SOME PRACTICAL INTERPOLATION FORMULAS

By John L. Roberts

Sometimes we wish to find by means of interpolation an approximation to a particular value of w_x in the interval between the known values, w_0 and w_1. But it also might be desirable in the interval from w_0 to w_1 to interpolate several approximations to w_x at equidistant values of x. It is very important to know that a formula which might be very satisfactory to interpolate a particular value in an interval might seriously fail to be the most satisfactory formula when it is desired to interpolate several values in the same interval. The range of this paper is so limited that we only wish to find by means of interpolation several approximations to the true value of w_x in the interval from w_0 to w_1 at equidistant values of x.

One way to perform an interpolation of this sort is to use osculatory interpolation.[1] The real function of osculatory interpolation is to secure smoothness at the known points, which are sometimes called pivotal points. By roughness is meant that one or more of the successive derivatives are discontinuous at the pivotal points. Experience proves that the osculatory formulas usually secure smoothness either at the expense of labor or by a loss of accuracies over the entire range from w_0 to w_1. Frequently the function of interpolation formulas is to save labor. In many cases it appears reasonable to save labor by a loss of both smoothness and accuracy. Formulas are herein selected, without direct regard for smoothness, so as to secure the best possible compromise between a maximum of accuracy and a minimum of labor. It appears that this results in many cases in a loss of smoothness that is no more objectionable than the loss in accuracy.

The actuarial profession, while trying to perfect their methods of constructing mortality tables, have made contributions of a high order of scholarship to the theory of osculatory interpolation. But since the statistician, the astronomer, the physicist, and other scientists also have occasions to make interpolations, it seems to be very important to discuss the problem of finding the most practical methods of interpolation, not only from the special viewpoint of the actuary, but also from the general viewpoint of mathematics.

Δw_x is called the first difference of w_x, and may be defined by $\Delta w_x = w_{x+1} - w_x$.

[1] Since this paper presupposes certain knowledge on the part of the reader, it may be worth while to indicate some sources of this knowledge. The elementary parts of this knowledge can be found in any good book on finite differences. "Population Statistics and Their Compilation" by Hugh H. Wolfenden, published by the Actuarial Society of America, contains an excellent summary of osculatory interpolation. This summary indicates some valuable sources of information.

Second, third, and higher differences are merely successive differences of the first. When use is made of central difference interpolation formulas, it is convenient to adopt Woolhouse's notation, which is defined by means of the following equations: $\Delta w_{-2} = a_{-2}$, $\Delta w_{-1} = a_{-1}$, $\Delta w_0 = a_1$, $\Delta w_1 = a_2$, $\Delta^2 w_{-2} = b_{-1}$, $\Delta^2 w_{-1} = b_0$, $\Delta^2 w_0 = b_1$, $\Delta^3 w_{-2} = c_{-1}$, $\Delta^3 w_{-1} = c_1$, $\Delta^4 w_{-2} = d_0$, $\Delta^5 w_{-2} = e_1$, $\Delta^6 w_{-3} = f_0$, etc.

An important family of curves can be represented by

$$u_x = u_0 + xa_1 + \frac{1}{2}x(x-1)B + \frac{1}{6}x(x-1)\left(x - \frac{1}{2}\right)C. \qquad (1)$$

Assume $u_0 = w_0$ and $\Delta u_0 = \Delta w_0$. Then a study of (1) shows that a_1, which has already been defined, must be a factor in the second term in order that (1) may be satisfied when $x = 1$. (1) is a third degree equation. However, if $C = 0$, (1) becomes a second degree equation; if both $B = 0$ and $C = 0$, (1) becomes a first degree equation. In other words, by giving B and C proper values, (1) can be made to become many different interpolation formulas.

For many purposes interpolation by a first degree formula is not sufficiently accurate. We, therefore, might wish to interpolate by either a second or a third degree formula. Since it is possible to draw an unlimited number of second degree curves or third degree curves between the points P_0 and P_1, the problem of selecting the best second degree interpolation curve and the best third degree curve is of great practical importance.

