Arbeidsnotater

S T A T I S T I S K S E N T R A L B Y R Å

WORKING PAPERS FROM THE CENTRAL BUREAU OF STATISTICS OF NORWAY

IO 68/~~16~~ 18 Oslo, 22. August 1968

FOUR DEMOGRAPHIC PAPERS

b y

Jan M. Hoem

C O N T E N T S

Purged and Partial Markov Chains

by Jan M. Hoem

C O N T E N T S

1. Introduction and summary

Consider a time-continuous Markov chain with a finite or countable state space I and with time-dependent transition probabilities (Feller, 1957, Chapter XVII.9). Assume that a number of independent sample paths of the chain are observed over a period $[0,x]$ with a view to statistical analysis of parameters of the process. One may sometimes like to remove such sample paths as end up in some closed subset R of the states in I by time x before carrying out the statistical analysis. We shall show that, on certain general conditions, the remaining paths may be regarded as realizations of a new Markov chain over the state space K=I-R, which we shall call the purged chain. By comparison to a corresponding partial Markov chain, which we will define, we shall see that the forces of transition of the purged chain generally differ from those of the original one. If the original chain is time-homogeneous, the purged one need not be so.

2. Blanket assumptions

If a sample path is in state i at time s, let $P_{ij}(s,t)$ be the probability that it will be in state j at time t>s. Let $P_{iA}(s,t) = \sum_{j\in A} P_{ij}(s,t)$ for $A\subseteq I$.

The following assumptions will be adopted to hold throughout the paper.

Assumption 1: For all i and j∈I,

$$P_{ij}(s,t)\geq 0,\ P_{iI}(s,t)=1 \qquad \text{for } 0\leq s<t\leq x,$$

$$\lim_{t\downarrow s} P_{ij}(s,t) = P_{ij}(s,s) = \delta_{ij} \qquad \text{for any } s\in[0,x\rangle,$$

and
$$P_{ij}(s,u) = \sum_{k\in I} P_{ik}(s,t)\, P_{kj}(t,u) \qquad \text{for } 0\leq s<t<u\leq x.$$

Assumption 2: To every pair (i,j) of states where $i\neq j$ there corresponds a time-dependent force of transition

(1) $$\mu_{ij}(s) = \lim_{t\downarrow s} P_{ij}(s,t)/(t-s) < \infty \quad \text{for } s\in[0,x\rangle.$$

Assumption 3: To every state i there corresponds a time-dependent total force of decrement

(2) $$\mu_i(s) = \lim_{t\downarrow s} \{1-P_{ii}(s,t)\}/(t-s) < \infty \quad \text{for } s\in[0,x\rangle$$

(3) where $\mu_i(s) = \sum_{j\in I-i} \mu_{ij}(s)$.

We permit the possibility $\mu_{ij}(s)\to\infty$ (or $\mu_i(s)\to\infty$) as $s\uparrow x$ for some state i and some $j\neq i$.

Let K and R be defined as in § 1, let H be the set of states in K from which R cannot be reached, and let J=K-H. Of course H may be empty. Otherwise H is closed. If $H\neq\emptyset$, we shall assume that also $J\neq\emptyset$ to avoid trivialities.

We wish to avoid that all sample paths end up in R by time x with probability 1. We also want to secure that some paths starting in K may enter R. We therefore make

<u>Assumption 4</u>: For any $i\in K$ and $s\in[0,x\rangle$, $P_{iK}(s,x)>0$. There exists an $i\in K$ and an $s\in[0,x\rangle$ such that $P_{iK}(s,x)<1$.

3. The purged chain

If all sample paths ending in R within time x are removed from the data, the remaining paths have transition probabilities

$$(4)\qquad Q_{ij}(s,t,x) = P_{ij}(s,t)\, P_{jK}(t,x)/P_{iK}(s,x)$$

for all $i\in K$, $j\in K$, $0\leq s<t\leq x$. Specifically

$$(5)\qquad Q_{ij}(s,t,x) = \begin{cases} P_{ij}(s,t)/P_{iK}(s,x) & \text{for } i\in J,\ j\in H, \\ P_{ij}(s,t) & \text{for } i\in H,\ j\in K. \end{cases}$$

It is easily proved that the $Q_{ij}(s,t,x)$ satisfy conditions similar to those in assumption 1 with I replaced by K provided the following assumption holds:

<u>Assumption 5</u>: For any given $j\in K$, $P_{jK}(\cdot,x)$ is continuous from the right in $[0,x\rangle$.

Under this assumption, forces of transition satisfy

$$(6)\qquad \lambda_{ij}(s,x) = \lim_{t\downarrow s} Q_{ij}(s,t,x)/(t-s)$$
$$= \mu_{ij}(s)\, P_{jK}(s,x)/P_{iK}(s,x) \quad \text{for } i\neq j,\ i\in K,\ j\in K,\ 0\leq s<x.$$

As in (5), this formula simplifies somewhat for $i\in J$, $j\in H$, and for $i\in H$, $j\in K$.

For each $i\in H$ a total force of decrement (relative to the $Q_{ij}(s,t,x)$) exists and equals $\mu_i(s)$. (Cf. the second member of (5).) To prove the existence of a finite total force of decrement satisfying a relation similar to (3) in the purged process for an $i\in J$, more restrictive conditions seem necessary. Specifically, if K is finite,

$$(7)\qquad \lambda_i(s,x) = \lim_{t\downarrow s} \{1-Q_{ii}(s,t,x)\}/(t-s) = \sum_{j\in K-i} \lambda_{ij}(s,x)$$

for all $i\in K$.

Obviously the sample paths removed from the data can be given a similar treatment.

4. The partial chain

A third Markov chain can be derived from the original one in the following way.

R is removed from the state space, which is thereby reduced to K. For each $i \in K$, $j \in R$, the function $\mu_{ij}(.)$ is substituted by 0. We make

Assumption 6: From the remaining forces of transition $\mu_{ij}(.)$ with i and $j \in K$ a Markov chain with state space K can be uniquely constructed.

We shall call this the partial chain corresponding to K. Its transition probabilities will be designated $\bar{P}_{ij}(s,t)$. These functions constitute a generalization of the partial probabilities of multiple decrement theory (Du Pasquier, 1913; Hoem, 1968a). (Other names for the same concept are "independent probabilities" (Zwinggi, 1945), "absolute probabilities" (Jordan, 1952), and "net probabilities" or "partial crude probabilities" (Chiang, 1961). Correspondingly the $P_{ij}(s,t)$ are called "influenced probabilities"(Sverdrup, 1961), "dependent probabilities" (Zwinggi, 1945; Jordan, 1952), or "crude probabilities" (Chiang, 1961).)

It has some interest to compare the purged and the partial Markov chain. Both are derived from the original chain, and both have state space K, but their transition probabilities will generally be different. By (6) and assumption 6 the purged and the partial process are identical if and only if $P_{iK}(.,x)$ is independent of i for $i \in K$. In that case H must be empty, since otherwise $\lambda_{ij}(s,x) = \mu_{ij}(s)/P_{iK}(s,x) > \mu_{ij}(s)$ for some $i \in J$, $j \in H$, $s \in [0,x>$, by assumption 4.

It is intuitively plausible that the two processes may be indentical if the $\mu_{ij}(s)$ are independent of i for $i \in K$, $j \in R$. We prove

Theorem 1: Let K be finite and assume that for each $j \in R$ there exists a function $\gamma_j(.)$ over $[0,x>$ such that $\mu_{ij}(.) = \gamma_j(.)$ for all $i \in K$. Let $\gamma_R = \sum_{j \in R} \gamma_j$, and assume that $\gamma_R(.)$ is continuous in $[0,x>$. Then

$$(8) \qquad P_{iK}(s,t) = \exp\{-\int_0^t \gamma_R(\tau)d\tau\}$$

for all $i \in K$, $0 \leq s < t < x$.

Proof: The Kolmogorov forward differential equation

$$\frac{\partial}{\partial t} P_{ij}(s,t) = -\mu_j(t) P_{ij}(s,t) + \sum_{k \in K} P_{ik}(s,t)\mu_{kj}(t)$$

holds for $i \in K$, $j \in K$, $0 \leq s < t < x$. Let $\mu_{kA}(t) = \sum_{j \in A} \mu_{kj}(t)$. Summation over all j in K gives

$$\frac{\partial}{\partial t} P_{iK}(s,t) = - \sum_{j \in K} \mu_j(t) P_{ij}(s,t) + \sum_{k \in K} P_{ik}(s,t)\mu_{kK}(t).$$

Since $\mu_{kK} = \mu_k - \mu_{kR} = \mu_k - \gamma_R$, we get

$$\frac{\partial}{\partial t} P_{iK}(s,t) = -\gamma_R(t)\, P_{iK}(s,t),$$

from which the theorem follows. □

The partial probability $\bar{P}_{ij}(s,t)$ need not have any interpretation at all as a probability within the original process (Hoem, 1968a and b). In a certain case this is different, however. Assume that the sample space has the form $I=M\times N$, where M is finite or countable, and where $N=\{1,2,\dots,n\}$ for some positive integer $n\geq 2$. Let $N^* = \{1,2,\dots,n-1\}$ and assume further that the forces of transition have the form

$$\mu_{(i,\alpha),(j,\beta)}(s) = \begin{cases} \phi_{ij}(s) & \text{for } i,j\in M, \alpha=\beta\in N, \\ \eta_{i\alpha}(s) & \text{for } i = j\in M, \alpha\in N^*, \beta=\alpha+1, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

(An example has been given by Hoem (1968b, § 3C).) We let $P_{(i,\alpha),(j,N)}(s,t) = \sum_{\beta\in N} P_{(i,\alpha),(j,\beta)}(s,t)$, and have

<u>Theorem 2</u>: Let $\alpha\in N^*$ and let $R = \{(j,\beta): j\in M, \beta>\alpha\}$. If the limit $\phi_{ij}(s) = \lim_{t\downarrow s} P_{(i,\alpha),(j,\alpha)}(s,t)/(t-s)$ exists uniformly in $i\in M$, then

$$\bar{P}_{(i,\alpha),(j,\alpha)}(s,t) = P_{(i,\alpha),(j,N)}(s,t).$$

<u>Proof</u>: Let $\phi_i(s) = \sum_{j\in M} \phi_{ij}(s)$. By the assumptions of the theorem we have

$$\frac{\partial}{\partial t}P_{(i,\alpha),(j,\beta)}(s,t) = -P_{(i,\alpha),(j,\beta)}(s,t)\,\phi_j(t) + \sum_{k\in M-j} P_{(i,\alpha),(k,\beta)}(s,t)\,\phi_{kj}(t)$$

$$-P_{(i,\alpha),(j,\beta)}(s,t)\eta_{j\beta}(t) + P_{(i,\alpha),(j,\beta-1)}(s,t)\eta_{j,\beta-1}(t),$$

for all $i\in M$, $j\in M$, $\alpha\in N$, $\beta\in N$, with suitable interpretations when $\beta=1$ or $\beta=n$. Summation over all β in N gives

$$\frac{\partial}{\partial t}P_{(i,\alpha),(j,N)}(s,t) = -P_{(i,\alpha),(j,N)}(s,t)\phi_j(t) + \sum_{k\in M-j} P_{(i,\alpha),(k,N)}(s,t)\phi_{kj}(t).$$

The theorem then follows from assumption 6. □

5. The time-homogeneous case

If the μ_i and the μ_{ij} are independent of s and $P_{ij}(s,s+t) = P_{ij}(t)$, we have

$$\lambda_{ij}(s,x) = \mu_{ij}\ P_{jK}(x-s)/P_{iK}(x-s),$$

which may genuinely depend on x-s. Even when the original Markov chain is time-homogeneous, the purged chain may thus be time-dependent.

In this case, $Q_{ij}(s,t,x)$ will depend on x-s and x-t, i.e. generally not on t-s only.

6. References

Chiang, C.L. (1961): "A Stochastic Study of the Life Table and its Applications. III. The Follow-up Study with the Consideration of Competing Risks." Biometrics, **17**, 57-78.

Du Pasquier, L.G. (1913): "Mathematische Theorie der Invaliditätsversicherung." Mitteil.Verein. schweitzer.Versicherungsmath., **8**, 1-153.

Feller, William (1957): "An Introduction to Probability Theory and Its Applications. Volume I. Second Edition." John Wiley & Sons, Inc.

Hoem, Jan M. (1968a): "Fertility Rates and Reproduction Rates in a Probabilistic Setting." Under publication.

Hoem, Jan M. (1968b): "A Probabilistic Approach to Nuptiality." Under publication.

Jordan, C.W. (1952): "Life Contingencies." The Society of Actuaries.

Sverdrup, Erling (1961): "Statistiske metoder ved dødelighetsundersøkelser." Mimeographed. Institute of Mathematics, University of Oslo.

Zwinggi, Ernst (1945): "Versicherungsmathematik." Verlang Birkhäuser, Basel.

Concepts of a Bisexual Theory of Marriage Formation.

By Jan M. Hoem.

CONTENTS

1. Introduction and summary

§ 1A. In the analysis of nuptiality it may be fruitful to regard a marriage as a contract established in a "marriage market" in which there is a demand for partners and a supply of partners. In many respects the two sexes enter symmetrically into this market. Rather than analysing marriage formation unisexually, i.e. from the point of view of one sex only, a bisexual approach therefore seems preferable.

§ 1B. In the present note we shall look into some of the basic concepts relevant to such an approach. Our terminology will be quite similar to that of birth-and-death processes. The most important concept introduced is a function $\nu_{ij}(\underset{\sim}{m},\underset{\sim}{k})$, called "the propensity to marry" for the age group conbination (i,j), conditional upon a marriageable population with size paramters $\underset{\sim}{m}$ and $\underset{\sim}{k}$. We suggest that a bisexual theory of marriage formation be formulated in terms of this concept.

§ 1C. Yntema (1954) has studied models closely related to ours.

2. A simple bisexual marriage-and-death process

§ 2A. Consider first a closed population observed during a period $[0,\zeta]$. Since real life nuptiality varies with age, any reasonably realistic nuptiality model would incorporate an age concept in some form. For our purposes it suffices to partition the population into suitable age groups.

Departing slightly from common terminology, we define an age group in the following way. We partition the marriageable age interval $[x_0,\omega\rangle$ for males into intervals $[x_0,x_1\rangle,[x_1,x_2\rangle,\ldots,[x_{I-1},x_I\rangle$, with $x_I=\omega$, and let $[y_0,y_1\rangle,[y_1,y_2\rangle,\ldots,[y_{J-1},y_J\rangle$ with $y_J=\omega$ be corresponding intervals for the marriageable females.(As usual ω is the highest possible live age.) The males who at time zero have ages in the interval $[x_{i-1},x_i\rangle$ will be taken to constitute "male age group i" throughout the period $[0,\zeta]$. At time ζ "male age group i" will then obviously consist of males at ages in the interval $[x_{i-1}+\zeta,x_i+\zeta\rangle$. Similarly for the females. In this way any person will remain within the same "age group" throughout the period of observation. This approach is well adapted to cohort analysis, and it will ease our presentation.

