Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 13–19

FOURIER METHODS FOR SMOOTH DISTRIBUTION FUNCTION ESTIMATION

JOSÉ E. CHACÓN, PABLO MONFORT AND CARLOS TENREIRO

ABSTRACT: In this paper we show how to use Fourier transform methods to analyze the asymptotic behavior of kernel distribution function estimators. Exact expressions for the mean integrated squared error in terms of the characteristic function of the distribution and the Fourier transform of the kernel are employed to obtain the limit value of the optimal bandwidth sequence in its greatest generality. The assumptions in our results are mild enough so that they are applicable when the kernel used in the estimator is a superkernel, or even the sinc kernel, and this allows to extract some interesting consequences, as the existence of a class of distributions for which the kernel estimator achieves a first-order improvement in efficiency over the empirical distribution function.

KEYWORDS: Fourier analysis, kernel distribution estimation, mean integrated squared error, optimal bandwidth, sinc kernel.
AMS SUBJECT CLASSIFICATION (2010): 62G05, 62G20.

1. Introduction

The kernel estimator of a distribution function was introduced independently by Tiago de Oliveira (1963), Nadaraya (1964) and Watson and Leadbetter (1964) as a smooth alternative to the empirical estimator. It is defined as the distribution function corresponding to the well-known kernel density estimator. Precisely, given independent real random variables $X_1, \ldots, X_n$ with common and unknown distribution function F, assumed to be absolutely continuous with density function f, the kernel estimator of $F(x)$ is

$$F_{nh}(x) = n^{-1} \sum_{j=1}^{n} K\big(h^{-1}(x - X_j)\big),$$

where $h > 0$ is the bandwidth and the function K will be referred to as the integrated kernel, since it is assumed that $K(x) = \int_{-\infty}^{x} k(y)dy$ for some integrable function k, called kernel, having unit integral over the whole real line.

Received May 17, 2013.
This work has been supported by grant MTM2010-16660 from the Spanish Ministerio de Ciencia e Innovación and by the Centro de Matemática da Universidade de Coimbra (CMUC), funded by the European Regional Development Fund through the program COMPETE and by the Portuguese Government through the FCT - Fundação para a Ciência e Tecnologia under the project PEst-C/MAT/UI0324/2011.

Classical references on kernel distribution function estimators include Yamato (1973), which provided mild necessary and sufficient conditions for its consistency in uniform norm, Azzalini (1981), Swanepoel (1988) and Jones (1990) on asymptotic squared error analysis of the estimator, or Sarda (1993), Altman and Léger (1995) and Bowman, Hall and Prvan (1998), and more recently Polansky and Baker (2000) and Tenreiro (2006), on data-driven bandwidth selection. There are also other recent papers on different aspects of kernel distribution function estimation, like Tenreiro (2003), Swanepoel and Van Graan (2005), Janssen, Swanepoel and Veraverbeke (2007), Giné and Nickl (2009), Berg and Politis (2009), Chacón and Rodríguez-Casal (2010) or Mason and Swanepoel (2012). See Servien (2009) for a detailed survey on distribution function estimation, not limited to kernel-type methods.

This paper is devoted to the study of the kernel distribution function estimator from the point of view of the mean integrated squared error,

$$\mathrm{MISE}(h) \equiv \mathrm{MISE}_n(h) = \mathbb{E} \int_{-\infty}^{\infty} \{F_{nh}(x) - F(x)\}^2 dx.$$

In this sense, the optimal bandwidth h_{0n} is the value of $h > 0$ minimizing $\mathrm{MISE}(h)$. The existence of such a bandwidth was proved in Theorem 1 of Tenreiro (2006) under very general assumptions, and Proposition 2 in the same paper showed that $h_{0n} \to 0$ whenever the Fourier transform of k is not identically equal to 1 on any neighbourhood of the origin. This condition can be considered mild as well, since it is satisfied for any finite-order kernel; however, it does not hold for a superkernel (see Chacón, Montanero and Nogales, 2007).

The purpose of this note is to show how to use Fourier transform techniques for the analysis of kernel distribution estimators. Particularly, expressing the MISE in terms of characteristic functions allows us to obtain a result on the limit behavior of the optimal bandwidth sequence in its most general form so that it also covers the case of a superkernel, and to explore its consequences showing the peculiar properties of the use of superkernels and the sinc kernel in kernel distribution function estimation. Precisely, it is shown in Section 2 that in some situations the sequence h_{0n} does not necessarily tend to zero. Moreover, we exhibit a class of distributions for which the kernel distribution estimator presents a first-order improvement over its empirical counterpart, opposite to the usual situation, where only second-order improvements are possible (see Remark 3). Our findings are illustrated in Section 3 through two representative examples.

