Parametrically guided nonparametric density and hazard estimation with censored data

Majda Talamakrouni [*] Ingrid Van Keilegom [*,§]
Anouar El Ghouch [*]

Abstract

The parametrically guided kernel smoother is a promising nonparametric estimation approach that aims to reduce the bias of the classical kernel density estimator without increasing its variance. Theoretically, the estimator is unbiased if a correct parametric guide is used, which can never be achieved by the classical kernel estimator even with an optimal bandwidth. The estimator is generalized to the censored data case and used for density and hazard function estimation. The asymptotic properties of the proposed estimators are established and their performance is evaluated via finite sample simulations. The method is also applied to data coming from a study where the interest is in the time to return to drug use.

Key Words: Cox model; Density estimation; Kaplan-Meier estimator; Kernel smoothing; Maximum likelihood; Right censoring.

[*]Institut de statistique, biostatistique et sciences actuarielles (ISBA), Université catholique de Louvain.

[§]Correspondence to: Institut de statistique, biostatistique et sciences actuarielles, Université catholique de Louvain, Voie du Roman Pays 20, B 1348 Louvain-la-Neuve, Belgium.

E-mail address: ingrid.vankeilegom@uclouvain.be (I. Van Keilegom).

1 Introduction

Censored data appear in a broad variety of research studies with practical applications. Random right censoring is one of the most common types of censoring. For example in medical, economic or engineering studies, it frequently happens that the variable of interest T is only partially observed due to the earlier occurrence of a censoring event. In such studies, the estimation of the probability density and hazard function of T has received considerable attention in the literature, as it allows to visualize and explore the distribution of data.

In this paper we wish to estimate the density and hazard function when T is subject to right censoring, by using a hybrid estimation method that has at the same time nonparametric and parametric ingredients. These two extremal estimation approaches have rather opposite characteristics. The fully parametric approach is accurate and powerful when the parametric family is correctly chosen, otherwise it can lead to incorrect inference. The fully nonparametric approach includes several methods, among which the popular kernel smoothing procedure. It is very flexible, since it does not rely on any restrictive assumptions about the form of the underlying density or hazard function. However, the resulting estimator has typically a slower rate of convergence.

In the case where the data are not subject to censoring, there is a large variety of approaches to estimate the density and the hazard function that are either semiparametric or that use aspects from both the nonparametric and the parametric school, and that are hence situated in between these two extreme approaches. One of these approaches is the parametrically guided nonparametric estimator proposed by Hjort and Glad (1995). Apart from reducing the bias compared to the classical kernel approach, the parametrically guided nonparametric approach allows for a theoretically unbiased estimator, which is impossible with the classical kernel approach. The basic idea of this approach is to start with any parametric density estimator and then to adjust this first stage parametric approximation using a nonparametric kernel-type estimator of a particular correction factor. More precisely, the key identity underlying the parametrically guided nonparametric approach is

$$f(t) = f_{\widehat{\theta}}(t) r_{\widehat{\theta}}(t),$$

where $r_{\widehat{\theta}}(t) = \frac{f(t)}{f_{\widehat{\theta}}(t)}$, $f_{\widehat{\theta}}(t)$ is a first stage parametric density approximation and $\widehat{\theta}$ is an estimator of the least false value θ^* according to a certain distance measure between $f(\cdot)$ and $f_\theta(\cdot)$ (see Assumption 3.3, below). Hjort and Glad (1995) defined the parametrically guided nonparametric estimator by

$$\widehat{f}_{\widehat{\theta}}(t) = f_{\widehat{\theta}}(t)\widehat{r}_{\widehat{\theta}}(t), \tag{1.1}$$

where $\widehat{r}_{\widehat{\theta}}(\cdot)$ is a kernel-type nonparametric estimator of the correction factor $r_{\widehat{\theta}}(\cdot)$. Essentially, this multiplicative correction does not affect the variance but can reduce the bias. The intuitive idea behind this approach is that if the parametric estimator $f_{\widehat{\theta}}(\cdot)$ is close to the true density $f(\cdot)$, the multiplicative correction function $r_{\widehat{\theta}}(\cdot)$ will be smoother than the true density $f(\cdot)$ and therefore simpler to estimate using kernel smoothing, resulting in an improved $\widehat{f}_{\widehat{\theta}}(\cdot)$ compared to the traditional kernel estimator. If the true density is far from the parametric estimator, then there is not much loss in accuracy for the parametrically guided nonparametric estimator.

The aim of this paper is to extend their method to the case of censored data. To the best of our knowledge, except for the recent work of Talamakrouni *et al.* (2014), who studied a guided local linear estimator of a regression function when the response is subject to censoring, the parametrically guided nonparametric method has never been investigated in the context of censored data. In addition to studying the estimation of the density function, we also propose and study a parametrically guided nonparametric estimator of the hazard rate function in the presence of censoring.

Apart from the above parametrically guided nonparametric estimator of Hjort and Glad (1995), there have been other proposals in the literature that combine the nice features of both the parametric and the nonparametric approach. These methods are quite different but can also achieve bias reduction compared to the fully nonparametric method. As far as we are aware of, except for the paper of Copas (1995) who adapted a local maximum likelihood estimator to censored data, none of them has been considered so far in the context of censored data. First of all, we find the projection pursuit density estimation developed by Friedman *et al.* (1984) for a multivariate density using a similar multiplicative correction. Hjort (1986) and Buckland (1992) introduced similar ideas using an estimated orthogonal expansion for the multiplicative correction factor. Hjort and Jones (1996) proposed a local parametric density estimator based on a local kernel smoothed likelihood function. This approach has a similar intention as the approach of Copas (1995) but is somehow more general. Another class of local likelihood methods has been discussed by Eguchi and Copas (1998). Efron and Tibshirani (1996) combined the maximum likelihood and the kernel estimator by putting an exponential family through a kernel estimator. Other semiparametric estimators involving an extra parameter have been proposed in the literature. For example, Olkin and Spiegelman (1987) and Faraway (1989) considered a convex combination of a parametric and a nonparametric estimate, and afterwards, Naito (2004) constructed a class of semi-parametric estimators using a local L_2-fitting criterion to estimate the correction factor. Finally, more recently, Veraverbeke *et al.* (2014) discussed a parametrically pre-adjusted nonparametric density estimator.

Parallel to this vast literature on parametrically guided density estimation, there also exists a large literature on parametrically guided nonparametric regression. We mention for example Glad (1998), Martin-Filho *et al.* (2008) and Fan *et al.* (2009), among others.

The paper is organized as follows. The next section explains in detail the proposed methodology. Section 3 provides some asymptotic results for the proposed estimators, while Section 4 investigates the finite sample properties of the new estimators. In Section 5 we apply the proposed method to data on the time to return to drug use from a study of the AIDS research unit of the University of Massachusetts. Finally, some general conclusions are drawn in Section 6. The proofs are collected in the Appendix.

2 Methodology

Let T be a variable of interest with density f and distribution function F, and let C be a censoring variable with continuous distribution function G. We assume throughout our paper that T is independent of C. Under random right censoring, the variable T is not completely observed. One can only observe (X, δ), where $X = \min(T, C)$, $\delta = I(T \leq C)$ and $I(\cdot)$ is the indicator function. Our first objective is to estimate the probability density function f using the observed i.i.d sample $(X_i, \delta_i), i = 1, \dots, n$ of (X, δ).

The kernel-based density estimator that we are currently investigating has been extended to censored data by Blum and Susarla (1980), among others. The estimator is based on the Kaplan-Meier (1958) estimator $\widehat{F}$ of the distribution function F and is defined as follows:

$$\widehat{f}(t) = \frac{1}{h} \int_{-\infty}^{+\infty} K\Big(\frac{t-s}{h}\Big) d\widehat{F}(s), \tag{2.2}$$

where K is a kernel function, $0 < h \equiv h_n$ is a bandwidth and $\widehat{F}(t)$ is defined by (in the absence of ties)

$$\widehat{F}(t) = 1 - \prod_{i:X_i \leq t} \Big(1 - \frac{1}{\sum_{j=1}^{n} \mathbf{1}_{\{X_j \geq X_i\}}}\Big)^{\delta_i}. \tag{2.3}$$

We also aim to estimate the hazard function $\lambda(\cdot)$ defined by $\lambda(t) = f(t)/(1 - F(t))$. A natural nonparametric estimator for the hazard function can be formed by dividing the kernel density estimator by the Kaplan-Meier estimator of the survival function $1 - F(\cdot)$:

$$\widehat{\lambda}(t) = \widehat{f}(t)/1 - \widehat{F}(t).$$

In this framework, the properties of the kernel density and hazard estimators have been studied by Blum and Susarla (1980), Földes *et al.* (1981), Tanner and Wong (1983), Padgett and McNichols (1984), Mielniczuk (1986), Lo *et al.* (1989), Xiang (1994) and Giné and Guillou (2001), among others.

Note that, the kernel estimators defined above are by construction completely nonparametric. In the uncensored data context, Hjort and Glad (1995) proposed a parametrically guided kernel density estimator (PGK) as an alternative to the traditional kernel density estimator (TK). As argued in the introduction, the PGK estimator combines the advantages of both parametric and nonparametric approaches and includes a prior information that allows the bias reduction of the PGK estimator compared to the TK estimator.

For censored data, we propose to multiply the first stage parametric estimator $f_{\widehat{\theta}}(t)$ in expression (1.1) with the following nonparametric kernel-type estimator of the correction function $r_{\widehat{\theta}}(t)$ adapted to censored data:

$$\widehat{r}_{\widehat{\theta}}(t) = \frac{1}{h}\int_{-\infty}^{+\infty} K\Big(\frac{t-s}{h}\Big)\frac{1}{f_{\widehat{\theta}}(s)}d\widehat{F}(s).$$

The ensuing PGK density estimator is

$$\begin{aligned}\widehat{f}_{\widehat{\theta}}(t) &= \frac{1}{h}\int_{-\infty}^{+\infty} K\Big(\frac{t-s}{h}\Big)\frac{f_{\widehat{\theta}}(t)}{f_{\widehat{\theta}}(s)}d\widehat{F}(s)\\ &= \frac{1}{h}\sum_{i=1}^{n} K\Big(\frac{t-X_i}{h}\Big)\frac{f_{\widehat{\theta}}(t)}{f_{\widehat{\theta}}(X_i)}W_i, \qquad (2.4)\end{aligned}$$

where W_i is the size of the jump of $\widehat{F}$ at X_i .
Note that when the data are completely observed, the weights $W_{(i)}$ are all equal to $1/n$ and the PGK estimator given above reduces to the estimator defined by Hjort and Glad (1995). Naturally the PGK estimator that we propose for the hazard function $\lambda(\cdot)$ is

$$\widehat{\lambda}_{\widehat{\theta}}(t) = \widehat{f}_{\widehat{\theta}}(t)/(1 - \widehat{F}(t)). \qquad (2.5)$$

As we will see in the following section, the multiplicative correction used in the PGK density and hazard function estimators does not affect the variance but can reduce the bias compared to the traditional kernel estimators defined above.

