S. Mitra

On the Efficiency of the Estimates of Life Table Functions

1. Introduction

Expected values and variances of several life table functions were obtained earlier for both the generational and the cross-sectional life tables (Wilson, 1938; Chiang, 1960; Keyfitz, 1968; Mitra, 1972) The former were based on probability estimates of proportions surviving to a given age from an initial sample cohort, while the latter were obtained from estimates of age-specific mortality rates from sample populations in different age groups. Obviously, the latter method is more realistic and can be put into use in cases where life table functions cannot be derived for the total population.

An alternative set of results for the expected values and variances of these functions is presented in this paper which is based on the transformation of age-specific mortality rates to probabilities of dying by means of Reed-MerrePs approximation. These results, judging from their algebraic expressions, appear to be simpler and operationally more convenient, when compared with those obtained earlier. The expressions also permit the development of a procedure that will enable the researcher, constrained usually by his limited resources, to maximize the efficiency of his sampling technique. Needless to say, these results will be particularly useful for countries where national vital statistics are not adequate for construction of life tables, and also for countries where such tables are usually obtained by comparing age distributions of successive censuses at intervals often years.

2. Expected values and variances

If number of deaths ${}_nD_x$ in the age-group x to $x+n$ is observed over a period of one year and compared against the corresponding sample population ${}_nP_x$, then the estimate of age-specific mortality rate

$$ {}_nm_x = \frac{{}_nD_x}{{}_nP_x} \qquad (1) $$

will be distributed approximately as a binomial variable with a variance

$$ V({}_nm_x) = \frac{{}_nm_x(1-{}_nm_x)}{{}_nP_x} \qquad (2) $$

If the probability of dying in the same age interval *i.e.,* ${}_nq_x$ is writtten as

$$1 - {}_nq_p = e^{-n\,{}_nm_x - an^3\,{}_nm_x^2 - } \tag{3}$$

following Reed and Merrell, so that the difference equation after taking the logarithm of both sides of (3) becomes

$$\triangle({}_nq_x) = (1 - {}_nq_x)\,[n\triangle({}_nm_x) + \dots\dots] \tag{4}$$

then

$$V({}_nq_x) = n^2\,(1 - {}_nq_x)^2\,V({}_nm_x) \tag{5}$$

when all terms after the first are ignored. Since ${}_nm_x$ values for the different age groups can be independently obtained, the expected value of the life table survivors l_x at age $\times$ can be written as

$$E(l_x) = E\left[\,l_0 \prod_{t=0}^{(x-n)/n} (1 - {}_nq_{+n}\right]$$

$$= l_0 \prod_{t=0}^{(x-n)/n} [1 - E({}_nq_{+n})] \tag{6}$$

The difference equation after dropping the E and taking logarithms of both sides of (6) can be written as

$$\frac{\triangle l_x}{l_x} = -\sum_{t=0}^{(x-n)/n} \frac{\triangle({}_nq_{+n})}{1 - {}_nq_{+n}} \tag{7}$$

so that

$$\frac{V(l_x)}{l_x^2} = \sum_{t=0}^{(x-n)/n} \frac{V({}_nq_{tn})}{(1 - {}_nq_{tn})^2} \tag{8}$$

$$= n^2 \sum_{t=0}^{(x-n)/n} V({}_nm_{tn}) \tag{9}$$

because of (5). For $y > x$, l_y will depend on l_x and estimates of the covariance and correlation coefficient is given by,

$$\text{Cov}(l_x, l_y) = V(l_x)\,E(l_y)/E(l_x) \tag{10}$$

and

$$r_{l_x l_y} = \sqrt{\sum_{n=0}^{(x-n)/n} V({}_nm_{tn}) \Big/ \sum_{n-0}^{(y-n)/n} V({}_nm_{tn})} \tag{11}$$

