JIA 72 (1946) 0482-0497

SOME NOTES ON INTERPOLATION

BY HUBERT VAUGHAN, F.I.A.

General Secretary and Actuary of the Mutual Life and Citizens' Assurance Company, Ltd., Sydney, Australia

I. THE CONNEXION BETWEEN CERTAIN INTERPOLATION AND GRADUATION FORMULAE

IT was shown by G. F. Hardy (*J.I.A.* Vol. XXIII, p. 351) that Woolhouse's system of interpolations, used in graduating the H^M Table, can be expressed in summation form; and other interpolation methods, such as Sprague's osculatory formula, were later made the basis of summation formulae of graduation. In the *Transactions of the Second International Actuarial Congress*, p. 78, Karup sketched a method by which a more general relation might be established; and this is further developed by Lidstone in *J.I.A.* Vol. XLII, p. 134.

The relation between interpolation and graduation by summation is, however, closer than may be generally realized. To illustrate this we show below the numerical coefficients when an interval is subdivided by five, using Lagrange's formula taken centrally:

$$\begin{aligned}
u_3 &= \cdot28u_0 + \cdot84u_5 - \cdot12u_{10},\\
u_4 &= \cdot12u_0 + \cdot96u_5 - \cdot08u_{10},\\
u_5 &= \;0 \quad + 1\cdot00u_5 + \;0,\\
u_6 &= -\cdot08u_0 + \cdot96u_5 + \cdot12u_{10},\\
u_7 &= -\cdot12u_0 + \cdot84u_5 + \cdot28u_{10},\\
u_8 &= \qquad\qquad \cdot28u_5 + \cdot84u_{10} - \cdot12u_{15},\\
u_9 &= \qquad\qquad \cdot12u_5 + \cdot96u_{10} - \cdot08u_{15},
\end{aligned}$$

and so on.

Inspecting the column showing the successive multipliers applied to u_{10}, it will be seen that u_{10} first enters into the calculation when it is multiplied by $-\cdot12$ to ascertain u_3. It is then, in the calculation of further interpolated values, multiplied successively by $-\cdot08$, zero, $+\cdot12$, etc. It is last used in ascertaining u_{17}, for which purpose it is again multiplied by $-\cdot12$. Exactly the same multipliers, fifteen in number, are applied successively to each of the given terms u_0, u_5, etc. These multipliers, written in the order of their appearance, are:

$$\ldots+0+0-\cdot12-\cdot08+0+\cdot12+\cdot28+\cdot84+\cdot96+1\cdot00+\cdot96+\cdot84+\cdot28$$
$$+\cdot12+0-\cdot08-\cdot12+0+0+\ldots.$$

This set of coefficients, especially if graphed out, will have a familiar appearance —that of the expanded coefficients of a summation formula of graduation. This is not surprising because they are in fact the coefficients of Woolhouse's formula multiplied by five.

It follows that this graduation formula could, if so desired, be used as an interpolation formula for the subdivision of intervals by five. If we write down every fifth term of a series, filling in the gaps with zeros, and apply Woolhouse's graduation formula multiplied by five, we shall be using only three of the expanded coefficients in ascertaining each value, the remainder being multiplied by zeros. In ascertaining u_3, for example, the only effective coefficients will be the fifth, tenth and fifteenth, viz. ·28, ·84 and −·12, which are the Lagrange coefficients as shown in the second paragraph above.

It is not merely the case that the graduation formula can be connected with the interpolation formula. In effect it is the interpolation formula. To show that this is general, a preliminary proposition will now be enunciated.

If u_x is a polynomial of the $(n-1)$th order (i.e. if nth differences vanish) and the operation $[k]^n$ is applied to a series consisting of every kth value of u_x with zeros interpolated, the result multiplied by k will exactly equal the result obtained by applying the same operation to the complete series of u_x.

For convenience a_x will be used to denote the terms of the series

$$u_0,\ 0,\ 0,\ 0,\ \ldots,\ u_k,\ 0,\ 0,\ \ldots,\ u_{2k},\ 0,\ \text{etc.},$$

so that a_x equals u_x when x is divisible by k, and otherwise $a_x = \text{zero}$.

δ is used to indicate the central difference for an interval of one, and δ_k the central difference for an interval of k. $[k]$ means, as usual, the sum of k consecutive terms taken centrally.

For any series of arithmetic numbers whatever, $\delta^n [k]^n \equiv \delta_k^n$. This is the symbolic expression of an identity arising from the nature of the operations. It is obvious that $\delta [k] \equiv \delta_k$, and the identity still holds if each operation is performed any number of times.

Now $\delta_k^n a_x$ is always zero, because for some values of x we are operating on terms of a polynomial of the $(n-1)$th order, and for others only on a series of zeros. Hence $\delta^n [k]^n a_x$ also is always zero, so that $[k]^n a_x$ must be a polynomial and of an order not higher than the $(n-1)$th. (It is, in fact, of the $(n-1)$th order, but it is not necessary to the proof to establish this.) We can therefore say that $[k]^n a_x$ may be expressed in the form $c_0 + c_1 x + \ldots + c_{n-1} x^{n-1}$, where the c's are the same for all values of x.

If now we apply the same operator $[k]^n$ to the series u_1, u_{k+1}, u_{2k+1} with zeros interpolated, the same reasoning holds good. The same operations are applied to a series of exactly the same mathematical form, $(x+1)$ taking the place of x. We must therefore reach a polynomial of the same form, with the same coefficients c_0, c_1, etc. In other words, the result of operating on u_1, u_{k+1}, etc., is identical with the result obtained from u_0, u_k, etc.

We can therefore say that

$$\begin{aligned} & [k]^n (u_0 + 0 + 0 + \ldots + u_k + 0 + \ldots + u_{2k} + \text{etc.}) \\ = {} & [k]^n (0 + u_1 + 0 + \ldots + 0 + u_{k+1} + \ldots + 0 + u_{2k+1} + \text{etc.}) \\ = {} & [k]^n (0 + 0 + u_2 + \ldots + 0 + 0 + u_{k+2} + \text{etc.}) \\ = {} & \text{etc.} \end{aligned}$$

Adding k successive expressions of the above form, the law of distribution applies, so that the total is

$$[k]^n (u_0 + u_1 + u_2 + u_3 + \text{etc.}),$$

i.e. the series of $[k]^n u_x$.

