# ON THE STATISTICAL THEORY OF ANALYTIC GRADUATION

JAN M. HOEM
THE CENTRAL BUREAU OF STATISTICS OF NORWAY

## 1. Introduction; motivating examples

1.1. *Example* 1. *Rates of mortality.* Standard methods for the investigation of human mortality will produce statistics such as those given in extract in Table I. The mortality rate at age $x$ is interpreted as a measure of the mortality risk for women born in the year 1968 minus $x$, and the corresponding number "exposed to risk" in Column 2 is used as a measure of the accuracy of this rate. (This will be made clearer later on.) Unless the population is substantially larger than the one producing these data. the diagram of the sequence of rates, plotted against age, will have a rather rugged appearance. Figure 1, based on the same data as Table I, shows the typical form of such diagrams. There seems to be a universal conviction, however, that "real mortality" would be portrayed by a smooth curve, and that any irregularities of curves of observed mortality rates are due to accidental circumstances. The observed rates are then regarded as "raw" or primary estimates of the underlying "real" rates, and graduation is employed to get a smoother curve.

A number of techniques have been developed to graduate age specific mortality rates, as can be seen from any text on the subject. (See, for instance, [55]; [59], pp. 145–197; [83], pp. 216–237, 243–244, and 251–252.) Most of these methods have been developed by intuitive arguments, at least initially, but investigations of statistical properties of some of them have also appeared [1]; [2]; [43]; [44]; [46]; [61]; [69]; [71]; [76]; [83], p. 252. One class of such methods consists in fitting a parametric function to the observed rates. We shall call this the class of *analytic graduation methods.*

Quite a number of functions have been suggested for analytic graduation of mortality rates [45], pp. 236–238; [67], pp. 453–454; [79], pp. 79–85; [83], pp. 56–60 and 243–244. By far the most commonly used for the adult ages is the Gompertz-Makeham formula

$$g_x(\alpha, \beta, c) = \alpha + \beta c^x \qquad \text{for} \quad \beta > 0, c > 1, \alpha > -\beta c^{x\min}, \tag{1.1}$$

where $x$ represents age attained. We have fitted this function to our data in

The present paper was written while the author visited the University of California, Berkeley, on a Ford Foundation fellowship.
This research was partially supported by National Science Foundation Grant GP 15283.

Figure 1 by minimum $\chi^2$. Other common methods are least squares and some moments methods. We shall describe each of these in turn.

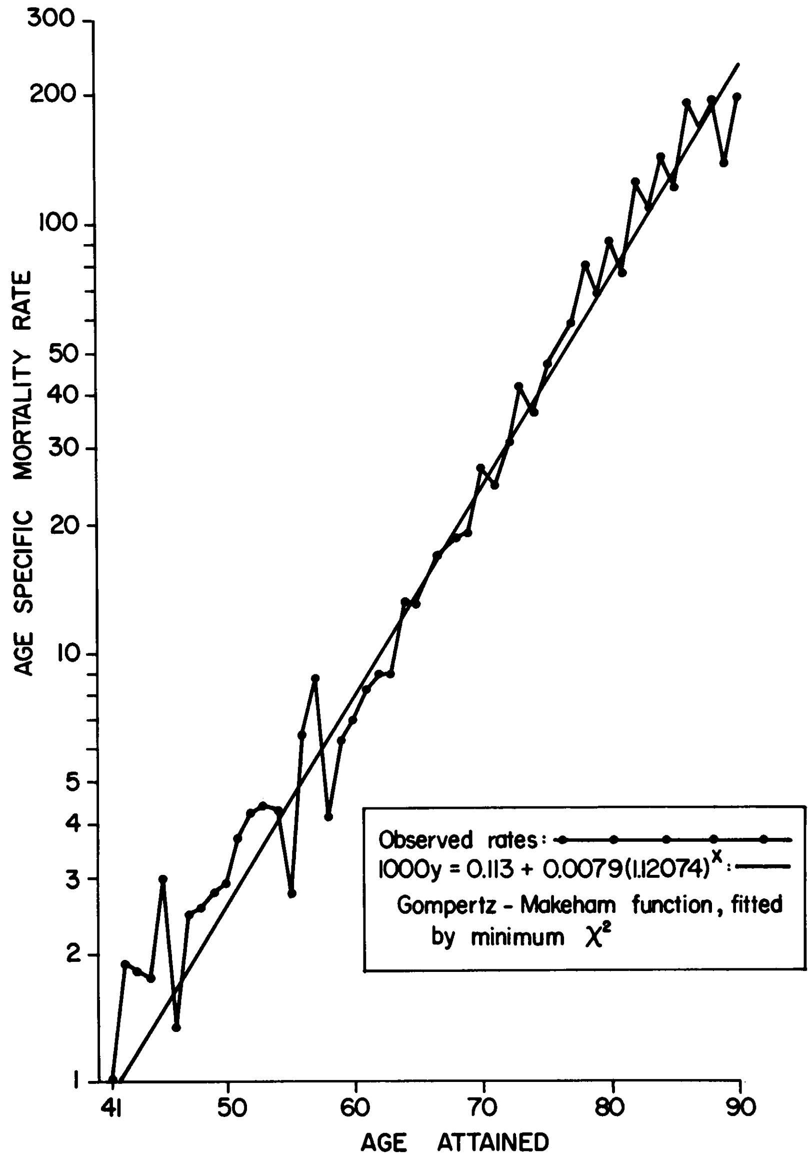


FIGURE 1
Age specific mortality rates per 1000.
Females, Oslo, 1968.

TABLE I

AGE SPECIFIC MORTALITY

FEMALES, MUNICIPALITY OF OSLO, NORWAY, 1968

Column 2 shows arithmetic mean of the number of persons at a given age as of January 1, 1968, and the corresponding number as of December 31, 1968.
Column 3 shows age at death taken as 1968 minus year of birth.
Column 4 shows ratio between entries in columns 3 and 2, multiplied by 1000.
Source: Central Bureau of Statistics of Norway.

| Age<br>1 | Exposed to risk<br>2 | Deaths<br>3 | Mortality rate per thousand<br>4 |
|---|---|---|---|
| 40 | 2798 | 1 | 0.357 |
| 41 | 2924.5 | 3 | 1.025 |
| 42 | 3156 | 6 | 1.901 |
| 43 | 3272.5 | 6 | 1.833 |
| 44 | 3465.5 | 6 | 1.731 |
| 45 | 3639 | 11 | 3.022 |
| 46 | 3770 | 5 | 1.326 |
| 47 | 4057 | 10 | 2.464 |
| 48 | 3886.5 | 10 | 2.573 |
| 49 | 3650.5 | 10 | 2.739 |
| .. | ...... | .. | ..... |
| .. | ...... | .. | ..... |
| 80 | 1204.5 | 111 | 92.154 |
| 81 | 1064 | 81 | 76.127 |
| 82 | 930 | 118 | 126.881 |
| 83 | 835 | 90 | 107.784 |
| 84 | 709.5 | 101 | 142.353 |
| 85 | 615.5 | 74 | 120.227 |
| 86 | 502 | 97 | 193.227 |
| 87 | 408 | 68 | 166.666 |
| 88 | 341 | 66 | 193.548 |
| 89 | 378 | 53 | 140.211 |
| 90 | 218.5 | 43 | 196.796 |

1.2 *Example* 2. *Rates of fertility.* A standard investigation of age specific human fertility would produce a table quite similar to Table I, except of course that column 3 would contain numbers of births (or usually numbers of liveborn children) by age of mother. A corresponding diagram would look something like the one in Figure 2, and graduation would again give a smoother curve.

A fertility curve of this sort closely resembles certain density functions, and one category of functions proposed for the analytic graduation of fertility curves consists of densities from the Pearson family [13]; [27]; [45], pp. 140–169; [53]; [56]; [73]; [77]; [80]; [81]; in particular Pearson type I, III, IV, VI, and the normal density, multiplied by a constant.

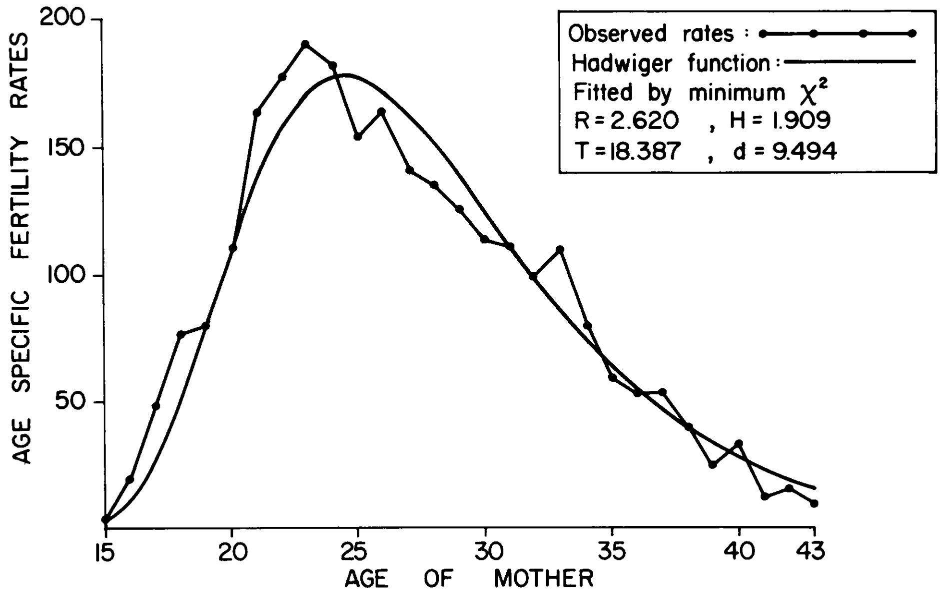


FIGURE 2
Age specific fertility rates per 1000.
Females, Stavanger (Norway), 1968.

Another category of graduating functions consists of polynomials in $x$ [11], [27], such as

$$b_x(a, b, c, d) = (x - \alpha + 1)(\beta - x)^2(a + bx + cx^2 + dx^3), \tag{1.2}$$

where $x$ stands for age of mother at childbearing, and where $[\alpha, \beta\rangle$ is the fertile period for females. It is customary to take $\alpha = 15$ and $\beta = 45$ or $\beta = 50$, but in certain cases $\alpha$ and $\beta$ occur as parameters which are estimated [13]; [54]; [77].

The Hadwiger function,

$$h_x(R, T, H, d) = \frac{RH}{T\sqrt{\pi}}\left(\frac{T}{x - d}\right)^{3/2} \exp\left\{- H^2\left(\frac{T}{x - d} + \frac{x - d}{T} - 2\right)\right\}, \tag{1.3}$$

with $R > 0$, $T > 0$, $H > 0$, $d < \alpha$, is a third type of graduating formula [27]; [28]; [31]; [45], pp. 149-169; [80]; [81]. (We follow Yntema's notation.) Other functions have also been suggested [12], [48], [49], [50], [54], [72].

Naturally, the same type of functions will be used for the graduation of other vital rates whose diagrams have the same general form as fertility rates, such as marriage rates [22], pp. 99-101, and [49].

1.3. The above age specific rates of mortality and fertility are examples of the kind of vital rates which occur in fields such as actuarial science, biostatistics, and demography. In the present paper we shall make a contribution to the statistical theory of curve fitting as applied to such rates in general. We shall suggest a probabilistic model within which the rates appear as estimators for certain parameters called forces of transition, and shall show how the analytic graduation can be interpreted as a procedure used to further estimate a set of "more basic" parameters, namely those of the graduating function.

The model will be introduced in Section 2. In Sections 3 and 4, we describe how the rates appear within the model, and Sections 5 to 9 are devoted to the study of analytic graduation methods. We shall be concerned mainly with the asymptotic statistical properties (as the population size $N$ increases) of the estimators. Most of our results are straightforward consequences of general asymptotic theory, and we shall often use standard theorems from that field, like those of Chapters 4 and 5 in [10] and Theorem 4.2.5 in [3], without explicit reference. We *shall* quote references whenever we use a deeper result.

Since we use standard theorems, it is not surprising that we can prove theorems which correspond to previous standard results. Thus (speaking informally here) we shall see that none of the general estimation procedures we study will be better than one of the maximum likelihood type, and that a (modified) minimum $\chi^2$ procedure is equally good, while moment methods will usually give less favorable results.

We feel that there may be a need for some explanation why procedures of the type which we shall describe are preferred to certain others. Rather than breaking up our presentation of the techniques involved by giving parts of this explanation as we go along, we have preferred to include it all in Section 10.

Apart from what is contained in Sections 1.1 and 1.2 above, no numerical examples will be given in this paper. Numerical investigations are planned and will be reported at a later date.

## 2. A Markov process model

2.1. *The general model.* To describe the phenomena at hand we shall use a Markov process model. Let $y_t$ be the sample function value at time $t$ of a time inhomogeneous Markov process with a denumerable state space $I$ and a continuous time parameter restricted to some finite time interval $[0, \zeta\rangle$. Let the transition probabilities be

$$P_{i,J}(s, t) = P\{y_t \in J \mid y_s = i\}, \tag{2.1}$$

for $0 \leqq s < t < \zeta$, and $i \in I$ and $J \subseteq I$, and assume that $P_{i,I}(s, t) \equiv 1$ and $\lim P_{i,j}(s, t) \equiv \delta_{i,j}$ (a Kronecker delta) as $t \downarrow s$. We introduce the *forces of transition*,

$$\mu_{i,j}(s) = \lim_{t \downarrow s} P_{i,j}(s, t)/(t - s) \qquad \text{for} \quad i \neq j, \tag{2.2}$$

and the *forces of decrement*,

$$\mu_i(s) = \lim_{t \downarrow s} \{1 - P_{i,i}(s, t)\}/(t - s), \tag{2.3}$$

for $0 \leqq s < \zeta$, and assume that all $\mu_i$ and $\mu_{i,j}$ are finite and integrable over $[0, \zeta\rangle$. We also assume that

$$\mu_i = \sum_{j \in I - i} \mu_{i,j} \qquad \text{for each} \quad i \in I. \tag{2.4}$$

We shall call a state $i$ absorbing if $\mu_i = 0$.