I

Suppose that w_{-2}, w_{-1}, w_0, w_1, w_2, and w_3 can be found in a table of values of the function w_x, and that we wish to find by means of interpolation several approximate values of w_x in the interval from w_0 to w_1. These six given values of w_x can be used to determine six pivotal points, which determine a fifth degree curve. Suppose this curve represents the function v_x. Then w_x and v_x would have exactly the same values at the six pivotal points, but would have values which are only approximately the same at other points. Using the first six terms of the Gauss central difference interpolation formula, we have

$$v_x = v_0 + xa_1 + \frac{1}{2!}x(x-1)b_0 + \frac{1}{3!}(x+1)x(x-1)c_1$$
$$+ \frac{1}{4!}(x+1)x(x-1)(x-2)d_0$$
$$+ \frac{1}{5!}(x+2)(x+1)x(x-1)(x-2)e_1.$$

It is proper to use in this formula the differences a_1, b_0, etc., which have already been defined as differences of w_x because these differences are exactly equal to the corresponding differences of v_x. Suppose P_0, $P_{\frac{1}{3}}$, $P_{\frac{2}{3}}$, and P_1 are four points

which are determined by v_x. Then B and C can be determined so that (1) will represent the curve which can go through these four points.
Then

$$u_{\frac{1}{3}} = u_0 + \frac{1}{3}a_1 - \frac{1}{9}\left(B - \frac{1}{18}C\right)$$

and

$$v_{\frac{1}{3}} = u_0 + \frac{1}{3}a_1 - \frac{1}{9}\left(b_0 + \frac{4}{9}c_1 - \frac{5}{27}d_0 - \frac{7}{81}e_1\right).$$

Also

$$u_{\frac{2}{3}} = u_0 + \frac{2}{3}a_1 - \frac{1}{9}\left(B + \frac{1}{18}C\right)$$

and

$$v_{\frac{2}{3}} = u_0 + \frac{2}{3}a_1 - \frac{1}{9}\left(b_0 + \frac{5}{9}c_1 - \frac{5}{27}d_0 - \frac{8}{81}e_1\right).$$

Since $u_{\frac{1}{3}} = v_{\frac{1}{3}}$ and $u_{\frac{2}{3}} = v_{\frac{2}{3}}$, we have two equations, which can be solved for B and C.

$$B = b - \frac{5}{27}d \text{ and } C = c_1 - \frac{1}{9}e_1 \tag{2}$$

where b and d are defined by

$$b = \frac{1}{2}(b_0 + b_1) \text{ and } d = \frac{1}{2}(d_0 + d_1).$$

A study of (1) shows that $u_{\frac{1}{2}}$ does not depend upon C because the term containing C becomes zero when $x = \frac{1}{2}$, and also shows that u_x over the entire range from u_0 to u_1 is more sensitive to errors in B than errors in C. The B in (2) usually contains some error because the six terms of the Gauss formula which were used in determining B usually produce results which are only approximate. Consequently a comparatively large error in C would not produce an important error.
Assume

$$B = b - \frac{5}{27}d \text{ and } C = c_1 - \frac{5}{27}e_1. \tag{3}$$

B is the same in both (2) and (3), but C is not the same. The accuracy of (2) and the accuracy of (3) do not differ by an important amount. On the other hand, if any attempt to apply (2) is compared with the working illustrations of (3) in this article, it will be found that (2) to an important extent is more laborious than (3). Therefore (3) is a better compromise between a maximum of accuracy and a minimum of labor than (2). For this reason (2) ought not to be regarded as a practical formula. On the other hand (2) because of its great accuracy serves as an ideal with which other formulas can be compared. In other words (2) is of theoretical importance.

In like manner another interpolation formula can be found if we use the first four terms of the Gauss formula to determine $P_{\frac{1}{2}}$. Then

$$u_{\frac{1}{2}} = u_0 + \frac{1}{2} a_1 - \frac{1}{8} B$$

and

$$v_{\frac{1}{2}} = u_0 + \frac{1}{2} a_1 - \frac{1}{8} \left(b_0 + \frac{1}{2} c_1 \right).$$

Since $u_{\frac{1}{2}} = v_{\frac{1}{2}}$, we can solve for B, and C is left arbitrary. If $C = 0$, we again get an excellent compromise between a maximum of accuracy and a minimum of labor. The following second degree formula results.

$$B = b \text{ and } C = 0. \tag{4}$$

In order that the value of (3) and (4) may be appreciated, they are herein compared with some other formulas which have been of historical importance.

