Arbeidsnotater

STATISTISK SENTRALBYRÅ

Dronningensgt. 16, Oslo-Dep., Oslo 1. Tlf. 41 38 20, 41 36 60

WORKING PAPERS FROM THE CENTRAL BUREAU OF STATISTICS OF NORWAY

IO 71/1 26 January 1971

ON THE INTERPRETATION OF CERTAIN VITAL RATES AS AVERAGES OF UNDERLYING FORCES OF TRANSITION

By

Jan M. Hoem[x)]

Contents

[x)] The present paper was written while the author visited the Department of Statistics, University of California, Berkeley, on a Ford Foundation award.

Summary

In the present paper, formulas are given which show how age-specific mortality rates and fertility rates appear as weighted averages of underlying forces of mortality and fertility. This is done separately for period rates and cohort rates, which turn out to have radically different properties. More general rates are also introduced and investigated. In a final chapter, it is shown how the theory could be extended to a simplistic description of school attendance.

A certain type of formula which has appeared in the demographic literature, turns out to have less general applicability than apparently believed before. It is valid in generalized stable populations, and as an approximation in connection with sectionally (generalized) stable populations, but not elsewhere.

1. Introduction

A. Consider a model population exposed to a mortality schedule given by a force of mortality function $\mu(\cdot)$. Let us define the period male death rate in the model for the ages x to x+n, say for a given year, as the number of male deaths at such ages divided by the aggregate number of person-years lived in the age segment during the year, both calculated according to the model. Such a rate measures the mean mortality risk over the given age interval, and it can be expressed as a weighted average of the instantaneous death rates $\mu(x+t)$ for $0 \leq t < n$ in the form

$$(1) \qquad \int_0^n w(x+t)\mu(x+t)dt \Big/ \int_0^n w(x+t)dt \,.$$

If we take the value $w(x+t)$ of the weight function at age x+t to represent the number of males in the population that will ever reach age x+t during the study period, then (1) is immediately seen to hold for any model population, whether open or closed.

The purpose of the present paper is to study formulas like (1), partly with different interpretations of $w(\cdot)$.

B. The idea that the death rate for an extended age interval can be regarded as a weighted average of corresponding rates for shorter intervals is quite familiar in demography, but apparently explicit formulas similar to (1) have only been published quite recently (except for very special cases).

Keyfitz [(1970), compare also (1968a), page 173, (7.1.7)] gives the formula (for n=5)

$$(2) \qquad \int_0^5 p(x+t)\mu(x+t)dt \;/\; \int_0^5 p(x+t)dt,$$

which he takes as valid if we "suppose a continuous function p(x+t) underlying the observed age distribution within the age group x to x+5. This means that the number of individuals in the exposed population between ages x+t and x+t+dt is designated p(x+t)dt."

Let us introduce

$$\overline{p}(x) = p(x) \;/\; \int_0^\omega p(u)\, du.$$

We can replace p by $\overline{p}$ in (2) without changig the value of the ratio. Let us imagine that this has been done. The function $\overline{p}$ represents the age distribution of the population at a particular moment. This distribution will change over time, except in stable populations. For a meaningful interpretation of formulas like (2), it is therefore necessary to specify, either

(i) at what moment $\overline{p}$ represents the population age distribution, or

(ii) that the population considered is taken to be stable.

We shall see how important it is not to omit such a specification. In the present paper, we take $\overline{p}$ to represent the age distribution at the beginning of the study period, which we call "time zero". (Compare Keyfitz (1968a), page 97.)

C. Through private conversations, I have found that many people consider the representation of the mortality rate by (2) as being self-evident. Unfortunately, it is not so. Quite contrary to this belif, not only is it not self-evident, but it is not even correct except in particular circumstances. Fortunately, it *is* correct in stable populations, and many papers on mathematical demography, like Keyfitz's from 1970, really confine themselves to such populations even though it is not always explicitly stated.

In fact, we shall show that in a closed population (2) is correct as a general formula, i.e., for all x simultaneously,
only when the population is of a generalized stable type, viz. when p has the form

$$(3a) \qquad p(u) = \sum_{k=0}^{\infty} A_k e^{-r_k u}\, \ell(u),$$

and at the same time

$$(3b) \qquad b(t) = \sum_{k=0}^{\infty} A_k\, e^{r_k t},$$

where b(t)dt represents the number of live babies (of the sex considered) born into the population between time t and t+dt. The constants $A_0, A_1, \ldots$ are the same in (3a) as in (3b), and so are $r_0, r_1, \ldots$. $\ell(x)$ is, of course, the survival function. We take $\ell(0)=1$.

D. We shall prove the above claims in section 2. In section 3 we study similar questions for cohort mortality, which turns out to be much simpler. We also introduce a general mortality rate and give corresponding formulas. In section 4 we show how this reasoning can be extended to cover age-specific fertility rates. The same theory can easily be formulated for the multiple decrement situation also, but we shall not do so.

