A FUNCTIONAL CLT FOR NONCONVENTIONAL POLYNOMIAL ARRAYS

YEOR HAFOUTA

DEPARTMENT OF MATHEMATICS
THE OHIO STATE UNIVERSITY AND THE HEBREW UNIVERSITY

ABSTRACT. In this paper we will prove a functional central limit theorem (CLT) for random functions of the form

\[ S_N(t) = N^{-\frac{1}{2}} \sum_{n=1}^{[Nt]} F(\xi_{q_1(n,N)}, \xi_{q_2(n,N)}, \ldots, \xi_{q_\ell(n,N)}) \]

where the \( q_i \)'s are certain type of bivariate polynomials, \( F = F(x_1, \ldots, x_{\ell}) \) is a locally Hölder continuous function and the sequence of random variables \( \{\xi_n\} \) satisfies some mixing and moment conditions. This paper continues the line of research started in \[15\] and \[17\], and it is a generalization of the results in \[8\] and Chapter 3 of \[10\]. We will also prove a strong law of large numbers (SLLN) for the averages \( N^{-\frac{1}{2}} S_N(1) \) which extends the results from the beginning of Chapter 3 of \[10\] to general bivariate polynomial functions \( q_i \). Our results hold true for sequences \( \{\xi_n\} \) generated by a wide class of Markov chains and dynamical systems. As an application we obtain functional CLT’s for expressions of the form \( N^{-\frac{1}{2}} M([Nt]) \), where \( M(N) \) counts the number of multiple recurrence of the sequence \( \{\xi_n\} \) to certain sets \( A_1, \ldots, A_\ell \) which occur at the times \( q_1(n,N), \ldots, q_\ell(n,N) \), as well as SLLN’s for these \( M(N) \)'s. One of the simplest examples is when \( \xi_n \) is \( n \)-the digit of a random \( m \)-base or continued fraction expansion, and each \( A_i \) is singleton (i.e. it represent one possible value of a digit).

1. INTRODUCTION

Since the ergodic theory proof of Szemerédi’s theorem due to Furstenberg (see \[13\]), ergodic theorems for “nonconventional” averages

\[ \frac{1}{N} \sum_{n=1}^{N} T^{q_1(n)} f_1 \cdots T^{q_\ell(n)} f_\ell \]

became a well established field of research (the term ”nonconventional” comes from \[14\]). Here \( T \) is a measure preserving transformation, \( f_1, \ldots, f_\ell \) are bounded measurable functions and the \( q_i \)'s are functions taking positive integer values on the set of positive integers. General polynomial \( q_i \)'s in this setup were first considered in

\[ \text{Date: March 18, 2022.} \]
\[ 2010 \text{ Mathematics Subject Classification.} \quad 60F10; 60J05; 37D99. \]
\[ \text{Key words and phrases.} \quad \text{Central limit theorems; Nonconventional setup, Ergodic theorems; Stein’s method, Markov chains, Dynamical systems.} \]
Taking $f_i$’s to be indicators of measurable sets asymptotic results on numbers of multiple recurrences follow, which was the original motivation for this study.

From a probabilistic point of view, ergodic theorems are laws of large numbers, and so the question about other probabilistic limit theorems is natural. In [15] and [17] central limit theorems for random functions of the form

$$N^{-\frac{1}{2}} \sum_{n=1}^{[Nt]} F(\xi_{q_1(n)}, \xi_{q_2(n)}, \ldots, \xi_{q_{\ell}(n)})$$

(1.1)

were obtained. Here $F = F(x_1, \ldots, x_{\ell})$ is a locally Hölder continuous function, $\{\xi_n\}$ is a sequence of random variables satisfying some mixing and moment conditions and the $q_i$’s are functions satisfying certain growth conditions, which take positive integer values on the set of positive integers. Considering polynomial $q_i$’s, the growth conditions [15] and [17] exclude the case when some of the non-linear functional CLT’s to the case when all of the $q_i$’s have the same degree. In [8] we extended the above functional CLT’s to the case when all of the $q_i$’s are polynomials, with no restrictions on their degrees.

In this paper we will consider more general random functions of the form

$$S_N(t) = N^{-\frac{1}{2}} \sum_{n=1}^{[Nt]} F(\xi_{q_1(n,N)}, \xi_{q_2(n,N)}, \ldots, \xi_{q_{\ell}(n,N)})$$

where each $q_i$ is a bivariate polynomial with integer coefficients. The motivation for considering this case where the $q_i$’s depend also on $N$ comes from applications to multiple recurrence (see the last two paragraphs of this section), and a relatively simple but still very interesting combinatorial number theoretic example can be described as follows. For each point $\omega \in [0,1)$ consider its base $m$ or continued fraction expansions with digits $\xi_k(\omega)$, $k = 1, 2, \ldots$. Next, count the number $S_N(\omega)$ of those $\ell$-tuples $q_1(n,N), \ldots, q_{\ell}(n,N)$, $n \leq N$ for which, say, $\xi_{q_j(n,N)}(\omega) = a_j$, $j = 1, \ldots, \ell$ for some fixed integers $a_1, \ldots, a_{\ell}$. We have

$$S_N(\omega) = \sum_{n=1}^{N} \prod_{j=1}^{\ell} \delta_{a_j, \xi_{q_j(n,N)}(\omega)}$$

where $\delta_{k,s} = 1$ if $k = s$ and $= 0$ otherwise. To make $\xi_k$’s random variable, we supply the unit interval with an appropriate probability measure such as the Lebesgue measure for base $m$ expansions and the Gauss measure for continued fraction expansions (see Section 8). Now we can view $S_N$ as a random variable and set $S_N(t) = S_{\lfloor Nt \rfloor}$.

In [18] several $L^2$ ergodic theorems were proved for the averages $N^{-\frac{1}{2}} S_N(1)$, when $F(x_1, \ldots, x_{\ell})$ has the form $F(x_1, \ldots, x_{\ell}) = \prod_{i=1}^{\ell} f_i(x_i)$. In [16] a strong law of large numbers (SLLN) was proved for $N^{-\frac{1}{2}} S_N(1)$ when the $q_i$’s depend only on $n$, while in Chapter 3 of [10], under certain mixing conditions, we proved an SLLN and a CLT for $S_N(1)$ for linear functions $q_i(n,N) = a_i n + b_i N$. The above results from [16] and [10] hold true for functions $F$ which do not necessarily have the above product form. In this paper we will prove an SLLN for general polynomials $q_i(n,N)$ and a functional CLT for certain classes of polynomials $q_i(n,N)$, and our results generalize both [8] and the above results from Chapter 3 in [10].
A crucial step in proving a functional central limit theorem (regardless of the proofs method) is to show that the asymptotic covariances

$$b(t,s) = \lim_{N \to \infty} E[\mathcal{S}_N(t)\mathcal{S}_N(s)]$$

exist. Using some mixing conditions, we will show that these limits exist for several classes of bivariate polynomials \(q_i(n,N)\). One of the main difficulties arising here is to understand the asymptotic behaviour as \(N \to \infty\) of certain sequences of sets \(A_N \subset [1,N]\) which are related to approximation properties of bivariate polynomial differences of the form \(|q(m,N) - p(n,N)|\). This type of behaviour is investigated independently in Section 5. In section 6.4 we address the question of the positivity of \(D^2 = \lim_{N \to \infty} \text{Var}(\mathcal{S}_N(1))\), which is important since \(\mathcal{S}_N\) converges towards the process which equals identically 0 when \(D^2 = 0\). When any two non-linear polynomials \(q_i\) and \(q_j\) do not satisfy that \(q_i(n,N) - q_j(cn + r,N) = z\), for some rational \(c, r, z, c \neq 1\), and all \(n\) and \(N\), or when the polynomials \(q_i(n,N)\) are ordered so that \(q_1(n,N) \leq q_2(n,N) \leq \ldots \leq q_i(n,N)\) for any sufficiently large \(n\) and \(N\) then we give a complete characterization of the positivity of \(D^2\). In more general circumstances we provide sufficient conditions for this positivity.

The CLT’s from [5] and Chapter 3 of [10] rely on classical martingale approximation techniques, which was shown in [17] to be effective in the nonconventional setup. In the past decades Stein’s method has become one of the main tools to prove central limit theorems. In [9] this method was applied successfully for nonconventional sums of the form (1.1), and in Chapter 1 of [10] we generalized [9], and, in particular, showed that Stein’s method yields a functional CLT in the case when the \(q_i\)’s depend only on \(n\). Our proof of the CLT for the random functions \(\mathcal{S}_N\) defined above will also rely on an appropriate functional version of Stein’s method. When the appropriate limiting covariances exist, Stein’s method in the functional setup can be applied successfully for random functions of the form \(\sum_{n=1}^{N} X_{n,N}\) when the triangular array \(\{X_{n,N} : 1 \leq n \leq N\}\) satisfies certain type of strong local dependence conditions, and our arguments will be based on showing that the summands in \(\mathcal{S}_N\) have such a local dependence structure. In fact, we will use this structure also to control the growth rate of the first four moments of \(\mathcal{S}_N(1)\), which is the key to the proof of the SLLN for \(N^{-1} \mathcal{S}_N(1)\).

Our results hold true when, for instance, \(\xi_n = T^n f\) where \(f = (f_1, \ldots, f_d)\), \(T\) is a topologically mixing subshift of finite type, a hyperbolic diffeomorphism or an expanding transformation taken with a Gibbs invariant measure, as well as in the case when \(\xi_n = f(\Upsilon_n)\), \(f = (f_1, \ldots, f_d)\) where \(\Upsilon_n\) is a Markov chain satisfying the Doeblin condition considered as a stationary process with respect to its invariant measure. In the dynamical systems case each \(f_i\) should be either Hölder continuous or piecewise constant on elements of Markov partitions. We can also extend our results to certain classes of dynamical systems \(T\) which can be modelled by a Young tower (even though the conditions specified in Section 2) do not seem to hold true, see Section 8.3. As a consequence, our results hold true for a variety of non-uniformly hyperbolic or distance expanding dynamical systems \(T\), as well. We refer the readers to Section 8 for a detailed description of the sequences \(\{\xi_n\}\) mentioned above.

As an application we can consider \(F(x_1, \ldots, x_d) = x_1^{(1)} \cdots x_d^{(1)}\), \(x_j = (x_j^{(1)}, \ldots, x_j^{(t)})\), \(\xi_n = (X_1(n), \ldots, X_d(n))\), \(X_j(n) = I_{A_j}(T^n z)\) in the dynamical systems case and \(X_j(n) = I_{A_j}(\Upsilon_n)\) in the Markov chain case where \(I_A\) is the indicator of a set \(A\).
Let $M(N)$ be the number of $i$’s between 0 and $N$ for which $T^{q_i,\ell}(l)(N)x \in A_j$ for $j = 0, 1, \ldots, \ell \ (\text{or} \ \Upsilon_{q_i,\ell}(N) \in A_j \ \text{in the Markov chains case})$, where we set $q_0 = 0$, namely the number of $\ell$-tuples of return times to $A_j$’s (either by $T^{q_i,\ell}(l)$ or by $\Upsilon_{q_i,\ell}(N)$). Then our results yield a functional central limit theorem for the number $M(\lfloor Nt \rfloor)$ and also an SLLN for $M(N)$.

Acknowledgement. I would like to thank professor Yuri Kifer for suggesting me the problem and for several helpful discussions, as well.

2. Limit theorems for nonconventional polynomial arrays

2.1. Preliminaries. Our setup consists of a $\varphi$-dimensional stochastic process $\{\xi_n, n \geq 0\}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a family of sub-$\sigma$-algebras $\mathcal{F}_{k,l}$, $-\infty \leq k \leq l \leq \infty$ such that $\mathcal{F}_{k,l} \subset \mathcal{F}_{k',l'} \subset \mathcal{F}$ if $k' \leq k$ and $l' \geq l$. We will impose restrictions on the mixing coefficients

$$\phi(n) = \sup\{\phi(\mathcal{F}_{-\infty,k}, \mathcal{F}_{k+n,\infty}) : k \in \mathbb{Z}\}$$

(2.1)

where for any two sub-$\sigma$-algebras $\mathcal{G}$ and $\mathcal{H}$ of $\mathcal{F}$

$$\phi(\mathcal{G}, \mathcal{H}) = \sup \left\{ \left| \frac{P(\Gamma \cap \Delta)}{P(\Gamma)} - P(\Delta) \right| : \Gamma \in \mathcal{G}, \Delta \in \mathcal{H}, P(\Gamma) > 0 \right\}.$$ 

(2.2)

The quantity $\phi(\mathcal{G}, \mathcal{H})$ measures the dependence between the $\sigma$-algebras $\mathcal{G}$ and $\mathcal{H}$; it vanishes if and only if $\mathcal{G}$ and $\mathcal{H}$ are independent.

In order to ensure some applications, in particular, to dynamical systems we will not assume that $\xi_n$ is measurable with respect to $\mathcal{F}_{n,n}$ but instead impose restrictions on the approximation rate

$$\beta_q(r) = \sup_{k \geq 0} \| \xi_k - E[\xi_k | \mathcal{F}_{k-r,k+r}] \|_{L^q}.$$ 

(2.3)

Our results will be obtained under the assumption that $\beta_q(r)$ and $\phi(n)$ converge sufficiently fast to 0 (see (2.2) and Assumption 2.1.3 below), for a certain choice of $q$’s. We note that when the sequence $\{\xi_n, n \geq 0\}$ itself satisfies some strong quantitative mixing conditions (e.g. when it forms a geometrically ergodic Markov chain or when $\xi_n$ depends only on the $n$-th coordinate of a topologically mixing subshift of finite type, see Section 5), then our results hold true when $\mathcal{F}_{k,l}$ is the $\sigma$-algebra generated by $\xi_{\max(0,k)}, \ldots, \xi_{\max(0,l)}$. In this case we have $\beta_q(r) = 0$ for any $q$ and $r$. When $\xi_n = f(T^n x)$ for several types of dynamical systems $T$ and functions $f$, the sequence $\{\xi_n, n \geq 0\}$ will only be mixing in an ergodic theoretic sense, but when the function $f$ is Hölder continuous we will have that $\beta_\infty(r) \leq c\delta^r$ for some $c > 0$ and $\delta \in (0,1)$ (see Section 5), where $\beta_\infty(r)$ is the approximation rate corresponding to some families of $\sigma$-algebras $\mathcal{F}_{k,l}$ for which $\phi(n) \leq c\delta^n$. For instance, when $T$ is a topologically mixing subshift of finite type then we can take $\mathcal{F}_{k,l}$ to be the $\sigma$-algebra generated by the projections of the coordinates indexed by $\max(k,0), \ldots, \max(l,0)$, while in the case when $T$ is some hyperbolic map we can take $\mathcal{F}_{k,l} = \bigwedge_{j=k}^l T^{-j} \mathcal{M}$, where $\mathcal{M}$ is a Markov partition with sufficiently small diameter (see Section 5).

Next, we will discuss our stationarity assumptions. We do not require the usual stationarity of the process $\{\xi_n, n \geq 0\}$, and instead we only assume that the distribution of $\xi_n$ does not depend on $n$ and that the joint distribution of $(\xi_n, \xi_m)$
depends only on \( n - m \), which we write for further reference by

\[ \xi_n \overset{d}{\sim} \mu \quad \text{and} \quad (\xi_n, \xi_m) \overset{d}{\sim} \mu_{m-n} \tag{2.4} \]

where \( Y \overset{d}{\sim} \mu \) means that \( Y \) has \( \mu \) for its distribution.

Let \( F = F(x_1, ..., x_\ell) \), \( x_j \in \mathbb{R}^p \) be a function on \((\mathbb{R}^p)^\ell\) which satisfies the following smoothness and growth conditions: for some \( K > 0 \), \( \iota \geq 0 \), \( \kappa \in (0,1] \) and all \( x_i, z_i \in \mathbb{R}^p \), \( i = 1, ..., \ell \), we have

\[ |F(x) - F(z)| \leq K[1 + \sum_{i=1}^{\ell} (|x_i|^\iota + |z_i|^\iota)] \sum_{i=1}^{\ell} |x_j - z_j|^\kappa \tag{2.5} \]

and

\[ |F(x)| \leq K[1 + \sum_{i=1}^{\ell} |x_i|^\iota] \tag{2.6} \]

where \( x = (x_1, ..., x_\ell) \) and \( z = (z_1, ..., z_\ell) \). We note that the role of (2.5) is to insure that we can approximate \( F(\xi_{n_1}, \xi_{n_2}, ..., \xi_{n_\ell}) \) by \( F(\xi_{n_1}, \xi_{n_2}, ..., \xi_{n_\ell}) \) where \( n_1, ..., n_\ell \) are arbitrary nonnegative integers, \( \xi_{n,r} = E[\xi_n | F_{n-r,n+r}] \) and \( r \) is sufficiently large, while the role of (2.6) is to insure that \( F(\xi_{n_1}, \xi_{n_2}, ..., \xi_{n_\ell}) \) will satisfy some moment conditions (see a discussion after Assumption 2.1.1). Such an approximations is possible, of course, only when \( \lim_{r \to \infty} \beta_q(r) = 0 \) (for some \( q \)) and the \( \xi_i \)'s satisfy certain moment conditions (such conditions are imposed in Assumptions 2.1.1 and 2.1.3 below). We also remark that in the case when \( \xi_n \) is measurable with respect to \( F_{n,n} \) then our results will follow with any Borel function \( F \) satisfying (2.6) without imposing (2.5), since then our proof do not require any approximations of the above form.

Our moment conditions on the \( \xi_n \)'s are summarized in the following

2.1.1. Assumption. There exist \( w > 4 \), \( q \geq 1 \) and \( v > 0 \) such that

\[ \frac{1}{w} > \frac{\iota}{v} + \frac{\kappa}{q} \quad \| \xi_n \|_{L^w} < \infty \quad \text{and} \quad \| \xi_n \|_{L^v} < \infty \tag{2.7} \]

where \( \iota \) and \( \kappa \) come from (2.5) and (2.6).

We note that Assumption 2.1.1 insures that \( \| F(\xi_{n_1}, \xi_{n_2}, ..., \xi_{n_\ell}) \|_{L^w} \leq C \) for some \( C > 0 \) and all nonnegative integers \( n_1, ..., n_\ell \). When \( \phi(n) \) and \( \beta_{q_0}(n) \) (for some \( q_0 \)) converge exponentially fast to 0 as \( n \to \infty \) then in order to get the strong law of large numbers (Theorem 2.1.2) and the functional central limit theorem (Theorem 2.3.3) we can take any \( w > 4 \), \( q = q_0 \) and any \( v > 0 \) so that \( \frac{1}{w} > \frac{\iota}{v} + \frac{\kappa}{q_0} \). We also remark that we have presented the moment conditions with some dependence on the regularity on \( F \) (i.e. on \( \kappa \) and \( \ell \)), but, of course, we could have first assumed that the properties described in Assumption 2.1.1 hold true with some \( w, q, v, \kappa \in (0,1] \) and \( \iota \geq 0 \), and then consider only functions satisfying (2.5) and (2.6) with these \( \kappa \) and \( \ell \).

Next, let \( q_1(n,N), ..., q_\ell(n,N) \) be polynomials with nonnegative integer coefficients which do not depend only on \( N \). We assume here, for the sake of convenience, that \( \deg q_i \leq \deg q_{i+1} \) for any \( i = 1, 2, ..., \ell - 1 \), where the degree of a bivariate polynomial \( p(x,y) \) is the degree of the univariate polynomial \( p(x,x) \), and that the differences \( q_i - q_j \) are not constants (the case of constant difference can be
treated as in Section 3 of [8]. For each \( N \) set

\[
S_N = \sum_{n=1}^{N} F(\xi_{q_1(n,N)}, \xi_{q_2(n,N)}, \ldots, \xi_{q_r(n,N)}).
\]

The first result we will prove in this paper is the following strong law of large numbers:

2.1.2. Theorem. Suppose that Assumption 2.1.1 holds true with numbers \( w \) and \( q \) so that

\[
\sum_{l=1}^{\infty} l\left((\phi(l))^{-\frac{1}{2}} + \beta_q(l)\right) < \infty
\]

where \( \kappa \) comes from 2.1.2. Then, \( P \)-almost surely we have

\[
\lim_{N \to \infty} \frac{1}{N} S_N = \bar{F} := \int F(x_1, \ldots, x_l) d\mu(x_1) \ldots d\mu(x_l).
\]

The reason that the mixing and approximation coefficients \( \phi(l) \) and \( \beta_q(l) \) appear additively in (2.8) is that in the proof of Theorem 2.1.2 we approximate each summand \( F(\xi_{q_1(n,N)}, \xi_{q_2(n,N)}, \ldots, \xi_{q_r(n,N)}) \) by \( F(\xi_{q_1(n,N), r(n)}, \xi_{q_2(n,N), r(n)}, \ldots, \xi_{q_r(n,N), r(n)}) \), where \( r(n) \) is some number which depends on \( n \) and \( \xi_{m,r} = E[\xi_{m,r} | F_{m-r,m+r}] \) for any \( m \) and \( r \geq 0 \) (or course, we will require that \( \lim_{n \to \infty} r(n) = \infty \) at a certain rate).

The main result in this section is a functional central limit theorem for the sequence of random functions \( S_N : [0, 1] \to \mathbb{R}, N \in \mathbb{N} \) given by

\[
S_N(t) = N^{-\frac{1}{2}} \sum_{n=1}^{[Nt]} F(\xi_{q_1(n,N)}, \xi_{q_2(n,N)}, \ldots, \xi_{q_r(n,N)}) = N^{-\frac{1}{2}} S_{[Nt]}.
\]

To simplify formulas we assume the (asymptotic) centering condition

\[
\bar{F} = \int F(x_1, \ldots, x_l) d\mu(x_1) \ldots d\mu(x_l) = 0
\]

which is not really a restriction since we can always replace \( F \) by \( F - \bar{F} \). We note that \( \bar{F} \) is not the expectation of \( F(\xi_{q_1(n,N)}, \xi_{q_2(n,N)}, \ldots, \xi_{q_r(n,N)}) \), but the arguments in our proofs show that, in the circumstances of Theorem 2.1.2 the expectation of \( \frac{1}{N} S_N \) converges to \( \bar{F} \) as \( N \to \infty \) (this is what expected to happen, in view of Theorem 2.1.2).

Our main mixing and approximation conditions for the central limit theorem is the following

2.1.3. Assumption. There exist \( d \geq 1 \) and \( \theta > 2 \) such that for any \( n \in \mathbb{N} \),

\[
\phi(n) + (\beta_q(n))^\frac{1}{2} \leq dn^{-\theta}.
\]
2.2. Classes of polynomials. We describe here several classes of polynomials for which we can derive the weak invariance principle for the random functions $\mathcal{S}_N$.

First, we assume here that the linear polynomials among the $q_i$’s have the form

$$q_i(n, N) = a_i n + b_i N$$

(2.11)

for some integers $a_i$ and $b_i$, namely that $q_i(0, 0) = 0$. Our additional requirements from the linear polynomials are described in the following

2.2.1. Assumption. For any linear $q_i$ and $q_j$ the difference $a_i - a_j$ is divisible by the greatest common divisor of $b_i$ and $b_j$ where the $a_i$’s and $b_i$’s are the same as in (2.11).