The problem which leads us to study analytic graduation consists in finding a method for estimating one or more of the $\mu_{i,j}(\cdot)$ from data of the type which one encounters within the fields of application mentioned at the beginning of Section 1.3.

2.2. *Examples.* We shall give some examples to show how models in the applications appear as particular cases of the general model in Section 2.1 above.

(i) Our simplest example will be a model with only two states, called "alive" (state 1) and "dead" (state 2). State 2 is absorbing, and there is only one nonzero force of transition and of decrement, namely,

$$\mu(\cdot) = \mu_1(\cdot) = \mu_{1,2}(\cdot), \tag{2.5}$$

called *the force of mortality*. The rates of Section 1.1 will be seen to appear within this model. The time parameter is represented by a person's age.

(ii) The age specific fertility rates of Section 1.2 can be interpreted within a model with a double infinity of states. A woman will be said to be in state $(k, 1)$ at age $x$ if she is alive then and her parity is $k$, that is, she has had $k$ births, $k = 0, 1, 2, \cdots$. She will be said to be in state $(k, 2)$ at age $x$ if she has died within age $x$ and her parity at death was $k$. All states of the form $(k, 2)$ are absorbing. We select two suitable functions, $\mu(\cdot)$ and $\varphi(\cdot)$, and set

$$\mu_{(k,1),(k,2)}(\cdot) = \mu(\cdot), \tag{2.6}$$

and

$$\mu_{(k,1),(k+1,1)}(\cdot) = \varphi(\cdot), \tag{2.7}$$

for $k = 0, 1, \cdots$, while all other $\mu_{i,j} = 0$. The function $\mu$ will be called the *force of mortality*, and $\varphi$ will be called the *force of fertility*. Again the time parameter is represented by the woman's age. This model, which we have studied in some detail previously [36], is not particularly "realistic", but it is probably the simplest one in which the rates of Section 1.2 can be meaningfully discussed. More realistic fertility models of this type have appeared elsewhere [37], [38].

(iii) To describe *marriage formation and dissolution*, we suggest a model with five states, called "never married" (state 1), "married" (state 2), "widowed" (state 3), "divorced" (state 4), and "dead" (state 5). State 5 is absorbing. The following forces correspond to impossible direct transitions, and are therefore identically equal to zero: $\mu_{i,1}$ for $i > 1$, $\mu_{3,4}$, $\mu_{4,3}$, $\mu_{5,j}$ for $j < 5$. The model applies to one sex only, while the other sex appears only implicitly, as a kind of shadow factor. We have also looked at marriage models elsewhere [41].

Other models of this type have been studied by, for instance, Du Pasquier [21]; Sverdrup [70]; Simonsen [66]; Chiang [15], Chapters 4, 5, and 7; and Hoem [35]. Compare also [25] and [64].

In each of these models, a state $i \in I$ corresponds to some vital status, that is, a marital status, a social status, a birth parity, and so on. A transition similarly corresponds to a vital event, such as a death, a birth, a marriage, a divorce, and so on. An individual sample path will be visualized intuitively as a person (or sometimes a group of persons, such as a household or a family) moving through some of the statuses of the system specified. The sample paths will be taken as stochastically independent.

2.3. *Seniority.* In demographic models one often wishes to distinguish between an age parameter (which may be actual age obtained, duration of marriage, interval since last previous birth, or the like), calendar time, and observational time. It is the age parameter which corresponds to the time parameter in the Markov process of Section 2.1. In a general model it may be useful to have a separate name for this (unspecified) age parameter, covering all interpretations which it may have in the applications. Following Henry [33] who calls it *ancienneté*, we shall use the name *seniority* for it.

2.4. *Some basic assumptions about the forces of transition.* In what follows, we shall disregard the forces of transition which are identically equal to zero because they correspond to impossible direct transitions by the definition of the model. Even if the state space $I$ can be (countably) infinite, there are many cases where only a finite number of the nonzero forces of transition are *distinct*. [Compare Example 2.2, (ii).] *We shall assume everywhere that there exist $A$ nonnegative real functions, $\lambda_1, \cdots, \lambda_A$, such that each $\mu_{i,j}$, not identically zero, equals some $\lambda_a$, and such that for each $\lambda_a$ there exists a $\mu_{i,j} = \lambda_a$.* By (2.4), we may then write

$$\mu_i = \sum_{a=1}^{A} c_a(i)\lambda_a \qquad \text{for each} \quad i \in I, \tag{2.8}$$

where

$$c_a(i) = \sum_{\{j \in I - i:\, \mu_{i,j} = \lambda_a\}} 1. \tag{2.9}$$

Each $c_a(i) < \infty$ because all $\mu_i < \infty$. Equation (2.8) shows that for any given $i \in I$, exactly $\Sigma_a c_a(i)$ of the $\mu_{i,j}$ can be positive, that is, a finite number of the $\mu_{i,j}$ only, while the rest are identically equal to zero.

Let us also assume everywhere that

$$\sup\{\mu_i(s): 0 \leqq s < \zeta, i \in I\} < \infty. \tag{2.10}$$

This assumption is not necessary for what follows, and it can be relaxed [39]; [40], Section 5, but one will expect it to hold in practice and it will simplify our exposition. It follows from (2.10) [40], Section 3.1, that there only exists a finite number of *distinct* vectors $\mathbf{c}(i) = (c_1(i), \cdots, c_A(i))$. Let us call these $(c_{b,1}, \cdots, c_{b,A})$ for $b = 1, 2, \cdots, B$, and let

$$\gamma_b = \sum_{a=1}^{A} c_{b,a}\lambda_a \qquad \text{for} \quad b = 1, 2, \cdots, B. \tag{2.11}$$

Then each nonzero $\mu_i$ equals some $\gamma_b$ and for each $\gamma_b$ there is a $\mu_i$ which equals it. There is, thus, only a finite number of distinct forces of decrement as well. (This need not be the case if (2.10) does not hold.) For each $i \in I$, let $b(i)$ be defined by $\mu_i = \gamma_{b(i)}$. Then, by (2.9),

$$c_{b(i),a} = \sum_{\{j \in I - i: \mu_{i,j} = \lambda_a\}} 1. \tag{2.12}$$

Let

$$\mathbf{c} = \begin{pmatrix} c_{1,1}, & \cdots, & c_{1,A} \\ \cdots & \cdots & \cdots \\ c_{B,1}, & \cdots, & c_{B,A} \end{pmatrix} \tag{2.13}$$

We shall finally assume everywhere that the rank of $\mathbf{c}$ is $B$. The extension to rank $\mathbf{c} < B$ is easy [40], Section 3.1, but it leads to slightly more complicated formulas.

## 3. The primary or "raw" estimates

3.1. *Approximation of the $\lambda_a$ by step functions.* As a first step in our description of the kind of estimation methods which have produced the rates of Sections 1.1 and 1.2, we shall approximate the $\lambda_a$ by step functions. The seniority interval $[0, \zeta\rangle$ is paritioned into $D$ subintervals, $[\zeta_0, \zeta_1\rangle$, $[\zeta_1, \zeta_2\rangle$, $\cdots$, $[\zeta_{D-1}, \zeta_D\rangle$, with $\zeta_0 = 0$ and $\zeta_D = \zeta$. Let $I_d(\cdot)$ be the indicator function of the interval $[\zeta_{d-1}, \zeta_d\rangle$, and let

$$\lambda_a^{\#}(\cdot) = \sum_{d=1}^{D} \lambda_{a,d} I_d(\cdot). \tag{3.1}$$

Here each $\lambda_{a,d}$ is a constant chosen in such a way that it can represent the values of $\lambda_a$ in $[\zeta_{d-1}, \zeta_d\rangle$. If $\lambda_a$ is assumed to be a nice and smooth function, with certain known monotonicity properties, say, then $\lambda_a^{\#}$ will inherit these properties, modified of course, by the fact that the latter is a step function.

*In what follows, we shall assume that the $\lambda_a^{\#}$ give an adequate representation of the $\lambda_a$, and our calculations will be made as if we actually had $\lambda_a = \lambda_a^{\#}$ for $a = 1, 2, \cdots, A$.*

3.2. *More about the $\zeta_d$.* In this presentation, we use the same partitioning $\{\zeta_d : d = 0, 1, \cdots, D\}$ for all $\lambda_a$. In certain situations one would rather use different partitionings for different $\lambda_a$. The results of this paper will continue to hold for such cases with only quite obvious modifications [37].

The approach sketched in Section 3.1 is closely related to histogram methods for the estimation of a probability density or a generalized failure rate [6], [74], [75]. Although the lengths of the histogram intervals are often made to converge to zero as the number of observations increase, this is not the case for the seniority subintervals above. The $\zeta_d$ will typically be selected according to conventional rules established with different considerations in mind than statistical convergence properties. When $\zeta$ is of the size order of several decades, as is often the case, the seniority interval $[0, \zeta\rangle$ will usually be partitioned into one year or five year intervals, possibly with a longer "tail" interval at the upper end. There is a tendency to use shorter subintervals in a large population than in a small one, at least if the data are reliable, but an interval length shorter than one year is commonly used only in certain standardized contexts, such as in investigations of infant mortality, where $\zeta$ equals one year of age [8], p. 211; [30], Tables 38 to 42; [67], p. 84. There seems to be no tendency toward letting such interval lengths decrease to zero.

3.3. *On the observational plan.* There are a number of observational plans (or ascertainment methods) in use in the fields of application we have in mind, and one could construct others from ideas used in life testing. (See, for example, [20], [32], [65], [82].

In the present paper we shall only consider observational plans where a group of people are followed continuously over some time interval $[0, T\rangle$. The data collection will consist of noting what happens to each person while under observation, that is, which states of $I$ he visits and just when vital events occur to him.

It is characteristic for the types of populations which occur in practice that some people enter them and others leave during the study period. They may also be heterogeneous with respect to seniority, in the sense that those who come under observation (whether they are in the population from the outset or enter later on) may have different seniorities at time $T$. We want to cover such possibilities. Let, therefore, $N$ be the number of individuals ever observed. Let us say that person number $k$ enters the population at some time $t_k \in [0, T\rangle$ with seniority $x_k$ and a status corresponding to state $r_k$, and that he stays there at least until time $t_k + z_k \in [0, T]$, when observation is discontinued. We shall take the entrance time $t_k$, the initial seniority $x_k$, the initial state $r_k$, and the exposure time $z_k$ to be preassigned, that is, not random. (Other possibilities are discussed in [40].) Any period spent in an absorbing state, for instance after the death of the individual, is included in the period of exposure $[t_k, t_k + z_k\rangle$, although of course no actual observation is made after a path has entered such a state.

We shall also take $N$ to be nonrandom.

3.4. *Estimation of the* $\lambda_{a,d}$. We get an estimator for $\lambda_a^{\#}$ by plugging estimators $\hat{\lambda}_{a,d}$ for the $\lambda_{a,d}$ into the right side of (3.1). (This is how the rugged curves in Figures 1 and 2 have arisen.) Standard estimators used for this purpose are occurrence/exposure rates, like those in Sections 1.1 and 1.2 [62], [63]. We shall see how these arise.

Some of the $\lambda_{a,d}$ may be known to be zero because they correspond to vital events which are impossible during $[\zeta_{d-1}, \zeta_d\rangle$, such as births after menopause. We will take the other $\lambda_{a,d}$ to be strictly positive. Let

$$\mathscr{G} = \{(a, d): \lambda_{a,d} > 0\}. \tag{3.2}$$

We can regard $\{\lambda_{a,d}: (a, d) \in \mathscr{G}\}$ as a point in the space

$$\Lambda_0 = \mathop{\times}_{(a,d)\in\mathscr{G}} \{x_{a,d} > 0\}. \tag{3.3}$$

The situation in hand will usually restrict the possible points we actually can have to a proper subset $\Lambda$ of $\Lambda_0$. We shall take $\Lambda$ to be open.

Now let $M_k(a, d)$ be the number of transitions observed for path number $k$ during the seniority interval $[\zeta_{d-1}, \zeta_d\rangle$, direct from any state $i$ to any state $j$ where $\mu_{i,j} = \lambda_a$. Let $U_k(i, d)$ be the total time spent in state $i$ during $[\zeta_{d-1}, \zeta_d\rangle$ by this path, and let

$$V_k(b, d) = \sum_{\{i\in I:\, \mu_i = \gamma_b\}} U_k(i, d). \tag{3.4}$$

Then $V_k(b, d)$ is the total time spent in any state $i$ where $\mu_i = \gamma_b$ by path $k$ during the interval mentioned. Finally, let

$$\begin{aligned} L_k(a, d) &= \sum_{b=1}^{B} c_{b,a} V_k(b, d) \\ &= \sum_{i\in I} c_{b(i),a} U_k(i, d), \end{aligned} \tag{3.5}$$

and let us use the notation $X = \sum_k X_k$, where $X_k$ is any quantity depending on $k$. Since the forces of transition are represented by step functions, we can then use the same method as in [37], Section 4.8 to write the likelihood in the form

$$\begin{aligned} &\prod_{(a,d)\in\mathscr{G}} \lambda_{a,d}^{M(a,d)} \exp\left\{-\sum_{d=1}^{D}\sum_{b=1}^{B} \gamma_{b,d} V(b, d)\right\} \\ &\qquad = \prod_{(a,d)\in\mathscr{G}} \lambda_{a,d}^{M(a,d)} \exp\left\{-\sum_{(a,d)\in\mathscr{G}} \lambda_{a,d} L(a, d)\right\}, \end{aligned} \tag{3.6}$$

where $\gamma_{b,d} = \sum_{a=1}^{A} c_{b,a}\lambda_{a,d}$. (Compare (2.11).) Thus we are dealing with a Darmois-Koopman class of probability distributions, and one may show [40], Section 3.1, that $\{M(a, d),\ V(b, d): a = 1, 2, \cdots, A;\ b = 1, 2, \cdots, B;\ d = 1, 2, \cdots, D\}$ is minimal sufficient for the $\lambda_{a,d}$. An unrestricted maximization of the likelihood function would give the estimators

$$\hat{\lambda}_{a,d} = M(a, d)/L(a, d), \tag{3.7}$$

which are the occurrence/exposure rates we mentioned. (We arbitrarily set $\hat{\lambda}_{a,d} = 0$ if $L(a, d) = 0$.) The point $\{\hat{\lambda}_{a,d}: (a, d) \in \mathscr{G}\}$ need not lie in $\Lambda$, nor even in $\Lambda_0$ if some $\hat{\lambda}_{a,d} = 0$. However, under certain conditions, spelt out in Theorem 1 below, the probability that the point lies in $\Lambda$ increases to 1 as $N \to \infty$.

