ILLINOIS JOURNAL OF MATHEMATICS
Volume 43, Number 3, 1999

FLUCTUATIONS OF ERGODIC AVERAGES

STEVEN KALIKOW AND BENJAMIN WEISS

ABSTRACT. We discuss universal estimates for the probability that there are many fluctuations in the ergodic averages of L^1 functions. Our methods involve an effective Vitali covering type of theorem and are valid for $\mathbb{Z}^d$ actions, for any $d \in \mathbb{N}$. For nonnegative functions we get an exponential decay for the probability of a large number of fluctuations.

It was E. Bishop who first showed that one can give universal estimates for the probability that there are many fluctuations in the ergodic averages of L_1 functions. His motivation was the desire to find a version of the ergodic theorem that would be valid in a constructive framework. This work is described in his book *Foundations of constructive analysis* [B].

Bishop's proof relies quite heavily, both on the linear ordering of $\mathbb{Z}$ and on the fact that he deals with one sided averages. In order to extend these results of $\mathbb{Z}^d$-actions it was necessary to develop a new method which gave new results even for $d = 1$ and which we shall now proceed to describe.

A real valued sequence $\{s(1), s(2), \ldots s(n)\}$ is said to *fluctuate* at least N-times across the gap (α, β) if there are N indices $j(0) < j(1) < j(2) < \cdots < j(N) \leq n$ such that for *odd* k, $s(j(k)) \leq \alpha$, while for *even* k $S(j(k)) \geq \beta$. We also speak of the *sequence having at least N-fluctuations*. For E. Bishop's theorem, let (X, B, μ, T) be a probability space with $T \circ \mu = \mu$ and let f be an L_1-function. Then we have:

THEOREM (E. BISHOP). *Let E_N^n denote the set of x for which the sequence $\left\{\frac{1}{m}\sum_0^{m-1} f(T^i x)\right\}_1^n$ has at least N-fluctuations across $\alpha < \beta$. Then*

$$\mu(E_N^n) \leq \frac{\|f\|_1}{(\beta - \alpha)N}.$$

This is an exact analogy of Doob's upcrossing inequality for martingales and indeed Bishop proves a very general lemma from which both Doob's inequality and the above follow. For integrable functions one cannot improve on the above, as simple examples show. However, if f is bounded, or even only semibounded, say $f \geq 0$, then our techniques yield:

THEOREM 1. *Suppose that $f \geq 0$ and $0 \leq \alpha < \beta < \infty$. Then there exist constants $c_1 > 0$, $c_2 < 1$ that depend only on β/α such that*

$$(*) \qquad \mu(E_N^n) \leq c_1 (c_2)^N \quad \text{for all } n,$$

Received April 3, 1998.
1991 Mathematics Subject Classification. Primary 28D05; Secondary 28D15.

where E_N^n, as above, is the set of x where the sequence $\left\{ \frac{1}{m} \sum_0^{m-1} f(T^i x) \right\}_1^n$ *has at least N-fluctuations across $\alpha < \beta$.*

Note that the analogous result for martingales was proved in 1962 by L. Dubins [D-1962].

This and most of the other results that are contained in this paper were obtained by us in 1993–94. At that time the only other work in this general direction that we were aware of was the paper of J. Bourgain [Bo] that dealt with other types of norms on the sequences of ergodic averages. Quite recently there has been a resurgence of interest in these matters. We refer to several recent publication and preprints, [Kh], [JKRW] and [CE] which give many more results of this general type. Our methods are rather different from what can be found in these papers and give results that are not contained in them. Since both [Kh] and [JKRW] also devote some attention to martingales we should point out that our methods work in the martingale context just as well. In fact we re-proved the main results of L. Dubins by these methods before learning of his work from D. Gilat.

1. Effective Vitali covering

We will write out a proof for dimension 2 but it is obvious that all statements and proofs carry over to any dimension. In addition, squares can be replaced by circles, or rectangles–as long as the ratio of the sides remain bounded–however in order to show how simple the ideas are we will keep to squares. For the sake of convenience we will call half the side length of a square its *radius*. Throughout this section, n and N are fixed integers with $N \gg n$. Let $\Omega_N = [-N, N] \times [-N, N] \subset \mathbb{Z}^2$ and normalize the area so that the area of Ω_N is 1, thus if $S \subset \Omega_N$ the area of S is $|S|/(2N+1)^2$. Denote this area by $\lambda(S)$, so

$$\lambda(S) = |S|/(2N+1)^2.$$

We call $\hat{C}_e$ the *e-enlargement* of a square C if $\hat{C}_e$ is the square concentric to C with radius equal to e times the radius of C.

