Law of Series

Tomasz Downarowicz
Institute of Mathematics and Computer Science
Wroclaw University of Technology

Extracted from
Entropy in Dynamical Systems
Cambridge University Press
New Mathematical Monographs 18
Cambridge 2011

1 Introduction

In this note we describe a relatively new result – a consequence of positive dynamical entropy of a process. It concerns the behavior of the return time random variables $\mathrm{R}_n(x)$ for large n (the same as treated by the Ornstein–Weiss Return Times Theorem in [Ornstein and Weiss, 1993], but in a complementary manner). The theorem has a very interesting interpretation, easy to articulate in a language accessible also to nonspecialists. Yet, as usually at such occasions, one has to be very cautious and not get enticed into pushing the conclusions too far. We begin this chapter with a short historical note concerning the debate on the Law of Series in the colloquial meaning. We explain how the *Ergodic Law of Series* contributes to this debate. Then we pass to the mathematical proof preceded by introducing a number of ergodic-theoretic tools.

2 History of the Law of Series

In the colloquial language, a "series" happens when a random event, usually extremely rare, is observed surprisingly often throughout a period of time. Even two repetitions, one shortly after another, are often interpreted as a "series". The Law of Series is the belief that such series happen more often than they should by "pure chance" (whatever that means). This belief is usually associated with another; that there exists some unexplained force or rule behind this "law". A number of idioms, such as "run of good luck" or "run of misfortune", or proverbs like "misfortune never comes alone", exist in nearly all languages, which confirms that people have been noticing this kind of mystery since a long time. The most commonly known examples of "series" are runs of good luck in gambling with the famous case of Charles Wells taking the lead (see e.g. *Charles Wells (gambler)* on Wikipedia).

Serial occurrences of certain types of events is perfectly understandable as a result of physical dependence. For example, volcanic eruptions appear in series during periods of increased tectonic activity. Another good example here are series of people falling ill due to a contagious disease, or very simply, returns of certain motifs in fashion design. The dispute around the Law of Series clearly concerns only such events for which there are no obvious clustering mechanisms, and which are expected to appear completely independently from each-other, and yet, they do appear in series. With this restriction the Law of Series belongs to the category of unexplained mysteries, such as synchronicity, telepathy or even Murphy's Law, and is often considered a manifestation of paranormal forces that exist in our world and escape scientific explanation. This might be the reason why, after the first burst of interest, serious scientists and journals refused to get involved in the investigations of this and related topics. Below we review the list of selected scientists involved in the debate.

Kammerer. An Austrian biologist *Paul Kammerer* (1880-1926) was the first scientist to study the Law of Series (law of seriality, in some translations). His book *Das Gesetz der Serie* [Kammerer, 1919] contains many examples from his and his nears' lives. Richard von Mises in his book [von Mises, 1981] describes that Kammerer conducted many (rather naive) experiments, spending hours in parks noting occurrences of pedestrians with certain features (glasses, umbrellas, etc.), or in shops, noting precise times of arrivals of clients, and the like. Kammerer "discovered" that the number of time intervals (of a fixed length) in which the number of objects under observation agrees with the average is by much smaller than the number of intervals, where that number is either zero or larger than the average. This, he argued, provided evidence for clustering. From today's perspective, Kammerer merely noted the perfectly normal spontaneous clustering of signals in the Poisson process. Nevertheless, Kammerer's book attracted some attention of the public, and even of some serious scientists, toward the phenomenon of clustering. Kammerer himself lost authority due to accusations of manipulating his biological experiments (unrelated to our topic), which eventually drove him to suicide.

Pauli and Jung. Examples of series are, in the popular culture, mixed with examples of other kinds of "unbelievable" coincidences. Pioneer theories about coincidences (including series) were postulated not only by Kammerer but also by a noted Swiss psychologist Carl Gustav Jung (1875-1961) and a Nobel prize winner in physics, Austrian, Wolfgang Pauli (1900-1958). They believed that there exist undiscovered physical "attracting" forces driving objects that are alike, or have common features, closer together in time and space (so-called synchronicity) [see e.g. Jung and Pauli, 1955; Jung, 1977].