Next, in order to describe our conditions regarding the non-linear polynomials among the $q_i$’s, we need the following definitions. Let $q(n, N)$ and $p(n, N)$ be two bivariate polynomials with nonnegative integer coefficients. We will say that $q$ and $p$ have exploding differences if for any $\delta \in (0, 1)$ there exist constants $C_\delta > 0$ and $N_\delta$ and sets $\Gamma_{N, \delta} \subset [1, N]$, whose cardinality does not exceed $\delta N$, so that for any $N > N_\delta$ and $n \in [\delta N, N] \setminus \Gamma_{N, \delta},$

$$\min_{m \in [\delta N, N]} |q(m, N) - p(n, N)| \geq C_\delta N.$$  

(2.12)

It is clear that any two polynomials $q$ and $p$ with different degrees have exploding differences and that two linear polynomials do not have exploding differences. In Section 3 we will give several classes of examples of polynomials $q$ and $p$ with the same non-linear degree which have exploding differences. Next, for any $1 \leq i \leq \ell$ such that $\deg q_i = k > 1$ write

$$q_i(n, N) = \sum_{u=0}^{k} N^u Q_{i,u}(y)$$

(2.13)

where $y = n/N$ and each $Q_{i,u}$ is a polynomial with non-negative integer coefficients whose degree does not exceed $u$. For any distinct $1 \leq i, j \leq \ell$ such that $\deg q_i = \deg q_j = k > 1$, we will say that $q_i$ and $q_j$ are linearly related if $Q_{i,k}$ and $Q_{j,k}$ are not constants and there exist constants $c_{i,j}, r_{i,j} \in \mathbb{R}$, $c_{i,j} > 0$ so that $Q_{j,k}(c_{i,j} y) = Q_{i,k}(y)$ and

$$Q_{i,k-1}(y) - Q_{j,k-1}(c_{i,j} y) = r_{i,j} Q_{j,k}(c_{i,j} y)$$

for any $y \in [0, 1]$. Then, any two polynomials $q_i$ and $q_j$ which do not depend on $N$ and have the same non-linear degree $k$ are linearly related. Indeed, in this case we have $Q_{i,u}(y) = a_{i,u} y^u$ and $Q_{j,u}(y) = a_{j,u} y^u$ for some integers $a_{i,u}$ and $a_{j,u}$ so that $a_{i,k}, a_{j,k} > 0$, and so we can take $c_{i,j} = (a_{i,k}/a_{j,k})^{1/k}$ and, with $c_{i,j,k} = (c_{i,j})^{k-1}$, $r_{i,j} = (a_{i,k-1} a_{j,k} - a_{i,k} a_{j,k-1})/c_{i,j,k} c_{i,j,k}$. This means that all the results obtained in this paper generalize the results from [8], in which a nonconventional polynomial CLT was obtained in the case when all the $q_i$’s are polynomial functions of the variable $n$. Observe also that the linear relation condition involves only the $Q_{i,k}$’s and $Q_{i,k-1}$’s and note that $q_i$ and $q_j$ (with the same non-linear degree $k$) are linearly related if

$$q_i(n, N) = \alpha_n n^s N^{k-s} + \beta_n n^{s-1} N^{k-s} + G_i(n, N)$$

and

$$q_j(n, N) = \alpha_n n^s N^{k-s} + \beta_n n^{s-1} N^{k-s} + G_j(n, N)$$

for some $0 < s \leq k$, polynomials $G_i$ and $G_j$ whose degree does not exceed $k - 2$ and positive integers $\alpha_i, \alpha_j, \beta_i$ and $\beta_j$. We refer the readers to Corollary 3.3.1 for a
characterization when two linearly related polynomials have exploding differences (see also Remark 2.3.2 below).

We will obtain our results under the following

2.2.2. Assumption. Any two non-linear polynomials $q_i$ and $q_j$ are either linearly related, or the differences of $q_i$ and $q_j$ explode.

2.3. Limiting covariances and the CLT. Our first CLT related result is the following

2.3.1. Theorem. Suppose that Assumptions 2.1.1 and 2.1.3 are satisfied with numbers $w$ and $\theta$ so that $\theta > \frac{4w}{w^2 - 2}$. Assume, in addition, that Assumptions 2.2.1 and 2.2.2 are satisfied. Then the limits

$$b(t, s) = \lim_{N \to \infty} E[S_N(t)S_N(s)]$$

exist, where $0 \leq t, s \leq 1$. In particular, the limit

$$D^2 = \lim_{N \to \infty} ES_N^2 = \lim_{N \to \infty} \text{Var}(S_N)$$

exists, where $S_N = S_N(1)$.

Note that in Section 6 we will also provide several formulas for the limits $b(t, s)$, as well as some conditions for the positivity of $D^2$ (when $D^2 = 0$ then $S_N$ converges to the process which equals $0$ identically). We refer the readers to Section 6.3 to a discussion about existence of $b(t, s)$ (or just $D^2$) for (more general) polynomials $q_i$’s satisfying certain number theory related conditions.

2.3.2. Remark. The property of being linearly related is, in fact, an equivalence relation. Indeed, if both pairs $(q_i, q_j)$ and $(q_j, q_l)$ are linearly related then we can always take $c_{i,l} = c_{i,j} \cdot c_{j,l}$ and $r_{i,l} = r_{j,l} + r_{i,j} \cdot c_{j,l}$.

We will say that the polynomials $q_i$ and $q_j$ are $Q$-equivalent if there exist rational $c$ and $r$ so that the difference $q_i(n, N) - q_j(cn + r, N)$ does not depend on $n$ and $N$. Then any two $Q$-equivalent polynomials are linearly related. In Corollary 3.3.1 we will show that any two linearly related polynomials which are not $Q$-equivalent have exploding difference. Therefore Assumption 2.2.2 means that any two non-linear polynomials among the $q_i$’s are either equivalent or have exploding differences, and under Assumption 2.2.2 having exploding differences is a symmetric relation.

When the asymptotic covariances $b(t, s)$ exist then, using a functional version of Stein’s method due to A.D. Barbour, we derive the following

2.3.3. Theorem. Suppose that Assumptions 2.1.1 and 2.1.3 are satisfied with numbers $w$ and $\theta$ so that $\theta > \frac{4w}{w^2 - 2}$. Assume, in addition, that the limiting covariances $b(t, s)$ exist. Then, the random functions $S_N : [0, 1] \to \mathbb{R}$ converge in distribution as $N \to \infty$ towards a centered Gaussian process $\eta(t)$ whose covariances are given by $E[\eta(t)\eta(s)] = b(t, s)$.

The arguments in the proof of Theorem 2.3.3 together with the arguments in Chapter 1 of [10] show that Stein’s method also yields almost optimal convergence rate in the CLT for the sequence of random variables $S_N = S_N(1)$, when $D^2 > 0$. These results are not included here in order not to overload this paper.
2.4. Outline of the proof of Theorems 2.1.2, 2.3.1 and 2.3.3. Consider first the case when the random variables \( \xi_n \) are independent. For any \( n \) and \( N \) set

\[
F_{n,N} = F(\xi_{q_1(n,N)}, \xi_{q_2(n,N)}, \ldots, \xi_{q_d(n,N)}).
\]

Fix some \( N \). Since the \( q_i \)'s are polynomials, each \( F_{n,N} \) can depend on at most \( d_0 \ell^2 \) of the random variables \( F_{m,N} \), \( m \geq 1 \), where \( d_0 \) is the maximal degree of the polynomials \( q_i \). We see then that the triangular array \( \{F_{n,N}, n = 1, 2, \ldots, N\} \) satisfies a “local dependence” type condition (here we view the indexes \( m \) of the random variables \( F_{m,N} \) which depend on \( F_{n,N} \) as a certain “neighborhood of dependence”). When the sequence \( \{\xi_n\} \) is only weakly dependent then we get a certain version of the above local dependence, but now, roughly speaking, the dependence is replaced with a certain type of “local strong dependence”, which essentially means that each \( F_{n,N} \) can “strongly depend” only a number of \( F_{m,N} \)'s whose magnitude is smaller than \( N \). Giving a precise meaning to this “strong local dependence” is one of the main ideas behind the proof of Theorem 2.3.3 (see the beginning of Section 7 for the precise definition of the underlying graph in the weakly depend case). This was done in Chapter 1 of [10] in the case when \( q_i \) depends only on \( n \) (in fact, the full details were given there only when \( q_i(n) = in \) for each \( i \)), but here the dependence on \( N \) causes additional difficulties. The resulting local dependence structure is a classical situation where we can use the functional version of Stein’s method for Gaussian approximation due to A.D. Barbour (see Theorem 7.0.1 and [1]).

Using the above strong local dependence structure together with some counting arguments, we get that the \( L^4 \) norms the sums \( N^{-\frac{1}{2}}(S_N - NF) \) are bounded in \( N \). This is the main idea of the proof of Theorem 2.1.2. Indeed, assume for the sake of convenience that \( F = 0 \). Then by the Markov inequality we have

\[
P\left( N^{-1}|S_N| \geq N^{-1/8} \right) \leq \frac{\|S_N\|_{L^4}}{N^{7/2}} \leq CN^{-3/2}
\]

which together with the Borel-Cantelli lemma yields that with probability one we have \( \lim_{N \to \infty} \frac{1}{N} S_N = 0 = \bar{F} \).

The proof of Theorem 2.3.1 is somehow less transparent even in the case of independent \( \xi_n \)'s (and it does not rely only on strong local dependence), but let us describe some of its key ingredients. In the case when the \( q_i \)'s depend only on \( n \), we can order them and just assume that \( q_1(n) < \ldots < q_d(n) \) for any sufficiently large \( n \). When the polynomials \( q_i \) depend on \( n \) and \( N \) then, in general, it is impossible to order them. The first step in the proof of Theorem 2.3.1 is to show that, after omitting a “small” number of \( n \)'s (in comparison to \( N \)) between 1 to \( N \), we can order the polynomials \( q_i(n, N) \) when \( n \in \Gamma_i, i = 1, 2, \ldots, d \), where \( \{\Gamma_i\} \) is a partition of the remaining \( n \)'s. After this is established, roughly speaking, our proof scheme requires to study certain combinatorial number theory related problems which are related to asymptotic densities of sets of the form

\[
\{aN \leq n \leq bN : q(n, N) \in \cup_{m \in \mathbb{N}} \{p(m, N)\} \} \subset [1, N]
\]

where \( a, b \) are positive numbers between 0 to 1 and \( q \) and \( p \) are bivariate polynomials with integer coefficients. Such sets were investigated in [8] when the polynomials \( q(x, y) \) and \( p(x, y) \) depend only on \( x \), but when they depend also on \( y \) many additional number theoretic difficulties arise.
3. Differences of bivariate non-linear polynomials

Let \( q(n, N) = q_N(N) \) and \( P(n, N) = P_N(n) \) be two polynomials in the variables \( n, N \) with nonnegative integer coefficients so that \( \deg q = \deg p = k > 1 \) for some \( k > 1 \). We will also assume here that the polynomials \( q \) and \( p \) do not depend only \( N \). In particular the functions \( q_N^{-1} : [q_N(0), \infty) \to [0, \infty) \) and \( p_N^{-1} : [p_N(0), \infty) \to [0, \infty) \) are well defined. The goal in this section is to investigate the asymptotic behaviour of the differences \( |q_N(m) - p_N(n)| \). In Sections 3.3 and 3.4 we will prove some general results, which will be applied in Sections 5.3 and 5.4 in more specific situations.

3.1. First estimate. Our first results is the following

3.1.1. Proposition. Suppose that \( q \) and \( p \) have the form

\[
p(n, N) = H(N)P(n, N) + r(n, N) \quad \text{and} \quad q(m, N) = H(N)Q(m, N) + s(m, N)
\]

for some non-constant polynomial \( H \), polynomials \( Q \) and \( P \) with non-negative integer coefficients and polynomials \( r \) and \( s \) so that \( \max(\deg r, \deg s) < \deg H \). Then for any \( 1 \leq n, m \leq N \) so that \( p_N(n) > q_N(0) \) and \( q_N(m) > p_N(0) \), either \( Q_N(m) = P_N(n) \), where \( Q_N(x) = Q(x, N) \) and \( P_N(x) = P(x, N) \), or

\[
|q_N(m) - p_N(n)| \geq \Delta_N(n, m)|H(N)| - |s(m, N) - r(n, N)|,
\]

where \( \Delta_N(n, m) \) is the minimum of \( Q_N^{-1}(Q_N^{-1}(P_N(n)))/Q_N^{-1}(m) \) and \( P_N^{-1}(P_N^{-1}(Q_N(n)))/P_N^{-1}(m) \) and \( P_N^{-1} \) and \( Q_N^{-1} \) are the inverse functions of the univariate functions \( P_N(\cdot) \) and \( Q_N(\cdot) \), respectively. In particular, for any \( \delta > 0 \) there exists a constant \( R_\delta > 0 \) so that for any sufficiently large \( N \) and \( \delta N \leq n, m \leq N \) so that \( Q_N(m) \neq P_N(n) \),

\[
|q_N(m) - p_N(n)| \geq R_\delta|H(N)|.
\]

As a consequence, \( q \) and \( p \) have exploding differences when \( Q \) and \( P \) have exploding differences or when the degrees of \( r \) and \( s \) are different.

The polynomials \( Q \) and \( P \) have exploding differences when they are linearly related, but not \( Q \)-equivalent (see Remark 2.3.2 and Corollary 3.3.1). They also have exploding differences in the circumstances of Corollary 3.4.3. Set

\[
d = \limsup_{N \to \infty} \frac{1}{N}\{|1 \leq n \leq N : P(n, N) \in \{Q(m, N) : m \in [1, N]\}\}.
\]

where \(|\Gamma|\) stands for the cardinality of a finite set \( \Gamma \). Then, it follows from Proposition 3.1.1 that the polynomials \( q \) and \( p \) have exploding differences also when \( d = 0 \). The upper limit \( d \) equals 0 when \( Q \) and \( P \) have exploding differences, but also when, for instance, \( P(n, N) \) and \( Q(m, N) \) take values at disjoint sets (e.g. when \( P(n, N) \) is odd and \( Q(m, N) \) is even etc.).

Proof of Proposition 3.1.1. It is clearly enough to prove 3.2 in the case when \( r \equiv 0 \) and \( s \equiv 0 \). Since \( p_N(n) > q_N(0) \) and \( q_N(n) > p_N(0) \), the numbers \( t_{n,N} = q_N^{-1}p_N(n) = Q_N^{-1}P_N(n) \) and \( s_{m,N} = p_N^{-1}q_N(m) = P_N^{-1}Q_N(m) \) are well defined. Suppose first that \( q_N(m) > p_N(n) \). Then \( Q_N(m) > P_N(n) \) and we can write \( m = t_{n,N} + x \), where here \( x \geq 0 \) is considered as a parameter. Define the function \( D_{n,N}(y) \) by

\[
D_{n,N}(y) = Q_N(t_{n,N} + y) - P_N(n) = Q_N(t_{n,N} + y) - Q_N(t_{n,N}).
\]
Then \(D_{n,N}(0) = 0\). Applying the mean value theorem with the function \(D_{n,N}\), taking into account that the derivative of \(Q_N\) is increasing and that \(x \geq 0\), we obtain that
\[
|Q_N(m) - P_N(n)| = |D_{n,N}(x) - D_{n,N}(0)| \geq Q'_N(t_{n,N}) \cdot |x - 0| = Q'_N(t_{n,N}) \cdot |m - t_{n,N}|.
\]
Next, we define the function \(g = g_{n,N}(\cdot)\) by \(g(t) = Q_N(t) - P_N(n)\). Then \(g(t_{n,N}) = 0\).

By the mean value theorem, there exists \(\xi\) between \(m\) and \(t_{n,N}\) so that
\[
|m - t_{n,N}| \cdot g'(\xi) = |g(m) - g(t_{n,N})| = |g(m)| \geq 1
\]
where we used that that \(m \neq t_{n,N}\). Since \(0 \leq g'(\xi) = Q'_N(\xi) \leq Q'_N(m)\) we obtain that
\[
|m - t_{n,N}| \geq \frac{1}{Q'_N(m)},
\]
which together with the previous estimates implies that
\[
|Q_N(m) - P_N(n)| \geq \frac{Q'_N(Q_N^{-1}P_N(n))}{Q_N(m)}.
\]
In the case when \(Q_N(m) < P_N(n)\) we obtain (3.2) by reversing the roles of \(Q\) and \(P\) and the above arguments.

We refer the readers to Corollary 3.4.4 in which we give a class of examples of polynomials \(Q\) and \(P\) with exploding differences.

In the case when \(r\) and \(s\) are polynomials of the same degree, we can check whether Proposition 3.1.1 can be applied with \(s\) and \(r\) in place of \(q\) and \(p\). Still, \(r\) and \(s\) (or even \(p\) and \(q\)) may, for instance, contain a monomial which does not depend on \(N\). In the next section we will estimate \(|Q_N(m) - P_N(n)|\) under somehow different type of conditions, which will have applications beyond the case considered in Proposition 3.1.1

### 3.2. Estimates using decompositions into homogeneous polynomials

Set \(y_1 = \frac{y}{N}\), \(y_2 = \frac{y}{N}\) and write
\[
p_N(n) = \sum_{j=0}^{k} N^j P_j(y_1) \quad \text{and} \quad q_N(m) = \sum_{j=0}^{k} N^j Q_j(y_2)
\]
where \(P_j\) and \(Q_j\) are polynomials whose degree does not exceed \(j\). We will also assume here that \(Q_k\) and \(P_k\) are not constant polynomials and that \(Q_k(0) \leq P_k(0)\).

In the above circumstances, the function \(\gamma_k(y) = Q_k^{-1}P_k(y)\) is well defined on \([0, \infty)\).

Next, for any \(y \in (0, 1]\), let the polynomial \(H_{N,y}\) be given by
\[
H_{N,y}(\xi) = \sum_{u=2}^{k-1} N^{-(u-1)} C_u(y) + (1 + \sum_{u=2}^{k} N^{-(u-1)} A_{1,u}(y)) \xi
\]
\[
+ \sum_{s=2}^{k} \left( \sum_{u=2}^{k} N^{-(u-1)} A_{s,u}(y) \right) \xi^s.
\]

Here \(A_{s,u}(y) = \sum_{j=s}^{u} \binom{j}{s} \frac{Q^{(j)}_{k-u}(\gamma_k(y))}{Q_k(\gamma_k(y))} \frac{Q^{(j)}_{k}(\gamma_k(y))}{Q_k(\gamma_k(y))} (R_k(y))^{j-s}\) and
\[
C_u(y) = \frac{Q_{k-u}(\gamma_k(y)) - P_{k-u}(y)}{Q_k(\gamma_k(y))}
\]
Therefore, there exist a constant $A_1$ which depends only on the polynomials $q$ and $p$ so that for any $y \in (0, 1]$,

$$
\left| H'_{N,y}(\xi) - 1 \right| \leq \frac{A_1}{NQ_k'/(\gamma_k(y))}.
$$

(3.5)

Therefore, there exist a constant $N_0$ so that if $Q_k'/(\gamma_k(y)) > \frac{A_1}{N}$ and $N > N_0$ is sufficiently large then the function $H_{N,y}$ is strictly increasing on $[-1, 1]$ and there exists a unique root $x_N(y)$ of $H_{N,y}$ in $[-1, 1]$, which, by the mean value theorem, satisfies that $H_{N,y}(0) = A_N(y)x_N(y)$, for some function $A_N(y)$ so that

$$
\left| A_N(y) - 1 \right| \leq \frac{A_1}{NQ_k'/\gamma_k(y)}.
$$

(3.6)

Observe that $H_{N,y}(0)$ is at most of order $N^{-1}$. When all of the functions $C_2, \ldots, C_{k-1}$ are identically 0 then $H_{N,y}(0) = x_N(y) = 0$. In general, we have the following

3.2.1. **Lemma.** Suppose that not all the $C_u$'s are identically zero. Let $s_0 \leq k - 1$ denote the first index $u$ so that the function $C_u(\cdot)$ does not equal 0 identically. Then for any $\delta \in (0, 1)$ there exist positive constants $B_1(\delta)$ and $B_2(\delta)$ and a set $\Gamma_{N,\delta} \subset [1, N]$, whose cardinality does not exceed $\delta N$, so that for any sufficiently large $N$ and $u \in [\delta N, N] \setminus \Gamma_{N,\delta}$,

$$
B_1(\delta)N^{-(s_0 - 1)} \leq |H_{N,N/|N(0)|} \leq B_2(\delta)N^{-(s_0 - 1)}.
$$

Proof. The function $C_{s_0}$ can have only a finite number of roots $y_1, \ldots, y_t$ in the interval $[0, 1]$ (since $C_{s_0}$ can be extended to an analytic function in a complex neighborhood of $[0, 1]$). Let us denote these roots by $y_1, \ldots, y_t$. Let $\delta > 0$. Then there exists a constant $C_\delta > 0$ so that for any $y \in [0, 1]$ which satisfy that

$$
\min_{1 \leq i \leq t} |y - y_i| \geq \frac{\delta}{2t} := \delta_t
$$

we have $|C_{s_0}(y)| \geq C_\delta$. Let $B$ be an upper bound of the absolute value of the functions $C_u : [0, 1] \to \mathbb{R}$. Set

$$
\Gamma_{N,\delta} = \bigcup_{i=1}^{t} [N(y_i - \delta_t), N(y_i + \delta_t)].
$$

If $y := n/N \in [\delta, 1] \setminus \Gamma_{N,\delta}$, then, with $s = s_0 - 1$

$$
(C_\delta - BkN^{-1})N^{-s} \leq |H_{N,y}(0)| = \left| \sum_{u=0}^{k} N^{-(u-1)}C_u(y) \right| \leq kBN^{-s}
$$

and the proof of the lemma is complete. 

Our next result is the following

3.2.2. **Lemma.** For any natural $n, m$ and $N$, with $y = \frac{n}{N}$, we have

$$
q_N(m) - p_N(n) = q_0(R_k(y)) - P_1 + Q_k'(\gamma_k(y))N^{k-1}H_{N,y}(m - N\gamma_k(y) - R_k(y)).
$$
As a consequence, if \( Q_k(y) = \frac{2A_1}{N} \) and \( N \) is sufficiently large, where \( A_1 \) comes from (3.5), then, with \( \alpha_N(y) = N\gamma_k(y) + R_k(y) + x_N(y) \),

\[
\frac{1}{2}Q'_k(\gamma_k(y))N^{k-1}|m - \alpha_N(y)| \\
\leq |q_N(m) - p_N(n) - (Q_0(R_k(y)) - P_0)| \leq 2Q'_k(\gamma_k(y))N^{k-1}|m - \alpha_N(y)|.
\]

In particular, for any constants \( s < k \) and \( 0 < B_1(y) < B_2(y) < \infty \) so that

\[
B_1(y)N^{-s} \leq |H_N,y(0)| \leq B_2(y)N^{-s}
\]

there exists a constant \( K(y) \) so that

\[|q_N(m) - p_N(n)| \geq K(y)N^{k-s} - |q_0(R_k(y)) - P_0|
\]

if

\[
|m - N\gamma_k(\frac{y}{N}) - R_k(\frac{y}{N})| \geq KN^{-(s-1)}
\]

for some \( K > 0 \). The constant \( K(y) \) depends only on \( B_1(y), B_2(y), K \) and \( s \).

Let \( \delta \in (0,1) \). If \( y = n/N \in \left[\delta,1\right] \) then \( Q'_k(y) \geq C_\delta \) for some \( C_\delta > 0 \) which depends only on \( \delta \), and so the magnitude of \( Q'_k(\gamma_k(y))N^{k-1} \) is \( N^{k-1} \) and the inequality \( Q'_k(y) > \frac{2A_1}{N} \) holds true, assuming that \( N \) is sufficiently large. Observe also that for such \( n \)’s we have

\[
\max_{y_0 \in [\delta N,1]}|q_0(R_k(y)) - P_0| = D_\delta < \infty
\]

and that \( D_\delta \) depends only on \( \delta \) (and on the polynomials \( q \) and \( p \)). Therefore, when \( n \in [\delta N, N] \setminus \Gamma_N, \delta \), where \( \Gamma_N, \delta \) comes from Lemma 3.2.1 with \( s_0 = s + 1 \), then

\[|q_N(m) - p_N(n)| \geq K_\delta N^{k-s}
\]

for any \( n \) and \( m \) satisfying (3.8), for some constant \( K_\delta > 0 \) (and so, the problem of verifying that \( q \) and \( p \) have exploding differences is reduced to the study of (3.8) see Corollary 3.4.2 for an application).

**Proof of Lemma 3.2.2.** Write \( y = \frac{n}{N} \) and \( m = N\gamma_k(y) + x \), where \( x \) is considered here as a parameter. Then, by considering the Taylor expansion of the polynomials \( Q_j \) around the point \( \gamma_k(y) \) we have

\[
q_N(m) - p_N(n) = \sum_{j=0}^{k} N^j \left( \sum_{s=0}^{j} \frac{Q_j^{(s)}(\gamma_k(y))}{s!} \left( \frac{x}{N} \right)^s \right) - P_j(y)
\]

\[
= \sum_{u=0}^{k} N^{k-u} \left( \sum_{s=0}^{u} \frac{Q_k^{(s)}(\gamma_k(y))x^s}{s!} \right) - P_{k-u}(y)
\]

\[
= N^{k-1}Q_k'(\gamma_k(y))(x - R_k(y)) + \sum_{u=2}^{k} N^{k-u} \left( \sum_{s=0}^{u} \frac{Q_k^{(s)}(\gamma_k(y))x^s}{s!} \right) - P_{k-u}(y)
\]

where we used that \( Q_k(\gamma_k(y)) = P_k(y) \). By considering the above expression as a (polynomial) function of \( x \) (where \( y \) is considered as a parameter), and then considering its Taylor polynomials around the point \( R_k(y) \) we arrive at

\[
q_N(m) - p_N(n) = Q'_k(\gamma_k(y))N^{k-1}H_{N,y}(x - R_k(y)) + \sum_{s=0}^{k} q_{s,n}(R_k(y))^s - P_0
\]

\[
= Q'_k(\gamma_k(y))N^{k-1}H_{N,y}(x - R_k(y)) + q_0(R_k(y)) - P_0
\]
where \( H_N \) was defined in the statement of the lemma \( q_{s,s} \) is the coefficient of monomial \( y^s \) in the polynomial \( Q_s(y) \).

