Ding-Geng (Din) Chen, Jianguo (Tony) Sun, and Karl E. Peace

Interval-Censored Time-to-Event Data: Methods and Applications

List of Figures

List of Tables

Contents

Chapter 1

Current Status Data in the 21st Century: Some Interesting Developments

Moulinath Banerjee

Department of Statistics, University of Michigan, Ann Arbor MI 48109, USA

Abstract

We revisit some important developments in the analysis of current status data and related interval censoring models over the past decade.

1.1 Introduction

This article aims to revisit some of the important advances in the analysis of current status data over the past decade. It is not my intention to be exhaustive since interest (and research) in current status and interval censored data has grown steadily in the recent past and it would be difficult for me to do justice to all the activity in this area in a chapter (of reasonable length) without being cursory. I will concern myself primarily with some problems that are closest to my own interests, describe some of the relevant results and discuss some open problems and conjectures. Before starting out, I would like to acknowledge some of the books and reviews in this area that I have found both enlightening and useful: the book on semiparametric information bounds and nonparametric maximum likelihood estimation by Groeneboom and Wellner (1992), the review by Huang and Wellner (1997), the review of current status data by Jewell and van der Laan (2003) and last but not least, the book on interval censoring by Sun (2006).

The current status model is one of the most well-studied survival models in statistics. An individual at risk for an event of interest is monitored at a particular observation time, and an indicator of whether the event has occurred is recorded. An interesting feature of this kind of data is that the NPMLE (nonparametric maximum likelihood estimator) of the distribution function (F) of the event time converges to the truth at rate $n^{1/3}$ (n, as usual, is sample size) when the observation time is a continuous random variable. Also, under mild conditions on the event-time distribution, the (pointwise) limiting distribution of the estimator in this setting is the non-Gaussian Chernoff's distribution. This is in contrast to right- censored data where the underlying survival function can be estimated nonparametrically at rate $\sqrt{n}$ under right-censoring and is pathwise norm-differentiable in the sense of van der Vaart (1991), admitting regular estimators and normal limits. On the other hand, when the status time in current status data has a distribution with finite support, the model becomes parametric (multinomial) and the event time distribution can be estimated at rate $\sqrt{n}$. The current status model which goes back to Ayer et al. (1955), van Eeden (1956), van Eeden (1957) was subsequently studied in Turnbull (1976) in a more general framework and asymptotic properties for the nonparametric maximum likelihood estimator (NPMLE) of the survival distribution were first obtained by Groeneboom (1987) (but see also Groeneboom and Wellner (1992)) and involved techniques radically different from those used in 'classical' survival analysis with right censored data.

In what follows, I will emphasize the following: the development of asymptotic likelihood ratio inference for current status data and its implications for estimating monotone functions in general, an area I worked on with Jon Wellner at the turn of the century and then on my own and with graduate students, extensions of these methods to more general forms of interval censoring, the technical challenges that come into play when there are multiple observation times on an individual and some of the (consequently) unresolved queries in these models, the recent developments in the study of current status data under competing risks, the development of smoothed procedures for inference in the current status model, adaptive estimation for current data on a grid, current status data with outcome mis-classification and semiparametric modeling of current status data.

1.2 Likelihood based inference for current status data

Consider the classical current status data model. Let $\{T_i, U_i\}_{i=1}^n$ be n i.i.d. pairs of non-negative random variables where T_i is independent of U_i. One can think of T_i as the (unobserved) failure time of the i'th individual, i.e. the time at which this individual succumbs to a disease or an infection. The individual is inspected at time U_i (known) for the disease/infection and one observes $\Delta_i = 1\{T_i \leq U_i\}$, their *cur-*

rent status. The data we observe is therefore $\{\Delta_i, U_i\}_{i=1}^n$. Let F be the distribution function of T and G that of U. Interest lies in estimating F. Let t_0 be an interior point in the support of F, assume that F and G are continuously differentiable in a neighborhood of t_0 and that $f(t_0), g(t_0) > 0$. Let $\hat{F}_n$ denote the NPMLE of F that can be obtained using the PAVA (Robertson et al. (1988)). Then, Theorem 5.1 of Groeneboom and Wellner (1992) shows that:

$$n^{1/3}(\hat{F}_n(t_0) - F(t_0)) \to_d \left(\frac{4\, F(t_0)\,(1 - F(t_0))\, f(t_0)}{g(t_0)} \right)^{1/3} \mathbb{Z}\,, \qquad (1.1)$$

where $\mathbb{Z} = \arg\min_{t \in \mathbb{R}} \{W(t) + t^2\}$, with $W(t)$ being standard two-sided Brownian motion starting from 0. The distribution of (the symmetric random variable) $\mathbb{Z}$ is also known as Chernoff's distribution, having apparently arisen first in the work of Chernoff (1964) on the estimation of the mode of a distribution. A so-called 'Wald-type' confidence interval can be constructed, based on the above result. Letting $\hat{f}(t_0)$ and $\hat{g}(t_0)$ denote consistent estimators of $f(t_0)$ and $g(t_0)$ respectively and $q(\mathbb{Z}, p)$ the p'th quantile of $\mathbb{Z}$, an asymptotic level $1 - \alpha$ CI for $F(t_0)$ is given by:

$$\left[\hat{F}_n(t) - n^{-1/3}\,\hat{C}\,q(\mathbb{Z}, \alpha/2)\;,\;\; \hat{F}_n(t) + n^{-1/3}\,\hat{C}\,q(\mathbb{Z}, \alpha/2)\right]\;. \qquad (1.2)$$

where

$$\hat{C} \equiv \left(\frac{4\,\hat{F}_n(t_0)\,(1 - \hat{F}_n(t_0))\,\hat{f}(t_0)}{\hat{g}(t_0)} \right)^{1/3}$$

consistently estimates the constant C sitting in front of $\mathbb{Z}$ in (1.1). One of the main challenges with the above interval is that it needs consistent estimation of $f(t_0)$ and $g(t_0)$. Estimation of g is possible via standard density estimation techniques since an i.i.d. sample of G is at our disposal. However, estimation of f is significantly more difficult. The estimator $\hat{F}_n$ is piecewise constant and therefore non-differentiable. One therefore has to smooth $\hat{F}_n$. As is shown in Groeneboom, Jongbloed and Witte (2010), a paper I will come back to later, even under the assumption of a second derivative for F in the vicinity of t_0, an assumption *not required* for the asymptotics of the NPMLE above, one obtains only an (asymptotically normal) $n^{2/7}$ consistent estimator of f. This is (unsurprisingly) much slower than the usual $n^{2/5}$ rate in standard density estimation contexts. Apart from the slower rate, note that the performance of $\hat{f}$ in a finite sample can depend heavily on bandwidth selection.

The above considerations then raise a natural question: can we prescribe confidence intervals that obviate the need to estimate these nuisance parameters? Indeed, this is what set Jon and me thinking of alternative solutions to the problem around 2000. That the usual Efron–type n out of n bootstrap is unreliable in this situation was already suspected; see, for example, the introduction of Delgado et al. (2001). While the m out of n bootstrap or its variant, subsampling, works in this situation, the selection of m is tricky and analogous to a bandwidth selection problem, which our goal was to avoid. As it turned out, in this problem likelihood ratios would come to the rescue.

The possible use of likelihood ratios in the current status problem was motivated by the then-recent work of Murphy and van der Vaart (1997) and Murphy and van der Vaart (2000) on likelihood ratio inference for the finite-dimensional parameter in regular semiparametric models. Murphy and van der Vaart showed that in semiparametric models, the likelihood ratio statistic (LRS) for testing $H_0 : \theta = \theta_0$ against its complement, θ being a pathwise norm-differentiable finite dimensional parameter in the model, converges under the null hypothesis to a χ^2 distribution with the number of degrees of freedom matching the dimensionality of the parameter. This result, which is analogous to what happens in purely parametric settings, provides a convenient way to construct confidence intervals via the method of inversion: an asymptotic level $1 - \alpha$ confidence set is given by the set of all $\theta^{'}$ for which the LRS for testing $H_{0,\theta'} : \theta = \theta^{'}$ (against its complement) is no larger than the $(1 - \alpha)$'th quantile of a χ^2 distribution. This is a clean method as nuisance parameters need not be estimated from the data; in contrast, the Wald type confidence ellipsoids that use the asymptotic distribution of the MLE would require estimating the information matrix. Furthermore, likelihood ratio based confidence sets are more 'data-driven' than the Wald type sets which necessarily have a pre-specified shape and satisfy symmetry properties about the MLE. An informative discussion of the several advantages of likelihood ratio based confidence sets over their competitors is available in Chapter 1 of Banerjee (2000).

In the current status model, the LRS relevant to constructing confidence sets for $F(t_0)$ would test $H_0 : F(t_0) = \theta_0$ against its complement. Is there an asymptotic distribution for the LRS in this problem? Is the distribution parameter-free? In particular, is it χ^2? As far as the last query is concerned, the χ^2 distribution for likelihood ratios is connected to the differentiability of the finite dimensional parameter in the sense of van der Vaart (1991); however, $F(t_0)$ is not a differentiable functional in the interval censoring model. But even if the limit distribution of the LRS (if one exists) is different, this does not preclude the possibility of it being parameter-free. Indeed, this is precisely what Jon and I established in Banerjee and Wellner (2001). We found that a particular functional of $W(t) + t^2$, which we call $\mathbb{D}$ (and which is therefore parameter-free) describes the large sample behavior of the LRS in the current status model. This asymptotic pivot can therefore be used to construct confidence sets for $F(t_0)$ by inversion. In subsequent work, I was able to show that the distribution $\mathbb{D}$ is a 'non-standard' or 'non-regular' analogue of the χ_1^2 distribution in nonparametric monotone function estimation problems and can be used to construct pointwise confidence intervals for monotone functions (via likelihood ratio based inversion) in a broad class of problems; see Banerjee (2000), Banerjee and Wellner (2001), Banerjee and Wellner (2005), Sen and Banerjee (2007), Banerjee (2007) and Banerjee (2009) for some of the important results along these lines. The first three references deal with the current status model in detail, the fourth to which I return later provides inference strategies for more general forms of interval-censoring and the last two deal with extensions to general monotone function models.

Let me now dwell briefly on the LRS for testing $F(t_0) = \theta_0$ in the current status model. The log-likelihood function for the observed data $\{\Delta_i, U_i\}_{i=1}^n$, up to an additive term not involving F, is readily seen to be:

$$
\begin{aligned}
L_n(F) &= \sum_{i=1}^{n} \left[\Delta_i \, \log F(U_i) + (1-\Delta_i) \, \log(1-F(U_i))\right] \\
&= \sum_{i=1}^{n} \left[\Delta_{(i)} \, \log F(U_{(i)}) + (1-\Delta_{(i)}) \, \log(1-F(U_{(i)}))\right],
\end{aligned}
$$

where $U_{(i)}$ is the i'th smallest of the U_j's and $\Delta_{(i)}$ its corresponding indicator. The LRS is then given by:

$$LRS(\theta_0) = 2\left[L_n(\hat{F}_n) - L_n(\hat{F}_n^0)\right],$$

where $\hat{F}_n$ is the NPMLE and $\hat{F}_n^0$ the *constrained* MLE under the null hypothesis $F(t_0) = \theta_0$. Let $\hat{F}_n(U_{(i)}) = v_i$ and $\hat{F}_n^0(U_{(i)}) = v_i^0$. It can be then shown, via the Fenchel conditions that characterize the optimization problems involved in finding the MLEs, that:

$$\{v_i\}_{i=1}^n = \arg \min_{s_1 \le s_2 \le \ldots \le s_n} \sum_{i=1}^{n} \left[\Delta_{(i)} - s_i\right]^2, \tag{1.3}$$

and that

$$\{v_i^0\}_{i=1}^n = \arg \min_{s_1 \le s_2 \le \ldots \le s_m \le \theta_0 \le s_{m+1} \le \ldots \le s_n} \sum_{i=1}^{n} \left[\Delta_{(i)} - s_i\right]^2, \tag{1.4}$$

where $U_{(m)} \le t_0 \le U_{(m+1)}$. Thus, $\hat{F}_n$ and $\hat{F}_n^0$ are also solutions to least squares problems. They are also extremely easy to compute using the PAV algorithm, having nice geometrical characterizations as *slopes of greatest convex minorants*.

