Ann. Inst. Henri Poincaré,

Vol. 30, n° 3, 1994, p. 353-395 *Probabilités et Statistiques*

Random ergodic theorems with universally representative sequences

by

Michael LACEY (1)

Department of Mathematics, Indiana University,
Bloomington, IN 47405, U.S.A.

Karl PETERSEN (2), **Mate WIERDL** (4)

Department of Mathematics, CB 3250, Phillips Hall,
University of North Carolina,
Chapel Hill, NC 27599, U.S.A.

and

Dan RUDOLPH (3)

Department of Mathematics, University of Maryland,
College Park, MD 20742, U.S.A.

ABSTRACT. – When elements of a measure-preserving action of $\mathbb{R}^d$ or $\mathbb{Z}^d$ are selected in a random way, according to a stationary stochastic process, a.e. convergence of the averages of an L^p function along the resulting orbits may almost surely hold, in every system; in such a case we call the sampling scheme *universally representative*. We show that i.i.d. integer-valued sampling schemes are universally representative (with $p > 1$) if and only if they have nonzero mean, and we discuss a variety of other sampling schemes which have or lack this property.

Key words: random ergodic theorems, random sampling of stationary processes.

(1) Research supported by NSF Grant DMS-9003245 and an NSF Postdoctoral Fellowship.
(2) Research supported by NSF Grants DMS-8900136, DMS-9103656, and DMS-9203489.
(3) Research supported by NSF Grant DMS-01524351.
(4) Research supported by NSF Grants DMS-9103656 and DMS-9203489.

Classification A.M.S.: ergodic theory 28 D 05, 28 D 10, 28 D 15.

RÉSUMÉ. – Si des eléments d'une action de $\mathbb{R}^d$ ou $\mathbb{Z}^d$ préservant une mesure sont choisis selon un processus stochastique stationnaire, il est possible que presque sûrement les moyennes d'une fonction de classe L^p selon les orbites generées convergent p.p., pour chaque système; en un tel cas nous appelons le schéma d'échantillon *universellement représentatif.* Nous montrons que les suites i.i.d. à valeurs entières forment un schéma universellement représentatif (pour $p > 1$) si et seulement si elles ont une espérance non nulle. Nous considérons plusieurs autres exemples des schémas d'échantillon qui ont ou n'ont pas cette propriété.

1. INTRODUCTION

Suppose that elements of a measure-preserving action are applied in a random but stationary way, and we are interested in the almost everywhere convergence of the averages of some function along the orbits so generated. The random sequence of elements is fixed in advance, and the same sequence is applied in all actions. If the sequence is very regular, for example periodic, we will always have a.e. convergence of the averages for all actions and all integrable functions, because of the pointwise ergodic theorem; it is natural to ask whether a sequence that is typical in some sense, or sufficiently chaotic, stochastic, or complex, will produce the same results. We have found some examples of stationary processes that do produce a.e. convergence in this universal manner, but we also show that in general there exist counterexamples, even for independent identically distributed sequences.

To establish notation and terminology, let I be a set, which will index families $\{S_i : i \in I\}$ of measure-preserving transformations acting on probability spaces $(Y, \mathcal{C}, \nu)$. A *scheme* for choosing commuting m.p.t.'s at random in a stationary way will be provided by a shift-invariant measure P on $\Omega = I^{\mathbb{N}}$. Thus if $\omega \in \Omega$, we will be interested in the a.e. $d\nu(y)$ convergence of the averages

$$A_n^\omega g(y) = \frac{1}{n} \sum_{k=1}^{n} g(S_{w_k} \dots S_{\omega_1} y)$$

for all systems $(Y, \mathcal{C}, \nu, \{S_i : i \in I\}, g)$. We have in mind the following examples:

$I = (0, \infty)$: we make measurements of continuous-parameter stationary processes, the gaps between measurements having been determined by another, positive-real-valued stationary process; (1)

$I = \mathbb{Z}$: at each time apply a power of a fixed m.p.t. T chosen in advance in a stationary way; (2)

$I = \{0, 1\}$: apply commuting m.p.t.'s S_0 and S_1, perhaps chosen independently, perhaps according to some other stationary measure; (3)

$I = \{-1, 1\}$: at each time, apply either a m.p.t. T or T^{-1}; (4)

$I = \{0, 1, \ldots, d-1\}$: at each time apply one of d commuting m.p.t.'s perhaps chosen independently; (5)

$I = \mathbb{Z}^d$: we apply a stationary sequence of elements of a $\mathbb{Z}^d$ action. (6)

We will say that the scheme (Ω, P) $(\Omega = I^{\mathbb{N}})$ is *universally representative for* L^p if for a.e. $\omega \in \Omega$, for every probability space $(Y, \mathcal{C}, \nu)$, every ν-preserving action $\{S_i : i \in I\}$ on Y, and every $g \in L^p(Y, \mathcal{C}, \nu)$, the averages $A_n^\omega g(y)$ converge a.e. $d\nu(y)$. (The Random Ergodic Theorem deals with a weaker property, since it fixes the process being sampled and states that almost all sampling sequences yield almost everywhere convergence. This follows immediately from the pointwise ergodic theorem applied to the skew-product transformation $(\omega, y) \to (\sigma\omega, S_{\omega_1} y)$ on $\Omega \times Y$, where σ is the shift on Ω.) The question of mean convergence (in $L^p(Y)$) for a.e. ω is also of interest; it is answered below (Corollary to Proposition 1) for examples of type (1) and mentioned from time to time in remarks about the other examples.

First, we treat the case $I = \mathbb{R}$, random sampling in continuous-parameter flows. The return-time theorem of Bourgain, Furstenberg, Katznelson, and Ornstein [8] corresponds to the special case when the sampling times take values in a lattice in $\mathbb{R}$. This return-time theorem, in the formulation involving averages of products of pairs of functions, extends to the real-parameter case (Theorem 1). This implies that any sampling scheme always works correctly on each function in a dense set in L^1, namely those that are smooth in the time parameter (Proposition 1); consequently convergence always holds in the mean of order $p \geqq 1$ (Corollary 1). In general, however, almost everywhere convergence can fail. For example, if the spacing process has a continuous distribution in some interval, then with probability 1 the sampling sequence will fail on some function in any real-parameter process (Example 1). On the other hand, there also exist nontrivial examples of universally representative sampling schemes. For example, if the spaces between sampling times take values 1 and α, where α may be irrational, the selections being made according to a suitable Markov measure, then the sampling scheme is universally representative (Example 2). By a modification of this idea, we can also construct universally representative schemes that allow arbitrarily small gaps between measurement times (Examples 3 and 4). Given a universally representative sampling scheme, certain kinds of perturbations of it will produce another universally representative scheme (Example 5). Although it seems likely that there are universally representative schemes in which the gaps between measurements are independent and take values in a countable set clustering at 0, so far we have not been able to construct a particular example.

The remainder of the paper concerns the random selection of transformations from a commuting family, perhaps from an aperiodic action of several commuting transformations or from the set of powers of a single transformation. The case when the selections are made according to a Bernoulli scheme (*i.e.*, in an i.i.d way), is already of considerable interest. Suppose first that we choose either a transformation or its inverse according to the Bernoulli measure $\mathcal{B}(1/2, 1/2)$; then because of the recurrence of the symmetric one-dimensional random walk (in n steps approximately $\sqrt{n}$ sites are visited, each one approximately $\sqrt{n}$ times), at first glance one might supose that when we look at the Cesaro averages we are seeing good weights on powers of a single m.p.t., and so we should obtain a good sequence with probability 1. We show (Theorem 4), by using Strassen's functional law of the iterated logarithm, that in fact this is *not* the case: in any aperiodic system, there is always a counterexample. On the other hand, whenever commuting m.p.t.'s are applied according to the Bernoulli scheme $\mathcal{B}(1/2, 1/2)$, we have a.e. convergence of the Cesaro averages along the subsequence $2^{n \log n}$ (Theorem 7); this is proved by the Fourier method introduced into ergodic theory by J. Bourgain. If powers of a single m.p.t. are to be applied according to the values of an integer-valued independent sequence of integrable random variables, then there is always a.e. convergence of the ergodic averages for L^p functions, $p > 1$, if and only if the sampling process has nonzero mean (Theorem 5 and 6, combined with Theorem 4). Choosing among commuting transformations according to the entries in a certain Sturmian sequence does provide an example of a universally representative scheme (Theorem 3).

The variety of behavior found in these examples makes the formulation of a condition that might be necessary and sufficient for a sampling scheme to be universally representative seem fairly remote at this time. The difficulty (and interest) of questions of this kind comes from two sources: inappropriate averaging, either the discrete sampling of a flow or the one-dimensional sampling of a two-dimensional action; and the need to deal simultaneously with uncountably many sampled processes, combining uncountably many exceptional sets of measure zero into a single set of measure zero. Sampling theorems of a different kind were considered in [14], [4].

The authors thank M. Akcoglu, M. Boshernitzan, and J. King for helpful suggestions and encouragement and the referee for many improvements in the presentation.

2. RANDOM SAMPLING OF CONTINUOUS-PARAMETER STATIONARY PROCESSES

Let $\{\delta_k(\omega) : k = 1, 2, \ldots\}$ be a positive real-valued integrable stationary process and $\tau_k(\omega) = \delta_1(\omega) + \ldots + \delta_k(\omega)$, $\tau_0 = 0$. The $\tau_k(\omega)$ form our sequence of sampling times, and the $\delta_k(\omega)$ are the gaps between measurements. The sampling scheme $\{\delta_k\}$ is *universally representative* (for L^1) if for a.e. ω, $\{\tau_k(\omega)\}$ is a good sequence, in the following sense: For every measure-preserving flow $\{S_t : -\infty < t < \infty\}$ on a measure space $(Y, \mathcal{C}, \nu)$ and integrable function g on Y,

$$A_n\, g(y) = \frac{1}{n} \sum_{k=1}^{n} g(S_{\tau_k(\omega)}\, y) \text{ converges a.e. } d\nu(y).$$

Because of the theorem on ergodic decompositions, we may as well assume that the sampling process as well as all of the processes being sampled are ergodic. By subtracting off a constant, we may also assume that the integral of g is 0. Letting $\Omega = (0, \infty)^{\mathbb{N}}$ with the measure P generated by $\{\delta_k\}$, we have the representation $\delta_k(\omega) = \delta(\sigma^{k-1}\omega)$, where $\sigma : \Omega \to \Omega$ is the shift transformation and $\delta : \Omega \to (0, \infty)$ reads off the first coordinate of ω.

Discrete samples and return times. – To see how sampling theorems are connected with return-time theorems in the particular case when δ is $\mathbb{N}$-valued (and similarly, with suitable adjustments, when δ takes values in a lattice in $\mathbb{R}$), we consider the primitive transformation $\tilde{\sigma} : \tilde{\Omega} \to \tilde{\Omega}$ built over the floor Ω with ceiling function $\delta - 1$ (*see* [15], p. 40). Then $\tau_k(\omega)$ is the time of the k'th entry of $\omega \in \Omega$ to Ω under $\tilde{\sigma}$. Thus if $J = \tau_n(\omega)$, then in the first J steps of its orbit under $\tilde{\sigma}$, the number of times that ω enters Ω is $N_J(\omega, \Omega) = n$, and we have

$$\begin{aligned}
\frac{1}{n} \sum_{k=1}^{n} g(S^{\tau_k(\omega)}\, y) &= \frac{1}{n} \sum_{j=1}^{\tau_n(\omega)} \chi_\Omega(\tilde{\sigma}^j\, \omega)\, g(S^j\, y) \\
&= \frac{1}{N_J(\omega, \Omega)} \sum_{j=1}^{J} \chi_\Omega(\tilde{\sigma}^j\, \omega)\, g(S^j\, y) \\
&= \frac{1}{N_J(\omega, \Omega)/J}\, \frac{1}{J} \sum_{j=1}^{J} \chi_\Omega(\tilde{\sigma}^j\, \omega)\, g(S^j\, y).
\end{aligned}$$

For a.a. ω, the first factor converges to $\frac{1}{\mu(\Omega)}$ (where μ is the usual extension of P to $\tilde{\Omega}$). That for a.a. ω the second factor converges a.e. $d\nu(y)$ for all $(Y, \mathcal{C}, \nu, g)$ is the return-time theorem of [8] (which applies just as well to bounded measurable functions as to characteristic function like χ_Ω).

THEOREM 1. – *If* $\{T_t : -\infty < t < \infty\}$ *is an ergodic measure-preserving flow on a finite measure space* $(X, \mathcal{B}, \mu)$ *and* $f \in L^\infty(X, \mathcal{B}, \mu)$, *then* a.e. $x \in X$ *has the following property: For each system* $(Y, \mathcal{C}, \nu, \{S_t\}, g)$, *where* $\{S_t : -\infty < t < \infty\}$ *is an ergodic measure-preserving flow on the finite measure space* $(Y, \mathcal{C}, \nu)$ *and* g *is an integrable function on* Y,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t f(T_s x)\, g(S_s y)\, ds \ \textit{exists for a.e. } y\,(d\nu).$$

Proof. – We assume that g has integral 0. Moreover, if f is in the closed linear span in L^2 of the eigenfunctions of $\{T_t\}$, then existence and identification of the limit both follow easily from the Ergodic Theorem, so by subtracting the projection of f onto this subspace (which is also bounded since it is a conditional expectation with respect to the corresponding factor algebra) we may assume that f is in the orthocomplement in L^2 of the eigenfunctions of $\{T_t\}$. In this case we want to show that the limit above is 0 for almost all y.

It is known that all but perhaps countably many of the maps T_t are ergodic; rescaling if necessary, we may assume that T_1 is ergodic, and hence $T_{1/r}$ is ergodic for each $r \in \mathbb{N}$. The set of good x whose existence is claimed in the theorem is formed by deleting the following sets of measure 0 from X. First, apply the return-time theorem of [8] to f and each of the maps $T_{1/r}$, $r \in \mathbb{N}$, discarding a bad set of measure 0 for each r. Next, notice that by the same Fubini argument that proves the Local Ergodic Theorem,

$$D_h f(x) = \frac{1}{h} \int_0^h |f(T_s x) - f(x)|\, ds \to 0 \quad \text{as} \quad h \to 0 \qquad \text{for a.e. } x.$$

For each $r \in \mathbb{N}$ and $q \in \mathbb{Q}$, discard the set of measure 0 of those x which are not generic for the set $\{u : D_{1/r} f(u) > q\}$, in the sense that the orbit of x under $T_{1/r}$ fails to visit this set with the correct limiting frequency.

