Thailand Statistician
April 2022; 20(2): 240-270
http://statassoc.or.th
Contributed paper

Deconvolving Cumulative Density from Associated Random Processes

Mohammed Es-salih Benjrada* **and Khedidja Djaballah**
Department of Probability and Statistics, University of Sciences and Technology Houari Boumediene, Algiers, Algeria.
*Corresponding author; e-mail: esslihm1@gmail.com

Received: 16 December 2019
Revised: 3 October 2020
Accepted: 12 December 2020

Abstract

The main purpose of the present paper is to discuss the problem of estimating the unknown cumulative density function $F(x)$ of X when only corrupted observations $Y = X + \varepsilon$ are present, where X and ε are independent unobservable random variables and ε is a measurement error with a known distribution. For a sequence of strictly stationary and positively associated random variables and assuming that the tail of the characteristic function of ε behaves either as super smooth or ordinary smooth errors, we obtain the precise asymptotic expressions, the bounds on the mean-square estimation error and the asymptotic normality.

Keywords: Deconvolution of cumulative densities, positively associated processes, quadratic-mean convergence, asymptotic normality.

1. Introduction

We consider the problem of estimation from observations that are contaminated by additive noise $\{\varepsilon_i\}_{i=1}^n$. Due to the nature of the experimental environment or the measuring tools, the random process $\{X_i\}_{i=1}^n$ is not available for direct observation. Instead of X_i, we observe the random variables Y_i given by

$$Y_i \triangleq X_i + \varepsilon_i, \;\; i = 1, ..., n. \tag{1}$$

In the present paper, the focus is to estimate nonparametrically the unknown common cumulative density function (c.d.f.) $F(x)$ of a process $\{X_i\}_{i=1}^n$ which is assumed to be strictly stationary and positively associated. In addition, we assume that the density function (p.d.f.) $f(.)$ of the process $\{X_i\}_{i=1}^n$ exists. Furthermore, the noise process $\{\varepsilon_i\}_{i=1}^n$ consists of independent and identically distributed (i.i.d.) random variables, and independent from $\{X_i\}_{i=1}^n$, with known density function $r(.)$. Thus the common probability density function $g(.)$ of the random variables Y_i is given by:

$$g(x) = \int_{-\infty}^{+\infty} f(x-t) r(t) dt. \tag{2}$$

Model (1) is called a convolution and the problem of estimating f with this model occurs in various domains. This model has been studied in Experimental Sciences. For example, Biological

Organisms (see Medgyessy, 1977; Rice and Rosenblatt, 1976); communication theory (see Wise et al., 1977, Snyder et al., 1988) and applied physics (see Jones and Misell, 1967).

The literature abounds of work devoted to the study of the p.d.f. in convolution problems. Zhang (1990) proposed a consistent estimator for the density based on grouped data for some cases of error density. Masry (1993) considered the estimation of the multivariate probability density functions under some structures of dependence. Lejeune and Sarda (1992) used the Moving Polynomial Regression (MPR) to smooth the empirical distribution function estimator. Fan (1990) considered the asymptotic uniform confidence bands.

The c.d.f. deconvolution has not attracted as many research. Ioannides and Papanastassiou (2001) developed the approach to examining the estimation of the c.d.f. and treated its corresponding asymptotic normality in the case where the joint random process $\{X_i, \varepsilon_i\}_{i=1}^{n}$ is stationary and satisfies the ρ-mixing condition and fulfilling some additional assumptions. Furthermore, the contaminated noises $\{\varepsilon_i\}_{i=1}^{n}$ are assumed to have a dependence structure and are either ordinary smooth or super smooth. Dattner et al. (2011) studied the minimax complexity of this problem when the unknown distribution has a density belonging to the Sobolev class and the error density is ordinary smooth. Cordy and Thomas (1997) considered the deconvolution when the unknown distribution is modeled as a mixture of p known distributions. Gaffey and William (1959) studied a consistent estimator of a distribution function from observations contaminated with additive Gaussian errors. Fan (1991) considered the estimate based on integration of the density deconvolution estimator. Wang et al. (2010) developed the estimation of the c.d.f. in the case where data are corrupted by heteroscedastic errors.

We study the quadratic mean convergence and deduce the mean-square convergence rate for the deconvolving cumulative density estimator under various assumptions on the characteristic function ϕ_r of the measurement error. The following two cases are generally distinguished:

- ϕ_r decays algebraically at infinity

$$|t|^{\beta} |\phi_r(t)| \underset{|t|\to+\infty}{\to} \beta_1 \text{ for some } \beta > 0 \text{ and } \beta_1 > 0.$$

  In this case, the error is called ordinary smooth.

- ϕ_r decays exponentially fast at infinity

$$\beta_2 e^{-m|t|^{\alpha}} |t|^{\beta} \leq |\phi_r(t)| \leq \beta_3 e^{-m|t|^{\alpha}} |t|^{\beta},$$

  for some positive constants α , m, real β, and positive constants β_2 and β_3.

This is called supersmooth error.

The parameter β is called the order of the noise density $r(x)$. Actually, it has a direct impact on the rate of convergence of the estimate $F_n(x)$. Particular examples of supersmooth distribution are Normal, Mixture Normal, Cauchy densities $r(x)$. The ordinary smooth distribution covers in particular the case of Gamma, Double Exponential, and Symmetric Gamma densities $r(x)$.

Next, it is of practical interest to show that the deconvolution difficulties are heavily related to the smoothness of the error distribution. Indeed, super smooth distributions are more difficult to deconvolve than ordinary smooth distributions, see for example the proofs in Masry (2003).

The infinite random process $\{X_i\}_{i=1}^{+\infty}$ is positively associated (PA for short), or just associated, if every finite subcollection $\{X_i\}_{i=1}^{n}, n \geq 1$ satisfies the property given in the following definition.

Definition 1 A finite family of random variables $\{X_i\}_{i=1}^{n}$ is said to be positively associated if

$$Cov\left[\Phi_1(X_i, i \in A_1), \Phi_2(X_j, j \in A_2)\right] \geq 0,$$

for every pair of disjoint subsets A_1 and A_2 of $\{1, 2, ..., n\}$, and Φ_l are coordinatewise increasing functions and this covariance exists for $l = 1, 2$.

Definition 1, which was introduced by Esary et al. (1967), includes several mixing process classes. Note that associated processes have attracted a lot of research attention since they arise in a variety of contexts. For instance, in Finance (see (Jiazhu 2002)), and in Applied physics (see Fortuin et al. (1971)), and even in Percolation theory. We may also cite the homogeneous Markov chains as a direct example of the association property and normal random vectors with nonnegative covariance sequences.

It is worthy to note that, if the underlying process $\{X_i\}_{i=1}^n$ is associated, then the process $\{Y_i\}_{i=1}^n$ involving the convolution model in (1) is a corrupted-associated random process. Actually, from Property P2 of Esary et al. (1967) (mentioned later), the independence between the processes $\{X_i\}_{i=1}^n$ and $\{\varepsilon_i\}_{i=1}^n$ ensures the association of the union $\{X_i\}_{i=1}^n \cup \{\varepsilon_i\}_{i=1}^n$. Fortunately, all dealing here is with a strictly stationary process. In fact, and as mentioned above, $\{\varepsilon_i\}_{i=1}^n$ consists of i.i.d. rvs. Since $\{X_i\}_{i=1}^n$ are independent from $\{\varepsilon_i\}_{i=1}^n$, it is clear that $\{Y_i = X_i + \varepsilon_i\}_{i=1}^n$ is a strictly stationary random process. Indeed, the best-known example of a strictly stationary process is the white noise process (i.i.d.). On the other hand, $H_{h_n}(\frac{\cdot}{h_n})$ (defined in (11)) is a measurable mapping, and so we deduce that $\left\{H_{h_n}(\frac{Y_i-x}{h_n})\right\}_{i=1}^n$ is a stationary process in a strict sense.

To the best of our knowledge, there are no papers dealing with the nonparametric estimation of the c.d.f. from corrupted-associated random variables, and this motivates the study in the present work. The layout of this paper is organized as follow. In Section 2, we develop the estimator and give some properties. In Section 3, we establish the mean-square error of our estimate. The asymptotic normality is shown in Section 4. In Section 5, we evaluate the performance of the estimator via simulated data, while Section 6 gives some additional proofs.

2. Estimation

First we denote by $\phi_g(\cdot)$, $\phi_f(\cdot)$, and $\phi_r(\cdot)$ the characteristic functions of $g(\cdot)$, $f(\cdot)$, and $r(\cdot)$ respectively, and let $\hat{\phi}_n(t)$ be the empirical characteristic function of $\{Y_j\}_{j=1}^n$, defined by

$$\hat{\phi}_n(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itY_j} \ . \tag{3}$$

Since the Y_i observations are identically distributed, it follows that $\hat{\phi}_n(t)$ is an unbiased estimator of $\phi_g(t)$

$$E\left(\hat{\phi}_n(t)\right) = \phi_g(t). \tag{4}$$

In practice, the noise density $r(\cdot)$ is usually unknown. In order to reduce the complexity of the statistical analysis, we assume as in all the literature mentioned earlier that $r(\cdot)$ is known, then we choose the kernel $k(\cdot)$ as a bounded even probability density function, and call $\phi_k(\cdot)$ its corresponding Fourier Transform.

As mentioned previously, only Y_i r.v.'s are available to observe. The usual non-parametric method for estimating the probability density for this case is the kernel estimation or the so called Parzen-Rosenblatt method (see Parzen (1962)). This method, which is based on a sample of the statistical population, allows one to estimate the density at any point of the support:

$$g_n(x) = \frac{1}{nh_n}\sum_{j=1}^{n} k(\frac{x-Y_j}{h_n}), \tag{5}$$

where $\{h_n\}_{n\geq 1}$ is a bandwith sequence of positive numbers converging to 0. Hence the cumulative density can be written as:

$$G_n(x) = \frac{1}{n}\sum_{j=1}^{n} K(\frac{x-Y_j}{h_n}), \tag{6}$$

where

$$K(x) = \int_{-\infty}^{x} k(t)dt.$$

To simplify the problem, we assume the independence between $\{X_j\}_{j=1}^{n}$ and $\{\varepsilon_j\}_{j=1}^{n}$, and assume $|\phi_r(t)| > 0$ for all $t \in \mathbb{R}$. Then the convolution equation (2) leads to $\phi_g(t) = \phi_f(t)\phi_r(t)$. On the other hand, if we assume that the chosen kernel k is bounded, even, and in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, then also the kernel type estimate $g_n(\cdot)$ lies in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Hence, its Fourier transform exists, and the calculations show that

$$\phi_{g_n}(t) = \hat{\phi}_n(t)\phi_k(th_n).$$

If we assume that $\dfrac{\phi_{g_n}(t)}{\phi_r(t)} \in L^1(\mathbb{R})$, then by the inverse Fourier Transform, the proposed estimate $\hat{f}_n(x)$ for $f(x)$ is

$$f_n(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-itx)\frac{\hat{\phi}_n(t)\phi_k(th_n)}{\phi_r(t)}dt. \tag{7}$$

We follow the same procedure as in Masry (1991), and using equation (3) we get the following alternative expression for $f(x)$:

$$f_n(x) = \frac{1}{nh_n}\sum_{j=1}^{n} w_{h_n}\left(\frac{Y_j - x}{h_n}\right), \tag{8}$$

where

$$w_{h_n}(s) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{its}\frac{\phi_k(t)}{\phi_r(\frac{t}{h_n})}dt. \tag{9}$$

Remark 1 Notice that the expression of $f_n(x)$ in (8) has a classical kernel form. However, for technical reasons (used in the proofs), we choose to apply the deconvolving kernel $w_{h_n}(.)$ on the term $\left(\frac{Y_j-x}{h_n}\right)$ instead of $\left(\frac{x-Y_j}{h_n}\right)$ which is usually used in kernel estimation. This explains the presence of the term e^{its} instead of e^{-its} in the expression (9).

We consider an estimator of the c.d.f. via integration of the density estimator. This approach was introduced by Zhang (1990). Fan (1991) proved that this type of estimator is minimax optimal for i.i.d. observations and the case of supersmooth noise destribution.

Using the assumptions on $\phi_r(t)$ and $\frac{\phi_{g_n}(t)}{\phi_r(t)}$ used to derive equation (7), we obtain the following deconvolving cumulative density estimator

$$F_n(x) = \frac{1}{n}\sum_{j=1}^{n} H_{h_n}\left(\frac{Y_j - x}{h_n}\right), \tag{10}$$

where

$$H_{h_n}(x) = \int_{-\infty}^{x} w_{h_n}(s)ds. \tag{11}$$

Recall the c.d.f estimator in (10) was suggested by Ioannides and Papanastassiou (2001) to deconvolve the c.d.f for ρ-mixing stochastic processes and the noise process $\{\varepsilon_i\}_{i=1}^{n}$ is assumed to have a dependence structure.

