



UNIVERSITÀ
di **VERONA**

Department
of **ECONOMICS**

Working Paper Series
Department of Economics
University of Verona

On asymptotically efficient maximum likelihood estimation of
linear functionals in Laplace measurement error models

Catia Scricciolo

WP Number: 9

December 2018

ISSN: 2036-2919 (paper), 2036-4679 (online)

On asymptotically efficient maximum likelihood estimation of linear functionals in Laplace measurement error models

Catia Scricciolo

31 December 2018

Abstract Maximum likelihood estimation of linear functionals in the inverse problem of deconvolution is considered. Given observations of a random sample from a distribution $P_0 \equiv P_{F_0}$ indexed by a (potentially infinite-dimensional) parameter F_0 , which is the distribution of the latent variable in a standard additive Laplace measurement error model, one wants to estimate a linear functional of F_0 . Asymptotically efficient maximum likelihood estimation (MLE) of integral linear functionals of the mixing distribution F_0 in a convolution model with the Laplace kernel density is investigated. Situations are distinguished in which the functional of interest can be consistently estimated at $n^{-1/2}$ -rate by the plug-in MLE, which is asymptotically normal and efficient, in the sense of achieving the variance lower bound, from those in which no integral linear functional can be estimated at parametric rate, which precludes any possibility for asymptotic efficiency. The \sqrt{n} -convergence of the MLE, valid in the case of a degenerate mixing distribution at a single location point, fails in general, as does asymptotic normality. It is shown that there exists no regular estimator sequence for integral linear functionals of the mixing distribution that, when recentered about the estimand and \sqrt{n} -rescaled, is asymptotically efficient, *viz.*, has Gaussian limit distribution with minimum variance. One can thus only expect estimation with some slower rate and, often, with a non-Gaussian limit distribution.

Keywords Asymptotic efficiency · Asymptotic normality · Laplace convolution model · Linear functionals · Non-parametric maximum likelihood estimation

Mathematics Subject Classification (2000) 62G05 · 62G20 · 62G30

The author gratefully acknowledges financial support from Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR), grant n° 2015SNS29B "Modern Bayesian nonparametric methods".

Catia Scricciolo
Dipartimento di Scienze Economiche (DSE), Università degli Studi di Verona, Via Cantarane 24, I-37129 Verona, Italy
Tel.: +39-045-8028341
E-mail: catia.scricciolo@univr.it

1 Introduction

The problem of asymptotically efficient estimation of integral linear functionals of the distribution of the latent variable in a standard additive Laplace measurement error model is considered. The focus is on establishing whether asymptotic normality and efficiency hold for the estimator obtained by plugging into the functional of interest the NPMLE of the mixing distribution in a convolution model with the Laplace kernel density. We study the behaviour of the plug-in NPMLE to answer the question of whether there exist integral linear functionals of the mixing distribution that can be consistently estimated by the maximum likelihood method at $n^{-1/2}$ -rate, the recentered and \sqrt{n} -rescaled version of the plug-in NPMLE being asymptotically normal with zero mean and minimum variance. Situations are distinguished in which the plug-in NPMLE is consistent at parametric rate and asymptotically efficient, albeit the mixing distribution itself can typically be estimated only at slower rates, from those in which there exists no regular sequence of estimators that can be asymptotically efficient. The model is described hereafter and the problem formally stated.

Model description Let X be a real-valued random variable (r.v.) with distribution P_0 defined, for every Borel set B on the real line, by the mapping $B \mapsto P_0(B) := \mathbb{P}(X \in B)$. Suppose that P_0 is dominated by Lebesgue measure λ on \mathbb{R} , with probability density function (p.d.f.) $p_0 := dP_0/d\lambda$. Let X satisfy the relationship

$$X = Y + Z, \quad (1)$$

where Y and Z are (stochastically) independent, *unobservable* random variables such that Y has *unknown* cumulative distribution function (c.d.f.) F_0 and Z has the *standard classical* Laplace¹ or double exponential² distribution with scale parameter $s = 1$, in symbols, $Z \sim \text{Laplace}(0, 1)$, whose density k has expression

$$k(z) = \frac{1}{2} e^{-|z|}, \quad z \in \mathbb{R}. \quad (2)$$

The density p_0 is therefore the convolution of F_0 and k or a location mixture of Laplace densities with mixing distribution F_0 supported on a subset $\mathcal{Y} \subseteq \mathbb{R}$, where \mathcal{Y} stands for a support of F_0 , see, e.g., Billingsley (1995), p. 23,

$$p_0(x) \equiv p_{F_0}(x) = \int_{\mathcal{Y}} k(x-y) dF_0(y) = \int_{\mathcal{Y}} \frac{1}{2} e^{-|x-y|} dF_0(y), \quad x \in \mathbb{R}.$$

For ease of exposition, the density of the standard Laplace distribution is considered as a kernel, but the density of *any* Laplace distribution centered at zero, with *known* scale parameter $s > 0$, in symbols, $Z \sim \text{Laplace}(0, s)$, whose variance is $\sigma_Z^2 = 2s^2$,³

¹ It is also known as the *first law of Laplace* to distinguish it from the *second law of Laplace*, as the normal distribution is sometimes called. It was named after Pierre-Simon Laplace (1749–1827) who, in 1774 (Laplace 1774), obtained $e^{-|z-\theta|}/2$, for $z, \theta \in \mathbb{R}$, as the density of the distribution whose likelihood is maximized when the location parameter θ is equal to the sample median.

² It is so called because it is formed by reflecting the exponential distribution around its mean.

³ To see that $\sigma_Z^2 = 2s^2$, one can take into account that, if V_1 and V_2 are independent r.v.'s, identically distributed as an exponential with parameter $1/s$, in symbols, $V_j \sim \text{Exp}(1/s)$, $j = 1, 2$, then $V_1 - V_2$ has a Laplace $(0, s)$ distribution. Consequently, $\sigma_Z^2 = 2\sigma_{V_1}^2 = 2(1/s)^{-2} = 2s^2$.

could be employed. Assume that X_1, \dots, X_n constitute a random sample from p_0 . Every X_i then satisfies

$$X_i = Y_i + Z_i, \quad i = 1, \dots, n, \quad (3)$$

where Y_1, \dots, Y_n and Z_1, \dots, Z_n are independent samples from the distributions with c.d.f. F_0 and p.d.f. k , respectively. The r.v.'s Y_1, \dots, Y_n are independent and identically distributed (i.i.d.) as Y and Z_1, \dots, Z_n are independent copies of Z . Realizations x_1, \dots, x_n of the noisy sample data X_1, \dots, X_n are observed instead of outcomes of the uncorrupted r.v.'s Y_1, \dots, Y_n . The r.v.'s Z_1, \dots, Z_n represent additive errors and their distribution is called the *error distribution*. In this model, the variable of interest Y cannot be directly observed and empirical access is limited to the sum of Y and the “noise” Z . Therefore, estimating the distribution function F_0 of Y , or related quantities like the p.d.f. f_0 (if it exists), based on a sample from P_0 , accounts for solving a particular *inverse* problem, called *deconvolution*, which consists in reconstructing (estimating) F_0 from indirect noisy observations X_1, \dots, X_n drawn from $P_0 \equiv P_{F_0}$, the latter being the image of F_0 under a known transformation that has to be “inverted”. As remarked in Groeneboom and Wellner (1992), p. 4, as well as in Bolthausen *et al.* (2002), p. 363, the problem can be viewed as a *missing data problem*: the complete observations would consist of the independent pairs $X_i^0 := (Y_i, Z_i)$, with $X_i^0 \sim Q := F_0 \times F_Z$,⁴ but part of the data is missing and only outcomes or realizations of the sums $X_i = T(X_i^0) := (Y_i + Z_i) \sim P_0 \equiv QT^{-1}$, viewed as transformations of the X_i^0 's through the function T , are observed.

The statistical model described by relationship (1), with a zero-mean Laplace measurement error r.v. Z independent of Y , is a special case of the *classical error model* $X = Y + Z$, in which X is a measurement of Y in the usual sense, Z has zero mean and is independent of Y , see, *e.g.*, Buzas *et al.* (2005), p. 733, and the references therein. Measurement errors with possibly different structures occur in nearly every discipline from medical statistics to astronomy and econometrics, *cf.* the monographs of Fuller (1987) and Buonaccorsi (2010). Furthermore, the Laplace distribution finds applications in a variety of disciplines, from image and speech recognition to ocean engineering, see Kotz *et al.* (2001), chpts. 7–10, pp. 343–397. An application to quality control of the classical Laplace measurement error model is outlined hereafter.

Application to steam generator inspection An application of the Laplace measurement error model to steam generator inspection and testing is described herein, see Easterling (1980) and Sollier (2017) for more details. Steam generators of pressurized water reactors contain many tubes through which heated water flows. For a variety of reasons, such as corrosion-induced wastage, the steam generator tube integrity can be degraded, the walls becoming thinned or cracked. Leaks may occur during normal operating conditions, thus requiring the plant to be shut down. In order to develop an inspection plan, a statistical model for tube degradation is considered. Experimental data evidenciate that measurements are affected by heavy-tailed and biased errors that can be represented by a r.v. E following a Laplace distribution with mean $\mu > 0$ and

⁴ The symbol F_Z denotes the c.d.f. of Z , that is, $F_Z(z) = \int_{-\infty}^z k(u) du$, with density k as in (2).

scale parameter $s > 0$, in symbols, $E \sim \text{Laplace}(\mu, s)$, with density

$$f_E(e) = \frac{1}{2s} \exp(-|e - \mu|/s), \quad e \in \mathbb{R}.$$

Denoted by D the *actual* degradation (extent of thinning) of a tube, expressed as a percentage of the initial tube wall thickness, the *measured* degradation M is modeled as

$$M = D + E,$$

where E is supposed to be independent of D . Assuming that the scale parameter s is known and the distribution of D possesses probability density function, say f_D , the interest is, in the first place, in estimating the p.d.f. of D , based on i.i.d. observations M_1, \dots, M_n drawn from the distribution of M . An exponential distribution for D , with scale parameter $\tau > 0$, in symbols, $D \sim \text{Exp}(1/\tau)$, whose density has expression

$$f_D(d) = \frac{1}{\tau} \exp(-d/\tau), \quad d > 0,$$

provides an exponential-double exponential model for the actual degradation M , which has proved to have an adequate fit on experimental data. Statistical procedures for fitness-for-service assessment are described in Carroll (2017).

