Lindeberg theorem for Gibbs–Markov dynamics

Manfred Denker¹, Samuel Senti² and Xuan Zhang²

¹ Department of Mathematics, The Pennsylvania State University, PA, United States of America
² Instituto de Matemática, Universidade Federal do Rio de Janeiro, Brazil

E-mail: denker@math.psu.edu, senti@im.ufrj.br and zhang@im.ufrj.br

Received 29 January 2017, revised 13 August 2017
Accepted for publication 14 September 2017
Published 16 November 2017

Recommended by Dr Dmitry Dolgopyat

Abstract

A dynamical array consists of a family of functions \( \{f_{ni} : 1 \leq i \leq k_n, n \geq 1 \} \) and a family of initial times \( \{\tau_{ni} : 1 \leq i \leq k_n, n \geq 1 \} \). For a dynamical system \((X,T)\) we identify distributional limits for sums of the form

\[
S_n = \frac{1}{s_n} \sum_{i=1}^{k_n} [f_{ni} \circ T^{\tau_{ni}} - a_{ni}] \quad n \geq 1
\]

for suitable (non-random) constants \( s_n > 0 \) and \( a_{ni} \in \mathbb{R} \). We derive a Lindeberg-type central limit theorem for dynamical arrays. Applications include new central limit theorems for functions which are not locally Lipschitz continuous and central limit theorems for statistical functions of time series obtained from Gibbs–Markov systems. Our results, which hold for more general dynamics, are stated in the context of Gibbs–Markov dynamical systems for convenience.

Keywords: central limit theorem, Gibbs–Markov dynamics, Lindeberg condition, dynamical array, transfer operator

Mathematics Subject Classification numbers: 37A50, 37A05, 37C30, 60F05, 62E20

1. Introduction

Probabilistic methods have been used for a long time in connection with number theory and some of these applications have formulations in terms of dynamical systems, but it took more than 50 years to realize the general importance for dynamics. Continued fraction is a typical example of such a common approach in number theory and dynamics.
While the ergodic theorem has a direct counterpart in Kolmogorov’s strong law of large numbers, the classical de Moivre–Laplace central limit theorem describing its fluctuation about the mean has none. Today, central limit theorems (CLTs) are widespread in the study of fluctuations of Birkhoff ergodic sums in dynamical systems. Ideas borrowed from probability theory such as stationary mixing processes [19] and Gordin’s martingale-coboundary method [15] are commonly used to prove central limit theorems. The survey paper [8] has a comprehensive list of references up to 1986. More recently, Chazottes [4] reviewed probabilistic laws for ergodic sums in dynamical systems modeled by Young towers ([24]). A central limit theorem for Markov fibred systems with the Schweiger property was proven in [3]. Examples of such systems include parabolic rational maps, Young towers and Gibbs–Markov maps. A CLT for general rational maps was proven in [11] using Gordin’s method. All these results are concerned with Birkhoff sums. In order to obtain a CLT, the observables in the Birkhoff sums are usually assumed to be Hölder continuous. This is partly due to the popular spectral gap method, which usually holds on Banach spaces endowed with Hölder norms. On the other hand it is still an open problem to determine the class of functions in $L^2$ satisfying the CLT.

From an applied viewpoint, ergodic sums provide only a limited method to draw conclusions on a dynamical system. A much wider approach is formulated in terms of design of experiments where different time series and their interplay are considered. This leads to the need of analyzing arrays containing different ergodic sums. To our knowledge this concept was first formulated for maps of the interval and some special statistical functionals in [7]. Recently, [17] used a special form of such an approach to obtain CLT for shrinking targets. In other directions, one should also note CLTs in the settings of random dynamics or sequential dynamics such as in [5, 6]. Lindeberg’s central limit theorem deals with arrays of independent random variables, i.e. families of random variables defined on row-wise different probability spaces. We formulate Lindeberg’s central limit theorem for dynamical arrays, and prove CLTs for arrays in dynamical systems, here Gibbs–Markov maps. Examples include certain countable state Markov chains and Markov maps of the unit interval given in [2] as well as parabolic rational maps in [3]. Other examples can be found in [1]. We use two classical methods: the characteristic operators approach as in [22] and Lindeberg’s method as in [21] for blocks to prove CLTs. It will be clear from the proof that our results can be extended to more general systems since only spectral properties of transfer operators and metric properties are taken from Gibbs–Markov systems. It is also clear how to extend the results to Young towers over Gibbs–Markov maps. Recent development can be found in [23]. For simplicity, we keep our discussions restricted to Gibbs–Markov systems.

Dynamical arrays have many applications. For instance, one may use an array of Hölder continuous observables to approximate Birkhoff sums of observables of lesser regularity. An example is given in corollary 4.2. In comparison, note that Gouëzel [16] proved a CLT for Birkhoff sums of observables with Hölder norm in $L^p$, where $0 < \eta < 1$. In another paper [12] we have used CLTs for arrays to study fluctuations for ergodic sums over periodic orbits. Another possible application is through coupling Birkhoff sums of different dynamical systems.

We recall some background material on Gibbs–Markov systems and spectral properties of their transfer operators in sections 2 and 3. Section 4 contains a CLT (theorem 4.1) for sequences of Birkhoff sums $\sum_{i=1}^{n} f \circ T^i$ ($n \geq 1$). Although this is a special dynamical array the central limit theorem is treated separately since the method of proof is different from the other main theorem and may have future applications to other dynamical systems. Such theorems provide central limit theorems for Birkhoff sums $\sum_{i=1}^{n} f \circ T^i$ for certain functions which are not Lipschitz (or Hölder) continuous. We provide one easy example and others are not
2. Gibbs–Markov systems

Gibbs–Markov systems were first formulated in [2]. Let \((\Omega, \mathcal{B}, \mu, T)\) denote a nonsingular transformation of a standard probability space. Consider a countable partition \(\alpha\) of \(\Omega\mod\mu\), \(\alpha = \{a_i : i \in I\}\), and denote the \(\sigma\)-algebra generated by \(\alpha\) by \(\sigma(\alpha)\). For \(x, y \in \Omega\), define
\[
s(x, y) := \min_{n \in \mathbb{N}} \{n + 1 : T^n(x) \text{ and } T^n(y) \text{ belong to different elements of } \alpha\}.
\]
For any \(r \in (0, 1)\), set \(r(x, y) := r^{s(x, y)}\) on \(\Omega\), which will become a metric. We use the same letter \(r\) to express the dependence of the metric on the choice of \(r\). It will be clear in the subsequent context when \(r\) represents a number or a metric.

**Definition 2.1.** A quintuple \((\Omega, \mathcal{B}, \mu, T, \alpha)\) is called a Gibbs–Markov map (or system) if the following four conditions hold modulo \(\mu\).

1. \(\alpha\) is a strong generator of \(\mathcal{B}\) by \(T\), i.e. \(\sigma(\{T^{-n}\alpha : n \geq 0\}) = \mathcal{B}\).
2. For every \(a \in \alpha\), \(Ta \in \sigma(\alpha)\) and the restriction \(T|_a\) is invertible and (two-sided) nonsingular.
3. \(\inf_{a \in \alpha} \mu(Ta) > 0\).
4. For each \(n \geq 1\) and \(a \in \bigvee_{i=0}^{n-1} T^{-i}\alpha\), denote the \(\mu\)-nonsingular inverse branch \(T^{-n}|_{Ta}\) by \(v_a : T^n a \to a\) and its Radon–Nikodym derivative by \(v_a'\). Then, there exist \(r \in (0, 1)\) and \(M > 0\) such that for any \(n \geq 1\), \(a \in \bigvee_{i=0}^{n-1} T^{-i}\alpha\) and \(x, y \in T^na\),
\[
\left|\frac{v_a'(x)}{v_a'(y)} - 1\right| \leq M \cdot r(x, y).
\]

**Remark 2.2.** Usually we do not specify \(\mathcal{B}\) and write only \((\Omega, \mu, T, \alpha)\). Also note that a number \(r\) and hence the metric \(r(\cdot, \cdot)\) are determined within the definition of a Gibbs–Markov map.

We will work with the following Banach spaces: given a Gibbs–Markov system \((\Omega, \mu, T, \alpha)\) and any partition \(\rho\) of \(\Omega\), the Hölder norm of a function \(f : \Omega \to \mathbb{C}\) subject to the partition \(\rho\) is defined by
\[
D_{\rho}f := \sup_{b \in \rho} \sup_{x, y \in b, x \neq y} \frac{|f(x) - f(y)|}{r(x, y)},
\]
where \(\sup\) is understood to be taken \(\mu\) almost everywhere. Denote the usual \(L^q\)-norm by \(\|\cdot\|_q\), \(1 \leq q \leq \infty\). Then
\[
\|f\|_{\infty, \rho} := \|f\|_\infty + D_{\rho}f
\]
defines a larger norm. Denote the subspace consisting of functions of finite \(\|\cdot\|_{\infty, \rho}\) norm by \(L^\infty_\rho\). It is standard to show that \(L^\infty_\rho\) is a Banach space.
Remark 2.3. Throughout this paper we will always assume that \((\Omega, \mu, T, \alpha)\) is a topologically mixing and measure-preserving Gibbs–Markov system, here topologically mixing means that for any \(a, b \in \alpha\), there is \(n_{a, b} \in \mathbb{N}\) such that for every \(n > n_{a, b}\), \(b \subset T^n a\).

3. Transfer operator and characteristic function operator

We continue setting up the theory for a Gibbs–Markov system \((\Omega, \mu, T, \alpha)\) by introducing its transfer operators. Since \(T\alpha \subset \sigma(\alpha)\), it follows that for every \(n \in \mathbb{N}\), \(T^n(\bigvee_{i=0}^{n-1} T^{-i} \alpha) = T\alpha\). Fix a partition \(\beta\) (which may be coarser than \(\alpha\)) such that \(\sigma(T\alpha) = \sigma(\beta)\).

The Perron–Frobenius transfer operator \(L_T : L^1(\mu) \to L^1(\mu)\) is defined by:

\[
L_T f := \sum_{b \in \beta} \int_{T^{-1} b} f \, d\mu,
\]

and the transfer operator for \(T^n\) is hence of the form

\[
L_{T^n} f = \sum_{b \in \beta} \int_{T^{-n} b} f \, d\mu.
\]

\(L_T\) satisfies and is uniquely characterized by:

\[
\int_{\Omega} L_T f \cdot g \, d\mu = \int_{\Omega} f \cdot T g \, d\mu, \quad \forall g \in L^\infty(\mu).
\]

It can be easily derived from the above equation that

\[
L_{T^n} = L_T^n
\]

and since \(\mu\) is \(T\)-invariant

\[
L_T 1 = 1.
\]

We will use \(L\) for \(L_T\) when \(T\) is fixed.