Next, using Fubini's theorem, we obtain

$$H_{h_n}(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{itx}\frac{\gamma_{h_n}(t)}{it}dt, \tag{12}$$

Table 1 $H_{h_n}(x)$ nature changes

$r(x)$	$\gamma_{h_n}(t)/(it)$	$H_{h_n}(x)$
real and even	purely imaginary and odd	real and odd
real and odd	real and even	real and even
purely imaginary and even	real and odd	purely imaginary and odd
purely imaginary and odd	purely imaginary and even	purely imaginary and odd

where

$$\gamma_{h_n}(t) \triangleq \frac{\phi_k(t)}{\phi_r(\frac{t}{h_n})}. \tag{13}$$

In the derivation of equations (12) and (13), we need to have

$$\frac{\phi_k(t)}{t\phi_r(\frac{t}{h_n})} \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R}). \tag{14}$$

For this end, we assume that

$$\phi_k(t) = o(t) \text{ as } t \to 0.$$

Also we note that $\phi_k(t)$ is real-valued and even since $k(x)$ is even. Via a simple calculation, we obtain the results in Table1, which shows the changes in the nature of $H_{h_n}(x)$ with respect to a general function $r(x)$. Hence, we restrict ourselves to the two first lines of Table1 since the function $r(x)$ in our case is the error density.

2.1. Notation and assumptions

2.1.1 Notation

Some notations and reasonable assumptions are needed in what follows.

$$Q_2 \triangleq \frac{1}{2\pi(\beta_1)^2}\int_{-\infty}^{+\infty} |t|^{2(\beta-1)} |\phi_k(t)|^2 dt, \tag{15}$$

where β is the order of the contaminating density $r(x)$, and β_1 is the positive quantity that the noise characteristic function $\phi_r(.)$ decays algebraically at infinity (ie the ordinary smooth case).

$$\chi_{h_n}(x) \triangleq \left[h_n^\beta H_{h_n}(x)\right]^2. \tag{16}$$

2.1.2 Assumptions

A1 The density $k(x)$ satisfies

- i) $\int_{-\infty}^{+\infty} sk(s)ds = 0$.
- ii) $\int_{-\infty}^{+\infty} s^2k(s)ds < \infty$.
- iii) $\phi_k(t) = o(t)$ as $t \to 0$.

A2

- i) $\int_{-\infty}^{+\infty} |t|^{2(\beta-1)} |\phi_k(t)|^2 dt < \infty$.
- ii) $\int_{-\infty}^{+\infty} |t|^{\beta} |\phi_k(t)| dt < \infty$.

- iii) $\int_{-\infty}^{+\infty} |t|^{\beta-1} |\phi_k(t)|^2 \, dt < \infty$.

A3

- $\int_{-\infty}^{+\infty} t^{-(3-j)} h_n^{\beta} \left| \gamma_{h_n}^{(j)}(t) \right| dt < \infty$ for $j = 0, 1, 2$.

A4

- i) $\phi_k(t)$ has a finite support $[-\tau, \tau]$ for some positive constant τ.
- ii) $|\phi_k(t)| \leq a_1(\tau - t)^p$ for $\tau - d \leq t \leq \tau$ for some positive constants p, d, and a_1.
- iii) $\phi_k(t) > a_2(\tau - t)^p$ for $\tau - d \leq t \leq \tau$ and a_2 is a positive constant.
- iv) If we note $R_{\phi_r(t)}(t)$ and $I_{\phi_r(t)}(t)$ the real and imaginary parts of $\phi_r(t)$ respectively, then when $t \to \infty$ either $I_{\phi_r(t)}(t) = o\left(R_{\phi_r(t)}(t)\right)$ or $R_{\phi_r(t)}(t) = o\left(I_{\phi_r(t)}(t)\right)$.

A5

- i) The common univariate probability density $g(x)$ of the observed random process $\{Y_i\}_{i=1}^n$ exists and is bounded for all $x \in \mathbb{R}$.
- ii) The $2-$dimensional density $g_{Y_1,Y_q}(x, y)$ of the random variables Y_1 and Y_q with $q > 0$, exists and is bounded for all $x, y \in \mathbb{R}$.
- iii) The random process $\{X_i\}_{i \geq 1}$ is positively associated. Moreover,

$$\sum_{j=0}^{+\infty} j^{\eta} cov(X_1, X_j) < \infty \text{ for some constant } \eta > 0.$$

Remark 2 To simplify the problem and under the light of conditions A1-A4, we choose to make use a kernel $k(x)$ in which its Fourier transform ϕ_k is compactly supported. Namely, the following kernel $k(x) = 48cos(x)(1 - 15x^{-2})/(\pi x^4) - 144sin(x)(2 - 5x^{-2})/(\pi x^5)$, with the characteristic function $\phi_k(t) = t1_{[-1,1]}(t)$ where $1_B(t)$ stands for the indicator function on a set B. Notice that this kernel is used in the calculation of our simulated estimator in Section 5, since it satisfies the basic assumptions in which our main results are stated.

Under the assumption that $\{\varepsilon_j\}_{j=1}^n$ and $\{X_j\}_{j=1}^n$ are independent and that the $\varepsilon_j^{\prime}s$ are independent among themselves, we can state that $cov(Y_1, Y_j) = cov(X_1, X_j)$. To simplify the notation we write

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} H_{h_n,j}(x),$$

where $H_{h_n,j}(x) = H_{h_n}(\frac{Y_j - x}{h_n})$ and $\upsilon_n \triangleq E(H_{h_n,j})$ for all j and $\tilde{H}_{h_n,j} \triangleq H_{h_n,j} - \upsilon_n$.

As shown before, $\left\{\tilde{H}_{h_n,j}(x)\right\}_{j=1}^n$ is a strictly stationary process. Now, we use Properties (P2 and P4) established by Esary et al. (1967), to show the positive association of $\left\{\tilde{H}_{h_n,j}(x)\right\}_{j=1}^n$.

P2: The union of two independent associated random processes is associated.

P4: Any non-decreasing functions applied on associated random variables are associated (ie non-decreasing functions remain the association).

In fact, the case of i.i.d. random variables represents one extreme of association, thus $\{\varepsilon_i\}_{i=1}^n$ is associated. Further, it is independent from the process $\{X_j\}_{j=1}^n$. Then we note that from P2, the union $\{X_i\}_{i=1}^n \cup \{\varepsilon_i\}_{i=1}^n$ is associated.

From P4 and by choosing $p_j(\{u_i\}_{i=1}^{N} \cup \{v_i\}_{i=1}^{N}) = u_j + v_j$ for any $j = 1, ..., N$, the random process $\{Y_j\}_{j=1}^{n}$ is associated. In our case, $Y_j = p_j(\{X_i\}_{i=1}^{n} \cup \{\varepsilon_i\}_{i=1}^{n})$ and

As reported by Rao (2012), if $\{Z_i\}_{i=1}^{n}$ is a sequence of associated rv's and α_i are positive numbers and $\tau_i \in \mathbb{R}$ for 1$\leq i \leq n$ then the rv's $\frac{Z_i-\alpha_i}{\tau_i}$ are associated. Thus the process $\left\{\frac{Y_j-x}{h_n}\right\}_{j=1}^{n}$ is associated. Since $H_h(x) = \int_{-\infty}^{x} \omega_h(t)dt$ is a nondecreasing function and by P4, the process $\{H_{h_n,i}(x)\}_{i=1}^{n}$ is associated.

The following lemma gives an important transformation used in Section 6.

Lemma 1 *Assume that $m(x)$ and $d(x)$ are two bounded densities and $\phi_m(t)$ and $\phi_d(t)$ are their Fourier transforms respectively, then*

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\phi_m(th)\phi_d(t)dt = \frac{(m_h * d)(x)}{h}, \tag{17}$$

where

$$m_h(x) = m(\frac{x}{h}) \textit{ for any } h \in \mathbb{R}_+^*.$$

To compute the exact asymptotic bias value, we first need to establish an approximation on the identity (see chapter 9 in Wheeden and Zygmund (1977)) and see what conditions should be imposed on the kernel $k(.)$. To this end, we make use of the following lemma due to Bochner (1959).

Lemma 2 (Bochner) *Suppose that $k \in L^1(\mathbb{R})$ is a bounded Borel function on $\mathbb{R}$, then, at every point x of continuity of $g(.)$, we have*

$$\lim_{n\to\infty}\frac{1}{h_n}\int_{-\infty}^{+\infty} k\left(\frac{u}{h_n}\right) g(x-u)\,du = g(x)\int_{-\infty}^{+\infty} k(u)\,du.$$

The density in this expression does not depend on the bandwidth $\{h_n\}_{n\geq 1}$. Thus the kernel $k(.)$ must satisfy the regularity conditions of ordinary density estimation.

Proposition 1 1) For all $x \in \mathbb{R}$, we have

$$\lim_{n\to\infty} E[F_n(x)] = F(x).$$

2) Assume that the kernel $k(.)$ satisfies A1-i) and A1-ii), we also assume that $F \in C_2(\mathbb{R})$ then we have

$$\lim_{n\to\infty}(h_n)^{-2}\, bias[F_n(x)] = \frac{1}{2}F''(x)\int_{-\infty}^{+\infty} s^2k(s)ds.$$

Remark 3 The correction of estimation bias plays a fundamental role in the measurement error model. Another interpretation of the results of Proposition 1 is that the biases of the estimator $F_n(x)$ either in the presence or the absence of the contaminating noise are the same and converge to zero regardless of the error smoothness type.

Notice that conditions A1-A4 have nothing to do with the dependence of the process $\{X_i\}_{i=1}^{+\infty}$. Moreover the result of Proposition 2 is standard and it is valid even for an i.i.d. case.

3. Quadratic-Mean Convergence

This section is divided into two parts. In the first part we assume that the characteristic function $\phi_r(t)$ of the contaminated errors ε decays algebrically at infinity, while in the second part we assume expononcial decay. In both cases we look for exact asymptotic expressions of the mean-square estimation error of $F_n(x)$. To this end we use the following bias-variance decomposition:

$$E\left[(F_n(x)-F(x))^2\right] = var(F_n(x)) + [bias\,(F_n(x))]^2. \tag{18}$$

For the expononcial decay case, we only provide tight bounds because the precise asymptotic expression of $\|H_{h_n}\|_2$ is not available in this case.

3.1. Ordinary smooth noise distribution

Let us first assume that $\phi_r(t)$ fulfills the following assumption:
Assumption B1 :
i) $|\phi_r(t)| > 0$ for all $t \in \mathbb{R}$,
ii) $|t|^\beta |\phi_r(t)| \to \beta_1$ as $|t| \to +\infty$, for some positive constants β and β_1.

The exact asymptotic bias given in Section 2 relied primarily on the identity approximation for the classical kernel-type density estimation with the help of lemma 2. For the asymptotic variance, an extra approximation for which the underlying function $\left(H^2_{h_n}(.)\right)$ depends on the bandwidth h_n is required. This may cause difficulties since the condition $(|t\phi_r(t)| > 0)$ is not necessarily met.

Generally, $H_{h_n}(.)$ is non-negative and strictly monotone. Under Assumption B1, Parsevals theorem and conditions A1-iii) and A2-i) ensure that $H_{h_n}(.)$ is always in $L^2(\mathbb{R})$. The next proposition gives a precise L_2-norm asymptotic expression for $H_{h_n}(.)$ and its proof is relegated to the last section.

Proposition 2 If we suppose that condition B1 is satisfied and

1) If conditions A1-iii) and A2-i) hold then

$$\lim_{n\to+\infty} h_n^{2\beta} \int_{-\infty}^{+\infty} |H_{h_n}(s)|^2 \, ds = Q_2 \ ,$$

where Q_2 defined in (15).

2) If A2-i) holds, we have

$$h_n^\beta \left\| H'_{h_n} \right\|_\infty \le C < \infty,$$

where H'_{h_n} is the first derivative of H_{h_n}.

3) If A3 holds, then we have

$$h_n^\beta \int_{-\infty}^{+\infty} |H_{h_n}(s)| \, ds \le C < \infty.$$

The following lemma gives an approximation of the identity under the smoothness assumptions on the characteristic functions $\phi_k(t)$ and $\phi_r(t)$. This lemma is needed to obtain the precise asymptotic variance of the estimator $F_n(x)$.

Lemma 3 *Under conditions B1, A1-iii), A2-i), A3 and A5-i), then*

$$\int_{-\infty}^{+\infty} \frac{1}{h_n} \chi_{h_n}\left(\frac{x-u}{h_n}\right) g(u) du \to g(x) Q_2.$$

at all points x of continuity of g, and the quantity Q_2 is defined in (15).

The next Lemma proposed by Birkel (1988) is crucial for what follows.

