AN IMPROVED METHOD OF CONSTRUCTING SHORTENED LIFE-TABLES.

By T. E. HAYWARD, M.B. (Lond.), F.R.C.S. (Eng.),
Medical Officer of Health of Haydock, Lancashire.

THE following is an addendum to a paper which appeared in the last number of this Journal.

The primary object of the paper was to give working formulæ for the construction of a shortened Life-Table, set forth with sufficient clearness to enable them to be practically used by anyone who understands ordinary arithmetic and the use of logarithms, without any necessary comprehension of the principles of the differential or of the integral calculus, by means of which the formulæ have been deduced.

However, there are probably but few who would undertake the task of constructing a Life-Table without interesting themselves in some degree concerning the mathematical principles involved, and, to those at least who may understand something of the method of "interpolating" terms in any given series by means of "Lagrange's formula," it is easily possible to demonstrate how the special formulæ which have been recommended may be very nearly arrived at by simple processes of interpolation only, and in so doing, those who take the trouble to follow the attempted explanations may be brought nearer to a knowledge of what is meant by the differential and integral calculi.

An attempt was made, by means of Fig. 1 and the references to it in the text of the paper (see pp. 89—90), to show how by interpolation and summation of intermediate ordinates at successively decreasing equal intervals the formula of "integration" to be used may be led up to, or in other words, how the "limit of summation" in the given series may be defined.

The following is an explanation of the method by which the first of the successive approximations to the true formula was obtained.

The first step in the problem is, having given the series $u_{-7\frac{1}{2}}$, $u_{-2\frac{1}{2}}$, $u_{2\frac{1}{2}}$, and $u_{7\frac{1}{2}}$, to interpolate $u_{-1\frac{1}{2}}$, $u_{-\frac{1}{2}}$, $u_{\frac{1}{2}}$, and $u_{1\frac{1}{2}}$.

According to "Lagrange's formula," having given any series of linear quantities (that is, such as are capable of being represented as ordinates of a curve), which may be denoted by the symbols u_a, u_b, u_c, u_d, etc., any other term u_x in the series may be interpolated thus (the special case of only four given quantities is now being considered, but it will be obvious how the formula may be applied to five or more):

$$u_x = \left[\frac{(x-b)(x-c)(x-d)}{(a-b)(a-c)(a-d)}u_a + \frac{(x-a)(x-c)(x-d)}{(b-a)(b-c)(b-d)}u_b + \frac{(x-a)(x-b)(x-d)}{(c-a)(c-b)(c-d)}u_c + \frac{(x-a)(x-b)(x-c)}{(d-a)(d-b)(d-c)}u_d\right].$$

By substituting $-7\frac{1}{2}$, $-2\frac{1}{2}$, $2\frac{1}{2}$ and $7\frac{1}{2}$ respectively for a, b, c, and d, and then in succession making $x = -1\frac{1}{2}$, $-\frac{1}{2}$, $\frac{1}{2}$, and $1\frac{1}{2}$, the required values can be readily obtained. (It is necessary to be careful about the signs of the individual factors, thus, $-1\frac{1}{2} - (-2\frac{1}{2}) = -1\frac{1}{2} + 2\frac{1}{2} = +1$, and $-1\frac{1}{2} - 2\frac{1}{2} = -4$, and also about counting the number of $-$ signs in the reduced factors of the numerators and denominators of the coefficients of u_a, u_b, etc., and if the number be odd to make the sign $-$, and if even $+$. It may also be noted that the sum of all the coefficients in the expression for u_x should $= 1$.)

It is found that

$$u_{-1\frac{1}{2}} = [-6u_{-7\frac{1}{2}} + 108u_{-2\frac{1}{2}} + 27u_{2\frac{1}{2}} - 4u_{7\frac{1}{2}}] \div 125,$$

$$u_{-\frac{1}{2}} = [-8u_{-7\frac{1}{2}} + 84u_{-2\frac{1}{2}} + 56u_{2\frac{1}{2}} - 7u_{7\frac{1}{2}}] \div 125,$$

$$u_{\frac{1}{2}} = [-7u_{-7\frac{1}{2}} + 56u_{-2\frac{1}{2}} + 84u_{2\frac{1}{2}} - 8u_{7\frac{1}{2}}] \div 125,$$

$$u_{1\frac{1}{2}} = [-4u_{-7\frac{1}{2}} + 27u_{-2\frac{1}{2}} + 108u_{2\frac{1}{2}} - 6u_{7\frac{1}{2}}] \div 125.$$

On taking the sum of these four values $+ u_{-2\frac{1}{2}} + u_{2\frac{1}{2}}$, dividing by 6, and reducing to denominator of 24, the result obtained is

$$\frac{12{\cdot}8\,(u_{-2\frac{1}{2}} + u_{2\frac{1}{2}}) - {\cdot}8\,(u_{-7\frac{1}{2}} + u_{7\frac{1}{2}})}{24}.$$

By proceeding similarly the results already given for interpolations at intervals of $\frac{1}{2}$ year and $\frac{1}{4}$ year may be obtained, but it is only necessary to have four successive terms, and therefore, as the first term is already given, only three terms need be interpolated; when this is done the series can be summed by means of a formula.

