THE BASIC ERGODIC THEOREMS, YET AGAIN

JAIRO BOCHI

ABSTRACT. A generalization of Rokhlin's Tower Lemma is presented. The Maximal Ergodic Theorem is then obtained as a corollary. We also use the generalized Rokhlin lemma, this time combined with a subadditive version of Kac's formula, to deduce a subadditive version of the Maximal Ergodic Theorem due to Silva and Thieullen.

In both the additive and subadditive cases, these maximal theorems immediately imply that "heavy" points have positive probability. We use heaviness to prove the pointwise ergodic theorems of Birkhoff and Kingman.

1. INTRODUCTION

Ergodic Theory is the subfield of Dynamical Systems concerned with measure-preserving dynamics, and it has applications throughout Mathematics. Its most fundamental result is the pointwise ergodic theorem proved by Birkhoff [Bi] in 1931. An important extension was obtained by Kingman [Ki] in 1968, and is known as the subadditive ergodic theorem. A multitude of other proofs of these basic results were obtained by many authors. One of the most popular methods of proof (especially in the case of Birkhoff's theorem) involves maximal inequalities, which have intrinsic interested by their own.

In this note, we provide self-contained proofs of Birkhoff's and Kingman's theorems by means of maximal inequalities; the only prerequisite is basic measure theory. Our approach has one novelty: it is based on a seemingly new extension of Rokhlin Tower Lemma, which is another basic tool used in many constructions in Ergodic Theory.

Let us proceed directly with the precise statements and proofs. We will provide further connections with the literature in the final Section 5.

Standing hypothesis: Let $(X, \mathcal{A}, \mu)$ be a Lebesgue probability space. Let $T\colon X \to X$ be an automorphism; this means that T and T^{-1} are measurable and preserve the measure μ. We assume that T is aperiodic, that is, the set of periodic points has zero measure.

2. A GENERALIZED ROKHLIN LEMMA

A measurable set $B \subseteq X$ is called *sweeping* if $\mu\left(\bigcup_{n \geqslant 0} T^n(B)\right) = 1$.

Theorem 2.1 (Generalized Rokhlin Lemma)**.** *For any $\varepsilon > 0$ and any measurable function $N\colon X \to \mathbb{Z}_+$, there exists a sweeping set $B \subseteq X$ such that*

   (i) *if $x \in B$ and $1 \leqslant i < N(x)$ then $T^i x \notin B$;*

(ii) $\int_B N \, \mathrm{d}\mu > 1 - \varepsilon$.

The case of constant N corresponds to the classical Rokhlin Lemma [Ro].

Let us introduce some useful terminology. If $B \subseteq X$ is any set of positive measure, Poincaré Recurrence Theorem says that a.e. $x \in B$ returns to B under iteration of T. So *return time* function

$$(2.1) \qquad R_B(x) := \min\{k \geqslant 1 \, ; \, T^k x \in B\} \quad \text{is finite for a.e. } x \in B.$$

If this function admits a lower bound n then the union $B \cup T(B) \cup \cdots \cup T^{n-1}(B)$ is disjoint; such a set is called a *tower* of *height* n with *base* B and *levels* B, $T(B)$, …, $T^{n-1}(B)$. A *skyscraper* is a countable union of disjoint towers; its base is defined as the union of the bases of the towers.

If $B \subseteq X$ is a set of positive measure, the *Kakutani skyscraper* with base B is the union of the towers C_i $(i = 1, 2, \dots)$ with respective bases $B_i := \{x \in B \, ; \, R_B(x) = i\}$. As a set, it equals $\bigcup_{n \geqslant 0} T^n(B)$. So the set B is sweeping if and only if the Kakutani skyscraper has full measure. In that case,

$$(2.2) \qquad \int_B R_B \, \mathrm{d}\mu = \sum_{i=1}^{\infty} i\mu(B_i) = \sum_{i=1}^{\infty} \mu(C_i) = 1;$$

this is *Kac's Lemma.*

A set B has property (i) in Theorem 2.1 if and only if $R_B \geqslant N$ on B. If this property is satisfied and moreover B is a sweeping set, then the following *error set*:

$$(2.3) \qquad E := \left\{ T^i x \, ; \, x \in B \text{ and } N(x) \leqslant i < R_B(x) \right\}.$$

has measure $\int_B (R_B - N) \, \mathrm{d}\mu$, which by Kac's Lemma equals $1 - \int_B N \, \mathrm{d}\mu$. So we can restate Theorem 2.1 in an equivalent form replacing conclusions (i) and (ii) by the following ones:

(i') $R_B \geqslant N$ on B;

(ii') the error set (2.3) has measure $\mu(E) < \varepsilon$.

Before proving Theorem 2.1, we need a preliminary result which is also used in the proof of the usual Rokhlin Lemma:

Lemma 2.2. *For any integer $m \geqslant 2$, there exists a sweeping set A with $R_A \geqslant m$.*

Proof. Since we are working on a non-atomic Lebesgue probability space, we can assume X is the unit interval and μ is Lebesgue measure. Since T is aperiodic, the function

$$\varphi(x) := \inf\{|T^j(x) - x| \, ; \, 1 \leqslant j < m\}$$

is positive a.e. As a consequence, any set E of positive measure contains another set F of positive measure such that $R_F \geqslant m$; indeed, it suffices to take a positive measure set $G \subseteq E$ where φ is bigger than some $\delta > 0$, and the take a positive measure subset $F \subseteq G$ with diameter less than this δ.

