Jump penalized L^1-Regression

I n a u g u r a l d i s s e r t a t i o n

zur

Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Ernst-Moritz-Arndt-Universität Greifswald

vorgelegt von
Nadiya Kosytsina
geboren am 24.10.1979
in Dnipropetrovsk, Ukraine

Greifswald, 5.07.2012

Dekan:	Prof. Dr. Klaus Fesser ....
1. Gutachter:	Prof. Dr. Volkmar Liebscher ....
2. Gutachter:	Prof. Dr. Axel Munk..
Tag der Promotion:	29.06.2012..

Contents

Chapter 1

Introduction

1.1 General concepts and definitions

In statistics, the term regression or, equivalently, regression model is used to describe a certain relationship between a scalar response variable Y and a vector-valued input variable $X = (X^{(1)}, \ldots, X^{(d)})$, where $d \geq 1$.
The main objective of regression analysis [18, 29, 39] is to determine such a relationship based on the data, which is available from the experimental observations. However, it should be emphasized that because we are always dealing with a finite number of observation points (X_i, Y_i), $i = 1 \ldots n$, $n < \infty$, the unknown dependence $Y = f(X)$ of the response variable Y on the input variable X is certainly not uniquely determined. In fact, in almost all practical cases the "true" function f does not exist as such, so that the purpose of the regression analysis reduces to determining a function $\widehat{f}$, which satisfies a certain number of criteria. The latter are usually designed to demand that the error between $\widehat{f}(X_i)$ and the experimental data Y_i is minimal and that $\widehat{f}$ is a relatively smooth function.
Inevitably, any experimental data is affected by some sort of randomness or fluctuations, which is usually referred to as noise. Therefore, any outcome of an experimental observation Y_i is itself a random variable.
Given n observation points, a regression model specifying the relationship between X_i and Y_i can be written as follows:

$$Y_i = f(X_i) + \xi_i, \;\; i = 1 \ldots n, \tag{1.1}$$

where ξ_i is a random variable, which determines the observation errors between the actual data Y_i and its prediction given by the unknown regression function f.
Thereby, when the variables X_i are of the non-stochastic nature, then the regression model (1.1) is called *fixed design model* [18, 29]. Usually in the case of fixed design model the variables $X_1, \ldots, X_n$ are taken as equidistant points from a certain interval $[a,\, b] \subset \mathbb{R}$, for instance $[a,\, b] = [0,\, 1]$ and $X_i = i/n$.
Historically, all regression models can be divided into two large classes: the *parametric* and the *non-parametric models* [41]. In case of the parametric model, the regression function f is assumed to belong to a certain class of functions, whose shape is determined by a set of scalar coefficients. For instance, f can be seen as a polynomial of a certain order, whose coefficients are to be determined. On the contrary, the non-parametric regression model does not make any assumptions on the shape of the function f, and, is therefore, widely regarded as a more robust, but also computationally more expensive approach.

The motivation behind a regression analysis may be manifold. First, the knowledge of regression estimator $\widehat{f}$ can be used to make forecasts as to the outcome of the experiment given some new value $\widetilde{X}$ of the input variable, which is not found in the set of currently available data X_i, $i = 1 \dots n$. Second, the function $\widehat{f}$ bears an essential information about the intrinsic dependence of Y on X, unknown a priori. For instance, one could deduce from $\widehat{f}$ whether or not the response Y should grow or shrink as the input X is being changed. Third, the computation of the regression function results in the smoothing of the response data Y [37, 10] and, therefore, can be seen as an effective tool for removing the so-called outliers, i.e. the data points Y_i, which are numerically distant from the rest of the data. The latter occur very often as a result of the measurement error or can be related to the heavy-tailed distribution [2] of the noise, i.e. a distribution whose tail decays slower than a Gaussian.

1.2 Estimation techniques: smoothing

The computation of $\widehat{f}$ from a given data set $(X_1, Y_1), ..., (X_n, Y_n)$, which obeys (1.1), can be regarded as *smoothing*. Indeed, because $f(X_i)$ differs from the output variable Y_i by a margin of the error ξ_i, the regression estimator will, in general, be much smoother than the initial data set (X_i, Y_i) itself.
There are many smoothing techniques available which allow us to compute the estimator $\widehat{f}$ of f. Below we give a brief overview of the most popular smoothing methods used in statistics nowadays. Namely we consider the following methods: *kernel smoothing*, *k-nearest neighbor smoothing*, *least square estimate*, *L^1- regression*, *spline smoothing* and *regressograms and median smoothing*. The detail description of these and other smoothing methods and approximation techniques can be found in numerous monographs, see for example Härdle [18], Schroeder [39] or Mendenhall [29].

Kernel smoothing belongs to the nonparametric regression models. The main idea of the kernel smoothing is to introduce a continuous and symmetric weight function K, called the kernel, which is normalized to one, i.e. $\int_{-\infty}^{\infty} K(u)\, du = 1$. Using the kernel function, the regression estimator $\widehat{f_h}$ is given by the Nadaraya-Watson formula [32]

$$\widehat{f_h}(x) = \frac{\sum_{i=1}^{n} K_h(x - X_i) Y_i}{\sum_{i=1}^{n} K_h(x - X_i)}, \tag{1.2}$$

where $K_h(x) = (1/h)K(x/h)$ is the kernel function scaled by the *bandwidth* h.
The meaning of equation (1.2) is rather simple. The value of the estimator function at any point x is given as the average of all the data points Y_i, each taken with a different weight $K_h(x - X_i)/\sum_{i=1}^{n} K_h(x - X_i)$.
An interesting property of the kernel smoothing is related to the choice of the band width parameter h. Namely this parameter controls the smoothness of the estimator. The larger is h, the smoother is the estimator $\widehat{f_h}$.

k-nearest neighbor smoothing also belongs to the nonparametric regression models. The estimator is constructed locally from k nearest neighbors of the point x. By defining the set of indexes $\mathcal{I}_x^k = \{i : X_i \text{ is one of the } k\text{−nearest neighbors of } x\}$, the estimator $\widehat{f}$ is given by

$$\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} W_i^k(x)\, Y_i, \tag{1.3}$$

where the weights $W_i^k(x) = \dfrac{n}{k}$ if $i \in \mathcal{I}_x^k$ and $W_i^k(x) = 0$ otherwise.

Least square estimates belong to the class of parametric models. This is a widely used technique, which is based on the minimization of the global error between a fit function in points X_i and the data Y_i. The global error in this case is defined as the total squared deviation, namely as follows:

$$S_2\,(g) = \sum_{i=1}^{n} \left(g(X_i) - Y_i\right)^2, \tag{1.4}$$

where g belongs to a certain class of functions, which is defined up to some unknown coefficients. These coefficients are to be determined by minimizing S_2.
Least square estimates can be applied in two different ways: globally and locally. In case of the global estimation, the estimator is computed from the whole set of data points. More frequently, however, the minimization of S_2 is done locally, i.e. for a certain window of the input variable. All local estimators are then joined together to give the estimator for the whole data set. A popular choice in the applied statistical analysis are the linear local estimators [18].

L^1**-regression** differs from least square estimators by the definition of the global error, which in this case is defined as the following total absolute deviation:

$$S_1\,(g) = \sum_{i=1}^{n} \left|g(X_i) - Y_i\right|. \tag{1.5}$$

Therefore in order to obtain the estimator $\widehat{f}$, one minimizes S_1, thereby, similarly to the least square estimates, g belongs to a certain class of functions, which is defined up to some unknown coefficients.

Spline smoothing technically belongs to the nonparametric models, although, as its name suggests, the estimator is ultimately given by, typically, cubic spline. The main idea behind this method is to approximate the data points by a smooth twice differentiable function in such a way as to minimize the least square error (1.4) by simultaneously ensuring that the estimator function $\widehat{f}$ is not too "rough". The measure of roughness is assumed to be given by the integral $\gamma \int (g''(x))^2\, dx$, where $\gamma \geq 0$ is a fixed smoothing parameter and g'' stands for the second derivative of g. As a result, the estimator function is given by the solution of the following minimization problem

$$\widehat{f} = \min_{g} \left[\sum_{i=1}^{n} \left(g(X_i) - Y_i\right)^2 + \gamma \int (g''(x))^2\, dx\right]. \tag{1.6}$$

It is well known that (1.6) has a unique solution $\widehat{f}$, which is given by a cubic spline [38, 35].

Regressograms and median smoothing. We will finish this section by introducing a special type of the smoothing technique, which technically belongs to the class of the local constant regression models. This local smoothing technique is called *regressogram*, the name which was first conceived by Tukey (1961) [42]. In principle, a regressogram is an approximation of the data points by a step function with a fixed or variable bin size and a certain number of jumps, no greater than the total number of bins. In every bin the value of the estimator is given by a simple arithmetic average of all the data points Y_i falling within the bin, this means that if we define

$$\mathcal{I}_x = \{i : X_i \text{ and } x \text{ are in the same bin}\}$$

and denote by $\#\mathcal{I}_x$ the total number of points in the corresponding bin, then

$$\widehat{f}(x) = \frac{1}{\#\mathcal{I}_x} \sum_{i \in \mathcal{I}_x} Y_i.$$

It should be emphasized that in this classical case, the smoothing by a regressogram coincides with the local least square estimate by a constant function. This immediately follows from the fact that the solution of minimization problem (1.4) with a constant function g is given by the mean of Y_i.
For a given partition of the observation interval by a certain number of bins, the definition of the local constant estimator is certainly not unique. Thus, instead of computing the average of Y_i within each bin, one may resolve to using a median $\text{med}_{\mathcal{I}_x}(Y_i)$ of Y_i.
A local *median* $\text{med}_{\mathcal{I}_x}(Y_i)$ can be determined as a real number, which satisfies the following conditions: at least half of all data in the current bin have values less than or equal to $\text{med}_{\mathcal{I}_x}(Y_i)$ and at least half of all data have values greater than or equal to $\text{med}_{\mathcal{I}_x}(Y_i)$ [14]. From the variational viewpoint a local median $\text{med}_{\mathcal{I}_x}(Y_i)$ is obtained by minimizing the sum

$$\sum_{i \in \mathcal{I}_x} |\, c - Y_i \,|, \quad c \in \mathbb{R} \tag{1.7}$$

and, therefore, a median $\text{med}(Y)$ of the whole data set $Y = (Y_1, \ldots, Y_n)$ is the minimizer of the functional S_1 from (1.5) in the case of constant g.
The most common way (see e.g. [14]) of computing the *median of an ordered sample* of n data points $Y = (Y_1, \ldots, Y_n)$ is as follows:

$$\text{med}(Y) = \begin{cases} Y_{\frac{n+1}{2}} & : \ n \text{ odd} \\ \frac{1}{2}\left(Y_{\frac{n}{2}} + Y_{\frac{n}{2}+1}\right) & : \ n \text{ even.} \end{cases} \tag{1.8}$$

It is well known that the median of a data set is generally not a unique number. Therefore, further on, we will often work with the set of medians, all corresponding to the same data set $Y = (Y_1, \ldots, Y_n)$ and will denote this set by $\mathit{med}(Y)$ or by $\mathit{med}_{\mathcal{I}}(Y)$, where $\mathcal{I}$ stands for a certain set of indexes of the data.

In many practical applications the smoothing using local medians has a certain advantage over the smoothing by local averages, or, equivalently, over the smoothing by the local least squares estimates. This mainly refers to the cases of the raw data, which may contain a certain amount of outliers. It is well known that the least squares estimates are very sensitive with respect to the presence of the outliers in the data. In fact, even a single spike Y_0 of the response data, which does not fall within the expected normal bounds, renders the local least squares estimate to vary linearly as $(1/N)Y_0$, where N is the number of data points in the bin. On the contrary, the local median estimate does not change at all with Y_0, as long as Y_0 stays always above or below the median (see Eq. (1.8)). In other words, the median estimate is much more robust with respect to the presence of the outliers than the least squares estimate.

1.3 Jump penalized estimators

As we saw in the previous section, the estimator function is often constructed in such a way as to minimize the deviation measure between the actual data points and the estimator function. If no further restrictions or penalties were applied, the solution of such a minimization problem would be any function, which passes through the data points Y_i, implying that the relative deviation error is strictly zero. Obviously, such a degenerate construction can hardly be regarded as smoothing, because the corresponding estimator function would have exactly same shape as the actual response data Y_i. Therefore, one needs to find some sort of balance between the roughness of the estimator function and the degree of its fidelity to the measured data. This brings us to the concept of *penalization* of the estimators. Depending on the smoothing method, there exist many different ways of how an estimator can be penalized to achieve the desired balance.
For instance, in case of the spline smoothing (see (1.6)) the degree of roughness of the estimator is assumed to be given by $\gamma \int (g''(x))^2\, dx$, with some fixed smoothing parameter γ. The estimator is then computed by balancing the mean square error and the roughness integral.
For the class of local constant estimators, or in other words, for the approximations by step functions, the degree of complexity can be characterized by the number of jumps of the estimator $\widehat{f}$.
Previously several different penalization criteria, which are based on the number of jumps of the estimator, have been studied in detail. Thus, in the work of Hinkley [20] it was assumed that the estimator is a step function with a single jump with unknown location of the change-point. More recently, by Yao and Au, in Refs. [48, 49], the number of change-points was assumed to have a fixed upper bound. The actual number of jumps was then determined using Schwarz' criterion [40].
A much more universal approach does not make any assumptions on the number of jumps a priori. Instead, the optimal number of jumps and the corresponding estimator are determined simultaneously by minimizing the so-called Potts functional. This functional was firstly used by R.B.Potts [34] as a generalization of Ising model [23] in statistical mechanics. Potts functional is a special case of the one-dimensional Mumford-Shah functional [30, 31], asymptotically behaviour of which was considered in detail for instance by Kempe *et al.* [26]. A comprehensive study of the regression analysis of one-dimensional data series based on Potts functional approach was conducted in Refs. [43, 44, 25, 45, 46, 16, 7, 8]. Complexity penalized sums of squares for the two-dimensional or more-dimensional data series have been analyzed in the Ph.D. thesis of Friedrich [15] and in the work of Demaret *et al.* [12], respectively.

Here we give a brief description of the approximation techniques based on the minimization of the one-dimensional Potts functionals.

Consider the following fixed design regression model:

$$y_i^n = f_i^n + \xi_i^n, \quad i = 1, \dots, n,\ n \in \mathbb{N}, \tag{1.9}$$

where f_i^n is the average value of a square integrable function f ($f \in L^2([0,1))$) over the interval $[(i-1)/n,\ i/n)$ and $(\xi_i^n)_{1\le i\le n,\, n\in\mathbb{N}}$ is a triangular array of random variables with zero expectation and finite variance.
The Potts functional $H^n_{\gamma,2}$ is defined on $\mathbb{R}^n$ as

$$H^n_{\gamma,2}(u, y^n) = \gamma \# J(u) + \frac{1}{n}\sum_{i=1}^{n}(u - y_i^n)^2, \tag{1.10}$$

where $\gamma \ge 0$ is the so-called regularization parameter, $J(u) := \{i : 1 \le i \le n-1,\ u_i \ne u_{i+1}\}$ is the set of "jumps" of $u \in \mathbb{R}^n$ and $\#J(u)$ denotes the number of elements of $J(u)$.

Note that the penalization term $\gamma\#J(u)$ for a special choice of γ plays the role of Schwarz' criterion for determining the number of change-points of the least square estimator of a step function as discussed by Yao and Au [48, 49].

Minimizers of $H^n_{\gamma,2}(\cdot,\ y^n)$ always exist, but need not be unique (see Kempe [25]). Embedding such minimizers into the space of step functions $S\big([0,1)\big) \subset L^2\big([0,1)\big)$ by the map $\iota^n : \mathbb{R}^n \to S\big([0,1)\big)$:

$$\iota^n\big((u_1, \dots, u_n)\big) := \sum_{i=1}^{n} u_i \cdot \mathbf{1}_{\left[\frac{i-1}{n},\, \frac{i}{n}\right)}(t), \tag{1.11}$$

yields an estimator of the unknown function f.
It is easy to see that the parameter γ regulates the complexity of the considered estimator. Namely, as mentioned before, if $\gamma = 0$, then the minimizer of Potts functional (1.10) would have exactly the same shape as the data $y^n = (y_1^n, \dots, y_n^n)$, implying that the number of bins of the considered estimator, which accords to (1.11), is automatically equal to the number of the data points y^n. On the other hand, if $\gamma \to \infty$, then it is clear that the minimizer of $H^n_{\gamma,2}(\cdot,\ y^n)$ reduces to the mean value of the given data and, consequently, the corresponding estimator will be a constant function without any jumps.

Winkler and Liebscher [43] have shown that the family of the minimizers of functional (1.10) can be computed in $O(n^3)$ steps and that the minimizer for a fixed γ can be computed in $O(n^2)$ steps.
The application of the above approach to the sets of data from real life problems, for example to MRI Brain Scans Data or to the data from Gene Expression Experiments has been discussed in detail in the works of Winkler *et al.* [44, 45].
The details of consistency of the estimator given by the minimizer of the Potts functionals was studied by Boysen *et al.* in Refs. [7, 8], which complement each other.
In these two works by Boysen *et al.* it has been shown that when $n \to \infty$, for any function $f \in L^2([0,1))$ and all sequences $(\gamma_n)_{n\in\mathbb{N}}$ such that $\gamma_n \underset{n\to\infty}{\longrightarrow} \gamma \ge 0$ and $\gamma_n \cdot (n/\log n) \underset{n\to\infty}{\longrightarrow} \infty$ the functional $H^n_{\gamma_n,2}(\cdot,\ y^n)$ converges in the sense of the so-called epi-convergence (see for

example Dal Maso [11] or later Definition 1, p.21) to the "continuous" Potts functional $H^{\infty}_{\gamma,2}(\cdot,\, f)$, which is defined for all step functions $g \in S\big([0,1)\big)$ and any $\gamma > 0$ as follows:

$$H^{\infty}_{\gamma,2}(g,f) \;\; := \;\; \gamma\#J(g) + \|g-f\|_2 - \|f\|_2, \tag{1.12}$$

where $J(g)$ is the set of discontinuity points of function g. It was also proved that the set of minimizers of $H^{\infty}_{\gamma,2}(\cdot,\, f)$ for any $f \in L^2([0,\,1))$ and all $\gamma \geq 0$ is not empty. One of the main results of these works is the theorem, which proves for the so-called uniform sub-Gaussian noise and a certain choice of γ_n the $\mathbb{P}$-almost surely convergence of a sequence of minimizers of the functionals $H^{n}_{\gamma_n,2}\,(\,\cdot\,,\,y^n)$, which is embedded in $S\big([0,\,1)\big)$, to the original function $f \in L^2\big([0,\,1)\big)$ in the L^2 metric, as n goes to ∞. For the case of a step function f, the authors found the optimal rate of this convergence. Moreover, the consistency results have also been obtained for various metrics and for different classes of the function f. For instance, consistency of the jump points in case of a picewise constant function f was shown with respect to the Hausdorff Metric (see for example Matheron [28] or later Definition 4, p.29). Simultaneous convergence of the jump points and the graph of the function was obtained in the Skorokhod topology (definition can be found in the book of Billingsley [6]). In case when the true regression function f is not a step function, the optimal rates of convergence were proved in certain classes of approximation spaces [13], especially the rates were given if f is the function of bounded variation, or if f is Hölder continuous on $[0,1]$ of order $\alpha \in (0,1]$.

1.4 Jump penalized L^1-estimators

As mentioned above, one of the shortcomings of the least squares estimators is their sensitivity to outliers. As we saw in the previous section, the Potts functional method is effectively based on the local least squares estimators, and, therefore, inherits all the disadvantages of the latter. On the other hand, we have learned that an extremely effective way of dealing with the outliers is to replace the least squares estimates by the median estimates. This brings us to the main motivation behind the present work.
Our first main objective is to develop a robust jump penalized regression model, which is much less sensitive to the outliers in the data set than the previously used model based on the least squares estimates [43, 44, 25, 45, 46, 16, 7, 8]. To this end we first modify Potts functional (1.10) replacing the mean squares deviation error by the mean absolute deviation error:

$$H^n_\gamma(u,y^n) = \gamma\#J(u) + \frac{1}{n}\sum_{i=1}^{n}|u-y^n_i|. \tag{1.13}$$

For further purposes it is reasonable to introduce the following definitions.
Let (Θ,μ) be a measure space and let

$$L^1(\Theta) := \Big\{\, f:\Theta\to\mathbb{R}\,:\, \int\limits_{\Theta} |f(x)|\,d\mu(x) < \infty \Big\}. \tag{1.14}$$

For any $f \in L^1(\Theta)$ the norm is defined by

$$\|f\| = \int\limits_{\Theta} |f(x)|\,d\mu(x). \tag{1.15}$$

For an arbitrary functional $F : \Theta \to \mathbb{R}$ we denote by $\operatorname{argmin} F$ or, equivalently, by $\underset{\Theta}{\operatorname{argmin}} F$ the entire set of its minimizers in Θ.
For an arbitrary $f \in L^1(\Theta)$ and some $\Theta_0 \subseteq \Theta$ we define *the set of medians of the function* f as follows:

$$med_{\Theta_0}(f) = \underset{c \in \mathbb{R}}{\operatorname{argmin}} \int\limits_{\Theta_0} |\, f(x) - c \,|\; d\mu(x). \tag{1.16}$$

In other words, a constant $\mathrm{med}_{\Theta_0}(f)$ is a *median of the function* f, i.e. $\mathrm{med}_{\Theta_0}(f) \in med_{\Theta_0}(f)$, if

$$\mu\Big(\big\{x \in \Theta_0 : f(x) \geq \mathrm{med}_{\Theta_0}(f)\big\}\Big) \geq \frac{\mu(\Theta_0)}{2}$$

and

$$\mu\Big(\big\{x \in \Theta_0 : f(x) \leq \mathrm{med}_{\Theta_0}(f)\big\}\Big) \geq \frac{\mu(\Theta_0)}{2}.$$

Next, we reformulate the fixed design regression model (1.9) for the class $L^1\big([0,\,1)\big)$ of absolutely integrable on the interval $[0,1)$ functions.
Let

$$y_i^n = f_i^n + \xi_i^n, \quad i \in \{1, \ldots, n\},\ n \in \mathbb{N}, \tag{1.17}$$

where a triangular array of random variables $(\xi_i^n)_{1 \leq i \leq n,\, n \in \mathbb{N}}$, which are centered around zero, represents noise and the numbers $f_1^n, \ldots, f_n^n$ approximate some function $f \in L^1([0,1))$ in the following way:

$$f(t) = \lim_{n \to \infty} \sum_{i=1}^{n} f_i^n \cdot \mathbf{1}_{\left[\frac{i-1}{n},\, \frac{i}{n}\right)}(t), \tag{1.18}$$

for example

$$f_i^n \in med_{\left[\frac{i-1}{n}, \frac{i}{n}\right)}(f). \tag{1.19}$$

For the given data points $y^n = (y_1^n, \ldots, y_n^n)$ as defined in (1.17) we want to estimate the function f by a step function. An estimating step function $\widehat{f^n}$ should obey certain properties, namely:

(i) $\widehat{f^n}$ should have as small as possible mean absolute deviation from the data points y_i^n

(ii) $\widehat{f^n}$ should be as smooth as possible, i.e. the number of jump points should be as small as possible.

Similar to the case of jump penalized least squares, in order to solve this problem, i.e. to find $\widehat{f^n}$, which fulfills the properties (i) and (ii), we will consider the set of minimizers of modified Potts functional (1.13), which is embedded into the set of step functions $S\big([0,1)\big) \subset L^1\big([0,1)\big)$ by the map ι^n given by (1.11).
Therefore, for the data $y^n := (y_1^n, \ldots, y_n^n)$ and some function $f \in L^1\big([0,1)\big)$, related to one another via regression model (1.17), the estimators of the function f, which we wish to study in details in this work, will be given by

$$\widehat{f^n} \in \iota^n\big(\operatorname{argmin} H_{\gamma_n}^n(\cdot\,, y^n)\big). \tag{1.20}$$

Note that the estimators $\widehat{f}^n$, obtained from (1.20), necessarily satisfy both above criteria (i) and (ii). Analogously to the L^2 case, the regularization parameter γ controls the number of jumps of $\widehat{f}^n$.
As we have already noticed before, for any $n \in \mathbb{N}$, any set of indexes $\mathcal{I} \subseteq \{1, \dots, n\}$ and data y_i^n, $i \in \mathcal{I}$ the set of minimizers of the sum $\sum_{i \in \mathcal{I}} |u - y_i^n|$ is equivalent to the set of all medians of the data y_i^n, $i \in \mathcal{I}$. Two important facts follow from this statement immediately. Firstly, for any $\gamma \geq 0$ and any finite $n \in \mathbb{N}$ a minimizer of the functional $H_\gamma^n(\cdot\,, y^n)$ always exists, but is not necessarily unique.
Secondly, the value of the estimator function $\widehat{f}^n$ from (1.20) on any interval I, where $\widehat{f}^n$ is constant, is a median $\mathrm{med}_I(y) \in med_I(y)$ of the points y_i^n, which lie within the interval I.

Moreover, according to the work of Friedrich *et al.* [16], the computation of minimizers of the functional $H_\gamma^n(\cdot\,, y^n)$ is also feasible. Thereby the minimum of functional (1.13) for a certain $\gamma \geq 0$ can be computed in $O(n^2 \ln n)$ steps, and for all γ simultaneously, in $O(n^3)$ steps.
Following the works of Boysen *et al.* [7, 8] discussed above, in this work we study the consistency properties of the estimator, which is based on the minimizers of modified Potts functional (1.13).
Our first goal is to investigate the limiting behaviour of $H_{\gamma_n}^n(\cdot\,, y^n)$ and its minimizers for $y^n = (y_1^n, \dots, y_n^n)$ from model (1.17), as n tends to infinity. This is done in Chapter 2.
Further on, in Chapter 3, we study the properties of the limiting functional, which is determined in Chapter 2. We will see certain differences between the L^2 and the L^1 cases. As the main result here, we show that the set of minimizers of the limiting functional is not empty.
In Chapters 2 and 3 we assume that noise is centered around zero. The more detailed conditions on the noise are given in the beginning of Chapter 2. In order to prove consistency of the considered estimate, the conditions on the noise term should be strengthened. Because our main objective is to study the data sets that contain outliers, it is reasonable, as mentioned before, to consider a distribution function of the random noise, which is taken from the family of the so-called heavy-tailed distributions, i.e. such distributions, whose tails are not exponentially bounded. For this purpose we will assume that the noise term is distributed according to the symmetrical Pareto distribution with the probability density given by $f(x) = \dfrac{\alpha - 1}{2\,(1+|x|)^\alpha}$, $\alpha \geq 9$, which is an example of a heavy-tailed distribution.
Chapter 4 contains some technical results about the properties of a Pareto distributed noise with the above probability density, which we use in the following chapter in order to show consistency of the estimators.
The main results and theorems about consistency of $\widehat{f}^n \in \iota^n\big(\operatorname{argmin} H_{\gamma_n}^n(\cdot\,, y^n)\big)$ are presented in Chapter 5. Here we prove that for a Pareto distributed noise $(\xi_i^n)_{1 \leq i \leq n,\, n \in \mathbb{N}}$ and for a suitable sequence of $(\gamma_n)_{n \in \mathbb{N}}$ the sequence $(\widehat{f}_n)_{n \in \mathbb{N}}$ converges $\mathbb{P}$-almost surely to the original function f with respect to the L^1 metric in case when f itself is a step function. This property is first proved for the case of a constant function f, following by the cases of a step function with one, two and, subsequently, an arbitrary number of jumps.

We complete this thesis with a summary and outlook.

In Appendices A and B we collect all important auxiliary propositions and lemmas as well as symbols and notations used in this work.

Chapter 2

Towards the continuous Model.

The main objective of this work is to show convergence of a sequence of estimators $(\widehat{f}^n)_{n\in\mathbb{N}}$, obtained by minimizing the functionals $H^n_{\gamma_n}(\cdot\,, y^n)$, towards the original function f. More precisely, one needs to show that for any function $f \in L^1\big([0,1)\big)$ and any given data $y^n = (y^n_1, \dots, y^n_n)$ satisfying the model (1.17), i.e. $y^n_i = f^n_i + \xi^n_i$, where $(\xi^n_i)_{1\le i\le n,\, n\in\mathbb{N}}$ is a triangular array of random variables and $(f^n_i)_{1\le i\le n,\, n\in\mathbb{N}}$ approximate the function f by means of (1.18), the piecewise constant function $\widehat{f}^n$ tends to f for sufficiently large n.
Taking this into account we will focus in this chapter on the properties of the functional $H^n_{\gamma_n}(\cdot\,, y^n)$, as $n \to \infty$.

To start with, we impose certain conditions on the noise variables ξ^n_i.

Condition 1. *The triangular array $(\xi^n_i)_{1\le i\le n,\, n\in\mathbb{N}}$ of random variables on a probability space $(\Omega, \mathbb{P})$ obeys the properties:*

- **(i)** *For all $n \in \mathbb{N}$ the random variables $\big(\xi^n_i\big)_{1\le i\le n}$ are independent and identically distributed and $\mathbb{E}|\xi^n_i| = \mathbb{E}|\xi| < \infty$ for each $i \in \{1, \dots, n\}$, where ξ is also an independent random variable on the space $(\Omega, \mathbb{P})$, having the same distribution as ξ^n_i, $i \in \{1, \dots, n\}$.*
- **(ii)** $\mathbb{P}\big(\xi^n_i \ge 0\big) \ge \dfrac{1}{2}$ *and* $\mathbb{P}\big(\xi^n_i \le 0\big) \ge \dfrac{1}{2}$ *for all $n \in \mathbb{N}$ and $i \in \{1, \dots, n\}$*

Further on, we introduce a new functional, which serves as a "continuous" analogue of the functional H^n_γ from (1.13) and is defined in the space of absolutely integrable functions $L^1\big([0,1)\big)$.

To this end we introduce the following auxiliary notations.

Recall that $S\big([0,1)\big)$ is the set of all piecewise constant functions on $[0,1)$ having a finite number of jumps.
Additionally, let $S_n\big([0,1)\big)$ be the set of all piecewise constant functions on $[0,1)$ with at most $n-1$ jumps, where all the discontinuity points are represented in the form $\dfrac{i}{n}$, $i = 1, \dots n$. Obviously, $S_n\big([0,1)\big) \subset S\big([0,1)\big)$.

Under a *partition* P of the interval $[0,1)$ we understand a finite set of non-overlapping subintervals of the form $[a,b)$, whose union coincides with $[0,1)$.

Let $\mathcal{P}$ denote the set of all partitions of the interval $[0,1)$.

For any step function $g \in S\big([0,1)\big)$ we denote by $J(g)$ the set of its discontinuity points, i.e.

$$J(g) := \Big\{ t \in [0,1) \ : \ g(t^+) \neq g(t^-) \Big\},$$

and by $P_{J(g)}$ the partition of the interval [0, 1), generated by the points of the set $J(g)$. In other words, let $g(x) = \sum_{i=1}^{m} g_i \cdot \mathbf{1}_{[a_{i-1},\, a_i)}(x)$, thereby $m \in \mathbb{N}$, $g_i \in \mathbb{R}$ for all $i \in \{1, \dots, m\}$ and $0 =: a_0 < a_1 < a_2 < \ldots < a_{m-1} < a_m := 1$. Then

$$J(g) = \Big\{ a_1, a_2, \dots, a_{m-1} \Big\} \ \text{ and } \ P_{J(g)} = \Big\{ [0,\, a_1), [a_1,\, a_2), \dots, [a_{m-1},\, 1) \Big\}.$$

For any $A \subseteq \mathbb{R}$ we denote by $l(A)$ the Lebesgue measure of A, implying that for any interval $[a,b) \subseteq [0,1)$, $l(\,[a,b)\,)$ is the length of $[a,b)$.

Now we define the functional $\tilde{H}^n_\gamma : L^1\big([0,1)\big) \times L^1\big([0,1)\big) \to \mathbb{R} \cup \{\infty\}$ for any $\gamma \geq 0$ and any $n \in \mathbb{N}$ as follows:

$$\tilde{H}^n_\gamma(g,f) \ := \ \begin{cases} \gamma \# J(g) + \|g - f\| - \|f\| & : \ g \in S_n\big([0,1)\big),\ \gamma > 0 \\ \|g - f\| - \|f\| & : \ g \in L^1\big([0,1)\big),\ \gamma = 0 \\ \infty & : \ \text{otherwise} \end{cases} \tag{2.1}$$

Here we notice that the continuous functional $\tilde{H}^n_\gamma$ differs from its discrete analogue $H^n_\gamma(u,Y) = \gamma \# J(u) + \frac{1}{n} \sum_{i=1}^{n} |u - Y_i|$, $Y \in \mathbb{R}^n$, $u \in \mathbb{R}^n$, by an extra term $\|f\|$. However, for convenience, we will use throughout the thesis the functional $\tilde{H}^n_\gamma(\cdot\,, f)$, as it is clear that the set of its minimizers is identical with the set of the minimizers of the functional $\tilde{H}^n_\gamma(\cdot\,, f) + \|f\|$. The reason for this minor modification of the continuous functional can be explained as follows.

Consider the function $\Phi : \mathbb{R}^2 \to \mathbb{R}$ defined by the rule

$$\Phi(a,x) := |a - x| - |x|. \tag{2.2}$$

It is easy to see that the function Φ, as a function of x, depending on whether a is positive or negative, can be rewritten as follows:
Let $a > 0$. Then it holds:

$$\Phi(a,x) = \begin{cases} a & , \quad x \leq 0 \\ a - 2\,x & , \quad 0 < x < a \\ -a & , \quad x \geq a \end{cases} \tag{2.3}$$

Let $a < 0$. Then we have:

$$\Phi(a,x) = \begin{cases} a & , \quad x \leq a \\ 2\,x - a & , \quad a < x < 0 \\ -a & , \quad x \geq 0 \end{cases} \tag{2.4}$$

For positive and negative a the function Φ is shown in Fig. 2.1 and Fig. 2.2, respectively.

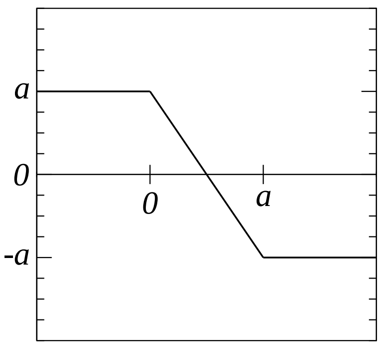

Figure 2.1: $y = |a - x| - |x|$, $a > 0$

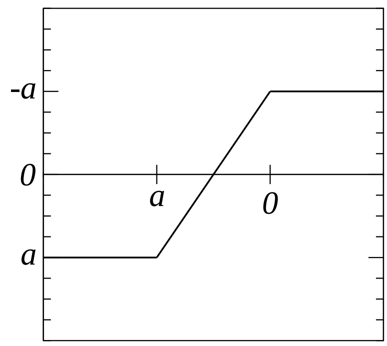

Figure 2.2: $y = |a - x| - |x|$, $a < 0$

According to this, we see that for any fixed a the function Φ has the following important properties:

1) Φ is a monotonic function, thereby, it is monotonically increasing if $a < 0$ and it is monotonically decreasing if $a > 0$;

2) Φ is a bounded function, namely for all $a \in \mathbb{R}$ and $x \in \mathbb{R}$ it holds

$$-|a| \leq |a - x| - |x| \leq |a|. \tag{2.5}$$

Consequently, for any two arbitrary functions $g_1 \in L^1\big([0,1)\big)$ and $g_2 \in L^1\big([0,1)\big)$ it holds:

$$-\infty < -\int\limits_{[0,1)} |g_1(x)|\, dx \leq \int\limits_{[0,1)} \big(|g_1(x) - g_2(x)| - |g_2(x)|\big)\, dx \leq \int\limits_{[0,1)} |g_1(x)|\, dx < \infty,$$

and, therefore, the functional

$$\Psi(g_1, g_2) := \|g_1 - g_2\| - \|g_2\| \tag{2.6}$$

is bounded and always attains finite values, while nothing concrete can be said about the integrability of $|g_1(x) - g_2(x)|$ and, hence, about the finiteness of the norm $\|g_1 - g_2\|$.

Therefore, it is in fact convenient to use in continuous case the functional $\tilde{H}^n_\gamma(\cdot\,, f)$, as defined in (2.1), instead of $\tilde{H}^n_\gamma(\cdot\,, f) + \|f\|$.

The next proposition explains the connection between the set of minimizers of the discrete and continuous functionals.

Proposition 1. *For all $n \in \mathbb{N}$, all $Y = (Y_1, \ldots, Y_n) \in \mathbb{R}^n$ and any $\gamma \geq 0$ it holds: $u \in \operatorname{argmin} H^n_\gamma(\cdot\,, Y)$ if and only if $\iota^n(u) \in \operatorname{argmin} \tilde{H}^n_\gamma(\cdot\,, \iota^n(Y))$, where ι^n is defined by (1.11).*

Proof:

Using (2.1) for all $\gamma > 0$ and all $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$ the functional $\tilde{H}^n_\gamma(\iota^n(u), \iota^n(Y))$ can be transformed as follows:

$$\tilde{H}^n_\gamma(\iota^n(u), \iota^n(Y)) = \gamma \# J(\iota^n(u)) + \|\iota^n(u) - \iota^n(Y)\| - \|\iota^n(Y)\|$$

$$
\begin{aligned}
&\underset{(1.11)}{=} \quad \gamma \# J(\iota^n(u)) + \int\limits_{[0,1)} \Big| \sum_{i=1}^{n} u_i \cdot \mathbf{1}_{[\frac{i-1}{n}, \frac{i}{n})}(x) - \sum_{i=1}^{n} Y_i \cdot \mathbf{1}_{[\frac{i-1}{n}, \frac{i}{n})}(x) \Big| \, dx - \|\iota^n(Y)\| \\
&= \quad \gamma \# J(\iota^n(u)) + \sum_{i=1}^{n} \int\limits_{[\frac{i-1}{n}, \frac{i}{n})} \Big| \sum_{i=1}^{n} u_i \cdot \mathbf{1}_{[\frac{i-1}{n}, \frac{i}{n})}(x) - \sum_{i=1}^{n} Y_i \cdot \mathbf{1}_{[\frac{i-1}{n}, \frac{i}{n})}(x) \Big| \, dx - \|\iota^n(Y)\| \\
&= \quad \gamma \# J(\iota^n(u)) + \sum_{i=1}^{n} \int\limits_{[\frac{i-1}{n}, \frac{i}{n})} |u_i - Y_i| \, dx - \|\iota^n(Y)\| \\
&= \quad \gamma \# J(u) + \frac{1}{n} \sum_{i=1}^{n} |u_i - Y_i| - \|\iota^n(Y)\| \\
&\underset{(1.13)}{=} \quad H_\gamma^n\,(u, Y) - \|\iota^n(Y)\|.
\end{aligned}
$$

Obviously, the same result is valid for $\gamma = 0$ and all $u \in \mathbb{R}^n$.
Thus we obtain that for all $\gamma \geq 0$ and all $u \in \mathbb{R}^n$ it holds:

$$
\tilde{H}_\gamma^n(\iota^n(u), \iota^n(Y)) = H_\gamma^n\,(u, Y) - \|\iota^n(Y)\|. \tag{2.7}
$$

Let $u = (u_1, \ldots, u_n) \in \operatorname{argmin} H_\gamma^n\,(\cdot\,, Y)$, which means that for all $u' \in \mathbb{R}^n$ we have

$$
H_\gamma^n\,(u, Y) \leq H_\gamma^n\,(u', Y). \tag{2.8}
$$

Now we consider $\tilde{H}_\gamma^n(\iota^n(u'), \iota^n(Y))$ for all $u' \in \mathbb{R}^n$:

$$
\tilde{H}_\gamma^n(\iota^n(u'), \iota^n(Y)) \underset{(2.7)}{=} H_\gamma^n\,(u', Y) - \|\iota^n(Y)\| \underset{(2.8)}{\geq} H_\gamma^n\,(u, Y) - \|\iota^n(Y)\| = \tilde{H}_\gamma^n(\iota^n(u), \iota^n(Y)).
$$

This shows that if $u \in \operatorname{argmin} H_\gamma^n(\cdot\,, Y)$ then $\iota^n(u) \in \operatorname{argmin} \tilde{H}_\gamma^n(\cdot\,, \iota^n(Y))$.
Conversely, let now $\iota^n(u) \in \operatorname{argmin} \tilde{H}_\gamma^n(\cdot\,, \iota^n(Y))$, i.e. $\forall\, u' \in \mathbb{R}^n$ it holds

$$
\tilde{H}_\gamma^n(\iota^n(u), \iota^n(Y)) \leq \tilde{H}_\gamma^n(\iota^n(u'), \iota^n(Y))
$$

or from (2.7)

$$
H_\gamma^n\,(u, Y) - \|\iota^n(Y)\| \leq H_\gamma^n\,(u', Y) - \|\iota^n(Y)\|.
$$

Since $\|\iota^n(Y)\| \geq 0$ we conclude from the last relation that $H_\gamma^n\,(u, Y) \leq H_\gamma^n\,(u', Y)$ and consequently, if $\iota^n(u) \in \operatorname{argmin} \tilde{H}_\gamma^n(\cdot\,, \iota^n(Y))$ then $u \in \operatorname{argmin} H_\gamma^n(\cdot\,, Y)$.
This completes the proof of Proposition 1.

□

Proposition 1 establishes the connection between the discrete and continuous cases for arbitrary $Y \in \mathbb{R}^n$. In the following corollary we rewrite the main statement of Proposition 1 for the data $y^n = (y_1^n, \ldots, y_n^n)$, which satisfy the regression relation (1.17), i.e.

$$
y_i^n = f_i^n + \xi_i^n.
$$

To this end we define the following two step functions $\xi^n \in S_n\big([0,1)\big)$ and $f^n \in S_n\big([0,1)\big)$, built on the basis of $\big(\xi_i^n\big)_{n\in\mathbb{N},1\le i\le n}$, which obeys Condition 1, and $\big(f_i^n\big)_{n\in\mathbb{N},1\le i\le n}$ fulfilling (1.17):

$$\xi^n \ := \ \iota^n\big((\xi_1^n,...,\xi_n^n)\big) = \sum_{i=1}^{n} \xi_i^n \cdot \mathbf{1}_{\left[\frac{i-1}{n},\frac{i}{n}\right)} \tag{2.9}$$

$$f^n \ := \ \iota^n\big((f_1^n,...,f_n^n)\big) = \sum_{i=1}^{n} f_i^n \cdot \mathbf{1}_{\left[\frac{i-1}{n},\frac{i}{n}\right)} \tag{2.10}$$

Taking these definitions and (1.18) into account, we set

$$f(t) = \lim_{n\to\infty} \sum_{i=1}^{n} f_i^n \cdot \mathbf{1}_{\left[\frac{i-1}{n},\frac{i}{n}\right)}(t)$$

and can claim that in the limit $n \to \infty$, the sequence $(f^n)_{n\in\mathbb{R}}$ converges to the unknown function $f \in L^1\big([0,1)\big)$, i.e.

$$\lim_{n\to\infty} f^n = f. \tag{2.11}$$

Corollary 1. *Let $y^n = (y_1^n,\dots,y_n^n)$, $n \in \mathbb{N}$, satisfy regression relation (1.17) with some $f \in L^1\big([0,1)\big)$ and let $f^n \in S_n\big([0,1)\big)$ and $\xi^n \in S_n\big([0,1)\big)$ obey (2.10) and (2.9), respectively. Then for any $\gamma \ge 0$ it holds:*

$$u \in \text{argmin } H_\gamma^n(\cdot\,, y^n) \ \Leftrightarrow \ \iota^n(u) \in \text{argmin } \tilde{H}_\gamma^n(\cdot\,,\, f^n + \xi^n). \tag{2.12}$$

Proof:
The sum $f^n + \xi^n$ can be explicitly written as follows:

$$f^n + \xi^n = \sum_{j=1}^{n} f_i^n \cdot \mathbf{1}_{[\frac{i-1}{n},\frac{i}{n})} + \sum_{j=1}^{n} \xi_i^n \cdot \mathbf{1}_{[\frac{i-1}{n},\frac{i}{n})} = \sum_{j=1}^{n} (f_i^n + \xi_i^n) \cdot \mathbf{1}_{[\frac{i-1}{n},\frac{i}{n})}$$
$$\underset{(1.17)}{=} \ \sum_{j=1}^{n} y_i^n \cdot \mathbf{1}_{[\frac{i-1}{n},\frac{i}{n})} \underset{(1.11)}{=} \iota^n(y^n).$$

So we have:

$$f^n + \xi^n = \iota^n(y^n). \tag{2.13}$$

From Proposition 1 we know that

$$u \in \text{argmin } H_\gamma^n(\cdot\,, y^n) \ \Leftrightarrow \ \iota^n(u) \in \text{argmin } \tilde{H}_\gamma^n(\cdot\,,\, \iota^n(y^n)).$$

Using (2.13) we conclude:

$$u \in \text{argmin } H_\gamma^n(\cdot\,, y^n) \ \Leftrightarrow \ \iota^n(u) \in \text{argmin } \tilde{H}_\gamma^n(\cdot\,,\, f^n + \xi^n).$$

□

For any $n \in \mathbb{N}$ and all $j,\, k \in \{1, \dots, n\}$, $j < k$, let us denote by $med_{[\frac{j}{n},\, \frac{k}{n})}(y^n)$ the set of all medians of points $(y_i^n)_{j \le i \le k-1}$, i.e.

$$med_{[\frac{j}{n},\, \frac{k}{n})}(y^n) = med(y_j^n, \dots, y_{k-1}^n). \tag{2.14}$$

From (2.13) and by taking into account the definitions (1.7) and (1.16) of the set of medians of a sample of points and of the set of medians of a function from $L^1\big([0,1)\big)$, respectively, it is clear that for all $n \in \mathbb{N}$ and for any interval $I = \left[\frac{i}{n}, \frac{j}{n}\right) \subseteq [0,1)$ it holds

$$med_I(y^n) = med_I(f^n + \xi^n). \tag{2.15}$$

Remark 1. *The main consequence of Corollary 1 can be summarized as follows. In case of the discrete functional, the estimator $\widehat{f^n}$ of the original function f based on the data y^n, which in turn satisfies the regression relation (1.17), is represented by*

$$\widehat{f^n} \in \iota^n\big(\operatorname{argmin} H_{\gamma_n}^n(\cdot\,, y^n)\big).$$

Whereas the very same estimator $\widehat{f^n}$ in case of the continuous functional is given by

$$\widehat{f^n} \in \operatorname{argmin}\, \tilde{H}_{\gamma_n}^n(\cdot\,,\, f^n + \xi^n). \tag{2.16}$$

From the technical viewpoint, it is easier to deal with the continuous functional. Therefore, from now on, we will be working with the estimator $\widehat{f^n}$, as given by (2.16).

Before we start a new subsection, which will be dealing with the analysis of asymptotic properties of the functional $\tilde{H}_{\gamma_n}^n$, we give an important remark about the structure of the minimizer $\widehat{f^n}$. This remark follows from (2.15) and from the fact mentioned in the introduction, namely that the value of the estimator function $\widehat{f^n} \in \iota^n\big(\operatorname{argmin} H_{\gamma_n}^n(\cdot\,, y^n)\big)$ on any interval $I = \left[\frac{i}{n}, \frac{j}{n}\right)$, where $\widehat{f^n}$ is constant, is given by a median $\mathrm{med}_I(y) \in med_I(y)$, as defined in (2.14).

Remark 2. *Let $f \in L^1\big([0,1)\big)$ and $(\xi_i^n)_{1 \le i \le n,\, n \in \mathbb{N}}$ fulfill Condition 1, $f^n \in S_n\big([0,1)\big)$ and $\xi^n \in S_n\big([0,1)\big)$ satisfy* (2.10) *and* (2.9), *respectively.*
For all $n \in \mathbb{N}$, any $\gamma > 0$, if

$$\widehat{f_n} \in \operatorname{argmin}\, \tilde{H}_{\gamma}^n(\cdot\,,\, f^n + \xi^n),$$

then there exist different natural numbers $\widehat{k}_1^n, \dots, \widehat{k}_{\#J(\widehat{f_n})}^n \in \{1, \dots, n-1\}$, $\widehat{k}_0^n := 0$, $\widehat{k}_{\#J(\widehat{f_n})+1}^n := n$ with the properties:
$J(\widehat{f_n}) := \left\{\frac{\widehat{k}_1^n}{n}, \dots, \frac{\widehat{k}_{\#J(\widehat{f_n})}^n}{n}\right\}$ is the set of the discontinuity points of $\widehat{f_n}$ and the intervals $\widehat{I}_i^n := \left[\frac{\widehat{k}_{i-1}^n}{n}, \frac{\widehat{k}_i^n}{n}\right)$ form a partition of the interval $[0,\, 1)$, i.e. $P_{J(\widehat{f_n})} := \big\{\widehat{I}_1^n, \dots, \widehat{I}_{\#J(\widehat{f_n})+1}^n\big\}$, such that

$$\widehat{f_n}(x) = \sum_{i=1}^{\#J(\widehat{f_n})+1} \widehat{f}_{n,i} \cdot \mathbf{1}_{\widehat{I}_i^n}(x), \tag{2.17}$$

where

$$\widehat{f}_{n,i} \in med_{\widehat{I}_i^n}(f^n + \xi^n), \;\; i \in \big\{1, \dots, \#J(\widehat{f_n}) + 1\big\}. \tag{2.18}$$

2.1 Limit functional

In this section we discuss the limiting behaviour of the continuous functionals $\tilde{H}^n_{\gamma_n}(\cdot\,,\, f^n+\xi^n)$ as n tends to infinity.
A very useful tool for studying the asymptotic behaviour of the minimizers of functionals is the so-called epi-convergence or Γ-convergence (see for example Dal Maso [11], or Hess [19]).

Definition 1. *Let $F_n : X \to \mathbb{R}\cup\{\infty\}$, $n \in \mathbb{N}$ and $F : X \to \mathbb{R}\cup\{\infty\}$ be functionals on a metric space (X, ρ). One says that the sequence $(F_n)_{n\in\mathbb{N}}$ epi-converges to F (symbol $F_n \underset{n\to\infty}{\overset{epi}{\longrightarrow}} F$), if the following two statements are valid:*

(i) *for any sequence $(x_n)_{n\in\mathbb{N}} \subset X$ with $x_n \underset{n\to\infty}{\longrightarrow} x$, $x \in X$ it holds*

$$\liminf_{n\to\infty} F_n(x_n) \geq F(x), \tag{2.19}$$

(ii) *for all $x \in X$ there exists a sequence $(x_n)_{n\in\mathbb{N}} \subset X$ with $x_n \underset{n\to\infty}{\longrightarrow} x$ such that*

$$\limsup_{n\to\infty} F_n(x_n) \leq F(x). \tag{2.20}$$

Derived from the Greek $\varepsilon\pi i$, which stands for "on the basis of " or "upon", the concept of the epi-convergence naturally co-exists with the standard pointwise convergence of sequences and complements the latter in the following important way. It is well known that the pointwise convergence does not necessarily preserve the minimum of a function or a functional, as the limit is taken, on the contrary, as shown for example by Beer [5, Theorem 5.3.6], if a sequence of functionals is said to be epi-convergent to a certain limiting functional, then the corresponding converging sequence of the minimizers necessarily approaches the minimizer of the limiting functional. This property of the epi-convergence together with the assumptions of the compactness and the almost surely uniqueness, has been used by Boysen *et al.* [7, 8] in order to show consistency of the estimators.

In the present work consistency of the estimators will be shown without involving the notion of the epi-convergence. However, the limiting behaviour of the functional $\tilde{H}^n_{\gamma_n}(\cdot\,,\, f^n+\xi^n)$ itself will be considered in the sense of epi-convergence. Namely, after introducing several auxiliary results, we will show that when $(\xi^n_i)_{1\leq i\leq n,\, n\in\mathbb{N}}$ fulfill Condition 1, $f \in L^1\big([0,1)\big)$, n is approaching infinity and $\gamma_n \underset{n\to\infty}{\longrightarrow} \gamma \geq 0$, then the functional $\tilde{H}^n_{\gamma_n}(\cdot\,,\, f^n+\xi^n)$ epi-converges to the functional $H^*_{\gamma,\xi}(\cdot\,,\, f)$, which for any $\gamma \geq 0$ and random variable $\xi : (\Omega, \mathbb{P}) \to \mathbb{R}$ satisfying Condition 1 is defined as follows:

$$H^*_{\gamma,\xi}(g, f) := \begin{cases} \gamma\#J(g) + \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du & : \ g \in S\big([0,1)\big),\ \gamma > 0; \\ \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du & : \ g \in L^1\big([0,1)\big),\ \gamma = 0; \\ \infty & : \ \text{otherwise}. \end{cases} \tag{2.21}$$

2.1.1 Limiting behaviour of the non-penalized functional

As the starting point, we omit the penalty term, which by $\tilde{H}^n_{\gamma_n}$ is proportional to γ_n, and only consider the remaining part of the functional, containing the difference of two norms. First, assume that the norm is applied to step functions.

Lemma 1. *Let $(\xi^n_i)_{1\leq i\leq n,\, n\in\mathbb{N}}$ fulfill Condition 1 and $\tilde{f}, \tilde{g} \in S\big([0,1)\big)$. Then it holds:*

$$\|\xi^n + \tilde{f} - \tilde{g}\| - \|\xi^n + \tilde{f}\| \underset{n\to\infty}{\overset{\mathbb{P}-a.s.}{\longrightarrow}} \int \mathbb{E}\big(|\xi + \tilde{f}(u) - \tilde{g}(u)| - |\xi + \tilde{f}(u)|\big)\, du \tag{2.22}$$

Proof:

We have $\tilde{f} \in S\big([0,1)\big)$ and $\tilde{g} \in S\big([0,1)\big)$, i.e. there exist two sets of finite constants $\tilde{f}_1, \ldots, \tilde{f}_{m_1}$ and $\tilde{g}_1, \ldots, \tilde{g}_{m_2}$, respectively, and two partitions $P_1 = \big\{[a_0, a_1), [a_1, a_2), \ldots, [a_{m_1-1}, a_{m_1})\big\}$ and $P_2 = \big\{[c_0, c_1), [c_1, c_2), \ldots, [c_{m_2-1}, c_{m_2})\big\}$ of the interval $[0,1)$ with $a_0 := 0$, $a_{m_1} := 1$, $c_0 := 0$, $c_{m_2} := 1$ so that

$$\tilde{f}(x) = \sum_{p=1}^{m_1} \tilde{f}_p \cdot \mathbf{1}_{[a_{p-1},\, a_p)}(x)$$

and

$$\tilde{g}(x) = \sum_{s=1}^{m_2} \tilde{g}_s \cdot \mathbf{1}_{[c_{s-1},\, c_s)}(x),$$

where m_1, m_2 are fixed and finite.

It is clear that there exist such integers $m \leq m_1 + m_2$ and such real numbers $b_0, b_1, \ldots, b_m$ with $b_0 := 0$, $b_m := 1$, $b_j \in \big\{a_1, \ldots, a_{m_1-1}, c_1, \ldots, c_{m_2-1}\big\}$ for $j = 1, \ldots, m-1$ that it holds

$$(\tilde{f} - \tilde{g})(x) = (\tilde{f} - \tilde{g}) \cdot \mathbf{1}_{[0,1)}(x) = \sum_{j=1}^{m} (\tilde{f} - \tilde{g})_j \cdot \mathbf{1}_{[b_{j-1},\, b_j)}(x)\ . \tag{2.23}$$

Moreover, there are two sets of real numbers $\tilde{f}'_1, \ldots, \tilde{f}'_m \in \big\{\tilde{f}_1, \ldots, \tilde{f}_{m_1}\big\}$ and $\tilde{g}'_1, \ldots, \tilde{g}'_m \in \big\{\tilde{g}_1, \ldots, \tilde{g}_{m_2}\big\}$, which allows us to rewrite the functions $\tilde{f}$ and $\tilde{g}$ as follows:

$$\tilde{f} \;=\; \sum_{j=1}^{m} \tilde{f}'_j \cdot \mathbf{1}_{[b_{j-1},\, b_j)}\ , \tag{2.24}$$

$$\tilde{g} \;=\; \sum_{j=1}^{m} \tilde{g}'_j \cdot \mathbf{1}_{[b_{j-1},\, b_j)}\ , \tag{2.25}$$

where $b_1, \ldots, b_m$ are points of discontinuity of function $(\tilde{f} - \tilde{g})$.
Hence,

$$\tilde{f} - \tilde{g} = \sum_{j=1}^{m} (\tilde{f} - \tilde{g})_j \cdot \mathbf{1}_{[b_{j-1},\, b_j)} = \sum_{j=1}^{m} (\tilde{f}'_j - \tilde{g}'_j) \cdot \mathbf{1}_{[b_{j-1},\, b_j)}. \tag{2.26}$$

For any $n \in \mathbb{N}$ we define the following subset of indexes

$$\mathbf{I}_n := \Big\{ i \in \{1,\dots,n\} : \exists\, j \in \{1,\dots,m-1\} \text{ and } M \in \{0,\dots,m-j-1\} \\ \text{such that } [b_j, b_{j+M}] \subseteq \Big(\frac{i-1}{n}, \frac{i}{n}\Big) \text{ and } b_{j-1},\, b_{j+M+1} \notin \Big(\frac{i-1}{n}, \frac{i}{n}\Big) \Big\},$$

thereby for each $i \in I_n$ the corresponding indexes j and M we denote by $j(i)$ and $M(i)$ respectively.

Now we consider $\|\xi^n + \tilde{f} - \tilde{g}\| - \|\xi^n + \tilde{f}\|$:

$$\begin{aligned}
&\|\xi^n + \tilde{f} - \tilde{g}\| - \|\xi^n + \tilde{f}\| \\
= &\int_{[0,1)} \Big| \sum_{i=1}^{n} \xi_i^n \cdot \mathbf{1}_{[\frac{i-1}{n},\frac{i}{n})}(u) + \sum_{j=1}^{m} \tilde{f}_j' \cdot \mathbf{1}_{[b_{j-1},b_j)}(u) - \sum_{j=1}^{m} \tilde{g}_j' \cdot \mathbf{1}_{[b_{j-1},b_j)}(u) \Big| \, du \\
&- \int_{[0,1)} \Big| \sum_{i=1}^{n} \xi_i^n \cdot \mathbf{1}_{[\frac{i-1}{n},\frac{i}{n})}(u) + \sum_{j=1}^{m} \tilde{f}_j' \cdot \mathbf{1}_{[b_{j-1},b_j)}(u) \Big| \, du \\
= &\sum_{j=1}^{m} \int_{[b_{j-1},b_j)} \Big(\Big| \sum_{i=1}^{n} \xi_i^n \cdot \mathbf{1}_{[\frac{i-1}{n},\frac{i}{n})}(u) + \tilde{f}_j' - \tilde{g}_j' \Big| - \Big| \sum_{i=1}^{n} \xi_i^n \cdot \mathbf{1}_{[\frac{i-1}{n},\frac{i}{n})}(u) + \tilde{f}_j' \Big| \Big) du \\
= &\sum_{j=1}^{m} \Bigg(\sum_{i:[\frac{i-1}{n},\frac{i}{n}) \subseteq [b_{j-1},b_j)} \int_{[\frac{i-1}{n},\frac{i}{n})} \Big(\big|\xi_i^n + \tilde{f}_j' - \tilde{g}_j'\big| - \big|\xi_i^n + \tilde{f}_j'\big| \Big) du \Bigg) \\
&+ \sum_{i \in \mathbf{I}_n} \Bigg[\int_{[\frac{i-1}{n}, b_{j(i)})} \Big(\big|\xi_i^n + \tilde{f}_{j(i)}' - \tilde{g}_{j(i)}'\big| - \big|\xi_i^n + \tilde{f}_{j(i)}'\big| \Big) du \\
&\quad + \int_{[b_{j(i)+M(i)}, \frac{i}{n})} \Big(\big|\xi_i^n + \tilde{f}_{j(i)+M(i)+1}' - \tilde{g}_{j(i)+M(i)+1}'\big| - \big|\xi_i^n + \tilde{f}_{j(i)+M(i)+1}'\big| \Big) du \\
&\quad + \sum_{k=j(i)+1,\, M(i)\neq 0}^{j(i)+M(i)} \int_{[b_{k-1}, b_k)} \Big(\big|\xi_i^n + \tilde{f}_k' - \tilde{g}_k'\big| - \big|\xi_i^n + \tilde{f}_k'\big| \Big) du \Bigg].
\end{aligned}$$

Using the function Φ defined in (2.2), we introduce the following new notations:

$$S_n^{(1)} := \sum_{j=1}^{m} \Bigg(\sum_{i:[\frac{i-1}{n},\frac{i}{n}) \subseteq [b_{j-1},b_j)} \int_{[\frac{i-1}{n},\frac{i}{n})} \Phi\big(\tilde{g}_j',\, \xi_i^n + \tilde{f}_j'\big)\, du \Bigg), \tag{2.27}$$

$$\begin{aligned}
S_n^{(2)} := &\sum_{i \in \mathbf{I}_n} \Bigg[\int_{[\frac{i-1}{n}, b_{j(i)})} \Phi\big(\tilde{g}_{j(i)}',\, \xi_i^n + \tilde{f}_{j(i)}'\big)\, du + \sum_{k=j(i)+1,\, M(i)\neq 0}^{j(i)+M(i)} \int_{[b_{k-1}, b_k)} \Phi\big(\tilde{g}_k',\, \xi_i^n + \tilde{f}_k'\big)\, du \\
&+ \int_{[b_{j(i)+M(i)}, \frac{i}{n})} \Phi\big(\tilde{g}_{j(i)+M(i)+1}',\, \xi_i^n + \tilde{f}_{j(i)+M(i)}'\big)\, du \Bigg].
\end{aligned}$$

Therefore, we obtain that for all $n \in \mathbb{N}$ it holds:

$$\|\xi^n + \tilde{f} - \tilde{g}\| - \|\xi^n + \tilde{f}\| = S_n^{(1)} + S_n^{(2)}. \tag{2.28}$$

First we estimate for all $n \in \mathbb{N}$ the sum $S_n^{(2)}$.
From property (2.6) of the function Φ we know that for all n, all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$ it is valid:

$$-|\tilde{g}_j'| \leq \left|\Phi\big(\tilde{g}_j',\, \xi_i^n + \tilde{f}_j'\big)\right| \leq |\tilde{g}_j'|.$$

Hence,

$$S_n^{(2)} \leq \sum_{i \in \mathbf{I}_n} \int\limits_{[\frac{i-1}{n},\, \frac{i}{n})} \max_{j \in \{1,\dots,m\}} |\tilde{g}_j'|\, du = \sum_{i \in \mathbf{I}_n} \frac{1}{n} \cdot \max_{j \in \{1,\dots,m\}} |\tilde{g}_j'| < m \cdot \max_{j \in \{1,\dots,m\}} |\tilde{g}_j'| \cdot \frac{1}{n} \underset{n \to \infty}{\longrightarrow} 0.$$

Consequently, according to (2.28), we can conclude that the sum $S_n^{(2)}$ does not affect in any way the limiting behaviour of the norm $\|\xi^n + \tilde{f} - \tilde{g}\| - \|\xi^n + \tilde{f}\|$ as $n \to \infty$.