THEOREM (EVC). *Let $B \subset \Omega_N$ with $n, N, r > 2$ fixed. Assume that to each $p \in B$ there are associated squares $A_1(p), \dots, A_n(p)$ such that:*

(1) *$p \in A_i(p) \subset \Omega_N$, $1 \le i \le n$.*
(2) *For all $1 \le i < n$ the r-enlargement of $A_i(p)$ is contained in $A_{i+1}(p)$.*

Set

$$S_i = \bigcup_{p \in B} A_i(p), \qquad 1 \le i \le n.$$

Then there is a disjoint subcollection of these squares such that

(a) *the union of their* $\left(1+\frac{4}{r-2}\right)$*-enlargements together with* $S_n \setminus S_1$ *covers all but at most* $(8/9)^n$ *of* S_n;
(b) *the measure of this union is at least* $(1-(8/9)^n)$ *times the measure of* S_1.

We remark that the $A_i(p)$ in the theorem need not be centered at p. Also note that (a) and (b) are really two separate results. During the proof we will call the squares of the form $A_i(p)$ for some $p \in B$ the *i-th level* squares. Such a square will be called *maximal* if it is not properly contained in any other i-th level square. Clearly S_i is also the union of the i-th level maximal squares.

Proof of the EVC Theorem. Using the standard Vitali covering argument we can find a disjoint collection of maximal n-th level squares, $A_n(p)$, that covers at least $1/9$ of S_n. Call this collection of squares $\mathcal{C}_n$ and let U_n denote the union of $\mathcal{C}_n$. Now we would like to restrict to $(n-1)$-level squares that are disjoint from U_n, but need not cover all of $S_{n-1} \setminus U_n$. The main point in the proof is the following:

$(*)$ *If an* $A_{n-1}(p)$ *has a non empty intersection with* U_n *then it is contained in the* $\left(1+\frac{2}{r-2}\right)$*-enlargement of some square* $A_n(p')$ *in* $\mathcal{C}_n$.

Assuming this, we restrict to those maximal $(n-1)$-level squares that are disjoint from U_n, and cover $1/9$ of their union by disjoint squares. This collection is called $\mathcal{C}_{n-1}$ and its union U_{n-1}. Analogous to $(*)$ we now have:

$(**)$ *If an* $A_{n-2}(p)$ *has a non-empty intersection with* $U_n \cup U_{n-1}$ *then it is contained in the* $\left(1+\frac{4}{r-2}\right)$*-enlargement of the square which it intersects.*

Again assuming $(**)$ we restrict attention to those maximal $(n-2)$-level squares that are disjoint from $U_n \cup U_{n-1}$ and cover at $1/9$ of their union by disjoint maximal squares, calling the collection $\mathcal{C}_{n-2}$ and its union U_{n-2}. We continue in the same way using the obvious generalization of $(**)$ to get $\mathcal{C}_{n-3}$, U_{n-3}, and so on up to $\mathcal{C}_1$ and U_1.

Clearly the union of all the $\left(1+\frac{4}{r-2}\right)$-enlargements of all the $\mathcal{C}_i$, $1 \le i \le n$, together with $S_n \setminus S_1$ covers all but $(8/9)^n$ of S_n and the measure of that union is at least $1-(8/9)^n$ times the measure of S_1. This will establish (a) and (b) once we verify $(*)$, $(**)$ and their generalizations. They are consequences of the claim that if

$$A_i(p) \cap A_j(q) \neq \emptyset, \quad i < j,$$

then the $\left(1+\frac{4}{r-2}\right)$-enlargement of $A_j(q)$ contains $A_i(p)$. To see this set

$$\begin{aligned} d &= \text{distance between the center of } A_i(p) \text{ and the center of } A_j(q) \\ d_1 &= \text{radius of } A_i(p) \\ d_2 &= \text{radius of } A_j(q). \end{aligned}$$

If our claims were to be false that would mean that the width of the new strip in the enlargement, $\left(\frac{4}{r-2}\right) d_2$ would be less than $2d_1$, which gives

$$rd_1 \geq 2d_1 + 2d_2.$$

But for the intersection to be non-trivial, $d \leq d_1 + d_2$ and it would follow that the r-enlargement of $A_i(p)$, would contain $A_j(q)$; thus by the hypotheses of the theorem, $A_{i+1}(p) \supset A_j(q)$ contradicting the maximality of $A_j(q)$ since $i+1 \leq j$. This completes the proof of the theorem. □

An important corollary of this result which in fact will be our main tool is an easy consequence of part (a):

Corollary. *Suppose in addition to hypotheses* (1) *and* (2) *we assume*:

(3) *Measure* $(S_n) \leq 10$ *measure* (S_1).

