NON-UNIFORM BERRY-ESSEEN THEOREM AND EDGEWORTH EXPANSIONS WITH APPLICATIONS TO TRANSPORT DISTANCES FOR WEAKLY DEPENDENT RANDOM VARIABLES

YEOR HAFOUTA

ABSTRACT. We obtain non-uniform Berry-Esseen type estimates and Edgeworth expansions for several classes of weakly dependent sequences of random variables, including uniformly elliptic inhomogeneous Markov chains, random and time-varying (partially) hyperbolic or expanding dynamical systems, products of random matrices and some classes of local statistics. In [19] S. Bobkov solved a conjecture by E. Rio [80] about optimal central limit theorem (CLT) rates in the transport distances for partial sums of independent random variables, and as an application of our results we will obtain similar results for weakly dependent processes. In particular, we obtain Gaussian coupling with optimal estimates on the error term in $L^p$ (i.e. a version of the Berry-Esseen theorem in $L^p$). Another application of the non uniform expansions is to average versions Berry-Esseen theorem and Edgeworth expansions, which provide estimates of the underlying distribution function in $L^p(dx)$ by the standard normal distribution function and its higher order corrections. An additional application is to expansions of expectations $E[h(S_n)]$ of functions $h$ of the underlying sequence $S_n$, whose derivatives grow at most polynomially fast. In particular we provide estimates on the moments of $S_n$ by means of the corresponding moments of a standard normal random variable and the variance of $S_n$, as well as appropriate expansions.

Contents

1. Introduction 2
1.1. The Berry-Esseen theorem and Edgeworth expansions 2
1.2. Main results: non-uniform Berry-Esseen theorems and expansions for weakly dependent processes 4
1.3. Immediate implications of the non-uniform estimates 5
1.4. Transport distances and their generalizations: Rio’s conjecture for weakly dependent sequences, a generalization of [19] to the weakly dependent case 6
1.5. Organization of the paper 7
2. Main results: abstract form 8
2.1. Non uniform Berry-Esseen theorems and Edgeworth expansions, and their applications to transport distances 8
2.2. Applications 10
2.3. Expansions of functions with polynomially fast growing derivatives 10
2.4. An explicit formula for the polynomials in the general self-normalized case 11
2.5. Weak forms of stationarity 12

Date: June 16, 2023.
1. Introduction

1.1. The Berry-Esseen theorem and Edgeworth expansions. Let $S_n$ be a sequence of centered random variables so that $\sigma_n = \|S_n\|_{L^2} \to \infty$. Recall that $S_n$ obeys the (centralized and self-normalized) central limit theorem (CLT) if $W_n = S_n/\sigma_n$ converges in distribution to the standard normal law, namely for every $x \in \mathbb{R},$

$$\lim_{n \to \infty} \mathbb{P}(W_n \leq x) = \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$  

While the classical CLT concerns partial sums $S_n = \sum_{j=1}^{n} X_j$ of independent random variables $X_j$ which satisfy certain regularity assumptions, by now the CLT has been extended to many classes of weakly dependent (aka “mixing”) summands $X_j$. We note

---

The classical Lindeberg condition, which is satisfied for uniformly bounded random variables.
that the CLT still holds true if we consider sequences of the form \( W_n = (S_n - A_n)/B_n \) for some sequence \( A_n = o(\sigma_n) \) and a sequence \( B_n \) so that \( B_n = \sigma_n(1 + o(1)) \).

Recall also that \( W_n \) obeys the Berry Esseen theorem if there is (an optimal) uniform rate in the CLT, namely

\[
\sup_{x \in \mathbb{R}} |\mathbb{P}(W_n \leq x) - \Phi(x)| = O(\sigma_n^{-1}).
\]

The first results of this kind were obtained independently by Berry and Esseen [14, 44] for partial sums of independent random variables, and by now optimal rates have been established for several classes of weakly dependent and locally dependent sequences for which \( \sigma_n^2 = \text{Var}(S_n) \) grows linearly fast in \( n \) (see [80, 91, 23, 88, 56, 89, 85, 66, 55, 69, 58] for a partial list). One exception is the recent result [34] which includes a Berry-Esseen theorem for certain classes of additive functionals of inhomogeneous Markov chains, for which the variance \( \sigma_n^2 \) of the underlying partial sums \( S_n \) can grow arbitrary slow.

A refinement of the Berry Esseen theorem is the, so-called, Edgeworth expansion. We say that \( W_n = (S_n - A_n)/B_n \) obeys the (uniform) Edgeworth expansion of order \( r \in \mathbb{N} \) if there are polynomials \( H_{j,n} \) with bounded coefficients whose degrees do not depend on \( n \) so that

\[
\sup_{x \in \mathbb{R}} \left| \mathbb{P}(W_n \leq x) - \Phi(x) - \varphi(x) \sum_{j=1}^{r} B_n^{-j} H_{j,n}(x) \right| = o(\sigma_n^{-r})
\]

where \( \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \). Note that the Berry-Esseen theorem corresponds to the case \( r = 0 \), and Edgeworth expansions can be viewed as an expansion of the distribution function of \( W_n \). When taking \( r = 1 \) the corresponding (first order) Edgeworth expansion (when it holds) reveals exactly when better than \( O(\sigma_n^{-1}) \) CLT rates are achieved. In applications (see [17, 68, 17, 34]) in the self-normalized case \( A_n = 0, B_n = \sigma_n \) the coefficients of \( H_{r,n} \) are defined by means of the first \( r + 2 \) moments of \( S_n \), and for \( r = 1 \) they are defined by means of \( \mathbb{E}[S_n^3]/\sigma_n^2 \). Thus, under the validity of the first order expansions we get better than the optimal \( O(\sigma_n^{-1}) \) rates if and only if \( \mathbb{E}[S_n^3] = o(\sigma_n^2) \), and so the second obstruction for better than optimal rates comes from the magnitude of the third moment. When \( S_n \) is a partial sum of a stationary sequence \( X_j \) or when it satisfies a certain “weak form of stationarity” (see Section 2.5), then it is natural also to consider only “stationary” expansions and require that \( H_{j,n} = H_j \) does not depend on \( n \). In this case the polynomials \( H_j \) must be unique (see, for instance, [48]).

1.1.1. Some relevant literature and historical comments. The idea of expanding distribution functions goes back to [28] and [41]. It seems that the first rigorous proof of an Edgeworth expansion was given by Cramér [30], who was motivated by applications in actuarial science. By now Edgeworth expansions have become an important tool in actuarial science and finance, see for instance [5, 50]. In a different direction, Efron [42] used Edgeworth expansions to study resampling techniques (e.g. bootstrapping) and demonstrated their significant superior performance compared to normal approximations, see [43, 64, 74] for an overview.

Since the early work of Cramér, Edgeworth expansions were obtained by many authors in many different setups. For independent random variables it was proven by Esseen in [14],

\[^2\text{An immediate obstruction for the first order expansions is the case when } S_n \text{ take values on some lattice, since then the distribution function of } W_n \text{ has jump of order } 1/\sigma_n.\]
that the expansion of order 1 holds iff the distribution of $S_N$ is non-lattice. The conditions for higher order expansions are not yet completely understood. Sufficient conditions for the Edgeworth expansions of an arbitrary order were first obtained in [30] under the assumption that the characteristic function of the sum $E(e^{itS_N})$ decays exponentially fast as $N \to \infty$ uniformly for large $t$ (this is the so-called, Cramer condition). Later on, the same expansions were obtained in [44, 47, 15, 21, 6] under weaker decay conditions\footnote{The decay conditions used in the above papers are optimal, since one can provide examples where the decay is slightly weaker and there are oscillatory corrections to Edgeworth expansion, see \cite{32, 33, 63}.}, where the second paper considered non identically distributed variables and the fourth and fifth considered random iid vectors. Edgeworth expansions were also proven for several classes of weakly dependent random variables including stationary Markov chains (\cite{78, 79, 48}), chaotic dynamical systems (\cite{26, 48, 49}) and certain classes of local statistics (\cite{16, 65, 11, 24}). In particular, Hervé-Pène proved in [67] that for several classes of stationary processes the first order Edgeworth expansion holds trye if $S_N$ is irreducible, in the sense that $S_N$ can not be represented as $S'_N + H_N$ where $S'_N$ is lattice valued and $H_N$ is bounded. Let us also mention that in \cite{9, 87} certain weak expansions were obtained, i.e. expansions of the form $E(\phi(S_N/\sigma_N))$ where $\phi$ is a smooth test function. Finally, in \cite{34} we obtained optimal conditions for Edgeworth expansions for additive functionals of uniformly elliptic inhomogeneous Markov chains (without assumptions on the growth rates of $\sigma_n$).

1.2. Main results: non-uniform Berry-Esseen theorems and expansions for weakly dependent processes. As opposed to the results discussed in the previous section, this paper concerning non-uniform estimates, which for the sake of convenience are described here as definitions.

1. Definition. We say that $W_n$ obeys the non-uniform Berry-Esseen theorem with power $s \geq 1$ if there is a constant $C = C_m > 0$ for that for every $x \in \mathbb{R}$ we have

$$|P(W_n \leq x) - \Phi(x)| \leq C(1 + |x|)^{-s}\sigma_n^{-1}.$$ 

Similarly we can also consider non-uniform expansions of order $r$:

2. Definition. We say that $W_n$ obeys the Edgeworth expansion of order $r$ with power $s \geq 1$ if for every $x \in \mathbb{R}$,

$$D_{r,n}(x) := |P(W_n \leq x) - \Phi(x) - \varphi(x) \sum_{j=1}^{r} B_n^{-j}H_{j,n}(x)| \leq C_n(1 + |x|)^{-s}\sigma_n^{-r}$$

where $C_n \to 0$ and $H_{j,n}$ are polynomials with the same properties as in the uniform expansions.

For partial sums of iid summands non-uniform Edgeworth expansions were obtained under various conditions, see for instance \cite{80, 81, 82} and \cite{84}. See also \cite{25, 19} for a non-uniform Berry-Esseen theorem for independent but not necessarily identically distributed summands (see also the recent papers \cite{10} and \cite{75}). However, despite the variety of applications of non uniform Berry-Esseen theorems and Edgeworth expansions (described in Sections \ref{sec1.3} and \ref{sec1.4.1} below), the literature about non-uniform estimates for weakly dependent random variables is not as vast as the literature in the uniform case. The goal of this paper is to extend the known results from the uniform case to the non-uniform case, as will be discussed in the following sections.
1.2.1. Main examples of the weakly dependent processes considered in this paper. The purpose of this paper is to obtain non-uniform Berry-Esseen theorems and Edgeworth expansions for the following types of sequences:

- Partial sums $S_n = \sum_{j=0}^{n-1} g \circ T^j$ generated by a chaotic dynamical system $T$ and a sufficiently regular observable $g$ (c.f. [48, 49]);
- Partial sums $S_n = \sum_{j=1}^n g(\xi_j)$ generated by an homogeneous geometrically ergodic Markov chain $\{\xi_j\}$ and a bounded function $g$ (c.f. [66]).
- Partial sums generated by a random (sequential) chaotic dynamical systems and a random (sequential) sufficiently regular observable (c.f. [59]);
- $S_n = \ln \|A_n A_{n-1} \cdots A_1 x\|$ where $A_j$ are iid (strongly irreducible) random matrices and $x$ is a unit vector (c.f. [49]);
- Partial sums $S_n = \sum_{j=1}^n f_j(\xi_j, \xi_{j+1})$ of additive functionals $f_j$ of uniformly elliptic (not necessarily homogeneous) Markov chains $\xi_j$ (c.f. [36, 34]).
- Some classes of local statistics (c.f. [37, Sections 3-5]);

As will be discussed in the next sections, once the appropriate non-uniform estimates are obtained several other results will follow (see also Remark 12).

1.3. Immediate implications of the non-uniform estimates.

1.3.1. $L^p$ type Gaussian estimates of distribution function. First, the non-uniform Berry Esseen theorem of power $s$ yields that for all $p > 1/s$ we have

$$\|F_n - \Phi\|_{L^p(dx)} = O(\sigma_n^{-1})$$

where $F_n(x) = \mathbb{P}(W_n \leq x)$. For independent summands, such optimal CLT rate in the $L^p$ norm is also a classical results, first obtained in [1, 2, 46] (see also [84, Chapter V]). Similarly, the non-uniform Edgeworth expansions of order $r$ and power $s$ yield that for all $p > 1/s$ we have

$$\|D_{r,n}\|_{L^p(dx)} = o(\sigma_n^{-r}).$$

Such results where obtained for partial sums of independent summands in [2], and here we consider the weakly dependent case.

1.3.2. Expansions of functions of $W_n$ with polynomially fast growing derivatives. Let $h : \mathbb{R} \to \mathbb{R}$ be an a.e. differentiable function so that $H_s = \int \frac{|h'(x)|}{(1+|x|)^s} dx < \infty$. Then, with $F_n(x) = \mathbb{P}(W_n \leq x)$ we have\(^4\)

$$\mathbb{E}[h(W_n)] = -\int_{-\infty}^{\infty} h'(x) F_n(x) dx$$

\(^{\text{\footnotesize 4}}\)Instead of geometric ergodicity, we can assume that the corresponding Markov operator has a spectral gap on an appropriate Banach space $B$, and that $g \in B$.

\(^{\text{\footnotesize 5}}\)Indeed,

$$\mathbb{E}[h(W_n)] = -\mathbb{E}\left[\int_{W_n} h'(x) dx\right] = -\int_{-\infty}^{\infty} h'(x) \mathbb{P}(W_n \leq x) dx = -\int_{-\infty}^{\infty} h'(x) F_n(x) dx.$$
and so
\[
\left| \mathbb{E}[h(W_n)] - \int h(x) dD_{r,n}(x) \right| \leq \int |h'(x)| D_{r,n}(x) dx \leq CH_n C_n \sigma_n^{-r}, \ C_n \to 0
\]
where we set \(D_{0,n}(x) = \Phi(x)\). In particular we obtain expansions for the moments of \(W_n\) by taking \(h(x) = x^p\) for all \(p < s\).

1.4. Transport distances and their generalizations: Rio’s conjecture for weakly dependent sequences, a generalization of [19] to the weakly dependent case. Given two Borel probability measures \(\mu\) and \(\nu\) on the real line \(\mathbb{R}\) with absolute moment of order \(p\), we set
\[
W_p(\mu, \nu) = \inf_{\pi} \left( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x-y|^p d\pi(x,y) \right)^{\frac{1}{p}}
\]
where the infimum is taken over all the probability measures on \(\mathbb{R} \times \mathbb{R}\) with marginals \(\mu\) and \(\nu\). Namely
\[
W_p(\mu, \nu) = \inf_{(X,Y)} \|X-Y\|_{L^p}
\]
where the infimum is taken over the class of random variables \((X,Y)\) on \(\mathbb{R} \times \mathbb{R}\) so that \(X\) is distributed according to \(\mu\) and \(Y\) is distributed according to \(\nu\). The function \((\mu, \nu) \rightarrow W_p(\mu, \nu)\) defines a metric on the set of Borel probability measure with absolute moment of order \(p\). Recently Bobkov [19] extended \(W_p\) to the class of finite signed measures with finite absolute moment of order \(p\) according to the formula
\[
\tilde{W}_p(\mu, \nu) = \sup_{u \in U_q} \int_{-\infty}^{\infty} |u(F_\mu(x)) - u(F_\nu(x))| dx
\]
where \(q\) is the conjugate exponent of \(p\), \(U_q\) is the class of all differentiable functions \(u\) so that \(\|u'\|_{L^q(dx)} \leq 1\) and \(F_\mu(x) = \mu((-\infty, x])\) and \(F_\nu = \nu((-\infty, x])\) are the, so-called, generalized distribution functions. Henceforth we will not distinguish between \(W_p\) and \(\tilde{W}_p\).

Let \(Y_1, Y_2, ..., Y_n\) be centered independent square integrable random variables and set \(S_n Y = \sum_{j=1}^{n} Y_j\). Let \(\mu_n\) be the law of \(S_n Y\) and let \(\gamma = d\Phi\) be the standard normal law. Then in [19] was shown (in particular) that for every \(p \geq 1\), if \(Y_i \in L^{p+2}\) then we have the following optimal CLT rate in the \(p\)-th transport distances:
\[
W_p(\mu_n, \gamma) \leq c_p L_{p+2,n}^{1/p}
\]
where \(L_{a,n} = \sum_{j=1}^{n} \mathbb{E}[|Y_j|^a] \) is the \(s\)-th Lyapunov coefficient. This proves the validity of Rio’s conjecture [86]. In the self-normalized case when \(Y_j\) has the form \(Y_j - Y_{j,n} = X_j/\|S_n\|_{L^2}\), with \(S_n = \sum_{j=1}^{n} X_j\), if \(\mathbb{E}[|X_j|^{p+2}] \leq C_p \mathbb{E}[|X_j|^2]\) (e.g. when \(\sup_j \|X_j\|_{L^\infty} < \infty\) the above result reads
\[
W_p(\mu_n, \gamma) = O(\|S_n\|_{L^2}^{-1})
\]
which, in general, is the best possible CLT rate (i.e. the optimal rate). Note that in the iid setup, when \(X_1 \in L^{p+2}\), both error terms \(L_{p+2,n}^{1/p}\) and \(O(\|S_n\|_{L^2}^{-1})\) yield the classical optimal rates
\[
W_p(\mu_n, \gamma) = O(n^{-1/2}).
\]
\(^6\)That is we have \(\frac{1}{q} + \frac{1}{p} = 1\).
Remark that showing that $W_p(\mu, \gamma) = O(n^{-1/2})$ was an old problem starting from Esseen (1958) in the case $p = 1$, and we refer to the introduction of [86] for more details.

A natural extension of Rio’s conjecture is whether the $W_p$-CLT rates $O(\|S_n\|_{L^2}^{-1})$ hold true for important classes of weakly dependent random variables. In this paper we will, in particular, obtain rates of the form $W_p(\mu, \gamma) = O(\|S_n\|_{L^2}^{-1})$ for the weakly dependent random variables described in Section 1.2.1. We would also like to mention the recent paper [7], in which a similar task was obtained for some classes of stationary $\alpha$-mixing random fields (in which $\|S_n\|_{L^2}$ grows linearly fast in $n$) using different methods (an appropriate adaptation of Stein’s method). While $\alpha$-mixing random variables include many important examples in statistics and probability most the examples we have in mind do not seem to be $\alpha$-mixing. Moreover, we do not require stationary of the underlying sequence, and we also obtain the more general results described in the previous sections.

In [19], under the classical Cramer condition, corresponding Edgeworth expansions in the transport distances were obtained, which in the iid case shows in which circumstances better than optimal CLT rates are achieved (see [19, Theorem 1.3]). Among the result obtained in this paper are extensions of such expansions to the aforementioned weakly dependent sequences.

1.4.1. **Estimating the transport distances via non-uniform Berry-Esseen estimates and Edgeworth expansions.** In [19, Corollary 3.2] Bobkov showed that for every finite measures $\mu, \nu$ on $\mathbb{R}$ so that
\[
\int_{-\infty}^{\infty} |x|^p d\mu(x) + \int_{-\infty}^{\infty} |x|^p d\nu(x) < \infty
\]
we have
\[
W_p(\mu, \nu) \leq \int_{-\infty}^{\infty} |F(x) - G(x)|^{1/p} dx
\] (1.2)
where $F(x) = \mu((-\infty, x])$ and $G(x) = \nu((-\infty, x])$ are the generalized distribution functions. Then, in order to solve Rio’s conjecture and the corresponding expansions Bobkov applied a certain type of non-uniform Edgeworth expansions for independent random variables, together with some smoothing argument. One of the goals of this paper is to generalized [19] to the weakly dependent case, and by (1.2) it is enough to obtain non-uniform Berry-Esseen theorems and Edgeworth expansions.

1.4.2. **Applications to coupling: $L^p$ versions of the CLT with optimal rates.** By [19, Lemma 11.1], in order to approximate a random vector $S$ in $L^p$ by a sum of independent Gaussian it is enough to estimate $W_p(dF, d\Phi)$, where $F$ is the distribution function of $S$. Using this result, for all the examples of processes described in Section 1.2.1 the non-uniform Berry Esseen estimates obtained in this paper yield (via the transport distances, using (1.2)) optimal $L^p$ rates of Gaussian approximation of $W_n$ of order $O(\|S_n\|_{L^2}^{-1})$. In all of our examples described in Section 1.2.1 we will get estimates in $L^p$ for all $p$.

1.5. **Organization of the paper.**
1.5.1. **Abstract results: Sections 2-3** In Section 2 we will present our abstract results, which hold true under certain type of logarithmic growth assumptions for general sequences $S_n$ of random variables. Section 2.6 is devoted to application of our abstract results to Gaussian coupling, still under the abstract growth assumptions. Section 3 deals with the special case when $S_n$ satisfies a certain “weak form of stationarity”, where in that case the expansions hold true with polynomials which do not depend on $n$.

1.5.2. **Examples: Sections 4-6** In Section 4 we will describe a few classes of processes which satisfy the above logarithmic growth assumptions and have the “weak form of stationarity”, including partial sums of deterministic (partially) expanding or hyperbolic dynamical systems, homogeneous Markov chains and products of random matrices. In Section 5 we will describe a few examples of non-stationary processes which satisfy our logarithmic growth assumptions including inhomogeneous Markov chains, random and time-varying expanding or hyperbolic dynamical systems and products of non-stationary matrices. In Section 6 we will briefly discuss other types of examples which satisfy some of our assumptions (e.g. local statistics).

1.5.3. **Proof of the abstract results: Sections 7-8** The rest of the sections will be dedicated to proving the abstract results. In Section 7 we will make a reduction to the self normalized case $W_n = S_n/\sigma_n$, which will be treated in Section 8.

2. **Main results: abstract form**

Let $(S_n)$ be a sequence of centered random variables so that, $\sigma_n = \|S_n\|_{L^2} \to \infty$. Let us define $W_n = \frac{S_n - A_n}{B_n}$, where $A_n$ is a bounded sequence and $B_n = \sigma_n + O(1)$. For instance, we can take $A_n = 0$ and $B_n = \sigma_n$ which corresponds to the centralized self-normalized case, but we will see in some of our application to stationary sequences and products of random matrices that the choice of $A_n = c + O(\delta n)$, for some $\delta \in (0, 1)$ and $B_n = \sigma \sqrt{n}$, $\sigma > 0$ will be natural. Let

$$F_n(x) = \mathbb{P}(W_n \leq x)$$

be the distribution function of $W_n$. In this section we will describe our non uniform Berry-Esseen theorems and Edgeworth expansions for $F_n$, under conditions which involve the difference between the logarithmic characteristic function of $S_n/\sigma_n$ and the standard normal one, which is given by

$$\Lambda_n(t) = \ln \mathbb{E}[e^{itS_n/\sigma_n}] + t^2/2.$$

2.1. **Non uniform Berry-Esseen theorems and Edgeworth expansions, and their applications to transport distances.** We consider here the following assumption (for different values of $m$).

3. **Assumption.** For some $m \geq 2$, for all $3 \leq j \leq m + 1$ there exist constants $C_j, \varepsilon_j > 0$ so that

$$\sup_{t \in [-\varepsilon_j, \varepsilon_j]} |\Lambda_n^{(j)}(t)| \leq C_j \sigma_n^{-(j-2)}.$$ (2.1)

4. **Remark.** Assumption 3 only concerns the derivatives of $\ln \mathbb{E}[e^{itS_n/\sigma_n}]$ (since for $j \geq 3$ the second term vanishes). However, it is more natural to present the results using the difference $\Lambda_n(t)$ since it measures the deviation from normality in an appropriate sense.
Our first result is the following non-uniform Berry-Esseen theorem:

5. **Theorem.** Under Assumption 3 the non-uniform Berry-Esseen theorem with power \( s = m \) holds true, namely for all \( x \in \mathbb{R} \) we have

\[
|F_n(x) - \Phi(x)| \leq C\sigma_n^{-1}(1 + |x|)^{-m}
\]

where \( C \) is some constant. In particular, for every \( p > 1/m \) we have

\[
\|F_n - \Phi\|_{L^p(dx)} = O(\sigma_n^{-1}).
\]

Next, to obtain Edgeworth expansions we consider the following assumption. Let \( f_n(t) = \mathbb{E}[e^{itS_n/\sigma_n}] \) and let \( f^{(m)}_n(t) \) be the \( m \)-th derivative of \( f_n \).

6. **Assumption.** For some \( m \geq 3 \), for every \( c > 0 \) (small enough) and \( B > 0 \) (large enough) we have

\[
\int_{c\sigma_n \leq |t| \leq B\sigma_n^{m-2}} \left| \frac{f^{(m)}_n(t)}{t} \right| dt = o(\sigma_n^{-(m-2)}).
\]

7. **Remark.** Let \( \psi_n(t) = \mathbb{E}[e^{itS_n}] \). Then \( f_n(t) = \mathbb{E}[e^{itS_n/\sigma_n}] = \psi_n(t/\sigma_n) \), and so \( f^{(m)}_n(t) = \sigma_n^{-m}\psi_n^{(m)}(t/\sigma_n) \). Hence, Assumption 6 means that for every \( c > 0 \) and \( B > 0 \) large enough we have

\[
(2.2) \quad \int_{c \leq |t| \leq B\sigma_n^{m-3}} \left| \frac{\psi_n^{(m)}(t)}{t} \right| dt = o(\sigma_n^2).
\]

8. **Theorem.** Let Assumptions 3 and 6 hold with the same \( m \). Then there are polynomials \( H_{j,n} \) with bounded coefficients whose degrees do not depend on \( n \) so that with

\[
(2.3) \quad \Psi_{r,n}(x) = \Phi(x) + \frac{r}{\sigma_n} \sum_{j=1}^{r} B_n^{-j} H_{j,n}(x)
\]

for all \( r = 1, 2, ..., m-2 \) we have

\[
|F_n(x) - \Psi_{r,n}(x)| \leq C_n\sigma_n^{-r}(1 + |x|)^{-m}, \quad C_n \to 0.
\]

In particular, for all \( p > 1/m \),

\[
\|F_n - \Psi_{k,n}\|_{L^p(dx)} = o(\sigma_n^{-k}).
\]

The coefficients of the polynomials are polynomial functions of \( A_n, B_n - \sigma_n \) and \( \Lambda^{(j)}_{n}(0)\sigma_n^{j-2} \) for \( j \geq 3 \).