Lemma 4 (Birkel) *If we suppose that $\{Y_j\}_{j\in I}$ is a finite random process of positively associated random variables and let A and B be subsets of I. Let $\Phi_1(.)$ and $\Phi_2(.)$ be bounded first order partial derivatives, then we have*

$$|cov\left[\Phi_1(Y_i, i \in A), \Phi_2(Y_j, j \in B)\right]| \le \sum_{i\in A}\sum_{j\in B} \left\| \frac{\partial \Phi_1}{\partial t_i} \right\|_\infty \cdot \left\| \frac{\partial \Phi_2}{\partial t_j} \right\|_\infty cov(Y_i, Y_j).$$

where $\left\| \frac{\partial \Phi_1}{\partial t_j} \right\|_\infty = \max\left\{ \left\| \frac{\partial^+ \Phi_i}{\partial t_j} \right\|_\infty, \left\| \frac{\partial^- \Phi_i}{\partial t_j} \right\|_\infty \right\}.$

Now, we are ready to treat the quadratic-mean convergence of $F_n(x)$.

Theorem 1 *Under conditions B1, A1-A3, and A5 we have*

$$\lim_{n\to\infty} nh_n^{2\beta-1} var(F_n(x)) = \sigma^2(x) \text{ at the points } x \text{ of continuity of } g,$$

where $\sigma^2(x) = Q_2 g(x)$.

For quadratic-mean convergence rates, we have the next corollary.

Corollary 1 Combining the bias given in Proposition 1 with the asymptotic variance found in Theorem 1 we find the quadratic-mean convergence of the c.d.f estimator. Next, by selecting an optimal value of the bandwidth parameter, $h_n \simeq n^{-1/(2\beta+3)}$, i.e. that minimizes this asymptotic mean-square error. Then, we have a mean-square convergence rate of:

$$E\left|F_n(x) - F(x)\right|^2 = O(n^{-2/(2\beta+3)}),$$

in the absence of contaminating noise and taking $h_n \simeq n^{-1/3}$, the mean square convergence rate is

$$E\left|F_n(x) - F(x)\right|^2 = O(n^{-2/3}).$$

Note that the presence of contaminating noise reduces the mean-square convergence rate of $F_n(x)$ by a factor that depends on the rate of decay of the tail characteristic function $\phi_r(t)$ of ε.

3.2. Supersmooth noise distribution

We now consider the quadratic mean convergence of $F_n(x)$ when the charateristic function $\phi_r(t)$ of the noise processes $\{\varepsilon_i\}_{i=1}^n$ has an exponential decay as $t \to \infty$ namely the super smooth case, and in particular when the following assumptions are met

Assumption B2

i) $|\phi_r(t)| > 0$ for all $t \in \mathbb{R}$,

ii) $\beta_2 e^{-m|t|^\alpha} |t|^\beta \leq |\phi_r(t)| \leq \beta_3 e^{-m|t|^\alpha} |t|^\beta$for some β real, and positive constants α, m, β_2 and β_3.

Super smooth errors are much harder to deconvolve than ordinary smooth errors, this may be due to the fact that impossible to find a simple expression and the exact order for the function $H_{h_n}(x)$ (even in i.i.d. case). As a consequence, the precise asymptotic rates and constants of $var(F_n(x))$ and $var(H_{h_n,1}(x))$ can not be obtained in this case.

We first derive a lower bound for $var(H_{h_n,1}(x))$ and then use it to establish the following asymptotic relationship

$$var(F_n(x)) = \frac{1}{n} var(H_{h_n,1}(x))(1 + o(1)).$$

Lemma 5 *1) Under conditions B2, A1-ii), A4-i), and A4-ii), we have*

$$\|H_{h_n}\|_\infty = O\left(h_n^{(p+1)\alpha+\beta}(\log(\frac{1}{h_n}))^p \exp(m(\frac{\tau}{h_n})^\alpha)\right),$$

$$\|H_{h_n}\|_2 = O\left(h_n^{(p+1/2)\alpha+\beta}(\log(\frac{1}{h_n}))^p \exp(m(\frac{\tau}{h_n})^\alpha)\right),$$

and

$$\left\|H'_{h_n}\right\|_\infty = O\left(h_n^{(p+1)\alpha+\beta}(\log(\frac{1}{h_n}))^p \exp(a(\frac{d}{h_n})^\beta)\right).$$

2) In addition if A4-iii) and A4-iv) hold, we have

$$|H_{h_n}(u)| \geq C_2\,|G(x)|\,h_n^{(p+1)\alpha+\beta}\exp(m(\frac{\tau}{h_n})^{\alpha})),$$

for a constant C_2 *and* $G(x) = \cos(\tau x)1\left(I_{\phi(\frac{t}{h_n})} = o(R_{\phi(\frac{t}{h_n})})\right) + \sin(\tau x)1\left(R_{\phi(\frac{t}{h_n})} = o(I_{\phi(\frac{t}{h_n})})\right)$, *where 1(.) is an indicator function.*

Lemma 6 *Under conditions B2 and A4, and as* $n \to \infty$

$$var(H_{h_n,1}(x)) \geq C_3 h_n^{2((p+1)\alpha+\beta+1/2)}\exp(2m(\frac{\tau}{h_n})^{\alpha}),$$

and

$$var(H_{h_n,1}(x)) \leq C_4 h_n^{2((p+1)\alpha+\beta-1)}\left(\log(\frac{1}{h_n})\right)^{2p}\exp(2m(\frac{\tau}{h_n})^{\alpha}),$$

for a positive constants C_3 *and* C_4.

Theorem 2 *Under conditions B2, A4, and A5 we have*

$$var(F_n(x)) = \frac{1}{n}var(H_{h_n,1}(x))(1+o(1)).$$

4. Asymptotic Normality

To discuss the asymptotic normality of the cumulative density estimate $F_n(x)$ in Eq. (10), we recall that $\{H_{n,j}(x)\}_{j=1}^{+\infty}$ is a positively associated random process as well as a strictly stationary sequence, since it involves monotonic transformations $H_{h_n}(.)$ of positively associated r.v.'s. Thus, following the same approach as used in Oliveira (2012), we will show that:

$$\frac{F_n(x) - E\left[F_n(x)\right]}{\sqrt{var\left[F_n(x)\right]}} \xrightarrow{L} N(0,1). \tag{19}$$

It was shown in Section 3 that, under suitable smoothness assumptions for the ordinary smooth case, we have

$$\lim_{n\to\infty} nh_n^{2\beta-1}var(F_n(x)) = \sigma^2(x).$$

Hence, the asymptotic distribution in (19) becomes

$$n^{1/2}h_n^{\beta-1/2}\left[F_n(x) - E\left[F_n(x)\right]\right] \xrightarrow{L} N(0,\sigma^2(x)).$$

On the other hand, when the noise characteristic function ϕ_r decays exponentially fast, it is found that:

$$var(F_n(x) = \frac{1}{n}var(H_{h_n,1})(x))(1+o(1)).$$

As already shown, the main problem with this type of measurement errors that it is difficult (or impossible) to find the limit of $E\left[H_{h_n,1}^2(x)\right]$ and the corresponding convergence rate. As a consequence, the central limit theorem (CLT) is not available in this case. Instead, we will prove that:

$$\sqrt{n}\frac{F_n(x) - E\left[F_n(x)\right]}{\sqrt{var(H_{h_n,1}(x))}} \xrightarrow{L} N(0,1).$$

It should be mentioned that, in the ordinary smooth case, the CLT for associated random variables is available only for a random process which is strictly stationary sequence and not a weakly stationary one.

Theorem 3 *1) Under conditions B1, A1-A3, and A5 we have*

$$n^{1/2} h_n^{\beta-1/2} \left[F_n(x) - E\left[F_n(x) \right] \right] \xrightarrow{L} N(0, \sigma^2(x)). \tag{20}$$

2)Under Assumption B2 and condition A4 we have that

$$\sqrt{n} \frac{F_n(x) - E\left[F_n(x) \right]}{\sqrt{var(H_{h_n,1}(x))}} \xrightarrow{L} N(0, 1), \tag{21}$$

as $n \to \infty$.

More precisely, the next corollary gives a better centering.

Corollary 2 If we consider that all assumptions provided in Theorem 3 hold in addition to $F(t) \in C_2(\mathbb{R})$, we have

1) For the ordinary smooth case:

$$n^{1/2} h_n^{\beta-1/2} \left[F_n(x) - F(x) \right] \xrightarrow{L} N(0, \sigma^2(x)).$$

2) For the supersmooth case:

$$\sqrt{n} \frac{F_n(x) - F(x)}{\sqrt{var(H_{h_n,1}(x))}} \xrightarrow{L} N(0, 1).$$

Remark 4 In the ordinary smooth case, we estimate the asymptotic variance $\sigma^2(x)$ by a plug-in-type estimator defined by

$$\sigma_n^2(x) = g_n(x) Q_2,$$

where $g_n(x)$ is a kernel estimator of $g(x)$ drawn from a sample of size n of Y_i. Thus, in the light of the results of the above Corollary, and for an asymptotic level $1 - \alpha$, we can establish an asymptotic confidence interval of $F(x)$ given by:

$$\left[F_n(x) - z_{1-\alpha/2} \sigma_n^2(x) (n h_n^{2\beta-1})^{-0.5}, F_n(x) + z_{1-\alpha/2} \sigma_n^2(x) (n h_n^{2\beta-1})^{-0.5} \right]$$

where $z_{1-\alpha/2}$ presents a quantile of order $1 - \alpha/2$ of $N(0, 1)$.

5. Numerical Experiments

This section is divided into two parts (subsections). In the first part, we mention an example of a convolution model and compare the performance of our simulated estimator from both direct and contaminated observations. This is for the goal of displaying the influence of ignoring the measurement errors. In the second part, we examine the behavior over finite samples of our conducted estimator via simulation experiments.

5.1. Example and comparisons

Convolution examples are extensive. But due to space restriction, we only mention one interesting example here which is in Communication Theory and more specifically in Signal Processing. Let Y_i stand for the voice heard when the $i-$th individual speaks and X_i be its pure voice (real) and ε_i some noise. In this example, we recognize two different cases:

(i) **direct communication:** where the audible sound is the spoken sound itself ($Y_i = X_i$)

(ii) **communication through the phone:** where ($Y_i = X_i + \varepsilon_i$), ε_i being here some perturbation due to a confusion of the phone-network.

For more details on the impact of the presence of measurement error in signal processing, we refer the interested reader to monographs by Mendelsohn and Rice (1982).

Our goal now is to compare the performance of the cumulative density estimators under different models and show the influence of ignoring the measurement errors. For that purpose, we conduct a simulated example in which the observations are contaminated by homoscedastic errors. We generate a pure random process $\{X_i\}_{i=1}^{1000}$ from an exponential distribution with parameter $\lambda = 0.25$ and the measurement errors from $N(0, 2)$. We consider the following two situations:

Presence's Impact: The goal in this situation is to examine the performance of the general estimation. For that purpose, we compare the classical kernel-type estimator (see (6)) under direct observations ($Y_i = X_i$) and the deconvolution estimator under convolution model (1). This will allow us to explain the effect of measurement errors in nonparametric estimation. The results obtained are displayed in Figure 1.

Ignoring's Impact: Consider the convolution model (1). In this situation, we compare the performance of the classical kernel c.d.f. estimator (neglecting the measurement errors) with the deconvolution-type estimator (which takes into account the noise). This is done in order to explain the impact of neglecting measurement errors.

From the outcome of our simulations, it is possible to see that the performance of estimating under the convolution model is very much inferior to that under the error-free model, ie no contaminated observations are present. The curves in Figure 2 reveal that the deconvolving estimator outperforms the usual kernel method. This is due to the fact that the latter procedure neglects the measurement errors and thus gives a biased estimator and may lead to wrong results.