In the following sections we will apply Lemma \( 3.2.2 \) in several situations, where \( \gamma_k \) and \( R_k \) are assumed to have certain structure. In Section \( 6.2 \) we will use the results from Sections \( 3.3 \) and \( 3.4 \) in order to prove Theorem \( 2.3.1 \). For some abstract application of Lemma \( 3.2.2 \) we refer the readers to Section \( 6.3 \).

### 3.3. Application I: linearly related polynomials

We begin with the following corollary.

**Corollary.** Suppose that \( \gamma_k(y) = cy, R_k(y) = r \) for some constants \( c \) and \( r \), namely that \( q \) and \( p \) are linearly related. Then either \( q \) and \( p \) have exploding differences, or \( q \) and \( p \) are \( \mathbb{Q} \)-equivalent (in the terminology of Remark \( 2.3.2 \)) and then for any \( \delta \in (0,1) \) there exist constants \( W_\delta \) and \( N_\delta \) so that for any \( N > N_\delta \), \( \delta N \leq n \leq N \) and \( m \in \mathbb{N} \) either \( m = cn + r \) (which happens on a finite union of arithmetic progressions) and \( q_N(m) - p_N(n) = q_0(r) - P_0 := d \) or \( |q_N(m) - p_N(n)| \geq W_\delta N^{k-1} \).

**Proof.** For each \( n \) and \( N \) set \( d_N(n) = \inf \{|q_N(m) - p_N(n)| : m \in \mathbb{N} \} \). Let \( \delta > 0 \). Suppose first that \( c \) is irrational. Then, using Weyl’s equidistribution theorem, we see that

\[
\limsup \frac{1}{N} |[1,N] \cap B(c,r,\delta)| \leq 2\delta
\]

where \( B(c,r,\delta) \) is the set of all natural numbers \( n \) so that \( |m - cn - r| < \delta \) for some integer \( m \), and \( |\Gamma| \) stands for the cardinality of a finite set \( \Gamma \). Set

\[
C_\delta = Q_k'(\gamma_k(\delta)) = \min_{y \in [0,1]} Q_k'(\gamma_k(y)) > 0
\]

and

\[
D = |q_0(r) - P_0| = |d|.
\]

Note that there exists a constant \( B_\delta > 0 \) so that for any sufficiently large \( N \) and \( n \in [\delta N, N] \) we have

\[
|d_N(n)| \leq B_\delta N^{-1}
\]

where \( X_N(n/N) \) was defined in Section \( 3.2 \). Relying now on Lemma \( 3.2.2 \) we obtain that for any \( n \in [\delta N, N] \cap B(c,r,\delta) \) and \( m \in \mathbb{N} \) we have

\[
|q_N(m) - p_N(n)| \geq \frac{1}{2} C_\delta N^{k-1}(\delta - B_\delta N^{-1}) - D.
\]

Therefore, for any sufficiently large \( N \) we have

\[
|[n \in [1,N] : d_N(n) \geq A_\delta N^{k-1}]| \geq (1 - 3\delta)N
\]

where \( A_\delta \) is a constant which depends only on \( \delta \), and thus the difference of \( q \) and \( p \) explode. Next, suppose that \( c = u/v \) is rational and that \( r \) is irrational. Then for any \( n \) and \( m \) we have

\[
|m - cn - r| = |v|^{-1} \cdot |rv - (mv - un)| \geq |v|^{-1} \inf_{l \in \mathbb{N}} |rv - l| := \delta_0 > 0,
\]

which as in the previous case is enough in order to derive that the differences of \( q \) and \( p \) explode. Note that when \( c \) and \( r \) are not both rational then the polynomials \( q \) and \( p \) are not \( \mathbb{Q} \)-equivalent.

Next, suppose that \( c = u/v \) and \( t = w/t \) are rational and that \( C_\delta \) does not equal identically 0 for some \( 2 \leq u \leq k - 1 \). Then \( q \) and \( p \) are not \( \mathbb{Q} \)-equivalent. Since \( c \)
and \( r \) are rational, either \( m = cn + r \) or \(|m - cn - r| = |tv|^{-1} \left| mtv - utn - vu \right| \geq |tv|^{-1} := \delta_1 > 0 \). In the case when \(|m - cn - r| \geq \delta_1 \) and \( n \in [N\delta, N] \), as in the first part of this proof, we have

\[
|q_N(m) - p_N(n)| \geq \frac{1}{2} C_\delta N^{k-1} (\delta_1 - B_\delta^{-1}) - D
\]

Assume now that \( m = cn + r \) (i.e. \( m = N\gamma_k(n/N) + R_k(n/N) \)) and \( n \in [N\delta, N] \).

Let \( s + 1 < k \) and \( \Gamma_{N,\delta} \) be as in Lemma 3.2.1. Then for any \( n \in \Gamma_{N,\delta} \) we have

\[
|q_N(m) - p_N(n)| \geq \frac{1}{2} B_\delta N^{k-1-s} - D.
\]

for some constant \( A_\delta > 0 \), which completes the proof that the differences of \( p \) and \( q \) explode in the case considered above.

Finally, suppose that \( c = u/v \) and \( t = w/t \) are rational and that \( C_u \equiv 0 \) for any \( u \) (i.e. that \( p \) and \( q \) and \( Q \)-equivalent). Then \( H_{N,y}(0) = \sum_{u=2}^{k-1} N^{-(u-1)} C_u(y) = 0 \) for any \( y \). As in the previous cases considered in this proof, when \( m \neq cn + r \) then \(|m - cn - r| \geq \delta_1 > 0 \) and so, by Lemma 3.2.2 if \( Q_k'(\gamma_k(cn/N)) > 2A_1 N^{-1} \) we have

\[
|q_N(m) - p_N(n)| \geq \frac{1}{2} Q_k'(\gamma_k(cn/N)) N^{k-1} (\delta_1 - BN^{-1}) - D
\]

for some constant \( B > 0 \). Here \( D = |q_0(r) - P_0| = |d| \) and we also used \( \text{(3.6)} \). Note that \( Q_k'(\gamma_k(y)) \geq C_\delta > 0 \) when \( y = n/N \in [\delta, 1] \). On the other hand, if \( m = cn + r \) then \( m = N\gamma_k(y) + R_k(y) \) and so

\[
q_N(m) - p_N(n) = d + Q_k'(\gamma_k(y)) N^{k-1} H_{N,y}(m - N\gamma_k(y) - R_k(y)) = q_0(r) - P_0 + Q_k'(\gamma_k(y)) N^{k-1} H_{N,0} = q_0(r) - P_0.
\]

\( \square \)

### 3.4. Application II: “fractional related” polynomials

In this section we will give several classes of examples for polynomials with exploding differences so that \( \gamma_k(y) \) is not a linear function of \( y \). We will rely on the following

#### 3.4.1. Lemma

Let \( a \) and \( b \) be positive coprime integers so that \( a > b \) and \( \alpha_0, \alpha_{b+1}, \ldots, \alpha_b \) be integers so that \( |\alpha_b| = 1 \). For each fixed \( N \), consider the equation

\[
m^a = \sum_{j=b}^{a} \alpha_j n^j N^{a-j} \quad \text{(3.9)}
\]

over the positive integers. Then for any \( \delta > 0 \) there exists a constant \( N_\delta \) and a set \( \Gamma_{N,\delta} \subseteq [1, N] \) whose cardinality does not exceed \( \delta N \) so that for any \( N > N_\delta \) and \( n \in [1, N] \setminus \Gamma_{N,\delta} \) there is no natural number \( m \) such that \( \text{(3.9)} \) holds true.

#### Proof

Fix a natural number \( N \). For any \( v \) that divides \( N \) set \( N_v = N/v \) and

\[
U_{v,N} = \{1 \leq n \leq N : \gcd(n, N) = v\} = \{1 \leq vn_v \leq N : \gcd(n_v, N_v) = 1\}
\]

and note that the cardinality \(|U_{v,N}|\) of \( U_{v,N} \) is just \( \varphi(N_v) \), where \( \varphi \) is the Euler totient function. Therefore

\[
\sum_{v|N} |U_{v,N}| = \sum_{v|N} \varphi(N_v) = N. \quad \text{(3.10)}
\]
Denote by $U'_{v,N}$ the set of all members of $U_{v,N}$ for which there exists $m$ satisfying (3.9). In order to find sets $\Gamma_{N,\delta}$ with the properties described in the statement of the lemma, it is enough to show that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{v|N} |U'_{v,N}| = 0$$

(3.11)

where $|U'_{v,N}|$ is the cardinality of $U'_{v,N}$, since the expression inside the limit is just $N^{-1}$ times the number of $n$'s in $[1, N]$ for which (3.9) holds with some $m$.

Let $n \in U'_{v,N}$ for some $v$ and $N$. Then (3.9) holds true for some $m$, and this $m$ must divide $v$. Therefore (3.9) also holds true with $n_v = n/v$ and $m_v = m/v$ and $N_v$ in place of $n, m$ and $N$, respectively. Let $p$ be a prime number that divides $n_v$ and let $e$ be the largest power of $p$ so that $p^e$ divides $n_v$. Then $eb$ is the largest power of $p$ that divides $n_v^b$. Write $eb = ka + w$ for some $k$ and $0 \leq w < a$. Then $p^{k+1}$ divides $m_v^a$ and so, if $w \neq 0$, then $p^{k+1}$ divides $m_v$, which implies that $p^a$ divides $n_v^b$ since

$$\sum_{j=b}^{a} \alpha_j n_v^{a-j} = n_v^b \left( \alpha_0 N_v^{a-b} + n_v \sum_{j=b+1}^{a} n_v^{j-1} N_v^{a-j} \right)$$

and the second factor of the right hand side is not divisible by $p$ (since gcd($n_v, N_v$) = 1 and $|\alpha_0| = 1$). This is clearly a contradiction since $ak + a > eb$. Therefore, $w = 0$ and so $eb = ka$ for some $k$. But $a$ and $b$ are coprime, and hence $k$ must have the form $k = k' b$ for some integer $k'$. Therefore we can write $e = k'a$, namely $n$ must have the form $n = vz^a$ for some integer $z$. Next, recall the following inequality (see Theorem 15 in [20]),

$$\varphi(M) \geq R(M) := \frac{M}{e^\gamma \ln \ln M + 3(\ln \ln M)}^n, \quad M > 2$$

(3.12)

where $\gamma$ is Euler’s constant. The equality $n_v = z^a$ clearly implies that gcd($z, N_v$) = 1 and therefore $vz$ is a member of $U_{v,N}$. The map $n \to z$ from $U'_{v,N}$ to $U_{v,N}$, where $n = vz^a$ is one to one, and its image is contained in the interval $[1, N_v^{\frac{1}{a}}]$ (since $z \leq N_v^{\frac{1}{a}}$), and hence $|U'_{v,N}| \leq N_v^{\frac{1}{a}}$. When $N_v > 2$ then, with $c = 3(\ln \ln 3)^{-1}$, applying (3.12) we derive that

$$|U'_{v,N}| \leq N_v^{\frac{1}{a}} = R(N_v) \cdot (e^\gamma \ln \ln N_v + c N_v^{1-\frac{1}{a}}) \leq |U_{v,N}|(e^\gamma \ln \ln (N_v) + c) N_v^{1-\frac{1}{a}}$$

which together with (3.10) yields (3.11).

The following result follows now from Lemma 3.2.2 and 3.4.1.

3.4.2. Corollary. Suppose that $R_k \equiv 0$ and that $\gamma_k$ has the form

$$\gamma_k(y) = \left( \sum_{j=b}^{a} \alpha_j y^j \right)^{\frac{1}{b}}$$

where $a, b$ and $\alpha_j$ satisfy the conditions of Lemma 3.4.1. Let $s \leq k - 1$ be so that $C_u \equiv 0$ for any $2 \leq u \leq s$. Suppose, in addition, that $a < s$. Then the polynomials $q$ and $p$ have exploding differences.
Remark. In order to prove the latter inequality, we define the polynomial $n$

Therefore, by Lemma 3.4.1, in order to complete the proof of Corollary 3.4.2, it is

Proof. Observe that $m = Nγ_k(n/N)$ if and only if the equation (3.10) holds true.

Therefore, by Lemma 3.4.1 in order to complete the proof of Corollary 3.4.2, it is

sufficient to show that for any $n$ and $m$ so that $m \neq Nγ_N(n/N)$ we have

\[ |q_N(m) − p_N(n)| ≥ Q_k(cγ_k(y))N^{k−1−n} \]  

(3.13)

where $y = n/N$ and $c$ is some constant. Since $x_N(y) ≤ BN^{−s}$ when $γ_k(y) > 2A_1 N^{−1}$, where $B > 0$ is some constant, then by Lemma 3.2.2 in order to show that (3.13) holds true, it is enough to show that for any $n$ and $m$ so that $m \neq Nγ_k(n/N)$ we have

In order to prove the latter inequality, we define the polynomial $W(x)$ by

\[ W(x) = x^a − \sum_{j=b}^a \alpha_j n^j N^{b−j}. \]

Then $W(Nγ_k(y)) = 0$ and

\[ |W(m)| = |m^a − \sum_{j=b}^a \alpha_j n^j N^{b−j}| ≥ 1 \]

since we have assumed that $m \neq Nγ_k(y)$. Therefore, by the mean value theorem, for some $ξ$ between $m$ and $Nγ_k(y)$,

\[ |m − Nγ_k(y)| = a^{−1}ξ^{−(a−1)}|W(m) − W(z)| = a^{−1}ξ^{−(a−1)}|W(m)| \]

\[ ≥ a^{−1}ξ^{−(a−1)} ≥ a^{−1}N^{−(a−1)} \]

and the proof of the corollary is complete. 

□

The following result also follows

3.4.3. Corollary. Suppose that the conditions of Proposition 3.1.1 hold and that $P(n, N) = \sum_{j=1}^a \alpha_j n^j N^{a−j}$ and $Q(m, N) = m^a$ for some positive coprime integers $a > b$ and integers $\alpha_j, ..., \alpha_a$ so that $|α_b| = 1$ (see Corollary 3.4.3). Then the polynomials $q$ and $p$ have exploding differences.

3.4.4. Remark. Suppose that $R_k ≡ 0$ and let $s$ be as in Corollary 3.4.2. Assume also that $γ_k(y)$ has the form $γ_k(y) = K^{−1}E(y)$ for some polynomials $K$ and $H$ with integer coefficients whose degrees do not exceed $d$, for some $d < s$. Consider the polynomial $W(x) = K(x/N) − E(n/N)$. Then $W(Nγ_k(n/N)) = 0$ and therefore, applying the mean value theorem yields that

\[ |m − NK^{−1}E(n/N)| = (W′(ξ))^{−1}|W(m)| \]

for some $ξ$ between $m$ and $Nγ_k(y)$ (which is of order $N$ when $n, m ∈ [δN, N]$). Notice that either $W(m) = 0$ or

\[ |W(m)| ≥ CN^{−d} \]

for some constant $C > 0$ (when $1 ≤ m ≤ N$). We conclude that for any $δ > 0$ there exists a constant $C_δ > 0$ so that for any $n, m ∈ [Nδ, N]$ we either have $m = NK^{−1}E(n/N)$ or

\[ |m − NK^{−1}E(n/N)| ≥ C_δ N^{−d−1}. \]

The difficulty in using the above estimates in order to determine whether the polynomials $q$ and $p$ have exploding differences arises here in determining for which $n$'s
there exists a solution \( m \) to the equation \( m = NK^{-1}E(n/N) \). When \( \gamma_k \) satisfies the conditions of Corollary 3.4.2 then we used Lemma 3.4.1 but for general polynomials \( K \) and \( H \) it does not seem possible to show that the equation \( m = NK^{-1}E(n/N) \) does not have a solution for any \( n \in [\delta N, N] \setminus \Gamma_{N,\delta} \) for sets \( \Gamma_{N,\delta} \) whose cardinality does not exceed \( \delta N \).

4. Expectation estimates

In this section we describe two results from [10] which are key ingredients in the proofs of Theorems 2.1.2, 2.3.1 and 2.3.3.

Let \( U_i, i = 1, 2, ..., k \) be \( d_i \)-dimensional random vectors defined on the probability space \((\Omega, \mathcal{F}, P)\) from Section 2 and \( \{C_j : 1 \leq j \leq s\} \) be a partition of \( \{1, 2, ..., k\} \). Consider the random vectors \( U(C_j) = \{U_i : i \in C_j\}, j = 1, ..., s \), and let

\[
U^{(j)}(C_j) = \{U_i^{(j)} : i \in C_j\}, j = 1, ..., s
\]

be independent copies of the \( U(C_j)'s \). For each \( 1 \leq i \leq k \) let \( a_i \in \{1, ..., s\} \) be the unique index such that \( i \in C_{a_i} \), and let \( H : \mathbb{R}^{d_1 + d_2 + ... + d_k} \to \mathbb{R} \) be such that \( \mathcal{D}(H) = |EH(U_1, U_2, ..., U_k) - EH(U_1^{(a_1)}, U_2^{(a_2)}, ..., U_k^{(a_k)})| \).

(4.1)

Of course, the above quantity is defined only when the expectations inside the absolute value exist.

The first result appears in [10] as Corollary 1.3.14, and it formulation is as follows:

4.0.1. Proposition. Let \( H : \mathbb{R}^{d_1 + d_2 + ... + d_k} \to \mathbb{R} \) be a function satisfying (2.6) and (2.7) with \( H \) in place of \( F \) and with \( u_i's \) in place of \( x_i's \). Let \( q, w > 1 \) and \( v > 0 \) be such that

\[
\frac{1}{w} > \frac{1}{v} + \frac{\kappa}{q}
\]

and set \( r_i = \lfloor \frac{1}{3}(m_i - n_{i-1}) \rfloor, i = 2, ..., k, r_1 = r_2 \) and \( r_{k+1} = r_k \). Suppose that \( \|U_i\|_{L^w} < \infty \) for any \( 1 \leq i \leq k \). Then \( \mathcal{D}(H) \) is well defined and

\[
\mathcal{D}(H) \leq 6R_0\Lambda_{w,\kappa}
\]

(4.2)

where

\[
\Lambda_{w,\kappa} = \left( \sum_{i=2}^{k} \phi(r_i) \right)^{1 - \frac{1}{w}} + \sum_{i=1}^{k} \left( \|U_i - \mathbb{E}[U_i|F_{m_i - r_i, n_i + r_{i+1}}]\|_{L^q} \right)^{\kappa}
\]

\( R_0 = K(1 + kg'_m) \) and \( g_m = \max\{\|U_i\|_{L^m} : 1 \leq i \leq k\} \). Here the \( \sigma \)-algebras \( \mathcal{F}_{k,l} \) are the ones specified in Section 2.

Next, recall that the \( \alpha \)-dependence coefficients of any two sub-\( \sigma \)-algebras \( \mathcal{G} \) and \( \mathcal{H} \) of \( \mathcal{F} \) is given by

\[
\alpha(\mathcal{G}, \mathcal{H}) = \sup \{\|P(\Gamma \cap \Delta) - P(\Gamma)P(\Delta)\| : \Gamma \in \mathcal{G}, \Delta \in \mathcal{H}\}.
\]

(4.3)

Another result we will need is the following proposition, which is stated in [10] as Corollary 1.3.11:
4.0.2. **Proposition.** Suppose that each \( U_i \) is \( \mathcal{F}_{m_i,n_i} \)-measurable, where \( n_{i-1} < m_i \leq n_i \), \( i = 1, \ldots, k \), \( n_0 = -\infty \) and \( m_{k+1} = \infty \). Then, for any bounded Borel function \( H : \mathbb{R}^{d_1+d_2+\ldots+d_k} \rightarrow \mathbb{R} \),

\[
\mathcal{D}(H) \leq 4 \sup |H| \sum_{i=2}^{k} \phi(m_i - n_{i-1})
\]

where \( \sup |H| \) is the supremum of \( |H| \). In particular, when \( s = 2 \) then

\[
\alpha(\sigma\{U(C_1)\}, \sigma\{U(C_2)\}) \leq 4 \sum_{i=2}^{k} \phi(m_i - n_{i-1})
\]

where \( \sigma\{X\} \) stands for the \( \sigma \)-algebra generated by a random variable \( X \) and \( \alpha(\cdot, \cdot) \) is given by (4.3).

In fact, Proposition 4.0.1 follows from Proposition 4.0.2 using standard approximations which involve the Hölder and the Markov inequalities (note the function \( H \) in Proposition 4.0.1 is locally Hölder continuous). The proof of Proposition 4.0.1 is based on the following lemma (see Lemma 1.3.10 in [10] or Lemma 3.11 in [9]).

4.0.3. **Lemma.** Let \( \mathcal{G}_1, \mathcal{G}_2 \subset \mathcal{F} \) be two sub-\( \sigma \)-algebras of \( \mathcal{F} \) and for \( i = 1, 2 \) let \( V_i \) be a \( \mathbb{R}^{d_i} \)-valued random \( \mathcal{G}_i \)-measurable vector with distribution \( \mu_i \). Set \( d = d_1 + d_2 \), \( \mu = \mu_1 \times \mu_2 \), denote by \( \nu \) the distribution of the random vector \((V_1, V_2)\) and consider the measure \( \nu = \frac{1}{2}(\kappa + \mu) \). Let \( B \) be the Borel \( \sigma \)-algebra on \( \mathbb{R}^d \) and \( H \in L^\infty(\mathbb{R}^d, \mathcal{B}, \nu) \). Then \( \mathbb{E}[H(V_1, V_2)|\mathcal{G}_1] \) and \( \mathbb{E}H(v, V_2) \) exist for \( \mu_1 \)-almost any \( v \in \mathbb{R}^{d_1} \) and

\[
|\mathbb{E}[H(V_1, V_2)|\mathcal{G}_1] - h(V_1)| \leq 2\|H\|_{L^\infty(\mathbb{R}^d, \mathcal{B}, \nu)} \phi(\mathcal{G}_1, \mathcal{G}_2), \quad P - \text{a.s.} \tag{4.4}
\]

where \( h(v) = \mathbb{E}H(v, V_2) \) and \( a.s. \) stands for almost surely.

For the sake of completeness, let us explain the idea behind the proof of Proposition 4.0.2. In the circumstances of Lemma 4.0.3, it follows by induction on \( k \) that

\[
|\mathbb{E}H(U_1, U_2, \ldots, U_k) - \int H(u_1, u_2, \ldots, u_k) d\nu_1(u_1) d\nu_2(u_2) \ldots d\nu_k(u_k)|
\]

\[
\leq 2 \sup |H| \sum_{i=2}^{k} \phi(m_i - n_{i-1}). \tag{4.5}
\]

where \( \nu_i \) is the distribution of \( U_i \). The proof of Proposition 4.0.2 is carried out by induction on the number of blocks \( s \), where the main ingredient in the induction step is (4.5) (applied with various functions appearing in the proof).

5. **First four moments growth and the SLLN**

5.1. **Linear growth of the variance.** For each \( n \) and \( N \) set

\[
F_{n,N} = F(\xi_{q_1(n,N)}, \ldots, \xi_{q_l(n,N)}).
\]

We begin with noting that by Assumption 2.1.1 and (2.0) there exists a constant \( B > 0 \) so that

\[
\sup_{n,N} \|F_{n,N}\|_{L^\infty} \leq B \tag{5.1}
\]

where \( w \) comes from Assumption 2.1.1 Next, for each \( l \) set

\[
a(l) = ((\phi(l/3))^{1-\frac{1}{\kappa}} + (\beta_q(l/3))^\kappa).
\]
here \( w \) and \( \theta \) come from Assumptions 2.1.1 and 2.1.3. In the circumstances of Theorems 2.3.1 and 2.3.3, we have \( \theta > \frac{1}{w} \) which implies that \( \sum_{l=0}^{\infty} (l+1) a(l) < \infty \).