3.5. *Asymptotic properties of the* $\hat{\lambda}_{a,d}$. For each $(a, d) \in \mathscr{G}$, the variables $L_1(a, d), \cdots, L_N(a, d)$ will not generally be identically distributed unless all $(x_k, z_k, r_k)$ are equal. Similarly for $M_1(a, d), \cdots, M_N(a, d)$. Nevertheless, one can prove the following consistency theorem [40], Section 4.2:

THEOREM 1. *Assume that* $P\{L(a, d) > 0\} \to 1$ *as* $N \to \infty$, *and that a finite positive limit*

$$L_{a,d} = \lim_{N \to \infty} EL(a, d)/N \tag{3.8}$$

*exists. Then* $\hat{\lambda}_{a,d}$ *converges to* $\lambda_{a,d}$ *in probability as* $N \to \infty$.

To arrive at a theorem concerning the asymptotic distribution of $\hat{\lambda}_{a,d}$ as $N \to \infty$, we make an additional set of assumptions, which establish a grouping of the $(x_k, z_k, r_k)$ at a finite set of strategic values. More precisely we make

ASSUMPTION 1. *There exists a finite set of possible initial seniorities,* $y_1, \cdots, y_H$, *a finite set of possible exposure times,* $w_1, \cdots, w_J$, *and a finite set of possible initial states,* $s_1, \cdots, s_Q$, *such that each* $(x_k, z_k, r_k)$ *must equal some* $(y_h, w_j, s_q)$. *We let*

$$\mathscr{K}_{h,j,q} = \{k\colon (x_k, z_k, r_k) = (y_k, w_j, s_q)\}, \tag{3.9}$$

*and let* $S_{h,j,q}(N)$ *be the number of elements in* $\mathscr{K}_{h,j,q}$. *We assume that*

$$\alpha_{h,j,q} = \lim_{N \to \infty} S_{h,j,q}(N)/N \tag{3.10}$$

*exists for each* $(h, j, q)$.

If $\varepsilon_{h,j,q}(a, d) = EL_k(a, d)$ for $k \in \mathscr{K}_{h,j,q}$, we get under Assumption 1 that the $L_{a,d}$ of (3.8) satisfy

$$L_{a,d} = \sum_{h,j,q} \alpha_{h,j,q}\varepsilon_{h,j,q}(a, d). \tag{3.11}$$

We may then prove ([40], Section 4.2; compare [47]) Theorem 2.

THEOREM 2. *Under Assumption* 1, *the variables* $N^{\frac{1}{2}}(\hat{\lambda}_{a,d} - \lambda_{a,d})$ *for which* $P\{L(a, d) > 0\} \to 1$ *as* $N \to \infty$, *are asymptotically independent and normally distributed with means* 0 *and asymptotic variances*

$$\sigma^2_{a,d} = \text{as. var}\, N^{1/2}(\hat{\lambda}_{a,d} - \lambda_{a,d}) = \lambda_{a,d}/L_{a,d}. \tag{3.12}$$

We note that under the assumptions of Theorem 2,

$$\hat{\sigma}^2_{a,d} = N\hat{\lambda}_{a,d}/L(a, d) \tag{3.13}$$

is a consistent estimator for $\sigma^2_{a,d}$. Thus we see a justification of the use of the number exposed to risk [which is $L(a, d)$ here] as an intuitive measure of the accuracy of the corresponding rate of transition $\hat{\lambda}_{a,d}$, as mentioned in Section 1.1 for a special case.

We also note that we do not need to know the value of $N$ in order to estimate the $\lambda_{a,d}$ and the asymptotic variance $\sigma^2_{a,d}/N$ of the $\hat{\lambda}_{a,d}$.

## 4. Non-observation of part of the state space

4.1. *A problem.* In Section 3 we assumed that one could observe what state a sample path visited at any time. This need not be the case in practice. Let us give two examples.

(i) Demographic studies will often be concerned with people living in a restricted area, such as a country or part of a country, and there will be some in and out migration. Say that a study of marriages is carried out, perhaps based on a model like the one in Section 2.2 (iii). If a person initially lives in the study area, then leaves and stays away for a while, and subsequently returns while the study is still being conducted, it rarely happens that his changes of marital status (if any) while outside the study area are traced. In many cases one will know his marital status on departure from the study area, as well as his status as he returns, but nothing more.

(ii) Similar problems occur in studies of the mortality of insured lives. A person may cancel his insurance policy and be uninsured for a while, then take out a new policy, which may be cancelled again after a while, and so on. The insurer will keep track of deaths among the persons covered by his policies, but will not usually know what happens to the uninsured.

The question is how one should take account of phenomena like these in the estimation procedures.

4.2. *Formalization of the two examples.*

(i) To describe the example in (i) above in terms of a probabilistic model, let $I_1 = \{1, 2, 3, 4, 5\}$ be the state space of the example in 2.2 (iii), let $I_2 = \{1, 2\}$, and let $I = I_1 \times I_2$. An individual with marital status $j$ will be said to be in state $(j, 1)$ if he lives in the study area, and in state $(j, 2)$ if he lives outside it. A migration *out of* the study area will correspond to a transition from a state $(j_1, 1)$ to a state $(j_2, 2)$. A migration *into* the study area will correspond to a transition from a state $(j_2, 2)$ to a state $(j_1, 1)$. In most cases, $j_1 = j_2$. In any case, we shall take $j_1$ to be observable. Whatever moves the sample path otherwise makes while in the subspace $\{(j, 2) : j \in I_1\}$ will not be observed.

(ii) To formalize the second example above, let us use four states, called "alive and insured" (state 1), "alive and uninsured" (state 2), "dead while insured" (state 3), and "dead while uninsured" (state 4). Which transitions are possible and which are not follows directly from the state names. Except for transitions from state 2 to state 4, all transitions (and the dates on which they occur) are recorded. (This is essentially the model studied by Du Pasquier [21], Fix and Neyman [25], and Sverdrup [70], except that they took all transitions as recorded. Recording problems different from the present one have been studied by Høyland [42], Kruopis [47], and others.)

Let us take all forces of transition to be constants. This will suffice for our purposes, which are those of illustration. Generalization to other cases is simple. We shall take the forces of mortality of the insured and the uninsured to be equal, and let $\mu = \mu_{1,3} = \mu_{2,4}$. Let $\nu = \mu_{1,2}$, $\rho = \mu_{2,1}$, $\alpha = \mu + \nu$,

$\beta = \mu + \rho$. Sample path number $k$ is followed over the period $[0, z_k]$, and we say that $k \in \mathscr{K}$ if this sample path is in state 2 or 4 at time $z_k$, that is, if person number $k$ is uninsured then. All $N$ paths start in state 1, and they make a total number of $M_{i,j}$ jumps from state $i$ to state $j$, for $(i, j) \in \{(1, 2), (1, 3), (2, 1)\}$. Let $W$ denote the total time spent in state 2 by the paths $k \notin \mathscr{K}$, and let $V$ be the total time spent in state 1 by all paths taken together. For $k \in \mathscr{K}$, let $z_k - U_k$ be the time of the last jump recorded from state 1 to state 2 for path $k$, that is, the time to last observed cancellation. Then the corresponding likelihood can be written as

$$e^{-\alpha V - \beta W} \nu^{M_{1,2}} \rho^{M_{2,1}} \mu^{M_{1,3}} \beta^{-K} \prod_{k \in \mathscr{K}} (\mu + \rho e^{-\beta U_k}), \tag{4.1}$$

where $K$ is the number of elements in $\mathscr{K}$. The maximum likelihood estimator of $\nu$ turns out to be

$$\hat{\nu} = M_{1,2}/V, \tag{4.2}$$

which is what (3.7) would have given. ($M_{1,2}$ is the number of cancellations observed.) Closed, explicit expressions for the maximum likelihood estimators of $\mu$ and $\rho$ do not exist in this case. We can still get an estimator of $\mu$, however, by letting

$$\hat{\mu} = M_{1,3}/V. \tag{4.3}$$

($M_{1,3}$ is the number of insured deaths.) The properties of $\hat{\nu}$ and $\hat{\mu}$ will appear by specialization of the results in Section 4.3 below.

4.3. *The general case.* Consider now the general model with the assumptions made in Sections 2 and 3.1. Let the state space $I$ be partitioned into two disjoint subsets, $H$ and $J$, and assume that all transitions between states in $H$ can be recorded, while no transitions between states in $J$ are recorded. Any transition from a state in $H$ to one in $J$ is recorded, as are all jumps from $J$ to $H$. For both kinds of jumps, one also records the state *to* which the jump is made.

We redefine the quantities $M(a, d)$ and $L(a, d)$, initially introduced in Section 3.4, as follows.

Let $\mathscr{A}$ be the set of the $a$ for which there exists a $\mu_{i,j}$, with $i \in H$, such that $\mu_{i,j} = \lambda_a$. For each $a \in \mathscr{A}$ and each $d \in \{1, 2, \cdots, D\}$ let $M(a, d)$ now be the total number of transitions observed during the seniority interval $[\zeta_{d-1}, \zeta_d\rangle$, for all paths taken together, direct from any state $i \in H$ to any state $j \in I - i$ such that $\mu_{i,j} = \lambda_a$. Furthermore, let

$$L(a, d) = \sum_{i \in H} c_{b(i),a} U(i, d), \tag{4.4}$$

with $c_{b(i),a}$ given by (2.12) and $U(i, d)$ defined as in Section 3.4. For $a \in \mathscr{A}$, let $\hat{\lambda}_{a,d}$ be given by (3.7) with the *new* definitions of $M(a, d)$ and $L(a, d)$. *Then Theorems* 1 *and* 2 *hold verbatim for the* $a \in \mathscr{A}$, even though the $\hat{\lambda}_{a,d}$ need not be maximum likelihood estimators, as demonstrated in the example in (ii) above. If there does not exist any $\mu_{i,j}$, with $i \in J$, such that $\mu_{i,j} = \lambda_a$ for any $a \in \mathscr{A}$, the $\hat{\lambda}_{a,d}$ *will* be maximum likelihood estimators, in the sense that they maximize the likelihood under free variation of the $\lambda_{a,d}$ in $\Lambda_0$.

If the state $j$ cannot be recorded when there is a jump from a state $i \in H$ to a state $j \in J$, the results above continue to hold, provided we again redefine the quantities involved in a natural way. In the definition of $M(a, d)$, we must only include jumps from $i$ to $j$ where both $i$ and $j \neq i$ belong to $H$, and where $\mu_{i,j} \doteq \lambda_a$. The set $\mathscr{A}$ is similarly reduced. This time we also redefine $c_{b,a}$ by letting

$$c_{b(i),a} = \sum_{\{j \in H - i:\, \mu_{i,j} = \lambda_a\}} 1 \qquad \text{for} \quad i \in H, a \in \mathscr{A}. \tag{4.5}$$

Using (4.5), we define $L(a, d)$ for $a \in \mathscr{A}$ by (4.4).

## 5. Conventions and notation relating to analytic graduation

5.1. *Analytic graduation.* Although an original $\lambda_a$ is assumed to be a nice and smooth function, the estimators $\hat{\lambda}_{a,d}$ now in use, such as those in (3.7), will typically produce a $\hat{\lambda}_a^{\#}$ which is considered too irregular, except in large populations. (Compare the account on page 561 in [18].) Analytic graduation then consists in selecting some nice, parametric function $g_a(\cdot, \boldsymbol{\theta}_a)$ and some representative seniority $\xi_d$ from each interval $[\zeta_{d-1}, \zeta_d\rangle$, and in getting an estimator $\hat{\boldsymbol{\theta}}_a$ for $\boldsymbol{\theta}_a$ by fitting the values $\{g_a(\xi_d, \boldsymbol{\theta}_a): d = 1, 2, \cdots, D\}$ to $\{\hat{\lambda}_{a,d}: d = 1, 2, \cdots, D\}$ by a suitable method. The function $g_a(\cdot, \hat{\boldsymbol{\theta}}_a)$, usually regarded as a function of a *continuous* seniority variable $x$, represents the final estimator for the function $\lambda_a(\cdot)$.

Most methods for constructing an estimator $\hat{\boldsymbol{\theta}}_a$ are based on analogies with estimation methods used in other contexts [1], [59], [83]. We shall study least squares and minimum $\chi^2$ methods in Section 6 [7], [12], [16], [27], [28], [48], [49], [50], [61]. (See also [60].) In Section 7, we shall discuss moment methods [11]; [13]; [28]; [45], pp. 140–169; [53]; [56]; [73]; [77]; [80]; [81]; and in Section 8 we shall introduce a technique of the maximum likelihood type. Some authors have also used methods involving the minimization of sums of absolute deviations [17], [28].

