

MISCELLANEA.

I. Some Notes on Interpolation in n -dimension Space.

By W. PALIN ELDERTON.

1. *Introductory.* The main idea of this paper is summed up in the principle, or the extension of the principle, that in certain cases the use of linear second differences will give a better result than the use of two-variable first differences, e.g. three points in the table chosen on a straight line may give a more accurate interpolation than the four "nearest" points. The usual formulae of interpolation assume that we require to find from tabulated values of u_x a value of the function corresponding to a certain value of x not among the tabulated cases. In some circumstances we have to deal with a function of two variables, and it was recently shown by Mr John Spencer* that in certain circumstances the simple one-variable interpolation formulae could be used. This saves a great deal of work in practice, and in the examples dealt with it was found to give accurate results. The object of the present paper is to supply a general method and to show that it is not necessary to limit the number of variables to two. Before proceeding to the subject itself it will be well to outline the alternative methods available. The most usual practice in two-variable interpolation is to interpolate between four values; thus if u_{rs} is required and u_{00} , u_{01} , u_{10} and u_{11} are given, then u_{r0} is found from u_{00} and u_{10} , and u_{r1} from u_{01} and u_{11} , and then u_{rs} from u_{r0} and u_{r1} . Graphically these processes might be represented by the following diagram, in which the dots show the positions of the given values, the crosses the positions of the interpolated values, and the lines connecting the points indicate the values used in each interpolation. Of course to give the values in a figure we should have to erect perpendiculars to the surface of the paper.

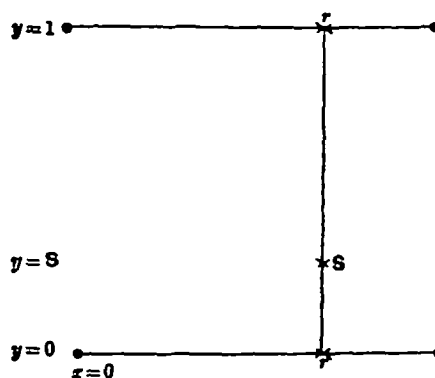


FIG. 1.

The general solution in terms of differences is obtained by expanding the right-hand side of

$$u_{rs} = (1 + \Delta_x)^r (1 + \Delta_y)^s u_{00},$$

* John Spencer, "Some practical hints on two-variable interpolation," *Journal of the Institute of Actuaries*, Vol. XL. pp. 293 et seq.

but it is, I think, better to use coefficients* instead of differences, just as one does when employing Lagrange's formula in the ordinary one-variable interpolation, and with the help of auxiliary tables such as those of Mr Herbert H. Edwards† very accurate results can be easily obtained. If however the intervals by which the given table of the function proceeds are different from those adopted by Mr Edwards, a large amount of arithmetical work is necessary. It is however worth while to recapitulate the method and increase the number of the independent variables from two to n . Assuming that we require to find $u_{x_{rel}y_{rel}}$; then by Lagrange's interpolation formula we have

$$u_{x_{rel}y_{rel}} = r_0 u_{0x_{rel}} + r_1 u_{1x_{rel}} + r_2 u_{2x_{rel}} \text{ etc.},$$

where r_0, r_1, \dots are numerical coefficients depending on the number of terms and the value of r . And similarly

$$u_{x_{rel}y_{rel}} = s_0 u_{0x_{rel}} + s_1 u_{1x_{rel}} + s_2 u_{2x_{rel}} \text{ etc.},$$

$$u_{x_{rel}y_{rel}} = t_0 u_{0x_{rel}} + t_1 u_{1x_{rel}} + t_2 u_{2x_{rel}} \text{ etc.},$$

and so on; hence

$$u_{x_{rel}y_{rel}} = r_0 s_0 t_0 \dots u_{000\dots} + r_0 s_0 t_1 \dots u_{001\dots} + \text{etc.} \\ = S(r_0 s_0 t_1 \dots u_{001\dots}),$$

where S stands for the summation of all combinations of λ, k, l, \dots for all values from 0 upwards.

