{2}, -\tfrac{1}{2}, +\tfrac{1}{2}, \epsilon, +\tfrac{1}{2}, -\tfrac{1}{2}, +\tfrac{1}{2}, \cdots$$

The corresponding differences on a level with the error ϵ are given in the second line of Table V.

TABLE V

Difference	2nd	4th	6th	8th	10th	12th
From sequence (S)	$1-2\epsilon$	$6\epsilon-5$	$22-20\epsilon$	$70\epsilon-93$	$386-252\epsilon$	$924\epsilon-1586$
Numerically largest legitimate errors $\pm L$	-2	$+8$	-32	$+128$	-512	$+2048$
Maximum error $\epsilon_{\max}$ that can escape detection	3/2	13/6	27/10	221/70	449/126	1817/462
	1.50	2.17	2.70	3.16	3.56	3.93

If these are just equal, numerically, to the maximum legitimate values due to rounding-off (given in the third line of Table V, with the appropriate sign to give maximum ϵ), then the corresponding maximum errors that might just escape detection result. These are given in the fourth and fifth lines of the table. In fact, for order $2k$,

$$\epsilon_{\max} = -\tfrac{1}{2} + 4^k \Big/ \binom{2k}{k}.$$

Larger blunders cannot escape detection.

4.3. The limits L and the blunders $\epsilon_{\max}$ that may just escape detection are, however, sometimes too great to be of practical use. Differences, due entirely to rounding-off, with magnitude approaching L, are so rarely met with that the occurrence of such a difference is a strong reason for suspecting a blunder. It is necessary, then, to choose a different limit K, as indicated in **4.1.**

Satisfactory practical limits K have been obtained, from experience in the examination of many tables, by Dr. L. J. Comrie. These limits are very roughly such that about 1 difference in 100 exceeds K numerically and requires more careful examination, and are as follows:

Difference	3rd	4th	5th	6th	8th	10th	12th	15th
Practical limit K	3	6	12	22	80	300	1100	8000

The determination of exact theoretical probabilities for differences of various sizes is a matter of some difficulty. A theory and technique have been devised, but results are not yet complete.[4] If we wish to choose the limit K so that the chance of a difference arising from rounding-off errors that exceeds K is less than .01, while the chance of an error exceeding $K - 1$ is greater than .01, the following results are relevant.

Order of Difference	Num. Value of Difference	Chance of Occurrence	Order of Difference	Num. Value of Difference	Chance of Occurrence
2	$\geqslant 1$	0.5	6	$\geqslant 21$	0.0128
	$\geqslant 2$	0.0		$\geqslant 22$	0.0079
3	$\geqslant 3$	0.04	7	$\geqslant 41$	0.0108
	$\geqslant 4$	0.00		$\geqslant 42$	0.0084
4	$\geqslant 6$	0.0130	8	$\geqslant 79$	0.0111
	$\geqslant 7$	0.0009		$\geqslant 80$	0.0099
5	$\geqslant 11$	0.0140	9	$\geqslant 155$	0.0103
	$\geqslant 12$	0.0052		$\geqslant 156$	0.0097

The probabilities serve to show the consistency of the practical limits given above, and to provide additional limits of 42 for the 7th difference and 156 for the 9th difference.

4.4. It is not to be supposed that the limits K of the last section must be adhered to rigidly. The major field for use of these precise limits is for differencing tables with a final figure that should be correct within half a unit. In this case, the original calculations will contain one or more extra figures, and these extra figures should be used in the examination of the one doubtful case in 100 previously mentioned, in order to verify that the actual rounding-off errors that occur do give rise to a difference of the right sign and about the right size.

If a printed table is differenced, the extra figures may not be available, while if the function is one difficult to compute, and if the table is a long one, the work of recomputing values to test one difference in 100 may be prohibitive. In such cases it may be necessary to adopt a higher limit than K, possibly even L may have to be used, in which case one would state, for example, that an examination of the 8th differences showed that no isolated end-figure error of 3.2 units or more could occur in the table. The possibility of systematic or coupled blunders remains.

An alternative plan is to difference the function values as computed, retaining all figures computed, including one or more guard figures. It is then unnecessary to examine marginal blunders or errors too closely, and the limit L, or even higher limits such as $2L$ or $3L$, might be adopted. This procedure has the advantage that blunders large enough to need correction will stand out prominently.

5. Part I of this paper has been concerned with the location and detection of isolated blunders. There remain several possibilities to be discussed in Part II. These include:

(i) The recognition of blunders made during the differencing.

(ii) The detection and location of coupled or multiple blunders, such as

(a) two equal blunders in successive values or (b) a systematic succession of erroneous values in a table.

It is also proposed to give error patterns, such as that in Table III, for tables of *divided* differences, for use with tables having certain common arrangements of arguments at unequal intervals, for example, with a table having arguments

$$0, \tfrac{1}{4}, \tfrac{1}{3}, \tfrac{1}{2}, \tfrac{2}{3}, \tfrac{3}{4}, 1, 1\tfrac{1}{4}, 1\tfrac{1}{3}, \cdots$$

J. C. P. MILLER

Scientific Computing Service
23 Bedford Square
London, W.C. 1, England

[1] The introduction of this useful distinction in name between rounding-off and true errors is due to C. R. G. COSENS.

[2] It must be remarked that the sequences of errors discussed here can arise from a cause other than the one indicated, though such causes are comparatively less common. For instance, if differencing is done on a calculating machine, a function value may be correctly recorded, but wrongly set on the machine. Likewise, a different sequence of differences indicates blunders of a different type. It is hoped to discuss some of these in Part II of the paper.

[3] In practice, a large blunder shows up well enough for location in earlier orders of differences, in fact, as soon as the largest of the differences due to the blunder sufficiently exceeds the true differences in magnitude, say in the ratio 5 to 1 or 10 to 1. Detection is possible in still earlier differences.

[4] A. VAN WIJNGAARDEN & W. L. SCHEEN of the Mathematisch Centrum of Amsterdam, Holland, have developed the theory independently and have obtained an asymptotic expansion. The result given for 9-th differences in our table was obtained by them and communicated to us for inclusion in this paper. Their 1 percent limit for 10-th differences is 303.

An ENIAC Determination of π and e to more than 2000 Decimal Places

Early in June, 1949, Professor JOHN VON NEUMANN expressed an interest in the possibility that the ENIAC might sometime be employed to determine the value of π and e to many decimal places with a view toward obtaining a statistical measure of the randomness of distribution of the digits, suggesting the employment of one of the formulas:

$$\pi/4 = 4 \arctan 1/5 - \arctan 1/239$$
$$\pi/4 = 8 \arctan 1/10 - 4 \arctan 1/515 - \arctan 1/239$$
$$\pi/4 = 3 \arctan 1/4 + \arctan 1/20 + \arctan 1/1985$$

in conjunction with the GREGORY series

$$\arctan x = \sum_{n=0}^{\infty} (-1)^n(2n + 1)^{-1}x^{2n+1}.$$

Further interest in the project on π was expressed in July by Dr. NICHOLAS METROPOLIS who offered suggestions about programming the calculation.

Since the possibility of official time was too remote for consideration, permission was obtained to execute these projects during two summer holiday week ends when the ENIAC would otherwise stand idle, and the planning and programming of the projects was undertaken on an extra-curricular basis by the author.

The computation of e was completed over the July 4th week end as a