This proves our theorem that the result of retaining only every kth term is to yield exactly $1/k$th of $[k]^n u_x$.

To illustrate numerically what has been proved so far, we show in Table A the result of summing three times in fives a series of the second order, selected arbitrarily. On the right of the table the same summations are applied to five times the quinquennial terms with o's inserted between. It will be seen that the columns of $[5]^3$ in each case are identical.

Table A

Series	[5]	$[5]^2$	$[5]^3$	Series	[5]	$[5]^2$	$[5]^3$
3				15			
1				0			
0	5			0	15		
0	5			0	15		
1	10	75		0	15	75	
3	20	125		15	15	75	
6	35	200	1125	0	15	200	1125
10	55	300	1625	0	15	325	1625
15	80	425	2250	0	140	450	2250
21	110	575	3000	0	140	575	3000
28	145	750		140	140	700	
36	185	950		0	140	950	
45	230			0	140		
55	280			0	390		
66				0			
78				390			

From the foregoing the connexion between interpolation and summation formulae follows easily.

It is known, from the work of G. F. Hardy, Todhunter and Higham, that with any given series of summations we can combine an 'operand' so that the result will reproduce a polynomial of any given order. For equal summations we have

$$k^n u_x = [k]^n \left(1 - \frac{n(k^2-1)}{24}\delta^2 + \text{etc.}\right) u_x,$$

which, in the historic case of Woolhouse's formula, becomes

$$125 u_x = [5]^3 (1 - 3\delta^2) u_x.$$

If we ascertain the consecutive terms of a polynomial series of the second order passing through u_0, u_5 and u_{10}, sum these terms in fives three times and apply the operand $(1-3\delta^2)$, dividing by 125, we shall reproduce the series. But the third sum in fives, obtained from u_0, u_5 and u_{10} with added zeros, is exactly one-fifth of that obtained as above, so that we can obtain the five central terms of the same series by applying the formula to

$$0,\ 0,\ 0,\ 0,\ u_0,\ 0,\ 0,\ 0,\ 0,\ u_5,\ 0,\ 0,\ 0,\ 0,\ u_{10},\ 0,\ 0,\ 0,\ 0,$$

dividing by 25 in lieu of 125. Seeing that only one polynomial curve of the second order will pass through three given points, the results so obtained must be identical with those of a central second-difference interpolation.

The preceding paragraph is for simplicity written with reference to the particular case of Woolhouse's formula, but the argument is general. It is known that a summation formula based on interpolations correct to an even order of differences will be correct to the next order, as the summation formula (applied to the whole series) takes terms equally from each side so that the coefficients of the odd orders of differences cancel out. Remembering this, we can at once write down the graduation formulae equivalent to subdividing an interval by ordinary finite difference methods. For example, for the interval 5:

$25u_x = [5]^2 u_x$	corresponds to 1st-difference interpolation,
$125u_x = [5]^3(1-3\delta^2)u_x$	corresponds to 2nd-difference interpolation,
$625u_x = [5]^4(1-4\delta^2)u_x$	corresponds to 3rd-difference interpolation,
$3125u_x = [5]^5(1-5\delta^2+14\delta^4)u_x$	corresponds to 4th-difference interpolation,
$15625u_x = [5]^6(1-6\delta^2+19\tfrac{4}{5}\delta^4)u_x$	corresponds to 5th-difference interpolation.

We have now to consider the case of special methods where additional terms are added to an ordinary interpolation formula for the sake of osculatory properties or for any other reason. The above proof does not apply, because the interpolated curve will not, if extended, pass through all the points from which it is calculated. Sprague's formula, for example, uses six points in obtaining each interpolated arc, but the arc, if extended, will not necessarily pass through more than two of those points. It will now be shown that if a term (or any number of terms) in a higher order of differences is added to an ordinary finite-difference interpolation, the process can still be expressed in summation form with an addition to the operand only, even if the added terms are purely arbitrary. This follows directly from the fact that Δ_k^{n+p} is always divisible by $[k]^n$.

To put the proof more explicitly, suppose that to an ordinary central second-difference interpolation between u_x, u_{x+5}, u_{x+10}, etc., we add $a\Delta_5^3 u_{x-5}$, $b\Delta_5^3 u_{x-5}$, $c\Delta_5^3 u_{x-5}$ and $d\Delta_5^3 u_{x-5}$ when interpolating for $u_{x+1}, u_{x+2}, u_{x+3}$ and u_{x+4} respectively. If we write out the coefficients as in our second paragraph adding $(-au_0+3au_5-3au_{10}+au_{15})$ to u_6 and so on, we find that to the series of multipliers we must add

$$+a+b+c+d+0-3a-3b-3c-3d+0+3a+3b+3c+3d+0-a-b-c-d.$$

The expression represented by these detached coefficients can obviously be written symbolically as

$$\delta_5^3(\ldots+0+a\mathrm{E}^{-1\frac{1}{2}}+b\mathrm{E}^{-\frac{1}{2}}+c\mathrm{E}^{\frac{1}{2}}+d\mathrm{E}^{1\frac{1}{2}}+0+\ldots),$$

which equals $\quad [5]^3\delta^3(\ldots 0+a\mathrm{E}^{-1\frac{1}{2}}+b\mathrm{E}^{-\frac{1}{2}}+c\mathrm{E}^{\frac{1}{2}}+d\mathrm{E}^{1\frac{1}{2}}+0\ldots).$

The special interpolation formula will therefore correspond to Woolhouse's summation formula with one-fifth of the above addition. Remembering that there is a numerical divisor of 125, this amounts to adding 25 times the third differences of 0, 0, 0, *a*, *b*, *c*, *d*, 0, 0, 0 to the Woolhouse operand.

Though the above is, for convenience, written for an extra term in δ^3, the argument is general. In practice the additional terms would be symmetrical: otherwise the interpolation and the summation formula would be lop-sided.

