On a New Method of Graduation.

By Professor E. T. Whittaker, F.R.S

(*Read 14th Nov. 1919. Received, in amended form, 8th Aug. 1923*).

§ 1. *Introductory.*

Suppose that as a result of observation or experience of some kind we have obtained a set of values of a variable u corresponding to equidistant values of its argument; let these be denoted by $u_1, u_2, \dots u_n$. If they have been derived from observations of some natural phenomenon, they will be affected by errors of observation; if they are statistical data derived from the examination of a comparatively small field, they will be affected by irregularities arising from the accidental peculiarities of the field; that is to say, if we examine another field and derive a set of values of u from it, the sets of values of u derived from the two fields will not in general agree with each other In any case, if we form a table of the differences $\Delta u_1 = u_2 - u_1$, $\Delta u_2 = u_3 - u_2$, ..., $\Delta^2 u_1 = \Delta u_2 - \Delta u_1$, etc., it will generally be found that these differences are so irregular that the difference-table cannot be used for the purposes to which a difference-table is usually put, viz., finding interpolated values of u, or differential coefficients of u with respect to its argument, or definite integrals involving u; before we can use the difference-tables we must perform a process of "smoothing," that is to say, we must find another sequence $u_1', u_2', u_3', \dots, u_n'$, whose terms differ as little as possible from the terms of the sequence $u_1, u_2, \dots u_n$, but which has regular differences. This smoothing process, leading to the formation of $u_1', u_2', \dots u_n'$, is called the *graduation* or *adjustment* of the observations.

Workers in experimental science generally deal with the problem by plotting the numbers $u_1, u_2, \dots u_n$, against the corresponding value of the argument, and drawing a freehand curve as nearly as possible through them. This somewhat arbitrary method is insufficient for the needs of Actuarial Science, and a large number of "graduation formulae" are to be found in the journals of the Actuarial Societies.

The standpoint of the present paper is, that the problem belongs essentially to the mathematical theory of Probability; we have the given observations, and they would constitute the "most probable" values of u for the corresponding values of the argument, were it not that we have *à priori* grounds for believing that the true values of u form a smooth series, the irregularities being due to accidental causes which it is desirable to eliminate. The problem is to combine all the materials of judgment—the observed values and the *à priori* considerations—in order to obtain the resulting "most probable" values of u.

§ 2. *The basis of the method in the theory of Probability.*

Let us then suppose that we are concerned with a number u_x which depends on an argument x, and suppose that we have n data $u_1, u_2, \ldots u_n$, which are affected with uncertainties or irregularities due, *e.g.*, to accidental errors of observation; so that when u_x is plotted as a function of x, the n points so obtained do not lie on a smooth curve, although there is a strong antecedent probability that if the observations had been more accurate the curve would have been smooth. We may make the somewhat vague word "smooth" more precise by interpreting it to mean that the third differences $\Delta^3 u_x$ are to be very small.

Now consider the following hypothesis; that the true value which should have been obtained by the observation for u_1 lies between u_1' and $u_1'+\sigma$ where σ is a small constant number; that the true value which should have been obtained by the observation for u_2 lies between u_2' and $u_2'+\sigma$, etc., and finally, that the true value which should have been obtained by the observation for u_n lies between u_n' and $u_n'+\sigma$. This hypothesis we shall call "hypothesis H."

Before the observations have been made we have nothing to guide us as to the probability of this hypothesis H except the degree of smoothness of the sequence $u_1', u_2', \ldots, u_n'$, which may be measured by the smallness of the sum of the squares of the third differences*

$$S=(u_4'-3u_3'+3u_2'-u_1')^2+(u_5'-3u_4'+3u_3'-u_2')^2+\ldots$$
$$+(u_n'-3u'_{n-1}+3u'_{n-2}-u'_{n-3})^2.$$

* The theory may be extended to the case when the observations are not taken at equidistant values of the argument, by taking instead of S the sum of the squares of the third *divided* differences of the graduated values.

We may therefore, by analogy with the normal law of frequency, suppose that the *à priori* probability of hypothesis H is $c\,e^{-\lambda^2 S}\,\sigma^n$ where c and λ denote constants.