Asymptotic efficiency of the NPMLE for linear functionals of the mixing distribution

For many purposes, interest can lie in only few aspects of the distribution of Y , key features of which can be represented as linear functionals of F_0 . In what follows, symbols F_0 and F will be used to indicate probability measures (p.m.'s) on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$, where $\mathcal{B}(\mathcal{Y})$ denotes the Borel σ -field on \mathcal{Y} , as well as the corresponding cumulative distribution functions, the correct meaning being clear from the context. Letting

$$\mathcal{P} := \{\text{all p.m.'s } F \text{ on } (\mathcal{Y}, \mathcal{B}(\mathcal{Y}))\}$$

be the collection of all probability measures F on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$, a functional is a mapping $\psi : \mathcal{P} \rightarrow \mathbb{R}$ that maps every $F \in \mathcal{P}$ to a real number $\psi(F)$. The focus is on estimating *integral* linear functionals

$$F \mapsto \psi(F) := \int_{\mathcal{Y}} a(y) dF(y) \quad (4)$$

at the “point” F_0 , where the function $a \in L^1(F_0)$ is given. The following examples illustrate choices of a for some common statistical functionals.

- *Distribution function at a point* If, for some fixed $y_1 \in \mathbb{R}$, the function $a(\cdot) = 1_{(-\infty, y_1]}(\cdot)$, then $\psi(F_0) = \int_{\mathcal{Y}} 1_{(-\infty, y_1]}(y) dF_0(y) = F_0(y_1)$ is the c.d.f. of Y at the point y_1 .
- *Probability of an interval* If, for fixed points $y_1, y_2 \in \mathbb{R}$, the function $a(\cdot) = 1_{(y_1, y_2]}(\cdot)$, then $\psi(F_0) = \int_{\mathcal{Y}} 1_{(y_1, y_2]}(y) dF_0(y) = F_0(y_2) - F_0(y_1) = \mathbb{P}(y_1 < Y \leq y_2)$ is the probability of the interval $(y_1, y_2]$.

- *Mean* If $a(\cdot) = \text{id}_{\mathcal{Y}}(\cdot)$ is the identity function on \mathcal{Y} , then $\psi(F_0) = \int_{\mathcal{Y}} y \, dF_0(y) = \mathbb{E}Y$ is the expected value of Y which, for *any* kernel density k (not necessarily the Laplace) with zero mean, $\mathbb{E}Z = 0$, is equal to $\mathbb{E}X$: in fact, from the relationship $X = Y + Z$ in (1), it follows that $\mathbb{E}X = \mathbb{E}Y + \mathbb{E}Z = \mathbb{E}Y$ by linearity of the expected value.
- *rth moment* If, for any positive integer r , the function $y \mapsto a(y) = y^r$, then $\psi(F_0) = \int_{\mathcal{Y}} y^r \, dF_0(y) = \mathbb{E}Y^r$ is the r th moment of Y .
- *Moment generating function* If, for some fixed point $t \in \mathbb{R}$ such that $0 < |t| < 1$, the mapping $y \mapsto a(y) = e^{ty}$, then $\psi(F_0)$ coincides with the moment generating function (m.g.f.) of F_0 at t , denoted by $M_{F_0}(t)$ or $M_Y(t)$, that is, $\psi(F_0) = \int_{\mathcal{Y}} e^{ty} \, dF_0(y) = M_{F_0}(t)$. Some features of the mixing distribution F_0 , like the mean or the variance, can be expressed in terms of the derivatives of the corresponding m.g.f. M_{F_0} evaluated at zero. Therefore, in principle, results for estimating aspects of F_0 can be obtained as by-products of the inference on M_{F_0} .

A standard and principled method for pointwise estimation of linear functionals consists in plugging *the*⁵ NPMLE \hat{F}_n of F_0 into $\psi(\cdot)$ to obtain the *plug-in estimator*

$$\psi(\hat{F}_n) := \psi(F) \Big|_{F=\hat{F}_n}.$$

A NPMLE \hat{F}_n of F_0 is a measurable function of the observations X_1, \dots, X_n taking values in \mathcal{P} , which is not necessarily uniquely defined by the relationship

$$\hat{F}_n \in \arg \max_{F \in \mathcal{P}} \frac{1}{n} \sum_{i=1}^n \log p_F(X_i),$$

equivalently written as

$$\hat{F}_n \in \arg \max_{F \in \mathcal{P}} \mathbb{P}_n \log p_F,$$

where

$$p_F(\cdot) := \int_{\mathcal{Y}} k(\cdot - y) \, dF(y)$$

is the generic location mixture of Laplace densities with mixing distribution F and

$$\mathbb{P}_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

is the empirical probability measure associated with the random sample X_1, \dots, X_n , namely, the discrete uniform distribution on the sample values that puts mass $1/n$ on each one of the observations. In the sequel, for a measurable function $f : \mathcal{X} \rightarrow \mathbb{R}$, where $\mathcal{X} \subseteq \mathbb{R}$ is specified at the different occurrences, the notation $\mathbb{P}_n f$ is used to abbreviate the empirical average $n^{-1} \sum_{i=1}^n f(X_i)$. Analogously, $P_0 f$ is used in lieu of $\int f \, dP_0$. Hereafter, unless it is necessary within the context to specify the integral domain, integration is understood to be performed over the entire natural domain

⁵ Uniqueness of \hat{F}_n is not guaranteed, it is therefore with an abuse of language that we refer to *the* NPMLE throughout the article.

of the integrand. Throughout the article, the probability measure \mathbf{P} stands for P_0^n , the joint law of the first n coordinate projections of the infinite product probability measure $P_0^{\mathbb{N}}$. Sequences of random variables are meant to convergence (in law or in probability) as the sample size n grows indefinitely large (as $n \rightarrow +\infty$).

Historical and conceptual background, overview of the results The deconvolution problem has been intensively studied over the last thirty years. There exists a vast literature on *density* deconvolution, which accounts for reconstructing/estimating the density f_0 of Y (if it exists) that satisfies the equation

$$p_0(x) = \int_{\mathcal{Y}} k(x-y)f_0(y) dy, \quad x \in \mathbb{R},$$

wherein the kernel density k (not necessarily a Laplace) is assumed to be known, based on outcomes x_1, \dots, x_n of i.i.d. r.v.'s X_1, \dots, X_n as in (3). We cite key articles of the early 90's like Carroll and Hall (1988), Stefanski and Carroll (1990), Fan (1991)⁶ that have been ground-breaking and have had a great impact on the area of measurement error, setting the general framework for attacking measurement error/deconvolution problems and developing an approach based on Fourier inversion techniques to construct a deconvolution kernel density estimator for recovering the density of the latent distribution, meanwhile showing how difficult it is to account for measurement errors: in fact, the smoother the error distribution, the stronger its confounding effect on the latent distribution, hence, the slower the optimal attainable rate of convergence for its estimators.

Far less instead seems to be known about *distribution function* deconvolution, keynote contributions, also based on Fourier inversion techniques, being those of Hall and Lahiri (2008), Dattner *et al.* (2011), the former article containing an illuminating critical analysis of the background to the problem of distribution estimation in deconvolution problems. Since the focus of this article is on the behaviour of the NPMLE \hat{F}_n of the mixing distribution F_0 , attention is hereafter restricted to review the theory of non-parametric maximum likelihood estimation in deconvolution problems. In general mixture models, the NPMLE \hat{F}_n of F_0 is *discrete*, with at most $l \leq n$ support points, l being the number of distinct values of the data points, *cf.* Lindsay (1983). In deconvolution problems with continuous and symmetric (about the origin) kernels decreasing on $[0, +\infty)$, a NPMLE \hat{F}_n always exist (uniqueness is not guaranteed), see Groeneboom and Wellner (1992), Lemma 2.1, pp. 57–58; for kernels that are also strictly convex on $[0, +\infty)$, like the (standard) Laplace, the NPMLE \hat{F}_n is supported on the set of observation points $\{X_1, \dots, X_n\}$, so that the corresponding probability measure, still denoted by \hat{F}_n consistently with the notational convention adopted throughout, is concentrated on the range of the data points $[X_{1:n}, X_{n:n}]$,⁷

$$\hat{F}_n([X_{1:n}, X_{n:n}]) = 1,$$

see *ibid.*, Corollary 2.2 and Corollary 2.3, p. 59 and p. 60, respectively. Consistency of \hat{F}_n at a *continuous* distribution function F_0 is proved in *ibid.*, § 4.2, pp. 79–81.

⁶ For a recent reference list, see also Davidian *et al.* (2014).

⁷ Following a common notational convention, we denote by $X_{r:n}$ the r th order statistic.

More is known about the *one*-parameter (location only) Laplace model, which can be viewed as a degenerate mixture with point mass mixing distribution at some fixed $\theta \in \mathbb{R}$, the Dirac measure at θ .⁸ A simple maximization argument to find a MLE of the location parameter θ , denoted by $\hat{\theta}_n$, is given in Norton (1984): the *sample median* is a MLE which is an *M-estimator*, see Huber (1967), solving the equation $\sum_{i=1}^n \text{sign}(X_i - \theta) = 0$.⁹ Therefore, a MLE exists, but may not be unique: if n is odd, that is, $n = 2m + 1$ for some $m \in \mathbb{N}$, then the sample median is uniquely defined as the middle observation $X_{m+1:n}$, while, if $n = 2m$ is even, then there are two middle observations $X_{m:n}$ and $X_{m+1:n}$ so that, in principle, any value in the interval $[X_{m:n}, X_{m+1:n}]$ could be chosen, even if the *canonical median* $(X_{m:n} + X_{m+1:n})/2$, the average of the middle observations, is typically used in practice. Therefore,

$$\hat{\theta}_n = \begin{cases} X_{m+1:n}, & \text{for } n = 2m + 1, \\ \frac{1}{2}(X_{m:n} + X_{m+1:n}), & \text{for } n = 2m. \end{cases} \quad m \in \mathbb{N},$$

The sample median $\hat{\theta}_n$ is a MLE and is *asymptotically efficient*, that is, consistent and, when recentered at θ and \sqrt{n} -rescaled, asymptotically normal with zero mean and variance equal to one, which is the information lower bound corresponding to the amount of information in a single observation,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

where “ $\xrightarrow{\mathcal{L}}$ ” denotes convergence in law. This has been established by Daniels (1961) who, motivated by the non-differentiability at zero of the (standard) Laplace density, proved a general theorem on the asymptotic efficiency of the MLE under conditions not involving the second and higher-order derivatives of the likelihood function. Even though, as later on noted by Huber (1967), a crucial step has been overlooked in Daniels’ proof, the assertion remains valid.

