Statistica Sinica **28** (2018), 2811-2822
doi:https://doi.org/10.5705/ss.202017.0057

EMPIRICAL FOURIER METHODS FOR INTERVAL CENSORED DATA

Peter G. Hall, W. John Braun and Thierry Duchesne

University of Melbourne, University of British Columbia and Université Laval

Abstract: Methods for estimating the probability density function are considered under the circumstance that the underlying measurements are interval-censored. Density and distribution function estimators are proposed under parametric and nonparametric assumptions on the censoring mechanism. Conditions for identifiability and consistency of the estimates are established theoretically, and it is shown that under such conditions, the estimates converge to the truth at a polynomial rate in the inverse sample size. An online supplement contains the technical arguments as well as practical guidelines for numerical implementation of the proposed methods. The core of the theory in this paper was originally drafted by Peter Hall in early 2010, following discussions at a workshop on mismeasured data held in Canada in December, 2009 at which Peter was the keynote speaker. The co-authors are grateful for the follow-up conversations held with Peter by long distance over the years prior to his regretful passing.

Key words and phrases: Characteristic functions, density estimation, kernel methods.

Preamble (by John Braun and Thierry Duchesne)

On December 10, 2009, Peter Hall arrived in Southern Ontario and expressed delight in seeing snow for the first time in several decades. It was an auspicious start to a three-day Fields Institute Workshop on indirectly or imprecisely observed data at which Peter was the principal keynote speaker. At the workshop, Peter shared the latest developments on Fourier deconvolution approaches to measurement error problems, and in discussions during and after the workshop, he became quite interested in how these approaches might be adapted to problems where the data were interval-censored. These discussions were immensely appreciated by all involved. By the end of the workshop, Peter indicated in his polite, but clear way, that he had quite seen enough of snow (it had been falling almost continuously for the entire event), and he happily boarded an airplane bound for Hong Kong.

Upon arrival in Hong Kong, Peter sent an email message to one of us (JB) containing an attachment of a carefully typed 6-page draft manuscript encapsulating some of the ideas discussed at the meeting. Much of that draft appears verbatim in Sections 2 and 3 of the present paper. Over the next few weeks, there were several emails back and forth concerning implementation of the proposed Fourier approach, and by January 8, 2010, Peter had completed most of the theory outlined in Sections 4 and 5.

Numerical issues continued to cause trouble over the ensuing months, seemingly contradicting the theoretical results concerning the consistency of the new estimator. The method seemed to require enormous sample sizes in order to work, so it did not appear to be a practical contribution to the literature on interval-censored density estimation. The work was abandoned, until TD was approached with questions about the consistency of a competing estimator (Braun, Duchesne and Stafford (2005)). Peter's theoretical ideas were brought into these discussions, and interest in implementing the Fourier method was rekindled. We had a few brief email conversations with Peter and discussed plans for the three of us to publish this paper, but Peter's illness brought those conversations to a close, and it was with sadness that we learned of his passing.

In October, 2016, we made one more attempt at numerically implementing the method, scanning Peter's carefully constructed theoretical arguments for clues that might assist us in practically implementing the method. Gradually, we began to see that our earlier numerical efforts had been based on unnecessary simplifications, leading to horribly suboptimal solutions; full implementation of the technique was, in fact, not only possible, but it also gave very good results. This paper sets out, then, to show that a Fourier method for kernel density and distribution function estimation for interval-censored data can work, both theoretically and practically.

Peter's original outline included a plan for numerical implementation; we have chosen instead to relegate that material as well as various extensions and the technical arguments to the supplementary material, so that Peter's voice can be heard in an almost continuous stream from Sections 2 through 4. We are honoured to have been able to interact with Peter on this problem, and we join the large chorus of other scientists who will miss him tremendously.

