Equidistribution of random walks on compact groups

Bence Borda
Alfréd Rényi Institute of Mathematics
1053 Budapest, Reáltaanoda u. 13-15, Hungary
Email: borda.bence@renyi.hu

Keywords: empirical distribution, ergodic theorem, strong law of large numbers, law of the iterated logarithm, central limit theorem, Wiener process

Mathematics Subject Classification (2010): 60G50, 60B15

Abstract

Let $X_1, X_2, \ldots$ be independent, identically distributed random variables taking values from a compact metrizable group $G$. We prove that the random walk $S_k = X_1X_2 \cdots X_k$, $k = 1, 2, \ldots$ equidistributes in any given Borel subset of $G$ with probability 1 if and only if $X_1$ is not supported on any proper closed subgroup of $G$, and $S_k$ has an absolutely continuous component for some $k \geq 1$. More generally, the sum $\sum_{k=1}^{N} f(S_k)$, where $f : G \to \mathbb{R}$ is Borel measurable, is shown to satisfy the strong law of large numbers and the law of the iterated logarithm. We also prove the central limit theorem with remainder term for the same sum, and construct an almost sure approximation of the process $\sum_{k \leq t} f(S_k)$ by a Wiener process provided $S_k$ converges to the Haar measure in the total variation metric.

1 Introduction

Let $G$ be a compact Hausdorff group with normalized Haar measure $\mu$, and let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed (i.i.d. for short) $G$-valued random variables. The random walk $S_k = \prod_{j=1}^{k} X_j = X_1X_2 \cdots X_k$ is a classical object in probability theory. Throughout we assume that the distribution of $X_1$ is a regular Borel probability measure $\nu$ on $G$; the distribution of $S_k$ is thus $\nu^k$, the $k$-fold convolution of $\nu$. Generalizing results of Lévy [12] on the circle group $G = \mathbb{R}/\mathbb{Z}$ and Kawada and Itô [9] on compact metrizable groups, it was Urbanik [20] and Kloss [10] who proved that if $\nu^k$ is weakly convergent, then its weak limit is the normalized Haar measure of a closed subgroup of $G$. Stromberg [19] gave the following necessary and sufficient condition for the weak limit to be the Haar measure on $G$ itself. We shall say that $\nu$ is adapted if the support of $\nu$ is not contained in any proper closed subgroup of $G$, and that $\nu$ is strictly aperiodic if the support of $\nu$ is not contained in a coset of any proper closed normal subgroup of $G$. 

**Theorem A.** Let $G$ be a compact Hausdorff group, and let $\nu$ be a regular Borel probability measure on $G$. The following are equivalent.

(i) $\nu$ is adapted and strictly aperiodic.

(ii) $\nu^k \to \mu$ weakly as $k \to \infty$.

A similar classical result gives a necessary and sufficient condition for convergence in the total variation metric $\| \cdot \|_{TV}$. The Lebesgue decomposition of $\nu^k$ with respect to the Haar measure $\mu$ will be written as $\nu^k = (\nu^k)_{\text{abs}} + (\nu^k)_{\text{sing}}$, where $(\nu^k)_{\text{abs}}$ is absolutely continuous and $(\nu^k)_{\text{sing}}$ is singular with respect to $\mu$.

**Theorem B.** Let $G$ be a compact Hausdorff group, and let $\nu$ be a regular Borel probability measure on $G$. The following are equivalent.

(i) $\nu$ is adapted and strictly aperiodic, and $(\nu^k)_{\text{abs}} \neq 0$ for some $k \geq 1$.

(ii) $\|\nu^k - \mu\|_{TV} \to 0$ as $k \to \infty$.

Moreover, if these equivalent conditions hold, then the convergence in (ii) is exponentially fast.

Special cases of Theorem B were proved by Bhattacharya [5]. For the general case and the history of related results see [1] and [16]. We mention that if $G$ is connected, then the assumption that $(\nu^k)_{\text{abs}} \neq 0$ for some $k \geq 1$ implies that $\nu$ is adapted and strictly aperiodic. This follows from the fact that in a connected, compact Hausdorff group any proper closed subgroup has Haar measure 0.

Theorems A and B concern the distribution of $S_k$ for a given $k \geq 1$. We can also view $S_k$, $k = 1, 2, \ldots$ as a random sequence in $G$, and consider the empirical distribution of the terms $S_1, S_2, \ldots, S_N$ for some $N \geq 1$. Under the technical assumption that $G$ is metrizable, Berger and Evans [2, Corollary 3.1] proved the following.

**Theorem C.** Let $G$ be a compact metrizable group. Let $X_1, X_2, \ldots$ be i.i.d. $G$-valued random variables with distribution $\nu$, and set $S_k = \prod_{j=1}^k X_j$. The following are equivalent.

(i) $\nu$ is adapted.

(ii) For any continuous function $f : G \to \mathbb{R}$

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^N f(S_k) = \int_G f \, d\mu \quad \text{a.s.}$$

(1)
Note that a.s. (almost surely) means that the given relation holds with probability 1. Since the Banach space of continuous, real-valued functions on $G$ (or indeed, on any compact metric space) is separable, condition (ii) in the previous theorem is equivalent to the property that with probability 1, (1) holds for all continuous functions $f : G \to \mathbb{R}$ simultaneously. A (deterministic) sequence $a_k, k = 1, 2, \ldots$ in $G$ is called uniformly distributed if $\lim_{N \to \infty} \left( \frac{1}{N} \sum_{k=1}^{N} f(a_k) \right) = \int_G f \, d\mu$ for any continuous function $f : G \to \mathbb{R}$. Theorem C thus states that the random sequence $S_k, k = 1, 2, \ldots$ is uniformly distributed with probability 1 if and only if $\nu$ is adapted. See [3], [4] and [17] for related results on the circle group $G = \mathbb{R}/\mathbb{Z}$, and [2] for the case of continuous time processes.

In this paper we consider $\sum_{k=1}^{N} f(S_k)$ for Borel measurable functions $f : G \to \mathbb{R}$, and we prove the following analogue of Theorem C.

**Theorem 1.** Let $G$ be a compact metrizable group. Let $X_1, X_2, \ldots$ be i.i.d. $G$-valued random variables with distribution $\nu$, and set $S_k = \prod_{j=1}^{k} X_j$. The following are equivalent.

(i) $\nu$ is adapted, and $(\nu^k)_{\text{abs}} \neq 0$ for some $k \geq 1$.

(ii) For any bounded, Borel measurable function $f : G \to \mathbb{R}$

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} f(S_k) = \int_G f \, d\mu \quad \text{a.s.} \quad (2)$$

(iii) For any bounded, Borel measurable function $f : G \to \mathbb{R}$

$$\limsup_{N \to \infty} \frac{\left| \sum_{k=1}^{N} f(S_k) - N \int_G f \, d\mu \right|}{\sqrt{N \log \log N}} < \infty \quad \text{a.s.}$$

The implications (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) are straightforward. Condition (ii) is in fact equivalent to the assumption that (2) holds for the indicator function $f = I_B$ of any Borel set $B \subseteq G$; indeed, a bounded, Borel measurable function can be uniformly approximated by finite linear combinations of such indicator functions. The equivalence (i) $\iff$ (ii) in Theorem 1 thus states that the random sequence $S_k, k = 1, 2, \ldots$ equidistributes in any given Borel set with probability 1 if and only if $\nu$ is adapted, and $(\nu^k)_{\text{abs}} \neq 0$ for some $k \geq 1$. Equidistribution of a random sequence in any given Borel set with probability 1 is sometimes called the “strong uniform distribution” property. In contrast, (ordinary) uniform distribution means equidistribution in any Borel set $B \subseteq G$ such that $\mu(\partial B) = 0$. Note that equidistribution in all Borel sets simultaneously is impossible; in particular, no deterministic sequence satisfies the strong uniform distribution property (unless $G$ is finite).
In Theorems C and 1 we did not assume that \( \nu \) is strictly aperiodic, whereas in Theorems A and B strict aperiodicity is required for the convergence of \( \nu^*k \). In the proof of the implication \( (i) \Rightarrow (iii) \) in Theorem 1 we will thus first assume that \( \nu \) is strictly aperiodic. In case the support of \( \nu \) is contained in a coset of a closed normal subgroup \( H \triangleleft G \), we will see that the factor group \( G/H \) is finite and cyclic, and we will argue by induction on the index \( |G:H| \). Surprisingly, in Theorem 1 the strong law of large numbers (condition (ii)) and the law of the iterated logarithm (condition (iii)) are equivalent. This is a consequence of the fact that whenever \( \nu^*k \) converges to the Haar measure \( \mu \) in the total variation metric, the convergence is necessarily exponentially fast. This fact does not have an analogue for weak convergence. We also prove the following central limit theorem under the technical assumption that \( \nu \) is a central measure. Note that condition (ii) below expresses convergence in distribution to the standard normal distribution.

**Theorem 2.** Let \( G \) be a compact metrizable group. Let \( X_1, X_2, \ldots \) be i.i.d. \( G \)-valued random variables with distribution \( \nu \), and set \( S_k = \prod_{j=1}^k X_j \). Assume that \( \nu \) is central. The following are equivalent.

(i) \( \nu \) is adapted and strictly aperiodic, and \( (\nu^*k)_{\text{abs}} \neq 0 \) for some \( k \geq 1 \).

(ii) For any bounded, Borel measurable function \( f : G \to \mathbb{R} \) such that \( \int_G f \, d\mu = 0 \) and \( f \) is not \( \mu \)-a.e. zero, we have

\[
\frac{\sum_{k=1}^N f(S_k)}{\sqrt{C(f, \nu)N}} \overset{d}{\to} N(0, 1)
\]

with some constant \( C(f, \nu) > 0 \) depending only on \( f \) and \( \nu \).

Recall that a compact Hausdorff topological space is metrizable if and only if it is second countable. We mention that in the proofs of Theorems 1 and 2 the choice of the metric on \( G \) is irrelevant; we only use the second countability of \( G \). Whether Theorems 1 and 2 are true for compact Hausdorff groups remains open.

## 2 Results

### 2.1 Preliminaries

For the rest of the paper we assume that \( G \) is a compact metrizable group. Let \( \mu \) denote the Haar measure on \( G \) normalized so that \( \mu(G) = 1 \). We will write \( L^p(G) = L^p(G, \mu) \) for the Lebesgue space of real-valued, Borel measurable functions with respect to \( \mu \), and \( \|f\|_p = \|f\|_{L^p(G, \mu)} \). In addition, \( \| \cdot \|_p \) will also denote the \( L^p \)-norm of (real-valued) random variables. Recall that \( \mu \) is both left and right invariant;
that is, for any Borel set $B \subseteq G$ and any $y \in G$ we have $\mu(yB) = \mu(By) = \mu(B)$.

Therefore for any $f \in L^1(G)$ we have

$$\int_G f(x) \, d\mu(x) = \int_G f(xy) \, d\mu(x) = \int_G f(yx) \, d\mu(x).$$

The total variation of a finite, signed Borel measure $\vartheta$ on $G$ is defined as

$$\|\vartheta\|_{TV} = \sup \left\{ \left| \int_G f \, d\vartheta \right| : f : G \to \mathbb{R} \text{ Borel measurable}, \sup_G |f| \leq 1 \right\}.$$

Given two finite, signed Borel measures $\vartheta_1$ and $\vartheta_2$ on $G$, their convolution $\vartheta_1 \ast \vartheta_2$ is the unique finite, signed Borel measure such that for any bounded, Borel measurable function $f : G \to \mathbb{R}$

$$\int_G f \, d(\vartheta_1 \ast \vartheta_2) = \int_G \int_G f(xy) \, d\vartheta_1(x) \, d\vartheta_2(y).$$

It is easy to see that $\|\vartheta_1 \ast \vartheta_2\|_{TV} \leq \|\vartheta_1\|_{TV} \cdot \|\vartheta_2\|_{TV}$. A finite, signed Borel measure $\nu$ on $G$ we have $\nu \ast \mu = \mu \ast \nu = \mu$. If $\nu_1$ and $\nu_2$ are Borel probability measures on $G$, then $\|\nu_1 - \nu_2\|_{TV} = 2 \sup_B |\nu_1(B) - \nu_2(B)|$, where the supremum is over all Borel sets $B \subseteq G$. The support of a Borel probability measure $\nu$ on $G$, denoted by $\text{supp} \, \nu$, is the smallest closed set $F \subseteq G$ such that $\nu(F) = 1$.

**Remark.** Every finite Borel measure on $G$ (or indeed, on any Polish space) is regular. Therefore in the definitions of total variation and convolution we could have equivalently used continuous functions $f : G \to \mathbb{R}$ instead of bounded, Borel measurable ones. The existence and uniqueness of the convolution thus follows from the Riesz representation theorem.

A $G$-valued random variable is a Borel measurable map $X$ from a probability space to $G$. Let $\nu_X$ denote the distribution of $X$; that is, $\nu_X(B) = \Pr(X \in B)$ for all Borel sets $B \subseteq G$. The variable $X$ is called uniformly distributed if $\nu_X = \mu$. If $X$ and $Y$ are independent $G$-valued random variables, then $\nu_{XY} = \nu_X \ast \nu_Y$. We shall write $X \overset{d}{=} Y$ if the (real-valued or $G$-valued) random variables $X$ and $Y$ have the same distribution.

