

Growth of normalizing sequences in limit theorems for conservative maps

Sébastien Gouëzel*

Abstract

We consider normalizing sequences that can give rise to nondegenerate limit theorems for Birkhoff sums under the iteration of a conservative map. Most classical limit theorems involve normalizing sequences that are polynomial, possibly with an additional slowly varying factor. We show that, in general, there can be no nondegenerate limit theorem with a normalizing sequence that grows exponentially, but that there are examples where it grows like a stretched exponential, with an exponent arbitrarily close to 1.

Keywords: infinite ergodic theory; limit theorems; normalization; transfer operator; renewal Markov chain.

AMS MSC 2010: 60F05; 37A40.

Submitted to ECP on March 29, 2018, final version accepted on November 12, 2018.

Supersedes arXiv:1803.11472.

Supersedes HAL : hal-01753889.

Let T be a non-singular map on a measure space (X, m) , and P a probability measure which is absolutely continuous with respect to m . Given a measurable function $f : X \rightarrow \mathbb{R}$, we say that its Birkhoff sums $S_n f = \sum_{k=0}^{n-1} f \circ T^k$ satisfy a limit theorem if there is a sequence $B_n \rightarrow \infty$ such that $S_n f / B_n$ converges in distribution with respect to P to a non-degenerate real random variable Z (by non-degenerate, we mean that Z is not constant almost surely). In this case, the asymptotic behavior of the sequence B_n and the limiting variable Z are defined uniquely, up to scalar multiplication (see for instance [5, Theorem 14.2]).

There are many examples of such non-degenerate limit theorems in the literature. Let us only mention the following ones:

1. Let Y_0, Y_1, \dots be a sequence of i.i.d. random variables which are in the domain of attraction of a stable law Z of index $\alpha \in (0, 1) \cup (1, 2]$, and are centered if integrable. Then, for some slowly varying function L , one has the convergence $(Y_0 + \dots + Y_{n-1}) / (n^{1/\alpha} L(n)) \rightarrow Z$. This can be recast in dynamical terms as follows. Let $X = \mathbb{R}^{\mathbb{N}}$ with the measure $m = \mathbb{P}_{Y_0}^{\otimes \mathbb{N}}$ and the left shift T . Define $f(x_0, x_1, \dots) = x_0$. Then $S_n f$ is distributed as $Y_0 + \dots + Y_{n-1}$, and $S_n f / (n^{1/\alpha} L(n))$ converges to Z . For $\alpha = 2$ and $L(n) = 1$, this is an instance of the classical central limit theorem. See for instance [6].

*Laboratoire Jean Leray, CNRS UMR 6629, Université de Nantes, 2 rue de la Houssinière, 44322 Nantes, France. E-mail: sebastien.gouezel@univ-nantes.fr

2. When T preserves m and m has infinite mass, one obtains different limits. Let $\alpha \in [0, 1)$ (we exclude $\alpha = 1$ to avoid degenerate limit laws). The Darling-Kac theorem gives examples of transformations T such that, for any function f which is integrable and has non-zero average for m , then $S_n f / (n^\alpha L(n))$ converges in distribution with respect to any probability measure $P \ll m$ to a Mittag-Leffler distribution of index α (where L is a slowly varying function). See [6, §8.11] and [1, 14]. This has been extended to functions with zero average by Thomine [15], where the normalization becomes $(n^\alpha L(n))^{1/2}$ and the limit distribution is modified accordingly. If one chooses carefully the system, one can obtain any nonnegative limit instead of the Mittag-Leffler distribution, see [3, Theorem 3].
3. In the same setting as in the second example, take a function $f = 1_E$ where E is a set of infinite measure. Thus, $S_n f / n$ records the proportion of time an orbit spends in E . Then one can devise such sets E for which $S_n f / n$ converges in distribution to generalized arcsine laws $\mathcal{L}_{\alpha, \beta}$, depending on α and an additional parameter $\beta \in (0, 1)$. (We exclude $\beta = 0$ and $\beta = 1$ to ensure that $\mathcal{L}_{\alpha, \beta}$ is nondegenerate). See [6, §8.6.2, §8.9.1] and [14]. Moreover, [2] includes an example where the normalization is of the form $c(n) = nL(n)$ where L is slowly varying and oscillates between 0 and 1.
4. Here is a nonergodic example. On $\mathbb{S}^1 \times \mathbb{S}^1$, consider the map $T(\alpha, x) = (\alpha, x + \alpha)$. Define a function $f(\alpha, x) = 1_{[0, r]}(x) - r$, where $r \in (0, 1)$ is fixed. Then, with respect to Lebesgue measure, $S_n f / \log n$ converges in distribution to a Cauchy random variable. This theorem is due to Kesten [12]. It has been generalized recently to higher dimensions by Dolgopyat and Fayad [9, 8]: in dimension d , replacing the characteristic function of the interval by the characteristic function of a nice subset \mathcal{C} and adding a randomization over r , they obtain limit theorems with normalizations $n^{(d-1)/2d}$ or $(\log n)^d$ depending on the geometric characteristics of \mathcal{C} .
5. It is possible to get any limiting distribution: Thouvenot and Weiss show in [16] that, for any aperiodic probability-preserving system T , and for any real random variable Z , there exists a function f such that $S_n f / n$ converges in distribution to Z . Moreover, if Z is nonnegative, one can also take f nonnegative if one is ready to replace the normalizing sequence by $nL(n)$ where L is slowly varying, see [3, Theorem 2].