Among many other results (Mitra, 1972), the following appear to be quite interesting. Using the standard approximation

$${}_nL_x = \frac{n}{2}\,(l_x + l_{x+n})$$

for the estimation of the stationary population T_x at and above age x, the covariance between l_x and T_x, namely,

$$\text{Cov}(l_x, T_x) = \text{Cov}\left(l_x, \frac{n}{2} l_x + n \sum_{t=1}^{(\alpha-x)/n} l_{x+tn}\right)$$

where α is the upper age limit. The above can be written as

$$\frac{n}{2} V(l_x) + \frac{nV(l_x)}{E(l_x)} \sum_{t=1}^{(\alpha-x)/n} E(l_{x+tn})$$

because of (10). Simplifying further,

$$\mathrm{Cov}\,(I_x, T_x) = V\,(I_x)\,\frac{E\,(T_x)}{E\,(I_x)}$$

$$= V\,(I_x)\,\mathring{e}_x \qquad (12)$$

approximately, substituting observed for the expected values. Similarly,

$$V\,(T_x + \frac{n}{2}\,I_x) = 2n \sum_{t=0}^{(\alpha-x)/n} V\,(I_{x+tn})\,\mathring{e}_x + tn \qquad (13)$$

so that

$$V\,(T_x) = 2n \sum_{t=0}^{(\alpha-v)/n} V\,(I_{x+tn})\,\mathring{e}_{x+tn} - nV\,(I_x)\,(\frac{n}{4} + \mathring{e}_x) \qquad (14)$$

In the difference equation

$$\triangle T_x = \mathring{e}_x \triangle I_x + I_x \triangle \mathring{e}_x \qquad (13)$$

I_x and $\mathring{e}_x$ are independent and therefore squaring both sides and taking expected values yields

$$V\,(\mathring{e}_x) = \frac{1}{I_x^2}\left[V\,(T_x) - \mathring{e}_x^2\;V\,(I_x)\right]$$

The expression $2n \sum_{t=0}^{(\alpha-x)/n} V\,(I_{x+tn})\,\mathring{e}_{x+tn}$ in (14) can be written as

$2n \sum_{t=0}^{(\alpha-x)/n} n^2\,I_{x+n}^2\,\mathring{e}_{x+tn} \sum_{j=0}^{(x+tn-n)/n} V\,({}_nm_{nj})$ from (9) after rearranging terms as

$$2n \left[n^2 \sum_{j=0}^{(x-n)/n} V\,({}_nm_{nj}) \sum_{t=0}^{(\alpha-x)/n} I_{x+tn}^2 \mathring{e}_{x+tn} \right.$$

$$\left. + n^2 \sum_{j=x/n}^{(\alpha-n)/n} V\,({}_nm_{nj}) \sum_{t=j}^{(\alpha-n)/n} I_{tn}^2\,\mathring{e}_{tn} \right] \qquad (17)$$

Observe that the integral approximation of

$2n \sum_{t=0}^{(\alpha-x)/n} I^2_{x+tn}\mathring{e}_{x+tn}$ Is $2\int_o^\alpha I^2_{x+t}\mathring{e}_{x+t}dt$ where the latter can be

written as

$$-\,2\int_o^\alpha T_{x+t}\,d(T_{x+t}) = T_x^2 \qquad (18)$$

Substituting (18) in (17), the expression simplifies into

$$n^2 \left[T_x^2 \sum_{j=0}^{(x-n)/n} V\,({}_nm_{nj}) + \sum_{j=x/n}^{(\alpha-n)/n} V\,({}_nm_{nj})\,T_{nj}^2 \right] \qquad (19)$$

$$= \mathring{e}_x^2\,V\,(I_x) + n^2 \sum_{j=x/n}^{(\alpha-n)/n} V\,({}_nm_{nj})\,T_{nj}^2 \qquad (20)$$

because of (9). Substitution of (20) into (14) and then into (16) results in

$$V\,(\mathring{e}_x) = \frac{1}{I_x^2}\left[n^2 \sum_{j=o}^{(\alpha-n)/n} V\,({}_nm_{nj})\,T_{nj}^2 \right] - n^3\,(\frac{n}{4} + \mathring{e}_x) \sum_{t=o}^{(x-n)/n} V\,({}_nm_{+n}) \qquad (21)$$

For $x = o$ and $I_o = I$, (21) reduces to

$$V\,(\mathring{e}_o) = n^2 \sum_{j=0}^{(\alpha-n)/n} V\,({}_nm_{nj})\,T_{nj}^2 \qquad (22)$$

$$n^2 \sum_{j=0}^{(\alpha-n)/n} T_{nj}^2 \frac{{}_nm_{nj}\,(1-{}_nm_{nj})}{{}_nP_{nj}} \tag{23}$$

substituting for $V({}_nm_{nj})$ from (2).

