Statistical Inference for Quasi-Infinitely Divisible Distributions via Fourier Methods

Vladimir Panov *[1] and Anton Ryabchenko †[2]

[1,2]HSE University, Pokrovsky boulevard 11, 109028 Moscow, Russia

May 21, 2025

Abstract

This study focuses on statistical inference for the class of quasi-infinitely divisible (QID) distributions, which was recently introduced by Lindner, Pan and Sato [12]. The paper presents a Fourier approach, based on the analogue of the Lévy-Khintchine theorem with a signed spectral measure. We prove that for some subclasses of QID distributions, the considered estimates have polynomial rates of convergence. This is a remarkable fact when compared to the logarithmic convergence rates of similar methods for infinitely divisible distributions, which cannot be improved in general. We demonstrate the numerical performance of the algorithm using simulated examples.

1 Introduction

A random variable X has quasi-infinitely divisible (QID) distribution if there exist two random variables Y and Z with infinitely divisible (ID) distributions such that

$$X + Y \stackrel{d}{=} Z, \tag{1}$$

where X and Y are independent. Trivially, this class includes all ID distributions, but it is actually significantly larger. In particular, it includes several notable examples of distributions, which are not ID, such as the Bernoulli distribution with parameter that is not equal to $1/2$ and some mixtures of normal distributions with different variances.

A complete characterization of quasi-infinite divisibility is not known yet, but there are some simple criteria for certain subclasses of distributions. For

*vpanov@hse.ru

†apryabchenko@hse.ru

instance, a distribution on integers is QID if and only if its characteristic function doesn't have real zeros (Lindner, Pan and Sato [12]). More generally, a discrete distribution is QID if and only if its characteristic function is separated from zero (Alexeev and Khartov [1]). This result also holds for mixtures of discrete and absolutely continuous distributions (Berger and Kutlu [5]). Along with the mentioned papers on the QID random variables, there are several papers dealing with the generalisation of this concept to random vectors (Berger et al. [6]) and stochastic processes (Passegeri [19]). Note that these notions are not only of theoretical interest, they have found applications in financial modelling (Madan et al. [14]), physics, specifically in the analysis of Landau levels (Chhaiba et al. [7], Demni and Mouayn [10]), number theory (Nakamura [16]) and insurance mathematics (Zhang et al. [22]).

It is a worth mentioning that this research area has emerged relatively recently, starting with the pioneer work [12]. Prior to that, QID had only been mentioned in some contexts without a systematic study (Cuppens [9], Linnik and Ostrovsky [13]).

Our research deals with the statistical inference for QID distributions. In contrast to the extensive literature on statistical estimation for ID distributions and Lévy-based models, this topic has been considered in only one paper by Passegeri [18] in the framework of Bayesian estimation. In the current paper, we consider a Fourier-based approach, which relies on the analogue of the well-known Lévy – Khintchine formula for the QID distributions. Note that methods of this type are widely used in the ID case, in particular, for the Lévy processes, see Belomestny [2], Comte and Genon-Catalot [8], Gugushvili [11], Neumann and Reiß [17].

Contribution. To the best of our knowledge, the application of Fourier-based methods for the statistical estimation of quasi-infinitely divisible distributions was never considered before. Our particular interest will be to the distributions, which are the mixtures of the normal distribution $\tilde{\mu}_\sigma$ with zero mean and variance equal to σ^2, and an absolutely continuous distribution μ°, that is,

$$\mu \quad = \quad p\tilde{\mu}_\sigma + (1-p)\mu^\circ, \qquad p \in (0,1). \tag{2}$$

This distribution is QID, if $p > 1/2$ and μ° satisfies one nonparametric condition, which we present and discuss below in Section 2.2. In our paper, we provide a semiparametric procedure for the estimation of unknown parameters p and σ, and unknown distribution μ°, from the observations of μ. This marks the first instance of the estimation in the model (2), despite parametric settings, e.g., EM algorithm dealing with the case when μ° is a normal distribution.

The parametric part of our algorithm (estimation of p and σ) is based on the representation of the characteristic function in a form similar to the

Lévy – Khintchine formula. We show that the estimates have polynomial rates of convergence if the distribution of μ° is supersmooth. This fact is remarkable when compared to the fact that the convergence rates of similar methods for ID distributions are logarithmic and can't be improved in general, see Theorems 4.3, 4.4, 5.7 in Belomestny and Reiss [4]. To estimate μ°, we propose the kernel estimator, which also has a polynomial rate of convergence, given our choice of the estimates of p and σ. Our approach provides a comprehensive semi-parametric estimation procedure for the model (2), which, to the best of our knowledge, does not have any analogues in the literature for the considered generality.

Structure. The paper is organised as follows. In the next section, we provide a brief overview of quasi-infinitely divisible distributions. Next, in Section 3 we propose a Fourier-based approach for all elements of the characteristic triplet of a QID distribution. Later, in Section 4 we describe a semiparametric estimation procedure for the model (2), which relies on the ideas presented in Section 3 and the kernel estimates for the density function of the measure μ°. Our main theoretical findings are given in Section 5. Comparison of Theorems 1 and 2 yields that the rates of convergence for the model (2) are essentially faster than in the general case if the distribution of μ° is supersmooth. Section 6 deals with the empirical study and includes, in particular, numerical comparison with the EM algorithm for the normal mixtures. The proofs are collected in Section 7.

2 Quasi-infinitely divisible distributions

2.1 Definition and spectral representation

The Lévy – Khintchine formula states that the characteristic function ϕ of an ID distribution can be represented in the following form

$$\phi(u) = \exp\Big\{ \mathrm{i}\gamma u - \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}\setminus\{0\}} \left(e^{\mathrm{i}ux} - 1 - \mathrm{i}uc(x)\right) \nu(dx) \Big\}, \qquad (3)$$

where $\gamma \in \mathbb{R}, \sigma \in \mathbb{R}_+$, $\nu : \mathcal{B}(\mathbb{R} \setminus \{0\}) \to \mathbb{R}_+$ is a measure such that $\int_{\mathbb{R}} \min(1, x^2)\nu(dx) < \infty$, and $c : \mathbb{R} \to \mathbb{R}$ is a bounded, Borel measurable function satisfying $c(x) = 1 + o(|x|)$ as $x \to 0$ and $c(x) = O(1/|x|)$ as $x \to \infty$. The function $c(x)$ is usually chosen as $c(x) = x1_{[-1,1]}(x)$.

To generalise (3) for the QID case, it is convenient to merge σ and ν into a single finite measure ζ on $\mathcal{B}(\mathbb{R})$:

$$\zeta(dx) \quad := \quad \sigma^2\delta_0(dx) + \min(1, x^2)1_{\{x\neq 0\}}\nu(dx),$$

which allows to rewrite (3) in the following form

$$\phi(u) \quad = \quad \exp\Big\{ \mathrm{i}\gamma u + \int_{\mathbb{R}} g(x, u)\zeta(dx) \Big\}, \qquad (4)$$

where

$$g(x,u) \quad := \quad \begin{cases} \frac{e^{\mathrm{i}ux}-1-\mathrm{i}uc(x)}{\min(1,x^2)}, & x \neq 0, \\ -u^2/2, & x = 0. \end{cases}$$

Trivially, there is a one-to-one correspondence between the pair (γ, ζ) and the Lévy triplet (γ, σ^2, ν), provided that the function c is fixed. In fact,

$$\sigma^2 = \zeta(\{0\}) \quad \text{and} \quad \nu(B) = \big(\min(1,x^2)\big)^{-1}\zeta(B), \qquad B \in \mathcal{B}(\mathbb{R}\setminus\{0\}). \quad (5)$$

Now we can formulate the following definition of a QID distribution.

Definition 1. A distribution μ is QID, if its characteristic function admits the representation (4) with some $\gamma \in \mathbb{R}$ and a finite signed measure $\zeta : \mathcal{B}(\mathbb{R}) \to [-\infty, +\infty]$.

Due to the Hahn - Jordan decomposition, for a finite signed measure ζ, there exist disjoint Borel sets C_+ and C_- and finite measures ζ_+ and ζ_- on $\mathcal{B}(\mathbb{R})$ such that

$$\zeta_+(\mathbb{R}\setminus C_+) = \zeta_-(\mathbb{R}\setminus C_-) = 0 \quad \text{and} \quad \zeta = \zeta_+ - \zeta_-,$$

and moreover the measures ζ_+ and ζ_- are uniquely determined by ζ. From (4) we get the representation

$$\phi(u) = \frac{\phi_1(u)}{\phi_2(u)}, \qquad (6)$$

where ϕ_1 and ϕ_2 are characteristic functions of ID distributions with characteristic pairs (γ, ζ_+) and $(0, \zeta_-)$. The equivalence with the definition given at the beginning of the introduction is now clear.

