Ann. I. H. Poincaré – PR 38, 6 (2002) 923–957

S0246-0203(02)01144-5/FLA

A FUNCTIONAL HUNGARIAN CONSTRUCTION FOR SUMS OF INDEPENDENT RANDOM VARIABLES

UNE CONSTRUCTION HONGROISE POUR DES SOMMES DE VARIABLES ALÉATOIRES INDÉPENDANTES

Ion GRAMA [a,1,*], **Michael NUSSBAUM** [b,2]

[a] *Université de Bretagne-Sud, laboratoire Sabres, rue Yves Mainguy, 56000 Vannes, France*
[b] *Department of Mathematics, Malott Hall, Cornell University, Ithaca, NY 14853-4201, USA*

Received 1 March 2001, revised 1 July 2002

ABSTRACT. – We develop a Hungarian construction for the partial sum process of independent non-identically distributed random variables. The process is indexed by functions f from a class $\mathcal{H}$, but the supremum over $f \in \mathcal{H}$ is taken outside the probability. This form is a prerequisite for the Komlós–Major–Tusnády inequality in the space of bounded functionals $l^\infty(\mathcal{H})$, but contrary to the latter it essentially preserves the classical $n^{-1/2} \log n$ approximation rate over large functional classes $\mathcal{H}$ such as the Hölder ball of smoothness $1/2$. This specific form of a strong approximation is useful for proving asymptotic equivalence of statistical experiments.

MSC: primary 60F17, 60F99; secondary 62G07

Keywords: Komlós-Major–Tusnády inequality; Partial sum process; Non-identically distributed variables; Function classes; Asymptotic equivalence of statistical experiments

RÉSUMÉ. – Nous développons une construction hongroise pour des sommes partielles de variables aléatoires indépendantes non identiquement distribuées. Le processus est indexé par les fonctions f d'une classe $\mathcal{H}$ mais le suprémum en $f \in \mathcal{H}$ est pris à l'extérieur de la probabilité. Cette forme est un prérequis pour l'inégalité de Komlós–Major–Tusnády dans l'espace des fonctionnelles bornées $l^\infty(\mathcal{H})$, mais contrairement à cette dernière, elle préserve pour l'essentiel la vitesse d'approximation classique en $n^{-1/2} \log n$ pour une large classe d'espace $\mathcal{H}$, y compris la boule hölderienne d'indice $1/2$. Cette forme spécifique d'approximation est utile pour démontrer l'équivalence asymptotique des expériences statistiques.

[*] Corresponding author.
E-mail addresses: ion.grama@univ-ubs.fr (I. Grama), nussbaum@math.cornell.edu (M. Nussbaum).
[1] Supported by Deutsche Forschungsgemeinschaft, by the German Academic Exchange Service (DAAD) and partly by the Alexander von Humboldt Schtiftung.
[2] Supported by NSF Grant DMS-0072162.

1. Introduction

Let X_i, $i = 1, \ldots, n$, be a sequence of independent random variables with zero means and finite variances. Let $\mathcal{H}$ be a class of real valued functions on the unit interval [0, 1] and $t_i = i/n$, $i = 1, \ldots, n$. The *partial sum process indexed by functions* is the process

$$X^n(f) = n^{-1/2} \sum_{i=1}^{n} f(t_i) X_i, \quad f \in \mathcal{H}.$$

Suppose $f \in \mathcal{H}$ are uniformly bounded; then $X^n = \{X^n(f),\ f \in \mathcal{H}\}$ may be regarded as a random element with values in $l^\infty(\mathcal{H})$- the space of real valued functionals on $\mathcal{H}$. The class $\mathcal{H}$ is Donsker if X^n converges weakly in $l^\infty(\mathcal{H})$ to a Gaussian process. We are interested in associated coupling results, i.e. in finding versions of X^n and of this Gaussian process on a common probability space which are close as random variables. The standard coupling results of the type "nearby variables with nearby laws" (cf. Dudley [3], Section 11.6) naturally refer to the sup-metric in $l^\infty(\mathcal{H})$: for an appropriate version of X^n ($\widetilde{X}^n = \{\widetilde{X}^n(f),\ f \in \mathcal{H}\}$, say) and of a Gaussian process $\widetilde{N}^n = \{\widetilde{N}^n(f),\ f \in \mathcal{H}\}$, we have

$$P^*\Big(\sup_{f \in \mathcal{H}} \big|\widetilde{X}^n(f) - \widetilde{N}^n(f)\big| > x\Big) \to 0, \quad x > 0, \tag{1.1}$$

where P^* is the outer probability on the common probability space (cf. van der Vaart and Wellner [20], 1.9.3, 1.10.4). Here we shall consider a different type of coupling. We are looking for versions $\widetilde{X}^n$, $\widetilde{N}^n$ such that

$$\sup_{f \in \mathcal{H}} P\big(\big|\widetilde{X}^n(f) - \widetilde{N}^n(f)\big| > x\big) \to 0, \quad x > 0, \tag{1.2}$$

and such that additional exponential bounds of the Komlós–Major–Tusnády type are valid. Note that (1.2) is weaker than (1.1) since the supremum is taken outside the probability. More specifically we are interested in a construction involving also a rate sequence $r_n \to 0$ such that

$$\sup_{f \in \mathcal{H}} P\big(r_n^{-1}\big|\widetilde{X}^n(f) - \widetilde{N}^n(f)\big| > x\big) \leqslant c_0 \exp\{-c_1 x\}, \quad x > 0. \tag{1.3}$$

Here c_0, c_1 are constants depending on the class $\mathcal{H}$.

The classical results of Komlós, Major and Tusnády ([9] and [10]) refer to a sup inside the probability for $\mathcal{H} = \mathcal{H}_0$, where $\mathcal{H}_0$ is the class of indicators $f(t) = \mathbf{1}(t \leqslant s)$, $s \in [0, 1]$. The following bound was established: for $r_n = n^{-1/2}$

$$P\Big(r_n^{-1} \sup_{f \in \mathcal{H}_0} \big|\widetilde{X}^n(f) - \widetilde{N}^n(f)\big| > x\Big) \leqslant c_0 \exp\{-c_1 x\}, \quad x \geqslant c_2 \log n, \tag{1.4}$$

provided $X_1, \ldots, X_n$ is a sequence of i.i.d. r.v.'s fulfilling Cramér's condition

$$E \exp\{t X_i\} < \infty, \quad |t| \leqslant t_0,\ i = 1, \ldots, n, \tag{1.5}$$

where c_0, c_1, c_2 are constants depending on the common distribution of the X_i. Note that r_n in (1.4) can be interpreted as a rate of convergence in the CLT over $l^\infty(\mathcal{H}_0)$. *The main reason for a construction with the supremum outside the probability is that an extension of* (1.4) *to larger functional classes* $\mathcal{H}$ *in general implies a substantial loss of approximation rate* r_n (cp. Koltchinskii ([8], Theorem 11.1). Our goal is a construction where the almost $n^{-1/2}$-rate of the original KMT result is preserved despite the passage to large functional classes $\mathcal{H}$ like Lipschitz classes.

Couplings of the type (1.3) have first been obtained by Koltchinskii ([8], Theorem 3.5) and Rio [18] for the empirical process of i.i.d. random variables, as intermediate results. They can be extended to a full functional KMT result, i.e. to a coupling in $l^\infty(\mathcal{H})$ with exponential bounds, but an additional control of the size of the functional class $\mathcal{H}$ is required, usually in terms of entropy conditions. A reduced approximation rate r_n may occur as a result.

We carry over the functional strong approximation result from the empirical process to the partial sum process under very general conditions: the distributions of X_i are allowed to be *nonidentical* and *nonsmooth*. That setting substantially complicates the task of a Hungarian construction. We can rely on the powerful methodology of Sakhanenko [19], who established the classical coupling (1.4) for nonidentical and nonsmooth summands. We stress however that for the functional version (1.3) we need to perform the construction entirely anew. Our results relate to Sakhanenko's [19] as Koltchinskii's Theorem 3.5 relates to Komlós, Major and Tusnády ([9] and [10]).

Further motivational discussion can be grouped under headings (A)–(C) below.

(A) *Statistical applications*. The Komlós–Major–Tusnády approximation has recently found an application in the asymptotic theory of statistical experiments. In [14] the classical KMT inequality for the empirical process was used to establish that a nonparametric experiment of i.i.d. observation on an interval can be approximated, in the sense of Le Cam's deficiency distance, by a sequence of signal estimation problems in Gaussian white noise. The two sequences of experiments are then asymptotically equivalent for all purposes of statistical decision with bounded loss. This appears as a generalization of Le Cam's theory of local asymptotic normality, applicable to ill-posed problems like density estimation. In particular it implies a nonparametric version of the Hàjek–Le Cam asymptotic minimax theorem. The control of the Le Cam distance is given by a relation to likelihood processes (see Le Cam and Yang [12]). Assume that there is an element $f_0 \in \Sigma$ such that the measures in the experiments $\mathcal{E}^n$ and $\mathcal{G}^n$ are absolutely continuous w.r.t. $P^n_{f_0}$ and $Q^n_{f_0}$ respectively. If there are versions $d\widetilde{P}^n_f/d\widetilde{P}^n_{f_0}$ and $d\widetilde{Q}^n_f/d\widetilde{Q}^n_{f_0}$ of the likelihood ratios $dP^n_f/dP^n_{f_0}$ and $dQ^n_f/dQ^n_{f_0}$ on a common probability space $(\Omega^n, \mathcal{F}^n, P^n)$, then

$$\Delta(\mathcal{E}^n, \mathcal{G}^n) \leqslant \sqrt{2} \sup_{f \in \Sigma} E^n_P \left(\sqrt{d\widetilde{P}^n_f/d\widetilde{P}^n_{f_0}} - \sqrt{d\widetilde{Q}^n_f/d\widetilde{Q}^n_{f_0}} \right)^2$$

(here the expected value on the right side coincides with the Hellinger distance between $\widetilde{P}^n_f$ and $\widetilde{Q}^n_f$). Thus asymptotic equivalence of experiments $\mathcal{E}^n$ and $\mathcal{G}^n$ requires a "good" coupling of the corresponding likelihood ratios $dP^n_f/dP^n_{f_0}$ and $dQ^n_f/dQ^n_{f_0}$ on a common probability space. This is achieved by constructing the linear terms (in $f - f_0$) in the expansions of the log-likelihoods such that they are close as random variables; hence the demand for an inequality (1.3) with the supremum outside the probability.

The Hungarian construction had been applied in statistics before, mostly for results on strong approximation of particular density and regression estimators (cf. Csörgő and Révész [2]). It is typical for these results that the "supremum inside the probability" is needed; for such an application of the functional KMT cf. Rio [18]. However for asymptotic equivalence of experiments, it turned out that it is sufficient, and indeed preferable, to have a coupling like (1.3) with the "supremum outside the probability". Applying theorem 3.5 of Koltchinskii [8], it became possible in [15] to extend the scope of asymptotic equivalence, for the density estimation problem, down to the limit of smoothness $1/2$. Analogously the present result is essential for establishing asymptotic equivalence of smooth nongaussian regression models to a sequence of Gaussian experiments, cf. Grama and Nussbaum [6]. The original result of Komlós, Major and Tusnády on the partial sum process [9] can be used for asymptotic equivalence in regression models, but presumably with a non-optimal smoothness limit as in [14].

(B) *Nonidentical and nonsmooth distributions*. The assumption of identically distributed r.v.'s substantially restricts the scope of application of the classical KMT inequality for partial sums. However this assumption happens to be an essential point in the original proof by Komlós, Major and Tusnády and also in much of the subsequent work. The original bound was extended and improved by many authors. Multidimensional versions were proved by Einmahl [4] and Zaitsev [22], [23] with a supremum over the class of indicators $\mathcal{H}_0$. A transparent proof of the original result was given by Bretagnolle and Massart [1]. We would like to mention the series of papers by Massart [13] and Rio [16], [17]. They treat the case of $\mathbf{R}^k$-valued r.v.'s X_i, indexed in $\mathbb{Z}_+^d$ with a supremum taken over classes $\mathcal{H}$ of indicator functions $f = \mathbf{1}_S$ of Borel sets S satisfying some regularity conditions. Condition (1.5) is also relaxed to moment assumptions, but identical distributions are still assumed.

Although there are no formal restrictions on the distributions of X_i when performing a Hungarian construction, it is not possible to get the required closeness between the constructed r.v.'s $\widetilde{X}_i \overset{d}{=} X_i$ and their normal counterparts N_i if the r.v.'s X_i are non-identically and non-smoothly distributed (see Section 4). This can be argued in the following way (see Sakhanenko [19]). Let us consider the sum $S = X_1 + \cdots + X_n$, where X_i takes values $\pm(1 + 2^{-i})$. Then we can identify each realization X_i by knowing only S. In the dyadic Hungarian scheme, the conditional distribution of $X_1 + \cdots + X_{[n/2]}$ given S is considered and used for coupling with a Gaussian random variable. However this distribution is now degenerate and hence not useful for coupling. This problem does not appear in the i.i.d. case, due to the exchangeability of the X_i.

We adopt a method to overcome this difficulty proposed by Sakhanenko [19]. In his original paper Sakhanenko treats the case of independent non-identically distributed r.v.'s for a class of indicators of intervals $\mathcal{H} = \mathcal{H}_0$. Here we consider the problem in another setting: $\mathcal{H} = \mathcal{H}(1/2, L)$ where $\mathcal{H}(1/2, L)$ is a Hölder ball with exponent $1/2$ and the sup is outside the probability, i.e. we give an exponential bound for the quantity (1.3) uniformly in f over the set of functions $\mathcal{H}(1/2, L)$. One complication which then appears is that the pairs $(\widetilde{X}_i, \widetilde{W}_i)$, $i = 1, \ldots, n$, of r.v.'s $\widetilde{X}_i \overset{d}{=} X_i$ and $\widetilde{W}_i \overset{d}{=} W_i$, $i = 1, \ldots, n$, constructed on the same probability space by the KMT method are no longer independent, even though $\widetilde{X}_i$, $i = 1, \ldots, n$, and $\widetilde{W}_i$, $i = 1, \ldots, n$, are sequences of independent r.v.'s. To deal with this we have to develop additional properties of the Hungarian construction which are not used in the classical setting (see Lemma 5.5 for details).

