ERGODIC THEOREMS AND THE BASIS OF SCIENCE

KARL PETERSEN

ABSTRACT. New results in ergodic theory show that averages of repeated measurements will typically diverge with probability one if there are random errors in the measurement of time. Since mean-square convergence of the averages is not so susceptible to these anomalies, we are led again to compare the mean and pointwise ergodic theorems and to reconsider efforts to determine properties of a stochastic process from the study of a generic sample path. There are also implications for models of time and the interaction between observer and observable.

1. Introduction

Continuing research in ergodic theory brings new heat, light, and shadow to debate on the relationship between mathematical models and physical reality. Much of science depends on the existence of limits, in several possible senses, of averages of sequences of measurements. But if random fluctuations are present in measurements of time, these limits often do not exist. In order to discuss the recent results and the problems that they raise, we try both to describe them and to state them with some precision, but for complete detail the original papers should be consulted.

2. Mathematics vs. Physics

The widely read and discussed essay "The unreasonable effectiveness of mathematics in the natural sciences" [24] raised—already in its title—intriguing questions about the surprising usefulness of mathematics in modeling and analyzing natural processes so as to be able to make successful predictions about their behavior. Since the admiration that physicists and mathematicians have for each others' achievements does not completely overwhelm all possible skepticism, Wigner's essay was soon countered—by a physicist lecturing on "The unreasonable effectiveness of physics in leading to good mathematics", and even by a mathematician, Jacob T. Schwartz, in his address "The pernicious influence of mathematics on science" [16]. (There is also Halmos' "Applied mathematics is bad mathematics" [10], but that is really about another question. See also [21] and the collection containing it.)

Schwartz suggests that awareness of hidden complexities may allow an investigator safely to use a simple explanation for complex behavior— that is to say, a theory. (So a theory is something analogous to a simple program that can generate sets of data that might seem complicated, as in the idea of computational complexity.) But he complains that the mathematical formulation of a theory can be deceptive, encouraging us to suppose there is more to it, in terms of predictive or explanatory power, than there really is.

Supported in part by NSF Grant DMS-9203489. The author thanks Klaus Schmidt and Máté Wierdl for illuminating discussions.

Moreover, characteristic mathematical arguments, which must be precise to be convincing, may be useless in forming real-world science: arguments that are *convincing even if vague* are probably superior, since they are more likely to be stable under changes of hypotheses and particular conditions. Can there be an argument that is convincing but vague? Clearly this depends on the standard one sets in order to be convinced. Is there a "structural stability" for arguments and theories, so that certain ones might not change their essentials under some kinds of small perturbations? Could truth itself be somehow structurally stable, so that once we arrive close enough to real understanding, further discoveries will produce only refinements and not qualitative changes? It is easy to raise such questions, but difficult even to begin to specify a framework in which they might be addressed. Recent developments in ergodic theory and dynamical systems raise more precise and limited, but still intriguing, questions about the processes of measurement and analysis—the relationship between the real world and the scientific study of it—and the relationship between mathematics and the physical ideas it seeks to model.

3. Averages

Science begins with observations, when possible quantitative ones, that is to say, measurements. Strangely enough, if a measurement is repeated, usually a different value is obtained. Thus the scientist makes a sequence of measurements and seeks to analyze the resulting time series, perhaps by looking for repetitions (exact or approximate) or trends, perhaps by taking averages of the successive measurements. Often the sequence itself might not show any trend, but as averages are taken of longer and longer blocks of measurements, the averages might; for example, they might appear to settle down toward a constant limiting value. Perhaps the process of averaging allows one to cancel out, in the limit, the effects of random noise or experimental error or other irrelevant (additive, mean-zero) factors. (To decide what is the actual target of observation and what is irrelevant noise may seem to present a circular problem. It can be addressed by, for example, conducting several experiments that are identical in many aspects of their design. Effects that seem to persist may provisionally be considered real.) If the averages do converge to a limit, this may reveal some underlying structure in the process under investigation. As an example, suppose that a coin is flipped repeatedly, and our measurement consists in writing down a 0 for heads and a 1 for tails on each flip. The sequence of 0's and 1's will in general be quite irregular, but their averages will converge to a number p which represents the probability of getting a 1 on any flip and which carries some information about the physical structure of the coin or the nature of the flipping mechanism. If the averages fail to converge, we will question whether the same coin is being used all the time or whether the process is indeed governed by constant laws.

