Sampling Bias in Survival Analysis — How to Use the Lexis Diagram

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

Preprint no. 4, 1998,
Department of Theoretical Statistics, University of Copenhagen

April 5, 1998

Abstract

In survival analysis, the data are often not analysed in the life time distribution. Because of the way individuals are selected for the study, the observed data must be analysed in a modified version of the lifetime distribution. We can use information on survival to the start of the study when we assume that the process of birthtimes is Poisson. In this case the relevant modifications are length bias and the stationary recurrence time distribution known from renewal theory and epidemiology. On the other hand, we can condition on the birthtimes for the individuals we observe. Then the analysis does not depend on the distribution of the process of birthtimes, and we cannot use information on survival to the start of the study. In this case the relevant modifications are left truncation, and right truncation or censoring.

The Lexis diagram is a coordinate system with calendar time in the horizontal direction and age of individuals in the vertical direction, and it is useful as a way to focus thinking of sampling patterns. We use a rigorous point process description of the Lexis diagram in order to make the preconditions and assumptions transparent. Six sampling patterns are systematically investigated and we find the likelihood functions. It is necessary to have the assumption that lifetimes are independent given all the birthtimes. We use either a Poisson model for the process of birthtimes or we condition on the birthtimes for the individuals we observe. The results in the Poisson setup is usually conditional on the number of individuals observed or the number of individuals alive and dead, respectively. The results and techniques are extended to more complicated situations than just lifetimes, and the disability model and the prevalent cohort study are used as examples.

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

Contents

1 Introduction

This paper introduces a rigorous framework for sampling of individuals in survival analysis. The Lexis diagram as reviewed in [Kei90] is used as the conceptual idea.

The Lexis diagram is just a way to depict the life of individuals. It is a coordinate system with calendar time in the horizontal direction and age of individuals in the vertical direction. This is illustrated in Figure 1. Mathematically we describe the Lexis diagram as a point process composed from the process of birthtimes and the lifetimes. The idea of a point process description is introduced in Brillinger [Bri86], and this paper is a systematic use of this idea. A basic assumption is that the lifetimes are independent given all the birthtimes.

We are interested in estimation of the distribution of the survival times, and for the six sampling patterns outlined in Figure 2 we find the likelihood function. The focus is on non parametric estimation, but since we find the likelihood function parametric models could be used as well. It is of extreme importance to carefully consider the way individuals are chosen for the study, since this selection often introduces biased or otherwise modified lifetime distributions. We consider the six sampling patterns in two different setups.

In the first setup we condition on the birthtimes for the individuals we observe. First of all this analysis does not depend on the distribution of the birthtimes. Furthermore this conditioning leads to left truncation and right censoring of the lifetimes when we observe deaths as well as individuals alive. We can e.g. use the Kaplan-Meier estimator for the survival function and the counting process techniques in [ABGK93].

In the second setup we assume that the process of birthtimes is a homogeneous Poisson process. This idea is first used in [Sim80] where Simon assumes that the times of diagnosis are described by a Poisson process. Brillinger [Bri86] emphasizes the Poisson assumption and discusses the applicability of the Poisson process. One of the main results in [Bri86] is that the points of death in the Lexis diagram is a Poisson process in the plane; this is my Example 4.2.

When we use this Poisson assumption we can use information on survival until the start of the study, and the lifetimes must be analysed with various forms of length bias. This is the only option when analyzing cross sectional data. These length biased distributions are the same as seen in renewal theory and epidemiology. Often we condition on a suitable number in the Poisson setup to obtain the desired likelihood functions, where the lifetimes are analysed as independent. The number we condition on is chosen in a way, so that it has a "nice" expected value.

We avoid a lot of asymptotic arguments compared to e.g. [Sim80] by letting the process of birthtimes be time homogeneous on the whole axis. This means that we have no limits of the type "in the distant past" or "as t tends to ∞".

We extend the general setup to more complicated situations than just the life time. The disability model is a special case of this generalization, and with the assumption that individuals are born as a Poisson process, we give some examples of the possible distributional results we can obtain, e.g. for the prevalent cohort study in Example 6.3. Keiding [Kei91] uses the Poisson assumption in the disability model too, but he focusses on the relation prevalence = incidence $\times$ duration and estimation of the intensity for the transition healthy $\rightarrow$ disabled.

The outline for the rest of this paper is as follows. Section 2 introduces notation, the Lexis diagram, the six basic sampling patterns considered, and the basic Assumption 2.2. Sections 3 and 4 systematically investigate the sampling patterns. In Section 3 we condition on the birthtimes for the individuals we observe, and in Section 4 we use the Poisson assumption. Section 5 discusses some of the assumptions made; mainly the Poisson assumption. The setup is generalized in Section 6, and the disability model is used as an example of a more complicated model. Finally some points are discussed in Section 7 and some open problems are outlined in Section 8.

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. Figure 1 is a sketch of a Lexis diagram. 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 death.

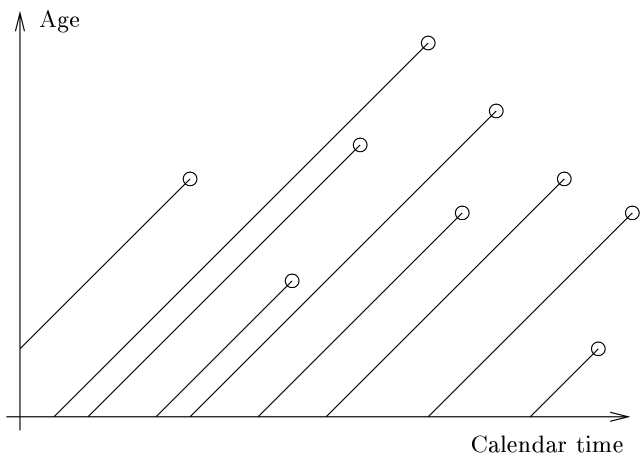


Figure 1: A Lexis diagram. A life line with slope 1 represents an individual. A circle denotes a death.

Let $(\sigma_i)_{i\in I}$ denote the collection of all birthtimes. We assume for convenience 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 will think of $(\sigma_i)_{i\in I}$ as a point process on $\mathbb{R}$ and denote by η the associated counting measure. 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. The index set I can e.g. 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 call the point process $(\sigma_i)_{i\in I}$ and the equivalent η the *process of birthtimes*. General point processes are treated in [DVJ88], [Kar91], and [SKM95].

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

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 $\mu = \sum_{i\in I} \varepsilon_{(\sigma_i, X_i)}$.

Note that μ is indeed a point process whenever η is a point process. 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 $(\sigma_i + X_i, X_i)_{i\in I}$ on $(\mathbb{R}\times\mathbb{R}_0, \mathcal{B}\otimes\mathcal{B}_0)$. We do not use this as our definition of the Lexis point process, since it is not appropriate in the general setup in Section 6. Also note that my Lexis point process corresponds to Lexis' original definition of the Lexis diagram in 1875 in the paper [Lex75].

2.2 Summary of Sampling Patterns

In this section we review some natural ways to sample individuals in the Lexis diagram. Given time points t_1, t_2, and t, where $t_1 < t_2$, define subsets A and B_t of the Lexis diagram by $A = [t_1, t_2] \times \mathbb{R}_0$ and $B_t = \{(s, a) \in \mathbb{R}\times\mathbb{R}_0 | s \geq t, a \geq s - t\}$. The set B_t is a triangle that contains the deaths of the individuals alive at calendar time t. The sampling patterns are illustrated in Figure 2.

Cohort study We observe all individuals born in $]t_1, t_2]$, i.e. we see the Lexis diagram on the set $O_{\text{cs}} = (B_{t_2} \cup A) \setminus B_{t_1} = \{(s, a) \in \mathbb{R}\times\mathbb{R}_0 | s - a \in]t_1, t_2]\}$. Observation might be stopped at a time t_3 after t_2. In this case individuals still alive are only observed to live longer than their age at time t_3. The actual lifetime is not observed.

Time window All deaths (times and ages) in the set $O_{\text{tw}} = A$ are observed. An observation of this type might e.g. arise in a study based on data from a register of deaths.

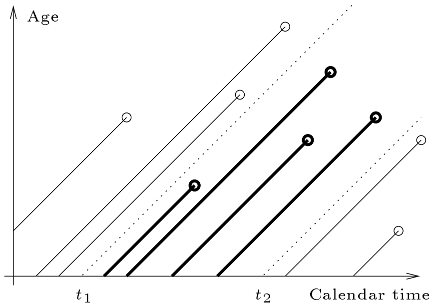
Cross sectional study The ages of individuals alive at time t are recorded.

Synthetic cohort All deaths and individuals alive during the time period $[t_1, t_2]$ are observed. Define $O_{\text{sc}} = A$. Especially we observe the current age for individuals alive at time t_1. 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; see Example 5.1.

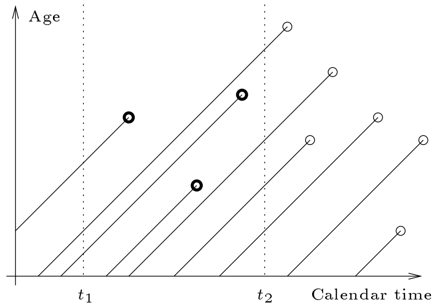
Follow up on cross sectional study At time t_2 we observe whether individuals alive at time t_1 are dead or not. If alive we observe the age, and if dead we observe the age and time of death. We observe deaths and individuals alive in $O_{\text{fcs}} = B_{t_1} \cap A$.

Current status data Every birthtime before time t is recorded and at time t it is observed whether individuals are alive or not. In this case we have extreme censoring; for individual i born at time σ_i it is only observed whether the lifetime is smaller or larger than $t - \sigma_i$. A recent reference on current status data with a lot of further references is [KBSH96]. I slightly misuse the term "current status data" since it is most often used in the context of the disability model in Section 6.2, where it is checked at a time point t whether an individual is in the disease state or not. A natural variation of current status data is that we only have data for individuals born between a time point $\tilde{t}$, where $\tilde{t} < t$, and t.