Our first result shows, in particular, that the variance of \( S_N = S_N(1) \) behaves as the expectation of \( S_N^2 \):

5.1.1. **Lemma.** Suppose that Assumption 2.1.1 holds true and that \( \sum_{n=0}^{\infty} a(n) < \infty \). Then there exists a constant \( C > 0 \) so that for any \( N \geq 1 \),

\[
\sum_{n=1}^{N} |\mathbb{E} F_{n,N}| \leq C.
\]

**Proof.** Fix \( N \geq 1 \) and \( u \geq 0 \) and set \( k^* = \max\{\deg q_i : 1 \leq i \leq \ell\} \). Then there exists at most \( 2\ell^2 k^* \) number of natural \( n \)'s so that

\[
\delta N(n,n) := \min_{1 \leq i \neq j \leq \ell} |q_i(n,N) - q_j(n,N)| = u.
\]

Let \( 1 \leq n \leq N \) be so that \( \delta N(n,n) = u \). Then there exists a permutation \( \sigma \) of \( \{1,...,\ell\} \) so that

\[
q_{\sigma(i)}(n,N) + u \leq q_{\sigma(i+1)}(n,N) \quad \text{for any} \quad i = 1,2,...,\ell - 1.
\]

Set \( U_i = \xi_{\sigma(i)} \) and \( H(x_1,...,x_{\ell}) = F(x_{\sigma(\ell)},x_1,...,x_{\sigma(\ell)}) \). Applying Proposition 4.0.1 with the function \( H \) and the random vectors \( U_i \), when \( \delta N(n,n) = u \) then

\[
|\mathbb{E} F_{n,N}| \leq R_0 (\phi(u/3))^{1/4} + (\beta q(u/3))^v \leq R_1 a(u)
\]

where \( R_0 \) and \( R_1 \) are some constants and we also used (2.9). Note that, in the terminology of Proposition 4.0.1, we used the partition of the index set \( \{1,...,\ell\} \) into points: \( C_i = \{i\} \). We conclude that

\[
\sum_{n=1}^{N} |\mathbb{E} F_{n,N}| \leq 2\ell^2 k^* \sum_{u=0}^{\infty} a(u) := C < \infty
\]

and the proof of the lemma is complete. \( \square \)

Now we will show that the variance of \( S_N \) grows at most linearly fast in \( N \). In fact, we will prove the following

5.1.2. **Lemma.** Suppose that Assumption 2.1.1 holds true and that \( \sum_{l=0}^{\infty} a(l) < \infty \). Then there exists a constant \( C > 0 \) so that for any positive integers \( n_1 < n_2 < ... < n_M \) and \( N \in \mathbb{N} \) we have

\[
\mathbb{E} (\sum_{k=1}^{M} F_{n_k,N})^2 \leq CM
\]

where for each random variable \( Z \) we set \( \mathbb{E} Z^2 = \mathbb{E} [Z^2] \).

**Proof.** Fix some \( N \) and let \( n_1 < ... < n_M \) be positive integers. Using Lemma 5.1.1 it is enough to show that the variance of the sum \( \sum_{k=1}^{M} F_{n_k,N} \) is bounded by \( CM \) for some constant \( C \) which does not depend on \( N \) and the choice of \( n_i \). For each \( n \) and \( m \) set

\[
d_N(n,m) = \min_{1 \leq i,j \leq \ell} |q_i(n,N) - q_j(m,N)|.
\]

For each \( l \geq 0 \) set

\[
\Gamma_{n,N,l} = \{m : d_N(n,m) = l\}.
\]
Then, since each \( q_i(x, N) \) is a polynomial function of the variable \( x \) (for any fixed \( N \)), whose degree is bounded by some constant which does not depend on \( N \), we have

\[
|\Gamma_{n,N,l}| \leq A
\]

for some constant \( A \) which does not depend on \( n, N \) and \( l \). Here \( |\Gamma_{n,N,l}| \) denotes the cardinality of the set \( \Gamma_{n,N,l} \). Now we can write

\[
\text{Var}\left( \sum_{k=1}^{M} F_{n_k,N} \right) = \sum_{k=1}^{M} \sum_{l=0}^{\infty} \sum_{n_x \in \Gamma_{n_k,N,l} \cap [1,N]} \text{Cov}(F_{n_k,N}, F_{n_x,N}). \tag{5.3}
\]

Let \( n, m \in \mathbb{N} \) and \( \ell \geq 0 \) be so that \( d_N(n, m) = \ell \), and consider the sets

\[
\Gamma_1 = \Gamma_1(n) = \{q_i(n, N) : 1 \leq i \leq \ell \} \quad \text{and} \quad \Gamma_2 = \Gamma_1(m) = \{q_i(m, N) : 1 \leq i \leq \ell \}.
\]

Then, there exist disjoint sets \( Q_1, ..., Q_L, \ L \leq 2\ell + 1 \) so that

\[
\Gamma := \Gamma_1 \cup \Gamma_2 = \bigcup_{i=1}^{L} Q_i,
\]

each one of the \( Q_i \)'s is contained in either \( \Gamma_1 \) or \( \Gamma_2 \) and for any \( q_i \in Q_i \) and \( q_{i+1} \in Q_{i+1}, \ i = 1, 2, ..., L - 1 \) we have

\[q_{i+1} \leq q_{i+1}.
\]

Consider now the random vectors \( U_1, ..., U_L \) given by

\[U_i = \{\xi_j : j \in Q_i\}.
\]

Consider the partition \( \{C_1, C_2\} \) of the index set \( \{1, ..., L\} \) given by

\[C_i = \bigcup_{j: Q_j \subset \Gamma_i} Q_j, \ i = 1, 2.
\]

Then \( F_{n,N} \) is a function of \( \{U_j : j \in C_1\} \) and \( F_{n,N} \) is a function of \( \{U_j : j \in C_2\} \). Applying Proposition 4.0.1 with the function \( H(x, y) = F(x)F(y) \) we obtain that for any \( n = n_k, \ l \geq 0 \) and \( m = n_x \in \Gamma_{n_k,N,l} \),

\[|\text{Cov}(F_{n_k,N}, F_{n_x,N})| \leq R_0 ((\phi(l/3))^{1-\frac{d}{2}} + (\beta_q(l/3)^{c}) = R_0 a(l)
\]

where \( R_0 > 0 \) is some constant. Therefore, using [5.3], we obtain that

\[
\text{Var}\left( \sum_{k=1}^{M} F_{n_k,N} \right) \leq \sum_{k=1}^{M} \sum_{l=0}^{\infty} \sum_{n_x \in \Gamma_{n_k,N,l} \cap [1,N]} |\text{Cov}(F_{n_k,N}, F_{n_x,N})| \leq (AR_0 \sum_{l=0}^{\infty} a(l)) M
\]

and the proof of the lemma is complete. \( \square \)

5.2. Fourth moments and the strong law of large numbers. We will prove here the following

5.2.1. Lemma. Set \( b(l) = (\phi(l))^{1-\frac{d}{2}} + (\beta_q(l)^{c}) \) and assume that \( \sum_{l=1}^{\infty} lb(l) < \infty \). Then there exists a constant \( C > 0 \) so that for any positive integers \( n_1 < n_2 < ... < n_M \) and \( N \in \mathbb{N} \) we have

\[E\left( \sum_{k=1}^{M} F_{n_k,N} \right)^4 \leq CM^2.
\]
Relying on Lemma 5.2.1 and using the Markov inequality, we obtain that
\[ P\{N^{-1}|S_N| \geq N^{-1/8}\} \leq \frac{\|S_N\|_4^4}{N^{1/2}} \leq CN^{-3/2} \]
which together with the Borel-Cantelli lemma yields that with probability one we have
\[ \lim_{N \to \infty} \frac{1}{N} S_N = 0 = \bar{F}. \]

Proof of Lemma 5.2.1 Relying on Lemmas 5.1.1 and 5.1.2, it is enough to prove
\[ \text{Var}\left( \left( \sum_{k=1}^{M} F_{n_k,N} \right)^2 \right) \leq CM \]
for some $C$ which do not depend on $N$ and the choice of $n_1, \ldots, n_M$, where $\bar{X} = X - \mathbb{E}X$ for any random variable $X$. For any $u_1, u_2, v_1, v_2$ set
\[ b_N(u_1, u_2, v_1, v_2) = \text{Cov}(F_{v_1,N} \cdot b_{v_2,N}, F_{u_1,N} \cdot b_{u_2,N}). \]
For each $u$ and $N$ set
\[ \Gamma_{u,N} = \{ q_i(u, N) : 1 \leq i \leq \ell \}. \]
For any two sets $A_1, A_2 \subset \mathbb{R}$ set $\text{dist}(A_1, A_2) = \inf\{|a_1 - a_2| : a_i \in A_i\}$. Since the $q_i$’s are polynomials, there exists a constant $A > 0$, which does not depend on $N$, so that for each nonnegative integers $k_1$ and $k_2$ and positive integers $u_1$ and $u_2$, there exist at most $A$ pairs $(v_1, v_2)$ of positive integers such that
\[ k_1 = \text{dist}(\Gamma_{u_1,N} \cup \Gamma_{u_2,N}, \Gamma_{v_1,N} \cup \Gamma_{v_2,N}) \]
and, if $k_1 = \text{dist}(\Gamma_{u_1,N} \cup \Gamma_{u_2,N}, \Gamma_{v_1,N})$ then
\[ k_2 = \text{dist}(\Gamma_{u_1,N} \cup \Gamma_{u_2,N}, \Gamma_{v_2,N}) \]
while when $k_1 > \text{dist}(\Gamma_{u_1,N} \cup \Gamma_{u_2,N}, \Gamma_{v_1,N})$ (and so $k_1 = \text{dist}(\Gamma_{u_1,N} \cup \Gamma_{u_2,N}, \Gamma_{v_2,N})$) then
\[ k_2 = \text{dist}(\Gamma_{u_1,N} \cup \Gamma_{u_2,N}, \Gamma_{v_2,N}, \Gamma_{v_1,N}). \]
Let us denote the set of all these indexes $(v_1, v_2)$ by $\Delta(u_1, u_2, k_1, k_2, N)$. Note that given $N, u_1, u_2, v_1, v_2$ there exist $k_1$ and $k_2$ so that $(v_1, v_2) \in \Delta(u_1, u_2, k_1, k_2, N)$. Next, we claim that for any $u_1, u_2$ and $(v_1, v_2) \in \Delta(u_1, u_2, k_1, k_2, N)$ we have
\[ |b_N(u_1, u_2, v_1, v_2)| \leq R_0 \tau(\max(k_1, k_2)) \tag{5.5} \]
where $\tau(l) = b(l/3)$, $R_0$ is some constant and $b(\cdot)$ was defined in the statement of the lemma. Indeed, first consider the case when $k_1 \geq k_2$, and set $\Delta_1 = \Gamma_{u_1,N} \cup \Gamma_{u_2,N}$ and $\Delta_2 = \Gamma_{v_1,N} \cup \Gamma_{v_2,N}$. Then we can write
\[ \Delta_1 \cup \Delta_2 = \bigcup_{i=1}^{L} Q_i, \ L \leq 4\ell + 4 \]
where each $Q_i$ is subsets of either $\Delta_1$ or $\Delta_2$, and the $Q_i$’s satisfy all the conditions appearing right after (5.4) with $k_1$ in place of $l$. Consider the random vectors $U_1, \ldots, U_L$ given by $U_i = \{ \xi_j : j \in Q_i \}$ and the partition $\{ C_1, C_2 \}$ of the index set $\{ 1, \ldots, L \}$ given by
\[ C_i = \bigcup_{j \in Q_j \cap \Gamma_i} Q_j, \ i = 1, 2. \]
Then \( F_{u_1,N} \cdot F_{u_2,N} \) is a function of \( \{ U_j : j \in C_1 \} \) and \( F_{v_1,N} \cdot F_{v_2,N} \) is a function of \( \{ U_j : j \in C_2 \} \). Applying Proposition 4.0.1 with the function \( H(x, y, z, w) = F(x) \overline{F(y)} F(z) F(w) \) we obtain that
\[
|b_N(u_1, u_2, v_1, v_2)| \leq R_0 \tau(k_1)
\]
for some \( R_0 \), which completes the proof of (5.5) in the above case.

Next, consider the case when \( k_1 < k_2 \) and \( k_2 = \text{dist}(\Gamma_{u_1,N} \cup \Gamma_{u_2,N} \cup \Gamma_{v_1,N}, \Gamma_{v_2,N}) \). Set \( \Delta_1 = \Gamma_{u_1,N} \cup \Gamma_{u_2,N} \cup \Gamma_{v_1,N} \) and \( \Delta_2 = \Gamma_{v_2,N} \). Then dist(\( \Delta_1, \Delta_2 \)) = \( k_2 \). As in the previous case, applying Proposition 4.0.1 yields that
\[
|E[F_{u_1,N} \cdot F_{u_2,N} \cdot F_{v_1,N} \cdot F_{v_2,N}] - E[F_{u_1,N} \cdot F_{u_2,N} \cdot F_{v_1,N}] \cdot E[F_{v_2,N}]| \leq R_1 \tau(k_2)
\]
for some constant \( R_1 \). An additional application of this corollary yields
\[
|E[F_{u_1,N} \cdot F_{u_2,N} \cdot F_{v_1,N}] \cdot E[F_{v_2,N}]| \leq R_2 a(k_2) \leq R_2 \tau(k_2)
\]
where \( a(\cdot) \) was defined at the beginning of this section and \( R_2 \) is some constant. Taking into account that \( E[X] = 0 \) for any random variable \( X \), combining the above estimates with (5.4) we obtain that
\[
|b_N(u_1, u_2, v_1, v_2)| \leq R_0 \tau(k_2)
\]
for some constant \( R_0 > 0 \). The proof of (5.5) in the case when \( k_1 < k_2 \) and \( k_2 = \text{dist}(\Gamma_{u_1,N} \cup \Gamma_{u_2,N}, \Gamma_{v_1,N}, \Gamma_{v_2,N}) \) proceed exactly in the same way by changing the roles for \( v_1 \) and \( v_2 \).

Finally, applying (5.5) we obtain that
\[
\text{Var}(\sum_{k=1}^{M} F_{nk,N}) = \sum_{1 \leq i,j \leq M} \sum_{k_1,k_2=0}^{\infty} \sum_{(n_i,n_j) \in \Delta(n_i,n_j,k_1,k_2,N)} b_N(n_i,n_j,n_i1,n_i2)
\]
\[
\leq A \sum_{1 \leq i,j \leq M} \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{k_1} \tau(k_2)
\]
\[
+ A \sum_{1 \leq i,j \leq M} \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{k_1} \sum_{k=0}^{k_1} \tau(k_1) = 2R_0 A M^2 \sum_{k} (k+1)b(k) \leq CM^2
\]
where \( C \) is some constant, and the proof of the lemma is complete. \( \square \)

5.2.2. Remark. Relying on conditional expectation type estimates, it is possible to prove Lemma 5.2.1 similarly to Chapter 3 in [10], using the functions \( F_{\tau,i} \) defined in Section 6. Moreover, it is also possible to derive this lemma using the method of cumulants, similarly to Section 6 in [11]. In fact, the arguments leading to the cumulants estimates obtained in [11] can be modified to the setup of this paper, which means that we can also obtain moderate deviations theorems and some concentration inequalities for the sums \( S_N \).

6. Limiting covariances

6.1. Ordering and decomposition. Consider the homogeneous decomposition
\[
q_i(n, N) = \sum_{j=0}^{k} N^j Q_{i,j}(n/N) \tag{6.1}
\]
where for each \(i\) and \(j\) the polynomial \(Q_{i,j}\) is of degree not exceeding \(j\). Let \(r_0\) be so for any \(1 \leq i, j \leq \ell\) with \(\deg q_i > \deg q_j\), for any sufficiently large \(N\) we have
\[
\min_{N^{r_0} \leq n, m \leq N} (q_i(n, N) - q_j(m, N)) \geq N.
\] (6.2)

We will first prove the following

6.1.1. **Proposition.** There exist constants \(r_1 \in (0, 1), c > 0\) and \(A, B > 0\), sets \(B_N, N \geq 1\) containing \([1, N^{r_0}]\) so that \(|B_N| \leq AN^{r_1}\) and disjoint sets \(I_\epsilon(N)\) of the form
\[
I_\epsilon(N) = \bigcup_{j=1}^{\ell} (a_{j,\epsilon}N + BN^{r_1}, b_{j,\epsilon}N - N^{r_1}) \subset [1, N]
\] (6.3)
whose union cover \([1, N] \setminus B_N\), where \(\epsilon\) ranges over all the permutations of \(\{1, \ldots, \ell\}\) and the sets \((a_{j,\epsilon}N, b_{j,\epsilon}N)\)'s are disjoint, so that for any sufficiently large \(N\) and \(n \in I_\epsilon(N)\) with \(n \geq N^r\) we have
\[
q_{\epsilon(1)}(n, N) < q_{\epsilon(2)}(n, N) < \ldots < q_{\epsilon(\ell)}(n, N)
\] (6.4)
and when \(q_{\epsilon(i+1)}\) and \(q_{\epsilon(i)}\) have the same non-linear degree then
\[
q_{\epsilon(i+1)}(n, N) - q_{\epsilon(i)}(n, N) \geq cN^{r^+}.
\] (6.5)

**Proof.** We will prove the proposition by induction on the maximal degree of the polynomials \(q_1, \ldots, q_\ell\). When the maximal degree is 1, i.e. when all the polynomials are linear, then exactly as in Chapter 3 of [10] there exist a finite union of intervals of the form \((a_{j,\epsilon}N, b_{j,\epsilon}N)\), \(\epsilon \in \mathcal{E}_\ell\) whose union cover \([1, N] \setminus B_N\), for some set \(B_N\) whose cardinality does not exceed \(2\ell^2\), so that (6.3) holds true for any \(n \in (a_{\epsilon}N, b_{\epsilon}N)\), and
\[
q_{\epsilon(i+1)}(n, N) - q_{\epsilon(i)}(n, N) \geq \min(n - a_{\epsilon}N - 1, B_{\epsilon}N - n - 1)
\]
for any \(n \in (a_{\epsilon}, b_{\epsilon})\) and \(1 \leq i < \ell\). Now we can just take \(r_1 = \frac{1}{\ell}\) (when all of the polynomials are linear, (6.2) is meaningless). For readers' convenience, let us repeat the details from [10]. Write \(q_i(n, N) = p_in + q_iN, i = 1, \ldots, \ell\), and assume without loss of generality that \(p_1 < p_2 < \ldots < p_\ell\). Since \(p_iN + q_iN \geq 0\) always then \(q_i \geq 0\) for \(i = 1, \ldots, \ell\). Let \(\mathcal{N}_N\) be the set of \(n \in \{1, 2, \ldots, N\}\) such that all \(p_iN + q_iN, i = 1, \ldots, \ell\) are distinct. For each \(n \in \mathcal{N}_N\) we define distinct integers \(\epsilon_i(n, N), i = 1, 2, \ldots, \ell\) such that
\[
p_{\epsilon_i(n, N)}N + q_{\epsilon_i(n, N)}N < p_{\epsilon_{i+1}(n, N)}N + q_{\epsilon_{i+1}(n, N)}N \quad \text{for all} \quad i = 1, 2, \ldots, \ell - 1.
\] (6.6)
For each \(\epsilon = (\epsilon_1, \ldots, \epsilon_\ell) \in \mathcal{E}_\ell\) set
\[
\mathcal{N}_{\epsilon,N} = \{n \in \{1, 2, \ldots, N\} : \epsilon_j(n, N) = \epsilon_j \quad \text{for each} \quad j = 1, \ldots, \ell\}.
\]
Some of the sets \(\mathcal{N}_{\epsilon,N}\) can be empty and for each \(n \in \mathcal{N}_{\epsilon,N}\),
\[
p_{\epsilon_i}n + q_{\epsilon_i}N < p_{\epsilon_{i+1}}n + q_{\epsilon_{i+1}}N, \quad \text{i.e.}
\]
\[
n > (q_{\epsilon_i} - q_{\epsilon_{i+1}})(p_{\epsilon_{i+1}} - p_{\epsilon_i})^{-1}N \quad \text{if} \quad \epsilon_{i+1} > \epsilon_i \quad \text{and}
\]
\[
n < (q_{\epsilon_i} - q_{\epsilon_{i+1}})(p_{\epsilon_{i+1}} - p_{\epsilon_i})^{-1}N \quad \text{if} \quad \epsilon_{i+1} < \epsilon_i.
\]
Hence,
\[
\mathcal{N}_{\epsilon,N} = \{n : a_{\epsilon}N < n < b_{\epsilon}N\}
\]
for some (not unique) \(a_{\epsilon} \geq 0\) and \(b_{\epsilon} \leq 1\), \(\mathcal{N}_{\epsilon,N}\) are disjoint for different \(\epsilon \in \mathcal{E}_\ell\) and, clearly,
\[
\mathcal{N}_N = \bigcup_{\epsilon \in \mathcal{E}_\ell} \mathcal{N}_{\epsilon,N}.
\] (6.7)
There is always \( \varepsilon = (\varepsilon_1, \ldots, \varepsilon_{\ell}) \) with \( \varepsilon_{\ell} = 0 \) and then \( \varepsilon_1 = \min\{i : q_i = \min_{1 \leq j \leq \ell} q_j\} \) and \( \varepsilon_\ell = \max\{i : q_i = \max_{1 \leq j \leq \ell} q_j\} \). Set \( B_N = \{1, 2, \ldots, N\} \setminus \mathcal{N}_N \). Then

\[
B_N \subset \{(q_i - q_j)(p_j - p_i)^{-1} : \text{for some } i, j = 1, \ldots, \ell, i \neq j\},
\]

and so the cardinality of \( B_N \) does not exceed \( \ell^2 \) (since the polynomials are linear).

In Lemma 3.2.6 in \( [10] \) it was proved that if \( a_{\varepsilon} N < n < b_{\varepsilon} N \) then

\[
(p_{\varepsilon_{\ell+1}} - p_{\varepsilon_1})n + (q_{\varepsilon_{\ell+1}} - q_{\varepsilon_1})N \geq \min(n - a_{\varepsilon} N - 1, b_{\varepsilon} N - n - 1).
\]

(6.8)

The proof of (6.8) is as follows. Let \( m_1, m_2 \geq 1 \) be integers satisfying \( a_{\varepsilon} N < n - m_1 \) and \( n + m_2 < b_{\varepsilon} N \). Then

\[
p_{\varepsilon_i}(n - m_1) + q_{\varepsilon_i}N < p_{\varepsilon_{\ell+1}}(n - m_1) + q_{\varepsilon_{\ell+1}}N
\]

and

\[
p_{\varepsilon_i}(n + m_2) + q_{\varepsilon_i}N < p_{\varepsilon_{\ell+1}}(n + m_2) + q_{\varepsilon_{\ell+1}}N.
\]

Hence,

\[
(p_{\varepsilon_{\ell+1}} - p_{\varepsilon_1})n + (q_{\varepsilon_{\ell+1}} - q_{\varepsilon_1})N \geq m_1
\]

if \( \varepsilon_{\ell+1} > \varepsilon_1 \) and if \( \varepsilon_{\ell+1} > \varepsilon_1 \) then

\[
(p_{\varepsilon_{\ell+1}} - p_{\varepsilon_1})n + (q_{\varepsilon_{\ell+1}} - q_{\varepsilon_1})N \geq m_2
\]

and so the assertion follows.

Now we will make the induction step. Suppose that the proposition holds true when the maximal degree does not exceed \( d \). Let \( q_1, \ldots, q_{\ell} \) be polynomials so that the maximal degree equals \( d + 1 \). Let \( k > \ell \) be so that \( q_1, \ldots, q_k \) are of degree strictly less than \( d + 1 \), and \( \deg q_i = d + 1 \) for any \( i > k \). By the induction hypothesis, there exist constants \( r_1 = r_1(H) \in (0, 1) \) and \( A = A_H, c = c_H > 0 \) and sets \( B_N = B_{H,N}, N \geq 1 \) and \( I_{\varepsilon'}(N) \) satisfying all the properties described in the statement of the proposition with the polynomials \( q_1, \ldots, q_k \), where \( \varepsilon' \) ranges over all the permutations of the set \( \{1, \ldots, k\} \).

In order to complete the induction hypothesis, it is enough to show that all the results stated in the proposition hold true for the family of polynomials \( q_{k+1}, \ldots, q_{\ell} \) (because of (6.2)). Indeed, assume that there exist constants \( r_1 = r_1(H) \in (0, 1) \) and \( c = c_{d+1} \) and sets \( B_N = B_{H,N,d+1} \) and \( I_{\varepsilon''}(N) \) with the properties described in the statement of Proposition \( [6.1.1] \) for the polynomials \( q_{k+1}, \ldots, q_{\ell} \), where \( \varepsilon'' \) ranges over all the permutations of the set \( \{k+1, \ldots, \ell\} \). Take \( 1 > r_1 > \max(r_1(H), r_{H,d+1}, r_0) \), where \( r_0 \) comes from (6.2). Consider all the endpoints of the intervals defining the sets \( I_{\varepsilon'}(N) \) and \( I_{\varepsilon''}(N) \) which are larger than \( N^{r_1} \). For any endpoint \( a \), consider the set \( E_a \) of all endpoints \( b \) so that \( |a - b| \) is sublinear in \( N \). Then we can partition the set of all endpoints to disjoint sets of the form \( E_a \). By omitting all the endpoints from each partition set \( E_a \) except \( a \), and then considering all the intervals generated by two consecutive (remaining) endpoints we get sets \( I_{\varepsilon}(N) \) with the desired properties (for the polynomials \( q_1, \ldots, q_{\ell} \)). Note that for any permutation of \( \{1, \ldots, \ell\} \) which does not have the form \( \varepsilon' \otimes \varepsilon'' \) we have \( I_{\varepsilon}(N) = \emptyset \).