3 Asymptotic results

This section is devoted to the development of the asymptotic normality of the PGK estimators $\widehat{f}_{\widehat{\theta}}(\cdot)$ and $\widehat{\lambda}_{\widehat{\theta}}(\cdot)$. For the PGK density estimator, we split the problem into two parts. First, we establish in Theorem 3.1 the asymptotic normality of $\widehat{f}_*(\cdot)$, an estimator of $f(\cdot)$ based on a given non-random guide $f_*(\cdot)$. Then, in Theorem 3.2, we extend this result to the case of a data-driven guide. Finally, in Theorem 3.3 we prove the asymptotic normality of the PGK estimator of the hazard function $\widehat{\lambda}_{\widehat{\theta}}(\cdot)$.

As stated in the previous section, under random right censoring the PGK estimator depends on the Kaplan-Meier estimator $\widehat{F}$, which is defined as a product (see expression (2.3)). This adds some extra complexity to the PGK estimation approach compared to the uncensored case. To circumvent these technical difficulties we mainly use the asymptotic i.i.d. representation of the Kaplan-Meier estimator investigated in Lo *et al.* (1989).

Let $\tau < \tau_H$, where $\tau_H = \sup\{t : H(t) < 1\}$ is the right endpoint of the distribution function $H(t) = P(X \leq t)$. Also, let $H_1(t) = P(X \leq t, \delta = 1)$, and define $\mu_K^2 = \int u^2 K(u) du$. The kernel function $K : \mathbb{R} \to \mathbb{R}$, the bandwidth h and the density $f(\cdot)$ are assumed to satisfy the following conditions for a fixed value $t \leq \tau$.

Assumption 3.1.

(A.1) *The kernel K is a symmetric, continuously differentiable probability density function with compact support* $[-1, 1]$.

(A.2) *The bandwidth sequence h satisfies $h \to 0$ and $nh^2(\log n)^{-2} \to \infty$.*

(A.3) *f is twice continuously differentiable in a neighborhood of t and $f(t) > 0$.*

3.1 Guided kernel density estimator with a fixed guide

Let $f_*(t)$ be a non-random density function that approximates $f(t)$, and let $\widehat{f}_*(t)$ be the corresponding PGK estimator defined as

$$\widehat{f}_*(t) = \frac{1}{h}\int_{-\infty}^{+\infty} K\left(\frac{t-s}{h}\right)\frac{f_*(t)}{f_*(s)}d\widehat{F}(s). \tag{3.6}$$

In the next section we will replace $f_*(\cdot)$ by the best approximation of $f(\cdot)$ within a certain parametric class, but for the time being $f_*(\cdot)$ can be any deterministic density.

Note that if $f_*(t)$ would be a uniform density, then $\widehat{f}_*(t)$ reduces to the TK estimator, which means that the PGK estimator is a generalization of the TK estimator. The following additional conditions are required for a fixed point $t \leq \tau$ at which we want to estimate the density.

Assumption 3.2.

(B.1) The density $f_*(\cdot)$ is twice continuously differentiable in a neighborhood of t.

(B.2) The density $f_*(\cdot)$ satisfies $f_*(t) > 0$.

The following theorem provides the asymptotic distribution of the PGK estimator $\widehat{f}_*(\cdot)$ using a non-random guide.

Theorem 3.1. *Suppose Assumptions 3.1 and 3.2 hold.*

1. *Then,*

$$\widehat{f}_*(t) - f(t) = \frac{1}{nh}\sum_{i=1}^{n} U_{in}(t) + \frac{1}{h}\int_{-\infty}^{+\infty} K\left(\frac{t-s}{h}\right)\frac{f_*(t)}{f_*(s)}dF(s) - f(t) + O_p(n^{-1/2}),$$

where

$$U_{in}(t) = \int_{-1}^{1} \xi_i(t-uh)K'(u)du,$$
$$\xi_i(t) = \int_{-\infty}^{X_i\wedge t} \frac{dH_1(s)}{(1-H(s))^2} + \frac{I\{X_i \leq t, \delta_i = 1\}}{1-H(X_i)}.$$

2. *Moreover,*

$$\sqrt{nh}\Big(\widehat{f}_*(t) - f(t) - B_*(t) + o(h^2)\Big) \xrightarrow{d} \mathcal{N}\Big(0, \sigma^2(t)\Big), \tag{3.7}$$

where

$$B_*(t) = \frac{1}{2}h^2\mu_K^2 r_*''(t) f_*(t),$$

$r_*(t) = f(t)/f_*(t)$ *and* $\sigma^2(t) = [f(t)/(1-G(t))]\int_{-1}^{1} K^2(u)du.$

Note that the choice of the guide has an obvious impact on the expression of the asymptotic bias $B_*(t)$, whilst the variance $\sigma^2(t)$ is invariant under this choice and is the same as for the TK estimator.

3.2 Guided kernel density estimator with an estimated guide

In this section, we investigate the situation where the guide is derived from the data by a first stage estimation procedure. We consider a possibly misspecified parametric model $\{f_\theta(\cdot) : \theta \in \Theta\}$ and assume that there exists an estimator $\widehat{\theta}$ that converges in probability to a finite limit θ^*. We need the following additional conditions for a fixed point $t \leq \tau$.

Assumption 3.3.

(C.1) *The parametric density function f_θ belongs to a parametrically indexed class defined by the following characteristics:*

1. *$\theta \in \Theta$, where Θ is a compact subset of $\mathbb{R}^p$.*
2. *The function $(t, \theta) \mapsto f_\theta(t)$ is twice continuously differentiable with respect to t and the components of θ in a neighborhood of t and θ^*.*

(C.2) *The parameter $\theta^* \in \Theta$ satisfies the following conditions:*

1. *$\widehat{\theta} - \theta^* = O_p(n^{-1/2})$.*
2. *The density $f_{\theta^*}(\cdot)$ satisfies $f_{\theta^*}(t) > 0$.*

In order to be as general as possible, in this paper, we don't restrict ourselves to a particular parametric estimation procedure. However, to illustrate the idea and give an example of an estimator that satisfies the conditions above, especially assumption (C.2.1), we discuss now the case of the maximum likelihood estimator (MLE). Define

$$\theta^* = \arg\max_{\theta\in\Theta} \int_{-\infty}^{+\infty} \log(f_\theta(t))dF(t).$$

This is the minimizer of $KL(f, f_\theta) = \int \log \frac{f(t)}{f_\theta(t)}\, dF(t)$, the Kullback-Leibler distance measure between the true density f and the parametric density model f_θ. If the parametric model is correct, i.e. if there exists a $\theta_0 \in \Theta$ such that $f(\cdot) = f_{\theta_0}(\cdot)$, then $\theta^* = \theta_0$. In the uncensored case, it is well known that the usual MLE given by $\widehat{\theta} = \arg\max_{\theta\in\Theta} n^{-1} \sum_i \log(f_\theta(X_i))$ is consistent for θ^*, even under misspecification. In the censored case, the analogue of $\widehat{\theta}$ is the approximate maximum likelihood estimator (AMLE) proposed by Oakes (1986) and defined as

$$\widehat{\theta} = \arg\max_{\theta\in\Theta} \int_{-\infty}^{+\infty} \log(f_\theta(t))\, d\widehat{F}(t),$$

where $\widehat{F}(\cdot)$ is the Kaplan-Meier estimator. Note that, in the uncensored case, the Kaplan-Meier estimator coincides with the empirical distribution function and therefore the AMLE reduces to the MLE. The properties of the AMLE estimator have been investigated by Suzukawa *et al.* (2001). Assuming that $\tau_F := \sup\{t : F(t) < 1\} \leq \tau_G := \sup\{t : G(t) < 1\}$ and under certain regularity assumptions, the authors prove that $\widehat{\theta}$ is $\sqrt{n}-$consistent.

Remark 3.1. *Even if the data are censored, the usual MLE estimator $\widehat{\theta}$ can still be used. However, under misspecification, this estimator does not convergence to θ^* but to another quantity; see Suzukawa et al. (2001) for more details.*

The following theorem is the most important result of the paper. It establishes that the PGK estimator with an estimated guide $f_{\widehat{\theta}}(t)$ is asymptotically equivalent to the PGK estimator with the fixed guide $f_{\theta^*}(t)$.

Theorem 3.2. *Suppose Assumptions 3.1 and 3.3 hold. Then,*

$$\sqrt{nh}\Big(\widehat{f}_{\widehat{\theta}}(t) - f(t) - B_{\theta^*}(t) + o(h^2)\Big) \xrightarrow{d} \mathcal{N}\Big(0, \sigma^2(t)\Big),$$

where $B_{\theta^*}(t) = \frac{1}{2}h^2\mu_K^2 r''_{\theta^*}(t) f_{\theta*}(t)$ and $r_{\theta^*}(t) = f(t)/f_{\theta*}(t)$.

First, notice that the expression of the asymptotic variance is independent of the parametric estimating procedure and is equal to that of the TK estimator. As revealed in the previous section by Theorem 3.1, the main difference between the behavior of the PGK estimator and the TK estimator appears in the term of the asymptotic bias $B_{\theta^*}(t)$, which depends on the parametric guide. Remind that, if the parametric guide is the uniform density, then the PGK estimator becomes the traditional kernel density estimator and $B_{\theta^*}(t)$ coincides with $B(t) = \frac{1}{2}h^2\mu_K^2 f''(t)$, the asymptotic bias of the TK estimator; see, for example, Lo *et al.* (1989). So, with an appropriate choice of the guide, i.e. when $|r''_{\theta^*}(t) f_{\theta*}(t)| \leq |f''(t)|$, the bias of the PGK estimator will be reduced in absolute value compared to that of the TK estimator, whilst the variance remains unchanged. If the parametric density is a good guess, then the correction function $r_{\theta^*}(\cdot)$ will be nearly constant and its second derivative $r''_{\theta^*}(\cdot)$ should be very small. In this case the bias reduction will be attained. Finally, in the ideal case when the parametric guide coincides with the true density we have that $B_{\theta^*}(t) = 0$. In such a case, the PGK estimator is unbiased and one can choose an arbitrarily large bandwidth to reduce the variance to its minimal possible value, which can never be achieved by the TK estimator even with an optimal bandwidth.