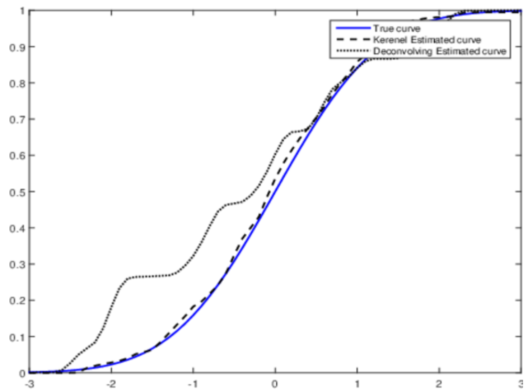


Figure 1 The blue line represents the true distribution function, the dashed line corresponds to the kernel estimates from uncorrupted observations, the dotted line to the deconvolving estimate

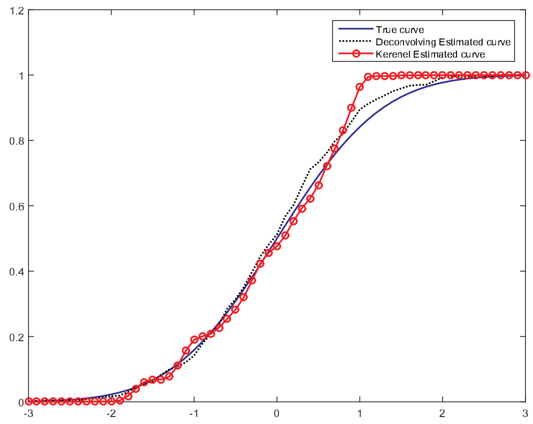


Figure 2 The blue line corresponds to the true distribution function, the red circled line to the kernel estimates from corrupted observations, the dotted line to the deconvolving estimate

5.2. Simulation study

In this subsection, we conducted the simulations using different sample sizes to quantify the performance of our estimator via the Global Mean Square Error $(GMSE)$ criterion computed using N Monte Carlo trials as below

$$GMSE(h) = \frac{1}{Nq} \sum_{j=1}^{N} \sum_{l=1}^{q} \left[F_{n,j}(x_l) - F(x_l)\right]^2,$$

where $F_{n,j}(x_l)$ is the estimated value of $F(x_l)$ at the j-th iteration, and q is the number of equidistant points x_l belonging to a given set.

Remark 5 The numerical implementations in this section illustrate the strong and weak points of the deconvolving approach related to the target c.d.f. and the measurement error classes.

5.2.1 Description and models

The next set-ups are used to generate our numerical experiments. We provide elaborated results correspond to four distinct target c.d.f.'s $F(X)$ which are supposed to come from the following models:

Unimodal Distributions:

- $X \rightsquigarrow N(0,1)$,
- $X \rightsquigarrow \chi^2(3)$.

Bimodal Distributions: We chose here the target c.d.f.'s to have distinctly separated modes.

- $X \rightsquigarrow 0.6N(-2,1) + 0.4N(3,1)$,
- $X \rightsquigarrow 0.5Gamma(4) + 0.5Gamma(14)$, where $Gamma(m)$ stands for $\Gamma(\alpha, \beta)$ with shape parameter $\alpha = m$ and scale parameter $\beta = 1$.

The target distributions have been considered as they satisfy a special features that can be found in practice, and they present an increasing order of deconvolving difficulty.

We control the variance of the errors $var(\varepsilon)$ to have particular values for the so-called noise to signal ratio (NSR) $\sigma_\varepsilon/\sigma_X$, where σ_X and σ_ε stand for the standard deviations of X and ε respectively. Particularly, we choose to have NSR= 0.1, 0.2 and 0.5 which is equivalent to 10%, 20% and 50% contaminating errors, respectively. Thus, for a better comparison, we define the next two error distribution scenarios:

1) **Normal distribution** $N(m, \sigma_\varepsilon^2)$ with $m = 0$ and $\sigma_\varepsilon = 1/10, 1/5, 1/2$

2) **Laplace distribution** $L(\mu, \sigma_\varepsilon^2)$ with $\mu = 0$ and $\sigma_\varepsilon = 2^{-0.5}(1/10, 1/5, 1/2)$

These have been chosen because they belong to the ordinary smooth and super smooth classes respectively. Thus, we have 8 combinations of convolution models. As shown through simulations in subsection 1, deconvolution recovers slowly the target distribution, thus we need large sample sizes in order to have an estimator that works well. For this end, we used sample sizes $n = 200$, 500 and 1000, and $N = 500$ replications for each model. Consequently, we have 24 different simulation set-ups.

Typically, measurement errors are supposed to have zero expectation; however, several cases violate this assumption. Nevertheless, numerous papers considered cases with non-zero expectation as a measurement error model. The general idea here is based on relocating these distributions to have zero expectation.

To simulate convolution sequences after a positive association, we generate the data as follows:

- Simulate $(n+1)$ iid rv's W_i from the distribution of the desired c.d.f.
- Simulate n iid rv's ε_i from the distribution of the considered errors.
- $Y_i = X_i + \varepsilon_i$ for $i = 1, ..., n$, where $X_i = (W_{i-1} + W_{i-2})/2$.

Generally, X_i are positively associated r.v.'s. and have the same distribution as that of W_i. At the end of this procedure, we selected the bandwidth h_n in the grid of values in the set $\Theta = \{1/n^{1/k}, k = 1, ..., 10\}$. Finally, the estimator $F_n(x)$ is calculated by varying x in the grid of points in $\Lambda = \{x \in [-8 : 0.01 : 8]\}$.

5.2.2 Simulation results

In this part, detailed results (Tables 2 and 3, and Figures 3-6) are presented for the purpose of illustrating the influence of the sample size, the distribution of the errors and the NSR rate on the performance quality of our estimator.

We report only on the results obtained using the standard normal unimodal (Table 2) and Gaussian mixture of bimodal (Table 3) cases, since they are similar to that of the other settings. Table 2 and 3 summarize the optimal global bandwidth $h_{opt} = \arg\min_{h\in\Theta} GMSE(h)$ together with its corresponding $GMES = \min_{h\in\Theta} GMSE(h)$ for the different experimental scenarios mentioned above. For each model, we display the estimation when the distribution of the errors is either Gaussian or Laplacian.

In Table 2, we see that $GMSE$ and h_{opt} values change in the same direction. First, note that our estimators work quite well irrespective the different scenarios. As one could expect, the deconvolution referring to a Gaussian noise provide very poor quality of estimation. In particular, the convergence rate in this situation is only $O\left((\ln n)^{-c}\right)$ for a positive constant c. Conversely, when the error is a Laplace distributed, we see a substantially improved performance.

Furthermore, we note however that as the sample size raised, the performance quality increase and becomes quite well whatever the error distributions. By contrast, the deconvolving estimation from small contamination (10%) provides much better quality and become slightly deteriorates as the contamination level increases, but still improved along with n increase.

For bimodal target c.d.f.'s, Table 3 illustrates the difficulty of recovering a distribution of two modes. We see that the effects of sample sizes and contamination levels on the performance quality

Table 2 Simulation results for unimodal distributions (standard normal): GMSE with its corresponding optimal bandwidth

Error distributions	NSR n	10%		25%		50%	
		$GMSE$	h_{opt}	$GMSE$	h_{opt}	$GMSE$	h_{opt}
Normal	200	2.04×10^{-2}	0.5157	3.11×10^{-2}	0.5179	4.11×10^{-2}	0.8441
	500	1.44×10^{-2}	0.4495	2.05×10^{-2}	0.4537	3.39×10^{-2}	0.5179
	1000	1.08×10^{-2}	0.3976	1.64×10^{-2}	0.4097	1.83×10^{-2}	0.4599
Laplace	200	1.35×10^{-2}	0.5174	1.89×10^{-2}	0.5157	2.45×10^{-2}	0.5303
	500	1.05×10^{-2}	0.4599	1.13×10^{-2}	0.4578	1.80×10^{-2}	0.3589
	1000	0.64×10^{-2}	0.4217	0.89×10^{-2}	0.4217	1.34×10^{-2}	0.4217

Table 3 Simulation results for bimodal distributions (Gaussian mixture): GMSE with its corresponding optimal bandwidth

Error distributions	NSR n	10%		25%		50%	
		$GMSE$	h_{opt}	$GMSE$	h_{opt}	$GMSE$	h_{opt}
Normal	200	8.33×10^{-2}	0.1303	10.5×10^{-2}	0.1243	11.8×10^{-2}	0.0447
	500	8.04×10^{-2}	0.0794	8.24×10^{-2}	0.0791	8.27×10^{-2}	0.0240
	1000	7.89×10^{-2}	0.0467	7.91×10^{-2}	0.0538	8.22×10^{-2}	0.0790
Laplace	200	1.64×10^{-2}	0.4902	2.79×10^{-2}	0.5157	3.98×10^{-2}	0.5257
	500	0.86×10^{-2}	0.4596	1.99×10^{-2}	0.4639	2.98×10^{-2}	0.4549
	1000	0.33×10^{-2}	0.4293	1.27×10^{-2}	0.4217	2.09×10^{-2}	0.4013

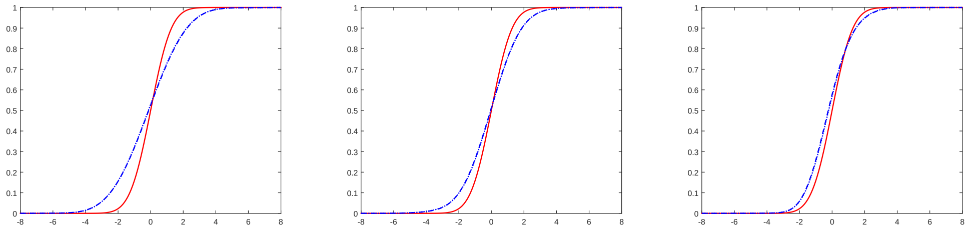
are similar to the unimodal setting. That being said, the overall performances have deteriorated compared to the unimodal case albeit less so for the Laplace errors.

Figures 3-6 display the quality of fit of our estimators, relatively to the results gathered from previous Tables. In each Figure, we plot side by side the c.d.f. estimates from different scenarios. For comparison purposes, we also plot the true X distributions, which we present on these Figures by a solid line.

From Figure 3, note that in terms of the rate of convergence, our estimator performs well under small contamination (10%) and has the same quality regardless of the distribution errors. This is essentially due to the small variance of the errors. Consequently, the estimation is comparable to a usual kernel approach when the errors are not considered.

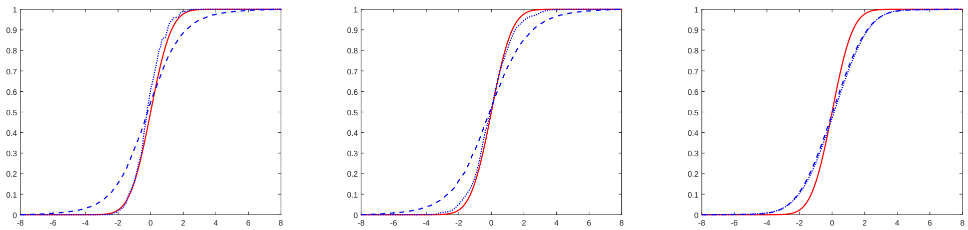
As it widely known, the deconvolution with supersmooth errors provides estimators with very poor convergence rates compared to the ordinary smooth errors. Thus for a reasonable contamination rate at 25%, the estimation from Laplace errors gives better results compared to Gaussian errors as shown in Figure 4.

Summarizing the simulation results from the previous Tables and Figures. The rate of convergence depends directly on the smoothness of the distribution of the errors: the smoother the distribution is, the slower rate of convergence will be. On the other hand the larger the sample size and the lower the NSR, the better the quality of performance will be. Furthermore, the quality of fit declines substantially from unimodal to bimodal distributions but it increases with a sufficiently smaller NSR value and a higher sample size.



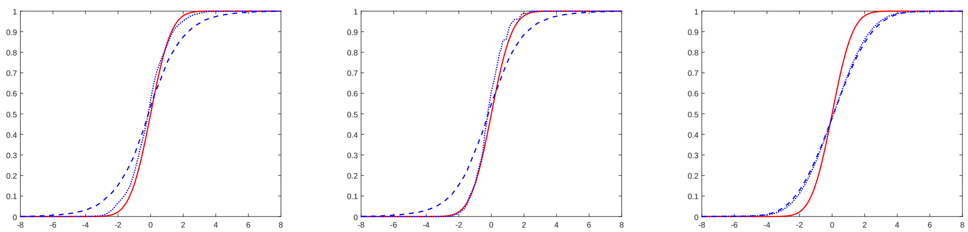
(a) $n = 200$ **(b)** $n = 500$ **(c)** $n = 1000$

Figure 3 Estimation of the normal c.d.f. for $n = 200$, 500 and 1000, and contaminating errors with $NSR = 10\%$. The solid line corresponds to the true distribution function, the dashed line to deconvolving cdf estimator for Normal errors, the dotted line to deconvolving cdf estimator for Laplace errors



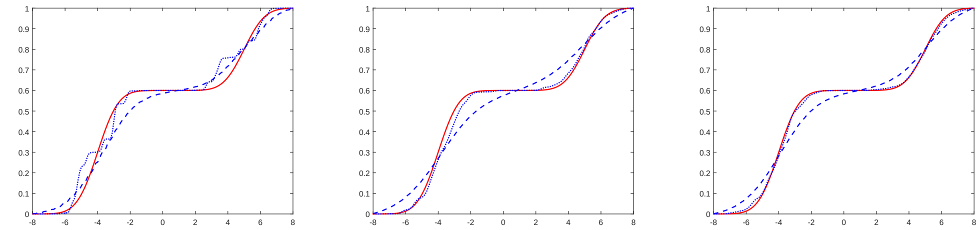
(a) $n = 200$ **(b)** $n = 500$ **(c)** $n = 1000$