Now we consider the behaviour of the sum $S_n^{(1)}$, as $n \to \infty$. But at first, for all $j \in \{1, \dots, m\}$ and all corresponding intervals $[b_{j-1}, b_j)$, we define the following two natural numbers l_j^n and r_j^n :

$$l_j^n := \min_{i : [\frac{i-1}{n}, \frac{i}{n}) \subseteq [b_{j-1}, b_j)} i \qquad \text{and} \qquad r_j^n := \max_{i : [\frac{i-1}{n}, \frac{i}{n}) \subseteq [b_{j-1}, b_j)} i\,. \tag{2.29}$$

From this it is clear that

$$r_j^n - (l_j^n - 1) = \#\Big\{i : \Big[\frac{i-1}{n}, \frac{i}{n}\Big) \subseteq \big[b_{j-1}, b_j\big)\Big\}$$

and

$$\frac{r_j^n - (l_j^n - 1)}{n} = l\,\Big(\bigcup_{i : [\frac{i-1}{n}, \frac{i}{n}) \subseteq [b_{j-1}, b_j)} \Big[\frac{i-1}{n}, \frac{i}{n}\Big)\Big).$$

Moreover, we can also conclude:

$$(b_j - b_{j-1}) - \frac{2}{n} \leq \frac{r_j^n - (l_j^n - 1)}{n} \leq (b_j - b_{j-1})$$

$$\Rightarrow \qquad r_j^n - l_j^n + 1 \;\geq n\,(b_j - b_{j-1}) - 2 \quad \text{and} \quad \frac{r_j^n - (l_j^n - 1)}{n} \underset{n \to \infty}{\longrightarrow} (b_j - b_{j-1}). \tag{2.30}$$

According to (2.29), the sum $S_n^{(1)}$ defined in (2.27), can be rewritten as follows:

$$S_n^{(1)} = \sum_{j=1}^{m} \Bigg(\sum_{i=l_j^n}^{r_j^n} \int\limits_{[\frac{i-1}{n},\, \frac{i}{n})} \Phi\big(\tilde{g}_j',\, \xi_i^n + \tilde{f}_j'\big)\, du \Bigg) = \sum_{j=1}^{m} \Bigg(\sum_{i=l_j^n}^{r_j^n} \frac{1}{n} \cdot \Phi\big(\tilde{g}_j',\, \xi_i^n + \tilde{f}_j'\big) \Bigg)$$

$$= \sum_{j=1}^{m} \Bigg(\frac{r_j^n - (l_j^n - 1)}{n} \sum_{i=l_j^n}^{r_j^n} \frac{1}{r_j^n - (l_j^n - 1)} \cdot \Phi\big(\tilde{g}_j',\, \xi_i^n + \tilde{f}_j'\big) \Bigg). \tag{2.31}$$

Due to this, on the way to prove (2.22), our next aim is to show that for all $j \in \{1, \dots, m\}$ and all corresponding intervals $[b_{j-1}, b_j)$ it holds:

$$\lim_{n\to\infty} \Big(\frac{1}{r_j^n - l_j^n + 1} \cdot \sum_{i=l_j^n}^{r_j^n} \Phi\big(\tilde{g}_j', \xi_i^n + \tilde{f}_j'\big)\Big) = \mathbb{E}\big(\Phi\big(\tilde{g}_j', \xi + \tilde{f}_j'\big)\big) \;\; \mathbb{P} - \text{a.s.} \tag{2.32}$$

This means that we need to show the validity of the following relation:

$$\mathbb{P}\Bigg(\forall j \in \{1, \dots, m\}, \lim_{n\to\infty} \Big(\sum_{i=l_j^n}^{r_j^n} \frac{1}{r_j^n - l_j^n + 1} \cdot \Phi\big(\tilde{g}_j', \xi_i^n + \tilde{f}_j'\big) - \mathbb{E}\big(\Phi\big(\tilde{g}_j', \xi + \tilde{f}_j'\big)\big)\Big) \neq 0 \Bigg) = 0.$$

To validate the last relation, using the well known Borel-Cantelli lemma (see e.g. Fisz [14, p. 266]), it is sufficient to show that for any $\varepsilon > 0$ it holds:

$$\sum_{n=1}^{\infty} \sum_{j=1}^{m} \mathbb{P}\Bigg(\Big| \sum_{i=l_j^n}^{r_j^n} \frac{1}{r_j^n - l_j^n + 1} \cdot \Phi\big(\tilde{g}_j', \xi_i^n + \tilde{f}_j'\big) - \mathbb{E}\big(\Phi\big(\tilde{g}_j', \xi + \tilde{f}_j'\big)\big) \Big| \geq \varepsilon \Bigg) < \infty. \tag{2.33}$$

In order to prove this, we consider the left hand side of (2.33). This quantity can be estimated by means of the inequality that directly follows from the well-known Hoeffding's inequality (see [33, p. 58]) and is considered in details in Appendix A, Lemma 22, p. 121.

$$\begin{aligned}
& \sum_{n=1}^{\infty} \sum_{j=1}^{m} \mathbb{P}\Bigg(\Big| \sum_{i=l_j^n}^{r_j^n} \frac{1}{r_j^n - l_j^n + 1} \cdot \Phi\big(\tilde{g}_j', \xi_i^n + \tilde{f}_j'\big) - \mathbb{E}\big(\Phi\big(\tilde{g}_j', \xi + \tilde{f}_j'\big) \Big| \geq \varepsilon \Bigg) \\
= \; & \sum_{n=1}^{\infty} \sum_{j=1}^{m} \mathbb{P}\Bigg(\Big| \sum_{i=l_j^n}^{r_j^n} \Phi\big(\tilde{g}_j', \xi_i^n + \tilde{f}_j'\big) - \sum_{i=l_j^n}^{r_j^n} \mathbb{E}\big(\Phi\big(\tilde{g}_j', \xi + \tilde{f}_j'\big)\big) \Big| \geq (r_j^n - l_j^n + 1) \cdot \varepsilon \Bigg) \\
= \; & \sum_{n=1}^{\infty} \sum_{j=1}^{m} \mathbb{P}\Bigg(\Big| \sum_{i=l_j^n}^{r_j^n} \Phi\big(\tilde{g}_j', \xi_i^n + \tilde{f}_j'\big) - \mathbb{E}\Big(\sum_{i=l_j^n}^{r_j^n} \Phi\big(\tilde{g}_j', \xi + \tilde{f}_j'\big)\Big) \Big| \geq (r_j^n - l_j^n + 1) \cdot \varepsilon \Bigg) \\
\underset{(A.22),(2.5)}{\leq} \; & \sum_{n=1}^{\infty} \sum_{j=1}^{m} 2 \exp\left\{ - \frac{2 \cdot (r_j^n - l_j^n + 1)^2 \cdot \varepsilon^2}{\sum\limits_{k=l_j^n}^{r_j^n} \big(|\tilde{g}_j'| - (-|\tilde{g}_j'|)\big)^2} \right\} \\
= \; & \sum_{n=1}^{\infty} \sum_{j=1}^{m} 2 \exp\left\{ - \frac{2 \cdot (r_j^n - l_j^n + 1)^2 \cdot \varepsilon^2}{(2\tilde{g}_j')^2 \, (r_j^n - l_j^n + 1)} \right\} \\
= \; & \sum_{n=1}^{\infty} \sum_{j=1}^{m} 2 \exp\left\{ - \frac{(r_j^n - l_j^n + 1) \cdot \varepsilon^2}{2 \cdot {\tilde{g}_j'}^2} \right\} \\
\underset{(2.30)}{\leq} \; & \sum_{n=1}^{\infty} \sum_{j=1}^{m} 2 \exp\left\{ - \frac{((b_j - b_{j-1}) \cdot n - 2) \cdot \varepsilon^2}{2 \cdot {\tilde{g}_j'}^2} \right\} \\
= \; & \sum_{j=1}^{m} 2 \exp\left\{ \frac{\varepsilon^2}{(\tilde{g}_j')^2} \right\} \sum_{n=1}^{\infty} \exp\left\{ - \frac{\varepsilon^2 (b_j - b_{j-1}) \cdot n}{2 \cdot {\tilde{g}_j'}^2} \right\}.
\end{aligned}$$

By definition, the number m is the number of jumps of the function $(\tilde{f} - \tilde{g})$, i.e. $m < \infty$. This means that the first sum in the last equation has the finite number of summands, therefore the series $\sum\limits_{j=1}^{m} 2 \exp\Big\{\frac{\varepsilon^2}{(\tilde{g}_j')^2}\Big\} \sum\limits_{n=1}^{\infty} \exp\Big\{-\frac{\varepsilon^2(b_j - b_{j-1}) \cdot n}{2 \cdot \tilde{g}_j'^2}\Big\}$ is finite, because for any $j \in \{1, \dots, m\}$ the sum $\sum\limits_{n=1}^{\infty} \exp\Big\{-\frac{\varepsilon^2(b_j - b_{j-1})\, n}{2 \cdot \tilde{g}_j'^2}\Big\}$ is a convergent geometric series of the form $S = \sum\limits_{n=0}^{\infty} a_0 \cdot q^n$ with $a_0 := \exp\Big\{-\frac{\varepsilon^2(b_j - b_{j-1})}{2 \cdot \tilde{g}_j'^2}\Big\}$ and $q := \exp\Big\{-\frac{\varepsilon^2(b_j - b_{j-1})}{2 \cdot \tilde{g}_j'^2}\Big\} \leq 1$.

Thus we obtain that (2.33) is valid, which allows us to deduce the validity of (2.32).

According to (2.31), (2.30) and (2.32), we get

$$S_n^{(1)} \underset{n\to\infty}{\overset{P-a.s.}{\longrightarrow}} \sum_{j=1}^{m} (b_j - b_{j-1}) \cdot \mathbb{E}\big(\Phi\big(\tilde{g}_j', \xi + \tilde{f}_j'\big)\big) = \sum_{j=1}^{m} \int\limits_{[b_{j-1}, b_j)} \mathbb{E}\big(\Phi\big(\tilde{g}_j', \xi + \tilde{f}_j'\big)\big) du$$
$$= \sum_{j=1}^{m} \int\limits_{[b_{j-1}, b_j)} \mathbb{E}\big(\Phi\big(\tilde{g}(u), \xi + \tilde{f}(u)\big)\big) du = \int \mathbb{E}\big(\Phi\big(\tilde{g}(u), \xi + \tilde{f}(u)\big)\big) du.$$

This completes the proof of (2.22) and therefore of Lemma 1.

□

In the next lemma, we generalize the results of Lemma 1 to the case of arbitrary functions from $L^1\big([0,1)\big)$.

Lemma 2. *Let $(\xi_i^n)_{1 \leq i \leq n,\, n \in \mathbb{N}}$ fulfill Condition 1. Then for any sequence $(g_n)_{n\in\mathbb{N}} \subset L^1\big([0,1)\big)$ with $g_n \underset{n\to\infty}{\longrightarrow} g \in L^1\big([0,1)\big)$ and any sequence $(f_n)_{n\in\mathbb{N}} \subset L^1\big([0,1)\big)$ with $f_n \underset{n\to\infty}{\longrightarrow} f \in L^1\big([0,1)\big)$ the following result is valid:*

$$\|\xi^n + f_n - g_n\| - \|\xi^n + f_n\| \underset{n\to\infty}{\overset{\mathbb{P}-a.s.}{\longrightarrow}} \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du \tag{2.34}$$

Proof:
Let $(g_n)_{n\in\mathbb{N}}$ and $(f_n)_{n\in\mathbb{N}}$ be two arbitrary sequences of functions from $L^1\big([0,1)\big)$ such that $\lim\limits_{n\to\infty} g_n = g$, $g \in L^1\big([0,1)\big)$ and $\lim\limits_{n\to\infty} f_n = f$, $f \in L^1\big([0,1)\big)$. This means that

$$\forall\, \varepsilon > 0 \;\; \exists\, N_0 \, : \, \forall\, n \geq N_0, \;\; \|f_n - f\| < \frac{\varepsilon}{18} \;\; \text{and} \;\; \|g_n - g\| < \frac{\varepsilon}{18}\,. \tag{2.35}$$

Moreover it is known that the set $S\big([0,1)\big)$ is dense in $L^1\big([0,1)\big)$ (see Kolmogorov [27, p. 432]), i.e. for any $\varepsilon > 0$ there exist such functions $\tilde{f} \in S\big([0,1)\big)$ and $\tilde{g} \in S\big([0,1)\big)$ that it holds

$$\|\tilde{f} - f\| < \frac{\varepsilon}{18} \quad \text{and} \quad \|\tilde{g} - g\| < \frac{\varepsilon}{18}. \tag{2.36}$$

From inequalities (2.35) and (2.36) follows that for all $n \geq N_0$ we have:

$$\|f_n - \tilde{f}\| \leq \|f_n - f\| + \|\tilde{f} - f\| < \frac{2\varepsilon}{18} \tag{2.37}$$

and similarly

$$\|g_n - \tilde{g}\| < \frac{2\varepsilon}{18}. \tag{2.38}$$

From the well-known generalized triangle inequality:

$$\big|(a-b-c)-(a'-b'-c')\big| \leq |a-a'|+|b-b'|+|c-c'|,$$

and generalized reverse triangle inequality:

$$\big|\,|a-b-c|-|a'-b'-c'|\,\big| \leq \big|(a-b-c)-(a'-b'-c')\big|$$

it follows that for all $a, b, c \in \mathbb{R}$ the following inequality is valid:

$$\big|\,|a-b-c|-|a'-b'-c'|\,\big| \leq |a-a'|+|b-b'|+|c-c'|. \tag{2.39}$$

Using inequality (2.39) and the following well-known inequalities

$$\big|\;\|f'\|-\|g'\|\;\big| \leq \|f'-g'\|, \quad f', g' \in L^1([0,1)) \tag{2.40}$$

$$|a|-|b| \leq |a-b| \leq |a|+|b|, \quad a, b \in \mathbb{R} \tag{2.41}$$

for any $n \geq N_0$ we obtain:

$$\begin{aligned}
&\Bigg|\;\Big|\,\|\xi^n+f_n-g_n\|-\|\xi^n+f_n\|-\int \mathbb{E}\big(|\xi+f(u)-g(u)|-|\xi+f(u)|\big)\,du\,\Big| \\
&\quad-\Big|\,\|\xi^n+\tilde{f}-\tilde{g}\|-\|\xi^n+\tilde{f}\|-\int \mathbb{E}\big(|\xi+\tilde{f}(u)-\tilde{g}(u)|-|\xi+\tilde{f}(u)|\big)\,du\,\Big|\;\Bigg| \\
\underset{(2.39)}{\leq}\;& \Big|\,\|\xi^n+f_n-g_n\|-\|\xi^n+\tilde{f}-\tilde{g}\|\,\Big|+\Big|\,\|\xi^n+f_n\|-\|\xi^n+\tilde{f}\|\,\Big| \\
+\;& \Big|\int \mathbb{E}\big(|\xi+f(u)-g(u)|-|\xi+f(u)|\big)\,du \\
&\quad-\int \mathbb{E}\big(|\xi+\tilde{f}(u)-\tilde{g}(u)|-|\xi+\tilde{f}(u)|\big)\,du\,\Big| \\
\underset{(2.40)}{\leq}\;& \|f_n-g_n-\tilde{f}+\tilde{g}\|+\|f_n-\tilde{f}\| \\
+\;& \int \mathbb{E}\big(\,|\xi+f(u)-g(u)|-|\xi+f(u)|-|\xi+\tilde{f}(u)-\tilde{g}(u)|+|\xi+\tilde{f}(u)|\,\big)\,du \\
\underset{(2.41)}{\leq}\;& 2\,\|f_n-\tilde{f}\|+\|\tilde{g}-g_n\| \\
+\;& \int \mathbb{E}\big(\,|\,\xi+f(u)-g(u)-\xi-\tilde{f}(u)+\tilde{g}(u)\,|+|\,\xi+\tilde{f}(u)-\xi-f(u)\,|\,\big)\,du \\
\leq\;& 2\,\|f_n-\tilde{f}\|+\|\tilde{g}-g_n\|+2\int|\tilde{f}(u)-f(u)|\,du+\int|\tilde{g}(u)-g(u)|\,du \\
=\;& 2\,\|f_n-\tilde{f}\|+\|\tilde{g}-g_n\|+2\,\|\tilde{f}-f\|+\|\tilde{g}-g\| < \frac{4\,\varepsilon}{18}+\frac{2\,\varepsilon}{18}+\frac{2\,\varepsilon}{18}+\frac{\varepsilon}{18} \;=\; \frac{\varepsilon}{2}.
\end{aligned}$$

Thus for any $\varepsilon > 0$ and $n \geq N_0$ we have shown that

$$\left| \; \left| \, \|\xi^n + f_n - g_n\| - \|\xi^n + f_n\| - \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du \,\right| \right. \\ \left. - \left| \, \|\xi^n + \tilde{f} - \tilde{g}\| - \|\xi^n + \tilde{f}\| - \int \mathbb{E}\big(|\xi + \tilde{f}(u) - \tilde{g}(u)| - |\xi + \tilde{f}(u)|\big)\, du \,\right| \; \right| < \frac{\varepsilon}{2}. \tag{2.42}$$

Moreover, from Lemma 1 we know that

$$\|\xi^n + \tilde{f} - \tilde{g}\| - \|\xi^n + f_n\| \overset{\mathbb{P}-\text{a.s.}}{\underset{n\to\infty}{\longrightarrow}} \int \mathbb{E}\big(|\xi + \tilde{f}(u) - \tilde{g}(u)| - |\xi + \tilde{f}(u)|\big)\, du.$$

This means, that for any $\epsilon > 0$ there exists N_1 such that for any $n \geq N_1$ it holds $\mathbb{P}$-almost surely:

$$\left| \, \|\xi^n + \tilde{f} - \tilde{g}\| - \|\xi^n + \tilde{f}\| - \int \mathbb{E}\big(|\xi + \tilde{f}(u) - \tilde{g}(u)| - |\xi + \tilde{f}(u)|\big)\, du \,\right| < \frac{\varepsilon}{2}. \tag{2.43}$$

Finally, taking (2.42) and (2.43) into account, for any $\epsilon > 0$ and all $n \geq \max\{N_0, N_1\}$ we obtain:

$$\begin{aligned} & \left| \, \|\xi^n + f_n - g_n\| - \|\xi^n + f_n\| - \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du \,\right| \\ \leq \; & \left| \; \left| \, \|\xi^n + f_n - g_n\| - \|\xi^n + f_n\| - \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du \,\right| \right. \\ & \left. - \left| \, \|\xi^n + \tilde{f} - \tilde{g}\| - \|\xi^n + \tilde{f}\| - \int \mathbb{E}\big(|\xi + \tilde{f}(u) - \tilde{g}(u)| - |\xi + \tilde{f}(u)|\big)\, du \,\right| \; \right| \\ + \; & \left| \, \|\xi^n + \tilde{f} - \tilde{g}\| - \|\xi^n + \tilde{f}\| - \int \mathbb{E}\big(|\xi + \tilde{f}(u) - \tilde{g}(u)| - |\xi + \tilde{f}(u)|\big)\, du \,\right| \\ < \; & \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \qquad \mathbb{P} - \text{a.s.} \end{aligned}$$

From this we deduce:

$$\|\xi^n + f_n - g_n\| - \|\xi^n + f_n\| \overset{\mathbb{P}-\text{a.s.}}{\underset{n\to\infty}{\longrightarrow}} \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du,$$

what completes the proof of Lemma 2.

□

Summarizing, we have studied the limiting behaviour of the non-penalized functional, i.e. the functional with the omitted penalty term, as n approaches infinity. In order to be able to conclude about the convergence of the full (penalized) functional $\tilde{H}^n_{\gamma_n}$, one needs to study the limiting behaviour of the penalty term itself, i.e. the term, which is proportional to the number of jumps $\#J(g)$ of a step function $g \in S\big([0,1)\big)$. Recalling Definition 1 of epi-convergence, we see that this task reduces to determining the inferior and the superior limits of the functional $\#J$. One of the most important concepts of the functional analysis, which is closely related to the inferior limits of a functional, is the concept of the lower semicontinuity. The next subsection is dedicated to studying the lower semicontinuity of the functional $\#J$.

2.1.2 Lower semicontinuity of the penalty term

We start with the detailed definition of lower semicontinuity, as well as with several auxiliary definitions.

Definition 2. *Let* (Θ, ρ) *be a metric space. We say that a functional* $F : \Theta \to \mathbb{R} \cup \{\infty\}$ *is lower semicontinuous with respect to* ρ*-convergence, if for all sequences of functions* $(g_n)_{n\in\mathbb{N}} \subset \Theta$ *such that* $g_n \xrightarrow[n\to\infty]{\rho} g$ *it holds:*

$$\lim_{n\to\infty} \inf F(g_n) \geq F(g). \tag{2.44}$$

Equivalently this can be expressed as follows:

Definition 3. *The functional* $F : \Theta \to \mathbb{R} \cup \{\infty\}$ *is lower semicontinuous with respect to* ρ*-convergence, if for all* $a \in \mathbb{R}$ *the set*

$$I_a := \big\{g \in \Theta : \, F(g) \leq a\big\} \tag{2.45}$$

is closed, i.e. for any sequence $(g_n)_{n\in\mathbb{N}} \subset I_a$ *such that* $g_n \xrightarrow[n\to\infty]{\rho} g$ *the limit function* g *is an element of* I_a.

These two definitions as well as the proof of their equivalence can be found for example in Rockafellar [36] .
Before proceeding, we remind, that $L^0([0,1))$ denotes the space of measurable functions on $[0,1)$ [24] and a sequence of functions $(g_n)_{n\in\mathbb{N}} \subset L^0([0,1))$ is said to converge in measure or in L^0 to some limiting function g, if for any $\varepsilon > 0$ it holds:

$$l\Big(\,\{\, x \in [0,1) \,:\, |g_n(x) - g(x)| > \varepsilon \,\}\,\Big) \xrightarrow[n\to\infty]{} 0, \tag{2.46}$$

where l is Lebesgue measure.
The goal of this subsection is to show that the functional $\#J : L^0([0,1)) \to \mathbb{N} \cup \{0,\infty\}$, which is defined in the following way:

$$\#J(g) = \begin{cases} \text{Number of Jumps of } g & : \quad g \in S\big([0,1)\big) \\ \infty & : \quad \text{otherwise} \end{cases} \tag{2.47}$$

is lower semicontinuous.
For our purposes it is sufficient to show that the functional $\#J$ is lower semicontinuous with respect to the L^1-convergence. However, it is clear that each function from $L^1([0,1))$ is a measurable functions on $[0,1)$, implying that $L^1([0,1)) \subset L^0([0,1))$. It is also known that if a sequence of functions is convergent in L^0, then this sequence is also convergent in L^1. Therefore, it is clear that if we show that $\#J$ is lower semicontinuios with respect to L^0- convergence, then $\#J$ is automatically lower semicontinuios with respect to L^1-convergence.
Next, we give the definition of the Hausdorff metric ρ_H on the space $CL([0,1])$ of closed subsets contained in $[0,1]$.

Definition 4. *[8] For nonempty closed sets* $A, B \subset [0,1]$ *the Hausdorff metric* ρ_H *is defined as follows:*

$$\rho_H(A,B) = \max\Big\{\, \max_{x\in A} \min_{y\in B} |x-y|,\ \max_{y\in B} \min_{y\in A} |x-y| \,\Big\}. \tag{2.48}$$

and $\rho_H(\emptyset, B) = \rho_H(A, \emptyset) = 1$.

Now we have all the information that we need in order to prove the lower semicontiniuty of $\#J$.

Lemma 3. *The functional $\#J$, as defined in (2.47), is lower semicontinuous with respect L^0- convergence.*

Proof:
Fix some finite number $N \in \mathbb{N} \cup \{0\}$ and consider an arbitrary sequence $(g_n)_{n\in\mathbb{N}}$ from the set

$$I_N := \big\{ g \in S\big([0,1)\big) \,:\, \#J(g) \leq N \big\}.$$

Additionally, assume that the sequence $(g_n)_{n\in\mathbb{N}}$ converges to some function g with respect to the L^0-convergence, i.e. $g_n \underset{n\to\infty}{\overset{L^0([0,1))}{\longrightarrow}} g$. In order to prove Lemma 3 by using Definition 3 of the lower semicontinuity, we need to show that the function g is also a piecewise constant function with at most N number of jumps, i.e. $\#J(g) \leq N$.
For any $n \in \mathbb{N}$ the set $J(g_n)$ is a finite set of points from the interval $(0,1)$, or in other words, it is a finite subset of $(0,1)$ and therefore, automatically is a closed subset of $[0,1]$, i.e. $\#J(g_n) \in CL\big([0,1]\big)$. It is known (see [7] or [28]) that the set $CL\big([0,1]\big)$ is compact, hence there exists a subsequence $(g_{n_k}) \subseteq (g_n)_{n\in\mathbb{N}}$ and a closed set $J \subset (0,1)$ such that

$$\rho_H\big(\, J(g_{n_k}) \cup \{0,1\},\, J \cup \{0,1\} \,\big) \underset{n\to\infty}{\longrightarrow} 0. \tag{2.49}$$

Due to the facts that $\#J(g_{n_k}) \leq N$ for any $n_k \in \mathbb{N}$ and that the cardinality is lower semicontinuous with respect to the Hausdorff metric ρ_H (see [7]) we conclude that

$$\#J \leq N. \tag{2.50}$$

Now we return to relation (2.49).
Using (2.49) and the definition of the limit, we obtain the following result:

$$\forall\, \varepsilon > 0 \;\; \exists\, N_0 \,:\, \forall\, n_k \geq N_0 \quad \rho_H\big(\, J(g_{n_k}) \cup \{0,1\},\, J \cup \{0,1\} \,\big) < \frac{\varepsilon}{2}$$

According to Definition 4, this implies that for any $\varepsilon > 0$ there exists such a natural number N_0 that for all $n_k \geq N_0$ it holds:

$$\max\Big\{ \max_{x\in J(g_{n_k})\cup\{0,1\}} \; \min_{y\in J\cup\{0,1\}} |x-y|,\; \max_{y\in J\cup\{0,1\}} \; \min_{x\in J(g_{n_k})\cup\{0,1\}} |x-y| \;\Big\} < \frac{\varepsilon}{2}. \tag{2.51}$$

The inequality (2.51) can be rewritten as follows:

$$\forall\, x \in J(g_{n_k}) \cup \{0,1\} \quad \text{it holds :} \; \min_{y\in J\cup\{0,1\}} |x-y| < \frac{\varepsilon}{2}$$

and

$$\forall\, y \in J \cup \{0,1\} \quad \text{it holds :} \; \min_{x\in J(g_{n_k})\cup\{0,1\}} |x-y| < \frac{\varepsilon}{2}.$$

These properties imply that for any $\varepsilon > 0$ there exits such a number $N_0 \in \mathbb{N}$ that for all $n_k \geq N_0$ it holds:

$$\forall\, x \in J(g_{n_k}) \cup \{0,1\} \quad \exists\, y' \in J \cup \{0,1\} : \;\; |x - y'| < \varepsilon$$

and conversely :

$$\forall\, y \in J \cup \{0,1\} \quad \exists\, x' \in J(g_{n_k}) \cup \{0,1\} : \quad |x' - y| < \varepsilon.$$

This also means that if for some interval $(s,t) \subseteq [0,1)$ the intersection with J is empty, i.e. if

$$(s,t) \cap J = \emptyset,$$

then for any $\varepsilon > 0$ there exits such a number $N_0 \in \mathbb{N}$ that for all $n_k \geq N_0$ it holds :

$$(s+\varepsilon,\, t-\varepsilon) \cap \big\{ J(g_{n_k}) \cup \{0,1\} \big\} = \emptyset.$$

We recall that for any $n_k \in \mathbb{N}$, $J(g_{n_k})$ is the set of jump points of the piecewise constant function g_{n_k}, which means that from the fact $(s+\varepsilon,\, t-\varepsilon) \cap \big\{ J(g_{n_k}) \cup \{0,1\} \big\} = \emptyset$ it follows that the function g_{n_k} is constant on the interval $(s+\varepsilon,\, t-\varepsilon)$. We know that the function g is the limiting function of the sequence $(g_n)_{n\in\mathbb{N}}$ with respect to the L^0- convergence. Moreover, due to Lemma 27 (see Appendix A, p.124) and the definition of the limit, we deduce that for any subsequence $(g_{n_k})_{n_k\in\mathbb{N}} \subseteq (g_n)_{n\in\mathbb{N}}$, any interval $(s,\, t) \subset [0,1)$ and any $\varepsilon > 0$ it holds: $g_{n_k} \cdot \mathbf{1}_{(s+\varepsilon,\, t-\varepsilon)} \xrightarrow[n\to\infty]{L^0([0,1))} g \cdot \mathbf{1}_{(s+\varepsilon,\, t-\varepsilon)}$. From this we see that if for any $\varepsilon > 0$ there exits such a number $N_0 \in \mathbb{N}$ that for all $n_k \geq N_0$ the function g_{n_k} is constant on some interval $(s+\varepsilon,\, t-\varepsilon)$, then the function g is also constant on the same interval $(s+\varepsilon,\, t-\varepsilon)$. Finally we conclude that the function g is constant on the interval $(s+\varepsilon,\, t-\varepsilon)$ for any $\varepsilon > 0$ and any interval $(s,t) \subseteq [0,1)$ with the property $(s,t) \cap J = \emptyset$, where J is a set of points from $[0,1)$ such that $\#J \leq N$ (see (2.50)). Since ε is arbitrary, we arrive at the conclusion that g is constant on all intervals (s,t) satisfying $(s,t) \cap J = \emptyset$. Consequently, the function g is a piecewise constant function with

$$J(g) \subseteq J$$

and

$$\#J(g) \leq \#J \leq N.$$

This completes the proof of Lemma 3.

□

2.1.3 Linking $S\big([0,1)\big)$, $L^1\big([0,1)\big)$ and $S_n\big([0,1)\big)$

The final auxiliary result, which is required in order to prove the epi-convergence of the functionals under consideration, is related to the possibility to approximate an arbitrary function from $S\big([0,1)\big)$ or from $L^1\big([0,1)\big)$ by functions from $S_n\big([0,1)\big)$ with a given number of jumps.

Lemma 4. **(i)** *For each function $g \in S\big([0,1)\big)$ there exists a sequence of functions $\big(g_n\big)_{n\in\mathbb{N}} \subset S_n\big([0,1)\big)$ such that $\#J(g_n) \leq \#J(g)$ for all $n \in \mathbb{N}$ and*

$$\|g_n - g\| \underset{n\to\infty}{\longrightarrow} 0. \tag{2.52}$$

(ii) *For each function $g \in L^1\big([0,1)\big)$ and any infinite sequence of positive numbers $(l_n)_{n\in\mathbb{N}}$, i.e. $l_n \underset{n\to\infty}{\longrightarrow} \infty$, there exists a sequence of step functions $(g_n)_{n\in\mathbb{N}} \subset S_n\big([0,1)\big)$ such that $\#J(g_n) \leq l_n$ for all $n \in \mathbb{N}$ and*

$$\|g_n - g\| \underset{n\to\infty}{\longrightarrow} 0\,. \tag{2.53}$$

Proof:
It is known that the set $S\big([0,1)\big)$ is dense in $L^1\big([0,1)\big)$ [27, p.432], this means that for each function $g \in L^1\big([0,1)\big)$ and any $\varepsilon > 0$ there exists the natural number $k_\varepsilon < \infty$ and a step function $g_{k_\varepsilon} \in S\big([0,1)\big)$ with k_ε number of steps :

$$g_{k_\varepsilon} = \sum_{i=1}^{k_\varepsilon} g_{(k_\varepsilon)_i} \cdot \mathbf{1}_{[a_{i-1},\, a_i)}, \quad J(g_{k_\varepsilon}) = \big\{a_1, \ldots, a_{k_\varepsilon - 1}\big\} \subset (0,1),\ a_0 := 0,\ a_{k_\varepsilon} := 1, \quad (2.54)$$

such that

$$\|g_{k_\varepsilon} - g\| < \frac{\varepsilon}{2}. \quad (2.55)$$

For all $n \in \mathbb{N}$ we take the following sequence of sets :

$$S_{n,\, k_\varepsilon}\big([0,1)\big) := \Big\{ g \in S_n\big([0,1)\big) : \#J(g) \le \#J(g_{k_\varepsilon}) \Big\}. \quad (2.56)$$

For any n we define a natural number

$$u_{(n,k_\varepsilon)} = \min\{n,\ k_\varepsilon\},$$

and points $t_0^n \le t_1^n \le \ldots \le t^n_{u_{(n,k_\varepsilon)}}$ as a points which satisfy the following property:

1) $t_0^n := 0,\ t^n_{u_{(n,\, k_\varepsilon)}} := 1,$ and $\forall\, i \in \{1, \ldots, u_{(n,k_\varepsilon)} - 1\},\ \ t_i^n \in (0,1),$

2) $\forall\, i \in \{1, \ldots, u_{(n,k_\varepsilon)} - 1\},\ \ n \cdot t_i^n \in \mathbb{N},$ i.e. $t_i^n = \dfrac{j}{n}$ for some $j \in \mathbb{N},\ j < n,$

3) $i \in \{1, \ldots, u_{(n,k_\varepsilon)}\}$ it is valid $0 \le (t_i - a_i) < \dfrac{1}{n}.$

Now we construct the following sequence of functions $\big(g_{n,\, k_\varepsilon}\big)_{n \in \mathbb{N}} \subset S_{n,\, k_\varepsilon}\big([0,1)\big)$:

$$g_{n,\, k_\varepsilon} := \sum_{i=1}^{u_{(n,k_\varepsilon)}} g_{k_\varepsilon}(t_{i-1}^n) \cdot \mathbf{1}_{[t_{i-1}^n,\, t_i^n)}. \quad (2.57)$$

Let the length of the shortest step of g_{k_ε} be $\triangle_{\min}(g_{k_\varepsilon})$. It is clear that $\triangle_{\min}(g_{k_\varepsilon}) < \dfrac{1}{k_\varepsilon}$.
If we define the natural number n_{k_ε} as the next smallest integer to the number $\dfrac{1}{\triangle_{\min}(g_{k_\varepsilon})}$, i.e. $n_{k_\varepsilon} = \left\lceil \dfrac{1}{\triangle_{\min}(g_{k_\varepsilon})} \right\rceil$, then for all $n \ge n_{k_\varepsilon}$ the function $g_{n,\, k_\varepsilon}$ can be rewritten as follows:

$$g_{n,\, k_\varepsilon} = \sum_{i=1}^{k_\varepsilon} g_{(k_\varepsilon)_i} \cdot \mathbf{1}_{[t_{i-1}^n,\, t_i^n)}. \quad (2.58)$$

In this case the function $g_{n,\, k_\varepsilon}$ is the the right-shift step function g_{k_ε} (Figure 2.3 represents an example of $g_{n,\, k_\varepsilon}$). Now for any $n \ge n_{k_\varepsilon}$ we consider the number $\|g_{n,\, k_\varepsilon} - g_{k_\varepsilon}\|$:

$$\|g_{n,\, k_\varepsilon} - g_{k_\varepsilon}\| = \int \Big| \sum_{i=1}^{k_\varepsilon - 1} g_{(k_\varepsilon)_i} \cdot \mathbf{1}_{[t_{i-1}^n,\, t_i^n)}(u) - \sum_{i=1}^{k_\varepsilon - 1} g_{(k_\varepsilon)_i} \cdot \mathbf{1}_{[a_{i-1},\, a_i)}(u) \Big| du$$

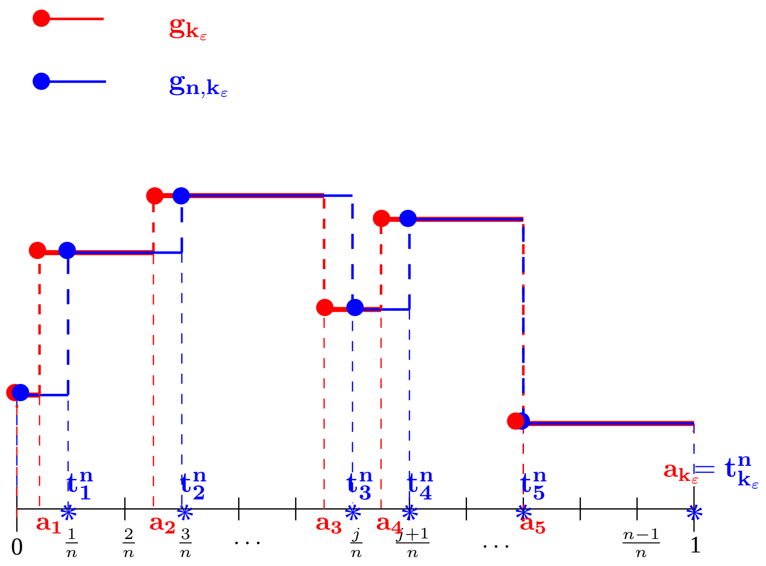

Figure 2.3: example of $g_{n,\,k_\varepsilon}$

$$= \int \sum_{i=1}^{k_\varepsilon - 1} |g_{(k_\varepsilon)_i} - g_{(k_\varepsilon)_{i+1}}| \cdot \mathbf{1}_{[a_i,\, t_i^n)}(u)\, du$$

$$= \sum_{i=1}^{k_\varepsilon - 1} \int\limits_{[a_i,\, t_i^n)} \sum_{i=1}^{k_\varepsilon} |g_{(k_\varepsilon)_i} - g_{(k_\varepsilon)_{i+1}}| \cdot \mathbf{1}_{[a_i,\, t_i^n)}(u)\, du = \sum_{i=1}^{k_\varepsilon - 1} \int\limits_{[a_i,\, t_i^n)} |g_{(k_\varepsilon)_i} - g_{(k_\varepsilon)_{i+1}}|\, du$$

$$= \sum_{i=1}^{k_\varepsilon - 1} |g_{(k_\varepsilon)_i} - g_{(k_\varepsilon)_{i+1}}| \cdot (t_i^n - a_i) \quad < \quad \frac{k_\varepsilon}{n} \cdot \max_{1 \le i \le k_\varepsilon - 1} |g_{(k_\varepsilon)_i} - g_{(k_\varepsilon)_{i+1}}|.$$

Since $k_\varepsilon < \infty$, $\max\limits_{1 \le i \le k_\varepsilon - 1} |g_{(k_\varepsilon)_{i+1}} - g_{(k_\varepsilon)_i}| < \infty$ and $\forall\, \varepsilon > 0\ \exists\, n_0 :\ \forall\, n \ge n_0\quad \dfrac{1}{n} < \dfrac{\varepsilon}{2}$, then for any $\varepsilon > 0$ and any $n \ge N_{k_\varepsilon} = \max\{n_0,\, n_{k_\varepsilon}\}$ we deduce that

$$\|g_{n,\,k_\varepsilon} - g_{k_\varepsilon}\| < \frac{\varepsilon}{2}. \tag{2.59}$$

The function g_{k_ε} is, in principle, an arbitrary function from the space $S\big([0,1)\big)$, therefore statement **(i)** of Lemma 4 is proved.

Now for any sequence of positive numbers $(l_n)_{n\in\mathbb{N}}$ such that $l_n \underset{n\to\infty}{\longrightarrow} \infty$ we consider the set

$$S_{n,\,l_n}\big([0,1)\big) := \Big\{g \in S_n\big([0,1)\big) : \#J(g) \le l_n\Big\}.$$

It is clear that because $l_n \underset{n\to\infty}{\longrightarrow} \infty$, then for all $k_\varepsilon < \infty$ there exists a number N_0 such that $k_\varepsilon < l_n$ for all $n \ge N_0$. From this it follows that for all $k_\varepsilon < \infty$ and for all $n \ge N_0$ the set $S_{n,\,k_\varepsilon}$, as defined in (2.56), is a subset of $S_{n,\,l_n}$ for all $n \ge N_0$ and consequently for any $g_{n,\,k_\varepsilon} \in S_{n,\,k_\varepsilon}$ it holds

$$\#J(g_{n,\,k_\varepsilon}) \le l_n.$$

Let $g \in L^1([0,1))$. For any $\varepsilon > 0$, the corresponding k_ε and $n \geq \max\{N_0, N_{k_\varepsilon}\}$ we consider for any $\|g_{n,k_\varepsilon} - g\|$:

$$\|g_{n,k_\varepsilon} - g\| \leq \|g_{n,k_\varepsilon} - g_{k_\varepsilon}\| + \|g_{k_\varepsilon} - g\| \underset{(2.55),(2.59)}{<} \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \tag{2.60}$$

This proves statement **(ii)** of Lemma 4.

□

2.1.4 Main result: epi-convergence

Theorem 1. *Let $(\xi_i^n)_{1\leq i\leq n, n\in\mathbb{N}}$ fulfill Condition 1, and $f \in L^1([0,1))$. Then for all sequences $(\gamma_n)_{n\in\mathbb{N}}$ with $\gamma_n \underset{n\to\infty}{\longrightarrow} \gamma$, $\gamma \geq 0$ it holds $\mathbb{P}$-almost surely*

$$\tilde{H}^n_{\gamma_n}(\cdot\,,\, f^n + \xi^n) \underset{n\to\infty}{\overset{epi}{\longrightarrow}} H^*_{\gamma,\xi}(\cdot\,,\, f). \tag{2.61}$$

The functionals $\tilde{H}^n_{\gamma_n}$ and $H^*_{\gamma,\xi}$ are defined by (2.1) and (2.21), respectively.

Proof:
First of all, for any $f \in L^1([0,1))$ we recall the definition of $\tilde{H}^n_\gamma(\cdot, f)$ and $H^*_{\gamma,\xi}(\cdot, f)$.

$$\tilde{H}^n_\gamma(g,f) := \begin{cases} \gamma\#J(g) + \|g-f\| - \|f\| & : \; g \in S_n([0,1)),\ \gamma > 0 \\ \|g-f\| - \|f\| & : \; g \in L^1([0,1)),\ \gamma = 0 \\ \infty & : \; \text{otherwise} \end{cases},$$

$$H^*_{\gamma,\xi}(g,f) := \begin{cases} \gamma\#J(g) + \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\,du & : \; g \in S([0,1)), \\ & \quad\; \gamma > 0 \\ \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\,du & : \; g \in L^1([0,1)), \\ & \quad\; \gamma = 0 \\ \infty & : \; \text{otherwise} \end{cases}$$

In order to prove this theorem we have to show that on a set with probability one it holds:

(i) If $g_n \underset{n\to\infty}{\longrightarrow} g$ then

$$\liminf_{n\to\infty} \tilde{H}^n_{\gamma_n}(g_n\,,\, f^n + \xi^n) \geq H^*_{\gamma,\xi}(g,\, f). \tag{2.62}$$

(ii) For all $g \in L^1([0,1))$ there exists $(g_n)_{n\in\mathbb{N}} \subset L^1([0,1))$ with $g_n \underset{n\to\infty}{\longrightarrow} g$ such that

$$\limsup_{n\to\infty} \tilde{H}^n_{\gamma_n}(g_n\,,\, f^n + \xi^n) \leq H^*_{\gamma,\xi}(g,\, f). \tag{2.63}$$

Proof of **(i)**:
Consider the following three cases:

1) If $g_n \notin S_n([0,1))$ and $\gamma_n > 0$ for infinitely many n, then according to (2.1), $\tilde{H}^n_{\gamma_n}(g_n\,,\, f^n + \xi^n) = \infty$ for infinitely many n. This implies that $\liminf\limits_{n\to\infty} \tilde{H}^n_{\gamma_n}(g_n\,,\, f^n + \xi^n) = \infty \geq H^*_{\gamma,\xi}(g,\, f)$, therefore relation **(i)** is valid.

2) If $g_n \in L^1\big([0,1)\big)$ and $\gamma_n = 0$ for infinitely many n, then $\lim\limits_{n\to\infty} \gamma_n = 0$, i.e. $\gamma = 0$.
From the definition of $\tilde{H}^n_\gamma$ and $H^*_{\gamma,\,\xi}$ we know:

$$\begin{aligned}
\tilde{H}^n_0(g_n,\, f^n + \xi^n) &= \|\xi^n + f^n - g_n\| - \|\xi^n + f^n\|, \;\; g_n \in L^1\big([0,1)\big),\\
H^*_{0,\,\xi}(g,\, f) &= \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du, \;\; g \in L^1\big([0,1)\big).
\end{aligned}$$

Let $g_n \underset{n\to\infty}{\longrightarrow} g$. Application of Lemma 2 yields:

$$\|\xi^n + f^n - g_n\| - \|\xi^n + f^n\| \underset{n\to\infty}{\overset{\mathbb{P}-\text{a.s.}}{\longrightarrow}} \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du.$$

This allows us to deduce:

$$\liminf_{n\to\infty} \tilde{H}^n_0(g_n\,,\, f^n + \xi^n) = H^*_{0,\,\xi}(g,\, f) \quad \mathbb{P} - \text{a.s.}$$

Hence, relation **(i)** in case 2) is also satisfied.

3) Now let $g_n \in S_n\big([0,1)\big)$ and $\gamma_n > 0$ for infinitely many n. Using the well-known inequality (superadditivity of limit inferior):

$$\liminf_{n\to\infty}\,(a_n + b_n) \geq \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n$$

and also Lemma 2 and 3 we obtain:

$$\begin{aligned}
& \liminf_{n\to\infty} \tilde{H}^n_{\gamma_n}(g_n\,,\, f^n + \xi^n)\\
\geq\;& \liminf_{n\to\infty} \gamma_n \# J(g_n) + \liminf_{n\to\infty} \big(\|\xi^n + f^n - g_n\| - \|\xi^n + f^n\|\big)\\
\underset{\mathbb{P}-\text{a.s.}}{\geq}\;& \gamma \# J(g) + \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du\\
=\;& H^*_{\gamma,\,\xi}(g,\, f).
\end{aligned}$$

Therefore **(i)** holds also in this case and, consequently, the relation **(i)** is proved.

Proof of **(ii)**:
Similarly to **(i)** we consider three different cases:

1) By definition, $H^*_{\gamma,\,\xi}(g,\, f) = \infty$ for any $\gamma > 0$ and $g \notin S_n\big([0,1)\big)$, this means that there is nothing to prove.

2) If $\gamma = 0$ and $g \notin S_n\big([0,1)\big)$ we have to show that for an arbitrary $g \in L^1\big([0,1)\big)$ there exists such sequence $(g_n)_{n\in\mathbb{N}} \subset L^1\big([0,1)\big)$ with $g_n \underset{n\to\infty}{\longrightarrow} g$ that

$$\limsup_{n\to\infty} \tilde{H}^n_{\gamma_n}(g_n\,,\, f^n + \xi^n) \leq \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du, \tag{2.64}$$

where $\gamma_n \underset{n\to\infty}{\longrightarrow} 0$.
By Lemma 4 **(ii)** we can choose g_n in this case as the best L^1-approximation of g in $S_n\big([0,1)\big)$ with at most $\dfrac{1}{\sqrt{\gamma_n}}$ jumps and claim that $\|g_n - g\| \underset{n\to\infty}{\longrightarrow} 0$.

Using the subadditivity of limit superior:

$$\limsup_{n\to\infty}(a_n+b_n)\le\limsup_{n\to\infty}a_n+\limsup_{n\to\infty}b_n, \tag{2.65}$$

we consider $\limsup_{n\to\infty}\tilde{H}^n_{\gamma_n}(g_n\,,\,f^n+\xi^n)$:

$$\begin{aligned}
&\limsup_{n\to\infty}\tilde{H}^n_{\gamma_n}(g_n\,,\,f^n+\xi^n)\\
\underset{(2.65)}{\le}\;&\limsup_{n\to\infty}\gamma_n\#J(g_n)+\limsup_{n\to\infty}\big(\|\xi^n+f^n-g_n\|-\|\xi^n+f^n\|\big)\\
\le\;&\limsup_{n\to\infty}\sqrt{\gamma_n}+\limsup_{n\to\infty}\big(\|\xi^n+f^n-g_n\|-\|\xi^n+f^n\|\big)\\
=\;&\int\mathbb{E}\big(|\xi+f(u)-g(u)|-|\xi+f(u)|\big)\,du\quad\mathbb{P}-\text{a.s.}
\end{aligned}$$

This proves **(ii)** in case 2).

3) Finally, we assume that $\gamma>0$ and $g\in S_n\big([0,1)\big)$. For such g and all $n\in\mathbb{N}$ we choose g_n as the best L^1 approximation of g in $S_n\big([0,1)\big)$ with at most $\#J(g)$ jumps, which fulfills $\|g_n-g\|\underset{n\to\infty}{\longrightarrow}0$. (The existence of such sequence g_n is shown in Lemma 4 **(i)**.) By means of (2.65) and Lemma 2 we obtain that $\mathbb{P}$-almost surely it holds:

$$\begin{aligned}
&\limsup_{n\to\infty}\tilde{H}^n_{\gamma_n}(g_n\,,\,f^n+\xi^n)\\
\le\;&\limsup_{n\to\infty}\gamma_n\#J(g_n)+\limsup_{n\to\infty}\big(\|\xi^n+f^n-g_n\|-\|\xi^n+f^n\|\big)\\
\le\;&\gamma\#J(g)+\int\mathbb{E}\big(|\xi+f(u)-g(u)|-|\xi+f(u)|\big)\,du\\
=\;&H^*_{\gamma,\xi}(g,\,f),
\end{aligned}$$

Hence we have proved statement **(ii)** in all three possible cases and, therefore, the proof of Theorem 1 is complete.

□

2.2 Conclusion

In this chapter it was shown that for any array $(\xi_i^n)_{1\le i\le n,\,n\in\mathbb{N}}$, which fulfills Condition 1, for any $f\in L^1\big([0,1)\big)$ and for the data $y^n=(y_1^n,\dots,y_1^n)$ such that $y_i^n=f_i^n+\xi_i^n$, the functional $\tilde{H}^n_{\gamma_n}(\cdot,\,f^n+\xi^n)$ has the same set of minimizers as the functional $H^n_{\gamma_n}(\cdot,y^n)$. Each estimator $\widehat{f}_n$ of the original function f, as given by a minimizer of $\tilde{H}^n_{\gamma_n}(\cdot,\,f^n+\xi^n)$, is shown to be represented by a step function, constructed using the local medians of the data points. Namely, the value of any estimator $\widehat{f}_n$ on such interval I, where $\widehat{f}_n$ is constant, is given by a local median $\mathrm{med}_I(f^n+\xi^n)\in med_I(f^n+\xi^n)$, or equivalently, by a median $\mathrm{med}_I(y^n)\in med_I(y^n)$ of the points y_i^n, which fall within the interval I. And finally, it was shown that if $n\to\infty$ and $\gamma_n\underset{n\to\infty}{\longrightarrow}\gamma$, then the functional $\tilde{H}^n_{\gamma_n}(\cdot,\,f^n+\xi^n)$ $\mathbb{P}$-almost surely converges to the functional $H^*_{\gamma,\xi}(\cdot\,,\,f)$ with respect to the epi-covergence.

In the following chapter, we investigate the limit functional $H^*_{\gamma,\xi}$ and the set of its minimizers. We also compare the properties of the limit functional $H^*_{\gamma,\xi}$ with the properties of the limit functional obtained in L^2 case by Boysen *et al.* [7, 8].

Chapter 3

Properties of the limit functional

In this chapter we continue working with the functional $H^*_{\gamma,\xi}$, which for any $f \in L^1\big([0,1)$, any random variable $\xi : (\Omega, \mathbb{P}) \to \mathbb{R}$ and $\gamma > 0$ is defined as follows:

$$H^*_{\gamma,\xi}(g\,,\,f) = \gamma\#J(g) + \int\limits_{[0,1)} \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\,du.$$

This functional is the limit functional of

$$\tilde{H}^n_{\gamma_n}(g,\, f^n + \xi^n) = \gamma\#J(g) + \|g - (f^n + \xi^n)\| - \|f^n + \xi^n\|,\ g \in S_n\big([0,1)\big)$$

in the sense of the epi-confergence, when $n \to \infty$ and $\gamma_n \underset{n\to\infty}{\longrightarrow} \gamma$ (for the complete definitions of these functionals see (2.21), p.21 and (2.1), p.16, respectively).

The main goal of this chapter is to show that for any $\gamma \geq 0$, any random variable $\xi : (\Omega, \mathbb{P}) \to \mathbb{R}$, which obeys $\mathbb{P}\{\xi \leq 0\} = \mathbb{P}\{\xi \geq 0\} = \frac{1}{2}$, and any $f \in L^1\big([0,1)\big)$ the set of the minimizers of the functional $H^*_{\gamma,\xi}(\cdot\,,\,f)$ is not empty.

First, however, we slightly rewrite the functional $H^*_{\gamma,\xi}$. To this end we recall some additionally well known definitions.
Let (Θ_1, μ_1) and (Θ_2, μ_2) be two arbitrary measure spaces. The set

$$\Theta_1 \times \Theta_2 := \Big\{(x,y)\ :\ x \in \Theta_1,\ y \in \Theta_2\Big\}. \tag{3.1}$$

is called the cross product of Θ_1 and Θ_2 [17].
For any two functions $g_1 : \Theta_1 \to \mathbb{R}$ and $g_2 : \Theta_2 \to \mathbb{R}$ the tensor product is a functional on the measure space $(\Theta_1 \times \Theta_2, \mu_1 \otimes \mu_2)$, defined by [17]:

$$(g_1 \otimes g_2)(x,y) := g_1(x) \cdot g_2(y), \quad x \in \Theta_1,\ y \in \Theta_2. \tag{3.2}$$

Moreover, in this chapter instead of $L^1\big([0,1)\big)$ we will work with $L^1\big([0,1) \times \Omega\big)$, where Ω is a probability space (definitions of $L^1(\Theta)$ for any measure space Θ and of the norm on $L^1(\Theta)$ are given in the introduction by (1.14) and (1.15), p.12).
At the beginning of Chapter 2 we have defined the functional $\Psi : L^1\big([0,1)\big) \times L^1\big([0,1)\big) \to \mathbb{R}$,

$$\Psi(g_1, g_2) = \|g_1 - g_2\| - \|g_2\|.$$

Now we consider the same functional, which is defined on the space $L^1\big([0,1) \times \Omega\big) \times L^1\big([0,1) \times \Omega\big)$. With the help of this functional we define a new functional H^∞_γ.

Definition 5. *Let $f \in L^1\big([0,1)\big)$ and $\xi : \Omega \to \mathbb{R}$ be an arbitrary random variable on a probability space $(\Omega, \mathbb{P})$. For any $\gamma \geq 0$ the functional $H_\gamma^\infty : L^1\big([0,1)\times\Omega\big)\times L^1\big([0,1)\times\Omega\big) \to \mathbb{R}\cup\{\infty\}$ is given by*

$$H_\gamma^\infty(g\otimes 1, f\otimes 1 + 1\otimes\xi) := \begin{cases} \gamma\#J(g) + \Psi(g\otimes 1,\, f\otimes 1 + 1\otimes\xi) & : \quad g \in S\big([0,1)\big),\, \gamma > 0 \\ \Psi(g\otimes 1,\, f\otimes 1 + 1\otimes\xi) & : \quad g \in L^1\big([0,1)\big),\, \gamma = 0 \\ \infty & : \quad \text{otherwise} \end{cases}$$

In the following lemma we determine the relation between the functional $H^*_{\gamma,\xi}$ and the functional H_γ^∞.

Lemma 5. *For any random variable $\xi : \Omega \to \mathbb{R}$, any $f \in L^1\big([0,1)\big)$, any $g \in L^1\big([0,1)\big)$ and any $\gamma \geq 0$ it holds:*

$$H^*_{\gamma,\xi}(g,f) = H_\gamma^\infty(g\otimes 1, f\otimes 1 + 1\otimes\xi) \tag{3.3}$$

Proof:
Using the definition of the expectation of a random variable, the integral $\int_{[0,1)} \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du$ can be written in the following form:

$$\begin{aligned} & \int_{[0,1)} \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du \\ = \quad & \int_{[0,1)} \int_{\Omega} \Big(\big(|\xi(\omega) + f(u) - g(u)| - |\xi(\omega) + f(u)|\big)\; d\,\mathbb{P}(\omega) \Big)\, du \\ \underset{(3.2)}{=} \quad & \int_{[0,1)\times\Omega} \big|(1\otimes\xi)(u,\omega) + (f\otimes 1)(u,\omega) - (g\otimes 1)(u,\omega)\big|\; du\; d\,\mathbb{P}(\omega) \\ - \quad & \int_{[0,1)\times\Omega} \big|(1\otimes\xi)(u,\omega) + (f\otimes 1)(u,\omega)\big|\; du\; d\,\mathbb{P}(\omega) \\ \underset{(1.14)}{=} \quad & \|(1\otimes\xi) + (f\otimes 1) - (g\otimes 1)\| - \|(1\otimes\xi) + (f\otimes 1)\| \\ \underset{(2.6)}{=} \quad & \Psi(g\otimes 1,\, f\otimes 1 + 1\otimes\xi) \end{aligned}$$

Consequently, according to the definitions of $H^*_{\gamma,\xi}$ and H_γ^∞, it is easy to see that Lemma 5 has been proved.

□

Further on, for any $f \in L^1\big([0,1)\big)$ and $\xi : (\Omega, \mathbb{P}) \to \mathbb{R}$ we use two equivalent forms of writing the sum of f and ξ, namely $(f + \xi)$, or, equivalently, $(f\otimes 1 + 1\otimes\xi)$.

According to the last lemma, the functional $H^*_{\gamma,\xi}$, which contains the expectation of the random variable ξ, is identical with the functional H_γ^∞. The later functional contains two terms, which affect its minimizer: the penalty term, which is responsible for smoothness of the minimizer, and the error term, given by the norm of the difference of two functions from the set $L^1\big([0,1)\times\Omega\big)$. Further, the definition (1.16) of the median of an arbitrary function from $L^1\big(\Theta\big)$ for an arbitrary metric space Θ and the definition (1.15) of the L^1 norm yield:

$$med_{I\times\Omega}(f+\xi) = \operatorname*{argmin}_{c\in\mathbb{R}} \|c\cdot \mathbf{1}_{I\times\Omega} - (f\otimes 1 + 1\otimes\xi)\cdot\mathbf{1}_{I\times\Omega}\|. \tag{3.4}$$

Taking the above into account, we immediately conclude that for any $f\in L^1\big([0,1)\big)$, any random variable $\xi:(\Omega,\mathbb{P})\to\mathbb{R}$ and all $\gamma>0$ the height of the step of each minimizer from $S\big([0,1)\big)$ of $H^\infty_\gamma(\cdot\otimes 1,\ f\otimes 1+1\otimes\xi)$ and, therefore, of $H^*_{\gamma,\xi}(\cdot,\ f)$, on any interval $I\subseteq[0,1)$, where the minimizer is constant, is given by a median $\mathrm{med}_{I\times\Omega}(f+\xi)\in med_{I\times\Omega}(f+\xi)$.
In other words, if

$$\widehat{g}\in\operatorname*{argmin}_{g\in S([0,1))} H^\infty_\gamma(g\otimes 1\,,f\otimes 1+1\otimes\xi)\neq\emptyset$$

then there exists a partition $P_{J(\widehat{g})}$ of the interval $[0,1)$ such that

$$\widehat{g}(x) = \sum_{I\in P_{J(\widehat{g})}} \mathrm{med}_{I\times\Omega}(f+\xi)\cdot\mathbf{1}_I(x), \tag{3.5}$$

where $\mathrm{med}_{I\times\Omega}(f+\xi)\in med_{I\times\Omega}(f+\xi)$.

Further on, for convenience, we consider the functional H^∞_γ instead of $H^*_{\gamma,\xi}$.

3.1 Comparison of L^2 and L^1 cases

At this stage it is interesting to compare the results obtained in this and previous chapters with the analogous results obtained for the L^2 case.

As mentioned in the introduction, it was shown by Boysen *et al.* [7] that for any function $f\in L^2([0,1))$, and for all sequences $(\gamma_n)_{n\in\mathbb{N}}$ with $\gamma_n \underset{n\to\infty}{\longrightarrow} \gamma\geq 0$ and $\gamma_n\cdot(n/\log n)\underset{n\to\infty}{\longrightarrow}\infty$ the functional

$$\tilde{H}_{\gamma_n,2}(\cdot,\ f+\xi^n) = \gamma_n\#J(g) + \|g-f+\xi^n\|_2 - \|f+\xi^n\|_2,\ g\in S_n\big([0,1)\big)$$

converges $\mathbb{P}$-almost surely to the functional

$$H^\infty_{\gamma,2}(g,f) = \gamma\#J(g)+\|g-f\|_2-\|f\|_2,\ g\in S\big([0,1)\big)$$

with respect to the epi-convergence.
Thereby it was assumed that the noise satisfies the following uniform *subgaussian Condition:*

- for all n the random variables $(\xi^n_i)_{1\leq i\leq n}$ are independent
- there exists a universal constant $\beta\in\mathbb{R}$ such that

$$\mathbb{E}e^{\nu\,\xi^n_i}\leq e^{\beta\,\nu^2}$$

  for all $\nu\in\mathbb{R}$, $1\leq i\leq n$ and $n\in\mathbb{N}$.

The first difference that we point out, before we compare the limit functionals in cases L^2 and L^1, is that for the validity of Lemma 1, i.e. for $\tilde{H}^n_{\gamma_n}(\cdot\,,\ f^n+\xi^n)$ to epi-converge to $H^\infty_\gamma(\cdot\otimes 1,\ f\otimes 1+1\otimes\xi)$, the random variables $(\xi^n_i)_{1\le i\le n}$ need to be independent and identical distributed. On the contrary, according to the above, in the L^2 case, the random variables $\xi^n_1,\dots,\xi^n_n$ must not necessary be identical distributed. For example, let $\xi^n_i \sim N(0,\sigma^2_{i,n})$ and $\sup\limits_{n\in\mathbb{N},\,1\le i\le n}\sigma^2_{i,n}<\infty$. Then for all $\beta\ge\frac{1}{2}\sup\limits_{n\in\mathbb{N}\,1\le i\le n}\sigma^2_{i,n}$ it holds: $\mathbb{E}e^{\nu\,\xi^n_i}=\exp\left\{\frac{\sigma^2_{i,n}\,\nu^2}{2}\right\}\le e^{\beta\,\nu^2}$, or in other words, $(\xi^n_i)_{1\le i\le n}$ satisfy the subgaussian condition.

Now we compare the corresponding limit functionals. The functionals $H^\infty_\gamma(\cdot\otimes 1,\ f\otimes 1+1\otimes\xi)$ and the functional $\gamma\#J(g)+\|g-f\|-\|f\|$, if both considered as functionals on $L^1([0,1))$, differ from each other in the following ways. First, they differ in their form and, secondly, the minimizers of these two functionals are not identical.

As we noticed in (3.5), the value of any minimizer of the functional $H^\infty_\gamma(\cdot\otimes 1\,,f\otimes 1+1\otimes\xi)$ on any such interval I, where this minimizer is constant, is given by a local median $\mathrm{med}_I(f+\xi)\in med_I(f+\xi)$. With regards to the functional $\gamma\#J(g)+\|g-f\|-\|f\|$, it is known that the heights of the steps of its minimizer on any interval I, where the minimizer is constant, are given by $\mathrm{med}_I(f)\in med_I(f)$. But crucially, as we will see in the next example, $med(f)\ne med(f+\xi)$.

Example. Let ξ be a random variable in the following probability space $(\Omega,\mathbb{P})$:
$\Omega=\{-1,1\}$, with $\mathbb{P}(\xi=-1)=\frac{1}{2}$ and $\mathbb{P}(\xi=1)=\frac{1}{2}$
and let for all $x\in[0,1)$

$$f(x)=\begin{cases}0, & x<\frac{1}{3}\\ a, & x\ge\frac{1}{3}\end{cases}.$$

Obviously, the set of images of $(f+\xi)$ is given by:

$$\mathrm{Im}(f+\xi)=\{-1,\ 1,\ a-1,\ a+1\},$$

thereby

$$\mathbb{P}(f+\xi=-1)=\frac{1}{6},\ \mathbb{P}(f+\xi=1)=\frac{1}{6},\ \mathbb{P}(f+\xi=a-1)=\frac{1}{3},\ \mathbb{P}(f+\xi=a+1)=\frac{1}{3}.$$

From this it follows that

$$med(f+\xi)=\begin{cases}[a-1,\ 1], & a<2\\ a-1, & a\ge 2\end{cases}$$

while $med(f)=a$ for all a.

This example demonstrates a very important fact, namely that in the L^1 case, the limit functional $H^\infty_\gamma(\cdot\otimes 1,\ f\otimes 1+1\otimes\xi)$ contains the information about the noise. This has far reaching consequences. Thus, since $\|g-f\|\ge 0$ for all $g,\ f\in L^2\big([0,1)\big)$, then the lower bound of the limit functional in the L^2 case, $H^\infty_{\gamma,2}(g,f)=\gamma\#J(g)+\|g-f\|_2-\|f\|_2$, $g\in S\big([0,1)\big)$, can obviously be estimated by the value $\gamma\#J(g)-\|f\|_2$. On the contrary, in the L^1 case, the intrinsic dependence of the limit functional on ξ, implies that the estimate of the lower bound retains the information on the noise term. This property is demonstrated in the next proposition.

