Sampling Bias in Population Studies — How to Use the Lexis Diagram

Jens Lund

The Royal Veterinary and Agricultural University,

Department of Mathematics and Physics, Denmark

June 7, 1999

Running headline: Sampling in the Lexis Diagram

Abstract

Modified versions of the life time distribution are often used in survival analysis. The modifications depend on how we choose individuals for the study and on the assumptions on the behavior of the population. A rigorous point process description of the Lexis diagram is used to make the sampling mechanisms and the preconditions transparent. The point process description gives a framework to handle all possible sampling patterns. The setup is generalized so it can handle more complicated life descriptions than just lifetimes, and the disability model is used as an example. Two setups can be used. Conditional on the birthtimes, the modifications of the life time distribution are left truncation and either right censoring or truncation. Assuming that the birthtimes can be described by a Poisson process the modifications are length bias and the recurrence time distribution known from renewal theory.

Keywords: Survival analysis, sampling bias, Lexis diagram, Poisson process, point process.

1 Introduction

The Lexis diagram (Lexis, 1875; Keiding, 1990) is a useful tool when analysing lifetimes as it is important carefully to consider the way individuals are chosen for the study. The Lexis diagram is a coordinate system with calendar time in the horizontal direction and age of individuals in the vertical direction, and each individual is represented by a line. The selection of individuals often introduces biased or otherwise modified lifetime distributions. This paper extends the point process description of the Lexis diagram introduced by Brillinger (1986) and gives a derivation of the likelihood functions associated with five sampling patterns. All these sampling patterns can be handled by the common point process framework provided. For four of the five patterns

the likelihood functions are derived in two contexts: conditional on the birthtimes and under the assumption that the process of birthtimes is a Poisson process.

The conditional approach is the one most often used in classical survival analysis (Andersen *et al.*, 1993), and the lifetime distributions are often left truncated and right censored in this setup. The usual non-parametric estimators such as the Kaplan-Meier estimator and the Nelson-Aalen estimator can be used here.

Further assumptions are required in cross sectional studies and in studies where for example the age at entry to the study is unobserved, or the age is observed but the time of death is unobserved. These situations have previously been handled in a renewal setup, and renewal processes are sometimes used in arguments for survival analysis likelihood functions, even though no "replacement" takes place (McClean & Devine, 1995; Winter & Földes, 1988; Denby & Vardi, 1986; Vardi, 1989, Problem B). Here we use an alternative approach based on the point process description of the Lexis diagram (Brillinger, 1986) and the assumption that the birthtimes can be described by a Poisson process (Simon, 1980; Brillinger, 1986; Keiding, 1991; Keiding, 1992). The Poisson assumption provides us with information on the survival until the start of the study. We now offer a framework based on the individual lifetimes to derive the same likelihood functions as in renewal theory, and we recognize the same forms of length bias as in renewal theory. The likelihoods are the same and so are the estimators, but the proofs of consistency and week convergence of estimators from renewal theory might not be valid in our setup.

We also consider a generalization of the Lexis diagram which expands the available class of models — and thus the usefulness of the mathematical description — considerably. The general setup provides a common framework for handling multi-state models, for example, and all sampling patterns can be handled in a unified way. The disability model (Keiding, 1991) is used as an example on how this "tool box" can be used. Also, the general setup allows for inclusion of covariates, as considered by Simon (1980) in a simple example. Analysis conditional on the birthtimes and covariates can be handled for quite complicated models (Andersen *et al.*, 1993). Our framework allows a more detailed study of the consequences of including covariates in models where the birthtimes are considered random.

The paper is organized as follows: In Section 2 we discuss the basic model and summarize the sampling patterns. The likelihood functions are derived in Section 3 (conditional on the birthtimes) and Section 4 (with the Poisson assumption). We discuss the Poisson assumption in Section 5. In Section 6 we generalize the setup and use the disability model as an example.

2 Sampling in the Lexis diagram

2.1 The Lexis diagram

The Lexis diagram is a coordinate system with calendar time in the horizontal direction and age

of individuals in the vertical direction. An example of a Lexis diagram is the top left sub-figure of Fig. 1. An individual is represented by a *life line* with slope 1 from the point of birth to the point of death. If σ is the birthtime and x is the lifetime, then the life line is from $(\sigma, 0)$ to $(\sigma + x, x)$. We often think of the Lexis diagram as the points $(\sigma + x, x)$ of deaths. Fig. 1 here

Let $(\sigma_i)_{i\in I}$ denote the collection of all birthtimes. The index set I can be taken as the integers $\mathbb{Z}$ with the conventions that $\cdots < \sigma_i < \sigma_{i+1} < \cdots$ and $\cdots \sigma_{-1} < 0 \leq \sigma_0 \cdots$. We assume that all birthtimes are distinct, but this is not critical. Assume further that there are only finitely many birthtimes in every bounded Borel set $B \subseteq \mathbb{R}$. We think of $(\sigma_i)_{i\in I}$ as a point process on $\mathbb{R}$ and denote by η the associated counting measure (Daley & Vere-Jones, 1988; Karr, 1991; Stoyan *et al.*, 1995). Let ε_c be the Dirac measure at c, then $\eta = \sum_{i\in I} \varepsilon_{\sigma_i}$ and $\eta(B)$ is the number of birthtimes in B. We call the point process $(\sigma_i)_{i\in I}$ or the equivalent η the *process of birthtimes.* The advantage of having the process of birthtimes defined on the whole axis $\mathbb{R}$ will be evident in Theorem 4.1 and its applications like Example 4.3. Compared to for example Simon (1980) we avoid asymptotic arguments like "in the distant past" or "as t tends to ∞".

For all $i \in I$, associate with the individual born at time σ_i a lifetime X_i in $\mathbb{R}_0 = [0, \infty[$.

Definition 2.1 The *Lexis point process* μ is the point process $(\sigma_i, X_i)_{i\in I}$ on $(\mathbb{R} \times \mathbb{R}_0, \mathcal{B} \otimes \mathcal{B}_0)$, that is $\mu = \sum_{i\in I} \varepsilon_{(\sigma_i, X_i)}$.

The distinction between the Lexis diagram and the Lexis point process is important. The points of deaths in the Lexis diagram are given by the point process $(\sigma_i + X_i, X_i)_{i\in I}$, but this cannot be used as a definition of a Lexis point process in the general setup in Section 6.

2.2 Summary of the sampling patterns

In this section we review some natural ways to sample individuals in the Lexis diagram as illustrated in Fig. 1. Let $t_1 < t_2 < t_3$ be given time points.

Cohort study We observe all individuals born in $]t_1, t_2]$. Observation might be stopped at a time t_3. In this case the actual lifetimes of individuals still alive at time t_3 are not observed.

Time window All deaths (times and ages) in the set $[t_1, t_2] \times \mathbb{R}_0$ are observed. An observation of this type could arise in register based study.

Cross sectional study The ages of individuals alive at time t are recorded.

Synthetic cohort All deaths and all individuals alive during the time period $[t_1, t_2]$ are observed. The observation corresponds to a time window and a cross section at time t_2. This is the classic case as synthetic cohorts are used in the construction of life tables.

Follow-up on cross sectional study At time t_2 we observe whether individuals alive at time t_1 are dead or alive. If alive we observe the age, and if dead we observe the age and time of death.

2.3 The basic assumption

In this section we introduce the basic probabilistic model and some notation. We need an assumption on conditional independence.

Assumption 2.2 The lifetimes X_i ($i \in I$) are independent conditional on the process η of birthtimes $(\sigma_i)_{i\in I}$. The distribution of X_i depends on η only through σ_i.