Supose now that x is in this remaining good set and that $(Y, \mathcal{C}, \nu, \{S_t\}, g)$ is given. It is sufficient to consider averages over intervals of length $t = n \in \mathbb{N}$. Upon dividing each interval $[j, j+1]$ into r subintervals $\left[\frac{k}{r}, \frac{k+1}{r}\right]$ we find that

$$\frac{1}{n}\int_0^n f(\mathrm{T}_s x)\, g(\mathrm{S}_s y)\, ds = \frac{1}{n}\sum_{k=0}^{rn-1}\left\{\int_{k/r}^{(k+1)/r} g(\mathrm{S}_s y)\,[f(\mathrm{T}_s x) - f(\mathrm{T}_{k/r} x)]\, ds + \int_{k/r}^{(k+1)/r} f(\mathrm{T}_{k/r} x)\,[g(\mathrm{S}_s y) - g(\mathrm{S}_{k/r} y)]\, ds + \frac{1}{r} f(\mathrm{T}_{k/r} x)\, g(\mathrm{S}_{k/r} y)\right\}. \quad (1)$$

Since $f \in \mathrm{L}^\infty$, we have a maximal inequality for

$$\sup_{t>0}\left|\frac{1}{t}\int_0^t f(\mathrm{T}_s x)\, g(\mathrm{S}_s y)\, ds\right| \leqq \|f\|_\infty \sup_{t>0}\frac{1}{t}\int_0^t |g(\mathrm{S}_s y)|\, ds$$

coming from the Maximal Ergodic Theorem, so it is enough to prove convergence a.e. for a dense set of functions $g \in \mathrm{L}^1(\mathrm{Y})$, such as those with $g(\mathrm{S}_s y)$ uniformly continuous in s, uniformly in y. (The familiar functions

$$g(y) = \int_{-\infty}^{\infty} h(\mathrm{S}_s y)\, \phi(s)\, ds$$

with $h \in \mathrm{L}^\infty(\mathrm{Y})$ and $\phi \in \mathcal{C}(\mathbb{R})$ with compact support have

$$|g(\mathrm{S}_s y) - g(\mathrm{S}_t y)| \leqq \|h\|_\infty \|\phi_{s-t} - \phi\|_{\mathrm{L}^1(\mathbb{R})}$$

where $\phi_s(u) = \phi(u+s)$.)

Dealing now with such a special g, choose r large enough that the second term in (1) is less than a given $1/\mathrm{K}$ for all n.

Since $\mathrm{D}_h f(u) \to 0$ a.e., we may choose r also large enough that $\mu\{u : \mathrm{D}_{1/r} f(u) > 1/\mathrm{K}\} < 1/\mathrm{K}$. Because the orbit of x under $\mathrm{T}_{1/r}$ visits $\{\mathrm{D}_{1/r} f > 1/\mathrm{K}\}$ with the right frequency, we can choose n large enough that

$$\frac{1}{rn}\sum_{k=0}^{rn-1} \chi_{(\mathrm{D}_{1/r} > 1/\mathrm{K}\}}(\mathrm{T}_{k/r} x) < 2\,\frac{1}{\mathrm{K}}.$$

Then the first term in (1) is bounded by

$$\|g\|_\infty \frac{1}{rn} \sum_{k=0}^{rn-1} \frac{1}{1/r} \int_{k/r}^{(k+1)/r} \left|f(\mathrm{T}_s x) - f(\mathrm{T}_{k/r} x)\right| ds$$

$$\leqq \|g\|_\infty \frac{1}{rn} \sum_{k=0}^{rn-1} \left[\frac{1}{\mathrm{K}} \chi_{\{\mathrm{D}_{1/r} \leqq 1/\mathrm{K}\}} (\mathrm{T}_{k/r} x) \right.$$

$$\left. + 2 \|f\|_\infty \chi_{\{\mathrm{D}_{1/r} > 1/\mathrm{K}\}} (\mathrm{T}_{k/r} x) \right]$$

$$\leqq \|g\|_\infty \left[\frac{1}{\mathrm{K}} + 2 \|f\|_\infty \frac{2}{\mathrm{K}} \right],$$

which is arbitrarily small for large K.

Finally, since r was fixed first, using the return-time theorem from [8] for a.e. y we may choose n large enough that the third term in (1) is also less than $1/\mathrm{K}$.

Remark. – E. Lesigne (personal communication) has observed that the argument of [8], when applied to integral averages, also yields this result.

In the case of a general sampling scheme, we may be considering returns to a set which has measure 0. With the notation above, if $\delta : \Omega \to (0, \infty)$ is the function generating the spacings between sampling times, consider the flow under the function δ with base (Ω, σ). That is, we now let $\mathrm{X} = \{(\omega, t) : 0 \leqq t < \delta(\omega)\}$ and define a flow $\{\mathrm{T}_s : -\infty < s < \infty\}$ on X by letting each point $(\omega, 0)$ flow up at unit speed until it hits the ceiling, at which point it moves to $(\sigma\omega, 0)$. Then $\delta_1(\omega), \delta_2(\omega), \ldots$ are the spacings between hits of the floor Ω, so that the $\tau_k(\omega)$ are exactly the return times of $(\omega, 0)$ to Ω under this flow.

In order to have a set of positive measure to return to, we take a narrow band below the ceiling and work with its characteristic function

$$f_h(x) = \frac{1}{h} \chi_{\{(\omega, t) :\, 0 \leqq \delta(\omega) - h \leqq t < \delta(\omega)\}}(x).$$

For small h, f_h has integral very near 1.

PROPOSITION 1. – *Any sampling scheme is representative for the (dense) set* $\mathcal{D}(\mathrm{Y})$ *of functions* $g \in \mathrm{L}^1(\mathrm{Y})$ *for which* $g(\mathrm{S}_s y)$ *is a uniformly continuous function of* s *uniformly in* y.

Proof. – As before, we may assume that g has mean 0. We consider h taking values in a sequence tending to 0. Suppose ω is such that a.e. point $x = (\omega, t)$ is good in the sense of the continuous-parameter return-time theorem (Theorem 1) for all the functions f_h as above, as h varies in this countable set. Suppose further that ω is generic for the function δ on Ω, in the sense of the ergodic theorem and the transformation σ. Then

$$\int_{\tau_{k-1}}^{\tau_k} f_h(\mathrm{T}_s(\omega, 0))\, g(\mathrm{S}_s y)\, ds = 0 \quad \text{unless}\, \sigma^{k-1}\omega \in \{\delta \geqq h\},$$

so that if $x = (\omega, \eta)$ we have

$$\begin{aligned} \mathrm{A}_n g(y) = {} & \frac{\tau_n(\omega)}{n} \frac{1}{\tau_n(\omega)} \int_0^{\tau_n(\omega)} f_h(\mathrm{T}_s x)\, g(\mathrm{S}_s y)\, ds \\ & - \frac{1}{n} \sum_{k=1}^{n} \chi_{\{\delta \geqq h\}}(\sigma^{k-1}\omega) \\ & \times \int_{\tau_k - h}^{\tau_k} \left[f_h(\mathrm{T}_s(\omega, 0))\, g(\mathrm{S}_{s-\eta} y) - \frac{1}{h} g(\mathrm{S}_{\tau_k} y) \right] ds \\ & + \frac{1}{n} \sum_{k=1}^{n} \chi_{\{\delta < h\}}(\sigma^{k-1}\omega)\, g(\mathrm{S}_{\tau_k} y) \\ & - \frac{1}{n} \int_{\tau_n}^{\tau_n + \eta} f_h(\mathrm{T}_s(\omega, 0))\, g(\mathrm{S}_{s-\eta} y)\, ds \\ & + \frac{1}{n} \int_0^{\eta} f_h(\mathrm{T}_s(\omega, 0))\, g(\mathrm{S}_{s-\eta} y)\, ds. \end{aligned} \tag{2}$$

Given $\varepsilon > 0$, choose $\eta \geqq 0$ very small so that g has oscillation less than $\varepsilon/2$ on any interval of length η and $x = (\omega, \eta)$ is good, in the sense of Theorem 1, for all the f_h with h in our countable set. Then choose h so small that the second term in (2) is less than ε for all n and y. For large n, the third term is of the order of $\mu\{\delta < h\}_{\|g\|_\infty}$; this can be made less than ε by choosing h small and n large. The fourth and fifth terms tend to 0 as $n \to \infty$ because they are bounded by $\eta\, \|g\|_\infty / nh$. Finally, by Theorem 1 we can choose n also large enough that the first term is also less than ε.

COROLLARY 1. – *Let* (Ω, P) *be a sampling scheme as above. Then for* a.e. ω, *for every measure-preserving flow* $\{\mathrm{S}_t : -\infty < t < \infty\}$ *on a measure space* $(\mathrm{Y}, \mathcal{C}, \nu)$, $p \geqq 1$, *and* $g \in \mathrm{L}^p(\mathrm{Y})$,

$$\mathrm{A}_n g(y) = \frac{1}{n} \sum_{k=1}^{n} g(\mathrm{S}_{\tau_k(\omega)} y) \text{ converges in } \mathrm{L}^p.$$

Proof. – Always assuming ergodicity and mean 0, for $h \in \mathcal{D}(\mathrm{Y})$,

$$\|\mathrm{A}_n g\|_p \leqq \|\mathrm{A}_n (g - h)\|_p + \|\mathrm{A}_n h\|_p \leqq \|g - h\|_p + \|\mathrm{A}_n h\|_p,$$

which can be made arbitrarily small by choosing first h and then n.

COROLLARY 2. – *If a sampling scheme* (Ω, P) *is such that for* a.e. ω, *for all* $(\mathrm{Y}, \mathcal{C}, \nu, \{\mathrm{S}_t\})$ *the averages satisfy a maximal inequality on* $\mathrm{L}^p(\mathrm{Y})$ $(p \geqq 1)$, *then the scheme is universally representative for* L^p.

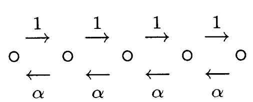
3. EXAMPLES OF BAD AND GOOD SAMPLING SCHEMES

Example 1. – Suppose that the spacing function δ is uniformly distributed in $[1/2, 1]$ and that the spaces $\delta\sigma^k$ are all independent. Then with probability 1 the sampling times τ_k and 1 are linearly independent over $\mathbb{Q}$, so it follows from [3] that the sampling scheme is not universally representative in this case: for a set of ω of probability 1, for every nontrivial $(\mathrm{Y}, \mathcal{C}, \nu, \{\mathrm{S}_s\})$ there is a bounded measurable g on Y for which the averages $\mathrm{A}_n\, g(y)$ of the samples at the times determined by ω fail to converge a.e. It can be argued that this is the type of sampling scheme that is necessarily employed by actual scientists, and that therefore, in view of this example and the Corollary to Proposition 1, von Neumann was right when he argued that the Mean Ergodic Theorem, rather than the pointwise one, was the real ergodic theorem.

Example 2. – We will show that there are some nontrivial Markov processes that do give rise to universally representative sampling schemes. Suppose that the sampling process $(\Omega, \mathcal{F}, \mathrm{P}, \delta)$ is such that δ takes values in $\{1, \alpha\}$ for some $\alpha > 0$ and that the difference between the number of 1's and the number of α's seen remains bounded: there is K such that

$$\left| \sum_{k=1}^{n} [\chi_{\{1\}}(\delta(\sigma^k \omega)) - \chi_{\{\alpha\}}(\delta(\sigma^k \omega))] \right| \leqq \mathrm{K}$$

for all $n = 1, 2, \ldots$ and a.e. $\omega \in \Omega$ (actually K could also depend on ω). Particular examples can be obtained from stationary Markov measures on run-limited graphs like



Then the sampling times $\tau_k(\omega)$ are in $\mathbb{Z} \cup \mathbb{Z}\alpha$, a countable dense set in $\mathbb{R}$, but of course the spaces between them are bounded below. If $0 \leqq g \in L^1(Y)$, then

$$A_n g(y) = \frac{1}{n} \sum_{k=1}^{n} g(S_{\tau_k(\omega)} y) \leqq \frac{1}{n} \sum_{k=1}^{n} (S_1 S_\alpha)^k \left(\sum_{r, s=-K}^{K} S_1^r S_\alpha^s \right) g(y)$$
$$\leqq (S_1 S_\alpha)^* G(y),$$

where

$$G(y) = \left(\sum_{r, s=-K}^{K} S_1^r S_\alpha^s \right) g(y)$$

and $(S_1 S_\alpha)^*$ denotes the usual ergodic maximal operator applied to $S_1 S_\alpha$. From this the weak 1,1 inequality for $\sup_n A_n g(y)$ follows immediately, for all $\omega \in \Omega$. Since by Proposition 1 a.e. ω yields a.e. convergence of $A_n g(y)$ for a dense set of $g \in L^1(Y)$, it follows that for almost all ω the sampled averages $A_n g(y)$ converge a.e. $dv(y)$ for all $g \in L^1(Y)$.

Example 3. – We will extend the previous example to one in which the spacing function δ takes arbitrarily small positive values, say in a sequence r_j that tends to 0. Under suitable conditions we will obtain a sampling scheme that is universally representative for $L^p(Y)$ for $p > 1$.

Let each $r_j \in (0, 1)$, and let P again have memory such that a step of size r_j is always followed immediately by one of size $1 - r_j$. For example, such a system can be constructed from an arbitrary measure-preserving system (Ω_0, P_0, T_0) and countable partition $\{A_1, A_2, \ldots\}$ as follows. Let (Ω, P, T) be the primitive transformation built by putting a "second floor" over (Ω_0, P_0), the second floor being merely a copy (Ω_0, P_0). (Points on the first floor are mapped up by the identification map to corresponding points in the second floor, and points in the second floor are mapped by T_0 and then to corresponding points in the first). We define $\delta \equiv r_j$ on A_j and $\delta \equiv 1 - r_j$ on TA_j.