In this article, we study the behaviour of the plug-in NPML $\psi(\hat{F}_n)$ to answer the question of whether there exist integral linear functionals $\psi(F_0)$ that can be consistently estimated by the maximum likelihood method at $n^{-1/2}$ -rate and for which

$$\sqrt{n}(\psi(\hat{F}_n) - \psi(F_0))$$

is *asymptotically efficient*, in the sense of Definition 2.8 in Bolthausen *et al.* (2002), p. 349, that is, asymptotically normal with zero mean and minimum variance. In fact, also in non-parametric problems estimation can be performed at $n^{-1/2}$ -rate and, in general mixture models $\int_{\mathcal{Y}} k(\cdot | y) dF_0(y)$, where k is any (not necessarily the Laplace) kernel density, there may exist linear functionals of F_0 that are estimable at parametric rate, even if F_0 itself can be pointwise estimated only at slower rates. Fundamental contributions developing the theory of information bounds are van der Vaart (1991), van der Vaart (1998), chpt. 25, pp. 358–432, Bolthausen *et al.* (2002), Part III, pp. 331–457, Groeneboom and Wellner (1992), with emphasis on non-parametric maximum likelihood estimation, and van de Geer (2000), chpt. 11, pp. 211–246, with a

⁸ The Dirac measure at θ , denoted by $\delta_\theta(\cdot)$, is defined on the Borel sets $B \in \mathcal{B}(\mathbb{R})$ by $\delta_\theta(B) = 1$, 0 if $B \ni \theta$ or $B \not\ni \theta$, respectively.

⁹ The *sign-function* is defined as $\text{sign}(x) = -1, 0, 1$ if $x < 0, x = 0$ or $x > 0$, respectively.

focus on asymptotic efficiency of the NPMLE in mixture models. To exemplify the issue, consider estimating the mean functional $\psi(F_0) = EY$ which, in a mixture model such that $E(X | Y) = Y$, is equal to EX . Then, the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ is a $n^{-1/2}$ -consistent and, after \sqrt{n} -rescaling, asymptotically normal estimator of EY , but may not be a MLE; furthermore, it does not take into account the information that the sampling density is a mixture. On other side, the MLE may be $n^{-1/2}$ -consistent and converge to a normal distribution with smaller variance than that of the sample mean. This is the case for the sample median in the single-parameter (location only) Laplace model, see van der Vaart (1998), Example 7.8 on location models, p. 96. Surprisingly, little is known in general about the asymptotic behaviour of the plug-in NPMLE for linear functionals in Laplace convolution mixtures, even if only for estimating the mean functional. Although the topic is useful, existence of this gap can be partially explained by the fact that the Laplace or double-exponential distribution is not an exponential family model so that standard results may not be valid or immediately available from the theory of exponential families.

In order to investigate whether integral linear functionals of the mixing distribution in a convolution model with the Laplace kernel density are estimable at $n^{-1/2}$ -rate, we appeal to van der Vaart's differentiability theorem, see Theorem 3.1 in van der Vaart (1991), p. 183, a general result that allows for a unified treatment of the information lower bound theory based on the concept of a differentiable functional, see, for a definition of the latter, display (2.2) in van der Vaart (1991), p. 180, or Definition 1.10 in Bolthausen *et al.* (2002), p. 343. The differentiability theorem characterizes differentiable functionals and, by combining the description of the set of differentiable functionals with a result stating that the existence of *regular* estimator sequences for a functional implies its differentiability, provides a way to distinguish situations in which the functional of interest is estimable at $n^{-1/2}$ -rate from situations in which this is not the case, see van der Vaart (1998), p. 365, for a definition of a regular estimator sequence. A necessary and sufficient condition for differentiability of a (not necessarily linear) functional is that its gradients are contained in the range of the adjoint of the score operator, where the score operator can be viewed as a derivative (in quadratic mean) of the map $F \mapsto P_F$, see (3.6) in van der Vaart (1991), p. 183, or (25.29) in § 25.5 of van der Vaart (1998), p. 372. As previously mentioned, differentiability is necessary for regular estimability of a functional or, equivalently, for the existence of regular estimator sequences, see Theorem 2.1 of van der Vaart (1991), p. 181, so that if the functional is not differentiable, then there exists no regular estimators and estimation at $n^{-1/2}$ -rate is impossible. Interestingly, for real-valued functionals, the differentiability condition is equivalent to having positive efficient information, see Theorem 4.1 of van der Vaart (1991), pp. 186–187. We find that the differentiability condition fails for integral linear functionals of the mixing distribution in a convolution model with the Laplace density, this implying that there exists no estimator sequence for $\psi(F_0)$ that is regular at F_0 and estimation at $n^{-1/2}$ -rate is impossible.

Organization The rest of the article is organized as follows. The main results are presented in Sect. 2, which is split into two parts. In the first one, asymptotic efficiency of the plug-in NPMLE for integral linear functionals of the mixing distribution in a

convolution model with the (one-sided) exponential kernel density is analysed and set in the affirmative. Construction of interval estimators and tests based on a Studentized version of the plug-in NPMLE, when the asymptotic variance is consistently estimated, is revisited. Conditions for extending results to non-linear functionals are discussed as a side-issue. In the second part, the focus is on asymptotically efficient estimation by the plug-in NPMLE for integral linear functionals of the mixing distribution in a convolution model with the double-exponential (Laplace) kernel density. It is shown that, except for the case of a degenerate mixing distribution at a single location point, maximum likelihood estimation completely fails, in the sense that no integral linear functional can be estimated at $n^{-1/2}$ -rate, which precludes any possibility for the NPMLE of being asymptotically efficient. Indeed, there exists no regular sequence of estimators for integral linear functionals of the mixing distribution that can be asymptotically efficient, therefore, estimation of linear functionals is impossible at $n^{-1/2}$ -rate. Final remarks and comments are exposed in Sect. 3. Proofs of the main results are deferred to the appendices: Appendix A reports the proof of the result for convolution with the exponential density, Appendix B reports the proof of the result for convolution with the Laplace density.

2 Main results

In this section, the main results of the article are presented. First, the case of a convolution model with an exponential kernel density is considered. Since, as previously noted, the Laplace density in (2) can be thought of as two exponential densities spliced together back-to-back, the positive half being a standard exponential density scaled by 1/2, it is reasonable to begin the analysis from the problem of asymptotically efficient non-parametric maximum likelihood estimation of linear functionals in a convolution model with the *exponential* kernel density. A preliminary study of the one-sided problem, beyond being of interest in itself, is useful to attack the two-sided one by partially reducing it to the previously solved case; it may furthermore provide insight for a better understanding of the reasons why symmetrization leads to a failure of asymptotically efficient estimation of linear functionals in the *double-exponential* (Laplace) case.

Convolution with the exponential density In this paragraph, a standard exponential kernel density on $[0, +\infty)$ is considered. This gives rise to a one-sided mixture density generating the data,

$$p_0(x) \equiv p_{F_0}(x) = \int_{\mathcal{Y}} e^{-(x-y)} 1\{y \leq x\} dF_0(y), \quad x \in \mathcal{X},$$

where $\mathcal{Y} := \text{support}(F_0)$ is assumed, without loss of generality, to be a proper, left-closed subset of the real line, and $\mathcal{X} := [y_{\min}, +\infty)$ is the support of P_0 , with $y_{\min} := \min \mathcal{Y} > -\infty$. Proposition 1 below establishes that, under sufficient conditions, an integral linear functional $\psi(F_0)$ can be consistently estimated at $n^{-1/2}$ -rate by the plug-in NPMLE $\psi(\hat{F}_n)$, which, when recentered about the estimand and \sqrt{n} -rescaled,

is asymptotically normal and efficient. In stating hereafter the assumptions, Newton's notation (or the dot notation) for differentiation is adopted, that is, $\dot{a}(y) := da(y)/dy$.

Assumptions

- (A0) $a \in L^1(F_0)$,
(A1) $M_{F_0}(1) := \int_{\mathcal{Y}} e^y dF_0(y) < +\infty$,
(A2) $|\psi(\hat{F}_n) - \psi(F_0)| = o_{\mathbf{P}}(1)$,
(A3) (i) a is continuous on \mathcal{Y} ,
(ii) either \mathcal{Y} is compact or a is bounded on \mathcal{Y} ,
(iii) $F_0(y_{\min}) = 0$,
(A4) there exists \dot{a} on \mathcal{Y} and $\sup_{y \in \mathcal{Y}} |\dot{a}(y)| < +\infty$,
(A5) there exists a constant $0 < c_0 < +\infty$ such that

$$\sup_{y \in \mathcal{Y}} \left| \frac{d(\dot{a}(y)e^{-y})}{dF_0(y)} \right| \leq c_0.$$

Some remarks and comments on the above listed assumptions are in order. Except for Assumption (A2), which concerns the plug-in NPML $\psi(\hat{F}_n)$, all other assumptions involve the function a and/or the mixing distribution F_0 that jointly define the functional $\psi(F_0)$. Specifically, Assumption (A0) guarantees that $\psi(F_0)$ is well defined. Assumption (A1) ensures the existence of the moment generating function of F_0 at the point $t = 1$. Assumption (A2) requires consistency of $\psi(\hat{F}_n)$ at $\psi(F_0)$. If \hat{F}_n converges weakly to F_0 in P_0^n -probability, then parts (i) and (ii) of Assumption (A3) together imply Assumption (A2) because a is continuous and bounded on \mathcal{Y} . Sufficient conditions for \hat{F}_n to converge weakly to F_0 in the convolution model with the standard exponential kernel density on $[0, +\infty)$ are stated in Groeneboom and Wellner (1992), p. 86; see also Theorem 2.3 of Chen (2017), p. 54, for sufficient conditions for strong consistency of \hat{F}_n in general mixture models. Part (ii) of Assumption (A3) postulates that either \mathcal{Y} is a closed and bounded interval $[y_{\min}, y_{\max}]$ or \mathcal{Y} is a right-unbounded interval $[y_{\min}, +\infty)$ and a is bounded. Assumption (A4) requires a to be differentiable and bounded on \mathcal{Y} , which, in particular, accounts for a to be right-differentiable at y_{\min} , that is, $\dot{a}_+(y_{\min}) < +\infty$,¹⁰ and, in the case where $\mathcal{Y} = [y_{\min}, y_{\max}]$, to be also left-differentiable at y_{\max} , that is, $\dot{a}_-(y_{\max}) < +\infty$.¹¹ Assumption (A5) plays its role in the proof of Proposition 1 when bounding the worst possible sub-directions, see (18) in the Appendix A.

