Non-parametric maximum likelihood estimation of interval censored failure time data subject to misclassification

Andrew C. Titman

September 30, 2016

Abstract

The paper considers non-parametric maximum likelihood estimation of the failure time distribution for interval censored data subject to misclassification. Such data can arise from two types of observation scheme; either where observations continue until the first positive test result or where tests continue regardless of the test results. In the former case, the misclassification probabilities must be known, whereas in the latter case joint estimation of the event-time distribution and misclassification probabilities is possible. The regions for which the maximum likelihood estimate can only have support are derived. Algorithms for computing the maximum likelihood estimate are investigated and it is shown that algorithms appropriate for computing non-parametric mixing distributions perform better than an iterative convex minorant algorithm in terms of time to absolute convergence. A profile likelihood approach is proposed for joint estimation. The methods are illustrated on a data set relating to the onset of cardiac allograft vasculopathy in post-heart-transplantation patients.

1   Introduction

Interval censored time-to-event data is commonly encountered in applications in which continual monitoring is not possible and instead the status of a subject, for instance whether they have developed a disease, is only known at intermittent observation or examination times. It is usually assumed that the status of the

process can be determined with complete accuracy at each of these examination times. However, in many medical applications where the event of interest is onset of a disease or onset of infection, diagnosis may be subject to classification error.

Time-to-event data subject to misclassification error has been considered previously. Espeland et al (1989) analyzed data relating to the diagnosis of onset of puberty in adolescents and considered a discrete time model where events could occur within 'sub epochs' of 'epochs' for which the incidence rate was assumed constant. They developed an EM algorithm that allowed for estimation of the underlying incidence rates and error probabilities. Balasubramanian and Lagakos (2003) considered estimation for a general continuous time model, but ensured identifiability by fixing the increments in the failure time distribution to be equal between some time points. Fully non-parametric estimation of the failure time distribution $F(t)$ has been considered in the case of current status (or Type I interval censored) survival data subject to misclassification of outcomes (McKeown and Jewell (2010); Sal y Rosas and Hughes (2011)). In that case, since each subject is observed precisely once, the misclassification probabilities must be known *a priori* in order to estimate the survivor distribution.

The purpose of this paper is to develop methods for fully non-parametric estimation of the failure time distribution, $F(t)$, for general (or mixed case) interval censored data subject to classification error, by extending existing methods for standard interval censored survival data. Two observation schemes are considered. In the first scenario, observations may be assumed to persist only until the first observed failure. Such an observation scheme has been considered in a discrete-time setting by Richardson and Hughes (2000) where patients were assessed based on an imperfect diagnostic test, with known error probabilities, repeatedly until their first positive test at which point they were treated and hence left the study. In this case, the misclassification probabilities must be known in order to estimate $F(t)$.

However, a second scenario is also considered where it is assumed that examinations of a subject continue to occur regardless of whether a positive result has been observed. This type of observation has been considered both for interval censored survival data without misclassification (Betensky et al, 2001; Zhu et al, 2008) and for misclassified data (Espeland et al, 1989). Such data may arise in studies where there are several outcomes of interest. In this case methods to jointly estimate $F(t)$ and the misclassification probabilities, which can be constant or depend parametrically on time, are considered.

The remainder of the paper is organized as follows. Section 2 formalizes the problem and characterizes the non-parametric maximum likelihood estimate (NPMLE), Section 3 discusses computation of the NPMLE, Section 4 considers methods for obtaining standard errors or confidence regions, Section 5 presents

2

a numerical study and Section 6 gives an illustrative example based on data from post-heart-transplantation patients. The paper concludes with a discussion of potential model extensions.

2 Non-parametric Maximum Likelihood Estimate

Consider failure times $T_j \sim F(t)$ for subjects $j = 1, \ldots, n$, indirectly observed through a series of observations $\mathbf{Y}_j = (Y_{1j}, \ldots, Y_{n_j j})$ at times $\mathbf{X}_j = (X_{1j}, \ldots, X_{n_j j})$ where Y_{ij} is a binary random variable such that

$$P(Y_{ij} = 1 | X_{ij} < T_j) = \alpha(X_{ij}; \boldsymbol{\theta})$$

and

$$P(Y_{ij} = 0 | X_{ij} \geq T_j) = \beta(X_{ij}; \boldsymbol{\theta}).$$

At each observation time, Y_{ij} is the outcome of an imperfect diagnostic test where the functions $\alpha(x; \boldsymbol{\theta})$ and $\beta(x; \boldsymbol{\theta})$ represent the probabilities of a false positive and false negative result, respectively, and may depend on the age of the process, x. It is assumed that $\alpha(x; \boldsymbol{\theta})$ and $\beta(x; \boldsymbol{\theta})$ are known functions satisfying $0 \leq \alpha(x; \boldsymbol{\theta}) < 1$, $0 \leq \beta(x; \boldsymbol{\theta}) < 1$ and $0 \leq \alpha(x; \boldsymbol{\theta}) + \beta(x; \boldsymbol{\theta}) < 1$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $x > 0$. The observation times X_{ij} are potentially random, but are assumed to be generated from a process that is non-informative for F and $\boldsymbol{\theta}$, as defined by Grüger et al (1991).

Interest lies in estimating F and possibly also $\boldsymbol{\theta}$. When observations stop at the first failure time, $\alpha(x; \boldsymbol{\theta})$ and $\beta(x; \boldsymbol{\theta})$ must be fully specified. However, if observations persist after a first positive result then $\boldsymbol{\theta}$ can also be estimated.

In the simplest case, the diagnostic test performed at each observation time would be assumed to have fixed sensitivity, $(1 - \alpha)$, and specificity $(1 - \beta)$. A convenient way to parameterize this is through a logistic transformation such that $\alpha(x; \boldsymbol{\theta}) = (1 + \exp(-\theta_1))^{-1}$ and $\beta(x; \boldsymbol{\theta}) = (1 + \exp(-\theta_2))^{-1}$ for $\boldsymbol{\theta} = (\theta_1, \theta_2)$. Alternatively if, as in Espeland et al (1989), the test accuracy is considered age-dependent, this could be represented by

$$\text{logit}\{\alpha(x; \boldsymbol{\theta})\} = \begin{cases} \theta_1 & \text{if } x < \tau_1 \\ \theta_1 + \theta_2 & \text{if } \tau_1 \leq x < \tau_2 \\ \quad \vdots & \quad \vdots \\ \theta_1 + \theta_m & x \geq \tau_{K-1}, \end{cases}$$

for fixed cut-points $0 < \tau_1 < \ldots < \tau_{K-1}$, with a similar expression for $\beta(x; \boldsymbol{\theta})$, implying that the sensitivity and specificity are piecewise constant functions.