We claim that if $p_j = P(A_j)$ decreases to 0 sufficiently rapidly, then for a.e. ω, for all Y, for all $g \in L^p(Y)$ the surpremum (over n) of the averages $A_n g(y)$ along the times $\tau_k(\omega)$ will be finite a.e. $d\nu(y)$, and hence, by Banach's Principle and Proposition 1, the sampling scheme will be universally representative for $L^p(Y)$. (For bounded g a more direct argument using only the return times theorem is possible). For simplicity, let us suppose first that $p > 2$ and $g \in L^p(Y)$. Then for $m = 0, 1, \ldots$, $\tau_{2m}(\omega) = m$ and $\tau_{2m+1}(\omega) = m + \delta(T^{2m}\omega)$ for all ω on the first floor of Ω, so that for $0 \leqq g \in L^p(Y)$

$$A_n g(y) = \frac{1}{n}\sum_{k=1}^{n} g(S_{\tau_k(\omega)} y) = \frac{1}{n}\sum_{k \leqq n/2} g(S_k y) + \frac{1}{n}\sum_{k<n/2} g(S_{k+\delta(T^{2k}\omega)} y).$$

The surpremum of the first term is controlled by the ergodic maximal function $S_1^* g$. To deal with the second term, for each j let $\chi_j = \chi_{\{\delta = r_j\}}$ and let χ_j^* denote the ergodic maximal function of χ_j under T^2. Then the surpremum of the second term is bounded by

$$\begin{aligned}&\sum_j \sup_n \frac{1}{n}\sum_{k<n/2} \chi_j(T^{2k}\omega) S_k S_{r_j} g(y)\\ &\leqq \sum_j \left[\sup_n \frac{1}{n}\sum_{k<n/2}\chi_j(T^{2k}\omega)\right]^{1/2} \times \left[\sup_n \frac{1}{n}\sum_{k<n/2} S_k S_{r_j} g^2(y)\right]^{1/2}\\ &\leqq \sum_j [\chi_j^*(\omega)]^{1/2}[S_1^*(S_{r_j} g^2)(y)]^{1/2}.\end{aligned}$$

We claim that it is possible to choose the probabilities p_j so that

$$\sum_j (\chi_j^*(\omega))^{1/2} < \infty \text{ almost surely;}$$

let us assume this for the moment and note that if $g \in L^p$ for some $p > 2$, then $S_1^*(S_{r_j} g^2) \in L^{p/2}$ for each j, by the Dominated Ergodic Theorem, so that its square root is again in $L^p \subset L^1$ (with L^1 norm bounded independently of j). Then the above estimate shows that

$$\int \sup_n A_n g(y)\, d\nu(y) \leqq \int S_1^* g(y)\, d\nu(y) + \sum_j [\chi_j^*(\omega)]^{1/2} \int [S_1^*(S_{r_j} g^2)(y)]^{1/2} d\nu(y) < \infty \quad \text{a.s.}$$

Therefore the same result also follows for a sequence of sampling times $\{\tau_k(\omega)\}$ generated by a point ω on the second floor of Ω.

It remains to show that one can choose the p_j so as to guarantee the almost sure convergence of the series $\sum (\chi_j^* (\omega))^{1/2}$; for example, $p_j = C\, 4^{-j}$ will work. For then, by the Maximal Ergodic Theorem,

$$P\{\chi_j^* > 2^{-j}\} \leqq 2^j \|\chi_j\|_1 = 2^j P\{\delta = r_j\} = 2^j p_j = C\, 2^{-j},$$

so that $\sum P\{\chi_j^* > 2^{-j}\} < \infty$, and hence with probability 1 we have $\chi_j^* \leqq 2^{-j}$ for large enough j.

In the case of an arbitrary $p > 1$ the same argument applies, except that the appropriate version of Hölder's inequality should be used. The requirement then becomes $\sum (\chi_j^* (\omega))^\alpha < \infty$ a.s., where $p > (1/\alpha)'$. Again this can be arranged by choosing the $p_j = P\{\delta = r_j\}$ suitably.

Example 4. – Upon deleting the times $\tau_k(\omega)$ that are in $\mathbb{Z}$ from the sampling scheme in Example 3, we obtain a scheme that might be termed an example of *random delay*: the k'th measurement is taken at time $\tau_k(\omega) = k + \delta(\sigma^k \omega)$, where $0 < \delta(\omega) < 1$ for all ω, so that the size of the gap between the k'th and $(k+1)$'st measurements is $\gamma_k(\omega) = 1 + \delta(\sigma^{k+1}\omega) - \delta(\sigma^k \omega) = \gamma(\sigma^k \omega)$, if $\gamma(\omega) = 1 + \delta(\sigma\omega) - \delta(\omega)$. Notice that again $\sum (\gamma - 1)(\sigma^k \omega)$ is bounded for each ω, exposing the connection of these examples with Example 2. Since $\mathbb{Z}$ is universally representative for $L^p(Y)$, this scheme will be also whenever p and the p_j are chosen so that the corresponding scheme in Example 3 is universally representative for $L^p(Y)$.

Question. – Is it possible to generate a universally representative sequence of sampling times if the spaces $\delta_k = \delta \circ \sigma^k$ between times are chosen independently from a set $\{r_j\}$ clustering at 0? Will the sequence automatically be universally representative if $P\{\delta = r_j\} \to 0$ fast enough?

Example 5. – The idea behind Examples 3 and 4 actually permits construction of a fairly wide class of examples. Roughly speaking, if a sampling scheme admits a return-times type theorem, then so does any countable-valued stochastic perturbation of the scheme, if its tail distribution vanishes sufficiently quickly. To make this precise, let $\{\tau_k(\omega)\}$ be a sequence of sampling times, i.e., an increasing sequence of nonnegative real-valued functions on a probability space $(\Omega, \mathcal{F}, P)$; we say that the sequence *has return times for* p $(\geqq 1)$ in case for all $f \in L^\infty(\Omega)$, for almost all ω, given a continuous-parameter measure-preserving flow $\{S_t\}$ on a probability space $(Y, \mathcal{C}, \nu)$ and $g \in L^p(Y)$,

$$A_n^\omega g(y) = \frac{1}{n} \sum_{k=1}^{n} f(T^k \omega)\, g(S_{\tau_k(\omega)}\, y) \quad \text{converges a.e. } d\nu(y).$$

For example, the sequence $\tau_k = k$ has return times for each $p \geqq 1$. As usual, let q denote the index dual to p.

THEOREM 2. – *Let* $p \geqq 1$. *Suppose that* $\{\tau_k\}$ *has return times for* p *and that* $\beta : \Omega \to \{\beta_1, \beta_2, \ldots\}$ *is a countable-valued measurable function for which there is a sequence* $\{a_j\}$ *of positive reals such that* $\sum 1/a_j < \infty$ *and*

$$\sum_j a_j \, \|\chi_{\{\beta=\beta_j\}}\|_q < \infty.$$

Then $\{\tau_k + \beta \, \mathrm{T}^k\}$ also has return times for p.

Remarks.

1. Similarly, for any measurable function γ on Ω, $\sigma_k(\omega) = \tau_k(\omega) + \beta(\mathrm{T}^k \omega) - \gamma(\omega)$ also has return times.

2. The situation of the theorem can be generalized still further. For example, in the definition of "has return times", we may suppose that instead of transformations $\mathrm{S}_{\tau_k(\omega)}$ sampled from a measure-preserving flow, we are dealing with a random sequence of contractions $\{\mathrm{U}_k(\omega)\}$ on $\mathrm{L}^p(y)$. The perturbation is formed by taking a sequence of contractions $\mathrm{V}_1, \mathrm{V}_2, \ldots$ on $\mathrm{L}^p(\mathrm{Y})$ and a measurable function $j : \Omega \to \mathbb{N}$ and replacing $\sigma_k(\omega) = \tau_k(\omega) + \beta \, \mathrm{T}^k(\omega)$ by $\mathrm{W}_k(\omega) = \mathrm{U}_k(\omega) \, \mathrm{V}_{j(\mathrm{T}^k \omega)}$. Under the condition that $\sum a_n \, \|\chi_{\{j(\omega)=n\}}\|_q < \infty$, the sequence $\{\mathrm{W}_k(\omega)\}$ will also have return times in this sense. (A particular case might involve the deflection of the action of a combination of commuting m.p.t.'s $(\mathrm{ST})^k$ off course, by applying powers of either S or T at random.)

3. If we are concerned only with $g \in \mathrm{L}^\infty$, we can dispense with the assumption involving the sequence $\{a_j\}$.

Proof of Theorem 2. – We may assume without loss of generality that $f, g \geqq 0$. Let $\chi_j = \chi_{\{\beta=\beta_j\}}$, and abbreviate $h_j = f \chi_j$, $g_j = g \, \mathrm{S}_{\beta_j}$, and $g_{j,k} = g_j \, \mathrm{S}_{\tau_k(\omega)}$. The good set of $\omega \in \Omega$ will consist of all those points for which (1) there is a $j_0(\omega)$ such that for $j \geqq j_0(\omega)$ we have

$$(h_j^q)^* \leqq a_j \, \|h_j^q\|_1,$$

where * denotes the ergodic maximal function for T $\Big($this is a set of full measure, since by the Maximal Ergodic Theorem and hypothesis,

$$\sum_j \mathrm{P}\,\{(h_j^q)^* > a_j \, \|h_j^q\|_1\} \leqq \sum_j \frac{1}{a_j \, \|h_j^q\|_1} \, \|h_j^q\|_1 < \infty \Big);$$

(2) ω is good, in the sense of the return-times hypothesis, for each h_j.

Notice that also for each k there is a set of full measure of y for which there is a $j_0(y)$ such that for $j \geqq j_0(y)$,

$$(g_{j,k}^p)^* \leqq a_j \, \|g_j^p\|_1;$$

in what follows, we will consider only those y that are in these sets of full measure for all k. Also, we let $j_0(\omega, y) = \max\{j_0(\omega), j_0(y)\}$. For $p > 1$ the interchange of limit and sum in the calculation

$$\begin{aligned}\lim_n \mathrm{A}_n^\omega g(y) &= \lim_n \sum_{j=1}^{\infty} \frac{1}{n} \sum_{k=1}^{n} h_j(\mathrm{T}^k \omega)\, g_{j,k}(y) \\ &= \sum_{j=1}^{\infty} \lim_n \frac{1}{n} \sum_{k=1}^{n} h_j(\mathrm{T}^k \omega)\, g_{j,k}(y)\end{aligned}$$

[in which the limits inside the sum exist a.e. $d\nu(y)$ for each j, by property (2)] is justified by the Dominated Convergence Theorem, since (using Hölder's Inequality)

$$\begin{aligned}&\sum_{j=1}^{\infty} \sup_n \frac{1}{n} \sum_{k=1}^{n} h_j(\mathrm{T}^k \omega)\, g_{j,k}(y) \\ &\leqq \sum_{j<j_0(\omega, y)} \sup_n \frac{1}{n} \sum_{k=1}^{n} h_j(\mathrm{T}^k \omega)\, g_{j,k}(y) \\ &+ \sum_{j \geqq j_0(\omega, y)} \left[\sup_n \frac{1}{n} \sum_{k=1}^{n} h_j^q(\mathrm{T}^k \omega)\right]^{1/q} \left[\sup_n \frac{1}{n} \sum_{k=1}^{n} g_{j,k}^p(y)\right]^{1/p} \\ &\leqq \sum_{j<j_0(\omega, y)} \sup_n \frac{1}{n} \sum_{k=1}^{n} h_j(\mathrm{T}^k \omega)\, g_{j,k} + \sum_{j \geqq j_0(\omega, y)} a_j \, \|h_j\|_q \, \|g\|_p < \infty.\end{aligned}$$

If $p = 1$, then the hypothesis $\sum a_j \, \|\chi_{\{\beta=\beta_j\}}\|_q < \infty$ only allows β to take finitely many values, and the interchange of limits is again permissible.

4. RANDOM APPLICATION OF COMMUTING M.P.T.'S – A GOOD EXAMPLE AND A KEY COUNTEREXAMPLE

In this section we investigate the problem of a.e. convergence of averages when the transformations being applied at each integer time are selected from an arbitrary set of commuting m.p.t.'s, not necessarily from a one-parameter flow. This is already an interesting question in the very simple case when at each time we choose independently one of two commuting m.p.t.'s, such as either a m.p.t. or its inverse; we will show below that sometimes this scheme admits pointwise ergodic theorems, but not always. If the two transformations are chosen according to a general shift-invariant measure, the principles that make possible the convergence of random averages are not at all clear; for two-dimensional averages (summing over rectanges in $\mathbb{Z}^2$) one can prove return-time theorems, but here we are dealing with something different – again some inappropriate averages, this time a sort of one-dimensional sampling, by means of a random walk, of a two-dimensional situation.

It will develop that for *independent* sequences (*i.e.*, if $P = \rho^\infty$ for a probability measure ρ on I) there is a difference between examples of the kinds (2) and (4) (powers of a single transformation) and those of kinds (3) and (5) (general aperiodic higher-dimensional actions). The first kind produces good schemes for a.e. convergence of the $A_n^\omega g$ if and only if the expectation of the powers is not zero. The second kind never produces good schemes; there is always, however, a.e. convergence of a fixed subsequence of the A_n^ω.

Sturmian samples. – We begin by considering a good scheme of type (3), in which two commuting m.p.t.'s chosen in a certain stationary (but highly dependent) way produce a good sampling scheme: we will show that whenever two commuting m.p.t.'s are applied according to the entries in certain Sturmian sequences, we will have a.e. convergence of the resulting ergodic averages for all functions in L^1. Fix an irrational α in $(0, 1)$, let $J = [0, \langle j\,\alpha\rangle]$ for some nonzero integer j (where $\langle x\rangle$ denotes the fractional part of x), and for each $x \in [0, 1]$ define $\phi(x) \in \{0, 1\}^{\mathbb{Z}}$ by

$$\phi(x)(k) = \chi_J(x + k\,\alpha) \qquad \text{and} \qquad \tau_k(x) = \sum_{j=0}^{k-1} \phi(x)(j).$$

When sampling a pair of commuting m.p.t.'s S_0, S_1, at each time k we will apply $S_{\phi(x)(k-1)}$.

THEOREM 3. – *Each $x \in [0, 1]$ generates a good $\{0, 1\}$-sequence, in the following sense: Given any probability space $(Y, \mathcal{C}, \nu)$, commuting* m.p.t.'s S_0 *and* S_1 *on* Y, *and* $g \in L^1(Y)$, *the averages*

$$A_n^x\, g(y) = \frac{1}{n} \sum_{k=1}^{n} g\,(S_0^{k-\tau_k(x)}\, S_1^{\tau_k(x)}\, y) \ \ \textit{converge a.e. } d\nu(y).$$

Proof. – Given any system $(Y, \mathcal{C}, \nu, S_0, S_1)$, by considering the skew-product transformation $(x, y) \to (x + \alpha, S_{\chi_J(x)}\, y)$ of the rotation by α on $[0, 1]$ with (Y, S_0, S_1) we can obtain, for each fixed $g \in L^1(Y)$, convergence of the averages $A_n^{x'} g(y)$ for a.e. (x', y). We will use this to show that in fact for all $x \in [0, 1]$, these averages converge a.e. on Y if $g \in L^\infty(Y)$.