§ 2B. Let us concentrate for the moment on the subpopulation consisting of all bachelors in a certain age group i and spinsters in an age group j. The number of marriages formed within this subpopulation during $[0,\zeta]$ depends on a series of factors, such as the number of marriageable males and females in other age groups and the number of previously married persons in male age group i and female age group j. To get a fairly simple mathematical model, however, we shall disregard for the moment all such factors except the number M_i of bachelors and the number K_j of spinsters in the subpopulation considered, their mortality, and their propensity to marry.

§ 2C. As the period of observation progresses, the number of bachelors in the subpopulation will decrease because some bachelors marry and others die. Similarly for the spinsters. Let $M_i(t)$ be the number of bachelors left in the subpopulation at time t, let $K_j(t)$ be the similar number for the spinsters, and let $N_{ij}(t)$ be the number of marriages contracted within the subpopulation during the period $[0,t]$.

We introduce a male force of mortality μ_i, a female force of mortality γ_j, and a "propensity to marry" $\nu_{ij}(m,k)$. These will be defined as follows:

$$\dot{P}\{M_i(t+\Delta t)=m-1, K_j(t+\Delta t)=k, \text{ and } N_{ij}(t+\Delta t)=n \mid M_i(t)=m, K_j(t)=k, \text{ and } N_{ij}(t)=n\}$$
$$=m\mu_i\Delta t+o(\Delta t); \text{ (a male death).}$$

$$\dot{P}\{M_i(t+\Delta t)=m, K_j(t+\Delta t)=k-1, \text{ and } N_{ij}(t+\Delta t)=n \mid M_i(t)=m, K_j(t)=k, \text{ and } N_{ij}(t)=n\}$$
$$=k\gamma_j\Delta t+o(\Delta t); \text{ (a female death).}$$

$$\dot{P}\{M_i(t+\Delta t)=m-1, K_j(t+\Delta t)=k-1, \text{ and } N_{ij}(t+\Delta t)=n+1 \mid M_i(t)=m, K_j(t)=k, \text{ and } N_{ij}(t)=n\}$$
$$=\nu_{ij}(m,k)\Delta t+o(\Delta t); \text{ (a marriage).}$$

The total probability for all other kinds of changes during $\langle t,t+\Delta t]$ is $o(\Delta t)$ when values for $M_i(t)$, $K_j(t)$, and $N_{ij}(t)$ have been assigned. (As usual $o(\Delta t)/\Delta t\to 0$ as $\Delta t\to 0$.) Suppressing the indices i and j for the rest of this paragraph, we let

$$P_{mkn}(t) = \dot{P}\{M(t) = m, K(t) = k, \text{ and } N(t) = n\},$$

$$P(t,x,y,z) = \sum_{m,k,n} P_{mkn}(t)\, x^m y^k z^n, \text{ and}$$

$$\Psi(t,x,y,z) = \sum_{m,k,n} \nu(m,k)P_{mkn}(t)x^m y^k z^n.$$

Standard methods will then give the following relations:

$$\frac{d}{dt}P_{mkn}(t) = -\{m\mu+k\gamma+\nu(m,k)\}P_{mkn}(t) + (m+1)\mu P_{m+1,k,n}(t)$$

$$+ (k+1)\gamma P_{m,k+1,n}(t) + \nu(m+1,k+1)P_{m+1,k+1,n-1}(t), \tag{2.1}$$

$$\frac{\partial}{\partial t}P(t,x,y,z) = \mu(1-x)\frac{\partial}{\partial x}P(t,x,y,z) + \gamma(1-y)\frac{\partial}{\partial y}P(t,x,y,z) + (\frac{z}{xy} -1)\Psi(t,x,y,z), \tag{2.2}$$

$$\frac{d}{dt}EM(t) = -\mu\ EM(t)- \frac{d}{dt}EN(t),\ \frac{d}{dt}EK(t) = -\gamma\ EK(t)- \frac{d}{dt}EN(t),$$

$$\text{and}\ \frac{d}{dt}EN(t) = \sum_{m,k,n} \nu(m,k)P_{mkn}(t). \tag{2.3}$$

The relations in (2.3) have an immediate interpretation in a deterministic model.

§ 2D. Our assumptions of constant forces of mortality are reasonably realistic provided ζ is not too large. To specify a set of values for $\nu(m,k)$ for the various (m,k) is in fact to give a theory of how people marry (in the subpopulation under consideration). To establish such a theory must be one of the main objectives of (bisexual) nuptiality analysis.

If such a specification were available, together with a specification of the values of μ and γ, relations (2.1) and (2.3) would make it possible to find the quantities $EM(t)$, $EK(t)$, $EN(t)$, and $P_{mkn}(t)$ for various (m,k,n,t) by numerical methods, although this would in general be quite a task due to the fact that there are four arguments in (2.1) (viz. m,k,n, and t).

If we wish to find closed mathematical formulae for $EM(t)$, $EN(t)$, etc., $\nu(m,k)$ must be an exceedingly simple function of m and k, as is shown by the following three examples.

§ 2E. Such a simple choice as $\nu(m,k) = \min(m,k)\ \nu$, where ν is a parameter, is outside our reach as Ψ and $\frac{d}{dt}EN(t)$ will be too complicated for serious treatment.

§ 2F. The choice $\nu(m,k) = mk\nu$ makes (2.2) reduce to

$$\frac{\partial}{\partial t}P(t,x,y,z) = \mu(1-x)\frac{\partial}{\partial x}P(t,x,y,z) + \gamma(1-y)\frac{\partial}{\partial y}P(t,x,y,z)+\nu(z-xy)\frac{\partial^2}{\partial x\partial y}P(t,x,y,z), \tag{2.4}$$

which appears insoluble by known methods. The last relation in (2.3) becomes $\frac{d}{dt}EN(t) = \nu E\{M(t)K(t)\}$, which makes (2.3) equally insoluble.

§ 2G. In case $\nu(m,k) = \frac{1}{2}(m+k)\nu$, (2.3) may be solved to give

$$EM(t) = \{M \cosh(\tfrac{1}{2}t\kappa) - \tfrac{1}{\kappa}[\nu K+(\mu-\gamma)M]\sinh(\tfrac{1}{2}t\kappa)\}e^{-\frac{1}{2}t(\nu+\mu+\gamma)},$$

a similar expression for EK(t), and

$$EN(t) = \tfrac{1}{4}\nu(M+K)\{(e^{rt}-1)/r+(e^{st}-1)/s\}$$

$$-\tfrac{1}{4}(\nu/\kappa)\{\nu(M+K)+(\mu-\gamma)(M-K)\}\{(e^{rt}-1)/r-(e^{st}-1)/s\},$$

where $\kappa = \sqrt{\nu^2+(\mu-\gamma)^2}$, $r = -\frac{1}{2}(\nu+\mu+\gamma)+\kappa$, and $s = -\frac{1}{2}(\nu+\mu+\gamma)-\kappa$. In this case, (2.2) reduces to an expression quite as forbidding as (2.4), and no closed expression for $P_{mkn}(t)$ has thus been found.

§ 2H. These three choices for $\nu(m,k)$ do not of course represent serious attempts at formulating a theory of how people marry. We have presented them only to give some indication of the mathematical difficulties involved.

3. Simultaneous analysis of all age groups

§ 3A. As we have already indicated, the model of chapter 2 has some defects, the most immediate of which probably being the fact that it does not take into consideration persons in age groups outside those on which interest is focused. If bachelors and spinsters in the various age groups are regarded as "goods" in a "marriage market", we have thus left out of account the possibility of substitution between the various kinds of "goods". A more realistic approach would consist in a simultaneous analysis of the nuptiality of all age groups of bachelors and spinsters. (For ease of exposition we shall disregard second and higher order marriages in chapter 3.)

§ 3B. Consider then our closed population with I age groups for bachelors and J age groups for spinsters, and let the forces μ_i and γ_j of mortality be defined as in chapter 2 along with the random variables $M_i(t)$, $K_j(t)$, and $N_{ij}(t)$, whose values $M_i(0)$, $K_i(0)$, and $N_{ij}(0) = 0$ are known. Let $\underset{\sim}{M}(t) = \{M_1(t),\ldots,M_I(t)\}$, let $\underset{\sim}{K}(t) = \{K_1(t),\ldots,K_J(t)\}$, and let $\underset{\sim}{N}(t)$ be the matrix $(N_{ij}(t))$. Given that $\underset{\sim}{M}(t) = \underset{\sim}{m}=(m_1,\ldots,m_I)$ and $\underset{\sim}{K}(t) = \underset{\sim}{k}=(k_1,\ldots,k_J)$, let $\nu_{ij}(\underset{\sim}{m},\underset{\sim}{k})\Delta t+o(\Delta t)$ be the probability of observing a marriage between a bachelor in age group i and a spinster in age group j during the period $\langle t,t,+\Delta t]$. We may call $\nu_{ij}(\underset{\sim}{m},\underset{\sim}{k})$ the "propensity to marry" for the age group combination (i,j), conditional upon the values $(m_1,\ldots, m_I)$ and $(k_1,\ldots,k_J)$ for the random vectors $\underset{\sim}{M}(t)$ and $\underset{\sim}{K}(t)$.

<u>To establish a theory for how bachelors and spinsters marry would then be to specify values for $\nu_{ij}(\underset{\sim}{m},\underset{\sim}{k})$ for the various $\underset{\sim}{m},\underset{\sim}{k}$,i, and j.</u> Given such a theory, one would be interested in finding quantities like the probability distribution and expected values of $\underset{\sim}{M}(t)$, $\underset{\sim}{K}(t)$, and $\underset{\sim}{N}(t)$.

§ 3C. One may always give some more or less vague general statement of probable effects of changes in the arguments $m_1,m_2,\dots,m_I,k_1,k_2,\dots,k_J$. If for instance $m_i''>m_i'$ and $k_j''>k_j'$ for all i and j, one would expect each $\nu_{ij}''=\nu_{ij}(\underset{\sim}{m}'',\underset{\sim}{k}'')$ to be larger than the corresponding $\nu_{ij}'=\nu_{ij}(\underset{\sim}{m}',\underset{\sim}{k}')$, simply because there are more marriageable people of both sexes in the larger population. Similarly if $m_i''=m_i'$ for all i while each k_j'' is considerably larger than the corresponding k_j', one would again expect each ν_{ij}' to be smaller than the corresponding ν_{ij}'', because the marriageable men would have greater difficulty in finding a wife in the smaller population.

These are scale effects. Substitution effects may be described analogously: If $m_i''=m_i'$ for each i and $k_j''=k_j'$ for each $j\neq s$, while $k_s''>k_s'$, each ν_{is}'' would be larger than the corresponding ν_{is}', while for $j\neq s$ each ν_{ij}'' might actually be smaller than the corresponding ν_{ij}', because in the smaller population many bachelors who might have married spinsters in age group s, must shift their demand for a wife to other (female) age groups.

§ 3D. Presumably a verbally formulated bisexual theory of marriage formation would include these as well as subtler effects. In that case it may be nice to have the aid of a precise concept like the function $\nu_{ij}(\underset{\sim}{m},\underset{\sim}{k})$ when the theory is established.

Perhaps such a theory is all that we can expect. It would not be sufficient, however, if a probabilistic characterization of $\underset{\sim}{M}(t)$, $\underset{\sim}{K}(t)$, and $\underset{\sim}{N}(t)$ is desired, as it would be e.g. in connection with a forecasting model. This would require the specification of a functional form for $\nu_{ij}(\underset{\sim}{m},\underset{\sim}{k})$. To give such a specification seems quite a task. In addition it is probably outside our reach to overcome the mathematical difficulties involved in the further analysis of $\underset{\sim}{M}(t)$, $\underset{\sim}{K}(t)$, and $\underset{\sim}{N}(t)$. It has seemed worthwhile nevertheless to give the formulation above so that we can see what the analyst is up against.

§ 3E. The concepts of § 3B invite a further interpretation of some functions of $\nu_{ij}(\underset{\sim}{m},\underset{\sim}{k})$. Let $\nu_{i.}(\underset{\sim}{m},\underset{\sim}{k})=\sum_{j=1}^{J}\nu_{ij}(\underset{\sim}{m},\underset{\sim}{k})$, $\nu_{.j}(\underset{\sim}{m},\underset{\sim}{k})=\sum_{i=1}^{I}\nu_{ij}(\underset{\sim}{m},\underset{\sim}{k})$, $r_{ij}(\underset{\sim}{m},\underset{\sim}{k})=\nu_{ij}(\underset{\sim}{m},\underset{\sim}{k})/\nu_{i.}(\underset{\sim}{m},\underset{\sim}{k})$, and $s_{ij}(\underset{\sim}{m},\underset{\sim}{k})=\nu_{ij}(\underset{\sim}{m},\underset{\sim}{k})/\nu_{.j}(\underset{\sim}{m},\underset{\sim}{k})$. Then $\nu_{i.}(\underset{\sim}{m},\underset{\sim}{k})\Delta t+o(\Delta t)$ is the probability of observing a marriage in $\langle t,t+\Delta t]$ with the bridegroom in age group i, given that $\mathbf{M}(t)=\mathbf{m}$ and $\mathbf{K}(t)=\mathbf{k}$. $r_{ij}(\mathbf{m},\mathbf{k})$ is the probability that the bride belongs to age group j, given that such a marriage has occurred and conditional on $\mathbf{M}(t)=\mathbf{m}$ and $\mathbf{K}(t)=\mathbf{k}$. $\nu_{.j}(\mathbf{m},\mathbf{k})$ and $s_{ij}(\mathbf{m},\mathbf{k})$ have similar interpretations.

§ 3F. An inspection of the argument in the present chapter shows that the definition of an age group introduced in chapter 2 is not essential to the general theory. If desired it may therefore be dropped in favour of the more common usage of the term. (The results of chapter 2 build on our special definition.)

4. Extensions

In the previous chapter we consciously disregarded the fact that a person who has been married, may return to the "marriage market" in seach of a new spouse after dissolution of the marriage. (Such a return would at the same time constitute an offer to supply a new "good" in substitution of a bachelor or a spinster, as the case may be.) Introduction of such returns poses no real difficulty for the concept formation. It is not essential to the theory that the indices i and j refer to age groups, the main point is the distinction between population groups with different mortality or nuptiality parameters. Thus i and j may be chosen to simultaneously represent age group and marital status. In principle other dimensions, like social status, area of residence, etc., may be incorporated similarly.

The restriction to a closed population is superfluous and may be removed.

5. Acknowledgement

I am grateful to Mr. Lars Bastiansen, who has proof-read the paper in manuscript.

6. Reference

Yntema, L. (1954): "Demographic Extensions of the Simple Birth- and Death-Process." Het Verzekerings-Archief, Actuarieel Bijvoegsel, **31**, 6*-20*.

A Probabilistic Approach to Nuptiality

by Jan M. Hoem.

CONTENTS

1. Introduction

§ 1A. In the present paper a hierarchic time-continuous age- and duration-dependent Markov process will be suggested as a nuptiality model. In the model each person is characterized at each moment by age, marital status, duration of current status, and number of marriages experienced.

We shall proceed as follows: The complete nuptiality model will be presented in chapter 2. Various sections of the model will then be described and discussed in chapters 3 to 5. In chapter 6 some further results on this model will be listed. Finally chapter 7 contains a suggestion for modifications of the model to comply with common data, as well as an indication of how nuptiality and fertility could be studied simultaneously.

Estimation problems have been treated in previous papers (Hoem, 1968a and b) and will not be considered here.