Given a measurable function \(f : \Omega \to \mathbb{R}\) and \(t \in \mathbb{R}\), the characteristic function operator \(L_{f, t} : L^1(\mu) \to L^1(\mu)\) is defined as:

\[
L_{f, t} g := L(e^{itf} \cdot g), \quad \forall g \in L^1(\mu).
\]

Then

\[
L_{f, 0} = L.
\]

Note that \(D_\alpha(\cdot) \leq D_\beta(\cdot) \leq D_\Omega(\cdot) \leq \max\{2\|\cdot\|_\infty, D_\alpha(\cdot)\}\), hence the functions in \(L^\infty_\beta\) have finite \(D_\alpha\) norms, and the norm \(\|\cdot\|_\infty,\beta\) is equivalent to the norm \(\|\cdot\|_{\infty, \alpha}\) and to the norm \(\|\cdot\|_{\infty, \Omega}\). From now on, we write for simplicity

\[
L := L^\infty_\beta, \quad \|\cdot\| := \|\cdot\|_{\infty, \beta}.
\]

As no confusion should appear, we use the same notation \(\|\cdot\|\) for the operator norm on \(L\). It is not hard to see that \(L\) and \(L_{f, t}\) are both bounded linear operators on \(L\). In fact, we have the following estimates.
Lemma 3.1 ([2, proposition 2.1, theorem 2.4]). There exist constants $M$ and $M_1$ such that for any $f, g \in L$ and $s, t \in \mathbb{R}$, we have

$$\|L_{f,s}^n g\| \leq (M + M_1 D_\alpha |s| D_\alpha f) \left( |t - s| D_\alpha e^{|t-s| f} - 1 \right) + |t - s| D_\alpha f, \tag{4}$$

and

$$\|L_{f,t} - L_{f,s}\| \leq (M + M_1 |s| D_\alpha f) (\|e^{(t-s)f} - 1\|_1 + |t - s| D_\alpha f).$$

In particular, let $s = 0$, we have for every $t \in \mathbb{R}$ and $f \in L$,

$$\|L_{f,t} - L\| \leq 2M |t| \|f\|.$$

The following lemma, adapted from [22, proposition 3], provides the Taylor expansion of $L_{f,t}$ around $L$.

Lemma 3.2. For any $f \in L$ and $t \in \mathbb{R}$, there exist bounded linear operators $L_{f,t}^{(n)}$ on $L$ such that

$$L_{f,t} = L + \sum_{n=1}^{\infty} \frac{t^n}{n!} L_{f,t}^{(n)}$$

converges absolutely with $\|L_{f,t}^{(n)}\| \leq \|L\| \|f\|^n$, and for every $m \in \mathbb{N}$,

$$L_{f,t} = L + t L_{f,t}^{(1)} + \cdots + \frac{t^{m-1}}{(m-1)!} L_{f,t}^{(m-1)} + L_{f,t}^{(m)} \tag{5}$$

with $\|L_{f,t}^{(m)}\| \leq \|L\| |t|^m \|f\|^m e^{|t| \|f\|}$.

**Remark 3.3.** $L_{f,t}^{(n)}$ are just derivatives of $L_{f,t}$ around $t = 0$ when $f$ is fixed.

**Proof.** Since for any $f, g \in L$,

$$\|f \cdot g\|_\infty \leq \|f\|_\infty \cdot \|g\|_\infty$$

and

$$D_\beta (f \cdot g) \leq D_\beta (f) \cdot \|g\|_\infty + \|f\|_\infty \cdot D_\beta (g),$$

we have

$$\|f \cdot g\| \leq \|f\| \cdot \|g\|. \tag{6}$$

Therefore $\|f^n\| \leq \|f\|^n$, and

$$\left\| \sum_{n=0}^{\infty} \frac{(it)^n}{n!} L(f^n g) \right\| \leq \sum_{n=0}^{\infty} \frac{1}{n!} \|L\| \|f\|^n |t|^n \|g\|$$

converges absolutely. Let $L_{f,t}^{(n)}(\cdot) := i^n L(f^n \cdot)$ and the expansion follows. \qed

One of the underlying tools throughout this paper is the spectral gap property of the transfer operator $L$, which has been proved in [2, theorem 1.6]. We explain it briefly here. Because that $L$-bounded sets are precompact in $L^1$ and because of lemma 3.1 and a theorem of Ionescu-Tulcea and Marinescu ([20]), $L$ is quasi-compact on $L$. Notice that 1 is an eigenvalue of $L$ and is a maximal eigenvalue of $L$ on $L$ by (3) since $L$ contracts $L^\infty$. Also it is known that a
(topologically) mixing Gibbs–Markov map is exact ([2, 3]) and hence is strong-mixing when it is measure-preserving. So 1 is the unique maximal eigenvalue and is simple. Hence the transfer operator $L$ of a mixing Gibbs–Markov map has the spectral gap property on $L$, namely one can decompose $L$ on $L$ as

$$L = P + N$$

so that $Pf = \int_{\Omega} f \, d\mu$, $PN = NP = 0$ and $\tau(N) < 1$, where $\tau(\cdot)$ denotes the spectral radius. $P$ is the eigenprojection of $L$ with respect to the eigenvalue 1 and the spectrum of $N$ is all the remaining spectrum of $L$. For any complex number $z$ not in the spectrum of $L$, denote by $R(z; L)$ the resolvent of $L$, $(zI - L)^{-1}$. According to [13, VII], one can calculate $P$ and $N$ by integrating the product of the resolvent and suitable analytic functions on neighborhoods of the spectrum of $L$. In fact,

$$P = \frac{1}{2\pi i} \int_{C_1} R(z; L) \, dz, \quad N = \frac{1}{2\pi i} \int_{C_2} zR(z; L) \, dz$$

where $C_1$ is a small circle around 1 of radius $\frac{1 - \tau(N)}{3}$ and $C_2$ is a circle around 0 of radius $\frac{1 + 2\tau(N)}{3}$ so that $C_1$ and $C_2$ are disjoint and that the spectrum of $L$ except for the eigenvalue 1 is totally contained within $C_2$. For every positive integer $k$,

$$N^k = \frac{1}{2\pi i} \int_{C_2} z^k R(z; L) \, dz.$$

We will call

$$\rho_1 := \frac{1 - \tau(N)}{3}, \quad \rho_2 := \frac{1 + 2\tau(N)}{3}.$$

By perturbation theory, the characteristic function operators also satisfy the above properties. **Lemma 3.4.** There exists a real number $a > 0$ such that if $|t| \cdot \|f\| < a$ then $L_{f,t}$ has the spectral gap property on $L$ with the decomposition:

$$L_{f,t} = \lambda_{f,t}P_{f,t} + N_{f,t}$$

where

1. $\lambda_{f,t}$ is the unique eigenvalue of the largest modulus of $L_{f,t}$, $\lambda_{f,t}$ is a simple eigenvalue and $|\lambda_{f,t}| \in (1 - \rho_1, 1 + \rho_1)$;
2. $P_{f,t}$ is the eigenprojection of $L_{f,t}$ with respect to $\lambda_{f,t}$, in the form

$$P_{f,t} = \frac{1}{2\pi i} \int_{C_1} R(z; L_{f,t}) \, dz;$$

3. $\tau(N_{f,t}) < \rho_2$ and $P_{f,t}N_{f,t} = N_{f,t}P_{f,t} = 0$, in fact,

$$N_{f,t} = \frac{1}{2\pi i} \int_{C_2} zR(z; L_{f,t}) \, dz;$$

4. fix an $f \in L$, then $t \mapsto \lambda_{f,t}, t \mapsto P_{f,t}$ and $t \mapsto N_{f,t}$ are analytic on $(-a/\|f\|, a/\|f\|)$.

Here $C_1, C_2$ are the same circles as in (8).

This lemma is essentially [22, proposition 4]. For our purpose we take expansions of the operators to higher orders in the next lemma.
Lemma 3.5. There exist constants $M > 0$ and $0 < a < M^{-1}$ such that if $|t| \cdot \|f\| < a$ is small enough, then

1. $P_{f,t}$ has an expansion

$$P_{f,t} = P + tP_{f}^{(1)} + \frac{t^2}{2} P_{f}^{(2)} + P_{f,t}^{(3)},$$

where $\|P_{f}^{(i)}\| \leq M\|f\|^i$, for $i = 1, 2$, and $\|P_{f,t}^{(3)}\| \leq M|t|^3\|f\|^3e^{|t|\|f\|}$.

2. Similarly, $\lambda_{f,t}$ expands as

$$\lambda_{f,t} = 1 + t\lambda_{t}^{(1)} + \frac{t^2}{2} \lambda_{t}^{(2)} + \lambda_{f,t}^{(3)},$$

where $|\lambda_{t}^{(i)}| \leq M\|f\|^i$, for $i = 1, 2$, and $|\lambda_{f,t}^{(3)}| \leq M|t|^3\|f\|^3e^{|t|\|f\|}$.

3. for all $n \in \mathbb{N}$,

$$\|N_{f}^{n}\|_1 \leq \rho_2^2 \frac{M|t|\|f\|}{1 - M|t|\|f\|}.$$  

Proof. We use notations from lemma 3.4. Notice that for any $z$ in the resolvent set of $\mathcal{L}$, if $|t| \cdot \|f\|$ is so small that $\|\mathcal{L}_{f,t} - \mathcal{L}\| \cdot \|R(z; \mathcal{L})\| < 1$ then $z$ is also in the resolvent set of $\mathcal{L}_{f,t}$ and

$$R(z; \mathcal{L}_{f,t}) = R(z; \mathcal{L}) \sum_{n=0}^{\infty} ((\mathcal{L}_{f,t} - \mathcal{L})R(z; \mathcal{L}))^n$$  

(9)

converges absolutely.

1. We use the resolvent equation (9) to calculate $P_{f,t}$ as follows. Choose $a$ small enough such that $\|\mathcal{L}_{f,t} - \mathcal{L}\| \cdot \sup_{z \in C_1} \|R(z, \mathcal{L})\| < 1$ whenever $|t|\|f\| < a$. Then, denoting by

$$R := R(z; \mathcal{L}),$$

$$P_{f,t} = \frac{1}{2\pi i} \int_{C_1} R(z; \mathcal{L}_{f,t})dz \overset{(9)}{=} \frac{1}{2\pi i} \int_{C_1} R \sum_{n=0}^{\infty} ((\mathcal{L}_{f,t} - \mathcal{L})R)^n dz$$

$$= P + \frac{1}{2\pi i} \int_{C_1} R \sum_{n=1}^{\infty} ((\mathcal{L}_{f,t} - \mathcal{L})R)^n dz$$

$$= P + \frac{1}{2\pi i} \int_{C_1} R \sum_{n=1}^{\infty} \left\{ R \mathcal{L}^{(1)} R + \frac{t^2}{2} \frac{1}{2\pi i} \int_{C_1} R \mathcal{L}^{(2)} R + 2R \left( \mathcal{L}^{(1)} R \right)^2 \right\} dz$$

$$+ \frac{1}{2\pi i} \int_{C_1} R \sum_{n=3}^{\infty} \left\{ R \mathcal{L}^{(1)} R + tR \mathcal{L}^{(2)} R + tR \mathcal{L}^{(1)} R \mathcal{L}^{(2)} R + R \left( \mathcal{L}^{(1)} R \right)^2 \right\} dz.$$

Define corresponding operators to write the last equation in the form.