Moisset. The Law of Series and synchronicity interests the investigators of spirituality, magic and parapsychology. It fascinates with its potential to generate "meaningful coincidences". A Frenchman Jean Moisset (born 1924), a self-educated specialist in parapsychology, wrote a number of books on synchronicity, Law of Series, and similar phenomena. He connects the Law of Series with psychokinesis and claims that it is even possible to use it for a purpose [Moisset, 2000].

Skeptics: Weaver, Kruskall, Diaconis and others. In opposition to the theory of synchronicity is the belief, represented by many statisticians, among others by Warren Weaver (closely collaborating with Claude Shannon), that any series, coincidences and the like, appear exclusively by pure chance and that there is no mysterious or unexplained force behind them. People's perception has the tendency to ignore all those sequences of events which do not posses the attribute of being unusual, so that we largely underestimate the size of the sample space, where the "unusual events" are observed. Human memory registers coincidencies as more frequent simply because they are more distinctive. This is the "mysterious force" behind synchronicity.

With regard to series of repetitions of identical or similar events, the skeptics' argumentation refers to the effect of spontaneous clustering. For an event, to repeat in time by "pure chance" means to follow a trajectory of a Poisson process. In a typical realization of a Poisson process the distribution of signals along the time axis is far from being uniform; the gaps between signals are sometimes bigger, sometimes smaller. Places where several smaller gaps accumulate (which obviously happens here and there along the time axis) can be interpreted as "spontaneous clusters" of signals. It is nothing but these natural clusters that are being observed and over-interpreted as the mysterious "series". Richard von Mises clearly indicates that it is this kind of "seriality" that has been seen by Kammerer in most of his experiments.

Yet another "cool-minded" explanation of synchronicity (including the Law of Series) asserts that very often events that seem unrelated (hence should appear independently of each-other) are in fact strongly related. Many "accidental" coincidencies or series of similar events, after taking a closer look at the mechanisms behind them, can be logically explained as "not quite accidental". Ordinary people simply do not bother to seek the logical connection. After all, it is much more exciting to "encounter the paranormal". This point of view is neatly described by Robert Matthews in some of his essays. Criticism of the ubiquitous assumption of independence in various experiments can be found in works of William Kruskal [e.g. Kruskal, 1988]. Percy Diaconis is famous for proving that coin tosses in reality do not represent an i.i.d. process [e.g. Diaconis et al., 2007].

Summarizing, the debate concentrates around the major question:

- *Does there indeed exist a Law of Series or is it just an illusion, a matter of our selective perception or memory?*

So far, this debate has avoided strict scientific language; even its subject is not precisely defined, and it is difficult to imagine appropriate repetetive experiments in a controlled environment. Thus, in this approach, the dispute is probably fated to remain an exchange of speculations.

3 The ergodic law of series

We will describe a rigorous approach embedded in the ergodic theory. Surprisingly, the study of stochastic processes supports the Law of Series against the skeptic point of view, of course, subject to correct interpretation.

We begin with definitions of *attracting* and *repelling*, the tools allowing to formalize the subject of study. Using the entropy theory we prove that in nondeterministic processes, for events of certain type (long cylinder sets), attracting prevails, while repelling (almost) does not exist – this is exactly how we understand the Ergodic Law of Series.

One has to be very wary about the applicability of this theory in reality. It concerns only events of a specific form (long cylinders) and it gives no quantitative lower bound on the time perspective at which the phenomenon becomes observable. Perhaps, it might be applied in genetics, computer science, or in data transmission, where one deals with really long blocks of symbols, but again, with extreme caution. The theory does not explain "runs of good luck", or why "misfortune never comes alone", because such "series" are not repetitions of one and the same long cylinder set. Nonetheless it contributes to the general debate at the philosophic level: Properly understood Law of Series is neither an illusion nor a paranormal phenomenon, but a rigorous mathematical law.

3.1 Attracting and repelling in signal processes

By a *signal process* we will understand a continuous time (also discrete time, when the increment of time is very small) stochastic process $(\mathrm{X}_t)_{t\geq 0}$ defined on a probability space $(\Omega, \mathfrak{A}, \mu)$ and assuming integer values, such that $\mathrm{X}_0 = 0$ a.s., and with nondecreasing and right-continuous trajectories $t \mapsto \mathrm{X}_t(\omega)$. We say that (for given $\omega \in \Omega$) a *signal* (or several simultaneous signals) occurs at time t if the trajectory $\mathrm{X}_t(\omega)$ jumps by a unit (or several units) at t.