The calculation of all the $r_\lambda, s_k, t_l, \dots$ terms is done directly from Lagrange's formula and is simple though laborious, and the remainder of the work is mechanical. When the function only depends on two variables the interpolation involves fairly heavy arithmetic if only three values are given to both λ and k , while if there are three variables the work could only be completed in two or three hours, and it is therefore clear that if the ordinary one-variable interpolation formula could be used we should save a great deal of arithmetical work.

The ideas underlying the two methods can be seen graphically.

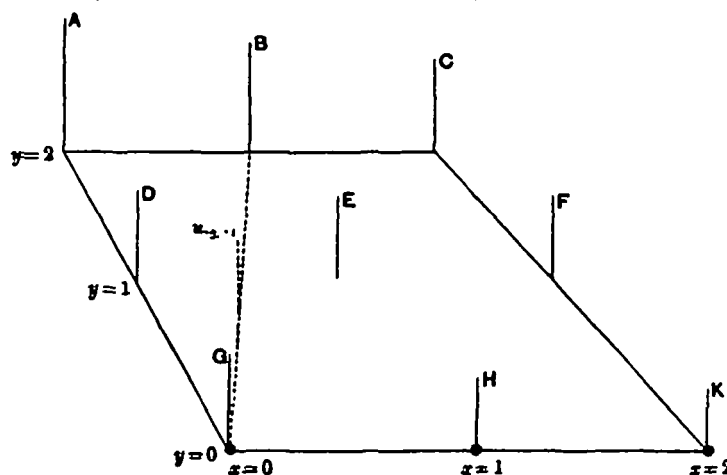


FIG. 2.

In this figure a surface $ACFG$ represents u_{xy} , so that when $x=0, y=2$, u_{xy} has the value represented by the height A , and when $x=1, y=1$, u_{xy} has the value represented by E . Now if we require to find $u_{2.4}$ we can approximate to the surface by the $(1+\Delta_x)(1+\Delta_y)$ method or by the Lagrange coefficient method and hence find the particular height we require, or we can choose the positions of two or more values so that the straight line passing through them also passes through the point $x=2, y=4$ and then interpolate by the one-variable formulae between the values of the function corresponding to the points on this straight line. In fig. 2 the line joining $x=0, y=0$ to $x=1, y=2$ passes through $x=2, y=4$ and by interpolation between B and G we can therefore approximate to $u_{2.4}$. The dotted lines show the process.

* W. Palin Elderton in *Biometrika*, Vol. II. Part 1.

† *Journal of the Institute of Actuaries*, Vol. XL. pp. 289–293.

2. *The Interpolation Line Method.* We may now attack the theory of this second method in greater detail.

Let us assume that from the following tabulated terms it is required to find $u_{h,k,l,\dots}$:

$$\begin{array}{lll} u_{0,0,0,\dots}, & u_{r,0,0,\dots}, & u_{2r,0,0,\dots}, \text{ etc.}, \\ u_{0,h,0,\dots}, & u_{r,h,0,\dots}, & u_{2r,h,0,\dots}, \text{ etc.}, \\ \text{etc.}, & \text{etc.} \end{array}$$

This scheme is general and there is no loss of generality in assuming that the interpolated values are for integers, i.e. if the function is tabulated for $x=0, 5, 10$, etc., $y=0, 8, 16$, etc., $z=0, 2, 4$, etc., the interpolated values required would be for $x=1, 2, 3, 4$, etc., $y=1, 2, 3$, ..., $z=1, 3, 5$, ...; because if fractional values are required we can assume that the x 's, y 's, z 's, etc. proceed by larger differences.

Now in order to get u_h we take $\frac{h}{r}$ and use the formula $(1+\Delta)^{\frac{h}{r}} u_0 = u_h$, or we can take $\frac{h}{2r}$ and use $(1+\Delta)^{\frac{h}{2r}} u_0 = u_{\frac{h}{2}}$, where $\Delta = u_r - u_0$ and $\Delta' = u_{2r} - u_0$. Again to get $u_{r+\frac{h}{2}}$ we take $(1+\Delta)^{\frac{h}{2r}} u_0$ or $(1+\Delta')^{\frac{h}{2r}} u_0$ or etc., where Δ and Δ' have similar meanings.