4593
\[ P_{f,j} = P + tP_f^{(1)} + \frac{t^2}{2}P_f^{(2)} + P_f^{(3)}. \]

Then there exist constants \( M_1, M_2 \) and \( M_3 \) such that when \( |t||f|| < a \),

\[
\|P_f^{(1)}\| \leq M_3\|L_f^{(1)}\| \leq M_1\|f\|,
\]

\[
\|P_f^{(2)}\| \leq M_3\|L_f^{(2)}\| + \|L_f^{(1)}\|^2 \leq M_1\|f\|,\]

\[
\|P_f^{(3)}\| \leq M_3(\|L_f^{(3)}\| + \|L_f^{(1)}\|\|L_f^{(2)}\| + \|L_f^{(1)}\|^2 + \sum_{n=3}^{\infty} M_2^n\|L_f^{(1)}\|^n)
\]

\[
\leq M_1|t|^3\|f\|^3e^{|t||f||}.\]

2. Let \( B \) be the Banach space of all bounded linear operators from \( L \) to itself. Take a linear functional \( \phi \in B^* \) such that \( \|\phi\|_{B^*} = 1 \) and \( \phi(P) = 1 \). Then because

\[
L_{f,j}P_{f,j} = \lambda_{f,j}P_{f,j},
\]

we have

\[
\lambda_{f,j} = \frac{\phi(L_{f,j}P_{f,j})}{\phi(P_{f,j})}.
\]

Define \( P_f^{(1)} = tP_f^{(1)} + \frac{t^2}{2}P_f^{(2)} + P_f^{(3)} \), then \( \|P_f^{(1)}\| \leq M_1|t||f||e^{|t||f||} \). Choose a small such that \( ae^a < M_1^{-1} \) then when \( |t||f|| < a \),

\[
\frac{1}{\phi(P_{f,j})} = \frac{1}{1 + \phi(P_{f,j}^{(1)})} = \sum_{n=0}^{\infty} (-1)^n\phi(P_{f,j}^{(1)})^n
\]

since \( \phi(P_{f,j}^{(1)}) \leq M_1|t||f||e^{|t||f||} \). Hence the expansions of \( L_{f,j} \) and \( P_{f,j} \) lead to the expansion of \( \lambda_{f,j} \).

3. The same resolvent equation is used in the calculation of \( N_{f,j} \). Choose a small enough such that \( \|L_{f,j} - L\| \cdot \sup_{z \in C}(||R(z; L)|| + ||R(z; L)||^2) < 1 \) whenever \( |t||f|| < a \), then

\[
N_{f,j}^n = \frac{1}{2\pi i} \int_{C_\varepsilon} z^n R(z; L_{f,j})dz
\]

\[
\overset{(9)}{=} \frac{1}{2\pi i} \int_{C_\varepsilon} z^n R(z; L) \sum_{m=0}^{\infty} ((L_{f,j} - L)R(z; L))^m dz
\]

\[
= N + \frac{1}{2\pi i} \int_{C_\varepsilon} z^n R(z; L) \sum_{m=1}^{\infty} ((L_{f,j} - L)R(z; L))^m dz.
\]

Notice that \( N1 = 0 \), whence we have:

\[
\|N_{f,j}^n1\| \leq \frac{1}{2\pi} \| \int_{C_\varepsilon} z^n R(z; L) \sum_{m=1}^{\infty} ((L_{f,j} - L)R(z; L))^m dz \|
\]

\[
\leq \rho^n \frac{M_2|t||f||}{1 - M_2|t||f||}
\]

for some constant \( M_2 \) with \( a < M_2^{-1} \). \( \square \)
When \( f \) is fixed, since \( \lambda_f \) is analytic with respect to \( t \) around 0 (lemma 3.4), the coefficients \( \lambda_f^{(j)} \) in the expansion of \( \lambda_f \) are just the corresponding derivatives of \( \lambda_f \) with respect to \( t \) at \( t = 0 \). They can be calculated in the following manner.

**Lemma 3.6 ([22, lemmes 2, 3, 6]).** Let \( f \in L \) with \( \int_{\Omega} f \, d\mu = 0 \). Then

\[
\lambda_f^{(1)} = 0, \quad \lambda_f^{(2)} = -\lim_{n \to \infty} \frac{1}{n} \int_{\Omega} (\sum_{m=0}^{n-1} f \circ T^m)^2 \, d\mu.
\]

The limit exists, and \( \lambda_f^{(2)} \neq 0 \) if and only if \( f \) is not of the form \( \varphi \circ T - \varphi \) for any \( \varphi \in L \).

**Proof.** One only needs to notice that the transfer operator and its spectral decomposition in [22] share the same spectral properties and expand in the same way as in our settings when \( f \) is fixed, to which the proofs of [22, lemmes 2, 3, 6] refer, hence the proofs carry over to our settings.

The asymptotic variance of \( f \) will be denoted by

\[
\sigma_f^2 := \lim_{n \to \infty} \frac{1}{n} \int_{\Omega} (\sum_{m=0}^{n-1} f \circ T^m)^2 \, d\mu.
\]

**4. A CLT for arrays of Birkhoff sums**

As mentioned in the introduction, in this section we prove a central limit theorem based on the tools developed in section 3 for arrays of functions in Gibbs–Markov systems. The following theorem appeared as part of the third author’s PhD thesis [25].

**Theorem 4.1.** Consider a Gibbs–Markov system \((\Omega, \mu, T, \alpha)\), a sequence \( \{f_n\} \subset L \) with \( \int_{\Omega} f_n \, d\mu = 0 \) and not of the form \( \varphi \circ T - \varphi \) for any \( \varphi \in L \) and a sequence of positive integers \( k_n \to \infty \). If

\[
\lim_{n \to \infty} \frac{\|f_n\|^3}{\sqrt{k_n} \sigma_n^2} = 0,
\]

where \( \sigma_n^2 := \sigma_{f_n}^2 \) is the asymptotic variance of \( f_n \), then

\[
\frac{f_n + f_n \circ T + \ldots + f_n \circ T^{k_n-1}}{\sqrt{k_n} \sigma_n^2}
\]

converges in distribution to the standard normal law \( N(0,1) \).

**Proof.** Let \( S_n = f_n + f_n \circ T + \ldots + f_n \circ T^{k_n-1} \). It can be easily verified that

\[
\mathcal{L}_{f_n}^{k_n} 1 = \mathcal{L}_{f_n}^{k_n} e^{iS_n}.
\]

Let \( t_n = \frac{i}{\sqrt{k_n} \sigma_n} \), then we have

\[
\int_{\Omega} e^{iS_n} \, d\mu = \int_{\Omega} e^{iS_n} \, d\mu = \int_{\Omega} \mathcal{L}_{f_n}^{k_n} e^{iS_n} \, d\mu
\]

\[
= \int_{\Omega} \mathcal{L}_{f_n}^{k_n} 1 \, d\mu.
\]
Since by assumption as \( n \to \infty \),
\[
|t_n| \|f_n\| = |t| \frac{\|f_n\|}{\sqrt{k_n} \sigma_n} \to 0,
\]
when \( n \) is large, according to lemma 3.4,
\[
\int_{\Omega} e^{i \frac{t_n}{\sqrt{k_n} \sigma_n}} d\mu = \lambda_{f_n, t_n}^{k_n} \int_{\Omega} P_{f_n, t_n} 1 d\mu + \int_{\Omega} N_{f_n, t_n}^{k_n} 1 d\mu.
\]
Lemma 3.5 implies that
\[
\lim_{n \to \infty} \int_{\Omega} P_{f_n, t_n} 1 d\mu = \int_{\Omega} P 1 d\mu = 1
\]
and for some constant \( M \),
\[
\left| \int_{\Omega} N_{f_n, t_n}^{k_n} 1 d\mu \right| \leq \rho_2 \frac{M|t_n| \|f_n\|}{1 - M|t_n| \|f_n\|} \to 0.
\]
By lemmas 3.5 and 3.6, when \( n \) is large,
\[
\lambda_{f_n, t_n} = 1 - \frac{1}{2} \sigma_n^2 + \lambda_{f_n, t_n}^{(3)}
\]
\[
= 1 - \frac{r^2}{2k_n} + \lambda_{f_n, t_n}^{(3)},
\]
where \( |\lambda_{f_n, t_n}^{(3)}| \leq M|t_n| \|f_n\| e^{3 |t_n| \|f_n\|} \). Hence the assumption that
\[
k_n|t_n|^3 \|f_n\|^3 = \frac{3 \|f_n\|^3}{\sqrt{k_n} \sigma_n^3} \to 0
\]
implies
\[
\lambda_{f_n, t_n}^{k_n} \to e^{-\frac{r^2}{2}}.
\]
This shows
\[
\lim_{n \to \infty} \int_{\Omega} e^{i \frac{t_n}{\sqrt{k_n} \sigma_n}} d\mu = e^{-\frac{r^2}{2}}
\]
finishing the proof of the theorem. \( \square \)

We wish to point out that the proofs of the CLT presented in this section were detailed in the context of Gibbs–Markov maps but also hold in other, more general settings of mixing dynamical systems. The main technique is the spectral gap property of Ionescu-Tulcea and Marinescu [20] which allows for the decomposition in equation (7). This property holds in all generality for maps which satisfy a Doeblin–Fortet inequality as in lemma 3.1. Our lemmas 3.4, 3.5 and 3.6 are instrumental in proving the CLT of theorem 4.1 and so this CLT holds in any setting in which the above mentioned lemmas are valid.

Note that identifying the context of a Banach space of functions along with a pair of norms satisfying our assumptions is a delicate but necessary task, without which the theorem lacks
relevant examples, and we refrain from formulating our theorems in an abstract albeit empty context. Other known examples of general settings to which theorem 4.1 applies, beyond the Gibbs–Markov systems presented in section 2, include maps of the interval endowed with the bounded variation norm, as well as Young towers endowed with the Hölder norm. For illustrative purposes we work out an example.

**Corollary 4.2.** Let \((\Omega, \mu, T, \alpha)\) be a Gibbs–Markov map. Then every function

\[
f = \sum_{n=1}^{\infty} \gamma_n g_n
\]

with \(g_n \in L, \gamma_n \in \mathbb{R}, \int_{\Omega} g_n d\mu = 0, \sup_{n \in \mathbb{N}} \|g_n\|_2 < \infty, \sup_{n \in \mathbb{N}} |\gamma_n| \|g_n\| < \infty\) and \(\sum_{k=n+1}^{\infty} |\gamma_k| \leq Kn^{-1-\eta}\) (for some constants \(K, \eta > 0\)) satisfies the central limit theorem in the form

\[
\frac{1}{\sqrt{n\sigma_n}} \sum_{j=0}^{n-1} f \circ T^j \Rightarrow \mathcal{N}(0, 1)
\]

for some sequence \(\sigma_n > 0\), provided the asymptotic variances of \(\sum_{k=1}^{n} \gamma_k g_k\) are bounded away from 0 uniformly.

**Remark 4.3.** If the asymptotic variances in the previous corollary converge to zero, then the ergodic sums normalized by \(\sqrt{n}\) converge to 0 stochastically.

**Proof.** Let \(l_n \in \mathbb{N}\) satisfy

\[
\lim_{n \to \infty} n^{-1} l_n^6 = 0 \quad \text{and} \quad \lim_{n \to \infty} l_n^{-6-2\eta} = 0.
\]

Define

\[
f_n = \sum_{j=1}^{l_n} \gamma_j g_j
\]

and denote by \(\sigma_n^2\) the asymptotic variance of \(f_n\). Note that \(f_n \in L\) with \(\|f_n\| \leq \sum_{k=1}^{l_n} |\gamma_k| \|g_k\| \leq Cl_n\) for some constant \(C > 0\) and hence (since \(\inf \sigma_n > 0\))

\[
\frac{\|f_n\|^3}{\sqrt{n\sigma_n^3}} = O \left( \frac{l_n^3}{\sqrt{n}} \right) \to 0.
\]

Take \(k_n = n\) in theorem 4.1 to deduce that

\[
\frac{f_n + f_n \circ T + \ldots + f_n \circ T^{n-1}}{\sqrt{n\sigma_n}}
\]

converges to the standard normal distribution.