Now consider the family $\mathcal{F}$ formed by the sets $F \subseteq X$ of positive measure that satisfy $R_F \geqslant m$, partially ordered as follows: $F_1 \prec F_2$ if $F_1 \subseteq F_2$ and $\mu(F_2 \smallsetminus F_1) = 0$. Increasing chains are at most countable, and so by Zorn's lemma $\mathcal{F}$ contains a maximal element A.

We claim that A is sweeping. Indeed, the Kakutani skyscraper S with base A is a mod 0 invariant set, and if its complement S^{c} had positive measure then, using

the fact established at the beginning, we could find a positive measure set $F \subseteq S^{\mathsf{c}}$ such that $R_F \geqslant m$. Then $A \prec A \cup F$, contradicting the maximality of A.[1] $\qquad \square$

The following proof is due to Anthony Quas [Qu]:

Proof of Theorem 2.1. Fix an integer $n > 1$ such that the set $\{x \in X \; ; \; N(x) \geqslant n\}$, which we will call *bad set*, has measure less than $\varepsilon/2$. Fix another integer $m > 2n/\varepsilon$. By Lemma 2.2, there exists a sweeping set A such that $R_A \geqslant m$. Consider the Kakutani skyscraper with base A. The set $\bigsqcup_{i=1}^{n} T^{-i}(A)$, which we will call *penthouse*, has measure $n\mu(A) < n/m < \varepsilon/2$, and consists on the n topmost levels of the skyscraper.

Let us define the set B. For each point $x \in A$, we follow the steps:

1. If x is in the penthouse, then we stop.
2. If x is in the bad set, then we replace x with $T(x)$, and go back to step 1.
3. Otherwise (i.e. x is neither in the penthouse nor in the bad set), then we put x inside B, replace x with $T^{N(x)}(x)$, and go back to step 1.

The set B constructed in this way is clearly measurable and satisfies $R_B \geqslant N$. The associated error set (2.3) is contained in the union of the bad set and the penthouse, and therefore has measure less than ε. $\qquad \square$

3. The Maximal and Birkhoff's Ergodic Theorems

3.1. **Maximal Ergodic Theorem.** As a first application of the Generalized Rokhlin Lemma, we will give a short proof of the Maximal Ergodic Theorem.

The Birkhoff sums of f are denoted as:

$$f^{(n)} := f + f \circ T + \cdots + f \circ T^{n-1} \,.$$

Theorem 3.1 (Wiener, Yosida, and Kakutani's Maximal Ergodic Theorem)**.** *Let $f \in L^1(\mu)$. Let P be the set of $x \in X$ such that $f^{(n)}(x) > 0$ for some $n \geqslant 1$. Then $\int_P f \, \mathrm{d}\mu \geqslant 0$.*

Proof. Let $L := X \smallsetminus P$ be the set where all Birkhoff sums are non-positive. Define a function $N \colon X \to \mathbb{Z}_+$ as follows: if $x \in P$, let $N(x)$ be the least $n \geqslant 1$ such that $f^{(n)}(x) > 0$, while if $x \in L$, let $N(x) := 1$. Apply Theorem 2.1 to the function N and a small $\varepsilon > 0$, obtaining a measurable set B whose return time function is $\geqslant N$, and such that the error set (2.3) has measure $\mu(E) < \varepsilon$. Then:

$$\int_X f = \int_B \left[f^{(N(x))}(x) + \sum_{i=N(x)}^{R_B(x)-1} f(T^i x) \right] \mathrm{d}\mu(x) = \int_B f^{(N(x))}(x) \, \mathrm{d}\mu(x) + \int_E f \,,$$

by invariance of the measure. The integrand $f^{(N(x))}(x)$ equals $f(x)$ if $x \in L$ and is positive otherwise, and so $\int_X f \geqslant \int_{B \cap L} f + \int_E f$. But $\int_{B \cap L} f \geqslant \int_L f$, since $f \leqslant 0$ on L. So $\int_P f = \int_X f - \int_P f \geqslant \int_E f$. Since f is integrable and E can be made of arbitrarily small measure, we conclude that $\int_P f \geqslant 0$. $\qquad \square$

---

[1]Incidentally, this construction yields a set A whose Kakutani skyscraper has all towers with heights between m and $2m - 1$.

3.2. **Heaviness.** We present a corollary of the Maximal Ergodic Theorem 3.1 that is sufficient for some applications.

Let a, $b \in \mathbb{R}$. We say that a point $x \in X$ is a-*heavy* (resp. b-*light* with respect to a function f if for all $n \geqslant 1$, we have $f^{(n)}(x) \geqslant an$ (resp. $f^{(n)}(x) \leqslant bn$).

A measurable function $f \colon X \to [-\infty, +\infty]$ is called *quasi-integrable* if at least one of the functions f^+ or f^- is integrable (where, as usual, $f^+ := \max(f, 0)$ and $f^- := \max(-f, 0)$), and so $\int f = \int f^+ - \int f^-$ is defined.