Our next result is a Berry-Esseen theorem in transport distances:

9. **Theorem.** Under Assumption 3 the Berry-Esseen theorem in the transport distance of any order \( p < m - 1 \) holds true. Namely, there exists a constant \( C > 0 \) so that

\[
W_p(dF_n, d\Phi) \leq C\sigma_n^{-1}.
\]

Next, as explained in Section 1.4, the following result follows from Theorem 8 together with [19, Corollary 3.2]:

\[8\] Namely, the non-uniform Edgeworth expansion or order \( r \) and power \( m \) holds true.
10. **Theorem.** Under Assumptions 3 and 6 we have the following: let $H_{r,n}$ be the polynomials from Theorem 8. Then for all $r = 1, 2, \ldots, m - 2$ and $p < m - 1$ we have

$$W_p(dF_n, d\Psi_{r,n}) = o(\sigma_n^{-r})$$

where $\Psi_{r,n}$ is defined in (2.3).

As described in Section 1.3, Theorem 9 follows from Theorem 10 together a certain smoothing argument.

2.2. **Applications.**

2.3. **Expansions of functions with polynomially fast growing derivatives.** As explained in Section 1.2, the following two results follow from Theorems 9 and 10.

11. **Corollary.** Let $h : \mathbb{R} \to \mathbb{R}$ be an a.e. differentiable function so that $H_m = \int \frac{|h'(x)|}{(1+|x|)^m} dx < \infty$.

(i) Under Assumption 3 there is a constant $R > 0$ which does not depend on $h$ so that

$$\left| \mathbb{E}[h(W_n)] - \int h(x)\varphi(x)dx \right| \leq RH_m\sigma_n^{-1}. $$

In particular, for every $q < m$ we have

$$\left| \mathbb{E}[W_n^q] - \int x^q\varphi(x)dx \right| \leq R_m\sigma_n^{-1}$$

and

$$\left| \mathbb{E}[|W_n|^q] - \int |x|^q\varphi(x)dx \right| \leq R_m\sigma_n^{-1}$$

where $R_m$ depends only on $m$.

(ii) Under Assumption 3 and 6 for all $1 \leq r \leq m - 2$ we have

$$\left| \mathbb{E}[h(W_n)] - \int h(x)d\Psi_{r,n}(x) \right| \leq C_nH_m\sigma_n^{-r}, \ C_n \to 0.$$ 

In particular, for every $q < m$ we have

$$\left| \mathbb{E}[W_n^q] - \int x^q\Psi_{r,n}(x) \right| \leq R_mC_n\sigma_n^{-r}, \ C_n \to 0.$$ 

and

$$\left| \mathbb{E}[|W_n|^q] - \int |x|^q\Psi_{r,n}(x) \right| \leq R_mC_n\sigma_n^{-r}, \ C_n \to 0.$$ 

12. **Remark.** Corollary 11 (ii) yields appropriate expansions for the moments like in [48], but in a more general setup, and for non-integer moments. Moreover, we are able to expand expectations of more general functions.

In comparison with Corollary 11 (ii), in [9] and [87] similar estimates were obtained for different classes of functions $h$. However, in general these results do not yield optimal rates $o(\sigma_n^{-r})$ and they also hold only for $m$ times differentiable functions $h$ satisfying some regularity conditions. On the other hand, the results in [9] and [87] apply to sufficiently
fast mixing sequences, which is a different class of processes than the one considered in this paper.

2.4. An explicit formula for the polynomials in the general self-normalized case. Let us denote by
\[ \gamma_j(W) = i^{-j} \left( \frac{d^j}{dt^j} \mathbb{E}[e^{itW}] \right) \bigg|_{t=0} \]
the \( j \)-th cumulant of a random variable \( W \) with finite absolute \( j \) moments. Let \( H_s \) be the \( s \)-th Hermite polynomials. Then in Section 8.1 we will see that in the self normalized case when \( A_n = 0 \) and \( B_n = \sigma_n = \|S_n\|_{L^2} \) our expansions hold true with \( \Psi_{k,n} = \Phi_{k+2,n} \), where for all \( m \geq 3 \),
\[ \Phi_{m,n}(x) = \Phi(x) - \varphi(x) \sum_{k} \frac{1}{k_1! \cdots k_{m-2}!} \left( \frac{\gamma_3(W_n)}{3!} \right)^{k_1} \cdots \left( \frac{\gamma_m(W_n)}{m!} \right)^{k_{m-1}} H_{k-1}(x) \]
where the summation is running over all tuples \( \bar{k} = (k_1, \ldots, k_{m-2}) \) of nonnegative integers so that \( \sum_j jk_j \leq m - 2 \), and \( k = k(k_1, \ldots, k_{m-2}) = 3k_1 + \cdots + mk_{m-2} \). Note that \( \gamma_j(W_n) = \Lambda_n^{(j)}(0) \) and so under Assumption 3 we have
\[ \gamma_j(W_n) = O(\sigma_n^{-(j-2)}) \text{, } j = 3, 4, \ldots, m + 1. \]
Observe that \( \Phi_{m,n}(x) \) can also be written in the form
\[ (2.4) \quad \Phi_{n,m}(x) = \Phi(x) - \varphi(x) \sum_{j=1}^{m-2} \sigma_n^{-j} H_{j,n}(x) \]
where
\[ (2.5) \quad H_{j,n}(x) = \sum_{k \in A_j} \prod_{j=1}^{s(k)-2} (\gamma_{j+2}(S_n)\sigma_n^{-2})^{k_j} H_{k-1}(x) \]
and \( A_j \) is the set of all tuples of nonnegative integers \( (k_1, \ldots, k_s) \), \( k_s \neq 0 \) for some \( s = s(\bar{k}) \geq 1 \) so that \( \sum_l lk_l = j \) (note that when \( j \leq m - 2 \) then \( s \leq m - 2 \) since \( k_s \geq 1 \)). Moreover,
\[ k = k(k_1, \ldots, k_s) = 3k_1 + \cdots + (s+2)k_s. \]

2.4.1. Better than optimal CLT rates in transport distances. The explicit formula for \( H_{j,n} \) together with [19, Proposition 5.1] yield the following result.

13. Theorem. Let the conditions of Theorem 10 hold. Then for all \( r = 1, 2, \ldots, m-2 \) and \( p < m-1 \) such that \( r \geq p \) we have the following. If \( \gamma_j(S_n)/\sigma_n^j = o(\sigma_n^{-r}) \) for all \( 3 \leq j \leq 2 + r/p \) then
\[ W_p(dF_n, d\Phi) = o(\sigma_n^{-r}). \]
In particular (by taking \( r = p \)),
\[ W_p(dF_n, d\Phi) = o(\sigma_n^{-1}) \]
if \( \gamma_3(S_n) = \mathbb{E}[S_n^3] = o(\sigma_n^2) \).

\[ \text{Note: The Markov chains considered in Section 5.1 will satisfy both our assumptions and the mixing assumptions in [10] and [87], but our results hold for a different class of functions } h, \text{ which might only be } C^1 \text{ with a sufficiently fast decaying derivative. Moreover, we obtain optimal rates.} \]
Recall that in [45], for partial sums with iid summands it was shown that we have uniform CLT rates of order $o\left(\sigma_n^{-1}\right)$ if the third moment of $X_1 - \mathbb{E}[X_1]$ vanishes. In [34] we provided an appropriate version of this phenomena for additive functionals of uniformly elliptic inhomogeneous Markov chains (where the condition about the third moment is replaced by $\mathbb{E}[S_n^3] = o(\sigma_n^2)$). The above theorem provides similar result but in the metric $W_p$.

2.5. **Weak forms of stationarity.** In this section we will introduce an assumption which will hold true in all of our applications to sequences which are generated by an underlying stationary process.

14. **Assumption.** There are numbers $p_k, q_k, k \geq 2$ and $\delta \in (0, 1)$ so that $p_2 > 0$ and for all $2 \leq k \leq m$ and $n \in \mathbb{N}$ we have

$$\gamma_k(S_n) = np_k + q_k + O(\delta^n).$$

In particular, $p_2 = \lim_{n \to \infty} \frac{1}{n} \text{Var}(S_n) > 0$ and $\sigma_n^2 = \gamma_2(S_n) = np_2 + q_2 + O(\delta^n)$.

15. **Remark.** Note that

$$p_k = \lim_{n \to \infty} \frac{1}{n} \gamma_k(S_n)$$

is the, so-called, $k$-th asymptotic cumulant of $S_n$.

16. **Remark.** All the results stated in this section hold true if $\gamma_k(S_n) = np_k + q_k + O(\sigma_n^{-w_k})$ for some $w_k$ which can be recovered from the proofs, but in all the application we have in mind the error term will be $O(\delta^n)$.

17. **Theorem.** Under Assumptions [9] and [14], Theorems 8 and 10 hold true with polynomials $H_{r,n} = H_r$ which do not depend on $n$ when $A_n = \text{const} + O(\delta^n)$ for some $\delta \in (0, 1)$ and either $B_n = \sigma_n$ or $B_n = \sqrt{p_2 n}$.

18. **Remark.** The polynomials $H_j$ are unique, and we refer to [3.3] for an explicit formula in the case $A_n = 0$ and $B_n = \sigma_n$ (in this case the coefficients of the polynomials depend only on $a_l = q_{l+2} - \frac{q_{p_{l+2}}}{p_{l+2}}$ and $\beta_l = p_{l+2}$ for $l \leq 3j$). We also note that the arguments in our proofs also yield a formula for $H_j$ when $B_n = \sqrt{p_2 n}$, and in this case the coefficients of the polynomials depend only on $p_k$ and $q_k$ and the constant part in $A_n = \text{const} + O(\delta^n)$.

2.6. **Applications to Gaussian coupling: A Berry-Esseen theorem in $L^p$.** The following result is a consequence of Theorem 9 together with [10] Lemma 11.1.

19. **Corollary.** Let Assumption [3] hold with some $m$ and suppose that $B_n$ is monotone increasing. Let $(Z_j)$ be a sequence of centered normal random variables so that $\text{Var}(Z_j) = B_j^2 - B_{j-1}^2$, where $B_0 = 0$.

Set $X_i = S_i - S_{i-1}, S_0 = 0$. Then for every $n$ there is a coupling of $(X_1, ..., X_n)$ with $(Z_1, ..., Z_n)$ so that

$$\left\| S_n - \sum_{j=1}^{n} Z_j \right\|_{m-1} \leq C$$

where $C$ is a constant which does not depend on $n$. 


20. **Remark.** In the stationary case we can just take $B_n = p_2 n$ for some $n$. In Sections 5.3 and 5.4 we will show that for certain classes of random and sequential dynamical systems we have $\sigma_n^2 = a_n + O(1)$ for an increasing sequence $(a_n)$. Thus, we can always take $B_n = \sqrt{a_n}$ in the above theorem. We will also see in Section 5.1 that $\sigma_n^2 = a_n + O(1)$ for an increasing sequence $(a_n)$ for uniformly bounded additive functionals of uniformly elliptic inhomogeneous Markov chains.

**Proof of Corollary 19.** For a fixed $n$, let $Y_j = Y_{j,n} = Z_j / B_n; j \leq n$. Then $Y_j$ is a centered normal random variable with variance $s_{j,n} = B_j^2 - B_{j-1}^2$. Note that $Y = \sum_{j=1}^n Y_j$ has the standard normal law. By Applying [19, Lemma 11.1] with $X_i = S_i - S_{i-1}$ (where $S_0 = 0$) and the above $Y_i$, together with Theorem 9, we see that we can couple $(X_1, \ldots, X_n)$ with $(Z_1, \ldots, Z_n)$ so that

$$\left\| S_{n} - \sum_{j=1}^n Z_j \right\|_{m-1} \leq C.$$ 

□

21. **Remark.** For each $n \in \mathbb{N}$, let $Z^{(n)} = (Z_j^{(n)})_{j=1}^\infty$ be copy of $(Z_j)_{j=1}^\infty$. Then, by the, so-called, Berkes-Philipp Lemma [12, Lemma A.1], we can redefine the sequence $(X_j)_{j=1}^\infty$ on a richer probability on which appropriate copies of $(Z_j^{(n)})_{j=1}^n$ are defined so that

$$\sup_n \left\| S_{n} - \sum_{j=1}^n Z_j^{(n)} \right\|_{m-1} < \infty.$$ 

Now, if $(u_n)$ is a subsequence such that $\text{Var}(S_{u_n})$ grows linearly fast in $n$ (e.g. we take $u_n = n$ in the stationary case) then by the Markov inequality and the Borel-Cantelli lemma we see that for every $\varepsilon > 0$ such that $\varepsilon(m - 1) > 1$ we have

$$S_{u_n} - \sum_{j=1}^{u_n} Z_j^{(u_n)} = O(n^\varepsilon), \ a.s.$$ 

When Assumption 3 holds true for every $m$ then we get the rates $O(n^\varepsilon)$ for all $\varepsilon > 0$. These rates of approximation match the best known rates for several types of weakly dependent random variables (see [13]), but the approximating Gaussian sequence depends on $n$, which is much weaker than the usual almost sure invariance principle (where we can take $Z_j^{(n)} = Z_j$).

3. **Forms of stationarity-stationary expansions: proof of Theorem 17 relying on Theorems 5 and 8**

We first need the following result.

3.1. **Stationary expansions in the centralized self-normalized case.** To prove Theorem 17 when $A_n = 0$ and $B_n = \sigma_n$, it is enough to prove the following result.
22. Proposition. Under Assumption [14] there are polynomials $H_j$ whose coefficients are polynomial functions of $\alpha_j = q_j + 2 - \frac{q_j p_j + 2}{p_2}$ and $\beta_j = \frac{p_j + 2}{p_2}$ for $j \leq m - 2$ so that

\begin{equation}
\Phi_{n,m}(x) = \Phi(x) - \varphi(x) \sum_{j=1}^{m-2} \sigma_n^{-j} H_j(x) + e^{-x^2/2} R_n(x) \sigma_n^{-(m-1)}
\end{equation}

where $R_n$ is a polynomial with bounded coefficients and degree depending only on $m$. As a consequence, if Theorems [7, 3, 10, 8] hold true with $A_n = 0$, $B_n = \sigma_n$ and $\Psi_{k,n} = \Phi_{k+2,n}$ given by (2.3), then the same results hold true for every $m$ with the function $\Phi_{m,n}$ given by

\begin{equation}
\Phi_{m,n}(x) = \Phi(x) - \varphi(x) \sum_{j=1}^{m-2} \sigma_n^{-j} H_j(x)
\end{equation}

instead of the function $\Phi_{m,n}$.

We refer the readers to Remark 23 for an explicit formula for $H_r$.

Proof of Proposition 22. Let us recall that

$$
\Phi_{m,n}(x) = \Phi(x) - \varphi(x) \sum_{j=1}^{m-2} \sigma_n^{-j} H_{j,n}(x)
$$

where $H_{r,n}(x)$ are defined in (2.3). Now, by Assumption [14] we have

$$
\gamma_{j+2}(S_n) = n p_{j+2} + q_{j+2} + O(\delta^n)
$$

and

$$
n/\sigma_n^2 = \frac{1}{p_2} - \frac{q_2}{p_2 \sigma_n^2} + O(\delta^n).
$$

Thus,

$$
\gamma_{j+2}(S_n) \sigma_n^{-2} = \sigma_n^{-2} \left(q_{j+2} - \frac{q_2 p_{j+2}}{p_2} \right) + \frac{p_{j+2}}{p_2} + O(\delta^n) =: \alpha_j \sigma_n^{-2} + \beta_j + O(\delta^n).
$$

Since $\sigma_n$ is of order $\sqrt{n}$ we can just disregard the $O(\delta^n)$ term and consider the polynomials

$$
\bar{H}_{r,n}(x) = \sum_{k \in A_r} \prod_{j=1}^{s(k)} (\alpha_j \sigma_n^{-2} + \beta_j)^{k_j} H_{k-1}(x)
$$

instead of $H_{r,n}$. By expanding the brackets

$$
(\alpha_j \sigma_n^{-2} + \beta_j)^{k_j},
$$

rearranging the negative powers of $\sigma_n$ and omitting all the terms which involve higher than $m - 2$ powers of $\sigma_n^{-1}$ we obtain (3.1). \qed

23. Remark. Proceeding as at the end of the proof, we get the following formula for $H_r$:

\begin{equation}
H_r(x) = \sum_{k \in \mathcal{J}_{j-1} A_j} \sum_{u \leq \frac{1}{2}(m-r-2)} \prod_{j=1}^{s(k)} \left( \begin{array}{l}
\ell_j \\
\ell_j
\end{array} \right) \alpha_j \beta_{\ell_j}^{k_j-\ell_j} H_{k-1}(x)
\end{equation}

where given a tuple $\bar{k} = (k_1, ..., k_{s(k)})$ we have that $A_{u, \bar{k}}$ is the set of all tuples $(\ell_1, ..., \ell_{s(\bar{k})})$ of nonnegative integers so that $\sum_j \ell_j = u$ and $\ell_j \leq k_j$, and as before $k = k(\bar{k}) = \sum_j (j+2)k_j$. 
3.2. Stationary expansions with the $\sqrt{n}$ normalization. Theorem 17 in the case when $B_n = \sqrt{p_2 n}$ will follow from the following result.

24. Proposition. Under Assumption 14, there are polynomials $H_j$ so that all the results stated in the previous sections (under the appropriate assumptions) hold true for every $m$ with $B_n = \sqrt{p_2 n}$ and with the function $\Phi_{m,n}$ given by

$$\Phi_{m,n} = \Phi(x) - \varphi(x) \sum_{j=1}^{m-2} n^{-j/2} H_j(x)$$

instead of the function $\Phi_{m,n}$ (and also with the latter function $\Phi_{m,n}(x)$).

3.3. Proof of Proposition 24. We first need the following result.

25. Lemma. There are polynomials $H_j$ whose coefficients are polynomial functions of $q_k$ and $p_k$ for $k \leq 3 j$ so that

$$\Phi_{m,n}(x) = \Phi(x) - \varphi(x) \sum_{j=1}^{m-2} n^{-j/2} H_j(x) + e^{-x^2/2} R_n(x) n^{-(m-1)/2}$$

where $R_n$ is a polynomial with bounded coefficients and degree depending only on $m$.

Proof. Starting from (3.1), we can expand $\sigma_n^r$ in powers of $n^{-1/2}$ using that $\sigma_n^2 = np_2 + q_2 + O(\delta^n)$. Then we can plug in the resulting expression (without the error term) inside (3.1) and obtain the desired result after rearranging the powers of $\sigma_n^{-1}$.

3.4. Completing the proof of Proposition 24. First, by Theorem 17 in the centered self-normalized case and Lemma 25 and arguing as in the proof of Proposition 22, all the result described in Section 2 hold true with the random variable $W_n = S_n/\sigma_n$ and the function

$$\tilde{\Phi}_{m,n}(x) = \Phi(x) - \varphi(x) \sum_{j=1}^{m-2} n^{-j/2} H_j(x)$$

instead of $\Phi_{m,n}$. To complete the proof let first explain how to pass from $W_n$ to $W_n = \frac{S_n - A_n}{\sqrt{p_2 n}}$. First, let $\rho_n = \frac{\sigma_n \sqrt{p_2}}{\sigma_n}$, where $\sigma = \sqrt{p_2} > 0$, and

$$F_n(x) = \mathbb{P}(W_n \leq x) = \mathbb{P}(W_n \leq x \rho_n + A_n/\sigma_n) = F_n(x \rho_n + A_n/\sigma_n).$$

Notice that $\rho_n \to 1$. As a consequence, if we define

$$\tilde{\Phi}_{m,n}(x) = \tilde{\Phi}_{m,n}(\rho_n x + A_n/\sigma_n) = \Phi(\rho_n x + A_n/\sigma_n) - \varphi(\rho_n x + A_n/\sigma_n) \sum_{j=1}^{m-2} n^{-j/2} H_j(\rho_n x + A_n/\sigma_n)$$

then all the results stated in Section 2 hold true with $W_n$ instead of $W_n$ and with $\tilde{\Phi}_{m,n}$ instead of $\Phi_{m,n}$. To complete the proof of Proposition 24 we need to prove the following result:

26. Lemma. There are polynomials $H_j$ whose coefficients are rational functions of $p_k$ and $q_k$ for $k \leq 3 j$ so that with

$$\tilde{\Phi}_{n,m}(x) = \Phi(x) - \varphi(x) \sum_{j=1}^{m-2} n^{-j/2} H_j(x)$$
uniformly in $x$ we have

$$\Phi_{n,m}(x) = \Phi_{n,m}(x) + R_n(x)e^{-x^2/4}O(n^{-m/2})$$

where $R_n$ is a polynomial with bounded coefficients whose degree does not depend on $n$.

**Proof of Lemma 2.4** Recall that

$$\Phi_{m,n}(x) = \Phi(\rho_n x + A_n/\sigma_n) - \varphi(\rho_n x + A_n/\sigma_n) \sum_{j=1}^{m-2} n^{-j/2} \tilde{H}_j(\rho_n x + A_n/\sigma_n).$$

First, let us suppose that $|x| \geq n^\varepsilon$ for some fixed $\varepsilon \in (0, 1/2)$. Then,

$$\max \left( |\Phi_{m,n}(x)|, |\tilde{\Phi}_{m,n}(x)| \right) \leq C_m(1 + |x|^m)e^{-x^2/2}$$

for any choice of polynomials $\tilde{H}_j$ for some $u_m$ which depends only on $m$ and the polynomials $\tilde{H}_j$. Thus, when $|x| \geq n^\varepsilon$ then for any choice of polynomials $\tilde{H}_j$, both $\tilde{\Phi}_{m,n}(x)$ and $\Phi_{m,n}(x)$ are of order $|x|^{u_m}e^{-x^2/4}e^{-n^{2\varepsilon}/4}$ for some $u_m$ which depends only on $m$ and the polynomials. Since

$$e^{-n^{2\varepsilon}/4} = O(n^{-m/2})$$

we see that it is enough to find polynomials $\tilde{H}_j$ which, given $n$ large enough, satisfy the desired properties for points $x$ so that $|x| \leq n^\varepsilon$, where again $\varepsilon$ is a fixed constant which can be chosen to be arbitrarily small.

Let us fix some small $\varepsilon > 0$ and let $n$ and $x$ be so that $|x| \leq n^\varepsilon$. The idea now is to expand all the functions of $\rho_n x + A_n/\sigma_n$ from the definition of $\tilde{\Phi}_{m,n}$ in powers of $n^{-1/2}$. Let us start with $\Phi(\rho_n x + A_n/\sigma_n)$. Using Assumption 14 with $\sigma = \sqrt{p_2} > 0$ we can write

$$\eta = \eta_{n,x} := x(\rho_n - 1) + A_n/\sigma_n = x \left( \frac{p_2 n - \sigma_n^2}{\sigma_n(\sigma_n + \sigma\sqrt{n})} \right) + A_n/\sigma_n$$

$$\frac{-xq_2}{\sigma_n(\sigma_n + \sigma\sqrt{n})} + A_n/\sigma_n + O(\delta^n) = \frac{-xq_2}{\sigma_n(\sigma_n + \sigma\sqrt{n})} + c/\sigma_n + O(\delta^n) := \eta_1 + O(\delta^n)$$

where the constant $c$ is the one satisfying $A_n = c + O(\delta^n)$. Then, using that $|x| \leq n^\varepsilon$, $\sigma_n^2 \asymp p_2 n$ and $A_n = O(1)$ we see that $|\eta| \leq Cn^{-1/2}$. Using now the formula for the Taylor remainder of $\Phi$ around $x$, and taking into account that $\rho_n = 1 + o(1)$ and $A_n = O(1)$ we see that for all $s \in \mathbb{N}$,

$$\Phi(x\rho_n + A_n/\sigma_n) = \Phi(x + \eta) = \Phi(x) + \sum_{j=1}^{s} \frac{\Phi^{(j)}(x)}{j!} \eta^j + r_n(x)$$

where

$$|r_n(x)| \leq C_s(1 + |x|^m)e^{-x^2/4}n^{-\frac{1}{2}(s+1)}$$

and $a_s$ and $C_s$ depend on $s$ but not on $x$ or $n$. Now, by expanding $\eta^j$ in powers of $\eta_1$, absorbing the powers of the term $O(\delta^n)$ times $\frac{\Phi^{(j)}(x)}{j!}$ in $r_n(x)$ and using that for $j \geq 2$ we have

$$\Phi^{(j)}(x) = \varphi^{(j-1)}(x) = (-1)^{j-1} \varphi(x) H_{j-1}(x)$$

we see that, possibly with different constants $a_s, C_s$, we have

$$\Phi(x\rho_n + A_n/\sigma_n) = \Phi(x) - \varphi(x) \sum_{j=1}^{s} \frac{(-1)^{j-1}}{j!}(\zeta^j_n x + c/\sigma_n)^j H_{j-1}(x) + r_n(x)$$
where with \( \sigma = \sqrt{p_2} \),

\[
\zeta_n = \frac{-q_2}{\sigma_n (\sigma_n + \sigma \sqrt{n})}.
\]

By taking \( s = s_m \) so that \( \frac{1}{2} (s + 1) > (m - 1)/2 \), expanding \((\zeta_n x + c/\sigma_n)^2\) using the Binomial formula and expanding \( \zeta_n^k, k \leq s \) in powers of \( n^{-1/2} \) (using that \( \sigma_n^2 = np_2 + q_2 + O(\delta^n) \) and the Taylor expansions of \( g_1(y) = \sqrt{1 + y} \) and \( g_2(y) = \frac{1}{1 + y} \) around the origin) we see that

\[
(3.7) \quad \left| \Phi(x\rho_n + A_n/\sigma_n) - \left( \Phi(x) - \varphi(x) \sum_{j=1}^{m-2} n^{-j/2} E_j(x) \right) \right| \leq C_m (1 + |x|^{v_m}) e^{-x^2/4} n^{-(m-1)/2}
\]

deep only on \( m \), \( E_j \) are polynomials whose coefficients depend only on \( \sigma = \sqrt{p_2}, p_k \) and \( q_k \) (the exact formula for \( E_r \) can be recovered from the latter two expansions).