1. Introduction

Methods to obtain smooth estimates of the probability density function of a random variable when the latter is observed subject to interval censoring have received considerable attention for many years. These methods are useful because in many applications the realized values of the variable of interest are not known exactly but only up to an interval. A review of methodologies for smooth estimation of the density, survival or hazard function given an interval censored sample can be found in Sun (2007, Chap. 3). To give a rough summary, we mention the logspline method of Kooperberg and Stone (1992), the local EM method of Betensky et al. (1999) and the kernel smoothing approach of Braun, Duchesne and Stafford (2005). Most of the aforementioned methods are connected with nonparametric maximum likelihood estimation and do not explicitly model the censoring time process. But as we shall demonstrate such modeling allows one to recast the problem of density estimation with interval-censored data as a deconvolution problem, which paves the way to approach these estimation problems with a new set of tools.

Density estimation based on deconvolution has been thoroughly studied in the literature, including seminal work by Peter Hall. We cannot reasonably list all of Peter's contributions to this field here, but his important contributions include his paper with R. J. Carroll (Carroll and Hall (1988)) on the optimal convergence rate for deconvolution density estimators and his proposals of new approaches based on truncated Fourier inversion in Diggle and Hall (1993) or on discrete Fourier transforms in 2005 (Hall and Qiu (2005)).

Peter has also investigated the use of deconvolution methods in measurement error problems (see for instance Delaigle, Hall and Meister (2008)). Though somewhat related, interval-censoring and measurement error are generally not equivalent. One case where they do coincide is when the variable of interest is measured with random uniform error. In this case, Groeneboom and Jongbloed (2003) proposed a deconvolution method for density estimation. But uniform measurement error is only a special case of interval censoring, and deconvolution methods for the general case do not seem to have been considered. Deconvolution methods are not the only option to obtain estimators from a Fourier transform. As Feuerverger and McDunnough (1981a); Feuerveger and McDunnough (1981b) have shown, estimating equations based on Fourier transforms can form a basis for inference in many problems. In this paper, we propose such a Fourier-based inference method for smooth nonparametric estimation of a density function when a parametric model for the process that generates the potential interval-censoring times is available. The interval-censoring model considered is presented in Section 2. A new density estimator is proposed in Section 3, and its convergence properties are discussed in Section 4. The supplementary material for the paper contains additional discussion, including the technical arguments justifying the methodology, and it provides guidelines for practical implementation of the method. References to equations and sections in the supplementary material are prefixed with the letter 'S'.

2. Model and Identifiability

We wish to make inference about the common distribution of random variables $X_1, X_2, \ldots$, which are interval censored. Specifically, we have access only to a sample of random intervals, $\mathcal{I} = \{[L_1, R_1], [L_2, R_2], \ldots\}$, often assumed to be generated as follows. For each i, a potentially infinite, stationary point process $\mathcal{T}_i = \{\ldots, T_{ij}, T_{i,j+1}, \ldots\}$ produces the interval endpoints $L_i \leq R_i$ defined by

$$L_i = \sup\{T_{ij} : T_{ij} \leq X_i\}, \qquad R_i = \inf\{T_{ij} : T_{ij} \geq X_i\}. \tag{2.1}$$

The pairs $(X_1, \mathcal{T}_1), (X_2, \mathcal{T}_2), \ldots$ are identically distributed, the quantities $X_1, X_2, \ldots$ and $\mathcal{T}_1, \mathcal{T}_2, \ldots$ are independent of one another, and the sequences $\mathcal{T}_i$ may or may not be observed. This amounts to the independent inspection process model of Lawless and Babineau (2006, Sec. 2) and to the case K interval censoring model (Sun (2007, Sec. 1.3)). Given the first n intervals in the set $\mathcal{I}$ we wish to estimate the distribution, and more particularly the probability density, of a generic value X of X_i.