Given $x \in [0, 1]$, notice that if x' is very near x, and say to the right of x, then $x + k\alpha$ and $x' + k\alpha$ are either both in J or both not in J, except when $x + k\alpha$ hits a very short interval to the left of an endpoint of J. However, if the length of J is α, when we create a discrepancy at time k by hitting the first of these intervals, we immediately correct it at time $k + 1$; if the length of J is $\langle j\alpha \rangle$, the resynchronization occurs after finitely many steps. Therefore $|A_n^x g(y) - A_n^{x'} g(y)|$ is bounded by $\|g\|_\infty$ times the frequency of visits of $x + k\alpha$ to a short interval, and so will be very small if x is close to x'. Thus, given $g \in L^\infty(Y)$, choose a countable dense set of x' for each of which there is a full set of $y \in Y$ with convergence of the associated averages; then given $x \in [0, 1]$ and $\varepsilon > 0$, if we choose one of our countably many x' sufficiently close to x, on the intersection of these countably many sets of y we will have

$$\liminf A_n^{x'} g(y) - \varepsilon \leqq \liminf A_n^x g(y) \leqq \limsup A_n^x g(y) \leqq \limsup A_n^{x'} g(y) + \varepsilon.$$

In order to lift convergence from $L^\infty(Y)$ to all of $L^1(Y)$, we will again use a bounded-deviation type of trick to establish a maximal inequality. Since $|\tau_k - k\alpha|$ is bounded (*see* [16]), we may write

$$\tau_k(x) = \lfloor k\alpha \rfloor + m(k, x),$$

where m is bounded (say by M) and integer-valued. Let S_t be the suspension (flow under the constant function 1) of $S_0^{-1} S_1$ on $\hat{Y} = Y \times [0, 1)$ and define $T(y, r) = (S_0 y, r)$ on $\hat{Y}$ and $S = S_1$; then T and S_t commute. Extend functions g on Y to $\hat{Y}$ by $\hat{g}(y, r) = g(y)$ for all $r \in [0, 1)$. Let g be nonnegative and integrable on Y, and define

$$g_M(y) = \sum_{j=-M}^{M} g(S_0^{-1} S_1^j y).$$

Then

$$\begin{aligned}
A_n^x g(Sy) &= \frac{1}{n} \sum_{k=1}^{n} S_0^k (S_0^{-1} S_1)^{\lfloor k\alpha \rfloor} (S_0^{-1} S_1)^{m(k,x)} g(Sy) \\
&\leqq \frac{1}{n} \sum_{k=1}^{n} g_M (S_{\lfloor k\alpha \rfloor} T^k S y) \\
&\leqq \frac{1}{n} \sum_{j=0}^{n-1} (TS_\alpha)^k [\hat{g}_M (y, t) + \hat{g}_M (Sy, t)],
\end{aligned}$$

since $(TS_\alpha)^k \hat{g}_M(y, t) = g_M(S_{\lfloor k\alpha \rfloor} T^k y)$ or $g_M(S_{\lfloor k\alpha \rfloor} T^k S y)$ depending on whether the fractional part of $k\alpha$ is to the left or right of $1 - t$. Then the Maximal Ergodic Theorem applied to the transformation TS_α on $\hat{Y}$ and Fubini's Theorem yield a maximal estimate for the averages $A_n^x g$.

Now we consider the problem of applying two commuting m.p.t.'s according to a Bernoulli (independent identically-distributed) sequence on two symbols. We will see that this sampling scheme is always universally representative for a fixed subsequence of the sequence of averages; but when the transformation are a m.p.t. and its inverse, it is universally representative if and only if the distribution is not symmetric.

First we show that making independent choices of 1 and -1, each with probability $\frac{1}{2}$ does *not* form a universally representative scheme (for applying a m.p.t. and its inverse).

THEOREM 4. – *Consider a sequence $\omega \in \{-1, 1\}^{\mathbb{N}}$ chosen according to the Bernoulli measure $\mathcal{B}\left(\frac{1}{2}, \frac{1}{2}\right)$ – that is, $\{\omega_k\}$ is a sequence of independent, identically-distributed choices of ± 1's, -1 and 1 each being chosen with probability $\frac{1}{2}$. As usual, for each $k = 1, 2, \ldots$ let $S_k(\omega) = \omega_1 + \ldots + \omega_k$ denote the position at time k of the random walk whose increment at time j is ω_j. Then with probability 1, ω is a bad sequence for U and U^{-1}, in the following sense: given any ergodic* m.p.t. U *on a nonatomic probability space $(Y, \mathcal{C}, \nu)$, there is $g \in L^1(Y, \mathcal{C}, \nu)$ such that the averages*

$$A_n^\omega g(y) = \frac{1}{n} \sum_{k=1}^{n} g(U^{S_k(\omega)} y)$$

diverge a.e. *In fact, with probability* 1 *the sequence* ω *will have* strong sweeping out: *given* $\varepsilon > 0$, *we can choose* g *to be the characteristic function of a set of measure less than* ε, *yet have*

$$\limsup_{n\to\infty} \mathrm{A}_n^\omega\, g(y) = 1 \text{ a.e.} \qquad \textit{and} \qquad \liminf_{n\to\infty} \mathrm{A}_n^\omega\, g(y) = 0 \text{ a.e.}$$

Proof. – The idea is to use Strassen's Functional Law of the Iterated Logarithm (*see* [11]) to find long stretches where the averages along the powers $\mathrm{S}_k(\omega)$ involved look like the averages along a sequences that is constant for very long stretches, except for relatively small fluctuations, and then to follow the approach used for example in [13], making slight adjustments so as to avoid any possible ameliorating effects of the fluctuations, to construct a small set that is visited too frequently.

It is enough (*see* [9]) to find, for a.e. ω, for each $\varepsilon > 0$, and $\mathrm{N} \in \mathbb{N}$, a set E of measure less than ε for which $\mu\{\sup_{n \geqq \mathrm{N}} \mathrm{A}_n^\omega\, \chi_{\mathrm{E}} > 1 - \varepsilon\}$. Because of the standard methods for transferring counterexamples by use of Rokhlin towers, we may work with the translation action of $\mathbb{Z}$ on itself and find $g : \mathbb{Z} \to \{0, 1\}$ that takes the value 1 only infrequently yet still achieves large values of the averages for most initial points (*cf.* [13]).

Given $\varepsilon > 0$, choose $q \in \mathbb{N}$ with $4/q < \varepsilon$, and fix a large positive integer n. Take $\alpha, \beta \in (0, 1)$ with $\beta(q+1)^{q-1} < 1$, let $\beta_{-1} = 0$ and for $i = 0, \ldots, q-1$ let $\beta_i = (q+1)^i \beta$. We define a continuous function F on $[0, 1]$ by $\mathrm{F}(0) = 0$, $\mathrm{F} = (i+1)\alpha$ on the interval $[\beta_{i-1} + \beta(q+1)^{i-1}, \beta_i]$ for $0 \leqq i \leqq q-1$, linearly in between, and constant elsewhere. The constant α is chosen small enough so that $\int |\mathrm{F}'|^2 \leqq 1$.

By the theorem of Strassen (*see* [11]), for a.e. ω there are large values of n for which the rescaled graph of the symmetric random walk uniformly approximates the graph of this function F: if $w_n = \sqrt{2n \log\log n}$, then we can find arbitrarily large n with

$$\left| \frac{\mathrm{S}_k(\omega)}{w_n} - \mathrm{F}\left(\frac{k}{n}\right) \right| < \alpha \qquad \text{for} \quad k = 1, 2 \ldots, n.$$

Fix such an ω and n.

For each $i = 1, \ldots, q$, let $\mathrm{I}_i = [(i-1)\alpha w_n, i\alpha w_n)$. We now define $g(x) = 1$ if $x \,(\mathrm{mod}\, q\alpha w_n) \in \mathrm{I}_q \cup \mathrm{I}_1 \cup \mathrm{I}_2$, $g(x) = 0$ otherwise. (We are mainly interested in $x \in \mathbb{Z}$, but it is convenient to work for the moment with $\mathbb{R}$.) Notice that if (a part of) g is transferred by means of the Rokhlin lemma to any aperiodic action of $\mathbb{Z}$, we will have $\mu\{g = 1\} < 4/q < \varepsilon$. Moreover, if $x \in \mathbb{Z}$, choosing i so that $x \,(\mathrm{mod}\, q\,\alpha\, w_n) \in \mathrm{I}_{q-i+1}$ shows that

$$\mathrm{A}^{\omega}_{\beta_{i-1} n}\, g(x) \geqq \frac{1}{n\,\beta_{i-1}} \sum_{n\,\beta_{i-2} \leqq k \leqq n\,\beta_{i-1}} g(x + \mathrm{S}_k(\omega))$$

$$= \frac{1}{n\,\beta_{i-1}} \sum_{n\,\beta_{i-2} \leqq k \leqq n\,\beta_{i-1}} g\left(x + \mathrm{F}\left(\frac{k}{n}\right) w_n + \delta_k\, w_n\right)$$

$$[\text{for some } \delta_k \in (-\alpha,\, \alpha)]$$

(dropping off contributions from the rise of F)

$$\geqq \frac{q-1}{qn\,\beta_{i-1}} \sum_{n\,\beta_{i-2} + n\,\beta\,(q+1)^{i-2} \leqq k \leqq n\,\beta_{i-1}} g\left(x + i\,\alpha\, w_n + \delta_k\, w_n\right)$$

$$\geqq \frac{q-1}{q}\, \frac{n\,\beta_{i-1} - n\,\beta_{i-2} - n\,\beta\,(q+1)^{i-2}}{n\,\beta_{i-1}} \geqq 1 - \frac{3}{q} \geqq 1 - \varepsilon,$$

since $x + i\,\alpha\, w_n \in \mathrm{I}_1$ and $|\delta_k\, w_n| \leqq \alpha\, w_n$.

Remarks.

1. By a similar technique, if powers of a single transformation are applied independently according to an integer-valued (not just ± 1-valued) i.i.d. process with finite second moment and *mean* 0, then we do not always have a.e. convergence. (We will show in the next section that such processes *are* universally representative if they have nonzero mean.)

2. This counterexample implies that sampling a flow by an i.i.d. sequence, for example applying S_1 or S_α with α irrational (members of a measure-preserving flow $\{\mathrm{S}_t\}$) according to whether the entries in a fixed i.i.d. (mean zero or not) Bernoulli sequence ω are 1 or -1, will not always produce a.e. convergence. (Contrast with Example 2 of Section 3, where the sampling process has some memory.) For otherwise we could deduce a maximal inequality in $\mathbb{Z}^2$ for sampling according to $\omega \in \mathbb{R}\,(p,\, 1-p)$; but in the flow $\mathrm{U}_t\, f(\gamma) = f(p + t)$ or $\mathbb{R}$, we can construct a counterexample ω above for U_1 and U_α if the expectation $p + (1-p)\,\alpha = 0$.

3. For the same reason, for an i.i.d. sampling scheme of type (3) (choosing one of a pair of commuting m.p.t.'s according to the entries in a *mean* 0 Bernoulli sequence on $\{-1,\, 1\}$), we cannot have a.e. convergence of the *full* sequence of averages in *any* a periodic $\mathbb{Z}^2$ action ($\mathrm{S}^m\,\mathrm{T}^n \neq \mathrm{S}^k\,\mathrm{T}^l$ unless $(m,\, n) = (k,\, l)$). We will see in Section 7 that we do always have a.e. convergence of a fixed subsequence of the averages.

5. SAMPLING WITH INDEPENDENT INTEGER-VALUED SEQUENCES WITH DRIFT – THE FOURIER ARGUMENT

In this section we show that schemes of type (2) ($I = \mathbb{Z}$), in which powers of a single m.p.t. are applied according to an i.i.d. sequence *with nonzero mean*, are universally representative for L^p for each $p > 1$. The proof uses the Fourier method – estimation of trigonometric sums, with a simplification and improvement by means of an exponential inequality for martingales. It is similar to the method used by Stout (*see* [11]) to prove an exponential inequality and law of the iterated logarithm for martingales. A previous argument along these lines due to Blum and Cogburn [5] yields mean convergence.

THEOREM 5. – *Let* $I = \mathbb{Z}$, $\Omega = I^{\mathbb{N}}$, *and* $P = \rho^{\infty}$, *where* ρ *is a probability measure on* $\mathbb{Z}$ *such that* ω_1 *has finite second moment and nonzero means* ξ *(that is,* $\omega_1, \omega_2, \ldots$ *is a sequence of i.i.d. integer-valued random variables with finite second moment and nonzero mean). Then the scheme* (Ω, P) *is universally representative for* L^p *for each* $p > 1$.

Proof. – We need to show that for every (σ-finite) invertible measure-preserving system $(Y, \mathcal{C}, \nu, U)$ and $g \in L^p$ the averages

$$A_n^{\omega} g(y) = \frac{1}{n} \sum_{k=1}^{n} g(U^{S_k(\omega)} y)$$

converge a.e. We can assume that the greatest common divisor of the essential range of ω_j is 1. We are going to compare A_N to another sequence of (nonrandom) operators V_N for which we already know a.e. convergence. Let $p_k = P\{w_j = k\}$, and let the $L^1 - L^{\infty}$ contraction R be defined by

$$R g(y) = \sum_{k=-\infty}^{\infty} p_k g(U^k y),$$

and let

$$V_N g(y) = \frac{1}{N} \sum_{n=1}^{N} R^n g(y).$$

By Hopf's ergodic theorem for positive contractions, for each $g \in L^p$ and for a.e. x there exists

$$\lim_{N \to \infty} V_N g(x) = g^*(x).$$

The limit function g^* is the projection of g onto the space of R-invariant functions. (In fact, using the assumption that the g.c.d. of the essential range of the ω_j is 1, it is not hard to see that the R-invariant functions coincide with the U-invariant functions.)

