Statistical Limit Laws for Equivariant Observations

Ian Melbourne Matthew Nicol
Department of Maths and Stats
University of Surrey
Guildford GU2 7XH, UK

2 April 2003; Revised 7 December 2003

Dedicated to the 60th birthday of Klaus Schmidt and Peter Walters

Abstract

We show that statistical limit laws for ergodic stationary sequences of G-equivariant observations ϕ on a probability space $\Omega \times G$ are inherited by sequences of observations $\phi(\,\cdot\,, g_0)$ on the probability space Ω for each fixed g_0. The statistical limit laws we consider are the central limit theorem, weak invariance principle and the law of the iterated logarithm.

1 Introduction

Let G be a compact Lie group with (normalised) Haar measure ν. We consider a fixed orthogonal representation of G on $\mathbb{R}^d$. Throughout this paper, all random variables will take values in $\mathbb{R}^d$.

Let (Ω, μ) be a probability space, $f : \Omega \to \Omega$ a measure-preserving transformation and $v : \Omega \to \mathbb{R}^d$ a measurable observation. Let $h : \Omega \to G$ be a measurable cocycle and define $h_j = h \cdot h \circ f \cdots h \circ f^{j-1}$. In this paper, we are interested in the asymptotic statistical properties of the sequence of partial sums

$$Y_N = \sum_{j=1}^{N} h_j \, v \circ f^j.$$

A related question is to consider the sequence defined on $\Omega \times G$ given by

$$Z_N(x, g) = \sum_{j=1}^{N} g h_j(x) v \circ f^j(x).$$

Observe that $Z_N(x,g) = gY_N(x)$ and $Y_N(x) = Z_N(x,e)$. A fruitful approach to studying the sequence $\{Y_N\}$ is to (i) study $\{Z_N\}$ and then (ii) deduce properties of $\{Y_N\}$ from the corresponding properties for $\{Z_N\}$. Step (i) has been the subject of recent work [4, 5, 6], and step (ii) is the subject of this paper.

The reason that $\{Z_N\}$ is more tractable than $\{Y_N\}$ is that the increments $\{gh_j\, v\circ f^j\}$ form a stationary sequence. That is, we can choose a measure preserving transformation T and a random variable ϕ such that

$$gh_j(x)\, v\circ f^j(x) = \phi\circ T^j(x,g). \tag{1.1}$$

Indeed, define

$$T(x,g) = (f(x), gh(x)), \qquad \phi(x,g) = gv(x).$$

Then the transformation $T : \Omega\times G \to \Omega\times G$ is measure-preserving with respect to the measure $\mu\times\nu$ (by translation-invariance of Haar measure ν), the map $\phi : \Omega\times G \to \mathbb{R}^d$ is measurable, and (1.1) is satisfied.

It is easily seen that if Y has a G-invariant distribution and $\{a_N\}$ is a convergent sequence of positive numbers such that $a_N Y_N \to_d Y$, then $a_N Z_N \to_d Y$. In general the converse is not true. The main result in this paper is that the converse is true when T is ergodic, $\phi\in L^p(\Omega\times G)$ and $a_N = O(N^{-1/p})$:

Theorem 1.1 *Let $T : \Omega\times G\to\Omega\times G$ be an ergodic measure-preserving transformation with respect to $\mu\times\nu$ and let $\phi : \Omega\times G\to\mathbb{R}^d$ be of the form $\phi(x,g) = gv(x)$ where $v\in L^p(\Omega,\mathbb{R}^d)$ for some $p\geq 1$. Let $\{a_N\}$ be a sequence of positive numbers satisfying $a_N = O(N^{-1/p})$. If $a_N\sum_{j=1}^N \phi\circ T^j \to_d Z$ with respect to the measure $\mu\times\nu$ on $\Omega\times G$, then $a_N\sum_{j=1}^N \phi\circ T^j(\,\cdot\,,g_0) \to_d Z$ with respect to the measure μ on Ω for each fixed $g_0\in G$.*

Remark 1.2 It suffices to prove that $a_N\sum_{j=1}^N \phi\circ T^j(\,\cdot\,,e) \to_d Z$ with respect to μ. To see this, note first by equivariance that $\phi\circ T^j(x,g_0 g) = g_0\,\phi\circ T^j(x,g)$ for all $g_0\in G$. By translation-invariance of ν, $g_0 a_N\sum_{j=1}^N \phi\circ T^j$ has the same distribution as $a_N\sum_{j=1}^N \phi\circ T^j$ with respect to $\mu\times\nu$. Hence $g_0 Z =_d Z$ for all $g_0\in G$.

Now suppose that $a_N\sum_{j=1}^N \phi\circ T^j(\,\cdot\,,e) \to_d Z$ and let $g_0\in G$ be fixed. Then $a_N\sum_{j=1}^N \phi\circ T^j(\,\cdot\,,g_0) = g_0 a_N\sum_{j=1}^N \phi\circ T^j(\,\cdot\,,e) \to_d g_0 Z =_d Z$ as required.

It turns out that the stationarity of the sequence $\{\phi\circ T^j\}$ does not play a large role in Theorem 1.1. We have the following generalisation for nonstationary sequences.

Theorem 1.3 *Let $T : \Omega\times G\to\Omega\times G$ be an ergodic measure-preserving transformation with respect to $\mu\times\nu$. Let $Z_N : \Omega\times G\to\mathbb{R}^d$ be a sequence of measurable* *functions of the form* $Z_N(x,g) = gV_N(x)$ *where* $V_N : \Omega \to \mathbb{R}^d$ *is measurable. Assume that for each fixed* $j \geq 1$

$$Z_N \circ T^j - Z_N \to 0 \ \ a.e. \tag{1.2}$$

If $Z_N \to_d Z$ *with respect to the measure* $\mu \times \nu$ *on* $\Omega \times G$*, then* $Z_N(\,\cdot\,, g_0) \to_d Z$ *with respect to the measure* μ *on* Ω *for each fixed* $g_0 \in G$.