Let $\hat{G}$ denote the unitary dual of $G$; that is, a complete set of pairwise unitarily inequivalent, irreducible unitary representations of $G$. Recall that every such representation is finite dimensional, and let $d_\pi$ denote the dimension of $\pi \in \hat{G}$. Thus $\pi(x)$ is a $d_\pi \times d_\pi$ unitary matrix with complex entries for any given $x \in G$. 

5
Let \( \pi_0 \in \hat{G} \), \( \pi_0 = 1 \) denote the trivial representation. Given \( f \in L^1(G) \) and \( \pi \in \hat{G} \), let \( f(\pi) = \int_G f(x) \pi(x)^* \, d\mu(x) \) denote the Fourier coefficients of \( f \). Here \( \pi(x)^* \) denotes the conjugate transpose of \( \pi(x) \), and the integral is taken entrywise. The Fourier coefficients of a finite, signed Borel measure \( \vartheta \) on \( G \) are defined similarly as \( \hat{\vartheta} = \int_G \pi(x)^* \, d\vartheta(x) \), \( \pi \in \hat{G} \). The Parseval formula states that for any \( f,g \in L^2(G) \) we have

\[
\int_G fg \, d\mu = \sum_{\pi \in \hat{G}} d_{\pi} \text{tr} \left( \hat{f}(\pi) \hat{g}(\pi)^* \right),
\]

where \( \text{tr} \) denotes trace. Given \( \pi \in \hat{G} \) the complex conjugate \( \bar{\pi} \) is also an irreducible unitary representation of \( G \), called the contragredient of \( \pi \). The contragredient \( \bar{\pi} \) may or may not be unitarily equivalent to \( \pi \). For the theory of Fourier analysis on compact groups we refer the reader to [7].

The notation \( a_n \ll b_n \) and \( a_n = O(b_n) \) mean that there exists an (implied) constant \( K \geq 0 \) such that \( |a_n| \leq K b_n \) for all \( n \geq 1 \). We write \( a_n = \Theta(b_n) \) if \( a_n \ll b_n \ll a_n \). We will use subscripts to denote dependence of the implied constant on certain parameters; thus e.g. \( a_n \ll_{f,\nu} b_n \) and \( a_n = O_{f,\nu}(b_n) \) mean that the implied constant may depend on \( f \) and \( \nu \). We emphasize that in the statement of all theorems, propositions and lemmas implied constants in the notation \( \ll \) and \( O \) are universal; in particular, they do not even depend on the group \( G \).

### 2.2 The main theorems

Let \( G \) be a compact metrizable group with normalized Haar measure \( \mu \), let \( X_1, X_2, \ldots \) be i.i.d. \( G \)-valued random variables with distribution \( \nu \), and set \( S_k = \prod_{j=1}^k X_j \). Let \( \Delta_k = \| \nu^{*k} - \mu \|_{TV} = 2 \sup_B |\Pr(S_k \in B) - \mu(B)| \) denote the total variation distance of the distribution of \( S_k \) from the uniform distribution. Note that

\[
\| \nu^{(k+1)} - \mu \|_{TV} = \| (\nu^{*k} - \mu) * \nu \|_{TV} \leq \| \nu^{*k} - \mu \|_{TV} \cdot \| \nu \|_{TV} = \| \nu^{*k} - \mu \|_{TV},
\]

hence \( \Delta_{k+1} \leq \Delta_k \).

The precise rate of convergence in the total variation metric was found by Anoussis and Gatzouras [1, Theorem 4.1]. Let

\[
q = \max \left\{ \sup_{\pi \in \hat{G}\setminus\{\pi_0\}} \rho(\hat{\nu}(\pi)), \inf_{k \geq 1} \| (\nu^{*k})_{\text{sing}} \|_{TV}^{1/k} \right\},
\]

where \( \rho(\hat{\nu}(\pi)) \) denotes the spectral radius of the matrix \( \hat{\nu}(\pi) \). Then \( \lim_{k \to \infty} \Delta_k^{1/k} = \inf_{k \geq 1} \Delta_k^{1/k} = q \); moreover, \( q < 1 \) if and only if \( \nu \) is adapted and strictly aperiodic,
and \((\nu^k)_{\text{abs}} \neq 0\) for some \(k \geq 1\). Thus, as already stated in Theorem B, whenever \(\|\nu^k - \mu\|_{TV} \to 0\), the convergence is necessarily exponentially fast; more precisely, we have \(q^k \leq \Delta_k\) for every \(k \geq 1\), and \(\Delta_k \leq (q + \varepsilon)^k\) for every \(k \geq k_0(\nu, \varepsilon)\). Let \(\Delta = 1 + 2 \sum_{k=1}^{\infty} \Delta_k\), and observe \(1/(1 - q) \leq \Delta\).

We now give a more quantitative form of Theorem 1. For any \(m \geq 1\) and \(\varepsilon > 0\) let \(c_{m, \varepsilon}(N) = N \prod_{i=1}^{m-1} \log_i N \left(\log_m N\right)^{1+\varepsilon}\), where \(\log_1 N = \log N\) and \(\log_m N = \log \log_{m-1} N\) denotes the \(i\)-fold iterated logarithm.

**Theorem 3.** Suppose that \(\nu\) is adapted, and that \((\nu^k)_{\text{abs}} \neq 0\) for some \(k \geq 1\). Let \(f : G \to \mathbb{R}\) be Borel measurable such that \(\int_G f \, d\mu = 0\).

(i) If \(\sup_{c \in G} \mathbb{E}|f(cX_1)|^p < \infty\) for some \(1 \leq p \leq 2\), then for any \(m \geq 1\) and \(\varepsilon > 0\)

\[
\lim_{N \to \infty} \frac{\sum_{k=1}^{N} f(S_k)}{c_{m, \varepsilon}(N)^{1/p}} = 0 \quad \text{a.s.} \tag{3}
\]

(ii) If \(\sup_{c \in G} \mathbb{E}|f(cX_1)|^p < \infty\) for some \(p > 2\), then

\[
\limsup_{N \to \infty} \frac{\left|\sum_{k=1}^{N} f(S_k)\right|}{\sqrt{N \log \log N}} < \infty \quad \text{a.s.} \tag{4}
\]

**Remark.** Note that \(\sup_{c \in G} \mathbb{E}|f(cX_1)|^p < \infty\) clearly implies \(f \in L^p(G)\). To mention a sufficient condition, suppose that \(\nu\) is absolutely continuous with density \(\frac{d\nu}{d\mu}\). If there exists a H"older conjugate pair \(r, s \in [1, \infty]\), \(1/r + 1/s = 1\) such that \(f \in L^r(G)\) and \(\frac{d\nu}{d\mu} \in L^s(G)\), then \(\sup_{c \in G} \mathbb{E}|f(cX_1)|^p \leq \|f\|_r^p \left\|\frac{d\nu}{d\mu}\right\|_s < \infty\).

Under the extra condition that \(\nu\) is strictly aperiodic, we will approximate \(\sum_{k=1}^{N} f(S_k)\) by a sum of independent random variables. Following Strassen [18], we can even construct an almost sure approximation by a Wiener process. To state the precise form of this result, let us introduce the following technical definition.

Given a function \(E(t)\) positive on \((t_0, \infty)\) for some \(t_0\), we shall say that two stochastic processes \(Y(t)\) and \(Z(t)\) in the Skorokhod space \(D[0, \infty)\), possibly defined on different probability spaces, are \(o(E(t))\)-equivalent if there exist finitely many processes \(Y_1(t), Y_2(t), \ldots, Y_m(t)\) in \(D[0, \infty)\) such that \(Y_1(t) = Y(t), Y_m(t) = Z(t),\) and for all \(1 \leq i \leq m - 1\) of the following hold:

(i) The processes \(Y_i(t)\) and \(Y_{i+1}(t)\), possibly defined on different probability spaces, have the same distribution.

(ii) The processes \(Y_i(t)\) and \(Y_{i+1}(t)\) are defined on the same probability space, and \(\lim_{t \to \infty}(Y_i(t) - Y_{i+1}(t))/E(t) = 0\) a.s.
Roughly speaking (ignoring the different probability spaces), \( Y(t) \) and \( Z(t) \) being \( o(E(t)) \)-equivalent thus means \( Y(t) = Z(t) + o(E(t)) \) a.s. Given \( f \in L^2(G) \) with \( \int_G f \, d\mu = 0 \), let

\[
C(f, \nu) = \mathbb{E}f(U)^2 + 2 \sum_{k=1}^{\infty} \mathbb{E}f(U)f(U)S_k, \tag{5}
\]

where \( U \) is a uniformly distributed \( G \)-valued random variable, independent of \( X_1, X_2, \ldots \). As we shall see, the series in (5) is absolutely convergent, and \( C(f, \nu) \geq 0 \).

**Theorem 4.** Suppose that \( \nu \) is adapted and strictly aperiodic, and that \( (\nu^k)_{\text{abs}} \neq 0 \) for some \( k \geq 1 \). Let \( f : G \to \mathbb{R} \) be Borel measurable such that \( \int_G f \, d\mu = 0 \). If \( \sup_{c \in G} \mathbb{E}|f(cX_1)|^{2+\delta} < \infty \) for some \( 0 < \delta < 2 \), then the processes \( \sum_{k \leq t} f(S_k) \) and \( \sqrt{C(f, \nu)}W(t) \) are \( o(t^{1/2-\delta/20}) \)-equivalent, where \( W(t) \) is a standard Wiener process.

The almost sure approximation by a Wiener process yields even more precise asymptotics than those in Theorem 3; for instance, it shows that for strictly aperiodic \( \nu \) the value of the limsup in (4) is \( \sqrt{2C(f, \nu)} \). The almost sure asymptotics as well as the limit distribution of continuous functionals of the process \( \sum_{k \leq t} f(S_k) \) also follow. Instead of the random step functions \( \sum_{k \leq t} f(S_k) \), we could have also used the piecewise linear functions \( \sum_{k \leq [t]} f(S_k) + (t - [t])f(S_{[t]+1}) \). In that case the \( o(t^{1/2-\delta/20}) \)-equivalence holds in the space \( C[0, \infty) \) as well.

In Theorem 4 in general we only know \( C(f, \nu) \geq 0 \); in the case \( C(f, \nu) = 0 \) the result simply states that \( \sum_{k \leq t} f(S_k) = o(t^{1/2-\delta/20}) \) a.s. The natural question whether \( C(f, \nu) > 0 \) is surprisingly delicate. We shall prove a necessary and sufficient condition in terms of irreducible unitary representations of \( G \), see Proposition 7 below. As this condition is rather cumbersome to use, we also give simpler criteria to ensure \( C(f, \nu) > 0 \). In particular, we will show that under mild technical assumptions (e.g. \( f \) is a class function or \( \nu \) is a central measure) we have \( C(f, \nu) = 0 \) if and only if \( f = 0 \) \( \mu \)-a.e. We work out the details in Section 3.

It clearly follows from Theorem 4 that \( N^{-1/2} \sum_{k=1}^{N} f(S_k) \) has a (possibly degenerate) Gaussian limit distribution. Under slightly weaker assumptions than those in Theorem 4, we also prove a Berry–Esseen type bound on its distance from the normal distribution in the Kolmogorov metric. Let \( \Phi(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} e^{-t^2/2} dt \) denote the standard normal distribution function.

**Theorem 5.** Suppose that \( \nu \) is adapted and strictly aperiodic, and that \( (\nu^k)_{\text{abs}} \neq 0 \) for some \( k \geq 1 \). Let \( f : G \to \mathbb{R} \) be Borel measurable such that \( \int_G f \, d\mu = 0 \), and assume \( f \in L^{2+\delta}(G) \) for some \( 0 < \delta \leq 1 \) and \( \sup_{c \in G} \mathbb{E}|f(cX_1)|^2 < \infty \).
(i) If \( C(f, \nu) > 0 \), then for any integer \( N \geq N_0(f, \nu, \delta) \) we have the Berry–Esseen type bound

\[
\sup_{x \in \mathbb{R}} \left| \Pr \left( \frac{\sum_{k=1}^{N} f(S_k)}{\sqrt{C(f, \nu)N}} < x \right) - \Phi(x) \right| \ll K \frac{\log^{\delta/(1+\delta)} N}{N^{\delta/(2+2\delta)}},
\]

where \( K = \Delta \left( \|f\|_{2+\delta}/\sqrt{C(f, \nu)} \right)^{(2+\delta)/(1+\delta)}. \)

(ii) If \( C(f, \nu) = 0 \), then \( N^{-1/2} \sum_{k=1}^{N} f(S_k) \to 0 \) in \( L^2 \).

Remark. If \( f \in L^3(G) \), then the right hand side of (6) becomes \( KN^{-1/4} \log^{1/2} N \) with \( K = \Delta \|f\|_3^3/C(f, \nu)^{3/4} \). We mention that if \( f \in L^4(G) \), then here \( \|f\|_3 \) can be replaced by \( \|f\|_2 \) (see the end of Section 5.2). As we will see in Proposition 8 below, if \( f \) is a class function or \( \nu \) is a central measure, then \( C(f, \nu) \geq \|f\|_2^2/(2\Delta) \). Thus in this case \( K \leq 2\Delta^{7/4} \), so the right hand side of (6) does not depend on \( f \). \( (N_0(f, \nu, \delta) \) always depends on \( f \), however.)