Lemma 3.2 (Heaviness). *Let f be a quasi-integrable function, and let a, $b \in \mathbb{R}$.*

(i) *If $a < \int f \, d\mu$ then the set of a-heavy points has positive measure.*
(ii) *If $b > \int f \, d\mu$ then the set of b-light points has positive measure.*

Proof. By symmetry, it is sufficient to prove part (ii). Adding a constant to f, we can assume that $b = 0$, so $\int f \, d\mu$ is strictly negative. Let L be the set of points that are 0-light w.r.t. f. First consider the case of integrable f. The set P in the statement of the Maximal Ergodic Theorem 3.1 is the complement of L and therefore $\int_L f = \int_X f - \int_P f$ is strictly negative. In particular, $\mu(L) > 0$, as we wanted to show. Now consider the case that f is not integrable, so $\int f = -\infty$. For sufficiently large $K > 0$, the integrable function $f_* := \max(f, -K)$ has $\int f_* < 0$. By the previous case, its set L_* of 0-light points has positive measure. But $L \supseteq L_*$, so L has positive measure as well. $\qquad\square$

3.3. **Birkhoff's Pointwise Ergodic Theorem.** The *conditional expectation* of a quasi-integrable function f with respect to a sub-σ-algebra $\mathcal{B} \subseteq \mathcal{A}$ is the $\mathcal{B}$-measurable quasi-integrable function denoted $\mathbb{E}(f \mid \mathcal{B})$ such that $\int_B \mathbb{E}(f \mid \mathcal{B}) \, d\mu = \int_B f \, d\mu$ for every $B \in \mathcal{B}$. Existence and essential uniqueness are immediate consequences of the Radon–Nikodym Theorem.

Let $\mathcal{I} \subseteq \mathcal{A}$ denote the σ-algebra of T-invariant sets.

Theorem 3.3 (Birkhoff's Ergodic Theorem). *If f is a quasi-integrable function then $\dfrac{f^{(n)}}{n} \to \mathbb{E}(f \mid \mathcal{I})$ a.e.*

Proof. Define functions $g \leqslant h$ respectively as the the $\liminf$ and the $\limsup$ of the sequence $f^{(n)}/n$. It follows from the equality $f^{(n)} = f_1 + f^{(n-1)} \circ T$ that the functions g and h are invariant. Let $\varphi := \mathbb{E}(f \mid \mathcal{I})$, which by definition is also invariant. We want to prove that $g = \varphi = h$ a.e. Our plan is to prove the following inequalities:

$$(3.1) \qquad\qquad\qquad g \geqslant \varphi \geqslant h \quad \text{a.e.}$$

In order to prove the first inequality, it is sufficient to show that for all real numbers $\alpha < \beta$, the set

$$E_{\alpha,\beta} := \{x \in X \; ; \; g(x) < \alpha < \beta < \varphi(x)\}$$

has zero measure; indeed in that case the functions g and φ coincide over the full-measure set $\bigcap_{\alpha,\beta \in \mathbb{Q}, \alpha < \beta} E_{\alpha,\beta}^{\mathsf{c}}$. So let us assume by contradiction that $\mu(E_{\alpha,\beta}) > 0$ for certain numbers $\alpha < \beta$. Let

$$c := \frac{1}{\mu(E_{\alpha,\beta})} \int_{E_{\alpha,\beta}} f \, d\mu = \frac{1}{\mu(E_{\alpha,\beta})} \int_{E_{\alpha,\beta}} \varphi \, d\mu \geqslant \beta \,,$$

where the equality between the integrals is due to the fact that the set $E_{\alpha,\beta}$ is invariant. Applying Lemma 3.2.(i) to the measurable dynamical system $(T|E_{\alpha,\beta}, \frac{\mu|E_{\alpha,\beta}}{\mu(E_{\alpha,\beta})})$,

we conclude that for any real $a < c$, there is a positive measure set of points $x \in E_{\alpha,\beta}$ that are a-heavy. Such points clearly satisfy $g(x) \geqslant a$. Therefore $\alpha > a$. Since this holds for every $a < c$, we conclude that $\alpha \geqslant c \geqslant \beta$. This is a contradiction, and the first inequality in (3.1) is therefore proved.

The second inequality in (3.1) follows from the first one applied to $-f$. $\qquad\square$

4. Subadditive Ergodic Theorems

4.1. **Subadditivity.** A sequence $(a_n)_{n \geqslant 1}$ taking values in $\mathbb{R} \cup \{-\infty\}$ is called *subadditive* if

$$a_{n+k} \leqslant a_n + a_k \quad \text{for all } n,\, k \geqslant 1.$$

By a well-known exercise (sometimes called Fekete Lemma), the limit $\lim_{n\to\infty} \frac{a_n}{n}$ exists in $\mathbb{R} \cup \{-\infty\}$ and equals $\inf_n \frac{a_n}{n}$. We will denote it by:

$$\operatorname{linf} \frac{a_n}{n}\,.$$

A sequence $(f_n)_{n \geqslant 1}$ of measurable functions is called *subadditive* with respect to T if, for all $n,\, k \geqslant 1$,

$$f_{n+k} \leqslant f_n + f_k \circ T^n \quad \text{for all } n,\, k \geqslant 1.$$

Suppose that the positive part f_1^+ is μ-integrable. Then we define the *asymptotic average* of the subadditive sequence by:

$$\lambda := \operatorname{linf} \int \frac{f_n}{n}\, \mathrm{d}\mu\,.$$

4.2. **Subadditive Kac's Formula.** Given a sweeping set $B \subseteq X$, let $\hat{T} \colon B \to B$ be the first-return map. It preserves $\hat{\mu} := \frac{\mu|B}{\mu(B)}$, the normalized restriction of μ. Kac's formula (2.2) becomes $\int_B R_B\, \mathrm{d}\hat{\mu} = \frac{1}{\mu(B)}$.