Similar arguments show that we can expand \( \varphi(x\rho_n + A_n/\sigma_n) \), namely there are polynomials \( U_j \) so that

\[
(3.8) \quad \left| \varphi(x\rho_n + A_n/\sigma_n) - \varphi(x) \sum_{j=1}^{m-2} n^{-j/2} U_j(x) \right| \leq C'_m (1 + |x|^{b_m}) e^{-x^2/4} n^{-(m-1)/2}
\]

with \( b_m \) and \( C'_m \) depending only on \( m \). Thus, so far we have managed to replace the terms \( \Phi(\rho_n x + A_n/\sigma_n) \) \( \varphi(\rho_n + A_n/\sigma_n) \) by \( \Phi(x) \) and \( \varphi(x) \) times polynomials of above form, and it remains to “handle” the terms \( \bar{H}_r(\rho_n x + A_n/\sigma_n) \).

Let us fix some \( j \). Then, using the same notations as above, since \( \bar{H}_j \) is a polynomial we have

\[
\bar{H}_j(\rho_n x + A_n/\sigma_n) = \bar{H}_j(x + \eta) = \sum_{l=0}^{w_j} \left( \sum_{k=l}^{w_j} a_k \binom{k}{j} \eta^{k-j} \right) x^l
\]

where \( w_j \) is the degree of \( \bar{H}_j \) and \( a_k = a_k(j) \) are its coefficients. Now we can further expand \( \eta^{k-l} \) in powers of \( n^{-1/2} \) and disregard the \( O(\delta^n) \) terms. Notice that since \( |x|^{w_j} \leq n^\varepsilon w_j \) then by taking \( \varepsilon \) small enough \( x^l \) times the error term in the above approximation would still be \( O(n^{-(m-1)}) \). By possibly omitting terms of order \( n^{-s/2} \) for \( s > m - 1 \), we conclude that there are polynomials \( V_{j,r} \) so that when when \( |x| \leq n^\varepsilon \) and \( \varepsilon \) is small enough then for all \( j \leq m - 2 \) we have

\[
(3.9) \quad \left| n^{-j/2} \bar{H}_j(\rho_n x + A_n/\sigma_n) - \sum_{r=j}^{m-2} n^{-r/2} V_{j,r}(x) \right| \leq C n^{-(m-1)/2}
\]

where \( C \) is some constant.

Plugging in (3.7), (3.8) and (3.9) inside (3.6), and possibly disregarding terms of order \( n^{-s/2} \) for \( s > m - 1 \), we arrive at (3.4) with some polynomials \( \bar{H}_j \) whose coefficients can be computed by keeping track of all the above expansions (one can first consider expansions in powers of \( n^{-1/2} \) smaller than \( m \) in each one of the above three expansions, and then disregard terms of order \( O(n^{-\frac{1}{2}(m-1)}) \) which come from multiplying all three expansions). \( \square \)
4. **Examples with some stationarity/homogeneity**

4.1. **General scheme.** In this section we describe a functional analytic framework that in the next section will be “verified” for several classes of stationary sequences.

27. **Assumption.** There are analytic functions \( \Pi(z) \), \( U(z) \) and \( \delta_n(z) \) of a complex variable \( z \) which are defined on a complex neighborhood of 0 so that:

1. \( \Pi(0) = 0, \ \Pi''(0) > 0 \) and \( U(0) = 1; \)
2. \( |\delta_n(z)| \leq C \delta^n \) for some \( \delta \in (0, 1) \) and \( C > 0; \)
3. We have

\[
\mathbb{E}[e^{z S_n}] = e^{n \Pi(z)(U(z) + \delta_n(z))}
\]

We note that since \( \delta_n(0) = 0 \) and \( \delta_n(z) \) is uniformly bounded it follows that \( |\delta_n(z)| \leq C |z|^\epsilon \delta^n \) for some \( C > 0 \).

28. **Lemma.** For all \( m \), all the conditions specified in Assumptions \( \mathbb{E} \) and \( 14 \) are met under Assumption \( 27 \). Moreover,

\[
\sigma^2_n = n \Pi''(0) + H''(0) + O(\delta^n).
\]

Furthermore, for all \( j \geq 3 \) we have

\[
\gamma_j(S_n) = n \Pi^{(j)}(0) + H^{(j)}(0) + O(\delta^n).
\]

In other words, Assumption \( 14 \) is in force with \( p_k = \Pi^{(k)}(0) \) and \( q_k = H^{(k)}(0) \).

**Proof.** By taking the logarithm of both sides of (4.1) we see that

\[
\Gamma_n(z) := \ln \mathbb{E}[e^{z S_n}] = n \Pi(z) + H(z) + r_n(z)
\]

where \( r_n(z) = O(|z|\delta^n) \) and \( H(z) = \ln U(z) \) are both analytic functions. Now (4.2) and (4.3) follow by differentiating \( \Gamma_n(z) \), using the Cauchy integral formula and plugging in \( z = 0 \). Note also that the same argument yields that for \( j \geq 3 \), for all \( |t| \) small enough we have

\[
\Lambda_n^{(j)}(t) = \sigma_n^{-j} \Gamma_n^{(j)}(it/\sigma_n) = \sigma_n^{-j} \left( n \Pi^{(j)}(it/\sigma_n) + H^{(j)}(it/\sigma_n) \right) + O(\delta^n)
\]

and since \( \sigma^2_n \simeq \Pi''(0)n \) we see that Assumption \( 8 \) holds true for every \( m \). \( \square \)

4.2. **A functional analytic setup.** Let \( (B, \| \cdot \|) \) be a Banach space of functions on some measurable space \( \mathcal{X} \) so that the constant functions are in \( B \) and \( \|g\| \geq \|g\|_{L^1(\mu)} \) for some probability measure \( \mu \) which, when viewed as linear functional, belongs to the dual \( B^* \) of \( B \). We also assume that \( \|fg\| \leq C \|f\| \|g\| \) for some \( C > 0 \) and all \( f, g \in B \). In addition, we suppose that there is an operator \( \mathcal{L} \) acting on \( B \) so that \( \mathcal{L}^* \mu = \mu \) and a sequence of \( \mathcal{X} \)-valued random variables \( \{U_j : j \geq 1\} \) so that for all \( f_1, \ldots, f_n \in B \) we have

\[
\mathbb{E} \left[ \prod_j f_j(U_j) \right] = \mu \left( \mathcal{L}_{f_n} \circ \cdots \circ \mathcal{L}_{f_2} \circ \mathcal{L}_{f_1} \cdot h_0 \right)
\]
where \( h_0 \in B \) is positive and satisfies \( \mu(h_0) = 1, \mathcal{L}h_0 = h_0 \) and
\[
\mathcal{L}_{f_j}(g) := \mathcal{L}(f_j g).
\]
We then take \( f \in B \) so that \( \mu(f) = 0 \) and set \( S_n = \sum_{j=1}^{n} X_j \), for \( X_j = f_j(U_j) \). For each complex number \( z \in \mathbb{C} \) we also set
\[
\mathcal{L}_z(g) = \mathcal{L}_{e^{zf}}(g) = \mathcal{L}(e^{zf} g).
\]
Let us also assume that\(^\text{10}\)
\[
p_2 = \sigma^2 = \lim_{n \to \infty} \frac{1}{n} \text{Var}(S_n) > 0.
\]
We have the following.

29. **Proposition.** In the above circumstances, Assumption\(^\text{27}\) is in force for all \( m \) if \( z \to \mathcal{L}_z \) is well defined analytic in \( z \) in a complex neighborhood of the origin, the operator \( \mathcal{L} \) is quasi-compact, the spectral radius of \( \mathcal{L} \) is 1, and up to a multiplicative constant, the function \( h_0 \) is the unique eigenfunction corresponding to an eigenvalue of modules one.

**Proof.** By applying an analytic perturbation theorem we get that there is a number \( r > 0 \) so that for every \( z \in \mathbb{C} \) with \( |z| < r \) there are analytic functions of \( z \to \lambda(z) \in \mathbb{C} \setminus \{0\} \), \( z \in h^{(z)} \in B \) and \( z \to \nu^{(z)} \in B^* \) so that
\[
\| \lambda(z)^{-n} \mathcal{L}_z^n - \nu^{(z)} \otimes h^{(z)} \| \leq C \delta^n
\]
where \( C > 0, \delta \in (0,1) \) and the operator \( \nu \otimes h \) is given by \( g \to \nu(g)h \). Moreover, \( \lambda(0) = 1, \nu^{(0)} = \mu \) and \( h^{(0)} = h_0 \). Thus,
\[
\mathbb{E}[e^{zS_n}] = \mu(\mathcal{L}_z^n h_0) = e^n \mathbb{E}[\nu^{\lambda(z)}(U(z) + \delta_n(z))]
\]
where \( U(z) = \nu^{(z)}(h_0)\mu(h^{(z)}) \) (which indeed takes the value 1 at \( z = 0 \)) and \( \delta_n(z) = O(\delta^n) \) for some \( \delta \in (0,1) \). \( \square \)

30. **Remark.** In order to verify Assumption\(^3\) with a given \( m \), it is enough to assume that the operators \( t \to \mathcal{L}_t \) are \( C^{m+1} \) in a (real) neighborhood of the origin. Then, under the same conditions on \( \mathcal{L} \) like in Assumption\(^\text{29}\) (see \[66\]) we get that
\[
(4.5) \quad \mathcal{L}_t^n = \lambda(t)^n \nu^{(it)} \otimes h^{(it)} + N^n(t) = \lambda(t)^n \left( \nu^{(it)} \otimes h^{(it)} + \tilde{N}^n(t) \right)
\]
where \( \tilde{N}(t) = N(t)/\lambda(t) \) and \( N(t) \) satisfies that \( \| N^n(t) \| \leq C \delta^n \) for some \( \delta \in (0,1) \) (all the expressions are \( C^{m+1} \) in \( t \)). Now, since \( \lambda(0) = 1 \), if \( |t| \) is small enough we have \( \| \tilde{N}^n(t) \| \leq \delta_1^n \) for some \( \delta_1 \in (0,1) \). Since all the first \( m+1 \) derivatives of \( \tilde{N}(t)^n \) decay exponentially fast to 0 as \( n \to \infty \) we can verify Assumption\(^3\) for the given \( m \) by differentiating the logarithms of both sides of \( (4.5) \). However, in the applications we have in mind the operators \( z \to \mathcal{L}_z \) will already be analytic.

Now, let us provide abstract conditions that yield estimates on the integral appearing in \[6\]
31. **Proposition.** Let $m \geq 3$. Suppose that for some $c$ domain of the form $\{ |t| \in [c, Bn^{(m-3)/2}] \}$, $c, B > 0$ we have

$$(4.6) \quad \| L_0^m \| \leq A(1 + C|t|) e^{-c|t|^{-\zeta}}$$

for some constants $A, C > 0$ and $0 \leq \zeta < \frac{2}{m-3}$ (where when $m = 3$ we set $\frac{2}{m-3} = \infty$). Then the estimate in Assumption 7 holds true with these $m, c$ and some $B' > 0$.

**Proof.** As explained in Remark 7, it is enough to verify the condition (2.2). First,

$$(4.7) \quad \frac{d^m}{dt^m} E[e^{itS_n}] = i^m E[(S_n)^m e^{itS_n}] = i^m \sum_{1 \leq \ell_1, \ldots, \ell_m \leq n} C_{\ell, m} E[X(\ell)e^{itS_n}]$$

where

$$C_{\ell, m} = \frac{m!}{\ell_1! \cdots \ell_n!}$$

and for a given $\ell = (\ell_1, \ldots, \ell_m)$, if $\ell_q, q \leq m(\ell)$ are the nonzero ones among $\ell$, with $s_j < s_{j+1}$ and $m(\ell) \leq m$ we have

$$X(\ell) = \prod_{q=1}^{m(\ell)} X_{s_q}.$$ 

Now, let us write $\{1, 2, \ldots, n\} \setminus \{s_1, \ldots, s_m(\ell)\}$ as a union of disjoint intervals $I_1, \ldots, I_d$ so that $I_i$ is to the left of $I_{i-1}$. Let $E_j = \{u_j, u_j + 1, \ldots, v_j\}$ be the gap between $I_j$ and $I_{j+1}$. Then $E_j$ is a union of $s_a$’s and hence its length is bounded by $m$. Moreover, the number of gaps is also bounded by $m$ (so $d \leq m$). Denote $e_r = e^{itf} f^r$, and let the operator $L_{j,t}$ be given by

$$L_{j,t} = L_{es} \circ \cdots \circ L_{es_{j+1}} \circ L_{es_j}. $$

Then, using (4.4),

$$E[X(\ell)e^{itS_n}] = \mu \left( L_{it}^{[d]} \circ \cdots \circ L_{it}^{[2]} \circ L_{it}^{[1]} \circ L_{it}^{[1]} 1 \right)$$

and hence

$$|E[X(\ell)e^{itS_n}]| \leq C' \left( \prod_j \| L_{j,t} \| \right) \left( \prod_a \| L_{it}^{[1]} \| \right).$$

Next, notice that $L_{es}(g) = L_{it}(f^{sa} g)$. Thus, since $s_a \leq m$ by using (4.6) with $n = 1$ we see that

$$\| L_{es} \| \leq C_0 \| L_{it} \| \leq AC_0(1 + C|t|)$$

for some $C_0$ which depends only on $f$ and $k$ and the norm $\| \cdot \|$. We thus conclude that

$$\prod_j \| L_{j,t} \| \leq \tilde{C}'(1 + C|t|)^m.$$ 

In order to bound $\prod_a \| L_{it}^{[1]} \|$ we use (4.6) and notice that at least one of the lengths $|I_a|$ of the $I_a$’s is at least of order $[n/2m]$. We conclude that

$$|E[X(\ell)e^{itS_n}]| \leq A'(1 + C|t|)^{2m} e^{-c|t|^{-\zeta}/m}.$$
By summing over all multi-indexes $\ell$ (and using (4.7)) we get that
\[
\left| \frac{d^n}{dt^n} \mathbb{E}[e^{itS_n}] \right| \leq A''_k (1 + C|t|)^{2m} n^m e^{-c_0 n |t| - \zeta/m}.
\]
Hence, on the domain $|t| \in [c, Bn^{(m-3)/2}]$ we see that
\[
\left| \frac{d^n}{dt^n} \mathbb{E}[e^{itS_n}] \right| |t|^{-1} \leq A''_m |t|^{2m-1} e^{-\zeta n^{-1} -(m-3)/2}, \varepsilon_m > 0
\]
which is enough for (2.2) to hold (recall that $1 - (m - 3)\zeta > 0$ and $\sigma_n \approx \sqrt{p_{2n}}$).

32. Remark. (i) When $m = 3$ then the conditions of Lemma 31 hold true for every $c$ and $B$ when the spectral radius of $L_{it}$, $t \neq 0$ is smaller than 1. Indeed, in this case for every compact subset $K$ of $\mathbb{R} \setminus \{0\}$ we have (see [66]),
\[
\sup_{t \in K} \|L^n_{it}\| \leq C(K) e^{-c(K)n}, c(K) > 0
\]
This is usually the case when the function $f$ is non-arithmetic (aperiodic) in an appropriate sense, see [56, 66].

(ii) $m > 3$, we note that if the assumptions of Lemma 31 do not hold true for every $c$ small enough then in order to verify Assumption 6 we also need to estimate the integrals over $[\alpha \sigma_n, c \sigma_n]$ with small $\alpha$’s. This is also requiring that the spectral radius of $L_{it}$ for $t \neq 0$ is smaller than 1.

4.3. Applications to homogeneous elliptic Markov chains. Let $(\xi_j)$ be an homogeneous (not necessarily stationary) Markov chain taking values on some measurable space $\mathcal{X}$. Let $R$ be the corresponding Markov operator, that is $R$ maps a bounded function $g$ on $\mathcal{X}$ to the function $Rg$ on $\mathcal{X}$ given by
\[
Rg(x) := \mathbb{E}[g(\xi_1)|\xi_0 = x].
\]
Let us assume that $R$ has a continuous action on a Banach space $(B, \| \cdot \|)$ of function on $\mathcal{X}$ with a norm $\| \cdot \|$ satisfying $\|g\| \geq \sup |g|$, and that for every function $f \in B$ and a complex number $z$ we have $e^{zf} \in B$ and that the function $z \rightarrow e^{zf}$ is analytic. Moreover, let us assume that $Rg$ is quasi compact and that the constant function 1 is the unique eigenfunction with eigenvalue of modulus 1. Namely, there is a unique stationary measure $\mu$ for the chain and constants $\delta \in (0, 1)$ and $C > 0$ so that for every function $g \in B$ we have
\[
\|R^n g - \mu(g)\| \leq C\|g\|\delta^n.
\]
33. Example. Suppose that the classical Doeblin condition holds true: there exists a probability measure $\nu$ on $\mathcal{X}$ and a constant $c > 0$ so that for every measurable set $\Gamma \subset \mathcal{X}$ we have
\[
\mathbb{P}(\xi_{n_0} \in \Gamma|\xi_0 = x) \geq c\nu(\Gamma).
\]
Then the Markov chain is geometrically ergodic, that is $R$ satisfies the above properties with the norm $\|g\| = \sup |g|$.

Next, let $f \in B$ and set
\[
S_n = \sum_{j=1}^n (f(\xi_j) - \mathbb{E}[f(\xi_j)]).
\]
For a complex parameter $z$, let

$$R_z(g) = R_z(e^{zf}).$$

Then $z \to R_z$ is analytic. Thus by applying a holomorphic perturbation theorem we get that in a neighborhood of the origin there are analytic functions $\lambda(z), \nu(z), h(z)$ of $z$, so that $\lambda(z) \in \mathbb{C} \setminus \{0\}$, $h(z) \in B$, $\nu(z) \in B^*$ ($B^*$ is the dual space) and $\lambda(0) = 1$, $h(0) = 1$ and $\nu(0) = \mu$, where $1$ is the constant function taking the value $1$. Moreover,

$$R^n_z = \lambda^n(z) \left( \nu(z) \otimes h(z) + O(\delta^n) \right)$$

where $(\nu \otimes h)(g) = \nu(g)h$ and $\delta \in (0, 1)$. Thus, with $\Pi(z) = \ln \lambda(z)$ and $U(z) = \mathbb{E}[h^{(z)}(X_1)]$ we have

$$\mathbb{E}[e^{\bar{z} S_n}] = \mathbb{E}[R^n_z \mathbb{1}(X_1)] = e^{n\Pi(z)} (U(z) + O(\delta^n)).$$

By applying Proposition 29 we get the following result.

34. Theorem. Assumption 27 is in force in the above circumstances. Therefore, let $A_n = O(1)$ and either $B_n = \sigma \sqrt{n} + O(1)$ or $B_n = \sigma n + O(1)$. Then the non-uniform Berry-Esseen theorem of any order holds true (Theorem 3 for all $m$), as well as the Berry Esseen theorem in the transport distances of all orders (Theorem 7 for all $m$) and the expectation estimates from Corollary 17 (i) with all $m$. Moreover, the coupling described in Corollary 19 exist for all $n$ and $m$.

35. Remark. We would like also to refer to the second example in Section 6 for related types of Markov chains for which Assumption 27 (and hence the above theorem) hold true.

4.3.1. Expansions of order 1. As mentioned in Remark 32 the conditions of Lemma 31 are always satisfied with $m = 3$ when $f$ is non-arithmetic. We thus conclude that in this case $W_n$ obeys all the types of Edgeworth expansions presented in Section 2 when $m = 3$ (i.e. all the first order expansions hold true).

4.3.2. Higher order expansions. In order to derive the non-uniform Edgeworth expansions (and their consequences) or order $r = m - 2 > 1$ we need to restrict ourselves to a more specific framework. Let us assume that $\mathcal{X}$ is a Riemannian manifold and that

$$Rg(x) = \int_{\mathcal{X}} p(x, y)g(y)dy$$

for some positive measurable function $p(x, y)$ which is bounded and bounded away from the origin, where $dy$ is the volume measure.

36. Theorem. If $f$ is Hölder continuous with exponent $\alpha \in (0, 1]$ then Assumption 6 holds true when $m - 2 < \frac{1 + \alpha}{1 - \alpha}$, where when $\alpha = 1$ we set $\frac{1 + \alpha}{1 - \alpha} = \infty$.

Proof. Combining [36, Lemma 4.3] with [34, Lemma 35] (restricted to the homogeneous case) we have

$$\|L^4_{it}\| \leq e^{-c_1 \delta^2(t)}, \ c > 0$$

where for every $\delta > 0$ on the domain $\{ |t| \in [\delta, \infty) \}$ we have $d^2(t) \asymp |t|^{1 - 1/\alpha}$. Hence, on every domain of the form $\{ |t| \in [\delta, \infty) \}$ we have

$$\|L^m_{it}\| \leq e^{-c \delta |t|^{1 - 1/\alpha} n}, \ c_1 > 0$$

This shows that the conditions of Proposition 31 hold true with every $c > 0$, and hence Assumption 6 is in force.
Remark. The condition \( m - 2 < \frac{1+\alpha}{1-\alpha} \) is optimal even for the uniform expansions (even in the iid case), see [34] (note that the role of \( m - 1 \) there is similar to the role of \( r \) in [34]).

4.4. Application to expanding or hyperbolic dynamical systems. Let \((X, \mathcal{B}, \mu)\) be a probability space and let \( f : \mathcal{X} \to \mathbb{R} \). Let \( T : \mathcal{X} \to \mathcal{X} \) be a measurable map and set
\[
U_j = T^jX_0
\]
where \( X_0 \) is a random element in \( \mathcal{X} \) whose distribution is \( \mu \). Let \( B \) be a Banach space of measurable functions on \( \mathcal{X} \) with the properties described in Section 4.2. Then the conditions of Proposition 29 hold true for a variety of (non-uniformly) expanding maps and measures, some of which listed below:

- \((X, \mathcal{B}, T)\) is a topologically mixing one sided subshift of finite type, \( \mu \) is a Gibbs measure corresponding to some Hölder continuous potential with exponent \( \alpha \) and \( B \) is the space of Hölder continuous functions, equipped with the norm
\[
\|g\| = \sup |g| + v_\alpha(g)
\]
where \( v_\alpha(g) \) is the Hölder constant of \( g \) corresponding to the exponent \( \alpha \). This has applications to Anosov maps \( T \) and invariant measure \( \mu \) which are equivalent to the volume measure, and we refer to [27] for the details.
- \((X, \mathcal{B}, T)\) is a locally expanding dynamical system on some compact manifold \( X \), \( \mu \) is the normalized volume measure and \( B \) is the space of functions with bounded variation (see [71]).
- \((X, \mathcal{B}, T)\) is an aperiodic non-invertible Young tower with exponential tails, \( B \) is the space (weighted) Lipschitz continuous functions and \( \mu \) is the lifted volume measure. This has a variety of applications to partially hyperbolic or expanding maps, and we refer to [93] for the details (see also [77] for a related setup).

4.5. Expansions of order 1. Let \( f \) be aperiodic in the sense of [56, 66] (see Remark 32). Then the spectral radius of \( \mathcal{L}_t \) is smaller than 1 for every \( t \neq 0 \). Moreover, \( f \) is not a coboundary with respect to the map \( T \) (i.e. \( f \) does not have the form \( f = r - r \circ T \)). As mentioned in Remark 32 this is already sufficient for all of the Edgeworth expansions stated in this paper to hold with \( m = 3 \) (i.e. for the first order Edgeworth expansions to hold).

4.5.1. Expansions of all orders. Let \( f \) be aperiodic. As mentioned in Remark 32 in order to verify Assumption 6 when \( m > 3 \) we need to show that the conditions of [31] hold for some \( c > 0 \) and all \( B \). Listed below are types of maps that satisfy these conditions for all functions \( f \in B \) which are not a coboundary with respect to the map \( T \).

- The one dimensional expanding maps with discontinuities as in [22] (for all \( m \) and functions \( f \) with bounded variation).
- The multidimensional expanding maps as in [31] (for all \( m \) and \( C^1 \)-functions \( f \), see [31, Section 3]).

4.6. Application to products of random matrices. Let \( V \) be a \( d \) dimensional vector space for some \( d > 1 \). Let us fix some scalar product on \( V \) and denote by \( \| \cdot \| \) the corresponding norm. Let us denote by \( X = \mathbb{P}V \) the projective space space on \( V \), equipped
with a suitable Remannian distance \(d(\cdot, \cdot)\) (see [20, Chapter II]). Given \(x \in V\) and a sequence \((g_n)\) of iid random variables which takes values on \(G = GL(V)\), we define
\[
S_n(x) = \log \frac{\|g_n \cdots g_1 x\|}{\|x\|}.
\]
Then, as usual, \(S_n(x)\) can be represented as the ergodic sum of \(\bar{\phi}(g, x) = \frac{\log \|g x\|}{\|x\|}\) (see [66]).

Let \(\mu\) be the common distribution of all \(g_n\). Let us consider the following assumptions.

38. Assumption. [Exponential moments] For some \(\delta > 0\) we have
\[
\int_G \max(\|g\|, \|g^{-1}\|)^\delta \, d\mu(g) < \infty.
\]

and

39. Assumption. [Strongly irreducibility and proximal elements] The semi group generated by the support \(\Gamma_{\mu}\) of \(\mu\) has the property that there is no finite union of proper subspaces which is \(\Gamma_{\mu}\)-invariant. Moreover, there exits \(g \in \Gamma_{\mu}\) which has a simple dominant eigenvalue.

As will be explained in what follows, using the arguments of [57], under the above two assumptions the following limits exist
\[
\lambda_1 = \lim_{n \to \infty} \frac{1}{n} \mathbb{E}[\log \|g_n \cdots g_1\|]
\]
and
\[
\sigma^2 = \lim_{n \to \infty} \frac{1}{n} \mathbb{E}[(\log \|g_n \cdots g_1\| - n\lambda_1)^2].
\]
Recall also that for every \(x \in \mathbb{P}V\) we have
\[
\lambda_1 = \lim_{n \to \infty} \frac{1}{n} S_n(x), \text{ a.s.}
\]
and that the above limit is also in \(L^1\), uniformly in \(x\). In particular,
\[
\lambda_1 = \lim_{n \to \infty} \frac{1}{n} \mathbb{E}[S_n(x)].
\]
Moreover, for every \(x \in \mathbb{P}V\) we have
\[
\sigma^2 = \lim_{n \to \infty} \frac{1}{n} \mathbb{E}[(S_n(x) - n\lambda_1)^2].
\]
The latter actually also follows from the CLT for \(n^{-1/2}(S_n(x) - \lambda_1 n)\) (see [49]) together with the existence of the limit \(\lim_{n \to \infty} \frac{1}{n} \mathbb{E}[(S_n(x) - n\lambda_1)^2]\) which will be proven in the proceeding arguments. Finally, note also that, for more general potentials \(\sigma^2 = 0\) if and only if the potential \(\bar{\phi}(g, x)\) admits an appropriate coboundary representation. For the above choice of \(\bar{\phi}\) we have that \(\sigma^2 > 0\) (see [66, 57]).

