NONPARAMETRIC INFERENCE UNDER DEPENDENT TRUNCATION

Ming-Yen Cheng[1,2] Peter Hall[1] You-Jun Yang[2]

Dedicated to Sándor Csörgő on the occasion of his 60th Anniversary

ABSTRACT. Data truncation is a problem in scientific investigations. So far, statistical models and inferences are mostly based on the assumption that the survival and truncation times are independent, which can be unrealistic in applications. In a nonparametric setting, we discuss identifiability of the conditional and unconditional survival and hazard functions when the survival times are subject to dependent truncation, namely, the survival time is dependent on the truncation time. Nonparametric kernel estimators of these unknowns are proposed. Usefulness of the nonparametric estimators are demonstrated through their theoretical properties, an application and a simulation study.

KEYWORDS. Conditional distribution, dependent truncation, hazard rate, identifiability, kernel, nonparametric estimation, survival function, truncation.

AMS SUBJECT CLASSIFICATION: Primary 62N02, Secondary 62F12.

SHORT TITLE. Dependent Truncation.

[1] Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia

[2] Department of Mathematics, National Taiwan University, Taipei 106, Taiwan

1. INTRODUCTION

We consider nonparametric inference in truncated survival-data problems, where the time of onset of a disease, and the time of death, are observable if and only if the onset time falls to the left of a time-point t, and the time of death lies to the right of t. Models of this type address settings, commonly seen in prevalent cohort studies, where a population is surveyed at time t, and only individuals who have experienced the disease and still have the disease at that time are followed up. Under this setting, the survival time between the two events is left truncated by the truncation time from the initiating event to enrollment.

In this context there exists a literature on parametric approaches to inference, and also on nonparametric methods, in cases where the onset time and survival time are stochastically independent or quasi-independent. We shall discuss shortly this work. However, this quasi-independent, or quasi-stationary, assumption may be violated in practice. For example, advances in medical treatment and patient management may reduce the risk of death for patients with later onset times. To cope with such dependence, Wang et al. (1993) included a parametric component in their hazard regression models. Nevertheless, the more general nonparametric setting, where t is fixed and no assumptions are made about the relationship between onset time and survival time, is not well understood.

At first sight it might appear that there is little hope of making progress with such a completely general nonparametric approach, since the distributions of onset time and survival time are distorted so much by the operation of truncation at t. However, a closer inspection reveals that a surprising amount of detail can be recovered.

For example, it is possible to nonparametrically identify, up to a constant of proportionality and to the right of t, the survival-time density and survival-time distribution function. This may be achieved without making any structural assumptions, such as the independence hypothesis mentioned earlier. However, the constant of proportionality, and the density to the left of t, are not identifiable from a nonparametric viewpoint. Nevertheless, a "working value" of the constant of proportionality is readily obtained by either fitting a model, or proceeding nonparametrically under the assumption of independence. In this way, fully nonparametric methods can be used to provide a check on the validity of parametric or structural approaches. For example, a model, or the independence assumption, can be used to estimate both the constant of proportionality and the density and distribution to the right of t. Then, nonparametric estimators of these quantities can be compared with the "predictions" of the model.

Moreover, the conditional hazard rate for survival time s, given that the onset time lies to the right of $t - s$, is fully identifiable in nonparametric terms, as too is a version of the cumulative hazard rate. We shall suggest fully nonparametric estimators of these quantities and those discussed in the previous paragraph, and outline their properties.

A related setting is that where t is random. Here, significantly more information is available for inference than in the fixed-t case, and greater progress can be made under relatively weak assumptions. See, for example, the work of Wang *et al.* (1993), which addressed models for prevalent cohort data where patients are recruited at different, random time points. In this research the survival-time distribution was modelled under both independence and dependence assumptions on the initiating time.

In many contexts the term "survival time" should be replaced by "incubation time," or "induction time," or "lag time." In statistical and mathematical terms these quantities have the same role in the model as survival time. Thus, the description given by Lagakos *et al.* (1988) of nonparametric inference for induction time in the setting of HIV infection, involves an assumption that is functionally equivalent to the independence of onset time and survival time in a survival-data problem. Kalbfleisch and Lawless (1989) presented parametric and nonparametric methodology under the same independence condition. The nonparametric induction-time distribution estimators in both papers are in the spirit of methods for nonparametric maximum-likelihood estimation discussed by Woodroofe (1985); see also Li (1995).

If a time-reversal argument is used (see e.g. Lagakos *et al.*, 1988; Kalbfleisch and Lawless, 1989) then, after appropriate adjustments, our methods for the conditional survival and hazard functions can be applied to problems arising from retrospective studies, where only individuals who have experienced the second event before time t are observed. Another closely related situation is that of estimating age-specific incidence, where the time horizon is age instead of calendar time. Outside the realm of biomedical and epidemiological studies, truncated data frequently occur in actuarial, demographic and physical research. Therefore, our results about identifiability, and our nonparametric methods, have a wide range of applications.

Keiding (1991) considered estimating age-specific incidence of certain diseases under independent truncation. Kalbfleisch and Lawless (1991) treated regression of the survival distribution on other covariates. There further exist many related parametric, semiparametric and nonparametric approaches, based on the independence assumption, for various problems of inference about the survival distribution, cumulative hazard and hazard functions, etc. See, for example, Becker and Marschner (1993), Pagano *et al.* (1994), Gross and Lai (1996), Cui (1999), Grigoletto and Akritas (1999), Sun and Zhou (2001) and Li *et al.* (2002). Tests for the independence assumption have been proposed by, for example, Tsai (1990), Kalbfkeisch and Lawless (1991), Efron and Petrosian (1994) and Martin and Betensky (2005).

Since our estimator of the conditional survival distribution conveys the shape of the true curve correctly, it can be used to check the quasi-independence condition and to offer a guild for parametric modeling of the dependence. Further, the proposed methods may be incorporated in semiparametric hazard or survival regression when nonparametric modeling of the dependence is desired.

Section 2 discusses identifiability of the conditional and unconditional survival and hazard functions for both the truncated and original populations. Estimators of these quantities are proposed in section 3. In section 4, an application to a real-data set arising from a cancer study and a simulation study are presented. There, data arise from the right truncation model mentioned two paragraphs before. Section 5 gives theoretical properties of estimators of the conditional distribution, hazard and density functions in the truncated population. Appendix I contains some notation and Appendix II gives a proof of the main theorem.

2. IDENTIFIABILITY ISSUES

2.1. Model, main problems, and definitions. Assume the random vector (X^*, Y^*) is distributed in the triangular region $0 < x < y < \infty$. We observe (X^*, Y^*) only if $X^* \leq t < Y^*$, where $t > 0$ is fixed. Thus, we have no access to information about the distribution of (X^*, Y^*) conditional on $X^* > t$, and so we shall condition on $X^* \leq t$. That is, we shall assume that $X^* \leq t$. This mathematical model describes data resulting from a medical survey, at time t, of a population, where X^* and Y^* denote respectively the time of onset, and time of cessation, of a particular medical condition in a given individual. Each member of the population who is observed at time t to exhibit the condition, is monitored until the condition ceases to be present.

Let (X, Y) have the distribution of (X^*, Y^*) conditional on $Y^* > t$, and let

$(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent random two-vectors distributed as (X, Y). Using these data we wish to estimate the distribution of the time duration, $S^* = Y^* - X^*$, of the medical condition. Below, we shall refer to S^* as the survival time; if the medical condition is fatal, in which case Y^* denotes the time of death, then this terminology is standard. We also seek estimators of the hazard function and cumulative hazard function, each of them conditional on onset time.

Put $S = Y - X$, and let $f_{S^*X^*}(s, x)$ and $f_{S^*|X^*}(s \,|\, x)$ denote, respectively, the joint density of S^* and X^* and the conditional density of S^* given X^*. Define $f_{X^*}(x)$, $f_X(x)$, etc, analogously. We shall replace f by F when referring to the corresponding distribution function. In this notation, the survival time distribution is F_{S^*}, and the conditional hazard function and its cumulative are, respectively,

$$\lambda(s \,|\, x) = \frac{f_{S^*|X^*}(s \,|\, x)}{1 - F_{S^*|X^*}(s \,|\, x)}\,, \quad \Lambda(s \,|\, x) = \int_0^s \lambda(u \,|\, x)\, du\,. \tag{2.1}$$

2.2. Formulae for the probability of observing (X^*, Y^*). The definition of the distribution of X, and the fact that $f_{S^*X^*} = f_{X^*}\, f_{S^*|X^*}$, entail, for a constant $q \geq 1$,

$$\begin{aligned} f_X(x) &= q \int_{t-x}^{\infty} f_{S^*X^*}(s, x)\, ds = q\, f_{X^*}(x) \int_{t-x}^{\infty} f_{S^*|X^*}(s \,|\, x)\, ds \\ &= q\, f_{X^*}(x)\, \{1 - F_{S^*|X^*}(t - x \,|\, x)\}\,, \end{aligned}$$

for $0 < x \leq t$. Therefore, provided

$$P(S^* > t - x \,|\, X^* = x) > 0 \quad \text{for all} \quad 0 < x \leq t\,, \tag{2.2}$$

we have, for $0 < x \leq t$,

$$f_{X^*}(x) = p\, \frac{f_X(x)}{1 - F_{S^*|X^*}(t - x \,|\, x)}\,, \tag{2.3}$$

where $p = q^{-1}$ is determined by the fact that $\int_{0<x\leq t} f_{X^*}(x)\, dx = 1$.

Note that

$$p = \int_0^t \{1 - F_{S^*|X^*}(t - x \,|\, x)\}\, f_{X^*}(x)\, dx = P\{(X^*, Y^*) \text{ is observed}\}\,.$$

That is, p equals the proportion of data (X^*, Y^*) that are observed as pairs (X, Y). In general the value of p is not identifiable merely from the distribution of (X, Y). Indeed, we may make p as close to zero, or as close to 1, as we like, by altering the distribution of (X^*, Y^*) but without affecting the (X, Y) distribution. However, under more stringent conditions on the distribution of (X^*, Y^*), p can be identified.