(C) *Coupling from marginals*. A weaker coupling of $\widetilde{X}^n$ and $\widetilde{N}^n$ can be obtained as follows. Assume for a moment that the r.v.'s X_i are uniformly bounded: $|X_i| \leqslant L$, $i = 1, \ldots, n$, and also that $\|f\|_\infty \leqslant L$, $f \in \mathcal{H}$. Take a finite collection of functions $\mathcal{H}_{00} = (f_j)_{j=1,\ldots,d} \subset \mathcal{H}$ and consider $Z_i = (f(t_i)X_i)_{f\in\mathcal{H}_{00}}$ as random vectors in $\mathbf{R}^d$. Reasoning as in Fact 2.2 of Einmahl and Mason [5] (using the result of Zaitsev [21] on the Prokhorov distance between the law of $\sum_{i=1}^n Z_i$ and a Gaussian law) we infer that for all such $\mathcal{H}_{00}$ there are versions $\widetilde{X}^n(f)$, $\widetilde{N}^n(f)$, $f \in \mathcal{H}_{00}$ (depending on x) such that

$$P\Big(n^{1/2} \max_{f\in\mathcal{H}_{00}} \big|\widetilde{X}^n(f) - \widetilde{N}^n(f)\big| \geqslant x\Big) \leqslant c_0 \exp\big(-c_1 x L^{-2}\big), \quad x \geqslant 0. \tag{1.6}$$

This yields (1.3) with rate $r_n = n^{-1/2}$ for every finite class $\mathcal{H}_{00} \subset \mathcal{H}$ of size d, but with constants c_0, c_1 depending on d. Hence any attempt to construct $\widetilde{X}^n(f)$ and $\widetilde{N}^n(f)$, on the full class $\mathcal{H}$ from (1.6) is bound to entail a substantial loss in rate r_n; but laws of the iterated logarithm can be established in this way (cf. Einmahl and Mason [5]). Thus, to obtain (1.3) for $r_n = n^{-1/2}\log^2 n$ and a full Hölder class $\mathcal{H}(1/2, L)$, the shortcut via (1.6) appears not feasible, and we revert to a direct KMT-type construction.

In order to keep the proof somewhat transparent we do not look for optimal logarithmic terms, but we believe that the optimal rate can be obtained by using the very delicate technique of the paper [19]. The main idea is, roughly speaking, to consider some *smoothed* sequences of r.v.'s instead of the initial *unsmoothed* sequence $X_1, \ldots, X_n$, and to apply the KMT construction for the smoothed sequences. This we perform by substituting normal r.v.'s N_i for the original r.v.'s X_i, for even indices $i = 2k$ in the initial sequence. Thus we are able to construct one half of our sequence and combine it with a Haar expansion of the function f. For the other half we apply the same argument which leads to a recursive procedure. It turns out that this kind of smoothing is enough to obtain "good" quantile inequalities although it gives rise to an additional $\log n$ term. On the other hand the usual smoothing technique (of each r.v. X_i individually) fails. Unfortunately even the above smoothing procedure applied with normal r.v.'s is not sufficient to obtain the best power for the $\log n$ in the KMT inequality for non-identically distributed r.v.'s. An optimal approach is developed in the paper of Sakhanenko [19] and uses r.v.'s constructed in a special way instead of the normal r.v.'s. Roughly speaking it corresponds to taking into consideration the higher terms in an asymptotic expansion for the probabilities of large deviations, which dramatically complicates the problem. For more details we refer the reader to this beautiful paper.

Nevertheless we would like to point out that the additional $\log n$ term which appears in our KMT result does not affect the eventual applications that we have in mind, i.e. asymptotic equivalence of sequences of nonparametric statistical experiments. We also believe that a stronger version of this result (with a supremum inside the probability) might be of use for constructing efficient kernel estimators in nonparametric models. But such an extension is beyond of the scope of the paper.

2. Notation and main results

Let $n \in \{1, 2, \ldots\}$. Suppose that on the probability space $(\Omega', \mathcal{F}', P')$ we are given a sequence of independent r.v.'s $X_1, \ldots, X_n$ such that

$$E' X_i = 0, \quad C_{\min} \leqslant E' X_i^2 \leqslant C_{\max}, \qquad i = 1, \ldots, n,$$

where $C_{\min} < C_{\max}$ are some positive absolute constants. Hereafter E' is the expectation under the measure P'. Assume also that the following extension of a condition due to Sakhanenko [19] holds true:

$$\lambda_n E' |X_i|^3 \exp\{\lambda_n |X_i|\} \leqslant E' X_i^2, \quad i = 1, \ldots, n, \tag{2.1}$$

where λ_n is a sequence of real numbers satisfying $0 < \lambda_n < \lambda$, $n \geqslant 1$, for some positive absolute constant λ. Along with this, assume that on another probability space $(\Omega, \mathcal{F}, P)$ we are given a sequence of independent normal r.v.'s $N_1, \ldots, N_n$ such that

$$E N_i = 0, \qquad E N_i^2 = E' X_i^2,$$

for all $i = 1, \ldots, n$. Hereafter E is the expectation under the measure P.

Let $\mathcal{H}(1/2, L)$ be the Hölder ball with exponent $1/2$, i.e. the set of real valued functions f defined on the unit interval $[0, 1]$ and satisfying the following conditions

$$\left| f(x) - f(y) \right| \leqslant L |x - y|^{1/2}, \qquad \|f\|_\infty \leqslant L/2,$$

where L is a positive absolute constant.

Let $t_i = i/n$, $i = 1, \ldots, n$, be a uniform grid in the unit interval $[0, 1]$. The notation $Y \overset{d}{=} X$ for random variables means equality in distribution. The symbol c (with possible indices) denotes a generic positive absolute constant (more precisely this means that it is a function only of the absolute constants introduced before).

The main result of the paper is the following.

THEOREM 2.1. – *Let $n \geqslant 2$. A sequence of independent r.v.'s $\widetilde{X}_1, \ldots, \widetilde{X}_n$ can be constructed on the probability space $(\Omega, \mathcal{F}, P)$ such that $\widetilde{X}_i \overset{d}{=} X_i$, $i = 1, \ldots, n$, and*

$$\sup_{f \in \mathcal{H}(1/2, L)} P\left(\left| \sum_{i=1}^n f(t_i)(\widetilde{X}_i - N_i) \right| > x \frac{\log^2 n}{\lambda_n} \right) \leqslant c_1 \exp\{-c_2 x\}, \quad x \geqslant 0.$$

Remark 2.1. – In the above theorem X_i, $i = 1, \ldots, n$, are not supposed to be identically distributed nor to have smooth distributions, although the result is new even in the case of i.i.d. r.v.'s. The r.v.'s $\widetilde{X}_1, \ldots, \widetilde{X}_n$ constructed are functions of the r.v.'s $N_1, \ldots, N_n$ only, so that no assumptions on the probability space $(\Omega, \mathcal{F}, P)$ are required other than existence of $N_1, \ldots, N_n$.

Remark 2.2. – The use of condition (2.1) instead of a more familiar Cramér type condition is motivated by the desire to cover also the case of non-identically distributed r.v.'s with subexponential moments, which corresponds to $\lambda_n \to 0$. This case cannot be treated under Cramér's condition, but it is important since it essentially includes the case of non-identically distributed r.v.'s with finite moments.

Theorem 2.1 can be formulated in the following equivalent form.

THEOREM 2.2. – *Let $n \geqslant 2$. A sequence of independent r.v.'s $\widetilde{X}_1, \dots, \widetilde{X}_n$ can be constructed on the probability space $(\Omega, \mathcal{F}, P)$ such that $\widetilde{X}_i \stackrel{d}{=} X_i$, $i = 1, \dots, n$, and for any t satisfying $|t| \leqslant c_1$*

$$\sup_{f \in \mathcal{H}(1/2,L)} E \exp\left\{ t \frac{\lambda_n}{\log^2 n} \sum_{i=1}^n f(t_i)(\widetilde{X}_i - N_i) \right\} \leqslant \exp\{c_2 t^2\}.$$

Let us formulate yet another equivalent version of Theorem 2.1. Assume that on the probability space $(\Omega', \mathcal{F}', P')$ we are given a sequence of independent r.v.'s $X_1, \dots, X_n$ such that for all $i = 1, \dots, n$

$$E' X_i = 0, \qquad \lambda_n^2 C_{\min} \leqslant E' X_i^2 \leqslant C_{\max} \lambda_n^2, \tag{2.2}$$

where $C_{\min} < C_{\max}$ are positive absolute constants and λ_n is a sequence of real numbers $0 < \lambda_n \leqslant 1$, $n \geqslant 1$. Assume also that the following condition due to Sakhanenko [19] holds true:

$$\lambda E' |X_i|^3 \exp\{\lambda |X_i|\} \leqslant E' X_i^2, \quad i = 1, \dots, n, \tag{2.3}$$

where λ is a positive absolute constant. Suppose that on another probability space $(\Omega, \mathcal{F}, P)$ we are given a sequence of independent normal r.v.'s $N_1, \dots, N_n$ such that for $i = 1, \dots, n$

$$E N_i = 0, \qquad E N_i^2 = E' X_i^2. \tag{2.4}$$

THEOREM 2.3. – *Let $n \geqslant 2$. A sequence of independent r.v.'s $\widetilde{X}_1, \dots, \widetilde{X}_n$ can be constructed on the probability space $(\Omega, \mathcal{F}, P)$ such that $\widetilde{X}_i \stackrel{d}{=} X_i$, $i = 1, \dots, n$, and for any t satisfying $|t| \leqslant c_1$*

$$\sup_{f \in \mathcal{H}(1/2,L)} E \exp\left\{ \frac{t}{\log^2 n} \sum_{i=1}^n f(t_i)(\widetilde{X}_i - N_i) \right\} \leqslant \exp\{c_2 t^2\}.$$

We shall give a proof of Theorem 2.3 in Section 6.

Now we turn to a particular case of the above results. Assume that the sequence of independent r.v.'s $X_1, \dots, X_n$ is such that

$$E' X_i = 0, \quad C_{\min} \leqslant E' X_i^2 \leqslant C_{\max}, \qquad i = 1, \dots, n, \tag{2.5}$$

for some positive absolute constants $C_{\min} < C_{\max}$. Assume also that the following Cramér type condition holds true:

$$E' \exp\{C_1 |X_i|\} \leqslant C_2, \quad i = 1, \dots, n, \tag{2.6}$$

where C_1 and C_2 are positive absolute constants.

THEOREM 2.4. – *Let $n \geqslant 2$. A sequence of independent r.v.'s $\widetilde{X}_1, \ldots, \widetilde{X}_n$ can be constructed on the probability space $(\Omega, \mathcal{F}, P)$ such that $\widetilde{X}_i \overset{d}{=} X_i$, $i = 1, \ldots, n$, and*

$$\sup_{f \in \mathcal{H}(1/2,L)} P\left(\left| \sum_{i=1}^{n} f(t_i)(\widetilde{X}_i - N_i) \right| > x \log^2 n \right) \leqslant c_1 \exp\{-c_2 x\}, \quad x \geqslant 0.$$

To deduce this result from Theorem 2.1, it suffices to note that Sakhanenko's condition (2.3) holds true with $\lambda_n = \text{const}$ depending on $C_{\min}$, C_1 and C_2, under (2.5) and (2.6).

Remark 2.3. – It should be mentioned that Sakhanenko's condition (2.3) holds true for the normal r.v.'s $N_1, \ldots, N_n$ only if the constant λ is small enough, namely if $\lambda \leqslant c\left(EN_i^2\right)^{-1/2}$. Since the function $\alpha|x|^3 \exp(\alpha|x|)$ is increasing in α, the condition (2.3) holds true for any $\lambda \leqslant \lambda'$ if it holds true with some $\lambda = \lambda'$. Therefore without loss of generality it can be assumed that the constant λ fulfills $\lambda \leqslant c/C_{\max} \leqslant c\left(E'X_i^2\right)^{-1/2}$, $i = 1, \ldots, n$, thus ensuring that (2.3) holds true also for $N_1, \ldots, N_n$.

3. Elementary properties of Haar expansions

For the following basic facts we refer to Kashin and Saakyan [7]). The Fourier–Haar basis on the interval $[0, 1]$ is introduced as follows. Consider the dyadic system of partitions by setting

$$s_{k,j} = j2^{-k},$$

for $j = 1, \ldots, 2^k$, and

$$\Delta_{k,1} = [0, s_{k,1}], \qquad \Delta_{k,j} = (s_{k,j-1}, s_{k,j}], \tag{3.1}$$

for $j = 2, \ldots, 2^k$, where $k \geqslant 0$. Define Haar functions via indicators $1(\Delta_{k,j})$

$$h_0 = 1(\Delta_{0,1}), \qquad h_{k,j} = 2^{k/2}\left(1(\Delta_{k+1,2j-1}) - 1(\Delta_{k+1,2j})\right), \tag{3.2}$$

for $j = 1, \ldots, 2^k$ and $k \geqslant 0$.

If f is a function from $\mathcal{L}_2([0, 1])$ then the following Haar expansion

$$f = c_0(f)h_0 + \sum_{k=0}^{\infty} \sum_{j=1}^{2^k} c_{k,j}(f)h_{k,j},$$

holds true with Fourier–Haar coefficients

$$c_0(f) = \int_0^1 f(u)h_0(u)\,du, \qquad c_{k,j}(f) = \int_0^1 f(u)h_{k,j}(u)\,du, \tag{3.3}$$

for $j = 1, \ldots, 2^k$ and $k \geqslant 0$. Along with this, consider the truncated Haar expansion

$$f_m = c_0(f)h_0 + \sum_{k=0}^{m-1} \sum_{j=1}^{2^k} c_{k,j}(f)h_{k,j}, \tag{3.4}$$

for some $m \geqslant 1$.

PROPOSITION 3.1. – *For $f \in \mathcal{H}(1/2, L)$ we have*

$$|c_0(f)| \leqslant L/2, \qquad |c_{k,j}(f)| \leqslant 2^{-3/2} L 2^{-k},$$

for $k = 0, 1, \ldots$ and $j = 1, \ldots, 2^k$.

Proof. – It is easy to see that

$$\begin{aligned} c_{k,j}(f) &= 2^{k/2} \left(\int_{\Delta_{k+1,2j-1}} f(u)\,du - \int_{\Delta_{k+1,2j}} f(u)\,du \right), \\ &= 2^{k/2} \int_{\Delta_{k+1,2j-1}} \left(f(u) - f\left(u + 2^{-(k+1)}\right) \right) du. \end{aligned}$$

Since f is in the Hölder ball $\mathcal{H}(\frac{1}{2}, L)$ we get

$$\begin{aligned} |c_{k,j}(f)| &\leqslant 2^{k/2} \sup_{u \in \Delta_{k+1,2j-1}} \left| f(u) - f\left(u + 2^{-(k+1)}\right) \right| \int_{\Delta_{k+1,2j-1}} du \\ &\leqslant 2^{k/2} L 2^{-(k+1)/2} 2^{-(k+1)} \leqslant 2^{-3/2} L 2^{-k}. \qquad \square \end{aligned}$$

Next we give an estimate for the uniform distance between f and f_m.

