Annals of the Institute of Statistical Mathematics (2022) 74:113–142
https://doi.org/10.1007/s10463-021-00790-7

Efficient estimation methods for non-Gaussian regression models in continuous time

Evgeny Pchelintsev[1] · Serguei Pergamenshchikov[2] · Maria Povzun[1]

Received: 21 May 2020 / Revised: 14 January 2021 / Accepted: 19 January 2021 /
Published online: 3 March 2021

Abstract
In this paper, we develop an efficient nonparametric estimation theory for continuous time regression models with non-Gaussian Lévy noises in the case when the unknown functions belong to Sobolev ellipses. Using the Pinsker's approach, we provide a sharp lower bound for the normalized asymptotic mean square accuracy. However, the main result obtained by Pinsker for the Gaussian white noise model is not correct without additional conditions for the ellipse coefficients. We find such constructive sufficient conditions under which we develop efficient estimation methods. We show that the obtained conditions hold for the ellipse coefficients of an exponential form. For exponential coefficients, the sharp lower bound is calculated in explicit form. Finally, we apply this result to signals number detection problems in multi-pass connection channels and we obtain an almost parametric convergence rate that is natural for this case, which significantly improves the rate with respect to power-form coefficients.

Keywords Regression model · Lévy process · Asymptotic efficiency · Weighted least squares estimates · Pinsker constant · Quadratic risk

This work was supported by RSF, Grant no 20-61-47043.

✉ Evgeny Pchelintsev
evgen-pch@yandex.ru

1 International Laboratory of Statistics of Stochastic Processes and Quantitative Finance, Tomsk State University, 36 prosp. Lenina, 634050 Tomsk, Russia

2 Laboratoire de Mathématiques Raphael Salem, Université de Rouen, Avenue de l'Université, BP.12, 76801 Saint-Etienne-du-Rouvray, France

1 Introduction

1.1 Problem

In this paper, we consider the estimation problem for the nonparametric regression model in continuous time:

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \varepsilon \mathrm{d}\xi_t\,, \quad 0 \le t \le 1\,, \tag{1}$$

where $S(t) \in \mathcal{L}_2[0,1]$ is an unknown function, the noise $\left(\xi_t\right)_{0\leqslant t\leqslant 1}$ is a Lévy process defined in Sect. 2, and $\varepsilon > 0$ is a noise intensity. The problem is to estimate the unknown function S in the model (1) on the basis of observations $(y_t)_{0\le t\le 1}$ as $\varepsilon \to 0$. Such models are widely used in statistical radio-physics to estimate the unknown signal S in connection channels in the case when ratio "signal/noise" tends to infinity (see, for example, Ibragimov and Khasminskii 1981; Kassam 1988; Kutoyants 1994 and the references therein). The main goal of this paper is to develop efficient estimation methods on the basis of the approach proposed by Pinsker (1981), i.e. we study the efficient estimation in the minimax sense for the unknown functions S from the Sobolev ellipse defined as

$$\Theta = \left\{ S \in \mathcal{L}_2[0,1] \;:\; \sum_{j=1}^{\infty} a_j \theta_j^2 \leqslant \mathbf{r} \right\}, \tag{2}$$

where $(a_j)_{j\in\mathbb{N}}$ is non-negative coefficients, $(\theta_j)_{j\in\mathbb{N}}$ is Fourier coefficients for the function S, and the radius $\mathbf{r} > 0$ is some fixed constant. In this paper, we use the quadratic risk

$$\mathcal{R}(\widehat{S}, S) := \mathbf{E}_S \|\widehat{S} - S\|^2 \quad \text{and} \quad \|S\|^2 = \int_0^1 S^2(t)\mathrm{d}t\,, \tag{3}$$

where $\mathbf{E}_S(\cdot) = \mathbf{E}_S^{\varepsilon}(\cdot)$ is the expectation over the distribution $\mathbf{P}_S = \mathbf{P}_S^{\varepsilon}$ of the process (1) corresponding to the function S .

Our goal is to minimize the maximal value of this risk for sufficient small parameter ε, i.e.

$$\inf_{\widehat{S}\in\Xi_\varepsilon} \sup_{S\in\Theta} \mathcal{R}(\widehat{S}, S) \quad \text{as} \quad \varepsilon \to 0\,, \tag{4}$$

where Ξ_ε is the class of all possible estimators for the function S, i.e. any functions measurable with respect to $\sigma\{y_t\,,\, 0 \le t \le 1\}$.

1.2 Motivations

In the particular case, when $(\xi_t)_{0\le t\le 1}$ is the Brownian motion, this estimation problem is very popular in the statistics of random processes (see, for example, Pinsker 1981; Kutoyants 1984; Tsybakov 2009 and the references therein) for the both parametric and nonparametric settings. Moreover, for the nonparametric Gaussian models (1), adaptive efficient methods have been proposed for different estimation problems (see, for example, Lepski 1990; Lepski and Spokoiny 1997; Tsybakov 1998 and the references therein). As to the non-Gaussian models, firstly for a parametric setting such problems were studied in Pchelintsev (2013), Konev et al. (2014) for the noises defined through the compound Poisson processes. For the non-Gaussian semimartingale nonparametric models (1), general adaptive efficient nonparametric estimation methods were developed for the risks (3) by Konev and Pergamenshchikov (2009a; b; 2012; 2015). Later, non-Gaussian models defined through stochastic differential equations were used in very important practical applications such as, for example, signal processing problems (Pchelintsev and Pergamenshchikov (2018; 2019); Pchelintsev et al. 2018; Beltaief et al. 2020), the analysis of neuron systems (Hodara et al. 2018). For non-Gaussian Lévy models, improved nonparametric estimation methods were suggested by Pchelintsev et al. (2018; 2019). Usually, to provide efficiency properties for the mean square accuracy (3), one uses the lower bounds methods based on the Bayesian risk approach developed by Pinsker for the nonparametric problems (see also Pinsker 1981; Nemirovskii 2000; Tsybakov 2009 and the references therein).

However, in all these papers efficiency properties are established only for power coefficients a_j in (2), i.e. $a_j = \mathrm{O}(j^\alpha)$ for some $\alpha > 0$. As it is turned out, this is very limited in various applications. For example, in the signals number detection problem in multi-pass connexion channels (Sect. 6), the efficient estimation for the ellipse with the power coefficients provides only power convergence rate $\varepsilon^{2\beta}$ for some $0 < \beta < 1$. But, in this case, the unknown signal S has some special parametric form with unknown parameter dimension. It is intuitively clear that the optimal convergence rate should be almost parametric, i.e. $\approx \varepsilon^2$. To obtain such rate, one needs to use the ellipse with the coefficients of some exponential form. In this paper, we develop efficient estimation procedures for such functional classes.

1.3 Main investments

To study efficiency properties, firstly, one obtains a lower bound for the risk (3), and then one needs to find an estimator whose risk matches this bound. As to a lower bound, Pinsker (1981) proposed a very nice method based on the Bayesian approach for the general functional class (2) for the model (1) with the Wiener noise process $\xi = (\xi_t)_{0\le t\le 1}$. The idea is to replace the nonparametric model with parametric one of large dimension and to estimate from below the supremum of the risk in (4) by a Bayesian risk with some prior distribution on the ellipse (2). To chose the prior distribution, Pinsker used the key idea of Hájek – Le Cam method based on the normal approximation, i.e. the local asymptotic normality (LAN) property (see, for example, Ibragimov and Khasminskii 1981; Le Cam 1990). Indeed, according to this approach to obtain a sharp lower bound for the maximal risk (4), one has to chose a prior distribution which asymptotically maximises its Bayesian risk. As it is known, such distribution has a Gaussian form. By this reason, the prior distribution is chosen as Gaussian in the main part. The problem is to plunge this prior distribution into the ellipse. To this end, Pinsker proposed to truncate the Gaussian distribution. Unfortunately, without additional conditions on the coefficients of the ellipse $(a_j)_{j\in\mathbb{N}}$, it is impossible to correctly pass to the truncated distribution (see Remark 3 below). To correct the proof of the main theorem in Pinsker (1981), we assume the conditions $\mathbf{A}_1$) and $\mathbf{A}_2$) in Sect. 2. Moreover, for the models with jumps, we cannot use this method; in this case to implement the Bayesian approach, we have to develop a special conditional Baysian risks tool with respect to the jumps of the process $(\xi_t)_{0\le t\le 1}$ in the model (1). It should be noted that similar lower bounds can be obtained through the Van Trees methods (see, for example, in Beltaief et al. 2020), but this method is useful only for the coefficients a_j of a power form. In this paper, first we find the constructive sufficient conditions on the ellipse coefficients $(a_j)_{j\in\mathbb{N}}$ under which we show the efficiency property for the weighted least squares estimates offered in Beltaief et al. (2020) for the non-Gaussian Lévy models (1) and for arbitrary basis functions in $\mathcal{L}_2[0,1]$. Then, we apply the obtained estimators to signals number detection problems in multi-pass connection channels, and finally, for this problem we get almost parametric convergence rate, i.e. $\varepsilon^2|\ln\varepsilon|^\nu$ for $\nu > 1$.

1.4 Organization of the paper

In Sect. 2, we describe the main conditions for the model (1). In Sect. 3, we construct the estimation procedures. In Sect. 4, we announce the main results. In Sect. 5, we consider the efficient estimation problems for the functional class (2) in two cases: power coefficients and exponential coefficients. In Sect. 6, we apply the constructed procedures to signal processing problems. In Sects. 7 and 8, we define the prior distribution and study its main properties. Section 9 gives the proofs of the main theorems. In Appendix, we prove the necessary auxiliary results.

2 Main conditions

First we precise the noise process in (1). Similar to Beltaief et al. (2020), Pchelintsev et al. (2019), we assume that $(\xi_t)_{0\le t\le 1}$ is the Lévy process defined as

$$\xi_t = W_t + z_t \quad \text{and} \quad z_t = x\,\mathbf{1}_{\{|x|\le\varrho\}} * (\mu - \widetilde{\mu})_t\,, \tag{5}$$

where $(W_t)_{0\le t\le 1}$ is a standard Brownian motion, "*" denotes the stochastic integral with respect to the compensated jump measure (see, for example, in Liptser and Shiryaev (1989) for details), $\mu(\mathrm{d}s\,\mathrm{d}x)$ is a jump measure with deterministic compensator $\widetilde{\mu}(\mathrm{d}s\,\mathrm{d}x) = \mathrm{d}s\,\Pi(\mathrm{d}x)$, $\Pi(\cdot)$ is the unknown Lévy measure, i.e. some non-negative measure on $\mathbb{R}_* = \mathbb{R}\setminus\{0\}$ for which we assume that

$$\Pi(x^2) < \infty\,, \tag{6}$$

where $\Pi(g(x)) = \int_{\mathbb{R}_*} g(z)\,\Pi(\mathrm{d}z)$. As to the threshold ϱ, we assume that it is a function of ε, i.e. $\varrho = \varrho_\varepsilon$ such that $\varrho \to 0$ as $\varepsilon \to 0$.

Remark 1 The assumption that in the model (1), the jumps are small, is not restriction, since we can remove the large jumps through the transformation proposed in Beltaief et al. (2020), i.e. we replace the observation model with

$$\widetilde{y}_t = y_t - \sum_{0 \le s \le t} \Delta y_s \, \mathbf{1}_{\{|\Delta y_s| \ge \varrho\}} \,.$$

The sum in this transformation is finite for every $\varrho > 0$, since this is cádlág process. Therefore, in the practice implementation when we have only the discrete observations (y_{t_j}) this transformation is defined through the replacing the jumps with the increments $y_{t_j} - y_{t_{j-1}}$.