In particular, if

$$S^* \text{ and } X^* \text{ are statistically independent} \tag{2.4}$$

then we may estimate F_{S^*} root-n consistently from n data on (X, Y) (see e.g. Woodroofe, 1985), and thence we may estimate

$$p = \left\{ \int_0^t \frac{f_X(x)}{1 - F_{S^*}(t - x)}\, dx \right\}^{-1} \tag{2.5}$$

at the same rate, from the same data. Result (2.5) follows from (2.3), provided (2.2) and (2.4) hold. In the context of independence, (2.2) is equivalent to $P(S^* > t) > 0$. He and Yang (1998) suggested an estimator of the truncation probability $1-p$ using a representation different from (2.5).

An alternative approach, more parametric in that it demands particular distributions for S^* and X^*, but less structured from the viewpoint that it does not assume independence, is to fit a model to (X^*, Y^*) and estimate model parameters from the data (X_i, Y_i). The value of p can then be computed for the model with parameters fitted.

In practice, estimating p under either (2.4) or a parametric model provides a "working guide" to choice of p, even if (2.4), or the model, may be false. A graph of nonparametric estimators of F_{S^*}, for a range of values of p centred around the working estimator of p, provides information where none might otherwise be available.

2.3. Identifying conditional hazard rate. It follows from the definition of (X, Y) that, provided that $s \geq t - x$,

$$P(S > s \mid X = x) = P(S^* > s \mid S^* > t - x \, , \; X^* = x) \, ,$$

or equivalently, that

$$\bar{F}_{S|X}(s \mid x) \equiv 1 - F_{S|X}(s \mid x) = \frac{1 - F_{S^*|X^*}(s \mid x)}{1 - F_{S^*|X^*}(t - x \mid x)} \, . \tag{2.6}$$

Noting the definitions of $\lambda(s \mid x)$ and $\Lambda(s \mid x)$ at (2.1), and that $\bar{F}_{S|X}(t - x \mid x) = 1$, we may deduce from (2.6) that for $s \geq t - x$,

$$\lambda(s \mid x) = -\frac{\partial \bar{F}_{S|X}(s \mid x)/\partial s}{\bar{F}_{S|X}(s \mid x)} \, , \tag{2.7}$$

$$L(s, x) \equiv \Lambda(s \mid x) - \Lambda(t - x \mid x) = -\log\left\{\bar{F}_{S|X}(s \mid x)\right\} . \tag{2.8}$$

Since, for this range of values of s, $\bar{F}_{S|X}(s \mid x)$ is identifiable from data on (X, Y), then it follows from (2.7) and (2.8) that within the same range, $\lambda(s \mid x)$ and $\Lambda(s \mid x) - \Lambda(t - x \mid x)$ are also identifiable from data on (X, Y). However, since we can alter the distribution of (X^*, S^*) in the range $s < t - x$ without affecting the distribution of (X, Y), then neither $\lambda(s \mid x)$ nor $\Lambda(s \mid x) - \Lambda(t - x \mid x)$ is identifiable for $s < t - x$, and $\Lambda(s \mid x)$ is not identifiable for any $s > 0$.

Of course, $\lambda(s \mid x)$ and $\Lambda(s \mid x)$ are both identifiable, for $s > 0$ and $0 < x \leq t$, under the independence assumption, (2.4). In this context the conditioning on x is irrelevant, and λ and Λ are functionals of F_{S^*}.

2.4. Identifying overall hazard and distribution functions. Assuming (2.2) holds, we may use (2.3) and (2.6) to obtain the formulae,

$$F_{S^*}(s) = p \int_0^t \frac{f_X(x) \, F_{S^*|X^*}(s \mid x)}{1 - F_{S^*|X^*}(t - x \mid x)} \, dx \quad \text{for} \quad s > 0 \tag{2.9}$$

$$= 1 - p \int_0^t f_X(x)\, \bar{F}_{S|X}(s \,|\, x)\, dx \quad \text{for} \quad s \ge t\,, \tag{2.10}$$

$$f_{S^*}(s) = p \int_0^t f_X(x)\, f_{S|X}(s \,|\, x)\, dx \qquad \text{for} \quad s \ge t\,. \tag{2.11}$$

Let us assume, for the time being, that p is known. Then, in view of (2.9), F_{S^*} is identifiable from data on (X, Y), if and only if the integral on the right-hand side of (2.9) is identifiable. Since knowing (X, Y) is equivalent to knowing (S, X), which is constrained by $S \ge t - X$, then we cannot, in general, identify $F_{S^*|X^*}(s \,|\, x)$ outside the range $s \ge t - x$. However, if $s < t$ then the integral at (2.9) requires knowledge of $F_{S^*|X^*}(s \,|\, x)$ for s to the left of $t - x$. Therefore, if the only information we have is in terms of the distribution of (X, Y), we cannot identify $F_{S^*}(s)$ outside the range $s \ge t$. Indeed, for the reasons just given it is possible to construct two examples of smooth, non-pathological distributions of (X^*, Y^*) for which p, and the distribution of (X, Y), are identical, but the values $F_{S^*}(s)$ differ for $s < t$. Within the identifiable range $s \ge t$, (2.10) provides an explicit formula for F_{S^*} in terms of p and the identifiable functions f_X and $F_{S|X}$.

Formula (2.10) may equivalently be written as

$$F_{S^*}(s) = 1 - \pi\, P(S > s \,|\, S > t) \quad \text{for} \quad s \ge t\,, \tag{2.12}$$

where the function $P(S > s \,|\, S > t)$ is of course identifiable for $s \ge t$, and the constant $\pi \equiv P(S^* > t)$ is not identifiable from data on (X, Y) alone. We argued in section 2.2 that a "working approximation" to p might be constructed by estimating p under the assumption that S^* and X^* are independent, or via a parametric model. The same approach can be taken for π, and the function $P(S > s \,|\, S > t)$ estimated root-n consistently from data on (X, Y).

Note particularly that (2.11), and also (2.12), imply that we can identify, up to a constant of proportionality and to the right of t, the survival-time density f_{S^*}.

In addition, from (2.10) and (2.11), the overall hazard rate

$$\lambda_{S^*}(s) = \frac{f_{S^*}(s)}{1 - F_{S^*}(s)}$$

is identifiable from data on (X, Y) for all $s \geq t$.

3. METHODOLOGY

3.1. Outline of methodology. We shall suggest kernel-based methods for estimating $F_{S|X}(s \,|\, x)$, using the data (X_i, Y_i). The estimators are of Nelson–Aalen or product-limit type. They can be applied directly to estimate $L(s, x)$, given at (2.8), and indirectly to estimate $\lambda(s \,|\, x)$ and $F_{S^*}(s)$. Our methods allow these quantities to be estimated throughout the range where they are identifiable, which for $L(s, x)$ and $\lambda(s \,|\, x)$ is $s \geq t - x$ and, for $F_{S^*}(s)$, is $s \geq t$.

The estimator of $F_{S^*}(s)$ may be smoothed, and then differentiated, to give an estimator of $f_{S^*}(s)$ for $s \geq t$. An alternative approach to estimating F_{S^*} and f_{S^*} is to proceed via the representation (2.12). That method too will be discussed. Our estimators enjoy optimal nonparametric convergence rates.

3.2. Estimators of $F_{S|X}$ and L. We shall treat only second-order methods, designed for estimating conditional distributions or densities with two bounded derivatives. Higher-order estimators may be treated similarly.

Let K, a kernel function, be a bounded, compactly supported, symmetric probability density, and let h denote a bandwidth. Put $K_h(u) = h^{-1} K(u/h)$. Two estimators of $F_{S|X}$, consistent in the range $s \geq t - x$ and $0 < x \leq t$, are:

$$\widehat{F}_{S|X}(s \,|\, x) = 1 - \exp\left\{ - \sum_{i=1}^{n} \frac{I(S_i \leq s)\, K_h(x - X_i)}{\sum_{j \leq n} I(t - X_j \leq S_i \leq S_j)\, K_h(x - X_j)} \right\}, \tag{3.1}$$

$$\widetilde{F}_{S|X}(s \,|\, x) = 1 - \prod_{i=1}^{n} \left\{ 1 - \frac{K_h(x - X_i)}{\sum_{j \leq n} I(t - X_j \leq S_i \leq S_j)\, K_h(x - X_j)} \right\}^{I(S_i \leq s)}.$$

The first estimator, based on (2.8), derives from a Nelson-Aalen estimator of $L(s,x)$:

$$\widehat{L}(s,x) = \sum_{i=1}^{n} \frac{I(S_i \le s)\, K_h(x - X_i)}{\sum_{j \le n} I(t - X_j \le S_i \le S_j)\, K_h(x - X_j)}\,.$$

The second estimator, $\widetilde{F}_{S|X}(s\,|\,x)$, is of product-limit type, and may be viewed as an approximation by Taylor expansion to $\widehat{L}(s,x)$. It is analogous to another estimator of $L(s,x)$, $\widetilde{L}(s,x)$ say.

Result (5.5), below, will show that $\widehat{L}$ converges to L at rate $(nh)^{-1/2} + h^2$, where the first term derives from error-about-the-mean and the second from bias. A central limit theorem for $\widehat{L}$ is obtainable from a bivariate form of (5.7).

3.3. Smoothing $\widehat{F}_{S|X}$ and $\widetilde{F}_{S|X}$, and estimating λ. We cannot immediately differentiate either $\widehat{F}_{S|X}(s\,|\,x)$ or $\widetilde{F}_{S|X}(s\,|\,x)$ in s, since they are not smooth functions of that variable. However, this difficulty can be eliminated by replacing the indicator function $I(S_i \le s)$, in the definitions of both estimators, by $M\{(s - S_i)/H\}$, where M denotes the distribution function corresponding to the density K, and H is another bandwidth. Provided K is continuous, the functions $\widehat{F}_{S|X}(s\,|\,x)$ and $\widetilde{F}_{S|X}(s\,|\,x)$ now have continuous derivatives with respect to s, and these derivatives can be viewed as estimators of $f_{S|X}(s\,|\,x)$.