Proposition 1 *Under Assumptions (A0)–(A5), we have*

$$\sqrt{n}(\psi(\hat{F}_n) - \psi(F_0)) = \sqrt{n}\mathbb{P}_n b_{F_0} + o_{\mathbf{P}}(1) \left(\xrightarrow{\mathcal{L}} \mathcal{N}(0, \|b_{F_0}\|_{2, P_0}^2) \right), \quad (5)$$

where the mapping $b_{F_0} : \mathcal{X} \rightarrow \mathbb{R}$, defined as

$$x \mapsto b_{F_0}(x) := a(x) - \dot{a}(x) - \psi(F_0), \quad (6)$$

¹⁰ The right-derivative of a at y_{\min} , denoted by $\dot{a}_+(y_{\min})$, is defined as the one-sided limit $\lim_{y \rightarrow y_{\min}^+} [a(y) - a(y_{\min})]/(y - y_{\min})$ if it exists as a real number.

¹¹ The left-derivative of a at y_{\max} , denoted by $\dot{a}_-(y_{\max})$, is defined as the one-sided limit $\lim_{y \rightarrow y_{\max}^-} [a(y_{\max}) - a(y)]/(y_{\max} - y)$ if it exists as a real number.

is the efficient influence function, whose squared $L^2(P_0)$ -norm

$$\|b_{F_0}\|_{2,P_0}^2 := \int b_{F_0}^2 dP_0 \quad (7)$$

is the efficient asymptotic variance.

Proposition 1 establishes that, under sufficient conditions listed as Assumptions **(A0)**–**(A5)**, cf. Lemma 4.6 of van de Geer (2003), p. 461, and van der Geer (2000), p. 231, the plug-in NPMLE $\psi(\hat{F}_n)$ consistently estimates $\psi(F_0)$ at $n^{-1/2}$ -rate; furthermore, when recentered at $\psi(F_0)$ and \sqrt{n} -rescaled, it is asymptotically distributed as a zero-mean Gaussian, with variance attaining the lower bound given by the squared $L^2(P_0)$ -norm of the efficient influence function, which plays here the same role as the normalized score function for the case of independent sampling from a parametric model $\{p_\theta, \theta \in \Theta \subseteq \mathbb{R}^d\}$, $d \geq 1$,

$$I_\theta^{-1} \dot{\ell}_\theta,$$

where $\dot{\ell}_\theta(\cdot) = \partial[\log p_\theta(\cdot)]/\partial\theta$ is the *score function* of the model and $I_\theta = E[\dot{\ell}_\theta \dot{\ell}_\theta^\top]$ the Fisher information matrix for θ . In a parametric set-up, the minimum variance lower bound reduces to the Cramér-Rao bound, which states that the inverse of the Fisher information matrix I_θ^{-1} is a lower bound on the variance of any \sqrt{n} -rescaled unbiased estimator $T_n \equiv T_n(X_1, \dots, X_n)$ of θ , in symbols, $\text{var}(\sqrt{n}T_n) \geq I_\theta^{-1}$. Therefore, the counterpart of $\|b_{F_0}\|_{2,P_0}^2 = P_0 b_{F_0}^2$ is, by symmetry of $(I_\theta$ hence of) I_θ^{-1} ,

$$E[(I_\theta^{-1} \dot{\ell}_\theta)(I_\theta^{-1} \dot{\ell}_\theta)^\top] = I_\theta^{-1} E[\dot{\ell}_\theta \dot{\ell}_\theta^\top] I_\theta^{-1} = I_\theta^{-1} I_\theta I_\theta^{-1} = I_\theta^{-1}.$$

In general, considered a function ψ that maps Θ into \mathbb{R}^m , $m \geq 1$, and denoted by $\dot{\psi}_\theta$ the derivative of $\theta \mapsto \psi(\theta)$, the matrix $\dot{\psi}_\theta I_\theta^{-1} \dot{\psi}_\theta^\top$ is a lower bound on the variance of any \sqrt{n} -rescaled unbiased estimator of $\psi(\theta)$.

Even if a statement of the result in Proposition 1 appears in Lemma 4.6 of van der Geer (2003), p. 461, as far as we are aware, a complete derivation of the assertion is not available in the literature, cf. also van der Geer (2000), p. 231, so the proof reported in the Appendix A might prove helpful. The underlying idea is outlined hereafter. A NPMLE \hat{F}_n solves the likelihood equation for every path $t \mapsto F_t$, with $dF_t := (1 + th_F) dF$, starting at a fixed point F corresponding to $t = 0$ and direction h_F such that $\int h_F dF = 0$, that is, for every parametric sub-model which passes (at $t = 0$) through it. In symbols,

$$\left. \frac{d}{dt} \mathbb{P}_n \log p_{\hat{F}_{n,t}} \right|_{t=0} = 0. \quad (8)$$

For ease of notation, let

$$A_F h_F(x) := \int_{\mathcal{Y}} h_F(y) \frac{e^{-(x-y)}}{p_F(x)} 1\{y \leq x\} dF(y).$$

Equation (8) reduces to $\mathbb{P}_n b_{\hat{F}_n} = 0$, where $b_{\hat{F}_n}(\cdot) = A_{\hat{F}_n} h_{\hat{F}_n}(\cdot)$ is the score function (at $t = 0$) in an “information loss model”, see, e.g., § 25.5.2 in van der Vaart (1998), pp. 374–375. If \hat{F}_n dominates F_0 , which, however, is seldom true, then $-P_0 b_{\hat{F}_n} = \psi(\hat{F}_n) - \psi(F_0)$ so that $\psi(\hat{F}_n) - \psi(F_0) = (\mathbb{P}_n - P_0) b_{\hat{F}_n}$. Asymptotic equicontinuity arguments

then yield that $(\mathbb{P}_n - P_0)b_{\hat{F}_n} = (\mathbb{P}_n - P_0)b_{F_0} + o_{\mathbf{P}}(n^{-1/2}) = \mathbb{P}_n b_{F_0} + o_{\mathbf{P}}(n^{-1/2})$ because $P_0 b_{F_0} = 0$, namely, the score has zero mean. So, $\sqrt{n}(\psi(\hat{F}_n) - \psi(F_0)) = \sqrt{n}\mathbb{P}_n b_{F_0} + o_{\mathbf{P}}(1)$. Asymptotic normality follows. The reader is referred to Sect. 11.2 of van de Geer (2000), pp. 211–246, for a more comprehensive treatment of the topic taking into account technical difficulties to which it cannot be here dedicated the necessary space.

Remark 1 For simplicity, a convolution model with an exponential kernel density on $[0, +\infty)$ having intensity $\lambda = 1$ has been considered (we warn the reader of the clash of notation with the symbol λ previously used to denote Lebesgue measure on the real line), but, as revealed by an inspection of the proof of Proposition 1, the assertion holds true for every $\lambda > 0$.

Remark 2 Part (i) of Assumption **(A3)** requires a to be continuous on \mathcal{Y} , which is not true for indicator functions, therefore Proposition 1 does not apply to pointwise estimation of the c.d.f. F_0 nor to the estimation of the probability of an interval, so that it cannot be concluded that these functionals are estimable at $n^{-1/2}$ -rate by the corresponding plug-in NPMLE's. Indeed, F_0 can be *pointwise* estimated only at $n^{-1/3}$ -rate, see Groeneboom and Wellner (1992), p. 121. Part (ii) of Assumption **(A3)** and Assumption **(A4)** require that both a and \dot{a} are bounded on \mathcal{Y} , which, for example, may not be true for the functions y^r and e^{ty} that define the r th moment and the moment generating function of Y at the point t , respectively: in fact, both y^r and e^{ty} , as well as their first derivatives, are continuous on the half-line $[y_{\min}, +\infty)$, but not bounded therein. Nonetheless, boundedness can be retrieved by restriction to a compact domain. Therefore, if, besides Assumptions **(A0)**–**(A2)** and **(A5)**, it also holds that F_0 has compact support, then, by Proposition 1, it can be concluded that EY^r and $M_{F_0}(t)$ are consistently and efficiently estimated at $n^{-1/2}$ -rate by their respective plug-in NPMLE's.

Although Proposition 1 asserts that certain integral linear functionals can be consistently estimated at $n^{-1/2}$ -rate by the plug-in NPMLE $\psi(\hat{F}_n)$, which, when recentered at $\psi(F_0)$ and \sqrt{n} -rescaled, is asymptotically normal and efficient, two orders of problems may arise that can make it difficult to employ the result for statistical inference:

- a) computation of the NPMLE \hat{F}_n ,
- b) dependence of the variance $\|b_{F_0}\|_{2, P_0}^2$ on the unknown sampling distribution P_0 .