Define $Z_1 = X_i - L_i$ and $Z_2 = R_i - X_i$, where we have suppressed the dependence of Z_1 and Z_2 on i. By definition, $P(Z_1 \geq 0) = P(Z_2 \geq 0) = 1$. Since the processes $\mathcal{T}_i$ are stationary, the distribution of (Z_1, Z_2), conditional on $X_i = x$, does not depend on x, and, since the pairs $(X_1, \mathcal{T}_1), (X_2, \mathcal{T}_2), \ldots$ are identically distributed, the distribution also does not depend on i. Bearing this in mind, we define the joint distribution of Z_1 and Z_2 to be the distribution conditional on X_i. In this notation, if (L, R, X) has the distribution of a generic triple (L_i, R_i, X_i) then (L, R, X) is distributed as $(X - Z_1, X + Z_2, X)$, where we take X to be independent of (Z_1, Z_2). Without loss of generality,

$$(L, R, X) = (X - Z_1, X + Z_2, X) \quad \text{where} \quad X \perp (Z_1, Z_2), \tag{2.2}$$

with $\perp$ denoting "is independent of."

The characteristic functions f_{LR}^{Ft}, f_X^{Ft}, and $f_{Z_1,Z_2}^{\mathrm{Ft}}$ of the distributions of (L, R), X and (Z_1, Z_2), respectively, satisfy

$$\begin{aligned} f_{LR}^{\mathrm{Ft}}(s,t) &= E\{\exp(isL + itR)\} = E\{\exp(is(X - Z_1) + it(X + Z_2))\} \\ &= E\{\exp(i(s+t)X - isZ_1 + itZ_2)\} = f_X^{\mathrm{Ft}}(s+t) f_{Z_1,Z_2}^{\mathrm{Ft}}(-s,t) \end{aligned}$$

where, on this occasion $i = \sqrt{-1}$ rather than denoting an index, and more generally, i has this interpretation in the expressions is and it. (The notation f_{LR}^{Ft} represents the Fourier transform, hence the superscript Ft, of the probability density f_{LR} of (L, R).) In summary,

$$f_{LR}^{\mathrm{Ft}}(s,t) = f_X^{\mathrm{Ft}}(s+t) f_{Z_1,Z_2}^{\mathrm{Ft}}(-s,t). \tag{2.3}$$

If (Z_1, Z_2) denotes a pair of nonnegative random variables for which (2.2) holds then, in general, neither the distribution of X nor that of (Z_1, Z_2) is identifiable from data on (L, R) alone. To appreciate why, let U, V, W, X be independent random variables for which $U, V, W \geq 0$, and put $(Z_1, Z_2) = (U - V, W + V)$. (Nonnegativity of (Z_1, Z_2) can be ensured under a side condition, for example by asking that, for a constant $c > 0$, $P(U \geq c) = P(W \geq 0) = P(0 \leq V \leq c) = 1$.) Then,

$$(L, R) = (X - Z_1, X + Z_2) = (X' - Z_1', X' + Z_2'),$$

where $X' = X + V$, $Z_1' = U$ and $Z_2' = W$. Moreover, the pair (Z_1', Z_2') is independent of X', and Z_1' and Z_2' are nonnegative. However, unless V is identically zero, the distributions of X and X' differ, as do those of (Z_1, Z_2) and (Z_1', Z_2'). Therefore, in the model at (2.2), even with the additional constraint that Z_1 and Z_2 are nonnegative, the distributions of X and (Z_1, Z_2) are not nonparametrically identifiable from data on (L, R) alone.

This lack of identifiability implies that, in a general interval-censoring problem, it is not possible to estimate either f_X or f_{Z_1,Z_2} nonparametrically using only data on (L, R). As is well documented in the interval-censoring literature (see Sun (2007, Chap. 3)) and references therein), data on (L, R) contain no information about f_X over so-called innermost intervals and the nonparametric maximum likelihood estimator of the corresponding cumulative distribution function is undefined over these intervals. In cases where L and R are defined more narrowly in terms of stationary point processes, as at (2.2), the ambiguity is less, since the class of possible distributions of (Z_1, Z_2) is restricted by that definition. If it is considered that the assumption of stationarity of the point processes $\mathcal{T}_i$ can be invalid (for example, because the point processes have not been run long enough before measurements are made), or that other assumptions are compromised, then inference is still vulnerable to problems caused by non-indentifiability.

