ADVANCES IN MATHEMATICS 63, 213–246 (1987)

Convex-Invariant Means and a Pathwise Central Limit Theorem

ALBERT FISHER*

Department of Mathematics, Northwestern University, Evanston, Illinois 60201

I. INTRODUCTION

How can one choose a point randomly from the real number line? An infinite-space Monte Carlo method suggests a possible way to make this choice. Begin a random walk at the origin, and stop at a random time, that is, at a time chosen randomly from the positive reals. But this method is circular! We have used some notion of random (or, one might say, uniform) time distribution to define a notion of uniform space distribution. Mathematically the question therefore becomes: are they (or can they be) the same distribution? As an application of the results of this paper, we will see that there do exist random distributions which are consistent in this sense. What is needed is a very strong notion of uniformity; the most fundamental notion, translation invariance, is not by itself enough. In this paper we study three additional sources of uniformity: convex scale invariance, consistency with averaging methods, and Mokobodzki's measure-linearity.

Recall that a *mean* is a positive, normalized, (continuous) linear functional on $L^\infty(\mathbb{R})$—see [Gr]. A mean determines a finitely additive probability measure on the Lebesgue measurable subsets of $\mathbb{R}$ and the set of means includes $P_1(\mathbb{R}) = \{\phi \in L^1(\mathbb{R})\colon \phi(x) \geqslant 0$ for a.e. x, and $\int \phi \, dx = 1\}$, which correspond to the countably additive measures.

One reasonable condition for a uniformly distributed mean λ to satisfy is *convolution-invariance* (called topological invariance in [Gr]), that $\lambda(f) = \lambda(\phi * f)$ for each $\phi \in P_1$. This is easily seen to imply translation-invariance, but the converse is false [Ru2; Ra, 5.1, 2]. Yet this is a natural requirement, viewing $(\phi * f)(t)$ in either of two ways, as the local smoothing $\int \phi(t-x) f(x)\, dx$, or as a weighted average of translates of $f(t)$, $\int \phi(x) f(t-x)\, dx$.

* Partially supported by the National Science Foundation under a Post-Doctoral Fellowship.

0001-8708/87 $7.50

Another natural requirement is dilation-invariance. This corresponds to changing the units of the number line, which should not change the average value of a function. But this is not guaranteed by convolution- (or translation-) invariance (Example 5.8).

Any invariant mean is a convex combination of two invariant pieces, weighted at plus and at minus infinity. Each piece extends the usual notion of limit, and can be viewed as a type of generalized limit (Banach limit). So one could hope that a uniform mean would also agree with various reasonable averaging methods.

The Cesaro average, for instance, is defined in a simple and natural way, and converges, to the "right" value, in many applications, such as for f periodic or almost periodic, or (more generally!) for bounded $f(t)$ randomly generated by a stationary stochastic process (by Birkhoff's ergodic theorem).

Here is a natural example where the Cesaro averaging method is, however, not sufficiently strong. Let $B(t)$ be a Brownian motion path in the reals, starting at zero. Let I be an interval in $\mathbb{R}$, and consider the function $f(t)=\chi_I((1/\sqrt{t})\, B(t))$. A pathwise Central Limit Theorem would state that the average time spent at value one is equal to $(1/\sqrt{2\pi}) \int_I e^{-x^2/2}\, dx$. But, as we show in Proposition 5.9, the Cesaro limit of $f(t)$ almost surely (with respect to Wiener measure) diverges.

We could instead try a sort of abstract extension of the Cesaro averages. Let $\lambda_t(f)=(1/t) \int_0^t f(x)\, dx$. This is a mean, and by the Banach–Alaoglu theorem the set of means is weak-$*$ compact. So there exists a limit point λ, which will necessarily be an invariant mean. It will be consistent with the Cesaro limit but defined everywhere. But as we will see, such a λ will also fail to give the pathwise CLT (Proposition 5.9).

What we really need for a satisfactory answer is a stronger averaging method. In fact, all we need is to move one stage beyond the Cesaro limit. We call the convolution operators $f \to \phi * f$ for $\phi \in P_1(\mathbb{R})$ *averaging operators of order zero.* Let $P_1(\mathbb{R}^+)$ denote those functions in $P_1(\mathbb{R})$ with support on the positive reals, and call an operator of the form $f(t) \to (1/t) \int \psi(x/t)\, f(x)\, dx$ for $\psi \in P_1(\mathbb{R}^+)$ a *Cesaro-type average.* Define the operator $E\colon L^\infty(\mathbb{R}) \to L^\infty(\mathbb{R})$ by $(Ef)(x)=f(\exp(x))$ which gives an exponential change of scale. Define $E^{-1}\colon L^\infty(\mathbb{R}) \to L^\infty(\mathbb{R})$ by $(E^{-1}f)(x)=f(\log(x))$ for $x>0$, zero for $x \leqslant 0$. Let E^0 be the identity and set $(E^{-n})=(E^{-1})^n$ for $n>0$. For $\phi \in P_1(\mathbb{R})$ define $A_\phi^0(f)=\phi * f$. For $n \in \mathbb{Z}$ *an averaging operator of order n* is of the form $A_\phi^n=E^{-n}A_\phi^0 E^n$, and an *averaging method of order n* is $\lim_{t \to +\infty} (A_\phi^n f)(t)$. Setting $\psi=(1/x)\, \phi(\log(1/x))$ for $x>0$, we see that the Cesaro-type operators are exactly the averaging operators of order one. For example, $\lambda_t=(1/t) \int_0^t$ corresponds to $\psi=\chi_{[0,1]}$ and $\phi(x)=e^{-x}\chi_{(0,\infty)}(x)$.

Perhaps the best way to understand this correspondence is to notice that since exp is a homomorphism from the additive to the multiplicative reals, convolution in $(\mathbb{R}, +)$ must be taken to convolution in $(\mathbb{R}^+, \cdot)$, which is the Cesaro-type average (since this is an average of dilations).

We will show that although the Cesaro-type averages almost surely diverge for the pathwise CLT, the order two (or higher) methods almost surely converge. There exists a mean consistent simutaneously with averaging methods of all orders, and moreover there exists a convolution-invariant mean which is invariant with respect to an exponential change of scale.

The proof of this turns out to be easy as well as instructive, so we suggest the reader begin there (Theorem 2.6). The more general formulation (all that is actually needed is the convexity of the scale change!) is harder to prove (Theorem 2.2).

In Section III we combine this scale-change approach with some ideas originally due to Mokobodzki [M, H-J]. His proof of the existence of measure-linear means on the integers used the Continuum Hypothesis, and it is now known that Martin's axiom is sufficient (proved by Norman [H-J]). Martin's axiom is, apparently, more readily accepted by set theorists—see the discussion in [H—J]. A sample application of measure-linearity, combined with exponential invariance, proves the Monte Carlo consistency discussed in the first paragraph (Theorem 4.2). Another application provides a second proof of the pathwise Central Limit Theorem (Theorem 4.1). A key step in the proof is an ergodic theoretic interpretation of Brownian motion: the scaling flow on Brownian paths is a Bernoulli flow of infinite entropy (Theorem 4.8). This leads to a new proof of the Blumenthal 0–1 law for Brownian motion (Lemma 4.10). These results also hold for Brownian motion in $\mathbb{R}^n$.

In the final section we give some examples and counterexamples and pose some questions. We also mention matters which will be developed further in two forthcoming papers: a pathwise CLT and Donsker's theorem for random walks in $\mathbb{R}^n$, and an extension of the Monte Carlo method to random walks on non-amenable groups and to their actions on a finite measure space.

In the non-amenable case the notion of average value is replaced by a projection onto the μ-harmonic functions, which is determined by the choice of a measure-linear Cesaro-consistent mean on the integers. The Monte Carlo method now produces boundary values which, when integrated, give the projection. Integration is necessary precisely because the Blumenthal 0–1 law no longer holds. For group actions the correspondence between boundary values and the harmonic projection has an interesting interpretation: the derivation of Kakutani and Yosida's ergodic theorem from the random ergodic theorem [Ka].

II. Convex Changes of Scale

Definition. Let $\mathscr{C}o$ denote the set of increasing convex functions c: $[0, +\infty) \to [0, +\infty)$ satisfying:

(1) c continuous (and $c(0)$ finite),

(2) $c'(x) > 0$ and $c''(x) \geqslant 0$ $\forall x$,

(3) $\lim_{x \to +\infty} c'(x) = +\infty$.

Partially order $\mathscr{C}o$ by $c < d$ if $\exists e \in \mathscr{C}o$ such that $c \circ e = d$, $c = d$ if $c(x) = d(x)$ for all x, and $c \leqslant d$ if $c < d$ or $c = d$.

Remarks. $\mathscr{C}o$ is closed under composition, so this relation is transitive. Smoothness of c provides a convenient way of specifying the sort of convexity the proof uses but is not necessary.

Definition. For $c \in \mathscr{C}o$, define C: $L^\infty(\mathbb{R}) \to L^\infty(\mathbb{R})$ by

$$(Cf)(x) = \begin{cases} f \circ c(x), & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

and C^{-1}: $L^\infty(\mathbb{R}) \to L^\infty(\mathbb{R})$ by

$$(C^{-1}f)(x) = \begin{cases} f \circ c^{-1}(x), & x \geqslant c(0) \\ 0, & x < c(0). \end{cases}$$

For $\phi \in P_1(\mathbb{R})$, with $A_\phi^0(f) = \phi * f$, define A_ϕ^c: $L^\infty \to L^\infty$ by $A_\phi^c(f) = (C^{-1}A_\phi^0 C)(f)$, and define $A_\phi^{c^{-1}}(f)$ analogously.

Theorem 2.1. *Let $\mathscr{G}$ be a directed subset of $\mathscr{C}o$. That is, any two elements of $\mathscr{G}$ are majorized by a third. Then there exists a mean λ on $L^\infty(\mathbb{R})$ such that for all $\phi \in P_1(\mathbb{R})$, for all $c \in \mathscr{G}$ (the limits being taken at $+\infty$),*

(a) $\underline{\lim}\, \phi * (f \circ c) \leqslant \lambda f \leqslant \overline{\lim}\, \phi * (f \circ c)$,

(b) $\lambda f = \lambda(A_\phi^c(f))$.

Theorem 2.2. *Let $c \in \mathscr{C}o$. Then there exists a mean λ such that $\lambda f = \lambda(\phi * f)$ for each $\phi \in P_1(\mathbb{R})$, and also $\lambda f = \lambda(f \circ c)$.*

We need first:

Lemma 2.3. *For $c \in \mathscr{C}o$, $f \in L^\infty(\mathbb{R})$, ϕ and $\psi \in P_1(\mathbb{R})$ we have $\overline{\lim}\, \psi * (f \circ c) \leqslant \overline{\lim}\, \phi * f$.*

Proof. Define the functional ψ_t: $L^\infty(\mathbb{R}) \to \mathbb{R}$ by $\psi_t(f) = A_\psi^c(f) = ((\psi * (f \circ c)) \circ c^{-1})(t)$. This is linear and continuous, and is explicitly given by integration against the L^1 function $\psi(c^{-1}(t) - c^{-1}(x))(c^{-1})'(x)$. We denote this function as $\psi_t(x)$, and can therefore write $\psi_t(f) = \int \psi_t(x) f(x)\, dx$. This agrees with what will be our usual notation for L^1 acting on L^∞. We will be done if we show

$$\psi_t(\phi * f - f) \to 0 \qquad \text{as} \quad t \to +\infty, \tag{$*$}$$

since if $(*)$ holds, $\overline{\lim}\, \psi_t(f) = \overline{\lim}\, \psi_t(\phi * f)$ and so

$$\overline{\lim}\, \psi * (f \circ c) = \overline{\lim}\, \psi_t(f) = \overline{\lim}\, \psi_t(\phi * f) \leqslant \overline{\lim}(\phi * f).$$

The last inequality uses the fact that, for $g \in L^\infty$ and $\psi \in P_1$, $\overline{\lim}\, \psi * g \leqslant \overline{\lim}\, g$, which is an immediate consequence of Hölder's inequality for L^∞ and L^1.

Now we prove $(*)$. We use the notation $\tilde{h}(x) = h(-x)$.

$$\begin{aligned}
\psi_t(\phi * f - f) &= [\tilde{\psi}_t * (\phi * f - f)](0)\\
&= [\tilde{\psi}_t * (\phi * f) - \tilde{\psi}_t * f](0)\\
&= ([(\tilde{\psi}_t * \phi) - \tilde{\psi}_t] * f)(0)\\
&= [(\psi_t * \tilde{\phi}) - \psi_t](f).
\end{aligned}$$

Therefore it is sufficient to show (replacing ϕ by $\tilde{\phi}$, also in P_1) that $\|\psi_t * \phi - \psi_t\|_1 \to 0$, which will imply $(*)$ by Hölder's inequality—in other words, since strong implies weak convergence in L^1.

Consider the set $E_\phi \subseteq P_1$, the set of ψ such that $\|\psi_t * \phi - \psi_t\|_1 \to 0$ for a fixed ϕ. This is a closed convex set in L^1, so it will be enough to prove this for ψ the normalized characteristic function of an interval. This follows, for instance, from the Martingale Convergence Theorem [P], by partitioning successively larger intervals in $\mathbb{R}$ into small ones, and averaging over each.