It is easy to see that Theorem 1.1 is a special case of Theorem 1.3 with $Z_N = a_N \sum_{j=1}^N \phi \circ T^j$. We must verify that condition (1.2) is valid. Since $\phi^p \in L^1$ it follows from the pointwise ergodic theorem that $\phi^p \circ T^N = o(N)$ *a.e.* Hence $a_N \phi \circ T^N \to 0$ *a.e.* But $Z_N \circ T^j - Z_N = a_N \sum_{i=1}^j \phi \circ T^{N+i}$ consists of finitely many such terms (for fixed j) so $Z_N \circ T^j - Z_N \to 0$ *a.e.*

In [4, 5, 6], we proved equivariant statistical limit theorems for the partial sums $\sum_{j=1}^N \phi \circ T^j$ in several settings, such as assuming that the transformation $f : \Omega \to \Omega$ is uniformly hyperbolic [4, 6] or more generally nonuniformly hyperbolic [5]. These results are phrased in terms of the measure $m = \mu \times \nu$ on $\Omega \times G$. It is natural to ask whether such limit theorems remain valid when restricted to the measure μ on Ω when $g_0 \in G$ is fixed. That is, does a statistical law for the sequence $\{\phi_N\}$ defined on $\Omega \times G$ imply the corresponding statement for $\{\phi_N(\,\cdot\,, g_0)\}$ defined on Ω? It turns out that the answer is affirmative. This is easily seen for the law of the iterated logarithm (LIL), but is less easily proved for the central limit theorem (CLT) or the weak invariance principle (WIP). Theorem 1.1 is the main ingredient required to obtain the CLT and WIP on Ω.

Thus the results in this paper can be viewed as being a nontrivial refinement of the statistical limit laws obtained in [4, 5, 6].

In a recent paper, we applied these refined limit laws to random iterations of Euclidean isometries [1]. The appropriate mathematical model to study random iterations of Euclidean isometries is to sample with respect to a measure space Ω (rather than $\Omega \times G$) and hence the results of [4, 5, 6] do not immediately apply. The stronger statements we present in this paper allow the circle of results in [4, 5, 6] to be applied in this context.

In Section 2, we consider the implications of Theorems 1.1 and 1.3 for statistical properties of observations on Ω. In Section 3, we prove Theorem 1.3.

2 Statistical limit laws on $\Omega \times G$ and on Ω

In this section we consider the implications of Theorems 1.1 and 1.3 for statistical properties of the sequence $\{\phi_N\}$ where $\phi_N = \sum_{j=1}^N \phi \circ T^j$.

A well-studied example is that of hyperbolic dynamical systems. For example, suppose that Ω is a hyperbolic basic set for an Axiom A diffeomorphism, μ is a Gibbs measure corresponding to a Hölder continuous potential, $h : \Omega \to G$ is a Hölder cocycle, and $\phi(x, g) = gv(x)$ where $v : \Omega \to \mathbb{R}^d$ is Hölder. Let $m = \mu \times \nu$ and suppose that $\int \phi \, dm = 0$. Suppose further that the transformation $T : \Omega \times G \to \Omega \times G$ induced by h is ergodic with respect to m. Under these assumptions, it is possible to establish the central limit theorem (CLT), weak invariance principle (WIP) and law of the iterated logarithm (LIL) [4, 5]. It follows from the results in this paper that these properties are inherited by the partial sums sampled from Ω for each fixed g_0.

Central limit theorem The CLT on Ω is immediate for L^2 observations ϕ by taking $p = 2$ and $a_N = 1/\sqrt{N}$ in Theorem 1.1. We record this result as a corollary:

Corollary 2.1 (To Theorem 1.1) *Suppose that $T : \Omega \times G \to \Omega \times G$ is ergodic and that $\phi : \Omega \to \mathbb{R}^d$ is of the form $\phi(x, g) = gv(x)$ where $v \in L^2(\Omega, \mathbb{R}^d)$.*

If the sequence $\{\frac{1}{\sqrt{N}}\phi_N\}$ defined on the probability space $(\Omega \times G, m)$ converges in distribution to a d-dimensional normal distribution with mean zero and covariance matrix Σ, then for any $g_0 \in G$, the sequence $\{\frac{1}{\sqrt{N}}\phi_N(\,\cdot\,, g_0)\}$ defined on the probability space (Ω, μ) converges in distribution to the same d-dimensional normal distribution with mean zero and covariance matrix Σ.

Weak invariance principle Set $W_N(0) = 0$, and $W_N(t) = \frac{1}{\sqrt{N}}\phi_{[Nt]}$, for $t = \frac{1}{N}, \frac{2}{N}, \ldots$ Linearly interpolating on each interval $[(r-1)/N, r/N]$, $r \geq 1$, we obtain a sequence of random elements $W_N \in C([0, \infty), \mathbb{R}^d)$ on the probability space $\Omega \times G$. In other words, W_N should be viewed as a measurable function $W_N : \Omega \times G \to C([0, \infty), \mathbb{R}^d)$. The weak invariance principle (which is a refinement of the CLT) is the property that the sequence $\{W_N\}$ converges weakly to Brownian motion.

Corollary 2.2 (To Theorem 1.3) *Suppose that $T : \Omega \times G \to \Omega \times G$ is ergodic and that $\phi : \Omega \to \mathbb{R}^d$ is of the form $\phi(x, g) = gv(x)$ where $v \in L^2(\Omega, \mathbb{R}^d)$.*

If the sequence $\{W_N(t)\}$ defined on the probability space $(\Omega \times G, m)$ converges weakly to a d-dimensional Brownian motion $W(t)$ with covariance matrix Σ, then for all $g_0 \in G$ the sequence $\{W_N(t)(\,\cdot\,, g_0)\}$ defined on the probability space (Ω, μ) also converges weakly to $W(t)$.

Proof Weak convergence is equivalent [2] to weak convergence of the finite dimensional distributions coupled with a tightness condition. By considering G-invariant subsets of $\mathbb{R}^d$, it is immediate that tightness for m is equivalent to tightness for μ. It remains to discuss the convergence of the finite-dimensional distributions

$$\Big(W_N(t_1),\, W_N(t_2) - W_N(t_1), \ldots, W_N(t_k) - W_N(t_{k-1})\Big)$$

where $0 \le t_1 \le \cdots \le t_k$ and $k \ge 1$ is fixed. Let $c_1, \dots, c_k \in \mathbb{R}$. By the Cramer-Wold technique (cf. [3, Theorem 29.4]), it equivalent to consider whether sequences of the form

$$Z_N = c_1 W_N(t_1) + c_2\big(W_N(t_2) - W_N(t_1)\big) + \cdots + c_k\big(W_N(t_k) - W_N(t_{k-1})\big)$$

converge in distribution to the appropriate d-dimensional normal distribution. Hence it suffices to show that if $Z_N \to_d Z$ on $\Omega \times G$, then $Z_N(\,\cdot\,, g_0) \to_d Z$ on Ω.