5.2. *Further assumptions and conventions.* We shall be working with a single, fixed value of $a$, and shall therefore suppress this subscript except where it may cause confusion.

In what follows, we shall disregard the fact that some of the $\lambda_d$ may be known to equal zero. The case where some $\lambda_d$ actually do equal zero needs only trivial notational modifications.

Let

$$g_d(\boldsymbol{\theta}) = g(\xi_d, \boldsymbol{\theta}), \tag{5.1}$$

$$\mathbf{g}(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), \cdots, g_D(\boldsymbol{\theta}))'. \tag{5.2}$$

(The prime denotes a transpose.)

ASSUMPTION 2. *We assume that $\boldsymbol{\theta}$ varies in an open subset $\Theta$ of the $G$-dimensional Euclidean space $\mathbf{R}_G$, where $G < D$. Let $\mathbf{g}$ be a one to one, bicontinuous, continuously differentiable mapping of $\Theta$ into*

$$(5.3) \qquad \Lambda_0 = \mathop{\times}_{d=1}^{D} \{x_d > 0\}.$$

*Define*

$$(5.4) \qquad \mathbf{J}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial}{\partial\theta_1} g_1(\boldsymbol{\theta}), \cdots, \frac{\partial}{\partial\theta_G} g_1(\boldsymbol{\theta}) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{\partial}{\partial\theta_1} g_D(\boldsymbol{\theta}), \cdots, \frac{\partial}{\partial\theta_G} g_D(\boldsymbol{\theta}) \end{pmatrix},$$

*and assume that $\mathbf{J}(\boldsymbol{\theta})$ has rank $G$ for each $\boldsymbol{\theta} \in \Theta$.*

We denote the true value of $\boldsymbol{\theta}$ by $\boldsymbol{\theta}^0$, and let

$$(5.5) \qquad \mathbf{J}_0 = \mathbf{J}(\boldsymbol{\theta}^0), \boldsymbol{\lambda}^0 = \mathbf{g}(\boldsymbol{\theta}^0), \quad \text{and} \quad L_d^0 = \lim_{N\to\infty} E_{\boldsymbol{\theta}^0} L(d)/N.$$

[Compare (3.8).] We also let

$$(5.6) \qquad \sigma_{d,0}^2 = \lambda_d^0/L_d^0 \quad \text{and} \quad \boldsymbol{\Sigma}^0 = \operatorname{diag}(\sigma_{1,0}^2, \cdots, \sigma_{D,0}^2)$$

[compare (3.12)], with the convention that we write $\mathbf{M} = \operatorname{diag}(m_1, \cdots, m_S)$ if $\mathbf{M}$ is a diagonal $S \times S$ matrix with the $m_s$ as diagonal elements.

Let us denote it by a right superscript $N$ if we want to stress that a quantity depends on $N$.

In Sections 3 and 4 we brought out some estimators $\hat{\boldsymbol{\lambda}}(N) = (\hat{\lambda}_1^{(N)}, \cdots, \hat{\lambda}_D^{(N)})'$ of the common occurrence/exposure type for the parameter $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_D)'$, and we stated some theorems concerning their asymptotic properties. In much of what follows, it is precisely these properties which are of interest, and not the form of the estimators themselves. In Sections 6 and 7, therefore, we shall take $\hat{\boldsymbol{\lambda}}^{(N)}$ to be any estimator for $\boldsymbol{\lambda}$, not necessarily the one given by (3.7), and we shall continuously make

ASSUMPTION 3.

$$(5.7) \qquad N^{1/2}(\hat{\boldsymbol{\lambda}}^{(N)}\mathscr{L}_{\boldsymbol{\theta}^0}\boldsymbol{\lambda}^0) \xrightarrow{\mathscr{L}_{\boldsymbol{\theta}^0}} \mathscr{N}(\mathbf{0}, \boldsymbol{\Sigma}_0),$$

*where $\mathscr{N}(\mathbf{0}, \boldsymbol{\Sigma}_0)$ is the multinormal distribution with mean $\mathbf{0}$ and a positive definite covariance matrix $\boldsymbol{\Sigma}_0$, which need not be the same as $\boldsymbol{\Sigma}^0$.*

## 6. Analytic graduation through minimization of a quadratic form

6.1. *The graduation method.* Let $\mathbf{M}$ be a positive definite, symmetric $D \times D$ matrix whose elements $m_{i,j}$ may (but need not) be random variables. Let

$$(6.1) \qquad Q(\boldsymbol{\theta}) = N(\hat{\boldsymbol{\lambda}} - \mathbf{g}(\boldsymbol{\theta}))'\mathbf{M}(\hat{\boldsymbol{\lambda}} - \mathbf{g}(\boldsymbol{\theta})).$$

Assume that there exists a $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$, which minimizes $Q(\boldsymbol{\theta})$. We shall then take $\hat{\boldsymbol{\theta}}$ to be our estimator for $\boldsymbol{\theta}$.

A whole class of graduation methods is generated by the various choices of the matrix $\mathbf{M}$. Thus if we take $\mathbf{M} = \mathbf{I}$, the identity matrix, we get

$$Q(\boldsymbol{\theta}) = N \sum_{d=1}^{D} (\hat{\lambda}_d - g_d(\boldsymbol{\theta}))^2, \tag{6.2}$$

and $\hat{\boldsymbol{\theta}}$ becomes a least squares estimator. An analogy with the modified minimum $\chi^2$ method results from setting

$$\mathbf{M} = \mathrm{diag}\,(1/\hat{\sigma}_1^2, \cdots, 1/\hat{\sigma}_D^2), \tag{6.3}$$

where the $\hat{\sigma}_d^2$ are given by (3.13). We then get

$$Q(\boldsymbol{\theta}) = \sum_{d=1}^{D} \{M(d) - L(d) g_d(\boldsymbol{\theta})\}^2 / M(d). \tag{6.4}$$

If, in particular, $\mathbf{g}(\boldsymbol{\theta})$ is a linear function of $\boldsymbol{\theta}$, say

$$\mathbf{g}(\boldsymbol{\theta}) = \mathbf{J}_0 \boldsymbol{\theta} + \mathbf{g}_0 \tag{6.5}$$

where $\mathbf{J}_0$ is a known $D \times G$ matrix of rank $G$, and $\mathbf{g}_0$ is a known $D \times 1$ vector, we get

$$\hat{\boldsymbol{\theta}} = (\mathbf{J}_0' \mathbf{M} \mathbf{J}_0)^{-1} \mathbf{J}_0' \mathbf{M} (\hat{\boldsymbol{\lambda}} - \mathbf{g}_0). \tag{6.6}$$

A particular case of (6.5) is given in (1.2).

6.2. *Asymptotic theory.* Let $\{\mathbf{M}^{(N)}\}$ be a sequence of positive definite, symmetric, possibly random, $D \times D$ matrices. For simplicity we assume that the $\mathbf{M}^{(N)}$ are not functions of $\boldsymbol{\theta}$. (This can be modified. Compare, for example, [14], Theorem 5.) Let $\hat{\boldsymbol{\theta}}^{(N)}$ be a value of $\boldsymbol{\theta}$. if any, which minimizes $Q(\boldsymbol{\theta})$ with $\mathbf{M} = \mathbf{M}^{(N)}$ and $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}^{(N)}$. We can then prove the following theorem by the methods of general asymptotic statistical theory. (See [51]. All the hard parts of the proof can be handled by the argument in [9].)

THEOREM 3. *Make Assumptions* 2 *and* 3, *and assume also that*

$$\mathrm{plim}\ \mathbf{M}^{(N)} = \mathbf{M}_0, \tag{6.7}$$

*where* $\mathbf{M}_0$ *is positive definite. With a probability increasing to* 1 *as* $N \to \infty$, *there then exists a value* $\hat{\boldsymbol{\theta}}^{(N)} \in \Theta$ *which minimizes* $Q(\boldsymbol{\theta})$, *and*

$$N^{1/2}(\hat{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta}^0) \xrightarrow{\mathscr{L}_{\theta^0}} \mathscr{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{6.8}$$

*where*

$$\boldsymbol{\Sigma} = (\mathbf{J}_0' \mathbf{M}_0 \mathbf{J}_0)^{-1} \mathbf{J}_0' \mathbf{M}_0 \boldsymbol{\Sigma}_0 \mathbf{M}_0 \mathbf{J}_0 (\mathbf{J}_0' \mathbf{M}_0 \mathbf{J}_0)^{-1} \tag{6.9}$$

*is positive definite.*

COROLLARY. $N^{1/2}\{\mathbf{g}(\hat{\boldsymbol{\theta}}^{(N)}) - \boldsymbol{\lambda}^0\} \xrightarrow{\mathscr{L}_{\theta^0}} \mathscr{N}(\mathbf{0}, \mathbf{J}_0' \boldsymbol{\Sigma} \mathbf{J}_0)$.

REMARK 1. If $\mathbf{M}_0 = \mathbf{\Sigma}_0^{-1}$, as is the case when we use (3.7) and (6.3), we get $\mathbf{\Sigma}$ equal to

$$\mathbf{\Sigma}_{0,0} = (\mathbf{J}_0'\mathbf{\Sigma}_0^{-1}\mathbf{J}_0)^{-1}. \tag{6.10}$$

REMARK 2. Since $G < D$, $\mathbf{J}_0'\mathbf{\Sigma}\mathbf{J}_0$ is singular.

REMARK 3. If we regard $\hat{\boldsymbol{\theta}}^{(N)}$ as a mapping from $\mathbf{R}_D$ to $\mathbf{R}_G$ (that is, a function of $\hat{\boldsymbol{\lambda}}$), we obviously have

$$\hat{\boldsymbol{\theta}}^{(N)}(\mathbf{g}(\boldsymbol{\theta})) = \boldsymbol{\theta} \qquad \text{for} \quad \boldsymbol{\theta} \in \Theta, \tag{6.11}$$

for any positive definite $\mathbf{M}^{(N)}$.

6.3. *The choice of* $\{\mathbf{M}^{(N)}\}$. Since different sequences $\{\mathbf{M}^{(N)}\}$ give rise to estimators $\{\hat{\boldsymbol{\theta}}^{(N)}\}$ which may have different asymptotic covariance matrices, one will want to know how to select a $\{\mathbf{M}^{(N)}\}$ so as to get a $\mathbf{\Sigma}$ which is as favorable as possible. Given two such matrices, $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, where $\mathbf{\Sigma}_2 - \mathbf{\Sigma}_1$ is positive semidefinite, we shall regard $\mathbf{\Sigma}_1$ as the more favorable, since each of the variances on its diagonal will be no greater than the corresponding variance on the diagonal of $\mathbf{\Sigma}_2$. At the same time, $\mathbf{J}_0'\mathbf{\Sigma}_1\mathbf{J}_0$ will be preferred to $\mathbf{J}_0'\mathbf{\Sigma}_2\mathbf{J}_0$ (compare the corollary to Theorem 3), since also $\mathbf{J}_0'(\mathbf{\Sigma}_2 - \mathbf{\Sigma}_1)\mathbf{J}_0$ will be positive semidefinite. The following theorem tells us that an $\{\mathbf{M}^{(N)}\}$ with $\mathbf{M}_0 = \mathbf{\Sigma}_0^{-1}$ will be optimal in this sense.

THEOREM 4. *Let* $\mathbf{\Sigma}$ *and* $\mathbf{\Sigma}_{0,0}$ *be given by* (6.9) *and* (6.10), *respectively, Then* $\mathbf{\Sigma} - \mathbf{\Sigma}_{0,0}$ *is positive semidefinite under the assumptions of Theorem* 3.

PROOF. (i) Let $\mathbf{A}$ be any $D \times G$ matrix of rank $G < D$. Then $\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ is idempotent, so all its characteristic roots equal 0 or 1. Thus $\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ has only 0 and 1 as characteristic roots, and this matrix, therefore, is positive semidefinite.

(ii) Let us then prove that $\mathbf{\Sigma}_0 - \mathbf{J}_0(\mathbf{J}_0'\mathbf{\Sigma}_0^{-1}\mathbf{J}_0)^{-1}\mathbf{J}_0'$ is positive semidefinite. Let $\mathbf{B}$ be a nonsingular matrix such that $\mathbf{B}'\mathbf{\Sigma}_0\mathbf{B} = \mathbf{I}$. Let $\mathbf{v}$ be an arbitrary $D \times 1$ vector, and let $\mathbf{w} = \mathbf{B}^{-1}\mathbf{v}$. Then

$$\begin{aligned} \mathbf{v}'\{\mathbf{\Sigma}_0 - \mathbf{J}_0(\mathbf{J}_0'\mathbf{\Sigma}_0^{-1}\mathbf{J}_0)^{-1}\mathbf{J}_0'\}\mathbf{v} &= \mathbf{w}'\{\mathbf{I} - \mathbf{B}'\mathbf{J}_0(\mathbf{J}_0'\mathbf{B}\mathbf{B}'\mathbf{J}_0)^{-1}\mathbf{J}_0'\mathbf{B}\}\mathbf{w} \\ &= \mathbf{w}'\{\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\}\mathbf{w}, \end{aligned} \tag{6.12}$$

with $\mathbf{A} = \mathbf{B}'\mathbf{J}_0$. Our assertion then follows from step (i) above.