But if $\frac{h}{r}$ be used for a two-variable, then the interpolated function found could be $u_{x+h, y+\frac{h}{r}}$, and therefore in order to get terms from which interpolation can be made we

must use, not u_{00} , u_{rs} , etc., but u_{00} , $u_{rs,as}$, so that when we take $(1+\Delta_a)^{\frac{h}{ra}} u_{00}$ we shall get $u_{h,k}$. It is fairly clear that if we can fix either ra or as the other can be obtained easily, and a little consideration of the way the ra and as arise will show the reader that the term can be fixed by the least common multiple of h , r , k and a .

A numerical example will make this clearer. Let $u_{0,0}$, $u_{4,8}$, $u_{10,16}$, etc. be given and assume that we require $u_{2,3}$; then the L.C.M. of 2, 3, 5, 8 being 120 we must interpolate between $u_{0,0}$, $u_{20,120}$, $u_{100,240}$, ..., for having fixed the 120, the proportion to be taken is $\frac{1}{120} = \frac{1}{120}$ and $2 \times 40 = 80$. If a third variable is introduced the same method can be used, and if in our last numerical example we take $u_{0,0,0}$, $u_{4,8,10}$, $u_{10,16,20}$, ... as given, and require $u_{2,3,7}$, we can use $u_{0,0,0}$, $u_{240,300,840}$, $u_{480,720,1680}$, etc. The numerical work would be as follows in an imaginary example:

$$\begin{array}{rcl} & \Delta & \Delta^2 \\ u_{0,0,0} & = 4872 & -2760 \quad +1656, \\ u_{240,300,840} & = 2112 & -1104, \\ u_{480,720,1680} & = 1008. \\ \therefore u_{2,3,7} & = 4872 - \frac{1}{120} \cdot 2760 - \frac{1}{2} \cdot \frac{1}{120} \cdot \frac{1}{2} \cdot 1656 \\ & = 4832. \end{array}$$

There are, of course, many objections to such a method in practice when it entails differences in the independent variable, such as 240, 360 and 840, but it can often be simplified considerably; for instance, in the last example, if we imagine the given values to extend to negative values of x , y and z , we can start from $u_{-5,0,10}$ and seek $u_{2,3,7}$. We notice that our fractions are now $\frac{2}{5}$, $\frac{3}{5}$, $\frac{7}{10}$ and the L.C.M. is reduced to 120. The terms to be used are $u_{-5,0,10}$, $u_{-115,120,-110}$ and $u_{-235,240,-230}$, and the value of $u_{2,3,7}$ is

$$u_{-5,0,10} + \frac{2}{5}(u_{-115,120,-110} - u_{-5,0,10}) - \frac{1}{2} \cdot \frac{2}{5} \cdot \frac{3}{5} \text{ (second diff.)}$$

In this way we can sometimes diminish the range of values considerably. Even here however a large table would be required for practical work, but it must be borne in mind that a three-variable table is very seldom made, and a two-variable table is frequently simplified by both the x 's and y 's proceeding by the same difference. The difficulty is however always present to some extent and it would be impossible to find values of u corresponding to a number of decimal places in x , y , ..., though the method might even in these cases be of help as a rough check.

Graphically these examples can be shown very simply by the following figures. Using dots to show the position of given values, then, if the required value is $u_{1,6}$, we join $(0, 0)$ to $(2, 6)$ and produce it till it passes through another point (x', y') , then we can interpolate for $u_{1,6}$ between $u_{0,0}$, $u_{x',y'}$, $u_{2x',2y'}$, etc.

