TFA 38 (1981-1983) 407-417

ON ESTIMATING THE FORCE OF MORTALITY

by

J. J. McCUTCHEON, M.A., Ph.D., F.F.A.

1. INTRODUCTION

In the a(55) Tables for Annuitants μ_x, the ultimate force of mortality at age x, was calculated by the formula

$$\mu_x = \frac{-d_{x-2}+7d_{x-1}+7d_x-d_{x+1}}{12l_x}. \qquad (1)$$

For the more recently published a(90) Tables the ultimate and select (duration 0) values of the force of mortality were found as

$$\mu_x = -\tfrac{1}{2}(\log p_{x-1}+\log p_x)+\tfrac{1}{12}(\Delta^2 \log p_{x-2}+\Delta^2 \log p_{x-1})$$
$$-\tfrac{1}{60}(\Delta^4 \log p_{x-3}+\Delta^4 \log p_{x-2})+\tfrac{1}{280}(\Delta^6 \log p_{x-4}+\Delta^6 \log p_{x-3}) \qquad (2)$$

and

$$\mu_{[x]} = -\tfrac{3}{2}\log p_{[x]}+\tfrac{1}{2}\log p_{x+1} \qquad (3)$$

respectively. Formula (2) was also used in relation to the PA(90) Tables. (An incorrect form of equation (2) seems to have been used in the A1967-70 Tables.)

The essential point about the above formulae is, of course, that under appropriate conditions each expresses the force of mortality in terms of the sequence of values $\{l_y\}$ (or the corresponding select values). At first sight the formulae appear to be fundamentally different and the form of equation (2), with finite difference operators, may be somewhat unfamiliar. We shall call equation (1) a " type A " formula and equations (2) and (3) formulae of " type B ", terms which we shall make more precise in §2 and §3 below.

The purpose of this note is to indicate that, while in an obvious sense formulae of type A and type B differ, the derivation of each type of formula depends essentially on the same mathematics. Without relying on finite difference methods we derive quite simply from first principles more general results than those above. Finally, we illustrate our remarks by some numerical results, obtained in experiments relating to the recently published FA1975-78 mortality table for female assured lives.

Although our remarks are necessarily of a somewhat elementary nature, we hope that the fairly general form of our results may be of use for future reference purposes.

2. FORMULAE OF TYPE A

Since, by definition,

$$\mu_x = -\frac{1}{l_x} . l'_x$$

and

$$d_x = l_x - l_{x+1},$$

it is readily verified that equation (1) above is equivalent to

$$l'_x = \frac{l_{x-2} - 8l_{x-1} + 8l_{x+1} - l_{x+2}}{12}.$$

This last equation is valid when l_y is a quartic in y on the age interval $[x-2, x+2]$. This remark indicates what we mean by a " type A " formula. Such a formula is defined to be one which expresses the force of mortality in terms of certain values of the function l on the assumption that over some specified age-range l_x *is a polynomial in* x of appropriate degree. We wish to consider type A formulae in greater generality.

We begin with one simple observation. If

$$f(x) = (x-x_1)(x-x_2)\dots(x-x_{r-1})(x-x_{r+1})\dots(x-x_n)$$

$$= \prod_{\substack{k=1 \\ k \neq r}}^{n} (x-x_k), \tag{4}$$

then (trivially by logarithmic differentiation)

$$f'(x) = \sum_{\substack{t=1 \\ t \neq r}}^{n} \left\{ \prod_{\substack{k=1 \\ k \neq r,t}}^{n} (x-x_k) \right\}. \tag{5}$$

Now suppose that

$$x_1 < x_2 < \dots < x_n$$

is a given increasing sequence of real numbers. Suppose also that $p(x)$ is a polynomial in x of degree $(n-1)$ and that, for $r = 1, 2, \dots, n$, the value of $p(x_r)$ is known. If

$$p(x_r) = p_r, \tag{6}$$

then clearly the polynomial p is determined by the n values p_1, $p_2, \dots, p_n$. For $i = 1, 2, \dots, n$ we wish to express $p'(x_i)$ in terms of these values.