3. Minimizing $V(\mathring{e}_0)$

It is quite natural for the variance of $\mathring{e}_0$ to depend on all the mortality rates as well as on the sizes of the samples on which these rates are based. If the resources are limited in the sense that the total sample size, *i.e.*,

$$\sum_{j=0}^{(\alpha-n)/n} {}_nP_{nj} = P \tag{24}$$

is to be held constant, then the efficiency of the estimate of $\mathring{e}_0$ can be maximized by decomposing P into the respective age-groups in a manner that will minimize the variance of $\mathring{e}_0$. The choice of $\mathring{e}_0$ over all other life table functions is based on the consideration that the expectation of life at birth is the most concise summary of the entire life table as it utilizes all the data necessary for the construction of such a table.

The problem then reduces to minimizing (23) subject to the restriction specified in (24) which is equivalent to minimizing

$$n^2 \sum_{j=0}^{(\alpha-n)/n} T^2_{nj} \frac{{}_nm_{nj}\,(1-{}_nm_{nj})}{{}_nP^2_{nj}} + \lambda \left(\sum_{j=0}^{(\alpha-n)/n} {}_nP_{nj} - P \right) \tag{25}$$

where λ is the undetermined multiplier. Differentiating (25) with respect to ${}_nP_{nj}$, and equating the result to zero produces the equation

$$n^2 T^2_{nj} \frac{{}_nm_{nj}\,(1-{}_nm_{nj})}{{}_nP^2_{nj}} = \lambda \tag{26}$$

or

$$\sqrt{\lambda}\;{}_nP_{nj} = nT_{nj} \sqrt{{}_nm_{nj}\,(1-{}_nm_{nj})} \tag{27}$$

so that λ can be obtained by summing (27) over j, which because of (24) reduces to

$$\sqrt{\lambda} = \frac{n}{P} \sum_{j=0}^{(\alpha-n)/n} T_{nj} \sqrt{{}_nm_{nj}\,(1-{}_nm_{xj})} \tag{28}$$

Eliminating λ from (27) gives

$${}_nP_{nj} = P \frac{T_{nj} \sqrt{{}_nm_{nj}\,(1-{}_nm_{nj})}}{\sum_{j=0}^{(\alpha-n)/n} T_{nj} \sqrt{{}_nm_{nj}\,(1-{}_nm_{nj})}} \tag{29}$$

In short, the sample size corresponding to an age-group should be made proportional to the population at and above that age group in the corresponding stationary population and also to the square roots of the corresponding age-specific mortality and survivorship rates.

The minimum value of $V(\mathring{e}_0)$ can now be rewritten as

$$\text{Min } V(\mathring{e}_0) = \frac{n^2}{P} \left[\sum_{j=0}^{(\alpha-n)/n} T_{nj} \sqrt{{}_nm_{nj}\,(1-{}_nm_{nj})} \right]^2 \tag{30}$$

4. Discussion of results

It is easy to see that the distribution of ${}_nP_{ni}$, among age groups will minimize V ($e°_0$) will not simultaneously minimize the variances of other life expectancies or other life table functions. The reader must also have noticed that the sampling technique resulting in the mini-mum variance of e^0_0, (in fact, all of the expressions derived earlier) depends on the knowledge of the functions like age-specific mortality rates that are yet to be estimated The results, however, are not expected, as optimum allocation of samples into subgroups. Accordingly, apriori information, such as the values of these functions as observed earlier or those from another but reasonably similar life tables will be required for such purposes.