To describe these characterizations we introduce some notation. First, for a function g from an interval I to $\mathbb{R}$, the greatest convex minorant or GCM of g will denote the supremum of all convex functions that lie below g. Note that the GCM is itself convex. Next, consider a set of points in $\mathbb{R}^2$, $\{(x_0, y_0), (x_1, y_1 1), \ldots, (x_k, y_k)\}$, where $x_0 = y_0 = 0$ and $x_0 < x_1 < \ldots < x_k$. Let $P(x)$ be the leftcontinuous function such that $P(x_i) = y_i$ and $P(x)$ is constant on (x_{i-1}, x_i). We will denote the vector of slopes (left-derivatives) of the GCM of $P(x)$, at the points $(x_1, x_2, \ldots, x_k)$, by slogcm$\{(x_i, y_i)\}_{i=0}^k$. The GCM of $P(x)$ is, of course, also the GCM of the function that one obtains by connecting the points $\{(x_i, y_i)\}_{i=0}^k$successively, by means of straight lines. Next, consider the so-called CUSUM (cumulative sum) 'diagram' given by $\{i/n\ ,\ \sum_{i=1}^n \Delta_{(i)}/n\}_{i=0}^n$. Then:

$$\{v_i\}_{i=1}^n = \text{slogcm}\{i/n\ ,\ \sum_{j=1}^{i} \Delta_{(j)}/n\}_{i=0}^n,$$

while

$$\{v_i^0\}_{i=1}^n = \left(\text{slogcm}\{i/n\ ,\ \sum_{j=1}^{i} \Delta_{(j)}/n\}_{i=0}^m \wedge \theta_0\ ,\ \text{slogcm}\{i/n\ ,\ \sum_{j=1}^{i} \Delta_{(m+j)}/n\}_{i=0}^{n-m} \vee \theta_0\right).$$

The maximum and minimum in the above display are interpreted as being taken component wise. The limiting versions of the MLEs (appropriately centered and scaled) have similar characterizations as in the above displays. It turns out that for determining the behavior of $LRS(\theta_0)$, only the behavior of the MLEs in a shrinking neighborhood of the point t_0 matters. This is a consequence of the fact that $D_n \equiv \{t : \hat{F}_n(t) \neq \hat{F}_n^0(t)\}$ is an interval around t_0 whose length is $O_p(n^{-1/3})$. Interest therefore centers on the processes:

$$X_n(h) = n^{1/3}(\hat{F}_n(t_0 + h\, n^{-1/3}) - F(t_0)) \;\text{ and }\; Y_n(h) = n^{1/3}(\hat{F}_n^0(t_0 + h\, n^{-1/3}) - F(t_0))$$

for h in compacts. The point h corresponds to a generic point in the interval D_n. The distributional limits of the processes X_n and Y_n are described as follows: For a real-valued function f defined on $\mathbb{R}$, let $\mathrm{slogcm}(f, I)$ denote the lefthand slope of the GCM of the restriction of f to the interval I. We abbreviate $\mathrm{slogcm}(f, \mathbb{R})$ to $\mathrm{slogcm}(f)$. Also define:

$$\mathrm{slogcm}^0(f) = (\mathrm{slogcm}(f, (-\infty, 0]) \wedge 0) 1(-\infty, 0] + (\mathrm{slogcm}(f, (0, \infty)) \vee 0) 1(0, \infty)\,.$$

For positive constants c, d let $X_{c,d}(h) = c\,W(h) + d\,h^2$. Set $g_{c,d}(h) = \mathrm{slogcm}(X)(h)$ and $g^0_{c,d}(h) = \mathrm{slogcm}^0(X)(h)$. Then, for every positive K,

$$(X_n(h), Y_n(h)) \to_d (g_{a,b}(h), g^0_{a,b}(h)) \text{ in } L_2[-K, K] \times L_2[-K, K]\,, \qquad (1.5)$$

where $L_2[-K, K]$ is the space of real-valued square-integrable functions defined on $[-K, K]$ while

$$a = \left(\sqrt{\frac{F(t_0)\,(1 - F(t_0))}{g(t_0)}}\right) \text{ and } b = \frac{f(t_0)}{2}\,.$$

These results can be proved in different ways, by using 'switching relationships' developed by Groeneboom (as in Banerjee (2000)) or through continuous mapping arguments (as developed in more general settings in Banerjee (2007)). Roughly speaking, appropriately normalized versions of the CUSM diagram converge in distribution to the process $X_{a,b}(h)$ in a strong enough topology that renders the operators slogcm and slogcm^0 continuous. The MLE processes X_n, Y_n are representable in terms of these two operators acting on the normalized CUSUM diagram and the distributional convergence then follows via continuous mapping. While I don't go into the details of the representation of the MLEs in terms of these operators, their relevance is readily seen by examining the displays characterizing $\{v_i\}$ and $\{v_i^0\}$ above. Note, in particular, the dichotomous representation of $\{v_i^0\}$ depending on whether the index is less than or greater than m and the constraints imposed in each segment via the max and min operations, which is structurally similar to the dichotomous representation of slogcm^0 depending on whether one is to the left or the right of 0.

I will not provide a detailed derivation of $LRS(\theta_0)$ in this review. Detailed proofs are available both in Banerjee (2000) and Banerjee and Wellner (2001) where it is shown that

$$LRS(\theta_0) \to_d \mathbb{D} \equiv \int \{(g_{1,1}(h))^2 - (g^0_{1,1}(h))^2\}\, dh\,.$$

However, I will illustrate why $\mathbb{D}$ has the particular form above by resorting to a residual sum of squares statistic (RSS) which leads naturally to this form. So, for the moment let's forget $LRS(\theta_0)$ and view the current status model as a *binary regression model*; indeed, the conditional distribution of Δ_i given U_i is Bernoulli(U_i) and given the U_i's (which we now think of as covariates), the Δ_i's are conditionally independent. Consider now the simple RSS for testing $H_0 : F(t_0) = \theta_0$. The least squares criterion is given by

$$LS(F) = \sum_{i=1}^{n} [\Delta_i - F(U_i)]^2 = \sum_{i=1}^{n} [\Delta_{(i)} - F(U_{(i)})]^2 \,.$$

The displays (1.3) and (1.4) show that $\hat{F}_n$ is the least squares estimate of F under no constraints apart from the fact that the estimate has to be increasing and that $\hat{F}_n^0$ is the least squares estimate of F under the additional constraint that the estimate assumes the value θ_0 at t_0. Hence, the RSS for testing $H_0 : F(t_0) = \theta_0$ is given by

$$RSS \equiv RSS(\theta_0) = \sum_{i=1}^{n} [\Delta_i - \hat{F}_n^0(U_i)]^2 - \sum_{i=1}^{n} [\Delta_i - \hat{F}_n(U_i)]^2 \,.$$

Before analyzing $RSS(\theta_0)$ we introduce some notation. Let $\mathcal{I}_n$ denote the set of indices such that $\hat{F}_n(U_{(i)}) \neq \hat{F}_n^0(U_{(i)})$. Then, note that the $U_{(i)}$'s in $\mathcal{I}_n$ live in the set D_n and are the only $U_{(i)}$'s that live in that set. Next, let $\mathbb{P}_n$ denote the empirical measure of $\{\Delta_i, U_i\}_{i=1}^n$. For a function $f(\delta, u)$ defined on the domain of (Δ_1, U_1), by $\mathbb{P}_n\, f$ we mean $n^{-1} \sum_{i=1}^n f(\Delta_i, U_i)$. The function f is allowed to be a random function. Similarly, if P denotes the joint distribution of (Δ_1, U_1) by Pf we mean $\int f\, dP$. This is *operator notation* and used extensively in the empirical process literature. Now,

$$\begin{aligned}
RSS(\theta_0) &= \sum_{i=1}^{n} [\Delta_{(i)} - \hat{F}_n^0(U_{(i)})]^2 - \sum_{i=1}^{n} [\Delta_{(i)} - \hat{F}_n(U_{(i)})]^2 \\
&= \sum_{i=1}^{n} [(\Delta_{(i)} - \theta_0) - (\hat{F}_n^0(U_{(i)}) - \theta_0)]^2 - \sum_{i=1}^{n} [(\Delta_{(i)} - \theta_0) - (\hat{F}_n(U_{(i)}) - \theta_0)]^2 \\
&= \sum_{i \in \mathcal{I}_n} (\hat{F}_n^0(U_{(i)}) - \theta_0)^2 - \sum_{i \in \mathcal{I}_n} (\hat{F}_n(U_{(i)}) - \theta_0)^2 \\
&\quad +2 \sum_{i \in \mathcal{I}_n} (\Delta_{(i)} - \theta_0)\, (\hat{F}_n(U_{(i)}) - \theta_0) - 2 \sum_{i \in \mathcal{I}_n} (\Delta_{(i)} - \theta_0)\, (\hat{F}_n^0(U_{(i)}) - \theta_0) \\
&= \sum_{i \in \mathcal{I}_n} (\hat{F}_n(U_{(i)}) - \theta_0)^2 - \sum_{i \in \mathcal{I}_n} (\hat{F}_n^0(U_{(i)}) - \theta_0)^2 \,,
\end{aligned}$$

where this last step uses the facts that:

$$\sum_{i \in \mathcal{I}_n} (\Delta_{(i)} - \theta_0)\, (\hat{F}_n(U_{(i)}) - \theta_0) = \sum_{i \in \mathcal{I}_n} (\hat{F}_n(U_{(i)}) - \theta_0)^2$$

and

$$\sum_{i \in \mathcal{I}_n} (\Delta_{(i)} - \theta_0)\,(\hat{F}_n^0(U_{(i)}) - \theta_0) = \sum_{i \in \mathcal{I}_n} (\hat{F}_n^0(U_{(i)}) - \theta_0)^2 \,.$$