(Note the mistaken discussion in [16], where the Ergodic Theorem is criticized for being deceptive on the basis that it deals with equilibrium situations, so there is no point in taking averages since all observables will have reached their constant equilibrium values. Of course there are many equilibrium situations in which interesting quantities are far from constant, such as the repeated independent flips of a coin above, or (frictionless, conservative) linked pendulums, or a constrained system of particles making elastic collisions. Even though frictionless cases may not be realizable in practice, they can be extremely well approximated. And the interest in the convergence of averages persists in many purely mathematical examples, including ones that model physical processes.)

A few words about a standard way to model physical systems, deterministic or stochastic, within the same framework, namely that of ergodic theory. Let us denote the set of all possible *states* of the system being studied by X. A point $x \in X$ gives all possible information about the system; even if the system consists of a huge number of particles, knowing x will bring with it absolutely perfect knowledge of, for example, the position and momentum of every particle in the system. Such perfect information is unattainable; instead of knowing the state x exactly, typically we will be able to know only whether or not x is consistent with some *observable event* B specified by a certain subset of X, that is, whether or not $x \in B$. The collection of all observable events forms a subfamily $\mathcal{B}$ of the family of all subsets of X. $\mathcal{B}$ is assumed to be a (nonempty) *σ-algebra*—closed under countable unions, countable intersections, and the taking of complements—so as to permit countably-infinite set-theoretic (equivalently logical) operations and hence analysis, which depends on taking limits. The *probability* of each observable event is determined by a function $\mu : \mathcal{B} \to [0,1]$ that is assumed to be countably additive ($\mu(\cup_{k=1}^{\infty} \mathcal{B}_k) = \sum_{k=1}^{\infty} \mu(B_k)$ if the B_k are pairwise disjoint) and such that $\mu(X) = 1$. A *measurement* on the system, or *random variable*, is a measurable function $f : X \to \mathbb{R}$. (For f to be measurable means that for each interval I in the real line $\mathbb{R}$, the event that f takes a value in I, namely $f^{-1}(I) = \{x : f(x) \in I\}$, is an observable event, i.e. is a member of $\mathcal{B}$.)

Now we get to the dynamics of the system, namely its development in time, presumably on the basis of some sort of fixed laws. This is modeled by a *measure-preserving transformation* $T : X \to X$. This means that T is a one-to-one onto map from X to X (possibly after a set of measure 0 has been excised from X) which preserves observability ($B \in \mathcal{B} \Leftrightarrow T^{-1}B \in \mathcal{B} \Leftrightarrow TB \in \mathcal{B}$) and probability ($\mu(B) = \mu(T^{-1}B) = \mu(TB)$ for each $B \in \mathcal{B}$). The idea is that if at any given instant the system is in state $x \in X$, then at the next instant the system is in state Tx. The invariance of μ under T—the assumption that probabilities of observable events do not change with time—reflects the fact that we are in an *equilibrium* situation. As mentioned above, this does not at all mean that the system is completely static! Think of several species competing or cooperating on an island when a new species is introduced. There can be a period of time when the populations fluctuate wildly before reaching a new equilibrium situation, which might itself involve fluctuations, periodic or even irregular, in the population sizes. This sort of approach toward equilibrium is sometimes represented mathematically as follows: an initial point is chosen according to Lebesgue probability measure in the manifold of possibilities available in the world of the experimenter (supposedly listing the actual real-world values of relevant variables, subject to errors in measurement); the trajectory of this point approaches an "attractor" in the manifold, on which the limiting dynamics can be quite complicated (but not necessarily—there could be just a single fixed point, or periodic orbit); and Lebesgue measure itself in time evolves towards a limiting invariant measure, supported on the attractor, which describes at least in statistical terms the dynamics of the equilibrium situation for the system being studied. Beginning with different initial points will in general lead in the long term to behavior that is quite different in its particulars but identical in a qualitative sense, for example in terms of the relative amounts of time that the system spends near any particular state.

