Arbeidsnotater

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

IO 68/12 Oslo, 17. juni 1968

A Probabilistic Model for Primary Marital Fertility

by Jan M. Hoem.

Page

1. Introduction

§ 1.1. The Population Register of the Central Bureau of Statistics of Norway keeps, in principle, an individual record of each person resident in the country, irrespective of nationality. The Register became operative as of October 1, 1964. The record contains the status on that date and all subsequent changes for each person then in the population. A complete record of his lifespan within the population is kept for any later arrival. Each record, continuously updated, contains, *inter alia*, an identification number (assigned as described by Selmer (1967), containing code for sex and date of birth), marital status with dates and kinds of possible changes, possible dates of immigration, emigration, re-immigration, re-emigration, or death, date(s) of migration with old and new address(es) (for persons with change(s) in postal address, also changes within a municipality), identification number of spouse (for persons married since October 1, 1964), and identification number of parents (also for persons born since October 1, 1964). Extensions of the activities are contemplated.

Further information on the Register has been given by Skaug (1967, 1968).

§ 1.2. The data compiled in a register like this are eminently suited for demographic analysis. In an attempt at turning such possibilities to good account the Norwegian CBS has established a Study Group for Population Models, to which the present author has the privilege of being advisor in mathematical statistics. The Study Group became operative in January/February, 1968, and a number of projects are in progress.

Projects taken up by the Study Group have a dual nature. On the one hand there is an independent interest in the analysis of the Population Register data, leading to some basic demographic research. And on the other hand each project serves the purpose of contributing to the improvement of the CBS population projection model being developed simultaneously by the Group.

§ 1.3. One project in hand is a probabilistic model for nuptiality and fertility. The present paper is devoted to one particular section of this model, viz. the study of first births in the first marriage of nulliparous women. There is a twofold reason for restricting ourselves in this way. First it will ease our exposition as it permits concrete examplification. Secondly there will be room for a fairly extensive description of the estimation techniques which we plan to employ.

We will suggest some of the various directions in which our model may be extended (chapter 3). Statistical techniques quite similar to those presented here may be applied in all cases mentioned. The models suggested will reflect the detail into which our data will permit us to go.

2. The model for primary marital fertility

§ 2.1. By a birth we shall mean a confinement resulting in the delivery of one or more children. Whether stillbirths will be counted along with the live births is here a matter of convenience or convention. In application to Norwegian population data we do not plan to count stillbirths because of incomplete registration.

§ 2.2. Consider a nulliparous woman in her first marriage whose age at marriage was y and who has been married for $u \geq 0$ years. We shall call such a female a (y,u)-woman.

Age y of marriage will conventionally be above the lower reproductive age. We shall similarly assume that the age $y+u$ from which the woman is taken under consideration, is below age at menopause, which we shall designate by ω_0.

Sooner or later the woman will experience one of the following five events.

(1) She has her first birth.

(2) She dies.

(3) Her husband dies.

(4) The marriage is dissolved by separation or divorce.

(5) She is lost from observation (e.g. through emigration.)

If more than one of these events occur simultaneously, we shall register only the one of lowest rank in the list.

§ 2.3. A population of nulliparous spinsters who marry and then experience events among those mentioned above, may be said to move through (part of) a system of six states as indicated in figure 2.3. We shall represent

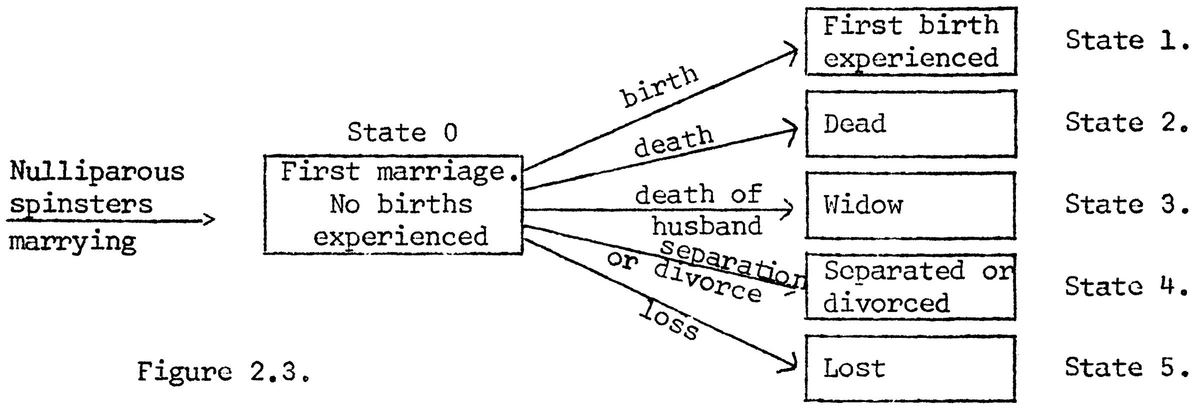

Figure 2.3.

the "carreer" of such a woman by a sample path in a six-state time-continuous age-dependent Markov chain where state 0 is transient and states 1 to 5 are absorbing. Sample paths of individual women will be regarded as stochastically independent.