If the interval be made $\frac{1}{10}$ year the coefficients in the numerator become 12·98 and − ·98; if $\frac{1}{100}$ year, they become 12·998 and − ·998; if $\frac{1}{1000}$ year, 12·9998 and − ·9998, and so on.

Thus by making the intervals less and less, nearer and nearer approximations are obtained to the limit

$$\frac{13\,(u_{-2\frac{1}{2}} + u_{2\frac{1}{2}}) - (u_{-7\frac{1}{2}} + u_{7\frac{1}{2}})}{24}.$$

(See Note 1.)

Fig. 1 having been intended to illustrate interpolation and *summation*, attention may now be directed to Fig. 2, which was designed to illustrate interpolation and *differencing*. (It may be remarked in passing that the only reason why the curve in one figure has been made convex and in the other concave has been to illustrate the point that some parts of Life-Table curves are convex and others concave.)

For the sake of simplicity let it now be supposed that the given five ordinates in Fig. 2 denoted by the general symbol u_x represent numbers of *population* at age x and upwards.

When five equidistant ordinates are given, separated by n units of interval, their relative distances apart measured from the central ordinate are $-2n$, $-n$, 0, n, and $2n$.

Now it can be shown by means of the differential calculus that the "differential coefficient," or in other words, the "limit of differencing" at the central point 0

$$= \frac{8\,(u_n - u_{-n}) - (u_{2n} - u_{-2n})}{12n}.$$

In the special instance now being considered $n = 5$.

It is now proposed to show how this limit may be approximately defined by simple interpolations.

x of course represents age in years. If x be increased by h, which may represent a number of years, a single year, or a fraction of a year, and if the ordinate u_{x+h} be interpolated, then the number belonging to the age-period x to $x+h$ will be measured by $u_{x+h} - u_x$, and $\frac{u_{x+h} - u_x}{h}$ will be the ratio of *number* to *length of period*, and may be considered to represent approximately the number living at the age which is the central point of the interval x to $x+h$.

Now it is obvious that if we wish to calculate thus the number of those supposed to be living near to the exact age x, the smaller we make h the more accurate will be the result.

The differential calculus measures the ratio $\frac{u_{x+h} - u_x}{h}$ when h vanishes, that is, when $h = 0$, and thus defines the "limit of differencing."

Similar reasoning applies to the ordinate u_{x-h}, and the smaller h is made the more nearly will $\frac{u_x - u_{x-h}}{h}$ approach $\frac{u_{x+h} - u_x}{h}$.

In order to obtain a symmetrical formula we may therefore take

$$\frac{1}{2}\left(\frac{u_{x+h} - u_x}{h} + \frac{u_x - u_{x-h}}{h}\right) = \frac{u_{x+h} - u_{x-h}}{2h}$$

as an approximation to the true "differential coefficient" which will become closer and closer to the true value as we make h smaller and smaller.

As a first approximation we may take $x = 0$ and $h = 1$; we have then to find the value of $\frac{u_1 - u_{-1}}{2}$.

By "Lagrange's formula" having given the series u_{-10}, u_{-5}, u_0, u_5, u_{10},

$$u_1 = [9u_{-10} - 66u_{-5} + 594u_0 + 99u_5 - 11u_{10}] \div 625,$$

$$u_{-1} = [-11u_{-10} + 99u_{-5} + 594u_0 - 66u_5 + 9u_{10}] \div 625,$$

$$\therefore\ \tfrac{1}{2}(u_1 - u_{-1}) = [165\,(u_5 - u_{-5}) - 20\,(u_{10} - u_{-10})] \div 1250,$$

and on reduction to denominator of 60 this expression becomes

$$[7{\cdot}92\,(u_5 - u_{-5}) - {\cdot}96\,(u_{10} - u_{-10})] \div 60.$$

By making h smaller and proceeding similarly the coefficients of the expressions arrived at are as follows:

when $h = \frac{1}{2}$, 7·98 and − ·99,

when $h = \frac{1}{4}$, 7·995 and − ·9975,

when $h = \frac{1}{10}$, 7·9992 and − ·9996,

when $h = \frac{1}{100}$, 7·999992 and − ·999996.