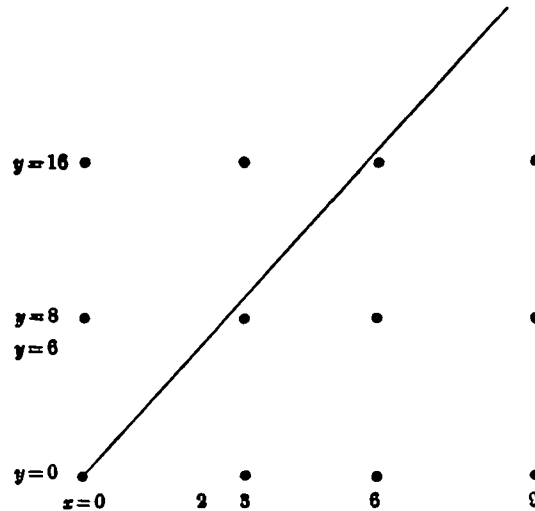


FIG. 3.

In a similar way if there are three variables the following diagram may be constructed, in which dotted lines show those parts of the figure that would not be viable if we had a solid, and the dots show the positions of given values; the "line of interpolation," as it might be called, is OB .

In fig. 2 we have represented the surface by giving the heights of the ordinates at the given points, and the same could have been done in fig. 3, but it would have been impossible in fig. 4

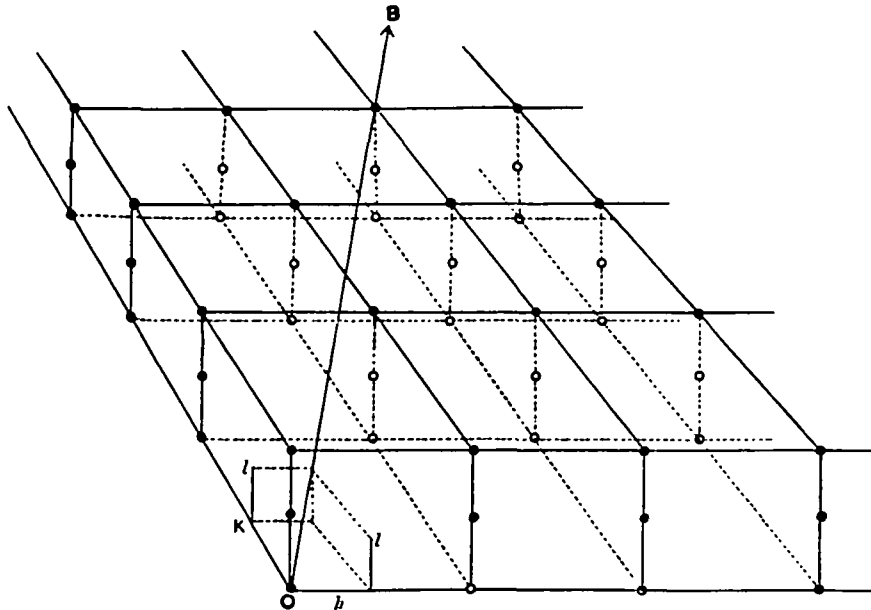


FIG. 4.

because as we are there dealing with a third-dimension body we should require a fourth dimension in order to enable us to represent the values of u_{xyz} , as well as their positions. For the same reason we cannot give a graphical illustration of interpolation with four variables.

3. *Some particular cases.* It will probably be useful to give the "interpolation-lines" in the case of a function of two variables where the values proceed by small differences.

We shall assume that only integral values are required, and in our diagrams a heavy dot will stand for the position of a given value and a small dot for the position of a value obtained by interpolation, while the lines show the "interpolation-lines."

Difference in $x = 3$.

" " $y = 3$.

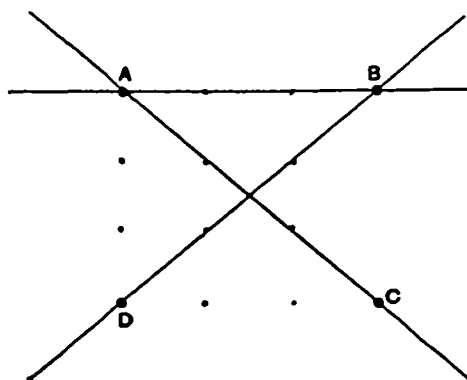


FIG. 5.

