

Article

The McMillan Theorem for Colored Branching Processes and Dimensions of Random Fractals

Victor Bakhtin ^{1,2}

¹ Department of Mathematics, IT and Landscape Architecture, John Paul II Catholic University of Lublin, Konstantynów Str. 1H, 20-708 Lublin, Poland

² Department of Mechanics and Mathematics, Belarusian State University, Nezavisimosti Ave. 4, 220030 Minsk, Belarus; E-Mail: bakhtin@tut.by

External Editor: Kevin H. Knuth

Received: 16 September 2014; in revised form: 18 November 2014 / Accepted: 12 December 2014 / Published: 19 December 2014

Abstract: For the simplest colored branching process, we prove an analog to the McMillan theorem and calculate the Hausdorff dimensions of random fractals defined in terms of the limit behavior of empirical measures generated by finite genetic lines. In this setting, the role of Shannon’s entropy is played by the Kullback–Leibler divergence, and the Hausdorff dimensions are computed by means of the so-called Billingsley–Kullback entropy, defined in the paper.

Keywords: colored branching process; random fractal; Hausdorff dimension; spectral potential; Kullback action; Billingsley–Kullback entropy; basin of a probability measure; maximal dimension principle

MSC classifications: 28A80; 37F35; 60J80

1. Introduction

Let us fix a finite set $\Omega = \{1, \dots, r\}$, whose elements denote different colors, and consider a finite random set X containing a random number of elements of colors $1, \dots, r$. The simplest colored branching process can be defined as an evolution of a population in which all individuals live the same fixed time, and then, when the lifetime ends, each individual generates (independently of others) a random set of “children” containing individuals of different colors and distributed as X . We will assume that the evolution starts with a unique initial individual. It is suitable to represent this process

as a random genealogical tree with individuals as vertices and each vertex connected with its “children” by edges. Denote by X_n the set of all genetic lines of length n starting from the initial individual. The colored branching process can degenerate (when, starting from some n , all of the sets X_n are empty) or, otherwise, evolve endlessly. Every genetic line $x = (x_1, \dots, x_n) \in X_n$ generates an empirical measure $\delta_{x,n}$ on the set of colors Ω in accordance with the following rule: for each $i \in \Omega$, the value of $\delta_{x,n}(i)$ is the quota in the string (x_1, \dots, x_n) of all elements having the color i .

Denote by $\mu(i)$, where $i \in \Omega$, the expectation of the number of all elements $x \in X$ having the color i . Let ν be an arbitrary probability measure on Ω . An analog to the McMillan theorem that will be proven below (see Theorem 11) asserts that under the condition of non-degeneracy of the colored branching process, the cardinality of the set $\{x \in X_n \mid \delta_{x,n} \approx \nu\}$ almost surely (*i.e.*, with probability one) has asymptotics (up to the logarithmic equivalence) of order $e^{-n\rho(\nu,\mu)}$, where:

$$\rho(\nu, \mu) = \sum_{i \in \Omega} \nu(i) \ln \frac{\nu(i)}{\mu(i)}. \quad (1)$$

Formally, the value of $\rho(\nu, \mu)$ coincides with the usual Kullback–Leibler divergence and differs from it only in the fact that in our setting, the measure μ is not a probability, and so $\rho(\nu, \mu)$ can be negative.

This is the first main result of the paper. It generalizes the usual McMillan theorem, which deals with Ω^n , rather than X^n . Its proof is based essentially on the Legendre duality between the Kullback–Leibler divergence and the so-called *spectral potential*, which was introduced first in [1] and then investigated profoundly in [2,3].

We investigate also Hausdorff dimensions of random fractals defined in terms of the limit behavior of the sequence of empirical measures $\delta_{x,n}$. Let X_∞ be the set of all infinite genetic lines. Fix an arbitrary vector $\theta = (\theta(1), \dots, \theta(r)) \in (0, 1)^r$ and define the following metric on X_∞ :

$$\text{dist}(x, y) = \prod_{t=1}^n \theta(x_t), \quad \text{where } n = \inf \{t \mid x_t \neq y_t\} - 1.$$

Denote by V any set of probability measures on Ω . The second main result of the paper (concerning Hausdorff dimensions of random fractals) is the following: under the condition of non-degeneracy of the colored branching process,

$$\dim_H \{x \in X_\infty \mid \delta_{x,n} \succ V\} = \sup_{\nu \in V} d(\nu, \mu, \theta) \quad (2)$$

almost surely, where $\delta_{x,n} \succ V$ means that the sequence $\delta_{x,n}$ has at least one limit point in V and $d(\nu, \mu, \theta)$ is the *Billingsley–Kullback entropy* defined as:

$$d(\nu, \mu, \theta) = \frac{\rho(\nu, \mu)}{\sum_{i=1}^r \nu(i) \ln \theta(i)} = \frac{\sum_{i=1}^r \nu(i) \ln \frac{\nu(i)}{\mu(i)}}{\sum_{i=1}^r \nu(i) \ln \theta(i)} \quad (3)$$

Moreover, almost surely:

$$\dim_H \{x \in X_\infty \mid \delta_{x,n} \rightarrow \nu\} = d(\nu, \mu, \theta). \quad (4)$$

These results are the essence of Theorems 12 and 17 below.

The systematic study of random self-similar fractals was initiated in the works by Falconer [4], Mauldin, Williams [5] and Graf [6]. They considered the attractor K of a random version of an iterated function system (IFS) in \mathbb{R}^n , letting the scaling factors $\theta(i) \in [0, 1)$ be independent and identically distributed (i.i.d.) random variables. The main result proven in [4–6] is the following: if the open set condition holds and $\sum_i \theta(i) > 1$, then $K \neq \emptyset$ with positive probability, and with the same probability (i.e., almost surely, provided $K \neq \emptyset$), its Hausdorff dimension s is a unique solution to the equation $\sum_i \mathbb{E} \theta(i)^s = 1$. Recently, the research was extended to the case of random self-affine fractals in \mathbb{R}^n (see [7,8]) and random affine code tree fractals [9].

On the other hand, Billingsley in [10,11] considered fractals in the space of infinite sequences $\Omega^{\mathbb{N}}$, where Ω is a finite set, defined in terms of the limit behavior of empirical measures. He investigated Hausdorff dimensions of such fractals and, in particular, obtained formulas analogous to Formulas (2) and (4) (with $\Omega^{\mathbb{N}}$ instead of X_{∞} and $\mu(i) \equiv 1$). Afterwards, Billingsley's ideas and methods were transferred to strange attractors of dynamical systems in \mathbb{R}^n (see [12,13]).

The present research combines both features of random self-similar fractals and subsets of nonrandom self-similar fractals defined in terms of the limit behavior of empirical measures. This combination provides its novelty.

The paper can be divided into two parts. The first one (Sections 3–7) contains preliminary information necessary for our research. All theorems in this part are known and supplied with appropriate references; however, some of them are modified in a certain way for the sake of convenience in further usage. Most of them are proven for completeness of presentation and the reader's convenience. The second part (Sections 2 and 8–11) contains essentially new results, which have been briefly formulated above.

2. Examples and Discussion

Example 1. Suppose that $\Omega = \{1, 2\}$, and the random set $X \subset \Omega$ contains the elements 1, 2 with probabilities $p(1)$, $p(2)$, respectively. Obviously, for $i = 1, 2$, the expectation of the number of elements in X that are equal to i is $p(i)$. Therefore, in the notation of the previous section, $\mu(i) = p(i)$.

Let us define a sequence of random sets $X_n \subset \Omega^n$ in the following way: each X_n contains each of 2^n strings $x = (x_1, \dots, x_n) \in \Omega^n$ with probability $p(x_1) \cdots p(x_n)$. In other words, under the condition $(x_1, \dots, x_n) \in X_n$, the set X_{n+1} contains the string $(x_1, \dots, x_n, 1)$ with probability $p(1)$ and the string $(x_1, \dots, x_n, 2)$ with probability $p(2)$. In particular, X_{n+1} contains both with probability $p(1)p(2)$ and contains neither of them with probability $(1 - p(1))(1 - p(2))$.

Let ν be the probability distribution on Ω , such that $\nu(1) = 1/3$ and $\nu(2) = 2/3$. Question: how many strings (x_1, \dots, x_n) in which the quota of elements that equal one is close to $1/3$ and the quota of elements that equal two is close to $2/3$ are there in the random set X_n ? According to Formula (1), under the condition of non-degeneracy of the stochastic process X_1, X_2, X_3, \dots , the number of those strings with probability one has an asymptotics of order $e^{-n\rho(\nu, \mu)}$, where:

$$\rho(\nu, \mu) = \sum_{i \in \Omega} \nu(i) \ln \frac{\nu(i)}{\mu(i)} = \frac{1}{3} \ln \frac{1/3}{p(1)} + \frac{2}{3} \ln \frac{2/3}{p(2)}.$$

In particular, if $p(1) = p(2) = 2/3$, then $\rho(\nu, \mu) = -\frac{1}{3} \ln 2$ and, respectively, $e^{-n\rho(\nu, \mu)} = 2^{n/3}$. On the other hand, if $p(1) = 1$ and $p(2) = 1/3$, then $\rho(\nu, \mu) = \frac{1}{3} \ln \frac{4}{3}$ and $e^{-n\rho(\nu, \mu)} = (3/4)^{n/3}$. The latter means that for all large enough n , there are no strings $x = (x_1, \dots, x_n)$ in the random set X_n , such that $\delta_{x,n}(1) \approx 1/3$ and $\delta_{x,n}(2) \approx 2/3$.

Example 2. (random Cantor dust). To each string $x = (x_1, \dots, x_n) \in X_n$ from Example 1, let us put in correspondence a line segment $I(x) \subset \mathbb{R}$ as follows:

$$\begin{aligned} I(\emptyset) &= [0, 1], \\ I(1) &= [0, 1/4], \quad I(2) = [1/2, 1], \\ I(1, 1) &= [0, 1/16], \quad I(1, 2) = [1/8, 1/4], \quad I(2, 1) = [1/2, 5/8], \quad I(2, 2) = [3/4, 1], \end{aligned}$$

and so on. If a certain segment $I(x) = I(x_1, \dots, x_n)$ has already been defined, then $I(x_1, \dots, x_n, 1)$ and $I(x_1, \dots, x_n, 2)$ are the left quarter and the right half of $I(x)$, respectively. The random Cantor dust C_∞ is defined as:

$$C_\infty = \bigcap_{n=1}^{\infty} C_n, \quad \text{where } C_n = \bigcup_{x \in X_n} I(x).$$

Put $\theta(1) = 1/4$ and $\theta(2) = 1/2$. Let the distance function on the random set of infinite strings X_∞ be defined as:

$$\text{dist}(x, y) = \prod_{t=1}^n \theta(x_t), \quad \text{where } n = \inf \{t \mid x_t \neq y_t\} - 1.$$

Consider the mapping $\pi: X_\infty \rightarrow C_\infty$ that takes any infinite string $x = (x_1, x_2, \dots) \in X_\infty$ to the point $\pi(x) = \bigcap_n I(x_1, \dots, x_n)$. Obviously, this mapping is bi-Lipschitz, and hence, the Hausdorff dimensions of any subset $A \subset X_\infty$ and its image $\pi(A) \subset \mathbb{R}$ are the same. Question: what is the Hausdorff dimension of the set $\pi(A)$, where:

$$A = \{x \in X_\infty \mid \delta_{x,n}(1) \rightarrow 1/3, \delta_{x,n}(2) \rightarrow 2/3\} \quad (5)$$

The answer is given by Formulas (3) and (4). For example, if for each segment $I(x_1, \dots, x_n)$, the “probabilities of survival” of its left quarter $I(x_1, \dots, x_n, 1)$ and its right half $I(x_1, \dots, x_n, 2)$ are $p(1) = p(2) = 2/3$, then under the condition $\pi(A) \neq \emptyset$ with probability one, we have:

$$\dim_H \pi(A) = d(\nu, \mu, \theta) = \frac{\sum_{i=1}^2 \nu(i) \ln \frac{\nu(i)}{\mu(i)}}{\sum_{i=1}^2 \nu(i) \ln \theta(i)} = \frac{\frac{1}{3} \ln \frac{1/3}{2/3} + \frac{2}{3} \ln \frac{2/3}{2/3}}{\frac{1}{3} \ln \frac{1}{4} + \frac{2}{3} \ln \frac{1}{2}} = \frac{1}{4}.$$

As far as I know, no theorem proven or published previously gives answers to the questions put in Examples 1 and 2. On the other hand, if, in the setting of Example 2, one replaces the set $\pi(A)$ by the whole of $\pi(X_\infty)$, then $\dim_H \pi(X_\infty)$ is computed, for instance, in [4–6] and Theorem 15.2 in [14]. The answer is given there as a solution to the so-called *Bowen equation*. Moreover, in those papers, $\dim_H \pi(X_\infty)$ was computed in a more general setting when the scaling factors $\theta(i) \in [0, 1)$ are random.