Immigration will be permitted. (This has not been indicated in the diagram.) Immigrants will be treated at par with "native" members of the population.

§ 2.4. Let "time 0" be the moment at which the (y,u)-woman of § 2.2 is taken under observation, and let S(t) be the state to which she belongs at time $t \geq 0$. Then S(0)=0. We let $P_{yu}(A)$ denote the probability that the event A will occur to such a woman, and introduce

$$P_k(y,u,t) = P_{yu}\{S(t)=k\}$$

as the probability that she will occupy state k at time t, for k=0,1,...,5. We also introduce the force of decrement from cause k,

$$\mu_k(y,u) = \lim_{t \downarrow 0} P_k(y,u,t)/t \text{ for } k=1,2,\dots,5,$$

and the total force of decrement

$$\mu(y,u) = \lim_{t \downarrow 0} [1-P_0(y,u,t)]/t = \sum_{k=1}^{5} \mu_k(y,u).$$

Here $\mu_1(y,u)$ is the force of fertility for a (y,u)-woman. There is every reason to believe that this function will be select, i.e. depend on y and u separately.

$\mu_2(y,u)$ is her force of mortality. In line with certain theories of marital mortality we allow for the possibility that this function also is select. Similarly for μ_3, μ_4, and μ_5.

In the present setting decrements 2 to 5 only represent nuisance disturbing our main interest, which is the primary fertility giving rise to cause 1 of decrement.

We take each $\mu_k(y,u)$ to be a continuous function of u for any given y.

§ 2.5. It is easily shown that the following relations hold:

$$P_0(y,u,t) = \exp\{-\int_0^t \mu(y,u+\tau)\, d\tau\}, \text{ and}$$

$$P_k(y,u,t) = \int_0^t P_0(y,u,\tau)\, \mu_k(y,u+\tau)\, d\tau$$

for $k = 1,2,\ldots,5$. We introduce a random variable T as the moment at which decrement from state 0 takes place. Since $P_{yu}\{T>t\} = P_0(y,u,t)$, T will have the density $\mu(y,u+\tau)\, P_0(y,u,t)$ for $t>0$, and the mean value

$$\overset{\circ}{e}(y,u) = \int_0^{\omega-y-u} P_0(y,u,t)\, d\tau,$$

where ω as usual designates the highest possible live age. Specifically $\overset{\circ}{e}(y,0)$ will be the mean length of the stay in state 0 of a newly married woman at age y.

§ 2.6. Since decrement 1 is naturally inoperative after age ω_0 at menopause, and since this decrement is our main interest, we shall concentrate on ages below ω_0. We introduce the waiting time V in state 0 up to menopause or decrement, and see that $V=T$ if $T<\omega_0-y-u$, $V=\omega_0-y-u$ if $T\geq\omega_0-y-u$. Using $F_{yu}(v)$ for the distribution function of V, we get

$$F_{yu}(v) = \begin{cases} 1-P_0(y,u,v) \text{ for } 0\leq v<\omega_0-y-u, \\ 1 \text{ for } v\geq\omega_0-y-u. \end{cases}$$

Thus V has a mixed distribution with a density for $0<v<\omega_0-y-u$, and an atom of size $P_0(y,u,\omega_0-y-u)$ at $v = \omega_0-y-u$. It is easily seen that V has a mean value of

$$e(y,u) = \int_0^{\omega_0-y-u} P_0(y,u,v)\, dv.$$

Specifically, $e(y,0)$ is the mean waiting time in state 0 until menopause or decrement for a newly married woman at age y.

§ 2.7. As is seen from the formulae above, each probability $P_k(y,u,t)$ will be influenced by the values of the functions $\mu_{k'}(y,u)$ for $k' \neq k$ as well as of $\mu_k(y,u)$. Using a terminology due to Sverdrup (1961) we shall therefore call the $P_k(y,u,t)$ influenced probabilities.

By suppressing one or more of the five decrements listed in § 2.2 and taking the corresponding force or forces μ_k to be identically equal to zero, semi-influenced or partial probabilities at various levels result. One interesting possibility is to suppress decrement 5 and study probabilities as they would be in the absence of emigration and other kinds of loss from observation. More central to our present subject would be the case where all decrements except the first one are eliminated up to age ω_0, so that $\mu_k(y,u)=0$ for all $y<\omega_0$, $u \in [0,\omega_0-y]$, $k \geq 2$.