Therefore, although the methods that we give in Section 3.1 can be modified so that, under the specific assumption of identifiability, they give consistent estimators of the density of X without using a model for the distribution of

(Z_1, Z_2), and employing only interval data, we instead discuss inference in cases where either:

(a) we have a parametric model for (Z_1, Z_2), for example derived when the point processes $\mathcal{T}_i$ are Poisson or, more generally, renewal processes; or

(b) the distribution of (Z_1, Z_2) is estimated nonparametrically from observations of the $\mathcal{T}_i$s.

These approaches alleviate the identifiability problem by, in case (a), greatly reducing the variety of options that are available for the distribution of (Z_1, Z_2) when using a parametric model, or, in case (b), specifying the distribution of (Z_1, Z_2) in terms of a consistent, nonparametric estimator of the true distribution. See also Lawless and Babineau (2006, Sec. 5) who discuss the estimation of the parameters of the inspection time process and propose a simulation-based inference method.

3. Estimators

3.1. General methodology

We either assume a parametric model for the distribution of (Z_1, Z_2), where the characteristic function is $f^{\mathrm{Ft}}_{Z_1,Z_2}(s, t|\theta)$, say, and θ is a finite vector of unknown parameters; or we estimate the distribution of (Z_1, Z_2) from point process data. These are the respectively cases (a) and (b) discussed in Section 3. In case (b), and conditional on $X_i = x$, the values of L_i and R_i are

$$L_i(x) = \sup\{T_{ij} : T_{ij} \le x\}, R_i(x) = \inf\{T_{ij} : T_{ij} \ge x\}$$

compare (2.1). Therefore a nonparametric estimator of $g(s, t) = f^{\mathrm{Ft}}_{Z_1,Z_2}(s, t)$ has the form

$$\widehat{g}(s, t) = \frac{1}{n(b-a)} = \sum_{i=1}^{n} \int_a^b \exp\left(is\{x - L_i(x)\} + it\{R_i(x) - x\}\right) dx, \tag{3.1}$$

where (a, b) denotes an interval (a little) shorter than the domain of the point processes $\mathcal{T}_i$.

In each of cases (a) and (b) our methodology for estimating the distribution F_X of X is based on approximating the density $f_X = F'_X$ by a histogram,

$$f_X(x|\mathcal{B}, \omega) = \sum_{j=1}^{m} \omega_j I(x \in \mathcal{B}_j), \tag{3.2}$$

where $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_m\}$ denotes a sequence of adjacent a histogram bins $\mathcal{B}_j$

of width $h > 0$, the nonnegative weights ω_j satisfy $h\sum_j \omega_j = 1$, and $\omega = (\omega_1, \ldots, \omega_m)$ is the vector of unknown parameters that we wish to estimate. The class of all distributions with densities of this type is dense in the class of all distributions, and so constructing a histogram density as at (3.2) should be seen as a method for general distribution estimation, not density estimation. See Section 3.2 for details. In Section 3.3 we suggest smoothing the resulting distribution estimator to obtain a density estimator, and discuss choice of the smoothing parameter.

In case (a), a common parametric model for (Z_1, Z_2) is that in which the point processes $\mathcal{T}_i$ are Poisson with intensity λ^{-1}. Here Z_1 and Z_2 are independent and identically distributed with density $\lambda \exp(-\lambda z)$, for $z > 0$. Thus, $\theta = \lambda$ is a scalar, and

$$f^{\mathrm{Ft}}_{Z_1,Z_2}(s,t|\lambda) = (1-\lambda^{-1}is)^{-1}(1-\lambda^{-1}it)^{-1} \tag{3.3}$$