Definition 3.1. A signal process is *homogeneous* if, for every $t_0 \geq 0$ and every finite collection $0 \leq t_1 < t_2 < \cdots < t_n$, the joint distribution of the increments

$$\mathrm{X}_{t_2} - \mathrm{X}_{t_1},\ \mathrm{X}_{t_3} - \mathrm{X}_{t_2},\ \dots,\ \mathrm{X}_{t_n} - \mathrm{X}_{t_{n-1}} \tag{3.2}$$

is the same as that of

$$\mathrm{X}_{t_2+t_0} - \mathrm{X}_{t_1+t_0},\ \mathrm{X}_{t_3+t_0} - \mathrm{X}_{t_2+t_0},\ \dots,\ \mathrm{X}_{t_n+t_0} - \mathrm{X}_{t_{n-1}+t_0}.$$

Assume that X_1 has an expected value $\mathsf{E}(\mathrm{X}_1) = \boldsymbol{\lambda} \in (0, \infty)$, which we call the *intensity* of the signals. Using homogeneity and a standard divisibility and monotonicity argument, one shows that then $\mathsf{E}(\mathrm{X}_t) = t\boldsymbol{\lambda}$ for every $t \in \mathbb{R}$.

With a homogeneous signal process we associate a random variable defined on Ω and called the *waiting time*:

$$\mathsf{W}(\omega) = \min\{t : \mathrm{X}_t(\omega) \geq 1\}.$$

The most basic example of a homogeneous signal process is the *Poisson process* [see e.g. Feller, 1968]). It is characterized by two properties: 1. the increments as described in (3.2) are independent, and 2. jumps by more than one unit have probability zero. These properties imply that the distribution of X_t is the Poisson distribution with the parameter $\boldsymbol{\lambda}t$, i.e., $P\{\mathrm{X}_t = k\} = e^{-\boldsymbol{\lambda}t}\frac{(\boldsymbol{\lambda}t)^k}{k!}$, $k = 0, 1, \dots$, where $\boldsymbol{\lambda} > 0$ coincides with the intensity. The waiting time in a Poisson process has the exponential distribution with the distribution function

$$\mathsf{F}(t) = 1 - e^{-\boldsymbol{\lambda}t}.$$

The independence between the increments means that the signals arriving before some fixed time do not influence the future signals, i.e., the signals arrive "independently from one-another". This pattern of signal arrivals is exactly what is intuitively described as "by pure chance". The Poisson process is the reference point while defining any deviation from the "by pure chance" scheme.

We will consider two such deviations: *attracting* and *repelling*. Intuitively, the signals *attract* each other if they have the tendency to occur in groups (also called *clusters* or *series*), separated by periods of absence. Likewise, the signals *repel* each other if they have the tendency to occur more evenly distributed along the time. We will put this intuition into a rigorous form. It turnes out that these properties depend solely on the distribution of the waiting time.

Definition 3.3. We say that the signals *attract* each other from a distance $t > 0$, if

$$\mathsf{F}_{\mathrm{W}}(t) < 1 - e^{-\boldsymbol{\lambda} t}.$$

where F_{W} is the distribution function of the waiting time W and $\boldsymbol{\lambda}$ is the intensity. Analogously, the signals *repel* each other from a distance t, if

$$\mathsf{F}_{\mathrm{W}}(t) > 1 - e^{-\boldsymbol{\lambda} t}.$$

The difference $|1 - e^{-\boldsymbol{\lambda} t} - \mathsf{F}_{\mathrm{W}}(t)|$ is called the *force* of attracting (or repelling) at t.

