INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0913

CONSISTENT DENSITY DECONVOLUTION UNDER PARTIALLY KNOWN ERROR DISTRIBUTION

SCHWARZ, M. and S. VAN BELLEGEM

Consistent density deconvolution under partially known error distribution

Maik Schwarz* Sébastien Van Bellegem⋆

July 27, 2009

We are interested in estimating the density f^X of a real-valued random variable X based on an i.i.d. sample from $Y = X + \varepsilon$, where ε is an independent additive error. In the literature, the density of the noise is usually supposed to be fully known. In contrast to this, we assume that ε is normally distributed, but with unknown variance $\sigma^2 > 0$. First, we show that (f^X, σ^2) can be identified from the observations when f^X vanishes on a set of positive Lebesgue measure. As opposed to standard procedures, this identification condition is not based on properties of the densities' characteristic functions. Deconvolving a density is well-known not to be continuous with respect to the L^2-norms, that is why it is called an ill-posed inverse problem. However, we show that deconvolution becomes continuous if we choose the topology of weak convergence for the deconvolution density and an appropriate topology for the observations' density. As a consequence, a minimum distance estimator of f^X will be weakly consistent imposing only a slightly stronger assumption than the identification condition. In particular, no further conditions on the densities' characteristic functions are required. The result remains true even if we do not require the involved probability distributions to have densities.

Keywords: deconvolution, measurement error, density estimation, minimum distance estimator, identification

Acknowledgments

The authors gratefully acknowledge the helpful suggestions of Jan Johannes during the preparation of the paper. They further wish to express their thanks to Michael Neumann for a stimulating discussion of some measure theoretic details in the proof of Theorem 3.5.

This work was supported by the IAP research network no P5/24 of the Belgian Government (Belgian Science Policy) and by the "Fonds Spéciaux de Recherche" from the Université catholique de Louvain.

*Université catholique de Louvain, maik.schwarz@uclouvain.be

⋆Toulouse School of Economics, svb@tse-fr.eu

1 Introduction

A standard problem in nonparametric statistics is to consistently estimate the distribution of some real random variable X from a statistical sample that is subject to an additive measurement error ε. The usual setting is to assume independent and identically distributed (iid) observations from a random variable Y that is such that $Y = X + \varepsilon$. Knowing the cumulative distribution function (cdf) of ε, a vast literature focusses on the accuracy of estimation of the cdf of X (e.g. Carroll and Hall (1988); Fan (1991)). The full knowledge of the cdf of ε is of course a strong assumption that is rarely encountered in real data analysis. However this assumption is also an identification condition: without the full information on the cdf of ε, the cdf of X cannot be identified from the observations of Y.

In order to circumvent that issue, some authors have worked under other sampling processes. In addition to the observation of a sample of Y some papers assume to observe another independent sample from the measurement error ε. That new sample allows to estimate the cdf of ε in a first step, and therefore to recover the cdf of X in a second step (e.g. Neumann (1997), Cavalier and Hengartner (2005), Johannes et al. (2009)). In other studies, it is instead assumed to observe longitudinal, or repeated, versions of Y (e.g. Li and Vuong (1998); Meister and Neumann (2009)).

The present paper develops another strategy to identify the cdf of X. We assume that the measurement error, ε, is normally distributed with mean zero and unknown variance σ^2. In this setting the cdf of X is of course not identified from the observation of Y. However if we restrict the class of cdfs of X we show below that the model becomes identified in spite of the partial knowledge on the cdf of ε. In Section 2 below we prove that if the real random variable X vanishes on some interval (or more generally on some set with nonzero Lebesgue measure) then its cdf is identified from the observation of $Y = X + \varepsilon$, where ε is now $\mathcal{N}(0, \sigma^2)$ with unknown σ^2.

The normality of ε is not a restrictive assumption in real data analysis, in which the main issue is often to evaluate the level of the noise σ^2. In Hall and Simar (2002) a similar setting is considered but under the additional assumption that σ^2 depends on the sample size n in such a way that $\sigma^2 \to 0$ as $n \to \infty$. Matias (2002) and Butucea and Matias (2005) also consider the consistent estimation of σ^2 and of the cdf of X under strong restrictions on the characteristic function of X. In Section 3 we show that if we slightly restrict the class of cdfs of X, then we can prove the consistency of a minimum penalized contrast estimator of the cdf and σ^2. The originality of our approach is to exploit the qualitative prior on X (which is vanishing on some interval) in order to separate it from the measurement error.