So far we have treated time as discrete or quantized, proceeding in a sequence of successive instants. Another popular model involves a continuous time parameter: we suppose that we have a family $\{T_t : t \in \mathbb{R}\}$ of measure-preserving transformations such that T_0 is the identity map on X, $T_{s+t} = T_s \circ T_t$ for all $s, t \in \mathbb{R}$, and $T_t x$ is a jointly measurable function of the two variables t and x. We think of $T_t x$ as specifying the state at time t of the system if at time 0 it is in state x. This is the model based on the solution of systems of differential equations (supposed to encapsulate the physical laws governing time development), with given initial conditions (supposed to contain the raw data corresponding to the contingencies of a particular physical realization). Although this model was long associated with the culmination of Newtonian mechanics in a deterministic "clockwork" view of the universe, we can also use it to model systems in which chance, indeterminacy, or hopelessly unanalyzable complexity are involved in the essentials.

The time development of the system, starting from state x at time 0, is given by the *orbit* or *trajectory* $\mathcal{O}(x) = \{x, Tx, T^2x, \dots\}$. The scientist seeking to measure an observable quantity f obtains the sequence of real numbers $\{f(x), f(Tx), f(T^2x), \dots\}$ and may compute the sequence of averages

$$A_n f(x) = \frac{1}{n} \sum_{k=1}^{n} f(T^k x) \ .$$

The existence of the *limit* of these averages as $n \to \infty$ is a question of fundamental importance for science: Is it true that for almost every initial state x (that is, with the possible exception of a set of probability 0) $\lim_{n\to\infty} A_n f(x) = \bar{f}(x)$ exists? If so, $\bar{f}(x)$ represents a mean or average value of the observable f that can carry some information about the underlying structure of the system. If not, (for example, if the averages continue to wander aimlessly as n increases), we have to question the meaning of the sequence of measurements or of the model itself.

It has been remarked since the beginnings of ergodic theory [17] that even an individual measurement (like our above $f(x)$) that is supposed to be instantaneous can in fact already be essentially a long-term average of many repeated measurements (like $\lim_{n\to\infty} A_n f(x) = \bar{f}(x)$). A needle on a gauge is averaging, through its momentum, the effects of many small influences that take place on a time scale much more rapid than ours. Now if the limit does not exist with probability 1, we cannot even write down a *single* measurement $f(x)$, and science can't even get started!

4. The Ergodic Theorems

The existence of the limits as $n \to \infty$ of these averages was proved in 1931 by von Neumann, for mean convergence, and Birkhoff, for almost everywhere convergence (see [14]). The theorem of von Neumann asserts convergence in the Hilbert space $L^2(X, \mathcal{B}, \mu)$ in the mean-square sense:

$$\|A_n f - \bar{f}\|_2^2 = \int_X |A_n f(x) - \bar{f}(x)|^2 \, d\mu(x) \to 0 \ ;$$

whereas Birkhoff's states that for all initial states x, with the possibility of an exceptional set of measure 0,

$$A_n f(x) \to \bar{f}(x) \text{ as } n \to \infty \ .$$

There has been some debate about which of these results should be called "the" Ergodic Theorem [4, 5, 6, 17, 19]. Birkhoff's theorem is mathematically deeper and more delicate. It seems to reflect reality in that a scientist in a laboratory is really presented with a sequence of measurements $\{f(x), f(Tx), \dots\}$ depending on a single world history starting from an initial state x; he must deal with this and does not have the possibility of accessing other possible initial states. On the other hand, von Neumann argues persuasively [17] that it may not be necessary or reasonable to ask for more than mean square convergence. When quantum effects are taken into account, it is not possible to make a sequence of measurements $\{f(x), f(Tx), \dots\}$, since making a measurement perturbs the system in an unpredictable way. Because of the Uncertainty Principle, we have even greater problems when trying to measure two observables simultaneously (f could be a vector-valued observable, too). But by making many measurements, each on a relatively small subsystem of a very large system, we could very well obtain estimates for quantities like the probability that $A_n f$ is in a certain Borel subset of $\mathbb{R}$ or $\mathbb{R}^d$, thereby making sense out of the (probabilistic) mean-square convergence statement. (See [18], pp. 299–300 and the surrounding discussion, also pp. 211–214 and 221–222.) Perhaps this operational viewpoint is closer to a structural stability of scientific conclusions—small perturbations or imprecisions are not likely to produce large qualitative changes—whereas in the pointwise approach a small (whatever that might mean) change in the initial state could lead to a vastly different development of the system. Moreover, von Neumann states [17] that estimates about the *speed* of convergence of $A_n f$ to its limit can be made, if we somehow have information about the spectral measure $E_T f$. This information can be extremely hard to come by, and may depend again on applying a convergence theorem, thereby introducing circularity to the argument. But again it could be the sort of circularity that can be avoided by controlled repetition, as is often done in physics, where a defining or calibrating step sets up measurements based on it. (Thus Newton's second law $F = ma$ first defines mass as a constant of proportionality, but afterwards it functions as a law, predicting what motions will occur under certain circumstances.)