Our results should be compared to classical ergodic theorems for Markov chains. See [13, Chapter 17] for the central limit theorem and the law of the iterated logarithm for discrete time Markov chains on a general state space; analogous results for continuous time Markov processes are given in [11]. Note that adaptedness and strict aperiodicity in our theorems play similar roles as irreducibility and aperiodicity in Markov chains. In addition, the definition (5) of the normalizing constant \( C(f, \nu) \) in the variance in our theorems is exactly the same as in analogous results for Markov chains. In the central limit theorem for Markov chains, under certain technical assumptions the Kolmogorov distance from the normal distribution is known to be of the optimal order \( O(N^{-1/2}) \); see [6] for the first result of this type, for Markov chains on a countable state space under a third moment condition on the return times. In contrast, we do not know if the Berry–Esseen bound \( O(N^{-1/4} \log^{1/2} N) \) in Theorem 5 is optimal or not. Whether this bound, perhaps under stronger conditions, can be improved to \( O(N^{-1/2}) \) remains open.

## 3 Moment estimates

Throughout this section we assume that \( \nu \) is adapted and strictly aperiodic, and that \( (\nu^k)_{abs} \neq 0 \) for some \( k \geq 1 \). Further, we fix a Borel measurable function \( f : G \to \mathbb{R} \) such that \( \int_G f \, d\mu = 0 \), and a uniformly distributed \( G \)-valued random variable \( U \) independent of \( X_1, X_2, \ldots \). We now prove moment estimates for the modified sum \( \sum_{k=1}^{N} f(U S_k) \). In Section 4 we shall give the counterparts of these estimates for shifted sums \( \sum_{k=M+1}^{M+N} f(S_k) \).
For every nonempty, finite interval of positive integers $J \subset \mathbb{N}$ let $S_J = \prod_{j \in J} X_j$. Note that $S_J$ has distribution $\nu^{S_J}$, hence by definition for any bounded, Borel measurable function $g : G \to \mathbb{R}$

$$|\mathbb{E} g(S_J) - \mathbb{E} g(U)| = \left| \int_G g \, d(\nu^{S_J} - \mu) \right| \leq \sup_G |g| \cdot \Delta_{|J|}. \quad (7)$$

**Proposition 6.** Assume $f \in L^2(G)$. The series in (5) is absolutely convergent, and $0 \leq C(f, \nu) \leq \|f\|_2^2 \Delta$. Further, for any integer $N \geq 1$

$$\mathbb{E} \left( \sum_{k=1}^N f(US_k) \right)^2 = C(f, \nu) N + O_\nu \left( \|f\|_2^4 \right) \leq \|f\|_2^2 \Delta N.$$  

**Proof.** Let $A_k = \mathbb{E} f(U) f(US_k)$. Since $U$ is independent of $S_k$, we have

$$A_k = \int_G \int_G f(u) f(ux) \, d\mu(u) \, d\nu^x(x). \quad (8)$$

The function $g(x) = \int_G f(u) f(ux) \, d\mu(u)$ satisfies $\int_G g \, d\mu = 0$ and $\sup_G |g| \leq \|f\|_2^2$. Applying (7) to $g$ we thus obtain $|A_k| \leq \|f\|_2^2 \Delta_k$. Since $\Delta_k \to 0$ exponentially fast, the series in (5) is absolutely convergent. As $\mathbb{E} f(U)^2 = \|f\|_2^2$, we have $C(f, \nu) \leq \|f\|_2^2 \Delta$. Finally, $C(f, \nu) \geq 0$ will follow from the second claim.

Expanding the square we have

$$\mathbb{E} \left( \sum_{k=1}^N f(US_k) \right)^2 = \sum_{k=1}^N \mathbb{E} f(US_k)^2 + 2 \sum_{1 \leq k < \ell \leq N} \mathbb{E} f(US_k) f(US_\ell). \quad (9)$$

Let us write $US_\ell = US_k S_{k+1, \ell}$. Since $\mu * \nu^x = \mu$, the variable $US_k$ is uniformly distributed on $G$ and independent of $S_{k+1, \ell}$; moreover, $S_{k+1, \ell} \overset{d}{=} S_{\ell-k}$. Thus $\mathbb{E} f(US_k)^2 = \mathbb{E} f(U)^2$ and $\mathbb{E} f(US_k) f(US_\ell) = \mathbb{E} f(U) f(US_{\ell-k})$, so (9) simplifies as

$$\mathbb{E} \left( \sum_{k=1}^N f(US_k) \right)^2 = N \mathbb{E} f(U)^2 + 2 \sum_{1 \leq k < \ell \leq N} \mathbb{E} f(U) f(US_{\ell-k})$$

$$= N \mathbb{E} f(U)^2 + 2 \sum_{d=1}^{N-1} (N-d) A_d$$

$$= C(f, \nu) N + O(\sum_{d=1}^{N-1} d |A_d| + N \sum_{d=N}^{\infty} |A_d|)$$

$$\leq N \mathbb{E} f(U)^2 + 2N \sum_{d=1}^{N-1} |A_d|.$$
The second claim thus follows from $|A_d| \leq \|f\|_2^2 \Delta_d$ and the fact that $\Delta_d \to 0$ exponentially fast. \qed

We now study the question whether the normalizing factor $C(f, \nu)$ in the variance is zero or positive. To this end, we derive an alternative formula for $C(f, \nu)$ in the form of an infinite series with nonnegative terms. Next, we will consider the special case when $f$ is a class function or $\nu$ is a central measure. As we shall see, the behavior of $C(f, \nu)$ then simplifies considerably, allowing for effective lower bounds.

**Proposition 7.** Assume $f \in L^2(G)$. We have

$$C(f, \nu) = \sum_{\pi \in \hat{G}\setminus\{\pi_0\}} d_\pi \text{tr} \left( \hat{f}(\pi)^* \hat{f}(\pi) B_\nu(\pi) \right),$$

(10)

where $B_\nu(\pi) = I_{d_\nu} + (I_{d_\nu} - \hat{\nu}(\pi))^{-1} \hat{\nu}(\pi) + (I_{d_\nu} - \hat{\nu}(\pi)^*)^{-1} \hat{\nu}(\pi)^*$ and $I_{d_\nu}$ denotes the $d_\pi \times d_\pi$ identity matrix. The series in (10) has nonnegative terms and is convergent. In particular, $C(f, \nu) > 0$ if and only if at least one term in (10) is nonzero.

**Proof.** Let $A_k = \mathbb{E} f(U) f(U S_k)$, and recall (8). The function $h_k(u) = \int_G f(ux) \, d\nu^k(x)$ is in $L^2(G)$, and its Fourier coefficients are $\hat{h}_k(\pi) = \hat{f}(\pi) (\hat{\nu}(\pi)^k)^*$. Applying the Parseval formula in (8) we thus obtain

$$A_k = \int_G f(u) h_k(u) \, d\mu(u) = \sum_{\pi \in \hat{G}} d_\pi \text{tr} \left( \hat{f}(\pi) \hat{\nu}(\pi)^k \hat{f}(\pi)^* \right).$$

(11)

Here $\hat{f}(\pi_0) = 0$ as $\int_G f \, d\mu = 0$. Fix $\pi \in \hat{G}$, $\pi \neq \pi_0$. For any $v \in \mathbb{C}^{d_\pi}$ we have $\langle \hat{\nu}(\pi)^k v, v \rangle = \int_G (\pi(x)^* v, v) \, d\nu^k(x)$, where $(a, b) = \sum_{i=1}^{d_\pi} a_i b_i$ denotes the standard sesquilinear form on $\mathbb{C}^{d_\pi}$. Since $\pi(x)^*$ is unitary, the integrand $g(x) = (\pi(x)^* v, v)$ satisfies $\sup_G |g| \leq |v|^2$. Further, $\int_G g \, d\mu = \langle \hat{\nu}(\pi)^k v, v \rangle = 0$ because $\pi \neq \pi_0$. Applying (7) to $g$ we thus obtain $|\langle \hat{\nu}(\pi)^k v, v \rangle| \leq |v|^2 \Delta_k$. In particular, for any $v \in \mathbb{C}^{d_\pi}$ we have $|\langle \hat{f}(\pi) \hat{\nu}(\pi)^k \hat{f}(\pi)^* v, v \rangle| \leq (\hat{f}(\pi) \hat{\nu}(\pi)^k \hat{f}(\pi)^*) \Delta_k$. Summing this over an orthonormal basis of $\mathbb{C}^{d_\pi}$ we get $|\text{tr} \left( \hat{f}(\pi) \hat{\nu}(\pi)^k \hat{f}(\pi)^* \right)| \leq \text{tr} \left( \hat{f}(\pi) \hat{\nu}(\pi)^k \hat{f}(\pi)^* \right) \Delta_k$. Hence

$$\sum_{k=1}^{\infty} \sum_{\pi \in \hat{G}\setminus\{\pi_0\}} d_\pi \left| \text{tr} \left( \hat{f}(\pi) \hat{\nu}(\pi)^k \hat{f}(\pi)^* \right) \right| \leq \|f\|_2^2 \sum_{k=1}^{\infty} \Delta_k < \infty,$$

justifying a change in the order of summation. Since $\rho(\hat{\nu}(\pi)) \leq q < 1$, we have $\sum_{k=1}^{\infty} \hat{\nu}(\pi)^k = (I_{d_\nu} - \hat{\nu}(\pi))^{-1} \hat{\nu}(\pi)$ in operator norm. We thus obtain

$$C(f, \nu) = \|f\|_2^2 + 2 \sum_{k=1}^{\infty} A_k = \sum_{\pi \in \hat{G}\setminus\{\pi_0\}} d_\pi \text{tr} \left( \hat{f}(\pi)^* \hat{f}(\pi) \left( I_{d_\nu} + 2(I_{d_\nu} - \hat{\nu}(\pi))^{-1} \hat{\nu}(\pi) \right) \right).$$

11
As \(C(f, \nu)\) is clearly real, we can take the real part of the series in the previous line, resulting in formula (10).

Finally, we prove that every term in (10) is nonnegative. Fix \(\pi \in \hat{G} \setminus \{\pi_0\}\). First, suppose that \(\pi\) and \(\pi\) are unitarily inequivalent. Then we may assume that \(\pi, \pi \in \hat{G}\). Since \(f\) and \(\nu\) are real-valued, we have \(\hat{f}(\pi) = \overline{f(\pi)}\) and \(\hat{\nu}(\pi) = \overline{\nu(\pi)}\). Hence \(B_\nu(\pi) = \overline{B_\nu(\pi)}\), and the terms in (10) indexed by \(\pi\) and \(\overline{\pi}\) are equal. Let \(F\) be the orthogonal projection of \(f\) in \(L^2(G)\) to the linear subspace spanned by the matrix elements \(\{\pi_{ij} : 1 \leq i, j \leq d_\pi\} \cup \{\overline{\pi}_{ij} : 1 \leq i, j \leq d_\pi\}\); that is, \(F(x) = d_\pi \text{tr} \left( \hat{f}(\pi) \pi(x) \right) + d_\pi \text{tr} \left( \hat{\nu}(\pi) \overline{\pi}(x) \right)\). Note that \(F\) is real-valued, \(\hat{F}(\pi) = \hat{f}(\pi)\), \(\hat{F}(\pi') = 0\) for all \(\pi' \neq \pi, \pi\). Therefore the terms in (10) indexed by \(\pi\) and \(\overline{\pi}\) are both \(C(F, \nu)/2\). But \(C(F, \nu) \geq 0\) from Proposition 6, and we are done. Next, suppose that \(\pi\) and \(\overline{\pi}\) are unitarily equivalent. Let \(F\) be the orthogonal projection of \(f\) in \(L^2(G)\) to the linear subspace spanned by the matrix elements \(\{\pi_{ij} : 1 \leq i, j \leq d_\pi\}\). Then \(\{\overline{\pi}_{ij} + \lambda \pi, 1 \leq i, j \leq d_\pi\}\) span the same linear subspace. Thus \(F(x) = d_\pi \text{tr} \left( \hat{f}(\pi) \pi(x) \right) = d_\pi \text{tr} \left( \hat{f}(\overline{\pi}) \overline{\pi}(x) \right)\). Again, \(F\) is real-valued, \(\hat{F}(\pi) = \hat{f}(\pi)\) and \(\hat{F}(\pi') = 0\) for all \(\pi' \neq \pi\). Therefore the term in (10) indexed by \(\pi\) is \(C(F, \nu) \geq 0\).

\[\text{Proposition 8. Assume } f \in L^2(G), \text{ and let } \nu^*(B) = \nu(B^{-1}) (B \subseteq G \text{ Borel}) \text{ denote the distribution of } X_1^{-1}. \text{ Suppose at least one of the following hold.}\]

(i) \(f\) is a class function

(ii) \(\nu \ast \nu^* = \nu^* \ast \nu\)

(iii) \(\nu\) is a central measure

Then \(\frac{1-a}{1-a^q} \|f\|^2 \leq C(f, \nu) \leq \frac{1+a}{1-a^q} \|f\|^2\). In particular, \(C(f, \nu) = 0\) if and only if \(f = 0\) \(\mu\)-a.e.

\[\text{Proof.} \text{ First, assume (i). It follows from Schur’s lemma that } \hat{f}(\pi) \text{ is a scalar multiple of the identity matrix. Hence (10) simplifies as}\]

\[C(f, \nu) = \sum_{\pi \in \hat{G}\setminus\{\pi_0\}} \text{tr} \left( \hat{f}(\pi)^* \hat{f}(\pi) \right) \text{tr} B_\nu(\pi).\]

Let \(\lambda_1, \lambda_2, \ldots, \lambda_d\) denote the eigenvalues of \(\hat{\nu}(\pi)\). Then

\[\text{tr} B_\nu(\pi) = \sum_{i=1}^{d_\pi} \left( 1 + \frac{\lambda_i}{1 - \lambda_i} + \frac{\overline{\lambda_i}}{1 - \overline{\lambda_i}} \right) = \sum_{i=1}^{d_\pi} \frac{1 - |\lambda_i|^2}{|1 - \lambda_i|^2}.\]
Since \( \rho(\hat{\nu}(\pi)) \leq q < 1 \), we have \(|\lambda_i| \leq q\), and so \( d_{\pi}^{1-q} \leq \text{tr} B_\nu(\pi) \leq d_{\pi}^{1+q} \). The claim thus follows from the Parseval formula.