The structure of the cases just mentioned is really very simple, and we get nice formulas. The same kind of theory can be brought to bear on more complicated situations, but then the complexity of the formulas increases rapidly. As an illustration, we use a simplistic model for school attendance, where the mathematics are still manageable (section 5).

Our account is written in the pseudo-probabilistic vein commonly employed in classical population mathematics. The possibility of immigration and emigration does not really throw any light on the questions we want to discuss here, but rather tends to detract attention from our main line of argument. In what follows, we shall therefore consider a closed population only.

In order to avoid burdering the account with mathematical niceties which really are beside the point, we shall assume that all functions appearing are continuous, and also state here once and for all that the set Ω appearing in sections 3B, 4C and 5C is taken to be Lebesgue measurable.

2. Period mortality rates

A. Let us designate our study period by $[0,T]$. We want to derive formulas for the mortality rate ${}_n\mathcal{M}_x$ for the age interval from x to x+n. Let ${}_n\mathcal{D}_x$ denote the number of deaths during $[0,T]$ with age at death between x and x+n, and let ${}_n\mathcal{L}_x$ denote the aggregate number of person-years lived (total lifetime) in the age interval $[x, x+n\rangle$ during the study period by people in the population. (We prefer to use script letters $\mathcal{M}$, $\mathcal{D}$, and $\mathcal{L}$ to avoid confusion with standard notation possibly having slightly different meaning.) Then ${}_n\mathcal{D}_x$ is the number of deaths in the region Ω_0 in the Lexis diagram in figure 1, ${}_n\mathcal{L}_x$ is the aggregate length of lifelines ever entering Ω_0, and we define