In this case the line between AB is an obvious interpolation, and similar lines will be neglected in other diagrams; in fact there is only one "interpolation-line," since AC and BD are the same in principle; they have merely different starting-points.

Difference in $x = 4$.

" " $y = 4$.

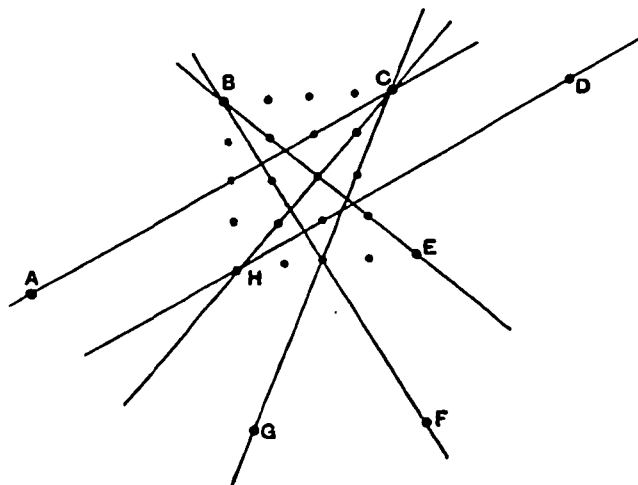


FIG. 6.

In this case there are only two "interpolation-lines," viz. BE and AC , for CH is the same as BE and HD , BF and CG are all the same as AC with different starting-points; in fact we can in general deal with any half quarter of the square $BCEH$ (i.e. half of any one of the triangles formed by two diagonals and a side) and use the results as applicable to all cases.

Example. If we wish to interpolate for $u_{2,1}$ we should take one quarter of the way between $u_{0,0}$ and $u_{4,4}$; if a third term were required we should take either $u_{1,3}$ or $u_{3,1}$, and the latter would generally be preferred as it is nearer the value required.

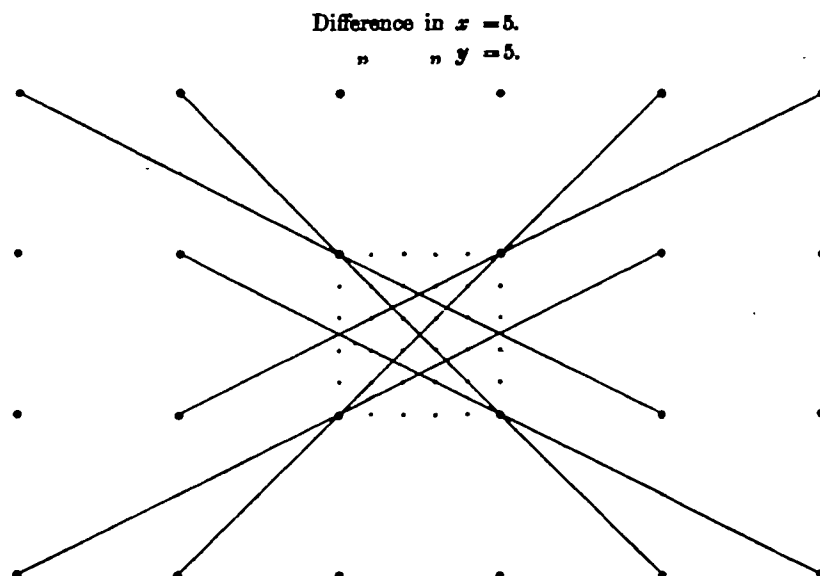


FIG. 7.

This is a very common case and there are only two necessary "interpolation-lines." It will be noticed that there is sometimes more than one line that could give the interpolation, but I have shown the one most likely to give a good result.

Difference in $x = 3$.
" " $y = 2$.