If we were interested only in smoothing $\widehat{F}_{S|X}$ and $\widetilde{F}_{S|X}$ to remove visually distracting jumps in graphs of function estimates, such as $\widehat{L}(s,x)$, we would choose H by eye. If we wished to estimate $f_{S|X}(s\,|\,x)$ or $\lambda_{S|X}(s\,|\,x)$, we would generally take $H = h$, so that only one bandwidth had to be chosen. This approach gives estimators, $\hat{f}_{S|X}$ and $\hat{\lambda}$, say, of $f_{S|X}$ and λ, respectively, with optimal convergence rates. Indeed, under the assumption that f_{SX} has two bounded derivatives, using the above smoothing technique, and choosing $H = h = h(n) \to 0$ and $nh^2 \to \infty$ as $n \to \infty$, the pointwise convergence rates of $\hat{f}_{S|X} \equiv \partial\widehat{F}_{S|X}/\partial s$ and $\partial\widetilde{F}_{S|X}/\partial s$ to $f_{S|X}(s\,|\,x)$ equal $O_p\{(nh^2)^{-1/2} + h^2\}$. Taking h to be of size $n^{-1/6}$, this gives the optimal mean-square convergence rate of $n^{-2/3}$, for functions with two bounded derivatives. In the case of $\hat{f}_{S|X}$ these results follow from (5.6) below, which also implies a central limit theorem for the estimator.

In more detail, substitution into (2.7) gives

$$\hat{\lambda}(s \,|\, x) = \frac{\partial \widehat{F}_{S|X}(s \,|\, x)/\partial s}{1 - \widehat{F}_{S|X}(s \,|\, x)} = \sum_{i=1}^{n} \frac{K_h(s - S_i)\, K_h(x - X_i)}{\sum_{j \le n} I(t - X_j \le S_i \le S_j)\, K_h(x - X_j)}\,, \tag{3.2}$$
$$\hat{f}_{S|X}(s \,|\, x) = \big\{1 - \widehat{F}_{S|X}(s \,|\, x)\big\}\, \hat{\lambda}(s \,|\, x)\,.$$

Here, different values of the smoothing parameter h could be used to construct $\hat{\lambda}(s \,|\, x)$ and $\widehat{F}_{S|X}(s \,|\, x)$; the bandwidths would be of sizes $n^{-1/6}$ and $n^{-1/5}$, respectively, reflecting the fact that the curve estimation problems are bivariate for the former and univariate for the latter.

3.4. Alternative estimator of $F_{S|X}$, and edge effects. Another approach to estimating $F_{S|X}$ is to proceed via a kernel estimator, $\check{f}_{SX}$ say, of f_{SX}:

$$\check{f}_{SX}(s, x) = \frac{1}{nh^2} \sum_{i=1}^{n} K\Big(\frac{s - S_i}{h}\Big)\, K\Big(\frac{x - X_i}{h}\Big)\,.$$

Since, for $s \ge t - x$,

$$F_{S|X}(s \,|\, x) = \frac{\int_{t-x<u<s} f_{SX}(u, x)\, du}{\int_{u>t-x} f_{SX}(u, x)\, du}\,,$$

then, for appropriate choice of the bandwidth h,

$$\check{F}_{S|X}(s \,|\, x) = \frac{\int_{t-x<u<s} \check{f}_{SX}(u, x)\, du}{\int_{u>t-x} \check{f}_{SX}(u, x)\, du}$$

is a consistent estimator of $F_{S|X}(s \,|\, x)$.

A disadvantage of this approach, however, is that it is relatively sensitive to the impact of edge effects, for example those that occur near the diagonal boundary defined by $s + x = t$. These effects can have substantial influence on the bias of $\check{F}_{S|X}$, with the result that $\check{F}_{S|X}(s\,|\,x)$ is, in general, not consistent at either the diagonal boundary or the vertical boundaries defined by $x=0$ and $x=t$. Corrections for edge effects need to be incorporated, in particular if $\check{F}_{S|X}$ is used as the basis for an estimator of F_{S^*}. Those adjustments can be tedious and cumbersome.

The estimators $\widehat{F}_{S|X}$ and $\widetilde{F}_{S|X}$ are substantially more robust than $\check{F}_{S|X}$ against edge effects due to bias. See, for example, the first part of the theorem in section 5, where the order of bias, $O(h^2)$, in (5.5) is valid along the diagonal boundary and within h of the vertical boundary. Along the vertical boundaries the order of bias is $O(h)$, rather than $O(1)$ as in the case of $\check{F}_{S|X}$, and so the estimators $\widehat{F}_{S|X}$ and $\widetilde{F}_{S|X}$ are consistent along the vertical boundaries.

However, $\widehat{F}_{S|X}$ and $\widetilde{F}_{S|X}$ can suffer problems in the upper tail of the distribution, in particular if it happens that for some (s,x),

$$F_{S|X}(s\,|\,x)=1\,. \tag{3.3}$$

Condition (5.1), in our discussion below of theoretical properties of $\widehat{F}_{S|X}(s\,|\,x)$, reflects this point. However, provided the distribution of S given X is unbounded to the right, the identity (3.3) will never arise, and the upper limit of values of s for which methods, based on $\widehat{F}_{S|X}$ or $\widetilde{F}_{S|X}$, can estimate $F_{S|X}$ effectively, will steadily increase with sample size.

3.5. Estimators of F_{S^*}, f_{S^*} and λ_{S^*}. One approach to estimating F_{S^*} is to construct an estimator of $F_{S|X}$, as suggested in section 3.2, then compute an estimator of $\hat{f}_X$ of f_X using standard kernel methods, and combine them into an estimator of F_{S^*}, for a particular value of p, by substituting into (2.10):

$$\widehat{F}_{S^*}(s)=1-p\int_0^t \hat{f}_X(x)\,\{1-\widehat{F}_{S|X}(s\,|\,x)\}\,dx$$

for $s\geq t$. Provided we appropriately undersmooth when constructing $\hat{f}_X$ and $\widehat{F}_{S|X}$, which means choosing a bandwidth h satisfying $n^{-1+\delta} \leq h = O(n^{-1/4})$ for some $\frac{1}{2} < \delta \leq \frac{3}{4}$, the estimator $\widehat{F}_{S^*}$ is root-n consistent for F_{S^*}.

It is simpler, however, to work from (2.12), and take

$$\widehat{F}_{S^*}(s) = 1 - \pi\, \widehat{G}(s,t) \tag{3.4}$$

for $s \geq t$, where

$$\widehat{G}(s,t) = \frac{\sum_i I(S_i > s)}{\sum_i I(S_i > t)} \tag{3.5}$$

estimates $G(s,t) \equiv P(S > s \,|\, S > t)$. Provided $P(S > s) > 0$, the estimator $\widehat{G}$ is root-n consistent for G.

Discontinuities in s can be removed using the method suggested in section 3.3. Here this amounts to replacing the indicator function $I(S_i > s)$ in (3.5) by $1 - M\{(s - S_i)/H\}$, where M is a distribution function and H denotes a bandwidth. As in section 3.3, if our reason for smoothing was to remove jumps in $\widehat{G}(\,\cdot\,, t)$ then H would be small, of order n^{-1}, and would be chosen by eye. On the other hand, if our aim was to estimate f_{S^*} then we would use a larger bandwidth. Taking $H = h$ in this case, we differentiate (3.4) with respect to s, obtaining a kernel estimator of $f_{S^*}(s)$:

$$\hat{f}_{S^*}(s) = \pi\, \frac{\sum_i K_h(s - S_i)}{\sum_i I(S_i > t)} \tag{3.6}$$

for $s \geq t$, where $K = M'$. Choosing h to be of size $n^{-1/5}$ gives an estimator $\hat{f}_{S^*}$ with optimal mean-square convergence rate $n^{-4/5}$, provided π assumes its true value. At the very least, without any attempt to evaluate π, (3.6) gives, for $s > t$, a completely nonparametric estimator of a function, $f_{S^*}(s)/\pi$, that is proportional to the true survival density, $f_{S^*}(s)$.

As suggested in section 2.4, one approach to choosing π would be to use either a parametric model, or the structural, but nonparametric, assumption that S^* and

X^* are independent, to produce a working value of the non-identifiable quantity $\pi = P(S^* > t)$; and to plot the estimates $\hat{f}_{S^*}(s)$, $s \geq t$, for values of π on either side of the working value.

Forming ratio of $\hat{f}_{S^*}(s)$ and $1 - \widehat{F}_{S^*}(s)$ we have an estimator

$$\hat{\lambda}_{S^*}(s) = \frac{\sum_i K_h(s - S_i)}{\sum_i I(S_i > s)} \tag{3.7}$$

of the overall hazard rate $\lambda_{S^*}(s)$ for all s to the right of t.

The estimators at (3.6) and (3.7) can suffer edge effects at the left- hand boundary, i.e. near $s = t$. These can be removed by conventional means, for example through using a boundary kernel there or by binning the data and fitting a local linear smoother.

4. NUMERICAL STUDY

4.1. An application. We present here an application to a dataset from a breast cancer study conducted by the National Cancer Institute of Canada Clinical Trials Group. Between April 1998 and July 1999, 305 eligible patients, who had inoperable metastatic or recurrent breast cancer with an Eastern Cooperative Oncology Group score of 0 to 2, were entered onto the study from 38 centres in Eastern Europe ($n = 192$), Canada ($n = 46$), South Africa ($n = 34$), Western Europe ($n = 24$), and Australia ($n = 9$).

Let $\widetilde{X}^*$, $\widetilde{Y}^*$ and t ($= 355$), all expressed in months and calculated from 1st January 1970, respectively denote the time a patient was first diagnosed with breast cancer, the time the patient had the first recurrence of breast cancer, and the time to the end of the study. Of major interest is the survival time, or progression free time, $\widetilde{S}^* = \widetilde{Y}^* - \widetilde{X}^*$ from the first incidence of breast cancer to the first recurrence.