Now, we need to precise that we consider the process (1) on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where Ω is the trajectory space, i.e. the Skorokhod space generated by the $[0, 1] \to \mathbb{R}$ functions, $\mathcal{F}$ is the cylindric field and $\mathbf{P}$ is the distribution of the noise process (5). Moreover, we denote by $\mathbf{P}_S$ the distribution of the process (1) corresponding to the unknown function S.

To study an asymptotical estimation accuracy, we need the following condition for the coefficients $(a_j)_{j \ge 1}$ in (2).

$\mathbf{A}_1$) *The sequence* $(a_j)_{j \geqslant 1}$ *is non-decreasing, i.e.* $a_j \geqslant a_{j-1}$, *and* $a_j \to \infty$ *as* $j \to \infty$.

Moreover, for $h > 0$ we set

$$N_h = \max \left\{ j \geqslant 1 \,:\, a_j \leqslant h \right\} . \tag{7}$$

Additionally, we will suppose that the coefficients $(a_j)_{j \geqslant 1}$ satisfy the following conditions.

$\mathbf{A}_2$) *There exists* $0 < \delta_0 < 1$ *such that for any* $0 < \delta \leqslant \delta_0$,

$$N_h - N_{\delta h} - \frac{1}{\sqrt{a_{N_h}}} \sum_{j=N_{\delta h}+1}^{N_h} \sqrt{a_j} \to \infty \,, \qquad \text{as} \qquad h \to \infty \,.$$

3 Estimation procedures

First, we fix some orthonormal basis $(\varphi_j)_{j \geqslant 1}$ in $\mathcal{L}_2[0, 1]$ with $\varphi_1 \equiv 1$. The main idea of a nonparametric estimation in $\mathcal{L}_2[0, 1]$ is to use the Fourier representation for the unknown function S, i.e.

$$S(t) = \sum_{j=1}^{\infty} \theta_j \varphi_j(t) \,, \tag{8}$$

where the Fourier coefficients $(\theta_j)_{j \geqslant 1}$ defined as

$$\theta_j = (S, \varphi_j) = \int_0^1 S(t) \varphi_j(t) \mathrm{d}t \,. \tag{9}$$

In this problem, there are two questions: how to estimate the coefficients $(\theta_j)_{j\geqslant 1}$ and how much terms one needs to use in the sum (8). To estimate the Fourier coefficients in (9), we use least squares estimates

$$\widehat{\theta}_j = \int_0^1 \varphi_j(t)\mathrm{d}y_t \tag{10}$$

which in view of (1) can be represented as

$$\widehat{\theta}_j = \theta_j + \varepsilon\xi_j \quad \text{and} \quad \xi_j = \int_0^1 \varphi_j(t)\mathrm{d}\xi_t\,. \tag{11}$$

As it is shown in Pinsker (1981) efficient estimates for the unknown function S have the following form

$$\widehat{S}_\gamma = \sum_{j=1}^{\infty} \gamma_j \widehat{\theta}_j \varphi_j\,, \tag{12}$$

where $\gamma = (\gamma_j)_{j\geqslant 1}$ is a weight sequence, $0 \leqslant \gamma_j \leqslant 1$, and beginning with some finite number all components are equal to zero. Now, we need to find the optimal estimator among all estimates of the form (12), i.e. we need to find $\gamma = (\gamma_j)_{j\geqslant 1}$ which minimize the risk (3). To this end, we set

$$\mathbf{D}_{*,\varepsilon} = \inf_{\gamma} \sup_{S\in\Theta} \mathcal{R}(\widehat{S}_\gamma, S)\,, \tag{13}$$

and, define the optimal number of nonzero terms in the sum (12)

$$n_* = n_*(\varepsilon, \mathbf{r}) = \max\left\{ l \geqslant 1\,:\, \sqrt{a_l}\sum_{j=1}^{l}\sqrt{a_j} - \sum_{j=1}^{l} a_j \leqslant \varepsilon^{-2}\mathbf{r} \right\}. \tag{14}$$

From the condition $\mathbf{A}_1$) it follows that $n_* < \infty$. So, for $1 \leq j \leq n_*$, we put

$$\theta_j^* = \varepsilon\left(\frac{\mu}{\sqrt{a_j}} - 1\right)^{1/2} \quad \text{with} \quad \mu = \frac{\varepsilon^{-2}\mathbf{r} + \sum_{j=1}^{n_*} a_j}{\sum_{j=1}^{n_*}\sqrt{a_j}}\,. \tag{15}$$

Using the Lagrange optimization method, one can show the following result.

Lemma 1 (Kuks and Olman 1971; Pinsker 1981) *Under the condition* $\mathbf{A}_1$)

$$\mathbf{D}_{*,\varepsilon} = \varepsilon^2 \sum_{j=1}^{n_*} \frac{(\theta_j^*)^2}{(\theta_j^*)^2 + \varepsilon^2} = n_*\varepsilon^2 - \frac{\varepsilon^2\left(\sum_{j=1}^{n_*}\sqrt{a_j}\right)^2}{\varepsilon^{-2}\mathbf{r} + \sum_{j=1}^{n_*} a_j} \tag{16}$$

and

$$n_* = N_{\mu^2}\,, \tag{17}$$

where n_*, μ *and* N_{μ^2} *are given in* (14), (15) *and* (7), *respectively.*

Remark 2 As already it is mentioned before, this lemma is shown through the Lagrange method. The coefficient μ is the optimal Lagrange coefficient for the condition optimization problem (13) for which the optimal function S belongs to functional class Θ. Note also that the right side of the equality (16) is the well-known Pinsker constant.

4 Main results

Now, to study the efficiency properties, we remind that Ξ_ε is the set of all possible estimators $\widetilde{S}$ for S, i.e. any $\sigma\{y_t\,,\, 0 \le t \le 1\}$ measurable functions.

Theorem 1 *Assume that for the model* (1) *the conditions* $\mathbf{A}_1$)–$\mathbf{A}_2$) *hold. Then,*

$$\underline{\lim}_{\varepsilon\to 0} \frac{\inf_{\widetilde{S}\in\Xi_\varepsilon} \sup_{S\in\Theta} \mathcal{R}(\widetilde{S},S)}{\mathbf{D}_{*,\varepsilon}} \ge 1\,, \tag{18}$$

where $\mathbf{D}_{*,\varepsilon}$ *is given in* (16).

Remark 3 In Pinsker (1981), author tried to show this theorem for general functional class (2). Unfortunately, without the condition $\mathbf{A}_2$), the proof proposed by Pinsker is not correct, (see Remark 10 below). May be one can show this theorem by the another way, using the Van Trees inequality (see, for example, Pchelintsev et al. 2019), but methodologically the Pinsker idea is very nice and we would like keep this proof in the general efficient statistical theory.

Further, we show that the lower bound obtained in Theorem 1 is sharp, i.e. there exists an estimator for which the inequality (18) is matched in the equality. We define the Pinsker estimate as

$$\widehat{S}(t) = \sum_{j=1}^{n_*} \gamma_j \widehat{\theta}_j \varphi_j(t) \quad \text{and} \quad \gamma_j = 1 - \frac{\sqrt{a_j}}{\mu}\,, \tag{19}$$

where the coefficient $\mu = \mu_\varepsilon$ is defined in (15) and $0 \le t \le 1$.

Theorem 2 *The risk for the estimate* (19) *is bounded from above as*

$$\overline{\lim}_{\varepsilon\to 0} \frac{\sup_{S\in\Theta} \mathcal{R}(\widehat{S},S)}{\mathbf{D}_{*,\varepsilon}} \leqslant 1\,. \tag{20}$$

Theorems 1 and 2 imply the efficiency property for the weighted least squares estimate (19).

Theorem 3 *Under conditions* $\mathbf{A}_1$) *and* $\mathbf{A}_2$), *the estimate* (19) *is efficient, i.e.*

$$\lim_{\varepsilon\to 0}\frac{\sup_{S\in\Theta}\mathcal{R}(\widehat{S},S)}{\mathbf{D}_{*,\varepsilon}}=1$$

and

$$\lim_{\varepsilon\to 0}\frac{\inf_{\widetilde{S}\in\Xi_\varepsilon}\sup_{S\in\Theta}\mathcal{R}(\widetilde{S},S)}{\sup_{S\in\Theta}\mathcal{R}(\widehat{S},S)}=1\,. \tag{21}$$

Remark 4 We provide the efficiency property (21) in Theorem 3 for the estimate (19) constructed through arbitrary basis $(\varphi_j)_{j\geq 1}$ in $\mathcal{L}_2[0,1]$.

5 Examples

In this section, we consider two main examples of the functional class (2) for which the conditions $\mathbf{A}_1$) and $\mathbf{A}_2$) hold true.

5.1 Power coefficients

First, we consider the functional set (2) with the coefficients

$$a_j=\mathbf{a}_* j^\alpha \tag{22}$$

with fixed constants $\mathbf{a}_*>0$ and $\alpha>0$.

Remark 5 If in (22) the coefficients $a_j=\pi^{2k}j^{2k}$ for some fixed integer $k\geq 1$, then the ellipse (2) defined through the trigonometric basis in $\mathcal{L}_2[0,1]$ coincides with the Sobolev class $W_{k,\mathbf{r}}$, i.e. the set of k times differentiable functions in $\mathcal{L}_2[0,1]$ with bounded in the norm of k-th derivative by $\mathbf{r}$. This is the usual setting for the efficient nonparametric estimation in $\mathcal{L}_2[0,1]$ (see, for example, in Tsybakov 2009 p.137).

We define

$$\widehat{S}(t)=\sum_{j=1}^{n_*}\gamma_j\widehat{\theta}_j\varphi_j(t)\,,\qquad \gamma_j=1-\left(\frac{j}{n_*}\right)^{\alpha/2} \tag{23}$$

and

$$n_*=\max\left\{l\geqslant 1\,:\,l^{\alpha/2}\sum_{j=1}^{l}j^{\alpha/2}-\sum_{j=1}^{l}j^\alpha\leqslant\varepsilon^{-2}\,\mathbf{r}/\mathbf{a}_*\right\}.$$

Theorem 4 *Let the coefficients in the class (2) are defined as in (22). Then, the estimator (23) is efficient, i.e. it satisfies the property (21) and*

$$\lim_{\varepsilon\to 0} \varepsilon^{-\frac{2\alpha}{\alpha+1}} \sup_{S\in\Theta} \mathcal{R}(\widehat{S}, S) = \mathbf{l}_*(\mathbf{r})\,, \tag{24}$$

where

$$\mathbf{l}_*(\mathbf{r}) = \left(\frac{(1+\alpha)\mathbf{r}}{\mathbf{a}_*}\right)^{\frac{1}{\alpha+1}} \left(\frac{\alpha}{\alpha+2}\right)^{\frac{\alpha}{\alpha+1}}.$$

Remark 6 For the "Gaussian white noise" model (1), i.e. when $\xi_t = W_t$, the proof of Theorem 4 is very clearly written in Tsybakov (2009) (Theorem 3.1, p.138).

5.2 Exponential coefficients

Now, we consider the functional set (2) with the coefficients

$$a_j = e^{2\kappa j^\alpha} \tag{25}$$

with fixed constants $0 < \alpha < 1$ and $\kappa > 0$. This coefficients will be used for the signals processing problems in Sect. 6. In this case, we put

$$\widehat{S}(t) = \sum_{j=1}^{n_*} \gamma_j \widehat{\theta}_j \varphi_j(t)\,, \quad \gamma_j = 1 - e^{-\kappa(n_*^\alpha - j^\alpha)} \tag{26}$$

and

$$n_* = \max\left\{ l \geqslant 1 \,:\, e^{\kappa l^\alpha} \sum_{j=1}^{l} e^{\kappa j^\alpha} - \sum_{j=1}^{l} e^{2\kappa j^\alpha} \leqslant \varepsilon^{-2}\,\mathbf{r} \right\}.$$

Theorem 5 *Let the coefficients in the class* (2) *are defined as in* (25). *Then, the estimator* (26) *is efficient, i.e. it satisfies the property* (21) *and*

$$\lim_{\varepsilon\to 0} \varepsilon^{-2} |\ln \varepsilon|^{-\nu} \sup_{S\in\Theta} \mathcal{R}(\widehat{S}, S) = \kappa^{-\nu} \tag{27}$$

with $\nu = 1/\alpha$.