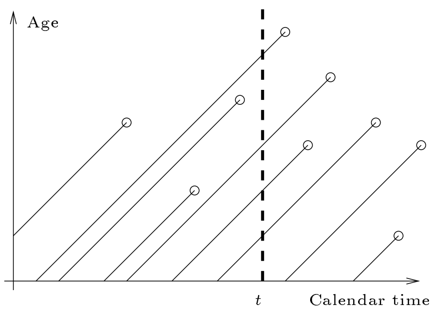
These are the six basic sampling patterns we will focus on. They can all be thought of as composed of three basic sets in the Lexis diagram from the context of epidemiology. These sets are e.g. discussed in [Kei90]. Several variations exist and some of these are discussed in Section 7.



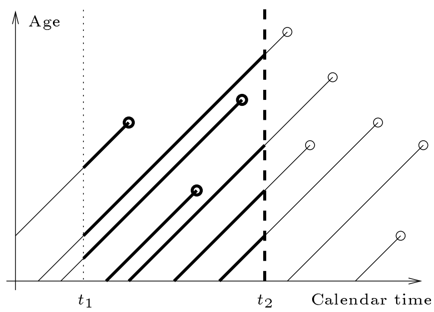
Cohort study Individuals born in $]t_1, t_2]$ are observed.



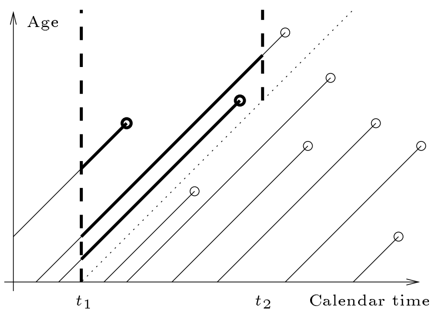
Time window All deaths in the time interval $[t_1, t_2]$ are observed.



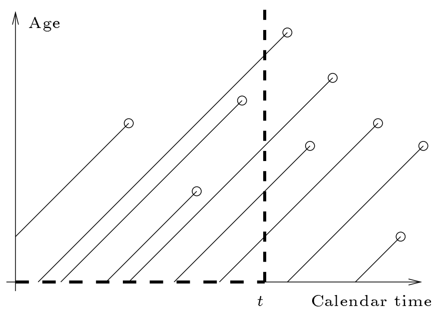
Cross sectional study Ages of individuals alive at time t are observed.



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.



Current status data Every birthtime before time t is known and at time t it is observed whether individuals are alive or not.

Figure 2: Sampling in the Lexis diagram. Bold lines indicate what we observe.

2.3 The Basic Assumption

In this section we discus the basic probabilistic structure of our model, that will be in force throughout the paper. We need some sort of independence assumption and we will introduce some further notation.

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.

Example 2.3 The lifetimes need not be independent in the marginal distribution. Take $I = \{1, 2\}$ and let U be a positive non degenerate random variable. Define $\sigma_1 = 1 + U$, $\sigma_2 = 2 + U$, and $X_1 = X_2 = U$. In the marginal distribution $X_1 = X_2$ almost surely, and conditional on (σ_1, σ_2) the two lifetimes are also degenerate, and thus independent. □

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 mapping 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 will 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. For convenience we will focus on the situation where the lifetime distribution has a density w.r.t. the Lebesgue measure. The results can be extended to the case with arbitrary lifetime distributions. Let f_σ be the density for P_σ w.r.t. the Lebesgue measure l_0 on $\mathbb{R}_0$.

By $\alpha(t, a)$ we denote the hazard $f_{t-a}(a)/\bar{F}_{t-a}(a)$ for the distribution P_{t-a} at age a. The interpretation of $\alpha(t, a)$ is local at the point (t, a) in the Lexis diagram; given an individual born at time $t - a$ has survived to age a, the probability of dying during the next h time units is $\alpha(t, a)h$. That is $P_{t-a}(a \le X < a + h | X \ge a) \approx \alpha(t, a)h$ for small h.

3 Analysis Conditional on Selected Birthtimes

In this section we will condition on the birthtimes for the individuals we observe.

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

The following theorem states that given the birthtimes for individuals we observe, the lifetimes X_i for these individuals are independent. Furthermore the theorem states that these lifetimes must be analysed conditional on that we observe the individual, and their distribution do not depend on the distribution of the process of birthtimes.

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

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 $X_i \sim P_{\sigma_i}(\,\cdot \mid h_{\sigma_i}(X_i))$. To see this let $\tilde{I}$ be any finite subset of I and let $\otimes$ denote product measure. We have that

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

The further conditioning on $(h_{\sigma_i}(X_i))_{i\in I}$ has a product structure, and thus by the principle of repeated conditioning, see e.g. [Jac82, App. 1], we get that the independence is preserved and

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

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. □

In the following O is a deterministic measurable subset of the Lexis diagram $\mathbb{R} \times \mathbb{R}_0$. We assume that all possible life lines intersect O at most once, and that the expected number of life lines intersecting O is finite. 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 $a_{\text{in},\sigma} = a_{\text{out},\sigma} = \infty$. See Figure 3.

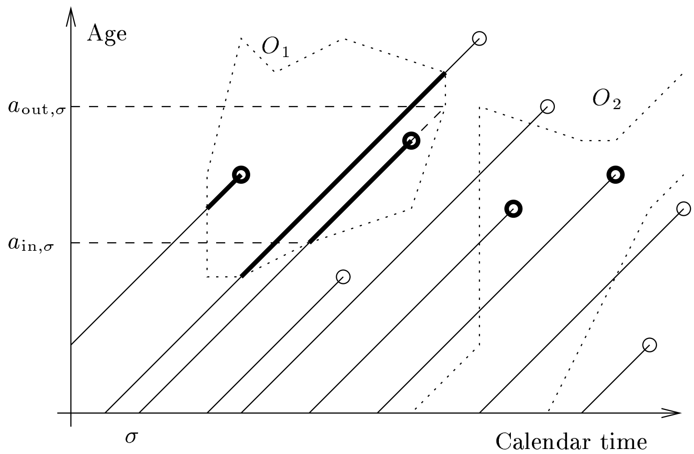


Figure 3: An example of the ages $a_{\text{in},\sigma}$ and $a_{\text{out},\sigma}$ is given. The set O_1 illustrates that we observe all life lines intersecting with O_1. The set O_2 illustrates that we observe deaths in the set O_2.

That each life line intersects O at most once is a condition on convexity within O for lines with slope 1. This means that whenever two points on a line with slope 1 both lie in O, then the line segment between the points lie in O. In case 2 below it will be enough that the expected number of deaths in O is finite. This is a weaker assumption than the expected number of life lines intersecting O being finite, but it is not a natural assumption in applications. The assumption that the expected number of life lines intersecting O being finite is really more a restriction on the parameter space, i.e. the Markov kernel P, than on the set O. The assumption ensures that we only observe finitely many individuals. We discuss this in more depth when considering current status data below.

The ages $a_{\mathrm{in},\sigma}$ and $a_{\mathrm{out},\sigma}$ are deterministic conditionally on σ, and thus the truncations and censoring in the following are independent. Truncation is when we make inference in a conditional distribution, and e.g. 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 is when we do only observe the lifetime to be larger than a certain age, and do not observe the lifetime itself. See [ABGK93, Sec. III.2, III.3] for a discussion of independent truncation and censoring.

We consider two situations that are illustrated in Figure 3.

1. We observe life lines intersecting with O. That means we see both individuals alive and those that die. Observing a life line going through O means observing the time and age when entering O, and the fact that the individual does not die while under observation. If the individual dies while under observation, i.e. the life line ends in O, we record the time and age of death. Let $(\sigma_j)_{j\in J}$ denote the $|J|$ birthtimes for individuals with life lines intersecting O. We use Theorem 3.1 with $h_\sigma(x)$ as the indicator function for the event $(x \geq a_{\mathrm{in},\sigma})$, and obtain the result that conditional on the birthtimes, the observed lifetimes are independent and X_j for $j \in J$ 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 $a_{\mathrm{out},\sigma_j}$ as well. 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. These results can e.g. be derived in the framework of counting processes, see e.g. [ABGK93].
2. We observe points of death in the set O. We use Theorem 3.1 with $h_\sigma(x)$ as the indicator function for the event $(a_{\mathrm{in},\sigma} \leq x < a_{\mathrm{out},\sigma})$. The theorem states that conditional on the birthtimes $\sigma_j = t_j - x_j$ for the $|J|$ deaths observed at $(t_j, x_j)_{j\in J}$, the lifetimes X_j for $j \in J$ are independent and must be analysed left truncated at age a_{in,σ_j} and right truncated at age $a_{\mathrm{out},\sigma_j}$. The nonparametric maximum likelihood estimator is a special case of the Turnbull estimator [Tur76], which in turn is an example of the EM algorithm.

We can use the above on four of the six sampling patterns:

Cohort study is both case 1 and 2 with the set $O = O_{\mathrm{cs}}$ because $a_{\mathrm{out},\sigma}$ equals ∞ for the relevant birthtimes. Notice that $a_{\mathrm{in},\sigma} = 0$ for individuals we observe. This means that we analyse the lifetimes X_j from the distribution f_{σ_j}. When we stop observing at time t_3 we are in case 1 with a slightly modified set O, and we analyse the lifetimes right censored at ages $a_{\mathrm{out},\sigma_j}$.

Time window is case 2 with $O = O_{\mathrm{tw}}$. The lifetimes are left and right truncated at age a_{in,σ_j} and $a_{\mathrm{out},\sigma_j}$ respectively.