Next, let us consider the *à priori* probability that the measures obtained by the observations will be $u_1, u_2, \ldots u_n$, on the assumption that hypothesis H is true. Since the true value of the first observed quantity is, on this hypothesis, u_1', the probability that a value between u_1 and $u_1+\sigma$ will actually be observed will (postulating the normal law of error) be

$$\frac{h_1}{\sqrt{\pi}}\,e^{-h_1^2(u_1-u_1')^2}\,\sigma\,,$$

where h_1 is a constant which measures the precision with which this observation can be made.

Similarly, the probability that a value between u_2 and $u_2+\sigma$ will actually be obtained for the second observed measure is

$$\frac{h_2}{\sqrt{\pi}}\,e^{-h_2^2(u_2-u_2')^2}\,\sigma,$$

where h_2 is the measure of precision of this observation.

Thus, on the assumption that hypothesis H is true, the *à priori* probability that the observed measure of the first observed quantity will be between u_1 and $u_1+\sigma$, the observed measure of the second observed quantity will be between u_2 and $u_2+\sigma$, etc., is

$$\frac{h_1, h_2, h_3, \ldots h_n}{\pi^{\frac{1}{2}n}}\,e^{-F}\,\sigma^n$$

where F denotes the sum

$$F=h_1^2\;(u_1-u_1')^2+h^2\,(u_2-u_2')^2+\ldots+h_n^2\;(u_n-u_n')^2.$$

The sums S and F enable us to express numerically the *smoothness* of the graduated values, and the *fidelity* of the graduated to the ungraduated values, respectively.

We must now make use of the fundamental theorem in the theory of Inductive Probability, which is as follows:—Suppose that a certain observed phenomenon may be accounted for by any one of a certain number of hypotheses, of which one, and not more than one, must be true: Suppose, moreover, that the probability of the s^{th} hypothesis, as based on information in our possession before the phenomenon is observed, is p_s, while the probability of the observed phenomenon on the assumption of the truth of the s^{th} hypothesis is P_s. Then when the observation of the phenomenon is taken into consideration, the probability of the s^{th} hypothesis is

$$\frac{p_s P_s}{\Sigma p_s P_s}$$

where the symbol Σ denotes summation over all the hypotheses.

It follows from this that whereas *before* the phenomenon was observed, the most probable hypothesis was that for which p_s was greatest, the most probable hypothesis *after* the phenomenon has been observed is that for which the product $P_s p_s$ is greatest. Applying this theorem to the case under consideration, we see that *the most probable hypothesis is that for which*

$$\frac{c\, h_1 h_2 \ldots h_n}{\pi^{\frac{1}{2}n}} e^{-\lambda^2 S - F} \sigma^{2n}$$

is a maximum, that is to say, *the most probable set* u_1', u_2', ... u_n' *of values of the quantities is that which makes*

$$\lambda^2 S + F$$

a minimum.

§ 3. *The analytical formulation.*

Writing down the ordinary conditions for a minimum, we obtain the equations

$$\begin{cases} h_1^2 u_1 = h_1^2 u_1' - \lambda^2 \Delta^3 u_1' \\ h_2^2 u_2 = h_2^2 u_2' + 3\lambda^2 \Delta^3 u_1' - \lambda^2 \Delta^3 u_2' \\ h_3^2 u_3 = h_3^2 u_3' - 3\lambda^2 \Delta^3 u_1' + 3\lambda^2 \Delta^3 u_2' - \lambda^2 \Delta^3 u_3' \\ h_4^2 u_4 = h_4^2 u_4' + \lambda^2 \Delta^3 u_1' - 3\lambda^2 \Delta^3 u_2' + 3\lambda^2 \Delta^3 u_3' - \lambda^2 \Delta^3 u_4' \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ h_n^2 u_n = h_n^2 u_n' + \lambda^2 \Delta^3 u'_{n-3}. \end{cases}$$

We shall now make the simplifying assumption that the measure of precision is the same for all the data, so

$$h_1 = h_2 = \ldots = h_n.*$$

* If this is not the case, we graduate some function of u, such as $\log u$, instead of u, choosing this function so that its measure of precision has nearly the same value for all values of the argument.