For each $r > 1$ let $I_r = \{[r^n] : n = 1, 2, \ldots\}$. The proof of our theorem consists of showing that for a.e. ω (independently of the choice of Y) and for each $r > 1$ we have

$$\| \sum_{N \in I_r} |A_N^\omega g(y) - V_N g(y)| \|_p \leqq C_{\omega, r} \|g\|_p. \tag{1}$$

For (1) implies that for a.e. $\omega \in \Omega$ and each $r > 1$ we have, for a.e. x,

$$\lim_{\substack{N \to \infty \\ N \in I_r}} A_N^\omega g(y) = g^*(y);$$

and then convergence to g^* along the full sequence follows from a well-known argument (*see*, for example, [12]).

We proceed by a sequence of reductions. We can assume that $1 < p < 2$. By the triangle inequality (1) will follow from

$$\sum_{N \in I_r} \|A_N^\omega g(y) - V_N g(y)\|_p \leqq C_{\omega, r} \|g\|_p.$$

But this in turn follows if we show that there is $\theta > 0$ such that

$$\|A_N^\omega g(y) - V_N g(y)\|_p \leqq \frac{C_\omega}{N^\theta} \|g\|_p, \quad N = 1, 2, \ldots \tag{2}$$

Since we can interpolate between L^2 and L^1, (2) follows from an estimate of the form

$$\|A_N^\omega g(y) - V_N g(y)\|_2 \leqq \frac{C_\omega}{N^\sigma} \|g\|_2, \quad N = 1, 2, \ldots \tag{3}$$

with a positive σ. An application of the spectral theorem shows that (3) follows from

$$\sup_{|\alpha| \leqq 1/2} |\hat{A}_N^\omega(\alpha) - \hat{V}_N(\alpha)| \leqq \frac{C_\omega}{N^\sigma}, \quad N = 1, 2, \ldots \tag{4}$$

Using the notation $e(\beta) = e^{2\pi i \beta}$ and letting $\phi(\alpha) = \sum_{k=-\infty}^{\infty} p_k e(k\alpha)$ (the characteristic function of ω_j), the Fourier transforms $\hat{A}_N^\omega$ and $\hat{V}_N$ (into $[-1/2, 1/2)$) are

$$\hat{A}_N^\omega(\alpha) = \frac{1}{N} \sum_{n=1}^{N} e(S_n(\omega)\alpha)$$

and

$$\hat{V}_N(\alpha) = \frac{1}{N}\sum_{n=1}^{N} \hat{R}^n(\alpha) = \frac{1}{N}\sum_{n=1}^{N} (\phi(\alpha))^n.$$

The estimate in (4) will be proved in two parts. First we will work with small α, showing that

$$\sup_{|\alpha| \leqq N^{-3/4}} |\hat{A}_N^{\omega}(\alpha) - \hat{V}_N(\alpha)| \leqq \frac{C_\omega}{N^\sigma}; \tag{5}$$

and then we will obtain bounds for the rest of the α's, namely

$$\sup_{N^{-3/4} \leqq |\alpha| \leqq 1/2} |\hat{V}_N(\alpha)| \leqq \frac{C_\omega}{N^\sigma}, \tag{6}$$

and

$$\sup_{N^{-3/4} \leqq |\alpha| \leqq 1/2} |\hat{A}_N^{\omega}(\alpha)| \leqq \frac{C_\omega}{N^\sigma}. \tag{7}$$

Proof of (5). – First we write

$$\begin{aligned} &\left| \frac{1}{N}\sum_{n=1}^{N} e(S_n(\omega)\alpha) - \frac{1}{N}\sum_{n=1}^{N} (\phi(\alpha))^n \right| \\ &\quad \leqq \frac{1}{N}\sum_{n=1}^{N} |e(S_n(\omega)\alpha) - (\phi(\alpha))^n| \\ &\quad \leqq \frac{1}{N}\sum_{n=1}^{N} |e(S_n(\omega)\alpha) - e(n\,\xi\alpha)| + |e(n\,\xi\alpha) - (\phi(\alpha))^n| = \otimes. \end{aligned} \tag{8}$$

By the (Hartman-Wintner) law of the iterated logarithm there is a set of full measure $\Omega_1 \subseteq \Omega$ so that if $\omega \in \Omega_1$ then $|S_n(\omega) - n \cdot \xi| \leqq C_\omega (n \log\log n)^{1/2}$. It follows, using $|\alpha| \leqq N^{-3/4}$, that

$$\begin{aligned} |e(S_n(\omega)\alpha) - e(n\,\xi\alpha)| &= |e((S_n(\omega) - n\,\xi)\alpha) - 1| \\ &\leqq 2\pi\, C_\omega\, |\alpha|\, (n \log\log n)^{1/2} \\ &\leqq C_\omega (\log\log N)^{1/2} N^{-1/4}. \end{aligned} \tag{9}$$

Next we show that we have the estimate

$$|e(n\,\xi\alpha) - (\phi(\alpha))^n| \leqq CN^{-1/2}. \tag{10}$$

To see this first note that $\phi(\alpha)$ is twice continuously differentiable (since ω_j has finite second moment); hence

$$\phi(\alpha) = 1 + \phi'(0)\,\alpha + O(\alpha^2) = 1 + 2\pi i \xi\alpha + O(\alpha^2). \tag{11}$$

We also have

$$e(\xi\alpha) = 1 + 2\pi i \xi\alpha + O(\alpha^2).$$

We can now estimate that

$$\begin{aligned}|(e(n\xi\alpha)) - (\phi(\alpha))^n| &= |(e(\xi\alpha))^n - (\phi(\alpha))^n| \\ &\leqq n\,|e(\xi\alpha) - \phi(\alpha)| \leqq n\,C\,\alpha^2 \leqq CN^{-1/2},\end{aligned}$$

since $\alpha^2 \leqq N^{-3/2}$. By (9) and (10) we can finish the estimation in (8): $\otimes \leqq C_\omega (\log\log N)^{1/2} N^{-1/4}$.

The above proof did not use the assumption that $\xi \neq 0$. On the other hand this is the only part of the theorem where we used that ω_j had finite second moment. From now on we will use only that ω_j has finite first moment.

Proof of (6). – We continue with the notation of the previous proof. By (11), since $\xi \neq 0$, there are $\delta > 0$ and $c > 0$ so that $|1 - \phi(\alpha)| \geqq c\,|\alpha|$ whenever $|\alpha| < \delta$. On the other hand, on the compact set $\frac{1}{2} \geqq |\alpha| \geqq \delta$ we have $|1 - \phi(\alpha)| \geqq \eta$ for some positive η. This is because the gcd of the essential range of ω_j is 1, and hence for $-1/2 \leqq \alpha \leqq 1/2$ we have $\phi(\alpha) = 1$ if and only if $\alpha = 0$. To sum up: there is $C > 0$ so that $|1 - \phi(\alpha)| \geqq C\,|\alpha|$ whenever $|\alpha| \leqq 1/2$. Now we can estimate that for $|\alpha| \geqq N^{-3/4}$

$$\begin{aligned}|\hat{V}_N(\alpha)| &= \left|\frac{1}{N}\sum_{n=1}^{N} (\phi(\alpha))^n\right| \leqq \frac{1}{N}\left|\frac{1 - (\phi(\alpha))^N}{1 - (\phi(\alpha))}\right| \\ &\leqq \frac{1}{N}\,\frac{2}{|1 - \phi(\alpha)|} \leqq \frac{1}{N}\,\frac{2}{C\,|\alpha|} \leqq CN^{-1/4}.\end{aligned}$$

Proof of (7). – We shall prove that there is a constant K (independent of N) so that for each $0 < |\alpha| \leqq 1/2$

$$P\left\{\omega \left| \left|\frac{1}{N}\sum_{n=1}^{N} e(S_n(\omega)\,\alpha)\right| \geqq K \cdot \sqrt{\frac{\log N}{N\,|1 - \phi(\alpha)|}}\right.\right\} \leqq CN^{-4}. \tag{12}$$

Let us see how (12) implies (7). During the course of the proof of (6) we have shown that

$$|1 - \phi(\alpha)| \geqq C\,|\alpha| \text{ whenever } |\alpha| \leqq 1/2.$$

Therefore for $|\alpha| \geqq N^{-3/4}$ (12) implies

$$P\left\{\omega \left| \left| \frac{1}{N} \sum_{n=1}^{N} e(S_n(\omega)\alpha) \right| \geqq C \cdot \sqrt{\log N} \cdot N^{-1/8} \right.\right\} \leqq CN^{-4}.$$

If follows that if we define $H_N = \{\alpha | 1/2 \geqq |\alpha| \geqq N^{-3/4},\ \alpha = k/N^2$ for some integer $k\}$, then

$$P\left\{\omega \left| \sup_{\alpha \in H_N} \left| \frac{1}{N} \sum_{n=1}^{N} e(S_n(\omega)\alpha) \right| \geqq C \cdot \sqrt{\log N} \cdot N^{-1/8} \right.\right\} \leqq CN^{-2},$$

since $\neq H_N \leqq N^2$. By the Borel-Cantelli lemma, for a.e. ω, for every N

$$\sup_{\alpha \in H_N} \left| \frac{1}{N} \sum_{n=1}^{N} e(S_n(\omega)\alpha) \right| \leqq C_\omega \cdot \sqrt{\log N} \cdot N^{-1/8}.$$

Now to get the estimate in (7) (in which the surpremum is taken over all $1/2 \geqq |\alpha| \geqq N^{-3/4}$) we just use the fact that for a.e. ω for every N we have the following uniform estimate for the derivative of $\hat{A}_N^\omega(\alpha)$:

$$\left| \frac{d\left(\frac{1}{N}\sum_{n=1}^{N} e(S_n(\omega)\alpha)\right)}{d\alpha} \right| = \left| \frac{1}{N} \sum_{n=1}^{N} 2\pi i\, S_n(\omega)\, e(S_n(\omega)\alpha) \right|$$

$$\leqq \frac{1}{N} \sum_{n=1}^{N} 2\pi n \frac{|S_n(\omega)|}{n} \leqq C_\omega N$$

(for a.e. ω, $\sup_n |S_n(\omega)|/n < \infty$).

So let us prove (12). We can assume that

$$\sqrt{\frac{\log N}{N\,|1 - \phi(\alpha)|}} \leqq 1. \tag{13}$$

We need to prove that

$$P\left\{\omega \left| \left| \frac{\sqrt{|1 - \phi(\alpha)| \cdot \log N}}{\sqrt{N}} \sum_{n=1}^{N} e(S_n(\omega)\alpha) \right| \geqq K \log N \right.\right\} \leqq CN^{-4},$$

or, with the notation $\Lambda = \sqrt{\frac{|1-\phi(\alpha)| \cdot \log N}{N}}$ and $F_N = \Lambda \sum_{n=1}^{N} e(S_n \alpha)$,

$$P\{\omega \,|\, |F_N(\omega)| \geqq K \cdot \log N\} \leqq CN^{-4}. \tag{14}$$

We can estimate the left-hand side of (14) as

$$P\{\omega \,|\, |F_N(\omega)| \geqq K \cdot \log N\} \leqq E(e^{|F_N|})\, e^{-K \log N}.$$

We see we just need to have the estimate

$$E(e^{|F_N|}) \leqq e^{C \log N}, \tag{15}$$

because then we just have to choose $K \geqq C + 4$. Let R_N and I_N denote the real and imaginary parts of F_N respectively. We can write, using $ab \leqq 1/2\,(a^2 + b^2)$,

$$E(e^{2 R_N}) \leqq e^{C \log N}; \qquad E(e^{-2 R_N}) \leqq e^{C \log N};$$
$$E(e^{2 I_N}) \leqq e^{C \log N}; \qquad E(e^{-2 I_N}) \leqq e^{C \log N}.$$

We will only show

$$E(e^{2 R_N}) \leqq e^{C \log N}, \tag{16}$$

the proofs of the other bounds being entirely similar. For $m \geqq 1$ let us denote by $\mathcal{B}_m$ the sub σ-algebra of $\mathcal{B}$ generated by $\{\omega_1, \ldots, \omega_m\}$, and let $\mathcal{B}_0 = \{\Omega, \emptyset\}$. We define the (finite) martingale $\{(F_m, \mathcal{B}_m) : m = 0, 1, \ldots, N\}$ and its real part $\{(R_m, \mathcal{B}_m) : m = 0, 1, \ldots, N\}$ by setting

$$F_m(\omega) = E^{\mathcal{B}_m} F_N(\omega) \quad \text{and} \quad R_m(\omega) = E^{\mathcal{B}_m} R_N(\omega), \quad m = 0, \ldots, N,$$

where $E^{\mathcal{B}_m}$ denotes the conditional expectation operator with respect to $\mathcal{B}_m$. We will show the estimates

$$|R_0| \leqq 2, \tag{17}$$

$$E^{\mathcal{B}_{m-1}}((R_m - R_{m-1})^2) \leqq C\,\frac{\log N}{N}, \quad m = 1, 2, \ldots, N, \tag{18}$$

and

$$E^{\mathcal{B}_{m-1}}(\exp(2(R_m - R_{m-1}))) \leqq \exp(C \cdot E^{\mathcal{B}_{m-1}}((R_m - R_{m-1})^2)), \quad m = 1, \ldots, N. \tag{19}$$

Then we would get (16) as follows: by (19) and (18) we have

$$E^{\mathcal{B}_{m-1}}(\exp(2(R_m - R_{m-1}))) \leqq e^{C \log N/N}, \; m = 1, \dots, N,$$

and hence

$$\begin{aligned} E(e^{2R_N}) &= E(E^{\mathcal{B}_{N-1}}(e^{2R_N})) = E(E^{\mathcal{B}_{N-1}}(e^{2R_N - R_{N-1}}) \cdot e^{2R_{N-1}}) \\ &\leqq e^{C \log N/N} \cdot E(e^{2R_{N-1}}) \leqq e^{2C \log N/N} \cdot E(e^{2R_{N-2}}) \\ &\leqq \dots \leqq e^{NC \log N/N} \cdot E(e^{C \log N + 4}), \end{aligned}$$

with the last inequality following from (17).