In this study, only individuals with $\widetilde{Y}^* \leq t$ can be observed. Equivalently, the survival time $\widetilde{S}^*$ is subject to right truncation by $t - \widetilde{X}^*$. Let $(\widetilde{X}, \widetilde{S})$ denote the truncated version of $(\widetilde{X}^*, \widetilde{S}^*)$, i.e. $(\widetilde{X}, \widetilde{S})$ equals $(\widetilde{X}^*, \widetilde{S}^*)$ conditional on $\widetilde{S}^* \le t - \widetilde{X}^*$. The left panel of Figure 1 plots $\widetilde{X} + \widetilde{S}$ against $\widetilde{X}$ for the 287 recurrent breast cancer patients, among the total of 305 metastatic or recurrent breast cancer cases.

If the time-reversal transformation

$$X^* = t - \widetilde{X}^*, \quad S^* = t - \widetilde{S}^*, \quad X = t - \widetilde{X}\,, \quad S = t - \widetilde{S}\,, \tag{4.1}$$

is employed, then S^* is subject to left truncation and (X, S) has the distribution of (X^*, S^*) conditional on $S^* \ge t - X^*$. Hence, the methods proposed in section 3 can be applied to analyse this data set. The right panel of Figure 1 shows the transformed data on the truncated population of (X, S). Note that one isolated point, corresponding to a much earlier initial time of original breast cancer than the others, has been deleted.

In section 4.1.1, the conditional survival and hazard functions of $\widetilde{S}$ given $\widetilde{X}$ are estimated using (4.1) and estimators of the conditional survival and hazard functions of S given X given in section 3.2. Section 4.1.2 studies the original population $(\widetilde{X}^*, \widetilde{S}^*)$. Under the model in section 2, $F_{S^*}(s)$ is identifiable up to a constant of proportionality and $\lambda_{S^*}(s)$ is identifiable for $s \ge t$; $F_{S^*|X^*}(s|x)$ is identifiable up to a constant of proportionality and $\lambda_{S^*|X^*}(s|x)$ is identifiable for $s > t - x$. Under the model here, neither $F_{\widetilde{S}^*}(\widetilde{s})$ nor $\lambda_{\widetilde{S}^*}(\widetilde{s})$ is identifiable for any $0 < \widetilde{s} < t$; $F_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s}|\widetilde{x})$ is identifiable up to a constant of proportionality and $\lambda_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s}|\widetilde{x})$ is identifiable up to a constant for all $0 < \widetilde{s} < t - \widetilde{x}$.

4.1.1. Truncated distribution. Following from the time-reversal transformation (4.1), for all $0 \le \widetilde{s} \le t - \widetilde{x}$,

$$F_{\widetilde{S}|\widetilde{X}}(\widetilde{s}\,|\,\widetilde{x}) = 1 - F_{S\,|\,X}(s^-\,|\,x)\,, \quad f_{\widetilde{S}|\widetilde{X}}(\widetilde{s}\,|\,\widetilde{x}) = f_{S|X}(s\,|\,x)\,,$$

where $s = t - \widetilde{s}$ and $x = t - \widetilde{x}$. Substituting into this formula our estimators of $F_{S|X}(s\,|\,x)$ and $f_{S|X}(s\,|\,x)$, given in (3.1), we obtain estimators of $F_{\widetilde{S}|\widetilde{X}}(\widetilde{s}\,|\,\widetilde{x})$ and $f_{\widetilde{S}|\widetilde{X}}(\widetilde{s}\,|\,\widetilde{x})$.

Observe from Figure 1 that truncation is heavy for large $\widetilde{x}$ values, and that there are very few points with small $\widetilde{x}$ values. Also, there are edge effects. Therefore, in the analysis of conditional survival distributions, $\widetilde{x}$ was fixed at ten middle values 220, 230,..., 310. The kernel function was $K(u) = \frac{15}{16}(1-u^2)^2\, I(|u| \leq 1)$. The bandwidth h was taken as 15 for the first five $\widetilde{x}$ values, and 7.5 for the others.

Figure 2 depicts the estimates of the truncated conditional distribution function, $F_{\widetilde{S}|\widetilde{X}}(\widetilde{s}\,|\,\widetilde{x})$, over the range $0 \leq \widetilde{s} \leq t-\widetilde{x}$ where it is identifiable. The truncated conditional distributions are noticeably different. Therefore, analysis based on the overall distribution of the true survival time $\widetilde{S}^*$, instead of on its conditional distribution on $\widetilde{X}^*$, could be quite misleading.

Note from (4.1) that the conditional hazard rate $\lambda_{\widetilde{S}|\widetilde{X}}(\widetilde{s}\,|\,\widetilde{x})$ is

$$\lambda_{\widetilde{S}|\widetilde{X}}(\widetilde{s}\,|\,\widetilde{x}) = \frac{f_{S|X}(s\,|\,x)}{F_{S|X}(s^-\,|\,x)} = \lambda_{S\,|\,X}(s\,|\,x)\;\frac{1-F_{S|X}(s\,|\,x)}{F_{S|X}(s^-\,|\,x)}\;,$$

for all $0 \leq \widetilde{s} \leq t-\widetilde{x}$. Hence, to estimate $\lambda_{\widetilde{S}|\widetilde{X}}(\widetilde{s}\,|\,\widetilde{x})$, one could replace the unknowns $F_{S|X}(s\,|\,x)$ and $\lambda_{S|X}(s\,|\,x)$ on the right-hand side by their estimators provided above and in (3.2).

4.1.2. Original distribution. The survival distribution in the original population $(\widetilde{X}^*, \widetilde{S}^*)$ is of significant interest. Lagakos et al. (1988) derived the product-limit estimator of $F_{\widetilde{S}^*}(\widetilde{s})$ using (4.1) and assuming independence between $\widetilde{S}^*$ and $\widetilde{X}^*$. Then one can obtain the corresponding Nelson-Aalen estimator of the cumulative hazard. The product-limit estimate of $F_{\widetilde{S}^*}(\widetilde{s})$ is shown in Figure 3. On the other hand, for all $0 < \widetilde{s} < t$,

$$F_{\widetilde{S}^*}(\widetilde{s}) = P(S^* \geq s) = p\int_0^t \frac{f_X(x)\bar{F}_{S^*|X^*}(s^-\,|\,x)}{1-F_{S^*|X^*}(t-x\,|\,x)}\,dx\,. \tag{4.2}$$

Hence, without the independence condition, neither $F_{\widetilde{S}^*}(\widetilde{s})$ nor $\lambda_{\widetilde{S}^*}(\widetilde{s})$ is identifiable using data on $(\widetilde{S}, \widetilde{X})$ for any $0 < \widetilde{s} < t$. Even with information on p, this difficulty remains as the integral in (4.2) cannot be changed to the form in (2.10) for any $0 < \widetilde{s} < t$. This in can be seen from the data structure $\widetilde{X}^* < \widetilde{X}^* + \widetilde{S}^* < t$ where t is a constant.

From (2.6) and (4.1), for $0 \le \widetilde{s} \le t - \widetilde{x}$,

$$F_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s} \,|\, \widetilde{x}) = \bar{F}_{S^*|X^*}(s^- \,|\, x) = \bar{F}_{S|X}(s^- \,|\, x)\, \bar{F}_{S^*|X^*}(t - x \,|\, x)\,. \tag{4.3}$$

The conditional truncation probability $F_{S^*|X^*}(t - x \,|\, x)$ is not identifiable from data on the truncated version $(\widetilde{S}, \widetilde{X})$. Under the independence assumption, $\bar{F}_{S^*|X^*}(t - x \,|\, x)$ equals $\bar{F}_{S^*}(t - x)$, or equivalently $F_{\widetilde{S}^*}(t - \widetilde{x})$, which can be estimated by the product-limit method described before. Combining this working value of $\bar{F}_{S^*|X^*}(t - x \,|\, x)$ and our estimator of $\bar{F}_{S|X}(s^- \,|\, x)$ we obtain an estimator of $F_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s} \,|\, \widetilde{x})$ over the range $0 \le \widetilde{s} \le t - \widetilde{x}$. The results are also depicted in Figure 3. It shows that the conditional distribution of the survival time $\widetilde{S}^*$ given the initial time $\widetilde{X}^*$ varies over the initial time. And the product-limit estimate of the unconditional distribution function differs from the estimates of the conditional distribution functions in shape.

Often, the hazard function gives better insight into the survival distribution than the survival function does. From (4.3), for $0 \le \widetilde{s} \le t - \widetilde{x}$,

$$\begin{aligned} \lambda_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s} \,|\, \widetilde{x}) &= \lambda_{S^*|X^*}(s \,|\, x)\, \frac{\bar{F}_{S^*|X^*}(s \,|\, x)}{F_{S^*|X^*}(s^- \,|\, x)} \\ &= \lambda_{S^*|X^*}(s \,|\, x)\, \frac{\bar{F}_{S|X}(s \,|\, x)\, \bar{F}_{S^*|X^*}(t - x \,|\, x)}{1 - \bar{F}_{S|X}(s^- \,|\, x)\, \bar{F}_{S^*|X^*}(t - x \,|\, x)}\,. \end{aligned}$$

Therefore, while $\lambda_{S^*|X^*}(s \,|\, x)$ is identifiable for all $s \ge t - x$, $\lambda_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s} \,|\, \widetilde{x})$ is not identifiable in the region $0 \le \widetilde{s} \le t - \widetilde{x}$. An estimator is available using our estimators of $F_{S|X}(s \,|\, x)$ and $\lambda_{S^*|X^*}(s \,|\, x)$ $(= \lambda_{S|X}(s \,|\, x)$ for $s \ge t - x)$ given in (3.1) and (3.2), and a working value of $F_{S^*|X^*}(t - x \,|\, x)$, e.g. the product-limit estimator of $F_{S^*}(t-x)$, by assuming independence. The results are depicted in Figure 4. When $\widetilde{x}=260$ or 270, the first peak in the conditional hazard estimates occurs at around $\widetilde{s}=40$ or 20, much earlier than the first peak (around $\widetilde{s}=65$) in the other curves. Alternatively, under independence of $\widetilde{S}^*$ and $\widetilde{X}^*$, kernel smoothing the Nelson-Aalen estimator yields an estimator of the overall hazard function $\lambda_{\widetilde{S}^*}(\widetilde{s})$. This overall hazard rate estimate, shown in Figure 4, does not resemble any of the conditional hazard rate estimates.