2. Main results

Recall from Chacón and Rodríguez-Casal (2010) that the kernel distribution function estimator admits the representation

$$F_{nh}(x) = \int F_n(x - hz)dK(z), \tag{1}$$

where F_n denotes the empirical distribution function (here and below integrals without integration limits are meant over the whole real line). Using this, and standard properties of the empirical process, it is possible to obtain a decomposition of $\mathrm{MISE}(h) = \mathrm{IV}(h) + \mathrm{ISB}(h)$, where the integrated variance $\mathrm{IV}(h) = \int \mathrm{Var}\{F_{nh}(x)\}dx$ and the integrated squared bias $\mathrm{ISB}(h) = \int\{\mathbb{E}[F_{nh}(x)] - F(x)\}^2dx$ can be expressed in the following exact form:

$$\mathrm{IV}(h) = n^{-1}\iiint\big\{F\big(x-h(y\vee z)\big) - F(x-hy)F(x-hz)\big\}dK(y)dK(z)dx, \tag{2}$$

$$\mathrm{ISB}(h) = \iiint\{F(x-hy) - F(x)\}\{F(x-hz) - F(x)\}dK(y)dK(z)dx, \tag{3}$$

with $y \vee z$ standing for $\max\{y, z\}$.

Note that the representation (1) and the exact expressions (2) and (3) also make sense for $h = 0$, implying that the kernel distribution estimator reduces to the empirical distribution function for $h = 0$, for which the well-known MISE formula reads $\mathrm{MISE}(0) = \mathrm{IV}(0) = n^{-1}\int F(1-F)$ whenever $\psi(F) = \int F(1-F)$ is finite. Moreover, it is not hard to check that $\int |x|dF(x) < \infty$ and $\int |y\,k(y)|dy < \infty$ ensure that $\mathrm{MISE}(h)$ is finite for all $h > 0$, so those two minimal conditions will be assumed henceforth. Note that the required condition that F have a finite mean is slightly stronger than $\psi(F) < \infty$ since $\psi(F) \leq 2\int |x|dF(x)$.

2.1. Limit behavior of the optimal bandwidth sequence. Denote by φ_g the Fourier transform of a function g, defined as $\varphi_g(t) = \int e^{itx}g(x)dx$. As in Chacón *et al.* (2007), the key to understand the limit behavior of the optimal bandwidth sequence is to use Fourier transforms to express the MISE criterion. Abdous (1993) provided a careful account of the necessary conditions under which the MISE can be expressed in terms of Fourier transforms. The proof of his Proposition 2 implicitly derives formulas for $\mathrm{ISB}(h)$ and $\mathrm{IV}(h)$ in terms of φ_k and φ_f for $h > 0$. We reproduce this result here for completeness, and show that it can be extended to cover the case $h = 0$ as well.

Theorem 1. *If* $\int |x|dF(x) < \infty$ *and* $\int |y\ k(y)|dy < \infty$ *then, for all* $h \geq 0$, *the IV and ISB functions can be written as*

$$\mathrm{IV}(h) = (2\pi)^{-1}n^{-1}\int t^{-2}|\varphi_k(th)|^2\{1 - |\varphi_f(t)|^2\}dt,$$
$$\mathrm{ISB}(h) = (2\pi)^{-1}\int t^{-2}|1 - \varphi_k(th)|^2|\varphi_f(t)|^2 dt.$$

Particularly, note that for $h = 0$ the previous result yields a Parseval-like formula for distribution functions,

$$\psi(F) = \int F(1 - F) = (2\pi)^{-1}\int t^{-2}\{1 - |\varphi_f(t)|^2\}dt, \tag{4}$$

which can be useful to compute errors in an exact way in cases where F does not have a close expression but φ_f does, as it happens for instance for the normal distribution (see also Section 3 below). Moreover, we show in Lemma 1 that (4) remains valid for integrated kernels K. In the following it will be assumed that $\psi(K) > 0$, a property that immediately holds, using (4), whenever $\varphi_k(t) \in [0, 1]$ for all t. Note that, for density estimation, admissible kernels are precisely those whose Fourier transform satisfies that restriction (see Cline, 1988).

The limit behavior of the optimal bandwidth sequence h_{0n} is determined in its greatest generality by the following constants, depending on the Fourier transforms of f and k: let C_f denote the smallest positive frequency from which φ_f is null along a proper interval and D_f the positive frequency from which φ_f is identically null (so that $C_f \leq D_f$, both possibly being infinite); also, denote S_k the greatest frequency such that φ_k is identically equal to one on $[0, S_k]$ and T_k the smallest frequency such that φ_k is not identically equal to one on a subinterval of $[T_k, \infty)$, and note that $S_k \leq T_k$ with both possibly being zero. In mathematical terms,

$$\begin{aligned}
C_f &= \sup\{r \geq 0\colon \varphi_f(t) \neq 0 \text{ a.e. for } t \in [0, r]\},\\
D_f &= \sup\{t \geq 0\colon \varphi_f(t) \neq 0\},\\
S_k &= \inf\{t \geq 0\colon \varphi_k(t) \neq 1\},\\
T_k &= \inf\{r \geq 0\colon \varphi_k(t) \neq 1 \text{ a.e. for } t \geq r\}.
\end{aligned}$$

Finally, define $h_* = \sup\{h \geq 0\colon \mathrm{ISB}(h) = 0\}$. The following result shows the limit of the optimal bandwidth sequence h_{0n} in the common case where $C_f = D_f$ and $S_k = T_k$.

Theorem 2. *Assume that* $\int |x|dF(x) < \infty$, $\int |y\,k(y)|dy < \infty$ *and* $\psi(K) > 0$, *and suppose that* $C_f = D_f$ *and* $S_k = T_k$. *Then,* $h_{0n} \to S_k/D_f$ *as* $n \to \infty$ *and also* $h_* = S_k/D_f$.