As we have seen, the limit when $h = 0$ is

$$[8\,(u_5 - u_{-5}) - (u_{10} - u_{-10})] \div 60.$$

(See Note 2.)

When the given series of u_x values represent the *logarithms* of the numbers of population, at age x and upwards, the logarithm of the "differential coefficient" at age 0 (which is strictly speaking the *ratio* $\frac{u_{x+h} - u_x}{h}$ when $h = 0$, but which may be taken as representing the number living at the exact age represented relatively by 0 in the given series) is found by the expression

$$u_0 + \log [8 (u_5 - u_{-5}) - (u_{10} - u_{-10})] - 1{\cdot}4159356,$$

where 1·4159356 = the logarithm of "M," the "modulus of the common logarithm," + the logarithm of 60.

In the actual work of calculating $\log p'_x$ values the factors $\frac{1}{M}$ and $\frac{1}{60}$ occur both in the numerator and denominator of the fraction $\frac{2P-d}{2P+d}$ and are cancelled out.

The explanation of the above formula in so far as it is modified in dealing with logarithms instead of numbers can only be given by referring to the "Exponential Theorem" and the "Theory of Logarithms," etc., and it is not proposed to attempt it here. Those who may be unable to verify it may take it upon trust, not merely on the authority of the writer, but as having been deduced by a master of the art of applying abstract mathematical principles to the elucidation of problems relating to vital statistics, Mr A. C. Waters.

Note 1.

This limit may be *exactly* obtained thus.

According to the integral calculus, having given a function of three orders of differences,

$$\int_{-n}^{n} u_x dx = 2nA + \frac{2n^3C}{3} = \frac{n}{3}(6A + 2n^2C).$$

The given series of u_x values being represented by u_a, u_b, u_c, and u_d, the value of $A = u_0$ may be found from the expression of Lagrange's formula already given on p. 185 by substituting 0 for x.

Thus $$u_0 = \frac{-bcd}{(a-b)(a-c)(a-d)} u_a + \ldots.$$

The value of C is expressed as follows.

$$C = \left[\frac{-(b+c+d)}{(a-b)(a-c)(a-d)}\right] u_a + \left[\frac{-(a+c+d)}{(b-a)(b-c)(b-d)}\right] u_b + \left[\frac{-(a+b+d)}{(c-a)(c-b)(c-d)}\right] u_c + \left[\frac{-(a+b+c)}{(d-a)(d-b)(d-c)}\right] u_d.$$

When the given terms of the series are equidistant, and when -3, -1, 1, and 3 are substituted respectively for a, b, c, and d,

$$A = \frac{9(u_{-1}+u_1) - (u_{-3}+u_3)}{16}.$$

$$C = \frac{-(u_{-1}+u_1) + (u_{-3}+u_3)}{16n^2}.$$

$$\therefore \frac{n}{3}(6A+2n^2C)=\frac{n}{3}\left[\frac{13(u_{-1}+u_1)-(u_{-3}+u_3)}{4}\right],$$

and on dividing this last expression by $2n$ it becomes

$$\frac{13(u_{-1}+u_1)-(u_{-3}+u_3)}{24}.$$

Note 2.

This limit may be *exactly* defined thus.

According to the differential calculus when

$$\phi(x)=A+Bx+Cx^2+\ldots,$$

$$\phi'(x)=B+2Cx+3Dx^2+\ldots,$$

$$\therefore \text{ when } x=0,\ \phi'(0)=B.$$

Having given a series of five linear quantities represented by u_0, u_a, u_b, u_c, and u_d,

$$B=\frac{-bcd}{a(a-b)(a-c)(a-d)}u_a+\frac{-acd}{b(b-a)(b-c)(b-d)}u_b+\frac{-abd}{c(c-a)(c-b)(c-d)}u_c$$
$$+\frac{-abc}{d(d-a)(d-b)(d-c)}u_d-\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}\right)u_0.$$

When the given terms are equidistant and when -2, -1, 1, and 2 are substituted respectively for a, b, c, and d, it is evident that the coefficient of u_0 is zero, and the formula will be found to work out to the value already given, the factor n coming into the denominator because in the expressions for the coefficients of u_a, u_b, etc., there are *three* factors in the numerators and *four* in the denominators.