Note that for any QID distribution

$$\sigma^2 := \zeta(\{0\}) = -2 \lim_{u\to\infty} \big(\log(\phi(u))/u^2\big) \geq 0,$$

see Lemma 2.7 in [12], and therefore the characteristic pair $(0, \zeta_-)$ is equivalent to the Lévy triplet in the form $(0, 0, \nu_-)$. Therefore, (6) yields the following analogue of the Lévy−Khintchine formula (3) for QID distributions

$$\phi(u) = \exp\Big\{\mathrm{i}\gamma u - \frac{1}{2}\sigma^2u^2 + \int_{\mathbb{R}\setminus\{0\}} \big(e^{\mathrm{i}ux} - 1 - \mathrm{i}uc(x)\big)\,\nu_+(dx) \\ - \int_{\mathbb{R}\setminus\{0\}} \big(e^{\mathrm{i}ux} - 1 - \mathrm{i}uc(x)\big)\,\nu_-(dx)\Big\}, \quad (7)$$

where ν_+ and ν_- are determined by ζ_+ and ζ_- as in (5). The triplet (γ, σ^2, ν) with $\nu = \nu_+ - \nu_-$ uniquely determines the QID distribution.

2.2 Examples

Many interesting examples can be constructed using the following lemma, which was proven in [12].

Lemma 1. Consider the mixture $\mu = p\mu_1 + (1-p)\mu_2$, where $p \in (1/2, 1)$, μ_1 is a QID distribution with triplet (γ, σ^2, ν), and μ_2 is some distribution on $\mathbb{R}$. Suppose that there exists a finite signed measure Λ on $\mathbb{R}$ such that its Fourier transform is equal to

$$\mathcal{F}[\Lambda](u) = \frac{\phi_2(u)}{\phi_1(u)}, \qquad u \in \mathbb{R}, \tag{8}$$

where ϕ_1 and ϕ_2 are the characteristic functions of μ_1 and μ_2, If this measure is finite with total variation $|\Lambda|(\mathbb{R}) < p/(1-p)$, then μ is QID with the characteristic triplet

$$\Big(\gamma + \int_{\mathbb{R}} c(x)\tilde{\nu}(\mathrm{d}x), \quad \sigma^2, \quad \nu + \tilde{\nu}\Big),$$

where $\tilde{\nu}$ is a finite signed measure defined as follows:

$$\tilde{\nu}(B) = \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \Big(\frac{1-p}{p}\Big)^m \Lambda^{*m}(B), \qquad B \in \mathcal{B}(\mathbb{R} \setminus \{0\}). \tag{9}$$

Note that the assumption $|\Lambda|(\mathbb{R}) < p/(1-p)$ is fulfilled if Λ is a probability measure, since $p \in (1/2, 1)$. The next two corollaries present specific cases when Λ is a probability measure, and therefore Lemma 2 can be applied.

Corollary 1. *(Cuppens theorem.) Any distribution μ with an atom of mass larger than 1/2 is QID.*

Proof. In fact, in this case $\mu = p\mu_1 + (1-p)\mu_2$, where $p = \mu(\{\lambda\}) > 1/2$, $\mu_1 = \delta_\lambda$, and $\mu_2 := (\mu - p\delta_\lambda)/(1-p)$. Note that μ_2 is a probability measure. In this case, measure Λ is equal to $\mu_2 * \delta_{-\lambda}$ and is also a probability measure. We conclude that μ is a QID distribution with triplet $(\lambda, 0, \nu)$, where

$$\nu = \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \Big(\frac{1-p}{p}\Big)^m \big(\mu_2 * \delta_{-\lambda}\big)^{*m}.$$

$\square$

Example 1. In particular, the Bernoulli distribution Bern(p) with probability of success $p \neq 1/2$ is QID. Recall that this distribution is not ID, since its support is bounded (see Proposition 2.3 from [20]). Note that Bern(1/2) is not QID, since its characteristic function has real zeros, $\phi(u) = (1+e^{iu})/2 = 0$ for $u = (2k+1)\pi$, what is not possible due to representation (6) and the fact that the characteristic function $\phi_1(\cdot)$ of an ID distribution doesn't have zeros (see Proposition 2.8 from [20]).

Corollary 2. *Consider the mixture* $\mu = p\tilde{\mu}_\sigma + (1-p)\mu^\circ$*, where* $\tilde{\mu}_\sigma$ *is a normal distribution with parameters* 0 *and* σ^2*, and* μ° *is the distribution with the characteristic function* ϕ° *such that*

$$\mathcal{H}(u) := \phi^\circ(u)e^{u^2\sigma^2/2}, \qquad u \in \mathbb{R}, \tag{10}$$

is the characteristic function of some probability measure Λ*. Then for any* $p \in (1/2, 1]$*,* μ *is a QID distribution with the characteristic triplet*

$$\Big(\int_{\mathbb{R}} c(x)\tilde{\nu}(\mathrm{d}x), \quad \sigma^2, \quad \tilde{\nu}\Big),$$

where $\tilde{\nu}$ *is given by* (9)*.*

Proof. For $p \in (1/2, 1)$, the corollary directly follows from Lemma 2, since the measure Λ is a probability measure. For $p = 1$ the statement follows from the fact that $\tilde{\mu}_\sigma$ is an ID distribution with triplet $(0, \sigma^2, 0)$. $\square$

Example 2. In particular, a mixture of two normal distributions

$$\mu = p\tilde{\mu}_{\sigma_1} + (1-p)\tilde{\mu}_{\sigma_2}$$

is QID if the main component has smaller variance, that is, $p > 1/2$ and $\sigma_1 < \sigma_2$ or, vice versa, $p < 1/2$ and $\sigma_1 > \sigma_2$. For instance, in the first case, the function $\mathcal{H}(u) = e^{-u^2(\sigma_2^2 - \sigma_1^2)}, u \in \mathbb{R}$, is a characteristic function of the normal distribution with zero mean and variance equal to $\sigma_2^2 - \sigma_1^2$. A mixture of two normal distributions with different variances is an example of a QID distribution, which is not ID, because its tail is too thin (see Chapter VI from [20]).

Example 3. More generally, the assumptions of Corollary 2 are fulfilled if μ° is ID with Lévy triplet $(\gamma, \bar{\sigma}^2, \nu)$ if $\bar{\sigma} > \sigma$. In fact, due to the Lévy–Khintchine formula (3), $\mathcal{H}(u) = \phi^\circ(u)e^{u^2\sigma^2/2}$ is again a characteristic function of an ID distribution with Lévy triplet $(\gamma, \bar{\sigma}^2 - \sigma^2, \nu)$. We will return to this example later in Section 4.

3 Fourier-based estimation procedure

In this section we extend the Fourier-based estimation procedure for the ID distributions described in [4] to the QID distributions.

For simplicity we consider the case when the jump parts of Y and Z in (1) have absolutely continuous distribution of compound Poisson type. In terms of the measures ν_+ and ν_-, this means that for any $B \in \mathcal{B}(\mathbb{R})$,

$$\nu_+(B) = \lambda_+ \int_B p_+(x)\mathrm{d}x, \qquad \nu_-(B) = \lambda_- \int_B p_-(x)\mathrm{d}x,$$

where $\lambda_\pm = \nu_\pm(\mathbb{R}) \geq 0$, and $p_\pm$ are some probability density functions. Note that the quasi-Lévy measure ν has a density $s(x) = \lambda_+ p_+(x) - \lambda_- p_-(x)$, $x \in \mathbb{R}$.

In our setup, formula (7) can be simplified to

$$\begin{aligned}\phi(u) &= \exp\Big\{\mathrm{i}\gamma^* u - \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}} \big(e^{\mathrm{i}ux} - 1\big)\,\nu_+(dx) - \int_{\mathbb{R}} \big(e^{\mathrm{i}ux} - 1\big)\,\nu_-(dx)\Big\} \\ &= \exp\Big\{\mathrm{i}\gamma^* u - \frac{1}{2}\sigma^2 u^2 + \mathcal{F}\big[s\big](u) - \lambda^*\Big\}, \qquad (11)\end{aligned}$$

where

$$\gamma^* = \gamma - \int_{\mathbb{R}} c(x)s(x)dx, \qquad \lambda^* = \nu(\mathbb{R}) = \lambda_+ - \lambda_-,$$

and $\mathcal{F}[s](u) = \int_{\mathbb{R}} e^{\mathrm{i}ux} s(x)dx$ is the Fourier transform of s. Since

$$\int_{\mathbb{R}} s(x)dx = \lambda^* < \infty,$$

the Riemann-Lebesgue lemma yields $\big|\mathcal{F}[s](u)\big| \to 0$ as $u \to \infty$. Therefore, as $u \to \infty$,

$$\mathrm{Re}\big(\log(\phi(u))\big) = -\frac{1}{2}\sigma u^2 - \lambda^* + o(1), \qquad (12)$$

$$\mathrm{Im}\big(\log(\phi(u))\big) = \gamma^* u + o(1). \qquad (13)$$

These ideas give rise for the following 4-step estimation procedure for the estimation of the parameters $\gamma^*, \sigma, \lambda^*$ and the density s from the observations $X_1, .., X_n$ of the corresponding QID distribution.