A number of consequences can be extracted from Theorem 2:

Remark 1. A kernel k with $S_k > 0$ is called a superkernel (see Chacón, Montanero and Nogales, 2007). If an integrated superkernel is used in the kernel distribution function estimator and the density f is such that $D_f < \infty$ (see Chacón *et al.*, 2007, and Section 3 below for examples of such distributions) then, contrary to the usual situation, the optimal bandwidth sequence h_{0n} does not tend to zero, but to the strictly positive constant S_k/D_f. Moreover, any positive constant can be the limit of an optimal bandwidth sequence, because modifying the scale of the density by taking $f_a(x) = f(x/a)/a$, for any $a > 0$, it follows that $D_{f_a} = D_f/a$, and hence the limit of the optimal bandwidth sequence equals aS_k/D_f.

Remark 2. Since $h_* = S_k/D_f$, the kernel estimator F_{nh} is unbiased for any fixed (i.e., not depending on n) choice of $h \in [0, S_k/D_f]$. If either K is not an integrated superkernel or the characteristic function φ_f does not have bounded support, then the only kernel distribution estimator with null ISB corresponds to $h = 0$, the empirical distribution function.

Remark 3. It is shown in the proof of Theorem 2 that for $h \in [0, S_k/D_f]$ the MISE of F_{nh} admits the exact expression $\mathrm{MISE}(h) = n^{-1}\psi(F) - n^{-1}\psi(K)h$. From this, it follows that for any fixed $h \in (0, S_k/D_f]$ the kernel estimator F_{nh} presents an asymptotic first-order reduction in MISE over the empirical estimator; that is, its MISE is of order n^{-1} as for the empirical estimator, yet with a strictly smaller constant (namely, $\psi(F) - \psi(K)h < \psi(F)$). As a result, over the class of distributions with D_f bounded by a constant (say, $D_f \le M$) the kernel estimator with bandwidth $h = S_k/M$ is strictly more efficient than the empirical distribution function F_n. This is in contrast with the more common case (i.e., $S_k = 0$ or $D_f = \infty$) where it is well-known that the asymptotic improvement of F_{nh} over F_n is only of second order, in the sense that $\mathrm{MISE}(h_{0n})$ admits the asymptotic representation $n^{-1}\psi(F) - cn^{-p} + o(n^{-p})$ for some $p > 1$ and $c > 0$ (see, e.g., Jones, 1990, and Shao and Xiang, 1997).

2.2. Sinc kernel distribution function estimator. In this section we consider the sinc kernel, defined by $\mathrm{sinc}(x) = \sin(x)/(\pi x)$ for $x \neq 0$ and $\mathrm{sinc}(0) = 1/\pi$. This function is not integrable, so the sinc kernel density estimator inherits this undesirable property, but such a defect can be corrected as described in Glad, Hjort and Ushakov (2003). Nevertheless, the sinc kernel is square integrable, and as such the sinc kernel density estimator achieves certain optimality properties with respect to the MISE (Davis, 1977), that make the sinc kernel useful for density estimation (see Glad, Hjort and Ushakov, 2007, or Tsybakov, 2009, Section 1.3).

Abdous (1993, Section 3) provided a careful study showing that it also makes sense to use the MISE criterion for kernel distribution function estimators based on the integrated sinc kernel. However, it is not so clear from his developments how the sinc kernel distribution function estimator is explicitly defined, nor the asymptotic properties of the optimal bandwidth sequence in this case, since Theorem 2 above can not be directly applied given that the sinc kernel is not integrable. This section contains a detailed treatment of these issues.

First, note that the definition of the integrated kernel $K(x) = \int_{-\infty}^{x} \mathrm{sinc}(z)dz$ has to be understood in the sense of Cauchy principal value, i.e. $K(x) = \lim_{M\to\infty} \int_{-M}^{x} \mathrm{sinc}(z)dz$, because the integral is not Lebesgue-convergent. A simpler way to express such principal value is $K(x) = \frac{1}{2} + \mathrm{Si}(x)$, where $\mathrm{Si}(x) = \int_0^x \mathrm{sinc}(z)dz$ is the sine integral function (with the usual convention that $\int_a^b = -\int_b^a$ if $b < a$). This yields the following explicit form for the sinc kernel distribution function estimator:

$$F_{nh}^{\mathrm{sinc}}(x) = \tfrac{1}{2} + n^{-1} \sum_{j=1}^{n} \mathrm{Si}\big(h^{-1}(x - X_j)\big). \tag{5}$$

An alternative, and perhaps more natural, derivation of (5) is found through the use of inversion formulas. The sinc kernel density estimator with bandwidth $h = 1/T$ is readily obtained from the inversion formula $f(x) = (2\pi)^{-1} \int e^{-itx} \varphi_f(t)dt$ by replacing φ_f with the empirical characteristic function $\varphi_n(t) = n^{-1} \sum_{j=1}^{n} e^{itX_j}$, conveniently truncated to get a finite integral $(2\pi)^{-1} \int_{-T}^{T} e^{-itx} \varphi_n(t)dt$ (see for instance Chiu, 1992, p. 774). An inversion formula relating F and φ_f is the so-called Gil-Pelaez formula $F(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty t^{-1} \Im\{e^{-itx}\varphi_f(t)\}dt$, with $\Im\{z\}$ standing for the imaginary part of a complex number z, which is valid for a continuous F in the principal value sense (Gurland, 1948). Reasoning as before, replacing φ_f with φ_n and restricting the domain of integration to $[0, 1/h]$, results in the same sinc kernel distribution function estimator shown in (5).