If we write $h_1^2 = h_2^2 = \ldots = \epsilon\lambda^2$ the equations may now be written

$$\left.\begin{array}{l} \epsilon u_1 = \epsilon u_1' - \Delta^3 u_1' \\ \epsilon u_2 = \epsilon u_2' + 3\Delta^3 u_1' - \Delta^3 u_2' \\ \epsilon u_3 = \epsilon u_3' - 3\Delta^3 u_1' + 3\Delta^3 u_2' - \Delta^3 u_3' \\ \epsilon u_4 = \epsilon u_4' + \Delta^3 u_1' - 3\Delta^3 u_2' + 3\Delta^3 u_3' - \Delta^3 u_4' \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \epsilon u_n = \epsilon u_n' + \Delta^3 u_{n-3}' . \end{array}\right\} \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

Now all the equations, except the three first and the three last, are of the form

$$\epsilon u_x = \epsilon u_x' - \Delta^6 u'_{x-3} .$$

Moreover, if we introduce a quantity u_0' such that $\Delta^3 u_0' = 0$, the third equation becomes

$$\epsilon u_3 = \epsilon u_3' - \Delta^6 u_0',$$

which is of the same form ; and similarly the first two and last three equations can be brought to the same form by introducing new quantities u'_{-1}, u'_{-2}, u'_{n+1}, u'_{n+2} , u'_{n+3}, such that

$$\Delta^3 u'_{-1} = 0, \quad \Delta^3 u'_{-2} = 0, \quad \Delta^3 u'_{n-2} = 0, \quad \Delta^3 u'_{n-1} = 0, \quad \Delta^3 u_n' = 0.$$

Thus *the graduated values u_x' satisfy the linear difference-equation*

$$\epsilon u_x' - \Delta^6 u'_{x-3} = \epsilon u_x , \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

being in fact the particular solution of this equation which satisfies the six terminal conditions

$$\Delta^3 u_0' = 0,\ \Delta^3 u'_{-1} = 0,\ \Delta^3 u'_{-2} = 0,\ \Delta^3 u'_{n-2} = 0,\ \Delta^3 u'_{n-1} = 0,\ \Delta^3 u_n' = 0 \ldots (3)$$

whence we have at once

$$\Delta^4 u'_{-2} = 0,\ \Delta^4 u'_{-1} = 0,\ \Delta^5 u'_{-2} = 0,\ \Delta^4 u'_{n-2} = 0,\ \Delta^4 u'_{n-1} = 0,\ \Delta^5 u'_{n-2} = 0 \quad (4)$$

§ 4. *The Theorems of Conservation.*

From (2) we have by summation

$$\begin{aligned} \epsilon(u_1' + u_2' + \ldots + u_n') - \epsilon(u_1 + u_2 + \ldots + u_n) &= \Delta^6 u'_{-2} + \Delta^6 u'_{-1} + \ldots + \Delta^6 u'_{n-} \\ &= \Delta^5 u'_{n-2} - \Delta^5 u'_{-2} \\ &= 0 \qquad \text{by (4).} \end{aligned}$$

Therefore $\qquad u_1' + u_2' + \ldots + u_n' = u_1 + u_2 + \ldots + u_n \quad \ldots\ldots\ldots(5)$

https://doi.org/10.1017/S0013091500077853 Published online by Cambridge University Press

Moreover, by (2)

$$\epsilon(u_1' + 2u_2' + 3u_3' + \ldots + nu_n') - \epsilon(u_1 + 2u_2 + \ldots + nu_n)$$
$$= \Delta^6 u'_{-2} + 2\Delta^6 u'_{-1} + \ldots + n\Delta^6 u'_{n-3}$$
$$= n\Delta^5 u'_{n-2} - \Delta^4 u'_{n-2} + \Delta^4 u'_{-2}$$
$$= 0 \qquad \text{by (4)}.$$

Therefore

$$u_1' + 2u_2' + 3u_3' + \ldots + nu_n' = u_1 + 2u_2 + 3u_3 + \ldots + nu_n \quad \ldots\ldots\ldots(6)$$