Now by Chebychev’s inequality with \(M = \sup_{n \in \mathbb{N}} \|g_n\|_2\), for any \(\epsilon > 0\)
\[ \mu \left( \left\{ x \in \Omega : \left| \sum_{j=0}^{n-1} \sum_{k=\ell_n+1}^{\infty} \gamma_k g_k(T^j(x)) \right| \geq \epsilon \sqrt{n \sigma_n} \right\} \right) \]

\[ \leq \frac{1}{\epsilon n \sigma_n} \int_\Omega \left( \sum_{j=0}^{n-1} \sum_{k=\ell_n+1}^{\infty} \gamma_k g_k \circ T^j \right)^2 d\mu \]

\[ \leq \frac{1}{\epsilon n \sigma_n} n^2 M^2 \left( \sum_{k=\ell_n+1}^{\infty} |\gamma_k| \right)^2 \]

\[ \leq O(n^{(1-3-\eta)^2}) , \]

which converges to 0. It follows that \( \frac{1}{\sqrt{n \sigma_n}} \sum_{j=0}^{n-1} f \circ T^j \) and \( \frac{1}{\sqrt{n \sigma_n}} \sum_{j=0}^{n-1} f \circ T^j \) have the same limiting distribution, whence the corollary.

**Example 4.4.** Let \((\Omega, B, \mu, T, \alpha)\) denote the continued fraction transformation with \( \Omega = (0, 1) \) and \( \mu \) the Gauss measure. For every irrational \( x \in (0, 1) \), denote by \((x_n)_{n \in \mathbb{N}}\) its continued fraction expansion. Let \( a_n := \{x : x_0 = n\} \) for every \( n \in \mathbb{N} \), the partition \( \alpha = \{a_n : n \in \mathbb{N}\} \). Let \( \eta \in (0, \frac{1}{2}) \), define

\[ m_n := \lceil - \log_2 n^2 \rceil, \quad \ell_n := r^{-m_n}, \]

\[ \gamma_n := \ell^{(2+\eta)m_n}, \quad g_n := \ell_n \mathbf{1}_{[a_n]} \circ T^m - \ell_n \mu(a_{\lfloor a_n \rfloor}). \]

Here we denote by \( \lfloor . \rfloor \) and \( \lceil . \rceil \) the usual floor and ceiling functions for real numbers. Recall that \( r \in (0, 1) \) is the constant in (1). In the current case, we can take \( r = 2/3 \) (see [2, example 2], noting \( |T|^{2} | \geq 9/4 \)). It is easy to see that \( g_n \in L \) and \( \int_\Omega g_n d\mu = 0 \).

\[ \|g_n\|_\infty = \ell_n (1 - \mu(a_{\lfloor a_n \rfloor})) , \quad D_{\Omega} g_n = \ell_n r^{-m_n} = r^{-2m_n} = n^4 , \]

\[ \|g_n\| g_n \| = O(1) r^{2m_n} = O(1) n^{-2\eta} , \]

\[ \|g_n\|_2^2 = \ell_n \sqrt{\mu(a_{\lfloor a_n \rfloor}) (1 - \mu(a_{\lfloor a_n \rfloor}))} = O(1) \ell_n / \sqrt{(\ell_n (\ell_n + 1))} , \]

\[ \sum_{k=\ell_n}^{\infty} |\gamma_k| \leq \sum_{d=m_n}^{\lfloor r-d/2 \rfloor} \sum_{\lfloor r-(d+1)/2 \rfloor}^{\infty} n^{2+\eta} \]

\[ \leq \sum_{d=m_n}^{\lfloor r-d/2 \rfloor} n^{2+\eta} (r^{-d+1/2} - r^{-d/2}) \leq O(n^{3-2\eta}) . \]

It follows that \( f = \sum_{j=1}^{\infty} \gamma_j g_j \) satisfies the assumptions in the corollary, hence the central limit theorem holds:

\[ \frac{1}{\sqrt{n \sigma_n}} \sum_{j=0}^{n-1} f \circ T^j \Rightarrow \mathcal{N}(0, 1). \]
We remark that \( f \in L^2(\mu) \) but \( f \notin L \). In fact, for large \( n \),
\[
D_n f \geq D_{\sum_{i=0}^{n-1} T^{-i} \circ f} \geq r^{-m_\alpha} \sum_{d \geq m_\alpha} \sum_{1 \leq k \leq \left\lfloor 1 - \frac{d}{d+1} \right\rfloor} r^{(2+\eta)d} - d = r^{-m_\alpha} O(r^{\frac{1}{2} + \eta} m_\alpha) = O(n^{1-2\eta}),
\]
hence \( D_n f \) is infinite. This calculation also indicates that for any \( n \), \( D_n f \) is infinite.

5. A CLT for dynamical arrays after Lindeberg

We prove a CLT for dynamical arrays, and later apply it to Birkhoff sums. The notion of dynamical array is also considered in [12].

**Definition 5.1.** A dynamical array is a sequence \( \{ (F_{n,i}, \tau_n) : i = 1, \ldots, k_n \}_{i \in \mathbb{N}} \) consisting of a family of real valued functions \( F_{n,i} \) defined on a dynamical system \((\Omega, T)\) and a family of initial times \( \tau_{n,i} \in \mathbb{N} \), where \( F_{n,i} \) is of form
\[
F_{n,i} = \sum_{j=1}^{l_{n,i}} f_{a,j} \circ T^{-j-1}, \quad i = 1, \ldots, k_n
\]
with the \( f_{n,i,j} : \Omega \to \mathbb{R} \) and \( l_{n,i} \in \mathbb{N} \) and where \( \tau_{n,i} \) satisfies \( \tau_{n,i-1} + l_{n,i-1} \leq \tau_{n,i} \) for all \( i = 2, \ldots, k_n \).

Such a dynamical array brings about a sequence of sums
\[
F_{n,1} \circ T^{\tau_{n,1}} + F_{n,2} \circ T^{\tau_{n,2}} + \cdots + F_{n,k_n} \circ T^{\tau_{n,k_n}} \quad n \in \mathbb{N}.
\]

Define
\[
m_n := \inf_{2 \leq r \leq k_n} \tau_{n,i} - \tau_{n,i-1} - l_{n,i-1}
\]
as the minimal spacing (of the \( n \)th row) of a dynamical array.

We recall some notations. Let \((\Omega, \mu, T, \alpha)\) be a mixing Gibbs–Markov system. \( \beta \) is a partition of \( \Omega \) satisfying (2), \( \sigma(\beta) = \sigma(T\alpha) \). A real number \( r \) inducing a metric on \( \Omega \) is given in definition 2.1. The transfer operator \( \mathcal{L} \) has the decomposition (7) on \( L = L_\beta^2 \), i.e.
\[
\mathcal{L} f = \int_{\Omega} f d\mu + N f
\]
for all \( f \in L \). Let \( \rho := \tau(\mathbb{N}) \in (0, 1) \).

**Remark 5.2.** It is known ([18, corollaire 1]) that the essential spectrum of \( \mathcal{L} \) is at most \( r \) due to lemma 3.1, but the relation between \( \rho \) and \( r \) is unclear.

**Theorem 5.3.** Let \( \{ (F_{n,i}, \tau_n) : i = 1, \ldots, k_n \}_{i \in \mathbb{N}} \) be a dynamical array defined on a mixing Gibbs–Markov system \((\Omega, \mu, T, \alpha)\) with \( F_{n,i} = \sum_{j=1}^{l_{n,i}} f_{n,i,j} \circ T^{-j-1} \) and minimal spacing \( m_n \).
Suppose every \( f_{n,i,j} \in L \) is centered, i.e. \( \int_{\Omega} f_{n,i,j} d\mu = 0 \). Let
\[
\bar{s}_{n,0}^2 := \text{Var}(F_{n,1} \circ T^{\tau_{n,1}} + F_{n,2} \circ T^{\tau_{n,2}} + \cdots + F_{n,k_n} \circ T^{\tau_{n,k_n}}).
\]
Assume the following properties for this array.

1. For every \( n \in \mathbb{N} \),
   \[ \hat{s}_n > 0. \]

2. \( m_n > 0 \) and 
   \[ \limsup_{n \to \infty} k_n^2 \rho_n m_n < \infty. \] \( (11) \)

3. 
   \[ \limsup_{n \to \infty} \rho_n^{m_n} \sum_{1 \leq i \leq k_n} \rho_i \| F_{n,i} \| \frac{\hat{s}_n}{s_n} < \infty. \] \( (12) \)

4. The Lindeberg condition holds, i.e. for every \( \epsilon > 0 \)
   \[ \limsup_{n \to \infty} \rho_n^{m_n} \sum_{1 \leq i \leq k_n} \rho_i \| F_{n,i} \| \frac{\hat{s}_n}{s_n} = 0 \]

Then, this array satisfies a CLT, i.e.

\[ F_{n,1} \circ T^{\tau_{n,1}} + F_{n,2} \circ T^{\tau_{n,2}} + \cdots + F_{n,k_n} \circ T^{\tau_{n,k_n}} \Rightarrow \mathcal{N}(0, 1). \]

Remark 5.4. If \( k_n \to \infty \), condition (11) implies that \( m_n \to \infty \).

Remark 5.5. It will become clear in the proof that we can replace condition 3. by condition 3’.

\[ \lim_{n \to \infty} \rho_n^{m_n} \sum_{1 \leq i \leq k_n} \rho_i \frac{\| F_{n,i} \|}{s_n} \sup_{1 \leq i \leq k_n} \frac{\| F_{n,i} \|}{s_n} = 0 \]

which will be handy to check in section 6.

Lemma 5.6. There exists a constant \( C \) independent of \( n \) such that with

\[ \rho_n := \rho_n^{m_n} \quad r_n := \rho_n, \]

for every \( f \in L, \ n \in \mathbb{N} \) and \( 1 \leq i < j \leq k_n \), if \( \tau_{n,i+1} - \tau_{n,j} > l_{n,j} \) then

\[ \| N^{\tau_{n,i} - \tau_{n,j}} f \| \leq C \rho_n \| f \|_1 + C \rho_n r_n^{\ell_i - 1} r_{l_{i+1}} \cdots r_{l_{i+1}} D_{\beta} f; \] \( (14) \)

and if \( \tau_{n,j} - \tau_{n,j} > l_{n,j} - 1 \) then

\[ \| N^{\tau_{n,j}} f \| \leq C \rho_n \| f \|_1 + C \rho_n r_{\tau_{n,i} + l_{n,j} - 1} D_{\beta} f. \] \( (15) \)

Proof. Because \( L = P + N \) with \( PN = NP = 0 \),

\[ N^{\tau_{n,j}} = N^{\tau_{n,i+1} - \tau_{n,j} - l_{n,j}} L^{\tau_{n,i} - \tau_{n,i+1} + l_{n,j}} f. \]
Therefore, by (4) (with \( t = 0 \)), we have
\[
\|N^{\tau_{j}-\tau_{n}}f\| \leq \|N^{\tau_{j+1}-\tau_{n}}\| \|L^{\tau_{n}-\tau_{n+1}+h_{j}}f\|
\leq O(1)\rho^{h_{j+1}}(\|f\|_{1} + \rho^{\tau_{n}-\tau_{n+1}+h_{j}}D_{j}f)
\leq O(1)\rho^{h_{j}}(\|f\|_{1} + \rho^{(j-1)h_{0}+\cdots+h_{j-1}}D_{j}f).
\]

Similarly, for \( 2 \leq j \leq k_{n} \), (15) follows from
\[
N^{\tau_{j}}f = N^{\tau_{n}-(\tau_{n}-\tau_{j-1})}L^{\tau_{n}-\tau_{j-1}+h_{j-1}}f.
\]

Lemma 5.7. Under the same assumptions as in theorem 5.3.

