Ergod. Th. & Dynam. Sys. (2011), **31**, 351–367
doi:10.1017/S0143385709001217

The law of series

T. DOWNAROWICZ† and Y. LACROIX‡

† *Institute of Mathematics and Computer Science, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland*
(e-mail: downar@pwr.wroc.pl, downar@im.pwr.wroc.pl)
‡ *Institut des Sciences de l'Ingénieur de Toulon et du Var, Avenue G. Pompidou, B.P. 56, 83162 La Valette du Var Cedex, France*
(e-mail: yves.lacroix@univ-tln.fr)

(*Received* 8 *June* 2009 *and accepted in revised form* 9 *December* 2009)

To Dan Rudolph

Abstract. We consider an ergodic process on finitely many states, with positive entropy. Our main result asserts that the distribution function of the normalized waiting time for the first visit to a small (i.e., over a long block) cylinder set B is, for majority of such cylinders and up to epsilon, dominated by the exponential distribution function $1 - e^{-t}$. That is, the occurrences of B along the time axis can appear either with gap sizes of nearly the exponential distribution (as in an independent and identically distributed process), or they attract each other and create 'series'. We recall that in [T. Downarowicz, Y. Lacroix and D. Leandri. Spontaneous clustering in theoretical and some empirical stochastic processes. *ESAIM Probab. Stat.* to appear] it is proved that in a typical (in the sense of category) ergodic process (of any entropy), all cylinders B of selected lengths (such lengths have upper density 1 in $\mathbb{N}$) reveal strong attracting. Combining this with the result of this paper, we obtain, globally in ergodic processes of positive entropy and for long cylinder sets, the prevalence of attracting and deficiency of repelling. This phenomenon resembles what in real life is known as *the law of series*; the common-sense observation that a rare event, having occurred, has a mysterious tendency to untimely repetitions.

1. *Introduction*

We prove a new theorem about ergodic processes with positive entropy, in the same category as the Shannon–McMillam–Breiman theorem or the theorem of Ornstein and Weiss [**OW**] (relating the return times of long blocks to entropy); we prove the L^1 convergence of a certain 'information-like' parameter associated with cylinder sets. (We leave as an open question whether our assertion can be strengthened to an almost everywhere statement.) Unlike in the two theorems mentioned above, our parameter neither describes an 'exponential rate' associated with longer and longer cylinder sets, nor provides (in the limit) an alternative formula for the entropy. Instead, it deals directly with the structure of the return times to the cylinder B. Our result implies that for the 'majority' of cylinders B, the return times to B may deviate from the behavior typical of an independent and identically distributed (i.i.d.) process in only one direction: toward stronger clustering (we call such deviation 'attracting'), while the opposite deviation ('repelling') cannot occur except for very few cylinders. This is a discovery of a new phenomenon common to all processes of positive entropy, an immense class for which one might expect all general properties of a similar kind to have been established a long time ago.

Our result complements another, proved in [**DLL**]. Combining these results together, one obtains the following picture of a typical process of positive entropy: if B is a long cylinder of a selected length (belonging to a rather large subset of $\mathbb{N}$; the upper density of this subset is 1) then B reveals very strong attracting (this is what [**DLL**] says). For other lengths, the majority of blocks either also attract, or at least do not repel (in which case they behave like in an i.i.d. process). Only extremely few blocks may exhibit the repelling behavior. Globally, for long cylinders there is an abundance of attracting and deficiency of repelling.

Here is an illustrative example: consider the i.i.d. process of coin tossing. It is clear that for any fixed long block B (string of zeros and ones) the gap lengths between its occurrences have exponential distribution. This means that the clustering is what we call 'neutral', i.e., we see long gaps and short gaps without any 'surprising' disproportion. Now suppose that we slightly perturb the generating 0–1 partition. This can be done by applying any kind of procedure causing errors in the perception of the outcomes, i.e., sometimes the heads are erroneously taken for tails and vice versa. The result of this work yields that, no matter how the process is perturbed, the majority of long blocks B will either continue to exhibit neutral clustering or tend toward stronger clustering. The result of [**DLL**] asserts that 'most likely' (in the sense of category), in the perturbed process many cylinders will choose the second option, moreover, in an extremely strong form. A typically perturbed (no matter how slightly) i.i.d. process of coin tosses is an example of a positive entropy process, where the distribution of repetitions of a long block B is (for many blocks) completely unlike in the i.i.d. process. In [**DLL**] we provide an explicit construction of such a perturbation.

We remark that the level of intricacy in proving the current result is much higher than in [**DLL**], addressing only typical processes, and cylinders of selected lengths. That result alone does not exclude repelling for cylinders of other lengths. In this work we prove a completely general property, valid for all processes of positive entropy and nearly all cylinders of all sufficiently large lengths. The proof is entirely contained within the classics of ergodic theory; it relies on basic facts on entropy for partitions and sigma-fields, some elements of the Ornstein theory (ϵ-independence), the Shannon–McMillan–Breiman theorem, the Ornstein–Weiss theorem, the ergodic theorem, basics of probability and calculus.

The result can also be viewed as a step forward in the study of the asymptotics of the hitting (or equivalently return) time statistics for cylinder sets. We refer the reader to the rich literature on the subject (e.g. [**AG, CK, C, DM, HLV, L**] and the references therein) for recent developments in this field. Many works concentrate on determining whether a process (or a class of processes) has 'exponential asymptotics' or not. These attempts have been successful only in somewhat restricted classes of processes. Our main theorem is the first fully general result saying something concrete about all ergodic processes with positive entropy, from this point of view. It implies that in such processes any essential limit distribution function for the hitting times is majorized by the exponential law $1 - e^{-t}$. In particular, this excludes many behaviors proved to exist in entropy zero, such as the presence of an essential limit law for the return times concentrated away from zero.

2. *Rigorous definitions and statements*

We establish the notation necessary to formulate the results. Let $(X, \Sigma, \mu, T, \mathcal{P})$ be an ergodic process on finitely many symbols, i.e., the process generated by a finite partition $\mathcal{P}$ of the standard probability space (X, Σ, μ) under the action of an automorphism T. The process can be identified with the action of the left shift map on the symbolic space $\mathcal{P}^{\mathbb{Z}}$ where μ is an ergodic shift-invariant probability measure on $\mathcal{P}^{\mathbb{Z}}$. Most of the time, we will identify finite blocks with their cylinder sets, i.e., we will identify the Cartesian product $\mathcal{P}^n$ with the partition $\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P})$. Depending on the context, a block $B \in \mathcal{P}^n$ is attached to some coordinates or it represents a 'word' which may appear starting at any coordinate in the $\mathcal{P}$-names. We will also use the probabilistic language of random variables. Then $\mu(\{x \in \mathcal{P}^{\mathbb{Z}} \mid R(x) \in A\})$ will be abbreviated as $\mu\{R \in A\}$ $(A \subset \mathbb{R})$. Recall that if the random variable R is non-negative and $F(t) = \mu\{R \le t\}$ is its distribution function, then the expected value of R equals $\int_0^\infty 1 - F(t)\, dt$.