Let P be a Markov kernel from $(\mathbb{R}, \mathcal{B})$ to $(\mathbb{R}_0, \mathcal{B}_0)$ describing the distribution of X_i given σ_i. That is, P is a map from $\mathbb{R} \times \mathcal{B}_0$ to $[0,1]$, such that $P(\sigma, \cdot)$ is a probability measure on $(\mathbb{R}_0, \mathcal{B}_0)$ for all $\sigma \in \mathbb{R}$, and the mapping $\sigma \mapsto P(\sigma, A)$ from $\mathbb{R}$ to $[0,1]$ is measurable for all $A \in \mathcal{B}_0$. We often write $P_\sigma(A)$ instead of $P(\sigma, A)$. Let F_σ denote the distribution function for P_σ, and $\bar{F}_\sigma = 1 - F_\sigma$ the corresponding survival function. We focus on the situation where the lifetime distribution has a density f_σ w.r.t. the Lebesgue measure l_0 on $\mathbb{R}_0$.

It is tacitly understood that all relevant expectations are finite. In particular the expected life time is finite and the mean number of individuals alive at any given time t should be finite.

3 Analysis conditional on selected birthtimes

In this section we condition on the birthtimes for the individuals we observe and prove a theorem stating that given the birthtimes for the individuals we observe, the lifetimes X_i for those individuals are independent. Furthermore the lifetimes must be analysed conditional on the event that we observe the individual, and their distributions do not depend on the distribution of the process of birthtimes. In most cases we get a traditional analysis with left truncation and right censoring of the lifetimes.

Let $(h_\sigma)_{\sigma\in\mathbb{R}}$ be a family of measurable functions $h_\sigma : \mathbb{R}_0 \to \{0,1\}$ indicating whether we observe the individual born at time σ with lifetime x or not. If $h_\sigma(x) = 1$ we get information on the individual, if $h_\sigma(x) = 0$ we do not get information on the individual.

Theorem 3.1 *Given the birthtimes σ_i for the individuals we observe ($i \in I : h_{\sigma_i}(X_i) = 1$), the lifetimes X_i for those individuals are independent, and the lifetime X_i for $i \in I$ with $h_{\sigma_i}(X_i) = 1$ has distribution $P_{\sigma_i}(\,\cdot \,|\, h_{\sigma_i}(X_i) = 1)$.*

Proof Assumption 2.2 states that, given $\eta = \sum_{i\in I} \varepsilon_{\sigma_i}$, the lifetimes $(X_i)_{i\in I}$ are independent and $X_i \sim P_{\sigma_i}$.

Now condition on both η and the family of functions $(h_{\sigma_i}(X_i))_{i\in I}$. Then $(X_i)_{i\in I}$ are independent and $X_i \sim P_{\sigma_i}(\,\cdot \,|\, h_{\sigma_i}(X_i))$. To see this, note that we have

$$\mathcal{L}\left((X_i)_{i\in\bar{I}} \middle| \eta = \sum_{i\in I} \varepsilon_{\sigma_i} \right) = \otimes_{i\in\bar{I}} P_{\sigma_i}(X_i \in \,\cdot\,)$$

where $\tilde{I}$ denotes any finite subset of I and $\otimes_{i\in\bar{I}} P_{\sigma_i}$ is the product measure. The further conditioning on $(h_{\sigma_i}(X_i))_{i\in I}$ has a product structure, and thus by the principle of repeated conditioning (Jacobsen, 1982, App. 1) the independence is preserved and

$$\mathcal{L}\left((X_i)_{i\in\bar{I}}\Big|\eta=\sum_{i\in I}\varepsilon_{\sigma_i},(h_{\sigma_i}(X_i))_{i\in I}\right)=\otimes_{i\in\bar{I}}P_{\sigma_i}(X_i\in\,\cdot\,|h_{\sigma_i}(X_i)).$$

Projection on the coordinates $i\in I: h_{\sigma_i}(X_i)=1$ gives that $(X_i)_{i\in I:h_{\sigma_i}(X_i)=1}$ are independent given η and $(h_{\sigma_i}(X_i))_{i\in I}$, and $X_i\sim P_{\sigma_i}(\,\cdot\,|h_{\sigma_i}(X_i)=1)$.

This conditional distribution does not depend on those σ_i and $h_{\sigma_i}(X_i)$ for which $h_{\sigma_i}(X_i)=0$, and they can be omitted from the conditioning. □

Let O be a deterministic measurable subset of the Lexis diagram $\mathbb{R}\times\mathbb{R}_0$ with the property that all possible life lines intersect O at most once. O indicates which area of the Lexis diagram we observe. Imagine an infinite life line attached to the individual born at time σ. Let $a_{\text{in},\sigma}$ be the age when the life line enters O, and let $a_{\text{out},\sigma}$ be the age when the life line leaves O again. If the life line does not leave O again, then $a_{\text{out},\sigma}=\infty$. In case the life line does not intersect O we define $a_{\text{in},\sigma}=a_{\text{out},\sigma}=\infty$. The ages $a_{\text{in},\sigma}$ and $a_{\text{out},\sigma}$ are deterministic conditionally on σ, so the truncations and censoring in the following are independent (Andersen *et al.*, 1993, Sec. III.2, III.3). Truncation is inference in a conditional distribution, and left truncation is when we condition on the lifetime being larger than a certain age. Censoring, on the other hand, is an incomplete observation. Right censoring occurs when we only observe the lifetime to be larger than a certain age, but do not observe the lifetime itself.

We consider two situations. First, let $h_\sigma(x)$ be the indicator function for the event $(x\geq a_{\text{in},\sigma})$. This means we observe life lines intersecting the observation set O. Conditional on the birthtimes, the observed lifetimes are independent and X_j for $j\in J=\{i: h_{\sigma_i}(x_i)=1\}$ must be analysed left truncated at ages a_{in,σ_j}. Because we only observe points of death in the set O, the lifetimes are right censored at ages a_{out,σ_j}. If the Markov kernel $P_\sigma(\,\cdot\,)$ does not depend on σ we immediately have the Kaplan-Meier estimator for the survival function, the Nelson-Aalen estimator for the integrated hazard, and asymptotic results of their behaviour (Andersen *et al.*, 1993, Sec. IV.1.5). Second, let $h_\sigma(x)$ be the indicator function for the event $(a_{\text{in},\sigma}\leq x<a_{\text{out},\sigma})$. This means we only observe individuals whose points of death are in the set O. Conditional on the birthtimes $\sigma_j=t_j-x_j$ for the $|J_1|$ deaths observed at $(t_j,x_j)_{j\in J_1}$, the lifetimes X_j for $j\in J_1$ are independent and must be analysed left truncated at age a_{in,σ_j} and right truncated at age a_{out,σ_j}. The nonparametric maximum likelihood estimator is a special case of the Turnbull estimator (Turnbull, 1976), which in turn is an example of the EM algorithm.

We are now ready to analyse four of the five sampling patterns. The lifetimes X_j in the *cohort study* should be analysed as independent and from the distribution f_{σ_j} as $a_{\text{in},\sigma_j}=0$ and $a_{\text{out},\sigma_j}=\infty$ for individuals we observe. If observation is stopped at time t_3 we analyse the lifetimes right censored at ages a_{out,σ_j}. In the *time window* the lifetimes are left and right truncated at age a_{in,σ_j} and

a_{out,σ_j} respectively. Both the *synthetic cohort* and the *follow up on cross sectional study* have the lifetimes left truncated and right censored at age a_{in,σ_j} and a_{out,σ_j} respectively.