§ 1B. The model presented is unisexual, i.e. only one sex is explicitly considered, except in one segment (marriage dissolution), where both sexes enter symmetrically during part of the argument. For the most part it is not necessary to specify which sex is considered, and we shall generally consider "persons" without further qualification. It is then understood that this refers to persons of a single but unspecified sex.

§ 1C. In a unisexual model the opposite sex enters only indirectly as a kind of shadow factor. In real life, however, the two sexes of course play rather symmetric parts in the marriage process. It would therefore be conceptually more satisfactory if this could be taken care of in the nuptiality model as well (Henry, 1959, pages 12-13). We have previously made an attempt at investigating a bisexual model for marriage formation (Hoem, 1968c). Unfortunately the technical difficulties involved proved insurmountable.

The bisexual approach led to "collective" treatment of the marriageable population in the sense that all persons of the population were considered simultaneously as interacting units. In contrast to this we shall now regard the individual person (or in one segment the individual married couple) as a unit operating independently of the other individuals in the population. Under this approach many of the real problems met with in the bisexual theory are glossed over in the hope that they may prove less important after all. In return the mathematics involved are greatly facilitated, and we are actually able to find a solution within the unisexual model to many questions which went unanswered or even unasked in the bisexual theory.

§ 1D. Among the many previous authors who have studied similar problems we may mention Winkler (1922), Wicksell (1931), Hyrenius (1948), Henry (1959, 1963), and Mertens (1965).

2. The complete nuptiality model

§ 2A. We classify the population by marital status and number of marriages experienced in the following way:

Spinsters and bachelors will be said to be in state M_0. Persons in their k-th marriage will similarly be classified as being in state M_k, for $k = 1,2,\ldots,m$. We will alow at most m marriages to a person, in the sense that marriages after the m-th will not be registered. Of course m will be chosen high enough to cover most marriages.

Persons divorced after their k-th marriage will be classified as in state D_k, and widows and widowers after the k-th marriage are in state W_k, for $k = 1,2,\ldots,m$. Persons who have reached D_m or W_m will be registered as staying there until death or emigration.

Let A be any of the states defined so far. A persons who dies while in state A will be said to move to a state DA. Similarly an emigration while in state A will be reflected in a registered transition into a state EA. Once a person has been registered as entering a state DA, he or she will obviously "stay" there forever after. Transitions into a state EA will be treated in the same way. Thus all states DA and EA are absorbing. All other states are transient.

Immigration into the population will be permitted, but will be determined extraneously to the model. For ease of exposition and although this is not always realistic immigrants will be assumed to have the same properties as the natives, and no distinction will be made in the model. Reimmigration of a previous emigrant is possible and will be treated on a par with an "ordinary" immigration.

§ 2B. As indicated already in connection with states of the form DA and EA, transitions between certain states of the system are possible. In fact this is what makes the model dynamic.

If a person marries for the first time, a transition from M_0 to M_1 will be recorded. A subsequent divorce will be recorded as a transition from state M_1 to D_1. Similarly for the other kinds of events registered.

Let B and C be two distinct states of the system. If it is possible to experience a transition direct from B to C, such as from M_0 to M_1, we shall designate this by writing $B \rightarrow C$. Otherwise $B \not\rightarrow C$.

The states of this system may be represented graphically as in diagram 2.1, where arrows indicate possible direct transitions. (Immigration has not been indicated, but it may occur into any transient state.)

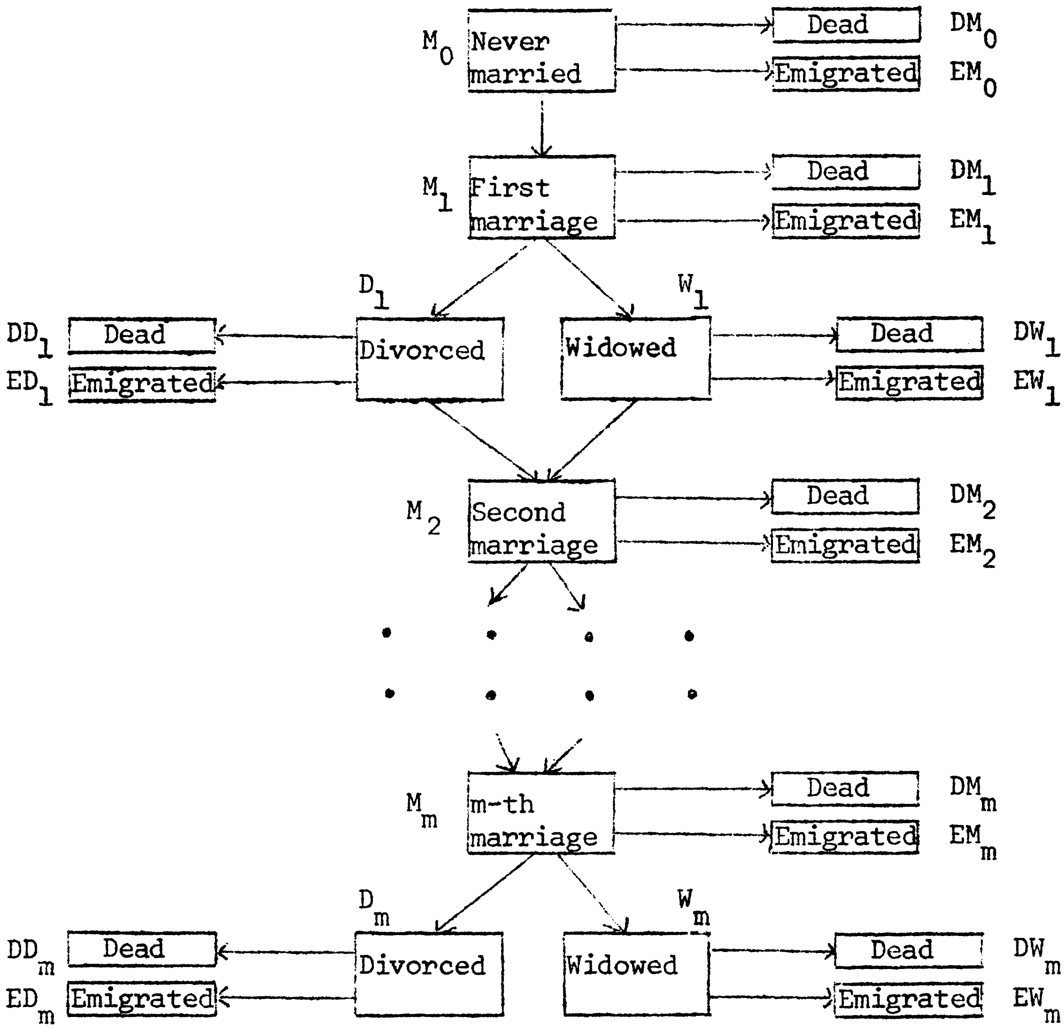


Diagram 2.1.

§ 2C. Even when $B\not\rightarrow C$ it may be possible to reach C from B by transitions via one or more intervening states. If $B\not\rightarrow C$ and C can be reached from B, whether direct or not, we shall write $B\rightsquigarrow C$. Otherwise $B\not\rightsquigarrow C$.

If $B\rightsquigarrow C$ in our model, then $C\not\rightsquigarrow B$. Thus we may call the system of states hierarchical.

§ 2D. Consider a person who at age y+u is in a transient state B and who entered this state at age y. We define $P_{BB}(y,u,t)$ as the probability that this person will stay in state B at least until age y+u+t. For any state C where $B\rightsquigarrow C$ we define $P_{BC}(y,u,t)$ as the probability that this person will be in state C at age y+u+t. (If C is absorbing, this will be taken to mean that state C has been entered within age y+u+t.)

Here $y>0$ unless $B=M_0$. In the latter case, y must obviously equal 0. It is then superfluous and will be suppressed, so that we write e.g. $P_{M_0C}(u,t)$ for $P_{M_0C}(0,u,t)$.

§ 2E. If $B\rightarrow C$, we shall assume that the limit

$$\lim_{t\downarrow 0} P_{BC}(y,u,t)/t$$

exists and is a continuous function of u for any y and u with $0\leq y<y+u<\omega$, where ω as usual designates the highest possible live age. The limit will be called <u>the force of transition</u> from state B to state C at age y+u and duration u.

We introduce the following names and designations for the forces of transition:

$\nu_0(u)$ is the force of primary nuptiality, i.e. the force of transition from state M_0 to M_1.

$\nu_{Dk}(y,u)$ and $\nu_{Wk}(y,u)$ are the forces of remarriage in states D_k and W_k, respectively, for $1\leq k<m$.

$\delta_k(y,u)$ is the force of divorce and $\omega_k(y,u)$ is the force of widowhood, respectively, in state M_k, for $1\leq k\leq m$.

$\mu_0(u)$ and $\eta_0(u)$ are the forces of mortality and emigration, respectively, in state M_0.

$\mu_k(y,u)$, $\eta_k(y,u)$, $\mu_{Dk}(y,u)$, $\eta_{Dk}(y,u)$, $\mu_{Wk}(y,u)$, and $\eta_{Wk}(y,u)$ are the corresponding forces for states M_k, D_k, and W_k, respectively, for $1\leq k\leq m$.

We interpret $\nu_0(u)\Delta u+o(\Delta u)$ as the probability that a u-year old unmarried person will marry within age $u+\Delta u$ and stay in M_1 until this age. The other forces have similar interpretations.

Forces of transition which depend on y and u, separately, rather than on age attained (y+u), will be called <u>select</u>.

§ 2F. Let B be transient, let $B \to C$, and let $\mathcal{C}$ be the set of states consisting of C and all states that can be reached from C. We introduce

$$P_{B\mathcal{C}}(y,u,t) = \sum_{A \in \mathcal{C}} P_{BA}(y,u,t).$$

Dividing by t, letting $t \downarrow 0$, and using the fact that $\lim_{t\downarrow 0} P_{BC}(y,u,t)/t = 0$ if $B \leadsto C$ but $B \not\to C$, we see that $\lim_{t\downarrow 0} P_{B\mathcal{C}}(y,u,t)/t$ equals the force of transition from B to C. We shall need this result at several occasions in later paragraphs.

§ 2G. For any transient state B the limit

$$\lim_{t\downarrow 0} \{1-P_{BB}(y,u,t)\}/t$$

exists and equals the sum of the forces of transition out of B, by the assumptions of § 2E. We shall call this limit <u>the total force of decrement</u> from state B at age y+u and duration u. We introduce the symbols $\alpha_0(u)$, $\alpha_k(y,u)$, $\alpha_{Dk}(y,u)$, and $\alpha_{Wk}(y,u)$ for the total forces of decrement from M_0, M_k, D_k, and W_k, respe   Then

$$\alpha_0(u) = \nu_0(u)+\eta_0(u)+\mu_0(u),$$

$$\alpha_k(y,u) = \delta_k(y,u)+\omega_k(y,u)+\eta_k(y,u)+\mu_k(y,u),$$

and similar relations hold for α_{Dk} and α_{Wk}. We interpret $\alpha_k(y,u)\Delta u+o(\Delta u)$ as the probability that a person in M_k at age y+u and duration u will have left M_k within age $y+u+\Delta u$. Similarly for the other total forces of decrement.

3. First marriage

§ 3A. The section of the model relevant for the never-married persons is given in diagram 3.1, where M_1 and all states which can be reached from M_1

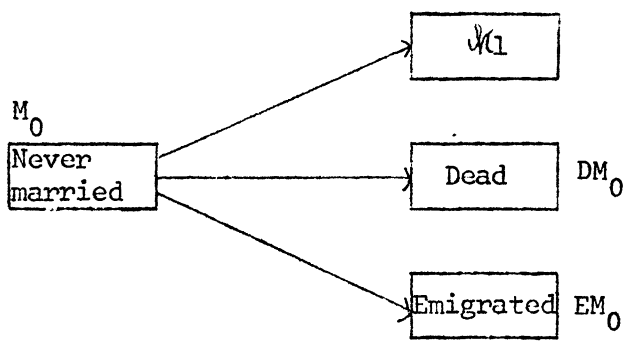


Diagram 3.1.

have been combined into a single state called $\mathcal{M}_1$. We may regard this section as a four-state Markov chain with the transient state M_0 and the absorbing states $\mathcal{M}_1$, DM_0, and EM_0. If we use the symbol x for age attained, the transition probabilities of this chain are

$$P_{M_0M_0}(x,t)=\exp\{-\int_0^t \alpha_0(x+\tau)d\tau\},\quad P_{M_0\mathcal{M}_1}(x,t)=\int_0^t P_{M_0M_0}(x,\tau)\nu_0(x+\tau)d\tau,$$

$$P_{M_0DM_0}(x,t)=\int_0^t P_{M_0M_0}(x,\tau)\mu_0(x+\tau)d\tau,\text{ and } P_{M_0EM_0}(x,t)=\int_0^t P_{M_0M_0}(x,\tau)\eta_0(x+\tau)d\tau.$$

By § 2F the forces of transition are ν_0, μ_0, and η_0.

Since the values of each of these probabilities are influenced by all three forces of transition, we shall call them influenced probabilities. (Sverdrup, 1961).

§ 3B. The sectional model of § 3A represents a common net nuptiality table (Spiegelman, 1955, 8.1.4; Grabill, 1945; Jacobson, 1959, pp. 76-82; Schwarz, 1965) with the addition of emigration as a cause of decrement. The various columns of a nuptiality table arise as follows.

From a cohort of 100 000 persons born alive, 100 000 $P_{M_0M_0}(0,x)$ will be expected to remain alive and single at age x. Out of these, a proportion $P_{M_0DM_0}(x,1)$ will be expected to die while single, a proportion $P_{M_0EM_0}(x,1)$ will be expected to emigrate while single, and a proportion $P_{M_0\mathcal{M}_1}(x,1)$ will be expected to marry for the first time during the age year following age x.

Of single persons at age x an expected proportion of $P_{M_0\mathcal{M}_1}(x,\omega-x)$ will ever marry (where ω is the highest possible live age). Similarly a proportion $P_{M_0\mathcal{M}_1}(0,x)$ of the new-born will be expected to marry within age x. This quantity must not of course be confused with the expected proportion of ever-married persons at age x, nor with the probability of being ever-married at this age.

§ 3C. Although emigration plays some part in the life of most national populations, this fact is commonly disregarded when marriage tables are set up. In fact all published tables known to this author implicitly build on a model where $\eta_0 \equiv 0$. If we let $\bar{\alpha}_0(x) = \nu_0(x)+\mu_0(x)$, $\bar{P}_{M_0M_0}(x,t) = \exp\{-\int_0^t \bar{\alpha}_0(x+\tau)d\tau\}$, $\bar{P}_{M_0\mathcal{M}_1}(x,t) = \int_0^t \bar{P}_{M_0M_0}(x,\tau)\nu_0(x+\tau)d\tau$, and $\bar{P}_{M_0DM_0}(x,t) = \int_0^t \bar{P}_{M_0M_0}(x,\tau)\mu_0(x+\tau)d\tau$, the $\bar{P}_{AB}(x,t)$ will represent the transition probabilities of this implicit model. Since the values of all $\bar{P}_{AB}(x,t)$ are influenced by both ν_0 and μ_0, but not by η_0, we shall call them <u>semi-influenced probabilities</u>. If the nuptiality and the mortality among the emigrants do not differ from those of the persons remaining in the population, the $\bar{P}_{AB}(x,t)$ may be given an interpretation in the model section of § 3A. While $P_{M_0\mathcal{M}_1}(x,t)$ is the probability of marrying within age x+t <u>while a member of the population</u>, $\bar{P}_{M_0\mathcal{M}_1}(x,t)$ is the probability of marrying within age x+t, no matter whether this occurs before or after a possible emigration. The other semi-influenced probabilities have similar interpretations. (For a proof, see Hoem (1968d).)