Let $\mathscr{C}(\mathbb{R})$ denote the bounded continuous functions. Its dual is $\mathscr{M}(\mathbb{R})$, the finite signed measures, which contains $L^1(\mathbb{R})$. Let $\mathscr{M}_1(\mathbb{R})$ denote the probability measures. This is convex and weak-$*$ compact in $\mathscr{M} = (\mathscr{C})^*$, and contains $P_1(\mathbb{R})$. Its extreme points are the point masses. Therefore $\phi \in P_1$ can be well-approximated (as integrated against continuous functions) by a finite convex combination of point masses, and such an approximation carries over to functions with the special form we are assuming for ψ_t, $\psi_t = (1/(\beta - \alpha))\, \chi_{[\alpha,\beta]}$ for $\alpha, \beta \in \mathbb{R}$. We conclude that it is enough to show that for $a \in \mathbb{R}$, $\|\psi_t * \delta_a - \psi_t\|_1 \to 0$ as $t \to +\infty$, i.e., $\int |\psi_t(x-a) - \psi_t(x)|\ dx \to 0$. Now $\psi_t(x) = \psi(c^{-1}(t) - c^{-1}(x))(c^{-1})'(x) = c^{-1\prime}(x)\, \chi_{[\bar{\alpha},\bar{\beta}]}$ where $\bar{\alpha} = c(c^{-1}(t) - \beta)$, $\bar{\beta} = c(c^{-1}(t) - \alpha)$. Note that $(c^{-1})'(x)$

is a decreasing function, due to the convexity of $c(x)$. Assume now, for instance, that $a \geqslant 0$, and define the function

$$\bar{\psi}_t^a = (\delta_a * \psi_t) + (\chi_{[\bar{\alpha}, \bar{\alpha}+a]})(\psi_t(\bar{\alpha})).$$

This is everywhere greater than ψ_t, so

$$\begin{aligned}\|\psi_t * \delta_a - \psi_t\| &\leqslant \|\psi_t * \delta_a - \bar{\psi}_t^a\| + \|\bar{\psi}_t^a - \psi_t\| \\ &\leqslant a\psi_t(\bar{\alpha}) + \int \bar{\psi}_t^a(x)\,dx - \int \psi_t(x)\,dx \\ &\leqslant 2a\psi_t(\bar{\alpha}) + \int \psi_t(x-a)\,dx - \int \psi_t(x)\,dx \\ &= 2a\psi_t(\bar{\alpha}) = 2a(c^{-1})'(\bar{\alpha}) \\ &= 2a(c^{-1})'(c(c^{-1}(t) - \beta)\end{aligned}$$

which $\to 0$ as $t \to \infty$ since the argument is a composition of increasing functions (since c', $(c^{-1})' \geqslant 0$), hence is increasing, and $\lim_{x \to \infty} (c^{-1})'(x) = \lim_{x \to \infty} (1/c'(x)) = 0$ by another property of $c \in \mathscr{C}_{\mathcal{O}}$. See also the easier proof of Theorem 2.6 for the specific case $c(x) = \exp(x)$. ∎

LEMMA 2.4. *For $c, d \in \mathscr{C}_{\mathcal{O}}$ with $c < d$, for $f \in L^\infty(\mathbb{R})$ and $\phi, \psi \in P_1(\mathbb{R})$ we have* $\overline{\lim}\, \psi * (f \circ d) \leqslant \overline{\lim}\, \phi * (f \circ c)$.

Proof. This follows immediately from Lemma 2.3 and the nature of the ordering on $\mathscr{C}_{\mathcal{O}}$. ∎

LEMMA 2.5. *For $c \in \mathscr{C}_{\mathcal{O}}$, and $\phi, \psi \in P_1(\mathbb{R})$:*

(a) *if λ is a mean such that $\forall f \in L^\infty$, $\underline{\lim}\, \psi * (f \circ c) \leqslant \lambda f \leqslant \overline{\lim}\, \psi * (f \circ c)$, then $\lambda(A_\phi^0 f) = \lambda f$ for all $f \in L^\infty$.*

(b) *If $c, d \in \mathscr{C}_{\mathcal{O}}$ with $c < d$, then if $\underline{\lim}\, \psi * (f \circ d) \leqslant \lambda f \leqslant \overline{\lim}\, \psi * (f \circ d)$, we have $\lambda(A_\phi^c f) = \lambda f$.*

Proof. In the proof of Lemma 2.3 we showed that $\psi_t(\phi * f - f) \to 0$ as $t \to \infty$. So, $\lambda(A_\phi^0 f - f) = \lambda(\phi * f - f) \leqslant \overline{\lim}\, \psi_t(\phi * f - f) = 0$. The same holds for $\underline{\lim}$. For part (b) see the proof of Lemma 2.4.

Proof of Theorem 2.1. Given $c \in \mathscr{C}_{\mathcal{O}}$, let $M_c(\phi)$ denote the set of means λ such that $\underline{\lim}\, \phi * (f \circ c) \leqslant \lambda f \leqslant \overline{\lim}\, \phi * (f \circ c)$ for a fixed $\phi \in P_1(\mathbb{R})$. We claim that $M_c(\phi)$ is non-empty, convex, and weak-$*$ compact. Define $\lambda_t^c : L^\infty \to \mathbb{R}$ by $\lambda_t^c(f) = (\phi * (f \circ c))(t)$. This is a mean. Let λ be a weak-$*$ limit point of the λ_t^c; λ is in $M_c(\phi)$ so this is non-empty. Clearly $M_c(\phi)$ is convex and closed, hence compact.

Next, let $M_c = \bigcap_{\phi \in P_1} M_c(\phi)$. This is compact and convex; we want to show it is non-empty. Let $d = c \circ c$ and let $\psi \in P_1$; $M_d(\psi)$ is non-empty as above. Let $\lambda \in M_d(\psi)$. By Lemma 2.4, since c and d are in $\mathscr{C}_\mathscr{O}$ and $d > c$, $\lambda \in M_c(\phi)$ for all ϕ, hence $\lambda \in M_c$ which is therefore nonempty.

Now let $M_{\mathscr{G}} = \bigcap_{c \in \mathscr{G}} M_c$. By Lemma 2.4, $c < d \Rightarrow M_c \supseteq M_d$. So compactness and the directed set condition imply M is non-empty and we have proved part (a). (Of course $M_{\mathscr{G}}$ is also compact and convex.)

Finally, the set of means which is A_ϕ^c-invariant is non-empty by the same reasoning plus (b) of Lemma 2.5, and is also compact (and convex), so we proceed as before to finish part (b). Note that by Lemma 2.5, this intersection is, in fact, exactly equal to $M_{\mathscr{G}}$. ∎

Remark. One could also prove (b) first and get (a) as a corollary. This is because for any mean λ, $\lambda f \leqslant \operatorname{ess\,sup} f$, so if λ is weighted at $+\infty$, $\lambda f \leqslant \overline{\lim}\, f$. This holds for λ which is A_ϕ^c-invariant for some $c \in \mathscr{C}_\mathscr{O}$. Hence, $\lambda f = \lambda(A_\phi^c f) \leqslant \overline{\lim}(C^{-1} A_\phi^0 C) f = \overline{\lim}\, A_\phi^0 C f$ as claimed.

Proof of Theorem 2.2. Let $\mathscr{G} = \{c^{(n)}(x) = c \circ \cdots \circ c(x)\ n \text{ times, for all } n \geqslant 1\}$. This set satisfies the hypothesis of Theorem 2.1, and $<$ is just the order of $\mathbb{N}$. By Theorem 2.1, $M_{\mathscr{G}}$ is non-empty and compact. The next step is closely related to the proof that a solvable group is amenable; see, e.g., [Gr]. Denote operators A_ϕ^d for $d = c^{(n)}$, $n \in \mathbb{Z}$, by A_ϕ^n. Any mean in $M_{\mathscr{G}}$ is also invariant under operators of the form A_ϕ^{-1}, and more generally A_ϕ^{-n} for $n \geqslant 1$, by Lemma 2.5. $M_{\mathscr{G}}$ is just the normalizer of the convolutions, with extending group $\mathbb{Z}$ (but note that the convolutions actually form a semigroup not a group). We want to average over $\mathbb{Z}$, which is amenable. Precisely, let m be any invariant mean on $l^\infty = L^\infty(\mathbb{Z})$, let $\lambda_0 \in M_{\mathscr{G}}$, and define $\lambda(f) = m(\cdots \lambda_0 C^{-1} f, \lambda_0 f, \lambda_0 C f \cdots)$. Because m is translation-invariant, $\lambda(Cf) = \lambda(f)$. Now we calculate $\lambda(A_\phi^0 f)$. It is easy to see that λ_0 is weighted at $+\infty$. For all $n \in \mathbb{Z}$, $A_\phi^n = C^{-n} A_\phi^0 C^n$ by definition. We claim that since λ_0 is a mean weighted at $+\infty$, $\lambda_0 C^n A_\phi^0 = \lambda_0 A_\phi^{-n} C^n$. This is because, for instance, $[C(C^{-1}(f))](x) = f(x)$ for $x \geqslant 0$, and $(C^{-1} C f)(x) = f(x)$ for $x \geqslant c(0)$, so $(A_\phi^0 C C^{-1} f)(x)$ is uniformly close to $(A_\phi^0 f)(x)$ for large x. Therefore,

$$\begin{aligned}\lambda(A_\phi^0 f) &= m(\ldots, \lambda_0 C^{-1} A_\phi f, \lambda_0 A_\phi^0 f, \lambda_0 C A_\phi^0 f, \ldots) \\ &= m(\ldots, \lambda_0 A_\phi^1 C f, \lambda_0 A_\phi^0 f, \lambda_0 A_\phi^{-1} C f, \ldots) = \lambda(f).\end{aligned}$$

This completes the proof of Theorem 2.2. ∎

The next theorem is exactly Theorem 2.2 for the case $c(x) = e^x$. The special nature of the exponential enables us to use familiar properties of L^1-approximate identities to simplify the proof, so we give this self-contained proof here.

THEOREM 2.6. *There exists a mean λ on $L^\infty(\mathbb{R})$ satisfying*:

(1) convolution-invariance: *for* $f \in L^\infty(\mathbb{R})$ and $\phi \in P_1(\mathbb{R})$, $\lambda(f) = \lambda(\phi * f)$;

(2) exponential-invariance: $\lambda(f(x)) = \lambda(f(e^x))$.

Note. *The following properties follow from* (1) *and* (2); *see the Introduction.*

(3) translation-invariance;

(4) dilation-invariance: $\lambda(f(x)) = \lambda(f(rx))$, $r > 0$;

(5) *invariance under Cesaro-type averages* (*see the Introduction*);

(6) *invariance under composition with* $c(x) = x^r$, $r > 1$.

One can directly prove the existence of a mean λ satisfying (3) and (4) in this simpler way: the $ax + b$ group is solvable hence amenable [Gr], and this carries over to the action on $\mathbb{R}$. Note (see the Introduction) that (1) is to (2) as (5) is to (4). Note that (2) implies λ is weighted at $+\infty$; to get a symmetric version of this mean one simply defines $\bar{\lambda}(f) = \frac{1}{2}\lambda(f) + \frac{1}{2}\lambda(\tilde{f})$. We remark that x^2-invariance is used to prove the Monte Carlo method (Theorem 4.2).[1]

Proof. Averaging operators A^n_ϕ for $n \in \mathbb{Z}$ are defined as in the Introduction. There exists a mean λ_0 which is convolution-invariant and weighted at $+\infty$; take any weak-$*$ limit point of $\lambda_t = (1/t)\int_0^t$ or see the proof of Theorem 2.1. Let m be an invariant mean on $l^\infty(\mathbb{Z})$ which is weighted at $+\infty$ (so we picture it as living on $\mathbb{N}$). We claim that λ defined by $\lambda(f) = m(\lambda_0 f, \lambda_0 Ef, \lambda_0 E^2 f, \ldots)$ satisfies the theorem. This is E-invariant by translation-invariance of m. We will prove A^0_ϕ-invariance by establishing a lemma: that A^0_ψ-invariance for a fixed $\psi \in P_1$ implies A^{-n}_ϕ-invariance for all $\phi \in P_1$, all $n \geqslant 1$. By conjugation with E, it is sufficient to prove that A^1_ψ-invariance implies A^0_ϕ-invariance for all ϕ. Define $\psi_t(x) = (1/x)\, \psi(\log(t/x))$ for $x > 0$, zero for $x \leqslant 0$. One checks that $A^1_\psi(f) = \int \psi_t(x) f(x)\, dx = \psi_t(f)$, with $\psi_t(x) \in P_1(\mathbb{R}^+)$. Defining $\hat{\psi}(x) = \psi_1(x)$, we see that $\psi_t(x) = (1/t)\, \hat{\psi}(x/t)$, a Cesaro-type operator.

Caution. The notation here differs slightly from that of the Introduction.

We need to show that $\lambda(\phi * f - f) = 0$. Since λ is E-invariant, it is weighted at $+\infty$, and since it is a mean, $\lambda f \leqslant \overline{\lim}_{+\infty} f$. Therefore

$$\lambda(\phi * f - f) = \lambda(A^1_\psi(\phi * f - f)) \leqslant \overline{\lim}\, A^1_\psi(\phi * f - f).$$

[1] Property (4) guarantees that λ gives the "correct" solution (Benford's law) to the first digit problem of statistics. See R. Rami, The first digit problem, *Amer. Math. Monthly* **83** (1976), 521–538. I am indebted to P. Diaconis for this reference.