Therefore, the following problem arises naturally: how can one compute $\dim_H A$ and $\dim_H \pi(A)$ for the set A defined in Formula (5) in the case of random scaling factors $\theta(i)$? The present work does not solve this problem.

Falconer and Miao started the investigation of random self-affine fractals [7]. They considered a random iterated function system (RIFS) generated by a finite collection of contracting affine maps $S_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and encoded by a random tree T_∞ that is defined exactly in the same manner as the random set of infinite genetic lines X_∞ above. It turned out that under the condition of non-degeneracy $T_\infty \neq \emptyset$ and some additional assumptions, the attractor of RIFS with probability one has a constant dimension (Hausdorff and box-counting), which can be found as a solution to a certain equation. Recently, analogous results were obtained for random affine code tree fractals [9] and for uniformly random self-similar fractals [8]. Thereby, the natural problem arises: is it possible to compute almost sure Hausdorff dimensions for subsets of random fractals like in [7–9], provided the subsets are defined in terms of the limit behavior of the corresponding empirical measures (say, as in Formula (5))?

One more possible direction to generalize the results of the paper is the case when the distribution ν , empirical measures $\delta_{x,n}$ and/or the scaling factors $\theta(i)$ are defined not on the set of colors Ω , but on the whole set of infinite strings $\Omega^\mathbb{N}$.

3. The Spectral Potential

The notion of spectral potential was introduced in [1–3]. It was defined there as a logarithm of the spectral radius of a weighted shift or transfer operator. We will use below a simplified version of the spectral potential that corresponds to an operator of weighted integration.

Let X be an arbitrary finite set. Denote by $B(X)$ the space of all real-valued functions on X , by $M(X)$ the set of all positive measures on X and by $M_1(X)$ the set of all probability distributions on X .

Every measure $\mu \in M(X)$ defines a linear functional on $B(X)$ of the form:

$$\mu[f] = \int_X f d\mu = \sum_{x \in X} f(x)\mu(x).$$

It is easily seen that this functional is positive (*i.e.*, it takes nonnegative values on nonnegative functions). If, in addition, the measure μ is a probability, then this functional is normalized (*i.e.*, $\mu[1] = 1$).

Consider the nonlinear functional:

$$\lambda(\varphi, \mu) = \ln \mu[e^\varphi], \quad (6)$$

where $\varphi \in B(X)$ and $\mu \in M(X)$. We will call it the *spectral potential*. It is monotone (if $\varphi \geq \psi$ then $\lambda(\varphi, \mu) \geq \lambda(\psi, \mu)$), additively homogeneous (that is, $\lambda(\varphi + t, \mu) = \lambda(\varphi, \mu) + t$ for each constant t) and analytic in φ .

Define a family of probability measures μ_φ on X , depending on the functional parameter $\varphi \in B(X)$, by means of the formula:

$$\mu_\varphi[f] = \frac{\mu[e^\varphi f]}{\mu[e^\varphi]}, \quad f \in B(X).$$

Evidently, each measure μ_φ is equivalent to μ and has the density $e^{\varphi - \lambda(\varphi, \mu)}$ with respect to μ .

Let us compute the first two derivatives of the spectral potential with respect to the argument φ . We introduce the notation:

$$\lambda'(\varphi, \mu)[f] = \left. \frac{d\lambda(\varphi + tf, \mu)}{dt} \right|_{t=0}.$$

This is nothing else than the derivative of the spectral potential in the direction f at the point φ . An elementary computation shows that:

$$\lambda'(\varphi, \mu)[f] = \left. \frac{d \ln \mu[e^{\varphi+tf}]}{dt} \right|_{t=0} = \frac{\mu[e^{\varphi} f]}{\mu[e^{\varphi}]} = \mu_{\varphi}[f]. \quad (7)$$

In other words, the derivative $\lambda'(\varphi, \mu)$ coincides with the probability measure μ_{φ} . Now, put:

$$\lambda''(\varphi, \mu)[f, g] = \left. \frac{\partial^2 \lambda(\varphi + tf + sg, \mu)}{\partial s \partial t} \right|_{s,t=0}$$

and compute this derivative using the just obtained Formula (7):

$$\begin{aligned} \lambda''(\varphi, \mu)[f, g] &= \left. \frac{\partial}{\partial s} \left(\frac{\mu[e^{\varphi+sg} f]}{\mu[e^{\varphi+sg}]} \right) \right|_{s=0} = \frac{\mu[e^{\varphi} f g]}{\mu[e^{\varphi}]} - \frac{\mu[e^{\varphi} f] \mu[e^{\varphi} g]}{(\mu[e^{\varphi}])^2} \\ &= \mu_{\varphi}[fg] - \mu_{\varphi}[f] \mu_{\varphi}[g]. \end{aligned}$$

In probability theory, the expression $\mu_{\varphi}[f]$ is usually called the expectation of a random variable f with respect to the probability distribution μ_{φ} , and the expression $\mu_{\varphi}[fg] - \mu_{\varphi}[f] \mu_{\varphi}[g]$ is called the covariance of random variables f and g . In particular, the second derivative:

$$\left. \frac{d^2 \lambda(\varphi + tf, \mu)}{dt^2} \right|_{t=0} = \mu_{\varphi}[f^2] - \mu_{\varphi}[f]^2 = \mu_{\varphi}[(f - \mu_{\varphi}[f])^2]$$

is equal to the variance of the random variable f with respect to the distribution μ_{φ} . Since the variance is nonnegative, it follows that the spectral potential is convex in φ .

4. The Kullback Action

Denote by $B^*(X)$ the space of all linear functionals on $B(X)$. Obviously,

$$M_1(X) \subset M(X) \subset B^*(X).$$

For any measure $\mu \in M(X)$ and nonnegative function $\varphi \in B(X)$, define a product $\varphi\mu \in M(X)$ by means of the formula $(\varphi\mu)(x) = \varphi(x)\mu(x)$.

The Kullback action is the following functional of two arguments $\nu \in B^*(X)$ and $\mu \in M(X)$:

$$\rho(\nu, \mu) = \begin{cases} \mu[\varphi \ln \varphi] = \nu[\ln \varphi], & \text{if } \nu \in M_1(X) \text{ and } \nu = \varphi\mu, \\ +\infty & \text{in all of the other cases.} \end{cases} \quad (8)$$

To be more precise, “all of the other cases” fit into at least one of the three categories: (1) probability measures ν , such that $\nu(x_0) > 0$ and $\mu(x_0) = 0$ for a certain $x_0 \in X$; (2) non-normalized functionals ν ; and (3) nonpositive functionals ν .

As far as we know, in the literature, this functional was defined only for probability measures ν and μ . Different authors call it differently: the relative entropy, the deviation function, the Kullback–Leibler information function and the Kullback–Leibler divergence.

When ν is a probability measure, the Kullback action can be defined by the explicit formula:

$$\rho(\nu, \mu) = \sum_{x \in X} \nu(x) \ln \frac{\nu(x)}{\mu(x)}. \quad (9)$$

In particular, if $\mu(x) \equiv 1$, then the Kullback action differs only in sign from Shannon's entropy:

$$H(\nu) = - \sum_{x \in X} \nu(x) \ln \nu(x). \quad (10)$$

In the case of a probability measure μ , the Kullback action is nonnegative and vanishes only if $\nu = \mu$. Indeed, if the functional ν is not a probability measure that is absolutely continuous with respect to μ , then $\rho(\nu, \mu) = +\infty$. Otherwise, if ν is a probability measure of the form $\nu = \varphi\mu$, then Jensen's inequality and strong convexity of the function $f(x) = x \ln x$ imply:

$$\rho(\nu, \mu) = \mu[f(\varphi)] \geq f(\mu[\varphi]) = 0$$

(since $\mu[\varphi] = \nu[1] = 1$) and the equality $\rho(\nu, \mu) = 0$ holds if and only if φ is constant almost everywhere, and therefore, ν coincides with μ .

Every measure $\mu \in M(X)$ can be written in the form $\mu = c\mu_1$, where $c = \mu[1]$ and $\mu_1 \in M_1(X)$. If $\nu \in M_1(X)$, then Formula (9) implies:

$$\rho(\nu, \mu) = \rho(\nu, \mu_1) - \ln c \geq -\ln \mu[1]. \quad (11)$$

In case $\nu \notin M_1(X)$, this inequality holds, as well, since the Kullback action is infinite.

Theorem 1. *The spectral potential and the Kullback action satisfy the Young inequality:*

$$\rho(\nu, \mu) \geq \nu[\psi] - \lambda(\psi, \mu), \quad (12)$$

that turns into equality if and only if $\nu = \mu_\psi$.

Proof. If $\rho(\nu, \mu) = +\infty$, then the Young inequality is trivial. If $\rho(\nu, \mu) < +\infty$, then by definition of the Kullback action, the functional ν is an absolutely continuous probability measure of the form $\nu = \varphi\mu$, where φ is a nonnegative density. In this case:

$$\begin{aligned} \lambda(\psi, \mu) &= \ln \mu[e^\psi] \geq \ln \sum_{\varphi(x) > 0} e^{\psi(x)} \mu(x) = \ln \sum_{\varphi(x) > 0} e^{\psi(x) - \ln \varphi(x)} \nu(x) \\ &= \ln \nu[e^{\psi - \ln \varphi}] \geq \nu[\psi - \ln \varphi] \end{aligned}$$

(where at the last step, we have used Jensen's inequality and concavity of the logarithm function). Since $\rho(\nu, \mu) = \nu[\ln \varphi]$, this formula implies inequality (12).

Recall that $\mu_\psi = e^{\psi - \lambda(\psi, \mu)} \mu$. Therefore, if $\nu = \mu_\psi$, then by definition:

$$\rho(\nu, \mu) = \nu[\psi - \lambda(\psi, \mu)] = \nu[\psi] - \lambda(\psi, \mu).$$

Vice versa, suppose that $\rho(\nu, \mu) = \nu[\psi] - \lambda(\psi, \mu)$. Then, subtracting from the above equality the Young inequality $\rho(\nu, \mu) \geq \nu[\varphi] - \lambda(\varphi, \mu)$, one obtains:

$$\lambda(\varphi, \mu) - \lambda(\psi, \mu) \geq \nu[\varphi - \psi]$$

which implies that $\nu = \lambda'(\psi, \mu)$. Finally, $\lambda'(\psi, \mu)$ coincides with μ_ψ . \square

Theorem 2. The Kullback action $\rho(\nu, \mu)$ is the Legendre transform of the spectral potential:

$$\rho(\nu, \mu) = \sup_{\psi \in B(X)} \{ \nu[\psi] - \lambda(\psi, \mu) \}, \quad \nu \in B^*(X), \quad \mu \in M(X). \quad (13)$$

Proof. By the Young inequality, the left-hand side of Formula (13) is not less than the right-hand one. Therefore, it is enough to associate with any functional $\nu \in B^*(X)$ a family of functions ψ_t , depending on the real-valued parameter t , on which the equality in Formula (13) is attained.

Suppose first that ν is a probability measure that is absolutely continuous with respect to μ . Therefore, ν has the form $\nu = \varphi\mu$, where φ is a nonnegative density. Consider the family of functions:

$$\psi_t(x) = \begin{cases} \ln \varphi(x), & \text{if } \varphi(x) > 0, \\ -t, & \text{if } \varphi(x) = 0. \end{cases}$$

When $t \rightarrow +\infty$, we have the following relations:

$$\begin{aligned} \mu[e^{\psi_t}] &= \sum_{\varphi(x)>0} \varphi(x)\mu(x) + \sum_{\varphi(x)=0} e^{-t}\mu(x) \longrightarrow \sum_{x \in X} \varphi(x)\mu(x) = 1, \\ \nu[\psi_t] &= \sum_{\varphi(x)>0} \varphi(x) \ln \varphi(x)\mu(x) - \sum_{\varphi(x)=0} t\varphi(x)\mu(x) = \mu[\varphi \ln \varphi], \\ \nu[\psi_t] - \lambda(\psi_t, \mu) &= \nu[\psi_t] - \ln \mu[e^{\psi_t}] \longrightarrow \mu[\varphi \ln \varphi] = \rho(\nu, \mu), \end{aligned}$$

and so, Formula (13) is proven.