To apply Theorem 1.3, it remains to verify the condition (1.2). By definition,

$$Z_N = \frac{1}{\sqrt{N}}\Big(c_1\phi_{[Nt_1]} + c_2\big(\phi_{[Nt_2]} - \phi_{[Nt_1]}\big) + \cdots + c_k\big(\phi_{[Nt_k]} - \phi_{[Nt_{k-1}]}\big)\Big).$$

All but finitely many of the terms in $Z_N \circ T^j - Z_N$ cancel, and the $2kj$ terms that remain are of the form $\frac{1}{\sqrt{N}} c_p \phi \circ T^q$ where $1 \le p \le k$ and $0 \le q \le Nt_k + j$. Since $\phi \in L^2$, each of these remaining terms converges to zero almost everywhere by the pointwise ergodic theorem. ∎

Law of iterated logarithm A version of the LIL is proved in [4, 5], whereby for each $c \in \mathbb{R}^d$, there exists $\sigma_c \ge 0$ such that

$$\limsup_{N\to\infty}\; c \cdot \phi_N(x,g)/\sqrt{2N \log\log N} = \sigma_c$$

for almost every $(x,g) \in \Omega \times G$. (In the situation of [4, 5], the covariance matrix $\Sigma = \lim_{N\to\infty} \frac{1}{N}\int_{X\times G} \phi_N \phi_N^T \, dm$ exists and σ_c is given by $\sigma_c = (c^T \Sigma c)^{1/2}$.) The analogous result holds when sampling from the measure space Ω. Theorem 1.1 is not required here, so this is a more elementary result.

We require a preliminary lemma.

Lemma 2.3 *Let $w_N : \Omega \to \mathbb{R}^d$ be a sequence of random variables. Suppose that for each $c \in \mathbb{R}^d$, there exists $\sigma_c \ge 0$ such that*

$$\limsup_{N\to\infty}\; c \cdot w_N(x) = \sigma_c \tag{2.1}$$

for almost every $x \in \Omega$. Then σ_c depends continuously on c, and there is a full measure subset $\widetilde{\Omega} \subset \Omega$ such that (2.1) holds for all $c \in \mathbb{R}^d$ and $x \in \widetilde{\Omega}$.

Proof Let C be a countable dense subset of $\mathbb{R}^d$. Then we can choose a single full measure subset $\widetilde{\Omega} \subset \Omega$ such that (2.1) holds for all $c \in C$ and $x \in \widetilde{\Omega}$.

Fix an $x \in \widetilde{\Omega}$. Then $\limsup_{N\to\infty} c \cdot w_N(x) = \sigma_c$ for all $c \in C$. In particular, $c \cdot w_N(x)$ is bounded and since C spans $\mathbb{R}^d$, $w_N(x)$ is bounded. Hence there exists a constant M (depending only on x) and an integer N_0 such that $|w_N(x)| \leq M$ for all $N \geq N_0$.

Next, we show that σ_c depends continuously on c. Suppose that $c_0 \in \mathbb{R}^d$ and choose $c_j \in C$ such that $c_j \to c_0$. There is a full measure set $\Omega_0 \subset \widetilde{\Omega}$ such that (2.1) holds for all $x \in \Omega_0$ and all $c \in \{c_1, c_2 \ldots\} \cup \{c_0\}$. In particular, $|w_N(x)|$ is bounded for fixed $x \in \Omega_0$. Hence $\limsup_{N\to\infty}(c_0 - c_j) \cdot w_N(x) \to 0$ as $j \to \infty$, and so $\sigma_{c_j} \to \sigma_{c_0}$.

Now we restrict attention again to $x \in \widetilde{\Omega}$. Suppose that $c \in \mathbb{R}^d$ and $\epsilon > 0$. Choose $c' \in C$ such that $|c - c'| < \epsilon/3M$ and $|\sigma_c - \sigma_{c'}| < \epsilon/3$. Increasing N_0 if necessary, we have $c' \cdot w_N(x) < \sigma_{c'} + \epsilon/3 < \sigma_c + 2\epsilon/3$ for all $N \geq N_0$ and hence

$$c \cdot w_N(x) = (c - c') \cdot w_N(x) + c' \cdot w_N(x) \leq \epsilon/3 + \sigma_c + 2\epsilon/3 = \sigma_c + \epsilon.$$

It follows that $\limsup_{N\to\infty} c \cdot w_N(x) \leq \sigma_c$.

To obtain the reverse inequality, observe that since $c' \cdot w_N(x) > \sigma_{c'} - \epsilon/3 > \sigma_c - 2\epsilon/3$ infinitely often, it follows that $c \cdot w_N(x) > \sigma_c - \epsilon$ infinitely often. ∎

Theorem 2.4 *Suppose that $T : \Omega \times G \to \Omega \times G$ is ergodic and that $\phi : \Omega \times G \to \mathbb{R}^d$ is of the form $\phi(x, g) = gv(x)$ where $v : \Omega \to \mathbb{R}^d$ is measurable. If*

$$\limsup_{N\to\infty} c \cdot \phi_N(x, g)/\sqrt{2N \log\log N} = \sigma_c \qquad (2.2)$$

holds for all $c \in \mathbb{R}^d$, for almost all $(x, g) \in \Omega \times G$, then (2.2) *holds for all $c \in \mathbb{R}^d$ and all $g_0 \in G$, for almost all $x \in \Omega$. Moreover, $\sigma_{gc} = \sigma_c$ for all $c \in \mathbb{R}^d$ and $g \in G$.*