Remark 7 First of all note that the efficiency property (27) is the new result even for the "Gaussian white noise" model. We recall that for the model (1), the optimal convergence rate for parametric problems is ε^2, here we obtained $\varepsilon^2|\ln\varepsilon|^\nu$, i.e. almost parametric convergence rate up to the logarithmic coefficient. The same effect was found in Pinsker (1981) (the example 2). However, the given proof in this paper for this example is not correct.

Remark 8 In Theorems 4 and 5, we use the efficient estimators in simplified asymptotic forms (23) and (26) obtained through the representations (7) and (17) by replacing in (19) the parameter $\mu \approx \sqrt{a_{n_*}}$.

Remark 9 We obtained the efficiency properties (24) and (27) using the $\alpha > 0$, $\kappa > 0$ and the radius $\mathbf{r} > 0$ in (23) and (26). To provide the same efficiency properties without using these parameters, one needs to use the adaptive model selection procedures developed for such model in Konev and Pergamenshchikov (2009a), Konev and Pergamenshchikov (2009b).

6 Signals detection in multi-path channels

Now, we apply the estimation procedure (25) to the signals number estimation problem in the multi-path connection channels considered in Beltaief et al. (2020). This problem is to estimate the summarized signal in the multi-path channel observed on the time interval [0, 1]:

$$\mathrm{d}y_t = \left(\sum_{j=1}^{q} \theta_j \varphi_j(t) \right) \mathrm{d}t + \varepsilon \mathrm{d}\xi_t \,, \tag{28}$$

where the energetic parameters $(\theta_j)_{j\geq 1}$, and the number of signals q are unknown, and the $(\varphi_j)_{j\geq 1}$ are known orthonormal signals, i.e.

$$\int_0^1 \varphi_i(t)\, \varphi_j(t)\, \mathrm{d}t = \mathbf{1}_{\{i \neq j\}} \,.$$

One needs to estimate q when ratio "signal/noise" goes to infinity, i.e. $\varepsilon \to 0$. If in the model (28) the noise is the Wiener process, i.e. $\xi_t = W_t$, then the logarithm of the likelihood ratio can be represented as

$$\ln L_\varepsilon = \frac{1}{\varepsilon^2} \sum_{j=1}^{q} \left(\theta_j \int_0^1 \varphi_j(t) \mathrm{d}y_t - \frac{1}{2} \sum_{j=1}^{q} \theta_j^2 \right)$$

and

$$\max_{\theta_j} \ln L_\varepsilon = \frac{1}{2\varepsilon^2} \sum_{j=1}^{q} \left(\int_0^1 \varphi_j(t) \mathrm{d}y_t \right)^2 .$$

Therefore, maximum over q gives us the trivial solution and cannot be used as an estimator for the number q, i.e. the parametric estimation approach does not work in this case. By these reasons, we propose to study the estimation problem for q for the process (28) in a nonparametric setting (1). Moreover, we consider this problem with non-Gaussian Lévy noises.

It is clear, that for any $1 \leq q < \infty$ the function $S = \sum_{j=1}^{q} \theta_j \varphi_j(t)$ belongs to the functional class (2) with the coefficients $a_j = e^{2\kappa j^\alpha}$ for any $\kappa > 0$ and $0 < \alpha < 1$. Therefore, using Theorem 5, we obtain that the estimator (26) has the almost parametric convergence rate, i.e. $\varepsilon^2 |\ln \varepsilon|^\nu$ for $\nu > 1$ which can be close to 1. Thus, we improve the detection quality of the signals number q with respect to the method proposed in Beltaief et al. (2020) based on the Sobolev functional class with the power coefficients.

7 Prior distribution

In this section, we introduce the prior distribution proposed in Pinsker (1981). For some fixed $q > 1$ we set

$$n_0 = N_{\delta h} \quad \text{and} \quad n_1 = N_h \,, \tag{29}$$

with $h = \mu^2 q^2/(1+q)^2$ and $\delta = q^{-2}$. Since $\mu \to \infty$ as $\varepsilon \to 0$, we obtain through the condition $\mathbf{A}_2$) and the equality (17), that

$$d = n_1 - n_0 \to \infty \quad \text{and} \quad d_* = n_* - n_1 \to \infty \tag{30}$$

as $\varepsilon \to 0$. From here and (15) one has

$$\min_{1\leqslant j\leqslant n_0} (\theta^*_j)^2 \geqslant \varepsilon^2 q \,, \qquad \max_{n_1+1\leqslant j\leqslant n_*} (\theta^*_j)^2 < q^{-1}\varepsilon^2$$

and

$$q^{-1}\varepsilon^2 \le \min_{n_0+1\leqslant j\leqslant n_1} (\theta^*_j)^2 \le \max_{n_0+1\leqslant j\leqslant n_1} (\theta^*_j)^2 < q\varepsilon^2 \,. \tag{31}$$

Also, we can deduce that

$$q^{-2} < \frac{a_j}{a_k} \leqslant q^2 \quad \text{for} \quad n_0 < j, k \le n_1 \,. \tag{32}$$

Now, on the space $(\mathbb{R}^{n_0}, \mathcal{B}(\mathbb{R}^{n_0}))$, we denote by Q_1 the distribution of the n_0 independent random variables uniformly distributed on the intervals $\left[-\mathbf{s}_j\,, \mathbf{s}_j\right]$ with $\mathbf{s}_j = (\theta^*_j)^2$, i.e. for any $\Gamma \in \mathcal{B}(\mathbb{R}^{n_0})$

$$Q_1(\Gamma) = \int_\Gamma \prod_{j=1}^{n_0} \frac{1}{2\mathbf{s}_j} \mathbf{1}_{[-\mathbf{s}_j\,,\,\mathbf{s}_j]}(x_j) \mathrm{d}x_1 \ldots \mathrm{d}x_{n_0} \,. \tag{33}$$

Moreover, on the space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, we denote by Q_2 the distribution of the d independent $(0, \mathbf{s}^*_j)$ Gaussian variables, with $\mathbf{s}^*_j = \psi \mathbf{s}_j$ and $0 < \psi < 1$, i.e.

$$Q_2(\Gamma) = \int_\Gamma \prod_{j=1}^{d} \frac{e^{-\frac{x_j^2}{2\mathbf{s}^*_j}}}{\sqrt{2\pi \mathbf{s}^*_j}} \mathrm{d}x_1 \ldots \mathrm{d}x_d \,, \quad \Gamma \in \mathcal{B}(\mathbb{R}^d) \,. \tag{34}$$

Finally, on the space $(\mathbb{R}^{d_*}, \mathcal{B}(\mathbb{R}^{d_*}))$, we denote by Q_3 distribution of d_* independent random variables with values in the set $\{-\mathbf{s}_j\,,\, \mathbf{s}_j\}$, i.e. for $\Gamma \in \mathcal{B}(\mathbb{R}^{d_*})$

$$Q_3(\Gamma) = \frac{1}{2^{d_*}} \sum_{x \in \Lambda_*} \mathbf{1}_{\{x \in \Gamma\}}\,, \tag{35}$$

where $\Lambda_* = \{-\mathbf{s}_1\,,\,\mathbf{s}_1\} \times \cdots \times \{-\mathbf{s}_{d_*}\,,\,\mathbf{s}_{d_*}\}$. Using these distributions, we introduce the probabilities Q^* and Q on the space $(\mathbb{R}^{n_*}, \mathcal{B}(\mathbb{R}^{n_*}))$ such that for any $\Gamma_1 \in \mathcal{B}(\mathbb{R}^{n_0})$, $\Gamma_2 \in \mathcal{B}(\mathbb{R}^d)$ and $\Gamma_3 \in \mathcal{B}(\mathbb{R}^{d_*})$

$$Q^*(\Gamma_1 \times \Gamma_2 \times \Gamma_3) = Q_1(\Gamma_1) Q_2(\Gamma_2) Q_3(\Gamma_3) \tag{36}$$

and

$$Q(\Gamma_1 \times \Gamma_2 \times \Gamma_3) = Q_1(\Gamma_1) Q_2(\Gamma_2 | B) Q_3(\Gamma_3)\,, \tag{37}$$

where $Q_2(\Gamma_2|B) = Q_2(\Gamma_2 \cap B)/Q_2(B)$ and

$$B = \left\{ y \in \mathbb{R}^d \,:\, \sum_{j=1}^{d} a_{n_0+j} y_j^2 \leqslant \mathbf{r}_1 \right\}, \quad \mathbf{r}_1 = \sum_{j=n_0+1}^{n_1} a_j (\theta_j^*)^2\,.$$

Now, on the space $(\Omega \times \mathbb{R}^{n_*}, \mathcal{F} \otimes \mathcal{B}(\mathbb{R}^{n_*}))$, we define two measures $\mathbf{P}^*$ and $\widetilde{\mathbf{P}}$ for any $\Gamma_1 \in \mathcal{F}$ and $\Gamma_2 \in \mathcal{B}(\mathbb{R}^{n_*})$ as

$$\mathbf{P}^*(\Gamma_1 \times \Gamma_2) = \int_{\Gamma_2} \mathbf{P}_{S_x}(\Gamma_1 | \mathcal{Z})\, Q^*(\mathrm{d}x) \tag{38}$$

and

$$\widetilde{\mathbf{P}}(\Gamma_1 \times \Gamma_2) = \int_{\Gamma_2} \mathbf{P}_{S_x}(\Gamma_1 | \mathcal{Z})\, Q(\mathrm{d}x)\,, \tag{39}$$

where $S_x = \sum_{j=1}^{n_*} x_j \varphi_j$ and $\mathcal{Z}$ is the σ-field generated by the jump component in noise process (5), i.e $\mathcal{Z} = \sigma\{z_t\,,\, 0 \le t \le 1\}$. One can check directly that for any $\Gamma \in \mathcal{F} \otimes \mathcal{B}(\mathbb{R}^{n_*})$

$$\widetilde{\mathbf{P}}(\Gamma) = \mathbf{P}^*(\Gamma | B^*) \quad \text{and} \quad B^* = \Omega \times \mathbb{R}^{n_0} \times B \times \mathbb{R}^{d_*}\,. \tag{40}$$

On the space $(\Omega \times \mathbb{R}^{n_*}, \mathcal{F} \otimes \mathcal{B}(\mathbb{R}^{n_*}))$, we set

$$\eta_j = \zeta_j + \varepsilon \xi_j \quad \text{for} \quad 1 \leqslant j \leqslant n_*\,, \tag{41}$$

where $(\zeta_j)_{1 \le j \le n_*}$ is coordinate random variables on $\mathbb{R}^{n_*}$, i.e. $\zeta_j(x) = x_j$ for $x = (x_1, \ldots, x_{n_*}) \in \mathbb{R}^{n_*}$. It should be noted that under the probabilities $\mathbf{P}^*$ and $\widetilde{\mathbf{P}}$, the random variables ξ_j are independent Gaussian with the parameters $(\nu_j, 1)$, where $\nu_j = \int_0^1 \varphi_j(s) \mathrm{d}z_s$. In the sequel, we need the following σ - fields

$$\mathcal{G}_1 = \sigma\{\eta_{n_0+1}, \ldots, \eta_{n_1}\} \quad \text{and} \quad \mathcal{G}_2 = \sigma\{\eta_j\,,\, j \ge 1\}\,, \tag{42}$$

where $\eta_j = \varepsilon \xi_j$ for $j > n_*$.