As a square integrable function, the Fourier transform of the sinc kernel is the indicator function of the interval $[-1, 1]$. In this sense, Abdous (1993) showed that the IV and ISB formulas of Theorem 1 above remain valid for the sinc kernel distribution estimator, as long as the square integrability of f is added to its assumptions, leading to the following simple exact MISE formula for $h > 0$:

$$\mathrm{MISE}(h) = (n\pi)^{-1} \int_0^{1/h} t^{-2}\{1 - |\varphi_f(t)|^2\}dt + \pi^{-1} \int_{1/h}^{\infty} t^{-2} |\varphi_f(t)|^2 dt. \tag{6}$$

Straightforward differentiation shows that the critical points of such a MISE function are located at any value $h_\diamond$ such that $|\varphi_f(1/h_\diamond)|^2 = (n+1)^{-1}$. This does not reveal, however, if such critical points are local minima or maxima. The following result shows the existence of a global minimizer h_{0n} of (6), and that Theorem 2 above remains valid for the sinc kernel estimator. Note, again, that Theorem 2 of Tenreiro (2006) on the existence of h_{0n} can not be directly applied here because it relies on the assumption that the kernel function is integrable.

Theorem 3. *Assume that f is square integrable and $\int |x| dF(x) < \infty$. Then, there exists a bandwidth h_{0n} that minimizes the MISE of the sinc kernel distribution function estimator. Moreover, if $C_f = D_f$ then $h_{0n} \to 1/D_f$ as $n \to \infty$ and also $h_* = 1/D_f$.*

If it were integrable, the sinc kernel could be considered as a superkernel with $S_{\text{sinc}} = T_{\text{sinc}} = 1$, so from Theorem 3 it follows that all the remarks above about the limit behavior of h_{0n} and the optimal MISE for superkernel distribution function estimators can be equally applied to the sinc kernel distribution estimator.

3. Numerical examples

In this section we present some examples to further illustrate the usefulness and consequences of Theorems 1, 2 and 3 above.

3.1. Example 1. In this example we consider the so-called Jackson-de la Vallé Poussin distribution F, with density function

$$f(x) = \frac{3}{4\pi}\left(\frac{\sin(x/2)}{x/2}\right)^4 = \frac{9 + 3\cos(2x) - 12\cos(x)}{2\pi x^4}$$

and whose characteristic function is shown in Butzer and Nessel (1971, p. 516) to be

$$\varphi_f(t) = \begin{cases} 1 - 3t^2/2 + 3|t|^3/4, & |t| \leq 1 \\ (2 - |t|)^3/4, & 1 \leq |t| \leq 2 \\ 0, & |t| \geq 2 \end{cases},$$

which implies that $C_f = D_f = 2$.

As shown in Theorems 2 and 3, since $C_f = D_f < \infty$ this distribution (or any of its rescalings $F_a(x) = F(x/a)$ with $a > 0$) represents a case where superkernel distribution function estimators are asymptotically more efficient than the empirical distribution function. To illustrate this fact, we include here a numerical comparison using two different superkernels: the sinc kernel and a proper superkernel, the trapezoidal superkernel given

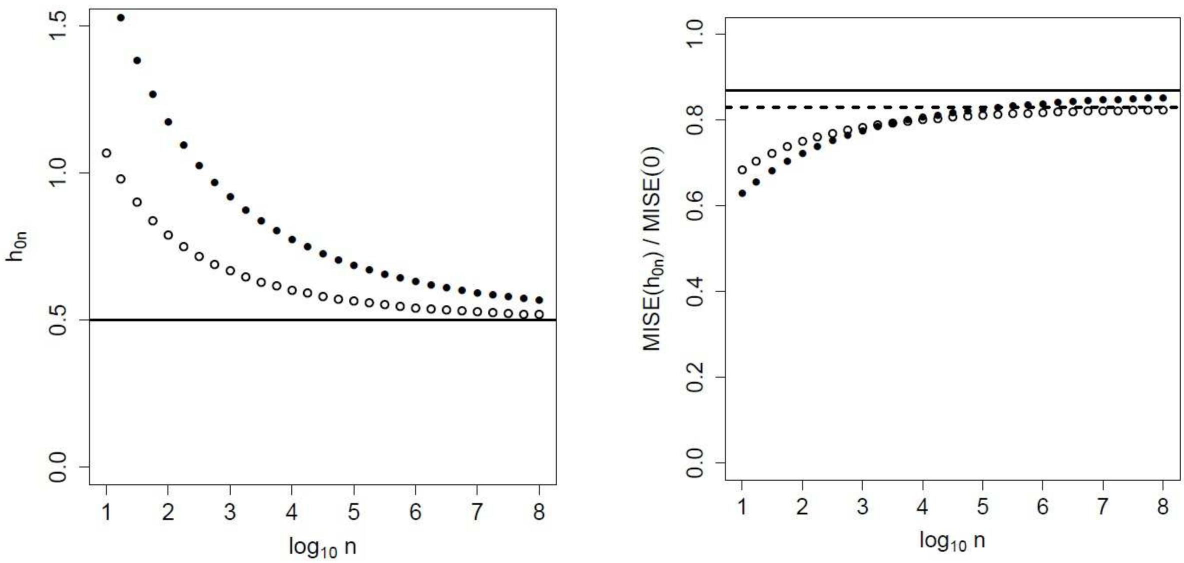

FIGURE 1. Optimal bandwidth sequence (left) and relative efficiency in MISE (right) for the estimation of the Jackson-de la Vallé Poussin distribution, as a function of $\log_{10} n$. The lines show the limit values. Solid circles and solid lines correspond to the trapezoidal superkernel and open circles and dashed lines correspond to the sinc kernel.