PROPOSITION 3.2. – *For $f \in \mathcal{H}(1/2, L)$ we have*

$$\sup_{0 \leqslant t \leqslant 1} |f(t) - f_m(t)| \leqslant L 2^{-m/2}.$$

Proof. – It is easy to check (see for instance Kashin and Saakyan [7], p. 81) that, whenever $t \in \Delta_{m,j}$,

$$f_m(t) = 2^m \int_{\Delta_{m,j}} f(s)\,ds,$$

for $j = 1, \ldots, 2^m$, which gives us $f_m(t) = f(\widetilde{t}_{m,j})$, with some $\widetilde{t}_{m,j} \in \Delta_{m,j}$. Since $f(t)$ is in the Hölder ball $\mathcal{H}(\frac{1}{2}, L)$, we obtain for any $j = 1, \ldots, 2^m$ and $t \in \Delta_{m,j}$

$$|f(t) - f_m(t)| = |f(t) - f_m(\widetilde{t}_{m,j})| \leqslant L|t - \widetilde{t}_{m,j}|^{1/2} \leqslant L 2^{-m/2}. \qquad \square$$

4. Background on quantile transforms

Let $(\Omega', \mathcal{F}', P')$ be a probability space. Let λ be a real number such that $0 < \lambda < \infty$. Denote by $\mathfrak{D}(\lambda)$ the set of all r.v.'s S on the probability space $(\Omega', \mathcal{F}', P')$ which can be represented as a sum $S = X_1 + \cdots + X_n$ of some independent r.v.'s on $(\Omega', \mathcal{F}', P')$ for some $n \geqslant 1$, satisfying relations (4.1), (4.2) below:

- The r.v.'s $X_1, \ldots, X_n$ have zero means and finite variances:

$$E'X_i = 0, \qquad 0 < E'X_i^2 < \infty \tag{4.1}$$

  for any $i = 1, \ldots, n$.
- Sakhanenko's condition

$$\lambda E'|X_i|^3 \exp\{\lambda |X_i|\} < E'X_i^2, \tag{4.2}$$

  is satisfied for all $i = 1, \ldots, n$.

Let μ be a real number satisfying $0 < \mu < \infty$. By $\mathfrak{D}_0(\lambda, \mu)$ we denote the subset of all r.v.'s $S \in \mathfrak{D}(\lambda)$ which additionally satisfy the following smoothness condition (4.3):

- For any $0 < \varepsilon < 1$, we have

$$\sup_{|h| \leqslant \varepsilon} \int_{|t| > \varepsilon} \left| \frac{E' \exp\{(\mathrm{i}t + h)S\}}{E' \exp\{hS\}} \right| dt \leqslant \frac{\mu}{\varepsilon E'S^2}, \tag{4.3}$$

  where $\mathrm{i} = \sqrt{-1}$.

Remark 4.1. – In the sequel we shall assume that μ is a positive absolute constant, and therefore, we shall drop the dependence on μ in the notation for $\mathfrak{D}_0(\lambda, \mu)$, i.e. we write for short $\mathfrak{D}_0(\lambda) = \mathfrak{D}_0(\lambda, \mu)$.

We now introduce the *quantile transformation* and the associated basic inequality (see Lemma 4.1). Assume that on probability space $(\Omega', \mathcal{F}', P')$ we are given an arbitrary r.v. X of mean zero and finite variance: $E'X = 0$ and $E'X^2 < \infty$. Assume that on another probability space $(\Omega, \mathcal{F}, P)$ we are given a normal r.v. N with the same mean and variance: $EN = 0$ and $EN^2 = E'X^2$. Let $F_X(x)$ and $\Phi_N(x)$ be the distribution functions of X and N respectively. Note that the r.v. $U = \Phi_N(N)$ is distributed uniformly on [0, 1]. Define the r.v. $\widetilde{X}$ to be the solution of the equation

$$F_X(\widetilde{X}) = \Phi_N(N) = U.$$

The r.v. $\widetilde{X}$ is called a quantile transformation of N. It is easy to see that a solution $\widetilde{X}$ always exists and has distribution function F, although it need not be unique. In the case of non-uniqueness, we choose one of the possible solutions.

The following assertion follows from the results in Sakhanenko [19] (see Theorem 4, p. 10).

LEMMA 4.1. – *Set $B^2 = E'X^2 = EN^2$. In addition to the above suppose that $X \in \mathfrak{D}_0(\lambda)$. Then*

$$|\widetilde{X} - N| \leqslant \frac{c_1}{\lambda} \left\{ 1 + \frac{\widetilde{X}^2}{B^2} \right\},$$

provided $|\widetilde{X}| \leqslant c_2 \lambda B^2$ and $\lambda B \geqslant c_3$, where c_1, c_2 and c_3 are positive absolute constants.

Let us now introduce the *conditional quantile transformation* and the associated basic inequality (Lemma 4.3 below).

Assume that on the probability space $(\Omega', \mathcal{F}', P')$ we are given two independent r.v.'s X_1, X_2 of means zero and finite variances: $E'X_i = 0$ and $E'X_i^2 < \infty$, for $i = 1, 2$. Assume further that on another probability space $(\Omega, \mathcal{F}, P)$ we are given two normal r.v.'s N_1, N_2 with the same means and variances: $EN_i = 0$ and $EN_i^2 = E'X_i^2$, for $i = 1, 2$. Set $X_0 = X_1 + X_2$ and $N_0 = N_1 + N_2$. Denote $B_i = E'X_i^2$, $\alpha_1 = B_1/B_2$, $\alpha_2 = B_2/B_1$. Suppose that we have constructed a $\widetilde{X}_0$ having the same distribution as X_0, and which depends only on N_0 and on some random vector W. Suppose that N_1 and N_2 do not depend on W. We wish to construct X_1 and X_2. Let $F_{T_0|X_0}(x|y)$ be the conditional distribution function of $T_0 = \alpha_2 X_1 - \alpha_1 X_2$ given $X_0 = y$ and $\Phi_{V_0}(x)$ be the distribution function of the normal r.v. $V_0 = \alpha_2 N_1 - \alpha_1 N_2$. Define $\widetilde{T}_0$ to be the solution of the equation

$$F_{T_0|X_0}(\widetilde{T}_0|\widetilde{X}_0) = \Phi_{V_0}(V_0) = U.$$

The r.v. $\widetilde{T}_0$ is called a conditional quantile transformation of V_0 given $\widetilde{X}_0$.

PROPOSITION 4.2. – *Set $\widetilde{X}_1 = \alpha_0^{-1}(T_0 + \alpha_1 \widetilde{X}_0)$ and $\widetilde{X}_2 = \alpha_0^{-1}(T_0 - \alpha_2 \widetilde{X}_0)$. Then $\widetilde{X}_1$ and $\widetilde{X}_2$ are independent and such that $\widetilde{X}_1 \stackrel{d}{=} X_1$, $\widetilde{X}_2 \stackrel{d}{=} X_2$. Moreover $\widetilde{X}_1$ and $\widetilde{X}_2$ are functions of the r.v.'s $\widetilde{X}_0$, N_1 and N_2 only.*

Proof. – Consider $U = \Phi(V_0)$. It is clear that the distribution of U is uniform on $[0, 1]$. Since $V_0 = \alpha_2 N_1 - \alpha_1 N_2$ and $N_0 = N_1 + N_2$ are normal and uncorrelated, U and N_0 are independent. Since (N_1, N_2) does not depend on W, we conclude that U does not depend on N_0 and W. But $\widetilde{X}_0$ is a function of N_0 and W only. Hence U and $\widetilde{X}_0$ are also independent.

Next, since the uniform r.v. U does not depend on $\widetilde{X}_0$, we easily check that the distribution of $\widetilde{T}_0$ given $\widetilde{X}_0 = y$, for any real y, is exactly $F_{T_0|X_0}(\cdot|y)$. Taking into account that $\widetilde{X}_0 \stackrel{d}{=} X_0$, we conclude that the two-dimensional distributions of the pairs $(\widetilde{T}_0, \widetilde{X}_0)$ and (T_0, X_0) coincide. From this we obtain in particular that $\widetilde{X}_1$ and $\widetilde{X}_2$ are independent and that $\widetilde{X}_1 \stackrel{d}{=} X_1$, $\widetilde{X}_2 \stackrel{d}{=} X_2$. Moreover it is obvious from the construction that $\widetilde{X}_1$ and $\widetilde{X}_2$ are functions of $\widetilde{X}_0$, N_1 and N_2 only. □

The following assertion follows from the results in Sakhanenko [19] (see Theorem 6, p. 20).

LEMMA 4.3. – *Set $B = B_1 B_2 / B_0$. In addition to the above suppose that $X_1, X_2 \in \mathfrak{D}_0(\lambda)$. Then*

$$|\widetilde{T}_0 - V_0| \leqslant \frac{c_1}{\lambda} \frac{B_0}{B} \left\{ 1 + \frac{\widetilde{T}_0^2}{B^2} + \frac{\widetilde{X}_0^2}{B^2} \right\},$$

provided $|\widetilde{T}_0| \leqslant c_2 \lambda B^2$, $|\widetilde{X}_0| \leqslant c_2 \lambda B^2$ and $\lambda B \geqslant c_3$, where c_1, c_2 and c_3 are absolute constants.

5. A construction for non-identically distributed r.v.'s

In this section we assume that we are given a sequence of independent r.v.'s X_i, $i = 1, \ldots, n$, satisfying (2.2) and (2.3). We shall construct a version of this sequence and an appropriate sequence of independent normal r.v.'s N_i, $i = 1, \ldots, n$, on the same probability space such that these are as close as possible. More precisely, the construction is performed so that the quantile inequalities in Section 4 are applicable. Of course the sequences which we obtain are dependent. To assure that this dependence remains under control, we partition the initial sequence into dyadic blocks with similar size of variances. Some prerequisites for this are given in the next section. The construction itself is performed in Section 5.2.

5.1. A dyadic blocking procedure

In this section we exhibit a special partition of the initial sequence into dyadic blocks so that the sums of the X_i inside the blocks at any dyadic level have approximately the same variances. This will be used for proving quantile inequalities in Section 5.4 and some exponential bounds in Section 6 (see Lemma 6.4 and Proposition 6.7).

Assume that $n > n_{\min} \geqslant 1$, where $n_{\min}$ is an absolute constant whose precise value will be indicated below. Set $M = [\log_2(n/n_{\min})]$. It is clear that $M \geqslant 0$ and $n_{\min}2^M \leqslant n < n_{\min}2^{M+1}$. Let $J_M = \{1, \ldots, n\}$ and define consecutively $J_m = \{i\colon\ 2i \in J_{m+1}\}$, for $m = 0, \ldots, M-1$. Alternatively, for any $m = 0, \ldots, M$ the set of indices J_m can be defined as follows:

$$J_m = \{i\colon\ 1 \leqslant i2^{M-m} \leqslant n\}.$$

Let n_m denote the last element in J_m i.e. $n_m = \# J_m$. It is not difficult to see that $n_{\min} \leqslant n_0 \leqslant 2n_{\min}$.

Recall that each r.v. X_i is attached to a design point $t_i = i/n$, $i = 1, \ldots, n$. Set

$$t_i^m = t_{i2^{M-m}}, \quad X_i^m = X_{i2^{M-m}}, \qquad m = 0, \ldots, M,\ i \in J_m. \tag{5.1}$$

Our next task is to split each sequence X_i^m, $i \in J_m$ into dyadic blocks so that the sums of X_i^m over blocks at a given resolution level m have approximately the same variances. To ensure this we shall introduce the strictly increasing function $b_m(t)\colon [0,1] \to [0,1]$, which is related to the variances of X_i^m as follows:

$$b_m(t) = \int_0^t \beta_m(s)\,ds \Big/ \int_0^1 \beta_m(s)\,ds, \quad t \in (0,1], \quad b_m(0) = 0,$$

where

$$\beta_m(s) = \begin{cases} E'(X_i^m)^2, & \text{if } s \in (t_{i-1}^m, t_i^m],\ i \in J_m, \\ E'(X_{n_m}^m)^2, & \text{if } s \in (t_{n_m}^n, 1]. \end{cases}$$

Let $a_m(t)$ be the inverse of $b_m(t)$, i.e.

$$a_m(t) = \inf\{s \in [0,1]\colon\ b_m(s) > t\}. \tag{5.2}$$

It is easy to see that condition (2.2) implies that both $b_m(t)$ and $a_m(t)$ are Lipschitz functions: for any $t_1, t_2 \in [0, 1]$, we have

$$|b_m(t_2) - b_m(t_1)| \leqslant L_{\max}|t_2 - t_1|, \qquad |a_m(t_2) - a_m(t_1)| \leqslant L_{\max}|t_2 - t_1|, \tag{5.3}$$

where $L_{\max} = C_{\max}/C_{\min}$. Consider the dyadic scheme of partitions

$$\Delta_{k,j}, \quad j = 1, \ldots, 2^k,\ k = 0, \ldots, M,$$

of the interval [0, 1] as defined by (3.1). For any $m = 0, \ldots, M$, denote by $I^m_{k,j}$ the set of those indices $i \in J_m$ for which $b_m(t^m_i)$ falls into $\Delta_{k,j}$, i.e.

$$I^m_{k,j} = \{i \in J_m\colon\ b_m(t^m_i) \in \Delta_{k,j}\}, \quad j = 1, \ldots, 2^k,\ k = 0, \ldots, m.$$

Since $\Delta_{k,j} = \Delta_{k+1,2j-1} + \Delta_{k+1,2j}$, it is clear that $I^m_{k,j} = I^m_{k+1,2j-1} + I^m_{k+1,2j-1}$, for $j = 1, \ldots, 2^k$. In particular $J_M = I^M_{0,1} = \{1, \ldots, n\}$. We leave to the reader to show that each set $I^m_{k,j}$ contains at least two elements, if the constant $n_{\min}$ is large enough.

PROPOSITION 5.1. – *Assume that $n_{\min} > 2C_{\max}/C_{\min} \geqslant 2$. Then for any $j = 1, \ldots, 2^k$, $k = 0, \ldots, m$, $m = 0, \ldots, M$, we have $\# I^m_{k,j} \geqslant 2$.*

In the sequel we shall assume that $n > n_{\min} \geqslant 2C_{\max}/C_{\min} \geqslant 2$. Now the sequence X^m_i, $i \in J_m$ can be split into dyadic blocks corresponding to the sets of indices $I^m_{k,j}$ as follows:

$$\{X^m_i\colon\ i \in J_m\} = \sum_{j=1}^{2^k} \{X_i\colon\ i \in I^m_{k,j}\}, \quad k = 0, \ldots, m.$$

Set

$$X^m_{k,j} = \sum_{i \in I^m_{k.j}} X^m_i, \qquad B^m_{k,j} = E'\big(X^m_{k,j}\big)^2 = \sum_{i \in I^m_{k.j}} E'\big(X^m_i\big)^2. \tag{5.4}$$

The following assertions are crucial in the proof of our results, as shall be seen later. The proofs being elementary are left to the reader.

PROPOSITION 5.2. – *For any $k = 0, \ldots, M-1$ and $j = 1, \ldots, 2^k$ we have*

$$\big|B^m_{k+1,2j-1} - B^m_{k+1,2j}\big| \leqslant c\lambda_n^2. \tag{5.5}$$

PROPOSITION 5.3. – *For any $k = 0, \ldots, M-1$ and $j = 1, \ldots, 2^k$ we have*

$$c^{-1} \leqslant B^m_{k+1,2j-1}/B^m_{k+1,2j} \leqslant c.$$

5.2. The construction

Recall that at this moment we are given just two sequences of independent r.v.'s: X_i, $i = 1, \ldots, n$, on the probability space $(\Omega', \mathcal{F}', P')$ and N_i, $i = 1, \ldots, n$, on the probability space $(\Omega, \mathcal{F}, P)$. We would like to construct a sequence of independent r.v.'s $\widetilde{X}_i$, $i = 1, \ldots, n$ on the probability space $(\Omega, \mathcal{F}, P)$ such that each $\widetilde{X}_i$ has the same distribution as X_i and the two sequences $\widetilde{X}_i$, $i = 1, \ldots, n$, and N_i, $i = 1, \ldots, n$, are as close as possible. Before proceeding with the construction we shall describe two necessary ingredients: the dyadic scheme of Komlós, Major and Tusnády [9] and an auxiliary construction.