Now

$$A^1_\psi(\phi * f - f) = \psi_t(\phi * f - f) = (\tilde{\psi}_t * \phi * f - \tilde{\psi}_t * f)(0)$$
$$= (\tilde{\psi}_t * \phi - \tilde{\psi}_t)(\tilde{f}) \leqslant \|\tilde{\psi}_t * \phi - \tilde{\psi}_t\|_1 \|f\|_\infty.$$

One can prove that the first factor goes to zero by an approximation argument (see proof of Theorem 2.1) or as follows.

$$\|\tilde{\psi}_t * \phi - \tilde{\psi}_t\|_1 = \|\tilde{\psi}_1 * (t\phi(tx)) - \tilde{\psi}_1\|_1$$

by a change of variables. And $t\phi(tx)$ is an L^1-approximate identity as $t \to \infty$ so we are done with the lemma.

We summarize the proof so far. We have chosen λ_0 to be A^0_ψ-invariant, and know λ_0 is A^n_ϕ-invariant for all $n \leqslant -1$, and that λ is E-invariant. We show now that λ is A^0_ϕ-invariant for all $\phi \in P_1$. Since $EA^0_\phi E^{-1} = A^{-1}_\phi$, $EA^0_\phi = A^{-1}_\phi E$ as applied to $f(x)$, for all arguments $x \geqslant \log(\alpha)$, if ϕ is compactly supported on, say, $[-\alpha, \alpha]$. No matter how large the support of ϕ, we have $\lambda_0 EA^0_\phi = \lambda_0 A^{-1}_\phi E$ since λ_0 is a mean weighted at $+\infty$, and the same for $n > 1$. So,

$$\lambda(A^0_\phi f) = m(\lambda_0 A^0_\phi f, \lambda_0 EA^0_\phi f, \lambda_0 E^2 A^0_\phi f, \ldots)$$
$$= m(\lambda_0 f, \lambda_0 Ef, \lambda_0 E^2 f, \ldots) = \lambda(f). \quad \blacksquare$$

Remark. Geometrically, we continuously expand our perspective as $t \to \infty$, so that ψ_t remains stationary and ϕ becomes an approximate identity. For general $c(x)$, something close to this is still true, but we no longer get a convolution operator and therefore a precise version of the argument becomes too complicated, and the other approach seems easiest.

Question. How large can a directed subset $\mathscr{G}$ of the set $\mathscr{C}_o$ be? The next proposition shows $\mathscr{C}_o$ itself is in one sense big enough, and the example which follows shows it is in another sense too big. I would like to thank M. Magidor and B. Weiss for conversations about the proposition.

PROPOSITION 2.7. *For $f \in L^\infty(\mathbb{R})$, there exists $c \in \mathscr{C}_o$ such that* $\lim_{t \to +\infty} A^0_\phi(f \circ c)$ exists.

Proof. Actually we construct c which is continuous and piecewise linear, and could then round off the corners to make it smooth. We will choose c linear on each $[n, n+1]$, $n \geqslant 0$. So

$$\int_{[n,n+1]} f \circ c = \frac{1}{c(n+1) - c(n)} \int_{[c(n),c(n+1)]} f.$$

Choose a sequence $m(i) \in \mathbb{N}$, $i = 1, 2, 3, \ldots$, such that $(1/m(i)) \int_0^{m(i)} f$ converges and such that $(m(i+1) - m(i))/m(i+1) \to 1$ as $i \to \infty$. Therefore the Cesaro average $A^1_\phi(f \circ c)$ converges, and to the same number. Now replace c by $c \circ \exp$, which is still in $\mathscr{C}_\mathcal{O}$, and A^0_ϕ applied to it converges. ∎

Remark. Notice the arbitrariness—any value between the lim sup and lim inf of the Cesaro averages of f can be achieved!

EXAMPLE 2.8. Define $[\exp](x)$ to equal $\exp(x)$ for $x \in \mathbb{N}$, and to be linearly interpolated in between. We claim that $[\exp]$ and exp cannot be majorized by any $k \in \mathscr{C}_\mathcal{O}$, thus they cannot both be in the same directed set $\mathscr{G}$. Furthermore we claim that a mean which is convolution-invariant, and exp-invariant, cannot be $[\exp]$-invariant. Proof of first statement: Suppose $k \geqslant \exp$, $[\exp]$ so $\exists c$, $d \in \mathscr{C}_\mathcal{O}$ such that $k = \exp \circ c = [\exp] \circ d$. Now define $f \in L^\infty$ as follows. Let $g(x)$ be the periodic function with period 1 which equals $\chi_{[0,1/2]}$ on the unit interval. Define $f(x) = g \circ \log x$ for $x > 0$, zero otherwise. So for positive x, $f \circ \exp$ is the periodic function g. And $h(x) = f \circ [\exp](x)$ is also periodic of period one—on the unit interval it is $\chi_{[0,\alpha]}$ where $\alpha = (e^{1/2} - 1)(e - 1)$ which is less than $\frac{1}{2}$. Now let $\phi(x) = \chi_{[0,1]}$ which is in P_1. So $\phi * g$ is the constant function $\frac{1}{2}$, and $\phi * h$ is the constant function α. If there were $k = \exp \circ c = [\exp] \circ d$ then $\frac{1}{2} \leqslant \overline{\lim}\, \psi * (g \circ c) \leqslant \overline{\lim}\, \phi * g = \frac{1}{2}$ by Lemma 2.4, and yet

$$g \circ c = f \circ \exp \circ c = f \circ k = f \circ [\exp] \circ d = h \circ d.$$

So $\lim \psi * (f \circ k) = \frac{1}{2}$ and also $= \alpha$, a contradiction. Proof of the second statement: We can use the same counterexample, or this related one—if λ is exp and $[\exp]$-invariant, then it is $l = [\exp]^{-1} \circ \exp$-invariant, and for h, g as above, $h \circ l = f \circ \exp = g$, providing a contradiction.

Remark. The general philosophy here is that the non-uniqueness of means essentially comes from increasingly slow oscillations at ∞. Ideally one would like to define a consistent family of scale-changes which are uniformly convex in some natural sense (perhaps with simple conditions on the higher-order derivatives), and which "grow fast enough" to pull in any of those oscillations to average them, thus giving the result of the Proposition. It is far from clear whether this can be done. It may be the case, for instance, that any directed set $\mathscr{G}$ can be extended to a maximal directed set. Does a totally ordered subset have an upper bound? This would require "diagonalizing" the tower, within $\mathscr{C}_\mathcal{O}$. Would a maximal family then guarantee the result of the proposition? It seems that an answer may get involved with interesting set-theoretical questions. See also part (1) of the Conclusions in Section V.

III. Measure-Linearity

1. *Preliminaries*

Let $(X, \mathscr{T})$ be a topological space with Borel σ-algebra $\mathscr{B}$. Given a probability measure μ on $\mathscr{B}$ (i.e., $\mu(X) = 1$), we define its *completion*—which will also be denoted by μ—as follows. Recall that an *outer measure* is just like a measure except (countably) *sub*additive. For any $S \subseteq X$, define $\mu^*(S) = \inf\{\mu^* B\colon B \in \mathscr{B}$ and $B \supseteq S\}$, which is an outer measure. Let $\mathscr{B}_\mu$ denote the σ-algebra generated by $\mathscr{B}$ and the μ^*-null sets; μ extends to this in the obvious way. These are also exactly the μ^*-Carathèodory-measurable sets. The *inner measure* of a set is the outer measure of its complement, and sets in $\mathscr{B}_\mu$ are those with inner and outer measure equal [Roy]. We call these sets the μ-measurable sets, or the $\mathscr{B}$-μ-measurable sets. They are measurable with respect to any subspace, in the sense that their restrictions are the measurable sets for μ^* restricted to any (possibly nonmeasurable) subset $S \subseteq X$. Also see [Ro, p. 5].

Let $\mathscr{M}_1(X)$ denote the set of completed probability measures on X. The *universally measurable* sets are sets in the σ-algebra $\mathscr{B}_u(X) = \bigcap_{\mu \in \mathscr{M}_1} \mathscr{B}_\mu(X)$. Let $(\bar{X}, \bar{\mathscr{T}}, \bar{\mathscr{B}})$ be another Borel space. A function $w\colon X \to \bar{X}$ is *Borel measurable* if $w^{-1}(\bar{\mathscr{T}}) \subseteq \mathscr{B}$, is *$\mu$-measurable* if $w^{-1}(\bar{\mathscr{T}}) \subseteq \mathscr{B}_\mu$, and is universally measurable if $w^{-1}(\bar{\mathscr{T}}) \subseteq \mathscr{B}_u$ or equivalently if w is μ-measurable for each μ. In each case $\bar{\mathscr{T}}$ could be replaced by $\bar{\mathscr{B}}$.

Any function $w\colon X \to \bar{X}$ maps the outer measures on X to those on $\bar{X}$. This gives a natural map W from $\mathscr{M}_1$ to the outer measures, leading us to an alternative characterization of universal measurability.

Proposition 3.1. *A function $w\colon X \to \bar{X}$ is universally measurable if and only if W maps $\mathscr{M}_1(X)$ to $\bar{\mathscr{M}}_1(\bar{X})$.*

Proof. The outer measure μ^* of $\mu \in \mathscr{M}_1$ pushes forward to an outer measure $W(\mu) = \nu$ on $\bar{\mathscr{B}}$, and if $w^{-1}(\bar{\mathscr{B}}) \subseteq \mathscr{B}_\mu$ then ν is a measure. Conversely if ν is a measure on $\bar{\mathscr{B}}$, then the outer and inner measures of $w^{-1}(\bar{B})$ are equal for $\bar{B} \in \bar{\mathscr{B}}$, and hence w is μ-measurable. ∎

Proposition 3.2. *The composition of two universally measurable maps is universally measurable.*

Proof. Let $f\colon X \to Y$ and $g\colon Y \to Z$ be universally measurable, and write $h = g \circ f$. Choose a measure $\mu \in \mathscr{M}_1(X)$; we are to show that for S Borel $\subseteq Z$, $h^{-1}(S)$ is μ-measurable. Push μ forward to a measure ν on Y, using the previous proposition. Now $g^{-1}(S) \Delta B = N$, for some B Borel and N ν^*-null. And there is a Borel set N_0 in Y, such that $N_0 \supseteq N$ and $\nu(N_0) = 0$.

This pulls back to a μ^*-null set, and the conclusion follows easily. ∎

Let E be a topological vector space. We will always assume that points are closed and that the dual space separates points, which is the case if, for instance, E is locally convex Hausdorff [Ru1]. If $X \subseteq E$ is convex compact, then we have the *barycenter map* b: $\mathscr{M}_1(X) \to X$ defined by $\alpha(b(\mu)) = \int_X \alpha(x)\, d\mu(x)$ for $\alpha \in E^*$. Let w: $X \to \bar{X}$ be universally measurable, with $\bar{X}$ a convex compact subset of a t.v.s. $\bar{E}$. We say w is *measure-affine* if w is universally measurable and if $b(w(\mu)) = w(b(\mu))$ for each $\mu \in \mathscr{M}_1(X)$. By considering point masses, measure-affine implies affine. If w: $E \to \bar{E}$ is linear and measure-affine for all compactly supported measures, we say w is *measure-linear.*

PROPOSITION 3.3. *The composition of two measure-affine maps is measure-affine.*

Proof. We know that the composition is universally measurable. The rest follows automatically. ∎

2. *Measure-Linear Means*

The appropriate Borel structure on l^∞ is given not by the sup norm topology but by the product topology $\mathscr{T}$. Note, for instance, that the norm unit ball $K = \prod_{-\infty}^{\infty} [-1, 1]$ is a nice measure space, being compact, convex, and metrizable. This suggests the definition of the Borel sets for $L^\infty(\mathbb{R})$, or indeed for L^∞ of any locally compact separable metric space. Let $\mathscr{T}$ be the weak-$*$ topology for $L^\infty = (L^1)^*$. By the Banach–Alaoglu theorem the unit ball K is again compact. It is convex and separable (since L^1 is separable) hence metrizable. This defines $\mathscr{B}$ and the notions of universally measurable and measure-affine. Note that the weak-$*$ topology on $l^\infty = L^\infty(\mathbb{Z})$ is just, once more, the product topology.

The following theorem is due to Mokobodzki [M].[2] His proof made use of the Continuum Hypothesis, and Norman showed Martin's axiom is in fact sufficient. Here $X = l^\infty = L^\infty(\mathbb{N})$.

THEOREM 3.4 (Mokobodzki). *There exists a mean λ on l^∞ such that*

(i) $\underline{\lim}(1/n) \sum_{i=1}^{n} x(i) \leqslant \lambda(x) \leqslant \overline{\lim}(1/n) \sum_{i=1}^{n} x(i)$ *for all* $x \in l^\infty$;

(ii) *λ is measure-linear, that is, λ is universally measurable and for μ compactly supported,* $\lambda(b\mu) = \int_X \lambda(x)\, d\mu(x)$.

[2] See also J. P. R. Christensen, "Topology and Borel Structure," Chap. 6, North-Holland, Amsterdam, 1974, and his 1971 paper cited there. (He proves universal measurability only.) I thank P. Diaconis for providing me this reference.

If a function $w: l^\infty \to \mathbb{R}$ is linear and $\mathcal{T}$*-continuous* then it is automatically measure-linear—see [Ru1, Theorem 3.27]. However, an invariant mean λ on l^∞ can certainly never be $\mathcal{T}$-continuous. Mokobodzki's result is remarkable in that λ also cannot even be Borel measurable. B. Weiss showed me a nice proof of this fact. One considers the restriction of λ to sequences of 0's and 1's. Every Borel set differs from an open set by a set of first category, and one uses the Baire category theorem plus symmetry to show that $\lambda^{-1}\{\frac{1}{2}\}$ must be residual, from which one derives a contradiction. A more abstract proof is in [H-J, Lemma 8.2 and remark before Theorem 9.3].