In all other cases, when ν is not an absolutely continuous probability measure, by definition, we have $\rho(\nu, \mu) = +\infty$. Let us consider these cases one after another.

If ν is a probability measure, which is singular with respect to μ , then there exists $x_0 \in X$, such that $\mu(x_0) = 0$ and $\nu(x_0) > 0$. In this case, consider the family of functions:

$$\psi_t(x) = \begin{cases} t, & \text{if } x = x_0, \\ 0, & \text{if } x \neq x_0. \end{cases}$$

It is easily seen that:

$$\nu[\psi_t] - \lambda(\psi_t, \mu) \geq t\nu(x_0) - \ln \mu[e^{\psi_t}] \geq t\nu(x_0) - \ln \mu[1].$$

The right-hand side of the above formula goes to $+\infty$, while t increases, and so, Formula (13) holds.

If the functional ν is not normalized, then put $\psi_t(x) \equiv t$. In this case, the expression:

$$\nu[\psi_t] - \lambda(\psi_t, \mu) = \nu[t] - \ln \mu[e^t] = t(\nu[1] - 1) - \ln \mu[1]$$

is unbounded from above, and hence, Formula (13) is still valid.

Finally, if the functional ν is not positive, then there exists a nonnegative function φ , such that $\nu[\varphi] < 0$. Consider the family $\psi_t = -t\varphi$, where $t > 0$. For this family:

$$\nu[\psi_t] - \lambda(\psi_t, \mu) \geq -t\nu[\varphi] - \lambda(0, \mu) \rightarrow +\infty$$

as $t \rightarrow +\infty$, and Formula (13) still remains in force. \square

Corollary 1. *The functional $\rho(\cdot, \mu)$ is convex and lower semicontinuous on the space $B^*(X)$.*

Proof. These are properties of the Legendre transform. \square

5. The Local Large Deviations Principle and the McMillan Theorem

As above, we keep the following notation: X is a finite set; $B(X)$ stands for the space of real-valued functions on X ; $B^*(X)$ is the space of linear functionals on $B(X)$; $M_1(X)$ is the set of all probability measures on X ; and $M(X)$ is the set of all positive measures on X .

With each finite sequence $x = (x_1, \dots, x_n) \in X^n$, we associate an *empirical measure* $\delta_{x,n} \in M_1(X)$, which is supported on the set $\{x_1, \dots, x_n\}$ and assigns to every point x_i the measure $1/n$. The integral of any function f with respect to $\delta_{x,n}$ looks like:

$$\delta_{x,n}[f] = \frac{f(x_1) + \dots + f(x_n)}{n}.$$

For any measure $\mu \in M(X)$, we denote by μ^n its Cartesian power, which is defined on X^n .

Theorem 3 (The local large deviations principle). *For any positive measure $\mu \in M(X)$, functional $\nu \in B^*(X)$ and $\varepsilon > 0$, there exists a neighborhood $O(\nu)$, such that:*

$$\mu^n \{x \in X^n \mid \delta_{x,n} \in O(\nu)\} \leq e^{-n(\rho(\nu, \mu) - \varepsilon)}. \quad (14)$$

On the other hand, for any $\varepsilon > 0$ and any neighborhood $O(\nu)$, the following asymptotic estimate holds:

$$\mu^n \{x \in X^n \mid \delta_{x,n} \in O(\nu)\} \geq e^{-n(\rho(\nu, \mu) + \varepsilon)}, \quad n \rightarrow \infty. \quad (15)$$

If $\rho(\nu, \mu) = +\infty$, then by the difference $\rho(\nu, \mu) - \varepsilon$ in Formula (14), we mean an arbitrary positive number.

In the case of probability measure μ , Theorem 3 is a partial case of Varadhan's large deviations principle (the explicit formulation of which can be found, e.g., in [15,16]). Therefore, this theorem can be deduced from Varadhan's large deviations principle by means of mere renormalization of μ . Nevertheless, we will prove it independently for the sake of completeness.

Proof. By Theorem 2, for any $\varepsilon > 0$, there exists $\psi \in B(X)$, such that:

$$\rho(\nu, \mu) - \varepsilon/2 < \nu[\psi] - \lambda(\psi, \mu). \quad (16)$$

Consider the probability measure $\mu_\psi = e^{\psi - \lambda(\psi, \mu)} \mu$. Obviously,

$$\frac{\mu^n(x)}{\mu_\psi^n(x)} = \prod_{i=1}^n \frac{\mu(x_i)}{\mu_\psi(x_i)} = \prod_{i=1}^n e^{\lambda(\psi, \mu) - \psi(x_i)} = e^{n(\lambda(\psi, \mu) - \delta_{x,n}[\psi])}. \quad (17)$$

Define a neighborhood of the functional ν as follows:

$$O(\nu) = \{\delta \in B^*(X) \mid \delta[\psi] > \nu[\psi] - \varepsilon/2\}.$$

Then, it follows from Formulas (17) and (16) that under the condition $\delta_{x,n} \in O(\nu)$,

$$\frac{\mu^n(x)}{\mu_\psi^n(x)} < e^{n(\lambda(\psi, \mu) - \nu[\psi] + \varepsilon/2)} < e^{n(-\rho(\nu, \mu) + \varepsilon)}.$$

Consequently,

$$\begin{aligned} \mu^n\{x \in X^n \mid \delta_{x,n} \in O(\nu)\} &= \sum_{\delta_{x,n} \in O(\nu)} \mu^n(x) \leq \sum_{\delta_{x,n} \in O(\nu)} e^{n(-\rho(\nu, \mu) + \varepsilon)} \mu_\psi^n(x) \\ &\leq e^{-n(\rho(\nu, \mu) - \varepsilon)}. \end{aligned}$$

Thus, the first part of Theorem 3 is proven.

If $\rho(\nu, \mu) = +\infty$, then estimating Formula (15) is trivial. Therefore, it is enough to prove it only for the case when ν is a probability measure of the form $\nu = \varphi\mu$, and the Kullback action $\rho(\nu, \mu) = \nu[\ln \varphi]$ is finite. Fix any number $\varepsilon > 0$ and any neighborhood $O(\nu)$. Define the sets:

$$Y_n = \{x \in X^n \mid \delta_{x,n} \in O(\nu), \quad |\delta_{x,n}[\ln \varphi] - \nu[\ln \varphi]| < \varepsilon/2\}$$

(the last inequality in the braces implies that $\varphi(x_i) > 0$ at each point of the sequence $x = (x_1, \dots, x_n)$). Note that for $x \in Y_n$,

$$\frac{\mu^n(x)}{\nu^n(x)} = \prod_{i=1}^n \frac{\mu(x_i)}{\nu(x_i)} = \prod_{i=1}^n \frac{1}{\varphi(x_i)} = e^{-n\delta_{x,n}[\ln \varphi]} > e^{-n(\nu[\ln \varphi] + \varepsilon/2)}.$$

Consequently,

$$\mu^n(Y_n) = \sum_{x \in Y_n} \mu^n(x) \geq \sum_{x \in Y_n} e^{-n(\nu[\ln \varphi] + \varepsilon/2)} \nu^n(x) = e^{-n\rho(\nu, \mu) - n\varepsilon/2} \nu^n(Y_n). \quad (18)$$

By the Law of large numbers $\nu^n(Y_n) \rightarrow 1$. Hence, Formula (18) implies Formula (15). \square

Corollary 2 (The McMillan theorem). *For any probability measure $\nu \in M_1(X)$ and $\varepsilon > 0$, there exists a neighborhood $O(\nu)$, such that:*

$$\#\{x = (x_1, \dots, x_n) \in X^n \mid \delta_{x,n} \in O(\nu)\} \leq e^{n(H(\nu) + \varepsilon)}.$$

On the other hand, for any neighborhood $O(\nu)$ and $\varepsilon > 0$:

$$\#\{x \in X^n \mid \delta_{x,n} \in O(\nu)\} \geq e^{n(H(\nu) - \varepsilon)} \quad \text{as } n \rightarrow \infty.$$

Here, $H(\nu)$ denotes Shannon's entropy defined in Formula (10).

Proof. This follows from Formulas (9) and (10) and Theorem 3 if we set $\mu(x) = 1$ for all $x \in X$. \square

6. Hausdorff Dimension and the Maximal Dimension Principle

Let us define the Hausdorff dimension of an arbitrary metric space Ω .

Suppose that Ω is covered by at most a countable collection of sets $\mathcal{U} = \{U_i\}$. Denote by $|\mathcal{U}|$ the diameter of this covering: $|\mathcal{U}| = \sup |U_i|$, where $|U_i|$ is the diameter of U_i . For every $\alpha \in \mathbb{R}$, put:

$$\text{mes}(\mathcal{U}, \alpha) = \sum_i |U_i|^\alpha.$$

The *Hausdorff measure* (of dimension α) of the metric space Ω is:

$$\text{mes}(\Omega, \alpha) = \liminf_{|\mathcal{U}| \rightarrow 0} \text{mes}(\mathcal{U}, \alpha),$$

where \mathcal{U} is at most a countable covering of Ω . Obviously,

$$\text{mes}(\mathcal{U}, \beta) \leq \text{mes}(\mathcal{U}, \alpha) |\mathcal{U}|^{\beta-\alpha} \quad \text{if } \beta \geq \alpha.$$

This implies the basic property of the Hausdorff measure: if $\text{mes}(\Omega, \alpha) < \infty$ for some α , then $\text{mes}(\Omega, \beta) = 0$ for all $\beta > \alpha$.

The *Hausdorff dimension* of the space Ω is the number:

$$\dim_H \Omega = \inf \{ \alpha \mid \text{mes}(\Omega, \alpha) = 0 \}. \quad (19)$$

In other words, $\dim_H \Omega = \alpha_0$ if $\text{mes}(\Omega, \alpha) = 0$ for all $\alpha > \alpha_0$ and $\text{mes}(\Omega, \alpha) = \infty$ for all $\alpha < \alpha_0$.

Note that, according to this definition, we obtain $\dim_H \Omega = -\infty$ when $\Omega = \emptyset$ and $\dim_H \Omega = 0$ when Ω is nonempty finite or countable.

Below, we will consider the space of sequences:

$$X^{\mathbb{N}} = \{x = (x_1, x_2, x_3, \dots)\}, \quad x_i \in X = \{1, \dots, r\}.$$

Let $x = (x_1, x_2, \dots) \in X^{\mathbb{N}}$. Denote by $Z_n(x)$ the set of all sequences $y = (y_1, y_2, \dots)$ whose first n coordinates coincide with the same coordinates of x . This set will be called a *cylinder of rank n* . In particular, $Z_0(x) = X^{\mathbb{N}}$. All cylinders generate the *Tychonoff topology* on the space $X^{\mathbb{N}}$ and the *cylinder σ -algebra* of subsets in $X^{\mathbb{N}}$.

On the set of all cylinders, we fix an arbitrary positive function η that possesses the following two properties: first, if $Z_n(x) \subset Z_m(y)$, then $\eta(Z_n(x)) \leq \eta(Z_m(y))$ and, second, $\eta(Z_n(x)) \rightarrow 0$ as $n \rightarrow \infty$ at each point $x \in X^{\mathbb{N}}$. Define the cylindrical metric on $X^{\mathbb{N}}$ by means of the formula:

$$\text{dist}(x, y) = \eta(Z_n(x)), \quad \text{where } n = \max \{m \mid Z_m(x) = Z_m(y)\}. \quad (20)$$

Evidently, the diameter of $Z_n(x)$ in this metric coincides with $\eta(Z_n(x))$.

Suppose that in addition to cylindrical metric (20) on $X^{\mathbb{N}}$, there is given a Borel measure μ on $X^{\mathbb{N}}$. The function:

$$d_\mu(x) = \liminf_{n \rightarrow \infty} \frac{\ln \mu(Z_n(x))}{\ln |Z_n(x)|}$$

is called the (*lower*) *pointwise dimension* of the measure μ .

The next theorem provides an effective tool for computing the Hausdorff dimensions of various subsets of $X^{\mathbb{N}}$.

Theorem 4. Suppose $A \subset X^{\mathbb{N}}$. If there exists a finite Borel measure μ on $X^{\mathbb{N}}$, such that $d_{\mu}(x) \leq d$ for each point $x \in A$, then $\dim_H A \leq d$. On the contrary, if $d_{\mu}(x) \geq d$ for each $x \in A$ and the outer measure $\mu^*(A)$ is positive, then $\dim_H A \geq d$.