Why is attracting and repelling defined in this way? Consider the random variable X_t (the number of signals in the time period $(0, t]$). As we know, $\mathsf{E}(\mathrm{X}_t) = \boldsymbol{\lambda} t$. On the other hand, $P\{\mathrm{X}_t > 0\} = P\{\mathrm{W} \leq t\} = \mathsf{F}_{\mathrm{W}}(t)$. Thus

$$\frac{\boldsymbol{\lambda} t}{\mathsf{F}_{\mathrm{W}}(t)} = \mathsf{E}(\mathrm{X}_t | \mathrm{X}_t > 0)$$

represents the conditional expected number of signals in the interval $(0, t]$ for these ω for which at least one signal occurs there. Attracting from the distance t, as defined above, means that $\mathsf{F}_{\mathrm{W}}(t)$ is smaller than the analogous distribution function (at t) evaluated for the reference Poisson process. This implies that the above conditional expected number is larger in our process than in the Poisson process (the numerators $\boldsymbol{\lambda} t$ are the same for both processes). This fact can be further expressed as follows: If we observe the signal process for time t and we happen to observe at least one signal, then the expected number of all observed signals is larger than as if they arrived "by pure chance". The first signal "attracts" further signals (within time length t). By homogeneity, the same happens in any interval $(s, s + t]$ of length t, contributing to an increased clustering effect. Repelling is the converse: the first signal lowers the expected number of signals in the observation period, contributing to a decreased clustering, and a more uniform distribution of signals in time (see Figure 1).

The force of attracting can be arbitrarily close to 1, which happens when the distribution function F_{W} remains near zero until large values of t (this implies attracting from all distances, except very small and very large ones, where marginal repelling can occur). Such F_{W} indicates that for most ω the waiting time is very long. In particular, $\mathrm{X}_1(\omega) = 0$. Because the intensity $\mathsf{E}(\mathrm{X}_1)$ is a fixed number $\boldsymbol{\lambda}$, there must be a small part of the space Ω, where many signals arrive within a unit of time. In other words, we

repelling .....|......|......|...|.....|......|.....|....|.....|....|..|....|......|..
Poisson .....|........|....|..|....|........|......|..|.......|..|.|......|......|..
attracting .....|..........|..|.|..|...........|.......||.........|.||.......|......|..
strong attracting ......|.||||..||.|||..|.|..

Figure 1: The distribution of signals along the time in processes with the same intensity.

observe two types of behavior: long lasting silence observed with very high probability and rarely a swarm of signals. This kind of behavior will be called *strong attracting* (we neglect to put sharp formal bounds on F_W for this new term).

On the other hand, it is not hard to see that the distribution function F_W cannot exceed the function $\min\{\boldsymbol{\lambda} t, 1\}$ ($t \geq 0$), which is attained for the process in which the signals arrive periodically in time (with gaps equal to $\frac{1}{\boldsymbol{\lambda}}$). This is the maximally repelling process, and the maximal force of repelling occurs at $t = \frac{1}{\boldsymbol{\lambda}}$ and equals e^{-1} (see Figure 2 below).

Figure 2: The distribution function F_W in the Poisson, strongly attracting and strongly repelling processes.

If a given process reveals attracting from some distance and repelling from another, the tendency to clustering is not clear and depends on the applied time perspective. However, if there is only attracting (without repelling), then at any time scale we shall see the increased clustering. This type of behavior is our subject of interest:

Definition 3.4. A homogeneous signal process *obeys the Law of Series* if

$$\mathsf{F}_\mathsf{W}(t) \leq 1 - e^{-\boldsymbol{\lambda} t},$$

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

In other words, the Law of Series is the conjunction of the following two postulates:

1. There is no repelling from any distance, and
2. there is attracting from at least one distance.

In practice, we agree to accept the presence of some "marginal" repelling with force much smaller than the force of the existing attracting as shown on the Figure 3. Let us explain at this point that the distribution function of the waiting time is always concave (this will become clear e.g. from the integral formula (3.6)), hence it cannot be drawn as just any distribution function.

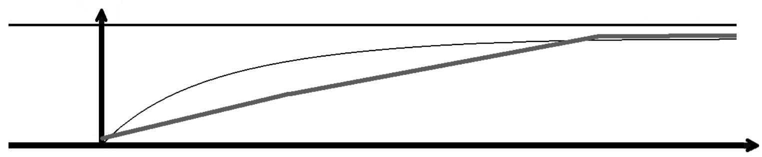

Figure 3: The distribution F_W in a process that "nearly" obeys the Law of Series.