by $k(x) = (\pi x^2)^{-1}\{\cos x - \cos(2x)\}$, for which $S_k = T_k = 1$ (see Chacón, Montanero and Nogales, 2007).

It is not hard from Theorem 1 (for the trapezoidal kernel) and (6) (for the sinc kernel) to come up with an explicit formula for the exact MISE function in each case. These exact MISE calculations allow to numerically compute the optimal bandwidth sequences h_{0n} and the minimum MISE values. The optimal bandwidth sequences for both superkernel estimators are shown in Figure 1 (left) as a function of $\log_{10} n$, where it is already noticeable that they both have limit 1/2, as predicted from theory.

The right graph in Figure 1 shows the relative efficiency of both superkernel estimators using optimal bandwidths with respect to the empirical estimator, namely $\text{MISE}(h_{0n})/\text{MISE}(0)$, together with their asymptotic values, given by $\text{MISE}(S_k/D_f)/\text{MISE}(0) = 1 - \psi(K)S_k/\{\psi(F)D_f\}$. Using (4) it follows that $\psi(F) = (96\log 2 - 43)/(8\pi)$, and $\psi(K)$ equals $(4\log 2 - 2)/\pi$ and $1/\pi$ for the trapezoidal and the sinc kernel, respectively, resulting in asymptotic relative efficiencies of approximately 0.87 and 0.83 for the two superkernel estimators, as reflected on Figure 1. For this distribution, the trapezoidal kernel is more efficient than the sinc kernel up to about sample size $n = 3000$, but asymptotically the sinc kernel is slightly more efficient.

Both are markedly more efficient than the empirical distribution as was to be expected from asymptotic theory; besides, the gains are even more substantial for low and moderate sample sizes.

3.2. Example 2. In this second example we make use of the MISE expressions in terms of characteristic functions to obtain exact MISE formulas for the case-study in which F corresponds to the $N(0, \sigma^2)$ distribution and the integrated kernel is either the standard normal distribution function Φ, or the integrated sinc kernel, and we compare both estimators.

For this specific example the exact MISE formula for the density estimation problem was provided in Fryer (1976) making use the convolution properties of the normal density function, which are also useful for deriving many other integral results for the normal density and its derivatives (see Aldershof *et al.*, 1995).

However, convolution techniques seem to be of little use to find exact MISE expressions for kernel distribution function estimators in the normal case, where not even the estimation goal F has an explicit formula. For this problem, it is convenient to work with exact expressions in terms of characteristic functions. For instance, using (4) it immediately follows that the MISE for the empirical distribution function equals $n^{-1}\pi^{-1/2}\sigma$ and, similarly, it is not hard to show that for the kernel estimator with the normal kernel

$$\pi^{1/2}\mathrm{MISE}(h) = n^{-1}\{(h^2+\sigma^2)^{1/2} - h\} + \{(2h^2+4\sigma^2)^{1/2} - (h^2+\sigma^2)^{1/2} - \sigma\}$$

and with the sinc kernel

$$\pi\mathrm{MISE}(h) = (1+n^{-1})\big\{he^{-\sigma^2/h^2} + 2\sigma\sqrt{\pi}\Phi\big(\sigma\sqrt{2}/h\big)\big\} - n^{-1}h - (2+n^{-1})\sigma\sqrt{\pi}.$$

In Figure 2 we show the relative efficiency $\mathrm{MISE}(h_{0n})/\mathrm{MISE}(0)$ as a function of $\log_{10} n$ for $\sigma = 1$ for both kernel estimators with respect to the empirical distribution function. Here, all the three estimators are asymptotically equally efficient, in the sense that the relative efficiency converges to 1 as $n \to \infty$. However, it is clear that this convergence is much slower for the sinc kernel estimator, which is more efficient that the normal kernel estimator for sample sizes as low as $n = 50$.