Next, by (2)

$$\epsilon(u_1' + 2^2 u_2' + 3^2 u_2' + \ldots + n^2 u_n') - \epsilon(u_1 + 2^2 u_2 + 3^2 u_3 + \ldots + n^2 u_n)$$
$$= \Delta^6 u'_{-2} + 2^2\Delta^6 u'_{-1} + \ldots + n^2\Delta^6 u'_{n-3}$$
$$= n^2\Delta^5 u'_{n-2} - (2n-1)\Delta^4 u'_{n-2} + 2\Delta^3 u'_{n-2} - \Delta^3 u'_{-2} - \Delta^3 u_{-1}$$
$$= 0 \qquad \text{by (3) and (4)}.$$

Therefore

$$u_1' + 2^2 u_2' + 3^2 u_3' + \ldots + n^2 u_2' = u_1 + 2^2 u_2 + 3^2 u_3 + \ldots + n^2 u_n \ldots\ldots\ldots(7)$$

Equations (5), (6), (7) show that *the moments of orders* 0, 1, 2 *are the same for the graduated data as for the original data.* This may be called the *Theorem of Conservation of Moments.* We may express it by saying that *the graph which represents the ungraduated data and the graph which represents the graduated data have the same area, the same x-coordinate of the centre of gravity, and the same moment of inertia about any line parallel to the axis of u.*

Thus by this method we secure that the total of the u's and their first and second moments shall be the same in the graduated table as in the actual statistics on which it is based.

§ 5. *The numerical process of graduation.*

The parameter ϵ is at our disposal, and measures the importance which we attach to keeping close to the original data, as weighed against our desire to attain perfect smoothness in the graduated curve. If ϵ were taken absolutely zero we should obtain a perfectly smooth graduated curve which would have the same moments of orders 0, 1 and 2 as the ungraduated curve, but in other respects might not fit the observed data closely. In practice therefore we do not take ϵ to be absolutely zero, but it may usually be taken to be a small number, so that it is convenient to expand the solution in ascending powers of ϵ and retain only the part which is independent of ϵ together with the part which involves the first power of ϵ: the parts involving higher powers of ϵ may be neglected.

Suppose then that the terms in the graduated value u_x' are arranged according to the powers of ϵ which they involve, thus

$$u_x' = u'_{x,0} + \epsilon u'_{x,1} + \epsilon^2 u'_{x,2} + \dots \qquad \dots\dots\dots\dots\dots\dots (8)$$

Substituting in (2) and equating the coefficients of ϵ, we have

$$\epsilon\, u'_{x,0} - \Delta^6 u'_{x-3,1} = \epsilon\, u_x \qquad \dots\dots\dots\dots\dots\dots (9)$$

which is a linear difference-equation to determine $u'_{x,1}$, if $u'_{x,0}$ can first be found.

Now $u'_{x,0}$ can be found without difficulty in the following way: From equation (1) it follows at once that, when ϵ is zero, the third differences of the graduated values are all zero: so $u'_{x,0}$ must be a polynomial of degree two in x, say,

$$u'_{x,0} = a + bx + cx^2 \qquad \dots\dots\dots\dots\dots\dots (10)$$

where a, b, c, are independent of x. Substituting in equations (5), (6), (7), they become

$$\left.\begin{aligned} &na + \tfrac{1}{2}n(n+1)\,b + \tfrac{1}{6}n(n+1)(2n+1)\,c = M_0 \\ &\tfrac{1}{2}n(n+1)\,a + \tfrac{1}{6}n(n+1)(2n+1)\,b + \tfrac{1}{4}n^2(n+1)^2\,c = M_1 \\ &\tfrac{1}{6}n(n+1)(2n+1)\,a + \tfrac{1}{4}n^2(n+1)^2\,b \\ &\qquad + \tfrac{1}{30}n(n+1)(6n^3+9n^2+n-1)\,c = M_2 \end{aligned}\right\}\dots\dots (11)$$

where M_0, M_1, M_2 denote the moments

$$(u_1 + u_2 + \dots + u_n),\ (u_1 + 2u_2 + \dots + nu_n),\ \text{and}\ (u_1 + 2^2u_2 + \dots + n^2u_n)$$

of the ungraduated data. The three equations (11) determine a, b, c; the solution may conveniently be performed as follows:—

Compute successively the numbers p, q, r, s, t, where

$$p = \frac{M_0}{n},\quad q = \frac{2M_1}{n(n+1)},\quad r = \frac{6M_2}{n(n+1)},\quad s = \frac{6(q-p)}{n-1},$$

$$t = \frac{2\{r-(2n+1)p\}}{n-1}.$$

Then c is given by

$$c = \frac{15\{t-(n+1)s\}}{(n+2)(n-2)}$$

b is then given by

$$b = s - (n+1)c$$

and a is then given by

$$a = q - \tfrac{1}{3}(2n+1)b - \tfrac{1}{2}n(n+1)c \qquad \dots\dots (12)$$

https://doi.org/10.1017/S0013091500077853 Published online by Cambridge University Press

The first of equations (11) may be used as a check.