If the point $P_{\frac{1}{2}}$ can first be accurately determined, a second degree curve through the points P_0, $P_{\frac{1}{2}}$, and P_1 would probably give more accurate results than such a curve through the points P_0, P_1, and P_2 because the first three points are in a smaller neighborhood; the second curve can be represented by the first three terms of the Gregory-Newton interpolation formula. The points P_{-1}, P_0, P_1, and P_2 determine a third degree curve, which can be represented by the first four terms of the Gauss central difference formula. It is probable that these terms would determine $P_{\frac{1}{2}}$ much more accurately than the first three terms of the Gregory-Newton formula because the latter is not a central difference formula with respect to $P_{\frac{1}{2}}$ and because four terms usually give more accurate results than only three terms. Consequently there is a strong probability that (4) is more accurate than the first three terms of the Gregory-Newton formula. In like manner (4) is more accurate than the first three terms of the Gauss formula. It is interesting to observe that (4) is the first three terms of the Newton-Bessel formula.

$$\text{If } B = b \text{ and } C = 3c_1,$$

then (1) is equivalent to Karup's osculatory interpolation formula in terms of differences taken centrally. B is the same in both (4) and Karup's formula. No interpolation formula can be very accurate unless C is about equal to c_1. Since, then, the error in C in Karup's formula is about twice as great as the error in C in (4), his formula is distinctly less accurate than (4). Since (4) is a second degree curve and Karup's formula is a third degree curve, his formula is very much more laborious. (4) is extremely accurate for a formula having its labor saving properties; for many purposes its roughness and inaccuracy appear to be in about the right proportion. On the other hand Karup's formula is extremely inaccurate for a formula so laborious; its only good point is its smoothness.

Changing somewhat the meanings of u and w, (3) may be written

$$u_{x+n} = u_n + x\Delta u_n + \frac{1}{2}x(x-1)\left[\frac{1}{2}(\Delta^2 w_n + \Delta^2 w_{n-1}) - \frac{5}{54}(\Delta^4 w_{n-1} + \Delta^4 w_{n-2})\right] + \frac{1}{6}x(x-1)\left(x-\frac{1}{2}\right)\left(\Delta^3 w_{n-1} - \frac{5}{27}\Delta^5 w_{n-2}\right).$$

If

$$\frac{du}{dx} = u'_{x+n},$$

then

$$u'_{0+0} - u'_{1-1} = \frac{1}{54}d_0 - \frac{5}{162}f_0,$$

which is the amount of discontinuity in $\frac{du}{dx}$ at P_0. (3) has greater smoothness than (4); in other words (3) is more like an osculatory formula. On the other hand

$$B = b - \frac{1}{6}d \text{ and } C = c_1 - \frac{1}{6}e_1, \tag{5}$$

which is equivalent to an important osculatory interpolation formula by Mr. Robert Henderson, compares much better with (3) from the viewpoint of labor saving and accuracy than Karup's formula does with (4).

II

An excellent formula can be easily spoiled if the method of applying it is not practical. Mr. Henderson, in the Transactions of the Actuarial Society of America, Vol. IX, applies (5) in such a way that the numerical work is very convenient. Some writers seem to have been very careless about this matter. A method intended to interpolate several values between w_0 and w_1 should provide that the end value w_1 shall be exactly reproduced if no error is made in the computation. In other words a good method should provide a check upon the work. At the same time, in order to avoid unnecessary labor, the work should not retain unnecessary decimal places or figures. In other words fictitious accuracy should be avoided. The following working illustrations are intended to show good methods of application of formulas and to show how much labor is necessary in order to apply them; also the size of the errors can be used to illustrate the theory.

When (4) is applied at either end of the table, where terms are not available for the calculation of the differences required, it should be assumed that the fourth differences that cannot be computed vanish and the required differences should be filled in consistently with that assumption. Δw_x represents the first differences. But it is convenient to have S represent the first differences in such a manner that they are arranged centrally in the working illustration. S^2 in like manner represents the second differences. The 2 in S^2 means S^2 is a second difference, and does not have the familiar meaning used in algebra. In the case of (4), $\Delta u_x = a_1 + xB$, $\Delta^2 u_x = B$, and the higher differences all equal zero. Since we wish in the working illustration of (4) to interpolate four values between w_0 and w_1, δ and δ^2 are defined by $\delta u_x = u_{x+.2} - u_x$ and $\delta^2 u_x = \delta u_{x+.2} - \delta u_x$. It is proved in any good book on finite differences that there are possibilities that Δ and δ, which are symbols of operation, can be separated from the functions upon which they operate, and they can be treated as if they were algebraic numbers. Consequently $1 + \delta = (1 + \Delta)^{\frac{1}{5}}$. In other words by means of the binomial law $\delta u_x = (.2\Delta - .08\Delta^2)u_x$, where all the terms within the parenthesis are to be considered as operating upon u_x. Also $\delta^2 u_x = .04\Delta^2 u_x \cdot$ s, s_x, and s^2 are defined by $s = s_x = \delta u_x$, and $s^2 = s^2_{.4} = \delta^2 u_x$. Therefore the middle $s = \delta u_{.4} = .2a_1$, and $s^2 = .04B = .02(b_0 + b_1)$. We are now in position to apply (4) to the case when $w_x = (1.04)^n$. It might prevent confusion if it is stated that x and n are related to each other in such a way that we always interpolate between w_0 and w_1.

