Published in final edited form as:

Lifetime Data Anal. 2010 April ; 16(2): 157–175. doi:10.1007/s10985-009-9122-8.

Nonparametric Inference and Uniqueness for Periodically-Observed Progressive Disease Models

Beth Ann Griffin and

RAND Corporation, 1200 South Hayes Street, Arlington, VA 22202, Tel.: +202-558-6879, Fax: +703-413-81111, bethg@rand.org

Stephen W. Lagakos

Department of Biostatistics, Harvard University, Boston, MA 02115

Abstract

In many studies examining the progression of HIV and other chronic diseases, subjects are periodically monitored to assess their progression through disease states. This gives rise to a specific type of panel data which have been termed “chain-of-events data”; e.g. data that result from periodic observation of a progressive disease process whose states occur in a prescribed order and where state transitions are not observable. Using a discrete time semi-Markov model, we develop an algorithm for nonparametric estimation of the distribution functions of sojourn times in a J state progressive disease model. Issues of uniqueness for chain-of-events data are not well-understood. Thus, a main goal of this paper is to determine the uniqueness of the nonparametric estimators of the distribution functions of sojourn times within states. We develop sufficient conditions for uniqueness of the nonparametric maximum likelihood estimator, including situations where some but not all of its components are unique. We illustrate the methods with three examples.

Keywords

Chain-of-events data; nonparametric maximum likelihood estimate (NPMLE); uniqueness; interval censoring; progressive disease model; Time-to-event data

1 Introduction

Chain-of-events data result from periodic monitoring of a progressive disease process whose states occur in a prescribed order and where each observation only provides information about the current state of the process. Progressive disease models with multiple states have been used to describe many different disease processes, especially HIV (De Gruttola and Lagakos, 1989; Gladman and Farewell, 1995; Jackson and Sharples, 2002; Sternberg and Satten, 1999). For example, Feibig *et al* (2003) examined the time course of viremia and antibody seroconversion during primary HIV infection from repeatedly-tested blood donors and characterized early HIV infection as having six distinct stages using this data.

Most statistical research for progressive disease models is based on continuously observed processes. Here, the transition times are known exactly or are subject to right censoring and/or truncation, and extensions of classical survival analysis methods are readily available (Lagakos, *et al.*, 1978; Kalbfleisch and Lawless, 1989; Sun, 2006). In settings where the process is observed only periodically, commonly called panel data, there have been two general approaches. In the first, dependencies among the sojourn times in states are allowed but a Markov assumption is made regarding state transitions (Kalbfleisch and Lawless, 1985; Sun and Kalbfleisch, 1995; Huzurbazar and Huzurbazar, 1999; Longini *et al.*, 1989). In the second,

the sojourn times in states are assumed to be independent. For example, De Gruttola and Lagakos (1989) consider nonparametric methods for a three-state model consisting of an uninfected state, an HIV-infected state prior to development of AIDS, and the state of having AIDS. Here, the independence assumption corresponds to a latency period between HIV-infection and AIDS that does not depend on the age at which the infection occurs. This 3-state process has also been examined by Gomez and Lagakos (1994), Gomez and Calle (1999), and Frydman (1995).

A common feature of methods for continuously-observed processes and interval censored data, including bivariate interval censoring, is that the uniqueness properties of the nonparametric maximum likelihood estimator (NPMLE) are data dependent but not difficult to establish (Kaplan and Meier, 1958; Turnbull, 1976; Gentleman and Geyer, 1994; Betensky and Finkelstein, 1999; Gentleman and Vandal, 2002). For example, for the univariate interval censored data problem, Turnbull (1976) describes intervals that are easily constructed from the observations for which the NPMLE places mass outside the intervals and for which the NPMLE is non-unique within the intervals. In contrast, as noted by De Gruttola and Lagakos (1989), Sternberg and Satten (1999) and others, issues of uniqueness for chain-of-events data are not well understood and cannot easily be characterized. Thus, for example, even when a maximum likelihood estimation algorithm converges to what appears to be a critical value, it is not necessarily clear whether one has reached a global maximum of the likelihood, whether the maximum is unique, and if it is not, how to find the set of maxima. In large part, the additional complexity in understanding the nature of the NPMLE for chain-of-events data arises from the greater number of states as compared to univariate failure time data and to the progressive nature of the underlying process which gives rise to observations that represent intervals of sums of sojourn times within states.

Despite this increased complexity, understanding whether or not the NPMLE is unique for a progressive disease model plays an important role in understanding the natural process of a disease and drawing inferences. For example, when the NPMLE of a distribution is not unique and an analyst does not recognize this to be the case, presentation of just one of many equally likely solutions to the likelihood function could be misleading because it does not represent the range of underlying distributions that are consistent with the observations. These different NPMLEs may show different patterns over time which could, in theory, lead to very different pictures of the biology of the disease (e.g. One solution may show a sharp increase at the beginning of a non-unique time interval and another could show a uniform growth over this time interval). For example, Prasitwattanaseree et al (2004) use Turnbull's (1976) nonparametric interval censored methods to analyze time to DNA polymerase chain reaction (PCR) positivity in infants with in utero/perinatal HIV infection. However, rather than interpreting the NPMLE as only unique in certain time intervals, the authors inappropriately plot the NPMLE as if it is uniquely estimable over the entire time scale and thus provide readers with only one (non-unique) solution to the likelihood equations, incorrectly giving the impression that many in utero/perinatal infections are not detectable within 2 weeks following birth.

The inference methods presented here for estimation can be viewed as generalizations of those developed by De Gruttola and Lagakos (1989) for doubly censored failure time data and which were initially developed by Turnbull (1976) for univariate interval censored failure time data. Sternberg and Satten (1999) have also extended Turnbull's results to progressive disease processes with an arbitrary number of states having a semi-Markov structure and refer to the observations as chain-of-events data. We extend their results by deriving conditions under which the parameters of interest have a unique NPMLE and present an alternative algorithm for estimation of the cumulative distribution functions.

In section 2, we introduce notation and describe the underlying model and observable quantities. In section 3, we introduce an algorithm for estimating the parameters of interests which is an extension of the algorithm described in De Gruttola and Lagakos (1989) for doubly censored survival data. Section 4 describes a series of conditions under which the nonparametric NPMLE will be unique and section 5 illustrates the results with a number of examples.

2 Notation, Observables, and Likelihood Construction

Assume we have a progressive disease model with J states where time spent in each state is a discrete random variable. (See Figure 1 for an illustration of a four-state progressive disease model.) Let X_j denote the chronological time of transition from state j to state $j + 1$ for $j = 1, \dots, J - 1$ and define $X_0 = 0$. Let $T_j = X_j - X_{j-1}$ denote the length of time spent in state j (e.g., the sojourn time for state j) for $j = 1, \dots, J$. Let $F_j(t)$ denote the cumulative distribution function of T_j for $j = 1, \dots, J$ and let the mass points of T_j be denoted $t_{j1}, \dots, t_{jK_j}$ where K_j denotes the total number of mass points for T_j . Define $f_{jk} = \Pr\{T_j = t_{jk}\}$ for $k = 1, \dots, K_j$. Our goal is to estimate $\boldsymbol{\eta} = (f_1, f_2, \dots, f_{J-1})$ where $f_j = (f_{j1}, \dots, f_{jK_j})$ for $j = 1, \dots, J - 1$. We do so assuming that T_j are independent for $j = 1, \dots, J$ and we note that that $\sum_{k=1}^{K_j} f_{jk} = 1$ for $j = 1, \dots, J - 1$.