EXAMPLE 3.5. Measure-linearity is a strong and useful property as these two examples illustrate.

(1) Let λ be a measure-linear invariant mean on l^∞. We will prove the Birkhoff ergodic theorem for λ—and the proof is absolutely trivial. Let T be a measure-preserving transformation of a measure space (Ω, m), and let $f: \Omega \to \mathbb{R}$ be bounded. Define $f^*(\omega) = \lambda(f(T^i\omega)))$ where $\omega \in \Omega$ and $i \in \mathbb{N}$. We see that f^* is measurable and T-invariant; we want to show that f^* has the same expected value as f. Define the map $\alpha: \Omega \to l^\infty$ by $\alpha(\omega) = x$ with $x(i) = f(T^i(\omega))$. Under α the measure m pushes forward to the (compactly supported) measure $\bar{m}$ in $\mathcal{M}_1(l^\infty)$. Its barycenter $b(\bar{m})$ is $y \in l^\infty$ with $y(i) = E(f \circ T^i) = E(f)$ for each $i \in \mathbb{N}$, so

$$E(f) = \lambda(b(\bar{m})) = \int \lambda(x)\, d\bar{m}(x) = \int f^*(\omega)\, dm(\omega) = E(f^*).$$

(2) For a mean λ on $L^\infty(\mathbb{R})$ which is measure-affine, we claim that translation-invariance implies convolution-invariance. Let $\phi \in P_1(\mathbb{R})$, and let $f \in K \subseteq L^\infty(\mathbb{R})$. Let m be the measure on $\mathbb{R}$ whose Radon–Nikodym derivative is ϕ. Defining the map $\alpha: \mathbb{R} \to K$ by $t \to f(x-t)$, α pushes m forward to a measure $\bar{m} \in \mathcal{M}_1(K)$. The barycenter of $\bar{m}$ is the function $\phi * f$, and so

$$\lambda(\phi * f) = \lambda(b\bar{m}) = \int \lambda(f(x-t))\, dm(t) = \lambda(f).$$

Now we are ready to state and prove these extensions of Theorems 2.1, 2.2, and 2.6. To prove Theorems 3.6, 3.7, and 3.8 we will use a key lemma of Mokobodzki and Norman. Theorem 3.4 is an easy corollary of Theorem 3.5, with, e.g., $c = \exp$.

THEOREM 3.6. *Let $\mathcal{G}$ be a directed subset of $\mathcal{C}_o$, with cardinality*

$\# \mathcal{G} < 2^{\aleph_0}$. Then under the assumption of Martin's axiom, there exists a mean λ on $L^\infty(\mathbb{R})$ such that for all $\phi \in P_1$, all $c \in \mathcal{G}$,

(a) $\underline{\lim}\, \phi * (f \circ c) \leqslant \lambda f \leqslant \overline{\lim}\, \phi * (f \circ c)$,

(b) $\lambda f = \lambda(A^c_\phi(f))$,

(c) *λ is measure-linear.*

THEOREM 3.7. *Let $c \in \mathscr{C}_\mathcal{O}$. Under Martin's axiom, there exists a mean λ which is convolution-invariant, c-invariant, and measure-linear.*

THEOREM 3.8. *Under the assumption of Martin's axiom, there exists a mean λ on $L^\infty(\mathbb{R}^+)$ satisfying:*

(1) *convolution-invariance,*

(2) *exponential-invariance,*

and therefore (3), (4), (5), *and* (6) *of Theorem* 3, *and also*

(7) *measure-linearity.*

As remarked in the example, $(7)+(3) \Rightarrow (1)$, and similarly $(7)+(4) \Rightarrow (5)$. To state the key lemma of Mokobodzki we need several definitions.

Let E be a t.v.s. (as in Section I) which is separable. Let K be a compact convex subset of E. $\Gamma_0^+(K)$ denotes the *bounded lower semicontinuous convex functions*: $K \to \mathbb{R}$, $\Gamma_0^- = -\Gamma_0^+$ *the upper semicontinuous concave functions.* For a family of convex functions $\{v_i\}_I$ on K, $\bigwedge_I v_i$ will denote the *convex infimum*, i.e., the largest convex function which is dominated by all the v_i. We define $\Gamma^+(K)$ to be the set of functions on K which are of the form $v = \bigwedge_I v_i$, for $v_i \in \Gamma_0^+(K)$ with $\# I < 2^{\aleph_0}$. The *concave supremum* $\bigvee u_i$ is defined similarly, giving us the class of functions $\Gamma^-(K)$ which equals $-\Gamma^+(K)$. To prove Theorems 3.6, 3.7, and 3.8, we need the following lemma. For proofs see [M], a short and sweet exposition which uses the Continuum Hypothesis, and [H-J], which presents valuable background information on the analysis and logic involved.

LEMMA 3.9. (In-Between Lemma [Mokobodzki and Norman]). *Let K be a compact convex subset of a separable t.v.s. E. Let $v \in \Gamma^+(K)$ and $u \in \Gamma^-(K)$ with $u(x) \leqslant v(x)$ for all $x \in K$. Then under Martin's axiom there exists w: $K \to \mathbb{R}$ measure-affine such that $u \leqslant w \leqslant v$.*

Proof of Theorem 3.6. Let E be $L^\infty(\mathbb{R})$ with the $\mathscr{T}$-topology, and let K be the unit ball. Define v: $K \to \mathbb{R}$ by

$$v(f) = \inf_{\phi \in P_1, c \in \mathcal{G}} \{\overline{\lim}\, \phi * (f \circ c)\}$$

and let $u(f)=\sup\{\underline{\lim}\phi * (f\circ c)\}$. We will show that $v\in\Gamma^+(K)$, $u\in\Gamma^-(K)$, and that $u\leqslant v$. The basic facts we will need are that the supremum of a family of uniformly bounded continuous functions is l.s.c., that the sup of a (bounded) family of convex functions is convex, and that a decreasing infimum of (bounded) convex functions is convex (and so is, in this case, the same as the convex infimum).

Using these facts, one sees that the map $\alpha: K\to\mathbb{R}$ defined by $\alpha(f)=\operatorname{ess\,sup}_{[0,+\infty]} f(x) = \sup_{i\in\mathbb{N}} (\operatorname{ess\,sup}_{x\in[i,i+1]} f(x))$ is $\mathscr{T}$-l.s.c. and convex. Note that $f\mapsto\phi * f$ for $\phi\in P_1$, and $f\mapsto f\circ c$ for $c\in\mathscr{G}$ are $\mathscr{T}$-continuous maps of L^∞. Now to each $c\in\mathscr{G}$ associate a function $\phi_c\in P_1$. By Lemma 4, using also the fact that the supremums decrease as n increases, and that $\mathscr{G}$ is a directed set, $v(f)=\inf_{(c,n)\in\mathscr{G}\times\mathbb{N}}\sup_{x\geqslant n} (\phi_c * (f\circ c))$. Therefore $v\in\Gamma^+(K)$. Finally, for each c, ϕ, and f, $\underline{\lim}\,\phi * (f\circ c)\leqslant\overline{\lim}\,\phi * (f\circ c)$, and therefore—using the fact that $\mathscr{G}$ is a directed set and Lemma 2.4—$u(f)\leqslant v(f)$. One could also use the conclusion of Theorem 2.1 to prove this last point. Now applying the in-between lemma, we are done. ∎

Proof of Theorem 3.7. Define $\mathscr{G}=\{c^{(n)}\}_{n\in\mathbb{N}}$ as in the proof of Theorem 2.2. By Theorem 3.6, there exists a mean λ_0 such that λ_0 is measure-linear and A^n_ϕ-invariant for all $n\in\mathbb{Z}$, $\phi\in P_1$. Now let m be an invariant mean on l^∞ which is also measure-linear. This exists by Mokobodzki's theorem (Theorem 3.4). Define the map $\lambda: L^\infty\to\mathbb{R}$ by $\lambda(f)=m(\ldots\lambda_0 C^{-1}f,\ \lambda_0 f,\ \lambda_0 Cf\ldots)$. By Proposition 3.2, λ is universally measurable and measure-linear. We conclude the argument just as in the proof of Theorem 2.2. ∎

Proof of Theorem 3.8. This is a corollary of the previous theorem. ∎

IV. Applications

THEOREM 4.1 (Pathwise Central Limit Theorem). *Let Ω be the space of Brownian motion paths on $\mathbb{R}$ for positive time, with initial condition $B(0)=0$ for $B\in\Omega$, and with Wiener measure ν. There exists an invariant mean λ on $L^\infty(\mathbb{R})$ such that for almost every B in Ω, for every Lebesgue measurable set $A\subseteq\mathbb{R}$, we have*

$$\lambda\left(\chi_A\left(\frac{1}{\sqrt{t}}B(t)\right)\right)=\frac{1}{\sqrt{2\pi}}\int_A e^{-x^2/2}\,dx.$$

In fact, it is sufficient for a mean λ to satisfy either:

(a) *convolution—and exponential—invariance, or*

(b) *measure-linearity and being weighted at* $+\infty$.

Let $(\hat{\Omega}, \hat{\nu})$ *denote the space of Brownian motion paths* $B(t)$ *defined for all times* $t \in \mathbb{R}$, *and satisfying* $B(0)=0$. *Recall (Theorem* 2.3) *that given a non-symmetric invariant mean* λ *on* $L^\infty(\mathbb{R})$, *we form the symmetric version by setting* $\tilde{\lambda}(f)=\lambda(\tilde{f})$ for $\tilde{f}(x)=f(-x)$, *and then defining* $\bar{\lambda}=\frac{1}{2}\lambda+\frac{1}{2}\tilde{\lambda}$.

THEOREM 4.2 (Monte Carlo method on $\mathbb{R}$). *With the assumption of Martin's axiom, there exists a mean* λ *on* $L^\infty(\mathbb{R})$ *which is translation-invariant and which is weighted at* $+\infty$, *such that for any* $f \in L^\infty$,

(a) *for* ν*-a.e.* $B \in \Omega$,

$$\bar{\lambda}(f)=\lambda(f(B(t)));$$

(b) *for* $\hat{\nu}$-a.e. $B \in \hat{\Omega}$,

$$\bar{\lambda}(f)=\bar{\lambda}(f(B(t))).$$

To prove both theorems we view Brownian motion from the perspective of ergodic theory. This approach requires us to be careful about the measure-theoretic foundations. There are many different ways to construct Brownian motion; for instance, at least three topologies, four σ-algebras, and three measure spaces may be brought into the act. Taking the perspective of isomorphism theory, and Rochlin's theory of Lebesgue spaces, seems to clarify some of the issues involved. We have found these references especially helpful: [Lo, Wi, Fr, RSII, Si, Bi1,2].

We will use the following setup for Brownian motion. $C(\mathbb{R}^+)$ denotes the space of continuous real-valued functions, and Ω is that subset such that the value at time 0 is 0. $\mathscr{T}$ denotes the topology on Ω of uniform convergence on compact subsets of $\mathbb{R}$, and $\mathscr{B}$ is the Borel σ-field of $\mathscr{T}$. *Wiener measure* ν will be complete, regular measure on the completion $\mathscr{B}_\nu$, with the correct Brownian motion joint distributions (on finite collections of times in $\mathbb{R}$), and by Brownian motion we mean the triple $(\Omega, \mathscr{B}_\nu, \nu)$. We also call Ω *Wiener space*. The topology $\mathscr{T}$ makes Ω a complete separable metric space. Therefore, for reasons we describe below, it is a well-behaved measure space from the point of view of ergodic theory. Also see [Wi] for related discussion.

DEFINITION. Let $(X, \mathscr{A}, \mu)$, $(Y, \mathscr{B}, \nu)$ be measure spaces with probability measures μ, ν and complete σ-algebras $\mathscr{A}$, $\mathscr{B}$. By a *homomorphism* $\alpha\colon X \to Y$ we mean a function α from $X \backslash N$ to Y with N a null set, such that α is $(\mathscr{A},\mathscr{B})$-measurable and measure-preserving. We say a homomorphism

$f: X \to Y$ is an *isomorphism* (measure-theoretic or (mod 0)-isomorphism) if there is a homomorphism $g: Y \to X$ with $gf(x) = x$ and $fg(y) = y$ for almost all x, y, i.e., except for (measurable) null sets. Equivalently, f is an isomorphism if it is a bimeasurable bijection between $X \backslash N$ and $Y \backslash M$ for null sets N, M. In general a homomorphism may have a non-measurable image; this does not happen for the following class of measure spaces.

DEFINITION. A probability space $(X, \mathscr{A}, \mu)$ (with $\mathscr{A}$ complete) is a *Lebesgue space* if it is isomorphic to the unit interval, or to a subinterval plus a sequence of point masses [Ro, p. 21]. See [Ro, CFS] for the axiomatic characterization.

These two theorems are direct corollaries of Rochlin's results [Ro, pp. 33,25].