So let us prove (17)-(19). We need explicit formulas for F_m. By the (complete) independence of the ω_j's,

$$E^{\mathcal{B}_m}(e(S_{m+1}(\omega)\alpha)) = e(S_m(\omega)\alpha) \cdot \phi(\alpha),$$

and therefore

$$F_N = \Lambda \sum_{n=1}^{N} e(S_n \alpha) = \Lambda \cdot \left(\sum_{n=1}^{N-1} e(S_n \alpha) + e(\omega_N \alpha) \cdot e(S_{N-1}\alpha)\right);$$

$$\begin{aligned} F_{N-1} &= \Lambda \cdot \left(\sum_{n=1}^{N-1} e(S_n \alpha) + \phi(\alpha) \cdot e(S_{N-1}\alpha)\right) \\ &= \Lambda \cdot \left(\sum_{n=1}^{N-2} e(S_n \alpha) + \frac{1 - \phi^2(\alpha)}{1 - \phi(\alpha)} e(\omega_{N-1}\alpha) \cdot e(S_{N-2}\alpha)\right); \end{aligned}$$

$$\begin{aligned} F_{N-2} &= \Lambda \cdot \left(\sum_{n=1}^{N-2} e(S_n \alpha) + \frac{1 - \phi^2(\alpha)}{1 - \phi(\alpha)} \phi(\alpha) \cdot e(S_{N-2}\alpha)\right) \\ &= \Lambda \cdot \left(\sum_{n=1}^{N-3} e(S_n \alpha) + \frac{1 - \phi^3(\alpha)}{1 - \phi(\alpha)} e(\omega_{N-2}\alpha) \cdot e(S_{N-3}\alpha)\right); \end{aligned}$$

$$\vdots$$

$$\begin{aligned} F_m &= \Lambda \cdot \left(\sum_{n=1}^{m} e(S_n \alpha) + \frac{1 - \phi^{N-m}(\alpha)}{1 - \phi(\alpha)} \phi(\alpha) \cdot e(S_m \alpha)\right) \\ &= \Lambda \cdot \left(\sum_{n=1}^{m-1} e(S_n \alpha) + \frac{1 - \phi^{N-m+1}(\alpha)}{1 - \phi(\alpha)} \cdot e(\omega_m \alpha) \cdot e(S_{m+1}\alpha)\right); \end{aligned}$$

$$\vdots$$

$$\mathrm{F}_1 = \Lambda \cdot \frac{1 - \phi^{\mathrm{N}}(\alpha)}{1 - \phi(\alpha)}\, e(\omega_1 \alpha);$$

$$\mathrm{F}_0 = \Lambda \cdot \frac{1 - \phi^{\mathrm{N}}(\alpha)}{1 - \phi(\alpha)}\, \phi(\alpha).$$

Now (17) follows immediately from (13) and the above expression for F_0. Using independence of ω_m and $\mathcal{B}^{m-1}$, we also have the following estimate for $m = 1, 2, \ldots, \mathrm{N}$:

$$\begin{aligned}
&\mathrm{E}^{\mathcal{B}_{m-1}}(|\mathrm{F}_m - \mathrm{F}_{m-1}|^2) \\
&= \frac{|1 - \phi(\alpha)| \cdot \log \mathrm{N}}{\mathrm{N}} \cdot \left| \frac{1 - \phi^{\mathrm{N}-m+1}(\alpha)}{1 - \phi(\alpha)} \right|^2 \cdot \mathrm{E}^{\mathcal{B}_{m-1}}(|e(\omega_m \alpha) - \phi(\alpha)|^2) \\
&\leqq \frac{|1 - \phi(\alpha)| \cdot \log \mathrm{N}}{\mathrm{N}} \, \frac{4}{|1 - \phi(\alpha)|}\,(1 - |\phi(\alpha)|^2) \leqq \mathrm{C}\, \frac{\log \mathrm{N}}{\mathrm{N}},
\end{aligned}$$

since $\dfrac{1 - |\phi(\alpha)|^2}{|1 - \phi(\alpha)|}$ is bounded uniformly in α (not an integer). This implies (18).

Finally we prove (19). Let $d_m = 2(\mathrm{R}_m - \mathrm{R}_{m-1})$. Then

$$|d_m| \leqq 2\,|\mathrm{F}_m - \mathrm{F}_{m-1}| \leqq 8.$$

This is because by (13)

$$\begin{aligned}
&|\mathrm{F}_m - \mathrm{F}_{m-1}| \\
&= \sqrt{\frac{|1 - \phi(\alpha)| \cdot \log \mathrm{N}}{\mathrm{N}}} \cdot \left| \frac{1 - \phi^{\mathrm{N}_{m+1}}(\alpha)}{1 - \phi(\alpha)} \right| \cdot |e(\omega_m \alpha) - \phi(\alpha)| \leqq 4.
\end{aligned}$$

Now by (20)

$$\begin{aligned}
\mathrm{E}^{\mathcal{B}_{m-1}}(e^{d_m}) &= \mathrm{E}^{\mathcal{B}_{m-1}}\left[1 + d_m + \sum_{j=2}^{\infty} \frac{d_m^j}{j!}\right] \\
&= \mathrm{E}^{\mathcal{B}_{m-1}}\left[1 + d_m + d_m^2 \sum_{j=2}^{\infty} \frac{d_m^{j-2}}{j!}\right] \\
&\leqq \mathrm{E}^{\mathcal{B}_{m-1}}\left(1 + d_m + d_m^2 \cdot \sum_{j=0}^{\infty} \frac{8^j}{j!}\right) \\
&= 1 + 0 + \mathrm{E}^{\mathcal{B}_{m-1}}(d_m^2) \cdot e^8 \leqq \exp(\mathrm{C}\mathrm{E}^{\mathcal{B}_{m-1}}(d_m^2)),
\end{aligned}$$

which finishes the proof.

Remarks. – 1. Combining this theorem with the counterexample that precedes it, we see that a square-integrable integer-valued i.i.d. process provides a universally representative sampling scheme *if and only if* it has nonzero mean.

2. The equivalent of Theorem 5 holds if the transformation U being sampled is a positive contraction of L^p, not necessarily a m.p.t. [12].

6. SAMPLING WITH ANY INTEGER-VALUED STATIONARY SEQUENCE WITH DRIFT – THE TOWER ARGUMENT

In this section we use a more ergodic-theoretic viewpoint (towers, orbit equivalence) to show that any (ergodic) integrable integer-valued stochastic process *with nonzero mean* is universally representative for *bounded* stationary processes. For the positive integer-valued case, the reduction of universal representation to the return-times theorem (by means of towers) was made in Section 1; with some more effort, we can accomplish the same thing in the positive expectation case.

THEOREM 6. – *Let* $I = \mathbb{Z}$, $\Omega = I^{\mathbb{Z}}$, *and let* P *be an ergodic shift-invariant probability measure on* Ω. *Suppose that* $f = \omega_1$ *is integrable and has nonzero mean* ξ. *Then the scheme* (Ω, P) *is universally representative for* L^∞.

Proof. – Assume that $\xi > 0$. We use the notations

$$S_k(\omega) = \omega_1 + \ldots + \omega_k = \sum_{j=0}^{k-1} f(\sigma^j \omega) \qquad \text{for} \quad k \geqq 1,$$

$$S_0(\omega) = 0, \quad \text{and}$$

$$S_k(\omega) = -(\omega_{k+1} + \ldots + \omega_0) = \sum_{j=k+1}^{0} f(\sigma^j \omega) \qquad \text{for} \quad k \leqq -1.$$

Also, for U a m.p.t. on a probability space $(Y, \mathcal{C}, \nu)$ and g a bounded measurable function on Y, continue to denote

$$A_n^\omega g(y) = \frac{1}{n} \sum_{k=1}^{n} g(U^{S_k(\omega)} y).$$

Define

$$M(\omega, n) = \max\{S_j(\omega) : 1 \leqq j \leqq n\},$$

$$m(\omega, n) = \min\{S_j(\omega) : 1 \leqq j \leqq n\},$$

$$N_n^+(\omega, k) = \operatorname{card}\{j : 1 \leqq j \leqq n,\ S_j(\omega) = k\} \quad (\text{for each } k \in \mathbb{Z}),$$

$$N^-(\omega, k) = \operatorname{card}\{j \leqq 0 : S_j(\omega) = k\} \quad (k \in \mathbb{Z}),$$

$$N_n(\omega, k) = \operatorname{card}\{j : j \leqq n,\ S_j(\omega) = k\},$$

and

$$N(\omega, k) = \operatorname{card}\{j \in \mathbb{Z} : S_j(\omega) = k\} \quad (k \in \mathbb{Z}).$$

We abbreviate $Z(\omega) = N(\omega, 0) = 1$. Then

$$A_n^\omega g(y) = \frac{M(\omega, n)}{n} \frac{1}{M(\omega, n)} \sum_{k=m(\omega, n)}^{M(\omega, n)} g(U^k y) N_n^+(\omega, k).$$

Assume that $g \geqq 0$. We will show that for a.e. $\omega \in \Omega$ (with the set of measure 0 not depending on $(Y, \mathcal{C}, \nu, U, g)$),

(1) $\dfrac{M(\omega, n)}{n} \to \xi$, and,

(2) $B_M = \dfrac{1}{M} \displaystyle\sum_{k=1}^{M} g(U^k y) N(\omega, k)$ converges for a.e. $y \in Y$ as $M \to \infty$.

Now

(3) $M(\omega, n) \to \infty$ as $n \to \infty$;

(4) for each ω, $m(\omega, n)$ is bounded as r varies;

(5) $N_n(\omega, k) = N_n^+(\omega, k) + N^-(\omega, k)$ and $N^-(\omega, k) = 0$ for $k \geqq k_0(\omega)$;

(6) By the Ergodic Theorem, given $\epsilon > 0$ for large n we have $S_n \in [(\xi - \epsilon)n, (\xi + \epsilon)n]$. Thus for large n (and $M = M(\omega, n)$), the difference between B_M and

$$B_M^n = \frac{1}{M} \sum_{k=1}^{M} g(U^k y) N_n^+(\omega, k)$$

is bounded by

$$\frac{2}{M} \sum_{k=(\xi-\epsilon)n}^{(\xi+\epsilon)n \wedge M} g(U^k y) N(\omega, k);$$

using (2), this can be seen to be on the order of ϵ.

In view of (3)-(6), in order to prove the theorem it is sufficient to establish (1) and (2).

To prove (1), take the sequence $n_1 < n_2 < \ldots$ of record times, i.e. such that $M(\omega, n_j) = S_{n_j}(\omega)$ for each j, and $S_n(\omega) \leqq M(\omega, n) = M(\omega, n_j)$ for $n_j \leqq n \leqq n_{j+1}$. Given $\epsilon > 0$, choose J large enough that for $n \geqq n_J$, $S_n(\omega)/n$ is within ϵ of ξ. Then for $n \geqq n_J$, say with $n_j \leqq n < n_{j+1}$, we have

$$\frac{S_n(\omega)}{n} \leqq \frac{M(\omega, n)}{n} = \frac{M(\omega, n_j)}{n} \leqq \frac{M(\omega, n_j)}{n_j} = \frac{S_{n_j}}{n_j},$$

all of which are within ϵ of ξ.

To prove (2), we will express the averages involved as those found in an application of the return-times theorem to another transformation related to the shift σ on Ω. The tower will have for its base

$$A = \{\omega \in \Omega : \text{ if } k < 0 \text{ then } S_k(\omega) \neq 0\}$$

with a column of height

$$h(\omega) = \inf\{S_k(\omega) : k \in \mathbb{Z} \text{ and } S_k(\omega) > 0\}$$

above each point $\omega \in A$. Notice that A has positive measure (for example, it contains the set A, for the transformation σ^{-1}, as defined in [15], p. 83).

Our set A is a section of the following equivalence relation: $\omega \sim \omega'$ if there is $j \in \mathbb{Z}$ with $\omega' = \sigma^j \omega$ and $S_j(\omega) = 0$. The cardinal of the equivalence class of ω is $Z(\omega)$; define $Z(\omega) = 0$ for $\omega \notin A$. If we let

$$k(\omega) = \inf\{j \in \mathbb{Z} : S_j(\omega) = h(\omega)\},$$

then

$$\psi(\omega) = \sigma^{k(\omega)} \omega$$

is a measure-preserving transformation $A \to A$. We define the measure-preserving system (X, T) to be the tower over (A, ψ) with height function h. Our result will be obtained by applying the return-times theorem to this system (X, T), the function Z, and the points of A.

Fix $\omega \in \Omega$ and spread out the σ-orbit of ω in $\mathbb{Z} \times \mathbb{N}$ as follows. Over each integer i in $\{S_k(\omega) : k \in \mathbb{Z}\}$ (the range of the cocyle) place a column (of height $N(\omega, i)$) of the points $\sigma^j \omega$ in the orbit of ω arrived at when this value is assumed (i.e., include $\sigma^{k-1} \omega$ in the stack over i if $S_k(\omega) = i$), ordered from bottom to top according to the order that the orbit of ω inherits from $\mathbb{Z}$. (This picture may usefully be arrived at by plotting the random walk $\{S_k(\omega) : k \in \mathbb{Z}\}$ as a "snake graph" over $\mathbb{Z}$). There are a few neat things to notice about this array of points:

(7) Any sum between a pair of elements in a stack is 0: If $j < k$ and $S_j(\omega) = S_k(\omega) = i$, then $S_{k-j}(\sigma^j \omega) = 0$.

(8) The bottom element of each stack is in A.

(9) If ω is replaced by another point in its orbit, we obtain essentially the same picture, just translated horizontally.

(10) The horizontal distance from the stack over $\sigma^j \omega \in A$ to the next one to the right of it is $h(\sigma^j \omega)$.

Further, looking at this array of points also allows one to see that *h is integrable*. Consider the region under the graph of $|f|$ on Ω to be made up of segments $D(x)$ of length $|f(x)|$ over the point $x \in \Omega$. We will cut up and rearrange this region in a measure-preserving way. Fix $\omega \in \Omega$ and spread out the orbit of ω in the array described above. For x in the orbit of ω, lay out a horizontal segment (of length $|f(\sigma x)|$) from the stack including x to the one including σx. This segment crosses a certain set of points $z \in A$, where "crosses" means that z lies under the left closed, right open segment. For each such z crossed by a segment, we cut off that part of the segment, of length $h(z)$, that extends from the stack over z to the next stack to its right, and we assign this piece of the segment to $z \in A$. Looking at the entire graph of the random walk, we see that in this way each $z \in A$ gets assigned to it the number of segments that cross its stack, which is

$$n(z) = \text{card}\{j \in \mathbb{Z} : \text{ one of } S_j(z),\ S_{j+1}(z) \text{ is nonpositive and the other is positive}\}.$$

Since the cutting is done measurably and the rearranging (the assigning of pieces of $D(x)$, for each $x \in \Omega$, to certain $z \in A$) is accomplished by powers of σ in the first coordinate and translation in the second, this process is a measure-preserving transformation of the region in $\Omega \times \mathbb{Z}$ under the graph of $|f|$ to the region in $A \times \mathbb{Z}$ under the graph of $n(z)\, h(z)$. Thus

$$\int_\Omega |f|\, dP = \int_A n(z)\, h(z)\, dP,$$

and in particular, since $n(z) \geqq 1$ a.e., the integrability of f implies that of h.