These examples show that the possible limit theorems are extremely diverse, even in natural examples. However, in all these examples, the normalizing sequence is of the form $n^\alpha L(n)$, where $\alpha \in [0, \infty)$ and L is a slowly varying sequence. All such α are realized in some examples above. A theorem of Lamperti (see [6, Theorem 8.5.3]) shows that, under the stronger assumption that $S_{\lfloor nt \rfloor} / B_n$ converges in distribution for all $t > 0$, then the normalizing sequence has to be of this form $n^\alpha L(n)$. However, under the sole assumption that $S_n f / B_n$ converges, the behavior of B_n might possibly be more exotic. Our goal in this short note is to discuss this question, and in particular the possible growth of a normalizing sequence B_n .

It is easy to see that, in probability-preserving systems, a normalizing sequence has to grow at most polynomially. This is proved in Proposition 1.1. Our main result is that, in conservative systems, a normalizing sequence cannot grow exponentially (Theorem 2.1), but that it can grow almost exponentially: we exhibit examples of non-degenerate limit theorems with normalizing sequence e^{n^α} for any $\alpha \in (0, 1)$, in some specific (renewal) Markov chains.

1 The case of probability-preserving systems

Proposition 1.1. *Let T be a probability-preserving map on a space X , and $f : X \rightarrow \mathbb{R}$ a function such that $S_n f / B_n \rightarrow Z$ for some normalizing sequence $B_n \rightarrow \infty$, where Z is a real random variable which is not almost surely 0. Then $B_{n+1} / B_n \rightarrow 1$, and B_n grows at most polynomially, i.e., there exists C such that $B_n = O(n^C)$.*

The proof is closely related to the proof of the normalizing constant proposition in [3, Page 5].

Proof. One has $S_n f / B_{n-1} = S_{n-1} f / B_{n-1} + f \circ T^{n-1} / B_{n-1}$. Moreover, $f \circ T^{n-1} / B_{n-1}$ has the same distribution as f / B_{n-1} as the probability is T -invariant. Since $B_n \rightarrow \infty$, one gets that $f \circ T^{n-1} / B_{n-1}$ tends in probability to 0. Since $S_{n-1} f / B_{n-1}$ tends in distribution to Z , we deduce that $S_n f / B_{n-1} \rightarrow Z$. Since $S_n f / B_n \rightarrow Z$ and Z is non-degenerate, it follows from the convergence-of-types theorem (see [5, Theorem 14.2]) that $B_n / B_{n-1} \rightarrow 1$.

Since Z is not concentrated at 0, there exist $\varepsilon > 0$ and $\alpha > 0$ such that $\mathbb{P}(|Z| > \varepsilon) > 3\alpha$. Then, for large enough n , we have

$$\mu(|S_n f| > B_n \varepsilon) > 3\alpha. \tag{1.1}$$

Consider also L such that $\mathbb{P}(|Z| \geq L) < \alpha$. Then, for large enough, we have $\mu(|S_n f| < LB_n) > 1 - \alpha$. Since the measure is invariant, we also have $\mu(|S_n f|(T^n x) < LB_n) > 1 - \alpha$. On the intersection of these two events, $|S_{2n} f| < 2LB_n$, and moreover this happens with probability at least $1 - 2\alpha$. Since $|S_{2n} f| > B_{2n} \varepsilon$ with probability at least 3α by (1.1), these two events intersect, and we obtain $B_{2n} \varepsilon \leq 2LB_n$. This shows that there exists a constant C with $B_{2n} \leq CB_n$.

We have shown that $B_{2n} \leq CB_n$ and $B_{n+1} \leq CB_n$, for some constant C . We deduce that $B_{2^{r+1}} \leq C^2 B_{2^r}$ and $B_{2^r} \leq C^2 B_{2^{r-1}}$. It follows by induction on r that, for $n \in [2^r, 2^{r+1})$, one has $B_n \leq C^{2r} B_1$, which is bounded by $2B_1 \cdot n^{2 \log C / \log 2}$. This shows that B_n grows at most polynomially. \square

Remark 1.2. The proof shows that the polynomial growth exponent for B_n is bounded solely in terms of the distribution of the limiting random variable Z .

Remark 1.3. The assumption that $B_n \rightarrow \infty$ in Proposition 1.1 is merely for convenience. Indeed, without this assumption, one still has that B_{n+1} / B_n tends to 1 along integers n where $B_n \rightarrow \infty$ (or $B_{n+1} \rightarrow \infty$), with the same proof. In particular, there exists a constant D such that $B_{n+1} \leq 2B_n$ if $B_{n+1} \geq D$. One deduces that $B_n \leq D \cdot C^{2^r}$ for $n \in [2^r, 2^{r+1})$, by separating the case where $B_n < D$ (for which one can stop directly, without using the induction) and $B_n \geq D$ (for which one uses the inductive assumption as in the proof of Proposition 1.1). This shows again that B_n grows at most polynomially.