3

Assuming the observation times, $\mathbf{X}_j$ are non-informative, the likelihood with respect to F for a single subject j can be expressed as

$$L_j = \sum_{i=1}^{n_j+1} \{F(X_{ij}+) - F(X_{i-1j}-)\} \times \prod_{k=1}^{i-1} \alpha(X_{kj}; \boldsymbol{\theta})^{Y_{kj}} (1 - \alpha(X_{kj}; \boldsymbol{\theta}))^{1-Y_{kj}}$$
$$\times \prod_{k=i}^{n_j} \beta(X_{kj}; \boldsymbol{\theta})^{1-Y_{kj}} (1 - \beta(X_{kj}; \boldsymbol{\theta}))^{Y_{kj}}$$

where $X_{0j} = 0$ and $X_{n_j+1j} = \infty$ and $F(\infty) = 1$. Note that in the case where $\alpha \equiv \beta \equiv 0$, such that there is no classification error, the expression reduces to the standard expression for interval censored survival data,

$$\begin{aligned} L_j &= \sum_{i=1}^{n_j+1} \{F(X_{ij}+) - F(X_{i-1j}-)\} I(Y_{ij} = 1, Y_{i-1j} = 0) \\ &= \{F(R_j+) - F(L_j-)\}. \end{aligned}$$

where $Y_{0j} = 0$ and $Y_{n_j+1j} = 1$ and $R_j = \min_i(\{X_{ij'} : Y_{ij'} = 1, j' = j\} \cup \{\infty\})$ and $L_j = \max_i(\{X_{ij'} : Y_{ij'} = 0, j' = j\} \cup \{\infty\})$.

2.1 Support of the NPMLE

Under standard interval censored survival data, the non-parametric maximum likelihood estimate (NPMLE) is only uniquely defined up to increments of mass within semi-closed intervals between unique observation times (Turnbull, 1976; Ng, 2002). The NPMLE for misclassified survival data also has this property and it is only the determination of the particular intervals of support which differs.

Initially assume that $\alpha(x; \boldsymbol{\theta}) > 0, \beta(x; \boldsymbol{\theta}) > 0$ for all $x \geq 0$ and that examination times are generated from a continuous distribution such that there are no tied observation times with probability 1. For notational convenience in the proof below the dependence of α and β on x and $\boldsymbol{\theta}$ is suppressed, but the argument remains valid for non-constant error probabilities.

Let $\mathcal{L} = \{0\} \cup \{X_{ij} : Y_{ij} = 0\}$ and $\mathcal{R} = \{X_{ij} : Y_{ij} = 1\} \cup \{\infty\}$ denote the sets of observation times corresponding to negative and positive test results, respectively.

A further set can be constructed consisting of disjoint intervals whose left and right end points lie in the sets $\mathcal{L}$ and $\mathcal{R}$ respectively and which do not contain any members of $\mathcal{L}$ or $\mathcal{R}$. Let $\{Q_i := (l_i, r_i], 1 \leq i \leq I\}$ be those intervals, where $l_1 < r_1 < l_2 < r_2 < \ldots < l_I < r_I \leq \infty$ and $Q = \cup_i Q_i$.

Theorem 1

The NPMLE of the cumulative distribution function, F, can only increase on Q and is unique only up to the values of $F(r_i)$ and $F(l_i-), (1 \leq i \leq I)$.

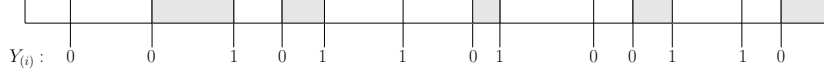$Y_{(i)}:$ 0   0   1   0   1   1   0   1   0   0   1   1   0

Figure 1: Example of the (shaded) support regions of the NPMLE for a particular set of pooled observations $Y_{(i)}$ corresponding to the ordered observation times $X_{(i)}$

Proof

Let $X_{(i)}$ denote the ith ordered observation time from the pooled set of observation times across all subjects, with corresponding observation $Y_{(i)}$. Suppose the NPMLE, $\hat{F}$, increases by $\epsilon > 0$ between times $X_{(i)}$ and $X_{(i+1)}$ corresponding to observations $Y_{(i)}$ and $Y_{(i+1)}$, where $Y_{(i)} = 1$. Construct $\tilde{F}$ to be identical to $\hat{F}$ except that the increment on $(X_{(i)}, X_{(i+1)}]$ is set to zero and the increment on $(X_{(i^*)}, X_{(i^*+1)}]$ is increased from π_0 to $\pi_0 + \epsilon$, where $i^* = \arg\max_j\{j : j < i, Y_{(j)} = 0, Y_{(j+1)} = 1\}$. The likelihood under $\hat{F}$ and $\tilde{F}$ is equal for all subjects except individuals observed at time points $X_{(i^*+1)}, \ldots, X_{(i)}$. For each of these subjects, assuming they are observed exactly once in $[X_{(i^*+1)}, X_{(i)}]$, their likelihood contributions are of the form $L(\hat{F}) = C + K(\pi_0(1 - \beta) + \epsilon\alpha)$ and $L(\tilde{F}) = C + K(\pi_0 + \epsilon)(1 - \beta)$, for some $C, K > 0$. Hence $L(\tilde{F}) - L(\hat{F}) = K\epsilon(1 - \beta - \alpha) > 0$, since $\alpha + \beta < 1$, and therefore $\hat{F}$ cannot be the MLE. If there are subjects observed $m > 1$ times within $[X_{(i^*+1)}, X_{(i)}]$ then $L(\hat{F}) = C + K(\pi_0(1 - \beta) + \epsilon\alpha)^m$ and $L(\tilde{F}) = C + K(\pi_0 + \epsilon)^m(1 - \beta)^m$ and $L(\tilde{F}) > L(\hat{F})$ provided $\alpha + \beta < 1$.