n	$(1.04)^n$	s	S	S^2	s^2
80	23.050	.9218		.845	
81	23.9718	.9603			
82	24.9321	.9988	4.994		.0385
83	25.9309	1.0373			
84	26.9682	1.0758			
85	28.044	1.1190		1.081	
86	29.1630	1.1670			
87	30.3300	1.2150	6.075		.0480
88	31.5450	1.2630			
89	32.8080	1.3110			
90	34.119	1.3636		1.317	
91	35.4826	1.4210			
92	36.9036	1.4784	7.392		.0574
93	38.3820	1.5358			
94	39.9178	1.5932			
95	41.511			1.553	

Some of the explanation of the application of (4) applies to (3) and does not need to be repeated. The method herein used of applying (3) is either the same as or a development of the Henderson method of applying (5). If it is desired to apply (3) at either end of the table, where terms are not available for the calculation of the differences required, it can be assumed that the sixth differences that can not be computed vanish and the required differences can be filled in consistently with that assumption. A study of the theory underlying this assumption shows that it does not result in a true central difference formula and that it consequently results usually in some loss of accuracy. In the case of (3) before the finding of the differences of (1), it is convenient to write it as follows:

$$u_x = u_0 + xa_1 + \frac{1}{2}x(x-1)\left(B + \frac{1}{2}C\right) + \frac{1}{6}x(x-1)(x-2)C.$$

Then

$$\Delta u_x = a_1 + x\left(B + \frac{1}{2}C\right) + \frac{1}{2}x(x-1)C,$$

$$\Delta^2 u_x = \left(B + \frac{1}{2}C\right) + xC, \text{ and } \Delta^3 u_x = C.$$

Suppose we wish to interpolate four values between w_0 and w_1. δ and δ^2 have already been defined. $\delta^3 u_x = \delta^2 u_{x+.2} - \delta^2 u_x$. Then $1 + \delta = (1 + \Delta)^{\frac{1}{5}}$, or $\delta u_x = (.2\Delta - .08\Delta^2 + .048\Delta^3)u_x$. Also $\delta^2 u_x = (.04\Delta^2 - .032\Delta^3)u_x$ and $\delta^3 u_x = .008\Delta^3$. s^2, s^2_x, and s^3 are defined by $s^2 = s^2_x = \delta^2 u_{x-.2}$, and $s^3 = s^3_{.4} = \delta^3 u_x$. The first

$$s^2 = \delta^2 u_{-.2} = .04\left(B - \frac{1}{2}C\right) = .04\left(b_0 - \frac{5}{27}d_0\right).$$

The last

$$s^2 = \delta^2 u_{.8} = .04\left(B + \frac{1}{2}C\right) = .04\left(b_1 - \frac{5}{27}d_1\right).$$

.1852 might be a useful approximation to $\frac{5}{27}$. The remaining s^2, s should be filled in so that they are in arithmetical progression with irregularities at the ends. If the irregularities can be distributed equally at both ends, the irregularities cause an error in C, but none in B. Errors in B are more important than those in C. The middle $s = \delta u_{.4} = .2a_1 - s^3$. In the following working illustration, $w_x = \sin n$.

n	sin n	S	S^2	S^3	S^4
−60	−.86603				
		.36603			
−30	−.50000		.13397		
		.50000		−.13397	
0	.00000		.00000		.00000
		.50000		−.13397	
30	.50000		−.13397		.03588
		.36603		−.09809	
60	.86603		−.23206		
		.13397			
90	1.00000				

n	sin n	s	s^2	s^3
0	.00000	.104498	.000000	
6	.104498	.103374	−.001124	
12	.207872	.101125	2249	−.001125
18	.308997	.097751	3374	
24	.406748	93252	4499	
30	.50000		−.005624	