Substituting the numerical values of $a\ b\ c$ thus found in equation (10) we obtain the formula for $u'_{x,0}$: and by substituting $x = 1,\ 2,\ 3,\ \ldots$ in it, we obtain the numerical values of $u'_{1,0},\ u'_{2,0},\ u'_{3,0},\ \ldots\ u'_{n,0}$.

Thus, performing the work in algebraical symbols for the case $n = 7$, we find in this way

$$\left.\begin{aligned} a &= \tfrac{1}{7} M_2 - \tfrac{9}{7} M_1 + \tfrac{17}{7} M_0 \\ b &= -\tfrac{8}{84} M_2 + \tfrac{67}{84} M_1 - \tfrac{9}{7} M_0 \\ c &= \tfrac{1}{84} M_2 - \tfrac{8}{84} M_1 + \tfrac{1}{7} M_0 \end{aligned}\right\}$$

Substituting these values of a, b, c, in (10), we obtain

$$u'_{10} = \tfrac{1}{42}\,(32u_1 + 15u_2 + 3u_3 - 4u_4 - 6u_5 - 3u_6 + 5u_7)$$

and, similarly, for u'_{20}, u'_{30}, … . A check is afforded by verifying that the last equation may be written

$$u'_{10} - u_1 = \tfrac{1}{42}\,(10\Delta^3 u_1 + 15\Delta^3 u_2 + 12\Delta^3 u_3 + 5\Delta^3 u_4),$$

since the values of $u_{10}' - u_1$, $u_{20}' - u_2$, … thus found must always be expressible as linear combinations of the third differences of the ungraduated data.

Having thus found $u'_{x,0}$ in terms of the ungraduated data, by substitution in equation (9) we obtain $\Delta^6 u'_{x-3,1}$ in terms of the ungraduated data. Denoting $\Delta^3 u'_{x,1}$ by v_x, we therefore have $\Delta^3 v_x$ known; and from (3) and (4) we have

$$\Delta^3 v_{-2} = 0, \qquad \Delta v_{-1} = 0, \qquad v_0 = 0,$$

so by mere summation of a difference-table we can obtain all the v's.

Then, continuing the literal working in the case $n = 7$, we have

$$\Delta^3 v_{-2} = u_{10}' - u_1 = \tfrac{1}{42}\,(10y_1 + 15y_2 + 12y_3 + 5y_4),$$

where y_x denotes $\Delta^3 u_x$; similarly

$$\begin{aligned} 42\ \Delta^3 v_{-1} &= -15y_1 - 15y_2 - 9y_3 - 3y_4 \\ 42\ \Delta^3 v_0 &= -3y_1 - 18y_2 - 15y_3 - 6y_4, \end{aligned}$$

and so on; writing these down in the Δ^3 column of a difference-table, and forming the Δ^2, Δ^1, and Δ^0 columns by summation, we obtain the complete difference-table as follows:—

	Δ	Δ^2	Δ^3
$42v_{-2}=0$			
	0		
$42v_{-1}=0$		0	
	0		$10y_1+15y_2+12y_3+5y_4$
$42v_0=0$		$10y_1+15y_2+12y_3+5y_4$	
	$10y_1+15y_2+12y_3+5y_4$		$-15y_1-15y_2-9y_3-3y_4$
$42v_1=10y_1+15y_2+12y_3+5y_4$		$-5y_1+3y_3+2y_4$	
	$5y_1+15y_2+15y_3+7y_4$		$-3y_1-18y_2-15y_3-6y_4$
$42v_2=15y_1+30y_2+27y_3+12y_4$		$-8y_1-18y_2-12y_3-4y_4$	
	$-3y_1-3y_2+3y_3+3y_4$		$4y_1+6y_2-6y_3-4y_4$
$42v_3=12y_1+27y_2+30y_3+15y_4$		$-4y_1-12y_2-18y_3-8y_4$	
	$-7y_1-15y_2-15y_3-5y_4$		$6y_1+15y_3+18y_3+3y_4$
$42v_4=5y_1+12y_2+15y_3+10y_4$		$2y_1+3y_2-5y_4$	
	$-5y_1-12y_2-15y_3-10y_4$		$3y_1+9y_2+15y_3+15y_4$
$42v_5=0$		$5y_1+12y_2+15y_3+10y_4$	
	0		$-5y_1-12y_2-15y_3-10y_4$
		0	