8 Properties of the prior distribution

We set the following normalizing coefficient

$$\mathbf{D}_{1,\varepsilon} = \sum_{j=n_0+1}^{n_1} \frac{(\theta_j^*)^2\varepsilon^2}{(\theta_j^*)^2 + \varepsilon^2} \,. \tag{43}$$

Lemma 2 *Under the condition* $\mathbf{A}_2$), *the following limit equality holds*

$$\lim_{\varepsilon\to 0} \frac{\mathbf{D}_{1,\varepsilon}}{\varepsilon^2} = +\infty \,. \tag{44}$$

Proof First, note that

$$\frac{\mathbf{D}_{1,\varepsilon}}{\varepsilon^2} = \sum_{j=n_0+1}^{n_1} \frac{(\theta_j^*)^2}{(\theta_j^*)^2 + \varepsilon^2} = n_1 - n_0 - \frac{1}{\mu} \sum_{j=n_0+1}^{n_1} \sqrt{a_j} \,, \tag{45}$$

and taking into account $\mu \geqslant \sqrt{a_{n_*}}$, we get

$$\frac{\mathbf{D}_{1,\varepsilon}}{\varepsilon^2} \geqslant n_1 - n_0 - \frac{1}{\sqrt{a_{n_1}}} \sum_{j=n_0}^{n_1} \sqrt{a_j} \,.$$

Using here the condition $\mathbf{A}_2$) and the definitions in (29), we have the property (44). Hence, Lemma 2. □

Proposition 1 *Let the conditions* $\mathbf{A}_1$)–$\mathbf{A}_2$) *hold. Then,*

$$\underline{\lim}_{\varepsilon\to 0} \frac{\sum_{j=n_0+1}^{n_1} \widetilde{\mathbf{E}}\left(\zeta_j - \widetilde{\mathbf{E}}(\zeta_j|\mathcal{G}_1)\right)^2}{\mathbf{D}_{1,\varepsilon}} \geq \psi \,, \tag{46}$$

where $\widetilde{\mathbf{E}}$ *is the expectation with respect to the probability measure* $\widetilde{\mathbf{P}}$.

Proof First, note that in view of the Doob–Dynkin Lemma, there exist $\mathbb{R}^d \to \mathbb{R}$ functions F_j such that

$$\overline{\zeta}_j = \widetilde{\mathbf{E}}(\zeta_j|\mathcal{G}_1) = \widetilde{\mathbf{E}}\,(\zeta_j|\eta^*) = F_j(\eta^*) \,, \tag{47}$$

where the random vector $\eta^* = (\eta_k)_{n_{n_0+1}\le k\le n_1}$ is defined in (41). Note that $\overline{\zeta}_j$ is the optimal $\mathcal{G}_1$-measurable estimate for ζ_j, i.e.

$$\inf_{\varsigma\in\mathcal{L}(\mathcal{G}_1)} \widetilde{\mathbf{E}}(\zeta_j - \varsigma)^2 = \widetilde{\mathbf{E}}(\zeta_j - \overline{\zeta}_j)^2 \,,$$

where $\mathcal{L}(\mathcal{G}_1)$ denotes all $\mathcal{G}_1$-measurable random variables.

By the Jensen inequality and (37) we get

$$\sum_{j=n_0+1}^{n_1} a_j \overline{\zeta}_j^2 \le \widetilde{\mathbf{E}}\left(\sum_{j=n_0+1}^{n_1} a_j \zeta_j^2 | \mathcal{G}_1 \right) \le \mathbf{r}_1 \quad \widetilde{\mathbf{P}} - \text{ a.s.} \tag{48}$$

In view of (34), under the probability $\mathbf{P}^*$, the random variables $(\zeta_j)_{n_0+1\le j\le n_1}$ are independent Gaussian. Therefore,

Lemma 4 implies

$$\sum_{j=n_0+1}^{n_1} \mathbf{E}^*\left(\zeta_j - \mathbf{E}^*(\zeta_j|\eta_j)\right)^2 = \psi \sum_{j=n_0+1}^{n_1} \frac{(\theta_j^*)^2\varepsilon^2}{(\theta_j^*)^2\psi + \varepsilon^2}$$

$$[2mm] \geqslant \psi \sum_{j=n_0+1}^{n_1} \frac{(\theta_j^*)^2\varepsilon^2}{(\theta_j^*)^2 + \varepsilon^2} = \psi \mathbf{D}_{1,\varepsilon}\,.$$

Moreover, taking into account that $\mathbf{E}^*(\zeta_j|\eta_{n_0+1}, \ldots, \eta_{n_1}) = \mathbf{E}^*(\zeta_j|\eta_j)$ and using the properties of the conditional expectations, we obtain

$$\sum_{j=n_0+1}^{n_1} \mathbf{E}^*(\zeta_j - \overline{\zeta}_j)^2 \geqslant \sum_{j=n_0+1}^{n_1} \mathbf{E}^*\left(\zeta_j - \mathbf{E}^*\left(\zeta_j|\eta_j\right)\right)^2 \geqslant \psi \mathbf{D}_{1,\varepsilon}\,. \tag{49}$$

It is clear that

$$\begin{aligned}\sum_{j=n_0+1}^{n_1} \mathbf{E}^*(\zeta_j - \overline{\zeta}_j)^2 =& \mathbf{P}^*(B^*) \sum_{j=n_0+1}^{n_1} \mathbf{E}^*\left((\zeta_j - \overline{\zeta}_j)^2 \mid B^*\right) \\ &+ \mathbf{P}^*\left((B^*)^{\mathbf{c}}\right) \sum_{j=n_0+1}^{n_1} \mathbf{E}^*\left((\zeta_j - \overline{\zeta}_j)^2 \mid (B^*)^{\mathbf{c}}\right),\end{aligned} \tag{50}$$

where the set B^* is defined in (40). For any bounded $\mathbb{R}^{2d} \to \mathbb{R}$ function H

$$\mathbf{E}^*\left(H(\zeta^*, \eta^*) \mid B^*\right) = \mathbf{E}^*\left(\overline{H}(\zeta^*) \mid B^*\right),$$

where $\zeta^* = (\zeta_j)_{n_0+1\le j\le n_1}$, $\overline{H}(x) = \mathbf{E}^*\, H(x, x + \varepsilon\xi^*) = \widetilde{\mathbf{E}}\, H(x, x + \varepsilon\xi^*)$, $x \in \mathbb{R}^d$ and $\xi^* = (\xi_j)_{n_0+1\le j\le n_1}$. So, from the the property (40), it follows that

$$\mathbf{E}^*\left(H(\zeta^*, \eta^*) \mid B^*\right) = \widetilde{\mathbf{E}}\, \overline{H}(\zeta^*) = \widetilde{\mathbf{E}}\, H(\zeta^*, \eta^*)\,.$$

Therefore, using the form (47), we get

$$\mathbf{E}^*\left(\sum_{j=n_0+1}^{n_1}(\zeta_j-\overline{\zeta}_j)^2 \ \mid \ B^*\right)=\sum_{j=n_0+1}^{n_1}\widetilde{\mathbf{E}}(\zeta_j-\overline{\zeta}_j)^2\,.$$

Now, from (49) and (50) one has

$$\sum_{j=n_0+1}^{n_1}\widetilde{\mathbf{E}}(\zeta_j-\overline{\zeta}_j)^2\geq\frac{\psi\mathbf{D}_{1,\varepsilon}}{\mathbf{P}^*(B^*)}-\frac{\mathbf{P}^*((B^*)^c)\sum_{j=n_0+1}^{n_1}\mathbf{E}^*\Big((\zeta_j-\overline{\zeta}_j)^2 \ \mid \ (B^*)^c\Big)}{\mathbf{P}^*(B^*)}\,.$$

Remind, that $B^*=\Omega\times\mathbb{R}^{n_0}\times B\times\mathbb{R}^{d_*}$, $(B^*)^c=\Omega\times\mathbb{R}^{n_0}\times B^c\times\mathbb{R}^{d_*}$ and B is defined in (37). Therefore, $\mathbf{P}^*(B^*)=Q_2(B)$, $\mathbf{P}^*((B^*)^c)=Q_2(B^c)$ and to prove the lower bound (46) it suffices to show

$$\lim_{\varepsilon\to 0}\frac{Q_2(B^c)\sum_{j=n_0+1}^{n_1}\mathbf{E}^*\Big((\zeta_j-\overline{\zeta}_j)^2|(B^*)^c\Big)}{\mathbf{D}_{1,\varepsilon}}=0 \tag{51}$$

and

$$\lim_{\varepsilon\to 0}Q_2(B^c)=0\,. \tag{52}$$

For this first note, that the property (48) yields

$$\sum_{j=n_0+1}^{n_1}\overline{\zeta}_j^2\leqslant a_{n_0}^{-1}\sum_{j=n_0+1}^{n_1}a_j\overline{\zeta}_j^2\leqslant a_{n_0}^{-1}\mathbf{r}_1\,,\quad \widetilde{\mathbf{P}}\,-\,\text{a.s.}$$

Note here that, using the definition (41) one can check directly that the distributions of the vector η^* under the probabilities $\widetilde{\mathbf{P}}$ and $\mathbf{P}^*$ are equivalent. Therefore, taking into account the representation (47) we can write that

$$\sum_{j=n_0+1}^{n_1}\overline{\zeta}_j^2\leq a_{n_0}^{-1}\mathbf{r}_1\,,\quad \mathbf{P}^*\,-\,\text{a.s.} \tag{53}$$

So, if $\sum_{j=n_0+1}^{n_1}a_j\zeta_j^2>\mathbf{r}_1$, then through the inequalities (32) we get

$$\sum_{j=n_0+1}^{n_1}\zeta_j^2\geqslant a_{n_1}^{-1}\sum_{j=n_0+1}^{n_1}a_j\zeta_j^2\geqslant a_{n_1}^{-1}\mathbf{r}_1\geqslant a_{n_0}^{-1}q^{-2}\mathbf{r}_1$$

and using here the bound (53) we can estimate this sum from below as

$$\sum_{j=n_0+1}^{n_1}\zeta_j^2\geq q^{-2}\sum_{j=n_0+1}^{n_1}\overline{\zeta}_j^2\,,\quad \mathbf{P}^*\,-\,\text{a.s.}$$

Therefore,

$$\sum_{j=n_0+1}^{n_1} \mathbf{E}^*\Big((\zeta_j - \overline{\zeta}_j)^2 | (B^*)^c\Big) \le 2(1+q^2) \sum_{j=n_0+1}^{n_1} \mathbf{E}^*\Big(\zeta_j^2 \mid (B^*)^c\Big)$$

and to prove the limit (51) it suffices to check that

$$\lim_{\varepsilon\to 0} \frac{Q_2(B^c) \sum_{j=n_0+1}^{n_1} \mathbf{E}^*\Big(\zeta_j^2 | (B^*)^c\Big)}{\mathbf{D}_{1,\varepsilon}} = 0\,. \tag{54}$$