Now suppose $(f_n)_{n \geqslant 1}$ is a subadditive sequence with respect to $T \colon (X, \mu) \leftrightarrow$, and $f_1^+ \in L^1(\mu)$. We define a sequence $(\hat{f}_n)_{n \geqslant 1}$ of functions on B by:

$$\hat{f}_n(y) := f_{R_B(y) + R_B(\hat{T}y) + \cdots + R_B(\hat{T}^{n-1}y)}(y)\,.$$

Then $(\hat{f}_n)_{n \geqslant 1}$ is a subadditive sequence with respect to $\hat{T}$; we call it the *induced subadditive sequence*. Note that $\hat{f}_1^+ \in L^1(\hat{\mu})$. Indeed, considering the Kakutani skyscraper and using invariance of μ, we see that in fact $\int_B \hat{f}_1^+\, \mathrm{d}\hat{\mu} \leqslant \frac{1}{\mu(B)} \int_X f_1^+\, \mathrm{d}\mu$.

Integrability allows us to define the asymptotic average of the induced subadditive sequence, that is,

$$\hat{\lambda} := \operatorname{linf} \int_B \frac{\hat{f}_n}{n}\, \mathrm{d}\hat{\mu}\,.$$

The next result relates the asymptotic averages $(\hat{f}_n)$: of the two subadditive sequences:

Theorem 4.1 (Subadditive Kac's Formula). $\hat{\lambda} = \dfrac{\lambda}{\mu(B)}\,.$

In the case of the additive sequence $f_n := n$, the result coincides with the usual Kac's formula.

In the case where the asymptotic average λ is a Lyapunov exponent, Theorem 4.1 appeared in [Wo, Lemma 2.2]; see also [Kn, Lemma 2.2].

Actually Theorem 4.1 is a easy consequence of Kingman's Subadditive Ergodic Theorem 4.5; this is the argument used in [Wo, Kn]. Here we go in the opposite direction; our ultimate aim is to prove Kingman's Theorem. So, to avoid circular reasoning, we should provide an independent proof of Theorem 4.1. Though this is possible, we won't do it, because the following weaker version is sufficient for our purposes:

Lemma 4.2 (Subadditive Kac Inequality). $\lambda \leqslant \int_B \hat{f}_1 \, d\mu.$

Proof. For each positive integer k, let $B_k := \{x \in B \; ; \; R_B(x) = k\}$, and $C_k := \bigsqcup_{j=0}^{k-1} T^j(B_k)$. So C_k is the tower of the Kakutani skyscraper with height k, and B_k is its base. Moreover, the B_k's form a mod 0 partition of B, and the C_k's form a mod 0 partition of X.

We claim that for every $m \geqslant 1$, the following inequality holds (the integrals being w.r.t. μ):

$$(4.1) \qquad \int_X f_m \leqslant \sum_{k=1}^m (m+1-k) \int_{B_k} \hat{f}_1 + \sum_{\ell=2}^\infty \min(\ell-1, m) \int_{C_\ell} f_1^+ \,.$$

In order to prove this, fix m and, for each point $x \in X$, consider all times $n_1 < n_2 < \cdots < n_p$ in the interval $\{0, 1, \ldots, m\}$ such that $T^{n_j} x \in B$. If $p \geqslant 1$ (i.e., the segment of orbit x, ..., $T^m x$ hits B at least once) then we use subadditivity to bound $f_m(x)$ by the following sum:

$$(4.2) \qquad \sum_{i=1}^{n_1-1} f_1^+(T^i x) + \sum_{j=1}^{p-1} \hat{f}_1(T^{n_j} x) + \sum_{i=n_p}^{m-1} f_1^+(T^i x) \,.$$

In the case that $p = 0$ (i.e., there are no hits), we bound $f_m(x)$ simply by

$$(4.3) \qquad \sum_{i=0}^{m-1} f_1^+(T^i x) \,.$$

Now we analyze the terms that appear in these sums:

- Given $k \geqslant 1$ and a non-periodic point $y \in B_k$, can a term $\hat{f}_1(y)$ appear in a sum (4.2)? The answer is clearly "no" if $k > m$. On the other hand, if $k \leqslant m$ then the term $\hat{f}_1(y)$ does appear in a sum (4.2): namely it appears once for each x in the set $\{y, T^{-1}y, \ldots, T^{-(m-k)}y\}$, and there is a total of $m - k + 1$ appearances.
- Similarly we ask: given $\ell \geqslant 1$ and a non-periodic point $z \in C_\ell$, how many times does a term $f_1^+(z)$ appear in a sum (4.2) or in a sum (4.3)? The answer is $\min(\ell-1, m)$. Indeed, if $\ell \geqslant m+1$ then the term $f_1^+(z)$ appears once for each x in the set $\{z, T^{-1}z, \ldots, T^{-(m-1)}z\}$, while if $\ell \leqslant m$ then $m+1-\ell$ of these points x do not contribute with a term of the form $f_1^+(z)$ and generate a term of the previous type $\hat{f}_1(z)$ instead.