2 Subexponential growth for normalizing sequences in conservative systems

The proof of Proposition 1.1 relies crucially on the fact that the measure is invariant and has finite mass. In the opposite direction, if one does not assume any kind of recurrence, then one can get any behavior. For instance, consider the right shift on \mathbb{Z} and define the function $f(k) = 2^k$. Then $S_n f(x) / 2^n$ converges to 2^x . This means that we get a limit theorem, but which depends on the starting measure: Starting from $\delta_0 / 3 + 2\delta_1 / 3$ or from $2\delta_0 / 3 + \delta_1 / 3$, say, we get two different limit distributions, for the normalizing sequence 2^n . For another silly example, on $\mathbb{Z} \times \{-1, 1\}$, let $T(n, a) = (n + 1, a)$ and $f(n, a) = a2^{2^n}$. Then $S_n f / 2^{2^n - 1}$ converges in distribution with respect to

$P = (\delta_{(0,1)} + \delta_{(0,-1)})/2$ towards a Bernoulli random variable. These examples show that there is nothing interesting to say in the nonconservative case.

When the map is conservative, on the other hand, we can use the conservativity to obtain some rigidity. The next theorem shows in particular that there can be no nontrivial limit theorem with a normalizing sequence 2^n .

Theorem 2.1. *Let T be a conservative map on the measure space (X, m) . Consider a probability measure P which is absolutely continuous with respect to m , and a measurable function f . Assume that $B_n > 0$ is such that $S_n f/B_n$ converges in distribution with respect to P towards a limiting random variable Z which is not almost surely 0. Then, for any $\rho > 0$, one has $B_n = o(e^{\rho n})$.*

Remark 2.2. As a test case for the usability of proof assistants for current mathematical research, Theorem 2.1 and its proof given below have been completely formalized and checked in the proof assistant Isabelle/HOL, see the file `Normalizing_Sequences.thy` in [11]. In particular, the correctness of this theorem is certified.

The main intuition behind the proof of Theorem 2.1 is the following. Assume by contradiction that B_{n+1} is much larger than B_n . Then

$$S_{n+1}f(x)/B_{n+1} = f(x)/B_{n+1} + S_n f(Tx)/B_n \cdot B_n/B_{n+1}. \tag{2.1}$$

The term $f(x)/B_{n+1}$ is small in distribution. Moreover, $S_n f/B_n$ is tight, so when multiplied by B_n/B_{n+1} one should get something small, and get a contradiction with the fact that the limit Z is not a Dirac mass at 0. There is a difficulty: in (2.1), there is an additional composition by T in $S_n f(Tx)/B_n$. Still, this heuristic argument can be made into a rigorous proof that $B_{n+1} \leq C \max_{i \leq n} B_i$, for some $C > 0$, in Lemma 2.3. This excludes superexponential growth for B_n , but not exponential growth. To improve this bound, we will take advantage of the opposite decomposition $S_{n+1}f = S_n f + f \circ T^n$, to obtain another growth control in Lemma 2.5, that only holds along a subsequence of density 1 but in which one loses a multiplicative constant L that only depends on the distribution of Z . The point is that this estimate also applies to the system T^j and the sequence B_{jn} for any j , replacing in essence L with $L^{1/j}$ and making it arbitrarily close to 1. The theorem then follows from the combination of these two lemmas.

An important tool in the proof is the transfer operator \hat{T} , i.e., the predual of the composition by T , acting on $L^1(m)$. It satisfies $\int f \cdot g \circ T \, dm = \int \hat{T}f \cdot g \, dm$. We will use the following characterizations of conservativity: for any nonnegative f , then $\sum_{n=0}^{\infty} f(T^n x)$ and $\sum_{n=0}^{\infty} \hat{T}^n f(x)$ are both infinite for almost every x with $f(x) > 0$, see Propositions 1.1.6 and 1.3.1 in [1].

We start with a lemma asserting that the growth from B_n to B_{n+1} is uniformly bounded.

Lemma 2.3. *Under the assumptions of Theorem 2.1, there exists C such that, for all n , one has $B_{n+1} \leq C \max(B_0, \dots, B_n)$.*

The idea of the proof is that $S_n f(x)/B_n = S_k f(x)/B_n + S_{n-k} f(T^k x)/B_n$. The first term is small if n is large and k is fixed, while the second term should be small with high probability if B_n/B_{n-k} is large, as $S_{n-k} f/B_{n-k}$ is distributed like Z with respect to P . The difficulty is that there is a composition with T^k , under which P is not invariant, so one can not argue that $S_{n-k} f(T^k x)/B_{n-k}$ is distributed like Z . But one can still take advantage of the conservativity to make the argument go through.

Proof. Replacing m with $m+P$, we can assume without loss of generality that $h = dP/dm$ is bounded by 1. We can also assume that $\max(B_0, \dots, B_n)$ tends to infinity, since the conclusion is obvious otherwise. Multiplying f and Z by a constant, we can also assume $\mathbb{P}(|Z| > 2) > 3\alpha > 0$ for some $\alpha > 0$ as Z is not concentrated at 0.

By conservativity, $\sum_{k=1}^K h(T^k x)$ tends m -almost everywhere to $+\infty$ on the set $\{h > 0\}$, and therefore P -almost everywhere. Therefore, we can fix K such that $P(\sum_{k=1}^K h(T^k x) \geq 1) > 1 - \alpha$.

Let $\delta > 0$ be small enough so that $K\delta < \alpha$. Let $\varepsilon > 0$ be small enough so that, for any $k \leq K$ and for any measurable set U with $P(U) < \varepsilon$, one has

$$\int 1_U(x) \hat{T}^k h(x) \, dP(x) < \delta. \tag{2.2}$$

This is possible since the function $\hat{T}^k h(x)$ is integrable with respect to m (its integral is $\int h \, dm = P(X) = 1$), and therefore with respect to P as $P \ll m$.