For a set B of positive measure let R_B and $\overline{R}_B$ denote the random variables defined on B (with the conditional measure $\mu_B = \mu/\mu(B)$) as the absolute and normalized first return time to B, respectively, i.e.,

$$R_B(y) = \min\{i > 0, T^i(y) \in B\}, \quad \overline{R}_B(y) = \mu(B)R_B(y).$$

Notice that, by the Kac theorem [**Kc**], the expected value of R_B equals $1/\mu(B)$, hence that of $\overline{R}_B$ is 1 (which is why we call it 'normalized'). We denote by $\tilde{F}_B(t)$ the distribution function of $\overline{R}_B$. We also define an auxiliary function

$$G_B(t) = \int_0^t 1 - \tilde{F}_B(s)\, ds.$$

Similarly, let V_B be the random variable defined on X as the *hitting time statistic*, i.e., the waiting time for the first visit in B. The defining formula is the same as for R_B, but this time it is considered on the whole space with the measure μ. Further, let $\overline{V}_B = \mu(B)V_B$, called, by analogy, *the normalized hitting time* (although the expected value of this variable need not be equal to 1). By ergodicity, V_B and $\overline{V}_B$ are well defined. By an elementary consideration of the skyscraper above B, one can easily verify, that the distribution function F_B of $\overline{V}_B$ satisfies, for every $t \ge 0$, the inequalities

$$G_B(t) - \mu(B) \le F_B(t) \le G_B(t)$$

(see [**HLV**] for more details). Because we deal with long blocks (so that, by the Shannon–McMillan–Breiman theorem, $\mu(B)$ is, with high probability, very small), we will often replace F_B by G_B.

The key notions of this work are defined below.

Definition 1. We say that the visits to B *repel* (*attract*) each other from a distance $t > 0$, if $F_B(t) > 1 - e^{-t}$ (if $F_B(t) < 1 - e^{-t}$). The parameters

$$\mathrm{REP}(B) = \max\left\{0, \sup_{t\geq 0}(F_B(t) - 1 + e^{-t})\right\} \quad \text{and}$$

$$\mathrm{ATT}(B) = \max\left\{0, \sup_{t\geq 0}(1 - e^{-t} - F_B(t))\right\}$$

represent the maximal intensity of repelling (attracting) from all distances and will be called the *repelling* (*attracting*) of B for short.

Let us explain why we use the terms 'attracting' and 'repelling'. We agree that in the i.i.d. process there is neither attracting nor repelling and $F_B(t) \approx 1 - e^{-t}$ for all long blocks. We use this as the reference point for other processes. We will compare the occurrences of a block B in $(X, \Sigma, \mu, T, \mathcal{P})$ with the same in an i.i.d. process, assuming that in both processes B has the same measure (i.e., the same overall frequency of occurrence). Fix some $t > 0$. Consider the random variable I counting the number of occurrences of B in the time period $[0, t/\mu(B)]$. The expected value of I equals t (up to the negligible error $\mu(B)$) in both processes. On the other hand, $\mu\{I > 0\} = \mu\{V_B \leq t/\mu(B)\} = F_B(t)$. The ratio $t/F_B(t)$ represents the conditional expected value of I on the set $\{I > 0\}$, i.e., the expected number of occurrences of B in all cases where at least one occurrence is observed. Attracting from the distance t, as defined above, means that $F_B(t)$ is smaller than $1 - e^{-t}$, hence the above conditional expected value is larger in $(X, \Sigma, \mu, T, \mathcal{P})$ than in the reference process. In other words, if we observe the process $(X, \Sigma, \mu, T, \mathcal{P})$ for time $t/\mu(B)$ and we happen to notice the event B before the end of this time, then we can expect a larger number of observed Bs than there would be in the reference i.i.d. process. The first occurrence of B 'attracts' its further repetitions. The interpretation of repelling is analogous, and can be viewed as a force driving the repetitions of B toward occurring with equal gaps. The strongest repelling occurs when the distribution function F_B reaches the largest possible function $\min\{t, 1\}$ $(t \geq 0)$. Then B occurs periodically with period $(\mu(B))^{-1}$. By way of an analogy, electrons on an electric wire, due to repelling, will distribute at equal distances.

If a given process exhibits attracting from some distance and repelling from another, the tendency to clustering (create series) is not clear and depends on the applied time perspective. We will be mostly interested in process with 'pure' attracting, not mixed with repelling from other distances, as defined below.

Definition 2. The event B *obeys the law of series* if

$$F_B(t) \leq 1 - e^{-t}$$

for all t, but the two functions are not equal.

In other words, the law of series is the conjunction of the following two postulates:

(1) the repelling $\mathrm{REP}(B)$ is zero;

(2) the attracting $\mathrm{ATT}(B)$ is positive.

In practice, we agree to accept the presence of some 'marginal' repelling if it is much smaller than the attracting.

Our main result is the following theorem.

THEOREM 1. *In an ergodic process of positive entropy, for every $\epsilon > 0$, the measure of the union of all n-blocks $B \in \mathcal{P}^n$ with $F_B(t) \leq 1 - e^{-t} + \epsilon$ for all t, tends to 1 as n grows to infinity.*

Because for bounded functions the convergence in measure is the same as L^1-convergence, the above can be equivalently phrased as follows.

THEOREM 1A. *If $(X, \Sigma, \mu, T, \mathcal{P})$ is ergodic and has positive entropy, and $x[0, n)$ denotes the cylinder of length n containing x, then*

$$\lim_{n\to\infty} \mathrm{REP}(x[0, n)) = 0,$$

where the limit is taken in $L_1(\mu)$.

Theorem 1A corresponds to the first postulate in Definition 2 for the law of series. Postulate 2 is fulfilled for a large collection of blocks with respect to a typical partition; this is subject of the paper [**DLL**]. The two results together imply that *in positive entropy processes, for long blocks, the law of series prevails.*

3. *More notation and preliminary facts*

We now establish further notation and preliminaries needed in the proofs. If $\mathbb{A} \subset \mathbb{Z}$ then we will write $\mathcal{P}^{\mathbb{A}}$ to denote the partition or sigma-field $\bigvee_{i\in\mathbb{A}} T^{-i}(\mathcal{P})$. We will abbreviate $\mathcal{P}^n = \mathcal{P}^{[0,n)}$, $\mathcal{P}^{-n} = \mathcal{P}^{[-n,-1]}$, $\mathcal{P}^- = \mathcal{P}^{(-\infty,-1]}$.

We assume that the reader is familiar with the basics of entropy for finite partitions and sigma-fields in a standard probability space. Our notation is compatible with [**P**] and we refer the reader to that book, as well as to [**Sh, Wa**], for background and proofs. In particular, we will be using the following:

- the entropy of a partition equals $H(\mathcal{P}) = -\sum_{A\in\mathcal{P}} \mu(A) \log_2(\mu(A))$;
- for two finite partitions $\mathcal{P}$ and $\mathcal{B}$, the conditional entropy $H(\mathcal{P}|\mathcal{B})$ is equal to $\sum_{B\in\mathcal{B}} \mu(B) H_B(\mathcal{P})$, where H_B is the entropy evaluated for the conditional measure μ_B on B;
- the same formula holds for conditional entropy given a sub-sigma-field $\mathcal{C}$, i.e.,
$$\sum_{B\in\mathcal{B}} \mu(B) H_B(\mathcal{P}|\mathcal{C}) = H(\mathcal{P}|\mathcal{B} \vee \mathcal{C});$$
- the entropy of the process is given by
$$h = H(\mathcal{P}|\mathcal{P}^-) = \frac{1}{n} H(\mathcal{P}^n|\mathcal{P}^-) = \lim_{n\to\infty} \frac{1}{n} H(\mathcal{P}^n).$$

We will exploit the notion of ϵ-independence for partitions and sigma-fields. The definition below is an adaptation from [**Sh**], where it concerns finite partitions only. See also [**Sm**] for treatment of countable partitions. Because in this work 'ϵ' is reserved for the intensity of repelling, we will talk about β-independence.