Using these counts and the T-invariance of the measure μ, we obtain the claimed inequality (4.1). Next, note the following two facts about series, which follow from

Fatou's Lemma and Dominated Convergence Theorem, respectively:

$$b_k \in [-\infty, +\infty), \ \sum_{k=1}^{\infty} b_k^+ < +\infty \ \Rightarrow \limsup_{m \to \infty} \frac{1}{m} \sum_{k=1}^{m} (m - k + 1) b_k \leqslant \sum_{k=1}^{\infty} b_k \,;$$

$$c_\ell \in [0, +\infty), \quad \sum_{\ell=1}^{\infty} c_\ell < +\infty \quad \Rightarrow \lim_{m \to \infty} \frac{1}{m} \sum_{\ell=2}^{\infty} \min(\ell - 1, m) c_\ell = 0 \,.$$

It follows that the $\limsup$ of the right hand side of (4.1) divided by m is at most $\int_B \hat{f}_1$. But the left hand side of (4.1) divided by m tends to λ. So $\lambda \leqslant \int_B \hat{f}_1$, as we wanted to show. □

4.3. **Subadditive Maximal Ergodic Theorem.** The following result extends the Maximal Ergodic Theorem 3.1 to the subadditive context:

Theorem 4.3 (Silva and Thieullen's Maximal Subadditive Ergodic Theorem). *Let (f_n) be a subadditive sequence of functions satisfying the integrability condition $f_1^+ \in L^1(\mu)$, and let λ be its asymptotic average. Let H be the set of $x \in X$ such that $f_n(x) \geqslant 0$ for all $n \geqslant 1$. Then $\lambda \leqslant \int_H f_1 \, d\mu$.*

Actually, the result is not stated (nor named) exactly in this form by Silva and Thieullen, but it is a corollary of [ST, Lemma 2.4(a)].

Using the Generalized Rokhlin Theorem 2.1 and the Subadditive Kac Inequality (Lemma 4.2), Theorem 4.3 becomes almost obvious; its proof is of course similar to the proof of Theorem 3.1:

Proof of Theorem 4.3. Define a function $N \colon X \to \mathbb{Z}_+$ as follows: if $x \notin H$ then $N(x)$ is the least $n \geqslant 1$ such that $f_n(x) < 0$, while if $x \in H$ then $N(x) := 1$. Apply Theorem 2.1 to the function N and a small $\varepsilon > 0$, obtaining a measurable set B whose return time satisfies $R_B \geqslant N$ on B, and such that the error set (2.3) has measure $\mu(E) < \varepsilon$. Then (all integrals are w.r.t. μ):

$$\lambda \leqslant \int_B \hat{f}_1 \qquad\qquad\qquad \text{(by Lemma 4.2)}$$

$$\leqslant \int_B \left[f_{N(x)}(x) + \sum_{i=N(x)}^{R_B(x)-1} f_1^+(T^i x) \right] d\mu(x) \quad \text{(by subadditivity)}$$

$$= \int_B f_{N(x)}(x) \, d\mu(x) + \int_E f_1^+ \qquad\qquad \text{(by invariance)}$$

$$\leqslant \int_{B \cap H} f_1 + \int_E f_1^+ \qquad\qquad\qquad \text{(by definition of } N)$$

$$\leqslant \int_H f_1 + \int_E f_1^+ \qquad\qquad\qquad \text{(since } f_1 \leqslant 0 \text{ on } H).$$

Since $\mu(E)$ can be made arbitrarily small, so can $\int_E f_1^+$. Therefore $\lambda \leqslant \int_H f_1$. □

4.4. **Subadditive Heaviness.** Let (f_n) be a subadditive sequence of functions, and let $a, b \in \mathbb{R}$. We say that a point $x \in X$ is *a-heavy*, or respectively *eventually b-light*, with respect to the sequence (f_n) if

$$f_n(x) \geqslant an \quad \text{for all } n \geqslant 1, \text{ or respectively}$$

$$f_n(x) \leqslant bn \quad \text{for all sufficiently large } n \geqslant 1.$$

The following is the subadditive version of Lemma 3.2.[2]

Lemma 4.4 (Subadditive Heaviness). *Let (f_n) be a subadditive sequence of functions satisfying the integrability condition $f_1^+ \in L^1(\mu)$, and let λ be its asymptotic average. Let a, $b \in \mathbb{R}$.*

   (i) *If $a < \lambda$ then the set of a-heavy points has positive measure.*
   (ii) *If $b > \lambda$ then the set of eventually b-light points has positive measure.*

Proof. Replacing f_n by $f_n - an$, we can assume that $a = 0$ and so $\lambda > 0$. Let H be the set of 0-heavy points. By Theorem 4.3, $\int_H f_1 \geq \lambda > 0$; in particular, $\mu(H) > 0$, proving part (i).