It follows that if $d_{\mu}(x) \equiv d$ on the whole of the subset $A \subset X^{\mathbb{N}}$, then its dimension is equal to d .

A weakened version of the second part of Theorem 4 in which the condition $d_{\mu}(x) \geq d$ is replaced by a more strong one $\mu(Z_n(x)) \leq |Z_n(x)|^d$ is usually called the *mass distribution principle*.

Theorem 4 was first proven by Billingsley in the case when the function η in Formula (20) is a probability measure on $X^{\mathbb{N}}$ (see Theorems 2.1 and 2.2 in [11]). An analog to this theorem for subsets $A \subset \mathbb{R}^r$ was proven in works [12,13].

Proof. Every cylinder $Z_n(x)$ is, in fact, a ball in metric (20), whose radius equals its diameter, and *vice versa*, any ball in this metric coincides with a cylinder. Besides, any two cylinders $Z_n(x)$ and $Z_m(y)$ either have an empty intersection or one of them is embedded in the other. Therefore, while computing the Hausdorff measure and the dimension of a subset $A \subset X^{\mathbb{N}}$, it is enough to operate only with disjoint coverings of A by cylinders.

Suppose first that $d_{\mu}(x) < \alpha$ for all points $x \in A$. Then, for each $x \in A$, there exist arbitrarily small cylinders $Z_n(x)$ satisfying the condition $|Z_n(x)|^{\alpha} < \mu(Z_n(x))$. Using cylinders of this kind, we can arrange a disjoint covering \mathcal{U} of the set A of arbitrarily small diameter. For this covering, we have:

$$\text{mes}(\mathcal{U}, \alpha) = \sum_{Z_n(x) \in \mathcal{U}} |Z_n(x)|^{\alpha} \leq \sum_{Z_n(x) \in \mathcal{U}} \mu(Z_n(x)) \leq \mu(X^{\mathbb{N}}),$$

and hence, $\dim_H A \leq \alpha$. Thus, the first part of the theorem is proven.

Suppose now that $d_{\mu}(x) > \alpha$ for all points $x \in A$. Define the sets:

$$A_{\varepsilon} = \{x \in A \mid |Z_n(x)|^{\alpha} > \mu(Z_n(x)) \text{ whenever } |Z_n(x)| < \varepsilon\}.$$

Obviously, $A = \bigcup_{\varepsilon > 0} A_{\varepsilon}$. Hence, there exists an ε , such that $\mu^*(A_{\varepsilon}) > 0$. Let \mathcal{U} be a disjoint covering of A by cylinders of diameters less than ε . From the definition of A_{ε} , it follows that $\text{mes}(\mathcal{U}, \alpha) \geq \mu^*(A_{\varepsilon})$. Therefore, $\dim_H A \geq \alpha$, and thus, the second part of the theorem is proven. \square

Each point $x = (x_1, x_2, \dots) \in X^{\mathbb{N}}$ generates a sequence of empirical measures $\delta_{x,n}$ on the set X :

$$\delta_{x,n}(i) = \frac{\#\{t \mid x_t = i, t \leq n\}}{n}, \quad i \in X.$$

In other words, $\delta_{x,n}(i)$ is the fraction of coordinates of the vector (x_1, \dots, x_n) coinciding with i .

For every probability measure $\nu \in M_1(X)$, let us define its *basin* $B(\nu)$ as the set of all points $x \in X^{\mathbb{N}}$, such that $\delta_{x,n}$ converges to ν :

$$B(\nu) = \{x \in X^{\mathbb{N}} \mid \delta_{x,n} \rightarrow \nu\}.$$

Evidently, basins of different measures do not intersect each other and are nonempty. If $x \in B(\nu)$ and $y \in X^{\mathbb{N}}$ differs from x in only finite number of coordinates, then $y \in B(\nu)$. This implies the density of each basin in $X^{\mathbb{N}}$.

Every measure $\nu \in M_1(X)$ generates the Bernoulli distribution $P_{\nu} = \nu^{\mathbb{N}}$ on $X^{\mathbb{N}}$. By the strong law of large numbers, the basin $B(\nu)$ has probability one with respect to Bernoulli distribution P_{ν} , and its complement has zero probability P_{ν} . In particular, any basin different from $B(\nu)$ has zero probability.

Points that do not belong to the union of all basins will be called *irregular*. The set of irregular points has zero probability with respect to any distribution P_ν , where $\nu \in M_1(X)$. As a result, $X^\mathbb{N}$ turns out to be decomposed into the disjoint union of different basins and the set of irregular points.

Let us fix some numbers $\theta(i) \in (0, 1)$ for all elements $i \in X = \{1, \dots, r\}$ and define a *cylindrical θ -metric* on $X^\mathbb{N}$ by the rule:

$$\text{dist}(x, y) = \prod_{t=1}^n \theta(x_t), \quad n = \inf \{t \mid x_t \neq y_t\} - 1. \quad (21)$$

It is a partial case of cylindrical metric (20).

For each measure $\nu \in M_1(X)$ and θ -metric (21), define the quantity:

$$S(\nu, \theta) = \frac{\sum_{i=1}^r \nu(i) \ln \nu(i)}{\sum_{i=1}^r \nu(i) \ln \theta(i)}. \quad (22)$$

We will call it the *Billingsley entropy*, because he was the first who wrote down this formula and applied it for the computation of the Hausdorff dimensions [10]. He also expressed this quantity in terms of Shannon's entropy and the Kullback action:

$$S(\nu, \theta) = \frac{H(\nu)}{H(\nu) + \rho(\nu, \theta)}.$$

Theorem 5. *The Hausdorff dimension of any basin $B(\nu)$ with respect to θ -metric (21) is equal to the Billingsley entropy $S(\nu, \theta)$.*

A partial case of this theorem, in which $\theta(1) = \dots = \theta(r) = 1/r$, was first proven by Eggleston [17]. In the complete form, this theorem and its generalizations were proven by Billingsley [10,11].

Proof. Assume first that $\nu(i) > 0$ for every $i = 1, \dots, r$. Obviously,

$$\frac{\ln P_\nu(Z_n(x))}{\ln |Z_n(x)|} = \frac{\sum_{t=1}^n \ln \nu(x_t)}{\sum_{t=1}^n \ln \theta(x_t)} = \frac{\sum_{i=1}^r n \delta_{x,n}(i) \ln \nu(i)}{\sum_{i=1}^r n \delta_{x,n}(i) \ln \theta(i)}.$$

Hence, for each point $x \in B(\nu)$, we have:

$$d_{P_\nu}(x) = \liminf_{n \rightarrow \infty} \frac{\ln P_\nu(Z_n(x))}{\ln |Z_n(x)|} = \frac{\sum_{i=1}^r \nu(i) \ln \nu(i)}{\sum_{i=1}^r \nu(i) \ln \theta(i)} = S(\nu, \theta). \quad (23)$$

Applying Theorem 4 to the set $A = B(\nu)$ and measure $\mu = P_\nu$, we obtain the statement of Theorem 5.

In the general case, the same argument provides only the lower bound $d_{P_\nu}(x) \geq S(\nu, \theta)$, which implies the lower bound $\dim_H B(\nu) \geq S(\nu, \theta)$. The inverse inequality is provided by the next lemma. \square

Lemma 1. *Suppose that $X^\mathbb{N}$ is equipped with metric (21). Then, for any measure $\nu \in M_1(X)$ and $\varepsilon > 0$, there exists a neighborhood $O(\nu)$, such that the Hausdorff dimension of the set:*

$$A = \{x \in X^\mathbb{N} \mid \forall N \exists n > N: \delta_{x,n} \in O(\nu)\}$$

does not exceed $S(\nu, \theta) + \varepsilon$.

Proof. Fix a measure $\nu \in M_1(X)$ and an arbitrary positive number \varkappa . By McMillan's theorem, there exists a neighborhood $O(\nu)$, such that for each positive integer n :

$$\#\{Z_n(x) \mid \delta_{x,n} \in O(\nu)\} \leq e^{n(H(\nu)+\varkappa)}. \quad (24)$$

Decrease this neighborhood in such a way that, in addition to Formula (24), for every measure $\delta \in O(\nu)$, the next inequality holds:

$$\sum_{i=1}^r \delta(i) \ln \theta(i) < \sum_{i=1}^r \nu(i) \ln \theta(i) + \varkappa.$$

Then, for every cylinder $Z_n(x)$ satisfying the condition $\delta_{x,n} \in O(\nu)$, we have the estimate:

$$\begin{aligned} |Z_n(x)| &= \prod_{t=1}^n \theta(x_t) = \exp\left\{\sum_{t=1}^n \ln \theta(x_t)\right\} = \exp\left\{n \sum_{i=1}^r \delta_{x,n}(i) \ln \theta(i)\right\} \\ &< \exp\left\{n \sum_{i=1}^r \nu(i) \ln \theta(i) + n\varkappa\right\}. \end{aligned} \quad (25)$$

For any positive integer N , the set A is covered by the collection of cylinders:

$$\mathcal{U}_N = \bigcup_{n=N}^{\infty} \{Z_n(x) \mid \delta_{x,n} \in O(\nu)\}.$$

Evidently, the diameter of this covering goes to zero when N increases. Now, we can evaluate $\text{mes}(\mathcal{U}_N, \alpha)$ by means of Formulas (24) and (25):

$$\begin{aligned} \text{mes}(\mathcal{U}_N, \alpha) &= \sum_{Z_n(x) \in \mathcal{U}_N} |Z_n(x)|^\alpha \\ &\leq \sum_{n=N}^{\infty} e^{n(H(\nu)+\varkappa)} \exp\left\{\alpha n \sum_{i=1}^r \nu(i) \ln \theta(i) + \alpha n \varkappa\right\} \\ &= \sum_{n=N}^{\infty} \exp\left\{n \left(-\sum_{i=1}^r \nu(i) \ln \nu(i) + \alpha \sum_{i=1}^r \nu(i) \ln \theta(i) + \varkappa + \alpha \varkappa\right)\right\}. \end{aligned} \quad (26)$$

If $\alpha > S(\nu, \theta)$, then we can choose $\varkappa > 0$ so small that the exponent in braces in Formula (26) is negative, and all of the sum in Formula (26) goes to zero as $N \rightarrow \infty$. Therefore, the Hausdorff measure (of dimension α) of the set A is zero, and hence, $\dim_H A$ does not exceed α . \square

We will say that a sequence of empirical measures $\delta_{x,n}$ *condenses* on a subset $V \subset M_1(X)$ (notation: $\delta_{x,n} \succ V$) if it has at least one limit point in V .

Similar to the known large deviations principle by Varadhan [15,16], it is natural to call the next theorem the *maximal dimension principle*.

Theorem 6. Let the space $X^{\mathbb{N}}$ be equipped with cylindrical θ -metric (21). Then, for any nonempty subset $V \subset M_1(X)$:

$$\dim_H \{x \in X^{\mathbb{N}} \mid \delta_{x,n} \succ V\} = \sup_{\nu \in V} S(\nu, \theta). \quad (27)$$

Proof. The set $A = \{x \in X^{\mathbb{N}} \mid \delta_{x,n} \succ V\}$ contains basins of all measures $\nu \in V$. Therefore, by Theorem 5, its dimension is not less than the right-hand side of Formula (27).

It is easily seen from the definition (22) of the Billingsley entropy $S(\nu, \theta)$ that it depends continuously on the measure $\nu \in M_1(X)$. Consider the closure \bar{V} of V . Obviously, it is compact. Fix any $\varepsilon > 0$. By Lemma 1 for any measure $\nu \in \bar{V}$, there exists a neighborhood $O(\nu)$, such that:

$$\dim_H \{x \in X^{\mathbb{N}} \mid \delta_{x,n} \succ O(\nu)\} \leq S(\nu, \theta) + \varepsilon \leq \sup_{\nu \in V} S(\nu, \theta) + \varepsilon. \quad (28)$$

Let us take a finite covering of \bar{V} composed of neighborhoods of this sort. Then, the set $A = \{x \in X^{\mathbb{N}} \mid \delta_{x,n} \succ V\}$ will be covered by a finite collection of sets of the form $\{x \in X^{\mathbb{N}} \mid \delta_{x,n} \succ O(\nu)\}$ satisfying inequality Formula (28). By the arbitrariness of ε , this implies the statement of Theorem 6. \square

A result very similar to Theorem 6 was proven by Billingsley in Theorem 7.1 in [10].

7. Branching Processes

First, let us recall the basic notions about the simplest Galton–Watson branching process.