The observations made from Figures 3 and 4 suggest that, for this data set, dependence between the survival time $\widetilde{S}^*$ and the truncation time $t-\widetilde{X}^*$ exists. In that case, traditional approaches that rely on the independence assumption, such as the product-limit and Nelson-Aalen estimators, may not be useful. By contrast, our methods provide a better description of the dependence by illustrating how the conditional survival and hazard change over the truncation time.

4.2. Simulation study. Here we use a simulation study to demonstrate that, in the case of dependent truncation, our methods indeed recover information about the original distribution from the truncated data more accurately than methods based on the independence assumption. We consider a setting similar to the breast-cancer data: $n=300$ observations were collected on $(\widetilde{X},\widetilde{Y})$, which equals $(\widetilde{X}^*,\widetilde{Y}^*)$ conditional on $\widetilde{X}^*\le\widetilde{Y}^*\le t$. Let the truncation time t be 350. Take $\widetilde{X}^*=350-Z$, where Z follows an Exponential distribution with rate 0.03 and is constrained by $Z\le t\,(=350)$. Suppose the conditional distribution of the survival time, $\widetilde{S}^*=\widetilde{Y}^*-\widetilde{X}^*$, given $\widetilde{X}^*=\widetilde{x}$ is a log-logistic distribution with density function

$$f_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s}\,|\,\widetilde{x})=\frac{\alpha(\widetilde{x})^{-1}\beta(\widetilde{x})\big\{\widetilde{s}/\alpha(\widetilde{x})\big\}^{\beta(\widetilde{x})-1}}{\Big[1+\big\{\widetilde{s}/\alpha(\widetilde{x})\big\}^{\beta(\widetilde{x})}\Big]^2}\,,\quad \widetilde{s}>0\,.$$

The survival and hazard functions are respectively

$$\bar{F}_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s}\,|\,\widetilde{x}) = \left[1+\{\widetilde{s}/\alpha(\widetilde{x})\}^{\beta(\widetilde{x})}\right]^{-1}, \ \lambda_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s}\,|\,\widetilde{x}) = \frac{\alpha(\widetilde{x})^{-1}\beta(\widetilde{x})\{\widetilde{s}/\alpha(\widetilde{x})\}^{\beta(\widetilde{x})-1}}{\left[1+\{\widetilde{s}/\alpha(\widetilde{x})\}^{\beta(\widetilde{x})}\right]}.$$

We set $\alpha(\widetilde{x}) = \exp(\widetilde{x}/70)$ and $\beta(\widetilde{x}) = (\widetilde{x}+100)/100$. The methods described in section 4.1 were applied to 1000 random samples generate from this model.

Panel (a) of Figure 5 plots the overall hazard function $\lambda_{\widetilde{S}^*}(\widetilde{s})$ and the conditional hazard function $\lambda_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s}\,|\,\widetilde{x})$ for $\widetilde{x} = 210, 230, 250$ and 270. Panel (b) of Figure 5 gives the pointwise 10th, 50th and 90th percentiles of the product-limit estimator (by independence assumption) of $\lambda_{\widetilde{S}^*}(\widetilde{s})$ and the true curve. As expected, this estimator fails here since independence does not hold and the overall hazard function does not resemble any of the conditional hazard functions. Given in panel (c) of Figure 5 are five typical realizations of the product-limit estimator that correspond to the 10th, 30th, 50th, 70th and 90th percentiles of the 1000 integrated squared errors, calculated from zero to the 95th quantile of the true distribution of $\widetilde{S}^*$. All of them depart from the true curve dramatically.

Figure 6 plots the pointwise 10th, 50th and 90th percentiles of our estimator of $\lambda_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s}\,|\,\widetilde{x})$, described in section 4.1.2, for $\widetilde{x} = 210, 230, 250$ and 270. Although the product-limit estimator, used as a working value of $F_{S^*|X^*}(t-x\,|\,x)$ in our method, is biased without the independence assumption, our estimator performs quite well. Comparing Figures 5 and 6, it is clear that our observation in section 4.1.2 are not just artifacts of random fluctuation. Instead, our estimators indeed recover shapes of the underlying curves accurately.

We also experimented with replacing the working value of $F_{S^*|X^*}(t-x\,|\,x)$ by the true value in the estimation of $\lambda_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s}\,|\,\widetilde{x})$. The outcomes are not given here, but this approach always does better. However, only for large values of $\widetilde{s}$ or $\widetilde{x}$ is it significantly better than our estimator.

As stated in section 4.1.1, the conditional hazard function $\lambda_{\widetilde{S}|\widetilde{X}}(\widetilde{s}\,|\,\widetilde{x})$ is identifiable; we gave an estimator of it there. Figure 7, presenting performance of our estimator of $\lambda_{\widetilde{S}|\widetilde{X}}(\widetilde{s}\,|\,\widetilde{x})$, is an analog of Figure 6. It validates our estimator for a relatively small sample size, $n = 300$.

5. THEORETICAL PROPERTIES

First we discuss the estimator $\widehat{F}_{S|X}$, introduced in section 3.2. Results similar to those below may be derived in the case of $\widetilde{F}_{S|X}$. Our main theoretical formula, from which our other results will follow, will describe uniform convergence properties of $\widehat{F}_{S|X}(s\,|\,x)$ for values of (s, x) satisfying $0 < t - x < s < s_0$, where s_0 has the property:

$$\sup_{0<x<t} F(s_0\,|\,x) < 1\,. \tag{5.1}$$

Since our estimators are based on second-order kernel methods then their biases are of order h^2. This is true for both density and distribution estimators. Concise expressions for the order h^2 bias terms are especially complex, however, and so we shall not give them here; they will be represented simply as $O(h^2)$ in asymptotic approximations. On the other hand, we shall be relatively concise about error-about-the mean terms.

Next we introduce notation. Let c_j and β_j denote functions, not depending on n or h, which are introduced in Appendix I. Put

$$a_1(s, x\,; s', x'\,|\,h) = K_h(x - x')\, c_1(s, x\,; s', x')\,,$$

$$a_2(s, x\,; s', x'\,|\,h) = K_h(x - x')\, K_h(s - s')\, c_2(s, x\,; s', x')$$

and $b_j(s, x\,; s', x'\,|\,h) = a_j(s, x\,; s', x'\,|\,h) - E\{a_j(s, x\,; S, X\,|\,h)\}$. We shall make the following assumptions:

the distribution of (S, X) is continuous, with density f_{SX}, which has two continuous derivatives on its support; the support of f_X equals the interval $[0, t]$, on which f_X is bounded away from zero; (5.2)

K is a bounded, symmetric, Hölder-continuous probability density, supported on $[-1, 1]$; (5.3)

$$\text{for some } \delta \in (0,1),\ h = h(n) \text{ satisfies } n^{\delta} h \to 0 \text{ and } n^{1-\delta} h \to \infty. \tag{5.4}$$

Theorem 1. *Assume* (5.2)–(5.4), *and that* s_0 *satisfies* (5.1). *Define* $\widehat{F}_{S|X}$ *by* (3.1). *Then, uniformly in* (s, x) *in the range that* $0 < t - x < s < s_0$ *with* $h \leq x \leq t - h$ *or* $s \leq \min(t, t - x + h)$

$$\frac{\widehat{F}_{S|X}(s \,|\, x) - F_{S|X}(s \,|\, x)}{\bar{F}_{S|X}(s \,|\, x)} = h^2 \, \beta_1(s, x) + \frac{1}{n} \sum_{i=1}^{n} b_1(s, x \,; S_i, X_i \,|\, h) + o_p\big\{(nh)^{-1/2} + h^2\big\} \,. \tag{5.5}$$

Theorem 2. *Assume the conditions in Theorem 1. If, in the definition at* (3.1), *we replace* $I(S_i \leq s)$ *in the numerator on the right-hand side by* $M\{(s - S_i)/h\}$, *where* M *denotes the distribution function corresponding to density* K*; if we then define* $\hat{f}_{S|X}(s \,|\, x)$ *as the derivative of* $\widehat{F}_{S|X}(s \,|\, x)$ *with respect to* s *and* $\hat{\lambda}(s \,|\, x)$ *as in* (3.2)*; and if we replace the assumption "*$n^{1-\delta} h \to \infty$*" in* (5.4) *by "*$n^{1-\delta} h^2 \to \infty$*"; then, for* (s, x) *satisfying* $0 < t - x < s < s_0$ *with* $h \leq x \leq t - h$ *and* $s > \min(t, t - x + h)$,

$$\hat{\lambda}(s \,|\, x) - \lambda(s \,|\, x) = \lambda(s \,|\, x) \left\{ h^2 \, \beta_2(s, x) + \frac{1}{n} \sum_{i=1}^{n} b_2(s, x \,; S_i, X_i \,|\, h) \right\} + o_p\big\{h^2 + (nh^2)^{-1/2}\big\} \,, \tag{5.6}$$

$$\hat{f}_{S|X}(s \,|\, x) - f_{S|X}(s \,|\, x) = f_{S|X}(s \,|\, x) \left\{ h^2 \, \beta_3(s, x) + \frac{1}{n} \sum_{i=1}^{n} b_2(s, x \,; S_i, X_i \,|\, h) \right\} + o_p\big\{h^2 + (nh^2)^{-1/2}\big\} \,. \tag{5.7}$$

Theorem 1 does not specifically address the order of the bias of $\widehat{F}_{S|X}$ within h of the vertical boundaries, where $x = 0$ or $x = t$. However, the bias there can be shown to be of order h. Formulae for the error-about-the-mean of $\widehat{F}_{S|X}$ in those places are given in (A.8).