Synthetic cohort is case 1 with $O = O_{\mathrm{sc}}$. The lifetimes are left truncated and right censored at age a_{in,σ_j} and $a_{\mathrm{out},\sigma_j}$ respectively.

Follow up on cross sectional study is case 1 with $O = O_{\mathrm{fcs}}$. The lifetimes are left truncated and right censored.

Some further work is necessary for the two remaining sampling patterns:

Cross sectional study We can say nothing when we have cross sectional data and condition on the birthtimes for the individuals we observe. Let $(a_j)_{j\in J}$ denote the $|J|$ ages of individuals alive at time t. Theorem 3.1 with $h_\sigma(x)$ as the indicator function for $((\sigma+x,x)\in B_t)$, or equivalent $(\sigma\le t, x\ge t-\sigma)$, states that conditional on the birthtimes $\sigma_j = t-a_j$ the lifetimes X_j are independent and must be analysed in the conditional distribution given $X_j \ge a_j$. Notice that $a_j = a_{\text{in},\sigma_j}$ if we define $O = B_t$. Unfortunately this is everything we observe about X_j, and thus the likelihood function equals 1. To say something from cross sectional data we need a model for the process of birthtimes η; see Section 4.

Current status data Enumerate the birthtimes as in Section 2.1 with the modification that σ_0 is the first birthtime after time t, so that $\cdots\sigma_{-1} < t \le \sigma_0 \cdots$ and $\cdots < \sigma_i < \sigma_{i+1} < \cdots$. Theorem 3.1 with $h_\sigma(x)$ as the indicator function for the event $(\sigma \le t)$ states that conditional on σ_j for $j<0$, the lifetimes X_j with $j<0$ are independent and $X_j \sim F_{\sigma_j}$. The likelihood for observation of survival status for individual j is

$$\bar{F}_{\sigma_j}(t-\sigma_j)^{1(X_j>t-\sigma_j)}(1-\bar{F}_{\sigma_j}(t-\sigma_j))^{1(X_j\le t-\sigma_j)}.$$

We have a likelihood for the observation if the product

$$\prod_{j:\sigma_j\in[s,t]} \bar{F}_{\sigma_j}(t-\sigma_j)^{1(X_j>t-\sigma_j)}(1-\bar{F}_{\sigma_j}(t-\sigma_j))^{1(X_j\le t-\sigma_j)} \tag{1}$$

with positive probability converges toward a positive number as s tends to $-\infty$.

It is sufficient for (1) to be convergent with a positive limit that the expected number of individuals alive at time t is finite. Conditional on σ_j for $j<0$ the expected number of individuals alive at time t is

$$\mathrm{E}\sum_{j=-1}^{-\infty} 1\,(X_j > t-\sigma_j) = \sum_{j=-1}^{-\infty} \bar{F}_{\sigma_j}(t-\sigma_j).$$

If this sum is finite, we will eventually only get factors of the form $1-\bar{F}_{\sigma_j}(t-\sigma_j)$ as s tends to $-\infty$. The product of these factors is convergent exactly when the sum is finite[1].

If we only observe the status for individuals born in $[\tilde{t}, t]$ with $\tilde{t} < t$ we just have a finite product in (1) and there is no problem.

The nonparametric maximum likelihood estimator is e.g. given in [KBSH96] which has a lot of further references. In [Kei91, pp. 386] there is an exposition. The estimator can be seen as a special case of the Turnbull estimator [Tur76].

In the sampling patterns above (except the cohort study), it is sufficient and necessary for the expected number of life lines intersecting O to be finite, that the expected number of individuals alive at time t_1 is finite.

When the lifetimes have an upper limit C, then $\bar{F}_\sigma(x) = 0$ for $x > C$. The expected number of individuals alive at time t is obviously finite in this situation. This condition

[1] Let $p_i \in]0,1[$ for $i\in\mathbb{N}$. Then $\prod_{i=1}^{\infty}(1-p_i)\in]0,1[\Leftrightarrow \sum_{i=1}^{\infty} p_i < \infty$.

is of course not satisfied for e.g. exponentially distributed lifetimes, although it seems to be fulfilled when talking about the lifetimes of people.

In the time homogeneous case where $\bar{F}_\sigma$ does not depend on σ and the expected lifetime is finite, the condition $\sum_{j=-1}^{-\infty} \bar{F}(t-\sigma_j) < \infty$ is a condition on the process of birthtimes η — the number of individuals born as s tends to $-\infty$ are not allowed to grow too fast. We make a few more comments on this in the Poisson case on page 13.

4 A Poisson Model for the Process of Birthtimes

When we have data from a cross sectional study we cannot use the approach in Section 3. We condition on too much information and have nothing left to say about the observed ages. In this situation we need a model for the birthtimes. This will utilize information on survival until the start of the study and we will see various forms of length bias instead of left truncation. We will most often condition on the number of individuals alive, dead, or the total number. A model for the birthtimes will also be useful if we do not know the ages at entrance to the study.

4.1 General Poisson Setup

In this section we model the process of birthtimes η as a Poisson process. That is, the number of births in disjoint sets are independent and the number of births $\eta(A)$ in a bounded Borel set $A \in \mathcal{B}$ is Poisson distributed with mean $\mathrm{E}\,\eta(A) = \Phi(A) < \infty$. We call Φ the intensity measure for the process of birthtimes.

This assumption about the process of birthtimes is mathematically convenient and we will discuss it in Section 5. A discussion in connection with the applicability of the Poisson process may be found in Brillinger [Bri86].

We state the following theorem in a general version that will be needed in Section 6. Until then, $(E, \mathcal{E})$ could be substituted by $(\mathbb{R}, \mathcal{B})$, $(G, \mathcal{G})$ could be substituted by $(\mathbb{R}_0, \mathcal{B}_0)$, 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 $\eta = \sum_{i \in I} \varepsilon_{\sigma_i}$ we have a family $(Y_i)_{i \in I}$ of independent random variables, Y_i having 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

$$\Lambda(A \times B) = \int_A P(\sigma, B)\, \Phi(d\sigma) \tag{2}$$

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

Proof Brillinger [Bri86] proved the result in Example 4.2 below, but he didn't state the result in the generality of the theorem. The proof carry over to the general situation and is fairly straightforward using Laplace functionals. The theorem is known as "positioning dependent marking" in [Kar91, Example 1.28 and 2.24, Exercise 1.11] and is mentioned in [DVJ88, pp. 205–206]. □

In Section 2.1 we assumed that all the birthtimes are distinct. For this to be true the intensity measure Φ for the process of birthtimes η are not allowed to have any atoms.

Make the mild assumption, that the intensity measure Φ has density φ w.r.t. the Lebesgue measure l on $\mathbb{R}$. 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 φ and λ the intensities for η and μ, respectively.

Example 4.2 (Points in the Lexis diagram) Recall the important difference between the Lexis point process, (σ, x), and the points of death in the Lexis diagram, $(\sigma + x, x)$. The points of death 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 $\tilde{\lambda}(t, a) = \varphi(t - a) f_{t-a}(a)$. That is, the expected number of deaths in calendar time $]t, t + dt]$ and at age $]a, a + da]$ is $\tilde{\lambda}(t, a)\, dt\, da$. This is one of the main results in [Bri86]. $\square$

Example 4.3 (Cross section) For fixed $t \in \mathbb{R}$ define $M : \mathcal{B}_0 \to \mathcal{B} \otimes \mathcal{B}_0$ by $M(A) = \{(\sigma, x) : \sigma < t, x > t - \sigma, t - \sigma \in A\}$. Then $N^t(A) = \mu(M(A))$ is the number of individuals alive at time t with age belonging to 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 e.g. seen from

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

by substitution of s by $u = t - s$. As far as I know this is the first time this is proofed rigourously with all the underlying assumptions clear.

Another approach, that perhaps seems more natural, is to consider the restriction of $\tilde{\mu} = \sum_{i \in I} \varepsilon_{(\sigma_i + X_i, X_i)}$ to the set B_t defined in Section 2.2. Then the process N^t is a transformation of $\tilde{\mu}\,|_{B_t}$ by the function $h : B_t \to \mathbb{R}_0$ given by $h(s, a) = a - (s - t)$. Note that h is *not* injective so we use the theorem in the appendix again. This time with $M = h^{-1}$ where h^{-1} is the inverse image mapping. $\square$

Example 4.4 (Independence) The result in Theorem 4.1 is also a result on independence. 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 is a Poisson process that is independent of deaths before time t. $\square$

We will now consider the parameter space. So far, the only restrictions are that the functions involved are positive and measurable, that $\int_A \varphi(t)\, dt < \infty$ for bounded sets $A \in \mathcal{B}$, and that f_σ is a density, $\int_{\mathbb{R}_0} f_\sigma(x)\, dx = 1$. If we use the hazard $\alpha(t, a)$ introduced on page 2.3 instead of f_σ, the condition is $\int_0^\infty \alpha(\sigma + x, x)\, dx = \infty$ for all σ.

As in Section 3 the expected number of individuals alive at time t must be finite, or else we cannot find a likelihood function. That is $\mathrm{E}\, N^t(\mathbb{R}_0) = \int_0^\infty \varphi(t - a)\bar{F}_{t-a}(a)\, da$ must be finite. One difference from Section 3 is that we here talk about the expected number of individuals alive at time t in the marginal distribution, whereas in Section 3 it was the expected number conditional on the birthtimes. In the time homogeneous case where $\varphi(t) = \varphi$ and $f_\sigma(x) = f(x)$ we have $\mathrm{E}\, N^t(\mathbb{R}_0) = \varphi M$; M denoting the expected lifetime. The equation $\mathrm{E}\, N^t(\mathbb{R}_0) = \varphi M$ is a simple version of the prevalence = incidence $\times$ duration relation, that is considered in depth in the disability model in [Kei91]. This motivates the natural, although not strictly necessary, condition, that the expected lifetime M_σ in the distribution F_σ is finite. In general the condition $\mathrm{E}\, N^t(\mathbb{R}_0) = \int_0^\infty \varphi(t-a)\bar{F}_{t-a}(a)\,da < \infty$ is a trade-off between the number of individuals born as $t - a \to -\infty$ and the tail $\bar{F}_{t-a}(a)$ in the lifetime distributions when $a \to \infty$.