40. Proposition. Under Assumptions [38] and [39] Assumption [3] is valid for all \(m\), Assumption [2] is valid with \(m = 3\) and Assumption [14] is in force. Moreover, we have \(B_n := \sigma \sqrt{n} = \sigma_n + O(1)\) and \(A_n := \lambda_1 n - \mathbb{E}[S_n] = O(1)\). Thus the sequence \(S_n(x) - n\lambda_1\), obeys the non-uniform Berry-Esseen theorem with all powers \(m\), as in Theorem [3] as well as Theorem [4] and Corollary [11] (i). Moreover, \(S_n - \lambda_1 n\) satisfies the results of Corollary
with all $m$ (with $S_n - \lambda_1 n$ instead of $S_n$). In addition, the non-uniform Edgeworth expansion of order 1 holds true (i.e. Theorem 8 with $m = 3$), as well as Theorem 14 and Corollary 17 (ii) hold true with $m = 3$ (and polynomials which do not depend on $n$).

**Proof.** Let $\| \cdot \|_\varepsilon$, $\varepsilon \in (0,1]$ denote the usual Hölder norm of functions on $X$ corresponding the the exponent $\varepsilon$. For every sufficiently small complex $z$ let us define the linear operator

$$\mathcal{L}_z h(x) = \int_G e^{z(\varphi(g,x) - \lambda_1)h(g \cdot x)}d\mu(g).$$

where $h$ is a bounded function. Since $\|g^{-1}\| x \leq \|g x\| \leq \|g\| x$ Assumption 38 implies that there is a constant $\delta_0 > 0$ so that if $|z| \leq \delta_0$ then $\mathcal{L}_z$ is well defined and acts continuously on the space of Hölder continuous functions (with respect to an arbitrary exponent $\varepsilon$). Moreover, the map $z \to \mathcal{L}_z$ is analytic. Now (see [66, 57]), under Assumptions 38 and 39 the operator $\mathcal{L}_0$ is quasi-compact, its spectral radius is 1, and the constant function 1 is the unique eigenfunction with eigenvalue of modulus 1 (note that $\mathcal{L}_0 1 = 1$). Now, using an analytic perturbation theorem (which follows from an analytic inverse function theorem for Banach space) we see that

$$\mathcal{L}_z^n = (\lambda(z))^n (\Pi_z + r_{z,n})$$

where $z \to \lambda_z$ is an analytic complex valued function, and $z \to \Pi_z$ and $z \to r_{z,n}$ are analytic operator-valued functions so that $\lambda(0) = 1$, $\Pi_0(g) = \nu(g) 1$ for some probability measure $\nu$, $|r_{z,n}| \leq Cr^n$ for some $r \in (0,1)$ and $C > 0$. Moreover, $\Pi_z$ has the form $\Pi_z(\nu) = \nu(h(z))$ for some analytic in $z$ element of the dual space of the space of Hölder continuous functions, and $h(z)$ is a complex valued Hölder continuous function, with the map $z \to h(z)$ being analytic and $h(0) = 1$.

Next, note that

$$E[\exp(z(S_n(x) - n\lambda_1))\xi(g_n \cdots g_1 x)] = \mathcal{L}_z^n \xi(x).$$

for every function $\xi$. Taking $\xi = 1$ we get that, in a complex neighborhood of the origin we have

$$E[\exp(z(S_n(x) - n\lambda_1))] = \mathcal{L}_z^n 1(x) = \lambda_z^n (U(z) + O(r^n))$$

where $U(z) = \Pi_z(1) = \nu(h(z)(\lambda(z))^n, U(0) = 1$. Since $\lambda(0) = 1$ we can write $\lambda(z) = e^{\Pi(z)}$, where $\Pi(z)$ is an analytic function so that $\Pi(0) = 0$.

Next, let us show that Assumption 3 holds true for all $m$. First, by taking the logarithms of both sides of (4.10) and differentiating $k$ times for $k \geq 1$ (using analyticity and the Cauchy integral formula) we see that

$$E[S_n(x)] = n(\Pi(0) + \lambda_1) + H'(0) + O(r^n),$$

where $H(z) = \ln U(z)$, and

$$\text{Var}(S_n(x)) = \Pi''(0)n + H''(0) + O(r^n).$$

By dividing both sides of the first equality by $n$, taking $n \to \infty$ and using (4.8) we see that

$$\Pi'(0) = 0.$$
Next, by dividing both sides of the second quality by \(n\), taking \(n \to \infty\) and using (4.9) and (4.11) we see that
\[
\Pi''(0) = \sigma^2
\]
and so
\[
\text{Var}(S_n(x)) = \sigma^2 n + O(1).
\]
Set \(S_n = S_n(x) - \mathbb{E}[S_n]\). The above estimates show that the choices \(B_n = \sigma \sqrt{n} = \text{Var}(S_n) + O(1)\), and \(A_n = n \lambda_1 - \mathbb{E}[S_n] = O(1)\) fit our general framework described in Section 2.

Next, by taking the logarithms of both sides of (4.10) and differentiating both sides \(j\) times for \(j \geq 3\) and taking into account that \(\sigma^2_n = \text{Var}(S_n(x)) \approx \sigma^2 n\) we see that Assumption 3 is in force with all \(m\). Moreover, by taking \(z = 0\) in the resulting equation (after the differentiation at \(z = 0\)) we get that for each \(j \geq 3\),
\[
\gamma_j(S_n(x)) = n \Pi^{(j)}(n) + H^{(j)}(0) + O(1)
\]
and we conclude that the conditions specified in Assumption 14 hold true with all \(m\). As a consequence, the sequence \(\frac{S_n - \lambda_1 n}{\sigma \sqrt{n}}\) obeys non-uniform Edgeworth expansions of all orders \(m\) with \(B_n = \sqrt{n} \sigma\) and stationary correction terms. Thus the corresponding (stationary) expansions in the transport distances as in Theorem 10 and expansions of the expectations as in (4.11) (ii) hold true for all \(m\).

Next, in order to obtain Edgeworth expansions of order \(r > 1\) let us consider the following condition.

41. **Assumption.** The support of \(\mu\) is Zariski dense in a connected algebraic subgroup of \(G_{L}(V)\).

42. **Proposition.** Under assumption 41, the integrability condition specified in Assumption 7 is met with all \(m\). As a consequence, the sequence \(\frac{S_n - \lambda_1 n}{\sigma \sqrt{n}}\) obeys non-uniform Edgeworth expansions of all orders \(m\) with \(B_n = \sqrt{n} \sigma\) and stationary correction terms. Thus the corresponding (stationary) expansions in the transport distances as in Theorem 10 and expansions of the expectations as in (4.11) (ii) hold true for all \(m\).

**Proof.** First, since \(\bar{\phi}\) is aperiodic for all \(0 < \delta < K\) we have that \(\sup_{\delta \leq |t| \leq K} \|\mathcal{L}_{it}\|_\varepsilon\) decays to 0 at exponential rate (see [57]). Now, by Assumption 41 (see [76, Theorem 4.19]) there exists \(K > 0\) so that if \(\varepsilon\) is small enough then \(\|\mathcal{L}_{it}\|_\varepsilon \leq C |t|^{2\varepsilon + c_0}\) for some \(C, c > 0\) and all \(t\) so that \(|t| \geq K\). Since
\[
|\mathbb{E}[e^{it(S_n(x) - \mathbb{E}[S_n(x)])}]| = |\mathbb{E}[e^{it(S_n(x) - n \lambda_1)}]| = \|\mathcal{L}_{it}\|_\varepsilon
\]
(for every \(\varepsilon\)), we see that the conditions of Lemma 31 hold with all \(m\). 

4.7. **Applications to skew products (annealed limit theorems): a brief discussion.** In [4] it was shown that for some classes of random iid expanding the corresponding skew product is controlled by means of an (“averaged”) quasi-compact operator \(A\). That is, if the underlying maps \(T_o\) map a space \(X\) to itself and the corresponding skew product is denoted by \(T\), then for every function \(\psi\) with bounded variation we have
\[
\mathbb{E}[e^{z S_n \psi}] = \mu(A_z^n 1)
\]
where \(S_n \psi = \sum_{j=0}^{n-1} \psi \circ T^j\) and \(A_z(g) = A(ge^{z \psi})\). Since \(A_z\) is analytic in \(z\), arguing as in the previous sections we see that Assumption 8 holds true with all \(m\).
5. Examples: Nonstationary dynamical systems and Markov chains

5.1. Inhomogenous uniformly elliptic Markov chains. Let \((X_i, \mathcal{F}_i), i \geq 1\) be a sequence of measurable spaces. For each \(i\), let \(R_i(x, dy), x \in X_i\) be a measurable family of (transition) probability measures on \(X_{i+1}\). Let \(\mu_1\) be any probability measure on \(X_1\), and let \(X_1\) be an \(X_1\)-valued random variable with distribution \(\mu_1\). Let \(\{X_j\}\) be the Markov starting from \(X_1\) with the transition probabilities

\[
P(X_{j+1} \in A | X_j = x) = R_j(x, A),
\]

where \(x \in \mathcal{X}_j\) and \(A \subset \mathcal{X}_{j+1}\) is a measurable set. Each \(R_j\) also gives rise to a transition operator given by

\[
R_jg(x) = \mathbb{E}[g(X_{j+1})|X_j = x] = \int g(y)R_j(x, dy)
\]

which maps an integrable function \(g\) on \(\mathcal{X}_{j+1}\) to an integrable function on \(\mathcal{X}_j\) (the integrability is with respect to the laws of \(X_{j+1}\) and \(X_j\), respectively). We assume here that there are probability measures \(\mathbf{m}_j, j > 1\) on \(\mathcal{X}_j\) and families of transition probabilities \(p_j(x, y)\) so that

\[
R_j g(x) = \int g(y)p_j(x, y)d\mathbf{m}_{j+1}(y).
\]

Moreover, there exists \(\varepsilon_0 > 0\) so that for any \(j\) we have

\[
(5.1) \quad \sup_{x,y}p_j(x, y) \leq 1/\varepsilon_0,
\]

and the transition probabilities of the second step transition operators \(R_j \circ R_{j+1}\) of \(X_{j+2}\) given \(X_j\) are bounded from below by \(\varepsilon_0\) (this is the uniform ellipticity condition):

\[
(5.2) \quad \inf \inf_{j \geq 1, x, z} \int p_j(x, y)p_{j+1}(y, z)d\mathbf{m}_{j+1}(y) \geq \varepsilon_0.
\]

Next, for a uniformly bounded sequence of measurable functions \(f_j : \mathcal{X}_j \times \mathcal{X}_{j+1} \to \mathbb{R}\) we set \(Y_j = f_j(X_j, X_{j+1})\) and

\[
(5.3) \quad S_n = \sum_{j=1}^{n} (Y_j - \mathbb{E}(Y_j)).
\]

Set \(V_n = \text{Var}(S_n)\) and \(\sigma_n = \sqrt{V_n}\). Then by [36, Theorem 2.2] we have \(\lim_{n \to \infty} V_n = \infty\) if and only if one can not decompose \(Y_j\) as

\[
Y_j = \mathbb{E}(Y_j) + a_{j+1}(X_{j+1}, X_{j+2}) - a_j(X_j, X_{j+1}) + g_n(X_j, X_{j+1})
\]

where \(a_n\) are uniformly bounded functions and \(\sum_j g_j(X_j, X_{j+1})\) converges almost surely.

Note that when the chain is one step elliptic (i.e. \(\varepsilon_0 \leq p_j(x, y) \leq \varepsilon_0^{-1}\)) and \(f_j(x, y) = f_j(x)\) depends only on the first variable then by [83, Proposition 13] we have

\[
C_1 \sum_{j=1}^{n} \text{Var}(f_j(X_j)) \leq \sigma_n^2 \leq C_2 \sum_{j=1}^{n} \text{Var}(f_j(X_j))
\]

for some constants \(C_1, C_2 > 0\), and so \(\sigma_n \to \infty\) iff the series \(\sum_{j=1}^{\infty} \text{Var}(f_j(X_j))\) diverges.
43. Lemma. There exist a monotone increasing sequence $a_n$ and a bounded sequence $b_n$ so that $\sigma_n^2 = a_n + b_n$ (and so $\sigma_n = \sqrt{a_n} + O(\sigma_n^{-1})$).

Proof. First, by [34] Lemma 27 there are real numbers $u_j$ so that $\sup_{n<k} |u_j| < \infty$ and

\[
C := \sup_{n<k} \left| \text{Var}(S_n - S_j) - \sum_{k=j+1}^n u_k \right| < \infty
\]

(in the notations of [34], $u_j = \Pi_{j}^n(0)$). Next, as in [34] Section 5, for every $A$ large enough, there is a sequence of intervals $I_j = \{a_j, a_j + 1, ..., b_j\}$ in the integers whose union covers $\mathbb{N}$ so that $I_j$ is to the left of $I_{j+1}$ and the variance of $S_{I_j} = \sum_{k \in I_j} Y_k$ is between $A$ to $2A$. Moreover, if we set $k_n = \max\{k : b_k \leq n\}$ then $k_n \asymp \sigma_n^2$ and

\[
\sup_n \left\| S_n - \sum_{j=1}^{k_n} S_{I_j} \right\|_{L^2} < \infty.
\]

Now, let us set $U_j = \sum_{k \in I_j} u_j$. Then by (5.4),

\[
\left| U_j - \text{Var}(S_{I_j}) \right| \leq C.
\]

Note that because of the properties of the blocks $S(I_j)$ and the exponential decay of correlation (see [36] Proposition 1.11 (2)) we have that

\[
\sigma_n^2 = \sigma_{k_n}^2 + O(1).
\]

Next, let us set $V_k = \text{Var} \left( \sum_{j=1}^{k} S_{I_j} \right)$. Then $\sigma_{k_n}^2 = V_{k_n}$. Moreover, by (5.4) we have

\[
\sup_k \left| V_{k+1} - \sum_{j=1}^{k+1} U_k \right| \leq C < \infty
\]

and so

\[
V_{k+1} - V_k \geq U_{k+1} - 2C \geq A - 2C.
\]

Thus when $A > 2C$ we see that $V_k$ is increasing. Since

\[
\sigma_n^2 = \sigma_{k_n}^2 + O(1) = V_{k_n} + O(1)
\]

and $k_n$ is increasing we see that we can take $a_n = \sigma_{k_n}^2 = V_{k_n}$.

Next, in [34] Proposition 24] we have shown that Assumption 3 holds true for all $m$. By applying Lemma 43 together with Theorem 9 and Corollary 19 we get the following result.

44. Theorem. The Berry-Esseen type theorems 3 and 9 hold true for every $m$. Moreover, the expectation estimates in Corollary 11 (i) hold true. Furthermore, for each $n$ and $m$ there is a coupling of $(Y_1, ..., Y_n)$ and a zero mean iid (finite) Gaussian sequence $(Z_1, ..., Z_n)$ such that $\text{Var}(Z_j) = b_j$ and

\[
\left\| S_n - \sum_{j=1}^{n} Z_j \right\|_{m-1} \leq C
\]

where $C$ is a constant which does not depend on $n$. 
45. **Remark.** When the variance of \( S_n \) grows linearly fast then Assumption 3 holds true for every \( m \) for any sufficiently fast mixing Markov chain (see [89]), and thus the above theorem holds. However, in this section we are interested in the situation when \( \sigma_n^2 \) can grow arbitrary slow.

### 5.2. Edgeworth expansions

We prove here the following:

46. **Theorem.** (i) Suppose that \( \|f_n\|_{L^\infty} = O(n^{-\beta}) \) for some \( \beta \in (0, 1/2) \). Then the Assumption 6 holds true when \( r := m - 2 < \frac{1}{1 - 2\beta} \), and therefore the non-uniform Edgeworth expansions of order \( r \) and the Edgeworth expansions in transport distances of order \( r \) (as in Theorems 3 and 10) and the expansions of the expectations as in Corollary 11 (ii) hold true (with such \( m \)). In particular, if \( \|f_n\|_{L^\infty} = O(n^{-1/2}) \) then \( S_n \) obeys expansions of all orders (in all of the above senses).

Moreover, the condition \( r = m - 2 < \frac{1}{1 - 2\beta} \) is optimal even for independent summands.

(ii) Suppose that \( X \) is a compact Riemannian manifold, and that \( f_n \) are uniformly Hölder continuous with some exponent \( \alpha \). Additionally, assume that all the measures \( m_j \) coincide with the volume measure and the transition densities \( p_j \) which are (uniformly) bounded and bounded away from the origin. Then the non-uniform Edgeworth expansions of order \( r = m - 2 \), the Edgeworth expansions in transport distances of order \( r \), and the expansions of the expectations as in Corollary 11 (ii) hold true for every \( m \) so that \( r = m - 2 < \frac{1 + \alpha}{1 - \alpha} \).

Moreover, the condition \( r = m - 2 < \frac{1 + \alpha}{1 - \alpha} \) is optimal even in the iid case.

**Proof.** Our goal is to verify condition (2.2) in both parts of Theorem 46. The beginning of the proof of both parts is identical.

First, similarly to [34, Section 5], there are intervals \( I_{j,n}, j \leq k_n \) in the positive integers whose union cover \( \{1, 2, ..., n\} \) and \( k_n \simeq \sigma_n^2 \). Moreover, for every \( p \geq 1 \) the \( L^p \)-norms of \( S_{I_{j,n}} \) are bounded by some constant \( C_p \), which does not depend on \( n \) and \( j \). Now, the \( k \)-th derivative of the characteristic function has the form

\[
\frac{d^m}{d t^m} \mathbb{E}[e^{itS_n}] = i^m \mathbb{E}[S_n e^{itS_n}] = i^m \mathbb{E} \left[ \sum_{j=1}^{k_n} S_{I_{j,n}} \right] e^{itS_n} = i^m \sum_{1 \leq \ell_1, ..., \ell_m \leq k_n} \mathbb{E}[X(\ell) e^{itS_n}]
\]

where for a given \( \ell = (\ell_1, ..., \ell_m) \), if \( \ell_{s_q}, q = 1, 2, ..., m(\ell) \) are the nonzero ones among \( \ell_s \) with \( s_j < s_{j+1} \) and \( m(\ell) \leq m \) then with \( \Xi_j = \Xi_{j,n} = S_{I_{j,n}} \):

\[
\Xi(\ell) = \prod_{q=1}^{m(\ell)} \Xi_{s_{q}}.
\]

Let \( J_i \) denote the gap between \( I_{s_i-1,n} \) and \( I_{s_i,n} \), where we set \( I_{0,n} = \{0\} \). With \( \bar{m} = m(\ell) \), let \( J_{\bar{m}+1} \) be the complement of the union of \( J_1, ..., J_{\bar{m}} \) and \( I_{s_1,n} \). For \( I = \{a, a+1, ..., a+b\} \) let

\[
R^I_{it} = R_{a,it} \circ R_{a+1,it} \circ \cdots \circ R_{a+n,it}
\]


where the operator $R_{j,it}$ is given by $R_{it,j}(x) = \mathbb{E}[e^{itJ_j(X_{j+1})}g(X_{j+1})|X_j = x]$. Using the Markov property we see that, for a fixed $\ell$, with $m = m(\ell)$ we have
\[
\mathbb{E}[\Xi(\ell)e^{itS_n}] = \mathbb{E}[R_{it}^{J_1} \circ L_{s_1} \circ R_{it}^{J_2} \circ L_{s_2} \circ \cdots \circ L_{s_{m-1}} \circ R_{it}^{J_m} \circ L_{s_m} \circ R_{it}^{J_{m+1}}1]
\]
where $L_s(g) = \mathbb{E}[g(\Xi_s)e^{it\Xi_s\Xi_s^\ell_s}|X_{a_s-1}]$ and $a_s$ is the left end point of $I_s$.

We claim next that the operator norms of the operators $L_{s_j}$ with respect the the appropriate essential supremum norms are bounded by some constant which does not depend on $\ell$ or $n$.

We first need the following. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Recall that the $\psi$-mixing coefficient of two sub $\sigma$-algebras $\mathcal{G}, \mathcal{H}$ of $\mathcal{F}$ is given by
\[
\psi(\mathcal{G}, \mathcal{H}) = \sup \left\{ \left| \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)\mathbb{P}(B)} - 1 \right| : A \in \mathcal{G}, B \in \mathcal{H}, \mathbb{P}(A)\mathbb{P}(B) > 0 \right\}.
\]

Then the $\psi$-mixing sequence $\psi(n)$ of the Markov chain $\{X_j\}$ is given by
\[
\psi(n) = \sup \{ \psi(\mathcal{F}_k, \mathcal{F}_{k+n,\infty}) : k \in \mathbb{N} \}
\]
where $\mathcal{F}_k$ is the $\sigma$-algebra generated by $\{X_1, \ldots, X_k\}$ and $\mathcal{F}_{k+n,\infty}$ is the $\sigma$-algebra generated by $\{X_j : j \geq k + n\}$. Then by [36, Proposition 1.22], $\psi(n) \leq Ce^{-cn}$ for some $c, C > 0$, and in particular $\psi(1) < \infty$. Now by [21, Ch. 4],
\[
\psi(\mathcal{G}, \mathcal{H}) = \sup \left\{ \|\mathbb{E}[h|\mathcal{G}] - \mathbb{E}[h]\|_{L^\infty} : h \in L^1(\Omega, \mathcal{H}, \mathbb{P}), \|h\|_{L^1} \leq 1 \right\}.
\]

Using the above representation of $\psi(\cdot, \cdot)$ and the definition of $\psi(1)$ we see that
\[
\left\| L_{s_j}(g) - \mathbb{E}[g(\Xi_{s_j})e^{it\Xi_j\Xi_j^\ell_{s_j}}]\right\|_{L^\infty} \leq \psi(1)\|g(\Xi_{s_j})e^{it\Xi_j\Xi_j^\ell_{s_j}}\|_{L^1} \leq \psi(1)\|\Xi_{s_j}\|_{L^\infty} \ell_{s_j}\|g(\Xi_{s_j})\|_{L^\infty}
\]
\[
\leq C_m'\|g(\Xi_{s_j})\|_{L^\infty}
\]
for some constant $C_m'$, where we have used that $\ell_{s_j} \leq m$ and all the $L^p$ norms of $\Xi_j$ are uniformly bounded in $j$. To complete the proof of the claim, note that
\[
\left| \mathbb{E}[g(\Xi_{s_j})e^{it\Xi_j\Xi_j^\ell_{s_j}}]\right| \leq \|g(\Xi_{s_j})\|_{L^\infty} \|\Xi_{s_j}\|_{L^\infty} \ell_{s_j} \leq C_m''\|g(\Xi_{s_j})\|_{L^\infty}.
\]

Using the above claim, we conclude that
\[
\left| \mathbb{E}[\Xi(\ell)e^{itS_n}]\right| \leq C_m''\prod_{s=1}^{m(\ell)+1} \|R_{it}^{J_s}\|_{L^\infty}.
\]

Let us now bound the norms $\|R_{it}^{J_s}\|_{L^\infty}$. First we note that these norms are always smaller than one. Second, since $m(\ell) \leq m$, the $L^2$-norm of the sum of $S_{I_{s_j}}, j \leq m(\ell)$ is bounded by some constant which does not depend on $\ell$ or $n$. Thus, the $L^2$ norm along the sum of the completing blocks $J_s$ is of order $\sigma_n$. Hence at one least of the these norms is of order $\sigma_n$ (there are at most $m_1 + 1$ blocks $J_{s_j}$). Let us denote this block by $J_u$. 
In the circumstances of part (i), in the beginning of Section 6 we showed that if $r < \frac{1}{1-2\beta}$ then when

$$|t| \leq C(\sigma(u))^{r-1}, \sigma(u) := \left\| \sum_{I_j \subset J_u} S_{I_j} \right\|_{L^2}$$

(for some constant $C$) we have

$$\| R_{it}^J \|_{L^\infty} \leq e^{-ct^2\sigma^2(u)}$$

where $c > 0$ is another constant. Note that in [34] we considered the supremum norms, but we can always replace the supremum norms on $X_j$ with the essential supremum norm with respect to the law of $X_j$. Recall now that, $\sigma(u) \geq c'\sigma_n$ for some positive constant $c'$, and we thus conclude that

$$\left| \mathbb{E}[\Xi(\xi) e^{itS_n}] \right| \leq C'' e^{-c''t^2\sigma_n^2}, \quad c'' > 0.$$ 

Starting from (5.6) and then combining this with the previous estimates we see that when $|t| \leq C''\sigma_n^{r-1}$ we have

$$\left| \frac{d^m}{dt^m} \mathbb{E}[e^{itS_n}] \right| \leq A'_m \sigma_n^{2m} e^{-c''t^2\sigma_n^2}$$

for some constant $A'_m$, which is enough to verify (2.2). Thus, as noted in Remark 7 Assumption 6 is satisfied when $r = m - 2 < \frac{1}{1-2\beta}$.

Next, in the circumstances of part (ii), in the proof of Proposition 34 we showed that when $|t|$ is large enough then for every $n$,

$$\| R_{it}^J \|_{L^\infty} \leq Ce^{-c|t|^{1-1/\alpha}\sigma^2(u)} \leq Ce^{-c'|t|^{1-1/\alpha}\sigma_n^2}$$

which, starting again from (5.6), yields that

$$\left| \frac{d^m}{dt^m} \mathbb{E}[e^{itS_n}] \right| \leq R_m \sigma_n^{2m} Ce^{-c'|t|^{1-1/\alpha}\sigma_n^2}$$

for some constant $R_m$, which is also enough for Assumption 6 to hold true when $r = m - 2 < \frac{1}{1-2\beta}$ (using again (2.2)).

Finally, the optimality of the conditions was already discussed in the context of uniform Edgeworth expansions in [34].