§ 3D. In a theory for the formation of first marriages, interest centers on transitions from M_0 to $\mathcal{M}_1$, and the decrements due to death and emigration represent elements of nuisance (Henry, 1959, 1963; Mertens, 1965). One may wish to produce some measure of nuptiality which is free from the influence of mortality and emigration. For this $\nu_0(x)$ is an obvious choice. It is sometimes desired to have a measure with the dimensions of a probability, however. Since both $P_{M_0\mathcal{M}_1}(x,t)$ and $\bar{P}_{M_0\mathcal{M}_1}(x,t)$ are obviously inadequate, multiple decrement theory then offers the <u>partial probability</u>

$$\bar{\bar{P}}_{M_0\mathcal{M}_1}(x,t) = 1-\exp\{-\int_0^t \nu_0(x+\tau)d\tau\} = 1-\bar{\bar{P}}_{M_0M_0}(x,t),$$

which represents the chances which a single x-year-old has of getting married whitin age x+t provided mortality and emigration are inoperative in the meantime. Gross nuptiality tables are based on this function.

Unfortunately $\bar{\bar{P}}_{M_0\mathcal{M}_1}(x,t)$ has no interpretation as a probability in the nuptiality model of § 3A. Instead it is a transition probability in a <u>different</u> Markov chain, viz. one with the two states M_0 and $\mathcal{M}_1$ only, and with ν_0 as the force of transition from M_0 to $\mathcal{M}_1$.

§ 3E. Let T be the remaining lifetime in state M_0 of a single person of age x. Then T is a random variable. Its probability distribution is

$$F(x,t) = 1-P_{M_0M_0}(x,t) \text{ in the model of § 3A,}$$

$$\bar{F}(x,t) = 1-\bar{P}_{M_0M_0}(x,t) \text{ in the model of § 3C, and}$$

$$\bar{\bar{F}}(x,t) = 1-\bar{\bar{P}}_{M_0M_0}(x,t) \text{ in the model of § 3D.}$$

Since $\bar{\bar{P}}_{M_0M_0}(x,\omega-x)$ will presumably be positive, $\bar{\bar{F}}(x,t)$ is an improper probability distribution.

We shall designate the mean values of T in the two first of these distributions by $\mathring{e}_0(x)$ and $\bar{e}_0(x)$, respectively. Then

$$\mathring{e}_0(x) = \int_0^{\omega-x} P_{M_0M_0}(x,t)dt = \int_0^{\omega-x} \{1-F(x,t)\}dt,$$

and a similar formula holds for $\bar{e}_0(x)$. Specifically $\mathring{e}_0(0)$ is the expected lifetime as a single person for a new-born baby. A mean value for T in the distribution $\bar{\bar{F}}$ will be introduced in § 3F.

§ 3F. The mean age at first marriage (of a given sex) will be calculated from population data by adding together ages at first marriages registered and dividing by the corresponding number of first marriages. The quantity ensuing is best interpreted as an estimate of a corresponding quantity within our probability model. One might perhaps believe that this estimand is either $\mathring{e}_0(0)$ or $\bar{e}_0(0)$, since mean age at first marriage is obviously equal to mean waiting time until first marriage for a new-born. It is noteworthy that neither of these suggestions is correct.

To get at the right estimand, we start by noting that $P_{M_0\mathcal{K}_1}(x,\omega-x)$ represents the probability that an x year old single person will ever get married. Given that such a person ever does get married,

$$F^*(x,t) = P_{M_0\mathcal{K}_1}(x,t)/P_{M_0\mathcal{K}_1}(x,\omega-x)$$

is the probability that this happens within age x+t. Thus $F^*(x,.)$ will be the distribution function of the waiting time until the marriage, and

$$e_0^*(x) = \int_0^{\omega-x} \{1-F^*(x,t)\}dt$$

is the mean value of this distribution. The quantity estimated by the observed mean age at marriage is $e_0^*(0)$.

Correspondingly, the median age at marriage $x+m_0^*(x)$ for an x-year-old single person who will marry, is defined by the relation $F^*\{x,m_0^*(x)\} = \frac{1}{2}$, or equivalently

$$P_{M_0\mathcal{K}_1}\{x,m_0^*(x)\} = \tfrac{1}{2}P_{M_0\mathcal{K}_1}(x,\omega-x).$$

$e_0^*(x)$ and $m_0^*(x)$ are influenced measures. Corresponding partial measures are $\bar{\bar{e}}_0(x)$ and $\bar{\bar{m}}_0(x)$, defined by

$$\bar{\bar{e}}_0(x) = \int_0^{\omega-x} \{1-\bar{\bar{F}}(x,t)/\bar{\bar{F}}(x,\omega-x)\}dt,$$

and $$\bar{\bar{P}}_{M_0\mathcal{K}_1}(x,\bar{\bar{m}}_0(x)) = \tfrac{1}{2}\bar{\bar{P}}_{M_0\mathcal{K}_1}(x,\omega-x).$$

§ 3G. Considerations like those of §§ 3C to F apply in each of the sectional models of the chapters below. To avoid being repetitive we shall not formulate them in each case.

4. Marriage dissolution

§ 4A. We shall open this chapter by suggesting a bisexual model for marriage dissolution, both because such a model has some independent interest, and also because it provides an introduction to the unisexual model, to which we shall return in § 4B.

If we want to preserve the feature of emigration in a bisexual model, we must take account of the fact that the spouses need not emigrate simultaneously. This will cause some trivial complication which have nothing to do with our present argument. We shall therefore disregard the possibility of emigration in §§ 4A and B.

Consider then a married couple where the husband is in his k-th and the wife is in her n-th marriage. Let ut describe this by saying that the couple is in state (M_k, M_n). If the husband dies while they are still married, the couple will be said to move to state $(DM_k, \mathcal{W}_n)$. Similarly they move to state $(\mathcal{W}_k, DM_n)$ if the wife dies and to state $(\mathcal{D}_k, \mathcal{D}_n)$ if they get divorced.

Let x and y denote the age at marriage of the husband and wife, respectively, and let u denote the current duration of their marriage. Transition probabilities will now have the form $P_{(M_k,M_n),(A,B)}(x,y,u,t)$. We define the following forces of transition from (M_k, M_n):

to $(DM_k, \mathcal{W}_n): \mu_k(x,u)$,

to $(\mathcal{W}_k, DM_n): \mu_n'(y,u)$, and

to $(\mathcal{D}_k, \mathcal{D}_n): \sigma_{kn}(x,y,u)$.

Thus the male force of mortality μ_k is assumed to be independent of characteristics of the wife. Similarly for the female force of mortality μ_n'.

Forces of mortality are well known to depend on marital status, (see e.g. Jacobson, 1959, p. 139), so we should at least distinguish between the forces of mortality for the single, the married, the widowed, and the divorced. Intuitively one would expect such forces to depend on the number of previous marriages as well, as we have specified.

The common explanation of the lower mortality generally observed in the married population than among single persons is twofold. In the first place, marriage is thought to be selective as regards both physical constitution and social adaptability. Secondly married life is considered to provide a better environment due to the greater regularity of living (Thompson and Lewis, 1965, pp. 364-368).

To the extent that the second explanation is correct, one would expect the effect of married life on mortality to be gradually more pronounced as the marriage proceeds. Thus there is a reason to use select forces of mortality for the married persons. No harm is done in doing so even if these forces prove to be non-select, for in that case one simply has $\mu_k(x,u) = \mu_k(x+u)$, and similarly for μ'_n.

There is every reason to believe that the force of divorce σ_{kn} is select (Jacobson, 1959, p. 149).

Formulae for the transition probabilities are established quite easily and will be omitted.

§ 4B. We now turn to the unisexual model. For an arbitrary woman in M_n whose age at last marriage was y and whose current marriage duration is u we define a random variable X as the age at marriage of her present husband, and another random variable K as the number of marriages which he has experienced. The $P_{(M_k,M_n),(A,B)}(x,y,u,t)$ defined in § 4A may now be regarded as conditional probabilities, given that X=x and K=k.

To characterize K and the age difference X-y we introduce the distribution

$$G_{kn}(z,y) = P\{X-y \leq z, K=k\}.$$

We then see that the transition probabilities for the wife have the form

$$(4.1) \qquad P_{M_n B}(y,u,t) = \sum_{k=0}^{m} \int P_{(M_k,M_n),(A,B)}(x,y,u,t)\, d_x\, G_{kn}(x-y,y)$$

where A is taken as M_k if $B=M_n$, $A=DM_k$ if $B=\mathcal{W}_n$, $A=\mathcal{W}_k$ if $B=DM_n$, and $A=\mathcal{D}_k$ if $B=\mathcal{D}_n$. The integral is taken over all possible values of x.

Dividing by t in (4.1) and letting $t\downarrow 0$, we get

$$(4.2) \qquad \omega_n(y,u) = \sum_{k=0}^{m} \int \mu_k(x,u)\, d_x\, G_{kn}(x-y,y), \text{ and}$$

$$(4.3) \qquad \delta_n(y,u) = \sum_{k=0}^{m} \int \sigma_{kn}(x,y,u)\, d_x\, G_{kn}(x-y,y),$$

while the force of female mortality will naturally turn out to be the $\mu'_n(y,u)$ of § 4A.

In the general case, these forces will be select, as one would expect. One may be more surprised to find that <u>the force ω_n of widowhood may be select even when the male force of mortality is nonselect and independent of marriage number</u>, as we now proceed to show:

Let $G_n(z,y)=P\{X-y\leq z\} = \sum_{k=0}^{m} G_{kn}(z,y)$, and assume that $\mu_k(x,u)=\mu(x+u)$. Then

$$(4.4) \qquad \omega_n(y,u) = \int \mu(y+u+z)d_z\, G_n(z,y),$$

which will depend on y and u, separately, rather than only on y+u, as long as the distribution function G_n genuinely depends on y. As long as the age difference X-y at marriage has a distribution which depends on the age y of the bride (Backer, 1965, p. 49), ω_n will therefore be select.

§ 4C. Of course the roles of the two sexes may be interchanged in the unisexual model of § 4B. The introduction of emigration is trivial. This gives the section pertaining to the persons in M_n in the general model of chapter 2.

5. Remarriage

§ 5A. For the persons divorced or widowed after their k-th marriage (k<m) the relevant model section is given in diagram 5.1.

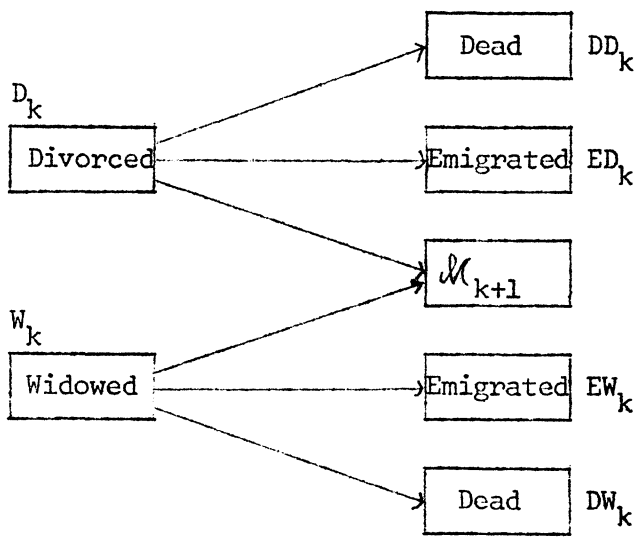


Diagram 5.1.

The transition probabilities of this chain are

$$P_{D_kD_k}(y,u,t) = \exp\{-\int_0^t \alpha_{Dk}(y,u+\tau)d\tau\},$$

$$P_{D_k\mathcal{M}_{k+1}}(y,u,t) = \int_0^t P_{D_kD_k}(y,u,\tau)\nu_{Dk}(y,u+\tau)d\tau,$$

$$P_{W_kDW_k}(y,u,t) = \int_0^t P_{W_kW_k}(y,u,\tau)\mu_{Wk}(y,u+\tau)d\tau,$$

and so on.

§ 5B. There is a lot of evidence to the effect that forces of remarriage depend on age at dissolution of (latest) marriage and duration since dissolution separately (Roeber and Marshall, 1933; Pressat, 1956; Niessen, 1960; Clarke, 1960). There are also many reports that remarrige rates among divorced men and women differ from those of widowed persons (Spiegelman, 1955, p. 137 ; Jacobson, 1959, pp. 70, 83-86; Backer, 1965, p. 60; Schwarz, 1965). Together with mortality considerations similar to those of § 4A this has notivated our model.

6. Some further results on the hierarchical model

§ 6A. While we have hitherto studied only sections of the nuptiality model of chapter 2, we shall now list a few results pertaining to the complete model. These results are self-evident, and they will be given without explanation.

§ 6B. For any C in $\mathcal{M}_1$,

$$P_{M_0C}(x,t) = \int_0^t P_{M_0M_0}(x,\tau)\nu_0(x+\tau)\, P_{M_1C}(x+\tau,0,t-\tau)d\tau.$$

For k=1,2,...,m,

$$P_{M_kD_k}(y,u,t) = \int_0^t P_{M_kM_k}(y,u,\tau)\delta_k(y,u+\tau)P_{D_kD_k}(y+u+\tau,0,t-\tau)d\tau,$$

and similarly if W_k is substituted for D_k. For any C which can be reached from D_k or W_k,

$$P_{M_kC}(y,u,t)=\int_0^t P_{M_kM_k}(y,u,\tau)\{\delta_k(y,u+\tau)P_{D_kC}(y+u+\tau,0,t-\tau)+\omega_k(y,u+\tau)P_{W_kC}(y+u+\tau,0,t-\tau)\}d\tau.$$

Similar relations are easily established for all relevant $P_{D_kC}(y,u,t)$ and $P_{W_kC}(y,u,t)$.

§ 6C. Consider a person who is in state A at age y+u and who entered this state at age y. The probability that this person will experience k marriages altogether within the population equals

$$\rho_k(y,u,A)=P_{A,DM_k}(y,u,\omega-y-u)+P_{A,EM_k}(y,u,\omega-y-u)+P_{A,DD_k}(y,u,\omega-y-u)+P_{A,ED_k}(y,u,\omega-y-u)$$

$$+ P_{A,DW_k}(y,u,\omega-y-u)+P_{A,EW_k}(y,u,\omega-y-u),$$

and the corresponding mean number of marriages equals $\sum_k k\rho_k(y,u,A)$. Various kinds of "partial means" corresponding to the reasoning of §§ 3C and D may be defined.

This concludes our study of the hierarchical nuptiality model.

7. Suggestions for modifications of the model

§ 7A. The model of chapter 2 may be modified in many directions to serve various purposes. To indicate some of the possibilities, we shall suggest two sets of such modifications, although without going into any detail.