Suppose that a random variable Z takes nonnegative values $k \in \mathbb{Z}_+$ with probabilities p_k . The *Galton–Watson branching process* is a sequence of random variables Z_0, Z_1, Z_2, \dots , such that $Z_0 \equiv 1$, $Z_1 = Z$, and further, each Z_{n+1} is defined as the sum of Z_n independent counterparts of the random variable Z . In particular, if $Z_n = 0$, then $Z_{n+1} = 0$, as well. Usually, Z_n is understood as the total number of descendants in the n -th generation of a unique common ancestor under the condition that each descendant independently of the others gives birth to Z children.

It is known that in some cases, the posterity of the initial ancestor may degenerate (when starting from a certain n , all Z_n are zeros), and in other cases, it can “flourish” (when Z_n grows exponentially). The type of behavior of a branching process depends on the mean number of children of any individual:

$$m = \mathbb{E}Z = \sum_{k=0}^{\infty} kp_k$$

and on the generating function of that number:

$$f(s) = f_1(s) = \sum_{k=0}^{\infty} p_k s^k.$$

Obviously, the restriction of the function $f(s)$ to the segment $[0, 1]$ is nonnegative, nondecreasing, convex and satisfies $f(1) = 1$ and $f'(1) = m$.

In the theory of branching processes (see, for instance, [18,19]), the following statements were proven.

Theorem 7. *The generating functions of the number of descendants in the n -th generation:*

$$f_n(s) = \sum_{k=0}^{\infty} \mathbb{P}\{Z_n = k\} s^k$$

satisfy the recursion relation $f_{n+1}(s) = f(f_n(s))$.

Theorem 8. If $m \leq 1$, then the branching process degenerates almost surely (except the case when each individual gives birth to exactly one child). If $m > 1$, then the probability of degeneration q is less than one and coincides with a unique non-unit root of the equation $f(s) = s$ on the segment $[0, 1]$.

Theorem 9. If $m > 1$ and $EZ^2 < \infty$, then the sequence $W_n = Z_n/m^n$ converges almost surely to a random variable W , such that $P\{W > 0\} = 1 - q$. If $m > 1$ and $EZ^2 = \infty$, then, for any number $m' < m$, with probability $1 - q$:

$$\lim_{n \rightarrow \infty} Z_n/m^n < \infty, \quad \lim_{n \rightarrow \infty} Z_n/(m')^n = \infty$$

(here, q is the probability of degeneration of the branching process).

Thereby, in the case $m > 1$, there is an alternative for the total number of descendants Z_n : either it vanishes at a certain moment n_0 (with probability $q < 1$) or it is asymptotically equivalent to Wm^n (with the complementary probability $1 - q$), where the random variable $W > 0$ does not depend on n (except the case $EZ^2 = \infty$, when only the logarithmic equivalence $\ln Z_n \sim \ln m^n$ is guaranteed). All of the other types of behavior of descendants' totality have zero probability.

We will exploit these theorems in the study of colored branching processes.

Suppose now that each individual may give birth to children of r different colors (or r different genders, if one likes). We will assume that the posterity of each individual in the first generation is a random set X containing a random number k_1 of children of the first color, a random number k_2 of children of the second color, and so on, up to a random number k_r of children of color r . All of the elements of X (including elements of the same color) are treated as different. The ordered array $k = (k_1, k_2, \dots, k_r) \in \mathbb{Z}_+^r$ will be called the *color structure* of the set X of children. Denote by p_k the probability of the birth of the set X with color structure $k = (k_1, k_2, \dots, k_r)$. Naturally, all of the probabilities p_k are nonnegative and:

$$\sum_{k \in \mathbb{Z}_+^r} p_k = 1.$$

If an individual x_1 gave birth to x_2 , then x_2 gave birth to x_3 , and so on, up to an individual x_n , then the sequence $x = (x_1, \dots, x_n)$ will be called a *genetic line* of length n .

Let us construct a new branching process taking into account not only the total number of descendants, but also the color of each individual and all its upward and downward lineal relations. This process may be treated as a random *genealogical tree* with a common ancestor in the root and all its descendants in the vertices, where each parent is linked with all its children. In the case of a degenerating population, its genealogical tree is finite, and in the case of a “flourishing” one, the tree is infinite.

Formally, it is convenient to define such a process as a sequence of random sets X_n containing all genetic lines of length n . As the first set X_1 , we take X . The subsequent sets X_n are built by induction: if X_n is already known, then for all genetic lines $(x_1, \dots, x_n) \in X_n$, one defines disjoint independent random sets of children $X(x_1, \dots, x_n)$, each with a color structure distribution as in X , and X_{n+1} is given by:

$$X_{n+1} = \{(x_1, \dots, x_n, x_{n+1}) \mid (x_1, \dots, x_n) \in X_n, x_{n+1} \in X(x_1, \dots, x_n)\}.$$

The stochastic process X_1, X_2, X_3, \dots built in this way will be referred to as a *colored branching process* (or *unconditional colored branching process*, if one wishes to emphasize that the posterity of any individual is independent of its color and genealogy).

8. The McMillan Theorem for Colored Branching Processes

Consider a colored branching process X_1, X_2, X_3, \dots with a given finite collection of colors $\Omega = \{1, \dots, r\}$ and a probability distribution $\{p_k \mid k \in \mathbb{Z}_+^r\}$, where $k = (k_1, \dots, k_r)$ is the color structure of each individual's set of children X . We will always assume that $X_1 = X$ is generated by a unique initial individual.

For any genetic line $x = (x_1, \dots, x_n) \in X_n$, define the *spectrum* $\delta_{x,n}$ as the corresponding empirical measure on Ω given by the rule:

$$\delta_{x,n}(i) = \frac{\#\{t \mid g(x_t) = i\}}{n}, \quad i \in \Omega, \quad (29)$$

where $g(x_t)$ denotes the color of x_t . In other words, $\delta_{x,n}(i)$ is the fraction of individuals of color i in the genetic line x . Our next goal is to obtain asymptotical estimates for cardinalities of the random sets:

$$\{x = (x_1, \dots, x_n) \in X_n \mid \delta_{x,n} \in O(\nu)\},$$

where $O(\nu)$ is a small neighborhood of the distribution ν on the set of colors Ω .

Denote by $\mu(i)$ the expectation of members of color i in X :

$$\mu(i) = \sum_{k \in \mathbb{Z}_+^r} k_i p_k, \quad i = 1, \dots, r. \quad (30)$$

Provided all $\mu(i)$ are finite, the vector $\mu = (\mu(1), \dots, \mu(r))$ can be regarded as a measure on the set of colors Ω . This measure generates the measure μ^n on Ω^n as a Cartesian product.

Define a mapping $G : X_n \rightarrow \Omega^n$ by means of the formula:

$$G(x_1, \dots, x_n) = (g(x_1), \dots, g(x_n)),$$

where $g(x_t)$ is the color of x_t .

Lemma 2. For any $\omega = (\omega_1, \dots, \omega_n) \in \Omega^n$, we have:

$$\mathbb{E} \#\{x \in X_n \mid G(x) = \omega\} = \prod_{t=1}^n \mu(\omega_t) = \mu^n(\omega). \quad (31)$$

Proof. Discard the last coordinate in ω , and let $\omega' = (\omega_1, \dots, \omega_{n-1})$. For every genetic line $(x_1, \dots, x_{n-1}) \in X_{n-1}$, by virtue of the definition of unconditional colored branching process, we have:

$$\mathbb{E} \#\{x_n \in X(x_1, \dots, x_{n-1}) \mid g(x_n) = \omega_n\} = \mu(\omega_n).$$

Evidently, this expression does not depend on $x' = (x_1, \dots, x_{n-1})$. Therefore,

$$\mathbb{E} \#\{x \in X_n \mid G(x) = \omega\} = \mathbb{E} \#\{x' \in X_{n-1} \mid G(x') = \omega'\} \mu(\omega_n).$$

Repeated application of the latter equality gives Formula (31). \square

For the measure μ from Formula (30), let us define the Kullback action:

$$\rho(\nu, \mu) = \sum_{i \in \Omega} \nu(i) \ln \frac{\nu(i)}{\mu(i)}, \quad \nu \in M_1(\Omega),$$

where $M_1(\Omega)$ is the set of all probability measures on Ω . This formula is a copy of Formula (9).

Theorem 10. Suppose X_1, X_2, X_3, \dots is an unconditional colored branching process with a finite collection of colors Ω . Then, for any $\varepsilon > 0$ and probability measure $\nu \in M_1(\Omega)$, there exists a neighborhood $O(\nu) \subset M_1(\Omega)$, such that for all natural n :

$$\mathbb{E} \# \{x \in X_n \mid \delta_{x,n} \in O(\nu)\} \leq e^{n(-\rho(\nu, \mu) + \varepsilon)}. \quad (32)$$

On the other hand, for any $\varepsilon > 0$ and any neighborhood $O(\nu)$:

$$\mathbb{E} \# \{x \in X_n \mid \delta_{x,n} \in O(\nu)\} \geq e^{n(-\rho(\nu, \mu) - \varepsilon)} \quad \text{as } n \rightarrow \infty. \quad (33)$$

If $\rho(\nu, \mu) = +\infty$, the expression $-\rho(\nu, \mu) + \varepsilon$ in Formula (32) should be treated as an arbitrary negative real number.

Proof. It follows from Formula (29) that, for every genetic line $x \in X_n$, its spectrum $\delta_{x,n}$ coincides with the empirical measure $\delta_{\omega,n}$, where $\omega = G(x)$. Therefore,

$$\# \{x \in X_n \mid \delta_{x,n} \in O(\nu)\} = \sum_{\omega \in \Omega^n: \delta_{\omega,n} \in O(\nu)} \# \{x \in X_n \mid G(x) = \omega\}. \quad (34)$$

It follows from Formulas (31) and (34) that:

$$\mathbb{E} \# \{x \in X_n \mid \delta_{x,n} \in O(\nu)\} = \mu^n \{\omega \in \Omega^n \mid \delta_{\omega,n} \in O(\nu)\}.$$

The latter equality converts estimate Formulas (32) and (33) into already proven estimate Formulas (14) and (15) from the large deviations principle. \square

It is remarkable that the last reference to the large deviations principle serves as a unique “umbilical cord” linking the first three Sections of the paper with the others.

Now, we are ready to state an analog of the McMillan theorem for colored branching processes. Let q^* be the probability of degeneration of the process (the probability of the event when starting from a certain number n all of the sets X_n turn out to be empty).

Theorem 11. Suppose that X_1, X_2, \dots is an unconditional colored branching process with a finite collection of colors Ω . Then, for any $\varepsilon > 0$ and any probability measure $\nu \in M_1(\Omega)$, there exists a neighborhood $O(\nu) \subset M_1(\Omega)$, such that almost surely:

$$\# \{x \in X_n \mid \delta_{x,n} \in O(\nu)\} < e^{n(-\rho(\nu, \mu) + \varepsilon)} \quad \text{as } n \rightarrow \infty. \quad (35)$$

On the other hand, if $\rho(\nu, \mu) < 0$, then, for any neighborhood $O(\nu)$ and positive ε , the estimate:

$$\# \{x \in X_n \mid \delta_{x,n} \in O(\nu)\} > e^{n(-\rho(\nu, \mu) - \varepsilon)} \quad \text{as } n \rightarrow \infty \quad (36)$$

holds with probability $1 - q^*$ (or almost surely under the condition that our branching process does not degenerate).

Proof. The application of Chebyshev’s inequality to Formula (32) gives:

$$P\{\#\{x \in X_n \mid \delta_{x,n} \in O(\nu)\} \geq e^{n(-\rho(\nu,\mu)+2\varepsilon)}\} \leq e^{-n\varepsilon}.$$

Summing up these inequalities over all $n \geq N$, one obtains:

$$P\{\exists n \geq N: \#\{x \in X_n \mid \delta_{x,n} \in O(\nu)\} \geq e^{n(-\rho(\nu,\mu)+2\varepsilon)}\} \leq \frac{e^{-N\varepsilon}}{1 - e^{-\varepsilon}}.$$

This implies Formula (35) with constant 2ε instead of ε , which does not change its sense.

Now, we proceed to the second part of the theorem. Let $\varkappa = -\rho(\nu, \mu) - \varepsilon$ and, at the same time, the number ε be so small that $\varkappa > 0$. By the second part of Theorem 10, for any neighborhood $O(\nu)$, there exists N , such that:

$$E \#\{x \in X_N \mid \delta_{x,N} \in O(\nu)\} > e^{N\varkappa}.$$

Without loss of generality, we may assume that $O(\nu)$ is convex.