1. First, the characteristic function of X is estimated by

$$\phi_n(u) = \frac{1}{n}\sum_{k=1}^{n} e^{iuX_k}. \qquad (14)$$

2. Then, motivated by (12), introduce the estimate

$$(\sigma_n^2, \lambda_n^*) = \operatorname*{arg\,min}_{\sigma^2,\lambda^*} \int_{\mathbb{R}_+} w^{U_n}(u)\Big[\mathrm{Re}\big(\log(\phi_n(u))\big) + \frac{1}{2}\sigma^2 u^2 + \lambda^*\Big]^2 du, \qquad (15)$$

where $w^{U_n}(u) = U_n^{-1} w(u/U_n)$ with some unbounded increasing sequence of positive numbers U_n, and a function w is supported on $[\varepsilon, 1]$, $\varepsilon > 0$, and belongs to the class $L^1([\varepsilon, 1])$. Direct calculations yield that the solution σ_n^2 of (15) can be represented in the form

$$\sigma_n^2 = \int_{\mathbb{R}_+} w_{\sigma^2}^{U_n}(u)\,\mathrm{Re}\big(\log(\phi_n(u))\big)du, \qquad (16)$$

where

$$w_{\sigma^2}^{U_n}(u) = \frac{2w^{U_n}(u)\Big(u^2\int_{\mathbb{R}_+} w^{U_n}(s)ds - \int_{\mathbb{R}_+} w^{U_n}(s)s^2ds\Big)}{\big(\int_{\mathbb{R}_+} w^{U_n}(s)s^2ds\big)^2 - \int_{\mathbb{R}_+} w^{U_n}(s)s^4ds \cdot \int_{\mathbb{R}_+} w^{U_n}(s)ds}.$$

Note that $w_{\sigma^2}^{U_n}(u) = U_n^{-3}w_{\sigma^2}^{1}(u/U_n)$, and moreover,

$$\int_{\mathbb{R}_+} w_{\sigma^2}^{U_n}(u)du = 0, \qquad \int_{\mathbb{R}_+} \frac{-u^2}{2} w_{\sigma^2}^{U_n}(u)du = 1. \tag{17}$$

Similarly,

$$\lambda_n^* = \int_{\mathbb{R}_+} w_{\lambda^*}^{U_n}(u)\,\mathrm{Re}\big(\log(\phi_n(u))\big)du, \tag{18}$$

where the function $w_{\lambda^*}^{U_n}(u) = U_n^{-1}w_{\lambda^*}^{1}(u/U_n)$ has the following properties

$$\int_{\mathbb{R}_+} -w_{\lambda^*}^{U_n}(u)du = 1, \qquad \int_{\mathbb{R}_+} \frac{u^2}{2} w_{\lambda^*}^{U_n}(u)du = 0.$$

3. Analogously, motivated by (13), introduce the estimate

$$\gamma_n^* = \arg\min_{\gamma^*} \int_{\mathbb{R}_+} w^{V_n}(u)\Big[\mathrm{Im}\big(\log(\phi_n(u))\big) - \gamma^* u\Big]^2 du, \tag{19}$$

where V_n is some unbounded increasing sequence of positive numbers, and the function w can be defined in the same way as in the previous step. The solution of this problem is given by

$$\gamma_n^* = \int_{\mathbb{R}_+} w_{\gamma^*}^{V_n}(u)\,\mathrm{Im}\big(\log(\phi_n(u))\big)du, \tag{20}$$

where the function $w_{\gamma^*}^{U_n}(u) = U_n^{-2}w_{\gamma^*}^{1}(u/U_n)$ satisfies

$$\int_{\mathbb{R}_+} u w_{\gamma^*}^{U_n}(u)du = 1.$$

4. Finally, we turn towards nonparametric estimation of the density s. From (11), we get that the natural estimator of s is the inverse Fourier transform, that is,

$$s_n(x) \quad = \quad \mathcal{F}^{-1}\Big[\big(\log(\phi_n(\cdot)) - \mathrm{i}\gamma_n^*(\cdot) + \frac{1}{2}\sigma_n^2(\cdot)^2 + \lambda_n^*\big)w_s(\cdot/T_n)\Big](x),$$

where $x \in \mathbb{R}$ and w_s is a weight function supported on $[-1, 1]$, and T_n is some unbounded increasing sequence of positive numbers.

4 Semiparametric inference for the model (2)

As we have seen in Corollary 2, the measure

$$\mu \quad = \quad p\tilde{\mu}_\sigma + (1-p)\mu^\circ. \tag{21}$$

is QID, if $p > 1/2$ and the function $\mathcal{H}$ defined by (10) is a characteristic function of some distribution. In this section, we propose a semiparametric estimation procedure for the estimation of the known parameters p and σ and unknown measure μ°, which is assumed to be absolutely continuous.

Parametric part. Let us recall two important issues related to the model (21).

1. In this particular case, the measure Λ, which plays essential role in Lemma 2, is a probability measure. Due to this fact, formula (9) yields

   $$\lambda^* = \tilde{\nu}(\mathbb{R}) := \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \left(\frac{1-p}{p}\right)^m \underbrace{\Lambda^{*m}(\mathbb{R})}_{=1} = -\log(p),$$

   where the last equality follows from $(1-p)/p \in (0,1)$.

2. The second element of the characteristic triplet, σ^2, coincides with the variance of the first (normal) component in (21).

Therefore, the application of the first and second steps from the previous section results into the estimation of $p = e^{-\lambda^*}$ by the natural plug-in estimate $p_n := e^{-\lambda_n^*}$, and σ^2 by σ_n^2.

Nonparametric part. Now we turn towards the non-parametric estimation of the measure μ° in (21). Assuming that this measure (and therefore μ) is absolutely continuous, we first estimate the density g of μ by the kernel estimate

$$g_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right), \qquad x \in \mathbb{R},$$

where $K : \mathbb{R} \to \mathbb{R}_+$ is a kernel function satisfying $\int_{\mathbb{R}} K(u)du = 1$. Then, we apply linear transform to estimate the density g° of μ°,

$$g_n^\circ(x) := \frac{g_n(x) - p_n \varphi_{\sigma_n}(x)}{1 - p_n}, \tag{22}$$

where φ_{σ_n} is the density of the normal distrbution with zero mean and variance equal to σ_n^2.

5 Convergence rates

In [4], it is shown that the rates of convergence of the algorithm presented in Section 3 for ID distributions are logarithmic if $\sigma > 0$, and cannot be improved in general. Below, in Theorem 1, we show similar result for the subclass $\mathcal{S}(r, \bar{\sigma}, C)$ of QID distributions with characteristic functions in the form (11), where the density s is r-times differentiable and moreover

$$\sigma \in (0, \bar{\sigma}], \qquad \gamma^* \in [-C, C], \qquad \lambda^* \in [0, C], \qquad \|s^{(r)}\|_\infty \leq C$$

with some $\bar{\sigma}, C > 0$.

It is convenient to establish the rates of convergence conditional to some event $\mathcal{A}_n$, which we define in the following lemma. The first part of this lemma shows that the probability of this event tends to 1 at a polynomial rate, if the constant χ° is small enough. The second part states that the sequence χ_n, which is essential in the definition of $\mathcal{A}_n$, converges to zero under certain choice of the sequence U_n.

Lemma 2. Denote the event

$$\mathcal{A}_n = \mathcal{A}_n(\chi^\circ) := \Big\{ \max_{u \in [-U_n, U_n]} \frac{|\phi_n(u) - \phi(u)|}{|\phi(u)|} \leq \chi_n \Big\}, \qquad n = 1, 2, ...$$

with

$$\chi_n := \chi^\circ \frac{\sqrt{\log(nU_n^2)/n}}{\inf_{u \in [-U_n, U_n]} |\phi(u)|},$$

where χ° is positive constant. The following statements hold.

(i) If $\chi^\circ < 1/16$, then

$$\mathrm{P}\{\mathcal{A}_n\} \geq 1 - c(\sqrt{n} U_n)^{-\varkappa}$$

with $\varkappa = (1/(2\chi^\circ)^2 - 64)/128 > 0$ and some positive constant c, which depends on $\mathrm{E}[|X_1|]$ only.

(ii) For any distribution from the class $\mathcal{S}(r, \bar{\sigma}, C)$, it holds

$$\chi_n \lesssim e^C \sqrt{\frac{\log(nU_n^2)}{n}} \exp(\frac{1}{2}\bar{\sigma}^2 U_n^2).$$

Therefore, $\chi_n \to 0$ if $U_n \lesssim \sqrt{\log n / \sigma_{max}^2}$ with any $\sigma_{max} > \bar{\sigma}$.

Proof. (i) Using a version of Hoeffding's inequality, given in Proposition 3.3 in [4], we get

$$\begin{aligned} \mathrm{P}\{\mathcal{A}_n^c\} &\leq \mathrm{P}\Big\{ \max_{u \in [-U_n, U_n]} |\phi_n(u) - \phi(u)| > \chi_n \inf_{u \in [-U_n, U_n]} |\phi(u)| \Big\} \\ &= \mathrm{P}\Big\{ \sqrt{n / \log(nU_n^2)} \max_{u \in [-U_n, U_n]} |\phi_n(u) - \phi(u)| > \chi^\circ \Big\} \leq c(\sqrt{n} U_n)^{-\varkappa}, \end{aligned}$$

provided $\chi^\circ < 1/16$.