When subjects undergo periodic visits, X_j will be subject to interval censoring for $j = 1, \dots, J - 1$. Thus, the observed data in our problem consists of N independent and identically distributed copies of $(X_{1L}, X_{1R}, X_{2L}, X_{2R}, \dots, X_{(J-1)L}, X_{(J-1)R})$ where $X_{jL} \leq X_j \leq X_{jR}$ for $j = 1, \dots, J - 1$. In the jargon used by De Gruttola and Lagakos (1989), T_j (the sojourn time for state j) will be doubly censored for each subject for $j = 2, \dots, J - 1$ in so far as neither the origin nor failure times will be observed exactly for these random variables. We assume that censoring acts noninformatively for X_j for $j = 1, \dots, J - 1$ (Oiler *et al.*, 2004). This would occur, for example, if the visit times were scheduled in advance.

Generalizing the approach of De Gruttola and Lagakos (1989), we denote the N copies of $(X_{1L}, X_{1R}, X_{2L}, X_{2R}, \dots, X_{(J-1)L}, X_{(J-1)R})$ with the superscript i and define

$$\alpha_{k_1 \dots k_{J-1}}^i = \begin{cases} 1, & \text{if } X_{1L}^i \leq t_{k_1} \leq X_{1R}^i, X_{2L}^i \leq t_{k_2} \leq X_{2R}^i, \dots, X_{(J-1)L}^i \leq \sum_{l=1}^{J-1} t_{k_l} \leq X_{(J-1)R}^i; \\ 0, & \text{otherwise} \end{cases}$$

as the indicator of whether $(t_{k_1}, t_{k_2}, \dots, t_{k_{J-1}})$ is an admissible value of $(T_1, T_2, \dots, T_{J-1})$ for subject i , given the observation $(X_{1L}^i, X_{1R}^i, X_{2L}^i, X_{2R}^i, \dots, X_{(J-1)L}^i, X_{(J-1)R}^i)$. Let

$\alpha^i = \{\alpha_{k_1 \dots k_{J-1}}^i : 1 \leq k_j \leq K_j \text{ for } j = 1, \dots, J - 1\}$ and note that knowing α^i is equivalent to knowing $(X_{1L}^i, X_{1R}^i, X_{2L}^i, X_{2R}^i, \dots, X_{(J-1)L}^i, X_{(J-1)R}^i)$ for $i = 1, \dots, N$. Thus, we can write the likelihood of the observed data as

$$L_1(\boldsymbol{\eta}) = \prod_{i=1}^N \left(\sum_{k_1 \dots k_{J-1}} \alpha_{k_1 \dots k_{J-1}}^i \prod_{j=1}^{J-1} f_{jk_j} \right). \quad (1)$$

3 Extended DGL Algorithm

Let $K = \sum_{j=1}^{J-1} K_j$. In order to maximize (1) with respect to $\boldsymbol{\eta}$ and subject to the constraints that $0 \leq \eta_c \leq 1$ for $c = 1, \dots, K$, and $\sum_{k=1}^{K_j} f_{jk} = 1$ for $j = 1, \dots, J - 1$, we extend the algorithm presented

in De Gruttola and Lagakos (1989) for the three-state progressive disease model and refer to our algorithm as the extended DGL algorithm.

Define $I_{k_1 \dots k_{J-1}}^i = 1$ if the true, unobserved value of $(T_1, \dots, T_{J-1})$ for the i^{th} subject is $(t_{k_1}, \dots, t_{k_{J-1}})$ and take $I_{k_1 \dots k_{J-1}}^i = 0$ otherwise. Then, the $E[I_{k_1 \dots k_{J-1}}^i | \alpha^i] = \mu_{k_1 \dots k_{J-1}}^i$ where

$$\mu_{k_1 \dots k_{J-1}}^i = \frac{\alpha_{k_1 \dots k_{J-1}}^i \prod_{j=1}^{J-1} f_{jk_j}}{\sum_{b_1 \dots b_{J-1}} \alpha_{b_1 \dots b_{J-1}}^i \prod_{j=1}^{J-1} f_{jb_j}}. \quad (2)$$

Define

$$f_{jc}^* = \sum_{k_1 \dots k_{j-1} c^{k_{j+1} \dots k_{J-1}}} \mu_{k_1 \dots k_{j-1} c^{k_{j+1} \dots k_{J-1}}}^i / N \quad (3)$$

for $c = 1, \dots, K_j$ and $j = 1, \dots, J-1$. Using this notation, we find a solution $(\hat{f}_1, \hat{f}_2, \dots, \hat{f}_{J-1})$ to equations (2)–(3) by using the following iterative scheme.

- i. Choose starting values for $(f_1, f_2, \dots, f_{J-1})$.
- ii. Compute $\mu_{k_1 \dots k_{J-1}}^i$ from (2).
- iii. Compute refined estimates of f_j for $j = 1, \dots, J-1$ using equation (3).
- iv. Repeat steps (ii) and (iii) until convergence.

This algorithm corresponds to that developed in Sternberg and Satten (1999) for the J state model. Sternberg and Satten (1999) derive their methods by extending Turnbull's (1976) self-consistency algorithm directly. Sternberg and Satten (1999) estimate η using a self-consistent estimation approach. They first define self-consistent distribution functions for the density functions f_j and their corresponding survival distributions conditional on the observed data and then iterate between these equations and the equation for the unconditional f_j to find the self-consistent solution for η . We find that the extended DGL algorithm presented above is more intuitive and computationally straight-forward and thus chose to proceed using this method as our basis.

When the extended DGL algorithm converges to a critical value that is not a saddle point, it may be a strict or nonstrict local or global maximum of the likelihood function and it may or may not be unique. We denote this critical value by $\hat{\eta} = (\hat{f}_1, \hat{f}_2, \dots, \hat{f}_{J-1})$ and note that estimates for $F_j(t)$ for $j = 1, \dots, J-1$ follow directly. In the ensuing sections, we focus on the nature of the solutions space of the likelihood function $L_1(\eta)$.

4 Uniqueness of NPMLE

In this section we develop methods for determining when the maximizing solution, say $\hat{\eta}$, of $L_1(\eta)$, is unique, including situations where some but not all of its components are unique. To establish whether $\hat{\eta}$ or some of its components are unique, we proceed in three steps, based on different types of reparameterizations of η . For each, we establish conditions for the uniqueness of $\hat{\eta}$ or some of its components and indicate when these might be expected to hold and how these can be assessed in practice. We end this section with a summary of the proposed techniques and the steps which we recommend be taken to determine if $\hat{\eta}$ or some of its

components are unique in practice. We note that the reparametrized likelihood functions given below are not used for estimation purposes but instead provide us with verifiable conditions that can be used to determine if $\hat{\boldsymbol{\eta}}$ is unique.

To proceed, we define the parameter space of $\boldsymbol{\eta}$ as

$$\mathbf{H} = \{\boldsymbol{\eta} = (\eta_1, \dots, \eta_K) : \sum_{a=1}^{K_1} \eta_a = \sum_{a=K_1+1}^{K_1+K_2} \eta_a = \dots = \sum_{a=K_1+\dots+K_{J-2}+1}^{K_1+\dots+K_{J-1}} \eta_a = 1, \eta_a \in [0, 1]\}. \quad (4)$$

We show (see Appendix B) that $L_1(\boldsymbol{\eta})$ is concave over $\mathbf{H}$, though not necessarily strictly concave. Thus, if the extended DGL algorithm converges to a maximum, it must be a global maximum. If the maximum is strict, then the NPMLE, $\hat{\boldsymbol{\eta}}$, of $\boldsymbol{\eta}$ is unique. Identifying a strict maximum of $L_1(\boldsymbol{\eta})$ in practice can be difficult because $\dim\{\mathbf{H}\}$ is typically very large, a maximizing solution can be on the boundary of $\mathbf{H}$, and the distribution of the observables are functions of products of the components of $\boldsymbol{\eta}$, such as $(f_{13}f_{24} \dots f_{(J-1)6})$. In addition, in some settings, only some of the components of the NPMLE, $\hat{\boldsymbol{\eta}}$, will be unique. Throughout this section, we assume that the extended DGL algorithm has converged to a maximum, namely $\hat{\boldsymbol{\eta}}$.