On this condition, the (partial) probability that a (y,u)-woman will have no birth before age y+u+t equals

$$\bar{P}_0(y,u,t) = \exp\{-\int_0^t \mu_1(y,u+\tau)\,d\tau\} \quad \text{for } 0 \leq t \leq \omega_0-y-u\,,$$

and the waiting time V will have a corresponding (partial) distribution function

$$\bar{F}_{yu}(v) = \begin{cases} 1-\bar{P}_0(y,u,v) & \text{for } 0 \leq v < \omega_0-y-u, \\ 1 & \text{for } v \geq \omega_0-y-u. \end{cases}$$

This probability distribution has the mean

$$\bar{e}(y,u) = \int_0^{\omega_0-y-u} \bar{P}_0(y,u,v)\,dv.$$

It should be noted that a quantity like $\bar{P}_0(y,u,t)$ has no interpretation as a probability within the original model. Specifically $\bar{P}_0(y,u,t)$ is not the conditional probability of being in state 0 at time t, given that a decrement from one of the causes 2 to 5 does not occur during the period $[0,\omega_0-y-u]$. The latter (conditional) probability equals

$$P_0^*(y,u,t) = P_0(y,u,t)/\{P_0(y,u,\omega_0-y-u)+P_1(y,u,\omega_0-y-u)\}$$

and is generally quite different from $\bar{P}_0(y,u,t)$.

§ 2.8. We shall conclude this chapter by defining a quantity called the mean age at first birth of a (y,u)-woman. Neither $y+u+\overset{\circ}{e}(y,u)$, $y+u+e(y,u)$, nor $y+u+\bar{e}(y,u)$ qualify for this name, because in practice the mean age mentioned would be estimated by averaging over those women only who actually experience a birth.

Let W be the waiting time until the occurrence of the first birth to a (y,u)-woman who will actually experience a birth. This random variable will have a distribution function

$$P_{yu}\{W \leq w\} = \tilde{F}_{yu}(w) = P_1(y,u,w)/P_1(y,u,\omega_0-y-u)$$

for $0 \leq w \leq \omega_0-y-u$, and its mean value equals

$$\tilde{e}(y,u) = \int_0^{\omega_0-y-u} w\, P_0(y,u,w)\mu_1(y,u+w)\, dw/P_1(y,u,\omega_0-y-u).$$

Then the mean age at first birth of a (y,u)-woman is $y+u+\tilde{e}(y,u)$.

3. Suggestions for extension of the model.

§ 3.1. The model of chapter 2 is a section of a larger model for marital fertility comprising all births rather than just the first one. In this larger model, the forces of transition may depend on number b of births experienced and, for $b \geq 1$, duration since the last birth, in addition to age at marriage and duration of marriage.

The model is otherwise sketched in figure 3.1. Although this is not indicated in the diagram, immigration may occur into any state of the system.

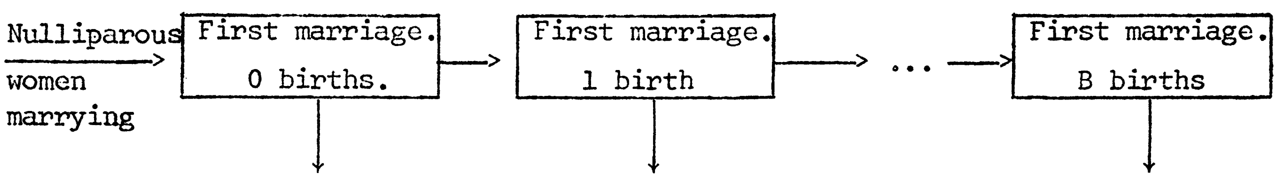

Figure 3.1.

Vertical arrows represent decrements 2 to 5.

When combined with a model with several marital states, this system can be further extended to permit a simultaneous analysis of fertility and nuptiality.

§ 3.2. Fertility generally depends on the area of residence of the woman. The model of chapter 2 can be extended to take care of this feature by splitting state 0 into a number, say S, of states, one for each area. (See figure 3.2.) Of course women may move direct from any one area to any other, but in the figure only arrows suggesting transitions between "neighbouring" areas have been drawn.

Such an extended model will permit a simultaneous analysis of migration and primary marital fertility.