Proposition 2. *Let* $(\Omega, \mathbb{P})$ *be a probability space and* $\xi : \Omega \to \mathbb{R}$ *be a random variable with* $0 < \mathbb{E}|\xi| < \infty$ *and*

$$\mathbb{P}\{\xi \leq 0\} = \mathbb{P}\{\xi \geq 0\} = \frac{1}{2}. \tag{3.6}$$

Then for all $h \in L^1\big([0,1)\big)$ *and any interval* $\widetilde{I} \subseteq [0,1)$ *it holds:*

$$\|(h \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{\widetilde{I} \otimes \Omega}\| \geq \|(1 \otimes \xi) \cdot \mathbf{1}_{\widetilde{I} \otimes \Omega}\| = \|\xi\|_\Omega \cdot l(\widetilde{I}). \tag{3.7}$$

In particular it holds

$$\|(h \otimes 1 + 1 \otimes \xi)\| \geq \|(1 \otimes \xi)\| = \|\xi\|_\Omega. \tag{3.8}$$

Proof:

First of all, we notice that for any arbitrary random variable $\xi : \Omega \to \mathbb{R}$ and any interval $\widetilde{I} \subseteq [0,1)$ it holds:

$$\begin{aligned} \|(1 \otimes \xi) \cdot \mathbf{1}_{\widetilde{I} \otimes \Omega}\| &= \int\limits_{\widetilde{I} \times \Omega} |(1 \otimes \xi)(x, \omega)|\, dx\; d\,\mathbb{P}(\omega) = \int\limits_{\widetilde{I}} dx \cdot \int\limits_{\Omega} |\xi(\omega)|\; d\,\mathbb{P}(\omega) \\ &= l(\widetilde{I}) \cdot \int\limits_{\{\omega\,:\,\xi(\omega) \geq 0\}} \xi(\omega)\; d\,\mathbb{P}(\omega) - l(\widetilde{I}) \cdot \int\limits_{\{\omega\,:\,\xi(\omega) \leq 0\}} \xi(\omega)\; d\,\mathbb{P}(\omega) \end{aligned} \tag{3.9}$$

Further on, we consider an arbitrary function $h \in L^1\big([0,1)\big)$ and an arbitrary interval $\widetilde{I}$ and denote by $\widetilde{\mathcal{I}}_+$ the set of all intervals $I \subseteq \widetilde{I}$, where the function h is not negative and by $\widetilde{\mathcal{I}}_-$ the set of all intervals $I \subseteq \widetilde{I}$, where the function h is negative, i.e.

$$\begin{aligned} \widetilde{\mathcal{I}}_+ &:= \big\{ I \subseteq \widetilde{I} \;:\; \forall\, x \in I \;\; h(x) \geq 0 \big\} \\ \widetilde{\mathcal{I}}_- &:= \big\{ I \subseteq \widetilde{I} \;:\; \forall\, x \in I \;\; h(x) < 0 \big\} \end{aligned}$$

Obviously, we have $\widetilde{I} = \widetilde{\mathcal{I}}_+ \cup \widetilde{\mathcal{I}}_-$.

Now we consider $\|(1 \otimes \xi) \cdot \mathbf{1}_{\widetilde{I} \otimes \Omega}\|$:

$$\begin{aligned} &\|(h \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{\widetilde{I} \otimes \Omega}\| \\ = &\int\limits_{\widetilde{I} \times \Omega} \big|(h \otimes 1)(x, \omega) + (1 \otimes \xi)(x, \omega)\big|\, dx\; d\,\mathbb{P}(\omega) = \int\limits_{\widetilde{I}} \Big(\int\limits_{\Omega} \big|h(x) + \xi(\omega)\big|\; d\,\mathbb{P}(\omega) \Big)\, dx \\ = &\int\limits_{\widetilde{I}} \Big(\int\limits_{\{\,\omega\,:\,h(x) + \xi(\omega) \geq 0\,\}} \big(h(x) + \xi(\omega)\big)\; d\,\mathbb{P}(\omega) \Big)\, dx \\ - &\int\limits_{\widetilde{I}} \Big(\int\limits_{\{\,\omega\,:\,h(x) + \xi(\omega) \leq 0\,\}} \big(h(x) + \xi(\omega)\big)\; d\,\mathbb{P}(\omega) \Big)\, dx \\ = &\sum_{I \in \widetilde{\mathcal{I}}_+} \int\limits_{I} \Big(\int\limits_{\{\,\omega\,:\,\xi(\omega) \geq -h(x)\,\}} \big(h(x) + \xi(\omega)\big)\; d\,\mathbb{P}(\omega) \Big)\, dx \end{aligned}$$

$$
\begin{aligned}
& - \sum_{I\in\widetilde{\mathcal{I}}_+} \int_I \Bigg(\int_{\{\,\omega\,:\,\xi(\omega)\le -h(x)\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega)\Bigg)\, dx \\
& + \sum_{I\in\widetilde{\mathcal{I}}_-} \int_I \Bigg(\int_{\{\,\omega\,:\,\xi(\omega)\ge -h(x)\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega)\Bigg)\, dx \\
& - \sum_{I\in\widetilde{\mathcal{I}}_-} \int_I \Bigg(\int_{\{\,\omega\,:\,\xi(\omega)\le -h(x)\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega)\Bigg)\, dx \\
= & \sum_{I\in\widetilde{\mathcal{I}}_+} \int_I \Bigg(\int_{\{\,\omega\,:\,\xi(\omega)\ge 0\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega) + \int_{\{\,\omega\,:\,-h(x)\le\xi(\omega)\le 0\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega)\Bigg)\, dx \\
& - \sum_{I\in\widetilde{\mathcal{I}}_+} \int_I \Bigg(\int_{\{\,\omega\,:\,\xi(\omega)\le 0\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega) - \int_{\{\,\omega\,:\,-h(x)\le\xi(\omega)\le 0\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega)\Bigg)\, dx \\
& + \sum_{I\in\widetilde{\mathcal{I}}_-} \int_I \Bigg(\int_{\{\,\omega\,:\,\xi(\omega)\ge 0\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega) - \int_{\{\,\omega\,:\,0\le\xi(\omega)\le -h(x)\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega)\Bigg)\, dx \\
& - \sum_{I\in\widetilde{\mathcal{I}}_-} \int_I \Bigg(\int_{\{\,\omega\,:\,\xi(\omega)\le 0\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega) + \int_{\{\,\omega\,:\,0\le\xi(\omega)\le -h(x)\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega)\Bigg)\, dx \\
= & \sum_{I\in\widetilde{\mathcal{I}}_+} \int_I h(x)\, dx \cdot \Bigg(\int_{\{\,\omega\,:\,\xi(\omega)\ge 0\,\}} d\,\mathbb{P}(\omega) - \int_{\{\,\omega\,:\,\xi(\omega)\le 0\,\}} d\,\mathbb{P}(\omega)\Bigg) \\
& + \sum_{I\in\widetilde{\mathcal{I}}_+} \int_I dx \cdot \Bigg(\int_{\{\,\omega\,:\,\xi(\omega)\ge 0\,\}} \xi(\omega)\; d\,\mathbb{P}(\omega) - \int_{\{\,\omega\,:\,\xi(\omega)\le 0\,\}} \xi(\omega)\; d\,\mathbb{P}(\omega)\Bigg) \\
& + 2 \sum_{I\in\widetilde{\mathcal{I}}_+} \int_I \Bigg(\int_{\{\,\omega\,:\,0\le\xi(\omega)\le -h(x)\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega)\Bigg)\, dx \\
& + \sum_{I\in\widetilde{\mathcal{I}}_-} \int_I h(x)\, dx \cdot \Bigg(\int_{\{\,\omega\,:\,\xi(\omega)\ge 0\,\}} d\,\mathbb{P}(\omega) - \int_{\{\,\omega\,:\,\xi(\omega)\le 0\,\}} d\,\mathbb{P}(\omega)\Bigg) \\
& + \sum_{I\in\widetilde{\mathcal{I}}_-} \int_I dx \cdot \Bigg(\int_{\{\,\omega\,:\,\xi(\omega)\ge 0\,\}} \xi(\omega)\; d\,\mathbb{P}(\omega) - \int_{\{\,\omega\,:\,\xi(\omega)\le 0\,\}} \xi(\omega)\; d\,\mathbb{P}(\omega)\Bigg) \\
& - 2 \sum_{I\in\widetilde{\mathcal{I}}_-} \int_I \Bigg(\int_{\{\,\omega\,:\,0\le\xi(\omega)\le -h(x)\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega)\Bigg)\, dx \\
\underset{(3.6)}{=} & \Bigg(\sum_{I\in\widetilde{\mathcal{I}}_+} l(I) + \sum_{I\in\widetilde{\mathcal{I}}_-} l(I)\Bigg) \cdot \Bigg(\int_{\{\,\omega\,:\,\xi(\omega)\ge 0\,\}} \xi(\omega)\; d\,\mathbb{P}(\omega) - \int_{\{\,\omega\,:\,\xi(\omega)\le 0\,\}} \xi(\omega)\; d\,\mathbb{P}(\omega)\Bigg) \\
& + 2 \sum_{I\in\widetilde{\mathcal{I}}_+} \int_I \Bigg(\int_{\{\,\omega\,:\,-h(x)\le\xi(\omega)\le 0\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega)\Bigg)\, dx \\
& - 2 \sum_{I\in\widetilde{\mathcal{I}}_-} \int_I \Bigg(\int_{\{\,\omega\,:\,0\le\xi(\omega)\le -h(x)\,\}} \big(h(x)+\xi(\omega)\big)\; d\,\mathbb{P}(\omega)\Bigg)\, dx
\end{aligned}
$$

$$\begin{aligned}
\geq \quad & l(\widetilde{I}) \cdot \Bigg(\int\limits_{\{\,\omega\,:\,\xi(\omega)\geq 0\,\}} \xi(\omega)\; d\,\mathbb{P}(\omega) - \int\limits_{\{\,\omega\,:\,\xi(\omega)\leq 0\,\}} \xi(\omega)\; d\,\mathbb{P}(\omega) \Bigg) \\
& + \; 2 \sum_{I\in\widetilde{\mathcal{I}}_{+}} \int\limits_{I} \Bigg(\int\limits_{\{\,\omega\,:\,-h(x)\leq\xi(\omega)\leq 0\,\}} \big(h(x)+(-h(x))\big)\; d\,\mathbb{P}(\omega) \Bigg)\, dx \\
& - \; 2 \sum_{I\in\widetilde{\mathcal{I}}_{-}} \int\limits_{I} \Bigg(\int\limits_{\{\,\omega\,:\,0\leq\xi(\omega)\leq -h(x)\,\}} \big(h(x)+(-h(x))\big)\; d\,\mathbb{P}(\omega) \Bigg)\, dx \\
\underset{(3.9)}{=} \quad & \|(1\otimes\xi)\cdot \mathbf{1}_{\widetilde{I}\otimes\Omega}\|.
\end{aligned}$$

From this it follows that for all $h \in L^1\big([0,1)\big)$ and any interval $\widetilde{I} \subseteq [0,1)$ we have:

$$\|(h\otimes 1 + 1\otimes\xi)\cdot \mathbf{1}_{\widetilde{I}\otimes\Omega}\| \geq \|(1\otimes\xi)\cdot \mathbf{1}_{\widetilde{I}\otimes\Omega}\|.$$

This proves the validity of (3.7). The validity of (3.8) is obtained if we assume that $\widetilde{I} = [0,1)$. Therefore the proof of Proposition 2 is complete.

□

Proposition 2 will also be used later in the following section to prove the existence of minimizer of the functional $H_\gamma^\infty(\cdot\,, f\otimes 1 + 1\otimes\xi)$.

3.2 Existence of a minimizer

In this section we will achieve the main goal of this chapter, namely, we show that for any $\gamma \geq 0$, any symmetrical random variable $\xi : (\Omega,\mathbb{P}) \to \mathbb{R}$, any $f \in L^1\big([0,1)\big)$ and any $\gamma \geq 0$, the set of minimizer of the functional $H_\gamma^\infty(\cdot\,, f\otimes 1 + 1\otimes\xi)$ is not empty.

We start with formulation of some auxiliary results and introduce several new notations and definitions.
Let $(\Omega,\mathbb{P})$ be an arbitrary probability space and $\mathcal{P}$ be the set of all partitions of the interval $[0,1)$. Further on, by Λ_Ω and $S(\Lambda_\Omega)$ we denote the two following sets:

$$\Lambda_\Omega := [0,1)\times\Omega. \tag{3.10}$$

$$S(\Lambda_\Omega) := \Big\{ h\in L^1\big(\Lambda_\Omega\big) \big| \exists\, P_{J(h)}\in\mathcal{P} : h(x,\omega) = \sum_{I\in P_{J(h)}} h_I\cdot \mathbf{1}_{I\times\Omega}(x,\omega),\ \forall I\in P_{J(h)}\ \ h_I\in\mathbb{R} \Big\}.$$

In other words, we say that a function $h \in L^1(\Lambda_\Omega)$ is a function from the set $S(\Lambda_\Omega)$, if for some partition $P_{J(h)}$ of the interval $[0,1)$ the function h can be represented as follows:

$$h(x,\omega) = \sum_{I\in P_{J(h)}} h_I\cdot \mathbf{1}_{I\times\Omega}(x,\omega), \quad h_I\in\mathbb{R}\ \ \forall\, I\in P_{J(h)} \tag{3.11}$$

Proposition 3. *Let* $\widehat{g} = \sum\limits_{I\in P_{J(\widehat{g})}} \widehat{g}_I\cdot\mathbf{1}_I$ *be an arbitrary step function on* $[0,1)$. *Then for any* $f\in L^1\big([0,1)\big)$, *any random variable* $\xi : (\Omega,\mathbb{P})\to\mathbb{R}$ *and all* $\gamma > 0$ *the two following statements are equivalent:*

$$\widehat{g} \in \underset{g\in S([0,1))}{\operatorname{argmin}} H_\gamma^\infty(g\otimes 1\,, f\otimes 1 + 1\otimes\xi) \tag{3.12}$$

and

$$\widehat{h} := \sum_{I \in P_{J(\widehat{g})}} \widehat{g}_I \cdot \mathbf{1}_{I\times\Omega} \in \underset{h\in S(\Lambda_\Omega)}{\operatorname{argmin}} H_\gamma^\infty(h, f\otimes 1 + 1\otimes\xi). \tag{3.13}$$

Proof:
The proof of this proposition immediately follows from the following fact:
for any $g = \sum_{I\in P_{J(g)}} g_I \cdot \mathbf{1}_I \in S([0,1))$ it holds:

$$(g\otimes 1)(x,\omega) = \Big(\sum_{I\in P_{J(g)}} g_I \cdot \mathbf{1}_I \otimes 1\Big)(x,\omega) = \sum_{I\in P_{J(g)}} g_I \cdot \mathbf{1}_{I\times\Omega}(x,\omega).$$

□

Using Proposition 3, it is clear that in oder to show that

$$\underset{g\in S([0,1))}{\operatorname{argmin}} H_\gamma^\infty(g\otimes 1\,, f\otimes 1 + 1\otimes\xi) \neq \emptyset$$

for any $\gamma > 0$, any symmetrical random variable $\xi : (\Omega,\mathbb{P}) \to \mathbb{R}$ and any $f \in L^1\big([0,1)\big)$, it is sufficient to show that

$$\underset{h\in S(\Lambda_\Omega)}{\operatorname{argmin}} H_\gamma^\infty(h, f\otimes 1 + 1\otimes\xi) \neq \emptyset. \tag{3.14}$$

Moreover, according to (3.5), it is also clear that, if

$$\widehat{h} \in \underset{h\in S(\Lambda_\Omega)}{\operatorname{argmin}} H_\gamma^\infty(h, f\otimes 1 + 1\otimes\xi) \neq \emptyset,$$

then there exists a partition $P_{J(\widehat{h})}$ of the interval $[0,1)$ such that

$$\widehat{h}(x,\omega) = \sum_{I\in P_{J(\widehat{h})}} \mathrm{med}_{I\times\Omega}(f+\xi)\cdot \mathbf{1}_{I\times\Omega}(x,\omega), \tag{3.15}$$

where $\mathrm{med}_{I\times\Omega}(f+\xi) \in med_{I\times\Omega}(f+\xi)$.

To show the validity of (3.14) we use the following well-known theorem:

Theorem 2. *(Weierstrass Theorem) [22, s.13] Let $F : \Theta \to \mathbb{R}$ be a lower semicontinuous functional, and K be a compact nonempty subset of Θ. Then it holds:*

$$\underset{K}{\operatorname{argmin}}\, F \neq \emptyset.$$

Especially, if for some $c \in \mathbb{R}$ the set $\{\theta : F(\theta) \leq c\}$ is relatively compact and nonempty, then

$$\underset{\Theta}{\operatorname{argmin}} F \neq \emptyset.$$

□

Due to this theorem and taking (3.15) into account, we conclude that in order to prove (3.14), one needs to prove two following facts:
(i) For any $f \in L^1\big([0,1)\big)$ any $\xi : \Omega \to \mathbb{R}$ and all $\gamma \geq 0$ the functional $H_\gamma^\infty(\cdot\,, f \otimes 1 + 1 \otimes \xi)$ is lower semicontinuous.
(ii) Let $\mathcal{P}$ be the set of all partitions of the interval $[0,1)$ and $(\Omega, \mathbb{P})$ an arbitrary probability space. For any $f \in L^1\big([0,1)\big)$ and any $\xi : \Omega \to \mathbb{R}$ the set:

$$\mathrm{Med}(f+\xi) := \Big\{ h \in L^1\big(\Lambda_\Omega\big) \,\big|\, \exists P \in \mathcal{P} : h = \sum_{I \in P} m_{I\times\Omega} \cdot \mathbf{1}_{I\times\Omega},\ \forall I \in P\ \ m_{I\times\Omega} \in med_{I\times\Omega}(f+\xi) \Big\}$$

is relatively compact in $L^1\big(\Lambda_\Omega\big)$, where $med_{I\times\Omega}(f+\xi)$ is defined by (3.4).

We start with the semicontiniuty of the functional H_γ^∞.

Proposition 4. *For any $\gamma \geq 0$ the functional H_γ^∞, as given in Definition 5, is lower semicontinuous.*

Proof:
From Lemma 3 (Chapter 2, p. 30), we known that $\#J$ for $g \in S\big([0,1)\big)$, which is defined by (2.47), is lower semicontinuous with respect to the L^1 convergence. Obviously, the L^1 norm on $[0,1) \times \Omega$, as well as the difference of two L^1 norms is continuous and therefore is lower semicontinuous. Thus H_γ^∞, which according to Definition 5, in case of $\gamma > 0$ is the sum of two terms, namely $\#J$ and the difference of two L^1 norms on $[0,1) \times \Omega$, is lower semicontinuous as a sum of two lower semicontinuous functionals.

□

The second fact, which is needed in order to prove the existence of a minimizer of the functional H_γ^∞, namely, the relative compactness of the set $\mathrm{Med}(f+\xi)$, is proved in the next subsection.

3.2.1 Compactness

Before we start the proof of the relative compactness of the set $\mathrm{Med}(f+\xi)$, we derive some auxiliary results.
First of all we recall the definition and the criteria of a relatively compact set.

Definition 6. *[27] Let Θ be a metric space. A set $M \subset \Theta$ is a relatively compact set, if for any sequence in M there is a convergent subsequence in Θ.*

Definition 7. *[27, 3] For some $\varepsilon > 0$ a set N_ε is an ε-net with respect to some set $M \subset \Theta$ if it holds:*

$$\forall\, x \in M\ \ \exists\, a \in N_\varepsilon :\ \ \|x - a\| < \varepsilon.$$

Theorem 3. *[27] If the set $M \subset \Theta$ is relatively compact then for any $\varepsilon > 0$ there exists a finite ε-net with respect to the set M.*
If the space Θ is complete and for any $\varepsilon > 0$ there exists a finite ε-net N_ε with respect to the set $M \subset \Theta$ then the set M is relatively compact.

□

Proposition 5. *[27, 3] Let Θ be a complete metric space. If for any $\varepsilon > 0$ there exists a relatively compact ε-net N_ε with respect to M, then the set M is relatively compact.*

□

Now we formulate several auxiliary results, which follow from Lemma 28 in Appendix A, p.125.

Lemma 6. *Let* $(\Omega, \mathbb{P})$ *be an arbitrary probability space,* $\xi : \Omega \to \mathbb{R}$ *is a random variable with* $0 < \mathbb{E}|\xi| < \infty$, $f \in L^1\big([0,1)\big)$ *and for any interval* $I \subseteq [0,1)$ *of the length* $l(I)$ *let*

$$\overline{f}_I := \frac{1}{l(I)} \int\limits_I f(x)\, dx.$$

Then for any $\varepsilon > 0$ *there exist such* $\delta > 0$ *that for all partitions* P *of the interval* $[0,1]$ *it holds:*

$$\sup_{P \in \mathcal{P}} \sum_{I \in P,\, l(I) < \delta} \| \, \overline{f}_I \cdot \mathbf{1}_{I \times \Omega} - (f \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{I \times \Omega} \, \| < \varepsilon. \tag{3.16}$$

Proof:

For any $I \subseteq [0,1)$ we consider $\| \, (\overline{f}_I \otimes 1) \cdot \mathbf{1}_{I \times \Omega} - (f \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{I \times \Omega} \, \|$:

$$\begin{aligned}
& \| \, \overline{f}_I \cdot \mathbf{1}_{I \times \Omega} - (f \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{I \times \Omega} \, \| \\
\leq \; & \| \, \overline{f}_I \cdot \mathbf{1}_{I \times \Omega} - (f \otimes 1) \cdot \mathbf{1}_{I \times \Omega} \, \| + \|(1 \otimes \xi) \cdot \mathbf{1}_{I \times \Omega} \, \| \\
= \; & \int\limits_{I \times \Omega} | \, \overline{f} - (f \otimes 1)(u, \omega) \, | \; du \; d\,\mathbb{P}(\omega) + \int\limits_{I \times \Omega} (1 \otimes \xi)(u, \omega)| \; du \; d\,\mathbb{P}(\omega) \\
\underset{(3.2)}{=} \; & \int\limits_{I \times \Omega} |\overline{f} - f(u)| \; du \; d\,\mathbb{P}(\omega) + \int\limits_{I \times \Omega} |\xi(\omega)| \; du \; d\,\mathbb{P}(\omega) \\
= \; & \int\limits_I |\overline{f} - f(u)|\, du \cdot \int\limits_\Omega d\,\mathbb{P}(\omega) + \int\limits_I du \cdot \int\limits_\Omega |\xi(\omega)| \; d\,\mathbb{P}(\omega) \\
= \; & \|\overline{f}_I \cdot \mathbf{1}_I - f \cdot \mathbf{1}_I\| + l(I) \cdot \mathbb{E}|\xi|
\end{aligned}$$

According to Lemma 28 (see Appendix A, p. 125) we have:

$$\forall \varepsilon > 0 \; \exists \delta_1 > 0 : \quad \sup_{P \in \mathcal{P}} \sum_{I \in P,\, l(I) < \delta} \| \, \overline{f}_I \cdot \mathbf{1}_I - f \cdot \mathbf{1}_I \, \| < \frac{\varepsilon}{2}.$$

Now for any $\varepsilon > 0$ we define δ_2 as follows:

$$\delta_2 := \frac{\varepsilon}{2\,\mathbb{E}|\xi|}.$$

It is obvious that for any interval $I \subset [0,1)$ such that $l(I) < \delta_2$ it holds:

$$l(I) \cdot \mathbb{E}|\xi| < \frac{\varepsilon}{2}.$$

Summarizing all facts listed above, we obtain that for any $\varepsilon > 0$ and for $\delta = \min\{\delta_1, \delta_2\}$ the following statement is valid:

$$\sup_{P \in \mathcal{P}} \sum_{I \in P,\, l(I) < \delta} \| \, \overline{f}_I \cdot \mathbf{1}_{I \times \Omega} - (f \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{I \times \Omega} \, \| \leq \|\overline{f}_I \cdot \mathbf{1}_I - f \cdot \mathbf{1}_I\| + l(I) \cdot \mathbb{E}|\xi| < \varepsilon.$$

This completes the proof of Lemma 6.

□

Lemma 7. *Let $(\Omega, \mathbb{P})$ be an arbitrary probability space, $f \in L^1\big([0,1)\big)$ and $\xi : \Omega \to \mathbb{R}$ be a random variable with $0 < \mathbb{E}|\xi| < \infty$. Then for any $\varepsilon > 0$ there exist such $\delta > 0$ that for all partitions P of the interval $[0,1]$ it holds:*

$$\sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta} \| \, m_{I\times\Omega} \cdot \mathbf{1}_{I\times\Omega} - (f\otimes 1 + 1\otimes \xi)\cdot \mathbf{1}_{I\times\Omega} \, \| < \varepsilon, \quad m_{I\times\Omega} \in med_{I\times\Omega}\,(f+\xi)$$

Proof:

From the definition of $med_{I\times\Omega}(f+\xi)$ (see (3.4)) we have:

$$med_{I\times\Omega}(f+\xi) = \operatorname*{argmin}_{c\in\mathbb{R}} \|c\cdot \mathbf{1}_{I\times\Omega} - (f\otimes 1 + 1\otimes\xi)\cdot\mathbf{1}_{I\times\Omega}\|.$$

Since $m_{I\times\Omega} \in med_{I\times\Omega}(f+\xi)$, it is clear that for any partition P of the interval $[0,1]$ it holds:

$$\begin{aligned} &\sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta} \|m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega} - (f\otimes 1 + 1\otimes\xi)\cdot\mathbf{1}_{I\times\Omega}\| \\ \le\; &\sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta} \|(\overline{f}_I\otimes 1 + 1\otimes\xi)\cdot\mathbf{1}_{I\times\Omega} - (f\otimes 1 + 1\otimes\xi)\cdot\mathbf{1}_{I\times\Omega}\|. \end{aligned}$$

Due to Lemma 6, we obtain:

$$\forall\varepsilon>0 \;\; \exists\delta>0: \quad \sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta} \| \, m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega} - (f\otimes 1 + 1\otimes\xi)\cdot\mathbf{1}_{I\times\Omega} \, \| < \varepsilon$$

□

Now we proceed to proving the main lemma of this section.

Lemma 8. *Let $(\Omega, \mathbb{P})$ be an arbitrary probability space, $\Lambda_\Omega = [0,1)\times\Omega$ and $\mathcal{P}$ is set of all partitions of the interval $[0,1)$. For any $f \in L^1\big([0,1)\big)$ and any random variable $\xi : \Omega\to\mathbb{R}$ with $0 < \mathbb{E}|\xi| < \infty$, the set:*

$$\mathrm{Med}(f+\xi) = \Big\{ h \in L^1(\Lambda_\Omega) \,\big|\, \exists P\in\mathcal{P} : h = \sum_{I\in P} m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega};\; \forall I\in P \;\; m_{I\times\Omega}\in med_{I\times\Omega}(f+\xi) \Big\}$$

is relatively compact in $L^1\big([0,1)\big)$, where $med_{I\times\Omega}(f+\xi)$ is defined by (3.4).

Proof:
Let $f\in L^1\big([0,1)\big)$ and $\xi:\Omega\to\mathbb{R}$ be a random variable with $0<\mathbb{E}|\xi|<\infty$.

We construct the proof of this lemma as follows:

Step 1:
For an arbitrary $M\in\mathbb{N}$ and $c<\infty$ we define a partition $P_{2M+1} := \{I_1,\ldots,I_{2M+1}\}$ of the interval $[0,1)$ and a set $\mathcal{F}_M^c$ as follows:

$$\mathcal{F}_M^c := \Big\{ \sum_{k=1}^{M} m_{Q_k}\cdot\mathbf{1}_{Q_k} + \sum_{k=M+1}^{2M+1} (f+\xi)\cdot\mathbf{1}_{Q_k} : \forall\, Q_k = I_k\times\Omega,\; m_{Q_k}\le c,\; m_{Q_k}\in med_{Q_k}(f+\xi) \Big\}$$

and show that this set is relatively compact in $L^1(\Lambda_\Omega)$ for all $M\in\mathbb{N}$ and $c<\infty$.

Step 2:

Further on, we show that it holds:

$$\forall \varepsilon > 0 \ \exists c,\, M: \ \forall h \in Med(f+\xi) \ \exists \widetilde{h} \in \mathcal{F}^c_M : \|\widetilde{h} - h\| < \varepsilon. \tag{3.17}$$

These two steps imply that for any $\varepsilon > 0$ there is a relatively compact set $\mathcal{F}^c_M$ that is an ε-net with respect to the set $Med(f+\xi)$. Moreover, it is known [27] that the space $L^1(\Lambda_\Omega)$ is complete and, according to Proposition 5, the set $Med(f+\xi)$ is also relatively compact.

Step 1:

Let $f \in L^1\big([0,1)\big)$ and $I = [a,b)$. It is clear that I is a relatively compact set in $\mathbb{R}$.

From Lemma 26, Appendix A, p. 123, we know that for all $a, b, c \in \mathbb{R}$, $c < \infty$, the mapping $f_0 : (a,b,c) \to c \cdot \mathbf{1}_{[a,b)}$ is continuous with respect to the L^1 norm. According to this, it is obvious that for any probability space Ω and all $a, b, c \in \mathbb{R}$, $c < \infty$, the function $h_0 : (a,b,c) \to c \cdot \mathbf{1}_{[a,b)\times\Omega}$ is also a continuous function.

Now we consider the mapping $h_1 : (a,b) \to (f \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{[a,b)\times\Omega}$, where $f \in L^1\big([0,1)\big)$, $\xi : \Omega \to \mathbb{R}$ and $[a,b) \subset [0,1)$. We demonstrate that h_1 is also a continuous function (even if f is an unbounded function).

For this purpose we have to show that $\lim\limits_{a\to a_0,\, b\to b_0} \|h_1(a,b) - h_1(a_0,b_0)\| = 0$.
Let a, b, a_0, b_0 be some real numbers such that $a \to a_0$ and $b \to b_0$. Without loss of generality we assume that $a > a_0$ and $b_0 > b$ and consider $\|h_1(a,b) - h_1(a_0,b_0)\|$:

$$\|h_1(a,b) - h_1(a_0,b_0)\| = \|(f \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{[a,b)\times\Omega} - (f \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{[a_0,b_0)\times\Omega}\|$$

$$= \|(f \otimes 1 + 1 \otimes \xi) \cdot (\mathbf{1}_{[a,b)\times\Omega} - \mathbf{1}_{[a_0,b_0)\times\Omega})\|$$

$$= \int\limits_{[0,1)\times\Omega} |\,(f\otimes 1)(x,\omega) + (1\otimes\xi)(x,\omega)\,| \cdot |\,\mathbf{1}_{[a,b)\times\Omega}(x,\omega) - \mathbf{1}_{[a_0,b_0)\times\Omega}(x,\omega)\,|\, dx\; d\,\mathbb{P}(\omega)$$

$$= \left| \int\limits_{[a,a_0)\times\Omega} (f\otimes 1)(x,\omega) + (1\otimes\xi)(x,\omega)|\, dx\; d\,\mathbb{P}(\omega) \right|$$

$$+ \left| \int\limits_{[b,b_0)\times\Omega} |(f\otimes 1)(x,\omega) + (1\otimes\xi)(x,\omega)|\, dx\; d\,\mathbb{P}(\omega) \right|.$$

Hence $\lim\limits_{a\to a_0,\, b\to b_0,} \|h_1(a,b) - h_1(a_0,b_0)\| = 0$ for any $f \in L^1\big([0,1)\big)$ and any $\xi : \Omega \to \mathbb{R}$.

Thus for any $f \in L^1\big([0,1)\big)$, any random variable $\xi : \Omega \to \mathbb{R}$ and all $0 < a < b < 1$ the mapping: $(a,b) \to (f \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{[a,b)\times\Omega}$ is continuous.

As a consequence of the above results, we have:

1) $\forall\, a, b, c \in \mathbb{R}$ $(c < \infty)$ the function: $[a,b,c) \to c \cdot \mathbf{1}_{[a,b)\times\Omega}$ is continuous.

2) The function: $(a,b) \to (f \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{[a,b)\times\Omega}$ is continuous for all $f \in L^1\big([0,1)\big)$, $\xi : \Omega \to \mathbb{R}$ and $0 < a < b < 1$,

3) The set $\mathcal{K} = \Big\{[-c,c] \times \big\{[a_i,b_i) : \ 0 \le a_i < b_i \le 1\big\}_{i=1}^{2M+1}\Big\}$ is relatively compact.

Let us define $h' := \sum\limits_{k=M+1}^{2M+1} (f \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{[a_k,b_k)\times\Omega}$, $M \in \mathbb{N}$. The function h' is a continuous mapping of the relatively compact set $\Big\{[a_{M+1}, b_{M+1}) \cup ... \cup [a_{2M+1}, b_{2M+1})\Big\}$. Hence the set $\Big\{\sum\limits_{k=M+1}^{2M+1} (f \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{[a_k,b_k)\times\Omega}\Big\}$, or in other words, the set $\Big\{\sum\limits_{k=M+1}^{2M+1} (f + \xi) \cdot \mathbf{1}_{[a_k,b_k)\times\Omega}\Big\}$ is also relatively compact as an image of a relatively compact set under a continuous function [27].

Because the mapping $[-c,c] \times [a_k,b_k) \longrightarrow m_{I_k\times\Omega} \cdot \mathbf{1}_{[a_k,b_k)\times\Omega}$ $(|m_{I_k\times\Omega}| \le c)$ is continuous, it is clear that the mapping $[-c,c] \times \big\{[a_i,b_i)\big\}_{k=1}^{M} \longrightarrow \sum\limits_{k=1}^{M} m_{I_k\times\Omega} \cdot \mathbf{1}_{[a_k,b_k)\times\Omega}$ is continuous as well. Consequently, the set $\sum\limits_{k=1}^{M} m_{I_k} \cdot \mathbf{1}_{[a_k,b_k)\times\Omega}$ is also relatively compact as an image of a relatively compact set under a continuous mapping.

Thus we have proved that the set

$$\mathcal{F}_M^c = \Big\{\sum_{k=1}^{M} m_{I_k\times\Omega} \cdot \mathbf{1}_{[a_k,b_k)\times\Omega} + \sum_{k=M+1}^{2M+1} (f+\xi) \cdot \mathbf{1}_{[a_k,b_k)\times\Omega}, \ \ |m_{I_k}| \le c\Big\}$$

is relatively compact.

Step 2:

From Lemma 7 we conclude that for any function $f \in L^1\big([0,1)\big)$, any continuous random variable $\xi : \Omega \to \mathbb{R}$ with $0 < \mathbb{E}|\xi| < \infty$, all partitions P of the interval $[0,1]$ and any $m_{I\times\Omega} \in med_{I\times\Omega}(f+\xi)$ the following result is valid:

$$\forall \varepsilon > 0 \ \ \exists \delta > 0 : \quad \sup_{P \in \mathcal{P}} \sum_{I \in P,\, l(I) < \delta} \big\| m_{I\times\Omega} \cdot \mathbf{1}_{I\times\Omega} - (f \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{I\times\Omega} \ \| < \varepsilon,$$

or, equivalently:

$$\forall \varepsilon > 0 \ \ \exists \delta > 0 : \quad \sup_{P \in \mathcal{P}} \sum_{I \in P,\, l(I) < \delta} \big\| m_{I\times\Omega} \cdot \mathbf{1}_{I\times\Omega} - (f + \xi) \cdot \mathbf{1}_{I\times\Omega} \ \| < \varepsilon. \tag{3.18}$$

Consider an arbitrary $\varepsilon > 0$ and a fixed δ such that inequality (3.18) is satisfied.
Further on, for any arbitrary partition P of the interval $[0,1)$ we define the function $\tilde{h}_P$ as follows:

$$\tilde{h}_P(x) := \sum_{I \in P,\, l(I) \ge \delta} m_{I\times\Omega} \cdot \mathbf{1}_{I\times\Omega}(x) \sum_{I \in P,\, l(I) < \delta} (f+\xi) \cdot \mathbf{1}_{I\times\Omega}(x) \tag{3.19}$$

Now we choose $M = \dfrac{1}{\delta}$, $c = \|f + \xi\| = \|f \otimes 1 + 1 \otimes \xi\|$, fix an arbitrary partition P of the interval $[0,1)$ and show that the corresponding function $\tilde{h}_P$ belongs to the set $\mathcal{F}_M^c$.

For this purpose we have to rewrite (3.19) in such a way that the following two properties are satisfied :

1) The number of summands in the first sum of (3.19) is not larger than M.

2) The number of summands in the second sum of (3.19) is not larger than $M+1$.

Firstly we show the validity of property **1)**.
P is some partition of the interval $[0,1)$, i.e $\sum\limits_{I\in P,\, l(I)\geq\delta} l(I) \leq 1 = l\big([0,1)\big)$.
On the other hand it holds:

$$\sum_{I\in P,\, l(I)\geq\delta} l(I) \geq \sum_{I\in P,\, l(I)\geq\delta} \delta = \delta \sum_{I\in P,\, l(I)\geq\delta} 1 = \delta\cdot\#\big\{I\in P:\ l(I)\geq\delta\big\}.$$

Therefore we obtain the following:

$$\delta\cdot\#\big\{I\in P:\ l(I)\geq\delta\big\} \leq \sum_{I\in P,\, l(I)\geq\delta} l(I) \leq 1 \tag{3.20}$$

$$\Rightarrow \qquad \#\big\{I\in P:\ l(I)\geq\delta\big\} \leq \frac{1}{\delta} = M. \tag{3.21}$$

This implies that the number of summands in the first sum of (3.19) is not larger than $M = \dfrac{1}{\delta}$. The proof of property **1)** is complete.

Next we transform (3.19) in such away that the number of summands in the second sum is not larger than $M+1$.
For the sake of convenience we denote the intervals $I \in P$, which fulfill $l(I) < \delta$, as *short* intervals and the intervals $I\in P$, which obey $l(I)\geq\delta$, as *long* intervals.
Without any loss of generality we consider the case of $\#\big\{I\in P:\ l(I)\geq\delta\big\} = M$. Further on, let the partition P be of the form $P=\big\{I_1, I_2, \dots, I_N\big\}$, where the intervals $I_{l_1}, I_{l_2}, \dots, I_{l_M}$ are long. This means l_k $(k\in\{1,\dots,M\})$ is the successive number of the k-th long interval. Let also $l_0 := 0$, $l_{M+1}-1 := N$. Next introduce the definition of the consecutively unified short intervals $\{\mathcal{I}_k\}_{k=1}^{M+1}$:

$$\mathcal{I}_1 := \begin{cases} \emptyset & : \ l_1 = 1 \\ \bigcup\limits_{i=l_0+1}^{l_1-1} I_i & : \ \text{otherwise} \end{cases}, \qquad \mathcal{I}_{M+1} := \begin{cases} \emptyset & : \ l_M = N \\ \bigcup\limits_{i=l_M+1}^{l_{M+1}-1} I_i & : \ \text{otherwise} \end{cases}$$

$$\mathcal{I}_k := \begin{cases} \emptyset & : \ l_{k-1} = i \ \wedge \ l_k = i+1, \\ & \quad \forall i\in\{1,\dots,N-1\} \\ \bigcup\limits_{i=l_{k-1}+1}^{l_k-1} I_i & : \ \text{otherwise} \end{cases}, \quad \forall\, k = 2,\dots,M$$

The last definition can be understood as follows.

- If the first interval I_1 of the partition P is long, then we define $\mathcal{I}_1 = \emptyset$, otherwise $\mathcal{I}_1$ is set to the union of all short intervals up to the first long interval.
- For the following $\mathcal{I}_k$ $(k = 2,\dots,M)$, if any two long intervals follow one another, then we set $\mathcal{I}_k = \emptyset$, otherwise $\mathcal{I}_k$ is again set to the union of all short intervals up to the next long interval.

- Finally, if the last interval I_N of the partition P is long, then we set $\mathcal{I}_{M+1} = \emptyset$, otherwise $\mathcal{I}_{M+1}$ is the union of all short intervals which follow the last long interval in the partition P.

Now we can transform the second sum of (3.19) as follows:

$$\sum_{I\in P,\, l(I)<\delta} (f+\xi)\cdot \mathbf{1}_{I\times\Omega} = \sum_{\mathcal{I}_k\neq\emptyset,\ k=\overline{1,M+1}} (f+\xi)\cdot \mathbf{1}_{\mathcal{I}_k\times\Omega}. \tag{3.22}$$

From the definition of the consecutively unified short intervals it follows that $\#\big\{\mathcal{I}_k :\ \mathcal{I}_k \neq \emptyset,\ k=1,\dots,M+1\big\} \le M+1$. This together with (3.22) allows us to deduce that the number of summands of the kind $(f+\xi)\cdot\mathbf{1}_{I\times\Omega}$ in (3.19) is not larger than $M+1$. Thus, we have shown that for any partition P of the interval $[0,1)$ the properties **1)** and **2)** are satisfied. This proves that for any $P\in\mathcal{P}$ there exist numbers M and c such that the function $\tilde{h}_P$ belongs to the set $\mathcal{F}^c_M$.

Finally we show that for any function $h\in \operatorname{Med}(f+\xi)$, $h=\sum\limits_{I\in P_{J(h)}} m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega}$, it holds $\|\tilde{h}_{P_{J(h)}}-h\|<\varepsilon$, where $\tilde{h}_{P_{J(h)}}$ is defined according to (3.19).

$$\begin{aligned}
&\big\|\tilde{h}_{P_{J(h)}}-h\big\| \\
&= \Big\|\sum_{I\in P_{J(h)},\, l(I)\ge\delta} m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega} + \sum_{I\in P_{J(h)},\, l(I)<\delta} (f+\xi)\cdot\mathbf{1}_{I\times\Omega} - \sum_{I\in P_{J(h)}} m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega}\Big\| \\
&= \Big\|\sum_{I\in P_{J(h)},\, l(I)\ge\delta} m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega} + \sum_{I\in P_{J(h)},\, l(I)<\delta} (f+\xi)\cdot\mathbf{1}_{I\times\Omega} \\
&\qquad - \sum_{I\in P_{J(h)}\, l(I)<\delta} m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega} - \sum_{I\in P_{J(h)}\, l(I)\ge\delta} m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega}\Big\| \\
&= \Big\|\sum_{I\in P_{J(h)},\, l(I)<\delta} (f+\xi)\cdot\mathbf{1}_{I\times\Omega} - \sum_{I\in P_{J(h)}\, l(I)<\delta} m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega}\Big\| \\
&= \sum_{I\in P_{J(h)},\, l(I)<\delta} \big\|(f+\xi)\cdot\mathbf{1}_{I\times\Omega} - m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega}\big\| \underset{(3.18)}{<} \varepsilon.
\end{aligned}$$

Summarizing, for any positive ε we have found such δ, $M=\dfrac{1}{\delta}$ and $c=\|f+\xi\|$ for which the following statement is valid. For each $h\in\operatorname{Med}(f+\xi)$ there exists a function $\tilde{h}=\tilde{h}_{P_{J(h)}}$ from the relatively compact set $\mathcal{F}^c_M$ such that $\big\|\tilde{h}-h\big\|<\varepsilon$. Therefore, due to Proposition 5, the set $\operatorname{Med}(f+\xi)$ is relatively compact and this completes the proof of Lemma 8.

□

3.2.2 Main result

After we have proved that the set $\operatorname{Med}(f+\xi)$ is relatively compact and that the functional H^∞_γ is lower semicontinuous, we have all the necessary tools in order to proceed to the proof of existence of a minimum of H^∞_γ.

Theorem 4. *Let $f \in L^1\big([0,1)\big)$, $(\Omega,\mathbb{P})$ be a probability space and $\xi : \Omega \to \mathbb{R}$ be a random variable with $0 < \mathbb{E}|\xi| < \infty$ and $\mathbb{P}\{\xi \le 0\} = \mathbb{P}\{\xi \ge 0\} = \frac{1}{2}$. Then for all $\gamma \ge 0$ it holds:*

$$\underset{g\in L^1([0,1))}{\operatorname{argmin}}\, H_\gamma^\infty(g\otimes 1\,, f\otimes 1 + 1\otimes\xi) \ne \emptyset. \tag{3.23}$$

Proof:

We consider three different cases.

1) For $\gamma = 0$ the functional $H_\gamma^\infty(g\otimes 1\,, f\otimes 1 + 1\otimes\xi)$ for all $g \in L^1([0,1))$ reduces to

$$H_\gamma^\infty(g\otimes 1\,, f\otimes 1 + 1\otimes\xi) = \|f\otimes 1 + 1\otimes\xi - g\otimes 1\| - \|f\otimes 1 + 1\otimes\xi\|.$$

It is clear that for $g = f$ the norm $\|f\otimes 1 + 1\otimes\xi - g\otimes 1\| = \|1\otimes\xi\|$. From Proposition 2 we know that $\|f\otimes 1 + 1\otimes\xi - g\otimes 1\| \le \|1\otimes\xi\|$ for all $f \in L^1([0,1))$ and $g \in L^1([0,1))$. This means that for all $g \in L^1([0,1))$ the norm $\|f\otimes 1 + 1\otimes\xi - g\otimes 1\|$ as well as, the functional $H_\gamma^\infty(g\otimes 1\,, f\otimes 1 + 1\otimes\xi)$ achieve its minimal value for $g = f$. Therefore,

$$\underset{g\in L^1([0,1))}{\operatorname{argmin}}\, H_\gamma^\infty(g\otimes 1\,, f\otimes 1 + 1\otimes\xi) = \big\{f\big\} \ne \emptyset.$$

2) Assume $\gamma > 0$ and $g \in S\big([0,1)\big)$ with $\#J(g) > \|f\otimes 1 + 1\otimes\xi\|/\gamma$. This yields:

$$\begin{aligned}
&H_\gamma^\infty(g\otimes 1\,, f\otimes 1 + 1\otimes\xi) = \gamma\#J(g) + \|f\otimes 1 + 1\otimes\xi - g\otimes 1\| - \|f\otimes 1 + 1\otimes\xi\| \\
> \;& \gamma\frac{\|f\otimes 1 + 1\otimes\xi\|}{\gamma} + \|f\otimes 1 + 1\otimes\xi - g\otimes 1\| - \|f\otimes 1 + 1\otimes\xi\| \\
= \;& \|f\otimes 1 + 1\otimes\xi - g\otimes 1\| \ge 0
\end{aligned}$$

On the other hand,

$$H_\gamma^\infty(0\otimes 1\,, f\otimes 1 + 1\otimes\xi) = \|f\otimes 1 + 1\otimes\xi\| - \|f\otimes 1 + 1\otimes\xi\| = 0$$

Therefore, we obtain, that for any $g \in S\big([0,1)\big)$ with $\#J(g) > \|f\otimes 1 + 1\otimes\xi\|/\gamma$ it holds:

$$H_\gamma^\infty(g\otimes 1\,, f\otimes 1 + 1\otimes\xi) > H_\gamma^\infty(0\otimes 1\,, f\otimes 1 + 1\otimes\xi).$$

Consequently, the number of jumps of a minimizer of $H_\gamma^\infty(g\otimes 1\,, f\otimes 1 + 1\otimes\xi)$ cannot exceed $\|f\otimes 1 + 1\otimes\xi\|/\gamma$.

3) Finally let $\gamma > 0$, $g \in S\big([0,1)\big)$ and $\#J(g) \le \|f\otimes 1 + 1\otimes\xi\|/\gamma$.

We have already noticed that due to Proposition 3, instead of showing that $\underset{g\in S([0,1))}{\operatorname{argmin}}\, H_\gamma^\infty(g\otimes 1\,, f\otimes 1 + 1\otimes\xi) \ne \emptyset$ one can equivalently prove that $\underset{h\in S(\Lambda_\Omega)}{\operatorname{argmin}}\, H_\gamma^\infty(h, f\otimes 1 + 1\otimes\xi) \ne \emptyset$ with $\Lambda_\Omega = [0,1)\times\Omega$, $S(\Lambda_\Omega)$ defined by (3.11).

For any $g = \sum_{I\in P_{J(g)}} g_I\cdot\mathbf{1}_I$, which fulfills $\#J(g) \le \|f\otimes 1 + 1\otimes\xi\|/\gamma$, we define the function $(f+\xi)_{J(g)} \in S(\Lambda_\Omega)$ as follows:

$$(f+\xi)_{J(g)} := \sum_{I\in P_{J(g)}} m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega}\,,\quad m_{I\times\Omega} \in med_{I\times\Omega}(f+\xi), \tag{3.24}$$

and consider $H_\gamma^\infty(g\otimes 1\,, f\otimes 1 + 1\otimes\xi)$.

$$
\begin{aligned}
& H_\gamma^\infty(g\otimes 1\,, f\otimes 1+1\otimes\xi) \\
= \quad & \gamma\#J(g)+\|f\otimes 1+1\otimes\xi-g\otimes 1\|-\|f\otimes 1+1\otimes\xi\| \\
= \quad & \gamma\#J(g)+\int\limits_{[0,1)\times\Omega}\big(\,|f(x)+\xi(\omega)-g(x)|-|f(x)+\xi(\omega)|\,\big)\,dx\ d\mathbb{P}(\omega) \\
= \quad & \gamma\#J(g)+\int\limits_{[0,1)\times\Omega}\Big(\Big|f(x)+\xi(\omega)-\sum_{I\in P_{J(g)}}g_I\cdot\mathbf{1}_I(x)\Big|-\Big|f(x)+\xi(\omega)\Big|\Big)\,dx\ d\mathbb{P}(\omega) \\
= \quad & \gamma\#J(g)+\sum_{I\in P_{J(g)}}\int\limits_{I\times\Omega}\Big(\Big|f(x)+\xi(\omega)-g_I\Big|-\Big|f(x)+\xi(\omega)\Big|\Big)\,dx\ d\mathbb{P}(\omega) \\
\underset{(3.4)}{\geq} \quad & \gamma\#J(g)+\sum_{I\in P_{J(g)}}\int\limits_{I\times\Omega}\Big(\Big|f(x)+\xi(\omega)-m_{I\times\Omega}\Big|-\Big|f(x)+\xi(\omega)\Big|\Big)\,dx\ d\mathbb{P}(\omega) \\
= \quad & \gamma\#J(g)+\sum_{I\in P_{J(g)}}\int\limits_{I\times\Omega}\Big(\Big|f(x)+\xi(\omega)-\sum_{I\in P_{J(g)}}m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega}\Big|-\Big|f(x)+\xi(\omega)\Big|\Big)\,dx\ d\mathbb{P}(\omega) \\
\underset{(3.24)}{=} \quad & \gamma\#J\big(\,(f+\xi)_{J(g)}\,\big)+\int\limits_{[0,1)\times\Omega}\Big(\Big|f(x)+\xi(\omega)-(f+\xi)_{J(g)}\Big|-\Big|f(x)+\xi(\omega)\Big|\Big)\,dx\ d\mathbb{P}(\omega) \\
= \quad & \gamma\#J\big(\,(f+\xi)_{J(g)}\,\big)+\|f\otimes 1+1\otimes\xi-(f+\xi)_{J(g)}\|-\|f\otimes 1+1\otimes\xi\| \\
= \quad & H_\gamma^\infty(\,(f+\xi)_{J(g)},\ f\otimes 1+1\otimes\xi\,).
\end{aligned}
$$

This shows that for any $g\in S\big([0,1)\big)$ it holds

$$H_\gamma^\infty(g\otimes 1\,, f\otimes 1+1\otimes\xi)\geq H_\gamma^\infty(\,(f+\xi)_{J(g)},\ f\otimes 1+1\otimes\xi\,).$$

From the definition of $(f+\xi)_{J(g)}$, it is clear that for any $g\in S\big([0,1)\big)$, which satisfies the condition $\#J(g)\leq\|f\otimes 1+1\otimes\xi\|/\gamma$, the function $(f+\xi)_{J(g)}$ belongs to the set $\mathrm{Med}'(f+\xi)=\Big\{h\in L^1(\Lambda_\Omega)\Big|\,\exists P\in\mathcal{P}:h=\sum\limits_{I\in P}m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega}$ with $\#J(h)\leq\|f+\xi\|,\ \forall I\in P\ \ m_{I\times\Omega}\in med_{I\times\Omega}(f+\xi)\Big\}$.

Due to Lemma 8, the set

$$\mathrm{Med}(f+\xi)=\Big\{h\in L^1\big(\Lambda_\Omega\big)\,\Big|\,\exists P\in\mathcal{P}:h=\sum_{I\in P}m_{I\times\Omega}\cdot\mathbf{1}_{I\times\Omega},\ \forall I\in P\ \ m_{I\times\Omega}\in med_{I\times\Omega}(f+\xi)\Big\}$$

is relatively compact. Consequently the set $\mathrm{Med}'(f+\xi)$ is also relatively compact. Moreover, Proposition 4 yields that for any $\gamma\geq 0$ the functional H_γ^∞ is lower semicontinuous. Finally, according to Theorem 2, which states the existence of the minimizer of a lower semicontinuous functional considered on a relatively compact set, we conclude: $\underset{h\in S(\Lambda_\Omega)}{\operatorname{argmin}}\,H_\gamma^\infty(h,f\otimes 1+1\otimes\xi)\neq\emptyset$, as well as, $\underset{g\in S([0,1))}{\operatorname{argmin}}\,H_\gamma^\infty(g\otimes 1\,,f\otimes 1+1\otimes\xi)\neq\emptyset$ for any $\gamma>0$ and all $g\in S\big([0,1)\big)$ with $\#J(g)\leq\|f\otimes 1+1\otimes\xi\|/\gamma$.
This completes the proof of Lemma 4.

□

3.3 Conclusion

The main result of this chapter can be formulated as the fact of existence of a minimizer $\widehat{g} \in L^1([0,1))$ of the functional $H_\gamma^\infty(\cdot \otimes 1\,, f \otimes 1 + 1 \otimes \xi)$ for and all $\gamma \geq 0$, any $f \in L^1\big([0,1)\big)$ and any continuous random variable $\xi : (\Omega, \mathbb{P}) \to \mathbb{R}$ with $\mathbb{P}\{\xi \leq 0\} = \mathbb{P}\{\xi \geq 0\} = \dfrac{1}{2}$ and $0 < \mathbb{E}|\xi| < \infty$. Thereby the functional $H_\gamma^\infty(g \otimes 1\,, f \otimes 1 + 1 \otimes \xi)$, as shown in Chapter 2, is the limit functional of the sequence of functionals $\big(\tilde{H}_{\gamma_n}^n(\cdot,\, f^n + \xi^n)\big)_{n \in \mathbb{N}}$ for infinitely large n, for all sequences γ_n such that $\gamma_n \underset{n\to\infty}{\longrightarrow} \gamma$, for any $f \in L^1\big([0,1)\big)$ and random noise $(\xi_i^n)_{1 \leq i \leq n,\, n \in \mathbb{N}}$, which fulfills Condition 1, according to which, the random variables ξ and $\xi_1^n, \ldots, \xi_n^n$ are independent, identically distributed and $\mathbb{E}|\xi| = \mathbb{E}|\xi_1^n| = \ldots = \mathbb{E}|\xi_n^n| < \infty$. Moreover, we have shown that for all $\gamma > 0$ and for any $\widehat{g} \in \underset{g \in S([0,1))}{\operatorname{argmin}} H_\gamma^\infty(g \otimes 1\,, f \otimes 1 + 1 \otimes \xi)$ there exists a partition $P_{J(\widehat{g})}$ of the interval $[0,1)$ such that

$$\widehat{g}(x) = \sum_{I \in P_{J(\widehat{g})}} \operatorname{med}_{I \times \Omega}(f + \xi) \cdot \mathbf{1}_I(x),$$

where $\operatorname{med}_{I \times \Omega}(f + \xi) \in \underset{c \in \mathbb{R}}{\operatorname{argmin}} \|c \cdot \mathbf{1}_{I \times \Omega} - (f \otimes 1 + 1 \otimes \xi) \cdot \mathbf{1}_{I \times \Omega}\|$.

Chapter 4

Properties of Pareto distributed noise

Up until now we have been dealing with the noise term, which satisfied Condition 1. Rephrasing, one can say that the distribution of the noise was assumed to be centered around zero. Condition 1 was sufficient in order to show that for an arbitrary function $f \in L^1\big([0,1)\big)$ there exist two distinct minimizers, each minimizing the respective functional: (i) a minimizer of the functional $\tilde{H}^n_{\gamma_n}(\cdot\,, f^n + \xi^n)$ with any large n and any non-negative γ_n, and, (ii) a minimizer of the limiting functional $H^\infty_\gamma(\cdot\,, f \otimes 1 + 1 \otimes \xi)$ with any $\gamma \geq 0$. Our next goal is to show that the minimizers of the functional $\tilde{H}^n_{\gamma_n}(\cdot\,, f^n + \xi^n)$ converge to the original function f when $n \to \infty$
To this end we need to specify the distribution function of the noise. As we have pointed out before, the case of the L^1 functionals stands out from the well-studied case of the L^2 functionals, because it is more robust when treating data with outliers. This property has been shown to be the direct consequence of the fact that a median, as computed from a set of a given data, is more stable to the outliers in the data than the average value. $f(x) = \dfrac{\alpha - 1}{2\,(1+|x|)^\alpha}$, $\alpha > 1$.
However we will actually work with $\alpha \geq 9$, as this chose of α is important for the validity of Lemma 11, which will be discussed later in this Chapter.

Specifically, throughout the following chapters the following condition on the noise $(\xi^n_i)_{1\leq i\leq n,\, n\in\mathbb{N}}$ will be used.

Condition 2. *The triangular array $(\xi^n_i)_{1\leq i\leq n,\, n\in\mathbb{N}}$ on a probability space $(\Omega, \mathbb{P})$ fulfills the properties:*
for all $n \in \mathbb{N}$ the random variables $\left(\xi^n_i\right)_{1\leq i\leq n}$ are independent identically distributed with respect to the symmetrical Pareto distribution with parameter $\alpha \geq 9$, this means the distribution function $F_{\xi^n_i}$ and the probability density $f_{\xi^n_i}$ of ξ^n_i are defined as follows:

$$F_{\xi^n_i}(x) \;=\; \begin{cases} \dfrac{1}{2\,(1-x)^{\alpha-1}} & ,\quad x < 0 \\ 1 - \dfrac{1}{2\,(1+x)^{\alpha-1}} & ,\quad x \geq 0 \end{cases} \tag{4.1}$$

$$f_{\xi^n_i}(x) \;=\; \begin{cases} \dfrac{\alpha-1}{2\,(1-x)^{\alpha}} & ,\quad x < 0 \\ \dfrac{\alpha-1}{2\,(1+x)^{\alpha}} & ,\quad x \geq 0 \end{cases} \tag{4.2}$$

It is obvious that for any symmetrical Pareto distributed random variable ξ, as defined in (4.1), it holds $F_\xi(x) = 1 - F_\xi(-x)$ or equivalently $\mathbb{P}\big(\xi \le 0\big) = \mathbb{P}\big(\xi \ge 0\big) = \frac{1}{2}$. Moreover, as it is shown in Appendix A , Eq. (A.47), p. 130, it also holds $\mathbb{E}|\,\xi\,| < \infty$. Consequently, it is clear that the noise, which fulfills Condition 2 also fulfills Condition 1.

Further on in this chapter we will investigate the properties of the noise, which fulfills Condition 2. Additionally, some helpful properties of a random variable, which is distributed according to (4.1), are summarized in Appendix A (see pp.130-135).

4.1 Boundedness of ξ^n

The first lemma of this chapter shows that for the noise $(\xi_i^n)_{1\le i\le n,\, n\in\mathbb{N}}$, which satisfies the Condition 2, the function $\xi^n = \sum\limits_{i=1}^{n} \xi_i^n \cdot \mathbf{1}_{\left[\frac{i-1}{n},\, \frac{i}{n}\right)}$ for all $n \in \mathbb{N}$ is a bounded function from the set $S_n\big([0,\,1)\big)$ in the sense that the norm of this function is $\mathbb{P}$-almost surely a finite number.

Lemma 9. *Let $(\xi_i^n)_{1\le i\le n,\, n\in\mathbb{N}}$ fulfill Condition 2. Then there exists a finite random constant C such that $\mathbb{P}$-almost surely for all $n \in \mathbb{N}$ it holds:*

$$\|\xi^n\| \le \widehat{C}. \tag{4.3}$$

Proof:
Using the definition of the norm and the fact that $\xi^n = \sum\limits_{i=1}^{n} \xi_i^n \cdot \mathbf{1}_{\left[\frac{i-1}{n},\, \frac{i}{n}\right)}$ is a step-function from $S_n\big([0\,1)\big)$ for all $n \in \mathbb{N}$, it is clear that

$$\|\xi^n\| = \frac{1}{n}\sum_{i=1}^{n} |\xi_i^n|.$$

According to Borel-Cantelli Lemma [14], in order to prove Lemma 9 we have to show the validity of the following property:

$$\sum_{n=1}^{\infty} P\Big(\|\xi^n\| > \widehat{C}_0\Big) < \infty \quad \text{for some } \ \widehat{C}_0 < \infty. \tag{4.4}$$

We choose $\widehat{C}_0 > \dfrac{1}{\alpha - 2}$ and consider the left-hand side of (4.4):

$$\begin{aligned}
&\sum_{n=1}^{\infty} P\left(\|\xi^n\| > \widehat{C}_0\right) = \sum_{n=1}^{\infty} P\left(\frac{1}{n}\sum_{i=1}^{n} |\xi_i^n| > \widehat{C}_0\right) \\
= \;& \sum_{n=1}^{\infty} P\left(\sum_{i=1}^{n} \big(|\xi_i^n| - \mathbb{E}\big(|\xi_i^n|\big)\big) > n\left(\widehat{C}_0 - \mathbb{E}\big(|\xi_i^n|\big)\right)\right) \\
\le \;& \sum_{n=1}^{\infty} \frac{\mathbb{E}\left|\sum\limits_{i=1}^{n} \big(|\xi_i^n| - \mathbb{E}\big(|\xi_i^n|\big)\big)\right|^3}{n^3\left(\widehat{C}_0 - \mathbb{E}\big(|\xi_1^n|\big)\right)^3}
\end{aligned} \tag{4.5}$$

The last inequality in formula (4.5) holds due to the well know standard Markov's inequality (see for example [14]).

Additionally we note that $\big(|\xi_1^n| - \mathbb{E}\big(|\xi_1^n|\big)\big), \ldots, \big(|\xi_n^n| - \mathbb{E}\big(|\xi_n^n|\big)\big)$ are independent random variables with zero means. Moreover, it is easy to show that $\mathbb{E}|\xi_1^n| = \ldots = \mathbb{E}|\xi_n^n| = \dfrac{1}{\alpha - 2}$ (see (A.47), Appendix A, p.130). Finally, it is easy to see that $\mathbb{E}|\xi_i^n|^p = 2 \int\limits_{[0,\infty)} \dfrac{x^p\,(\alpha-1)}{2\,(1+x)^\alpha}$ is finite for $\alpha - p > 1$ and for all $i \in \{1, \ldots n\}$. Consequently, according to Petrov [33, p. 60], for all $2 \le p < \alpha - 1$, all $n \in \mathbb{N}$ and all $i \in \{1, \ldots, n\}$ the following inequality is valid:

$$\mathbb{E}\Big|\sum_{i=1}^{n} \big(|\xi_i^n| - \mathbb{E}\big(|\xi_i^n|\big)\big)\Big|^p \le \tilde{C}_p\, n^{p/2-1} \sum_{i=1}^{n} \mathbb{E}\big|\,|\xi_i^n| - \mathbb{E}\big(|\xi_i^n|\big)\big|^p, \tag{4.6}$$

where

$$\tilde{C}_p = \frac{1}{2}\, p\,(p-1)\, \max\{1, 2^{p-3}\} \Big(1 + \frac{2}{p}\, K_{2m}^{(p-2)/2m}\Big),$$

the integer m satisfies the condition $2m \le p < 2m+2$ and the constant is defined by $K_{2m} = \sum\limits_{r=1}^{m} \dfrac{r^{2m-1}}{(r-1)!}$.

Due to Condition 2, the parameter α is greater than 9, therefore for all $n \in \mathbb{N}$ and all $i \in \{1, \cdots, n\}$ the value $\mathbb{E}\big|\,|\xi_i^n| - \mathbb{E}\big(|\xi_i^n|\big)\big|^3$ is finite. Moreover, for $p = 3$ it is clear that $m = 1$, $K_{2m} = 1$ and $\tilde{C}_p = 5$.
Hence inequality (4.6) for $p = 3$ becomes:

$$\begin{aligned} \mathbb{E}\Big|\sum_{i=1}^{n} \big(|\xi_i^n| - \mathbb{E}\big(|\xi_i^n|\big)\big)\Big|^3 &\le 5\, n^{1/2} \sum_{i=1}^{n} \mathbb{E}\big|\,|\xi_i^n| - \mathbb{E}\big(|\xi_i^n|\big)\big|^3 \\ &= 5\, n^{3/2} \mathbb{E}\big|\,|\xi_1^n| - \mathbb{E}\big(|\xi_1^n|\big)\big|^3. \end{aligned} \tag{4.7}$$

After substituting (4.7) in (4.5) and also due to the fact that $\mathbb{E}\big(|\xi_i^n|\big) = \dfrac{1}{\alpha-2}$ for any $i \in \{1, \ldots, n\}$, we obtain

$$\sum_{n=1}^{\infty} P\Big(\|\xi^n\| > \widehat{C}_0\Big) \le \sum_{n=1}^{\infty} \frac{5\,\mathbb{E}\big|\,|\xi_1^n| - \mathbb{E}\big(|\xi_1^n|\big)\big|^3}{n^{3/2}\,\big(\widehat{C}_0 - \frac{1}{\alpha-2}\big)^3}.$$

Taking into account that $\mathbb{E}\big|\,|\xi_i^n| - \mathbb{E}\big(|\xi_i^n|\big)\big|^3$ are finite for $\alpha \ge 9$, $\widehat{C}_0 > \dfrac{1}{\alpha-2}$ and $\sum\limits_{n=1}^{\infty} \dfrac{1}{n^{3/2}} < \infty$, we conclude that $\sum\limits_{n=1}^{\infty} P\Big(\|\xi^n\| > \widehat{C}_0\Big) < \infty$. Hence (4.4) is valid, which proves the validity of (4.3).