3.2 Decay of repelling in positive entropy

In an ergodic nonperiodic process $(X, \mathcal{P}, \mu, T, \mathbb{S})$ (with $\mathcal{P}$ finite) fix a measurable set B and consider the signal process defined on the probability space (X, μ), where signals are occurrences of the event B, i.e.,

$$\mathrm{X}_t(x) = \#\{n \in (0, t] : T^n x \in B\}.$$

This is a *discrete time homogeneous process*; the homogeneity (see Definition 3.1) holds for integer t_0. By the Ergodic Theorem, the intensity $\boldsymbol{\lambda}$ equals $\mu(B)$, and $\mathrm{E}(\mathrm{X}_t) = \boldsymbol{\lambda} t$ holds for integer t. Since every nonatomic standard probability space is isomorphic to the unit interval (and the measure in an ergodic nonperiodic process is nonatomic), we can draw B (equipped with the meaure μ_B) as the interval $[0, 1]$ and we can arrange that the return time R_B defined for $x \in B$ as

$$\mathrm{R}_B(x) = \min\{n > 0 : T^n x \in B\}$$

increases from left to right. Then the graph of the return time R_B coincides with the roof of the skyscraper over B representing the entire space X. Now, the same graph reflected about the diagonal represents the distribution function G_B of R_B.

Notice that there is a relation between G_B and the distribution function F_B of the waiting time W_B in this process; by an elementary consideration of the skyscraper (which we leave to the reader) one easily verifies that, for any integer t,

$$\mathsf{F}_B(t) = \mu(B) \sum_{i \le t} (1 - \mathsf{G}_B(i)) \tag{3.5}$$

(thus $\mathsf{G}_B(t) = 1 - \frac{\mathsf{F}_B(t) - \mathsf{F}_B(t-1)}{\mu(B)}$). Both functions are determined by their values at integer arguments. Thus it is completely equivalent whether we study the distribution of the return time variable (defined on B), or of the waiting time variable (defined on X).

The Law of Series in occurrences of the event B can be nicely expressed in terms of the shape of the skyscraper above B; the formula (3.5) translates the inequality $\mathsf{F}_B \leq 1 - e^{-\boldsymbol{\lambda} t}$ into the following property of the shape of the skyscraper:

- At any point $t \in B$ the area above the graph of $-\frac{\log(1-s)}{\boldsymbol{\lambda}}$ and below the roof function to the left of t (i.e., for $s \leq t$) must not exceed the area below the graph of $-\frac{\log(1-s)}{\boldsymbol{\lambda}}$ and above the roof function to the left of t.

This property is explained graphically on the Figure 4. In particular, the graph

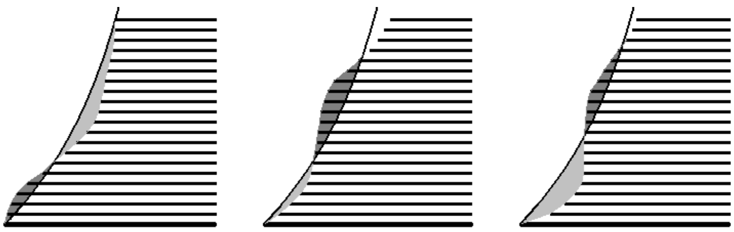

Figure 4: The first two skyscrapers are not admitted by the Law of Series, the last one is. The dark-grey area must be smaller than or equal to the light-grey area to the left.

of the roof function must start at zero tangentally to or below the line $s \mapsto \frac{s}{\boldsymbol{\lambda}}$. For instance, the return time cannot be bounded below by a positive value.

Although the Ornstein-Weiss Theorem (see [Ornstein and Weiss, 1993]) provides some information about the return time R_B, where B is a "typical" long cylinder, its precise distribution on B, i.e., the shape of the skyscraper over B is by no means captured. Small deviations of the value $\frac{1}{n}\log \mathrm{R}_B(x)$ as x ranges over B (allowed in the Ornstein-Weiss Theorem mean, for large n, huge deviations of $\log \mathrm{R}_B(x)$ i.e., huge freedom in the proportions between $\mathrm{R}_B(x)$ (hence also of W_B) at different points. In order to be able to compare the distribution function of W_B with the exponential distribution function $1 - e^{-\boldsymbol{\lambda} t}$ we will need completely different tools.