A demonstration of the results is shown in Table 1 in which variances have first been obtained by allocating samples of sizes 500 in each five year age-group. Next, the total sample has been redistributed to minimize the variance of e^0_0 which shows a considerable improvement in the efficiency of the estimates. Comparison of the variances of life expectancies at other ages shows increased efficiency at earlier ages and reduction elsewhere. This is quite expected because the samples at later ages continue to decline in size very rapidly. For reasons of operational simplicity, however, sample sizes were kept at a minimum of 10, although, theoretically the sample sizes reduced to a fraction at extreme ages.

Table 1: Variances of life expectancies by age and sample size

Age (1)	Ix (2)	o ex (3)	* $V_1(e^0x)$ (4)	nx(5)	$V_2(e^0x)$
0	100,000	43.06	6.97	1,106	2.48
1	81,162	52.00	5.88	1.546	3.34
5	72,813	53.83	3.97	680	3.42
10	71,584	49.72	3.64	600	3.20
15	70,420	45.49	3.39	779	3.01
20	68,090	41.95	2.97	797	2.80
25	65,127	38.75	2.49	720	2.58
33	62,144	35.49	2.07	635	2 37
35	59,210	32.13	1.70	564	2.16
40	56,2 12	28.71	1.38	498	1.95
45	53,088	25.25	1.11	433	1.75
50	49,785	21.75	.89	388	1.56
55	45,893	18.38	.70	345	1.39
60	41,075	15.23	.53	291	1.24
65	35,251	12.32	.39	239	1.11
70	28.004	9.85	.27	174	1.05
75	19,996	7.78	.19	111	1.07
80	12,145	6.23	.12	57	1.22
85	6,230	4.90	.08	27	1.38
90	2,401	4.14	.06	10	1.12

* Variance based on sample of size 100 for single year of age or 500 for 5 year age group. For 90 and above the sample size is approximately 1,000 (a hard to find sample) which shows that samples of equal sizes are not only unnecessary but also operationally unsound.

Summary

Application of sampling techniques for estimating life table functions has received little attention by population scientists. A few researchers, namely, Wilson (1938) and Koop and more recently Chiang (1960) and Keyfitz have investigated the sampling distributions of a number of biometric functions. It is somewhat surprising to note that the agencies in charge of compilation of life tables await the availability of a complete vital registration statistics or of census distributions, when estimates of these tables can be derived at a reasonably low cost by the use of statistical sampling.

In this paper, estimates of the life table functions and their variances have been obtained by utilizing sample estimates of age-specific mortality rates and their subsequent conversion to probabilities of dying through Reed-MerrelPs formula. The results are different (from those obtained earlier) because of the particular transformation and are also simple in algebraic forms as well as from operational point of view, particularly so in ;the case of life expectancies. Investigation has also been carried out to determine the effects of differential allocation of fixed resources (in age-groups for estimating the mortal ity-rates) on the variances of life expectancies. In particular, the technique for the allocation of samples by age-groups has been outlined which, for a fixed siza of the total sample, minimizes the variance of the expectation of life at birth. .

References ;

Chiang, C. L. "Stochastic study of the life table and its applications: I. Probability distribution biometric functions," *Biometrics,* Vol. 16 (1960), pp. 618-635.

Chiang. C. L. "Stochastic study of the life table aud its application II. Sample variance of the observed expectation of life and other biometric functions", *Human Biology.* Vol. 32 (1960).

Irwin, J. O., "The standard error of an estimate of exceptional life" Journal of Hygiene, Vol. 47 (1949) pp. 188-189.

Keyfitz, N , *Introduction to the Mathematics of Population,* Addison-Wesley Publishing Company, Massachusetts, Ch. 15 1968.

Koop, J. C., "Notes on the estimation of gross and net reproduction rates by statistical sampling", Biometrics: Vol. 7(1951), pp. 155-166.

Mitra. S., "On Some Characteristics of the Distributions of Life Table Functions" presented at the 1972 meetings of the Population Association of America, Toronto.

Nilson, E. B. "The standard deviation of sampling for life expectancy", *Journal of the American Statistical Association,* Vol. 33 (1938), pp. 705-708.