Proof By Lemma 2.3, there is a full measure set $\widetilde{Y} \subset \Omega \times G$ such that (2.2) holds for all $(x, g) \in \widetilde{Y}$ and $c \in \mathbb{R}^d$. By Fubini's theorem, there exists $\widetilde{\Omega} \subset \Omega$ of full measure and $g_1 \in G$ such that (2.2) holds for $g = g_1$ and all $x \in \widetilde{\Omega}$. Now, compute that for each $g_0 \in G$ and all $x \in \widetilde{\Omega}$,

$$c \cdot \phi_N(x, g_0) = (g_1 g_0^{-1} c) \cdot (g_1 g_0^{-1} \phi_N(x, g_0)) = c' \cdot \phi_N(x, g_1),$$

where $c' = g_1 g_0^{-1} c$. Hence $\limsup_{N\to\infty} c \cdot \phi_N(x, g_0)/\sqrt{2N \log\log N} = \sigma_{c'}$. If $\sigma_{gc} = \sigma_c$ for all $g \in G$ and $c \in \mathbb{R}^d$ then $\sigma_{c'} = \sigma_c$. Hence we need only prove the last statement of Theorem 2.4, that $\sigma_{gc} = \sigma_c$ for all $g \in G$ and all $c \in \mathbb{R}^d$.

Let $\sigma(c, g)$ denote the lim sup, which is independent of $x \in \widetilde{\Omega}$ and almost independent of g (constant on a full measure subset of G). Furthermore $\sigma(c, g_1) = \sigma_c$ depends continuously on c. Since

$$c \cdot \phi_N(x, g) = (g_1 g^{-1} c) \cdot (g_1 g^{-1} \phi_N(x, g)) = (g_1 g^{-1} c) \cdot \phi_N(x, g_1),$$

it follows that $\sigma(c, g) = \sigma(g_1 g^{-1} c, g_1) = \sigma_{g_1 g^{-1} c}$. Hence $\sigma(c, g)$ depends continuously also on g and so $\sigma(c, g) = \sigma_c$ for all g. But now we have $\sigma_c = \sigma_{g_1 g^{-1} c}$ for all g as required. ∎

Remark 2.5 The dynamical setting under which our hypotheses hold is not limited to uniformly hyperbolic systems. The conditions on hyperbolicity are relaxed substantially in [5], so that statistical properties on $\Omega \times G$ are obtained in many situations when Ω is nonuniformly hyperbolic, including Young towers [8]. In particular, the CLT and WIP on $\Omega \times G$ are proved for certain nonuniformly hyperbolic systems in [5] and again these restrict to Ω by the results in this paper. In general, additional hypotheses are required for the LIL to be valid, though in [5] the upper LIL (where $=$ is replaced by $\leq$ in (2.1)) is obtained under certain assumptions on ϕ. Again, the LIL or upper LIL, whenever valid on $\Omega \times G$, is valid on Ω also.

3 Proof of Theorem 1.3

It simplifies our argument considerably to assume from the start that f and hence T are invertible measure-preserving transformations. This assumption is with no loss of generality as we could always form the natural extension of f [7]. By Remark 1.2 it suffices to prove the result for $g_0 = e$.

Throughout, we write $Y_N(x) = Z_N(x, e)$. Also, we write $\mu(Y_N \in E)$ instead of $\mu\{x \in \Omega : Y_N(x) \in E\}$ and we write $m(Z_N \in E)$ instead of $m\{(x,g) \in \Omega \times G : Z_N(x,g) \in E\}$.

(a) Sketch of the proof

Let $C_\rho \subset \mathbb{R}^d$ be a ball of radius ρ.

Our proof is based upon the observation that for any set $B \subset G$, for μ a.e $y \in \Omega$ the sequence $\{h_j(f^{-j}y)\}$ is 'well-distributed' in G in the sense that

$$\lim_{J\to\infty} \frac{1}{J} \sum_{j=0}^{J-1} I_{gB} \circ h_j(f^{-j}y) = \nu(B),$$

for ν a.e. $g \in G$. This is a consequence of the ergodicity of T. Choose g_0 from this set of full measure in G.

Let $\{B_i\}_{i=1}^{M}$ be a cover of G by sets of small diameter and equal Haar measure $\frac{1}{M}$. We may choose J large enough that for most points in Ω, for $1 \leq u, v \leq M$

$$\frac{1}{J} \sum_{j=0}^{J-1} I_{g_0 B_u} \circ h_j(f^{-j}y) \simeq \frac{1}{M} \simeq \frac{1}{J} \sum_{j=0}^{J-1} I_{g_0 B_v} \circ h_j(f^{-j}y),$$

where $h_j(x) = h(x)h(fx)\cdots h(f^{j-1}x)$. Translating the cover $\{B_i\}$ by g_0^{-1} and relabelling, we may assume without loss that

$$\frac{1}{J}\sum_{j=0}^{J-1} I_{B_u}\circ h_j(f^{-j}y) \simeq \frac{1}{J}\sum_{j=0}^{J-1} I_{B_v}\circ h_j(f^{-j}y)$$

for most $y \in \Omega$. Define $\Omega_j = T^j(\Omega \times \{e\})$. Then the set $\cup_{j=0}^{J-1} T^j(\Omega\times\{e\}) = \cup_{j=0}^{J-1}\Omega_j$ is 'well-distributed' in $\Omega \times G$.

Next, we show that for large N and most points in Ω, $Z_N \circ T^j(x,e) - Z_N(x,e) \simeq 0$ for $0 \le j \le J-1$. Hence, if we define

$$\sigma(Z_N \in E) = \frac{1}{J}\sum_{j=0}^{J-1}\mu\{x : Z_N \circ T^j(x,e) \in E\}$$

then $\sigma(Z_N \in E) \simeq \mu(Y_N \in E)$.

Since points in $\cup_{j=0}^{J-1}\Omega_j$ are well-distributed, and the measure σ averages over this set, we have $\sigma(Z_N \in B_i C_\rho) \simeq \sigma(Z_N \in B_j C_\rho)$. But since $\sigma(Z_N \in E) \simeq \mu(Y_N \in E)$ we also have $\mu(Y_N \in B_i C_\rho) \simeq \mu(Y_N \in B_j C_\rho)$. This shows that the equivariance of Z_N is inherited by Y_N in the limit as $N \to \infty$. The remainder of the proof is straightforward.

(b) Details of the proof

We break the proof of Theorem 1.3 into 6 steps.

Step 1: Main steps of proof.