The polynomial p may be expressed directly in Lagrange's form as

$$p(x) = \sum_{j=1}^{n} p_j . \left\{ \prod_{\substack{k=1 \\ k \neq j}}^{n} \left(\frac{x-x_k}{x_j-x_k} \right) \right\}$$

$$= \sum_{j=1}^{n} z_j . \left\{ \prod_{\substack{k=1 \\ k \neq j}}^{n} (x - x_k) \right\}, \tag{7}$$

where

$$z_j = \frac{p_j}{\prod\limits_{\substack{k=1 \\ k \neq j}}^{n} (x_j - x_k)}. \tag{8}$$

By picking out the term with $j = i$, we may write equation (7) as

$$p(x) = z_i . \prod_{\substack{k=1 \\ k \neq i}}^{n} (x - x_k) + \sum_{\substack{j=1 \\ j \neq i}}^{n} z_j . \left\{ \prod_{\substack{k=1 \\ k \neq j}}^{n} (x - x_k) \right\}.$$

Using equations (4) and (5) above to differentiate this last equation and the letting $x = x_i$, we obtain

$$p'(x_i) = z_i . \sum_{\substack{t=1 \\ t \neq i}}^{n} \left\{ \prod_{\substack{k=1 \\ k \neq i,t}}^{n} (x_i - x_k) \right\} + \sum_{\substack{j=1 \\ j \neq i}}^{n} z_j . \left[\sum_{\substack{t=1 \\ t \neq j}}^{n} \left\{ \prod_{\substack{k=1 \\ k \neq j,t}}^{n} (x_i - x_k) \right\} \right]. \tag{9}$$

Brief consideration shows that the term within square brackets in the second expression of this last equation can be simplified considerably. If $t \neq i$, then (since from the outer summation $j \neq i$) the product term contains the factor $(x_i - x_i)$ and is thus zero. The only non-zero term in the inner summation therefore occurs when $t = i$, so that equation (9) reduces to

$$p'(x_i) = z_i \sum_{\substack{t=1 \\ t \neq i}}^{n} \left\{ \prod_{\substack{k=1 \\ k \neq i,t}}^{n} (x_i - x_k) \right\} + \sum_{\substack{j=1 \\ j \neq i}}^{n} z_j . \left\{ \prod_{\substack{k=1 \\ k \neq j,i}}^{n} (x_i - x_k) \right\}.$$

Now substitute in this last equation the values of z_i and z_j from equation (8). This gives

$$p'(x_i) = p_i \frac{\sum\limits_{\substack{t=1 \\ t \neq i}}^{n} \left\{ \prod\limits_{\substack{k=1 \\ k \neq i,t}}^{n} (x_i - x_k) \right\}}{\prod\limits_{\substack{k=1 \\ k \neq i}}^{n} (x_i - x_k)} + \sum_{\substack{j=1 \\ j \neq i}}^{n} p_j \frac{\prod\limits_{\substack{k=1 \\ k \neq j,i}}^{n} (x_i - x_k)}{\prod\limits_{\substack{k=1 \\ k \neq j}}^{n} (x_j - x_k)}$$

or

$$p'(x_i) = p_i\left[\sum_{\substack{t=1\\t\neq i}}^{n} \frac{1}{x_i - x_t}\right] + \sum_{\substack{j=1\\j\neq i}}^{n} p_j \cdot \left[\frac{\prod\limits_{\substack{k=1\\k\neq i,j}}^{n} (x_i - x_k)}{\prod\limits_{\substack{k=1\\k\neq j}}^{n} (x_j - x_k)}\right]$$

Replacing the summation variable in the first term of the right-hand side by a new variable k and taking out the factor $(x_j - x_i)$ in the denominator of second term, we may write the equation in the form

$$p'(x_i) = p_i\left[\sum_{\substack{k=1\\k\neq i}}^{n} \frac{1}{x_i - x_k}\right] + \sum_{\substack{j=1\\j\neq i}}^{n} p_j \cdot \left[\frac{1}{x_j - x_i} \cdot \prod_{\substack{k=1\\k\neq i,j}}^{n} \left(\frac{x_i - x_k}{x_j - x_k}\right)\right]$$

or (from our definition (6))

$$p'(x_i) = \sum_{j=1}^{n} c_{ij} p(x_j), \tag{10}$$

where, for $1 \leqslant i, j \leqslant n$,

$$\left.\begin{aligned} c_{ii} &= \sum_{\substack{k=1\\k\neq i}}^{n} \frac{1}{x_i - x_k} \\ \text{and} \quad c_{ij} &= \frac{1}{x_j - x_i} \cdot \prod_{\substack{k=1\\k\neq i,j}}^{n} \left(\frac{x_i - x_k}{x_j - x_k}\right) \quad (i \neq j). \end{aligned}\right\} \tag{11}$$