§ 7B. It is one of the nice features of the above model that it is easy to establish simple formulae for the transition probabilities, and that one easily sees the intuitive content of these formulae. Some further simplification will result if the feature of select forces of transition is dropped so that the attained age y+u may be substituted for the pair of variables (y,u) everywhere. In fact non-select forces will probably be used more often (even though this gives a less "realistic" model) because it is more expedient or because the data contain no information about duration and thus make estimation of select forces impossible.

The simplicity of the mathematics of this model rests heavily on the fact that it is hierarchical. In a nuptiality model where marriage dissolution and remarriage may occur, this feature is provided by the registration of the number k of marriages experienced for each person. If this variable is deleted, a non-hierarchical model results, and in such a model no closed formulae can be established for the transition probabilities in the general case.

The sample space of such a nuptiality model has been indicated in figure 7.1. The forces of transition of this model may be regarded as averages of corresponding forces in an underlying, more refined model, much in the same way as δ_n and ω_n in § 4B (Stolnitz and Ryder, 1949).

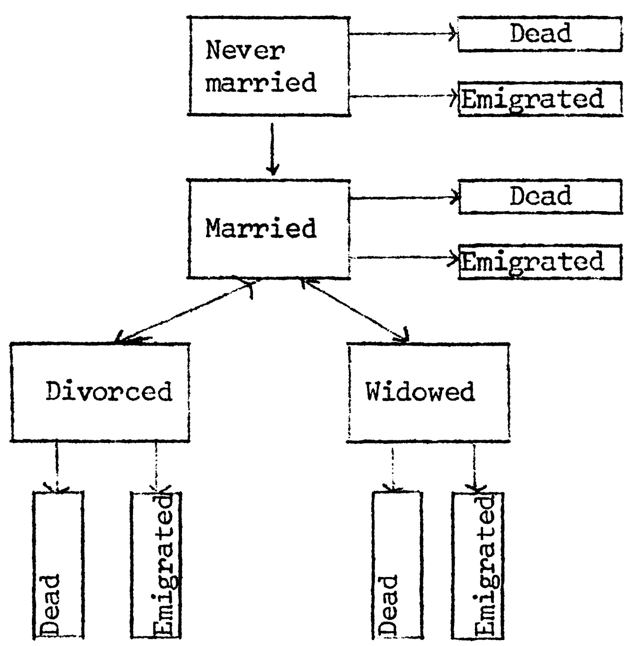


Figure 7.1.

§ 7C. One often wishes to study nuptiality in conjunction with some other phenomenon, such as fertility, migration, social status, or religious denomination. As an example we shall suggest some basic concepts of a simultaneous study of nuptiality and fertility. For simplicy we shall use non-select forces of transition, although in the case of fertility there are stronger arguments than ever for using select forces.

Let some nuptiality model with state space I be given. Although I will contain states like "dead" and "emigrated", we shall call any state in I a marital state. We define $P_{ij}(x,t;k,n)$ as the probability that a person who at age x has had k births and is in marital state i, will have n-k further births within age x+t ($n \geq k$), and will be in marital state j at that age. We introduce the force of fertility at age x in marital state i and with k births as the quantity

$$\phi_{ik}(x) = \lim_{t \downarrow 0} P_{ii}(x,t;k,k+1)/t,$$

and a corresponding force of change of marital status

$$\lambda_{ijk}(x) = \lim_{t \downarrow 0} P_{ij}(x,t;k,k)/t.$$

Of course many λ_{ijk} will be identically equal to zero, since direct transition is impossible between many marital states.

A simultaneous analysis of nuptiality and fertility then consists in a study of the forces ϕ_{ik} and the positive λ_{ijk}.

8. Acknowledgement

I am grateful to Mr. Tore Schweder, who has proof-read this paper in manuscript.

9. References.

[1] Backer, Julie E. (1965): "Marriages, Births and Migrations in Norway, 1856-1960." Oslo: Statistisk Sentralbyrå.

[2] Clarke, E.E. (1960): "Remarriage Experience under the Pension Act of Canada." (With discussion.) Trans.Society of Actuaries, 12, 449-465.

[3] Dahlberg, Gunnar (1951): "Marriage and Divorce Risks." Acta Genetica et Statistica Medica, 2, 318-330.

[4] Grabill, W.H (1945): "Attrition Life Tables for the Single Population." Journ.Am.Stat.Assn. 40, 364-375.

[5] Henry, Louis (1959): "D'un problème fondamental de l'analyse démographique." Population, 14, 9-32.

[6] Henry, Louis (1963): "Approximations et erreurs dans les tables de nuptialité de générations." Population, 18, 737-776.

[7] Hoem, Jan M. (1968a): "Time-Continuous Markov Chain Estimation Techniques in Demographic Models." Under publication.

[8] Hoem, Jan M. (1968b): "A Probabilistic Model for Primary Marital Fertility." Under publication.

[9] Hoem, Jan M. (1968c): "Concepts of a Bisexual Theory of Marriage Formation." Under publication.

[10] Hoem, Jan M. (1968d): "Purged and Partial Markov Chains." Under publication.

[11] Hyrenius H.(1948): "La mesure de la reproduction et de l'accroissement naturel." Population, 3, 271-292.

[12] Jacobson, Paul H. (1959): "American Marriage and Divorce." Rinehart & Co., Inc., New York.

[13] Mertens, Walter (1965): "Methodological Aspects of the Construction of Nuptiality Tables." Demography, 2, 317-348.

[14] Niessen, A.M. (1960): "A New Remarriage Table." (With discussion.) Trans.Society of Actuaries, 12, 1-21.

[15] Pressat, Roland (1956): "Le remariage des veufs et des veuves." Population, 11, 47-58.

[16] Roeber, W.F. and R.M. Marshall (1933): "An American Remarriage Table." Proc.Casualty Actuarial Society, 19, 272-349.

[17] Schwarz, Karl (1965): "Heiratstafeln für Ledige, Verwitwete und Geschiedene, 1960/62." Wirtschaft und Statistik, 11, 709-715, 730*-733*.

[18] Spiegelman, M.(1955): "Introduction to Demography." The Society of Actuaries.

[19] Stolnitz, G.J. and N.B. Ryder (1949): "Recent Discussion of the Net Reproduction Rate." Population Index, 15, 114-128. Reprinted in J.J. Spengler and O.D. Duncan (1956): "Demographic Analysis. Selected Readings." The Free Press, Glencoe.

[20] Sverdrup, Erling (1961): "Statistiske metoder ved dødelighetsundersøkelser." Mimeographed, Institutt for matem.fag, Univ. i Oslo.

[21] Thompson, W.S. and D.T. Lewis (1965): "Population Problems." Mc Graw-Hill Book Co.

[22] Wicksell, S.D. (1931): "Nuptiality, Fertility, and Reproductivity." Skandinavisk Aktuarietidskrift 1931, 125-157.

[23] Winkler, Wilhelm (1922): "Neuere Heiratstafeln." Jahrbücher für Nationalekonomie und Statistik, 118 (III, 63), 64-70.

TIME-CONTINUOUS MARKOV CHAIN ESTIMATION TECHNIQUES IN DEMOGRAPHIC MODELS

B y

Jan M. Hoem

CONTENTS

A previous draft of this paper has been mimeographed as a Memorandum dated April 24, 1968, from the Institute of Economics at the University of Oslo.

1. Introduction

§ 1.1. Time-continuous Markov chain models with a finite or countably infinite number of states can be applied to many problems in demographic analysis. If i and j are two states (not necessarily distinct) of such a chain, let $P_{ij}(x,t)$ be the probability that a person who is in state i at age x, will be in state j at age x+t. For $i \neq j$ we define

$$\mu_{ij}(x) = \lim_{t \downarrow 0} P_{ij}(x,t)/t$$

provided a finite limit exists. We shall call this quantity <u>the force of transition</u> from state i to state j at age x. Then $\mu_{ij}(x)\Delta x + o(\Delta x)$ may be interpreted as the probability that a person who at age x is in state i, will move to state j within age $x+\Delta x$.

As indicated by the notation, the force of transition may depend on the age of the person in question. In demographic applications of models which should be realistic over broader age ranges, it is probably nearly always necessary to use age-dependent forces of transition. In applications to a restricted age interval, however, it may often be possible to remove the technical complication of age-dependence and to work with models where the forces of transition are constant parameters. In fact, even the calculation of such quantities as age-specific fertility and mortality rates may be regarded as an application of <u>age-homogeneous</u> time-continuous Markov chain estimation techniques to each of a number of age classes. The parameter values are then assumed to be constant within each age class, but they may differ from one class to another. After the estimation of the parameter values of each class, they are sometimes graduated by some method, such as the Gompertz-Makeham procedures in mortality investigations.

In the present paper we shall concentrate on age-homogeneous models. Omitting the age x from our notation, we let $P_{ij}(t)$ be the probability that a person will be in state j at time $t>0$, given that he is in state i at time zero. Similarly the forces of transition are $\mu_{ij} = \lim_{t \downarrow 0} P_{ij}(t)/t$ for $i \neq j$. (For existence theorems, see Chung (1960).)

§ 1.2. We shall start by giving two examples of simple age-homogeneous Markov chain models with applications in demography. We shall then formulate a general model which will have these examples and a great many others as special cases. In the general model we shall derive maximum likelihood estimators and shall investigate some of their large-sample properties. Finally we shall give some consideration to parameter transformations.

§ 1.3. Estimation of the forces of transition has been studied by Zahl (1955), Meier (1955), Billingsley (1961), Albert (1962), Sverdrup (1965), and others. Many of our results are straightforward generalizations of similar results given by Sverdrup (1965) for the special case of a three-state disability process.

2. Introductory examples

§ 2.1. (Work-force participation.)

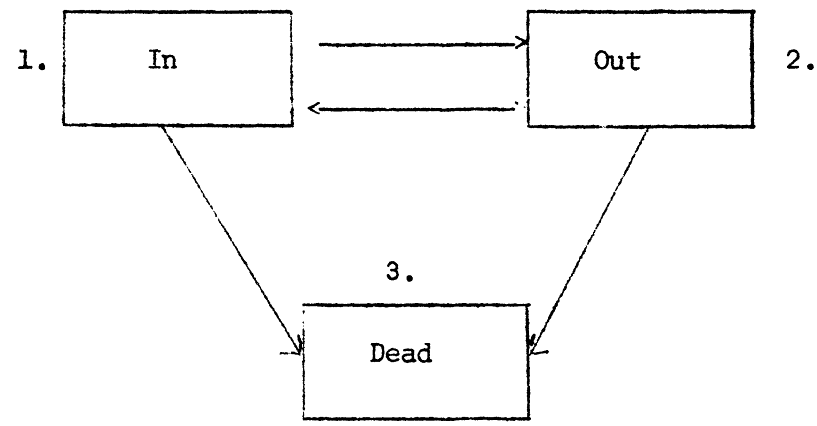


Figure 2.1.

Our first example will be a three-state time-homogeneous Markov chain, which we suggest as a model for work-force participation. A person is said to belong to state 1 if he is a member of the work-force, he will belong to state 2 if he does not work, and he moves to state 3 if he dies. Possible transitions are suggested by arrows in figure 2.1. We assume that all forces of transition between the states may be regarded as constant parameters for the age range studied. These forces are

the force of "unemployment" ν,
the force of "re-employment" ρ, and
the force of mortality μ.

It will be understood that the names of these parameters have been introduced as mnemotechnical devices only, and that e.g. the definition of an unemployed person does not have to coincide with that of ordinary labour force statistics.

For various reasons we have defined a single force of mortality valid for persons both in state 1 and 2. Thus we assume that there is a common mortality for the employed and the unemployed. It is not difficult to generalize to a model in which a distinction is made between the mortality of the two groups (Sverdrup, 1965).

We introduce

$$\bar{P}_{11}(t) = (\rho+\nu e^{-(\nu+\rho)t})/(\nu+\rho),\ \bar{P}_{12}(t) = \nu(1-e^{-(\nu+\rho)t})/(\nu+\rho),$$

$$\bar{P}_{21}(t) = \rho(1-e^{-(\nu+\rho)t})/(\nu+\rho), \text{ and } \bar{P}_{22}(t) = (\nu+\rho e^{-(\nu+\rho)t})/(\nu+\rho).$$

It may be shown that $P_{13}(t) = P_{23}(t) = 1-e^{-\mu t}$, and that

$$P_{ij}(t) = \bar{P}_{ij}(t)e^{-\mu t} \quad \text{for i and j} = 1,2. \tag{1}$$

Here $\bar{P}_{ij}(t)$ would be the probability that a person will be in state j at time t, given that he starts in state i at time zero, provided there is no mortality, so that $\mu=0$. By analogy with the terminology from mortality investigations, we may call the $\bar{P}_{ij}(t)$ partial probabilities. In contrast to the $\bar{P}_{ij}(t)$, the $P_{ij}(t)$ are influenced by the value of μ, and we may call the $P_{ij}(t)$ influenced probabilities. (This terminology is a natural generalization of definitions due to Du Pasquier (1913) and Sverdrup (1961). In more common but less fortunate terminology, the $\bar{P}_{ij}(t)$ and the $P_{ij}(t)$ would be called independent and dependent probabilities, respectively.)

Formula (1) shows that the $P_{ij}(t)$ are derived from the $\bar{P}_{ij}(t)$ by means of a simple mortality correction. This nice property is due to the fact that we consider mortality to be equal for the employed and the unemployed. When a distinction is made between the mortality of the two groups, no such simple relation exists.

§ 2.2. (A simple fertility model.)

In applications to a narrow age range, it may not be too unrealistic to assume that fertility and mortality are age-independent. Various models will then be applicable. Hoem (1968) studies in some detail a particularly simple one, which we shall use more briefly as our second example.

Let $P_k(t)$ be the probability that a parent will have k births during time $[0,t]$ and still be alive at time t, and let $Q_k(t)$ be the probability that a parent will have k births during time $[0,t]$ and die within time t. We assume that mortality is independent of the number of births experienced, and let μ be the force of mortality. Similarly we assume that fertility is independent of the parity and spacing of the births, and designate the force of fertility by ϕ. Thus, no matter how many births a parent has had during $[0,t]$ and regardless of when they have arrived, the probability of another birth during the time interval $\langle t,t+\Delta t]$ is $\phi\Delta t+o(\Delta t)$, provided the parent is alive at time t. It may be shown (Hoem, 1968, § 4.1) that

$$P_k(t) = \frac{(\phi t)^k}{k!} e^{-(\mu+\phi)t} \qquad \text{for } k = 0,1,\ldots, \text{ and}$$

$$Q_k(t) = \frac{1}{\phi} \left(\frac{\mu\phi}{\mu+\phi}\right)^{k+1} \left\{1 - \sum_{\nu=0}^{k} \frac{1}{\nu!} \left(\frac{(\mu+\phi)t}{\mu}\right)^{\nu} e^{-(\mu+\phi)t}\right\} \qquad \text{for } k = 0,1,\ldots .$$

In the present model there is a double infinity of states. We may say that a parent is in state (k,1) at time t if the parent has had k births during $[0,t]$ and is alive at time t, and that the parent is in state (k,2) at time t if he or she has had the same number of births during $[0,t]$ and is no longer alive at time t. Then $P_k(t) = P_{(0,1),(k,1)}(t)$, and $Q_k(t) = P_{(0,1),(k,2)}(t)$. One-step transitions are possible from (k,1) only to (k+1,1) and to (k,2). The forces of these transitions are ϕ and μ, respectively.