Remark 3.2. *In practice, the choice of the bandwidth is a crucial issue in kernel-based density estimation. The theoretical optimal bandwidth that minimizes the asymptotic mean integrated squared error (MISE) criterion is given by*

$$h_{opt} = \left(\frac{\int \sigma^2(t) dt}{\mu_K^4 \int (r''_{\theta^*}(t) f_{\theta*}(t))^2 dt} \right)^{1/5} n^{-1/5}.$$

This expression can hardly be used in practice, since it depends on many unknown components. To select the bandwidth h in our case, one can use for example the least squares cross validation method or the bootstrap method discussed in Sánchez-Sellero et al. (1999). In our data analysis, see Section 5, we adopt the cross validation method as discussed in Wang and Wang (2007).

We also point out that, the PGK method will work even if the parametric guide is not optimally chosen. However, an optimal choice of the parametric guide will improve the quality of the PGK estimator. One can for example use goodness-of-fit tests to choose the parametric guide; see for example Castro-Kuriss (2011).

3.3 Guided kernel hazard estimator with an estimated guide

The hazard function $\lambda(\cdot)$ has been extensively studied in the literature. The estimation by means of kernel methods has been investigated by Gefeller and Dette (1992), Gefeller and Michels (1992), Patil (1993), Müller and Wang (1994) and González-Manteiga *et al.* (1996), among others. The PGK estimator for the hazard function $\lambda(\cdot)$ that we proposed in the previous section is

$$\widehat{\lambda}_{\widehat{\theta}}(t) = \widehat{f}_{\widehat{\theta}}(t) / [1 - \widehat{F}(t)],$$

where $\widehat{f}_{\widehat{\theta}}(t)$ is the PGK density estimator given in (2.4) and $\widehat{F}(t)$ is the Kaplan-Meier estimator. Note that one can also replace the Kaplan-Meier estimator by a parametrically guided non-parametric version of the distribution function F or by any other estimator that has parametric and nonparametric ingredients (see e.g. Veraverbeke *et al.* (2014), Section 7.1, for an overview of possible estimators). However, given that the rate of convergence of the estimator of $F(t)$ will always be faster than the rate of convergence of the density estimator $\widehat{f}_{\widehat{\theta}}(t)$, the choice of the estimator of $F(t)$ has no impact on the asymptotic distribution of the estimator of $\lambda(t)$. For simplicity we therefore estimate $F(t)$ by the Kaplan-Meier estimator $\widehat{F}(t)$.

The following theorem deals with the asymptotic normality of the PGK hazard rate estimator.

Theorem 3.3. *Suppose Assumptions 3.1 and 3.3 hold. Then,*

$$\sqrt{nh}\Big(\widehat{\lambda}_{\widehat{\theta}}(t) - \lambda(t) - \beta_{\theta^*}(t) + o(h^2)\Big) \xrightarrow{d} \mathcal{N}\Big(0, \tau^2(t)\Big),$$

where

$$\beta_{\theta^*}(t) = \frac{1}{2}h^2\mu_K^2 r''_{\theta^*}(t) f_{\theta^*}(t)/[1-F(t)],$$

and $\tau^2(t) = [\lambda(t)/(1-H(t))]\int K^2(u)du$.

As for the density estimator, the asymptotic bias of the PGK hazard rate estimator depends on the parametric guide, while the asymptotic variance remains unchanged compared to the TK hazard estimator.

4 Simulation results

In this section we evaluate the finite sample performance of the PGK estimator by means of Monte Carlo simulations. To check the theoretical results and compare the PGK estimator with the TK estimator we investigate two examples. In the first example we study the classical class of normal mixture densities of Marron and Wand (1992), and in the second example we investigate the Weibull density and perform a comparison with the logspline approach of Kooperberg and Stone (1991, 1992). Along the simulations we consider the Epanechnikov kernel function K, and, for every estimator, we only show the results corresponding to the optimal tuning parameters, i.e. those which minimize the empirical mean squared error (MSE).

4.1 Normal mixture model

The class of normal mixture densities of Marron and Wand (1992) includes fifteen densities that cover a broad variety of shapes. In the context of uncensored guided density estimation, this class was investigated by Hjort and Glad (1994) and Naito (2004), among others. We studied all the fifteen densities but for the sake of brevity we only show the results of the following ones: the normal density, #1, the bimodal density, #6, the separated bimodal density, #7, the asymmetric bimodal density, #8, the trimodal density, #9, the claw density, #10, the double claw density, #11, the asymmetric double claw density, #13, and the smooth comb density, #14. See Figure 1 for a plot of all these densities. In each case, independent and identically distributed variables T_i, $i = 1, \ldots, 400$, are drawn. Independently, we drawn the censoring variables C_i from the same distribution. This leads to 50% rate of censoring. As for the parametric guide, we consider a standard normal density whose parameters are estimated by maximum likelihood. So the first case (#1) is the only situation where the guide is correctly specified. We compute the PGK and the TK density estimators at $t = 0$ taking 100 equally spaced bandwidths over the interval $[0.01, 4]$. The squared bias ($\text{Bias}^2 \times 10^4$), the variance ($\text{Var}\times 10^4$) and the empirical mean squared error ($\text{MSE} \times 10^4$) of each estimator were computed. Table 1 provides the results with 1000 replications for cases #1, #6, #7, #8, #9, #10, #11, #13, and #14. As already mentioned, #1 corresponds to the case where the guide is a good guess of the true density. As expected, in this case, the bias of the PGK estimator is almost zero and is substantially reduced compared to that of the TK estimator. For cases #8, #9 and #11 the PGK estimator is significantly better than the TK estimator. For case #10 the MSE becomes very large and reveals unstable behavior for both estimators. In addition, although the MSE is not greatly enhanced, the bias and the MSE of the PGK estimator remain the smallest. Cases #7 and #14 show a quite similar behavior of the PGK and the TK estimators. However, even if the bias is still reduced, the TK estimator beats the PGK estimator in terms of the MSE, for cases #6 and #13. Finally, we point out that the selected bandwidths for both competitors were the same for most of the non-normal situations.

Figure 2 shows how the MSE changes with h for three cases: #1, #10 and #13. As expected, in case #1, where the guide is correct, the bias is almost zero. In this situation increasing the bandwidth reduces the MSE of the PGK estimator compared to that of the TK estimator. For case #10 the parametric guide is well specified and the MSE of the PGK estimator is significantly reduced especially for large bandwidths. For case #13 the parametric guide is completely misspecified and the MSE of the PGK estimator is now larger than that of the TK estimator for most choices of h. Finally, in all cases, similar behavior of both competitors is observed for small values of h. This is not surprising since with a small bandwidth only the $X_i's$ quite close to t are used and the ratio $f_{\widehat{\theta}}(t)/f_{\widehat{\theta}}(X_i)$ defined in expression (2.4) is close to 1 in that case.

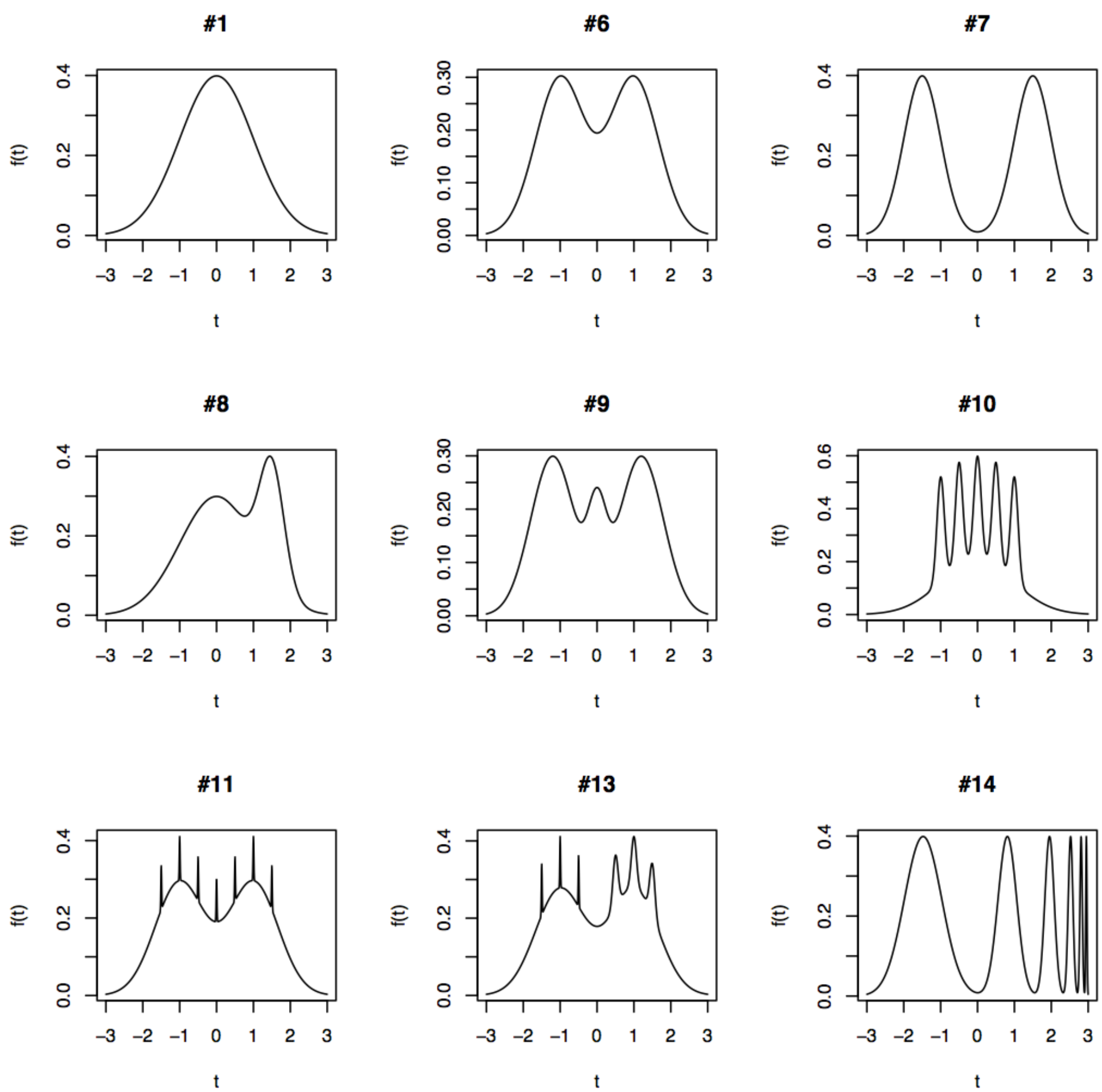

Figure 1: Standard normal and normal mixture target densities.