Equations (10) and (11) are the key to further progress. In relation to a polynomial of degree $(n-1)$ which takes known values at given points $x_1, x_2, \ldots, x_n$, the equations express the value of the derivative of the polynomial at each of the given points as an appropriate linear combination of the known function values. For any given sequence $\{x_1, x_2, \ldots, x_n\}$ the weights $\{c_{ij}\}$ are easily calculated.

By considering a polynomial which is identically equal to 1 (with derivative everywhere zero), note that equation (10) implies that

$$\sum_{j=1}^{n} c_{ij} = 0 \quad (i = 1, 2, \ldots, n). \tag{12}$$

By combining our previous results we immediately obtain the equation

$$\mu_{x_i} = -\left[\frac{\sum\limits_{j=1}^{n} c_{ij} l_{x_j}}{l_{x_i}}\right] \quad (i = 1, 2, \ldots, n), \tag{13}$$

which is valid if l_x is a polynomial of degree $(n-1)$ on the age-interval $[x_1, x_n]$. This is the most general formula of type A.

A fairly common application of this formula is to be found when l_x is known at n consecutive ages and it is desired to estimate the force of mortality at one or more of these ages. For the remainder of this section we restrict our attention to this special situation.

If the ages at which l_x is known are $x_1 = \alpha+1$, $x_2 = \alpha+2$, ..., $x_n = \alpha+n$, then $(x_r - x_s) = r-s$ and the coefficients $\{c_{ij}\}$ defined by equation (11) above do *not* depend on α. Then, if l_x is a polynomial in x of degree $(n-1)$ on the interval $[\alpha+1, \alpha+n]$, we have as a special case of equation (13) (for future reference)

$$\mu_{\alpha+i} = -\frac{\sum_{j=1}^{n} c_{ij} l_{\alpha+j}}{l_{\alpha+i}} \quad (i = 1, 2, \ldots, n), \tag{14}$$

where in this last equation c_{ij} is defined by equation (11) with $x_r = r$ $(r = 1, 2, \ldots, n)$. [More precisely we should write $c_{ij}(n)$, but in any given situation the value of n will be known and no confusion will arise, if we simply write c_{ij}.] It is not necessary to evaluate all the coefficients since a skew-symmetry exists. In the special case when $x_r = r$ $(r = 1, 2, \ldots, n)$ it is easily verified from equation (11) that

$$c_{n+1-i,\, n+1-j} = -c_{ij} \quad (1 \leqslant i, j \leqslant n). \tag{15}$$

Thus in this special case the $n \times n$ array $\{c_{ij}\}$ is such that the entries of row $(n+1-i)$ are the entries of row i *in reverse order and with opposite signs*. This result is obvious by general reasoning.

Since $(x_r - x_s) = r-s$, each of the coefficients $\{c_{ij}\}$ is a rational number. Moreover it is easily seen that $(n-1)!c_{ij}$ is an integer. Thus each of the coefficients $\{c_{ij}\}$ can be expressed as a fraction with denominator $(n-1)!$. (In fact for most values of n it is possible to simplify this denominator further.)

In practice it is usual to consider those situations for which n is odd and for reasons of space we restrict our attention to such cases, although, of course, the above results are valid also if n is even. For $n = 3$, 5, 7, and 9 we give in Tables 1—4 below appropriate multiples of the coefficients $\{c_{ij}\}$, calculated by equation (11).