It is easily seen that $\mathrm{E}\, N^t(\mathbb{R}_0)$ is finite if the lifetimes are bounded, or if $\varphi(t-a)$ is bounded as a function of $a \geq 0$ and the lifetimes are stochastically dominated by a distribution G with finite mean. That is, the tail in G is larger than the tail in F_s: for all $s < t, u \in \mathbb{R}_0$: $\bar{F}_s(u) \leq \bar{G}(u)$.

This gives us conditions on specific *pairs* of parameters (φ, f_σ) or equivalently (φ, α). We cannot describe the relevant *family* of parameters. If e.g. the lifetimes are exponentially distributed with $\alpha(t,a) = \alpha > C_2$ for a fixed C_2, then $\varphi(t) = C_1 \exp(-C_2 t)$ is allowed for any C_1. If, on the other hand, we want all exponential lifetimes $\alpha(t,a) = \alpha > 0$ to be allowed, then exponential trend in φ is impossible, and only polynomial trend $\varphi(t) = C_1|t|^{C_2}$ is allowed, C_1 and C_2 arbitrary.

We need a few general results about Poisson processes. See e.g. [DVJ88].

- 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 $S_1 \dot{\cup} S_2 = \{(j,s) \mid (j = 1 \wedge s \in S_1) \vee (j = 2 \wedge s \in S_2)\}$ in an obvious manner.

- 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 $\mu(A) = n$ the $n = |J|$ points $(x_j)_{j \in J}$ of μ on A are iid with distribution $\pi(\,\cdot\,) = \frac{\Lambda(\,\cdot\,\cap A)}{\Lambda(A)}$.

- Let Λ_1 and Λ_2 be intensity measures on $(S, \mathbb{S})$, P_1 and P_2 the distributions of the corresponding Poisson processes. Let $|_A$ denote restriction. If $A \in \mathbb{S}$ satisfies $\Lambda_1(A) < \infty$ and $\Lambda_2(A) < \infty$, and $\Lambda_1 \mid_A \ll \Lambda_2 \mid_A$ then $P_1 \mid_A \ll P_2 \mid_A$ with $P_i \mid_A$ the distribution of the Poisson process with intensity measure $\Lambda_i \mid_A$. If the observation of μ with distribution P_1 on A is $(x_j)_{j \in J}$ where $|J| < \infty$, then the likelihood function is
$$\frac{dP_1 \mid_A}{dP_2 \mid_A}(\mu) = \exp\bigl(-(\Lambda_1(A) - \Lambda_2(A))\bigr) \prod_{j \in J} \frac{d\Lambda_1 \mid_A}{d\Lambda_2 \mid_A}(x_j).$$

For each of the sampling patterns in Section 2.2 we will now investigate how to analyse the observations. The process of birthtimes is Poisson with intensity $\varphi(t)$. The starting point will most often be the Poisson process of deaths, $\tilde{\mu}$, with intensity $\tilde{\lambda}(t,a) = \varphi(t-a) f_{t-a}(a)$.

In the rest of this Section 4 we will 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 $\int_0^\infty \bar{F}(a)\,da$. In Section 5 we return to the non-homogeneous case.

4.2 Cohort Study

We observe deaths in the Lexis diagram on the set O_{cs}. The expected number of deaths is $\int_{O_{\mathrm{cs}}} \tilde{\lambda}(t,a)\, d(t,a) = \varphi \cdot (t_2 - t_1)$, which equals the expected number of births in $]t_1, t_2]$.

Let $(t_j, x_j)_{j \in J}$, where $|J| = \tilde{\mu}(O_{\mathrm{cs}})$, denote the observed times and ages of deaths. Ignoring the factors from the reference measure, that can be taken as the Poisson measure with $\varphi(t) \equiv 1$ and $\alpha(t,a) \equiv 1$, the likelihood is

$$L \propto \exp(-\varphi \cdot (t_2 - t_1))\varphi^{\tilde{\mu}(Ocs)} \prod_{j \in J} f(x_j). \tag{3}$$

We see that the birth intensity is estimated by $\hat{\varphi} = \tilde{\mu}(O_{\mathrm{cs}})/(t_2 - t_1)$ and that the lifetime distribution f must be estimated by $\tilde{\mu}(O_{\mathrm{cs}})$ independent observations from f. As a nonparametric maximum likelihood estimator we can use the Kaplan-Meier estimator for the survival function and the equivalent Nelson-Aalen estimator for the hazard function. See e.g. [ABGK93, Sec. IV.1.5]. The reason that we are able to analyse the lifetimes from the distribution f, is that the expected number of deaths does not depend on f. The expected number of "points" depending on f will be a complication in some of the other sampling patterns.

When conditioning on the number of births $\tilde{\mu}(O_{\mathrm{cs}}) = n$ the points $(t_j, x_j)_{j \in J}$ are iid with density

$$\frac{f(x_j)}{t_2 - t_1} 1\left((t_j, x_j) \in O_{\mathrm{cs}}\right). \tag{4}$$

We are not able to estimate the birth intensity φ since $\tilde{\mu}(O_{\mathrm{cs}})$ is sufficient for φ as seen from (3). The likelihood function has the same dependence on f as (3) because the expected number of deaths does not depend on f. In other words, the expected number of deaths is ancillary for f.

It is tempting to analyse the marginal distribution of the observed ages $(x_j)_{j \in J}$. Define the Poisson counting process $N_{\mathrm{age}}(a) = \tilde{\mu}(O_{\mathrm{cs}} \cap (\mathbb{R} \times [0,a]))$ counting the number of deaths in the cohort up to age a. The intensity process for N_{age} is $\varphi \cdot (t_2 - t_1) f(a)$. If we conditioned on $\tilde{\mu}(O_{\mathrm{cs}}) = n$, the marginal distribution of the ages are n independent observations from f. In Section 4.6 we discuss a situation where considering the calendar times of death does matter.

Sometimes the cohort study is stopped at a time point t_3 after time t_2, and individuals alive at t_3 are only observed to live longer than their age at time t_3. We will go through this situation in some detail since it reveals some important aspects of the Poisson model. At least three different ways of handling this censoring is possible.

Referring to Example 4.2, 4.3, and 4.4, the observation is composed of two independent observations when we stop the study at time t_3:

- An observation of the deaths in the Lexis diagram on the set $O_{\mathrm{cs}} \cap (]-\infty, t_3] \times \mathbb{R}_0)$. Denote the $|J_1| = \tilde{\mu}(O_{\mathrm{cs}} \cap]-\infty, t_3] \times \mathbb{R}_0)$ deaths by $(t_j, x_j)_{j \in J_1}$.
- An observation of the ages of individuals alive at time t_3. This is an observation of N^{t_3} on the set $[t_3 - t_2, t_3 - t_1[$. Let $(a_j)_{j \in J_2}$ be the $|J_2| = N^{t_3}([t_3 - t_2, t_3 - t_1[)$ observed ages.

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 thus the likelihood function is

$$\begin{aligned} L \propto \exp\left(-\int_{t_1}^{t_3}\int_{0\vee s-t_2}^{s-t_1}\varphi f(a)\,da\,ds\right)\prod_{j\in J_1}\varphi f(x_j)\cdot \\ \exp\left(-\int_{t_3-t_2}^{t_3-t_1}\varphi\bar{F}(a)\,da\right)\prod_{j\in J_2}\varphi\bar{F}(a_j) \\ = \exp\left(-\varphi\cdot(t_2-t_1)\right)\varphi^{|J_1|+|J_2|}\prod_{j\in J_1}f(x_j)\prod_{j\in J_2}\bar{F}(a_j). \end{aligned}$$

We see, that the ages must be analysed as independent with ordinary right censoring when we observe the individual alive at time t_3. This is also the case when we condition on the *total* number of individuals observed. In that case the likelihood function is

$$L = \prod_{j\in J_1}\frac{f(x_j)}{t_2-t_1}\prod_{j\in J_2}\frac{\bar{F}(a_j)}{t_2-t_1}. \tag{5}$$

Instead of conditioning on the total number of individuals observed we can condition on the number of deaths, $|J_1|= n_1$, and the number of individuals alive, $|J_2|= n_2$. This will e.g. be reasonable in the synthetic cohort study in Section 4.5, but here the likelihood is

$$L = \prod_{j\in J_1}\frac{f(x_j)}{\int_{t_1}^{t_3}\int_{0\vee s-t_2}^{s-t_1}f(a)\,da\,ds}\prod_{j\in J_2}\frac{\bar{F}(a_j)}{\int_{t_3-t_2}^{t_3-t_1}\bar{F}(a)\,da}.$$

This is a complicated likelihood and we cannot interpret it due to the two expected numbers in the denominator not having any separate meaning. It does not seem reasonable to try to maximize it. In contrast, we will see in Section 4.5 a situation where the most reasonable thing to do, is to condition on the number of individuals alive and death respectively, rather than the total number of individuals.

The third possible way of handling the right censoring is to condition on the total number of individuals and then make a transformation of the distribution (4). The transformation we have in mind is the identity on $O_{\text{cs}}\cap(]-\infty,t_3]\times\mathbb{R}_0)$ and $h(t,a) = (t_3, a-(t-t_3))$ on $O_{\text{cs}}\setminus(]-\infty,t_3]\times\mathbb{R}_0)$. The first of these two sets is the set on which we observe the deaths, and the second set contains the deaths in the Lexis diagram that we do not observe because the lifetimes are censored. The censored points of deaths are transformed to the ages at time t_3. It is easy to show that this yields the distribution (5). (Here we ignore the burden of specifying the exact reference measure.) This is quite general for Poisson processes on a space with a finite number of expected points — making a transformation of the Poisson process and conditioning on the number of points yields the same distribution, as if we condition on the number of points and then make a transformation of the distribution. This means, that we are free to choose which one of these two approaches that seems most natural in the context.