The next section addresses the identification issue and in Section 3 we prove the consistency of the minimum penalized contrast estimator. A technical lemma on measure theory is deferred to an appendix.

2 Identification

Suppose we want to recover the probability distribution P^X of a random variable X that is observed with an additive random contamination error ε. This measurement error is assumed to be normally distributed with mean zero and unknown variance σ^2. The resulting observational model is

$$Y = X + \varepsilon. \tag{1}$$

The distribution of Y is the convolution $P^Y = P^X * \mathcal{N}_\sigma$, where $\mathcal{N}_\sigma$ denotes the probability distribution of ε. Writing φ^X, φ^Y and φ_σ for the characteristic functions of P^X, P^Y and $\mathcal{N}_\sigma$, respectively, the convolution equation is equivalent to $\varphi^Y = \varphi^X \varphi_\sigma$ by virtue of the convolution theorem. Because of the uncertainty about the variance of the measurement error, not all probability distributions can be recovered from the model. Define the set of distributions

$$\mathcal{P}_0 = \left\{ P \in \mathcal{P} \;\middle|\; \exists\, A \in \mathcal{B}(\mathbb{R}) \,:\, |A| > 0 \;\wedge\; P(A) = 0 \right\},$$

where $\mathcal{B}(\mathbb{R})$ denotes the set of Borel sets in $\mathbb{R}$ and $\mathcal{P}$ the set of all probability distributions, and $|A|$ the Lebesgue measure of A. In the following theorem, we show that all distributions belonging to $\mathcal{P}_0$ are identifiable from the observational model.

Theorem 2.1 (Identification) *The model defined by (1) is identifiable for the parameter space $\mathcal{P}_0 \times (0, \infty)$, that is, for any two probability measures $P^1, P^2 \in \mathcal{P}_0$ and $\sigma_1, \sigma_2 > 0$, we have that $P^1 * \mathcal{N}_{\sigma_1} = P^2 * \mathcal{N}_{\sigma_2}$ implies $P^1 = P^2$ and $\sigma_1 = \sigma_2$.*

The proof of this theorem is based on the following lemma.

Lemma 2.2 *Let P^1 and P^2 be probability distributions and $0 < \sigma_1 < \sigma_2$. Then,*

$$P^1 * \mathcal{N}_{\sigma_1} = P^2 * \mathcal{N}_{\sigma_2} \implies P^1 = P^2 * \mathcal{N}_{\sigma_3}, \qquad \textit{where } \sigma_3 = \sqrt{\sigma_2^2 - \sigma_1^2}.$$

Proof. First, apply the convolution theorem on both sides of the equation, then divide by φ_{σ_1} which is positive everywhere. To conclude, it suffices to remark that $\varphi_{\sigma_3} = (\varphi_{\sigma_2}/\varphi_{\sigma_1})$. □

Proof of Theorem 2.1. Suppose that $(P^1, \sigma_1), (P^2, \sigma_2) \in \mathcal{P}_0 \times (0, \infty)$ are chosen in such a way that $P^1 * \mathcal{N}_{\sigma_1} = P^2 * \mathcal{N}_{\sigma_2}$. It has to be shown that this implies $(P^1, \sigma_1) = (P^2, \sigma_2)$. First, we prove by contradiction that $\sigma_1 = \sigma_2$. Suppose that $\sigma_1 \neq \sigma_2$. Without loss of generality, say $\sigma_1 < \sigma_2$. By virtue of Lemma 2.2, this implies $P^1 = P^2 * \mathcal{N}_{\sigma_3}$.

We show now that this is only possible if P^1 is not in $\mathcal{P}_0$ which contradicts the assumption. Indeed, let $A = [a_1, a_2]$ be some interval of positive length $|A| = a_2 - a_1$ and $B = [b_1, b_2]$ another interval with $|B| < |A|$ and $P^2(B) > 0$. By definition of the convolution and in view of the independence of X and ε in our model, we can write $(P^2 * \mathcal{N}_{\sigma_3})(A) = (P^2 \otimes \mathcal{N}_{\sigma_3})(S_A)$, where $S_A = \{(x, y) \in \mathbb{R}^2 \mid x + y \in A\}$ and $\otimes$ denotes the product measure. We have that $a_1 - b_1 < a_2 - b_2$ because of $|B| < |A|$. It is easily verified that $B \times [a_1 - b_1, a_2 - b_2] \subset S_A$ and hence

$$P^1(A) = (P^2 * \mathcal{N}_{\sigma_3})(A) \geq P^2(B)\, \mathcal{N}_{\sigma_3}([a_1 - b_1, a_2 - b_2]) > 0.$$

This contradicts the assumption that $P^1 \in \mathcal{P}_0$, showing that $\sigma_1 = \sigma_2$.