5. Inferring the Structure of a Process from an Individual Sample Path

Norbert Wiener, in making key contributions to ergodic theory as well as probability, seemed to base a sizable chunk of his work on the effort to determine information about the structure of a stochastic process (a sequence $\{f_1, f_2, \dots\}$ of random variables) from a *sample path* $\{f_1(x), f_2(x), \dots\}$. Now *stationary* stochastic processes—the ones whose probability distributions do not change with time—coincide exactly with the sequences $\{f, fT, fT^2, \dots\}$ considered above. In an extension of Birkhoff's theorem, Wiener and Wintner [23] showed how to get at the discrete spectrum (the set of eigenvalues of the associated unitary operator) of a stationary stochastic process: there is a single set of measure 0 in X outside of which

$$\xi_x(\lambda) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} e^{-ik\lambda} f(T^k x) \text{ exists for all } \lambda \in \mathbb{R} .$$

Thus there is defined a *spectral function* ξ_x; it is non-zero only for at most countably many λ, namely the elements of the discrete spectrum. For each fixed λ the existence of the limit for almost every x is an immediate consequence of Birkhoff's theorem; the interest and novelty here is that the *uncountably many* sets of measure 0, one for each λ, have been combined into a single bad set—still of measure 0!—that works simultaneously for all λ. (The union of countably many sets of measure 0 still has measure 0, but the property of having measure 0 need not be preserved under uncountable unions.)

In [9] a relationship was established between the entire *spectral measure* E_T of a measure-preserving transformation T (that is, of its associated unitary operator) and of another type of ergodic-theoretic limit, the *helical transform*

$$H_\epsilon f(x) = \lim_{n\to\infty} \sum_{k=-n}^{n}{}' e^{-ik\epsilon} \frac{f(T^k x)}{k} ,$$

the $'$ meaning that the term $k = 0$ is omitted:

$$\frac{1}{i}(H_\epsilon f - H_0 f) = -\epsilon f + \pi[E_T\{-\epsilon\}f + E_T\{0\}f] + 2\pi E_T(-\epsilon, 0)f .$$

There a maximal inequality (usually necessary for handling difficult almost everywhere convergence results) was proved and the Wiener-Wintner property for the helical transform was conjectured. Substantial progress for establishing this property has been made by I. Assani and by M. Lacey. This property would permit analysis of a single sample path of a process to obtain detailed information about its entire spectral measure, not just about its point spectrum.

In this connection we emphasize one great advantage of the pointwise, or sample path, approach: we can carry out our computations and deductions without knowing anything about the actual probability μ that describes the relative probabilities of events. We can even find out intimate information about the spectral measure E_T without even knowing the Hilbert space $L^2(X, \mathcal{B}, \mu)$ on which it acts. This is done by breaking the expressions involved up into their real and imaginary parts, then each of those into their positive and negative parts, and in the limits (in the ergodic theorem or for the helical transform) replacing the limit by a limit superior, which always exists for *all* x, if we allow $+\infty$ as a limit. Regardless of μ, we will *almost surely* (with respect to μ)—that is to say, *certainly*—get the correct value of the limit for that part. This provides some justification for the working process of science—stumbling around in the dark, averaging measurements, without any real idea of what we're doing, with no knowledge of the situation that we're actually in, success is nevertheless guaranteed: the average is *bound* to exist and equal what it should. Here is a kind of stability that does not depend on vagueness.