Now let us prove part (ii). Suppose $b > \lambda$, and take $\varepsilon > 0$ such that $b - \varepsilon > \lambda$. Fix $m \geq 1$ such that $\int \frac{f_m}{m} < b - \varepsilon$. Let $\psi := \max(f_1^+, f_2^+, \ldots, f_{m-1}^+)$. By subadditivity,

$$f_n(x) \leq \sum_{i=0}^{\lfloor n/m \rfloor - 1} f_m(T^{mi}x) \; + \; \psi(T^{m\lfloor n/m \rfloor}x) =: \text{\textcircled{1}} \; + \; \text{\textcircled{2}}$$

We deal with these two terms as follows:

   • Let L be the set of points that that are $(b - \varepsilon)m$-light with respect to the function f_m and the dynamics T^m. So $\mu(L) > 0$ by Lemma 3.2.(ii). Note that for all $x \in L$ and all $n \geq 1$ we have $\text{\textcircled{1}} \leq (b - \varepsilon)n$.
   • Since $\psi \in L^1(\mu)$, by Birkhoff $\lim_{k \to \infty} \frac{1}{k} \varphi \circ T^k = 0$ a.e. [3] In particular, for almost every x and all sufficiently large n, we have $\text{\textcircled{2}} \leq \varepsilon n$.

It follows that for almost every $x \in L$ and all sufficiently large n, we have $f_n(x) \leq \text{\textcircled{1}} + \text{\textcircled{2}} \leq bn$. This proves part (ii). □

4.5. **Kingman's Subadditive Ergodic Theorem.** Finally, let us use Lemma 4.4 to prove the following fundamental result:

Theorem 4.5 (Kingman's Subadditive Ergodic Theorem). *Let (f_n) be a subadditive sequence of functions satisfying the integrability condition $f_1^+ \in L^1(\mu)$. Let $\varphi := \liminf \mathbb{E}\left(\frac{f_n}{n} \,\Big|\, \mathcal{I}\right)$. Then $\frac{f_n}{n} \to \varphi$ a.e.*

Proof. If E is an invariant (or mod 0 invariant) set of positive measure, let $\lambda(E)$ denote the asymptotic average of the subadditive sequence restricted to E, with respect to the restricted system $(T|E, \frac{\mu|E}{\mu(E)})$, that is,

$$\lambda(E) := \frac{1}{\mu(E)} \liminf \int_E \frac{f_n}{n} \, \mathrm{d}\mu \,.$$

We claim that

$$(4.4) \qquad\qquad \lambda(E) = \frac{1}{\mu(E)} \int_E \varphi \, \mathrm{d}\mu \,.$$

Indeed, on one hand, for every n we have $\mathbb{E}(\frac{f_n}{n} \mid \mathcal{I}) \geq \varphi$ and in particular $\int \frac{f_n}{n} \geq \int \varphi$; taking limits we obtain the $\geq$ inequality in (4.4). On the other hand, by subadditivity, each $\mathbb{E}(\frac{f_n}{n} \mid \mathcal{I})$ can be bounded from above by $\mathbb{E}(f_1^+ \mid \mathcal{I})$, which is a integrable function. Using Fatou's Lemma we obtain the $\leq$ inequality in (4.4).

---

[2]See the discussion in the Comments (Section 5) about the occurrence of Lemma 4.4 in the literature.

[3]For a simple proof of this fact that does not rely on Birkhoff, see [AB, Lemma 2].

The rest of the proof of Kingman's Theorem 4.5 is analogous to our proof of Birkhoff's Theorem 3.3, using Lemma 4.4 instead of Lemma 3.2.

Define functions $g \leqslant h$ respectively as the the $\liminf$ and the $\limsup$ of the sequence f_n/n. It follows from the inequality $f_n \leqslant f_1 + f_{n-1} \circ T$ that the functions g and h are *sub-invariant*, that is, $g \leqslant g \circ T$ and $h \leqslant h \circ T$. By invariance and finiteness of the measure μ, every sub-invariant function is a.e. invariant.

We must prove that $g = \varphi = h$ a.e., and our plan is to show that:

$$(4.5) \qquad\qquad g \geqslant \varphi \geqslant h \quad \text{a.e.}$$

Assume by contradiction that the inequality $g \geqslant \varphi$ fails on a positive measure set; then there exist real numbers $\alpha < \beta$ such that the (mod 0 invariant) set

$$E_{\alpha,\beta} := \{x \in X \; ; \; g(x) < \alpha < \beta < \varphi(x)\}$$

has positive measure. Applying Lemma 4.4.(i) to the system restricted to $E_{\alpha,\beta}$, we conclude that for any real $a < \lambda(E_{\alpha,\beta})$, there is a positive measure set of points $x \in E_{\alpha,\beta}$ that are a-heavy. Such points satisfy $g(x) \geqslant a$. Therefore $\alpha > a$. Since this holds for every $a < \lambda(E_{\alpha,\beta})$, we conclude that α is at least $\lambda(E_{\alpha,\beta})$, which by (4.4) is at least β. So $\alpha \geqslant \beta$, a contradiction. The first inequality in (4.5) is therefore proved.