To this end, note that

$$Q_2(B^c) \sum_{j=n_0+1}^{n_1} \mathbf{E}^*\Big(\zeta_j^2 | (B^*)^c\Big) = \upsilon_\varepsilon Q_2(B^c) - \upsilon_\varepsilon Q_2(B)\, \mathbf{E}^*\big(\varDelta_\varepsilon \mid B^*\big)\,, \tag{55}$$

where $\upsilon_\varepsilon = \mathbf{E}^* \sum_{j=n_0+1}^{n_1} \zeta_j^2$ and $\varDelta_\varepsilon = \sum_{j=n_0+1}^{n_1} \zeta_j^2/\upsilon_\varepsilon - 1$. Now we show the limit equality (52) and

$$\mathbf{P}^* - \lim_{\varepsilon\to 0} \varDelta_\varepsilon = 0\,. \tag{56}$$

By the Chebyshev inequality we obtain

$$Q_2(B^c) \le \frac{\mathbf{E}^*\Big(\sum_{j=n_0+1}^{n_1} a_j\, \check{\zeta}_j\Big)^2}{\Big(\mathbf{r}_1 - \sum_{j=n_0+1}^{n_1} a_j\, \mathbf{E}^*\zeta_j^2\Big)^2}\,,$$

where $\check{\zeta}_j = \zeta_j^2 - \mathbf{E}^*\zeta_j^2$. Similarly, for any $\epsilon > 0$ we get

$$\mathbf{P}^*\big(|\varDelta_\varepsilon| > \epsilon\big) \le \frac{\mathbf{E}^*\Big(\sum_{j=n_0+1}^{n_1} \check{\zeta}_j\Big)^2}{\epsilon^2 \upsilon_\varepsilon^2}\,. \tag{57}$$

Since

$$\mathbf{E}^*\zeta_j^2 = \psi(\theta_j^*)^2 \quad \text{and} \quad \mathbf{E}^*\check{\zeta}_j^2 = 2\psi^2(\theta_j^*)^4\,, \tag{58}$$

then

$$Q_2(B^c) \leqslant \frac{2\psi^2 \sum_{j=n_0+1}^{n_1} a_j^2(\theta_j^*)^4}{(1-\psi)^2\mathbf{r}_1^2} \leqslant \frac{2\psi^2 \max_{n_0<j\le n_1} a_j\,(\theta_j^*)^2}{(1-\psi)^2\mathbf{r}_1}\,.$$

The properties (31) and (32) imply

$$q^{-1} a_{n_0+1}\varepsilon^2 < a_j(\theta_j^*)^2 < q^3 a_{n_0+1}\varepsilon^2\,.$$

Therefore, taking into account that

$$\sum_{j=n_0+1}^{n_1} (\theta_j^*)^2 \geqslant \sum_{j=n_0+1}^{n_1} \frac{(\theta_j^*)^2 \varepsilon^2}{(\theta_j^*)^2 + \varepsilon^2} = \mathbf{D}_{1,\varepsilon}\,,$$

we have

$$\frac{\max_{n_0<j\le n_1}\, a_j\,(\theta_j^*)^2}{\mathbf{r}_1} = \frac{\max_{n_0<j\le n_1}\, a_j\,(\theta_j^*)^2}{\sum_{k=n_0+1}^{n_1} a_k \theta_k^{*2}} \le \frac{\varepsilon^2 q^3}{\sum_{k=n_0+1}^{n_1} \theta_k^{*2}} = \frac{\varepsilon^2 q^3}{\mathbf{D}_{1,\varepsilon}}\,.$$

Lemma 2 implies directly the property (52). Next, from (58) and (57), we find

$$\mathbf{P}^*\left(|\Delta_\varepsilon| > \epsilon\right) \le \frac{2\sum_{j=n_0+1}^{n_1} (\theta_j^*)^4}{\epsilon^2 \left(\sum_{j=n_0+1}^{n_1} (\theta_j^*)^2\right)^2} \le \frac{2\max_{n_0<j\le n_1}\,(\theta_j^*)^2}{\epsilon^2 \left(\sum_{j=n_0+1}^{n_1} (\theta_j^*)^2\right)}\,.$$

Using here the inequalities (31) and again Lemma 2, we obtain

$$\frac{\max_{n_0<j\le n_1} (\theta_j^*)^2}{\sum_{j=n_0+1}^{n_1} (\theta_j^*)^2} \leqslant \frac{q\varepsilon^2}{\sum_{j=n_0+1}^{n_1} (\theta_j^*)^2} \leqslant \frac{q\varepsilon^2}{\mathbf{D}_{1,\varepsilon}} \to 0\,, \tag{59}$$

as $\varepsilon \to 0$. Furthermore, note that on the set B, using the bounds (32), we can estimate Δ_ε from above as

$$|\Delta_\varepsilon| \le 1 + \frac{\sum_{j=n_0+1}^{n_1} \zeta_j^2}{\psi \sum_{j=n_0+1}^{n_1} (\theta_j^*)^2} \le 1 + \frac{\mathbf{r}_1}{\psi\, a_{n_0+1} \sum_{j=n_0+1}^{n_1} (\theta_j^*)^2} \le 1 + \frac{q^2}{\psi}\,.$$

Therefore, the dominated convergence theorem together with the properties (52) and (56) yields

$$\lim_{\varepsilon\to 0} \mathbf{E}^*\left(\Delta_\varepsilon \mid B\right) = 0\,.$$

Moreover, using the inequalities (31), we can estimate the term υ_ε in (55) as

$$\upsilon_\varepsilon = \psi(1+q) \sum_{j=n_0+1}^{n_1} (\theta_j^*)^2 \frac{1}{1+q}$$

$$[2mm] \le \psi \sum_{j=n_0+1}^{n_1} (\theta_j^*)^2 \frac{\varepsilon^2}{\varepsilon^2 + (\theta_j^*)^2} (1+q) = \psi(1+q)\mathbf{D}_{1,\varepsilon}\,.$$

Thus, we come to the limit property (54) which implies the bound (46). □

Remark 10 From the representation (45), it follows that if the condition $\mathbf{A}_2$) doesn't hold (as, for example, for $a_j = e^j$), then

$$\overline{\lim_{\varepsilon\to 0}}\,\frac{\mathbf{D}_{1,\varepsilon}}{\varepsilon^2}<\infty\,,$$

i.e. the upper bound in (59) doesn't go to 0 and, therefore, the lower bound (46) doesn't hold, and as a consequence, Theorem 2 and the example 2 in Pinsker (1981) do not hold true. This means that the condition $\mathbf{A}_2$) is crucial for this proof.

9 Proofs

9.1 Proof of Theorem 1

Lemma 3 implies that the field $\sigma\{y_t\,,\,0\le t\le 1\}$ coincides with the field $\mathcal{G}=\sigma\left\{\widehat{\theta}_j\,,\,j\ge 1\right\}$. Thus, any estimator from Ξ_ε can be represented as

$$\widetilde{S}(t)=\sum_{j=1}^{\infty}\widetilde{\theta}_j\,\varphi_j(t)\,,$$

where the coefficients $\widetilde{\theta}_j=\int_0^1\widetilde{S}(t)\,\varphi_j(t)\,\mathrm{d}t$ are measurable with respect to the σ-field $\mathcal{G}$. This implies that there exist $\mathbb{R}^{\mathbb{N}}\to\mathbb{R}$ functions $\mathbf{t}_j$ such that

$$\widetilde{\theta}_j=\mathbf{t}_j(\widehat{\theta})\quad\text{and}\quad\widehat{\theta}=(\widehat{\theta}_j)_{j\ge 1}\,.$$

We will use the natural extension of the measure Q defined in Sect. 7 on the space $(\mathbb{R}^{\mathbb{N}},\mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ as the distribution of the random variables $\zeta=(\zeta_j)_{j\ge 1}$ with $\zeta_j=0$ for $j\ge n_*$. It is clear that

$$\begin{aligned}\sum_{j=1}^{n_*}\zeta_j^2a_j&=\sum_{j=1}^{n_0}\zeta_j^2a_j+\sum_{j=n_0+1}^{n_1}\zeta_j^2a_j+\sum_{j=n_1+1}^{n_*}\zeta_j^2a_j\\&\le\sum_{j=1}^{n_0}(\theta_j^*)^2a_j+\mathbf{r}_1+\sum_{j=n_1+1}^{n_*}\zeta_j^2a_j=\mathbf{r},\end{aligned}$$

i.e. $Q(\zeta\in\Theta)=1$. Using this property, we can estimate from below the maximum value of the risk (3) as

$$\begin{aligned}\sup_{S\in\Theta}\mathcal{R}(\widetilde{S},S)&=\sup_{S\in\Theta}\sum_{j=1}^{\infty}\mathbf{E}_S(\widetilde{\theta}_j-\theta_j)^2\ge\sum_{j=1}^{n_*}\int_\Theta\mathbf{E}_{S_x}(\widetilde{\theta}_j-\theta_j)^2\,Q(\mathrm{d}x)\\&=\mathbf{E}\sum_{j=1}^{n_*}\int_\Theta\mathbf{E}_{S_x}\left((\widetilde{\theta}_j-\theta_j)^2|\mathcal{Z}\right)Q(\mathrm{d}x)\\&=\mathbf{E}\sum_{j=1}^{n_*}\widetilde{\mathbf{E}}(\mathbf{t}_j-\zeta_j)^2\,,\end{aligned}\tag{60}$$

where the random variables $\mathbf{t}_j$ are measurable with respect to the σ-field $\mathcal{G}_2$ defined in (42). Note that $\widetilde{\mathbf{E}}(\mathbf{t}_j - \zeta_j)^2 \geq \widetilde{\mathbf{E}}\left(\widetilde{\mathbf{E}}(\zeta_j|\mathcal{G}_2) - \zeta_j\right)^2$. To clarify the lower bound in (60), we show that

$$\underline{\lim}_{\varepsilon\to 0} \frac{\sum_{j=1}^{n_*} \widetilde{\mathbf{E}}(\zeta_j - \widetilde{\mathbf{E}}(\zeta_j|\mathcal{G}_2))^2}{\mathbf{D}_{*,\varepsilon}} \geqslant 1\,. \tag{61}$$

In view of Lemmas 5 and 6, for any $\mathbf{g} > 0$ there exists q_0 such that for any $q \geq q_0$

$$\inf_{1\leq j\leq n_0} \widetilde{\mathbf{E}}\left(\zeta_j - \widetilde{\mathbf{E}}(\zeta_j|\eta_j)\right)^2 \geq (1-\mathbf{g})\varepsilon^2$$

and

$$\inf_{n_0+1\leq j\leq n_1} \widetilde{\mathbf{E}}\left(\zeta_j - \widetilde{\mathbf{E}}(\zeta_j|\eta_j)\right)^2 \geq (1-\mathbf{g})(\theta_j^*)^2\,.$$

By the definition of the prior distribution in Sect. 7, the conditional expectation $\widetilde{\mathbf{E}}(\zeta_j|\mathcal{G}_2) = \widetilde{\mathbf{E}}(\zeta_j|\eta_j)$ for any $1 \leq j \leq n_0$ and $n_0 + 1 \leq j \leq n_1$. Then, choosing here $\mathbf{g} = 1 - \psi^2$ with the parameter ψ defined in (37), we get

$$\sum_{j=1}^{n_0} \widetilde{\mathbf{E}}(\zeta_j - \mathbf{E}(\zeta_j|\mathcal{G}_2))^2 + \sum_{j=n_1+1}^{n_*} \widetilde{\mathbf{E}}(\zeta_j - \widetilde{\mathbf{E}}(\zeta_j|\mathcal{G}_2))^2 \geq \psi^2\mathbf{D}_{2,\varepsilon}\,,$$

where

$$\mathbf{D}_{2,\varepsilon} = \varepsilon^2 \sum_{j=1}^{n_0} \frac{(\theta_j^*)^2}{(\theta_j^*)^2 + \varepsilon^2} + \varepsilon^2 \sum_{j=n_1+1}^{n_*} \frac{(\theta_j^*)^2}{(\theta_j^*)^2 + \varepsilon^2}\,.$$