□

4.2 Quantiles properties

We continue this chapter with a brief account of the properties of the medians and other quantiles of the symmetrical Pareto distributed noise.
However, first, we recall the definition of a quantile for an arbitrary random variable and for a random sample.

Definition 8. *[14] Let η be an arbitrary random variable. The number $\widetilde{\eta}_q$ will be called q-quantile of η, where $q \in (0,1)$, if it holds*

$$\mathbb{P}(\eta \leq \widetilde{\eta}_q) \geq q \quad \textit{and} \quad \mathbb{P}(\eta \geq \widetilde{\eta}_q) \geq 1-q.$$

Here we also notice that if η is a continuous random variable, symmetrical distributed with the density function f_η, i.e. if $f_\eta(x) = f_\eta(-x)$, then $\widetilde{\eta}_q$ satisfies the following obvious properties:

$$\mathbb{P}(\eta \leq \widetilde{\eta}_q) = q \quad \text{and} \quad \mathbb{P}(\eta \geq \widetilde{\eta}_q) = 1-q \tag{4.8}$$

and

$$\widetilde{\eta}_{(1-q)} = -(\widetilde{\eta}_q). \tag{4.9}$$

Definition 9. *[14] Let $\eta_1, \ldots, \eta_n$, $n \in \mathbb{N}$, be a random sample. A number* q-quantile$(\eta_1, \ldots, \eta_n)$ *is q-quantile of the data* $\eta_1, \ldots, \eta_n$*, if at least* $q \cdot 100\%$ *of the data is less than or equal to* q-quantile$(\eta_1, \ldots, \eta_n)$ *and at least* $(1-q) \cdot 100\%$ *of the data is greater than or equal to* q-quantile$(\eta_1, \ldots, \eta_n)$*.*

Any q-quantile of the data set has an obvious property, which will be used later on:

$$q\text{-quantile}\,(-\eta_1, \ldots, -\eta_n) = -\,(1\text{-}q)\text{-quantile}\,(\eta_1, \ldots, \eta_n). \tag{4.10}$$

This can be explained as follows: if a real number is the q-quantile of the data $(\eta_1, \ldots, \eta_n)$, then the same number, taken with the opposite sign, is a $(1-q)$-quantile of the data $(-\eta_1, \ldots, -\eta_n)$.

Notice that similarly to the median, i.e. the 0.5-quantile, the general q-quantile $(q \in (0,1))$ of an arbitrary sample is not uniquely defined. Due to this fact, the following important remark should be made.

Remark 3. *Recall that for any sample of data $\eta_1, \cdots \eta_n$ by $med(\eta_1, \ldots, \eta_n)$ we will denote the set of all medians of the data $(\eta_1, \ldots, \eta_n)$. For the sake of brevity, all the results presented below are formulated for a single median* $\mathrm{med}(\cdots)$ *which is taken from the set $med(\cdots)$ of all medians of the same data points. It should be emphasized however that the choice of this median is arbitrary, implying that all the results remain valid for any median* $\mathrm{med}(\cdots)$ *from $med(\cdots)$.*
This remark is generalized to an arbitrary quantile, i.e. we understand that the validity of the results, which involve a q-quantile q, remain unchanged for any such quantile from the corresponding set of all quantile of this order.

We proceed this section by formulating a technical lemma, which will help us to prove the subsequent lemmas of this section. Before stating the lemma we introduce one new definition.

Definition 10. *For some arbitrary random variable X and one arbitrary real constant c we define the random variable δ_X^c as follows:*

$$\delta_X^c := \mathbf{1}_{(c,+\infty)}(X) = \begin{cases} 1 & , \quad X > c \\ 0 & , \quad \textit{else} \end{cases} \tag{4.11}$$

It is obvious that the random variable δ_X^c is distributed according to the Bernoulli distribution with parameter $\mathbb{P}(X > c) := P_X^c$.

Lemma 10. *Let $\left(\eta_i^n\right)_{n\in\mathbb{N},\,1\le i\le n}$ be an arbitrary triangular array of independent identically distributed random variables. If for some real constant c it holds:*

$$\sum_{n=1}^{\infty}\sum_{l=1}^{n}\sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta^c_{\eta_k^n}\ge\frac{l}{2}\Big)<\infty, \tag{4.12}$$

where $\delta^c_{\eta_k^n}$ is defined according to (4.11), then $\mathbb{P}$-almost surely for all sufficiently large $n\in\mathbb{N}$, for all $l\in\{1,\dots,n\}$ and all $i\in\{0,\dots,n-l\}$ it holds:

$$\operatorname{med}(\,\eta_{i+1}^n,\dots,\eta_{i+l}^n\,)\le c. \tag{4.13}$$

Proof:
In order to show the validity of (4.13), using Borel-Cantelli Lemma [14], it is sufficient to show that

$$\sum_{n=1}^{\infty}\mathbb{P}\Big(\exists\, l\in\{1,\dots,n\},\,\exists\, i\in\{0,\dots,n-l\}\ :\operatorname{med}\left(\eta_{i+1}^n,\dots,\eta_{i+l}^n\right)>c\Big)<\infty, \tag{4.14}$$

For some fixed constant c, for all $n\in\mathbb{N}$ and all $l\in\{1,\dots,n\}$ we consider $\mathbb{P}\Big(\exists\, i\in\{0,\dots,n-l\}\ :\operatorname{med}(\,\eta_{i+1}^n,\dots,\eta_{i+l}^n\,)>c\Big)$:

$$\begin{aligned}
&\mathbb{P}\Big(\exists\, i\in\{0,\dots,n-l\}\ :\operatorname{med}\left(\,\eta_{i+1}^n,\dots,\eta_{i+l}^n\,\right)>c\Big)\\
=\;&\mathbb{P}\Big(\exists\, i\in\{0,\dots,n-l\}\ :\#\big\{k\in\{i+1,\dots,i+l\}\ :\eta_k^n>c\big\}\ge\frac{l}{2}\Big)\\
\underset{(4.11)}{=}\;&\mathbb{P}\Big(\exists\, i\in\{0,\dots,n-l\}\ :\sum_{k=i+1}^{i+l}\delta^c_{\eta_k^n}\ge\frac{l}{2}\Big).
\end{aligned} \tag{4.15}$$

With (4.15) in mind we consider the left-hand side of (4.14)

$$\begin{aligned}
&\sum_{n=1}^{\infty}\mathbb{P}\Big(\exists\, l\in\{1,\dots,n\},\,\exists\, i\in\{0,\dots,n-l\}\ :\operatorname{med}\left(\eta_{i+1}^n,\dots,\eta_{i+l}^n\right)>c\Big)\\
=\;&\sum_{n=1}^{\infty}\mathbb{P}\left(\bigcup_{l=1}^{n}\Big\{\exists\, i\in\{0,\dots,n-l\}\ :\operatorname{med}\left(\,\eta_{i+1}^n,\dots,\eta_{i+l}^n\,\right)>c\Big\}\right)\\
\le\;&\sum_{n=1}^{\infty}\sum_{l=1}^{n}\mathbb{P}\Big(\,\exists\, i\in\{0,\dots,n-l\}\ :\operatorname{med}\left(\,\eta_{i+1}^n,\dots,\eta_{i+l}^n\,\right)>c\,\Big)\\
\underset{(4.15)}{=}\;&\sum_{n=1}^{\infty}\sum_{l=1}^{n}\mathbb{P}\Big(\,\exists\, i\in\{0,\dots,n-l\}\ :\sum_{k=i+1}^{i+l}\delta^c_{\eta_k^n}\ge\frac{l}{2}\Big)\\
=\;&\sum_{n=1}^{\infty}\sum_{l=1}^{n}\mathbb{P}\left(\Big(\sum_{k=1}^{l}\delta^c_{\eta_k^n}\ge\frac{l}{2}\Big)\cup\ldots\cup\Big(\sum_{k=n-l+1}^{n}\delta^c_{\eta_k^n}\ge\frac{l}{2}\Big)\right)\\
=\;&\sum_{n=1}^{\infty}\sum_{l=1}^{n}\mathbb{P}\left(\bigcup_{i=0}^{n-l}\Big(\sum_{k=i+1}^{i+l}\delta^c_{\eta_k^n}\ge\frac{l}{2}\Big)\right)\\
\le\;&\sum_{n=1}^{\infty}\sum_{l=1}^{n}\sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta^c_{\eta_k^n}\ge\frac{l}{2}\Big).
\end{aligned}$$

Thus condition (4.14) can be equivalently rewritten as follows:

$$\sum_{n=1}^{\infty}\sum_{l=1}^{n}\sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta^c_{\eta^n_k}\geq\frac{l}{2}\Big)<\infty. \tag{4.16}$$

This implies that (4.16), as well as (4.14), are sufficient conditions for the validity of (4.13). Therefore Lemma 10 is proved.

□

The following two corollaries generalize Lemma 10 to other q-quantiles of the noise as well as to the so called mixed median, i.e. the median of the set of data points, some of which are shifted by a certain positive constant. The proofs of these corollaries are omitted, as they are similar to the proof to Lemma 10.

Corollary 2. *Let $\left(\eta^n_i\right)_{n\in\mathbb{N},\,1\leq i\leq n}$ be an arbitrary triangular array of independent identically distributed random variables. If for some real constant c and for any $q\in(0,1)$ it holds:*

$$\sum_{n=1}^{\infty}\sum_{l=1}^{n}\sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta^c_{\eta^n_k}\geq(1-q)\cdot l\Big)<\infty, \tag{4.17}$$

where $\delta^c_{\eta^n_k}$ is defined according to (4.11), then $\mathbb{P}$-almost surely for all sufficiently large $n\in\mathbb{N}$, all $l\in\{1,\ldots,n\}$ and all $i\in\{0,\ldots n-l\}$ it holds :

$$q\text{-quantile}\,(\,\eta^n_{i+1},\ldots,\eta^n_{i+l}\,)\leq c. \tag{4.18}$$

□

Corollary 3. *Let $\left(\eta^n_i\right)_{n\in\mathbb{N},\,1\leq i\leq n}$ be an arbitrary triangular arrays of independent identically distributed random variables. If for some real constant c and for some $\Delta>0$ it holds:*

$$\sum_{n=1}^{\infty}\sum_{l=l_1}^{l_2}\sum_{k=k_1}^{n-l}\sum_{i=0}^{n-l-k}\mathbb{P}\Big(\sum_{j=i+1}^{i+k}\delta^c_{\eta^n_j}+\sum_{s=i+k+1}^{i+k+l}\delta^c_{\eta^n_s+\Delta}\geq\frac{k+l}{2}\Big)<\infty, \tag{4.19}$$

where $1\leq l_1<l_2\leq n,\ 1\leq k_1\leq n-l$ and $\delta^c_{\eta^n_k}$ or $\delta^c_{\eta^n_k+\Delta}$ are defined according to (4.11), then $\mathbb{P}$-almost surely for all sufficiently large $n\in\mathbb{N}$, all $l\in\{l_1,\ldots,l_2\}$, all $k\in\{k_1,\ldots,n-l\}$ and all $i\in\{0,\ldots,n-l-k\}$ it holds:

$$\mathrm{med}(\,\eta^n_{i+1},\ldots,\eta^n_{i+k},\,\eta^n_{i+k+1}+\Delta,\ldots,\eta^n_{i+k+l}+\Delta)\leq c. \tag{4.20}$$

□

4.2.1 Estimate of median

The next lemmas yield estimates of a median of the noise, which satisfy Condition 2. Recall that the array of the noise data $(\xi^n_i)_{1\leq i\leq n,\,n\in\mathbb{N}}$ is triangular. That means, every single row of this array represents sample from the Pareto distributed random variables, with the number of elements given by n. For an arbitrary sufficiently large n, consider a median of a certain subset of this set of noise data, which contains $l\leq n$ elements. Obviously, if one fixes l and increases n, the median of l elements stays not necessarily bounded, implying

that the median diverges as n tends to infinity. On the contrary, if the number l of elements in the subset is of the order n and n tends to infinity, then in analogy to the law of large numbers, it is reasonable to assume that the limiting value of the median tends to the median of the Pareto distributed random variable ξ, i.e. it tends to zero.
In order to make any assumptions on the rate of convergence of the median to zero, we recall the strong law of large numbers, according to which, the average of a given set of i.i.d. data converges to the mean no faster than $(1/\sqrt{n})$, where n represents the number of elements in the corresponding set (see [4]).
This motivates the following assumption on the upper bound of the median.

Lemma 11. *Let $(\xi_i^n)_{1\leq i\leq n,\, n\in\mathbb{N}}$ fulfill Condition 2. Then $\mathbb{P}$-almost surely for all sufficiently large $n \in \mathbb{N}$, all $j \in \{1,\ldots,n\}$ and all $i \in \{1,\ldots,j\}$ it holds:*

$$\text{med}(\,\xi_i^n,\ldots,\xi_j^n\,) \leq \frac{\sqrt{n}\,\ln n}{j-i+1}\ . \tag{4.21}$$

Proof:
Let us define

$$C(n) := \sqrt{n}\,\ln n. \tag{4.22}$$

The statement of Lemma 11 can be equivalently rewritten in the following way.
$\mathbb{P}$-almost surely for all sufficiently large $n \in \mathbb{N}$, all $l \in \{1,\ldots,n\}$ and $i \in \{1,\ldots,n-l+1\}$ it holds:

$$\text{med}(\,\xi_{i+1}^n,\ldots,\xi_{i+l}^n\,) \leq \frac{C(n)}{l}\ . \tag{4.23}$$

Using Lemma 10 we will prove the last inequality and respectively Lemma 11, if we will show that $(\xi_i^n)_{1\leq i\leq n,\, n\in\mathbb{N}}$ satisfy the following property:

$$\sum_{n=1}^{\infty}\sum_{l=1}^{n}\sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi_k^n}^{C(n)/l} \geq \frac{l}{2}\Big) < \infty, \tag{4.24}$$

where, according to (4.11), $\delta_{\xi_k^n}^{C(n)/l}$ for all possible n, l, i and k is Bernoulli distributed random variable with $\mathbb{P}\Big(\delta_{\xi_k^n}^{C(n)/l} = 1\Big) = \mathbb{P}\Big(\xi_k^n > \dfrac{C(n)}{l}\Big)$ and $\sum\limits_{k=i}^{i+l-1}\delta_{\xi_k^n}^{C(n)/l}$ is a binomial distributed random variable with the parameters l and $\mathbb{P}\Big(\xi_k^n > \dfrac{C(n)}{l}\Big)$.
Let us introduce the following new notations:

$$p_{l,n} := \mathbb{P}\Big(\xi_1^n > \frac{C(n)}{l}\Big) = \ldots = \mathbb{P}\Big(\xi_n^n > \frac{C(n)}{l}\Big) \tag{4.25}$$

and

$$X_1^l := \sum_{k=1}^{l}\delta_{\xi_k^n}^{C(n)/l} \sim Bin\big(l,\, p_{l,n}\big), \tag{4.26}$$

then for all $n \in \mathbb{N}$ we have:

$$\sum_{l=1}^{n}\sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi_k^n}^{C(n)/l} \geq \frac{l}{2}\Big) \leq \sum_{l=1}^{n}(n-l+1)\,\mathbb{P}\Big(X_1^l \geq l/2\Big). \tag{4.27}$$

Due to inequality (A.26) in Lemma 23 of Appendix A, p.121, for all $n \in \mathbb{N}$ and all $l \leq n$ the probability $\mathbb{P}\Big(X_1^l \geq \dfrac{l}{2}\Big)$ can be estimated as follows:

$$\mathbb{P}\Big(X_1^l \geq l/2\Big) \leq 2^l \big((1-p_{l,n})\, p_{l,n}\big)^{l/2}.$$

Therefore inequality (4.27) can be rewritten as follows:

$$\sum_{l=1}^{n} \sum_{i=0}^{n-l} \mathbb{P}\Big(\sum_{k=i+1}^{i+l} \delta_{\xi_k^n}^{C(n)/l} \geq \frac{l}{2}\Big) \leq \sum_{l=1}^{n} (n-l+1)\, 2^l \big((1-p_{l,n})\, p_{l,n}\big)^{l/2}, \quad n \in \mathbb{N}. \qquad (4.28)$$

$\Rightarrow$

$$\begin{aligned}
\sum_{n=1}^{\infty}\sum_{l=1}^{n} \sum_{i=0}^{n-l} \mathbb{P}\Big(\sum_{k=i+1}^{i+l} \delta_{\xi_k^n}^{C(n)/l} \geq \frac{l}{2}\Big) &\underset{(4.28)}{\leq} \sum_{n=1}^{\infty}\sum_{l=1}^{n} (n-l+1)\, 2^l \big((1-p_{l,n})\, p_{l,n}\big)^{l/2} \\
&\underset{(4.25)}{=} \sum_{n=1}^{\infty}\sum_{l=1}^{n} (n-l+1)\, 2^l \Big(\big(1-\mathbb{P}(\xi_1^n > C(n)/l)\big)\, \mathbb{P}(\xi_1^n > C(n)/l)\Big)^{l/2} \\
&= \sum_{n=1}^{\infty}\sum_{l=1}^{n} (n-l+1) \left(2\sqrt{F_{\xi_1^n}(C(n)/l)\cdot\Big(1-F_{\xi_1^n}(C(n)/l)\Big)}\right)^{l} \\
&\leq \sum_{n=1}^{\infty}\sum_{l=1}^{n} n \left(2\sqrt{F_{\xi_1^n}(C(n)/l)\cdot\Big(1-F_{\xi_1^n}(C(n)/l)\Big)}\right)^{l} \\
&\underset{(4.1)}{=} \sum_{n=1}^{\infty}\sum_{l=1}^{n} n \left(\frac{2\big(1+C(n)/l\big)^{\alpha-1}-1}{\big(1+C(n)/l\big)^{2(\alpha-1)}}\right)^{l/2}, \quad \alpha \geq 9.
\end{aligned}$$

Denoting by $G_\alpha(l,n)$ for a fix $\alpha \geq 9$ the following expression:

$$G_\alpha(l,n) := \left(2\sqrt{F_{\xi_1^n}(C(n)/l)\cdot\Big(1-F_{\xi_1^n}(C(n)/l)\Big)}\right)^{l} = \left(\frac{2\big(1+C(n)/l\big)^{\alpha-1}-1}{\big(1+C(n)/l\big)^{2(\alpha-1)}}\right)^{l/2}, \qquad (4.29)$$

we rewrite the estimate for the sum $\sum\limits_{n=1}^{\infty}\sum\limits_{l=1}^{n}\sum\limits_{i=0}^{n-l} \mathbb{P}\Big(\sum\limits_{k=i+1}^{i+l} \delta_{\xi_k^n}^{C(n)/l} \geq \dfrac{l}{2}\Big)$ in the form:

$$\sum_{n=1}^{\infty}\sum_{l=1}^{n} \sum_{i=0}^{n-l} \mathbb{P}\Big(\sum_{k=i+1}^{i+l} \delta_{\xi_k^n}^{C(n)/l} \geq \frac{l}{2}\Big) < \sum_{n=1}^{\infty}\sum_{l=1}^{n} n \cdot G_\alpha(l,n). \qquad (4.30)$$

Next, we define a natural number N_0 such that for all $n \geq N_0$ the following conditions are satisfied:

Conditions (B)

(B1) $\sqrt{n} > \ln n$,

(B2) $\ln(\ln n) + 1 - \dfrac{\ln 2}{(\alpha-1)} < \dfrac{1}{2}\ln n$,

(B3) $\frac{1}{8}\Big(\ln C(n)+1-\frac{\ln 2}{\alpha-1}\Big)-2>1,$

(B4) $\frac{1}{2}\Big(\ln C(n)+1-\frac{\ln 2}{\alpha-1}\Big)\leq \ln C(n)-1-\frac{\ln 2}{\alpha-1}.$

We notice here that the definition of N_0 is correct due to the following four relations:

$$\begin{aligned}
&\lim_{n\to\infty}\frac{\sqrt{n}}{\ln n}=\infty,\\
&\lim_{n\to\infty}\frac{\frac{1}{2}\ln n}{\ln(\ln n)+1-\frac{\ln 2}{(\alpha-1)}}=\Big(\frac{\infty}{\infty}\Big)=\lim_{n\to\infty}\frac{\frac{1}{2\,n}}{\frac{1}{n\,\ln n}}=\lim_{n\to\infty}\frac{1}{2}\ln n=\infty,\\
&\lim_{n\to\infty}\frac{1}{8}\Big(\ln C(n)+1-\frac{\ln 2}{\alpha-1}\Big)-2\underset{(4.22)}{=}\infty,\\
&\lim_{n\to\infty}\frac{\frac{1}{2}\Big(\ln C(n)+1-\frac{\ln 2}{\alpha-1}\Big)}{\ln C(n)-1-\frac{\ln 2}{\alpha-1}}=\frac{1}{2}.
\end{aligned}$$

Now we split the series $\sum\limits_{n=1}^{\infty}\sum\limits_{l=1}^{n} n\cdot G_\alpha(l,n)$ into two parts, namely $\sum\limits_{n=1}^{N_0-1}\sum\limits_{l=1}^{n} n\cdot G_\alpha(l,n)$ and $\sum\limits_{n=N_0}^{\infty}\sum\limits_{l=1}^{n} n\cdot G_\alpha(l,n)$. Since the first sum $\sum\limits_{n=1}^{N_0}\sum\limits_{l=1}^{n} n\cdot G_\alpha(l,n)$ is finite, in order to prove the validity of (4.24), i.e.

$$\sum_{n=1}^{\infty}\sum_{l=1}^{n}\sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi_k^n}^{C(n)/l}\geq\frac{l}{2}\Big)<\infty,$$

according to (4.30), it is enough to show that the second sum is finite, i.e.

$$\sum_{n=N_0}^{\infty}\sum_{l=1}^{n} n\cdot G_\alpha(l,n)<\infty. \tag{4.31}$$

As next, we define the following number:

$$l^*(n):=\frac{C(n)}{2}. \tag{4.32}$$

It is obvious that for all $l\leq l^*(n)$ or $l\geq l^*(n)$ we have $\frac{l}{C(n)}\leq\frac{1}{2}$ or $\frac{l}{C(n)}\geq\frac{1}{2}$, respectively.

Taking into account that $G_\alpha(l,n)\geq 0$ for all n and all l, we rearrange the series $\sum\limits_{n=N_0}^{\infty} n\sum\limits_{l=1}^{n}G_\alpha(l,n)$ as follows:

$$\sum_{n=N_0}^{\infty} n\sum_{l=1}^{n}G_\alpha(l,n)\leq\sum_{n=N_0}^{\infty} n\sum_{l=1}^{l^*(n)}G_\alpha(l,n)+\sum_{n=N_0}^{\infty} n\sum_{l=l^*(n)}^{n}G_\alpha(l,n) \tag{4.33}$$

and estimate the first summand in the right-hand side of (4.33):

$$\begin{aligned}
\sum_{n=N_0}^{\infty} n \sum_{l=1}^{l^*(n)} G_\alpha(l,n) &\underset{(4.29)}{=} \sum_{n=N_0}^{\infty} n \sum_{l=1}^{l^*(n)} \left(2\sqrt{F_{\xi_1^n}\big(C(n)/l\big)\cdot\Big(1-F_{\xi_1^n}\big(C(n)/l\big)\Big)} \right)^l \\
\leq \sum_{n=N_0}^{\infty} n \sum_{l=1}^{l^*(n)} \left(2\sqrt{1-F_{\xi_1^n}\big(C(n)/l\big)} \right)^l &\underset{(4.1)}{=} \sum_{n=1}^{\infty} n \sum_{l=1}^{l^*(n)} \left(\frac{2}{\sqrt{2\big(1+C(n)/l\big)^{\alpha-1}}} \right)^l \\
< \sum_{n=N_0}^{\infty} n \sum_{l=1}^{l^*(n)} \big(\sqrt{2}\big)^l \left(\frac{C(n)}{l}\right)^{-\frac{\alpha-1}{2}l}.
\end{aligned}$$

By $\tilde{G}_\alpha(l,n)$ we denote $\big(\sqrt{2}\big)^l \left(\frac{C(n)}{l}\right)^{\frac{\alpha-1}{2}l}$, i.e

$$\tilde{G}_\alpha(l,n) := \big(\sqrt{2}\big)^l \left(\frac{C(n)}{l}\right)^{-\frac{\alpha-1}{2}l}, \tag{4.34}$$

this leads to the following relation:

$$\sum_{n=N_0}^{\infty} n \sum_{l=1}^{l^*(n)} G_\alpha(l,n) < \sum_{n=N_0}^{\infty} n \sum_{l=1}^{l^*(n)} \tilde{G}_\alpha(l,n). \tag{4.35}$$

According to (4.33) and (4.35) we see that in order to show the validity of (4.31) and respectively to complete the proof of Lemma 11 it is sufficient to show the validity of the following two statements:

(i)
$$\sum_{n=N_0}^{\infty} n \sum_{l=1}^{l^*(n)} \tilde{G}_\alpha(l,n) < \infty \tag{4.36}$$

(ii)
$$\sum_{n=N_0}^{\infty} n \sum_{l=l^*(n)}^{n} G_\alpha(l,n) < \infty \tag{4.37}$$

Proof of **(i)** :
First, for all $n \geq N_0$ we consider $\tilde{G}_\alpha(x,n)$ in more details.
Noticing that $\ln x \leq x-1$, we have:

$$\begin{aligned}
\tilde{G}_\alpha(x,n) &= \exp\left\{\frac{x}{2}\ln 2 - \frac{(\alpha-1)}{2}\big(x\ln C(n) - x\ln x\big)\right\} \\
&\leq \exp\left\{\frac{x}{2}\ln 2 - \frac{(\alpha-1)}{2}\big(x\ln C(n) - x(x-1)\big)\right\} =: G^*_\alpha(x,n). \qquad (4.38)
\end{aligned}$$

Due to the following facts:

$$\begin{aligned}
\frac{d}{dx}\Big(\ln \tilde{G}_\alpha(x,n)\Big) &= \frac{1}{2}\Big(\ln 2 - (\alpha-1)\ln C(n) + (\alpha-1)\ln x + (\alpha-1)\Big), && (4.39)\\
\frac{d^2}{dx^2}\Big(\ln \tilde{G}_\alpha(x,n)\Big) &= \frac{1}{2x}(\alpha-1) > 0 \text{ for } \alpha>1 \text{ and } x \geq 1, && (4.40)\\
\frac{d}{dx}\Big(\ln G^*_\alpha(x,n)\Big) &= \frac{1}{2}\Big(\ln 2 - (\alpha-1)\ln C(n) + 2x(\alpha-1) - (\alpha-1)\Big), && (4.41)\\
\frac{d^2}{dx^2}\Big(\ln G^*_\alpha(x,n)\Big) &= (\alpha-1) > 0 \text{ for } \alpha > 1, && (4.42)
\end{aligned}$$

it is clear that the functions $\ln \tilde{G}_\alpha(\cdot,\, n)$ and $\ln G^*_\alpha(\cdot,\, n)$ are convex.

Moreover, both functions $\tilde{G}$ and G^*_α as function of x for some fixed n are functions of the type $\exp\{h(x)\}$ for some function h, implying that $\ln\big(\exp\{h(x)\}\big) = h(x)$. From the relation $\frac{d^2}{dx^2}\exp\{h(x)\} = \exp\{h(x)\}\cdot\Big(\frac{d}{dx}h(x)\Big)^2 + \exp\{h(x)\}\cdot\frac{d^2}{dx^2}h(x)$ it follows that if $\frac{d^2}{dx^2}h(x) > 0$, or equivalently, if $\ln\big(\exp\{h(\cdot)\}\big)$ is convex, then it holds $\frac{d^2}{dx^2}\exp\{h(x)\} > 0$ and therefore the function $\exp\{h(\cdot)\}$ is also convex.
According to this we deduce that $\tilde{G}_\alpha(\cdot,\, n)$ and $G^*_\alpha(\cdot,\, n)$ are both convex functions together with the functions $\ln\tilde{G}_\alpha(\cdot,\, n)$ and $\ln G^*_\alpha(\cdot,\, n)$.
From (4.39) and (4.41) it is easy to see that for

$$\ln l_0(n) = \ln C(n) - 1 - \frac{\ln 2}{\alpha - 1} \tag{4.43}$$

or

$$l_0(n) = \frac{1}{e\cdot 2^{\frac{1}{\alpha-1}}}\cdot C(n) \tag{4.44}$$

it holds:

$$\min_x \tilde{G}_\alpha(x,n) = \tilde{G}_\alpha(l_0(n),n) \tag{4.45}$$

and for

$$l_1(n) = \frac{1}{2}\Big(\ln C(n) + 1 - \frac{\ln 2}{\alpha - 1}\Big) \tag{4.46}$$

we have:

$$\min_x G^*_\alpha(x,n) = \tilde{G}_\alpha(l_1(n),n). \tag{4.47}$$

Thereby by means of **(B4)**, **(B1)** and (4.32), it is obvious that for all $n \geq N_0$

$$l_1(n) \leq \ln l_0(n) < l_0(n) < l^*(n). \tag{4.48}$$

Hence, we can say that for all $x \in [1,\, l_1(n))$ the two functions $\tilde{G}_\alpha(\cdot,\, n)$ and $G^*_\alpha(\cdot,\, n)$ are monotonically decreasing. Further on, for all $x \in [l_1(n),\, l_0(n))$ the function $\tilde{G}_\alpha(\cdot,\, n)$ is monotonically decreasing but the function $G^*_\alpha(\cdot,\, n)$ is monotonically increasing. Moreover, for all $x \geq l_0(n)$ both functions $\tilde{G}_\alpha(\cdot,\, n)$ and $G^*_\alpha(\cdot,\, n)$ are monotonically increasing. Therefore, it is convenient to estimate the series $\sum\limits_{n=N_0}^{\infty} n \sum\limits_{l=1}^{l^*(n)} \tilde{G}_\alpha(l,n)$ by a sum of three series:

$$\begin{aligned}
&\sum_{n=N_0}^{\infty} n \sum_{l=1}^{l^*(n)} \tilde{G}_\alpha(l,n) \\
\leq\; &\sum_{n=N_0}^{\infty} n \sum_{l=1}^{l_1(n)} \tilde{G}_\alpha(l,n) + \sum_{n=N_0}^{\infty} n \sum_{l=l_1(n)}^{l_0(n)} \tilde{G}_\alpha(l,n) + \sum_{n=N_0}^{\infty} n \sum_{l=l_0(n)}^{l^*(n)} \tilde{G}_\alpha(l,n)
\end{aligned} \tag{4.49}$$

and analyze each of this series separately (the last inequality holds due to the fact that $\tilde{G}_\alpha(l,n) \geq 0$ for all n and all l).

We start with the series $\sum\limits_{n=N_0}^{\infty} n \sum\limits_{l=1}^{l_1(n)} \tilde{G}_\alpha(l,n)$.

By the comparison criteria of convergence [47] and by taking (4.38) into account, we conclude that the series $\sum\limits_{n=N_0}^{\infty} n \sum\limits_{l=1}^{l_1(n)} \tilde{G}_\alpha(x,n)$ converges together with the series $\sum\limits_{n=N_0}^{\infty} n \sum\limits_{l=1}^{l_1(n)} G^*_\alpha(x,n)$.

As we have noticed above, for all $n \geq N_0$ and for all $1 \leq l \leq l_1(n)$ the function $\tilde{G}_\alpha(\cdot,\, n)$ is a monotonically decreasing function. Moreover, from Lemma 20, Appendix A, p.116 we know that for $\alpha \geq 9$ it holds:

$$\sum_{n=1}^{\infty} n \int\limits_{1}^{l_1(n)} \exp\left\{\frac{x}{2}\ln 2 - \frac{(\alpha-1)}{2}\big(x \ln C(n) - x(x-1)\big)\right\} dx < \infty.$$

Hence, according (4.38) and to the integral criteria [47], which is valid for the monotonically decreasing functions, we deduce that

$$\sum_{n=N_0}^{\infty} n \sum_{l=1}^{l_1(n)} G^*_\alpha(l,n) < \infty$$

and therefore

$$\sum_{n=N_0}^{\infty} n \int\limits_{1}^{l_1(n)} \tilde{G}_\alpha(x,n)\, dx = \sum_{n=N_0}^{\infty} n \int\limits_{1}^{l_1(n)} \big(\sqrt{2}\big)^x \left(\frac{C(n)}{x}\right)^{-\frac{\alpha-1}{2}x} dx < \infty\,. \tag{4.50}$$

Now we study the convergence of the series $\sum\limits_{n=N_0}^{\infty} n \sum\limits_{l=l_1(n)}^{l_0(n)} \tilde{G}_\alpha(l,n)$.

As we know (see(4.45), (4.48), (4.40)), for all $n \geq N_0$ the function $\tilde{G}_\alpha(\cdot,\, n)$ on the interval $[l_1(n),\, l_0(n)]$ is a monotonically decreasing function. This implies that for all $n \geq N_0$ and all $x \in [l_1(n),\, l_0(n)]$

$$\begin{aligned} &\tilde{G}_\alpha(x,n) \leq \max_{x\in[l_1(n),\, l_0(n)]} \tilde{G}_\alpha(x,n) = \tilde{G}_\alpha(l_1(n),n) \\ \Rightarrow\quad &\sum_{n=N_0}^{\infty} n \sum_{l=l_1(n)}^{l_0(n)} \tilde{G}_\alpha(l,n) \leq \sum_{n=N_0}^{\infty} n \sum_{l=l_1(n)}^{l_0(n)} \tilde{G}_\alpha(l_1(n),n) < \sum_{n=N_0}^{\infty} n^2 \cdot \tilde{G}_\alpha(l_1(n),n). \end{aligned} \tag{4.51}$$

Further on we consider $\tilde{G}_\alpha(l_1(n),n)$, but prior to this we recall that $C(n) = \sqrt{n}\,\ln n$,

$$\tilde{G}_\alpha(l,n) = \big(\sqrt{2}\big)^l \left(\frac{C(n)}{l}\right)^{-\frac{\alpha-1}{2}l}$$

and

$$l_1(n) \underset{(4.46)}{=} \frac{1}{2}\Big(\ln C(n) + 1 - \frac{\ln 2}{\alpha-1}\Big) = \frac{1}{4}\ln n + \frac{1}{2}\ln(\ln n) + \frac{1}{2} - \frac{\ln 2}{2\,(\alpha-1)}.$$

Due to **(B2)**, it is clear that $\frac{1}{2}\Big(\ln(\ln n)+1-\frac{\ln 2}{(\alpha-1)}\Big) < \frac{1}{4}\ln n$ for all $n \geq N_0$. This yields that for all $n \geq N_0$ it holds:

$$l_1(n) < \frac{1}{2}\ln n. \tag{4.52}$$

According to (4.52) and also to the fact that $\alpha \geq 9$, for all $n \geq N_0$ we have:

$$\begin{aligned}\tilde{G}_\alpha(l_1(n),n) &= \big(\sqrt{2}\big)^{l_1(n)}\left(\frac{C(n)}{l_1(n)}\right)^{-\frac{\alpha-1}{2}l_1(n)} = \big(\sqrt{2}\big)^{l_1(n)}\left(\frac{l_1(n)}{C(n)}\right)^{\frac{\alpha-1}{2}l_1(n)} \\ &< 2^{\frac{l_1(n)}{2}}\cdot\left(\frac{1}{2\sqrt{n}}\right)^{\frac{\alpha-1}{2}l_1(n)} < 2^{\frac{l_1(n)}{2}}\cdot\left(\frac{1}{2\sqrt{n}}\right)^{\frac{l_1(n)}{2}} = \left(\frac{1}{n}\right)^{\frac{l_1(n)}{4}}.\end{aligned} \tag{4.53}$$

By substituting (4.53) into (4.51) we obtain the following result:

$$\sum_{n=N_0}^{\infty} n \sum_{l=l_1(n)}^{l_0(n)} \tilde{G}_\alpha(l,n) < \sum_{n=N_0}^{\infty}\left(\frac{1}{n}\right)^{\frac{l_1(n)}{4}-2} \underset{\mathbf{(B3)}}{<} \infty. \tag{4.54}$$

Thus, to accomplish the proof of (i), we have to show that the sum $\sum_{n=N_0}^{\infty} n \sum_{l=l_0(n)}^{l^*(n)} \tilde{G}_\alpha(l,n)$ is also finite.
For all $n \geq N_0$ we estimate $\tilde{G}_\alpha(l,n)$ for all $l_0(n) \leq l \leq l^*(n)$ taking into account that $\tilde{G}_\alpha(\cdot,n)$ on the interval $[l_0(n),\, l^*(n)]$ is a monotonically increasing function (see(4.40), (4.45)):

$$\begin{aligned}\tilde{G}_\alpha(l,n) < \tilde{G}_\alpha(l^*(n),n) &\underset{(4.34),\,(4.32)}{=} \left(\frac{1}{2^{(\alpha-2)/2}}\right)^{\frac{C(n)}{2}} = \left(\frac{1}{2^{(\alpha-2)/2}}\right)^{\frac{\sqrt{n}\,\ln n}{2}} \\ \underset{\mathbf{(B1)}}{<} \left(\frac{1}{2^{(\alpha-2)/2}}\right)^{\frac{\ln^2 n}{2}} &= \exp\left\{-\frac{(\alpha-2)\,\ln 2}{4}\cdot\ln^2 n\right\}.\end{aligned} \tag{4.55}$$

Hence,

$$\begin{aligned}\sum_{n=N_0}^{\infty} n \sum_{l=l_0(n)}^{l^*(n)} \tilde{G}_\alpha(l,n) &< \sum_{n=N_0}^{\infty} n \sum_{l=l_0(n)}^{l^*(n)} \exp\left\{-\frac{(\alpha-2)\,\ln 2}{4}\cdot\ln^2 n\right\} \\ &< \sum_{n=N_0}^{\infty} n^2\cdot\exp\left\{-\frac{(\alpha-2)\,\ln 2}{4}\cdot\ln^2 n\right\} \underset{(A.2)}{<} \infty \quad (\alpha \geq 9).\end{aligned} \tag{4.56}$$

Consequently, from relations (4.49), (4.50), (4.54) and (4.56) we deduce that for $\alpha \geq 9$ the sum $\sum_{n=N_0}^{\infty} n \sum_{l=1}^{l^*(n)} \tilde{G}_\alpha(l,n)$ is finite, and therefore the proposition (i) has been proved.

Proof of **(ii)** :
Now we analyze the convergence of the series $\sum_{n=N_0}^{\infty} n \sum_{l=l^*(n)}^{n} G_\alpha(l,n)$, thereby as before we consider $\alpha \geq 9$.

First, for all $n \geq N_0$ and for all $x \geq l^*(n)$, i.e. for all $x \geq \dfrac{C(n)}{2}$ or equivalently, for all $\dfrac{C(n)}{x} \leq 2$, we consider the function

$$G_\alpha(x,n) \underset{(4.29)}{=} \left(\frac{2\big(1 + C(n)/x\big)^{\alpha-1} - 1}{\big(1 + C(n)/x\big)^{2\big(\alpha-1\big)}} \right)^{x/2}.$$

From Lemma 21, Appendix A, p.119 we know that for $\alpha \geq 1$,

$$k = \frac{1}{4}\Big(1 - \frac{1}{3^{\alpha-1}}\Big)^2 > 0 \tag{4.57}$$

and all $u \in (0,\, 2]$ it holds:

$$\frac{2(1+u)^{\alpha-1} - 1}{(1+u)^{2(\alpha-1)}} \leq (1 - ku^2). \tag{4.58}$$

According to this and to the fact that for all $u \in (0,\, 2]$ both sides of inequality (4.58) are nonnegative, it is clear that for all $n \geq N_0$ and all $x \in [\,l^*(n),\, n\,]$, $l^*(n) = \dfrac{C(n)}{2}$ we have:

$$G_\alpha(x,n) \leq \left(1 - k\Big(\frac{C(n)}{x}\Big)^2\right)^{x/2}.$$

Using the inequality $1 + x \leq e^x$, which is valid for all $x \in \mathbb{R}$ as it is shown in Proposition 7, Appendix A, p.120, and due to the fact that $\left(1 - k\Big(\dfrac{C(n)}{x}\Big)^2\right) \geq 0$ for all $x \in [\,l^*(n),\, n\,]$, for all $n \geq N_0$ we conclude:

$$G_\alpha(x,n) \leq \left(1 - k\Big(\frac{C(n)}{x}\Big)^2\right)^{x/2} \leq \exp\Big\{ -k \cdot \frac{C^2(n)}{2x} \Big\}, \;\; x \in [\,l^*(n),\, n\,].$$

$\Rightarrow$

$$\begin{aligned} &\sum_{n=N_0}^{\infty} n \sum_{l=l^*(n)}^{n} G_\alpha(l,n) \leq \sum_{n=N_0}^{\infty} n \sum_{l=l^*(n)}^{n} \exp\Big\{ -k \cdot \frac{C^2(n)}{2l} \Big\} \\ <\; &\sum_{n=N_0}^{\infty} n \sum_{l=l^*(n)}^{n} \exp\Big\{ -k \cdot \frac{C^2(n)}{2n} \Big\} \underset{(4.22)}{=} \sum_{n=N_0}^{\infty} n \sum_{l=l^*(n)}^{n} \exp\Big\{ -k \cdot \ln^2 n \Big\} \\ <\; &\sum_{n=N_0}^{\infty} n^2 \exp\Big\{ -\frac{k}{2} \cdot \ln^2 n \Big\} \underset{(A.2)}{<} \infty \end{aligned}$$

Therefore we have proved the validity of statement **(ii)** and respectively the proof of Lemma 11 is complete.

□

Corollary 4. *Let $(\xi_i^n)_{1 \leq i \leq n,\, n \in \mathbb{N}}$ satisfy Condition 2. Then $\mathbb{P}$-almost surely for all sufficiently large $n \in \mathbb{N}$, all $j \in \{1, \ldots, n\}$ and all $i \in \{1, \ldots, j\}$ it holds :*

$$|\,\mathrm{med}(\,\xi_i^n, \ldots, \xi_j^n\,)\,| \leq \frac{\sqrt{n}\,\ln n}{j - i + 1}\;. \tag{4.59}$$

Proof:
It is easy to see that if some random variable ξ is symmetrical Pareto distributed, then the random variable $-\xi$ is also symmetrical Pareto distributed with the identical distribution function. Consequently, the inequality (4.21) is valid not just for the random variables $(\xi_i^n)_{1\leq i\leq n,\, n\in\mathbb{N}}$, which satisfy Condition 2, but also for the random variables $\left(-\xi_i^n\right)_{n\in\mathbb{N},\, 1\leq i\leq n}$. Thus, if the sequence of random variables $(\xi_i^n)_{1\leq i\leq n,\, n\in\mathbb{N}}$ fulfills Condition 2, then $\mathbb{P}$-almost surely for all sufficiently large $n \in \mathbb{N}$, all $j \in \{1, \dots, n\}$ and all $i \in \{1, \dots, j\}$ alongside with the inequality

$$\operatorname{med}(\, \xi_i^n, \dots, \xi_j^n \,) \leq \frac{\sqrt{n}\, \ln n}{j - i + 1}$$

it holds :

$$\operatorname{med}(\, -\xi_i^n, \dots, -\xi_j^n \,) \leq \frac{\sqrt{n}\, \ln n}{j - i + 1} \; .$$

The last inequality is equivalent to:

$$-\operatorname{med}(\, \xi_i^n, \dots, \xi_j^n \,) \leq \frac{\sqrt{n}\, \ln n}{j - i + 1} \; ,$$

or :

$$\operatorname{med}(\, \xi_i^n, \dots, \xi_j^n \,) \geq -\, \frac{\sqrt{n}\, \ln n}{j - i + 1} \; . \tag{4.60}$$

Both inequalities, (4.21) and (4.60), prove the validity of inequality (4.59).

□

4.2.2 Estimate of quantiles of arbitrary order

Now we generalize the result of Corollary 4 and Lemma 11 towards an arbitrary quantile of the noise, which satisfy Condition 2. It is reasonable to assume that the estimate for the generalized quantile should be shifted by the theoretical value of this quantile of the Pareto distributed random variable.

Lemma 12. *Let $(\xi_i^n)_{1\leq i\leq n,\, n\in\mathbb{N}}$ fulfill Condition 2. Then $\mathbb{P}$-almost surely for all sufficiently large $n \in \mathbb{N}$, all $l \in \left\{ \left\lceil \frac{n}{\ln n} \right\rceil, \dots, n \right\}$ and all $i \in \{1, \dots, n-l\}$ it holds :*

$$\widetilde{\xi}_q - \frac{\sqrt{n}\, \ln n}{l} \leq q\text{-quantile}\,(\, \xi_{i+1}^n, \dots, \xi_{i+l}^n \,) \leq \widetilde{\xi}_q + \frac{\sqrt{n}\, \ln n}{l} \; , \tag{4.61}$$

where $\widetilde{\xi}_q$ is q-quantile of a random variable ξ, which is identically distributed with ξ_i^n for all $n \in \mathbb{N}$ and all $i \in \{1, \dots, n\}$.

Proof:
First of all, we notice that for the case $q = \frac{1}{2}$ the lemma was proved by Corollary 4.

Therefore we have to prove this lemma for the cases of $q < \frac{1}{2}$ and $q > \frac{1}{2}$, respectively. It is convenient to consider these cases separately.

We start with the case $q > \frac{1}{2}$ and consider the second inequality in (4.61).
This means that we fix an arbitrary $q > \frac{1}{2}$ and show the validity of the following inequality

$$q\text{-quantile}\,(\,\xi^n_{i+1},\dots,\xi^n_{i+l}\,) \le \widetilde{\xi}_q + \frac{C(n)}{l} \qquad \mathbb{P} - \text{a.s.}, \tag{4.62}$$

where $n \in \mathbb{N}$, $l \in \left\{\left\lceil \frac{n}{\ln n}\right\rceil,\dots,n\right\}$, $i \in \{1,\dots,n-l\}$ and $C(n)$ was defined in (4.22), namely

$$C(n) = \frac{\sqrt{n}\,\ln n}{l}.$$

From Corollary 2 we know that in order to do this it is sufficient to show that

$$\sum_{n=1}^{\infty}\sum_{l=\lceil\frac{n}{\ln n}\rceil}^{n}\sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi^n_k}^{\widetilde{\xi}_q+C(n)/l} \ge (1-q)\cdot l\Big) < \infty, \tag{4.63}$$

where for all $n \in \mathbb{N}$ and $k \in \{1,\dots,n\}$ the random variable

$$\delta_{\xi^n_k}^{\widetilde{\xi}_q+C(n)/l} = \begin{cases} 1 & , \quad \xi^n_k > \widetilde{\xi}_q + C(n)/l \\ 0 & , \quad \text{else} \end{cases}$$

is Bernoulli distributed with the parameter $\mathbb{P}\Big(\xi^n_k > \widetilde{\xi}_q + \frac{C(n)}{l}\Big)$.
Since for all $n \in \mathbb{N}$ and $k \in \{1,\dots,n\}$ the random variables ξ^n_k are identically distributed, the probability $\mathbb{P}\Big(\xi^n_k > \widetilde{\xi}_q + \frac{C(n)}{l}\Big)$ is the same for all $k \in \{1,\dots n\}$. For the sake of simplicity further on we denote this probability by $\widetilde{p}^{\,+}_{n,l}$, i.e.

$$\widetilde{p}^{\,+}_{n,l} := \mathbb{P}\Big(\xi^n_1 > \widetilde{\xi}_q + \frac{C(n)}{l}\Big) = \ldots = \mathbb{P}\Big(\xi^n_n > \widetilde{\xi}_q + \frac{C(n)}{l}\Big). \tag{4.64}$$

According to this, it is clear that for all $n \in \mathbb{N}$, all $l \in \left\{\left\lceil \frac{n}{\ln n}\right\rceil,\dots,n\right\}$ and $i \in \{0,\dots,n-l\}$ it holds:

$$\mathbb{E}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi^n_k}^{\widetilde{\xi}_q+C(n)/l}\Big) = l\cdot\widetilde{p}^{\,+}_{n,l}. \tag{4.65}$$

With (4.65) in mind we consider now the sum in the left-hand side of (4.63):

$$\begin{aligned}&\sum_{n=1}^{\infty}\sum_{l=\lceil\frac{n}{\ln n}\rceil}^{n}\sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi^n_k}^{\widetilde{\xi}_q+C(n)/l} \ge (1-q)\cdot l\Big)\\ =\ &\sum_{n=1}^{\infty}\sum_{l=\lceil\frac{n}{\ln n}\rceil}^{n}\sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi^n_k}^{\widetilde{\xi}_q+C(n)/l} - \mathbb{E}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi^n_k}^{\widetilde{\xi}_q+C(n)/l}\Big) \ge l\cdot\big((1-q)-\widetilde{p}^{\,+}_{n,l}\big)\Big).\end{aligned} \tag{4.66}$$

From Lemma 29, Appendix A, p.132 we know that there exists $N_0 \in \mathbb{N}$ such that for all $n > N_0$ and all $l > \frac{n}{\ln n}$ it holds:

$$((1-q)-\widetilde{p}^{\,+}_{n,l}) \ge \frac{\alpha-1}{2\,(\,1+\widetilde{\xi}_q\,)^{\alpha}}\cdot\frac{C(n)}{l} > 0, \quad \alpha \ge 9. \tag{4.67}$$

This means that in order to estimate the sum in the left-hand side of (4.66) we can apply Hoeffding's inequality (A.23), p.121.
Therefore we obtain:

$$\begin{aligned}
&\sum_{n=N_0}^{\infty}\ \sum_{l=\lceil \frac{n}{\ln n}\rceil}^{n}\ \sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi_k^n}^{\widetilde{\xi}_q+C(n)/l}\geq(1-q)\cdot l\Big)\\
\underset{(A.23)}{\leq}\ &\sum_{n=N_0}^{\infty}\ \sum_{l=\lceil \frac{n}{\ln n}\rceil}^{n}\ \sum_{i=0}^{n-l}\exp\Big\{-\frac{2\,l^2\left((1-q)-\widetilde{p}_{n,l}^{+}\right)^2}{l}\Big\}\\
\leq\ &\sum_{n=N_0}^{\infty}\ \sum_{l=\lceil \frac{n}{\ln n}\rceil}^{n} n\cdot\exp\Big\{-2\,l\,((1-q)-\widetilde{p}_{n,l}^{+})^2\Big\}\\
\underset{(4.67)}{\leq}\ &\sum_{n=N_0}^{\infty}\ \sum_{l=\lceil \frac{n}{\ln n}\rceil}^{n} n\cdot\exp\Big\{-2\,l\Big(\frac{\alpha-1}{2\,(\,1+\widetilde{\xi}_q\,)^{\alpha}}\cdot\frac{C(n)}{l}\Big)^2\Big\}\\
=\ &\sum_{n=N_0}^{\infty}\ \sum_{l=\lceil \frac{n}{\ln n}\rceil}^{n} n\cdot\exp\Big\{-\frac{(\alpha-1)^2}{(\,1+\widetilde{\xi}_q\,)^{2\,\alpha}}\cdot\frac{(C(n))^2}{2l}\Big\}\\
\leq\ &\sum_{n=N_0}^{\infty} n^2\cdot\exp\Big\{-\frac{(\alpha-1)^2}{(\,1+\widetilde{\xi}_q\,)^{2\,\alpha}}\cdot\frac{(C(n))^2}{2n}\Big\}\\
=\ &\sum_{n=N_0}^{\infty} n^2\cdot\exp\Big\{-\frac{(\alpha-1)^2}{2\,(\,1+\widetilde{\xi}_q\,)^{2\,\alpha}}\cdot\ln^2 n\Big\}\underset{(A.2)}{<}\infty.
\end{aligned}$$

Hence, taking into account that N_0 is a finite number, we conclude that the sum $\sum_{n=1}^{\infty}\sum_{l=\lceil \frac{n}{\ln n}\rceil}^{n}\sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi_k^n}^{\widetilde{\xi}_q+C(n)/l}\geq(1-q)\cdot l\Big)$ is finite as well.

According to (4.63) we have proved that for $q>\frac{1}{2}$, for all sufficiently large $n\in\mathbb{N}$, all $l\in\Big\{\Big\lceil\frac{n}{\ln n}\Big\rceil,\dots,n\Big\}$ and all $i\in\{1,\dots,n-l\}$ it holds

$$q\text{-quantile}\,(\,\xi_{i+1}^n,\dots,\xi_{i+l}^n\,)\leq\widetilde{\xi}_q+\frac{C(n)}{l}\qquad\mathbb{P}-\text{a.s.}$$

Now we show the validity of the same inequality for $q<\frac{1}{2}$.
Analogue to the case of $q>\frac{1}{2}$, it is sufficient to show the convergence of the series

$$\sum_{n=1}^{\infty}\ \sum_{l=\lceil \frac{n}{\ln n}\rceil}^{n}\ \sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi_k^n}^{\widetilde{\xi}_q+C(n)/l}-\mathbb{E}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi_k^n}^{\widetilde{\xi}_q+C(n)/l}\Big)\geq l\cdot\big((1-q)-\widetilde{p}_{n,l}^{+}\big)\Big).\qquad(4.68)$$

Lemma 29, Appendix A, p.132, yields that for any $q<\frac{1}{2}$ there exists a natural finite number N' such that for all $n>N'$ and all $l>\frac{n}{\ln n}$ it holds:

$$(1-q)-\widetilde{p}_{n,l}^{+}\geq\frac{\alpha-1}{2\,(\,1-\widetilde{\xi}_q\,)^{\alpha}}\cdot\frac{C(n)}{l}>0\quad(\alpha\geq 9,\ \widetilde{\xi}_q<0).$$

Consequently, the sum

$$\sum_{n=N'}^{\infty}\sum_{l=\lceil\frac{n}{\ln n}\rceil}^{n}\sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi_k^n}^{\widetilde{\xi}_q+C(n)/l}-\mathbb{E}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi_k^n}^{\widetilde{\xi}_q+C(n)/l}\Big)\geq l\cdot\big((1-q)-\widetilde{p}_{n,l}^{+}\big)\Big)$$

can be estimated with help of Hoeffding's inequality similarly to the case of $q>\dfrac{1}{2}$:

$$\begin{aligned}&\sum_{n=N'}^{\infty}\sum_{l=\lceil\frac{n}{\ln n}\rceil}^{n}\sum_{i=0}^{n-l}\mathbb{P}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi_k^n}^{\widetilde{\xi}_q+C(n)/l}-\mathbb{E}\Big(\sum_{k=i+1}^{i+l}\delta_{\xi_k^n}^{\widetilde{\xi}_q+C(n)/l}\Big)\geq l\cdot\big((1-q)-\widetilde{p}_{n,l}^{+}\big)\Big)\\ \leq\;&\sum_{n=N'}^{\infty}n^2\cdot\exp\Big\{-\frac{(\alpha-1)^2}{2\,(1-\widetilde{\xi}_q)^{2\,\alpha}}\cdot\ln^2 n\Big\}\underset{(A.2)}{<}\infty.\end{aligned}$$

N' is a finite number, therefore we have shown the convergence of the series $\sum\limits_{n=1}^{\infty}\sum\limits_{l=\lceil\frac{n}{\ln n}\rceil}^{n}\sum\limits_{i=0}^{n-l}\mathbb{P}\Big(\sum\limits_{k=i+1}^{i+l}\delta_{\xi_k^n}^{\widetilde{\xi}_q+C(n)/l}\geq(1-q)\cdot l\Big)$ in the case of $q<\dfrac{1}{2}$.

This proves the following inequality for an arbitrary $q\in(0,1)$:

$$q\text{-quantile}\,(\,\xi_{i+1}^n,\ldots,\xi_{i+l}^n\,)\leq\widetilde{\xi}_q+\frac{C(n)}{l}\quad\mathbb{P}-\text{a.s.},\tag{4.69}$$

where n is sufficiently large, $l\in\Big\{\Big\lceil\dfrac{n}{\ln n}\Big\rceil,\ldots,n\Big\}$ and $i\in\{1,\ldots,n-l\}$.

In order to complete the proof of this lemma we have to show that $\mathbb{P}$-almost surely for all sufficiently large $n\in\mathbb{N}$, all $l\in\Big\{\Big\lceil\dfrac{n}{\ln n}\Big\rceil,\ldots,n\Big\}$ and all $i\in\{1,\ldots,n-l\}$ it holds :

$$q\text{-quantile}\,(\,\xi_{i+1}^n,\ldots,\xi_{i+l}^n\,)\geq\widetilde{\xi}_q-\frac{\sqrt{n}\,\ln n}{l}.\tag{4.70}$$

This inequality follows directly from inequality (4.69), because the latter is valid not just for the random variables $(\xi_i^n)_{1\leq i\leq n,\,n\in\mathbb{N}}$, which satisfy Condition 2, but also for the random variables $\big(-\xi_i^n\big)_{n\in\mathbb{N},\,1\leq i\leq n}$, which also fulfill Condition 2. More precisely, for any $q\in(0,1)$, all sufficiently large $n\in\mathbb{N}$, all $l\in\Big\{\Big\lceil\dfrac{n}{\ln n}\Big\rceil,\ldots,n\Big\}$ and all $i\in\{1,\ldots,n-l\}$ it holds:

$$q\text{-quantile}\,(\,-\xi_{i+1}^n,\ldots,-\xi_{i+l}^n\,)\leq\widetilde{\xi}_q+\frac{C(n)}{l}\quad\mathbb{P}-\text{a.s.}$$

According to (4.10) we know:

$$q\text{-quantile}\,(\,-\xi_{i+1}^n,\ldots,-\xi_{i+l}^n\,)=-\,(1\text{-}q)\text{-quantile}\,(\,\xi_{i+1}^n,\ldots,\xi_{i+l}^n\,).$$

From this it follows:

$$(1\text{-}q)\text{-quantile}\,(\,\xi_{i+1}^n,\ldots,\xi_{i+l}^n\,)\geq-(\widetilde{\xi}_q)-\frac{C(n)}{l}\quad\mathbb{P}-\text{a.s.}$$

Due to the fact that $-(\widetilde{\xi}_q)=\widetilde{\xi}_{(1-q)}$ (see (4.9)), for any $q\in(0,1)$, all sufficiently large $n\in\mathbb{N}$, all $l\in\Big\{\Big\lceil\dfrac{n}{\ln n}\Big\rceil,\ldots,n\Big\}$ we obtain:

$$(1\text{-}q)\text{-quantile}\,(\,\xi_{i+1}^n,\ldots,\xi_{i+l}^n\,)\geq\widetilde{\xi}_{(1-q)}-\frac{C(n)}{l}\quad\mathbb{P}-\text{a.s.}$$

Since the last inequality is valid for an arbitrary $q \in (0,1)$, we conclude that the next inequality is also $\mathbb{P}$-almost surely valid for any $q \in (0,1)$, all sufficiently large $n \in \mathbb{N}$, all $l \in \Big\{ \Big\lceil \frac{n}{\ln n} \Big\rceil, \dots, n \Big\}$:

$$q\text{-quantile}\,(\,\xi^n_{i+1}, \dots, \xi^n_{i+l}\,) \geq \widetilde{\xi}_q - \frac{C(n)}{l}.$$

Hence, we have proved inequality (4.70) and therefore completed the proof of Lemma 12.

□

The statement of Lemma 12 can be rewritten in a slightly different form. This is presented in the following corollary.

Corollary 5. *Let $(\xi^n_i)_{1 \leq i \leq n,\, n \in \mathbb{N}}$ fulfill Condition 2 and $q \in (0,1)$. Then $\mathbb{P}$-almost surely for all sufficiently large $n \in \mathbb{N}$, all $l \in \Big\{ \Big\lceil \frac{n}{\ln n} \Big\rceil, \dots, n \Big\}$ and all $i \in \{1, \dots, n-l\}$ it holds:*

$$|\; q\text{-quantile}\,(\,\xi^n_{i+1}, \dots, \xi^n_{i+l}\,)\,| \leq \max\Big\{ -\big(\tilde{\xi}_q - \frac{\sqrt{n}\,\ln n}{l}\big),\; \tilde{\xi}_q + \frac{\sqrt{n}\,\ln n}{l} \Big\}, \qquad (4.71)$$

where $\widetilde{\xi}_q$ is q-quantile of a random variable ξ, which is identically distributed with ξ^n_i for all $n \in \mathbb{N}$ and all $i \in \{1, \dots, n\}$.

□

4.2.3 Estimate of mixed median

In the previous subsections we have estimated various quantiles of a set of data of an arbitrary size of identical symmetrical Pareto distributed random variables. In particular we gave an estimate for the median of a set of the symmetrical Pareto distributed variables. Now we consider a set of data of the symmetrical Pareto distributed random variables, in which some of the elements are shifted by a certain positive constant. In other words, our aim is to derive an estimate for the following median of the noise $(\xi^n_i)_{1 \leq i \leq n,\, n \in \mathbb{N}}$, which fulfills Condition 2 :

$$\text{med}(\,\xi^n_{i+1}, \dots, \xi^n_{i+k},\, \xi^n_{i+k+1} + \Delta, \dots, \xi^n_{i+k+l} + \Delta\,), \qquad (4.72)$$

where Δ is an arbitrary positive constant, $n \in \mathbb{N}$, $l \in \{1, \dots, n\}$, $k \in \{1, \dots, n-l\}$ and $i \in \{0, \dots, n-l-k\}$. This estimate depends on the size of a particular realization of the random variables, i.e. it depends on the parameters k and l. We are particularly interested in the case, when at least one of these two numbers is large. By large we mean that k or l should be not smaller than $\frac{n}{\ln n}$, where n is an arbitrary large integer.
The following two lemmas represent two different estimates of a median of the type (4.72), one is for the case when either k or l is large, and the other estimate is for the case when both these numbers are large. Clearly, in case when one of the numbers k or l is large and the other one is small, the median of the $k+l$ elements is expected to depend more strongly on the characteristics of the elements of the longer subset.

For the sake of simplicity we introduce here some new notations.