(iii) Finally, let $\mathbf{v}$ be as above, and let

$$\mathbf{w} = \mathbf{M}_0\mathbf{J}_0(\mathbf{J}_0'\mathbf{M}_0\mathbf{J}_0)^{-1}\mathbf{v}. \tag{6.13}$$

Then $\mathbf{J}_0'\mathbf{w} = \mathbf{v}$ and so

$$\mathbf{v}'\{\mathbf{\Sigma} - \mathbf{\Sigma}_{0,0}\}\mathbf{v} = \mathbf{w}'\{\mathbf{\Sigma}_0 - \mathbf{J}_0\mathbf{\Sigma}_{0,0}\mathbf{J}_0'\}\mathbf{w} \geqq 0 \tag{6.14}$$

by step (ii) above. Thus $\mathbf{\Sigma} - \mathbf{\Sigma}_{0,0}$ is positive semidefinite. *Q.E.D.*

6.4. *The choice of* $\{\hat{\boldsymbol{\lambda}}^{(N)}\}$. In Sections 6.1 to 6.3 above, we have focused on a single estimator $\{\hat{\boldsymbol{\lambda}}^{(N)}\}$. Assume now that two such sequences are proposed, say $\{\hat{\boldsymbol{\lambda}}_1^{(N)}\}$ and $\{\hat{\boldsymbol{\lambda}}_2^{(N)}\}$, both satisfying the assumptions of Theorem 3, with asymptotic covariance matrices $\mathbf{\Sigma}_1/N$ and $\mathbf{\Sigma}_2/N$, respectively. Say that $\mathbf{\Sigma}_2 - \mathbf{\Sigma}_1$ is positive semidefinite. Intuitively one would expect $\{\hat{\theta}_1^{(N)}\}$ to have a more favorable asymptotic covariance matrix than $\{\hat{\theta}_2^{(N)}\}$, when $\{\hat{\theta}_i^{(N)}\}$, for $i = 1, 2$, is produced from $\{\hat{\lambda}_i^{(N)}\}$ by the method of Section 6.1 with a choice of $\mathbf{M}_i^{(N)}$ which is optimal according to Theorem 4. This turns out to be correct.

THEOREM 5. *If $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are positive definite and $\mathbf{\Sigma}_2 - \mathbf{\Sigma}_1$ is positive semidefinite, then $\mathbf{\Sigma}_{0,2} - \mathbf{\Sigma}_{0,1}$ is positive semidefinite, where*

$$\mathbf{\Sigma}_{0,i} = (\mathbf{J}_0'\mathbf{\Sigma}_i^{-1}\mathbf{J}_0)^{-1} \tag{6.15}$$

*for $i = 1, 2$. Here $\mathbf{J}_0$ is any $D \times G$ matrix of rank $G < D$.*

PROOF. If $\mathbf{A}$ and $\mathbf{B}$ are positive definite $D \times D$ matrices with positive semidefinite $\mathbf{A} - \mathbf{B}$, then $\mathbf{B}^{-1} - \mathbf{A}^{-1}$ will also be positive semidefinite [26], page 55, Theorem 2.5. From this the theorem easily follows. *Q.E.D.*

## 7. Moment methods

7.1. *The graduation method.* A moment method estimator $\tilde{\theta}^{(N)}$ of $\theta$ is defined as a solution of the system of equations

$$\sum_{d=1}^{D} \xi_d^r\{\hat{\lambda}_d^{(N)} - g_d(\hat{\theta}^{(N)})\} = 0 \qquad \text{for} \quad r = 0, 1, \cdots, G-1, \tag{7.1}$$

if it exists. Let

$$\mathbf{M} = \begin{pmatrix} 1, & 1, & \cdots, & 1 \\ \xi_1, & \xi_2, & \cdots, & \xi_D \\ \xi_1^2, & \xi_2^2, & \cdots, & \xi_D^2 \\ \cdot\ \cdot & \cdot\ \cdot & \cdot\ \cdot & \cdot\ \cdot \\ \xi_1^{G-1}, & \xi_2^{G-1}, & \cdots, & \xi_D^{G-1} \end{pmatrix}. \tag{7.2}$$

Then (7.1) can be rewritten as

$$\mathbf{M}\{\hat{\lambda}^{(N)} - \mathbf{g}(\tilde{\theta}^{(N)})\} = 0. \tag{7.3}$$

We shall extend this definition, and shall call $\tilde{\theta}^{(N)}$ a *generalized moment method estimator* for $\theta$ if it is a solution of (7.3), where $\mathbf{M}$ here can be any $G \times D$ matrix, that is, $\mathbf{M}$ need not be given by (7.2).

To give an example of an estimator generated by (7.3) but not satisfying (7.1), we shall consider the King-Hardy method of estimating the three parameters, $\alpha$, $\beta$, and $c$, of the Gompertz-Makeham function in (1.1). Say that we can take $[0, \zeta\rangle$ to be the age interval $[x_0, x_0 + 3h\rangle$ for some integer $h$, and that $\zeta_x = x_0 + x$ for $x = 0, 1, \cdots, 3h$, so that we have one year age intervals. Then the King-Hardy estimators are the solution $(\tilde{\alpha}, \tilde{\beta}, \tilde{c})$ of the equations

$$\sum_{x=x_0+(k-1)h}^{x_0+kh-1} (\tilde{\alpha} + \tilde{\beta}\tilde{c}^x) = \tilde{H}_k \qquad \text{for} \quad k = 1, 2, 3, \tag{7.4}$$

where

$$\tilde{H}_k = \sum_{x=x_0+(k-1)h}^{x_0+kh-1} \hat{\lambda}_x.$$

We get [59], p. 167,

$$\tilde{c}^h = (\tilde{H}_3 - \tilde{H}_2)/(\tilde{H}_2 - \tilde{H}_1), \tag{7.5}$$

and similar formulas for $\tilde{\alpha}$ and $\tilde{\beta}$. If we let $\mathbf{m}_x$ be a $1 \times h$ vector where all elements equal $x$, and let

$$\mathbf{M} = \begin{pmatrix} \mathbf{m}_1 & \mathbf{m}_0 & \mathbf{m}_0 \\ \mathbf{m}_0 & \mathbf{m}_1 & \mathbf{m}_0 \\ \mathbf{m}_0 & \mathbf{m}_0 & \mathbf{m}_1 \end{pmatrix}, \tag{7.6}$$

then (7.3) reduces to (7.4) in the case where

$$g_x(\alpha, \beta, c) = \alpha + \beta c^{(x_0+x)}. \tag{7.7}$$

In applications to analytic graduation, the matrix $\mathbf{M}$ is usually nonrandom and not a function of $N$ or $\boldsymbol{\theta}$. For simplicity we shall only study this case, but generalization to possibly random $\mathbf{M}$, possibly depending on $N$ and $\boldsymbol{\theta}$, can be made by standard methods [24], [78].

If, in particular, $\mathbf{g}(\boldsymbol{\theta})$ is given by (6.5), we get

$$\tilde{\boldsymbol{\theta}}^{(N)} = (\mathbf{M}\mathbf{J}_0)^{-1}\mathbf{M}(\hat{\boldsymbol{\lambda}}^{(N)} - \mathbf{g}_0), \tag{7.8}$$

provided $\mathbf{M}\mathbf{J}_0$ is nonsingular.

7.2. *Asymptotic theory.*

THEOREM 6. *Make Assumptions* 2 *and* 3, *and assume also that* $\mathbf{MJ}(\theta)$ *is nonsingular for any* $\boldsymbol{\theta} \in \Theta$. *There then exists a neighborhood* $\Omega$ *of* $\mathbf{g}(\Theta)$ *and a one to one mapping* $\tilde{\boldsymbol{\theta}}^{(N)}$ *from* $\mathbf{R}_D$ *to* $\mathbf{R}_G$, *continuous in* $\Omega$, *such that*

$$\tilde{\boldsymbol{\theta}}^{(N)}(\mathbf{g}(\boldsymbol{\theta})) = \boldsymbol{\theta} \tag{7.9}$$

*for* $\boldsymbol{\theta} \in \Theta$ *and* (7.3) *holds for all* $\hat{\boldsymbol{\lambda}}^{(N)} \in \Omega$. *Furthermore,*

$$N^{1/2}(\tilde{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta}^0) \xrightarrow{\mathscr{L}_{\theta^0}} \mathscr{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{7.10}$$

*where*

$$\boldsymbol{\Sigma} = (\mathbf{M}\mathbf{J}_0)^{-1}\mathbf{M}\boldsymbol{\Sigma}_0\{(\mathbf{M}\mathbf{J}_0)^{-1}\mathbf{M}\}'. \tag{7.11}$$

$\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{0,0}$ is positive semidefinite. ($\boldsymbol{\Sigma}_{0,0}$ is given in (6.10).)

PROOF. By Theorem 1 in [24], p. 1054, we need only prove the final assertion above. Let $\mathbf{v}$ be an arbitrary $G \times 1$ vector, and let $\mathbf{w} = (\mathbf{J}_0'\mathbf{M}')^{-1}\mathbf{v}$. Then

$$\mathbf{v}'\{\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{0,0}\}\mathbf{v} = (\mathbf{M}\mathbf{w})'\{\boldsymbol{\Sigma}_0 - \mathbf{J}_0\boldsymbol{\Sigma}_{0,0}\mathbf{J}_0'\}(\mathbf{M}\mathbf{w}) \geqq 0 \tag{7.12}$$

by step (ii) of the proof of Theorem 4. *Q.E.D.*

REMARK 4. By the final assertion of the theorem, the generalized moment method can never give a more favorable asymptotic covariance matrix for the estimatior of $\boldsymbol{\theta}$ than the corresponding "optimal" estimator found in Section 6.

REMARK 5. Since $\hat{\boldsymbol{\lambda}}^{(N)}$ is $N^{1/2}$-consistent for $\boldsymbol{\lambda}^0$,

$$P\{\hat{\boldsymbol{\lambda}}^{(N)} \in \Omega\} \to 1 \qquad \text{as} \quad N \to \infty. \tag{7.13}$$

REMARK 6. The analogues of Theorem 5 and the corollary to Theorem 3 hold in the present situation.

When the generalized moment method is applied to a particular case, it is frequently modified to suit the characteristics of the situation in hand. We shall give examples of this in Sections 7.3 and 7.4.

7.3. *Modifications, Example* 1: *Gompertz-Makeham graduation.* In mortality studies using the Gompertz-Makeham formula (1.1), one will frequently find that $c$ is estimated by (7.5), but that estimators for $\alpha$ and $\beta$ are subsequently found by some other method, for instance by minimizing

$$\sum_{x=x_0}^{x_0+3h-1} (\hat{\lambda}_x^{(N)} - \alpha - \beta \tilde{c}^x)^2 \tag{7.14}$$

[83], p. 225. Let us consider a slightly more general case, and let us estimate $\alpha$ and $\beta$ by minimizing

$$Q(\alpha, \beta) = N(\hat{\boldsymbol{\lambda}}^{(N)} - \alpha \mathbf{e} - \beta \tilde{\boldsymbol{\psi}}^{(N)})' \mathbf{M}^{(N)} (\hat{\boldsymbol{\lambda}}^{(N)} - \alpha \mathbf{e} - \beta \tilde{\boldsymbol{\psi}}^{(N)}). \tag{7.15}$$

Here $\{\mathbf{M}^{(N)}\}$ is a sequence of matrices of the kind studied in Section 6, $\mathbf{e}$ is a $3h \times 1$ vector where all elements equal 1, and

$$\tilde{\boldsymbol{\psi}}^{(N)} = (\tilde{c}^{x_0}, \tilde{c}^{x_0+1}, \cdots, \tilde{c}^{x_0+3h-1})'. \tag{7.16}$$

Assuming that $\mathbf{M}^{(N)}$ is positive definite, and letting

$$\mathbf{K}^{(N)} = (\mathbf{e}, \tilde{\boldsymbol{\psi}}^{(N)}), \tag{7.17}$$

we get the estimators

$$\begin{pmatrix} \hat{\hat{\alpha}} \\ \hat{\hat{\beta}} \end{pmatrix} = (\mathbf{K}^{(N)\prime} \mathbf{M}^{(N)} \mathbf{K}^{(N)})^{-1} \mathbf{K}^{(N)\prime} \mathbf{M}^{(N)} \hat{\boldsymbol{\lambda}}^{(N)}. \tag{7.18}$$

Now let $\mathbf{m}_x$ be defined as below (7.5), let $\alpha_0$, $\beta_0$, and $c_0$ be the true values of the parameters, let

$$\begin{aligned} \boldsymbol{\psi}_0 &= (c_0^{x_0}, \cdots, c_0^{x_0+3h-1})', \qquad \mathbf{K}_0 = (\mathbf{e}, \boldsymbol{\psi}_0), \\ \gamma &= (c_0 - 1)/\{h\beta_0 c_0^{x_0+h-1}(c_0^h - 1)^2\}, \end{aligned} \tag{7.19}$$

and

$$\boldsymbol{\Phi} = \begin{pmatrix} (\mathbf{K}_0' \mathbf{M}_0 \mathbf{K}_0)^{-1} \mathbf{K}_0 \mathbf{M}_0 \\ (\mathbf{m}_0, \mathbf{m}_{-\gamma}, \mathbf{m}_\gamma) \end{pmatrix}. \tag{7.20}$$

We then get

THEOREM 7. *Let* $p \lim \mathbf{M}^{(N)} = \mathbf{M}_0$, *where* $\mathbf{M}_0$ *is positive definite, and make Assumption* 3. *Then*

$$N^{1/2}\{(\hat{\alpha}, \hat{\beta}, \tilde{c})' - (\alpha_0, \beta_0, c_0)'\} \xrightarrow{\mathscr{L}_{(\alpha_0,\beta_0,c_0)}} \mathscr{N}(\mathbf{0}, \mathbf{\Phi\Sigma}_0\mathbf{\Phi}'). \tag{7.21}$$

Stevens [68], Patterson [57], Lipton and McGilchrist [52], and others in their references have studied the estimation of the Gompertz-Makeham parameters in a regression model. Stevens [68] found that King-Hardy's method may be very inefficient there. The estimators developed for the regression model can also be used for purposes of analytic graduation, and it would be interesting to see an investigation of their merits in that context.