$${}_n\mathcal{M}_x = {}_n\mathcal{D}_x \,/\, {}_n\mathcal{L}_x .$$

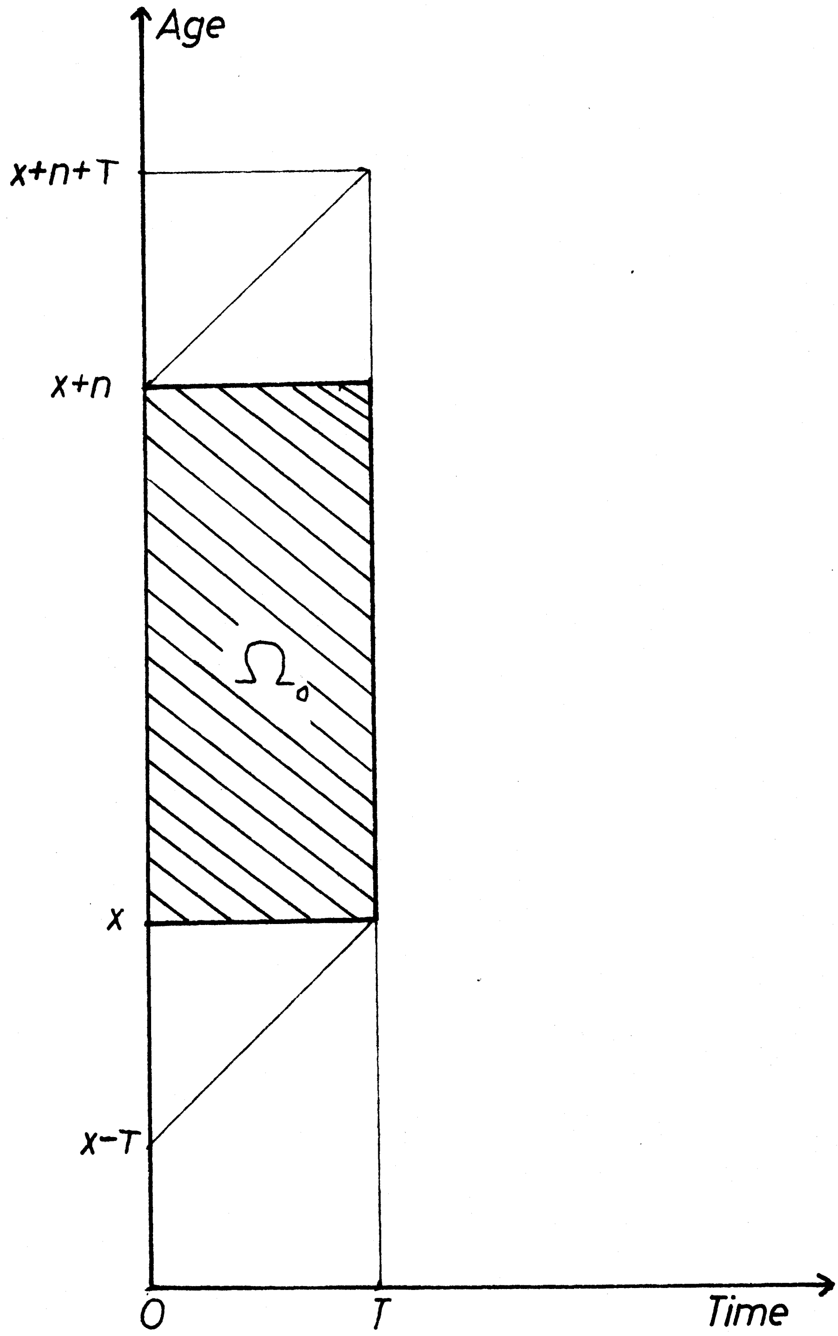

Figure 1

The Lexis diagram. A region Ω_0 of interest in studies of period mortality

Assume for the time being that $x \geq T$. This means that persons born into the population during the study period cannot contribute deaths or lifetime in Ω_0, because their lifelines never enter this region. Then

$$(4) \qquad {}_n\mathcal{D}_x = \int_{x-T}^{x+n} \int_{\max(0,x-u)}^{\min(T,x+n-u)} p(u)\,\frac{\ell(u+t)}{\ell(u)}\,\mu(u+t)\,dt\,du.$$

This formula is valid both for $T \leq n$ and for $T > n$.

The proof of (4) goes like this: Assume first that $T \leq n$, as in figure 1. The individuals who can contribute to ${}_n\mathcal{D}_x$, are those who at time zero have ages in the interval from $x-T$ to $x+n$. Let us split this interval into the three subintervals $[x-T,x\rangle$, $[x,x+n-T\rangle$, and $[x+n-T,x+n]$, and let us first consider the $p(u)du$ individuals at a particular age u, say for $x-T \leq u < x$. They will start contributing deaths from the moment their lifelines enter Ω_0, which is at time $x-u$. At time t, for $x-u \leq t < T$, $p(u)du\{\ell(u+t)/\ell(u)\}$ of the original $p(u)du$ individuals will still remain alive, and during the period $(t,t+dt)$ they will contribute $p(u)du\{\ell(u+t)/\ell(u)\}\mu(u+t)dt$ deaths. The entire contribution to ${}_n\mathcal{D}_x$ from people staring in the age group from $x-T$ to x at time zero is, therfore,

$$\int_{x-T}^{x} \int_{x-u}^{T} p(u)\,\frac{\ell(u+t)}{\ell(u)}\,\mu(u+t)\,dt\,du.$$

Similar arguments for those who have ages between x and $x+n-T$ at time zero, and separately for those who then are in the age bracket from $x+n-T$ to $x+n$, give (4) for $T \leq n$.

In the case where $T > n$, (4) is seen to hold by considering each of the age groups $[x-T, x+n-T\rangle$, $[x+n-T,x\rangle$, and $[x,x+n]$, separately. This establishes (4). $\square$

Introducing $y=u+t$ for t and changing the order of integration, we get

$$(5a) \qquad {}_n\mathcal{D}_x = \int_x^{x+n} \ell(y)\mu(y) \int_{y-T}^{y} \frac{p(u)}{\ell(u)}\,du\,dy \quad \text{for } x \geq T.$$

By a similar argument, we get

$$(5b) \qquad {}_n\mathcal{L}_x = \int_x^{x+n} \ell(y) \int_{y-T}^{y} \frac{p(u)}{\ell(u)}\,du\,dy \quad \text{for } x \geq T.$$

If we let

$$w(y) = \int_{y-T}^{y} p(u)\,\frac{\ell(y)}{\ell(u)}\,du \quad \text{for } y \geq T,$$

then $w(\cdot)$ has the verbal interpretation given at the end of section 1A, and we see that ${}_n\mathcal{M}_x$ equals the expression in (1). It does not generally reduce to (2), so <u>the latter is not a correct</u> general <u>representation of</u> ${}_n\mathcal{M}_x$.

B. If $x<T$, individuals born during the period $[0,T-x]$ will contribute to ${}_n\mathcal{D}_x$ and ${}_n\mathcal{L}_x$. We let a function $b(\cdot)$ be defined verbally as underneath (3b),

without necessarily assuming that (3b) holds. If we introduce

$$(6) \qquad \Lambda(u) = \begin{cases} p(u)/\ell(u) & \text{for } u \geq 0, \\ b(-u) & \text{for } u < 0, \end{cases}$$

then one can establish the following two formulas by an argument quite similar to the one for (4):

$$(7a) \qquad {}_n\mathcal{D}_x = \int_x^{x+n} \ell(y)\mu(y) \int_{y-T}^{y} \Lambda(u)du\, dy,$$

and

$$(7b) \qquad {}_nL_x = \int_x^{x+n} \ell(y) \int_{y-T}^{y} \Lambda(u)du\, dy.$$

Note that (7) reduces to (5) if $x \geq T$, so (7) is valid whether $x<T$ or $x \geq T$. ${}_nM_x$ is seen to equal (1) with $w(y) = \ell(y) \int_{y-T}^{y} \Lambda(u)du$ for all $y \geq 0$.