For the case of $\hat{F}_n^0$ this equality is an outcome of the fact that $\mathcal{I}_n$ can be decomposed into a number of consecutive blocks of indices say $B_1, B_2, \ldots, B_r$ on each of which $\hat{F}_n^0$ is constant (denote the constant value ob B_j by w_j) and furthermore, on each block B_j such that $w_j \neq \theta_0$, we have for each $k \in B_j$,

$$w_j = \hat{F}_n^0(U_{(k)}) = \frac{\sum_{i \in B_j} \Delta_{(i)}}{n_j} \,,$$

where n_j is the size of B_j. A similar phenomenon holds for $\hat{F}_n$. The equalities in the above two displays then follow by writing the sum over $\mathcal{I}_n$ as a double sum where the outer sum is over the blocks and the inner sum over the i's in a single block. We have

$$\begin{aligned}
RSS(\theta_0) &= \sum_{i \in \mathcal{I}_n} (\hat{F}_n(U_{(i)}) - \theta_0)^2 - \sum_{i \in \mathcal{I}_n} (\hat{F}_n^0(U_{(i)}) - \theta_0)^2 \\
&= n\,\mathbb{P}_n \left[\{(\hat{F}_n(u) - F(t_0))^2 - (\hat{F}_n^0(u) - F(t_0))^2\}\, 1\{u \in D_n\} \right] \\
&= n^{1/3}\,(\mathbb{P}_n - P) \left[\{(n^{1/3}(\hat{F}_n(u) - F(t_0)))^2 - (n^{1/3}(\hat{F}_n^0(u) - F(t_0)))^2\}\, 1\{u \in D_n\} \right] \\
&\quad + n^{1/3}\, P \left[\{(n^{1/3}(\hat{F}_n(u) - F(t_0)))^2 - (n^{1/3}(\hat{F}_n^0(u) - F(t_0)))^2\}\, 1\{u \in D_n\} \right] .
\end{aligned}$$

Empirical processes arguments show that the first term is $o_p(1)$ since the random function that sits as the argument to $n^{1/3}(\mathbb{P}_n - P)$ can be shown to be eventually contained in a Donsker class of functions with arbitrarily high pre-assigned probability. Hence,

$$\begin{aligned}
RSS(\theta_0) &= n^{1/3} \int_{D_n} \{(n^{1/3}(\hat{F}_n(u) - F(t_0)))^2 - (n^{1/3}(\hat{F}_n^0(u) - F(t_0)))^2\}\, dG(t) \\
&= n^{1/3} \int_{D_n} \{(n^{1/3}(\hat{F}_n(u) - F(t_0)))^2 - (n^{1/3}(\hat{F}_n^0(u) - F(t_0)))^2\}\, g(t) dt \\
&= \int_{n^{1/3}(D_n - t_0)} \Big\{ (n^{1/3}(\hat{F}_n(t_0 + h\, n^{-1/3}) - F(t_0)))^2 \\
&\qquad - (n^{1/3}(\hat{F}_n^0(t_0 + h\, n^{-1/3}) - F(t_0)))^2 \Big\}\, g(t_0 + h\, n^{-1/3}) dh \\
&= \int_{n^{1/3}(D_n - t_0)} (X_n^2(h) - Y_n^2(h))\, g(t_0)\, dh + o_p(1) \,.
\end{aligned}$$

Now, using (1.5) along with the fact that the set $n^{1/3}(D_n - t_0)$ is eventually contained in a compact set with arbitrarily high probability, conclude that

$$RSS(\theta_0) \to_d g(t_0) \int \{(g_{a,b}(h))^2 - (g_{a,b}^0(h))^2\}\, dh \,.$$

There are some nuances involved in the above distributional convergence which we skip. The next step is to invoke Brownian scaling to relate $g_{a,b}$ and $g^0_{a,b}$ to the 'canonical' slope-of-convex-minorant processes $g_{1,1}$ and $g^0_{1,1}$ and use this to express the limit distribution aboove in terms of these canonical processes. See Pages 1724-1725 of Banerjee and Wellner (2001) for the exact nature of the scaling relations from which it follows that

$$\int \{(g_{a,b}(h))^2 - (g^0_{a,b}(h))^2\}\, dh \equiv_d a^2 \int \{(g_{1,1}(h))^2 - (g^0_{1,1}(h))^2\}\, dh\,.$$

It follows from the definition of a^2 that

$$RSS(\theta_0) \to_d \theta_0\,(1-\theta_0)\,\mathbb{D}\,.$$

Thuss $RSS/\theta_0(1-\theta_0)$ is an asymptotic pivot and confidence sets can be obtained via inversion in the usual manner. Note that the inversion does not involve estimation of $f(t_0)$. Now, the RSS is not quite the LRS for testing $F(t_0) = \theta_0$ though it is intimately connected to it. Firstly, the RSS can be interpreted as a *working likelihood ratio statistic* where instead of using the binomial log-likelihood we use a normal log-likelihood. Secondly, up to a scaling factor, $RSS(\theta_0)$ is asymptotically equivalent to $LRS(\theta_0)$. Indeed, from the derivation of the asymptotics for $LRS(\theta_0)$ which involves Taylor expansions one can see that:

$$\frac{RSS(\theta_0)}{\theta_0\,(1-\theta_0)} = LRS(\theta_0) + o_p(1)\,;$$

the Taylor expansions give a second order quadratic approximation to the Bernoulli likelihood, effectively reducing $LRS(\theta_0)$ to RSS. The third order term in the expansion can be neglected as in asymptotics for the MLE and likelihood ratios in classical parametric settings.

I should point out here that the form of $\mathbb{D}$ also follows from considerations involving an asymptotic testing problem where one observes a process $X(t) = W(t)+F(t)$ with $F(t)$ being the primitive of a monotone function f and W being standard Brownian motion on $\mathbb{R}$. This is an asymptotic version of the 'signal + noise' model where the 'signal' corresponds to f and 'noise' can be viewed as $d\,W(t)$ (the point of view is that Brownian motion is generated by adding up little bits of noise; think of the convergence of a random walk to Brownian motion under appropriate scaling). Taking $F(t) = t^2$, which gives Brownian motion plus quadratic drift and corresponds to $f(t) = 2t$, consider the problem of testing $H_0 : f(0) = 0$ against its complement based on an observation of a sample path of X. Thus, the null hypothesis constrains a monotone function at a point, similar to what we have considered thus far. Wellner (2003) shows that an appropriately defined likelihood ratio statistic for this problem is given *precisely* by $\mathbb{D}$ using Cameron-Martin-Girsanov's theorem followed by an integration by parts argument.

On the methodological front, a detailed investigation of the likelihood ratio based intervals in comparison to other methods for current status data was undertaken in Banerjee and Wellner (2005) and their behavior was seen to be extremely satisfactory. Among other things, the simulations strongly indicate that in a zone of rapid change of the distribution function the likelihood ratio method is significantly more reliable than competing methods (unless good parametric fits to the data were available). As subsequent investigation in the current status and closely related models has shown, if the underlying distribution function is expected to be fairly erratic, the likelihood ratio inversion method is generally a very reliable choice.

1.3 More general forms of interval censoring

With current status data, each individual is tested only once to ascertain whether the event of interest has transpired. However, in many epidemiological studies there are multiple follow up times for each individual and, in fact, the number of follow up times may vary from individual to individual. Such models are called *mixed-case interval censoring* models, a term that seems to have originated in the work of Schick and Yu (2000) who dealt with the properties of the NPMLE in these models. In this section I will describe to what extent the ideas of the previous section for current status data extend to mixed-case interval censoring models and what challenges remain. It turns out that one fruitful way to view mixed-case models is through the notion of panel count data which is described below.

Suppose that $N = \{N(t) : t \geq 0\}$ is a counting process with mean function $EN(t) = \Lambda(t)$, K is an integer-valued random variable and $T = T_{k,j},\ j = 1, ..., k, k = 1, 2, ...$ is a triangular array of potential observation times. It is assumed that N and (K, T) are independent, that K and T are independent and $T_{k,j-1} \leq T_{k,j}$ for $j = 1, ..., k$, for every k; we interpret $T_{k,0}$ as 0. Let $X = (N_K, T_K, K)$ be the observed random vector for an individual. Here K is the number of times that the individual was observed during a study, $T_{K,1} \leq T_{K,2} \leq \ldots \leq T_{K,K}$ are the times when they were observed and $N_K = \{N_{K,j} \equiv N(T_{K,j})\}_{j=1}^{K}$ are the observed counts at those times. The above scenario specializes easily to the mixed-case interval censoring model, when the counting process is $N(t) = 1(S \leq t)$, S being a positive random variable with distribution function F and independent of (T, K). To understand the issues with mixed-case interval censoring it is best to restrict to Case-2 interval censoring where K is identically 2. For this case, I use slightly different notation, denoting $T_{2,1}$ and $T_{2,2}$ by U and V respectively. With n individuals, our (i.i.d) data can be written as $\{\Delta_i, U_i, V_i\}_{i=1}^{n}$ where $\Delta_i = (\Delta_i^{(1)}, \Delta_i^{(2)}, \Delta_i^{(3)})$ and $\Delta_i^{(1)} = 1(S_i \leq U_i)$, $\Delta_i^{(2)} = 1(U_i < S_i \leq V_i)$ and $\Delta_i^{(3)} = 1(V_i < S_i) \equiv 1 - \Delta_i^{(1)} - \Delta_i^{(2)}$. Here, S_i is the survival time of the i'th individual. The likelihood function for Case 2 censoring is given by:

$$L_n = \prod_{i=1}^{n} F(U_i)^{\Delta_i^{(1)}} \left(F(V_i) - F(U_i)\right)^{\Delta_i^{(2)}} \left(1 - F(V_i)\right)^{\Delta_i^{(3)}},$$

and the corresponding log-likelihood by

$$l_n = \sum_{i=1}^{n} \left\{ \Delta_i^{(1)} \log F(U_i) + \Delta_i^{(2)} \log(F(V_i) - F(U_i)) + \Delta_i^{(3)} \log(1 - F(V_i)) \right\} .$$

Now, let $t_0 \equiv 0 < t_1 < t_2 \ldots < t_J$ denote the ordered distinct observation times. If (U, V) has a continuous distribution then, of course, $J = 2n$ but in general this may not be the case. Now, consider the rank function R on the set of U_i's and V_i's, i.e. $R(U_i) = s$ if $U_i = t_s$ and $R(V_i) = p$ if $V_i = t_p$. Then,

$$l_n = \sum_{i=1}^{n} \left\{ \Delta_i^{(1)} \log F(t_{R(U_i)}) + \Delta_i^{(2)} \log(F(t_{R(V_i)}) - F(t_{R(U_i)})) + \Delta_i^{(3)} \log(1 - F(t_{R(V_i)})) \right\} .$$

Now, l_n as a function in $(F(t_1), F(t_2), \ldots, F(t_J))$ is concave and thus, finding the NPMLE of F boils down to maximizing a concave function over a convex cone as in the current status problem. However, the structure of l_n is now considerably more involved than in the current status model. If we go back to l_n in the current status model we see that it is the sum of n univariate concave functions with the i'th function involving $F(U_{(i)})$ and the corresponding response $\Delta_{(i)}$; thus we have a separation of variables. With the Case 2 log-likelihood this is no longer the case as terms of the form $\log(F(t_i) - F(t_j))$ enter the likelihood and l_n no longer has an additive (separated) structure in the $F(t_i)$'s. The non-separated structure in Case 2 interval censoring leads to some complications: firstly, there is no longer an explicit solution to the NPMLE via the PAVA; rather, $\hat{F}_n$ has a *self-induced* characterization as the slope of the GCM of a stochastic process that depends on $\hat{F}_n$ itseld. See, for example, Chapter 2 of Groeneboom and Wellner (1992). The computation of $\hat{F}_n$ relies on the ICM (iterative convex minorant) algorithm that is discussed in Chapter 3 of the same book and was subsequently modified for effective implementation in Jongbloed (1998) where the Case 2 log-likelihood was used as a test example. Secondly, and more importantly, the non-separated log-likelihood is quite difficult to handle. Groeneboom (1996) had to use some very hard analysis to get around the lack of separation and establish the pointwise asymptotic distribution of $\hat{F}_n$ in the Case 2 censoring model. Under certain regularity conditions for which we refer the reader to the original manuscript, the most critical of which is that $V - U$ is larger than some positive number with probability one (this condition is very natural in practical applications since there is always a minimal gap between the first and second inspection times), Groeneboom (1996) shows that $n^{1/3}(\hat{F}_n(t_0) - F(t_0))$ converges in distribution to a constant times $\mathbb{Z}$.