Let $C > 0$ be such that $P(|Z| \geq C) < \varepsilon$. We will show that $B_n \leq C \max_{i < n} B_k$ for large enough n , by contradiction. Assume instead that n is large and $B_n > CB_k$ for any $k < n$. In particular, since $\max(B_0, \dots, B_{n-1})$ tends to infinity with n , this implies that B_n is very large. With probability at least 3α , we have $|S_n f / B_n| > 2$, as $P(|Z| > 2) > 3\alpha$ and $S_n f / B_n$ tends in distribution to Z . Moreover, $P(\forall k \leq K, |S_k f| / B_n \leq 1)$ is arbitrarily close to 1, as B_n is very large while K is fixed. In particular, it is $\geq 1 - \alpha$. This shows that the event

$$V = \left\{ x : \sum_{k=1}^K h(T^k x) \geq 1 \text{ and } |S_n f(x) / B_n| > 2 \text{ and } \forall k \in [1, K], |S_k f(x)| / B_n \leq 1 \right\}$$

has probability at least α , as it is the intersection of three sets whose measures are at least $1 - \alpha$ and 3α and $1 - \alpha$. For $x \in V$, we have for any $k \in [1, K]$ the inequality $|S_{n-k} f(T^k x)| / B_n \geq 1$, and therefore $|S_{n-k} f(T^k x)| / B_{n-k} > C$. We get

$$\alpha \leq P(V) = \int 1_V(x) \, dP(x) \leq \int \left(\sum_{k=1}^K h(T^k x) 1_{|S_{n-k} f| / B_{n-k} > C}(T^k x) \right) \, dP(x).$$

Writing $dP = h \, dm$ and changing variables by $y = T^k x$ in the k -th term of the sum, we obtain

$$\begin{aligned} \alpha &\leq \sum_{k=1}^K \int h(y) 1_{|S_{n-k} f| / B_{n-k} > C}(y) \cdot \hat{T}^k h(y) \, dm(y) \\ &= \sum_{k=1}^K \int 1_{|S_{n-k} f| / B_{n-k} > C}(y) \cdot \hat{T}^k h(y) \, dP(y). \end{aligned}$$

As $P(|Z| \geq C) < \varepsilon$, we have $P(|S_m f| / B_m > C) < \varepsilon$ for large enough m . Then the property of δ given in (2.2) ensures that $\int 1_{|S_{n-k} f| / B_{n-k} > C}(y) \cdot \hat{T}^k h(y) \, dP(y) \leq \delta$. This gives

$$\alpha \leq \sum_{k=1}^K \delta = K\delta.$$

This is a contradiction as $K\delta < \alpha$ by construction. □

If we could show that C in Lemma 2.3 can be taken arbitrarily close to 1, then Theorem 2.1 would follow directly. However, we do not know if this is true. What we will show instead is that such an inequality is true along a sequence of integers of density 1. This will be enough to conclude the proof. To proceed, we will need the following technical lemma, relying on conservativity.

Lemma 2.4. *Let T be a conservative dynamical system on the measure space (X, m) . Consider a finite measure P which is absolutely continuous with respect to m , and disjoint measurable sets A_n . Then $P(T^{-n} A_n)$ tends to 0 along a set of integers of density 1.*

Proof. Replacing m by a finite measure which is equivalent to it, we can assume that m is finite. Replacing P by $P + m$ (which only makes the conclusion stronger), we can also assume that P is equivalent to m . Finally, replacing m by $P/P(X)$, we can even assume that $P = m$, and that it is a probability measure.

Let $\varepsilon > 0$. Let L be a large constant. Let \hat{T} be the transfer operator associated to T , i.e., the adjoint (for the measure m) of the composition by T . We have $\sum_{j=0}^{k-1} \hat{T}^j 1(x) \rightarrow \infty$ almost everywhere, by conservativity. In particular, one can find a set U with $m(X \setminus U) \leq \varepsilon/2$ and an integer K such that, for all $x \in U$, one has $\sum_{j=0}^{K-1} \hat{T}^j 1(x) \geq L$. This gives in particular for all n

$$m(T^{-n}A_n) \leq \varepsilon/2 + m(T^{-n}A_n \cap U). \tag{2.3}$$

Let $n \geq K$. Since the sets A_{n-i} are disjoint, we have $\sum_{i=0}^{K-1} 1_{A_{n-i}}(T^n x) \leq 1$. Integrating this inequality, and applying the transfer operator \hat{T}^i to the i -th term in the sum, we obtain

$$\sum_{i=0}^{K-1} \int \hat{T}^i 1(x) \cdot 1_{A_{n-i}}(T^{n-i}x) \, dm(x) \leq 1.$$

Let us sum this inequality over $n \in [K, N-1]$. Changing variables $p = n-i$ and discarding the terms with $p < K$ or $p > N-K$, we obtain

$$\sum_{p=K}^{N-K} \int \left(\sum_{i=0}^{K-1} \hat{T}^i 1(x) \right) 1_{A_p}(T^p x) \, dm(x) \leq N.$$

Since $\left(\sum_{i=0}^{K-1} \hat{T}^i 1(x) \right) \geq L$ on U , this gives

$$\sum_{p=K}^{N-K} m(T^{-p}A_p \cap U) \leq N/L.$$

Therefore, thanks to (2.3),

$$\sum_{p=0}^{N-1} m(T^{-p}A_p) \leq N\varepsilon/2 + 2K + \sum_{p=K}^{N-K} m(T^{-p}A_p \cap U) \leq N\varepsilon/2 + 2K + N/L.$$