5.3. Random (partially expanding or hyperbolic) dynamical systems. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\theta : \Omega \rightarrow \Omega$ be an ergodic invertible map. Let $(\mathcal{X}, d)$ be a metric space and let $E_\omega, \omega \in \Omega$ be a measurable family of compact subsets (see [58], Ch. 6). Set $\mathcal{E} = \{(\omega, x) : \omega \in \Omega, x \in E_\omega\}$ and let $T : \mathcal{E} \rightarrow \mathcal{E}$ be a measurable map of the form

$$T(\omega, x) = (\theta \omega, T_\omega x)$$

\(11\)The upper bounds on the characteristic function obtained there relied on appropriate estimates on the norms of the operators.
where $T_\omega : \mathcal{E}_\omega \to \mathcal{E}_{\theta \omega}$ is a random family of maps ($T$ is the so-called skew product induced by $\{T_\omega\}$). Let $\mu$ be a $T$-invariant probability measure on $\mathcal{E}$, and let us represent it in the form

$$\mu = \int \mu_\omega d\mathbb{P}(\omega)$$

where for $\mathbb{P}$-almost every $\omega$ the probability measure $\mu_\omega$ on $\mathcal{E}_\omega$ satisfies $(T_\omega)^\ast \mu_\omega = \mu_{\theta \omega}$.

Next, let $\mathcal{L}_\omega$ denote the dual of $T_\omega$ with respect to the measures $\mu_\omega$ and $\mu_{\theta \omega}$, namely the unique operator so that

$$\int g \cdot (f \circ T_\omega) d\mu_\omega = \int f \cdot (\mathcal{L}_\omega g) d\mu_{\theta \omega}$$

for all bounded functions $g$ and $f$ on the appropriate domains. Let $(B_\omega, \| \cdot \|_\omega)$ be a norm on functions on $\mathcal{E}_\omega$ and let $f : \mathcal{E} \to \mathbb{R}$ be a measurable function so that $\|f\| := \text{ess-sup}\|f(\omega, \cdot)\|_\omega < \infty$. In what follows we will introduce assumptions on $\mathcal{L}_\omega$ and their complex perturbations which will guarantee that Assumption 47 holds true for

$$S_n^\omega = \sum_{j=0}^{n-1} f_{\theta^j \omega} \circ T_{\theta^j -1 \omega} \circ \cdots \circ T_\omega, \ f_\omega(\cdot) = f(\omega, \cdot)$$

for $\mathbb{P}$-a.e. $\omega$, when considered as a random variables on the space $(\mathcal{E}_\omega, \mu_\omega)$.

For a complex parameter $z$, let us consider the operator $\mathcal{L}_\omega^z$ given by $\mathcal{L}_\omega^z(g) = \mathcal{L}_\omega(ge^{zf_\omega})$. For it to be well defined, we assume that $\mathcal{L}_\omega$ is a continuous operator between $B_\omega$ and $B_{\theta \omega}$ and that the map $z \to e^{zf_\omega} \in B_{\omega}$ is analytic in $z$, uniformly in $\omega$.

47. **Assumption.** There is a constant $r_0 > 0$ so that $(\mathbb{P}\text{-a.s.})$ for every complex $z$ with $|z| < r_0$ there is a triplet $\lambda_\omega(z) \in \mathbb{C} \setminus \{0\}$, $h^{(z)}_\omega \in B_\omega$ and $\nu^{(z)}_\omega \in B^*_\omega$ which is measurable in $\omega$, analytic in $z$ and:

1. $\lambda_\omega(0) = 1$, $h^{(0)}_\omega = 1$, $\nu^{(0)}_\omega = \mu_\omega$, $\nu^{(z)}_\omega(h^{(z)}_\omega) = 1$;
2. with

$$\mathcal{L}_\omega^{z,n} = \mathcal{L}_{\theta^{n-1} \omega} \circ \cdots \circ \mathcal{L}_\omega^z$$

and

$$\lambda_{\omega, n}(z) = \lambda_{\theta^{n-1} \omega}(z) \cdots \lambda_\omega(z)$$

there are $C > 0$ and $\delta \in (0, 1)$ so that

$$\left\| (\lambda_{\omega, n}(z))^{-1} \mathcal{L}_\omega^{z,n} - \nu^{(z)}_\omega \otimes h^{(z)}_{\theta^{n} \omega} \right\|_{\theta^n \omega} \leq C\delta^n$$

where the operator $\nu \otimes h$ is given by $g \to \nu(g) \cdot h$.

48. **Example.** Let us list a few types of random maps $T_\omega$ and measures $\mu_\omega$ for which Assumption 47 holds true.

1. The random expanding maps considered in [38, Ch. 6], where $\| \cdot \|_\omega$ is a H"older norm of a function on $\mathcal{E}_\omega$ with respect to some exponent $\alpha \in (0, 1]$ and $\mu_\omega$ is a random Gibbs measure;
2. The random expanding maps considered in [38], where $\| \cdot \|_\omega$ is the bounded-variation norm and $\mu_\omega$ is the unique random absolutely continuous equivariant measure.
(3) The hyperbolic maps considered in [39] (see also [40]), with $\| \cdot \|_\omega$ being the appropriate strong norm and $\mu_\omega$ being the unique random fiberwise absolutely continuous equivariant measure.

(4) The random partially expanding or hyperbolic maps considered in [61], with $\| \cdot \|_\omega$ being the appropriate “weighted” Hölder norm and $\mu_\omega$ being a sampling measure.

(5) The uniformly random version (as in [63]) of the random partially expanding maps considered in [92] with $\| \cdot \|_\omega$ being a Hölder norm with respect to some exponent $\alpha \in (0, 1]$ and $\mu_\omega$ being a random Gibbs measure corresponding to a potential with a sufficiently small oscillation (see [63]).

49. Remark. We note that in [38, 39] Assumption 47 appears as

$$\| \mathcal{L}^n_{\omega} - \lambda_{\omega,n}(z) \left( \nu_\omega(z) \otimes h_{\theta_n,\omega}(z) \right) \|_{\mu_\omega} \leq C \delta^n$$

but since $\lambda_\omega(z) = 1 + O(z)$ upon decreasing $r_0$ we get that $(1 - \varepsilon)^n \leq |\lambda_{\omega,n}(z)| \leq (1 + \varepsilon)^n$ for an arbitrary small $\varepsilon$, and hence the above two forms are equivalent.

50. Proposition. Under assumption 47 we we have tha following.

(i) There is a nonnegative number $\sigma$ so that $\mathbb{P}$-a.s.

$$\sigma^2 = \lim_{n \to \infty} n^{-1} \text{Var}_{\omega}(S_n^\omega).$$

Moreover, $\sigma^2 = 0$ iff $f(\omega, x) - \int f(\omega, y) d\mu_\omega(y) = r(\omega, x) - r(\theta \omega, T_\omega x)$ for some function $r$ so that $\int |r(\omega, y)|^2 d\mu_\omega(y) d\mathbb{P}(\omega) < \infty$.

(ii) Suppose that $\sigma^2 > 0$. Then, under Assumption 47, for $\mathbb{P}$-a.e. $\omega$ we have that $S_n = S_n^\omega - \mu_\omega(S_n^\omega)$ verifies Assumption 3 with every $m$. Thus, all the results stated in Theorems 5, 8 and Corollary 11 (i) hold true for all $m$.

Proof. We have

$$\mathbb{E}[e^{zS_n^\omega}] = \mu_\theta \omega(\mathcal{L}^n_{\omega} 1).$$

Now, like in the stationary case, we have

$$\ln \mathbb{E}[e^{zS_n^\omega}] = \Pi_{\omega,n}(z) + H_{\omega,n}(z) + O(|z| \delta^n).$$

where

$$\Pi_{\omega,n}(z) = \sum_{j=0}^{n-1} \ln \lambda_{\theta^j \omega}(z)$$

and $H_{\omega,n}(z)$ is bounded by some constant $C$ and it vanishes at $z = 0$. Now, by differentiating both sides of (5.7), using the Cauchy integral formula, and plugging in $z = 0$ we see that

$$\sigma^2_{\omega,n} = \text{Var}_{\mu_\omega}(S_n^\omega) = \Pi''_{\omega,n}(0) + O(1) = \sum_{j=0}^{n-1} \Pi''_{\theta^j \omega}(0) + O(1).$$

Next, by ergodicity of $(\Omega, \mathcal{F}, \mathbb{P}, \theta)$ we conclude that, $\mathbb{P}$-a.s. we have

$$\lim_{n \to \infty} \frac{1}{n} \sigma^2_{\omega,n} = \int_{\Omega} \Pi''_{\omega}(0) d\mathbb{P}(\omega) := \sigma^2.$$

The proof of the characterization of positivity of $\sigma^2$ proceeds similarly to [72, Section 3].

---

12 ln $\lambda_\omega(z)$ is an analytic branch which vanishes at the origin and is uniformly bounded around the origin.
Finally, in order to show that Assumption 3 is satisfied for all \( m \) we differentiate both sides of (5.7) \( j \) times and then use the Cauchy integral formula to conclude that for all \( j \geq 3 \) there are constants \( r_j, C_j > 0 \) so that (P.a.s.) on the domain \( \{ |z| \leq r_j \} \) we have

\[
\tilde{A}_n^{(j)}(z) \leq C_j n, \quad \tilde{A}_n(z) = \tilde{A}_{\omega,n}(z) = \ln \mathbb{E}[e^{zS_n}]
\]

where we have used that \( \ln \lambda_{\theta_j\omega}(z) \) and \( H_{\omega,n}(z) \) and all their derivatives are uniformly bounded in \( z, \omega, j \) and \( n \) (when \( |z| \) is small enough). To complete the proof that Assumption 3 is satisfied we use that \( \sigma_{\omega,n}^2 \) grows linearly fast in \( n \).

\[ \square \]

5.3.1. Coupling: on the verification of the conditions of Corollary 19

51. Lemma. Suppose that \( \sigma > 0 \). For \( \mathbb{P} \)-a.a. \( \omega \) there exists an increasing sequence \((B_n(\omega))_{n=1}^{\infty}\) so that

\[
\sigma_{\omega,n}^2 = \|S_n^\omega - \mu_\omega(S_n^\omega)\|_{L^2(\mu_\omega)}^2 = B_n(\omega) + O(1).
\]

As a consequence, the couplings with the properties described in Corollary 19 always exist (for all \( m \)) for \( S_n = S_n^\omega - \mu_\omega(S_n^\omega) \).

Proof. As in the proof of Proposition 50, by differentiating twice both sides of (5.7) and using the Cauchy integral formula we see that \( \sigma_{\omega,n}^2 = \|S_n^\omega - \mu_\omega(S_n^\omega)\|_{L^2(\mu_\omega)}^2 \) has the form

\[
\sigma_{\omega,n}^2 = \sum_{j=0}^{n-1} u_{\theta_j \omega} + O(1), \quad u_{\omega} = \Pi_\omega(0)
\]

with \( \text{ess-sup}|u_\omega| < \infty \). Moreover, \( \sigma_{\omega,n}^2 \leq \sigma^2 n \) for some \( \sigma > 0 \). Thus, for \( \mathbb{P} \)-a.a. \( \omega \) there are intervals in the integers \( I_{\omega,1}, I_{\omega,2}, \ldots \) whose union covers \( \mathbb{N}_+ = \{0, 1, 2, \ldots\} \) and \( I_{\omega,j} \) is to the left of \( I_{\omega,j+1} \) and with \( U_{\omega,k} = \sum_{j \in I_{\omega,k}} u_{\theta_j \omega} \) we have \( A \leq U_{\omega,k} \leq 2A \) for some constant \( A > 1 \). Moreover, for every interval \( J \) in the integers with the same left end point as \( I_{\omega,j} \) so that \( J \subset I_{\omega,j} \) and \( J \neq I_{\omega,j} \) we have \( \sum_{s \in I_{\omega,j}} u_{\theta_s \omega} \leq A \).

Let \( k_n(\omega) = \max\{k : I_{\omega,k} \subset \{0, 1, \ldots, n - 1\} \} \). Then \((k_n(\omega))_{n=1}^{\infty}\) is an increasing sequence and

\[
\sigma_{\omega,n}^2 = \sum_{j=1}^{k_n(\omega)} U_{\omega,j} + O(1)
\]

where the \( O(1) \) term is actually bounded in absolute value by \( A \). Now we can take \( B_n = B_n(\omega) = \sum_{j=1}^{k_n(\omega)} U_{\omega,j} \).

\[ \square \]

5.3.2. On the verification of Assumption 6 with \( m = 3 \). We first need the following assumption.

52. Assumption. For every compact subset \( K \) of \( \mathbb{R} \setminus \{0\} \) we have

\[
(5.8) \quad \sup_{t \in K} \sup_{n \geq 1} \|L^{t,n}_{\omega}\|_{\theta^n_{\omega}} \leq C(K)
\]

where \( C(K) \) depends only on \( K \) and

\[
(5.9) \quad \sup_{t \in K} \|L^{t,n}_{\omega}\|_{\theta^n_{\omega}} \leq C_\omega(K)e^{-c_K n}
\]

where \( C_\omega(K) > 0 \) and \( c_K > 0 \) depend on \( K \) and \( \omega \rightarrow C_\omega(K) \) is measurable.
53. **Proposition.** Under the additional Assumption 52 for \( \mathbb{P} \)-a.e. \( \omega \) for every \( m \) and all \( b > a > 0 \) we have
\[
\int_{a \sigma_n \leq |t| \leq b \sigma_n} \left| \frac{f_{n}^{(m)}(t)}{t} \right| \, dt = O(\delta^n)
\]
where \( f_n(t) = \mathbb{E}_{\mu_\omega}[e^{itS_n^\omega} - S_n^\omega] \) and \( \delta \in (0, 1) \) depends on \( a, b \) and \( \omega \). Thus, under Assumptions 47 and 52 all the results stated in Theorems 10, 18 and Corollary 11 (ii) hold true with \( m = 3 \).

**Proof.** Let us take \( K = [-b, -a] \cup [a, b] \). Let \( A \subset \Omega \) be a measurable set with positive probability so that on \( A \) we have \( C_K(\omega) \leq C_0 \) for some constant \( C_0 \). Since \( \theta \) is ergodic it follows from Kac’s formula that for \( \mathbb{P} \)-a.e. \( \omega \) there is an infinite sequence of natural numbers \( n_1(\omega) < n_2(\omega) < \ldots \) so that \( \theta^{n_i(\omega)} \omega \in A \) and
\[
(5.10) \quad \lim_{k \to \infty} \frac{n_k(\omega)}{k} = \frac{1}{\mathbb{P}(A)}.
\]

Now, after replacing \( f(\omega, x) \) with \( f(\omega, y) \) (i.e. assuming \( \mu_\omega(f_\omega) = 0 \)) we have
\[
\frac{d^k}{dt^k} \mathbb{E}[e^{itS_n^\omega}] = \frac{d^k}{dt^k} \mathbb{E}[(S_n^\omega)^k e^{itS_n^\omega}] = i^k \sum_{1 \leq \ell_1 \leq \ell_2 \leq \ldots \leq \ell_k \leq n} \mathbb{E}[X_{\ell_1,\omega}^{k}(\omega) e^{itS_n^\omega}]
\]
where
\[
X_{\ell,\omega} = \prod_{s=1}^{k} f_{\theta^{s} \omega} \circ T_{\omega}^{\ell_s}.
\]

Let us fix some \( \ell = (\ell_1, \ldots, \ell_k) \). Let \( j_1 < j_2 < \ldots < j_r, r \leq k \) be the distinct indexes among \( \ell_1, \ldots, \ell_k \), and suppose that \( j_s \) appears \( a_s \) times in \( \ell \). Let us define an operator by \( L_{j_s, r_s, it} \) by
\[
L_{j_s, r_s, it}(g) = \mathcal{L}_{\theta^{s} \omega} (g f_{\theta^{s} \omega} e^{itf_{\theta^{s} \omega}}).
\]

Then the norms of the operators \( L_{j_s, r_s, it} \) are uniformly bounded by some constant \( R(K) \geq 1 \) when \( t \in K \). Notice now that
\[
\mathbb{E}[X_{\ell,\omega}^{k}(\omega) e^{itS_n^\omega}] = \mu_{\theta^n \omega}(L_{j_r, j_{r-1}, it} \circ L_{j_{r-1}, j_{r-2}, it} \circ \cdots \circ L_{j_2, j_1, it} \circ L_{j_1, it} \circ L_{j_1, it}^{1}).
\]

Now, since \( \sum_{j=1}^{r}(j_r+1-j_{r-1}) = n - r \), where \( j_r+1 = n \) and \( j_0 = -1 \), at least one of the iterates \( L_{j_r, j_{r-1}, it} \) is a composition of at least \( \lfloor n/r \rfloor - 1 \) operators. Now, by (5.10) for every interval in the integers \( I \subset \{1, 2, \ldots, n\} \), whose left end point is large enough and whose length is at least \( \lfloor n/3r \rfloor - 3 \), contains at least one point \( a \) so that \( \theta^a \omega \in A \). By taking \( I \) to be the middle third of the interval \( I = \{j_s+1, \ldots, j_{s+1}\} \) whose length is at least \( \lfloor n/r \rfloor - 1 \), using (5.3) with \( \theta^a \omega \) instead of \( \omega \) and \( \lfloor n/3r \rfloor - 2 \) instead of \( n \) and using (5.8) to bound the norms of the other “big” blocks we see that \( \mathbb{P} \)-a.s. for every \( n \) large enough we have
\[
\left\| L_{j_r, j_{r-1}, it} \circ L_{j_{r-1}, j_{r-2}, it} \circ \cdots \circ L_{j_1, it} \circ L_{j_1, it}^{1} \right\|_{\theta^n \omega} \leq (R(K))^k (1 + C) e^{-\frac{1}{C}cK^n}
\]
for some \( C > 0 \), where we have used that \( r \leq k \). Hence, for all \( k \) and positive \( b > a \) we have
\[
\int_{a \leq |t| \leq b} \left| \psi_{\omega,n}^{(k)}(t) \right| \, dt = O(\delta^n)
\]
for some $\delta \in (0,1)$, where $\psi_{\omega,n}(t) = \mathbb{E}[e^{it(S_n^\omega - \bar{\mu}(S_n^\omega))}]$. As noted in Remark 7, this is enough for Assumption 6 to hold true with $m = 3$.

□

54. Example. We refer to [58, Lemma] and [58, Ch.6] for sufficient conditions for Assumption 52.

5.3.3. On the verification of Assumption 6 with $m > 3$. In order to control the integral of $|f_n(m)(t)|/|t|$ over domains of the form $\{|t| \in [c_\sigma, B\sigma_n^{-m-2}]\}$ for large $c$‘s we need to introduce a few additional assumptions. Let $\| \cdot \|_{1,\omega}$ be (possibly) another norm on the space of functions on $\mathcal{E}_\omega$ so that $\| \cdot \|_{1,\omega} \geq \| \cdot \|_{\infty}$. For every fixed $t \neq 0$ let us take a norm $\| \cdot \|_{\omega,(t)}$ so that

\[ c_1 \| g \|_{\omega,(t)} \leq \| g \|_{1,\omega} \leq c(1 + |t|) \| g \|_{\omega,(t)} \]

and that

\[ \| \mathcal{L}_{\omega}^{t,n} \|_{\theta^n,\omega,(t)} \leq C \]

for some constant $C$ which does not depend on $\omega, n$ and $t$.

55. Proposition. Let Assumptions 47 and 52 hold, and suppose that $\sigma > 0$. Assume also that there are positive random variables $\rho_\omega, \gamma(\omega)$ and $b_\omega$ so that, $\mathbb{P}$-a.s. for every $t \in \mathbb{R}$ such that $|t| \geq b_\omega$ we have

\[ \| \mathcal{L}_{\omega}^{t,n_\omega(t)} \|_{\theta^n,\omega,(t)} \leq e^{-n_\omega(t)\gamma(\omega)} \]

where $n_\omega(t) = [\rho_\omega \ln |t|]$. Then for $\mathbb{P}$-a.e. $\omega$ we have that $S_n^\omega$ satisfies the integrability conditions in Assumption 6 for every $m$. Thus, under Assumptions 47, 52 and the above condition all the Edgeworth expansions stated in Theorems 8, 10 and Corollary 11 (ii) hold true for every $m$.

Proof. First, by Proposition 52 it is enough to show that the integral of $|f_n(m)(t)|/|t|$ over $\{|t| \in [b_1\sigma, B\sigma_n^{-m-2}]\}$ is $o(\sigma_n^2)$ for all $b_1 > 0$ large enough and $B > 0$. Next, by Remark 7 it is enough to verify (2.2) with $c$ large enough.

Let $A$ be a measurable subset of $\Omega$ with positive probability so that $b_\omega \leq b_0$ and $\rho_1 \leq \rho_\omega \leq \rho_2$ and $\gamma(\omega) \geq \gamma$ for all $\omega \in A$, where $b_0, \rho_1, \rho_2$ and $\gamma$ are positive constants. Since $\theta$ is ergodic we know from Kac’s formula that for $\mathbb{P}$-a.e. $\omega$ there is an infinite sequence of natural numbers $n_1(\omega) < n_2(\omega) < ...$ so that $\theta^{n_i(\omega)} \omega \in A$ and

\[ \lim_{k \to \infty} \frac{n_k(\omega)}{k} = \frac{1}{\mathbb{P}(A)}. \]

Thus, with

\[ m_n(\omega) = \max\{m : n_m(\omega) \leq n\} \]

we have

\[ \lim_{n \to \infty} \frac{m_n(\omega)}{n} = \mathbb{P}(A). \]

Now, as in the proof of Proposition 53 we have

\[ \frac{d^k}{dt^k} \mathbb{E}[e^{itS_n^\omega}] = i^k \mathbb{E}[(S_n^\omega)^k e^{itS_n^\omega}] = i^k \sum_{1 \leq \ell_1 \leq \ell_2 \leq ... \leq \ell_k \leq n} \mathbb{E}[X_{\ell,\omega}(\omega)e^{itS_n^\omega}]. \]
where

\[ X_{\ell,\omega} = \prod_{s=1}^{k} f_{g^{s+1}} \circ T_{\omega}^{s}. \]

As in the proof of Proposition \[ \text{53} \] let us fix some \( \ell = (\ell_1, ..., \ell_k) \) and assume that \( j_1 < j_2 < ... < j_r \) are the distinct indexes among \( \ell_1, ..., \ell_k \). Let \( a_s \) be the number of times that \( j_s \) appears in \( \ell \). Then we have

\[ (5.15) \quad \mathbb{E}[X_{\ell,\omega}(\omega)e^{itS_n}] = \mu_{\theta^{c_s}(\omega)}(L_{\theta^{c_j+1}} \circ L_{\theta^{c_j-1}} \circ \cdots \circ L_{\theta^{c_{j_r+1}}(\omega)}} \circ L_{\theta^{c_{j_r-1}}(\omega)}} \circ \cdots \circ L_{\theta^{c_{j_1+1}}(\omega)}} \circ L_{\theta^{c_{j_1-1}}(\omega)}}(\omega) \).

Let us suppose now that \( |t| \leq c_0 n^{(m-3)/2} \) for some \( c_0 > 0 \). In order to bound the norm of the above composition, let us note that there is an \( \varepsilon_0 > 0 \) so that for every \( n \) large enough there are at least \( \varepsilon_0 n \) indexes \( k \) between 0 and \( n - 1 \) so that \( \theta^k \omega \in A \). Let us take some \( D > 0 \), and let \( k_1, k_2, ..., k_d, d_n \approx \frac{n}{\ln n} \) be indexes so that \( \theta^{k_i} \omega \in A \) and \( k_{j+1} - k_j \geq D \ln n \). By omitting the \( k_j \)'s which belong to \( \mathcal{J}_s = [j_s - D \ln n, j_s] \) for some \( s \), we can always assume that all \( k_j \)'s are not in the union of \( \mathcal{J}_s \). Now, if \( D \) is large enough then by using that

\[ \left\| L_{\theta^{k_j-1}\omega}(t) \right\|_{\theta^{k_j-1}\omega(t)} \leq e^{-\gamma |t|} \]

and \( (5.12) \) to bound the other big blocks, taking into account that \( \rho_1 \leq \rho_{\theta^{k_j}\omega} \leq \rho_2 \) we conclude that the \( \| \cdot \|_{\theta^{k_j-1}\omega(t)} \) norm of the product of the operators does not exceed

\[ C_0 R^{n/\ln n} e^{-c_1 \ln |t| n/\ln n} \]

where \( C_0, R \) and \( c_1 \) are constants which do not depend neither \( t \) nor \( n \), and we have assumed that \( |t| \geq b_0 \). Now, using also \( (5.11) \) we conclude that if \( b_1 \) is large enough and \( b_1 \leq |t| \leq c_0 e^{m-3} \) then

\[ \left| \frac{d^k}{dt^k} \mathbb{E}[e^{itS_n}] \right| \leq C(1 + |t|) n^k e^{-c_2 n/\ln n} \]

for some constant \( c_2 > 0 \). Hence if \( c \geq b_1 \) then \( (2.2) \) is valid for all \( m \) and an arbitrary large \( B \).

56. Example.

- The conditions of Proposition \[ \text{55} \] hold true for the random expanding interval maps considered in \[ \text{59} \] Section 5.1.3, with \( \| \cdot \|_{\omega} \) being the BV norm (note that the maps considered in \[ \text{59} \] Section 5.1.3 can have discontinuities).