As an example, Karup's formula of interpolation, used for several English Life Tables, is equivalent to giving *a*, *b*, *c*, *d* the values $-\cdot 016$, $-\cdot 048$, $+\cdot 048$, $+\cdot 016$; this adds the coefficients $(-\cdot 4+0+3\cdot 6-6\cdot 4+3\cdot 6+0-\cdot 4)$ to the Woolhouse operand $(0+0-3+7-3+0+0)$, giving

$$(-\cdot 4+0+\cdot 6+\cdot 6+\cdot 6+0-\cdot 4) \quad \text{or} \quad (-2+0+3+3+3+0-2)\div 5,$$

which is the operand found by Karup and King.

To summarize, the above enables us to say explicitly:

(1) Subdivision of an interval by k, using ordinary central differences correct to Δ^n, corresponds to a summation formula of graduation with the summations $[k]^{n+1}$. The operand will be the expression necessary to make the graduation correct to $(n+1)$th differences when n is even, and to nth differences when n is odd.

(2) If additional special terms are added to such an interpolation formula (for the purpose of osculation or other reason), the summations will still be $[k]^{n+1}$ but the operand will alter.

(3) In any case, the graduation expression, divided by k, can be used as an interpolation formula.

II. INTERPOLATING FOR SMOOTHNESS

The origin of the foregoing investigation was the thought that interpolation formulae designed for smoothness might be subjected to the test customary for summation formulae.

If we consider the original terms as made up of two parts, a polynomial and a series of errors or departures from the true values, we can regard an interpolation formula from two angles. The order of differences to which it is correct indicates its capacity to reproduce a polynomial, and the smoothness in the flow of the coefficients might measure its capacity to pass smoothly through an irregular series of 'errors'. As the coefficients of an interpolation formula come into the calculation of successive terms in the same order as in the related summation formula, the smoothing coefficient of the latter seems a reasonable index for comparing the relative smoothing properties of two interpolation formulae.

The theory is, of course, open to criticism on the ground that the errors are not independent, and have not the same standard deviation, etc.; but, as with summation formulae, there is ground for expecting the comparison to give results approximating to those of a series of trials.

When Sprague's formula is used for an interval of five, we have six successive points on a polynomial arc of the fifth order; but, since six points supply only one fifth difference, no two successive fifth differences of the final series need actually be equal. At the junction points only two differential coefficients are equalized; so that, while a degree of smoothness is ensured, its measure is not obvious. There are later formulae, some equalizing one differential coefficient only, some more than Sprague's, others based on different principles. The method now used adopts an independent standard for comparing them.

The smoothing coefficients quoted below do not imply that an interpolation method of graduation gives results of the same smoothness as the summation formula; that depends on other factors. The suggestion is only that when one of two graduation formulae has the better smoothing coefficient, the related interpolation formula will probably give the smoother interpolation on an irregular series.

The following applies in particular to subdivision of an interval by five. The relative smoothing index for another interval would differ slightly. The smoothing coefficients have been calculated for Δ^3, according to the usual practice.

Taking fifth-difference formulae first, we write the ordinary Everett formula below for reference:

$$u_x = xu_1 + \frac{x(x^2-1)}{6}\delta^2 u_1 + \frac{x(x^2-1)(x^2-4)}{120}\delta^4 u_1$$
$$+\epsilon u_0 + \frac{\epsilon(\epsilon^2-1)}{6}\delta^2 u_0 + \frac{\epsilon(\epsilon^2-1)(\epsilon^2-4)}{120}\delta^4 u_0,$$

where $\epsilon = (1-x)$. This, as already shown, corresponds to the summation formula

$$15625 u_x = [5]^6 (1 - 6\delta^2 + 19\tfrac{4}{5}\delta^4) u_x.$$

The smoothing coefficient of this is found to be 1/67.

The next step is to calculate, as a standard of comparison, the formula of the same range, correct to fourth differences, with the smoothest possible coefficients for subdividing an interval by five. We have shown that the process is equivalent to ascertaining the best operand for a summation formula with the summations $[5]^5$. The method of doing this can be found elsewhere (Larus, *Transactions of the Actuarial Society of America*, Vol. XIX; Vaughan, *J.I.A.* Vol. LXV); and the formula which has been ascertained is not very convenient for use, so that it is perhaps sufficient to say that the smoothing coefficient is found to be 1/117. This is the best value that can be obtained by any formula of the stated range and accuracy, and it can be used as a standard for assessing the various methods.

Sprague's original osculatory formula (*J.I.A.* Vol. XXII), which replaces the Everett coefficients of δ^4 by $x^3(x-1)(5x-7)\div 24$ and a similar term in ϵ, has been put in summation form by Karup and by King. The smoothing coefficient turns out to be 1/103. The classic formula therefore passes the present test with honours.

As a matter of interest it may be mentioned that after calculating the 'ideal' formula above, the fractional coefficients in the operand were, by methods described elsewhere, replaced by the simplest integral coefficients that would give the same approximate smoothing effect. The result was to rediscover Sprague's formula from an entirely different approach.

Shovelton's formula (*J.I.A.* Vol. XLVII) is based on somewhat different principles and replaces the Everett δ^4 coefficients by $x^2(x-1)(x-5)\div 48$ and a similar ϵ term. It is found that the smoothing coefficient of the related graduation formula is 1/116, which is practically as good as possible.

The above formulae are both correct to Δ^4. The three next mentioned are correct only to Δ^3.

A formula of Henderson, published in *T.A.S.A.* Vol. XXII and quoted in *Mathematics for Actuarial Students* (Freeman), Part II, p. 153, replaces the Everett terms in δ^2 and δ^4 by the attractively simple form $\frac{1}{6}x(x^2-1)(\delta^2-\frac{1}{6}\delta^4)$. However, the smoothing coefficient has the disappointing value of 1/73. It may be said that this formula was a by-product of Henderson's difference-equation method, which is excellent. The latter method requires an artificial second difference B_x to be ascertained from the difference equation

$$B_x + \tfrac{1}{6}\delta^2 B_x = \delta^2 u_x.$$

If B_x is taken at the approximate value of $\delta^2(1-\frac{1}{6}\delta^2)u_x$, the interpolation formula under consideration is reached; but the approximation does not seem to retain the smoothness. The coefficients of the simple formula are very close to those of a straight fifth-difference interpolation, and the formula might therefore be regarded as one for quickly approximating to an ordinary fifth-difference interpolation, but not as replacing an osculatory formula.