THEOREM 4.3. *Let X, Y be Lebesgue spaces and let α: $X \to Y$ be a homomorphism. Then for any set A of positive measure, $\alpha(A)$ is measurable in Y. In particular, $\alpha(X)$ is measurable.*

THEOREM 4.4. *Let X be a separable, complete metric space with Borel sets $\mathscr{B}$, measure μ, and completion $\mathscr{B}_\mu$. The space $(X, \mathscr{B}_\mu, \mu)$ is a Lebesgue space.*

COROLLARY 4.5. *The space $(\Omega, \mathscr{B}_\nu, \nu)$ of Brownian paths is a Lebesgue space.*

DEFINITION. A *transformation* T: $X \to X$ of a probability measure space $(X, \mathscr{A}, \mu)$ is a 1–1, bimeasurable, measure-preserving bijection. A *flow* is a family $S_t: X \to X$ of transformations indexed by $t \in \mathbb{R}$, such that the map $(t, x) \mapsto S_t(x)$ is jointly measurable and an $\mathbb{R}$-action. A *homomorphism* of flows (X, S_t) and (Y, T_t) is a homomorphism α from $X \backslash N$ to Y such that $T_t \alpha = \alpha S_t$ (mod 0), with the requirement that the null set is the same for all t, i.e., that orbits go to complete orbits. Thus the image is a flow invariant set of full outer measure. If (X, S_t) and (Y, T_t) are continuous flows of topological spaces, then a measure-theoretic isomorphism α is *finitary* if it is a homeomorphism off of the flow invariant null sets. This terminology [Fi] is motivated by coding theory [KS]. Let $\{X_i\}_I$ for $I = \mathbb{Z}$ or $\mathbb{R}$ be a family of real-valued random variables with consistent (finite) joint distributions [La]. One models this stochastic process as a measured space of sample paths as follows: let $\Omega = \mathbb{R}^I$, and let $\mathscr{F}$ be the algebra generated by the finite cylinder sets (finite intersections of sets of the form $\{f \in \Omega$: $f(i) \in B$ for $B \in \mathscr{B}(\mathbb{R})\}$ for some $i \in I$). Since the joint distributions are consistent, they determine a measure μ on $\mathscr{F}$; by Kolmogorov's extension theorem [Bi1] this has a unique extension to $\mathscr{F}_0$, the complete σ-algebra generated by $\mathscr{F}$. Points in Ω are regarded as sample paths for the process $\{X_i\}_I$. For $I=\mathbb{Z}$, Ω is a Lebesgue space and thus, for instance, for $\{X_i\}_I$ stationary, we have modeled the process by a measure-preserving transformation of a Lebesgue space. One can then study isomorphisms of this model. For $I=\mathbb{R}$, however, $(\Omega, \mathscr{F}_0, \mu)$ is *not* a Lebesgue space. We will describe the sort of problems that can arise and how they can be resolved for the particular case of Brownian motion.

Now let $\hat{\Omega}=\mathbb{R}^{\mathbb{R}^+}$ and let $\mathscr{F}$ be the finite cylinder algebra. Let μ be the measure on $\mathscr{F}$ given by the joint distributions of Brownian motion [Bi2], and let $\mathscr{F}_0$ be the σ-algebra generated by $\mathscr{F}$. Any (possibly non-measurable) set $S\subseteq\hat{\Omega}$ forms a measure space $(S, \mathscr{A}, \mu^*)$, where $\mathscr{A}$ is the collection of all sets of the form $A\cap S$ for $A\in(\mathscr{F}_0)_\mu$ [Ro, p. 5]. If $S\subseteq\hat{\Omega}$ has outer measure 1 we call the space $(S, \mathscr{A}, \mu^*)$ a *version* of the Brownian motion process. Note that the random variables defined by $X_i=X(i)$ for $i\in\mathbb{R}^+$ and $X\in S$ give the correct joint distributions. And yet the inclusion map $S\hookrightarrow\hat{\Omega}$, which *is* a homomorphism, may not be an isomorphism. This unsettling behavior is in distinct contrast to the discrete time case. Here is a specific example.

Let $S=\Omega$, the continuous paths with $B(0)=0$. It can be proved that $\mu^*(\Omega)=1$ [Bi2, Lo], so one can say that Brownian motion has a *continuous version*. But it also has a discontinuous version. Let $\bar{S}$ be the set of paths $\bar{B}(t)$ such that $\bar{B}(t)=B(t)$ for $B\in S$, except at time $t=\tau_1(B)$, where τ_1 is the first positive time such that $B(\tau_1)=1$, and set $\bar{B}=0$ there. Note that, since $\bar{S}$ gives the same joint distributions as S, it has outer measure 1, and yet $S\cap\bar{S}=0$. Neither set is $(\mathscr{F}_0)_\mu$-measurable.

This trouble is avoided if instead we use the Borel σ-algebra $\mathscr{B}_p$ generated by the product topology $\mathscr{T}_p$ on $\hat{\Omega}$. The point is that $\mathscr{B}_p$ is much larger than $\mathscr{F}_0$—which does not even include, for instance, the $\mathscr{T}_p$-closed singleton $\{B\}$. The reasonable way to extend a measure to all of $\mathscr{T}_p$, and beyond to $\mathscr{B}_p$, is to have the extension be regular. In fact there exists a unique regular extension $\hat{\nu}$ [N, RSII]. We will now see that:

PROPOSITION 4.6. *$(\hat{\Omega}, (\mathscr{B}_p)_{\hat{\nu}}, \hat{\nu})$ is a Lebesgue space and is isomorphic to $(\Omega, \mathscr{B}_\nu, \nu)$.*

Proof. The set $S=\Omega\subseteq\hat{\Omega}$ is $(\mathscr{B}_p)_{\hat{\nu}}$-measurable and of measure 1; see [Bi2, RSII]. The topologies $\mathscr{T}$ on Ω of uniform convergence and $\mathscr{T}_p$ relative to Ω (pointwise convergence) are not the same, but since paths in Ω are continuous, they generate the same Borel σ-algebras. Therefore the inclusion map from Ω to $\hat{\Omega}$ is an isomorphism (mod 0), and since $(\Omega, \mathscr{B}_\nu, \nu)$ is a Lebesgue space, so is $(\hat{\Omega}, (\mathscr{B}_p)_{\hat{\nu}}, \hat{\nu})$. ∎

Remarks. (1) Since a continuous path is determined by its values on a dense set, relative to Ω, $\mathscr{F}_0$ also generates the σ-algebra $\mathscr{B}$.

(2) We remark that $(\bar{S}, (\mathscr{F}_0)_\mu, \mu)$ is also a Lebesgue space, but that $\nu(\bar{S}) = 0$.

(3) If $(\hat{\Omega}, (\mathscr{F}_0)_\mu, \mu)$ had been a Lebesgue space, then the injection $\bar{\alpha}: \bar{S} \subsetneq \hat{\Omega}$ would have been an isomorphism by Rochlin's theorem—and therefore $(\bar{\alpha})^{-1}$ would have been an inverse. And yet $(\bar{\alpha})^{-1} \circ \alpha$ is defined nowhere. This exhibits why for isomorphism theory one prefers to use Lebesgue spaces.

(4) The issue of non-regular extensions also occurs for us in Example 5.1, for finitely additive measures.

We are now ready to describe a certain flow on the Wiener space Ω.

DEFINITION. A *Bernoulli flow* is a flow S_t of a Lebesgue space $(\Omega, \mathscr{A}, \mu)$ such that each time-t map S_t is isomorphic to a Bernoulli shift.

THEOREM 4.7 (Ornstein [O, Sh]). *Two Bernoulli flows of equal entropy are isomorphic.*

Define the map $\Delta_a: \Omega \to \Omega$, for $a > 0$, by $(\Delta_a(B))(t) = (1/\sqrt{a})\, B(at)$. The scaling property of Brownian motion [Lo] says that this map preserves all (finite) joint distributions. Therefore, it preserves the measure ν. Note that $\Delta_a \circ \Delta_b = \Delta_{ab}$, so we have an action of the multiplicative positive reals on Ω. Each Δ_a is a homeomorphism of Ω, and since paths are continuous, the map $(B, a) \to \Delta_a(B)$ is jointly continuous. We define $\tau_s = \Delta_{\exp(s)}$ for $s \in \mathbb{R}$; τ_s is a continuous measure-preserving flow ($\mathbb{R}$-action) on the Lebesgue space (Ω, ν).

We call this the *scaling flow*. We will now see what it is measure theoretically.

Define the map D: $\hat{\Omega} \to \hat{\Omega}$ by $(Df)(t) = (1/\sqrt{t})\, f(t)$, and $E: \hat{\Omega} \to \mathbb{R}^{\mathbb{R}}$ by $(Ef)(t) = f(\exp(t))$. These are homeomorphisms, and $E \circ D$ maps Ω into $C(\mathbb{R})$. Let $\tilde{\Omega} = D(\Omega)$ and $\bar{\Omega} = E(\bar{\Omega})$. We write $\bar{B} = D(B)$ for $B \in \Omega$, and $\tilde{B} = E(D(B))$. Define $\sigma_s: \tilde{\Omega} \to \tilde{\Omega}$ to make this diagram commute:

$$\begin{tikzcd} \Omega \arrow[r, "\tau_s"] \arrow[d, "E \circ D"'] & \Omega \arrow[d, "E \circ D"] \\ \tilde{\Omega} \arrow[r, "\sigma_s"] & \tilde{\Omega} \end{tikzcd}$$

This is a topological and measure-theoretic flow isomorphism, and σ_s is just the shift $(\sigma_s(\tilde{B}))(t) = \tilde{B}(t+s)$ on $\tilde{\Omega}$. Therefore $(\tilde{\Omega}, \sigma_s, \tilde{\nu})$ is a stationary process. It is Gaussian and Markov. One can see directly that it is very weak Bernoulli, hence a Bernoulli flow by [O], and it has infinite entropy. Or one can note that it is mixing, and of kernel type, hence Bernoulli by [OS]. It is also a familiar object in this presentation: the integral of $\tilde{B}(t)$ gives the Ornstein–Uhlenbeck model of Brownian motion, and $\tilde{\Omega}$ is known as the Ornstein–Uhlenbeck velocity process.[3] Also see [Si], where it is called the oscillator process. We conclude:

THEOREM 4.8. *The scaling flow σ_s on $(\Omega, \mathscr{B}_\nu, \nu)$ is a Bernoulli flow of infinite entropy. So is the scaling flow on $(\hat{\Omega}, (\mathscr{B}_p)_{\hat{\nu}}, \hat{\nu})$, and the inclusion map of Ω into $\hat{\Omega}$ is a measure-theoretic but not a finitary isomorphism.*

A Bernoulli flow is, in particular, ergodic and Kolmogorov. It is these properties which will be used in the proof of Theorems 4.1 and 4.2.

The next lemma provides the main technical tools needed in the proofs of the theorems. We write $\mathscr{T}(\Omega)$ for the topology of uniform convergence on compacts, and $\mathscr{T}(K)$ for the weak-$*$ topology on the norm-unit ball K of $L^\infty(\mathbb{R})$. $\mathscr{B}(\Omega)$ and $\mathscr{B}(K)$ denote the Borel sets. Lebesgue measure on $\mathbb{R}$ is denoted m and $\mathscr{B}_m$ is the completed σ-algebra (Lebesgue measurable sets) on $\mathbb{R}$. $\mathscr{B}_\nu$ is $(\mathscr{B}(\Omega))_\nu$, the completion of Wiener measure.

LEMMA 4.9. *Let $f \in L^\infty(\mathbb{R})$.*

(1) *For ν-a.e. B in Ω, $f(B(t))$ is a well-defined element of $L^\infty(\mathbb{R}^+)$.*

(2) *The map $(B, t) \mapsto f(B(t))$ from $\Omega \times \mathbb{R}^+$ to $\mathbb{R}$ is measurable from the completed σ-algebra*

$$(\mathscr{B}(\Omega) \times \mathscr{B}(\mathbb{R}^+))_{\nu \times m} \text{ to } \mathscr{B}(\mathbb{R}).$$

(3) *For an interval $I \subseteq \mathbb{R}$,*

$$\int_\Omega \left(\int_I f(B(t))\, dt \right) d\nu(B) = \int_I \left(\int_\Omega f(B(t))\, d\nu(B) \right) dt.$$

(4) *The map $B \mapsto f(B(t))$ from $(\Omega, \mathscr{T}(\Omega))$ to $(K, \mathscr{T}(K))$ is continuous off a set $N \subseteq \Omega$ of ν-measure 0. (This is also true for the pointwise topology on Ω.)*

Proof. We first prove:

(a) The map $(B, t) \mapsto B(t)$ from $\Omega \times \mathbb{R}^+$ to $\mathbb{R}$ is Borel measurable, i.e., it is $(\mathscr{B}(\Omega) \times \mathscr{B}(\mathbb{R}^+), \mathscr{B}(\mathbb{R}))$-measurable. The proof of (a) follows [Bi2, p. 448]. For $B \in \Omega$, define $B^n(t) = B(k2^{-n})$ for all $t \in [k2^{-n}, (k+1)2^{-n})$, for $k = 1, 2, \ldots$. The mapping $(B, t) \mapsto B^n(t)$ from $\Omega \times \mathbb{R}^+$

[3] See Cox and Miller, "The Theory of Stochastic Processes," p. 229, Chapman & Hall, London, 1972.

to $\mathbb{R}$ is Borel measurable: For instance, for an interval $I \subseteq \mathbb{R}$, $\{(B, t): B^n(t) \in I\} = \bigcup_k(\{B: B(k2^{-n}) \in I\} \times [k, k+1)\, 2^{-n})$ which is Borel measurable. Since paths in Ω are continuous, the mapping $(B, t) \mapsto B^n(t)$ converges pointwise to the mapping $(B, t) \mapsto B(t)$ for all t (and all B). Therefore, the limit is Borel and (a) is proved.

(b) For $f: \mathbb{R} \to \mathbb{R}$ Borel measurable, the map $(B, t) \mapsto f(B(t))$ is Borel measurable from $\Omega \times \mathbb{R}^+$ to $\mathbb{R}$.