Similarly, suppose the NPMLE, $\hat{F}$, increases by $\epsilon > 0$ between times $X_{(i)}$ and $X_{(i+1)}$ corresponding to observations $Y_{(i)}$ and $Y_{(i+1)}$, where $Y_{(i+1)} = 0$. Construct $\tilde{F}$ to be identical to $\hat{F}$ except the increment on $(X_{(i)}, X_{(i+1)}]$ is set to zero and the increment on $(X_{(i^*)}, X_{(i^*+1)}]$ is increased from π_0 to $\pi_0 + \epsilon$, where $i^* = \arg\min_j\{j : j > i, Y_{(j)} = 0, Y_{(j+1)} = 1\}$. The likelihood contributions under $\hat{F}$ and $\tilde{F}$ are equal for all subjects except individuals observed at time points $X_{(i+1)}, \ldots, X_{(i^*+1)}$. For each of these subjects, assuming they are observed exactly once in $[X_{(i+1)}, X_{(i^*+1)}]$ their likelihood contributions are of the form $L(\hat{F}) = C + K(\pi_0(1 - \beta) + \epsilon\alpha)$ while $L(\tilde{F}) = C + K(\pi_0 + \epsilon(1 - \beta))$, for some $C, K > 0$. Hence $L(\tilde{F}) - L(\hat{F}) = K\epsilon(1 - \beta - \alpha) > 0$ since $\alpha + \beta < 1$, and therefore $\hat{F}$ cannot be the MLE. Again, a similar argument follows for subjects observed $m > 1$ times within $[X_{(i+1)}, X_{(i^*+1)}]$.

$\square$

Figure 1 illustrates the implied support regions for a particular set of pooled observations ordered by observation time.

Special cases arise when either one of $\alpha(x; \boldsymbol{\theta})$ or $\beta(x; \boldsymbol{\theta})$ is identically equal to 0. If $\alpha(x; \boldsymbol{\theta}) \equiv 0$, an

individual's failure time can be determined to have occurred before the first positive observation. Define $X_j^- = \inf(\{X_{ij'} : Y_{ij'} = 1, j' = j\} \cup \{\infty\})$ for $j = 1, \ldots, n$ and let $\mathcal{L}^- = \cup_j \{X_{ij'} : Y_{ij'} = 0, X_{ij'} < X_j^-, j' = j\} \cup \{0\}$ and $\mathcal{R}^- = \cup_j \{X_{ij'} : X_{ij'} = X_j^-, j' = j\} \cup \{\infty\}$. The set of support, Q, is then constructed from the disjoint closed sets $\mathcal{L}^-$ and $\mathcal{R}^-$.

If $\beta(x; \boldsymbol{\theta}) \equiv 0$, an individual's failure time can be determined to have occurred after the last observed 0. Define $X_j^+ = \sup(\{0\} \cup \{X_{ij'} : Y_{ij'} = 0, j' = j\})$ and let $\mathcal{L}^+ = \cup_j \{X_{ij'} : X_{ij'} = X_j^+, j' = j\}\{0\}$ and $\mathcal{R}^+ = \cup_j \{X_{ij'} : Y_{ij'} = 1, X_{ij'} > X_j^+, j' = j\} \cup \{\infty\}$. The set of support Q is then constructed from $\mathcal{L}^+$ and $\mathcal{R}^+$.

If $\alpha(x; \boldsymbol{\theta}) \not\equiv 0$ and $\beta(x; \boldsymbol{\theta}) \not\equiv 0$ but are equal to zero for some observation times, define $X_j^- = \inf(\{X_{ij'} : Y_{ij'} = 1, \alpha(X_{ij'}; \theta) = 0, j' = j\} \cup \{\infty\})$ and $X_j^+ = \sup(\{0\} \cup \{X_{ij'} : Y_{ij'} = 0, \beta(X_{ij'}; \theta) = 0, j' = j\})$ and $\mathcal{L}^\pm = \cup_j \{X_{ij'} : Y_{ij'} = 0, X_j^+ \le X_{ij'} < X_j^-, j' = j\} \cup \{0\}$ and $\mathcal{R}^\pm = \cup_j \{X_{ij'} : Y_{ij'} = 1, X_j^+ < X_{ij'} \le X_j^-, j' = j\} \cup \{\infty\}$. Again the set of support Q is constructed from $\mathcal{L}^\pm$ and $\mathcal{R}^\pm$.

When $\alpha(x; \boldsymbol{\theta}) \equiv \beta(x; \boldsymbol{\theta}) \equiv 0$, the data are standard interval censored and the support set is defined as in Turnbull (1976) in the case where there is no truncation.

The above assumes the observation times are unique. If there are ties we can define $X_{(i)}$ to be the ith ordered unique observation time and an analogous argument can be applied to show that a necessary condition for the NPMLE to increase on $(X_{(i)}, X_{(i+1)}]$ is that at least one observed state at time $X_{(i)}$ must be 0 and at least one observed state at time $X_{(i+1)}$ must be 1.

2.2 Optimality conditions

Böhning et al (1996) noted the relationship between the non-parametric estimates of the survivor function for interval censored data and the NPMLE of a mixing distribution. This relationship also exists in the case of misclassified failure time data. Again assuming the X_{ij} are non-informative, the likelihood with respect to F for a single individual can also be expressed as

$$L_j = \int_0^\infty \prod_{i=1}^{n_j} P(Y_{ij}|X_{ij}, T_j = t) dF(t),$$

where the unobserved failure time T_j acts as a mixing variable.