Let C_ρ be a closed ball of radius $\rho > 0$ such that $P(Z \in \partial C_\rho) = 0$. Let $\epsilon > 0$, and choose $\delta < \rho/4$ sufficiently small, so that

$$0 \le P(Z \in C_{\rho+4\delta}) - P(Z \in C_{\rho-4\delta}) < \epsilon.$$

Since Z_N converges in distribution to Z (on $\Omega \times G$), we may choose $N(\delta)$ sufficiently large that for all $N \ge N(\delta)$

$$0 \le m(Z_N \in C_{\rho+4\delta}) - m(Z_N \in C_{\rho-4\delta}) < 2\epsilon. \tag{3.1}$$

If $B \subset G$ define $B(\alpha) = \{g \in G : d(g,B) \le \alpha\}$. We claim that for all sufficiently small $r > 0$ there exists α independent of r and a collection of compact connected sets $\{B_i\}_{i=1}^M$ in G, such that for all $1 \le i,j,k,l \le M$,

(a) $\mathrm{diam}(B_i) < r$ and $\bigcup_{i=1}^M B_i = G$.

(b) $\nu(B_i) = \frac{1}{M}$ and $\nu(B_i \cap B_j) = 0$ unless $i = j$.

(c1) If $\nu(B_i \cap B_j B_k B_l) \neq 0$ then $B_i \subset B_j(\alpha r) B_k B_l$.

(c2) If $\nu(B_i \cap B_j B_k^{-1} B_l) \neq 0$ then $B_j B_k^{-1} B_l \subset B_i(\alpha r)$.

(d1) $B_i(\alpha r) C_{\rho - \frac{\delta}{2}} \subset B_i C_\rho$.

(d2) $B_i(\alpha r) C_\rho \subset B_i(r) C_{\rho + \frac{3\delta}{2}}$.

In other words $\{B_i\}_{i=1}^M$ form a cover of G, and a measurable partition of G, satisfying certain additional properties. Clearly these properties are shared by any translated collection $\{gB_i\}$. Without loss we assume $e \in B_1$.

Furthermore, we claim that this collection of sets $\{B_i\}$ has the property that

$$m(Z_N \in C_{\rho+4\delta}) \geq \sum_{i=1}^{M} \nu(B_i)\mu(Y_N \in B_i C_{\rho+3\delta}) \tag{3.2}$$

$$\frac{1}{M} \sum_{i=1}^{M} \mu(Y_N \in B_i C_{\rho-3\delta}) \geq m(Z_N \in C_{\rho-4\delta}) \tag{3.3}$$

and that there exists $N(\epsilon)$ such that

$$\mu(Y_N \in B_i C_{\rho-3\delta}) \leq \mu(Y_N \in B_1 C_\rho) + 5\epsilon \tag{3.4}$$

$$\mu(Y_N \in B_1 C_\rho) \leq \mu(Y_N \in B_i C_{\rho+3\delta}) + 5\epsilon \tag{3.5}$$

for all $N \geq N(\epsilon)$ and $1 \leq i \leq M$.

Suppose that the claim is valid. Then by (3.4) and (3.5),

$$\frac{1}{M} \sum_i \mu(Y_N \in B_i C_{\rho-3\delta}) \leq \mu(Y_N \in B_1 C_\rho) + 5\epsilon \tag{3.6}$$

$$\mu(Y_N \in B_1 C_\rho) \leq \frac{1}{M} \sum_i \mu(Y_N \in B_i C_{\rho+3\delta}) + 5\epsilon. \tag{3.7}$$

Hence,

$$\begin{aligned} m(Z_N \in C_{\rho+4\delta}) &\geq \frac{1}{M} \sum_{i=1}^{M} \mu(Y_N \in B_i C_{\rho+3\delta}) \text{ from } (3.2) \\ &\geq \mu(Y_N \in B_1 C_\rho) - 5\epsilon \text{ from } (3.7) \\ &\geq \frac{1}{M} \sum_{i=1}^{M} \mu(Y_N \in B_i C_{\rho-3\delta}) - 10\epsilon \text{ from } (3.6) \\ &\geq m(Z_N \in C_{\rho-4\delta}) - 10\epsilon \text{ from } (3.3). \end{aligned}$$

In particular,

$$m(Z_N \in C_{\rho+4\delta}) + 5\epsilon \geq \mu(Y_N \in B_1 C_\rho) \geq m(Z_N \in C_{\rho-4\delta}) - 5\epsilon.$$

Combining this with (3.1), since ϵ and δ are arbitrarily small,

$$\lim_{N\to\infty} \mu(Y_N \in C_\rho) = \lim_{N\to\infty} m(Z_N \in C_\rho)$$

proving the theorem.

The remainder of the proof is structured as follows. In Step 2, we prove (3.2) and (3.3). The sets $\{B_i\}$ are constructed in Step 3. Steps 4–6 contain the proof of (3.4) and (3.5).

Step 2: Proof of (3.2) **and** (3.3)**.**

For sufficiently small $r > 0$ if $B \subset G$ is a set of diameter less than r then

$$gC_{\rho+4\delta} \supset BC_{\rho+3\delta} \supset BC_{\rho-3\delta} \supset gC_{\rho-4\delta} \tag{3.8}$$

for all $g \in B$.

If $A \subset \mathbb{R}^d$, then $m(Z_N \in A) = \int_G \mu(Y_N \in gA)d\nu(g) = \sum_{i=1}^M \int_{B_i} \mu(Y_N \in gA)d\nu(g)$. It follows from the inequalities (3.8) that

$$\begin{aligned} m(Z_N \in C_{\rho+4\delta}) &= \sum_{i=1}^M \int_{B_i} \mu(Y_N \in gC_{\rho+4\delta})d\nu(g) \geq \sum_{i=1}^M \nu(B_i)\mu(Y_N \in B_i C_{\rho+3\delta}) \\ &\geq \sum_{i=1}^M \nu(B_i)\mu(Y_N \in B_i C_{\rho-3\delta}) \geq \sum_{i=1}^M \int_{B_i} \mu(Y_N \in gC_{\rho-4\delta})d\nu(g) \\ &= m(Z_N \in C_{\rho-4\delta}). \end{aligned}$$

This proves (3.2) and (3.3).

Step 3: Construction of the sets B_i.