Table 1: Squared bias ($\times 10^4$), Variance ($\times 10^4$), MSE ($\times 10^4$) and the optimal bandwidth h of the estimators of several normal mixture densities, for samples of size $n = 400$, with a censoring rate of 50% and $N = 1000$ replications.

	Method	Bias^2	Var	MSE	h
#1	PGK	0.015	4.376	4.391	0.332
	TK	3.764	8.556	12.32	0.736
#6	PGK	0.003	2.352	2.355	3.234
	TK	0.008	0.350	0.358	1.582
#7	PGK	0.207	0.866	1.073	0.171
	TK	0.196	0.860	1.056	0.171
#8	PGK	0.008	3.067	3.075	3.435
	TK	2.467	5.372	7.839	0.777
#9	PGK	0.001	2.663	2.664	2.307
	TK	3.479	1.161	4.640	0.856
#10	PGK	18.72	135.3	154.0	0.050
	TK	19.40	135.1	154.5	0.050
#11	PGK	0.006	2.977	2.983	0.775
	TK	37.99	1.590	39.58	0.735
#13	PGK	0.001	1.563	1.564	3.476
	TK	0.012	0.305	0.317	1.783
#14	PGK	0.381	1.210	1.591	0.131
	TK	0.368	1.204	1.572	0.131

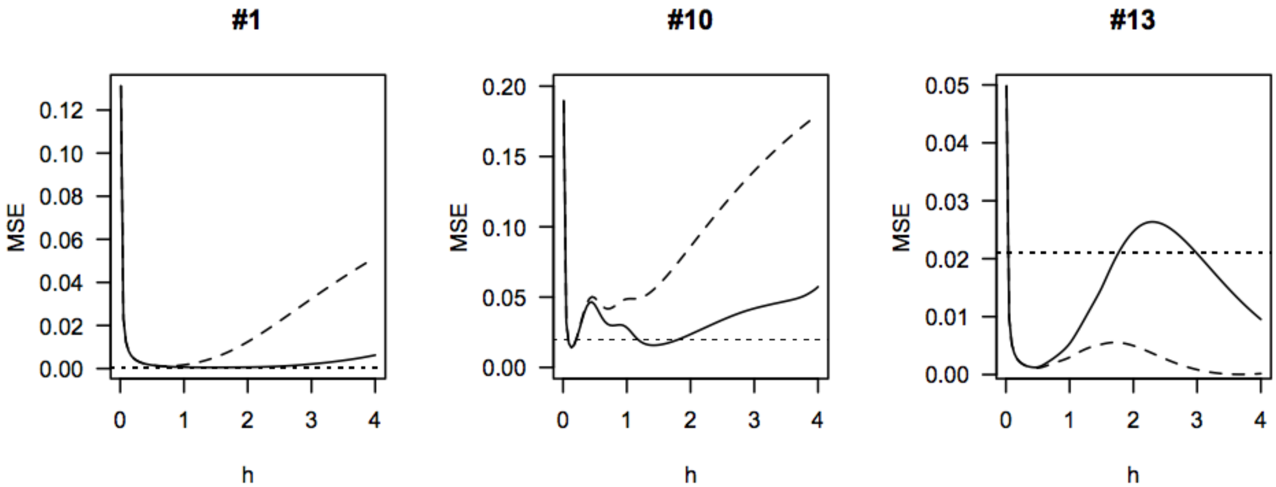

Figure 2: The MSE as a function of h for cases #1, #10 and #13 for the PGK estimator (solid curve), the TK estimator (dashed curve), and the parametric guide (dotted curve).

4.2 Weibull density with an exponential guide

In this model, the variable of interest T is generated from a Weibull distribution with a scale parameter $b = 4$ and a shape parameter taking three values $a = 1, 2, 4$. The graphs of the resulting densities are plotted in Figure 3. The censoring variable is also drawn from a Weibull distribution with shape parameter a and scale parameter given by $b((1-p)/p)^{1/a}$, ensuring a degree of censoring equal to p. We consider two censoring rates $p = 10\%$ and $p = 40\%$, and two sample sizes $n = 150$ and $n = 400$. As a parametric guide we use the exponential density $f_\theta(t) = \theta \exp(-\theta t)$, where θ is estimated using the approximated maximum likelihood estimator given by $\hat{\theta} = 1/\sum_{i=1}^{n} W_{(i)} X_{(i)}$, where $X_{(i)}$ are the ordered values of the observed variables $X_i = \min(T_i, C_i)$ and $W_{(i)}$ is the size of the jump of the Kaplan-Meier estimator at $X_{(i)}$ (see Suzukawa *et al.* (2001)). Note that the case $a = 1$ is the only situation where the guide is correctly specified. If $a \neq 1$ then the parametric guide is incorrect and deviates gradually from the true density. Our goal is to compare the performance of the PGK estimator with that of the TK estimator and the logspline estimator (LSP) (Kooperberg and Stone (1992)) for both the density and the hazard function. To this end, we run 1000 simulations and for every generated data set we calculate the estimators at the point $t = 3$. We look for the optimal bandwidths via a grid on $[0.1, 5]$. For the LSP estimator we select the optimal number of knots which minimizes the MSE over a set on $\{3, \ldots, 10\}$. The results are summarized in Tables 2 and 3 for the density and the hazard function, respectively.

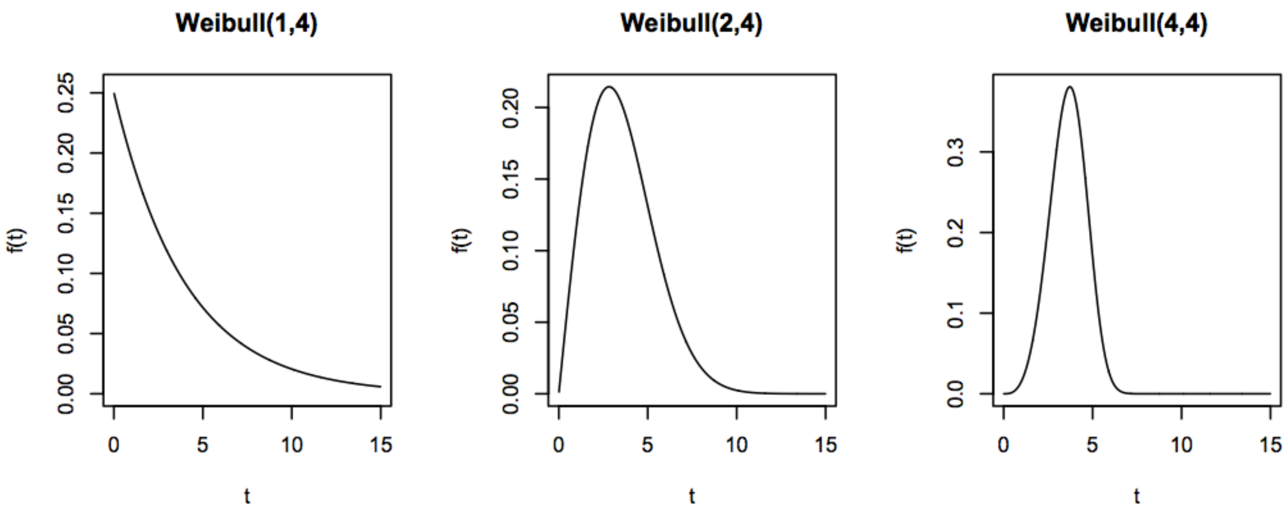

Figure 3: Weibull density with shape parameters $a = 1, 2, 4$ and scale parameter $b = 4$.

We start with the simulation results for the density estimators. As expected, with a correct parametric guide ($a = 1$) we get the best results for the PGK estimator. The bias of the PGK estimator is significantly reduced compared to that of the TK and the LSP estimator. Regarding the MSE, it is also reduced for the PGK estimator compared to the MSE of the LSP and the TK estimator, except for the case of sample size 150 and censoring rate 40%, where we observe a slightly larger variance and MSE for the PGK estimator compared to the TK estimator, but this is corrected with a larger sample size $n = 400$. For $a = 2$ and $a = 4$, even if the parametric guide is incorrect, the PGK estimator remains significantly better than the TK estimator, while the LSP estimator has a significantly smaller bias than both the PGK and the TK estimator. Regarding the variance, as expected, the PGK and the TK estimator have similar behavior except for the case $a = 4$ where the TK estimator has a larger variance. The LSP estimator has in general substantially larger variance and MSE, compared to the two kernel-based estimators, except for the case $a = 2$ with sample size 400, where the

LSP estimator outperforms both kernel-based estimators. In general, the MSE of the PGK estimator is not substantially reduced compared to that of the TK estimator, because in this example the variance dominates the bias.

For the hazard function, we computed the PGK, the TK and the LSP estimator using the same data generating procedure as for density function. The results are summarized in Table 3 and show that the PGK estimator generally outperforms the TK and the LSP estimator even if the parametric guide is not correctly specified. Note that for both the density and the hazard function, increasing the sample size enhances the performance of the PGK estimator. Another point to remark is that for the density and the hazard function, the selected optimal bandwidths for the PGK and the TK estimator are often close. In addition, we compared the PGK estimator based on the MLE and the PGK estimator based on the approximate MLE (AMLE). Simulations not given here show that when the guide is correct the PGK estimator based on the MLE outperforms the PGK estimator based on the AMLE. This is to be expected, since in this case the MLE is consistent, while the PGK estimator based on the AMLE behaves better when the guide is misspecified (see Remark 3.2 and Suzukawa *et al.* (2001)).

Finally, we compared the performance of the PGK and the TK density estimator at different time points. We computed both competitors at 100 different equally spaced time points from 0 to 7. We compared the performance for all cases given in Table 2, but, for sake of brevity, we only show three cases. Figure 4 gives the squared bias, the variance and the MSE as a function of t for the cases $a = 1, 2$ and 4, respectively. When the guide is correct ($a = 1$) the bias of the PGK estimator is reduced to zero for most values of t. This is not the case for the MSE because as said before, the variance dominates the bias. For the case $a = 2$, the bias and the MSE are reduced considerably for many values of t. In the last case $a = 4$, the PGK and the TK estimator behave similarly for most values of t.