An earlier formula of Henderson (*T.A.S.A.* Vol. IX) gave the very satisfactory result of 1/116. In this case the Everett coefficients of δ^4 are replaced by $x^2(1-x)\div 12$ and $\epsilon^2(1-\epsilon)\div 12$. This formula was successfully devised to fill an intermediate position between those of Sprague and Karup. It is correct to one more order of difference than Karup's formula while retaining some of its simplicity.

Jenkins's interesting osculatory formula with final coefficients of form $x^3(1-x)\div 12$ is related to a smoothing coefficient of 1/113. This is quite satisfactory, but the formula is not quite as simple as that last mentioned, and is correct to one order of differences less than Sprague's or Shovelton's.

We now come to the simplest formula, that of Karup, used for recent English Life Tables. This equalizes one differential coefficient only at each end, and is correct only to second differences. The smoothing coefficient of the graduation formula related to a 'straight' third-difference interpolation is 1/52. Karup's formula replaces the Everett coefficients of δ^2 by $\frac{1}{2}x^2(x-1)$ and $\frac{1}{2}\epsilon^2(\epsilon-1)$, δ^4 of course not being used. The relative smoothing coefficient is 1/105, and it is found that nothing better can be done with any similar formula of the same range. This confirms King's remark that the simpler formula sometimes gives a curve nearly as smooth as the longer forms. The experiment has been made of interpolating between terms of the utterly irregular series quoted in *J.I.A.* Vol. XLII, p. 112, and it is found that here the shorter formula gives third differences quite as small as Sprague's method. However, it must be remembered that the smoothing coefficient does not tell the whole story; it supplies a guide to the result from an utterly irregular series, but correctness to a higher order of differences is valuable when the differences show some regularity. It is also the case that the smallness of Δ^3 is not a complete test of smoothness; and if the smoothing coefficient were computed from say Δ^5 a formula such as Sprague's would show to more advantage.

The verdict from this test is that existing formulae are very good of their type and no improvement can be expected from any new method on the same lines, but also that the gain in smoothness is definitely limited.

It is interesting to note that, among ordinary interpolation formulae, those correct to an odd order of differences are the smoother. A second-difference interpolation corresponds to Woolhouse's formula with a smoothing coefficient of 1/15. Third-difference interpolation would bring this to 1/52 (so that Woolhouse would no doubt have attained a material improvement in smoothing by using another order of differences). For fourth differences the figure is 1/23, and for fifth 1/67. The reason for this is probably that for the odd orders the interpolated arcs each pass through two pivot points so that adjoining arcs have a pivot point in common. This suggests that in completing a series by interpolation, if we go beyond first differences, we should go to third rather than second, adopting the convenient Everett formula.

Interpolation combined with graduation

There is an inherent limitation to the smoothing power of the above methods, due to the condition that the pivot values must be unaltered. In the series used by Sprague to illustrate his formula, the first five values of Δ_5^3 are $+548$, -1128,

$+1322$, -56, $+1605$; and it is obvious that in such a case any interpolated curve passing through the assigned points must either have breaks of continuity or waves.

Turning to the coefficients of an interpolation formula, set out in order as described in these notes, it will be seen that to reproduce given quinquennial values it is necessary that the central coefficient be unity and every fifth coefficient from the centre zero. This condition restricts the flow of the coefficients.

When we abandon ordinary finite-difference methods for other formulae it may be that we are assuming the series to be other than a polynomial form. In that case we should logically search out the true form and use it. If we are assuming the series to be approximately a polynomial but subject to errors, why not retain liberty to alter the given values?

In *J.I.A.* Vol. L, p. 126, 'R.T.' (presumably Todhunter) described a method for use 'if the retention of the given values is not material'. With present knowledge better formulae can be deduced, but the method is mentioned as perhaps the first on such lines.

Jenkins (*T.A.S.A.* Vol. XXVIII) gave the following osculatory formula, which is both simple and effective:

$$y_x = xu_1 + \tfrac{1}{6}x(x^2-1)\,\delta^2u_1 - \tfrac{1}{36}x^3\delta^4u_1$$
$$+\epsilon u_0 + \tfrac{1}{6}\epsilon(\epsilon^2-1)\,\delta^2u_0 - \tfrac{1}{36}\epsilon^3\delta^4u_0.$$

This corresponds to the ordinary Everett form down to δ^2. The coefficients of δ^4 do not both vanish when $x=0$ or 1, so the formula replaces an original u_x by $(u_x - \tfrac{1}{36}\delta^4u_x)$.

The smoothing coefficient of the corresponding graduation formula for subdivision in fives is $1/197$. The introduced 'error' has a slight graduating effect on the original values; and at this expense, if it be an expense, the smoothness of the interpolated series is materially improved in comparison with Sprague's and similar formulae.

Now we can, from the angle of the present paper, deduce the formula corresponding to the best possible smoothing coefficient. Such a formula, calculated to reproduce the assigned values, has already been mentioned as corresponding to a smoothing coefficient of $1/117$. Granted freedom from the restriction, it has been found that this can be improved to $1/417$.