Next, consider the decompositions \( [6.11] \) of the polynomials \( q_{k+1}, \ldots, q_{\ell} \) (whose degree is \( d + 1 \)). Let \( Q_{i_1,d+1}, \ldots, Q_{i_\ell,d+1} \) be the distinct polynomials among the polynomials \( Q_1,d+1, \ldots, d+1 \). Let \( 0 \leq y_1 < y_2 < \ldots < y_s \leq 1 \) be the set of all points \( y \) in \([0, 1]\) so that \( Q_{i_j,d+1}(y) = Q_{i_{j'},d+1}(y) \) for some \( j \neq j' \). On each interval of the form \( (y_a, y_{a+1}) \) we can order the polynomials \( Q_{i_j,d+1} \). In fact, since
the degrees of the polynomials $Q_{i_j,d+1}$ is at most $d + 1$, there exists a permutation $\sigma_a$ of $\{i_1,\ldots, i_u\}$ so that for any
\[
y \in [y_a + N^{-\frac{d}{d+1}}, y_{a+1} - N^{-\frac{d}{d+1}}] =: J_{a,n}
\]
and $1 \leq j < u$ we have
\[
Q_{\sigma_a(i_{j+1}),d}(y) > Q_{\sigma_a(i_j),d}(y) + CN^{-\frac{1}{2}}
\]
for some constant $C > 0$. Since the degree of $q_i$, $i > k$ is $d + 1$, it follows that for any $j$ and $n$ that when $n/N \in J_{a,N}$ we have
\[
q_{\sigma_a(i_{j+1})}(n,N) - q_{\sigma_a(i_j)}(n,N) \geq C_1 N^{d+1-\frac{1}{2}}
\]
where $C_1 > 0$ is some constant. In the case when all of the $Q_{i_j,d+1}$’s are distinct, we have completed the induction step. Otherwise, set $\Gamma_j = \{i : Q_{i_j,d+1} = Q_{i_k,d+1}\}$. Then, for any $a$ and $n/N \in J_{a,N}$, $j < j'$ and $i \in \Gamma_{\sigma_a(j)}$, $i' \in \Gamma_{\sigma_a(j')}$ we have
\[
q_i(n,N) - q_i(n,N) \geq C' N^{d+1-\frac{1}{2}}
\]
for some constant $C' > 0$. Finally, for each $j$ and $i \in \Gamma_j$ we define
\[
\tilde{q}_i(n,N) = q_i(n,N) - N^{d+1} Q_{i_j,d+1}(n/N) = q_i(n,N) - N^{d+1} Q_{i_j,d+1}(n/N).
\]
Then the degrees of the polynomials $\{\tilde{q}_i : i \in \Gamma_j\}$ do not exceed $d$, and so we can apply the induction hypothesis with them. By intersecting the resulting sets $I_{\varepsilon,j}(N)$ with each interval $J_{a,N}$, where $\varepsilon^{(j)}$ ranges over all the permutations of the set $\Gamma_j$, omitting the intervals whose length is a sublinear function of $N$ and taking a sufficiently large $r \in (0,1)$ we get disjoint sets $I_{\varepsilon''}(N)$ which have the properties described in the statement of the proposition for the polynomials $q_k + 1, \ldots, q_\ell$ (the ones with degree $d + 1$), where $\varepsilon''$ now ranges over the permutations of the set $\{k + 1, \ldots, \ell\}$. As we have explained before, it is enough in order to complete the induction step. The proof of the proposition is complete. 

Let $\mathcal{E}_\ell$ be the set of all permutations of the set $\{1, 2, \ldots, \ell\}$. Let $0 = s_0 < s_1 < s_2, \ldots < s_{z-1} < s_z = \ell$ and $d_1 < \ldots < d_z$ be so that the degree of $q_i$, $s_i < i \leq s_{i+1}$ is $d_i$, for any $i = 0, 1, \ldots, z - 1$. Then, as can bee seen from the proof of Proposition 6.1.1, the sets $I_{\varepsilon}(N)$ are not empty only when the permutation $\varepsilon$ preserves the sets $D_i = \{s_i + 1, \ldots, s_{i+1}\}$, where $i = 0, 1, \ldots, z - 1$. In other words
\[
\varepsilon = \varepsilon^{(0)} \otimes \varepsilon^{(1)} \otimes \cdots \otimes \varepsilon^{(z-1)}
\]
is the tensor product of $z$ permutations $\varepsilon^{(i)}$ of the sets $D_i$, where $i = 0, 1, \ldots, z - 1$. We denote by $\mathcal{E}'_{\ell}$ the set of all permutations with the above product structure.

Next, let $I_{\varepsilon}(N)$, $\varepsilon \in \mathcal{E}'_{\ell}$, be the sets from Proposition 6.1.1 and set
\[
G_N = \bigcup_{\varepsilon \in \mathcal{E}'_{\ell}} I_{\varepsilon}(N), \quad S_{1,N} = N^{-\frac{1}{2}} \sum_{n \in G_N} F(\xi_{q_1(n,N)}, \xi_{q_2(n,N)}, \ldots, \xi_{q_\ell(n,N)})
\]
and $I_{\varepsilon}(t,N) = I_{\varepsilon}(N) \cap [1, NT]$, $t \in [0,1]$.

Since the cardinality of the set $B_N = N \setminus G_N$ does not exceed $AN^r$ for some $A > 0$ and $r \in (0,1)$, using Lemma 6.1.2 we see that if the limits
\[
b(t,s) = \lim_{N \to \infty} E[B_{s_1[N]}B_{s_1[N]s}]
\]
exist, then we also have $b(t,s) = \lim_{N \to \infty} E[S_{N}(t)S_{N}(t)]$. 

\[26\] Y. Hafouta
For each \( \varepsilon \in \mathcal{E}_t^1 \) set
\[
S_{1, \varepsilon, N} = N^{-\frac{1}{2}} \sum_{n \in I_\varepsilon(N)} F_\varepsilon(x_{\varepsilon(1)}(n, N), x_{\varepsilon(2)}(n, N), \ldots, x_{\varepsilon(\ell)}(n, N))
\]
where \( F_\varepsilon(x_{\varepsilon(1)}, x_{\varepsilon(2)}, \ldots, x_{\varepsilon(\ell)}) = F(x_1, x_2, \ldots, x_\ell) \). Then \( S_{1, N} = \sum_{\varepsilon \in \mathcal{E}_t^1} S_{1, \varepsilon, N} \). For each \( \varepsilon \), we consider the decomposition of \( F_\varepsilon \)
\[
F_\varepsilon(x_{\varepsilon(1)}, \ldots, x_{\varepsilon(\ell)}) = \sum_{j=1}^\ell F_{\varepsilon, j}(x_{\varepsilon(1)}, \ldots, x_{\varepsilon(j)})
\]
where
\[
F_{\varepsilon, j}(x_{\varepsilon(1)}, \ldots, x_{\varepsilon(j)}) = F_\varepsilon(x_1, \ldots, x_\ell) - \int F(x_1, \ldots, x_\ell) d\mu(x_{\varepsilon(\ell)})
\]
and for all \( j = \ell - 1, \ell - 2, \ldots, 1, \)
\[
F_{\varepsilon, j}(x_{\varepsilon(1)}, \ldots, x_{\varepsilon(j)}) = \int F(x_1, \ldots, x_\ell) d\mu(x_{\varepsilon(\ell)}) d\mu(x_{\varepsilon(\ell-1)}) \cdots d\mu(x_{\varepsilon(j+1)})
- \int F(x_1, \ldots, x_\ell) d\mu(x_{\varepsilon(\ell)}) d\mu(x_{\varepsilon(\ell-1)}) \cdots d\mu(x_{\varepsilon(j)}).
\]
Observe that for any \( \varepsilon, i \) and \( y_{\varepsilon(i)}, \ldots, y_{\varepsilon(i-1)} \) we have
\[
\int F_{\varepsilon, i}(y_{\varepsilon(1)}, \ldots, y_{\varepsilon(i-1)}, x) d\mu(x) = 0. \quad (6.10)
\]
Set
\[
S_{\varepsilon, j, N} = \sum_{n \in I_\varepsilon(N)} F_{\varepsilon, j}(x_{\varepsilon(1)}(n, N), \ldots, x_{\varepsilon(\ell)}(n, N)).
\]
Then the limit \( b(t, s) \) exists if the limits
\[
D_{\varepsilon, \tau, i, j}(t, s) = \lim_{N \to \infty} \mathbb{E}[S_{\varepsilon, i, [N]j} S_{\tau, j, [N]i}]
\]
exist, for any \( \sigma, \tau \in \mathcal{E}_t^1 \) and \( 1 \leq i, j \leq \ell \).

6.2. Existence of the limiting covariances: proof of theorem 2.3.1. In the course of the proof of Theorem 2.3.1 we will need the following

6.2.1. Lemma. For any \( N, \varepsilon, \tau \in \mathcal{E}_t, 1 \leq i, j \leq \ell, n \in I_\varepsilon(N) \) and \( m \in I_\tau(N) \) set
\[
b_{\varepsilon, \tau, i, j, N}(n, m) = \mathbb{E}[F_{\varepsilon, i}(x_{\varepsilon(1)}(n, N), \ldots, x_{\varepsilon(\ell)}(n, N)) F_{\tau, j}(x_{\tau(1)}(m, N), \ldots, x_{\tau(\ell)}(m, N))].
\]
Then
\[
b_{\varepsilon, \tau, i, j, N}(n, m) \leq C v_N(|q_{\varepsilon(i)}(n, N) - q_{\tau(j)}(m, N)|)
\]
where \( v_N(l) = \left( \phi \left( \min([N^{1/2}], [l/3]) \right) \right)^{1 - \frac{1}{\omega}} + \left( \beta \left( \min([N^{1/2}], [l/3]) \right) \right)^{k} \) and \( C \) is some constant.

Suppose that \( \theta(1 - \frac{1}{\omega}) > 4 \) (as in Theorems 2.3.1 and 2.3.3). Then, under Assumption 2.1.3 we have
\[
v_N(l) \leq C \max(N^{-2}, l^{-4})
\]
where \( C \) is some constant. In particular
\[
\lim_{N \to \infty} N v_N(N) = 0 \quad (6.11)
\]
and
\[ Y := \sup_N \sum_{l=0}^{\infty} (l+1)v_N(l) < \infty. \] (6.12)

**Proof of Lemma 6.2.1.** Relying on (5.1), the statement of the lemma clearly holds true when \( q_{c(i)}(n, N) = q_{r(j)}(m, N) \). Suppose now that \( l = q_{c(i)}(n, N) - q_{r(j)}(m, N) > 0 \). Consider the variables \( u_c = (x_{c(1)}, \ldots, x_{c(i)}) \), \( u_r = (y_{r(1)}, \ldots, y_{r(j)}) \) and \( u = (u_c, u_r) \). Consider also the function \( H = H(u) \) given by
\[ H(u) = F_{r,j}(u_r)F_{c,i}(u_c). \]

Set
\[ A_1 = \{ q_{c(w)}(m, N) : 1 \leq w \leq j \} \cup \{ q_{c(w)}(n, N) : 1 \leq w < i \} \quad \text{and} \quad A_2 = \{ q_{r(i)}(n, N) \}. \]

Then by (6.2), (6.5) for any \( a \in C_1 \) and \( b \in C_2 \) we have
\[ a \leq b - \min(l, c_1 N^{\frac{1}{2}}) \]
where \( c_1 = \min(c, 1) \). Consider the random vectors \( U_1 \) and \( U_2 \) given by
\[ U_i = \{ \xi_d : d \in A_i \}. \]

Then
\[ b_{\varepsilon,\tau,i,j,N}(n, m) = EH(U_1, U_2). \]

Let \( \hat{U}_2 \) be a copy of \( U_2 \) which is independent of \( U_1 \). Then by (6.10),
\[ \boxed{ \begin{align*} &EH(U_1, \hat{U}_2) \\ &= EH(F_{r,j}(\xi_{q_{r(j)}(m, N)}), \ldots, \xi_{q_{r(j)}(m, N)}) \cdot EH(F_{r,j}(\xi_{q_{r(i)}(n, N)}), \ldots, \xi_{q_{r(i)}(n, N)}, U_2|U_1) \end{align*}} \]
where we used that the law of \( U_2 \) is \( \mu \). Taking into account Assumption 2.1.1 (2.5) and (2.6), applying Proposition 4.0.1 we obtain that
\[ |b_{\varepsilon,\tau,i,j,N}(n, m)| = |EH(U_1, U_2) - EH(U_1, \hat{U}_2)| \leq C\nu_N(l) \]
where \( C \) is some constant and \( \nu_N(l) \) was defined in the statement of the lemma. The proof of the lemma in the case when \( q_{c(i)}(n, N) - q_{r(j)}(m, N) < 0 \) is analogous. \( \square \)

Now we will prove

**6.2.2. Proposition.** Suppose that the conditions of Theorem 2.2.1 hold true. Then the limits \( b(t, s) \) exist. In fact:
(i) If \( \deg q_{c(i)} \neq \deg q_{r(j)} \) then the limits \( D_{\varepsilon,\tau,i,j}(t, s) \) exist and equals 0.
(ii) If \( \varepsilon = \tau, i = j \) and \( \deg q_{c(i)} > 1 \) then
\[ D_{\varepsilon,\tau,i,j}(t, s) = \min(s, t) c \int F_{c,i}(x_{c(1)}, \ldots, x_{c(i)}) d\mu(x_{c(1)}) \cdot d\mu(x_{c(2)}) \cdots d\mu(x_{c(i)}) \]
where
\[ c = \lim_{N \to \infty} N^{-1} |I_{\varepsilon}(N)| = \sum_{j=1}^{l_{\varepsilon}} \max(b_j, \varepsilon - a_j, 0) \]
and \( b_j, \varepsilon \) and \( a_j, \varepsilon \) are defined in Proposition 6.1.1.
(iii) If \( q_{c(i)} \) and \( q_{r(j)} \) have exploding differences then the limit \( D_{\varepsilon,\tau,i,j}(t, s) \) exists and equals 0.
(iv) If \( q_{\varepsilon(i)} \) and \( q_{\varepsilon(j)} \) are linearly related, \( \deg q_{\varepsilon(i)} = \deg q_{\varepsilon(j)} = k > 1 \) and the differences of \( q_{\varepsilon(i)} \) and \( q_{\varepsilon(j)} \) do not explode, then the limit \( D_{\varepsilon,\tau,i,j}(t,s) \) exists and has the form

\[
c_{\varepsilon,i,\tau,j}(s,t) \int F_{\varepsilon,i}(\tilde{x}) F_{\varepsilon,j}(\tilde{y}) d\mathcal{M}(\tilde{x}, \tilde{y})
\]

where \( \mathcal{M} \) is defined by (6.12) and the number \( c_{\varepsilon,i,\tau,j}(s,t) \) can be recovered from the proof.

Proof. As we have explained before, the limits \( b(t,s) \) exist if the limits

\[
D_{\varepsilon,\tau,i,j}(t,s) = \lim_{N \to \infty} \mathbb{E}[S_{\varepsilon,i,[N]}S_{\tau,j,[N]}]
\]

exist, for any \( \varepsilon, \tau \in \mathcal{E}'_l \) and \( 1 \leq i, j \leq \ell \).

Let \( \varepsilon, \tau \in \mathcal{E}'_l \) and \( 1 \leq i, j \leq \ell \). When \( \deg q_{\varepsilon(i)} = \deg q_{\varepsilon(j)} = 1 \), then existence of the limits \( D_{\varepsilon,\tau,i,j}(t,s) \) is obtained exactly as in Chapter 3 of [10], taking into account that \( \varepsilon, \tau \) have the tensor product structure (6.9). For readers’ convenience, let us explain the main idea of the proof in [10]. First, it enough to show that the limit

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{a_{\varepsilon,N} < m < b_{\varepsilon,N}, \alpha_{\varepsilon,N} < n < b_{\varepsilon,N}} b_{i,j,\varepsilon,\tau}(N, m, n)
\]

exists for each integer \( u \) (where the sum over the empty set is considered to be zero) and since then we can just over all \( u \)'s. The idea behind proving (6.13) is to show that if \( N \to \infty, m - a_{\varepsilon,N} \to \infty, b_{\varepsilon,N} - m \to \infty, n - a_{\varepsilon,N} \to \infty, b_{\varepsilon,N} - n \to \infty \) and \( \rho_{\varepsilon}(m,N) - \rho_{\varepsilon}(n,N) = p_{\varepsilon,m} - p_{\varepsilon,n} + N(q_{\varepsilon_i} - q_{\varepsilon_j}) = u \) then \( b_{i,j,\varepsilon,\tau}(N, m, n) \) converges to a limit. After that, using Assumption [2.2.1] we estimated the number of solutions of the equation

\[
p_{\varepsilon,m} - p_{\varepsilon,n} + N(q_{\varepsilon_i} - q_{\varepsilon_j}) = u
\]

in \( m \in \mathcal{N}_{\varepsilon,N} \) and \( n \in \mathcal{N}_{\varepsilon,N} \), divided this number by \( N \) and showed that this ratio converges to a limit which will give the limit in (6.13). We refer the readers’ to the Section 3.3.2 in Chapter 3 of [10] for the precise details.

Next, suppose that \( \max(\deg q_{\varepsilon(i)}, \deg q_{\varepsilon(j)}) > 1 \). In what follows, we will always assume that \( \deg q_{\varepsilon(i)} \geq \deg q_{\varepsilon(j)} \). We first write

\[
\mathbb{E}[S_{\varepsilon,i,[N]}S_{\tau,j,[N]}] = N^{-1} \sum_{n \in I_{\varepsilon}(N,t)} \sum_{m \in I_{\tau}(N,t)} b_{\varepsilon,\tau, i, j,N}(m,n)
\]

where \( I_{\varepsilon}(N,t) = I_{\varepsilon}(N) \cap [1,Nt] \) and \( I_{\tau}(N,s) = I_{\tau}(N) \cap [1,Ns] \). Suppose first that \( \deg q_{\varepsilon(i)} > \deg q_{\varepsilon(j)} \). Then by (6.2) we have

\[
\min_{N^0 \leq n,m \leq N} (q_{\varepsilon(i)}(n,N) - q_{\varepsilon(j)}(m,N)) \geq N
\]

and therefore, by Lemma [6.11] for any \( n \in I_{\varepsilon}(N) \) and \( m \in I_{\tau}(N) \),

\[
|b_{\varepsilon,\tau, i, j,N}(n,m)| \leq C\nu_{N}(N).
\]

Hence, by (6.11) we have

\[
|\mathbb{E}[S_{\varepsilon,i,[N]}S_{\tau,j,[N]}]| \leq \nu_{N}(N) \to 0 \quad \text{when} \quad N \to \infty.
\]

Now we will consider the case when \( \deg q_{\varepsilon(i)} = \deg q_{\varepsilon(j)} > 1 \). Suppose that \( \varepsilon = \tau \) and \( i = j \). If \( n \neq m \) then

\[
|q_{\varepsilon(i)}(n,N) - q_{\varepsilon(i)}(m,N)| \geq C_1 \min(n,m)
\]
where \( C_1 > 0 \) is some constant, and therefore by Lemma \( 6.2.1 \) we have

\[
|b_{\varepsilon, \tau, i,j,N}(n, m)| = |b_{\varepsilon, \tau, i,i,N}(n, m)| \leq CV_N(C_1 \min(n, m))
\]

which implies that

\[
\lim_{N \to \infty} N^{-1} \sum_{n \in I_s(N, t) \setminus \Gamma_N \setminus \delta N} \sum_{m \in I_s(N, s) \cap [\delta N, N]} |b_{\varepsilon, \tau, i,i,N}(n, m)| = 0
\]

where we used (6.11). On the other hand, when \( n = m \) we have

\[
b_{\varepsilon, \tau, i,i,N}(n, n) = \int F_{\varepsilon,i}(x_{\varepsilon(1)}, \ldots, x_{\varepsilon(i)})d\mu(x_{\varepsilon(1)}) \cdots d\mu(x_{\varepsilon(i)})
\]

and therefore by (6.15),

\[
\lim_{N \to \infty} E[S_{\varepsilon,i,[N]}S_{\varepsilon,j,[N]}] = \int F_{\varepsilon,i}(x_{\varepsilon(1)}, \ldots, x_{\varepsilon(i)})d\mu(x_{\varepsilon(1)}) \cdots d\mu(x_{\varepsilon(i)})
\]

and the proof of Proposition \( 6.2.2 \) (ii) is completed.

Suppose next that the differences of \( q_{\varepsilon(i)} \) and \( q_{\varepsilon(j)} \) explode. Fix some \( \delta \in (0, 1) \). Then there exist constants \( C_\delta > 0 \), \( N_\delta > 0 \) and sets \( \Gamma_{N, \delta} \subset [1, N] \) of integers whose cardinality does not exceed \( \delta N \) so that for any \( N > N_\delta \) and \( n \in [\delta N, N] \setminus \Gamma_{N, \delta}, \)

\[
\min_{m \in [\delta N, N]} |q_{\varepsilon(i)}(n, N) - q_{\varepsilon(j)}(m, N)| \geq C_\delta N.
\]

Write

\[
E[S_{\varepsilon,i,[N]}S_{\varepsilon,j,[N]}] = J_1 + J_2 + J_3
\]

where \( J_i = J_i(N, \varepsilon, \tau, i,j, s, t, \delta) \) are given by

\[
J_1 = N^{-1} \sum_{n \in I_s(N, t) \setminus \Gamma_N \setminus \delta N} \sum_{m \in I_s(N, s) \cap [\delta N, N]} b_{\varepsilon, \tau, i,j,N}(n, m),
\]

\[
J_2 = N^{-1} \sum_{n \in I_s(N, t) \setminus \Gamma_N \setminus \delta N} \sum_{m \in I_s(N, s) \cap [1, \delta N]} b_{\varepsilon, \tau, i,j,N}(n, m) \quad \text{and}
\]

\[
J_3 = N^{-1} \sum_{n \in I_s(N, t) \cap \Gamma_N \setminus \delta N} \sum_{m \in I_s(N, s)} b_{\varepsilon, \tau, i,j,N}(n, m).
\]

We will show that the upper limit as \( N \to \infty \) of each one of the \( J_i \)'s does not exceed \( \delta \), which by taking \( \delta \to 0 \) will complete the proof of Proposition \( 6.2.2 \) (iii). First by Lemma \( 6.2.1 \)

\[
|J_1| \leq N^{-1} \sum_{n \in I_s(N, t) \setminus \Gamma_N \setminus \delta N} \sum_{m \in I_s(N, s) \cap [\delta N, N]} |b_{\varepsilon, \tau, i,j,N}(n, m)| \leq C'_\delta N V_N([C_\delta N])
\]

for some constant \( C'_\delta > 0 \), and so by (6.11) we have \( \lim_{N \to \infty} J_1 = 0 \). Next,

\[
|J_2| \leq N^{-1} \sum_{m=1}^{\infty} \sum_{k=0}^{\delta N} \sum_{n \in A_k(m, N)} |b_{\varepsilon, \tau, i,j,N}(n, m)|
\]

where \( A_k(m, N) \) is the set of all \( n \)'s so that \( |q_{\varepsilon(i)}(n, N) - q_{\varepsilon(j)}(m, N)| = k \). Since the \( q_i \)'s are polynomials, \( |A_k(m, N)| \leq C_2 \) for some constant \( C_2 \) which does not depend on \( m \) and \( k \). It follows from Lemma \( 6.2.1 \) that

\[
|J_2| \leq C_2 N^{-1} \sum_{m=1}^{\delta N} \sum_{k=0}^{\infty} V_N(k) \leq C_2 Y\delta
\]
where $\Upsilon$ was defined in (6.12). Since the cardinality of the set $\Gamma_{N,\delta}$ does not exceed $\delta$, similar arguments show that for any sufficiently large $N$,

$$|J_3| \leq A \delta$$

for some constant $A$.