Now we define a map $\tau : \Omega \to \Omega$ as follows. Given $\omega \in \Omega$, form the array corresponding to it as above. If ω is not the top element in a stack, let $\tau\omega$ be the point immediately above it; otherwise, let $\tau\omega$ be the point on the bottom of the next stack to the right. Then τ is a well-defined, measurable, one-to-one, onto map of the form $\tau(\omega) = \sigma^{j(\omega)}(\omega)$ which has the same orbits as does σ; therefore τ is an ergodic m.p.t. on Ω.

Let $\psi : \mathrm{A} \to \mathrm{A}$ be the first-return map (under τ) to A. Thus for each $z \in \mathrm{A}$, ψz is the element on the bottom of the stack to the right of z. Further, for any $x \in \Omega$, let πx be the bottom element of the stack to which x belongs (this is independent of the starting point in the σ-orbit of x used to form the array). Finally, as mentioned above, (X, T) is the measure-preserving system obtained by building a tower with height function h over (A, ψ), and we are interested in the function $\mathrm{Z}(\omega) = \mathrm{card}\{j \in \mathbb{Z} : \mathrm{S}_j(\omega) = 0\}$ on X. Then

$$\frac{1}{\mathrm{M}} \sum_{k=1}^{\mathrm{M}} g(\mathrm{U}^k y)\,\mathrm{N}(\omega, k) = \frac{1}{\mathrm{M}} \sum_{k=1}^{\mathrm{M}} g(\mathrm{U}^k y)\,\mathrm{Z}(\mathrm{T}^k \pi\omega).$$

By the return-times theorem, for almost all $\pi\omega$, and hence for almost all ω, for any measure-preserving system $(\mathrm{Y}, \mathcal{C}, \nu, \mathrm{U})$ and any $g \in \mathrm{L}^\infty(\mathrm{Y})$ these averages converge a.e. $d\nu(y)$.

Remark. – If $\mathrm{Z} \in \mathrm{L}^p(\mathrm{A})$ for some $1 \leqq p \leqq \infty$, and $\frac{1}{p} + \frac{1}{q} = 1$, then for a.a. ω we will obtain a.e. convergence of $\mathrm{A}_n^\omega g(y)$ for all systems $(\mathrm{Y}, \mathcal{C}, \nu, \mathrm{U})$ and all $g \in \mathrm{L}^q(\mathrm{Y})$. This follows from the above argument and the L^p, L^q version of the return-times theorem.

7. INDEPENDENT APPLICATIONS OF ELEMENTS OF A HIGHER-DIMENSIONAL ACTION – CONVERGENCE OF A SUBSEQUENCE

In this section we consider sampling schemes of type (6), applications of elements of a higher-dimensional measure-preserving action selected at random independently (for example, at each time we apply one of a finite set of commuting m.p.t.'s): d is a positive integer, ρ is a probability measure on $\mathbb{Z}^d$ with finite second moment (i.e. $\sum |k|^2 \rho(k) < \infty$), and $(\Omega, \mathrm{P}) = (\mathbb{Z}^d, \rho)^{\mathbb{N}}$, the product measure space of infinitely many copies of $(\mathbb{Z}^d, \rho)$. We continue to let $\mathrm{S}_n(\omega) = \omega_1 + \ldots + \omega_n$, where ω_j is the projection of ω onto the j'th coordinate space of $\Omega = (\mathbb{Z}^d)^{\mathbb{N}}$. The p-th coordinate of a vector $v \in \mathbb{R}^d$ is denoted by $v^{(p)}$. We use the absolute value sign two (not really different) ways: $|x|$ means the absolute value of the real number x, but for a vector $x \in \mathbb{R}^d$ we understand $|x| = \max_{1 \leqq p \leqq d} |x^{(p)}|$.

The Fourier-transform variables $\alpha, \beta \in \mathbb{R}^d$ always satisfy $|\alpha|, |\beta| \leqq 1/2$. The constants mentioned below can depend on the dimension d.

THEOREM 7. – *For* $t = 1, 2, \ldots$ *let* $N_t = [2^{t \log t}]$. *Then* a.e. $\omega \in \Omega$ *has the following property: for every measure-preserving action* U *of* $\mathbb{Z}^d$ *on a* σ*-finite measure space* $(Y, \mathcal{C}, \nu)$ *and every* $g \in L^2(Y)$, *the limit*

$$\lim_{t \to \infty} \frac{1}{N_t} \sum_{n=1}^{N_t} g(U^{S_n(\omega)} y)$$

exists for a.e. $y \in Y$.

Remarks.

(1) We will see from the proof that if the support of ρ generates (with respect to addition) the whole of $\mathbb{Z}^d$, then the above limit is ν-a.e. equal to the projection of g on the space of U-invariant functions.

(2) Probably the methods of [2] can be used to prove a.e. convergence for $g \in L^p(Y)$, $p > 1$, but we do not know what happens for $g \in L^1(Y)$.

(3) U can be replaced by a $\mathbb{Z}^d$-action of positive L^2-contractions. (The spectral theorem is valid for L^2-contractions, not just for isometries.)

(4) There is an L^2-dense class of g's for which the averages converge along the full sequence $N = 1, 2, \ldots$, and not just along the subsequence N_t. (Take the ranges of the images of the spectral measures of the d-torus with a neighborhood of the origin removed. The Fourier transforms of the kernels involved tend to 0 uniformly on each such set. If the support of ρ fails to generate, we may need to remove a finite number of d-dimensional cubes.)

(5) If we had convergence along the sequence $N_t = 2^t$ then we would also have convergence for the full sequence. (This is because, for positive operators, existence of a maximal inequality along such a subsequence implies the maximal inequality along the full sequence, and we have convergence on a dense set.)

(6) It is easy to modify our proof below to get a.e. convergence of the averages along the subsequence $N_t = t!$ or $N_t = [2^{ct \log t}]$, where c is any fixed positive constant. The secret is that $\{N_t\}$ should satisfy: $N_t > 1$ and N_{t+1}/N_t should be greater than a fixed positive power of $\log(N_t)$. Any such sequence has the following three key properties:

(i) There is a positive integer s so that for any given nonnegative t_0, if $t \geqq t_0 + s$ then

$$\frac{\log N_t}{N_t} \leqq \frac{1}{N_{t_0+1} \log \log N_{t_0+1}}.$$

(ii) For each t_0,

$$\sum_{t<t_0} \mathrm{N}_t < \mathrm{N}_{t_0}.$$

(iii) If $\mathrm{M}_t = \log \mathrm{N}_t/\mathrm{N}_t$, then

$$\sum_{t \geqq t_0} \mathrm{M}_t \leqq \mathrm{CM}_{t_0}.$$

(7) We have *mean convergence* along the full sequence $\mathrm{N} = 1, 2 \ldots$ (since we have a.e. convergence for a dense set).

Proof of Theorem 7. – The ideas are similar to the proof of Theorem 5. We can assume that the support of ρ generates $\mathbb{Z}^d$. (Otherwise consider a subaction of $\mathbb{Z}^d$ which is necessarily isomorphic to a $\mathbb{Z}^{d'}$-action for some $d' \leqq d$.) For $\omega \in \Omega$, continue to denote

$$\mathrm{A}_{\mathrm{N}}^{\omega} g(y) = \frac{1}{\mathrm{N}} \sum_{n=1}^{\mathrm{N}} g(\mathrm{U}^{\mathrm{S}_n(\omega)} y).$$

We are going to compare $\mathrm{A}_{\mathrm{N}}^{\omega}$ to another sequence of operators V_{N}. First we define

$$\mathrm{R} g(y) = \sum_{k \in \mathbb{Z}^d} \rho(k) g(\mathrm{U}^k y),$$

and then we let

$$\mathrm{V}_{\mathrm{N}} g(y) = \frac{1}{\mathrm{N}} \sum_{n=1}^{\mathrm{N}} \mathrm{R}^n g(y).$$

By the Hopf ergodic theorem, for each $g \in \mathrm{L}^2$ and a.e. y we have

$$\lim_{\mathrm{N} \to \infty} \mathrm{V}_{\mathrm{N}} g(y) = g^*(y),$$

where g^* is the projection of g on the space of R-invariant functions. In fact, it is not hard to see that the R-invariant functions coincide with the U-invariant functions. We want to show that there is $\Omega' \subseteq \Omega$ with $\mathrm{P}(\Omega') = 1$ such that if $\omega \in \Omega'$ then

$$\int_{\mathrm{Y}} \left(\sum_{t=1}^{\infty} |\mathrm{A}_{\mathrm{N}_t}^{\omega} g(y) - \mathrm{V}_{\mathrm{N}} g(y)|^2 \right) dy \leqq \mathrm{C}_{\omega} \|g\|_{\mathrm{L}^2(\mathrm{Y})}^2. \tag{8}$$

An application of the spectral theorem shows that (8) follows from

$$\sup_{\alpha} \sum_{t=1}^{\infty} |\hat{\mathrm{A}}_{\mathrm{N}_t}^{\omega}(\alpha) - \hat{\mathrm{V}}_{\mathrm{N}_t}(\alpha)|^2 \leqq \mathrm{C}_{\omega}, \tag{9}$$

where the Fourier transforms $\hat{A}_N^{\omega}(\alpha)$ and $\hat{V}_N(\alpha)$ are

$$\hat{A}_N^{\omega}(\alpha) = \frac{1}{N} \sum_{n=1}^{N} e(S_n(\omega) \cdot \alpha)$$

and

$$\hat{V}_N(\alpha) = \frac{1}{N} \sum_{n=1}^{N} (\phi(\alpha))^n,$$

and $\phi(\alpha) = \sum_{k \in \mathbb{Z}^d} \rho(k)\, e(k \cdot \alpha)$.

The inequality in (9) will be a consequence of the following estimates (compare with inequalities 5-7 in the proof of Theorem 5):

$$|\hat{A}_N^{\omega}(\alpha) - \hat{V}_N(\alpha)| \leqq \min\{2,\, C_{\omega}\, |\alpha| \sqrt{N \log\log N}\}; \tag{10}$$

$$|\hat{V}_N(\alpha)| \leqq \min\left\{1,\, \frac{C}{N\,|\alpha|^2}\right\}; \tag{11}$$

and

$$|\hat{A}_N^{\omega}(\alpha)| \leqq \min\left\{1,\, C_{\omega} \sqrt{\frac{\log N}{N\,|\alpha|^2}}\right\}. \tag{12}$$

Let us deduce a single inequality that is a consequence of (10)-(12) (using the fact that the square root of a number less than one is greater than the number):

$$|\hat{A}_N^{\omega}(\alpha) - \hat{V}_N(\alpha)| \leqq \min\left\{2,\, C_{\omega}\, |\alpha| \sqrt{N \log\log N},\, C_{\omega} \frac{\sqrt{\frac{\log N}{N}}}{|\alpha|}\right\} \tag{13}$$

Note that for fixed N the above estimate gives only the trivial bound 2 if

$$\frac{2}{C_{\omega}} \frac{1}{\sqrt{N \log\log N}} \leqq |\alpha| \leqq \frac{C_{\omega}}{2} \sqrt{\frac{\log N}{N}}.$$

This is the main cause of difficulty; however, along the highly lacunary subsequence $\{N_t\}$, each α is in this undesirable region for only a finite number of steps, uniformly bounded in α. Before we prove the estimates (10)-(12), let us see how they–that is, (13)–imply (9). We will use the properties mentioned in Remark (6).

Fix $\alpha \neq 0$. Let t_0 be the largest positive integer t satisfying $|\alpha| \leqq \frac{1}{\sqrt{N_t \log \log N_t}}$. (If there is no t like this then we take $t_0 = 0$.) Note that

$$\frac{1}{\sqrt{N_{t_0+1} \log \log N_{t_0+1}}} < |\alpha|.$$

Using the s from Remark 6 (i), we now split the sum on the left of (9) into three subsums:

$$\sum_{t=1}^{\infty} |\hat{A}^{\omega}_{N_t}(\alpha) - \hat{V}_{N_t}(\alpha)|^2$$
$$= \sum_{1 \leqq t \leqq t_0} |\cdot|^2 + \sum_{t_0 < t < t_0+s} |\cdot|^2 + \sum_{t_0+s \leqq t} |\cdot|^2.$$

Using (13) and the definition of N_t we can estimate the first sum as

$$\sum_{1 \leqq t \leqq t_0} |\cdot|^2 \leqq \sum_{1 \leqq t \leqq t_0} (C_\omega |\alpha| \sqrt{N_t \log \log N_t})^2$$
$$\leqq \sum_{1 \leqq t \leqq t_0} \left(\frac{C_\omega \sqrt{N_t \log \log N_t}}{\sqrt{N_{t_0+1} \log \log N_{t_0+1}}} \right)^2 \leqq C_\omega;$$

then the second as

$$\sum_{t_0 < t < t_0+s} |\cdot|^2 \leqq \sum_{t_0 < t < t_0+s} 2^2 \leqq 4\, s = C;$$

and finally the third as

$$\sum_{t_0+s \leqq t} |\cdot|^2 \leqq \sum_{t_0+s \leqq t} \left(C_\omega \frac{\sqrt{\frac{\log N_t}{N_t}}}{|\alpha|} \right)^2$$

$$= \sum_{t_0+s \leqq t} \left(\frac{C_\omega \sqrt{\frac{\log N_t}{N_t}}}{\sqrt{\frac{\log N_{t_0+s}}{N_{t_0+s}}}} \right)^2 \leqq C_\omega,$$

which implies (9).