It is, then, natural to be curious as to whether the LRS for testing $F(t_0) = \theta_0$ is again asymptotically characterized by $\mathbb{D}$. Unfortunately, this has still not been established. One key reason behind this is the fact that the computation of the constrained MLE of F under the hypothesis $F(t_0) = \theta_0$ can no longer be decomposed into two separate optimization problems, in contrast to the current status model in the previous section or the monotone response models in Banerjee (2007). A self-induced characterization of the constrained NPMLE is still available but computationally more difficult to implement. Furthermore the techniques for separated monotone function models that enable us to get a handle on the relationships between the unconstrained and constrained MLEs of F and, in particular the set on which they differ (which plays a crucial role in studying the LRS) do not seem to work either. Nevertheless, some rough heuristics (which involve some conjectures about the relation of $\hat{F}_n$ to $\hat{F}_n^0$) indicate that $\mathbb{D}$ may, yet again, be the distributional limit of the LRS. As a first step, one would want to implement a progam to compute the LRS in the Case 2 model and check whether there is empirical agreement between the quantiles from its distribution and the quantiles of $\mathbb{D}$.

The complexities with the Case 2 model are of course present with mixed-case censoring. Song (2004) studies estimation with mixed-case interval censored data, characterizes and computes the NPMLE for this model and establishes asymptotic properties like consistency, global rates of convergence and an asymptotic minimax lower bound but does not have a pointwise limit distribution result analogous to that in Groeneboom (1996). The question then is whether one can postulate an asymptotically pivotal method (as in the current status case) for estimation of $F(t_0)$ in the mixed case model. Fortunately, Bodhi Sen and I were able to provide a positive answer to this question in Sen and Banerjee (2007) by treating the mixed case model as a special case of the panel data model introduced at the beginning of this section.

Our approach was to think of mixed-case interval censored data as data on a one-jump counting process with counts available only at the inspection times and to use a pseudo-likelihood function based on the marginal likelihood of a Poisson process to construct a pseudo-likelihood ratio statistic for testing null hypotheses of the form $H_0 : F(t_0) = \theta_0$. We showed that under such a null hypothesis the statistic converges to a pivotal quantity. Our method was based on an estimator originally proposed by Sun and Kalbfleisch (1995) whose asymptotic properties, under appropriate regularity conditions, were studied in Wellner and Zhang (2000). Indeed, our point of view, that the interval censoring situation can be thought of as a one-jump counting process to which, consequently, the results on the pseudo-likelihood based estimators can be applied, was motivated by the latter work.

The pseudo-likelihood method starts by *pretending* that $N(t)$, the counting process introduced above is a non-homogeneous Poisson process. Then the marginal distribution of N(t) is given by $prN(t) = k = \exp\{-\Lambda(t)\}\Lambda(t)^k/k!$ for non-negative integers k. Note that, under the Poisson process assumption, the successive counts on an individual $(N_{K,1}, N_{K,2}, ...)$, conditional on the $T_{K,j}$s, are actually dependent. However, we ignore the dependence in writing down a likelihood function for the data. Letting $\{N_{K_i}, T_{K_i}, K_i\}_{i=1}^n$ denote our data $\underline{X}$, our likelihood function, conditional on the T_{K_i}'s and K_i's (whose distributions do not involve Λ) is:

$$L_n^{ps}(\Lambda \mid \underline{X}) = \prod_{i=1}^{n} \prod_{j=1}^{K_i} \exp\{-\Lambda(T_{K_i,j}^{(i)})\} \frac{\Lambda(T_{K_i,j}^{(i)})^{N_{K_i,j}^{(i)}}}{N_{K_i,j}^{(i)}!} ,$$

and the corresponding log-likelihood up to an irrelevant additive constant is

$$l_n^{ps}(\Lambda \mid \underline{X}) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \left\{ N_{K_i,j}^{(i)} \, \log \Lambda(T_{K_i,j}^{(i)}) - \Lambda(T_{K_i,j}^{(i)}) \right\} .$$

Denote by $\hat{\Lambda}_n$ and $\hat{\Lambda}_n^0$ respectively the unconstrained and constrained pseudo-MLEs of Λ with the latter MLE computed under the constraint $\Lambda(t_0) = \theta_0$. As Λ is increasing, isotonic estimation techniques apply; furthermore, it is easily seen that the log-likelihood has an additive separated structure in terms of the ordered distinct observation times for the n individuals. Techniques similar to the previous section can therefore be invoked to study the behavior of the pseudo-LRS. Theorem 1 of Sen and Banerjee (2007) shows that

$$2 \left\{ l_n^{ps}(\hat{\Lambda}_n \mid \underline{X}) - l_n^{ps}(\hat{\Lambda}_n^0 \mid \underline{X}) \right\} \to_d \frac{\sigma^2(t_0)}{\Lambda(t_0)} \, \mathbb{D} \, .$$

The above result provides an easy way of constructing a likelihood-ratio-based confidence set for $F(t_0)$ in the mixedcase interval censoring model. This is based on the observation that under the mixed-case interval censoring framework, where the counting process $N(t)$ is $1(S \leq t)$ with S following F independently of (K, T), the pseudo-likelihood ratio statistic in the above display converges to $(1 - \theta_0) \, \mathbb{D}$ under the null hypothesis $F(t_0) = \theta_0$. Thus, an asymptotic level $(1 - \alpha)$ confidence set for $F(t_0)$ is $\{\theta : (1 - \theta) \, \mathrm{PLRS}_n(\theta) \leq q(\mathbb{D}, 1 - \alpha)\}$, where $q(\mathbb{D}, 1 - \alpha)$ is the $(1 - \alpha)$th quantile of $\mathbb{D}$ and $\mathrm{PLRS}_n(\theta)$ is the pseudo-likelihood ratio statistic computed under the null hypothesis $H_{0,\theta} : F(t_0) = \theta$. Once again, nuisance parameter estimation has been avoided. An alternative confidence interval could be constructed by considering the asymptotic distribution of $n^{1/3}(\hat{F}_{n,pseudo}(t_0) - F(t_0))$ where $\hat{F}_{n,pseudo}$ is the pseudo-MLE of F but this involves a very hard to estimate nuisance parameter; see the remarks following Theorem 4.4 in Wellner and Zhang (2000).

Relying, as it does, on the marginal likelihoods, the pseudo-likelihood approach *ignores* the dependence among the counts at different times points. An alternative approach is based on considering the full likelihood for a non-homogeneous Poisson process as studied in Section 2 of Wellner and Zhang (2000). The MLE of Λ based on the full likelihood was characterized in this paper; owing to the lack of separation of variables in the full Poisson likelihood (similar to the true likelihood for mixed case interval censoring), the optimization of the likelihood function as well as its analytical treatment are considerably more complicated. In particular, the analytical behavior of the MLE of Λ based on the full likelihood does not seem to be known. Wellner and Zhang (2000) prove an asymptotic result for a 'toy' estimator obtained by applying one step of the iterative convex minorant algorithm starting from the true Λ; while an asymptotic equivalence between the MLE and the toy estimator is conjectured, it remains to be proved. Simulation studies show that the MSE of the MLE is smaller than that of the pseudo-MLE (unsurprisingly) when the underlying counting process is Poisson. A natural query in the context of the above discussion is the behavior of the LRS for testing an hypothesis of the form $\Lambda(t_0) = \theta_0$ using the full Poisson likelihood ratio statistic. This has not been studied either computationally or theoretically. Once again, one is tempted to postulate $\mathbb{D}$ up to a constant but whether one gets an asymptotically pivotal quantity in the mixed-case model with this alternative statistic is unclear.

Thus, there are three conceivably different ways of constructing CIs via likelihood ratio inversion for $F(t_0)$ in the mixed-case model. The first is based on the the true likelihood ratio for this model, the second on the pseudo-likelihood method of Sen and Banerjee (2007) and the third on the full Poisson likelihood; in the last two cases, we think of mixed-case interval censored data as panel count data from a counting process. As of now, the second method is the only one that has been implemented and theoretically validated and appears to be the only asymptotically pivotal method for nonparametric estimation of F in the mixed-case model. However, there is need to investigate the other two approaches, as these may produce alternative and, potentially, better pivots in the sense that inversion of such pivots may lead to sharper confidence intervals as compared to the pseudo-LRS based ones.

1.4 Current status data with competing risks

The current status model in its simplest form, as discussed above, deals with failure of an individual or a system but does not take into account the cause of failure. However, data is often available not only on the status of an individual, i.e. whether they have failed or not at the time of observation, but also on the cause of failure. A classic example in the clinical setting is that of a woman's age at menopause, where the outcome of interest Δ is whether menopause has occurred, U is the age of the woman and the two competing causes for menopause are either natural or operative. More generally, consider a system with K (finite) components that will fail as soon as one of its component fails. Let T be time to failure, Y be the index of the component that fails and U the (random) observation time. Thus (T, Y) has a joint distribution that is completely specified by the sub-distribution functions $\{F_{0i}\}_{i=1}^K$ where $F_{0i}(t) = P(T \leq t, Y = i)$. The distribution function of T, say F_+ is simply $\sum_{i=1}^K F_{0i}$ and the survival function of T is $S(t) = 1 - F_+(t)$. Apart from U, we observe a vector of indicators $\Delta = (\Delta_1, \Delta_2, \ldots, \Delta_{K+1})$ where $\Delta_i = 1\{T \leq U, Y = i\}$ for $i = 1, 2, \ldots, K$ and $\Delta_{K+1} = 1 - \sum_{j=1}^K \Delta_j = 1\{T > U\}$. A natural goal is to estimate the sub-distribution functions, as well as F_+. Competing risks in the more general setting of interval-censored data was considered by Hudgens et al. (2001) and the more specific case of current status data was investigated by Jewell et al. (2003) . In what follows, I will restrict to two competing causes ($K = 2$) for simplicity of notation and understanding; everything extends readily to more general (finite) K but the case of infinitely many competing risks, the so-called 'continuous marks model', is dramatically different and I will touch upon it later.