Moreover, taking into account that for $n_0 + 1 \leq j \leq n_1$, the conditional expectations $\widetilde{\mathbf{E}}(\zeta_j|\mathcal{G}_2) = \widetilde{\mathbf{E}}(\zeta_j|\mathcal{G}_1)$, we obtain through Proposition 1, that for sufficiently small ε

$$\sum_{j=n_0+1}^{n_1} \widetilde{\mathbf{E}}(\zeta_j - \widetilde{\mathbf{E}}(\zeta_j|\mathcal{G}_2))^2 \geqslant \psi^2\mathbf{D}_{1,\varepsilon}$$

and, therefore,

$$\sum_{j=1}^{n_*} \widetilde{\mathbf{E}}(\zeta_j - \mathbf{E}(\zeta_j|\mathcal{G}_2))^2 \geq \psi^2(\mathbf{D}_{1,\varepsilon} + \mathbf{D}_{2,\varepsilon}) = \psi^2\mathbf{D}_{*,\varepsilon}\,.$$

Taking here $\lim_{\psi\to 1}\lim_{\varepsilon\to 0}$, we get (61). Then, the using this bound in (60) with the help of the Fatou Lemma implies the lower bound (18). □

9.2 Proof of Theorem 2

For the noise sequence in (11), one can calculate the variance

$$\mathbf{E}\,\xi_j^2 = \sigma^2 = \varepsilon^2(1 + \Pi(x^2\mathbf{1}_{\{|x|\le\varrho\}}))\,.$$

The condition (6) through the uniform integrability property implies

$$\lim_{\varepsilon\to 0}\,\Pi(x^2\mathbf{1}_{\{|x|\le\varrho\}}) = 0\,. \tag{62}$$

Moreover, from (19), we get

$$\mathbf{E}_S\|\widehat{S} - S\|^2 = \sum_{j=1}^{n_*}(1-\gamma_j)^2\theta_j^2 + \sum_{j=n_*+1}^{\infty}\theta_j^2 + \sigma^2\sum_{j=1}^{n_*}\gamma_j^2\,. \tag{63}$$

The using here the condition $\mathbf{A}_1$) and the property (17) yields

$$\sum_{j=1}^{n_*}(1-\gamma_j)^2\theta_j^2 + \sum_{j=n_*+1}^{\infty}\theta_j^2 \leqslant \frac{1}{\mu^2}\sum_{j=1}^{n_*}a_j\theta_j^2 + \frac{1}{a_{n_*+1}}\sum_{j=n_*+1}^{\infty}\theta_j^2 a_j$$
$$\le \mathbf{r}\max\left(\frac{1}{\mu^2}, \frac{1}{a_{n_*+1}}\right) = \frac{\mathbf{r}}{\mu^2}\,.$$

Then, the term (16) can be represented as

$$\mathbf{D}_{*,\varepsilon} = \frac{\mathbf{r}}{\mu^2} + \varepsilon^2\sum_{j=1}^{n_*}\gamma_j^2 = \varepsilon^2\sum_{j=1}^{n_*}\gamma_j\,,$$

and

$$\mathbf{E}_S\|\widehat{S} - S\|^2 \le \mathbf{D}_{*,\varepsilon}\left(1 + \Pi(x^2\mathbf{1}_{\{|x|\le\varrho\}})\right)\,.$$

Therefore, the property (62) implies the upper bound (20). □

9.3 Proof of Theorem 4

It is clear that the condition $\mathbf{A}_1$) holds. To check the condition $\mathbf{A}_2$), we denote

$$Y_h = N_h - N_{\delta h} - \frac{1}{\sqrt{a_{N_h}}}\sum_{j=N_{\delta h}+1}^{N_h}\sqrt{a_j}\,. \tag{64}$$

In this case $N_h = \left[\widetilde{h}^{1/\alpha}\right]$, where $[x]$ denotes the integer part of the number x and $\widetilde{h} = h/\mathbf{a}_*$. Therefore, asymptotically as $h\to\infty$, we obtain

$$Y_h = \left(\mathbf{b}\left(\delta^{\frac{1}{\alpha}}\right) + \mathrm{o}(1)\right)\widetilde{h}^{1/\alpha},$$

where

$$\mathbf{b}(z) = 1 - z - \frac{2 - 2z^{\alpha/2+1}}{2+\alpha} > 0 \quad \text{for} \quad 0 \leqslant z < 1.$$

This implies $\mathbf{A}_2$). Moreover, in view of (14) and (16), we can get that

$$\lim_{\varepsilon \to 0} \varepsilon^{\frac{2}{1+\alpha}} n_* = \mathbf{c}_\alpha \quad \text{and} \quad \lim_{\varepsilon \to 0} \varepsilon^{-\frac{2\alpha}{\alpha+1}} \mathbf{D}_{*,\varepsilon} = \mathbf{l}_*(\mathbf{r}), \tag{65}$$

where

$$\mathbf{c}_\alpha = \left(\frac{\mathbf{r}(\alpha+1)(\alpha+2)}{\mathbf{a}_* \alpha}\right)^{\frac{1}{\alpha+1}}$$

and $\mathbf{l}_*(\mathbf{r})$ is given in (24). Now, in view of (63) we obtain

$$\mathbf{E}_S \|\widehat{S} - S\|^2 \leq \frac{\mathbf{r}}{a_* n_*^\alpha} + \sigma^2 \sum_{j=1}^{n_*} \left(1 - \left(\frac{j}{n_*}\right)^{\alpha/2}\right)^2.$$

The properties (62) and (65) imply the equality (24), and through the lower bound (18), we come to Theorem 4. □

9.4 Proof of Theorem 5

It is clear that the condition $\mathbf{A}_1$) holds. To check the condition $\mathbf{A}_2$), we note that, in this case,

$$N_h = [\kappa_0 (\ln h)^\nu], \quad \nu = \alpha^{-1} \quad \text{and} \quad \kappa_0 = (2\kappa)^{-\nu}.$$

For any $0 < \delta < 1$, the term (64) can be represented as

$$Y_h = d\big(1 - T_h\big), \tag{66}$$

where $d = N_h - N_{\delta h}$ and

$$T_h = \frac{1}{d} \sum_{j=N_{\delta h}+1}^{N_h} \exp\left\{-\kappa(N_h^\alpha - j^\alpha)\right\}.$$

Taking into account that $\nu > 1$, we can obtain that

$$\lim_{h\to\infty} \frac{d}{(\ln h)^{\nu-1}} = \nu\kappa_0\delta_1 \quad \text{and} \quad \delta_1 = -\ln\delta. \tag{67}$$

One has $d/N_h \to 0$ as $h \to \infty$. Moreover, asymptotically as $h \to \infty$

$$T_h = \frac{1 + \mathrm{o}(1)}{d} \sum_{l=1}^{d} \exp\left\{ -\kappa N_h^{\alpha} g\left(\frac{l}{N_h}\right) \right\},$$

where $g(x) = 1 - (1-x)^{\alpha}$. For $0 < x \le d/N_h$ this function can be written as

$$g(x) = \frac{\alpha}{(1 - \vartheta x)^{1-\alpha}} x \quad \text{and} \quad 0 \le \vartheta \le 1\,. \tag{68}$$

Using here that $g(x) \ge \alpha x$ for $0 \le x \le 1$, we get

$$T_h \le \frac{1}{d} \sum_{l=1}^{d} \exp\left\{ -\alpha\kappa \frac{l}{N_h^{1-\alpha}} \right\}.$$

From (67) we find directly that

$$\lim_{h\to\infty} \frac{d}{N_h^{1-\alpha}} = \frac{\nu\delta_1}{2\kappa}\,.$$

Therefore, we have for any $0 < \delta < 1$

$$\overline{\lim_{h\to\infty}}\, T_h \le \lim_{d\to\infty} \frac{1}{d} \sum_{l=1}^{d} \exp\left\{ -\frac{\delta_1 l}{4d} \right\} = \int_0^1 e^{-\frac{\delta_1}{4}z} \mathrm{d}z < 1\,.$$

From here and (66) we obtain the condition $\mathbf{A}_2$). Now, we study the asymptotic properties of n_*. To this end, one notes that from (14) we find that

$$a_{n_*}\rho_{n_*} \le \varepsilon^{-2}\mathbf{r} \quad \text{and} \quad a_{n_*+1}\rho_{n_*+1} > \varepsilon^{-2}\mathbf{r}\,, \tag{69}$$

where

$$\rho_n = \sum_{j=1}^{n-1} \sqrt{\frac{a_j}{a_n}} \left(1 - \sqrt{\frac{a_j}{a_n}} \right).$$

It is clear that $\rho_n \le n$, and we can estimate it from below as

$$\rho_n = \sum_{l=1}^{n-1} e^{-\kappa n^{\alpha} g(l/n)} \left(1 - e^{-\kappa n^{\alpha} g(l/n)} \right) \ge \sum_{l=1}^{\mathbf{m}_n} e^{-\kappa n^{\alpha} g(l/n)} \left(1 - e^{-\kappa n^{\alpha} g(l/n)} \right),$$

where $\mathbf{m}_n = [n^{1-\alpha}]$. From the representation (68) we can obtain that

$$\alpha x \le g(x) \le 2\alpha x \quad \text{for} \quad 0 < x < 1 - \frac{1}{2^{1/(1-\alpha)}}\,.$$

Therefore,

$$\rho_n \geq \sum_{l=1}^{\mathbf{m}_n} e^{-2\alpha\kappa x_l}(1 - e^{-\alpha\kappa x_l}) \quad \text{and} \quad x_l = \frac{l}{n^{1-\alpha}},$$

i.e.

$$\underline{\lim}_{n\to+\infty} \frac{\rho_n}{m_n} \geq \int_0^1 e^{-2\alpha\kappa x}(1 - e^{-\alpha\kappa x})\mathrm{d}x > 0\,.$$

This implies, that for sufficiently large n

$$\rho_n > \rho_* \, n^{1-\alpha} \quad \text{and} \quad \rho_* = \frac{1}{2}\int_0^1 e^{-2\alpha\kappa x}(1 - e^{-\alpha\kappa x})\mathrm{d}x\,.$$

Using this in (69), we obtain that

$$\lim_{\varepsilon\to 0} \frac{n_*}{|\ln \varepsilon|^\nu} = \kappa^{-\nu}\,.$$

From (14) and (16) it follows that

$$n_*\varepsilon^2 \geq \mathbf{D}_{*,\varepsilon} = n_*\varepsilon^2 - \frac{\varepsilon^2\left(\sum_{j=1}^{n_*}\sqrt{a_j}\right)^2}{\varepsilon^{-2}\mathbf{r} + \sum_{j=1}^{n_*} a_j} \geq n_*\varepsilon^2 - \varepsilon^2 \sum_{j=1}^{n_*} \sqrt{\frac{a_j}{a_{n_*}}}\,.$$

To provide the property (27) it suffices to show that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \sqrt{\frac{a_j}{a_n}} = 0\,. \tag{70}$$

Indeed, for any $0 < \delta < 1$

$$\sum_{j=1}^{n} \sqrt{\frac{a_j}{a_n}} \leq (1-\delta)n + 1 + \sum_{j=1}^{[\delta n]} \sqrt{\frac{a_j}{a_n}} \leq (1-\delta)n + 1 + ne^{-\kappa(1-\delta)^\alpha n^\alpha}\,.$$

Therefore, for any $0 < \delta < 1$

$$\overline{\lim_{n\to n}} \frac{1}{n}\sum_{j=1}^{n} \sqrt{\frac{a_j}{a_n}} \leq 1 - \delta$$

and making tend $\delta \to 1$, we get the equality (70), which implies

$$\lim_{\varepsilon\to 0} \frac{\mathbf{D}_{*,\varepsilon}}{n_*\varepsilon^2} = 1\,. \tag{71}$$