For *cross sectional data* the situation is different. Indeed, let $h_\sigma(x)$ be the indicator function for the event $(\sigma \le t, x \ge t - \sigma)$ and let $(a_j)_{j\in J_2}$ denote the $|J_2|$ ages of individuals alive at time t. Conditionally on the birthtimes $\sigma_j = t - a_j$ the lifetimes X_j are independent and must be analysed in the conditional distribution given $X_j \ge a_{\text{in},\sigma_j} = a_j$. Unfortunately this is everything we observe about X_j, and thus the likelihood function equals 1. However, with a model for the birthtimes η, cross sectional data can be analysed.

4 A Poisson model for the process of birthtimes

In the cross sectional study we need a model for the birthtimes so we can use information on survival until the start of the study. We see various forms of length bias instead of left truncation in all sampling patterns. The Poisson process is mathematically convenient as a process of birthtimes (Brillinger, 1986) and we discuss it further in Section 5.1.

4.1 General Poisson setup

In this section we model the process of birthtimes η as a Poisson process. That is, the numbers of births in disjoint sets are independent and the number of births $\eta(B)$ in a bounded Borel set $B \in \mathcal{B}$ is Poisson distributed with mean $\mathrm{E}\,\eta(B) = \Phi(B) < \infty$. The intensity measure Φ for the process of birthtimes is not allowed to have any atoms, and we assume it has density φ w.r.t. the Lebesgue measure l on $\mathbb{R}$. We call φ the intensity of η.

The following theorem states that the Lexis point process is a Poisson process when the process of birthtimes is a Poisson process. The theorem is known as "positioning dependent marking" (Karr, 1991, Example 1.28 and 2.24, Exercise 1.11) and is mentioned by Daley & Vere-Jones (1988, pp. 205–206). It is formulated in a general version that will be needed in Section 6. For now, the measurable spaces $(E, \mathcal{E})$ and $(G, \mathcal{G})$ could be substituted by $(\mathbb{R}, \mathcal{B})$ and $(\mathbb{R}_0, \mathcal{B}_0)$ respectively, and Y_i could be substituted by X_i.

Theorem 4.1 *Let η be a Poisson process on $(E, \mathcal{E})$ with intensity measure Φ and let P be a Markov kernel from $(E, \mathcal{E})$ to $(G, \mathcal{G})$. Assume that, given $\eta = \sum_{i\in I} \varepsilon_{\sigma_i}$, $(Y_i)_{i\in I}$ is a family of independent random variables, Y_i with distribution $P(\sigma_i, \cdot\,)$ on $(G, \mathcal{G})$ for all $i \in I$. Then the Lexis point process $\mu = \sum_{i\in I} \varepsilon_{(\sigma_i, Y_i)}$ is a Poisson process on $(E \times G, \mathcal{E} \otimes \mathcal{G})$ with intensity measure Λ given by*

$$\Lambda(A \times B) = \int_A P(\sigma, B)\, \Phi(d\sigma) \tag{1}$$

for $A \in \mathcal{E}$ and $B \in \mathcal{G}$.

Proof Brillinger (1986) proved the result in the case of Example 4.2 below, but he did not state the result in the generality of the theorem. His proof carries over to the general situation and is fairly straightforward using Laplace functionals. □

With the notation from Section 2.3, the intensity measure Λ for the Lexis point process μ has density $\lambda(\sigma, x) = \varphi(\sigma) f_\sigma(x)$ w.r.t. $l \otimes l_0$ according to the theorem. We call λ the intensity for μ. The following two examples introduce the important point processes $\tilde{\mu}$ and N^t.

Example 4.2 (Points in the Lexis diagram) The points of deaths in the Lexis diagram are a bijective transformation $(t, x) \overset{h}{\mapsto} (t + x, x)$ of the points in the Lexis point process. The map h transforms μ to the point process of deaths in the Lexis diagram $\tilde{\mu} = h(\mu) = \sum_{i \in I} \varepsilon_{(\sigma_i + X_i, X_i)}$, and $\tilde{\mu}$ is a Poisson process with intensity measure $\tilde{\Lambda} = h(\Lambda)$ that has intensity $\tilde{\lambda}(t, a) = \varphi(t-a) f_{t-a}(a)$ (Brillinger, 1986). That is, the expected number of deaths in calendar time $]t, t + dt]$ and at age $]a, a + da]$ is $\tilde{\lambda}(t, a)\, dt\, da$. □

Example 4.3 (Cross section) For fixed $t \in \mathbb{R}$ define $K : \mathcal{B}_0 \to \mathcal{B} \otimes \mathcal{B}_0$ by $K(A) = \{(\sigma, x) : \sigma < t, x > t - \sigma, t - \sigma \in A\}$. Then $N^t(A) = \mu(K(A))$ is the number of individuals alive at time t with age in A. The theorem in the appendix states that for t fixed, $N^t(\,\cdot\,)$ has a version that is a Poisson process on $\mathbb{R}_0$. The intensity is $\lambda^t(a) = \varphi(t-a)\bar{F}_{t-a}(a)$. This is seen from

$$\begin{aligned} \mathrm{E}\, N^t([0,a]) = \Lambda^t([0,a]) = \Lambda(K([0,a])) &= \int_{K([0,a])} \lambda(s,x)\, d(s,x) \\ &= \int_{t-a}^{t} \int_{t-s}^{\infty} \varphi(s) f_s(x)\, dx\, ds \\ &= \int_0^a \varphi(t-u) \bar{F}_{t-u}(u)\, du, \end{aligned}$$

by substitution of s by $u = t - s$. (The second equality is due to the appendix, the fourth equality is due to Theorem 4.1.) Keiding (1991, p. 379) stated a somewhat similar result without making the underlying assumptions clear. □

Example 4.4 (Independence) We can use Theorem 4.1 and the basic independence property of the Poisson process to infer results about independence of relevant observations in survival analysis. The two point processes N^t and $\tilde{\mu}$ are of course not independent. However N^t and the restriction of $\tilde{\mu}$ to $]-\infty, t[\times\mathbb{R}_0$ have independent versions because they concern disjoint parts of the Lexis diagram. This means that the number and ages of individuals alive at time t are independent of deaths before time t. Similar arguments were also used by Brillinger (1986). □

We need a few general results about Poisson processes (Daley & Vere-Jones, 1988). If μ_1 and μ_2 are independent Poisson processes on $(S_1, \mathbb{S}_1)$ and $(S_2, \mathbb{S}_2)$ respectively, they correspond to one Poisson process on the disjoint union $S_1 \dot{\cup} S_2 = \{(j, s) \,|\, (j = 1, s \in S_1) \text{ or } (j = 2, s \in S_2)\}$. Let μ be a Poisson process on $(S, \mathbb{S})$ with intensity measure Λ and let $A \in \mathbb{S}$ such that $0 < \Lambda(A) < \infty$. Given $\mu(A) = n = |J|$, the points $(x_j)_{j \in J}$ of μ on A are i.i.d. with distribution $\pi(\,\cdot\,) = \Lambda(\,\cdot \cap A)/\Lambda(A)$. Assume that Λ has intensity λ w.r.t. a suitable reference measure. The likelihood for the observation $(x_j)_{j \in J}$ of μ on A is $L \propto \exp(-\Lambda(A)) \prod_{j \in J} \lambda(x_j)$.

These properties of Poisson processes together with the three basic examples 4.2, 4.3, and 4.4 make up the basic parts in the following. For each of the sampling patterns in Section 2.2 we investigate how to analyse the observations. The procedure is to identify the sampling pattern as an observation of $\tilde{\mu}$ and N^t on suitable sets and for a suitable t. The observations are independent according to Example 4.4 and we have an expression for the likelihood function. In case the expected number of observed points ($\Lambda(A)$ above) depends on the life time distribution f we often have problems interpreting the likelihood function. We condition on the observed number of individuals to obtain an alternative likelihood in this situation. Sections 4.4 and 4.5 illustrate an important point in this respect — whether to condition on the total number of individuals or to condition on the numbers of individuals from each of the two processes.