As for the former difficulty, although the NPMLE \hat{F}_n can be found by a one-step procedure computing the slope of the convex minorant of a certain function, *cf.* Groeneboom and Wellner (1992), pp. 62–63 (see also Vardi (1989) for a different approach), as a by-product of Theorem 11.8 of van de Geer (2000), p. 217, which the assertion of Proposition 1 relies on, the recentered and \sqrt{n} -rescaled plug-in NPMLE $\sqrt{n}(\psi(\hat{F}_n) - \psi(F_0))$ is equivalent, in the sense of being asymptotically approximable, up to an $o_{\mathbf{P}}(1)$ -error term, by the empirical average of the efficient influence function. This is part of a general issue concerning the fact that sequences of efficient estimators for functionals are asymptotically approximable by an empirical average of the

efficient influence function, see, *e.g.*, Lemma 2.9 in Bolthausen *et al.* (2002), p. 349. In fact, set the position

$$\tilde{\psi}_n := \frac{1}{n} \sum_{i=1}^n [a(X_i) - \dot{a}(X_i)] = \mathbb{P}_n(a - \dot{a}) \quad (9)$$

and noted that, from the definition of b_{F_0} in (6), the term $\mathbb{P}_n b_{F_0}$ appearing in (5) writes as $\tilde{\psi}_n - \psi(F_0)$, we have

$$\begin{aligned} \sqrt{n}(\psi(\hat{F}_n) - \psi(F_0)) &= \sqrt{n}\mathbb{P}_n b_{F_0} + o_{\mathbf{P}}(1) \\ &= \sqrt{n}(\tilde{\psi}_n - \psi(F_0)) + o_{\mathbf{P}}(1) \left(\xrightarrow{\mathcal{L}} \mathcal{N}(0, \|b_{F_0}\|_{2, P_0}^2) \right). \end{aligned}$$

Thus, both $\sqrt{n}(\psi(\hat{F}_n) - \psi(F_0))$ and $\sqrt{n}(\tilde{\psi}_n - \psi(F_0))$ are asymptotically normal and efficient. Moreover, estimators arising from $\tilde{\psi}_n$ may coincide with simple *naïve* estimators. For example,

- if $\psi(F_0) = \mathbb{E}Y$, from (9), for $(a - \dot{a})(y) = y - 1$, we get the estimator $\tilde{\psi}_n = \bar{X}_n - 1$, which is the one we would suggest considering that $\mathbb{E}Y = \mathbb{E}X - \mathbb{E}Z = \mathbb{E}X - 1$;
- if $\psi(F_0) = M_Y(t)$ for any fixed $t < 1$ such that $\int_{\mathcal{Y}} e^{ty} dF_0(y) < +\infty$, then the estimator derived from (9), for $(a - \dot{a})(y) = (1 - t)e^{ty}$, is $\tilde{\psi}_n = (1 - t) \times n^{-1} \sum_{i=1}^n e^{tX_i}$, which is the one we would suggest taking into account that $M_Y(t) = M_X(t)/M_Z(t) = (1 - t)M_X(t)$, where $M_Z(t) = (1 - t)^{-1}$, $t < 1$, is the m.g.f. of a standard exponential r.v. Z . So, letting $M_n(t) := n^{-1} \sum_{i=1}^n e^{tX_i}$, $t \in \mathbb{R}$, be the empirical m.g.f. for the random sample X_1, \dots, X_n , it turns out that $\tilde{\psi}_n = M_n(t)/M_Z(t)$.

As for the difficulty listed in b), by the plug-in approach, replacing the asymptotic variance $\|b_{F_0}\|_{2, P_0}^2$ with a consistent estimator S_n^2 leads to the following assertion.

Corollary 1 *Under the conditions of Proposition 1, if, in addition, $S_n^2 \xrightarrow{\mathbf{P}} \|b_{F_0}\|_{2, P_0}^2$, where “ $\xrightarrow{\mathbf{P}}$ ” denotes convergence in \mathbf{P}_0^n -probability, then*

$$\sqrt{n} \frac{\tilde{\psi}_n - \psi(F_0)}{S_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Replaced the efficient asymptotic variance in (7) with a consistent sequence of estimators, asymptotic normality of the Studentized version of $\tilde{\psi}_n$ allows to carry out *pointwise* inference on linear functionals by interval estimation or hypotheses testing constructing confidence intervals or tests, respectively. For every $0 < \alpha < 1$, let $z_{\alpha/2}$ be the $(1 - \alpha/2)$ -quantile of a standard normal distribution, *i.e.*, $\Phi(z_{\alpha/2}) = 1 - \alpha/2$, where $\Phi(\cdot)$ stands for the c.d.f. of a standard normal. Then,

$$C_n := [\tilde{\psi}_n - z_{\alpha/2} S_n / \sqrt{n}, \tilde{\psi}_n + z_{\alpha/2} S_n / \sqrt{n}], \quad \text{with } \mathbf{P}(C_n \ni \psi(F_0)) = 1 - \alpha + o(1),$$

is an approximate $(1 - \alpha)$ -level confidence interval for $\psi(F_0)$.

Remark 3 Asymptotic normality of the plug-in NPMLE for linear functionals of the mixing distribution can be employed to establish asymptotic normality for *non*-linear functionals. Suppose, for instance, that $F \mapsto \varphi(F)$ is defined as

$$\varphi(F) := (g \circ \psi)(F) = g(\psi(F)),$$

where the function $g : \mathbb{R} \rightarrow \mathbb{R}$ has non-zero derivative at $\psi(F_0)$ denoted by $\dot{g}(\psi(F_0))$. Asymptotic normality of $\sqrt{n}(\varphi(\hat{F}_n) - \varphi(F_0))$ then follows from asymptotic normality of $\sqrt{n}(\psi(\hat{F}_n) - \psi(F_0))$ by the delta method, see, *e.g.*, chpt. 3 in van der Vaart (1998), pp. 25–34. So, if the convergence in (5) takes place, then

$$\sqrt{n}(\varphi(\hat{F}_n) - \varphi(F_0)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\varphi^2), \quad \text{with } \sigma_\varphi := \dot{g}(\psi(F_0)) \|b_{F_0}\|_{2, P_0},$$

where efficiency of $\psi(\hat{F}_n)$ carries over into efficiency of $\varphi(\hat{F}_n)$, see *ibid.*, p. 386, for details.

Alternatively, set the position

$$R_n := \frac{\varphi(\hat{F}_n) - \varphi(F_0)}{\psi(\hat{F}_n) - \psi(F_0)},$$

under the condition

$$|R_n - 1| = o_{\mathbf{P}}(1),$$

which requires that, in probability, $\varphi(\hat{F}_n) - \varphi(F_0)$ behaves asymptotically as $\psi(\hat{F}_n) - \psi(F_0)$, after \sqrt{n} -rescaling, the two differences have the same limiting distribution. In fact, if the convergence in (5) takes place, then Slutsky's lemma implies that

$$\begin{aligned} \sqrt{n}(\varphi(\hat{F}_n) - \varphi(F_0)) &= [(R_n - 1) + 1] \sqrt{n}(\psi(\hat{F}_n) - \psi(F_0)) \\ &= (o_{\mathbf{P}}(1) + 1) \sqrt{n}(\psi(\hat{F}_n) - \psi(F_0)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \|b_{F_0}\|_{2, P_0}^2), \end{aligned}$$

see also the Remark of van de Geer (2000) on p. 223.

Convolution with the double-exponential (Laplace) density In this paragraph, the case of main interest of the article concerning asymptotically efficient maximum likelihood estimation of linear functionals of the mixing distribution in a convolution model with the (standard) Laplace kernel density is considered. It has been recalled in Sect. 1 that, for a *one*-parameter θ (location only) Laplace model, the sample median $\hat{\theta}_n$ is a MLE, consistent and asymptotically efficient, even if, for small sample sizes, it may not be the best estimator to use because there exist other unbiased estimators with smaller variances, which are therefore more efficient, see, *e.g.*, Remark 2.6.2 in Kotz *et al.* (2001), p. 82. More precisely, for a sample of odd size n from a general Laplace(θ, s) distribution, the variance of $\hat{\theta}_n$ is equal to

$$\frac{1}{4(n+2)[k(0)/s]^2} = \frac{s^2}{(n+2)},$$

where $k(\cdot)$ is the density of a standard Laplace distribution as defined in (2), while the asymptotic variance is equal to

$$\frac{1}{4n[k(0)/s]^2} = \frac{s^2}{n}.$$

It is just the case to observe that also the sample mean \bar{X}_n is asymptotically normal with mean θ , but the asymptotic relative efficiency (ARE) of the median to the mean, namely, the ratio of the variance of the sample mean to the asymptotic variance of the sample median equals 2:

$$\frac{2s^2/n}{s^2/n} = 2.$$

On a side note, we recall that, for any function g differentiable at θ , with derivative $\dot{g}(\theta)$, the plug-in MLE $g(\hat{\theta}_n)$ is also asymptotically efficient, with

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, [s\dot{g}(\theta)]^2),$$

see, *e.g.*, Lehmann and Casella (1998), p. 440.

In what follows, we aim at giving results on asymptotically efficient maximum likelihood estimation of linear functionals of the mixing distribution, beyond the case of a degenerate mixing distribution localized at a point θ on the real line. As recalled in Sect. 1, in the deconvolution problem with the Laplace kernel density, a NPML \hat{F}_n always exists and consistency at a *continuous* distribution function F_0 holds, but little is known about the asymptotic behaviour of the plug-in NPML for linear functionals. The following proposition states that, except for the above recalled degenerate case, estimation of integral linear functionals at $n^{-1/2}$ -rate is impossible.

Proposition 2 *Let F_0 be a non-degenerate probability measure supported on \mathcal{Y} . Let $\psi(F_0)$ be any integral linear functional evaluated at F_0 . Then, there exists no estimator sequence for $\psi(F_0)$ that is regular at F_0 .*

Some comments on Proposition 2, whose proof is deferred to the Appendix B, are in order. It states that no integral linear functional is estimable at parametric rate, in particular, by the plug-in NPML $\psi(\hat{F}_n)$. One can thus expect estimation, performed by any method, only at slower rates and, possibly, with a non-Gaussian limiting distribution, even if the theorem we invoke to establish Proposition 2 does not give any indication about which rates to expect when estimation at $n^{-1/2}$ -rate fails, an issue that requires further investigation. A related open question concerns the possible extension of the negative result of Proposition 2 to convolution models with general kernel densities that are symmetric about zero, but not differentiable at it, a feature that seems to play a crucial role in causing failure of estimation at parametric rate. To sum-up, only in the case of a degenerate mixing distribution at a point θ , the MLE $\hat{\theta}_n$ is asymptotically efficient for the location parameter and the plug-in MLE $g(\hat{\theta}_n)$ is asymptotically efficient for any $g(\theta)$, with g differentiable at θ .

3 Final remarks

In this article, we have studied asymptotically efficient maximum likelihood estimation of linear functionals of the mixing distribution in a standard additive measurement error model, when the error has either the exponential or Laplace distribution. In the former case, the plug-in NPMLE of certain linear functionals is \sqrt{n} -consistent, asymptotically normal, efficient and equivalent to naïve estimators that are empirical averages of a given transformation of the observations. In the latter case, instead, even if the kernel is generated by symmetrization about the origin of the exponential density, left aside the degenerate case of a single Laplace model in which the MLE, the sample median, is asymptotically efficient for the location parameter, asymptotically efficient estimation of linear functionals completely fails, in the sense that estimation at $n^{-1/2}$ -rate is impossible for linear functionals of non-degenerate mixing distributions. An open question then is whether this negative result extends to general kernel densities symmetric about zero, but not differentiable at zero, a feature that seems to play a crucial role in causing the failure.