The characteristic function of the normal distribution being positive everywhere, an application of the convolution theorem completes the proof. □

It is worth noticing that the identification holds on $\mathcal{P}_0$ only. Indeed, if in Theorem 2.1 we do not suppose both P^1 and P^2 to belong to $\mathcal{P}_0$, the conclusion is false in general as the following counterexample illustrates. Let P^1 be the uniform distribution on the unit interval and φ^1 its characteristic function. Clearly, $P^1 \in \mathcal{P}_0$. If we let P^2 be the probability distribution with characteristic function $\varphi^2 := \varphi^1\, \varphi_\sigma / \varphi_{(\sigma/2)}$, then, in view of the convolution theorem, we have $P^1 * \mathcal{N}_\sigma = P^2 * \mathcal{N}_{(\sigma/2)}$, but $P^1 \neq P^2$.

We conclude this section by remarking that the probability distributions in question are not required to have densities. For those having one, the identification condition can be equivalently expressed by requiring that the density has to vanish on a set of positive Lebesgue measure.

3 Estimation

Now suppose we have an i.i.d. sample $\{Y_1, \ldots, Y_n\}$ from the model (1). Let $\xrightarrow{\mathcal{D}}$ denote convergence in distribution. An estimator $(\widehat{P}_n^X, \widehat{\sigma}_n)$ of (P^X, σ) is called *consistent* if, almost surely, $\widehat{P}_n^X \xrightarrow{\mathcal{D}} P^X$ and $\widehat{\sigma}_n \to \sigma$ as $n \to \infty$. For a consistent estimator, we always have $\widehat{P}_n^X * \mathcal{N}_{\widehat{\sigma}_n} \xrightarrow{\mathcal{D}} P^Y$, which is hence a necessary condition of consistency. An estimator satisfying this condition is called *admissible*.

3.1 Minimum distance estimation

Let $\widehat{\varphi}_n^Y(t) = \frac{1}{n}\sum_{k=1}^n \exp(itY_k)$ be the empirical characteristic function of the observations. By the Glivenko-Cantelli theorem, it converges almost surely uniformly to the true characteristic function φ^Y. For characteristic functions $\widetilde{\varphi}^X, \widetilde{\varphi}_\sigma$, and $\widetilde{\varphi}^Y$ let us define a distance ρ,

$$\rho(\widetilde{\varphi}^X, \widetilde{\varphi}_\sigma; \widetilde{\varphi}^Y) := \int_{\mathbb{R}} |\widetilde{\varphi}^X(t)\, \widetilde{\varphi}_\sigma(t) - \widetilde{\varphi}^Y(t)|\; h(t)\; dt, \tag{2}$$

where h is some continuous and strictly positive probability density ensuring the existence of the integral. The estimation consists in choosing $\widehat{P}_n^X$ and $\mathcal{N}_{\widehat{\sigma}_n}$ such that their characteristic functions $\widehat{\varphi}_n^X$ and $\varphi_{\widehat{\sigma}_n}$ minimize $\rho(\cdot\,,\cdot\,; \widehat{\varphi}_n^Y)$. Since this minimum is not necessarily attained, we give the following definition.