7.4. *Modifications, Example* 2: *Hadwiger graduation.* Consider now the problem of graduating a set $\{\hat{\lambda}_x : x = \alpha, \alpha + 1, \cdots, \beta - 1\}$ of female fertility rates calculated for single year age groups by fitting the Hadwiger function (1.3) to the rates. If we regard $h_x$ as a function of a continuous $x$, and define

$$R'_k(R, T, H, d) = \int_d^\infty x^k h_x(R, T, H, d)\, dx, \tag{7.22}$$

then

$$R'_0(R, T, H, d) = R,\ R'_1(R, T, H, d) = R(T + d), \tag{7.23}$$

and the formulas for $R'_k$ for $k \geqq 2$ can be found from the fact that the corresponding cumulants for $d = 0$ are

$$\kappa_k = (1)(3)\cdots(2k - 3)2^{k-1}H^{2k-2}/T^{3k-4} \qquad \text{for} \quad k \geqq 2. \tag{7.24}$$

(Compare [45], pp. 150, 151, 160.) No such nice formulas are known for the discrete case, that is, for

$$R_k(R, T, H, d) = \sum_{x=\alpha}^{\beta-1} x^k h_x(R, T, H, d), \tag{7.25}$$

where only integer values of $x$ are used in the summation. Rather than attempting cumbersome calculations with the $R_k$, and acting on the analogy between the $R_k$ and the $R'_k$, Yntema [28], [81] has suggested an estimation procedure which amounts to the following: regard $h_x$ as a function of a *continuous* $x$. Let $U = T + d$.

Then

$$h_U(R, T, H, d) = \frac{RH}{T\sqrt{\pi}}, \tag{7.26}$$

and the mode of the function is

$$M = d + 3T\{(1 + 16H^4/9)^{1/2} - 1\}/(4H^2). \tag{7.27}$$

One easily sees that $M < U$. Solve (7.27) with respect to $H$, introduce the result into (7.26), let

$$a = \tfrac{4}{3}\pi(Th_U/R)^2, \qquad b = (M - d)/T, \tag{7.28}$$

and get

$$b = \{(1 + a^2)^{1/2} - 1\}/a. \tag{7.29}$$

For the range of values in which $a$ will usually lie, the right side here is approximately equal to $1 - a^{-1}$. Solving $b \approx 1 - a^{-1}$ with respect to $T$ after substituting $U - T$ for $d$, we get

$$T \approx R^2/\{\tfrac{4}{3}\pi(U - M)h_U^2\}. \tag{7.30}$$

Now introduce the estimators as follows: Let

$$\hat{R} = \sum_{x=\alpha}^{\beta-1} \hat{\lambda}_x, \qquad \hat{U} = \sum_{x=\alpha}^{\beta-1} x\hat{\lambda}_x/\hat{R}. \tag{7.31}$$

(Compare (7.23).) If $[y]$ denotes the integer value of $y$, let

$$V = [\hat{U} + \tfrac{1}{2}], \qquad \hat{h} = \hat{\lambda}_V, \qquad \hat{M} = \min\{x: \hat{\lambda}_x \geqq \hat{\lambda}_y \text{ for all } y\} \tag{7.32}$$

Finally, let

$$\hat{T} = \hat{R}^2/\{\tfrac{4}{3}\pi(\hat{U} - \hat{M})\hat{h}^2\}, \qquad \hat{d} = \hat{U} - \hat{T}, \tag{7.33}$$

[compare (7.30)], and let

$$\hat{H} = \hat{h}\hat{T}\sqrt{\pi}/\hat{R} = \tfrac{3}{4}\hat{R}/\{\sqrt{\pi}\hat{h}(\hat{U} - \hat{M})\}. \tag{7.34}$$

(Compare (7.26).) Then $\hat{\boldsymbol{\theta}} = (\hat{R}, \hat{T}, \hat{H}, \hat{d})$ is an estimator for $\boldsymbol{\theta} = (R, T, H, d)$. To study its asymptotic properties, we let $\boldsymbol{\theta}^0 = (R^0, T^0, H^0, d^0)$ be the true value of $\boldsymbol{\theta}$, and introduce

$$\begin{aligned}
&R_0 = R_0(\boldsymbol{\theta}^0), \qquad U_0 = R_1(\boldsymbol{\theta}^0)/R_0, \qquad V_0 = [U_0 + \tfrac{1}{2}], \\
&\qquad\qquad h_0 = h_{V_0}(\boldsymbol{\theta}^0), \\
&M_0 = \min\{x \in \{\alpha, \alpha + 1, \cdots, \beta - 1\}: h_x(\boldsymbol{\theta}^0) \geqq h_y(\theta^0) \\
&\qquad\qquad \text{for all } y \in \{\alpha, \alpha + 1, \cdots, \beta - 1\}\}, \\
&T_0 = R_0^2/\{\tfrac{4}{3}\pi(U_0 - M_0)h_0^2\}, \qquad d_0 = U_0 - T_0, \\
&\qquad\qquad H_0 = h_0T_0\sqrt{\pi}/R_0.
\end{aligned} \tag{7.35}$$

Let $\mathbf{e}$ be a $(\beta - \alpha) \times 1$ vector where all elements equal 1, let

$$\boldsymbol{\psi} = R_0^{-1}(\alpha - T_0, \alpha + 1 - T_0, \cdots, \beta - 1 - T_0)', \tag{7.36}$$

and let

$$\boldsymbol{\Phi} = (\mathbf{e}, \mathbf{e}T_0/R_0, \mathbf{e}T_0h_0\sqrt{\pi}/R_0^2, \boldsymbol{\psi}). \tag{7.37}$$

We then have

THEOREM 8. *Under Assumption* 3,

$$(7.38)\qquad N^{1/2}\{(\hat{R}, \hat{T}, \hat{H}, \hat{d})' - (R_0, H_0, T_0, d_0)'\} \xrightarrow{\mathscr{L}_{\theta^0}} \mathscr{N}(\mathbf{0}, \mathbf{\Phi}'\Sigma_0\mathbf{\Phi}).$$

REMARK 7. Note that $\mathbf{\Phi}'\Sigma_0\mathbf{\Phi}$ is singular.

REMARK 8. Note also that $(R_0, T_0, H_0, d_0)$ is not the true value of the parameters here. No one seems to have looked into the difference $(R^0, T^0, H^0, d^0)$ minus $(R_0, T_0, H_0, d_0)$ in any detail.

## 8. A maximum likelihood method

In Sections 6 and 7, the estimators for $\boldsymbol{\theta}$ appear as functions of the "raw" estimator $\hat{\boldsymbol{\lambda}}$ for $\boldsymbol{\lambda}$. If one may really assume that $\boldsymbol{\lambda}_a = \mathbf{g}_a(\boldsymbol{\theta}_a)$ for $a = 1, 2, \cdots, A$, different approaches may be at least as efficient, however. One obvious possibility is to enter the $\mathbf{g}_a(\boldsymbol{\theta}_a)$ into the likelihood function and maximize with respect to the $\boldsymbol{\theta}_a$. In the situation of Section 3.4, this will amount to maximizing $\Sigma_a \hat{\eta}_a(\boldsymbol{\theta}_a)$, where

$$(8.1)\qquad \hat{\eta}_a(\boldsymbol{\theta}_a) = \sum_d M(a, d) \log g_{a,d}(\boldsymbol{\theta}_a) - \sum_d L(a, d) g_{a,d}(\boldsymbol{\theta}_a).$$

For simplicity, we shall assume that $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_A$ are functionally independent, so that we can maximize the likelihood function (if at all) by maximizing each $\hat{\eta}_a$ separately.

In the situation of Section 4.3, the log likelihood function is of a different form, but we shall still construct an estimator $\boldsymbol{\theta}_a^*$ for $\boldsymbol{\theta}_a$ by maximizing $\hat{\eta}_a$.

The following theorem holds.

THEOREM 9. *Fix $a \in \{1, 2, \cdots, A\}$ and let the $M(a, d)$ and $L(a, d)$ be given as in Section 3.4 or 4.3. Assume that $P\{L(a, d) > 0\} \to 1$ as $N \to \infty$ for all $d$ where $(a, d) \in \mathscr{G}$. Make Assumptions 1 and 2.*

*With a probability increasing to 1 as $N \to \infty$, there will then exist a value $\boldsymbol{\theta}_a^{*(N)} \in \Theta_a$ which maximizes $\hat{\eta}_a(\boldsymbol{\theta}_a)$, and*

$$(8.2)\qquad N^{1/2}(\boldsymbol{\theta}_a^{*(N)} - \boldsymbol{\theta}_a^0) \xrightarrow{\mathscr{L}_{\theta^0}} \mathscr{N}(\mathbf{0}, \Sigma_{0,0}),$$

*where $\Sigma_{0,0}$ is given by (6.10), provided there exist constants $k_a$ and $k_a'$ such that*

$$(8.3)\qquad k_a' \geqq g_{a,d}(\boldsymbol{\theta}_a) \geqq k_a > 0 \qquad \text{for all} \quad \boldsymbol{\theta}_a \in \Theta_a.$$

PROOF. (i) *Preliminaries.* Suppress the subscript $a$ and fix the true value $\boldsymbol{\theta}^0$. Let

$$(8.4)\qquad \ell_d = \ell_d^{(N)} = E_{\theta^0} L(d), \qquad L_d = \lim_{N\to\infty} \ell_d^{(N)}/N,$$

and note that [40], (10),

$$(8.5)\qquad E_{\theta^0} M(d) = \lambda_d^0 \ell_d.$$

Let

$$(8.6)\qquad \hat{Q}(\boldsymbol{\lambda}) = \sum_d M(d) \log \lambda_d - \sum_d L(d) \lambda_d,$$

so that

$$\hat{\eta}(\boldsymbol{\theta}) = \hat{Q}(\mathbf{g}(\boldsymbol{\theta})), \tag{8.7}$$

and let

$$Q(\boldsymbol{\lambda}) = E_{\boldsymbol{\theta}^0}\hat{Q}(\boldsymbol{\lambda}) = \sum_d \ell_d\{\lambda_d^0 \log \lambda_d - \lambda_d\}, \tag{8.8}$$

and

$$\eta(\boldsymbol{\theta}) = Q(\mathbf{g}(\boldsymbol{\theta})) = E_{\boldsymbol{\theta}^0}\hat{\eta}(\boldsymbol{\theta}). \tag{8.9}$$

Finally, let

$$\Delta(\boldsymbol{\lambda}) = Q(\boldsymbol{\lambda}^0) - Q(\boldsymbol{\lambda}) = \sum_d \ell_d\{\lambda_d^0(\log \lambda_d^0 - \log \lambda_d) - (\lambda_d^0 - \lambda_d)\}. \tag{8.10}$$

For large enough $N$, each $\ell_d$ will be positive. For such $N$, we will have $\Delta(\boldsymbol{\lambda}) > 0$ for all $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}^0$, and $\Delta(\boldsymbol{\lambda})$ will strictly increase as each $|\lambda_d - \lambda_d^0|$ increases. For every $\varepsilon > 0$ there then exists a $\delta_1(\varepsilon)$ such that if $\Delta(\mathbf{g}(\boldsymbol{\theta})) \leqq \varepsilon$ then $|\mathbf{g}(\boldsymbol{\theta}) - \mathbf{g}(\boldsymbol{\theta}^0)| \leqq \delta_1(\varepsilon)$, and by the bicontinuity of $\mathbf{g}$ there further exists a $\delta(\varepsilon)$ such that $|\boldsymbol{\theta} - \boldsymbol{\theta}^0| \leqq \delta(\varepsilon)$. Conversely there exists a $\delta_0(\varepsilon)$ such that $\Delta(\mathbf{g}(\boldsymbol{\theta})) \leqq \delta_0(\varepsilon)$ if $|\boldsymbol{\theta} - \boldsymbol{\theta}^0| \leqq \varepsilon$. Let

$$S_\varepsilon = \{\boldsymbol{\theta} \in \Theta : |\boldsymbol{\theta} - \boldsymbol{\theta}^0| \leqq \varepsilon\}, \tag{8.11}$$

and choose $\varepsilon$. Choose $\varepsilon' > 0$, and let $\varepsilon''$ be so small that $\varepsilon'' \leqq \delta_0(\varepsilon)$, and $\delta(2\varepsilon'') \leqq \varepsilon'$, and

$$0 < 2\varepsilon'' < \eta(\boldsymbol{\theta}^0) - \inf\{\eta(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}, \tag{8.12}$$

and let

$$\Theta_{\varepsilon''} = \{\boldsymbol{\theta} \in \Theta : \Delta(\mathbf{g}(\boldsymbol{\theta})) \leqq 2\varepsilon''\}. \tag{8.13}$$

(ii) *Existence of* $\boldsymbol{\theta}^*$. Let

$$\gamma = \sup \sum_d \{|\log g_d(\boldsymbol{\theta})| + g_d(\boldsymbol{\theta})\}. \tag{8.14}$$

By (8.3), $\gamma < \infty$. Let $A_{\varepsilon''}^{(N)}$ be the event that

$$|N^{-1}M(d) - \lambda_d^0 \ell_d| < \varepsilon''/\gamma, \qquad |N^{-1}L(d) - \ell_d| < \varepsilon''/\gamma. \tag{8.15}$$

Then $P_{\boldsymbol{\theta}^0}(A_{\varepsilon''}^{(N)}) \to 1$. Assume that (8.15) holds. Then

$$|\hat{\eta}(\boldsymbol{\theta}) - \eta(\boldsymbol{\theta})| < \varepsilon'' \qquad \text{for all} \quad \boldsymbol{\theta} \in \Theta. \tag{8.16}$$

If $\boldsymbol{\theta} \in \Theta - \Theta_{\varepsilon''}$, we therefore get

$$\hat{\eta}(\boldsymbol{\theta}) < \eta(\boldsymbol{\theta}) + \varepsilon'' < \eta(\boldsymbol{\theta}^0) - \varepsilon'' < \hat{\eta}(\boldsymbol{\theta}^0), \tag{8.17}$$

so in maximizing $\hat{\eta}(\boldsymbol{\theta})$ we need not take such $\boldsymbol{\theta}$ into account. Since $\Theta_{\varepsilon''}$ is closed and $\hat{\eta}(\cdot)$ is continuous, there exists a maximizing value $\boldsymbol{\theta}^* \in \Theta_{\varepsilon''}$.