Figure 4 Estimation of the normal c.d.f. for $n = 200$, 500 and 1000, and contaminating errors with $NSR = 25\%$. The solid line corresponds to the true distribution function, the dashed line to deconvolving cdf estimator for Normal errors, the dotted line to deconvolving cdf estimator for Laplace errors



(a) $n = 200$ **(b)** $n = 500$ **(c)** $n = 1000$

Figure 5 Estimation of the normal c.d.f. for $n = 200$, 500 and 1000, and contaminating errors with $NSR = 50\%$. The solid line corresponds to the true distribution function, the dashed line to deconvolving cdf estimator for Normal errors, the dotted line to deconvolving cdf estimator for Laplace errors

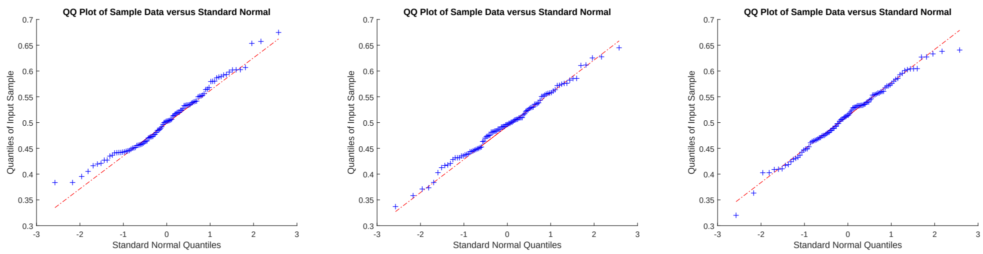


(a) $n = 200$ **(b)** $n = 500$ **(c)** $n = 1000$

Figure 6 Estimation of the mixed normal c.d.f. for $n = 200$, 500 and 1000, and contaminating errors with $NSR = 25\%$. The solid line corresponds to the true distribution function, the dashed line corresponds to deconvolving cdf estimator for Normal errors, the dotted line corresponds to deconvolving cdf estimator for Laplace errors

5.2.3 Asymptotic normality and confidence intervals

In this part, we study the asymptotic normality of the c.d.f. estimator through normal-probability plots. For this goal, we only examine the unimodal distribution from a Laplacian errors case. The deconvolving c.d.f estimation was implemented here for $NSR = 25\%$, $N = 1000$ replications, and $n = 200$, 500 and 1000. This NSR value was preferred since it gives a distinct and reasonable performance for different distribution of the errors, encountered in previous simulations. The results of this practical implementation are summarized on Figure 7.



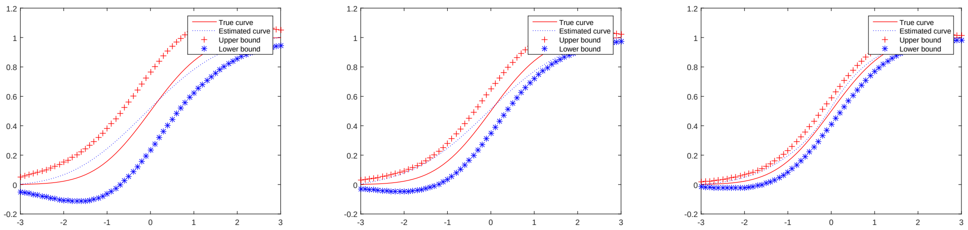
(a) $n = 200$ **(b)** $n = 500$ **(c)** $n = 1000$

Figure 7 The normal-probability plots of the gaussian c.d.f. estimator for $n = 200$, 500 and 1000, and contaminating errors with $NSR = 25\%$.

From Figure 7, we see again that for the asymptotic normality, the estimator provides good performance for a Laplacian error distribution with a chosen NSR rate. Fortunately, these optimistic results become more visible for a large sample size. This indicates that the impact of the NSR on the convergence on distribution becomes fast and faster along with $n \to \infty$.

To present the results of confidence intervals of our estimator $F_n(x)$ for different values of x in Λ, we propose simulated examples implementaing the results of Remark 4 in Section 4. For this end, we give a 95% confidence interval for the c.d.f. estimator when the distribution of the errors is Laplacian and we consider $n = 200$, 500 and 1000 for comparison purposes.

We first note that in this case, we have $\beta_1 = \frac{1}{\sigma_\varepsilon^2}$ and $\beta = 2$. The simulations here are done for $NSR = 25\%$ and their corresponding optimal global bandwidths h_{opt} found in Table 2 and $x = 0.5$. The results are displayed in Figure 8.



(a) $n = 200$ **(b)** $n = 500$ **(c)** $n = 1000$

Figure 8 The 95% confidence intervals of the gaussian c.d.f. for $n = 200$, 500 and 1000, and contaminating errors with $NSR = 25\%$. The solid line corresponds to the true distribution function, the dashed line to deconvolving cdf estimator, the $++$ line to upper bound, the $**$ line to lower bound.

6. Proofs

Proof: [Proof of Lemma 1]

Let $q(x) = (m_h * d)(x)$ be the convolution between the defined functions $m_h(.)$ and $d(.)$. It is easily found that $\phi_{m_h}(t) = h\phi_m(th)$. Thus, applying the forward Fourier Transform gives the next equality

$$\begin{aligned} \phi_q(t) &= \phi_{m_h}(t)\phi_d(t) \\ &= h\phi_m(th)\phi_d(t). \end{aligned}$$

On the one hand, we have

$$\begin{aligned} q(x) &= (m_h * d)(x) \\ &= \int_{-\infty}^{+\infty} d(x-u)m_h(u)du \\ &= \int_{-\infty}^{+\infty} d(x-u)m(\frac{u}{h})du. \end{aligned}$$

On the other hand, we get

$$\begin{aligned} q(x) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-itx)\phi_q(t)dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-itx)\phi_{m_h}(t)\phi_d(t)dt \\ &= \frac{h}{2\pi}\int_{-\infty}^{+\infty} \exp(-itx)\phi_m(th)\phi_d(t)dt. \end{aligned}$$

The desired conclusion is obtained by identification.

Proof: [Proof of Proposition 1] We start by proving the point 1. By Fubini's theorem

$$\begin{aligned} E\left[F_n(x)\right] &= \int_{-\infty}^{x} E\left[\frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-its)\frac{\phi_k(th_n)}{\phi_r(t)}\hat{\phi}_n(t)dt\right] ds \\ &= \int_{-\infty}^{x} \left[\frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-its)\frac{\phi_k(th_n)}{\phi_r(t)} E\left[\hat{\phi}_n(t)\right] dt\right] ds\,. \end{aligned}$$

It is clear from (4) that $E\left[\hat{\phi}_n(t)\right] = \phi_f(t).\phi_r(t)$, and by Lemma 1, it follows that

$$\begin{aligned} E\left[F_n(x)\right] &= \int_{-\infty}^{x}\left[\frac{1}{2\pi}\int_{-\infty}^{+\infty}\exp(-its)\phi_k(th_n)\phi_f(t)dt\right]ds \\ &= \int_{-\infty}^{x}\left[\frac{1}{h_n}\int_{-\infty}^{+\infty}k(\frac{u}{h_n})f(s-u)du\right]ds\ . \end{aligned}$$

Again by Fubini's theorem

$$E\left[F_n(x)\right] = \int_{-\infty}^{+\infty}\frac{1}{h_n}k(\frac{u}{h_n})F(x-u)du\ .$$

The first part is obtained by Bochner's identity in Lemma 2.
Now, from the last expression, and considering the variable change $y = \frac{u}{h_n}$, we can see that

$$E\left[F_n(x)\right] = \int_{-\infty}^{+\infty}k(y)F(x-yh_n)dy\ .$$

Then, we write $F(x-yh_n)$ as a Taylor series

$$F(x-yh_n) = F(x) - yh_nF'(x) + \frac{1}{2}y^2h_n^2F''(x) + o(h_n^2).$$

Point 2 follows directly by using the conditions A1-i) and A1-ii), and the fact that $F(t) \in C_2(\mathbb{R})$.

Proof: [Proof of Proposition 2]
By Parseval's theorem we have

$$h_n^{2\beta}\int_{-\infty}^{+\infty}\left|H_{h_n}(s)\right|^2ds = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left|\varphi_{h_n}(t)\right|^2dt,$$

where

$$\varphi_{h_n}(t) = \frac{t^{\beta-1}\phi_k(t)}{\left(\frac{t}{h_n}\right)^{\beta}\phi_r(\frac{t}{h_n})}.$$

Indeed, $\frac{t}{h_n} \to +\infty$ if and only if $n \to +\infty$ for some $|t| > 0$ or $t \to \infty$ for some $h_n > 0$.
By condition B1-ii) we have

$$\lim_{n\to\infty}\left|\varphi_{h_n}(t)\right|^2 = \frac{1}{\beta_1^2}\left|t\right|^{2(\beta-1)}\left|\phi_k(t)\right|^2.$$

Furthermore, condition B1 means that for a large R we have

$$\left|t\right|^{\beta}\left|\phi_r(t)\right| > \frac{\beta_1}{2} \text{ for any } t \geq R.$$

Thus, we have

$$\left|\varphi_{h_n}(t)\right|^2 \leq \left|\frac{\phi_k(t)}{t}\right|^2\frac{1}{\left|\phi_r(\frac{t}{h_n})\right|^2}1_{[|t|<h_nR]} + \frac{4}{\beta_1^2}\left|t\right|^2\left|\phi_k(t)\right|^2 1_{[|t|\geq h_nR]}.$$

By conditions A1-iii) and B1 it might be clear to see

$$\left|\varphi_{h_n}(t)\right|^2 \leq \frac{4}{\beta_1^2}\left|t\right|^2\left|\phi_k(t)\right|^2 1_{[|t|\geq h_nR]}.$$

Finally, the first conclusion follows by dominated convergence. For part 2), we use the fact that

$$\left\|H'_{h_n}\right\|_\infty = \|\omega_{h_n}\|_\infty .$$

Hence, the second assertion follows using similar arguments as that used for Lemma 4-b) in Masry (2003). Thus, we omit the details.

Now we deal with the last conclusion, for this we put $H_{h_n}(x) = \int\limits_{-\infty}^{+\infty} \frac{1}{2\pi} \exp(itx)\, \ell_{h_n}(t)\, dt$ where $\ell_{h_n}(t) \triangleq \frac{\gamma_{h_n}(t)}{it}$. Then, if we suppose that $\ell_{h_n} \in C_2(\mathbb{R}) \cap L^1(\mathbb{R})$ then $H_{h_n} \in L_1(\mathbb{R})$. In fact, we repeat integration by parts twice to find

$$-x^2 H_{h_n}(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} \exp(itx)\, \ell_{h_n}^{(2)}(t)\, dt,$$

where $\ell_{h_n}^{(2)}(t) = \frac{2}{it^3}\gamma_{h_n}^{(2)}(t) - \frac{2}{it^2}\gamma'_{h_n}(t) + \frac{1}{it}\gamma_{h_n}(t)$. Thus, condition A3 assure that

$$\frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} \left| h_n^\beta \ell_{h_n}^{(2)}(t) \right| dt \le C < \infty.$$

Then, we deduce that

$$h_n^\beta \left| H_{h_n}(x) \right| \le \frac{C}{x^2}. \tag{22}$$

Hence, the upper bound on the L_1- norm of $h_n^\beta H_{h_n}(x)$ follows immediately by applying Riemann's Theorem (with $\alpha = 2$).

Proof: [Proof of Lemma 3] Write the quantity $g(x)Q_2$ as a limit

$$\begin{aligned} g(x)Q_2 &= \lim_{n\to+\infty} \int_{-\infty}^{+\infty} \left[h_n^\beta H_{h_n}(t)\right]^2 g(x)dt \\ &= \lim_{n\to+\infty} \int_{-\infty}^{+\infty} \chi_{h_n}(t) g(x)dt \\ &= \lim_{n\to+\infty} \int_{-\infty}^{+\infty} \frac{1}{h_n}\chi_{h_n}(\frac{t}{h_n}) g(x)dt. \end{aligned}$$

It remains to show that

$$\lim_{n\to\infty} \int_{-\infty}^{+\infty} \frac{1}{h_n}\chi_h\left(\frac{t}{h_n}\right)\left[g(x-t) - g(x)\right] dt = 0.$$

Then, we split appropriately the interval of integration as follow:

$$\begin{aligned} \int_{-\infty}^{+\infty} \frac{1}{h_n}\chi_h(\frac{t}{h_n})\left[g(x-t) - g(x)\right] dt &= \int_{|t|>\theta} \frac{1}{h_n}\chi_{h_n}(\frac{t}{h_n})\left[g(x-t) - g(x)\right] dt \\ &\quad + \int_{|t|\le\theta} \frac{1}{h_n}\chi_{h_n}(\frac{t}{h_n})\left[g(x-t) - g(x)\right] dt, \end{aligned}$$

Since x is a point of continuity of $g(.)$, then for every $b > 0$ and fixed positive θ, we have $|g(x-t) - g(x)| \le b$ for any $-\theta \le t \le \theta$. Thus, we can see that

$$\int_{|t|\le\theta} \frac{1}{h_n}\chi_{h_n}(\frac{t}{h_n})\left[g(x-t) - g(x)\right] dt \le b \int_{-\infty}^{+\infty} \chi_{h_n}(t)dt.$$

For any chosen value of b, the fixed constant θ remains positive. Hence, letting b go to zero, leads to $\int_{|t|\leq\theta}\frac{1}{h_n}\chi_{h_n}(\frac{t}{h_n})\,[g(x-t)-g(x)]\,dt=0$.