Then the union of the $\left(1 + \frac{4}{r-2}\right)$*-enlargements covers at least* $\left(1 - 10 \cdot \left(\frac{8}{9}\right)^n\right)$ *of* S_1.

2. Exponential decay for nonnegative functions

Our goal here is to show that if $f \geq 0$ and $A_m f(x)$ represent the average of $f(T_u x)$ as u ranges over the square $[-m, m] \times [-m, m]$ for a $\mathbb{Z}^2$ measure-preserving action $\{T_n \colon u \in \mathbb{Z}^2\}$ then for any fixed $0 < \alpha < \beta \leq 1$, there are constants c_1, c_2 $(c_2 < 1)$ depending only on α/β such that

$$\mu\{x\colon \text{the sequences } A_m f(x) \text{ has more than } N \text{ successive downcrossings from above } \beta \text{ to below } \alpha\} \leq c_1 c_2^N.$$

To gain a better understanding of the argument we shall first establish the theorem without taking care to estimate c_2. Later we will see how it converges to 1 as α/β tends to 1. This estimate will be important for our application to γ-fluctuations but we postpone a discussion of it.

Our first observation is that because the function f is nonnegative if $A_n f(x) \geq \beta$ while $A_m f(x) \leq \alpha$ with $m > n$ then $m/n > \sqrt{\beta/\alpha}$. It follows that by skipping every L downcrossings we can assume that we get the rapid growth of the size of the squares as required in the proof of the effective Vitali covering theorem. Clearly this L depends only on α/β and the desired r (cf. (r) in the EVC theorem).

Technically, we will carry out the proof by estimating the size of the set where N downcrossings take place along orbits of the action. To carry this out we introduce:

$$E_{N,M} = \{x\colon \text{the sequence } \{A_m f(x)\colon 0 \leq m \leq M\} \text{ undergoes at least } N \text{ downcrossings from above } \beta \text{ to below } \alpha\}.$$

The set whose measure we are trying to estimate is $\bigcup_M E_{N,M}$ so that it will suffice to get an estimate for fixed $E_{N,M}$ that is independent of M. To do this we will choose W to be very large compared to M, fix an $x_0 \in X$ and define a set

$$B \subset [-W, W] \times [-W, W]$$

by declaring that $u \in B$ if both coordinates of u do not exceed $W - M$ in absolute value and $T_u x_0 \in E_{N,M}$. The first condition ensures that the squares centered at n over which the downcrossings are taking place lie within $[-W, W] \times [-W, W]$. Any upper bound for the relative density of B will give the same upper bound for the measure of $E_{N,M}$ because of the invariance of the measure under the action T. This is essentially the content of the easy direction of A. Calderon's transfer principle. For the reader's convenience we give a brief description of this.

By the invariance of the measure μ under T_u we have the equality

$$\sum_{\substack{|u_1| \le W \\ |u_2| \le W}} \int 1_{E_{N,M}}(T_u x)\, d\mu(x) = (2W+1)^2 \mu(E_{N,M}).$$

Divide both sides by $(2W+1)^2$; then we see that $\mu(E_{N,M})$ is estimated by an upper bound for the relative density of B plus M/W. Letting W tend to infinity gives the desired result.

Each time we want to use the EVC in the form of the corollary we must check to see if conditions (2) and (3) are satisfied. For (2) we merely skip in steps of size L. If (3) doesn't hold, then by applying the standard Vitali covering argument we get disjoint squares of the highest level with a total area increase of $10/9$ over the previous stage. The main interest is in α/β close to 1 so we might as well assume $\beta/\alpha \le \frac{10}{9}$ and then we don't need to carry out any complicated argument–we just record this increase and proceed to the next step. It follows without loss of generality that we may assume that both (2) and (3) of the EVC hold.

Now, fix r and n for use in the effective Vitali covering (EVC), then L as above, and divide N into groups of nL successive downcrossings. These r and n depend on α/β and will be specified below. Using the first group we can replace B by B_1 which is a union of disjoint squares over each of which the average of f is above β and the normalized area or λ-measure of B_1 is almost that of B. Next we apply the EVC to B_1 using the next group of nL downcrossings, to obtain B_2, a union of disjoint squares over each of which the average of f is now below α. Furthermore B_2 almost covers B_1.

We would like to say that this forces the λ-measure of B_2 to have increased over that of B_1 by a factor of essentially β/α. We would be correct in this assertion if the squares in B_1 were to lie completely within the squares of B_2. We do not know this, and so we argue as follows.