4. Proofs

For $h > 0$, the statement of Theorem 1 is contained within the proof of Proposition 2 in Abdous (1993). Therefore, it only remains to show the case $h = 0$; i.e., Equation (4). This formula is valid in the more general situation where F is not necessarily a distribution function, but an

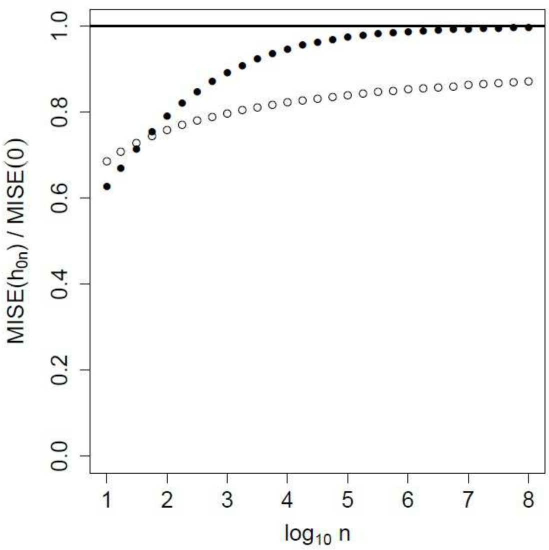

FIGURE 2. Relative efficiency in MISE for the estimation of standard normal distribution, as a function of $\log_{10} n$. The line shows the limit value. Solid circles correspond to the normal kernel and open circles correspond to the sinc kernel.

integrated kernel with finite first order moment, as shown in the following lemma.

Lemma 1. *Suppose that* $K(x) = \int_{-\infty}^{x} k(y)dy$, *where* k *is an integrable function with* $\int k(y)dy = 1$ *and* $\int |yk(y)|dy < \infty$. *Then,*

$$\int K(x)\{1 - K(x)\}dx = (2\pi)^{-1} \int t^{-2}\{1 - |\varphi_k(t)|^2\}dt.$$

Proof: It is not hard to show that

$$K(x)\{1 - K(x)\} = \int \{I_{[y,\infty)}(x) - K(x)\}^2 k(y)dy,$$

where I_A stands for the indicator function of a set A. Moreover, reasoning as in the proof of Proposition 2 in Abdous (1993), it follows that the condition $\int |yk(y)|dy < \infty$ guarantees that $\int |I_{[y,\infty)}(x) - K(x)|dx < \infty$ for all y, which implies that the function $G_y(x) = I_{[y,\infty)}(x) - K(x)$ is square integrable, since K is bounded (because $|K(x)| \leq \int |k(y)|dy$ for all x). Therefore, by Parseval's identity,

$$\int \{I_{[y,\infty)}(x) - K(x)\}^2 dx = (2\pi)^{-1} \int |\varphi_{G_y}(t)|^2 dt.$$

The Fourier transform of G_y is shown to be $(-it)^{-1}\{e^{ity} - \varphi_k(t)\}$, since splitting the integration region and using integration by parts,

$$\begin{aligned}-it\varphi_{G_y}(t) &= it\int_{-\infty}^{y} e^{itx}K(x)dx - it\int_{y}^{\infty} e^{itx}\{1-K(x)\}dx\\ &= K(y)e^{ity} - \int_{-\infty}^{y} e^{itx}k(x)dx + \{1-K(y)\}e^{ity} - \int_{y}^{\infty} e^{itx}k(x)dx\\ &= e^{ity} - \varphi_k(t).\end{aligned}$$

Thus,

$$\int \{I_{[y,\infty)}(x) - K(x)\}^2 dx = (2\pi)^{-1}\int t^{-2}\big[1 + |\varphi_k(t)|^2 - 2\Re\big\{e^{-ity}\varphi_k(t)\big\}\big]dt,$$

where $\Re\{z\}$ denotes the real part of a complex number z. This finally leads to

$$\begin{aligned}&\int K(x)\{1-K(x)\}dx\\ &= \iint \{I_{[y,\infty)}(x) - K(x)\}^2 k(y)dxdy\\ &= (2\pi)^{-1}\iint t^{-2}\big[1 + |\varphi_k(t)|^2 - 2\Re\big\{e^{-ity}\varphi_k(t)\big\}\big]k(y)dydt\\ &= (2\pi)^{-1}\int t^{-2}\{1 - |\varphi_k(t)|^2\}dt,\end{aligned}$$

where the last line follows from the fact that

$$\int e^{-ity}\varphi_k(t)k(y)dy = \varphi_k(-t)\varphi_k(t) = |\varphi_k(t)|^2.$$

■

The proof of Theorem 2 is immediate from the following lemma.

Lemma 2. *Assume that F and K satisfy the assumptions of Theorem 2. Then,*

$$S_k/D_f \le \inf_{n\in\mathbb{N}} h_{0n} \le \limsup_{n\to\infty} h_{0n} \le h_* \le \min\{S_k/C_f, T_k/D_f\}.$$

Proof: First notice that $\mathrm{ISB}(h) = 0$ for all $h \in [0, S_k/D_f]$, since using Theorem 1

$$\begin{aligned}0 \le \pi\,\mathrm{ISB}(h) &= \int_0^\infty t^{-2}|1-\varphi_k(th)|^2|\varphi_f(t)|^2 dt\\ &\le \int_0^{S_k/h} t^{-2}|1-\varphi_k(th)|^2|\varphi_f(t)|^2 dt + \int_{D_f}^\infty t^{-2}|1-\varphi_k(th)|^2|\varphi_f(t)|^2 dt = 0,\end{aligned}$$

with the last equality due to the facts that $\varphi_k(th) = 1$ for $t \in [0, S_k/h]$ and $\varphi_f(t) = 0$ for $t \geq D_f$ by definition of S_k and D_f, respectively.