First of all, it will be convenient to change the time unit to $\frac{1}{\boldsymbol{\lambda}}$, i.e., to replace R_B by $\overline{\mathrm{R}}_B = \mu(B)\mathrm{R}_B$ (and W_B by $\overline{\mathrm{W}}_B = \mu(B)\mathrm{W}_B$). We call this step *normalization* because the *normalized return time* has expected value 1 (although the *normalized waiting time* $\overline{\mathrm{W}}_B$ may even have infinite expected value). This trick has many advantages: (1) the signal process in this new time scale has intensity 1, hence the parameter $\boldsymbol{\lambda}$ disappears from the calculations, (2) the time of the signal process becomes nearly continuous (the increment of time is now $\boldsymbol{\lambda} = \mu(B)$, which is very small), (3) the formula (3.5) takes on, for the distribution functions $\overline{\mathsf{F}}_B$ of $\overline{\mathrm{W}}_B$ and $\overline{\mathsf{G}}_B$ of $\overline{\mathrm{R}}_B$, the integral form

$$\overline{\mathsf{F}}_B(t) \approx \int_0^t 1 - \overline{\mathsf{G}}_B(s)\, ds \tag{3.6}$$

(up to accuracy $\mu(B)$), and (4) we can compare the behaviors of signal processes obtained for sets B of different measures. In particular, we can see what happens in the limit when B represents longer and longer cylinders.

Rich literature is devoted to the subject of the limit distributions of the normalized return (and waiting) time variables as the length of the cylinders grow, in specific types of processes [see Coelho, 2000; Abadi, 2001; Abadi and Galves, 2001; Durand and Maass, 2001; Hirata et al., 1999; Haydn et al., 2005, and the reference therein]. Here we will be mainly interested in consequences of the sole assumption of positive entropy. For each x define

$$\mathsf{Rep}_n(x) = \sup_{t\geq 0}(\overline{\mathsf{F}}_{A_x^n}(t) - 1 + e^{-t}),$$

the maximal force of repelling of the cylinder $A_x^n \in \mathcal{P}^n$ containing x. The main theorem in the area is this this [Downarowicz and Lacroix, 2010]:

Theorem 3.7 (The Ergodic Law of Series)**.** *Let* $(X, \mathcal{P}, \mu, T, \mathbb{S})$ *be an ergodic process with positive entropy, where* $\mathcal{P}$ *is finite. Then*

$$\mathsf{Rep}_n \underset{n\to\infty}{\longrightarrow} 0 \quad \textit{in } L^1(\mu).$$

Because for functions bounded by a common bound the L^1-convergence is the same as the convergence in measure, the above can be equivalently expressed as follows: for every $\varepsilon > 0$ the measure of the union of all blocks of length n, $B \in \mathcal{P}^n$ which repel with force ε, converges to zero as n grows to infinity.

The above theorem asserts that the majority of sufficiently long cylinders reveals almost no repelling, in which they satisfy the first postulate of the Law of Series (phrased next to Definition 3.4). Examples show that arbitrarily strong attracting is admitted by such cylinders, (and it is proved that in the majority of processes it indeed occurs; see the last section of this chapter), hence they satisfy also the second postulate.

Question 3.8. It is unknown whether Theorem 3.7 holds also in the almost everywhere convergence.

3.3 The idea of the proof

The formal proof of Theorem 3.7 is too large for this note. Nonetheless we will sketch the idea behind the proof. First of all, by applying the natural extension, we will assume that the process is invertible, i.e., its symbolic representation is bilateral. We intend to estimate (from above, by $1 - e^{-t} + \varepsilon$) the function $\overline{\mathsf{F}}_B$ for a long cylinder $B \in \mathcal{P}^n$. Instead of B, we can consider a concatenation $BA \in \mathcal{P}^{[-n,r)}$ (i.e., the cylinder set $B \cap A$ with $B \in \mathcal{P}^{-n}, A \in \mathcal{P}^r$), where the "positive" part A has a fixed length r, while we allow the "negative" part B to be (sufficiently) long.