The series on the right-hand side of (5.5) is of course a sum of independent and identically distributed random variables, and so (5.5) can be used readily to derive a central limit theorem for $\widehat{F}_{S|X}$. Indeed, $n^{-1} \sum_i b_1(x, s; S_i, X_i \,|\, h)$ is asymptotically normally distributed with zero mean and variance $(nh)^{-1} \sigma_1(s,x)^2$, where $\sigma_1(s,x)^2 = \kappa\, f_X(x)\, E\{c_1(s,x;S,X)^2\}$ and $\kappa = \int K^2$. Therefore, (5.5) implies that

$$\widehat{F}_{S|X}(s \,|\, x) - F_{S|X}(s \,|\, x) = \bar{F}_{S|X}(s \,|\, x) \big\{(nh)^{-1/2}\, \sigma_1(s,x)\, N_n(s,x) + h^2\, \beta_1(s,x)\big\} + o_p\big(h^2\big)\,, \quad (5.8)$$

where the random variable $N_n(s,x)$ is asymptotically normally distributed with zero mean and unit variance.

Standard arguments may be used to prove that, under the conditions of the theorem, $n^{-1} \sum_i b_1(x, s; S_i, X_i \,|\, h)$ is of order $(nh)^{-1/2} (\log n)^{1/2}$, uniformly in (s,x) in the range $0 < t - x < s < s_0$. Hence, (5.5) may also be used to prove that

$$\sup_{\substack{(s,x)\,:\,0<t-x<s<s_0,\\ h\leq x\leq t-h}} \Big|\widehat{F}_{S|X}(s \,|\, x) - F_{S|X}(s \,|\, x)\Big| = O_p\big\{(nh)^{-1/2} (\log n)^{1/2} + h^2\big\}\,.$$

Analogously to the way in which (5.8) follows from (5.5), (5.6) and (5.7) imply that $\big\{\lambda_{S|X}(s \,|\, x)\big\}^{-1}\big\{\hat{\lambda}_{S|X}(s \,|\, x) - \lambda_{S|X}(s \,|\, x)\big\}$ and $\big\{f_{S|X}(s \,|\, x)\big\}^{-1}\big\{\hat{f}_{S|X}(s \,|\, x) - f_{S|X}(s \,|\, x)\big\}$ are each asymptotically normally distributed with respective asymptotic means $h^2\, \beta_2(s,x)$ and $h^2\, \beta_3(s,x)$ and common variance $(nh^2)^{-1}\sigma_2^2(s,x)$, where $\sigma_2^2(s,x) = \kappa^2\, f_{SX}(s,x) E\big\{c_2(s,x;S,X)^2\big\}$.

APPENDIX I: DEFINITIONS OF FUNCTIONS c_j AND β_j

Put

$$\kappa_j(h,s,x) = \int_{v\,:\,\min(0,t-s)<x-vh<t} v^j\, K(v)\, dv\,, \quad \rho_j(h,s,x) = \frac{\kappa_j(h,s,x)}{\kappa_0(h,s,x)}\,,$$

$$c_j(s,x\,;s',x') = \frac{I(j=1)\, I(s' \leq s) + I(j=2)}{\kappa_0(h,s',x)\, r(s',x)} - \big\{I(j=1)\, I(t-x' \leq s) + I(j=2)\big\} \times I(t-x' \leq s') \int_{t-x'}^{s'} \frac{\big\{I(j=1)\, I(u \leq s) + I(j=2)\big\}\, f_{SX}(u,x)}{\kappa_0(h,u,x)\, r(u,x)^2}\, du\,,$$

where $r(s,x) = \int_{u>s} f_{SX}(u,x)\,du$. Let $f_{SX}^{(0,j)}(s,x) = (\partial^j/\partial x^j)\, f_{SX}(s,x)$,

$$g_j(s,x) = \frac{\int_{u>s} f_{SX}^{(0,j)}(u,x)\,du}{r(s,x)}, \quad L_j(s,x) = \frac{f_{SX}(s,x)\, g_j(s,x) - f_{SX}^{(0,j)}(s,x)}{r(s,x)},$$

$$\beta_1(s,x) = L_1(t-x,x) \int_0^{(s-t+x)/h} \rho_1(h, t-x+uh, x)\,du$$
$$- \tfrac{1}{2}\, \rho_2(h,t,x) \left\{ \int_0^s L_2(u,x)\,du \right\} I\big\{s > \min(t, t-x+h)\big\}.$$

Define $\kappa_2 = \int v^2 K(v)\,dv$, $\beta_2(s,x) = \kappa_2 \{g_1(s,x) L_1(s,x) - L_2(s,x)\}$ and $\beta_3(s,x) = \beta_2(s,x) - \beta_1(s,x)$.

APPENDIX II: PROOF OF THEOREM 1

We shall derive only (5.5); proofs of (5.6) and (5.7) are similar. Put

$$r_h(s,x) = \iint_{t-x+hv \le s \le u} K(v)\, f_{SX}(u, x-hv)\,du\,dv,$$

which, for $s + x > t$, converges to $r(s,x)$ as the bandwidth, h, decreases to zero. Note that, with

$$R(s,x) = \frac{1}{n} \sum_{i=1}^{n} I(t - X_i \le s \le S_i)\, K_h(x - X_i),$$

we have $r_h(s,x) = E\{R(s,x)\}$. Without loss of generality, the support of K is confined to $[-1,1]$. Given $\epsilon > 0$, let $\mathcal{R} = \mathcal{R}(\epsilon)$ denote the set of points (s,x) such that $t - x' < s < s_0$ for some $x' \in [0,t]$ with $|x - x'| \le \epsilon$. Under the conditions of the theorem, there exists $\eta > 0$ such that, if $\epsilon > 0$ is sufficiently small and $h \le \epsilon$, then $r_h(s,x) \ge \eta$ for all $(s,x) \in \mathcal{R}$.

Put $\Delta_R(s,x) = R(s,x) - r_h(s,x)$. Standard methods can be used to prove that, for each $\delta > 0$, $\Delta_R(s,x) = O_p\{n^\delta\,(nh)^{-1/2}\}$ uniformly in $(s,x) \in \mathcal{R}$. Therefore,

$$\sup_{0<u<s} \Big| R(u,x)^{-1} - \big\{ r_h(u,x)^{-1} - r_h(u,x)^{-2}\, \Delta_R(u,x) \big\} \Big| = O_p\big\{ n^\delta\, (nh)^{-1} \big\},$$

uniformly in $(s, x) \in \mathcal{R}$. From this result it may be proved that, uniformly in $(s, x) \in \mathcal{R}$,

$$T(s, x) \equiv \frac{1}{n} \sum_{i=1}^{n} \frac{I(S_i \leq s)\, K_h(x - X_i)}{R(S_i, x)} = T_1(s, x) - T_2(s, x) + O_p\{n^{\delta}\,(nh)^{-1}\}\,, \tag{A.1}$$

where

$$T_1(s, x) = \frac{1}{n} \sum_{i=1}^{n} \frac{I(S_i \leq s)\, K_h(x - X_i)}{r_h(S_i, x)}\,,$$

$$T_2(s, x) = \frac{1}{n} \sum_{i=1}^{n} \frac{I(S_i \leq s)\, K_h(x - X_i)\, \Delta_R(S_i, x)}{r_h(S_i, x)^2}\,.$$

Note that

$$\begin{aligned} A(s, x) \equiv E\{T_1(s, x)\} &= E\bigg\{\frac{I(S \leq s)\, K_h(x - X)}{r_h(S, x)}\bigg\} \\ &= \int_0^s \frac{\kappa_0(h, u, x)\, f_{SX}(u, x)}{r_h(u, x)}\, du + \sum_{j=1}^{2} \frac{(-h)^j}{j!} \int_0^s \frac{\kappa_j(h, u, x)\, f_{SX}^{(0,j)}(u, x)}{r_h(u, x)}\, du \\ &\quad + o\big(h^2\big)\,, \\ r_h(s, x) &= \kappa_0(h, s, x)\, r(s, x) \left\{1 + \sum_{j=1}^{2} \frac{(-h)^j\, \rho_j(h, s, x)\, g_j(s, x)}{j!} + o\big(h^2\big)\right\}, \end{aligned}$$

where, here and below, $o(h^2)$ remainder terms are of that order uniformly in $(s, x) \in \mathcal{R}$. We may Taylor-expand $A(s, x)$ in increasing powers of h, obtaining

$$A(s, x) = -\log\left\{\frac{r(s, x)}{r(0, x)}\right\} + h\, A_1(s, x) + h^2\, A_2(s, x) + o\big(h^2\big)\,, \tag{A.2}$$

say, where A_1 and A_2 do not depend on h. From (A.2) it may be proved that if K vanishes outside $[-1, 1]$ then for $0 < t - x < s < s_0$,

$$\exp\{-A(s, x)\} = \bar{F}_{S|X}(s \,|\, x) \left\{1 - h\, C_1(s, x) - h^2\, \beta_1(s, x) + o\big(h^2\big)\right\}, \tag{A.3}$$

where $C_1(s, x) = C_2(s, x) + O(h)$ and

$$C_2(s, x) = I\Big\{|2x - t| > t - 2h,\; s > \min(t, t - x + h)\Big\}\, \rho_1(h, t, x) \int_0^s L_1(u, x)\, du\,.$$

Define

$$B(s,x\,;s_1,s_2,x_1,x_2)=\frac{I(s_1\le s)\,K_h(x-x_1)}{r_h(s_1,x)^2}\,I(t-x_2\le s_1\le s_2)\,K_h(x-x_2)\,,$$

$$B_1(s,x\,;s_1,x_1)=E\{B(s,x\,;s_1,S,x_1,X)\}\,,$$

$$B_2(s,x\,;s_2,x_2)=E\{B(s,x\,;S,s_2,X,x_2)\}$$

and $B_3(s,x)=E\{B_1(s,x\,;S,X)\}=E\{B_2(s,x\,;S,X)\}$. Note that

$$\begin{aligned}T_2(s,x)&=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{I(S_i\le s)\,K_h(x-X_i)}{r_h(S_i,x)^2}\\&\qquad\qquad\times\{I(t-X_j\le S_i\le S_j)\,K_h(x-X_j)-r_h(S_i,x)\}\\&=T_{21}(s,x)+\big(1-n^{-1}\big)\,T_{22}(s,x)+\frac{1}{n^2}\sum_{i=1}^{n}B_1(s,x;S_i,X_i)^2\\&\qquad\qquad-n^{-1}\,T_1(s,x)\,,\end{aligned}\tag{A.4}$$

where $T_{22}(s,x)=n^{-1}\sum_j\{B_2(s,x\,;S_j,X_j)-B_3(s,x)\}$ and

$$\begin{aligned}T_{21}(s,x)=\frac{1}{n^2}\sum\sum_{i\neq j}\Big\{&B(s,x\,;S_i,S_j,X_i,X_j)-B_1(s,x\,;S_i,X_i)\\&-B_2(s,x\,;S_j,X_j)+B_3(s,x)\Big\}\,.\end{aligned}$$