Next, suppose that $q_{\varepsilon(i)}$ and $q_{\tau(j)}$ are linearly related and that their differences do not explode. Then, by Corollary 3.3.1 they are equivalent, namely there exists a constant $C_0$ so that

$$q_{\varepsilon(i)}(n, N) = q_{\tau(j)}(cn + r, N)$$

for any natural $n$. If $m \neq cn + r$ then by Corollary 6.3.1 we have

$$|q_N(m) - p_N(n)| \geq Q(cn/N)N^{k-1}C_0$$

for some continuous function $Q$ which is strictly positive on $(0,1)$, and so by Lemma 6.2.1

$$|b_{\varepsilon,\tau,i,j,N}(n, m)| \leq C_{UV}(Q(cn/N)N^{k-1}C_0).$$

Therefore, taking into account (6.11), in order to prove that the limit $D_{\varepsilon,\tau,i,j}(t, s)$ exists, it is enough to show that the limit

$$\lim_{N \to \infty} N^{-1} \sum_{n \in I_r(N,t)} b_{\varepsilon,\tau,i,j,N}(n, cn + r)\mathbb{I}(cn + r \in I_r(N) \cap N)$$

(6.17)

exists, where $x \to I(x \in A)$ is the indicator function of a set $A$. Since the sets $I_r(N)$ and $I_r(N)$ are unions of intervals whose length is proportional to $N$ (i.e. they have the form (6.3)), and the set

$$\{n \in \mathbb{N} : cn + r \in \mathbb{N}\}$$

is a finite union of arithmetic progressions, in order to show that the above limit exists it is enough to show that there exist sets $\Delta_{N,\delta}$ whose cardinalities do not exceed $\delta N$ so that for any $\delta \in (0,1)$ the limit

$$\lim_{N \to \infty, n \in I_r(N) \setminus \Delta_{N,\delta}, cn + r \in I_r(N)} b_{\varepsilon,\tau,i,j,N}(n, cn + r)$$

(6.18)

exists, and that this limit does not depend on $\delta$. Fix some $\delta \in (0,1)$. Suppose that $q_{\varepsilon(u)}(x, N) = a_{\varepsilon(u)}x + b_{\varepsilon(u)}N$ and $q_{\tau(v)}(x, N) = a_{\tau(v)}x + b_{\tau(v)}N$ are linear polynomials where $u < i$ and $v < j$. Then

$$q_{\varepsilon(u)}(n, N) - q_{\tau(v)}(cn + r, N) = a_{\varepsilon(u)}r + (a_{\varepsilon(u)}c - a_{\tau(v)})n + (b_{\varepsilon(u)} - b_{\tau(v)})N.$$

If $ca_{\varepsilon(u)} \neq a_{\tau(v)}$ but $b_{\varepsilon(u)} = b_{\tau(v)}$ then for any $n \in [\delta N, N]$ we have

$$|q_{\varepsilon(u)}(n, N) - q_{\tau(v)}(cn + r, N)| \geq C_\delta N$$

for some constant $C_\delta > 0$. When $ca_{\varepsilon(u)} \neq a_{\tau(v)}$ and $b_{\varepsilon(u)} \neq b_{\tau(v)}$, there exists a constant $w = w_{u,v,N} \in \mathbb{R}$ so that for any sufficiently large $N$ and $n \in [\delta N, N] \setminus [N(w - \delta), N(w + \delta)] := J(w, \delta, N)$,

$$|q_{\varepsilon(u)}(n, N) - q_{\tau(v)}(cn + r, N)| \geq C'_\delta N$$

where $C'_\delta > 0$ is some constant. Let $(u_j, v_j)$, $j = 1, ..., L_1$ be the indexes so that $q_{\varepsilon(u_j)}$ and $q_{\tau(v_j)}$ are linear, $ca_{\varepsilon(u_j)} \neq a_{\tau(v_j)}$ and $b_{\varepsilon(u_j)} \neq b_{\tau(v_j)}$. Let $q_{\varepsilon(1)}, ..., q_{\varepsilon(L)}$
and \(q_{\varepsilon(1)}, \ldots, q_{\varepsilon(V)}\) be the linear polynomials among the \(q_{\varepsilon(u)}\)'s and the \(q_{\varepsilon(v)}\)'s, respectively, where \(1 \leq u < i \) and \(1 \leq v < j\). We conclude that for any

\[
n \in \bigcup_{j=1}^{L_1} J(w_{u_j,v_j,N},\delta,N) := \Theta_{N,\delta}
\]

we can order the numbers \(q_{\varepsilon(u)}(n,N)\) and \(q_{\varepsilon(v)}(cn+r,N)\), where \(1 \leq u \leq U\) and \(1 \leq v \leq V\), so that the differences between each consecutive numbers in this ordering is either a constant which does not depend on \(n\) and \(N\), or it is not less than \(B_{\delta} N^2\) for some \(B_{\delta} > 0\), where we also used (6.5). Combining this with Assumption 2.2.2, we deduce from Corollary 3.3.1 that the terms

\[
q_{\varepsilon(1)}(n,N) - q_{\varepsilon(z_j)}(cn+r,N)
\]

where \(1 \leq l \leq i\) and \(1 \leq z \leq j\), either have absolute values bounded from below by some \(E_\delta \), \(E_\delta > 0\) or they are constants. Set \(\Delta_{N,\delta} = [\delta N, N] \setminus \Theta_{N,\delta}\). Using (6.2) and (6.3), applying Proposition 4.0.1 we derive that there exist constants \(d_1, \ldots, d_L\) \((L \geq 0)\) and indexes \((l_e,z_e)\), \(e = 1, 2, \ldots, L\) so that

\[
\lim_{N \to \infty, n \in I_e(N) \setminus \Delta_{N,\delta}, cn+r \in I_e(N)} \frac{b_{\varepsilon,\tau,i,j,N}(n, cn+r)}{d_1, \ldots, d_L} = \int F_{\varepsilon,i}(\bar{x}) F_{\varepsilon,j}(\bar{y}) d\mathcal{M}
\]

where \(\bar{x} = (x_{z_1}, \ldots, x_{z_j})\), \(\bar{y} = (y_{r(1)}, \ldots, y_{r(j)})\) and \(\mathcal{M}\) has the form

\[
dl{\mathcal{M}(\bar{x}, \bar{y}) = \prod_{l \notin \{l_e\}} dl_{\varepsilon(x_{l])] \prod_{z \notin \{z_e\}} dl_{\varepsilon(y_{z])} \prod_{e=1}^{L} d\mu_{\delta_e}(x_{l_e}, y_{r(z_e)}).}
\]

(6.19)

The indexes \((l_e, z_e)\) are exactly the ones for which

\[
q_{\varepsilon(l_e)}(n,N) - q_{\varepsilon(z_e)}(cn+r,N) = d_e = q_{\varepsilon(l_e)}(r,0) - q_{\varepsilon(z_e)}(0,0)
\]

where \(d_e\) does not depend on \(n\) and \(N\). \(\square\)

6.3. Existence under abstract number theory type conditions. Let \(\varepsilon, \tau \in \mathcal{E}'\) and \(1 \leq i, j \leq \ell\) be so that \(\deg q_{\varepsilon(i)} = \deg q_{\varepsilon(j)} = k > 1\). Then

\[
\mathbb{E}[S_{\varepsilon,i,[N]} S_{\tau,j,[N]}] = \frac{1}{N} \sum_{n \in I_e(N,t)} \sum_{v \in Z} \sum_{m \in J_e(N,s,v,i,e)} b_{\varepsilon,\tau,i,j,N}(n,m)
\]

where \(J_e(N,s,v,u,\varepsilon)\) is the set of all \(m\)'s in \(I_e(N,s)\) so that \(q_{\varepsilon(j)}(m,N) - q_{\varepsilon(i)}(n,N) = v\). Taking into account Lemma 6.2.1 if the limit

\[
\Gamma_v = \lim_{N \to \infty} \frac{1}{N} \sum_{n \in I_e(N,t)} \sum_{m \in J_e(N,s,v,i,e)} b_{\varepsilon,\tau,i,j,N}(n,m)
\]

exists for each \(v\), then

\[
D_{\varepsilon,\tau,i,j}(t,s) = \lim_{N \to \infty} \mathbb{E}[S_{\varepsilon,i,[N]} S_{\tau,j,[N]}] = \sum_{v \in Z} \Gamma_v.
\]

Using Lemma 3.2.2 with \(q = q_{\varepsilon(j)}\) and \(p = q_{\varepsilon(i)}\), the equality \(q_{\varepsilon(j)}(m,N) - q_{\varepsilon(i)}(n,N) = v\) means that, with \(y = n/N\),

\[
H_{N,y}(m - N\gamma_k(y) - R_k(y)) = \frac{v + P_k + Q(R_k(y),0)}{N^{k-1}Q_k'(\gamma_k(y))}.
\]
Fix some $\delta \in (0,1)$. When $|u| \leq \sqrt{N}$ and $y \geq \delta$ then, since $H_{N,y}$ is one to one on intervals of the form $[-a,a]$, $a > 0$ (when $N$ is large enough), we obtain that for any sufficiently large $N$ there exists a unique solution $x = x_{n,N,v}$ to the equation

$$H_{N,y}(x) = v + P_0 + Q(R_k(y), 0) \frac{N^{k-1}Q'_k(\gamma_k(y))}{N^{k-1}Q'_k(\gamma_k(y))}.$$ 

Using Lemma 6.2.1 the following proposition follows similarly to the proof of Proposition 6.2.2.

6.3.1. Proposition. The limit $D_{\varepsilon,\tau,i,j}(t, s)$ exists if for any $v$ and sufficiently large $n$ and $m$ so that $q_{\varepsilon(j)}(m, N) - q_{\varepsilon(i)}(n, N) = v$ the differences $q_{\varepsilon(j)}(m, N) - q_{\varepsilon(i)}(n, N)$, $i' \leq i$, $j' \leq j$

are either constants or they converge to $\infty$ (in absolute value) as $n \to \infty$ and the limit

$$\lim_{N \to \infty} \frac{1}{N} |\{n \in I_{\varepsilon}(N, t) : x_{n,N,v} + \gamma_k(n/N) + R_k(n/N) \in N\}|$$

exists, where $|\Gamma|$ stands for the cardinality of a finite set $\Gamma$.

Note that when the above conditions hold true only with $s = t = 1$ then we obtain that the limit $D^2$ exists, which is enough in order to derive that $S_N(1)$ converges in distribution towards a centered normal random variable whose variance is $D^2$.

6.4. Positivity of the asymptotic variance. We assume here that the conditions of Theorem 2.3.1 are satisfied. Set $S_N = S_N(1)$ and

$$D^2 = \lim_{N \to \infty} E S_N^2 = \lim_{N \to \infty} \text{Var}(S_N).$$

We will say that the polynomials $q_1(x, N)$ and $q_j(x, N)$ are equivalent if there exist rational $c$ and $r$ so that $q_i(n, N) - q_j(cn+r, N)$ is a constant function of $n$ and $N$ (in Remark 2.3.2 we called this equivalence relation $Q$-equivalence). Then any two equivalent polynomials have the same degree. Let $q_{s+1}, \ldots, q_w$ be the non-linear polynomials among the $q_i$’s and consider the decomposition of $\{q_i : i > s\}$ into equivalence classes $A_1, A_2, \ldots, A_w$, ordered so that the degree of each member of $A_i$ does not exceed the degree of each member of $A_{i+1}$, $i = 1, \ldots, w-1$. For any $\varepsilon, \tau, i, j$ so that $q_{\varepsilon(i)}$ and $q_{\tau(j)}$ are not both linear and not equivalent we have

$$D_{\varepsilon,\tau,i,j} := D_{\varepsilon,\tau,i,j}(1,1) = 0.$$ 

Therefore, $D^2 = \sum_{t=0}^w D_t^2$ where for $t = 1, 2, \ldots, w,$

$$D_t^2 = \sum_{\varepsilon, \tau, i,j} \sum_{q_{\varepsilon(i)}, q_{\tau(j)} \in A_t} D_{\varepsilon,\tau,i,j}.$$

and $D_0^2 := D^2 - \sum_{t=1}^w D_t^2$ (we will see soon that it also nonnegative). Hence, $D^2 = 0$ if and only if $D_t^2 = 0$ for any $t$.

When $q_1$ is linear and $q_1, \ldots, q_s$ are the linear polynomials among the $q_i$’s, we define the function $G$ by

$$G(x_1, \ldots, x_s) = \int F(x_1, \ldots, x_s, y_{s+1}, \ldots, y_t) d\mu(y_{s+1}) \cdots d\mu(y_t).$$
Then, taking into account the tensor product structure of the set of permutations \( \mathcal{E}' \) from (6.9) we have
\[
\sum_{n=1}^{N} G(\xi_{q_1(n,N)}, \ldots, \xi_{q_s(n,N)}) = \sum_{e \in \mathcal{E}'} \sum_{n \in \mathcal{E}(N)} \sum_{i=1}^{k} F_{e,i}(\xi_{q_{e(i)}(n,N)}, \ldots, \xi_{q_{e(i)}(n,N)})
\]
\[
+ \sum_{n \in B_N} G(\xi_{q_1(n,N)}, \ldots, \xi_{q_s(n,N)}) := G_{N,1} + G_{N,2}
\]
where \( B_N \) is the set from Proposition 6.1.1. By Lemma 5.1.2 we have
\[
\lim_{N \to \infty} N^{-1} \mathbb{E} Q^2_N = 0 \quad \text{since the cardinality of } B_N \text{ grows sublinearly in } N
\]
hence
\[
D_0^2 = \lim_{N \to \infty} N^{-1} \mathbb{E} Q^2_N
\]

where
\[
G_N = \sum_{n=1}^{N} \sum_{i=1}^{s} \sum_{b=1}^{b_i} G(\xi_{q_i(n,N)}, \ldots, \xi_{q_i(n,N)})
\]

Recall next the following result from Chapter 3 in [10], which was stated there as Theorem 3.3.4, and is reformulated here for the function \( G \):

6.4.1. Theorem. Let \( \{ \xi_{n}^{(i)} : n \geq 0 \}, i = 1, 2, \ldots, s \) be independent copies of the random sequence \( \xi = \{ \xi_n, m \geq 0 \} \). Suppose that the conditions of Theorem 6.4.1 hold true and \( \sigma^2 = \lim_{N \to \infty} N^{-1} \mathbb{E} G_N^2 \) exists. Let \( a_1, b_i \) be the numbers from (6.7). Set
\[
U_{n,N} = G(\xi_{a_1,n+b_1,N}, \xi_{a_2,n+b_2,N}, \ldots, \xi_{a_s,n+b_s,N})
\]
\[
\Sigma_N = \sum_{1 \leq n \leq N} U_{n,N}
\]
and
\[
v^2 = \lim_{N \to \infty} N^{-1} \mathbb{E} (\Sigma_N)^2.
\]

Then, the limit \( v^2 \) exists, and \( \sigma^2 > 0 \) if and only if \( v^2 > 0 \). The latter condition holds true if and only if there exists no representation of the form
\[
U_n = Z_{n+1} - Z_n, \quad n = 0, 1, 2, \ldots
\]

where \( U_n = G(\xi_{a_1,n}, \ldots, \xi_{a_s,n}) \), \( \{ Z_n \}_{n=0}^{\infty} \) is a square integrable weakly (i.e. in the wide sense) stationary process. Furthermore, \( v^2 = 0 \) if and only if \( \text{Var} \Sigma_N \) is bounded.

For readers’ convenience, let us give some of the ideas behind the proof of Theorem 6.4.1. Let us first explain (6.20). Since the above copies are independent and the distribution of \( (\xi_n, \xi_m) \) depends only on \( n - m \) we have
\[
\mathbb{E} \Sigma_N^2 = \mathbb{E} \left( \sum_{n=1}^{N} U_n \right)^2.
\]

Observe next that the sequence \( \{ G(\xi_{a_1,n}, \ldots, \xi_{a_s,n}) \} \) is centered and stationary in the wide sense. We recall now the following result, which is a combination of Proposition 8.3 and Theorem 8.6 of [5] (modified for a one sided sequence): let \( \{ U_n \} \) be a centered sequence of weakly stationary random variables with correlations \( b_n = \text{Cor}(U_n, U_0) \) satisfying \( \sum_{n=0}^{\infty} (n+1)|b_n| < \infty \). Then the limit
\[
v^2 = \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left( \sum_{j=0}^{n-1} U_n \right)^2
\]
exists, and it is positive if and only if there exists
a weakly stationary process $Z_n$ so that $U_n = Z_{n+1} - Z_n$ for any $n \geq 0$. We conclude that $\sigma^2 = 0$ if and only there exists a stationary in the wide sense sequence \( \{Z_n\} \) of random variables so that for any $n$,

\[
G(\xi_{a_1n}, \ldots, \xi_{a_sn}) = Z_{n+1} - Z_n.
\]  

(6.21)

We note that the correlations of \( \{U_n\} \) indeed satisfy the above conditions because of the definition of $G$, \([\text{2.5.4.}]\) and Assumption \([\text{2.5.3.}]\). We also remark that $U_n$ is indeed centered because of our assumption that $\bar{F} = 0$.

Next, for readers’ convenience we will give the idea behind the equivalence $\sigma^2 = 0 \iff \sigma^2 = 0$ in the case when $s = 2$, namely when there are only two linear polynomials $q_1(n, N) = a_1n + b_1N$ and $q_2(n, N) = a_2n + b_2N$. In the case when $b_1 = b_2$ we have

\[
\mathbb{E}[G(\xi_{q_1(n, N)}, \xi_{q_1(n, N)})G(\xi_{q_1(m, N)}, \xi_{q_2(m, N)})] = \mathbb{E}[G(\xi_{a_1n}, \xi_{a_2n})G(\xi_{a_1m}, \xi_{a_2m})]
\]

and therefore the whole problem is reduced to the case when $q_1$ and $q_2$ do not depend on $N$, and this case was covered in Section 4 of \([\text{7.}]\). We assume without a loss of generality that $b_2 > b_1$. The proof in the case when $a_1 \leq a_2$ proceeds exactly as in \([\text{7.}]\). Indeed, we can use the same block partitions which were constructed in \([\text{7.}]\) in the case when $q_1(n) = (a_1 + b_1)n$ and $q_2(n) = (a_2 + b_2)n$. The distance between $(a_2 + b_2)n$ and $(a_1 + b_1)n$ is smaller than the distance between $a_2n + b_2N$ and $a_1n + b_1N$, and so the proof in \([\text{7.}]\) proceeds similarly. We note that only the case when $q_1(n) = in$ was considered in \([\text{7.}]\), but the general case when all $q_i$’s have the form $q_i(n) = l_in$, $i = 1, 2, \ldots, s$, $l_i \in \mathbb{N}$ follows from this case by considering functions $F(x_1, \ldots, x_s)$ which depend only on the variables $x_{i_1}$.

In view of the above, we will focus here on the case when $a_1 > a_2$. We will show that, in fact, in this case we have $\sigma^2 = \sigma^2$. Consider the functions $G_1$ and $G_2$ given by

\[
G_1(x) = \int G(x, y)d\mu(y), \text{ and } G_2(x, y) - G_1(x).
\]

Then

\[
\int G_2(x, y)d\mu(y) = 0 \text{ for any } x.
\]  

(6.22)

Set $\alpha = \frac{b_2 - b_1}{a_2 - a_1}$. Then $\alpha$ satisfies that $q_1(n, N) \leq q_2(n, N)$ for any $1 \leq n \leq \alpha N$, while $q_2(n, N) \leq q_1(n, N)$ for any $N \geq n \geq \alpha N$. Using the above notations we have

\[
\sum_{n=1}^{N} G(\xi_{q_1(n, N)}, \xi_{q_1(n, N)}) = \sum_{1 \leq i, j \leq 2} S_i^{(2)}(N)
\]

where

\[
S_1^{(1)}(N) = \sum_{n=1}^{\alpha N} G_1(\xi_{a_1n+b_1N}), \quad S_2^{(1)}(N) = \sum_{n=1}^{\alpha N} G_2(\xi_{a_1n+b_1N}, \xi_{a_2n+b_2N})
\]

\[
S_1^{(2)}(N) = \sum_{\alpha N < n \leq N} G_1(\xi_{a_2n+b_2N}), \quad S_2^{(2)}(N) = \sum_{\alpha N < n \leq N} G_2(\xi_{a_2n+b_2N}, \xi_{a_1n+b_1N}).
\]

Using \([\text{6.22.}]\) we get that for $i = 1, 2$

\[
\|\mathbb{E}[S_i^{(1)}(N)S_i^{(2)}(N)]\| \leq C
\]  

(6.23)
for some constant $C > 0$. In the case when $\xi_n$’s were independent for any $n < \alpha N$ and $m > \alpha N$ we have

$$
\mathbb{E}[G_1(\xi_{a_1n+b_1N})G_2(\xi_{a_2m+b_2N}, \xi_{a_1m+b_1N})]
$$

(6.24)

and

$$
\mathbb{E}[G_1(\xi_{a_1n+b_1N})\mathbb{E}[G_2(\xi_{a_2m+b_2N}, \xi_{a_1m+b_1N})|\xi_{a_1n+b_1N}]] = 0
$$

(6.25)
since the conditional expectation inside the expectation vanishes due to (6.22). This shows that we can take $C = 0$ in the independent case when $i = 1$, and similar idea yields the same with $i = 2$. When $\xi_n$’s are not independent we can use our mixing and approximation conditions from Assumption 2.1.3 in order to derive (6.23) (applying Lemma 6.2.4). Similar arguments yield that $|\mathbb{E}[S_1^{(1)}(N)S_2^{(1)}(N)]|$ and $|\mathbb{E}[S_2^{(1)}(N)S_2^{(2)}(N)]|$ are bounded in $N$. We conclude that

$$
\mathbb{E}S_2^{(2)} = c_N + \sum_{1 \leq j \leq 2} \mathbb{E}(S_1^{(j)})^2
$$

where $c_N$ is some bounded sequence. The above equality also holds true when we replace $\xi_{a_1n+b_1N}$ with $\tilde{\xi}_{a_1n+b_1N}$, where $\tilde{\xi}$ is an independent copy of $\xi$. It is enough to show that $E(S_2^{(2)})^2$ behaves the same as in the case when the left and right coordinates are independent. We will explain why this is true for $j = 1$. Write

$$
S_2^{(1)}(N) = A_N + B_N
$$

where $A_N = \sum_{n=1}^{M_N} G_2(\xi_{a_1n+b_1N}, \xi_{a_2n+b_2N})$ and $B_N = S_2^{(1)}(N) - A_N$. Here $M_N$ satisfies that $a_1 M_N + b_1 N = (b_2 - \varepsilon)N$, where $\varepsilon > 0$ is sufficiently small (we can just take $M = \beta N$ for an appropriate $\beta < \alpha$). This means that the difference between the largest index appearing in the left coordinate of $G_2$ in the sum $A_N$ is much smaller than the minimal index appearing in the right coordinate. Therefore,

$$
E A_N^2 = c_N + E(\bar{A}_N)^2
$$

where $c_N$ is some bounded sequence and $\bar{A}_N$ is defined similarly to $A_N$ but with $\tilde{\xi}_{a_1n+b_1N}$ in place of $\xi_{a_1n+b_1N}$. Using the same reasoning which lead to (6.23), we obtain that $|E A_N b_N|$ is bounded in $N$. Therefore,

$$
E(S_2^{(1)}(N))^2 = E(\bar{A}_N)^2 + E B_N^2.
$$

Now we decompose $B_N$ into two blocks. On the first block the maximal index in the left coordinate will is smaller than the minimal index in the right coordinate, and the expectation of the product of the two blocks is bounded in $N$. Proceeding this way we derive that $\sigma^2 = \eta^2$. □

The next result we have is the following

6.4.2. Theorem. Suppose that any two non-linear polynomials $q_i$ and $q_j$ are not equivalent. Then $D^2 = 0$ if and only if (6.21) holds true and for any $\varepsilon$ and $i$ so that deg $q_{\varepsilon(i)} > 0$ we have $F_{\varepsilon,i}^2(x_1, ..., x_i) = 0$ for $\mu^i = \mu \times \cdots \mu$ almost any $(x_1, ..., x_i)$.