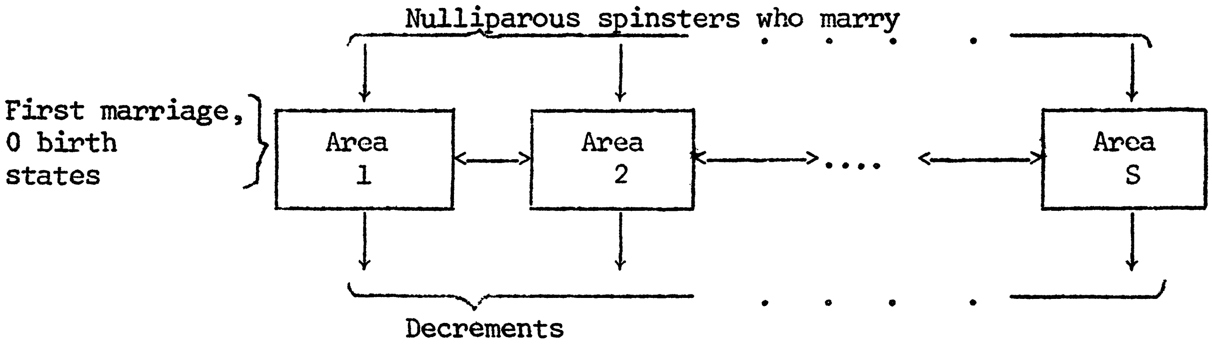

Figure 3.2.

§ 3.3. Unfortunately our data will not permit analysis by social status. Otherwise we could have used the model of § 3.2 directly by substituting "social stratum" for "area of residence" everywhere.

4. A general model for several decrements.

§ 4.1. Our next concern is a time-continuous Markov chain with one transient and K absorbing states. As in chapter 2 the transient state has the label "state 0" while the absorbing states are "state 1", "state 2",..., "state K". Observation starts at some "time zero".

If a persons has age x and belongs to state 0 at time 0, let $P_k(x,t)$ be the probability that he will be in state k at time $t \geq 0$. The force of decrement from cause k will be defined as

$$\mu_k(x) = \lim_{t \downarrow 0} P_k(x,t)/t \quad \text{for } k=1,2,\ldots,K,$$

and the total force of decrement from state 0 is

$$\mu(x) = \lim_{t \downarrow 0} \left[1-P_0(x,t)\right]/t = \sum_{k=1}^{K} \mu_k(x).$$

§ 4.2. Although we have spoken of the parameter x as the age of an individual and shall continue to do so, this parameter may have a quite different interpretation in some applications of the model.

We see that for any given y the model of chapter 2 may be regarded as a special case of the general model of § 4.1 with K=5 and with the marriage duration u corresponding to the age parameter x. Most of the results which we shall need for the fertility model are straightforward consequences of corresponding results for the general decrement model. In order not to unduly restrict the generality of our exposition we shall therefore concentrate on the general model in chapter 4, and shall subsequently indicate how results obtained may be applied to fertility.

§ 4.3. We turn first to the estimation of the forces μ_k of decrement over an age interval $[0,\zeta>$.

If it were known that

$$\mu_k(x) = g_k(x;\theta_1,\ldots,\theta_r) \quad \text{for } k=1,2,\ldots,K \text{ and for } x \in [0,\zeta>$$

for some set of functions $g_1,\ldots,g_K$ which were completely specified except for a set of parameters $\theta_1, \theta_2,\ldots, \theta_r$, our job would be to estimate the θ_i from the data. (One example of such a function is the Gompertz-Makeham formula $\alpha+\beta c^{\gamma x}$ for the force of mortality.) We shall leave such a situation aside and shall rather look into the case where, although some prior information is at hand about each of the μ_k, it is not sufficient for a parametrical formulation.

§ 4.4. Let $\mathcal{K} = \{1,2,\ldots,K\}$. For each $k \in \mathcal{K}$, we partition the age interval $[0,\zeta>$ into subintervals $[\zeta_k(0), \zeta_k(1)>$, $[\zeta_k(1), \zeta_k(2)>,\ldots,$ $[\zeta_k(I_k-1), \zeta_k(I_k)>$, where $\zeta_k(0) = 0$ and $\zeta_k(I_k) = \zeta$. (The number I_k of intervals as well as the points $\{\zeta_k(i):i=1,2,\ldots,I_k-1\}$ may be different for different k.) We will represent the values of μ_k in the age interval $[\zeta_k(i-1), \zeta_k(i)>$ by a simple parameter μ_{ki}.

Such prior knowledge as we have about each μ_k will be used in choosing the points $\zeta_k(i)$ with a view to achieving a reasonable approximation by the above technique.

§ 4.5. In some cases certain relationships may be known to exist among the μ_{ki}. If such knoewledge can be given a sufficiently precise form, it should be included into the analysis. In the case of primary marital fertility, however, such prior knowledge as the present author has of the subject is too vague to permit a precise mathematical formulation. We shall therefore proceed as when we do not recognize any relationships at all between the μ_{ki}. Specifically they will be taken to be functionally independent.

Our next step is to suggest estimators for the μ_{ki}.