The calculated formula was then 'touched up' to obtain coefficients to only three decimal places. The result is

$$y_0 = u_0 - \cdot061\delta^4u_0,$$
$$y_{\frac{1}{5}} = -\cdot045u_{-2} + \cdot127u_{-1} + \cdot614u_0 + \cdot366u_1 - \cdot057u_2 - \cdot005u_3,$$
$$y_{\frac{2}{5}} = -\cdot028u_{-2} + \cdot034u_{-1} + \cdot560u_0 + \cdot476u_1 - \cdot028u_2 - \cdot014u_3,$$
$$y_{\frac{3}{5}} = -\cdot014u_{-2} - \cdot028u_{-1} + \cdot476u_0 + \cdot560u_1 + \cdot034u_2 - \cdot028u_3,$$
$$y_{\frac{4}{5}} = -\cdot005u_{-2} - \cdot057u_{-1} + \cdot366u_0 + \cdot614u_1 + \cdot127u_2 - \cdot045u_3.$$

It will be seen that the 'error' is $-\cdot061\delta^4$. The related smoothing coefficient, after the reduction to three decimal places, is $1/413$.

The above can be expressed in an Everett form, but it is necessary to quote the numerical coefficients of δ^4 because the general expression in terms of x

is complicated. For the common case of subdivision by five, the values for $x = \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}$ are $-\cdot005$, $-\cdot014$, $-\cdot028$, $-\cdot045$ respectively. While this is simple enough, it was desired to find a formula that could be expressed in a simple mathematical form for any interval of interpolation. This was found in an interesting way.

An attempt was made to deduce interpolation formulae by starting from summation forms; and, for certain reasons, a trial was made with

$$[k]^4 \left\{1 - \frac{1}{6}\left(1 - \frac{1}{k^2}\right)\delta_k^2\right\} \div k^4.$$

The corresponding interpolation method turned out to be good but so close in effect to Jenkins's formula that it was not worth suggesting as an alternative. The interesting fact was then noticed that if k were made infinite (i.e. the case of infinite subdivision where we would be interpolating a continuous curve and in effect aiming to minimize the differential coefficient in lieu of the finite difference) the formula would become exactly the same as Jenkins's—a quite unexpected result.

This suggested that the 'ideal' formula also might be simplified by calculating the general form for an interval of k, and putting $k = \infty$. This led to formula C quoted below. The related smoothing coefficient, corresponding to subdivision in fives, is $1/366$, which is sufficiently close to the best result obtainable.

It was then noticed that this formula would be osculatory, although obtained without reference to that property; and that therefore it should be one of the family of formulae covered by the general form given by Reid and Dow (*T.F.A.* Vol. XIV, p. 188). Their general expression for fourth and sixth differences of an osculatory formula in Everett form is:

$$\{bx - \tfrac{1}{36}(1 + 48e + 60f)x^3 + ex^4 + fx^5\}\delta^4 u_1 + \tfrac{1}{6}bx^3\delta^6 u_1,$$

where b, e and f may have any numerical values. The form we had reached is $-\frac{1}{36}x^3(5-3x)\delta^4$, and it will be seen that this can be obtained from the general form by putting $b = 0$, $f = 0$, $e = \frac{1}{12}$.

The work, however, was not wasted as it indicated which simple cases of the general form are likely to be effective. Failing such a guide it would be possible to select a form of the general expression that would have increased error coupled with comparative loss of smoothing.

Knowing the approximate form for maximum smoothing, two simple intermediate formulae were obtained. These, with Jenkins's formula, form a series of four with graded degrees of error and smoothing power. They are listed in the following table, and the suggestion is that from them a formula can be selected which would limit change in the pivot values to the degree considered desirable in any case.

Reid and Dow made a similar suggestion, involving the use of terms in δ^6. This has high smoothing power, but the wide range of the process may be thought an objection as eight pivot values (covering nearly forty ages when subdividing by five) are required for each interpolated term. The method now described provides a good measure of flexibility without going beyond δ^4.

In the following table of formulae the expression in the second column replaces the coefficient of $\delta^4 u_1$ in Everett's formula, the coefficient of $\delta^4 u_0$ being of the same form in ϵ.

Table of formulae

	Coefficient of $\delta^4 u_1$	Error	Smoothing coefficient of graduation formula corresponding to sub-division by five
Formula A	$+x^3(2-3x)\div 72$	$-1/72\delta^4$	1/144
Jenkins's formula	$-x^3 \div 36$	$-1/36\delta^4$	1/197
Formula B	$-x^3(2-x)\div 24$	$-1/24\delta^4$	1/285
Formula C	$-x^3(5-3x)\div 36$	$-1/18\delta^4$	1/366

III. DIFFERENCE-EQUATION INTERPOLATION

In the usual methods a direct expression for each interpolated value is obtained in terms of a few adjoining given values, two for first-difference interpolation, three for second-difference interpolation, and so on. It is possible to adopt a different approach, and ascertain definitely the interpolated terms which will have over the whole series the smallest possible sum of squared differences of any given order. The solution is quite simple, but requires the use of a difference equation which is of a form less generally familiar than the direct algebraic formulae. The arithmetical work need be no greater than for the older methods.

Henderson (*The Record, American Institute of Actuaries*, Vol. XIII) obtained a simple difference-equation solution to the problem of interpolating an osculatory series; and, in *T.A.S.A.* Vol. XLIV, Spoerl has given a solution of the problem mentioned in the preceding paragraph. Spoerl's approach was to connect the problem with the Whittaker-Henderson graduation formula B and, until the following work was completed, it was not noticed that the result would be identical. In the following the approach is direct, and the suggestions for a solution by factorization and for the method of constructing the series are thought to be new.

Given u_0, u_k, u_{2k}, etc., it is required to find u_1, u_2, etc., so that the summed squares of Δ^n shall be a minimum.

The necessary condition is that, if x is not divisible by k,

$$\frac{\partial}{\partial u_x}\Sigma(\Delta^n u)^2 = 0.$$

Now u_x is contained only in $(n+1)$ successive values of Δ^n, so that (ignoring the ends of the series for the moment) we have

$$\frac{\partial}{\partial u_x}\{(\Delta^n u_{x-n})^2 + (\Delta^n u_{x-n+1})^2 + \ldots + (\Delta^n u_x)^2\} = 0.$$

When n is odd, $\Delta^n u_x$ contains $-u_x$ so that, by the ordinary rules of differentiation,

$$\frac{\partial}{\partial u_x}(\Delta^n u_x)^2 = -2\Delta^n u_x;$$

$\Delta^n u_{x-1}$ contains $+nu_x$ so that

$$\frac{\partial}{\partial u_x}(\Delta^n u_{x-1})^2 = 2n\Delta^n u_{x-1};$$

and so on. Hence the first equation becomes

$$2\left(\Delta^n u_{x-n} - n\Delta^n u_{x-n+1} + \frac{n(n-1)}{2}\Delta^n u_{x-n+2} - \ldots - \Delta^n u_x\right) = 0,$$

i.e. $\Delta^{2n} u_{x-n} = 0$. When n is even the signs will be different, but the result is the same.