Definition 3.1 *Let $(\delta_n)_{n\in\mathbb{N}}$ be a vanishing sequence and $\mathcal{C}$ a set of probability distributions. In the context of the deconvolution model, we call a random sequence $(\widehat{P}_n^X, \widehat{\sigma}_n)$ depending on the observations $\{Y_1, \dots, Y_n\}$ a* minimum distance estimator *on $\mathcal{C}$ if it is such that the corresponding characteristic functions yield*

$$\rho(\widehat{\varphi}_n^X, \varphi_{\widehat{\sigma}_n}; \widehat{\varphi}_n^Y) \leq \inf_{\substack{\widetilde{\varphi}^X \in \Phi_{\mathcal{C}} \\ \widetilde{\sigma} \geq 0}} \rho(\widetilde{\varphi}^X, \varphi_{\widetilde{\sigma}}; \widehat{\varphi}_n^Y) + \delta_n, \tag{3}$$

where we denote by $\Phi_{\mathcal{C}}$ the set of all characteristic functions of some class of distributions $\mathcal{C}$. Let further $\widehat{P}_n^{X+\varepsilon} := \widehat{P}_n^X * \mathcal{N}_{\widehat{\sigma}_n}$, the characteristic function of which is $\widehat{\varphi}_n^X \varphi_{\widehat{\sigma}_n}$.

Our aim is to prove the consistency of this estimator. Obviously, this requires further assumptions on the class $\mathcal{C}$. In the first instance, we show that the minimum distance estimator is always admissible.

Lemma 3.2 (Admissibility) *Any minimum distance estimator $(\widehat{P}_n^X, \widehat{\sigma}_n)$ on the set $\mathcal{P}$ of all probability distributions is admissible.*

Proof. The empirical characteristic function $\widehat{\varphi}_n^Y$ converges pointwise to φ^Y. By Lebesgue's Theorem, this implies $\rho(\varphi^X, \varphi_\sigma; \widehat{\varphi}_n^Y) \to 0$ almost surely. Applying the triangle inequality and using $\varphi^X \varphi_\sigma = \varphi^Y$, we obtain

$$\rho(\widehat{\varphi}_n^X, \varphi_{\widehat{\sigma}_n}; \varphi^Y) \leq \rho(\widehat{\varphi}_n^X, \varphi_{\widehat{\sigma}_n}; \widehat{\varphi}_n^Y) + \rho(\varphi^X, \varphi_\sigma; \widehat{\varphi}_n^Y).$$

Because of (3), we can write $\rho(\widehat{\varphi}_n^X, \varphi_{\widehat{\sigma}_n}; \widehat{\varphi}_n^Y) \leq \rho(\varphi^X, \varphi_\sigma; \widehat{\varphi}_n^Y) + \delta_n$, so that we conclude that, almost surely,

$$\rho(\widehat{\varphi}_n^X, \varphi_{\widehat{\sigma}_n}; \varphi^Y) \leq 2\rho(\varphi^X, \varphi_\sigma; \widehat{\varphi}_n^Y) + \delta_n \to 0.$$

We choose an element $\omega \in \Omega$ of the underlying probability space such that this convergence holds. As the integrand in (2) is non-negative, it follows that

$$\int_0^a \widehat{\varphi}_n^X(t) \varphi_{\widehat{\sigma}_n}(t) dt \longrightarrow \int_0^a \varphi^Y(t) dt$$

for all $a \in \mathbb{R}$ as $n \to \infty$. We have shown that the integrated characteristic functions of the measures $\widehat{P}_n^{X+\varepsilon}$ converge to the integrated characteristic function of the probability measure P^Y, so that, applying Theorem 6.3.3 from Chung (1968), we get $\widehat{P}_n^{X+\varepsilon} \xrightarrow{\mathcal{V}} P^Y$, where $\xrightarrow{\mathcal{V}}$ denotes vague convergence. Knowing that the measures $\widehat{P}_n^{X+\varepsilon}$ as well as their vague limit P^Y are probability distributions, the Portmanteau Theorem implies that we do in fact have weak convergence, which means that the estimator is admissible. $\square$

Remark 3.3 We have seen that the minimum contrast estimator is always admissible. Next, we determine classes of distributions on which it is also consistent. One might wonder if the identification condition alone is sufficient to guarantee consistency, that is, if minimum distance estimators on $\mathcal{P}_0$ are consistent. This is not the case, as the following counterexample illustrates. Let

$$\widehat{P}_n^X(A) := \widehat{P}_n^Y(A \cap [-n, n])/\widehat{P}_n^Y(A)$$

for any Borel set A, where $\widehat{P}_n^Y$ is any consistent estimator of P^Y. Note that $\widehat{P}_n^X \in \mathcal{P}_0$ for every $n \geq 1$. Let further $\widehat{\sigma}_n := (1/n)$. It is easily verified that this estimator is admissible but not consistent. The following consideration shows in which way we have to restrict the class $\mathcal{P}_0$ in order to obtain consistency. Assume that $P^X \in \mathcal{P}_0$ and let $(\widehat{P}_n^X, \widehat{\sigma}_n)$ be an admissible estimator. By virtue of Lemma A.1 below, admissibility implies the existence of an increasing sequence $(n_k)_{k\in\mathbb{N}}$, some probability measure P_∞^X, and $\sigma_\infty \geq 0$ such that