4.1 Reparametrization 1: Using the Joint Distribution of $(T_1, T_2, \dots, T_{J-1})$

Let $\boldsymbol{\theta}$ denote the $(p \times 1)$ vector of all the possible combinations $(f_{1k_1}f_{2k_2} \dots f_{(J-1)k_{J-1}})$ where $p = K_1 K_2 \dots K_{J-1}$. We define $\boldsymbol{\theta}$ so that the entries first change over f_{1k_1} then over f_{2k_2} and so on so to finally change over $f_{(J-1)k_{J-1}}$. Define $\boldsymbol{\Theta}$ to be the set of all $(\theta_1, \dots, \theta_p)$ such that $0 \leq \theta_a \leq 1$ for $a = 1, \dots, p$ and for which $\sum_{a=1}^p \theta_a = 1$.

We can rewrite the likelihood of our observed data in (1) as

$$L_2(\boldsymbol{\theta}) = \prod_{i=1}^N \left(\sum_{a=1}^p B_{ia} \theta_a \right) \quad (5)$$

where B_{ia} is an indicator of whether the i^{th} subject contributes the a^{th} combination of $(f_{1k_1}f_{2k_2} \dots f_{(J-1)k_{J-1}})$ to the likelihood (e.g. has $(t_{1k_1}, t_{2k_2}, \dots, t_{(J-1)k_{J-1}})$ as an admissible value). Let B denote the $N \times p$ matrix with elements (B_{ia}) .

Consider the function $\mathbf{g} : \boldsymbol{\eta} \rightarrow \boldsymbol{\theta}$ and define $\boldsymbol{\Theta}_g = \{\mathbf{g}(\boldsymbol{\eta}) : \boldsymbol{\eta} \in \mathbf{H}\}$ to be the range of $\mathbf{g}$. Unlike with univariate interval censored problems, $\boldsymbol{\Theta}_g$ is a strict subset of $\boldsymbol{\Theta}$. Thus, the maximum likelihood estimator, $\hat{\boldsymbol{\theta}}$, of $\boldsymbol{\theta}$ over $\boldsymbol{\Theta}$ need not maximize $L_2(\boldsymbol{\theta})$ over the space $\boldsymbol{\Theta}_g$. It is important to note the difference between the maximizing solution, $\hat{\boldsymbol{\theta}}$, over $\boldsymbol{\Theta}$ and the NPMLE, $\hat{\boldsymbol{\theta}}_g$, over $\boldsymbol{\Theta}_g$. $\hat{\boldsymbol{\theta}}_g = \mathbf{g}(\hat{\boldsymbol{\eta}})$ and thus, since $\mathbf{g}$ is one-to-one, $\hat{\boldsymbol{\eta}}$ is unique if $\hat{\boldsymbol{\theta}}_g$ is unique. We define the larger rectangular space, $\boldsymbol{\Theta}$, to allow us to utilize some existing results in the process of assessing the uniqueness of $\hat{\boldsymbol{\eta}}$.

Theorem 1: If $L_2(\boldsymbol{\theta})$ is strictly concave over $\boldsymbol{\Theta}$, then $\hat{\boldsymbol{\eta}}$ is unique.

The proof of Theorem 1 is given in Appendix A. Theorem 1 shows that if $L_2(\boldsymbol{\theta})$ is strictly concave over $\boldsymbol{\Theta}$ and the extended DGL algorithm converges to a maximum then it must be the unique global maximum. $L_2(\boldsymbol{\theta})$ will be strictly concave when its Hessian matrix is negative definite. Gentleman and Geyer (1994) show that the Hessian matrix for $L_2(\boldsymbol{\theta})$ can be written as $B'DB$ where D is a diagonal matrix with negative entries. Thus, when $N \geq p$, the Hessian

matrix will be negative definite if and only if B is of full rank. When B is full rank, $\hat{\theta}$ and $\hat{\theta}_g$ are necessarily unique.

In practice, the conditions of Theorem 1 will be met for a sufficiently large sample size if, with positive probability, a visit can occur between all possible values of $t_{1k_1}, t_{1k_1} + t_{2k_2}, \dots, t_{1k_1} + t_{2k_2} + \dots + t_{(J-1)k_{J-1}}$ for $k_j = 1, \dots, K_j$ and $j = 1, \dots, J-1$. As $N \rightarrow \infty$, every possible admissible value for $(T_1, T_2, \dots, T_{J-1})$ and its corresponding combination $(f_{1k_1} f_{2k_2} \dots f_{(J-1)k_{J-1}})$ will be observed exactly making $\text{rank}(B) \rightarrow p$. Theorem 1 is a sufficient but not necessary condition for $\hat{\eta}$ to be unique and is introduced primarily because its conditions are straight-forward to assess.

4.2 Reparametrization 2: Using a Subset of the Joint Distribution of $(T_1, T_2, \dots, T_{J-1})$

The conditions of Theorem 1 may not hold in settings where p is large, even with large N , since for each of the p combinations $(f_{1k_1} f_{2k_2} \dots f_{(J-1)k_{J-1}})$, there must exist at least one subject whose likelihood contribution contains that combination for B to be of full column rank. We now consider the situation when some of the elements (combinations) of θ do not appear in the likelihood because none of the subjects have the corresponding $(t_{1k_1}, t_{2k_2}, \dots, t_{(J-1)k_{J-1}})$ admissible value for $(T_1, T_2, \dots, T_{J-1})$. Combinations that do not appear in the likelihood correspond to the columns in B which have only zero entries.

Let B^* denote the $N \times p^*$ matrix resulting from eliminating the zero columns of B , and let θ^* denote the corresponding p^* vector of combinations. Define Θ^* to be the set of $(\theta_1^*, \dots, \theta_{p^*}^*)$ such that $0 \leq \theta_a^* \leq 1$ for $a = 1, \dots, p^*$ and for which $\sum_{a=1}^{p^*} \theta_a^* \leq 1$. The likelihood in (5) can now be written

$$L_2^*(\theta^*) = \prod_{i=1}^N \left(\sum_{a=1}^{p^*} B_{ia}^* \theta_a^* \right). \quad (6)$$

As in section 4.1, we consider the function $g^* : \eta \rightarrow \theta^*$ and define $\Theta_{g^*} = \{g^*(\eta) : \eta \in H\} \subset \Theta^*$. Again, the maximizing solution, say $\hat{\theta}^*$, of θ^* over Θ^* need not be the NPMLE of $L_2^*(\theta^*)$ over Θ_{g^*} . Unless all the combinations of θ are unique, we cannot directly establish that η will have a unique maximum simply by showing $L_2^*(\theta^*)$ is strictly concave. In order to guarantee $L_1(\eta)$ has a unique maximum over H , we need the additional condition that the mapping $g^* : \eta \rightarrow \theta^*$ is one-to-one. We have the following result.

Theorem 2: If $L_2^*(\theta^*)$ is strictly concave over Θ^* and $g^* : \eta \rightarrow \theta^*$ is one-to-one, then $\hat{\eta}$ is unique.

The proof of Theorem 2 follows directly from the proof of Theorem 1 given in Appendix A. As with Theorem 1, establishing whether $L_2^*(\theta^*)$ is strictly concave can be done by checking if B^* is of full rank when $N \geq p^*$ (Gentleman and Geyer, 1994). The more difficult challenge in verifying whether Theorem 2 can be applied in a particular setting is determining whether g^* is one-to-one as this will depend on the specific observation patterns.

Below, we provide two lemmas that provide sufficient conditions for $g^* : \eta \rightarrow \theta^*$ to be one-to-one. When g^* is one-to-one and B^* is of full rank, $\hat{\eta}$ is unique. When g^* is not one-to-one but B^* is of full rank, a subset of the components of $\hat{\eta}$ may still be unique. Definition A and Theorem A below are introduced to distinguish between cases where all of $\hat{\eta}$ or only some of its components are unique.