We have therefore established, for example, that, when subdividing by five to minimize the square of third differences, the interpolated series must be such that sixth differences are zero except at quinquennial points.

At the beginning of the series some of the terms contained in $\Delta^{2n} u_{x-n}$ will be lacking, and the last equation becomes successively

$$\Delta^n u_1 - n\Delta^n u_0 = 0,$$

$$\Delta^n u_2 - n\Delta^n u_1 + \frac{n(n-1)}{2}\Delta^n u_0 = 0,$$

etc.,

from which we can deduce

$$\Delta^n u_1 = n\Delta^n u_0,$$

$$\Delta^n u_2 = \frac{n(n+1)}{2}\Delta^n u_0,$$

and generally, when x is less than k,

$$\Delta^n u_x = \frac{n(n+1)\ldots(n+x-1)}{x!}\Delta^n u_0.$$

At the other end of the series similar conditions apply.

As the result of the factorial form of these coefficients, n, $\frac{n(n+1)}{2}$, etc., it will be seen that, if we treat Δ^n beyond the end of the series as zero and difference n times to reach Δ^{2n}, we will here also reach a series of zeros except for $\Delta^{2n} u_{-n}$; so that we can legitimately regard the values of Δ^n beyond the end as zeros, and on this understanding the statement that Δ^{2n} is zero (except for $\Delta^{2n} u_{-n}$, $\Delta^{2n} u_{k-n}$, etc.) applies to the whole series including the ends. This is analogous to the position in difference-equation graduation as explained in *The Calculus of Observations* (Whittaker and Robinson), p. 306.

We have still to find the values of Δ^{2n} when x is divisible by k. After ascertaining these it will be possible to complete the interpolation by continuous addition, if desired. It will not be necessary to calculate fresh leading differences for each interval, as the proposition covers the whole series.

Before taking the next step, however, the above will be confirmed by an alternative proof to meet the possible objection that the vanishing of the first differential coefficient is not a complete test for a minimum.

Alternative proof. Given that a series of u_x has been interpolated following the above conditions, it is required to prove that any change will increase the sum of $(\Delta^n u_x)^2$.

For simplicity the proof, though general, is written for the case when $n = 3$.

If any positive or negative values, a_1, a_2, etc., be added to the series, a_0, a_k, etc., must, of course, be zero (since u_0, u_k, etc., are the given values), and for all other values of x, $\Delta^6 u_{x-3}$ is zero; so that $a_x \times \Delta^6 u_{x-3}$ is always zero, and we can write, for all values of x,

$$a_x(\Delta^3 u_x - 3\Delta^3 u_{x-1} + 3\Delta^3 u_{x-2} - \Delta^3 u_{x-3}) = 0.$$

If from all the terms of this form we gather the coefficients of $\Delta^3 u_x$ we get

$$(a_x - 3a_{x+1} + 3a_{x+2} - a_{x+3})\Delta^3 u_x, \quad \text{i.e.} \quad -\Delta^3 a_x \Delta^3 u_x.$$

After doing this for all values of x there will be no surplus terms because $\Delta^3 u_x$ vanishes at the ends, so we find that

$$\Sigma\Delta^3 a_x . \Delta^3 u_x = -\Sigma a_x \Delta^6 u_x$$
$$= 0.$$

Hence $$\Sigma\{\Delta^3(u_x + a_x)\}^2 = \Sigma(\Delta^3 u_x)^2 + \Sigma(\Delta^3 a_x)^2,$$

which proves the proposition that the summed squares for the altered series must exceed those of u_x, and also shows that the excess is the exact total of the summed squares for the additions of a_x taken separately.

We turn back to the problem of ascertaining δ^{2n} at the points where x is divisible by k, and for this it will be more convenient to write central differences. We require δ^{2n} for u_0, u_k, u_{2k}, etc., and commence by writing the identity

$$\delta_k^{2n} = [k]^{2n}\delta^{2n}.$$

Since u_x is given at intervals of k, we know the left-hand side; but the known value is expressed as a linear compound of the unknown, so that the relation is a difference equation of the first order which we have to solve. A difference equation is, of course, subject to a multiplicity of numerical solutions as the process is an inverse one, but in this case we need the particular solution for which Δ^n becomes zero at the ends. With fixed end conditions the solution is unique.

As δ^{2n} is zero, except at intervals of k, the only effective coefficients of $[k]^{2n}$ are at the same intervals. If c_0 represent the central coefficient, c_1 the coefficient k places from the centre, and so on, the above identity in this case becomes

$$\delta_k^{2n} u_x = \ldots + c_1\delta^{2n} u_{x-k} + c_0\delta^{2n} u_x + c_1\delta^{2n} u_{x+k} + \ldots.$$

The c's are expanded coefficients of $(1 + E + \ldots + E^{k-1})^{2n}$, and for such an expansion an expression is given in standard works on algebra; but perhaps the easiest method of obtaining the numerical values is to construct the coefficients of $[k]^{2n}$ by summation for any case required, e.g.

$$\begin{array}{ccccccccccccccccc}
 & & & & & & 1 & 1 & 1 & 1 & 1 & & & & & & \\
 & & & & 1 & 2 & 3 & 4 & 5 & 4 & 3 & 2 & 1 & & & & \\
 & & 1 & 3 & 6 & 10 & 15 & 18 & 19 & 18 & 15 & 10 & 6 & 3 & 1 & & \\
1 & 4 & 10 & \underline{20} & 35 & 52 & 68 & 80 & \underline{85} & 80 & 68 & 52 & 35 & \underline{20} & 10 & 4 & 1
\end{array}$$