5.2.1. The Komlós–Major–Tusnády dyadic scheme

In this section we shall describe a version of the construction appropriate for our purposes.

Let $\xi_{m,j}$, $j = 1, \ldots, 2^m$, be a sequence of r.v.'s of zero means and finite variances given on a probability space $(\Omega', \mathcal{F}', P')$, and let $\eta_{m,j}$, $j = 1, \ldots, 2^m$, be a sequence of normal r.v.'s with the same means and variances given on a probability space $(\Omega, \mathcal{F}, P)$. At this moment it is not necessary to assume that these are sequences of independent r.v.'s. The goal is to construct a version of $\xi_{m,j}$, $j = 1, \ldots, 2^m$, on the probability space $(\Omega, \mathcal{F}, P)$. The new sequence will be denoted $\widetilde{\xi}_{m,j}$, $j = 1, \ldots, 2^m$.

Set $\xi_{k,j} = \xi_{k+1,2j-1} + \xi_{k+1,2j}$ and $\eta_{k,j} = \eta_{k+1,2j-1} + \eta_{k+1,2j}$, for $j = 1, \ldots, 2^k$ and $k = 0, \ldots, m-1$. First define $\widetilde{\xi}_{0,1}$ to be the quantile transformation of $\eta_{0,1}$, i.e. define $\widetilde{\xi}_{0,1}$ to be the solution of the equation

$$F_{\xi_{0,1}}(\widetilde{\xi}_{0,1}) = \Phi_{\eta_{0,1}}(\eta_{0,1})$$

where $F_{\xi_{0,1}}(x)$ is the distribution function of $\xi_{0,1}$, and $\Phi_{\eta_{0,1}}(x)$ is the distribution function of $\eta_{0,1}$ (see Section 4). Suppose that for some $k = 0, \ldots, m-1$ the r.v.'s $\widetilde{\xi}_{k,j}$, $j = 1, \ldots, 2^k$, have already been constructed, and the goal is to construct $\widetilde{\xi}_{k+1,j}$, $j = 1, \ldots, 2^{k+1}$. To this end set for $j = 1, \ldots, 2^k$

$$V_{k,j} = \alpha_{k+1,2j}\eta_{k+1,2j-1} - \alpha_{k+1,2j-1}\eta_{k+1,2j}, \tag{5.6}$$

where

$$\alpha_{k+1,2j-1} = \left(\frac{B_{k+1,2j-1}}{B_{k+1,2j}}\right)^{1/2}, \qquad \alpha_{k+1,2j} = \left(\frac{B_{k+1,2j}}{B_{k+1,2j-1}}\right)^{1/2}$$

and

$$B_{k+1,2j-1} = E\xi^2_{k+1,2j-1}, \qquad B_{k+1,2j} = E\xi^2_{k+1,2j}.$$

Define $\widetilde{T}_{k,j}$ to be the conditional quantile transformation of $V_{k,j}$ given $\widetilde{\xi}_{k,j}$, i.e. for $j = 1, \ldots, 2^k$ define $\widetilde{T}_{k,j}$ as the solution of the equation

$$F_{T_{k,j}|\xi_{k,j}}(\widetilde{T}_{k,j}|\widetilde{\xi}_{k,j}) = \Phi_{V_{k,j}}(V_{k,j}) \tag{5.7}$$

where $F_{T_{k,j}|\xi_{k,j}}(x|y)$ is the conditional distribution function of $T_{k,j}$ given $\xi_{k,j} = y$, and $\Phi_{V_{k,j}}(x)$ is the distribution function of $V_{k,j}$ (see Section 4). For any $j = 1, \ldots, 2^k$, the desired r.v.'s $\widetilde{\xi}_{k+1,2j-1}$ and $\widetilde{\xi}_{k+1,2j}$ are defined as the solution the linear system

$$\begin{cases} \widetilde{T}_{k,j} = \alpha_{k+1,2j}\widetilde{\xi}_{k+1,2j-1} - \alpha_{k+1,2j-1}\widetilde{\xi}_{k+1,2j}, \\ \widetilde{\xi}_{k,j} = \widetilde{\xi}_{k+1,2j-1} + \widetilde{\xi}_{k+1,2j}, \end{cases} \tag{5.8}$$

the determinant of which is obviously strictly positive. This completes description of the dyadic procedure.

The following result concerns basic properties of the resulting sequence $\widetilde{\xi}_{m,j}$, $j = 1, \ldots, 2^m$.

LEMMA 5.4. – *Assume that* $\xi_{m,j}$, $j = 1, \ldots, 2^m$, *and* $\eta_{m,j}$, $j = 1, \ldots, 2^m$, *are sequences of independent r.v.'s. Then for any* $k = 0, \ldots, m$, *the r.v.'s* $\widetilde{\xi}_{k,j}$, $j = 1, \ldots, 2^k$, *are independent and such that* $\widetilde{\xi}_{k,j} \stackrel{d}{=} \xi_{k,j}$, $j = 1, \ldots, 2^k$. *Moreover* $\widetilde{\xi}_{k,j}$, $j = 1, \ldots, 2^k$, *are functions of the sequence* $\eta_{k,j}$, $j = 1, \ldots, 2^k$, *only.*

Proof. – The proof is similar to statements in Komlós, Major and Tusnády [9] (see also Sakhanenko [19], Einmahl [4], Zaitsev [24]) and therefore will not be detailed here. □

It turns out that the properties of the Komlós–Major–Tusnády dyadic construction established in Lemma 5.4 are sufficient for proving a strong approximation result if the index functions of the process belong to the class of indicators. However for proving our functional version we need one more property of this construction, which we formulate below. Recall that $V_{k,j}$ and $\widetilde{T}_{k,j}$ are defined by(5.6) and (5.8).

LEMMA 5.5. – *If* $\xi_{m,j}$, $j = 1, \ldots, 2^m$, *and* $\eta_{m,j}$, $j = 1, \ldots, 2^m$, *are sequences of independent r.v.'s, then, for any* $k = 0, \ldots, m$, *the r.v.'s* $\widetilde{T}_{k,j} - V_{k,j}$, $j = 1, \ldots, 2^k$, *are independent.*

Proof. – For the proof of this statement it suffces to note that for any $k = 0, \ldots, m$,

$$\{\widetilde{\xi}_{k,j}, V_{k,j}\colon\ j = 1, \ldots, 2^k\}$$

is a collection of of jointly independent random variables. □

5.2.2. An auxiliary construction

In the sequel we shall need also an auxiliary procedure which is not as powerful as the KMT construction, but which permits us to construct somehow the components inside an already constructed arbitrary sum of independent r.v.'s. Below we present one of the possible methods.

We start from an arbitrary sequence of r.v.'s $\xi_1, \ldots, \xi_n$ (not necessarily independent) given on $(\Omega', \mathcal{F}', P')$. Set $S_k = \xi_1 + \cdots + \xi_k$, $k = 1, \ldots, n$. Suppose that on another probability space $(\Omega, \mathcal{F}, P)$ we have constructed only the r.v. $\widetilde{S}_n \stackrel{d}{=} S_n$, which corresponds to the sum S_n and we wish to construct its components, i.e. $\widetilde{\xi}_1, \ldots, \widetilde{\xi}_n$ such that $\widetilde{\xi}_1 \stackrel{d}{=} \xi_1, \ldots, \widetilde{\xi}_n \stackrel{d}{=} \xi_n$ and $\widetilde{S}_n = \widetilde{\xi}_1 + \cdots + \widetilde{\xi}_n$. As a prerequisite we assume that on the probability space $(\Omega, \mathcal{F}, P)$ we are given a sequence of nondegenerate normal r.v.'s $\eta_1, \ldots, \eta_n$ (not necessarily independent). First we define $\widetilde{\xi}_n$ to be the conditional quantile transformation of η_n given $\widetilde{S}_n$, i.e. we define $\widetilde{\xi}_n$ to be the solution of the equation

$$F_{\xi_n|S_n}(\widetilde{\xi}_n|\widetilde{S}_n) = \Phi_{\eta_n}(\eta_n)$$

where $F_{\xi_n|S_n}(x|y)$ is the conditional distribution of ξ_n given $S_n = y$, and $\Phi_{\eta_n}(x)$ is the distribution function of η_n. Set $\widetilde{S}_{n-1} = \widetilde{S}_n - \widetilde{\xi}_n$. If for some $2 \leqslant k \leqslant n-1$ the r.v.'s

$\widetilde{\xi}_n, \dots, \widetilde{\xi}_{k+1}$ and $\widetilde{S}_k$ are already constructed, we define $\widetilde{\xi}_k$ to be the conditional quantile transformation of η_k given $\widetilde{S}_k$, i.e. we define $\widetilde{\xi}_k$ to be the solution of the equation

$$F_{\xi_k|S_k}(\widetilde{\xi}_k|\widetilde{S}_k) = \Phi_{\eta_k}(\eta_k)$$

where $F_{\xi_k|S_k}(x|y)$ is the conditional distribution of ξ_k given $S_k = y$, and $\Phi_{\eta_k}(x)$ is the distribution function of η_k. Set $\widetilde{S}_{k-1} = \widetilde{S}_k - \widetilde{\xi}_k$. Finally, for $k = 1$, we define $\widetilde{\xi}_1 = \widetilde{S}_1$, this completing our procedure.

The easy proof of the following assertion is left to the reader.

LEMMA 5.6. – *Assume that $\xi_1, \dots, \xi_n$ and $\eta_1, \dots, \eta_n$ are sequences of independent r.v.'s. Then $\widetilde{\xi}_1, \dots, \widetilde{\xi}_n$ are independent, $\widetilde{\xi}_i \stackrel{d}{=} \xi_i$, $i = 1, \dots, n$, and $\widetilde{\xi}_1 + \dots + \widetilde{\xi}_n = \widetilde{S}_n$. Moreover $\widetilde{\xi}_1, \dots, \widetilde{\xi}_n$ are functions of $\eta_1, \dots, \eta_n$ and $\widetilde{S}_n$ only.*

5.2.3. The main construction

Our next step is to describe a construction which will result in the desired sequence $\widetilde{X}_i$, $i = 1, \dots, n$. It should be noted that although both the dyadic procedure and the auxiliary construction described above work with arbitrary distributions, in order to use the quantile inequalities stated in Section 4 (which actually will provide the desired closeness of $\widetilde{X}_i$, $i = 1, \dots, n$, and N_i, $i = 1, \dots, n$), one has to assume the r.v.'s X_i, $i = 1, \dots, n$, to be in the class $\mathfrak{D}_0(r)$, for some $r > 0$, or to be identically distributed (as in Komlós, Major and Tusnády [9], [10]). In order to avoid such assumptions we shall employ an inductive procedure which goes back to the paper of Sakhanenko [19]. The idea is first to substitute the initial sequence with some smoothed sequences, and then to apply the dyadic procedure described in Section 5.2.1 to the smoothed sequences. Below we formally describe this construction.

Consider the product probability space $(\Omega'', \mathcal{F}'', P'') = (\Omega', \mathcal{F}', P') \times (\Omega, \mathcal{F}, P)$, where $P'' = P' \times P$. It is obvious that the sequences X_i, $i = 1, \dots, n$, and N_i, $i = 1, \dots, n$, are independent on the probability space $(\Omega'', \mathcal{F}'', P'')$.

Recall that above we introduced the sets of indices $J_m = \{i\colon\ 1 \leqslant i2^{M-m} \leqslant n\}$. For each $m = M, \dots, 0$, the set J_m can be decomposed as $J_m = J_m^1 + J_m^2$, where

$$J_m^1 = \{i\text{-odd: } i \in J_m\}, \quad J_m^2 = \{i\text{-even: } i \in J_m\}, \qquad m = M, \dots, 1,$$

and $J_0^1 = J_0$, $J_0^2 = \emptyset$. It is clear that

$$J_{m-1} = \{i\colon\ 2i \in J_m\}, \quad m = 1, \dots, M.$$

To start our iterative construction, for any $i \in J_M = \{1, \dots, n\}$, define the following r.v.'s:

$$X_{2i}^{M+1} = X_i, \qquad \widetilde{Y}_{2i}^{M+1} = N_i. \tag{5.9}$$

We proceed to describe the m-th step of our construction which is performed consecutively for all $m = M, \dots, 0$.

• *mth step*. For any $i \in J_m$, define the following r.v.'s:

$$X_i^m = X_{2i}^{m+1}, \qquad W_i^m = \widetilde{Y}_{2i}^{m+1}, \tag{5.10}$$

and

$$Y_i^m = \begin{cases} X_i^m, & \text{if } i \in J_m^1, \\ W_i^m, & \text{if } i \in J_m^2. \end{cases} \tag{5.11}$$

Note that the r.v.'s W_i^m, $i \in J_m$, are defined on the probability space $(\Omega, \mathcal{F}, P)$, while the r.v.'s Y_i^m, $i \in J_m$, are defined on the probability space $(\Omega'', \mathcal{F}'', P'')$. Here X_i^m, $i \in J_m$, is the part of the initial sequence X_i, $i \in J_M = \{1, \ldots, n\}$, given on the probability space $(\Omega', \mathcal{F}', P')$, which is not yet constructed on the probability space $(\Omega, \mathcal{F}, P)$; W_i^m, $i \in J_m$, is the corresponding sequence of normal r.v.'s and Y_i^m, $i \in J_m$ is the smoothed sequence which is constructed at this step. Consider the following sums: for $j = 1, \ldots, 2^k$ and $k = 0, \ldots, m$ set

$$Y_{k,j}^m = \sum_{i \in I_{k,j}^m} Y_i^m, \qquad W_{k,j}^m = \sum_{i \in I_{k,j}^m} W_i^m. \tag{5.12}$$

Then obviously for $j = 1, \ldots, 2^k$ and $k = 0, \ldots, m-1$

$$Y_{k,j}^m = Y_{k+1,2j-1}^m + Y_{k+1,2j}^m, \qquad W_{k,j}^m = W_{k+1,2j-1}^m + W_{k+1,2j}^m. \tag{5.13}$$