$$\widehat{P}_{n_k}^X \xrightarrow{\mathcal{D}} P_\infty^X \quad \text{and} \quad \widehat{\sigma}_{n_k} \longrightarrow \sigma_\infty \tag{4}$$

as $n \to \infty$, which implies $\widehat{P}_{n_k}^X * \mathcal{N}_{\widehat{\sigma}_{n_k}} \xrightarrow{\mathcal{D}} P_\infty^X * \mathcal{N}_{\sigma_\infty}$, and hence, due to admissibility and by uniqueness of the weak limit,

$$P_\infty^X * \mathcal{N}_{\sigma_\infty} = P^X * \mathcal{N}_\sigma. \tag{5}$$

It follows from (4) that a necessary condition for $\widehat{P}_n^X$ to be consistent is $P_\infty^X = P^X$. In view of (5) and Theorem 2.1, this is equivalent to $P_\infty^X \in \mathcal{P}_0$. But this may not be the case in spite of all $\widehat{P}_{n_k}^X$ lying in $\mathcal{P}_0$ as this class is not closed under convergence in distribution as the above counterexample shows.

3.2 Consistency

In Remark 3.3 we have seen that in order to show consistency of the minimum distance estimator, we need to restrict the set of considered distributions. For $R, \eta > 0$, let

$$\mathcal{P}_R^\eta := \{P \in \mathcal{P} \mid \exists A = (a_1, a_2) \subset [-R, R] \ : \ |A| \geq \eta \ \wedge \ P(A) = 0\}.$$

Indeed, this choice avoids the problem of possibly obtaining a sequence of estimators the weak limit of which lies outside the identified class $\mathcal{P}_0$, as the following lemma shows. Note that for a positive random variable X, we have $P^X \in \mathcal{P}_R^\eta$ for any choice of η and R.

Lemma 3.4 *For any $R, \eta > 0$, weakly convergent sequences in $\mathcal{P}_R^\eta$ have their limit in $\mathcal{P}_0$.*

Proof. Let $(P_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_R^\eta$. Then, we have

$$\forall n \in \mathbb{N} \ \exists \text{ interval } A_n \subset [-R, R] \ : \ P_n(A_n) = 0 \wedge |A_n| \geq \eta. \tag{6}$$

Suppose further that $(P_n)_{n\in\mathbb{N}}$ converges in distribution to some P_∞. We have to show that there is some $A_\infty \in \mathcal{B}([-R, R])$ of positive Lebesgue measure with $P_\infty(A_\infty) = 0$, that is $P_\infty \in \mathcal{P}_0$.

Firstly, we deduce from (6) that there exists an $x_0 \in [-R, R]$ which lies in infinitely many A_n, or in other words,

$$\exists x_0 \in [-R, R] \ \exists (n_k)_{k\in\mathbb{N}} \ \forall k \in \mathbb{N} \ : \ x_0 \in A_{n_k}.$$

As all A_{n_k} are intervals of length at least η, there is an interval containing x_0 which is a null set for infinitely many measures of the sequence P_{n_k}. More precisely, there is a subsequence n'_k of n_k such that

$$(x_0 - \eta/2, x_0] \subset \bigcap_{k\in\mathbb{N}} A_{n'_k} \quad \text{or} \quad [x_0, x_0 + \eta/2) \subset \bigcap_{k\in\mathbb{N}} A_{n'_k}.$$

Hence, we can choose $A_\infty = (x_0 - \eta/2, x_0)$ or $A_\infty = (x_0, x_0 + \eta/2)$ satisfying $|A_\infty| = \eta/2 > 0$ and $P_{n'_k}(A_\infty) = 0$ for all $k \in \mathbb{N}$.
The latter assertion implies that $\liminf_{n\to\infty} P_n(A_\infty) = 0$. Recall that the P_n converge weakly to P_∞ and that A_∞ is an open set. Therefore, the Portmanteau theorem allows us to conclude that $P_\infty(A_\infty) = 0$. □