Definition 3. Fix $\beta > 0$. A partition $\mathcal{P}$ is said to be *β-independent* of a sigma-field $\mathcal{B}$ if, for any $\mathcal{B}$-measurable countable partition $\mathcal{B}'$,

$$\sum_{A \in \mathcal{P}, B \in \mathcal{B}'} |\mu(A \cap B) - \mu(A)\mu(B)| \leq \beta.$$

A process $(X, \Sigma, \mu, T, \mathcal{P})$ is called a *β-independent process* if $\mathcal{P}$ is β-independent of the past $\mathcal{P}^-$.

A partition $\mathcal{P}$ is independent of another partition or a sigma-field $\mathcal{B}$ if and only if $H(\mathcal{P}|\mathcal{B}) = H(\mathcal{P})$. The following approximate version of this fact holds (see [**Sh**, Lemma 7.3] for finite partitions, from which the case of a sigma-field is easily derived).

FACT 1. *A partition $\mathcal{P}$ is β-independent of another partition or a sigma-field $\mathcal{B}$ if $H(\mathcal{P}|\mathcal{B}) \geq H(\mathcal{P}) - \xi$, for ξ sufficiently small.*

In the course of proving Theorem 1, we will make frequent use of a certain lengthy condition, abbreviated in the following definition.

Definition 4. Given a partition $\mathcal{P}$ of a space with a probability measure μ and $\delta > 0$, we will say that a property $\Phi(A)$ *holds for $A \in \mathcal{P}$ with μ-tolerance δ* if

$$\mu\left(\bigcup\{A \in \mathcal{P} \mid \Phi(A)\}\right) \geq 1 - \delta.$$

We shall also need an elementary estimate, whose proof is an easy exercise.

FACT 2. *For each $A \in \mathcal{P}$, $H(\mathcal{P}) \leq (1 - \mu(A)) \log_2(\#\mathcal{P}) + 1$.*

In addition to the random variables of the absolute and normalized return times R_B and $\overline{R}_B$, we will also use the analogous notions of the kth absolute return time

$$R_B^{(k)} = \min\{i : \#\{0 < j \leq i : T^j(y) \in B\} = k\},$$

and of the normalized kth return time $\overline{R}_B^{(k)} = \mu(B) R_B^{(k)}$ (both defined on B), with $\tilde{F}_B^{(k)}$ denoting the distribution function of the latter. Clearly, the expected value of $\overline{R}_B^{(k)}$ equals k.

4. *The idea of the proof and the basic lemma*

Before we turn to the formal proof of Theorem 1 we would like to fill in some of the details of the idea behind it. We intend to estimate (from above, by $1 - e^{-t} + \epsilon$) the function G_{BA} (replacing F_{BA}), for long blocks of the form $BA \in \mathcal{P}^{[-n,r)}$. The 'positive' part A has a fixed length r, while we allow the 'negative' part B to be arbitrarily long. There are two key ingredients leading to the estimation. The first one, contained in Lemma 3, is the observation that for a fixed typical $B \in \mathcal{P}^{-n}$, the part of the process induced on B (with the measure μ_B) generated by the partition $\mathcal{P}^r$ is not only a β-independent process but also β-independent of many return times $R_B^{(k)}$ of the cylinder B (see Figure 1).

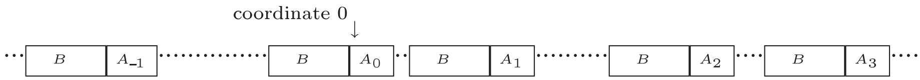

FIGURE 1. The process $\ldots A_{-1}A_0A_1A_2\ldots$ of r-blocks following the copies of B is a β-independent process with additional β-independence properties of the positioning of the copies of B.

This allows us to decompose (with high accuracy) the distribution function $\tilde{F}_{BA}$ of the normalized return time of BA as follows:

$$\begin{aligned}
\tilde{F}_{BA}(t) &= \mu_{BA}\{\overline{R}_{BA} \le t\} = \mu_{BA}\left\{R_{BA} \le \frac{t}{\mu(BA)}\right\} \\
&= \sum_{k\ge 1} \mu_{BA}\left\{R_A^{(B)} = k,\, R_B^{(k)} \le \frac{t}{p\mu(B)}\right\} \\
&\approx \sum_{k\ge 1} \mu_{BA}\{R_A^{(B)} = k\} \cdot \mu_B\left\{\overline{R}_B^{(k)} \le \frac{t}{p}\right\} \\
&\approx \sum_{k\ge 1} p(1-p)^{k-1} \cdot \tilde{F}_B^{(k)}\left(\frac{t}{p}\right),
\end{aligned}$$

where $R_A^{(B)}$ denotes the first (absolute) return time of A in the process induced on B, and $p = \mu_B(A)$. Because this last process is β-independent, the distribution of the kth return time is nearly geometric with parameter p—this explains the occurrence of the term $p(1-p)^{k-1}$ above.

The second key observation is contained in the elementary Lemma 0 below, in which, for simplicity, we assume full independence in place of β-independence. The idea behind this lemma is as follows: the repelling for BA is strongest when the repelling of B is strongest, i.e., when B occurs periodically. But if B does appear periodically, the return time of BA has nearly geometric distribution, because it is a return time in a β-independent process (only the increment of time is now equal to the constant gap between the occurrences of B). If p is small, this geometric distribution, after normalization, is nearly the exponential law $1 - e^{-t}$. We will regulate the smallness of p by the choice of the parameter r (see Lemma 1). Lemma 0 is constructed to be useful also in the rigorous proof.

LEMMA 0. *Fix some $p \in (0, 1)$. Let $\tilde{F}^{(k)}$ ($k \ge 1$) be a sequence of distribution functions on $[0, \infty)$ such that the expected value of the distribution associated with $\tilde{F}^{(k)}$ equals k. Define*

$$\tilde{F}(t) = \sum_{k\ge 1} p(1-p)^{k-1}\tilde{F}^{(k)}\left(\frac{t}{p}\right) \quad \text{and} \quad G(t) = \int_0^t 1 - \tilde{F}(s)\, ds.$$

Then $G(t) \le (1 - e_p^{-t})/\log e_p$, where $e_p = (1-p)^{-1/p}$.

Proof. We have

$$G(t) = \sum_{k\ge 1} p(1-p)^{k-1} \int_0^t 1 - \tilde{F}^{(k)}\left(\frac{s}{p}\right) ds.$$

We know that $\tilde{F}^{(k)}(t) \in [0, 1]$ and that $\int_0^\infty 1 - \tilde{F}^{(k)}(s)\, ds = k$ (the expected value). With such constraints, it is the indicator function $1_{[k,\infty)}$ that maximizes the integrals from 0 to t simultaneously for every t (because the 'mass' k above the graph is, for such choice of the function $\tilde{F}^{(k)}$, swept maximally to the left). The rest follows by direct calculations:

$$\begin{aligned} G(t) &\leq \sum_{k\geq 1} p(1-p)^{k-1} \int_0^t 1_{[0,k)}\left(\frac{s}{p}\right) ds = \int_0^t \sum_{k=\lceil s/p \rceil}^{\infty} p(1-p)^{k-1}\, ds \\ &= \int_0^t (1-p)^{\lceil s/p \rceil}\, ds \leq \frac{(1-p)^{t/p} - 1}{\log(1-p)^{1/p}}. \end{aligned}$$

□

Notice that the maximizing distribution functions $\tilde{F}_B^{(k)} = 1_{[k,\infty)}$ occur, for the normalized return times of a set B, precisely when B is visited periodically.