Having established the support of the NPMLE in Section 2.1, the likelihood can be expressed in terms of a finite mixture. Let

$$\gamma_{ij}(\alpha, \beta) = P(\mathbf{Y}_j | \mathbf{X}_j, T_j \in (l_i, r_i])$$

represent the joint probability of subject j's outcome vector $\mathbf{Y}_j$ given observation times $\mathbf{X}_j$ and that the failure time for subject j occurs within the ith support region, $Q_i = (l_i, r_i]$. Note that $\gamma_{ij}(\alpha, \beta)$ depends on the full classification probability functions $\alpha(x; \boldsymbol{\theta})$ and $\beta(x; \boldsymbol{\theta})$ but this dependence is suppressed for notational convenience. Denote the probability that the failure occurs in the ith support region as $\pi_i = F(r_i) - F(l_i-)$ with $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_I)$ then the likelihood for the data can be written as

$$L(\boldsymbol{\pi}) = \prod_{j=1}^{n} \sum_{i=1}^{I} \gamma_{ij}(\alpha, \beta) \pi_i.$$

Let $\mathbf{S}$ be the $I \times n$ matrix with (i, j) entry

$$d_{ij}(\pi) = \frac{\gamma_{ij}(\alpha, \beta)}{\sum_{k=1}^{I} \gamma_{ik}(\alpha, \beta) \pi_k},$$

the first and second derivatives of the log-likelihood can be expressed as $\nabla \log L = (\mathbf{S1})'$ and $\nabla^2 \log L = \mathbf{SS}'$ respectively, hence the Hessian is positive-semi-definite and $\log L$ is concave.

In general, $\gamma_{ij}(\alpha, \beta)$ can be computed by considering which of subject j's observations, $\mathbf{Y}_j$, would be false positives, true positives, false negatives and true negatives given the failure occurred in $(l_i, r_i]$ and computing the relevant probability at the corresponding event times. Hence the general expression is of the form

$$\gamma_{ij}(\alpha, \beta) = \prod_{j=1}^{n} \prod_{k=1}^{n_j} \alpha(X_{kj}; \boldsymbol{\theta})^{Y_{kj} I(X_{kj} < l_{i-1})} (1 - \alpha(X_{kj}; \boldsymbol{\theta}))^{(1 - Y_{kj}) I(X_{kj} < l_{i-1})}$$
$$\times \beta(X_{kj}; \boldsymbol{\theta})^{(1 - Y_{kj}) I(X_{kj} \geq r_i)} (1 - \beta(X_{kj}; \boldsymbol{\theta}))^{Y_{kj} I(X_{kj} \geq r_i)}.$$

However, in the special case where the misclassification probabilities are time constant it is only necessary to count the number of false and true negatives and positives, so that the expression can be simplified to

$$\gamma_{ij}(\alpha, \beta) = \alpha^{n_{1j}^-(l_{i-1})} (1 - \alpha)^{n_{0j}^-(l_{i-1})} \beta^{n_{0j}^+(r_i)} (1 - \beta)^{n_{1j}^+(r_i)} \tag{2.1}$$

where

$$n_{1j}^-(t) = \sum_{k=1}^{n_j} Y_{kj} I(X_{kj} < t), \qquad n_{1j}^+(t) = \sum_{k=1}^{n_j} Y_{kj} I(X_{kj} \geq t),$$

$$n_{0j}^-(t) = \sum_{k=1}^{n_j} (1 - Y_{kj}) I(X_{kj} < t), \quad n_{0j}^+(t) = \sum_{k=1}^{n_j} (1 - Y_{kj}) I(X_{kj} \geq t).$$

Finally, in the case where there is no classification error, such that $\alpha = \beta = 0$ and meaning a subject's failure time can be bounded by $(L_j, R_j]$, the expression reduces to

$$\gamma_{ij}(0, 0) = I((l_i, r_i] \subseteq (L_j, R_j]).$$

This expression can be derived from (2.1) by adopting the convention that $0^0 = 1$ and noting that when there is no misclassification, $I(n_{1j}^-(t) = 0) = I(R_j \leq t)$ and $I(n_{0j}^+(t) = 0) = I(L_j \geq t)$.

The NPMLE of F can be obtained by considering the constrained maximization problem $\max_{\boldsymbol{\pi}} \log(L(\boldsymbol{\pi}))$ subject to $\sum_{i=1}^{I} \pi_i = 1$ and $\pi_i \geq 0$ for all i. The Kuhn-Tucker conditions for $\boldsymbol{\pi}$ to be the NPMLE relate to the directional derivatives. The directional derivative of $L(\boldsymbol{\pi})$ from $\boldsymbol{\pi}$ to a mixing distribution with all mass in the interval $(l_i, r_i]$ is given by $d_i(\boldsymbol{\pi}) = \sum_{j=1}^{n} d_{ij}(\boldsymbol{\pi}) - n$. The Kuhn-Tucker conditions for optimality can then be expressed as $\max_i d_i(\boldsymbol{\theta}) = 0$. The equivalence of maximizing the likelihood with minimizing the largest directional derivative and finding the solution with $\max_i d_i(\boldsymbol{\theta}) = 0$ can also be shown by using a fundamental theorem of mixture estimation in Lindsay (1983), which is analogous to the general equivalence theorem of optimal design.

3 Computation

Initial proposals for algorithms for finding the NPMLE both in the case of interval censored survival data (Turnbull, 1976) and also in the case of non-parametric mixing distributions (Laird, 1978), were based upon EM algorithms. However, in each case other algorithms have been devised with substantially better properties. This section considers applying two of the best performing algorithms to the problem of finding the NPMLE for misclassified data.

3.1 Iterative convex minorant algorithm

The best performing algorithms for computing the NPMLE for interval censored survival data are the iterative convex minorant (ICM) algorithm (Jongbloed, 1998) or hybrid ICM-EM algorithm (Wellner and Zhan, 1997). In principle, an ICM type algorithm could also be used for misclassified survival data.

The ICM approach parameterizes by F_i where $\pi_i = F_i - F_{i-1}$ and uses a quadratic function about F_i based on its gradient and approximate Hessian (taken as the diagonal of the full Hessian) and maximizes this quadratic form with respect to monotonicity constraints on F. Under this parameterization, the first derivatives of the log-likelihood have the form

$$g_i = \frac{\partial \log L}{\partial F_i} = \sum_k \frac{\gamma_{ik} - \gamma_{i+1k}}{L_k}$$

and the diagonal elements of the Hessian have the form

$$H_{ii} = -\frac{\partial^2 \log L}{\partial F_i^2} = \sum_k \frac{(\gamma_{ik} - \gamma_{i+1k})^2}{L_k^2}.$$

Each step of the ICM algorithm then involves updating $\tilde{F}^{(s)}$ by taking $\tilde{F}^{(s+1)}$ to be the solution, $\hat{F}$, to the

isotonic regression problem

$$\min \sum_i H_{ii}(\hat{F} - \tilde{F}_i^{(s)} - \frac{g_i}{H_{ii}})^2 \qquad (3.1)$$

subject to monotonicity constraints on $\hat{F}$. This can be achieved by using the pool-adjacent-violators (PAV) algorithm (Barlow et al, 1972).