Section S2.2 in the online supplementary material discusses generalisations of this model when the processes $\mathcal{T}_i$ are renewal processes. More generally in case (a), the parameter vectors θ and ω can be estimated by least squares, minimising

$$S_q(\theta,\mathcal{B},\omega) = \left\{ \int\int \left| f^{\mathrm{Ft}}_{LR}(s,t) - f^{\mathrm{Ft}}_X(s+t|\mathcal{B},\omega) f^{\mathrm{Ft}}_{Z_1,Z_2}(-s,t|\theta) \right|^q w(s,t)^q ds dt \right\}^{1/q}, \tag{3.4}$$

where $q \geq 1$,

$$\widehat{f^{\mathrm{Ft}}_{LR}}(s,t) = \frac{1}{n}\sum_{j=1}^{n} \exp(isL_j + itR_j)$$

is a conventional nonparametric, unbiased estimator of the characteristic function of (L,R), $f^{\mathrm{Ft}}_X(\cdot|\mathcal{B},\omega)$ denotes the characteristic function of the distribution with histogram density $f^{\mathrm{Ft}}_X(\cdot|\mathcal{B},\omega)$, defined at (3.2), and w in (3.4) is an integrable, nonnegative weight function. The criterion function $S_q(\theta,\mathcal{B},\omega)$ is motivated by (2.3). One may be tempted to compute $\widehat{f^{\mathrm{Ft}}_X}(s,t) = \widehat{f^{\mathrm{Ft}}_{LR}}(s,t)/\widehat{f^{\mathrm{Ft}}_{Z_1,Z_2}}(-s,t)$ and then invert $\widehat{f^{\mathrm{Ft}}_X}(s,t)$ to obtain an estimate of f_X, but $\widehat{f^{\mathrm{Ft}}_X}(s,t)$ does not depend on (s,t) only through $s+t$ for finite samples, except in very specific cases; this issue is avoided when minimizing (3.4).

3.2. Distribution estimation

In numerical practice we suggest choosing the histogram to minimise the distance from the empirical characteristic function $\widehat{f^{\mathrm{Ft}}_{LR}}$ to our model for this characteristic function, differing somewhat in cases (a) and (b).

In case (a) we take $(\widehat{\theta}, \widehat{\mathcal{B}}, \widehat{\omega})$ to minimise $S_q(\theta, \mathcal{B}, \omega)$ at (3.4) (or to minimise a similar quantity such as that at (S2.4)— these approaches are both part of case (a)), and in case (b) we take $(\widehat{\mathcal{B}}, \widehat{\omega})$ to minimise $S_q(\mathcal{B}, \omega)$ at (S2.5). Importantly, in both settings the minimum is taken over h, m and choices of $\omega_1, \ldots, \omega_m$, not just over the latter. We could also take it over choices of the bin centres, although this is generally not necessary. The result is a histogram,

$$\widehat{f}_X(x) \equiv f_X(x|\widehat{\mathcal{B}}, \widehat{\omega}) = \sum_{j=1}^{m} \widehat{\omega}_j I(x \in \widehat{\mathcal{B}}_j),$$

that is generally too rough to be a useful estimator of f_X, but its integral is appropriate as an estimator $\widehat{F}_X = F(\cdot|\mathcal{B}, \widehat{\omega})$ of the distribution F_X with density f_X.

In theoretical terms, $\widehat{F}_X$ can be taken to be any weak limit of any sequence of histogram distributions with densities $f_X(\cdot|\mathcal{B}, \omega)$ along which the minimum of (3.4) in case (a), or of (S2.5) in case (b), is obtained. (Any sequence of distributions, here a sequence of histogram distributions, has a convergent subsequence.) In numerical terms, $\widehat{F}_X$ is constructed to be the member of a sequence of histogram distributions, defined iteratively, that results when the algorithm for minimising either (3.4) or (S2.5) terminates.