C. We have shown that (2) is not generally valid, and turn now to the case of a stable population, in which case we shall see that (2) is correct. Let

$$(8a) \qquad p(u) = Ae^{-ru}\ell(u),$$

$$(8b) \qquad b(u) = Ae^{ru},$$

where A is any positive constant. Then

$$(9) \qquad \Lambda(u) = Ae^{-ru},$$

and

$$(10) \qquad \int_{y-T}^{y} \Lambda(u)du = \alpha\, \Lambda(y),$$

with

$$(11) \qquad \alpha = \begin{cases} T & \text{if } r=0, \\ (e^{rT}-1)/r & \text{if } r \neq 0. \end{cases}$$

We also get

$$(12) \qquad {}_nM_x = \int_0^n e^{-rt}\ell(x+t)\mu(x+t)dt \Big/ \int_0^n e^{-rt}\ell(x+t)dt,$$

which means that (2) is valid when (8) holds.

In fact, (2) is correct under more general conditions than this. In proving (12), we did not use (8) itself. We rather used (10), which in effect tells us that the double integrals in (7) can be reduced to single integrals. We see that (2) is valid if and only if (10) holds for some $\alpha>0$, not necessarily the α given by (11).

It is a nice fact that we can solve (10) and find what form Λ must have to satisfy this relation. We differentiate (10) and get

$$\alpha \frac{d}{dy} \Lambda(y) - \Lambda(y) + \Lambda(y-T) = 0 \qquad \text{for } y \geq 0$$

This is a differential-difference equation studied in some detail by Bellman and Cooke (1963). It turns out that all solutions of (10) are of the form

(13) $\quad \Lambda(y) = \sum_{k=0}^{\infty} A_k e^{-r_k y}$,

where the A_k are arbitrary constants. $r_1, r_2, \ldots$ are all the non-real solutions of the characteristic equation

(14) $\quad \alpha r = e^{rT} - 1.$

(Compare (11). Note the following difference: In (11), r was given, and we derived α. In (14), α is given, and we solve for r.)

(14) has two real roots, ρ_1 and ρ_2, say. Here, $\rho_1 = 0$, while $\rho_2 \gtreqless 0$ as $\alpha \gtreqless T$. In (13), $r_0 = \rho_2$. Thus

$$r_0 \gtreqless 0 \quad \text{as} \quad \alpha \gtreqless T.$$

(13) is, of course, equivalent to (3).

D. We have investigated, among other things, the validity of (2) as a general formula, i.e., for all x simultaneously, and we found that (2) is valid if and only if (10) holds for all $y \geq 0$. If we are only interested in a particular value of x, it is of course sufficient that (10) holds for $x \leq y < x+5$. This will be the case, e.g., when (8) holds for $x-T \leq u < x+5$. Similarly, (12) will hold for particular x and n if (8) is valid for $x-T \leq u < x+n$. (Note that it is not enough that (8) holds for $x \leq u < x+n$.)

If (8) is approximately valid for $x-T \leq u < x+n$, then of course (12) is also approximately correct. On this basis, Keyfitz (1968 b, 1970) has developed a method of retrieving values of the $\ell(\cdot)$-function from a series of known values ${}_5M_0$, ${}_5M_5$, ${}_5M_{10}, \ldots$ for the mortality rates on the assumption that the population is (approximately) what he calls sectionally stable (Keyfitz, 1968 b, page 1260).

E. In many situations, a demographer does not have ${}_nL_x$ available, but must use some approximation, such as the midperiod population segment or the average between the initial and final population segments. This gives rise to mortality rates which are slightly different from ${}_nM_x$. The size of the relevant population segment at time t is

$${}_nP_x(t) = \int_{x-t}^{x+n-t} \Lambda(u)\ell(u+t)du.$$

In particular, ${}_nP_x(\frac{1}{2}T)$ is the size of the mid-period population segment. We define

$${}_nM_x^{*} = {}_nD_x / \{{}_nP_x(\tfrac{1}{2}T)\cdot T\}$$

and

$${}_nM_x^{\#} = {}_nD_x / \{\tfrac{1}{2}({}_nP_x(0) + {}_nP_x(T)).T\}.$$

If (8) holds, with $r=0$, the three rates coincide. If (8) holds with $r\neq 0$, we get

(15) $${}_nM_x^{*} = {}_nM_x\cdot(e^{\frac{1}{2}rT} - e^{-\frac{1}{2}rT})/rT$$

and

$${}_nM_x^{\#} = {}_nM_x\cdot\frac{2(e^{rT}-1)}{rT(e^{rT}+1)}\,.$$

The correction factor in (15) is approximately equal to

$$1+r^2T^2/24$$

as has been pointed out by Robert Retherford.