Let us comment a bit more on the first key ingredient, the β-independence. Establishing this is the most complicated part of the argument. The idea is to prove conditional (given a 'finite past' $\mathcal{P}^{-n}$) β-independence of the 'present' $\mathcal{P}^r$ jointly from the full past and a large part of the future, responsible for the return times of majority of the blocks $B \in \mathcal{P}^{-n}$. This is done in Lemmas 2 and 3: we succeed in finding a sigma-field (containing the full past and a part of the future), of which $\mathcal{P}^r$ is conditionally β-independent, and which 'nearly determines', for a majority of blocks $B \in \mathcal{P}^{-n}$, some finite number of their sequential return times (probably not all of them). This finite number is sufficient to allow the described earlier decomposition of the distribution function $\tilde{F}_{BA}$.

5. *The proof of Theorem 1*

Throughout the sequel we assume ergodicity of the process $(X, \Sigma, \mu, T, \mathcal{P})$ and that its entropy $h = h_\mu(\mathcal{P})$ is positive. We begin our computations with an auxiliary lemma allowing us to assume (by replacing $\mathcal{P}$ by some $\mathcal{P}^r$) that the elements of the 'present' partition are small, relatively in most of $B \in \mathcal{P}^n$ and for every n. Note that the Shannon–McMillan–Breiman theorem is insufficient: for the conditional measure the error term in that theorem depends increasingly on n, which we do not fix.

LEMMA 1. *For each δ there exists an $r \in \mathbb{N}$ such that for every $n \in \mathbb{N}$ the following holds for $B \in \mathcal{P}^{-n}$ with μ-tolerance δ:*

$$\text{for every } A \in \mathcal{P}^r, \quad \mu_B(A) \leq \delta.$$

Proof. Let α be so small that

$$\sqrt{\alpha} \leq \delta \quad \text{and} \quad \frac{h - 3\sqrt{\alpha}}{h + \alpha} \geq 1 - \frac{\delta}{2},$$

and set $\gamma = \alpha / \log_2(\#\mathcal{P})$. Let r be so large that

$$\frac{1}{r} \leq \alpha, \qquad \frac{1}{r(h+\alpha)} \leq \frac{\delta}{2},$$

and that there exists a collection $\overline{\mathcal{P}^r}$ of no more than $2^{r(h+\alpha)} - 1$ elements of $\mathcal{P}^r$ whose joint measure μ exceeds $1 - \gamma$ (by the Shannon–McMillan–Breiman theorem).

************oo..************.........************.........************.........

FIGURE 2. The circles indicate the coordinates 0 through $r-1$, the conditioning sigma-field is over the coordinates marked by stars, which includes the entire past and part of the future with gaps of size $2\alpha M$ repeated periodically with period M (the first gap is half the size).

Let $\widetilde{\mathcal{P}^r}$ denote the partition into the elements of $\overline{\mathcal{P}^r}$ and the complement of their union, and let $\mathcal{R}$ be the partition into the remaining elements of $\mathcal{P}^r$ and the complement of their union, so that $\mathcal{P}^r = \widetilde{\mathcal{P}^r} \vee \mathcal{R}$. For any n, we have

$$\begin{aligned} rh &= H(\mathcal{P}^r|\mathcal{P}^-) \le H(\mathcal{P}^r|\mathcal{P}^{-n}) = H(\widetilde{\mathcal{P}^r} \vee \mathcal{R}|\mathcal{P}^{-n}) \\ &= H(\widetilde{\mathcal{P}^r}|\mathcal{R} \vee \mathcal{P}^{-n}) + H(\mathcal{R}|\mathcal{P}^{-n}) \le H(\widetilde{\mathcal{P}^r}|\mathcal{P}^{-n}) + H(\mathcal{R}) \\ &\le \sum_{B\in\mathcal{P}^{-n}} \mu(B) H_B(\widetilde{\mathcal{P}^r}) + \gamma r \log_2(\#\mathcal{P}) + 1 \end{aligned}$$

(we have used Fact 2 for the last part). After dividing by r, we obtain

$$\sum_{B\in\mathcal{P}^{-n}} \mu(B)\frac{1}{r} H_B(\widetilde{\mathcal{P}^r}) \ge h - \gamma \log_2(\#\mathcal{P}) - \frac{1}{r} \ge h - 2\alpha.$$

Because each term $(1/r)H_B(\widetilde{\mathcal{P}^r})$ is not larger than $(1/r)\log_2(\#\widetilde{\mathcal{P}^r})$, which was set to be at most $h+\alpha$, we deduce that

$$\frac{1}{r} H_B(\widetilde{\mathcal{P}^r}) \ge h - 3\sqrt{\alpha}$$

holds for $B \in \mathcal{P}^{-n}$ with μ-tolerance $\sqrt{\alpha}$, hence also with μ-tolerance δ. On the other hand, by Fact 2, for any B and $A \in \widetilde{\mathcal{P}^r}$,

$$H_B(\widetilde{\mathcal{P}^r}) \le (1-\mu_B(A))\log_2(\#\widetilde{\mathcal{P}^r}) + 1 \le (1-\mu_B(A)) r(h+\alpha) + 1.$$

Combining the last two displayed inequalities we establish that, with μ-tolerance δ for $B \in \mathcal{P}^{-n}$ and then for every $A \in \widetilde{\mathcal{P}^r}$,

$$1 - \mu_B(A) \ge \frac{h - 3\sqrt{\alpha}}{h+\alpha} - \frac{1}{r(h+\alpha)} \ge 1 - \delta.$$

So, $\mu_B(A) \le \delta$. Because $\mathcal{P}^r$ refines $\widetilde{\mathcal{P}^r}$, the elements of $\mathcal{P}^r$ are also not larger than δ. $\square$

We continue the proof with a lemma which can be deduced from [**Ru**, Lemma 3]. We provide a direct proof. For $\alpha > 0$ and $M \in \mathbb{N}$, let

$$S(M,\alpha) = \bigcup_{m\in\mathbb{Z}} [mM + \alpha M, (m+1)M - \alpha M) \cap \mathbb{Z}.$$

LEMMA 2. *For fixed α and r there exists M_0 such that, for every $M \ge M_0$,*

$$H(\mathcal{P}^r|\mathcal{P}^- \vee \mathcal{P}^{S(M,\alpha)}) \ge rh - \alpha$$

(see Figure 2).

Proof. First assume that $r = 1$. Write

$$S'(M, \alpha) = \bigcup_{m \in \mathbb{Z}} [mM + \alpha M, (m+1)M) \cap \mathbb{Z}.$$

Let M be so large that $H(\mathcal{P}^{(1-\alpha)M}) < (1-\alpha)M(h+\gamma)$, where $\gamma = \alpha^2/(2(1-\alpha))$. Then, for any $m \geq 1$,

$$H(\mathcal{P}^{S'(M,\alpha)\cap[0,mM)}|\mathcal{P}^-) \leq H(\mathcal{P}^{S'(M,\alpha)\cap[0,mM)}) < (1-\alpha)mM(h+\gamma).$$

Because $H(\mathcal{P}^{[0,mM)}|\mathcal{P}^-) = mMh$, the complementary part of entropy must exceed $mMh - (1-\alpha)mM(h+\gamma)$ (which equals $\alpha mM(h - \alpha/2)$), i.e., we have

$$H(\mathcal{P}^{[0,mM)\setminus S'(M,\alpha)}|\mathcal{P}^- \vee \mathcal{P}^{S'(M,\alpha)\cap[0,mM)}) > \alpha mM\left(h - \frac{\alpha}{2}\right).$$