We will apply the dyadic procedure described in Section 5.2.1, with $\xi_{m,j} = Y_{m,j}^m$ and $\eta_{m,j} = W_{m,j}^m$, $j = 1, \ldots, 2^m$, to construct a doubly indexed sequence $\widetilde{Y}_{kj}^m$, $j = 1, \ldots, 2^m$, $k = 0, \ldots, m$. Let $\widetilde{Y}_{0,1}^m$ be the quantile transformation of $W_{0,1}^m$, i.e. let $\widetilde{Y}_{0,1}^m$ be the solution of the equation

$$F_{Y_{0,1}^m}\big(\widetilde{Y}_{0,1}^m\big) = \Phi_{W_{0,1}^m}\big(W_{0,1}^m\big), \tag{5.14}$$

where $F_{Y_{0,1}^m}(x)$ is the distribution function of $Y_{0,1}^m$ and $\Phi_{W_{0,1}^m}(x)$ is the distribution function of $W_{0,1}^m$. The solution exists since $W_{0,1}^m$ is a nondegenerate normal r.v. Assume that we have already constructed $\widetilde{Y}_{k,j}^m$, $j = 1, \ldots, 2^k$, for some $k = 0, \ldots, m-1$. We shall construct such an array with $k+1$ replacing k. To this end set, for $j = 1, \ldots, 2^k$,

$$V_{k,j}^m = \alpha_{k+1,2j}^m W_{k+1,2j-1}^m - \alpha_{k+1,2j-1}^m W_{k+1,2j}^m, \tag{5.15}$$

where

$$\alpha_{k+1,2j-1}^m = \left(\frac{B_{k+1,2j-1}^m}{B_{k+1,2j}^m}\right)^{1/2}, \qquad \alpha_{k+1,2j}^m = \left(\frac{B_{k+1,2j}^m}{B_{k+1,2j-1}^m}\right)^{1/2}$$

and

$$B_{k+1,2j-1}^m = E\big(\widetilde{Y}_{k+1,2j-1}^m\big)^2, \qquad B_{k+1,2j}^m = E\big(\widetilde{Y}_{k+1,2j}^m\big)^2. \tag{5.16}$$

Let $\widetilde{T}_{k,j}^m$ be the conditional quantile transformation of $V_{k,j}^m$, given $\widetilde{Y}_{k,j}^m$, for $j = 1, \ldots, 2^k$, i.e. let $\widetilde{T}_{k,j}^m$ be the solution of the equation

$$F_{T_{k,j}^m | Y_{k,j}^m}\big(\widetilde{T}_{k,j}^m \big| \widetilde{Y}_{k,j}^m\big) = \Phi_{V_{k,j}^m}\big(V_{k,j}^m\big), \tag{5.17}$$

where $F_{T_{k,j}^m | Y_{k,j}^m}(x|y)$ is the conditional distribution function of $T_{k,j}^m$, given $Y_{k,j}^m$, and $\Phi_{W_{k,j}^m}(x)$ is the distribution function of $W_{k,j}^m$. The solution exists, since $V_{k,j}^M$ is a nondegenerate normal r.v. For any $j = 1, \ldots, 2^k$ we define the desired r.v.'s $\widetilde{Y}^m_{k+1,2j-1}$ and $\widetilde{Y}^m_{k+1,2j}$ as the solution of the linear system

$$\begin{cases} \widetilde{T}^m_{k,j} = \alpha^m_{k+1,2j} \widetilde{Y}^m_{k+1,2j-1} - \alpha^m_{k+1,2j-1} \widetilde{Y}^m_{k+1,2j}, \\ \widetilde{Y}^m_{k,j} = \widetilde{Y}^m_{k+1,2j-1} + \widetilde{Y}^m_{k+1,2j}. \end{cases} \tag{5.18}$$

Thus the r.v.'s $\widetilde{Y}^m_{k,j}$, $j = 1, \ldots, 2^k$, are constructed for all $k = 0, \ldots, m$ on the probability space $(\Omega, \mathcal{F}, P)$. It remains to construct the components inside each sum $\widetilde{Y}^m_{m,j}$, $j = 1, \ldots, 2^m$. For this we make use of the auxiliary construction described in Section 5.2.2, with $\xi_i \equiv Y^m_i$ and $\eta_i \equiv W^m_i$, $i \in I^m_{m,j}$. For each fixed j and m it provides a sequence of r.v.'s $\widetilde{Y}^m_i \equiv \widetilde{\xi}_i$, $i \in I^m_{m,j}$, such that

$$\widetilde{Y}^m_{m,j} = \sum_{i \in I^m_{m,j}} \widetilde{Y}^m_i. \tag{5.19}$$

This completes the mth step of our construction.

Let us recall briefly some notation associated with the construction, which will also be used in the sequel. For any $m = M, \ldots, 0$ we have defined the r.v.'s Y^m_i, W^m_i, $\widetilde{Y}^m_i$, $i \in J_m$, and $Y^m_{k,j}$, $W^m_{k,j}$, $\widetilde{Y}^m_{k,j}$, $j = 1, \ldots, 2^k$, $k = 0, \ldots, m$, such that, by (5.12) and (5.19) (cp. with (5.4)),

$$Y^m_{k,j} = \sum_{i \in I^m_{k,j}} Y^m_i, \qquad W^m_{k,j} = \sum_{i \in I^m_{k,j}} W^m_i, \qquad \widetilde{Y}^m_{k,j} = \sum_{i \in I^m_{k,j}} \widetilde{Y}^m_i, \tag{5.20}$$

for $k = 0, \ldots, m$, $j = 1, \ldots, 2^k$, $m = 0, \ldots, M$.

5.3. Correctness and some useful properties

In fact implicitly the construction of the desired sequence $\widetilde{X}_i$, $i = 1, \ldots, n$, has already been carried out; it remains to select the appropriate components from the sequences $\{\widetilde{Y}^m_i\colon i \in J_m\}$ found above. But before this step we need to show that the construction is performed correctly, and we shall also discuss some properties of the r.v.'s $\widetilde{Y}^m_i$ and W^m_i introduced. The proofs of the following assertions are left to the reader.

In analogy to X^m_i (see (5.1)), set $N^m_i = N_{i2^{M-m}}$, where $m = 0, \ldots, M$, $i \in J_m$.

LEMMA 5.7. – *For any $m = 0, \ldots, M$ the following statements hold true*:

(a) *The r.v.'s W^m_i, $i \in J_m$, are independent and satisfy $W^m_i \overset{d}{=} N^m_i = N_{i2^{M-m}}$, $i \in J_m$.*

(b) *The r.v.'s $\widetilde{Y}^m_i$, $i \in J_m$, are independent, are functions of W^m_i, $i \in J_m$, only and satisfy, for $i \in J_m$,*

$$\widetilde{Y}^m_i \overset{d}{=} Y^m_i \overset{d}{=} \begin{cases} X^m_i & \text{if } i \in J^1_m, \\ N^m_i & \text{if } i \in J^2_m. \end{cases}$$

Remark 5.1. – Since by Proposition 5.1 $\# I^m_{k,j} \geqslant 2$, from Lemma 5.7 and from (5.20) it follows that $W^m_{k,j}$, $j = 1, \ldots, 2^k$, $k = 0, \ldots, m$, are nondegenerate normal r.v.'s which ensures that the solutions of Eqs. (5.14), (5.17) exist. This proves the correctness of the main construction.

PROPOSITION 5.8. – *The vectors* $\{\widetilde{Y}_i^m\colon i \in J_m^1\}$, $m = M, \ldots, 0$, *are independent.*

Now finally we are able to present the sequence $\widetilde{X}_i$, $i = 1, \ldots, n$. It is defined on the probability space $(\Omega, \mathcal{F}, P)$ in the following way:

$$\widetilde{X}_{i2^{M-m}} = \widetilde{Y}_i^m, \quad \text{where } i \in J_m^1,\ 0 \leqslant m \leqslant M. \tag{5.21}$$

PROPOSITION 5.9. – $\widetilde{X}_i$, $i = 1, \ldots, n$, *are independent and such that* $\widetilde{X}_i \stackrel{d}{=} X_i$, $i = 1, \ldots, n$.

Proof. – The required assertion follows from Lemma 5.7 and Proposition 5.8. □

In the proof of our main result Theorem 2.3, the following elementary representation is essential. Recall that $t_i^m = t_\nu = \nu/n$ where $\nu = i2^{M-m}$, $i \in J_m$, $m = 0, \ldots, M$ (see Section 5.1).

PROPOSITION 5.10. – *For any real valued function* $f(t)$ *on the interval* $[0, 1]$, *we have*

$$\sum_{i=1}^{n} f(t_i)(\widetilde{X}_i - N_i) = \sum_{m=0}^{M} \sum_{j \in J_m} f\left(t_j^m\right)\left(\widetilde{Y}_j^m - W_j^m\right).$$

5.4. Quantile inequalities

In this section we shall establish so-called quantile inequalities (see Lemma 5.12 and Lemma 5.13), which will ensure the required closeness of the r.v.'s $\widetilde{X}_i$, $i = 1, \ldots, n$, and N_i, $i = 1, \ldots, n$.

The following lemma shows that the r.v.'s $Y_{k,j}^m$, $j = 1, \ldots, 2^k$, are smooth enough to allow application of the quantile inequalities stated in Section 4.

LEMMA 5.11. – *For* $m = 0, \ldots, M$, $k = 0, \ldots, m$, $j = 1, \ldots, 2^k$ *the r.v.* $Y_{k,j}^m$ *is in the class* $\mathfrak{D}(r)$, *for some positive absolute constant* r.

Proof. – We shall check conditions (4.1), (4.2) and (4.3) in Section 4. Toward this end fix m, k, j as in the condition of the lemma and note that

$$\zeta_0 \equiv Y_{k,j}^m = \sum_{i \in I_{k,j}^m} Y_i^m = \sum_{i \in I_1} Y_i^m + \sum_{i \in I_2} Y_i^m \equiv \zeta_1 + \zeta_2,$$

where I_1 and I_2 are the sets of all odd and even indices in $I_{k,j}^m$ respectively. By Lemma 5.7, we have $Y_i^m \stackrel{d}{=} N_i$, for any $i \in I_2$. Thus ζ_2 is actually a sum of independent normal r.v.'s. Since $n_{\min}$ is large enough, the set $I_{k,j}^m$ has at least two elements (see Proposition 5.1), from which we conclude that I_2 has at least one element. Next, taking into account (2.2) and the obvious inequality $\#I_2 \geqslant \frac{1}{3}\#I_{k,j}^m$, we get

$$E\zeta_2^2 \geqslant C_{\min}\lambda_n \#I_2 \geqslant \frac{C_{\min}}{3}\lambda_n \#I_{k,j}^m \geqslant cE\zeta_0^2.$$

For $|h| \leqslant \lambda$ and $t \in R$, let

$$f_{\zeta_i}(t,h) = E\exp\{(\mathrm{i}t+h)\zeta_i\}/E\exp\{h\zeta_i\}$$

be the conjugate characteristic function of the r.v. ζ_i, $i=0,1,2$. Since ζ_1 and ζ_2 are independent and ζ_2 is normal,

$$\begin{aligned}\left|f_{\zeta_0}(t,h)\right| &= \left|f_{\zeta_1}(t,h)f_{\zeta_2}(t,h)\right| \leqslant \left|f_{\zeta_2}(t,h)\right|\\ &\leqslant \exp\left\{-\frac{t^2}{2}E\zeta_2^2\right\} \leqslant \exp\left\{-\frac{t^2}{2}cE\zeta_0^2\right\},\end{aligned}$$

for $|h| \leqslant \lambda$, $t \in \mathbf{R}^1$. With this bound we have

$$\int_{|t|>\varepsilon} \left|f_{\zeta_0,h}(t)\right| dt \leqslant \int_{|t|>\varepsilon} \exp\left\{-\frac{t^2}{2}cE\zeta_0^2\right\} dt \leqslant \frac{\mu}{\varepsilon E\zeta_0^2},$$

where μ is some absolute constant, which proves that $\zeta_0 = Y_{k,j}^m$ satisfies condition (4.3).

It remains only to show that conditions (4.1) and (4.2) are satisfied. The first condition follows from (2.2) as soon as $Y_i^m \overset{d}{=} X_i$ or $Y_i^m \overset{d}{=} N_i$ for any $i \in I_{k,j}^m \subseteq J_m$, by Lemma 5.7. For the second we make use of (2.3) and of the elementary fact that Sakhanenko's condition (4.2) holds true for any normal r.v. N if λ is small enough: $\lambda \leqslant c(\mathrm{Var}N)^{-1/2}$ (see Remark 2.3). □

Recall that for any $m = 0,\dots,M$, $k=1,\dots,m$ and $j=1,\dots,2^k$, by (5.18),

$$\widetilde{T}_{k,j}^m = \alpha_{k,2j}^m \widetilde{Y}_{k,2j-1}^m - \alpha_{k,2j-1}^m \widetilde{Y}_{k,2j}^m, \tag{5.22}$$

and by (5.15),

$$V_{k,j}^m = \alpha_{k,2j}^m W_{k,2j-1}^m - \alpha_{k,2j-1}^m W_{k,2j}^m. \tag{5.23}$$

Recall also that $B_{k,j}^m = E'(X_{k,j}^m)^2$ (see (5.4)).

The following quantile inequalities show that the r.v.'s $\widetilde{T}_{k,j}^m$ and $W_{k,j}^m$ are close enough. These statements are crucial for our results.

LEMMA 5.12. – *For any $m = 0,\dots,M$, we have*

$$|\widetilde{Y}_{0,1}^m - W_{0,1}^m| \leqslant c_1\left\{1 + \frac{(\widetilde{Y}_{0,1}^m)^2}{B_{0,1}^m}\right\},$$

provided $|\widetilde{Y}_{0,1}^m| \leqslant c_2 B_{0,1}^m$ and $B_{0,1}^m \geqslant c_3$, where c_1, c_2 and c_3 are positive absolute constants.