3. Cohort mortality rates. General mortality rates

A. Let $p(\cdot)$ and $b(\cdot)$ be defined as above. Cohort mortality functions will not coincide with the corresponding period mortality functions, so let us designate the former by a prime. Thus, for instance, the cohort survival function value at age x is $\ell'(x)$, and the corresponding force of mortality is $\mu'(x)$. (The prime must not be confused with the symbol for a derivative. We shall not use the prime in the latter meaning in this paper.)

We shall now give formulas for the mortality rate ${}_nM_x'$ for the age interval from x to x+n for a cohort born during the period from $T'-x$ to $T'+T''-x$. Compare figure 2. The number of deaths in Ω_0' is

$${}_n\mathcal{D}_x' = B_x\cdot\int_0^n \ell(x+t)\mu(x+t)\,dt,$$

where

$$B_x = \int_{x-T'-T''}^{x-T'} \Lambda(u)du.$$

The corresponding exposure time is

$${}_nL_x' = B_x\cdot\int_0^n \ell(x+t)dt.$$

Thus,

(16) $${}_nM_x' = {}_n\mathcal{D}_x'/{}_nL_x' = \int_0^n \ell(x+t)\mu(x+t)dt/\int_0^n \ell(x+t)dt.$$

Note that these formulas hold for all $n>0$, all $T''>0$, and all T', not only when the configuration is as in figure 2.

<u>Note also that ${}_nM_x'$ is entirely independent of</u> p <u>and</u> b, <u>in striking contrast to ${}_nM_x$. Thus (16) holds in a stable population also</u>. One should not insert the factor e^{-rt} in the integrals in the formula for ${}_nM_x'$ as we did in (12). (16) can also be written as