Expressing the last entropy term as a sum over $j \in [0, mM)\setminus S'(M, \alpha)$ of the conditional entropies of $T^{-j}(\mathcal{P})$ given the sigma-field over all coordinates left of j and all coordinates from $S'(M, \alpha) \cap [0, mM)$ right of j, and because every such term is at most h, we deduce that more than half of these terms are greater than or equal to $h - \alpha$. So, a term not smaller than $h - \alpha$ occurs for a j within one of the gaps in the left half of $[0, mM)$. Shifting by j, we obtain $H(\mathcal{P}|\mathcal{P}^- \vee T^i(\mathcal{P}^{S'(M,\alpha)\cap[0,mM/2)})) \geq h - \alpha$, where $i \in [0, \alpha M)$ denotes the relative position of j in the gap. As we increase m, one value i will repeat in this role along a subsequence m'. The operation $\vee$ is continuous for increasing sequences of sigma-fields, hence $\mathcal{P}^- \vee T^i(\mathcal{P}^{S'(M,\alpha)\cap[0,m'M/2)})$ converges over m' to $\mathcal{P}^- \vee T^i(\mathcal{P}^{S'(M,\alpha)})$. The entropy is continuous for such convergence, hence $H(\mathcal{P}|\mathcal{P}^- \vee T^i(P^{S'(M,\alpha)})) \geq h - \alpha$. The assertion now follows because $S(M, \alpha)$ is contained in $S'(M, \alpha)$ shifted to the left by any $i \in [0, \alpha M)$.

Finally, if $r > 1$, we can simply argue with $\mathcal{P}^r$ replacing $\mathcal{P}$. This will impose divisibility of M_0 and M by r, but it is not hard to see that for large M the argument works without divisibility at the cost of a slight adjustment of α. □

For a block $B \in \mathcal{P}^{-n}$ consider the process $(X, \Sigma, \mu_B, T_B, \mathcal{P}^r)$ generated by $\mathcal{P}^r$ under the induced transformation T_B (and with the measure μ_B). It can easily be proved that for a fixed $\beta > 0$ and n large enough, the above is a β-independent process for $B \in \mathcal{P}^{-n}$ with μ-tolerance β. The following lemma proves a stronger result: this process is β-independent of the past joined with a finite number of future return times. This fact is the crucial and most difficult item in the proof of Theorem 1.

LEMMA 3. *For every $\beta > 0$, $r \in \mathbb{N}$ and $K \in \mathbb{N}$ there exists n_0 such that for every $n \geq n_0$, with μ-tolerance β for $B \in \mathcal{P}^{-n}$, with respect to μ_B, $\mathcal{P}^r$ is β-independent of jointly the past $\mathcal{P}^-$ and the first K return times to B, $R_B^{(k)}$ $(k \in [1, K])$. In particular, $(X, \Sigma, \mu_B, T_B, \mathcal{P}^r)$ is a β-independent process.*

Proof. We choose ξ according to Fact 1, so that $(\beta/2)$-independence is implied. Let α satisfy

$$0 < \frac{2\alpha}{h-\alpha} < 1, \quad 18K\sqrt{\alpha} < 1, \quad \sqrt{2\alpha} < \xi, \quad K\sqrt[4]{\alpha} < \frac{\beta}{2}.$$

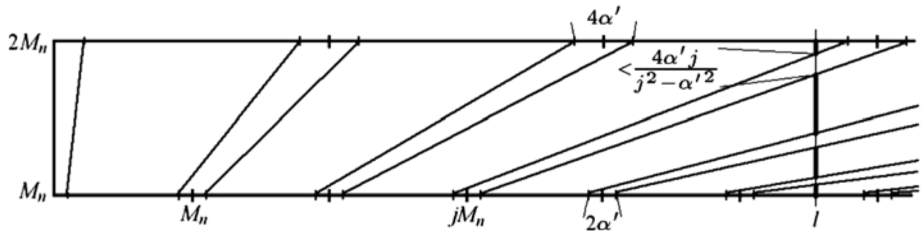

FIGURE 3. The complement of S splits into thin skew strips as shown. The normalized Lebesgue measure of any vertical section of the jth strip (starting at jM_n with $j \geq 1$) is at most $4\alpha' j/(j^2 - \alpha'^2) \leq 5\alpha'/j \leq 10\alpha/j$. Each vertical line at $l \geq M_n$ intersects strips with indices j, $j+1$, $j+2$ up to at most $2j$ (for some j), so the joint measure of the complement of the section of S does not exceed 15α.

Let n_0 be so large that $H(\mathcal{P}^r|\mathcal{P}^{-n}) < rh + \alpha$ for every $n \geq n_0$ and that for every $k \in [1, K]$ with μ-tolerance α for $B \in \mathcal{P}^{-n}$,

$$\mu_B\{2^{n(h-\alpha)} \leq R_B^{(k)} \leq 2^{n(h+\alpha)}\} > 1 - \alpha$$

(we are using the Ornstein–Weiss theorem [**OW**]; the multiplication by k, which should appear for the kth return time, can be included in α in the exponent). Let $M_0 \geq 2^{n_0(h-\alpha)}$ be so large that the assertion of Lemma 2 holds for α, r and M_0, and that for every $M \geq M_0$,

$$(M+1)^{1+2\alpha/(h-\alpha)} < \alpha M^2 \quad \text{and} \quad \frac{\log_2(M+1)}{M(h-\alpha)} < \alpha.$$

We can now redefine (enlarge) n_0 and M_0 so that $M_0 = \lfloor 2^{n_0(h-\alpha)} \rfloor$. Similarly, for each $n \geq n_0$ we set $M_n = \lfloor 2^{n(h-\alpha)} \rfloor$. Observe, that the interval where the first K returns of most n-blocks B may occur (up to probability α) is contained in $[M_n, \alpha M_n^2]$ (because $2^{n(h+\alpha)} \leq (M_n+1)^{1+2\alpha/(h-\alpha)} < \alpha M_n^2$).

At this point we fix some $n \geq n_0$. The idea is to carefully select an M between M_n and $2M_n$ (hence not smaller than M_0), such that the initial K returns of nearly every n-block happen most likely inside (with all its n symbols) the set $S(M, \alpha)$, so that they are 'controlled' by the sigma-field $\mathcal{P}^{S(M,\alpha)}$. Let $\alpha' = \alpha + n/M_n$, so that every n-block overlapping with $S(M, \alpha')$ is completely covered by $S(M, \alpha)$. By the second assumption on $M \geq M_0$ and by the formula connecting M_n and n, we have $\alpha' < 2\alpha$. To define M we will invoke the triple Fubini theorem. Fix $k \in [1, K]$ and consider the probability space

$$\mathcal{P}^{-n} \times [M_n, 2M_n] \times \mathbb{N}$$

equipped with the (discrete) measure $\mathcal{M}$ whose marginal on $\mathcal{P}^{-n} \times [M_n, 2M_n]$ is the product of μ (more precisely, of its projection onto $\mathcal{P}^{-n}$) with the uniform distribution on the integers in $[M_n, 2M_n]$, while, for fixed B and M, the measure on the corresponding $\mathbb{N}$-section is the distribution of the random variable $R_B^{(k)}$. In this space let S be the set whose $\mathbb{N}$-section for a fixed M (and any fixed B) is the set $S(M, \alpha')$. We claim that for every $l \in [M_n, \alpha M_n^2] \cap \mathbb{N}$ (and any fixed B) the $[M_n, 2M_n]$-section of S has measure exceeding $1 - 16\alpha$. This is quite obvious (even for every $l \in [M_n, \infty)$ and with $1 - 15\alpha$) if $[M_n, 2M_n]$ is equipped with the normalized Lebesgue measure (see Figure 3).