Next, assume (ii). Since \( \hat{\nu}^*(\pi) = \hat{\nu}(\pi)^* \), the condition \( \nu * \nu^* = \nu^* * \nu \) implies that the matrix \( \hat{\nu}(\pi) \) is normal. Therefore there exists an orthonormal basis \( v_1, v_2, \ldots, v_{d_\pi} \) of \( \mathbb{C}^{d_\pi} \) comprised of eigenvectors of \( \hat{\nu}(\pi) \); say, \( \hat{\nu}(\pi)v_i = \lambda_i v_i \). It follows that \( \hat{\nu}(\pi)^* v_i = \overline{\lambda_i} v_i \), and hence

\[
B_\nu(\pi)v_i = \left(1 + \frac{\lambda_i}{1 - \lambda_i} + \frac{\overline{\lambda_i}}{1 - \overline{\lambda_i}}\right) v_i = 1 - \frac{|\lambda_i|^2}{|1 - \lambda_i|^2} v_i.
\]

The eigenvalues of \( B_\nu(\pi) \) again satisfy \( \frac{1-q}{1+q} \leq \frac{1-|\lambda_i|^2}{|1-\lambda_i|^2} \leq \frac{1+q}{1-q} \). It is now easy to see that

\[
\frac{1-q}{1+q} \text{tr} \left(\hat{f}(\pi)^* \hat{f}(\pi)\right) \leq \text{tr} \left(\hat{f}(\pi)^* \hat{f}(\pi) B_\nu(\pi)\right) \leq \frac{1+q}{1-q} \text{tr} \left(\hat{f}(\pi)^* \hat{f}(\pi)\right),
\]

and the claim follows. Finally, note that condition (iii) implies condition (ii). \( \square \)

We conclude this section with an estimate of the \( L^p \)-norm for \( 1 \leq p \leq 4 \). These estimates, combined with the Erdős–Stechkin and the Rademacher–Menshov inequalities will help us bound the fluctuations of \( \sum_{k=1}^N f(S_k) \) as \( N \) runs in a short interval. Additionally, we will also use them to verify the Lyapunov condition in the proof of Theorem 5.

**Proposition 9.** Assume \( f \in L^p(G) \) for some \( 1 \leq p \leq 4 \). For any integer \( N \geq 1 \)

\[
\left\| \sum_{k=1}^N f(U S_k) \right\|_p \ll \left\{ \begin{array}{ll}
\|f\|_p (\Delta N)^{1/p} & \text{if } 1 \leq p < 2, \\
\|f\|_p \sqrt{\Delta N} & \text{if } 2 \leq p \leq 4.
\end{array} \right.
\]

In the case \( p = 4 \) we also have

\[
\left\| \sum_{k=1}^N f(U S_k) \right\|_4 \ll \|f\|_2 \sqrt{\Delta N} + \|f\|_4 \Delta^{3/4} N^{1/4}. \tag{12}
\]

**Proof.** First, assume \( p = 4 \). Expanding the fourth power we get

\[
\mathbb{E} \left( \sum_{k=1}^N f(U S_k) \right)^4 \ll \sum_{1 \leq k_1 \leq k_2 \leq k_3 \leq k_4 \leq N} |\mathbb{E} f(U S_{k_1}) f(U S_{k_2}) f(U S_{k_3}) f(U S_{k_4})|.
\]

Fix \( 1 \leq k_1 \leq k_2 \leq k_3 \leq k_4 \leq N \). Since \( U S_{k_i} \) is uniformly distributed on \( G \) and independent of \( X_{k_1+1}, X_{k_1+2}, \ldots \), we have

\[
\mathbb{E} f(U S_{k_1}) f(U S_{k_2}) f(U S_{k_3}) f(U S_{k_4}) = \\
\int_G \int_G \int_G \int_G f(u) f(u x) f(u x y) f(u x y z) d\mu(u) d\nu^{(k_2-k_1)}(x) d\nu^{(k_3-k_2)}(y) d\nu^{(k_4-k_3)}(z).
\]
Here we use the convention that $\nu^*0$ is the Dirac measure concentrated on the identity element of $G$, and $\Delta_0 = \|\nu^*0 - \mu\|_{TV} \leq 2$. Let

$$g(z) = \int_G \int_G \int_G f(u) f(wx) f(uxy) f(uxyz) d\mu(u) d\nu^*(k_2-k_1)(x) d\nu^*(k_3-k_2)(y).$$

As $\int_G g(z) d\mu(z) = 0$, applying (7) to $g$ we obtain

$$|\mathbb{E} f(U S_{k_1}) f(U S_{k_2}) f(U S_{k_3}) f(U S_{k_4})| \leq \sup_G |g| \cdot \Delta_{k_4-k_3}. \quad (14)$$

Fix $z \in G$, and let $h_z(y) = \int_G \int_G f(u) f(wx) f(uxy) f(uxyz) d\mu(u) d\nu^*(k_2-k_1)(x)$. Note that

$$\int_G h_z(y) d\mu(y) = \int_G \int_G f(u) f(wx) w_z d\mu(u) d\nu^*(k_2-k_1)(x),$$

where $w_z = \int_G f(uxy) f(uxyz) d\mu(y) = \int_G f(y) f(yz) d\mu(y)$ does not depend on $u$ and $x$. Applying (7) to $h_z$ we get

$$|g(z)| = \left| \int_G h_z(y) d\nu^*(k_3-k_2)(y) \right| \leq |w_z| \left| \int_G \int_G f(u) f(wx) d\mu(u) d\nu^*(k_2-k_1)(x) \right| + \sup_G |h_z| \cdot \Delta_{k_3-k_2}.$$

Here $|w_z| \leq \|f\|_2^2$, and the double integral in the previous line is $\leq \|f\|_2^2 \Delta_{k_2-k_1}$, as seen in the proof of Proposition 6. Hence

$$|g(z)| \leq \|f\|_2^2 \Delta_{k_2-k_1} + \sup_G |h_z| \cdot \Delta_{k_3-k_2}. \quad (15)$$

Now fix $y, z \in G$, and let $r_{y,z}(x) = \int_G f(u) f(wx) f(uxy) f(uxyz) d\mu(u)$. Note that $\sup_G |r_{y,z}| \leq \|f\|_4^4$, and that $\int_G r_{y,z}(x) d\mu(x) = 0$. Applying (7) we thus get

$$|h_z(y)| = \left| \int_G r_{y,z}(x) d\nu^*(k_2-k_1)(x) \right| \leq \|f\|_4^4 \Delta_{k_2-k_1}. \quad (16)$$

Combining (14), (15) and (16) we finally obtain

$$|\mathbb{E} f(U S_{k_1}) f(U S_{k_2}) f(U S_{k_3}) f(U S_{k_4})| \leq \|f\|_2^2 \Delta_{k_2-k_1} \Delta_{k_4-k_3} + \|f\|_4^4 \Delta_{k_2-k_1} \Delta_{k_3-k_2} \Delta_{k_4-k_3}.$$

Summing over $1 \leq k_1 \leq k_2 \leq k_3 \leq k_4 \leq N$, (12) follows.

On the other hand, we can use $\Delta_{k_3-k_2} \leq 2$ to deduce the simpler estimate

$$|\mathbb{E} f(U S_{k_1}) f(U S_{k_2}) f(U S_{k_3}) f(U S_{k_4})| \leq 3 \|f\|_4^4 \Delta_{k_2-k_1} \Delta_{k_4-k_3}.$$
and by summing over $1 \leq k_1 \leq k_2 \leq k_3 \leq k_4 \leq N$ we get $\|\sum_{k=1}^{N} f(U S_k)\|_4 \ll \|f\|_4 \sqrt{\Delta N}$. Proposition 6 shows that if $f \in L^2(G)$, the same estimate holds with $\|\cdot\|_4$ replaced by $\|\cdot\|_2$ on both sides. Moreover, we also have the trivial estimate $\left\| \sum_{k=1}^{N} f(U S_k) \right\|_1 \leq \|f\|_1 N$ for any $f \in L^1(G)$. This settles the endpoints of the intervals $1 \leq p \leq 2$ and $2 \leq p \leq 4$. The cases $1 < p < 2$ and $2 < p < 4$ follow from the Riesz–Thorin interpolation theorem applied to the linear operator $f \mapsto \sum_{k=1}^{N} f(U S_k) - N \int_{G} f \, d\mu$. 

\section{Approximation by independent variables}

Assume again, that $\nu$ is adapted and strictly aperiodic, and that $(\nu^*k)_{\text{abs}} \neq 0$ for some $k \geq 1$. Fix a Borel measurable function $f : G \to \mathbb{R}$ such that $\int_{G} f \, d\mu = 0$. In this section we approximate the shifted sum $\sum_{k=M+1}^{M+N} f(S_k)$ by a sum of independent random variables. The main tool of this approximation is a coupling between $\nu^*k$ and $\mu$, which we will construct using Strassen’s theorem. We mention that this is the only step of the proof where we use the fact that $G$ is metrizable. A similar approach was used by Schatte [17] on the circle group $G = \mathbb{R}/\mathbb{Z}$, with a different type of coupling based on the Kolmogorov metric instead of the total variation metric.

Recall that for any two Borel probability measures $\nu_1$ and $\nu_2$ on $G$ (or indeed, on any Polish space) we have $\|\nu_1 - \nu_2\|_{TV} = 2 \inf_{\vartheta} \vartheta(\{(x, y) \in G \times G : x \neq y\})$, where the infimum is over all Borel probability measures $\vartheta$ on $G \times G$ whose marginals are $\vartheta(B \times G) = \nu_1(B)$ and $\vartheta(G \times B) = \nu_2(B)$. This fact follows from Strassen’s theorem, which in turn is a special case of the Kantorovich duality theorem in the theory of optimal transportation (see e.g. [21, Chapter 1]). In particular, for any $k \geq 1$ there exists a Borel probability measure $\vartheta_k$ on $G \times G$ with marginals $\nu^*k$ and $\mu$ such that $\vartheta_k(\{(x, y) \in G \times G : x \neq y\}) \leq \Delta_k$. After a suitable extension of the probability space, for any nonempty, finite interval of positive integers $J \subseteq \mathbb{N}$ we may therefore introduce auxiliary $G$-valued random variables $T_J, U_J$ whose joint distribution is $\vartheta_{|J}$; that is, $T_J \overset{d}{=} S_J$, $U_J$ is uniformly distributed on $G$, and $\Pr(T_J \neq U_J) \leq \Delta_{|J|}$. Moreover, we may assume $(T_J, U_J)$, $J \subseteq \mathbb{N}$ and $X_1, X_2, \ldots$ are independent. The independence of the approximating variables will follow from the following observation.

\textbf{Lemma 1.} Let $G$ be a compact metrizable group, and let $(S, \mathcal{A})$ be a measurable space. Let $U$ be a $G$-valued, and let $V$ be an $S$-valued random variable. If $U$ and $V$ are independent and $U$ is uniformly distributed on $G$, then for any Borel measurable function $g : S \to G$ the variables $g(V)U$ and $V$ are also independent.
**Proof.** Note that $g(V)U$ is uniformly distributed on $G$. Let $\gamma$ denote the distribution of $V$. For any Borel set $B \subseteq G$ and any $A \in \mathcal{A}$ we have

$$
\Pr (g(V)U \in B, V \in A) = \int_G \int_S I_B(g(v)u)I_A(v) \, d\mu(u) \, d\gamma(v) \\
= \int_G \int_S I_{g(v)-1}(u)I_A(v) \, d\mu(u) \, d\gamma(v) \\
= \int_S \mu(B)I_A(v) \, d\gamma(v) \\
= \Pr (g(V)U \in B) \Pr(V \in A).
$$

\qed

We construct the approximating variables as follows. Fix an integer $M \geq 0$, and let us decompose the infinite set $\{M + 1, M + 2, \ldots\}$ into consecutive, nonempty, finite intervals of integers $H_1, J_1, H_2, J_2, \ldots$. For all $i \geq 1$ and $k \in J_i$, let

$$
W_k = S_M \prod_{j=1}^{i-1} (T_{H_j}S_{J_j}) T_{H_i} \prod_{\ell \in J_i \atop \ell \leq k} X_{\ell}, \quad W_k^* = S_M \prod_{j=1}^{i-1} (T_{H_j}S_{J_j}) U_{H_i} \prod_{\ell \in J_i \atop \ell \leq k} X_{\ell}.
$$

Similarly, for all $i \geq 2$ and $k \in H_i$, let

$$
W_k = S_M \prod_{j=1}^{i-2} (S_{H_j}T_{J_j}) S_{H_{i-1}} T_{J_{i-1}} \prod_{\ell \in H_i \atop \ell \leq k} X_{\ell}, \quad W_k^* = S_M \prod_{j=1}^{i-2} (S_{H_j}T_{J_j}) S_{H_{i-1}} U_{J_{i-1}} \prod_{\ell \in H_i \atop \ell \leq k} X_{\ell}.
$$

Note that here the case $i = 1$ is excluded to ensure that $H_i$ is preceded by an interval $J_{i-1}$. See Figure 1 for an illustration of the construction. Let us also introduce the variables

$$
Y_i = \sum_{k \in J_i} f(W_k), \quad Y_i^* = \sum_{k \in J_i} f(W_k^*) \quad (i \geq 1),
$$

$$
Z_i = \sum_{k \in H_i} f(W_k), \quad Z_i^* = \sum_{k \in H_i} f(W_k^*) \quad (i \geq 2).
$$

Observe that the random sequence $W_k$, $k \in \bigcup_{i=1}^\infty J_i$ has the same distribution as $S_k$, $k \in \bigcup_{i=1}^\infty J_i$. Similarly, $W_k$, $k \in \bigcup_{i=2}^\infty H_i$ has the same distribution as $S_k$, $k \in \bigcup_{i=2}^\infty H_i$.