Before proving consistency, recall the definition of the Lévy distance: For probability distributions P^1, P^2 with cumulative distribution functions F^1, F^2, define

$$d(P^1, P^2) := \inf\{\delta > 0 \mid F^1(x-\delta) - \delta \le F^2(x) \le F^1(x+\delta) + \delta \quad \forall x \in \mathbb{R}\}.$$

For a sequence P_n of probability distributions, one has that $P_n \xrightarrow{\mathcal{D}} P$ if and only if $d(F_n, F) \to 0$ as $n \to \infty$. In other words, d metrizes the weak convergence. Now define, for probability distributions $\widetilde{P}^X$ and real numbers $\widetilde{\sigma}$,

$$\Delta(\widetilde{P}^X, \widetilde{\sigma}; P^X, \sigma) = d(\widetilde{P}^X, P^x) + |\widetilde{\sigma} - \sigma|.$$

Remark that $\Delta(\mathbf{P}_n^X, \sigma_n; P^X, \sigma) \to 0$ if and only if $P_n^X \xrightarrow{\mathcal{D}} P^X$ and $\sigma_n \to \sigma$ (and hence $\mathcal{N}_{\sigma_n} \xrightarrow{\mathcal{D}} \mathcal{N}_\sigma$).

Theorem 3.5 (Consistency) *Let $R, \eta > 0$ and suppose that in the deconvolution model (1), we have $P^X \in \mathcal{P}_R^\eta$. Then, any estimator $(\widehat{P}_n^X, \widehat{\sigma}_n)$ on $\mathcal{P}_R^\eta$ is consistent, that is, we have $\Delta(\widehat{P}_n^X, \widehat{\sigma}_n; P^X, \sigma) \to 0$ almost surely.*

Proof. We have seen in Lemma 3.2 that the considered estimator is admissible. Now we show that this implies $\Delta(\widehat{P}_n^X, \widehat{\sigma}_n; P^X, \sigma) \to 0$ under the assumptions of the theorem. The proof is by contradiction. Assume there is a $\delta > 0$ and an increasing sequence $(n_k)_{k\in\mathbb{N}}$ in $\mathbb{N}$ such that $\Delta(\widehat{P}_{n_k}^X, \widehat{\sigma}_{n_k}; P^X, \sigma) \ge \delta \quad \forall\, k \in \mathbb{N}$. Lemma A.1 furnishes a subsequence $(n'_k)_{k\in\mathbb{N}}$ of $(n_k)_{k\in\mathbb{N}}$, a probability measure P_∞^X and a constant $\sigma_\infty \ge 0$ such that

$$\widehat{P}_{n'_k}^X \xrightarrow{\mathcal{D}} P_\infty^X \qquad \text{and} \qquad \mathcal{N}_{\widehat{\sigma}_{n'_k}} \xrightarrow{\mathcal{D}} \mathcal{N}_{\sigma_\infty}. \tag{7}$$

Denote the characteristic functions of P_∞^X and $\mathcal{N}_{\sigma_\infty}$ by φ_∞^X and φ_{σ_∞}, respectively. Since weak convergence implies pointwise convergence of the corresponding characteristic functions, we obtain by Fatou's Lemma that

$$\rho(\varphi_\infty^X, \varphi_{\sigma_\infty}; \varphi^Y) \le \lim\inf_{k\to\infty} \rho(\widehat{\varphi}_{n'_k}^X, \varphi_{\widehat{\sigma}_{n'_k}}; \varphi^Y) = 0,$$

that is, $\int_{\mathbb{R}} |\varphi_\infty^X(t)\,\varphi_{\sigma_\infty}(t) - \varphi^Y(t)|\; h(t)\; dt = 0$. As h is strictly positive and characteristic as functions are uniformly continuous on $\mathbb{R}$, we conclude that

$$\varphi_\infty^X(t)\,\varphi_{\sigma_\infty}(t) = \varphi^Y(t) = \varphi^X(t)\,\varphi_\sigma(t) \qquad \forall\, t \in \mathbb{R}.$$

Lemma 3.4 allows us to deduce from (7) that $P_\infty^X \in \mathcal{P}_0$. Hence, Theorem 2.1 ensures that $P_\infty^X = P^X$ and $\sigma_\infty = \sigma$. Together with (7), this implies $\Delta(\widehat{P}_{n'_k}^X, \widehat{\sigma}_{n'_k}; P^X, \sigma) \to 0$. This contradicts the assumption and completes the proof. □

4 Conclusion

We have considered the problem of density deconvolution from one single contaminated sample with uncertainty in the error distribution and we have shown a minimum distance estimator to be consistent in this model. The estimation procedure presented here is inspired by a similar estimator suggested by Neumann (2007) in the context of panel data. Neumann proposes an identification assumption which also implies consistency. This condition is expressed in terms of characteristic functions.