Appendix A

In this section, we present the proof of Proposition 1 on the asymptotic efficiency of the plug-in NPMLE for integral linear functionals of the mixing distribution in a convolution model with the *exponential* kernel density on $[0, +\infty)$.

Proof of Proposition 1 We appeal to Theorem 2.1 of van de Geer (1997), p. 21 (see also Theorem 11.8 of van de Geer (2000), pp. 217–220, for a slightly more general version) and, in showing that Conditions 1–4 are satisfied, we follow the indications exposed in Sect. 3, *ibid.*, pp. 24–27.

Verification of Condition 1. (Consistency and rates).

Under Assumption **(A1)** that $M_{F_0}(1) < +\infty$, the MLE $p_{\hat{F}_n}$ converges in the Hellinger distance d_H , defined as the L^2 -distance between the square-root densities, at the rate $O_{\mathbf{P}}(n^{-1/3})$. In symbols, for $\delta_n := n^{-1/3}$,

$$d_H(p_{\hat{F}_n}, p_0) := \|p_{\hat{F}_n}^{1/2} - p_0^{1/2}\|_2 = O_{\mathbf{P}}(\delta_n).$$

The result can be obtained by applying Theorem 7.4 in van de Geer (2000), pp. 99–100, see also *ibid.*, p. 124. As a consequence, see, *e.g.*, Corollary 7.5, *ibid.*, p. 100,

$$\mathbb{P}_n \log \frac{2p_{\hat{F}_n}}{p_{\hat{F}_n} + p_0} = O_{\mathbf{P}}(\delta_n^2),$$

where $\delta_n^2 = o(n^{-1/2})$. Consistency of $\psi(\hat{F}_n)$ is guaranteed by Assumption **(A2)**.

Verification of Condition 2. (Existence of the worst possible sub-directions and efficient influence functions. Differentiability of ψ in a neighborhood of F_0).

For real numbers $M > M_{F_0}(1) > 0$ and $r > 0$, let

$$\mathcal{P}_0 := \{F \in \mathcal{P} : \text{support}(F) = \mathcal{Y}, M_F(1) < M, d_H(p_F, p_0) \leq r\} \quad (10)$$

be a Hellinger-type ball centered at p_0 with radius $r > 0$. For every $\alpha \in [0, 1)$ and $F \in \mathcal{P}_0$, let $F_\alpha := \alpha F + (1 - \alpha)F_0$. We prove

- (a) *existence of the worst possible sub-directions h_{F_α} such that $h_{F_\alpha} \in L^2(F_\alpha)$ and $\int h_{F_\alpha} dF_\alpha = 0$;*
- (b) *existence of the efficient influence functions $b_{F_\alpha} := A_{F_\alpha} h_{F_\alpha}$, where $A_{F_\alpha} h_{F_\alpha}(\cdot) := \mathbb{E}[h_{F_\alpha}(Y) | X = \cdot]$;*
- (c) *differentiability of ψ at F_α :*

$$A^* b_{F_\alpha}(Y) = a(Y) - \psi(F_\alpha) \quad a.s. [F_\alpha], \quad (11)$$

where $A^* b_{F_\alpha}(\cdot) := \mathbb{E}[b_{F_\alpha}(X) | Y = \cdot]$.

For every $\alpha \in [0, 1)$, we prove the existence of h_{F_α} such that the corresponding $b_{F_\alpha} = A_{F_\alpha} h_{F_\alpha}$ satisfies $A^* b_{F_\alpha}(y) = a(y) - \psi(F_\alpha)$ for F_α -almost all y 's. We proceed by first deriving the expression of b_{F_α} as a solution of (11) and then proving the existence of the corresponding worst possible sub-direction h_{F_α} as required in (a) and (b). The function b_{F_α} has to satisfy

$$A^* b_{F_\alpha}(y) := \mathbb{E}[b_{F_\alpha}(X) | Y = y] = \int_{\mathcal{X}} b_{F_\alpha}(x) e^{-(x-y)} 1\{x \geq y\} dx = a(y) - \psi(F_\alpha), \quad (12)$$

where $\mathcal{X} = [y_{\min}, +\infty)$, for F_α -almost all y 's. Differentiating both sides of (12) with respect to y , we get

$$\int_{\mathcal{X}} b_{F_\alpha}(x) e^{-(x-y)} 1\{x \geq y\} dx - b_{F_\alpha}(y) = \dot{a}(y). \quad (13)$$

Using constraint (12) in (13), we obtain that $a(y) - \psi(F_\alpha) - b_{F_\alpha}(y) = \dot{a}(y)$, whence $b_{F_\alpha}(y) = a(y) - \dot{a}(y) - \psi(F_\alpha)$. The solution is unique up to sets of F_α -measure zero. By an extension to \mathcal{X} , b_{F_α} is then defined as in (6).

- (a) *Existence of $h_{F_\alpha} \in L^2(F_\alpha)$ such that $\int h_{F_\alpha} dF_\alpha = 0$.*

Recall that

$$p_{F_\alpha}(x) := \int_{\mathcal{Y}} e^{-(x-y)} 1\{y \leq x\} dF_\alpha(y), \quad x \in \mathcal{X}.$$

Defined the function $I_{F_\alpha} : \mathcal{X} \rightarrow \mathbb{R}$ as

$$x \mapsto I_{F_\alpha}(x) := \int_{\mathcal{Y}} \dot{a}(y) \int_{\mathcal{Y}} \frac{e^{-(x-u)}}{p_{F_\alpha}(x)} 1\{u \leq y\} dF_\alpha(u) 1\{y \leq x\} dy,$$

integration by parts yields that

$$\begin{aligned}
& \int_{\mathcal{Y}} a(y) \frac{e^{-(x-y)}}{p_{F_\alpha}(x)} 1\{y \leq x\} dF_\alpha(y) \\
&= \int_{\mathcal{Y}} a(y) \frac{d}{dF_\alpha(y)} \left(\int_{\mathcal{Y}} \frac{e^{-(x-u)}}{p_{F_\alpha}(x)} 1\{u \leq y\} dF_\alpha(u) \right) 1\{y \leq x\} dF_\alpha(y) \\
&= a(y) \int_{\mathcal{Y}} \frac{e^{-(x-u)}}{p_{F_\alpha}(x)} 1\{u \leq y\} dF_\alpha(u) \Big|_{y_{\min}}^x - I_{F_\alpha}(x) \\
&= a(x) \int_{\mathcal{Y}} \frac{e^{-(x-u)}}{p_{F_\alpha}(x)} 1\{u \leq x\} dF_\alpha(u) - I_{F_\alpha}(x) \\
&= a(x) - I_{F_\alpha}(x)
\end{aligned}$$

because $a(y_{\min}) < +\infty$ and $F_\alpha(y_{\min}) = 0$ by part (iii) of Assumption **(A3)** combined with the fact that $F \in \mathcal{P}_0$. Analogously, since $\dot{a}_+(y_{\min}) < +\infty$,

$$\begin{aligned}
& - \int_{\mathcal{Y}} \frac{d}{dF_\alpha(y)} \left(\dot{a}(y) \int_{\mathcal{Y}} e^{-(y-u)} 1\{u \leq y\} dF_\alpha(u) \right) \frac{e^{-(x-y)}}{p_{F_\alpha}(x)} 1\{y \leq x\} dF_\alpha(y) \\
&= -\dot{a}(y) \int_{\mathcal{Y}} e^{-(y-u)} 1\{u \leq y\} dF_\alpha(u) \times \frac{e^{-(x-y)}}{p_{F_\alpha}(x)} \Big|_{y_{\min}}^x \\
&\quad + \int_{\mathcal{Y}} \dot{a}(y) \int_{\mathcal{Y}} \frac{e^{-(x-u)}}{p_{F_\alpha}(x)} 1\{u \leq y\} dF_\alpha(u) 1\{y \leq x\} dy \\
&= -\dot{a}(x) \int_{\mathcal{Y}} \frac{e^{-(x-u)}}{p_{F_\alpha}(x)} 1\{u \leq x\} dF_\alpha(u) + I_{F_\alpha}(x) \\
&= -\dot{a}(x) + I_{F_\alpha}(x).
\end{aligned}$$

Then, defined the mapping $h_{F_\alpha} : \mathcal{Y} \rightarrow \mathbb{R}$ as

$$y \mapsto h_{F_\alpha}(y) := \left[a(y) - \frac{d}{dF_\alpha(y)} \left(\dot{a}(y) \int_{\mathcal{Y}} e^{-(y-u)} 1\{u \leq y\} dF_\alpha(u) \right) - \psi(F_\alpha) \right], \quad (14)$$

by previous computations, we have that

$$\begin{aligned}
\forall x \in \mathcal{X}, \quad A_{F_\alpha} h_{F_\alpha}(x) &:= \mathbb{E}[h_{F_\alpha}(Y) | X = x] \\
&= \int_{\mathcal{Y}} h_{F_\alpha}(y) \frac{e^{-(x-y)}}{p_{F_\alpha}(x)} 1\{y \leq x\} dF_\alpha(y) \\
&= a(x) - \dot{a}(x) - \psi(F_\alpha) = b_{F_\alpha}(x). \quad (15)
\end{aligned}$$

In order to check that h_{F_α} has expected value $\int h_{F_\alpha} dF_\alpha = 0$, it suffices to note that, by applying twice the tower rule and using equalities (15) and (12),

$$\begin{aligned}
\int h_{F_\alpha} dF_\alpha &= \mathbb{E}[h_{F_\alpha}(Y)] = \mathbb{E}[\mathbb{E}[h_{F_\alpha}(Y) | X]] \\
&= \mathbb{E}[A_{F_\alpha} h_{F_\alpha}(X)] \\
&= \mathbb{E}[b_{F_\alpha}(X)] \\
&= \mathbb{E}[\mathbb{E}[b_{F_\alpha}(X) | Y]] \quad (16)
\end{aligned}$$

$$\begin{aligned}
&= E[A^* b_{F_\alpha}(Y)] \\
&= E[a(Y) - \psi(F_\alpha)] = \int_{\mathcal{Y}} [a(y) - \psi(F_\alpha)] dF_\alpha(y) = 0.
\end{aligned}$$