- Let \( T_1, T_2, ..., T_d \) be smooth expanding maps on some compact Riemannian manifold with the same properties as in \[ \text{31} \] Section 3.4 and let \( f_1, ..., f_d \) be smooth functions so that \( f_1 \) is not infinitesimally integrable (see also \[ \text{48} \] Section 6.5). Now let us suppose that \( (\Omega, \mathcal{F}, \mathbb{P}, \theta) \) is the shift system generated by some ergodic sequence of random variable \( \xi_j \) taking values on \( \{1, 2, ..., d\} \) so that \( \mathbb{P}(\xi_1 = \xi_2 = ... = \xi_n = 1) > 0 \) for some \( n_0 \) that will be described below, and assume that for \( \omega = (\xi_j) \) we have \( T_{\omega} = T_{\xi_0} \) and \( f_\omega = f_{\xi_0} \) (e.g. we can take a stationary finite state Markov chain with positive transition probabilities).
Next, by [31, Lemma 3.18] the, so-called Dolgopyat type inequality ([48, (6.9)]) holds true for the transfer operators $L_{1,tt}$ and the norms $\| \cdot \| = \| \cdot \|_{C^1}$ and

$$\|g\|_{(t)} = \max \left( \sup |g|, \frac{\sup |Dg|}{C(1 + |t|)} \right),$$

where the complex transfer $L_{j,z}$ operators are given by $g \to L_{T_j}(e^{zJ}g)$ and $C$ is a sufficiently large constant. Namely, $L_{1,tt}$ obeys a nonrandom version of (5.13) with these norms. Now, note that in view of the Lasota-York inequality (see [58, Chapter 6] for the random case) if $C$ is large enough then

$$\|L_{1,tt}\|_{(t)} \leq 1 \text{ and } \|L_{j,tt}\|_{(t)} \leq 1.$$ 

Now, as argued in [48, Section 5.4], there are constants $c_0, K_0 > 0$ ($K_0$ can be taken to be arbitrary small) and $r_0 \in (0, 1)$ so that if $|t| \geq K_0$ then

$$\|L_{1,tt}\|_{(t)} \leq c_0 r_0^n.$$

Next, let us take $n_0$ large enough so that $c_0 r_0^{n_0} \leq \frac{1}{2}$. Let $X_k = X_k(\omega) = (\xi_{k0}, \ldots, \xi_{kn_0+n_0-1})$ and set

$$Q_n = Q_n(\omega) = \sum_{k=0}^{[n/n_0]-1} \mathbb{I}(X_k = (1, 1, \ldots, 1)).$$

Then by the mean ergodic theorem $Q_n/n \to \mathbb{P}(X_1 = (1, 1, \ldots, 1)) = p_0 > 0$ (almost surely). Thus, for almost every realization of the random dynamical system we get that the operators on the right hand side of (5.14) are composed of $k$ blocks with $\| \cdot \|_{(t)}$-norms of order $O(1)$, at least $[c_1 n]$ blocks $c_1 = \frac{1}{4} a_0 > 0$ with $\| \cdot \|_{(t)}$-norms less or equal to $\frac{1}{2}$ and the rest of the blocks have $\| \cdot \|_{(t)}$-norms less or equal to 1. Starting from (5.14), using (5.15) and the above block decomposition, we see that $\mathbb{P}$-a.s. for all $n$ large enough and $|t| \geq K_0$ we have

$$\left| \frac{d^k}{dt^k} \mathbb{E}[e^{itS_n}] \right| \leq C(1 + |t|) n^k \left( \frac{1}{2} \right)^{c_1 n}.$$

Hence Assumption (i) is in force for all $m$.

Finally, let us note that the same argument works for a random dynamical systems $T_\omega$ and a random function $f_\omega$ so that $\mathbb{P}(A) > 0$, $A = \{ \omega : T_\theta \omega = T_1, f_\theta \omega = f_1, \forall 0 \leq j \leq n_0 \}$ since then we can consider the number of visiting times to $A$.

5.4. Sequential dynamical systems in a neighborhood of a single system. Let $(\mathcal{X}_j, \mathcal{B}_j, \mu_j)$ be a sequence of probability space, and let $T_j : \mathcal{X}_j \to \mathcal{X}_{j+1}$ be a sequence of measurable maps. Let $\mathcal{L}_j$ an operator which is defined by the duality relation:

$$\int_{\mathcal{X}_j} g \cdot (f \circ T_j) d\mu_j = \int (\mathcal{L}_j g) \cdot f d\mu_{j+1}$$

for all bounded functions $g : \mathcal{X}_j \to \mathbb{R}$ and $f : \mathcal{X}_{j+1} \to \mathbb{R}$. Let $B_j$ be a Banach space of measurable functions on $\mathcal{X}_j$, equipped with a norm $\| \cdot \|_j$ so that $sup |g| \leq C \|g\|_j$ for some constant $C$ which does not depend on $j$. We assume here that there are constants $A$ and $\delta \in (0, 1)$ and strictly positive functions $h_j \in B_j$ so that for every $g \in B_j$ and all $n \geq 1$ we have

$$\|\mathcal{L}_{j+n-1} \circ \cdots \circ \mathcal{L}_{j+1} \circ \mathcal{L}_j g - \mu_j(g) h_{j+n} \|_{j+n} \leq A \|g\|_j \delta^n.$$
Let take now a sequence of real-valued functions \( f_j \in B_j \) so that \( \|f\| := \sup_j \|f\|_j < \infty \), \( \mu_j(f_j) = 0 \) and define
\[
S_n = \sum_{j=0}^{n-1} f_j \circ T_{j-1} \circ \cdots \circ T_1 \circ T_0(x_0)
\]
where \( x_0 \) is distributed according to \( \mu_0 \). Let us define \( \mathcal{L}_j(z)(g) = \mathcal{L}_j(e^{zf_j}) \), where \( z \in \mathbb{C} \). We further assume here that the map \( z \to \mathcal{L}_j(z) \) is analytic around the complex origin (with values in the space of bounded linear operators between \( B_j \) and \( B_{j+1} \)), uniformly in \( j \).

57. Example.

(i) These conditions hold true in the setup of \([60]\), where the maps \( T_j \) are locally expanding, the space \( B_j \) is the space of Hölder continuous functions and \( \mu_j \) are measures so that \((T_j)_* \mu_j = \mu_{j+1} \). Note that when the maps are absolutely continuous with respect to some reference measure (e.g. a volume measure) then \( \mu_j \) is equivalent to that measure.

(ii) These conditions hold true in the setup of \([29]\), where \( \mathcal{X}_j = \mathcal{X} \) and \( \mu_j = \mu \) coincide with the same manifold and (normalized) volume measure, respectively, each \( T_j \) is a locally expanding map and \( B_j = B \) is the space of functions with bounded variation.

Next, by applying a sequential perturbation theorem based on \([51]\) (or, in the setup of \([60]\) using complex cones contractions) we see that there is a constant \( r_0 \) so that for every complex parameter \( z \) with \( |z| \leq r_0 \) there are analytic in \( z \) triplets \((\lambda_j(z), h_j(z), \nu_j(z))\) consisting of a complex non-zero random variable \( \lambda_j(z) \), a complex-valued function \( h_j(z) \in B_j \) and a complex continuous linear functional \( \nu_j(z) \in B^*_j \) so that \( \lambda_j(0) = 1, h_j(0) = h_j \) and \( \nu_j(0) = \mu_j, \nu_j(z)(h_j(z)) = \nu_j(z)(1) = 1 \). Moreover, there are constants \( A_1 > 0 \) and \( \delta_1 \in (0, 1) \) so that for all \( j, n \) and a function \( g \in B_j \) we have
\[
\left\| \mathcal{L}_{j+n-1} \circ \cdots \circ \mathcal{L}_{j+1} \circ \mathcal{L}_j g - \nu_j(z)(g)h_j(z) \right\| \leq A_1 \|g\| \delta_1^n.
\]
The following result follows exactly like the corresponding result about random dynamical systems in the previous section.

58. Theorem. Suppose that \( \text{Var}_{\mu_0}(S_n) \geq c_0 n \) for some \( c_0 > 0 \) and all \( n \) large enough. Then Assumption \( \mathcal{A} \) is in force, and so Theorems \( \mathcal{A} \) \( \mathcal{B} \) and Corollary \( \mathcal{C} \) (with some \( B_n \)), as well as Corollary \( \mathcal{D}(i) \) hold true.

5.4.1. Linearly growing variances: the perturbative approach. Let us suppose that \( \mathcal{X}_j = \mathcal{X} \) and \( B_j = B \) coincide with a single space \( \mathcal{X} \) and a Banach space, respectively. Let \( T: \mathcal{X} \to \mathcal{X} \) be a map so that all the previous properties are satisfied with the sequence \( T_j = T \), and let \( f \in B \). Then the corresponding triplet \( \lambda_{j,T}(z) = \lambda(z; T), h_{j,T}(z) = h_T(z) \) and \( \nu_{j,T}(z) = \nu_T(z) \) does not depend on \( j \). Next, let us assume that \( f \) is not an \( L^2(\mu_T) \) coboundary with respect to \( T \), where \( \mu_T = \nu_T(0) \). Then, with \( S_n f = \sum_{j=0}^{n-1} f \circ T^j \) we have
\[
\sigma^2 = \lim_{n \to \infty} \frac{1}{n} \text{Var}_{\mu_T}(S_n f) > 0.
\]
Let us denote the dual operator corresponding to \( T \) by \( \mathcal{L}_T \).
59. Proposition. Set  
\[ \varepsilon_0 = \sup_j \max (\|L_j - L_T\|, \|f - f_j\|). \]

If \( \varepsilon_0 \) is small enough then for all \( n \) large enough we have \( \sigma_n^2 \geq \frac{1}{2} \sigma^2 \).

Proof. First, as was shown in [60] (or using a sequential inverse function theorem in the more general setup), we have that
\[ \sup_{j, |z| \leq r_1} \max \left( |\lambda_j(z) - \lambda_T(z)|, \|h_j^{(z)} - h_T^{(z)}\|, \|\nu_j^{(z)} - \nu_T^{(z)}\| \right) \leq \delta(\varepsilon_0) \]
where \( \delta(\varepsilon_0) \to 0 \) as \( \varepsilon_0 \to 0 \), and \( r_1 \) is a constant. Next, as was shown in Section 4.2 in the stationary case we have \( \sigma^2 = \Pi_T(0) \) where \( \Pi_T(z) \) is a branch of \( \ln \lambda_T(z) \). Thus, arguing as in the proof of [60] Theorem 2.4 (b) we see that
\[ |\sigma_n^2 - \text{Var}_{\mu_T}(S_n f)| \leq n \delta_1(\varepsilon_0) \]
where \( \delta_1(\varepsilon_0) \to 0 \) as \( \varepsilon_0 \to 0 \). Thus by taking \( \varepsilon_0 \) small enough (so that \( \delta_1(\varepsilon_0) \leq \sigma^2/4 \)) we conclude that for all \( n \) large enough we have \( \sigma_n^2 \geq c_0 n \), where \( c_0 = \frac{1}{2} \sigma^2 \). \( \square \)

60. Remark. The transfer operators \( T \to L_T \) are continuous with respect to the dynamics \( T \) for a variety of piecewise monotone one dimensional interval maps when all \( T \)'s are close to a single map \( T \) in the Keller-Liverani norm (see [70]). In particular, we can consider the case when \( T_j x = \beta_j x \) and \( T x = \beta x \). Then the conditions of Proposition 59 hold true when \( \sup_j |\beta_j - \beta| \) and \( \sup_j |f_j - f| \) are small enough. In the multidimensional case, the conditions of Proposition 59 hold true when each \( T_j \) is obtained from \( T \) by a small perturbation of each inverse branch of \( T \). For instance, let \( X = [0,1)^d \) for some \( d \in \mathbb{N} \) and suppose that there is a partition of \( [0,1)^d \) into cubes of the form \([a_1, b_1) \times \cdots \times [a_d, b_d]\) so that on each cube \( T \) is expanding (and say, has a full image). Now, the conditions of Proposition 59 are in force if each \( T_j \) is generated by gluing sufficiently small perturbations of each one of the latter expanding \( |I| \) maps on the elements \( I \) of the partition.

5.4.2. Products of random non-stationary matrices. In this section we will briefly discuss how to extend the ideas from the previous section to uniformly bounded products of independent but not identically distributed random matrices, whose probability laws belong to a sufficiently small neighborhood of a single probability law with the properties described in Section 4.6.

Let us recall that the transfer operator corresponding to a probability distribution on \( G = GL(V) \) is given by
\[ L_{\mu} h(x) = \int_G \mu(g \cdot x) d\mu(g). \]

Then, for every two probability laws \( \mu_1, \mu_2 \) on \( G \) we have
\[ \sup_x |L_{\mu_1} h(x) - L_{\mu_2} h(x)| \leq \sup h \|\mu_1 - \mu_2\|_{TV}. \]

Moreover, given two vectors \( x, y \in V \), for every Hölder continuous function \( h \) with exponent \( \alpha \) we have
\[ |(L_{\mu_1} h(x) - L_{\mu_1} h(y)) - (L_{\mu_2} h(x) - L_{\mu_2} h(y))| \leq \sup_{g \in S(\mu_1) \cup S(\mu_2)} \|h(g \cdot x) - h(g \cdot y)\| \mu_1 - \mu_2\|_{TV} \leq \sup_{g \in S(\mu_1) \cup S(\mu_2)} \|g(x - y)\|^{\alpha} \nu_0(h) \|\mu_1 - \mu_2\|_{TV} \]
where $S(\mu)$ denotes the support of $\mu$. We thus see that
\[ v_\alpha(\mathcal{L}_{\mu_1}h - \mathcal{L}_{\mu_1}h) \leq \|\mu_1 - \mu_2\|_{TV} v_\alpha(h) \sup_{g \in S(\mu_1) \cup S(\mu_2)} \|g\|^\alpha. \]
We conclude that
\[ (5.16) \quad \|\mathcal{L}_{\mu_1} - \mathcal{L}_{\mu_2}\|_\alpha \leq \max \left(1, \sup_{g \in S(\mu_1) \cup S(\mu_2)} \|g\|^\alpha \right) \|\mu_1 - \mu_2\|_{TV}. \]

Now, let us take an independent sequence $g_1, g_2, \ldots$ uniformly bounded of random matrices which are not necessarily uniformly distributed. Let us denote law of $g_i$ by $\mu_i$. Let $\mu$ be a probability distribution $\mu$ with the properties described at the beginning of Section 4.6 (so that Assumptions 38 and 39 are in force and Proposition 40 holds true). Let us further assume that $S(\mu)$ is bounded. Set
\[ \varepsilon = \sup_i \|\mu_i - \mu\|_{TV}. \]
Then by (5.16),
\[ \limsup_{\varepsilon \to 0} \|\mathcal{L}_{\mu_i} - \mathcal{L}_{\mu_i}\|_\alpha = 0. \]
Thus, as in the previous section, when $\varepsilon$ is small enough we get an appropriate complex sequential spectral gap for the complex perturbations of the operators. This leads to the non-uniform Berry-Esseen theorem of all powers $m$ (and all their applications). That is, Theorems 5 and 9 are in force with every $m$, as well as the $L^{m-2}$ estimates in Corollary 19 (with all $m$) and Corollary 11 (i) with all $m$.

6. ADDITIONAL EXAMPLES

Let us note that Assumption 3 holds true for every $m$ when
\[ \gamma_j(S_n/\sigma_n) \leq C^j(j!)\sigma_n^{-j-2} \]
for some $C \geq 1$ and all $j \geq 3$. Indeed, this condition insures that the Taylor series of the function $\Lambda_n(t)$ and all of its derivatives converge in a neighborhood of the origin (which may depend on the order of differentiation).

61. Example.

- Nondegenerate U-Statistics, Characteristic Polynomials in the Circular Ensembles and Determinantal Point Processes as in [37, Section 3-5].
- Partial sums $S_n$ of exponentially fast $\phi$ mixing uniformly bounded Markov chains, with linearly fast growing variances (see [89, Theorem 4.26]).

7. “Reduction” to the self normalized case: proof of Theorems 5 and 8 relying on the self-normalized case

In this section we will prove Theorems 5 and 8 based on the validity of these theorems in the self-normalized case when $B_n = \sigma_n$ and $A_n = 0$. Let $W_n = S_n/\sigma_n$ and $Z_n = \frac{W_n - A_n}{\sigma_n}$. Then for every $x \in \mathbb{R}$,
\[ \mathbb{P}(Z_n \leq x) = \mathbb{P}(W_n \leq a_n x + v_n). \]
where $a_n = \frac{a_n}{\sigma_n} = 1 + O(\sigma_n^{-1})$ and $v_n = \frac{a_n}{\sigma_n} = O(\sigma_n^{-1})$. Therefore, all the results stated in Theorems 2.5 and 2.8 hold true with the generalized distribution function

$$\Phi_{m,n}(x) = \Phi_{m,n}(a_n x + v_n)$$

instead of $\Phi_{m,n}$, where $\Phi_{m,n}$ is defined in (2.3). The next step of the proof will be to pass from $\Phi_{m,n}(x)$ to a function defined similarly to $\Phi_{m,n}(x)$ but possibly with different polynomials. This is achieved in the following result.

62. **Lemma.** There are polynomials $U_{r,n}$ with bounded coefficients and degrees depending only on $r$ so that

$$|\Phi_{m,n}(x) - \left(\Phi(x) + \varphi(x) \sum_{r=1}^{m-2} B_n^{-r} U_{r,n}(x)\right) | \leq C_m(1 + |x|^m) e^{-x^2/4} \sigma_n^{-(m-1)}$$

where $C_m$ and $e_m$ are constants which depend only on $m$.

**Proof.** First, for any choice of polynomials $U_{r,n}$, if we choose $|x| \geq \sigma_n^\varepsilon$ for some $\varepsilon \in (0,1)$ then, regardless of the choice of $U_{r,n}$, since $a_n \to 1$ and $v_n \to 0$ both terms inside the absolute value on the left hand side of (7.1) are at most of order $|x|^m e^{-x^2/3}$ for some $c_m$. Let us write

$$|x|^m e^{-x^2/3} = |x|^m e^{-x^2/12} e^{-x^2/4} \leq |x|^m e^{-\sigma_n^{2\varepsilon}/12} e^{-x^2/4} \leq C |x|^m e^{-x^2/4} \sigma_n^{-(m-1)}.$$

Thus, by it is enough to find polynomials which satisfy (7.1) for all $n$ and $x$ so that $|x| \leq \sigma_n^\varepsilon$, where $\varepsilon$ can be arbitrarily close to 0.

Let us fix some $n$ and $x$ so that $|x| \leq \sigma_n^\varepsilon$. Let us first suppose that $|x| \leq C$ for some constant $C > 0$. Let us write

$$a_n x + v_n = x + (u_n x + v_n) := x + \eta$$

where $u_n = a_n - 1 = \frac{a_n}{\sigma_n} - 1$ and $\eta = \eta(n, x) = u_n x + v_n$. Then our assumption about the sequences $B_n$ and $\sigma_n$ insures that $u_n = \frac{a_n}{\sigma_n}$ for some bounded sequence $D_n$. Now, using that $\varphi^{(k)}(x) = (-1)^k \varphi(x) H_k(x)$, by the Lagrange form of the Taylor remainders of $\Phi$ around the point $x$ for all $s$ we have

$$\Phi(a_n x + v_n) = \Phi(x + \eta) = \Phi(x) + \varphi(x) \sum_{k=1}^{s} \frac{(-1)^{k-1} H_{k-1}(x)}{k!} \eta^k + O(\eta^{s+1}).$$

Now, since $|x| \leq C$ our assumption about $A_n$ and $B_n$ guarantee that $\eta = O(\sigma_n^{-1})$ and so the above remainder is of order $e^{-x^2/4}(1 + |x|^m)^{-1} \sigma_n^{-(m-1)}$ if $s = m - 1$. Similarly,

$$\varphi(a_n x + v_n) = \varphi(x + \eta) = \varphi(x) + \varphi(x) \sum_{k=1}^{s} \frac{(-1)^k H_k(x)}{k!} \eta^k + O(\eta^{s+1})$$

and the remainder is of order $e^{-x^2/4}(1 + |x|^m)^{-1} \sigma_n^{-(m-1)}$. Finally, let us expand the polynomials. Now, let us write $H_{r,n}(x) = \sum_{k=0}^{d_r} a_{k,r,n} x^k$, where $H_{r,n}$ are defined in (2.3). Then

$$H_{r,n}(a_n x + v_n) = H_{r,n}(x + \eta) = \sum_{q=0}^{d_r} \left( \sum_{k=q}^{d_r} \binom{k}{q} \eta^{k-q} \right) x^q.$$
Since \( u_n = \frac{D_n}{B_n} \), \( v_n = O(\sigma_n^{-1}) \) and \(|x| \leq C\) we see that \( \eta^j \) has the form \( \eta^j = B_n^{-j}(D_n x + L_n)^j \) for some bounded \( L_n \) and \( D_n \). Thus \( \eta^j \) is a polynomial in \( x \) whose coefficients are of order \( B_n^{-j} \). Combining this with the previous estimates and an ignoring terms of order \( B_n^{-u} \) for \( u > (m-1) \) we see obtain (7.1) with some polynomials \( U_{r,n} \) on bounded domains of \( x \).

Next, let us show that the above three Taylor estimates yield (7.1) with the same polynomials on domains of the form \( \{C \leq |x| \leq \sigma_n^+\} \), with \( C \) large enough. To achieve that we use the Lagrange form of the Taylor remainder to get that

\[
\Phi(a_n x + v_n) = \Phi(x) + \varphi(x) \sum_{k=1}^{\infty} (-1)^{k-1} \frac{H_{k-1}(x)}{k!} \eta^k + \frac{(-1)^s \eta^{s+1}}{(s+1)!} H_s(x + \zeta_1) \varphi(x + \zeta_1)
\]

and

\[
\varphi(a_n x + v_n) = \varphi(x + \eta) = \varphi(x) + \varphi(x) \sum_{k=1}^{\infty} (-1)^k \frac{H_k(x)}{k!} \eta^k + \frac{(-1)^s \eta^{s+1}}{(s+1)!} H_{s+1}(x + \zeta_2) \varphi(x + \zeta_2)
\]

where \( \max(|\zeta_1|, |\zeta_2|) \leq \eta \). Next, since \( |x| \leq \sigma_n^\varepsilon \) we get that \( |\eta| = O(\sigma_n^{-(1-\varepsilon)}) \) and hence when \((s+1)(1-\varepsilon) > m-1\) we have \(|\eta|^{s+1} = o(B_n^{-m-1})\). Next, since

\[
|\zeta_i| \leq |\eta| = o(1)
\]

and \( x^2 \leq \sigma_n^{2\varepsilon} \), if \( \varepsilon \) is small enough we get that

\[
\varphi(x + \zeta_i) \leq e^{-x^2/3}
\]

and so for \( i = 1, 2 \)

\[
|H_{s+i}(x)| \varphi(x + \zeta_i) \leq (1 + |x|^{u_n}) e^{-x^2/4}
\]

for some constant \( u_s \). By using again (7.2) and expressing \( \eta^j \) as a polynomial in \( x \) whose coefficients are of order \( \sigma_n^{-j} \), the proof of (7.1) is completed now also in the case \( C \leq |x| \leq \sigma_n^+ \).

\[
\square
\]

8. The self normalized case: proofs

8.1. The Edgeworth polynomials. Recall that \( \gamma_j(W) \) denotes the \( j \)-th cumulant of a random variable \( W \) with finite absolute \( j \)'s moment. Then for all \( 3 \leq j \leq m+1 \),

\[
\gamma_j(W_n) = \Lambda_n^{(j)}(0) = O(\sigma_n^{-(j-2)})
\]

where \( \Lambda_n \) comes from the main Assumption [3]. Let us consider the following polynomials

\[
P_{m,n}(z) = \sum_k \frac{1}{k_1! \cdots k_{m-2}!} \left( \frac{\gamma_3(W_n)}{3!} \right)^{k_1} \cdots \left( \frac{\gamma_m(W_n)}{m!} \right)^{k_{m-2}} z^{3k_1 + \cdots + mk_{m-2}}
\]

where the summation runs over the collection of \( m-2 \) tuples of nonnegative integers \((k_1, \ldots, k_{m-2})\) that are not all 0 so that \( \sum_j j k_j \leq m - 2 \). Let \( \nu_{m,n} \) be the signed measure on \( \mathbb{R} \) whose Fourier transform is

\[
g_{m,n}(t) = e^{-t^2/2} (1 + P_{m,n}(it)).
\]

Let \( H_k(z) \) be the \( k \)-th Hermite polynomial, which is defined through the identity

\[
(-1)^k H_k(x) \varphi(x) = \varphi^{(k)}(x).
\]

Then \( \mu_{m,n} \) is absolutely continuous with respect to the
Lebesgue measure with density

$$\varphi_{m,n}(x) = \varphi(x) \left( 1 + \sum_k \frac{1}{k_1! \cdots k_{m-2}!} \left( \frac{\gamma_3(W_n)}{3!} \right)^{k_1} \cdots \left( \frac{\gamma_m(W_n)}{m!} \right)^{k_{m-2}} H_k(x) \right)$$

where $$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$ is the standard normal density function and $$k = k(k_1, \ldots, k_{m-2}) = 3k_1 + \cdots + mk_{m-2}$$. Using that $$H_k(x) \varphi(x) = -(H_{k-1}(x) \varphi(x))'$$ we see that the corresponding generalized distribution function is given by

$$(8.3) \Phi_{m,n}(x) = \int_{-\infty}^{x} \varphi_{m,n}(x) dx = \Phi(x) - \varphi(x) \sum_k \frac{1}{k_1! \cdots k_{m-2}!} \left( \frac{\gamma_3(W_n)}{3!} \right)^{k_1} \cdots \left( \frac{\gamma_m(W_n)}{m!} \right)^{k_{m-2}} \Phi_{k-1}(x).$$

Note that $$\Phi_{m,n}(-\infty) = 0$$ and $$\Phi_{m,n}(\infty) = \int_{-\infty}^{\infty} \varphi_{m,n}(x) dx = g_{m,n}(0) = 1$$.

Notice also that for $$m = 2$$ we have $$P_{2,n} = 0$$ and so $$\nu_{2,n}$$ is the standard normal law and $$\Phi_{2,n} = \Phi$$ is the standard normal distribution function.

Next, observe that the function $$\Phi_{n,m}(x)$$ can also be written in the form

$$(8.4) \Phi_{n,m}(x) = \Phi(x) - \varphi(x) \sum_{r=1}^{m-2} \sigma_n^{-r} H_{r,n}(x)$$

where

$$(8.5) H_{r,n}(x) = \sum_{\vec{k} \in A_r} C_\vec{k} \prod_{j=1}^{s} (\gamma_{j+2}(S_n)\sigma_n^{-2})^{k_j} H_{k-1}(x)$$

and $$A_r$$ is the set of all tuples of nonnegative integers $$\vec{k} = (k_1, \ldots, k_s)$$, $$k_s \neq 0$$ for some $$s = s(\vec{k}) \geq 1$$ so that $$\sum_j jk_j = r$$ (note that when $$r \leq m-2$$ then $$s \leq m-2$$ since $$k_s \geq 1$$). Moreover, $$k = 3k_1 + \cdots (s+2)k_s$$ and

$$C_\vec{k} = \prod_{j=1}^{s} \frac{1}{k_j!(j+2)^{k_j}}.$$

We note that the polynomials $$H_{r,n}$$ have bounded coefficients (because of (8.1)) and that their degrees does not depend on $$n$$. The polynomials $$H_{r,n}$$ coincide with the “Edgeworth polynomials” defined in [34] (denoted there by $$P_{r,n}$$).