Definition A: For any vector-valued function $\mathbf{g}^*$ with domain $\mathbf{H}$ and range $\Delta \subseteq E^k$, define the partition of $\mathbf{H}$ into the equivalence sets S_λ , indexed by $\lambda \in \Lambda$, such that $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \in S_\lambda \Rightarrow \mathbf{g}^*(\boldsymbol{\eta}_1) = \mathbf{g}^*(\boldsymbol{\eta}_2)$, and $\boldsymbol{\eta}_1 \in S_\lambda, \boldsymbol{\eta}_2 \in S_{\lambda'}, \text{ where } \lambda \neq \lambda' \Rightarrow \mathbf{g}^*(\boldsymbol{\eta}_1) \neq \mathbf{g}^*(\boldsymbol{\eta}_2)$. That is, the S_λ denote the sets in $\mathbf{H}$ that map to distinct points in Δ . In addition, partition Δ into $\{\Delta_0, \Delta_0^c\}$, where Δ_0 consists of the elements of Δ for which the corresponding sets S_λ are singletons. Thus, when $\Delta_0 = \Delta$, S_λ is a singleton for all $\lambda \in \Lambda$, and $\mathbf{g}^*(\boldsymbol{\eta}) = \boldsymbol{\delta}$ is a one-to-one mapping.

Theorem A: Define $\mathcal{C} \subseteq \{\emptyset, 1, 2, \dots, K\}$ to be the set whose elements satisfy $c \in \mathcal{C}$ if and only if for every $\lambda \in \Lambda, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \in S_\lambda \Rightarrow \boldsymbol{\eta}_{1c} = \boldsymbol{\eta}_{2c}$ where $\boldsymbol{\eta}_{ic}$ denotes the c^{th} component of $\boldsymbol{\eta}_i$. When $\mathcal{C} = \{1, 2, \dots, K\}$, each S_λ is a singleton and $\mathbf{g}^*$ is one-to-one.

When $\mathbf{H}$ and Δ represent the domains of likelihood functions, say $L_1(\boldsymbol{\eta})$ and $L_0(\boldsymbol{\delta}) = L_0[\mathbf{g}^*(\boldsymbol{\eta})]$, it follows that when the NPMLE of $\boldsymbol{\delta}$ is unique, the components of $\boldsymbol{\eta}$ corresponding to the elements in $\mathcal{C}$ will also have unique maximum likelihood estimates. In addition, when Δ_0 is a strict subset of Δ and $\mathbf{g}^*$ is not one-to-one, the uniqueness of $\hat{\boldsymbol{\eta}}$ will depend on whether $\hat{\boldsymbol{\delta}}$ is in Δ_0 or not. Below we describe ways to identify the components of $\mathcal{C}$ as well as ways to determine when $\hat{\boldsymbol{\delta}} = \mathbf{g}^*(\hat{\boldsymbol{\eta}}) \in \Delta_0$.

Note we can write $\boldsymbol{\eta} = A\boldsymbol{\theta}$ where A is the $K_1 + K_2 + \dots + K_{J-1} \times p$ matrix of zeros and ones indicating how to obtain an element of $\boldsymbol{\eta}$ by summing elements of $\boldsymbol{\theta}$. For example,

$$f_{ab} = \sum_{j=1; j \neq a}^{J-1} \sum_{k_j=1; k_a \neq b}^{K_j} f_{1k_1} f_{2k_2} \dots f_{ab} \dots f_{(j-1)k_{j-1}} \text{ since } \sum_{k_j=1}^{K_j} f_{jk_j} = 1 \text{ for } j = 1, \dots, J-1. \text{ Let } A_{(b)} \text{ denote the } b_{th} \text{ row of the matrix } A.$$

Lemma 2.1: If for some $b \in \{1, 2, \dots, K\}$, $A_{(b)}\boldsymbol{\theta}$ is a sum of the elements in $\boldsymbol{\theta}^*$ and if the elements of $\boldsymbol{\theta}^*$ include all the elements in $\boldsymbol{\eta}$, then $\mathbf{g}^* : \boldsymbol{\eta} \rightarrow \boldsymbol{\theta}^*$ is one-to-one.

Lemma 2.2: If at least $J-2$ of the combinations $(f_{1k_1} \dots f_{(j-1)k_{j-1}} f_{(j+1)k_{j+1}} \dots f_{(J-1)k_{J-1}})$ for different j 's are sums of the elements in $\boldsymbol{\theta}^*$ and if the elements in $\boldsymbol{\theta}^*$ include all the elements in $\boldsymbol{\eta}$, then $\mathbf{g}^* : \boldsymbol{\eta} \rightarrow \boldsymbol{\theta}^*$ is one-to-one.

The proofs for Lemma 2.1 and 2.2 are given in section A of the supplementary materials. When $L_2^*(\boldsymbol{\theta}^*)$ is strictly concave and either Lemma 2.1 or Lemma 2.2 holds, $\hat{\boldsymbol{\eta}}$ will be unique. If neither Lemma 2.1 nor Lemma 2.2 hold, $\mathbf{g}^*$ can still be shown to be one-to-one by showing that $\mathcal{C} = \{1, 2, \dots, K\}$ and invoking Theorem A. An algorithm is available in Table 1 of the supplementary material to help determine which components of $\hat{\boldsymbol{\eta}}$ are elements of $\mathcal{C}$ and hence are unique by applying the Lemmas and Theorems given above. If B^* is full rank and the extended DGL algorithm has converged to a local maximum, say $\hat{\boldsymbol{\eta}}$, the subset of elements in $\boldsymbol{\eta}$ for which $c \in \mathcal{C}$ will be uniquely estimated.

Both Theorem 1 and Theorem 2 tend to hold in data sets where subjects have a common visit schedule and where some individual visits may be missed, but where few drop out, causing right censoring. When this is not the case and neither B nor B^* are of full rank, $\boldsymbol{\eta}$ can still have a unique NPMLE.

4.3 Reparametrization 3: Using the Unique Observations in the Data

In this section, we reparametrize the likelihood function in terms of the number of possible unique observations $(X_{1L}, X_{1R}, X_{2L}, X_{2R}, \dots, X_{(J-1)L}, Y_{(J-1)R})$ that occur in the data. This can range from 1 (e.g. If everyone has the same observed interval censoring pattern) to N (e.g. if everyone in the data has a different observed interval censoring pattern). We have

$$L_3(\boldsymbol{\varphi}) = \prod_{u=1}^U \varphi_u^{n_u} \quad (7)$$

where u indexes distinct observation patterns, φ_u is the probability of observation pattern u , and n_u denotes the number of subjects with observation pattern u . Let $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_U)$. Note that each φ_u is simply a sum of elements in $\boldsymbol{\theta}$, and $\boldsymbol{\varphi} = M\boldsymbol{\theta}$ where M is matrix of zeros and ones indicating which elements of $\boldsymbol{\theta}$ are included in each $\boldsymbol{\varphi}$.

Theorem 3: If $\mathbf{g}^\varphi : \boldsymbol{\eta} \rightarrow \boldsymbol{\varphi}$ is one-to-one, then $\hat{\boldsymbol{\eta}}$ is unique.

Since $L_3(\boldsymbol{\varphi})$ is strictly concave, the proof for Theorem 3 follows directly from the proofs of Theorems 1 and 2. As with Theorem 2, the challenge in applying Theorem 3 is establishing whether $\mathbf{g}^\varphi : \boldsymbol{\eta} \rightarrow \boldsymbol{\varphi}$ is one-to-one. Definition A and Theorem A are again useful for this purpose.

The vector $\boldsymbol{\varphi}$ yields a nonlinear system of U equations in $K - 3$ unknowns since

$\sum_{k_j=1}^{K_j} f_{jk_j} = 1$ for $j = 1, \dots, J - 1$. When U is small, the system of equations may be straightforward to solve for the elements in $\boldsymbol{\eta}$ and thereby establish if $\mathbf{g}^\varphi : \boldsymbol{\eta} \rightarrow \boldsymbol{\varphi}$ is one-to-one. We provide an example below to help illustrate cases where the system of equations can be solved directly and what implications the solutions have in regards to the uniqueness of $\hat{\boldsymbol{\eta}}$.