In the discrete case, however, *a priori* it might happen that the integers along some $[M_n, 2M_n]$-section often 'miss' the section of S leading to a decreased measure value.

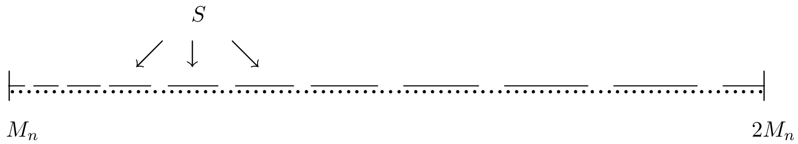

FIGURE 4. The discretization replaces the Lebesgue measure by the uniform measure on M_n integers, thus the measure of any interval can deviate from its Lebesgue measure by at most $1/M_n$. For $l \le \alpha M_n^2$ the corresponding section of S (in this picture drawn horizontally) consists of at most αM_n intervals, so its measure can deviate by no more than α.

(For example, it is easy to see that for $l = (2M_n)!$ the measure of the section of S is zero.) But because we restrict to $l \le \alpha M_n^2$, the discretization does not affect the measure of the section of S by more than α, and the estimate with $1 - 16\alpha$ holds (see Figure 4).

Taking into account all other inaccuracies (the smaller than α part of S outside $[M_n, \alpha M_n^2]$ and the smaller than α part of S projecting onto blocks B which do not obey the Ornstein–Weiss return time estimate) it is safe to claim that

$$\mathcal{M}(S) > 1 - 18\alpha.$$

This implies that for every M from a set of measure at least $1 - 18\sqrt{\alpha}$ the measure of the $(\mathcal{P}^{-n} \times \mathbb{N})$-section of S is larger than or equal to $1 - \sqrt{\alpha}$. For every such M, with μ-tolerance $\sqrt[4]{\alpha}$ for $B \in \mathcal{P}^{-n}$, the probability μ_B that the kth repetition of B falls in $S(M, \alpha')$ (hence with all its n terms inside the set $S(M, \alpha)$) is at least $1 - \sqrt[4]{\alpha}$.

Because $18K\sqrt{\alpha} < 1$, there exists at least one M for which the above holds for every $k \in [1, K]$. This is our final choice of M which from now on remains fixed. For this M, and for cylinders B chosen with μ-tolerance $K\sqrt[4]{\alpha}$, each of the considered K returns of B with probability $1 - \sqrt[4]{\alpha}$ falls (with all its coordinates) inside $S(M, \alpha)$. Thus, for such a B, with probability $1 - K\sqrt[4]{\alpha}$ the same holds simultaneously for all K return times. In other words, there is a set U_B of measure not exceeding $K\sqrt[4]{\alpha}$ outside of which $R_B^{(k)} = \tilde{R}_B^{(k)}$, where $\tilde{R}_B^{(k)}$ is defined as the time of the kth return of B fully visible inside $S(M, \alpha)$. Notice that $\tilde{R}_B^{(k)}$ is $\mathcal{P}^{S(M,\alpha)}$-measurable.

Let us return to our entropy estimates. By Lemma 2,

$$\begin{aligned}
&\sum_{B \in \mathcal{P}^{-n}} \mu(B) H_B(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{S(M,\alpha)}) \\
&\quad = H(\mathcal{P}^r | \mathcal{P}^{-n} \vee \mathcal{P}^- \vee \mathcal{P}^{S(M,\alpha)}) = H(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{S(M,\alpha)}) \\
&\quad \ge rh - \alpha \ge H(\mathcal{P}^r | \mathcal{P}^{-n}) - 2\alpha = \sum_{B \in \mathcal{P}^{-n}} \mu(B) H_B(\mathcal{P}^r) - 2\alpha.
\end{aligned}$$

Because $H_B(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{S(M,\alpha)}) \le H_B(\mathcal{P}^r)$ for every B, we deduce that with μ-tolerance $\sqrt{2\alpha}$ for $B \in \mathcal{P}^{-n}$,

$$H_B(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{S(M,\alpha)}) \ge H_B(\mathcal{P}^r) - \sqrt{2\alpha} \ge H_B(\mathcal{P}^r) - \xi.$$

Combining this with the preceding arguments, with μ-tolerance $K\sqrt[4]{\alpha} + \sqrt{2\alpha} < \beta$ for $B \in \mathcal{P}^{-n}$ both the above entropy inequalities hold, and we have the estimates of the measures of sets U_B. By the choice of ξ, we obtain that with respect to μ_B, $\mathcal{P}^r$ is jointly $(\beta/2)$-independent of the past and the modified return times $\tilde{R}_B^{(k)}$ $(k \in [1, K])$.

Because $\mu(U_B) \leq K\sqrt[4]{\alpha} < \beta/2$, this clearly implies β-independence if each $\tilde{R}_B^{(k)}$ is replaced by $R_B^{(k)}$. □

To complete the proof of Theorem 1 it now remains to put the items together.

Proof of Theorem 1. Fix an $\epsilon > 0$. On $[0, \infty)$, the functions

$$g_p(t) = \min\left\{1, \frac{1}{\log e_p}(1 - e_p^{-t}) + pt\right\},$$

where $e_p = (1-p)^{-1/p}$, decrease uniformly to $1 - e^{-t}$ as $p \to 0^+$. So, let δ be such that $g_\delta(t) \leq 1 - e^{-t} + \epsilon$ for every t. We also assume that

$$(1 - 2\delta)(1 - \delta) \geq 1 - \epsilon.$$

Let r be specified by Lemma 1, so that $\mu_B(A) \leq \delta$ for every $n \geq 1$, every $A \in \mathcal{P}^r$ and for $B \in \mathcal{P}^{-n}$ with μ-tolerance δ. On the other hand, once r is fixed, the partition $\mathcal{P}^r$ has at most $(\#\mathcal{P})^r$ elements, so with μ_B-tolerance δ for $A \in \mathcal{P}^r$, $\mu_B(A) \geq \delta(\#\mathcal{P})^{-r}$. Let $\mathcal{A}_B$ be the subfamily of $\mathcal{P}^r$ (depending on B) where this inequality holds. Let K be so large that for any $p \geq \delta(\#\mathcal{P})^{-r}$,

$$\sum_{k=K+1}^{\infty} p(1-p)^k < \frac{\delta}{2},$$

and choose $\beta < \delta$ so small that

$$(K^2 + K + 1)\beta < \frac{\delta}{2}.$$

The application of Lemma 3 now provides an n_0 such that for any $n \geq n_0$, with μ-tolerance β for $B \in \mathcal{P}^{-n}$, the process induced on B generated by $\mathcal{P}^r$ has the desired β-independence properties involving the initial K return times of B. So, with tolerance $\delta + \beta < 2\delta$ we have both the above β-independence and the estimate $\mu_B(A) < \delta$ for every $A \in \mathcal{P}^r$. Let $\mathcal{B}_n$ be the subfamily of $\mathcal{P}^{-n}$ where these two conditions hold. Fix some $n \geq n_0$.

Let us consider a cylinder set $B \cap A \in \mathcal{P}^{[-n,r)}$ (or, equivalently, the block BA), where $B \in \mathcal{B}_n$, $A \in \mathcal{A}_B$. The length of BA is $n + r$, which represents an arbitrary integer larger than $n_0 + r$. Notice that the family of such sets BA covers more than $(1-2\delta)(1-\delta) \geq 1 - \epsilon$ of the space.