3. A general model

§ 3.1. Let us study a time-continuous finite or countably infinite Markov chain with constant forces of transition. To get a reasonable model, we shall make the standard assumptions that $\sum_j P_{ij}(t) = 1$ for all $t \geq 0$, and $\lim_{t \downarrow 0} P_{ij}(t) = \delta_{ij}$, where δ_{ij} is a Kronecker delta.

The force of decrement from any state i will be defined by $\mu_i = \lim_{t \downarrow 0} \{1-P_{ii}(t)\}/t$. We shall assume that for all i, $0 \leq \mu_i < \infty$ and

$$\mu_i = \sum_{j \neq i} \mu_{ij}. \tag{2}$$

Such assumptions automatically hold in a finite-state chain, but not necessarily in a chain with a countable infinity of states.

A state is called absorbing if $\mu_i = 0$. It is impossible to leave an absorbing state.

§ 3.2. Intuitively an individual sample path may be visualized as a person moving through (some of) the states of the system. We shall consider a situation where K persons are kept under observation, not necessarily simultaneously, nor need the periods of observation of all persons be equally long. For convenience we shall designate the period of exposure of person no. k by the time interval $[0,z_k]$, where the exposure time z_k for the time being will be taken as a preassigned finite positive number. This should be understood as follows: At some time z_k' (measured by "ordinary" time, e.g. in years and parts of a year A.D.) we start observing what happens to this person, and we go on observing right up to time $z_k' + z_k$. At this moment observation ends, either because the observer decides to stop, or because the person under observation is removed from observation by some cause extraneous to our model. Periods spent in absorbing states (e.g. after the death of the person observed) is included in the interval $[z'_k, z'_k + z_k]$, although of course no actual observation is made after a person has entered such a state.

In many cases there is a certain state of the chain from which all sample paths must start at time zero. In our fertility model (§ 2.2), for instance, every path starts in state (0,1). This feature is not common to all the models which we wish to cover by our theory, however, and we shall designate by r_k the state from which sample path no. k starts out. For the time being we shall take also all r_k as preassigned.

The sample paths corresponding to the K persons will be assumed independent.

§ 3.3. Under the assumptions mentioned above, it is possible to prove that the number of states visited (and thus the number of transitions experienced) by any person during his period of observation, is finite with probability 1. Let $N_k(i,j)$ be the number of transitions direct from state i to state j experienced by person no. k, and let $U_k(i)$ be the total time spent by him in state i. Then

$$P\{\bigcap_{k=1}^{K} \bigcap_{i} \bigcap_{j} [N_k(i,j) = n_k(i,j) \text{ and } u_k(i) < U_k(i) < u_k(i) + du_k(i)]\}$$
$$= \exp\{-\sum_i \mu_i u(i)\} \prod_i \{\prod_j \mu_{ij}^{n(i,j)}\} du_1(i) \dots du_K(i)\} \tag{3}$$

with $n(i,j) = \sum_{k=1}^{K} n_k(i,j)$ and $u(i) = \sum_{k=1}^{K} u_k(i)$. Here $\mu_{ij}^{n(i,j)}$ is intepreted as 1 whenever $\mu_{ij} = 0$. In what follows we shall disregard those μ_{ij} that are identically equal to zero by the definition of the model, such as μ_{31} and μ_{32} in our model for work-force participation.

As shown in the examples of chapter 2, all μ_{ij} need not be distinct. We shall assume that there exists a finite system of parameters $\{\lambda_a : a = 1,2,\dots,A\}$, such that each μ_{ij} equals some λ_a, and such that for each $a \in \{1,2,\dots,A\}$, $\mu_{ij} = \lambda_a$ for some (i,j). For simplicity λ_a and $\lambda_{a'}$ will be assumed functionally independent when $a \neq a'$. The specification of the λ_a is not necessarily unique.

We let $\mathcal{A} = \{1,2,\dots,A\}$. For each $a \in \mathcal{A}$ there is a set of pairs (i,j) such that $\mu_{ij} = \lambda_a$. We introduce

$$M_k(a) = \sum_{\{(i,j):\mu_{ij}=\lambda_a\}} N_k(i,j), \text{ and } M(a) = \sum_{k=1}^{K} M_k(a).$$

§ 3.4. By (2) and the assumptions of § 3.3 there exists a finite system of parameters $\{\gamma_b : b=1,2,\dots,B\}$ with $B \leq A$, such that each μ_i equals some γ_b, and such that for each $b \in \mathcal{B} = \{1,2,\dots,B\}$, $\mu_i = \gamma_b$ for some i. Each parameter γ_b will be the sum of a selected number of the parameters λ_a. We may write

$$\gamma_b = \sum_{a=1}^{A} c_{ba}\lambda_a \qquad \text{for each b,} \tag{4}$$

where each c_{ba} equals the number of times (possibly zero) which λ_a is a member of the sum constituting the μ_i that equals γ_b. We shall assume that for any given specification of the λ_a, the c_{ba} are uniquely defined finite numbers. The matrix (c_{ba}) will be designated by c.

For each i, let b(i) be defined by $\mu_i = \gamma_{b(i)}$. Then $c_{b(i),a}$ is the number of times that $\mu_{ij} = \lambda_a$ when j runs through all its possible values while i and a are kept fixed.

For each $b \in \mathcal{B}$ we introduce

$$V_k(b) = \sum_{\{i:\mu_i=\gamma_b\}} U_k(i), \text{ and } V(b) = \sum_{k=1}^{K} V_k(b).$$

§ 3.5. Passing to random variables, we see that the likelihood in (3) may be written in the form

$$\Lambda = \prod_{a=1}^{A} \lambda_a^{M(a)} \cdot \exp\{-\sum_{b=1}^{B} \gamma_b V(b)\}. \tag{5}$$

It will prove convenient to introduce

$$L_k(a) = \sum_{b=1}^{B} c_{ba} V_k(b) = \sum_i c_{b(i),a} U_k(i), \tag{6}$$

and $$L(a) = \sum_{k=1}^{K} L_k(a) = \sum_{b=1}^{B} c_{ba} V(b), \text{ whereby}$$

$$\Lambda = \prod_{a=1}^{A} \lambda_a^{M(a)} \cdot \exp\{-\sum_{a=1}^{A} \lambda_a L(a)\}.$$

The next three paragraphs contain examples for illustration.

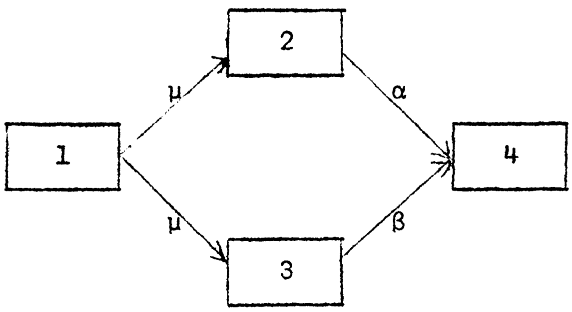
§ 3.6. In the model for work-force participation (§ 2.1) $A = 3$, $\mu_{12} = \nu$, $\mu_{21} = \rho$, $\mu_{13} = \mu_{23} = \mu$, and $\mu_{31} = \mu_{32} = 0$. (State 3 is absorbing.) Thus the set of parameters λ_a consists of ν, ρ, and μ. M(1) is the number of transitions (among the K persons) from state 1 to state 2, M(2) is the number of transitions in the opposite direction, and M(3) is the total number of deaths observed. We have $B = 2$, V(b) is the total living time observed in state b (for $b = 1,2$),

$\gamma_1 = \nu+\mu$, and $\gamma_2 = \rho+\mu$. The matrix $c = (c_{ba})$ may thus be written $c = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$. Finally $L(1) = V(1)$, $L(2) = V(2)$, $L(3) = V(1) + V(2)$.

§ 3.7. In the fertility model of § 2.2, $A = 2$, $\mu_{(k,1),(k+1,1)} = \phi$, and $\mu_{(k,1)(k,2)} = \mu$ for $k = 1,2,\ldots$, $\mu_{ij} = 0$ otherwise. Thus $\lambda_1 = \phi$, $\lambda_2 = \mu$. $M(1)$ is the number of births and $M(2)$ is the number of deaths observed. We have $B = 1$, and $V(1) = L(1) = L(2)$ is the total lifetime observed. $\gamma_1 = \mu$, and $c = (0,1)$. All states $(k,2)$ are absorbing.

§ 3.8. Without offering any "real-life" interpretation, we shall also study the following simple example.

Figure 3.8.



There are four states. The possible direct transitions and the corresponding forces are indicated by arrows and greek letters in figure 3.8. If person no. k starts in state 1 at time 0, his likelihood is

$$\Lambda_k = \mu^{N_k(1,2)+N_k(1,3)} \alpha^{N_k(2,4)} \beta^{N_k(3,4)} \exp\{-2\mu U_k(1)-\alpha U_k(2)-\beta U_k(3)\}. \text{ Thus}$$

$A=3$, $\lambda_1=\mu$, $\lambda_2=\alpha$, $\lambda_3=\beta$; $B=3$, $\gamma_1=2\mu$, $\gamma_2=\alpha$, $\gamma_3=\beta$; $M(1)=N(1,2)+N(1,3)$, $M(2)=N(2,4)$, $M(3)=N(3,4)$; $V(b)=U(b)$ for $b=1,2,3$; and $c = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Finally $L(1)=2U(1)$, $L(2)=U(2)$, and $L(3)=U(3)$.

§ 3.9. If $r_k = r$, the expected total time spent by person no. k in state j equals

$$E\, U_k(j) = \int_0^{z_k} P_{rj}(t)dt,$$

and his expected number of transitions from state i to state j equals

$$E\, N_k(i,j) = \int_0^{z_k} P_{ri}(t)\mu_{ij}\, dt = \mu_{ij} E\, U_k(i).$$

The values of $E\, V_k(b)$ and $E\, M_k(a)$ are obtained by summation over the corresponding values of $E\, U_k(i)$ and $E\, N_k(i,j)$, respectively. We get

$$E\, M_k(a) = \lambda_a \cdot \sum_{\{(i,j):\mu_{ij}=\lambda_a\}} E\, U_k(i) = \lambda_a \sum_i c_{b(i),a} E\, U_k(i).$$

By (6), therefore,

$$E\, M_k(a) = \lambda_a\, E\, L_k(a) \tag{8}$$

for all k and a.

§ 3.10. In our model for work-force participation we get

$$\int_0^z P_{11}(t)dt = \{\rho(1-e^{-\mu z})/\mu+\nu(1-e^{-(\nu+\rho+\mu)z})/(\nu+\rho+\mu)\}/(\nu+\rho),$$

$$\int_0^z P_{12}(t)dt = \nu\{(1-e^{-\mu z})/\mu-(1-e^{-(\nu+\rho+\mu)z})/(\nu+\rho+\mu)\}/(\nu+\rho), \text{ and}$$

$$\int_0^z P_{13}(t)dt = z - (1-e^{-\mu z})/\mu.$$

Formulae for $\int_0^z P_{2j}(t)dt$ are obtained by exchanging ν and ρ above. (8) also gives the relation $EM(1) = \nu \cdot EL(1)$, $EM(2) = \rho \cdot EL(2)$, and $EM(3) = \mu \cdot EL(3)$.

We shall not care to consider our fertility model at this point.

4. A further study of the derived model

§ 4.1. We shall now concentrate on the model defined by (5). This is a Darmois-Koopman class of distributions. To avoid unnecessary peculiarities, we state as an assumption that all M(a) and V(b) be essentially linearly independent. (This can always be achieved, by reparametrisation if necessary.) We also assume that the matrix $c = (c_{ba})$ has rank B. (We note that $B \leq A$, so the rank of c cannot exceed B.) We introduce

$$\beta_\alpha = \begin{cases} \ln \lambda_\alpha & \text{for } \alpha = 1,2,\ldots,A, \\ -\gamma_{\alpha-A} & \text{for } \alpha = A+1,\ldots,A+B, \end{cases} \tag{9}$$

where ln stands for the natural logarithm, and let $\beta = (\beta_1,\ldots,\beta_{A+B})$.

Theorem: The β_α are linearly independent.

Proof: We wish to prove that if there exists a set of coefficients $x_1, x_2,\ldots,x_{A+B}$ such that $\sum_{\alpha=1}^{A+B} x_\alpha\beta_\alpha = 0$ identically in β, then $x_1=x_2=\ldots=x_{A+B}=0$. Equivalently this identity can also be written as an identity in the λ_a in the form

$$\sum_{a=1}^{A} x_a \ln \lambda_a \equiv \sum_{a=1}^{A} \lambda_a \sum_{b=1}^{B} c_{ba}x_{A+b}.$$

Differentiation with respect to λ_a gives

$$x_a/\lambda_a \equiv \sum_{b=1}^{B} c_{ba} x_{A+b}$$

as an identity in λ_a. Since the right hand side here is independent of λ_a, we must have $x_a = 0$, and

$$\sum_{b=1}^{B} c_{ba} x_{A+b} = 0$$

for $a = 1,2,\ldots,A$. In matrix form this is

$$(x_{A+1},\ldots,x_{A+b})c = 0. \tag{10}$$

Since c has rank B, there exist B distinct elements $i_1,\ldots,i_B$ in $\mathcal{B}$ such that the matrix

$$c^* = \begin{pmatrix} c_{1i_1},\ldots,c_{1i_B} \\ \cdots\cdots\cdots \\ c_{Bi_1},\ldots,c_{Bi_B} \end{pmatrix}$$

has rank B. (10) implies $(x_{A+1},\ldots,x_{A+B})c^* = 0$, whereby $x_{A+b} = 0$ for all $b \in \mathcal{B}$. □

By general properties of Darmois-Koopman classes of distributions, $\{M(1),\ldots,M(A), V(1),\ldots,V(B)\}$ is then a minimal sufficient statistic.

§ 4.2. It is typical for demographic models that they must account for mortality. In our general model, this feature would be formalized by the introduction of the requirement that for each non-absorbing state i there exists an absorbing state i' such that $\mu_{ii'}>0$. In such a case, $N_k(i,i')$ equals 0 or 1 for each k.

Obviously i' need not be uniquely defined. We may for instance distinguish between various causes of mortality, between mortality and emigration as causes of decrement, etc., in which case i' may be chosen at will from a number of states.

The model of § 3.8 does not fulfil this requirement, since from state 1 one may pass direct only to states 2 and 3, none of which are absorbing. On the other hand any sample path may pass from state 1 to state 2 (or state 3) at most once during any period of observation. It will turn out (in the proof of the theorem of § 4.4) that we shall frequently prefer such a feature in our models to the more restrictive requirement that it always be possible to pass direct into an absorbing state.

Passing to the λ_a, we shall therefore often make the following

Assumption: For each $b \in \mathcal{B}$ there exists a known, finite integer $\kappa(b)$ and an $a \in \mathcal{a}$, say $a(b)$, such that $c_{b,a(b)} \geq 1$ and such that $\lambda_{a(b)}$ only represents one or more forces of transitions that can be made at most $\kappa(b)$ times altogether in any sample path. In other words $0 \leq M_k(a(b)) \leq \kappa(b)$ for any k.