To illustrate the nuisances that can arise when solving a system of equations $\boldsymbol{\varphi}$ for $\boldsymbol{\eta}$, suppose we have a three state progressive disease model such that T_1 has mass points at 1, 2, and 3 and T_2 and T_3 have mass points at 1 and 2. Thus, $\boldsymbol{\eta} = (f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{31}, f_{32})$. Suppose next that the observations arise from subject visits at times 2.5, 4.5, 6.5 and 8.5. From the 12 possible values of (X, T, C) and this fixed visit pattern, we obtain six unique observation patterns $(X_{1L}, X_{1R}, X_{2L}, X_{2R}, X_{3L}, X_{3R})$:

$$(X_{1L}, X_{1R}, X_{2L}, X_{2R}, X_{3L}, X_{3R}) = \begin{cases} (1, 2, 1, 2, 3, 4), & \text{if } (T_1, T_2, T_3) = (1, 1, 1) \text{ or } (1, 1, 2); \\ (1, 2, 3, 4, 3, 4), & \text{if } (T_1, T_2, T_3) = (1, 2, 1) \text{ or } (2, 1, 1); \\ (1, 2, 3, 4, 5, 6), & \text{if } (T_1, T_2, T_3) = (1, 2, 2), (2, 1, 2), (2, 2, 1) \text{ or } (2, 2, 2); \\ (3, 4, 3, 4, 5, 6), & \text{if } (T_1, T_2, T_3) = (3, 1, 1) \text{ or } (3, 1, 2); \\ (3, 4, 5, 6, 5, 6), & \text{if } (T_1, T_2, T_3) = (3, 2, 1); \\ (3, 4, 5, 6, 7, 8), & \text{if } (T_1, T_2, T_3) = (3, 2, 2). \end{cases}$$

Thus, the likelihood can be written

$$L_3(\boldsymbol{\varphi}) = \prod_{u=1}^6 \varphi_u^{n_u}$$

Where

$$\begin{aligned} \varphi_1 &= f_{11}f_{21}f_{31} + f_{11}f_{21}f_{32} \\ \varphi_2 &= f_{11}f_{22}f_{31} + f_{12}f_{21}f_{31} \\ \varphi_3 &= f_{11}f_{22}f_{32} + f_{12}f_{21}f_{32} + f_{12}f_{22}f_{31} + f_{12}f_{22}f_{32} \\ \varphi_4 &= f_{13}f_{21}f_{31} + f_{13}f_{21}f_{32} \\ \varphi_5 &= f_{13}f_{22}f_{31} \\ \varphi_6 &= f_{13}f_{22}f_{32} \end{aligned}$$

For any N , B will have rank at most 6, since the rank of B will equal the number of unique, linearly independent observation patterns. For sufficient N , the six unique, linearly independent rows of B are

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The conditions of Theorem 1 will not hold here since B is not of full column rank. However, $\hat{\eta}$ is still unique since the condition of Theorem 3 holds. One can easily solve the system of equations ϕ given above for each component of η , thereby showing that g^ϕ is one-to-one.

In most cases, solving the system of nonlinear equations ϕ will not be simple. Table 2 in the supplementary material provides an algorithm that can be helpful in identifying the elements of $\mathcal{C}$ and thus, the corresponding unique components of $\hat{\eta}$ in more complex situations than illustrated above. It works by exploiting right-censoring in a dataset to simplify the system of equations being solved and isolate pieces of the NPMLE that are uniquely estimable. If the algorithm finds that $\mathcal{C} = \{1, 2, \dots, K\}$, then g^ϕ is one-to-one and $\hat{\eta}$ is unique. The algorithm is not useful for data sets with a small number of right-censored observations.

Until now, we have focused on identifying elements of $\mathcal{C}$ which can be used to determine the unique components of $\hat{\eta}$. When $\mathcal{C} = \{1, 2, \dots, K\}$, g^ϕ is a one-to-one mapping and the previous results can be used to show that all of the components of $\hat{\eta}$ are unique. However, it is possible that $\hat{\eta}$ will be unique even when $g^\phi(\eta)$ is not a one-to-one mapping, or equivalently, when $\mathcal{C} \neq \{1, 2, \dots, K\}$. Specifically, when the NPMLE, $\hat{\phi} = g^\phi(\hat{\eta})$, of $L_3(\phi)$ satisfies $\hat{\phi} \in \Delta_0$ (see Definition A), $\hat{\eta}$ is unique regardless of whether $g^\phi(\eta)$ is one-to-one. Below we indicate how the uniqueness of $\hat{\eta}$ may be established in these situations.

4.4 Summary of Uniqueness Results

To implement the preceding results in practice, we recommend the following steps to determine if $\hat{\eta}$ is unique:

1. Check if B from (5) is full rank. If so, $L_2(\theta)$ is strictly concave, the condition of Theorem 1 holds, and $\hat{\eta}$ is unique. Otherwise, proceed to step 2.
2. Check if B^* from (6) is full rank and if $g^*: \eta \rightarrow \theta^*$ is one-to-one. If so, the conditions of Theorem 2 hold and $\hat{\eta}$ is unique. Otherwise, proceed to step 3.
3. Check if $g^\phi: \eta \rightarrow \phi$ is one-to-one. If so, the conditions of Theorem 3 holds and $\hat{\eta}$ is unique.

In cases where neither Theorem 1 nor Theorem 2 hold and it cannot be established that $g^\phi: \eta \rightarrow \phi$ is one-to-one, $\hat{\eta}$ may still be unique.

Theorem 4: If $\hat{\eta}$ is a strict local maximum of $L_1(\eta)$, then $\hat{\eta}$ is the unique.

The proof of Theorem 4 is given in Appendix B. Unfortunately, the condition of Theorem 4 can be difficult to verify in practice without a thorough grid search in the neighborhood of $\hat{\eta}$. We illustrate below when Theorem 4 is useful in practice.

Even in cases where $\hat{\eta}$ is not unique, it is likely that there will be a subset of elements in η which have unique maximum likelihood estimates. The algorithms in Tables 1 and 2 of the supplementary material can be used to form this subset of elements. As with univariate interval censored data (Turnbull 1976), the algorithms determine which regions of the distribution functions for $T_1, T_2, \dots, T_{J-1}$ are not unique and are therefore useful results in establishing the uniqueness properties for a given chain-of-events data set. However, every data set will be different making it difficult to find one solution that helps in all cases.

Finally, let $t_1^* = \max_i \{X_{1R}^i : X_{1R}^i < \infty\}$ and $t_j^* = \max_i \{X_{jR}^i - X_{(j-1)L}^i : X_{jR}^i < \infty\}$ for $j = 2, \dots, J-1$. As with typical right censored univariate data, $\hat{F}_1(t)$ will place all of its mass at values of t_{1k} no greater than t_1^* provided $\max_i \{X_{1L}^i : X_{1L}^i = \infty\} \leq t_1^*$ and $\hat{F}_j(t)$ will place all of its mass at values of t_{jk} no greater than t_j^* provided $\max_i \{X_{jL}^i - X_{(j-1)R}^i : X_{1R}^i < \infty, \dots, X_{(j-1)R}^i < \infty, X_{jR}^i = \infty\} \leq t_j^*$ for $j = 2, \dots, J-1$ (De Gruttola and Lagakos, 1989). Otherwise, $\hat{F}_1(t)$ will not be uniquely estimable for values $t_{1k} \geq \max_i \{X_{1L}^i : X_{1L}^i = \infty\}$ and $\hat{F}_j(t)$ for $t_{jk} \geq \max_i \{X_{jL}^i - X_{(j-1)L}^i : X_{1R}^i < \infty, \dots, X_{(j-1)R}^i < \infty, X_{jR}^i = \infty\}$ for $j = 2, \dots, J-1$.