We also note that the Fourier transform of the derivative of $$\Phi_{m,n}(x)$$ has the form

$$g_{m,n}(x) = e^{-x^2/2} \left( 1 + \sum_{r=1}^{m-2} \sigma_n^{-r} P_{j,n}(x) \right)$$

where

$$P_{r,n}(x) = \sum_{\vec{k} \in A_r} C_\vec{k} \prod_{j=1}^{s} (\gamma_{j+2}(S_n)\sigma_n^{-2})^{k_j} (ix)^{3k_1 + \cdots (s+2)k_s}.$$
8.2. Proof of Theorems 5 and 8. The proof of Theorems 5 and 8 rely on the following two results, whose proof is postponed to the next sections.

63. Proposition. Under Assumption 3 if \(|t| \geq C \sigma_n^{1/3}\) for some \(C > 0\) then for \(p = 0, 1, 2, \ldots, m\) we have

\[
|j_{m,n}^{(p)}(t)| \leq C_m \sigma_n^{-(m+1)} e^{-ct^2}
\]

where \(C_m\) depends only on \(C, m\) and the constants in Assumption 3 and \(c_0 \in (0, 1/2)\) is a constant (which, upon increasing \(C_m\), can be made arbitrarily close to 1/2).

64. Proposition. Under Assumption 3 there are constants \(c, c_0 > 0\) and \(C_m > 0\) so that if \(|t| \leq c \sigma_n\) then for \(p = 0, 1, 2, \ldots, m\) we have

\[
\left| \frac{d^p}{dt^p} (f_n(t) - g_{m,n}(t)) \right| \leq C_m e^{-ct^2} \min\{1, |t|^{m+1-p}\} \sigma_n^{-(m-1)}.
\]

8.3. Proof of Theorems 5 and 8 relying on relying on Propositions 63 and 64.

65. Lemma. Under Assumption 3 for all \(x \in \mathbb{R}\) and \(3 \leq p \leq m\) we have

\[
|\Phi_{p,n}(x) - \Phi_{p-1,n}(x)| \leq C_m (1 + |x|)^{-m-1} \sigma_n^{-(p-2)}.
\]

As a consequence, for all \(2 \leq m_1 < m\),

\[
|\Phi_{m,n}(x) - \Phi_{m_1,n}(x)| \leq A_m (1 + |x|)^{-m-1} \sigma_n^{-(m_1-1)}.
\]

Here \(A_m, C_m\) are constants that do not depend on \(n\) or \(x\).

Proof. Notice that the difference between the derivatives of \(\Phi_{p,n}(y)\) and \(\Phi_{p-1,n}(y)\) has the form

\[
\varphi_{p,n}(y) - \varphi_{p-1,n}(y) = \frac{e^{-y^2/2}}{\sqrt{2\pi}} \left( \sum_{\bar{k}} \prod_{j=1}^{p-2} \frac{1}{k_j! j!} \prod_{j=p}^{p-2} \left( \gamma_{j+2}(W_n) \right)^{k_j} H_k(y) \right)
\]

where the summation is running over all tuples \(\bar{k} = (k_1, \ldots, k_{p-2})\) of nonnegative integers so that \(\sum_j jk_j = p - 2\). Using (2.1) with \(t = 0\), we see that the product of the cumulants in the above difference is of order \(\sigma_n^{-(p-2)}\). Therefore,

\[
|\Phi_{p,n}(x) - \Phi_{p-1,n}(x)| \leq \int_{-\infty}^{x} |\varphi_{p,n}(y) - \varphi_{p-1,n}(y)| dy \leq C \sigma_n^{-(p-2)} \int_{-\infty}^{x} e^{-y^2/2} (1 + |y|^s) dy
\]

where \(s = s_n\) is some positive integer. This proves that the lemma when \(x \leq 0\) since the integral on the right hand side decays exponentially fast in \(|x|\) as \(x \to -\infty\). To prove the statement for positive \(x\)’s we recall that \(\Phi_{j,n}(\infty) = 1\), and so

\[
\Phi_{p,n}(x) - \Phi_{p-1,n}(x) = (1 - \Phi_{p-1,n}(x)) - (1 - \Phi_{p,n}(x)) = \Phi_{p-1,n}(\infty) - \Phi_{p-1,n}(x) - (\Phi_{p,n}(\infty) - \Phi_{p,n}(x)) = \int_{x}^{\infty} (\varphi_{p-1,n}(y) - \varphi_{p,n}(y)) dy.
\]

Using this inequality the proof proceeds similarly to the case when \(x \to -\infty\). □

Proof of Theorems 5 and 8. By the results described in Ch. VI, Lemma 8], if \(F\) is a distribution function and \(G\) is a generalized distribution function so that \(G(-\infty) = \)
Then for all $x \in \mathbb{R}$ and $T > 1$,

\begin{equation}
|F(x) - G(x)| \leq c(s)(1 + |x|)^{-s} \left( \int_{-T}^{T} \left| f(t) - g(t) \right| dt + \int_{-T}^{T} \left| a_s(t) \right| dt + \frac{K}{T} \right)
\end{equation}

where $f(t) = \int_{-\infty}^{\infty} e^{itx} dF(x)$, $g(t) = \int_{-\infty}^{\infty} e^{itx} dG(x)$ and

$$a_s(t) = \int_{-\infty}^{\infty} e^{itx} \left( a^s(F(x) - G(x)) \right).$$

Here $c(s)$ is an absolute constant which depends only on $s$. Notice that for $t \neq 0$ we have (see [17]),

\begin{equation}
(i)^{-s} \frac{d^s}{dt^s} a_0(t) = i^{-s} \int_{0}^{1} \left( a_{0}^{(s)}(t) - a_{0}^{(s)}(\eta t) \right) s\eta^{s-1} d\eta.
\end{equation}

Note also that $a_0(y) = f_n(y) - g_{m,n}(y)$. Moreover, observe that for every measurable nonnegative function $q : \mathbb{R} \to \mathbb{R}$ and all $T > 0$ and $s \geq 1$ we have

$$\int_{0}^{T} \frac{1}{t} \int_{0}^{1} q(\eta t) s\eta^{s-1} d\eta dt = \int_{0}^{T} q(x) \left( \frac{1}{x} - \frac{1}{T} \cdot \frac{x}{T} \right)^{s-1} dx \leq \int_{0}^{T} \frac{q(x)}{x} dx.$$

Indeed, the above follows by making the change of variables $tn = x$ in the inner integral and then changing the order of integration. Therefore, using also (8.7) we see that

\begin{equation}
\int_{-T}^{T} \left| a_s(t) \right| dt \leq 2 \int_{-T}^{T} \left| a_0^{(s)}(t) \right| dt.
\end{equation}

Now, in order to prove the non-uniform Berry-Esseen theorem (Theorem 5) we take $F$ to be the distribution function of $W_n$, $G = \Phi_{m,n}$, $T = O(\sigma_n)$ and $s = m$. Note that the derivative of $G$ is uniformly bounded in $n$ because of Assumption 3 which guarantees that $G'(x) = E_n(x)e^{-x^2/2}$ for some polynomial $E_n(x)$ whose degree does not depend on $n$ and its coefficients are uniformly bounded in $n$. Next, by taking $p = 0$ in Proposition 64 we see that the first integral in the brackets on the right hand side of (8.6) is of order $O(\sigma_n^{-m-1})$. In order to estimate the second integral in the brackets on the right hand side of (8.6) let us take $p = m$ in Proposition 64 which together with (8.8) yields that

$$\int_{-T}^{T} \left| a_s(t) \right| dt = O(\sigma_n^{-m-1}).$$

To complete the proof of Theorem 5 we need to pass from $\Phi_{m,n}$ to $\Phi = \Phi_{2,n}$, and this is done by applying Lemma 65 with $m_1 = 2$.

To prove the non-uniform Edgeworth expansions (Theorem 8) of order $r \leq m - 2$ and power $m$, by using Lemma 65 it is enough to consider the case $r = m - 2$. Let us take again $F$ to be the distribution function of $W_n$ and $G = \Phi_{m,n}$. Let us also take $s = m$. Let us fix some $\varepsilon \in (0,1)$ and let $T = B\sigma_n^{m-2}$ for some $B > \frac{1}{\varepsilon}$. We thus see that the contribution of the term $c(s)K/T$ is at most $c\varepsilon\sigma_n^{-(m-2)} = c\varepsilon\sigma_n^{-r}$ for some constant $c$ which does not depend on $\varepsilon$. Hence, it remains to show that for this fixed $B$ all the rest of
the expressions on the right hand side of (8.6) are $o(\sigma_n^{-m-2}) = o(\sigma_n^{-r})$. To estimate the first integral in the brackets on the right hand side of (8.6), we split it to the integrals on two domains. The first is a domain of the form $\{|t| = O(\sigma_n)\}$, and the corresponding integral is of order $O(\sigma_n^{-(m-1)})$ because of Proposition 64. The second domain is of the form $\{c_0 \sigma_n \leq |t| \leq T\}$, and to estimate the corresponding integral we use Proposition 63 together with Assumption 6, which yield that

$$
\int_{-T}^{T} \left| \frac{f_n(t) - g_{m,n}(t)}{n} \right| dt = o(\sigma_n^{-m-2}) = o(\sigma_n^{-r}).
$$

To estimate the second integral in the brackets of the right hand side of (8.6), we first use (8.8), and then we split the integral

$$
\int_{-T}^{T} \left| \frac{a_s(t)}{t} \right| dt
$$

into two domains. The first is of the form $\{|t| \leq c_0 \sigma_n\}$ for a sufficiently small $c_0$. By taking $p = m$ in Proposition 64 we see that the contribution of this domain of integration is $O(\sigma_n^{-(m-1)})$. The second domain has the form $\{c_0 \sigma_n \leq |t| \leq T\}$, and by combining Proposition 63 with Assumption 6, we see that the contribution coming from this domain is $o(\sigma_n^{-(m-2)})$. We conclude that

$$
\int_{-T}^{T} \left| \frac{a_s(t)}{t} \right| dt = o(\sigma_n^{-m-2}) = o(\sigma_n^{-r})
$$

and the proof of Theorem 8 is complete. \(\square\)

8.4. **Proof of Theorem 9 via smoothing.** The proof begins with a classical smoothing argument. Similarly to [19], let $\xi_m$ be a centered random variable with finite $m$ absolute moments whose characteristic function $h_m$ is supported on some interval $[-a, a]$ for some $a > 0$. For example, we can first take $\xi_m$ to be the random variable with the density $w_m(x) = \Lambda^{-1} (\sin x/x)^{2m}$, where $\Lambda$ is an appropriate normalizing constant (so that the characteristic function is proportional to $(1 - |t|)_+$). Then we can normalize $\tilde{\xi}_m$ so that its second moment is 1. Let us denote the normalized version by $\xi_m$. By enlarging the underlying probability space if necessary, we can always assume that $\xi_m$ is defined on this space and it is independent of $S_n$. Fix some constant $c > 0$, let $n$ be so that $c \sigma_n > 1$ and set

$$
\tilde{S}_n = \sqrt{1 - (c \sigma_n)^{-2}} S_n + c^{-1} \xi_m.
$$

Then, because $\xi_m$ is independent of $S_n$ we have

$$
\bar{\psi}_n(t) := E[e^{it \tilde{S}_n}] = \psi_n(t) h(t/c)
$$

and so the characteristic function $\bar{\psi}_n$ of $\tilde{S}_n$ vanishes outside the interval $[-ca, ca]$. Notice also that

$$
\sigma_n^2 = \|S_n\|_{L^2} = \|\tilde{S}_n\|_{L^2}.
$$

In view of (8.9) and (8.10), it is also clear that Assumption 3 is satisfied for $\tilde{S}_n$ if it holds for $S_n$, and with the same $m$. Now, if $c$ is small enough then Assumption 4 trivially holds.
true for \( \tilde{S}_n \) for every \( m \), and so, by applying Theorem \[8\] with \( \tilde{S}_n \) instead of \( S_n \) (in the self normalized case) we see that, with \( \tilde{F}_n(x) = \mathbb{P}(\tilde{S}_n/\sigma_n \leq x) \) we have

\[
\left| \tilde{F}_n(x) - \tilde{\Phi}_{m-2,n}(x) \right| \leq C(1 + |x|)^{-m} \sigma_n^{-(m-1)}
\]

where \( \tilde{\Phi}_{m-2,n} \) is defined similarly to \( \Phi_{m-2,n} \) but with the cumulants of \( \tilde{S}_n \) instead of the cumulants of \( S_n \), and \( C \) is some constant. By applying \[19\] Corollary 3.2 we see that for every \( p < m - 1 \) we have

\[
W_p(d\tilde{F}_n, d\tilde{\Phi}_{m-2,n}) = O(\sigma_n^{-1}).
\]

Next, by applying \[19\] Proposition 5.1 (with \( \varepsilon = O(\sigma_n^{-1}) \)) together with Lemma \[65\] we see that

\[
W_p(d\tilde{\Phi}_{m-2,n}, d\Phi) = O(\sigma_n^{-1}).
\]

Finally, by the definition of \( W_p \) we have

\[
W_p(dF_n, d\tilde{F}_n) \leq \|\tilde{S}_n - S_n\|_{L^p} \leq (1 - \sqrt{1 - (\sigma_n)^{-2}}) \|S_n\|_{L^p} + \|\xi_m\|_{L^p} (\sigma_n)^{-1} \leq (\sigma_n)^{-2} \|S_n\|_{L^p} + \|\xi_m\|_{L^p} (\sigma_n)^{-1}.
\]

In order to to complete the proof of Theorem \[9\] we apply the last three \( W_p \)-estimates together with the triangle inequality for the metric \( W_p \) and the following result:

66. **Lemma.** Under Assumption \[3\] we have that for all \( p \leq m \),

\[
\|S_n\|_{L^p} \leq C_p \sigma_n
\]

for some constant \( C_p \).

**Proof of Lemma 66.** Let \( k \in \{m, m+1\} \) be an even number. Then

\[
\|S_n\|_{L^p} \leq \|S_n\|_{L^k} = \left( \mathbb{E}[S_n^k] \right)^{1/k}.
\]

Next, by \[39\] (1.53) (which in our case is also a consequence of Faà Di Bruno’s formula) we see that \( \mathbb{E}[S_n^k] \) is a linear combination (with coefficients not depending on \( n \)) of products of the form \( \prod_{j=1}^d (\gamma_{k_j}(S_n))^{r_j} \) with \( \sum_{j=1}^d r_j k_j = k \) and \( r_j > 0 \) (\( k_j \) and \( r_j \) are nonnegative integers). Note that \( \Gamma_1(S_n) = \mathbb{E}[S_n] = 0 \), and so we can assume that \( k_j > 1 \) for all \( j \). Now, by Assumption \[3\] we have that \( \gamma_{k_j}(S_n) = O(\sigma_n^2) \) for all \( j \) (as \( k_j \leq k \leq m+1 \)). Thus the contribution coming from the above product is \( O(\sigma_n^{2\sum r_j}) \). Notice now that since \( k_j \geq 2 \) we have \( 2 \sum_j r_j \leq \sum_j r_j k_j = k \). We thus conclude that

\[
\mathbb{E}[S_n^k] \leq C_k \sigma_n^k
\]

as claimed. \( \square \)

8.5. **Proof of the main Propositions 63 and 64.** The general structure of both proofs share some similarities with existing proofs for partial sums of independent random variables (see the recent survey \[18\]), but many arguments are different since the desired error term is \( o(\sigma_n^{-r}) \), \( r = m - 2 \) and not the Lyapunov coefficient of order \( m \), and instead of independence of \( S_{j+1} - S_j \) we only use Assumption \[3\]
Proof of Proposition 63. Fix some \( t \) so that \( |t| \geq C\sigma_n^{1/3} \). Now, since \( g_m = g_{m,n} \) can be naturally extended to a complex function \( g_m(z) \) on the complex plane, using the Cauchy integral formula (or the maximum modulus principle) we see that for all \( \eta \in (0,1/4) \) and all \( p = 0,1,2,...,m \) we have

\[
|g_m^{(p)}(t)| \leq C_m \max_{|z-t| \leq \eta |t|} |z|^{-p} |g_m(z)|
\]

where have used that \( |t|/2 \leq |z| \leq 2|t| \) which follows from the inequality \( |z-t| \leq \eta |t| \). Thus, it is enough to show that when \( |t| \geq C\sigma_n^{1/3} \),

\[
\max_{|z-t| \leq \eta |t|} |z|^{-p} |g_m(z)| \leq C_m \sigma_n^{-(m+1)} e^{-c_0 t^2}
\]

where \( c_0 \to 1/2 \) as \( \eta \to 0 \) (and it does not depend on \( t \)) and \( C_m \) is a constant.

Next, let us consider the entire functions \( R_k(z) = e^{-z^2/2} z^k \). Then, using that

\[
|\gamma_j(W_n)| \leq C_j \sigma_n^{-(j-2)}, \quad 3 \leq j \leq m
\]

we see that

\[
(8.12) \quad |g_m(z)| \leq C_m \left( R_0(z) + \sum_k \sigma_n^{-d_k} |R_{s_k}(z)| \right)
\]

where \( \bar{k} = (k_1,k_2,...,k_{m-2}) \) ranges over all the tuples so that \( d_k := \sum_j j k_j \leq m - 2 \) and \( s_k = \sum_j (j+2) k_j \). Notice that \( \Re(z^2) \leq a \Re(t^2) \) for some constant\footnote{We can take \( a = a(\eta) = 1 - 2\eta - \eta^2 \)} \( a > 0 \) which converges to 1 as \( \eta \to 0 \) (and it does not depend on \( t \)). Thus,

\[
|R_{s_k}(z)| \leq 2e^{-at^2/2} |t|^{s_k}.
\]

Now, since \( |t| \geq C\sigma_n^{1/3} \), we have that \( |t| \) it is bounded away from the origin, and we conclude that there is an integer \( u_m \), which depends only on \( m, C \) and the first index \( n_0 \) so that \( \sigma_{n_0} \geq 1 \) with the property that for all relevant tuples \( \bar{k} \) and for all complex \( z \) with \( |z-t| \leq \eta |t| \) we have

\[
|R_{s_k}(z)| \leq B |t|^{p+n_0} e^{-at^2/2} \leq B(m,a,a') |t|^{p-3(m+1)} e^{-a't^2/2}
\]

where \( a' < a \) is arbitrarily close to \( a \), \( B \) is some constant and \( B_m(m,a,a') \) depends only on \( m,a,a' \) and \( C \) and \( n_0 \) as above. Finally, since \( |t| \geq C\sigma_n^{1/3} \) we have

\[
|t|^{-3(m+1)} \leq C^{-(m+1)} \sigma_n^{-(m+1)}
\]

and hence there is a constant \( A \) so that

\[
|z|^{-p} |R_{s_k}(z)| \leq A \sigma_n^{-(m+1)} e^{-a't^2/2}.
\]

Now the proposition follows from \( (8.11) \) and \( (8.12) \). \( \square \)

8.6. Proof of Proposition 64. The proof is somehow lengthy, and we will split it into several parts.
8.6.1. Bounds on the characteristic function and its derivatives.

67. **Lemma.** (i) Under Assumption 3 we have the following. For every \( c \in (0, 1) \) there exists a constant \( a > 0 \) so that on the domain \( \{ |t| \leq a\sigma_n \} \) we have

\[
|f_n(t)| \leq e^{-ct^2}.
\]

(ii) Under Assumption 3 (with some \( m \)) we have the following. For every \( c \in (0, 1) \) there exist constants \( a, b > 0 \) so that on the domain \( \{ |t| \leq a\sigma_n \} \) for all \( p \leq m \) we have

\[
\left| \frac{d^p}{dt^p} f_n(t) \right| \leq b(1 + |t|^p)e^{-ct^2}.
\]

**Proof.** We have

\[
f_n(t) = e^{\Lambda_n(t) - t^2/2}.
\]

To prove (i) we notice that \( \Lambda_n(0) = \Lambda'_n(0) = \Lambda''_n(0) = 0 \). Hence, by taking \( j = 3 \) in (2.1) we see that the second order Taylor reminder of the function \( \Lambda_n \) around the origin is of order \( O(|t|^3\sigma_n^{-1}) \). Thus, if \( |t|/\sigma_n \) is small enough we have

\[
|\Lambda_n(t)| \leq C|t|^3\sigma_n^{-1} \leq \varepsilon t^2
\]

where \( \varepsilon \) is an arbitrary small positive number.

In order to prove the second part, recall that by Faà Di Bruno’s formula, for a \( p \)-times differential function \( h(t) \) we have

\[
(8.13) \quad \frac{d^p}{dt^p} \left( e^{h(t)} \right) = e^{h(t)} \sum_k C_{p,k} \prod_{j=1}^p (h^{(j)}(t))^{k_j}
\]

where \( \tilde{k} = (k_1, \ldots, k_p) \) ranges over all \( p \)-tuples of nonnegative integers so that \( \sum_j jk_j = p \) and

\[
C_{k,p} = \frac{p!}{\prod_{j=1}^p k_j!(j!)^{k_j}}.
\]

Here \( h^{(j)} \) denotes the \( j \)-th derivative of a function \( h \). Let us apply the above formula with the function \( h(t) = \Lambda_n(t) - t^2/2 \). We first claim that all the derivatives of orders \( 1 \leq j \leq p \) of \( h \) are bounded by some constant \( C \) which does not depend on \( n \), when \( |t| \leq a\sigma_n \) for some \( a \) small enough. Plugging in these upper bounds in (8.13) (i.e. replacing the derivatives with the upper bounds) and using part (i) to bound \( |f_n(t)| = |e^{h(t)}| \) we obtain the estimates described in part (ii).

Now, let us shows that there is an \( a > 0 \) so that on \( \{ |t| \leq a\sigma_n \} \) we have \( \max_{1 \leq j \leq p} |h^{(j)}(t)| = O(1) \). First, by Assumption 3 on a domain of the form \( \{ |t| \leq a\sigma_n \} \) for all \( 3 \leq j \leq m \) we have

\[
|h^{(j)}(t)| = |\Lambda^{(j)}_n(t)| \leq C\sigma_n^{-j} = O(1).
\]

For \( j = 2 \) we have

\[
|h''(t)| = |1 + \Lambda''_n(t)|.
\]

Since \( \Lambda''_n(0) = 0 \) we have

\[
\Lambda''_n(t) = \int_0^t \Lambda'''_n(s)ds
\]

and so by Assumption 3

\[
|\Lambda''_n(t)| \leq C|t|\sigma_n^{-1} = O(1).
\]
Combining this with the fact that $\Lambda_n(0) = 0$ we get that when $|t| = O(\sigma_n)$ then
\[ |h'(t)| = |\Lambda_n'(t) - t| \leq |t| + \left| \int_0^t \Lambda_n''(s) ds \right| \leq |t| + C|t|^2\sigma_n^{-1} \leq C'|t|, \]

\[ 8.6.2. \text{Expansions of the characteristic function.} \text{ We first need is the following simple observation, which for the sake of convenience if formulated as a lemma.} \]

\[ \text{68. Lemma. Let } h : (-a,a) \to \mathbb{R}, a > 0 \text{ be a function which is differentiable } k \text{ times for some } k. \text{ Let } R_k \text{ be the Taylor remainder of } h \text{ of order } k \text{ around the origin, and let } u \leq k. \text{ Then the } u \text{-th derivative } R_k^{(u)} \text{ of } R_k \text{ is the Taylor remainder of the function } h^{(u)} \text{ of order } k - u \text{ around the origin.} \]

\[ \text{69. Proposition. Under Assumption } \text{ we have the following. There is a constant } c_1 > 0 \text{ so that for all real } t \text{ with } |t| \leq c_1 \sigma_n^{m+1} \text{ we have} \]
\[ e^{t^2/2} f_n(t) = e^{Q_{m,n}(it)} (1 + \varepsilon_{n,m}(t)) \]
where
\[ Q_{m,n}(z) = \sum_{\ell=3}^m \gamma_{\ell}(W_n) \frac{z^\ell}{\ell!} = \sum_{\ell=3}^m \sigma_{n-\ell} \gamma_{\ell}(S_n) \frac{z^\ell}{\ell!} \]
and for all $p \leq m$,
\[ \left| \frac{d^p(\varepsilon_{n,m}(t))}{dt^p} \right| \leq C_m |t|^{m+1-p} \sigma_n^{-(m-1)} \]
where $C_m$ depends only on $m$ and the constants in Assumption \text{.}

\[ \text{Proof. Set} \]
\[ r(t) = \Lambda_n(t) - Q_{m,n}(it) \]
which is the Taylor remainder of order $m$ of $\Lambda_n(t)$ around the origin. Then
\[ e^{t^2/2} f_n(t) = e^{\Lambda_n(t)} = e^{Q_{m,n}(t)} e^{r(t)}. \]
Now, by Lemma \text{ for all } j \leq m \text{ we have that } r^{(j)}(t) \text{ is the Taylor remainder of order } m - j \text{ of the function } \Lambda_n^{(j)}(t) \text{ around the origin. Using now the Lagrange form of the remainder (of order } m - j \text{) together with Assumption } \text{ we see that when } |t| \leq c \sigma_n \text{ for } c \text{ small enough then} \]
\[ |r^{(p)}(t)| \leq C|t|^{m+1-j} \sigma_n^{-(m+1-2)} = C|t|^{m+1-j} \sigma_n^{-(m-1)} \]
for some constant $C$.