The process of birthtimes is Poisson with intensity $\varphi(t)$. The starting point is the Poisson process of deaths, $\tilde{\mu}$, with intensity $\tilde{\lambda}(t,a) = \varphi(t-a)f_{t-a}(a)$. We focus on the time homogeneous case where neither the process of birthtimes nor the lifetimes are allowed to depend on calendar time; $\varphi(t) = \varphi$ and $f_{t-a}(a) = f(a)$. Thus $\tilde{\lambda}(t,a) = \varphi f(a)$. Let M denote the expected lifetime $\int_0^\infty af(a)\,da = \int_0^\infty \bar{F}(a)\,da$. The assumption of time homogeneity is discussed in Section 5.2.

The following notation is used in the rest of Section 4. Let $(t_j, x_j)_{j\in J_1}$ be an observation of $\tilde{\mu}$ on a suitable set C depending on the sampling pattern. That is, (t_j, x_j) is the time and age of death for individual j, and we have observed $|J_1| = \tilde{\mu}(C)$ deaths. (The death times t_j should be distinguished from the sampling constants t_1 and t_2 without problems.) In a similar way, let $(a_j)_{j\in J_2}$ be an observation of N^t on a set C. That is, a_j is the age of an individual *alive* at time t and $|J_2| = N^t(C)$ is the number of individuals alive at time t with age in C. The individuals alive are censored at the observed age.

4.2 Cohort study and time window

We observe deaths in the Lexis diagram on the set $O_{\text{cs}} = \{(s,a) \in \mathbb{R}\times\mathbb{R}_0 | s-a \in]t_1,t_2]\}$ in the cohort study. The expected number of deaths is $(t_2-t_1)\varphi$, which equals the expected number of births in $]t_1,t_2]$. The likelihood is $L \propto \exp(-(t_2-t_1)\varphi)\varphi^{\tilde{\mu}(Ocs)}\prod_{j\in J_1} f(x_j)$. The birth intensity is estimated by $\hat{\varphi} = \tilde{\mu}(O_{\text{cs}})/(t_2-t_1)$ and the lifetime distribution f must be estimated by $\tilde{\mu}(O_{\text{cs}})$ independent observations from f.

In case the study is stopped at time t_3, the observation is composed of two independent observations. An observation of the deaths in the Lexis diagram on the set $O_{\text{cs}} \cap (]-\infty,t_3] \times \mathbb{R}_0)$ and an observation of the ages of individuals alive at time t_3. The latter is an observation of N^{t_3} on the set $[t_3-t_2, t_3-t_1[$. The expected number of points are $\int_{t_1}^{t_3}\int_{0\vee s-t_2}^{s-t_1}\varphi f(a)\,da\,ds$ and $\int_{t_3-t_2}^{t_3-t_1}\varphi\bar{F}(a)\,da$ respectively, and the likelihood function is thus proportional to
$\exp\left(-(t_2-t_1)\varphi\right)\varphi^{|J_1|+|J_2|}\prod_{j\in J_1} f(x_j)\prod_{j\in J_2}\bar{F}(a_j)$. We see, that the ages must be analysed as independent with ordinary right censoring. This is also the case when we condition on the *total* number of individuals observed.

In the time window we observe the deaths in the set $[t_1, t_2] \times \mathbb{R}_0$. This is, word for word, the same as the cohort study except that we do not have the notion of censoring here. The expected number of individuals is the same, the likelihoods are the same etc.

4.3 Cross sectional study

In a cross sectional study at time t we observe N^t on the whole axis $\mathbb{R}_0$. The expected number of individuals alive at time t is φM, and thus the likelihood function equals $L \propto \exp(-\varphi M)\varphi^{N^t(\mathbb{R}_0)} \prod_{j \in J_2} \bar{F}(a_j)$. Since the mean lifetime M depends on f, it is not obvious how to maximize the likelihood function. Conditioning on the observed number of individuals $N^t(\mathbb{R}_0) = n$, the observed ages are i.i.d. with density $\bar{F}(a)/M$. In epidemiology it has been known for a long time that the density of ages in a cross section in a "stable" population is proportional to the survival function, see for example Pressat (1995, p. 150). The density $\bar{F}(a)/M$ is the stationary recurrence distribution known from renewal theory and it reflects the fact that we observe the ages of individuals alive and thus are subject to interception bias. Feller (1971) page 369 (4.6), page 371 (4.16), and Problem 10 page 386 gives an overview of biased distributions in renewal theory, and Vardi (1988) advocates the terminology "interception bias" for the stationary recurrence time distribution. The nonparametric maximum likelihood estimator from n i.i.d. replications from the recurrence time distribution is known in the literature as the Grenander estimator, see Barlow *et al.* (1972, pp. 223), Denby & Vardi (1986), and Vardi (1989, Problem D).

4.4 Synthetic cohort

Observing deaths and individuals alive in a time window is equivalent to observing deaths in a time window and a cross section at time t_2. The two observations are independent. For the cross sectional study we know that it is most sensible to condition on the number of individuals alive at time t_2. By symmetry it seems natural to condition on the number of deaths observed as well. Conditional on $\tilde{\mu}([t_1, t_2] \times \mathbb{R}_0) = n_1$ and $N^{t_2}(\mathbb{R}_0) = n_2$ the likelihood equals $L \propto \prod_{j \in J_1} f(x_j) \prod_{j \in J_2} \bar{F}(a_j)/M$. The nonparametric maximum likelihood estimator is considered by Vardi (1982) and Soon & Woodroofe (1996). Conditional on the total number $\tilde{\mu}([t_1, t_2] \times \mathbb{R}_0) + N^{t_2}(\mathbb{R}_0)$ being equal to n, we obtain a likelihood with factors $f(x_j)/(M + (t_2 - t_1))$ and $\bar{F}(a_j)/(M + (t_2 - t_1))$. This likelihood has no nice interpretation and thus it does not seem reasonable to condition on the total number of individuals.

4.5 Follow-up on cross sectional study

We follow the individuals alive at time t_1. At time t_2 the observation is stopped and we have censoring. This situation is very much like the censoring in the cohort study in Section 4.2. Simon (1980) considers the follow up on a cross sectional study in a slightly different setup.

Assume that we follow individuals alive at time t_1 until death. That is, we observe the deaths $(t_j, x_j)_{j \in J_1}$ in the Lexis diagram on the set $O_{\text{fu}} = \{(s,a) \in \mathbb{R} \times \mathbb{R}_0 | s \geq t_1, a \geq s - t_1\}$. The expected number of individuals is the expected number of individuals alive at time t_1, i.e. φM. The likelihood function is $L \propto \exp(-\varphi M)\varphi^{\tilde{\mu}(O_{\text{fu}})} \prod_{j \in J_1} f(x_j)$. Conditioning on $\tilde{\mu}(O_{\text{fu}}) = n$, the points $(t_j, x_j)_{j \in J_1}$ are i.i.d. with density

$$\frac{f(x_j)}{M} 1\left((t_j, x_j) \in O_{\text{fu}}\right). \tag{2}$$

The distribution (2) is known from renewal theory as the simultaneous distribution of a forward and backward recurrence time. The point (t_j, x_j) is simply a bijective transformation of $(b_j, c_j) = (x_j - (t_j - t_1), t_j - t_1)$, the time lived before and after time t_1 respectively, so an interpretation of $x_j = b_j + c_j$ as the sum of the two recurrence times is natural. The marginal distribution of x_j has the length biased distribution with density $(xf(x))/M$. The likelihood function based on the marginal distribution of x depends on f in the same way as the simultaneous likelihood (2).