If L is large enough, this is bounded by $N\varepsilon$ for large N . We have proved that $m(T^{-p}A_p)$ tends to zero in Cesàro average. This is equivalent to convergence to zero along a density one set of integers (see for instance [17, Theorem 1.20]). \square

We can take advantage of the previous lemma to obtain a bound on B_{n+1}/B_n which is only true along a subsequence of density 1, but in which we only lose an explicit multiplicative constant. The difference between this lemma and Lemma 2.3 is that the constant L below only depends on the distribution of Z , unlike the constant C of Lemma 2.3. This means that it will be possible to apply this lemma to an iterate T^j of T with the same L , replacing in essence L with $L^{1/j}$ and making it arbitrarily close to 1.

Lemma 2.5. *Let T be a conservative dynamical system on the measure space (X, m) . Consider a probability measure P which is absolutely continuous with respect to m , and two measurable functions f and g . Assume that $(g + S_n f)/B_n$ converges in distribution with respect to P towards a limiting random variable Z which is not concentrated at 0. Then there exists $L > 1$ depending only on the distribution of Z such that $B_{n+1} \leq L \max(B_0, \dots, B_n)$ along a set of integers of density 1.*

Proof. Multiplying if necessary Z , f and g by a suitable constant that only depends on the distribution of Z , we can assume without loss of generality that $\mathbb{P}(|Z| > 2) = 3\alpha > 0$

for some $\alpha > 0$, as Z is not concentrated at 0. Choose $L > 3$ such that $\mathbb{P}(|Z| \geq L - 1) < \alpha$. Note that L only depends on the distribution of Z . Let $M_n = \max(B_0, \dots, B_n)$. Let us show that $\{n : B_{n+1} > LM_n\}$ has zero density. Let $A_n = \{x : |f(x)| \in (LM_n, LM_{n+1})\}$. These sets are disjoint as M is nondecreasing. Consider n such that $B_{n+1} > LM_n$.

Since $(g + S_n f)/B_n$ has a distribution close to that of Z , one has $P(|(g + S_n f)/B_n| < L) \geq 1 - \alpha$. Therefore, $P(|(g + S_n f)/B_{n+1}| < 1) \geq 1 - \alpha$ since $B_{n+1} > LB_n$. On the other hand, $P(|(g + S_{n+1} f)/B_{n+1}| \in [2, L - 1]) \geq 3\alpha - \alpha = 2\alpha$. Hence,

$$P\{x : |(g(x) + S_n f(x))/B_{n+1}| < 1 \text{ and } |(g(x) + S_{n+1} f(x))/B_{n+1}| \in [2, L - 1]\} \geq \alpha. \tag{2.4}$$

On this event, $f \circ T^n/B_{n+1} = (g + S_{n+1} f)/B_{n+1} - (g + S_n f)/B_{n+1}$ has an absolute value in $(1, L)$, i.e., $f(T^n x) \in (B_{n+1}, LB_{n+1})$. Since $B_{n+1} > LM_n$, this shows that $x \in T^{-n} A_n$. Hence, the event in (2.4) is contained in $T^{-n} A_n$. Since the probability of $T^{-n} A_n$ tends to 0 along a set of density 1 by Lemma 2.4, this shows that the inequality $B_{n+1} > LM_n$ can only hold along a set of density 0. \square

We can now conclude the proof of Theorem 2.1.

Proof of Theorem 2.1. Let $j > 0$ and $k \in [0, j)$. We claim that, along a subsequence of density 1, we have

$$B_{nj+k} \leq L \max(B_k, B_{j+k}, \dots, B_{(n-1)j+k}), \tag{2.5}$$

where $L = L(Z)$ is given by Lemma 2.5. Indeed, consider the conservative dynamical system $\tilde{T} = T^j$, the function $\tilde{f} = (S_j f) \circ T^k$, and the function $\tilde{g} = S_k f$. Denoting by \tilde{S} the Birkhoff sums for the transformation \tilde{T} , we have $\tilde{g} + \tilde{S}_n \tilde{f} = S_{nj+k} f$. Therefore, $(\tilde{g} + \tilde{S}_n \tilde{f})/B_{nj+k}$ converges in distribution to Z . Lemma 2.5 then implies (2.5).

Define $M_n = \max(B_0, B_1, \dots, B_{(n+1)j-1})$. Along a subsequence of density 1, all the inequalities (2.5) hold for $k \in [0, j)$. We get that $M_{n+1} \leq LM_n$ along a subsequence of density 1. Moreover, Lemma 2.3 shows that $M_{n+1} \leq C^j M_n$ for all n . We obtain $M_n \leq L^n \cdot (C^j)^{o(n)} M_0$. Therefore, for large enough n , we have $M_n \leq L^{2n} M_0$. In particular, for any $m \in [nj, (n+1)j)$, we get

$$B_m \leq M_n \leq L^{2n} M_0 \leq L^{2m/j} M_0.$$

The exponential growth rate of the term on the right hand side, in terms of m , can be made arbitrarily small by taking j large enough. \square

3 An example with a stretched exponential normalizing sequence

Let us now describe an example showing that Theorem 2.1, which excludes the exponential growth of normalizing sequences, is almost sharp. Indeed, for any $\alpha < 1$, we construct a conservative dynamical system (more specifically a renewal Markov chain, which preserves an infinite measure) and a function $f > 0$ such that $S_n f/e^{n^\alpha}$ converges in distribution to a non-degenerate limit.