Suppose we wish to interpolate nine values between w_0 and w_1 by the use of (3). Then $\delta u_x = u_{x+.1} - u_x$, $\delta^2 u_x = \delta u_{x+.1} - \delta u_x$, and $\delta^3 u_x = \delta^2 u_{x+.1} - \delta^2 u_x$. Consequently $1 + \delta = (1 + \Delta)^{\frac{1}{10}}$, or $\delta u_x = (.1\Delta - .045\Delta^2 + .0285\Delta^3)u_x$. Then $\delta^2 u_x = (.01\Delta^2 - .009\Delta^3)u_x$ and $\delta^3 u_x = .001\Delta^3$. $s^2 = s^2_x = \delta^2 u_{x-.1}$ and $s^3 = s^3_{.4} = \delta^3 u_x$. The first

$$s^2 = \delta^2 u_{-.1} = .01\left(B - \frac{1}{2}C\right) = .01\left(b_0 - \frac{5}{27}d_0\right).$$

The last

$$s^2 = \delta^2 u_{.9} = .01\left(B + \frac{1}{2}C\right) = .01\left(b_1 - \frac{5}{27}d_1\right).$$

$\delta u_{.4} = (.1a_1 - 4s^3) - \frac{1}{2}\delta^2 u_{.4}$ and $\delta u_{.5} = (.1a_1 - 4s^3) + \frac{1}{2}\delta^2 u_{.4}$.

n	$\sin n$	s	s^2	s^3
0	.00000	52318	.000000	
3	.052318	52179	−.000139	
6	.104497	51899	280	
9	.156396	51478	421	
12	.207874	.050916	562	−.000141
15	.258790	.050212	703	
18	.309002	49368	844	
21	.358370	48383	985	
24	.406753	47257	1126	
27	.454010	45990	1267	
30	.50000		−.001406	

Suppose we wish to interpolate five values between w_0 and w_1. The first $s^2 = \frac{1}{36}\left(b_0 - \frac{5}{27}d_0\right)$ and the last $s^2 = \frac{1}{36}\left(b_1 - \frac{5}{27}d_1\right)$.

$$\delta u_{\frac{1}{2}} = \frac{1}{6}(a_1 - 8\delta u_x) - \frac{1}{2}\delta^2 u_{\frac{1}{2}}$$

and

$$\delta u_{\frac{1}{2}} = \frac{1}{6}(a_1 - 8\delta^3 u_x) + \frac{1}{2}\delta^2 u_{\frac{1}{2}}.$$

In the following working illustration the given values of sin n are written correct to five decimal places; in other words after each decimal point there are five symbols or digits representing numbers; also each of these symbols is written in the scale of ten. It can be observed that some values of u_x, s, s^2, and s^3 in the working illustration have six symbols to the right of the decimal point, and that some values have seven symbols to the right of the decimal point. In all cases the sixth symbol to the right of the decimal point is written in the scale of ten, and the seventh symbol is written in the scale of six. This procedure provides a check by exactly reproducing w_1. Also this procedure does not cause much fictitious accuracy, and can be quickly used after a little practice.

n	$\sin n$	s	s^2	s^3
0	.00000	87130	.000000	
5	.0871305	86479	−.000651	
10	.1736104	.0851775	1302	
15	.2587883	.0832245	1953	−.000651
20	.3420132	80620	2604	
25	.4226341	77365	3255	
30	.50000		−.003906	

In general if we wish to interpolate $i - 1$ values between w_0 and w_1 when i is neither five nor ten, w_1 can be exactly reproduced if some of the symbols are written in the scale of i. If $i = 12$, it is evident that we need two extra symbols, say t and e, to stand for ten and eleven respectively. If we wish to interpolate $i - 1$ values between w_0 and w_1 by the use of (4), in the computation each of u_x, s and s^2 except the given values should contain one more symbol than each given value contains, and the extra symbol should be written in the scale of i.

ERRATA

THE ANNALS OF MATHEMATICAL STATISTICS

Volume VI, No. 3, September, 1935

The eleventh line on page 137 should read

$$u'_{0+0} - u'_{1-1} = \frac{1}{54} d_0 + \frac{5}{162} f_0.$$

In the sixth line from bottom of page 139, read s^2's, i.e. the plural of s^2.

About the middle of page 141 the formula $\delta u_{\frac{1}{3}}$ should read

$$\delta u_{\frac{1}{3}} = \frac{1}{6} (a_1 - 8\, \delta^3 u_x) - \frac{1}{2} \delta^2 u_{\frac{1}{3}}.$$