Proof. – According to the construction, $\widetilde{Y}_{0,1}^m$ is the quantile transformation of $W_{0,1}^m$ (see (5.14)). Then it suffices to note that, by Lemma 5.11, the r.v. $\widetilde{Y}_{0,1}^m$ is in the class $\mathfrak{D}(\lambda_0)$ and to apply Lemma 4.1 with $X = Y_{0,1}^m$, $N = W_{0,1}^m$ and $\widetilde{X} = \widetilde{Y}_{0,1}^m$. □

LEMMA 5.13. – *Let* $m = 0, \ldots, M$, $k = 0, \ldots, m-1$, $j = 1, \ldots, 2^k$. *Then*

$$|\widetilde{T}_{k,j}^m - V_{k,j}^m| \leqslant c_1 \left\{ 1 + \frac{(\widetilde{Y}_{k+1,2j-1}^m)^2}{B_{k+1,2j-1}^m} + \frac{(\widetilde{Y}_{k+1,2j}^m)^2}{B_{k+1,2j}^m} \right\},$$

provided $|\widetilde{Y}_{k+1,2j-1}^m| \leqslant c_2 B_{k+1,2j-1}^m$, $|\widetilde{Y}_{k+1,2j}^m| \leqslant c_2 B_{k+1,2j}^m$ *and* $B_{k+1,2j-1}^m \geqslant c_3$, $B_{k+1,2j}^m \geqslant c_3$, *where* c_1, c_2 *and* c_3 *are positive absolute constants.*

Proof. – Fix m, k, and j as in the condition of the lemma. We are going to make use of Lemma 4.3 with

$$\widetilde{X}_1 = \widetilde{Y}_{k+1,2j-1}^m, \qquad \widetilde{X}_2 = \widetilde{Y}_{k+1,2j}^m, \qquad \widetilde{X}_0 = \widetilde{X}_1 + \widetilde{X}_2 = \widetilde{Y}_{k,j}^m, \tag{5.24}$$

and

$$N_1 = W_{k+1,2j-1}^m, \qquad N_2 = W_{k+1,2j}^m, \qquad N_0 = N_1 + N_2 = W_{k,j}^m. \tag{5.25}$$

Note that, by Lemma 5.11, the r.v.'s $\widetilde{X}_0$, $\widetilde{X}_1$ and $\widetilde{X}_2$ are in the class $\mathfrak{D}(r)$ for some absolute constant $r > 0$. Since by construction $\widetilde{T}_{k,j}^m$ is the conditional quantile transformation of $V_{k,j}^m$ (see (5.17)), Lemma 4.3 implies

$$|\widetilde{T}_{k,j}^m - V_{k,j}^m| \leqslant c_1 \frac{B_0}{B} \left\{ 1 + \frac{1}{B^2} (\widetilde{X}_1^2 + \widetilde{X}_2^2) \right\}, \tag{5.26}$$

provided

$$|\widetilde{T}_{k,j}^m| \leqslant c_2 B^2, \qquad |\widetilde{Y}_{k,j}^m| \leqslant c_2 B^2, \tag{5.27}$$

and $B \geqslant c_3$, where

$$B_1^2 = B_{k+1,2j-1}^m, \qquad B_2^2 = B_{k+1,2j}^m, \qquad B_0^2 = B_1^2 + B_2^2, \qquad B^2 = \frac{B_1 B_2}{B_0}.$$

By Proposition 5.3, we have

$$c_4^{-1} \leqslant B_1^2 / B_2^2 \leqslant c_4. \tag{5.28}$$

Now we check that (5.27) holds true if $|\widetilde{X}_1| \leqslant c_5 B_1^2$ and $|\widetilde{X}_2| \leqslant c_5 B_2^2$, where c_5 is a sufficiently small constant. Indeed

$$|\widetilde{T}_{k,j}^m| \leqslant \frac{B_2}{B_1} |\widetilde{X}_1| + \frac{B_1}{B_2} |\widetilde{X}_2| \leqslant 2 c_5 B_1 B_2.$$

By (5.28), we get $|\widetilde{T}_{k,j}^m| \leqslant c_5 c_6 B^2$. Choosing the constant c_6 such that $c_5 c_6 \leqslant c_2$, we see that (5.27) is satisfied. Exactly in the same way we show that the second inequality in (5.27) holds true. Condition $B \geqslant c_3$ follows easily from (5.28). □

6. Proof of the main results

6.1. An auxiliary exponential bound

We keep the same notation as in the previous section. In addition set for brevity

$$\widetilde{S}_0^m = \widetilde{Y}_{0,1}^m - W_{0,1}^m, \quad \widetilde{S}_{k,j}^m = \widetilde{T}_{k,j}^m - V_{k,j}^m, \qquad j=1,\ldots,2^k,\ k=0,\ldots,m, \tag{6.1}$$

where $\widetilde{T}_{k,j}^m$ and $V_{k,j}^m$ are defined by (5.22) and (5.23). The main result of this section is Lemma 6.1 which establishes an exponential type bound for the differences $\widetilde{S}_{k,j}^m$ and $\widetilde{S}_0^m$. Because of the special construction of $\widetilde{T}_{k,j}^m$ and $V_{k,j}^m$ on the same probability space, this bound is much better that the usual exponential bounds (cf. Lemma 6.3 below). This statement plays a crucial role in establishing our functional version of the Hungarian construction. It is the only place where the quantile inequalities are used.

LEMMA 6.1. – *For any $m=0,\ldots,M$, $k=0,\ldots,m-1$, $j=1,\ldots,2^k$,*

$$E\exp\{t\widetilde{S}_0^m\} \leqslant \exp\{c_1t^2\}, \quad E\exp\{t\widetilde{S}_{k,j}^m\} \leqslant \exp\{c_1t^2\}, \qquad |t|\leqslant c_0.$$

We postpone the proof of the lemma to the end of this section; it will be based on some estimates stated and proved below.

LEMMA 6.2. – *For any $\varepsilon>0$ there is a constant $c(\varepsilon)$ depending only on ε, such that for any $m=0,\ldots,M$, $k=0,\ldots,m$ and $j\in J_k$,*

$$P\big(\big|\widetilde{Y}_{k,j}^m\big| > \varepsilon B_{k,j}^m\big) \leqslant 2\exp\{-c(\varepsilon)B_{k,j}^m\}.$$

Proof. – By Chebyshev's inequality, we have for $t>0$

$$P\big(\widetilde{Y}_{k,j}^m > \varepsilon B_{k,j}^m\big) \leqslant \exp\{-t\varepsilon B_{k,j}^m\}E\exp\{tY_{k,j}^m\}. \tag{6.2}$$

Note that by (5.20) and by Lemma 5.7, the r.v. $\widetilde{Y}_{k,j}^m$ is the sum of independent r.v.'s $\widetilde{Y}_i^m$, $i\in I_{k,j}^m$. Then by (2.3) and Lemma A.1, we obtain for $|t|\leqslant\lambda/3$,

$$E\exp\{t\widetilde{Y}_{k,j}^m\} = \prod_{i\in I_{k,j}^m} E\exp\{t\widetilde{Y}_i^m\} \leqslant \exp\{t^2B_{k,j}^m\}.$$

Inserting this bound into (6.2), with an appropriate choice of t (depending on ε), we get

$$E\big(\widetilde{Y}_{k,j}^m > \varepsilon B_{k,j}^m\big) \leqslant \exp\{-c(\varepsilon)B_{k,j}^m\}.$$

In the same way one can show that

$$E\big(\widetilde{Y}_{k,j}^m < -\varepsilon B_{k,j}^m\big) \leqslant \exp\{-c(\varepsilon)B_{k,j}^m\},$$

which in conjunction with the previous bound proves the lemma. □

LEMMA 6.3. – *Let* $m = 0, \ldots, M$, $k = 0, \ldots, m-1$, $j = 1, \ldots, 2^k$. *Then for any* $0 \leqslant t \leqslant c_1$ *we have*

$$E\exp\{t|\widetilde{S}^m_{k,j}|\} \leqslant c_2 \exp\{t^2 B^m_{k,j}\}.$$

Proof. – Fix m, k and j as in the condition of the lemma. From (6.1) and from the Hölder inequality one gets, for $0 \leqslant t \leqslant \lambda$,

$$E\exp\{t|\widetilde{S}^m_{k,j}|\} \leqslant \left(E\exp\{t|\widetilde{T}^m_{k,j}|\} E\exp\{t|V^m_{k,j}|\}\right)^{1/2}. \tag{6.3}$$

The r.v. $\widetilde{Y}^m_{k+1,2j-1}$ and $\widetilde{Y}^m_{k+1,2j}$ are independent, hence by (5.22)

$$E\exp\{t|\widetilde{T}^m_{k,j}|\} \leqslant E\exp\{t\alpha^m_{k+1,2j}|\widetilde{Y}^m_{k+1,2j-1}|\} E\exp\{t\alpha^m_{k+1,2j-1}|\widetilde{Y}^m_{k+1,2j}|\}. \tag{6.4}$$

Since by (5.20) and by Lemma 5.7, $\widetilde{Y}^m_{k+1,2j-1}$ is exactly the sum of independent r.v.'s $\widetilde{Y}^m_i$, $i \in I^m_{k+1,2j-1}$, one has

$$E\exp\{\pm t\alpha^m_{k+1,2j}\widetilde{Y}^m_{k+1,2j-1}\} = \prod_{i \in I^m_{k+1,2j-1}} E\exp\{\pm t\alpha^m_{k+1,2j}\widetilde{Y}^m_i\}.$$

Taking into account (2.3) and choosing t small enough ($t \leqslant \lambda/3$), by Lemma A.1 one obtains

$$\begin{aligned} E\exp\{\pm t\alpha^m_{k+1,2j}\widetilde{Y}^m_{k+1,2j-1}\} &\leqslant \prod_{i \in J^m_{k+1,2j-1}} E\exp\{t^2(\alpha^m_{k+1,2j})^2 E(X^m_i)^2\} \\ &\leqslant \exp\{t^2(\alpha^m_{k+1,2j})^2 B^m_{k+1,2j-1}\}. \end{aligned}$$

Since $(\alpha^m_{k+1,2j})^2 = B^m_{k+1,2j}/B^m_{k+1,2j-1}$,

$$E\exp\{t\alpha^m_{k+1,2j}|\widetilde{Y}^m_{k+1,2j-1}|\} \leqslant 2\exp\{t^2 B^m_{k+1,2j}\}.$$

For the second expectation on the right hand side of (6.4) one gets an analogous bound. Then

$$E\exp\{t|\widetilde{T}^m_{k,j}|\} \leqslant 4\exp\{t^2 B^m_{k+1,2j} + t^2 B^m_{k+1,2j-1}\} = 4\exp\{t^2 B^m_{k,j}\}. \tag{6.5}$$

A similar bound holds for the second expectation on the right-hand side of (6.3), i.e.

$$E\exp\{t|V^m_{k,j}|\} \leqslant 4\exp\{t^2 B^m_{k,j}\}. \tag{6.6}$$

Now the lemma follows from (6.5), (6.6) and (6.3). □

Now we are prepared to show that $\widetilde{S}^m_{k,j}$ has a bounded exponential moment uniformly in m, k and j.

LEMMA 6.4. – *For any* $m = 0, \ldots, M$, $k = 0, \ldots, m-1$, $j = 1, \ldots, 2^k$

$$E\exp\{c_1|\widetilde{S}^m_{k,j}|\} \leqslant c_2.$$

Proof. – Fix m, k and j as in the condition of the lemma. It is enough to consider the case where $B^m_{k+1,2j-1}$ and $B^m_{k+1,2j}$ are greater than c' only, where c' is the absolute constant c_3 in Lemma 5.13; otherwise, by Proposition 5.3, we have $B^m_{k+1,2j-1}, B^m_{k+1,2j} \leqslant c_1$ (thus $B^m_{k,j} = B^m_{k+1,2j-1} + B^m_{k+1,2j} \leqslant 2c_1$) and the claim follows from Lemma 6.3.

Set for brevity

$$G^m_{k+1,l} = \{|\widetilde{Y}^m_{k+1,l}| \leqslant c'' B^m_{k+1,l}\}, \quad l = 2j-1, 2j, \tag{6.7}$$

where $c'' = \min\{1, c_2\}$ and c_2 is the absolute constant in Lemma 5.13. Denote by $G^{m,c}_{k+1,l}$ the complement of the set $G^m_{k+1,l}$. It is easy to see that, for $0 \leqslant t \leqslant \lambda$,

$$E \exp\{t|\widetilde{S}^m_{k,j}|\} = Q_1 + Q_2, \tag{6.8}$$

where

$$Q_1 = E \exp\{t|\widetilde{S}^m_{k,j}|\}\mathbf{1}(G^{m,c}_{k+1,2j-1} \cup G^{m,c}_{k+1,2j-1}), \tag{6.9}$$

$$Q_2 = E \exp\{t|\widetilde{S}^m_{k,j}|\}\mathbf{1}(G^{m}_{k+1,2j-1} \cap G^{m}_{k+1,2j-1}). \tag{6.10}$$

First we give an estimate for Q_1. Applying Hölder's inequality, we obtain from (6.9),

$$Q_1 \leqslant (\exp\{2t|\widetilde{S}^m_{k,j}|\})^{1/2}(P(G^{m,c}_{k+1,2j-1})^{1/2} + P(G^{m,c}_{k+1,2j})^{1/2}). \tag{6.11}$$

By Lemma 6.2 we have with $l = 2j-1, 2j$

$$P(G^{m,c}_{k+1,l}) = P(|\widetilde{Y}^m_{k+1,l}| > c'' B^m_{k+1,l}) \leqslant 2\exp\{-c_2 B^m_{k+1,l}\}. \tag{6.12}$$

Note that by Proposition 5.3, we have $c_3^{-1} \leqslant B^m_{k+1,2j-1}/B^m_{k+1,2j} \leqslant c_3$, which implies $B_{k+1,l} \geqslant c_4 B_{k,j}$ for $l = 2j-1, 2j$. Then from (6.12) it follows that

$$P(G^{m,c}_{k+1,l}) \leqslant 2\exp\{-c_5 B^m_{k,j}\}, \quad l = 2j-1, 2j. \tag{6.13}$$

Inserting the bound provided by Lemma 6.3 and the inequality (6.13) into (6.11) and choosing t sufficiently small we obtain

$$Q_1 \leqslant c_6 \exp\{(c_7 t^2 - c_8)B^m_{k,j}\} \leqslant c_6 \exp\left\{-\frac{1}{2}c_8 B^m_{k,j}\right\} \leqslant c_6.$$

Now we shall give a bound for Q_2. Recall that the r.v.'s $\widetilde{Y}^m_{k+1,l}$, $l = 2j-1, 2j$ are smooth (belong to the class $\mathfrak{D}(r)$), by Lemma 5.11. By virtue of Lemma 5.13 and of the assumption $B^m_{k+1,2j-1} \geqslant c'$ and $B^m_{k+1,2j} \geqslant c'$, on the set $G^m_{k+1,2j-1} \cap G^m_{k+1,2j}$ we have

$$|\widetilde{S}^m_{k,j}| \leqslant c_9\{1 + U^m_{k+1,2j-1} + U^m_{k+1,2j}\}, \tag{6.14}$$

where for $l = 2j-1, 2j$

$$U^m_{k+1,l} = (\widetilde{Y}^{m,*}_{k+1,l})^2/B^m_{k+1,l}, \qquad \widetilde{Y}^{m,*}_{k+1,l} = \widetilde{Y}^m_{k+1,l}\mathbf{1}(|\widetilde{Y}^m_{k+1,l}| \leqslant B^m_{k+1,l}).$$

According to (6.10) and (6.14)

$$Q_2 \leqslant E\exp\{tc_{10}(1+U^m_{k+1,2j-1}+U^m_{k+1,2j})\}$$
$$=\exp\{tc_{10}\}E\exp\{tc_{10}U^m_{k+1,2j-1}\}E\exp\{tc_6U^m_{k+1,2j}\}. \tag{6.15}$$

By Lemma A.3 (see Appendix A) we have

$$E\exp\{c_{10}U^m_{k+1,2j-1}\}\leqslant 1+2/c_{10} \tag{6.16}$$

and a similar bound holds true for $U^m_{k+1,2j-1}$. Taking t sufficiently small, from (6.15) and (6.16) we obtain $Q_2\leqslant c_{11}$. Combining the estimates $Q_1\leqslant c_6$ and $Q_2\leqslant c_{11}$ obtained above with (6.8) yields the lemma. □

LEMMA 6.5. – *For any* $m=0,\ldots,M$

$$E\exp\{c_1|\widetilde{S}^m_0|\}\leqslant c_2.$$

Proof. – The argument is similar to that for Lemma 6.4, and therefore will not be given here. The only difference is that instead of Lemma 5.13 we make use of Lemma 5.12. □

Now Lemma 6.1 follows easily from Lemmas 6.4, 6.5 and Lemma A.1 in Appendix A.