5 Examples

In this section we present four examples to illustrate how the results in section 4 might be used in practice. The examples were chosen to illustrate the different methods that can be used to show $\hat{\eta}$ or a subset of its components are unique, based on the proposed strategy given in section 4.4. The first example is based on simulated data while the second and third are based on real data.

5.1 Example 1: Simulated Data

Suppose we have a four state progressive disease model with T_1, T_2 , and T_3 each having mass points at 1, 2, 3, 4, and 5 months and $\eta_c = 1/5$ for $c = 1, 2, \dots, 15$. Assume that subjects undergo monthly visits with probability v ($0 < v \leq 1$). Thus, subjects may have a different number of visits but each will be contained in $\{1, 2, \dots\}$. When $N = 900$ and $v = 1$, the B matrix from (5) of our simulated data set was a 900×125 matrix and the condition of Theorem 1 held. Thus, $\hat{\eta}$ was unique. We also examined cases where $v = .9$ and $N = 600$, $v = .75$ and $N = 400$, and $v = .5$ and $N = 300$. For each, B was of full column rank and the condition of Theorem 1 held. Thus, $\hat{\eta}$ was unique for each case.

Next, we considered the same model but with $v = 1$ and $N = 500, 100, 75$, and 50, respectively. With smaller values of N , the condition of Theorem 1 will not hold since B will no longer be of full column rank due to having columns with all zero entries. Here, we focus on using Theorem 2 to show that $\hat{\eta}$ is unique. Details can be found in section C of the supplementary material. For each case, B^* from (6) was of full rank and thus, only the one-to-oneness of $g^* : \eta \rightarrow \theta^*$ had to be established. When $N = 500$ and $N = 100$, the conditions of Lemma 2.1 and 2.2 held, respectively. When $N = 75$, the algorithm in Table 1 was used to establish that g^* was one-to-one. Finally, when $N = 50$, g^* could not be shown to be one-to-one and Theorems 1 – 3 failed to demonstrate that $\hat{\eta}$ was unique. Here, Theorem 4 was easily invoked to prove that $\hat{\eta}$ was still unique.

5.2 Example 2: FEV1 in Patients with Bronchiolitis Obliterans Syndrome

Jackson and Sharples (2002) present results from a study of Bronchiolitis obliterans syndrome (BOS) in patients who had a lung transplant. The data consist of longitudinal measurements on the forced expiratory volume in one second (FEV1) from the transplant recipients, from six months after their transplant onwards. BOS is denned as the clinical syndrome of irreversible, progressive airway obstruction caused by the presence of Obliterative bronchiolitis in lung

transplant recipients and is diagnosed based on FEV1 measurements. Here, the four-state progressive disease model is such that subjects are BOS free during the first state and then travel through three progressive states of BOS which are clinically defined by having FEV1 measurements below 80%, 65%, and 50% of the baseline FEV1 measurement, respectively. FEV1 measurements were scheduled to be taken every three to six months and there are a total of $N = 203$ subjects. The visit schedules vary greatly among patients.

We let X_1 , X_2 , and X_3 denote time until first FEV1 measurement less than or equal to 80%, 65%, and 50% of the baseline FEV1 measurement, respectively. We assume that T_1 , T_2 , and T_3 have mass points every six months such that the points $t_{10}, t_{11}, t_{12}, \dots$ denote time 0, 6, 12, ... months where time 0 is taken to be 182 days after lung transplant and the point from which subjects are at risk for BOS. We discretize the data into six month intervals such that $(X_{1L}, X_{1R}, X_{2L}, X_{2R}, X_{3L}, X_{3R})$ give us closed intervals on X_1 , X_2 and X_3 and assume that when transition times between states occurred at two consecutive visits within a six month interval, the observation is assigned to the left-hand endpoint of the interval. $\hat{F}_1(t)$ does not place all of its mass at points for $t_{1k} \leq 25 = t_1^*$ since $t_1^* > \max_i \{X_{1R}^i : X_{1R}^i < \infty\} = 11$. In fact, $\hat{F}_1(x)$ cannot be uniquely estimated for points $t_{1k} \geq 25$. For computational purposes, we let T_1 take on mass points between 0 and 25 and note that $f_{1,25}$ denotes the probability that $T_1 \geq 25$. Since $\hat{F}_2(t)$ places all of its mass at points t_{2k} for which $t_{2k} \leq 14 = t_2^*$ and $\hat{F}_3(t)$ places all of its mass at point t_{3k} for which $t_{3k} \leq t_3^* = 17$, we took $\boldsymbol{\eta} = (f_{10}, \dots, f_{1,25}, f_{20}, \dots, f_{2,14}, f_{30}, \dots, f_{3,17})$.

In this example, B from (5) is a 203×7020 matrix and the conditions for Theorems 1 and 2 do not hold, so we focus on Theorem 3 and the algorithm in Table 2 of the supplementary material for solving the system of equations created by the data. We find that $\mathcal{C}$ contains the subscripts corresponding to the elements $\{f_{10}, \dots, f_{1,15}, f_{1,18}, f_{1,19}, f_{1,25}, f_{20}, \dots, f_{2,9}, f_{30}, \dots, f_{3,11}\}$. We need to interpret the uniqueness of the components of $\hat{\boldsymbol{\eta}}$ by examining $\mathcal{C}$ together with the value of $\hat{\boldsymbol{\eta}}$ obtained from the extended DGL algorithm. Immediately, we know that $\hat{\eta}_c$ for $c \in \mathcal{C}$ are unique. In addition, we find that $\hat{f}_{1,16} = \hat{f}_{1,17} = \hat{f}_{1,20} = \dots = \hat{f}_{1,24} = 0$ since $\hat{f}_{1,16} + \hat{f}_{1,17} + \dots + \hat{f}_{1,24}$ are uniquely estimated to be zero and that $\hat{f}_{2,10} = \hat{f}_{2,11} = \hat{f}_{2,12} = \hat{f}_{2,13} = \hat{f}_{2,15} = \hat{f}_{2,16} = 0$ for similar reasons. Thus, these elements are uniquely estimated despite not being contained in $\mathcal{C}$. Lastly, we find that $\hat{f}_{3,12} + \hat{f}_{3,13} + \hat{f}_{3,14} = 0.044$ is uniquely estimated to have nonzero mass. Thus, $\hat{f}_{3,k}$ for $k = 12, 13, 14$ are not unique here. There are an infinite number of solutions for $\hat{f}_{3,12}, \hat{f}_{3,13}$, and $\hat{f}_{3,14}$ for which $\hat{f}_{3,12} + \hat{f}_{3,13} + \hat{f}_{3,14} = 0.044$. Thus, $\hat{\boldsymbol{\eta}}$ is not unique in this example.

Note the importance of t_1^* in this data. Since $t_1^* = 25 < 11 = \max_i \{X_{1R}^i : X_{1R}^i < \infty\}$, we cannot uniquely estimate $\hat{F}_1(t)$ for $t_{1k} \geq 25$. While the algorithm shows that $\hat{f}_{1,25}$ is unique, this represents the probability that $T_1 \geq 25$. Figure 2 shows the estimated cumulative distribution functions for T_1 , T_2 and T_3 in the FEV1 data set. The blank regions in the top graph for $T_1 \geq 25$ and in the bottom graph for $12 \leq T_3 < 14$ reflect the fact that $\hat{F}_1(t)$ and $\hat{F}_3(t)$ are not unique in these regions, respectively.