Let S denote one of the squares of B_2, with radius $2e$ and let $\partial_r S$ denote those points within $(10/r) \cdot e$ of the boundary of S.

Lemma. *If a square S' from B_1 intersects $S \setminus \partial_r S$ then it is contained in S.*

Proof. This follows immediately from the fact that only maximal squares were used and the squares of B_2 come from a higher level than the ones in B_1. □

There are now two possibilities. Either most of B_1 lies within B_2 or the boundaries of the squares in B_2 contain a significant fraction of B_1. The latter possibility immediately gives $\lambda(B_2) \geq 2\lambda(B_1)$ while the former would give essentially $\lambda(B_2) \geq (\beta/\alpha)\lambda(B_1)$ except for small errors controlled by $1/r$ and $(8/9)^k$.

The next step is to apply EVC again to get B_3, a union of disjoint squares over which the averages of f are $\geq \beta$ with $\lambda(B_3)$ being not much less than $\lambda(B_2)$ so as not to lose what was gained in the previous step. After that all succeeding steps are carried out in the same way with a minimum gain of essentially β/α every two steps. Since we never can exceed a λ-value of 1, this gives the desired exponential upper bound for $E_{N,M}$.

After this qualitative discussion we must now specify the choice of constants. Let us do this in terms of the distance between β/α and 1, so set $\beta/\alpha - 1 = \frac{1}{K}$. Our expansion being only $1 + \frac{1}{K}$ we cannot afford to lose more than a $1/K$ fraction when using EVC and when seeing how much might be in $\partial_r S$. This dictates that r should be large compared to K, say $r = 10^3 \cdot K^2$ and n should be chosen in a similar way so that $10^{10} \cdot (8/9)^n = 1/K^2$. Finally this dictates a choice of L on the order of $10 \cdot K \log K$ so that the order of magnitude of c_2 as a function of K is

$$c_2 = \left(1 - \frac{1}{K}\right)^{c_3/K(\log K)^2}$$

where c_3 is some universal constant not less than 10^{-4}. This discussion gives the following theorem.

THEOREM. *If $(X, \mathcal{B}, \mu, \{T_u\colon u \in \mathbb{Z}^2\})$ is a probability preserving action of $\mathbb{Z}^2$, $f \geq 0$ a nonnegative $\mathcal{B}$-measurable function, and $\{A_n f(x)\}$ represents the averages of $f(T_u x)$ as u ranges over $[-n, n] \times [-n, n]$, then for universal constants c_1 and c_2 and for $\frac{\alpha}{\beta} = 1 - \frac{1}{K}$ we have*

$\mu\{x$: *the sequences $A_n f(x)$ fluctuates more than N-times across the interval*

$$(\alpha, \beta)\} \leq c_1 \left\{\left(1 - \tfrac{1}{K}\right)^{\frac{c_2}{K(\log K)^2}}\right\}^N.$$

3. Estimates for integrable functions

A numerical sequence $\{b_j\}$ is said to have N γ-fluctuations if there are increasing indices $n(1), n(2), \ldots, n(2N)$, such that for $1 \leq i \leq N$,

$$|b_{n(2i-1)} - b_{n(2i)}| \geq \gamma.$$

The advantage of γ-fluctuations over (α, β)-crossings is the observation that for the sum of two numerical sequences to have N γ-fluctuations at least one of the sequences must have at least $N/2$ $\gamma/2$-fluctuations. From this simple observation it follows that if we can get estimates for nonnegative L^1-functions then we automatically get essentially the same estimates for all L^1 functions. Now suppose that f is a nonnegative L^1-function and we want to estimate the size of the set where $A_n f(x)$ has N γ-fluctuations. For any level v we could use the maximal ergodic theorem to estimate the set

$$V = \{x\colon \text{ for some } n,\ A_n f(x) \geq v\}$$

by $\|f\|_1/v$. Then the interval $[0, v]$ could be divided into $2v/\gamma$ intervals of size $\gamma/2$ and then on at least one of them, say on (α, β), $\{A_n f(x)\}$ would have at least $\gamma/10 \cdot N/v$ downcrossings. Anyone of these could be estimated by our earlier results and it remains only to choose v in the best way possible. The worst case is when the (α, β) is at the very end in which case for $\beta/\alpha - 1$ we get essentially a constant divided by v. Recalling the estimate for the rate of exponential decay this dictates a choice of $N^{1/3}/(\log N)^3$ up to constant factors and we then get the following theorem:

THEOREM. *There is a constant c so that for any integrable function f in a measure preserving $\mathbb{Z}_2$ action, $(X, \mathcal{B}, \mu, T)$, for all $\gamma > 0$, $N \in \mathbb{N}$ we have*

$$\mu\{x : \{A_n f(x)\} \text{ has at least } N,\ \gamma\text{-fluctuations}\} \leq$$
$$\leq \left(c \cdot \frac{\|f\|_1}{\gamma}\right) / N^{1/3} (\log N)^{-3}.$$

For a fixed (α, β) we can improve the exponent $1/3$ to $1/2$ (but we still need the logarithmic terms). For this we must reexamine the proof that we gave for nonnegative functions. Since f is no longer nonnegative it is convenient to fix the interval (α, β) to be $(-1, +1)$. In order to find squares that grow at a definite rate so that the methods of EVC would be applicable we proceed as follows. Suppose that the number of fluctuations of $A_n f(x)$ across $(-1, +1)$ is M^2 and denote by $e_i(x)$, $1 \leq i \leq M^2$, the radii of the squares at level i. Divide the places where these crossing are taking place into M equal groups. For each such x there are two possibilities:

(I) $e_{(i+1)M}(x)/e_{iM+1}(x) \geq 2$ for all $0 \leq i < M$.
(II) $e_{(i+1)M}(x)/e_{iM+1}(x) < 2$ for all $0 \leq i < M$.

At least one of these must hold for at least half of the x's where there are more than M^2 fluctuations. In case this happens for (I), we will be able to use EVC. It turns out that case (II) is even easier! Focus on M successive fluctuations across $(-1, +1)$ where the size of the largest square is no more than twice the size of the smallest at each x. Using the very first step of EVC we can get disjoint squares of the Mth level that cover at least $1/9$ of the set of x's where we have more than M^2 fluctuations. On each such square we see that there must be a mass of $|f|$ that is proportional to M-times the area of the square, since each time we go between -1 and $+1$ or 1 and -1 new mass of $|f|$ must outbalance the previous average. This gives an immediate upper bound for the measure of such x's as

$$\frac{1}{9}\|f\|_1/M.$$

We have glided over the reduction that we discussed in detail before about how to estimate the measure of a set by computing averages along an orbit. Returning now to the dichotomy let us see what happens in case (I). Skipping every ℓ_0 steps we get an expansion rate of 2^{ℓ_0}. The proof now proceeds as before in the case of (α, β) fluctuations for nonnegative functions. There are several points to observe.

1. Our goal at each step is to show that there is a quantity of new mass of $|f|$ almost equal to the measure of the set B, where the fluctuations are taking place. Since the measure of B doesn't get bigger each time the amount of error that is allowable in EVC must now be small compared to M - the number of times the procedure will be carried out. This means that r should now be chosen large compared to the number of fluctuations. Such a requirement introduces only a further logarithmic term in the estimate.

2. If condition (1) of the EVC fails to hold, we get an increase in the measure of the set B, by simple disjointification. If this were to happen $\log N$ times we would get our target estimate for the measure of B directly. We can therefore assume without loss of generality that (1) always holds.

3. The argument involving maximal squares and the r-boundaries of the squares is valid here as well so that if we are not covering completely almost all of the squares of the previous stage there would be a definite increase in the measure of the set B. Once again this can only happen a logarithmic number of times and may be ignored.

References

[B] E. Bishop, *Foundations of constructive analysis*, McGraw-Hill, NY, 1967.

[Bo] J. Bourgain, *Pointwise ergodic theorem for arithmetic sets*, Publ. Math. IHES, **69**(1989), 5–45.

[CE] P. Collet and J. P. Eckmann, *Oscillations of observables in 1-dimensional lattice systems*, Mathematical Physics Electronics Journal **3** (1997), paper #3.

[D] L. Dubins, *Rises and upcrossings of nonnegative martingales*, Illinois J. of Math., **6**(1962), 226–241.

[JKRW] R. L. Jones, R. Kaufman, J. Rosenblatt, M. Wierdl, *Oscillation in ergodic theory*, Ergodic Theory Dynamical Systems **18** (1998), 889–935.

[Kh] A. G. Khachurovskii, *The rate of convergence in ergodic theorems*, Russian Math. Surveys **51**(1996), 653–703.

Steven Kalikow, School of Mathematics, Tel Aviv University, Ramat Aviv, 69978, Israel

72604.3435@compuserve.com

Benjamin Weiss, Institute of Mathematics, The Hebrew University of Jerusalem, Givat Ram, Jerusalem, 91904, Israel
weiss@math.huji.ac.il