3.3. Density estimation

We can smooth $\widehat{F}_X$ to an estimator $\widetilde{f}_X$ of f_X in many ways. For example, if we favour kernel methods we can take

$$\begin{aligned}\widetilde{f}_X(x) &= \int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) d\widehat{F}_X(y) = \int_{-\infty}^{\infty} K(y) d\widehat{F}_X(x - hy) \\ &= \frac{1}{h} \sum_{j=1}^{m} \widehat{\omega}_j \int_{\mathcal{B}_j} K\left(\frac{x-y}{h}\right) dy, \qquad (3.5)\end{aligned}$$

where h is a bandwidth and K a kernel function.

4. Convergence Properties

In Theorem 1 we state conditions that are sufficient for the estimator $F_X(\cdot|\widehat{\mathcal{B}}, \widehat{\omega})$ of the distribution function F_X of X to converge to F_X at a polynomial rate in n^{-1}. Theorem 2 observes that those assumptions, together with a minor smoothness constraint on the density $f_X = F_X'$, are also sufficient for the kernel density estimator based on $F_X(\cdot|\widehat{\mathcal{B}}, \widehat{\omega})$ to converge uniformly to f_X. It is convenient here to work with criterion functions S_q where $q = \infty$, although convergence rates can also be derived in the case of finite q.

We separately treat the cases (a) and (b), introduced in Section 3, representing parametric and nonparametric settings, respectively. In case (a) we have a model $f_{Z_1,Z_2}(\cdot|\theta)$ for the joint density of (Z_1, Z_2), and it is helpful to define

$$s(\theta,\mathcal{B},\omega) = \sup_{-\infty<s,t<\infty} |f_X^{\mathrm{Ft}}(s+t) f_{Z_1,Z_2}^{\mathrm{Ft}}(-s,t|\theta_0) - f_X^{\mathrm{Ft}}(s+t|\mathcal{B},\omega) f_{Z_1,Z_2}^{\mathrm{Ft}}(-s,t|\theta)| w(s,t), \tag{4.1}$$

where θ_0 denotes the true value of θ. We quantify the identifiability of both f_X and the finite parameter vector θ by assuming that for constants $C_1, C_2 > 0$, for all values of θ in some neighbourhood of θ_0, and for all values of $(\mathcal{B},\omega)$ for which the supremum on the right-hand side of (4.2) does not exceed some given positive number,

$$s(\theta,\mathcal{B},\omega) \geq C_1 \left\{ ||\theta-\theta_0|| + \sup_{-\infty<s,t<\infty} |f_X^{\mathrm{Ft}}(s+t) - f_X^{\mathrm{Ft}}(s+t|\mathcal{B},\omega)| w_1(s,t) \right\}^{C_2}, \tag{4.2}$$

where w_1 is a nonnegative weight function and $||\cdot||$ is the usual Euclidean norm. The inequality (4.2) is readily shown to be satisfied with $C_2 = 1$ in many cases of practical interest, for example where $f_{Z_1,Z_2}^{\mathrm{Ft}}(\cdot|\theta)$ is a differentiable function of θ.

In addition we ask that:

(i) for a constant $C_3 > 0, E|L|^{C_3} + E|R|^{C_3} + E|X|^{C_3} + EZ_1^{C_3} + EZ_2^{C_3} < \infty$; (ii) the infimum in the definition $(\widehat{\theta},\widehat{\mathcal{B}},\widehat{\omega}) = \mathrm{arginf}_{(\theta,\mathcal{B},\omega)} S_\infty(\theta,\mathcal{B},\omega)$ is, for each fixed θ, taken over all distributions having density $f_X(\cdot|\mathcal{B},\omega)$ and satisfying $\int |x|^{C_3} f_X(x|\mathcal{B},\omega)dx \leq C_4$, where $C_4 > 0$ also has the property $E|X|^{C_3} < C_4$; (iii) for constants $C_5, C_6 > 0$, $\inf_{|s|,|t|\leq u} |f_{Z_1,Z_2}^{\mathrm{Ft}}(s,t)| \geq C_5(1+|u|)^{-C_6}$ for all $u > 0$; (iv) the true density f_X of X is uniformly bounded; (v) for all real s and $t, w(s,t) = w(t,s) = w(-s,t)$; and (vi) for constants $C_7,\ldots,C_{10} > 0$ and all real s and $t, C_7(1+|s|)^{-C_8}(1+|t|)^{-C_8} \leq \min\{w_1(s,t), w(s,t)\} \leq \max\{w_1(s,t), w(s,t)\} \leq C_9(1+|s|)^{-C_{10}} \times (1+|t|)^{-C_{10}}$. (4.3)