6.2. Proof of Theorem 2.3

The idea of the proof is to decompose the function f into a Haar expansion and then to make use of the closeness properties of the sequences $\widetilde{X}_i$, $i=1,\ldots,n$, and N_i, $i=1,\ldots,n$, over the dyadic blocks. For this the representation provided by Proposition 5.10 and the exponential inequalities in Lemma 6.1 are crucial.

For the sake of brevity set

$$S_n(f)=\sum_{i=1}^n f(t_i)(\widetilde{X}_i-N_i).$$

What we have to show is that for any t satisfying $|t|\leqslant c_0$,

$$E\exp\{t(\log n)^{-2}S_n(f)\}\leqslant\exp\{t^2c_1\}. \tag{6.17}$$

Toward this end let $M=[\log_2(n/n_0)]$ and note that according to Proposition 5.10,

$$S_n(f)=\sum_{m=0}^M S^m \quad \text{where } S^m=\sum_{i\in J_m} f(t^m_i)(\widetilde{Y}^m_i-W^m_i).$$

By Hölder's inequality

$$E\exp\{t(\log n)^{-2}S_n(f)\}\leqslant\prod_{m=0}^M\left(E\exp\{t(M+1)(\log n)^{-2}S^m\}\right)^{1/(M+1)}. \tag{6.18}$$

Set for brevity

$$u_n = (M+1)(\log n)^{-2}. \tag{6.19}$$

Obviously $u_n \leqslant 1$ for n large enough (such that $\log n \geqslant 2$).

It is easy to see that inequality (6.17) follows from (6.18) if we prove that for $m = 0, \ldots, M$ and any t satisfying $|t| \leqslant c_0$

$$E\exp\{tu_n S^m\} \leqslant \exp\{t^2 c_1\}. \tag{6.20}$$

In the sequel we will give a proof of (6.20).

First we consider the case $m = 0$. By Hölder's inequality,

$$E\exp\{tu_n S^0\} \leqslant \left(E\exp\left\{2tu_n \sum_{i\in J_0} f\left(t_i^0\right)\widetilde{Y}_i^0\right\} E\exp\left\{2tu_n \sum_{i\in J_0} f\left(t_i^0\right)W_i^0\right\}\right)^{1/2}. \tag{6.21}$$

Since $\widetilde{Y}_i^0$, $i \in J_0$, are independent we have

$$E\exp\left\{2tu_n \sum_{i\in J_0} f(t_i^0)\widetilde{Y}_i^0\right\} = \prod_{i\in J_0} E\exp\{2tu_n f(t_i^0)\widetilde{Y}_i^0\}.$$

By choosing the constant c_0 small enough we can easily guarantee that $|2tu_n f(t_i^0)| \leqslant \lambda/3$, and by Lemma A.1 we obtain

$$E\exp\left\{2tu_n \sum_{i\in J_0} f\left(t_i^0\right)\widetilde{Y}_i^0\right\} \leqslant \exp\left\{c_2 t^2 \sum_{i\in J_0} E\left(\widetilde{Y}_i^0\right)^2\right\}. \tag{6.22}$$

Since $E(\widetilde{Y}_i^0)^2 = E'(X_{i2^M})^2 \leqslant C_{\max}$ for $i \in J_0$, and $\#J_0 \leqslant 2n_{\min}$ (see Section 5.1), we have $\sum_{i\in J_0} E(\widetilde{Y}_i^0)^2 \leqslant c_3$, which in conjunction with (6.22) yields

$$E\exp\left\{2tu_n \sum_{i\in J_0} f\left(t_i^0\right)\widetilde{Y}_i^0\right\} \leqslant \exp\{c_4 t^2\}.$$

An analogous bound holds true for the second expectation in (6.21). From these bounds and from (6.21) we obtain (6.20) for $m = 0$.

For the case $m \geqslant 1$ introduce the function $g(s) = f(a(s))$, $s \in [0,1]$, where $a(s)$ is defined by (5.2). Set for brevity $s_i^m = b(t_i^m)$, $i \in J_m$. Then for the sum S^m we get the following representation:

$$S^m = \sum_{i\in J_m} g\left(s_i^m\right)\left(\widetilde{Y}_i^m - W_i^m\right).$$

Let g_m be the truncated Haar expansion of g for $m \geqslant 1$ (see (3.4):

$$g_m = c_0(g)h_0 + \sum_{k=0}^{m-1}\sum_{j=1}^{2^k} c_{k,j}(g)h_{k,j}, \tag{6.23}$$

where $c_0(g)$ and $c_{k,j}(g)$ are the corresponding Fourier–Haar coefficients defined by (3.3) with g replacing f. Then obviously

$$S^m = S_1^m + S_2^m,$$

where

$$S_1^m = \sum_{i\in J_m} \big(g(s_i^m) - g_m(s_i^m)\big)\big(\widetilde{Y}_i^m - W_i^m\big),$$
$$S_2^m = \sum_{i\in J_m} g_m(s_i^m)\big(\widetilde{Y}_i^m - W_i^m\big). \tag{6.24}$$

By Hölder's inequality

$$E\exp\{tu_nS^m\} \leqslant \big(E\exp\{2tu_nS_1^m\}E\exp\{2tu_nS_2^m\}\big)^{1/2}. \tag{6.25}$$

Now the inequality (6.20) for $m \geqslant 1$ will be established if we prove that both expectations on the right-hand side of (6.25) are bounded by $\exp\{t^2c\}$. These inequalities are the subject of Propositions 6.6 and 6.7 below. This completes the proof of Theorem 2.3.

First we prove the bound for the first expectation on the right hand side of (6.25).

PROPOSITION 6.6. – *For any $m = 1, \ldots, M$ and t satisfying $|t| \leqslant c_0$ we have*

$$E\exp\{tu_nS_1^m\} \leqslant \exp\{t^2c_1\}.$$

Proof. – Since by (5.3) the function $a(s)$ is Lipschitz and $f \in \mathcal{H}(\frac{1}{2}, L)$, it is easy to see that the function $g(s) = f(a(s))$ is also in a Hölder ball $\mathcal{H}(\frac{1}{2}, L_0)$ but with another absolute constant L_0. By Hölder's inequality

$$E\exp\{tu_nS_1^m\} \leqslant \left(E\exp\left\{\sum_{i\in J_m}\rho_i\widetilde{Y}_i^m\right\}E\exp\left\{-\sum_{i\in J_m}\rho_iW_i^m\right\}\right)^{1/2}, \tag{6.26}$$

where $\rho_i = 2tu_n(g(s_i^m) - g_m(s_i^m))$ and $|t| \leqslant c_0$ for some sufficiently small absolute constant c_0. Note that by Proposition 3.2 we have $\|g - g_m\|_\infty \leqslant L_0 2^{-m/2}$. Therefore for $|t| \leqslant c_0$ (where c_0 is small)

$$|\rho_i| \leqslant c_1|t|u_n2^{-m/2} \leqslant c_1|t|2^{-m/2} \leqslant \lambda/3.$$

Then according to Lemma A.1 we get for $i \in J_m$

$$E\exp\{\rho_i\widetilde{Y}_i^m\} \leqslant \exp\{\rho_i^2E(\widetilde{Y}_i^m)^2\} \leqslant \exp\{c_2t^22^{-m}E(X_i^m)^2\}. \tag{6.27}$$

An analogous bound holds true for the normal r.v.'s W_i^m, $i \in J_m$:

$$E\exp\{-\rho_iW_i^m\} \leqslant \exp\{c_2t^22^{-m}E(X_i^m)^2\}. \tag{6.28}$$

Taking into account that $\widetilde{Y}_i^m$, $i \in J_m$, and W_i^m, $i \in J_m$, are sequences of independent r.v.'s and inserting (6.27) and (6.28) into (6.26), we obtain

$$E \exp\{t u_n S_1^m\} \leqslant \exp\left\{c_3 t^2 2^{-m} \sum_{i \in J_m} E\left(X_i^m\right)^2\right\}. \tag{6.29}$$

Now we remark that $\# J_m \leqslant 2^{m+1}$. Hence by (2.2)

$$\sum_{i \in J_m} E\left(X_i^m\right)^2 \leqslant \# J_m C_{\max} \leqslant 2^{m+1} C_{\max}. \tag{6.30}$$

Inserting (6.30) into (6.29), we obtain the result. □

Now we will find the bound for the second expectation on the right hand side of (6.25).

PROPOSITION 6.7. – *For any $m = 1, \ldots, M$ and t satisfying $|t| \leqslant c_0$ we have*

$$E \exp\{t u_n S_2^m\} \leqslant \exp\{t^2 c_1\}.$$

Proof. – From (6.24), (6.23) and (3.2) we obtain

$$S_2^m = c_0(g)\left(\widetilde{Y}_{0,1}^m - W_{0,1}^m\right) + \sum_{k=0}^{m-1} 2^{k/2} \sum_{j=1}^{2^k} c_{k,j}(g)\left(T_{k,j}^{*,m} - V_{k,j}^{*,m}\right)$$

where

$$T_{k,j}^{*,m} = \widetilde{Y}_{k+1,2j-1}^m - \widetilde{Y}_{k+1,2j}^m, \qquad V_{k,j}^{*,m} = W_{k+1,2j-1}^m - W_{k+1,2j}^m \tag{6.31}$$

(compare with (5.22) and (5.23)). Here $\widetilde{Y}_{k,j}^m$ and $W_{k,j}^m$ are defined by (5.20). Set in analogy to (6.1)

$$S_0^m = \widetilde{Y}_{0,1}^m - W_{0,1}^m, \qquad S_{k,j}^m = T_{k,j}^{*,m} - V_{k,j}^{*,m}, \quad j = 1, \ldots, 2^k,\ k = 0, \ldots, m-1. \tag{6.32}$$

Since the function $g(s)$ is in the Hölder ball with a Hölder constant L_0, according to Proposition 3.1 we have the following bounds for the Fourier–Haar coefficients:

$$c_0(g) \leqslant L_0/2, \qquad \left|c_{k,j}(g)\right| \leqslant 2^{-3/2} L_0 2^{-k}, \quad j = 1, \ldots, 2^k,\ k = 0, \ldots, m-1. \tag{6.33}$$

Note also that by Lemma 6.1 there is an absolute constant t_0 sufficiently small such that for $|v| \leqslant t_0$

$$E \exp\{v \widetilde{S}_0^m\} \leqslant \exp\{c_1 v^2\}, \qquad E \exp\{v \widetilde{S}_{k,j}^m\} \leqslant \exp\{c_1 v^2\} \tag{6.34}$$

for $j = 1, \ldots, 2^k$ and $k = 0, \ldots, m-1$, where $\widetilde{S}_0^m$ and $\widetilde{S}_{k,j}^m$ are defined by (6.1).

By Hölder's inequality we have, for any t satisfying $|t| \leqslant c_0 \leqslant t_0$,

$$E \exp\{t u_n S_2^m\} \leqslant \left(E \exp\{t(m+1) u_n c_0(g) S_0^m\} \prod_{k=0}^{m-1} E \exp\{t(m+1) u_n U_k\}\right)^{1/(m+1)}, \tag{6.35}$$

where

$$U_k = 2^{k/2} \sum_{j=1}^{2^k} c_{k,j}(g) S_{k,j}^m, \quad k = 0, \ldots, m-1. \tag{6.36}$$

The claim will be established, if we show that the constant c_0 can be chosen such that for t satisfying $|t| \leqslant c_0$,

$$E \exp\{t u_{m,n} c_0(g) S_0^m\} \leqslant \exp\{c_2 t^2\} \tag{6.37}$$

and

$$E \exp\{t u_{m,n} U_k\} \leqslant \exp\{c_2 t^2\}, \tag{6.38}$$

where for the sake of brevity we set $u_{m,n} = (m+1)u_n$.

It is easy to show (6.37). For this we note that by (6.33) and (6.19), for $|t| \leqslant c_0$ we have

$$\left|t u_{m,n} c_0(g)\right| \leqslant c_3 |t| (m+1)(M+1) L_0 / \log^2 n \leqslant c_4 c_0 \leqslant t_0, \tag{6.39}$$

if the constant c_0 is small enough. Then the inequality (6.37) follows from (6.34) and from (6.39).

The proof of (6.38) is somewhat more involved. The main problem is that $S_{k,j}^m$, $j = 1, \ldots, 2^k$, are dependent and therefore we cannot make use of the product structure of the exponent $\exp\{tU_k\}$ directly. However Proposition 5.2 ensures that the components of the sum U_k (see (6.36)) are *almost* independent, which allows to exploit the product structure in an implicit way. The main idea is to "substitute" $S_{k,j}^m$, $j = 1, \ldots, 2^k$, by $\widetilde{S}_{k,j}^m$, $j = 1, \ldots, 2^k$, which are independent. With this in mind we write

$$U_k = U_k^1 + U_k^2,$$

where

$$U_k^1 = 2^{k/2} \sum_{j=1}^{2^k} c_{k,j}(g) \widetilde{S}_{k,j}^m, \qquad U_k^2 = 2^{k/2} \sum_{j=1}^{2^k} c_{k,j}(g) \left(S_{k,j}^m - \widetilde{S}_{k,j}^m\right).$$

Then by Hölder's inequality,

$$E \exp\{t u_{m,n} U_k\} \leqslant \left(E \exp\{2t u_{m,n} U_k^1\} E \exp\{2t u_{m,n} U_k^2\}\right)^{1/2}. \tag{6.40}$$

Now we proceed to estimate the first expectation on the right-hand side of (6.40). We make use of the independence of $\widetilde{S}_{k,j}^m$, $j = 1, \ldots, 2^k$ (see Lemma 5.5), to get

$$E \exp\{2t u_{m,n} U_k^1\} = \prod_{j=1}^{2^k} E \exp\{t q_j \widetilde{S}_{k,j}^m\}, \tag{6.41}$$

where $q_j = q_{m,n,k,j} = 2u_{m,n} 2^{k/2} c_{k,j}(g)$. Note that by (6.33) and (6.19)

$$|t q_j| \leqslant \left|2t u_{m,n} 2^{k/2} c_{k,j}(g)\right| \leqslant c_5 |t| 2^{-k/2} \leqslant t_0,$$

provided c_0 is small enough. It then follows from (6.34) that for $j=1,\dots,2^k$

$$E\exp\{tq_j\widetilde{S}^m_{k,j}\}\leqslant\exp\{c_6t^2 2^{-k}\}. \tag{6.42}$$

Inserting (6.42) into (6.41) we find the bound

$$E\exp\{2tu_{m,n}U^1_k\}\leqslant\exp\{c_7t^2\}. \tag{6.43}$$

Thus we have estimated the first expectation on the right hand side of (6.40). It remains to estimate the second one.