Let

$$p_n = c/(n(\log n)^2) \tag{3.1}$$

for $n \geq 2$ and $p_0 = p_1 = 0$, where c is adjusted so that p is a probability distribution. We consider a recurrent Markov chain on a set A for which the excursion away from a reference point 0 has length n with probability p_n , and for which all trajectories from 0 to any point x , avoiding 0 after time 1, have the same number of steps $h(x)$. We say that $h(x)$ is the *height* of x .

The simplest such example is on the set $A = \{0\} \cup \{(n, i) : n > 1, i \in [1, n-1]\}$ in which one jumps from 0 to $(n, 1)$ with probability p_n , from (n, i) to $(n, i+1)$ with probability 1 if

$i < n - 1$, and from $(n, n - 1)$ to 0 with probability 1. The height of (n, i) is i . One can also realize such an example on $A = \mathbb{N}$, with jumps from n to 0 with probability q_n and from n to $n + 1$ with probability $1 - q_n$, where q_n is defined inductively by the relation $p_{n+1} = (1 - q_0) \cdots (1 - q_{n-1})q_n$. The height of n is just n in this example. This example is called a renewal Markov chain, see [7, 1].

Let X be the space of possible trajectories of the Markov chain, m the corresponding (infinite) measure on trajectories, $\pi : X \rightarrow A$ the projection associating to a trajectory its starting point, and T the shift map forgetting the first point of the trajectory. It is conservative as the Markov chain is recurrent. Abusing notations, if g is a function on A , we will write $g(x)$ for $g(\pi x)$. For instance, $h(x)$ will denote the height of the starting point of the trajectory x .

Let P be the Markov measure on paths starting from 0. Then, with respect to P , the height $h(T^n x)$ is by definition distributed like $h(X_n)$ where X_n is the random walk starting from 0. Since p_n in (3.1) tends very slowly to 0, one expects most excursions away from 0 to be extremely long, so that $h(X_n)$ should be of the order of magnitude of n with large probability. This is true, but we will need the following more precise estimate.

Lemma 3.1. *When $k \rightarrow \infty$ and $n - k \rightarrow \infty$, one has*

$$P(h(T^n x) = k) \sim \frac{1}{(\log k) \cdot (n - k)}.$$

Proof. Denote by Y_i the distribution of the length of the i -th excursion away from 0 of the random walk. These variables are independent, and share the same distribution Y given by $\mathbb{P}(Y = s) = p_s$. Then X_n is at height k exactly if there is an index j such that $Y_1 + \cdots + Y_j = n - k$ and moreover $Y_{j+1} > k$. Since the distribution of the excursion after $n - k$ is independent of the fact that $n - k$ was reached by such a sum, we get

$$P(h(T^n x) = k) = \mathbb{P}(Y > k) \cdot \mathbb{P}(\exists j, Y_1 + \cdots + Y_j = n - k).$$

The last event in this equation is a *renewal event*. Local asymptotics of renewal probabilities are known for p distributed as in (3.1): by [13] (see also [4]), one has when $s \rightarrow \infty$

$$\mathbb{P}(\exists j, Y_1 + \cdots + Y_j = s) \sim \frac{\mathbb{P}(Y = s)}{\mathbb{P}(Y > s)^2}.$$

Since $\mathbb{P}(Y > s) \sim c/\log s$, we obtain

$$P(h(T^n x) = k) \sim \frac{c}{\log k} \cdot \frac{c/(n - k)(\log(n - k))^2}{(c/\log(n - k))^2} = \frac{1}{(\log k) \cdot (n - k)}. \quad \square$$

Lemma 3.2. *Let $\beta \in (0, 1)$. When n tends to infinity,*

$$P(h(T^n x) \in [n - n^\beta, n]) \rightarrow \beta.$$

Proof. Let us first note that, with high probability, $k = h(T^n x)$ is bounded away from 0 and n . Indeed, if $C > 0$ is fixed, the approximations for the probabilities in Lemma 3.1 satisfy

$$\sum_{k=n/2}^{n-C} \frac{1}{(\log k) \cdot (n - k)} \sim \frac{1}{\log n} \sum_{k=n/2}^{n-C} \frac{1}{n - k} \sim \frac{1}{\log n} \sum_{m=C}^{n/2} \frac{1}{m} \rightarrow 1.$$

Thanks to Lemma 3.1, this shows that the contribution of this range of values of $h(T^n x)$ has probability close to 1. Therefore, we may without loss of precision replace $P(h(T^n x) = k)$ with the approximating probabilities given in this lemma, and even by $1/((\log n) \cdot (n - k))$.

Therefore,

$$P(h(T^n x) \in [n - n^\beta, n]) \sim \frac{1}{\log n} \sum_{k=n-n^\beta}^{n-1} \frac{1}{n-k} = \frac{1}{\log n} \sum_{m=1}^{n^\beta} \frac{1}{m} \sim \frac{\log(n^\beta)}{\log n} = \beta. \quad \square$$

Let $\alpha \in (0, 1)$. Let $s(n) = e^{(n+1)^\alpha} - e^{n^\alpha} \geq 0$. This is equivalent to $\tilde{s}(n) = \alpha n^{\alpha-1} e^{n^\alpha}$ at infinity, so we could use \tilde{s} instead of s in what follows, but the point of the construction is clearer with s . Define a function f by $f(x) = s(h(x)) \geq 0$. In this way, the sum of the values of f along an excursion up to height $k - 1$ is $e^{k^\alpha} - 1$.