Let $(\xi_i^n)_{1\le i\le n,\, n\in\mathbb{N}}$ fulfill Condition 2. Further on, let ξ be a random variable having the same distribution as ξ_i^n for all $n \in \mathbb{N}$ and all $i \in \{1,\dots,n\}$ and $\widetilde{\xi}_q$ is q-quantile of ξ, $q \in (0,1)$. Then for any $\nu < \dfrac{1}{2}$ by z_ν^- and z_ν^+ we denote $\Big(\dfrac{1}{2}-\nu\Big)$-quantile of ξ and $\Big(\dfrac{1}{2}+\nu\Big)$-quantile of ξ, respectively, i.e.

$$z_\nu^- := \widetilde{\xi}_{\frac{1}{2}-\nu} \tag{4.73}$$

$$z_\nu^+ := \widetilde{\xi}_{\frac{1}{2}+\nu} \tag{4.74}$$

And by $C(n)$ we still denote the number $\sqrt{n}\,\ln n$, i.e.

$$C(n) = \sqrt{n}\,\ln n.$$

Lemma 13. *Let $(\xi_i^n)_{1\le i\le n,\, n\in\mathbb{N}}$ fulfill Condition 2 and Δ be an arbitrary real constant. Then $\mathbb{P}$-almost surely for all sufficiently large $n \in \mathbb{N}$ the following two properties are valid:*

(i) for all $l \in \Big\{\Big\lceil \dfrac{n}{\ln n}\Big\rceil,\dots,n\Big\}$, all $k \in \{1,\dots,n-l\}$ such that $k<l$ and all $i \in \{0,\dots,n-l-k\}$ it holds:

$$\begin{aligned} &|\operatorname{med}(\,\xi_{i+1}^n,\dots,\xi_{i+k}^n,\,\xi_{i+k+1}^n+\Delta,\dots,\xi_{i+k+l}^n+\Delta\,)\,| \\ \le\ &\max\Big\{-\Big(z^-_{\frac{k}{k+l}}+\Delta-\frac{C(n)}{l}\Big),\ z^+_{\frac{k}{k+l}}+\Delta+\frac{C(n)}{l}\Big\} \end{aligned} \tag{4.75}$$

(ii) for all $k \in \Big\{\Big\lceil \dfrac{n}{\ln n}\Big\rceil,\dots,n\Big\}$, $l \in \{0,\dots,n-k\}$ such that $l<k$ and all $i \in \{0,\dots,n-l-k\}$ it holds:

$$\begin{aligned} &|\operatorname{med}(\,\xi_{i+1}^n,\dots,\xi_{i+k}^n,\,\xi_{i+k+1}^n+\Delta,\dots,\xi_{i+k+l}^n+\Delta\,)\,| \\ \le\ &\max\Big\{-\Big(z^-_{\frac{l}{k+l}}-\frac{C(n)}{k}\Big),\ z^+_{\frac{l}{k+l}}+\frac{C(n)}{k}\Big\} \end{aligned} \tag{4.76}$$

Proof:
We prove here only statement (i), as the proof of statement (ii) can be constructed by analogy.
We have to show that the following inequality, which is equivalent to (4.75), is valid $\mathbb{P}$-almost surely for all sufficiently large n, all $l \ge \dfrac{n}{\ln n}$, all $k<l$ and all $i \in \{0,\dots,n-l-k\}$:

$$z^-_{\frac{k}{k+l}}+\Delta-\frac{C(n)}{l} \le \operatorname{med}\big(\,\xi_{i+1}^n,\dots,\xi_{i+k}^n,\,\xi_{i+k+1}^n+\Delta,\dots,\xi_{i+k+l}^n+\Delta\,\big) \le z^+_{\frac{k}{k+l}}+\Delta+\frac{C(n)}{l}.$$

Clearly, for all considered n, l, k and i it holds:

$$\begin{aligned} &\operatorname{med}\big(\,\xi_{i+1}^n,\dots,\xi_{i+k}^n,\,\xi_{i+k+1}^n+\Delta,\dots,\xi_{i+k+l}^n+\Delta\,\big) \\ \ge\ &\operatorname{med}\big(\underbrace{-\infty,\dots,-\infty}_{k},\,\xi_{i+k+1}^n+\Delta,\dots,\xi_{i+k+l}^n+\Delta\,\big) \\ =\ &\Big(\frac{1}{2}-\frac{k}{k+l}\Big)\text{-quantile}\,(\xi_{i+k+1}^n+\Delta,\dots,\xi_{i+k+l}^n+\Delta\,) \\ =\ &\Big(\frac{1}{2}-\frac{k}{k+l}\Big)\text{-quantile}\,(\xi_{i+k+1}^n,\dots,\xi_{i+k+l}^n\,)+\Delta\,. \end{aligned} \tag{4.77}$$

For the estimate of $\Big(\dfrac{1}{2}\text{-}\dfrac{k}{k+l}\Big)$-quantile $(\xi^n_{i+k+1},\dots,\xi^n_{i+k+l})$ we can use the left part of inequality (4.61) in Lemma 12, which yields that $\mathbb{P}$-almost surely for all sufficiently large $n\in\mathbb{N}$ all $l\geq \dfrac{n}{\ln n}$, all $k<l$ and all $i\in\{0,\dots,n-l-k\}$ it holds:

$$\Big(\frac{1}{2}\text{-}\frac{k}{k+l}\Big)\text{-quantile}\,(\xi^n_{i+k+1},\dots,\xi^n_{i+k+l}) \geq \widetilde{\xi}_{\frac{1}{2}-\frac{k}{k+l}} - \frac{C(n)}{l} \underset{(4.73)}{=} z^-_{\frac{k}{k+l}} - \frac{C(n)}{l}. \quad (4.78)$$

After substituting (4.78) into (4.77), $\mathbb{P}$-almost surely for all sufficiently large $n\in\mathbb{N}$ and $l\geq \dfrac{n}{\ln n}$, all $k<l$ and all $i\in\{0,\dots,n-l-k\}$ we obtain:

$$\text{med}\,\big(\,\xi^n_{i+1},\dots,\xi^n_{i+k},\,\xi^n_{i+k+1}+\Delta,\dots,\xi^n_{i+k+l}+\Delta\,\big) \geq z^-_{\frac{k}{k+l}} + \Delta - \frac{C(n)}{l}. \quad (4.79)$$

On the other hand, $\mathbb{P}$-almost surely for all sufficiently large $n\in\mathbb{N}$ and all $l\geq\dfrac{n}{\ln n}$, all $k<l$ and all $i\in\{0,\dots,n-l-k\}$ it holds:

$$\begin{aligned}
& \text{med}\,\big(\,\xi^n_{i+1},\dots,\xi^n_{i+k},\,\xi^n_{i+k+1}+\Delta,\dots,\xi^n_{i+k+l}+\Delta\,\big)\\
\leq\ & \text{med}(\,\xi^n_{i+k+1}+\Delta,\dots,\xi^n_{i+k+l}+\Delta,\,\underbrace{+\infty,\dots,+\infty}_{k})\\
=\ & \Big(\frac{1}{2}+\frac{k}{k+l}\Big)\text{-quantile}\,(\xi^n_{i+k+1}+\Delta,\dots,\xi^n_{i+k+l}+\Delta\,)\\
=\ & \Big(\frac{1}{2}+\frac{k}{k+l}\Big)\text{-quantile}\,(\xi^n_{i+k+1},\dots,\xi^n_{i+k+l}\,)+\Delta\\
\underset{(4.61)}{\leq}\ & \widetilde{\xi}_{\frac{1}{2}+\frac{k}{k+l}} + \frac{C(n)}{l} + \Delta \underset{(4.74)}{=} z^+_{\frac{k}{k+l}} + \Delta + \frac{C(n)}{l}\,.
\end{aligned} \quad (4.80)$$

Inequalities (4.79) and (4.80) yield the $\mathbb{P}$-almost surely validity of (4.75), which completes the proof of statement (i). Statement (ii) is obtained in the analogues way by swapping k with l. Therefore the proof of Lemma 13 is complete.

□

Moreover, Lemma 13 can be generalized for medians of the following type:

$$|\ \text{med}(\,\xi^n_{i+1}+\Delta_1,\dots,\xi^n_{i+k}+\Delta_1,\xi^n_{i+k+1},\dots,\xi^n_{i+k+l},\,\xi^n_{i+k+l+1}+\Delta_2,\dots,\xi^n_{i+k+l+s}+\Delta_2\,)\,|,$$

where $\Delta_1,\Delta_2\in\mathbb{R}$, $n\in\mathbb{N}$, $l\in\Big\{\Big\lceil\dfrac{n}{\ln n}\Big\rceil,\dots,n\Big\}$, $k\in\{1,\dots,n-l\}$, $s\in\{1,\dots,n-l-k\}$, $s+k<l$ and $i\in\{0,\dots,n-l-k-s\}$.

This will be presented in the following lemma.

Lemma 14. *Let $(\xi^n_i)_{1\leq i\leq n,\,n\in\mathbb{N}}$ fulfill Condition 2 and Δ_1 and Δ_2 be arbitrary real constants. Then $\mathbb{P}$-almost surely for all sufficiently large $n\in\mathbb{N}$ for all $l\in\Big\{\Big\lceil\dfrac{n}{\ln n}\Big\rceil,\dots,n\Big\}$, all $k\in\{1,\dots,n-l\}$ and $s\in\{1,\dots,n-l-k\}$ such that $s+k<l$ and all $i\in\{0,\dots,n-l-k-s\}$ it holds:*

$$\begin{aligned}
&|\ \text{med}(\,\xi^n_{i+1}+\Delta_1,\dots,\xi^n_{i+k}+\Delta_1,\xi^n_{i+k+1},\dots,\xi^n_{i+k+l},\,\xi^n_{i+k+l+1}+\Delta_2,\dots,\xi^n_{i+k+l+s}+\Delta_2\,)\,|\\
&\leq \max\Big\{-\Big(z^-_{\frac{k+s}{k+s+l}}-\frac{C(n)}{l}\Big),\ z^+_{\frac{k+s}{k+s+l}}+\frac{C(n)}{l}\Big\},
\end{aligned} \quad (4.81)$$

where $z^-_{\frac{k+s}{k+l+s}}$, $z^+_{\frac{k+s}{k+l+s}}$ are defined according to (4.73) and (4.74).

Proof:
We prove this lemma in the analogous way with Lemma 14.
We fix arbitrary $\Delta_1 \in \mathbb{R}$ and $\Delta_2 \in \mathbb{R}$. Our goal is to show that the following inequality, which is equivalent to (4.81), is valid $\mathbb{P}$-almost surely for all sufficiently large n, all $l > \dfrac{n}{\ln n}$, all k and s with $k+s<l$ and all $i \in \{0,\dots,n-l-k-s\}$:

$$\begin{aligned}
& z^-_{\frac{k+s}{k+s+l}} - \frac{C(n)}{l} \\
\leq \; & |\operatorname{med}(\xi^n_{i+1}+\Delta_1,\dots,\xi^n_{i+k}+\Delta_1,\xi^n_{i+k+1},\dots,\xi^n_{i+k+l},\, \xi^n_{i+k+l+1}+\Delta_2,\dots,\xi^n_{i+k+l+s}+\Delta_2)| \\
\leq \; & z^+_{\frac{k+s}{k+s+l}} + \frac{C(n)}{l},
\end{aligned}$$

where $z^-_{\frac{k+s}{k+l+s}} = \widetilde{\xi}_{\frac{1}{2}-\frac{k+s}{k+l+s}}$, $z^+_{\frac{k+s}{k+l+s}} = \widetilde{\xi}_{\frac{1}{2}+\frac{k+s}{k+l+s}}$.

It is not difficult to see that for all considered n, l, k, s and i it holds:

$$\begin{aligned}
& \operatorname{med}(\xi^n_{i+1}+\Delta_1,\dots,\xi^n_{i+k}+\Delta_1,\xi^n_{i+k+1},\dots,\xi^n_{i+k+l},\, \xi^n_{i+k+l+1}+\Delta_2,\dots,\xi^n_{i+k+l+s}+\Delta_2) \\
\geq \; & \operatorname{med}\big(\underbrace{-\infty,\dots,-\infty}_{k},\, \xi^n_{i+k+1},\dots,\xi^n_{i+k+l}\,\underbrace{-\infty,\dots,-\infty}_{s}\big) \\
= \; & \Big(\frac{1}{2} - \frac{k+s}{k+l+s}\Big)\text{-quantile}\,(\xi^n_{i+k+1},\dots,\xi^n_{i+k+l}) \qquad (4.82)
\end{aligned}$$

To estimate $\Big(\dfrac{1}{2} - \dfrac{k+s}{k+l+s}\Big)$-quantile $(\xi^n_{i+k+1},\dots,\xi^n_{i+k+l})$ we apply the left part of inequality (4.61) in Lemma 12, and obtain that $\mathbb{P}$-almost surely for all sufficiently large n, all $l > \dfrac{n}{\ln n}$, all k and s with $k+s<l$ and all $i \in \{0,\dots,n-l-k-s\}$ it holds:

$$\begin{aligned}
& \Big(\frac{1}{2} - \frac{k+s}{k+l+s}\Big)\text{-quantile}\,(\xi^n_{i+k+1},\dots,\xi^n_{i+k+l}) \\
\overset{\mathbb{P}-\text{a.s.}}{\geq} \; & \widetilde{\xi}_{\frac{1}{2}-\frac{k+s}{k+l+s}} - \frac{C(n)}{l} = z^-_{\frac{k+s}{k+l+s}} - \frac{C(n)}{l} \qquad (4.83)
\end{aligned}$$

Substituting (4.83) into (4.82) yields the validity of the following inequality $\mathbb{P}$-almost surely for all sufficiently large n, all $l > \dfrac{n}{\ln n}$, all k and s with $k+s<l$ and all $i \in \{0,\dots,n-l-k-s\}$:

$$\begin{aligned}
& \operatorname{med}(\xi^n_{i+1}+\Delta_1,\dots,\xi^n_{i+k}+\Delta_1,\xi^n_{i+k+1},\dots,\xi^n_{i+k+l},\, \xi^n_{i+k+l+1}+\Delta_2,\dots,\xi^n_{i+k+l+s}+\Delta_2) \\
\geq \; & z^-_{\frac{k+s}{k+l+s}} - \frac{C(n)}{l}. \qquad (4.84)
\end{aligned}$$

On the other hand, $\mathbb{P}$-almost surely for all sufficiently large n, all $l > \dfrac{n}{\ln n}$, all k and s with $k+s<l$ and all $i \in \{0,\dots,n-l-k-s\}$ it holds:

$$\begin{aligned}
& \operatorname{med}(\xi^n_{i+1}+\Delta_1,\dots,\xi^n_{i+k}+\Delta_1,\xi^n_{i+k+1},\dots,\xi^n_{i+k+l},\, \xi^n_{i+k+l+1}+\Delta_2,\dots,\xi^n_{i+k+l+s}+\Delta_2) \\
\leq \; & \operatorname{med}\big(\underbrace{+\infty,\dots,+\infty}_{k},\, \xi^n_{i+k+1},\dots,\xi^n_{i+k+l},\, \underbrace{+\infty,\dots,+\infty}_{s}\big)
\end{aligned}$$

$$= \quad \Big(\frac{1}{2} + \frac{k}{k+l}\Big)\text{-quantile}\,(\xi^n_{i+k+1},\ldots,\xi^n_{i+k+l}\,)$$
$$\underset{(4.61)}{\overset{\mathbb{P}-\text{a.s.}}{\leq}} \quad \widetilde{\xi}_{\frac{1}{2}+\frac{k+s}{k+l+s}} + \frac{C(n)}{l} = z^+_{\frac{k+s}{k+l+s}} + \frac{C(n)}{l}\,. \qquad (4.85)$$

From (4.84) and (4.85) we deduce the $\mathbb{P}$-almost surely validity of (4.81). This completes the proof of Lemma 14.

□

Next step is to estimate a mixed median from (4.72) in case when both numbers k and l are large. Clearly, because the number of the non-shifted as well as the number of the shifted elements tends to infinity as $n \to \infty$, the respective medians of these sets tend to the corresponding theoretical medians. Taking into account that the theoretical median of the non-shifted Pareto random variable is zero and the theoretical median of the shifted Pareto random variable is given by Δ, the mixed median cannot be large than Δ.

Before moving on to the main result regarding the mixed medians, we derive one auxiliary result.

Lemma 15. *Let ξ be a symmetrical Pareto distributed variable with parameter $\alpha > 9$, i.e. the distribution function F_ξ is defined by (4.1). Moreover, let $P^c_\xi := \mathbb{P}(\xi > c),\ c \in \mathbb{R}$. Then for any finite $\Delta > 0$ and any finite $\beta > 0$ there exists a number $\chi = \chi(\Delta,\beta) > 0$ and $\chi < \Delta$ such that the following two equivalent inequalities are valid:*

$$\frac{1}{(1+\Delta-\chi)^{\alpha-1}} - \frac{\beta}{(1+\chi)^{\alpha-1}} > (1-\beta) \qquad (4.86)$$

and

$$\beta\cdot(\,1-2\,P^{\chi}_{\xi}\,) + (\,1-2\,P^{\chi-\Delta}_{\xi}\,) > 0. \qquad (4.87)$$

Proof:
In Lemma 24, Appendix A, p.122 we show that for any symmetrical continuous random variable ξ, for any $\Delta > 0$ and any $\beta > 0$ there exists a finite number $\chi = \chi(\Delta,\beta) > 0$ such that $\chi < \Delta$, which satisfies inequality (4.87). Moreover, in the proof of Lemma 24 we also show that inequality (4.87) and inequality:

$$\beta\,F_\xi(\chi) + F_\xi(\chi-\Delta) > \frac{(\beta+1)}{2}$$

are equivalent.
Taking into account that $\chi > 0$ and $\chi - \Delta < 0$, we substitute formula (4.1), which defines F_ξ, into the last inequality:

$$\beta\Big(1 - \frac{1}{2\,(1+\chi)^{\alpha-1}}\Big) + \frac{1}{2\,(1-(\chi-\Delta))^{\alpha-1}} > \frac{(\beta+1)}{2}.$$

This can be equivalently rewritten as follows:

$$\frac{1}{2\,(1+\Delta-\chi)^{\alpha-1}} - \frac{\beta}{2\,(1+\chi)^{\alpha-1}} > \frac{(1-\beta)}{2}$$

Therefore, we deduce that for an arbitrary Pareto distributed variable ξ with parameter $\alpha > 9$, there exist $\chi < \Delta$ such that both equivalent inequalities (4.87) and (4.86) are valid. Consequently, we have proved Lemma 15.

□

Lemma 16. *Let $(\xi_i^n)_{1\le i\le n,\, n\in\mathbb{N}}$ fulfill Condition 2. Then for some finite $\Delta > 0$, for all finite $\beta > 0$, any $0 < \chi < \Delta$, which fulfills inequality (4.86), the following statement is valid $\mathbb{P}$-almost surely for all sufficiently large $n \in \mathbb{N}$ such that $n > 10$, all $l \in \left\{\left\lceil \frac{n}{\ln n}\right\rceil, \ldots, \frac{n-1}{2}\right\}$, $k \in \{\beta \cdot l, \ldots, n-l\}$ and all $i \in \{0, \ldots, n-l-k\}$:*

$$\operatorname{med}(\,\xi_{i+1}^n, \ldots, \xi_{i+k}^n,\, \xi_{i+k+1}^n + \Delta, \ldots, \xi_{i+k+l}^n + \Delta\,) \le \chi. \tag{4.88}$$

Proof:
We fix an arbitrary $\Delta > 0$, arbitrary $\beta < \infty$ and arbitrary $\chi = \chi(\Delta, \beta)$, $\chi < \Delta$, which satisfies inequality (4.86).
Application of Corollary 3 (p.60) yields that in order to prove this lemma we have to show:

$$\sum_{n=11}^{\infty} \sum_{l=\lceil \frac{n}{\ln n}\rceil}^{\frac{n-1}{2}} \sum_{k=\beta\cdot l}^{n-l} \sum_{i=0}^{n-l-k} \mathbb{P}\Big(\sum_{j=i+1}^{i+k} \delta_{\xi_j^n}^{\chi} + \sum_{s=i+k+1}^{i+k+l} \delta_{\xi_s^n+\Delta}^{\chi} \ge \frac{k+l}{2}\Big) < \infty,$$

where, due to (4.11), for all $n \in \mathbb{N}$ and for any natural $j \le n$

$$\delta_{\xi_j^n}^{\chi} = \begin{cases} 1 & , \quad \xi_j^n > \chi \\ 0 & , \quad \text{else} \end{cases} \quad \text{and} \quad \delta_{\xi_j^n+\Delta}^{\chi} = \begin{cases} 1 & , \quad \xi_j^n + \Delta > \chi \\ 0 & , \quad \text{else} \end{cases}.$$

Taking into account that $\delta_{\xi_j^n+\Delta}^{\chi} = \delta_{\xi_j^n}^{\chi-\Delta}$, we conclude that in order to prove Lemma 16, one needs to show:

$$\sum_{n=11}^{\infty} \sum_{l=\lceil \frac{n}{\ln n}\rceil}^{\frac{n-1}{2}} \sum_{k=\beta\cdot l}^{n-l} \sum_{i=0}^{n-l-k} \mathbb{P}\Big(\sum_{j=i+1}^{i+k} \delta_{\xi_j^n}^{\chi} + \sum_{s=i+k+1}^{i+k+l} \delta_{\xi_s^n}^{\chi-\Delta} \ge \frac{k+l}{2}\Big) < \infty. \tag{4.89}$$

Again we notice that for all $n \in \mathbb{N}$ and for any $j \le n$ the random variables $\delta_{\xi_j^n}^{\chi}$, $\delta_{\xi_j^n}^{\chi-\Delta}$ are distributed according to the Bernoulli distribution with the parameters $P_{\xi_j^n}^{\chi}$ and $P_{\xi_j^n}^{\chi-\Delta}$ respectively, where $P_\eta^c = \mathbb{P}(\eta > c)$ for any random η and any real c.
For the sake of simplicity, let

$$p_1 \;:=\; P_\xi^\chi = \mathbb{P}\Big(\xi > \chi\Big), \tag{4.90}$$

$$p_2 \;:=\; P_\xi^{\chi-\Delta} = \mathbb{P}\Big(\xi > \chi - \Delta\Big), \tag{4.91}$$

where ξ is a random variable having the same distribution as ξ_i^n for all $n \in \mathbb{N}$ and all $i \in \{1, \ldots, n\}$. Due to these facts we see that

$$\mathbb{E}\Big(\sum_{j=i+1}^{i+k} \delta_{\xi_j^n}^{\chi}\Big) = k \cdot p_1, \quad i,\, k,\, n \in \mathbb{N},$$

$$\mathbb{E}\Big(\sum_{s=i+k+1}^{i+k+l} \delta_{\xi_s^n}^{\chi-\Delta}\Big) = l \cdot p_2, \quad i,\, k,\, l,\, n \in \mathbb{N}.$$

From this it follows:

$$\begin{aligned} &\mathbb{P}\Big(\sum_{j=i+1}^{i+k} \delta_{\xi_j^n}^{\chi} + \sum_{s=i+k+1}^{i+k+l} \delta_{\xi_s^n}^{\chi-\Delta} \ge \frac{k+l}{2}\Big) \\ =\; &\mathbb{P}\Big(\sum_{j=i+1}^{i+k} \delta_{\xi_j^n}^{\chi} + \sum_{s=i+k+1}^{i+k+l} \delta_{\xi_s^n}^{\chi-\Delta} - \mathbb{E}\Big(\sum_{j=i+1}^{i+k} \delta_{\xi_j^n}^{\chi} - \sum_{s=i+k+1}^{i+k+l} \delta_{\xi_s^n}^{\chi-\Delta}\Big) \ge \frac{k+l}{2} - k\cdot p_1 - l\cdot p_2\Big), \end{aligned}$$

where $n > 10$, $l \in \left\{ \left\lceil \frac{n}{\ln n} \right\rceil, \ldots, \frac{n-1}{2} \right\}$, $k \in \{\beta \cdot l, \ldots, n-l\}$ and $i \in \{0, \ldots, n-l-k\}$.
Because the considered χ $(0 < \chi < \Delta)$ fulfills inequality (4.86), then according to Lemma 15 it holds:

$$\beta\,(1-2p_1) + (1-2p_2) > 0. \tag{4.92}$$

This fact and the fact that $k \geq \beta \cdot l$ yield:

$$\begin{aligned} & \frac{k+l}{2} - k \cdot p_1 - l \cdot p_2 = \frac{1}{2} \cdot \big(k\,(1-2p_1) + l\,(1-2p_2) \big) \\ \geq\ & \frac{l}{2} \cdot \big(\beta\,(1-2p_1) + (1-2p_2) \big) > 0. \end{aligned} \tag{4.93}$$

Consequently, according to this, the probability

$$\mathbb{P}\Big(\sum_{j=i+1}^{i+k} \delta^{\chi}_{\xi^n_j} + \sum_{s=i+k+1}^{i+k+l} \delta^{\chi-\Delta}_{\xi^n_s} - \mathbb{E}\Big(\sum_{j=i+1}^{i+k} \delta^{\chi}_{\xi^n_j} - \sum_{s=i+k+1}^{i+k+l} \delta^{\chi-\Delta}_{\xi^n_s} \Big) \geq \frac{k+l}{2} - k \cdot p_1 - l \cdot p_2 \Big)$$

and, correspondingly, the probability $\mathbb{P}\Big(\sum\limits_{j=i+1}^{i+k} \delta^{\chi}_{\xi^n_j} + \sum\limits_{s=i+k+1}^{i+k+l} \delta^{\chi-\Delta}_{\xi^n_s} \geq \frac{k+l}{2} \Big)$ can be estimated with the help of Hoeffding's inequality (A.23), p.121.
Namely,

$$\begin{aligned} & \sum_{n=11}^{\infty} \sum_{l=\lceil \frac{n}{\ln n} \rceil}^{\frac{n-1}{2}} \sum_{k=\beta \cdot l}^{n-l} \sum_{i=0}^{n-l-k} \mathbb{P}\Big(\sum_{j=i+1}^{i+k} \delta^{\chi}_{\xi^n_j} + \sum_{s=i+k+1}^{i+k+l} \delta^{\chi-\Delta}_{\xi^n_s} \geq \frac{k+l}{2} \Big) \\ \leq\ & \sum_{n=11}^{\infty} \sum_{l=\lceil \frac{n}{\ln n} \rceil}^{\frac{n-1}{2}} \sum_{k=\beta \cdot l}^{n-l} \sum_{i=0}^{n-l-k} \exp\Bigg\{ -2\, \frac{\frac{1}{4} \cdot \big(k\,(1-2p_1) + l\,(1-2p_2) \big)^2}{k+l} \Bigg\} \\ \leq\ & \sum_{n=11}^{\infty} \sum_{l=\lceil \frac{n}{\ln n} \rceil}^{\frac{n-1}{2}} \sum_{k=\beta \cdot l}^{n-l} n \cdot \exp\Bigg\{ -2\, \frac{\frac{1}{4} \cdot \big(k\,(1-2p_1) + l\,(1-2p_2) \big)^2}{n} \Bigg\} \\ \underset{(4.93)}{\leq}\ & \sum_{n=11}^{\infty} \sum_{l=\lceil \frac{n}{\ln n} \rceil}^{\frac{n-1}{2}} \sum_{k=\beta \cdot l}^{n-l} n \cdot \exp\Bigg\{ -\frac{l^2 \big(\beta\,(1-2p_1) + (1-2p_2) \big)^2}{2\,n} \Bigg\} \\ \leq\ & \sum_{n=11}^{\infty} \sum_{l=\lceil \frac{n}{\ln n} \rceil}^{\frac{n-1}{2}} n^2 \cdot \exp\Bigg\{ -\frac{n^2}{\ln^2 n} \cdot \frac{\big(\beta\,(1-2p_1) + (1-2p_2) \big)^2}{2\,n} \Bigg\} \\ \leq\ & \sum_{n=11}^{\infty} n^3 \cdot \exp\Bigg\{ -\frac{n \big(\beta\,(1-2p_1) + (1-2p_2) \big)^2}{2\,\ln^2 n} \Bigg\} \underset{(4.92),\,(A.1)}{<} \infty, \end{aligned}$$

Consequently, according to (4.89), we have completed the proof of Lemma 16.

□

4.3 Auxiliary result

We finish this chapter with one auxiliary lemma, that yields an estimate for certain sums, which contain the variables $(\xi_i^n)_{1\le i\le n,\, n\in\mathbb{N}}$. This lemma will be used later in Chapter 5.

Lemma 17. *Let $(\xi_i^n)_{1\le i\le n,\, n\in\mathbb{N}}$ fulfill Condition 2. Then for all sequences $(\gamma_n)_{n\in\mathbb{N}}$ such that $\ln n\cdot\gamma_n \underset{n\to\infty}{\longrightarrow} 0$, for any finite $\Delta'>0$, for any finite $\beta>0$ the following statement is $\mathbb{P}$-almost surely valid for all sufficiently large $n\in\mathbb{N}$ such that $n>10$, all $l\in\left\{\left\lceil\frac{n}{\ln n}\right\rceil,\dots,\frac{n-1}{2}\right\}$, $k\in\{\beta\cdot l,\dots,n-l\}$ and all $i\in\{0,\dots,n-l-k\}$:*

$$\sum_{j=i+k+1}^{i+k+l}\big(|\Delta'+\xi_j^n|-|\xi_j^n|\big)>3\cdot n\cdot\gamma_n.$$

Proof:
We fix an arbitrary sequence $(\gamma_n)_{n\in\mathbb{N}}$ such that $\ln n\cdot\gamma_n \underset{n\to\infty}{\longrightarrow} 0$, an arbitrary $\Delta'>0$ and an arbitrary finite $\beta>0$.
In order to proof this lemma, by applying Borel-Cantelli Lemma, we need to show:

$$\sum_{n=11}^{\infty}\sum_{l=\lceil\frac{n}{\ln n}\rceil}^{\frac{n-1}{2}}\sum_{k=\beta\cdot l}^{n-l}\sum_{i=0}^{n-l-k}\mathbb{P}\Big(\sum_{j=i+k+1}^{i+k+l}\big(|\xi_j^n|-|\Delta'+\xi_j^n|\big)\ge-3\cdot n\cdot\gamma_n\Big)<\infty. \tag{4.94}$$

We start by considering $\mathbb{P}\Big(\sum_{j=i+k+1}^{i+k+l}\big(|\xi_j^n|-|\Delta'+\xi_j^n|\big)\ge-3\cdot n\cdot\gamma_n\Big)$ for arbitrary natural n, l, k and i :

$$\begin{aligned}&\mathbb{P}\Big(\sum_{j=i+k+1}^{i+k+l}\big(|\xi_j^n|-|\Delta'+\xi_j^n|\big)\ge-3\cdot n\cdot\gamma_n\Big)\\ =\;&\mathbb{P}\Big(\sum_{j=i+k+1}^{i+k+l}\big(|\xi_j^n|-|\Delta'+\xi_j^n|\big)-l\cdot\mathbb{E}\big(|\xi|-|\Delta'+\xi|\big)\ge-3\cdot n\cdot\gamma_n-l\cdot\mathbb{E}\big(|\xi|-|\Delta'+\xi|\big)\Big)\end{aligned}$$

From (2.5) we know that for all $n,\, j\in\mathbb{N}$ it holds:

$$-\Delta'\le|\xi_j^n|-|\Delta'+\xi_j^n|\le\Delta'.$$

Furthermore, because $E(\Delta',\alpha):=\mathbb{E}\big(|\Delta'+\xi|-|\xi|\big)>0$ (see Proposition 8, Appendix A, p.130) and because we consider such l and such γ_n that satisfy the conditions $l\ge\frac{n}{\ln n}$ and $\ln n\cdot\gamma_n\underset{n\to\infty}{\longrightarrow}0,$ we can conclude that there exists such a natural number $10<N_1<\infty$ that for some fixed ε_0 with

$$0<\varepsilon_0<E(\Delta',\alpha) \tag{4.95}$$

and for all $n\ge N_1$, $n\in\mathbb{N}$ the following statement is valid:

$$l\cdot E(\Delta',\alpha)-3\cdot n\cdot\gamma_n\ge\frac{n}{\ln n}\Big(E(\Delta',\alpha)-3\cdot\ln n\cdot\gamma_n\Big)\ge\frac{n}{\ln n}\Big(E(\Delta',\alpha)-\varepsilon_0\Big)>0. \tag{4.96}$$

Therefore in order to estimate the quantity

$$\mathbb{P}\Big(\sum_{j=i+k+1}^{i+k+l}\big(|\xi_j^n|-|\Delta'+\xi_j^n|\big)-l\cdot\mathbb{E}\big(|\xi|-|\Delta'+\xi|\big)\geq -3\cdot n\cdot\gamma_n+l\cdot E(\Delta',\alpha)\Big),\ n\geq N_1$$

we can apply Hoeffding's inequality (A.23).
As the result we obtain :

$$\begin{aligned}
&\sum_{n=N_1}^{\infty}\sum_{l=\lceil\frac{n}{\ln n}\rceil}^{\frac{n-1}{2}}\sum_{k=\beta\cdot l}^{n-l}\sum_{i=0}^{n-l-k}\mathbb{P}\Big(\sum_{j=i+k+1}^{i+k+l}\big(|\xi_j^n|-|\Delta'+\xi_j^n|\big)\geq -3\cdot n\cdot\gamma_n\Big)\\
\leq\ &\sum_{n=N_1}^{\infty}\sum_{l=\lceil\frac{n}{\ln n}\rceil}^{\frac{n-1}{2}}\sum_{k=\beta\cdot l}^{n-l}\sum_{i=0}^{n-l-k}\exp\Big\{-2\,\frac{\big(-3\cdot n\cdot\gamma_n+l\cdot E(\Delta',\alpha)\big)^2}{4\cdot l\cdot(\Delta')^2}\Big\}\\
\leq\ &\frac{1}{2\,(\Delta')^2}\sum_{n=N_1}^{\infty}\sum_{l=\lceil\frac{n}{\ln n}\rceil}^{\frac{n-1}{2}}n^2\exp\Big\{-\frac{\big(-3\cdot n\cdot\gamma_n+l\cdot E(\Delta',\alpha)\big)^2}{l}\Big\}.
\end{aligned}\tag{4.97}$$

Now for an arbitrary $n\geq N_1$ we consider $h_n(l):=\dfrac{\big(l\cdot E(\Delta',\alpha)-3\cdot n\cdot\gamma_n\big)^2}{l}$.

$$\begin{aligned}
h_n'(l)\ &=\ \frac{2\cdot l\cdot E(\Delta',\alpha)\cdot\big(l\cdot E(\Delta',\alpha)-3\cdot n\cdot\gamma_n\big)-\big(l\cdot E(\Delta',\alpha)-3\cdot n\cdot\gamma_n\big)^2}{l^2}\\
&=\ \frac{\big(l\cdot E(\Delta',\alpha)-3\cdot n\cdot\gamma_n\big)\cdot\big(l\cdot E(\Delta',\alpha)+3\cdot n\cdot\gamma_n\big)}{l^2}\underset{(4.96)}{>}0.
\end{aligned}$$

Consequently, for all $n\geq N_1$ the function h_n is monotonically increasing and, respectively, the function $\exp\Big\{-\dfrac{\big(-3\cdot n\cdot\gamma_n+l\cdot E(\Delta',\alpha)\big)^2}{l}\Big\}$ as the function of l is monotonically decreasing, which means that for all $l\in\Big\{\Big\lceil\dfrac{n}{\ln n}\Big\rceil,\dots,\dfrac{n-1}{2}\Big\}$ $(n\geq N_1)$

$$\exp\Big\{-\frac{\big(-3\cdot n\cdot\gamma_n+l\cdot E(\Delta',\alpha)\big)^2}{l}\Big\}\leq\exp\Big\{-\frac{\big(-3\cdot n\cdot\gamma_n+\frac{n}{\ln n}\cdot E(\Delta',\alpha)\big)^2}{\frac{n}{\ln n}}\Big\}.\tag{4.98}$$

Hence,

$$\begin{aligned}
&\sum_{n=N_1}^{\infty}\sum_{l=\lceil\frac{n}{\ln n}\rceil}^{\frac{n-1}{2}}\sum_{k=\beta\cdot l}^{n-l}\sum_{i=0}^{n-l-k}\mathbb{P}\Big(\sum_{j=i+k+1}^{i+k+l}\big(|\xi_j^n|-|\Delta'+\xi_j^n|\big)\geq -3\cdot n\cdot\gamma_n\Big)\\
\underset{(4.97),(4.98)}{\leq}\ &\frac{1}{2\,(\Delta')^2}\sum_{n=N_1}^{\infty}n^3\exp\Big\{-\frac{\big(-3\cdot n\cdot\gamma_n+\frac{n}{\ln n}\cdot E(\Delta',\alpha)\big)^2}{\frac{n}{\ln n}}\Big\}\\
=\ &\frac{1}{2\,(\Delta')^2}\sum_{n=N_1}^{\infty}n^3\exp\Big\{-\frac{n\,\big(E(\Delta',\alpha)-3\cdot\ln n\cdot\gamma_n\big)^2}{\ln n}\Big\}\\
\underset{(4.95),(4.96)}{\leq}\ &\frac{1}{2\,(\Delta')^2}\sum_{n=N_1}^{\infty}n^3\exp\Big\{-\frac{n\,\big(E(\Delta',\alpha)-\varepsilon_0\big)^2}{\ln n}\Big\}
\end{aligned}\tag{4.99}$$

Due to (4.96), the term $(E(\Delta', \alpha) - \varepsilon_0)$ is strictly positive, which means that according to (A.1) in Proposition 6, the series $\frac{1}{2\,(\Delta')^2} \sum\limits_{n=N_1}^{\infty} n^3 \exp\left\{-\frac{n\,\left(E(\Delta', \alpha) - \varepsilon_0\right)^2}{\ln n}\right\}$ is convergent.
From (4.99) it follows immediately that the series

$$\sum_{n=11}^{\infty} \sum_{l=\lceil \frac{n}{\ln n} \rceil}^{\frac{n-1}{2}} \sum_{k=\beta \cdot l}^{n-l} \sum_{i=0}^{n-l-k} \mathbb{P}\Big(\sum_{j=i+k+1}^{i+k+l} \left(|\xi_j^n| - |\Delta' + \xi_j^n| \right) \geq -3 \cdot n \cdot \gamma_n \Big)$$

is also convergent.
Therefore the validity of (4.94) and, consequently, of Lemma 17 is proved.

□

Lemma 17 concludes this chapter. Below we briefly summarize all the results obtained in this chapter.

4.4 Conclusion

In this chapter we have defined new condition on the noise $(\xi_i^n)_{1 \leq i \leq n,\, n \in \mathbb{N}}$, namely we assumed that for all $n \in \mathbb{N}$ and all $i \in \{1, \ldots, n\}$ the random variables ξ_i^n are independent identically distributed according to the symmetrical Pareto distribution with the distribution function

$$F_{\xi_i^n}(x) = \begin{cases} \dfrac{1}{2\,(1-x)^{\alpha-1}} & , \quad x < 0 \\ 1 - \dfrac{1}{2\,(1+x)^{\alpha-1}} & , \quad x \geq 0 \end{cases} .$$

This distribution was chosen as an example of heavy tailed distributions, which is typically used to describe random data that contains outliers.

Further on, we have studied the properties of the noise, which satisfies the above condition. More precisely, we have firstly shown that the function $\xi^n = \sum\limits_{i=1}^{n} \xi_i^n \cdot \mathbf{1}_{\left[\frac{i-1}{n}, \frac{i}{n}\right)}$ has $\mathbb{P}$-almost surely finite norm.
Secondly, for all sufficiently large n and for an arbitrary sample of data noise $(\xi_1^n, \ldots, \xi_n^n)$ we derived an estimate of the medians and other quantiles of arbitrary subsets of these data with different number of elements. The choice of these estimates was motivated by the strong law of large numbers [4], namely any quantile, in particular any median, as computed from a data set with n elements was assumed to converge to the theoretical value of the corresponding quantile no faster than $\frac{1}{\sqrt{n}}$. More specifically, the rate of convergence was estimated to be given by $\frac{\ln n}{\sqrt{n}}$.
We have shown that these estimates of the medians and other quantiles can be validated if the parameter α in the Pareto distribution is chosen to be $\alpha \geq 9$. This choice of α together with the assumption on the distribution function of the noise is referred to throughout Chapter 4 and the rest of the thesis as Condition 2.

Finally, we have estimated the so called mixed medians. These are determined as the medians of the type $\operatorname{med}(\xi_{i+1}^n, \ldots, \xi_{i+k}^n, \xi_{i+k+1}^n + \Delta, \ldots, \xi_{i+k+l}^n + \Delta)$, where Δ is an arbitrary real constant, n is a sufficiently large natural number, $l \in \{1, \ldots, n\}$, $k \in \{1, \ldots, n-l\}$ and $i \in \{0, \ldots, n-l-k\}$. When estimating the mixed medians we have considered two different cases: in case (i) only one of the two numbers k and l is greater than or equal to $\frac{n}{\ln n}$, in case (ii), one of the two numbers k and l is not smaller than $\frac{n}{\ln n}$ and the other one is of the same order or greater. In case (i) we have estimated the mixed medians with the help of characteristics of the elements of the longer subset and in case (ii) we have shown that the mixed median can be estimated by a certain constant that depends on Δ and is always smaller than Δ. Moreover, case (i) was generalized for the medians of the type $\operatorname{med}(\xi_{i+1}^n + \Delta_1, \ldots, \xi_{i+k}^n + \Delta_1, \xi_{i+k+1}^n, \ldots, \xi_{i+k+l}^n, \xi_{i+k+l+1}^n + \Delta_2, \ldots, \xi_{i+k+l+s}^n + \Delta_2)$, where $\Delta_1, \Delta_2 \in \mathbb{R}$, $n \in \mathbb{N}$, $l \in \left\{\left\lceil \frac{n}{\ln n} \right\rceil, \ldots, n\right\}$, $k \in \{1, \ldots, n-l\}$, $s \in \{1, \ldots, n-l-k\}$, $s + k < l$ and $i \in \{0, \ldots, n-l-k-s\}$.

All the obtained estimates are valid for any arbitrary median or q-quantile taken from the sets of all possible medians or q-quantiles of the corresponding data sets, respectively.

All results obtained in Chapter 4 will be used in the next chapter to prove that for the data $y_i^n = f_i^n + \xi_i^n$, $(n \in \mathbb{N}, i \in \{1, \ldots, n\})$ the minimizers of the functional $\tilde{H}_{\gamma_n}^n(\cdot, f^n + \xi^n)$, which is firstly defined in Chapter 2 by (2.1), approach the original function f in the limit of infinitely large n.

Chapter 5

Consistency and rates

In this chapter we investigate the asymptotic behaviour of the minimizers of the functional $\tilde{H}^n_\gamma$.
Again recall that according to the formulation of the problem, for some function $f \in L^1((0,1))$ one observes the noisy data $y^n_i = f^n_i + \xi^n_i$, $(n \in \mathbb{N},\ i \in \{1, \dots, n\})$, which are assumed to be known. Further, $\tilde{H}^n_{\gamma_n}(g\,, f^n + \xi^n)$ in case, when $g \in S_n[(0,1))$, is defined according to (2.1) as follows:

$$\tilde{H}^n_{\gamma_n}(g\,, f^n + \xi^n) = \gamma \# J(g) + \|g - (f^n + \xi^n)\| - \|f^n + \xi^n\|,$$

where $f^n(x) = \sum\limits_{i=1}^{n} f^n_i \cdot \mathbf{1}_{\left[\frac{i-1}{n}, \frac{i}{n}\right)}(x)$ and $\xi^n(x) = \sum\limits_{i=1}^{n} \xi^n_i \cdot \mathbf{1}_{\left[\frac{i-1}{n}, \frac{i}{n}\right)}(x)$.

The main objective of this chapter is to determine conditions on the sequence of $(\gamma_n)_{n\in\mathbb{N}}$ such that the corresponding sequence $(\widehat{f^n})_{n\in\mathbb{N}}$, which fulfills $\widehat{f^n} \in \operatorname{argmin} \tilde{H}^n_{\gamma_n}(\cdot\,, f^n + \xi^n)$, $\mathbb{P}$-almost surely tends to the function $f \in S([0,\,1))$. This property is first proved for the case when f is a constant function, following by the cases of a step function with one, two and, subsequently, an arbitrary number of jumps.

We also remark that similar to the previous chapter, we continue to work in this chapter with the noise, which satisfies Condition 2, i.e with the noise, which is symmetrical Pareto distributed with $\alpha \geq 9$, as stated in (4.1). Therefore, all the properties of the noise, obtained in the previous chapter can also be used here.
As in Chapter 4, for a given data set $\eta_1, \dots, \eta_n$ by med $(\eta_1, \dots, \eta_n)$ or q-quantile $(\eta_1, \dots, \eta_n)$ we imply an arbitrary median or quantile, from the set $med(\eta_1, \dots, \eta_n)$ of all medians, or from the set of all q-quantiles of this data set, respectively. The same convention is valid for the median of a functions, i.e. by $\operatorname{med}_I(f)$ for some function f and some $I \subseteq [0,1)$ we imply an arbitrary element from the set $med(f) = \operatorname*{argmin}\limits_{c\in\mathbb{R}} \int\limits_I |c - f(x)|\, dx$.

5.1 Number of jumps of $\widehat{f^n}$

In this section we consider the number of jumps of the minimizers of the functional $\tilde{H}_\gamma(\cdot\,, f^n + \xi^n)$.
Before moving on to the main results of this section we return to the function f^n from $S_n\big([0,1)\big)$, which, according to (2.10) and (2.11), is an approximation of the true function

f in $S_n\big([0,1)\big)$ and satisfies the condition $\lim\limits_{n\to\infty} f^n = f$. Namely we give a more precise description of the structure of f^n in case, when $f \in S\big([0,1)\big)$.

Notice that in the proof of Lemma 4, we have already used a particular choice of f^n, used below. However, for convenience we recall here the details of this construction.

Definition 11. *Let* $f(x) := \sum\limits_{i=1}^{\#J(f)+1} f_i \cdot \mathbf{1}_{[a_{i-1},\,a_i)}(x)$ *be a function from the set* $S\big([0,1)\big)$, *where* $\#J(f)$ *is the number of jumps of* f, $f_1,\dots,f_{\#J(f)+1}$ *are finite real numbers, and* $a_0 := 0$, $a_{\#J(f)+1} := 1$, $a_i \in (0,1)$ *with* $a_{i-1} < a_i$ *for all* $i \in \{1,\dots \#J(f)+1\}$.
For any $n \in \mathbb{N}$ *we define a natural number*

$$u_n = \min\big\{n,\ \#J(f)+1\big\},$$

and the points $m_0^n \le m_1^n \le \dots \le m_{u_n}^n$ *as some points from the set* $\{0,\dots n\}$, *which obey the following property:*

$$\begin{aligned} &m_0^n := 0,\ \ m_{u_n}^n := n\\ &0 \le \Big(\frac{m_i^n}{n} - a_i\Big) < \frac{1}{n}, \quad i \in \{1,\dots u_n\}. \end{aligned} \tag{5.1}$$

Finally, the function $f^n \in S_n\big([0,1)\big)$ *is defined as follows:*

$$f^n(x) = \sum_{i=1}^{u_n} f\Big(\frac{m_{i-1}^n}{n}\Big) \cdot \mathbf{1}_{\big[\frac{m_{i-1}^n}{n},\,\frac{m_i^n}{n}\big)}(x). \tag{5.2}$$

We notice here that for all $n \in \mathbb{N}$ the number of jumps of the function f^n is no larger than the number of jumps of the function f.
Moreover, if we define the following number N_f :

$$N_f := \Bigg\lceil \frac{1}{\min\limits_i\,(a_i - a_{i-1})} \Bigg\rceil \tag{5.3}$$

then for all $n \ge N_f$ the definition of f^n can be rewritten as follows:

$$f^n(x) = \sum_{i=1}^{\#J(f)+1} f_i \cdot \mathbf{1}_{\big[\frac{m_{i-1}^n}{n},\,\frac{m_i^n}{n}\big)}(x). \tag{5.4}$$

This means for all $n \ge N_f$ the function $f^n \in S_n\big([0,1)\big)$ is the right-shift step function $f \in S\big([0,1)\big)$. Clearly in this case the number of jumps of both function is identical, i.e.

$$\#J(f^n) = \#J(f). \tag{5.5}$$

The next lemma yields the upper bound for the number of jumps of $\widehat{f}_n \in \operatorname{argmin} \tilde{H}_\gamma(\cdot\,, f^n + \xi^n)$, which depends on the number of jumps of f^n.

Lemma 18. *Let* $f \in L^1\big([0,1)\big)$, $(\xi_i^n)_{1\le i\le n,\, n\in\mathbb{N}}$ *fulfill Condition 2 and* $\widehat{f}_n \in \operatorname{argmin} \tilde{H}_\gamma\,(\cdot\,, f^n + \xi^n)$.
Then for all $n \in \mathbb{N}$ *and any* $\gamma > 0$ *the following inequality is valid:*

$$\#J(\widehat{f}_n) \le \frac{\|\xi^n\|}{\gamma} + \#J(f^n). \tag{5.6}$$

Proof:
For any $\gamma > 0$ we consider $\tilde{H}^n_\gamma\,(\widehat{f}_n\,, f^n + \xi^n)$ and $\tilde{H}^n_\gamma\,(f^n\,, f^n + \xi^n)$.

$$\begin{aligned}
\tilde{H}^n_\gamma\,(\widehat{f}_n\,, f^n + \xi^n) &= \gamma\#J(\widehat{f}_n) + \|\widehat{f}_n - (f^n + \xi^n)\| - \|f^n + \xi^n\|, \\
\tilde{H}^n_\gamma\,(f^n\,, f^n + \xi^n) &= \gamma\#J(f^n) + \|f^n - (f^n + \xi^n)\| - \|f^n + \xi^n\| \\
&= \gamma\#J(f^n) + \|\xi^n\| - \|f^n + \xi^n\|.
\end{aligned}$$

Since $\widehat{f}_n \in \operatorname{argmin} \tilde{H}_\gamma\,(\cdot\,, f^n + \xi^n)$, it is clear that it holds :

$$\tilde{H}_\gamma\,(\widehat{f}_n\,, f^n + \xi^n) \leq \tilde{H}_\gamma\,(f^n\,, f^n + \xi^n).$$

Hence it also holds:

$$\gamma\#J(\widehat{f}_n) + \|\widehat{f}_n - f^n - \xi^n\| \leq \gamma\#J(f^n) + \|\xi^n\|$$

$$\Rightarrow \qquad \gamma\#J(\widehat{f}_n) \leq \gamma\#J(f^n) + \|\xi^n\|$$

$$\Rightarrow \qquad \#J(\widehat{f}_n) \leq \frac{\|\xi^n\|}{\gamma} + \#J(f^n)$$

Consequently we have proved Lemma 18.

□

In the following lemma we show that in certain cases the value of the functional $\tilde{H}^n_\gamma\,(\cdot\,, f^n + \xi^n)$ can be decreased if the number of jumps of its minimizer is increased.

Prior to this, we introduce some auxiliary notation.

First of all, we define some condition on the sequence of parameters $(\gamma_n)_{n\in\mathbb{N}}$.

Condition 3. *The sequence* $(\gamma_n)_{n\in\mathbb{N}} \subset (0,\infty)$ *satisfies*

$$\ln n \cdot \gamma_n \underset{n\to\infty}{\longrightarrow} 0.$$

This condition should be met in order to ensure the validity of the next lemma and, therefore, the validity of the lemmas and theorems, which follow from this lemma.

Further on, for the sake of brevity, for any $n \in \mathbb{N}$, an arbitrary $k \in \{1, \ldots, n\}$ and arbitrary $i \in \{0, \ldots, n-k\}$ by $I^n_{i,k}$ we denote the following interval:

$$I^n_{i,k} := \Big[\frac{i}{n}, \frac{i+k}{n}\Big) \tag{5.7}$$

Notice that from this notation it is obvious that for any $n \in \mathbb{N}$, any $l \in \{1, \ldots, n\}$ any $k \in \{0, \ldots, n-l\}$ and any $i \in \{0, \ldots, n-l-k\}$ it holds:

$$I^n_{i,k+l} = I^n_{i,k} \cup I^n_{i+k,l}\,. \tag{5.8}$$

Finally, for any $I \subseteq [0,1)$ by I^C we denote in a standard way the complement of I, i.e

$$I^C = [0,1) \setminus I.$$

Lemma 19. *Let* $f \in S\big([0,1)\big)$, $(\xi_i^n)_{1\le i\le n,\, n\in\mathbb{N}}$ *fulfill Condition 2 and* $\widehat{f}_n \in \operatorname{argmin}\ \tilde{H}^n_{\gamma_n}(\cdot\,, f^n+\xi^n)$.
For all sequences $(\gamma_n)_{n\in\mathbb{N}}$*, which satisfy Condition 3, and for any finite* $\beta>0$*, the following statement is* $\mathbb{P}$*-almost surely valid for all sufficiently large* $n\in\mathbb{N}$*, all* $l \ge \dfrac{n}{\ln n}$, $k\ge \beta\cdot l$ *and all* $i\in\{0,\dots,n-l-k\}$ *such that the function* $f^n=\sum_{i=1}^{n} f_i^n\cdot \mathbf{1}_{\left[\frac{i-1}{n},\,\frac{i}{n}\right)}$ *on the interval* $I^n_{i,\,k+l}$*, as defined in (5.7), has exactly one point of discontinuity at* $\dfrac{i+k}{n}$*:*
If one considers the two following functions $\widehat{f}_n^1$ *and* $\widehat{f}_n^2$ *:*

$$\widehat{f}_n^1 \;:=\; \widehat{f}_n\cdot \mathbf{1}_{(I^n_{i,\,k+l})^C} + \operatorname{med}_{I^n_{i,\,k+l}}(f^n+\xi^n)\cdot \mathbf{1}_{I^n_{i,\,k+l}}\,, \tag{5.9}$$

$$\widehat{f}_n^2 \;:=\; \widehat{f}_n\cdot \mathbf{1}_{(I^n_{i,\,k+l})^C} + \operatorname{med}_{I^n_{i,\,k}}(f^n+\xi^n)\cdot \mathbf{1}_{I^n_{i,\,k}} + \operatorname{med}_{I^n_{i+k,\,l}}(f^n+\xi^n)\cdot \mathbf{1}_{I^n_{i+k,\,l}} \tag{5.10}$$

then it holds

$$\tilde{H}^n_{\gamma_n}(\widehat{f}_n^1\,, f^n+\xi^n) > \tilde{H}^n_{\gamma_n}(\widehat{f}_n^2\,, f^n+\xi^n). \tag{5.11}$$

Proof:

First of all we fix an arbitrary sequence $(\gamma_n)_{n\in\mathbb{N}}$, which satisfies the Condition 3, and an arbitrary finite $\beta>0$.
For all sufficiently large $n\in\mathbb{N}$, all $l\ge \dfrac{n}{\ln n}$, $k\ge\beta\cdot l$ and $i\in\{0,\dots,n-l-k\}$ we consider an interval $I^n_{i,\,k+l}$ such that the function f^n on this interval has exactly one jump at $\dfrac{i+k}{n}$.
According to this and also taking (5.7) into account, the function f^n on the interval $I^n_{i,\,k+l}$ can be represented as follows:

$$\forall\, x\in I^n_{i,\,k} \qquad f^n(x)=\widetilde{f}_1,\quad \widetilde{f}_1\in\mathbb{R}, \tag{5.12}$$

$$\forall\, x\in I^n_{i+k,\,l} \quad f^n(x)=\widetilde{f}_2,\quad \widetilde{f}_2\in\mathbb{R},\quad \widetilde{f}_2\neq\widetilde{f}_1. \tag{5.13}$$

Notice that due to the fact that $f\in S\big([0,1)\big)$, i.e. f is a step function with a finite number of jumps, and also taking definition 11 of f^n into account, for all n the function f^n is also a step function with a finite number of jumps. Consequently, $\widetilde{f}_1$ and $\widetilde{f}_2$ both belong to the finite set $f\big([0,1)\big)$.
Now let us discuss the difference between the number of jumps of the functions $\widehat{f}_n^1$ and $\widehat{f}_n^2$. For our calculation it will be helpful to determine the largest possible value of the difference $\#(\widehat{f}_n^2)-\#(\widehat{f}_n^1)$. It is not difficult to see, that in the majority of cases, the function $\widehat{f}_n^1$ has one jump less than $\widehat{f}_n^2$. However, it may also happen that $\widehat{f}_n^1$ has three jumps less than $\widehat{f}_n^2$, if for instance $\widehat{f}_n^1$ is a constant function and $\widehat{f}_n^2$ has exactly three jumps. Conversely, in case when $\widehat{f}_n^1$ is constant in $\Big(\dfrac{i-1}{n}, \dfrac{i+k+l+1}{n}\Big]$, according to (5.9), (5.10), the function $\widehat{f}_n^2$ can have no more than three jumps.
Therefore, without loss of generality we can conclude that $\widehat{f}_n^1$ has at most 3 jumps less than $\widehat{f}_n^2$, i.e. it holds:

$$\#J(\widehat{f}_n^1)-\#J(\widehat{f}_n^2)\ge -3. \tag{5.14}$$

With (5.9), (5.10), (5.12), (5.13) and (5.14) in mind, we consider $\tilde{H}^n_{\gamma_n}(\widehat{f}^1_n, f^n+\xi^n) - \tilde{H}^n_{\gamma_n}(\widehat{f}^2_n, f^n+\xi^n)$:

$$
\begin{aligned}
&\tilde{H}^n_{\gamma_n}(\widehat{f}^1_n, f^n+\xi^n) - \tilde{H}^n_{\gamma_n}(\widehat{f}^2_n, f^n+\xi^n) \\
\underset{(2.1)}{=}\ & \gamma_n \# J(\widehat{f}^1_n) + \|\widehat{f}^1_n - (f^n+\xi^n)\| - \|f^n+\xi^n\| \\
& - \gamma_n \# J(\widehat{f}^2_n) - \|\widehat{f}^2_n - (f^n+\xi^n)\| + \|f^n+\xi^n\| \\
\geq\ & -3\cdot\gamma_n + \|\widehat{f}^1_n - (f^n+\xi^n)\| - \|\widehat{f}^2_n - (f^n+\xi^n)\| \\
=\ & -3\cdot\gamma_n + \|(\widehat{f}_n - (f^n+\xi^n))\cdot \mathbf{1}_{(I^n_{i,\,k+l})^C}\| + \sum_{j\in I^n_{i,\,k+l}} \frac{1}{n}\cdot |\operatorname{med}_{I^n_{i,\,k+l}}(f^n+\xi^n) - (f^n_j+\xi^n_j)| \\
& - \|(\widehat{f}_n - (f^n+\xi^n))\cdot \mathbf{1}_{(I^n_{i,\,k+l})^C}\| - \sum_{j\in I^n_{i,\,k}} \frac{1}{n}\cdot |\operatorname{med}_{I^n_{i,\,k}}(f^n+\xi^n) - (f^n_j+\xi^n_j)| \\
& - \sum_{j\in I^n_{i+k,\,l}} \frac{1}{n}\cdot |\operatorname{med}_{I^n_{i+k,\,l}}(f^n+\xi^n) - (f^n_j+\xi^n_j)| \\
=\ & -3\cdot\gamma_n + \frac{1}{n}\sum_{j=i+1}^{i+k}\Big(|\operatorname{med}_{I^n_{i,\,k+l}}(f^n+\xi^n) - \widetilde{f}_1 - \xi^n_j| - |\operatorname{med}_{I^n_{i,\,k}}(f^n+\xi^n) - \widetilde{f}_1 - \xi^n_j|\Big) \\
& + \frac{1}{n}\sum_{j=i+k+1}^{i+k+l}\Big(|\operatorname{med}_{I^n_{i,\,k+l}}(f^n+\xi^n) - \widetilde{f}_2 - \xi^n_j| - |\operatorname{med}_{I^n_{i+k,\,l}}(f^n+\xi^n) - \widetilde{f}_2 - \xi^n_j|\Big)
\end{aligned}
$$

From the definitions of the intervals $I^n_{i,\,k},\ I^n_{i+k,\,l},\ I^n_{i,\,k+l}$ and the function f^n on this intervals we have:

$$
\begin{aligned}
\operatorname{med}_{I^n_{i,\,k}}(f^n+\xi^n) &= \operatorname{med}\big(\xi^n_{i+1}+\widetilde{f}_1,\ldots,\xi^n_{i+k}+\widetilde{f}_1\big) = \operatorname{med}\big(\xi^n_{i+1},\ldots,\xi^n_{i+k}\big)+\widetilde{f}_1 \\
\operatorname{med}_{I^n_{i+k,\,l}}(f^n+\xi^n) &= \operatorname{med}\big(\xi^n_{i+k+1}+\widetilde{f}_2,\ldots,\xi^n_{i+k+l}+\widetilde{f}_2\big) = \operatorname{med}\big(\xi^n_{i+k+1},\ldots,\xi^n_{i+k+l}\big)+\widetilde{f}_2 \\
\operatorname{med}_{I^n_{i,\,k+l}}(f^n+\xi^n) &= \operatorname{med}\big(\xi^n_{i+1}+\widetilde{f}_1,\ldots,\xi^n_{i+k}+\widetilde{f}_1,\ \xi^n_{i+k+1}+\widetilde{f}_2,\ldots,\xi^n_{i+k+l}+\widetilde{f}_2\big)
\end{aligned}
$$

Without loss of generality we can assume that $\widetilde{f}_1 < \widetilde{f}_2$. Now we define

$$\Delta := \widetilde{f}_2 - \widetilde{f}_1 \neq 0.$$

Observe that because, as mentioned above, $\widetilde{f}_2$ and $\widetilde{f}_1$ belong to the finite set $f\big([0,\,1)\big)$, it is clear that Δ can be chosen only from the finite set $\big\{f(x)-f(x-0),\ x\in J(f)\big\}$.

Taking into account the definition of Δ, it is easy to see that it holds:

$$\operatorname{med}_{I^n_{i,\,k+l}}(f^n+\xi^n) = \operatorname{med}\,(\,\xi^n_{i+1},\ldots,\xi^n_{i+k},\ \xi^n_{i+k+1}+\Delta,\ldots,\xi^n_{i+k+l}+\Delta\,)+\widetilde{f}_1.$$

Thus, we obtain:

$$\begin{aligned}
& \tilde{H}^n_{\gamma_n}(\widehat{f}^1_n, f^n+\xi^n) - \tilde{H}^n_{\gamma_n}(\widehat{f}^2_n, f^n+\xi^n) \\
\geq \quad -3\cdot\gamma_n \quad & + \frac{1}{n}\sum_{j=i+1}^{i+k} |\operatorname{med}(\xi^n_{i+1},\ldots,\xi^n_{i+k},\, \xi^n_{i+k+1}+\Delta,\ldots,\xi^n_{i+k+l}+\Delta) - \xi^n_j| \\
& - \frac{1}{n}\sum_{j=i+1}^{i+k} |\operatorname{med}(\xi^n_{i+1},\ldots,\xi^n_{i+k}) - \xi^n_j| \\
& + \frac{1}{n}\sum_{j=i+k+1}^{i+k+l} |\operatorname{med}(\xi^n_{i+1},\ldots,\xi^n_{i+k},\, \xi^n_{i+k+1}+\Delta,\ldots,\xi^n_{i+k+l}+\Delta) - \Delta - \xi^n_j| \\
& - \frac{1}{n}\sum_{j=i+k+1}^{i+k+l} |\operatorname{med}(\xi^n_{i+k+1},\ldots,\xi^n_{i+k+l}) - \xi^n_j|.
\end{aligned}$$

We recall that in order to proof this Lemma we have to show that $\mathbb{P}$-almost surely for all sufficiently large $n\in\mathbb{N}$, all $l\geq \dfrac{n}{\ln n}$, $k\geq\beta\cdot l$ and $i\in\{0,\ldots,n-l-k\}$ it holds :

$$\tilde{H}^n_{\gamma_n}(\widehat{f}^1_n, f^n+\xi^n) - \tilde{H}^n_{\gamma_n}(\widehat{f}^2_n, f^n+\xi^n) > 0. \tag{5.15}$$

Due to the properties of the median:

$$\min_{a\in\mathbb{R}}\sum_{j=i+1}^{i+k} |a-\xi^n_j| = \sum_{j=i+1}^{i+k} |\operatorname{med}(\xi^n_{i+1},\ldots,\xi^n_{i+k}) - \xi^n_j|,$$

it is clear that

$$\frac{1}{n}\sum_{j=i+1}^{i+k}\Big(|\operatorname{med}(\xi^n_{i+1},\ldots,\xi^n_{i+k},\,\xi^n_{i+k+1}+\Delta,\ldots,\xi^n_{i+k+l}+\Delta)-\xi^n_j| - |\operatorname{med}(\xi^n_{i+1},\ldots,\xi^n_{i+k})-\xi^n_j|\Big)\geq 0.$$

Consequently, in order to prove (5.15) it is sufficient to show the validity of the next inequality:

$$\begin{aligned}
& \sum_{j=i+k+1}^{i+k+l} |\operatorname{med}(\xi^n_{i+1},\ldots,\xi^n_{i+k},\,\xi^n_{i+k+1}+\Delta,\ldots,\xi^n_{i+k+l}+\Delta) - \Delta - \xi^n_j| \\
- & \sum_{j=i+k+1}^{i+k+l} |\operatorname{med}(\xi_{i+k+1},\ldots,\xi^n_{i+k+l}) - \xi^n_j| > 3\cdot n\cdot\gamma_n \qquad \mathbb{P}-\text{a.s.}
\end{aligned}\tag{5.16}$$

Further on, according to Lemma 15, p.77, we fix χ such that $0<\chi<\Delta$ and which fulfills inequality (4.86), i.e.

$$\frac{1}{(1+\Delta-\chi)^{\alpha-1}} - \frac{\beta}{(1+\chi)^{\alpha-1}} > (1-\beta).$$

In Lemma 16 of the previous chapter (see p.78), we have proved that $\mathbb{P}$-almost surely for all sufficiently large $n\in\mathbb{N}$ ($n>10$), all $l\in\left\{\left\lceil\dfrac{n}{\ln n}\right\rceil,\ldots,\dfrac{n-1}{2}\right\}$, all $k\in\{\beta\cdot l,\ldots,n-l\}$ and all $i\in\{0,\ldots,n-l-k\}$ it holds:

$$\operatorname{med}(\xi^n_{i+1},\ldots,\xi^n_{i+k},\,\xi^n_{i+k+1}+\Delta,\ldots,\xi^n_{i+k+l}+\Delta)\leq\chi.$$

Therefore the following inequality is also $\mathbb{P}$-almost surely valid:

$$\operatorname{med}\left(\xi_{i+1}^n, \ldots, \xi_{i+k}^n, \xi_{i+k+1}^n + \Delta, \ldots, \xi_{i+k+l}^n + \Delta \right) - \Delta \leq \chi - \Delta < 0 \tag{5.17}$$

Moreover, from Corollary 4, Inequality (4.59), we know that:

$$-\frac{\sqrt{n}\,\ln n}{l} \leq \operatorname{med}\left(\xi_{i+k+1}^n, \ldots, \xi_{i+k+l}^n \right) \leq \frac{\sqrt{n}\,\ln n}{l} \qquad \mathbb{P}-\text{a.s.}$$

For all $l \geq \dfrac{n}{\ln n}$ is clear that $\dfrac{\sqrt{n}\,\ln n}{l} \leq \dfrac{\ln^2 n}{\sqrt{n}} \underset{n\to\infty}{\longrightarrow} 0$. This together with (5.17) yields that $\mathbb{P}$-almost surely for all sufficiently large n and all considered l, k and i we have:

$$\operatorname{med}\left(\xi_{i+1}^n, \ldots, \xi_{i+k}^n, \xi_{i+k+1}^n + \Delta, \ldots, \xi_{i+k+l}^n + \Delta \right) - \Delta \leq \chi - \Delta < \operatorname{med}\left(\xi_{i+k+1}^n, \ldots, \xi_{i+k+l}^n \right).$$

Let us define:

$$\Delta' := \Delta - \chi > 0. \tag{5.18}$$

Due to the fact that the function $h(a) = \sum\limits_{j=i+k+1}^{i+k+l} |\, a - \xi_j^n \,|$ as a linear combination of convex functions is also convex and the point of the minimum of $h(a)$ is $\operatorname{med}\left(\xi_{i+k+1}^n, \ldots, \xi_{i+k+l}^n \right)$, we conclude:

$$\begin{aligned}
& \sum_{j=i+k+1}^{i+k+l} |\operatorname{med}\left(\xi_{i+1}^n, \ldots, \xi_{i+k}^n, \xi_{i+k+1}^n + \Delta, \ldots, \xi_{i+k+l}^n + \Delta \right) - \Delta - \xi_j^n \,| \\
- & \sum_{j=i+k+1}^{i+k+l} |\operatorname{med}\left(\xi_{i+k+1}, \ldots, \xi_{i+k+l}^n \right) - \xi_j^n \,| \\
\geq & \sum_{j=i+k+1}^{i+k+l} |-\Delta' - \xi_j^n \,| - \sum_{j=i+k+1}^{i+k+l} |\xi_j^n \,| = \sum_{j=i+k+1}^{i+k+l} \left(|\Delta' + \xi_j^n \,| - |\xi_j^n \,| \right).
\end{aligned} \tag{5.19}$$

Moreover, from Lemma 17 (p. 80) we know that for any sequences $(\gamma_n)_{n\in\mathbb{N}}$, which satisfies Condition 3, for any finite $\Delta' > 0$, for any finite $\beta > 0$, $\mathbb{P}$-almost surely for all sufficiently large $n \in \mathbb{N}$ $(n > 10)$, all $l \in \left\{ \left\lceil \dfrac{n}{\ln n} \right\rceil, \ldots, \dfrac{n-1}{2} \right\}$, all $k \in \{\beta \cdot l, \ldots, n-l\}$ and all $i \in \{0, \ldots, n-l-k\}$ it holds:

$$\sum_{j=i+k+1}^{i+k+l} \left(|\Delta' + \xi_j^n \,| - |\xi_j^n \,| \right) > 3 \cdot n \cdot \gamma_n \qquad \mathbb{P}-\text{a.s.} \tag{5.20}$$

Combination of (5.20) with (5.19) yields the validity of (5.16).
Therefore the proof of Lemma 19 is complete.