There are other aspects of the global structure of a system that can be determined by studying typical sample paths: its entropy, by means of the Shannon-McMillan-Breiman Theorem (see [14]); the algorithmic complexity of individual orbits [8, 22]; and, for Bernoulli (independent identically-distributed) processes, the entire process itself [13].

6. Time Fluctuations

In the process of measurement, that is, in arriving at the number $f(T^k x)$ or $f(T_t x)$, there can be fluctuations and errors not only in f, but also in k or in t, since time itself is measured, or at least somehow counted or kept track of. What happens if we try to take measurements once every 100 ticks of our clock, but lose count now and then? Or if we try to take measurements, say of the earth's magnetic field, every midnight, but actually make small errors with a Gaussian distribution around those target times? It turns out that in the case of discrete time there is wide latitude for error—the averages will still converge almost surely to the correct value. But in the case of continuous time, there will always be observable quantities whose averages *diverge* almost everywhere. On the other hand, mean convergence always works. These statements are based on results in [3, 7, 12]; we proceed to give a bit more detail.

First of all, we emphasize that we are dealing with the same kind of universality or decoupling that is found in the Wiener-Wintner Theorem. The measurement times $\tau_k(\omega)$ and spaces between them $\delta_k(\omega) = \tau_k(\omega) - \tau_{k-1}(\omega)$ are determined in advance, and the same sequence of times is then to be used in all laboratories on all experiments. We assume that we are in an equilibrium situation, so that the spaces between measurements form a stationary stochastic process: $\delta_k(\omega) = \delta(T^k\omega)$ for a measure-preserving transformation $T : \Omega \to \Omega$ on a probability space $(\Omega, \mathcal{F}, P)$ and measurable function $\delta : \Omega \to [0, \infty)$. The question is whether or not for all ω, with the possible exception of a set of measure 0, we obtain a sequence of sampling times $\tau_k(\omega)$ such that if any stationary process is observed at this sequence of times, then the resulting sequence of averages

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

converges for almost every y in a probability space $(Y, \mathcal{C}, \nu)$, whatever the measure-preserving transformation $S : Y \to Y$ and measurable (perhaps also bounded, or integrable, or p'th-power integrable) function g on Y.

First, the positive statement for the discrete-valued case. So long as the spacing function δ is integrable and has positive finite expectation (mean), we *will* have almost everywhere convergence of the averages of these randomly time-perturbed measurements, for every *bounded* measurable g [12]. Possibly the boundedness restriction is removable, but so far we only know that, in case the δ_k are independent and identically distributed with finite second moment and nonzero mean, then we have almost everywhere convergence for each g whose p'th power is integrable ($p > 1$). On the other hand, in case the process $\{\delta_k\}$ giving the waiting times between measurements has mean 0, there will in every system Y be bounded (even 0,1-valued) observables g whose averages along the sequence of random times will *diverge* almost surely [7, 12]. Suppose that our scientist has been making measurements since time immemorial and has listed them in order as $\{\dots, m_{-1}, m_0, m_1, m_2, \dots\}$. We start at position 0 on the list and move either one spot forward or backward on the list, depending on whether a sequence of fair coin tosses that was made and written down previously came up heads or tails at the corresponding time. We obtain a sequence of numbers $\{n_0, n_1, \dots\}$ that wanders back and forth in the original list. The statement is that quite possibly $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} n_k$ will not exist.

In the continuous-time case, if the waiting function δ has continuous distribution in some interval, say a sharply peaked normal distribution centered at 1, then the sequence of sampling times that it generates will be bad: for almost every ω, in every nontrivial system Y there will be bounded observables g for which the averages $A_n^\omega g(y)$ diverge for almost every y [3].