$${}_nM_x' = \{\ell(x) - \ell(x+n)\}/\int_0^n \ell(x+t)dt.$$

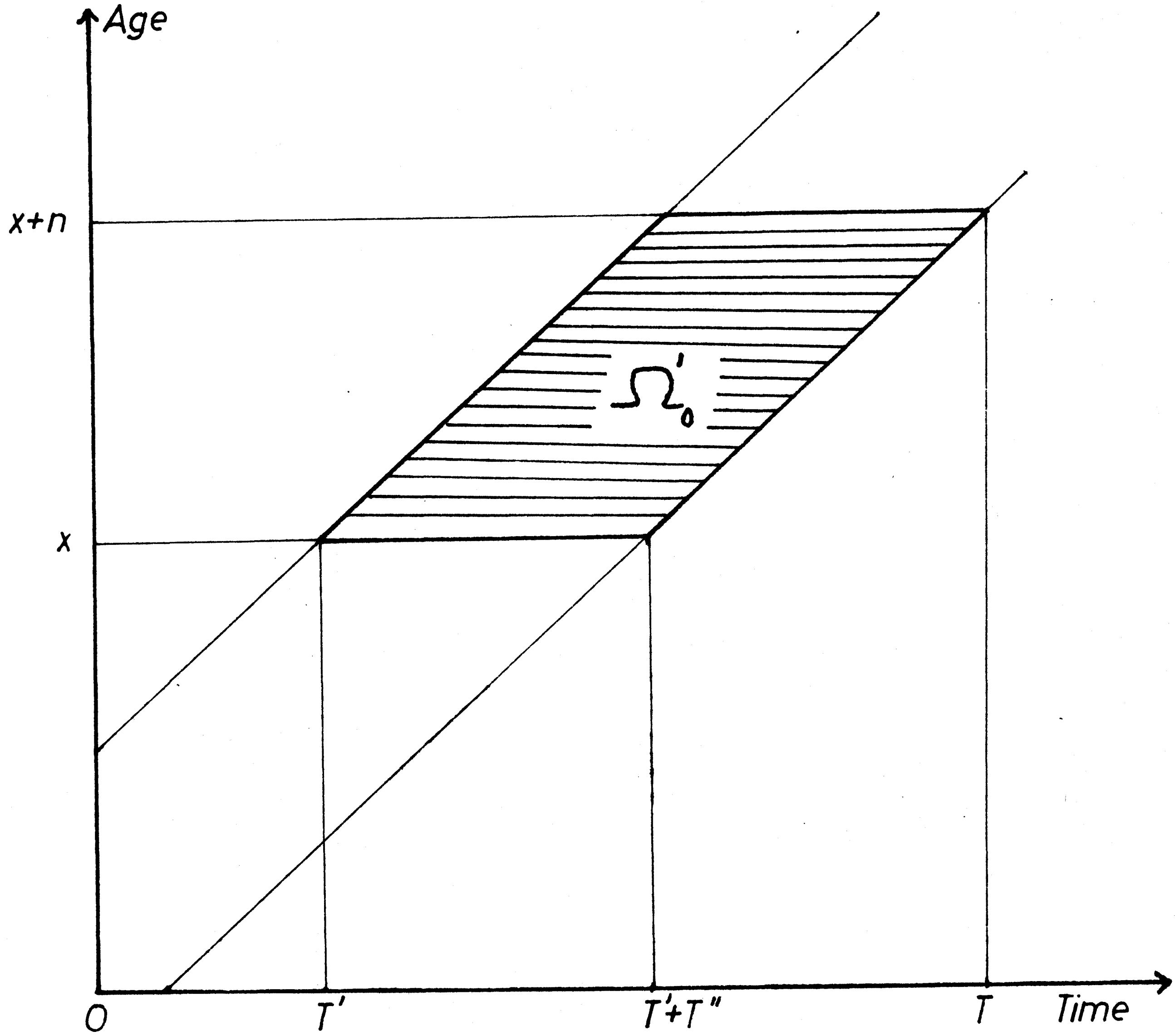

Figure 2

The Lexis diagram. A region Ω_0' of interest in studies of cohort mortality.

B. Ω_0 and Ω_0' in figures 1 and 2, respectively, are areas in the first quadrant of the plane, and we have defined mortality rates ${}_nM_x$ and ${}_nM_x'$ corresponding to these areas. We can go further and define a mortality rate $M(\Omega)$ corresponding to an arbitrary area Ω in the first quadrant.

We designate a point in the plane by (t,x), and let

$$T = \sup\{t : (\exists x)((t,x)\epsilon\Omega)\},$$

i.e. T is the latest moment for which observation in Ω is carried out. We assume that $0<T<\infty$.

The number of deaths in Ω is

$$\mathcal{D}(\Omega) = \int_{-T}^{\omega}\int_{0}^{T}\Lambda(x)\ell(x+t)\mu(x+t)I_\Omega(t,x+t)dt\ dx,$$

where

$$I_\Omega(t,y) = \begin{cases}1 & \text{if } (t,y)\epsilon\Omega,\\ 0 & \text{otherwise.}\end{cases}$$

The corresponding total lifetime is

$$L(\Omega) = \int_{-T}^{\omega}\int_{0}^{T}\Lambda(x)\ell(x+t)I_\Omega(t,x+t)dt\ dx,$$

and the mortality rate is

$$M(\Omega) = \mathcal{D}(\Omega)/L(\Omega). \tag{17}$$

(4), (5), (7), and (16) are particular cases of these formulas.

4. Fertility rates

A. Let us designate the force of fertility at age x by $\phi(x)$. Thus the probability that a woman at age x will have a birth within age $x+dx$ is $\phi(x)dx$. (For a more detailed discussion, see Hoem (1970). Keyfitz (1968a) and many others call this quantity $m(x)$, but we prefer our own notation both because we consistently use Greek letters for forces of transition, and because the letter m is used for so many quantities connected with mortality. Note that ϕ is not the net maternity function (Keyfitz, 1968a, page 140). The latter equals $\phi(x)\ell(x)$ at age x.) We shall now see how one can define fertility rates for the female population in analogy with the mortality rates we have studied above. The similarity with the mortality formulas is striking, and the arguments are essentially the same, so we can be very brief.

B. Period fertility rates. Let T be less than the age ω_0 at puberty, so that no one born during the study period can also bear children during this period. The number of births in the study period between ages x and x+n are then

$$_nF_x = \int_{x-T}^{x+n} \int_{\max(0,x-u)}^{\min(T,x+n-u)} p(u)\,\frac{\ell(u+t)}{\ell(u)}\,\phi(u+t)dt\,du$$

$$= \int_{x}^{x+n} \ell(y)\phi(y) \int_{y-T}^{y} \frac{p(u)}{\ell(u)}du\,dy,$$

in analogy with (4) and (5a). The period fertility rate is $_nF_x/_nL_x$. If, in particular, (10) holds, the rate is

$$\int_0^n p(x+t)\phi(x+t)dt/\int_0^n p(x+t)dt.$$

(Compare Keyfitz (1968a), page 173, (7.1.8) and the lines below it.)

C. Cohort and general fertility rates. In analogy with (16), we get the cohort fertility rate for the ages from x to x+n to be

$$\int_0^n \ell(x+t)\phi(x+t)dt/\int_0^n \ell(x+t)dt.$$

A general fertility rate, corresponding to the arbitrary area Ω, is defined as $F(\Omega)/L(\Omega)$, where

(18) $$F(\Omega) = \int_{-T}^{\omega} \int_0^T \Lambda(x)\ell(x+t)\phi(x+t)I_\Omega(t,x+t)dt\,dx$$

is the number of births in Ω.

5. A model for school attendance

A. Preliminaries. The situations giving rise to the mortality and fertility rates we have discussed above, are very simple, and we get nice formulas. As an example of a case which is more complex but still manageable, we shall briefly consider a model for school attendance, and shall give formulas for school entrance and drop-out rates.