The quantity $T_{21}(s,x)$ can be viewed as the difference between a U-statistic and its projection. In particular, it has zero mean and variance of order $(nh)^{-2}$. By using a lattice approximation we may show that for all $\delta>0$,

$$\sup_{(s,x)\in\mathcal{R}}|T_{21}(s,x)|=O_p\big\{n^{\delta}\,(nh)^{-1}\big\}\,.\tag{A.5}$$

Similarly, defining $T_3(s,x)=T_1(s,x)-E\{T_1(s,x)\}$, it may be shown that for all $\delta>0$,

$$\sup_{(s,x)\in\mathcal{R}}\left\{|T_3(s,x)|+|T_{22}(s,x)|+\left|\frac{1}{n^2}\sum_{i=1}^{n}B_1(s,x;S_i,X_i)^2\right|\right\}=O_p\big\{n^{\delta}\,(nh)^{-1/2}\big\}\,.\tag{A.6}$$

Combining (A.1) and (A.3)–(A.6) we deduce that for all $\delta > 0$,

$$\begin{aligned}\exp\{-T(s,x)\} &= \bar F_{S|X}(s\,|\,x)\,\exp\Big[-T_3(s,x)+T_{22}(s,x)-h\,C_1(s,x)+h^2\,\beta_1(s,x)\\ &\qquad\qquad + o_p\big(h^2\big)+O_p\big\{n^\delta (nh)^{-1}\big\}\Big]\\ &= \bar F_{S|X}(s\,|\,x)\,\Big[1-T_3(s,x)+T_{22}(s,x)-h\,C_1(s,x)-h^2\,\beta_1(s,x)\\ &\qquad\qquad + o_p\big(h^2\big)+O_p\big\{n^\delta (nh)^{-1}\big\}\Big]\,, \end{aligned} \tag{A.7}$$

uniformly in $(s,x) \in \mathcal{R}$.

Define

$$\begin{aligned}\xi_1(s,x\,;s',x') = K_h(x-x')\,\bigg[&\frac{I(s' \le s)}{r_h(s',x)} - I\{t-x' \le \min(s,s')\}\\ &\times \int_{0<x-hv<t} K(v)\,dv \int_{t-x'}^{\min(s,s')} \frac{f_{SX}(u,x-hv)}{r_h(u,x)^2}\,du\bigg]\end{aligned}$$

and $\xi_2(s,x\,;s',x') = \xi_1(s,x\,;s',x') - E\{\xi_1(s,x\,;S,X)\}$. Then, $T_3(s,x) - T_{22}(s,x) = n^{-1}\sum_i \xi_2(s,x\,;S_i,X_i)$, and so (A.7) is equivalent to:

$$\begin{aligned}\frac{\widehat F_{S|X}(s\,|\,x) - F_{SX}(s\,|\,x)}{\bar F_{S|X}(s\,|\,x)} &= h\,C_1(s,x) + h^2\,\beta_1(s,x) + n^{-1}\sum_{i=1}^{n} \xi_2(s,x\,;S_i,X_i)\\ &\qquad + o_p\big(h^2\big) + O_p\big\{n^\delta (nh)^{-1}\big\}\,.\end{aligned} \tag{A.8}$$

For small h, $\xi_1(s,x\,;s',x')$ is well approximated by $a_1(s,x\,;s',x'\,|\,h)$, and therefore $\xi_2(s,x\,;s',x')$ is well approximated by $b_1(s,x\,;s',x'\,|\,h)$. Making these approximations concise, we obtain (5.5) from (A.8).

ACKNOWLEDGEMENTS

The authors thank Dongsheng Tu (National Cancer Institute of Canada Clinical Trials Group and Queen's University) for providing the breast cancer data set.

REFERENCES

BECKER, N.G. AND MARSCHNER, I.C. (1993). A method for estimating the age-specific relative risk of HIV infection from AIDS incidence data. *Biometrika* **80**, 165–178.

EFRON, B. AND PETROSIAN, V. (1994). Survival analysis of the Gamma-ray burst data. *J. Amer. Statist. Assoc.* **89**, 452–462.

CUI, J. (1999). Nonparametric estimation of a delay distribution based on left-censored and right-truncated data. *Biometrics* **55**, 345–349.

GRIGOLETTO, M. AND AKRITAS, M.G. (1999). Analysis of covariance with incomplete data via semiparametric model transformation. *Biometrics* **55**, 1177–1187.

GROSS, S.T. AND LAI, T.L. (1996). Nonparametric estimation and regression analysis with left-truncated and right-censored data. *J. Amer. Statist. Assoc.* **91**, 1166–1180.

HE, S. AND YANG, G.L. (1998). Estimation of the truncation probability in the random truncation model. *Ann. Statist.* **26**, 1011–1027.

HU, X.J. AND LAWLESS, J.F. (1996). Estimation from truncated lifetime data with supplementary information on covariates and censoring times. *Biometrika* **83**, 747–761.

KALBFLEISCH, J.D. AND LAWLESS, J.F. (1989). Inference based on restrospective ascertainment: an analysis of the data on transfusion-related AIDS. *J. Amer. Statist. Assoc.* **84**, 360–372.

KALBFLEISCH, J.D. AND LAWLESS, J.F. (1991). Regression models for right truncated data with applications to AIDS incubation times and reporting lags . *Statistica Sinica* **1**, 19–32.

KEIDING, N. (1991). Age-specific incidence and prevalence: a statistical perspective. *J. Royal Statist. Soc. Ser. A* **154**, 371–412.

LAGAKOS, S.W., BARRAJ, L.M. AND DE GRUTTOLA, V. (1988). Nonparametric analysis of truncated survival data, with applications to AIDS. *Biometrika* **75**, 515–523.

LI, G. (1995). Nonparametric likelihood ratio estimation of probabilities for truncated data. *J. Amer. Statist. Assoc.* **90**, 997–1003.

LI, Y., BETENSKY, R.A., LOUIS, D.N. AND CAIRNCROSS (2002). The use of frailty hazard models for unrecognized heterogeneity that interacts with treatment: considerations of efficiency and power. *Biometrics* **58**, 232–236.

MARTIN, E.C. AND BETENSKY, R.A. (2005). Testing quasi-indpendence of failure and truncation times via conditional Kendall's tau. *J. Amer. Statist. Assco.* **100**, 484–492.

PAGANO, M., TU, X.M., DE GRUTTOLA, V. AND MAWHINNEY, S. (1994). Regression analysis of censored and truncated data: estimating reporting-delay distributions and AIDS incidence from surveillance data. *Biometrics* **50**, 1203–1214.

SUN, L. AND ZHOU, X. (2001). Survival function and density estimation for estimation for truncated dependent data. *Statist. Probab. Letters* **52**, 47–57.

TSAI, W.Y. (1990). Testing the assumption of independence of truncation time and failure time. *Biometrika* 77, 169–177.

WANG, M.-C., BROOKMEYER, R. AND JEWELL, N.P. (1993). Statistical models for prevalent cohort data. *Biometrics* **49**, 1-11.

WOODROOFE, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13**, 163–177.

CAPTIONS FOR FIGURES

Caption for Figure 1: *Data.* In panel (a) the truncated data are illustrated by plotting $\widetilde{X} + \widetilde{S}$, the time of first recurrence of breast cancer, against $\widetilde{X}$, the time of first diagnosis of breast cancer, for the same patient. Panel (b) shows a scatter plot of $(X, S) = (t - \widetilde{X}, t - \widetilde{S})$, i.e. the truncated data after applying the time-reversal transformation (4.1).

Caption for Figure 2: *Conditional distribution in the truncated population.* In panel (a) (respectively, panel (b)) our estimates of $F_{\widetilde{S}|\widetilde{X}}(\widetilde{s}|\widetilde{x})$ for $x = 220$, 230, 240, 250, 260 (respectively, $\widetilde{x} = 270$, 280, 290, 300, 310) are represented by the solid, dashed, dotted, dot-dashed and long-dashed lines, respectively.

Caption for Figure 3: *Conditional and unconditional distribution functions in the original population.* Panels (a) and (b) plot the estimates of $F_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s}|\widetilde{x})$ when $\widetilde{x} = 220, 230, 240, 250, 260$ and $\widetilde{x} = 270$, 280, 290, 300, 310, respectively. Line types for different values of $\widetilde{x}$ are as in the case of Figure 2. Panel (c) graphs the product-limit estimator of $F_{\widetilde{S}^*}(\widetilde{s})$.

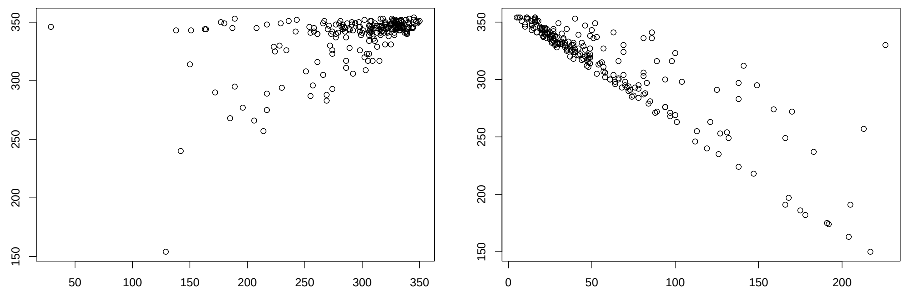
Caption for Figure 4: *Conditional and unconditional hazard rates in the original population.* Panels (a) and (b) plot the estimates of $\lambda_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s}|\widetilde{x})$ when $\widetilde{x} = 220$, 230, 240, 250, 260 and $\widetilde{x} = 270$, 280, 290, 300, 310, respectively. Line types for different values of $\widetilde{x}$ are as in the case of Figure 2. Panel (c) shows an estimate of $\lambda_{\widetilde{S}^*}(\widetilde{s})$, based on smoothing the Nelson-Aalen estimator.