When stopping the observation at time t_2, we transform the points (t_j, x_j) with $t_j > t_2$ to $(t_2, x_j - (t_j - t_2))$. (This is a non-injective transformation of a Poisson process and we use the Theorem in the appendix. We get the same result if we consider the censoring in the same way as in Section 4.2.) A censored point contributes to the likelihood by a factor $\bar{F}(a)/M$ from the recurrence time distribution. This means that if we observe the deaths $(t_j, x_j)_{j \in J_1}$ and the ages $(a_j)_{j \in J_2}$ of individuals alive at time t_2, the likelihood is

$$L \propto \prod_{j \in J_1} \frac{f(x_j)}{M} \prod_{j \in J_2} \frac{\bar{F}(a_j)}{M} \tag{3}$$

conditional on the total number of individuals $|J_1| + |J_2| = n$. As noted the contributions from censored observations to the likelihood function is interception biased. This can be interpreted as a censoring of the forward time in the simultaneous distribution of backward and forward times (2). Note that right censoring in the marginal distribution of x_j gives a different result!

Winter & Földes (1988) find the nonparametric maximum likelihood estimator for the likelihood (3) in a renewal process setup and it can be re-derived in a counting process framework (Keiding & Gill, 1988, Section 7). The estimator is found in yet another setup by Vardi (1989, Problem B).

5 Discussion of some assumptions

In Section 5.1 we discuss the Poisson assumption from Section 4, and in Section 5.2 we consider the situation where the process of birthtimes is non time-homogeneous. Remember that conditional on the birthtimes for observed individuals the distribution of the process of birthtimes is irrelevant. The survey paper Keiding (1990) discusses estimation of the lifetime distribution conditional on the birthtimes when the lifetime distribution depends on calendar time as well as age. A simple example is when the hazard $\alpha(t,a)$ for $P_{t-a}(\cdot)$ at age a is piecewise constant. As in the ordinary

case with the age axis only, the maximum likelihood estimator of each of these constants is an "occurrence exposure rate". That is, the number of deaths divided by the total time under risk.

5.1 The Poisson assumption

The advantage of an assumption on the distribution of the process of birthtimes is that we can use information on survival until the start of the study. This is most predominant in the cross sectional study where such an assumption is mandatory. The Poisson process is a rough, but very useful approximation of real open populations. The most important aspect of the Poisson assumption is the independence between individuals. Note, that we most often condition on a suitable number of individuals before the lifetimes can be analysed as independent. Brillinger (1986), Keiding (1991), references therein, and the accompanying discussions have a lot of comments on the applicability of the Poisson assumptions in real life setups. The conclusion is that the Poisson approximation usually performs well. Another approach would be to model changes in a general population with the general framework in Section 6. The Poisson process assumption is the most simple and non-informative assumption we can make.

The Poisson assumption is also mandatory in for example the time window (Section 4.2) when the time of death t_j or the age at entry is unobserved. See also references on renewal processes given in the introduction.

When both the conditional approach and the Poisson approach are available, which one should we use? When M is small compared to $]t_1, t_2]$ the synthetic cohort and the time window have almost identical likelihoods as the fraction of individuals dying in the interval tends to one. Furthermore the fraction of individuals that are both born and die in $]t_1, t_2]$ tends to one. In this case, the likelihood functions conditional on the birthtimes and with the Poisson assumption is almost identical since the individuals that are both born and die in $]t_1, t_2]$ contribute by a factor $f(x)$ in both cases. The conclusion is that the Poisson assumption is useful when M is large compared to $]t_1, t_2]$, and in the cross sectional study and the follow-up on a cross sectional study. As will be seen in Example 6.2, the mean lifetime M could for example be the mean time spent in a disease state. I am not aware of any work comparing actual estimates and their variances conditional on the birthtimes and with the Poisson assumption.

5.2 Process of birthtimes depends on calendar time

Consider a non time-homogeneous Poisson process with intensity $\varphi(t)$ describing the births and the lifetimes being time-homogeneous; $f_s(a) = f(a)$. Assume that the birth intensity $\varphi(t)$ is *known*. In a cross sectional study we observe the $|J_2| = N^t(\mathbb{R}_0)$ ages $(a_j)_{j\in J_2}$ of individuals alive at time t. The expected number of individuals is $\int_0^\infty \varphi(t-a)\bar{F}(a)\,da$ and thus the likelihood is $L \propto \exp\left(-\int_0^\infty \varphi(t-a)\bar{F}(a)\,da\right) \prod_{j\in J_2} \varphi(t-a_j)\bar{F}(a_j)$. Conditional on $N^t(\mathbb{R}_0) = n$, the ages $(a_j)_{j\in J_2}$ are i.i.d. with density $\varphi(t-a)\bar{F}(a)/\int_0^\infty \varphi(t-a)\bar{F}(a)\,da$. In both cases φ appears as a

weight function for the observed ages. If for example the mean number of newborn is increasing, more weight is given to young individuals.

The lifetimes can be analysed by the age transformation $h(a) = \int_0^a \varphi(t-s)\,ds$. Let $\tilde{a}_j = h(a_j)$ and note that the transformed ages $(\tilde{a}_j)_{j\in J_2}$ are i.i.d. with density $\bar{F}(h^{-1}(\tilde{a}))/\tilde{M}$. Here $\tilde{M}$ is the mean lifetime in the distribution with survival function $\bar{F}\circ h^{-1}$. As in Section 4.3 we obtain an estimate for $\bar{F}\circ h^{-1}$, and h is known so we get an estimate of $\bar{F}$.

A similar age transformation can be used in the other sampling patterns, except the cohort study where no special tricks are needed. When φ is not known in advance one might use an estimate of φ. Keiding (1992, p. 319) gives the advice to avoid the reliance on the stationarity assumption when ever possible. The alternatives are to condition on the birthtimes or to use information on the non-homogeneity as demonstrated here.

6 Generalizations

In applications there would often be complications due to for example migration and population heterogeneity. This section extends the previous presented model to allow for covariates and more complicated life descriptions. The general setup gives a common framework to handle a multitude of complicated sampling schemes. We use the disability model as an example, and we focus on probabilistic statements about the disability model, whereas Keiding (1991) focuses on estimation.

6.1 Generalizations of the Lexis diagram

We extend the notation from Section 2.1 and restate some earlier results. Let $\eta = \sum_{i\in I} \varepsilon_{\sigma_i}$ be the point process of birthtimes $(\sigma_i)_{i\in I}$ on a measurable space $(E, \mathcal{E})$. For example, if we want the covariate gender in our model, we take E as $\mathbb{R}\times\{\text{male},\text{female}\}$. We call a description of a whole life Y and assume it has values in a space $(G, \mathcal{G})$. Such a description can be a stochastic process $(Y_a)_{a\geq 0}$ describing the status Y_a of the individual at age a. Marked point processes on the real line is one possible class of processes (Andersen *et al.*, 1993). They can be illustrated in the Lexis diagram by adding marks on the life lines at the time points of events. Let Y_i be the life description for the individual born at "time" σ_i and let $(Y_i)_{i\in I}$ be the collection of life descriptions. As in Definition 2.1 we define the *Lexis point process* μ as the point process $(\sigma_i, Y_i)_{i\in I}$ on $(E\times G, \mathcal{E}\otimes\mathcal{G})$, that is $\mu = \sum_{i\in I} \varepsilon_{(\sigma_i, Y_i)}$. Similar to Assumption 2.2 we assume that the life descriptions Y_i $(i \in I)$ are independent conditional on the process η of birthtimes $(\sigma_i)_{i\in I}$ and that the distribution of Y_i depends on η only through σ_i. Furthermore, as in Section 2.3, let P be a Markov kernel from $(E, \mathcal{E})$ to $(G, \mathcal{G})$ describing the distribution of Y_i given σ_i and use similar notation. Theorem 3.1 carries over without problems, and Theorem 4.1 is already formulated in the general context. As in Section 4.1 the intensity measure Λ for μ has intensity $\lambda(\sigma, y) = \varphi(\sigma)g(\sigma, y)$ w.r.t. $\pi_E \otimes \pi_G$, whenever the intensity measure Φ for the process of birthtimes η has density φ w.r.t. a measure π_E on E and the measures $P(\sigma, \cdot)$ have density $g(\sigma, y)$ w.r.t. a measure π_G on G.