As a description of real-life school attendance, the model suffers from the weakness of being rather too simple, in that a person's previous school history is not taken into account, except insofar as this is reflected in his current status as being in school out of school. We include it here, nonetheless, because

(i) it may have some pedagogical merit as a relatively simple illustration of how the theory of the previous chapters may be extended,

(ii) it does have some of the more important features that a school model should have, and

(iii) it is not more "unrealistic" than many of the other models commonly used as abstractions of population phenomena. In this respect, it is in fact just about on the same level as the type of very standard fertility model which we discussed in chapter 4. In the latter, an individual's previous fertility history is disregarded. It is easy to conceive of practical situations where the lack of availability of data on individual histories forces the investigator to work with models of this type.

B. Some basic concepts and notation. We start by introducing some basic forces of transition (ρ, ν, and μ), some transition probabilities (r, s, p, $\bar{r}$ and $\bar{s}$), and some survival type functions (ℓ_r and ℓ_s).

Let $\rho(x)dx$ be the probability that a member of the population who is not in school at age x, will enter school within age x+dx.

Let $\nu(x)dx$ be the probability that an individual who attends school at age x, will leave school within age x+dx.

Let $\mu(x)$ be the force of mortality at age x, both for individuals attending school and for those not doing so.

Furthermore, let ${}_tr_x$ be the probability that a person who is not in school at age x, will attend school at age x+t, and let ${}_ts_x$ be the probability that someone who attends school at age x, will also be in school at age x+t. Let

$${}_tp_x = \ell(x+t)/\ell(x),$$

$$\ell_s(x) = {}_xr_0, \text{ and } \ell_r(x) = \ell(x)-\ell_s(x).$$

Then $\ell_s(x)/\ell(x)$ is the probability that an x-year-old attends school, and $\ell_r(x)/\ell(x)$ is the probability that he does not do so.

Finally, let ${}_t\bar{r}_x$ and ${}_t\bar{s}_x$ be the values ${}_tr_x$ and ${}_ts_x$ would have if there were no mortality, i.e. if we replace μ by zero everywhere.

One may prove that the following formulas hold (Hoem, 1970, (5.4)).

(19a) $\quad {}_t r_x = {}_t \bar{r}_x \cdot {}_t p_x,$

(19b) $\quad {}_t s_x = {}_t \bar{s}_x \cdot {}_t p_x.$

The Kolmogorov forward differential equation (Feller, 1957, page 426) for $\bar{s}$ is

$$\frac{\partial}{\partial t}\, {}_t \bar{s}_x = -{}_t \bar{s}_x \{\nu(x+t)+\rho(x+t)\}+\rho(x+t),$$

which has the solution

(20a)
$$ {}_t \bar{s}_x = \exp\{-\int_0^t [\nu(x+\tau)+\rho(x+\tau)]d\tau\} + \int_0^t \rho(x+\tau) \exp\{-\int_\tau^t [\nu(x+\xi)+\rho(x+\xi)]d\xi\}d\tau .$$

Similarly,

(20b)
$$ {}_t \bar{r}_x = \int_0^t \rho(x+\tau) \exp\{-\int_\tau^t [\nu(x+\xi)+\rho(x+\xi)]d\xi\}d\tau ,$$

so that

$$ {}_t \bar{s}_x = {}_t \bar{r}_x + \exp\{-\int_0^t [\nu(x+\tau)+\rho(x+\tau)]d\tau\}.$$

We conclude that

(21)
$$ \ell_s(x) = \ell(x) \int_0^x \rho(t) \exp\{-\int_t^x [\nu(\tau)+\rho(\tau)]d\tau\}dt,$$

and get

$$ {}_x \bar{r}_0 = \ell_s(x) \,/\, \ell(x).$$

C. <u>General school entrance rates and drop-out rates</u>. In analogy with the previous notation, we let $p_r(x)dx$ denote the number of individuals in the age bracket from x to x+dx who at time zero do not attend school, and let $p_s(x)dx$ denote the corresponding number in school. Then, of course, $p = p_r + p_s$.

The number of times one observes that some person enters school in the arbitrary area Ω in the Lexis diagram, is

$$\begin{aligned} R(\Omega) = & \int_0^\omega \int_0^T p_r(y)({}_t p_y - {}_t r_y)\rho(y+t)\, I_\Omega(t,y+t)dt\, dy \\ & + \int_0^\omega \int_0^T p_s(y)({}_t p_y - {}_t s_y)\rho(y+t)\, I_\Omega(t,y+t)dt\, dy \\ & + \int_0^\omega \int_0^T b(t)\ell_r(y)\rho(y)\, I_\Omega(y+t,y)dt\, dy. \end{aligned}$$

The aggregate not-in-school lifetime in Ω is

$$\begin{aligned} L_r(\Omega) = & \int_0^\omega \int_0^T p_r(y)({}_t p_y - {}_t r_y)\, I_\Omega(t,y+t)dt\, dy \\ & + \int_0^\omega \int_0^T p_s(y)({}_t p_y - {}_t s_y)\, I_\Omega(t,y+t)dt\, dy \\ & + \int_0^\omega \int_0^T b(t)\ell_r(y)\, I_\Omega(y+t,y)dt\, dy. \end{aligned}$$