4.3 Time Window

We are observing the deaths in the set O_{tw}. This is, word for word, the same as in Section 4.2 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.

It is, however, important to notice that we are analysing the lifetimes *as if* we had $\tilde{\mu}(O_{\mathrm{tw}})$ independent observations from f. Actually, this is not the case since individuals are represented by life lines with slope 1 that cross the area on which we observe the Lexis diagram.

To understand this, take a look at Figure 4. The two sets O_{tw} and O_{cs} are composed of similar elementary triangles. By similar triangles we mean triangles with the same size, shape, orientation, and position in the age direction. It is allowed to translate the triangles in calendar time because this does not alter the contribution to the likelihood function and thus similar triangles give the same contribution to the likelihood function. In the discussion [Chi86] of [Bri86] there are some remarks not unlike these. In [Kei90] there is a more thorough examination of elementary sets in the Lexis diagram.

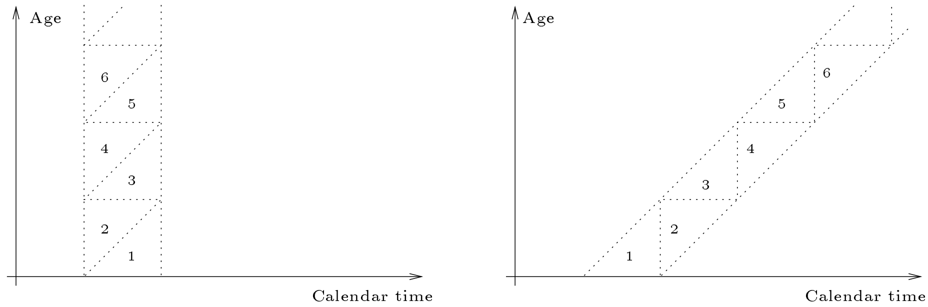


Figure 4: Observing deaths in a time window and a cohort study is the same because the two sets are composed of the same elementary triangles.

4.4 Cross Sectional Study

In a cross sectional study at time t we observe N^t on the whole axis $\mathbb{R}_0$. Let $(a_j)_{j\in J}$ denote the $|J|= N^t(\mathbb{R}_0)$ observed ages. The expected number of individuals alive at time t is φM, and thus the likelihood function equals

$$L \propto \exp(-\varphi M)\varphi^{N^t(\mathbb{R}_0)} \prod_{j\in J} \bar{F}(a_j).$$

Since the mean lifetime M depends on f, it is not obvious how to maximize this likelihood function.

Conditioning on the observed number of individuals $N^t(\mathbb{R}_0) = n$, the observed ages are iid with density $\bar{F}(a)/M$. This is the stationary recurrence distribution known from renewal theory. See e.g. [Fel71] page 369 (4.6), 371 (4.16), and Problem 10 page 386 for an overview of biased distributions in renewal theory. This reflects the fact that we observe the ages of individuals alive and thus are subject to interception bias. (Vardi [Var88] advocates the terminology “interception bias” for the stationary recurrence time distribution.) The density is, not surprisingly, decreasing, as the fraction of young individuals in the population is larger than the fraction of old individuals. It has been known for a long time, that the density of ages of individuals alive at a fixed time point in a stationary population is proportional to the survival function for the lifetime distribution.

The nonparametric maximum likelihood estimator from n iid replications from the recurrence time distribution is known in the literature as the Grenander estimator. The estimator is e.g. described in [BBBB72, pp. 223], [DV86], and [Var89, Problem D].

4.5 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 according to Example 4.4.

Let $(t_j, x_j)_{j\in J_1}$, where $|J_1|= \tilde{\mu}(O_{\mathrm{tw}})$, denote the observed deaths and $(a_j)_{j\in J_2}$ the $|J_2|= N^{t_2}(\mathbb{R}_0)$ observed ages of individuals alive. For the cross sectional study it was seen, 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. When conditioning on $\tilde{\mu}(O_{\mathrm{tw}}) = n_1$ and $N^{t_2}(\mathbb{R}_0) = n_2$ the likelihood equals

$$L \propto \prod_{j\in J_1} f(x_j) \prod_{j\in J_2} \frac{\bar{F}(a_j)}{M}. \tag{6}$$

The nonparametric maximum likelihood estimator when we have observations from f and the corresponding interception biased distribution $\bar{F}/M$ is considered in Vardi [Var82]. Vardi develops the RT algorithm to maximize a likelihood function from renewal theory that has factors of the form in (6). Unfortunately the algorithm is for distributions on the integers $\mathbb{N}$. Soon and Woodroofe [SW96] extends the RT algorithm to arbitrary distributions.

Note that the expected total number of individuals observed is $\varphi \cdot (M + (t_2 - t_1))$. If we condition on the total number $\tilde{\mu}(O_{\mathrm{tw}}) + N^{t_2}(\mathbb{R}_0)$ being equal to n, we obtain the likelihood

$$L \propto \prod_{j\in J_1} \frac{f(x_j)}{M + (t_2 - t_1)} \prod_{j\in J_2} \frac{\bar{F}(a_j)}{M + (t_2 - t_1)}.$$

We cannot interpret this likelihood and thus it does not seem reasonable to condition on the total number of individuals.

4.6 Follow Up on Cross Sectional Study

We are following 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. Here we choose to condition on the total number of individuals and then make a transformation (censoring) of the observations, as this gives the easiest interpretation. Simon [Sim80] 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 $|J|= \tilde{\mu}(B_{t_1})$ deaths $(t_j, x_j)_{j\in J}$ in the Lexis diagram on the set B_{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

$$L \propto \exp(-\varphi M)\varphi^{\tilde{\mu}(B_{t_1})} \prod_{j\in J} f(x_j)$$

and we notice that the mean lifetime M depends on f. Conditioning on $\tilde{\mu}(B_{t_1}) = n$, the points $(t_j, x_j)_{j\in J}$ are iid with density

$$\frac{f(x_j)}{M} 1\left((t_j, x_j) \in B_{t_1}\right). \tag{7}$$

The distribution (7) is known from renewal theory as the simultaneous distribution of a forward and backwards recurrence time; see the references to [Fel71] in Section 4.4. 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 a interpretation of x_j as the sum of the two recurrence times is natural. Note that $x_j = b_j + c_j$, and that 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 based on (7).

When stopping observation at time t_2, we transform the points (t_j, x_j) 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 then the likelihood is

$$L \propto \prod_{j\in J_1} \frac{f(x_j)}{M} \prod_{j\in J_2} \frac{\bar{F}(a_j)}{M} \tag{8}$$

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 (7).

Note that right censoring in the marginal distribution of x_j gives a different result! This shows, that it is important to keep information on the calendar time t_j of death. The difference between the two approaches is whether the censoring regards the total lifetime x or only the lifetime lived after time t_1. To obtain the lifetime lived after time t_1 we need the time t_j of death.

Winter and Földes [WF88] find the nonparametric maximum likelihood estimator for the likelihood (8) in a renewal process setup. In [KG88, Section 7] the estimator is re-derived in a counting process framework, and in [Var89, Problem B] the estimator is found in yet another setup.

4.7 Current Status Data

Current status data in the general setup from Section 2.2 does not really fit into the Poisson setup. The reason is that we observe a Poisson process on a set with an infinite expected number of points and thus we cannot write down a likelihood. We observe a cross section at time t. Independent hereof we see the (non injective) transformation $h(s, a) = s - a$ of the points in the Lexis diagram on the set $]-\infty, t] \times \mathbb{R}_0$. This transformation yields a Poisson process on $]-\infty, t]$ with intensity $\varphi F(t - s)$ at time s for $s < t$. The total expected number of points is of course infinite. We see that as time s tends to $-\infty$ the intensity tends to that of the process of birthtimes because no individuals survive until time t.

We can modify the setup slightly to obtain more reasonable results.

Assume that we only make an observation corresponding to the individuals born in $]\tilde{t}, t]$ for some $\tilde{t} < t$. Let $(\sigma_j)_{j \in J}$ denote the $|J|$ observed birthtimes in $]\tilde{t}, t]$. Make a disjoint union of J in two sets J_1 and J_2, where J_1 are the individuals that survive until time t and J_2 are the individuals that die before time t. The likelihood is

$$L \propto \exp\left(-\varphi \cdot (t - \tilde{t})\right) \prod_{j \in J_1} \bar{F}(t - \sigma_j) \prod_{j \in J_2} (1 - \bar{F}(t - \sigma_j)). \tag{9}$$

This likelihood is very similar to the likelihood (1), so estimation of the lifetime distribution is the same whether we condition on the birthtimes or not. We can condition on the total number of individuals without changing the way (9) depends on the lifetime distribution f, but it will not work to condition on the numbers $|J_1|$ and $|J_2|$.

The modification we obtain by assuming that we have bounded survival times, $\bar{F}(x) = 0$ for all $x > C$ is not substantially different from the above modification. We still need to restrict the observation to a bounded interval of birthtimes.

5 Discussion of some Assumptions

In this section we discuss some of the assumptions we use. In Section 5.1 we discuss the Poisson assumption from Section 4, in Section 5.2 what happens when the process of birthtimes is non time-homogeneous, and in Section 5.3 estimation of the lifetime distribution when this distribution depends on calendar time.

Remember from Section 3, that if we condition on the birthtimes for observed individuals the distribution of the process of birthtimes does not matter at all.

5.1 The Poisson Assumption

The main advantage of having 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. If we condition on the birthtimes as in Section 3 we can say nothing in this case, so an assumption on the process of birthtimes is necessary.