In the examples of §§ 3.6 and 3.7, μ has this property. In § 3.8, both μ, α, and β have this property. In § 3.6, $a(1)=a(2)=3$; in § 3.7, $a(1) = 2$; and in § 3.8, $a(b) = b$ for $b = 1,2,3$. In each case, all $\kappa(b) = 1$.

Under the above assumption,

$$0 \leq M(a(b)) \leq K\kappa(b) \text{ for all } b. \tag{11}$$

§ 4.3. In the proof of the theorem of § 4.4 it will also turn out that we shall sometimes want to make the following

Assumption: For each b and $b' \in \mathcal{B}$, $c_{b',a(b)} \leq c_{b,a(b)}$.

Thus we assume that it is possible to choose each $a(b)$ in such a way that $c_{b,a(b)}$ is the maximal member of the column $(c_{1,a(b)}, c_{2,a(b)},\ldots,c_{B,a(b)})'$. This is obviously the case in our three examples.

Many of the results below do not use the assumptions of §§ 4.2 and 4.3. We shall therefore explicitly state when a proof is based on these assumptions.

§ 4.4. As can be seen e.g. in (3), the likelihood in (5) is a density with respect to some measure σ over the sample space Ω. We will take it to be a proper probability density so that

$$\int_\Omega \prod_{a=1}^{A} \lambda_a^{m(a)} \exp\{-\sum_{b=1}^{B} \gamma_b v(b)\} d\sigma = 1 \tag{12}$$

for all $\lambda_a > 0$, $a \in \mathcal{a}$. Introducing

$$W(\alpha) = \begin{cases} M(\alpha) & \text{for } \alpha = 1,2,\ldots,A, \\ V(\alpha-A) & \text{for } \alpha = A+1,\ldots,A+B, \end{cases}$$

we see that (12) may also be written in the form

$$\int_\Omega \exp\{\sum_{\alpha=1}^{A+B} \beta_\alpha w(\alpha)\} d\sigma = 1 \tag{13}$$

for all $\beta \in \mathfrak{B}$, where $\mathfrak{B}$ is the parameter space of the present model. Thus $\mathfrak{B}$ is the part defined by (9) and (4) of the Euclidean space R_{A+B}. Let us now disregard (9) and (4) for a moment. The integral in (13) may converge (not necessarily to 1) also for $\beta \notin \mathfrak{B}$. We define $\mathfrak{B}^*$ as the set of those $\beta \in R_{A+B}$ for which the integral converges. Obviously $\mathfrak{B} \subseteq \mathfrak{B}^*$.

Theorem: Under the assumptions stated in §§ 4.2. and 4.3, all points in $\mathfrak{B}$ are interior points in $\mathfrak{B}^*$.

Proof: As proved by Lehmann (1959, p. 51), $\mathfrak{B}^*$ is a convex region. We choose a fixed set of positive values for the λ_a, and let $\beta'_{A+b} = -\sum_{a=1}^{A} c_{ba}\lambda_a$ for all $b \in \mathfrak{B}$.

(i) Let $\beta = (\beta_1,\ldots,\beta_{A+B})$, with $\beta_a = \ln \lambda_a$ for all $a \in \mathfrak{A}$, $\beta_{A+b_0} < \beta'_{A+b_0}$ for a $b_0 \in \mathfrak{B}$, $\beta_{A+b} = \beta'_{A+b}$ otherwise. Then $\exp\{\sum_{\alpha=1}^{A+B} \beta_\alpha w(\alpha)\} \leqq \exp\{\sum_{\alpha \neq A+b_0} \beta_\alpha w(\alpha) + \beta'_{A+b_0} w(A+b_0)\}$.

As the right hand side here is integrable with respect to σ, so is the left hand side, and $\beta \in \mathfrak{B}^*$.

(ii) Then let $\beta''_{A+b_0} = \beta'_{A+b_0} + c_{b_0,a(b_0)}\lambda_{a(b_0)} = -\sum_{a \neq a(b_0)} c_{b_0 a}\lambda_a$, and let β be as under (i) with the single exception that $\beta''_{A+b_0} > \beta_{A+b_0} > \beta'_{A+b_0}$. Let

$\lambda'_{a(b_0)} = (\beta''_{A+b_0} - \beta_{A+b_0})/c_{b_0,a(b_0)}$, $\lambda'_a = \lambda_a$ otherwise, and let $\psi = \lambda_{a(b_0)}/\lambda'_{a(b_0)}$.

Then $\psi > 1$, and $\exp\{\sum_{\alpha=1}^{A} \beta_\alpha w(\alpha)\} = \prod_{a=1}^{A} \lambda_a^{m(a)} = \psi^{m(a(b_0))} \prod_{a=1}^{A} (\lambda'_a)^{m(a)} \leqq$ $\psi^{K\kappa(b_0)} \prod_{a=1}^{A} (\lambda'_a)^{m(a)}$, by (11). Furthermore, $\sum_{a=1}^{A} c_{ba}\lambda'_a = \frac{c_{b,a(b_0)}}{c_{b_0,a(b_0)}}(\beta''_{A+b_0} - \beta_{A+b_0}) - \beta''_{A+b_0}$ $\leqq -\beta_{A+b}$ by the assumption of § 4.3. Thus $\exp\{\sum_{\alpha=1}^{A+B} \beta_\alpha w(\alpha)\} \leqq$ $\psi^{K\kappa(b_0)} \prod_{a=1}^{A} (\lambda'_a)^{m(a)} \exp\{-\sum_{b=1}^{B} w(A+b) \sum_{a=1}^{A} c_{ba}\lambda'_a\}$ since $\beta_{A+b} = -\sum_{a=1}^{A} c_{ba}\lambda_a = -\sum_{a=1}^{A} c_{ba}\lambda'_a$ when $b \neq b_0$. By (i) above we therefore get $\beta \in \mathfrak{B}^*$. From this the theorem follows. □

§ 4.5. The following result is a simple consequence of theorem 2.7.9 by Lehmann (1959, p. 52):

Theorem: Let h be any bounded real integrable function over the sample space Ω, and let

$$h^*(\bar{\beta}) = E_{\beta} h = \int_{\Omega} h(\omega) \exp\{\sum_{\alpha=1}^{A+B} \bar{\beta}_\alpha w(\alpha,\omega)\} d\sigma(\omega)$$

be considered as a function of the complex variables $\bar{\beta}_\alpha = \beta_\alpha + i\beta^*_\alpha$ (for $\alpha = 1,2,\ldots,A+B$) with $\beta \in \mathfrak{B}^*$. Then $h^*(\bar{\beta})$ is an analytic function in each argument for which β is an interior point of $\mathfrak{B}^*$. The derivatives of all orders with respect to the $\bar{\beta}_\alpha$ may be computed under the integral sign.

By introducing various choices of h, differentiation of h^* may give a series of interesting formulae. We shall restrict ourselves to the case of $h \equiv 1$. By (7), formula (12) may also be written in the form

$$\int_\Omega \prod_{a=1}^{A} \lambda_a^{m(a)} \exp\{-\sum_{a=1}^{A} \lambda_a \ell(a)\} d\sigma \equiv 1.$$

Under the assumptions of §§ 4.2 and 4.3, differentiation with respect to the λ_a then gives

$$E\, M(a) = \lambda_a\, E\, L(a), \tag{14}$$

$$E\, \{M(a) - \lambda_a L(a)\}^2 = E\, M(a), \text{ and} \tag{15}$$

$$E\, \{M(a) - \lambda_a L(a)\}\, \{M(a') - \lambda_{a'} L(a')\} = 0 \quad \text{for } a \neq a'. \tag{16}$$

Formula (14) also follows from (8), which holds under more general assumptions.

5. Estimation of the λ_a

§ 5.1. If we take all M(a) and L(a) as given, $\mathfrak{A}$ may be partitioned into the set $\mathfrak{A}_0$ of those $a \in \mathfrak{A}$ for which $L(a) = 0$, and the set $\mathfrak{A}_+$ of those a for which $L(a) > 0$. Since $M(a) = 0$ if $L(a) = 0$, the Λ of (7) may be written in the form

$$\Lambda = \prod_{a \in \mathfrak{A}_+} \lambda_a^{M(a)} \exp\{-\sum_{a \in \mathfrak{A}_+} \lambda_a L(a)\}.$$

Thus Λ is maximized by

$$\hat{\lambda}_a = M(a)/L(a) \quad \text{for all } a \in \mathfrak{A}_+,$$

while for $a \in \mathfrak{A}_0$ there is no particular value for λ_a which maximizes Λ. We shall arbitrarily choose $\hat{\lambda}_a = 0$ if $a \in \mathfrak{A}_0$.

If for an $a \in \mathfrak{A}$ there exists a k such that $\lambda_a = \mu_{r_k j}$ for some j, then $P\{L(a) > 0\} = 1$, and the possibility that $a \in \mathfrak{A}_0$ poses no real problem. In this case $\hat{\lambda}_a$ is a maximum likelihood estimator. Otherwise we may find that $P\{L(a) > 0\} < 1$, in which case a maximum likelihood estimator for λ_a is not defined.

The following paragraphs study asymptotic properties of the $\hat{\lambda}_a$ as $K \to \infty$ under the blanket assumption that no z_k exceeds some finite positive number z_0.

§ 5.2. Let $\bar{M}(a) = M(a)/K$ and $\bar{L}(a) = L(a)/K$. For each a, $L_1(a),\ldots,L_K(a)$ will not generally be identically distributed unless all z_k have the same value and all r_k are equal. We shall not make such an assumption. Nevertheless the following theorem holds.

Theorem: Let $a \in \mathcal{A}$. Assume that

(i) a finite positive limit $L_a = \lim_{K\to\infty} E\bar{L}(a)$ exists, and

(ii) $P\{L(a) > 0\} \to 1$ as $K \to \infty$.

Then $\hat{\lambda}_a$ is consistent as $K \to \infty$.

Proof: By (6), $L_k(a) \leqq z_k \sum_{b=1}^{B} c_{ba} \leqq z_0 \sum_{b=1}^{B} c_{ba}$, so $EL_k(a)$ and var $L_k(a)$ exist and the latter is majored by some finite value σ^2. By Tchebycheff's inequality,

$$P\{|\bar{L}(a) - E\bar{L}(a)| < \varepsilon\} \geqq 1 - \text{var } \bar{L}(a)/\varepsilon^2 \geqq 1 - \sigma^2/(K\varepsilon^2)$$

for any $\varepsilon > 0$. For $K \geqq K_0(\varepsilon)$, $|E\bar{L}(a) - L_a| < \frac{\varepsilon}{2}$, and then

$$P\{|\bar{L}(a) - L_a| < \varepsilon\} \geqq P\{|\bar{L}(a) - E\bar{L}(a)| < \tfrac{\varepsilon}{2}\} \geqq 1 - 4\sigma^2/(K\varepsilon^2).$$

Letting $K \to \infty$ we see that plim $\bar{L}(a) = L_a$.

It is possible to show that also var $M_k(a)$ will exist for any k. We invoke (8) to see that $E\bar{M}(a) = \lambda_a E\bar{L}(a)$. Thus $M_a = \lim_{K\to\infty} E\bar{M}(a) = \lambda_a L_a$ exists as a finite limit. By an argument quite similar to the one above, plim $\bar{M}(a) = \lambda_a L_a$.

Thus, for arbitrary positive η and δ, $P\{F\} > 1 - \delta/3$ and $P\{G\} > 1 - \delta/3$ for $K > K_1(\eta,\delta)$, where F is the event that $|\bar{L}(a) - L_a| < \eta$ and G is the event that $|\bar{M}(a) - \lambda_a L_a| < \eta$.

Let H be the event that $L(a) > 0$. By assumption, $P(H) > 1 - \delta/3$ for $K > K_2(\delta)$.

Since the real function $\phi(x,y) = x/y$ is continuous in (x,y) for all $y \neq 0$, the event $F \cap G \cap H$ implies the event $H \cap I$ with $I = \{|\bar{M}(a)/\bar{L}(a) - \lambda_a| < \varepsilon\}$, if $\eta \leqq \eta_0(\varepsilon)$ for any $\varepsilon > 0$. With such a choice of η, therefore,
$P\{|\hat{\lambda}_a - \lambda_a| < \varepsilon\} \geqq P\{L(a) > 0 \text{ and } |\bar{M}(a)/\bar{L}(a) - \lambda_a| < \varepsilon\} = P\{H \cap I\} \geqq P\{F \cap G \cap H\} =$
$1 - P\{\bar{F} \cup \bar{G} \cup \bar{H}\} \geqq 1 - \{P(\bar{F}) + P(\bar{G}) + P(\bar{H})\} = P(F) + P(G) + P(H) - 2 > 1 - \delta$,
provided $K > K_3(\varepsilon,\delta) = \max\{K_1(\eta_0(\varepsilon),\delta), K_2(\delta)\}$. □

§ 5.3. In all previous paragraphs the exposure times $z_1,\ldots,z_K$ were regarded as fixed numbers. In certain situations it is possible to interpret the z_k as values of random variables $Z_1,Z_2,\ldots,Z_K$ which are stochastically independent and identically distributed with some distribution function G, where $G(0)=0$ and $G(z_0)=1$ for some finite $z_0>0$. We will then say that the z_k <u>are G-random</u> for brief. A special case is the situation where it is known beforehand that all z_k will equal some preassigned positive number z, as we may then let $P\{Z_k=z\}=1$ for all k.

§ 5.4. When the z_k are G-random while the r_k are preassigned, the likelihood will have the form $\Lambda^* = \Lambda dG(Z_1)dG(Z_2)\ldots dG(Z_K)$, where Λ is given by (7). If G is completely specified, the "maximum likelihood" estimators of the λ_a are still given as in § 5.1. If G is only partly specified, maximum likelihood estimators for the λ_a may or may not exist. As long as G is independent of the λ_a, however, "partial maximation" of Λ^* with respect to the λ_a may be carried out as in § 5.1.

$\{M(1),\ldots,M(A),\ V(1),\ldots,V(B),\ Z_1,\ldots,Z_K\}$ will be sufficient, but not necessarily minimal sufficient, no matter what is known about G.

§ 5.5. Even when the z_k are preassigned, it may be possible to interpret the initial states $r_1,\ldots,r_K$ as values of random (not necessarily real-valued) variables $R_1,\ldots,R_K$ which are stochastically independent and identically distributed with some distribution $\pi_r = P\{R_k = r\}$. We will then say that the r_k are π-random. Again $r_k \equiv r$ is a special case.

Finally the pairs (Z_k,R_k) may be independent and identically distributed random variables with a distribution given by $H(z,r) = P\{Z_k \leq z \text{ and } R_k = r\}$, in which case we shall say that the (z_k,r_k) are H-random.

Considerations quite similar to those of § 5.4 pertain if the r_k are π-random and if the (z_k,r_k) are H-random.

We note that formula (8), the results of §§ 4.4 and 4.5, and the theorem of § 4.1 hold as stated even when the z_k are G-random, when the r_k are π-random, and when the (z_k,r_k) are H-random. So does the theorem of § 5.2, but in the case of H-randomness its assumptions (i) and (ii) hold almost automatically, and under G-randomness they follow from some other assumptions, as we shall see below.