We then define the school entrance rate for Ω as

$$R(\Omega)/L_r(\Omega).$$

The corresponding drop-out rate is

$$S(\Omega)/L_s(\Omega),$$

with

$$S(\Omega) = \int_0^\omega \int_0^T p_r(y) \cdot {}_t r_y \cdot \nu(y+t)\, I_\Omega(t,y+t)dt\, dy$$
$$+ \int_0^\omega \int_0^T p_s(y) \cdot {}_t s_y \cdot \nu(y+t)\, I_\Omega(t,y+t)dt\, dy$$
$$+ \int_0^\omega \int_0^T b(t)\ell_s(y)\nu(y)\, I_\Omega(y+t,y)dt\, dy$$

and

$$L_s(\Omega) = \int_0^\omega \int_0^T p_r(y) \cdot {}_t r_y \cdot\, I_\Omega(t,y+t)dt\, dy$$
$$+ \int_0^\omega \int_0^T p_s(y) \cdot {}_t s_y \cdot\, I_\Omega(t,y+t)dt\, dy$$
$$+ \int_0^\omega \int_0^T b(t)\ell_s(y)\, I_\Omega(y+t,y)dt\, dy.$$

D. Generation rates. By specialization of the above results to the Ω_0' in section 3, we get the following generation rates:

Entrance rate: $\int_0^n \ell_r(x+t)\rho(x+t)dt \,/ \int_0^n \ell_r(x+t)dt$

Drop-out rate: $\int_0^n \ell_s(x+t)\nu(x+t)dt \,/ \int_0^n \ell_s(x+t)dt.$

Note that neither of these depend on p_r, p_s, and b.

E. Period rates. For simplicity, let Ω be such that no one born during the study period can enter (or leave) school during $[0,T]$. We can then disregard the third integral in each of the integral formulas of section 5C. In this case,

$$R(\Omega_0) = \int_x^{x+n} \ell(y)\rho(y) \int_{y-T}^{y} \frac{p_r(u)}{\ell(u)} (1-{}_{y-u}\bar{r}_u)du\, dy$$
$$+ \int_x^{x+n} \ell(y)\rho(y) \int_{y-T}^{y} \frac{p_s(u)}{\ell(u)} (1-{}_{y-u}\bar{s}_u)du\, dy,$$

and

$$S(\Omega_0) = \int_x^{x+n} \ell(y)\nu(y) \int_{y-T}^{y} \frac{p_r(u)}{\ell(u)}\, {}_{y-u}\bar{r}_u du\, dy$$
$$+ \int_x^{x+n} \ell(y)\nu(y) \int_{y-T}^{y} \frac{p_s(u)}{\ell(u)}\, {}_{y-u}\bar{s}_u\, du\, dy,$$

and similar formulas hold for $L_r(\Omega_0)$ and $L_s(\Omega_0)$. In the particularly nice situation where (8a) holds, and where we also have

$$p_s(u)/p(u) = \ell_s(u)/\ell(u)$$

for all u, we get the following rates:

Entrance rate: $\int_0^n e^{-rt}\ell_r(x+t)\rho(x+t)dt/\int_0^n e^{-rt}\ell_r(x+t)dt.$

Drop-out rate: $\int_0^n e^{-rt}\ell_s(x+t)\nu(x+t)dt/\int_0^n e^{-rt}\ell_s(x+t)dt.$

Note the many similarities between the results of the present chapter and the previous ones.

Acknowledgements

I wish to acknowledge my great indebtedness to Nathan Keyfitz for many conversations on the subject matter of the present paper. I am also grateful to Nader Fergany, Samuel Preston, Robert Retherford, and Griffith Feeney, discussions with whom inspired many of the formulations here; and to Mr. Bjørn Tønnesen, who has checked all arguments and formulas.

References

[1] Bellman, R. and Cooke, K.L. 1963. "Differential-Difference Equations", Academic Press, New York.

[2] Feller, W. 1957. "An Introduction to Probability Theory and Its Applications, Volume I, 2nd Edition", Wiley, New York.

[3] Hoem, J.M. 1970. Probabilistic fertility models of the life table type, Theor. Popul. Biology, **1** (1), 12-38.

[4] Hoem, J.M. 1969. Purged and partial Markov chains, Skand. Aktuarietidskr., **52**, 147-155.

[5] Keyfitz, N. 1968a. "Introduction to the mathematics of population", Addison-Wesley, Reading, Mass.

[6] Keyfitz, N. 1968b. A life table that agrees with the data: II, J. Am. Statist. Ass., **63**, 1253-1268.

[7] Keyfitz, N. 1970. Finding probabilities from observed rates or How to make a life table, The American Statistician, **24** (1), 28-33.