(ii) The lower bound for $\inf_{u\in[-U_n,U_n]}|\phi(u)|$ follows from the representation (11)

$$\begin{aligned}\inf_{u\in[-U_n,U_n]}|\phi(u)| &= \inf_{u\in[-U_n,U_n]}\Big|\exp\big\{\mathrm{i}\gamma^* u-\frac{1}{2}\sigma^2u^2+\mathcal{F}\big[s\big](u)-\lambda^*\big\}\Big|\\ &= \inf_{u\in[-U_n,U_n]}\Big[\exp\big\{-\lambda^*-\frac{1}{2}\sigma^2u^2+\mathrm{Re}(\mathcal{F}[s](u))\big\}\Big]\\ &\gtrsim e^{-C}\exp(-\frac{1}{2}\bar{\sigma}^2U_n^2),\end{aligned}$$

where we use that $|\,\mathrm{Re}(\mathcal{F}[s](u))|\leq 1$. □

Theorem 1. Let the weight function $w\in L^1([\varepsilon,1])$ with $\varepsilon>0$, and moreover, the functions $w^1_{\sigma^2}, w^1_{\lambda}, w^1_{\gamma}$ defined by (16), (18), (20), belong to the class

$$\mathcal{P}_r := \Big\{f:\ \mathcal{F}\big[\frac{f(x)}{x^r}\big](\cdot)\in L^1\big(\mathbb{R}\big)\Big\}.$$

Let the sequence $U_n=V_n$ and the constant $\chi_\circ$ be such that $\mathrm{P}\{\mathcal{A}_n\}\to 1$ and $\chi_n\to 0$ as $n\to\infty$, see Lemma 2. Then on the event $\mathcal{A}_n$,

$$\sup_{\mathcal{S}(r,\bar{\sigma},C)}\big|\sigma_n^2-\sigma^2\big| \lesssim\ e^C\frac{\sqrt{\log(nU_n^2)}}{\sqrt{n}U_n^2}e^{\frac{1}{2}\bar{\sigma}^2U_n^2}+\frac{C}{U_n^{r+3}},$$

$$\sup_{\mathcal{S}(r,\bar{\sigma},C)}\big|\lambda_n^*-\lambda^*\big| \lesssim\ e^C\frac{\sqrt{\log(nU_n^2)}}{\sqrt{n}}e^{\frac{1}{2}\bar{\sigma}^2U_n^2}+\frac{C}{U_n^{r+1}},$$

$$\sup_{\mathcal{S}(r,\bar{\sigma},C)}\big|\gamma_n^*-\gamma^*\big| \lesssim e^C\frac{\sqrt{\log(nU_n^2)}}{\sqrt{n}U_n}e^{\frac{1}{2}\bar{\sigma}^2U_n^2}+\frac{C}{U_n^{r+2}},$$

where the supremum is taken over all distributions from the considered class $\mathcal{S}(r,\bar{\sigma},C)$.

Remark 1. Note that each first summand in the bounds presented above, appears due to the term

$$\int_{\mathbb{R}} a_n(u)\,\mathrm{Re}\big(\log(\phi_n(u))-\log(\phi(u))\big)du,$$

where $a_n(\cdot)$ is either $w^{U_n}_{\sigma^2}(\cdot), w^{U_n}_{\lambda}(\cdot)$ or $w^{U_n}_{\gamma}(\cdot)$. This term is associated with the bias of the estimate. The second summand appear due to

$$\int_{\mathbb{R}} a_n(u)\mathcal{F}\big[s\big](u)du,$$

which is closely related to the variance. The choice

$$U_n=V_n=\sqrt{\frac{\log n}{\sigma^2_{max}}},$$

where $\sigma_{max} > \overline{\sigma}$, balances these terms and yields the logarithmic rates of convergence on $\mathcal{A}_n$,

$$\sup_{\mathcal{S}(r,\bar{\sigma},C)} \left|\sigma_n^2 - \sigma^2\right| \lesssim C\left(\frac{\log n}{\sigma_{max}^2}\right)^{\frac{-(r+3)}{2}},$$

$$\sup_{\mathcal{S}(r,\bar{\sigma},C)} \left|\lambda_n^* - \lambda^*\right| \lesssim C\left(\frac{\log n}{\sigma_{max}^2}\right)^{\frac{-(r+1)}{2}},$$

$$\sup_{\mathcal{S}(r,\bar{\sigma},C)} \left|\gamma_n^* - \gamma^*\right| \lesssim C\left(\frac{\log n}{\sigma_{max}^2}\right)^{\frac{-(r+2)}{2}}.$$

As we have seen in Lemma 2(ii), this choice also guarantees that $\mathrm{P}\{\mathcal{A}_n\}$ converges to 1 at a polynomial rate.

Now we turn towards the convergence analysis of the estimates σ_n^2 and p_n introduced in Section 4. The next theorem aims to show that the convergence rates are polynomial, if the distribution of μ° is supersmooth.

Theorem 2. Assume that the assumptions of the Theorem 1 are fulfilled. Consider the subclass $\mathcal{S}^*(r,\bar{\sigma},C,\underline{p}) \subset \mathcal{S}(r,\bar{\sigma},C)$ of QID distributions in the form (2) such that $p \geq \underline{p} > 1/2$, and the function $\mathcal{H}$ defined by (10) is a characteristic function of some probability measure Λ. Moreover assume that the characteristic function ϕ° of μ° has exponential tails, i.e.

$$\left|\phi^\circ(u)\right| \leq c_1 e^{-c_2|u|^\gamma}, \qquad u \in \mathbb{R}, \tag{23}$$

with some $c_1 > 0, c_2 > (\sigma^2/2)1_{\gamma=2}$ and $\gamma \geq 2$. Then on the event $\mathcal{A}_n$,

$$\sup_{\mathcal{S}^*}\left|\sigma_n^2 - \sigma^2\right| \lesssim \frac{\mathcal{R}_n}{U_n^2}, \qquad \sup_{\mathcal{S}^*}\left|\lambda_n^* - \lambda^*\right| \lesssim \mathcal{R}_n, \qquad \sup_{\mathcal{S}^*}\left|\gamma_n^* - \gamma^*\right| \lesssim \frac{\mathcal{R}_n}{U_n},$$

where

$$\mathcal{R}_n = \mathcal{R}_n(C,\tilde{c}_1,\tilde{c}_2,\gamma) := e^C \frac{\sqrt{\log(nU_n^2)}}{\sqrt{n}} e^{\frac{1}{2}\bar{\sigma}^2 U_n^2} + \tilde{c}_1 e^{-\tilde{c}_2 \varepsilon^\gamma U_n^\gamma},$$

and $\tilde{c}_1 = -c_1 \log\left(2\underline{p} - 1\right) > 0$, $\tilde{c}_2 = c_2 - (\sigma^2/2)1_{\gamma=2} > 0$.

Remark 2. Let us again, as in Lemma 2(ii) and Remark 1, choose

$$U_n = V_n = \sqrt{\frac{\log n}{\sigma_{max}^2}},$$

where $\sigma_{max} > \overline{\sigma}$. Then

$$\mathcal{R}_n \lesssim C_1 \sqrt{\log(n)} n^{-C_2},$$

where

$$C_1 = \max\{e^{1+C}, \tilde{c}_1\}, \qquad C_2 = \min\Big(\frac{\sigma_{max}^2 - \overline{\sigma}^2}{2\sigma_{max}^2}, \frac{\tilde{c}_2 \varepsilon^\gamma}{\sigma_{max}^\gamma}\Big) > 0.$$

Therefore, under this choice of the sequences U_n and V_n, the rates of convergence presented in Theorem 2 are polynomial. In particular, on the event $\mathcal{A}_n$,

$$\sup_{\mathcal{S}^*} \big|\sigma_n^2 - \sigma^2\big| \lesssim C_1 \sigma_{max}^2 n^{-C_2}.$$

and, moreover,

$$\sup_{\mathcal{S}^*} \big|p_n - p\big| \lesssim C_1 \sqrt{\log(n)} n^{-C_2}.$$

Remark 3. The most crucial difference from the general case is in the particular form (9) of the measure ν and the condition (23) on ϕ°. Note that this condition with $\gamma = 2$ is fulfilled for our key examples 2 and 3. Assumptions of this type are widely used in nonparametric statistics, see [15].

The last result of this section shows that the nonparametric estimate of the second component of the mixture (21) has also polynomial rate of convergence, provided that the parameters are estimated with polynomial accuracy.