The second inequality is proved similarly: Assume by contradiction that there are real numbers $\alpha < \beta$ such that the (mod 0 invariant) set

$$F_{\alpha,\beta} := \{x \in X \; ; \; \varphi(x) < \alpha < \beta < h(x)\}$$

has positive measure. Applying Lemma 4.4.(ii) to the system restricted to $F_{\alpha,\beta}$, we conclude that for any real $b > \lambda(F_{\alpha,\beta})$, there is a positive measure set of points $x \in F_{\alpha,\beta}$ that are eventually b-light. Such points satisfy $h(x) \leqslant a$. Therefore $\beta < b$. Since this holds for every $b > \lambda(F_{\alpha,\beta})$, we conclude that β is at most $\lambda(F_{\alpha,\beta})$, which by (4.4) is at most β. So $\beta \leqslant \beta$, a contradiction. This proves the second inequality in (4.5). □

5. COMMENTS

To summarize, our approach to prove Birkhoff's and Kingman's ergodic theorems was:

Generalized Rokhlin Lemma	$\Rightarrow$	maximal inequality	$\Rightarrow$	heaviness lemma	$\Rightarrow$	ergodic theorem

In the subadditive case, the first arrow also relies on a generalization of Kac's Lemma.

These intermediate results are also interesting by themselves. This path to the ergodic theorems is not the shortest one[4], but we hope that it has a gentle slope.

There are many extensions and variations of Rokhlin Lemma (see [We, Ko]), but nevertheless Theorem 2.1 appears to be new.

There are other proofs of the Maximal Ergodic Theorem 3.1 using towers: see [Pe, p.27ff]. Garsia's celebrated short proof uses a different idea; see e.g. [Pe, p.75ff].[5] Quoting Steele [St], *the proof has become a textbook standard, but the inequality*

---

[4]The proof in [KH, p. 136] is unbeatable.

[5]As made clear by Steele, Garsia's proof boils down to a pointwise inequality involving a coboundary: see inequality (3) in [St].

and its proof are widely regarded as mysterious. It is our hope that the Generalized Rokhlin Lemma makes the Maximal Ergodic Theorem more plain to see[6].

Silva and Thieullen deduce their Subadditive Ergodic Theorem [ST, Lemma 2.4(a)] (which implies Theorem 3.1 in this note) from a pointwise inequality. This type of proof is probably the shortest in this case, and appears in other proofs of the ergodic theorems [KW, Jo, KP, AB].[7]

While Kac's formula and Rokhlin Lemma also hold for non-invertible T (see [Kr, HJ]), it turns out that the generalized Rokhlin Lemma introduced here is false for non-invertible T: see Proposition 5.1 below. On the other hand, we can easily drop the invertibility assumption in the heaviness Lemmas 3.2 and 4.4, by considering the natural extension of T [Pe, p.13].

The "heaviness" terminology comes from Ralston [Ra] (who used it in a slightly different context). Lessa [Le] also uses heaviness (without this terminology) to prove Birkhoff's and Kingman's theorems. Lessa's work is perhaps the first place where the statement of Lemma 4.4 appears explicitly. Lemma 4.4.(i), which as we have seen follows immediately from Silva–Thieullen's result, is also contained in a deeper result by Karlsson and Margulis, namely [KM, Lemma 4.1]. Lemma 4.4.(ii) is [Le, Teorema 3.10].

Neither Karlsson–Margulis [KM] nor Lessa [Le] use maximal inequalities to obtain heaviness; instead they use Riesz' combinatorial lemma about leaders. See also Karlsson [Ka] for a related approach.

For another version of heaviness in a subadditive context, see [Mo, p. 144ff].

We conclude with the following example, also due to Quas, which shows that invertibility of T is necessary for the validity of Theorem 2.1:

Proposition 5.1 (Quas)**.** *Let T the shift on the space $X := \{1,2\}^{\mathbb{N}}$, and let μ be the $(\frac{1}{2}, \frac{1}{2})$-Bernoulli measure. Consider the function $N(x) := x_0$. Then for any measurable set $B \subseteq X$ such that $R_B \geqslant N$ on B, we have $\int_B (R_B - N)\, \mathrm{d}\mu \geqslant \frac{1}{9}$.*

Proof. If $\mu(B) < \frac{4}{9}$, then by Kac's formula $\int_B (R_B - N)d\mu = 1 - \int_B N d\mu \geqslant 1 - 2\mu(B) > \frac{1}{9}$. So let us assume that $\mu(B) \geqslant \frac{4}{9}$.

Consider $S := [1] \cap T^{-1}([2]) \cap T^{-2}(B)$, so that $\mu(S) = \frac{1}{4}\mu(B) \geqslant \frac{1}{9}$. Note that $T^{-1}(S) \subseteq B^{\mathsf{c}}$, as a consequence of the hypothesis $R_B \geqslant N$.

Now let $f := \mathbb{1}_B \cdot (R_B - N)$. We claim that for any $x \in S$ that returns to S in finite time $n := R_S(x)$, the Birkhoff sum $f^{(n)}(T(x)) = f(Tx) + \cdots + f(T^n x)$ is at least 1. Indeed, consider the biggest $k \in \{2, 3, \ldots, n\}$ such that $T^k(x) \in B$; such k exists because $n \geqslant 2$ and $T^2(x) \in B$. Since $T^{n+1}(x) \notin B$ and $T^{n+2}(x) \in B$, we have $R_B(T^k(x)) = n + 2 - k$. Now, if $k = n$ then $N(T^k(x)) = 1$, while if $k < n$ then $R_B(T^k(x)) \geqslant 3$. In either case, $f(T^k(x)) \geqslant 1$, proving the claim.