6.2 Example: the disability model

The disability model in Fig. 2 has three states, H (healthy), I (invalid), and D (dead). Assume that the process of birthtimes is Poisson with intensity $\varphi(\sigma)$, the transition intensities from state H are $\alpha_H(t,a)$ and $\gamma(t,a)$, and that the death intensity $\alpha_I(t,a,d)$ from state I depends on calendar time $t = \sigma + a$, age a and duration d. No reactivation is allowed, that is, the transition $I \to H$ is not possible. The space G of life descriptions can be taken as all right continuous piecewise constant functions $(Y_a)_{a\geq 0}$ with state space $\tilde{G} = \{H, I, D\}$ and only finitely many jumps in finite time. Fig. 2 here

The transition intensities specify the Markov kernels $P_\sigma(\,\cdot\,)$ and for example

$$
\begin{aligned}
&P_\sigma(\{Y_0 = H\}) = 1,\\
&P_\sigma(\{Y_a = H\}) = \exp\left(-\int_0^a (\alpha_H + \gamma)(\sigma+u,u)\,du\right),\\
&P_\sigma(\{Y_a = I\}) = \int_0^a \left[e^{-\int_0^u (\alpha_H+\gamma)(\sigma+v,v)\,dv}\gamma(\sigma+u,u)e^{-\int_u^a \alpha_I(\sigma+v,v,v-u)\,dv}\right] du.
\end{aligned}
$$

Note that the processes Y are not Markov processes because the intensity from state I depends on how long the individual has been in state I.

For $Y \in G$ define T_1 as the age at the jump out of state H and T_2 as the age at the jump into state D. If the first jump is $H \to D$ then $T_2 = T_1$, else the first jump is $H \to I$ and T_2 is the age at the second jump $I \to D$. Let $\#$ denote "number of", and define for $t \in \mathbb{R}$, $A, B \in \mathcal{B}_0$ and $C \in \mathcal{B}$ the following quantities:

- $N^{t,H}(A) = \#$ individuals in state H at time t and age in A.
- $N^{t,I}(A \times B) = \#$ individuals in state I at time t, age in A and duration in B.
- $N^{t,H,D}(C \times A) = \#$ individuals in state D at time t, who made the transition $H \to D$ at calendar time in C and at age T_1 in A.
- $N^{t,I,D}(C \times A \times B) = \#$ individuals in state D at time t, who made the transition $I \to D$ at calendar time in C, at age T_2 in A and duration $T_2 - T_1$ in B.
- $N^{H,I}(C \times A) = \#$ individuals making the transition $H \to I$ at age in A and calendar time in C.

The processes $N^{t,H}$ and $N^{t,I}$ count the number of individuals in states H and I, whereas $N^{t,H,D}$ and $N^{t,I,D}$ describes the time point for the transition to state D. We now find the intensities for the N processes and some independence relations.

Theorem 6.1 *The processes $N^{t,H}(\cdot)$, $N^{t,I}(\cdot)$, $N^{t,H,D}(\cdot)$, $N^{t,I,D}(\cdot)$, and $N^{H,I}(\cdot)$ all have versions that are Poisson processes, and the versions of $N^{t,H}$, $N^{t,I}$, $N^{t,H,D}$ and $N^{t,I,D}$ are independent. The intensity processes are given by*

$$\lambda^{t,H}(a) = \varphi(t-a)\exp\left(-\int_0^a (\alpha_H + \gamma)(t-a+u,u)\,du\right),$$

$$\lambda^{t,I}(a,d) = 1\,(d \leqq a)\,\varphi(t-a)\exp\left(-\int_0^{a-d}(\alpha_H+\gamma)(t-a+u,u)\,du\right)$$
$$\times\,\gamma(t-d,a-d)\exp\left(-\int_{a-d}^{a}\alpha_I\,(t-a+u,u,u-(a-d))\,du\right),$$
$$\lambda^{t,H,D}(s,a) = 1\,(s \leqq t)\,\varphi(s-a)\exp\left(\int_0^{a}(\alpha_H+\gamma)(s-a+u,u)\,du\right)\alpha_H(s,a),$$
$$\lambda^{t,I,D}(s,a,d) = 1\,(s \leqq t, d \leqq a)\,\varphi(s-a)\exp\left(-\int_0^{a-d}(\alpha_H+\gamma)(s-a+u,u)\,du\right)$$
$$\times\,\gamma(s-d,a-d)\exp\left(-\int_{a-d}^{a}\alpha_I(s-a+u,u,u-(a-d))\,du\right)\alpha_I(s,a,d),$$
$$\lambda^{H,I}(t,a) = \varphi(t-a)\exp\left(-\int_0^{a}(\alpha_H+\gamma)(t-a+u,u)\,du\right)\gamma(t,a).$$

Proof The processes have versions that are Poisson processes according to the theorem in the appendix: The space $\mathbb{R}\times G$ is a Borel space because it can be written as a sequence space. Bounded sets in the arguments of the N processes relate to bounded sets for the process of birthtimes, and hence the expected values are finite. The lifetime distributions are continuous so singleton sets have zero measure. For $N^{t,H,D}$, $N^{t,I,D}$, and $N^{H,I}$, we do not have disjoint sets in $\mathbb{R}\times G$ for disjoint sets of the argument, but the intersections have intensity measure zero because each individual P_σ-a.s. makes the transitions 0 or 1 time. Similarly the processes $N^{t,H}$, $N^{t,I}$, $N^{t,H,D}$ and $N^{t,I,D}$ have independent versions according to the corollary in the appendix.

The calculation of the intensities is straightforward, using the definitions of the processes and the intensity measure (1). □

Theorem 6.1 states for example that the number of individuals in state H respectively I at time t are independent and both Poisson distributed. Conditional on $N^{t,H}(\mathbb{R}_0)+N^{t,I}(\mathbb{R}_0\times\mathbb{R}_0)=n$, the distribution on state, age, and duration for a random individual is given by

$$P(\text{individual in } H) = 1 - P(\text{individual in } I) = \frac{\mathrm{E}\,N^{t,H}(\mathbb{R}_0)}{\mathrm{E}\,N^{t,H}(\mathbb{R}_0)+\mathrm{E}\,N^{t,I}(\mathbb{R}_0\times\mathbb{R}_0)},$$
$$P(\text{age } \in da|\text{individual in } H) = \frac{\lambda^{t,H}(a)}{\mathrm{E}\,N^{t,H}(\mathbb{R}_0)}\,da,$$
$$P(\text{age } \in da, \text{duration } \in dv|\text{individual in } I) = \frac{\lambda^{t,I}(a,v)}{\mathrm{E}\,N^{t,I}(\mathbb{R}_0\times\mathbb{R}_0)}\,da\,dv.$$

Example 6.2 (Prevalent cohort study) The prevalent cohort study (Simon, 1980; Keiding, 1992, Sec. 7) is a sample of the individuals in the disease state I at time t_1, and at time t_2 we follow up on the sampled individuals. We are interested in the intensity α_I, that determines the duration of the stay in state I.