Proof. Let $\deg q_{\varepsilon(i)}$ and $\deg q_{\varepsilon(j)}$ be two non-linear polynomials. If they not equivalent then by Proposition 6.2.2 we have

$$
D_{\varepsilon,\tau,i,j}(s,t) = 0
$$

for any $s$ and $t$. Suppose that they are equivalent. Then, using the assumptions of Theorem 6.4.2 we must have $q_{\varepsilon(i)} = q_{\varepsilon(j)}$. We claim that in this case $\varepsilon = \tau$. Indeed, since $q_{\varepsilon(i)} = q_{\varepsilon(j)}$ then we have $c = 1$ in (6.18). The sets $I_\varepsilon(N)$ and $I_\tau(N)$ are disjoint when $\varepsilon \neq \tau$, and in this case the sum in (6.17) is over a set of $n$’s whose
asymptotic density is 0. Therefore, either $D_{ε,τ,i,j}(t,s) = 0$ for any $t,s$ or $ε = τ$, $i = j$ and

$$D_{ε,τ,i,j}(1,1) = D_{ε,ε,ι,i}(1,1) = s_ε \int F_{ε,i}^2 dμ^i.$$  

We conclude that

$$D^2 = \sum_{ε,i: \deg q_i(ε) > 1} s_ε \int F_{ε,i}^2 dμ^i + D_0^2$$

and the proof of the proposition is completed using Theorem 6.4.1 □

6.4.3. Remark. The proof of Theorem 6.4.2 shows that if there exists one non-linear polynomial $q_{i_0}$ which is not equivalent to any other of the $q_j$’s (i.e. there exists a non-linear equivalence class which is a singleton), then

$$D^2 = D_0^2 + \sum_ε s_ε \int F_{ε,ε^{-1}(i_0)}^2(x_1, ..., x_{ε^{-1}(i_0)})dμ(x_1) \cdots dμ(x_{ε^{-1}(i_0)}) + δ^2$$

where $δ^2 > 0$ is the part of the variance which comes from the non-linear polynomials $q_{i_0}$ and $q_{τ(j)}$ which do not equal $q_{i_0}$. Therefore, we derive that $D^2 > 0$ if one of the functions $F_{ε,ε^{-1}(i_0)}$ is not identically 0 ($μ^{ε^{-1}(i_0)}$-almost surely). Moreover, when there exists one class of equivalent non-linear polynomials for which $c = 1$ in (6.18) for any $q_ε(i), q_τ(j)$ in this class, then the proof of Theorem 6.4.2 shows that $D^2 > 0$ if for any $ε$ the functions $\{u_ε^{(ε)}\} → s_ε \sum_{i,q_ε(i)ε} F_{ε,i}(u_ε^{(ε)})$ vanish with respect to an appropriate probability measure $κ_A,ε$, where $A$ is the underlying class and $\{u_ε^{(ε)}\}$ is a set of variables which depends on $ε$. Indeed, when $q_ε(i)$ and $q_τ(j)$ are equivalent and non-linear and $c = 1$ in (6.18) we get that

$$D_{ε,τ,i,j} = s_ε \int F_{ε,i}F_{ε,j}dM_{ε,i,j,A}$$

for some probability measure $M_{ε,i,j,A}$ (which was constructed in the proof of Proposition 6.4.2). It remains to show that the measures $M_{ε,i,j}$ are an appropriate marginal of a measure $M_{ε,A}$ which does not depend on $i$ and $j$ (this is done as in Lemma 7.1 in [8], where we recall that $ε$ is fixed). We conclude that if any two equivalent non-linear polynomials $q_i$ and $q_j$ satisfy $q_i(x, y) = q_j(x + r, y) + z$ for some $r, z ∈ \mathbb{Z}$, then $D^2 = 0$ if and only if (6.21) holds true and the functions $\sum_{i,q_ε(i)ε} F_{ε,i}(u_ε^{(ε)})$ vanish with respect to $M_{ε,A}$ for any $ε$ and a non-linear class $A$.

6.4.4. Remark. When the polynomials $q_i$ are ordered so that $q_1(n, N) ≤ q_2(n, N) ≤ ... ≤ q_ε(n, N)$ for any sufficiently large $n$ and $N$, then Theorem 2.3 in [8] is proved exactly in the same way. This means that there exists a family of functions $G_α$ and a family of probability measures $κ_α$ so that $D^2 = 0$ if and only if (6.21) holds true and $G_α^2 = 0, κ_α$ almost surely, for any $α$. The measures and the functions can be constructed explicitly, see Section 7 in [8].

7. A FUNCTIONAL CLT VIA STEIN’S METHOD

Let $(Ω, F, P), F_{n,m}, \{ξ_α : n ≥ 0\}$ and $F$ be as described in Section 2 and Consider the random function $S_N$ given by

$$S_N(t) = N^{-\frac{1}{2}} \sum_{n=1}^{[Nk]} F(ξ_n, N)$$

where $ξ_n, N = (ξ_{q_1(n, N)}, ξ_{q_2(n, N)}, ..., ξ_{q_ε(n, N)})$.
and assume that the limiting covariances \( b(t, s) \) from Theorem \( 2.3.1 \) exist. For each \( n, m \) and \( N \) set
\[
d_N(n, m) = \min_{1 \leq i,j \leq \ell} |q_i(n, N) - q_j(m, N)|.
\]

Set \( l(N) = 3N^{\zeta_1} + 3 \) where \( \zeta_1 = \frac{3w}{4w - 2} \) and the numbers \( w \) and \( \theta \) come from Assumptions \( 2.1.1 \) and \( 2.1.3 \). Recall our assumption in Theorem \( 2.3.3 \) that \( \theta > \frac{4w}{w - 2} \), which implies that
\[
(\theta(1 - \frac{2}{w}))^{-1} < \zeta_1 < \frac{1}{4}.
\]

The proceeding arguments will be true for any \( \zeta_1 \) satisfying \( (7.1) \). Define \( r(N) = [l(N)/3] \) and for each \( m \) and \( r \) set
\[
\xi_{m,r} = E[\xi_m F_{m-r,m+r}] \quad \text{and} \quad \Xi_{m,N,r} = (\xi_{q_1(n,N),r}, \xi_{q_2(n,N),r}, ..., \xi_{q_l(n,N),r}).
\]

We also set
\[
Z_N(t) = N^{-\frac{1}{2}} \sum_{n=1}^{[NT]} F(\Xi_{n,N,r(N)}) \quad \text{and} \quad W_N(t) = Z_N(t) - EZ_N(t).
\]

Using \( (2.5) \) and Assumption \( 2.1.3 \) we can approximate \( F(\Xi_{n,N}) \) by \( F(\Xi_{n,N,r(N)}) \). Taking into account Lemmas \( 5.1.2 \) and \( 5.1.1 \) this yields that
\[
\lim_{N \to \infty} E[W_N(t)W_N(s)] = \lim_{N \to \infty} E[S_N(t)S_N(s)] = b(t, s)
\]
and that the weak convergence of \( S_N(\cdot) \) follows from the weak convergence of \( W_N(\cdot) \). We refer the readers to the beginning of the proof of Theorem 1.6.2 from [10] for the precise details.

In the rest of this section we will prove the weak invariance principle for \( W_N(\cdot) \) using a functional version of Stein’s method. For each \( N \) we define a graph \( (V_N, E_N) \) on \( V_N = \{1, ..., N\} \) by declaring that \( (n, m) \in E_N \) if
\[
d_N(n, m) = \min_{1 \leq i,j \leq \ell} |q_i(n, N) - q_j(m, N)| \leq l(N) = 3N^{\zeta_1} + 3.
\]

Then the size of a ball of radius 1 in this graph does not exceed \( K_1 l(N) \) for some constant \( K_1 \) which does not depend on \( N \). Indeed let \( n \in V_N \) and let \( m \) be a neighbor of \( n \). Then \( d_N(n, m) \leq 3N^{\zeta_1} + 3 \) and therefore there exists \( 1 \leq i, j \leq \ell \) so that
\[
q_i(n, N) - q_j(m, N) \in \{-l(N), -l(N) + 1, ..., l(N) - 1, l(N)\} =: \Gamma_N.
\]

Hence \( m \) solves one of the equations
\[
q_i(n, N) - q_j(m, N) = g, \quad g \in \Gamma_N.
\]

Each equation has at most \( d_j = \deg q_j \) solutions, \( m_{g,1}, ..., m_{g,u_g} \), where \( u_g \leq d_j \). Therefore, the unit ball around \( n \) is contained in the set
\[
\{n\} \cup \bigcup_{g \in \gamma} \{m_{g,1}, ..., m_{g,u_g}\}
\]
whose cardinality does not exceed \( 2d^{*}(l(N) + 1) + 1 \), where \( d^{*} \) is the maximal degree of the polynomials \( q_1, ..., q_\ell \).

Now, for each \( N \) and \( n \in V_N \) set
\[
X_{n,N} = \frac{F(\Xi_{n,N,r}) - E(\Xi_{n,N,r(N)})}{\sqrt{N}}.
\]
Then $\mathcal{W}_N(t) = \sum_{n=1}^{[Nt]} X_{n,N}$. We also set
\[ b_N(t,s) = \mathbb{E}[\mathcal{W}_N(t)\mathcal{W}_N(s)] = \text{Cov}(\mathcal{W}_N(t),\mathcal{W}_N(s)). \]

Next, for any two sub-$\sigma$-algebras $\mathcal{G}, \mathcal{H}$ of $\mathcal{F}$ we will measure the dependence between $\mathcal{G}$ and $\mathcal{H}$ via the quantities
\[ \varepsilon_{p,q}(\mathcal{G}, \mathcal{H}) = \sup \left\{ |\text{Cov}(g,h)| : \max(\|g\|_{L^p}, \|h\|_{L^q}) \leq 1 \right\} \]
where in the definition of $\varepsilon_{p,q}(\mathcal{G}, \mathcal{H})$ we consider function $g$ and $h$ so that $g$ is measurable with respect to $\mathcal{G}$ and $h$ is measurable with respect to $\mathcal{H}$. We recall the following classical relations between $\alpha(\mathcal{G}, \mathcal{H})$ (defined in 4.3) and $\varepsilon_{p,q}(\mathcal{G}, \mathcal{H})$ (see, for instance, Theorem A.5, Corollary A.1 and Corollary A.2 in [6]): for any $p > 1$, we have
\[ \varepsilon_{p,\infty}(\mathcal{G}, \mathcal{H}) \leq 6\left(\alpha(\mathcal{G}, \mathcal{H})\right)^{1-\frac{1}{p}} \text{ and } \varepsilon_{p,q}(\mathcal{G}, \mathcal{H}) \leq 8\left(\alpha(\mathcal{G}, \mathcal{H})\right)^{1-\frac{1}{p} - \frac{1}{q}} \]
for any $q > 1$ such that $\frac{1}{p} + \frac{1}{q} < 1$.

Now, for any $A \subset V_N$, let $\mathcal{G}_A$ be the $\sigma$-algebra generated by the random vector $X_A = \{X_{n,N} : n \in A\}$. For any $A, B \subset V_N$ and $1 \leq p, q \leq \infty$ we will measure the dependence between $X_A$ and $X_B$ via the quantities
\[ \alpha(A, B) := \alpha(\mathcal{G}_A, \mathcal{G}_B) \text{ and } \varepsilon_{p,q}(A, B) := \varepsilon_{p,q}(\mathcal{G}_A, \mathcal{G}_B). \]

Then by (7.3) for any $A, B \subset V_N$ and $p > 1$,
\[ \varepsilon_{p,\infty}(A, B) \leq 6\left(\alpha(\mathcal{G}_A, \mathcal{G}_B)\right)^{1-\frac{1}{p}} \text{ and } \varepsilon_{p,q}(A, B) \leq 8\left(\alpha(\mathcal{G}_A, \mathcal{G}_B)\right)^{1-\frac{1}{p} - \frac{1}{q}} \]
for any $q > 1$ such that $\frac{1}{p} + \frac{1}{q} < 1$.

Now, for reader's convenience, we will restate Theorem 1.5.1 from [10]:

7.0.1. Theorem. Let $p_0, q_0 \geq 1$ and set
\[ \tau_N = d_1 + d_2 + d_3 + d_4 \]
where with $X_n = X_{n,N}$, $\sigma_{n,m} = \mathbb{E}X_n X_m$, $N_n = \{n\} \cup \{m : (n, m) \in E_N\}$ and $N_n^c = V_N \setminus N_n$,
\[ d_1 = \sum_{n=1}^{N} \|X_n\|_{L^{p_0} \varepsilon_{p_0,\infty}}(\{n\}, N_n^c), \]
\[ d_2 = \sum_{n=1}^{N} \sum_{m \in N_n} \|X_n X_m - \sigma_{n,m}\|_{L^{q_0} \varepsilon_{q_0,\infty}}(\{n, m\}, N_n^c \cap N_m^c), \]
\[ d_3 = \sum_{n=1}^{N} \sum_{m,k \in N_n} \left(\mathbb{E}|X_n X_m X_k| + \mathbb{E}|X_n X_m |X_k|\right), \]
\[ d_4 = \sum_{n=1}^{N} \sum_{m \in N_n^c} \|X_n\|_{L^{p_0}} \|X_m\|_{L^{q_0} \varepsilon_{p_0,q_0}}(\{n\}, \{m\}). \]

Suppose that there exists $\Gamma > 0$ such that
\[ \|\mathcal{W}_N(s) - \mathcal{W}_N(t)\|_{L^2}^2 \leq \Gamma \frac{[N_s] - [Nt]}{N} \]
for any $N \in \mathbb{N}$ and $0 \leq t \leq s \leq 1$. Furthermore assume that the limits
\[ \lim_{N \to \infty} b_N(t, s) = b(t, s), s, t \in [0, 1] \]
exist and that
\[ \lim_{N \to \infty} \tau_N \ln^2 N = 0. \]
Then \( W_N \) weakly converge in the Skorokhod space \( D \) to a continuous centered Gaussian process \( G \) with covariance function \( b(\cdot, \cdot) \).

Theorem 7.0.1 is essentially due to A.D. Barbour, and it follows from the arguments in [1] and [2] (see the proof of Theorem 1.5.1 in [10]).

In the rest of this section we will show that all the conditions of Theorem 7.0.1 are satisfied with

\[ p_0 = 2q_0 = w \]

where \( w \) comes from Assumption 2.1.1. Then there exists a constant \( C > 0 \) so that for any \( n \) we have

\[ \|X_n\|_w \leq CN^{-\frac{1}{2}}. \] (7.11)

That the covariances converge we have already shown. We claim next that there exists \( C > 0 \) such that

\[ \text{Var}(W_N(s) - W_N(t)) \leq C^2 \frac{[Ns] - [Nt]}{N} \] (7.12)

for any \( N \in \mathbb{N} \) and \( 0 \leq t \leq s \leq 1 \). The arguments proceeding (7.2) yield that it is enough to prove that (7.12) holds true with \( S_N \) in place of \( W_N \). As in (7.2), we refer the readers to the proof of Theorem 1.6.2 in [10] for the exact technical details.

By Proposition 4.0.1 there exists a constant \( c_0 > 0 \) such that for any \( n, m \in \mathbb{N} \),

\[ |\text{Cov}(F(\Xi_{n,N}), F(\Xi_{m,N}))| \leq c_0 \tau(\frac{1}{3}d_N(n, m)) \]

where \( \tau(k) = \phi_k^{1-\frac{1}{w}} + \beta_{\epsilon,k}^\nu \). Observe that by Assumption 2.1.3 we have \( \tau(k) \leq dk^{-(1-\frac{1}{w})^\theta} \), \( k \geq 0 \). The assumptions in Theorem 2.3.3 imply that \( \theta(1-\frac{1}{w}) \) > 1, and therefore

\[ \sum_{k=0}^{\infty} \tau(k) < \infty. \]

Let \( d^* \) be the maximal of the polynomials \( q_i \). Then for any \( k \geq 0 \) and \( n \in \mathbb{N} \) there exist at most \( 2\ell^2 d^* \) natural \( m \)'s such that \( d_N(n, m) = k \). Hence, for any \( n \in \mathbb{N} \) we have

\[ \sum_{m=1}^{\infty} |\text{Cov}(F(\Xi_{n,N}), F(\Xi_{m,N}))| \leq 6\ell^2 d^* c_0 \sum_{s=0}^{\infty} \tau(s) := A < \infty \]

for any natural \( n \), which implies that

\[ \text{Var}(S_N(s) - S_N(t)) \leq N^{-1} \sum_{n=[Nt]}^{[Ns]} \sum_{m=[Ns]}^{[Nt]} |\text{Cov}(F(\Xi_{n,N}), F(\Xi_{m,N}))| \leq A \frac{[Ns] - [Nt]}{N} \]

implying (7.12) with \( S_N \) in place of \( W_N \). In that above arguments we have used Assumption 2.1.3 (and that \( \theta(1-\frac{1}{w}) \) > 1) in order to derive that \( A \) is indeed finite.

It remains to show that \( \lim_{N \to \infty} \tau_N \ln^2 N = 0 \). We begin with estimating \( d_1 \) and \( d_4 \). Let \( n \in \mathbb{N} \) and set

\[ A = \{n\} \quad \text{and} \quad B = N_n^c. \]

Let us estimate \( \alpha(A, B) \). Consider the sets

\[ \Gamma_1 = V = \{q_i(n, N) : 1 \leq i \leq \ell\}, \quad \text{and} \quad \Gamma_2 = \{q_i(m, N) : 1 \leq i \leq \ell, m \in N_n^c\}. \]

Then for each \( \gamma_1 \in \Gamma_1 \) and \( \gamma_1 \in \Gamma_2 \) we have

\[ |\gamma_1 - \gamma_2| > l \]
and therefore, there exist disjoint sets \( Q_1, \ldots, Q_L \), \( L \leq 2\ell + 1 \) so that

\[
\Gamma_1 \cup \Gamma_2 = \bigcup_{i=1}^{L} Q_i,
\]
each one of the \( Q_i \)'s is contained either in \( \Gamma_1 \) or in \( \Gamma_2 \) and

\[
q_i + r \leq q_{i+1}
\]
for any \( 1 \leq i \leq L - 1 \), \( q_i \in Q_i \) and \( q_{i+1} \in Q_{i+1} \). Consider the random vector \( U = (U_1, \ldots, U_L) \) where for each \( i \),

\[
U_i = \{ \xi_{m,r(N)} : m \in Q_i \}.
\]
Let \( \{C_1, C_2\} \) be the partition of \( \Gamma_1 \cup \Gamma_2 \) given by

\[
C_i = \bigcup_{j : q_j \subset \Gamma_i} Q_j, \quad i = 1, 2.
\]
Then \( \alpha(A, B) = \alpha(\sigma(U(C_1), U(C_2))) \) and by Proposition 4.0.2 we have

\[
\alpha(A, B) = \alpha(\sigma(U(C_1), U(C_2))) \leq L \phi(r(N))
\]
where \( U(C_i) = \{ U_j : j \in C_i \} \) and \( \sigma(U(C_i)) \) is the \( \sigma \)-algebra generated by \( U(C_i) \), \( i = 1, 2 \). We conclude from (7.3), (7.5), (7.8) and (7.11) that there exists a constants \( C_1 \) and \( C_4 \) so that

\[
d_1 \leq C_1 N^{\frac{1}{2}} (\phi(r(N)))^{1 - \frac{2}{\ell}} \quad \text{and} \quad d_4 \leq C_4 N (\phi(r(N)))^{1 - \frac{2}{\ell}}. \tag{7.13}
\]
Next, we will estimate \( d_2 \). Let \( n, m \in V_N \) be so that \( m \in N_n \). Set

\[
A = \{ n, m \} \quad \text{and} \quad B = N^{c}_n \cap N^{c}_m.
\]
Let use first estimate \( \alpha(A, B) \). Consider the sets \( \Gamma_1 = \{ q_i(s, N) : 1 \leq i \leq \ell, s = n, m \} \) and

\[
\Gamma_2 = \{ q_i(m, N) : 1 \leq i \leq \ell, m \in N^{c}_n \cap N^{c}_m \}.
\]
Then for any \( \gamma_1 \in \Gamma_1 \) and \( \gamma_2 \in \Gamma_2 \) we have

\[
|\gamma_1 - \gamma_2| > l
\]
and therefore, there exist disjoint sets \( Q_1, \ldots, Q_L \), \( L \leq 4\ell + 1 \) so that

\[
\Gamma_1 \cup \Gamma_2 = \bigcup_{i=1}^{L} Q_i,
\]
each one of the \( Q_i \)'s is contained either in \( \Gamma_1 \) or in \( \Gamma_2 \) and

\[
q_i + r(N) \leq q_{i+1}
\]
for any \( 1 \leq i \leq L - 1 \), \( q_i \in Q_i \) and \( q_{i+1} \in Q_{i+1} \). Using this partition, exactly as in the estimates of \( d_1 \) and \( d_4 \) we obtain that

\[
\alpha(A, B) \leq L \phi(r(N)). \tag{7.14}
\]
We conclude from (7.3), (7.6) and (7.11) that there exists a constant \( C_2 \) so that

\[
d_2 \leq C_2 N (\phi(r(N)))^{1 - \frac{2}{\ell}}. \tag{7.15}
\]
Finally, by (7.11) and the Hölder inequality, each one of the summands in \( d_3 \) does not exceed \( CN^{-\frac{1}{2}} \) for some \( C > 0 \) and therefore

\[
d_3 \leq C_3 N^{-\frac{1}{2}} (l(N))^2 = C_3 N^{-\frac{1}{2} + 2\xi_1}. \tag{7.16}
\]
where we used that \( |N_n| \leq K_1 l(N) \) for some \( K_1 > 0 \) and any \( n \) and \( N \). Relying on (7.11), (7.13), (7.15), (7.16) and on the inequality
\[
\phi(r(N)) \leq d(r(N))^{-q} \leq dN^{-q1}
\]
where \( d \) and \( \theta \) come from Assumption \( [4.1.3] \) we conclude that (7.10) holds true and the proof of Theorem 2.3.3 is complete.

8. Applications

In this section we will describe several type of processes \( \{\xi_n\} \) for which all the results stated in Section 2 hold true.

8.1. Hidden Markov chains and related processes. Let \( X \) be a topological space and let \( B \) be the space of all bounded measurable functions on \( X \), equipped with the supremum norm \( \| \cdot \|_\infty \). Let \( R : \mathcal{B} \to \mathcal{B} \) be a positive operator so that \( R1 = 1 \), where \( 1 \) is the function which takes the constant value 1 (i.e. \( R \) is a Markov-Feller operator). We assume here that \( R \) has a stationary probability measure \( \mu \) so that for any \( n \geq 1 \) and \( g \in \mathcal{B} \) we have
\[
\| R^n g - \mu(g) \|_\infty \leq \| g \|_\infty \tau(n)
\] (8.1)
for some sequence \( \tau(n) \) which converges to 0 as \( n \to \infty \). Let \( \{\Upsilon_n\} \) be the stationary Markov chain with initial distribution \( \mu \), whose transition operator is \( R \). Then the inequality (8.1) holds true with \( \tau(n) \) of the form \( \tau(n) = Cn^{-c} \), \( c, C > 0 \) for an aperiodic Markov chain \( \{\Upsilon_n\} \) if, for instance, a version of the Doeblin condition holds true (see, for instance, Section 21.23 in [5]).

Next, for any \( 0 \leq n \leq m \) let \( F_{n,m} = \sigma\{\Upsilon_n, ..., \Upsilon_m\} \) be the \( \sigma \)-algebra generated by the random variables \( \Upsilon_n, \Upsilon_{n+1}, ..., \Upsilon_m \). When \( n \) is negative we set \( F_{n,m} = F_{0,\max(0,m)} \). The following result is well known (see [5]), but it has a short proof which is given here for readers’ convenience.

8.1.1. Lemma. Suppose that (8.7) holds true. Then the mixing coefficients \( \phi(n) \) corresponding to the \( \sigma \)-algebras \( F_{n,m} \), defined above satisfy \( \phi(n) \leq 2\tau(n) \).