□

In the following remark we give an interpretation of Lemma 19 in terms of the distances from a jump point of f^n or of the function $f \in S([0, 1))$ to the two neighboring jump points of $\widehat{f}_n$.

Prior to this, we recall the well-known definition.

Definition 12. *Let* $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ *and* $(b_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ *be two arbitrary sequences such that* $\lim_{n\to\infty} a_n = 0$ *and* $\lim_{n\to\infty} b_n = 0$ *Then one says:*

(i) $(a_n)_{n\in\mathbb{N}}$ *is of order of* $(b_n)_{n\in\mathbb{N}}$ *(symbol:* $a_n = \mathrm{O}(b_n)$*), if* $\limsup_{n\to\infty} \dfrac{a_n}{b_n} < \infty$.

(ii) $(a_n)_{n\in\mathbb{N}}$ *is of smaller order than* $(b_n)_{n\in\mathbb{N}}$ *(symbol:* $a_n = \mathrm{o}(b_n)$*), if* $\lim_{n\to\infty} \dfrac{a_n}{b_n} = 0$.

Remark 4. *(i) Let* $f \in S\big([0,1)\big)$, $f^n = \sum_{i=1}^{n} f_i^n \cdot \mathbf{1}_{\left[\frac{i-1}{n},\frac{i}{n}\right)}$, $n \in \mathbb{N}$, *is an approximation of* f *in* $S^n\big([0,1)\big)$ *with* $\lim_{n\to\infty} f^n = f$, $(\xi_i^n)_{1\le i\le n,\, n\in\mathbb{N}}$ *fulfills Condition 2 and* $\widehat{f}_n$ *for any* $n \in \mathbb{N}$ *is one possible minimizer of the functional* $\tilde{H}^n_{\gamma_n}(\cdot\,, f^n + \xi^n)$, *where* $(\gamma_n)_{n\in\mathbb{N}}$ *satisfy Condition 3. Then it will* $\mathbb{P}$*-almost surely not happen that one of the two distances from an arbitrary jump point of* f^n *to the two neighboring jump points of* $\widehat{f}_n$ *is larger than or equal to* $\dfrac{1}{\ln n}$ *and the other one is* $\dfrac{\beta}{\ln n}$ *for any finite* $\beta > 0$ *or larger.*

(ii) From Definition 11 we know that for any n *the distances between the neighboring jump points of* $f \in S\big([0,1)\big)$ *and the corresponding* f^n *is no larger than* $\dfrac{1}{n}$*. Taking this into account and also due to the obvious fact that* $\left(\dfrac{\beta}{\ln n} \pm \dfrac{1}{n}\right) = \mathrm{O}\left(\dfrac{1}{\ln n}\right)$ *for any* $\beta < \infty$, *we deduce that the statement (i) is also valid with regard to the distance between any jump point of the function* $f \in S\big([0,1)\big)$ *and the two neighboring jump points of* $\widehat{f}_n$.

(iii) In particular, if the function $f \in S([0,\,1])$ *or the function* $f^n \in S^n([0,\,1])$ *have, for the sake of argument, exactly one jump point* a_0 *then, according to Lemma 19,* $\mathbb{P}$*-almost surely for all sufficiently large* n *it holds that either the lengths of both of the two intervals between* a_0 *and its neighboring jump points of* $\widehat{f}_n$ *are of the order* $\mathrm{o}\left(\dfrac{1}{\ln n}\right)$ *or one of the intervals is larger than or equal to* $\dfrac{1}{\ln n}$ *and the other one is* $\mathrm{o}\left(\dfrac{1}{\ln n}\right)$*. Consequently, in the latter case, the ratio of the lengths of the shorter and the longer intervals vanishes as* n *tends to infinity.*

5.2 Consistency

Now we are ready to show that for a certain sequence of $(\gamma_n)_{n\in\mathbb{N}}$, a sequence of the minimizers $\widehat{f}_n$ of the functional $\tilde{H}^n_{\gamma_n}(\cdot\,, f^n + \xi^n)$ converges to the original function $f \in S\big([0,1)\big)$ or in other words, an arbitrary minimizer of $\tilde{H}^n_{\gamma_n}(\cdot\,, f^n + \xi^n)$ is a consistent estimate of the function $f \in S\big([0,1)\big)$.

It is convenient to split the proof of consequence into several different scenario, as dictated by the total number of jumps of f. Namely, we start by proving consistency of the minimizers for the trivial case, i.e. of a constant function f. After that the proof will be given for a step function f with the gradually increasing number of jumps.

Prior to this, we notice that Condition 3 on the sequence $(\gamma_n)_{n\in\mathbb{N}}$, i.e. $\ln n \cdot \gamma_n \underset{n\to\infty}{\longrightarrow} 0$, turned out to be insufficient to prove consistency of $\widehat{f}_n$. Therefore, below we formulate a set of additional conditions on $(\gamma_n)_{n\in\mathbb{N}}$.

Condition 4. *The sequence* $(\gamma_n)_{n\in\mathbb{N}} \subset (0,\infty)$ *satisfies*

$$\gamma_n \geq \frac{\ln^2 n}{\sqrt{n}} \,. \tag{5.21}$$

Condition 5. *The sequence* $(\gamma_n)_{n\in\mathbb{N}} \subset (0,\infty)$ *obeys*

$$\gamma_n \geq \frac{1}{\ln^2 n} \,. \tag{5.22}$$

It is obvious that Condition 5 implies Condition 4.

In the remainder of the thesis we use sequences $(\gamma_n)_{n\in\mathbb{N}} \subset (0,\infty)$, which fulfill at least Condition 4.
In case when the original function f is a constant function, Condition 4 is the sufficient condition on $(\gamma_n)_{n\in\mathbb{N}}$, which ensures consistency of $\widehat{f}_n \in \operatorname{argmin} \tilde{H}^n_{\gamma_n}(\cdot\,, f^n + \xi^n)$. With the help of Condition 5 we improve the rate of convergence of $\widehat{f}_n$ to f in this case.
In the more general case, i.e. when f is not necessarily constant, in order to prove consistency we require that Condition 4 and Condition 3 are satisfied. The latter allows us to use Lemma 19.
For the sake of simplicity we additionally introduce a new notation, which we will use in the subsequent sections:

Remark 5. *For all* $n \in \mathbb{N}$*, any arbitrary* i_1, i_2, j_1, j_2 *such that* $1 \leq i_1 \leq i_2 < j_1 \leq j_2 \leq n$ *and any real finite* c_1, c_2 *(even for* $c_1 = 0$ *or* $c_2 = 0$*) let*

$$\operatorname{med}\big(\,\xi^n_{i_1:i_2} + c_1,\, \xi^n_{j_1:j_2} + c_2\,\big) := \operatorname{med}\big(\,\xi^n_{i_1} + c_1, \dots, \xi^n_{i_2} + c_1,\, \xi^n_{j_1} + c_2, \dots, \xi^n_{j_2} + c_2\,\big). \tag{5.23}$$

In an analogous way, for all $n \in \mathbb{N}$*, any arbitrary* $i_1, i_2, j_1, j_2, s_1, s_2$ *such that* $1 \leq i_1 \leq i_2 < j_1 \leq j_2 < s_1 \leq s_2 \leq n$ *and any real finite* c_1, c_2, c_3 *(even for* $c_1 = 0$ *or* $c_2 = 0$ *or* $c_2 = 0$*) let*

$$\begin{aligned} &\operatorname{med}\big(\,\xi^n_{i_1:i_2} + c_1,\, \xi^n_{j_1:j_2} + c_2,\, \xi^n_{s_1:s_2} + c_3\big) \\ := \;&\operatorname{med}\big(\,\xi^n_{i_1} + c_1, \dots, \xi^n_{i_2} + c_1,\, \xi^n_{j_1} + c_2, \dots, \xi^n_{j_2} + c_2,\, \xi^n_{s_1} + c_3, \dots, \xi^n_{s_2} + c_3\,\big). \end{aligned} \tag{5.24}$$

5.2.1 Case of constant f

Theorem 5. *Let* $(\xi^n_i)_{1\leq i\leq n,\, n\in\mathbb{N}}$ *fulfill Condition 2,* $f \in L^1\big([0,1)\big)$ *be a constant function and for all* $n \in \mathbb{N}$ *let* $\widehat{f}_n \in \operatorname{argmin} \tilde{H}^n_{\gamma_n}(\cdot\,, f^n + \xi^n)$*. Then*

(i) for all sequences $(\gamma_n)_{n\in\mathbb{N}}$*, which fulfill Condition 4, it holds* $\mathbb{P}$*-almost surely:*

$$\|\,\widehat{f}_n - f\,\| = \mathrm{O}\Big(\frac{1}{\ln n}\Big) \tag{5.25}$$

(ii) for all sequences $(\gamma_n)_{n\in\mathbb{N}}$*, which satisfy Condition 5, it holds* $\mathbb{P}$*-almost surely:*

$$\|\,\widehat{f}_n - f\,\| = \mathrm{O}\Big(\frac{\ln^3 n}{\sqrt{n}}\Big) \tag{5.26}$$

Proof:
Let for all $x \in [0,1)$ and some constant $f \in \mathbb{R}$

$$f(x) = f.$$

According to Remark 2, p.20, $\widehat{f}_n \in \operatorname{argmin} \tilde{H}^n_{\gamma_n}(\cdot\,, f^n + \xi^n)$ for all $n \in \mathbb{N}$ can be represented by the following expression:

$$\widehat{f}_n(x) = \sum_{i=1}^{\#J(\widehat{f}_n)+1} \widehat{f}_{n,i} \cdot \mathbf{1}_{\widehat{I}^n_i}(x),$$

where

$J(\widehat{f}_n) := \left\{ \frac{\widehat{k}^n_1}{n}, \ldots, \frac{\widehat{k}^n_{\#J(\widehat{f}_n)}}{n} \right\}$, $\widehat{k}^n_1 \neq \ldots \neq \widehat{k}^n_{\#J(\widehat{f}_n)} \in \{1, \ldots n-1\}$, $\widehat{k}^n_0 := 0$, $\widehat{k}^n_{\#J(\widehat{f}_n)+1} := n$,
$\widehat{I}^n_i := \left[\frac{\widehat{k}^n_{i-1}}{n}, \frac{\widehat{k}^n_i}{n} \right)$, $P_{J(\widehat{f}_n)} := \left\{ \widehat{I}^n_1, \ldots, \widehat{I}^n_{\#J(\widehat{f}_n)+1} \right\}$ and
$\widehat{f}_{n,i} = \operatorname{med}_{\widehat{I}^n_i}(f^n + \xi^n) \in med_{\widehat{I}^n_i}(f^n + \xi^n) \quad \forall\, i \in \left\{1, \ldots, \#J(\widehat{f}_n) + 1\right\}$.

We note also that because in this case the function f is constant, then it holds

$$f^n + \xi^n = f + \xi^n$$

and for any $I \subseteq [0,\, 1)$

$$\operatorname{med}_I(f^n + \xi^n) = f + \operatorname{med}_I({\xi_i}^n).$$

In other words, for any $i \in \left\{1, \ldots, \#J(\widehat{f}_n) + 1\right\}$ and any $\widehat{f}_{n,i} = \operatorname{med}_{\widehat{I}^n_i}(f^n + \xi^n)$ we can write:

$$\widehat{f}_{n,i} = f + \operatorname{med}\left(\xi^n_{\widehat{k}^n_{i-1}+1}, \ldots, \xi^n_{\widehat{k}^n_i} \right) \tag{5.27}$$

Now we consider $\| \widehat{f}_n - f \|$:

$$\| \widehat{f}_n - f \| = \sum_{i=1}^{\#J(\widehat{f}_n)+1} l(\widehat{I}^n_i) \cdot | \widehat{f}_{n,i} - f | \underset{(5.27)}{=} \sum_{i=1}^{\#J(\widehat{f}_n)+1} \frac{\widehat{k}^n_i - \widehat{k}^n_{i-1}}{n} \cdot \left| \operatorname{med}\left(\xi^n_{\widehat{k}^n_{i-1}+1}, \ldots, \xi^n_{\widehat{k}^n_i} \right) \right|$$

$$\underset{(4.59)}{\overset{\mathbb{P}-\text{a.s.}}{\leq}} \sum_{i=1}^{\#J(\widehat{f}_n)+1} \frac{\widehat{k}^n_i - \widehat{k}^n_{i-1}}{n} \cdot \frac{\sqrt{n}\, \ln n}{\widehat{k}^n_i - \widehat{k}^n_{i-1}} = (\#J(\widehat{f}_n) + 1) \cdot \frac{\ln n}{\sqrt{n}} \underset{(5.6)}{\leq} \frac{\|\xi^n\|}{\gamma_n} \cdot \frac{\ln n}{\sqrt{n}} + \frac{\ln n}{\sqrt{n}} \tag{5.28}$$

If we substitute Condition 4 into (5.28) and additionally recall from Lemma 9 that $\mathbb{P}$-almost surely for all $n \in \mathbb{N}$ there exists some finite random constant $\widehat{C}$ such that $\|\xi^n\| \leq \widehat{C}$, we obtain:

$$\| \widehat{f}_n - f \| \leq \frac{\|\xi^n\|}{\ln n} + \frac{\ln n}{\sqrt{n}} \underset{n \to \infty}{\longrightarrow} 0.$$

Moreover,

$$\limsup_{n \to \infty} \frac{\| \widehat{f}_n - f \|}{(1/\ln n)} = \limsup_{n \to \infty} \left(\|\xi^n\| + \frac{\ln^2 n}{\sqrt{n}} \right) = \widehat{C} < \infty,$$

shows the validity of statement (i), i.e.

$$\| \widehat{f}_n - f \| = \mathrm{O}\left(\frac{1}{\ln n}\right) \qquad \mathbb{P}-\text{a.s.}$$

Similarly, in order to prove statement (ii) it is sufficient to substitute Condition 5 into (5.28):

$$\| \widehat{f}_n - f \| \leq \|\xi^n\| \cdot \frac{\ln^3 n}{\sqrt{n}} + \frac{\ln n}{\sqrt{n}} = \mathrm{O}\left(\frac{\ln^3 n}{\sqrt{n}}\right) \qquad \mathbb{P}-\text{a.s.}$$

This completes the proof of Theorem 5.

□

5.2.2 Case of a piecewise constant f

In this section we consider the relationships between a piecewise constant function f and the minimizer of $\tilde{H}^n_{\gamma_n}(\cdot\,, f^n + \xi^n)$. We start with the case, when f has exactly one jump.

Theorem 6. *Consider $f \in S\big([0,1)\big)$ with $\#J(f) = 1$. Moreover, let $(\xi^n_i)_{1\leq i\leq n,\, n\in\mathbb{N}}$ fulfill Condition 2 and*

$$\widehat{f}_n \in \operatorname{argmin} \tilde{H}^n_{\gamma_n}(\cdot\,, f^n + \xi^n).$$

Then for all $(\gamma_n)_{n\in\mathbb{N}}$, which fulfill Condition 3 and Condition 4, it holds $\mathbb{P}$-almost surely:

$$\| \widehat{f}_n - f \| \underset{n\to\infty}{\longrightarrow} 0, \tag{5.29}$$

in particular

$$\| \widehat{f}_n - f \| = \mathrm{O}\left(\frac{1}{\ln n}\right). \tag{5.30}$$

Proof:
Let f be a step function with exactly one jump. Denote by $a \in (0,1)$ the jump point of f and by $f_1,\ f_2 \in \mathbb{R}$ $(f_1 \neq f_2)$ the two different values of the function f, i.e. $J(f) = \{a\}$ and

$$f(x) = f_1 \cdot \mathbf{1}_{[0,\,a)}(x) + f_2 \cdot \mathbf{1}_{[a,\,1)}(x). \tag{5.31}$$

In this theorem we use again the representation of $\widehat{f}_n$ from Remark 2, implying

$$\widehat{f}_n(x) = \sum_{i=1}^{\#J(\widehat{f}_n)+1} \widehat{f}_{n,i} \cdot \mathbf{1}_{\widehat{I}^n_i}(x),$$

where

$J(\widehat{f}_n) := \Big\{\frac{\widehat{k}^n_1}{n}, \ldots, \frac{\widehat{k}^n_{\#J(\widehat{f}_n)}}{n}\Big\}$, $\widehat{k}^n_1 \neq \ldots \neq \widehat{k}^n_{\#J(\widehat{f}_n)} \in \{1, \ldots n-1\}$, $\widehat{k}^n_0 := 0$, $\widehat{k}^n_{\#J(\widehat{f}_n)+1} := n$,
$\widehat{I}^n_i := \Big[\frac{\widehat{k}^n_{i-1}}{n}, \frac{\widehat{k}^n_i}{n}\Big)$, $P_{J(\widehat{f}_n)} := \big\{\widehat{I}^n_1, \ldots, \widehat{I}^n_{\#J(\widehat{f}_n)+1}\big\}$ and
$\widehat{f}_{n,i} = \operatorname{med}_{\widehat{I}^n_i}(f^n + \xi^n) \in med_{\widehat{I}^n_i}(f^n + \xi^n) \quad \forall\, i \in \big\{1, \ldots, \#J(\widehat{f}_n) + 1\big\}$.

We know that the jump point a of the function f belongs to $(0,\,1)$, implying that there is a number $j \in \big\{1, \dots, \#J(\widehat{f}_n)+1\big\}$ such that $a \in \Big[\frac{\widehat{k}^n_{j-1}}{n},\, \frac{\widehat{k}^n_j}{n}\Big) = \widehat{I}^n_j$.
We fix this particular j and introduce the following notations:

$$\widehat{I}^n_{j-1,\,a} \;:=\; \Big[\frac{\widehat{k}^n_{j-1}}{n},\; a\Big) \tag{5.32}$$

$$\widehat{I}^n_{a,\,j} \;:=\; \Big[a,\; \frac{\widehat{k}^n_{j}}{n}\Big) \tag{5.33}$$

Moreover, according to (5.31) and Definition 11 of the function f^n in case of $f \in S\big([0,1)\big)$, we known that for all $n \ge \Big\lceil \frac{1}{\min\{a,\,1-a\}} \Big\rceil =: N^0_f$ the function f^n, similarly to f, has exactly one jump, which coordinate we denote by $\frac{\widehat{k}^n_{j-1}+m^n}{n}$, and can be represented in the form

$$f^n(x) = f_1 \cdot \mathbf{1}_{\big[0,\,\frac{\widehat{k}^n_{j-1}}{n}+\frac{m^n}{n}\big)}(x) + f_2 \cdot \mathbf{1}_{\big[\frac{\widehat{k}^n_{j-1}}{n}+\frac{m^n}{n},\,1\big)}(x). \tag{5.34}$$

We also remark that for all $n \ge N^0_f$ the number $m^n \in \{1,\dots,n-1\}$ is in the following relationship with the point a:

$$0 \le \Big(\frac{\widehat{k}^n_{j-1}+m^n}{n} - a\Big) \le \frac{1}{n}\,. \tag{5.35}$$

Moreover, we assume that for all $n \ge N^0_f$ the length of the interval $\widehat{I}^n_j$ is represented as follows:

$$l(\widehat{I}^n_j) = \frac{1}{n}(\widehat{k}^n_j - \widehat{k}^n_{j-1}) =: \frac{m^n}{n} + \frac{\widehat{l}^n}{n}\,, \qquad \widehat{l}^n \in \{1,\dots,n-m^n\}, \tag{5.36}$$

this means that

$$\widehat{k}^n_j = \widehat{k}^n_{j-1} + m^n + \widehat{l}^n. \tag{5.37}$$

Therefore, for all $n \ge N^0_f$ the real number $\widehat{f}_{n,j} = \mathrm{med}_{\widehat{I}^n_j}(f^n+\xi^n)$ can be written in the following form:

$$\widehat{f}_{n,j} = \mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n} + f_1\,,\; \xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} + f_2\,\big), \tag{5.38}$$

where $\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n} + f_1\,,\; \xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} + f_2\,\big)$ is defined analogue to (5.23).

We note also that $\widehat{f}_{n,i} = \mathrm{med}_{\widehat{I}^n_i}(f^n+\xi^n)$ for all $n \in \mathbb{N}$, $i \in \big\{1,\dots,\#J(\widehat{f}_n)+1\big\}$ and for all $n \ge N^0_f$ the function f^n is constant on every interval $\widehat{I}^n_i$, $i \ne j$. Thus for all $n \ge N^0_f$ the constants $\widehat{f}_{n,i}$, $i \ne j$, can be represented as follows:

$$\widehat{f}_{n,i} \;=\; f_1 + \mathrm{med}\,(\,\xi^n_{\widehat{k}^n_{i-1}+1},\dots,\xi^n_{\widehat{k}^n_i}\,),\;\; i \in \{1,\dots,j-1\} \tag{5.39}$$

$$\widehat{f}_{n,i} \;=\; f_2 + \mathrm{med}\,(\,\xi^n_{\widehat{k}^n_{i-1}+1},\dots,\xi^n_{\widehat{k}^n_i}\,),\;\; i \in \{j+1,\dots,\#J(\widehat{f}_n)+1\} \tag{5.40}$$

With (5.38), (5.39), (5.40) in mind, we consider $\| \widehat{f}_n - f \|$:

$$\| \widehat{f}_n - f \| = \int\limits_{[0,1)} \Big| \sum_{i=1}^{\#J(\widehat{f}_n)+1} \widehat{f}_{n,i} \cdot \mathbf{1}_{\widehat{I}_i^n}(x) - f_1 \cdot \mathbf{1}_{[0,a)}(x) - f_2 \cdot \mathbf{1}_{[a,1)}(x) \Big| \, dx$$

$$= \sum_{i=1}^{\#J(\widehat{f}_n)+1} \int\limits_{\widehat{I}_i^n} \big| \widehat{f}_{n,i} - f_1 \cdot \mathbf{1}_{[0,a)}(x) - f_2 \cdot \mathbf{1}_{[a,1)}(x) \big| \, dx$$

$$\underset{(5.32),(5.33)}{=} \sum_{i=1}^{j-1} \int\limits_{\widehat{I}_i^n} | \widehat{f}_{n,i} - f_1 | \, dx + \sum_{i=j+1}^{\#J(\widehat{f}_n)+1} \int\limits_{\widehat{I}_i^n} | \widehat{f}_{n,i} - f_2 | \, dx$$

$$+ \int\limits_{\widehat{I}_{j-1,a}^n} | \widehat{f}_{n,j} - f_1 | \, dx + \int\limits_{\widehat{I}_{a,j}^n} | \widehat{f}_{n,j} - f_2 | \, dx$$

$$= \sum_{i=1}^{j-1} l(\widehat{I}_i^n) \cdot | \operatorname{med}(\xi^n_{\widehat{k}_{i-1}^n+1}, \ldots, \xi^n_{\widehat{k}_i^n})| + \sum_{i=j+1}^{\#J(\widehat{f}_n)+1} l(\widehat{I}_i^n) \cdot | \operatorname{med}(\xi^n_{\widehat{k}_{i-1}^n+1}, \ldots, \xi^n_{\widehat{k}_i^n})|$$

$$+ \; l(\widehat{I}_{j-1,a}^n) \cdot \big| \operatorname{med}\big(\xi^n_{\widehat{k}_{j-1}^n+1 : \widehat{k}_{j-1}^n+m^n} + f_1 \, , \, \xi^n_{\widehat{k}_{j-1}^n+m^n+1 : \widehat{k}_{j-1}^n+m^n+\widehat{l}^n} + f_2 \big) - f_1 \big|$$

$$+ \; l(\widehat{I}_{a,j}^n) \cdot \big| \operatorname{med}\big(\xi^n_{\widehat{k}_{j-1}^n+1 : \widehat{k}_{j-1}^n+m^n} + f_1 \, , \, \xi^n_{\widehat{k}_{j-1}^n+m^n+1 : \widehat{k}_{j-1}^n+m^n+\widehat{l}^n} + f_2 \big) - f_2 \big|$$

$$\underset{(4.59)}{\overset{\mathbb{P}-\text{a.s.}}{\leq}} \sum_{i=1,\, i \neq j}^{\#J(\widehat{f}_n)+1} \frac{\widehat{k}_i^n - \widehat{k}_{i-1}^n}{n} \cdot \frac{\sqrt{n} \, \ln n}{\widehat{k}_i^n - \widehat{k}_{i-1}^n}$$

$$+ \; l(\widehat{I}_{j-1,a}^n) \cdot \big| \operatorname{med}\big(\xi^n_{\widehat{k}_{j-1}^n+1 : \widehat{k}_{j-1}^n+m^n} + f_1 \, , \, \xi^n_{\widehat{k}_{j-1}^n+m^n+1 : \widehat{k}_{j-1}^n+m^n+\widehat{l}^n} + f_2 \big) - f_1 \big|$$

$$+ \; l(\widehat{I}_{a,j}^n) \cdot \big| \operatorname{med}\big(\xi^n_{\widehat{k}_{j-1}^n+1 : \widehat{k}_{j-1}^n+m^n} + f_1 \, , \, \xi^n_{\widehat{k}_{j-1}^n+m^n+1 : \widehat{k}_{j-1}^n+m^n+\widehat{l}^n} + f_2 \big) - f_2 \big|$$

$$\underset{(5.6)}{\leq} \Big(\frac{\|\xi^n\|}{\gamma_n} + 1 \Big) \cdot \frac{\ln n}{\sqrt{n}}$$

$$+ \; l(\widehat{I}_{j-1,a}^n) \cdot \big| \operatorname{med}\big(\xi^n_{\widehat{k}_{j-1}^n+1 : \widehat{k}_{j-1}^n+m^n} + (f_1 - f_1) \, , \, \xi^n_{\widehat{k}_{j-1}^n+m^n+1 : \widehat{k}_{j-1}^n+m^n+\widehat{l}^n} + (f_2 - f_1) \big) \big|$$

$$+ \; l(\widehat{I}_{a,j}^n) \cdot \big| \operatorname{med}\big(\xi^n_{\widehat{k}_{j-1}^n+1 : \widehat{k}_{j-1}^n+m^n} + (f_1 - f_2) \, , \, \xi^n_{\widehat{k}_{j-1}^n+m^n+1 : \widehat{k}_{j-1}^n+m^n+\widehat{l}^n} + (f_2 - f_2) \big) \big|$$

$$\underset{(5.21)}{\leq} \frac{\|\xi^n\|}{\ln n} + \frac{\ln n}{\sqrt{n}}$$

$$+ \; l(\widehat{I}_{j-1,a}^n) \cdot \big| \operatorname{med}\big(\xi^n_{\widehat{k}_{j-1}^n+1 : \widehat{k}_{j-1}^n+m^n} \, , \, \xi^n_{\widehat{k}_{j-1}^n+m^n+1 : \widehat{k}_{j-1}^n+m^n+\widehat{l}^n} + (f_2 - f_1) \big) \big|$$

$$+ \; l(\widehat{I}_{a,j}^n) \cdot \big| \operatorname{med}\big(\xi^n_{\widehat{k}_{j-1}^n+1 : \widehat{k}_{j-1}^n+m^n} + (f_1 - f_2) \, , \, \xi^n_{\widehat{k}_{j-1}^n+m^n+1 : \widehat{k}_{j-1}^n+m^n+\widehat{l}^n} \big) \big|$$

If we define

$$\Delta := f_2 - f_1 \neq 0 \tag{5.41}$$

then for all $n \geq N_f^0$ we obtain :

$$\begin{aligned}
&\| \widehat{f}_n - f \| \\
\leq\ & \Big(\frac{\|\xi^n\|}{\ln n} + \frac{\ln n}{\sqrt{n}}\Big) + l(\widehat{I}^n_{j-1,a}) \cdot \big|\, \mathrm{med}\,\big(\, \xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n}\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} + \Delta \,\big)\,\big| \\
+\ & l(\widehat{I}^n_{a,j}) \cdot \big|\, \mathrm{med}\,\big(\, \xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n} - \Delta\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} \,\big)\,\big|. \qquad (5.42)
\end{aligned}$$

Due to Lemma 9 the function ξ^n is $\mathbb{P}$-almost surely bounded. Similar to the proof of Theorem 5, the first component of the last inequality converges $\mathbb{P}$-almost surely to 0 with the convergence rate $\mathrm{O}\Big(\frac{1}{\ln n}\Big)$, as n goes to infinity.

Now we investigate the behaviour in the limit $n \to \infty$ of the two other components of (5.42).
Numerical estimates of these two terms depend on the lengths of the respective intervals $\widehat{I}^n_{j-1,a} = \Big[\frac{\widehat{k}^n_{j-1}}{n},\ a\Big)$ and $\widehat{I}^n_{a,j} = \Big[a,\ \frac{\widehat{k}^n_j}{n}\Big)$. More precisely, in order to determine these estimates one needs to know how the lengths of the respective intervals compare with $\frac{1}{\ln n}$ for infinitely large n.
As we have already noticed in Remark 4 (iii), which follows from Lemma 19, it is $\mathbb{P}$-almost surely impossible that for sufficiently large n the lengths of one of the two intervals $\widehat{I}^n_{j-1,a}$ and $\widehat{I}^n_{a,j}$ is at least $\frac{1}{\ln n}$ and the length of the other one is of the same order or larger. Moreover, for all $n \geq N^0_f$ the same statement is valid also for the intervals $\Big[\frac{\widehat{k}^n_{j-1}}{n},\ \frac{\widehat{k}^n_{j-1}+m^n}{n}\Big)$ and $\Big[\frac{\widehat{k}^n_{j-1}+m^n}{n},\ \frac{\widehat{k}^n_{j-1}+m^n+\widehat{l}^n}{n}\Big)$.

Consequently, there exists a natural number $N_f \geq N^0_f$ such that for all $n \geq N_f$ it is sufficient to consider the two following cases:

1) $l(\widehat{I}^n_{j-1,a}) < \frac{1}{\ln n},\quad \frac{m^n}{n} < \frac{1}{\ln n}$ and $l(\widehat{I}^n_{a,j}) < \frac{1}{\ln n},\quad \frac{\widehat{l}^n}{n} < \frac{1}{\ln n}$,

2) $l(\widehat{I}^n_{j-1,a}) \geq \frac{1}{\ln n},\quad \frac{m^n}{n} \geq \frac{1}{\ln n}$ and $l(\widehat{I}^n_{a,j}) = \mathrm{o}\Big(\frac{1}{\ln n}\Big),\quad \frac{\widehat{l}^n}{n} = \mathrm{o}\Big(\frac{1}{\ln n}\Big)$

The case $l(\widehat{I}^n_{j-1,a}) = \mathrm{o}\Big(\frac{1}{\ln n}\Big)$, and $l(\widehat{I}^n_{a,j}) \geq \frac{1}{\ln n}$ is similar to the case **2)**.

Case **1)** Let $l(\widehat{I}^n_{j-1,a}) < \frac{1}{\ln n}$ and $l(\widehat{I}^n_{a,j}) < \frac{1}{\ln n},\ n \geq N_f$.

The goal is to show that it holds $\mathbb{P}$-almost surely:

$$l(\widehat{I}^n_{j-1,a}) \cdot |\, \mathrm{med}\,(\, \xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n}\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} + \Delta \,)\,| \underset{n\to\infty}{\longrightarrow} 0 \qquad (5.43)$$

$$\text{and}\quad l(\widehat{I}^n_{a,j}) \cdot |\, \mathrm{med}\,(\, \xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n} - \Delta\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} \,)\,| \underset{n\to\infty}{\longrightarrow} 0 \qquad (5.44)$$

Application of Lemma 25 (see Appendix A, p.123) yields:

$$\begin{aligned} & \left| \operatorname{med}\left(\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n}\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} + \Delta \right) \right| \\ \leq \ & \left| \operatorname{med}\left(\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} \right) \right| + |\Delta| \end{aligned} \tag{5.45}$$

and

$$\begin{aligned} & \left| \operatorname{med}\left(\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n} - \Delta\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} \right) \right| \\ \leq \ & \left| \operatorname{med}\left(\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} \right) \right| + |\Delta| \end{aligned} \tag{5.46}$$

According to notations (5.23), (5.37) and to Corollary 4, p.68, $\mathbb{P}$-almost surely for all sufficiently large n we have:

$$|\operatorname{med}\left(\xi^n_{\widehat{k}^n_{j-1}+1}, \ldots, \xi^n_{\widehat{k}^n_{j-1}+m^n+\widehat{l}} \right)| = |\operatorname{med}\left(\xi^n_{\widehat{k}^n_{j-1}+1}, \ldots, \xi^n_{\widehat{k}^n_j} \right)| \underset{(4.59)}{\leq} \frac{\ln n}{\sqrt{n}} \cdot \frac{1}{l(\widehat{I}^n_j)} \tag{5.47}$$

Taking (5.45), (5.46), (5.47) into account and due to the facts that $l(\widehat{I}^n_{j-1,a}) < \dfrac{1}{\ln n}$ and $l(\widehat{I}^n_{a,j}) < \dfrac{1}{\ln n}$, we obtain that it holds $\mathbb{P}$-almost surely:

$$\begin{aligned} & l(\widehat{I}^n_{j-1,a}) \cdot |\operatorname{med}\left(\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n}\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} + \Delta \right)| \\ \leq \ & \frac{l(\widehat{I}^n_{j-1,a})}{l(\widehat{I}^n_j)} \cdot \frac{\ln n}{\sqrt{n}} + l(\widehat{I}^n_{j-1,a}) \cdot |\Delta| \leq \frac{\ln n}{\sqrt{n}} + \frac{|\Delta|}{\ln n} = \mathrm{O}\left(\frac{1}{\ln n} \right) \underset{n\to\infty}{\longrightarrow} 0 \end{aligned}$$

and similarly

$$\begin{aligned} & l(\widehat{I}^n_{a,j}) \cdot |\operatorname{med}\left(\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n} - \Delta\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} \right)| \\ \leq \ & \frac{l(\widehat{I}^n_{a,j})}{l(\widehat{I}^n_j)} \cdot \frac{\ln n}{\sqrt{n}} + l(\widehat{I}^n_{a,j}) \cdot |\Delta| \leq \frac{\ln n}{\sqrt{n}} + \frac{|\Delta|}{\ln n} = \mathrm{O}\left(\frac{1}{\ln n} \right) \underset{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

Consequently we have shown the validity of (5.43) and (5.44). Hence, according to (5.42), in the Case **1)** we have proved the fact:

$$\| \widehat{f}_n - f \| = \mathrm{O}\left(\frac{1}{\ln n} \right) \quad \mathbb{P} - \text{a.s.}$$

Case **2)** Suppose that for all $n \geq N_f$

$$l(\widehat{I}^n_{j-1,a}) \geq \frac{1}{\ln n},\ \frac{m^n}{n} \geq \frac{1}{\ln n} \ \text{ and } \ l(\widehat{I}^n_{a,j}) = \mathrm{o}\left(\frac{1}{\ln n} \right),\ \frac{\widehat{l}^n}{n} = \mathrm{o}\left(\frac{1}{\ln n} \right) \tag{5.48}$$

This implies:

$$\lim_{n\to\infty} \frac{l(\widehat{I}^n_{a,j})}{l(\widehat{I}^n_{j-1,a})} = 0,$$

as well as,

$$m^n \geq \frac{n}{\ln n} \quad (n \geq N_f) \quad \text{and} \quad \lim_{n\to\infty} \frac{\widehat{l}^n}{m^n} = 0. \tag{5.49}$$

In this case it becomes immediately clear that due to the same arguments as in Case **1)**, the number $l(\widehat{I}^n_{a,j}) \cdot |\, \mathrm{med}\,(\, \xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n} - \Delta\,,\, \xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n}\,)\,|$ converges $\mathbb{P}$-almost surely to 0, as n approaches ∞. Thereby the rate of convergence is given by $\mathrm{O}\left(\frac{1}{\ln n}\right)$.

Hence we have to show here that even if $l(\widehat{I}^n_{j-1,a}) \geq \frac{1}{\ln n}$, $m^n \geq \frac{n}{\ln n}$, but $\lim_{n\to\infty} \frac{\widehat{l}^n}{m^n} = 0$, the following result it still valid:

$$l(\widehat{I}^n_{j-1,a}) \cdot |\, \mathrm{med}\,(\, \xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n}\,,\, \xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} + \Delta\,)\,| \underset{n\to\infty}{\longrightarrow} 0 \quad \mathbb{P}-\text{a.s.} \tag{5.50}$$

Conditions (5.49) allow us to use statement (ii) of Lemma 13, p.74 in order to estimate $|\, \mathrm{med}\,(\, \xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n}\,,\, \xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} + \Delta\,)\,|$. This statement under the use of the notations

$$\widehat{\nu}^n := \frac{\widehat{l}^n}{m^n + \widehat{l}^n} \tag{5.51}$$

for sufficiently large n yields the following result:

$$\begin{aligned} &|\, \mathrm{med}\,(\, \xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n}\,,\, \xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} + \Delta\,)\,| \\ \leq\ & \max\left\{ -\left(z^-_{\widehat{\nu}^n} - \frac{\sqrt{n}\,\ln n}{m^n}\right),\ z^+_{\widehat{\nu}^n} + \frac{\sqrt{n}\,\ln n}{m^n} \right\} \quad \mathbb{P}-\text{a.s.}, \end{aligned} \tag{5.52}$$

where $z^-_{\widehat{\nu}^n} = \widetilde{\xi}_{\frac{1}{2}-\widehat{\nu}^n}$ and $z^+_{\widehat{\nu}^n} = \widetilde{\xi}_{\frac{1}{2}+\widehat{\nu}^n}$ are the appropriate quantiles of a random variable ξ, which is identically distributed with ξ^n_i for all n and all $i \in \{1,\ldots,n\}$.

Next we consider $z^-_{\widehat{\nu}^n}$ and $z^+_{\widehat{\nu}^n}$ taking into account that according to (5.49) and (5.51) it holds:

$$\widehat{\nu}^n \underset{n\to\infty}{\longrightarrow} 0. \tag{5.53}$$

Due to (A.49) and (A.50) (p.130), we have:

$$\begin{aligned} z^-_{\widehat{\nu}^n} &= 1 - \frac{1}{(2\,(\frac{1}{2} - \widehat{\nu}^n)\,)^{\frac{1}{\alpha-1}}} = 1 - \frac{1}{(1 - 2\,\widehat{\nu}^n)^{\frac{1}{\alpha-1}}} \\ z^+_{\widehat{\nu}^n} &= \frac{1}{(2\,(1 - (\frac{1}{2} + \widehat{\nu}^n)))^{\frac{1}{\alpha-1}}} - 1 = \frac{1}{(1 - 2\,\widehat{\nu}^n)^{\frac{1}{\alpha-1}}} - 1 \end{aligned}$$

The term $\frac{1}{(1 - 2\,\widehat{\nu}^n)^{\frac{1}{\alpha-1}}}$, due to (5.53), can be rewritten by Taylor's expansion:

$$\frac{1}{(1 - 2\,\widehat{\nu}^n)^{\frac{1}{\alpha-1}}} = 1 + \frac{2}{\alpha - 1} \cdot \widehat{\nu}^n + \mathrm{O}\left((\widehat{\nu}^n)^2\right)$$

From this it follows:

$$-z^-_{\widehat{\nu}^n} = z^+_{\widehat{\nu}^n} = \frac{2}{\alpha-1} \cdot \widehat{\nu}^n + \mathrm{O}\left((\widehat{\nu}^n)^2\right) = \frac{2}{\alpha-1} \cdot \frac{\widehat{l}^n}{m^n + \widehat{l}^n} + \mathrm{O}\left(\left(\frac{\widehat{l}^n}{m^n + \widehat{l}^n}\right)^2\right). \tag{5.54}$$

Consequently, after substituting (5.54) into (5.52) we obtain that the following relationships are $\mathbb{P}$-almost surely valid :

$$\begin{aligned} & l(\widehat{I}^n_{j-1,a}) \cdot |\, \mathrm{med}\,(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n}\,,\,\xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n} + \Delta\,)\,| \\ \leq\quad & l(\widehat{I}^n_j) \cdot \Big(\frac{\sqrt{n}\,\ln n}{m^n} + \frac{2}{\alpha-1} \cdot \frac{\widehat{l}^n}{m^n+\widehat{l}^n} + \mathrm{O}\Big(\Big(\frac{\widehat{l}^n}{m^n+\widehat{l}^n}\Big)^2\Big)\Big) \\ =\quad & \frac{m^n+\widehat{l}^n}{n} \cdot \Big(\frac{\sqrt{n}\,\ln n}{m^n} + \frac{2}{\alpha-1} \cdot \frac{\widehat{l}^n}{m^n+\widehat{l}^n} + \mathrm{O}\Big(\Big(\frac{\widehat{l}^n}{m^n+\widehat{l}^n}\Big)^2\Big)\Big) \\ =\quad & \frac{\ln n}{\sqrt{n}} + \frac{\ln n}{\sqrt{n}} \cdot \frac{\widehat{l}^n}{m^n} + \frac{2}{\alpha-1} \cdot \frac{\widehat{l}^n}{n} + \mathrm{O}\Big(\frac{\widehat{l}^n}{n} \cdot \frac{\widehat{l}^n}{m^n+\widehat{l}^n}\Big) \\ \underset{(5.48),(5.49)}{<}\quad & \frac{\ln n}{\sqrt{n}} + \frac{\ln n}{\sqrt{n}} + \frac{2}{\alpha-1} \cdot \frac{1}{\ln n} + \mathrm{O}\Big(\frac{1}{\ln n}\Big) = \mathrm{O}\Big(\frac{1}{\ln n}\Big) \underset{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

Therefore we have proved the validity of (5.50), which means that in Case **2)** it holds $\|\,\widehat{f}_n - f\,\| \underset{n\to\infty}{\longrightarrow} 0$ $\mathbb{P}$-almost surely and the rate of convergence remains unchanged because of the last relationship. Summarizing, we deduce:

$$\|\,\widehat{f}_n - f\,\| = \mathrm{O}\Big(\frac{1}{\ln n}\Big) \quad \mathbb{P}-\text{a.s.}$$

Thus, the proof of Theorem 6 is complete.

□

Remark 6. *Notice that because in the equation (5.42) at least one of the two terms, which contain medians, in all considered cases tends to zero as* $\mathrm{O}\Big(\frac{1}{\ln n}\Big)$ *and because both these terms do not depend on* γ_n*, we can conclude that by changing solely* γ_n *one cannot improve the rate of convergence, as it could be done in case of the constant* f*.*
Furthermore, from the proof of Theorem 6 we can also see that the jumps of estimators $\widehat{f}^n$ *approach the jumps of the original function* f *with convergence rate of at least* $\mathrm{O}\Big(\frac{1}{\ln n}\Big)$*. This rate does not seem optimal. In* L^2 *case the corresponding rate is given by* $\mathrm{O}\Big(\frac{\ln n}{n}\Big)$ *[8].*

By Theorems 5 and 6 we have proved consistence of the considered estimators for a constant function and for a piecewise constant function with exactly one jump. Additionally, we have found rate of convergence in these two cases. Now we prove the validity of the same results in case of a piecewise constant function with two jumps.

Theorem 7. *Let* $f \in S\big([0,1)\big)$ *and* $\#J(f) = 2$*,* $(\xi^n_i)_{1\leq i\leq n,\, n\in\mathbb{N}}$ *fulfill Condition 2 and*

$$\widehat{f}_n \in \mathrm{argmin}\ \tilde{H}^n_{\gamma_n}(\,\cdot\,, f^n + \xi^n).$$

Then for all $(\gamma_n)_{n\in\mathbb{N}}$*, which satisfy Condition 3 and Condition 4, it holds* $\mathbb{P}$*-almost surely:*

$$\|\,\widehat{f}_n - f\,\| = \mathrm{O}\Big(\frac{1}{\ln n}\Big). \tag{5.55}$$

Proof:
The condition $f \in S\big([0,1)\big)$ and $\#J(f) = 2$ implies that there exist $a_1,\, a_2 \in (0,1)$, $a_1 \neq a_2$, representing the jump coordinates of f, i.e. $J(f) := \{a_1, a_2\}$, and three real finite constants $f_1,\ f_2$ and f_3 with $f_1 \neq f_2,\ f_2 \neq f_3$ such that

$$f(x) = f_1 \cdot \mathbf{1}_{[0,\,a_1)}(x) + f_2 \cdot \mathbf{1}_{[a_1,\,a_2)}(x) + f_3 \cdot \mathbf{1}_{[a_2,\,1)}(x). \tag{5.56}$$

The minimizer $\widehat{f}_n$ of $\tilde{H}^n_{\gamma_n}(\cdot\,, f^n + \xi^n)$ for any $n \in \mathbb{N}$ is given by a similar expression as above, namely:

$$\widehat{f}_n(x) = \sum_{i=1}^{\#J(\widehat{f}_n)+1} \widehat{f}_{n,i} \cdot \mathbf{1}_{\widehat{I}^n_i}(x),$$

where
$J(\widehat{f}_n) := \Big\{\frac{\widehat{k}^n_1}{n}, \dots, \frac{\widehat{k}^n_{\#J(\widehat{f}_n)}}{n}\Big\}$, $\widehat{k}^n_1 \neq \ldots \neq \widehat{k}^n_{\#J(\widehat{f}_n)} \in \{1, \dots n-1\}$, $\widehat{k}^n_0 := 0$, $\widehat{k}^n_{\#J(\widehat{f}_n)+1} := n$,
$\widehat{I}^n_i := \Big[\frac{\widehat{k}^n_{i-1}}{n}, \frac{\widehat{k}^n_i}{n}\Big)$, $P_{J(\widehat{f}_n)} := \big\{\widehat{I}^n_1, \dots, \widehat{I}^n_{\#J(\widehat{f}_n)+1}\big\}$ and
$\widehat{f}_{n,i} = \mathrm{med}_{\widehat{I}^n_i}(f^n + \xi^n) \in med_{\widehat{I}^n_i}(f^n + \xi^n) \quad \forall\, i \in \big\{1, \dots, \#J(\widehat{f}_n) + 1\big\}$.

Let us now discuss the possible positions of the jump coordinates a_1 and a_2 of the function f.

Case 1. For any $n \in \mathbb{N}$ suppose that there exist two numbers $j \in \big\{1, \dots, \#J(\widehat{f}_n)\big\}$ and $s \in \big\{j+1, \dots, \#J(\widehat{f}_n)+1\big\}$ such that $a_1 \in \Big[\frac{\widehat{k}^n_{j-1}}{n}, \frac{\widehat{k}^n_j}{n}\Big) = \widehat{I}^n_j$ and $a_2 \in \Big[\frac{\widehat{k}^n_{s-1}}{n}, \frac{\widehat{k}^n_s}{n}\Big) = \widehat{I}^n_s$.

In this case, we will operate analogous to the proof of Theorem 5.
Let us introduce the following notations:

$$\begin{aligned}
\widehat{I}^n_{j-1,\,a_1} &:= \Big[\frac{\widehat{k}^n_{j-1}}{n},\ a_1\Big),\\
\widehat{I}^n_{a_1,\,j} &:= \Big[a_1,\ \frac{\widehat{k}^n_j}{n}\Big),\\
\widehat{I}^n_{s-1,\,a_2} &:= \Big[\frac{\widehat{k}^n_{s-1}}{n},\ a_2\Big),\\
\widehat{I}^n_{s-1,\,a_2} &:= \Big[a_2,\ \frac{\widehat{k}^n_s}{n}\Big).
\end{aligned}$$

Using Definition 11 we determine the functions f^n, $n \in \mathbb{N}$, which correspond to the function f defined by (5.56). For all $n \in \mathbb{N}$ the number of jumps of f^n is less or equal to two. If we define:

$$N'_f := \Big\lceil \frac{1}{\min\{a_1,\ a_2 - a_1,\ 1 - a_2\}} \Big\rceil, \tag{5.57}$$

then for any $n \geq N'_f$ the function f^n has exactly two jumps.

The coordinates of these jumps in Case **1** will be denoted by $\dfrac{\widehat{k}^n_{j-1} + m^n_1}{n}$ and $\dfrac{\widehat{k}^n_{s-1} + m^n_2}{n}$

for some $m_1^n \in \{1, \dots, n-1\}$ and $m_2^n \in \{1, \dots, n - m_1^n\}$. It is known that these points satisfy the following conditions:

$$0 \le \frac{\widehat{k}_{j-1}^n + m_1^n}{n} - a_1 \le \frac{1}{n},$$
$$0 \le \frac{\widehat{k}_{s-1}^n + m_2^n}{n} - a_2 \le \frac{1}{n}.$$

According to this, it is clear that for all $n \ge N_f'$ in Case **1** the function f^n can be represented as follows:

$$f^n(x) = f_1 \cdot \mathbf{1}_{\left[0,\, \frac{\widehat{k}_{j-1}^n}{n} + \frac{m_1^n}{n}\right)}(x) + f_2 \cdot \mathbf{1}_{\left[\frac{\widehat{k}_{j-1}^n}{n} + \frac{m_1^n}{n},\, \frac{\widehat{k}_{s-1}^n}{n} + \frac{m_2^n}{n}\right)}(x) + f_2 \cdot \mathbf{1}_{\left[\frac{\widehat{k}_{s-1}^n}{n} + \frac{m_2^n}{n},\, 1\right)}(x).$$

Moreover, we assume that it holds:

$$\widehat{k}_j^n = \widehat{k}_{j-1}^n + m_1^n + \widehat{l}_1^n,$$
$$\widehat{k}_s^n = \widehat{k}_{s-1}^n + m_2^n + \widehat{l}_2^n,$$

thereby $l_1 \in \{1, \dots, n - m_1^n\}$, $m_2^n \in \{1, \dots, n - m_1^n - l_1\}$ and $l_2 \in \{1, \dots, n - m_1^n - l_1 - m_2^n\}$. From this it follows:

$$\widehat{f}_{n,i} = f_1 + \operatorname{med}\,(\xi^n_{\widehat{k}_{i-1}^n+1}, \dots, \xi^n_{\widehat{k}_i^n}), \quad i \in \{1, \dots, j-1\}$$
$$\widehat{f}_{n,i} = f_2 + \operatorname{med}\,(\xi^n_{\widehat{k}_{i-1}^n+1}, \dots, \xi^n_{\widehat{k}_i^n}), \quad i \in \{j+1, \dots, s-1\}$$
$$\widehat{f}_{n,i} = f_3 + \operatorname{med}\,(\xi^n_{\widehat{k}_{i-1}^n+1}, \dots, \xi^n_{\widehat{k}_i^n}), \quad i \in \{s+1, \dots, \#J(\widehat{f}_n) + 1\}$$

and

$$\widehat{f}_{n,j} = \operatorname{med}\left(\xi^n_{\widehat{k}_{j-1}^n+1\,:\,\widehat{k}_{j-1}^n+m_1^n} + f_1\,,\ \xi^n_{\widehat{k}_{j-1}^n+m_1^n+1\,:\,\widehat{k}_{j-1}^n+m_1^n+l_1} + f_2\right)$$
$$\widehat{f}_{n,s} = \operatorname{med}\left(\xi^n_{\widehat{k}_{s-1}^n+1\,:\,\widehat{k}_{s-1}^n+m_2^n} + f_2\,,\ \xi^n_{\widehat{k}_{s-1}^n+m_2^n+1\,:\,\widehat{k}_{s-1}^n+m_2^n+l_2} + f_3\right).$$

Then for all $n \ge N_f'$ the norm of the difference between the estimate $\widehat{f}_n$ and the original function f can be represented in the form:

$$\begin{aligned}
&\|\,\widehat{f}_n - f\,\| \\
=\ & \int_{[0,1)} \Big|\sum_{i=1}^{\#J(\widehat{f}_n)+1} \widehat{f}_{n,i} \cdot \mathbf{1}_{\widehat{I}_i^n}(x) - f_1 \cdot \mathbf{1}_{[0,\,a_1)}(x) - f_2 \cdot \mathbf{1}_{[a_1,\,a_2)}(x) - f_3 \cdot \mathbf{1}_{[a_2,\,1)}(x)\,\Big|\,dx \\
=\ & \sum_{i=1}^{\#J(\widehat{f}_n)+1} \int_{\widehat{I}_i^n} \big|\,\widehat{f}_{n,i} - f_1 \cdot \mathbf{1}_{[0,\,a_1)}(x) - f_2 \cdot \mathbf{1}_{[a_1,\,a_2)}(x) - f_3 \cdot \mathbf{1}_{[a_2,\,1)}(x)\,\big|\,dx \\
=\ & \sum_{i=1}^{j-1} \int_{\widehat{I}_i^n} |\,\widehat{f}_{n,i} - f_1\,|\,dx + \sum_{i=j+1}^{s-1} \int_{\widehat{I}_i^n} |\,\widehat{f}_{n,i} - f_2\,| + \sum_{i=s+1}^{\#J(\widehat{f}_n)+1} \int_{\widehat{I}_i^n} |\,\widehat{f}_{n,i} - f_3\,|\,dx
\end{aligned}$$

$$
\begin{aligned}
&+ \int\limits_{\widehat{I}^n_{j-1,\,a_1}} |\,\widehat{f}_{n,\,j} - f_1\,|\,dx + \int\limits_{\widehat{I}^n_{a_1,\,j}} |\,\widehat{f}_{n,\,j} - f_2\,|\,dx \\
&+ \int\limits_{\widehat{I}^n_{s-1,\,a_2}} |\,\widehat{f}_{n,\,s} - f_2\,|\,dx + \int\limits_{\widehat{I}^n_{a_2,\,s}} |\,\widehat{f}_{n,\,s} - f_3\,|\,dx \qquad (5.58)\\
= &\sum_{i=1,\,i\neq j,\,i\neq s}^{\#J(\widehat{f}_n)+1} l(\widehat{I}^n_i)\cdot|\,\mathrm{med}\,(\,\xi^n_{\widehat{k}^n_{i-1}+1},\ldots,\xi^n_{\widehat{k}^n_i}\,)|\\
&+ l(\widehat{I}^n_{j-1,\,a_1})\cdot\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1}+f_1\,,\,\xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+\widehat{l}^n_1}+f_2\,\big)-f_1\,\big|\\
&+ l(\widehat{I}^n_{a_1,\,j})\cdot\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1}+f_1\,,\,\xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n_1+\widehat{l}_1{}^n}+f_2\,\big)-f_2\,\big|\\
&+ l(\widehat{I}^n_{s-1,\,a_2})\cdot\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{s-1}+1\,:\,\widehat{k}^n_{s-1}+m^n_2}+f_2\,,\,\xi^n_{\widehat{k}^n_{s-1}+m^n_2+1\,:\,\widehat{k}^n_{s-1}+m^n_2+\widehat{l}^n_2}+f_3\,\big)-f_2\,\big|\\
&+ l(\widehat{I}^n_{a_2,\,s})\cdot\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{s-1}+1\,:\,\widehat{k}^n_{s-1}+m^n_2}+f_2\,,\,\xi^n_{\widehat{k}^n_{s-1}+m^n_2+1\,:\,\widehat{k}^n_{s-1}+m^n_2+\widehat{l}^n_2}+f_3\,\big)-f_3\,\big|\\
\underset{(4.59)}{\overset{\mathbb{P}-\text{a.s.}}{\leq}} &\sum_{i=1,\,i\neq j,\,i\neq s}^{\#J(\widehat{f}_n)+1} \frac{\widehat{k}^n_i-\widehat{k}^n_{i-1}}{n}\cdot\frac{\sqrt{n}\,\ln n}{\widehat{k}^n_i-\widehat{k}^n_{i-1}}\\
&+ l(\widehat{I}^n_{j-1,\,a_1})\cdot\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1}\,,\,\xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+\widehat{l}^n_1}+(f_2-f_1)\,\big)\,\big|\\
&+ l(\widehat{I}^n_{a_1,\,j})\cdot\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1}+(f_1-f_2)\,,\,\xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+\widehat{l}_1{}^n}\,\big)\,\big|\\
&+ l(\widehat{I}^n_{s-1,\,a_2})\cdot\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{s-1}+1\,:\,\widehat{k}^n_{s-1}+m^n_2}\,,\,\xi^n_{\widehat{k}^n_{s-1}+m^n_2+1\,:\,\widehat{k}^n_{s-1}+m^n_2+\widehat{l}^n_2}+(f_3-f_2)\,\big)\,\big|\\
&+ l(\widehat{I}^n_{a_2,\,s})\cdot\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{s-1}+1\,:\,\widehat{k}^n_{s-1}+m^n_2}+(f_2-f_3)\,,\,\xi^n_{\widehat{k}^n_{s-1}+m^n_2+1\,:\,\widehat{k}^n_{s-1}+m^n_2+\widehat{l}^n_2}\,\big)\,\big|
\end{aligned}
$$

Taking into account (5.6), $\mathbb{P}$-almost surely boundedness of ξ^n, Condition 4 and the fact that $\#J(f^n)=2$ for all $n\geq N'_f$, consider $\sum\limits_{i=1,\,i\neq j,\,i\neq s}^{\#J(\widehat{f}_n)+1}\dfrac{\ln n}{\sqrt{n}}$:

$$
\begin{aligned}
&\sum_{i=1,\,i\neq j,\,i\neq s}^{\#J(\widehat{f}_n)+1}\frac{\ln n}{\sqrt{n}} = \Big(\#J(\widehat{f}_n)-1\Big)\,\frac{\ln n}{\sqrt{n}} \leq \Big(\frac{\|\xi^n\|}{\gamma_n}+\#J(f^n)-1\Big)\cdot\frac{\ln n}{\sqrt{n}}\\
=\;&\Big(\frac{\|\xi^n\|}{\gamma_n}+1\Big)\cdot\frac{\ln n}{\sqrt{n}} \leq \frac{\|\xi^n\|}{\ln n}+\frac{\ln n}{\sqrt{n}}.
\end{aligned}
$$

According to this and using the following notations:

$$\Delta_1 := f_2 - f_1,$$

$$\Delta_2 := f_3 - f_2,$$

we conclude:

$$
\begin{aligned}
&\|\,\widehat{f}_n - f\,\| \leq \frac{\|\xi^n\|}{\ln n}+\frac{\ln n}{\sqrt{n}}\\
+\;& l(\widehat{I}^n_{j-1,\,a_1})\cdot\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n}\,,\,\xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}^n_1}+\Delta_1\,\big)\,\big|\\
+\;& l(\widehat{I}^n_{a_1,\,j})\cdot\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n}-\Delta_1\,,\,\xi^n_{\widehat{k}^n_{j-1}+m^n+1\,:\,\widehat{k}^n_{j-1}+m^n+\widehat{l}_1{}^n}\,\big)\,\big|
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad l(\widehat{I}^n_{s-1,\,a_2}) \cdot \big|\ \mathrm{med}\left(\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_0}\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_0+1\,:\,\widehat{k}^n_{j-1}+m^n_0+\widehat{l}^n_2} + \Delta_2\right) \big| \\
&+ \quad l(\widehat{I}^n_{a_2,\,s}) \cdot \big|\ \mathrm{med}\left(\xi^n_{\widehat{k}^n_{s-1}+1\,:\,\widehat{k}^n_{s-1}+m^n_0} - \Delta_2\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_0+1\,:\,\widehat{k}^n_{j-1}+m^n_0+\widehat{l}^n_2}\right) \big| \qquad (5.59)
\end{aligned}
$$

In the proof of Theorem 6, we have already shown that all terms in the right-hand side of inequality (5.59) $\mathbb{P}$-almost surely tend to zero as n tends to infinity with the rate of convergence of at least $\dfrac{1}{\ln n}$. Consequently, the left-hand side of inequality (5.59) $\mathbb{P}$-almost surely also tends to zero as n tends to infinity. The rate of convergence remains unchanged.