These discrete averages in a continuous-time system are somehow *inappropriate*—usually in such a case one considers integral averages; this is a second difficulty to pile on top of universality (the attempt to organize uncountably many bad sets of measure 0 into a single set of measure 0). An analogous situation is found in sampling a higher-dimensional discrete ($\mathbb{Z}^d$) action along a one-dimensional random walk. Suppose that we have two commuting measure-preserving transformations S_0 and S_1; this could represent a *random field* in (two-dimensional) space, as considered in statistical mechanics (say random magnetizations, up or down, at each site in a crystal), or perhaps two different principles of development that could operate on a system. At each discrete time instant $k = 1, 2, \ldots$ we apply one or the other of these transformations, depending on whether heads or tails came up on our previously-observed sequence of coin flips. Again in every nontrivial system there will always be bounded observables for which the sequence of averages of measurements diverges almost everywhere. Similarly for any $d = 3, 4, \ldots$.

7. Possible Solutions

(1) *Mean convergence.* In all of the situations described above, mean convergence (in L^2) continues to hold. So von Neumann's theorem, though cruder than Birkhoff's, has robustness not only against the difficulties caused by quantum theory but also those arising from random time fluctuations.

(2) *Subsequences of averages.* In case our random process giving the waiting times between measurements takes positive integer values, we are dealing with a subsequence of the set of times. There is a large literature of positive and negative results about the existence of averages of measurements along subsequences of the integers, including some striking and important recent positive results for the sequences of squares and primes; see [15] for a comprehensive survey.

From [7, 12] it follows (for $\{\delta T^k\}$ i.i.d.) that provided we look at the *averages* $A_n^\omega g(y)$ only infrequently, say for $n = r^r, r = 1, 2, \ldots$, we *will* get almost everywhere convergence for higher-dimensional actions (including the case of T and T^{-1} being applied according to a sequence of coin flips), even though the full sequence of averages can diverge. Thus a very careful scientist can analyze the sequence of observations by forming the sequence of averages and discarding most of them (provided the time fluctuations are i.i.d.).

(3) *Fewer observables.* In any system Y there are many observables g whose averages *do* converge as they're supposed to, although in the sense of Baire category most g might be ill-behaved. Conceivably it is only these good functions (which typically have some "smoothness" to them) that are relevant for modeling the real world.

(4) *Exceptional sets.* In the delicate theorems of analysis that treat pointwise convergence, usually an exceptional set of measure 0 must be avoided. But how do we know that our system is not following a history determined by one of these bad initial states? After all, in reality we have only a single world history available, just one point x; how can it be known whether it is good or bad? We are supposed only to have confidence that our model faithfully reflects what will occur with overwhelming probability—which we take as certainty—when no cheating by rigging up of the initial conditions into peculiar settings is allowed.

(5) *Multiple averaging.* As mentioned above, $f(T^k x)$ might actually be not a single measurement, but already an average, or even nearly a limit of averages, "computed" for the experimenter by his measuring devices or process. This pre-averaging can provide the smoothing discussed in (3). Nonexistence of the first averages could account for some of the fluctuations in our measurements.

(6) *Quantized time.* If time comes in discrete ticks rather than a continuous stream, then it can in principle be measured perfectly precisely. And even if we make random errors in counting ticks, so long as our measurement times continue to advance at least

on the average (recall that δ only had to have *mean* $\neq 0$), we will obtain universal convergence of averages of all bounded measurements.

(7) *No decoupling.* The universality, in the style of the Wiener-Wintner Theorem, sought for the existence of averages amounted to a decoupling of the process that determines the measurement times from the processes being measured: a random sequence of times was determined in advance, then the same sequence of times was used for all measurements in all laboratories. This might be based on the suspicion that the experimenter is part of a different system, and so subject to different laws, than the systems he is studying. Perhaps such a decoupling into different levels of reality is impossible—the sequence of random time spacings $\{\delta_k\}$ *must* be determined in concert with the performance of the experiment and the readings of measurements at those times. Parallel experimenters attempting to make measurements at exactly midnight will each make their own slight time errors, no matter how hard we try to synchronize their measurements. Thus these abstract mathematical investigations give some theoretical support to the parts of relativity and quantum mechanics that argue the impossibility of enforcing or determining simultaneity or of separating the experimenter from the system being studied.