The results $\Delta^2 v_5=0$, $\Delta v_5=0$, $v_5=0$, furnish a check on the accuracy of the working: and as a further check we may form the columns Δ^4, Δ^5, Δ^6 in this difference-table: the column Δ^6 should give simply $-42y_1$, $-42y_2$, $-42y_3$, $-42y_4$.

Having now obtained the numbers v_1, v_2, v_3, v_{n-3} we have to find the numbers $u'_{1,1}, u'_{2,1}, \ldots u'_{n,1}$ from them. For this we use the conditions that (1) $u'_{x,1}$ satisfies the difference-equation

$$\Delta^3 u'_{x,1} = v_x \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(13)$$

and (2) that it is the particular solution of this difference-equation for which the moments of orders 0, 1 and 2 vanish. So in order to compute $u'_{x,1}$, we write down $v_1, v_2, \ldots v_{n-3}$ as the third column of a difference-table, and form the second, first and zero columns by summation, taking any arbitrary numbers whatever for the entries at the top of the columns. In this way we obtain in the zero column a set of numbers $w_1, w_2, \ldots w_n$ which satisfy the difference-equation (13), but which are not the particular solution we require. However, any two solutions of (13) differ only by a solution of the difference-equation $\Delta^3 y = 0$, *i.e.* they differ only by a quadratic function of x. So we can write

$$u'_{x,1} = w_x - A - Bx - Cx^2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(14)$$

and we have now only to determine A, B and C. For this we use the second of the above conditions; denoting the sums

$$(w_1 + w_2 + \ldots + w_n),\ (w_1 + 2w_2 + \ldots + nw_n),\ (w_1 + 2^2 w_2 + \ldots + n^2 w_n)$$

by N_0, N_1, N_2 respectively, we have by summing equation (14)

$$\left.\begin{aligned} &nA + \tfrac{1}{2}n(n+1)B + \tfrac{1}{6}n(n+1)(2n+1)C = N_0 \\ &\tfrac{1}{2}n(n+1)A + \tfrac{1}{6}n(n+1)(2n+1)B \\ &\qquad\qquad + \tfrac{1}{4}n^2(n+1)^2 C = N_1 \\ &\tfrac{1}{6}n(n+1)(2n+1)A + \tfrac{1}{4}n^2(n+1)^2 B \\ &\qquad\qquad + \tfrac{1}{30}n(n+1)(6n^3 + 9n^2 + n - 1)C = N_2 \end{aligned}\right\} \ldots\ldots(15)$$

These equations are of the same type as equations (11), and are solved in the same way; that is we compute successively

$$P = \frac{N_0}{n},\quad Q = \frac{2N_1}{n(n+1)},\quad R = \frac{6N_2}{n(n+1)},\quad S = \frac{6(Q-P)}{n-1},$$
$$T = \frac{2\{R - (2n+1)P\}}{n-1},$$

then C is given by

$$C = \frac{15\{T - (n+1)S\}}{(n+2)(n-2)}.$$

B is given by $\qquad B = S - (n+1)C$

and A is given by

$$A = Q - \tfrac{1}{3}(2n+1)B - \tfrac{1}{2}n(n+1)C.$$

The first of equations (15) may be used as a check.

https://doi.org/10.1017/S0013091500077853 Published online by Cambridge University Press

Having thus found A, B, C we substitute in equation (14), and so calculate $u'_{x,1}$ for $x=1, 2, 3, \ldots n$. Lastly, from the equation

$$u_x' = u'_{x,0} + \epsilon u'_{x,1}$$

we compute the graduated values u'_x for $x=1, 2, \ldots, n$. The graduation is thus completed.