Table 2: Squared bias ($\times 10^6$), Variance ($\times 10^6$), MSE ($\times 10^6$), the optimal bandwidth h and the optimal number of knots $nknot$ for the estimators of several Weibull densities for $a = (1, 2, 4)$, two censoring rates $p = (10\%, 40\%)$, two sample sizes $n = (150, 400)$ and $N = 1000$ replications.

p			10%				40%			
n	a	Method	Bias^2	Var	MSE	$h/nknot$	Bias^2	Var	MSE	$h/nknot$
150										
	1	PGK	0.009	114.6	114.6	5	0.579	194.2	194.8	5
		TK	23.60	99.70	123.3	5	26.60	153.7	180.3	5
		LSP	22.95	162.2	185.2	6	20.97	251.6	272.6	6
	2	PGK	99.60	251.1	350.7	3.367	123.7	334.2	457.9	3.713
		TK	114.5	274.1	388.6	3.020	152.9	345.7	498.6	3.268
		LSP	81.90	332.7	414.6	3	10.90	627.9	638.8	6
	4	PGK	98.91	258.1	357.0	3.169	94.40	377.0	471.4	3.268
		TK	246.6	514.4	761.0	2.228	303.7	629.2	932.9	2.327
		LSP	21.10	835.0	856.1	5	410.2	900.8	1311	4
n	a	Method	Bias^2	Var	MSE	$h/nknot$	Bias^2	Var	MSE	$h/nknot$
400										
	1	PGK	0.020	40.58	40.60	5	0.01	68.66	68.67	5
		TK	20.63	39.89	60.52	4.802	22.35	58.59	80.94	5
		LSP	3.36	55.81	59.17	6	8.340	85.10	93.44	6
	2	PGK	50.70	124.8	175.5	2.872	62.80	177.7	240.5	3.020
		TK	54.20	146.0	200.2	2.525	75.00	189.3	264.3	2.723
		LSP	0.450	150.4	150.9	5	0.778	207.5	208.3	6
	4	PGK	44.20	136.3	180.5	2.822	51.80	187.6	239.4	2.921
		TK	117.5	293.8	411.3	1.882	141.8	330.5	472.3	1.981
		LSP	20.90	319.1	340.0	5	0.523	520.2	520.7	6

Table 3: Squared bias ($\times 10^5$), Variance ($\times 10^5$), MSE ($\times 10^5$), the optimal bandwidth h and the optimal number of knots *nknot* for the estimators of several Weibull hazards for $a = (1, 2, 4)$, two censoring rates $p = (10\%, 40\%)$, two sample sizes $n = (150, 400)$ and $N = 1000$ replications.

p			10%				40%			
n	a	Method	Bias^2	Var	MSE	$h/nknot$	Bias^2	Var	MSE	$h/nknot$
150										
	1	PGK	0.360	90.77	91.13	5	1.300	145.0	146.3	5
		TK	14.69	113.0	127.7	4.901	18.60	168.6	187.2	5
		LSP	20.91	73.22	94.13	5	5.430	189.5	194.9	6
	2	PGK	37.10	128.0	165.1	3.664	43.90	167.1	211.0	4.010
		TK	35.40	166.4	201.8	3.168	52.40	206.9	259.3	3.515
		LSP	0.140	179.2	179.3	5	1.105	325.3	326.4	6
	4	PGK	24.20	91.4	115.6	3.366	24.70	119.6	144.2	3.465
		TK	48.00	169.2	218.2	2.327	71.90	189.4	261.3	2.525
		LSP	6.343	258.4	264.7	5	4.548	411.3	415.8	5
n	a	Method	Bias^2	Var	MSE	$h/nknot$	Bias^2	Var	MSE	$h/nknot$
400										
	1	PGK	0.190	34.91	35.10	5	0.220	57.13	57.35	5
		TK	8.200	47.40	55.60	4.307	13.30	68.50	81.80	4.901
		LSP	0.810	41.42	42.23	6	1.860	75.25	77.11	6
	2	PGK	17.20	63.50	80.70	3.020	21.80	85.10	106.9	3.218
		TK	15.80	80.90	96.70	2.574	22.90	102.3	125.1	2.822
		LSP	0.340	66.92	67.26	5	0.870	120.4	121.3	6
	4	PGK	10.36	46.39	56.75	2.971	12.06	56.66	68.72	3.070
		TK	24.50	87.10	111.6	1.981	26.40	98.70	125.1	2.030
		LSP	5.300	101.8	107.1	5	0.800	153.2	154.0	5

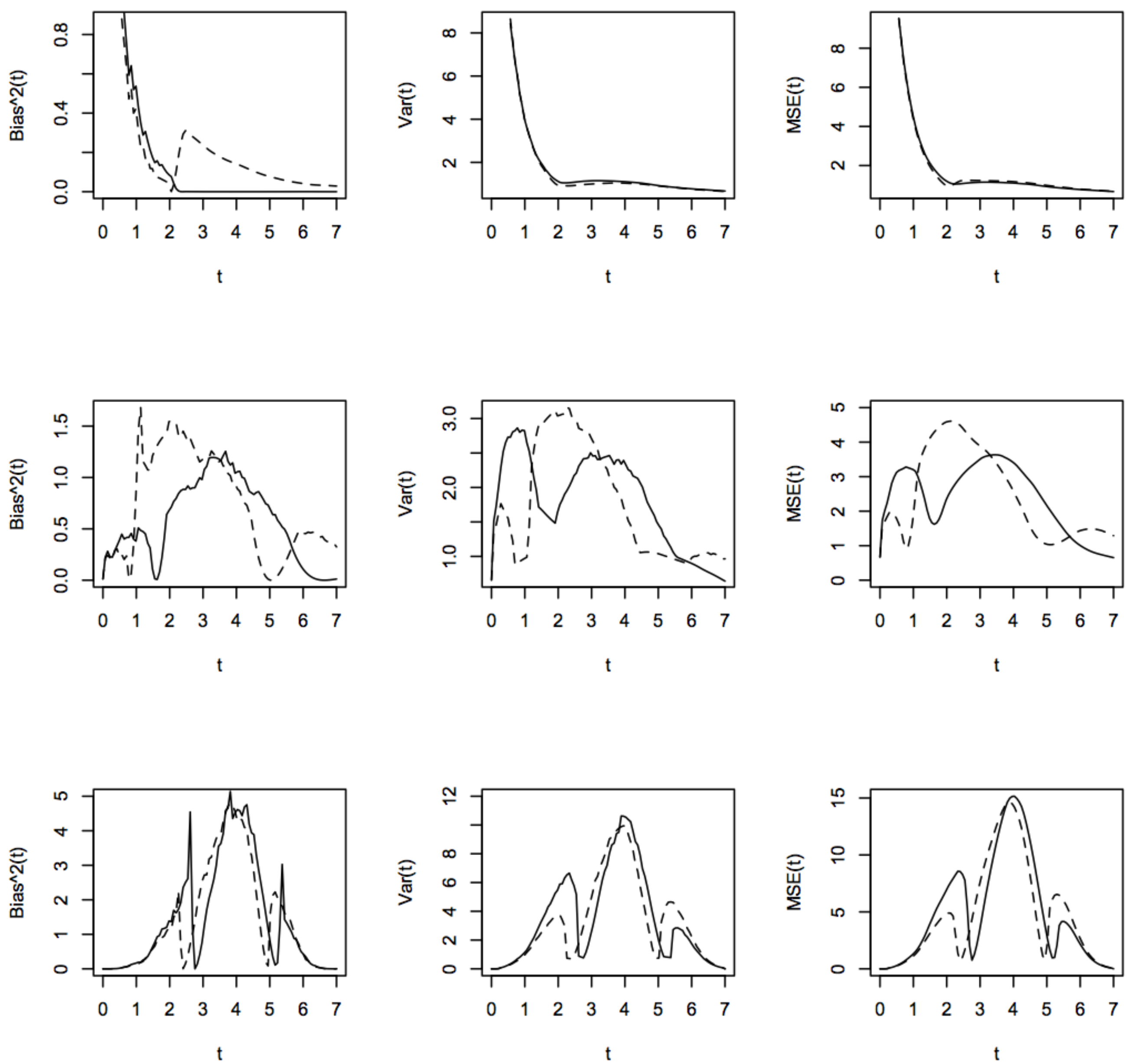

Figure 4: Squared bias, variance and MSE ($\times 10^4$) as a function of t, for the PGK estimator (solid curve) and the TK estimator (dotted curve). The figures correspond to $a = 1$ (top), $a = 2$ (middle) and $a = 4$ (bottom). The sample size is $n = 150$ and the proportion of censoring is $p = 0.10$.

5 Application

To illustrate our method with a real dataset, we consider the UIS dataset from the University of Massachusetts Aids Research Unit (UMARU) IMPACT Study. The goal of this study is to model time until return to drug use for patients enrolled in two different residential treatment programs that differed in length (treat=0 is the short program and treat=1 is the long program). Among a total of 628 observations, there are 120 censored observations, which corresponds to a censoring rate of 19.12%. The data as well as a detailed description of the study can be found in Section 1.3 of Hosmer *et al.* (2008).

As a first step, before proceeding to more complicated analyses, it is always useful to have an idea about the distribution of the variable of interest without considering explanatory variables. This basic univariate analysis is given in Section 5.1 and is in general informative but not sufficient in practice. A more complete analysis including explanatory variables is performed in a second step as an extension of the PGK method to the Cox model, see Section 5.3 for more details.

We also explain in Section 5.2 how to obtain confidence intervals for the density and hazard function.

5.1 Density and hazard estimation for the UIS data

In this section, we use the PGK, the TK and the logspline (LSP) estimator (Kooperberg and Stone (1992)) for the analysis of the time to return to drug use in the UIS dataset. We consider the PGK estimator using two different guides, an exponential density and a Weibull density for which the parameters are estimated using maximum likelihood. For the kernel-based estimators, we use the Epanechnikov kernel and the choice of the bandwidths is achieved by a data driven bandwidth selection based on the least squares cross-validation method adapted to each estimator (see for example Wang and Wang (2007)). The selected bandwidths are: $h = 671.21$ for the PGK estimator with an exponential guide, $h = 716.66$ for the PGK estimator with a Weibull guide and $h = 663.63$ for the TK estimator. For the LSP estimator, we specify that the density equals zero to the left of 0, we use a BIC penalty and a stepwise knot deletion procedure (see Kooperberg and Stone (1991)), which selected four knots of the original 10 knots.

The plots of the different estimators of the density and the hazard function are given in Figure 5. From the plot of the hazard function it can be seen that the risk to return to drug use increases slowly during the first 200-300 days, after which it decreases to zero. While the overall shape is the same for all four estimators, there are some absolute differences especially during the first 400 days. We do not know which estimator is closer to the true curve, but given the results of our simulation study, we suspect that the TK method has tendency to either overestimate or underestimate the hazard function and so the real risks to return to drug use. Finally, concerning the density function we see that the PGK and LSP estimators are quite close. Moreover, it seems that our PGK estimators remove a considerable part of the boundary effect on the left endpoint of the distribution compared to the TK estimator.