Under the assumption that U is independent of (T, Y), the likelihood function for the data which comprises n i.i.d. observations $\{\Delta^j, U_j\}_{j=1}^n$, in terms of generic sub-distribution functions F_1, F_2 is

$$L_n(F_1, F_2) = \prod_{i=1}^{n} F_1(U_i)^{\Delta_1^i}, F_2(U_i)^{\Delta_2^i}\, S(U_i)^{\Delta_3^i}\,;$$

this follows easily from the observation that the conditional distribution of Δ given U is multinomial. Maximization of the above likelihood function is somewhat involved; as Jewell and van der Laan (2003) note, the general EM algorithm can be used for this purpose but is extremely slow. Jewell and Kalbfleisch (2004) developed a much faster iterative algorithm which generalizes the PAVA; the pooling now involves solving a polynomial equation instead of simple averaging, the latter being the case with standard current status data. We denote the MLE of (F_1, F_2) by $(\hat{F}_1, \hat{F}_2)$. A competing estimator is the so-called 'naive estimator' which was also studied in Jewell et al. (2003) and we denote this by $(\tilde{F}_1, \tilde{F}_2)$. Here $\tilde{F}_i = \max_F\, L_{ni}(F)$ where F is a generic sub-distribution function and

$$L_{ni}(F) = \prod_{k=1}^{n} F(U_k)^{\Delta_i^k}\,(1 - F(U_k))^{1-\Delta_i^k}\,. \tag{1.6}$$

Thus, the naive estimator separates the estimation problem into 2 separate well-known univariate current status problems and the properties of the naive estimator follow from the same arguments that work in the simple current status model. The problem, however, lies in that by treating Δ_1 and Δ_2 separately, a critical feature of the data is ignored and the natural estimate of F_+, $\tilde{F}_+ = \tilde{F}_1 + \tilde{F}_2$, may no longer be a proper distribution function (it can be larger than 1). Both the MLE and the naive estimator are consistent but the MLE turns out to be more efficient than the naive estimator as we will see below.

Groeneboom et al. (2008a) and Groeneboom et al. (2008b) develop the full asymptotic theory for the MLE and the naive estimators. The naive estimator, of course, converges pointwise at rate $n^{1/3}$ but figuring out the local rates of convergence of the MLEs of the sub-distribution functions takes much work. As Groeneboom et al. (2008a) note in their introduction, the proof of the local rate of convergence of $\hat{F}_1$ and $\hat{F}_2$ requires new ideas that go well beyond those needed for the simple current status model or general monotone function models. One of the major difficulties in the proof lies in the delicate handling of the system of sub-distribution functions. This requires an initial result on the convergence rate of $\hat{F}_+$ uniformly on a fixed neighborhood of t_0 and is accomplished in Theorem 4.10 of their paper. It shows that under mild conditions – $F_{0i}(t_0) \in (0, F_{0i}(\infty))$ for $i = 1, 2$ and the F_{0i}'s and G are continuously differentiable in a neighborhood of t_0 with the derivatives at t_0, $\{f_{0i}(t_0)\}_{i=1}^2$ and $g(t_0)$, being positive – for any $\beta \in (0, 1)$, there is a constant $r > 0$ such that

$$\sup_{t \in [t_0 - r, t_0 + r]} \frac{|\hat{F}_+(t) - F_+(t)|}{v_n(t - t_0)} = O_p(1)\,,$$

where $v_n(s) = n^{-1/3}\, 1(|s| \le n^{-1/3}) + n^{-(1-\beta)/3}|t|^\beta\, 1(|s| > n^{-1/3})$. Thus, the local rate of $\hat{F}_+$, the MLE of the distribution function, is the same as in the current status model (as the form of $v_n(s)$ for $|s| \leq n^{-1/3}$ shows), but outside of the local $n^{-1/3}$ neighborhood of t_0, the normalization changes (as the altered form of v_n shows). This result leads to some crucial bounds that are used in the proof of Theorem 4.17 of their paper where it is shown that given $\epsilon, M_1 > 0$, one can find $M, n_1 > 0$ such that for each i,

$$P\left(\sup_{h\in[-M_1,M_1]} n^{1/3}\,|\hat{F}_i(s+n^{-1/3}\,h) - F_{0i}(s)| > M\right) < \epsilon\,,$$

for all $n > n_1$ and s varying in a small neighborhood of t_0.

Groeneboom et al. (2008b) make further inroads into the asymptotics: they determine the pointwise limit distributions of the MLEs of the F_{0i}'s in terms of completely new distributions, the characterizations of which again require much difficult work. Let W_1 and W_2 denote a couple of correlated Brownian motions originating from 0 with mean 0 and covariances:

$$E(W_j(t)\,W_k(s)) = (|s| \wedge |t|)\,1\{st > 0\}\,\Sigma_{jk}, \quad s,t \in \mathbb{R}, 1 \leq j,k \leq 2\,,$$

with $\Sigma_{jk} = g(t_0)^{-1}\,[1\{j = k\}\,F_{0k}(t_0) - F_{0k}(t_0)\,F_{0j}(t_0)]$. Note the multinomial covariance structure of Σ; this is not surprising in the light of the observation that the conditional distribution of Δ given $U = t_0$ is $Multinomial(1, F_{01}(t_0), F_{02}(t_0), S(t_0))$. Consider the drifted Brownian motions (V_1, V_2) given by $V_i(t) = W_i(t) + (f_{0i}(t_0)/2)\,t^2$ for $i = 1, 2$. The limit distribution of the MLEs can be described in terms of certain complex functionals of the V_i's which have *self-induced characterizations*. Before describing the characterization, we introduce some notation: let $F_{03}(t) = 1 - F_+(t)$, $a_k = (F_{0k}(t_0))^{-1}$ for $k = 1, 2, 3$ and for a finite collection of functions $g_1, g_2, \ldots$, let g_+ denote the sum. Groeneboom et al. (2008b) show that there exist almost surely a unique pair of convex functions $(\hat{H}_1, \hat{H}_2)$ with right continuous derivatives $(\hat{S}_1, \hat{S}_2)$ satisfying:

(1) $$a_i\,\hat{H}_i(h) + a_3\,\hat{H}_+(h) \leq a_i\,V_i(h) + a_3\,V_+(h)\ , \quad i = 1,2 \ \text{ and } \ h \in \mathbb{R}$$

(2) $$\int \{a_i\,\hat{H}_i(h) + a_3\,\hat{H}_+(h) - a_i\,V_i(h) - a_3\,V_+(h)\}\,d\hat{F}_i(h) = 0 \quad i = 1,2\,,$$

and

(3) For each $M > 0$ and $i = 1, 2$, there are points $\tau_{1i} < -M$ and $\tau_{2i} > M$ such that $a_i\,\hat{H}_i(h) + a_3\,\hat{H}_+(h) = a_i\,V_i(h) + a_3\,V_+(h)$ for $h = \tau_{i1}, \tau_{i2}$.

The self-inducedness is clear from the above description as the defining properties of $\hat{H}_1$ and $\hat{H}_2$ have to be written in terms of their sum. The random functions $\hat{S}_i$ are the limits of the normalized sub-distribution functions as Theorem 1.8 of Groeneboom et al. (2008b) shows:

$$\{n^{1/3}(\hat{F}_i(t_0 + h\,n^{-1/3}) - F_{0i}(t_0))\}_{i=1}^{2} \to_d (S_1(h), S_2(h))$$

in the Skorohod topology on $D(\mathbb{R})^2$. Here $D(\mathbb{R})$ is the space of real-valued cadlag functions on $\mathbb{R}$ equipped with the topology of convergence in the Skorohod metric on compact sets. In particular, this yields convergence of finite-dimensional distributions: thus, $n^{1/3}(\hat{F}_i(t_0) - F_{0i}(t_0)) \to_d S_i(0)$ for each i. The proof of the above process convergence requires the local rate of convergence of the MLEs of F_{01} and F_{02} discussed earlier. It is somewhat easier to characterize the asymptotics of the naive estimator. Let $\tilde{H}_i$ denote the GCM of V_i and let $\tilde{S}_i$ denote the right derivative of $\tilde{H}_i$. Then,

$$\{n^{1/3}(\tilde{F}_i(t_0 + h\,n^{-1/3}) - F_{0i}(t_0))\}_{i=1}^2 \to_d (\tilde{S}_1(h), \tilde{S}_2(h))\,.$$

Groeneboom et al. (2008b) compare the efficiency of the MLE with respect to the naive estimator and a ‘scaled naive estimator’ which makes a scaling adjustment to the naive estimator when the sum of the components, i.e. $\tilde{F}_1 + \tilde{F}_2$ exceeds one at some point (see Section 4 of their paper). It is seen that the MLE is more efficient than its competitors, so the hard work in computing and studying the MLE pays off. It should be noted that while the MLE beats the naive estimators for the point wise estimation of the sub-distribution functions, estimates of smooth functionals of F_{0i}’s based on the MLEs and the naive estimators are both asymptotically efficient – see, Jewell et al. (2003) and Maathuis (2006). The discrepancy between the MLE and the naive estimator therefore manifests itself only in the estimation of non-smooth functionals like the value of the sub-distribution functions at a point.

Maathuis and Hudgens (2011) extend the work in the paper discussed above to current status competing risks data with discrete or grouped observation times. In practice, recorded observation times are often discrete, making the model with continuous observation times unsuitable. This leads them to investigate the limit behavior of the maximum likelihood estimator and the naive estimator in a discrete model in which the observation time distribution has discrete support, and a grouped model in which the observation times are assumed to be rounded in the recording process, yielding grouped observation times. They establish that the large sample behavior of the estimators in the discrete and grouped models is critically different from that in the smooth model (the model with continuous observation times): the maximum likelihood estimator and the naive estimator both converge locally at $\sqrt{n}$ rate, and have limiting Gaussian distributions. The Gaussian limits in their setting arise because they consider discrete distributions with a fixed countable support and in the case of grouping, a fixed countable number of groups irrespective of sample size n. A similar phenomenon in the context of simple current status data was observed in Yu et al. (1998). However, if the support of the discrete distribution or the number of groupings (in the grouped data case) are allowed to change with n, the properties of the estimators can be quite different, a point I will return to later.

Maathuis and Hudgens (2011) also discuss the construction of pointwise confidence intervals for the sub-distribution functions in the discrete and grouped models as well as the smooth model. They articulate several difficulties with using the limit distribution of the MLEs in the smooth model for setting confidence intervals, like nuisance parameter estimation as well as the lack of scaling properties of the limit. The usual n out of n or model-based bootstrap are both suspect, though the

m out of n bootstrap as well as subsampling can be expected to work. Maathuis and Hudgens suggest using inversion of the (pseudo) likelihood ratio statistic for testing $F_{0i}(t_0) = \theta$ using the pseudo-likelihood function in (1.6). This is based on the naive estimator and its constrained version under the null hypothesis. The likelihood ratio statistic can be shown to converge to $\mathbb{D}$ under the null hypothesis by methods similar to Banerjee and Wellner (2001). The computational simplicity of this procedure makes it attractive even though owing to the inefficiency of the naive estimator with respect to the MLE, these inversion based intervals will certainly not be optimal in terms of length. The behavior of the true likelihood ratio statistic in the smooth model for testing the value of a sub-distribution function at a point remains completely unknown and it is unclear whether it will be asymptotically pivotal. More recently, Werren (2011) has extended the results of Sen and Banerjee (2007) to mixed-case interval censored data with competing risks. She defines naive pseudo-likelihood estimator for the sub-distribution functions corresponding to the various risks using a working Poisson-process (pseudo) likelihood, proves consistency, derives the asymptotic limit distribution of the naive estimators and presents a method to construct pointwise confidence intervals for these sub-distribution functions using a pseudo-likelihood ratio statistic in the spirit of Sen and Banerjee (2007).