(c) Let $f(x) = \chi_A$ for $A \in \mathscr{B}(\mathbb{R})$. Then

$$\int_\Omega f(B(1))\, dv(B) = \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2}\, dx.$$

This is because the left-hand side is equal to $v\{B: B(1) \in A\}$, corresponding to a cylinder set in $\mathbb{R}^{\mathbb{R}^+}$ whose μ-measure we know from the construction of μ and v.

(d) Let $N = \mathscr{B}_m$ be a Lebesgue measurable null set. Define $\alpha: (B, t) \mapsto B(t) \in \mathbb{R}$. We claim that $\alpha^{-1}(N)$ is of $(v \times m)$-outer measure zero. $N \subseteq \tilde{N}$, a Borel measurable m-null set. Since α is Borel by (a), we can apply Fubini's theorem:

$$v \times m(\alpha^{-1}(\tilde{N})) = \int_{\mathbb{R}} \left(\int_\Omega \chi_{\tilde{N}} \alpha(B, t)\, dv(B) \right) dt.$$

The inner integral is, for instance, for time $t = 1$, equal to $\int_\Omega \chi_{\tilde{N}}(B(1))\, dv(B))$, which is zero by (c). Therefore $\alpha^{-1}(N) \subseteq \alpha^{-1}(\tilde{N})$ is null in the $v \times m$-completion of $\mathscr{B}(\Omega) \times \mathscr{B}(\mathbb{R})$.

(e) Now consider f and $\hat{f}$ in the same L^∞-equivalence class, that is, equal off of some m-null set $N \subseteq \mathbb{R}$. $f(B(t)) = \hat{f}(B(t))$ for all (B, t) not in $\alpha^{-1}(N) \subseteq \Omega \times \mathbb{R}^+$. This has $v \times m$-measure zero by (d), and so for v-a.e. B in Ω by Fubini's theorem, $m\{t: f(B(t)) \neq \hat{f}(B(t))\} = 0$, i.e., $f(B(t))$ and $\hat{f}(B(t))$ are in the same $L^\infty(\mathbb{R}^+)$ equivalence class, which proves (1).

(f) Next let $f \in L^\infty$. We can find f_n continuous such that $f_n(x) \to f(x)$ except for $x \in N$ with $m(N) = 0$. Since $B \in \Omega$ is continuous, $f_n(B(t)) \to f(B(t))$ for all $t \in B^{-1}(N)$, or for all $(B, t) \in \alpha^{-1}(N)$. By (d) this has $v \times m$-measure zero. Therefore, $(B, t) \mapsto f_n(B(t))$ converges $v \times m$-almost everywhere, and the limit function $(B, t) \mapsto f(B(t))$ is measurable with respect to the completion, proving (2). Now (3) follows by Fubini's theorem for completed product measure [Roy]. To prove (4) we need to show that for any $g \in L^1(\mathbb{R})$, if $B \to B_n$ either uniformly on compacts or pointwise, then for $f \in L^\infty$,

$$\int g(t)\, f(B_n(t))\, dt \to \int g(t)\, f(B(t))\, dt.$$

First assume f is continuous. In that case,

$$g(t)\, f(B_n(t)) \to g(t)\, f(B(t)) \qquad \text{as} \quad n \to \infty$$

almost surely in t. Therefore, by the Lebesgue dominated convergence theorem, the integrals converge. Next, let f be an arbitrary L^∞ function, and let $f_m \to f$ be a sequence of continuous functions converging pointwise except on a set $N \subseteq \mathbb{R}$ with $m(N)=0$. We claim that $\int g(t)\, f_m(B(t))\, dt \to \int g(t)\, f(B(t))\, dt$ for ν-almost all B. But $f_m(B(t)) \to f(B(t))$ except for $t \in B^{-1}(N)$. Let $S = \{B \in \Omega : m(B^{-1}(N) \neq 0)\}$; $\nu(S)=0$ by (e), so again by the dominated convergence theorem we have the result. Therefore, for m and n large ($\sim^\varepsilon$ means "within ε"),

$$\int g(t)\, f(B_n(t))\, dt \overset{\varepsilon}{\sim} \int g(t)\, f_m(B_n(t))\, dt \overset{\varepsilon}{\sim} \int g(t)\, f_m(B(t))\, dt$$

$$\overset{\varepsilon}{\sim} \int g(t)\, f(B(t))\, dt,$$

for all B and B_n not in S. That is, the map $B \mapsto \int g(t)\, f(B(t))\, dt$ is continuous from $\Omega \setminus S$ to $\mathbb{R}$, with the same set S for all $g \in L^1$, so we are done.

An alternate proof of (3) can be constructed along lines similar to the proof of the Feynman–Kăc formula in [RSII].

First Proof of Theorem 4.1. Consider the function $\beta : \Omega \to \mathbb{R}$ defined by $\beta(B) = \chi_A(B(1))$ for A Borel measurable in $\mathbb{R}$. By (c) in the proof of the lemma, β has expected value $\int_\Omega \beta(B)\, d\nu(B) = (1/\sqrt{2\pi}) \int_A e^{-x^2/2}\, dx$. For $A \subseteq \mathbb{R}$ Lebesgue measurable, β is measurable by part (1) of the lemma, or directly by approximating with Borel sets, and we have the same formula. The same is true for $\alpha : \hat{\Omega} \to \mathbb{R}$ for the Ornstein–Uhlenbeck velocity process $(\hat{\Omega}, \sigma_s)$. This is a Bernoulli flow and so is, in particular, ergodic. Therefore the time and space averages are equal by the Birkhoff ergodic theorem, i.e.,

$$\lim \frac{1}{t} \int_0^t \alpha(\sigma_s(\hat{B}))\, ds = \int_\Omega \alpha(B)\, d\nu(\hat{B})$$

for $\hat{\nu}$-a.e. $\hat{B} \in \hat{\Omega}$, and therefore for λ exponentially and convolution-invariant,

$$\lambda\left(\chi_A\left(\frac{1}{\sqrt{t}} B(t)\right)\right) = \lambda(\chi_A(\bar{B}(t))) = \lambda(\chi_A(\hat{B}(t)))$$

by exponential-invariance, and since such a λ agrees with Cesaro limits when they exist, this equals, for $\hat{\nu}$-a.e. $\hat{B}$, or equivalently for ν-a.e. B,

$$\lim \frac{1}{t} \int_0^t \chi_A(\hat{B}(s))\, ds = \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2}\, dx. \quad \blacksquare$$

Second Proof of Theorem 4.1 (using assumption (b)). Let λ be a measure-linear mean which is weighted at $+\infty$. It could, for instance, be translation-invariant. Define $D\colon \Omega \to \bar{\Omega}$ as before, with $\bar{\nu} = D(\nu)$. Define $w\colon \Omega \to K$ by $(w(B))(t) = \chi_A(\bar{B}(t))$. This is continuous, off a ν-null set, as a map to the weak-$*$ topology by (4) of Lemma 4.9. Therefore it is ν-measurable and so ν pushes forward to $w(\nu) = \mu$; this is in $\mathcal{M}_1(K)$ and must have a barycenter $b(\mu)$ in K. For each time t the expected value of the process (K, μ) is $(1/\sqrt{2\pi}) \int_A e^{-x^2/2}\, dx = a \in \mathbb{R}$ by the same reasoning as in the first proof. Now for any $g \in L^1(\mathbb{R})$, we have

$$\begin{aligned}\int_K \left(\int_{\mathbb{R}} h(t)\, g(t)\, dt \right) d\mu(h) &= \int_{\mathbb{R}} g(t) \left(\int_K h(t)\, d\mu(h) \right) dt \\ &= \int_{\mathbb{R}} a g(t)\, dt.\end{aligned}$$

The exchange of integrals is justified by (3) of the lemma. Therefore the function which is constantly equal to the expected value a is the only function in L^∞ which gives this result when tested against every g. We conclude that this is the barycenter $b(\mu)$.

So far we know that

$$\begin{aligned}\int_\Omega \lambda \left(\chi_A \frac{1}{\sqrt{t}} B(t) \right) d\nu(B) &= \int_{\bar{\Omega}} \lambda(\chi_A(\bar{B}))\, d\bar{\nu}(\bar{B})) \\ &= \int_K \lambda(h)\, d\mu(h),\end{aligned}$$

which equals $\lambda(b\mu) = a = (1/\sqrt{2\pi}) \int_A \lambda(h)\, d\mu(h)$ by measure-linearity and the fact that λ is a mean. So the expected value is correct, and if we can show that we are integrating a constant function we will be done. We need the next lemma, and provide a new proof which ties back in nicely with the ergodic theory of the first proof of Theorem 4.1.

DEFINITION. For a family $\{\mathcal{A}_i\}_I$ of σ-algebras, $\bigvee_I \mathcal{A}_i$ denotes the smallest σ-algebra containing each, and $\bigwedge_I \mathcal{A}_i$ the largest contained in each. Define $\mathcal{B}_t = \bigcup_{B \in \mathcal{B}(\mathbb{R})} \{f \in \Omega\colon f(t) \in B,$ some $B \in \mathcal{B}(\mathbb{R})\}$ which is a sub-σ-algebra of $\mathcal{B}(\Omega)$. Let $\mathcal{B}_{\geqslant r} = \bigvee_{t \geqslant r} \mathcal{B}_t$, and $\mathcal{B}_{+\infty} = \bigwedge_{r=0}^{\infty} \mathcal{B}_{\geqslant r}$ so $\mathcal{B}_{+\infty} = \bigcap \mathcal{B}_{\geqslant r} = \bigwedge_r \bigvee_{t \geqslant r} \mathcal{B}_t$. We call this the *tail algebra of* Ω.

LEMMA 4.10 (Blumenthal 0–1 Law). *If* $E \in \mathscr{B}_{+\infty}$, *then* $\nu(E)=0$ *or* 1.

Proof. Define τ_s, $\Delta_s\colon \Omega \to \Omega$ as in the first proof, and set $\mathscr{B}^s = \Delta_s(\mathscr{B}_1)$. One sees that $\mathscr{B}^s = \mathscr{B}_s$. Therefore, $\mathscr{B}_{\geqslant r} = \bigvee_{s \geqslant r} \Delta_s(\mathscr{B}_1)$ and $\mathscr{B}_{+\infty} = \bigcap \bigvee \Delta_s(\mathscr{B}_1) = \bigcap \bigvee \tau_s(\mathscr{B}_1)$. In other words, the time-shift tail algebra for the (non-stationary) process of Brownian motion equals the scaling flow tail algebra generated by the σ-algebra $\mathscr{B}_1$ of the time-one cylinders.

Let I denote the interval $[0, 1]$, and set $\mathscr{B}_I = \bigvee_{t \in I} \mathscr{B}_t$. We see that $\mathscr{B}_{+\infty} = \bigwedge \bigvee \tau_1^i \mathscr{B}_I$. The transformation (Ω, τ_1) is Bernoulli since the flow is a Bernoulli flow, hence it is Kolmogorov [CFS, O], and we would like to conclude that the tail algebra $\mathscr{B}_{+\infty}$ is therefore trivial. This is immediate for any factors generated by finite subalgebras; see [CFS, Sect. 10.8, Theorem 1]. But $\mathscr{B}_I$ is not of this form so we need to be more careful. One sees easily from the form of the Gaussian Markov process (Ω, τ_1) that for $A \in \mathscr{B}_I$, given $\varepsilon > 0$, there exists n such that for any $B \in \mathscr{B}_I$, $\nu(\tau_1^i B \mid A) - \nu(B) < \varepsilon$ for $i > n$. Therefore $\mathscr{B}_I$ is independent of $\mathscr{B}_{+\infty}$, and the triviality of $\mathscr{B}_{+\infty}$ follows in a familiar way. ∎

This completes the second proof of Theorem 4.1. ∎

Remark 4.11. For non-amenable groups the Blumenthal 0–1 law does not hold—see the comments at the end of Section V. In $\mathbb{R}^n$, however, the tail algebra is still trivial, even though Brownian motion is not recurrent for $n \geqslant 3$. Therefore we have:

THEOREM 4.12 (Pathwise CLT for $\mathbb{R}^n$). *The statement is the analogue of that for* $n = 1$.

Proof. Brownian motion in $\mathbb{R}^n$ is 1-dimensional Brownian motion in each coordinate. Therefore one has the same scaling property, and so the scaling flow, which is a product of Bernoulli flows, is Bernoulli. Again, the tail algebra is trivial. So both the first and second proofs go through just as before. ∎

Proof of Theorem 4.2. Let λ be a measure-linear mean which is exp- and convolution-invariant (which exists by Theorem 3.8). It is weighted at $+\infty$. Let $\bar{\lambda}$ denote the symmetrized version. This is also invariant under Cesaro-type operators in the following sense: let $\psi \in P_1(\mathbb{R})$, and define

$$A_\psi^1\colon L^\infty(\mathbb{R}) \to L^\infty(\mathbb{R}^+)$$

by

$$f \mapsto \frac{1}{t} \int \psi\left(\frac{x}{t}\right) f(x)\, dx.$$

We have, then,

$$\bar{\lambda}(f) = \lambda(A^1_\psi(f)).$$

Now consider, for $f \in K \subseteq L^\infty(\mathbb{R})$, the map $\alpha(B) = f(B(t))$ from Ω to K. By Lemma 4.9, this is ν-a.s. continuous to the weak-$*$ topology of K. In particular, it is ν-measurable, hence it maps ν to a measure $\mu \in \mathscr{M}_1(K)$. What is its barycenter?