The assumption that the process of birthtimes is a Poisson process is the most simple and non-informative assumption we can make. To highlight the point that the Poisson assumption enables the utilization of information on survival until the start of the study, notice that if the mean lifetime M is small compared to the size of the window, then several of the situations considered are nearly identical. First of all the synthetic cohort and time window are very much alike. If the observation interval $]t_1, t_2]$ is large compared to M, the fraction of individuals that die in $]t_1, t_2]$ compared to the number of individuals alive at time t_2 tends to one. Furthermore the fraction of individuals that are both born and die in $]t_1, t_2]$ tends to one. This means that it will make no difference, if we use the Poisson assumption or condition on the birthtimes when M is small compared to $]t_1, t_2]$ — the likelihood functions will be almost identical since the individuals that are both born and die in $]t_1, t_2]$ in all cases contribute by a factor $f(x)$ to the likelihood function. In e.g. the case when we condition on the birthtimes, nearly no lifetimes are left truncated or right censored. 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.

The most important aspect of the Poisson assumption is the independence between individuals. Without this independence, the results in Section 4 would be different. Note, that we most often condition on a suitable number of individuals before the lifetimes can be analysed as independent. The marginal and the conditional likelihoods are the same if the number on which we condition does not depend on the lifetime distribution f. When this number is not ancillary for f, then the expected value must be a "nice" function of f. Otherwise we cannot interpret the likelihood function. Intuition tells us that if the mean number of newborn is correctly specified and the individuals are close to being independent then the results of Section 4 will be approximately correct.

In Brillinger [Bri86] and the accompanying discussion there are various comments on the applicability of the Poisson assumptions in real life setups.

The assumption in Section 4 about stationarity of the Poisson process of birthtimes is discussed in the next Section 5.2.

5.2 Process of Birthtimes Depends on Calendar Time

We use the assumption that the process of birthtimes is time-homogeneous when we use the Poisson assumption in Section 4. We will now discuss what happens when the process of birthtimes is non time-homogeneous.

Consider the case of a non time-homogeneous Poisson process describing the births and the lifetimes being time-homogeneous; $f_s(a) = f(a)$. *Assume* that the birth intensity $\varphi(t)$ is known. In the case of a cross sectional study (as in Example 4.3) we observe the $|J| = N^t(\mathbb{R}_0)$ ages $(a_j)_{j\in J}$ of individuals alive at time t. The expected number of individuals is $\int_0^\infty \varphi(t-a)\bar{F}(a)\,da$ according to Example 4.3 and thus the likelihood is

$$L \propto \exp\left(-\int_0^\infty \varphi(t-a)\bar{F}(a)\,da\right)\prod_{j\in J}\varphi(t-a_j)\bar{F}(a_j)$$

and, conditional on $N^t(\mathbb{R}_0) = n$, the ages $(a_j)_{j\in J}$ are iid with density

$$\frac{\varphi(t-a)\bar{F}(a)}{\int_0^\infty \varphi(t-a)\bar{F}(a)\,da}. \tag{10}$$

In both cases φ appears as a weight function for weighting the observed ages in a suitable manner. If e.g. the mean number of newborn is increasing, more weight is given to young individuals.

Now define the age transformation $h(a) = \int_0^a \varphi(t-s)\,ds$ and let $\tilde{a}_j = h(a_j)$. Then by transformation of (10) with h, the transformed ages $(\tilde{a}_j)_{j\in J}$ are iid with density $\bar{F}(h^{-1}(\tilde{a}))/\tilde{M}$ where $\tilde{M}$ is the mean lifetime in the distribution with survival function $\bar{F}\circ h^{-1}$. As in Section 4.4 we obtain an estimate for $\bar{F}\circ h^{-1}$, and h is known so we get an estimate of $\bar{F}$.

A similar approach can be used in the other sampling patterns. When φ is not known in advance one might use an estimate of φ in some suitable manner.

Keiding [Kei92, p. 319] gives the advice to avoid the reliance on the stationarity assumption when it is possible. The alternative is to condition on the birthtimes.

5.3 Lifetime Distribution Depends on Calendar Time

Keiding discusses in the survey paper [Kei90] approaches to estimation of the lifetime distribution when the distribution depends on calendar time as well as age. We will not go into details here, but rather refer to that paper. Keiding only considers the case when we condition on the birthtimes for individuals observed. One main example is the following.

Example 5.1 (Piecewise constant hazards) One particularly simple example is when the hazard $\alpha(t, a)$ is piecewise constant, i.e. we have a tessellation of the Lexis plane and $\alpha(t, a)$ is constant on each cell. As in the ordinary case with the age axis only, the maximum likelihood estimator of each of these constants is an "occurrence exposure" rate — the number of deaths divided by the total time under risk. This is used in the synthetic cohort study for the construction of life tables. Explanatory variables can easily be incorporated into the model, and if the effect on α is log-linear, software for the analysis of log-linear Poisson models can be used. In [Kei90] the problem of identifying the parameters in a age-period-cohort model is discussed and further references are given. $\square$

6 Generalizations

In this section we consider some more advanced examples. In Section 6.1 we generalize the Lexis diagram to a more complicated setup than just lifetimes. The concepts introduced is exemplified by the disability model in Section 6.2. We will focus on probabilistic statements about the disability model, whereas [Kei91] focuses on estimation. Finally in Section 6.3 we state some simple theorems in a more general case than the disability model.

The advantage of the general setup in this section is that we get a formal common framework to handle a multitude of ways to sample individuals in complicated models in a rigourous way. Moreover, the Poisson assumption on the process of birthtimes leads to new Poisson processes and independence results.

6.1 Generalizations of the Lexis Diagram

We often want to include background variables in our model. Inspired by [Sim80] we can take the process of birthtimes on a larger space $(E, \mathcal{E})$ than the reals $\mathbb{R}$. If e.g. we want the covariate gender in our model, we can take $(E, \mathcal{E})$ as $\mathbb{R} \times \{\text{male}, \text{female}\}$ and the process of birthtimes η is a point process on $\mathbb{R} \times \{\text{male}, \text{female}\}$ instead of $\mathbb{R}$.

Likewise we will need a more complicated description of an entire life history than just the lifetime X. We call a description of a whole life Y and assume it has values in a space $(G, \mathcal{G})$. Such a description can e.g. be a stochastic process $(Y_a)_{a \geq 0}$ describing the status Y_a of the individual at age a. By status we think of e.g. whether the individual has some disease or not. As demonstrated in [ABGK93] marked point processes on the real line form a flexible class of processes for the description of lives. This can be illustrated in the Lexis diagram by adding marks on the life lines at the time points where events occur. There will most often be an age given as some function of the life history Y, which we can call the lifetime. Let X denote this lifetime.

We will now 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 the space $(E, \mathcal{E})$. 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 on the space $(G, \mathcal{G})$. 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 $\mu = \sum_{i \in I} \varepsilon_{(\sigma_i, Y_i)}$. In the same way as 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 a notation similar to that section. Theorem 3.1 carries over without problems; the only change is the change of the spaces. We have already formulated Theorem 4.1 in the general context here.

If we want to incorporate covariates in the distribution of the process of birthtimes, we can use Theorem 4.1 twice in the Poisson setup. First generate the birthtimes and variables using one Markov kernel. Second use the resulting Poisson process as a process of birthtimes, associating a description Y of the life with each combination of birthtime and variable.

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

Consider the disability model with the three states H (healthy), I (invalid, inactive) and D (dead):

$$\begin{tikzcd} {} \arrow[r, "\varphi(t)"] & H \arrow[rr, "\gamma(t,a)"] \arrow[dr, "\alpha_H(t,a)"'] & & I \arrow[dl, "\alpha_I(t,a,d)"] \\ & & D & \end{tikzcd}$$

We assume that the process of birthtimes is Poisson with intensity $\varphi(t)$, and that 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, age a and duration d. No reactivation is allowed, i.e. the transition $I \to H$ is not possible.

To elaborate a little bit, the process of birthtimes is on the real line $\mathbb{R}$ and the lives of individuals are independent given the birthtimes. The space G of life descriptions can be taken as all piecewise constant cadlag functions $(Y_a)_{a \geq 0}$ with state space $\tilde{G} = \{H, I, D\}$ and only finitely many jumps in finite time.

The transition intensities specify the Markov kernels $P_\sigma(\,\cdot\,)$ and e.g.

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

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.

Definition 6.1 Define for $Y \in G$

$$T_1 = \inf_a (Y_a \neq H)$$
$$T_2 = \inf_a (Y_a = D)$$

with the convention that $\inf \emptyset = +\infty$. Now a.s., T_1 is the age at the time of the first jump. 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 $\mu = \sum_{i \in I} \varepsilon_{(\sigma_i, Y_i)}$ be the Lexis process and let # denote "number of". Now define for $t \in \mathbb{R}$, $A, B \in \mathcal{B}_0$ and $C \in \mathcal{B}$ the following quantities:

$$\begin{aligned}
N^{t,H}(A) &= \{\#\text{ individuals in state } H \text{ at time } t \text{ and age} \in A \\
&= \mu((\sigma, Y) : \sigma < t, Y_{t-\sigma} = H, t - \sigma \in A) \\
N^{t,I}(A \times B) &= \begin{cases} \#\text{ individuals in state } I \text{ at time } t\text{, age} \in A \text{ and duration} \\ \in B \end{cases} \\
&= \mu((\sigma, Y) : \sigma < t, Y_{t-\sigma} = I, t - \sigma \in A, (t - \sigma) - T_1 \in B) \\
N^{t,H,D}(C \times A) &= \begin{cases} \#\text{ individuals in state } D \text{ at time } t\text{, coming from state} \\ H\text{, who made the transition } H \to D \text{ at calendar time} \\ \in C \text{ and at age } T_1 \in A \end{cases} \\
&= \mu((\sigma, Y) : \sigma < t, Y_{t-\sigma} = D, Y_{T_1} = D, T_1 \in A, \sigma + T_1 \in C) \\
N^{t,I,D}(C \times A \times B) &= \begin{cases} \#\text{ individuals in state } D \text{ at time } t\text{, coming from state} \\ I\text{, who made the transition } I \to D \text{ at calendar time} \\ \in C\text{, at age } T_2 \in A \text{ and duration } T_2 - T_1 \in B \end{cases} \\
&= \mu((\sigma, Y) : \sigma < t, Y_{t-\sigma} = D, Y_{T_1} = I, T_2 \in A, \\
&\qquad T_2 - T_1 \in B, \sigma + T_2 \in C) \\
N^{H,I}(C \times A) &= \begin{cases} \#\text{ individuals making the transition } H \to I \text{ at age} \in A \\ \text{and calendar time} \in C \end{cases} \\
&= \mu((\sigma, Y) : T_1 \in A, \sigma + T_1 \in C, Y_{T_1} = I)
\end{aligned}$$