Lemma 3.1 *For all sufficiently small $r > 0$ there exists α, independent of r, and a collection $\{B_i\}_{i=1}^M$ of compact connected subsets of G satisfying $(a), (b), (c1)(c2), (d1), (d2)$ of Step 1.*

Proof Since G is a compact Riemannian manifold G has a triangulation into sets $\{A_i\}_{i=1}^m$ with $\mathrm{diam}(A_i) < r$ for all i. By perturbing the boundaries of the triangulation we may assume that the numbers $\{\nu(A_i)\}$ are rational, and hence have a greatest common denominator q. Express the numbers $\nu(A_i)$ with respect to the common denominator so that $\nu(A_i) = \frac{p_i}{q}$. Then subdivide each set A_i into p_i compact connected subsets $A_{i,j}, j = 1, \ldots, p(i)$ of equal measure. Finally relabel the sets $\{A_{i,j}\}$ as sets

$B_1, \ldots, B_M$. Given a set B_i we let $A_{r(i)}$ denote the unique triangle which contains B_i. Since G is compact and acts smoothly on itself there exists a α' such that for all $r > 0$ if $\operatorname{diam}(A_i) < r$, $i = 1, \ldots, M$ then $\operatorname{diam}(A_j A_k^{-1} A_l) < \alpha' r$. Suppose that $\nu(B_i \cap B_j B_k^{-1} B_l) \neq 0$. Then $B_i \cap B_j B_k^{-1} B_l \neq \emptyset$. Since $B_j B_k^{-1} B_l \subset A_{r(j)} A_{r(k)}^{-1} A_{r(l)}$ and $\operatorname{diam}(A_{r(j)} A_{r(k)}^{-1} A_{r(l)}) \leq \alpha' r$ we have $A_{r(j)} A_{r(k)}^{-1} A_{r(l)} \subset B_i(2\alpha' r)$. Thus $B_j B_k^{-1} B_l \subset B_i(2\alpha' r)$. We now define $\alpha = 2\alpha'$. This proves (c1), and (c2) is proved similarly.

For fixed ρ, δ we may choose $0 < r < \epsilon$ sufficiently small such that if $B \subset G$ is a set of diameter less than r then $B_i(\alpha r) C_{\rho - \frac{\delta}{2}} \subset B_i C_\rho$ and $B_i(\alpha r) C_\rho \subset B_i C_{\rho + \frac{3\delta}{2}}$. This proves (d1) and (d2). ∎

Step 4: Ergodicity.

Let B_i be a cover element, so $\nu(B_i) = \frac{1}{M}$. Since T^{-1} is ergodic,

$$\lim_{J \to \infty} \frac{1}{J} \sum_{j=1}^{J} I_{\Omega \times B_i^{-1}} \circ T^{-j}(x, g) = \frac{1}{M},$$

for a.e. $(x, g) \in \Omega \times G$. By Fubini's theorem, there exists $g_0 \in G$ such that the limit holds with $g = g_0$ for μ a.e. $x \in \Omega$. By equivariance, $I_{\Omega \times B_i^{-1}} \circ T^{-j}(x, g_0) = I_{\Omega \times g_0^{-1} B_i^{-1}} \circ T^{-j}(x, e)$. It is no loss of generality and also notationally simpler to redefine B_i as $B_i g_0^{-1}$. Then we have $\lim_{J \to \infty} \frac{1}{J} \sum_{j=1}^{J} I_{\Omega \times B_i^{-1}} \circ T^{-j}(x, e) = \frac{1}{M}$ for *a.e.* $x \in \Omega$.

Next recall that $T^j(x, e) = (f^j x, h_j(x))$ where $h_j(x) = h(x) h(fx) \cdots h(f^{j-1} x)$. Similarly, $T^{-j}(x, e) = (f^{-j} x, [h_j(f^{-j} x)]^{-1})$, and so we have $\lim_{J \to \infty} \frac{1}{J} \sum_{j=1}^{J} I_{B_i} \circ h_j(f^{-j} x) = \frac{1}{M}$ for *a.e.* $x \in \Omega$. Hence, there exists an integer J and a subset $\Omega' \subset \Omega$ with $\mu(\Omega') > 1 - \epsilon/2$ such that

$$\left| \frac{1}{J} \sum_{j=1}^{J} I_{B_i} \circ h_j(f^{-j} x) - \frac{1}{M} \right| \leq \epsilon$$

for all $x \in \Omega'$ and $i = 1, \ldots, M$.

Step 5: A new measure on $\Omega \times G$.

Let J be the integer fixed in Step 4. Define the measure σ on $\Omega \times G$ by

$$\sigma(E) = \frac{1}{J} \sum_{j=0}^{J-1} \mu\{x : T^j(x, e) \in E\}.$$

Again, if $C \subset \mathbb{R}^d$, we write $\sigma(Z_N \in C)$ instead of $\sigma\{(x, g) \in \Omega \times G : Z_N(x, g) \in C\}$.

Let Ω' be the subset of Ω defined in Step 4. By condition (1.2), there exists $N(\delta)$ and a subset $\widetilde{\Omega} \subset \Omega'$ with $\mu(\widetilde{\Omega}) > 1 - \epsilon$ such that for all $N \geq N(\delta)$, for all $0 \leq j \leq J$, and for all $x \in \widetilde{\Omega}$,

$$\left|Z_N \circ T^j(x,e) - Z_N(x,e)\right| < \delta.$$

As subsets of $\widetilde{\Omega} \times \{e\}$,

$$\{Z_N \in B_i C_{\rho-3\delta}\} \subset \{Z_N \circ T^j \in B_i C_{\rho-2\delta}\} \subset \{Z_N \circ T^j \in B_i C_{\rho-\delta}\} \subset \{Z_N \in B_i C_\rho\},$$

so by definition of σ,

$$\begin{aligned}\mu(Y_N \in B_i C_{\rho-3\delta}) \leq \sigma(Z_N \in B_i C_{\rho-2\delta}) + \epsilon \leq \sigma(Z_N \in B_i C_{\rho-\delta})\\ + \epsilon \leq \mu(Y_N \in B_i C_\rho) + 2\epsilon. \end{aligned} \tag{3.9}$$

We claim that for all $N > N(\delta)$ and $1 \leq i \leq M$,

$$\sigma(Z_N \in B_i C_{\rho-2\delta}) \leq \sigma(Z_N \in B_1 C_{\rho-\delta}) + 3\epsilon. \tag{3.10}$$

By (3.9) and (3.10),

$$\begin{aligned}\mu(Y_N \in B_i C_{\rho-3\delta}) &\leq \sigma(Z_N \in B_i C_{\rho-2\delta}) + \epsilon \leq \sigma(Z_N \in B_1 C_{\rho-\delta}) + 4\epsilon\\ &\leq \mu(Y_N \in B_1 C_\rho) + 5\epsilon\end{aligned}$$

for all i proving (3.4). The proof of (3.5) is similar, and we omit the details. It now remains to prove (3.10).