§ 4.6. The set $\{\zeta_k(i):i=1,2,\ldots,I_k-1;k\in K\}$ of points in $\langle 0,\zeta]$ may be arranged in increasing order. The resulting set of points will be designated $\theta(1),\theta(2),\ldots,\theta(J-1)$, and we will let $\theta(0)=0,\theta(J)=\zeta$, so that $\theta(0)<\theta(1)<\ldots<\theta(J)$. Then for each (k,i) there exists a $j(k,i)$ such that $\zeta_k(i)=\theta\{j(k,i)\}$. Conversely for each $j\in\{0,1,\ldots,J\}$ there is at least one $\zeta_k(i)$ which equals $\theta(j)$.

For each $k\in K$ and $j\in \mathcal{J}=\{1,2,\ldots,J\}$ there exists an $i=i(k,j)$ such that $[\theta(j-1),\ \theta(j)\rangle \subseteq [\zeta_k(i-1),\ \zeta_k(i)\rangle$. The parameter value which represents the function $\mu_k(.)$ in the interval $[\theta(j-1),\ \theta(j)\rangle$ will then be $\mu_{k,i(k,j)}$, which we will designate $\lambda_k(j)$. Thus for given (k,i),

$$\lambda_k(j) = \mu_{ki} \quad \text{for } j=j(k,i-1)+1,j(k,i-1)+2,\ldots,j(k,i). \tag{1}$$

§ 4.7. We introduce $\lambda(j) = \sum_{k=1}^{K} \lambda_k(j)$ and shall prove that <u>the $\lambda(j)$ are linearly independent</u>.

Assume that there exist constants g_j for $j\in \mathcal{J}$ such that $\sum_{j=1}^{J} g_j\ \lambda(j)=0$. By (1) the sum here equals

$$\sum_{k=1}^{K} \sum_{i=1}^{I_k} \mu_{ki}\, G_{ki} \quad \text{with} \quad G_{ki} = \sum_{\nu=j(k,i-1)+1}^{j(k,i)} g_\nu .$$

Since the μ_{ki} are linearly independent, all $G_{ki}=0$. Let $j\in \mathcal{J}$. There then exists a pair (k_j,i_j) such that $j=j(k_j,i_j)$, since otherwise $\theta(j)$ would be a superfluous partitioning point. Thus

$$\sum_{\nu=j(k_j,i_j-1)+1}^{j} g_\nu=0 \quad \text{for all } j\in \mathcal{J},$$

from which follows that all $g_\nu=0$, as was to be proved. □

§ 4.8. Assume that observations are given for n persons each of whom moves through one or more states of our system. The histories observed of the individuals will be regarded as independent sample paths of the Markov chain.

Person no. ν is observed continuously from some age $x_\nu \in [0,\zeta\rangle$, at which he belongs to state 0, and until he leaves state 0, or until some age $x_\nu+z_\nu \in [0,\zeta\rangle$ if earlier. For the time being x_ν and z_ν will be taken as non-random quantities.

Let $x_\nu+W_\nu$ be his age at the moment at which observation is actually terminated. We also introduce binary variables $A_{\nu 1},\ldots, A_{\nu K}$, where $A_{\nu k}=1$ if person no. ν is observed to leave state 0 of cause k, $A_{\nu k}=0$ otherwise.

The age x_ν will belong to some interval $[\theta(s_\nu-1),\ \theta(s_\nu)\rangle$, and the age $x_\nu+W_\nu$ will belong to some interval $[\theta(R_\nu-1),\ \theta(R_\nu)\rangle$, where R_ν is random while s_ν is non-random. The likelihood corresponding to person no. ν may then be written in the form

$$\Lambda_\nu=\exp\{-\sum_{\alpha=s_\nu}^{R_\nu} \lambda(\alpha)[\theta(\alpha)-\theta(\alpha-1)] + \lambda(s_\nu)[x_\nu-\theta(s -1)] + \lambda(R_\nu)[\theta(R_\nu)-(x_\nu+W_\nu)]\} \prod_{k=1}^{K} \lambda_k(R_\nu)^{A_{\nu k}} .$$

Forming $\Lambda = \prod_{\nu=1}^{n} \Lambda_\nu$ and rearranging, we get

$$\Lambda = \exp\{-\sum_{j=1}^{J} \lambda(j)\, L(j)\} \prod_{k=1}^{K} \prod_{j=1}^{J} \lambda_k(j)^{B(k,j)} \tag{2}$$

where L(j) is the aggregated lifetime observed in state 0 for the age interval $[\theta(j-1),\ \theta(j)\rangle$, and B(k,j) is the number of decrements of cause k observed in this interval.

Our sampling distribution thus belongs to a Darmois-Koopman class.