There are two key ingredients leading to the estimation. The first one is the (rather nontrivial) observation that for a fixed typical $B \in \mathcal{P}^{-n}$ the process induced on B (with the conditional measure μ_B) generated by the partition $\mathcal{P}^r$ is nearly an independent process and also nearly independent of the process on $(B, \mathcal{Q}, \mu_B, T_B, \mathbb{Z})$ generated by the partition $\mathcal{Q}$ depending on the return time (see the Figure 5). For the expository purposes of this note we will skip the precise meaning of "nearly" and we skip any traces of proof of this statement.

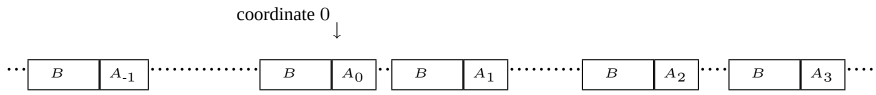

Figure 5: The process $\dots A_{-1}A_0A_1A_2\dots$ of blocks of length r following the copies of B is a nearly independent process, nearly independent of the positioning of the copies of B.

The second key observation is explained below. We assume, for simplicity, full independences in the preceding statement. Then, it is easy to show, that the strongest repelling for BA occurs when the repelling of B is the strongest, i.e., when B occurs periodically. But if B does appear periodically, the return time of BA has nearly the 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 $1-e^{-t}$. (The smallness of p is regulated by the choice of the parameter r.)

3.4 Typicality of attracting for long cylinders

The preceding section provides evidence that in positive entropy processes the occurrences of a selected long cylinder, in principle, do not repel. This corresponds to the first postulate in the interpretation of Definition 3.4 of the Law of Series. As to the second postulate (presence of attracting), of course, it cannot be satisfied by long cylinders in all positive entropy processes. For example, in the independent process all long cylinders occur with neither attracting nor repelling. The same holds in sufficiently fast mixing processes (see [Abadi, 2001] or [Hirata et al., 1999]). But such processes are in fact exceptional; in a "typical" process many blocks reveal strong attracting. We know that a fixed dynamical system $(X, \mathfrak{A}, \mu, T, \mathbb{S})$ gives rise to many processes $(X, \mathcal{P}, \mu, T, \mathbb{S})$, each generated by some partition $\mathcal{P}$. We can thus parametrize the processes by the partitions and use the complete metric structure that exists on the space of partitions to determine the meaning of "typicality":

Definition 3.9. We say that a property Υ of a process is *typical* in a certain class of measure-preserving transformations, if for every $(X, \mathfrak{A}, \mu, T, \mathbb{S})$ in this class, the set of partitions $\mathcal{P}$ of cardinality $m \geq 2$, such that the generated process $(X, \mathcal{P}, \mu, T, \mathbb{S})$ has the property Υ, is *residual* (i.e., contains a dense G_δ set) in the space $\mathfrak{P}_m$ of all partitions into at most m elements.

The theorem below captures the typicality of strong attracting:

Theorem 3.10. *The following property of a process is typical in the class of all ergodic measure-preserving transformations: There exists a set of lengths $N \subset \mathbb{N}$ with upper density 1, such that for every ε and sufficiently large $n \in N$, with tolerance ε every block of length n reveals strong attracting (with force $1-\varepsilon$) of its occurrences.*

Recall that strong attracting automatically eliminates repelling other than marginal. So, this theorem alone, implies that the majority of blocks of selected lengths obey the Law of Series. Nevertheless, blocks of other lengths may strongly repel (but only if the entropy is zero). Examples of such systems have been built by Paulina Grzegorek and Michal Kupsa [Grzegorek and Kupsa, 2009]. In such systems, in the overall picture, where all long cylinders are taken into account, we can still see a mixed behavior without decisive domination of attracting over repelling.