(iii) *Consistency of* $\boldsymbol{\theta}^*$. By the definition of $\Theta_{\varepsilon''}$, we have $\Delta(\mathbf{g}(\boldsymbol{\theta}^*)) \leqq 2\varepsilon''$.

Thus $|\boldsymbol{\theta}^* - \boldsymbol{\theta}^0| \leqq \delta(2\varepsilon'') \leqq \varepsilon'$, and the consistency of $\boldsymbol{\theta}^*$ follows.

The theorem now follows from some general results due to LeCam [51]. *Q.E.D.*

REMARK 9. Note that $\boldsymbol{\Sigma}_{0,0}$ is the most favorable asymptotic covariance matrix we can get by the procedures in Section 6 when $\hat{\boldsymbol{\lambda}}$ is given by (3.7). In this sense, therefore, the method of the present section is at least as good as any of the other general methods we have studied.

## 9. The choice of a graduating function

9.1. In previous sections, it was presupposed that the applicability of a particular graduating function $\mathbf{g}(\boldsymbol{\theta})$ had been established, and the problem was to estimate $\boldsymbol{\theta}$. In many practical cases, the situation will be different. Instead of a single function $\mathbf{g}$, there is often a finite family $\mathscr{F} = \{\mathbf{g}(s, \boldsymbol{\theta}): s = 1, 2, \cdots, S\}$ of candidates for a graduating function, and one is required to choose one of these on the basis of the data. In the case of human fertility rates, for instance, it is seldom given which function to use, and one may have to select one from among the Pearson family, the Hadwiger function, and the Brass polynomial (1.4), say.

We shall assume that all functions $\mathbf{g}(s, \boldsymbol{\theta})$ have the same parameter space $\Theta$. This need not be the case originally, but it can be achieved by the introduction of dummy parameters if necessary.

9.2. To describe what it means to choose a function from the class $\mathscr{F}$ "on the basis of the data," we shall assume that there is a member $\mathbf{g}(s^0, \cdot)$ of $\mathscr{F}$ which is the "true" graduating function. The choice of a member of $\mathscr{F}$ then amounts to estimating $s^0$ as an extra parameter. A number of estimation procedures are in use (compare, for example, [45], Section 6.5), but their statistical properties do not seem to have been much investigated, except that one may know something about their consistency as $N \to \infty$. We list some of these procedures.

(i) For choosing among the members of the Pearson family, there exist standard methods [22], [58], [34], [13] based on the first four empirical moments. Keyfitz, [45], p. 160, suggests that this type of criterion can also be used when the Hadwiger function (1.3) is included in $\mathscr{F}$ along with the Pearson type functions.

(ii) In connection with the methods of Section 6, an obvious procedure is to set

$$\hat{Q}_s(\boldsymbol{\theta}) = N(\hat{\boldsymbol{\lambda}} - \mathbf{g}(s, \boldsymbol{\theta}))'\mathbf{M}(\hat{\boldsymbol{\lambda}} - \mathbf{g}(s, \hat{\boldsymbol{\theta}})), \tag{9.1}$$

let $\hat{\boldsymbol{\theta}}^{(s)}$ be a value of $\boldsymbol{\theta}$ which minimizes $\hat{Q}_s(\boldsymbol{\theta})$ for $s = 1, 2, \cdots, S$, and define $\hat{s}$ as the value of $s$ that subsequently minimizes $\hat{Q}_s(\hat{\boldsymbol{\theta}}^{(s)})$ [27], [73], [80], [81].

(iii) A similar criterion can be used in connection with the method of Section 8. Let

$$\hat{\eta}(s, \boldsymbol{\theta}) = \sum_d \{M(d) \log g_d(s, \hat{\boldsymbol{\theta}}) - L(d) g_d(s, \boldsymbol{\theta})\}, \tag{9.2}$$

let $\boldsymbol{\theta}^{*(s)}$ maximize this quantity, and let $\hat{s}$ be the value of $s$ that subsequently maximizes $\hat{\eta}(s, \boldsymbol{\theta}^{*(s)})$.

(iv) Yntema [28], [81] has suggested calculating

$$\Delta_s = \sum_{d=1}^{D} |\hat{\lambda}_d - g_d(s, \hat{\boldsymbol{\theta}}^{(s)})| \tag{9.3}$$

and

$$\Delta_s' = \max \{|\hat{\lambda}_d - g_d(s, \hat{\boldsymbol{\theta}}^{(s)})| : d = 1, 2, \cdots, D\}, \tag{9.4}$$

and taking $\hat{s}$ as the $s$-value that minimizes $\Delta_s$ or $\Delta_s'$. Here $\hat{\boldsymbol{\theta}}^{(s)}$ is any suitable estimator for $\boldsymbol{\theta}$ based on $\mathbf{g}(s, \boldsymbol{\theta})$.

9.3. We shall take a look at the consistency properties of $\hat{s}$ as defined in Sections 9.2 (ii) and (iii) above.

By a proper specification of $\mathscr{F}$ and $\Theta$ we should be able to get $\mathbf{g}(s', \Theta) \cap \mathbf{g}(s'', \Theta)$ to be empty whenever $s' \neq s''$. (Otherwise, part of the values $\mathbf{g}(s'', \boldsymbol{\theta})$ would be redundant). To prove consistency, however, we need the stronger assumption that

$$|\mathbf{g}(s', \Theta) - \mathbf{g}(s'', \Theta)| > 0 \qquad \text{for} \quad s' \neq s'', \tag{9.5}$$

where $|A - B| = \inf \{|a - b| : a \in A, b \in B\}$ denotes the Euclidean distance between two subsets $A$ and $B$ of $\mathbf{R}_D$.

*If* (9.5) *holds, if* $\hat{\boldsymbol{\lambda}}$ *is consistent for* $\boldsymbol{\lambda}^0$, *and if* plim $\mathbf{M} = \mathbf{M}_0$ *as* $N \to \infty$, *with* $\mathbf{M}_0$ *positive definite, then* $\hat{s}$ *is consistent in Section* 9.2 (ii) *above.*

Similarly, by step (i) in the proof of Theorem 9, $\hat{s}$ *is consistent in Section* 9.2 (iii) *above* when (9.5) holds.

9.4. Let $\{\hat{\boldsymbol{\theta}}^{(s)}\}$ be some estimator which we would use for $\boldsymbol{\theta}$ if it were known that $s^0 = s$, and assume that

$$P_{s^0, \boldsymbol{\theta}^0}\{N^{1/2}(\hat{\boldsymbol{\theta}}^{(s^0)} - \boldsymbol{\theta}^0) \in B\} \to \Phi(B) \text{ as } N \to \infty, \tag{9.6}$$

where $\Phi$ is a limiting probability measure and $B$ is any $\Phi$-continuous measurable set. Let $\hat{s}$ be a consistent estimator for $s^0$. Then it is easy to show that

$$P_{s^0, \boldsymbol{\theta}^0}\{N^{1/2}(\hat{\boldsymbol{\theta}}^{(\hat{s})} - \boldsymbol{\theta}^0) \in B\} \to \Phi(B) \text{ as } N \to \infty \tag{9.7}$$

for the same $B$. Of course, $\hat{\boldsymbol{\theta}}^{(\hat{s})}$ is our estimator for $\boldsymbol{\theta}$ and (9.7) tells us that its limiting distribution is the one we would get if $s^0$ were known. Similarly, $\mathbf{g}(\hat{s}, \hat{\boldsymbol{\theta}}^{(\hat{s})})$ will be our estimator for $\boldsymbol{\lambda}^0$, and its asymptotic properties follow directly from (9.7).

## 10. Concluding remarks

10.1. In the models described in Sections 2.1 and 2.2 above, the seniority parameter is continuous. If it is known (or if one assumes) that one of the forces of transition can be represented by a nice and smooth parametric function, say

$$\lambda(x) = g(x; \boldsymbol{\theta}), \tag{10.1}$$

and if one is faced with the problem of estimating $\lambda$, using analytic graduation is not necessarily the most obvious line of attack. In fact, it seems more natural to try to construct an estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ directly, without going the way via the $\lambda^{\#}$, as described in Section 3. Grenander [29], pp. 76–91, has shown how this might be done for the force of mortality in the example in 2.2 (i) when the Gompertz-Makeham formula (1.1) (with continuous $x$) applies. A similar investigation could be carried out for other forces of transition, like the forces of fertility of the example in 2.2 (ii).

If one does not know enough about the function $\lambda(\cdot)$ to specify a parametric $g(\cdot, \boldsymbol{\theta})$ which can represent it, one may turn to nonparametric methods, such as those developed within reliability theory [5], [6]. The force of mortality in Section 2.2 (i) appears there under the name of failure rate or hazard rate, and quite a lot of energy has gone into finding suitable methods of estimating this function.

Although both of these types of approach were initiated by Grenander's paper [29] on mortality measurement, such techniques do not seem to be much in use in demography and related fields. One would be curious to know why this is so. Part of the explanation is, no doubt, that these developments are largely unknown among people working in those fields of application, but there are more valid reasons. We shall suggest some of them.

10.2. The following types of argument seem to be among the ones leading people to base their inferences from the data on the $M(a, d)$ and $L(a, d)$ only, and sometimes on the $\hat{\lambda}_{a,d}$ only. (Note that least squares and moments method estimation procedures of $\boldsymbol{\theta}$ only require knowledge of the $\hat{\lambda}_{a,d}$.)

(i) In Section 3.2 we described how the points $\{\zeta_d : d = 1, 2, \cdots, D\}$ partitioning the seniority interval were selected according to conventional rules. Similarly, it is standard procedure to calculate "occurrence/exposure" rates of the kind developed in Sections 3.4 and 4.3. The use of standard techniques, standard tabulations, and so on, facilitates comparison with other investigations of the same subject matter. This encourages the continued use of techniques which are already widely known and widely applied even when other methods may be known to a few people.

(ii) The reliability of the data which demographers have to work with, can be very weak due to phenomena such as age misreporting, underenumeration, and so on. Also, one frequently does not know more than approximate dates (for example, the calendar year only) of occurrences of the events studied. This calls for the application of rather robust statistical techniques, such as those which we have described. Even though demographic data may be deficient, they may still be reliable enough to permit the use of the aggregated values $M(a, d)$ and $L(a, d)$ or at least the $\hat{\lambda}_{a,d}$.

In many cases, the investigator does not even have access to the original data, but only to standard tabulations made from them. Such tabulations will often permit the use of methods described here, and rule out others.

(iii) A similar argument applies to the reliability of the *models* used. For example, most current models, including those considered in this paper, leave seasonal variation over the calendar year out of account. There is plenty of evidence of the importance of such variation in the occurrence of vital events, but in many cases this is just a nuisance factor which one wants to eliminate. Current methods relying on seniority interval lengths of at least a year seem to effectively do so.

(iv) Even in cases where the data are reliable and sufficiently detailed (and the present author believes that not nearly enough attention has been given to such cases), the information extracted by a statistical procedure should be geared to the needs of the user. It seems that a standard table of rates, like Table I, and certain other tables derived from it, contains just about as much information as can be handled in a substantive study. In fact, the prevalence of summary indices derived from such tables, and the extent to which argumentation is carried out in terms of such indices, suggests that the standard tables contain even too much information. The use of analytic graduation can be seen as another piece of evidence in the same direction, since it enables one to substitute the formula of a function and a (small) set of parameter values for a whole table. (This argument does not rule out the parametric procedure suggested at the beginning of Section 10.1.)

(v) Each of the estimators $\hat{s}$ listed in Section 9.2 is a function of the data via $\hat{\boldsymbol{\lambda}}$ only. This reflects the fact that an investigator faced with the problem of selecting a graduating function from a class $\mathscr{F}$ of candidates is likely to calculate $\hat{\boldsymbol{\lambda}}$, plot the corresponding diagram, and use this to decide which member to choose from $\mathscr{F}$. In fact, this is the way in which certain graduating functions historically have been pinpointed as more suitable than others.

Once $\hat{s}$ has been determined, however, the investigator should not necessarily continue to use $\hat{\boldsymbol{\lambda}}$ in the estimation of $\boldsymbol{\theta}$, but should feel free to choose among all available procedures as far as the quality of his data permits.

10.3. It is probably appropriate to underline once more (compare Section 1.1) that there exist many types of graduation methods in addition to analytic graduation techniques. Most of them were first developed for use in mortality studies, and in that context they are apparently applied at least as often as analytic methods are. Many of them must have been intended for use in other connections as well, for example in fertility studies. With the exception of graphic methods, however, their application to other types of vital rates than mortality rates seems much less popular. (Compare [54], p. 53.)

◇ ◇ ◇ ◇ ◇

I wish to express my great indebtedness to L. LeCam and P. J. Bickel for many helpful discussions on the large sample theory used in this paper, and to M. L. Eaton, who has read the paper in manuscript and has given valuable comments. I am also grateful to W. Simonsen and L. Boza who independently

suggested problems which led to the formulations in Section 4 above. The data for the diagrams were furnished by the Central Bureau of Statistics of Norway, and I carried out their graduation with help from the Department of Statistics, University of California, Berkeley.

## REFERENCES


[1] H. AMMETER, "Wahrscheinlichkeitstheoretische Kriterien für die Beurteilung der Güte der Ausgleichung einer Sterbetafel," *Mitt. Verein. Schweiz. Versich.-Math.*, Vol. 52 (1952), pp. 19–72.