Note that

$$\widetilde{S}^m_{k,j}-S^m_{k,j}=\big(\widetilde{T}^m_{k,j}-T^{*,m}_{k,j}\big)-\big(V^m_{k,j}-V^{*,m}_{k,j}\big).$$

Hence

$$U^2_k=U^3_k+U^4_k,$$

where

$$U^3_k=2^{k/2}\sum_{j=1}^{2^k}c_{k,j}(g)\big(\widetilde{T}^m_{k,j}-T^{*,m}_{k,j}\big),$$

$$U^4_k=2^{k/2}\sum_{j=1}^{2^k}c_{k,j}(g)\big(V^m_{k,j}-V^{*,m}_{k,j}\big).$$

By Hölder's inequality we obtain

$$E\exp\{2tu_{m,n}U^2_k\}\leqslant\big(E\exp\{4tu_{m,n}U^3_k\}E\exp\{4tu_{m,n}U^4_k\}\big)^{1/2}. \tag{6.44}$$

Since $\widetilde{T}^m_{k,j}-T^{*,m}_{k,j}$, $j=1,\dots,2^k$, is a sequence of independent r.v.'s, we get

$$E\exp\{4tu_{m,n}U^3_k\}\leqslant\prod_{j=1}^{2^k}E\exp\{2tq_j\big(\widetilde{T}^m_{k,j}-T^{*,m}_{k,j}\big)\} \tag{6.45}$$

where q_j is defined above (see (6.41)). The definitions of $\widetilde{T}^m_{k,j}$ and of $T^{*,m}_{k,j}$ (see (5.22) and (6.31)) imply

$$\widetilde{T}^m_{k,j}-T^{*,m}_{k,j}=\beta_{2j}\widetilde{Y}^m_{k+1,2j-1}-\beta_{2j-1}\widetilde{Y}^m_{k+1,2j}.$$

Hereafter we abbreviate $\beta_i=\alpha^m_{k+1,i}-1$, $B_i=B^m_{k+1,i}$. Then

$$\begin{aligned}E\exp\{2tq_j\big(\widetilde{T}^m_{k,j}-T^{*,m}_{k,j}\big)\}&=E\exp\{tq_j\beta_{2j}\widetilde{Y}^m_{k+1,2j-1}\}\\&\quad\times E\exp\{-tq_j\beta_{2j-1}\widetilde{Y}^m_{k+1,2j}\}.\end{aligned} \tag{6.46}$$

Since by Proposition 5.3 $B_{2j} \leqslant c_8 B_{2j-1}$, we have $\beta_{2j} \leqslant 1 + c_8$. Hence by (6.33) and (6.19)

$$|tq_j\beta_{2j}| \leqslant c_9|t|2^{-k/2}\beta_{2j} \leqslant \lambda/3 \tag{6.47}$$

for t sufficiently small. By (5.20) and by Lemma 5.7, $\widetilde{Y}^m_{k+1,2j-1}$ is a sum of independent r.v.'s which satisfy Sakhanenko's condition (2.3). Hence using Lemma A.1 we obtain

$$\begin{aligned} E\exp\{tq_j\beta_{2j}\widetilde{Y}^m_{k+1,2j-1}\} &= \prod_{i\in I^m_{k+1,2j-1}} E\exp\{tq_j\beta_{2j}\widetilde{Y}^m_i\} \\ &\leqslant \prod_{i\in I^m_{k+1,2j-1}} \exp\{t^2q_j^2\beta_{2j}^2E(\widetilde{Y}^m_i)^2\}. \end{aligned}$$

By (6.47)

$$\begin{aligned} E\exp\{tq_j\beta_{2j}\widetilde{Y}^m_{k+1,2j-1}\} &\leqslant \prod_{i\in I^m_{k+1,2j-1}} \exp\{c_{10}t^2 2^{-k/2}\beta_{2j}^2E(\widetilde{Y}^m_i)^2\} \\ &= \exp\{c_{10}t^2 2^{-k/2}\beta_{2j}^2B_{2j-1}\}. \end{aligned}$$

Taking into account Proposition 5.2, we obtain

$$\beta_{2j}^2B_{2j-1} = \left(\sqrt{B_{2j}} - \sqrt{B_{2j-1}}\right)^2 \leqslant |B_{2j} - B_{2j-1}| \leqslant c_{11}.$$

This proves that

$$E\exp\{tq_j\beta_{2j}\widetilde{Y}^m_{k+1,2j-1}\} \leqslant \exp\{c_{12}t^2 2^{-k/2}\}.$$

For the second expectation on the right hand side of (6.46) we prove an analogous bound. Invoking these bounds in (6.46) we get

$$E\exp\{2tq_j(\widetilde{T}^m_{k,j} - T^{*,m}_{k,j})\} \leqslant \exp\{c_{13}t^2 2^{-k/2}\}. \tag{6.48}$$

Inserting in turn (6.48) into (6.45) we arrive at

$$E\exp\{4tu_{m,n}U_k^3\} \leqslant \exp\{c_{14}t^2\}.$$

In the same way we prove an inequality for U_k^4. Then by (6.44) we have

$$E\exp\{2tu_{m,n}U_k^2\} \leqslant \exp\{c_{14}t^2\}. \tag{6.49}$$

From (6.40), (6.49) and (6.43) we obtain inequality (6.38), this completing the proof of the proposition. □

Acknowledgement

We would like to thank A. Zaitsev for useful discussions and the referee for remarks which helped to improve the presentation of the paper.

Appendix A

In the course of the reasoning we made use of the following simple auxiliary results.

LEMMA A.1. – *Let* ξ *be a real valued r.v. with mean* 0 *and finite variance*: $E\xi = 0$, $0 < E\xi^2 < \infty$. *Assume that Sakhanenko's condition*

$$\lambda E|\xi|^3 \exp\{\lambda|\xi|\} \leqslant E\xi^2$$

holds true for some $\lambda > 0$. *Then for all* $|t| \leqslant \lambda/3$

$$E \exp\{t\xi\} \leqslant \exp\{t^2 E\xi^2\}.$$

Proof. – Let $\mu(t) = E \exp(t\xi)$ and $\psi(t) = \log \mu(t)$ be the moment and cumulant generating functions respectively. The conditions of the lemma imply that $\mu(t) \leqslant c_1$ for any real $|t| \leqslant \lambda/3$. Using a three term Taylor expansion we obtain for $0 \leqslant \nu \leqslant 1$

$$\psi(t) = \psi(0) + \psi'(0)t + \psi''(0)\frac{t^2}{2} + \psi'''(\nu t)\frac{t^3}{6}.$$

Note that $\psi(0) = 0$, $\psi'(0) = 0$, $\psi''(0) = E\xi^2$ and $\mu(t) \geqslant 1$ by Jensen's inequality, while for the third derivative we have for any real s satisfying $|s| \leqslant \lambda/3$,

$$\psi'''(s) = \mu'''(s)\mu(s)^{-1} - 3\mu''(s)\mu'(s)\mu(s)^{-2} + 2\mu'(s)^3\mu(s)^{-3}.$$

Using Hölder's inequality and $\mu(s) \geqslant 1$ we obtain the bound

$$\left|\psi'''(s)\right| \leqslant 6E|\xi|^3 \exp\left(\lambda|\xi|\right).$$

Since $|t| \leqslant \lambda/3$, by Sakhanenko's condition we have

$$0 \leqslant \psi(t) \leqslant \frac{t^2}{2}E\xi^2 + t^3 E|\xi|^3 \exp\left(\lambda|\xi|\right) \leqslant t^2 E\xi^2. \qquad \square$$

LEMMA A.2. – *Let* ξ *be a real valued r.v. such that* $E\xi = 0$ *and*

$$E \exp\{\lambda|\xi|\} \leqslant c_1$$

for some $\lambda \geqslant 0$ *and* $c_1 \geqslant 1$. *Then for all* $|t| \leqslant \lambda/2$ *we have*

$$E \exp\{t\xi\} \leqslant \exp\{c_2 t^2\},$$

where $c_2 = 4c_1/\lambda^2$.

Proof. – The argument is similar to Lemma A.1. We use the same notations. A two term Taylor expansion yields, for $0 \leqslant \nu \leqslant 1$,

$$\psi(t) = \psi(0) + \psi'(0)t + \psi''(\nu t)\frac{t^2}{2}.$$

Since $x^2 \leqslant 2\exp(|x|)$ for any real x, we have for any s satisfying $|s| \leqslant \lambda/2$

$$0 \leqslant \psi''(s) = \mu(s)^{-2}\left\{E\xi^2\exp(s\xi) - \left(E\xi\exp(s\xi)\right)^2\right\}$$
$$\leqslant E\xi^2\exp(s\xi) \leqslant E\xi^2\exp\left(\frac{\lambda}{2}|\xi|\right) \leqslant 8\frac{c_1}{\lambda^2}.$$

Consequently

$$0 \leqslant \psi(t) = \psi''(\nu t)\frac{t^2}{2} \leqslant 4\frac{c_1}{\lambda^2}t^2. \qquad \square$$

LEMMA A.3. – *Let ξ_i, $i = 1, \ldots, n$, be a sequence of independent r.v.'s such that for all $i = 1, \ldots, n$ we have $E\xi_i = 0$, $0 < E\xi_i^2 < \infty$ and*

$$\lambda E|\xi_i|^3\exp\{\lambda|\xi_i|\} \leqslant E\xi_i^2$$

for some positive constant λ. Set $S_n = \xi_1 + \cdots + \xi_n$, $B_n^2 = ES_n^2$ and $S_n^* = S_n\mathbf{1}(|S_n| \leqslant B_n^2)$. Then

$$E\exp\{c_1(S_n^*/B_n)^2\} \leqslant 1 + 2/c_1,$$

where $c_1 = \frac{1}{4}\min\{\lambda/3, 1/2\}$.

Proof. – Denote

$$F(x) = P\left((S_n^*/B_n)^2 > x\right).$$

First we shall prove that

$$F(x) \leqslant 2\exp\{-c_2x\}, \quad x \geqslant 0, \tag{A.1}$$

where $c_2 = 2c_1$. For this we note that

$$F(x) = P\left(S_n^*/B_n > \sqrt{x}\right) + P\left(S_n^*/B_n < -\sqrt{x}\right).$$

It suffices to estimate only the first probability on the right hand side of the above equality; the second can be treated in the same way. If $x > B_n^2$ then

$$P\left(S_n^*/B_n > \sqrt{x}\right) = 0;$$

thus there is nothing to prove in this case. Let $x \leqslant B_n^2$. Denoting $t = 2c_2\sqrt{x}$, one obtains

$$P\left(S_n^* > \sqrt{x}\right) \leqslant P\left(S_n > \sqrt{x}\right) \leqslant \exp\{-t\sqrt{x}\}E\exp\{tS_n/B_n\}$$
$$= \exp\{-t\sqrt{x}\}\prod_{i=1}^n E\exp\{t\xi_i/B_n\}. \tag{A.2}$$

Note that $t/B_n = 2c_2\sqrt{x}/B_n \leqslant 2c_2 \leqslant \lambda/3$. Hence by Lemma A.1

$$E\exp\{t\xi_i/B_n\} \leqslant \exp\{t^2E\xi_i^2/B_n^2\}.$$

Inserting this into (A.2) we get

$$P\left(S_n^*/B_n > \sqrt{x}\right) \leqslant \exp\left\{-t\sqrt{x}\right\} \prod_{i=1}^{n} \exp\left\{t^2 E\xi_i^2/B_n^2\right\}$$
$$= \exp\left\{-t\sqrt{x} + t^2\right\} \leqslant \exp\{-c_2 x\}$$

which proves (A.1). Integrating by parts we obtain

$$E\exp\left\{c_1(S_n^*)^2/B_n\right\} = \int_0^\infty \exp\{c_1 x\}\, dF(x)$$
$$= 1 + \int_0^\infty F(x)\exp\{c_1 x\}\, dx$$
$$\leqslant 1 + 2\int_0^\infty \exp\{c_1 x - c_2 x\}\, dx$$
$$\leqslant 1 + 2/c_1.$$

REFERENCES

[1] J. Bretagnolle, P. Massart, Hungarian constructions from the nonasymptotic viewpoint, Ann. Probab. 17 (1989) 239–256.

[2] M. Csörgő, P. Révész, Strong Approximations in Probability and Statistics, Academic Press, New York, 1981.

[3] R. Dudley, Real Analysis and Probability, Wadsworth and Brooks/Cole, Pacific Grove, CA, 1989.

[4] U. Einmahl, Extensions on results of Komlós, Major and Tusnády to the multivariate case, J. Multivariate Anal. 28 (1989) 20–68.

[5] U. Einmahl, D.M. Mason, Gaussian approximation of local empirical processes indexed by functions, Probab. Theory Related Fields 107 (1997) 283–311.

[6] I. Grama, M. Nussbaum, Asymptotic equivalence for nonparametric generalized linear models, Probab. Theory Related Fields 111 (1998) 167–214.

[7] B.S. Kashin, A.A. Saakyan, Orthogonal Series, American Mathematical Society, Providence, RI, 1989.

[8] V.I. Koltchinskii, Komlós–Major–Tusnády approximation for the general empirical process and Haar expansions of classes of functions, J. Theoret. Probab. 7 (1994) 73–118.

[9] J. Komlós, P. Major, G. Tusnády, An approximation of partial sums of independent rv's and the sample df. I, Z. Wahrsch. verw. Gebiete 32 (1975) 111–131.

[10] J. Komlós, P. Major, G. Tusnády, An approximation of partial sums of independent rv's and the sample df. II, Z. Wahrsch. verw. Gebiete 34 (1976) 33–58.

[11] L. Le Cam, Asymptotic Methods in Statistical Decision Theory, Springer-Verlag, New York, 1986.

[12] L. Le Cam, G.L. Yang, Asymptotics in Statistics: Some Basic Concepts, 2nd Ed., Springer-Verlag, New York, 2000.

[13] P. Massart, Strong approximation for multivariate empirical and related processes, via KMT constructions, Ann. Probab. 17 (1989) 266–291.

[14] M. Nussbaum, Asymptotic equivalence of density estimation and white noise, Preprint No. 35, Institute of Applied Analysis and Stochastics, Berlin, 1993.
[15] M. Nussbaum, Asymptotic equivalence of density estimation and Gaussian white noise, Ann. Statist. 24 (1996) 2399–2430.
[16] E. Rio, Strong approximation for set-indexed partial sum processes, via K.M.T. constructions I, Ann. Probab. 21 (1993) 759–790.
[17] E. Rio, Strong approximation for set-indexed partial sum processes, via K.M.T. constructions II, Ann. Probab. 21 (1993) 1706–1727.
[18] E. Rio, Local invariance principles and their applications to density estimation, Probab. Theory Related Fields 98 (1994) 21–45.
[19] A. Sakhanenko, The rate of convergence in the invariance principle for non-identically distributed variables with exponential moments, Limit theorems for sums of random variables. Trudy Inst. Matem., Sibirsk. Otdel. AN SSSR. 3 (1984) 3–49 (in Russian).
[20] A. van der Vaart, J. Wellner, Weak Convergence and Empirical Processes, Springer-Verlag, New York, 1996.
[21] A. Zaitsev, On the Gaussian approximation of convolutions under multidimensional analogues of S.N. Bernstein's inequality conditions, Probab. Theory Related Fields 74 (1987) 534–566.
[22] A. Zaitsev, Estimates for quantiles of smooth conditional distributions and multidimensional invariance principle, Siberian Math. J. 37 (1996) 807–831.
[23] A. Zaitsev, Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments, ESAIM: P & S 2 (1998) 41–108.
[24] A. Zaitsev, Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. I, Theory Probab. Appl. 45 (2001) 624–641.