§ 4.9. We shall make the additional assumption that

$$P\{L(j)>0\}=1 \quad \text{for } j=1,2,\ldots,J. \tag{3}$$

This will be achieved e.g. if each interval $[\theta(j-1),\ \theta(j)\rangle$ contains some x_ν.

By (3), L(1),...,L(J) and the B(k,j) are linearly independent with probability 1. Thus by § 4.7 and by the properties of Darmois-Koopman classes of sampling distributions the set $\{L(j):j\in \mathcal{J}\} \cup \{B(k,j):j\in\mathcal{J},\ k\in\mathcal{K}\}$ is a minimal sufficient statistic.

§ 4.10. We introduce

$$M(k,i) = \sum_{j=j(k,i-1)+1}^{j(k,i)} L(j) \quad (4)$$

and $$D(k,i) = \sum_{j=j(k,i-1)+1}^{j(k,i)} B(k,j). \quad (5)$$

$M(k,i)$ will be the aggregated lifetime observed in state 0 for the age interval $[\zeta_k(i-1), \zeta_k(i)\rangle$, and $D(k,i)$ will be the number of decrements from cause k observed in this interval. (2) may then be written in the form

$$\Lambda = \exp\{-\sum_{k=1}^{K} \sum_{i=1}^{I_k} \mu_{ki} M(k,i)\} \prod_{k=1}^{K} \prod_{i=1}^{I_k} \mu_{ki}^{D(k,i)}.$$

Thus the maximum likelihood estimators for the μ_{ki} will be

$$\hat{\mu}_{ki} = D(k,i)/M(k,i). \quad (6)$$

Our estimators are occurrence/exposure rates, as is quite similar to results obtained in other models of the same kind. (Sverdrup (1961, 1965).)

§ 4.11. Let $D_\nu(k,i)$ be the number of decrements of cause k in $[\zeta_k(i-1), \zeta_k(i)\rangle$ for person no. ν, and let $M(k,i)$ be his observed lifetime in state 0 for this interval. Furthermore let $\phi(x_\nu, z_\nu, t)=1$ for $t \in [x_\nu, x_\nu+z_\nu]$, $\phi(x_\nu, z_\nu, t) = 0$ otherwise. Then

$$ED_\nu(k,i) = \int_{\zeta_k(i-1)}^{\zeta_k(i)} \phi(x_\nu, z_\nu, t)\, \mu_k(t)\, P_0(x_\nu, t)\, dt, \text{ and}$$

$$EM_\nu(k,i) = \int_{\zeta_k(i-1)}^{\zeta_k(i)} \phi(x_\nu, z_\nu, t)\, P_0(x_\nu, t)\, dt.$$

Since the function $\mu_k(.)$ is represented by the parameter μ_{ki} in $[\zeta_k(i-1), \zeta_k(i)\rangle$, we therefore get $ED_\nu(k,i) = \mu_{ki} EM_\nu(k,i)$. By summation over ν we obtain

$$ED(k,i) = \mu_{ki} EM(k,i). \quad (7)$$

§ 4.12. We shall assume that for each (k,i) there exists a positive constant $\alpha(k,i)$ such that

$$\lim_{n\to\infty} E\bar{M}(k,i) = \alpha(k,i) \tag{8}$$

where $\bar{M}(k,i) = M(k,i)/n$. Since var $\{\bar{M}(k,i)\}$ exists and approaches zero as $n\to\infty$, we get $\operatorname{plim}_{n\to\infty} \bar{M}(k,i) = \alpha(k,i)$. (The proof is trivial. See e.g. Hoem (1968, § 5.2).) By (7) and a theorem attributed to Slutsky, we conclude that the $\hat{\mu}_{ki}$ are consistent for the μ_{ki} as $n\to\infty$.

§ 4.13. When the (x_ν, z_ν) are given quantities and not all equal to each other, the pairs $\{D_\nu(k,i), M_\nu(k,i)\}$ for $\nu = 1,2,\ldots,n$ are not generally identically distributed. In certain cases it is possible to regard the (x_ν, z_ν) as values of random variables $\{(X_\nu, Z_\nu): \nu = 1,2,\ldots,n\}$, where the pairs (X_ν, Z_ν) are independent and identically distributed with some distribution function $G(x,z)$. Situations quite similar to this have been considered previously by Sverdrup (1961, 1965) and Hoem (1968). We shall not go deeply into the matter, therefore, but shall only list the following results which hold in such cases:

(i) The $\hat{\mu}_{ki}$ are still consistent as $n\to\infty$.