Simon (1980) assumes individuals enter state I as a time homogeneous Poisson process in calendar time and does not consider age at all. So "when living in a world of only diseased" we can apply the results in Section 4.5 to the duration in state I. We now comment on the situation where age is considered.

Keiding (1992) states in an informal way that conditional on suitable times the hazard for the age at death is α_I. We now state a similar result and assume all three variables time, age, and duration known. Define $(t_e, a_e) = (\sigma + T_1, T_1)$, the calendar time and age at entry to state I, and the duration $d = T_2 - T_1$ of the stay in state I. Note that the Lexis point process is a bijective transformation of $((\sigma_i + T_{1,i}, T_{1,i}), (T_{2,i} - T_{1,i}))_{i \in I}$. Note furthermore that $d \geq 0$, and that $d > 0$ if and only if the transition $H \to I$ is made. Conditional on $(t_{e,i}, a_{e,i})_{i \in I}$ the durations $(d_i)_{i \in I}$ are independent since we condition on the birthtimes $\sigma_i = t_{e,i} - a_{e,i}$, and the duration d has distribution:

$$P(d > 0) = 1 - P(d = 0) = \frac{\gamma(t_e, a_e)}{(\alpha_H + \gamma)(t_e, a_e)},$$
$$\text{hazard for } d \,|\, d > 0: \quad \alpha_I(t_e + d, a_e + d, d) = \alpha_{I, t_e, a_e}(d).$$

We now consider the pair (t_e, a_e) as the birthtime and d as the description of life. The arguments just given show that Assumption 2.2 is fulfilled. Theorem 3.1, with $h_{(t_e, a_e)}$ as the indicator function for individuals in state I at time t_1, states that the durations must be analysed as independent and left truncated at the duration at time point t_1 conditional on (t_e, a_e). But left truncation preserves the hazard $\alpha_{I, t_e, a_e}(d)$ and we have independent delayed entry (Keiding, 1992).

Information on the duration up to the sampling point t_1 can be used if we assume that the process of birthtimes σ_i is a Poisson process. Assume that all intensities are independent of calendar time t, and that the death intensity $\alpha_I(t, a, d) = \alpha_I(d)$ for diseased individuals only depends on duration d. This assures that individuals enter state I as a homogeneous Poisson process in calendar time t, even though the intensity $\gamma(a)$ is allowed to depend on age a. This is the case of Simon (1980) and Section 4.5. We can easily show that conditional on the number of sampled individuals, the joint distribution of age a_e at entry to state I and total duration d has a density proportional to

$$\frac{\exp\left(-\int_0^{a_e} (\alpha_H + \gamma)(u)\, du\right) \gamma(a_e)}{\int_0^\infty \exp\left(-\int_0^a (\alpha_H + \gamma)(u)\, du\right) \gamma(a)\, da} \frac{f_I(d)}{M_I}.$$

Here f_I denotes the density with hazard $\alpha_I(d)$, and M_I the mean time spent in state I. The duration d must be analysed as in Section 4.5 and independent of the age at entry to state I. □

Acknowledgments

I am very indebted to Martin Jacobsen, Helle Sørensen, Niels Keiding, and Torben Martinussen for their many helpful comments. Also detailed comments from the referees and an associate editor improved the presentation. Part of this work was done while I visited the department of mathematical statistics, Chalmers University of Technology, Göteborg, Sweden, with support from NorFA.

Author's present address: Jens Lund, The Royal Veterinary and Agricultural University, Department of Mathematics and Physics, Thorvaldsensvej 40, DK – 1871 Frederiksberg C, Denmark, e-mail: `jlund@dina.kvl.dk`.

References

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993). *Statistical models based on counting processes*. Springer-Verlag, New York.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. & Brunk, H. D. (1972). *Statistical inference under order restrictions*. Wiley, London.

Brillinger, D. R. (1986). The natural variability of vital rates and associated statistics (with discussion). *Biometrics* **42**, 693–734.

Daley, D. J. & Vere-Jones, D. (1988). *An introduction to the theory of point processes*. Springer-Verlag, New York.

Denby, L. & Vardi, Y. (1986). The survival curve with decreasing density. *Technometrics* **28**(4), 359–367.

Feller, W. (1971). *An introduction to probability theory and its applications*. Vol. II. Second edn. Wiley, New York.

Hoffmann-Jørgensen, J. (1994). *Probability with a view toward statistics*. Vol. II. Chapman & Hall, New York.

Jacobsen, M. (1982). *Statistical analysis of counting processes*. Vol. 12 of *Lecture Notes in Statist.*. Springer-Verlag, New York.

Karr, A. F. (1991). *Point processes and their statistical inference*. Probab.: Pure and Appl.. Second edn. Marcel Dekker, New York.

Keiding, N. (1990). Statistical inference in the Lexis diagram. *Philos. Trans. Roy. Soc. London Ser. A* **332**, 487–509.

Keiding, N. (1991). Age-specific incidence and prevalence: a statistical perspective (with discussion). *J. Roy. Statist. Soc. Ser. A* **154**(3), 371–412.

Keiding, N. (1992). Independent delayed entry. *In* J. P. Klein & P. K. Goel, eds, Survival Analysis: State of the Art. Vol. 211 of *NATO ASI Series, Series E: Applied Sciences*. Kluwer Acad. Publ.. pp. 309–326.

Keiding, N. & Gill, R. (1988). Random truncation models and Markov processes. Report MS-R8817. Centrum voor Wiskunde en Informatica. Department of Mathematical Statistics.

Lexis, W. (1875). Einleitung in die theorie der bevölkerungs-statistik. *In* D. Smith & N. Keyfitz, eds, Mathematical Demography (selected papers) (1977). Vol. 6 of *Biomathematics*. Springer-Verlag. Berlin. pp. 39–41.

McClean, S. & Devine, C. (1995). A nonparametric maximum likelihood estimator for incomplete renewal data. *Biometrika* **82**(4), 791–803.

Pressat, R. (1995). *Eléments de démographie mathématique.* Association Internationale des Demographes de Langue Française, Paris.

Simon, R. (1980). Length biased sampling in etiologic studies. *Amer. J. Epidem.* **111**(4), 444–452.

Soon, G. & Woodroofe, M. (1996). Nonparametric estimation and consistency for renewal processes. *J. Statist. Plann. Inference* **53**, 171–195.

Stoyan, D., Kendall, W. S. & Mecke, J. (1995). *Stochastic geometry and its applications.* Second edn. Wiley, Chichester.

Turnbull, B. W. (1976). The emperical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser. B* **38**, 290–295.

Vardi, Y. (1982). Nonparametric estimation in renewal processes. *Ann. Statist.* **10**(3), 772–785.

Vardi, Y. (1988). Statistical models for intercepted data. *J. Amer. Statist. Assoc., Theory and Methods* **83**(401), 183–197.

Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika* **76**(4), 751–761.

Winter, B. B. & Földes, A. (1988). A product-limit estimator for use with length-biased data. *Canad. J. Statist.* **16**(4), 337–355.

Appendix: non-injective transformation of Poisson processes

This appendix describes how we can transform a Poisson process by a class of non-injective transformations and still get a Poisson process. This extends well-known results on transformation of Poisson processes by injective maps $h : (S, \mathbb{S}) \to (\tilde{S}, \tilde{\mathbb{S}})$ corresponding to $K = h^{-1}$ below.