1. the array is asymptotically negligible,
\[
\lim_{n \to \infty} \sup_{1 \leq i \leq k_{n}} \int_{\Omega} \frac{F_{n,i}^{2}}{s_{n}^{2}} \, d\mu = 0;
\tag{16}
\]

2. an asymptotic variance formula holds,
\[
\lim_{n \to \infty} \sum_{i=1}^{k_{n}} \int_{\Omega} \frac{F_{n,i}^{2}}{s_{n}^{2}} \, d\mu = 1.
\tag{17}
\]

Proof.

1. This is implied by the Lindeberg condition (13), for
\[
\int_{\Omega} \frac{F_{n,i}^{2}}{s_{n}^{2}} \, d\mu \leq \int_{\Omega} \frac{F_{n,i}^{2}}{s_{n}^{2}} \cdot 1_{\left(\frac{|F_{n,i}|}{s_{n}} < \epsilon\right)} \, d\mu + \sum_{i=1}^{k_{n}} \int_{\Omega} \frac{F_{n,i}^{2}}{s_{n}^{2}} \cdot 1_{\left(\frac{|F_{n,i}|}{s_{n}} \geq \epsilon\right)} \, d\mu
\leq c^{2} + I_{n,\epsilon}.
\tag{18}
\]

2. Use the transfer operator \( L \) to expand the total variance
\[
\tilde{s}_{n}^{2} = \int_{\Omega} (F_{n,1} \circ T_{\tau_{1}} + \cdots + F_{n,k_{n}} \circ T_{\tau_{k_{n}}})^{2} \, d\mu
\]
\[
= \sum_{i=1}^{k_{n}} \int_{\Omega} F_{n,i}^{2} \, d\mu + 2 \sum_{1 \leq i < j \leq k_{n}} \int_{\Omega} F_{n,i} \cdot F_{n,j} \circ T_{\tau_{i}+\cdots+\tau_{j}} \, d\mu
\]
\[
= \sum_{i=1}^{k_{n}} \int_{\Omega} F_{n,i}^{2} \, d\mu + 2 \sum_{1 \leq i < j \leq k_{n}} \int_{\Omega} L^{\tau_{i}+\cdots+\tau_{j}} F_{n,i} \cdot F_{n,j} \, d\mu
\]
\[
(10) \quad \sum_{i=1}^{k_{n}} \int_{\Omega} F_{n,i}^{2} \, d\mu + 2 \sum_{1 \leq i < j \leq k_{n}} \int_{\Omega} N^{\tau_{i}+\cdots+\tau_{j}} F_{n,i} \cdot F_{n,j} \, d\mu.
\]

The last equality holds because \( \int_{\Omega} F_{n,i} \, d\mu = 0 \). Estimate
\[ \frac{1}{\sqrt{n}} \sum_{1 \leq i < j \leq k_n} \int_{\Omega} N^{\tau_{n,i} - \tau_{n,j}} F_{n,i} \cdot F_{n,j} \, d\mu \]

\[ \leq \frac{1}{\sqrt{n}} C \sum_{1 \leq i < j \leq k_n} \left( \rho_n \| F_{n,i} \|_1 + \rho_n r_n^{i-1} \rho_n r_n^{j-1} \rho_n r_n^{i-j} \rho_n r_n^{j-1} \right) \cdot \| F_{n,j} \|_1 \]

\[ \leq C \sup_{1 \leq i \leq k_n} \left( \rho_n k_n^2 \frac{\| F_{n,i} \|_2^2}{s_n^2} + \rho_n \frac{1}{1 - r_n} \sum_{1 \leq n \leq k_n} r_n^{i-1} \frac{\| F_{n,i} \|_1 \| F_{n,j} \|_2}{s_n} \right) \]

\[ \leq C \rho_n k_n^2 \left( \epsilon^2 + L_{n,k} \right) + \frac{\rho_n}{1 - r_n} \sum_{1 \leq n \leq k_n} r_n^{i-1} \frac{\| F_{n,i} \|_2}{s_n} \left( \epsilon^2 + L_{n,k} \right)^{1/2}. \]

Now the assumptions (11)–(13) imply that the lim sup of the upper bound is bounded by $\zeta K$ for some $K > 0$, hence (17) follows. \hfill $\square$

**Proof of theorem 5.3.** Extending our probability space if necessary, we may assume that there exists an array of random variables $\{X_n\}_{i=1}^{k_n}$ such that $X_n, i = 1, \ldots, k_n$, are independent normal random variables and

\[ \mathbb{E}X_{n,i} = 0 \quad \text{and} \quad \text{Var}X_{n,i} = \text{Var}F_{n,i}. \quad (19) \]

Without loss of generality we may as well assume that for each $n$, $\{X_{n,i}\}_{i=1}^{k_n}$ and $\{F_{n,i} \circ T_{\tau_{n,i}}\}_{i=1}^{k_n}$ are independent. Define two random variables

\[ F_n = \frac{F_{n,1} \circ T_{\tau_{n,1}} + F_{n,2} \circ T_{\tau_{n,2}} + \cdots + F_{n,k_n} \circ T_{\tau_{n,k_n}}}{\tilde{s}_n}, \]

\[ X_n = \frac{X_{n,1} + \cdots + X_{n,k_n}}{\tilde{s}_n}. \]

$X_n$ is a normal random variable for being a sum of independent normal random variables and converges weakly to $\mathcal{N}(0, 1)$ because of (17). Since $F_n$ has variance 1, the set of distributions of $F_n$ is mass-preserving. To show that $F_n$ also converges weakly to $\mathcal{N}(0, 1)$, it suffices to prove that for any $h$ in the separating class $C^\infty_c(\mathbb{R})$,

\[ \mathbb{E}h(F_n) - \mathbb{E}h(X_n) \to 0. \]

Letting for $2 \leq i \leq k_n - 1$

\[ U_{n,i} := \frac{F_{n,1} \circ T_{\tau_{n,1}} + \cdots + F_{n,i-1} \circ T_{\tau_{n,i-1}} + X_{n,i+1} + \cdots + X_{n,k_n}}{\tilde{s}_n}, \]

and

\[ U_{n,1} = \frac{X_{n,2} + \cdots + X_{n,k_n}}{\tilde{s}_n}, \quad U_{n,k_n} = \frac{F_{n,1} \circ T_{\tau_{n,1}} + \cdots + F_{n,k_n-1} \circ T_{\tau_{n,k_n-1}}}{\tilde{s}_n}, \]
we can write, noting that $F_n = U_{n,k_n} + \frac{1}{s_n} F_{n,k_n} \circ T_{\tau_{k_n}}$ and $X_n = U_{n,1} + \frac{1}{s_n} X_{n,1}$,

$$h(F_n) - h(X_n) = \sum_{i=1}^{k_n} h \left( U_{n,i} + \frac{F_{n,i} \circ T_{\tau_{n,i}}}{s_n} \right) - h \left( U_{n,i} + \frac{X_{n,i}}{s_n} \right).$$

Use Taylor expansion to deduce that

$$h(F_n) - h(X_n) = \sum_{i=1}^{k_n} h'(U_{n,i}) \left( \frac{F_{n,i} \circ T_{\tau_{n,i}}}{s_n} - \frac{X_{n,i}}{s_n} \right)$$

$$+ h'' \left( U_{n,i} + \hat{\theta}_{n,i} \frac{F_{n,i} \circ T_{\tau_{n,i}}}{s_n} \right) \left( \frac{F_{n,i} \circ T_{\tau_{n,i}}}{2s_n^2} - \frac{X_{n,i}}{2s_n^2} \right) - h'' \left( U_{n,i} + \bar{\theta}_{n,i} \frac{X_{n,i}}{s_n} \right) \frac{X_{n,i}^2}{2s_n^2},$$

where $\theta_{n,i}, \hat{\theta}_{n,i} : \Omega \to [0, 1]$. Rewrite the right-hand side as

$$\sum_{i=1}^{k_n} h'(U_{n,i}) \left( \frac{F_{n,i} \circ T_{\tau_{n,i}}}{s_n} - \frac{X_{n,i}}{s_n} \right) + h'' \left( U_{n,i} \right) \left( \frac{F_{n,i} \circ T_{\tau_{n,i}}}{2s_n^2} - \frac{X_{n,i}}{2s_n^2} \right)$$

$$+ \left\{ h'' \left( U_{n,i} + \hat{\theta}_{n,i} \frac{F_{n,i} \circ T_{\tau_{n,i}}}{s_n} \right) \frac{F_{n,i} \circ T_{\tau_{n,i}}}{2s_n^2} - h'' \left( U_{n,i} \right) \frac{F_{n,i} \circ T_{\tau_{n,i}}}{2s_n^2} \right\}$$

$$- \left\{ h'' \left( U_{n,i} + \bar{\theta}_{n,i} \frac{X_{n,i}}{s_n} \right) \frac{X_{n,i}^2}{2s_n^2} - h'' \left( U_{n,i} \right) \frac{X_{n,i}^2}{2s_n^2} \right\}. \quad (20)$$

We are about to show that the expectation of (20) vanishes asymptotically. Denote by

$$\mathbb{E}_{n,i}(\cdot) := \mathbb{E}(\cdot | F_{n,1} \circ T_{\tau_1}, \ldots, F_{n,i} \circ T_{\tau_{i}})$$

the corresponding conditional expectation. To estimate the expectation of the first summand in (20), we write

$$\mathbb{E} \left( \sum_{i=1}^{k_n} h'(U_{n,i}) \left( \frac{F_{n,i} \circ T_{\tau_{n,i}}}{s_n} - \frac{X_{n,i}}{s_n} \right) \right)$$

$$= \sum_{i=1}^{k_n} \mathbb{E} \left( \mathbb{E}_{n,i}(h'(U_{n,i})) \cdot \frac{F_{n,i} \circ T_{\tau_{n,i}}}{s_n} \right) - \mathbb{E} h'(U_{n,i}) \mathbb{E} \frac{X_{n,i}}{s_n}$$

$$= \frac{1}{s_n} \sum_{i=2}^{k_n} \int_{\Omega} \mathbb{T}_{\tau_{n,i}} \mathbb{E}_{n,i}(h'(U_{n,i})) \cdot F_{n,i} d\mu$$

$$= \frac{1}{s_n} \sum_{i=2}^{k_n} \mathbb{E} h'(U_{n,i}) \int_{\Omega} F_{n,i} d\mu + \int_{\Omega} \mathbb{N}_{\tau_{n,i}} \mathbb{E}_{n,i}(h'(U_{n,i})) \cdot F_{n,i} d\mu$$

$$= \frac{1}{s_n} \sum_{i=2}^{k_n} \int_{\Omega} \mathbb{N}_{\tau_{n,i}} \mathbb{E}_{n,i}(h'(U_{n,i})) \cdot F_{n,i} d\mu, \quad (21)$$

where in the first equality we use the independence between $U_{n,i}$ and $X_{n,i}$ and

$$\mathbb{E}_{n,i}(h'(U_{n,i}) \cdot F_{n,i} \circ T_{\tau_{n,i}}) = \mathbb{E}_{n,i}(h'(U_{n,i})) \cdot F_{n,i} \circ T_{\tau_{n,i}}.$$
in the second equality we also use that \( U_{n,1} \) is independent with \( F_{n,1} \circ T^{\tau_n} \) and the last equality is due to \( \int_{\Omega} F_{n} \, d\mu = 0 \). Observe the following inequalities.