Next, we show that, for every $\alpha \in [0, 1)$ and $F \in \mathcal{P}_0$,

$$\sup_{y \in \mathcal{Y}} |h_{F_\alpha}(y)| < +\infty,$$

which implies that $h_{F_\alpha} \in L^2(F_\alpha)$. Noting that

$$\begin{aligned}
&\frac{d}{dF_\alpha(y)} \left(\dot{a}(y) \int_{\mathcal{Y}} e^{-(y-u)} 1\{u \leq y\} dF_\alpha(u) \right) \\
&= \dot{a}(y) e^{-y} \times \frac{d}{dF_\alpha(y)} \left(\int_{\mathcal{Y}} e^u 1\{u \leq y\} dF_\alpha(u) \right) \\
&\quad + \frac{d(\dot{a}(y) e^{-y})}{dF_\alpha(y)} \times \int_{\mathcal{Y}} e^u 1\{u \leq y\} dF_\alpha(u) \\
&= \dot{a}(y) + \frac{d(\dot{a}(y) e^{-y})}{dF_\alpha(y)} \times \int_{\mathcal{Y}} e^u 1\{u \leq y\} dF_\alpha(u),
\end{aligned}$$

we can rewrite h_{F_α} in (14) as

$$\begin{aligned}
h_{F_\alpha}(y) &= \left\{ a(y) - \dot{a}(y) - \alpha \frac{d(\dot{a}(y) e^{-y})}{dF_0(y)} \times \frac{dF_0(y)}{dF_\alpha(y)} \int_{\mathcal{Y}} e^u 1\{u \leq y\} dF(u) \right. \\
&\quad - (1 - \alpha) \frac{d(\dot{a}(y) e^{-y})}{dF_0(y)} \times \frac{dF_0(y)}{dF_\alpha(y)} \int_{\mathcal{Y}} e^u 1\{u \leq y\} dF_0(u) \\
&\quad \left. - \alpha \psi(F) - (1 - \alpha) \psi(F_0) \right\}.
\end{aligned}$$

To conclude that h_{F_α} is bounded on \mathcal{Y} , we observe two facts. First,

$$\begin{aligned}
|\alpha \psi(F) + (1 - \alpha) \psi(F_0)| &< |\psi(F_0)| + |\psi(F) - \psi(F_0)| \\
&\leq |\psi(F_0)| + \int_{\mathcal{Y}} |a(y)| d(F + F_0)(y) \\
&\leq |\psi(F_0)| + 2 \sup_{y \in \mathcal{Y}} |a(y)| < +\infty,
\end{aligned}$$

where $|\psi(F_0)| < +\infty$ by Assumption **(A0)** and $\sup_{y \in \mathcal{Y}} |a(y)| < +\infty$ by parts (i) and (ii) of Assumption **(A3)**. Second, for every $\alpha \in [0, 1)$ and $F \in \mathcal{P}_0$,

$$0 < \sup_{y \in \mathcal{Y}} \frac{dF_0}{dF_\alpha}(y) \leq \frac{1}{(1 - \alpha)}, \quad (17)$$

where (dF_0/dF_α) exists because F_α dominates F_0 . The bound in (17) holds uniformly over \mathcal{P}_0 . Therefore,

$$\sup_{y \in \mathcal{Y}} |h_{F_\alpha}(y)| < \sup_{y \in \mathcal{Y}} |a(y)| + \sup_{y \in \mathcal{Y}} |\dot{a}(y)| + \frac{1}{1 - \alpha} \sup_{y \in \mathcal{Y}} \left| \frac{d(\dot{a}(y) e^{-y})}{dF_0(y)} \right| [M_F(1) + M_{F_0}(1)] \quad (18)$$

$$+ |\psi(F_0)| + 2 \sup_{y \in \mathcal{Y}} |a(y)| < +\infty$$

by Assumptions **(A0)**, **(A1)**, **(A3)**–**(A5)** and the fact that $M_F(1)$ is bounded by a constant M on \mathcal{P}_0 .

(b)–(c) *Definition of b_{F_α} and differentiability of ψ at F_α .*

The function b_{F_α} defined in (6), which solves equation (12), is such that $A_{F_\alpha} h_{F_\alpha}(x) = b_{F_\alpha}(x)$ for every $x \in \mathcal{X}$, in virtue of (15).

Verification of Condition 3. (Control on the worst possible sub-directions h_{F_α}).

Recall that $M_F(1)$ in (18) is bounded by M on \mathcal{P}_0 . Besides, the factor $(1 - \alpha)^{-1}$, which diverges to $+\infty$ as $\alpha \rightarrow 1$, is counterbalanced by $1 - \alpha$. There thus exists a positive constant $B \equiv B(M, r) < +\infty$ such that

$$\sup_{F \in \mathcal{P}_0} \sup_{0 \leq \alpha < 1} \sup_{y \in \mathcal{Y}} (1 - \alpha) |h_{F_\alpha}(y)| \leq B.$$

Verification of Condition 4. (Control on the efficient influence functions b_{F_α}).

The information for estimating $\psi(F_0)$ is positive and finite, $0 < \|b_{F_0}\|_{2, P_0}^2 < +\infty$. Also, the influence functions are uniformly bounded. In fact, for every $x \in \mathcal{X}$, we have $|b_{F_\alpha}(x)| < |b_{F_0}(x)| + |\psi(F) - \psi(F_0)|$ so that

$$\sup_{F \in \mathcal{P}_0} \sup_{0 \leq \alpha < 1} \sup_{x \in \mathcal{X}} |b_{F_\alpha}(x)| < 3 \sup_{y \in \mathcal{Y}} |a(y)| + \sup_{y \in \mathcal{Y}} |\dot{a}(y)| + |\psi(F_0)| < +\infty$$

by Assumptions **(A0)**, **(A3)** (parts (i) and (ii)) and **(A4)**.

Next, to show that relationships (2.10) and (2.11) in van de Geer (1997), p. 21, are satisfied, we follow the reasoning illustrated in Sect. 3.4, *ibid.*, pp. 26–27, and check that, for some positive sequence $r_n \rightarrow 0$,

$$\lim_{n \rightarrow +\infty} \sup_{F \in \mathcal{P}_n} \sup_{0 \leq \alpha < 1} \|b_{F_\alpha} - b_{F_0}\|_{2, P_0}^2 = 0, \quad (19)$$

where \mathcal{P}_n is the set obtained from \mathcal{P}_0 in (10) by replacing r with r_n . Note that $b_{F_\alpha} - b_{F_0} = \alpha[\psi(F) - \psi(F_0)] = \alpha \int_{\mathcal{Y}} a(y) d(F - F_0)(y)$. Using integration by parts, together with conditions (i) and (ii) of Assumption **(A3)**, which jointly guarantee that a is bounded on \mathcal{Y} , as well as the fact that every $F \in \mathcal{P}_n$ has the same support as F_0 , we find that $\int_{\mathcal{Y}} a(y) d(F - F_0)(y) = - \int_{\mathcal{Y}} \dot{a}(y)(F - F_0)(y) dy$. The latter integral can be bounded above by applying inequality (30) in Scricciolo (2018), p. 358, which relates the L^1 -Wasserstein or Kantorovich distance $W_1(F, F_0) = \|F - F_0\|_1$ between distribution functions F and F_0 to the Hellinger distance between the corresponding mixtures (of exponential densities) $d_H \equiv d_H(p_F, p_0) = \|p_F^{1/2} - p_0^{1/2}\|_2$,

$$W_1(F, F_0) \lesssim \sqrt{d_H} \log^{3/4}(1/d_H), \quad (20)$$

where “ \lesssim ” indicates inequality valid up to a constant multiple that is universal or fixed within the context, but anyway inessential for our purposes because the bound is uniform over \mathcal{P}_n . The inequality is obtained by setting $p = 1$ and $\beta = 1$, the latter

value being determined by condition (29), *ibid.*, p. 358, on the Fourier transform of a standard exponential density. By Assumption (A4), which guarantees that \dot{a} is bounded on \mathcal{Y} , and inequality (20), we have

$$\begin{aligned} \|b_{F_\alpha} - b_{F_0}\|_{2,P_0}^2 &< \left| \int_{\mathcal{Y}} a(y) d(F - F_0)(y) \right|^2 \\ &= \left| \int_{\mathcal{Y}} \dot{a}(y)(F - F_0)(y) dy \right|^2 \\ &\leq \left(\sup_{y \in \mathcal{Y}} |\dot{a}(y)| \right)^2 W_1^2(F, F_0) \lesssim d_H \log^{3/2}(1/d_H), \end{aligned}$$

where $\lim_{n \rightarrow +\infty} d_H \log^{3/2}(1/d_H) = 0$ because $d_H \leq r_n$ on \mathcal{P}_n . The limit in (19) follows.

It remains to check that, for the collection of functions $\mathcal{S} := \{b_{F_\alpha} : d_H(p_F, p_0) \leq r, 0 \leq \alpha < 1\}$, the bracketing integral

$$\int_0^1 \sqrt{\log N_{[]}(\varepsilon, \mathcal{S}, L^2(P_0))} d\varepsilon < +\infty, \quad (21)$$

where $N_{[]}(\varepsilon, \mathcal{S}, L^2(P_0))$ is the ε -bracketing number of \mathcal{S} for the $L^2(P_0)$ -metric, namely, the smallest number of ε -brackets needed to cover \mathcal{S} , see, *e.g.*, 2.1.6 Definition (Bracketing numbers) in van der Vaart and Wellner (1996), p. 83, or Definition 2.2 in van der Geer (2000), p. 16. Under Assumption (A3) (parts (i) and (ii)) and Assumption (A4), by the same arguments as before, the $L^2(P_0)$ -distance between the lower and upper functions $b_{F_\alpha}^L$ and $b_{F_\alpha}^U$ of every bracket $[b_{F_\alpha}^L, b_{F_\alpha}^U]$ can be bounded above as follows:

$$\|b_{F_\alpha}^U - b_{F_\alpha}^L\|_{2,P_0} \lesssim \|F^U - F^L\|_1.$$

By 2.7.5 Theorem in van der Vaart and Wellner (1996), pp. 159–162, the bracketing entropy of the class of all uniformly bounded, monotone functions on the real line is of the order $O(1/\varepsilon)$. Therefore, $\log N_{[]}(\varepsilon, \mathcal{S}, L^2(P_0)) = O(1/\varepsilon)$ and the integral in (21) is finite. The proof of Condition 4 is thus complete.