We will examine the distribution of the normalized first return time for BA. In addition to our customary return-time notation, let $R_A^{(B)}$ be the first (absolute) return time of A in $(X, \Sigma, \mu_B, T_B, \mathcal{P}^r)$, i.e., the variable defined on BA, counting the number of visits to B until the first return to BA. Let $p = \mu_B(A)$ (recall that this is not smaller than $\delta(\#\mathcal{P})^{-r}$). We have

$$\begin{aligned}\tilde{F}_{BA}(t) &= \mu_{BA}\{\overline{R}_{BA} \leq t\} = \mu_{BA}\left\{R_{BA} \leq \frac{t}{\mu(BA)}\right\}\\ &= \sum_{k\geq 1} \mu_{BA}\left\{R_A^{(B)} = k, R_B^{(k)} \leq \frac{t}{p\mu(B)}\right\}.\end{aligned}$$

The kth term of this sum equals

$$\frac{1}{p}\mu_B\bigg(\{A_k = A\} \cap \{A_{k-1} \neq A\} \cap \cdots \cap \{A_1 \neq A\} \cap \{A_0 = A\} \\ \cap \left\{R_B^{(k)} \leq \frac{t}{p\mu(B)}\right\}\bigg),$$

where A_i is the r-block following the ith copy of B (the counting starts from 0 at the copy of B positioned at $[-n, -1]$).

By Lemma 3, for $k \leq K$, in this intersection of sets each term is β-independent of the intersection to its right. So, proceeding from the left, we can replace the probabilities of the intersections by products of probabilities, allowing an error of β. Note that the last term equals $\mu_B\{\overline{R}_B^{(k)} \leq t/p\} = \tilde{F}_B^{(k)}(t/p)$. Jointly, the inaccuracy will not exceed $(K+1)\beta$:

$$\left|\mu_{BA}\left\{R_A^{(B)} = k,\, R_B^{(k)} \leq \frac{t}{p\mu(B)}\right\} - p(1-p)^{k-1}\tilde{F}_B^{(k)}\left(\frac{t}{p}\right)\right| \leq (K+1)\beta.$$

Similarly, we also have $|\mu_{BA}\{R_A^{(B)} = k\} - p(1-p)^{k-1}| \leq K\beta$, hence the tail of the series $\mu_{BA}\{R_A^{(B)} = k\}$ above K is smaller than $K^2\beta$ plus the tail of the geometric series $p(1-p)^{k-1}$, which, by the fact that $p \geq \delta(\#\mathcal{P})^{-r}$, is smaller than $\delta/2$. Therefore

$$\tilde{F}_{BA}(t) \approx \sum_{k\geq 1} p(1-p)^{k-1}\tilde{F}_B^{(k)}\left(\frac{t}{p}\right),$$

up to $(K^2 + K + 1)\beta + \delta/2 \leq \delta$, uniformly for every t. By the application of Lemma 0, G_{BA} satisfies

$$G_{BA}(t) \leq \min\left\{1, \frac{1}{\log e_p}(1 - e_p^{-t}) + \delta t\right\} \leq g_\delta(t) \leq 1 - e^t + \epsilon$$

(because $p \leq \delta$). We have proved that for our choice of ϵ and an arbitrary length $m \geq n_0 + r$, with μ-tolerance ϵ for the cylinders $C \in \mathcal{P}^m$, the intensity of repelling between visits to C is at most ϵ. This concludes the proof of Theorem 1. □

6. *An example*

It is important not to be misled by an oversimplified approach to Theorem 1. The 'decay of repelling' in positive entropy processes appears to agree with the intuitive understanding of entropy as chaos: repelling is a 'self-organizing' property; it leads to a more uniform, hence less chaotic, distribution of an event along a typical orbit. Thus one might expect that repelling with intensity ϵ exhibited by a fraction ξ of all n-blocks contributes to lowering the entropy by some percentage proportional to ξ and depending increasingly on ϵ. If this happens for infinitely many lengths n with the same parameters ξ and ϵ, the entropy should be driven to zero by a geometric progression. Surprisingly, it is not quite so, and the phenomenon of Theorem 1 has more subtle grounds. We will present an example which exhibits the incorrectness of such an intuition. Note also that in the proof of Theorem 1 the entropy is 'killed completely in one step', which means that positive entropy and persistent repelling lead to a contradiction by examining the blocks of one sufficiently large length n; we do not use any iterated procedure requiring repelling for infinitely many lengths.

The construction below shows that for each $\delta > 0$ and $n \in \mathbb{N}$ there exist an $N \in \mathbb{N}$ and an ergodic process on N symbols with entropy $\log_2 N - \delta$, such that some n-blocks of joint measure equal to $1/n$ repel with nearly the maximal possible intensity (i.e., occur nearly periodically). Because δ can be extremely small compared to $1/n$, this construction illustrates that there is no 'reduction of entropy' by an amount proportional to the fraction of blocks which exhibit strong repelling.

Example 1. Let $\mathcal{P}$ be an alphabet of a large cardinality N. Divide $\mathcal{P}$ into two disjoint subsets, one, denoted $\mathcal{P}_0$, of cardinality $N_0 = N2^{-\delta}$ and the relatively small (but still very large) remained which we denote by $\{1, 2, \ldots, r\}$ (we will refer to these symbols as 'markers'). For $i = 1, 2, \ldots, r$, let $\mathcal{B}_i$ be the collection of all n-blocks whose first $n-1$ symbols belong to $\mathcal{P}_0$ and the terminal symbol is the marker i. The cardinality of $\mathcal{B}_i$ is N_0^{n-1}. Let $\mathcal{C}_i$ be the collection of all blocks of length nN_0^{n-1} obtained as concatenations of blocks from $\mathcal{B}_i$ using each of them exactly once. The cardinality of $\mathcal{C}_i$ is $(N_0^{n-1})!$. Let X be the subshift whose points are infinite concatenations of blocks from $\bigcup_{i=1}^r \mathcal{C}_i$, in which every block belonging to $\mathcal{C}_i$ is followed by a block from $\mathcal{C}_{i+1}$ $(1 \le i < r)$ and every block belonging to $\mathcal{C}_r$ is followed by a block from $\mathcal{C}_1$. Let μ be the shift-invariant measure of maximal entropy on X. It is immediately evident that the entropy of μ is $(1/nN_0^{n-1}) \log_2((N_0^{n-1})!)$, which, for large N, nearly equals $\log_2 N_0 = \log_2 N - \delta$. Finally, observe that the measure of each $B \in \mathcal{B}_i$ equals $1/nrN_0^{n-1}$, the joint measure of $\bigcup_{i=1}^r \mathcal{B}_i$ is exactly $1/n$, and every block B from this family appears in any $x \in X$ with gaps ranging between $(1 - 1/r)/\mu(B)$ and $(1 + 1/r)/\mu(B)$, exhibiting strong repelling.

Remark 1. Viewing the blocks of length nrN_0^{n-1} starting with a block from $\mathcal{C}_1$ as a new alphabet, and repeating the above construction inductively, we can produce an example (with the measure of maximal entropy on the intersection of systems created in consecutive steps) with entropy $\log_2 N - 2\delta$, in which strong repelling will occur with probability $1/n_k$ for infinitely many lengths n_k.