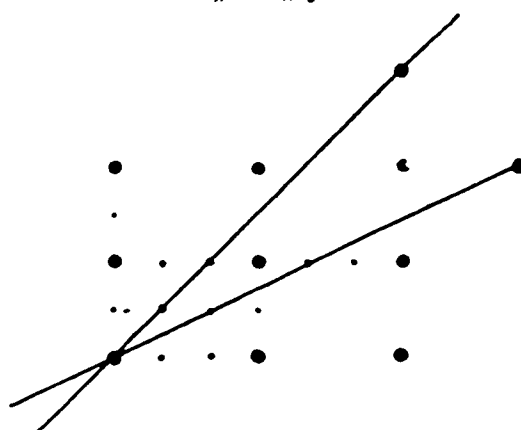


FIG. 8.

In this case in order to obtain an interpolated value for $u_{2,1}$ we require $u_{0,0}$, $u_{4,4}$ and $u_{-4,-4}$; clearly a less useful tabulation than differences of 3 in both x and y , and not much improvement (sometimes it is worse) than differences of 4 in both x and y . The lesson to be learnt from this is that if one can choose one's differences in tabulation it is best to take the same difference in both x and y .

Difference in $x = 10$.

" " $y = 10$.

The diagram in this case becomes rather troublesome, but the following table answers the same purpose.

x	y	Interpolate between	Taking $\frac{\pi}{10}$ of distance where π is
1	1	0, 0 and 10, 10 I	1
1	2	0, 0 and 10, 20 II	1
1	3	0, 0 and 10, 30 III	1
1	4	10, 10 and -20, -10 IV	3
1	5	0, 0 and 10, 50 V	1
2	2	0, 0 and 10, 10 I	2
2	3	0, 0 and 20, 30 IV	1
2	4	0, 0 and 10, 20 II	2
2	5	0, 0 and 20, 50 VI	1
3	3	0, 0 and 10, 10 I	3
3	4	0, 10 and 10, -10 II	3
3	5	0, -10 and 10, 40 V	3
4	4	0, 0 and 10, 10 I	4
4	5	10, -10 and -10, 40 VI	3
5	5	0, 0 and 10, 10 I	5

This would of course be applicable where one decimal place is required and the tabulation is for integral values, but as will be noticed immediately, the number of independent "interpolation-lines" (see Roman numbers) is very great, and it is not very difficult to show that the decimal system is by no means the best method of tabulating functions. We shall however return to this after showing the application of the above diagrams to numerical work.

4. *Numerical Examples.* The following table gives the values of a function of three-variables, which will enable us to give a few examples.

z	$x = 45$				$x = 50$				$x = 55$				$x = 60$				
	y				y				y				y				
	10	15	20	25	10	15	20	25	10	15	20	25	10	15	20	25	30
5	7.24	9.51	12.74	17.71	5.74	7.39	9.60	12.74	4.69	5.93	7.53	9.65	3.95	4.94	6.14	7.68	9.71
10	16.06	21.04	28.20	39.36	12.70	16.27	21.11	28.11	10.34	13.01	16.45	21.13	8.69	10.77	13.33	16.62	21.19
15	26.76	35.03	47.09	66.18	21.08	26.94	35.01	46.90	17.09	21.41	27.06	34.94	14.30	17.61	21.74	27.23	34.97
20	39.73	52.12	70.49	100.00	31.13	39.81	51.99	70.27	25.09	31.39	39.84	51.84	20.86	25.61	31.70	39.95	51.74

Let us first find a value for $x=45$, $y=17$, $z=12$; this only involves two variables, and fig. 7 tells us to interpolate between 45, 10, 5 and 45, 15, 10, etc.; hence the interpolated value required is obtained by ordinary interpolation in the following way:

	Δ	Δ^2	Δ^3
7.24	13.80	12.25	13.61
21.04	28.05	26.88	
47.09	52.91		
100.00			

therefore value is

$$7.24 + 1.4 \times 13.80 + \frac{(1.4)(.4)}{2} 12.25 - \frac{(1.4)(.4)(.6)}{2.3} 13.61 = 29.13,$$

while with only three terms we should have obtained 29.89. If we had used the ordinary double-interpolation method with the terms 21.04, 28.20, 47.09 and 35.03, we should have obtained 30.28, and as the true result is 29.48, we have in this case obtained a better approximation by the "interpolation-line" method. The data are not good for interpolation because the differences do not decrease, but this is no disadvantage for our present purpose.