However to guarantee convergence, it is necessary to include a line search step where $\tilde{F}^{(s+1)}$ is taken to be the point on the line $\tilde{F}^{(s)} + (\hat{F} - \tilde{F}^{(s)})d$, $d \in (0, 1]$ for which the likelihood is maximized. In practice, as suggested by Jongbloed (1998), the line search only needs to be applied if taking the step length $d = 1$, i.e. applying $\tilde{F}^{(s+1)} = \hat{F}$ does not result in an increase in the likelihood or increases the likelihood by less than some specified threshold.

The effectiveness of (3.1) as a quasi-Newton step and hence the effectiveness of ICM as a whole, depends on the extent to which the full Hessian can be approximated by its diagonal. For standard interval censored data the approximation is usually quite effective since each subject's failure time can be observed to have only occurred in a small proportion of the support regions. In contrast for misclassified data, with imperfect sensitivity and specificity, no support regions can be ruled out for any subject. The effectiveness of a diagonal approximation will depend on the error probabilities α and β, with better performance expected when these are smaller.

3.2 Constrained Newton algorithm

Given the representation as a non-parametric mixture problem, an alternative approach to take is to use an algorithm appropriate for finding the NPMLE of a mixing distribution. Several such algorithms exist, for instance the vertex exchange method (Böhning (1985)), the intra-simplex direction method (Lesperance and Kalbfleisch (1992)) and the constrained Newton method of Wang (2007). Here an adapted version of this latter algorithm is proposed.

The main idea behind the constrained Newton algorithm is to use the optimality condition based upon the directional derivatives, to determine the best locations to add new support points or regions. Specifically it seeks the maximum, or maxima, of the directional derivatives, with respect to the mixing parameter π and proposes these points as additions to the finite mixture.

The constrained Newton algorithm can be difficult to implement for a general non-parametric mixture where as it requires a line search over the continuous space of π to find the maxima. However, the problem is simplified for the case of misclassified interval censored data, as the support reduction step ensures that only a finite number of locations, corresponding to the support intervals $Q_i = (l_i, r_i]$, need to be considered.

The proposed algorithm involves collecting a set of support intervals $\mathcal{I} \subseteq Q$, where $\pi_i > 0$ for all $i \in \mathcal{I}$ and $\pi_i = 0$ for all $i \notin \mathcal{I}$. Set $s = 0$ and an initial feasible $\boldsymbol{\pi}^{(s)}$ implying a support set $\mathcal{I}^{(s)}$, such that $\log L(\boldsymbol{\pi}^{(s)}) > -\infty$.

1. Find the set $\mathcal{I}^*$ of $i \in \mathcal{I}^{(s)}$ such that $d_i(\boldsymbol{\pi}^{(s)}) > d_{i-1}(\boldsymbol{\pi}^{(s)})$ and $d_i(\boldsymbol{\pi}^{(s)}) > d_{i+1}(\boldsymbol{\pi}^{(s)})$ where, by convention, $d_0(\boldsymbol{\pi}^{(s)}) = d_{I+1}(\boldsymbol{\pi}^{(s)}) = 0$. If $\max_i d_i(\boldsymbol{\pi}^{(s)}) = 0$ stop.

2. Set $\mathcal{I}^+ = \mathcal{I} \cup \mathcal{I}^*$. Compute $\boldsymbol{\pi}^{(s+1)}$ by performing one constrained Newton iteration.

3. Discard all points in $\mathcal{I}^+$ with zero entries in $\boldsymbol{\pi}^{(s+1)}$ to obtain $\mathcal{I}^{(s+1)}$. Set $s =: s + 1$ and return to 1.

Step 2 of the algorithm involves maximizing a quadratic approximation to $\log L(\boldsymbol{\pi})$ in a neighbourhood of $\boldsymbol{\pi}^{(s)}$. Wang (2007) shows that this is equivalent to solving $\min_{\boldsymbol{\pi}^*} ||\mathbf{S}\boldsymbol{\pi}^* - 2||^2$ subject to $\boldsymbol{\pi}^{*'}\mathbf{1} = 1, \pi_i^* \geq 0$ for all $i \in \mathcal{I}^+$, which in turn can be solved using the non-negative least squares (NNLS) algorithm (Lawson and Hanson (1974)).

A simplified procedure, which retains all possible support points for $\boldsymbol{\pi}$ at each iteration of the algorithm, would also be effective but somewhat slower due to requiring a larger NNLS problem to be solved at each iteration.

In Section 5 the performance of the constrained Newton algorithm is compared with the ICM algorithm.

3.3 Semi-parametric estimation

In the scenario in which observation continues after the first observed failure, there is scope for joint estimation of F and the misclassification probabilities $\alpha(x; \boldsymbol{\theta})$ and $\beta(x; \boldsymbol{\theta})$. In this case the model can be expressed as a semi-parametric mixture, where the mixture distribution - defined through F - is non-parametric, but there are additional parameters determining the response conditional on the mixture component.

Provided $\alpha(x; \boldsymbol{\theta}), \beta(x; \boldsymbol{\theta}) > 0$ and $\alpha(x; \boldsymbol{\theta}) + \beta(x; \boldsymbol{\theta}) < 1$ for all $x > 0$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the potential support points of F will remain fixed. This ensures that the profile likelihood is differentiable with respect to $\boldsymbol{\theta}$, provided $\alpha(x; \boldsymbol{\theta})$ and $\beta(x; \boldsymbol{\theta})$ are themselves differentiable. The maximum likelihood estimate can be found by maximizing the profile log-likelihood

$$pl(\boldsymbol{\theta}) = l(\boldsymbol{\theta}, \hat{\boldsymbol{\pi}}(\boldsymbol{\theta})).$$