Next, by condition A5-i), we get $|g(x-t)-g(x)|\leq 2c$. Thus

$$\int_{|t|>\theta}\frac{1}{h_n}\chi_{h_n}(\frac{t}{h_n})\,[g(x-t)-g(x)]\,dt\leq 2c\int_{|t|>\frac{\theta}{h_n}}\chi_{h_n}(t)dt.$$

In the proof of the third part in Proposition 2, we find $|H_{h_n}(t)|\leq\frac{C}{t^2}$. This means that $\chi_{h_n}(t)\leq C\frac{h_n^{2\beta}}{t^4}$. Then, using the fact $\frac{\theta}{h_n}\to+\infty$ as $n\to\infty$, we can conclude that

$$\begin{aligned}\int_{|t|>\theta}\frac{1}{h_n}\chi_{h_n}(\frac{t}{h_n})\,[g(x-t)-g(x)]\,dt &\leq Ch_n^{2\beta}\int_{|t|>\frac{\theta}{h_n}}\frac{1}{t^4}dt\\ &= Ch_n^{2\beta}\left\{\left.\frac{1}{t^3}\right|_{|t|>\frac{\theta}{h_n}}\right\}\to 0 \text{ as } n\to\infty.\end{aligned}$$

Proof: [Proof of Theorem 1]

As seen before, $\{H_{h_n,j}(x)\}_{j=1}^{n}$is a strictly stationary random process for all $n>1$, thus

$$var(F_n(x))=\frac{1}{n}I_{n,1}+\frac{2}{n}\sum_{j=2}^{n}(1-\frac{j}{n})I_{n,j},$$

where $I_{n,j}=cov(H_{h_n,1}(x),H_{h_n,j}(x))$. We first need to demonstrate the following results:

- $\lim_{n\to\infty}h_n^{2\beta-1}I_{n,1}=\sigma^2(x)$,
- $h_n^{2\beta-1}\sum_{j=2}^{n}|I_{n,j}|=o(1)$, for a large n.

Firstly, we deal with the first point. On the one hand, we have

$$E\left(H_{h_n}\left(x\right)\right)=E\left(F_n\left(x\right)\right),$$

since the data are identically distributed. On the other hand, using the results of Proposition 1-2), we find

$$E\left(H_{h_n}\left(x\right)\right)=O\left(h_n^2\right).$$

Thus

$$I_{n,1}=\int_{-\infty}^{\infty}H_{h_n}^2(\frac{u-x}{h_n})g(u)du+O(h_n^4).$$

Then, Lemma 3 leads to

$$\begin{aligned}h_n^{2\beta-1}I_{n,1} &= \int_{-\infty}^{\infty}\frac{1}{h_n}\left[h_n^{\beta}H_{h_n}(\frac{u-x}{h_n})\right]^2g(u)du+O(h_n^{2\beta+3})\\ &= \int_{-\infty}^{\infty}\frac{1}{h_n}\chi_{h_n}(\frac{u-x}{h_n})g(u)du+O(h_n^{2\beta+3})\to\sigma^2(x).\end{aligned}$$

For the second part and for more simplicity, we consider the next decomposition

$$\sum_{j=2}^{n}|I_{n,j}|=\sum_{j=2}^{\theta_n}|I_{n,j}|+\sum_{j=\theta_n+1}^{n}|I_{n,j}|,$$

where $\theta_n \to \infty$ and $\theta_n h_n \to 0$ as $n \to \infty$. For $2 \le j \le \theta_n$ we have

$$I_{n,j} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} H_{h_n}(\frac{t-x}{h_n})H_{h_n}(\frac{u-x}{h_n})\left[g_{Y_1,Y_j}(t,u) - g(t)g(u)\right] dtdu.$$

Conditions A5-i), A5-ii) and Proposition 2-3) show that

$$\begin{aligned} |I_{n,j}| &\le Ch_n^2 \int_{-\infty}^{+\infty} |H_{h_n}(u)|\, du \int_{-\infty}^{+\infty} |H_{h_n}(t)|\, dt \\ &= Ch_n^2 \left(\int_{-\infty}^{+\infty} |H_{h_n}(u)|\, du\right)^2 \\ &= O(h_n^{2-2\beta}). \end{aligned}$$

Thus, uniformly for $2 \le j \le \theta_n$ we can conclude that

$$\begin{aligned} h_n^{2\beta-1}\sum_{j=2}^{\theta_n} |I_{n,j}| &= O(\theta_n h_n) \\ &= o(1). \end{aligned}$$

Next, we consider the contribution when $\theta_n + 1 \le j \le n$. Actually the process $\{H_{h_n,j}(x)\}_{j=1}^{+\infty}$ is positively associated, and its covariance sequence is obtained by using Lemma 4. In our case $\Phi_1(.)$ and $\Phi_2(.)$ are identical and equal to $H_{h_n}(\frac{.-x}{h_n})$, and the subsets A and B consist of a single random variable 1 and j respectively, thus

$$|cov(H_{h_n,1}(x), H_{h_n,j}(x))| \le \left\|H'_{h_n,1}\right\|_\infty^2 cov(Y_1, Y_j).$$

From the definition of $H_{h_n,1}(x)$, we can see that

$$\left\|H'_{h_n,1}\right\|_\infty = \frac{1}{h_n}\left\|H'_{h_n}\right\|_\infty.$$

Next, by Proposition 2, and the fact that $\{X_j\}_{j=1}^n$and $\{\varepsilon_j\}_{j=1}^n$ are independent, we can see

$$|cov(H_{h_n,1}(x), H_{h_n,j}(x))| \le \frac{C}{h_n^{2\beta+2}} cov(X_1, X_j).$$

We see evidently that $\left(\frac{j}{\theta_n}\right)^\eta \ge 1$ for all $j \in [\theta_n + 1, n]$ and a positive constant η. Then, it follows that

$$h_n^{2\beta-1}\sum_{j=\theta_n+1}^{n} |I_{n,j}| \le \frac{C}{h_n^3\theta_n^\eta}\sum_{j=1}^{n} j^\eta cov(X_1, X_j).$$

From condition A5-iii), and by choosing $\theta_n = h_n^{\frac{-\alpha}{\eta}}$ for some $\alpha > 3$, we have

$$h_n^{2\beta-1}\sum_{j=\theta_n+1}^{n} |I_{n,j}| \to 0 \text{ as } n \to \infty.$$

Proof: [Proof of Lemma 5] Most arguments and procedures here are inspired by the proof of Lemma 3.1 in Fan and Masry (1992). For a positive constant c we consider

$$\delta_n = ch_n^\alpha \log(\frac{1}{h_n}). \tag{23}$$

By symmetry and condition A4-i)

$$\begin{aligned}\|H_{h_n}\|_\infty &\leq \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{|\phi_k(t)|}{|t|\left|\phi_r(\frac{t}{h_n})\right|}dt\\ &= \left(\int_0^{\tau-\delta_n}+\int_{\tau-\delta_n}^{\tau}\right)\frac{|\phi_k(t)|}{|t|\left|\phi_r(\frac{t}{h_n})\right|}dt\\ &= L_1+L_2.\end{aligned}$$

First, we deal with L_1. For this we choose D as a large enough but fixed number. Then,

$$\begin{aligned}L_1 &= \left(\int_0^{Dh_n}+\int_{Dh_n}^{\tau-\delta_n}\right)\left|\frac{\phi_k(t)}{t}\right|.\left|\frac{1}{\phi_r(\frac{t}{h_n})}\right|dt\\ &= L_{1,1}+L_{1,2}.\end{aligned}$$

Firstly, we use condition A1-iii) to get

$$\begin{aligned}L_{1,1} &= \int_0^{Dh_n}\left|\frac{\phi_k(t)}{t}\right|.\left|\frac{1}{\phi_r(\frac{t}{h_n})}\right|dt\\ &\leq \frac{1}{\min_{0\leq u\leq D}\phi_r(u)}\int_0^{Dh_n}\left|\frac{\phi_k(t)}{t}\right|.dt\\ &= o(1)\,. \end{aligned} \tag{24}$$

Concerning $L_{1,2}$, condition B2 with $\beta_2=\beta_3$ leads to

$$\begin{aligned}L_{1,2} &= \int_{Dh_n}^{\tau-\delta_n}|\phi_k(t)|\,.\left|\frac{1}{t\phi_r(\frac{t}{h_n})}\right|dt\\ &\leq C\int_{Dh_n}^{\tau-\delta_n}\left|\frac{1}{t}\right|\left(\frac{t}{h_n}\right)^{-\beta}\exp(m(\frac{t}{h_n})^\alpha)dt\\ &= Ch_n^\beta\int_{Dh_n}^{\tau-\delta_n}|t|^{-(\beta+1)}\exp(m(\frac{t}{h_n})^\alpha)dt.\end{aligned}$$

Next, by considering the derivative of the function $W\,(t)$=$|t|^{-(\beta+1)}\exp(m(\frac{t}{h_n})^\alpha)$ with respect to t, we can clearly see that the integrand in the latter inequality is an increasing function in the interval $Dh_n\leq t\leq\tau-\delta_n$. Hence, it achieves its top when $t=\tau-\delta_n$. This together with (24) give the following:

$$\begin{aligned}L_1 &\leq C\,(\tau-\delta_n)^{-(\beta+1)}\,h_n^\beta\exp(m(\frac{\tau-\delta_n}{h_n})^\alpha).\\ &= O(h_n^\beta\exp(m(\frac{\tau}{h_n})^\alpha(1-\frac{\delta_n}{\tau})^\alpha)).\end{aligned}$$

By applying a Taylor expansion, we have $(1-\frac{\delta_n}{\tau})^\alpha = 1-\alpha\frac{\delta_n}{\tau}+O\left(\delta_n^2\right)$. Thus, a simple calculation using the quantity of δ_n defined in (23) gives

$$L_1 = O(h_n^{\beta+mc\alpha\tau^{\alpha-1}}\exp(m(\frac{\tau}{h_n})^\alpha)).$$

Next, we consider L_2. We note that $(\tau - t) \leq \delta_n$ for any $\tau - \delta_n \leq t \leq \tau$, then by condition A4-ii) we find $|\phi_k(t)| \leq a_1 \delta_n^p$. Thus

$$L_2 \leq c h_n^{\beta} a_1 \delta_n^p \int_{\tau-\delta_n}^{\tau} t^{-(\beta+1)} \exp(m(\frac{t}{h_n})^{\alpha}) dt.$$

With the same argument, we finish this bound by founding

$$L_2 = O(h_n^{\beta+\alpha(1+p)} (\log(\frac{1}{h_n}))^p \exp(m(\frac{\tau}{h_n})^{\alpha}).$$

We note that L_1 is dominated by L_2. Hence, we obtain the first conclusion.