References

1. Akcoglu, M., del Junco, A., Jones, R., Losert, V., Reinhold-Larsson, K., and Wierdl, M. *The strong sweeping out property for lacunary sequences, Riemann sums, convolution powers and related matters.* to appear in Erg. Th. Dyn. Sys.
2. Akcoglu, M., Bellow, A., del Junco, A., and Jones, R. 1993. Divergence of averages obtained by sampling a flow. *Proc. Amer. Math. Soc.*, **118**, 499–505.
3. Bergelson, V., Boshernitzan, M., and Bourgain, J. 1994. Some results on non-linear recurrence. *J. d'Analyse Math.*, **62**, 29–46.
4. Birkhoff, G. D. 1942a. The ergodic theorems and their importance in statistical mechanics. *Revista Ci., Lima*, **44**, 4–116.
5. Birkhoff, G. D. 1942b. What is the Ergodic Theorem? *Amer. Math. Monthly*, **49**, 222–226.
6. Birkhoff, G. D. 1943. The mathematical nature of physical theories. *Amer. Scientist*, **31**, 281–310.
7. Bourgain, J. 1993. Convergence of ergodic averages on lattice random walks. *Ill. J. Math.*, **37**, 624–636.
8. Brudno, A. A. 1983. Entropy and the complexity of the trajectories of a dynamical system. *Trans. Moscow Math. Soc.*, **2**, 127–151.
9. Campbell, J., and Petersen, K. 1989. The spectral measure and Hilbert transform of a measure-preserving transformation. *Trans. Amer. Math. Soc.*, **313**, 121–129.
10. Halmos, P. R. 1981. Applied mathematics is bad mathematics. *In:* Steen, L. A. (ed), *Mathematics Tomorrow.* Springer-Verlag, N.Y.
11. Jones, R. L., and Wierdl, M. 1994. Convergence and divergence of ergodic averages. *Erg. Th. Dyn. Sys.*, **14**(3), 515–535.
12. Lacey, M., Petersen, K., Rudolph, D., and Wierdl, M. 1994. Random sampling with universally representative sequences. *Ann. Inst. H. Poincaré, Prob. et Stat.*, **30**, 353–395.
13. Ornstein, D., and Weiss, B. 1990. How sampling reveals a process. *Ann. of Prob.*, **18**, 905–930.
14. Petersen, K. 1983, 1989. *Ergodic Theory.* Cambridge University Press.
15. Rosenblatt, J., and Wierdl, M. 1995. Pointwise ergodic theorems via harmonic analysis. *In:* Petersen, K. E., and Salama, I. A. (eds), *Ergodic Theory and Its Connections with Harmonic Analysis: Proceedings of the 1993 Alexandria Conference.* Cambridge University Press.
16. Schwartz, J. T. 1986. The pernicious influence of mathematics on science. *Pages 19–25 of:* Kac, M., Rota, G.-C., and Schwartz, J. T. (eds), *Discrete Thoughts.* Birkhäuser, Boston.
17. von Neumann, J. 1932. Physical applications of the Ergodic Hypothesis. *Proc. Nat. Acad. Sci. U.S.A.*, **18**, 263–266.

18. von Neumann, J. 1955a. *Mathematical Foundations of Quantum Mechanics*. Princeton University Press.
19. von Neumann, J. 1955b. Method in the physical sciences. *Pages 157–164 of:* Leary, L. (ed), *The Unity of Knowledge*. New York: Doubleday.
20. von Neumann, J., and Koopman, B.O. 1932. Dynamical systems of continuous spectrum. *Proc. Nat. Acad. Sci. U.S.A.*, **18**, 255–263.
21. West, B. J. 1990. The disproportionate response. *Pages 258–290 of:* Mickens, R. E. (ed), *Mathematics and Science*. World Scientific, Singapore.
22. White, H. S. 1993. Algorithmic complexity of points in dynamical systems. *Erg. Th. Dyn. Sys.*, **13**, 807–830.
23. Wiener, N., and Wintner, A. 1941. Harmonic analysis and ergodic theory. *Amer. J. Math.*, **63**, 415–426.
24. Wigner, E. P. 1960. The unreasonable effectiveness of mathematics in the natural sciences. *Comm. Pure Appl. Math.*, **13**, 1–14.

Department of Mathematics, CB 3250, Phillips Hall, University of North Carolina, Chapel Hill, NC 27599 USA

E-mail address: `petersen@@math.unc.edu`