I end this section with a brief note on the current status continuous marks model. Let X be an event time and Y a jointly distributed continuous 'mark' variable with joint distribution F_0. In the current status continuous mark model, instead of observing (X, Y), we observe a continuous censoring variable U , independent of (X, Y) and the indicator variable $\Delta = 1\{X \leq U\}$. If $\Delta = 1$, we also observe the mark variable Y ; in case $= 0$ the variable Y is not observed. Note that this is precisely a continuous version of the competing risks model: the discrete random variable Y in the usual competing risks model has now been changed to a continuous variable. Maathuis and Wellner (2008) consider a more general version of this model where instead of current status censoring they have general case k censoring. They derive the MLE of F_0 and its almost sure limit which leads to necessary and sufficient conditions for consistency of the MLE. However, these conditions force a relation between the unknown distribution F_0 and G, the distribution of U. Since such a relation is typically not satisfied, the MLE is inconsistent in general in the continuous marks model. Inconsistency of the MLE can be removed by either discretizing the marks as in Maathuis and Wellner (2008); an alternative strategy is to use appropriate kernel smoothed estimators of F_0 (instead of the MLE) which will be discussed briefly in the next section.

1.5 Smoothed estimators for current status data

I first deal with the simple current status model that has been the focus of much of the previous sections using the same notation as in Section 1.2. While the MLE of F in the current status model does not require bandwidth specification and achieves the best possible pointwise convergence rate under minimal smoothness (F only needs to be continuously differentiable in a neighborhood of t_0, the point of interest), it is not the most optimal estimate in terms of convergence rates if one is willing to assume stronger smoothness conditions. If F is twice differentiable around t_0, it is not unreasonable to expect that appropriate estimators of $F(t_0)$ will converge faster than $n^{1/3}$. This is suggested, firstly, by results in classical nonparametric kernel estimation of densities and regression functions where kernel estimates of the functions of interest exhibit the $n^{2/5}$ convergence rate under a (local) twice-differentiability assumption on the functions and, secondly, by the work of Mammen (1991) on kernel based estimation of a smooth monotone function while respecting the monotonicity constraint. In a recent paper, Groeneboom et al. (2010) provide a detailed analysis of smoothed kernel estimates of F in the current status model.

Two competing estimators are proposed by Groeneboom et al. (2010): the MSLE, originally introduced by Eggermont and LaRiccia (2001) in the context of density estimation, which is a general likelihood-based M estimator and turns out to be automatically smooth, and the SMLE, which is obtained by convolving the usual MLE with a smooth kernel. If $\mathbb{P}_n$ denotes the empirical measure of the $\{\Delta_i, U_i\}$'s, the log-likelihood function can be written as:

$$l_n(F) = \int \{\delta \, \log F(u) + (1-\delta) \, \log(1 - F(u))\} \, d\mathbb{P}_n(\delta, u) \, .$$

For $i \in \{0, 1\}$ define the empirical sub-distribution functions

$$\mathbb{G}_{n,i}(u) = \frac{1}{n} \sum_{j=1}^{n} 1_{[0,u]\times\{i\}}(T_j, U_j) \, .$$

Note that $d\mathbb{P}_n(u, \delta) = \delta \, d\mathbb{G}_{n,1}(u) + (1-\delta) \, d\mathbb{G}_{n,0}(u)$. Now, consider a probability density k that has support $[-1, 1]$, is symmetric and twice continuously differentiable on $\mathbb{R}$, let K denote the corresponding distribution function, let $K_h(u) = K(u/h)$ and $k_h(u) = (1/h) \, k(u/h)$, where $h > 0$. Consider now, kernel-smoothed versions of the $\mathbb{G}_{n,i}$'s given by: $\hat{G}_{n,i}(t) = \int_{[0,t]} \hat{g}_{n,i}(u) \, du$ for $i = 0, 1$, where

$$\hat{g}_{n,i}(t) = \int k_h(t-u) \, d\mathbb{G}_{n,i}(u) \, .$$

Some minor modification is needed for $0 < t < h$ but as h is the bandwidth and will go to 0 with increasing n, the modification is on a vanishing neighborhood of 0. These smoothed versions of the $\mathbb{G}_{n,i}$'s lead to a smoothed version of the empirical measure given by

$$d\hat{P}_n(u, \delta) = \delta \, d\hat{G}_{n,1}(u) + (1-\delta) \, d\hat{G}_{n,0}(u) \, .$$

This can be used to define a smoothed version of the log-likelihood function, namely

$$l_n^S(F) = \int \{\delta \, \log F(u) + (1-\delta) \, \log(1 - F(u))\} \, d\hat{P}_n(\delta, u) \, .$$

The MSLE (maximum smoothed likelihood estimator), denoted by $\hat{F}_n^{MS}$, is simply the maximizer of l_n^S over all sub-distribution functions and has an explicit characterization as the slope of a convex minorant as shown in Theorem 3.1 of Groeneboom et al. (2010). Theorem 3.5 of this paper provides the asymptotic distribution of $\hat{F}_n^{MS}(t_0)$ under certain assumptions, which, in particular, require F and G to be three times differentiable at t_0: under a choice of bandwidth of the form $h \equiv h_n = c\,n^{-1/5}$, it is shown that $n^{2/5}(\hat{F}_n^{MS}(t_0) - F(t_0))$ converges to a normal distribution with a non-zero mean. Explicit expressions for this asymptotic bias as well as the asymptotic variance are provided. The asymptotics for $\hat{f}_n^{MS}$, the natural estimate of f which is obtained by differentiating $\hat{F}_n^{MS}$, are also derived; with a bandwidth of order $n^{-1/7}$, $n^{2/7}(\hat{f}_n^{MS}(t_0) - f(t_0))$ converges to a normal distribution with non-zero mean.

The construction of the SMLE (smoothed maximum likelihood estimator) simply alters the steps of smoothing and maximization. The raw likelihood, $l_n(F)$ is first maximized to get the MLE $\hat{F}_n$ which is then smoothed to get the SMLE:

$$\hat{F}_n^{SM}(t) = \int K_h(t-u)\, d\hat{F}_n(u)\,.$$

Again, under appropriate conditions and in particular, twice differentiability of F at t_0, $n^{2/5}(\hat{F}_n^{MS}(t_0) - F(t_0))$ converges to a normal limit with a non-zero mean when a bandwidth of order $n^{-1/5}$ is used and the asymptotic bias and variance can be explicitly computed. A comparison of this result to the asymptotics for $\hat{F}_n^{MS}$ shows that the asymptotic variance is the same in both cases; however, the asymptotic biases are unequal and there is no monotone ordering between the two. Thus, in some situations the MSLE may work better than the SMLE and vice-versa. Groeneboom et al. (2010) discuss a bootstrapped based method for bandwidth selection but do not provide simulation based evidence of the performance of their method.

The work in Groeneboom et al. (2010) raises an interesting question. Consider a practitioner who wants to construct a confidence interval for $F(t_0)$ in the current status model and let us suppose that the practitioner is willing to assume that F around t_0 is reasonably smooth (say three times differentiable). She could either use the likelihood ratio technique from Banerjee and Wellner (2005) or the smoothed likelihood approach of Groeneboom et al. (2010). The former technique would avoid bandwidth specification and also the estimation of nuisance parameters. The latter would need active bandwidth selection and also nuisance parameter estimation. In this respect, the likelihood inversion procedure is methodologically cleaner. On the other hand, because the smoothed estimators achieve a higher convergence rate ($n^{2/5}$ as opposed to $n^{1/3}$ obtained through likelihood based procedures), the CIs based on these estimators would be asymptotically shorter than the ones based on likelihood ratio inversion. So, there is a trade-off here. The $n^{2/15}$ faster rate of convergence of the smoothed MLE will start to show at large sample sizes, but at smaller sample sizes, bandwidth selection and the estimation of nuisance parameters from the data would introduce much more variability in the intervals based on the smoothed MLE. There is, therefore, a need for a relative study of these two procedures in terms of actual performance at different sample sizes.

Groeneboom et al. (2011) also use kernel smoothed estimators to remedy the inconsistency of the MLE in the continuous marks model under current status censoring. They develop a version of the MSLE in this model following similar ideas as in the above paper: the log-likelihood function for the observed data in this model can be written as an integral of a deterministic function involving f (the joint density of the event time and the continuous mark) and various operators acting on f, with respect to the empirical measure of the observed data. As before, the idea is to replace the empirical measure by a smoothed version to obtain a smoothed log-likelihood function which is then maximized over f to obtain $\hat{f}_n^{MS}$ and the corresponding joint distribution $\hat{F}_n^{MS}$. Consistency results are obtained for the MSLE using histogram-type smoothers for the observation time distribution but rigorous asymptotic results are unavailable. Heuristic considerations suggest yet again the $n^{2/5}$ rate of convergence with a normal limit under an appropriate decay-condition on the bin-width of the histogram smoother.

Smoothing methods have also been invoked in the study of current status data in the presence of covariate information. Van der Laan and Robins (1998) studied locally efficient estimation with current status data and time-dependent covariates. They introduce an inverse probability of censoring weighted estimator of the distribution of the failure time and of smooth functionals of this distribution which involves kernel smoothing. More recently, van der Vaart and van der Laan (2006) have studied estimation of the survival distribution in the current status model when high-dimensional and/or time-dependent covariates are available, and/or the survival events and censoring times are only conditionally independent given the covariate process. Their method of estimation consists of regularizing the survival distribution by taking the primitive function or smoothing, estimating the regularized parameter by using estimating equations, and finally recovering an estimator for the parameter of interest. Consider, for example, a situation where the event time T and the censoring time C are conditionally independent given a vector of covariates L; time dependence is not assumed here but the number of covariates can be large. Let $F(t \mid L)$ and $G(t \mid L)$ denote the conditional distributions of T and C given L, and $g(t \mid L)$ the density of T given L.The goal is to estimate $S(t) = 1 - E_L(F(t \mid L))$, the survival function of T based on i.i.d. realizations from (Δ, C, L) where $\Delta = 1\{T \leq C\}$. This is achieved via estimating equations of the form

$$\psi(F, g, r)(c, \delta, l) = \frac{r(c)(F(c \mid l) - \delta)}{g(c \mid l)} + \int_0^\infty r(s)\overline{F}(s \mid l)\, ds\,,$$

for some real-valued function r defined on $[0, \infty)$. Up to a constant, this is the efficient influence function for estimating the functional $\int_0^\infty r(s)S(s)\, ds$ in the model where $F(t \mid l) \equiv 1 - \overline{F}(t \mid l)$ and the distribution of L are left fully unspecified. One estimate of S suggested by van der Vaart and van der Laan is based on pure smoothing; namely:

$$S_{n,b}(t) = \mathbb{P}_n\, \psi(F_n, g_n, k_{b,t})$$

where F_n and g_n are preliminary estimates of F and g and $k_{b,t}(s) = k((s - t)/b)$, where k is a probability density supported on $[-1, 1]$ and $b \equiv b_n$ is a bandwidth that goes to 0 with increasing n. Th estimator $S_{n,b}(t)$ should be viewed as estimating $P_{F,g}\,\psi(F, g, k_{b,t}) = \int_0^\infty k_{b,t}(s)S(s)\,ds$ which converges to $S(t)$ as $b \to 0$. Under appropriate conditions on F_n and g_n as discussed in Section 2.1 of this paper and which should not be hard to satisfy, as well as mild conditions on the underlying parameters of the model, Theorem 2.1 of van der Vaart and van der Laan (2006) shows that with $b_n = b_1\,n^{-1/3}$, $n^{1/3}(S_{n,b_n}(t) - S(t))$ converges to a mean 0 normal distribution. Sections 2.2 and 2.3 of the paper discuss variants based on the same estimating equation; while Section 2.2 relies only on isotonization, Section 2.3 proposes an estimator combining isotonization and smoothing. This leads to estimators with lower asymptotic variance than in Section 2.1, but there are caveats as far as practical implementation is concerned and the authors note that more refined asymptotics would be needed to understand the bias-variance trade-off. Some discussion on constructing F_n and g_n is also provided, but there is no associated computational work to illustrate how these suggestions work for simulated and real data sets. It seems to me that there is scope here for investigating the implementability of the proposed ideas in practice.