Case 2. Next, we have to consider the situation when there exists only one number $j \in \big\{1, \ldots, \#J(\widehat{f}_n){+}1\big\}$ such that both points a_1 and a_2 belong to the interval $\Big[\dfrac{\widehat{k}^n_{j-1}}{n},\ \dfrac{\widehat{k}^n_j}{n}\Big) = \widehat{I}^n_j$. Further on, we fix this integer j. It is clear that the points a_1 and a_2 split the interval $\widehat{I}^n_j$ into three intervals, which we will denote as follows:

$$
\begin{aligned}
\widehat{I}^n_{j-1,\,a_1} &:= \Big[\frac{\widehat{k}^n_{j-1}}{n},\ a_1\Big), \\
\widehat{I}^n_{a_1,\,a_2} &:= [a_1,\ a_2), \\
\widehat{I}^n_{a_2,\,j} &:= \Big[a_2,\ \frac{\widehat{k}^n_j}{n}\Big).
\end{aligned}
$$

Similarly to Case **1**, also in this case, for all $n \geq N'_f$, where N'_f is defined by (5.57), the function f^n has exactly two jumps. However, the coordinates of these two jumps in Case **2** will be denoted by $\dfrac{\widehat{k}^n_{j-1} + m^n_1}{n}$ and $\dfrac{\widehat{k}^n_{j-1} + m^n_1 + m^n_2}{n}$ for some $m^n_1 \in \{1, \ldots, n-1\}$ and $m^n_2 \in \{1, \ldots, n - m_1\}$, respectively. These numbers obey the following conditions:

$$
0 \leq \frac{\widehat{k}^n_{j-1} + m^n_1}{n} - a_1 \leq \frac{1}{n}, \qquad (5.60)
$$

$$
0 \leq \frac{\widehat{k}^n_{j-1} + m^n_1 + m^n_2}{n} - a_2 \leq \frac{1}{n}. \qquad (5.61)
$$

Therefore, for any $n \geq N'_f$ in Case **2** the function f^n has the form:

$$
f^n(x) = f_1 \cdot \mathbf{1}_{\left[0,\ \frac{\widehat{k}^n_{j-1}}{n} + \frac{m^n_1}{n}\right)}(x) + f_2 \cdot \mathbf{1}_{\left[\frac{\widehat{k}^n_{j-1}}{n} + \frac{m^n_1}{n},\ \frac{\widehat{k}^n_{j-1}}{n} + \frac{m^n_1}{n} + \frac{m^n_2}{n}\right)}(x) + f_3 \cdot \mathbf{1}_{\left[\frac{\widehat{k}^n_{j-1}}{n} + \frac{m^n_1}{n} + \frac{m^n_2}{n},\, 1\right)}(x),
$$

Similar to the above for all $n \geq N'_f$ we define a number $\widehat{l}^n$ according to:

$$
\widehat{l}^n := \widehat{k}^n_j - m^n_1 - m^n_2 - \widehat{k}^n_{j-1},
$$

which means that:

$$
\widehat{k}^n_j = \widehat{k}^n_{j-1} + m^n_1 + m^n_2 + \widehat{l}^n, \qquad (5.62)
$$

or, alternatively:

$$
l(\widehat{I}^n_j) = \frac{1}{n}(\widehat{k}^n_j - \widehat{k}^n_{j-1}) = \frac{m^n_1 + m^n_2 + \widehat{l}^n}{n}. \qquad (5.63)
$$

From the last relation it is clear, that the number $\widehat{l}^n$ is a natural number from the set $\{1, \ldots, n - m_1^n - m_2^n\}$.
Taking all these facts into account, as well as using notation (5.24), it is clear that $\widehat{f}_{n,j} = \mathrm{med}_{\widehat{I}_j^n}(f^n + \xi^n) \in med_{\widehat{I}_j^n}(f^n + \xi^n)$ for all $n \geq N_f'$ can be represented in the following form:

$$\widehat{f}_{n,i} = f_1 + \mathrm{med}\,(\,\xi^n_{\widehat{k}^n_{i-1}+1}, \ldots, \xi^n_{\widehat{k}^n_i}\,), \quad i \in \{1, \ldots, j-1\} \tag{5.64}$$

$$\widehat{f}_{n,i} = f_3 + \mathrm{med}\,(\,\xi^n_{\widehat{k}^n_{i-1}+1}, \ldots, \xi^n_{\widehat{k}^n_i}\,), \quad i \in \{j+1, \ldots, \#J(\widehat{f}_n) + 1\} \tag{5.65}$$

and

$$\widehat{f}_{n,j} = \mathrm{med}\left(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m_1^n} + f_1\,,\; \xi^n_{\widehat{k}^n_{j-1}+m_1^n+1\,:\,\widehat{k}^n_{j-1}+m_1^n+m_2^n} + f_2,\; \xi^n_{\widehat{k}^n_{j-1}+m_1^n+m_2^n+1\,:\,\widehat{k}^n_j} + f_3\,\right),$$

thereby m_1^n, m_2^n and $\widehat{l}^n$ can be seen as the orders of the three samples of symmetrical Pareto distributed random variables shifted by f_1, f_2 and f_3, respectively.
Now we are ready to consider $\|\,\widehat{f}_n - f\,\|$:

$$\begin{aligned}
\|\,\widehat{f}_n - f\,\| &= \sum_{i=1}^{\#J(\widehat{f}_n)+1} \int\limits_{\widehat{I}_i^n} \Big|\, \widehat{f}_{n,i} - f_1 \cdot \mathbf{1}_{[0,\,a_1)}(x) - f_2 \cdot \mathbf{1}_{[a_1,\,a_2)}(x) - f_3 \cdot \mathbf{1}_{[a_2,\,1)}(x)\,\Big|\, dx \\
&= \sum_{i=1}^{j-1} \int\limits_{\widehat{I}_i^n} |\,\widehat{f}_{n,i} - f_1\,|\, dx + \sum_{i=j+1}^{\#J(\widehat{f}_n)+1} \int\limits_{\widehat{I}_i^n} |\,\widehat{f}_{n,i} - f_3\,|\, dx \\
&\quad + \int\limits_{\widehat{I}^n_{j-1,\,a_1}} |\,\widehat{f}_{n,j} - f_1\,|\, dx + \int\limits_{\widehat{I}^n_{a_1,\,a_2}} |\,\widehat{f}_{n,j} - f_2\,|\, dx + \int\limits_{\widehat{I}^n_{a_2,\,j}} |\,\widehat{f}_{n,j} - f_3\,|\, dx \\
&= \sum_{i=1,\; i\neq j}^{\#J(\widehat{f}_n)+1} l(\widehat{I}_i^n) \cdot |\,\mathrm{med}\,(\,\xi^n_{\widehat{k}^n_{i-1}+1}, \ldots, \xi^n_{\widehat{k}^n_i}\,)| \\
&\quad + l(\widehat{I}^n_{j-1,\,a_1}) \cdot |\,\widehat{f}_{n,j} - f_1\,| + l(\widehat{I}^n_{a_1,a_2}) \cdot |\,\widehat{f}_{n,j} - f_2\,| + l(\widehat{I}^n_{a_2,\,j}) \cdot |\,\widehat{f}_{n,j} - f_3\,|
\end{aligned} \tag{5.66}$$

Below we separately analyze different terms of the last sum.
We start with the sum $\sum\limits_{\substack{i=1\\ i\neq j}}^{\#J(\widehat{f}_n)+1} l(\widehat{I}_i^n) \cdot |\,\mathrm{med}\,(\,\xi^n_{\widehat{k}^n_{i-1}+1}, \ldots, \xi^n_{\widehat{k}^n_i}\,)|$, which has been estimated in Theorem 6 or also in Case **1** using inequalities (4.59), (5.6), Condition 4 and $\mathbb{P}$-almost surely boundedness of ξ^n. As a result we have:

$$\begin{aligned}
\sum_{\substack{i=1\\ i\neq j}}^{\#J(\widehat{f}_n)+1} l(\widehat{I}_i^n) \cdot |\,\mathrm{med}\,(\,\xi^n_{\widehat{k}^n_{i-1}+1}, \ldots, \xi^n_{\widehat{k}^n_i}\,)| &\leq \left(\frac{\|\xi^n\|}{\gamma_n} + \#J(f^n)\right) \cdot \frac{\ln n}{\sqrt{n}} \\
&\leq \frac{\|\xi^n\|}{\ln n} + 2\,\frac{\ln n}{\sqrt{n}} = \mathrm{O}\left(\frac{1}{\ln n}\right) \quad \mathbb{P} - \text{a.s.}
\end{aligned} \tag{5.67}$$

The quantities $|\,\widehat{f}_{n,j} - f_1\,|, |\,\widehat{f}_{n,j} - f_2\,|$ and $|\,\widehat{f}_{n,j} - f_3\,|$ will be estimated in the analogous way to the corresponding estimations in Theorem 6.

Namely, if we define

$$\begin{aligned}\Delta_1 &:= f_2 - f_1,\\ \Delta_2 &:= f_3 - f_2,\\ \Delta_3 &:= f_3 - f_1,\end{aligned}$$

then similarly to (5.42) in the Theorem 6, for all $n \geq N'_f$ it holds on one hand:

$$\begin{aligned}&\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1} + f_1\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+m^n_2} + f_2,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+m^n_2+1\,:\,\widehat{k}^n_j} + f_3\,\big)\\ \leq\ &\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1}\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+m^n_2} + \Delta_1,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+m^n_2+1\,:\,\widehat{k}^n_j} + \Delta_3\,\big) + f_1\big|,\end{aligned}$$

on the other hand:

$$\begin{aligned}&\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1} + f_1\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+m^n_2} + f_2,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+m^n_2+1\,:\,\widehat{k}^n_j} + f_3\,\big)\\ \leq\ &\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1} - \Delta_1\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+m^n_2},\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+m^n_2+1\,:\,\widehat{k}^n_j} - \Delta_2\,\big) + f_2\big|\end{aligned}$$

and finally:

$$\begin{aligned}&\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1} + f_1\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+m^n_2} + f_2,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+m^n_2+1\,:\,\widehat{k}^n_j} + f_3\,\big)\\ \leq\ &\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1} - \Delta_3\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+m^n_2} - \Delta_2,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+m^n_2+1\,:\,\widehat{k}^n_j}\,\big) + f_3\big|.\end{aligned}$$

From this results it follows:

$$|\,\widehat{f}_{n,j} - f_1\,| \leq \big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1}\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+m^n_2} + \Delta_1,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+m^n_2+1\,:\,\widehat{k}^n_j} + \Delta_3\,\big)\big|,$$

$$|\,\widehat{f}_{n,j} - f_2\,| \leq \big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1} - \Delta_1\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+m^n_2},\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+m^n_2+1\,:\,\widehat{k}^n_j} - \Delta_2\,\big)\big|$$

and

$$|\,\widehat{f}_{n,j} - f_3\,| \leq \big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1} - \Delta_3\,,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+m^n_2} - \Delta_2,\ \xi^n_{\widehat{k}^n_{j-1}+m^n_1+m^n_2+1\,:\,\widehat{k}^n_j}\,\big)\big|.$$

First of all, we notice that since $l(\widehat{I}^n_{a_1,a_2}) = (a_2 - a_1) > 0$, as we know from Definition (5.56) of f, then there exist a natural number N^1_f such that for all $n \geq N^1_f$ it holds:

$$l(\widehat{I}^n_{a_1,a_2}) \geq \frac{1}{\ln n}. \tag{5.68}$$

Moreover, we recall that for all $n \geq N'_f$, the function f^n has the same number of jumps as the function f, namely it has two jumps. This means that for all $n \geq N'_f$ the length of the second step of the function f^n, which in our notations equals $\dfrac{m^n_2}{n}$ is greater than 0, i.e. $\dfrac{m^n_2}{n} > 0$. Consequently, there exists a number $N^2_f \geq N'_f$ such that for all $n \geq N^2_f$ it holds:

$$\frac{m^n_2}{n} \geq \frac{1}{\ln n}. \tag{5.69}$$

Thus, if we define

$$N_f = \max\{N_f^1, N_f^2\},$$

then it is clear that for all $n \geq N_f$ the both inequalities (5.68) and (5.69) are valid.
Because of Condition 3 on the sequences $(\gamma_n)_{n\in\mathbb{N}}$ we can use Lemma 19 and Remark 4, according to which it is $\mathbb{P}$-almost surely impossible that the intervals $l(\widehat{I}^n_{j-1,\,a_1})$ and $l(\widehat{I}^n_{a_2,\,j})$, as well as $\frac{m_1^n}{n}$ and $\frac{\widehat{l}^n}{n}$, are of the order $\left(\frac{1}{\ln n}\right)$ or larger.
Therefore, we can restrict ourselves to the following case :

$$l(\widehat{I}^n_{a_1,\,a_2}) \geq \frac{1}{\ln n}, \qquad l(\widehat{I}^n_{j-1,\,a_1}) = \mathrm{o}\left(\frac{1}{\ln n}\right) \quad \text{and} \quad l(\widehat{I}^n_{a_2,\,j}) = \mathrm{o}\left(\frac{1}{\ln n}\right) \tag{5.70}$$

and respectively

$$\frac{m_2^n}{n} \geq \frac{1}{\ln n}, \quad \frac{m_1^n}{n} = \mathrm{o}\left(\frac{1}{\ln n}\right), \quad \frac{\widehat{l}^n}{n} = \mathrm{o}\left(\frac{1}{\ln n}\right). \tag{5.71}$$

In what follows, we proceed analogously to Theorem 6.
We begin with $l(\widehat{I}^n_{j-1,\,a_1}) \cdot |\,\widehat{f}_{n,j} - f_1\,|$. Here we use Lemma 25 from Appendix A, p.123, which yields the following result:

$$\begin{aligned}
&\Big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m_1^n}\,,\ \xi^n_{\widehat{k}^n_{j-1}+m_1^n+1\,:\,\widehat{k}^n_{j-1}+m_1^n+m_2^n} + \Delta_1,\ \xi^n_{\widehat{k}^n_{j-1}+m_1^n+m_2^n+1\,:\,\widehat{k}^n_j} + \Delta_3\,\big)\Big| \\
\leq\ &|\,\mathrm{med}\,(\,\xi^n_{\widehat{k}^n_{j-1}+1}, \ldots, \xi^n_{\widehat{k}^n_j}\,)\,| + \max\big\{|\Delta_1|,\, |\Delta_3|\big\},
\end{aligned}$$

Hence,

$$\begin{aligned}
& l(\widehat{I}^n_{j-1,\,a_1}) \cdot |\,\widehat{f}_{n,j} - f_1\,| \\
\leq\ & l(\widehat{I}^n_{j-1,\,a_1}) \cdot \Big(\,|\,\mathrm{med}\,(\,\xi^n_{\widehat{k}^n_{j-1}+1}, \ldots, \xi^n_{\widehat{k}^n_j}\,)\,| + \max\big\{|\Delta_1|,\, |\Delta_3|\big\}\Big) \\
\underset{(4.59)}{\overset{\mathbb{P}-\text{a.s.}}{\leq}}\ & \frac{l(\widehat{I}^n_{j-1,\,a})}{l(\widehat{I}^n_j)} \cdot \frac{\ln n}{\sqrt{n}} + l(\widehat{I}^n_{j-1,\,a}) \cdot \max\big\{|\Delta_1|,\, |\Delta_3|\big\} \\
\underset{(5.70)}{\leq}\ & \frac{\ln n}{\sqrt{n}} + \frac{\max\big\{|\Delta_1|,\, |\Delta_3|\big\}}{\ln n} = \mathrm{O}\left(\frac{1}{\ln n}\right) \underset{n\to\infty}{\longrightarrow} 0 \quad \mathbb{P}-\text{a.s.}
\end{aligned} \tag{5.72}$$

For $l(\widehat{I}^n_{a_2,\,j}) \cdot |\,\widehat{f}_{n,j} - f_3\,|$ a similar estimate is valid $\mathbb{P}$-almost surely :

$$l(\widehat{I}^n_{a_2,\,j}) \cdot |\,\widehat{f}_{n,j} - f_3\,| \;\leq\; \frac{\ln n}{\sqrt{n}} + \frac{\max\big\{|\Delta_1|,\, |\Delta_2|\big\}}{\ln n} = \mathrm{O}\left(\frac{1}{\ln n}\right) \underset{n\to\infty}{\longrightarrow} 0 \tag{5.73}$$

Finally, all that remains is to show:

$$l(\widehat{I}^n_{a_1,\,a_2}) \cdot |\,\widehat{f}_{n,j} - f_2\,| \;\leq\; \mathrm{O}\left(\frac{1}{\ln n}\right) \underset{n\to\infty}{\longrightarrow} 0. \quad \mathbb{P}-\text{a.s.} \tag{5.74}$$

The conditions (5.71) allow us firstly, to conclude the following facts:

$$m_2^n \geq \frac{n}{\ln n}, \tag{5.75}$$

$$\lim_{n\to\infty} \frac{m_1^n}{m_2^n} = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{\widehat{l}^n}{m_2^n} = 0, \tag{5.76}$$

and secondly, to apply Lemma 14 in order to estimate the quantity
$\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1} + \Delta_1\,,\,\xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+m^n_2},\,\xi^n_{\widehat{k}^n_{j-1}+m^n_1+m^n_2+1\,:\,\widehat{k}^n_j} + \Delta_2\,\big)\big|$.
Lemma 14, under the notations:

$$\widehat{\nu}^n := \frac{m^n_1 + \widehat{l}^n}{m^n_1 + m^n_2 + \widehat{l}^n}. \tag{5.77}$$

yields the following result:

$$\begin{aligned}
&\big|\,\mathrm{med}\,\big(\,\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1} + \Delta_1\,,\,\xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+m^n_2},\,\xi^n_{\widehat{k}^n_{j-1}+m^n_1+m^n_2+1\,:\,\widehat{k}^n_j} + \Delta_2\,\big)\big| \\
\le\ & \max\Big\{ -\Big(z^-_{\widehat{\nu}^n} - \frac{\sqrt{n}\,\ln n}{m^n_2}\Big),\ z^+_{\widehat{\nu}^n} + \frac{\sqrt{n}\,\ln n}{m^n_2}\Big\} \quad \mathbb{P}-\text{a.s.},
\end{aligned} \tag{5.78}$$

here $z^-_{\widehat{\nu}^n} = \widetilde{\xi}_{\frac{1}{2}-\widehat{\nu}^n}$ and $z^+_{\widehat{\nu}^n} = \widetilde{\xi}_{\frac{1}{2}+\widehat{\nu}^n}$ are the appropriate quantiles of a random variable ξ, which is identically distributed with ξ^n_i for all n and all $i \in \{1,\ldots,n\}$.
From (5.76) and (5.77) it is obvious that $\widehat{\nu}^n \underset{n\to\infty}{\longrightarrow} 0$, this means that the quantiles $z^-_{\widehat{\nu}^n}$ and $z^+_{\widehat{\nu}^n}$ can be represented using Taylor's formula similarly to (5.54):

$$\begin{aligned}
-z^-_{\widehat{\nu}^n} = z^+_{\widehat{\nu}^n} &= \frac{1}{(1-2\,\widehat{\nu}^n)^{\frac{1}{\alpha-1}}} - 1 = \frac{2}{\alpha-1}\cdot\widehat{\nu}^n + \mathrm{O}\left((\widehat{\nu}^n)^2\right) \\
&= \frac{2}{\alpha-1}\cdot\frac{m^n_1+\widehat{l}^n}{m^n_1+m^n_2+\widehat{l}^n} + \mathrm{O}\Big(\Big(\frac{m^n_1+\widehat{l}^n}{m^n_1+m^n_2+\widehat{l}^n}\Big)^2\Big).
\end{aligned} \tag{5.79}$$

Taking into account equations (5.78) and (5.79) and additionally (5.63) and (5.75) we obtain that it holds $\mathbb{P}$-almost surely:

$$\begin{aligned}
& l(\widehat{I}^n_{a_1,a_2})\cdot|\,\widehat{f}_{n,j} - f_2\,| \\
\le\ & l(\widehat{I}^n_{a_1,a_2})\cdot\big|\,\mathrm{med}\,\big(\xi^n_{\widehat{k}^n_{j-1}+1\,:\,\widehat{k}^n_{j-1}+m^n_1} + \Delta_1,\,\xi^n_{\widehat{k}^n_{j-1}+m^n_1+1\,:\,\widehat{k}^n_{j-1}+m^n_1+m^n_2},\,\xi^n_{\widehat{k}^n_{j-1}+m^n_1+m^n_2+1\,:\,\widehat{k}^n_j} + \Delta_2\big)\big| \\
\le\ & l(\widehat{I}^n_j)\cdot\Big(\frac{\sqrt{n}\,\ln n}{m^n_2} + \frac{2}{\alpha-1}\cdot\frac{m^n_1+\widehat{l}^n}{m^n_1+m^n_2+\widehat{l}^n} + \mathrm{O}\Big(\Big(\frac{m^n_1+\widehat{l}^n}{m^n_1+m^n_2+\widehat{l}^n}\Big)^2\Big)\Big) \\
=\ & \frac{m^n_1+m^n_2+\widehat{l}^n}{n}\cdot\Big(\frac{\sqrt{n}\,\ln n}{m^n_2} + \frac{2}{\alpha-1}\cdot\frac{m^n_1+\widehat{l}^n}{m^n_1+m^n_2+\widehat{l}^n} + \mathrm{O}\Big(\Big(\frac{m^n_1+\widehat{l}^n}{m^n_1+m^n_2+\widehat{l}^n}\Big)^2\Big)\Big) \\
\le\ & \Big(\frac{\sqrt{n}\,\ln n}{m^n_2} + \frac{2}{\alpha-1}\cdot\frac{m^n_1+\widehat{l}^n}{n} + \mathrm{O}\Big(\frac{m^n_1+\widehat{l}^n}{n}\cdot\frac{m^n_1+\widehat{l}^n}{m^n_1+m^n_2+\widehat{l}^n}\Big)\Big) \\
<\ & \frac{\ln^2 n}{\sqrt{n}} + \frac{4}{\alpha-1}\cdot\frac{1}{\ln n} + \mathrm{O}\Big(\frac{1}{\ln n}\Big) = \mathrm{O}\Big(\frac{1}{\ln n}\Big)\underset{n\to\infty}{\longrightarrow} 0
\end{aligned}$$

Consequently, we have proved the validity of (5.74). This, together with (5.67), (5.72), (5.73) and (5.66) leads us to conclude the validity of the fact:

$$\|\,\widehat{f}_n - f\,\| = \mathrm{O}\Big(\frac{1}{\ln n}\Big)\underset{n\to\infty}{\longrightarrow} 0 \quad \mathbb{P}-\text{a.s.}$$

and therefore of Theorem 7.

□

All theorems, which we have proved until now in this section give us the same information about consistence and convergence rate of the considered estimator for a piecewise constant function with zero, one or two jumps.
In principle consistency and convergence rate of the estimator for a step function with arbitrary number of jumps can be derived from Theorems 5, 6, 7 and Lemma 19. This result is presented in the last theorem of this thesis.

Theorem 8. *Let $f \in S\big([0,1)\big)$, $(\xi_i^n)_{1\le i\le n,\, n\in\mathbb{N}}$ fulfill Condition 2 and*

$$\widehat{f}_n \in \operatorname{argmin}\, \tilde{H}^n_{\gamma_n}(\cdot\,, f^n+\xi^n).$$

Then for all $(\gamma_n)_{n\in\mathbb{N}}$, which satisfy Condition 3 and Condition 4, it holds $\mathbb{P}$-almost surely:

$$\|\,\widehat{f}_n - f\,\| = \mathrm{O}\Big(\frac{1}{\ln n}\Big). \tag{5.80}$$

Proof:
Let f be an arbitrary step function on the interval [0, 1). We assume that the points $a_i \in (0,1),\ i \in \{1,\dots,\#J(f)\}$, which obey $a_1 \neq a_2 \neq \dots \neq a_{\#J(f)}$, are the jump coordinates of the function f and $f_i \in \mathbb{R}$ with $f_i < \infty$ and $f_i \neq f_{i+1}$ for any $i \in \{1,\dots,\#J(f)\}$ are the values of the function f on the intervals $[a_{i-1},\, a_i)$, i.e.

$$f(x) := \sum_{i=1}^{\#J(f)+1} f_i \cdot \mathbf{1}_{[a_{i-1},\,a_i)}(x), \quad a_0 := 0,\ a_{\#J(f)+1} := 1.$$

We start the proof with the discussion of the relative position of the jump points of the function f with respect to the position of the jump points of the function $\widehat{f}_n$, which for any $n \in \mathbb{N}$ is a minimizer of the functional $\tilde{H}^n_{\gamma_n}(\cdot\,, f^n+\xi^n)$.
Before doing this, we recall that for all $n \in \mathbb{N}$ the representation of the function $\widehat{f}_n$ is defined as given above in Remark 2, i.e.

$$\widehat{f}_n(x) = \sum_{i=1}^{\#J(\widehat{f}_n)+1} \widehat{f}_{n,i}\cdot \mathbf{1}_{\widehat{I}_i^n}(x),$$

where
$J(\widehat{f}_n) := \Big\{\dfrac{\widehat{k}_1^n}{n},\dots,\dfrac{\widehat{k}^n_{\#J(\widehat{f}_n)}}{n}\Big\}$, $\widehat{k}_1^n \neq \dots \neq \widehat{k}^n_{\#J(\widehat{f}_n)} \in \{1,\dots n-1\}$, $\widehat{k}_0^n := 0$, $\widehat{k}^n_{\#J(\widehat{f}_n)+1} := n$,
$\widehat{I}_i^n := \Big[\dfrac{\widehat{k}^n_{i-1}}{n}, \dfrac{\widehat{k}^n_i}{n}\Big)$, $P_{J(\widehat{f}_n)} := \big\{\widehat{I}_1^n,\dots,\widehat{I}^n_{\#J(\widehat{f}_n)+1}\big\}$ and
$\widehat{f}_{n,i} = \mathrm{med}_{\widehat{I}_i^n}(f^n+\xi^n) \in \mathit{med}_{\widehat{I}_i^n}(f^n+\xi^n) \quad \forall\, i \in \big\{1,\dots,\#J(\widehat{f}_n)+1\big\}$.

We note that for a sufficiently large $n \in \mathbb{N}$ it is not difficult to show that the situation, when there exists a number $j \in \big\{1,\dots,\#J(\widehat{f}_n)+1\big\}$ such that any three neighboring jump points of the function f belong to the same interval $\widehat{I}_j^n$ is $\mathbb{P}$-almost surely impossible.
In fact, assume the opposite, namely assume that for all $n \in \mathbb{N}$ there exist some $j \in \big\{1,\dots,\#J(\widehat{f}_n)+1\big\}$ and $s \in \big\{1,\dots,\#J(\widehat{f})-2\big\}$ such that $a_s, a_{s+1}, a_{s+2} \in \widehat{I}_j^n$. According to Lemma 19, $\mathbb{P}$-almost surely for all sufficiently large n the length of the intervals $[a_s,\, a_{s+1})$ and $[a_{s+1},\, a_{s+2})$ cannot be simultaneously longer than or equal to $\dfrac{1}{\ln n}$. On the other

hand, we know that $a_s \neq a_{s+1}$ and $a_{s+1} \neq a_{s+2}$, which means that $l[a_s,\, a_{s+1}) > 0$ and $l[a_{s+1},\, a_{s+2}) > 0$ and therefore there always exist such N_0 that for all $n > N_0$ both inequalities: $l[a_s,\, a_{s+1}) \geq \dfrac{1}{\ln n}$ and $l[a_{s+1},\, a_{s+2}) \geq \dfrac{1}{\ln n}$ are valid. Therefore we have shown that $\mathbb{P}$-almost surely for all sufficiently large n and all $j \in \big\{1, \ldots, \#J(\widehat{f}_n)+1\big\}$ it holds that at most two jump points of f can belong to one interval $\widehat{I}^n_j$.
According to this, for all sufficiently large n the norm $\|\widehat{f}_n - f\|$ can be represented in the following way:

$$
\begin{aligned}
\|\widehat{f}_n - f\| &= \sum_{i=1}^{\#J(\widehat{f}_n)+1} \int\limits_{\widehat{I}^n_i} \big|\, \widehat{f}_{n,i} - f(x) \,\big|\, dx \\
&= \sum_{\substack{i=1,\\ \widehat{I}^n_i \,\cap\, J(f)=\emptyset}}^{\#J(\widehat{f}_n)+1} \int\limits_{\widehat{I}^n_i} \big|\, \widehat{f}_{n,i} - f(x) \,\big|\, dx + \sum_{\substack{i=1,\\ \widehat{I}^n_i \,\cap\, J(f)\neq\emptyset}}^{\#J(\widehat{f}_n)+1} \int\limits_{\widehat{I}^n_i} \big|\, \widehat{f}_{n,i} - f(x) \,\big|\, dx
\end{aligned}
\tag{5.81}
$$

First of all, we note that the number of jumps of f is finite, that means, firstly, $\#J(f^n) < \infty$ for all $n \in \mathbb{N}$ and, secondly, the number of summands of the second sum in the right-hand side of (5.81) is finite as well.
The first sum in the right-hand side of (5.81) can be estimated by means of Conditions 4, inequalities (4.59), (5.6) and (4.3), similarly to the estimations of analogous sum in Theorem 6 or in Theorem 7 (see for instance (5.67), p. 105). Namely,

$$
\begin{aligned}
&\sum_{\substack{i=1,\\ \widehat{I}^n_i \cap J(f)=\emptyset}}^{\#J(\widehat{f}_n)+1} \int\limits_{\widehat{I}^n_i} \big|\, \widehat{f}_{n,i} - f(x) \,\big|\, dx = \sum_{\substack{i=1,\\ \widehat{I}^n_i \cap J(f)=\emptyset}}^{\#J(\widehat{f}_n)+1} l(\widehat{I}^n_i) \cdot |\, \mathrm{med}\,(\, \xi^n_{\widehat{k}^n_{i-1}+1}, \ldots, \xi^n_{\widehat{k}^n_i}\,)| \\
\overset{\mathbb{P}-\text{a.s.}}{\leq}\;& \Big(\frac{\|\xi^n \cdot\|}{\gamma_n} + \#J(f^n)\Big)\cdot \frac{\ln n}{\sqrt{n}} \leq \Big(\frac{\|\xi^n\| \cdot \sqrt{n}}{\ln^2 n} + \#J(f^n)\Big)\cdot \frac{\ln n}{\sqrt{n}} \\
=\;& \frac{\|\xi^n\|}{\ln n} + \#J(f^n)\cdot \frac{\ln n}{\sqrt{n}} = \mathrm{O}\Big(\frac{1}{\ln n}\Big).
\end{aligned}
\tag{5.82}
$$

Further on we consider in details the second sum in the right-hand side of (5.81). According to above, $\mathbb{P}$-almost surely we have:

$$\#\{\widehat{I}^n_i \,\cap\, J(f) \neq \emptyset\} \leq 2.$$

From this it follows:

$$
\begin{aligned}
&\sum_{\substack{i=1,\\ \widehat{I}^n_i \cap J(f)\neq\emptyset}}^{\#J(\widehat{f}_n)+1} \int\limits_{\widehat{I}^n_i} \big|\, \widehat{f}_{n,i} - f(x) \,\big|\, dx \\
=\;& \sum_{\substack{i=1,\\ \#\{\widehat{I}^n_i \cap J(f)\}=1}}^{\#J(\widehat{f}_n)+1} \int\limits_{\widehat{I}^n_i} \big|\, \widehat{f}_{n,i} - f(x) \,\big|\, dx + \sum_{\substack{i=1,\\ \#\{\widehat{I}^n_i \cap J(f)\}=2}}^{\#J(\widehat{f}_n)+1} \int\limits_{\widehat{I}^n_i} \big|\, \widehat{f}_{n,i} - f(x) \,\big|\, dx \quad \mathbb{P}-\text{a.s.}
\end{aligned}
\tag{5.83}
$$

The summands of the first sum of (5.83) are similar to the components of (5.58), p.103 in the proof of Theorem 7 and summands of the second sum are similar to the components in (5.66), p.105 in the same proof. In the proof of Theorem 7 we have shown that each of this components $\mathbb{P}$-almost surely tends to 0 as n tends to infinity with the rate of convergence $\frac{1}{\ln n}$. According to this, we deduce that each component of each sums in the right-hand side of (5.83) converges to 0 as n goes to infinity with the rate of convergence $\frac{1}{\ln n}$ and thereby the number of all summands is finite. Therefore we conclude:

$$\sum_{\substack{i=1,\\ \widehat{I}_i^n \,\cap\, J(f)\neq\emptyset}}^{\#J(\widehat{f}_n)+1} \int_{\widehat{I}_i^n} \big|\, \widehat{f}_{n,i} - f(x) \,\big|\, dx = \mathrm{O}\Big(\frac{1}{\ln n}\Big) \quad \mathbb{P}-\text{a.s.} \tag{5.84}$$

After substituting of (5.82) and (5.84) into (5.81) we obtain:

$$\|\widehat{f}_n - f\| = \mathrm{O}\Big(\frac{1}{\ln n}\Big) \quad \mathbb{P}-\text{a.s.} \tag{5.85}$$

Thus the proof of Theorem 8 is complete.

□

5.3 Conclusion

In this chapter we have derived the main result of this work. Namely, for an arbitrary function $f \in S\big([0,1)\big)$ and for any given data $y = (y_1^n, \dots, y_n^n)$, which satisfies the model $y_i^n = f_i^n + \xi_i^n$, where $(\xi_i^n)_{1\leq i\leq n,\, n\in\mathbb{N}}$ fulfills Condition 2 and f_i^n approximate the function f by means of $f(t) = \lim\limits_{n\to\infty} \sum\limits_{i=1}^{n} f_i^n \cdot \mathbf{1}_{\left[\frac{i-1}{n},\, \frac{i}{n}\right)}(t)$, we have shown that the minimizers $\widehat{f}^n$ of the functional $\tilde{H}^n_{\gamma_n}(\cdot\,, f^n + \xi^n)$, for a certain choice of the sequence $(\gamma_n)_{n\in\mathbb{N}}$, converge $\mathbb{P}$-almost surely in the L^1 metric to the original function f, if n approaches infinity.

It was shown that for this convergence to occur, it is sufficient to impose the following conditions on the sequence $(\gamma_n)_{n\in\mathbb{N}} \subset (0,\infty)$:

$$\gamma_n \geq \frac{\ln^2 n}{\sqrt{n}} \qquad \text{and} \qquad \ln n \cdot \gamma_n \underset{n\to\infty}{\longrightarrow} 0.$$

In case, when the original function f is constant, the sufficient condition on the sequence $(\gamma_n)_{n\in\mathbb{N}} \subset (0,\infty)$ is given by the first of these two conditions.

Moreover, we have shown that in the case of the piecewise constant f, the rate of convergence is at least $\mathrm{O}\Big(\frac{1}{\ln n}\Big)$. This rate is not as optimal as the corresponding rate in L^2 case, which is given by $\mathrm{O}\Big(\frac{\ln n}{n}\Big)$. However in the case, when f is constant, we could show that if stronger conditions are applied on $(\gamma_n)_{n\in\mathbb{N}}$, namely if $\gamma_n \geq \frac{1}{\ln^2 n}$, the convergence rate is $\mathrm{O}\Big(\frac{\ln^3 n}{\sqrt{n}}\Big)$.

Additionally, we derived several important facts about the number and location of jumps of the estimators $\widehat{f}^n$, which have also been used in all pivotal theorems.

Chapter 6

Summary and outlook

In this thesis we have considered a regression model of one dimensional noisy data with the regression function taken from the class $L^1([0,1))$ of absolutely integrable on the interval $[0,1)$ functions. In particular, we focused on the piecewise constant jump penalized estimators, which minimize the L^1 version of the so called Potts functional (see e.g. [8]). Originally, the Potts functional is constructed as the jump penalized least square estimator of a function from the class $L^2([0,1))$ of square integrable on the interval $[0,1)$ functions.

More precisely, in this work, the measure of fidelity of the estimator to the given data is assumed to be given by the average absolute deviation of the data points from the estimator. Additionally, the roughness of the estimator is controlled by the total number of its jumps. The underlying objective of this work is to develop a robust regression model, capable of dealing with data, which contains outliers. Following this, the reason behind the particular choice of the class of the estimator functions is dictated by the well known fact that the median is much less sensitive to the outliers than the average.

Technically, the considered regression model is formulated in a very similar way to the model of the jump penalized least square estimators, as studied in details in Refs. [43, 44, 25, 45, 46, 16, 7, 8]. The major difference is that the L^1 estimators are found as a step functions, constructed on the local medians of the data points, contrary to the L^2 estimators, which are given by a step functions, constructed on the local averages. This difference is reflected in the modified Potts functional, which in our case is given by the jump penalized L^1 norm of the difference between the data and the estimator instead of the jump penalized L^2 norm.

Our main result concerns consistency of the minimizers of the modified Potts functional $H^n_{\gamma_n}$, that means the convergence of the corresponding minimizers to the original function f, which is unknown a priori. The only information available about this function is contained in the given noisy data.
On the way to achieve this goal, we have studied the limiting behaviour of $H^n_{\gamma_n}$ and of the respective minimizers as n approaches infinity. This is done in Chapter 2. In Chapter 3 we have demonstrated that the set of minimizers of the limiting functional H^∞_γ is not empty. It should be emphasized that the only condition on the noise term, used in Chapters 2 and 3 is the center symmetry of its distribution.
To prove consistency of the considered estimator the conditions on the noise term had to be strengthened, or in other words the distribution of the random noise had to be specified. Because our main objective was to study data sets that contain outliers, our choice of the

distribution was symmetrical Pareto distribution with the density $f(x) = \dfrac{\alpha - 1}{2\,(1 + |x|)^{\alpha}}$, $\alpha > 1$, as an example of the heavy tailed distribution, which is typically used to describe random data with outliers.
In Chapter 4 we have studied in details the properties of a Pareto distributed noise with the above probability density. In particular, for all sufficiently large n and for an arbitrary sample of the noise data we have obtained certain estimates of the medians and other quantiles of arbitrary subsets of these data with the different number of elements. For the validity of these estimates we had to assumed that the parameter α in the Pareto distribution is not smaller than 9. For consistency, this choice of α is used throughout Chapters 4 and 5.
The final chapter, Chapter 5, contains all details about consistency and rates of convergence of the estimators, which are the minimizers of the modified Potts functional, or in other words are step functions, constructed on the local medians of the data points.

Notice that the study of the jump penalized L^1 estimators for one-dimensional data is far from being complete. In fact, this thesis should be considered as a first step towards a comprehensive theory of L^1 estimators. The list of open questions includes many important issues.
First of all it must be pointed out that consistency of considered estimators was shown only for the case, when the original function f is itself a step function. To prove consistency in the general case, i.e. when $f \in L^1([0,1))$, we could possibly use the fact that the set $S\big([0,1)\big)$ is dense in $L^1([0,1))$.
In more details, we know that for any $\varepsilon > 0$ and for any $f \in L^1\big([0,1)\big)$ there exist functions $\tilde{f} \in S\big([0,1)\big)$ such that it holds:

$$\|\tilde{f} - f\| \leq \varepsilon$$

Moreover, in Theorem 8 we have shown that for any $\tilde{f} \in S\big([0,1)\big)$, any $\varepsilon > 0$, there exists such a natural number N_0 that $\mathbb{P}$-almost surely for all $n \geq N_0$ and for all sequences $(\gamma_n)_{n\in\mathbb{N}}$, which satisfy $\ln n \cdot \gamma_n \underset{n\to\infty}{\longrightarrow} 0$ and $\gamma_n \geq \dfrac{\ln^2 n}{\sqrt{n}}$, it holds:

$$\|\,\widehat{f}_n^{\,0} - \tilde{f}\,\| \leq \varepsilon,$$

where $(\widehat{f}_n^{\,0})_{n\in\mathbb{N}}$ is a sequence of considered estimators of $\tilde{f}$.
Hence, it may make sense to estimate the norm of difference between the original function f and its estimator $\widehat{f}_n$ for all sufficiently large n in the following way:

$$\|\,\widehat{f}_n - f\,\| \leq \|\,\widehat{f}_n - \widehat{f}_n^{\,0}\,\| + \|\,\widehat{f}_n^{\,0} - \tilde{f}\,\| + \|\,\tilde{f} - f\,\|.$$

The open problem that arises here is to show that the term $\|\,\widehat{f}_n - \widehat{f}_n^{\,0}\,\|$ approaches 0. In other words, one needs to show that the estimators of the original function f and the estimators of its approximation $\tilde{f}$ in $S((0,1])$ lie not too far away from each other.

Moreover, it is interesting to determine the rate of convergence of the estimators in this case. As discussed by Boysen *et al.* [8], the rate of convergence of the L^2 Potts minimizers in case of arbitrary $f \in L^2((0,1])$ is essentially different from the case when f is a step function. It is therefore intriguing to estimate the rate of convergence of the L^1 estimators in order to compare the performance of the L^2 and L^1 models.

Furthermore, our convergence rates do not seem optimal. These rates are less optimal than the corresponding rates in the L^2 case, implying that the convergence could be slower than in the L^2 case. Thus the question remains open about the choice of $(\gamma_n)_{n\in\mathbb{N}}$, which leads to the optimal rates of convergence.

The next open question is related to the results on consistency and convergence for other, non Pareto distributions of the noise.

Besides this, a very interesting point to pursue is the comparative study of the L^2 and the L^1 models in case of the noisy signal with outliers. Here, the analytical results on the rate of convergence can be numerically verified on numerous examples of real data and the comparison of both techniques can be visualized.

Finally, in order to generalize the theoretical framework to an even larger class of functions, one could attempt to develop the similar regression model for the class of the regression function, which are just measurable on $[0, 1)$.

Appendix A

Auxiliary Results

In this chapter we collect some auxiliary propositions and lemmas, which have been used in the main text without proof.

A.1 Convergence of the series of a particular type

Proposition 6. *Let $a > 0$ be an arbitrary real number and $k \geq 0$, $m \geq 0$, $l \geq 2$ are arbitrary natural numbers. Then it holds:*

$$\sum_{n=1}^{\infty} n^k \exp\Big\{ -\frac{a \cdot n}{(\ln n)^m} \Big\} < \infty \tag{A.1}$$

$$\sum_{n=1}^{\infty} n^k \exp\big\{ -a \cdot (\ln n)^l \big\} < \infty. \tag{A.2}$$

Proof of $(A.1)$ **:**

It is obvious that for all real $a > 0$ and natural $k \geq 0$, $m \geq 0$ it holds:

$$\sum_{n=1}^{\infty} n^k \exp\Big\{ -\frac{a \cdot n}{(\ln n)^m} \Big\} = \sum_{n=1}^{\infty} \exp\Big\{ -\frac{a \cdot n - k \cdot (\ln n)^{m+1}}{(\ln n)^m} \Big\} \tag{A.3}$$

Now for all real $a > 0$ and natural $k \geq 0$, $m \geq 0$ we consider the ratio of $\Big(\frac{a \cdot n - k \cdot (\ln n)^{m+1}}{(\ln n)^m}\Big)$ to $(2 \cdot \ln n)$:

$$\lim_{n\to\infty} \frac{a \cdot n - k \cdot (\ln n)^{m+1}}{2\,(\ln n)^m\,(\ln n)} = \lim_{n\to\infty} \frac{a \cdot n}{2\,(\ln n)^{m+1}} - \frac{k}{2} = \infty$$

From this it follows that for all real $a > 0$ and natural $k \geq 0$, $m \geq 0$ there exists finite natural number N_0 such that for all $n \geq N_0$ it holds:

$$\frac{a \cdot n - k \cdot (\ln n)^{m+1}}{(\ln n)^m} > 2 \cdot \ln n$$

and therefore

$$\exp\Big\{ -\frac{a \cdot n - k \cdot (\ln n)^{m+1}}{(\ln n)^m} \Big\} < \exp\big\{ -2 \cdot \ln n \big\} = \frac{1}{n^2} \tag{A.4}$$

According to (A.3) and (A.4) we conclude:

$$\sum_{n=1}^{\infty} n^k \exp\Big\{-\frac{a\cdot n}{(\ln n)^m}\Big\} < \sum_{n=1}^{N_0-1} n^k \exp\Big\{-\frac{a\cdot n}{(\ln n)^m}\Big\} + \sum_{n=N_0}^{\infty}\frac{1}{n^2} < \infty$$

Therefore, the series $\sum\limits_{n=1}^{\infty} n^k \exp\Big\{-\frac{a\cdot n}{(\ln n)^m}\Big\}$ converges for all real $a > 0$ and natural $k \geq 0$, $m \geq 0$. This means, we have proved the validity of $(A.1)$.

Proof of $(A.2)$ **:**
Similarly to the proof of $(A.1)$, for all real $a > 0$ and natural $k \geq 0$, $l \geq 2$ we consider the ratio of $(a\cdot(\ln n)^l - k\cdot \ln n)$ to $(2\cdot \ln n)$:

$$\lim_{n\to\infty}\frac{a\cdot(\ln n)^l - k\cdot\ln n}{2\ln n} = \lim_{n\to\infty} a\cdot(\ln n)^{l-1} - \frac{k}{2} = \infty$$

Hence for all real $a > 0$ and natural $k \geq 0$, $l \geq 2$ there exists a finite natural number N_0 such that for all $n \geq N_0$ it holds:

$$\exp\big\{-(a\cdot(\ln n)^l - k\cdot\ln n)\big\} < \exp\big\{-2\cdot\ln n\big\} = \frac{1}{n^2}.$$

From this we conclude:

$$\sum_{n=1}^{\infty} n^k \exp\big\{-a\cdot(\ln n)^m\big\} < \sum_{n=1}^{N_0-1}\exp\big\{-(a\cdot(\ln n)^l - k\cdot\ln n)\big\} + \sum_{n=N_0}^{\infty}\frac{1}{n^2} < \infty$$

Therefore $(A.2)$ is also valid and proof of Proposition 6 is complete.

□

Lemma 20. *Let* $C(n) = \sqrt{n}\ln n,\ n \in \mathbb{N}\ ,\ \alpha \geq 9,$

$$G^*_\alpha(x,n) = \exp\Big\{\frac{x}{2}\ln 2 - \frac{(\alpha-1)}{2}\big(x\ln C(n) - x(x-1)\big)\Big\} \tag{A.5}$$

and

$$l_1(n) = \frac{1}{2}\Big(\ln C(n) + 1 - \frac{\ln 2}{\alpha - 1}\Big). \tag{A.6}$$

Then it holds:

$$\sum_{n=N_0}^{\infty} n \int_1^{l_1(n)} G^*_\alpha(x,n)\,dx < \infty. \tag{A.7}$$

Proof:

First, for any $n \in \mathbb{N}$ we consider the integral $\int\limits_1^{l_1(n)} G^*_\alpha(x,n)\,dx$.

$$\int_1^{l_1(n)} G^*_\alpha(x,n)\,dx = \int_1^{l_1(n)} \exp\Big\{\frac{x}{2}\ln 2 - \frac{(\alpha-1)}{2}\big(x\ln C(n) - x(x-1)\big)\Big\}\,dx$$

$$= \int_1^{l_1(n)} \exp\Big\{x\Big(\frac{1}{2}\big(\ln 2 - (\alpha-1)\ln C(n) - (\alpha-1)\big)\Big) + \frac{\alpha-1}{2}x^2\Big\}\,dx.$$

If we introduce the following new notations:

$$a(n) := \frac{1}{2}\big(\ln 2 - (\alpha-1)\ln C(n) - (\alpha-1)\big) \quad \text{and} \quad b := \frac{\alpha-1}{2}\,, \tag{A.8}$$

then it is clear that

$$\int\limits_{1}^{l_1(n)} G^*_\alpha(x,n)\,dx = \int\limits_{1}^{l_1(n)} \exp\big\{a(n)\cdot x + bx^2\big\}\,dx. \tag{A.9}$$

From the definitions of $l_1(n)$, a and b it is obvious that

$$l_1(n) = -\frac{a(n)}{2b}. \tag{A.10}$$

Moreover, it is known (see [1, p. 303]) that for the error function erf, defined by

$$\operatorname{erf}(x) \quad := \quad \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt \tag{A.11}$$

it holds:

$$\int \exp\big\{ax + bx^2\big\}\,dx = \sqrt{\frac{\pi}{4(-b)}}\cdot e^{-\frac{a^2}{4b}}\cdot \operatorname{erf}\Big(\sqrt{-b}x - \frac{a}{2\sqrt{-b}}\Big). \tag{A.12}$$

It is also known that the function erf fulfills the following properties:

$$\operatorname{erf}(ix) \quad = \quad i\frac{2}{\sqrt{\pi}}\int_0^x e^{t^2}\,dt, \tag{A.13}$$

$$\operatorname{erf}(-x) \quad = \quad -\operatorname{erf}(x), \tag{A.14}$$

$$\operatorname{erf}(0) \quad = \quad 0. \tag{A.15}$$

Summarizing, we obtain :

$$\begin{aligned}
\int\limits_{1}^{l_1(n)} G^*_\alpha(x,n)\,dx &\underset{(A.9),(A.10)}{=} \int\limits_{1}^{l_1(n)} \exp\big\{-2\,b\,l_1(n)\,x + b\,x^2\big\}\,dx \\
&\underset{(A.12)}{=} \frac{1}{i}\sqrt{\frac{\pi}{4\,b}}\cdot\exp\big\{-b\,l_1(n)^2\big\}\Big[\operatorname{erf}\big(\underbrace{i\sqrt{b}\,l_1(n) - i\sqrt{b}\,l_1(n)}_{=0}\big) - \operatorname{erf}\big(i\sqrt{b} - i\sqrt{b}\,l_1(n)\big)\Big] \\
&\underset{(A.15),(A.8)}{=} -\frac{1}{i}\sqrt{\frac{\pi}{2(\alpha-1)}}\cdot\exp\Big\{-l_1(n)^2\cdot\frac{(\alpha-1)}{2}\Big\}\cdot\operatorname{erf}\Big(i\sqrt{\frac{\alpha-1}{2}}\cdot(1-l_1(n))\Big) \\
&\underset{(A.14)}{=} \frac{1}{i}\sqrt{\frac{\pi}{2(\alpha-1)}}\cdot\exp\Big\{-l_1(n)^2\cdot\frac{(\alpha-1)}{2}\Big\}\cdot\operatorname{erf}\Big(i\sqrt{\frac{\alpha-1}{2}}\cdot(l_1(n)-1)\Big) \\
&\underset{(A.13)}{=} \sqrt{\frac{2}{(\alpha-1)}}\cdot\exp\Big\{-l_1(n)^2\cdot\frac{(\alpha-1)}{2}\Big\}\int_0^{\sqrt{\frac{\alpha-1}{2}}\cdot(l_1(n)-1)} e^{t^2}\,dt\,.
\end{aligned} \tag{A.16}$$

The integral $F(x) := e^{-x^2} \int_0^x e^{t^2}\, dt$ is known as the Dawson's Integral (see [1]). Because of the following result:

$$\lim_{x\to\infty} \frac{\int_0^x e^{t^2}\, dt}{e^{x^2}/inparx} = \lim_{x\to\infty} \frac{x\int_0^x e^{t^2}\, dt}{e^{x^2}} = \Big(\frac{\infty}{\infty}\Big) = \lim_{x\to\infty} \frac{\int_0^x e^{t^2}\, dt + x\, e^{x^2}}{2\,x\,e^{x^2}}$$

$$= \lim_{x\to\infty} \frac{\int_0^x e^{t^2}\, dt}{2\,x\,e^{x^2}} + \frac{1}{2} = \Big(\frac{\infty}{\infty}\Big) = \lim_{x\to\infty} \frac{e^{x^2}}{2\,e^{x^2}(1+2\,x^2)} + \frac{1}{2} = \frac{1}{2},$$

we deduce that for a sufficiently large $|x|$ it holds:

$$\int_0^x e^{t^2}\, dt \le \frac{e^{x^2}}{x}. \tag{A.17}$$

Due to the facts that $\lim_{n\to\infty} l_1(n) = \lim_{n\to\infty} \frac{1}{2}\Big(\ln(\sqrt{n}\,\ln n) + 1 - \frac{\ln 2}{\alpha-1}\Big) = \infty$ and $\lim_{n\to\infty} \sqrt{\frac{\alpha-1}{2}} \cdot (l_1(n)-1) = \infty$, it is clear that there exists a natural number $N_1 \in \mathbb{N}$ such that for all $n \ge N_1$ we can use the result (A.17) for calculation of (A.16). This yields:

$$\int_1^{l_1(n)} G^*_\alpha(x,n)\, dx \le \sqrt{\frac{2}{(\alpha-1)}} \cdot \frac{\exp\Big\{-l_1(n)^2\cdot\frac{(\alpha-1)}{2} + \frac{\alpha-1}{2}\cdot(l_1(n)-1)^2\Big\}}{\sqrt{\frac{\alpha-1}{2}}\cdot(l_1(n)-1)}$$

$$= \frac{2\,\exp\Big\{-2\,l_1(n)\cdot\frac{(\alpha-1)}{2} + \frac{\alpha-1}{2}\Big\}}{(\alpha-1)\cdot(l_1(n)-1)} = \frac{2\,\exp\Big\{\frac{\alpha-1}{2}\Big\}\cdot\exp\{-l_1(n)\,(\alpha-1)\}}{(\alpha-1)\cdot(l_1(n)-1)}$$

$$\underset{(A.6)}{=} \frac{2\,\exp\Big\{\frac{\alpha-1}{2}\Big\}\cdot C(n)^{-\frac{\alpha-1}{2}}\cdot\exp\Big\{-\frac{\alpha-1}{2}\Big\}\cdot\sqrt{2}}{\frac{(\alpha-1)}{2}\ln C(n) - \frac{(\alpha-1)}{2} - \frac{\ln 2}{2}}$$

$$= \frac{2\sqrt{2}\,\big(\sqrt{n}\,\ln n\big)^{-\frac{\alpha-1}{2}}}{\frac{(\alpha-1)}{2}\ln\big(\sqrt{n}\,\ln n\big) - \frac{(\alpha-1)}{2} - \frac{\ln 2}{2}}$$

$$= \frac{\sqrt{2}\,n^{-\frac{\alpha-1}{4}}\,\big(\ln n\big)^{-\frac{\alpha-1}{2}}}{(\alpha-1)\Big(\frac{1}{2}\ln n + \ln(\ln n)\Big) - (\alpha-1) - \ln 2}. \tag{A.18}$$

Taking into account that $\lim_{n\to\infty} \ln(\ln n) = \infty$, it is obvious that there exist a natural number N_2 such that for all $n \ge N_2$ the following relations are valid:

$$(\alpha-1)\ln(\ln n) - (\alpha-1) - \ln 2 > 0$$

and

$$(\alpha-1)\Big(\frac{1}{2}\ln n + \ln(\ln n)\Big) - (\alpha-1) - \ln 2 > \frac{\alpha-1}{2}\ln n.$$

This together with (A.18) yields that for all $n \ge N_0 := \max(N_1, N_2)$ it holds:

$$\int_1^{l_1(n)} G^*_\alpha(x,n)\, dx < \frac{\sqrt{2}\,n^{-\frac{\alpha-1}{4}}\,\big(\ln n\big)^{-\frac{(\alpha-1)}{2}}}{\frac{(\alpha-1)}{2}\ln n} = \frac{2\sqrt{2}}{\alpha-1}\Big(n^{-\frac{\alpha-1}{4}}\cdot(\ln n)^{-\frac{\alpha+1}{2}}\Big).$$

From this it follows:

$$\sum_{n=N_0}^{\infty} n \int_1^{l_1(n)} G_\alpha^*(x,n)\,dx < \frac{2\sqrt{2}}{\alpha-1} \sum_{n=N_0}^{\infty} \frac{n}{n^{\frac{\alpha-1}{4}} \cdot (\ln n)^{\frac{\alpha+1}{2}}} = \frac{2\sqrt{2}}{\alpha-1} \sum_{n=N_0}^{\infty} \frac{1}{n^{\frac{\alpha-5}{4}} \cdot (\ln n)^{\frac{\alpha+1}{2}}}.$$

Due to the fact that the series $\sum\limits_{n=1}^{\infty} \frac{1}{n^\lambda \cdot (\ln n)^\beta}$ converges either for $\lambda > 1$ or for $\lambda = 1$ and $\beta > 1$, we deduce that for $\alpha \geq 9$ the sum $\sum\limits_{n=N_0}^{\infty} n \int\limits_1^{l_1(n)} G_\alpha^*(x,n)\,dx$ is finite and, therefore, the sum $\sum\limits_{n=1}^{\infty} n \int\limits_1^{l_1(n)} G_\alpha^*(x,n)\,dx$ is finite as well. This completes the proof of Lemma 20.

□

A.2 Some important inequalities

Lemma 21. *Let $\alpha \geq 1$ and*

$$k := \frac{1}{4}\Big(1 - \frac{1}{3^{\alpha-1}}\Big)^2.$$

Then for all $u \in (0,\, 2]$ it holds:

$$\frac{2(1+u)^{\alpha-1} - 1}{(1+u)^{2(\alpha-1)}} \leq (1 - ku^2).$$

Proof:
We are looking for such constant k that for all $u \in (0,\, 2]$ and $\alpha \geq 1$ it holds:

$$\frac{2(1+u)^{\alpha-1} - 1}{(1+u)^{2(\alpha-1)}} \leq (1 - ku^2).$$

The last inequality can be transformed as follows:

$$\begin{aligned}
& 2(1+u)^{\alpha-1} - 1 \leq (1-ku^2)\,(1+u)^{2(\alpha-1)} \\
\Leftrightarrow\quad & ku^2\,(1+u)^{2(\alpha-1)} \leq (1+u)^{2(\alpha-1)} - 2(1+u)^{\alpha-1} + 1 \\
\Leftrightarrow\quad & ku^2\,(1+u)^{2(\alpha-1)} \leq \big((1+u)^{(\alpha-1)} - 1\big)^2 \\
\Leftrightarrow\quad & k \leq \left(\frac{(1+u)^{(\alpha-1)} - 1}{u\,(1+u)^{(\alpha-1)}}\right)^2
\end{aligned} \tag{A.19}$$

We continue by computing $\min\limits_{u\in(0,\,2]} \frac{(1+u)^{(\alpha-1)} - 1}{u\,(1+u)^{(\alpha-1)}}$.

$$\begin{aligned}
& \frac{d}{du}\left(\frac{(1+u)^{(\alpha-1)} - 1}{u\,(1+u)^{(\alpha-1)}}\right) \\
= \quad & \frac{(\alpha-1)\,(1+u)^{(\alpha-2)}\,u\,(1+u)^{(\alpha-1)}}{u^2\,(1+u)^{2(\alpha-1)}}
\end{aligned}$$

$$
\begin{aligned}
& - \quad \frac{\left((1+u)^{(\alpha-1)}-1\right)\left((1+u)^{(\alpha-1)}+u\,(\alpha-1)\,(1+u)^{(\alpha-2)}\right)}{u^2\,(1+u)^{2(\alpha-1)}} \\
= & \quad \frac{(1+u)^{(\alpha-1)}-(1+u)^{(2\,\alpha-2)}+(\alpha-1)\,u\,(1+u)^{(\alpha-2)}}{u^2\,(1+u)^{2(\alpha-1)}} \\
= & \quad \frac{(1+u)^{(\alpha-2)}\cdot\left((1+u)-(1+u)^{\alpha}+(\alpha-1)\,u\right)}{u^2\,(1+u)^{2(\alpha-1)}} \\
= & \quad \frac{(1+\alpha\,u)-(1+u)^{\alpha}}{u^2\,(1+u)^{\alpha}}.
\end{aligned}
$$

From Bernoulli's inequality [9]:

$$(1+r\,x) \leq (1+x)^r,$$

which is valid for all $x > -1$, $r \leq 0$ or $r \geq 1$, we obtain that that for all $u \in (0,\,2]$ and for $\alpha \geq 1$ it holds:

$$\frac{d}{du}\left(\frac{(1+u)^{(\alpha-1)}-1}{u\,(1+u)^{(\alpha-1)}}\right) = \frac{(1+\alpha\,u)-(1+u)^{\alpha}}{u^2\,(1+u)^{\alpha}} \leq 0. \tag{A.20}$$

Consequently, for all $u \in \mathbb{R}$ and for $\alpha \geq 1$ the function $f(u) := \dfrac{(1+u)^{(\alpha-1)}-1}{u\,(1+u)^{(\alpha-1)}}$ is a monotonically decreasing function. Therefore

$$\min_{u\in(0,\,2]} \frac{(1+u)^{(\alpha-1)}-1}{u\,(1+u)^{(\alpha-1)}} = f(2) = \frac{1}{2}\left(1-\frac{1}{3^{\alpha-1}}\right).$$

Hence, according to (A.19), it is clear that for

$$k = \frac{1}{4}\left(1-\frac{1}{3^{\alpha-1}}\right)^2$$

and all $u \in (0,\,2]$ it holds:

$$\frac{2(1+u)^{\alpha-1}-1}{(1+u)^{2(\alpha-1)}} \leq (1-ku^2).$$

This means that Lemma 21 is proved.

□

Proposition 7. *For all $x \in \mathbb{R}$ it holds:*

$$(1+x) \leq e^x \tag{A.21}$$

Proof:
For all $x \in \mathbb{R}$ we consider the function $f(x) = e^x - (1+x)$.

$$f'(x) = e^x - 1$$

From this it is clear that

$$f'(x) = 0 \;\Leftrightarrow\; x = 0.$$

Moreover $f''(x) = e^x > 0$, this means that the function f is a convex function and

$$\min_{x\in\mathbb{R}} f(x) = f(0) = 0.$$

Consequently we deduce that $f(x) \geq 0$ for all $x \in \mathbb{R}$. In other words, for all $x \in \mathbb{R}$ it holds:

$$(1 + x) \leq e^x.$$

□

Lemma 22. *Let $X_1, \ldots, X_n$ be independent random variables such that $a_k \leq X_k \leq b_k$ for all $k = 1, \ldots, n$. Then for any $x > 0$ it holds:*

$$\mathbb{P}\left(\Big| \sum_{k=1}^{n} X_k - \mathbb{E}\Big(\sum_{k=1}^{n} X_k\Big) \Big| \geq n \cdot x \right) \leq 2\exp\left\{ -\frac{2\,n^2\,x^2}{\sum\limits_{k=1}^{n}(b_k - a_k)^2} \right\}. \tag{A.22}$$

Proof:
In order to prove inequality(A.22) we will use the following standard Hoeffding's inequality (see [33, p. 58]), which is valid for any $x > 0$ and for all independent random variables $X_1, \ldots, X_n$ such that for all $k = 1, \ldots, n$ it holds $a_k \leq X_k \leq b_k$:

$$\mathbb{P}\left(\sum_{k=1}^{n} X_k - \mathbb{E}\Big(\sum_{k=1}^{n} X_k\Big) \geq n \cdot x \right) \leq \exp\left\{ -\frac{2\,n^2\,x^2}{\sum\limits_{k=1}^{n}(b_k - a_k)^2} \right\}. \tag{A.23}$$

It is clear that the similar inequality can be written for the independent random variables $-X_1, \ldots, -X_n$, which are also bounded, namely $-b_k \leq -X_k \leq -a_k$, for all $k = 1, \ldots, n$. Thus it holds:

$$\mathbb{P}\left(-\sum_{k=1}^{n} X_k - \mathbb{E}\Big(-\sum_{k=1}^{n} X_k\Big) \geq n \cdot x \right) \leq \exp\left\{ -\frac{2\,n^2\,x^2}{\sum\limits_{k=1}^{n}(-a_k - (-b_k))^2} \right\}. \tag{A.24}$$

Now, combining the inequalities (A.23) and (A.24), we obtain:

$$\mathbb{P}\left(\Big| \sum_{k=1}^{n} X_k - \mathbb{E}\Big(\sum_{k=1}^{n} X_k\Big) \Big| \geq n \cdot x \right) \leq 2\exp\left\{ -\frac{2\,n^2\,x^2}{\sum\limits_{k=1}^{n}(b_k - a_k)^2} \right\},$$

which completes the proof of Lemma 22.

□

Lemma 23. *Let X be a binomial distributed random variable with parameter (l, p), i.e. $X \sim Bin\big(l,\, p\big)$. Then for any m with $0 \leq m < l$ it holds:*

$$\mathbb{P}\Big(X \geq m \Big) \leq \frac{l^l}{m^m \cdot (l-m)^{l-m}}\, p^m\, (1-p)^{l-m}. \tag{A.25}$$

In particular, it holds:

$$\mathbb{P}\Big(X \geq \frac{l}{2} \Big) \leq 2^l \big((1-p)\,p\big)^{l/2}. \tag{A.26}$$

Proof:
This lemma is a particular case of Theorem 1, which is contained in Hoeffding's work [21, p. 15]. According to that theorem, for any sequence $(X_1, X_2, \dots, X_l)$ of independent random variables and any $0 < t < 1 - m$ it holds:

$$\mathbb{P}(\overline{X} - \mu \geq t) = \mathbb{P}(\overline{X} \geq \mu + t) \leq \left\{ \Big(\frac{\mu}{\mu + t}\Big)^{\mu+t} \Big(\frac{1-\mu}{1-\mu-t}\Big)^{1-\mu-t} \right\}^{l}, \tag{A.27}$$

where $\overline{X} = \dfrac{X_1 + \ldots + X_l}{l}, \quad \mu = \mathbb{E}\overline{X} = \dfrac{\mathbb{E}(X_1 + \ldots + X_l)}{l}$.

Ii is clear that if $X \sim Bin\big(l,\, p\big)$ then $X = X_1 + \ldots + X_l$, where $X_1, \dots, X_l$ are independent and $X_i \sim Bin\big(1,\, p\big)\ \forall\, i \in \{1, \dots, l\}$. Moreover, it is obvious that in this case, $\mu = p$.
Now for $0 \leq m < l$ we consider $\mathbb{P}(X \geq m)$:

$$\mathbb{P}\big(X \geq m\big) = \mathbb{P}\Big(\overline{X} \geq \frac{m}{l}\Big) \underset{(A.27)}{\leq} \frac{l^{\,l}}{m^m \cdot (l-m)^{l-m}}\, p^m\, (1-p)^{l-m}.$$

In particular for $m = \dfrac{l}{2}$ we have:

$$\mathbb{P}\Big(X \geq l/2\Big) = \mathbb{P}\Big(\overline{X} \geq 1/2\Big) \leq 2^l \big((1-p)\, p\big)^{l/2}.$$

This completes the proof of Lemma 23. □

Lemma 24. *Let η be an arbitrary continuous random variable with the distribution function F_η and $\mathbb{P}(\eta \leq 0) = \mathbb{P}(\eta \geq 0) = \frac{1}{2}$. Then for any finite $\Delta > 0$ and finite $\beta > 0$ there exists a number $\chi = \chi(\Delta,\, \beta) > 0$ and $\chi < \Delta$ such that it holds*

$$\beta \cdot (\, 1 - 2\, P_\eta^{\chi}\,) + (\, 1 - 2\, P_\eta^{\chi - \Delta}\,) > 0, \tag{A.28}$$

where $P_\eta^{\chi} := \mathbb{P}(\eta > \chi)$ for any real χ and any random η.