TABLE 1

$(n = 3)$

Values of $2c_{ij}$

	j		
i	1	2	3
1	−3	4	−1
2	−1	0	1

TABLE 2

($n = 5$)

Values of $12c_{ij}$

i	j = 1	2	3	4	5
1	−25	48	−36	16	−3
2	−3	−10	18	−6	1
3	1	−8	0	8	−1

TABLE 3

($n = 7$)

Values of $60c_{ij}$

i	j = 1	2	3	4	5	6	7
1	−147	360	−450	400	−225	72	−10
2	−10	−77	150	−100	50	−15	2
3	2	−24	−35	80	−30	8	−1
4	−1	9	−45	0	45	−9	1

TABLE 4

($n = 9$)

Values of $840c_{ij}$

i	j = 1	2	3	4	5	6	7	8	9
1	−2,283	6,720	−11,760	15,680	−14,700	9,408	−3,920	960	−105
2	−105	−1,338	2,940	−2,940	2,450	−1,470	588	−140	15
3	15	−240	−798	1,680	−1,050	560	−210	48	−5
4	−5	60	−420	−378	1,050	−420	140	−30	3
5	3	−32	168	−672	0	672	−168	32	−3

As an illustration we consider three examples of the use of equation (14), with $n = 5$ in each case.

Putting $\alpha = x-3$ and $i = 3$ in equation (14), we obtain immediately from Table 2

$$\mu_x = \frac{-l_{x-2}+8l_{x-1}-8l_{x+1}+l_{x+2}}{12l_x},$$

which, as we have already remarked, is equivalent to formula (1) above. Likewise putting $\alpha = x-1$ and $i = 1$ in equation (14), we obtain (again from Table 2)

$$\mu_x = \frac{25l_x-48l_{x+1}+36l_{x+2}-16l_{x+3}+3l_{x+4}}{12l_x}. \tag{16}$$

Alternatively, if in equation (14) we put $\alpha = x-1$ and $i = 2$, we obtain (still from Table 2)

$$\mu_{x+1} = \frac{3l_x+10l_{x+1}-18l_{x+2}+6l_{x+3}-l_{x+4}}{12l_{x+1}}. \tag{17}$$

The " select " equivalents of these last equations (16) and (17) could be used to estimate the force of mortality at durations 0 and 1 respectively for a select mortality table.

Finally note how equation (15) may be used. If, for example, it is desired to estimate μ_{x+3} in terms of $l_x, l_{x+1}, \ldots, l_{x+4}$, we consider equation (14) with $n = 5$, $\alpha = x-1$, and $i = 4$. This gives

$$\begin{aligned} \mu_{x+3} &= -\left[\sum_{j=1}^{5} c_{4j} l_{x-1+j}\right] / l_{x+3} \\ &= \left[\sum_{j=1}^{5} c_{2,6-j}\, l_{x-1+j}\right] / l_{x+3} \qquad \text{(by (15))} \\ &= \frac{l_x-6l_{x+1}+18l_{x+2}-10l_{x+3}-3l_{x+4}}{12l_{x+3}}, \end{aligned}$$

from Table 2.

Tables 1, 2, 3, or 4 may be used in various circumstances to obtain estimates for the force of mortality based on l-values at 3, 5, 7, or 9 consecutive ages.

3. FORMULA OF TYPE B

An alternative to the type A formula of §2 is provided by assuming that, over an appropriate age-interval, *the force of mortality itself is a polynomial function of age.* We call any formula for the force of mortality based on this alternative hypothesis a formula of " type B ".

The definition

$$\mu_x = -\frac{d}{dx}\log l_x$$

leads immediately to the well-known equation

$$\int_a^b \mu_y dy = \log l_a - \log l_b. \tag{18}$$

Equation (18) provides the starting-point for the derivation of formulae of type B.

Suppose, as before, that

$$x_1 < x_2 < \ldots < x_n$$

and that the value of l_x is known for $x = x_1, x_2, \ldots, x_n$. Making the further assumption that over the age interval $[x_1, x_n]$ μ_y is a polynomial in y of degree $(n-2)$, we wish to express μ_{x_i} (for $i = 1, 2, \ldots, n$) in terms of the n known values of l_x.