Therefore, for $h \in [0, S_k/D_f]$ the MISE reduces to the IV, and admits the exact expression $\mathrm{MISE}(h) = n^{-1}\{\psi(F) - \psi(K)h\}$ because, again using Theorem 1, noting the expression (4) for $\psi(F)$, taking into account the definition of S_k and D_f and making the change of variable $s = th$, we obtain

$$\begin{aligned}
\mathrm{IV}(h) &= (n\pi)^{-1} \int_0^\infty t^{-2} |\varphi_k(th)|^2 \{1 - |\varphi_f(t)|^2\} dt \\
&= (n\pi)^{-1} \int_0^\infty t^{-2} \{1 - |\varphi_f(t)|^2\} dt \\
&\quad - (n\pi)^{-1} \int_0^\infty t^{-2} \{1 - |\varphi_k(th)|^2\}\{1 - |\varphi_f(t)|^2\} dt \\
&= n^{-1}\psi(F) - (n\pi)^{-1} \int_{S_k/h}^\infty t^{-2} \{1 - |\varphi_k(th)|^2\} dt \\
&= n^{-1}\psi(F) - n^{-1}\psi(K)h.
\end{aligned}$$

Since the MISE function is linear in h with negative slope in $[0, S_k/D_f]$, its minimum has to be attached at some point greater than S_k/D_f, hence we obtain the first inequality.

On the other hand, reasoning as in Chacón *et al.* (2007) it is possible to show that $\mathrm{ISB}(h) > 0$ for $h > S_k/C_f$ and for $h > T_k/D_f$, thus yielding the last inequality. Finally, denote $h_L = \limsup_{n\to\infty} h_{0n}$ and assume that $h_L > h_*$. Then, the continuity of $\mathrm{ISB}(h)$ with respect to h (Tenreiro, 2006, Proposition 1) entails that there is a subsequence h_{0n_k} such that, as $k \to \infty$, $\mathrm{ISB}(h_{0n_k}) \to \mathrm{ISB}(h_L)$ with $\mathrm{ISB}(h_L) > 0$ since we are assuming $h_L > h_*$. But from (2) and (3) it immediately follows that, for every fixed h, $\mathrm{MISE}_{n_k}(h) \to \mathrm{ISB}(h)$ as $k \to \infty$, so we obtain that the following chain of inequalities

$$\begin{aligned}
\mathrm{ISB}(h) = \lim_{k\to\infty} \mathrm{MISE}_{n_k}(h) &\geq \lim_{k\to\infty} \mathrm{MISE}_{n_k}(h_{0n_k}) \\
&\geq \lim_{k\to\infty} \mathrm{ISB}(h_{0n_k}) = \mathrm{ISB}(h_L) > 0,
\end{aligned}$$

is valid for every fixed h, implying that $\lim_{h\to 0} \mathrm{ISB}(h) \geq \mathrm{ISB}(h_L) > 0$, which contradicts Proposition 1 in Tenreiro (2006), where it is shown that $\mathrm{ISB}(h) \to 0$ as $h \to 0$. Hence, it should be $h_L \leq h_*$, as desired. ■

Finally, we show the proof of Theorem 3. We focus only on the statement about the existence of the optimal bandwidth sequence, since the arguments showing the limit behavior can be adapted from the proof of Lemma 2 above.

Proof of Theorem 3: It is clear from (6) and (4) that $\lim_{h\to 0} \mathrm{MISE}(h) = n^{-1}\psi(F)$. Moreover, $\lim_{h\to\infty} \mathrm{MISE}(h) = \infty$, since $\varphi_f(0) = 1$ and by continuity it is possible to take $\delta > 0$ such that $|\varphi_f(t)|^2 > \frac{1}{2}$ for all $0 \leq t \leq \delta$, so this yields $\int_0^\infty t^{-2}|\varphi_f(t)|^2 dt \geq \frac{1}{2}\int_0^\delta t^{-2}dt = \infty$. These two limit conditions, together with the fact that $\mathrm{MISE}(h)$ is a continuous function, imply that the existence of a minimizer of the MISE is guaranteed if there is some $h_1 > 0$ such that $\mathrm{MISE}(h_1) < n^{-1}\psi(F)$. But from (6) we have

$$\mathrm{MISE}(h) - n^{-1}\psi(F) = -(n\pi)^{-1}h + (1+n^{-1})\pi^{-1}\int_{1/h}^\infty t^{-2}|\varphi_f(t)|^2 dt$$

so that using the Riemann-Lebesgue lemma and the dominated convergence theorem, it follows that

$$\lim_{h\to 0} h^{-1}\{\mathrm{MISE}(h) - n^{-1}\psi(F)\} = -(n\pi)^{-1} < 0,$$

which entails that there is some $h_1 > 0$ fulfilling the aforementioned desired property. ∎

References

Abdous B. (1993) Note on the minimum mean integrated squared error of kernel estimates of a distribution function and its derivates. *Communications in Statistics Theory and Methods*, **22**, 603–609.

Aldershof, B., Marron, J.S., Park, B.U. and Wand, M.P. (1995) Facts about the Gaussian probability density function. *Applicable Analysis*, **59**, 289–306.

Altman, N. and Léger, C. (1995) Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference*, **46**, 195–214.

Azzalini, A. (1981) A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, **68**, 326–328.

Berg, A. and Politis, D. (2009) CDF and survival function estimation with infinite-order kernels. *Electronic Journal of Statistics*, **3**, 1436–1454.