Unlike this, the focus of the present note is on a weak identification condition in the time domain which reflects the properties of the involved distributions in a more natural way. In Section 2 we have proposed such an assumption. An additional difficulty arises from the fact that this condition is too weak to imply consistency, which motivates the definition of admissibility. However, a slight restriction of the considered class of distributions is sufficient to conclude, as the theorem shows.

As far as convergence rates under additional assumptions on the characteristic functions are concerned, the reader may refer to the work of Butucea and Matias (2005) in which rates are developed and shown to be minimax-optimal.

As for practical computability, the infinite-dimensional minimization problem (3) could be reduced to a finite-dimensional one by considering $\widetilde{\sigma} \in \Sigma_n$ and $\widetilde{\varphi}^X \in \Phi_{\mathcal{P}_n}$ only, where Σ_n becomes dense in $\mathbb{R}$ as $n \to \infty$ and $\mathcal{P}_n = \{\sum_{j=1}^{N_n} \alpha_j \delta_{x_j} \mid \alpha_j \in \mathbb{R}\}$ is a collection of purely atomic probability distributions which grows with n.

A Technical lemma

Lemma A.1 *Let Q_n be a sequence of probability distributions and σ_n a sequence of positive real numbers. Suppose further that $(Q_n * \mathcal{N}_{\sigma_n})_{n\in\mathbb{N}}$ converges weakly to some probability distribution. Then, there exist an increasing sequence $(n_k)_{k\in\mathbb{N}}$, a probability distribution Q_∞, and a constant $\sigma_\infty \geq 0$ such that*

$$Q_{n_k} \xrightarrow{\mathcal{D}} Q_\infty \quad \text{and} \quad \mathcal{N}_{\sigma_{n_k}} \xrightarrow{\mathcal{D}} \mathcal{N}_{\sigma_\infty}$$

as $n \to \infty$, where $\mathcal{N}_0 := \delta_0$ denotes the Dirac measure by convention.

Proof. By Helly's selection theorem, there is a subsequence $(n_k)_{k\in\mathbb{N}}$ and a subprobability measure Q_∞ such that $Q_{n_k} \xrightarrow{\mathcal{V}} Q_\infty$, where $\xrightarrow{\mathcal{V}}$ denotes vague convergence (e.g. Chung (1968)). We show below that the σ_{n_k} are bounded from above such that they have a convergent subsequence; without loss of generality, say $\sigma_{n_k} \to \sigma_\infty$ for some $\sigma_\infty \geq 0$. Proposition 3.1 from Jain and Orey (1979) states that if $R_n \xrightarrow{\mathcal{V}} R$ and $S_n \xrightarrow{\mathcal{D}} S$, then $R_n * S_n \xrightarrow{\mathcal{V}} R * S$, so we have $Q_{n_k} * \mathcal{N}_{\sigma_{n_k}} \xrightarrow{\mathcal{V}} Q_\infty * \mathcal{N}_{\sigma_\infty}$. By assumption, the same sequence converges weakly, and hence vaguely, to some distribution, so the uniqueness of the vague limit of measures on locally compact spaces implies $Q_\infty(\mathbb{R}) = 1$ because $(\mu * \nu)(\mathbb{R}) = \mu(\mathbb{R})\,\nu(\mathbb{R})$ for any two finite measures μ and ν on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The Portmanteau Theorem then implies $Q_{n_k} \xrightarrow{\mathcal{D}} Q_\infty$, which was our claim.