It remains to show that when $|t| = O(\sigma_n^{m+1})$ we can replace $e^{r(t)}$ with $1 + r(t)$. Let $\varepsilon(t) = e^{r(t)} - 1$. Then
\[ e^{t^2} f_n(t) = e^{Q_{m,n}(t)} e^{r(t)} = e^{Q_{m,n}(t)}(1 + \varepsilon(t)) \]
and we need to show that the function $\varepsilon(t) = \varepsilon_{n,m}(t) = \varepsilon(t)$ satisfies the properties described in Proposition \text{. By taking } p = 0 \text{ in } (8.15) \text{ we see that } |r(t)| \leq C |t|^{m+1} \sigma_n^{-(m-1)} = O(1) \text{ and so} \]
\[ |\varepsilon(t)| = |e^{r(t)} - 1| \leq C' |r(t)| \leq C'' |t|^{m+1} \sigma_n^{-(m-1)}. \]
To get the desired estimates for higher order derivatives, by (8.13), for \( p \geq 1 \) we have

\[
\varepsilon^{(p)}(t) = e^{r(t)} \sum_{k} C_{p, k} \prod_{j=1}^{p} (r^{(j)}(t))^{k_j} = (1 + \varepsilon(t)) \sum_{k} C_{p, k} \prod_{j=1}^{p} (r^{(j)}(t))^{k_j}
\]

and so, using also (8.15) we see that there is a constant \( C_m > 0 \) so that

\[
|\varepsilon^{(p)}(t)| \leq C_m |t|^{-p} \sum_{m_{j=3}}^{m} |t|^{m+1} \sigma_n^{-m+1} \sum_{j=1}^{k_j} \prod_{j=1}^{k_j} (r^{(j)}(t))^{k_j} \]

where we have taken into account that \( \sum_{j=1}^{k_j} j k_j = p \). Finally, since \( \sum_{j=1}^{k_j} j k_j \geq 1 \) and \( |t|^{m+1} \sigma_n^{-m+1} \leq c_1 \) for some constant \( c_1 \) we see that

\[
|\varepsilon^{(p)}(t)| \leq A_m |t|^{m+1-p} \sigma_n^{-(m-1)}
\]

where \( A_m \) is another constant which depends on \( m \) and \( c_1 \), and the proof of the proposition is complete. \( \square \)

8.6.3. Estimates on the cumulants polynomials \( Q_m \).

70. **Lemma.** Under Assumption 3 we have

\[
|Q_{m,n}(z)| \leq C_m |z|^3 \sigma_n^{-1} \sum_{\ell=3}^{m} |z|^{\ell-3} \sigma_n^{-(\ell-3)}.
\]

In particular, if \( |z| \leq C \sigma_n \) then

\[
|Q_{m,n}(z)| \leq C' |z|^3 \sigma_n^{-1}
\]

for some constant \( C' \), and so when \( |z| \leq C_1 \sigma_n^{1/3} \) then \( |Q_{m,n}(z)| \leq C'_1 \).

**Proof.** The lemma follows from the definition of \( Q_{m,n} \) together with the estimates

\[
|\gamma_j(W_n)| \leq C_j \sigma_n^{-(j-2)}, \quad 3 \leq j \leq m + 1.
\]

on the cumulants of \( W_n = S_n/\sigma_n \) (which come from Assumption 3). \( \square \)

The next result we need is as follows.

71. **Lemma.** Under Assumption 3, if \( |t| \leq C \sigma_n^{1/3} \) for some constant \( C \) then

\[
e^{Q_{m,n}(it)} = \sum_{k=0}^{m-2} \frac{Q_{m}(it)^k}{k!} + \varepsilon(t)
\]

where for all \( p \leq m \) we have

\[
|\varepsilon^{(p)}(t)| \leq A |t|^{3(m-1)-p} \sigma_n^{-(m-1)}
\]

for some constant \( A \) that may depend on \( C \) and \( m \).

**Proof.** By Lemma 70 when \( |z| = O(\sigma_n^{1/3}) \) then \( Q_{m,n}(z) \) is uniformly bounded in a ball of radius \( M \) which is independent of \( n \). Let \( \Psi(z) = e^z - \sum_{j=0}^{m-2} \frac{z^j}{j!} \). Then by the Lagrange form of the Taylor remainder of order \( m - 2 \) of the function \( e^z \) we have

\[
\sup_{|z| \leq M} |\Psi(z)| \leq C_M |z|^{m-1}.
\]
Thus, by using the penultimate in Lemma 70 we see that if $|z| = O(\sigma_n^{1/3})$ then
\[ |\varepsilon(z)| = |\Psi(Q_{m,n}(iz))| \leq C_M|Q_{m,n}(iz)|^{m-1} \leq C'|z|^{3(m-1)}\sigma_n^{-(m-1)}. \]
This proves the lemma for $p = 0$. The proof when $p > 0$ is completed by using that by the Cauchy integral formula we have
\[ |\varepsilon^{(p)}(t)| \leq p!(|t|/4)^{-p} \max_{|z-t|=|t|/4} |\varepsilon(z)|. \]
\[ \square \]

The following result is an immediate corollary of Lemma 70 and Lemma 71.

72. **Corollary.** Under Assumption \( \exists \) on domains of the form \( \{|t| \leq C\sigma_n^{1/3}\} \) all the first \( m \) derivatives of \( e^{Q_{m(it)}} \) are bounded by a constant which does not depend on \( n \) and \( t \).

8.6.4. Passing from \( Q \) to \( P \).

73. **Proposition.** Let \( P_{m,n} \) be as in (8.2). Under Assumption \( \exists \), if \( |t| \leq C\sigma_n^{1/3} \) then
\[ e^{Q_{m(it)}} = 1 + P_{m,n}(it) + \delta(t) \]
where for all \( p \leq m \) we have
\[ |\delta^{(p)}(t)| \leq A \max(|t|^{m+1-p}, |t|^{3(m-1)-p})\sigma_n^{-(m-1)} \]
for some constant \( A \).

**Proof.** Using Lemma 71 we only need to bound the first \( p \) derivatives of
\[ d(t) := \sum_{k=0}^{m-2} \frac{Q_{m,n}(it)^k}{k!} - P_{m,n}(it). \]
Notice that by the definitions (8.2) and (8.14) of \( P_{m,n} \) and \( Q_{m,n} \), respectively, we have
\[ d(t) = \sum_k D_{k,m-2} \prod_{j=1}^{m-2} \left( (\gamma_{j+2}(W_n))^{k_j} \right) (it)^{3k_1+4k_2+\ldots+mk_{m-2}} \]
where
\[ D_{k,m-2} = \frac{1}{\prod_{j=1}^{m-2} k_j!(j+2)!^{k_j}} \]
and \( k = (k_1, \ldots, k_{m-2}) \) ranges over all the tuples of nonnegative integers so that \( \sum_j k_j \leq m - 2 \) and \( k_1 + 2k_2 + \ldots + (m-2)k_{m-2} \geq m - 1 \). It is enough to show that the \( p \)-th derivatives of each monomial
\[ w_k(t) = \prod_{j=1}^{m-2} \left( (\gamma_{j+2}(W_n))^{k_j} \right) (it)^{3k_1+4k_2+\ldots+mk_{m-2}} \]
admit the desired upper bounds. Let us fix such a monomial. Then
\[ \left| \frac{d^p}{dt^p} w_k(t) \right| \leq \prod_{j=1}^{m-2} \left( (\gamma_{j+2}(W_n))^{k_j} \right) |t|^{3k_1+4k_2+\ldots+mk_{m-2}-p}. \]
Notice that
\[(8.16) \quad 3k_1 + 4k_2 + \ldots + mk_{m-2} = 2 \sum_j k_j + \sum_j jk_j \geq m + 1\]
and so the above power of \(|t|\) is positive.

Next, let us suppose that \(|t| \leq 1\). Since
\[3k_1 + 4k_2 + \ldots + mk_{m-2} - p \geq m + 1 - p\]
and by Assumption 3
\[(8.17) \quad |\gamma_j(W_n)| \leq C \sigma_n^{j-2}, \quad 3 \leq j \leq m + 1\]
and \(k_1 + 2k_2 + \ldots + (m-2)k_{m-2} \geq m - 1\) we conclude that
\[
\left| \frac{d^p}{dt^p} w_k(t) \right| \leq C|t|^{m+1-p} \sigma_n^{-(m-1)}.
\]

To complete the proof, let us consider the case when \(1 \leq |t| \leq c \sigma_n^{1/3}\) for some \(c > 0\). Then, using that \(\sum_j jk_j \geq m - 1 > \sum_j k_j\) together with (8.17) we have
\[
\left| \frac{d^p}{dt^p} w_k(t) \right| \leq C|t|^{-p}\left|t\right|^{\sum_j jk_j + 2\sum_j k_j} \sigma_n^{-\sum_j jk_j}
\]
\[= C|t|^{3(m-1)-p}\left(|t|^{\sum_j jk_j + 2\sum_j k_j} \sigma_n^{-\sum_j jk_j}\right)\sigma_n^{-(m-1)}
\]
\[\leq C'\left|t\right|^{3(m-1)-p}\left(|t|^{\sum_j jk_j + 2\sum_j k_j} \sigma_n^{-\sum_j jk_j}\right)\sigma_n^{-(m-1)}
\]
\[= C'|t|^{3(m-1)-p}\left(|t|^{2\sum_j jk_j + 2\sum_j k_j} \sigma_n^{-\sum_j jk_j}\right)\sigma_n^{-(m-1)} \leq C'|t|^{3(m-1)-p}\sigma_n^{-(m-1)}.
\]

\[\square\]

74. Corollary. Under Assumption 3, if \(|t| \leq C \sigma_n^{1/3}\) then there is a constant \(A > 0\) so that for all \(p \leq m\),
\[|P_{m,n}(it)| \leq A \min(1, |t|^{-p})\]

Proof. Notice that \(P_{m,n}(it) = e^{Q_{m,n}(it)} - 1 - \delta(t)\). Now the corollary follows from the combination of Lemma 70, Lemma 71 and Proposition 73. The upper bound \(A\) when \(|t| \leq 1\) follows directly from the definition (8.2) of \(P_{m,n}\) together with Assumption 3 which yields that the coefficients of \(P_{m,n}\) are bounded. \(\square\)

8.6.5. Completing the proof of Proposition 64

75. Lemma. Under Assumption 3, if \(|t| \leq C \sigma_n^{1/3}\) then
\[f_n(t) = e^{-t^2/2} (1 + P_m(it) + r(t))\]
with
\[|r^{(p)}(t)| \leq C_m \sigma_n^{(m-1)} \max(|t|^{m+1-p}, |t|^{3(m-1)-p}), \quad p = 0, 1, \ldots, m.\]

Proof. Using Propositions 69 and 73 we have
\[f_n(t) = e^{-t^2/2} e^{Q_{m}(it)} (1 + \varepsilon_{m,n}(t)) = e^{-t^2/2} (1 + P_m(it) + \delta(t)) (1 + \varepsilon_{m,n}(t)).\]
Thus
\[r(t) = (1 + P_m(it)) \varepsilon_{m,n}(t) + \delta(t) (1 + \varepsilon_{m,n}(t)).\]
Now the bounds on the derivative of \( r(t) \) follow from Propositions 69 and 73 and Corollary 74 together with the binomial formula for the derivatives of products of two functions.

76. **Corollary.** If \( |t| \leq C \sigma_n^{1/3} \) then

\[
\left| \frac{d^p}{dt^p} (f_n(t) - g_{m,n}(t)) \right| \leq C_m \sigma_n^{-(m-1)} \max(|t|^{m+1-p}, |t|^{3(m-1)+p}) e^{-c_0 t^2}
\]

for some constants \( C_m, c_0 > 0 \).

**Proof.** In the notations of Lemma 75 we have

\[ f_n(t) - g_{m,n}(t) = e^{-t^2/2} r(t). \]

Now we can use the binomial formula for the derivatives of a product of two functions and Lemma 75.

**Completion of the proof of Proposition 64.** On the interval \( |t| \leq C \sigma_n^{1/3} \) for some \( C > 0 \) the desired estimate follows from Corollary 76. On the intervals \( C \sigma_n^{1/3} \leq |t| \leq C_1 \sigma_n \) (for some \( C_1 \) small enough), the desired estimate follows from Lemma 67 together with Proposition 63.

**References**

[1] Agnew, R.P.: Global versions of the central limit theorem. Proc. Natl. Acad. Sci. USA 40, 800–804 (1954)

[2] Agnew, R.P.: Estimates for global central limit theorems. Ann. Math. Stat. 28, 26–42 (1957)

[3] Agnew, R.P.: Asymptotic expansions in global central limit theorems. Ann. Math. Stat. 30, 721–737 (1959)

[4] R. Aimino, M. Nicol and S. Vaienti, **Annealed and quenched limit theorems for random expanding dynamical systems**, Probab. Th. Rel. Fields 162, 233-274, (2015).

[5] E. Alós. A generalization of the Hull and White formula with applications to option pricing approximation. Finance Stoch., 10(3) (2006), 353–365.

[6] J. Angst, G. Poly, **A weak Cramér condition and application to Edgeworth expansions**, Electron. J. Probab. 22 (2017) # 59, 1-24.

[7] M. Austern, T. Liu, **Wasserstein-p Bounds in the Central Limit Theorem under Weak Dependence**, preprint, https://arxiv.org/abs/2209.09377v1

[8] P. Baldi, Y. Rinott **On normal approximations of distributions in terms of dependency graphs**, Ann. Prob. 17 (1989) 1646–1650.

[9] A.D. Barbour, **Asymptotic expansions based on smooth functions in the central limit theorem**, Probab. Th. Rel. Fields 72 (1986) 289–303.

[10] P.Beckedorf, A.Rohde, Non-uniform bounds and Edgeworth expansions in self-normalized limit theorems, preprint, https://arxiv.org/abs/2207.14402v2

[11] V. Bentkus, F. Gotze, W. Van Zvet **An Edgeworth Expansion for Symmetric Statistics**. Ann. Stat. 25 (1997) 851–896.

[12] I. Berkes and W. Philipp, **Approximation theorems for independent and weakly dependent random vectors**. Ann. Probab. 7 29–54 (1979).

[13] Berkes, I., Liu, W., Wu, W.: **Komlós–Major–Tusnády approximation under dependence**. Ann. Probab. 42, 794–817 (2014)

[14] W. Berry, **The accuracy of the Gaussian approximation to the sum of independent variates**, Trans. AMS 49 (1941) 122–136.
[15] R.N. Bhattacharya, R. Ranga Rao Normal Approximation and Asymptotic Expansions, Wiley, New York-London-Sydney-Toronto (1976) xiv+274 pp.

[16] P. Bickel P., F. Gotze, W. Van Zwet The Edgeworth Expansion for U-Statistics of Degree Two, Ann. Stat. 14 (1986) 1463–1484.

[17] Bobkov, S.G. Closeness of probability distributions in terms of Fourier–Stieltjes transforms (English translation: Russian Math. Surveys). (Russian) Uspekhi Mat. Nauk 71(6), 37–98 (2016)

[18] S.G. Bobkov Asymptotic Expansions for Products of Characteristic Functions Under Moment Assumptions of Non-integer Orders, In: Carlen, E., Madiman, M., Werner, E. (eds) Convexity and Concentration. The IMA Volumes in Mathematics and its Applications, vol 161. Springer, New York, NY (2017)

[19] S.G. Bobkov, Berry–Esseen bounds and Edgeworth expansions in the central limit theorem for transport distances, Probab. Theory Relat. Fields (2018) 170:229–262.

[20] Bougerol, Ph., Lacroix, J. Products of Random Matrices with Applications to Schrödinger Operators. Birkhäuser, Boston, Basel, Stuttgart (1985)

[21] E. Breuillard, Distributions diophantiennes et theoreme limite local sur $\mathbb{R}^d$, Probab. Th. Rel. Fields 132 (2005) 39–73.

[22] Butterley, O., Peyman, E. Exponential mixing for skew products with discontinuities. Trans. Am. Math. Soc. 369(2), 783–803 (2017).

[23] H. Callaert, P. Janssen, The Berry-Esseen theorem for U statistics, Ann. Stat. 6 (1978) 417–421.

[24] H. Callaert, P. Janssen, N. Veraverbeke An Edgeworth Expansion for U-Statistics. Ann. Stat. 8 (1980) 299–312.

[25] Louis H.Y. Chen and Qi-Man Shao, A non-uniform Berry–Esseen bound via Stein’s method, Probab Theory Relat Fields 120, 236–254.

[26] Z. Coelho and Qi-Man Shao, A non-uniform Berry–Esseen bound via Stein’s method, Probab Theory Relat Fields 120, 236–254.

[27] R. Bowen, Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Lecture Notes in Mathematics, volume 470, Springer Verlag, 1975.

[28] P. Chebyshev. Sur deux the'or'emes relatifs aux probabilités. Acta Math., 14(1) (1890), 305–315.

[29] J. P. Conze and A. Raugi, Limit theorems for sequential expanding dynamical systems on $[0, 1]$, In: Ergodic Theory and Related Fields., Contemp. Math. 430 (2007), 89–121.

[30] H. Cramer On the composition of elementary errors, Skand. Aktuarietid skr. 1 (1928) 13–74; 141–180.

[31] D. Dolgopyat, On mixing properties of compact group extensions of hyperbolic systems, Isr. J. Math. 130, 157–205 (2002).

[32] D. Dolgopyat, K. Fernando An error term in the Central Limit Theorem for sums of discrete random variables, preprint.

[33] D. Dolgopyat, Y. Hafouta Edgeworth expansion for independent bounded integer valued random variables, Stoch. Proc. App. https://doi.org/10.1016/j.spa.2022.07.001

[34] D. Dolgopyat, Y. Hafouta, A Berry-Esseen theorem and Edgeworth expansions for uniformly elliptic inhomogeneous Markov chains, Probab. Theory Relat. Fields (2022). https://doi.org/10.1007/s00440-022-01177-2

[35] D. Dolgopyat, Y. Hafouta, A Berry-Esseen theorem and Edgeworth expansions for uniformly elliptic inhomogeneous Markov chains, preprint arXiv:2111.03738

[36] D. Dolgopyat, O. Sarig, Local limit theorems for inhomogeneous Markov chains, arXiv:2109.05560

[37] H. Döring, P. Eichelsbacher, Moderate deviations via cumulants, J. Theor. Probab. 26 (2013) 360–385.

[38] D. Dragičević, G. Froyland, C. González-Tokman, and S. Vaienti. A Spectral Approach for Quenched Limit Theorems for Random Expanding Dynamical Systems, Comm. Math. Phys. 360 (2018), 1121–1187.
[39] D. Dragićević, G. Froyland, C. González-Tokman, and S. Vaienti. A Spectral Approach for Quenched Limit Theorems for Random Expanding Dynamical Systems, Tran. Amer. Math. Soc. 360 (2018), 1121–1187.

[40] D. Dragićević, Y. Hafouta, Limit theorems for random expanding or Anosov dynamical systems and vector valued observables, Ann. Henri Poincaré 21 (2020) 3869–3917.

[41] F.Y. Edgeworth. The asymmetrical probability curve. Proceedings of the Royal Society of London, 56(336-339) (1894), 271–272

[42] B. Efron, ‘Bootstrap methods: Another look at the jackknife’, Ann. Statist. 7(1) (1979), 1–26.

[43] B. Efron. The jackknife, the bootstrap and other resampling plans, volume 38 of CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1982.

[44] C.-G. Esseen Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law, Acta Math. 77 (1945) 1–125.

[45] C.-G. Esseen, A moment inequality with an application to the central limit theorem, Skand. Aktuarietidskr. 39, (1956) 160–170.

[46] Esseen, C.-G.: On mean central limit theorems. Kungl. Tekn. Högsk. Handl. Stockholm, 121 (1958)

[47] W. Feller, An introduction to probability theory and its applications, Vol. II., 2d edition, John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[48] K. Fernando, C. Liverani Edgeworth expansions for weakly dependent random variables, Ann. de l’Institut Henri Poincare Prob. & Stat. 57 (2021) 469–505.

[49] K. Fernando, F. Pêne Expansions in the local and the central limit theorems for dynamical systems, arXiv:2008.08726.

[50] J. Fouque, G. Papanicolaou, R. Sircar, and K. Solna. Singular perturbations in option pricing. SIAM Journal on Applied Mathematics, 63(5) (2003), 1648–1665

[51] H Frankowska, Some inverse mapping theorems, Annales de l’Institut Henri Poincare (C) Non Linear, 1990.

[52] R.E. Gilman, A class of functions continuous but not absolutely continuous, Ann. Math. 33 (1932) 433–442.

[53] B.V. Gnedenko, A.N. Kolmogorov Limit distributions for sums of independent random variables, Cambridge, Addison-Wesley, 1954.

[54] W.F. Grams, R.J. Serfling, Convergence rates for U-statistics and related statistics, Ann. Stat. 1 (1973) 153–160.

[55] S. Gouëzel. Berry–Esseen theorem and local limit theorem for non uniformly expanding maps, Annales IHP Prob. & Stat. 41 (2005), 997–1024.

[56] Y. Guivarc’h, J. Hardy. Théorèmes limites pour une classe de chaînes de markov et applications aux difféomorphismes d’anosov, Annales IHP Prob. & Stat. 24 (1988), 73–98.

[57] Y. Guivarc’h Spectral gap properties and limit theorems for some random walks and dynamical systems, Proc. Symp. Pure Math. 89, 279–310 (2015).

[58] Y. Hafouta and Yu. Kifer, Nonconventional limit theorems and random dynamics, World Scientific, Singapore, 2018.

[59] Y. Hafouta, On the asymptotic moments and Edgeworth expansions of some processes in random dynamical environment, J. Stat Phys 179 (2020) 945–971.

[60] Y. Hafouta, Limit theorems for some time dependent expanding dynamical systems, Nonlinearity 33 (2020) 6421–6460.

[61] Y. Hafouta, Limit Theorems for Random Non-uniformly Expanding or Hyperbolic Maps with Exponential Tails, Ann. Henri Poincaré 23, 293–332 (2022).

[62] Y. Hafouta, An almost sure invariance principle for some classes of non-stationary mixing sequences, arXiv:2005.02915, 12 pages
Y. Hafouta, *Explicit conditions for the CLT and related results for non-uniformly partially expanding random dynamical systems via effective RPF rates*, preprint, https://arxiv.org/abs/2208.00518

P. Hall. The bootstrap and Edgeworth expansion. Springer Series in Statistics. Springer-Verlag, New York, 1992.

P. Hall, *Edgeworth expansion for Student’s t statistic under minimal moment conditions*, Ann. Probab. 15 (1987) 920–931.

H. Hennion, L. Hervé. *Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness*, Springer Lecture Notes in Math. 1766 (2001).

L. Hervé, F. Pène, *The Nagaev–Guivarch method via the Keller-Liverani theorem*, Bull. Soc. Math. France 138 (2010) 415–489.

I. A. Ibragimov, Yu. V. Linnik *Independent and stationary sequences of random variables*, Wolters-Noordhoff Publishing, Groningen, 1971. 443 pp.

M. Jirak *Berry-Esseen theorems under weak dependence*, Ann. Prob. 44 (2016) 2024–2063.

G. Keller, C. Liverani, *Stability of the spectrum for transfer operators*, Annali della Scuola Normale Superiore di Pisa - Classe di Scienze, Serie 4, Volume 28 (1999) no. 1, pp. 141-152.

G. Keller, C. Liverani, *A Spectral Gap for a One-dimensional Lattice of Coupled Piecewise Expanding Interval Maps*, In: Dynamics of Coupled Map Lattices and of Related Spatially Extended Systems. Lecture Notes in Physics, vol 671. Springer, Berlin, Heidelberg (2005).

Yu. Kifer, *Limit theorems for random transformations and processes in random environments*, Trans. Amer. Math. Soc. 350 (1998), 1481-1518.

A. Korepanov, *Rates in almost sure invariance principle for Young towers with exponential tails*, Comm. Math. Phys. 363 (2018), 173–190.

S.N. Lahiri, Resampling methods for dependent data (Springer Series in Statistics. Springer-Verlag, New York, 2003).

D. Leung, Q.M. Shao, *Nonuniform Berry-Esseen bounds for Studentized U-statistics*, preprint, https://arxiv.org/abs/2303.08619v1

Yu. Kifer, *Limit theorems for random transformations and processes in random environments*, Trans. Amer. Math. Soc. 350 (1998), 1481-1518.

I. Melbourne and M. Nicol, *Almost Sure Invariance Principle for Nonuniformly Hyperbolic Systems*. Commun. Math. Phys. 260, 131-146 (2005).

S.V. Nagaev, *Some limit theorems for stationary Markov chains*, Theory Probab. Appl. 2 (1957) 378-406.

S.V. Nagaev, *More exact statements of limit theorems for homogeneous Markov chains*, Theory Probab. Appl. 6 (1961), 62-81.

S.V. Nagaev, *Some limit theorems for large deviations*. Theory Probab. Appl. 10(1), 214–235 (1965)

L.V Osipov, *Asymptotic expansions in the central limit theorem*. (Russian) Vestnik Leningrad. Univ. 19, 45–62 (1967)

L.V Osipov, *On asymptotic expansions of the distribution function of a sum of random variables with non uniform estimates for the remainder term*. (Russian) Vestnik Leningrad. Univ. 1, 51–59 (1972)

M. Peligrad, Central limit theorem for triangular arrays of non-homogeneous Markov chains, Probab. Theory Relat. Fields (2012) 154:409–428.

V.V. Petrov, *Sums of independent random variables*. Translated from the Russian by A. A. Brown. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82. Springer, New York-Heidelberg, 1975. x+346 pp. Russian ed.: Moscow, Nauka (1972)

E. Rio, *Sur le théorème de Berry-Esseen pour les suites faiblement dépendantes*, Probab. Th. Relat. Fields 104 (1996), 255-282
[86] E. Rio, Upper bounds for minimal distances in the central limit theorem, Annales de l’Institut Henri Poincaré - Probabilités et Statistiques 2009, Vol. 45, No. 3, 802–817.

[87] Y. Rinott and V. Rotar On Edgeworth expansions for dependency-neighborhoods chain structures and Stein’s method, Probab. Theory Rel. Fields 126 (2003) 528–570.

[88] J. Rousseau-Egele. Un théorème de la limite locale pour une classe de transformations dilatantes et monotones par morceaux, Ann. Prob. 11 (1983) 772–788.

[89] L. Saulis, V.A. Statulevicius, Limit Theorems for Large Deviations, Kluwer, Dordrecht, Boston, 1991.

[90] S. Sethuraman, S. R. S. Varadhan A martingale proof of Dobrushin’s theorem for non-homogeneous Markov chains, Electron. J. Probab. 10 (2005) paper 36, 1221–1235.

[91] C. Stein, A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, Berkeley Symposium on Mathematical Statistics and Probability, 1972: 583-602 (1972)

[92] M. Stadlbauer, S. Suzuki, P. Varandas Thermodynamic formalism for random non-uniformly expanding maps, Commun. Math. Phys. 385, 369–427 (2021)

[93] L.S. Young, Statistical properties of dynamical systems with some hyperbolicity, Ann. Math. 7 (1998) 585-650.

The University of Maryland
Email address: yeor.hafouta@mail.huji.ac.il