1. By (15),
\[
\frac{1}{\delta_n} \sum_{i=2}^{k_n} \int_{\Omega} N^{\tau_n} E_{n,j} h'(U_{n,j}) \cdot F_{n,j} \, d\mu \leq C \sum_{i=2}^{k_n} (\rho_n \| E_{n,i} h'(U_{n,i}) \|_1 + \rho_n r^{\tau_{n-i+1+\epsilon}} D_\beta(E_{n,i} h'(U_{n,i}))) \cdot \| F_{n,i} \|_{\delta_n}.
\]

2. \( \| E_{n,i} h'(U_{n,i}) \|_1 \leq \mathbb{E}(E_{n,i} | h'(U_{n,i})) \leq \| h' \|_\infty \).

3. Recall that \( U_{n,j} = \frac{1}{k_n} \left( \sum_{i=1}^{j-1} F_{n,j} \circ T^{\tau_n} + \sum_{j=i+1}^{k_n} X_{n,j} \right) \). Because \( \{ X_{n,j} \}_{j=1}^{k_n} \) and \( \{ F_{n,j} \circ T^{\tau_n} \}_{j=1}^{k_n} \) are independent,
\[
D_\beta(E_{n,i} h'(U_{n,i})) \leq \| h' \|_\infty D_\beta(F_{n,1} \circ T^{\tau_1} + \cdots + F_{n,i-1} \circ T^{\tau_{i-1}}).
\]

4. For any \( f \in L \) and \( m \in \mathbb{N} \),
\[
D_\beta(f \circ T^m) = \sup_{b \in \beta, x,y \in b} \frac{|f \circ T^m(x) - f \circ T^m(y)|}{r(x,y)}
\leq \sup_{b \in \beta, x,y \in b} \frac{|f \circ T^m(x) - f \circ T^m(y)| r(T^m x, T^m y)}{r(x,y)}
\leq D_\beta(f) \cdot r^{-m}
\leq \max \left\{ \frac{2 \| f \|_\infty}{r}, D_\beta(f) \right\} \cdot r^{-m} = O(1) \| f \| r^{-m}.
\]

We use these inequalities to estimate (21),
\[
\frac{1}{\delta_n} \sum_{i=2}^{k_n} \int_{\Omega} N^{\tau_n} E_{n,i} h'(U_{n,i}) \cdot F_{n,i} \, d\mu
\leq O(1) \sum_{i=2}^{k_n} \left( \rho_n \| h' \|_\infty + \rho_n r^{\tau_{n-i+1+\epsilon}} \| h' \|_\infty \sum_{j=1}^{i-1} \frac{1}{\delta_n} \rho_n \| F_{n,j} \| \right) \cdot \| F_{n,i} \|_{\delta_n}
\leq O(1) \left( k_n \rho_n + \rho_n \sum_{i=2}^{k_n} \sum_{j=1}^{i-1} r^{-j-1} \rho_n^{\epsilon} \| F_{n,j} \|_{\delta_n} \right) \cdot \sup_{1 \leq i \leq k_n} \| F_{n,i} \|_{\delta_n}
\leq O(1) \left( k_n \rho_n + \rho_n \frac{1}{r_n} \sum_{1 \leq i \leq k_n} \rho_n^{\epsilon} \| F_{n,i} \|_{\delta_n} \right) \cdot \sup_{1 \leq i \leq k_n} \| F_{n,i} \|_{\delta_n}.
\]

The bound tends to 0 as \( n \to \infty \) because of (16) and assumptions (11) and (12).
The expectation of the second summand in (20) is estimated in a similar way. We rewrite
\[ \mathbb{E} \sum_{i=1}^{k_2} h''(U_{n,i}) \left( F_{n,j}^2 \circ T^{\tau_{n,j}} - X_{n,i}^2 \right) \]
\[ = \sum_{i=2}^{k_2} \int_{\Omega} \mathbb{E}_n h''(U_{n,i}) \cdot F_{n,j}^2 \circ T^{\tau_{n,j}} \, d\mu - \mathbb{E} h''(U_{n,i}) \text{Var} X_{n,i} \]
\[ \overset{(10)}{=} \sum_{i=2}^{k_2} \mathbb{E} h''(U_{n,i}) \text{Var} F_{n,j}^2 + \int_{\Omega} N^{\tau_{n,j}} \mathbb{E}_n h''(U_{n,i}) \cdot F_{n,j}^2 \, d\mu - \mathbb{E} h''(U_{n,i}) \text{Var} X_{n,i} \]
\[ \overset{(19)}{=} \sum_{i=2}^{k_2} \int_{\Omega} N^{\tau_{n,j}} \mathbb{E}_n h''(U_{n,i}) \cdot F_{n,j}^2 \, d\mu. \]

Then we can repeat the estimate for (21) to deduce an upper-bound similar to (23).

The expectation of the third summand in (20) is equal to
\[ \sum_{i=1}^{k_2} \int_{\Omega} \left( h'' \left( U_{n,i} + \theta_{n,i} F_{n,j} \circ T^{\tau_{n,j}} / \delta_n \right) - h''(U_{n,i}) \right) \frac{F_{n,j}^2}{2\delta_n^2} \, d\mu \]
\[ = \sum_{i=1}^{k_2} \int_{\Omega} \left( h'' \left( U_{n,i} + \theta_{n,i} F_{n,j} \circ T^{\tau_{n,j}} / \delta_n \right) - h''(U_{n,i}) \right) \frac{F_{n,j}^2}{2\delta_n^2} \, d\mu \]
\[ \leq \epsilon \|h''\|_{\infty} \sum_{i=1}^{k_2} \int_{\Omega} \frac{F_{n,j}^2}{\delta_n} \, d\mu + \|h''\|_{\infty} L_{n,\epsilon} \]

for any \( \epsilon > 0 \). This expectation converges to 0 in view of (17) and (13). The expectation of the last summand in (20) is controlled in the same way as the third summand. \( \square \)

Applying this theorem to Birkhoff sums, we obtain the following result.

**Corollary 5.8.** Given a Gibbs–Markov system \((\Omega, \mu, T, \alpha)\) and a sequence of centered functions \(\{f_n\}\) in \(L\). Let \(s_n^2 := \text{Var}(f_n + \cdots + f_n \circ T^{n-1})\). Assume that there are sequences of integers \(l_n > m_n > 0\) with the following properties.

1. \(\limsup_{n \to \infty} k_n^2 \rho^{m_n} < \infty\),

where \(k_n := \left[ \frac{n}{l_n + m_n} \right]\).

2. \(\limsup_{n \to \infty} k_n \rho^{m_n} \frac{\|f_n\|}{s_n} < \infty\). \quad (24)

3. For every \(1 \leq i \leq l_n\)

\[ \frac{1}{s_n} \int_{\Omega} (f_n + \cdots + f_n \circ T^{i-1})^2 \, d\mu \to 0. \] \quad (25)
4. 

\[ \frac{k_n}{s_n^2} \int_{\Omega} (f_n + \cdots + f_n \circ T^{m_n-1})^2 \, d\mu \to 0. \]  

(26)

5. 

\[ \frac{k_n}{s_n^2} (f_n + \cdots + f_n \circ T^{k_n-1})^2 \text{ is uniformly integrable.} \]  

(27)

Then \( \frac{L + \cdots + f_n \circ T^{m_n-1}}{s_n} \Rightarrow N(0, 1). \)

**Proof.** Let \( F_{n,i} := f_n + \cdots + f_n \circ T^{i-1} \) \( g_{n,i} := f_n + \cdots + f_n \circ T^{m_n-1} \) and \( \tau_{n,i} := (i - 1) \) \((l_n + m_n)\) for \( 1 \leq i \leq k_n \), then

\[ f_n + \cdots + f_n \circ T^{k_n(l_n + m_n) - l_n} = \sum_{i=1}^{k_n} F_{n,i} \circ T^{\tau_{n,i}} + g_{n,i} \circ T^{\tau_{n,i} + l_n}. \]

To complete the ergodic sum, let \( F_{n,k_n+1} := f_n + \cdots + f_n \circ T^{\min\{l_n, n-k_n(l_n + m_n)\}-1} \) \( \tau_{n,k_n+1} := k_n(l_n + m_n) \) and \( g_{n,k_n+1} := f_n + \cdots + f_n \circ T^{n-k_n(l_n + m_n)-l_n-1} \) if necessary. The following two properties ensure that the dynamical array \( \{(F_{n,i}, \tau_{n,i}) : i = 1, \ldots, k_n + 1\} \) has the same distributional limit as the ergodic sum.

1. 

\[ \frac{\hat{s}_n}{s_n} \to 1, \]  

(28)

where \( \hat{s}_n^2 = \text{Var} \sum_{i=1}^{k_n+1} F_{n,i} \circ T^{\tau_{n,i}}. \)

2. 

\[ \frac{\sum_{i=1}^{k_n+1} g_{n,i} \circ T^{\tau_{n,i} + l_n}}{s_n} \to 0. \]  

(29)

In fact, to see (28) first note that

\[ s_n^2 = \hat{s}_n^2 + \text{Var} \sum_{i=1}^{k_n+1} g_{n,i} \circ T^{\tau_{n,i} + l_n} + 2 \int_{\Omega} \sum_{i=1}^{k_n+1} F_{n,i} \circ T^{\tau_{n,i} + l_n} \cdot \sum_{i=1}^{k_n+1} g_{n,i} \circ T^{\tau_{n,i} + l_n} \, d\mu. \]