It rests to prove that σ_{n_k} is bounded from above. We show that otherwise the sequence $(Q_{n_k} * \mathcal{N}_{\sigma_{n_k}})_{k\in\mathbb{N}}$ would not be tight, which contradicts its weak convergence. Random variable notation is more convenient for this argument, so let $U_k \sim Q_{n_k}$ and $V_k \sim \mathcal{N}_{\sigma_{n_k}}$ be i.i.d. random variables and $W_k := U_k + V_k$. We have to show the non-tightness of the distributions of $\{W_k\}_{k\in\mathbb{N}}$, that is

$$\exists\, \delta \in (0,1)\ \forall\, J > 0\ \exists\, k \in \mathbb{N}\ :\ \mathbf{P}\big[W_k \in [-J, J]\big] < 1 - \delta.$$

Fix $\delta = (1/12)$ and $J > 0$, and let $\mathcal{J} = [-J, J]$. Put $I_j^+ = [3jJ, (3j+1)J]$ and $I_j^- = [-(3j+2)J, -(3j-1)J]$, and let $\mathcal{I} := \biguplus_{j\geq 0} I_j^+$ be the disjoint union of the I_j^+. Because of the monotony of the normal density on $[0,\infty)$, we have

$$\mathbf{P}[V_k \in I_j^+] > (1/3)\mathbf{P}\big\{V_k \in [3jJ, 3(j+1)J)\big\}.$$

The disjoint union over $j \geq 0$ of the intervals on the right hand side of this inequality is $[0,\infty)$, and $\mathbf{P}[V_k \geq 0] = (1/2)$. Thus, we have $\mathbf{P}[V_k \in \mathcal{I}] > (1/6)$. We can now write $\mathbf{P}[W_k \in \mathcal{J}] < (5/6) + (1/6)\mathbf{P}[W_k \in \mathcal{J} \mid V_k \in \mathcal{I}]$, and it is sufficient to prove that the conditional probability appearing in this inequality is less than $(1/2)$ for some k.

It is easy to check that

$$\mathbf{P}[W_k \in \mathcal{J} \mid V_k \in \mathcal{I}] = \sum_{j=0}^{\infty} \mathbf{P}[W_k \in \mathcal{J} \mid V_k \in I_j^+]\, \mathbf{P}[V_k \in I_j^+ \mid V_k \in \mathcal{I}].$$

By construction, $V_k \in I_j^+$ and $W_k \in \mathcal{J}$ together imply $U_k \in I_j^-$. Using further the monotony of the normal density on $[0,\infty)$, we deduce that

$$\mathbf{P}[W_k \in \mathcal{J} \mid V_k \in \mathcal{I}] \leq 6\mathbf{P}[V_k \in I_0^+] \sum_{j=0}^{\infty} \mathbf{P}[U_k \in I_j^-].$$

As the I_j^- are pairwise disjoint, the sum is bounded from above by 1, and hence we have $\mathbf{P}[W_k \in \mathcal{J} \mid V_k \in \mathcal{I}] \leq 6\mathbf{P}[V_k \in I_0^+]$. If σ_{n_k} is unbounded, k can be chosen in such a way that the right hand side of this inequality is less than $(1/2)$, which completes the proof. $\square$

References

Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric deconvolution model. *Bernoulli*, 11:309–340.

Carroll, R. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, 83:1184–1186.

Cavalier, L. and Hengartner, N. (2005). Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21:1345–1361.

Chung, K. (1968). *A course in probability theory.* Harcourt, Brace and World, Inc.

Fan, J. (1991). On the optimal rate of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19:1257–1272.

Hall, P. and Simar, L. (2002). Estimating a changepoint, boundary, or frontier in the presence of observation error. *J. Amer. Statist. Assoc.*, 97:523–533.

Jain, N. C. and Orey, S. (1979). Vague convergence of sums of independent random variables. *Isr. J. Math.*, 33:317–348.

Johannes, J., Van Bellegem, S., and Vanhems, A. (2009). Convergence rates for ill-posed inverse problems with an unknown operator. Technical report, IDEI Working Paper.

Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivar. Anal.*, 65:139–165.

Matias, C. (2002). Semiparametric deconvolution with unknown noise variance. *ESAIM: Probability and Statistics*, 6:271–292.

Meister, A. and Neumann, M. (2009). Deconvolution from non-standard error densities under replicated measurements. *Statist. Sinica.* to appear.

Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametric Stat.*, 7(4):307–330.

Neumann, M. H. (2007). Deconvolution from panel data with unknown error distribution. *J. Multivariate Anal.*, 98(10):1955–1968.