Assumption (4.3)(i) asks only that the random variables under consideration have a moment of some positive order and, since C_3 can be arbitrarily small, it is particularly weak; (4.3)(ii) asserts that the approximating distribution is constructed so that it also satisfies (4.3)(i), which in practice is readily imposed by constraining the locations and the number, m, of histogram blocks. Condition (4.3)(iii) ensures that the tails of the characteristic function of the bivariate error distribution do not decay at a faster rate than the inverse of a polynomial in s and t, and is commonly imposed in deconvolution problems, for example using an exponential model; such a model is suggested in the present setting by (3.3). (The converse case, where the rate of decrease of the tails of the characteristic function is exponentially fast, is termed "supersmooth" in the context of deconvolution problems, and results in convergence rates that are slower than any polynomial in n^{-1}.) Assumption (4.3)(iv) requires only that f_X be bounded; and (4.3)(v) and (4.4)(vi) are weak conditions on the weight function w_1.

In case (b) there there is no model for the joint distribution of (Z_1, Z_2), and we estimate $g = f^{\mathrm{Ft}}_{Z_1,Z_2}$ using g, defined at (3.1). We define $S_\infty(\mathcal{B}, \omega)$ as at (S2.6), and impose an analogue of (4.3):

$$\begin{aligned}&\text{Assumptions (i) and (iii)–(v) are as in (4.3); (vi) is as in (4.3) but with the}\\&\text{function } w_1 \text{ dropped; (ii) is replaced by the property: (ii), the infimum}\\&\text{in the definition } (\widehat{B}, \widehat{\omega}) = \mathrm{arginf}_{(B,\omega)} S_\infty(B, \omega) \text{ is taken over all distributions}\\&\text{having density } f_X(\cdot|\mathcal{B}, \omega) \text{ and satisfying } \int |x|^{C_3} f_X(x|\mathcal{B}, \omega)dx \le C_4, \text{where}\\&C_4 > 0 \text{ has the property } E|X|^{C_3} < C_4.\end{aligned} \tag{4.4}$$

Throughout we use the objective function S_∞, defined at (S2.3) in case (a) and at (S2.6) in case (b), to define estimators.

Theorem 1. *(a) In the parametric case, if* (4.2) *and* (4.3) *hold then there exists* $\varepsilon > 0$ *such that*

$$||\widehat{\theta} - \theta_0|| = O_p(n^{-\varepsilon}), \sup_{-\infty<x<\infty} |F_X(x|\widehat{\mathcal{B}}, \widehat{\omega}) - F_X(x)| = O_p(n^{-\varepsilon}). \tag{4.5}$$

(b) In the nonparametric case, if (4.4) *holds then there exists* $\varepsilon > 0$ *such that*

$$\sup_{-\infty<x<\infty} |F_X(x|\widehat{\mathcal{B}}, \widehat{\omega}) - F_X(x)| = O_p(n^{-\varepsilon}). \tag{4.6}$$

A kernel density estimator, $\widetilde{f}_X$, of f_X, derived from the distribution estimator $F_X(x|\widehat{\mathcal{B}}, \widehat{\omega})$ and based on a kernel K and bandwidth h, is given by (3.5). We permit h to decrease to zero as n increases; the performance of $\widetilde{f}_X$ depends on choice of h, discussed in Section 3.3. To apply Theorem 1 to the problem of convergence of $\widetilde{f}_X$ we impose a conventional Hölder smoothness condition on f_X:

$$\text{for constants } C_{11}, \delta > 0, \sup_{-\infty < u, x < \infty} |u|^{-\delta} |f_X(x+u) - f_X(x)| \leq C_{11}. \tag{4.7}$$

We ask that K satisfy

$$\int (1+|u|)^{\delta} |K(u)| du < \infty, \int K = 1, \int |K'| < \infty, \tag{4.8}$$

where $\delta > 0$ is as in (4.7).