1.6 Inference for current status data on a grid

While the literature on current status data is large, somewhat surprisingly, the problem of making inference on the event time distribution, F, when the observation times lie on a grid with multiple subjects sharing the same observation time had never been satisfactorily addressed. This important scenario, which transpires when the inspection times for individuals at risk are evenly spaced and multiple subjects can be inspected at any inspection time, is completely precluded by the assumption of a continuous observation time. One can also think of a situation where the observation times are all distinct but cluster into a number of distinct well-separated clumps with very little variability among the observation times in a single clump. For making inference on F in this situation, the assumption of a continuous observation time distribution would not be ideal and a better approximation might be achieved by considering all points within one clump to correspond to the same observation time; say, the mean observation time for that clump.

For simple current status data on a regular grid, say with K grid-points, sample size n and n_i individuals sharing the i'th grid-point as the common observation time, how does one construct a reliable confidence interval for the value of F at a grid-point of interest? What asymptotic approximations should the statistician use for $\hat{F}(t_g)$, where t_g is a grid-point and $\hat{F}$ the MLE? Some thought shows that this hinges critically on the size of n relative to K. If n is much larger than K and the number of individuals per time is high, the problem can be viewed as a parametric one and a normal approximation should be adequate. If n is 'not too large' relative to K, the normal approximation would be suspect and the usual Chernoff approximation may be more ideal. As Tang et al. (2011) show, one can view this problem in an asymptotic framework and the nature of the approximation depends heavily on how large $K = K(n)$ is, relative to n. Unfortunately, the rate of growth of $K(n)$ is unknown in practice; Tang et al. (2011) suggest a way to circumvent this problem by using a family of 'boundary distributions' indexed by a scale parameter $c > 0$ which provide correct approximations to the centered and scaled MLE: $n^{1/3}(\hat{F}(t_g) - F(t_g))$. Below, I briefly describe the proposed method.

Let $[a, b]$ $(a \geq 0)$ be the interval on which the time grid $\{a + \delta, a + 2\,\delta, \ldots, a + K\delta\}$ is defined (K is such that $a + (K + 1)\delta > b$) and let n_i be the number of individuals whose inspection time is $a + i\,\delta$. The MLE, $\hat{F}$, is easily obtained as the solution to a weighted isotonic regression problem with the n_i's acting as weights. Now, find $\hat{c}$ such that K is the largest integer not exceeding $(b - a)/\hat{c}n^{-1/3}$; this roughly equates the spacing of the grid-points to $\hat{c}\,n^{-1/3}$. Then, the distribution of $n^{1/3}(\hat{F}(t_g) - F(t_g))$ can be approximated by that of the distribution of a random variable that is characterized as the left-slope of the GCM of a real-valued stochastic process defined on the grid $\{\hat{c}j\}_{j \in \mathcal{Z}}$ (where $\mathcal{Z}$ is the set of integers) and depending on positive parameters α, β that can be consistently estimated from the data. Section 4 of Tang et al. (2011) provides the details. The random variable that provides the approximation is easy to generate since the underlying process that defines it is the restriction of a quadratically drifted Brownian motion to the grid $\{\hat{c}j\}_{j \in \mathcal{Z}}$. Tang et al. (2011) demonstrate the effectiveness of their proposed method through a variety of simulation studies.

As Tang et al. (2011) point out, the underlying principle behind their 'adaptive' method that adjusts to the intrinsic resolution of the grid can be extended to a variety of settings. Recall that Maathuis and Hudgens (2011) studied competing risks current status data under grouped or discrete observation times but did not consider settings where the size of the grid could depend on the sample size n. As a result, they obtained Gaussian-type asymptotics. But again, if the number of discrete observation times is large relative to the sample size, these Gaussian approximations become unreliable, as demonstrated in Section 5.1 of their paper. It would therefore be interesting to develop a version of the adaptive procedure in their case, a point noted both in Section 6 of Maathuis and Hudgens (2011) as well as in Section 6 of Tang et al. (2011). Extensions to more general forms of interval censoring as well as models incorporating covariate information should also be possible.

1.7 Current status data with outcome misclassification

There has been recent interest in the analysis of current status data where outcomes may be mis-classified. McKeown and Jewell (2010) discuss a number of biomedical and epidemiological studies where the current status of an individual, say a patient, is determined through a test which may not have full precision. In this case, the real current status is perturbed with some probability depending on the sensitivity and specificity of the test. We use their notation for this section. So, let T be the event time and C the censoring time. If perfect current status were available, we would have a sample from the distribution of (Y, C) where $Y = 1(T \leq C)$. Consider now, the misclassification model that arises from the following specifications:

$$P(\Delta = 1 \mid Y = 1) = \alpha \quad \text{and} \quad P(\Delta = 0 \mid Y = 0) = \beta\,.$$

The probabilities α, β are each assumed greater than 0.5, as will be the case for any realistic testing procedure. The observed data $\{\Delta_i, C_i\}_{i=1}^n$ is a sample from the distribution of (Δ, C). Interest lies, as usual, in estimating F, the distribution of T. The log-likelihood function for the observed data is:

$$\tilde{l}_n(F) = \sum_{i=1}^n \Delta_i \, \log(\gamma\, F(C_i) + (1-\beta)) + \sum_{i=1}^n (1-\Delta_i)\, \log\left(\beta - \gamma\, F(C_i)\right),$$

where $\gamma = \alpha + \beta - 1 > 0$. McKeown and Jewell provide an explicit characterization of $\hat{F}$, the MLE, in terms of a max-min formula. From the existing results in the monotone function literature it is clear that $n^{1/3}(\hat{F}(t_0) - F(t_0))$ is distributed asymptotically like a multiple of $\mathbb{Z}$, so they resort to construction of confidence intervals via the m out of n bootstrap. More recently, Sal y Rosas and Hughes (2011) have proposed new inference schemes for $F(t_0)$. They observe that the model of McKeown and Jewell (2010) is a *monotone response model* in the sense of Banerjee (2007) and therefore likelihood ratio inversion using the quantiles of $\mathbb{D}$ can be used to set confidence intervals for $F(t_0)$. Sal y Rosas and Hughes (2011) extend this model to cover situations where the current status of an individual may be determined using any one of k available laboratory tests with differing sensitivities and specificities. This introduces complications in the structure of the likelihood and the MLE must now be computed via the modified ICM of Jongbloed (1998). Confidence intervals for $F(t_0)$ in this model can be constructed via likelihood ratio inversion as before as this model is also, essentially, a monotone response model.

McKeown and Jewell (2010) also consider time varying misclassification as well as versions of these models in a regression setting while Sal y Rosas and Hughes (2011) deal with extensions to two-sample problems and a semiparametric regression version of the misclassification problem using the Cox proportional hazards model.

1.8 Semiparametric models and other work

The previous sections have dealt, by and large, with fully nonparametric models. There has, of course, been significant progress in semiparametric modeling of current status data over the past 10 years. One of the earliest papers is that of Shen (2000) who considers linear regression with current status data. The general linear regression model is of the form $Y_i = \beta^T X_i + \epsilon_i$ (an intercept term is included in the vector of regressors) and Shen deals with a situation where one observes $\Delta_i = 1\{Y_i \leq C_i\}$, C_i being an inspection time. The error ϵ_i is assumed independent of C_i and X_i while

C_i and Y_i are assumed conditionally independent given X_i. Based on observations $\{\Delta_i, C_i, X_i\}_{i=1}^n$, Shen (2000) develops a random-sieve likelihood based method to make inference on β and the error variance σ^2 without specifying the error distribution X; in fact an asymptotically efficient estimator of β is constructed. This model has close connections to survival analysis as a variety of survival models can be written in the form $h(Y_i) = \beta^T X_i + \epsilon_i$ for a monotone transformation h of the survival time Y_i. With ϵ_i following the extreme-value distribution $F(x) = 1 - e^{-e^x}$, the above model is simply the Cox PH model, where the function h determines the baseline hazard. When $F(x) = e^x/(1 + e^x)$, i.e. the logistic distribution, one gets the proportional odds model. Such models are also known as semiparametric linear transformation models and have been studied in the context of current status data by other authors.

Sun and Sun (2005) deal with the analysis of current status data under semiparametric linear transformation models for which they propose a general inference procedure based on estimating functions. They allow time-dependent covariates $Z(t)$ and model the conditional survival function of the failure time T as $S_Z(t) = g(h(t) + \beta^T Z(t))$ for a known continuous strictly decreasing function g and an unknown function h. This is precisely an extension of the models in the previous paragraph to the time-dependent covariate setting; with time-independent covariates, setting $g(t) = e^{-e^t}$ gives the Cox PH model and setting g to be the logistic distribution gives the proportional odds model. As in the previous paragraph $\Delta_i = 1\{T_i \leq C_i\}$ is recorded. Sun and Sun use counting process based ideas to construct estimates in situations where C is independent of (T, Z) and also when T and C are conditionally independent given Z. A related paper by Zhang et al. (2005) deals with regression analysis of interval-censored failure time data with linear transformation models. Ma and Kosorok (2005) consider a more general problem where a continuous outcome U is modeled as $H(U) = \beta^T Z + h(W) + e$, with H being an unknown monotone transformation, h an unknown smooth function, e has a known distribution function and $Z \in \mathbb{R}^d$, $W \in \mathbb{R}$ are covariates. The observed data is $X = (V, \Delta, Z, W)$ where $\Delta = 1(U \leq V)$. It is easily seen that this extends the models in Shen (2000) to incorporate a nonparametric covariate effect. Note however that in the more restricted set-up of Shen (2000), the distribution of e is not assumed known. Ma and Kosorok develop a maximum penalized log-likelihood estimation method for the parameters of interest and demonstrate, in particular, the asymptotic normality and efficiency of their estimate of β. A later paper, Ma and Kosorok (2006), studies adaptive penalized M-estimation with current status data. More recently, Cheng and Wang (2011) have generalized the approach of Ma and Kosorok (2005) to cover additive transformation models. In their model, $H(U) = \beta^T Z + \sum_{j=1}^d h_j(W_j) + \epsilon$ where the h_j's are smooth and can have varying degrees of smoothness and U is subjected to current status censoring by a random examination time V. In contrast to the approach adopted in Ma and Kosorok (2005), Cheng and Wang (2011) consider a B-spline based estimation framework and establish asymptotic normality and efficiency of their estimate of β.

Banerjee et al. (2006) study the Cox PH regression model with current status data. They develop an asymptotically pivotal likelihood ratio method to construct pointwise confidence sets for the conditional survival function of the event time T given time-independent covariates Z. In related work, Banerjee et al. (2009) study binary regression models under a monotone shape constraint on the nonparametric component of the regression function using a variety of link functions and develop asymptotically pivotal methods for constructing confidence sets for the regression function. Through the connection of these models to the linear transformation models with current status data as in Shen (2000), the techniques of Banerjee et al. (2009) can be used to prescribe confidence sets for the conditional survival function of T given X for a number of different error distributions which correspond to the link functions in the latter paper.