(ii) As $n\to\infty$, the $\hat{\mu}_{ki}$ are asymptotically normal and independent with means μ_{ki}, and

$$\text{as.var. } \hat{\mu}_{ki} = \frac{1}{n}\frac{\mu_{ki}}{\Delta_{ki}} \tag{9}$$

with $\Delta_{ki} = \int_\Omega \int_{\zeta_k(i-1)}^{\zeta_k(i)} \phi(x,z,t)\, P_0(x,t)\, dt\, dG(x,z)$,

where Ω is the area of variations for the (X_ν, Z_ν).

(iii) If G is fully known, the $\hat{\mu}_{ki}$ are optimal Fisher consistent estimators for the μ_{ki}.

§ 4.14. We note that if the value of some μ_{ki} is known beforehand, the results of the present chapter still hold with only obvious modifications. Specifically (6) applies for unknown μ_{ki} even if it is given that $\mu_{k'i}(x) \equiv 0$ for some $k' \neq k$. (Cf. § 2.7.)

§ 4.15. One may want to compare values of forces of decrement for two or more populations. Let there be S such populations, let the age interval $[0,\zeta\rangle$ be common to them all, and assume for simplicity that we may use the same partitioning points $\zeta_k(i)$ for all populations. The k-th force of decrement in population s will be represented by a parameter $\mu_{ki}^{(s)}$ in the age interval $[\zeta_k(i-1), \zeta_k(i)\rangle$. One may wish to test a hypothesis of the form

$$H_0: \mu_{k(r),i(r)}^{(1)} = \mu_{k(r),i(r)}^{(2)} = \dots = \mu_{k(r),i(r)}^{(S)} \quad \text{for } r=1,2,\dots,R,$$

against an alternative where at least one of the equality signs specified does not hold.

Let $M^{(s)}(k,i)$ and $D^{(s)}(k,i)$ be defined for population s by relations similar to (4) and (5), and let

$$M(k,i) = \sum_{s=1}^{S} M^{(s)}(k,i), \; D(k,i) = \sum_{s=1}^{S} D^{(s)}(k,i),$$

$$\hat{\mu}_{ki}^{(s)} = D^{(s)}(k,i)/M^{(s)}(k,i), \text{ and } \hat{\mu}_{ki} = D(k,i)/M(k,i)$$

in the present paragraph. Designating the likelihood ratio by Q, we then get

$$-2 \ln Q = 2 \sum_{r=1}^{R} \{ \sum_{s=1}^{S} D^{(s)}(k(r),i(r)) \ln \hat{\mu}_{k(r),i(r)}^{(s)} - D(k(r),i(r) \ln \hat{\mu}_{k(r),i(r)} \}.$$

For sufficiently large n, -2 ln Q will be approximately χ^2-distributed with R(S-1) degrees of freedom under H_0, at least if the (x_ν,z_ν) behave as indicated in § 4.13. This gives a test criterion for H_0.

5. Application to the model for primary marital fertility.

§ 5.1. As indicated in § 4.2 already, the general theory of chapter 4 is easily adapted to our model for primary marital fertility. Consider the fertility of women whose age at marriage was y, say, and assume that an open sub-population of such women is kept under observation over some time period $[0,\tau]$.

At time 0 there is a certain number, say n_0, of women in the sub-population at ages $y+u_1$, $y+u_2, \dots,$ $y+u_{n_0}$, where all $u_\nu \in [0,\zeta\rangle$. For the ν-th of these women, the maximal length of the period of observation is $z_\nu = \tau-u_\nu$.

At certain times $\tau_1, \tau_2, \ldots$, women enter the sub-population through marriage at age y. We enumerate these newly-weds as women no. n_0+1, $n_0+2, \ldots$, n_0+n_1, say, and see that the maximal period of observation for woman no. $n_0+\nu$ is $z_{n_0+\nu} = \tau-\tau_\nu$. We also introduce the "durations" $u_{n_0+\nu} = 0$ of the marriages at the moment when these women are taken under observation.

Finally women immigrate into the sub-population from abroad at times $\tau_1', \ldots, \tau_{n_2}'$ with marriage durations $u_{n_0+n_1+1}, \ldots, u_{n_0+n_1+n_2}$ and maximal periods f observation $z_{n_0+n_1+\nu} = \tau-\tau_\nu'$ for $\nu = 1,2,\ldots,n_2$.

For each woman in the sub-population there is thus a pair (u_ν, z_ν) corresponding to the pair (x_ν, z_ν) of § 4.8. Whether it is possible to regard the (u_ν, z_ν) as values of independent and identically distributed random variables, depends on the circumstances.

§ 5.2. Altogether $n = n_0+n_1+n_2$ women are observed. In some cases it may be natural to regard n_1 and n_2 as random variables. This will make n random. Presumably the value of n will be determined exogeneously to the fertility model in such a case. When this value is given, the theory of chapter 4 may then be applied as if n were a known parameter.