Theorem 3. Assume that the assumptions of Theorem 2 are fulfilled, and the parameter U_n, constants C_1 and C_2 are fixed as in Remark 2. Consider the subclass $\mathcal{S}^{**} = \mathcal{S}^{**}(r, \underline{\sigma}, \overline{\sigma}, C, \underline{p}, \overline{p}) \subset \mathcal{S}^*(r, \overline{\sigma}, C, \underline{p})$ consisting of the distributions with $\sigma \in \big[\underline{\sigma}, \overline{\sigma}\big]$ and $p \in \big[\underline{p}, \overline{p}\big]$ for some $0 < \underline{\sigma} < \overline{\sigma}$ and $1/2 < \underline{p} < \overline{p} < 1$. Moreover, assume that g° is twice differentiable and $g^\circ, (g^\circ)'' \in L^2(\mathbb{R})$. Let K be a kernel of order 1, that is,

$$\int_{\mathbb{R}} K^2(u) du < \infty, \qquad \int_{\mathbb{R}} u^2 K(u) du < \infty.$$

Then

$$\sup_{\mathcal{S}^{**}} \int_{\mathbb{R}} \mathrm{E}\Big[\big(g_n^\circ(x) - g^\circ(x)\big)^2 | \mathcal{A}_n\Big] \lesssim \frac{1}{(1-\overline{p})^4} \Big(\frac{1}{nh} \int_{\mathbb{R}} K^2(u) du + \frac{h^4}{4} \big(\int_{\mathbb{R}} |u^2 K(u)| du \big)^2 \int_{\mathbb{R}} \big(g''(x)\big)^2 dx + C_3 \log(n) n^{-2C_2} \Big),$$

where $C_3 = C_1^2 \big(\underline{\sigma}^{-1} + \int_{\mathbb{R}} g^2(x)\, dx\big)$.

Remark 4. Due to the fact that the bandwidth parameter h does not appear in the last term of the r.h.s., the approach for selecting the optimal bandwidth value remains the same as in the kernel density estimation problem, see [21]. With the optimal choice of the parameter h, the first two terms are of the order $n^{-4/5}$, yielding the polynomial rate of convergence for g_n°.

6 Numerical study

In this section we apply our algorithm to several models of the form (2). First, we study the classic variance mixture model of two normal distributions. Then we proceed to the Bart Simpson distribution (known also as "the claw"), which is a mixture of 6 normal distributions. For both models we compare our approach with the well-known EM algorithm. Finally we analyse the performance of our approach for Example 3, which involves the convolution of a normal distribution and another ID distribution as a second component of the mixture.

6.1 Two-component normal mixture

The first considered model is a mixture of two normal distributions, where the resulting density has the following form

$$g(x) = p\varphi_{\sigma_1}(x) + (1-p)\varphi_{\sigma_2}(x),$$

where $0 < \sigma_1 < \sigma_2$, $\frac{1}{2} < p < 1$, $x \in \mathbb{R}$. The resulting distribution is QID as it was discussed in Example 2. For the numerical example, we fix the parameters of this distribution as $p = 0.75$, $\sigma_1^2 = 0.1$, $\sigma_2^2 = 0.5$.

Figure 1 shows the plot of real parts of logarithms of $\phi(u)$ (orange line) and several realisations of its estimate $\phi_n(u)$ given by (14) for $n = 1000$. As it can be seen in the graph, the deviation increases with the growth of u. This observation influences the choice of $U_n = V_n$, because on the interval $[\varepsilon U_n, U_n]$, one should simultaneously have $\phi_n(u) \approx \phi(u)$ and $\phi(u) \approx \mathrm{i}\gamma^*(u) - (1/2)\sigma^2 u^2 - \lambda^*$. For instance, for $n = 1000$ these conditions are fulfilled with $U_n = V_n = 8$.

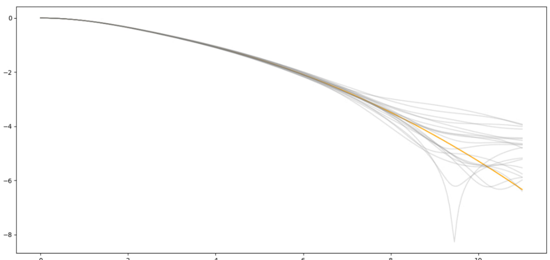
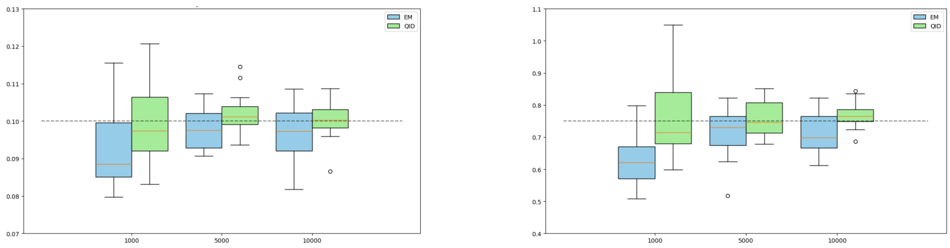


Figure 1: The plot of $\mathrm{Re}(\log(\phi(u)))$ (orange line) and plots of $N = 20$ realisations of its estimate $\mathrm{Re}(\log(\phi_n(u)))$ (grey lines).

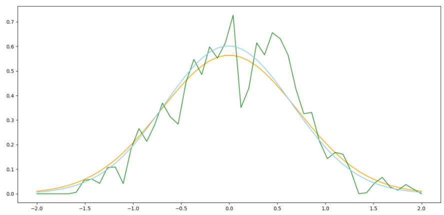
Figure 2 shows the quality of the estimators of σ_1^2 and p obtained by our method and by the EM algorithm based on n observations from $N = 20$ simulation runs. The number n takes the values $1000, 5000, 10000$. Both estimates converge to the true values of the parameters, but the accuracy of our estimate increases with the growth of n, compared to the estimates obtained by the EM algorithm.



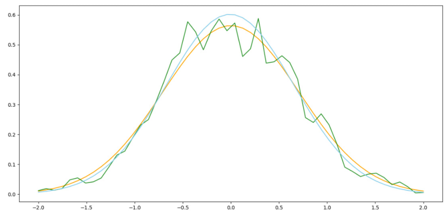
(a) $\widehat{\sigma}_1^2$ (b) $\widehat{p}$

Figure 2: Boxplots for the estimates of σ_1^2 and p obtained by the EM algorithm (blue boxes) and our approach (green boxes) for the two-component normal mixture model.

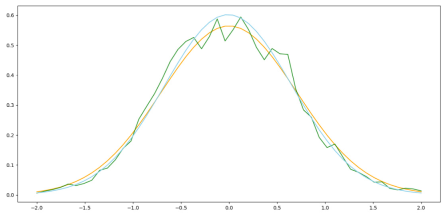
Lastly, Figure 3 shows the estimates of $g^\circ(\cdot)$ obtained by both algorithms for different number of observations n. The quality of our algorithm improves as n increases, resulting in a performance comparable to that of the EM algorithm for large sample sizes.



(a) $n = 1000$



(b) $n = 5000$



(c) $n = 10000$

Figure 3: Plots of the true function $g^\circ(x)$ (orange line) and its estimates, obtained by the EM algorithm (blue line) and our inference method (green line) for the two-component normal mixture model.

6.2 Bart Simpson distribution

In Example 6.1 from [23], the Bart Simpson distribution is defined via the density function

$$g(x) = p\varphi_{(0,\sigma_1)}(x) + \frac{1-p}{5}\sum_{j=0}^{4}\varphi_{((j/2)-1,\,\sigma_2)}(x),$$

where $p = 0.5, \sigma_1 = 0.1, \sigma_2 = 0.01$, and $\varphi_{(m,\sigma)}$ is a density of the normal distribution with parameters m and σ^2. The name of this distribution is derived from the shape of its probability density function, which resembles the hair on Bart Simpson's head.

With this choice of parameters p, σ_1, σ_2, the conditions of Corollary 2 are not satisfied. Nevertheless, if $p = 0.5 + \delta$ with any (small) δ and $\sigma_1 < \sigma_2$, then the conditions hold and the distribution is QID.

For this simulation study we take $\delta = 0.001, \sigma_1 = 0.05, \sigma_2 = 0.1$. Figure 4 depicts the histogram of the simulated data for $n = 1000$ observations. Note that there are 5 distinct "spikes" in the density function, although they are not as pronounced as in the original example. Below we call this distribution "the modified Bart Simpson model".