Clearly, the profile likelihood can be computed by applying the algorithm for non-parametric estimation for a fixed value of $\boldsymbol{\theta}$. Following Wang (2010), standard unconstrained quasi-Newton type algorithms can be

used to maximize the profile likelihood, using finite-difference approximations to evaluate

$$\frac{\partial pl(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left.\frac{\partial l(\boldsymbol{\theta}, \boldsymbol{\pi})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\pi}=\hat{\boldsymbol{\pi}}(\boldsymbol{\theta})}.$$

4 Estimates of uncertainty

Estimation of standard errors or the construction of confidence intervals for F is not straightforward since, in common with other NPMLE for mixing distributions or interval censored data standard $n^{1/2}$ asymptotics do not apply and instead the estimator converges at a slower rate (van de Geer, 2003). In the context of current status data, some progress has been made in developing asymptotic theory. For instance, if the observation times are random and generated from a continuous distribution, the convergence rate of the NPMLE is $n^{1/3}$ (Groeneboom and Wellner, 1992). Banerjee and Wellner (2001, 2005) derive the limiting form of the likelihood ratio and Tang et al (2012) have developed an adaptive inference procedure suitable when the nature of the sampling times distribution, and hence the convergence rate, is unknown. However, these results are not easily extendable either to general (mixed case) interval censored data or to misclassified data. Abrevaya and Huang (2005) showed that standard non-parametric bootstrap approaches are not consistent for estimators with $n^{1/3}$ convergence implying that the standard non-parametric bootstrap is unlikely to be consistent for misclassified survival data.

Possible approaches to quantifying the uncertainty in $\hat{F}$ could involve either a parametric bootstrap, though this requires estimation of the inspection process, subsampling (Politis and Romano, 1994) if $n^{1/3}$ convergence can be assumed, or the use of a smoothed bootstrap (Sen and Xu, 2015). If the distribution of examination times can be assumed to have finite support, then standard $n^{1/2}$ asymptotics will apply, but may nevertheless have poor performance in finite samples (Maathuis and Hudgens (2011)).

For the semi-parametric case, if interest is in estimation of the parameters $\boldsymbol{\theta}$ that determine α and β, with F as a nuisance parameter, then through direct application of results from Murphy and van der Vaart (1997), profile likelihood ratio based confidence intervals can be obtained. An asymptotic $100(1-\nu)\%$ confidence region for $\boldsymbol{\theta}$ can be found by taking

$$\{\boldsymbol{\theta} : 2\{l(\hat{\boldsymbol{\theta}}, \hat{F}_{\hat{\boldsymbol{\theta}}}) - l(\boldsymbol{\theta}, \hat{F}_{\boldsymbol{\theta}})\} \le \chi^2_{|\boldsymbol{\theta}|}(\nu)\}. \tag{4.1}$$

A confidence interval for an individual component, θ_k of $\boldsymbol{\theta}$ can be found in a similar way by inverting a profile likelihood ratio test with respect to θ_k, which profiles out the other components of $\boldsymbol{\theta}$ as well as F. Alternatively, standard errors can be obtained by computing the Hessian of the profile likelihood through a finite difference approximation.
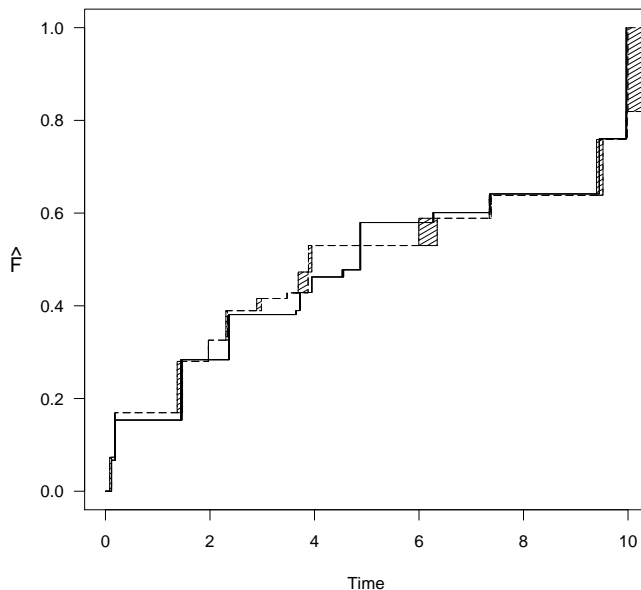
Figure 2: Comparison of the NPMLE for the misclassified data (solid line and solid shading) and the NPMLE for the interval censored data (dashed line and cross hatched shading).

5   Numerical Investigation

To illustrate the differences between the estimation of the NPMLE for interval censored data and the NPMLE for misclassified data a dataset of $n = 100$ subjects is simulated. The sampling times are random with independent gap times distributed $U(0, 2)$ up to a fixed administrative censoring time at $t = 10$. From the true classifications, $I(T_i < X_{ij})$, standard interval censored survival data, consisting of pairs $(L_j, R_j]$, can be obtained. In addition misclassified data, taking the misclassification probabilities to be fixed in time as $\alpha = 0.05$ and $\beta = 0.1$, is generated. The failure time distribution is estimated, first assuming (α, β) are fixed and known and second by jointly estimating (α, β). Figure 2 shows the estimates for a particular simulated dataset.

Applying the ICM and proposed constrained Newton algorithms for α and β at their true values, the convergence of the constrained Newton algorithm was considerably faster than that of ICM. Convergence, assessed based upon a criterion of $\max d_i < 1 \times 10^{-8}$, was reached in 27 iterations using constrained Newton compared to 281 iterations with ICM. This also corresponded to a CPU time of 0.28s for constrained Newton compared to 10.59s for ICM. The precise comparison may depend on the efficiency of the implementations

Table 1: Proportion of simulated likelihood ratio statistics above nominal χ_2^2 upper quantiles

		Nominal upper quantile			
n	0.5	0.2	0.1	0.05	0.01
100	0.562	0.256	0.138	0.070	0.014
500	0.536	0.255	0.134	0.067	0.016

of the PAV and NNLS algorithms. Here, there should be reasonable comparability since in each case Fortran routines ported into **R** via the packages *Iso* and *nnls*, respectively. Similarly, in each case the starting point was chosen such that all mass in F was in the single support region j for which $\sum_{i=1}^{n} \gamma_{ij}$ was maximized.