Moreover, from (63), we have

$$\begin{aligned}\mathbf{E}_S\|\widehat{S}-S\|^2 &\le \frac{\mathbf{r}}{a_{n_*}} + \sigma^2 \sum_{j=1}^{n_*}\left(1 - 2\frac{\sqrt{a_j}}{\sqrt{a_{n_*}}} + \frac{a_j}{a_{n_*}}\right)\\ &= \sigma^2 n_* - \sigma^2\left(2 - \frac{\varepsilon^2\mu}{\sigma^2\sqrt{a_{n_*}}}\right)\sum_{j=1}^{n_*}\frac{\sqrt{a_j}}{\sqrt{a_{n_*}}}\,.\end{aligned}$$

In this case $a_n/a_{n+1} \to 1$ as $n \to \infty$, and in view of the (17), $\mu/\sqrt{a_{n_*}} \to 1$ as $\varepsilon \to 0$. Therefore, taking into account the properties (62), (70) and (71), and through the lower bound (18) we obtain Theorem 5. □

10 Conclusion

In conclusion, we would like to emphasize that in this paper we find the sufficient conditions $\mathbf{A}_1$) and $\mathbf{A}_2$) which provide the efficient property for the least squares estimate (19). Moreover, we calculated the asymptotic sharp lower bound called the Pinsker constant for the quadratic risk for the exponential ellipse coefficients in (2). It should be emphasized here, that for the "signal plus Gaussian white noise" model, Pinsker considered the exponential coefficients also (see, example 2 in Pinsker 1981). Unfortunately, for this example, the condition $\mathbf{A}_2$) doesn't hold true, and therefore, the efficiency property in this case doesn't hold true as well. In this paper, we corrected the exponential form for the ellipse coefficients in (25). For this case, we show in Theorem 5 that the weighted least squares estimate (26) is efficient. Then, using this result, we obtained an improved almost parametric minimax convergence rate for the problem of determining the number of signals in a multi-pass connection channel, which allows us to very quickly provide efficient nonparametric statistical inference in signal processing problems.

Auxiliary results

Proof of Lemma 1

First, note that the mean square error (13) can be represented as

$$\mathbf{D}_{*,\varepsilon} = \inf_{\gamma}\sup_{S\in W^k_{\mathbf{r}}} \mathcal{R}(\widehat{S}_\gamma, S) = \inf_{\gamma}\sup_{\theta\in\Theta}\sum_{j=1}^{\infty}\mathbf{E}_\theta\left(\gamma_j\widehat{\theta}_j - \theta_j\right)^2.$$

Recall that

$$\inf_{\gamma}\sup_{\theta\in\Theta}\sum_{j=1}^{\infty}\mathbf{E}_\theta\left(\gamma_j\widehat{\theta}_j - \theta_j\right)^2 \geqslant \sup_{\theta\in\Theta}\inf_{\gamma}\sum_{j=1}^{\infty}\mathbf{E}_\theta\left(\gamma_j\widehat{\theta}_j - \theta_j\right)^2.$$

In view of (10) and $\mathbf{E}_\theta \xi_j = 0$, one has

$$\mathbf{E}_\theta \left(\gamma_j \widehat{\theta}_j - \theta_j\right)^2 = (1-\gamma_j)^2 \theta_j^2 + \gamma_j \varepsilon^2 \,, \tag{72}$$

and

$$\inf_{\gamma_j} \mathbf{E}_\theta \left(\gamma_j \widehat{\theta}_j - \theta_j\right)^2 = \frac{\theta_j^2 \varepsilon^2}{\theta_j^2 + \varepsilon^2} \quad \text{with} \quad \gamma_j = \frac{\theta_j^2}{\theta_j^2 + \varepsilon^2} \,. \tag{73}$$

Furthermore, using (73) and (72), we can rewrite

$$\mathbf{D}_{*,\varepsilon} \leqslant \sup_{\theta \in \Theta} \sum_{j=1}^{\infty} \frac{\theta_j^2 \varepsilon^2}{\theta_j^2 + \varepsilon^2} \,,$$

and, therefore,

$$\mathbf{D}_{*,\varepsilon} = \sup_{\theta \in \Theta} \sum_{j=1}^{\infty} \frac{\theta_j^2 \varepsilon^2}{\theta_j^2 + \varepsilon^2} \,.$$

Note that γ_j from (73) cannot be used in (12), because they depend of unknown parameters, and the estimate $\widehat{S}_\gamma(t)$ cannot be calculated. By the Lagrange method, we obtain that

$$\sup_{\theta \in \Theta} \sum_{j=1}^{\infty} \frac{\varepsilon^2 \theta_j^2}{\theta_j^2 + \varepsilon^2} = \sum_{j=1}^{\infty} \frac{\varepsilon^2 (\theta_j^*)^2}{(\theta_j^*)^2 + \varepsilon^2} \quad \text{and} \quad (\theta_j^*)^2 = \varepsilon^2 \left(\frac{\mu}{\sqrt{a_j}} - 1 \right)_+ \,,$$

where $(x)_+ = \max(0, x)$ and the Lagrange coefficient μ is the solution of the following equation

$$f(\mu) = \varepsilon^{-2} \mathbf{r} \quad \text{with} \quad f(\mu) = \sum_{j=1}^{\infty} a_j \left(\frac{\mu}{\sqrt{a_j}} - 1 \right)_+ \,.$$

If the condition $\mathbf{A}_1$) holds, then the function $f(\mu)$ is continuous increasing function with $f(0) = 0$ and $\lim_{\mu \to \infty} f(\mu) = \infty$, and we can deduce that this equation has an unique solution

$$\mu = \frac{\varepsilon^{-2} \mathbf{r} + \sum_{j=1}^{\mathbf{m}} a_j}{\sum_{j=1}^{\mathbf{m}} \sqrt{a_j}} \,, \tag{74}$$

where $\mathbf{m} = N_{\mu^2}$ is defined in (7). This implies that

$$\sqrt{a_{\mathbf{m}}} \leqslant \mu \quad \text{and} \quad \sqrt{a_{\mathbf{m}+1}} > \mu \,.$$

Setting $g(n) = \sqrt{a_n}\sum_{j=1}^{n}\sqrt{a_j} - \sum_{j=1}^{n} a_j$ and using the definition (74), we obtain that

$$g(\mathbf{m}) \leqslant \varepsilon^{-2}\mathbf{r} \quad \text{and} \quad \sqrt{a_{\mathbf{m}+1}}\sum_{j=1}^{\mathbf{m}}\sqrt{a_j} - \sum_{j=1}^{\mathbf{m}} a_j = g(\mathbf{m}+1) > \varepsilon^{-2}\mathbf{r}\,.$$

Therefore, in view of the definition (14), we find

$$\mathbf{m} = \max\left\{n \geqslant 1 \,:\, g(n) \leqslant \varepsilon^{-2}\mathbf{r}\right\} = n^*\,.$$

Hence, Lemma 1. □

Representation for the σ-field generated by $(\xi_t)_{0\le t\le 1}$.

Lemma 3 *Let $(\varphi_k)_{k\ge 1}$ be arbitrary orthonormal basis in $\mathcal{L}_2[0,1]$ with $\varphi_1 \equiv 1$. Then,*

$$\sigma\{\xi_t\,,\ 0 \le t \le 1\} = \sigma\{\xi_k\,,\ k \ge 1\}\,,$$

where $\xi_k = \int_0^1 \varphi_k(t)\mathrm{d}\xi_t$.

Proof Let $(\,\mathrm{Tr}_j)_{j\ge 1}$ be the trigonometric basis in $\mathcal{L}_2[0,1]$. Taking into account that any trajectory of the process ξ belongs to $\mathcal{L}_2[0,1]$, we can represent it as

$$\xi_t = \sum_{j=1}^{\infty}\tau_j\ \mathrm{Tr}_j(t) \quad \text{and} \quad \tau_j = \int_0^1 \xi_s\ \mathrm{Tr}_j(s)\mathrm{d}s\,.$$

Using here the Ito formula, we obtain that

$$\tau_j = \xi_1\,\widetilde{\mathrm{Tr}}_j(1) - \int_0^1 \widetilde{\mathrm{Tr}}_j(s)\mathrm{d}\xi_s \quad \text{and} \quad \widetilde{\mathrm{Tr}}_j(s)\mathrm{d}s = \int_0^t \mathrm{Tr}_j(s)\mathrm{d}s\,.$$

Note now that the functions $\widetilde{\mathrm{Tr}}_j$ can be represented as

$$\widetilde{\mathrm{Tr}}_j(s) = \sum_{l=1}^{\infty}\mathbf{k}_{j,l}\varphi_l(s) \quad \text{and} \quad \mathbf{k}_{j,l} = \int_0^1 \widetilde{\mathrm{Tr}}_j(u)\varphi_l(u)\mathrm{d}u\,.$$

In view of $\xi_1 = \int_0^1 \varphi_1(s)\mathrm{d}\xi_s$, we can rewrite the coefficients τ_j as

$$\tau_j = \xi_1\,\widetilde{\mathrm{Tr}}_j(1) - \sum_{l=1}^{\infty}\mathbf{k}_{j,l}\,\xi_l\,.$$

So, the coefficients τ_j are measurable with respect to the σ-field $\sigma\{\xi_k\,,\ k \ge 1\}$, and therefore, the Brownian motion is measurable with respect to this σ-field also, i.e. $\sigma\{\xi_t\,,\ 0 \le t \le 1\} \subseteq \sigma\{\xi_k\,,\ k \ge 1\}$. The inverse inclusion is obvious. Hence, Lemma 3. □

Conditional distribution tool

Lemma 4 *Let ζ and ξ be independent Gaussian random variables with the parameters $(0, \theta^2)$ and (ν, σ^2), respectively, and let $\eta = \zeta + \xi$. Then,*

$$\mathbf{E}(\zeta - \mathbf{E}(\zeta|\eta))^2 = \frac{\theta^2\sigma^2}{\theta^2 + \sigma^2}\,.$$

Proof Note that η is Gaussian random variable with the parameters $(\nu, \theta^2 + \sigma^2)$. By the definition of the conditional expectation

$$\mathbf{E}\,(\zeta|\eta = y) = \int_{\mathbb{R}} x\,\mathbf{p}_{\zeta|\eta}(x|y)\mathrm{d}\,x$$

and $\mathbf{p}_{\zeta|\eta}(x|y)$ is the corresponding conditional distribution density

$$p_{\zeta|\eta}(x|y) = \frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{1}{2\sigma_1^2}(x - \mathbf{m}(y))^2\right),$$

where

$$\sigma_1^2 = \frac{\theta^2\sigma^2}{\theta^2 + \sigma^2} \quad \text{and} \quad \mathbf{m}(y) = \frac{(y - \nu)\theta^2}{\theta^2 + \sigma^2}\,.$$

This implies that

$$\mathbf{E}(\zeta|\eta) = \mathbf{m}(\eta) \quad \text{and} \quad \mathbf{E}(\zeta - \mathbf{E}(\zeta|\eta))^2 = \frac{\theta^2\sigma^2}{\theta^2 + \sigma^2}\,.$$

Hence Lemma 4. □

Lemma 5 *Let ζ and ξ be two independent random variables such that ζ is uniformly distributed on $(-\theta, \theta)$ for some $\theta > 0$ and ξ is Gaussian with the parameters (ν, σ^2), and let $\eta = \zeta + \xi$. Then,*

$$\lim_{L\to\infty}\sup_{\nu\in\mathbb{R}}\left|\frac{\mathbf{E}(\zeta - \mathbf{E}(\zeta|\eta))^2}{\sigma^2} - 1\right| = 0\,,$$

where $L = \theta/\sigma$.