5.3 Example 3: Primary HIV Infection

Feibig *et al.* 2003 present results from a retrospective study involving archived samples from HIV infected plasma donors. The data set contains 435 samples from 51 donors detected by anti-HIV enzyme immunoassays donated between 1984 and 1994. In their paper, they show that primary HIV infection can be characterized by as many as six states. For the sake of illustration, we assume there are three states of primary infection which we refer to as follows: acute infection corresponding to Feibig states 1 and 2, recent infection corresponding to Feibig states 3 and 4, and early infection corresponding to Feibig states 5 and 6. Thus, our four state progressive disease model begins with an uninfected state, followed by the acute infection, recent infection, and early infection states. The median interval between samples was 5 days

(range 1 to 204 days). The time of the last visit in which a subject was uninfected was not available to us; instead, we set each person to have an uninfected visit 14 days prior to their first visit with a positive test result and do not draw any inferences about the distribution of time spent in the uninfected state for this sample.

We let X_1 , X_2 , and X_3 denote time until acute infection, recent infection, and early infection, respectively. We assume that T_1 , T_2 and T_3 have mass points every week such that the points $t_{10}, t_{11}, t_{12}, \dots$ denote time 0, 1, 2, ... weeks where time 0 is taken to be time of first sample donation. We discretize the data into one week intervals such that $(X_{1L}, X_{1R}, X_{2L}, X_{2R}, X_{3L}, X_{3R})$ give us closed intervals on X_1 , X_2 , and X_3 and, as above, we assume that when transition times between states occurred at two consecutive visits within a one week interval, the observation is assigned to the left-hand endpoint of the interval. We take $K_j = 20$ for $j = 1, 2, 3$, and without loss of generality let $\boldsymbol{\eta} = (f_{10}, f_{11}, \dots, f_{1,20}, f_{20}, f_{21}, \dots, f_{2,20}, f_{30}, f_{31}, \dots, f_{3,20})$ where $f_{jk} = \Pr\{T_j = k\}$ for $j = 1, \dots, 3$. Due to a small number of right-censored observations in the data, we found that $\hat{F}_j(t)$ places all of its mass at points $t_{jk} \leq 20$ for $j = 1, 2$ and 3.

In this example, B from (5) is a 51×9261 matrix and the conditions for Theorems 1 and 2 do not hold, so we next focus on Theorem 3. Unfortunately, there are only 4 right-censored observations in the Feibig data making the algorithm in Table 2 of the supplementary material unapplicable here. Careful examination of the system of equations in this example did not reveal any components of $\boldsymbol{\eta}$ that could be easily isolated and the formation of $\mathcal{C}$ is therefore not straightforward.

Figure 3 shows the estimated cumulative distribution functions for T_2 and T_3 in the data set. A grid search was performed and multiple starting values were used to help inform whether or not the estimated cumulative distribution functions shown are unique. As shown, $\boldsymbol{\eta}$ is not a strict maximum since there is a clear set of values for which the NPMLE is not unique; in particular, f_{3k} for $k = 3, \dots, 16$ are not unique (e.g. $\hat{F}_3(t)$ was not uniquely estimable between days 21 and 112). We are unable to make a formal proof in this case as with our examples above but the other elements of $\boldsymbol{\eta}$ appear to be uniquely estimated based on our analyses.

Figure 3 also shows how the distribution functions for acute and recent HIV infection are well fit by exponential distributions, which were assumed in Feibig *et al* 2003.

6 Discussion

The methods presented here focus on estimation and uniqueness of the NPMLE for the distributions of sojourn times in a J state progressive disease model when transition times between states are interval censored. The extended DGL algorithm is a computational straightforward technique for obtaining the NPMLE and the results of section 4 help determine the unique aspects of the distributions of sojourn times for chain-of-events data.

In practice, the uniqueness of $\hat{\boldsymbol{\eta}}$ will depend on the choice of mass points for $T_1, T_2, \dots, T_{J-1}$. As the number of mass points increase, uniqueness will be more difficult to establish. Too fine a selection of mass points can yield a loss in important features of the distribution functions. It is recommended that one choose mass points for $T_1, T_2, \dots, T_{J-1}$ based on the planned visit schedule for patients and other known biological features of the disease being studied.

In practice, many of the components of $\hat{\boldsymbol{\eta}}$ will have unique maximum likelihood estimates even when the whole vector $\hat{\boldsymbol{\eta}}$ is not unique. The results in this paper provide useful tools for determining this set of components. Between constructing $\mathcal{C}$ and examining the value of $\hat{\boldsymbol{\eta}}$, the components with unique NPMLEs can easily be separated from those which are not uniquely identifiable. In practice, the steps outlined in the algorithms in Tables 1 and 2 of the supplementary material can be facilitated by automated software, which is being developed

and will be made available by the authors. Given that right-censoring is very common in longitudinal studies of progressive diseases, we anticipate that the algorithm in Table 2 of the supplementary material will be useful in a number of circumstances. However, we also acknowledge that it is not exhaustive since it can only be utilized in data sets with right-censoring. More research needs to be done to determine if there are other useful algebraic techniques for solving the system of equations in Theorem 3.

Ideally, one will be able to use Theorem 1 to establish the uniqueness of $\hat{\eta}$. However, the conditions of Theorem 1 are much stronger than is necessary for $\hat{\eta}$ to be unique. For example, by examining the system of equations ϕ given in Theorem 3, the uniqueness of $\hat{\eta}$ can be established without having strict concavity of $L_2(\theta)$ over Θ . Moreover, as shown in our examples, it will not always be straightforward to determine whether η or some of its components are unique. In these cases, thorough grid searches and multiple starting values will need to be utilized to help analyst feel more secure about the estimated cumulative distribution functions.

In many settings, it is likely that having interval-censored chain-of-event data on many states will increase the number of unique components in $F_1(\cdot)$ when compared to estimating $F_1(\cdot)$ using only (X_{1L}^i, X_{1R}^i) for $i = 1, \dots, N$. When one estimates $F_1(\cdot)$ using only the univariate interval censored observations on X_1 , the unique regions of $F_1(\cdot)$ can be constructed as described in Turnbull (1976). Namely, one forms the set of ordered disjoint intervals whose left and right hand endpoints come from $\{X_{1L}^i\}$ and $\{X_{1R}^i\}$, respectively and the NPMLE of $F_1(\cdot)$ will be non-unique within these intervals but unique everywhere else. When one simultaneously analyzes $(X_{1L}^i, X_{1R}^i, X_{2L}^i, X_{2R}^i, \dots, X_{(J-1)L}^i, X_{(J-1)R}^i)$ for $i = 1, \dots, N$, the information added from knowledge about the distributions of $T_2, \dots, T_{J-1}$ aids in estimation and uniqueness of the distribution of T_1 within some of these regions. Similar gains in uniqueness can be expected for $F_2(\cdot), \dots, F_{J-1}(\cdot)$. We note that the example in section 4.3 of the paper illustrates that gains in uniqueness are achievable when analyzing chain-of-events data as opposed to ordinary univariate interval censored data. For example, applying interval censored methods to $[X_L, X_R]$ leads to $f_{11} + f_{12}$ and f_{13} having unique MLEs, whereas when simultaneously analyzing $(X_L, X_R, Z_L, Z_R, Y_L, Y_R)$, f_{11}, f_{12} , and f_{13} all have unique MLEs. Such gains in uniqueness have also been noted by Sternberg and Satten (1999).

Future areas of interest include models which allow for dependence between $T_1, T_2, \dots, T_{J-1}$ as well as uniqueness of the NPMLE when $T_1, T_2, \dots, T_{J-1}$ are assumed to be continuous random variables.

Acknowledgments

The authors are very grateful to Michael Busch, Eberhard Fiebig, and David Wright for providing the data for Example 3, and thank Tianxi Cai and Victor De Gruttola for their helpful suggestions.