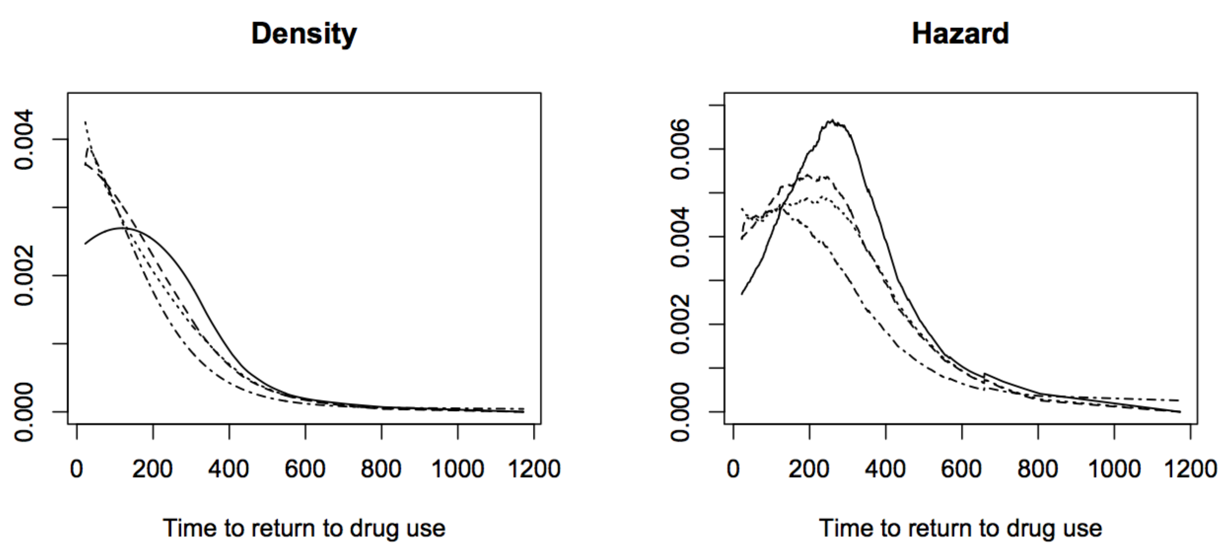

Figure 5: Estimation of the density and hazard function for the UIS data by means of the TK estimator (solid curve), the LSP estimator (dot-dashed curve), the PGK estimator with exponential guide (dashed curve), and the PGK estimator with Weibull guide (dotted curve).

5.2 Asymptotic confidence intervals

Based on the asymptotic normality of our PGK estimators, we construct confidence intervals for the density and the hazard function of the time to return to drug use for the UIS data. In order to remove the asymptotic bias, we use the undersmoothing technique; see for example Horowitz (2001) and Fiorio (2004). The pointwise $100(1-\alpha)\%$ confidence interval for the density function $f(t)$ and the hazard function $\lambda(t)$ are given respectively by

$$\widehat{f}_{\widehat{\theta}}(t) \pm z_{1-\alpha/2}\widehat{\sigma}(t)/\sqrt{nh}, \quad \widehat{\lambda}_{\widehat{\theta}}(t) \pm z_{1-\alpha/2}\widehat{\tau}(t)/\sqrt{nh},$$

where $\widehat{\sigma}^2(t) = [\widehat{f}_{\widehat{\theta}}(t)/(1-\widehat{G}(t))]\int K^2(u)du$, $\widehat{\tau}^2(t) = [\widehat{\lambda}_{\widehat{\theta}}(t)/(1-\widehat{H}(t))]\int K^2(u)du$, and $h = h_{opt}n^{1/5}/n^{\tau} = 486.36$. Here, $\tau = 1/4$ is the undersmoothing parameter, h_{opt} is the optimal bandwidth selected via the cross-validation method used in the previous section, $\widehat{G}(\cdot)$ is the Kaplan Meier estimator of $G(\cdot)$ and $\widehat{H}(\cdot)$ is the empirical distribution function of $H(\cdot)$. The undersmoothed PGK density and hazard estimator with exponential guide and their respective confidence intervals are plotted in Figure 6. After the 500th day, the intervals become larger because most censoring occurs after this date.

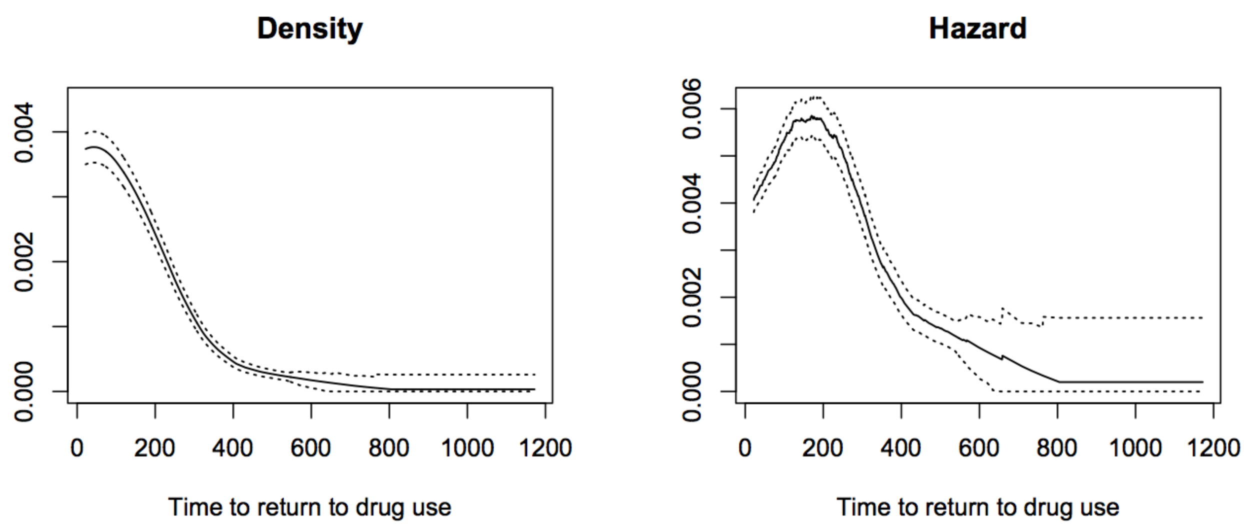

Figure 6: Estimation of the density and hazard function for the UIS data by means of the undersmoothed PGK estimator with exponential guide (solid curve), together with 95% pointwise asymptotic confidence intervals (dotted curve).

5.3 Extending the PGK estimation to the Cox model

In this section, we extend the PGK method to the estimation of the baseline and the conditional density and hazard functions in the Cox model. First, the PGK estimator that we propose for the baseline hazard function is $\widehat{\lambda}_{\widehat{\theta},B}(t) = \widehat{f}_{\widehat{\theta},B}(t)/(1 - \widehat{F}_B(t))$, where $\widehat{F}_B(\cdot)$ is the baseline distribution function estimated after fitting a semiparametric Cox model and $\widehat{f}_{\widehat{\theta},B}(t)$ is the PGK estimator of the baseline density defined as

$$\widehat{f}_{\widehat{\theta},B}(t) = \frac{1}{h}\sum_{i=1}^{n} K\Big(\frac{t - X_i}{h}\Big)\frac{f_{\widehat{\theta},B}(t)}{f_{\widehat{\theta},B}(X_i)} W_{Bi},$$

where $f_{\widehat{\theta},B}(\cdot)$ is a parametric baseline guide estimated via a parametric proportional hazards model (PPH) and W_{Bi} is the jump size of $\widehat{F}_B(\cdot)$ at X_i.

Now, to extend the PGK estimator to the conditional survival and hazard function, we propose the following estimators:

$$\widehat{\lambda}_{\widehat{\theta}}(t \mid Z) = \widehat{\lambda}_{\widehat{\theta},B}(t)\exp(Z^t\widehat{\beta}), \quad \widehat{S}_{\widehat{\theta}}(t \mid Z) = \exp\Big(-\int_0^t \widehat{\lambda}_{\widehat{\theta}}(s \mid Z)ds\Big),$$

where Z is a vector of covariates and $\widehat{\beta}$ are the estimated parameters of the Cox model using the partial likelihood maximization.

We apply the proposed estimators to the UIS data. In addition to the time to return to drug use, we now consider $Z = (Z_1, Z_2, Z_3, Z_4)$ a vector of four covariates: Z_1 is the age in years, Z_2 is the drug use history (1=never, 2=previous, 3=recent), Z_3 is the number of prior drug treatments at admission, and Z_4 is the treatment randomization assignment (0=short, 1=long). As parametric guide, we use a Weibull baseline density estimated after fitting a PPH model. The bandwidth is selected as before using the cross-validation method. The selected bandwidth is $h_{cv} = 590$. Figure 7 shows the plots of the PGK baseline density and hazard estimators. We see that, except for the beginning of the study, the estimators are monotonically decreasing, meaning that at the start of the study the participants are at high risk to restart using drugs. The estimated parameters for the fitted Cox model are $\widehat{\beta} = (-0.030, 0.208, 0.029, -0.235)$. Based on this model, we conclude that increasing the participant age or the treatment period reduces the risk of returning to drug use, while increasing the drug use history or the number of previous treatments increases the risk of returning to drug use. Since the parametric guide has no effect on the asymptotic distribution of $\hat{\beta}$, asymptotic confidence intervals can be computed as in the classical Cox model. We investigated pointwise confidence intervals for every parameter. Note that all parameters are significant but for sake of brevity we only give the 95% pointwise confidence interval for the parameter β_4, which is $[-0.42, -0.05]$ and shows a significant effect of the treatment randomization assignment on the risk to return to drug use. Figure 8 shows the plots of the PGK conditional survival and hazard function estimators for $Z = (36, 2, 1, 1)$.

It appears from the plot of the conditional survival function that for a participant aged 36 years, receiving a long treatment, with a previous history of drug use and one prior treatment, the median time to return to drug use is about 200 days.

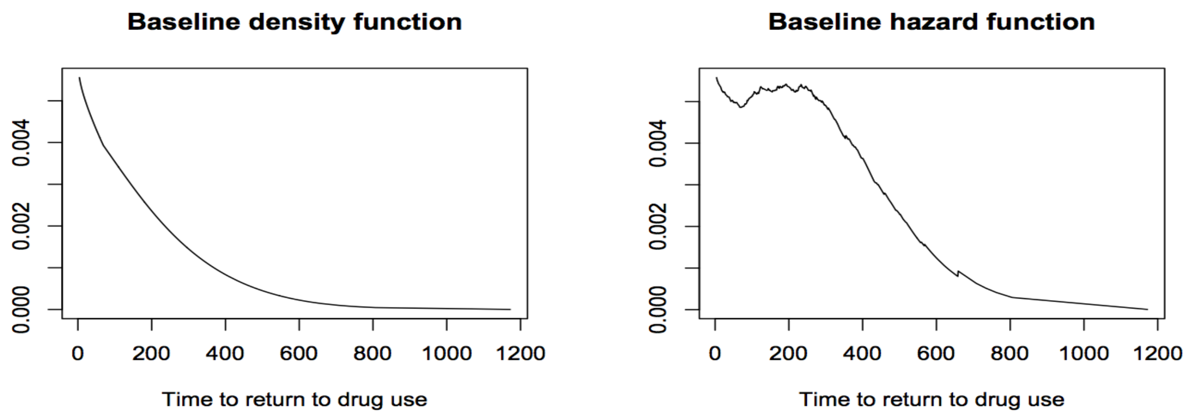

Figure 7: Estimation of the baseline density and hazard function in the Cox model for the UIS data, by means of the PGK estimator with Weibull guide.