7. *Consequences for limit laws*

Studies of limit laws for return/hitting time statistics are based on the following approach: for $x \in \mathcal{P}^{\mathbb{Z}}$ define $F_{x,n} = F_B$ (and $\tilde{F}_{x,n} = \tilde{F}_B$), where B is the block $x[0, n)$ (the cylinder in $\mathcal{P}^n$ containing x). Because for non-decreasing functions $F : [0, \infty) \to [0, 1]$, the weak convergence coincides with the convergence at continuity points, and it makes the space of such functions metric and compact, for every x there exists a well-defined collection of limit distributions for $F_{x,n}$ (and for $\tilde{F}_{x,n}$) as $n \to \infty$. They are called *limit laws for the hitting (return) times at* x. Due to the integral relation ($F_B \approx G_B$), a sequence of return time distributions converges weakly if and only if the corresponding hitting time distributions converge pointwise (see [**HLV**]), so the limit laws for the return times completely determine those for the hitting times and vice versa. A limit law is *essential* if it appears along some subsequence (n_k) for xs in a set of positive measure. In particular, the strongest situation occurs when there exists an almost sure limit law along the full sequence (n). In case this unique limit law is the exponential distribution $1 - e^{-t}$, the process is said to have *exponential asymptotics*. Most of the results concerning the limit laws obtained so far can be classified into three major groups:

(a) characterizations of possible essential limit laws for specific zero entropy processes (e.g. [**CK, DM**]; these limit laws are usually atomic for return times or, equivalently, piecewise linear for hitting times);
(b) finding classes of processes with exponential asymptotics (e.g. [**AG, HSV**]); and
(c) results concerning non-essential limit laws, limit laws along sets other than cylinders (see [**L**]; every probabilistic distribution with expected value not exceeding 1 can occur in any process as the limit law for such general return times), or other very specific topics.

As a consequence of our Theorem 1, we obtain, for the first time, a serious bound on the possible essential limit laws for the hitting time statistics along cylinders in the general class of ergodic positive entropy processes. Statement (1) below is even slightly stronger, because we require, for a subsequence, convergence on a positive measure set, but not necessarily to a common limit.

THEOREM 2. *Assume ergodicity and positive entropy of the process* $(X, \Sigma, \mu, T, \mathcal{P})$.
(1) *If a subsequence* (n_k) *is such that* $\tilde{F}_{x,n_k}$ *converge pointwise to some limit laws* $\tilde{F}_x$ *on a positive measure set* A *of points* x*, then almost surely on* A, $\tilde{F}_x(t) \leq 1 - e^{-t}$ *at each* $t \geq 0$.
(2) *If* (n_k) *grows sufficiently fast, then there is a full measure set such that, for every* x *in this set,* $\limsup_k \tilde{F}_{x,n_k}(t) \leq 1 - e^{-t}$ *at each* $t \geq 0$.

Proof. The implication of Theorems 1 to 2 is obvious and we leave it to the reader. For (2) we hint that (n_k) must grow fast enough to ensure summability of the measures of the sets where the intensity of repelling persists, and then the Borel–Cantelli lemma applies. □

8. *Questions*

Question 1. Can one strengthen Theorem 1A as follows:

$$\lim_{n\to\infty} \mathrm{REP}(x[0, n)) = 0 \quad \mu\text{-almost everywhere?}$$

Question 2. Is there a speed of the convergence to zero of the joint measure of the 'bad' blocks in Theorem 1? More precisely, does there exist a positive function $s(n, \epsilon, \#\mathcal{P})$ converging to zero as n grows, such that if, for some ϵ and infinitely many ns, the joint measure of the n-blocks which repel with intensity ϵ exceeds $s(n, \epsilon, \#\mathcal{P})$, then the process has necessarily entropy zero? (By Example 1, $1/n$ is not enough.)

Question 3. In Lemma 3, can one obtain $\mathcal{P}^r$ conditionally β-independent jointly of the past and *all* the return times $R_B^{(k)}$ $(k \geq 1)$ (for sufficiently large n, with μ-tolerance β for $B \in \mathcal{P}^{-n}$)? In other words, can the β-independent process $(X, \Sigma, \mu_B, T_B, \mathcal{P}^r)$ be obtained β-independent of the factor process generated by the partition into B and its complement?

Question 4. (suggested by J.-P. Thouvenot) Find a purely combinatorial proof of Theorem 1, by counting the cardinality of very long strings (of length m) inside which a positive fraction (in measure) of all n-blocks repel with a fixed intensity. For sufficiently large n this quantity should eventually (as $m \to \infty$) be smaller than h^m for any preassigned positive h.

Acknowledgements. The authors would like to thank Dan Rudolph for a hint leading to the first example of a positive entropy process with attracting, and, in effect, to the discovery of the attracting/repelling asymmetry. We are grateful to Jean-Paul Thouvenot for his interest in the subject, substantial help, and the challenge to find a purely combinatorial proof of the main theorem (which we hope to address in the future). The first version of this paper was written during the first author's several visits to CPT/ISITV, supported by CNRS and ISITV. The research of the first author is supported by resources for science in years 2009–2012 as research project (grant MENII N N201 394537, Poland).

REFERENCES

[AG] M. Abadi and A. Galves. Inequalities for the occurrence times of rare events in mixing processes. The state of the art. Inhomogeneous random systems. *Markov Process. Related Fields* **7** (2001), 97–112.

[CK] V. Chaumoître and M. Kupsa. Asymptotics for return times of rank one systems. *Stoch. Dyn.* **5** (2005), 65–73.

[C] Z. Coelho. Asymptotic laws for symbolic dynamical systems. *Topics in Symbolic Dynamics and Applications (London Mathematical Society Lecture Note Series, 279)*. Cambridge University Press, Cambridge, 2000, pp. 123–165.

[DM] F. Durand and A. Maass. Limit laws for entrance times for low-complexity Cantor minimal systems. *Nonlinearity* **14** (2001), 683–700.

[DLL] T. Downarowicz, Y. Lacroix and D. Leandri. Spontaneous clustering in theoretical and some empirical stochastic processes. *ESAIM Probab. Stat.* to appear.

[HLV] N. Haydn, Y. Lacroix and S. Vaienti. Hitting and return times in ergodic dynamical systems. *Ann. Probab.* **33** (2005), 2043–2050.

[HSV] N. Hirata, B. Saussol and S. Vaienti. Statistics of return times: a general framework and new applications. *Commun. Math. Phys.* **206** (1999), 33–55.

[Kc] M. Kac. On the notion of recurrence in discrete stochastic processes. *Bull. Amer. Math. Soc.* **53** (1947), 1002–1010.

[L] Y. Lacroix. Possible limit laws for entrance times of an ergodic aperiodic dynamical system. *Israel J. Math.* **132** (2002), 253–263.

[OW] D. S. Ornstein and B. Weiss. Entropy and recurrence rates for stationary random fields. *IEEE Trans. Inform. Theory* **48** (2002), 1694–1697.

[P] K. Petersen. *Ergodic Theory*. Cambridge University Press, Cambridge, 1983.

[Ru] D. J. Rudolph. If a two-point extension of a Bernoulli shift has an ergodic square, then it is Bernoulli. *Israel J. Math.* **30** (1978), 159–180.

[Sh] P. Shields. *The Theory of Bernoulli Shifts (Chicago Lectures in Mathematics)*. University of Chicago Press, Chicago, 1973.

[Sm] M. Smorodinsky. On Ornstein's isomorphism theorem for Bernoulli shifts. *Adv. Math.* **9** (1972), 1–9.

[Wa] P. Walters. *Ergodic Theory—Introductory Lectures (Lecture Notes in Mathematics, 458)*. Springer, Berlin, 1975.