Define $\psi(x) = (1/\sqrt{2\pi})\, e^{-x^2/2}$, which is in $P_1(\mathbb{R})$. Define $\alpha_t : \Omega \to \mathbb{R}$ by $\alpha_t(B) = f(B(t))$. This is measurable by the lemma, and its expected value, as in (c) of the lemma, is, for $t = 1$, $(1/\sqrt{2\pi}) \int f(x)\, e^{-x^2/2}\, dx$. By the scaling property of Brownian motion, the expected value of $\alpha_t(B)$ is

$$\int_\Omega \alpha_t(B)\, d\nu(B) = \int \frac{1}{\sqrt{t}}\, \psi\left(\frac{x}{t}\right) f(x)\, dx.$$

Call this function $\hat{f}(t)$. We claim that $\hat{f}(t)$ is the barycenter $b(\mu)$. We need to show that, for $g \in L^1(\mathbb{R})$,

$$\int_{\mathbb{R}} \hat{f}(t)\, g(t)\, dt = \int_\Omega \left(\int_{\mathbb{R}} f(B(t))\, g(t)\, dt\right) d\nu(B).$$

By Fubini's theorem—the use of which is justified by the lemma—the right-hand side is equal to $\int_{\mathbb{R}} (\int_\Omega f(B(t))\, d\nu(B))\, g(t)\, dt$, and this is exactly $\int \hat{f}(t)\, g(t)\, dt$ as we hoped it would be.

Now we prove the theorem. Define $\psi(x)$ as above, and define $\psi_t(x) = (1/t)\, \psi(x/t)$, and set $h(t) = \int \psi_t(x) f(x)\, dx$. By measure-linearity,

$$\int \lambda(f(B(t)))\, d\nu(B) = b(\lambda u) = \lambda(b\mu) = \lambda\left(\int \psi_{\sqrt{t}}(x) f(x)\, dx\right) = \lambda(h(\sqrt{t})).$$

By $c(x) = x^2$-invariance (Theorem 2.6), this equals $\lambda(h(t))$. Then by Cesaro-type invariance,

$$\lambda(h(t)) = \lambda\left(\frac{1}{t}\int \psi\left(\frac{x}{t}\right) f(x)\, dx\right) = \bar{\lambda}(f).$$

Therefore the expected value of $\lambda(f(B))$ is just what we wanted it to be. Furthermore, by the lemma, and since λ is universally measurable, the map $B \mapsto \lambda(f(B))$ from Ω to $\mathbb{R}$ is measurable. Its values only depend on the infinite future of $B(t)$, and hence they define sets measurable with respect to the tail algebra of Brownian motion. Therefore by the Blumenthal 0–1 law, the function is a.s. (ν) constant, and since we know its expected value, we are done. The second part of the theorem follows immediately. ∎

Now we extend this idea to $\mathbb{R}^n$. Given a mean λ on $L^\infty(\mathbb{R})$ which is weighted at $+\infty$, define an associated mean $\bar{\lambda}$ on $L^\infty(\mathbb{R}^n)$ by

$$\bar{\lambda}(\hat{f}) = \lambda(f(r))$$

for

$$\hat{f}(r) = \int_{S_r} f(x)\, dm_r(x),$$

whose m_r is normalized Lebesgue measure on the $(n-1)$-sphere of radius r centered at the origin. If λ is invariant under Cesaro-type averages, then $\bar{\lambda}$ is invariant under the obvious $\mathbb{R}^n$-analogue, i.e.,

$$\bar{\lambda}(f) = \lambda\left(\int_{\mathbb{R}^n} f(x)\, \psi_t(x)\, dx\right) \qquad \text{for} \quad \psi(x) \in P_1(\mathbb{R}^n)$$

and $\psi_t(x) = (1/t^n)\ \psi(x/t)$. This implies that, e.g., $\bar{\lambda}(f) = \lambda((1/|B_r|) \int_{B_r} f(x)\, dx)$ for B_r the ball of radius r and $|B_r|$ its volume, and also dilation-invariance: $\bar{\lambda}(f) = \lambda(f(rx))\ \forall r > 0$. As is the case for $\mathbb{R}$, this condition implies translation-invariance of $\bar{\lambda}$. Things are quite different for a non-amenable manifold or group: see the Conclusion in Section V.

THEOREM 4.13 (Monte Carlo Theorem for $\mathbb{R}^n$). *With the assumption of Martin's axiom, there exists a translation-invariant mean λ on $L^\infty(\mathbb{R}^+)$, and an associated translation-invariant mean $\bar{\lambda}$ on $\mathbb{R}^n$, such that for any $f \in L^\infty(\mathbb{R}^n)$, for ν-a.e. $B(t)$ (Brownian motion in $\mathbb{R}^n$),*

$$\bar{\lambda}(f) = \lambda(f(B(t))).$$

Proof. Let λ be measure-linear and exponentially and convolution-invariant. The proof, as it was for the extension of the PCLT to $\mathbb{R}^n$, is just like that for dimension one. ∎

V. COUNTEREXAMPLES AND QUESTIONS

EXAMPLE 5.1. A translation-invariant mean on $L^\infty(\mathbb{R})$ which is not convolution-invariant.

Rudin [Ru2] constructs such an example on any non-discrete locally compact infinite amenable group. For the circle S^1, his construction roughly goes as follows. Let $A \subseteq S^1$ be a dense, open set of measure $< \varepsilon$. Let $f = 1 - \chi_A$. There is a translation-invariant ideal J in the Banach algebra $L^\infty(S^1)$ which contains f and is annihilated by a complex homomorphism. There is an associated invariant mean λ such that $\lambda f = 0$, and yet for appropriately chosen $\phi \in P_1(S^1)$, $\phi * f(x) \geqslant \frac{1}{2}$ for all x, which proves that $\lambda(f) \neq \lambda(\phi * f)$. Now, λ defines a finitely additive measure μ on the Lebesgue measurable subsets of S^1. Therefore μ cannot be regular, since on a compact domain a finitely additive regular measure is countably additive [DS], and it would be Lebesgue measure. So this also provides a good example of a non-regular measure. Also see [Wa] on regularity.

Such a phenomenon *cannot* happen on $C(S^1)$, since these are the uniformly continuous functions, and there, $\lambda(f) = \lambda(f * \phi)$ [Gr]. The extension of a mean from $C(X)$ to $L^\infty(X)$ for a space X is via the Hahn–Banach theorem [DS] and is non-unique for "non-regular" extensions as this example shows. What should be the proper definition and role of regularity for a mean?

EXAMPLE 5.2. The same property but on $\mathbb{R}$.

This example [Ra] is quite different and works on $C(\mathbb{R})$: Raimi produces $f_0 \in C(\mathbb{R})$ such that $\lambda(f_0) \neq \lambda(f_0 * \phi)$ for any $\phi \in P_1(\mathbb{R})$. The function f_0 is a piecewise linear sawtooth with values 1 and -1 at increasingly closely spaced points. His construction of such a λ uses rational approximations of real numbers in an interesting way.

Means and the Ergodic Theorem. For $c \in \mathscr{C}_{\mathscr{O}}$, we say the averaging method $\lim \phi * (f \circ c)$ is *Birkhoff* if it gives the result of the Birkhoff ergodic theorem (for bounded functions). That is, let μ be a measure on $K \subseteq L^\infty$ which represents an ergodic flow; what we require is that the limit exists for μ-a.e. f and gives the expected value. (We have throughout this paper restricted our attention to bounded f; one should try to extend it—see [M, H-J].) We say a mean λ on L^∞ or l^∞ is Birkhoff if it gives the expected value for μ-a.e. f in K. We have shown (Example 3.5) that any measure-linear mean which is invariant is Birkhoff. Birkhoff's ergodic theorem says exactly that for $c(x) = \exp(x)$, the averaging method satisfies this property (since it is just the Cesaro average). But how little is actually needed? We make this wild guess:

Conjecture 5.3. For any $c \in \mathscr{C}_{\mathscr{O}}$, $\lim \phi * (f \circ c)$ is Birkhoff. The reasoning is that the finite measure space on which $\mathbb{R}$ is acting should force enough "almost periodicity" or recurrence that things average out.

EXAMPLE 5.4. A mean on $L^\infty(\mathbb{R})$ which is translation-invariant but not Birkhoff.

Consider the Rudin mean of Example 5.1, except on $\mathbb{R}$ instead of S^1. Let (S^1, m, T_t) be the flow of rotation by angle t on S^1 with Lebesgue measure m. Let f on S^1 be as in Example 5.1. For every x, $\lambda(f(T_t x)) = 0$ and yet the expected value is $\geqslant \varepsilon$.

PROPOSITION 5.5. *Let $\Omega = \prod_0^\infty \{1, 2,..., k\}$ be a Bernoulli shift with μ an independent infinite product measure. Then any mean λ on l^∞ which is measurable, linear, and weighted at $+\infty$ is Birkhoff for all such shifts, for a cylinder-set observable.*

Sketch of Proof. For measures $(1/k, 1/k,...)$ one uses the Kolmogorov 0–1 law, linearity and measurability, and the fact that permutations are automorphisms to see one gets the correct value (for a cylinder-set observable). This extends to rational probabilities. Considering the shift as a factor of the infinite-entropy Bernoulli shift $\prod_1^\infty [0, 1]$, with Lebesgue measure on $[0, 1]$, one sees that generic-point sequences of 0's and 1's can be nested as the probabilities change. Thus we can interpolate from the rational case. This proof came about in conversations with Jon Aaronson, Eli Glasner, and Hillel Furstenberg. It indicates that the independent case is quite special. Probably (since the O–U velocity process is close to independent) any such mean gives the PCLT with, e.g., no translation-invariance required.

EXAMPLE 5.6. A mean on l^∞ which is invariant but not measurable with respect to the product measure $\mu = \bigotimes_1^\infty (\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1)$ on $\prod_1^\infty \{0, 1\} \subseteq l^\infty$.

This is an example due to Sierpinski; see [H-J, p. 72].

EXAMPLE 5.7. A mean on $L^\infty(\mathbb{R})$ which is convolution-invariant but not Birkhoff.

We extend the idea in Example 5.6 to L^∞ and achieve this. It gives another example of Example 5.4. Cesaro averages often exist, in the sense that (on l^∞) they exist μ-a.e. for any stationary μ. But (this was pointed out to me by Mike Boyle) they only exist on a topologically small set, a set of first category [Ox]. The same is true for any countable family of averaging methods given by $\mathscr{G} \subseteq \mathscr{C}_\mathscr{O}$. Does there exist $\mathscr{G}$ (consistent) such that the mean value is precisely determined on a non-residual set? (Even our exponentially invariant mean is not nearly strong enough to get this.)

EXAMPLE 5.8. A mean which is convolution- (or translation-) but not dilation-invariant.

This follows Example 2.8. Let $c(x) = [\exp](x)$ as in the example. By Theorem 2.2 there exists a mean λ which is $c(x)$- and convolution-invariant. But for f as in the remark, $f(\exp(x))$ is periodic and piecewise continuous. This implies that its mean value is fixed for any translation-invariant mean. The same number would be the mean value of f, if the mean were dilation-invariant, since f is dilation-periodic. Yet, $f([\exp](x))$

is also periodic and has a *different* mean value. This example exhibits some of the rigidity involved in choosing a consistent family.

PROPOSITION 5.9. (a) *Any Cesaro-type averaging method diverges a.s.* (ν) *for the Pathwise CLT.* (b) *Abstract extensions (see the Introduction) also fail in general.*

Proof. Given $\psi \in P^1(\mathbb{R}^+)$ we want to show that for ν-a.e. B, the lim sup and lim inf of $(1/s) \int \psi(t/s)\, \chi_I(1/\sqrt{t})\, \chi_I(B(t))\, dt$ disagree. This is the same as applying a convolution $\phi*(\chi_I(\hat{B}(t)))$ to some $\hat{B}$ in the 0–U velocity process $\hat{\Omega}$, which clearly continues to oscillate. (In fact the paths $\phi * (\chi_I(\hat{B}(t)))$ form a (finitary) factor of a Bernoulli flow—and so this is also a Bernoulli flow.) To prove (b) we need to show that there is a weak-$*$ limit point of the net of means $\alpha_t\colon L^\infty \to \mathbb{R}$ defined by $\alpha_t(g) = (\phi * g)(t)$, which diverges. Let $t \in \mathbb{N}$, and then take some (non-measurable) ultrafilter limit applied to these α_t; this gives the result.

EXAMPLE 5.10. A mean on L^∞ which is exp- but not convolution-invariant.

This can be constructed using an idea similar to Rudin's. However, here also the failure is essentially local. It would be much nicer to have a positive answer to:

Question 5.11. Is there an exp-invariant mean on L^∞ which is not translation-invariant?

It is easy to produce exp-invariant means, and it is easy to produce such an example on $C(\mathbb{R})$. This is an important question and perhaps not too hard.

Question 5.12. Is there a convolution-invariant mean which is not Birkhoff?

I think I have an example but have to check it further. If not, there would be an easy (too easy) proof of Conjecture 5.3 via Lemma 2.5.

Conclusions (1) One direction for further investigation is to find out more about the partial ordering on $\mathscr{C}_{\mathcal{O}}$. How large can a directed subset $\mathscr{G}$ be—does there exist a maximal family? One might try to use Zorn's lemma: can one find an upper bound for a tower? To do this, one would like to "diagonalize" within $\mathscr{C}_{\mathcal{O}}$. Does a maximal family $\mathscr{G}$ have, for any $f \in L^\infty$, some element $c \in \mathscr{G}$ such that $\lim \phi * (f \circ c)$ exists? Compare Proposition 2.7. Are there many maximal families, and if so, how are they classified? Are there natural conditions on higher-order derivatives which specify a maximal family? Then one would have a candidate for a unique mean value on L^∞. If so, is it still universally measurable? (Does the existence of a measure-linear mean depend on Martin's axiom?) If the program we are suggesting cannot be pushed so far as to get a unique natural mean, then what are the obstructions and how can they be formulated in a precise way?