With conditions (24) and (26), arguments involving the transfer operator similar to those used in proving (17) indicate that

\[ \lim_{n \to \infty} \frac{\text{Var} \sum_{i=1}^{k_n+1} g_{n,i} \circ T^{\tau_{n,i} + l_n}}{s_n^2} = \lim_{n \to \infty} \frac{\int_{\Omega} \sum_{i=1}^{k_n+1} f_{n,i} \circ T^{\tau_{n,i} + l_n} \, d\mu}{s_n^2}. \]
which is 0 by (25) and (26). As we can separate
\[
\int_{\Omega} F_{n,i} \circ T^{\tau_n} \cdot \sum_{j=1}^{k_n+1} g_{n,j} \circ T^{\tau_n + l_n - \tau_n} \, d\mu = \sum_{j=1}^{k_n+1} \int_{\Omega} F_{n,i} \cdot g_{n,j} \circ T^{\tau_n + l_n - \tau_n} \, d\mu
\]
\[+ \sum_{j \leq i-2} \int_{\Omega} g_{n,j} \cdot F_{n,i} \circ T^{\tau_n - \tau_n - l_n} \, d\mu + \int_{\Omega} F_{n,i} \cdot g_{n,j} \circ T^{l_n} \, d\mu + \int_{\Omega} g_{n,i-1} \cdot F_{n,i} \circ T^{m_i} \, d\mu,
\]
\[
\int_{\Omega} F_{n,i} \cdot g_{n,i} \circ T^{l_n} \, d\mu = \left( f_{n} + \cdots + f_{n} \circ T^{l_n - m_n - 1} \right) \cdot g_{n,i} \circ T^{l_n} \, d\mu + \int_{\Omega} g_{n,i-1} \cdot F_{n,i} \circ T^{m_i} \, d\mu
\]
and similarly for \( \int_{\Omega} g_{n,i-1} \cdot F_{n,i} \circ T^{m_i} \, d\mu \), the techniques of transfer operator can be used again to show that
\[
\frac{1}{s_n^2} \sum_{i=1}^{k_n+1} \int_{\Omega} F_{n,i} \circ T^{\tau_n} \cdot \sum_{j=1}^{k_n+1} g_{n,j} \circ T^{\tau_n + l_n} \, d\mu \to 0.
\]
Thus (28) holds. The previous arguments also imply that (29) is just a consequence of (26). Hence we only need to verify the conditions in theorem 5.3 for the dynamical array \( \{(F_{n,i}, \tau_n) : i = 1, \ldots, k_n + 1\} \), (12) is taken care of by (24) since
\[
\limsup_{n \to \infty} \frac{\rho_m}{s_n} \sum_{1 \leq i \leq k_n} \rho_i \|F_{n,i}\| \leq O(1) \limsup_{n \to \infty} \frac{\rho_m}{s_n} k_n \frac{\|F_{n}\|}{s_n}.
\]
Note that
\[
\sum_{i=1}^{k_n} \int_{\Omega} F_{n,i}^2 \mathbf{1}_{\{|F_{n,i}| \geq \varepsilon_n\}} \, d\mu = k_n \int_{\Omega} F_{n,i}^2 \mathbf{1}_{\{|F_{n,i}| \geq \varepsilon_n\}} \, d\mu,
\]
the Lindeberg condition (13) follows from (27), (25) and (28).

**Remark 5.9.** Theorem 5.3 also can be generalized with the same assumptions to more general dynamical systems. It is in fact holds for any system for which the transfer operator satisfies the Doeblin-Fortet inequality (4) and for which the composition operator satisfies the inequality (22) (or in the case of Lipschitz norm, \( r(T^{m_i}, T^{m_j}) \leq \frac{r(x,y)}{r^{1-m}} \)). Note that the inequalities (4) and (22) are bounded at the rates of \( r \) and \( r^{-1} \) respectively.

6. Applications to the large sample theory in statistics

The CLT under the Lindeberg condition has many applications, in particular in nonparametric statistics. The book [14] provides a glimpse on these applications, though it is not a complete list. Here we restrict to one particular case, the famous Behrens–Fisher problem. In what follows, consider the setup for the two sample problem in a Gibbs–Markov dynamical system \((\Omega, \mu, T, \alpha)\).
Definition 6.1. Denote by \( \tilde{L} \) the set of all measurable functions \( f : \Omega \to \mathbb{R} \) for which there exists a sequence of functions \( \{f_n : n \geq 1\} \) in \( L \) such that \( \|f - f_n\| \to 0 \).

Consider two functions \( \phi, \psi : \Omega \to \mathbb{R} \) in the class \( \tilde{L} \), which determine two stationary sequences \( X_n = \phi \circ T^n \) and \( Y_k = \psi \circ T^k \). For simplicity we assume that the distributions \( \mu_\phi \) of \( \phi \) and \( \mu_\psi \) of \( \psi \) have no atoms. Based on observations \( X_1, \ldots, X_m \) and \( Y_1, \ldots, Y_n \), the Behrens–Fisher problem is to determine whether the distributions of \( \phi \) and \( \psi \) are different or not in a statistical sense. We shall deal with this problem when the distributions differ in their means, are different or not in a statistical sense. We shall deal with this problem when the distributions differ in their means, normalized by \( \sqrt{n} \).

We first give conditions under which the second moment of \( A \) is the two sample Wilcoxon rank sum test. In order to solve the problem in a nonparametric setup one needs to determine the asymptotic distribution of \( W_{m,n} \).

\[
W_{m,n} = \sum_{i=1}^{m} R_i
\]

is the two sample Wilcoxon rank sum test. In order to solve the problem in a nonparametric setup one needs to determine the asymptotic distribution of \( W_{m,n} \).

\[
W_{m,n} = \sum_{i=1}^{m} \sum_{k=1}^{n} 1_{\{Y_k \leq X_i\}} + \sum_{i=1}^{m} \sum_{k=1}^{n} 1_{\{X_i \leq Y_k\}}
\]

\[
= \sum_{i=1}^{m} \sum_{k=1}^{n} 1_{\{Y_k \leq X_i\}} + \frac{m(m-1)}{2}
\]

\[
= \left( \sum_{i=1}^{m} \sum_{k=1}^{n} \left( 1_{\{Y_k \leq X_i\}} - \int 1_{\{Y_k \leq t\}} \mu_\phi(dt) - \int 1_{\{t \leq X_i\}} \mu_\psi(dt) + \int \int 1_{\{s \leq t\}} \mu_\phi(dt) \mu_\psi(ds) \right) \right)
\]

\[
+ m \left( \sum_{i=1}^{m} \int 1_{\{Y_k \leq t\}} \mu_\phi(dt) - n \int \int 1_{\{t \leq X_i\}} \mu_\phi(dt) \mu_\psi(ds) \right)
\]

\[
+ n \left( \sum_{i=1}^{m} \int 1_{\{t \leq X_i\}} \mu_\psi(ds) - m \int \int 1_{\{s \leq t\}} \mu_\phi(dt) \mu_\psi(ds) \right)
\]

\[
+ \left( mn \int \int 1_{\{s \leq t\}} \mu_\psi(ds) \mu_\phi(ds) + \frac{m(m-1)}{2} \right)
\]

\[
= A + mB_n + nC_m + D.
\]

We first give conditions under which the second moment of \( A \), normalized by \( m^3 \), converges to zero as \( m \to \infty \) and \( n/m \to \lambda \in (0, 1) \). This can be seen directly or by applying [9] when \( (x,y) \mapsto 1_{\{\psi(y) \leq \phi(x)\}} \) approximately belongs to the projective tensor product \( L_2(\pi^2) \) over \( L_2(\mu^2) \) and therefore the variance of the approximation \( A \) to \( A \) increases like \( \sqrt{mn} \|A\|_{L_2(\mu^2)} \).

We refer to [9] for the definitions and properties of projective tensor products. Alternatively, assuming that the distributions of \( \psi \) and \( \phi \) are absolutely continuous with respect to Lebesgue measure and have a bounded density, one could use [10, theorem 1 or lemma 3] to show that the variance of \( A \) is of smaller order. As an example, we prove...
Proposition 6.2. Assume that the distributions $\mu_\phi$ and $\mu_\psi$ satisfy
\[
\mu_\phi(I) \leq K\eta^r \quad \text{and} \quad \mu_\psi(I) \leq K\eta^r \quad \forall \eta > 0, \forall \text{ interval } I \text{ of length } \eta
\]
for some $K > 0$ and $r \in (\frac{r}{4}, 1]$ and assume that $\|\phi\|_\infty$ and $\|\psi\|_\infty$ are finite. Then, as $n/m \to \lambda \in (0, 1)$, $A$ has a representation $A = A_1 + A_2$ so that
\[
\operatorname{Var}_1 = o(m^3) \quad \text{and} \quad \mathbb{E}[A_2] = o(m^{3/2}).
\]
Therefore, normalized by $m^{3/2}$, $A$ does not contribute to the distributional limit of $W_{m,n}$.

Proof. Let $m \in \mathbb{N}$ and choose $q$ which depends on $m$ and is chosen below. Let $M = \max\{\|\phi\|_\infty, \|\psi\|_\infty\}$ and $h(x, y) = 1_{\{-M \leq x \leq M\}}$. Divide the interval $[-M, M]$ into $q$ subintervals $J_1, ..., J_q$ of equal length $2Mq^{-1}$ and let
\[
I_j = \{ (x, y) : x \in J_j, -M \leq y \leq \min J_j \}, \quad I = \bigcup_{j=1}^q I_j.
\]
Then $\mu_\phi \times \mu_\psi(\{(x, y) : -M \leq x \leq M \} \setminus I) \leq K(2M)^r q^{-r}$ and the projective norm of $(u, v) \mapsto \hat{h}_q(u, v) = 1_I(\phi(u), \psi(v))$ is bounded by (see [9, lemma 1])
\[
\|\hat{h}_q\|_{L_2(\mu^2)} \leq \sum_{j=1}^q \|1_{I_j}(\phi, \psi)\|_{L_2(\mu^2)} \leq q \left[K(2M)^r q^{-r}\right]^{1/4}.
\]
Write
\[
\hat{h}_q(u, v) = \hat{h}_q(u, v) - \int \hat{h}_q(w, v)\mu(dw) - \int \hat{h}_q(u, w)\mu(dw) + \int \int \hat{h}_q(w, w')\mu(dw)\mu(dw')
\]
and
\[
A_1 = \sum_{i=1}^n \sum_{j=1}^m \hat{h}_q(T^i(u), T^j(v)).
\]
Then $\|\hat{h}_q\|_{L_2(\mu^2)} = O(\|\hat{h}_q\|_{L_2(\mu^2)}^2)$ and applying lemma 4 in [9] with $d = m = 2$ and $p = 4$ (one can verify the assumption in this lemma for $h_q$) it follows that there is a constant $C$ (independent of $q$ and $(n, m)$) such that
\[
\|A_1\|_{L_2(\mu)} \leq C \sqrt{nm} \|\hat{h}_q\|_{L_2(\mu^2)} = O(mq^{1-r/4}).
\]
Moreover, we get
\[
\int \int \left| \sum_{k=1}^n \sum_{i=1}^m h(\phi(T^i(u), \psi(T^j(v))) - \hat{h}_q(T^i(u), T^j(v))) \right| \mu(du)\mu(dv) \leq K(2M)^r q^{-r} nm.
\]
Similar estimates hold for the other summand $A_2 = A - A_1$.  