For every $R \geq 1$ let $N_R$ be such that $M + N_R = \max J_R$. Then $\sum_{k=M+1}^{M+N_R} f(S_k)$ along the subsequence $N_R$ satisfies

$$
\sum_{k=M+1}^{M+N_R} f(S_k) = \sum_{k \in H_1} f(S_k) + \sum_{i=1}^R \sum_{k \in J_i} f(S_k) + \sum_{i=2}^R \sum_{k \in H_i} f(S_k).
$$
Here the sequence $\sum_{i=1}^{R} \sum_{k \in J_{i}} f(S_{k})$, $R = 1, 2, \ldots$ has the same distribution as $\sum_{i=1}^{R} Y_{i}$, $R = 1, 2, \ldots$; similarly, the sequence $\sum_{i=2}^{R} \sum_{k \in H_{i}} f(S_{k})$, $R = 2, 3, \ldots$ has the same distribution as $\sum_{i=2}^{R} Z_{i}$, $R = 2, 3, \ldots$.

The main idea is to replace $Y_{i}$ by $Y_{i}^{\ast}$, and $Z_{i}$ by $Z_{i}^{\ast}$. By construction, $W_{k}^{\ast}$ is obtained from $W_{k}$ by replacing a factor of type $T$ by a factor of type $U$; the uniform distribution of $U$ will ensure the independence of the approximating variables via Lemma 1. On the other hand, the pair $(T, U)$ was constructed so that $T \neq U$ with exponentially small probability; this will ensure that the error of the approximation is small. We formalize this intuition in Lemmas 2 and 3 below.

Our strategy will be to choose the size of $J_{i}$ exponentially larger than that of $H_{i}$, so that the contribution of the variables $Z_{i}^{\ast}$ is negligible. The sum $\sum_{k=1}^{M+N_{R}} f(S_{k})$ will then be well approximated by the sum of independent variables $\sum_{i=1}^{R} Y_{i}^{\ast}$. Applying classical limit theorems to the latter sum, we obtain our results along the subsequence $N_{R}$. The final step in our proofs will be an upper estimate to the fluctuation $\max_{N \in H_{R} \cup J_{R}} \left| \sum_{k=1}^{N} f(S_{k}) \right|$ of our sum between $N_{R}$ and $N_{R+1}$.

**Lemma 2.** $Y_{1}^{\ast}, Y_{2}^{\ast}, \ldots$ are independent, and $\mathbb{E}Y_{i}^{\ast} = 0$.

(i) If $f \in L^{2}(G)$, then $\mathbb{E}(Y_{i}^{\ast})^{2} = C(f, \nu) |J_{i}| + O_{\nu}(\|f\|_{2}^{2})$.

(ii) If $f \in L^{p}(G)$ for some $1 \leq p \leq 4$, then for any $0 \leq R < S$

$$\left\| \sum_{i=R+1}^{S} Y_{i}^{\ast} \right\|_{p} \ll \begin{cases} \|f\|_{p} \left( \Delta \sum_{i=R+1}^{S} |J_{i}| \right)^{1/p} & \text{if } 1 \leq p < 2, \\ \|f\|_{p} \sqrt{\Delta \sum_{i=R+1}^{S} |J_{i}|} & \text{if } 2 \leq p \leq 4. \end{cases}$$

In the case $p = 4$ we also have

$$\left\| \sum_{i=R+1}^{S} Y_{i}^{\ast} \right\|_{4} \ll \|f\|_{2} \sqrt{\Delta \sum_{i=R+1}^{S} |J_{i}|} + \|f\|_{4} \Delta^{3/4} \left( \sum_{i=R+1}^{S} |J_{i}| \right)^{1/4}. \quad (17)$$

The same hold for $Z_{2}^{\ast}, Z_{3}^{\ast}, \ldots$ with $|J_{i}|$ replaced by $|H_{i}|$. 

---

**Figure 1:** The construction of $W_{k}$ and $W_{k}^{\ast}$

| [1, M] | $H_{1}$ | $J_{1}$ | $H_{2}$ | $J_{2}$ | $\ldots$ |
|-------|--------|--------|--------|--------|--------|
| $W_{k}$: | $S$ | $T$ | $S$ | $T$ | $X$ |
| $W_{k}^{\ast}$: | $S$ | $T$ | $S$ | $U$ | $X$ |
yielding (17). On the other hand, Proposition 9 also gives
\[ \| Y_i^* \|_{4} \]
and so endpoints of the intervals \( 1 \leq R < S \), the sum
\[ \| Y_i^* \|_4 = \left( \sum_{k=1}^{[J_i]} f(U_k) \right) \| f \|_2 \Delta \| J_i \| \]
yields (17). On the other hand, Proposition 9 also gives \( \| Y_i^* \|_4 \ll \| f \|_4 \Delta \| J_i \| \),
and so \( \left( \sum_{i=R+1}^{S} Y_i^* \right) \ll \| f \|_4 \Delta \sum_{i=R+1}^{S} |J_i| \)
follows as well. This settles the independence of the endpoints 1 \( \leq p \leq 2 \) and 2 \( \leq p \leq 4 \).

Observe that for a given integer \( M \geq 0 \), given intervals \( H_1, J_1, \ldots \) and given 0 \( \leq R < S \), the sum \( \sum_{i=R+1}^{S} Y_i^* \)
is linear in \( f \). Applying the Riesz–Thorin interpolation theorem to the linear operator
\[ f \mapsto \sum_{i=R+1}^{S} Y_i^* - \left( \sum_{i=R+1}^{S} |J_i| \right) \int_{\mathcal{J}_i} f \, d\mu, \]
the cases \( 1 < p < 2 \) and 2 \( \leq p < 4 \) follow. The proof for \( Z_2^*, Z_3^*, \ldots \) is analogous.

**Lemma 3.** If \( L_p = \sup_{c \in G} \mathbb{E}|f(cX_1)|^p < \infty \) for some \( p \geq 1 \), then \( \| Y_i - Y_i^* \|_p \leq 2|J_i| \left( L_p \Delta |J_i| \right)^{1/p} \) and \( \| Z_i - Z_i^* \|_p \leq 2|H_i| \left( L_p \Delta |J_i| \right)^{1/p} \).

**Proof.** We have \( \| Y_i - Y_i^* \|_p \leq \sum_{k \in J_i} |f(W_k) - f(W_k^*)| \). Let \( \mathcal{F} \) be the \( \sigma \)-algebra generated by \( S_M \prod_{j=1}^{[J_i]} (T_{H_j}, S_{J_i}^j), T_{H_i}, U_{H_i} \) and \( X_{\ell}, \ell \in J_i, \ell < k \). Then \( W_k = aX_k \) and \( W_k^* = a^*X_k \) with some \( \mathcal{F} \)-measurable random variables \( a, a^* \). Note that if \( T_{H_i} = U_{H_i}, \) then \( W_k = W_k^* \). Therefore \( \mathbb{E}(|f(W_k) - f(W_k^*)|^p \mid \mathcal{F}) \leq 2^p L_p I_{\{T_{H_i} \neq U_{H_i}\}}. \)
Taking the (total) expectation we get
\[ E |f(W_k) - f(W_k^*)|^p \leq 2^p L_p \Pr(T_{H_i} \neq U_{H_i}) \leq 2^p L_p \Delta_{|H_i|}, \]
and the result follows. The proof for \( \|Z_i - Z_i^*\|_p \) is analogous.

As a simple application of the approximating variables constructed above, we deduce moment estimates for shifted sums \( \sum_{k=M+1}^{M+N} f(S_k) \) from the results of Section 3.

**Corollary 10.**

(i) If \( \operatorname{sup}_{c \in \mathbb{C}} \mathbb{E} f(cX_1)^2 < \infty \), then for any integers \( M \geq 0 \) and \( N \geq 1 \)
\[ \left\| \sum_{k=M+1}^{M+N} f(S_k) \right\|_2 = \sqrt{C(f, \nu)N + O_{f, \nu} (\log(N + 1))}. \]

(ii) If \( \operatorname{sup}_{c \in \mathbb{C}} \mathbb{E} |f(cX_1)|^p < \infty \) for some \( 1 \leq p \leq 4 \), then for any integers \( M \geq 0 \) and \( N \geq 1 \)
\[ \left\| \sum_{k=M+1}^{M+N} f(S_k) \right\|_p \ll K_{f, \nu, p} \log(N + 1) + \begin{cases} \|f\|_p (\Delta N)^{1/p} & \text{if } 1 \leq p < 2, \\ \|f\|_p \sqrt{\Delta N} & \text{if } 2 \leq p \leq 4. \end{cases} \]

with some constant \( K_{f, \nu, p} > 0 \). In the case \( p = 4 \) we also have
\[ \left\| \sum_{k=M+1}^{M+N} f(S_k) \right\|_4 \ll K_{f, \nu, p} \log(N + 1) + \|f\|_2 \sqrt{\Delta N} + \|f\|_4 \Delta^{3/4} N^{1/4}. \]

**Proof.** We may assume that \( N \) is large enough in terms of \( f, \nu \) and \( p \). Let us decompose the index set \([M + 1, M + N]\) into two consecutive intervals of integers \( H_1 \) and \( I_1 \) such that \( |H_1| = [4\Delta \log N] \). We then have \( \sum_{k=M+1}^{M+N} f(S_k) = \sum_{k \in H_1} f(S_k) + \sum_{k \in I_1} f(S_k) \), where \( \sum_{k \in I_1} f(S_k) \overset{d}{=} Y_1 \). To see (i), let us write
\[ \left\| \sum_{k=M+1}^{M+N} f(S_k) \right\|_2 = \|Y_1\|_2 + O \left( \|Y_1 - Y_1^*\|_2 + \sum_{k \in H_1} \|f(S_k)\|_2 \right). \]

By Lemma 2 (i), here \( \|Y_1\|_2 = \sqrt{C(f, \nu)N + O_{f, \nu}(1)} \). Since, say, \( \Delta_k^{1/k} \leq (1 + q)/2 \) for \( k \geq k_0(\nu) \) and \((1 + q)/2\Delta \leq ((1 + q)/2)^{1/(1-q)} \leq e^{-1/2} \), we have \( \Delta_{|H_1|} \leq N^{-2} \). Lemma 3 thus gives \( \|Y_1 - Y_1^*\|_2 \ll_{f, \nu} 1 \). Finally, note that \( \operatorname{sup}_{c \in \mathbb{C}} \mathbb{E} f(cX_1)^2 < \infty \)
implies \( \operatorname{sup}_{c \geq 1} \mathbb{E} f(S_k)^2 < \infty \). Hence \( \sum_{k \in H_1} \|f(S_k)\|_2 \ll_{f, \nu} \log N \), and (i) follows. If we use Lemma 2 (ii) instead of Lemma 2 (i), similar arguments show (ii).

**Remark.** We could easily improve the error term \( K_{f, \nu, p} \log(N + 1) \) in (ii) by decomposing \([M + 1, M + N]\) into more than 2 consecutive intervals of exponentially increasing sizes.
5 Proof of the theorems

We prove Theorem 3 (i) for strictly aperiodic measures and Theorem 4 in Section 5.1; the general case of Theorem 3 and Theorem 1 in Section 5.2; finally, Theorem 5 and the Remark thereafter, and Theorem 2 in Section 5.3.

5.1 Almost sure asymptotics, strictly aperiodic case

Suppose that \( \nu \) is adapted and strictly aperiodic, and that \((\nu^* k)_{\text{abs}} \neq 0\) for some \( k \geq 1 \). Let \( f : G \to \mathbb{R} \) be Borel measurable such that \( \int_G f \, d\mu = 0 \), and assume \( \sup_{c \in G} E|f(cX_1)|^p < \infty \). In this section we prove the strong law of large numbers (3) in the case \( 1 \leq p \leq 2 \), and the almost sure approximation by a Wiener process in Theorem 4 in the case \( 2 < p < 4 \). For the sake of brevity, in the proofs of this section implied constants are allowed to depend on \( f, \nu \) and \( p \).

First, assume \( 1 \leq p \leq 2 \). We start by estimating the fluctuations. Recall that \( \log m \) denotes the \( m \)-fold iterated logarithm.