The number of increments of the NPMLE is similar for the interval censored and the misclassified data. However, the regions of support are typically much narrower for the misclassified data because they are always bounded by consecutive unique ordered sampling times.

Over 1000 samples of $n = 100$, the median number of increments of the NPMLE was 15 for both the standard interval censored data and the misclassified data. The mean width of support regions, excluding those bounded by ∞, was 0.07 for standard interval censored data and 0.01 for the misclassified data.

Table 1 compares the distribution of the profile likelihood ratio statistic, $2\{l(\hat{\boldsymbol{\theta}}, \hat{F}_{\hat{\boldsymbol{\theta}}}) - l(\boldsymbol{\theta}_0, \hat{F}_{\boldsymbol{\theta}_0})\}$, testing a null hypothesis that $\boldsymbol{\theta}_0 = (\alpha, \beta) = (0.05, 0.1)$, across 1000 samples of $n = 100$ and $n = 500$ against the asymptotic χ_2^2 distribution. This shows that the likelihood ratio test is slightly anti-conservative which implies that constructed confidence regions using (4.1) would tend to have lower than nominal coverage.

6   Example: Onset of Cardiac Allograft Vasculopathy

As an illustrative example, an analysis of data on the time of onset of cardiac allograft vasculopathy (CAV) among post-heart-transplantation patients is presented. CAV is the narrowing of arteries of the transplanted heart and is asymptomatic in its early stages. Patients are diagnosed through angiographic surveillance which is known to misdiagnose in some cases.

The data consist of the apparent presence or absence of CAV at a series of patient specific examination times. Diagnoses were made on the basis of 1304 angiograms taken from 562 patients in their first 6 years after transplanation, with individual patients having between 1 and 5 angiograms. It is assumed that the

donated organs are healthy such that the patients will be CAV free at the time of transplantation.

Observations continue to take place after the first positive diagnosis so it is possible to jointly estimate the time-to-onset distribution as well as the false positive and false negative rates of the angiogram.

Table 2: Estimates and confidence intervals of false positive and false negative CAV diagnosis probabilities for models based on a non-parametric failure time distribution and an exponential failure time distribution.

	NPMLE		Exponential	
	Estimate	95% CI	Estimate	95% CI
α	0.022	(0.007, 0.045)	0.021	(0.008, 0.053)
β	0.073	(0.015, 0.147)	0.090	(0.038, 0.197)

Table 2 gives the estimated diagnostic error rates using the NPMLE and also based on a parametric model assuming a constant hazard of progression, implying exponentially distributed onset times. The estimated error probabilities are broadly similar - in each case there is estimated to be a greater risk that the angiogram will fail to detect the presence of CAV, than to over-diagnose it. The estimated false negative probability, β, is slightly lower for the non-parametric model.

Figure 3 gives the NPMLE of the onset distribution compared to the exponential estimate from the parametric model. There is some suggestion of an increasing hazard of onset since the non-parametric estimate is somewhat lower than the parametric in the period 1 to 2 years post-transplantation.

7   Discussion

In addition to allowing the misclassification probabilities to depend on time, a relatively straightforward extension would be to allow dependence on explanatory variables. A more challenging extension is to incorporate the effect of explanatory variables into the failure time distribution F. One approach would be to adopt an accelerated failure time model where $\log T_j = \mathbf{Z}_j' \boldsymbol{\theta} + \epsilon_j$ for a p-dimensional covariate $\mathbf{Z}_j$ with corresponding p-dimensional regression coefficient $\boldsymbol{\theta}$, where ϵ_j are independent and identically distributed residuals with unknown distribution $\tilde{F}$. In the standard interval censored case, Rabinowitz et al (1995) considered estimation via a family of score statistics. This approach can be adapted for misclassified data. Let $X_{ij,\mathbf{u}} = \log X_{ij} - \mathbf{Z}_j' \mathbf{u}$ and $\tilde{\gamma}_{ij} = P(\mathbf{Y}_j | \epsilon_j \in (l_i, r_i])$, where $(l_i, r_i]$ denote the left and right boundaries of
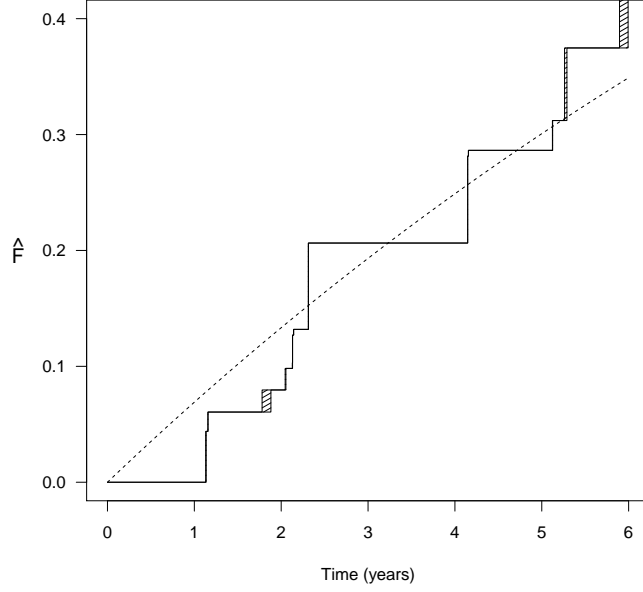
Figure 3: NPMLE for probability of disease onset (solid line) and parametric exponential estimate (dashed line) for CAV data.

the support regions of the NPMLE for $\tilde{F}$ given a particular value, $\mathbf{u}$, of the regression coefficient. Estimation can be based upon a score statistic of the form

$$\mathbf{S}(\mathbf{u}) = \sum_{j=1}^{n} \mathbf{Z}_j \frac{\sum_{i=1}^{n_j+1} \{g(F(X_{ij,\mathbf{u}})) - g(F(X_{ij-1,\mathbf{u}}))\} \tilde{\gamma}_{ij}}{\sum_{i=1}^{n_j+1} (\{F(X_{ij,\mathbf{u}}) - F(X_{ij-1,\mathbf{u}})\} \tilde{\gamma}_{ij})}$$

where $X_{i0,\mathbf{u}} = -\infty$ and $X_{in_i+1,\mathbf{u}} = \infty$ and g is some function with domain $[0,1]$ satisfying $g(0) = g(1) = 0$, with this condition ensuring $E_{\mathbf{u}}(\mathbf{S}(\mathbf{u})) = \mathbf{0}$. The efficiency of this estimator will depend on the choice of g. An adaptive procedure for choosing g, similar to that proposed by Rabinowitz et al (1995), may be possible, but is beyond the scope of this paper.