(2) There are several directions one can go in extending the results of Section IV. Random walks in $\mathbb{R}$ approximate Brownian motion under scaling in the sense that the cumulative distributions converge. This has an attractive pathwise version: under scaling, the polygonal path of a random walk is a generic point, in the ergodic theoretic sense, for Brownian motion. The same result holds in $\mathbb{R}^n$. This enables us to prove pathwise versions of Donsker's theorem, and of the pathwise CLT for random walks. These ideas will be carefully developed in a second paper.

Measure-linear means have special additional interest for more general locally compact, non-amenable groups and their actions. For a probability measure μ on a group G, the convolution $\mu * \mu * \mu$, for instance, is the distribution of an independent random walk on G, with distribution μ, after three steps. Given X_i an independent sequence of G-valued random variables with distribution μ, and $f \in C(G)$, a natural definition of boundary value for f is $\lambda(f(X_1), f(X_1 X_2), f(X_1 X_2 X_3), ...) = g(X_1 X_2 ...)$ on the boundary point $X_1 X_2 ...$ [F, KV], where λ is a measure-linear, Cesaro-invariant mean on $l^\infty(\mathbb{N})$. By measure-linearity, when one integrates this over the boundary measure, one gets exactly $\lambda(\mu(f), \mu * \mu(f), ...)$ which can be interpreted as the average of f as seen from the identity. For amenable G this average-from-a-point is constant; for non-amenable G it varies as a μ-harmonic function. Thus we have a projection from $L^\infty(G)$ to the μ-harmonic functions, which depends on the choice of the mean λ, but which is then consistent with the corresponding boundary values. This correspondence is just the Monte-Carlo method applied to general G. One sees that the Blumenthal 0–1 Law is what fails for the non-amenable case (and that in fact the measurable tail algebra sets are the measurable sets of the boundary). One can do this for a group action on a finite measure space (X, m). Here the correspondence is strikingly familiar: for $f \in L^\infty(X, m)$,

$$\frac{1}{n} \sum_{i=1}^{n} \underbrace{\mu * \cdots * \mu}_{i \text{ times}} f(x) \to \hat{f}(x)$$

by Kakutani and Yosida's Ergodic Theorem [Y, p. 388], and on the other hand, for a.e. sequence $X_1 X_2, ...,$ and a.e. x,

$$\frac{1}{n} \sum_{i=1}^{n} f(X_1 X_2 \cdots X_i(x)) \to h(X_1 X_2 \cdots)(x),$$

by the Random Ergodic Theorem of Pitt, Ulam, von Neumann, and

Kakutani [Ha], [Ka]. Integrated over the boundary, this gives the μ-harmonic function $\hat{f}$. If the group is amenable, this is a.s. constant on the ergodic components of the action. But what about for other specific cases? These theorems say that for an action on a finite measure space, the measure-linear mean collapses down to a simple Cesaro average of a special type. Ornstein and Weiss have proved a pointwise ergodic theorem for actions of amenable groups of polynomial growth. Are averaging sets of the form supp($\mu * \cdots * \mu$) "Følner" sets for, e.g., symmetric μ, even for the non-amenable cases? There are many fascinating questions here. The general philosophy is that, although "average" makes no sense for general groups, the harmonic projection does—and this is just the average as seen from a point. When one factors onto other groups, or onto infinite measure spaces (via an action), one may again recover a constant—that is, a global sense of average. Then the question becomes, how little can one get away with? Sometimes a Cesaro-type average is enough, and this is the real content of the various ergodic theorems. This work will be treated in a third paper. I am indebted to B. Weiss and H. Furstenberg in particular for a conversation concerning the harmonic projection, in which Weiss observed that the average as seen from a point *can* vary, and Furstenberg then noticed that it should vary harmonically.

Comments on References

Precursors to this paper in the general realm of Banach limits, invariant means, and summability methods are far too many to list here, but see [Gr]. McShane, Warfield, and Warfield in [MWW] did produce a mean on l^∞ which is Cesaro-invariant and iteration-invariant (where iteration is the analogue, for sequences, of dilation). This paper was influenced in places by their algebraic approach.

We have, throughout this paper, been asking a perhaps unanswerable question: what are the strongest possible natural axioms one can require of a mean? A deeper fundamental problem is to investigate the same open-ended question for a conditional probability $p(A, B)$ defined on pairs of subsets of $\mathbb{R}$. This has been studied by Parikh and Parnes [PP1, PP2] using the tools of non-standard analysis. Indeed, this is one case where it is hard to imagine an approach which does not make use of non-standard analysis. Their axioms include dilation-invariance (but see the remark following Theorem 1.2 of [PP2]). How much further can their axioms be strengthened? Is there a mutual extension of their results and those of this paper?

The general idea that non-amenability is intimately related to the existence of a non-trivial boundary at infinity, and to the existence of nonconstant bounded harmonic functions, goes back at least to the 1962–1963 papers of Furstenberg. Garnett [Ga1, Ga2] discusses a Monte Carlo method for Brownian motion on foliations of a compact manifold, making use the Kakutani–Yosida theorem (this is where I learned of that tool). Since (hopefully) foliations are to a compact manifold as orbits are to a finite measure space, this is, from our point of view, much like taking the harmonic projection of the group action. The boundary of the action is just the product of the group boundary and the measure space, and the boundary values are a function on this boundary. Zimmer [Zi] has made a similar definition for the foliation case. Sometimes the action is such that the boundary squashes down to something much smaller. Much interesting work remains on what happens for specific examples.

Lyons and Sullivan [LS] have just published a paper in which they describe a harmonic projection for Brownian motion on an infinite Riemannian manifold. They do not have available the tool of measure-linear means and so do not discuss boundary values. We use random walks (see [F, KV, LE] for more recent material) because the boundary theory is easier there; ideally, this theory should be worked out simultaneously for random walks and Brownian motion, and on trees, groups, and Riemannian manifolds. One can think of geometrical definitions of amenability for a manifold (are balls "Følner sets"?). When does this agree with the algebraic definition? There are many remaining unanswered questions here as well.

Acknowledgments

This project really began when I was an undergraduate at the University of California at Berkeley. It spanned graduate school in Seattle, postdoctoral appointments at the Hebrew University of Jerusalem, Tel Aviv University, and Northwestern University, as well as visits to MSRI Berkeley and Warwick University; I wish to thank these institutions and their staffs for their support. Without the interest, inspiration, and encouragement of the following teachers and colleagues, this paper would never have been written: listed in chronological order, David Salisbury, Andrew Ogg, Robert Gulliver, Brian Marcus, Robert Moore, Douglas Lind (my doctoral adviser—see [Fi]), Eli Glasner, Meir Smorodinsky, Jon Aaronson, and Alexandra Bellow. The following people contributed in a special way, reading portions of the manuscript and offering their comments on it: chronologically, John Kelley, Isaac Namioka, Hillel Furstenberg, and Benjamin Weiss. Finally, I wish to dedicate this paper to the memory of my undergraduate instructors, M. Loève and R. Bowen, who, as teachers, truly brought mathematics to life in the classroom. It was a privilege to have known them as teachers and as human beings, and it continues to be a privilege to learn from their published work.

References

[Bi1] P. Billingsley, "Probability and Measure," Wiley, New York, 1979

[Bi2] P. Billingsley, "Convergence of Probability Measures," Wiley, New York, 1968.

[CFS] I. Cornfeld, S. Fomin, and Ya. Sinai, "Ergodic Theory," Springer–Verlag, New York, 1982.

[DS] N. Dunford and J. Schwartz, "Linear Operators," Vol. I, Interscience, New York, 1958.

[F] H. Furstenberg, Random walks and discrete subgroups of Lie groups, *in* "Advances in Probability and Related Topics I" (P. Ney, Ed.), pp. 1–63, Dekker, New York, 1971.

[Fi] A. Fisher, "Gurevich Flows of Equal Entropy Are Finitarily Isomorphic," Thesis, University of Washington, Seattle, 1983.

[Fr] D. Freedman, "Brownian Motion and Diffusion," Holden–Day, San Francisco, 1971.

[Ga1] L. Garnett, "Functions and Measures Harmonic along Leaves of a Foliation," Thesis, Dartmouth College, Hanover, N.H., 1981.

[Ga2] L. Garnett, Foliations, the ergodic theorem and Brownian motion, *J. Functional Anal.* **51** (1983), 285–311.

[Gr] F. Greenleaf, "Invariant Means on Topological Groups and Their Applications," Van Nostrand–Reinhold, New York, 1969.

[Ha] P. R. Halmos, "Lectures on Ergodic Theory," Chelsea, New York, 1956.

[H-J] J. Hoffman–Jorgensen, How to make a divergent sequence convergent by Martin's axiom, Aarhus University Preprint Series No. 21, 1977/1978, Aarhus, Denmark.

[Ka] S. Kakutani, Random ergodic theorems and Markoff processes with a stable distribution, *in* "2nd Berkeley Symposium on Mathematical Statistics and Probability," (J. Neyman, Ed.), pp. 247–261, Univ. of California Press, Berkeley/Los Angeles, 1951.

[KS] M. Keane and M. Smorodinsky, Bernoulli schemes of the same entropy are finitarily isomorphic, *Ann. of Math.* (2) **109**, No. 2 (1979), 397–406.

[KV] V. Kaimanovitch and A. Vershik, Random walks on discrete groups: Boundary and entropy, *Ann. Probab.* **11** (3) (1983), 457–490.

[La] J. Lamperti, "Probability," Benjamin–Cummings, Reading, Mass., 1966.

[Le] F. Ledrappier, Poisson boundaries of discrete groups of matrices, to appear.

[Lo] M. Loève, "Probability Theory," Vol. II, 4th ed., Springer–Verlag, New York, 1977.

[LS] T. Lyons and D. Sullivan, Function theory, random paths and covering spaces, *J. Diff. Geom.* **19** (1984), 299–323.

[M] P. A. Meyer, Limite mediales d'apres Mokobodzki, *in* "Sem. Pr. VII Strasbourg," pp. 198–204, Lecture Notes in Mathematics, Vol. 321, Springer–Verlag, New York/Berlin, 1973.

[MWW] E. J. McShane, R. B. Warfield, Jr., and V. M. Warfield, Invariant extensions of linear functionals, with applications to measures and stochastic processes, *Pacific J. Math.* **28** (1969), 121–142.

[N] E. Nelson, Regular probability measures on function space, *Ann. of Math.* **69** (1959), 630–643.

[O] D. Ornstein, "Ergodic Theory, Randomness, and Dynamical Systems," Yale Univ. Press, New Haven, Conn., 1973.

[OS] D. Ornstein and P. Shields, Mixing Markov shifts of kernel type are Bernoulli, *Adv. in Math.* **10** (1973), 143–146.

[OW] D. Ornstein and B. Weiss, The Shannon–McMillan–Breiman theorem for a class of amenable groups, *Illinois J. Math.* **44**, No. 1 (1983), 53–60.

[Ox] J. Oxtoby, "Measure and Category," Springer–Verlag, New York, 1971.

[PP1] R. PARIKH AND M. PARNES, Conditional probability can be defined for all pairs of sets of reals, *Adv. in Math.* **9** (1972), 313–315.

[PP2] R. PARIKH AND M. PARNES, Conditional probabilities on uniform sets, *in* "Victoria Symposium on Nonstandard Analysis (1972)," pp. 180–194, Lecture Notes in Mathematics, Vol. 369, Springer-Verlag, New York, 1974.

[P] W. PARRY, "Topics in Ergodic Theory," Cambridge Tracts in Mathematics No. 75, Cambridge Univ. Press, New York/London, 1981.

[Ra] R. RAIMI, *Duke Math. J.* **26** (1959), 17–28.

[RSII] M. REED AND B. SIMON, "Methods of Modern Mathematical Physics," Vol. II, Academic Press, New York, 1975.

[Roy] H. L. ROYDEN, "Real Analysis," 2nd ed., Macmillan & Co., London, 1968.

[Ro] V. A. ROCHLIN, On the fundamental ideas of measure theory, *Mat. Sb.* **25** (67) (1949), 107–150; Trans. Math. Monographs No. 71, pp. 1–55, Amer. Math. Soc., Providence, R.I., 1952.

[Ru1] W. RUDIN, "Functional Analysis," McGraw-Hill, New York, 1973.

[Ru2] W. RUDIN, Invariant means on L^∞, *Studia Math.* **44** (1972), 219–227.

[Sh] P. SHIELDS, "The Theory of Bernoulli Shifts," Univ. of Chicago Press, Chicago, 1973.

[Si] B. SIMON, "Functional Integration and Quantum Physics," Academic Press, New York, 1979.

[Wa] P. WALTERS, "An Introduction to Ergodic Theory," Springer-Verlag, New York, 1982.

[Wi] D. WILLIAMS, "Diffusions, Markov Processes, and Martingales," Vol. I, Wiley, New York, 1979.

[Y] K. YOSIDA, "Functional Analysis," 6th ed., Springer-Verlag, New York, 1980.

[Zi] R. ZIMMER, Ergodic theory, semisimple Lie groups, and foliations by manifolds of negative curvature, to appear.