Bowman, A., Hall, P. and Prvan, T. (1998) Bandwidth selection for the smoothing of distribution functions. *Biometrika*, **85**, 799–808.

Butzer, P.L. and Nessel, R.J. (1971) *Fourier analysis and approximation.* Academic Press, New York.

Chacón, J.E., Montanero, J. and Nogales, A.G. (2007) A note on kernel density estimation at a parametric rate. *Journal of Nonparametric Statistics*, **19**, 13–21.

Chacón, J.E., Montanero, J., Nogales, A.G. and Pérez, P. (2007) On the existence an limit behavior of the optimal bandwidth for kernel density estimation. *Statistica Sinica*, **17**, 289–300.

Chacón, J.E. and Rodríguez-Casal, A. (2010) A note on the universal consistency of the kernel distribution function estimator. *Statistics and Probability Letters*, **80**, 1414–1419.

Chiu, S.-T. (1992) An automatic bandwidth selector for kernel density estimation. *Biometrika*, **79**, 771–782.

Cline, D.B.H. (1988) Admissible kernel estimators of a multivariate density. *Annals of Statistics*, **16**, 1421–1427.

Davis, K.B. (1977) Mean integrated square error properties of density estimates. *Annals of Statistics*, **5**, 530–535.

Fryer, M.J. (1976) Some errors associated with the nonparametric estimation of density functions. *IMA Journal of Applied Mathematics*, **18**, 371–380.

Giné, E. and Nickl, R. (2009) An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probability Theory and Related Fields*, **143**, 569–596.

Glad, I.K., Hjort, N.L. and Ushakov, N.G. (2003) Correction of density estimators that are not densities. *Scandinavian Journal of Statistics*, **30**, 415–427.

Glad, I.K., Hjort, N.L. and Ushakov, N.G. (2007) Density estimation using the sinc kernel. Preprint Statistics No. 2/2007, Norwegian University of Science and Technology, Trondheim, Norway. Available at `http://www.math.ntnu.no/preprint/statistics/2007/`

Gurland, J. (1948) Inversion formulae for the distribution of ratios. *Annals of Mathematical Statistics*, **19**, 228–237.

Janssen, P., Swanepoel, J. and Veraverbeke, N. (2007) Modifying the kernel distribution function estimator towards reduced bias. *Statistics*, **41**, 93–103.

Jones, M.C. (1990) The performance of kernel density functions in kernel distribution function estimation. *Statistics and Probability Letters*, **9**, 129–132.

Mason, D.M. and Swanepoel, J.W.H. (2012) Uniform in bandwidth limit laws for kernel distribution function estimators. In *From Probability to Statistics and Back: High-Dimensional Models and Processes*, IMS Collections **9**, 241–253.

Nadaraya, E.A. (1964) Some new estimates for distribution functions. *Theory of Probability and Its Applications*, **15**, 497–500.

Polansky, A.M. and Baker, E.R. (2000) Multistage plug-in bandwidth selection for kernel distribution function estimates. *Journal of Statistical Computation and Simulation*, **65**, 63–80.

Sarda, P. (1993) Smoothing parameter selection for smooth distribution functions. *Journal of Statistical Planning and Inference*, **35**, 65–75.

Servien, R. (2009) Estimation de la fonction de répartition: revue bibliographique. *Journal de la Société Française de Statistique*, **150**, 84–104.

Shao, Y. and Xiang, X. (1997) Some extensions of the asymptotics of a kemel estimator of a distribution function. *Statist. Probab. Lett.*, **34**, 301–308.

Swanepoel, J.W.H. (1988) Mean integrated square error properties and optimal kernels when estimating a distribution function. *Communications in Statistics Theory and Methods*, **17**, 3785–3799.

Swanepoel, J.W.H. and Van Graan, F.C. (2005) A new kernel distribution function estimator based on a non-parametric transformation of the data. *Scandinavian Journal of Statistics*, **32**, 551–562.

Tenreiro, C. (2003) On the asymptotic behavour of the ISE for automatic kernel distribution estimators. *Journal of Nonparametric Statistics*, **15**, 485–504.

Tenreiro, C. (2006) Asymptotic behaviour of multistage plug-in bandwidth selections for kernel distribution function estimators. *Journal of Nonparametric Statistics*, **18**, 101–116.

Tiago de Oliveira, J. (1963) Estatística de densidades: resultados assintóticos. *Revista da Faculdade de Ciências de Lisboa*, **9**, 111–206.

Tsybakov, A.B. (2009) *Introduction to Nonparametric Estimation.* Springer Science+Business Media, New York.

Watson, G.S. and Leadbetter, M.R. (1964) Hazard analysis II. *Sankhy ā Series A*, **26**, 101–116.

Yamato, H. (1973) Uniform convergence of an estimator of a distribution function. *Bulletin of Mathematical Statistics*, **15**, 69–78.

José E. Chacón
Departamento de Matemáticas, Universidad de Extremadura, E-06006 Badajoz, Spain
E-mail address: jechacon@unex.es

Pablo Monfort
Departamento de Matemáticas, Universidad de Extremadura, E-06006 Badajoz, Spain
E-mail address: pabmonf@unex.es

Carlos Tenreiro
CMUC, Department of Mathematics, University of Coimbra, 3001–501 Coimbra, Portugal
E-mail address: tenreiro@mat.uc.pt