Step 6: Proof of (3.10)**.**

Since f is invertible and measure preserving,

$$\mu\{x \in \Omega : T^j(x,e) \in E\} = \mu\{y \in \Omega : (y, h_j(f^{-j}y)) \in E\}.$$

Hence,

$$\sigma(E) = \frac{1}{J}\sum_{j=0}^{J-1} \mu\{y \in \Omega : (y, h_j(f^{-j}y)) \in E\}.$$

The group elements $h_j(f^{-j}y)$ induce a coding of the point y in the symbol space $\{1,\ldots,M\}^J$ by $w(y) = (w_0(y), w_1(y), \ldots, w_J(y))$ where $w_j(y) = s \in \{1,\ldots,M\}$ if $h_j(f^{-j}y) \in B_s$. We define cylinder sets $\{S_i\}$, $i = 1,\ldots,r$, say, where each cylinder set consists of points in $\widetilde{\Omega}$ with the same coding, i.e. the cylinder sets are elements of the partition of $\widetilde{\Omega}$ induced by the equivalence relation $y \simeq z$ if and only if $w(y) = w(z)$. Each cylinder set is defined by a word in $\{1,\ldots,M\}^J$ and $\widetilde{\Omega}$ is partitioned by $\{S_i\}_{i=1}^r$. Note that to each letter $w_j(y)$ of the word $w(y)$ there corresponds the point $(y, h_j(f^{-j}y)) \in \Omega_j$ where $h_j(f^{-j}y) \in B_s$ if and only if $w_j(y) = s$. Furthermore by assumption, for all $p, q \in \{1,\ldots,M\}, y \in \widetilde{\Omega}$

$$\frac{1}{J}|\#\{j : w_j(y) = p\} - \#\{j : w_j(y) = q\}| \leq \epsilon.$$

Our key observation is that by G-equivariance for any set U, $Z_N(y, h_j(f^{-j}y)) \in U$ implies $Z_N(y, h_k(f^{-k}y)) \in h_k(f^{-k}y)h_j^{-1}(f^{-j}y)U$.

Lemma 3.2 *Fix B_p. We claim that there is a permutation $\{t(1),\ldots,t(M)\}$ of $\{1,\ldots,M\}$ such that for each $s\in\{1,\ldots,M\}$, $B_{t(s)}B_s^{-1}B_p\subset B_1(\alpha r)$.*

Proof Let $g_p\in B_p$ and consider the cover of G given by the sets $\{g_p^{-1}B_1, g_p^{-1}B_2,\ldots,g_p^{-1}B_M\} = \{C_1,\ldots,C_M\}$. We say a collection of sets $\{B_{i_1},\ldots,B_{i_r}\}$ is path connected if any sets B_{i_u},B_{i_v} may be connected by a path contained in $\cup_{j=1}^r B_{i_j}$.

Our proof is by induction and starts at $k=2$. Choose B_{i_1} and B_{i_2} which have a common boundary and hence are path connected. Now $\nu(B_{i_1}\cap C_j)\neq 0$ for some j. Choose such a j and define $t(i_1)=j$. Similarly $\nu(B_{i_2}\cap C_r)\neq 0$ for some r. We may choose $r\neq t(i_1)$ since otherwise $B_{i_2}\subset B_{t(i_1)}$ and because $\nu(B_{i_2})=\nu(B_{t(i_1)})$ this implies $\nu(B_{t(i_1)})\cap B_{i_1})=\nu(B_{i_2}\cap B_{i_1})=0$, a contradiction.

Suppose we have chosen a path connected set $\{B_{i_1},\ldots,B_{i_k}\}$ with corresponding permutation t defining a set $\{C_{t(i_1)},\ldots,C_{t(i_k)}\}$, such that $\nu(B_{i_j}\cap C_{t(i_j)})\neq 0$. We say B_{i_j} is labelled by $C_{t(i_j)}$. Now we choose $B_{i_{k+1}}$, path connected to $\{B_{i_1},\ldots,B_{i_k}\}$. If $B_{i_{k+1}}\not\subset\cup_{j=1}^k C_{t(i_j)}$ then there exists C_s such that $\nu(C_s\cap B_{i_{k+1}})\neq 0$ and we may choose $t(i_{k+1})=s$ to extend the set to $k+1$ members. A problem arises only if $B_{i_{k+1}}\subset\cup_{j=1}^k C_{t(i_j)}$.

So suppose $B_{i_{k+1}}\subset\cup_{j=1}^k C_{t(i_j)}$. We define $S_k=\cup_{j=1}^k C_{t(i_j)}$. Although we will redefine the map t later we keep S_k fixed.