The figures underlined are the coefficients for subdivision of an interval by five when $2n = 4$, so that in minimizing the squares of second differences, the difference equation becomes

$$\delta_5^4 u_x = 20\delta^4 u_{x-5} + 85\delta^4 u_x + 20\delta^4 u_{x+5}.$$

Continuing for a few more lines we would find the coefficients, when $n = 3$, to be

$$21, \quad 666, \quad 1751, \quad 666, \quad 21,$$

and, when $n = 4$,

$$8, \quad 1652, \quad 18320, \quad 38165, \quad 18320, \quad 1652, \quad 8.$$

In the case where $n = 2$, the equation is simple, especially as the right-hand side expressed in E's can be factorized

$$20\mathrm{E}^{-5} + 85\mathrm{E}^0 + 20\mathrm{E}^5 = 5(\mathrm{E}^{-5}+4)(4+\mathrm{E}^5),$$

leading to the simple process of solution described later in detail.

When $n = 3$, the factors are less simple, but still quite practical with the help of a multiplying machine. They are

$$(\cdot 014440\mathrm{E}^{-10} + \cdot 451425\mathrm{E}^{-5} + 1)(1 + \cdot 451425\mathrm{E}^5 + \cdot 014440\mathrm{E}^{10}) \div \cdot 0006876.$$

As one objection to the method of minimizing differences has been that at the ends of the series the order of differences concerned tends to vanish, making for an 'artificial situation at the ends', it may now be pointed out that these end conditions may, if desired, be ignored. By a modification of the 'alternative proof' above, it can be shown that, if we adopt end differences to suit our own idea, the method will yield the minimum values obtainable with those end terms. It is possible, if desired, to apply similar initial conditions to those put forward for Henderson's osculatory method. (In fact, as Spoerl shows, the latter is a limiting case of the present method.)

Under Henderson's and Spoerl's methods for second differences, a function B_x is ascertained which replaces δ_n^2 in Everett's formula; but, as already pointed out, after ascertaining all the differences of one order we need only one initial set of differences to enable the series to be completed by continuous addition. The latter method is therefore now suggested. It will be found that owing to the nature of the method the accumulation of error is not troublesome.

Taking the case where $n = 2$ and $k = 5$, and using v_x to denote the given series of quinquennial values with intermediate zeros, we have for every value of x

$$(20\mathrm{E}^{-5} + 85 + 20\mathrm{E}^5)\,\delta^4 u_x = \delta_5^4 v_x = \delta^2 [5]^2 \delta_5^2 v_x.$$

Taking the finite integral of each side twice,

$$(20\mathrm{E}^{-5} + 85 + 20\mathrm{E}^5)\,\delta^2 u_x = [5]^2\,\delta_5^2 v_x.$$

We need not concern ourselves with constants of integration, as the solutions of the last difference equation will include those of the previous one.

Since $\delta_5^2 v_x$ is zero except at quinquennial points, $[5]^2$ is at such points equal to a multiplication by 5 and of course $v_x = u_x$; so at the quinquennial points we have simply

$$(4\mathrm{E}^{-5} + 17 + 4\mathrm{E}^5)\,\delta^2 u_x = \delta_5^2 u_x. \qquad (1)$$

By solving this we obtain $\delta^2 u_x$ at intervals of 5; and the intermediate values we can fill in by first-difference interpolation, because, as previously seen, the operation $[5]^2 \div 5$ is equivalent to this.

Having found the whole series of $\delta^2 u_x$, we need only one value of Δ to enable the whole of the u's to be constructed. Since each set of five successive values of δ^2 is in arithmetical progression, we can write accurately (when x is divisible by 5) the ordinary advancing difference formula

$$u_{x+5} = u_x + 5\Delta u_x + 10\Delta^2 u_x + 10\Delta^3 u_x,$$

and, since $5\Delta^3 u_x = \delta^2 u_{x+5} - \delta^2 u_x$, and $\Delta^3 u_{x-1} = \Delta^3 u_x$, this becomes

$$u_{x+5} = u_x + 5\Delta u_x + 6\delta^2 u_x + 4\delta^2 u_{x+5},$$

so that $$\Delta u_x = \tfrac{1}{5}(u_{x+5} - u_x - 6\delta^2 u_x - 4\delta^2 u_{x+5}), \qquad (2)$$

and, when $\delta^2 u_0$ is zero, $$\Delta u_0 = \tfrac{1}{5}(\Delta_5 u_0 - 4\delta^2 u_5). \qquad (3)$$

Equations (1) and (3) are all we need to obtain an interpolation with the smallest possible squares of second difference. Equation (2) can be used, if desired, to check the Δ at any point or to ascertain Δu_0 if we prefer in any case that δ^2 should not tend to zero at the ends.

We can minimize the squares of third differences by a similar process, the calculation required being about the same as for a Sprague interpolation. The work for the second-difference case is about the same as for Karup's simple formula, or for an ordinary third-difference interpolation.

It may be remarked, in passing, that we could also work in summations because it can be deduced that

$$20u_{x-5} + 85u_x + 20u_{x+5} = [5]^4 v_x,$$

leading to another simple process. We could solve the difference equation for the given quinquennial u's, make a first-difference interpolation (to replace $[5]^2$), and sum twice in fives to get the interpolated values without using differences at all. The process above suggested is, however, perhaps slightly shorter, though not so compact in expression.

An actual example of the suggested procedure may clarify the explanation. In Table B the series shown is that of $q_{[x]+3}$, used by Sprague (*J.I.A.* Vol. XXII, p. 281) to illustrate his osculatory formula, the figures being multiplied by 10^6 to save writing the decimal points. The next column contains the second central differences.