§ 5.3. Primary fertility for women with different ages at marriage may be compared by the method of § 4.15. Consider for instance the comparison of the primary fertility of women with ages y_1 and y_2 at marriage, where $y_1 < y_2$. Since $\mu_1(y_2,u) = 0$ for all $u > \omega_0 - y_2$ while $\mu_1(y_1,u)$ may be positive for $\omega_0 - y_2 < u < \omega_0 - y_1$, comparison is actually meaningful only for durations less than $\zeta = \omega_0 - y_2$. For such durations, let the partitioning $\{\zeta_1(i): i=0,1,2,\ldots,I_1\}$ be common to the two sub-populations. The hypothesis then takes the form

$$H_0^* : \mu_{1i}^{(1)} = \mu_{1i}^{(2)} \quad \text{for } i=1,2,\ldots,I_1,$$

where $\mu_{1i}^{(s)}$ is the force of fertility parameter in the duration interval $[\zeta_1(i-1), \zeta_1(i)\rangle$ for the woman with age y_s at marriage, s=1,2. The χ^2-statistic will have the form

$$2 \sum_{i=1}^{I_1} \{ \sum_{s=1}^{2} D^{(s)}(1,i) \ln \hat{\mu}_{1i}^{(s)} - D(1,i) \ln \hat{\mu}_{1i} \}.$$

There are I_1 degrees of freedom.

§ 5.4. If the hypothesis in § 5.3 is rejected in a set of data, the conclusion to be drawn is that the two groups of women have different primary fertility for durations in $[0,\zeta>$. This is probably not always a quite satisfactory answer, as one would presumably also like to know something about the structure of the difference.

As an example consider the case where one wishes to investigate a theory to the effect that women with age y_1 at marriage get their first births quicker than those who marry at age y_2. One might formulate this theory more precisely in terms of the distribution functions introduced in § 2.8, and test the hypothesis.

$$H_0' : \tilde{F}_{y_1 0}(w) \leqq \tilde{F}_{y_2 0}(w) \quad \text{for } 0<w<\zeta = w_0 - y_2$$

against the alternative $\tilde{F}_{y_1 0}(w) > \tilde{F}_{y_2 0}(w)$ for such ω.

If each woman has been observed from marriage and until age ω_0 or until decrement from state 0, one of the standard non-parametric tests (Wilcoxon two-sample test; Smirnov-Kolmogorov) may be applied. If the data consist of observations taken only during fractions of this period, as we have allowed for previously in this paper, existing standard methods unfortunately seem inadequate, and it appears to be an open question how one would test H_0'.

There are several other ways in which the theory mentioned may be formulated, e.g. in terms of the functions $P_0^*(y_1,0,.)$ and $P_0^*(y_2,0,.)$ of § 2.7, in terms of $\tilde{e}(y_1,0)$ and $\tilde{e}(y_2,0)$ of § 2.8, or in terms of the medians or other fractiles of the distributions $\tilde{F}_{y_1 0}$ and $\tilde{F}_{y_2 0}$. In any case one will have difficulties if observations have been made only during a part of the lifespan of each person.

No known test seems to be known for a hypothesis like H_0' formulated in terms of the $\bar{F}_{y_1 0}$ and $\bar{F}_{y_2 0}$ of § 2.7.

6. Concluding remarks.

§ 6.1. One of the merits of the probabilistic approach chosen in the present paper is its precise mathematical formulation of ideas which are generally expressed only verbally. This formulation permits a rational derivation of results which would otherwise have to be established by intuition and common sense, like the occurrence/exposure rates of § 4.10. It discloses some possible pitfalls along the road, as suggested in §§ 2.7 and 2.8, and it enables us to arrive at results which could never have been found by non-mathematical arguments. (§§ 4.11 to 4.13,etc.)

It certainly does not solve all the problems of analysis which the demographer faces, however, not even all those that can be given a mathematical formulation. (§ 5.4.) There also remains the task of giving such a formulation to all elements in the techniques developed, such as how "prior information" will actually influence the choice of the partitioning points $\zeta_k(i)$ of § 4.4.

7. Acknowledgement.

I am grateful to cand.act. Bjørn Tønnesen, who has proof-read the present paper in manuscript.

8. References.

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

Selmer, Ernst S. (1967): "Registration Numbers in Norway: Some Applied Number Theory and Psychology." Journ. Roy. Stat. Soc., Ser. A, **130**, 225-231.

Skaug, Helge (1967): "The Norwegian Population Register System." Proc. Int. Sympos. on Autom. of Popul. Register Systems, Jerusalem 1967, **1**, 3-39.

Skaug, Helge (1968): "Oversikt over Det sentrale personregister i Statistisk Sentralbyrå." Mimeographed. Oslo: The CBS of Norway.

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

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.