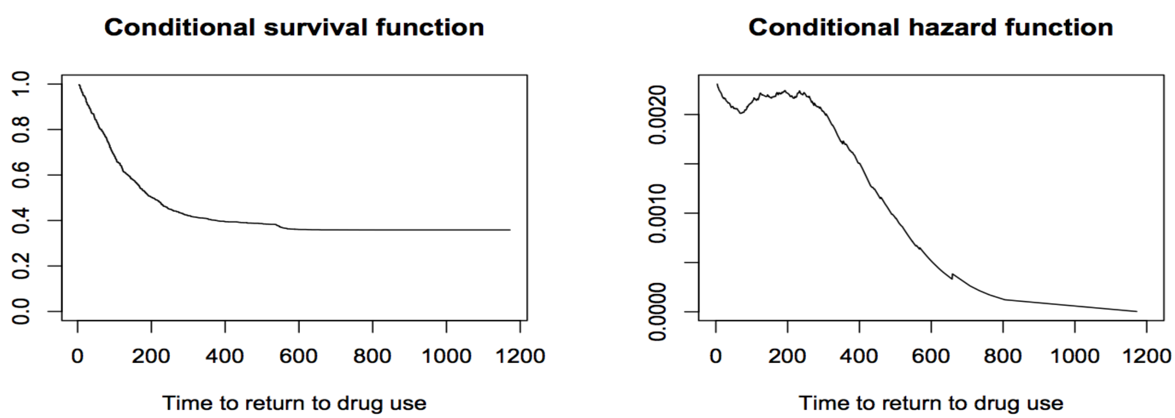

Figure 8: Estimation of the conditional survival and hazard function in the Cox model for the UIS data, by means of the PGK estimator with Weibull guide. The covariate equals $Z = (36, 2, 1, 1)$.

6 Conclusion

In this paper, we extended the parametrically guided kernel density and hazard estimators to the censored data framework. The proposed estimators are obtained by multiplying an initial parametric estimator by a nonparametric kernel type estimator of a certain correction function. We established the asymptotic normality of the proposed estimators and obtained asymptotic expressions of the bias and variance. Under certain regularity conditions, we proved that the bias of the proposed estimator can be reduced compared to that of the traditional kernel estimator, while the variance does not change. Simulations confirmed the theoretical results and provide the following remarks for the density and the hazard functions: the PGK estimator with censored data outperforms the TK estimator if the parametric guide is equal or close to the true target function and performs as the TK estimator if the parametric guide is misspecified. The comparison to the logspline estimator shows that the PGK estimator is generally better in terms of the MSE. The application of the PGK estimator to the UIS dataset reveals that, in addition to bias reduction, the estimator also seems to correct in an automatic way for possible boundary effects. Moreover, confidence intervals and an extension to the Cox model are developed and applied to the UIS data.

Finally, as pointed out by Hjort and Glad (1995) in the uncensored case, the advantages of the multiplicative PGK method come with some drawback caused by the correction factor $f_{\widehat{\theta}}(t)/f_{\widehat{\theta}}(X_{(i)})$, see equation (2.4). Small values of $f_{\widehat{\theta}}(X_{(i)})$ may affect the numerical stability of the resulting estimator especially with a "large" bandwidth and this affects also the MSE. One may correct for this by adding a small ϵ to both the numerator and the denominator or by adopting an additive parametric correction (instead of a multiplicative one). This method is under investigation and will be the subject of a future publication.

7 Appendix

Proof of Theorem 3.1. The PGK density estimator based on the non random parametric guide $f_*(\cdot)$ can be decomposed as follows:

$$\widehat{f}_*(t) - f(t) = (\widehat{f}_*(t) - \widetilde{f}(t)) + (\widetilde{f}(t) - f(t)), \tag{7.8}$$

where

$$\widetilde{f}(t) = \frac{1}{h}\int_{-\infty}^{+\infty} K\left(\frac{t-s}{h}\right)\frac{f_*(t)}{f_*(s)}dF(s).$$

1. For $t \le \tau$, we have

$$\begin{aligned}
\widehat{f}_*(t) - \widetilde{f}(t) &= \frac{f_*(t)}{h}\int_{t-h}^{t+h} K\left(\frac{t-s}{h}\right)\frac{1}{f_*(s)}d(\widehat{F}(s)-F(s)) \\
&= \frac{f_*(t)}{h}\int_{-1}^{1}(\widehat{F}(t-uh)-F(t-uh))d\big(K(u)/f_*(t-uh)\big) \\
&= \frac{f_*(t)}{h}\int_{-1}^{1}(\widehat{F}(t-uh)-F(t-uh))\frac{K'(u)}{f_*(t-uh)}du \\
&\quad + f_*(t)\int_{-1}^{1}(\widehat{F}(t-uh)-F(t-uh))\frac{K(u)f_*'(t-uh)}{f_*^2(t-uh)}du \\
&= A_{1,n} + A_{2,n}.
\end{aligned}$$

First, write

$$\begin{aligned}
A_{1,n} &= \frac{1}{h}\int_{-1}^{1}(\widehat{F}(t-uh)-F(t-uh))K'(u)du \\
&\quad + \frac{f_*(t)}{h}\int_{-1}^{1}(\widehat{F}(t-uh)-F(t-uh))\left(\frac{K'(u)}{f_*(t-uh)} - \frac{K'(u)}{f_*(t)}\right)du \\
&= A_{11,n} + A_{12,n}.
\end{aligned}$$

We start with $A_{12,n}$. We have

$$\begin{aligned}
A_{12,n} &= \frac{1}{h}\int_{-1}^{1}(\widehat{F}(t-uh)-F(t-uh))\frac{f_*(t)-f_*(t-uh)}{f_*(t-uh)}K'(u)du \\
&= \frac{1}{h}\int_{-1}^{1}(\widehat{F}(t-uh)-F(t-uh))\frac{f_*'(t+\rho)uh}{f_*(t-uh)}K'(u)du,
\end{aligned}$$

for some ρ between 0 and $-uh$. Therefore,

$$|A_{12,n}| \le \sup_{s\in\aleph_t}|\widehat{F}(s)-F(s)|\sup_{s\in\aleph_t}|f_*'(s)|(\inf_{s\in\aleph_t} f_*(s))^{-1}\int_{-1}^{1}|K'(u)||u|du,$$

where $\aleph_t$ is a small neighborhood around t. Hence, under assumptions 3.1 and 3.2, and using the uniform rate of the Kaplan-Meier estimator (see e.g. Theorem 1 in Lo and Singh (1986)) we have that $A_{12,n} = O_p(n^{-1/2})$. Now, we treat the term $A_{11,n}$. We consider the i.i.d. decomposition of $\widehat{F}$ given in Lemma 2.1 in Lo *et al.* (1989):

$$\widehat{F}(s) - F(s) = n^{-1}\sum_{i=1}^{n}\xi_i(s) + r_n(s),$$

where $\xi_i(s) = -\int_{-\infty}^{X_i \wedge s} \big(1 - H(x)\big)^{-2} dH_1(x) + \big(1 - H(X_i)\big)^{-1} I\{X_i \leq s, \delta_i = 1\}$ and $\sup_{s \in \aleph_t} |r_n(s)| = O_p(n^{-1} \log n)$. Note that Lo *et al.* (1989) assume that the lifetimes are non-negative, whereas we work with random variables defined on the whole real line. However, it can be easily seen that their resuls remain valid in our setup. Then,

$$A_{11,n} = \frac{1}{nh} \sum_{i=1}^{n} U_{in}(t) + O_p((nh)^{-1} \log n),$$

where $U_{in}(t) = \int_{-1}^{1} \xi_i(t - uh) K'(u) du$. Therefore,

$$A_{1,n} = \frac{1}{nh} \sum_{i=1}^{n} U_{in}(t) + O_p(n^{-1/2}), \tag{7.9}$$

thanks to assumption (A.2).

Finally, we consider the term $A_{2,n}$. Under assumptions 3.1 and 3.2, we have

$$\begin{aligned} |A_{2,n}| &\leq f_*(t) \sup_{s \in \aleph_t} |\widehat{F}(s) - F(s)| \sup_{s \in \aleph_t} |f_*'(s)| (\inf_{s \in \aleph_t} f_*(s))^{-2} \\ &= O_p(n^{-1/2}). \end{aligned}$$

Therefore,

$$\widehat{f}_*(t) - \tilde{f}(t) = \frac{1}{nh} \sum_{i} U_{in}(t) + O_p(n^{-1/2}). \tag{7.10}$$

The result now follows from expressions (7.8) and (7.10).