4609
Since \(1 \geq r > \frac{1}{2}\), we have that \(0 < 2 - \frac{r}{2} < 2r\) hence can pick 
\[
\frac{1}{2r} < \tau < \frac{1}{2 - \frac{1}{2}}
\]
and \(q = m^\tau\) to obtain 
\[
m^{-3/2} E(A_1^2) = O(m^{-3} m^2 q^{2 - r/2}) = O(m^{-1 + 2r - \frac{r}{2}}) = o(1)
\]
and 
\[
m^{-3/2} |A - A_1| = O(m^{-3/2} m^2 q^{-r}) = O(m^{1 - r\tau}) = o(1).
\]
\(\square\)

Since \(\phi, \psi \in \tilde{L}\) they are approximated in \(L^2\) by functions \(\phi_m, \psi_n \in L\). Set \(F_\phi\) and \(F_\psi\) for the respective distribution functions of \(\phi\) and \(\psi\). Denote 
\[
\tilde{B}_n := \sum_{k=1}^{n} (1 - F_\phi(\psi_n \circ T^k)) - n \int (1 - F_\phi(s)) \mu_\psi(ds)
\]
\[
\tilde{C}_m := \sum_{k=1}^{m} F_\psi(\phi_m \circ T^k) - m \int F_\psi(s) \mu_{\phi_n}(ds)
\]
and 
\[
\sigma_m^2 := m^2 \text{Var}(\tilde{B}_n) + n(m)^2 \text{Var}(\tilde{C}_m) + 2n(m)m \text{Cov}(\tilde{B}_n, \tilde{C}_m).
\]

We are not developing more details and extensions of the forgoing discussions, instead we assume that 
\[
\text{Var}(A) = o(\sigma_m^2) \tag{30}
\]
\[
||\phi - \phi_m||_1 = o(n^{-1} m^{-1} \sigma_m) \quad ||\psi - \psi_n||_1 = o(n^{-1} m^{-1} \sigma_m) \tag{31}
\]
\[
D_\alpha \phi_m = o(m^2) \quad D_\alpha \psi_n = o(n^2) \tag{32}
\]
and that \(F_\phi\) and \(F_\psi\) are Lipschitz continuous. Under these simplifying assumptions the following argument becomes short and shows the pattern of the proof under more general assumptions.

**Proposition 6.3.** Under the assumptions (30) and (31) and (32) and if \(n = n(m)\) so that for some \(\lambda \in (0, 1)\), \(\lambda \leq n/m \leq \lambda^{-1}\) and that 
\[
\liminf \sigma_m m^{-3/2} > 0,
\]
then as \(m \to \infty\) 
\[
\frac{1}{\sigma_m^2} \left( W_{m,n(m)} - D \right) \Rightarrow \mathcal{N}(0,1).
\]

Note that in case the distributions of \(\phi\) and \(\psi\) are equal, then 
\[
D = \frac{mn}{2} + \frac{m(m-1)}{2} = \frac{m}{2}(n+m-1).
\]
This shows that the two sample Wilcoxon rank sum test checks whether the distributions of \( \phi \) and \( \psi \) differ by a location alternative.

**Proof.** We first show that \( \frac{1}{\sigma_n}(W_{m,n(m)} - D) \) and \( \frac{1}{\sigma_n}(m\tilde{B}_{n(m)} + n(m)\tilde{C}_m) \) have the same limiting distribution. This follows from the assumption (30) and (using (31))

\[
m\|B_{n(m)} - B_{n(m)}\|_1 \leq 2n(m)\|\psi - \psi_{n(m)}\|_1 = o(\sigma_n)
\]

where \( D_{F_\phi} \) denotes the corresponding Lipschitz constant and from a similar inequality for \( n\|C_m - \tilde{C}_m\|_1 \).

Hence the assertion of the proposition follows if

\[
V_m = \frac{1}{\sigma_m}(m\tilde{B}_{n(m)} + n(m)\tilde{C}_m)
\]

converges weakly to the standard normal distribution.

We apply theorem 5.3. Let w.l.o.g. \( n \leq m \), \( \theta_{\phi_n} = \int \mathbb{1}_{\{\psi \leq \phi_n\}} d\mu_{\psi} d\mu_{\phi_n} \), \( \theta_{\psi_n} = \int \mathbb{1}_{\{\phi \leq \psi_n\}} d\mu_{\psi} d\mu_{\phi_n} \), \( n = k_n(p + q) + q_n \) and \( m = k_m(p + q) + q_m \) where \( 0 \leq q_n, q_m < p + q \). Denote

\[
F_{n,i} = \sum_{l=0}^{p-1} \left[ n(F_{\psi}(\phi_n \circ T^l) - \theta_{\phi_n}) - m(F_{\phi}(\psi_n \circ T^l) - \theta_{\psi_n}) \right] \quad 1 \leq i \leq k_n,
\]

\[
F_{n,i} = \sum_{l=0}^{p-1} \left[ n(F_{\psi}(\phi_n \circ T^l) - \theta_{\phi_n}) \right] \quad k_n + 1 \leq i \leq k_m,
\]

\[
\tau_{n,i} = (i-1)(p+q) \quad i = 1, ..., k_m, \quad s_n^2 = \text{Var}\left( \sum_{i=1}^{k_n} F_{n,i} \circ T^{\tau_{n,i}} \right).
\]

We check next that conditions 1.–4. in theorem 5.3 hold with an appropriate choice of \( p \) and \( q \). Let \( r \) and \( \rho \) be the constants related to the decomposition (10) which are given by the system. Choosing \( p = O(m^{1/2}) \) for some \( 0 < \delta < \frac{1}{2} \) and \( q = \frac{2 \log m}{\log \rho} \) it follows that

\[
k_n^2 \rho^p = O(m^{1+2\delta}),
\]

hence 2. holds. Since \( ||F_{n,i}|| \leq e^{-r} \left( nD_\phi(F_{\psi} \circ \phi_n) + mD_\phi(F_{\phi} \circ \psi_n) \right) + O(mp) \) for \( i = 1, ..., k_n \) and similarly for \( i = k_n + 1, ..., k_m \) we calculate:

\[
\rho^q k_n^q r^p ||F_{n,i}|| ||F_{n,i}|| \sigma_m^{-2} = O(m^{-1/2}(m(D_{\phi_m} + D_{\psi_n}) + r^q mp)mq^2)
\]

and

\[
\int F_{n,i}^2 1_{(F_{n,i} \leq \epsilon_n)} d\mu = O(m^{-1/2}) \int F_{n,i}^2 d\mu = O(m^{-1/2}) \int F_{n,i}^2 d\mu
\]

and hence

\[
\sigma_m^{-2} \sum_{i=1}^{k_n} F_{n,i}^2 1_{(F_{n,i} \leq \epsilon_n)} d\mu = O(m^{-2\delta}) \int F_{n,i}^2 d\mu
\]
It is straightforward to show using the calculus developed in this article (see (17) and (28) and observing (33) and that \( |F_\psi \circ \phi_\theta| \) and \( |F_\phi \circ \psi_\theta| \) are bounded) that 
\[
\hat{s}_n / \sigma_m = 1 + O(\sqrt{D\phi_m + D\psi_n / m}) \overset{(32)}{\to} 1
\]
and
\[
\hat{s}_n^{-2} \sum_{i=1}^{k_m} \int F_{n,i}^2 d\mu \to 1.
\]
Therefore conditions 1., 3’, and 4. hold. It is proved now that
\[
\frac{1}{\sigma_m} \sum_{i=1}^{k_m} F_{n,i}
\]
converges weakly to the standard normal distribution. We finally remark that \( \frac{1}{\sigma_m} \sum_{i=1}^{k_m} F_{n,i} \) is stochastically equivalent to \( V_m \), since the variance of the difference is bounded by
\[
\sigma_m^{-2} k_m q^2 m^2 = O(m^{-\frac{1}{2} + \delta (\log m)^2}) = o(1)
\]
were one uses the same estimates as for the comparison of \( \hat{s}_n \) and \( \sigma_m \). This finishes the proof.

\[\square\]

Acknowledgment

The research was supported by número 158/2012 de Pesquisador Visitante Especial de CAPES. XZ was also supported by PNPD/CAPES.

References

[1] Aaronson J and Denker M 1999 The Poincaré series of \( \mathbf{C} \setminus \mathbf{Z} \) Ergod. Theor. Dynam. Syst. 19 1–20
[2] Aaronson J and Denker M 2001 Local limit theorems for partial sums of stationary sequences generated by Gibbs–Markov maps Stoch. Dyn. 1 193–237
[3] Aaronson J, Denker M and Urbaniński M 1993 Ergodic theory for Markov fibred systems and parabolic rational maps Trans. Am. Math. Soc. 337 495–548
[4] Chazottes J-R 2015 Fluctuations of observables in dynamical systems: from limit theorems to concentration inequalities Nonlinear dynamics new directions (Nonlinear System Complex vol 11) (Cham: Springer) pp 47–85
[5] Cohen G and Conze J-P 2017 CLT for random walks of commuting endomorphisms on compact abelian groups J. Theor. Probab. 30 143–95
[6] Conze J-P and Raugi A 2007 Limit theorems for sequential expanding dynamical systems on \([0, 1]\) Ergodic Theory and Related Fields (Contemporary Mathematics vol 430) (Providence, RI: American Mathematical Society) pp 89–121
[7] Denker M 1982 Statistical decision procedures and ergodic theory Ergodic Theory and Related Topics (Vite, 1981) (Mathematical Research vol 12) (Berlin: Akademie) pp 35–47
[8] Denker M 1989 The central limit theorem for dynamical systems Dynamical Systems and Ergodic Theory (Warsaw, 1986) (Banach Center Publications vol 23) (Warsaw: PWN) pp 33–62
[9] Denker M and Gordin M 2014 Limit theorems for von Mises statistics of a measure preserving transformation Probab. Theor. Relat. Fields 160 1–45
[10] Denker M and Keller G 1986 Rigorous statistical procedures for data from dynamical systems
J. Stat. Phys. 44 67–93
[11] Denker M, Przytycki F and Urbański M 1996 On the transfer operator for rational functions on the
Riemann sphere Ergod. Theor. Dynam. Syst. 16 255–66
[12] Denker M, Senti S and Zhang X 2017 Fluctuations of ergodic sums over periodic orbits under
specification preprint
[13] Dunford N and Schwartz J T 1958 Linear Operators. I. General Theory (Pure and Applied
Mathematics vol 7) (New York: Interscience Publishers, Inc.) (with the assistance of W G Bade
and R G Bartle)
[14] Ferguson T S 1996 A Course in Large Sample Theory (Texts in Statistical Science Series) (London:
Chapman and Hall)
[15] Gordin M I 1969 The central limit theorem for stationary processes Dokl. Akad. Nauk SSSR
188 739–41
[16] Gouëzel S 2010 Characterization of weak convergence of Birkhoff sums for Gibbs–Markov maps
Isr. J. Math. 180 1–41
[17] Haydn N, Nicol M, Vaienti S and Zhang L 2013 Central limit theorems for the shrinking target
problem J. Stat. Phys. 153 864–87
[18] Hennion H 1993 Sur un théorème spectral et son application aux noyaux lipchitziens Proc. Am.
Math. Soc. 118 627–34
[19] Ibragimov I A and Linnik Y V 1971 Independent and Stationary Sequences of Random Variables
(Groningen: Wolters-Noordhoff) (with a supplementary chapter by I A Ibragimov and V V
Petrov, Translation from the Russian edited by J F C Kingman)
[20] Tulcea C T I and Marinescu G 1950 Théorie ergodique pour des classes d’opérations non
complètement continues Ann. Math. (2) 52 140–7
[21] Lindeberg J W 1922 Ein neuer Herleitung des Exponentialgesetzes der Wahrscheinlichkeitsrechnung
Math. Z. 15 211–25
[22] Rousseau-Egele J 1983 Un théorème de la limite locale pour une classe de transformations
dilatantes et monotones par morceaux Ann. Probab. 11 772–88
[23] Thomine D 2014 A generalized central limit theorem in infinite ergodic theory Probab. Theor.
Relat. Fields 158 597–636
[24] Young L-S 1998 Statistical properties of dynamical systems with some hyperbolicity Ann. Math.
(2) 147 585–650
[25] Zhang X 2015 Studies on the weak convergence of partial sums in Gibbs–Markov dynamical
systems PhD Thesis The Pennsylvania State University