Theorem 2. *Assume* (4.7) *and* (4.8)*, and that either* (4.2) *and* (4.3) *in the parametric case (a), or* (4.4) *in the nonparametric case (b), hold. Let* $\varepsilon > 0$ *be as at* (4.5) *or* (4.6) *in those two respective cases. Then*

$$\sup_{-\infty < x < \infty} |\widetilde{f}_X(x) - f_X(x)| = O_p(n^{-\varepsilon} h^{-1} + h^{\delta}). \tag{4.9}$$

An immediate corollary of (4.9) is that if $h = h(n) \to 0$ sufficiently slowly to ensure that $n^{\varepsilon} h \to \infty$, then the density estimator $\widetilde{f}_X$ is uniformly consistent for f_X.

Supplementary Materials

Additional information is provided in the Supplementary Materials section. In particular, there is a consideration of identifiability for the models given in Section 2, and further insight into the choice of weight function in Section 3. Bandwidth selection is briefly considered. Generalization to renewal processes for the monitoring process is also discussed. The technical arguments behind the proofs of Theorems 1 and 2 are provided. This section concludes with advice on numerically implementing the technique.

Acknowledgment

Support from the Canadian Statistical Sciences Institute is gratefully acknowledged.

References

Betensky, R. A., Lindsey, J. C., Ryan, L. M. and W and, M. P. (1999). Local EM estimation of the hazard function for interval-censored data. *Biometrics* **55**, 238–245.

Braun, W. J., Duchesne, T. and Stafford, J. E. (2005). Local likelihood density estimation for interval censored data. *Canadian Journal of Statistics* **33**, 39–60.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Associations* **83**, 1184–1186.

Delaigle, A., Meister, A. and Hall, P. (2008). On deconvolution with repeated measurements. *The Annals of Statistics* **36**, 665–685.

Diggle, P. J. and Hall, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *Journal of the Royal Statistical Society, series B (Statistical Methodology)* **55**, 523–531.

Feuerverger, A. and McDunnough, P. (1981a). On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **43**, 20–27.

Feuerverger, A. and McDunnough, P. (1981b). On some Fourier methods of inference. *Journal of the American Statistical Association* **76**, 379–387.

Groeneboom, P. and Jongbloed, G. (2003). Density estimation in the uniform deconvolution model. *Statistica Neerlandica* **57**, 136–157.

Hall, P. and Qiu, P. (2005). Discrete-transform approach to deconvolution problems. *Biometrika* **92**, 135–148.

Kooperberg, C. and Stone, C. J. (1992). Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics* **1**, 301–328.

Lawless, J. F. and Babineau, D. (2006). Models for interval censoring and simulation-based inference for lifetime distributions. *Biometrika* **93**, 671–686.

Sun, J. (2007). *The Statistical Analysis of Interval-Censored Failure Time Data*. Springer, New York.

School of Mathematics and Statistics, University of Melbourne, Victoria 3010, Australia.

SCI 388 - 1177 Research Road, University of British Columbia (Okanagan Campus), Kelowna, BC V1V 1V7, Canada.

E-mail: john.braun@ubc.ca

Département de mathématiques et de statistique, Université Laval, 1045, avenue de la Médecine, Québec, QC, G1V 0A6, Canada.

E-mail: Thierry.Duchesne@mat.ulaval.ca

(Received January 2017; accepted June 2017)