2. We have

$$\begin{aligned} \widetilde{f}(t) - f(t) &= \frac{f_*(t)}{h} \int_{t-h}^{t+h} K\left(\frac{t-s}{h}\right) \frac{f(s)}{f_*(s)} ds - f(t) \\ &= f_*(t) \int_{-1}^{1} K(u) r_*(t - uh) du - f(t) \\ &= f_*(t) \int_{-1}^{1} K(u) \big(r_*(t) - r_*'(t) uh + \frac{1}{2} h^2 r_*''(t) u^2 + o(h^2)\big) du - f(t) \\ &= \frac{1}{2} h^2 r_*''(t) f_*(t) \mu_K^2 + o(h^2). \end{aligned}$$

Now, the result is an immediate consequence of the first point, and of Theorem 3.2 and Corollary 3.3 in Lo *et al.* (1989). □

Proof of Theorem 3.2. Write

$$(nh)^{1/2}(\widehat{f}_{\widehat{\theta}}(t) - f(t)) = (nh)^{1/2}(\widehat{f}_{\widehat{\theta}}(t) - \widehat{f}_{\theta *}(t)) + (nh)^{1/2}(\widehat{f}_{\theta^*}(t) - f(t)),$$

where $\widehat{f}_{\theta^*}(t)$ is the PGK density estimator based on the parametric guide $f_{\theta^*}(\cdot)$. From the second point in Theorem 3.1 it follows that

$$\sqrt{nh}\Big(\widehat{f}_{\theta^*}(t) - f(t) - \frac{1}{2}h^2\mu_K^2 r''_{\theta^*}(t) f_{\theta^*}(t) + o(h^2)\Big) \xrightarrow{d} \mathcal{N}\Big(0, \sigma^2(t)\Big). \tag{7.11}$$

On the other hand, we have

$$\begin{aligned}
\widehat{f}_{\widehat{\theta}}(t) - \widehat{f}_{\theta^*}(t) &= \frac{1}{h}(f_{\widehat{\theta}}(t) - f_{\theta^*}(t)) \int_{t-h}^{t+h} K\Big(\frac{t-s}{h}\Big)\frac{1}{f_{\widehat{\theta}}(s)} d\widehat{F}(s) \\
&\quad - \frac{f_{\theta^*}(t)}{h}\int_{t-h}^{t+h} K\Big(\frac{t-s}{h}\Big)\frac{f_{\widehat{\theta}}(s) - f_{\theta^*}(s)}{f_{\widehat{\theta}}(s) f_{\theta^*}(s)} d\widehat{F}(s) \\
&= I_{1n} - I_{2n}.
\end{aligned}$$

By a Taylor expression we have, for an intermediate point θ_m between $\widehat{\theta}$ and θ^* and a constant $C < \infty$,

$$\begin{aligned}
|I_{1n}| &\leq \frac{C}{h}\big\|\nabla_\theta f_{\theta_m}(t)\big\|\, |\widehat{\theta} - \theta^*|\, |\widehat{F}(t+h) - \widehat{F}(t-h)| \\
&\leq \frac{C}{h}\big\|\nabla_\theta f_{\theta_m}(t)\big\|\, |\widehat{\theta} - \theta^*|\, |\widehat{F}(t+h) - \widehat{F}(t-h) - F(t+h) + F(t-h)| \\
&\quad + \frac{C}{h}\big\|\nabla_\theta f_{\theta_m}(t)\big\|\, |\widehat{\theta} - \theta^*|\, |F(t+h) - F(t-h)|,
\end{aligned}$$

where $\nabla_\theta f_\theta(t) = (\partial f_\theta(t)/\partial\theta_j)_{j=1}^p$. From Lemma 3 in Gijbels and Veraverbeke (1989), we have

$$\widehat{F}(t+h) - \widehat{F}(t-h) - F(t+h) + F(t-h) \;=\; O_p(n^{-1/2}h^{1/2}(\log n)^{1/2}).$$

Hence, $I_{1n} = o_p((nh)^{-1/2})$. In similar way it can be shown that $I_{2n} = o_p((nh)^{-1/2})$, and so

$$\widehat{f}_{\widehat{\theta}}(t) - \widehat{f}_{\theta^*}(t) = o_p((nh)^{-1/2}). \tag{7.12}$$

The result of Theorem 3.2 now follows from equations (7.11) and (7.12). $\square$

Proof of Theorem 3.3. We have

$$(nh)^{1/2}(\widehat{\lambda}_{\widehat{\theta}}(t) - \lambda(t))) = (nh)^{1/2}\widehat{f}_{\widehat{\theta}}(t)\Bigg[\frac{\widehat{F}(t) - F(t)}{(1-\widehat{F}(t))(1-F(t))}\Bigg] + (nh)^{1/2}\Bigg[\frac{\widehat{f}_{\widehat{\theta}}(t) - f(t)}{1-F(t)}\Bigg].$$

Since the first term on the right hand side converges to zero in probability, the result of Theorem 3.3 is a direct consequence of Theorem 3.2. □

Acknowledgements

We thank the associate editor and the referees for their valuable comments. This research was supported by IAP research network grant nr. P07/06 of the Belgian government (Belgian Science Policy), and by the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'. The research of the second author was also supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650.

References

Blum, J.R. and Susarla, V. (1980). Maximal deviation theory of density and failure rate estimates based on censored data. *J. Multiv. Anal.* **5**, 213-222.

Buckland, S.T. (1992). Maximum likelihood fitting of Hermite and simple polynomial densities. *J. Roy. Statist. Soc. Ser. C* **41**, 241-266.

Castro-Kuriss, C. (2011). On a goodness-of-fit test for censored data from a location-scale distribution with applications. *Chil. J. Stat.* **2**, 115-136.

Copas, J.B. (1995). Local likelihood based on kernel censoring. *J. Roy. Statist. Soc. Ser. B* **57**, 221-235.

Efron, B. and Tibshirani, R. (1996). Using specially designed exponential families for density estimation. *J. Amer. Statist. Assoc.* **24**, 2431-2461.

Eguchi, S. and Copas, J.B. (1998). A class of local likelihood methods and near-parametric asymptotics. *J. Roy. Statist. Soc. Ser. B* **60**, 709-724.

Fan, J., Wu, Y. and Feng, Y. (2009). Local quasi-likelihood with a parametric guide. *Ann. Statist.* **37**, 4153-4183.

Faraway, J. (1989). Implementing semiparametric density estimation. *Statist. Probab. Lett.* **10**, 141-143.

Földes, A., Rejtö, L. and Winter, B.B. (1981). Strong consistency properties of nonparametric estimators for randomly censored data. II: Estimation of density and failure rate. *Period. Math. Hung.* **12**, 15-29.

Friedman, J.H., Stuetzle, W. and Schroeder, A. (1984). Projection pursuit density estimation. *J. Amer. Statist. Assoc.* **79**, 599-608.

Fiorio, C. V. (2004). Confidence intervals for kernel density estimation. *Stata J.* **4(2)**, 168-179.

Gefeller, O. and Dette, H. (1992). Nearest neighbor kernel estimation of the hazard function from censored data. *J. Statist. Comput. Simul.,* **43**, 93-101.

Gefeller, O. and Michels, P. (1992). A review on smoothing methods for the estimation of the hazard rate based on kernel functions, in Dodge, Y. and Whittaker, J. (eds). *Computational Statistics*, Physica-Verlag, Switzerland, 459-464.

Gijbels, I. and Veraverbeke, N. (1989). Sequential fixed width confidence intervals for quantiles in the presence of censoring. *J. Statist. Plan. Infer.* **22**, 213-222.

Giné, E. and Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data : rates holding uniformly over adaptive intervals. *Ann. Inst. H. Poincaré Probab. Statist.* **37**, 503-522.

Glad, I. (1998), Parametrically guided non-parametric regression, *Scand. J. Stat.* **25**, 649-668.

González-Manteiga, W., Cao, R. and Marron, J.S. (1996). Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation. *J. Amer. Statist. Assoc.* **91**, 1130-1140.

Hjort, N.L. (1986). *Theory of Statistical Symbol Recognition.* Norwegian Computing Centre, Oslo.

Hjort, N.L. and Glad, I.K. (1994). *Nonparametric density estimation with parametric start.* Statistical research report, Dept. Mathematics, Univ. Oslo.

Hjort, N.L. and Glad, I.K. (1995). Nonparametric density estimation with parametric start. *Ann. Statist.* **23**, 882-904.

Hjort, N.L. and Jones, M.C. (1996). Locally parametric nonparametric density estimation. *Ann. Statist.* **24**, 1619-1647.

Horowitz, J. L. (2001). *The bootstrap.* In Handbook of Econometrics, ed. J. J. Heckeman and E. Leamer. **5**, 3159-3228. Amsterdam: North-Holland.

Hosmer, D.W., Lemeshow, S. and May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time to Event Data: Second Edition*, John Wiley and Sons Inc., New York.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

Klein, J.P. and Moeschberger, M.L. (1997). *Survival Analysis. Techniques for Censored and*

Truncated Data. Springer, New York.

Kooperberg, C. and Stone, C. J. (1991). A study of logspline density estimation. *Comput. Stat. Data Anal.*, **12**, 327-347.

Kooperberg, C. and Stone, C. J. (1992). Logspline density estimation for censored data. *J. Comp. Graph. Stat.*, **1**, 301-328.

Lo, S.H., Mack, Y.P. and Wang, J.L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Probab. Theory Relat. Fields* **80**, 461-473.

Lo, S.H. and Singh, K. (1986). The product-limit estimator and the bootstrap: Some asymptotic representation. *Probab. Theory Relat. Fields,* **71**, 455-465.

Marron, J.S. and Wand, M.P. (1992). Exact mean integrated squared error. *Ann. Statist.* **20**, 712-736.

Martins-Filho, C., Mishra, S. and Ullah, A. (2008). A class of improved parametrically guided nonparametric regression estimators. *Econometric Reviews* **27**, 542-573.

Mielniczuk, J. (1986). Some asymptotic properties of kernel estimators of a density function in the case of censored data. *Ann. Statist.* **14**, 766-773.

Müller, H.G. and Wang, J.L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* **50**, 61-76.

Naito, K. (2004). Density estimation by local L2-fitting. *Ann. Statist.* **32**, 1162-1191.

Oakes, D. (1986). An approximate likelihood procedure for censored data. *Biometrics* **42**, 177-182.

Olkin, I. and Spiegelman, C.H. (1987). A semiparametric approach to density estimation. *J. Amer. Statist. Assoc.* **82**, 858-865.

Padgett, W.J. and McNichols, D.T. (1984). Nonparametric density estimation from censored data. *Commun. Statist.-Theory Meth.* **13**, 1581-1611.

Patil, P.N. (1993). Bandwidth choice for nonparametric hazard rate estimation. *J. Statist. Plann. Inf.* **35**, 15-30.

Sánchez-Sellero, C., González-Manteiga, W. and Cao, R. (1999). Bandwidth selection in density estimation with truncated and censored data. *Ann. Inst. Statist. Math.* **51**, 51-70.

Suzukawa, A., Imai, H. and Sato, Y. (2001). Kullback-Leibler information consistent estimation for censored data. *Ann. Inst. Statist. Math.* **53**, 262-276.

Talamakrouni, M., El Ghouch, A. and Van Keilegom, I. (2014). Guided censored regression. *Scand. J. Stat.* (to appear).

Tanner, M. and Wong, W.H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Statist.* **11**, 989-993.

Veraverbeke, N., Gijbels, I. and Omelka, M. (2014). Preadjusted non-parametric estimation of a conditional distribution function. *J. Roy. Statist. Soc. Ser. B* **76**, 399-438.

Wang, B. and Wang, X-F. (2007). *Bandwidth selection for weighted kernel density estimation* (unpublished manuscript).

Xiang, X.J. (1994). Law of the logarithm for density and hazard rate estimation for censored data, *J. Multiv. Anal.* **49**, 278-286.