**Lemma 4.** For any integers \( m \geq 1, M \geq 0 \) and \( N \geq N_0(f, \nu, p, m) \), and any \( \lambda > \lambda_0(f, \nu, p, m) \)

\[
\Pr \left( \max_{1 \leq n \leq N} \left| \sum_{k=M+1}^{M+n} f(S_k) \right| \geq \lambda N^{1/p} \right) \ll f, \nu, p, m \frac{(\log m \, N)^p}{\lambda^p}.
\]

**Proof.** We use induction on \( m \). Recall that the Rademacher–Menshov inequality \([14, \text{Theorem F}]\) states that if the sequence of (not necessarily independent) random variables \( \zeta_1, \zeta_2, \ldots \) satisfy \( \left| \sum_{k=M+1}^{M+N} \zeta_k \right|_p \leq KN^{1/p} \) for all integers \( M \geq 0 \) and \( N \geq 1 \) with some constant \( K > 0 \) and some \( 1 \leq p \leq 2 \), then

\[
\left| \max_{1 \leq n \leq N} \left| \sum_{k=M+1}^{M+n} \zeta_k \right| \right|_p \leq KN^{1/p} \log(4N)
\]

for all integers \( M \geq 0 \) and \( N \geq 1 \). Corollary 10 (ii) and the Rademacher–Menshov inequality thus give \( \left| \max_{1 \leq n \leq N} \left| \sum_{k=M+1}^{M+n} f(S_k) \right| \right|_p \ll N^{1/p} \log(4N) \). The \( m = 1 \) case follows from the Markov inequality.

Now assume the claim holds for some \( m \geq 1 \), and let us prove it for \( m + 1 \). Let us decompose the index set \([M + 1, M + N]\) into consecutive intervals of integers \( H_1, J_1, H_2, J_2, \ldots, H_R, J_R \), as in Section 4, such that \( |H_i|, |J_i| \geq 4 \Delta \log N \) for all \( i \), and \( R = \Theta(N/\log N) \). Similarly to the proof of Corollary 10 we have \( \Delta_{|H_i|}, \Delta_{|J_i|} \leq N^{-2} \). Let \( M + n_r = \max J_r \), and recall that for any \( 2 \leq r \leq R \)

\[
\sum_{M+1}^{M+n_r} f(S_k) = \sum_{k \in H_1 \cup J_1} f(S_k) + \sum_{i=2}^{r} \sum_{k \in J_i} f(S_k) + \sum_{i=2}^{r} \sum_{k \in H_i} f(S_k).
\]
Here the variables $\sum_{i=2}^r \sum_{k \in I_i} f(S_k), 2 \leq r \leq R$ have the same joint distribution as $\sum_{i=2}^r Y_i, 2 \leq r \leq R$; similarly, $\sum_{i=2}^r \sum_{k \in H_i} f(S_k), 2 \leq r \leq R$ have the same joint distribution as $\sum_{i=2}^r Z_i, 2 \leq r \leq R$. Let us introduce the random events

\[
A = \left\{ \max_{2 \leq r \leq R} \left| \sum_{i=1}^r Y_i \right| \geq \frac{\lambda}{4} N^{1/p} \right\},
\]

\[
B = \left\{ \max_{2 \leq r \leq R} \left| \sum_{i=2}^r Z_i \right| \geq \frac{\lambda}{4} N^{1/p} \right\},
\]

\[
C_i = \left\{ \max_{n \in I_i \cup J_i} \left| \sum_{k=\min I_i}^n f(S_k) \right| \geq \frac{\lambda}{4} N^{1/p} \right\}.
\]

The event in the claim of the lemma is a subset of $A \cup B \cup \bigcup_{i=1}^R C_i$. Applying the inductive hypothesis on the interval $H_i \cup J_i$ of length $\ll \log N$ we get $\Pr(C_i) \ll \lambda^{-p}(\log N) \log N)^p$, and hence $\Pr\left( \bigcup_{i=1}^R C_i \right) \ll \lambda^{-p}(\log_{m+1} N)^p$.

Recall from Lemma 2 (ii) that $\left\| \sum_{i=r+1}^s Y_i^* \right\| \ll (\sum_{i=r+1}^s |J_i|)^{1/p} \ll N^{1/p}$ for any $1 \leq r < s \leq R$. It follows that the median of $\sum_{i=r+1}^s Y_i^*$ is also $\ll N^{1/p}$, and hence from Lévy’s inequality (see e.g. [15, p. 51]) we get

\[
\Pr \left( \max_{2 \leq r \leq R} \left| \sum_{i=2}^r Y_i^* \right| \geq \frac{\lambda}{8} N^{1/p} \right) \leq 2 \Pr \left( \left| \sum_{i=2}^R Y_i^* \right| \geq \frac{\lambda}{16} N^{1/p} \right) \ll \frac{1}{\lambda^p} \tag{18}
\]

provided $\lambda$ is large enough. On the other hand, Lemma 3 gives $\mathbb{E}|Y_i - Y_i^*|^p \ll |J_i|^p \Delta_{|H_i|} \ll N^{-2}(\log N)^p$, and thus

\[
\Pr \left( |Y_i - Y_i^*| \geq \frac{\lambda}{8R} N^{1/p} \right) \ll \frac{(\log N)^p}{N^2} \cdot \frac{R^p}{\lambda^p N} \ll \frac{1}{\lambda^p N}.
\]

Therefore $\Pr\left( \sum_{i=2}^R |Y_i - Y_i^*| \geq (\lambda/8) N^{1/p} \right) \ll \lambda^{-p}$. This relation, together with (18) shows $\Pr(A) \ll \lambda^{-p}$. Repeating the same arguments for $Z_i$ and $Z_i^*$ we get $\Pr(B) \ll \lambda^{-p}$. Hence $\Pr\left( A \cup B \cup \bigcup_{i=1}^R C_i \right) \ll \lambda^{-p}(\log_{m+1} N)^p$, as claimed. \hfill \Box

We are now ready to prove (3). Fix $m \geq 1$ and $\varepsilon > 0$. Let us decompose the set of positive integers into consecutive intervals of integers $H_1, J_1, H_2, J_2, \ldots$, as in Section 4 (with the choice $M = 0$), such that, say, $|H_i| = |J_i| = i$ for all $i \geq 1$. Similarly to the proof of Corollary 10 we have $i \geq 16 \Delta \log i$, and so $\Delta_{|H_i|} = \Delta_{|J_i|} \leq \varepsilon^{-8}$ for all integers $i$ large enough in terms of $\nu$.

Consider (3) along the subsequence $N_R = \max J_R = \Theta(R^2)$. We have

\[
\sum_{k=1}^{N_R} f(S_k) = \sum_{k \in H_1 \cup J_1} f(S_k) + \sum_{i=2}^R \sum_{k \in J_i} f(S_k) + \sum_{i=2}^R \sum_{k \in H_i} f(S_k).
\]
Here the sequences $\sum_{i=2}^{R} \sum_{k \in J_i} f(S_k)$ and $\sum_{i=2}^{R} \sum_{k \in H_i} f(S_k)$, $R = 2, 3, \ldots$ have the same distribution as $\sum_{i=2}^{R} Y_i$ and $\sum_{i=2}^{R} Z_i$, $R = 2, 3, \ldots$, respectively. Using Lemma 2 (ii) we get

$$\sum_{i=1}^{\infty} \frac{\mathbb{E}[Y_i^*]^p}{i \varphi_{m, \varepsilon}(i)} \ll \sum_{i=1}^{\infty} \frac{1}{\varphi_{m, \varepsilon}(i)} < \infty.$$  

By a classical form of the strong law of large numbers (see e.g. [15, p. 209]) and $R^{1/p} \varphi_{m, \varepsilon}(R)^{1/p} = \Theta(\varphi_{m, \varepsilon}(NR)^{1/p})$, we have

$$\lim_{R \to \infty} \frac{\sum_{i=1}^{R} Y_i^*}{\varphi_{m, \varepsilon}(NR)^{1/p}} = 0 \quad \text{a.s.} \quad (19)$$

Lemma 3 gives $\mathbb{E}[Y_i - Y_i^*]^p \ll |J_i|^p \Delta_{|H_i|} \ll i^{p-8}$, and hence $\Pr(|Y_i - Y_i^*| \geq 1/i^2) \ll i^{3p-8} \leq i^{-2}$. By the Borel–Cantelli lemma $\sum_{i=2}^{\infty} |Y_i - Y_i^*| < \infty$ a.s., and consequently (19) remains true if we replace $Y_i^*$ by $Y_i$. Repeating the same arguments for $Z_i$ and $Z_i^*$, we obtain (3) along the subsequence $N_R = \max J_R$.

On the other hand, applying Lemma 4 with $m + 2$ on the interval $H_R \cup J_R$ of length $2R$, we get

$$\Pr \left( \max_{N \in H_R \cup J_R} \left| \sum_{k=\min H_R}^{N} f(S_k) \right| \geq R^{1/p} \varphi_{m+1, \varepsilon}(R)^{1/p} \right) \ll \left( \frac{\log_{m+2}(R)}{\varphi_{m+1, \varepsilon}(R)} \right)^p.$$  

The Borel–Cantelli lemma shows that with probability 1, for any $R \geq 1$ and any $N \in H_R \cup J_R$ the fluctuation satisfies $\left| \sum_{k=\min H_R}^{N} f(S_k) \right| \ll_{\omega} \varphi_{m+1, \varepsilon}(N)^{1/p}$ with an implied constant depending on the point $\omega$ of the probability space. Therefore (3) holds along all $N$. This finishes the proof of Theorem 3 (i) under the extra condition that $\nu$ is strictly aperiodic.

**Proof of Theorem 4.** Assume $p = 2 + \delta$ for some $0 < \delta < 2$, and $C(f, \nu) > 0$. Let us decompose the set of positive integers into consecutive intervals of integers $H_1, J_1, H_2, J_2, \ldots$, as in Section 4 (with the choice $M = 0$), such that $|H_i| = \lceil 22 \Delta \log(i + 1) \rceil$ and $|J_i| = \lceil i^{\delta/(4+2\delta)} \rceil$ for all $i \geq 1$. As before, we have $\Delta_{|H_i|} \leq i^{-11}$ and $\Delta_{|J_i|} \leq i^{-11}$ for all integers $i$ large enough in terms of $\nu$ and $\delta$.

Recall that the Erdős–Stechkin inequality [14, Theorem A] states that if the sequence of (not necessarily independent) random variables $\zeta_1, \zeta_2, \ldots$ satisfy $\left\| \sum_{k=M+1}^{M+N} \zeta_k \right\|_p \leq K \sqrt{N}$ for all integers $M \geq 0$ and $N \geq 1$ with some constant $K > 0$ and some $p > 2$, then

$$\left\| \max_{1 \leq \eta \leq N} \sum_{k=M+1}^{M+n} \zeta_k \right\|_p \leq K_n K \sqrt{N}$$

22
for all integers $M \geq 0$ and $N \geq 1$, where $K_p > 0$ is a constant depending only on $p$. Corollary 10 (ii) and the Erdős–Stečkin inequality thus give

$$\left\| \max_{1 \leq n \leq N} \sum_{k=M+1}^{M+n} f(S_k) \right\|_{2+\delta} \ll \sqrt{N}$$

for any $M \geq 0$ and $N \geq 1$. Therefore for any $R \geq 1$ we have

$$\Pr \left( \max_{N \in H_R \cup J_R} \left| \sum_{k=\min H_R}^{N} f(S_k) \right| \geq R^{1/2} \right) \ll \frac{|J_R|^{1+\delta/2}}{R^{1+\delta/2}} \ll \frac{1}{R^{1+\delta/4}}.$$  

The Borel–Cantelli lemma shows that with probability 1, for any $R \geq 1$ and any $N \in H_R \cup J_R$ the fluctuation satisfies $\left| \sum_{k=\min H_R}^{N} f(S_k) \right| \ll R^{1/2}$ with an implied constant depending on the point $\omega$ of the probability space. For any $t \geq 1$ let $R(t)$ denote the positive integer for which $[t] \in H_{R(t)} \cup J_{R(t)}$. Summing over $\min J_1 \leq k \leq \max J_{R(t)}$ instead of $1 \leq k \leq t$, we thus obtain

$$\sum_{k \leq t} f(S_k) = \sum_{i=1}^{R(t)} \sum_{k \in J_i} f(S_k) + \sum_{i=2}^{R(t)} \sum_{k \in H_i} f(S_k) + O_\omega \left( R(t)^{1/2} \right) \quad \text{a.s.} \quad (20)$$

Here $\sum_{i=1}^{R(t)} \sum_{k \in J_i} f(S_k) = \sum_{i=1}^{R(t)} Y_i$ and $\sum_{i=2}^{R(t)} \sum_{k \in H_i} f(S_k) = \sum_{i=2}^{R(t)} Z_i$ in the Skorokhod space $D[0, \infty)$. From Lemma 3 we get

$$\mathbb{E} |Y_i - Y_i^*|^2+\delta \ll |J_i|^{2+\delta} \Delta_{|H_i|} \ll i^{-11+\delta/2}.$$  

Hence

$$\Pr(|Y_i - Y_i^*| \geq 1/i^2) \ll i^{-7+5\delta/2} \leq i^{-2},$$

so by the Borel–Cantelli lemma $\sum_{i=1}^{\infty} |Y_i - Y_i^*| < \infty$ a.s. Clearly the same holds for $Z_i - Z_i^*$.

By Lemma 2 we have $\|Z_i^*\|_{2+\delta} \ll \sqrt{\log i}$, and $\sum_{i=2}^{R} \mathbb{E} |Z_i^*|^2 = \Theta (R \log R)$. It follows (see e.g. [15, p. 246]) that $\sum_{i=2}^{R} Z_i^*$ satisfies the law of the iterated logarithm; in particular, $\sum_{i=2}^{R} |Z_i^*| \ll \sqrt{R \log R \log \log R}$ a.s. Note that $R(t)^{1/2} \ll t^{(2+\delta)/(4+3\delta)}$ and $(2+\delta)/(4+3\delta) < 1/2 - \delta/20$ whenever $0 < \delta < 2$. Thus the second double sum on the right hand side of (20) is $o(t^{1/2-\delta/20})$ a.s., and consequently the processes $\sum_{k \leq t} f(S_k)$ and $\sum_{i=1}^{R(t)} Y_i^*$ are $o \left( t^{1/2-\delta/20} \right)$-equivalent.