Now, we wish to construct a Galton–Watson branching process Z_1, Z_2, \dots satisfying the conditions:

$$Z_1 = \#\{x \in X_N \mid \delta_{x,N} \in O(\nu)\}, \quad (37)$$

$$Z_n \leq \#\{x \in X_{nN} \mid \delta_{x,nN} \in O(\nu)\}, \quad n = 2, 3, \dots \quad (38)$$

Let the random variable Z_1 be defined by Formula (37). For $n > 1$, set Z_n to be the total number of genetic lines:

$$(x_1, \dots, x_N, \dots, x_{(n-1)N+1}, \dots, x_{nN}) \in X_{nN}$$

such that the spectrum of each segment $(x_{kN+1}, \dots, x_{(k+1)N})$ belongs to $O(\nu)$. In other words, we treat as “individuals” of the process Z_1, Z_2, Z_3, \dots those segments $(x_{kN+1}, \dots, x_{(k+1)N})$ of genetic lines of the initial process whose spectrum lies in $O(\nu)$. Now, Formula (38) follows from convexity of $O(\nu)$; and from the unconditionality of the initial colored branching process, it can be concluded that the sequence Z_1, Z_2, \dots in fact forms a Galton–Watson branching process.

By construction, one has $E Z_1 > e^{N\varkappa}$. In this setting, Theorem 9 asserts that there is an alternative for the sequence Z_n : either it tends to zero with a certain probability $q < 1$ or it grows faster than $e^{nN\varkappa}$ with probability $1 - q$. In the second case, Formula (38) implies:

$$\#\{x \in X_{nN} \mid \delta_{x,nN} \in O(\nu)\} e^{-nN\varkappa} \rightarrow \infty, \quad n \rightarrow \infty. \quad (39)$$

To finish the proof, we have to do two things: verify that Formula (39) is valid with probability $1 - q^*$ and get rid of the multiplier N there. To do this, we will exploit two ideas. First, if the colored branching process X_1, X_2, \dots had been generated by m initial individuals instead of a unique one, then Formula (39) would be valid with a probability of at least $1 - q^m$. Second, if one genetic line is a part of another one and the ratio of their lengths is close to one, then their spectra are close to each other, as well.

Obviously, the total number of individuals in the n -th generation of the initial branching process X_1, X_2, X_3, \dots equals $|X_n|$. The sequence of random variables $|X_n|$ forms a Galton–Watson branching process with the probability of degeneration q^* , that does not exceed q . Therefore, the sequence $|X_n|$ grows exponentially with probability $1 - q^*$.

Consider the colored branching process $X_{k+1}, X_{k+2}, X_{k+3}, \dots$ obtained from the initial one by truncation of the first k generations. It can be thought of as a union of $|X_k|$ independent branching processes generated by all individuals of k -th generation. Therefore, it satisfies Formula (39) with probability at least $1 - q^{|X_k|}$. Hence, for the initial process with even greater probability, we obtain the condition:

$$\#\{x \in X_{k+nN} \mid \delta_{x,k+nN} \in O^*(\nu)\} e^{-nN\epsilon} \rightarrow \infty, \quad n \rightarrow \infty, \quad (40)$$

where $O^*(\nu)$ is an arbitrary neighborhood of ν containing the closure of $O(\nu)$.

Suppose that the sequence $|X_n|$ grows exponentially. For every $m \in \mathbb{N}$, we set the numbers:

$$k_i = \min\{k : |X_k| \geq m, \quad k = i \bmod N\}, \quad i = 0, 1, \dots, N-1.$$

For each $k = k_i$, condition (40) holds with probability at least $1 - q^m$, and together, they give the estimate:

$$\#\{x \in X_n \mid \delta_{x,n} \in O^*(\nu)\} > e^{n\epsilon}, \quad n \rightarrow \infty,$$

with probability at least $1 - Nq^m$. By the arbitrariness of m , this estimate is valid almost surely (under the condition $|X_n| \rightarrow \infty$, which takes place with probability $1 - q^*$). It is equivalent to Formula (36). \square

9. Dimensions of Random Fractals (Upper Bounds)

Here, we continue the investigation of a colored branching process X_1, X_2, \dots with a finite collection of colors $\Omega = \{1, \dots, r\}$. Let us consider the corresponding set of infinite genetic lines:

$$X_\infty = \{x = (x_1, x_2, x_3, \dots) \mid (x_1, \dots, x_n) \in X_n \quad \forall n \in \mathbb{N}\}.$$

Define the cylindrical θ -metric on X_∞ :

$$\text{dist}(x, y) = \prod_{t=1}^n \theta(x_t), \quad n = \inf\{t \mid x_t \neq y_t\} - 1, \quad (41)$$

where the numbers $\theta(1), \dots, \theta(r)$ are taken from $(0, 1)$.

We will be interested in the Hausdorff dimensions of both the space X_∞ and its various subsets, defined in terms of the partial limits of empirical measures on Ω (we call these measures spectra; see Formula (29), and denote by $\delta_{x,n}$). If the colored branching process degenerates, then $X_\infty = \emptyset$. Therefore, the only case of interest is when $m = E|X_1| > 1$, and the cardinality of X_n increases with a rate of order m^n .

As before, we denote by $\mu(i)$, where $i \in \Omega$, the expectation of individuals of color i in the random set X_1 . It will be always assumed that $\mu(i) < \infty$. Consider any probability measure $\nu \in M_1(\Omega)$. It will be proven below that the dimension of the set $\{x \in X_\infty \mid \delta_{x,n} \rightarrow \nu\}$ can be computed by means of the function:

$$d(\nu, \mu, \theta) = \frac{\rho(\nu, \mu)}{\sum_{i=1}^r \nu(i) \ln \theta(i)} = \frac{\sum_{i=1}^r \nu(i) \ln \frac{\nu(i)}{\mu(i)}}{\sum_{i=1}^r \nu(i) \ln \theta(i)}. \quad (42)$$

We will call it the *Billingsley–Kullback entropy*.

The numerator in Formula (42) is the Kullback action, and the denominator is negative. If μ is a probability measure on Ω , then the Kullback action is nonnegative. However, in our setting, this is not the case, since $m = \mu(1) + \cdots + \mu(r) > 1$. In particular, if $\mu(i) > \nu(i)$, for all $i \in \Omega$, then the Kullback action is negative, and the Billingsley–Kullback entropy is positive. Note, in addition, that if $\mu(1) = \cdots = \mu(r) = 1$, then $-\rho(\nu, \mu)$ is equal to Shannon’s entropy $H(\nu)$, and the Billingsley–Kullback entropy turns into the Billingsley entropy (22).

Lemma 3. *Let the space X_∞ of infinite genetic lines be equipped with metric (41). Then, for any probability measure $\nu \in M_1(\Omega)$ and any $\varepsilon > 0$, there exists a neighborhood $O(\nu)$, such that the Hausdorff dimension of the set:*

$$A = \{x \in X_\infty \mid \forall N \exists n > N: \delta_{x,n} \in O(\nu)\}$$

does not exceed $d(\nu, \mu, \theta) + \varepsilon$ almost surely.

The proof is carried out in the same manner as in Lemma 1. Take any $\varkappa > 0$. By Theorem 32, there exists a neighborhood $O(\nu)$, such that almost surely:

$$\#\{x \in X_n \mid \delta_{x,n} \in O(\nu)\} \leq e^{n(-\rho(\nu, \mu) + \varkappa)} \quad \text{as } n \rightarrow \infty. \quad (43)$$

Let us reduce this neighborhood in such a way that in addition to Formula (43), for all $\delta \in O(\nu)$,

$$\sum_{i=1}^r \delta(i) \ln \theta(i) < \sum_{i=1}^r \nu(i) \ln \theta(i) + \varkappa.$$

Then, for each cylinder $Z_n(x)$ satisfying the condition $\delta_{x,n} \in O(\nu)$, the following estimate holds:

$$\begin{aligned} |Z_n(x)| &= \prod_{t=1}^n \theta(x_t) = \exp\left\{\sum_{t=1}^n \ln \theta(x_t)\right\} = \exp\left\{n \sum_{i=1}^r \delta_{x,n}(i) \ln \theta(i)\right\} \\ &< \exp\left\{n \sum_{i=1}^r \nu(i) \ln \theta(i) + n\varkappa\right\}. \end{aligned} \quad (44)$$

For every natural N , the set A is covered by the collection of cylinders:

$$\mathcal{U}_N = \bigcup_{n=N}^{\infty} \{Z_n(x) \mid \delta_{x,n} \in O(\nu)\}.$$

Evidently, the diameter of this covering tends to zero as $N \rightarrow \infty$. Hence, $\text{mes}(\mathcal{U}_N, \alpha)$ can be estimated by means of Formulas (43) and (44):

$$\begin{aligned} \text{mes}(\mathcal{U}_N, \alpha) &= \sum_{Z_n(x) \in \mathcal{U}_N} |Z_n(x)|^\alpha \\ &\leq \sum_{n=N}^{\infty} e^{n(-\rho(\nu, \mu) + \varkappa)} \exp\left\{\alpha n \sum_{i=1}^r \nu(i) \ln \theta(i) + \alpha n \varkappa\right\} \\ &= \sum_{n=N}^{\infty} \exp\left\{n \left(-\sum_{i=1}^r \nu(i) \ln \frac{\nu(i)}{\mu(i)} + \alpha \sum_{i=1}^r \nu(i) \ln \theta(i) + \varkappa + \alpha \varkappa\right)\right\}. \end{aligned} \quad (45)$$

If $\alpha > d(\nu, \mu, \theta)$, then \varkappa can be chosen so small, that the last exponent in braces is negative, and all of the sum (45) tends to zero as $N \rightarrow \infty$. Thus, the Hausdorff measure (of dimension α) of the set A is zero, and its dimension does not exceed α . \square

As before, we say that a sequence of empirical measures $\delta_{x,n}$ condenses on a subset $V \subset M_1(\Omega)$ (notation $\delta_{x,n} \succ V$) if it has a limit point in V .

Theorem 12. Let X_1, X_2, \dots be an unconditional colored branching process with a finite set of colors Ω , and let the set X_∞ of all infinite genetic lines be equipped with a cylindrical θ -metric (41). Then, for any subset $V \subset M_1(\Omega)$, almost surely:

$$\dim_H \{x \in X_\infty \mid \delta_{x,n} \succ V\} \leq \sup_{\nu \in V} d(\nu, \mu, \theta). \quad (46)$$

In particular, $\dim_H X_\infty \leq s$ almost surely, where s is a unique root of the “Bowen equation”:

$$\sum_{i=1}^r \mu(i) \theta(i)^s = 1. \quad (47)$$

Proof. It follows from the definition of the Billingsley–Kullback entropy $d(\nu, \mu, \theta)$ that it depends continuously on the measure $\nu \in M_1(\Omega)$. Let \bar{V} be the closure of V . Obviously, it is compact. Take an arbitrary $\varepsilon > 0$. By Lemma 3, for any measure $\nu \in \bar{V}$, there exists a neighborhood $O(\nu)$, such that almost surely:

$$\dim_H \{x \in X_\infty \mid \delta_{x,n} \succ O(\nu)\} \leq d(\nu, \mu, \theta) + \varepsilon \leq \sup_{\nu \in V} d(\nu, \mu, \theta) + \varepsilon. \quad (48)$$

Choose a finite covering of the set \bar{V} by neighborhoods of this kind. Then, the set $\{x \in X_\infty \mid \delta_{x,n} \succ V\}$ will be covered by a finite collection of sets of the form $\{x \in X_\infty \mid \delta_{x,n} \succ O(\nu)\}$ satisfying Formula (48). By the arbitrariness of ε , this implies the first statement of Theorem 12.

Let s be a solution to Equation (47). Note that for any $\nu \in M_1(\Omega)$, the concavity of the logarithm function implies that:

$$\begin{aligned} s \sum_{i=1}^r \nu(i) \ln \theta(i) - \sum_{i=1}^r \nu(i) \ln \frac{\nu(i)}{\mu(i)} &= \sum_{i=1}^r \nu(i) \ln \frac{\mu(i) \theta(i)^s}{\nu(i)} \\ &\leq \ln \left\{ \sum_{\nu(i) > 0} \nu(i) \frac{\mu(i) \theta(i)^s}{\nu(i)} \right\} \leq 0. \end{aligned}$$

Consequently, $d(\nu, \mu, \theta) \leq s$. Now, the second part of the theorem follows from the first one if we take $V = M_1(\Omega)$. \square

Remark. Usually, the term “Bowen equation” is used for the equation $P(s\varphi) = 0$, where $P(s\varphi)$ is the topological pressure of a weight function $s\varphi$ in some dynamical system (more detailed explanations can be found in [13]). If we replace the topological pressure $P(s\varphi)$ by the spectral potential:

$$\lambda(s\varphi, \mu) = \ln \sum_{i=1}^r e^{s\varphi(i)} \mu(i), \quad \text{where } \varphi(i) = \ln \theta(i),$$

then the Bowen equation turns into the equation $\lambda(s\varphi, \mu) = 0$, which is equivalent to Formula (47).