Proof:
For any $\Delta > 0$ and any $\beta > 0$ inequality (A.28) can be rewritten in the following equivalent forms:

$$\begin{aligned}
& \beta \cdot (\, 1 - 2\,(1 - F_\eta(\chi))\,) + (\, 1 - 2\,(1 - F_\eta(\chi - \Delta))\,) > 0 \\
\Leftrightarrow \quad & \beta \cdot (\, 2\, F_\eta(\chi) - 1\,) + (2\, F_\eta(\chi - \Delta) - 1) > 0 \\
\Leftrightarrow \quad & 2\,(\beta\, F_\eta(\chi) + F_\eta(\chi - \Delta)) - (\beta + 1) > 0 \\
\Leftrightarrow \quad & \beta\, F_\eta(\chi) + F_\eta(\chi - \Delta) > \frac{(\beta + 1)}{2}
\end{aligned} \tag{A.29}$$

Now we consider $\beta\, F_\eta(\chi) + F_\eta(\chi - \Delta)$.
$\beta\, F_\eta(\chi) + F_\eta(\chi - \Delta)$ is a continuous and monotonically increasing function as a function of χ, since

$$\frac{\partial}{\partial \chi}\big(\beta\, F_\eta(\chi) + F_\eta(\chi - \Delta)\big) = \beta f_\eta(\chi) + f_\eta(\chi - \Delta) \geq 0.$$

Moreover, for $\chi = 0$ and for $\chi = \Delta$ it holds:

$$\begin{aligned} \beta\, F_\eta(0) + F_\eta(-\Delta) &< \frac{\beta}{2} + \frac{1}{2} \\ \beta\, F_\eta(\Delta) + F_\eta(0) &> \frac{\beta}{2} + \frac{1}{2} \end{aligned}$$

From this it follows that there exists a unique number $0 < \chi_0 < \Delta$ such that

$$\beta\, F_\eta(\chi_0) + F_\eta(\chi_0 - \Delta) = \frac{(\beta + 1)}{2}.$$

Thus for all $\chi > \chi_0$ and in particular for all $\chi \in (\chi_0,\, \Delta)$ inequality(A.29) and respectively inequality(A.28) are valid. Therefore we have completed the proof of Lemma 24.

□

Lemma 25. *For any real numbers $\eta_1, \ldots, \eta_n$ and $x_1, \ldots, x_n$ it holds:*

$$|\operatorname{med}(\eta_1 + x_1, \ldots, \eta_n + x_n)| \le |\operatorname{med}(\eta_1, \ldots, \eta_n)| + \max_{1 \le i \le n} |x_i| \tag{A.30}$$

Proof:
It is not difficult to see that from componentwise monotonicity of the median it holds:

$$\begin{aligned} & \operatorname{med}\big(\eta_1 - \max_{1\le i\le n} |x_i|, \ldots, \eta_n - \max_{1\le i\le n} |x_i|\big) \le \operatorname{med}\big(\eta_1 + x_1, \ldots, \eta_n + x_n\big) \\ \le\; & \operatorname{med}(\eta_1 + \max_{1\le i\le n} |x_i| \ldots, \eta_n + \max_{1\le i\le n} |x_i|\big). \end{aligned}$$

From this it follows:

$$\begin{aligned} - \quad & |\operatorname{med}(\eta_1, \ldots, \eta_n)| - \max_{1\le i\le n} |x_i| \le \operatorname{med}(\eta_1, \ldots, \eta_n) - \max_{1\le i\le n} |x_i| \\ \le \quad & \operatorname{med}(\eta_1 + x_1, \ldots, \eta_n + x_n) \le \operatorname{med}(\eta_1, \ldots, \eta_n) + \max_{1\le i\le n} |x_i| \\ \le \quad & |\operatorname{med}(\eta_1, \ldots, \eta_n)| + \max_{1\le i\le n} |x_i|. \end{aligned}$$

Consequently we deduce:

$$|\operatorname{med}(\eta_1 + x_1, \ldots, \eta_n + x_n)| \le |\operatorname{med}(\eta_1 \ldots, \eta_n)| + \max_{1\le i\le n} |x_i|.$$

This completes the proof of Lemma 25.

□

A.3 Some properties of the indicator function

Lemma 26. *For all $a, b, c \in \mathbb{R}$, $c < \infty$ the function $f_0 : (a, b, c) \longrightarrow c \cdot \mathbf{1}_{[a,b)}$ is continuous with respect to the L^1 norm.*

Proof:
According to the definition of a continuous function for all $a, b \in \mathbb{R}$ and $c < \infty$ we have to show that:

$$\lim_{a\to a_0,\, b\to b_0,\, c\to c_0} f_0(a, b, c) = f_0(a_0, b_0, c_0).$$

Let a, b, c, a_0, b_0, c_0 be some real numbers such that $a \to a_0,\ b \to b_0,\ c \to c_0\ (c < \infty)$. Consider $\|f_0(a,b,c) - f_0(a_0,b_0,c_0)\|$:

$$\|f_0(a,b,c) - f_0(a_0,b_0,c_0)\| = \|c \cdot \mathbf{1}_{[a,b)} - c_0 \cdot \mathbf{1}_{[a_0,b_0)}\| = \int\limits_{[0,1)} |c \cdot \mathbf{1}_{[a,b)} - c_0 \cdot \mathbf{1}_{[a_0,b_0)}|\, dx$$

$$= \begin{cases} \underbrace{\int_a^{a_0} |c|\, dx}_{(a\to a_0)\ \longrightarrow 0} + \underbrace{\int_b^{b_0} |c_0|\, dx}_{(b\to b_0)\ \longrightarrow 0} + \int_{a_0}^{b} \underbrace{|c - c_0|}_{(c\to c_0)\ \longrightarrow 0}\, dx & : \ a \le a_0 < b \le b_0 \\ \underbrace{\int_{a_0}^{a} |c_0|\, dx}_{(a\to a_0)\ \longrightarrow 0} + \underbrace{\int_{b_0}^{b} |c|\, dx}_{(b\to b_0)\ \longrightarrow 0} + \int_{a}^{b_0} \underbrace{|c - c_0|}_{(c\to c_0)\ \longrightarrow 0}\, dx & : \ a_0 \le a < b_0 \le b \\ \underbrace{\int_a^{a_0} |c|\, dx}_{(a\to a_0)\ \longrightarrow 0} + \underbrace{\int_{b_0}^{b} |c|\, dx}_{(b\to b_0)\ \longrightarrow 0} + \int_{a_0}^{b_0} \underbrace{|c - c_0|}_{(c\to c_0)\ \longrightarrow 0}\, dx & : \ a \le a_0 < b_0 \le b \\ \underbrace{\int_{a_0}^{a} |c_0|\, dx}_{(a\to a_0)\ \longrightarrow 0} + \underbrace{\int_b^{b_0} |c_0|\, dx}_{(b\to b_0)\ \longrightarrow 0} + \int_{a}^{b} \underbrace{|c - c_0|}_{(c\to c_0)\ \longrightarrow 0}\, dx & : \ a_0 \le a < b \le b_0 \end{cases}$$

$$\Rightarrow\ \forall\, a,b,c \in \mathbb{R},\ \ c < \infty \qquad \lim_{a\to a_0,\, b\to b_0,\, c\to c_0} \|f_0(a,b,c) - f_0(a_0,b_0,c_0)\| = 0,$$

i.e. the function f_0 is indeed a continuous function. This completes the proof of Lemma 26. □

Lemma 27. *Let* $g_n \xrightarrow[n\to\infty]{L^0([0,1))} g$. *Then for any interval* $(s,t) \subseteq [0,1)$ *it holds:*

$$g_n \cdot \mathbf{1}_{(s,t)} \xrightarrow[n\to\infty]{L^0([0,1))} g \cdot \mathbf{1}_{(s,t)}. \tag{A.31}$$

Proof:
To prove Lemma 27 we need to show that from the statement that for any $\varepsilon > 0$

$$l\Big(\, \{\, x \in [0,1)\, :\, |g_n(x) - g(x)| > \varepsilon \,\} \Big) \underset{n\to\infty}{\longrightarrow} 0,$$

it follows that for any $\varepsilon > 0$ and any interval $(s,t) \subseteq [0,1)$

$$l\Big(\, \{\, x \in [0,1)\, :\, |g_n(x) \cdot \mathbf{1}_{(s,t)}(x) - g(x) \cdot \mathbf{1}_{(s,t)}(x)| > \varepsilon \,\} \Big) \underset{n\to\infty}{\longrightarrow} 0.$$

For an arbitrary $\varepsilon > 0$ and an arbitrary interval $(s,t) \subseteq [0,1)$ we consider the left hand side of the last equation :

$$l\Big(\, \{\, x \in [0,1)\, :\, |g_n(x) \cdot \mathbf{1}_{(s,t)}(x) - g(x) \cdot \mathbf{1}_{(s,t)}(x)| > \varepsilon \,\} \Big)$$

$$
\begin{aligned}
&= l\Big(\big\{\, x \in [0,1) \,:\, |g_n(x) - g(x)| \cdot \mathbf{1}_{(s,\,t)}(x) > \varepsilon \,\big\}\Big) \\
&= l\Big(\big\{\, x \in (s,t) \,:\, |g_n(x) - g(x)| > \varepsilon \,\big\}\Big) \\
&\leq l\Big(\big\{\, x \in [0,1) \,:\, |g_n(x) - g(x)| > \varepsilon \,\big\}\Big) \underset{n\to\infty}{\longrightarrow} 0
\end{aligned}
$$

This proves Lemma 27.

□

A.4 Relationship between a function and its mean

Lemma 28. *Suppose that $f \in L^1\big([0,1)\big)$ and for any interval $I \subseteq [0,1)$ of the length $l(I)$ let $\overline{f}_I := \frac{1}{l(I)} \int_I f(x)\,dx$. Then for any $\varepsilon > 0$ there exists such $\delta > 0$ that for all partitions P of the interval $[0,1]$ it holds*

$$
\sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta} \|\overline{f}_I \cdot \mathbf{1}_I - f \cdot \mathbf{1}_I\| < \varepsilon. \tag{A.32}
$$

Proof:

We define a set $\mathcal{H}_0$ as follows:

$$
\mathcal{H}_0 = \Big\{ f \in L^1\big([0,1)\big) \;\mid\; \forall\,\varepsilon > 0 \;\; \exists\,\delta > 0 : \sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta} \|\overline{f}_I \cdot \mathbf{1}_I - f \cdot \mathbf{1}_I\| < \varepsilon \Big\}.
$$

To accomplish the proof we need to show the validity of the following statements:

(i) $\mathcal{H}_0$ is a linear subspace of $L^1\big([0,1)\big)$.

(ii) $\mathcal{H}_0$ is closed.

(iii) $\mathcal{H}_0$ is dense in $L^1\big([0,1)\big)$.

Proof of **(i)** :
In order to prove that $\mathcal{H}_0$ is a linear subspace of $L^1\big([0,1)\big)$ we need to show that:

1) $\forall f_1,\ f_2 \in \mathcal{H}_0$ it holds: $f_1 + f_2 \in \mathcal{H}_0$;

2) $\forall \alpha \in \mathbb{R},\ \forall f \in \mathcal{H}_0$ it holds: $\alpha \cdot f \in \mathcal{H}_0$.

1): Let $f_1 \in \mathcal{H}_0$ and $f_2 \in \mathcal{H}_0$, i.e.

$$
\forall \varepsilon > 0 \;\exists\, \delta_1 : \quad \sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta_1} \|(\overline{f_1}_I - f_1) \cdot \mathbf{1}_I\| < \frac{\varepsilon}{2} \tag{A.33}
$$

and

$$
\forall \varepsilon > 0 \;\exists\, \delta_2 : \quad \sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta_2} \|(\overline{f_2}_I - f_2) \cdot \mathbf{1}_I\| < \frac{\varepsilon}{2}. \tag{A.34}
$$

Suppose that $g := f_1 + f_2$. We have to show that $g \in \mathcal{H}_0$, i.e.

$$\forall \varepsilon > 0 \, \exists \, \delta : \quad \sup_{P \in \mathcal{P}} \sum_{I \in P, \, l(I) < \delta} \|(\overline{g}_I - g) \cdot \mathbf{1}_I\| < \varepsilon. \tag{A.35}$$

For any interval $I \subset [0,1)$ consider $\|(\overline{g}_I - g) \cdot \mathbf{1}_I\|$:

$$\|(\overline{g}_I - g) \cdot \mathbf{1}_I\| = \int_I |\overline{g}_I - g(x)| \, dx = \int_I \left| \overline{f_1}_I + \overline{f_2}_I - f_1(x) - f_2(x) \right| dx$$

$$\leq \int_I \left| \overline{f_1}_I - f_1(x) \right| dx + \int_I \left| \overline{f_2}_I - f_2(x) \right| dx = \|(\overline{f_1}_I - f_1) \cdot \mathbf{1}_I\| + \|(\overline{f_2}_I - f_2) \cdot \mathbf{1}_I\|.$$

Consequently, $\forall I \subset [0,1) \quad \|(\overline{g}_I - g) \cdot \mathbf{1}_I\| \leq \|(\overline{f_1}_I - f_1) \cdot \mathbf{1}_I\| + \|(\overline{f_2}_I - f_2) \cdot \mathbf{1}_I\|$.

Therefore, for all partitions P of the interval $[0,1]$ and for any $\delta > 0$ it holds:

$$\sum_{I \in P, \, l(I) < \delta} \|(\overline{g}_I - g) \cdot \mathbf{1}_I\| \leq \sum_{I \in P, \, l(I) < \delta} \|(\overline{f_1}_I - f_1) \cdot \mathbf{1}_I\| + \sum_{I \in P, \, l(I) < \delta} \|(\overline{f_2}_I - f_2) \cdot \mathbf{1}_I\|.$$

We choose $\delta := \min\{\delta_1, \delta_2\}$ and according to the conditions (A.33) and (A.34) obtain that for all partitions P of the interval $[0,1]$ it holds

$$\sum_{I \in P, \, l(I) < \delta} \|(\overline{g}_I - g) \cdot \mathbf{1}_I\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Because this condition is satisfied for any partition P of the interval $[0,1]$, we conclude that:

$$\forall \varepsilon > 0 \, \exists \, \delta : \quad \sup_{P \in \mathcal{P}} \sum_{I \in P, \, l(I) < \delta} \|(\overline{g}_I - g) \cdot \mathbf{1}_I\| < \varepsilon \quad \text{for } \delta := \min\{\delta_1, \delta_2\}.$$

Thus the proof of (A.35) is complete. **2):** For any $\alpha \in \mathbb{R}$ and $f \in \mathcal{H}_0$ we have to show that $\alpha f \in \mathcal{H}_0$, i.e.

$$\forall \varepsilon > 0 \, \exists \, \delta : \quad \sup_{P \in \mathcal{P}} \sum_{I \in P, \, l(I) < \delta} \|(\overline{\alpha f}_I - \alpha f) \cdot \mathbf{1}_I\| < \varepsilon. \tag{A.36}$$

Let α be an arbitrary real number and $f \in \mathcal{H}_0$. From this it follows:

$$\forall \, \varepsilon > 0 \;\; \exists \, \delta : \quad \sup_{P \in \mathcal{P}} \sum_{I \in P, \, l(I) < \delta} \|(\overline{f}_I - f) \cdot \mathbf{1}_I\| < \frac{\varepsilon}{|\alpha|}. \tag{A.37}$$

Using the properties of the norm and of the mean value of a function we obtain that for any $I \subset [0,1)$ it holds:

$$\|(\overline{\alpha f}_I - \alpha f) \cdot \mathbf{1}_I\| = |\alpha| \cdot \|(\overline{f}_I - f) \cdot \mathbf{1}_I\|.$$

Consequently for all partitions P of the interval $[0,1]$ and for δ from (A.37) we obtain:

$$\sum_{I \in P, \, l(I) < \delta} \|(\overline{\alpha f}_I - \alpha f) \cdot \mathbf{1}_I\| = |\alpha| \sum_{I \in P, \, l(I) < \delta} \|(\overline{f}_I - f) \cdot \mathbf{1}_I\| < |\alpha| \cdot \frac{\varepsilon}{|\alpha|} = \varepsilon.$$

This inequality is valid for all partitions P of the interval $[0,1]$, therefore we conclude that:

$$\sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta} \|(\overline{\alpha f}_I - \alpha f)\cdot \mathbf{1}_I\| < \varepsilon.$$

This completes the proof of (A.36) and therefore of statement **(i)**.

Proof of **(ii)** :
According to the definition of a closed set, a set is closed if it contains all of its accumulation points [3, p. 45].
Suppose that f^0 is an accumulation point of the set $\mathcal{H}_0$, i.e. there exists a sequence of functions $\{f_n\}_{n=1}^{\infty}$ such that $\forall n\in\mathbb{N}\ f_n\in\mathcal{H}_0$ and

$$\forall\varepsilon>0\ \ \exists\, n_0:\ \forall n\geq n_0\ \ \|f_n - f^0\| < \frac{\varepsilon}{3}. \tag{A.38}$$

In order to prove **(ii)** we have to show that $f^0\in\mathcal{H}_0$, i.e.

$$\forall\varepsilon>0\ \ \exists\,\delta_0:\ \ \forall\,\delta<\delta_0\ \ \sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta} \|(\overline{f^0}_I - f^0)\cdot \mathbf{1}_I\| < \varepsilon. \tag{A.39}$$

Let us first recall all the relations that we have used so far.

The sequence $\{f_n\}_{n=1}^{\infty}$ is a sequence of functions from $\mathcal{H}_0$, i.e. for each function f_n we have:

$$\forall\varepsilon>0\ \ \exists\,\delta':\ \ \forall\,\delta<\delta'\ \ \sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta} \|(\overline{f_n}_I - f_n)\cdot \mathbf{1}_I\| < \frac{\varepsilon}{3}. \tag{A.40}$$

Further on for all functions f and g from $L^1([0,1])$ and for any interval $I\subset[0,1)$ we consider $\big\|\overline{f}_I\cdot\mathbf{1}_I - \overline{g}_I\cdot\mathbf{1}_I\big\|$:

$$\begin{aligned}
\big\|\overline{f}_I\cdot\mathbf{1}_I - \overline{g}_I\cdot\mathbf{1}_I\big\| &= \int_I \big|\overline{f}_I - \overline{g}_I\big|\,dt = \int_I \left|\frac{1}{l(I)}\int_I f(x)\,dx - \frac{1}{l(I)}\int_I g(x)\,dx\right| dt \\
&= \left|\int_I \Big(f(x)-g(x)\Big)\,dx\right|\cdot\frac{1}{l(I)}\int_I dt = \left|\int_I \Big(f(x)-g(x)\Big)\,dx\right|\cdot\frac{1}{l(I)}\cdot l(I) \\
&\leq \int_I \big|f(x)-g(x)\big|\,dx = \big\|f\cdot\mathbf{1}_I - g\cdot\mathbf{1}_I\big\|.
\end{aligned}$$

Thus we have shown that for all functions f and g from $L^1([0,1])$ and for any interval $I\subset[0,1)$ it holds:

$$\big\|\overline{f}_I\cdot\mathbf{1}_I - \overline{g}_I\cdot\mathbf{1}_I\big\| \leq \big\|f\cdot\mathbf{1}_I - g\cdot\mathbf{1}_I\big\|. \tag{A.41}$$

From this result we can deduce that for each $\delta>0$, in particular for each $\delta<\delta'$, we have:

$$\sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta} \big\|\overline{f}_I\cdot\mathbf{1}_I - \overline{g}_I\cdot\mathbf{1}_I\big\| \leq \sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta} \big\|f\cdot\mathbf{1}_I - g\cdot\mathbf{1}_I\big\|. \tag{A.42}$$

Our next goal is to proof the following statement:

$$\forall\varepsilon>0\ \ \text{and}\ \ \forall\,\delta<\delta'\ \ \sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta} \big\|\overline{f_n}_I\cdot\mathbf{1}_I - \overline{f^0}_I\cdot\mathbf{1}_I\big\| < \frac{\varepsilon}{3}. \tag{A.43}$$

From (A.38) we obtain that for any $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ the following chain of relationships is valid:

$$\int\limits_{[0,1]} \left|f_n(x) - f^0(x)\right| dx < \frac{\varepsilon}{3} \;\Rightarrow\; \sup_{P \in \mathcal{P}} \sum_{I \in P} \int\limits_I \left|f_n(x) - f^0(x)\right| dx < \frac{\varepsilon}{3} \;\Rightarrow\; \sup_{P \in \mathcal{P}} \sum_{I \in P} \left\|(f_n - f^0) \cdot \mathbf{1}_I\right\| < \frac{\varepsilon}{3}.$$

Since for all $\delta > 0$ it holds

$$\sup_{P \in \mathcal{P}} \sum_{I \in P,\, l(I) < \delta} \left\|f_n \cdot \mathbf{1}_I - f^0 \cdot \mathbf{1}_I\right\| \leq \sup_{P \in \mathcal{P}} \sum_{I \in P} \left\|f_n \cdot \mathbf{1}_I - f^0 \cdot \mathbf{1}_I\right\|,$$

we conclude:

$$\forall \varepsilon > 0 \quad \text{and} \quad \forall\, \delta < \delta' \quad \sup_{P \in \mathcal{P}} \sum_{I \in P,\, l(I) < \delta} \left\|f_n \cdot \mathbf{1}_I - f^0 \cdot \mathbf{1}_I\right\| < \frac{\varepsilon}{3}. \tag{A.44}$$

Using (A.42) we deduce that (A.43) is satisfied.
Finally we show the validity of (A.39).

$$\begin{aligned}
& \sup_{P \in \mathcal{P}} \sum_{I \in P,\, l(I) < \delta} \left\|\overline{f^0}_I \cdot \mathbf{1}_I - f^0 \cdot \mathbf{1}_I\right\| \\
= & \sup_{P \in \mathcal{P}} \sum_{I \in P,\, l(I) < \delta} \left\|\overline{f^0}_I \cdot \mathbf{1}_I - f^0 \cdot \mathbf{1}_I + f_n \cdot \mathbf{1}_I - f_n \cdot \mathbf{1}_I + \overline{f_n}_I \cdot \mathbf{1}_I - \overline{f_n}_I \cdot \mathbf{1}_I\right\| \\
\leq & \sup_{P \in \mathcal{P}} \Bigg(\sum_{I \in P,\, l(I) < \delta} \left\|\overline{f^0}_I \cdot \mathbf{1}_I - \overline{f_n}_I \cdot \mathbf{1}_I\right\| + \sum_{I \in P,\, l(I) < \delta} \left\|f_n \cdot \mathbf{1}_I - f^0 \cdot \mathbf{1}_I\right\| \\
& \qquad + \sum_{I \in P,\, l(I) < \delta} \left\|\overline{f_n}_I \cdot \mathbf{1}_I - f_n \cdot \mathbf{1}_I\right\| \Bigg) \\
\leq & \sup_{P \in \mathcal{P}} \sum_{I \in P,\, l(I) < \delta} \left\|\overline{f^0}_I \cdot \mathbf{1}_I - \overline{f_n}_I \cdot \mathbf{1}_I\right\| + \sup_{P \in \mathcal{P}} \sum_{I \in P,\, l(I) < \delta} \left\|f_n \cdot \mathbf{1}_I - f^0 \cdot \mathbf{1}_I\right\| \\
& + \sup_{P \in \mathcal{P}} \sum_{I \in P,\, l(I) < \delta} \left\|\overline{f_n}_I \cdot \mathbf{1}_I - f_n \cdot \mathbf{1}_I\right\| \underset{(A.43),\,(A.44),\,(A.40)}{<} \varepsilon.
\end{aligned}$$

Thus for all $\varepsilon > 0$ we have found such $\delta_0 = \delta'$ that for any $\delta < \delta_0$ the following inequality is valid:

$$\sup_{P \in \mathcal{P}} \sum_{I \in P,\, l(I) < \delta} \left\|\overline{f^0}_I \cdot \mathbf{1}_I - f^0 \cdot \mathbf{1}_I\right\| < \varepsilon.$$

Consequently we have shown the correctness of (A.39) and therefore of the fact that $\mathcal{H}_0$ is closed. Thus statement **(ii)** is proved.

Proof of **(iii)** :
In order to prove that $\mathcal{H}_0$ is dense in $L^1\big([0,1)\big)$ it is sufficient to show that $\mathcal{H}_0$ is dense in $S\big([0,1)\big)$ since $\mathcal{H}_0$ is closed and the set $S\big([0,1)\big)$ is dense in $L^1\big([0,1)\big)$ [27]. For this purpose we show firstly that an arbitrary indicator function belongs to $\mathcal{H}_0$.
Let $[a,b)$ be an arbitrary interval from $[0,1)$ and $f = \mathbf{1}_{[a,b)}$. We have to prove that $f \in \mathcal{H}_0$, i.e.

$$\forall \varepsilon > 0 \quad \exists \delta_0 : \quad \sup_{P \in \mathcal{P}} \sum_{I \in P,\, l(I) < \delta_0} \left\|\big(\overline{\mathbf{1}}_{[a,b)_I} - \mathbf{1}_{[a,b)}\big) \cdot \mathbf{1}_I\right\| < \varepsilon. \tag{A.45}$$

We fix an arbitrary ε and choose $\delta = \frac{\varepsilon}{4}$. Further on, we denote the intervals $I \subset (0,1]$ with $l(I) < \delta$ as *short* intervals and the intervals $I \subset (0,1]$ with $l(I) \geq \delta$ as *long* intervals.

Now consider some partition $P := \{I_1, \ldots, I_k, \ldots, I_m, \ldots, I_N\}$ of the interval $[0,1)$.
Suppose $a \in I_k$, $b \in I_m$ for some k and m from $\{1, \ldots, N\}$. It is clear that, if either $I_k = I_m$, or the both intervals I_k and I_m are long, i.e. $l(I_k) \geq \delta$ and $l(I_m) \geq \delta$, then for all short intervals I from P, i.e. for all $I \in P$ with $l(I) < \delta$, it holds: $\overline{\mathbf{1}}_{[a,b)_I} = \mathbf{1}_{[a,b)} \cdot \mathbf{1}_I$. If one of the two intervals is long and one is short, then there is only one interval I from P with $l(I) < \delta$, so that $\overline{\mathbf{1}}_{[a,b)_I} \neq \mathbf{1}_{[a,b)} \cdot \mathbf{1}_I$. Finally, if $l(I_k) < \delta$ and $l(I_m) < \delta$ then there exist two short intervals of partition P, where the mean $\overline{\mathbf{1}}_{[a,b)_I}$ and the function $\mathbf{1}_{[a,b)} \cdot \mathbf{1}_I$ are different. These intervals are I_k and I_m.
Consequently for the partition P of the interval $[0,1)$ it is valid:

$$\sum_{I \in P,\, l(I) < \delta} \left\| (\overline{\mathbf{1}}_{[a,b)_I} - \mathbf{1}_{[a,b)}) \cdot \mathbf{1}_I \right\| \leq \left\| (\overline{\mathbf{1}}_{[a,b)_{I_k}} - \mathbf{1}_{[a,b)}) \cdot \mathbf{1}_{I_k} \right\| + \left\| (\overline{\mathbf{1}}_{[a,b)_{I_m}} - \mathbf{1}_{[a,b)}) \cdot \mathbf{1}_{I_m} \right\|, \quad \text{(A.46)}$$

where $l(I_k) < \delta$ and $l(I_m) < \delta$.

First we compute $\left\| (\overline{\mathbf{1}}_{[a,b)_{I_k}} - \mathbf{1}_{[a,b)}) \cdot \mathbf{1}_{I_k} \right\|$.

Let $I_k = [a_k, b_k),\ 0 \leq a_k \leq a < b_k < b < 1$.Consider $\int_{I_k} \mathbf{1}_{[a,b)}(x)\, dx$:

$$\int_{I_k} \mathbf{1}_{[a,b)}\, dx = \int_{[a_k,b_k)} \mathbf{1}_{[a,b]}\, dx = \int_{[a_k,a)} \mathbf{1}_{[a,b)}\, dx + \int_{[a,b_k)} \mathbf{1}_{[a,b)}\, dx = \int_{[a,b_k)} dx = l\big([a,b_k)\big)$$

From this it follows:

$$\begin{aligned}
\left\| (\overline{\mathbf{1}}_{[a,b)_{I_k}} - \mathbf{1}_{[a,b)}) \cdot \mathbf{1}_{I_k} \right\| &= \int_{I_k} \left| \Big(\frac{1}{l(I_k)} \int_{I_k} \mathbf{1}_{[a,b)}(x)\, dx \Big) - \mathbf{1}_{[a,b)}(x) \right| dx = \\
= \int_{[a_k,b_k)} \left| \frac{l\big([a,b_k)\big)}{l\big([a_k,b_k)\big)} - \mathbf{1}_{[a,b)}(x) \right| dx &= \int_{[a_k,a)} \left| \frac{l\big([a,b_k)\big)}{l\big([a_k,b_k)\big)} - 0 \right| dx + \int_{[a,b_k)} \left| \frac{l\big([a,b_k)\big)}{l\big([a_k,b_k)\big)} - 1 \right| dx \\
&= \int_{[a_k,a)} \frac{l\big([a,b_k)\big)}{l\big([a_k,b_k)\big)}\, dx + \int_{[a,b_k)} \left| \frac{l\big([a,b_k)\big) - l\big([a_k,b_k)\big)}{l\big([a_k,b_k)\big)} \right| dx \\
&= \frac{l\big([a,b_k)\big) \cdot l\big([a_k,a)\big)}{l\big([a_k,b_k)\big)} + \int_{[a,b_k)} \frac{l\big([a_k,a)\big)}{l\big([a_k,b_k)\big)}\, dx \\
&= 2\frac{l\big([a,b_k)\big) \cdot l\big([a_k,a)\big)}{l\big([a_k,b_k)\big)} \leq 2\frac{l\big([a_k,b_k)\big) \cdot l\big([a_k,b_k)\big)}{l\big([a_k,b_k)\big)} = 2\, l\big([a_k,b_k)\big) = 2\, l(I_k).
\end{aligned}$$

Since $l(I_k) < \delta$, we obtain the following result:

$$\forall a \in I_k,\ \ b \notin I_k: \quad \left\| (\overline{\mathbf{1}}_{[a,b)_{I_k}} - \mathbf{1}_{[a,b)}) \cdot \mathbf{1}_{I_k} \right\| \leq 2\, l(I_k) < 2\,\delta.$$

Similarly it holds:

$$\forall b \in I_m,\ \ a \notin I_m: \quad \left\| (\overline{\mathbf{1}}_{[a,b)_{I_m}} - \mathbf{1}_{[a,b)}) \cdot \mathbf{1}_{I_m} \right\| \leq 2\, l(I_m) < 2\,\delta.$$

After substituting these two relations into (A.46) one obtains:

$$\forall P \quad \sum_{I\in P,\, l(I)<\delta} \left\|(\overline{\mathbf{1}}_{[a,b)_I} - \mathbf{1}_{[a,b)}) \cdot \mathbf{1}_I\right\| < 4\,\delta = 4 \cdot \frac{\varepsilon}{4} = \varepsilon$$

The last inequality holds for an arbitrary partition P of the interval $[0,1)$. Therefore the following is also valid:

$$\sup_{P\in\mathcal{P}} \sum_{I\in P,\, l(I)<\delta} \left\|(\overline{\mathbf{1}}_{[a,b)_I} - \mathbf{1}_{[a,b)}) \cdot \mathbf{1}_I\right\| < \varepsilon, \quad \text{where } \delta = \frac{\varepsilon}{4}.$$

Consequently any indicator function $f = \mathbf{1}_{[a,b)},\ [a,b) \subset [0,1)$ belongs to $\mathcal{H}_0$.

According to **(i)**, $\mathcal{H}_0$ is a linear subset of $L^1\big([0,1)\big)$. Consequently together with any indicator function, $\mathcal{H}_0$ also contains any linear combination of such functions. This means that any arbitrary step function on $[0,1)$, i.e. any arbitrary function from $S\big([0,1)\big)$, belongs to $\mathcal{H}_0$. From this we deduce, that for each function f from the space $S\big([0,1)\big)$ one can always find a sequence $\{f_n\}_{n=1}^{\infty}$ from the set $\mathcal{H}_0$ with $\lim\limits_{n\to\infty} f_n = f$ (e.g. $\forall n,\ f_n = f$). According to the definition of the dense subset (s. [3, p. 53-54]), $\mathcal{H}_0$ is dense in $S\big([0,1)\big)$ and therefore it is also dense in $L^1\big([0,1)\big)$. This completes the proof of **(iii)**.

From **(ii)** and **(iii)** we have: $\mathcal{H}_0 = \overline{\mathcal{H}}_0$ and $\overline{\mathcal{H}}_0 = L^1\big([0,1)\big)$, respectively. Summarizing we conclude that $\mathcal{H}_0 = L^1\big([0,1)\big)$, i.e. any function f from $L^1\big([0,1)\big)$ obeys the property (A.32).

Thus the proof of Lemma 28 is complete.

□

A.5 Properties of Pareto distributed random variable

Proposition 8. *Let Δ be an arbitrary positive constant and ξ be a random variable which is distributed according to the symmetrical Pareto distribution with parameter $\alpha \geq 9$, as defined in (4.1), (4.2), Chapter 4. Then the following statements are valid:*

$$\mathbb{E}|\,\xi\,| \;=\; \frac{1}{\alpha-2} > 0 \tag{A.47}$$

$$E(\Delta,\alpha) \;:=\; \mathbb{E}\big(\,|\Delta+\xi\,| - |\xi\,|\,\big) = \Delta - \frac{1}{\alpha-2} + \frac{1}{\big(1+\Delta\big)^{\alpha-2}\,(\alpha-2)} > 0 \tag{A.48}$$

$$q < \frac{1}{2} \quad\Leftrightarrow\quad \widetilde{\xi}_q = 1 - \frac{1}{(2\,q\,)^{\frac{1}{\alpha-1}}} < 0 \tag{A.49}$$

$$q > \frac{1}{2} \quad\Leftrightarrow\quad \widetilde{\xi}_q = \frac{1}{(2\,(\,1-q\,))^{\frac{1}{\alpha-1}}} - 1 > 0, \tag{A.50}$$

where $\widetilde{\xi}_q$ is q-quantile of ξ.

Proof of (A.47):

$$\mathbb{E}\,(\,|\,\xi\,|\,) = \int\limits_{(-\infty,+\infty)} |x| \cdot f_{\xi}(x)\,dx$$

$$
\begin{aligned}
&\underset{(4.2)}{=} \int\limits_{(-\infty,0)} -x\cdot\frac{\alpha-1}{2\,(1-x)^{\alpha}}\,dx \;+ \int\limits_{[0,+\infty)} x\cdot\frac{\alpha-1}{2\,(1+x)^{\alpha}}\,dx\\
&= \int\limits_{(-\infty,0)} \frac{\alpha-1}{2}\cdot\frac{(1-x)-1}{(1-x)^{\alpha}}\,dx \;+ \int\limits_{[0,+\infty)} \frac{\alpha-1}{2}\cdot\frac{(1+x)-1}{(1+x)^{\alpha}}\,dx\\
&= \int\limits_{(-\infty,0)} \frac{\alpha-1}{2\,(1-x)^{\alpha-1}}\,dx \;- \int\limits_{(-\infty,0)} \frac{\alpha-1}{2\,(1-x)^{\alpha}}\,dx\\
&\quad+ \int\limits_{[0,+\infty)} \frac{\alpha-1}{2\,(1+x)^{\alpha-1}}\,dx \;- \int\limits_{[0,+\infty)} \frac{\alpha-1}{2\,(1+x)^{\alpha}}\,dx\\
&\underset{(4.1)}{=} \frac{\alpha-1}{2\,(\alpha-2)\,(1-x)^{\alpha-2}}\Bigg|_{-\infty}^{0} -\; F_{\xi}\big((-\infty,\,0\,)\big)\\
&\quad- \frac{\alpha-1}{2\,(\alpha-2)\,(1+x)^{\alpha-2}}\Bigg|_{0}^{+\infty} -\; F_{\xi}\big([\,0,\,+\infty)\big)\\
&= \frac{\alpha-1}{\alpha-2}-1 \;=\; \frac{1}{\alpha-2} > 0 \;\text{ for } \alpha\geq 9.
\end{aligned}
$$

Thus statement (A.47) is proved.

Proof of (A.48):

We fix an arbitrary $\Delta > 0$ and consider $\mathbb{E}\left(\,|\Delta+\xi| - |\,\xi\,|\,\right)$, which we denote by $E(\Delta,\alpha)$:

$$
E(\Delta,\alpha) = \mathbb{E}\left(\,|\Delta+\xi|-|\,\xi\,|\,\right) = \int\limits_{(-\infty,+\infty)} \left(\,|\Delta+x|-|x|\,\right)\cdot f_{\xi}(x)\,dx
$$

$$
\begin{aligned}
&= \int\limits_{(-\infty,-\Delta)} -\Delta\cdot f_{\xi}(x)\,dx + \int\limits_{[\,-\Delta,0\,)} (\Delta+2\,x)\cdot f_{\xi}(x)\,dx + \int\limits_{[0,+\infty)} \Delta\cdot f_{\xi}(x)\,dx\\
&= -\Delta\cdot F_{\xi}(-\Delta)+\Delta\cdot\left(\frac{1}{2}-F_{\xi}(-\Delta)\,\right) - \int\limits_{[\,-\Delta,0\,)} \frac{(\alpha-1)\,(1-x-1)}{(1-x)^{\alpha}}\,dx + \frac{\Delta}{2}\\
&= -2\,\Delta\cdot F_{\xi}(-\Delta) + \Delta - \frac{\alpha-1}{(\alpha-2)\,(1-x)^{\alpha-2}}\Bigg|_{-\Delta}^{0} + 2\,F_{\xi}(0) - 2\,F_{\xi}(\,-\Delta\,)\\
&= -2\,F_{\xi}(-\Delta)\,(\,\Delta+1\,)+(\,\Delta+1\,)+\frac{\alpha-1}{(\alpha-2)\,(1+\Delta)^{\alpha-2}} - \frac{\alpha-1}{\alpha-2}\\
&= (\,\Delta+1\,)\Big(1-\frac{1}{(1+\Delta)^{\alpha-1}}\Big) - \frac{\alpha-1}{\alpha-2} + \frac{\alpha-1}{\alpha-2}\cdot\frac{1}{(1+\Delta)^{\alpha-2}}\\
&= \Delta-\Big(\frac{\alpha-1}{\alpha-2}-1\Big) + \frac{1}{(1+\Delta)^{\alpha-2}}\cdot\Big(\frac{\alpha-1}{\alpha-2}-1\Big)\\
&= \Big(\Delta-\frac{1}{\alpha-2}\Big) + \frac{1}{(1+\Delta)^{\alpha-2}\,(\alpha-2)}.
\end{aligned}
$$

Consequently, for any $\Delta > 0$ and any Pareto distributed random variable ξ with parameter α $(\alpha \geq 9)$, we obtain that

$$E(\Delta,\alpha) = \Big(\Delta - \frac{1}{\alpha-2}\Big) + \frac{1}{(1+\Delta)^{\alpha-2}\,(\alpha-2)}.$$

By means of the strict version of Bernoulli's inequality [9]:

$$(1+r\,x) < (1+x)^r, \tag{A.51}$$

which is valid for any $x \geq -1$ with $x \neq 0$ and for any $r > 1$ or $r < 0$, it is easy to show that $E(\Delta,\alpha)$ is strictly positive for any $\Delta > 0$ and any $\alpha \geq 9$, indeed

$$E(\Delta,\alpha) > \Delta - \frac{1}{\alpha-2} + \frac{1-(\alpha-2)\,\Delta}{\alpha-2} = 0.$$

This completes the proof of $(A.48)$.

Proof of (A.49) and (A.50) :
We compute q-quantile $\widetilde{\xi}_q$ of ξ, $q \in (0,1)$. According to the definition of the q-quantile it holds:

$$(1-q) = \mathbb{P}(\xi > \widetilde{\xi}_q)$$

1) Let $\widetilde{\xi}_q > 0$. Then due the definition of symmetrical Pareto distribution (see (4.1)) it holds:

$$\begin{aligned} 1-q &= \frac{1}{2\,(1+\widetilde{\xi}_q\,)^{\alpha-1}}, \\ \Rightarrow \qquad \widetilde{\xi}_q &= \frac{1}{(2\,(1-q\,))^{\frac{1}{\alpha-1}}} - 1. \end{aligned} \tag{A.52}$$

2) Let $\widetilde{\xi}_q < 0$. Then

$$\begin{aligned} 1-q &= 1 - \frac{1}{2\,(1-\widetilde{\xi}_q\,)^{\alpha-1}}, \\ \Rightarrow \qquad \widetilde{\xi}_q &= 1 - \frac{1}{(2\,q\,)^{\frac{1}{\alpha-1}}}. \end{aligned} \tag{A.53}$$

Since ξ is a symmetrical random variable, it is obvious that $\widetilde{\xi}_q > 0$ if and only if $q > 0$ and $\widetilde{\xi}_q < 0$ if and only if $q < 0$. This completes the proof of (A.49) and (A.50) and, consequently, of Proposition 8.

□

Lemma 29. *Let ξ be a symmetrical Pareto distributed random variable with parameter $\alpha \geq 9$, as defined in (4.1), (4.2), Chapter 4. Moreover, let $\widetilde{\xi}_q$ be q-quantile of ξ and $C(n) = \sqrt{n}\,\ln n$. Then for sufficiently large $n \in \mathbb{N}$ and all $l \geq \frac{n}{\ln n}$ it holds:*

$$(1-q) - \mathbb{P}\Big(\xi > \widetilde{\xi}_q + \frac{C(n)}{l}\Big) \geq \frac{\alpha-1}{2\,(1+\widetilde{\xi}_q\,)^{\alpha}} \cdot \frac{C(n)}{l} > 0, \quad q > \frac{1}{2}, \tag{A.54}$$

$$(1-q) - \mathbb{P}\Big(\xi > \widetilde{\xi}_q + \frac{C(n)}{l}\Big) \geq \frac{\alpha-1}{2\,(1-\widetilde{\xi}_q\,)^{\alpha}} \cdot \frac{C(n)}{l} > 0, \quad q < \frac{1}{2}. \tag{A.55}$$

Proof:
At first we consider the case when $q > \frac{1}{2}$.
In the proof of Proposition 8 (see (A.52) it was shown that for any $q > \frac{1}{2}$ it holds:

$$(1-q) = \frac{1}{2\,(1+\widetilde{\xi}_q\,)^{\alpha-1}}, \tag{A.56}$$

thereby $\widetilde{\xi}_q > 0$.
According to this and also to formula (4.1) of the distribution function of the symmetrical Pareto distributed random variable ξ we have:

$$\mathbb{P}\Big(\xi > \widetilde{\xi}_q + \frac{C(n)}{l}\Big) = \frac{1}{2\,\Big(1+\widetilde{\xi}_q + C(n)/l\Big)^{\alpha-1}}, \quad \alpha \geq 9. \tag{A.57}$$

Since for any $l > \dfrac{n}{\ln n}$ it holds:

$$\frac{C(n)}{l} = \frac{\sqrt{n}\,\ln n}{l} < \frac{\ln^2 n}{\sqrt{n}} \underset{n\to\infty}{\longrightarrow} 0, \tag{A.58}$$

then there exists a natural finite number N_0 such that for all $n \geq N_0$ and all $l \geq \dfrac{n}{\ln n}$ the function $\dfrac{1}{2\,\Big(1+\widetilde{\xi}_q + C(n)/l\Big)^{\alpha-1}}$ can be approximated with the help of Taylor's formula near the point $\widetilde{\xi}_q$.
As the result, for all $n \geq N_0$ and all $l \geq \dfrac{n}{\ln n}$ we obtain:

$$\mathbb{P}\Big(\xi > \widetilde{\xi}_q + \frac{C(n)}{l}\Big) = \frac{1}{2\,(1+\widetilde{\xi}_q\,)^{\alpha-1}} - \frac{\alpha-1}{2\,(1+\widetilde{\xi}_q\,)^{\alpha}} \cdot \frac{C(n)}{l} + \mathrm{O}\Big(\frac{(C(n))^2}{l^2}\Big)$$

This equation together with (A.56) yields :

$$(1-q) - \mathbb{P}\Big(\xi > \widetilde{\xi}_q + \frac{C(n)}{l}\Big) = \frac{\alpha-1}{2\,(1+\widetilde{\xi}_q\,)^{\alpha}} \cdot \frac{C(n)}{l} + \mathrm{O}\Big(\frac{(C(n))^2}{l^2}\Big) \geq \frac{\alpha-1}{2\,(1+\widetilde{\xi}_q\,)^{\alpha}} \cdot \frac{C(n)}{l} > 0$$

Therefore the validity of (A.54) has been proved.

Next we consider the case when $q < \frac{1}{2}$.
The quantity $(1-q) - \mathbb{P}\Big(\xi > \widetilde{\xi}_q + \frac{C(n)}{l}\Big)$ for $q < \frac{1}{2}$ will be slightly different from its value in the case $q > \frac{1}{2}$ because for $q < \frac{1}{2}$ the q-quantile $\widetilde{\xi}_q < 0$.
We know (s. (A.53)) that for any $q < \frac{1}{2}$ it holds:

$$(1-q) = \mathbb{P}(\xi > \widetilde{\xi}_q) = 1 - \frac{1}{2\,(1-\widetilde{\xi}_q\,)^{\alpha-1}}.$$

Now we consider $\mathbb{P}\Big(\xi > \widetilde{\xi}_q + \dfrac{C(n)}{l}\Big)$, but first, we remark that for $l > \dfrac{n}{\ln n}$, according to (A.58), we have $\dfrac{C(n)}{l} \underset{n\to\infty}{\longrightarrow} 0$, which implies that $\widetilde{\xi}_q + \dfrac{C(n)}{l} \le 0$ for $\widetilde{\xi}_q < 0$, or in other words $\widetilde{\xi}_q + \dfrac{C(n)}{l} \le 0$ for $q < \dfrac{1}{2}$. This means that there exists a natural finite number N' such that for all $n > N'$ and all $l > \dfrac{n}{\ln n}$ it holds:

$$\mathbb{P}\Big(\xi > \widetilde{\xi}_q + \frac{C(n)}{l}\Big) = 1 - \frac{1}{2\Big(1-(\widetilde{\xi}_q + C(n)/l)\Big)^{\alpha-1}},$$

$$\Rightarrow \qquad (1-q) - \mathbb{P}\Big(\xi > \widetilde{\xi}_q + \frac{C(n)}{l}\Big) = \frac{1}{2\Big(1-(\widetilde{\xi}_q + C(n)/l)\Big)^{\alpha-1}} - \frac{1}{2\,(1-\widetilde{\xi}_q\,)^{\alpha-1}}$$

$$= \frac{1}{2\,(1-\widetilde{\xi}_q\,)^{\alpha-1}} + \frac{\alpha-1}{2\,(1-\widetilde{\xi}_q\,)^{\alpha}} \cdot \frac{C(n)}{l} + \mathrm{O}\Big(\frac{(C(n))^2}{l^2}\Big) - \frac{1}{2\,(1-\widetilde{\xi}_q\,)^{\alpha-1}}$$

$$\ge \frac{\alpha-1}{2\,(1-\widetilde{\xi}_q\,)^{\alpha}} \cdot \frac{C(n)}{l} > 0$$

The last relation is valid since $\widetilde{\xi}_q < 0$ for $q < \dfrac{1}{2}$ and $\alpha \le 9$.
Consequently, we have proved the validity of (A.55). This means the proof of Lemma 29 is complete.

□

Appendix B

Symbols and Notations

- $\lceil a \rceil = \min\limits_{k\in\mathbb{Z},\, k\geq a} k$
- $C(n) = \sqrt{n}\,\ln n,\ \ n \in \mathbb{N}$
- (Θ, μ), (Θ_1, μ_1) and (Θ_2, μ_2) some arbitrary measure spaces
- $\Theta_1 \times \Theta_2 := \Big\{(x, y) \,\big|\, x \in \Theta_1,\ y \in \Theta_2\Big\}$ cross product of Θ_1 and Θ_2
- $(g_1 \otimes g_2)(x, y) := g_1(x) \cdot g_2(y), \quad x \in \Theta_1,\ y \in \Theta_2$ tensor product
- $(\Omega, \mathbb{P})$ some arbitrary probability space
- $\widetilde{\eta}_q$ q-quantiel of a random variable $\eta,\ q \in (0, 1)$
- $L^p(\Theta) := \Big\{\, f : \Theta \to \mathbb{R} \ \Big|\ \int\limits_{\Theta} |f(x)|^p \, d\mu(x) < \infty \Big\},\ \ p \geq 1$
- $L^1\big([0, 1)\big) := \Big\{\, f : [0, 1) \to \mathbb{R} \ \Big|\ \int\limits_{[0,1)} |f(x)| \, dx < \infty \,\Big\}$
- $L^1(\Theta_1 \times \Theta_2) := \Big\{\, f : \Theta_1 \times \Theta_2 \to \mathbb{R} \ \Big|\ \int\limits_{\Theta_1\times\Theta_2} |f(x, y)| \, d\mu_1(x)\, \mu_2(y) < \infty \,\Big\}$
- $\|f\|_p := \Bigg(\int\limits_{\Theta} |f(x)|^p \; d\mu(x)\Bigg)^{1/p}$ the norm of $f \in L^p(\Theta)$ for $1 \leq p < \infty$
- $\|f\| = \int\limits_{[0,1)} |f(x)| \; dx$ L^1 norm of $f \in L^1\big([0, 1)\big)$
- $\|h\| = \int\limits_{\Theta_1\times\Theta_2} |h(x, y)| \; d\mu_1(x) \, d\mu_2(y)$ L^1 norm of $h \in L^1\big(\Theta_1 \times \Theta_2\big)$
- $\|f_n - f\| \underset{n\to\infty}{\longrightarrow} 0$ convergence of a sequence $(f_n)_{n\in\mathbb{N}} \subset L^1\big([0, 1)\big)$ to f with respect to the L^1 convergence
- $l(A)$ Lebesgue measure of a set A

- $\operatorname{argmin} F$ or $\operatorname*{argmin}_{\Theta} F$ or $\operatorname*{argmin}_{g\in\Theta} F(g)$ the whole set of minimizers in Θ for a functional $F:\Theta\to\mathbb{R}$

- $med_{\,\Theta_0}(h) := \Big\{ c\in\mathbb{R} \;\Big|\; \forall x\in\Theta_0 : \;\; \mu\big(h(x)\geq c\big) \geq \frac{1}{2} \text{ and } \mu\big(h(x)\leq c\big)\geq \frac{1}{2} \Big\}$ the set of all medians of $h:\Theta\to\mathbb{R}$ on a set $\Theta_0\subseteq\Theta$

- $med_{\Theta_0}(f) = \operatorname*{argmin}_{c\in\mathbb{R}} \int_{\Theta_0} |\, f(x)-c \,|\; d\mu(x)$ the set of all median of $f\in L^1(\Theta)$ on $\Theta_0\subseteq\Theta$

- $\mathbf{1}_A(x) = \begin{cases} 1, & \text{if } x\in A, \\ 0, & \text{if } x\notin A \end{cases}$ the indicator function of a subset $A\subseteq\Theta$

- $\iota^n:\mathbb{R}^n\to L^1\big([0,1)\big), \quad \iota^n\big((u_1,\dots,u_n)\big) := \sum_{i=1}^{n} u_i\cdot \mathbf{1}_{\left[\frac{i-1}{n},\frac{i}{n}\right)}(t)$

- $S\big([0,1)\big) := \operatorname{span}\Big\{ \mathbf{1}_{[s,t)} \;:\; 0\leq s<t\leq 1 \Big\}$ set of all piecewise constant functions on $[0,1)$ with finite number of jumps.

- $S_n\big([0,1)\big) := \iota^n\big(\mathbb{R}^n\big)$ set of all piecewise constant functions on $[0,1)$ with at most $n-1$ jumps, where all discontinuity points are the points of the form $\frac{i}{n}$, $i=1,\dots,n$

- $f^n := \iota^n\big(\,(f_1^n,\dots,f_n^n)\,\big)$ an approximation in $S_n\big([0,1)\big)$ of a function $f\in L^1\big([0,1)\big)$

- $\xi^n := \iota^n\big(\,(\xi_1^n,\dots,\xi_n^n)\,\big)$ embedding into $S_n\big([0,1)\big)$ of random variables $\xi_1^n,\dots,\xi_n^n$

- P partition of the interval $[0,1)$

- $\mathcal{P}$ set of all partitions of the interval $[0,1)$

- $J(g)$ set of the discontinuity points of the function $g\in S\big([0,1)\big)$, i.e. $J(g) := \Big\{ t\in[0,1) \;:\; g(t^+)\neq g(t^-) \Big\}$

- $P_{J(g)}$ the partition of interval [0, 1) generated by the points of the set $J(g)$

- $H_\gamma^n:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}_+$
  $H_\gamma^n(u,Y) := \gamma\#J(u) + \frac{1}{n}\sum_{i=1}^{n}\big|u_i-y_i\big|,$
  with $Y=(y_1,\dots,y_n)$, $u=(u_1,\dots,u_n)$ and $J(u) := \Big\{ i \;:\; 1\leq i\leq n-1,\; u_i\neq u_{i+1} \Big\}$

- $\tilde{H}_\gamma^n : L^1\big([0,1)\big)\times L^1\big([0,1)\big)\to\mathbb{R}\cup\{\infty\}$

$$\tilde{H}_\gamma^n(g,f) := \begin{cases} \gamma\#J(g) + \|g-f\| - \|f\| & : \; g\in S_n\big([0,1)\big),\ \gamma>0 \\ \|g-f\| - \|f\| & : \; g\in L^1\big([0,1)\big),\ \gamma=0 \\ \infty & : \; \text{otherwise} \end{cases}$$

- $H^*_{\gamma,\xi} : L^1\big([0,1)\big) \times L^1\big([0,1)\big) \to \mathbb{R} \cup \{\infty\}$

$$H^*_{\gamma,\xi}(g,f) := \begin{cases} \gamma \# J(g) + \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du & : \; g \in S\big([0,1)\big), \\ & \;\;\; \gamma > 0; \\ \int \mathbb{E}\big(|\xi + f(u) - g(u)| - |\xi + f(u)|\big)\, du & : \; g \in L^1\big([0,1)\big), \\ & \;\;\; \gamma = 0; \\ \infty & : \; \text{otherwise} \end{cases}$$

- $\Phi(x,y) := |x-y| - |y|$
- $\Psi(x,y) := \|x-y\| - \|y\|$
- $H^\infty_\gamma : L^1\big([0,1) \times \Omega\big) \times L^1\big([0,1) \times \Omega\big) \to \mathbb{R} \cup \{\infty\}$

$$H^\infty_\gamma(g \otimes 1,\, h) := \begin{cases} \gamma \# J(g) + \Psi(g \otimes 1,\, h) & : \; g \in S\big([0,1)\big),\, \gamma > 0 \\ \Psi(g \otimes 1,\, h) & : \; g \in L^1\big([0,1)\big),\, \gamma = 0 \\ \infty & : \; \text{otherwise} \end{cases}$$

- distribution function of symmetrical Pareto distribution with parameter $\alpha > 1$:

$$F(x) = \begin{cases} \dfrac{1}{2\,(1-x)^{\alpha-1}} & , \quad x < 0 \\ 1 - \dfrac{1}{2\,(1+x)^{\alpha-1}} & , \quad x \geq 0 \end{cases}$$

- probability density of symmetrical Pareto distribution with parameter $\alpha > 1$:

$$f_\xi(x) = \begin{cases} \dfrac{\alpha-1}{2\,(1-x)^{\alpha}} & , \quad x < 0 \\ \dfrac{\alpha-1}{2\,(1+x)^{\alpha}} & , \quad x \geq 0 \end{cases}$$

- $\operatorname{med}\big(\xi^n_{i_1:i_2} + c_1,\, \xi^n_{j_1:j_2} + c_2\big) := \operatorname{med}\big(\xi^n_{i_1} + c_1, \ldots, \xi^n_{i_2} + c_1,\, \xi^n_{j_1} + c_2, \ldots, \xi^n_{j_2} + c_2\big)$
- $\operatorname{med}\big(\xi^n_{i_1:i_2} + c_1,\, \xi^n_{j_1:j_2} + c_2,\, \xi^n_{s_1:s_2} + c_3\big) :=$
  $\operatorname{med}\big(\xi^n_{i_1} + c_1, \ldots, \xi^n_{i_2} + c_1,\, \xi^n_{j_1} + c_2, \ldots, \xi^n_{j_2} + c_2,\, \xi^n_{s_1} + c_3, \ldots, \xi^n_{s_2} + c_3\big)$

Bibliography

[1] M. Abramowitz, I. A. Stegun, *Handbook of mathematical Functions*, 9th Printing, Dover Publications, Inc., New York (1972).

[2] R. J. Adler, R. E. Feldman, M. S. Taqqu (Editors), *A Practical Guide to Heavy Tails*, Series: Statistical Techniques and Applications, Birkhäuser (1998)

[3] G. Bachman, L. Narici, *Functional Analysis* Dover Publications, Inc, Mineola, New York (2000)

[4] L. E. Baum, M. Katz, *Convergence rates in the low of large numbers*, *Transactions of the American Mathematical Society* **120** (1), 108-123 (1965).

[5] G. Beer, *Topologies on Closed and Closed Convex Sets, 5.Auflage*, vol. 268 of *Mathematics and its Applications*, Kluwer Academic Publishers Group, Dordrecht (1993)

[6] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York (1968).

[7] L. Boysen, V. Liebscher, A. Munk and O. Wittich, *Scale space consistency of piecewise constant least squares estimators - another look at the regressogram*, *IMS Lecture Notes-Monograph Series*, **55**, 65-84 (2007).

[8] L. Boysen, A. Kempe, V. Liebscher, A. Munk and O. Wittich, *Consistencies and rates of convergence of jump-penalized least squares estimators*, *The Annals of Statistics*, **107**(1), 157-183 (2009).

[9] N. Carothers, *Real Analysis*, Cambridge: Cambridge University Press (2000)

[10] W. S. Cleveland, *Robust Locally Weighted Regression and Smoothing Scatterplots*, Journal of the American Statistical Association **74** (368): 829836 (1979)

[11] G. Dal Maso, *An Introduction to* Γ*-convergence*, vol. 8 of *Progress in Nonlinear Differential Equations and their Applications*, Birkhuser, Boston Inc., Boston, MA (1993).

[12] L. Demaret, F. Friedrich, V. Liebscher, G. Winkler, *Complexity Penalized M-Estimation: Consistency in More Dimensions*, submited (August, 2011)

[13] R. A. DeVore, G. G. Lorentz, *Constructive Approximation*, Springer, Berlin (1993).

[14] M. Fisz, *Wahrscheinlichkeitsrechnung und mathematische Statistik*, 5.Auflage, VEB Deutscher Verlag der Wissenschften, Berlin (1970).

[15] F. Friedrich, *Complexity penalized segmentations in 2D*, *PhD thesis*, Institut für Biomathematik und Biometrie an der Gesellschaft fr̈ Umwelt und Gesundheit, München-Neuherberg (2005).

[16] F. Friedrich, A. Kempe, V. Liebscher and G. Winkler, *Complexity Penalised M-Estimation: Fast Computation*, *J. Comp. Graph. Stattist.*, **17**, 1-24 (2008).

[17] P. Halmos, *Finite Dimensional Vector Spaces*, Springer (1974).

[18] W. Härdle, *Applied nonparametric regression*, Cambridge University Press, Cambridge New York (1990).

[19] C. Hess, *Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator*, *Ann. Statist.*, **24**, 1298-1315 (1996).

[20] D. V. Hinkley, *Inference about the change-point in a sequence of random variables.*, *Biometrica*, **57**, 1-17 (1970).

[21] W. Hoeffding, *Probability Inequalities for Sum of Bounded Random Variables*, *Journal of American Statistical Association*, **58**(301), 13-30 (March 1963).

[22] A. D. Ioffe, V. M. Tihomirov, *Theory of Extremal Problems*, North-Holland Publishing Company, Amsterdam-New York-Oxford(1979).

[23] E. Ising, *Beitrag zur Theorie des Ferromagnetismus.*, *Z. Physik*, **31**, 253 (1925).

[24] N. J. Kalton, N. T. Peck, J. W. Roberst, *An F-space sampler*, London Mathematical Society, Lecture Note Series, **89**, Cambridge: Cambridge University Press (1984).

[25] A. Kempe, *Statistical analysis of discontinuous phenomena with Potts functionals*, *PhD thesis*, Institute of Biomathematics and Biometry, National Research Center for Environment and Health, Munich, Germany (2004).

[26] A. Kempe, V. Liebscher, G. Winkler and O. Wittich, *Segmentation of Time Series: Parameter Dependence of Blake Zisserman Functionals and Mumford-Shah Functionals and the Transition from Discrete to Continuous*, *Schriftereihe des IBB*, 05-02 (March 2005).

[27] A. N. Kolmogorov, S. V. Fomin, *Elements of the Theory of Functions and Functional Analysis*,(Russian edition), 4.Edition, Nauka, Moskva, (1989).

[28] G. Matheron, *Random Sets and Integral Geometry*, in Wiley Series in Probability and Mathematical Statistics, John Wiley&Sons., New York etc. (1975).

[29] W. Mendenhall, T. Sincich, *A Second Course in Statistics: Regression Analysis*, (7th Edition) Prentice Hall; 7 edition (January 15, 2011).

[30] D. Mumford and J. Shah, *Boundary detection by minimizing functionals I*, In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 22-26, Washington (June 1985). CVPR '85, San Francisco, IEEE Computer Society Press (1985).

[31] D. Mumford and J. Shah, *Optimal approximation by piecewise smooth functions and associated variational problem*, *Comm. Pure Appl.Math.,*, **42**, 577-685, (1989).

[32] E. A. Nadaraya, *On Estimating Regression Theory of Probability and its Applications*, **9** (1), 141142 (1964).

[33] V. V. Petrov, *Sums of Independent Random Variables*, Springer-Verlag, Berlin, Heidelberg, New York (1975).

[34] R. Potts, *Some generalized order-disorder transitions*, *Proc. Camb. Philos. Soc*, **48**, 106-109 (1952).

[35] H. Reinsch, *Smoothing by spline functions*, *Numerische Mathematik*, **10**, 177-83 (1967).

[36] R. T. Rockafellar, *Convex Analysis*, Paperback, Princeton University Press, Princeton (1997).

[37] M. G. Schimek (Editor), *Smoothing and Regression: Approaches, Computation, and Application*, A Wiley-Interscience Publication, JOHN WILEY and SONS (2000)

[38] I. J. Schoenberg, *Spline functions and the problem of graduation*, *Mathematics*, **52**, 947-50 (1964).

[39] L. D. Schroeder, D. L. Sjoquist, P. E. Stephan, *Understanding Regression Analysis: an introductory guide*, Series: Quantitative Applications in the Sociale Sciences, a SAGE University Paper, 1986

[40] G. Schwarz, *Estimating the Dimension of a Model*, *Annals of Statistics*, 461-464 (2/6/1978).

[41] K. Takezawa, *Introduction to Nonparametric Regression*, John willey and Sons, Inc. (2006).

[42] J. W. Tukey, *Curves as parameters, and touch estimation*, *In Proc. 4th Berkeley Sympos. Math. Statist. and Prob*, **I**, Univ. California Press, Berkeley, Calif., 681-694 (1961).

[43] G. Winkler and V. Liebscher, *Smoothers for Discontinuous Signals*, *J.Nonpar. Statist.*, **14**(1-2) 203-222 (2002).

[44] G. Winkler, A. Kempe, V. Liebscher and O. Wittich, *Parsimonious Segmentation of Time Series' by Potts Models*, *Schriftereihe des IBB*, 03-03 (March 2003).

[45] G. Winkler, O. Wittich, V. Liebscher and A. Kempe, *Don't shed Tears over Break*, *Jahresbericht der DMV*, **107**(2), 57-87 (2005).

[46] O. Wittich, A. Kempe, V. Liebscher and G. Winkler, *Complexity Penalized Sums of Squares for Time Series: Rigorous Analytical Results*, *Schriftereihe des IBB*, 05-01 (February 2005).

[47] E. T. Whittaker, G. N. Watson, *A Course in Modern Analysis*, fourth edition, Cambridge University Press, Cambridge University Press (1963)

[48] Y. -C. Yao, *Estimating the number of change-points via Schwarz' criterion*, *Statist. Probab. Lett.*, **6**, 181-189 (1988).

[49] Y. -C. Yao and S. T. Au, *Least-squares estimation of a step function*, *Sankhya Ser. A*, **51**, 370-381 (1989).

Acknowledgements

First of all I thank my supervisor Prof. Dr. Volkmar Liebscher for offering an excellent supervision over the entire course of my Phd study. I also thank him for his thorough attitude to everything and for the numerous scientific discussions that we have had while I was working at the department and also during the time, when I left Greifswald. I enjoyed the friendly atmosphere at our department, which was created not least due to his efforts. I am grateful for his understanding and constant support in questions other than work-related.

My special thank goes to all my colleagues, especially, Evelyn Herrholz, Robert Bialowons, Michael Kläre, Mirco Schultka, Claudia Schurmann, Peter Nestler and Nora Stahnke. I thank them for their friendship and readiness to help not only by discussing mathematics. Thank to them, I was really enjoying my stay in Greifswald, and they became a substantial part of my life.

Thank to all my friends for being a very kind and useful distraction from work. I am very thankful to my family, my son Anton, my husband Andrey, my parents and my brothers for the inspiration and emotional support. I always new that no matter what happens, my family and my friends will always be there for me.

I also thank Andrey and my good friend, Valeriya Lykina, for proofreading the thesis and for many helpful comments.

Finally, I thank the Department of Mathematics and Computer Science, Ernst-Moritz-Arndt-University Greifswald for the financial support.