The quantity ϵ, which is at our disposal, is not selected until the end of the process, when we try two or three different values and see which gives the most satisfactory result. By increasing ϵ we bring the graduated values into close fidelity to the ungraduated values, while by diminishing ϵ we make the sequence of graduated values smoother. There is not much labour involved in these trials as they merely amount to multiplying the column of known values of $u'_{x,1}$ by the final value of ϵ, and adding to the column of known values of $u'_{x,0}$.

The advantages of this method of graduation seem to be

(1) Its elasticity, due to the freedom of choice of ϵ. A satisfactory method of graduation *ought* to possess such elasticity, because the degree to which we are justified in sacrificing fidelity in order to obtain smoothness varies greatly from one problem to another.

(2) Its more logical basis in the mathematical theory of Probability.

(3) The total of the u's and their first and second moments are the same in the graduated table as in the actual statistics on which it is based. (These conditions are not satisfied in methods such as Sheppard's or Spencer's, which depend on formulae for graduating individual values.)

(4) It makes use of the whole material available to obtain each graduated value, whereas in *e.g.* Spencer's formula each value is graduated by using only it and its ten nearest neighbours on either side, and therefore the material used in order to graduate one value is slightly different from the material used in order to graduate the next member in the sequence.

(5) There is no difficulty near the beginning and end of the sequence, whereas Spencer's formula cannot be applied when we are within ten places of either terminal.

(6) These advantages are not counterbalanced by greater labour in the computations.

§ 12. *An example.*

A short section of the Government Female Annuitants (1883) Ultimate Table is here graduated by (i) Spencer's formula (*Journal of the Institute of Actuaries 38* (1904), p 334, *41* (1907), p 361), by (ii) Todhunter's method of interlaced parabolas (*ibid 53* (1922) p. 92), (iii) by the method of the present paper, taking $\epsilon = 0$ (iv), and taking $\epsilon = 0{\cdot}01$, and (v) taking $\epsilon = 0{\cdot}08$.

x (= Age)	Ungraduated		Graduated by Spencer's formulae		Graduated by Todhunter's method		Graduated by the method of this paper with $\epsilon=0$		Graduated by the method of this paper with $\epsilon=0{\cdot}01$		Graduated by the method of this paper with $\epsilon=0{\cdot}08$	
	$q\times10^5$	Δ^3	q	Δ^3	q	Δ^3	q	Δ^3	q	Δ^3	q	Δ^3
50	1019		1278		1298		1244		1219		1048	
51	1550		1382		1391		1379		1386		1436	
		551		−9		−13		1		4		16
52	1611		1494		1497		1504		1525		1673	
		−204		−8		−8		0		3		42
53	1753		1605		1603		1620		1640		1775	
		−120		−5		6		−1		8		58
54	1772		1707		1701		1727		1734		1784	
		+941		2		−4		0		9		60
55	1548		1795		1797		1824		1815		1758	
		−1271		5		1		1		3		47
56	2022		1871		1887		1911		1892		1757	
		591		10		10		−1		7		34
57	1923		1940		1972		1989		1968		1828	
		550		8		3		0		1		17
58	1842		2012		2062		2057		2050		2005	
59	2329		2095		2160		2115		2139		2305	
Sums	17369		17179		17368		17370		17368		17369	

https://doi.org/10.1017/S0013091500077853 Published online by Cambridge University Press

The sum of the absolute values of the differences between the graduated and the corresponding ungraduated members is 1576 for Spencer's method, 1595 for Todhunter's, 1563 for the present method with $\epsilon = 0$, 1438 for the present method with $\epsilon = 0{\cdot}01$, and 996 for the present method with $\epsilon = 0{\cdot}08$. An examination of the figures shows that so far as smoothness alone is concerned the best result is obtained from the present method with $\epsilon = 0$ (as of course is inevitable) and that the graduated values thus obtained show a slightly greater fidelity to the ungraduated values than is attained by Spencer's or Todhunter's formula. If, however, we wish to attach more importance to fidelity, the new method with $\epsilon = 0{\cdot}08$ yields graduated values which are very much closer to the ungraduated values, and whose third differences are fairly regular and not very large.