10. Block Selections of Colored Branching Processes

Let $\xi_1, \xi_2, \xi_3, \dots$ be a sequence of independent identically distributed random variables taking values of zero or one (independent Bernoulli trials).

Lemma 4. *If $0 < p' < p < 1$ and $P\{\xi_i = 1\} \geq p$, then:*

$$P\{\xi_1 + \dots + \xi_k \geq p'k\} \rightarrow 1 \quad \text{as } k \rightarrow \infty \quad (49)$$

uniformly with respect to the probability $P\{\xi_i = 1\} \geq p$.

Proof. In the case $P\{\xi_i = 1\} = p$, this follows from the law of large numbers. If $P\{\xi_i = 1\}$ increases, then the probability in the left-hand side of Formula (49) increases, as well. \square

Consider a colored branching process X_1, X_2, \dots with a finite set of colors $\Omega = \{1, \dots, r\}$. Each X_n consists of genetic lines (x_1, x_2, \dots, x_n) of length n , in which every subsequent individual is born by the previous one. Fix a (large enough) natural N . We will split genetic lines of length divisible by N into blocks of length N :

$$(x_1, x_2, \dots, x_{nN}) = (y_1, \dots, y_n), \quad \text{where } y_k = (x_{(k-1)N+1}, \dots, x_{kN}).$$

Each block y_k generates an empirical measure δ_{y_k} (spectrum) on Ω by the rule:

$$\delta_{y_k}(i) = \frac{\#\{t \mid g(x_t) = i, (k-1)N < t \leq kN\}}{N},$$

where $g(x_t)$ denotes the color of x_t .

A *block selection of order N* from a colored branching process X_1, X_2, \dots is any sequence of random subsets $Y_n \subset X_{nN}$ with the following property: if $(y_1, \dots, y_{n+1}) \in Y_{n+1}$, then $(y_1, \dots, y_n) \in Y_n$. In this case, the sequence of blocks (y_1, \dots, y_{n+1}) will be called a *prolongation* of the sequence (y_1, \dots, y_n) .

As above (see Formula (30)), denote by $\mu(i)$ the expectation of children of color i born by each individual and by μ the corresponding measure on Ω .

Theorem 13. *Let X_1, X_2, \dots be an unconditional colored branching process with a finite set of colors Ω and the probability of degeneration $q^* < 1$. If a measure $\nu \in M_1(\Omega)$ satisfies the condition $\rho(\nu, \mu) < 0$, then for any neighborhood $O(\nu) \subset M_1(\Omega)$ and any number $\varepsilon > 0$ with probability $1 - q^*$, one can extract from the branching process a block selection Y_1, Y_2, \dots of an order N , such that each sequence of blocks $(y_1, \dots, y_n) \in Y_n$ has at least $l(N)$ prolongations in Y_{n+1} , where:*

$$l(N) = e^{N(-\rho(\nu, \mu) - \varepsilon)}, \quad (50)$$

and the spectra of all blocks belong to $O(\nu)$.

Proof. Fix any numbers p and ε satisfying the conditions:

$$0 < p < p + \varepsilon < 1 - q^*, \quad \rho(\nu, \mu) + \varepsilon < 0.$$

By the second part of Theorem 11 for all large enough N , we have:

$$P\{\#\{x \in X_N \mid \delta_{x,N} \in O(\nu)\} > e^{N(-\rho(\nu,\mu)-\varepsilon/2)}\} \geq p + \varepsilon. \quad (51)$$

Further, we will consider finite sequences of random sets X_1, \dots, X_{nN} and extract from them block selections Y_1, \dots, Y_n of order N , such that the spectra of all of their blocks belong to $O(\nu)$ and each sequence of blocks $(y_1, \dots, y_k) \in Y_k$ has at least $l(N)$ prolongations in Y_{k+1} . Denote by A_n the event of the existence of a block selection with these properties. Define one more event A by the condition:

$$\#\{x \in X_N \mid \delta_{x,N} \in O(\nu)\} > l(N)e^{N\varepsilon/2}.$$

It follows from Formulas (50) and (51) that $P(A) \geq p + \varepsilon$. Evidently, $A \subset A_1$. Therefore, $P(A_1) \geq p + \varepsilon$. Now, we are going to prove by induction that $P(A_n) \geq p$ whenever the order N of our block selection is large enough. Let us perform the step of induction. Suppose that $P(A_n) \geq p$ is valid for some n . Consider the conditional probability $P(A_{n+1}|A)$. By the definition of events A_{n+1} and A , it cannot be less than the probability of the following event: there are at least $l(N)$ wins in a sequence of $[l(N)e^{N\varepsilon/2}]$ independent Bernoulli trials with the probability of win $P(A_n)$ in each. Choosing N large enough and using Lemma 4 (with $p' = p/2$ and $k = [l(N)e^{N\varepsilon/2}]$), one can make this probability greater than $1 - \varepsilon$. Then,

$$P(A_{n+1}) \geq P(A)P(A_{n+1}|A) > (p + \varepsilon)(1 - \varepsilon) > p.$$

Thus, the inequality $P(A_n) > p$ is proven for all n .

This means that with probability greater than p , one can extract from the sequence X_1, \dots, X_{nN} a block selection Y_1, \dots, Y_n of order N , such that the spectra of all blocks belong to the neighborhood $O(\nu)$ and each sequence of blocks $(y_1, \dots, y_k) \in Y_k$ has at least $l(N)$ prolongations in Y_{k+1} .

To obtain a block selection of infinite length with the same properties, we will construct finite block selections Y_1, \dots, Y_n in the following manner. Initially, suppose that every Y_k , where $k \leq n$, consists of all of the sequences of blocks $(y_1, \dots, y_k) \in X_{kN}$, such that the spectrum of each block lies in $O(\nu)$. At the first step, we exclude from Y_{n-1} all of the sequences of blocks having less than $l(N)$ prolongations in Y_n and then exclude from Y_n all prolongations of the sequences that were excluded from Y_{n-1} . At the second step, we exclude from Y_{n-2} all of the sequences of blocks having after the first step less than $l(N)$ prolongations in the modified Y_{n-1} , and then, exclude from Y_{n-1} and Y_n all prolongations of the sequences that were excluded from Y_{n-2} . Proceeding further in the same manner, after n steps, we will obtain a block selection Y_1, \dots, Y_n , such that each sequence of blocks from any Y_k has at least $l(N)$ prolongations in Y_{k+1} . Evidently, this selection will be the maximal one among all of the selections of order N having the mentioned “ $l(N)$ -prolongations” property. Therefore, with probability at least p , all of the sets Y_k are nonempty.

For every n , let us construct, as is described above, the maximal block selection $Y_1^{(n)}, \dots, Y_n^{(n)}$ possessing the mentioned “ $l(N)$ -prolongations” property. From the maximality of these selections, it follows that:

$$Y_n^{(n)} \supset Y_n^{(n+1)} \supset Y_n^{(n+2)} \supset \dots$$

Define the sets $Y_n = \bigcap_{k \geq n} Y_n^{(k)}$. Then, with probability at least p , all of them are nonempty and compose an infinite block selection from Theorem 13. Since p may be chosen arbitrarily close to $1 - q^*$, such selections do exist with probability $1 - q^*$. \square

Theorem 13 can be strengthened by taking several measures in place of a unique measure $\nu \in M_1(\Omega)$.

Theorem 14. *Let X_1, X_2, X_3, \dots be an unconditional colored branching process with a finite set of colors Ω and probability of degeneration $q^* < 1$. If a finite collection of measures $\nu_i \in M_1(\Omega)$, where $i = 1, \dots, k$, satisfy the inequalities $\rho(\nu_i, \mu) < 0$, then for any neighborhoods $O(\nu_i) \subset M_1(\Omega)$ and any $\varepsilon > 0$ with probability $1 - q^*$, one can extract from the branching process a block selection Y_1, Y_2, \dots of an order N , such that for every $i = 1, \dots, k$, each sequence of blocks $(y_1, \dots, y_n) \in Y_n$ has at least:*

$$l_i(N) = e^{N(-\rho(\nu_i, \mu) - \varepsilon)}$$

prolongations $(y_1, \dots, y_n, y) \in Y_{n+1}$ with the property $\delta_y \in O(\nu_i)$.

It can be proven in the same manner as the previous one. Here, the event A_n should be understood as the existence of a finite block selection Y_1, \dots, Y_n satisfying the conclusion of Theorem 14, and the event A should be defined by the system of inequalities:

$$\#\{x \in X_N \mid \delta_{x,N} \in O(\nu_i)\} > l_i(N) e^{N\varepsilon/2}, \quad i = 1, \dots, k.$$

We leave the details to the reader.

11. Dimensions of Random Fractals (Lower Bounds)

Now, we continue the investigation of the space of infinite genetic lines:

$$X_\infty = \{x = (x_1, x_2, x_3, \dots) \mid (x_1, \dots, x_n) \in X_n \ \forall n \in \mathbb{N}\},$$

which is generated by an unconditional colored branching process X_1, X_2, \dots with a finite set of colors $\Omega = \{1, \dots, r\}$. It is assumed that there is a measure:

$$\mu = (\mu(1), \dots, \mu(r))$$

on Ω , where $\mu(i)$ denote the expectation of children of color i born by each individual, and X_∞ is equipped with cylindrical θ -metric (41).

Theorem 15. *Let X_1, X_2, X_3, \dots be an unconditional colored branching process with a finite set of colors Ω and probability of degeneration $q^* < 1$. If a measure $\nu \in M_1(\Omega)$ satisfies the condition $d(\nu, \mu, \theta) > 0$, then with probability $1 - q^*$ for any neighborhood $O(\nu)$, we have the lower bound:*

$$\dim_H \{x \in X_\infty \mid \exists N \ \forall n > N \ \delta_{x,n} \in O(\nu)\} \geq d(\nu, \mu, \theta). \quad (52)$$

Proof. Let us fix any number $\alpha < d(\nu, \mu, \theta)$ and $\varepsilon > 0$ so small that:

$$d(\nu, \mu, \theta) = \frac{\rho(\nu, \mu)}{\sum_{i=1}^r \nu(i) \ln \theta(i)} > \frac{\rho(\nu, \mu) + 2\varepsilon}{\sum_{i=1}^r \nu(i) \ln \theta(i) - \varepsilon} > \alpha. \quad (53)$$

Then, we choose a convex neighborhood $O^*(\nu)$, whose closure lies in $O(\nu)$ and such that, for any measure $\delta \in O^*(\nu)$:

$$\sum_{i=1}^r \delta(i) \ln \theta(i) > \sum_{i=1}^r \nu(i) \ln \theta(i) - \varepsilon. \quad (54)$$

By Theorem 13 with probability $1 - q^*$, one can extract from the branching process under consideration a block selection Y_1, Y_2, \dots of an order N , such that any sequence of blocks (y_1, \dots, y_n) from Y_n has at least $l(N)$ prolongations in Y_{n+1} , where:

$$l(N) = e^{N(-\rho(\nu, \mu) - \varepsilon)},$$

and for each block $y_k = (x_{(k-1)N+1}, \dots, x_{kN})$, the corresponding empirical measure δ_{y_k} (spectrum) belongs to $O^*(\nu)$. Let us exclude from this selection a certain part of genetic lines in such a way that each of the remaining sequences of blocks $(y_1, \dots, y_n) \in Y_n$ has exactly $[l(N)]$ prolongations in Y_{n+1} .