Funding

National Institute of Health (T32-AI007358, AI24643)

References

1. Betensky RA, Finkelstein DM. A non-parametric maximum likelihood estimator for bivariate interval censored data. *Statistics in Medicine* 1999;18:3089–3100. [PubMed: 10544308]
2. DeGruttola V, Lagakos SW. Analysis of doubly-censored survival data, with application to AIDS. *Biometrics* 1989;45:1–11. [PubMed: 2497809]

3. Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, Heldebrant C, Smith R, Conrad A, Kleinman SH, Busch MP. Dynamics of hiv viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary hiv infection. *AIDS* 2003;17:1871–1879. [PubMed: 12960819]
4. Frydman H. Semiparametric estimation in a three-state duration-dependent Markov model from interval-censored observations with application to AIDS data. *Biometrics* 1995;51:502–511. [PubMed: 7662839]
5. Gentleman R, Geyer CJ. Maximum likelihood for interval censored data: Consistency and computation. *Biometrika* 1994;81:618–623.
6. Gentleman R, Vandal AC. Nonparametric estimation of the bivariate CDF for arbitrarily censored data. *The Canadian Journal of Statistics* 2002;30(4):557–571.
7. Gomez G, Calle ML. Non-parametric estimation with doubly censored data. *Journal of Applied Statistics* 1999;26:45–58.
8. Gomez G, Lagakos SW. Estimation of the infection time and latency distribution of AIDS with doubly censored data. *Biometrics* 1994;50:204–212. [PubMed: 8086603]
9. Huzurbazar S, Huzurbazar AV. Survival and hazard functions for progressive diseases using saddlepoint approximations. *Biometrics* 1999;55:198–203. [PubMed: 11318155]
10. Jackson CH, Sharples LD. Hidden markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. *Statistics in Medicine* 2002;21:113–128. [PubMed: 11782054]
11. Kalbfleisch JD, Lawless JF. The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* 1985;80:863–871.
12. Kalbfleisch, JD.; Lawless, JF. Some statistical methods for panel life history data; *Statistics Canada Symposium*; 1989. p. 185-192.
13. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 1958;53:457–481.
14. Lagakos SW, Sommer CJ, Zelen M. Semi-Markov models for partially censored data. *Biometrika* 1978;65:311–318.
15. Longini IM, Clark JSW, Byers RH, Ward JW, Darrow WW, Lemp GF, Hethcote HW. Statistical analysis of the stages of HIV infection using a Markov model. *Statistics in Medicine* 1989;8:831–843. [PubMed: 2772443]
16. Oiler R, Gomez G, Calle LM. Interval censoring: model characterizations for the validity of the simplified likelihood. *The Canadian Journal of Statistics* 2004;32(3):315–326.
17. Prasitwattanaseree S, Lallemand M, Costagliola D, Jourdain G, Mary J. Influence of mother and infant zidovudine treatment duration on the age at which hiv infection can be detected by polymerase chain reaction in infants. *Antiviral Therapy* 2004;9:179–185. [PubMed: 15134179]
18. Sternberg MR, Satten GA. Discrete-time nonparametric estimation for semi-Markov models of chain-of-events data subject to interval censoring and truncation. *Biometrics* 1999;55:514–522. [PubMed: 11318208]
19. Sun, J. *The Statistical Analysis of Interval-Censored Failure Time Data*. New York, NY: Springer; 2006.
20. Sun J, Kalbfleisch J. Estimation of the mean function of point processes based on panel count data. *Statistica Sinica* 1995;5:279–290.
21. Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society, Series B, Methodological* 1976;38:290–295.

Appendix A: Proof of Theorem 1

We know that Θ_g is a convex, compact set because H is a convex and compact set and g is a continuous mapping of H to Θ_g . Since $L_2(\theta)$ is strictly concave over Θ , it is also strictly concave over Θ_g . It follows that $L_2(\theta)$ has a unique maximum, say $\hat{\theta}_g$, over Θ_g . We now show that $g : \eta \rightarrow \theta$ is one-to-one and thus, that only one value of η maps to $\hat{\theta}_g$. Consider two elements of H , say η^* and η^{**} , such that $g(\eta^*) = g(\eta^{**}) = \hat{\theta}_g$ where

$\eta^* = (f_{11}^*, \dots, f_{1K_1}^*, f_{21}^*, \dots, f_{2K_2}^*, \dots, f_{(J-1)1}^*, \dots, f_{(J-1)K_{J-1}}^*)$ and $\eta^{**} = (f_{11}^{**}, \dots, f_{1K_1}^{**}, f_{21}^{**}, \dots, f_{2K_2}^{**}, \dots, f_{(J-1)1}^{**}, \dots, f_{(J-1)K_{J-1}}^{**})$. By assumption,

$$\begin{aligned} f_{11}^* f_{21}^* \cdots f_{(J-1)1}^* &= f_{11}^{**} f_{21}^{**} \cdots f_{(J-1)1}^{**} \\ f_{11}^* f_{21}^* \cdots f_{(J-1)2}^* &= f_{11}^{**} f_{21}^{**} \cdots f_{(J-1)2}^{**} \\ &\vdots \\ f_{11}^* f_{21}^* \cdots f_{(J-1)K_{J-1}}^* &= f_{11}^{**} f_{21}^{**} \cdots f_{(J-1)K_{J-1}}^{**} \end{aligned}$$

Summing these equations together, we find $f_{11}^* = f_{11}^{**}$. Using similar arguments, we can show that $f_{jk_1}^* = f_{jk_1}^{**}$ for $k_1 = 1, \dots, K_1$ and $j = 1, \dots, J-1$. Thus, $\eta^* = \eta^{**}$ and hence $L_1(\eta)$ has a unique maximum, $\hat{\eta}$.

Appendix B: Proof of Theorem 4

It is shown in Gentleman and Geyer (1994) that $L_2(\theta)$ is concave over Θ . Thus, $L_2(\theta)$ is also concave over Θ_g . For any $\eta, \eta' \in H$, define $\theta = g(\eta)$ and $\theta' = g(\eta')$, and note that $L_2(\theta) = L_2(g(\eta)) = L_1(\eta)$. For any $\lambda \in (0, 1)$,

$$\begin{aligned} L_2((1-\lambda)\theta + \lambda\theta') &= L_2((1-\lambda)g(\eta) + \lambda g(\eta')) \\ &= L_1((1-\lambda)\eta + \lambda\eta') \end{aligned} \quad (8)$$

and

$$(1-\lambda)L_2(\theta) + \lambda L_2(\theta') = (1-\lambda)L_1(\eta) + \lambda L_1(\eta'). \quad (9)$$

Since $L_2(\theta)$ is concave,

$$L_1((1-\lambda)\eta + \lambda\eta') \geq (1-\lambda)L_1(\eta) + \lambda L_1(\eta'). \quad (10)$$

and hence $L_1(\eta)$ is concave. Now suppose that the extended DGL algorithm converges to a strict local maximum, say $\hat{\eta}$. Since $L_1(\eta)$ is concave, $\hat{\eta}$ must be a global maximum. We show by contradiction that $\hat{\eta}$ is the unique global maximum. Take $\hat{\eta}^*$ to be another global maximum of $L_1(\eta)$ so that $L_1(\hat{\eta}) = L_1(\hat{\eta}^*)$. Since $\hat{\eta}$ is a strict global maximum, $L_1(\hat{\eta}) > L_1(\eta)$ for η in a sufficiently small neighborhood of $\hat{\eta}$. However, since $L_1(\eta)$ is also concave,

$$L_1(\lambda\hat{\eta} + (1-\lambda)\hat{\eta}^*) \geq \lambda L_1(\hat{\eta}) + (1-\lambda)L_1(\hat{\eta}^*) = L_1(\hat{\eta}) \quad \text{for all } \lambda \in (0, 1). \quad (11)$$

This holds only if $\hat{\eta} = \hat{\eta}^*$ and hence $\hat{\eta}$ is the unique maximum of $L_1(\eta)$.

Fig. 1.
Four-state progressive disease model.

Fig. 2.
Estimated cumulative distribution functions for time spent in States 1, 2, and 3 of BOS infection from six months after lung transplant.

Fig. 3. Estimated nonparametric (solid lines) and parametric (dashed lines) cumulative distribution functions for time spent in the acute and recent infection states of primary HIV.