Theorem 3.3. *The sequence $S_n f(x)/e^{n^\alpha}$ converges in distribution with respect to P towards a non-trivial random variable, equal to 0 with probability α and to 1 with probability $1 - \alpha$.*

Proof. Let $\varepsilon > 0$. We claim that, if $h(T^n x) \in [n - n^{1-\alpha-\varepsilon}, n]$ (which happens asymptotically with probability $1 - \alpha - \varepsilon$ by Lemma 3.2) then $S_n f(x)/e^{n^\alpha}$ is close to 1. Moreover, we claim that, if $h(T^n x) \in [n - n^{1-\varepsilon}, n - n^{1-\alpha+\varepsilon}]$ (which happens asymptotically with probability $(1 - \varepsilon) - (1 - \alpha + \varepsilon) = \alpha - 2\varepsilon$ by Lemma 3.2), then $S_n f(x)/e^{n^\alpha}$ is close to 0. As the probabilities of these two events add up to one to within 3ε , this shows the desired convergence in distribution by letting ε tend to 0.

It remains to prove the claims. For this, consider $k = n - n^\beta$ for some β . Then

$$k^\alpha = n^\alpha (1 - n^{\beta-1})^\alpha = n^\alpha - \alpha n^{\alpha+\beta-1} (1 + o(1)).$$

Therefore,

$$\frac{e^{k^\alpha}}{e^{n^\alpha}} = e^{-\alpha n^{\alpha+\beta-1} (1+o(1))} \rightarrow \begin{cases} 0 & \text{if } \beta > 1 - \alpha, \\ 1 & \text{if } \beta < 1 - \alpha. \end{cases} \quad (3.2)$$

Let us now prove the claims. Assume first that $h(T^n x) \in [n - n^{1-\alpha-\varepsilon}, n]$. Then the sum of the values of f along the last excursion up to time $n - 1$ is equal to $\sum_{j=0}^{h(T^n x)-1} s(j) = e^{h(T^n x)^\alpha} - 1$. By (3.2) with $\beta < 1 - \alpha - \varepsilon$, we deduce that this sum divided by e^{n^α} is close to 1. We should then add the contributions of the first excursions. Since their total length is bounded by $n^{1-\alpha-\varepsilon}$, the maximal height they could have reached is $n^{1-\alpha-\varepsilon}$, and their total contribution is at most $n \cdot e^{(n^{1-\alpha-\varepsilon})^\alpha}$. This is negligible. This proves the first claim.

The second claim, for $h(T^n x) \in [n - n^{1-\varepsilon}, n - n^{1-\alpha+\varepsilon}]$, is proved analogously. Indeed, by (3.2) with $\beta > 1 - \alpha + \varepsilon$, the contribution of the last excursion is negligible. And the contribution of the other excursions is bounded by $n e^{n^{(1-\varepsilon)\alpha}}$ and is also negligible. This concludes the proof. \square

Remark 3.4. In our example, the sequence $p_n = c/(n(\log n)^2)$ tends very slowly to 0. If one tries to build a less extreme example, by taking $p_n = c/n^\gamma$ for some $\gamma > 1$, then the construction fails. Indeed, one can check that in this case $h(T^n x)$ is distributed over the whole interval $[0, n]$ (instead of being very concentrated around n). More precisely, $h(T^n x)/n$ converges in distribution to a random variable with support equal to $[0, 1]$ (this follows from direct computations, or from the arcsine law for waiting times [14]). Then $e^{h(T^n x)^\alpha}/e^{n^\alpha}$ converges in distribution to 0, giving a degenerate limit theorem.

Remark 3.5. The above example is quite flexible. For instance, instead of obtaining the limit $\alpha\delta_0 + (1 - \alpha)\delta_1$, one can obtain any limiting distribution Z which is equal to 0 with probability α and to Z' with probability $(1 - \alpha)$ where Z' is an arbitrary real random variable, with an ergodic conservative map, as follows. We describe it in the setting of the random walk on \mathbb{N} above, with return probabilities q_n from n to 0. Replace the space \mathbb{N} of the random walk with $\mathbb{N} \times \mathbb{R}$, with the following transition probabilities. Jump from (n, t) to $(n + 1, t)$ with probability $1 - q_n$, and from (n, t) to $\{0\} \times \mathbb{R}$ with probability

$q_n \otimes \mathbb{P}_{Z'}$ where $\mathbb{P}_{Z'}$ is the distribution of Z' . If one starts from $P = \delta_0 \otimes \mathbb{P}_{Z'}$, then the first component of this new random walk is distributed like the original random walk on \mathbb{N} , and the second component is always distributed like Z' , with independence every times the walk returns to 0. Define a function $g(n, t) = te^{n\alpha}$, and let $f(x) = g(Tx) - g(x)$. Then the Birkhoff sum $S_n f(x)$ is equal to $te^{k\alpha} - g(x)$ if $\pi(T^n x) = (k, t)$. The proof of Theorem 3.3 then shows that $S_n f / e^{n\alpha}$ converges to Z .