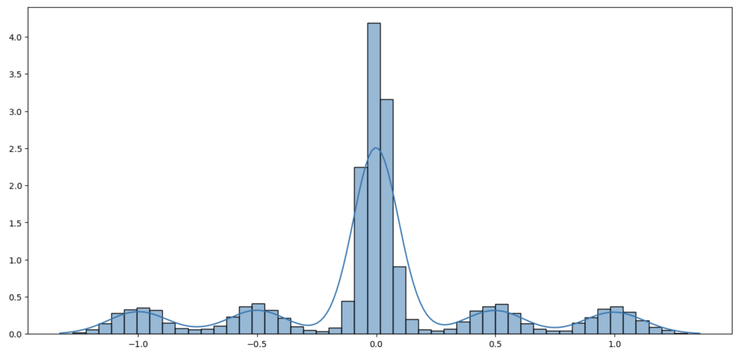
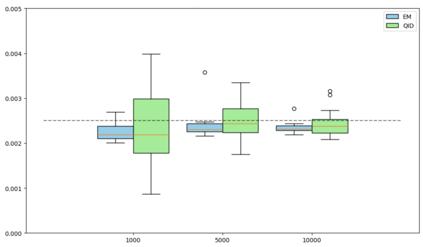
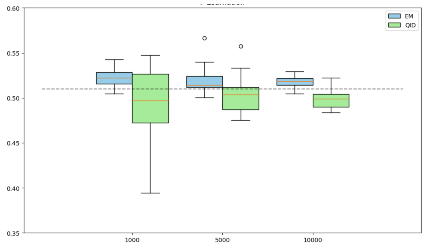


Figure 4: Histogram from the modified Bart Simpson model and the graph of the true density $g(x)$.

Figure 5 shows the comparison of the estimates of σ_1^2 and p^* between the EM algorithm and our method based on n observations from $N = 20$ simulation runs. In contrast to the previous example, the estimate obtained by the EM algorithm has a noticeably smaller deviation and converges more quickly. However, our inference approach also leads to a good quality of estimation.



(a) $\widehat{\sigma}_1^2$



(b) $\widehat{p^*}$

Figure 5: Boxplots for the estimates of σ_1^2 and p obtained by the EM algorithm (blue boxes) and our approach (green boxes) for the modified Bart Simpson model.

Figure 6 presents the estimate of the normal component $p\varphi_{(0,\sigma_1)}(x)$ and the "rest part" $(1-p_n)g_n^\circ(x)$, which in this particular case is the estimate of the mixture of 5 normal components. We see that our estimate of $(1-p_n)g^\circ(x)$ replicates all 5 spikes of the true density.

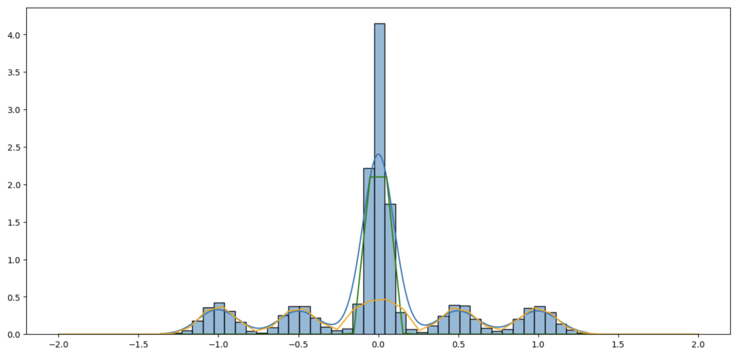


Figure 6: The histogram from the simulated data, plot of the true density function $g(x)$ (blue line) and plots of the estimates $p_n\varphi_{\sigma_n}(x)$, $(1-p_n)g_n^\circ(x)$ (green and orange lines respectively) for the modified Bart Simpson model.

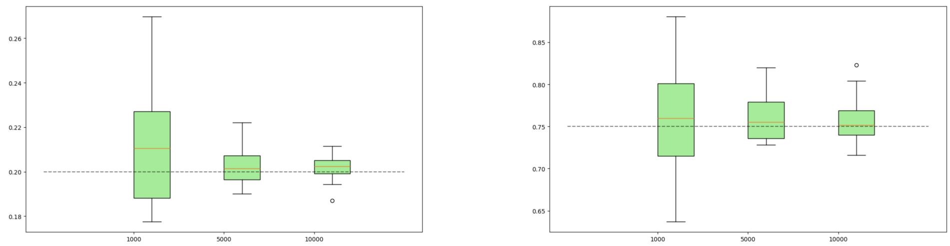
6.3 Another example of model (2)

In Example 3, we have considered the distribution

$$\mu = p\tilde{\mu}_{\sigma_1} + (1-p)\mu^\circ,$$

where μ° is ID with Lévy triplet $(\gamma, \bar{\sigma}^2, \nu)$. Recall that this distribution is QID, if $p > 1/2$, and $\bar{\sigma} > \sigma_1$. For this numerical example, we fix μ° as the convolution of standard Student's distribution with d degrees of freedom and a normal distribution with zero mean and variance equal to σ_2^2, where $0 < \sigma_1 < \sigma_2$. In this study we fix $d = 3, p = 0.75$, $\sigma_1^2 = 0.2$, $\sigma_2^2 = 0.5$.

Figure 7 depicts the estimates of σ_1^2 and p based on n observations from $N = 20$ simulation runs. The number n takes the values $1000, 5000, 10000$. Here we again conclude that the estimates converge to the true values and the variances of the estimates decrease with the growth of n.



(a) $\widehat{\sigma}_1^2$ (b) $\widehat{p}$

Figure 7: Boxplots for the estimates of σ_1^2 and p for the case when the second component of the mixture is the convolution of Student's and normal distributions.

Analogously to Figure 6, Figure 8 depicts the decomposition of the density g into two parts. It can be seen that the first part, which corresponds to the normal component (green line), is the main part of the resulting mixture density, while the second part (orange line) makes the quality of the estimate better.

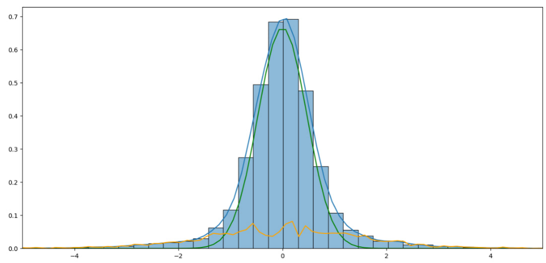
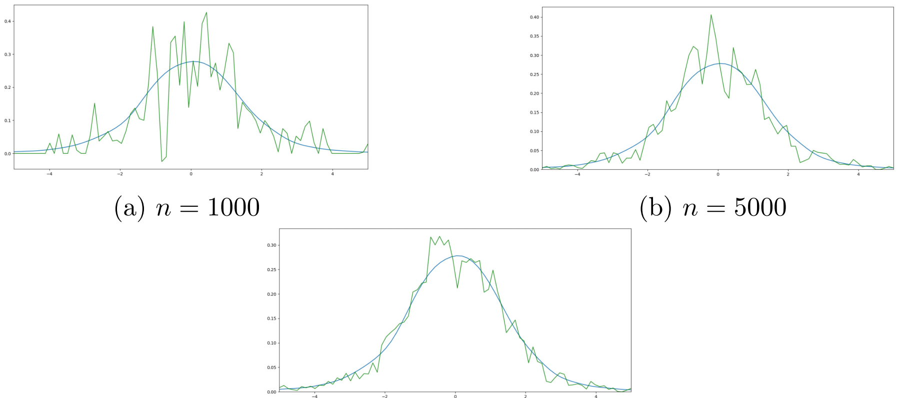


Figure 8: Histogram of the simulated data, plot of the true density $g(x)$ (blue line) and plots of the first and the second components (green and orange lines respectively) for the case when the second component of the mixture is the convolution of Student's and normal distributions.

Lastly, we consider more precisely the estimation of the density g° of the second component. Figure 9 illustrates the accuracy of our estimate $g_n^\circ(x)$ for different n. The quality of this estimate increases with the growth of n.



(a) $n = 1000$ (b) $n = 5000$

(c) $n = 10000$

Figure 9: Plots of the true functions $g^\circ(x)$ (blue line) and their estimates $g_n^\circ(x)$ (green line) based on n observations for the case when the second component of the mixture is the convolution of Student's and normal distributions.

7 Proofs

7.1 Proof of Theorem 1

Below we will prove the upper bound for the mean-squared error of σ_n^2. The proofs for λ_n^* and γ_n^* follow the same lines.