Caption for Figure 5: *True hazard rates in the original distribution.* The solid line is the true overall hazard $\lambda_{\widetilde{S}^*}(\widetilde{s})$. In panel (a), the dashed, dotted, dot-dashed and long-dashed lines respectively depict the true conditional hazard functions $\lambda_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s}|\widetilde{x})$ for $\widetilde{x} = 210$, 230, 250 and 270. The dashed lines in panel (b) are the 10%, 50% and 90% pointwise quantiles of the product-limit estimator of $\lambda_{\widetilde{S}^*}(\widetilde{s})$

(under the independence assumption). Panel (c) depicts five typical realizations of the product-limit estimator that have the 10% (dashed), 30% (dotted), 50% (dot-dashed), 70% (long-dashed) and 90% (dotted-long-dashed) integrated squared error values.

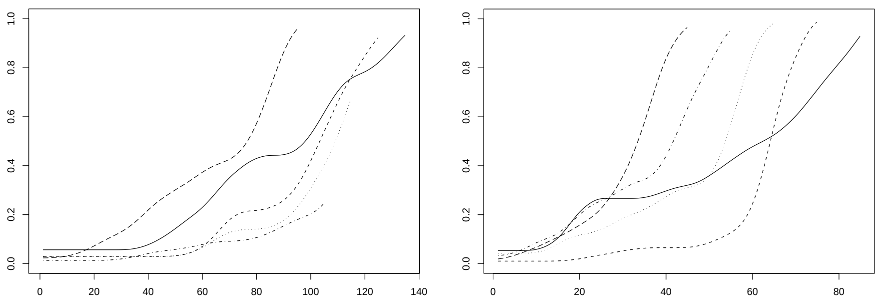
Caption for Figure 6: *Variability plot of the estimator of* $\lambda_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s}|\widetilde{x})$. The dashed lines in panels (a)–(d) are the 10th, 50th and 90th pointwise percentiles of our estimator of $\lambda_{\widetilde{S}^*|\widetilde{X}^*}(\widetilde{s}|\widetilde{x})$ for $\widetilde{x} = 210$, 230, 250 and 270, respectively. The solid lines represent the corresponding true conditional hazard functions.

Caption for Figure 7: *Variability plot of the estimator of* $\lambda_{\widetilde{S}|\widetilde{X}}(\widetilde{s}|\widetilde{x})$. Interpretation is as for Figure 6.



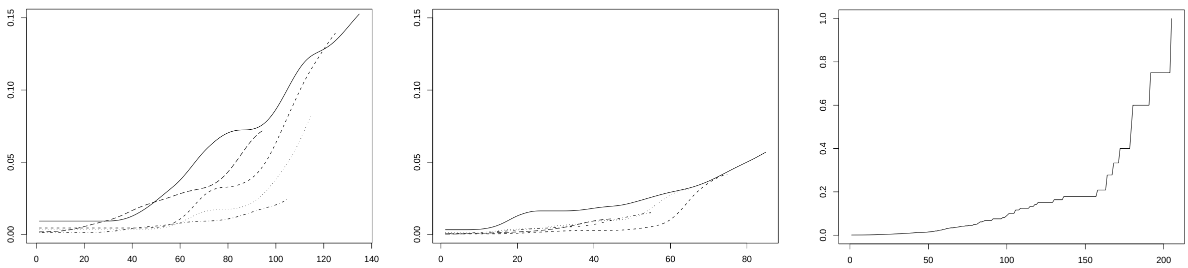
(a) (b)

Figure 1.



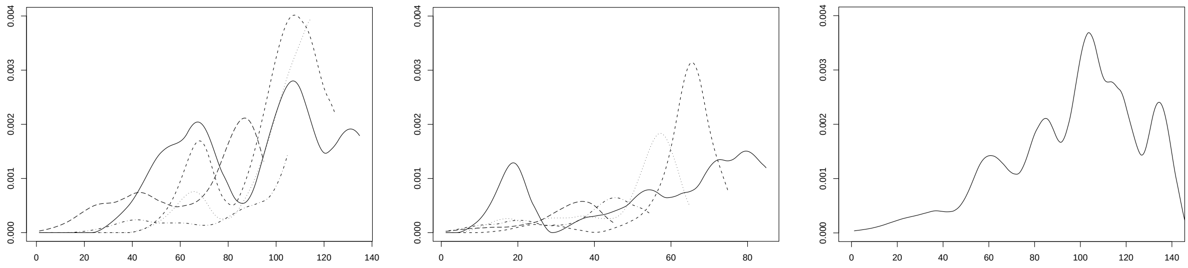
(a) (b)

Figure 2.



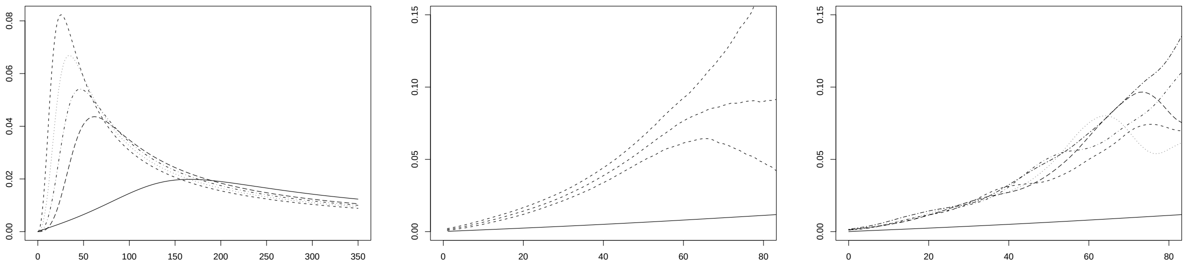
(a) (b) (c)

Figure 3.



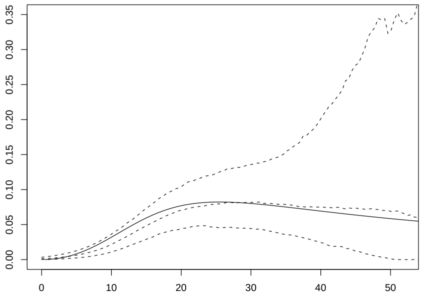
(a) (b) (c)

Figure 4.

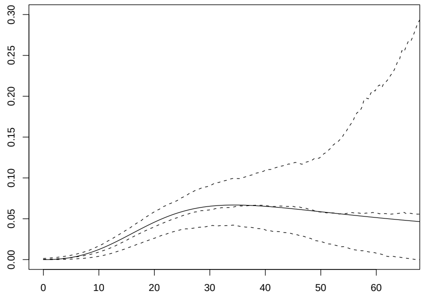


(a) (b) (c)

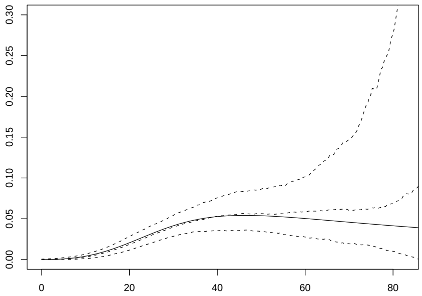
Figure 5.



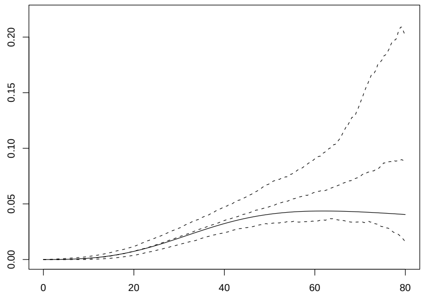
(a)



(b)

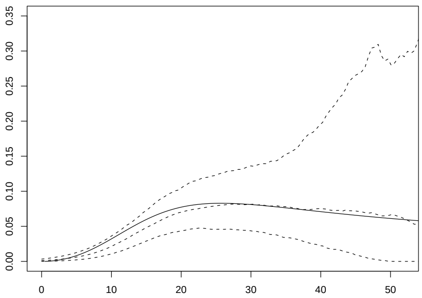


(c)

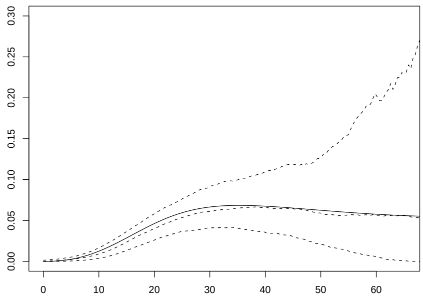


(d)

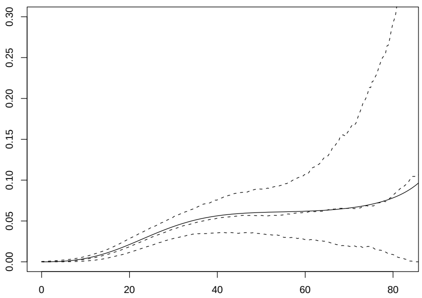
Figure 6.



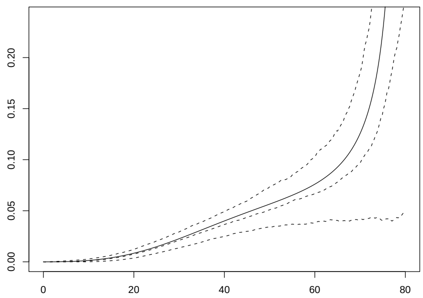
(a)



(b)



(c)



(d)

Figure 7.