A special case of a theorem of Strassen [18, Theorem 4.4] states the following. Given independent random variables $\xi_i$, $i = 1, 2, \ldots$ with $\mathbb{E} \xi_i = 0$ and $V_R = \sum_{i=1}^{R} \mathbb{E} |\xi_i|^2 \rightarrow \infty$, for any $t \geq V_1$ let $R'(t)$ denote the positive integer for which $V_{R'(t)} \leq t < V_{R'(t)+1}$. If $\sum_{i=1}^{\infty} \mathbb{E} |\xi_i|^p / V_i^{\theta p/2} < \infty$ for some $p > 2$ and $0 < \theta \leq 1$, then the processes $\sum_{i=1}^{R'(t)} \xi_i$ and $W(t)$ are $o \left( t^{(1+\theta)/4 \log t} \right)$-equivalent, where $W(t)$ is a standard Wiener process.
We apply Strassen’s theorem to \( \zeta_i = Y_i^* / \sqrt{C(f, \nu)} \), \( i = 1, 2, \ldots \). By Lemma 2 we have
\[
V_R = \sum_{i=1}^{R} \mathbb{E}|\zeta_i|^2 = \sum_{i=1}^{R} |J_i| + O(R) = \Theta \left( R^{1+\delta/(4+2\delta)} \right)
\]
and \( \mathbb{E}|\zeta_i|^{2+\delta} \ll |J_i|^{1+\delta/2} \ll \delta^4/4 \). Hence \( \sum_{i=1}^{\infty} \mathbb{E}|\zeta_i|^{2+\delta}/V_i^{\theta(1+\delta/2)} < \infty \) for any \( \theta > (4 + \delta)/(4 + 3\delta) \). Choosing \( \theta \) close enough to \( (4 + \delta)/(4 + 3\delta) \), we have \( (1 + \theta)/4 < 1/2 - \delta/20 \), and so the processes \( \sum_{i=1}^{R(t)} Y_i^* / \sqrt{C(f, \nu)} \) and \( W(t) \) are \( o(\ell^{1/2-\delta/20}) \)-equivalent; clearly so are \( \sum_{i=1}^{R(t)} Y_i^* \) and \( \sqrt{C(f, \nu)}W(t) \).

Finally, we show that the processes \( Y(t) = \sum_{i=1}^{R(t)} Y_i^* \) and \( \sum_{i=1}^{R(t)} Y_i^* \) are \( o(\ell^{1/2-\delta/20}) \)-equivalent. Clearly \( \max J_R = \sum_{i=1}^{R} (|H_i| + |J_i|) \), and recall that \( V_R = \sum_{i=1}^{R} |J_i| + O(R) \). Therefore for all large enough integer \( r \), on the interval \( V_r \leq t < V_{r+1} \) we have \( R'(t) = r \) and \( R(t) = R'(V_r - s) \) for some \( 0 \leq s \ll V_r^{(4+2\delta)/(4+3\delta)} \log V_r \), and hence \( \sum_{i=1}^{R(t)} Y_i^* = Y(V_r) \) and \( \sum_{i=1}^{R(t)} Y_i^* = Y(V_r - s) \). Letting \( K_r = cV_r^{(4+2\delta)/(4+3\delta)} \log V_r \) with a large enough constant \( c > 0 \), it will thus be enough to prove that
\[
\sup_{0 \leq s \leq K_r} |Y(V_r) - Y(V_r - s)| = o \left( V_r^{1/2-\delta/20} \right) \quad \text{a.s.} \tag{21}
\]
Recalling the distribution of the running maximum of a Wiener process, we have
\[
\Pr \left( \sup_{0 \leq s \leq K_r} |W(V_r) - W(V_r - s)| \geq \lambda \sqrt{K_r} \right) \ll e^{-\lambda^2/2}.
\]
Choosing, say, \( \lambda = 2\sqrt{\log V_r} \) and noting \( (2 + \delta)/(4 + 3\delta) < 1/2 - \delta/20 \), the Borel–Cantelli lemma shows that the process \( W(t) \) satisfies the property in (21); clearly so does \( \sqrt{C(f, \nu)}W(t) \). Since (21) is invariant under \( o(\ell^{1/2-\delta/20}) \)-equivalence, \( Y(t) \) also satisfies (21). This finishes the proof in the case \( C(f, \nu) > 0 \).

If \( C(f, \nu) = 0 \), the proof is much simpler. In this case Lemma 2 gives \( \mathbb{E}|Y_i^*|^2 \ll 1 \). Therefore \( \sum_{i=1}^{\infty} \mathbb{E}|Y_i^*|^2 / i^{1+2\varepsilon} \ll \infty \) for any \( \varepsilon > 0 \), and by the strong law of large numbers \( \sum_{i=1}^{R(t)} Y_i^* = o \left( R(t)^{1/2+\varepsilon} \right) \) a.s. Similarly, \( \sum_{i=2}^{R(t)} Z_i^* = o \left( R(t)^{1/2+\varepsilon} \right) \) a.s. Using these relations instead of the law of the iterated logarithm and Strassen’s theorem and noting that \( R(t)^{1/2+\varepsilon} \ll t^{1/2-\delta/20} \) for small enough \( \varepsilon > 0 \), we get \( \sum_{k \leq t} f(S_k) = o \left( t^{1/2-\delta/20} \right) \) a.s., as claimed.

5.2 Almost sure asymptotics, general case

Proof of Theorem 3. Under the extra condition that \( \nu \) is strictly aperiodic, we proved claim (i) in Section 5.1, whereas claim (ii) follows from Theorem 4. We now show that the condition of strict aperiodicity can be removed, and prove the general case of Theorem 3.
Assume that the pair \((G, \nu)\) satisfies the conditions of Theorem 3; that is, \(G\) is a compact metrizable group, and \(\nu\) is a Borel probability measure on \(G\) such that \(\nu\) is adapted, and \((\nu^k)_{\text{abs}} \neq 0\) for some \(k \geq 1\). We shall use the notation \(\mu_G\) for the normalized Haar measure on \(G\). It is not difficult to see that if \(\nu_1\) and \(\nu_2\) are Borel probability measures on \(G\), then \(\operatorname{supp}(\nu_1 * \nu_2) = (\operatorname{supp} \nu_1)(\operatorname{supp} \nu_2)\) (see e.g. [20, Lemma 2]). Therefore \(\operatorname{supp} \nu^k = (\operatorname{supp} \nu)^k\), where we use the notation \(A^k = \{a_1a_2 \cdots a_k : a_1, a_2, \ldots, a_k \in A\}\). In particular, \(\operatorname{supp} \nu^{k+1}\) contains a translate of \(\operatorname{supp} \nu^k\), so the sequence \(\mu_G(\operatorname{supp} \nu^k)\) is nondecreasing. Let \(\alpha(G, \nu) = \lim_{k \to \infty} \mu_G(\operatorname{supp} \nu^k)\). Note that \(\alpha(G, \nu) > 0\) because \((\nu^k)_{\text{abs}} \neq 0\) for some \(k \geq 1\).

The following simple observation is a special case of [8, Theorem 14]. For the sake of completeness we include a short proof.

**Lemma 5.** Let \(G\) be a compact metrizable group. If \(K \subseteq G\) is nonempty and closed, and \(K^2 \subseteq K\), then \(K\) is a subgroup.

**Proof.** Let \(a \in K\) be arbitrary. By assumption \(a^n \in K\) for all \(n \geq 1\). Using the compactness of \(K\) we have \(a^n \to b \in K\) as \(k \to \infty\) for some subsequence \(a^{n_k}\).

For any fixed \(n \geq 1\) we have \(a^{n_k-n} \to a^{-n}b \in K\) as \(k \to \infty\). After replacing \(n_k\) by another subsequence we may assume that \(a^{n_k} \to b \in K\) and \(a^{-n_k}b \to c \in K\) as \(k \to \infty\) for some \(c\).

Then \(b = a^{n_k}a^{-n_k}b \to bc\), hence \(c = 1 \in K\). It remains to prove that for any \(a \in K\) we have \(a^{-1} \in K\). But \(aK\) is also nonempty and closed, and \((aK)^2 \subseteq aK\). By the previous argument \(b \in aK\), therefore \(a^{-1} \in K\). \(\square\)

Assume now, that there exists a proper closed normal subgroup \(H < G\) such that \(\operatorname{supp} \nu \subseteq aH\) for some coset \(aH\). Since \(H\) is normal, we have \(\operatorname{supp} \nu^k \subseteq a^kH\) for all \(k \geq 1\). Thus \(\mu_G(H) = \mu_G(a^kH) \geq \mu_G(\operatorname{supp} \nu^k)\), and so \(\mu_G(H) \geq \alpha(G, \nu) > 0\).

In particular, \(aH\) has finite order \(d\) in the factor group \(G/H\). Since \(\bigcup_{i=1}^d aH\) is a closed subgroup of \(G\) containing \(\nu\), and \(\nu\) is assumed to be adapted, we have \(G = \bigcup_{i=1}^d aH\) and \(|G : H| = d\). As \(\operatorname{supp} \nu^d \subseteq H\), we can view \(\nu^d\) as a Borel probability measure on the compact metrizable group \(H\). Note that \(\mu_H(B) = d \cdot \mu_G(B)\) \((B \subseteq H\) Borel\) is the normalized Haar measure on \(H\). Clearly \((\nu^d)^{\#k}\) has an absolutely continuous component with respect to \(\mu_H\) for some \(k \geq 1\).

It is also not difficult to see that \(\nu^d\) is adapted on \(H\). Indeed, suppose \(H < H\) is a proper closed subgroup for which \(\operatorname{supp} \nu^d \subseteq K\). Consider \(C = \bigcup_{i=1}^d \operatorname{supp} \nu^iK\), and note that here \(\operatorname{supp} \nu^iK \subseteq a^iH\); in particular, \(C \neq G\). On the other hand, writing an arbitrary integer \(k \geq 1\) in the form \(k = nd + i, 1 \leq i \leq d\) we have \(\operatorname{supp} \nu^{nd + i} = (\operatorname{supp} \nu^i)(\operatorname{supp} \nu^{nd})^n \subseteq \operatorname{supp} \nu^iK\). Therefore the topological closure \(\bigcup_{k=1}^\infty \operatorname{supp} \nu^k\) is a subset of \(C \neq G\). Using Lemma 5 we get that \(\bigcup_{k=1}^\infty \operatorname{supp} \nu^k\) is a proper closed subgroup of \(G\), contradicting the adaptedness of \(\nu\). Altogether, we find that the pair \((H, \nu^d)\) satisfies the conditions of Theorem 3. Observe, moreover, that \(\alpha(H, \nu^d) = d \cdot \alpha(G, \nu)\).
Assume in addition, that the pair \((H, \nu^d)\) satisfies the claims of Theorem 3. We now prove that under all these assumptions \((G, \nu)\) also satisfies the claims of Theorem 3. Fix a Borel measurable function \(f : G \to \mathbb{R}\) such that \(\sup_{x \in G} |f(cX_1)|^p < \infty\) for some \(p \geq 1\). It will be enough to prove that for any \(1 \leq i \leq d\) we have
\[
\lim_{N \to \infty} \frac{\sum_{k=1}^N f(S_{i+kd}) - dN \int_{a'_i} f \, d\mu_G}{\varphi_{m,\varepsilon}(N)^{1/p}} = 0 \quad \text{a.s.} \quad (22)
\]
for any \(m \geq 1\) and \(\varepsilon > 0\) in the case \(1 \leq p \leq 2\), and
\[
\limsup_{N \to \infty} \frac{\sum_{k=1}^N f(S_{i+kd}) - dN \int_{a'_i} f \, d\mu_G}{\sqrt{N \log \log N}} < \infty \quad \text{a.s.} \quad (23)
\]
in the case \(p > 2\). Fix \(1 \leq i \leq d\), and let \(\mathcal{F}_i\) denote the \(\sigma\)-algebra generated by \(X_1, X_2, \ldots, X_i\). Letting \(Y_n = \prod_{j=i+(n-1)d}^{i+nd} X_j\), the variables \(Y_1, Y_2, \ldots\) are i.i.d. \(H\)-valued random variables with distribution \(\nu^d\), independent of \(\mathcal{F}_i\). Let \(b = X_1X_2\cdots X_i\), and note \(b \in a'^i H\) a.s. Let \(g : H \to \mathbb{R}, g(x) = f(bx)\), and observe \(\sup_{c \in H} \mathbb{E}(|g(cY_1)|^p | \mathcal{F}_i) < \infty\) a.s. and \(\int_H g \, d\mu_H = d \int_{a'_i} f \, d\mu_G\) a.s. We thus have
\[
\sum_{k=1}^N f(S_{i+kd}) - dN \int_{a'_i} f \, d\mu_G = \sum_{k=1}^N g(Y_1Y_2\cdots Y_k) - N \int_H g \, d\mu_H \quad \text{a.s.}
\]
By the assumption that \((H, \nu^d)\) satisfies the claims of Theorem 3, we have
\[
\Pr \left( \lim_{N \to \infty} \frac{\sum_{k=1}^N g(Y_1Y_2\cdots Y_k) - N \int_H g \, d\mu_H}{\varphi_{m,\varepsilon}(N)^{1/p}} = 0 \quad \bigg| \mathcal{F}_i \right) = 1
\]
in the case \(1 \leq p \leq 2\), and
\[
\Pr \left( \limsup_{N \to \infty} \frac{\sum_{k=1}^N g(Y_1Y_2\cdots Y_k) - N \int_H g \, d\mu_H}{\sqrt{N \log \log N}} < \infty \quad \bigg| \mathcal{F}_i \right) = 1.
\]
in the case \(p > 2\). Taking the (total) probability, (22) and (23) follow.