Proof First, note that

$$\mathbf{E}(\zeta - \mathbf{E}(\zeta|\eta))^2 = \mathbf{E}(\xi - \mathbf{E}(\xi|\eta))^2 = \sigma^2 - \mathbf{E}\,\mathbf{m}^2(\eta)\,,$$

where $\mathbf{m}(\eta) = \mathbf{E}(\tilde{\xi}|\eta)$ and $\tilde{\xi} = \xi - \nu$. It is clear that

$$m(z)=\frac{\int_{-\infty}^{+\infty} x\mathbf{p}_{\eta|\tilde{\xi}}(z|x)\mathbf{p}_{\tilde{\xi}}(x)\mathrm{d}x}{\mathbf{p}_{\eta}(z)}=\frac{\int_{-\infty}^{+\infty} x\,\mathbf{1}_{\langle|z-x-\nu|\leqslant\theta\rangle}\mathbf{p}_{\tilde{\xi}}(x)\mathrm{d}x}{2\theta\mathbf{p}_{\eta}(z)}\,,$$

where $\mathbf{p}_{\eta|\tilde{\xi}}(z|x)$, $\mathbf{p}_{\tilde{\xi}}$ and $\mathbf{p}_{\eta}$ are the corresponding distribution densities. Since the random variables ζ and $\tilde{\xi}$ are independent and ζ is uniform on the interval $(-\theta,\theta)$, then

$$\mathbf{p}_{\eta}(z)=\int_{-\infty}^{+\infty}\mathbf{p}_{\zeta}(z-x)\mathbf{p}_{\xi}(x)\mathrm{d}x=\frac{1}{2\theta}\int_{-\infty}^{+\infty}\mathbf{1}_{\langle|z-x-\nu|\leq\theta\rangle}\mathbf{p}_{\tilde{\xi}}(x)\mathrm{d}x\,.$$

Here $\mathbf{p}_{\tilde{\xi}}(x)=\sigma^{-1}\phi(x/\sigma)$, where ϕ is the (0, 1)-Gaussian density. Now, we have

$$m(z)=\sigma\,\frac{\int_{-\infty}^{+\infty} y\mathbf{1}_{\Gamma}(y)\,\phi(y)\mathrm{d}y}{\int_{-\infty}^{+\infty}\mathbf{1}_{\Gamma}(y)\,\phi(y)\mathrm{d}y}\,,$$

where

$$\Gamma=\left\{y\;:\;\left|y-\frac{\tilde{z}}{\sigma}\right|\leq L\right\}\quad\text{and}\quad\tilde{z}=z-\nu\,.$$

For $|\tilde{z}|<(1-\epsilon)\theta$ with $\epsilon=1/\sqrt{L}$, the indicator $\mathbf{1}_{\Gamma}\to 1$ as $L\to\infty$ and, therefore, $m(z)/\sigma\to 0$. Let now $\rho_L=m^2(\eta)/\sigma^2=(\mathbf{E}(\overline{\xi}|\eta))^2$ and $\overline{\xi}=\widetilde{\xi}/\sigma\sim\mathcal{N}(0,1)$. Then,

$$\mathbf{E}\rho_L=\mathbf{E}\rho_L\mathbf{1}_{\langle|\widetilde{\eta}|<(1-\epsilon)\theta\rangle}+\mathbf{E}\rho_L\mathbf{1}_{\langle(1-\epsilon)\theta\leqslant|\widetilde{\eta}|\leqslant(1+\epsilon)\theta\rangle}+\mathbf{E}\rho_L\mathbf{1}_{\langle|\widetilde{\eta}|>(1+\epsilon)\theta\rangle}\,,$$

where $\widetilde{\eta}=\eta-\nu$. By the Jensen inequality $\rho_L^2\leq\mathbf{E}\left(\overline{\xi}^4\mid\eta\right)$, and, therefore,

$$\mathbf{E}\rho_L^2\leq\mathbf{E}\mathbf{E}\left(\overline{\xi}^4\mid\eta\right)=\mathbf{E}\,\overline{\xi}^4=3\,,$$

i.e. $(\rho_L)_{L\geq 1}$ is uniformly integrable. Since the random variables $\rho_L\mathbf{1}_{\langle|\widetilde{\eta}|<(1-\epsilon)\theta\rangle}\to 0$ as $L\to\infty$ almost sure, then $\mathbf{E}\rho_L\mathbf{1}_{\langle|\widetilde{\eta}|<(1-\epsilon)\theta\rangle}\to 0$ as $L\to\infty$. Moreover, taking into account that $\mathbf{p}_{\eta}(z)\leq 1/2\theta$, we get

$$\mathbf{P}\big((1-\epsilon)\theta\leq|\widetilde{\eta}|\leq(1+\epsilon)\theta\big)\leqslant 2\epsilon=\frac{2}{\sqrt{L}}\to 0\quad\text{as}\quad L\to\infty\,.$$

Further, we have

$$\mathbf{P}\big(|\widetilde{\eta}|>(1+\epsilon)\theta\big)\leq\mathbf{P}(|\bar{\xi}|>\sqrt{L})\to 0\quad\text{as}\quad L\to\infty\,.$$

Hence, Lemma 5. □

Lemma 6 *Let ζ and ξ be two independent random variables, such that* $\mathbf{P}(\zeta=-\theta)=\mathbf{P}(\zeta=\theta)=1/2$ *for some* $\theta>0$ *and ξ is Gaussian with the parameters* (ν,σ^2), *and let* $\eta=\zeta+\xi$. *Then*

$$\lim_{L\to 0}\sup_{v\in\mathbb{R}}\left|\frac{\mathbf{E}(\zeta-\mathbf{E}(\zeta|\eta))^2}{\theta^2}-1\right|=0\,, \tag{75}$$

where $L=\theta/\sigma$.

Proof First, note that in this case

$$\mathbf{E}(\zeta|\eta)=\theta\rho_L(\widetilde{\eta})\quad\text{and}\quad\rho_L(x)=\frac{\phi(x-L)-\phi(x+L)}{\phi(x-L)+\phi(x+L)}\,,$$

where $\widetilde{\eta}=(\eta-v)/\sigma$ and ϕ is the (0, 1)-Gaussian density. It is clear that $|\rho_L(x)|\le 1$ and

$$\lim_{L\to 0}\sup_{|x|\le M}|\rho_L(x)|=0\quad\text{for any}\quad M>0\,.$$

Therefore,

$$\left|\frac{\mathbf{E}(\zeta-\mathbf{E}(\zeta|\eta))^2}{\theta^2}-1\right|=\mathbf{E}\rho_L^2(\widetilde{\eta})\le\sup_{|x|\le M}\rho_L^2(x)+\mathbf{P}(|\widetilde{\eta}|>M)\,.$$

Taking into account here that $\mathbf{E}\,\widetilde{\eta}^2\le 2L^2+2$ and passing to the limit as $\lim_{M\to\infty}\lim_{L\to 0}$, we obtain (75). Hence, Lemma 6. □

Acknowledgements The authors are grateful to the anonymous referees and to the AE for careful reading and for helpful comments. The results of Sect. 6 have been obtained under Grant of the President of the Russian Federation no. MK-834.2020.9.

References

Beltaief, S., Chernoyarov, O., Pergamenshchikov, S. M. (2020). Model selection for the robust efficient signal processing observed with small Levy noise. *Annals of the Institute of Statistical Mathematics, 72,* 1205–1235.

Hodara, P., Krell, N., Löcherbach, E. (2018). Non-parametric estimation of the spiking rate in systems of interacting neurons. *Statistical Inference for Stochastic Processes, 21,* 81–111.

Ibragimov, I. A., Khasminskii, R. Z. (1981). *Statistical estimation*: *Asymptotic theory*. New York: Springer.

Kassam, S. A. (1988). *Signal detection in non-Gaussian noise. IX*. New York: Springer.

Konev, V. V., Pergamenshchikov, S. M. (2009a). Nonparametric estimation in a semimartingale regression model. Part 1. Oracle Inequalities. *Vestnik Tomskogo Gosudarstvennogo Universiteta. Matematika i Mekhanika, 3*(7), 23–41.

Konev, V. V., Pergamenshchikov, S. M. (2009b). Nonparametric estimation in a semimartingale regression model. Part 2. Robust asymptotic efficiency. *Vestnik Tomskogo Gosudarstvennogo Universiteta. Matematika i Mekhanika, 4*(8), 31–45.

Konev, V. V., Pergamenshchikov, S. M. (2012). Efficient robust nonparametric in a semimartingale regression model. *Annales de lInstitut Henri Poincare, 48*(4), 1217–1244.

Konev, V. V., Pergamenshchikov, S. M. (2015). Robust model selection for a semimartingale continuous time regression from discrete data. *Stochastic Processes and Their Applications, 125,* 294–326.

Konev, V. V., Pergamenshchikov, S. M., Pchelintsev, E. A. (2014). Estimation of a regression with the pulse type noise from discrete data. *Theory of Probability & its Applications, 58*(3), 442–457.

Kuks, I. A., Olman, V. (1971). A minimax linear estimator of regression coefficients. *Izv. Akad. Nauk Eston. SSR, 20,* 480–482.

Kutoyants, Yu. A. (1984). *Parameter Estimation for Stochastic Processes*. Berlin: Heldeman-Verlag.

Kutoyants, Yu. A. (1994). *Identification of dynamical systems with small noise*. Dordrecht: Kluwer Academic Publishers Group.

Le Cam, L. (1990). *Asymptotic methods in statistical decision theory. Springer series in statistics*. New York: Springer.

Lepski, O. V. (1990). One problem of adaptive estimation in Gaussian white noise. *Theory of Probability & Its Applications, 35,* 459–470.

Lepski, O. V., Spokoiny, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *The Annals of Statistics, 25,* 2512–2546.

Liptser, R. S., Shiryaev, A. N. (1989). *Theory of martingales*. New York: Springer.

Nemirovskii, A. (2000). Topics in non-parametric statistics. *Lecture Notes in Mathematics, 1738,* 85–277.

Pchelintsev, E. (2013). Improved estimation in a non-Gaussian parametric regression. *Statistical Inference for Stochastic Processes, 16*(1), 15–28.

Pchelintsev, E., Pergamenshchikov, S. (2018). Oracle inequalities for the stochastic differential equations. *Statistical Inference for Stochastic Processes, 21*(2), 469–483.

Pchelintsev, E., Pergamenshchikov, S. (2019). Adaptive model selection method for a conditionally Gaussian semimartingale regression in continuous time. *Vestnik Tomskogo Gosudarstvennogo Universiteta. Matematika i Mekhanika, 58,* 14–31.

Pchelintsev, E. A., Pchelintsev, V. A., Pergamenshchikov, S. M. (2018). Non asymptotic sharp oracle inequality for the improved model selection procedures for the adaptive nonparametric signal estimation problem. *Communications—Scientific Letters of the University of Zilina, 20*(1), 72–76.

Pchelintsev, E. A., Pchelintsev, V. A., Pergamenshchikov, S. M. (2019). Improved robust model selection methods for a Lévy nonparametric regression in continuous time. *Journal of Nonparametric Statistics, 31*(3), 612–628.

Pinsker, M. S. (1981). Optimal filtration of square integrable signals in gaussian white noise. *Problems of Transimission Information, 17,* 120–133.

Tsybakov, A. B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Annals of Statistics, 26,* 2420–2469.

Tsybakov, A. B. (2009). *Introduction in nonparametric estimation*. New York: Springer.