It follows the asymptotic average of f along almost every orbit is at least the frequency that the set S is visited. Since μ is ergodic, this means that $\int f\, \mathrm{d}\mu \geqslant \mu(B) \geqslant \frac{1}{9}$, as we wanted to prove. $\square$

---

[6] In the case of finite measure, at least.

[7] Incidentally, as remarked by Karlsson [Ka], Garsia's argument has a minor subadditive extension which is unfortunately insufficient to prove Kingman's theorem.

an extra step is required. I thank Anthony Quas for removing this assumption, for telling me that invertibility of T cannot be relaxed (Proposition 5.1), and for several comments and corrections. I thank Godofredo Iommi for stimulating conversations. Finally, I thank the referee for suggestions that improved the presentation.

References

[AB] Avila, A.; Bochi, J. – *On the subadditive ergodic theorem.* Manuscript, 2009. www.mat.uc.cl/~jairo.bochi/docs/kingbirk.pdf

[Bi] Birkhoff, G.D. – Proof of the ergodic theorem. *Proc. Nat. Acad. Sci. USA*, 17 (1931), 656–660.

[HJ] Heinemann, S.-M.; Schmitt, O. – Rokhlin's lemma for non-invertible maps. *Dynam. Systems Appl.* 10 (2001), no. 2, 201–213.

[Jo] Jones, R.L. – New proofs for the maximal ergodic theorem and the Hardy-Littlewood maximal theorem. *Proc. Amer. Math. Soc.* 87 (1983), no. 4, 681–684.

[Ka] Karlsson, A. – A proof of the subadditive ergodic theorem. *Groups, graphs and random walks*, 343–354, London Math. Soc. Lecture Note Ser., 436, Cambridge Univ. Press, Cambridge, 2017.

[KM] Karlsson, A.; Margulis, G. – A multiplicative ergodic theorem and nonpositively curved spaces. *Comm. Math. Phys.* 208 (1999), no. 1, 107–123.

[KH] Katok, A.; Hasselblatt, B. – *Introduction to the modern theory of dynamical systems.* Encyclopedia of Mathematics and its Applications, 54. Cambridge Univ. Press, Cambridge, 1995.

[KW] Katznelson, Y.; Weiss, B. – A simple proof of some ergodic theorems. *Israel J. Math.* 42 (1982), no. 4, 291–296.

[KP] Keane, M.; Petersen, K. – Easy and nearly simultaneous proofs of the ergodic theorem and maximal ergodic theorem. *Dynamics & stochastics*, 248–251. IMS Lecture Notes Monogr. Ser., 48. Inst. Math. Statist., Beachwood, OH, 2006.

[Ki] Kingman, J.F.C. – The ergodic theory of subadditive stochastic processes. *J. Roy. Statist. Soc. Ser. B*, 30 (1968), 499–510.

[Kn] Knill, O. – The upper Lyapunov exponent of $SL(2, \mathbb{R})$ cocycles: discontinuity and the problem of positivity. *Lyapunov exponents (Oberwolfach, 1990)*, 86–97, Lecture Notes in Math., 1486, Springer, Berlin, 1991.

[Ko] Kornfeld, I. – Some old and new Rokhlin towers. *Chapel Hill Ergodic Theory Workshops*, 145–169, Contemp. Math., 356, Amer. Math. Soc., Providence, RI, 2004.

[Kr] Krengel, U. – *Ergodic theorems.* De Gruyter Studies in Mathematics, 6. Walter de Gruyter & Co., Berlin, 1985.

[Le] Lessa, P. – *Teoremas ergódicos en espacios hiperbólicos.* Master's thesis, Universidad de la República, Montevideo (2009). www.cmat.edu.uy/~lessa/masters.html

[Mo] Morris, I.D. – Mather sets for sequences of matrices and applications to the study of joint spectral radii. *Proc. Lond. Math. Soc.* 107 (2013), no. 1, 121–150.

[Pe] Petersen, K. – *Ergodic theory.* Corrected reprint of the 1983 original. Cambridge Studies in Advanced Mathematics, 2. Cambridge Univ. Press, Cambridge, 1989.

[Qu] Quas, A. – mathoverflow.net/q/279635

[Ra] Ralston, D. – Heaviness: an extension of a lemma of Y. Peres. *Houston J. Math.* 35 (2009), no. 4, 1131–1141.

[Ro] Rokhlin, V. – A "general" measure-preserving transformation is not mixing. (Russian) *Doklady Akad. Nauk SSSR* 60, (1948), 349–351.

[ST] Silva, C.E.; Thieullen, P. – The subadditive ergodic theorem and recurrence properties of Markovian transformations. *J. Math. Anal. Appl.* 154 (1991), no. 1, 83–99.

[St] Steele, J.M. – Explaining a mysterious maximal inequality – and a path to the law of large numbers. *Amer. Math. Monthly* 122 (2015), no. 5, 490–494.

[We] Weiss, B. – On the work of V. A. Rokhlin in ergodic theory. *Ergodic Theory Dynam. Systems* 9 (1989), no. 4, 619–627.

[Wo] Wojtkowski, M. – Invariant families of cones and Lyapunov exponents. *Ergodic Theory Dynam. Systems* 5 (1985), no. 1, 145–161.

Facultad de Matemáticas, Pontificia Universidad Católica de Chile.