[2] ———, "Der doppelseitige und die einseitigen ($I\chi^2$)-Tests und ihre Leistungsfähigkeit für die wahrscheinlichkeitstheoretische Überprüfung von Sterbetafeln," *Blätter der Deutschen Gesellschaft für Versicherungsmathematik* (*Deutscher Aktuarverein*) *E. V.*, Vol. 1 (1953), pp. 39–60.

[3] T. W. ANDERSON, *An Introduction to Multivariate Statistical Analysis*, New York, Wiley, 1958.

[4] E. W. BARANKIN and F. GURLAND, "On asymptotically normal, efficient estimators: I," *Univ. California Publ. Statist.*, Vol. 1 (1951), pp. 89–130. Compare [23].

[5] R. E. BARLOW, "Some recent developments in reliability theory," *Amsterdam Mathematisch Centrum, Mathematical Centre Tracts*, No. 27 (1968), pp. 49–65.

[6] R. E. BARLOW and W. R. VAN ZWET, "Asymptotic properties of isotonic estimators for the generalized failure rate function. Parts I and II," University of California, Berkeley, Operations Research Center Reports, No. ORC 69-5 and 69-10, 1969.

[7] H. A. R. BARNETT, "Graduation tests and experiments," *J. Inst. Actuaries*, Vol. 72 (1951), pp. 15–74.

[8] B. BENJAMIN, *Health and Vital Statistics*, London, George Allen & Unwin Ltd., 1968.

[9] P. BICKEL, "Lecture notes for asymptotic theory, Statistics 217B, Winter 1970," University of California, Berkeley, Department of Statistics (mimeographed).

[10] P. BILLINGSLEY, *Convergence of Probability Measures*, New York, Wiley, 1968.

[11] W. BRASS, "The graduation of fertility distributions by polynomial functions," *Population Studies*, Vol. 14 (1960), pp. 148–162.

[12] J. M. CALLIES, "Utilisation de modèles mathématiques pour l'estimation des données démographiques dans les pays en voie de developpement," *Rev. Inst. Internat. Statist.*, Vol. 34 (1966), pp. 341–359.

[13] C. CHANDRASEKARAN and P. P. TALWAR, "Forms of age-specific birth rates by orders of birth in an Indian community," *Eugenics Quarterly*, Vol. 15 (1968), pp. 264–272.

[14] C. L. CHIANG, "On regular best asymptotically normal estimates," *Ann. Math. Statist.*, Vol. 27 (1956), pp. 336–351.

[15] ———, *Introduction to Stochastic Processes in Biostatistics*, New York, Wiley, 1968.

[16] H. CRAMÉR and H. WOLD, "Mortality variations in Sweden. A study in graduation and forecasting," *Skand. Aktuarietidskr.*, Vol. 18 (1935), pp. 161–240.

[17] M. DAVIES, "Linear approximation using the criterion of least total deviations," *J. Roy. Statist. Soc. Ser. B*, Vol. 29 (1967), pp. 101–109.

[18] J. M. DICKEY, "Smoothed estimates for multinomial cell probabilities," *Ann. Math. Statist.*, Vol. 39 (1968), pp. 561–566.

[19] ———, "Smoothing by cheating," *Ann. Math. Statist.*, Vol. 40 (1969), pp. 1477–1482.

[20] H. F. DORN, "Methods of analysis for follow-up studies," *Human Biology*, Vol. 22 (1950), pp. 238–248.

[21] L. G. DU PASQUIER, "Mathematische Theorie der Invaliditätsversicherung," *Mitt. Verein. Schweiz. Versich.-Math.*, Vol. 7 (1912), pp. 1–7, and Vol. 8 (1913), pp. 1–153.

[22] W. P. ELDERTON and N. L. JOHNSON, *Systems of Frequency Curves*, London, Cambridge University Press, 1969.

[23] G. U. FENSTAD, "A note on a theorem of Barankin and Gurland," University of Oslo, Institute of Mathematics, Statistical Research Report, No. 2, 1965.

[24] T. S. FERGUSON, "A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities," *Ann. Math. Statist.*, Vol. 29 (1953), pp. 1046–1062.

[25] E. FIX and J. NEYMAN, "A simple stochastic model of recovery, relapse, death and loss of patients," *Human Biology*, Vol. 23 (1951), pp. 205–241.

[26] D. A. S. FRASER, *Nonparametric Methods in Statistics*, New York, Wiley, 1957.

[27] E. GILJE, "Fitting curves to age-specific fertility rates. Some examples," *Statistical Review, Ser.* III, Vol. 7 (1969), pp. 118–134 (Swedish Central Bureau of Statistics, Stockholm).

[28] E. GILJE and L. YNTEMA, "The shifted Hadwiger fertility function," to appear.

[29] U. GRENANDER, "On the theory of mortality measurement," *Skand. Aktuarietidskr.*, Vol. 39 (1956), pp. 70–96 and 125–153.

[30] R. D. GROVE and A. M. HETZEL, *Vital Statistics Rates in the United States, 1940–1960*, Washington, D.C., National Center for Health Statistics, U.S. Department of Health, Education and Welfare, 1968.

[31] H. HADWIGER, "Eine analytische Reproduktions-funktion für biologische Gesamtheiten," *Skand. Aktuarietidskr.*, Vol. 23 (1940), pp. 101–113.

[32] T. E. HARRIS, P. MEIER, and J. W. TUKEY, "Timing of the distribution of events between observations," *Human Biology*, Vol. 22 (1950), pp. 249–270.

[33] L. HENRY, "D'un problème fondamental de l'analyse démographique," *Population*, Vol. 14 (1959), pp. 9–32.

[34] A. B. HOADLEY, "Use of the Pearson densities for approximating a skew density whose left terminal and first three moments are known," *Biometrika*, Vol. 55 (1968), pp. 559–563.

[35] J. M. HOEM, "Markov chain models in life insurance," *Blätter der Deutschen Gesellschaft für Versicherungsmathematik (Deutscher Aktuarverein) E. V.*, Vol. 9 (1969), pp. 91–107.

[36] ———, "Fertility rates and reproduction rates in a probabilistic setting," *Biométr.-Praxim.*, Vol. 10 (1969), pp. 38–66.

[37] ———, "A probabilistic model for primary marital fertility," *Yearbook of Population Research in Finland*, Vol. 11 (1969), pp. 73–86.

[38] ———, "Probabilistic fertility models of the life table type," *Theoretical Population Biology*, Vol. 1 (1970), pp. 12–38.

[39] ———, "Some results concerning the number of jumps in a finite time interval of a Markov chain with a continuous time parameter," University of Oslo, Institute of Economics, Memorandum, February 2, 1970.

[40] ———, "Point estimation of forces of transition in demographic models," *J. Roy. Statist. Soc. Ser. B*, Vol. 32 (1970), to appear.

[41] ———, "A probabilistic approach to nuptiality," *Biométr.-Praxim.*, to appear.

[42] L. HØYLAND, "Estimation in follow-up studies," University of Oslo, Institute of Mathematics, Statistical Research Report, No. 4, 1967.

[43] H. JECKLIN and P. STRICKLER, "Wahrscheinlichkeitstheoretische Begründung mechanischer Ausgleichung und deren praktische Anwendung," *Mitt. Verein. Schweiz. Versich.-Math.*, Vol. 54 (1954), pp. 125–161.

[44] D. A. JONES, "Bayesian statistics," *Trans. Soc. Actuaries*, Vol. 17 (1965), pp. 33–57.

[45] N. KEYFITZ, *Introduction to the Mathematics of Population*, Reading, Addison-Wesley, 1968.

[46] G. S. KIMELDORF and D. A. JONES, "Bayesian graduation," *Trans. Soc. Actuaries*, Vol. 19 (1967), pp. 66–112.

[47] YU. L. KRUOPIS, "On estimates of transition intensities with migration," *Theor. Probability Appl.*, Vol. 14 (1969), pp. 219–228.

[48] I. LAH, "Generalization of Yastremsky's formula for analytical graduation of fertility rates," *J. Roy. Statist. Soc. Ser. A*, Vol. 121 (1958), pp. 100–104.

[49] ———, "Analytische Ausgleichung der aus den Ergebnissen der Volkszählungen berechneten demographischen Tafeln," *International Population Conference, Wien, 1959*, International Union for the Scientific Study of Population, pp. 192–201.

[50] ———, "Analytically graduated fertility of married women in Australia with respect to the duration of marriage," *Contributed Papers, Sydney Conference, 1967*, International Union for the Scientific Study of Population, pp. 266–276.

[51] L. LECAM, "Asymptotic least squares theory," mimeographed notes, University of California, Berkeley, 1955.

[52] S. LIPTON and C. A. MCGILCHRIST, "The derivation of methods for fitting exponential regression curves," *Biometrika*, Vol. 51 (1964), pp. 504–507.

[53] A. J. LOTKA, *Théorie Analytique des Associations Biologiques. Part II. Analyse Démographique avec Application Particulière à l'Espèce Humaine*, Paris, Hermann & Cie, 1939.

[54] D. P. MAZUR, "The graduation of age-specific fertility rates by order of birth of child," *Human Biology*, Vol. 39 (1967), pp. 53–64.

[55] M. D. MILLER *et al.*, *Elements of Graduation*, New York, Actuarial Society of America, 1942.

[56] S. MITRA, "The pattern of age-specific fertility rates," *Demography*, Vol. 4 (1967), pp. 894–906.

[57] H. D. PATTERSON, "A further note on a simple method for fitting an exponential curve," *Biometrika*, Vol. 47 (1960), pp. 177–180.

[58] E. S. PEARSON and H. O. HARTLEY, *Biometrika Tables for Statisticians*, Cambridge University Press, 1966.

[59] W. SAXER, *Versicherungsmathematik. Zweiter Teil*, Berlin, Springer-Verlag, 1958.

[60] B. SCHNEIDER, "Die Bestimmung der Parameter im Ertragsgesetz von E. A. Mitscherlich," *Biometr. Z.*, Vol. 5 (1963), pp. 78–95.

[61] H. L. SEAL, "Tests of a mortality table graduation," *J. Inst. Actuaries*, Vol. 71 (1941), pp. 5–67.

[62] M. C. SHEPS, "Characteristics of a ratio used to estimate failure rates: occurrences per person year of exposure," *Biometrics*, Vol. 22 (1966), pp. 310–321.

[63] ———, "On the person year concept in epidemiology and demography," *Milbank Memorial Fund Q.*, Vol. 44 (1966), pp. 69–91.

[64] M. C. SHEPS, J. A. MENKEN, and A. P. RADICK, "Probability models for family building, an analytical review," *Demography*, Vol. 6 (1969), pp. 161–183.

[65] M. C. SHEPS and J. A. MENKEN, "On closed and open birth intervals in a stable population," paper prepared for the Segunda Conferencia Regional de Poblacion, Mexico City, August, 1970.

[66] W. SIMONSEN, *Forsikringsmatematik Hefte I og II*, København Universitets Fond til Tilvejebringelse af Laeremidler, 1966–67.

[67] M. SPIEGELMAN, *Introduction to Demography*, Cambridge, Harvard University Press, 1968 (revised ed.).

[68] W. L. STEVENS, "Asymptotic regression," *Biometrics*, Vol. 7 (1951), pp. 247–267.

[69] E. SVERDRUP, "Basic concepts in life assurance mathematics," *Skand. Aktuarietidskr.*, Vol. 35 (1952), pp. 115–131.

[70] ———, "Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between different states of health," *Skand. Aktuarietidskr.*, Vol. 46 (1965), pp. 184–211.

[71] ———, *Laws and Chance Variations*, Vol. 1, Amsterdam, North Holland, 1967.

[72] P. P. TALWAR, "Age patterns of fertility," University of North Carolina, Chapel Hill, Institute of Statistics Mimeo Series, No. 656, 1970.

[73] K. TEKSE, "On demographic models of age-specific fertility rates," *Statistical Review*, *Ser.* III, Vol. 5 (1967), pp. 189–207 (Swedish Central Bureau of Statistics, Stockholm).

[74] E. J. WEGMAN, "Maximum likelihood histograms," University of North Carolina, Chapel Hill, Institute of Statistics Mimeo Series, No. 629, 1969.

[75] ———, "Nonparametric probability density estimation," University of North Carolina, Chapel Hill, Institute of Statistics Mimeo Series, No. 638, 1969.

[76] K. WEICHSELBERGER, "Über eine Theorie der gleitende Durchschnitte und verschiedene Anwendungen dieser Theorie," *Metrika*, Vol. 8 (1964), pp. 185–230.

[77] S. D. WICKSELL, "Nuptiality, fertility, and reproductivity," *Skand. Aktuarietidskr.*, Vol. 14 (1931), pp. 125–157.

[78] R. A. WIJSMAN, "On the theory of BAN-estimates," *Ann. Math. Statist.*, Vol. 30 (1959), pp. 185–191. (Correction note, pp. 1268–1270.)

[79] H. H. WOLFENDEN, *Fundamental Principles of Mathematical Statistics with Special Reference to the Requirements of Actuaries and Vital Statisticians, and an Outline of a Course in Graduation*, New York, Actuarial Society of America, 1942.

[80] L. YNTEMA, "The graduation of net fertility tables," *Bolm Instit. Actuários Portugueses*, Vol. 8 (1953), pp. 29–43.

[81] ———, "On Hadwiger's fertility function," *Statistical Review, Ser.* III, Vol. 7 (1969), pp. 113–117 (Swedish Central Bureau of Statistics, Stockholm).

[82] S. ZAHL, "A Markov process model for follow-up studies," *Human Biology*, Vol. 27 (1955), pp. 90–120.

[83] E. ZWINGGI, *Versicherungsmathematik*, Basel, Birkhäuser, 1958.