As a second example we will take $x=60$, $y=18$, $z=11$, and we can use 14.30, 13.33 and 9.71 and obtain

$$14.30 - .8 \times .97 + \frac{(.8)(.2)}{2} \times 2.65 = 13.73,$$

and the true value is 13.69. Two terms only would have given 13.58. With a little practice the values to be used can be easily picked out.

We will take as our final examples two three-variable cases. To obtain $x=49$, $y=18$, $z=14$ we can use 45, 20, 10 and 55, 15, 20 and take two-fifths of the difference, i.e. $28.20 + \frac{2}{5}(31.39 - 28.20) = 29.48$, or we can use 50, 20, 15 and 45, 10, 10 thus getting 31.21 while if we add 55, 30, 20 which is 70.02 we get 29.93; the true value is 30.31. To obtain $x=52$, $y=14$, $z=14$ we can use 50, 10, 15 and 60, 30, 10 and obtain 21.10, or we can take 50, 15, 15 and 60, 10, 10 but the former is distinctly preferable and if we add the term 70, 50, 5 which is 17.63 it becomes practically exact.

The latter arrangement, however, shows the method at its worst. The differences run extremely awkwardly and the interpolation by first differences leads to an untrustworthy result. When using the "interpolation-line" method we are, as it were, picking out a level piece of ground on our surface before interpolating; if we choose an uneven piece of ground, i.e. if the differences are large, the result will be inaccurate.

5. *Connection of Method with the Construction of Tables.* An interesting point of some practical importance may now be examined. If we were to make a new two-variable table, what would be the best interval to use if we intended to apply the "interpolation-line" method?

We must clearly use an interval so that the terms used in the interpolations are not far apart, and in order to answer the question it becomes necessary to consider what intervals are most satisfactory from this point of view. We will first see how the various interpolations can be built up and will take the interval 5 to start with. The terms we require to find by interpolation are 1, 1; 1, 2 and 2, 2, because 1, 3 is just the same as 1, 2 but with a different starting point, for it is merely distance 1, 2 from the given term 0, 5. There are therefore only two lines necessary, one symbolised by 1, 1 and showing interpolation between terms like 0, 0 and 5, 5, and the other by 1, 2. The former gives 2, 2; we should have to take two-fifths of the distance between the 0, 0 and 5, 5 given terms instead of the one-fifth that gave 1, 1. Now let us take 7 as an interval. The terms wanted are 1, 1; 1, 2; 1, 3; 2, 2; 2, 3; 3, 3. The 1, 1 line gives 1, 1; 2, 2 and 3, 3; and the 1, 2 line gives 1, 2 and 2, 4, but the second of these is the same as 2, 3 with a different starting-point, and we only therefore require two lines. The idea is easily continued,

and it follows that if we use a prime number (p) as our interval, we get with each interpolation line $\frac{p-1}{2}$ cases; but as we require

$$\frac{p-1}{2} + \left(\frac{p-1}{2} - 1\right) + \left(\frac{p-1}{2} - 2\right) + \dots = \frac{\left(\frac{p-1}{2}\right)\left(\frac{p-1}{2} + 1\right)}{2}$$

cases in all, the number of interpolation-lines required is $\frac{1}{2}\left(\frac{p-1}{2} + 1\right)$. This gives the following table:

Interval	2	3	5	7	11	13	17	19
No. of lines	1	1	2	2	3	4	5	5

Turning from prime numbers to those containing factors we find at once that many more lines are required. Thus we have already seen that an interval of 10 calls for no fewer than six interpolation-lines, and the intervals of 6 and 8 will both be found to require four. The interval 9 requires three lines.

Let us now compare the 11 interval with the 8 and 10 intervals and see which is the best grouping.