Let

$$p(x) = \int_{x_1}^{x} \mu_y dy, \tag{19}$$

from which it follows, if μ is continuous, that

$$p'(x) = \mu_x\,. \tag{20}$$

Note also that equations (18) and (19) imply that, for $j = 1, 2, \ldots, n$,

$$p(x_j) = \log l_{x_1} - \log l_{x_j}\,. \tag{21}$$

Since, by hypothesis, μ_y is a polynomial in y of degree $(n-2)$, it follows directly from equation (19) that $p(x)$ is a polynomial in x of degree $(n-1)$. We are therefore in a position to apply the results of §2. Hence, if c_{ij} is defined by equation (11), we have (for $i = 1, 2, \ldots, n$)

$$\begin{aligned} \mu_{x_i} &= p'(x_i) && \text{(by equation (20))} \\ &= \sum_{j=1}^{n} c_{ij} p(x_j) && \text{(by equation (10))} \\ &= \sum_{j=1}^{n} c_{ij}(\log l_{x_1} - \log l_{x_j}) && \text{(by equation (21))} \\ &= \left(\sum_{j=1}^{n} c_{ij}\right)\log l_{x_1} - \sum_{j=1}^{n} c_{ij}\log l_{x_j}\,. \end{aligned}$$

Equation (12) implies that the first term on the right-hand side of this last equation is zero, so that finally we obtain

$$\mu_{x_i} = -\sum_{j=1}^{n} c_{ij}\log l_{x_j} \quad (i = 1, 2, \ldots, n)\,. \tag{22}$$

The fact that the same coefficients $\{c_{ij}\}$ occur in both equations (13) and (22) should be noted. The former equation holds if l is a polynomial function of age of degree $(n-1)$ while the latter equation is valid if μ is a polynomial of degree $(n-2)$.

As before, the particular case when $x_i = \alpha + i$ $(i = 1, 2, \ldots, n)$ is of special interest and we devote the remainder of this section to this situation. Our last equation then becomes

$$\mu_{\alpha+i} = -\sum_{j=1}^{n} c_{ij}\log l_{\alpha+j}\,, \tag{23}$$

the coefficients $\{c_{ij}\}$ now being given by equation (11) with $x_r = r$ $(r = 1, 2, \ldots, n)$.

We conclude this section with some examples.

First suppose that $n = 5$. Letting $\alpha = x-3$ and $i = 3$ in equation (23), we obtain (from Table 2)

$$\mu_x = \frac{-\log l_{x-2}+8\log l_{x-1}-8\log l_{x+1}+\log l_{x+2}}{12}.$$

This is the type B formula corresponding to the type A formula given immediately following Table 4.

When $n = 3$, $\alpha = x-1$, and $i = 1$ equation (23) gives (from Table 1)

$$\mu_x = \frac{3\log l_x-4\log l_{x+1}+\log l_{x+2}}{2}. \tag{24}$$

The " select " equivalent of this equation was used to estimate the force of mortality at duration 0 for the FA1975-78 mortality tables. Note the equation (24) may be written as

$$\mu_x = \frac{-3(\log l_{x+1}-\log l_x)+(\log l_{x+2}-\log l_{x+1})}{2}$$

or

$$\mu_x = -\tfrac{3}{2}\log p_x+\tfrac{1}{2}\log p_{x+1}\,.$$

Formula (3) is simply the select form of this last equation (for a mortality table with a one-year select period).

Finally consider equation (23) with $n = 9$, $\alpha = x-5$, and $i = 5$. Combined with Table 4, this gives

$$\begin{aligned}\mu_x = \tfrac{1}{840}\{&-3\log l_{x-4}+32\log l_{x-3}-168\log l_{x-2}+672\log l_{x-1}\\ &-672\log l_{x+1}+168\log l_{x+2}-32\log l_{x+3}+3\log l_{x+4}\},\end{aligned} \tag{25}$$

which is valid if μ is a polynomial of degree 7 on the age interval $[x-4, x+4]$.

A trivial re-expression of equation (25) is

$$\begin{aligned}\mu_x = \tfrac{1}{840}\{&3(\log l_{x-3}-\log l_{x-4})-29(\log l_{x-2}-\log l_{x-3})\\ &+139(\log l_{x-1}-\log l_{x-2})-533(\log l_x-\log l_{x-1})\\ &-533(\log l_{x+1}-\log l_x)+139(\log l_{x+2}-\log l_{x+1})\\ &-29(\log l_{x+3}-\log l_{x+2})+3(\log l_{x+4}-\log l_{x+3})\}\end{aligned}$$

or

$$\begin{aligned}\mu_x = \tfrac{1}{840}\{&3\log p_{x-4}-29\log p_{x-3}+139\log p_{x-2}-533\log p_{x-1}\\ &-533\log p_x+139\log p_{x+1}-29\log p_{x+2}+3\log p_{x+3}\}.\end{aligned}$$

The strong-willed reader may verify that formula (2) is a restatement of this last equation. Thus formula (2) is equivalent to equation (25) above.