Now we involve the following fact concerning entropy:

Theorem 3.11. *Positive entropy is Rokhlin-typical and typical in the class of measure-preserving transformations with positive Kolmogorov-Sinai entropy.*

Combining the above two facts (recall that the intersection of two residual sets is residual) with the Thoerem 3.7 of the preceding section we obtain that

- **in the class of ergodic measure-preserving transformations with positive entropy, in a typical finitely generated process, long cylinders reveal almost no repelling, while many of them reveal strong attracting.**

This time we do have decisive domination of attracting over repelling. This is the full strength of the Ergodic Law of Series.

The following example shows how the Ergodic Law of Series can manifest itself in reality. Of course, it should be treated with due reserve.

Example 3.12. Consider the experiment of randomly generating independent ASCII characters (the monkey typing). In theory this is an independent process hence every possible long block should appear with positive probability and it should reveal neither repelling nor attracting. In reality, however, the independence of the consecutive outcomes is imperfect (there is no perfect physical independence between any events in reality). We can thus consider the process as being generated by a slightly perturbed partition corresponding to the alphabet. Then there are high chances that the process falls in the class of typical processes (of positive entropy) described in the above theorems. If so, then majority of blocks will obey the Law of Series and if we focus on one particular long block (say the tex file of this note) it is quite likely that once it occurs it will occur again very "soon" (compared with the expected waiting time, which is unimaginably large).

References

Abadi, M. 2001. Exponential approximation for hitting times in mixing processes. *Math. Phys. Electron. J.*, **7**, Paper 2, 19 pp. (electronic).

Abadi, M., and Galves, A. 2001. Inequalities for the occurrence times of rare events in mixing processes. The state of the art. *Markov Process. Related Fields*, **7**(1), 97–112. Inhomogeneous random systems (Cergy-Pontoise, 2000).

Coelho, Z. 2000. Asymptotic laws for symbolic dynamical systems. Pages 123–165 of: *Topics in symbolic dynamics and applications (Temuco, 1997)*. London Math. Soc. Lecture Note Ser., vol. 279. Cambridge: Cambridge Univ. Press.

Diaconis, P., Holmes, S., and Montgomery, R. 2007. Dynamical bias in the coin toss. *SIAM Rev.*, **49**(2), 211–235.

Downarowicz, T., and Lacroix, Y. 2010. The law of series. *Ergodic Theory Dynam. Systems*, **31**, 351–357.

Durand, F., and Maass, A. 2001. Limit laws of entrance times for low-complexity Cantor minimal systems. *Nonlinearity*, **14**(4), 683–700.

Feller, W. 1968. *An introduction to probability theory and its applications. Vol. I*. Third edition. New York: John Wiley & Sons Inc.

Grzegorek, P., and Kupsa, M. 2009. Return times in a process generated by a typical partition. *Nonlinearity*, **22**(2), 371–379.

Haydn, N., Lacroix, Y., and Vaienti, S. 2005. Hitting and return times in ergodic dynamical systems. *Ann. Probab.*, **33**(5), 2043–2050.

Hirata, M., Saussol, B., and Vaienti, S. 1999. Statistics of return times: a general framework and new applications. *Comm. Math. Phys.*, **206**(1), 33–55.

Jung, C. G. 1977. *Jung on Synchronicity and the Paranormal: Key Readings*. Routledge.

Jung, C. G., and Pauli, W. 1955. *The Interpretation of Nature and Psyche*. New York: Pantheon Books.

Kammerer, P. 1919. *Das Gesetz der Serie. Eine Lehre von den Wiederholungen im Lebens- und Weltgeschehen*. Stuttgart/Berlin: Deutsche Verlags-Anstalt. Mit 8 Tafeln und 26 Abb.

Kruskal, W. 1988. Miracles and statistics: the casual assumption of independence. *J. Amer. Statist. Assoc.*, **83**(404), 929–940.

Moisset, J. 2000. *La loi des séries*. Ed. JMG Editions.

Ornstein, D. S., and Weiss, B. 1993. Entropy and data compression schemes. *IEEE Trans. Inform. Theory*, **39**(1), 78–83.

von Mises, R. 1981. *Probability, statistics and truth*. English edn. New York: Dover Publications Inc.

Institute of Mathematics and Computer Science
Wroclaw University of Technology
Wybrzee Wyspiaskiego 27, 50-370 Wroclaw, POLAND
e-mail: downar@pwr.wroc.pl