Semiparametric models for current status data in the presence of a 'cured' proportion in the population has also attracted interest. Lam and Xue (2005) use a mixture model that combines a logistic regression formulation for the probability of cure with a semiparametric regression model belonging to the flexible class of partly linear models for the time to occurrence of the event and propose sieved likelihood estimation. Ma (2009) has also considered current status data in the presence of a cured subgroup assuming that the cure probability satisfies a generalized linear model with a known link function, while for susceptible subjects, the event time is modeled using linear or partly linear Cox models. Likelihood based strategies are used. An extension, along very similar lines, to mixed case interval censored data is developed in Ma (2010). An additive risk model for the survival hazard for subjects susceptible to failure in the current status cure rate model is studied in Ma (2011).

The above survey should give an ample feel for the high level of activity in the field of current status (and more generally interval-censored) data in recent times. As I mentioned in the introduction, the goal of the exposition was not to be exhaustive but to be selective and as was admitted, the selection-bias was driven to some extent by my personal research interests. A substantial body of research in this area therefore remains uncovered; some examples include work on additive hazards regression with current status data initially studied in Lin et al. (1998) and pursued subsequently by Ghosh (2001) and Martinussen and Scheike (2002); computational algorithms for interval censored problems as developed by Gentlemen and Vandal (2001) and Vandal et al. (2005); inference for two sample problems with current status data and related models as developed by Zhang et al. (2001), Zhang (20006), Tong et al. (2007) and most recently in Groeneboom (2011) using a likelihood ratio based approach; current status data in the context of multistage/multistate models as studied in Datta and Sundaram (2006) and Lan and Datta (2010) and finally Bayesian approaches to the problem where interesting research has been carried out by D.B. Dunson, Bo Cai and Lianming Wang among others.

Bibliography

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *The Annals of Mathematical Statistics* **24**, 641–647.

Banerjee, M. (2000). *Likelihood Ratio Inference in Regular and Non-regular Problems*. Ph.D. thesis, University of Washington.

Banerjee, M. (2007). Likelihood based inference for monotone response models. *Annals of Statistics* **35**, 931–956.

Banerjee, M. (2009). Inference in exponential family regression models under certain shape constraints. In *Advances in Multivariate Statistical Methods, Statistical Science and Interdisciplinary Research*, pages 249–272. World Scientific.

Banerjee, M., Biswas, P., and Ghosh, D. (2006). A semiparametric binary regression model involving monotonicity constraints. *Scandinavian Journal of Statistics* **33**, 673–697.

Banerjee, M., Mukherjee, D., and Mishra, S. (2009). Semiparametric binary regression models under shape constraints with an application to indian schooling data. *Journal of Econometrics* **149**, 101–117.

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

Banerjee, M. and Wellner, J. A. (2005). Confidence intervals for current status data. *Scand. J. Statist.* **32**, 405–424. ISSN 0303-6898.

Cheng, G. and Wang, X. (2011). Semiparametric additive transformation models under current status data. *Electronic Journal of Statistics, to appear* .

Chernoff, H. (1964). Estimation of the mode. *Ann. Inst. Stat. Math* **16**, 31–41.

Datta, S. and Sundaram, R. (2006). Nonparametric marginal estimation in a multistage model using current status data. *Biometrics* **62**, 829–837.

Delgado, M., Rodriguez-Poo, J., and Wolf, M. (2001). Subsampling inference in cube root asymptotics with an application to manski's maximum score estimator. *Economics Letters* **73**, 241–250.

Eggermont, P. and LaRiccia, V. (2001). *Maximum Penalized Likelihood Estimation*. Springer, New York.

Gentlemen, A. and Vandal, A. (2001). Computational algorithms for censored data problems using intersection graphs. *JCGS* **10**, 403–421.

Ghosh, D. (2001). Efficiency considerations in the additive hazards model with current status data. *Statistica Neerlandica* **55**, 367–376.

Groeneboom, P. (1987). *Asymptotics for incomplete censored observations. Report 87-18*. Faculteit Wiskunde and Informatica, Universiteit van Amsterdam.

Groeneboom, P. (1996). Lectures on inverse problems. In *Lectures on Probability Theory and Statistics. Lecture Notes in Math, 1648*, pages 67–164. Springer, Berlin.

Groeneboom, P. (2011). Likelihood ratio type two-sample tests for current status data. *Scandinavian Journal of Statistics, submitted* .

Groeneboom, P., Jongbloed, G., and Witte, B. I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *The Annals of Statistics* **38**, 352–387.

Groeneboom, P., Jongbloed, G., and Witte, B. I. (2011). A maximum smoothed likelihood estimator in the current status continuous mark model. *Journal of Nonparametric Statistics, to appear* .

Groeneboom, P., Maathuis, M. H., and Wellner, J. A. (2008a). Current status data with competing risks: Consistency and rates of convergence of the mle. *The Annals of Statistics* **36**, 1031–1063.

Groeneboom, P., Maathuis, M. H., and Wellner, J. A. (2008b). Current status data with competing risks: Limiting distribution of the mle. *The Annals of Statistics* **36**, 1064–1089.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhaüser, Basel.

Huang, J. and Wellner, J. A. (1997). *Interval Censored Survival Data: A Review of Recent Progress. Eds. D. Lin and T. Fleming.* Springer-Verlag, New York.

Hudgens, M., Satten, G., and Longini, I. (2001). Nonparametric maximum likelihood estimation for competing risks survival data subject to interval-censoring and truncation. *Biometrics* **57**, 74–80.

Jewell, N. P. and Kalbfleisch, J. D. (2004). Maximum likelihood estimation of ordered multinomial parameters. *Biostatistics* **23**, 625–642.

Jewell, N. P. and van der Laan, M. (2003). Current status data: Review, recent developments and open problems. *Handbook of Statistics* **5**, 291–306.

Jewell, N. P., van der Laan, M., and Henneman, T. (2003). Nonparametric estimation from current status data with competing risks. *Biometrika* **90**, 183–197.

Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *JCGS* **7**, 310–321.

Lam, K. and Xue, H. (2005). A semiparametric regression cure model with current status data. *Biometrika* **92**, 573–586.

Lan, L. and Datta, S. (2010). Comparison of state occupation, entry, exit and waiting times in two or more groups based on current status data in a multistate model. *Statistics in Medicine* **29**, 906–914.

Lin, D. Y., Oakes, D., and Ying, Z. (1998). Additive hazards regression with current status data. *Biometrika* **85**, 289–298.

Ma, S. (2009). Cure model with current status data. *Statistica Sinica* **19**, 233–249.

Ma, S. (2010). Mixed case interval censored data with a cured subgroup. *Statistica Sinica* **20**, 1165–1181.

Ma, S. (2011). Additive risk model for current status data with a cured subgroup. *Ann. Inst. Stat. Math* **63**, 117–134.

Ma, S. and Kosorok, M. R. (2005). Penalized log-likelihood estimation for partly linear transformation models with current status data. *The Annals of Statistics* **33**, 2256–2290.

Ma, S. and Kosorok, M. R. (2006). Adaptive penalized m-estimation with current status data. *Annals of the Institute of Statistical Mathematics* **58**, 511–526.

Maathuis, M. (2006). *Nonparametric estimation for current status data with competing risks*. Ph.D. thesis, University of Washington.

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

Maathuis, M. and Wellner, J. A. (2008). Inconsistency of the mle for the joint distribution of interval censored survival times and continuous marks. *Scandinavian Journal of Statistics* **35**, 83–103.

Mammen, E. (1991). Estimating a smooth monotone regression function. *Annals of Statistics* **19**, 724–740.

Martinussen, T. and Scheike, T. H. (2002). Efficient estimation in additive hazards regression with current status data. *Biometrika* **8**, 649–658.

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

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

Murphy, S. A. and van der Vaart, A. (2000). On profile likelihood. *JASA* **95**, 449–485.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order restricted statistical inference*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Chichester: John Wiley & Sons Ltd. ISBN 0-471-91787-7.

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

Schick, A. and Yu, Q. (2000). Consistency of the gmle with mixed case interval-censored data. *Scandinavian Journal of Statistics* **27**, p. 45 Length: 11 pages.

Sen, B. and Banerjee, M. (2007). A pseudo–likelihood method for analyzing interval censored data. *Biometrika* .

Shen, X. (2000). Linear regression with current status data. *JASA, Theory and Methods* **95**, 842–852.

Song, S. (2004). Estimation with univariate 'mixed-case' interval censored data. *Statistica Sinica* **14**, 269–282.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer-Verlag, first edition.

Sun, J. and Kalbfleisch, J. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica* **5**, 279–290.

Sun, J. and Sun, L. (2005). Semiparametric linear transformation models for current status data. *The Canadian Journal of Statistics* **33**, 85–96.

Tang, R., Banerjee, M., and Kosorok, M. R. (2011). Likelihood inference for current status data on a grid: A boundary phenomenon and an adaptive inference procedure. *Annals of Statistics, to appear* .

Tong, X., Zhu, C., and Sun, J. (2007). Semiparametric regression analysis of two-sample current status data, with applications to tumorogenicity experiments. *The Canadian Journal of Statistics* **35**, 575–584.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological)* **38**, 290–295. ISSN 00359246.
URL `http://www.jstor.org/stable/2984980`

Van der Laan, M. and Robins, J. (1998). Locally efficient estimation with current status data and time-dependent covariates. *JASA* **93**, 693–701.

van der Vaart, A. (1991). On differentiable functionals. *The Annals of Statistics* **19**, 178–204.

van der Vaart, A. and van der Laan, M. (2006). Current status data with high-dimensional covariates. *The International Journal of Biostatistics* **2**, Issue 1, Article 9.

van Eeden, C. (1956). Maximum likelihood estimation of ordered probabilities. *Proceedings Koninklijke Nederlandse Akademie van Wetenschappen A* pages 444–455.

van Eeden, C. (1957). Maximum likelihood estimation of partially or completely ordered parameters. *Proceedings Koninklijke Nederlandse Akademie van Wetenschappen A* pages 128–136.

Vandal, A., Gentleman, R., and X, L. (2005). Constrained estimation and likelihood intervals for censored data. *The Canadian Journal of Statistics* **33**, 71–83.

Wellner, J. A. (2003). Gaussian white noise models: some results for monotone functions. In *Crossing Boundaries: Statistical Essays in Honor of Jack Hall*, pages 87–104. IMS Lecture Notes Monograph Series, 43.

Wellner, J. A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Annals of Statistics* **28**, 779–814.

Werren, S. (2011). *Pseudo-Likelihood Methods for the Analysis of Interval-Censored Data*. Ph.D. thesis, ETH, Zurich.

Yu, Q., Schick, A., Li, L., and Wong, G. Y. C. (1998). Asymptotic properties of the gmle in the case 1 interval-censorship model with discrete inspection times. *The Canadian Journal of Statistics* **26**, 619–627.

Zhang, Y. (20006). Nonparametric k-sample tests with panel count data. *Biometrika* **93**, 777–790.

Zhang, Y., Liu, W., and Zhan, Y. (2001). A nonparametric two-sample test of the failure function with interval censoring case 2. *Biometrika* **88**, 677–686.

Zhang, Z., Sun, L., Zhao, X., and Sun, J. (2005). Regression analysis of interval- censored failure time data with linear transformation models. *The Canadian Journal of Statistics* **33**, 61–70.