Next, symmetry and Parseval's theorem lead to

$$\|H_{h_n}\|_2^2 = \frac{1}{\pi} \int_0^{\tau} \frac{|\phi_k(t)|^2}{|t|^2 \left|\phi_r(\frac{t}{h_n})\right|^2} dt.$$

Then, the second conclusion is attained by a similar treatment. Concerning the third assertion, we can see that

$$\left\|H'_{h_n}(u)\right\|_{\infty} = \|\omega_{h_n}(u)\|_{\infty}.$$

The proceedings here are identical to those used for the first point in Lemma 3.1 of Fan and Masry (1992). Thus, we omit the details. Now, we deal the last conclusion. We write

$$\begin{aligned} H_{h_n}(x) &= \frac{1}{2\pi} \left(\int_{-\tau}^{-(\tau-\delta_n)} + \int_{-(\tau-\delta_n)}^{\tau-\delta_n} + \int_{\tau-\delta_n}^{\tau} \right) \exp(itx) \frac{\phi_k(t)}{it\phi_r(\frac{t}{h_n})} dt \\ &= J_1 + J_2 + J_3. \end{aligned}$$

Notice that

$$|J_2| \leq O(L_1) = O(h_n^{\beta + mc\alpha\tau^{\alpha-1}} \exp(m(\frac{\tau}{h_n})^{\alpha})).$$

It is clear that $\exp(itx) = \cos(tx) + i\sin(tx)$ and $\phi_r(\frac{t}{h_n}) = R_{\phi_r}(\frac{t}{h_n}) + iI_{\phi_r}(\frac{t}{h_n})$. Note that under condition A4 -iv), we have $\phi_r(\frac{t}{h_n}) = R_{\phi_r}(\frac{t}{h_n})(1 + io(1))$ or $\phi_r(\frac{t}{h_n}) = I_{\phi_r}(\frac{t}{h_n})(o(1) + i)$. As a consequence, $R_{\phi_r}(\frac{t}{h_n})$ and $I_{\phi_r}(\frac{t}{h_n})$ can't change their signs, otherwise it will be a contradiction with Assumption B2. By symmetry we find

$$J_1 + J_3 = \frac{1}{\pi} \int_{\tau-\delta_n}^{\tau} \frac{i\phi_k(t)}{t\left|\phi_r(\frac{t}{h_n})\right|^2} \left[\sin(tx) R_{\phi_r}(\frac{t}{h_n}) - \cos(tx) I_{\phi_r}(\frac{t}{h_n})\right] dt. \tag{25}$$

We only treat the case where $R_{\phi_r}(\frac{t}{h_n}) = o\left(I_{\phi_r}(\frac{t}{h_n})\right)$. The case where $I_{\phi_r}(\frac{t}{h_n}) = o\left(R_{\phi_r}(\frac{t}{h_n})\right)$ can be obtained by the same steps. Now, from the fact that $\tan(0) = 0$ we can deduce that $sin(x) = o(cos(x))$ for $x \to 0$. Thus, (25) becomes

$$J_1 + J_3 = \frac{1}{\pi} \int_{\tau-\delta_n}^{\tau} \frac{i\phi_k(t)}{t\left|\phi_r(\frac{t}{h_n})\right|^2} \cos(tx) I_{\phi_r}(\frac{t}{h_n}) (o(1) - 1) dt.$$

If $x = \frac{(q+0.5)\pi}{\tau}$ and $q \in \mathbb{N}$, then the situation becomes evident since $cos(\tau x) = 0$. If we consider $x \neq \frac{(q+0.5)\pi}{\tau}$, then

$$\begin{aligned} J_1 + J_3 &= \frac{1}{\pi} \left(\int_{\tau-\delta_n}^{\tau-h_n^{\alpha}} + \int_{\tau-h_n^{\alpha}}^{\tau} \right) \frac{i\phi_k(t)}{t\left|\phi_r(\frac{t}{h_n})\right|^2} \cos(tx) I_{\phi_r}(\frac{t}{h_n}) (o(1) - 1) dt \qquad (26) \\ &= J_{4,1} + J_{4,2}. \end{aligned}$$

The function $\cos(tx)$ has the same sign in $[\tau-\delta_n,\tau]$. Moreover it can be written as

$$\cos(tx)=\cos(\tau x)+o(\cos(\tau x)),$$

Consequently $J_{4,1}$ and $J_{4,2}$ have the same sign. Then (26) leads to

$$|J_1+J_3|\geq |J_{4,2}|.$$

By Assumption B2 and condition A4-iii)

$$|J_1+J_3|\geq a_2h_n^{\beta}\left|\cos(\tau x)(1+o(1))\right|\int_{\tau-h_n^{\alpha}}^{\tau}(\tau-t)^p t^{-(\beta+1)}\exp(m(\frac{t}{h_n})^{\alpha})dt.$$

It might be clear to see that $t^{(\beta+1)}\exp(m(\frac{t}{h_n})^{\alpha}$ is an increasing function. Then

$$|J_1+J_3|\geq a_2h_n^{\beta}\left|\cos(\tau x)(1+o(1))\right|(\tau-h_n^{\alpha})^{-(\beta+1)}\exp(m\left(\frac{\tau}{h_n}\right)^{\alpha}(1-\frac{h_n^{\alpha}}{\tau}))\int_{\tau-h_n^{\alpha}}^{\tau}(\tau-t)^p dt. \tag{27}$$

By simple calculation we find

$$\int_{\tau-h_n^{\alpha}}^{\tau}(\tau-t)^p dt=\frac{1}{p}h_n^{\alpha(p+1)},$$

and

$$(1-\frac{h_n^{\alpha}}{\tau})\leq 1.$$

Then, (27) becomes

$$|J_1+J_3|\geq Ch_n^{\beta+\alpha(p+1)}\left|\cos(\tau x)\right|\exp(m(\frac{\tau}{h_n})^{\alpha}).$$

We finish this conclusion by choosing c as a large enough number to be J_2 dominated by J_1+J_3.

Proof: [Proof of Lemma 6]
By Proposition 1, we see that

$$\begin{aligned}var(H_{h_n,1}(x)) &= E\left(H_{h_n}^2(\frac{Y_1-x}{h_n})\right)+O(1)\\ &= h_n\int_{-\infty}^{+\infty}H_{h_n}^2(u)g(uh_n+x)du+O(1).\end{aligned}$$

By condition A4-iii), A4-iv) and the result of Lemma 5-2), we can write

$$var(H_{h_n,1}(x))\geq C_2^2h_n^{2((p+1)\alpha+\beta+1/2)}\exp(2m(\frac{\tau}{h_n})^{\alpha})\int_{-\infty}^{\infty}g(uh_n+x)\left|G(u)\right|^2du+O(1).$$

Recall that $h_n\to 0$ as $n\to\infty$. This together with fact that $g\left(.\right)$ is a continuous function lead to

$$\begin{aligned}var(H_{h_n,1}(x)) &\geq C_3h_n^{2((p+1)\alpha+\beta+1/2)}\exp(2m(\frac{\tau}{h_n})^{\alpha})g(x)\int_{-1}^{1}\left|G(u)\right|^2du\left(1+o\left(1\right)\right)\\ &\geq C_3h_n^{2((p+1)\alpha+\beta+1/2)}\exp(2m(\frac{\tau}{h_n})^{\alpha}).\end{aligned}$$

Thus the first part is obtained. The second part is obtained by using the upper bound on the L_{∞}-norm found in Lemma 5.

Proof: [Proof of Theorem 2] The stationarity of the process $\{H_{h_n,j}(x)\}_{j=1}^n$ implies that

$$var(F_n(x)) = \frac{1}{n}I_{n,1} + \frac{2}{n}\sum_{j=2}^{n}(1-\frac{j}{n})I_{n,j},$$

with $I_{n,j} = cov(\ H_{h_n,1}(x), H_{h_n,j}(x))$. Thus the main task in this proof is to show that $\sum_{j=2}^n I_{n,j} = o(I_{n,1})$, or by equivalent way: $\frac{1}{var(H_{h_n,1}(x))}\sum_{j=2}^n I_{n,j} = o(1)$. To this end, we write

$$\sum_{j=2}^{n} I_{n,j} = \sum_{j=2}^{\rho_n} I_{n,j} + \sum_{j=\rho_n+1}^{n} I_{n,j}, \tag{28}$$

where $\rho_n \to \infty$ as $n \to \infty$. Next, we set

$$\begin{aligned} I_{n,j} &= E(H_{h_n}(\frac{Y_1-x}{h_n})H_{h_n}(\frac{Y_j-x}{h_n})) + O(1) \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} H_{h_n}(\frac{s-x}{h_n})H_{h_n}(\frac{t-x}{h_n})g_{Y_1,Y_j}(t,s)dtds + O(1). \end{aligned}$$

Condition A5 leads to

$$\begin{aligned} |I_{n,j}| &\leq Ch_n^2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} H_{h_n}(t)H_{h_n}(t)dtds \\ &\leq Ch_n^2\left\|H_{h_n}\right\|_1^2 \\ &= O(h_n^2). \end{aligned}$$

Now, select $\rho_n = \exp(m(\frac{\tau}{h_n})^\alpha)$. Then, by using the bounds on $var(H_{h_n,j}(x))$ established in Lemma 6, we get

$$\begin{aligned} \frac{1}{var(H_{h_n,1}(x))}\sum_{j=2}^{\rho_n}|I_{n,j}| &\lesssim h_n^{-2((p+1)\alpha+\beta+1/2)}\exp(-2m(\frac{\tau}{h_n})^\alpha)\rho_n h_n^2 \\ &= O(h_n^{-2((p+1)\alpha+\beta-1/2)}\exp(-m(\frac{\tau}{h_n})^\alpha)). \end{aligned} \tag{29}$$

Consider now the second contribution of $I_{n,j}$. Note that the choice of ρ_n is the same and $\left(\frac{j}{\rho_n}\right)^\eta \geq 1$ for any j in $[\rho_n+1, n]$. Next, from the third point of Lemma 5-1) we can deduce that that:

$$\left\|H'_{h_n}(\frac{\cdot}{h_n})\right\|_\infty = O(h_n^{(p+1)\alpha+\beta-1}(\log(\frac{1}{h_n}))^p\exp(m(\frac{\tau}{h_n})^\alpha)).$$

Now, we use Lemma 4 to get

$$|I_{n,j}| \leq Ch_n^{2((p+1)\alpha+\beta-1)}(\log(\frac{1}{h_n}))^{2p}\exp(2m(\frac{\tau}{h_n})^\alpha)cov(Y_1,Y_j).$$

As mentioned early, $cov(Y_1,Y_j) = cov(X_1,X_j)$, and using the lower bound established in Lemma 6, we find

$$\begin{aligned} \frac{1}{var(H_{h_n,1}(x))}\sum_{j=\rho_n+1}^{n}|I_{n,j}| &\leq Ch_n^{-3}(\log(\frac{1}{h_n}))^{2p}\sum_{j=\rho_n+1}^{n}cov(X_1,X_j) \\ &\leq Ch_n^{-3}\exp(-\eta m(\frac{\tau}{h_n})^\alpha)(\log(\frac{1}{h_n}))^{2p}\sum_{j=1}^{n}j^\eta cov(X_1,X_j) \to 0. \end{aligned}$$

Then, the final result follows from (28) and (29). The proof of Theorem 2 is finish.

The next Lemma which was introduced by Newman and Wright (1981), is considered as a key element for the CLT when estimating under associated concepts. Thus, it is essential in the proof of Theorem 3.

Lemma 7 (Newman) Let $X_1, X_2, ..., X_N$ be a sequence of positively associated random variables, then for all $(t_1, ..., t_N) \in \mathbb{R}^N$ we have

$$\left|\phi_{(X_1,X_2,...,X_N)}(t) - \prod_{j=1}^{N} \phi_{X_j}(t_j)\right| \leq \sum_{i<j}^{N} |t_j t_i| \ |cov(X_i, X_j)|,$$

where $\phi_{(X_1,X_2,...,X_i)}(t)$ is the characteristic function of the subset $(X_1, X_2, ..., X_i)$ for $i \geq 1$.

Proof: [Proof of Theorem 3]

First, we deal with assertion (20). To this end, we set $\hat{H}_{n,j}(x) := h_n^{\beta-1/2}\tilde{H}_{n,j}(x)$ and $S_n := \sum_{j=1}^{n} \hat{H}_{n,j}(x)$. The previous analysis indicates that

$$\lim_{n\to\infty} var(\frac{S_n}{\sqrt{n}}) = \sigma^2(x) < \infty. \tag{30}$$

Hence it suffices to confirm that

$$\frac{S_n}{\sqrt{n}} \to N(0, \sigma^2(x)).$$

The procedure used here is based on decomposing the sum S_n into appropriate blocks and dealing with these blocks as if they were independent. Typically, this requires controlling the approximation between the real associated blocks and its counterpart independent blocks. This approximation is essentially established using characteristic functions and Lemma 7.

We define the characteristic function of $\frac{S_n}{\sqrt{n}}$ as $\Psi_n(t) = E(e^{it\frac{S_n}{\sqrt{n}}})$ for all $t \in \mathbb{R}$ and let $s \in \mathbb{N}$ and $k = \left[\frac{n}{s}\right]$. Hence, $ks \leq n \leq ks + s$. The blocks are defined as follows:

$$S_n = \sum_{j=1}^{k+1} Z_{j,s},$$

where

$$Z_{j,s} = \sum_{i=(j-1)s+1}^{js} \hat{H}_{n,i}(x) \text{ for all } j = 1, ..., k \text{ and } Z_{k+1,s} = \sum_{i=sk+1}^{n} \hat{H}_{n,i}(x). \tag{31}$$

Thus, we need to have

$$\lim_{n\to\infty} \left|\Psi_n(t) - \exp\left(-\frac{\sigma^2(x)t^2}{2}\right)\right| = 0. \tag{32}$$

Now, we divide our proof into four main steps. In the first three steps we consider s to be a fixed finite number. To get the conclusion in our last step, we will let s go to infinity.