Interval 11.

Ordinary interpolation gives	Line 1, 1 gives	Line 1, 2 gives	Line 1, 3 gives
0, 1	1, 1	1, 2	1, 3
0, 2	2, 2	2, 4	2, 6=2, 5
0, 3	3, 3	3, 6=3, 5	3, 9=3, 2
0, 4	4, 4	4, 8=4, 3	4, 12=4, 1
0, 5	5, 5	5, 10=5, 1	5, 15=5, 4

Average interval in interpolation in $x = 8.25$

" " " " " $y = 15.4$

Maximum interval in ... $x = 11$

" " " " " $y = 33$

Interval 10.

Ordinary interpolation gives	Line 1, 1 gives	Line 1, 2 gives	Line 1, 3 gives	Line 1, 4* gives	Line 1, 5 gives	Line 2, 5 gives
0, 1	1, 1	1, 2	1, 3	1, 4	1, 5	2, 5
0, 2	2, 2	2, 4	2, 6=2, 4	{also by 1, 2 given above	2, 8=2, 2	
0, 3	3, 3	3, 6=3, 4	3, 9=3, 1		3, 12=3, 2	3, 5
0, 4	4, 4	4, 8=4, 2				6, 5=4, 5
0, 5	5, 5	which has already been obtained				

Average interpolation interval in $x = 8.5$ or 9.5

" " " " " $y = 23.5$ or 22.5

Maximum interval in ... $x = 20$

" " " " " $y = 50$

* 2, 3 would give the same results and the averages resulting in each case are given.

Interval 8.

Ordinary interpolation gives	Line 1, 1 gives	Line 1, 2 gives	Line 1, 3 gives	Line 1, 4 gives
0, 1	1, 1	1, 2	1, 3	1, 4
0, 2	2, 2	2, 4	2, 6=2, 3	
0, 3	3, 3	3, 6=3, 2		
0, 4	4, 4			

Average interval in $x = 5.71$

" " " $y = 13.71$

Maximum interval in $x = 8$

" " " $y = 32$

The interval 9 gives quite a good result, certainly better than 8, but considering the fewer number of values that would be required 11 seems preferable to either, and it is far better than the common decimal interval. When the number representing the interval has factors the "interpolation-lines" do not give as many values as they do when a prime number is used because, owing to the factors, there is repetition.

6. *Conclusions.* The conclusions to which these notes lead us would seem to be

- (1) that interpolation can be effected by means of the ordinary one-variable formulae in n -dimension tables;
- (2) that the method can give reasonably accurate results;
- (3) that if the intervals can be chosen, (a) the same interval in all the variables will generally be better than different intervals even if the latter are rather smaller, and (b) the interval should be given by a prime number.

II. Note on the Relative Variability of the Sexes in *Carabus auratus*, L.

By H. G. KRIBS.

(From the Zoological Laboratory of the University of Pennsylvania.)

The majority of biometrical investigations, which have been made regarding the comparative variability of the sexes as such, seem to have been confined to the human species*. Many observations have been made among lower organisms however, on the existence and variability of secondary sexual characters, or those morphological differentiations, apart from the sexual organs *per se*, by which the male is easily distinguished from the female of the same species. The possible significance of secondary sexual characters for evolution was first pointed out by Darwin†, as suggested by the disproportionate development of these characters in the males and females of the same species where sexual dimorphism is present, and by the difference in behaviour seemingly correlated with it. The data furnished by Darwin covered a wide range in the animal kingdom, and his conclusions have received very general acceptance.

One conclusion reached by Darwin was that in a bisexual species the female ordinarily lies closer to the morphological norm of the species than the male, and that her contribution in reproduction is inherently conservative and tends to maintain the organic stability of the norm. The male, on the other hand, was thought to vary considerably more about the species, norm.

[* The relative variability of the sexes has been dealt with in crabs, wasps, toads and a variety of other published cases, not hitherto collected together, but all tending in much the same direction as in *Carabus auratus*. Ed.]

† *Descent of Man*, Chap. VIII.