The conclusion of Theorem 2.1 follows:

$$\psi(\hat{F}_n) - \psi(F_0) = \int b_{F_0} d(\mathbb{P}_n - P_0) + o_{\mathbf{P}}(n^{-1/2}) = \mathbb{P}_n b_{F_0} + o_{\mathbf{P}}(n^{-1/2}),$$

where $\mathbb{P}_n b_{F_0}$ has expected value $P_0 b_{F_0} = 0$, as it can be deduced from (16) when $\alpha = 0$. Hence,

$$\sqrt{n}(\psi(\hat{F}_n) - \psi(F_0)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \|b_{F_0}\|_{2,P_0}^2)$$

and the proof is complete. \square

Appendix B

In this section, we present the proof of Proposition 2 which states that no integral linear functional of a non-degenerate mixing distribution in a convolution model with the *Laplace* kernel density is estimable at parametric rate, in particular, by the maximum likelihood method.

Proof of Proposition 2 We let, at the outset, $\psi(F_0)$ be any integral linear functional, as defined in (4), evaluated at the “point” F_0 . Arguments are laid down to identify functions a (if any) whose corresponding functionals are estimable at $n^{-1/2}$ -rate. To the aim, we appeal to van der Vaart’s differentiability theorem, which provides a necessary and sufficient condition for pathwise differentiability of a (not necessarily linear) functional, see Theorem 3.1, Corollaries 3.2, 3.3 and Lemma 3.4 of van der Vaart (1991), pp. 183–185, or Theorem 3.1, Corollaries 3.1, 3.2 and Proposition 3.1 in Groeneboom and Wellner (1992), pp. 24–28. If differentiability of a functional fails, then, by Theorem 2.1 of van der Vaart (1991), p. 181, the functional is not estimable at $n^{-1/2}$ -rate, see also chpt. 25 in van der Vaart (1998), pp. 358–432. A necessary and sufficient condition for differentiability of an integral linear functional $\psi(F_0)$ is that, for $\mathcal{X} = \mathbb{R}$, there exists a function $b : \mathcal{X} \rightarrow \mathbb{R}$, with $b \in L^2(P_0)$, satisfying

$$\forall y \in \mathcal{Y}, \quad E[b(X) | Y = y] = a(y) - \psi(F_0),$$

explicitly,

$$\forall y \in \mathcal{Y}, \quad \int_{\mathcal{X}} b(x) \frac{1}{2} e^{-|x-y|} dx = a(y) - \psi(F_0), \quad (22)$$

where the conditional density of X , given $Y = y$, is $k(x-y) = e^{-|x-y|}/2$, see § 7 in van der Vaart (1991), pp. 189–191, or Example 3.2 in Groeneboom and Wellner (1992), pp. 30–31. If an integral linear functional $\psi(F_0)$ is regularly estimable, then the condition in (22) must be necessarily satisfied and a regular estimator for $\psi(F_0)$ is given by $\mathbb{P}_n b = n^{-1} \sum_{i=1}^n b(X_i)$. The following arguments are aimed at deriving the expression of b . Let $y \in \mathcal{Y}$ be fixed. For a function $a : \mathcal{Y} \rightarrow \mathbb{R}$ such that

$$\lim_{u \rightarrow -\infty} a(u)e^u = 0, \quad (23)$$

where, in the case when \mathcal{Y} is bounded, a (hence its derivative \dot{a}) is taken to be identically equal to zero on \mathcal{Y}^c so that the limit is automatically verified, integration by parts yields that

$$\begin{aligned} \int_{\mathcal{X}} \dot{a}(x) \frac{1}{2} e^{-(y-x)} 1\{x \leq y\} dx &= a(x) \frac{1}{2} e^{-(y-x)} \Big|_{-\infty}^y - \int_{\mathcal{X}} a(x) \frac{1}{2} e^{-(y-x)} 1\{x \leq y\} dx \\ &= \frac{1}{2} a(y) - \int_{\mathcal{X}} a(x) \frac{1}{2} e^{-(y-x)} 1\{x \leq y\} dx, \end{aligned} \quad (24)$$

whence

$$\int_{\mathcal{X}} [a(x) + \dot{a}(x)] \frac{1}{2} e^{-(y-x)} 1\{x \leq y\} dx = \frac{1}{2} a(y). \quad (25)$$

The integral analogous to the one on the left-hand side of (24), but with the right branch of the Laplace density, can be dealt with similarly. For some a satisfying the limit in (23) and also

$$\lim_{u \rightarrow +\infty} a(u)e^{-u} = 0,$$

where the same proviso on a and \dot{a} applies for the case when \mathcal{Y} is bounded, we get

$$\begin{aligned} \int_{\mathcal{X}} \dot{a}(x) \frac{1}{2} e^{-(x-y)} 1\{x > y\} dx &= a(x) \frac{1}{2} e^{-(x-y)} \Big|_y^{+\infty} + \int_{\mathcal{X}} a(x) \frac{1}{2} e^{-(x-y)} 1\{x > y\} dx \\ &= -\frac{1}{2} a(y) + \int_{\mathcal{X}} a(x) \frac{1}{2} e^{-(x-y)} 1\{x > y\} dx, \end{aligned}$$

whence

$$\int_{\mathcal{X}} [a(x) - \dot{a}(x)] \frac{1}{2} e^{-(x-y)} 1\{x > y\} dx = \frac{1}{2} a(y). \quad (26)$$

Summing side by side (25) and (26) and subtracting $\psi(F_0)$ on both sides of the resulting equation, we obtain

$$\int_{\mathcal{X}} [a(x) - \operatorname{sgn}(x-y)\dot{a}(x) - \psi(F_0)] \frac{1}{2} e^{-|x-y|} dx = a(y) - \psi(F_0).$$

In order to get rid of the dependence of the function $a(\cdot) - \operatorname{sgn}(\cdot - y)\dot{a}(\cdot) - \psi(F_0)$ on y , the derivative $\dot{a}(\cdot)$ must be equal to zero, which means that $a(\cdot)$ is identically equal to a constant on \mathcal{Y} and the functional is trivially equal to the constant itself. Conclude that there exists no integral linear functional $\psi(F_0)$ of a non-degenerate mixing distribution F_0 that can be estimated at $n^{-1/2}$ -rate. This completes the proof. \square

References

- Billingsley P (1995) Probability and measure. Wiley, New York, 3rd edition
- Bolthausen E, Perkins E, van der Vaart A (2002) Lectures on probability theory and statistics. Ecole d'Été de Probabilités de Saint-Flour XXIX – 1999. Bernard P (ed) Lecture Notes in Mathematics, Vol 1781. Springer-Verlag, Berlin, pp 331–457
- Buonaccorsi, JP (2010) Measurement error: models, methods, and applications. Chapman & Hall/CRC Press, Boca Raton, FL
- Buzas JS, Stefanski LA, Tosteson TD (2005) Measurement error. In: Ahrens W, Pigeot I (eds) Handbook of epidemiology. Springer-Verlag, Berlin, Heidelberg, pp 729–765
- Carroll LB (2017) Nuclear steam generator fitness-for-service assessment. In: Riznic J (ed) Steam generators for nuclear power plants. Woodhead Publishing, pp 511–523
- Carroll RJ, Hall P (1988) Optimal rates of convergence for deconvolving a density. *J Amer Statist Assoc* 83:1184–1186
- Chen J (2017) Consistency of the MLE under mixture models. *Stat Sci* 32:47–63
- Daniels HE (1961) The asymptotic efficiency of a maximum likelihood estimator. In: Proc Fourth Berkeley Symp on Math Statist and Prob, Vol 1. Univ of Calif Press, pp 151–163
- Dattner I, Goldenshluger A, Juditsky A (2011) On deconvolution of distribution functions. *Ann Stat* 39:2477–2501
- Davidian M, Lin X, Morris JS, Stefanski LA (2014) The work of Raymond J. Carroll: The impact and influence of a statistician. Springer International Publishing, Switzerland
- Easterling RG (1980) Statistical analysis of steam generator inspection plans and eddy current testing. Washington, DC: Division of Operating Reactors, Office of Nuclear Reactor Regulation, US Nuclear Regulatory Commission
- Fan J (1991) On the optimal rates of convergence for nonparametric deconvolution problems. *Ann Stat* 19:1257–1272
- Fuller WA (1987) Measurement error models. John Wiley, New York
- Groeneboom P, Wellner JA (1992) Information bounds and nonparametric maximum likelihood estimation. Birkhäuser, Basel
- Hall P, Lahiri SN (2008) Estimation of distributions, moments and quantiles in deconvolution problems. *Ann Stat* 36:2110–2134
- Huber PJ (1967) The behavior of maximum likelihood estimates under nonstandard conditions. In: Proc Fifth Berkeley Symp on Math Statist and Prob, Vol 1. Univ of Calif Press, pp 221–233
- Kotz S, Kozubowski TJ, Podgórski K (2001) The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance. Birkhäuser, Boston
- Laplace P-S (1774) Mémoire sur la probabilité des causes par les événements. *Mém Acad Roy Sci Paris (Savants étrangers)* Tome VI:621–656
- Lehmann EL, Casella G (1998) Theory of point estimation, 2nd ed. Springer-Verlag, New York
- Lindsay BG (1983) The geometry of mixture likelihoods: a general theory. *Ann Stat* 11:86–94

- Norton RM (1984) The double exponential distribution: using calculus to find a maximum likelihood estimator. *Am Stat* 38:135–136
- Scricciolo C (2018) Bayes and maximum likelihood for L^1 -Wasserstein deconvolution of Laplace mixtures. *Stat Methods Appl* 27:333–362
- Sollier T (2017) Nuclear steam generator inspection and testing. In: Riznic J (ed) *Steam generators for nuclear power plants*. Woodhead Publishing, pp 471–493
- Stefanski L, Carroll RJ (1990) Deconvoluting kernel density estimators. *Statistics* 21:169–184
- van de Geer S (1997) Asymptotic normality in mixture models. *ESAIM Probab Stat* 1:17–33
- van de Geer SA (2000) *Empirical processes in M-estimation*. Cambridge University Press, Cambridge
- van de Geer S (2003) Asymptotic theory for maximum likelihood in nonparametric mixture models. *Comput Stat Data An* 41:453–464
- van der Vaart A (1991) On differentiable functionals. *Ann Stat* 19:178–204
- van der Vaart AW (1998) *Asymptotic statistics*. Cambridge University Press, Cambridge
- van der Vaart AW, Wellner JA (1996) *Weak convergence and empirical processes*. Springer-Verlag, New York
- Vardi Y (1989) Multiplicative censoring, renewal processes, deconvolution and decreasing density: non-parametric estimation. *Biometrika* 76:751–761