In applications of the model, the assumption of independence of the observations conditional on the failure time is likely to be key. In principle, a more general form for $\gamma_{ij}(\alpha, \beta)$ could be adopted allowing the individual classifications to be correlated, for instance via a shared multivariate binary copula. However, it is not clear how this would affect the support set of F or the joint identifiability of F and $\boldsymbol{\theta}$.

In the application on onset of CAV, an implicit assumption that censoring due to death was made. The data used was restricted to the first 6 years of follow-up, where death from CAV is relatively rare. However, a more appropriate method of dealing with this issue, particularly if data with greater follow-up were used,

would be to adopt an illness-death type model with misclassified illness status (Sharples et al, 2003; Teeple and Brown, 2015). A fully non-parametric approach in this case would require the extension of the methods of Frydman and Szarek (2009) for three-state illness-death Markov models.

References

Abrevaya J, Huang J (2005) On the bootstrap of the maximum score estimator. Econmetrica 73:1175–1204

Balasubramanian R, Lagakos S (2003) Estimation of a failure time distribution based on imperfect diagnostic tests. Biometrika 90:171–182

Banerjee M, Wellner J (2001) Likelihood ratio tests for monotone functions. Annals of Statistics 29:1699–1731

Banerjee M, Wellner J (2005) Confidence intervals for current status data. Scandinavian Journal of Statistics 32:405–424

Barlow R, Bartholomew D, Bremner J, Brunk H (1972) Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression. Wiley

Betensky R, Rabinowitz D, Tsiatis A (2001) Computationally simple accelerated failure time regression for interval censored data. Biometrika 88:703–711

Böhning D (1985) Numerical estimation of a probability measure. Journal of Statistical Planning and Inference 11:57–69

Böhning D, Schlattmann P, Dietz E (1996) Interval censored data: A note on the nonparametric maximum likelihood estimator of the distribution function. Biometrika 83:462–466

Espeland M, Platt O, Gallagher D (1989) Joint estimation of incidence and diagnostic error rates from irregular longitudinal data. Journal of the American Statistical Association 84:972–979

Frydman H, Szarek M (2009) Nonparametric estimation in a Markov 'illness-death' process from interval censored observations with missing intermediate transition status. Biometrics 65:143–151

van de Geer S (2003) Asymtotic theory for maximum likelihood in nonparametric mixture models. Computational Statistics and Data Analysis 41:453–464

Groeneboom P, Wellner JA (1992) Information Bounds and Nonparametric Maximum Likelihood Estimation, Springer Science & Business Media, vol 19. Birkhäuser, Basel, Switzerland

Grüger J, Kay R, Schumacher M (1991) The validity of inferences based on incomplete observations in disease state models. Biometrics 47:595–605

Jongbloed G (1998) The iterative convex minorant algorithm for nonparametric estimation. Journal of Computational and Graphical Statistics 7(3):310–321

Laird N (1978) Nonparametric maximum likelihood estimation of a mixing distribution. Journal of the American Statistical Association 73:805–811

Lawson C, Hanson R (1974) Solving Least Squares Problems. Englewood Cliffs: Prentice-Hall

Lesperance M, Kalbfleisch J (1992) An algorithm for computing the nonparametric mle of a mixing distribution. Journal of the American Statistical Association 87:120–126

Lindsay BG (1983) The geometry of mixture likelihoods: A general theory. Annals of Statistics 11(1):86–94

Maathuis M, Hudgens M (2011) Nonparametric inference for competing risks current status data with continuous, discrete or grouped observation times. Biometrika 98:325–340

McKeown K, Jewell N (2010) Misclassification of current status data. Lifetime Data Analysis 16:215–230

Murphy S, van der Vaart A (1997) Semiparametric likelihood ratio inference. Annals of Statistics 25:1471–1509

Ng MP (2002) A modification of Peto's nonparametric estimation of survival curves for interval-censored data. Biometrics 58:439–442

Politis D, Romano J (1994) Large sample confidence regions based on subsamples under minimal assumptions. Annals of Statistics 22:2031–2050

Rabinowitz D, Tsiatis A, Aragon J (1995) Regression with interval-censored data. Biometrika 82:501–513

Richardson B, Hughes J (2000) Product limit estimation for infectious disease data when the diagnostic test for the outcome is measured with uncertainty. Biostatistics 1:341–354

Sal y Rosas V, Hughes J (2011) Nonparametric and semiparametric analysis of current status data subject to outcome misclassification. Statistical Communications in Infectious Diseases 3(1):7

Sen B, Xu G (2015) Model based bootstrap mmethod for interval censored data. Computational Statistics and Data Analysis 81:121–129

Sharples LD, Jackson CH, Parameshwar J, Wallwork J, Large SR (2003) Diagnostic accuracy of coronary angiography and risk factors for postheart-transplant cardiac allograft vasculopathy. Transplantation 76(4):679–682

Tang R, Banerjee M, Kosorok M (2012) Likelihood based inference for current status data on a grid: A boundary phenomenon and an adaptive inference procedure. Annals of Statistics 40:45–72

Teeple EA, Brown ER (2015) Adjusting for time-dependent sensitivity in an illness-death model, with application to mother-to-child transmission of hiv. Statistics in Medicine 34:1277–1292

Turnbull B (1976) The empirical distribution function with arbitrarily grouped, censored and truncated data. Journal of the Royal Statistical Society: Series B 38:290–295

Wang Y (2007) On fast computation of the non-parametric maximum likelihood estimate of a mixing distribution. Journal of the Royal Statistical Society: Series B 69:185–198

Wang Y (2010) Maximum likelihood computation for fitting semiparametric mixture models. Statistics and Computing 20:75–86

Wellner J, Zhan Y (1997) A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. Journal of the American Statistical Association 92:945–959

Zhu L, Tong X, Sun J (2008) A transformation approach for the analysis of interval-censored failure time data. Lifetime Data Analysis 14:167–178