To solve the difference equation (1), we calculate the next column, a_x, so that $a_{x+5} = \frac{1}{4}(\delta^2 u_{x+5} - a_x)$ starting with $a_{20} = 0$. For example,

$$a_{30} = (1312 - 191) \div 4 = 280{\cdot}25.$$

To obtain the next column, b_x, we start at the foot with a zero and apply the same process backwards to a_x; so that b_{60}, for example,

$$= (745{\cdot}75 - 579{\cdot}52) \div 4 = 41{\cdot}56.$$

Since $\delta^2 u_x = 4a_x + a_{x-5}$, and $a_x = 4b_x + b_{x+5}$,

we now have $\delta_5^2 u_x = 4b_{x-1} + 17b_x + 4b_{x+1}$.

so that b_x is a solution of difference equation (1).

This process is equivalent to solving the equation by factorization, since we have in effect replaced $(4E^{-5} + 17 + 4E^5)$ by its factors $(E^{-5} + 4)(4 + E^5)$, and effected the solution in two steps.

However, b_x is not quite the solution that we want, and we can obtain a solution starting with zero by a simple adjustment. To obtain other solutions we may add to b_x any solution of the difference equation $4E^{-5} + 17 + 4E^5 = 0$, the solutions being of the form $(p4^x + q4^{-x})$, where p and q are any numbers. We can take p as zero if we wish to adjust the beginning of the series, or q as zero if we wish to adjust the end. Hence in the next column we write $+7{\cdot}13$, in the next line one-fourth of this with the sign changed, and continue until in a few lines the adjustment becomes negligible.

Table B

x	u_x	δ_5^2	a_x	b_x	Adjustment	δ^2	x
20	10709		0	−7·13	7·13	0	20
25	9585	764	191·00	28·52	−1·78	26·74	25
30	9225	1312	280·25	76·90	·44	77·34	30
35	10177	184	−24·06	−27·37	−·11	−27·48	35
40	11313	1506	382·52	85·41	·03	85·44	40
45	13955	1450	266·87	40·88	−·01	40·87	45
50	18047	3055	697·03	103·36		103·36	50
55	25194	5401	1175·99	283·61		283·61	55
60	37742	4159	745·75	41·56		41·56	60
65	54449	12324	2894·56	579·52		579·52	65
70	83480	12118	2305·86	576·46		576·46	70
75	124629			0		0	75

x	Δ^2	Δ	Interpolated series	x
	0			
20	5·35	−246·20	10709	20
	10·70	−240·85	10462·80	
	16·05	−230·15	10221·95	
	21·40	−214·10	9991·80	
	26·75	−192·70	9777·70	
25	36·87	−165·95	9585	25
	46·99	−129·08	9419·05	
	57·11	−82·09	9289·97	
	67·23	−24·98	9207·88	
	77·35	42·25	9182·90	
30	56·38	119·60	9225	30
	35·41	175·98	9344·60	
	14·44	211·39	9520·58	
	−6·53	225·83	9731·97	
	−27·50	219·30	9957·80	
35	−4·91	191·80	10177	35
	17·68	186·89	10368·80	
	40·27	204·57	10555·69	
	62·86	244·84	10760·26	
	85·45	307·70	11005·10	
40			11313	40

Adding the adjustment to b_x, we reach in the final column the values of $\delta^2 u_x$ for our interpolation. It will be found that any three consecutive lines of this column multiplied in order by 4, 17 and 4, will add to the corresponding δ_5^2 in the third column.

The explanation may seem rather laboured, but it is desired to make the process clear to a reader who may not be versed in difference equations. The actual work consisted only of a few subtractions and divisions by four, and took less time to do than to describe. This work replaces that of calculating the leading differences for a Karup interpolation by King's method.

$$\text{Now from equation (3), } \Delta u_{20} = \tfrac{1}{5}(\Delta_5 u_{20} - 4\delta^2 u_{25})$$
$$= \tfrac{1}{5}(-1124 - 4\times 26{\cdot}74)$$
$$= -246{\cdot}2.$$

The lower part of Table B shows the completion of the interpolation. The values of δ^2 which we have obtained are written at quinquennial points and underlined. They have been taken to the nearest five or zero in the last decimal place to simplify the arithmetic, and appear one line higher as we are now using advancing in lieu of central differences. The intervening values of Δ^2 are filled in by first-difference interpolation, and, starting with $\Delta u_{20} = -246{\cdot}20$ as above, we can complete the series of u's. The work has been carried only to u_{40} to compare with Sprague's example.

It will be noticed that the original values are in each case reproduced correctly to an integer. If sufficient decimal places had been used the reproduction would have been exact. We have carried the work for twenty terms without needing an adjustment for accumulated error, although δ^2 is correct only to ·05. It was not thought worth while to use more decimal places as Sprague took the q's to the sixth decimal, and further extension seemed superfluous. Any accumulation of error that might appear could be met by adjusting a Δ by say ·05, or a Δ half-way down the series could be checked by equation (2) if it were thought worth while.

It will be seen that the successive arcs of the interpolated series overlap, u_{24} to u_{31}, for example, being on one polynomial arc, u_{29} to u_{36} on another, and so on, so that adjoining arcs have three points in common. This incidentally explains why in the limiting case, when the interval of interpolation is infinitesimal, this process becomes identical with Henderson's osculatory method.

It will be noted that the present method covers the whole series while Sprague's interpolation commences at age 30.

An interpolation of the same series has also been made for third differences, using the factorization method quoted earlier. With a multiplying machine, the process is not unduly onerous.

Comparative differences for the small portion of the series quoted by Sprague (ages 30 to 40) are as follows:

Sprague			Difference equation for second differences			Difference equation for third differences		
Δ^2	Δ^3	Δ^4	Δ^2	Δ^3	Δ^4	Δ^2	Δ^3	Δ^4
+74	−26	−17	+56	−21	+ 1	+54	−20	− 3
+48	−43	+20	+35	−20	− 2	+34	−23	+ 4
+ 5	−23	+27	+15	−22	+ 2	+11	−19	+12
−18	+ 4	− 5	− 7	−20	+42	− 8	− 7	+15
−14	− 1	+24	−27	+22	0	−15	+ 8	+11
−15	+23	+19	− 5	+22	+ 2	− 7	+19	+ 4
+ 8	+42	−15	+17	+24	− 2	+12	+23	0
+50	+27		+41	+22		+35	+23	
+77			+63			+58		