Theorem 6.2 *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 on age axis $\mathbb{R}_0$, age axis $\times$ duration $\mathbb{R}_0 \times \mathbb{R}_0$, calendar time $\times$ age axis $\mathbb{R} \times \mathbb{R}_0$, calendar time $\times$ age axis $\times$ duration $\mathbb{R} \times \mathbb{R}_0 \times \mathbb{R}_0$, and calendar time $\times$ age axis $\mathbb{R} \times \mathbb{R}_0$ respectively.*

These versions of $N^{t,H}$, $N^{t,I}$, $N^{t,H,D}$ and $N^{t,I,D}$ are independent.

The intensity processes are

- *for $N^{t,H}$:*

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

- *for $N^{t,I}$:*

$$\begin{aligned}
\lambda^{t,I}(a,d) = 1\,(d \leqq a)\, \varphi(t-a) &\exp\left(-\int_0^{a-d} (\alpha_H + \gamma)(t - a + u, u)\, du\right) \cdot \\
&\gamma(t-d, a-d) \exp\left(-\int_{a-d}^{a} \alpha_I(t - a + u, u, u - (a-d))\, du\right)
\end{aligned}$$

- *for* $N^{t,H,D}$*:*

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

- *for* $N^{t,I,D}$*:*

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

- *for* $N^{H,I}$*:*

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

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 relates to bounded sets for the process of birthtimes, and hence the expected values are finite. The lifetime distributions are continuous so one point sets have measure 0. 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 0 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 definition of the processes and the intensity measure (2). □

Several remarks can be made:

- The problem with sets of intensity measure 0 and versions of processes can be avoided by restricting the space G (and the probability measure etc.) to the set of processes with the interesting sample paths.
- 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 $N^{t,I,D}$ is more about the time point for the transition to state D.
- The process $N^{H,I}$, "disease onset", is independent of $N^{t,H}$ and $N^{t,H,D}$ as long as we only consider transitions $H \to I$ before time t.
- In the case of reactivation allowed, the processes $N^{t,H}$, $N^{t,I}$, $N^{t,H,D}$ and $N^{t,I,D}$ are still independent. (Suitable redefinitions considering duration have to be made.) The process $N^{H,I}$ is not a simple point process in this setting.

If the expected number of individuals alive at time t is finite, that is $\mathrm{E}\,N^{t,H}(\mathbb{R}_0) < \infty$ and $\mathrm{E}\,N^{t,I}(\mathbb{R}_0\times\mathbb{R}_0) < \infty$, we immediately see 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

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

Example 6.3 (Prevalent cohort study) The prevalent cohort study is considered in [Sim80] and [Kei92, Sec. 7]. At a certain time t_1 we sample individuals in the disease state I, 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 [Sim80] assumes that individuals enter state I as a time homogeneous Poisson process in calendar time and he does not consider age at all. So "when living in a world of only diseased" we can apply the results in Section 4.6 to the duration in state I. We will comment on the situation where age is considered.

Keiding [Kei92] states in a rather informal way that conditional on the age at disease onset, the time of death, and that the individual attains an age larger than the age at disease onset, the hazard for the age at death is α_I. We will now state a similar result in the rigorous setup developed so far. Note that we 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 the 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 thus we have independent delayed entry in the terminology of [Kei92].

If we want to use information on the duration up to the sampling point t_1 we must use the Poisson assumption of the process of birthtimes σ_i. 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 are entering state I as a homogeneous Poisson process in calendar time t, even though the intensity $\gamma(a)$ is allowed to depend on age a. We are in the case of [Sim80] and Section 4.6. More formally, we observe (a slight variation of) the process $N^{t,I,D}$ on the set of diseased individuals alive at time t_1. We can easily show that conditional on the number of sampled individuals, the 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. We see, that the duration d must be analyzed as in Section 4.6 and that we also have the distribution of age at entry to state I. □

6.3 Piecewise Constant Life Descriptions

Let the space $(G,\mathcal{G})$ of life descriptions be the space of all cadlag processes $(Y_a)_{a\geq 0}$ that are piecewise constant with state space $(\tilde{G},\tilde{\mathcal{G}})$ and finitely many jumps in finite time. Define for $t\in\mathbb{R}$, $B\in\tilde{\mathcal{G}}$ and $A\in\mathcal{B}_0$

$$\begin{aligned}N^{t,B}(A) &= \{\#\text{ individuals at time } t \text{ with state} \in B \text{ and age} \in A\\ &= \mu((\sigma,Y):\sigma<t, Y_{t-\sigma}\in B, t-\sigma\in A),\end{aligned}$$

and similarly for $B_1,B_2\in\tilde{\mathcal{G}}$ disjoint, $C\in\mathcal{B}$ and $A\in\mathcal{B}_0$ define

$$\begin{aligned}N^{B_1,B_2}(C\times A) &= \begin{cases}\#\text{ individuals that in calendar time} \in C \text{ and age} \in A\\ \text{makes a transition from a state} \in B_1 \text{ to a state} \in B_2\end{cases}\\ &= \mu((\sigma,Y):\exists t\in C \text{ so } t-\sigma\in A, Y_{(t-\sigma)-}\in B_1, Y_{t-\sigma}\in B_2).\end{aligned}$$

Then, in the spirit of Theorem 6.2 and with the same proof, we may formulate the following two theorems.

Theorem 6.4 *For $t\in\mathbb{R}$ and $B_1,\ldots,B_n\in\tilde{\mathcal{G}}$ pairwise disjoint, the processes $N^{t,B_1},\ldots,N^{t,B_n}$ have independent versions that are Poisson processes with intensity processes $\lambda^{t,B_j}(a)=\varphi(t-a)P_{t-a}(Y_a\in B_j)$.*

Theorem 6.5 *Let $B_1,B_2\in\tilde{\mathcal{G}}$ be disjoint and assume that for Φ-a.e. $\sigma\in\mathbb{R}$*

$$P_\sigma(Y_{a_1}\in B_2, Y_{a_2}\in B_1 \textit{ for some } a_1,a_2\,:\,0\leq a_1<a_2)=0.$$

Then N^{B_1,B_2} has a version that is a Poisson process on calendar time $\times$ age axis $\mathbb{R}\times\mathbb{R}_0$.

The assumption in Theorem 6.5 assures that for Λ-a.e. (birthtime, life) a transition from B_2 to B_1 does not happen.

It is essential for N^{B_1,B_2} to be a counting measure that the number of *individuals* making the transition from B_1 to B_2 equals the number of *transitions* from B_1 to B_2. If we redefined N^{B_1,B_2} in a suitable manner, so we count the number of transitions instead and dropped the assumption in the theorem, then N^{B_1,B_2} would become a counting measure. However, in this situation N^{B_1,B_2} is in general not a Poisson process because an individual can make several transitions from B_1 to B_2.

7 Discussion

We have seen that length biased distributions from renewal theory arise in many cases when we assume that the process of birthtimes is a Poisson process. In fact, renewal processes are sometimes used in survival analysis in the arguments for the likelihood function, instead of arguments based on individuals. The renewal setup only seems reasonable when an actual replacement takes place, although it is used anyway — see e.g. [MD95] and [WF88]. Maximisation of the likelihood functions with biased lifetime distributions is so far considered in the context of renewal theory. The estimators will be the same as here, but it might be that the proofs for consistency and week convergence are no longer valid.

The synthetic cohort is the classical case. We observe individuals alive as well as dead during a time period. The synthetic cohort is not unlike the line-segment problem in [vdL96]. The difference is that in the line-segment problem, we do not see the ages of individuals alive at time t_1, i.e. we only see the possible censored time lived after time t_1. Line-segment similar problems are sometimes put into a renewal process setup, see e.g. [Var82]. In [MD95] the renewal process setup is used in nurse manpower planning. One might argue that the renewal setup is a bit artificial in this application since nurses do not live forever, and we only see certain renewals. A model along the line of the disability model in Section 6.2 will probably be more natural, although the likelihoods are the same. Another difference between the nurse planning problem and the synthetic cohort is that individuals alive at both time t_1 and t_2 are not observed in [MD95].

The follow up on a cross sectional study has also been treated in a renewal process setup, see [WF88] and [Var89, problem B]. The variant where the age at time t_1 is unobservable is treated in a renewal process setup in [DV86]. This is quite general, the variations with age at entrance unobserved must be handled in the Poisson setup in Section 4.

8 Open Problems

We have developed a rigorous framework to handle sampling problems in survival analysis in calendar time. In Section 3 and 4 we systematically investigated six sampling patterns in the basic setup with only lifetimes. There is still room for further improvement, however.

It will be worth doing a more systematic study of the inclusion of covariates in the models. Simon [Sim80] uses a dichotomous covariate as gender, and compares the survival in the two groups. It is interesting to extend this scheme. Furthermore semi parametric models such as the Cox model would be of interest. In the Poisson setup it might be that we must condition on the number of individuals in different (say) treatment groups. Note that the general setup allows time dependent covariates by letting a part of the space of birthtimes E be a stochastic process. One can discuss if time dependent covariates should be modeled as a part of the life in the space of life descriptions G instead — this will allow the covariates to change as a function of the development of the life and give new variations of the sampling patterns as well.