Choose $C_{t(i_{s_1})}$ such that $C_{t(i_{s_1})}\cap B_{i_{k+1}}\neq 0$ and label $B_{i_{k+1}}$ with this set by defining $t(i_{k+1})=t(i_{s_1})$ and keeping $t(i_j)$ fixed for $j\notin\{k+1,s_1\}$. Now $B_{i_{s_1}}$ is unlabelled. If $B_{i_{s_1}}\not\subset S_k$ then there exists $C_t\subset S_k^c$ such that $\nu(C_t\cap B_{i_s})\neq 0$ and we may define $t(i_s)=t$ to extend the number of labelled sets to $k+1$. If however $B_{i_{s_1}}\subset S_k$ we choose $C_{t(i_{s_2})}$ such that $C_{t(i_{s_2})}\cap B_{i_{s_1}}\neq 0$ and relabel $t(i_{s_1})=t(i_{s_2})$. We continue this process of relabelling along all possible paths, hoping to find a set $B_{i_{s_r}}\not\subset S_k$, in which case we may extend the number of labelled sets to $k+1$. A possible problem is that the set of all such paths formed in this way starting at $B_{i_{k+1}}$ does not include a set $B_{i_{s_r}}\not\subset S_k$. To see that this is not the case suppose we encountered only sets $\{B_{i_{s_1}},\ldots,B_{i_{s_t}}\}$ covered wholly by S_k. Then in fact

$$B_{i_{k+1}}\cup\cup_{j=1}^t B_{i_{s_j}}\subset\cup_{j=1}^k C_{t(i_{s_j})}$$

since otherwise we could form a path including another set $B_{i_j}\notin\{B_{i_{s_1}},\ldots,B_{i_{s_t}}\}$. However $B_{i_{k+1}}\cup\cup_{j=1}^t B_{i_{s_j}}\subset\cup_{j=1}^k C_{t(i_{s_j})}$ leads to a contradiction as $B_{i_{k+1}}\cup\cup_{j=1}^t B_{i_{s_j}}$ has strictly greater measure than $\cup_{j=1}^k C_{t(i_{s_j})}$. Thus we may find a set $B_{i_{s_r}}\not\subset S_k$ and extend the number of labelled sets to $k+1$.

Thus we have inductively defined the bijection t such that $\nu(g_p^{-1}B_s\cap B_{t(s)})\neq 0$. Recall that the cover $\{B_1,\ldots,B_M\}$ has the property that if $\nu(B_i\cap B_jB_k^{-1}B_l)\neq 0$ then $B_jB_k^{-1}B_l\subset B_j(\alpha r)$. Since $\nu(g_p^{-1}B_s\cap B_{t(s)})\neq 0$, $g_p^{-1}\in B_{t(s)}B_s^{-1}$ and hence

$e \in B_{t(s)}B_s^{-1}B_p$. Thus $B_1 \cap B_{t(s)}B_s^{-1}B_p \neq 0$ and hence $B_{t(s)}B_s^{-1}B_p \subset B_1(\alpha r)$ as required. ∎

To sum up: for all j such that $w_j = t(s)$, for all $y \in S_i$ we have $h_j(f^{-j}y) \in B_{t(s)}$. Thus if $Z_N(y, h_j(f^{-j}y)) \in B_pC_{\rho-2\delta}$ and $w_j = s$ then if $w_k = t(s)$ it follows that $Z_N(y, h_k(f^{-k}y)) \in B_1(\alpha r)C_{\rho-2\delta} \subset B_1C_{\rho-\delta}$. Hence

$$\begin{aligned}
&\frac{1}{J}\sum_{j=0}^{J-1} \mu\{y \in S_i : Z_N(y, h_j(f^{-j}y)) \in B_pC_{\rho-2\delta}\} \\
&\quad = \frac{1}{J}\sum_{s=1}^{M}\sum_{w_j=s} \mu\{y \in S_i : Z_N(y, h_j(f^{-j}y)) \in B_pC_{\rho-2\delta}\} \\
&\quad \leq \frac{1}{J}\sum_{s=1}^{M}\sum_{w_j=t(s)} \mu\{y \in S_i : Z_N(y, h_j(f^{-j}y)) \in B_1(\alpha r)C_{\rho-2\delta}\} + \epsilon\mu(S_i) \\
&\quad = \frac{1}{J}\sum_{j=0}^{J-1} \mu\{y \in S_i : Z_N(y, h_j(f^{-j}y)) \in B_1(\alpha r)C_{\rho-2\delta}\} + \epsilon\mu(S_i) \\
&\quad \leq \frac{1}{J}\sum_{j=1}^{J} \mu\{y \in S_i : Z_N(y, h_j(f^{-j}y)) \in B_1C_{\rho-\delta}\} + \epsilon\mu(S_i).
\end{aligned}$$

This holds for all cylinder sets S_i so by summing over cylinder sets in $\widetilde{\Omega}$ we have

$$\begin{aligned}
&\frac{1}{J}\sum_{j=0}^{J-1} \mu\{y \in \widetilde{\Omega} : Z_N(y, h_j(f^{-j}y)) \in B_pC_{\rho-2\delta}\} \\
&\qquad \leq \frac{1}{J}\sum_{j=0}^{J-1} \mu\{y \in \widetilde{\Omega} : Z_N(y, h_j(f^{-j}y)) \in B_1C_{\rho-\delta}\} + \epsilon.
\end{aligned}$$

Finally including the error from the set $\widetilde{\Omega}$ we have

$$\sigma(Z_N \in B_pC_{\rho-2\delta}) \leq \sigma(Z_N \in B_1C_{\rho-\delta}) + 3\epsilon.$$

Since B_p is arbitrary, we have proved (3.10) as required. This completes the proof of Theorem 1.3.

Acknowledgements We thank Mike Field and Andrew Török for very helpful discussions and an anonymous referee for helpful comments. Both authors thank the University of Houston for hospitality and MN wishes to thank the Leverhulme Foundation and the Royal Society for financial support.

References

[1] M. Adahl, I. Melbourne and M. Nicol. Random iteration of Euclidean isometries. *Nonlinearity* **16** (2003) 977–987.

[2] P. Billingsley. *Convergence of Probability Measures.* Wiley, New York, 1968.

[3] P. Billingsley. *Probability and Measure.* Wiley, New York, 1986.

[4] M. J. Field, I. Melbourne, and A. Török. Decay of correlations, central limit theorems and approximation by Brownian motion for compact Lie group extensions. *Ergod. Th. & Dynam. Sys.* **23** (2003) 87–110.

[5] I. Melbourne and M. Nicol. Statistical properties of endomorphisms and compact group extensions. *J. London Math. Soc.* To appear.

[6] M. Nicol, I. Melbourne and P. Ashwin. Euclidean extensions of dynamical systems. *Nonlinearity* **14** (2001) 275–300.

[7] V. A. Rohlin. Exact endomorphisms of a Lebesgue space. *Izv. Akad. Nauk SSSR Ser. Mat.* **25** (1961) 499–530.

[8] L.-S. Young. Statistical properties of systems with some hyperbolicity including certain billiards. *Ann. of Math.* **147** (1998) 585–650.