Remark 3.6. One can also construct an example where $S_n f / B_n$ converges to a non-trivial distribution, but B_{n+1} / B_n does not tend to 1. Consider the same example as in Theorem 3.3, but where the probabilities p_n are nonzero only for even n : there is a periodicity phenomenon in this Markov chain, of which we will take advantage. Define $g(n) = e^{n\alpha}$ for even n , and $g(n) = e^{n\alpha} / 2$ for odd n . Let $f(x) = g(Tx) - g(x)$. Then $S_n f / B_n$ converges in distribution with respect to P towards $Z = \alpha\delta_0 + (1 - \alpha)\delta_1$, where $B_n = e^{n\alpha}$ for even n and $B_n = e^{n\alpha} / 2$ for odd n . In this example, if one denotes by Q the measure on trajectories starting at 1, then $S_n f / B_n$ does not converge to Z with respect to Q , while it does with respect to P , whereas P and Q are both absolutely continuous with respect to the invariant infinite measure m . This phenomenon does not happen for probability-preserving maps, by Eagleson's Theorem [10]. This shows that Eagleson's Theorem is only valid in full generality for probability-preserving maps, and not for conservative maps in general. Zweimüller has proved in [18] that Eagleson's Theorem holds for Birkhoff sums in conservative maps under an additional assumption of asymptotic invariance, which in our setting translates to the fact that $f \circ T^n / B_n$ converges to 0 in distribution. Indeed, this is not the case in the previous counterexample.

References

- [1] Jon Aaronson, *An introduction to infinite ergodic theory*, Mathematical Surveys and Monographs, vol. 50, American Mathematical Society, Providence, RI, 1997. MR-MR1450400
- [2] Jon Aaronson, Maximilian Thaler, and Roland Zweimüller, *Occupation times of sets of infinite measure for ergodic transformations*, Ergodic Theory Dynam. Systems **25** (2005), no. 4, 959–976. MR-2158393
- [3] Jon Aaronson and Benjamin Weiss, *Distributional limits of positive, ergodic stationary processes and infinite ergodic transformations*, Ann. Inst. Henri Poincaré Probab. Stat. **54** (2018), no. 2, 879–906. MR-3795070
- [4] Kenneth S. Alexander and Quentin Berger, *Local limit theorems and renewal theory with no moments*, Electron. J. Probab. **21** (2016), Paper No. 66, 18. MR-3580032
- [5] Patrick Billingsley, *Probability and measure*, third ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1995, A Wiley-Interscience Publication. MR-MR1324786
- [6] Nicholas H. Bingham, Charles M. Goldie, and Jozef L. Teugels, *Regular variation*, Encyclopedia of Mathematics and its Applications, vol. 27, Cambridge University Press, Cambridge, 1987. MR-MR898871
- [7] Kai Lai Chung, *Markov chains with stationary transition probabilities*, Second edition. Die Grundlehren der mathematischen Wissenschaften, Band 104, Springer-Verlag New York, Inc., New York, 1967. MR-0217872
- [8] Dmitry Dolgopyat and Bassam Fayad, *Deviations of ergodic sums for toral translations II. Boxes*, Preprint, 2012.
- [9] Dmitry Dolgopyat and Bassam Fayad, *Deviations of ergodic sums for toral translations I. Convex bodies*, Geom. Funct. Anal. **24** (2014), no. 1, 85–115. MR-3177379
- [10] G. K. Eagleson, *Some simple conditions for limit theorems to be mixing*, Teor. Veroyatnost. i Primenen. **21** (1976), 653–660. MR-MR0428388
- [11] Sebastien Gouezel, *Ergodic theory*, Archive of Formal Proofs (2018), http://devel.isa-afp.org/entries/Ergodic_Theory.html, Formal proof development.

- [12] Harry Kesten, *Uniform distribution mod 1. II*, Acta Arith **7** (1961/1962), 355–380. MR-0142532
- [13] S. V. Nagaev, *A renewal theorem without power moments*, Teor. Veroyatn. Primen. **56** (2011), no. 1, 188–197. MR-2848426
- [14] Maximilian Thaler and Roland Zweimüller, *Distributional limit theorems in infinite ergodic theory*, Probab. Theory Related Fields **135** (2006), no. 1, 15–52. MR-2214150
- [15] Damien Thomine, *A generalized central limit theorem in infinite ergodic theory*, Probab. Theory Related Fields **158** (2014), no. 3-4, 597–636. MR-3176360
- [16] Jean-Paul Thouvenot and Benjamin Weiss, *Limit laws for ergodic processes*, Stoch. Dyn. **12** (2012), no. 1, 1150012, 9. MR-2887924
- [17] Peter Walters, *An introduction to ergodic theory*, Graduate texts in mathematics, vol. 79, Springer, 1982.
- [18] Roland Zweimüller, *Mixing limit theorems for ergodic transformations*, J. Theoret. Probab. **20** (2007), 1059–1071. MR-MR2359068

Electronic Journal of Probability

Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS¹)
- Easy interface (EJMS²)

Economical model of EJP-ECP

- Non profit, sponsored by IMS³, BS⁴, ProjectEuclid⁵
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund⁶ (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

¹LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

²EJMS: Electronic Journal Management System <http://www.vtex.lt/en/ejms.html>

³IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

⁴BS: Bernoulli Society <http://www.bernoulli-society.org/>

⁵Project Euclid: <https://projecteuclid.org/>

⁶IMS Open Access Fund: <http://www.imstat.org/publications/open.htm>