It is also of interest to extend the study of more complicated descriptions of lives than just lifetimes. A good starting point is the article [Kei91] by Keiding and the disability model in Section 6.2. A lot of problems concerning the sampling of the individuals can arise.

A line of further development, not touched upon so far, is the possibility of observing the Lexis diagram on *random* sets O, rather than the deterministic sets in the sampling patterns in Figure 2. This will permit a lot of new interesting sampling patterns and some problems. We could allow for different stopping times for different individuals, stopping of observation when certain conditions (on the Lexis process) are fulfilled etc. One of the problems we could encounter is that we would no longer have independent censoring and truncation.

We use the assumption, that the process of birthtimes is a stationary Poisson process throughout Section 4. It is of interest to elaborate on this assumption in more practical situations. In the discussion [Cox91] of [Kei91], Cox suggests that the Poisson assumption is not that important in many situations. It will also be of interest to study the stability of the obtained estimates, when the Poisson assumption is not quite fulfilled. In e.g. population dynamics the current birthrate depends on the size of the population, but it might be that the Poisson assumption is approximately fulfilled. See also the discussion in [Bri86]. Alho [Alh92] extends the relation prevalence = incidence $\times$ duration from the time homogeneous case in [Kei91] to the situation with exponentially increasing or decreasing populations. Some of these methods can perhaps be used in case of non stationarity in the Poisson setup.

It might also be of interest to relax the basic Assumption 2.2 of independence. The following example with phone calls show, that it can be desirable to model some dependence between lifetimes. Let the lifetimes be the length of phone calls and suppose we have n phone lines available. This means, that we can have a maximum of n calls going on at any time point t. Call number $n + 1$ will be refused and the length of calls thus have a point mass at 0. A dependence between the length of calls is introduced, both unconditional and conditional on the initiation times. This is in connection with the development of the filtration in time: the filtration for the process of birthtimes develops in a natural way, but we do not observe the marks (the life description Y) at the time of birth. As calendar time evolves we see more and more of the marks associated with earlier birthtimes.

Acknowledgments

This work is based on my master's thesis [Lun96]. I am very indebted to my supervisor Martin Jacobsen, with whom I have spent many hours discussing this work. I also owe a debt of gratitude to Helle Sørensen who have read this manuscript at various stages and have given many valuable comments. Niels Keiding has kindly given a lot of references and many helpful remarks, which has improved the paper. Part of this work was done while I visited the department of mathematical statistics, Chalmers University of Technology, Göteborg, Sweden. This visit was supported by a grant from NorFA.

References

[ABGK93] Per Kragh Andersen, Ørnulf Borgan, Richard D. Gill, and Niels Keiding, *Statistical models based on counting processes*, Springer Series in Statistics, Springer-Verlag, 1993.

[Alh92] Juha M. Alho, *On prevalence, incidence, and duration in general stable populations*, Biometrics **48** (1992), 587–592.

[BBBB72] R. E. Barlow, D. J. Bartholomew, J. M. Bremner, and H. D. Brunk, *Statistical inference under order restrictions*, John Wiley & Sons, 1972.

[Bri86] David R. Brillinger, *The natural variability of vital rates and associated statistics*, Biometrics **42** (1986), 693–734, with discussion.

[Chi86] Chin Long Chiang, *Discussion on the paper [Bri86]*, Biometrics **42** (1986), 714–717.

[Cox91] David Cox, *Discussion of the paper by Keiding*, Journal of the Royal Statistical Society A **154** (1991), no. 3, 398–399, discussion of [Kei91].

[DV86] L. Denby and Y. Vardi, *The survival curve with decreasing density*, Technometrics **28** (1986), no. 4, 359–367.

[DVJ88] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes*, Springer Series in Statistics, Springer-Verlag, 1988.

[Fel71] William Feller, *An introduction to probability theory and its applications*, second ed., vol. II, John Wiley & Sons, 1971.

[Jac82] Martin Jacobsen, *Statistical analysis of counting processes*, Lecture Notes in Statistics, vol. 12, Springer-Verlag, 1982.

[Kar91] Allan F. Karr, *Point processes and their statistical inference*, 2. ed., Probability: Pure and Applied, Marcel Dekker, 1991.

[KBSH96] Niels Keiding, Kamilla Bergtrup, Thomas H. Scheike, and Günther Hasibeder, *Estimation from current-status data in continous time*, Lifetime Data Analysis **2** (1996), 119–129.

[Kei90] Niels Keiding, *Statistical inference in the Lexis diagram*, Philosophical Transactions of the Royal Society London A **332** (1990), 487–509.

[Kei91] Niels Keiding, *Age-specific incidence and prevalence: a statistical perspective*, Journal of the Royal Statistical Society A **154** (1991), no. 3, 371–412, with discussion.

[Kei92] Niels Keiding, *Independent delayed entry*, Survival Analysis: State of the Art (J. P. Klein and P. K. Goel, eds.), NATO ASI Series, Series E: Applied Sciences, vol. 211, Kluwer Academic Publishers. Printed in the Netherlands, 1992, pp. 309–326.

[KG88] Niels Keiding and Richard Gill, *Random truncation models and Markov processes*, Report MS-R8817, Centrum voor Wiskunde en Informatica, Department of Mathematical Statistics, 1988.

[Lex75] Wilhelm Lexis, *Einleitung in die theorie der bevölkerungs-statistik*, Strassbourg: Trübner (1875), pp. 5–7 translated to English by Nathan Keyfitz in [SK77].

[Lun96] Jens Lund, *Analyse af ventetidsaksen i en fornyelsesproces og sampling i en poissonprocesmodel for Lexis diagrammet*, Master's thesis, University of Copenhagen, Institute of Mathematical Statistics, November 1996, in danish. Title translates to "Analysis of the Interrenewal Times in a Renewal Process and Sampling in a Poisson Process Model for the Lexis Diagram". Available from `ftp://ftp.dina.kvl.dk/pub/Staff/Jens.Lund/master_thesis/thesis.ps`.

[MD95] Sally McClean and Colum Devine, *A nonparametric maximum likelihood estimator for incomplete renewal data*, Biometrika **82** (1995), no. 4, 791–803.

[Sim80] Richard Simon, *Length biased sampling in etiologic studies*, American Journal of Epidemiology **111** (1980), no. 4, 444–452.

[SK77] David Smith and Nathan Keyfitz, *Mathematical demography*, Biomathematics, vol. 6, Springer-Verlag, 1977, selected papers.

[SKM95] Dietrich Stoyan, Wilfrid S. Kendall, and Joseph Mecke, *Stochastic geometry and its applications*, sec. ed., Wiley, 1995.

[SW96] Guoxing Soon and Michael Woodroofe, *Nonparametric estimation and consistency for renewal processes*, Journal of Statistical Planning and Inference **53** (1996), 171–195.

[Tur76] B. W. Turnbull, *The emperical distribution function with arbitrarily grouped, censored and truncated data*, Journal of the Royal Statistical Society B **38** (1976), 290–295.

[Var82] Yehuda Vardi, *Nonparametric estimation in renewal processes*, The Annals of Statistics **10** (1982), no. 3, 772–785.

[Var88] Yehuda Vardi, *Statistical models for intercepted data*, Journal of the American Statistical Association, Theory and Methods **83** (1988), no. 401, 183–197.

[Var89] Yehuda Vardi, *Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation*, Biometrika **76** (1989), no. 4, 751–761.

[vdL96] Mark J. van der Laan, *Efficiency of the* NPMLE *in the line-segment problem*, Scandinavien Journal of Statistics **23** (1996), 527–550.

[WF88] B. B. Winter and A. Földes, *A product-limit estimator for use with length-biased data*, The Canadian Journal of Statistics **16** (1988), no. 4, 337–355.

Appendix: Non-Injective Transformation of Poisson Processes

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 $A \in \tilde{\mathbb{S}}$.

In the following theorem we talk about bounded sets $A \in \tilde{\mathbb{S}}$. For this to be valid we need some structure on $\tilde{S}$, e.g. we could require $\tilde{S}$ to be a complete separable metric space or a product of such a space and a space without any special structure. In the last case the bounded sets are (subsets of) a bounded set $\times$ the space without structure.

Theorem *Let $(S, \mathbb{S})$ and $(\tilde{S}, \tilde{\mathbb{S}})$ be measurable spaces, let M 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 one point sets are measurable. Then $\tilde{\mu} = \mu \circ M$ 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.*
- *$M(\emptyset) = \emptyset$.*
- *M is countably set additive: for $A_1, A_2, \dots \in \tilde{\mathbb{S}}$ we have $M(A_1 \cup A_2 \cup \cdots) = M(A_1) \cup M(A_2) \cup \cdots$.*
- *For $a \in \tilde{S}$ we have $\Lambda(M(\{a\})) = 0$.*
- *For $A \in \tilde{\mathbb{S}}$ bounded we have $\Lambda(M(A)) < \infty$.*
- *For $A_1, A_2 \in \tilde{\mathbb{S}}$ and $A_1 \cap A_2 = \emptyset$ we have $\Lambda(M(A_1) \cap M(A_1)) = 0$.*

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

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

We omit the proofs. The hard part in the proof of the theorem is to show that $\tilde{\mu}$ is a counting measure; it is easy to show e.g. σ-additivity almost surely for a fixed sequence of sets $A_1, A_2, \dots$, but the exception set depend on the sequence of sets $A_1, A_2, \dots$.

Note that we allow more general mappings than the inverse image mapping $M = h^{-1}$ for some suitable $h : (S, \mathbb{S}) \to (\tilde{S}, \tilde{\mathbb{S}})$.