When $(\tilde{S}, \tilde{\mathbb{S}})$ is a measurable space we equip the function space $\overline{\mathbb{R}}_0^{\tilde{\mathbb{S}}}$ of functions $\mu : \tilde{\mathbb{S}} \to \overline{\mathbb{R}}_0$ with the σ-algebra $\sigma(\overline{\mathbb{R}}_0^{\tilde{\mathbb{S}}})$ induced by the projections $\mu \mapsto \mu(A)$ for all $A \in \tilde{\mathbb{S}}$. Two random variables μ, $\tilde{\mu}$ defined on $(\Omega, \mathcal{F}, P)$, taking values in $(\overline{\mathbb{R}}_0^{\tilde{\mathbb{S}}}, \sigma(\overline{\mathbb{R}}_0^{\tilde{\mathbb{S}}}))$ are versions of each other if $P(\mu(A) = \tilde{\mu}(A)) = 1$ for all $A \in \tilde{\mathbb{S}}$.

To talk about bounded set $A \in \tilde{\mathbb{S}}$ we need some structure on $\tilde{S}$. We could for example require $\tilde{S}$ to be a complete separable metric space (Daley & Vere-Jones, 1988) or a product of such a space and a space without any special structure. In the last case a bounded set is (a subset of) a bounded set $\times$ the space without structure.

Theorem *Let $(S, \mathbb{S})$ and $(\tilde{S}, \tilde{\mathbb{S}})$ be measurable spaces, let K be a mapping from $\tilde{\mathbb{S}}$ to $\mathbb{S}$, and let μ be a Poisson process on $(S, \mathbb{S})$ with intensity measure Λ. Assume that all singleton sets in $\tilde{S}$ are measurable. Then $\tilde{\mu} = \mu \circ K$ is a random variable taking values in $(\overline{\mathbb{R}}_0^{\tilde{\mathbb{S}}}, \sigma(\overline{\mathbb{R}}_0^{\tilde{\mathbb{S}}}))$. If*

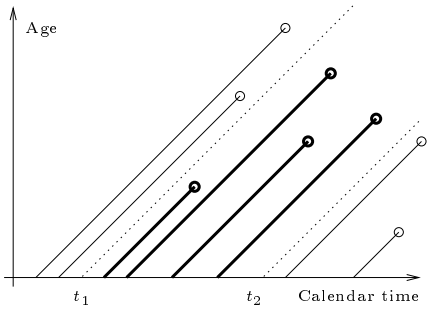
- *S is a Borel space.*

- $K(\emptyset) = \emptyset$.
- *K is countably set additive; that is for $A_1, A_2, \cdots \in \tilde{\mathbb{S}}$ we have* $K(A_1 \cup A_2 \cup \cdots) = K(A_1) \cup K(A_2) \cup \cdots$.
- *For $a \in \tilde{S}$ we have* $\Lambda(K(\{a\})) = 0$.
- *For $A \in \tilde{\mathbb{S}}$ bounded we have* $\Lambda(K(A)) < \infty$.
- *For $A_1, A_2 \in \tilde{\mathbb{S}}$ and $A_1 \cap A_2 = \emptyset$ we have* $\Lambda(K(A_1) \cap K(A_1)) = 0$.

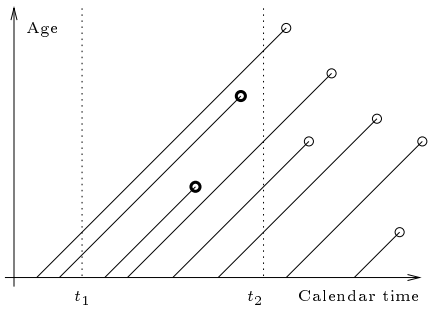
then $\tilde{\mu}$ has a version that is a Poisson process on $(\tilde{S}, \tilde{\mathbb{S}})$ with intensity measure $\tilde{\Lambda} = \Lambda \circ K$.

Corollary *Let $(S, \mathbb{S})$ and $(\tilde{S}_j, \tilde{\mathbb{S}}_j)$ where $j = 1, \ldots, n$, be measurable spaces, let K_j be a mapping from $\tilde{\mathbb{S}}_j$ to $\mathbb{S}$ for $j = 1, \ldots, n$, and let μ be a Poisson process on $(S, \mathbb{S})$ with intensity measure Λ. Assume that all singleton sets in $\tilde{S}_j$ are measurable, that S is a Borel space and that the mappings K_j, $j = 1, \ldots, n$, fulfill the conditions of the above theorem. If all the intersections $K_{j_1}(\tilde{S}_{j_1}) \cap K_{j_2}(\tilde{S}_{j_2})$ for $j_1 \neq j_2$, have Λ-measure 0 the mappings $\tilde{\mu}_j = \mu \circ K_j$, $j = 1, \ldots, n$, have versions that are independent Poisson processes.*

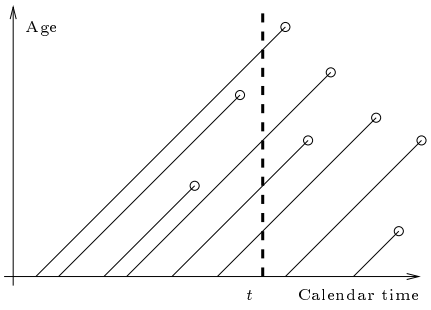
We omit the proofs. The hard part of the proof of the theorem is to show that $\tilde{\mu}$ is a counting measure; it is easy to show σ-additivity almost surely for any fixed sequence of sets $A_1, A_2, \ldots$, but the exception set depends on the sequence of sets $A_1, A_2, \ldots$. For similar proofs see, e.g., Hoffmann-Jørgensen (1994, Sec. 10.29) for existence of regular conditional distributions.



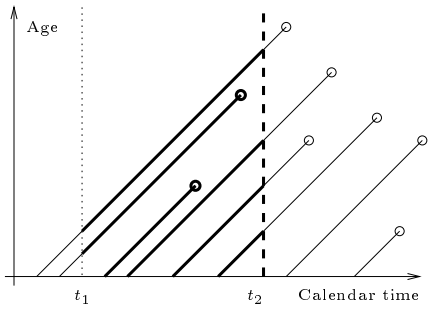
Cohort study Individuals born in $]t_1, t_2]$ are observed.



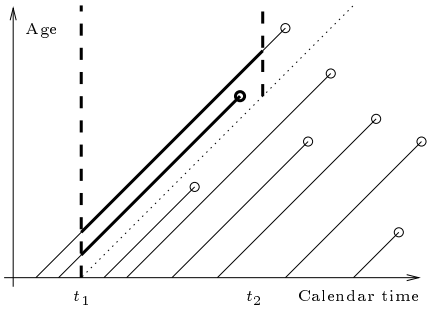
Time window All deaths in the time interval $[t_1, t_2]$ are observed.



Cross sectional study Ages of individuals alive at time t are observed.



Synthetic cohort All deaths and individuals alive in the time interval $[t_1, t_2]$ are observed.



Follow-up on cross sectional study At time t_2 we follow up on individuals alive at time t_1.

Figure 1: Sampling in the Lexis diagram. A life line with slope 1 represents an individual. A circle denotes a death. Bold lines indicate what we observe.

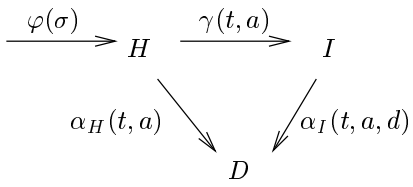


Figure 2: The disability model. The intensities describe the movements between the three states H (healthy), I (invalid) and D (dead) as a function of the birthtime σ, age a, duration d, and calendar time $t = \sigma + a$.