4. PRACTICAL APPLICATION OF THE ABOVE RESULTS

In relation to the recently published FA1975-78 mortality tables for female assured lives, preliminary experiments were carried out to determine the values to be published for the force of mortality. Estimates, based on formulae of both type A and type B, were obtained at each single age.

For $n = 3, 5, 7$, and 9 the value of μ_x for $x \geqslant (n-1)/2$ was estimated in terms of $l_{x-(n-1)/2}, \ldots, l_{x+(n-1)/2}$ by the " central " forms of equations (14) and (23) with $\alpha = x-(n+1)/2$ and $i = (n+1)/2$—the coefficients $\{c_{ij}\}$ being given in Tables 1—4. For $0 \leqslant x < (n-1)/2$ the value of μ_x was estimated in terms of $l_0, l_1, \ldots, l_{n-1}$ by the same equations with $\alpha = -1$ and $i = x+1$.

Specimen values of the resulting estimates of μ_x for $x = 20, 40, 60, 80$ and 100 are given in our next table.

TABLE 5

FA1975-78: Alternative estimates for the force of mortality

Values of $10^6.\mu_x$

Age	Type A estimates				Type B estimates			
x	$(n = 3)$	$(n = 5)$	$(n = 7)$	$(n = 9)$	$(n = 3)$	$(n = 5)$	$(n = 7)$	$(n = 9)$
20	268	268	268	268	268	268	268	268
40	1,026	1,025	1,025	1,025	1,026	1,025	1,025	1,025
60	6,737	6,729	6,729	6,729	6,740	6,729	6,729	6,729
80	50,750	50,757	50,757	50,757	50,868	50,757	50,757	50,757
100	622,126	608,216	607,840	607,819	609,554	607,818	607,817	607,817

It is immediately clear from the above Table that for practical purposes there is little difference between the two types of estimate—at least at all but the most advanced ages. No clear choice of which particular estimate to adopt is indicated. However, scrutiny of the alternative values at very advanced ages provides further insight into the relative merits of the different estimates—at least from a theoretical viewpoint. As an extreme (and somewhat artificial) illustration, consider for example, the alternative estimates for μ_{109} as indicated (to 3 decimal places) by the following table.

TABLE 6

FA1975-78: Alternative estimates for μ_{109}

Type A estimate				Type B estimate			
$(n = 3)$	$(n = 5)$	$(n = 7)$	$(n = 9)$	$(n = 3)$	$(n = 5)$	$(n = 7)$	$(n = 9)$
2·826	1·450	1·941	1·858	1·857	1·855	1·855	1·855

As n increases, the unstable nature of the type A estimates should be noted. The estimated value for the force of mortality with $n = 3$ is almost double the estimated value with $n = 5$, while the estimated value with $n = 7$ exceeds this latter value by 34%. In contrast the sequence of values estimated by type B formulae is much more stable.

From a theoretical viewpoint there may therefore be reasons for preferring formulae of type B. For practical purposes either type of formulae seems satisfactory. There is clearly no merit in using an excessively large value of n, if virtually the same estimates can be obtained with smaller values.

For the FA1975-78 Tables the force of mortality was calculated by type B formulae, with $n = 5$ for the ultimate values μ_x and $n = 3$ for the select values $\mu_{[x]}$ and $\mu_{[x]+1}$.

5. SUMMARY

If $x_1 < x_2 < \ldots < x_n$ and the values of $l_{x_1}, l_{x_2}, \ldots, l_{x_n}$ are given, we may wish to estimate the value of the force of mortality at a particular age x_i. Two alternative methods of estimation, each used in standard tables, are discussed in some detail.

The most general estimates for each method are given by equations (13) and (22) above, the coefficients $\{c_{ij}\}$ in these equations being defined by equation (11).

In the particular case of consecutive ages, the relevant estimates are given by equations (14) and (23). In this case the values of $\{c_{ij}\}$ for $n = 3$, 5, 7, and 9 are contained in Tables 1—4 above.

Acknowledgment

I am grateful to Dr J. C. Eilbeck for an interesting discussion on the topic of this note.