Proof of (10). – Let us put $\xi = \sum_{k \in \mathbb{Z}^d} k\,\rho(k)$, the mean element of the random action. First we write

$$\left| \frac{1}{N} \sum_{n=1}^{N} e(S_n(\omega) \cdot \alpha) - \frac{1}{N} \sum_{n=1}^{N} (\phi(\alpha))^n \right|$$
$$\leqq \frac{1}{N} \sum_{n=1}^{N} |e(S_n(\omega) \cdot \alpha) - (\phi(\alpha))^n|$$
$$\leqq \frac{1}{N} \sum_{n=1}^{N} \{|e(S_n(\omega) \cdot \alpha) - e(n\,\xi \cdot \alpha)|$$
$$+ |e(n\,\xi \cdot \alpha) - (\phi(\alpha))^n|\} = \otimes.$$

By the law of the iterated logarithm there is a set $\Omega_1 \subseteq \Omega$ of full measure such that if $\omega \in \Omega_1$ then $|S_n(\omega) - n\,\xi| \leqq C_\omega \sqrt{n \log\log n}$ for all n. (The d-dimensional estimate follows by using the one-dimensional one on each coordinate.) As in the proof of (5.5) we get

$$|e(S_n(\omega) \cdot \alpha) - e(n\,\xi \cdot \alpha)| \leqq C_\omega |\alpha| \sqrt{n \log \log n}. \tag{14}$$

Next we show that

$$|e(n\,\xi \cdot \alpha) - (\phi(\alpha))^n| \leqq C\,|\alpha|^2\,n. \tag{15}$$

To see this first note that (since ρ has finite second moment)

$$\phi(\alpha) = 1 + 2\pi i\,\xi \cdot \alpha + O(|\alpha|^2).$$

We also have

$$e(\xi \cdot \alpha) = 1 + 2\pi i\,\xi \cdot \alpha + O(|\alpha|^2).$$

As in the proof of (5.5) the above two estimates imply (15). By (14) and (15) we can finish our proof:

$$\otimes \leqq C_\omega |\alpha| \sqrt{N \log \log N} + C\,|\alpha|^2\,N$$
$$\leqq C_\omega |\alpha| \sqrt{N \log \log N}$$

(using $C_\omega |\alpha| \sqrt{N \log \log N} < 2$ and that the square root of a number less then one is greater than the number).

Proof of (11). – We can assume that $\alpha \neq 0$. It follows that $\phi(\alpha) \neq 1$, since the support of ϕ generates $\mathbb{Z}^d$. Summing the geometric progression, we get

$$|\hat{\mathrm{V}}_{\mathrm{N}}(\alpha)| \leqq \frac{2}{\mathrm{N}\,|1-\phi(\alpha)|}.$$

We see that we just have to prove that

$$|1-\phi(\alpha)| \geqq \mathrm{C}\,|\alpha|^2, \tag{16}$$

with C independent of α and N. (Recall the two meanings of $|\cdot|$.) As in the proof of (5.6) it is enough to prove (16) in a neighborhood of 0. First we write

$$\begin{aligned}|1-\phi(\alpha)| \geqq |1-\mathcal{R}(\phi(\alpha))| &= |1-\sum_{k\in\mathbb{Z}^d}\phi(k)\cos(2\pi\, k\cdot\alpha)|\\ &= 2\pi^2\sum_{k\in\mathbb{Z}^d}\phi(k)(k\cdot\alpha)^2 + o(|\alpha|^2),\end{aligned}$$

since ρ has finite second moment.

We see we just need to show that

$$\sum_{k\in\mathbb{Z}^d}\phi(k)(k\cdot\alpha)^2 \geqq \mathrm{C}\,|\alpha|^2. \tag{17}$$

Let us define the quadratic form Q on $\mathbb{R}^d$ by

$$\mathrm{Q}(x, x) = \sum_{k\in\mathbb{Z}^d}\phi(k)(k\cdot x)^2.$$

Since the support of ϕ generates $\mathbb{Z}^d$ it follows that Q is positive definite, and (17) is just a norm equivalence result on $\mathbb{R}$. [By classical results, there is a basis $\{v_1, \ldots, v_d\}$ of $\mathbb{R}^d$, and positive numbers $\lambda_1, \ldots, \lambda_d$, so that if $x = \sum_{q=1}^{d} x_{(q)}\, v_q$ then

$$\mathrm{Q}(x, x) = \sum_{q=1}^{d} x_{(q)}^2\,\lambda_q.$$

Therefore

$$\mathrm{Q}(x, x) \geqq \mathrm{C}\max_{1\leqq q\leqq d} x_{(q)}^2.$$

To prove (17) we just need to prove that $|\alpha| \leqq \mathrm{C}\max_{1\leqq q\leqq d}|\alpha_{(q)}|$.

Let $\{w_1, \ldots, w_d\}$ be the orthonormal basis for which we defined the coordinates $\alpha^{(p)}$. We have

$$\sum_{p=1}^{d} \alpha^{(p)} w_p = \sum_{q=1}^{d} \alpha_{(q)} v_q.$$

It follows that

$$\begin{aligned} |\alpha^{(p)}| &= \left| \sum_{q=1}^{d} \alpha_{(q)} (v_q \cdot w_p) \right| \leqq \mathrm{C}_p \max_{1 \leqq q \leqq d} |\alpha_{(q)}| \\ &\leqq \mathrm{C} \max_{1 \leqq q \leqq d} |\alpha_{(q)}| \end{aligned}$$

and we are done.]

Proof of (12). – An argument identical with the proof of (5.7) gives a constant K so that for every α and every N

$$\mathrm{P}\left\{\omega \,|\, |\hat{\mathrm{A}}_{\mathrm{N}}^{\omega}(\alpha)| \geqq \mathrm{K}\sqrt{\frac{\log \mathrm{N}}{\mathrm{N}\,|1-\phi(\alpha)|}}\right\} \leqq \mathrm{CN}^{-2-2d}$$

(just choose K large enough, d being fixed). By the estimate in (16) this implies that (with a different K)

$$\mathrm{P}\left\{\omega \,|\, |\hat{\mathrm{A}}_{\mathrm{N}}^{\omega}(\alpha)| \geqq \mathrm{K}\sqrt{\frac{\log \mathrm{N}}{\mathrm{N}\,|\alpha|^2}}\right\} \leqq \mathrm{CN}^{-2-2d}. \tag{18}$$

Let $\mathrm{H}_{\mathrm{N}} = \{\alpha | \alpha^{(p)} = k^{(p)}/\mathrm{N}^2,\ p = 1, \ldots, d, \text{ for some } k \in \mathbb{Z}^d\}$. By (18) we have

$$\mathrm{P}\left(\bigcup_{\alpha \in \mathrm{H}_{\mathrm{N}}} \left\{\omega \,|\, |\hat{\mathrm{A}}_{\mathrm{N}}^{\omega}(\alpha)| \geqq \mathrm{K}\sqrt{\frac{\log \mathrm{N}}{\mathrm{N}\,|\alpha|^2}}\right\}\right) \leqq \mathrm{CN}^{-2}.$$

By the Borel-Cantelli Lemma there is a set $\Omega_2 \subseteq \Omega$ of full measure so that if $\omega \in \Omega^2$ then for $\alpha \in \mathrm{H}_{\mathrm{N}}$

$$|\hat{\mathrm{A}}_{\mathrm{N}}^{\omega}(\alpha)| \leqq \mathrm{C}_\omega \sqrt{\frac{\log \mathrm{N}}{\mathrm{N}\,|\alpha|^2}}. \tag{19}$$

But we want (19) to hold for every α. An application of the Law of Large Numbers shows that there is a set $\Omega_3 \subseteq \Omega$ of full measure so that if $\omega \in \Omega_3$ then for $|\alpha - \beta| \leqq \mathrm{N}^{-2}$

$$|\hat{\mathrm{A}}_{\mathrm{N}}^{\omega}(\alpha) - \hat{\mathrm{A}}_{\mathrm{N}}^{\omega}(\beta)| \leqq \frac{\mathrm{C}_\omega}{\mathrm{N}} \tag{20}$$

(estimate the differences of exponentials, or take the gradient). Now (19) and (20) imply (12), since the right-hand side of (19) is at least $C_\omega \sqrt{\log N/N}$ and we may assume that $C_\omega \sqrt{\log N/N} \, \frac{1}{|\alpha|} < 1$.

THEOREM 8. – *A.e.* $\omega \in \Omega$ *has the following property: for every measure-preserving action of* $\mathbb{Z}^d$ *on a* σ*-finite measure space* $(Y, \mathcal{C}, \nu)$ *and every* $g \in L^2(Y)$,

$$\lim_{N\to\infty} \frac{1}{N(\log N \log\log N)^{1/4}} \sum_{n=1}^{N} g(U^{S_n(\omega)} y) = 0 \quad \text{a.e. } d\nu(y).$$

Proof. – We may assume that the support of ρ generates $\mathbb{Z}^d$. Denote

$$B_N^\omega g(y) = \frac{1}{N(\log N \log\log N)^{1/4}} \sum_{n=1}^{N} g(U^{S_n(\omega)} y).$$

Fix $r > 1$ and let $N_t = N_t(r) = \lfloor r^t \rfloor$. It is enough to prove that for each $r > 1$,

$$\lim_{t\to\infty} B_{N_t}^\omega g = 0.$$

We are going to compare B_N^ω to

$$W_N g(y) \frac{1}{N(\log N \log\log N)^{1/4}} \sum_{n=1}^{N} R^n g(y),$$

where R is as in the proof of Theorem 5. By Hopf's theorem, for each $g \in L^2$ and a.e. y we have

$$\lim_{N\to\infty} W_N g(y) = 0.$$

As before, it is enough to show that for a.e. $\omega \in \Omega$

$$\sup_\alpha \sum_{t=1}^{\infty} |\hat{B}_{N_t}^\omega(\alpha) - \hat{W}_{N_t}(\alpha)|^2 \leqq C_\omega, \tag{21}$$

where the Fourier transforms of the kernels are

$$\hat{B}_N^\omega(\alpha) = \frac{1}{N(\log N \log\log N)^{1/4}} \sum_{n=1}^{N} e(S_n(\omega)\cdot\alpha)$$

and

$$\hat{W}_N(\alpha) = \frac{1}{N(\log N \log\log N)^{1/4}} \sum_{n=1}^{N} (\phi(\alpha))^n,$$

with $\phi(\alpha)$ as before. If

$$u_{\mathrm{N}} = \frac{2}{(\log \mathrm{N} \log \log \mathrm{N})^{1/4}}$$

and

$$v_{\mathrm{N}} = \frac{\sqrt{\mathrm{N}}\,(\log \log \mathrm{N})^{1/4}}{\log^{1/4} \mathrm{N}},$$

then the estimate (13) gives

$$|\hat{\mathrm{B}}_{\mathrm{N}}^{\omega}(\alpha) - \hat{\mathrm{W}}_{\mathrm{N}}(\alpha)| \leqq \min \left\{ u_{\mathrm{N}},\, \mathrm{C}_{\omega} |\alpha| v_{\mathrm{N}},\, \frac{\mathrm{C}_{\omega}}{v_{\mathrm{N}} |\alpha|} \right\}.$$

This time we obtain the "trivial" bound u_{N} when α is inside a small cube but outside a really small cube at stage N $\left(\frac{u_{\mathrm{N}}}{\mathrm{C}_{\omega}} \frac{1}{v_{\mathrm{N}}} \leqq |\alpha| \leqq \frac{\mathrm{C}_{\omega}}{u_{\mathrm{N}}} \frac{1}{v_{\mathrm{N}}} \right)$. Again, along the lacunary sequence $\{\mathrm{N}_t\}$, the two sequences of squares pass each other after a fixed delay: there is an s such that.

$$\frac{\mathrm{C}_{\omega}}{u_{\mathrm{N}_t}} \frac{1}{v_{\mathrm{N}_t}} \leqq \frac{u_{\mathrm{N}_{t+s}}}{\mathrm{C}_{\omega}} \frac{1}{v_{\mathrm{N}_{t+s}}} \qquad \text{for all } t,$$

and so we may again select t_0 to be the largest t with

$$\frac{u_{\mathrm{N}_t}}{\mathrm{C}_{\omega}} \frac{1}{v_{\mathrm{N}_t}} \geqq |\alpha|$$

and decompose the sum into three pieces as before.

REFERENCES

[1] A. BELLOW and V. LOSERT, The weighted pointwise ergodic theorem and the individual ergodic theorem along subsequences, *Trans. Amer. Math. Soc.*, **288**, 1985, p. 307-345.

[2] A. BELLOW, R. L. JONES and J. ROSENBLATT, Almost everywhere convergence of weighted averages, *Math. Ann.*, **293**, 1992, p. 399-426.

[3] V. BERGELSON, M. BOSHERNITZAN and J. BOURGAIN, *Some results on non-linear recurrence*, Preprint.

[4] N. H. BINGHAM and Ch. M. GOLDIE, Probabilistic and deterministic averaging, *Trans. Amer. Math. Soc.*, **269**, 1982, p. 453-480.

[5] J. R. BLUM and R. COGBURN, On ergodic sequences of measures, *Proc. Amer. Math. Soc.*, **51**, 1975, p. 359-365.

[6] J. BOURGAIN, On the maximal ergodic theorem for certain substets of the integers, *Israel J. Math.*, **61**, 1988, p. 39-72.

[7] J. BOURGAIN, An approach to pointwise ergodic theorems, *Springer Lecture Notes in Math.*, **1317**, 1988, p. 224-223.

[8] J. BOURGAIN, H. FURSTENBERG, Y. KATZNELSON and D. ORNSTEIN, Return times of dynamical systems, appendix to Pointwise ergodic theorems for arithmetic sets, by J. Bourgain, *Pub. Math. I.H.E.S.*, **69**, 1989, p. 5-45.

[9] A. DEL JUNCO and J. ROSENBLATT, Counterexamples in ergodic theory and number theory, *Math. Ann.*, **245**, 1979, p. 185-197.

[10] R. DURRETT, *Probability theory and examples*, Wadsworth & Brooks/Cole, Belmont, CA, 1991, p. 394.

[11] P. HALL and C. C. HEYDE, *Martingale limit theory and its applications*, Academic Press, New York, 1980.

[12] R. JONES, J. OLSEN and M. WIERDL, Subsequence ergodic theorems for L^p contractions, *Trans. Amer. Math. Soc.*, To appear.

[13] R. JONES and M. WIERDL, *Examples of a.e. convergence and divergence of ergodic averages*, Preprint.

[14] L. MEILIJSON, The average of the value of a function at random points, *Israel J. Math.*, **15**, 1973, p. 193-203.

[15] K. PETERSEN, *Ergodic Theory*, Cambridge University Press, Cambridge and New York, 1983 and 1989.

[16] K. PETERSEN, On a series of cosecants related to a problem in ergodic theory, *Comp. Math.*, **26**, 1973, p. 313-317.

[17] M. WIERDL, Pointwise ergodic theorem along the prime numbers, *Israel J. Math.*, **64**, 1988, p. 315-336.

(Manuscript received October 8, 1992;
revised March 17, 1993.)