Step 1: As a first step, we take n to be a multiple of s. If this is not the case, then:

$$\begin{aligned}
|\Psi_n(t) - \Psi_{ks}(t)| &= \left|E(e^{it\frac{S_n}{\sqrt{n}}}) - E(e^{it\frac{S_{ks}}{\sqrt{ks}}})\right| \\
&= \left|E\left[e^{it\frac{S_{ks}}{\sqrt{ks}}}(e^{it\left(\frac{S_n}{\sqrt{n}} - \frac{S_{ks}}{\sqrt{ks}}\right)} - 1)\right]\right| \\
&\leq E\left|e^{it\left(\frac{S_n}{\sqrt{n}} - \frac{S_{ks}}{\sqrt{ks}}\right)} - 1\right|.
\end{aligned}$$

Using Cauchy-Schwarzs inequality and the fact that $\left|e^{it}-1\right| \leq |t|$

$$\begin{aligned} |\Psi_n(t) - \Psi_{ks}(t)| &\leq \left| E^{1/2}\left[|t| \left(\frac{S_n}{\sqrt{n}} - \frac{S_{ks}}{\sqrt{ks}} \right) \right]^2 \right|. \\ &= |t|\, var \left(\frac{S_n}{\sqrt{n}} - \frac{S_{ks}}{\sqrt{ks}} \right)^{1/2}. \end{aligned}$$

Now, let us define $\Gamma(m) \triangleq var\left(\frac{S_m}{\sqrt{m}}\right)$ for any $m \geq 1$. From the positive association and equation (30), we see that $\Gamma(m) < \infty$. Then

$$var\left(\frac{S_n}{\sqrt{n}} - \frac{S_{ks}}{\sqrt{ks}}\right) = var\left(\frac{S_n - S_{sk}}{\sqrt{n}} - \frac{\sqrt{n}-\sqrt{sk}}{\sqrt{nsk}} S_{sk}\right).$$

Thus, by positive association, we can write

$$var(\frac{S_n}{\sqrt{n}} - \frac{S_{ks}}{\sqrt{ks}}) \leq \frac{1}{n} var\left(S_{n-sk}\right) + \left(\frac{\sqrt{n}-\sqrt{sk}}{\sqrt{n}}\right)^2 \Gamma(sk).$$

Again by association, and since $0 < n - sk < s$

$$\begin{aligned} \frac{1}{n} var\left(S_{n-sk}\right) &< \frac{1}{n} var\left(S_s\right) \\ &= \frac{s}{n}\Gamma\left(S_s\right). \end{aligned}$$

We may note that $\sqrt{n} - \sqrt{sk} < \sqrt{n-sk}$. Thus

$$var\left(\frac{S_n}{\sqrt{n}} - \frac{S_{ks}}{\sqrt{ks}}\right) \leq \frac{s}{n}\left[\Gamma\left(S_s\right) + \Gamma(sk)\right].$$

Finally, we get

$$\lim_{n\to\infty} |\Psi_n(t) - \Psi_{ks}(t)| = 0. \tag{33}$$

Step 2: This step is devoted to controlling the approximation of the joint distribution of the underlying blocks and see what we get if we assume that we have independent blocks.

Notice that, for a fixed s we can write $Z_{js} = S_{js} - S_{(j-1)s}$. The latter is a sum of s strictly stationary r.v.'s. Then, the distribution of Z_{js} and S_s are the same. Thus, we can consider Ψ_s as a characteristic function of $\frac{1}{\sqrt{s}} Z_{js}$. Actually, only monotone transformations of the original variables can keep the association, which is the case with Z_{js}. Then, we can apply Newman's inequality (see Lemma 7) to show that

$$\left|\Psi_{ks}\left(t\right) - \Psi_s^k\left(\frac{t}{\sqrt{s}}\right)\right| \leq \frac{t^2}{2k} \sum_{\substack{j,j'=1 \\ j\neq j'}}^{k} cov\left(Z_{js}, Z_{js}\right).$$

Due to the stationarity again, we get

$$\begin{aligned} \frac{1}{k}\sum_{\substack{j,j'=1 \\ j\neq j'}}^{k} cov\left(Z_{js}, Z_{js}\right) &= \frac{1}{k} var\left(\sum_{j=1}^{k} Z_{j,s}\right) - \frac{1}{k}\sum_{j=1}^{k} var\left(Z_{j,s}\right) \\ &= \frac{1}{k} var\left(S_{ks}\right) - var\left(S_s\right) \\ &= s\left[\Gamma\left(ks\right) - \Gamma\left(s\right)\right]. \end{aligned}$$

By simple algebra and using the stationarity, we can conclude that

$$\left|\Psi_{ks}\left(t\right)-\Psi_{s}^{k}\left(\frac{t}{\sqrt{s}}\right)\right|\to 0. \tag{34}$$

Step 3: In this step, we suppose the independence between the blocs $Z_{j,s}$. Applying the usual CLT for i.i.d. parts, and the fact $n\to\infty$ (and so $k\to\infty$) we see that

$$\left|\Psi_{s}^{k}\left(\frac{t}{\sqrt{k}}\right)-\exp\left(-\frac{t^{2}\Gamma\left(s\right)}{2}\right)\right|\to 0. \tag{35}$$

Step 4: The limits in equations (33), (34) and (35) imply that

$$\left|\Psi_{n}\left(t\right)-\exp\left(-\frac{t^{2}\sigma^{2}}{2}\right)\right|\to\left|\exp\left(-\frac{t^{2}\sigma^{2}}{2}\right)-\exp\left(-\frac{t^{2}\Gamma(s)}{2}\right)\right|,$$

where $\sigma^{2}=\lim_{n\to\infty}\Gamma(s)$. Using the inequality $|\exp\left(t\right)-\exp\left(t'\right)|\leq|t-t'|$ for any $t,t'\in\mathbb{R}$, we have

$$\lim_{n\to\infty}\sup\left|\Psi_{n}\left(t\right)-\exp\left(-\frac{t^{2}\sigma^{2}}{2}\right)\right|\leq\frac{t^{2}}{2}\left|\sigma^{2}-\Gamma(s)\right|.$$

We complete this step by letting $s\to\infty$. The proof of the first point is finished.

For the second conclusion we normalize our r.v.'s by considering $\hat{H}_{n,j}(x)\triangleq\frac{var(\tilde{H}_{n,j}(x))}{\tilde{H}_{n,j}(x)}$. Here, we need the same arguments to show that

$$\frac{S_{n}}{\sqrt{n}}\to N(0,1).$$

Thus, we would like to have

$$\lim_{n\to\infty}\left|\Psi_{n}(t)-\exp\left(-\frac{t^{2}}{2}\right)\right|=0. \tag{36}$$

It is note worthy that, regardless of the type of errors, whether ordinary smooth or super smooth, positive association and strict stationarity properties stay unaltered. Therefore, only a few modifications are required to adapt the previous steps to the current proof. Indeed, it is easy to see that

$$\lim_{n\to\infty}\Gamma(n)=1.$$

We follow the same steps as before and start modifications at Eq. (34) in **Step 2**. We notice that $var(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\xi_{i})\simeq var(\frac{S_{ms}}{ms})=\sigma_{ms}^{2}$, and by using the fact that $\{\xi_{i}\}_{i=1}^{i=m}$ is associated and identically distributed we have $\frac{1}{m}\sum_{i=1}^{m}var(\xi_{i})\asymp var(\xi_{1})=\sigma_{s}^{2}$. Then

$$\left|\Psi_{ms}(t)-\left(\Psi_{s}(\frac{t}{\sqrt{m}})\right)^{m}\right|\leq\frac{t^{2}}{m}(\sigma_{ms}^{2}-\sigma_{s}^{2}),$$

by letting $n\to\infty$ and $s\to\infty$ (and so $m\to\infty$), we get

$$\lim_{m\to\infty}\left(\Psi_{s}(\frac{t}{\sqrt{m}})\right)^{m}\to e^{-\sigma_{s}^{2}t^{2}/2}.$$

Then it comes out that

$$\lim_{n\to\infty}\left|\Psi_{ms}(t)-e^{-\sigma^{2}t^{2}/2}\right|=0.$$

Proof: [Proof of Corollary 2]

We consider the following decomposition:

$$F_n(x) - E(F_n(x)) = F_n(x) - F(x) - bias\left[F_n(x)\right].$$

Theorem 3 ensures distribution convergence for $F_n(x) - F(x)$. Then, the second conclusion in Proposition 1 shows that

$$\lim_{n\to\infty} h_n^{-2} bias\left[F_n(x)\right] = C < \infty.$$

Hence, we only have to show that $\frac{h_n^2}{n^{1/2}h_n^{\beta-1/2}} \underset{n\to\infty}{\to} 0$ and $h_n^2 n^{1/2} \underset{n\to\infty}{\to} 0$ for the ordinary smooth and super-smooth cases respectively. But this is clearly latent in the assumptions of Theorem 3.

Acknowledgements

The authors are grateful to the referees for his comments and suggestions which helped us to improve the quality and the presentation of this manuscript.

References

Birkel T. On the convergence rate in the central limit theorem for associated processes. Ann Prob. 1988; 16(4): 1685-1698.

Bochner S. Lectures on Fourier integrals. New York: Princeton University Press; 1959.

Cordy CB, Thomas DR. Deconvolution of a distribution function. J Amer Stat Assoc. 1997; 92(440): 1459-1465.

Dattner I, Goldenshluger A, Juditsky A. On deconvolution of distribution functions. Ann Stat. 2011; 39(5): 2477-2501.

Esary JD, Proschan F, Walkup DW. Association of random variables, with applications. Ann Math Stat. 1967; 3(1): 1466-1474.

Fan J. Asymptotic normality for deconvolving kernel density estimators. Sankhya. 1990; 53(2): 1-31.

Fan J. On the optimal rates of convergence for nonparametric deconvolution problems. Ann Stat. 1991; 23(1): 1257-1272.

Fan J and Masry E. Multivariate regression estimation with errors-invariables: Asymptotic normality for mixing processes. J Multivariate Anal. 1992; 43(1): 237-271.

Fortuin CM, Kasteleyn PW, Ginibre J. Correlation inequalities on some partially ordered sets. Comm Math Phys. 1971; 22(2): 89-103.

Gaffey WR. A consistent estimator of a component of a convolution. Inst Math Stat. 1959; 30(1): 198-205.

Jiazhu P. Tail dependence of random variables from ARCH and heavy-tailed bilinear models. Sciences in China. 2002; 45(1): 749-760.

Jones AF, Misell DL. A practical method for deconvolution of experimental curves. Brit J Appl Phys. 1967; 18(1): 1479-1483.

Ioannides DA, Papanastassiou, DP. Estimating the distribution function of a stationary process involving measurement errors. Stat Inference Stoch Processes. 2001; 2(4): 181-198.

Lejeune M, Sarda P. Smooth estimators of distribution and density functions. Comput Stat Data Anal. 1992; 14(1): 457-471.

Masry E. Multivariate probability density deconvolution for stationary random processes. IEEE Trans Inform Theory. 1991; 37(4): 1105-1115.

Masry E. Strong consistency and rates for deconvolution of multivariate densities of stationary processes. Stochast Process Appl. 1993; 47(1): 53-74.

Masry E. Deconvolving multivariate kernel density estimates from contaminated associated observations. IEEE Trans Inform Theory. 2003; 49(11): 2941-2952.

Medgyessy P. Decomposition of superpositions of density functions and discrete distributions. New York: John Wiley & Sons; 1977.

Mendelsohn J, Rice J. Deconvolution of microfluorometric histograms with B-splines. J Amer Stat Assoc. 1982; 77(1): 748-753.

Newman CM, Wright AL. An invariance principle for certain dependent sequences. Ann Prob. 1981; 9(4): 671-675.

Oliveira P E. Asymptotics for associated random variables. New York: Springer Science & Business Media; 2012.

Parzen E. Estimation of a probability density function and mode. Ann Math Stat. 1962; 33(1): 145-156.

Rao BLSP. Associated sequences, demimartingales and nonparametric inference. Prob its Appl; 2012.

Rice J, Rosenblatt M. Estimation of the log survivor function and hazard function. Sankhya. 1976; 38(1): 60-78.

Snyder DL, Miller MI, Schultz TJ. Constrained probability density estimation from noisy data. Proc: 22nd Annual Conference on Information Sciences and System. 1988.

Wang X, Fan Z, Wangc B. Estimating smooth distribution function in the presence of heteroscedastic measurement errors. Comput Stat Data Anal. 2010; 54(1): 25-36.

Watson GS, Leadbetter MR. Hazard analysis I. Biometrika. 1964; 51(1/2): 175-184.

Wheeden RL, Zygmund A. Measure and integral : an introduction to real analysis. New York: Marcel Dekker; 1977.

Wise G, Traganitis A, Thomas J. The estimation of a probability density function from measurements corrupted by Poisson noise. IEEE Trans Inform Theory. 1977; 23(6): 764-766.

Zhang CH. Fourier methods for estimating mixing densities and distributions. Ann Stat. 1990; 18(1): 806-830.