Finally, we prove that the pair \((G, \nu)\) satisfies the claims of Theorem 3. Let \(H_0 = G\). If \(\nu\) is not strictly aperiodic in \(H_0\), then let \(H_1 \lhd H_0\) be a proper closed normal subgroup such that \(\text{supp} \, \nu\) is contained in a coset of \(H_1\), and let \(d_1 = |H_0 : H_1|\). As seen above, the pair \((H_1, \nu^{d_1})\) satisfies the conditions of Theorem 3, hence we can iterate this procedure. We obtain a sequence \(H_0 \lhd H_1 \lhd \cdots \lhd H_j\), where \(H_i\) is a proper closed normal subgroup of \(H_{i-1}\) with index \(d_i = |H_{i-1} : H_i|\), and
supp $\nu^{*^{(d_1\cdots d_{i-1})}}$ is contained in a coset of $H_i$ for all $1 \leq i \leq j$. The procedure ends after step $j$ if $\nu^{*^{(d_1\cdots d_j)}}$ is strictly aperiodic in $H_j$. Note that $1 \geq \alpha(H_i, \nu^{*^{(d_1\cdots d_i)}}) = \alpha(G, \nu)$, therefore the procedure terminates after finitely many steps. We prove the claims by induction on $j$. If $j = 0$, that is, $\nu$ is strictly aperiodic, the claims have already been proved. To prove the inductive step from $j - 1$ to $j$, we first apply the inductive hypothesis to $(H_1, \nu^{*^{d_1}})$, then the arguments above to conclude that $(G, \nu)$ satisfies the claims of Theorem 3.

Proof of Theorem 1. The implication (iii)$\Rightarrow$(ii) is trivial, whereas (i)$\Rightarrow$(iii) is a special case of Theorem 3. Let us finally prove (ii)$\Rightarrow$(i). First, suppose that $\nu^{*k}$ is singular with respect to $\mu$ for every $k \geq 1$. Then there exists a Borel set $B \subseteq G$ such that $\mu(B) = 0$ and $\Pr(S_k \in B) = 1$ for every $k \geq 1$. Hence the indicator function $f = I_B$ does not satisfy (ii), giving a contradiction. Suppose next, that $\nu$ is not adapted; that is, there exists a proper closed subgroup $H < G$ such that $\Pr(X_1 \in H) = 1$. Then $\Pr(S_k \in H) = 1$ for all $k \geq 1$. Since every nonempty open subset of $G$ has positive Haar measure, we have $\mu(H) < 1$. Therefore $f = I_H$ does not satisfy (ii), giving a contradiction.

5.3 Central limit theorem

Proof of Theorem 5. In this proof implied constants will be universal. Claim (ii) follows from Corollary 10 (i). To see (i), fix a positive integer $N$ large enough in terms of $f$, $\nu$ and $\delta$, and let us prove (6). Let $E_N = N^{-\delta/(2+2\delta)} \log^{\delta/(1+\delta)} N$ and $K = \Delta \left( \| f \|_{2+\delta} / \sqrt{C(f, \nu)} \right)^{(2+\delta)/(1+\delta)}$. Let us decompose the set $\{1, 2, \ldots, N\}$ into consecutive intervals of integers $H_1, J_1, \ldots, H_R, J_R$, as in Section 4 (with the choice $M = 0$), such that $|H_i| = \lfloor \Delta N \log N \rfloor$ and $|J_i| = \Theta \left( (\Delta/K^2) N^{\delta/(1+\delta)} \log^{2/(1+\delta)} N \right)$ for all $1 \leq i \leq R$. As in the proof of Corollary 10, we have $\Delta_{|J_i|} \leq N^{-2}$, and clearly the same holds for $\Delta_{|H_i|}$.

Recall that

$$\sum_{k=1}^N f(S_k) = \sum_{k \in H_1} f(S_k) + \sum_{i=1}^R \sum_{k \in J_i} f(S_k) + \sum_{i=2}^R \sum_{k \in H_i} f(S_k),$$

where $\sum_{i=1}^R \sum_{k \in J_i} f(S_k) \overset{d}{=} \sum_{i=1}^R Y_i$ and $\sum_{i=2}^R \sum_{k \in H_i} f(S_k) \overset{d}{=} \sum_{i=2}^R Z_i$. From Lemma 2 and the classical Lyapunov condition (see e.g. [15, p. 154]) we get

$$\Pr \left( \frac{\sum_{i=1}^R Y_i^*}{\sqrt{\sum_{i=1}^R \mathbb{E}(Y_i^*)^2}} < x \right) = \Phi(x) + O(K E_N). \quad (24)$$
Here $\sum_{i=1}^{R} E(Y_i)^2 = C(f, \nu)N + O_f(\nu(N^{1/(1+\delta)})$, therefore the error of replacing the normalizing factor on the left hand side of (24) by $\sqrt{C(f, \nu)N}$ is $o(E_N)$. Similarly, $\sum_{i=2}^{R} Z_i^*$ also satisfies the central limit theorem with remainder term $O(KE_N)$. In particular,

$$\sup_{x \in \mathbb{R}} \left| \Pr \left( \frac{\sum_{i=2}^{R} Z_i^*}{\sqrt{\sum_{i=2}^{R} E(Z_i^*)^2}} \geq x \right) - 2(1 - \Phi(x)) \right| \ll KE_N.$$ 

Applying this with $x = \sqrt{\log N}$ and noting that $1 - \Phi(\sqrt{\log N}) = O(N^{-1/2}) = o(E_N)$, we obtain

$$\Pr \left( \frac{\sum_{i=2}^{R} Z_i^*}{\sqrt{C(f, \nu)N}} \gg KE_N \right) \ll KE_N. \quad (25)$$

From Lemma 3 we get $\|\sum_{i=1}^{R} (Y_i - Y_i^*)\|_2 \ll_{f, \nu} \sum_{i=1}^{R} |J_i| \sqrt{\Delta_i} \ll_{f, \nu} 1$, hence the Chebyshev inequality gives

$$\Pr \left( \frac{\sum_{i=1}^{R} (Y_i - Y_i^*)}{\sqrt{C(f, \nu)N}} \geq KE_N \right) \ll_{f, \nu} \frac{1}{N^E_N} = o(E_N). \quad (26)$$

We similarly deduce

$$\Pr \left( \frac{\sum_{i=2}^{R} (Z_i - Z_i^*)}{\sqrt{C(f, \nu)N}} \geq KE_N \right) \ll_{f, \nu} \frac{1}{N^E_N} = o(E_N). \quad (27)$$

Finally, note that $\sup_{c \in G} \mathbb{E} f(cX_1)^2 < \infty$ implies $\sup_{k \geq 1} \mathbb{E} f(S_k)^2 < \infty$. Therefore $\|\sum_{k \in H} f(S_k)\|_2 \ll_{f, \nu} \log N$, and the Chebyshev inequality gives

$$\Pr \left( \frac{\left| \sum_{k \in H} f(S_k) \right|}{\sqrt{C(f, \nu)N}} \geq KE_N \right) \ll_{f, \nu} \frac{\log N}{N^E_N} = o(E_N). \quad (28)$$

Combining (24)–(28) we thus have

$$\Pr \left( \frac{\sum_{k=1}^{N} f(S_k)}{\sqrt{C(f, \nu)N}} < x \right) = \Pr \left( \frac{\sum_{i=1}^{R} Y_i}{\sqrt{C(f, \nu)N}} < x + O(KE_N) \right) + O(KE_N)$$

$$= \Phi(x) + O(KE_N).$$

\[\square\]
We now prove the Remark made after Theorem 5. If \( f \in L^4(G) \), then instead of \( \|Y_i^*\|_3 \ll \|f\|_3 \sqrt{|J_i|} \), Lemma 2 gives the slightly better estimate \( \|Y_i^*\|_3 \leq \|f\|_2 \sqrt{|J_i|} + \|f\|_4 \Delta^{3/4} |J_i|^{1/4} \). Therefore if in the definition of \( K \) we replace \( \|f\|_2+\delta \) by \( \|f\|_2 \), the Lyapunov condition gives that (24) and (25) hold with error terms \( O(KE_N) + o(E_N) = O(KE_N) \). The rest of the proof remains unchanged.

**Proof of Theorem 2.** The implication (i)\( \Rightarrow \) (ii) follows from Theorem 5 and Proposition 8. The latter is needed to ensure \( C(f, \nu) > 0 \). We now prove (ii)\( \Rightarrow \) (i). The facts that \( \nu \) is adapted, and that \( (\nu^{\geq k})_{abs} \neq 0 \) for some \( k \geq 1 \) follow similarly to the proof of Theorem 1. Suppose that \( \text{supp} \ \nu \) is contained in a coset \( aH \) of some proper closed normal subgroup \( H < G \). We have seen in Section 5.2 that the index \( d = |G : H| \) is finite, and \( G = \bigcup_{i=1}^d a^iH \). Note that if \( k = nd+i \) for some \( 1 \leq i \leq d \), then \( S_k \in a^iH \) a.s. Letting \( f = I_{H} - \mu(H) \), we thus have \( \sum_{k=1}^N f(S_k) \leq 1 - 1/d \) a.s. Hence \( N^{-1/2} \sum_{k=1}^N f(S_k) \) cannot have a nondegenerate limit distribution. \( \square \)

**Acknowledgments**

The author is grateful to István Berkes for helpful discussions about random walks, and to the referee for a careful reading of the manuscript and valuable comments.

**References**

[1] M. Anoussis and D. Gatzouras: A spectral radius formula for the Fourier transform on compact groups and applications to random walks. Adv. Math. 188 (2004), no. 2, 425–443.

[2] A. Berger and S. N. Evans: A limit theorem for occupation measures of Lévy processes in compact groups. Stoch. Dyn. 13 (2013), no. 1, 1250008, 16 pp.

[3] I. Berkes and B. Borda: On the law of the iterated logarithm for random exponential sums. Trans. Amer. Math. Soc. 371 (2019), no. 5, 3259–3280.

[4] I. Berkes and M. Raseta: On the discrepancy and empirical distribution function of \( \{n_k \alpha\} \). Unif. Distrib. Theory 10 (2015), no. 1, 1–17.

[5] R. N. Bhattacharya: Speed of convergence of the \( n \)-fold convolution of a probability measure on a compact group. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 25 (1972/73), 1–10.

[6] E. Bolthausen: The Berry–Esseen theorem for functionals of discrete Markov chains. Z. Wahrsch. Verw. Gebiete 54 (1980), no. 1, 59–73.

[7] G. B. Folland: A Course in Abstract Harmonic Analysis. Second edition. CRC Press, Boca Raton, FL, 2016.
[8] B. Gelbaum, G. K. Kalisch and J. M. H. Olmsted: *On the embedding of topological semi-groups and integral domains*. Proc. Amer. Math. Soc. 2 (1951), 807–821.

[9] Y. Kawada and K. Itô: *On the probability distribution on a compact group. I*. Proc. Phys.-Math. Soc. Japan (3) 22 (1940), 977–998.

[10] B. M. Kloss: *Probability distributions on bicom pact topological groups*. Theor. Probability Appl. 4 (1959), 237–270.

[11] T. Komorowski, C. Landim and S. Olla: *Fluctuations in Markov Processes. Time Symmetry and Martingale Approximation*. Grundlehren der Mathematischen Wissenschaften, 345. Springer, Heidelberg, 2012.

[12] P. Lévy: *L’addition des variables aléatoires définies sur un cercle*. Bull. Soc. Math. France 67 (1939), 1–41.

[13] S. Meyn and R. Tweedie: *Markov Chains and Stochastic Stability*. Communications and Control Engineering Series. Springer-Verlag, London, 1993.

[14] F. Móricz: *Moment inequalities and the strong laws of large numbers*. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 35 (1976), no. 4, 299–314.

[15] V. V. Petrov: *Limit Theorems of Probability Theory. Sequences of Independent Random Variables*. Oxford Studies in Probability, 4. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1995.

[16] K. A. Ross and D. Xu: *Norm convergence of random walks on compact hypergroups*. Math. Z. 214 (1993), no. 3, 415–423.

[17] P. Schatte: *On a law of the iterated logarithm for sums mod 1 with application to Benford’s law*. Probab. Theory Related Fields 77 (1988), no. 2, 167–178.

[18] V. Strassen: *Almost sure behavior of sums of independent random variables and martingales*. 1967 Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66) Vol. II: Contributions to Probability Theory, Part 1, pp. 315–343 Univ. California Press, Berkeley, Calif.

[19] K. Stromberg: *Probabilities on a compact group*. Trans. Amer. Math. Soc. 94 (1960), 295–309.

[20] K. Urbanik: *On the limiting probability distribution on a compact topological group*. Fund. Math. 44 (1957), 253–261.

[21] C. Villani: *Topics in Optimal Transportation*. Graduate Studies in Mathematics, 58. American Mathematical Society, Providence, RI, 2003.