Proof. First (see [5], Ch. 4), for any two sub-\( \sigma \)-algebras \( \mathcal{G}, \mathcal{H} \subset \mathcal{F} \) we have
\[
\phi(\mathcal{G}, \mathcal{H}) = \frac{1}{2} \sup\{\|E[h|\mathcal{G}] - Eh\|_\infty : h \in L^\infty(\Omega, \mathcal{H}, P), \|h\|_\infty \leq 1\}
\] (8.2)
where \( \phi(\mathcal{G}, \mathcal{H}) \) is defined in (2.2) (so \( \phi(n) \) is given by (2.1)). Let \( k \) and \( n \) be nonnegative integers, and set
\[
\mathcal{G} = \sigma\{\Upsilon_0, \Upsilon_1, ..., \Upsilon_k\}, \quad \text{and} \quad \mathcal{H} = \sigma\{\Upsilon_m : m \geq k + n\}.
\]
Let \( h \) be a \( \mathcal{H} \)-measurable random variable which is bounded \( P \)-a.s. by 1. Then we can write \( h = H(\Upsilon_{k+n}, \Upsilon_{k+n+1}, ...) \) for some measurable function \( H \) so that \( |H| \leq 1 \). Set \( H_{n+k} = E[h|\Upsilon_0, \Upsilon_1, ..., \Upsilon_{n+k}] \). Then \( H_{n+k} = H_{n+k}(\Upsilon_{n+k}) \) is a function of \( \Upsilon_{n+k} \), and
\[
E[h|\mathcal{G}] = E[H_{n+k}(\Upsilon_{n+k})|\mathcal{G}] = R^n H_{n+k}(\Upsilon_k)
\]
where we used that \( \sigma\{\Upsilon_0, \Upsilon_1, ..., \Upsilon_{n+k}\} \) is finer than \( \mathcal{G} \) (and the tower property of conditional expectations). Using (8.1) and taking into account that \( |H_{n+k}| \leq 1 \) and that \( EH = E[H_{n+k} H_{n+k}] = \mu_{n+k}(H_{n+k}) \) we obtain that
\[
\|E[h|\mathcal{G}] - Eh\|_L^\infty \leq \tau(n).
\]
Taking the supremum over all the above functions \( h \) and using (8.2) we obtain that
\[
\phi(\mathcal{G}, \mathcal{H}) \leq 2\tau(n).
\]

Taking the supremum over all choices of \( k \) completes the proof of the lemma. \( \square \)

Let \( f = (f_1, \ldots, f_d) \) be a measurable bounded function and for each \( n \geq 0 \) set \( \xi_n = f(\Upsilon_n) \). Then, in the notations of Section 2 we have \( \beta_q(r) = 0 \) for any \( r \geq 0 \) and \( q \). Therefore, all he results stated there hold true with the stationary sequence \( \{\xi_n\} \) when \( \tau(n) \) satisfies \( \tau(n) \leq dn^{-\theta} \) for some \( d > 0 \) and \( \theta > 0 \). When the random variables \( \xi_n \) above only belong to \( L^w \) for some \( w > 2 \) then we our results hold true when \( \tau(n) \leq d^{-\theta} \) for some \( d > 0 \) and \( \theta > \frac{4w}{w-2} \). In particular we can consider the case when \( \tau(n) \) converges exponentially fast to 0. Suppose now that \( (\mathcal{X}, \rho) \) is a metric space. Let \( \rho_\infty \) be the metric on \( \mathcal{X}^\mathbb{N} \) given by
\[
\rho_\infty(x, y) = \sum_{j=1}^{\infty} 2^{-j} \frac{\rho(x_j, y_j)}{1 + \rho(x_j, y_j)}, \quad x = (x_j), \quad y = (y_j).
\]
Let \( f : \mathcal{X}^\mathbb{N} \to \mathbb{R}^d \) be a H"older continuous function with respect to the metric \( \rho_\infty \), and for each \( n \geq 0 \) set \( \xi_n = f(\Upsilon_n, \Upsilon_{n+1}, \Upsilon_{n+2}, \ldots) \). In these circumstances, it is clear that
\[
\beta_\infty(r) = \sup_{n \geq 1} \|E[\xi_n | \mathcal{F}_{n-r,n+r}]\|_{L^\infty} \leq A\delta^r
\]
for some constants \( A > 0 \) and \( \delta \in (0, 1) \). Hence, \( \tau(n) \leq dn^{-\theta} \), then all the results stated in Section 2 hold true for the stationary sequence \( \{\xi_n\} \) defined above. Here \( \theta > \frac{4w}{w-2} \), \( d > 0 \) and \( w > 2 \) satisfies that \( \|\xi_n\|_{L^w} < \infty \) (assuming that such a \( w \) exists).

We can also consider several types of linear Markov processes, described in what follows. Let \( \{\Upsilon_n : n \in \mathbb{Z}\} \) be a two sided stationary Markov chain with transition operator \( R \) and stationary distribution \( \mu \). Then Lemma 8.1.1 also holds true with the \( \sigma \)-algebras \( \mathcal{F}_{n,m} = \sigma(\Upsilon_{n}, \ldots, \Upsilon_{m}) \). Let \( (a_n) \) be a two sided sequence such that \( \sum_{n \in \mathbb{Z}} |a_n| < \infty \) and let \( f : \mathcal{X} \to \mathbb{R} \) be a bounded function so that \( \int f d\mu = 0 \). For each \( i \) set
\[
\xi_i = \sum_{n \in \mathbb{Z}} a_n f(\Upsilon_{n+i}).
\]
Then \( \{\xi_i : i \in \mathbb{Z}\} \) is a bounded stationary sequence of random variables. Observe that for each \( n \) and \( k \geq 0 \),
\[
|E[f(\Upsilon_n) f(\Upsilon_{n+k})]| = |E[R_k f(\Upsilon_n) f(\Upsilon_n)]| \leq \|f\|_{L^\infty} \|R_k f\|_{L^\infty} = \|f\|_{L^\infty} \|R_k f - \mu(f)\|_{L^\infty} \leq \|f\|_{L^\infty}^2 \tau(k)
\]
where in the last inequality we used (8.1). Suppose that \( \sum_{n=0}^{\infty} \tau(n) < \infty \). Then, using (8.3), a direct calculation shows that for any \( r > 0 \),
\[
\left\|\xi_i - \sum_{|n| \leq r} a_n f(\Upsilon_{n+i})\right\|_{L^2} \leq C \sum_{|n| > r} |a_n|
\]
where \( C \) is some constant. Therefore, the approximation coefficients \( \beta_2(r) \) defined in Section 2 satisfy
\[
\beta_2(r) \leq C \sum_{|n| > r} |a_n|.
\]
Thus, when \( \sum_{|n| > r} |a_n| \) and \( \tau(r) \) converge to 0 sufficiently fast as \( r \to \infty \), all of the results stated in Section 2 (with \( q = 2 \)) hold true with the stationary sequence \( \{\xi_i : i \geq 0\} \).

8.2. Subshifts of finite type (and uniformly hyperbolic and distance-expanding maps) and continued fraction expansions. Next, we recall the definition of a (topologically mixing) subshift of finite type. Let \( d > 1 \) be a positive integer and set \( \mathcal{A} = \{1, 2, ..., d\} \). We consider here \( \mathcal{A} \) as a discrete topological space, and let \( \mathcal{X} = \mathcal{A}^{\mathbb{N}} \) be the product (topological) space. We define a metric on \( \mathcal{X} \) by

\[
d(x, y) = 2^{-\inf\{n : x_n \neq y_n\}}
\]

where we set \( \inf \emptyset = \infty \) and \( 2^{-\infty} = 0 \). Then the product topology is generated by this metric. Let \( A = (A_{i,j}) \) be a \( d \times d \) matrix with \( 0 - 1 \) entries. Suppose that \( (A^M)_{i,j} > 0 \) for some \( M \) and all \( 1 \leq i, j \leq d \), and set

\[
\Sigma(A) = \{(x_i) \in \mathcal{X} : A_{x_i, x_{i+1}} = 1 \forall i \geq 0\}.
\]

Let \( T : \Sigma(A) \to \Sigma(A) \) be the left shift given by

\[
(Tx)_i = x_{i+1}, \quad i \geq 0.
\]

Then \( \Sigma(A) \) is \( T \)-invariant. Let \( \mu \) be any invariant Gibbs measure (see [4]). For each finite word \( (a_0, ..., a_r) \in \mathcal{A}^{r+1} \) we define its corresponding cylinder set \( [a_0, ..., a_r] \) to be the set of all \( x \in \Sigma(A) \) so that \( x_i = a_i \) for any \( 1 \leq 0, 1, ..., r \). The length of such a set is defined to be \( r + 1 \). Let \( \mathcal{F}_{0,n} \) be the \( \sigma \)-algebra generated by all cylinder sets of length \( n \) and for each \( 0 \leq n \leq m \) set \( \mathcal{F}_{n,m} = T^{-n}\mathcal{F}_{0,m-n} \). When \( n \) is negative we set \( \mathcal{F}_{n,m} = \mathcal{F}_{0,\max(0,m)} \). Then (see [4]), these \( \sigma \)-algebras satisfy that \( \phi(n) \leq A\delta^n \) for some \( A > 0 \) and \( \delta \in (0, 1) \) (in fact, we also have exponentially fast \( \psi \)-mixing).

For each \( n \geq 0 \) set \( \xi_n(x) = T^n f(x) \), where \( f = (f_1, ..., f_d) \) is an \( \mathbb{R}^d \)-valued function, each \( f_i \) is a Hölder continuous function, and \( x \) is distributed according to \( \mu \). Then

\[
\beta_\infty(r) = \sup_{n \geq 0} \|\xi_n - \mathbb{E}[\xi_n | \mathcal{F}_{n-r, n}]\|_{L^\infty} \leq Ca^r
\]

for some \( C > 0 \) and \( a \in (0, 1) \). Note that when \( f \) is constant on cylinder sets (and hence Hölder continuous) then \( \beta_\infty(r) = 0 \) for any sufficiently large \( r \). We conclude that all the results stated in Section 2 hold true for the sequences \( \{\xi_n\} \) defined above. Using [4], we obtain that the results from Section 2 also hold in the case when \( \xi_n = T^n f(x) \), where \( f = (f_1, ..., f_d) \), \( T \) is a hyperbolic diffeomorphism or an expanding transformation taken with a Gibbs invariant measure, and each \( f_i \) is either Hölder continuous or piecewise constant on elements of Markov partitions.

Next, set \( \mathcal{X} = (0, 1) \setminus \mathbb{Q} \), let \( T : \mathcal{X} \to \mathcal{X} \) be the Gauss map which is given by \( Tx = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor \), and let \( \mu \) be the unique absolutely continuous \( T \)-invariant probability measure given by \( \mu(A) = \frac{1}{\ln 2} \int_A \frac{1}{x} dx \). Let \( A \) be the partition of \( \mathcal{X} \) into the intervals \( I_n = \left( \frac{1}{n+1}, \frac{1}{n} \right) \), where \( n = 1, 2, ..., \) for each \( i = 0, 1, 2, ..., \) let \( n_i(x) \) be the unique positive integer so that \( T^i x \in I_{n_i(x)} \). Then the map \( x \to (n_0(x), n_1(x), ...) \) represents the continued fraction expansion of \( x \). Set \( \mathcal{F}_{n,m} = \bigvee_{j=n}^{m} T^{-j}A \), where \( n_0 = \max(0, n) \) and \( n_0 = \max(0, m) \). Then these \( \sigma \)-algebras are exponentially fast \( \psi \)-mixing, and, in particular \( \phi(n) \leq A\delta^n \) for some \( A > 0 \), \( \delta \in (0, 1) \) and all nonnegative integers \( n \). Moreover, the partition \( \mathcal{F}_{0,n} \) is a partition into intervals whose lengths do not exceed \( Ce^{-cn} \) for some constants \( c, C > 0 \). Therefore, all the conditions from Section 2 also holds true for stationary sequences of the form
$\xi_n = f \circ T^n(x)$, where $x$ is distributed according to $\mu$ and $f : [0, 1] \to \mathbb{R}^d$ is either a Hölder continuous function or a function which is constants of the elements of the partition $\mathcal{F}_{0,r}$ for some fixed $r$.

8.3. Extension to Young towers. Let $(\Delta, \nu, T)$ be the noninvertible and mixing Young tower considered in [23] (or the projected tower considered in Section 3 of [22]). Let $\Delta_0$ be the base of the tower, $R : \Delta_0 \to \Delta_0$ be the return time function and let $d(x, y) = \beta^{s(x,y)}$, $\beta \in (0, 1)$ be the dynamical distance defined by the separation time $s(x,y)$ from [23] (or the one on the projected tower in [22]). In this section we will denote the levels of the tower by $\Delta_\ell$, $\ell \geq 0$ where $\Delta_0$ is identified with $\Delta_0 \times \{0\}$ and for each $\ell > 0$, 

$$\Delta_\ell = \{(x, \ell) : x \in \Delta_0, R(x) > \ell\}.$$ 

Let $\mathcal{L}$ be the transfer operator associated with the tower $T$ (so $\nu$ is its conformal measure- the eigen-measure of $\mathcal{L}$) and $h$ be the eigenfunction of $\mathcal{L}$ (see [22] and [23]). Let us denote by $\mathcal{A}^n$ the $\sigma$-algebra generated by all cylinder sets of length $n$, and let the $\sigma$-algebras $\mathcal{F}_{n,m}$ be given by $\mathcal{F}_{n,m} = T^{-n}\mathcal{A}^{m-n}$, $0 \leq n \leq m$ while when $n < 0$ we set $\mathcal{F}_{n,m} = \mathcal{F}_{0, \max(0,m)}$.

We will consider here processes of the form $\xi_n(x) = f \circ T^n(x)$ where $f : \Delta \to \mathbb{R}^d$ is a Hölder continuous function so that $\|f\|_{L^2(\nu)} < \infty$ for some $q \geq 1$, and $x$ is distributed according to the absolutely continuous invariant measure $\mu$ given by $d\mu = h d\nu$. In this case, it is clear that

$$\beta_\infty(r) = \sup_{n \geq 0} \|\xi_n - E[\xi_n | \mathcal{F}_{n-r,n+r}]\|_{L^\infty} \leq C e^{-cr}$$

for some $C, c > 0$, since we can approximate $f$ uniformly by functions which depend only on elements of the partition $\mathcal{A}^r$, and $E[\xi_n | \mathcal{F}_{n-r,n+r}] = E[f | \mathcal{A}^r] \circ T^n$.

The family of $\sigma$-algebras $\mathcal{F}_{n,m}$ does not seem to be $\phi$-mixing in the sense of Section 2 and so we can not apply the results from Section 2. Note that when the tails $\mu\{R > j\}$ decay polynomially sufficiently fast to 0 then the map $T^R : \Delta_0 \to \Delta_0$ is $\phi$-mixing (see Lemma 2.4 (b) in [19]), in the sense that its corresponding family of $\sigma$-algebras $\mathcal{F}_{n,m}$ is (left) exponentially $\phi$-mixing. Relying on this we could probably extend the results from Section 2 under certain restriction on the behaviour of the nonconventional sums between two consecutive returns to the base $\Delta_0$. Still, we claim that all the results stated in Section 2 hold true for the above sequence $\{\xi_n\}$ when $\nu\{R > n\} \leq An^{-d}$ for some $A > 0$ and a sufficiently large $d > 0$, without restrictions of that kind.

Using tower extensions, our results also hold true in the case when $\xi_n = f \circ T^n$, for several classes of map which satisfy a certain hyperbolicity or expansion conditions only on some peices of the underlying manifold (or, for any non-invertible dynamical systems that can be modeled by a Young tower). Results in the invertible case also follow (in the exponential tails case), by considering first the projected tower (see Section 3 in [22]) and then proceeding essentially as in Section 4 in [22] in order to derive Theorems 2.3.1 and 2.3.3 for the original system from the corresponding limit theorems on the projected tower. We refer the readers to [22], [23], [19] and [12] for examples of maps $T$ which can be modelled by towers.

In the rest of the section we will explain how to obtain the results from Section 2 in the above Young tower setup. First, let $v > 0$ be a function which is constant
on the levels $\Delta_\ell$ of the tower and define a transfer operator $L$ by

$$Lf = \frac{L(fv)}{v}.$$ 

Then the measure $\nu_L$ given by $d\nu_L = vdv$ is conformal with respect to $L$ and the function $h_L = h/v$ is preserved under $L$. For any measurable set $A$, let us denote by $I_A$ its indicator function. We will rely on

8.3.1. **Lemma.** There exists a constant $C > 0$ so that for any $n \geq 0$, $A \in \mathcal{A}^n$ and an arbitrary measurable set $B$,

$$|\nu(hI_A \cdot 1_B \circ T^{k+n}) - \nu(hI_A)\nu(h1B)| \leq C\nu_L(B) \sup_{g: \|g\| = 1} \|L^k(g/v) - \nu_L(g/v)h_L\|_\infty$$

$$= C\nu_L(B) \sup_{g: \|g\| = 1} \|L^k(g)/v - \nu(g)h/v\|_\infty.$$ 

Here $\|\|_\infty$ stands for the supremum norm and $\|g\| := \|g\|_\infty + K_g$, where $K_g$ is the infimum of the set of values $K$ so that $|g(x) - g(y)| \leq Kd(x, y)$ for any $x, y$ in the same level of the tower.

Before proving this lemma we will first explain how it will be used. In the setup of Section 3 of [22] (the projected tower setup), for some $v$ so that $v_\ell := v|\Delta_\ell = e^{\varepsilon\ell}$ (where $\Delta_\ell$ is the $\ell$-th floor and $\varepsilon > 0$ is some constant), in Section 3 of [22] (Proposition A) it was proved that

$$\sup_{g: \|g\| = 1} \|L^k(g)/v - \nu(g)h/v\| \leq A\delta^k$$

for some $A > 0$ and $\delta \in (0, 1)$. In the setup of [23], when $\nu\{R > j\} \leq Cj^{-d}$ for some $C > 0$ and $d > 2$ then by taking $v_\ell = \ell^{d-2} > 0$ (so that $\sum_\ell v_\ell \nu(R > \ell) < \infty$) we obtain from Proposition 3.13 in [21] that

$$\sup_{g: \|g\| = 1} \|L^n(g/v) - \nu_L(g/v)h_L\|_\infty \leq An^{-(d-2)}$$

for some constant $n$. In fact, also the exponential case is considered in [21] and it is possible also to get the same estimates with the norm $\|\cdot\|$ in place of the supremum norm.

When $v|\Delta_\ell = e^{\varepsilon\ell}$ and $B$ is contained in a the union of the first $j$ floors we get that

$$\nu_L(B) = \nu(v \cdot 1_B) \leq e^{\varepsilon j} \nu(B).$$

Hence, there exists a constant $c > 0$ so that for any $j$ and $k$ satisfying $j \leq ck$ we have

$$\phi(T^{-(n+k)}G_j, \mathcal{A}^n) \leq A e^{-ak} \quad (8.5)$$

where $A, a > 0$ and $G_j$ is the induced $\sigma$-algebra on the union of the first $j$ floors. Similarly, when $v_\ell = \ell^{d-2}$, then for any $j, k$ and $\alpha \in (0, 1)$ so that $j \leq \ell^\alpha$ we have

$$\phi(T^{-(n+k)}G_j, \mathcal{A}^n) \leq Ak^{-(d-2)(1-\alpha)} \quad (8.6)$$

for some constant $A > 0$. We will show after the proof of Lemma 8.3.1 how to use 8.6 in order to derive all the results stated in Section 2 in the Young tower case.
Proof of the Lemma 8.3.1 Let $n, A$ and $B$ be as in the statement of the lemma. Write
\[
\nu(h_B \cdot I_B \circ T^{k+n}) - \nu(h_B) = \nu_L(h_B \cdot I_B \circ T^{k+n}) - \nu_L(h_B) = \int_B (L^k(f_n) - \nu_L(f_n)h_B) d\nu_L
\]
where
\[
f_n = L^n(h_B) = \frac{L^n(h_B)}{v}.
\]
Observe that
\[
\|v f_n\|_\infty \leq \|L^n h\|_\infty = \|h\|_\infty < \infty.
\]
Moreover, by Lemma 4 in [12], for any $x, y$ in the same floor we have
\[
|(v f_n)(x) - (v f_n)(y)| \leq c_1 L^n(h_B)(x, y) \leq c_1 \|h\|_\infty d(x, y) =: C_1 d(x, y)
\]
where $c_1$ is some constant which does not depend on $n$ and $A$ (note that similar estimates appear in the Sublemma at the beginning of Section 4.2 in [22]). We conclude that
\[
|\nu(h_B \cdot I_B \circ T^{n+k}) - \nu(h_B)| \leq C \nu_L(B) \sup_{g \|g\|_1} \|L^k(g/v) - \nu_L(g/v)h_B\|_\infty = C \nu_L(B) \sup_{g \|g\|_1} \|L^k(g)/v - \nu(g)h/v\|_\infty
\]
where $C$ is some constant.

Next, we will explain how to use Lemma 8.3.1 in order to obtain functional central limit theorems for nonconventional polynomial arrays in the case when $\xi_n(x) = f \circ T^n(x)$ discussed at the beginning of this section. For any $j \geq 0$, let $\Delta^{(j)}$ be the union of the first $j$ floors, and let $\chi_j$ be its indicator set. Then there exists a constant $C > 0$ so that for any $r \geq 0$,
\[
\|E[f|A^r](1 - \chi_j)\|_q \leq \|f\|_{2q} (\mu\{R > j\})^{\frac{1}{2q}}
\]
\[
\leq \|g\|_{2q} \cdot (\|h\|_\infty)^{\frac{1}{2q}} (\nu\{R > j\})^{\frac{1}{2q}} \leq C j^{-d/2q}
\]
where we assumed that the tails decay at least as fast as $j^{-d}$, that $\|f\|_{2q} < \infty$ and we write $\|g\|_q = \|g\|_{L^q(\Delta, \mu)}$. Taking into account (8.4) with $n = 0$ we conclude that
\[
\beta_q(r, j) := \sup_{n \geq 0} \|\xi_n - (E[f|A^r] \cdot \chi_j) \circ T^n\|_q = \sup_{n \geq 0} \|f - E[f|A^r] \chi_j] \circ T^n\|_q
\]
\[
= \|f - E[f|A^r] \chi_j\|_q \leq \|f - E[f|A^r]\|_q + \|E[f|A^r](1 - \chi_j)\|_q \leq C_1 (e^{-cr} + j^{-d/2q})
\]
for some constant $C_1 > 0$. Using the approximation coefficients $\beta_q(r, j)$ in place of $\beta_q(r)$ from Section 2 and Lemma 8.3.1 the proofs of all the results stated in Section 2 proceed essentially in the same way when $d/2q$ is sufficiently large. Indeed, all the results from Sections 3 and 4 rely only on Proposition 4.0.1 together with several combinatorial arguments. This corollary has an appropriate version which involves the approximation coefficients $\beta_q(r, j)$ (instead of $\beta_q(r)$), since Proposition 4.0.2 can be derived also using the right $\phi$-mixing coefficients. Using this version of Proposition 4.0.1 we obtain all the results from Sections 5 and 6 also in the Young tower case. Relying on the above version of Proposition 4.0.2 Theorem 7.0.1 can be applied successfully also in the Young tower case similarly to Section 7 using the sequence $\beta_q(r(N), j(N))$ instead of $\beta_q(r(N))$, where $r(N) = \lfloor l(N)/3 \rfloor$ is the same as in Section 7 and $j(N) = \lceil (r(N))^\alpha \rceil$ for a sufficiently small $\alpha \in (0, 1)$. Note that
in the appropriate applications of (8.6) we have to take $k = r(N)$ and so, $j = j(N)$ will indeed satisfy $j \leq k^\alpha$.

References

[1] A.D. Barbour, Stein’s Method for Diffusion Approximations, Probab. Th. Rel. Fields 84 (1990), 297-322.
[2] A.D. Barbour and S. Janson, A functional combinatorial central limit theorem, Electron. J. Probab. 14 (2009), 2352-2370.
[3] V. Bergelson, Weakly mixing PET, Ergod. Th.& Dynam. Sys 7(1987), 337-349.
[4] R. Bowen, Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Lecture Notes in Mathematics, volume 470, Springer Verlag, 1975.
[5] R.C. Bradley, Introduction to Strong Mixing Conditions, Volume 1, Kendrick Press, Heber City, 2007.
[6] P.G. Hall and C.C. Hyde, Martingale central limit theory and its application, Academic Press, New York, 1980.
[7] Y. Hafouta and Yu. Kifer, Berry-Esseen type estimates for nonconventional sums, Stoch. Proc. Appl. 126 (2016), 2430-2464.
[8] Y. Hafouta and Yu. Kifer, Nonconventional polynomial CLT, Stochastics, 89 (2017), 550-591.
[9] Y. Hafouta, Stein’s method for nonconventional sums, Electron. Commun. Probab., Volume 23 (2018), paper no. 38, 14 pp.
[10] Y. Hafouta and Yu. Kifer, Nonconventional limit theorems and random dynamics, World Scientific, Singapore, 2018.
[11] Y. Hafouta, Nonconventional moderate deviations theorems and exponential concentration inequalities, to appear in Ann. Inst. H. Poincaré Probab. Statist.
[12] N Haydn and Y Psiloyenis Return times distribution for Markov towers with decay of correlations, Nonlinearity 27(6) (2014), 1323-1349.
[13] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, J. d’Analyse Math., 31 (1977), 204-256.
[14] H. Furstenberg, H, Nonconventional ergodic averages, Proc. Symp. Pure Math. 50 (1990), 43-56.
[15] Yu. Kifer, Nonconventional limit theorems, Probab. Th. Rel. Fields, 148 (2010), 71-106.
[16] Yu. Kifer, A nonconventional strong law of large numbers and fractal dimensions of some multiple recurrence sets, Stoch. Dyn., 12 (2012).
[17] Yu. Kifer and S.R.S Varadhan, Nonconventional limit theorems in discrete and continuous time via martingales, Ann. Probab. 42 (2014), 649-688.
[18] Yu. Kifer, Ergodic theorems for nonconventional arrays and an extension of the Szemerédi theorem, Discrete and Continuous Dynamical Systems-A 38 (2018), 2687-2716.
[19] I. Melbourne and M. Nicol, Almost sure invariance principle for nonuniformly hyperbolic systems, Commun. Math. Phys. 260 (2005), 131-146.
[20] J.B. Rosser and L. Schoenfeld, Approximate Formulas for Some Functions of Prime Numbers, Illinois J. Math. 6 (1962), 64-97.
[21] Véronique Maume-Deschamps, Projective metrics and mixing properties on towers, Trans. Amer. Math. Soc. 353 (2001), 3371-3389.
[22] L.S. Young, Statistical properties of dynamical systems with some hyperbolicity, Ann. Math. 7 (1998) 585-650.
[23] L.S. Young, Recurrence time and rate of mixing, Israel J. Math. 110 (1999) 153-88.