Using the representation (16) and the properties (17), we get

$$\begin{aligned}\sigma_n^2 - \sigma^2 &= \int_{\mathbb{R}} w_{\sigma^2}^{U_n}(u)\,\mathrm{Re}\big(\log(\phi_n(u)) - \log(\phi(u))\big)du + \sigma^2 \int_{\mathbb{R}} \frac{u^2}{2} w_{\sigma^2}^{U_n}(u)du \\ &\qquad + \int_{\mathbb{R}} w_{\sigma^2}^{U_n}(u)\,\mathrm{Re}\big(\log(\phi(u))\big)du + \lambda^* \int_{\mathbb{R}} w_{\sigma^2}^{U_n}(u)du \\ &= \int_{\mathbb{R}} w_{\sigma^2}^{U_n}(u)\,\mathrm{Re}\big(\log(\phi_n(u)) - \log(\phi(u))\big)du \\ &\qquad + \int_{\mathbb{R}} w_{\sigma^2}^{U_n}(u)\mathcal{F}[s](u)du,\end{aligned}$$

where the second equality follows from (11). Since $|\log(1+z) - z| \leq 2|z|^2$ for any $|z| < 1/2$, we have on the event $\mathcal{A}_n$ for n large enough

$$\log(\phi_n(u)) - \log(\phi(u)) = \log\Big(\frac{\phi_n(u) - \phi(u)}{\phi(u)} + 1\Big) = \frac{\phi_n(u) - \phi(u)}{\phi(u)} + R_n(u),$$

where $R_n(u) = O\Big(|(\phi_n(u) - \phi(u))/\phi(u)|^2\Big)$, $u \in \mathbb{R}$. Therefore, on $\mathcal{A}_n$ we get

$$\Big|\sigma_n^2 - \sigma^2\Big| \leq |I_1| + |I_2| + |I_3|,$$

where

$$
\begin{aligned}
I_1 &:= \Big|\int_{\mathbb{R}} w_{\sigma^2}^{U_n}(u)\,\mathrm{Re}\Big(\frac{\phi_n(u)-\phi(u)}{\phi(u)}\Big)du\Big|,\\
I_2 &:= \Big|\int_{\mathbb{R}} w_{\sigma^2}^{U_n}(u)\,\mathrm{Re}(\mathrm{R}_n(u))du\Big|,\\
I_3 &:= \Big|\int_{\mathbb{R}} w_{\sigma^2}^{U_n}(u)\mathcal{F}\big[s\big](u)du\Big|.
\end{aligned}
$$

Below we separately consider these three terms. For the terms I_1 and I_2 we have

$$
\begin{aligned}
I_j &\le \max_{u\in[-U_n,U_n]}\Big|\frac{\phi_n(u)-\phi(u)}{\phi(u)}\Big|^j\Big(\int_{\mathbb{R}}|w_{\sigma^2}^{U_n}(u)|du\Big)\\
&= \frac{\chi_n^j}{U_n^2}\Big(\int_{\mathbb{R}}|w_{\sigma^2}^{1}(u)|du\Big)\lesssim\frac{\chi_n^j}{U_n^2},\qquad j=1,2.
\end{aligned}
$$

Therefore, $I_2 \lesssim I_1$, and due to Lemma 2 (ii),

$$
I_1 \lesssim \exp(C)\frac{\sqrt{\log(nU_n^2)}}{\sqrt{n}U_n^2}\exp\big(\frac{\overline{\sigma}^2U_n^2}{2}\big).
$$

For I_3, we apply the Plancherel identity

$$
\begin{aligned}
I_3 &= |\int_{\mathbb{R}} w_{\sigma^2}^{U_n}(u)\mathcal{F}\big[s\big](u)du|\\
&= 2\pi|\int_{\mathbb{R}_+} s^{(r)}(x)\overline{\mathcal{F}^{-1}}\big[\frac{w_{\sigma^2}^{U_n}(t)}{(it)^r}\big](x)dx|\\
&\le U_n^{-(r+3)}||s^{(r)}||_\infty\cdot||\mathcal{F}\big[\frac{w_{\sigma^2}^{1}(u)}{u^r}\big]||_{L^1},
\end{aligned}
$$

yielding the upper bound $I_3 \lesssim CU_n^{-(r+3)}$. Combining these bounds, we arrive at the desired result.

7.2 Proof of Theorem 2

The proof is similar to the proof of Theorem 1 with the different approach for the estimation of the term I_3. Using the properties of the Lévy measure for the QID distributions from the class $\mathcal{S}^*(r,\bar{\sigma},C,\underline{p})$, we get the following upper bound

$$
\begin{aligned}
I_3 &= \Big| \int_{\mathbb{R}_+} w_{\sigma^2}^{U_n}(u) \mathcal{F}\Big[\sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} (q/p)^m \Lambda^{*m} \Big](u) du \Big| \\
&= \Big| \int_{\mathbb{R}_+} w_{\sigma^2}^{U_n}(u) \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} (q/p)^m \big(\mathcal{F}[\Lambda](u)\big)^m du \Big| \\
&= \Big| U_n^{-2} \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} (q/p)^m \int_{\varepsilon}^{1} w_{\sigma^2}^{1}(u) \big(\mathcal{F}[\Lambda](uU_n)\big)^m du \Big| \\
&\leq U_n^{-2} \sum_{m=1}^{\infty} \frac{(q/p)^m}{m} \int_{\varepsilon}^{1} |w_{\sigma^2}^{1}(u)| \cdot \big|\mathcal{F}[\Lambda](uU_n)\big|^m du.
\end{aligned}
$$

Due to our assumption (23), we get $\big|\mathcal{F}[\Lambda](u)\big| \leq c_1 e^{-\tilde{c}_2 |u|^\gamma}$ for $|u|$ large enough, where $\tilde{c}_2 = c_2 - (\sigma^2/2)1_{\gamma=2} > 0$. Then for large n

$$
\begin{aligned}
I_3 &\leq c_1 U_n^{-2} \int_{\varepsilon}^{1} |w_{\sigma^2}^{1}(u)| e^{-\tilde{c}_2 (uU_n)^\gamma} du \cdot \sum_{m=1}^{\infty} \frac{(q/p)^m}{m} \\
&\leq c_1 U_n^{-2} e^{-\tilde{c}_2 \varepsilon^\gamma U_n^\gamma} \Big(\int_{\varepsilon}^{1} |w_{\sigma^2}^{1}(u)| du \Big) \cdot \log\Big(\frac{p}{p-q}\Big) \\
&\lesssim c_1 \log\Big(\frac{1}{2\underline{p}-1}\Big) U_n^{-2} e^{-\tilde{c}_2 \varepsilon^\gamma U_n^\gamma}.
\end{aligned}
$$

Combining this inequality with the upper bounds for I_1 and I_2 obtained in the previous section completes the proof.

7.3 Proof of Remark 2

To show the upper bound for the estimate $p_n = e^{-\lambda_n^*}$, we use the Taylor series decomposition with the stochastic remainder term. For a k-times differentiable function $f : \mathbb{C} \to \mathbb{C}$, this decomposition reads as

$$
\begin{aligned}
f(z) = &\sum_{j=0}^{k-1} \frac{f^{(j)}(a)}{j!} (z-a)^j \\
&+ \frac{1}{(k-1)!} \mathrm{E}_\tau \Big[(1-\tau)^{k-1} f^{(k)}(a + \tau(z-a)) \Big] (z-a)^k, \quad (24)
\end{aligned}
$$

where τ is uniformly distributed on $[0,1]$, see [3], Appendix A.1. Therefore,

$$
\begin{aligned}
p_n - p &= e^{-\lambda_n^*} - e^{-\lambda^*} \\
&= -e^{-\lambda^*} \Big((\lambda_n^* - \lambda^*) - (\lambda_n^* - \lambda^*)^2 \mathrm{E}_\tau \Big[(1-\tau) e^{-\tau(\lambda_n^* - \lambda^*)} \Big] \Big),
\end{aligned}
$$

and we conclude that

$$
\big| p_n - p \big| \leq |\lambda_n^* - \lambda^*| + |\lambda_n^* - \lambda^*|^2 e^{|\lambda_n^* - \lambda^*|} \lesssim \mathcal{R}_n, \quad (25)
$$

and the result follows.

7.4 Proof of Theorem 3

We consider the convergence rate for estimator (22) of $g_n^\circ(x)$. Firstly, we rewrite the estimation error as following

$$
\begin{aligned}
&\int_{\mathbb{R}} \mathrm{E}\Big[\big(g_n^\circ(x) - g^\circ(x)\big)^2 |\mathcal{A}_n\Big] dx \\
&\quad = \int_{\mathbb{R}} \mathrm{E}\Big[\Big(\frac{g_n(x) - p_n\varphi_{\sigma_n}(x)}{1-p_n} - \frac{g(x) - p\varphi_\sigma(x)}{1-p}\Big)^2 |\mathcal{A}_n\Big] dx \\
&\quad = \int_{\mathbb{R}} \mathrm{E}\Big[\frac{\Big((1-p)(I_1+I_2) + I_3 + I_4\Big)^2}{(1-p_n)^2(1-p)^2} |\mathcal{A}_n\Big] dx \\
&\quad \lesssim \frac{1}{\left(1-\overline{p}\right)^4} \int_{\mathbb{R}} \mathrm{E}\Big[\big(I_1^2 + I_2^2 + I_3^2 + I_4^2\big)|\mathcal{A}_n\Big] dx,
\end{aligned}
$$

where

$$
\begin{aligned}
I_1 &= |g_n(x) - g(x)|, \qquad & I_2 &= p_n|\varphi_{\sigma_n}(x) - \varphi_\sigma(x)|, \\
I_3 &= |p - p_n|\varphi_\sigma(x), \qquad & I_4 &= |p - p_n| g(x),
\end{aligned}
$$

and we use that $\underline{p} \le p \le \overline{p}$ and

$$
|1 - p_n| \ge |1-p| - |p - p_n| \ge |1 - \overline{p}| + o(1), \qquad n \to \infty,
$$

on the event $\mathcal{A}_n$. Below we consider the terms $I_1, ..., I_4$ separately. For the first term, we use Theorem 1.3 from [21], which yields

$$
\mathrm{E}\big[I_1^2|\mathcal{A}_n\big] \lesssim \frac{1}{nh}\int_{\mathbb{R}} K^2(u)du + \frac{h^4}{4}\Big(\int_{\mathbb{R}} |u^2K(u)|du\Big)^2 \int_{\mathbb{R}} (g''(x))^2 dx. \tag{26}
$$