The two first formulae of § 3.9 give conditional expected values.

In the two following paragraphs some effects of G- and H-randomness are investigated. We shall not give π-randomness any further consideration.

§ 5.6. (<u>G-randomness.</u>) In the case of G-random z_k and preassigned r_k, we shall assume that there is a finite number of possible initial states $s_1,\ldots,s_C$ so that each r_k must equal some s_c. We let $\mathcal{K}_c = \{k: r_k = s_c\}$ and let $S_c(K)$ be the number of elements in $\mathcal{K}_c$ when there are K sample paths in all. We note that for each $a \in \mathfrak{a}$ and each c, all $L_k(a)$ with $k \in \mathcal{K}_c$ will be identically distributed. Similarly for the $M_k(a)$.

We shall assume that $\alpha_c = \lim_{K\to\infty} S_c(K)/K$ exists for all c.

<u>Theorem 1</u>: Under the assumptions stated above, $\hat{\lambda}_a$ is consistent as $K \to \infty$ provided there exists a c for which $P\{L_k(a) > 0\} > 0$ if $k \in \mathcal{K}_c$ and where $S_c(K) \to \infty$ as $K \to \infty$.

<u>Proof</u>: Letting

$$\varepsilon_c(a) = EL_k(a) \quad \text{for } k \in \mathcal{K}_c \tag{17}$$

we get

$$E\bar{L}(a) = \sum_{c=1}^{C} \varepsilon_c(a)\, S_c(K)/K \to \sum_{c=1}^{C} \alpha_c \varepsilon_c(a) = L_a, \tag{18}$$

so (i) of the theorem in § 5.2 holds. Let $p_c(a) = P\{L_k(a) = 0\}$ for $k \in \mathcal{K}_c$. Then $P\{L(a) = 0\} = \prod_{c=1}^{C} \{p_c(a)\}^{S_c(K)}$ (with 0^0 interpreted as 1). Let c be given as in the theorem. Then $p_c(a) < 1$. If $p_c(a) = 0$, then $P\{L(a) = 0\} = 0$ for large enough K. If $p_c(a) > 0$, $\{p_c(a)\}^{S_c(K)} \to 0$ as $K \to \infty$. In any case $P\{L(a) > 0\} \to 1$ as $K \to \infty$. Theorem 1 then follows from the theorem of § 5.2. □

We now turn to the asymptotic distribution of the $\hat{\lambda}_a$ as $K \to \infty$. Let $\mathcal{Q}'$ consist of those $a \in \mathcal{Q}$ for which $P\{L(a) > 0\} \to 1$ as $K \to \infty$. By a suitable enumeration we may assume that $\mathcal{Q}' = \{1,2,\ldots,A'\}$, with $A' \leq A$. We shall not consider the $\hat{\lambda}_a$ for which $a \in \mathcal{Q} - \mathcal{Q}'$.

<u>Theorem 2</u>: Under the assumptions stated above theorem 1, the vector $\sqrt{K}\,(\hat{\lambda}_1 - \lambda_1, \ldots, \hat{\lambda}_{A'} - \lambda_{A'})$ is asymptotically multinormal with mean **0** and some covariance matrix Σ.

<u>Proof</u>: Let $L(a,c) = \sum_{k \in \mathcal{K}_c} L_k(a)$, let $\bar{L}(a,c) = 0$ if $S_c(K) = 0$, $\bar{L}(a,c) = L(a,c)/S_c(K)$ otherwise, and let $M(a,c)$ and $\bar{M}(a,c)$ be defined similarly. Then

$$\sqrt{K}\{\bar{M}(a) - \lambda_a \bar{L}(a)\} = \sum_c S_c(K)\{\bar{M}(a,c) - \lambda_a \bar{L}(a,c)\}K^{-\frac{1}{2}}$$

$$= \sum_c \{S_c(K)/K\}^{\frac{1}{2}} \cdot \{S_c(K)\}^{\frac{1}{2}} \{\bar{M}(a,c) - \lambda_a \bar{L}(a,c)\}.$$

Since $S_c(K) \to \infty$ as $K \to \infty$ for the c where $\alpha_c > 0$, the central limit theorem, combined with (8) and a limit theorem due to Cramér (1946, Chapter 20.6), gives the result that the simultaneous distribution function of the vector

$$\sqrt{K}\,\{\bar{M}(1) - \lambda_1 \bar{L}(1), \ldots, \bar{M}(A') - \lambda_{A'} \bar{L}(A')\}$$

converges at all points to the distribution function of a multinormal distribution with mean 0 and some covariance matrix Σ^*.

By (18) and the first part of the proof of the theorem in § 5.2, plim $\bar{L}(a)=L_a$ for all $a \in \mathcal{A}'$, with L_a given in (18). Since

$$\sqrt{K}(\hat{\lambda}_a - \lambda_a) = \sqrt{K}\ \{\bar{M}(a) - \lambda_a\bar{L}(a)\}/\bar{L}(a) \quad \text{for } L(a) > 0,$$

a new application of the Cramérian limit theorem gives the theorem. □

When the assumptions of §§ 4.2 and 4.3 hold, we may use (15) and (16) to assert that the variables $\sqrt{K}(\hat{\lambda}_a - \lambda_a)$ for $a \in \mathcal{A}$ are asymptotically independent and normally distributed with means 0 and asymptotic variances,

$$\text{as. var } \sqrt{K}(\hat{\lambda}_a - \lambda_a) = \lambda_a/L_a \tag{19}$$

for $a \in \mathcal{A}'$, with L_a given by (17) and (18).

All these results are independent of any specific assumptions on the distribution G.

§ 5.7. (H-randomness.) In the case of H-random (z_k, r_k), the variables $L_1(a),\dots,L_K(a)$ are independent and identically distributed for each $a \in \mathcal{A}$. In this case, therefore, $L_a = E\bar{L}(a) = EL_k(a)$, and plim $\bar{L}(a) = L_a$ by the law of large numbers. Similarly for the $M_k(a)$.

To avoid unnecessary peculiarities, we shall assume that $P\{L_k(a) > 0\} > 0$ for all $a \in \mathcal{A}$.

Theorem 1: Under the assumptions stated above, all $\hat{\lambda}_a$ are consistent.

Proof: By the teorem of §5.2 it suffices to show that $P\{L(a) > 0\} \to 1$ as $K \to \infty$. Let $p(a) = P\{L_k(a) = 0\}$. Since $p(a) < 1$ by assumption, $P\{L(a) = 0\} = \{p(a)\}^K \to 0$ as $K \to \infty$. □

Theorem 2: Under the assumptions of the present paragraph, the vector $\sqrt{K}(\hat{\lambda}_1 - \lambda_1,\dots,\hat{\lambda}_A - \lambda_A)$ is asymptotically multinormal with mean 0 and some covariance matrix Σ.

The proof is a simplified version of the proof of theorem 2 in § 5.6. The remarks after that proof still hold with $\mathcal{A}'$ replaced by $\mathcal{A}$ and with $L_a = EL_k(a)$.

When in addition to this H is completely specified, results due to Sverdrup (1965, Appendix B) may be used to show that the rates $\hat{\lambda}_a$ have asymptotic variances at least as small as any other Fisher-consistent estimators.

§ 5.8. It is possible to construct confidence intervals for the λ_a and to test hypotheses about them e.g. by simple extensions of methods described by Sverdrup (1965). We will not take up these questions here.

§ 5.9. All previous results hold with only obvious modifications if the values of some of the parameters λ_a happen to be known. If for some a, λ_a has a given positive value and $P\{L(a) > 0\} > 0$, formula (8) shows that our family of probability distributions is not complete.

6. Transformations of the parameter space

§ 6.1. It may sometimes be convenient to consider transformations of the parameter space. (Examples of this are abundant. See e.g. Hoem (1968).) Although we have applications to demographic models specifically in mind, we shall give a more general formulation of our reasoning and results, as there is no need to unduly restrict their scope.

Let $\Lambda(\theta)$ be a random (scalar) variable, the value of which depends on the parameter vector $\theta = (\theta_1,\ldots,\theta_s) \in \Theta$. Assume that

$$E\,\frac{\partial\Lambda(\theta)}{\partial\theta_i} = 0 \qquad \text{for } i = 1,2,\ldots,s, \tag{20}$$

and let $\Sigma = (\sigma_{ij})$, with $\sigma_{ij} = -E\,\frac{\partial^2\Lambda(\theta)}{\partial\theta_i\partial\theta_j}$, where all σ_{ij} are assumed to exist as finite functions of θ. We introduce a new parameter vector $\alpha = (\alpha_1,\ldots,\alpha_s) \in \mathcal{A}$ by a one-to-one transformation ψ from $\mathcal{A}$ to Θ, so that $\theta_i = \psi_i(\alpha_1,\ldots,\alpha_s)$ for $i = 1,2,\ldots,s$. The differential quotients $\frac{\partial}{\partial\alpha_j}\psi_i(\alpha_1,\ldots,\alpha_s)$ are assumed to exist for all i and j and for all $\alpha \in \mathcal{A}$.

Let
$$J = \begin{Bmatrix} \frac{\partial\psi_1(\alpha)}{\partial\alpha_1}, & \ldots\ldots, & \frac{\partial\psi_1(\alpha)}{\partial\alpha_s} \\ \cdots & \cdots & \cdots \\ \frac{\partial\psi_s(\alpha)}{\partial\alpha_1}, & \ldots\ldots, & \frac{\partial\psi_s(\alpha)}{\partial\alpha_s} \end{Bmatrix}$$

be the matrix of differential quotients. We define $\Lambda_1(\alpha) = \Lambda[\psi(\alpha)]$,

$$\gamma_{ij} = -E\,\frac{\partial^2\Lambda_1(\alpha)}{\partial\alpha_i\partial\alpha_j}, \text{ and } \Gamma = (\gamma_{ij}).$$

<u>Theorem</u>: Under the above assumptions, $\Gamma = J'\Sigma J$.

<u>Proof</u>: The formula of the lemma is equivalent to

$$\gamma_{ij} = \sum_{\nu=1}^{s}\sum_{k=1}^{s}\frac{\partial\psi_\nu}{\partial\alpha_i}\,\sigma_{\nu k}\,\frac{\partial\psi_k}{\partial\alpha_j} \qquad \text{for } i \text{ and } j = 1,2,\ldots,s. \tag{21}$$

This is what we proceed to prove. By the definition of $\Lambda_1(\alpha)$, we get

$$\frac{\partial}{\partial\alpha_i}\Lambda_1(\alpha) = \sum_{\nu=1}^{s}\frac{\partial}{\partial\theta_\nu}\Lambda[\psi(\alpha)]\,\frac{\partial\psi_\nu}{\partial\alpha_i}, \text{ and}$$

$$\frac{\partial^2}{\partial\alpha_i\partial\alpha_j}\Lambda_1(\alpha) = \sum_{\nu=1}^{s}\sum_{k=1}^{s}\frac{\partial^2}{\partial\theta_\nu\partial\theta_k}\Lambda[\psi(\alpha)]\,\frac{\partial\psi_\nu}{\partial\alpha_i}\frac{\partial\psi_k}{\partial\alpha_j} + \sum_{\nu=1}^{s}\frac{\partial}{\partial\theta_\nu}\Lambda[\psi(\alpha)]\,\frac{\partial^2\psi_\nu}{\partial\alpha_i\partial\alpha_j}.$$

Using (20) and the definition of $\sigma_{\nu k}$, we get (21). □

The obvious application of this theorem is to transformations of parameters $\theta_1,\ldots,\theta_r$ for which maximum likelihood estimators $\hat{\theta}_1,\ldots,\hat{\theta}_r$ exist. Under very general conditions, $\hat{\theta}$ will be asymptotically multinormal with mean θ and a covariance matrix $\frac{1}{K}\Sigma^{-1}$, where Σ is as defined above with $\Lambda(\theta)$ as the logarithm of the likelihood function. Applications to our $\hat{\lambda}_a$ are immediate. Relation (20) becomes $E\{M(a) - \lambda_a L(a)\} = 0$ for all $a\in\mathfrak{A}$, which follows from (8).

§ 6.2. It may often occur that one of the parameters, say θ_1, is actually not changed at all by the transformation ψ, and thus $\theta_1 = \alpha_1$. It would be rather unfortunate if in such a case the asymptotic variance of the estimator $\hat{\theta}_1$ were different from the one of $\hat{\alpha}_1$. In fact it is easily shown that they are equal. We let σ^{11} denote element no. (1.1) in Σ^{-1}, and similarly for γ^{11}.

Theorem: If in § 6.1 $\psi_1(\alpha_1,\ldots,\alpha_s) = \alpha_1$, then $\gamma^{11} = \sigma^{11}$.

Proof: In this case J has the form

$$J = \begin{bmatrix} 1 & 0 \\ J_{21} & J_{22} \end{bmatrix}$$

where the zero stands for a $1\times(s-1)$ matrix of zeroes, J_{21} stands for some $(s-1)\times 1$ matrix, and J_{22} stands for some $(s-1)\times(s-1)$ matrix. Partitioning Σ and Γ in the same way, some simple arithmetic gives $\Gamma_{22} = J'_{22}\ \Sigma_{22}\ J_{22}$. Thus $\gamma^{11} = |\Gamma_{22}|/|\Gamma| = |J_{22}|^2.|\Sigma_{22}|/|J|^2.|\Sigma| = |\Sigma_{22}|/|\Sigma| = \sigma^{11}$, since $|J| = |J_{22}|$. □

The results of the present chapter have been used in Hoem (1968), but it has seemed preferable to present them in the more general setting of the present paper.

7. Acknowledgement

I am grateful to Mr. Thorleif Borge, who has proof-read much of this paper in manuscript.

8. References

[1] Albert, Arthur (1962): "Estimating the Infinitesimal Generator of a Continuous Time, Finite State Markov Process." Ann.Math.Stat., **33**, 727-753.

[2] Billingsley, Patrick (1961): "Statistical Inference for Markov Processes." Univ. of Chicago Press.

[3] Chung, Kai Lai (1960): "Markov Chains with Stationary Transition Probabilities." Springer-Verlag.

[4] Cramér, Harald (1946): "Mathematical Methods of Statistics." Princeton University Press.

[5] Du Pasquier, L.G. (1913): "Mathematische Theorie der Invaliditätsversicherung." Mitteil.Verein.schweitzer.Versicherungsmath., 8, 1-153.

[6] Hoem, Jan M. (1968): "Fertility Rates and Reproduction Rates in a Probabilistic Setting." Under publication.

[7] Lehmann, E.L. (1959): "Testing Statistical Hypotheses." John Wiley & Sons, Inc.

[8] Meier, Paul (1955): "Note on Estimation in a Markov Process with Constant Transition Rates." Human Biology, 27, 121-124.

[9] Sverdrup, Erling (1961): "Statistiske metoder ved dødelighetsundersøkelser." Mimeographed. Inst. of Math., University of Oslo.

[10] Sverdrup, Erling (1965): "Estimates and Test Procedures in Connection with Stochastic Models for Deaths, Recoveries and Transfers between different States of Health." Skand.Aktuarietidskr. 1965, 184-211.

[11] Zahl, Samuel (1955): "A Markov Process Model for Follow-up Studies." Human Biology, 27, 90-120.