Note that here we use that $g'' \in L^2(\mathbb{R})$ due to the condition $g''_\circ \in L^2(\mathbb{R})$. For the second term, we use the decomposition (24) with the function $f(z) = \varphi_z(x)$ and any fixed $x \in \mathbb{R}$. Note that on the event $\mathcal{A}_n$,

$$
\xi := \sigma^2 + \tau(\sigma_n^2 - \sigma^2) \in \big[\underline{\sigma}^2, \overline{\sigma}^2\big]
$$

for n large enough, see Remark 2. This observation yields the following upper bound for the difference between φ_{σ_n} and φ_σ:

$$
\begin{aligned}
|\varphi_{\sigma_n}(x) - \varphi_\sigma(x)| &= \Big|\frac{1}{\sqrt{2\pi}\sigma_n}\exp\Big(\frac{-x^2}{2\sigma_n^2}\Big) - \frac{1}{\sqrt{2\pi}\sigma}\exp\Big(\frac{-x^2}{2\sigma^2}\Big)\Big| \\
&= \Big|\mathrm{E}_\tau\Big[\frac{1}{\xi}\Big(\frac{x^2}{\xi} - 1\Big)\varphi_\xi(x)\Big](\sigma_n^2 - \sigma^2)\Big| \\
&\lesssim \frac{1}{\underline{\sigma}^3}\Big(\frac{x^2}{\underline{\sigma}^2} + 1\Big)\exp\Big(-\frac{x^2}{2\overline{\sigma}^2}\Big)\frac{\mathcal{R}_n}{U_n^2}.
\end{aligned}
$$

Therefore,

$$\begin{aligned}
\mathrm{E}\big[I_2^2|\mathcal{A}_n\big] &\lesssim \mathrm{E}\big[p_n^2|A_n\big]\int_{\mathbb{R}}\Big(\frac{1}{\underline{\sigma}^3}\big(\frac{x^2}{\underline{\sigma}^2}+1\big)\exp\big(-\frac{x^2}{2\overline{\sigma}^2}\big)\Big)^2dx\cdot\frac{\mathcal{R}_n^2}{U_n^4}\\
&= \mathrm{E}\big[p_n^2|A_n\big]\cdot\frac{\overline{\sigma}\sqrt{\pi}}{\underline{\sigma}^6}\Big(\frac{3}{4}\big(\frac{\overline{\sigma}}{\underline{\sigma}}\big)^4+\big(\frac{\overline{\sigma}}{\underline{\sigma}}\big)^2+1\Big)\cdot\frac{\mathcal{R}_n^2}{U_n^4}\\
&\lesssim \frac{\overline{\sigma}}{\underline{\sigma}^6}\Big(\frac{3}{4}\big(\frac{\overline{\sigma}}{\underline{\sigma}}\big)^4+\big(\frac{\overline{\sigma}}{\underline{\sigma}}\big)^2+1\Big)\cdot\frac{\mathcal{R}_n^2}{U_n^4},
\end{aligned}\tag{27}$$

since $\mathrm{E}\big[p_n^2|\mathcal{A}_n\big]\lesssim\mathrm{E}\big[(p_n-p)^2|\mathcal{A}_n\big]+p^2\leq 1$ for n large enough and

$$\int_{\mathbb{R}}x^ne^{-x^2}dx=\frac{n!\sqrt{\pi}}{2^n(n/2)!}$$

for even n. Next, we get for the third term

$$\begin{aligned}
\mathrm{E}\big[I_3^2|\mathcal{A}_n\big] &= \mathrm{E}\big[(p_n-p)^2|\mathcal{A}_n\big]\cdot\int_{\mathbb{R}}\varphi_\sigma^2(x)dx\\
&= \frac{1}{2\sqrt{\pi}\sigma}\mathrm{E}\big[(p_n-p)^2|\mathcal{A}_n\big]\lesssim\frac{1}{\underline{\sigma}}\mathcal{R}_n^2.
\end{aligned}\tag{28}$$

Analogously, for the term $\mathrm{E}I_4^2$ we get

$$\mathrm{E}\big[I_4^2|\mathcal{A}_n\big]=\mathrm{E}\big[(p_n-p)^2|\mathcal{A}_n\big]\cdot\int_{\mathbb{R}}g^2(x)dx\lesssim\int_{\mathbb{R}}g^2(x)dx\cdot\mathcal{R}_n^2.\tag{29}$$

Combining these bounds, we arrive at desired result.

Acknowledgment

The article was prepared within the framework of the HSE University Basic Research Program.

References

[1] Alexeev, I., and Khartov, A. (2023). Spectral representations of characteristic functions of discrete probability laws. *Bernoulli*, 29(2), 1392 – 1409.

[2] Belomestny, D. (2010). Spectral estimation of the fractional order of a Lévy process. *Ann. Stat.* 38(1): 317 – 351.

[3] Belomestny, D., Morozova, E., and Panov, V. (2024). Decompounding under general mixing distributions. *ArXiv preprint*, 2405.05419.

[4] Belomestny, D. and Reiss, M. (2015). Estimation and calibration of Lévy models via Fourier methods. In *Lévy Matters IV: Estimation for Discretely Observed Lévy Processes*, 1 – 76.

[5] Berger, D., and Kutlu, M. (2023). Quasi-infinite divisibility of a class of distributions with discrete part. *Proceedings of the American Mathematical Society*, 151(05), 2211 – 2224.

[6] Berger, D., Kutlu, M. and Lindner, A. (2022). On multivariate quasi-infinitely divisible distributions. *A Lifetime of Excursions Through Random Walks and Lévy Processes: A Volume in Honour of Ron Doney's 80th Birthday*, 87 – 120.

[7] Chhaiba, H., Demni, N., and Mouayn, Z. (2016). Analysis of generalized negative binomial distributions attached to hyperbolic Landau levels. *Journal of Mathematical Physics*, 57 (7): 072103.

[8] Comte, F., and Genon-Catalot, V. (2010). Nonparametric adaptive estimation for pure jump Lévy processes. *Annales de l'Institut Henri Poincaré Probabilités et Statistique,* 46 (3): 595 – 617.

[9] Cuppens, R. (1970) Quelques nouveaux résultats en arithmétique des lois de probabilité (French, with English summary). *Les probabilités sur les structures algébriques (Actes Colloq. Internat. CNRS, No. 186, Clermont-Ferrand, 1969), Éditions Centre Nat. Recherche Sci.,* pp. 97 – 112.

[10] Demni, N., and Mouayn, Z. (2015). Analysis of generalized Poisson distributions associated with higher Landau levels. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 18 (04), 1550028.

[11] Gugushvili, S. (2012). Nonparametric inference for discretely sampled Lévy processes. *Annales de l'Institut Henri Poincaré Probabilités et Statistique,* 48 (1): 282 – 307.

[12] Lindner, A., Pan, L., and Sato, K. I. (2018). On quasi-infinitely divisible distributions. *Transactions of the American Mathematical Society*, 370(12), 8483 – 8520.

[13] Linnik, Y. and Ostrovskii, I. (1977). *Decomposition of Random Variables and Vectors*, American Mathematical Society, Providence, R. I.Translated from the Russian; Translations of Mathematical Monographs, Vol. 48.

[14] Madan, D. B., Schoutens, W., and Wang, K. (2023). Modeling the bid and ask prices of options. *Journal of Computational Finance,* 26(4), 1 – 36.

[15] Meister, A. (2019). *Deconvolution Problems in Nonparametric Statistics*, Springer Science and Business Media.

[16] Nakamura, T. (2022). Zeta distributions generated by Dirichlet series and their (quasi) infinite divisibility. *ArXiv preprint*, 2209.13257.

[17] Neumann, M. H., and Reiß, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15(1): 223 – 248.

[18] Passeggeri, R. (2023). On quasi-infinitely divisible random measures. *Bayesian Analysis*, 18(1), 253 – 286.

[19] Passeggeri, R. (2020). Spectral representations of quasi-infinitely divisible processes. *Stochastic Processes and their Applications*, 130(3), 1735 – 1791.

[20] Steutel, F.W. and Van Harn, K. (2004). *Infinite Divisibility of Probability Distributions on the Real Line*. Marcel Dekker, Inc, New York.

[21] Tsybakov, A. (2008) *Introduction to Nonparametric Estimation.* Springer Science and Business Media.

[22] Zhang, H., Liu, Y., and Li, B. (2014). Notes on discrete compound Poisson model with applications to risk theory. *Insurance: Mathematics and Economics*, 59: 325 – 336.

[23] Wasserman, L. (2006). All of Nonparametric Statistics. *Springer Texts in Statistics.*