Define the random set:

$$Y_\infty = \{y = (y_1, y_2, \dots) \mid (y_1, \dots, y_n) \in Y_n, n = 1, 2, \dots\}.$$

Any sequence $y = (y_1, y_2, \dots) \in Y_\infty$ consists of blocks of length N . Writing down in order the elements of all of these blocks, we obtain from y an infinite genetic line $x = (x_1, x_2, \dots) \in X_\infty$. Denote it as $\pi(y)$. By the definition of Y_∞ , the spectrum of each block y_k belongs to $O^*(\nu)$. For every point $x = \pi(y)$, where $y \in Y_\infty$, the empirical measure $\delta_{x, nN}$ is the arithmetical mean of the empirical measures corresponding to the first n blocks of y , and so, it belongs to $O^*(\nu)$, as well. It follows that:

$$\pi(Y_\infty) \subset \{x \in X_\infty \mid \exists N \forall n > N \delta_{x, nN} \in O(\nu)\}. \quad (55)$$

The family of all cylinders of the form $Z_{nN}(x)$, where $x \in \pi(Y_\infty)$, generates some σ -algebra on $\pi(Y_\infty)$. Define a probability measure P on this σ -algebra, such that:

$$P(Z_{nN}(x)) = [l(N)]^{-n}.$$

Then, for all large enough N , all $x \in \pi(Y_\infty)$ and all natural n :

$$P(Z_{nN}(x)) \leq e^{nN(\rho(\nu, \mu) + 2\varepsilon)}.$$

On the other hand, by Formula (54):

$$\begin{aligned} |Z_{nN}(x)| &= \prod_{t=1}^{nN} \theta(x_t) = \exp \left\{ \sum_{t=1}^{nN} \ln \theta(x_t) \right\} = \exp \left\{ \sum_{i=1}^r nN \delta_{x, nN}(i) \ln \theta(i) \right\} \\ &\geq \exp \left\{ nN \left(\sum_{i=1}^r \nu(i) \ln \theta(i) - \varepsilon \right) \right\}. \end{aligned}$$

It follows from the last two formulas and Formula (53) that:

$$\begin{aligned} |Z_{nN}(x)|^\alpha &\geq \exp \left\{ nN \alpha \left(\sum_{i=1}^r \nu(i) \ln \theta(i) - \varepsilon \right) \right\} \\ &\geq e^{nN(\rho(\nu, \mu) + 2\varepsilon)} \geq P(Z_{nN}(x)). \end{aligned}$$

Now, we are ready to compute the Hausdorff measure of dimension α of the set $\pi(Y_\infty)$. If, while computing the Hausdorff measure, we used coverings of $\pi(Y_\infty)$ not by arbitrary cylinders, but only by cylinders of orders divisible by N , then the last formula would imply that such a measure is not less than $P(\pi(Y_\infty)) = 1$. Any cylinder can be embedded in a cylinder of order divisible by N , such that the difference of their orders is less than N and the ratio of their diameters greater than $\min \theta(i)^N$. Therefore,

$$\text{mes}(\pi(Y_\infty), \alpha) \geq \min \theta(i)^{N\alpha}$$

and hence, $\dim_H \pi(Y_\infty) \geq \alpha$.

The set defined in the right-hand part of Formula (55) contains $\pi(Y_\infty)$. Thus, its dimension is at least α . Recall that we have proven this fact by means of a block selection that exists with probability $1 - q^*$. By the arbitrariness of $\alpha < d(\nu, \mu, \theta)$, this implies the desired bound (52) with the same probability. \square

Theorem 16. *Let s be a root to the “Bowen equation”:*

$$\sum_{i=1}^r \mu(i) \theta(i)^s = 1.$$

If $s \leq 0$, then $X_\infty = \emptyset$ almost surely. Otherwise, if $s > 0$, then X_∞ is nonempty with a positive probability, and with the same probability, its dimension equals s .

We note that a more general version of Theorem 16, in which the similarity coefficients $\theta(i)$ are random, is proven in [4–6,14].

Proof. The expectation of the total number of children born by each individual in the branching process generating the set X_∞ is equal to $m = \mu(1) + \dots + \mu(r)$. If $s \leq 0$, then $m \leq 1$. In this case, by Theorem 8, our branching process degenerates almost surely, and $X_\infty = \emptyset$.

If $s > 0$, then $m > 1$. In this case, by Theorem 8, our branching process does not degenerate with a positive probability, and X_∞ is nonempty. Define a measure $\nu \in M_1(\Omega)$ by means of the equality:

$$\nu(i) = \mu(i) \theta(i)^s, \quad i \in \Omega.$$

Then, evidently, $d(\nu, \mu, \theta) = s$. By the previous theorem, $\dim_H X_\infty \geq s$ with the same probability with which $X_\infty \neq \emptyset$. On the other hand, by Theorem 12, the inverse inequality holds almost surely. \square

For every probability measure $\nu \in M_1(\Omega)$, we define its basin $B(\nu) \subset X_\infty$ as the set of all infinite genetic lines $x = (x_1, x_2, x_3, \dots)$, such that the corresponding sequence of empirical measures $\delta_{x,n}$ converges to ν . What is the dimension of $B(\nu)$? By Theorem 12, it does not exceed the Billingsley–Kullback entropy $d(\nu, \mu, \theta)$ with probability one. On the other hand, the inverse inequality does not follow from the previous results (and, in particular, from Theorem 15). To obtain it, we ought to enhance the machinery of block selections.

Lemma 5. *Let Q_1, \dots, Q_{2^r} be the vertices of a cube in \mathbb{R}^r . Then, there exists a law of choice $i: \mathbb{R}^r \rightarrow \{1, \dots, 2^r\}$, such that if neighborhoods $O(Q_i)$ are small enough, and sequences:*

$$\delta_n \in \mathbb{R}^r \quad \text{and} \quad \Delta_n = \frac{\delta_1 + \dots + \delta_n}{n}$$

satisfy the conditions $\delta_{n+1} \in O(Q_{i(\Delta_n)})$ and $\delta_1 \in O(Q_1) \cup \dots \cup O(Q_{2^r})$, then the sequence Δ_n converges to the center of the cube.

Proof. First, we consider the case $r = 1$, when the cube is a segment. Let, for definiteness, $Q_1 = -1$ and $Q_2 = 1$. Set:

$$i(\Delta) = \begin{cases} 1, & \text{if } \Delta \geq 0, \\ 2, & \text{if } \Delta < 0. \end{cases} \quad (56)$$

Take any neighborhoods $O(Q_1)$ and $O(Q_2)$ with radii at most one. Then, for any sequence δ_n satisfying the conditions $\delta_{n+1} \in O(Q_{i(\Delta_n)})$ and $|\delta_1| < 2$, we have the estimate $|\Delta_n| < 2/n$. It may be easily proven by induction. Thus, in the one-dimensional case, the lemma is proven. To prove it in the multidimensional case, one should choose a coordinate system with the origin at the center of the cube and the axes parallel to its edges and apply the law of choice (56) to each of the coordinates independently. \square

Theorem 17. Let X_1, X_2, X_3, \dots be an unconditional colored branching process with a finite set of colors Ω and probability of degeneration $q^* < 1$. If a measure $\nu \in M_1(\Omega)$ satisfies the condition $d(\nu, \mu, \theta) > 0$, then with probability $1 - q^*$:

$$\dim_H B(\nu) = d(\nu, \mu, \theta).$$

Proof. Fix any number $\alpha < d(\nu, \mu, \theta)$ and $\varepsilon > 0$ so small, that:

$$d(\nu, \mu, \theta) = \frac{\rho(\nu, \mu)}{\sum_{i=1}^r \nu(i) \ln \theta(i)} > \frac{\rho(\nu, \mu) + 3\varepsilon}{\sum_{i=1}^r \nu(i) \ln \theta(i) - \varepsilon} > \alpha.$$

The set $M_1(\Omega)$ is, in fact, a simplex of dimension $r - 1$, where $r = |\Omega|$. Suppose first that ν is an inner point of this simplex (in other words, $\nu(i) > 0$ for all $i \in \Omega$). Take a small convex neighborhood $O(\nu) \subset M_1(\Omega)$, such that for any measure $\delta \in O(\nu)$:

$$\begin{aligned} \rho(\delta, \mu) &< \rho(\nu, \mu) + \varepsilon, \\ \sum_{i=1}^r \delta(i) \ln \theta(i) &> \sum_{i=1}^r \nu(i) \ln \theta(i) - \varepsilon. \end{aligned}$$

Let $Q_1, \dots, Q_{2^{r-1}}$ be the vertices of some cube in $O(\nu)$ with the center at ν . Define for them small neighborhoods $O(Q_i) \subset O(\nu)$ as in Lemma 5. By Theorem 14 with probability $1 - q^*$, one can extract from the colored branching process X_1, X_2, \dots a block selection Y_1, Y_2, \dots of an order N , such that for every $i \leq 2^{r-1}$, each sequence of blocks $(y_1, \dots, y_n) \in Y_n$ has at least:

$$l(N) = e^{N(-\rho(\nu, \mu) - 2\varepsilon)}$$

prolongations $(y_1, \dots, y_n, y) \in Y_{n+1}$ possessing the property $\delta_y \in O(Q_i)$.

We exclude from this block selection a certain part of the genetic lines, so that each of the remaining sequences of blocks $(y_1, \dots, y_n) \in Y_n$ has exactly $[l(N)]$ prolongations $(y_1, \dots, y_n, y) \in Y_{n+1}$, and all of these prolongations satisfy the law of choice from Lemma 5, namely,

$$\delta_y \in O(Q_{i(\Delta_n)}), \quad \text{where} \quad \Delta_n = \frac{\delta_{y_1} + \dots + \delta_{y_n}}{n}.$$

Let $\pi(Y_\infty)$ be the set of all infinite genetic lines $(x_1, x_2, \dots) \in X_\infty$ for which every initial segment of length nN , being partitioned into blocks of length N , turns into an element of Y_n . Then, by Lemma 5, we have the inclusion $\pi(Y_\infty) \subset B(\nu)$.

Reproducing the reasoning from the proof of Theorem 15, one can verify that the dimension of $\pi(Y_\infty)$ is greater than α . Since α can be taken arbitrarily close to $d(\nu, \mu, \theta)$, we obtain the lower bound $\dim_H B(\nu) \geq d(\nu, \mu, \theta)$. The inverse inequality, as was mentioned above, follows from Theorem 12. Thus, in the case of an inner point $\nu \in M_1(\Omega)$, the theorem is proven.

If the measure ν belongs to the boundary of the simplex $M_1(\Omega)$, then one should exclude from Ω all elements i with $\nu(i) = 0$ and consider the set:

$$\Omega' = \{i \in \Omega \mid \nu(i) > 0\}.$$

Let us exclude from the branching process X_1, X_2, \dots all of the genetic lines containing elements of colors not in Ω' and denote by X'_1, X'_2, \dots the resulting branching process (with the set of colors Ω'). The corresponding set of infinite genetic lines X'_∞ is contained in X_∞ . It follows from the definition of the Billingsley–Kullback entropy $d(\nu, \mu, \theta)$ that it is the same for the sets of colors Ω and Ω' . Besides, the measure ν lies in the interior of the simplex $M_1(\Omega')$. Therefore, $\dim_H B(\nu) \cap X'_\infty = d(\nu, \mu, \theta)$ with the same probability as the probability of the event $X'_\infty \neq \emptyset$.

The theorem would be completely proven if the probability of the event $X'_\infty \neq \emptyset$ was equal to $1 - q^*$. However, it may be less than $1 - q^*$. This obstacle can be overcome as follows.

Let $m' = \sum_{i \in \Omega'} \mu(i)$. This is nothing else than the expectation of each individual's number of children in the branching process X'_1, X'_2, \dots . If $m' \leq 1$, then Formula (11) implies the inequality $\rho(\nu, \mu) \geq 0$, which contradicts the condition $d(\nu, \mu, \theta) > 0$ of our theorem. Therefore, $m' > 1$, and respectively, the probability of the event $X'_\infty = \emptyset$ is strictly less than one. Let us denote it by q' .

If the branching process X'_1, X'_2, \dots was generated not by a unique initial element, but by k initial elements, then the probability of the event $X'_\infty = \emptyset$ would be equal to $(q')^k$. Recall that the cardinality of X_n grows exponentially with probability $1 - q^*$. If this is the case, one can first wait for the event $|X_n| \geq k$ and then consider separately $|X_n|$ independent counterparts of the branching process X'_1, X'_2, \dots generated by different elements of X_n . This trick allows one to obtain the bound $\dim_H B(\nu) \geq d(\nu, \mu, \theta)$ with conditional probability at least $1 - (q')^k$ under the condition $|X_n| \rightarrow \infty$. Since k is arbitrary, the above-mentioned conditional probability is, in fact, equal to one, and the complete probability cannot be less than $1 - q^*$. \square

Conflicts of Interest

The author declares no conflict of interest.

References

1. Bakhtin, V.I. Positive processes. *Ergod. Theor. Dyn. Syst.* **2007**, *27*, 639–670.
2. Antonevich, A.B.; Bakhtin, V.I.; Lebedev, A.V. On t-entropy and variational principle for the spectral radius of transfer and weighted shift operators. *Ergod. Theor. Dyn. Syst.* **2011**, *31*, 995–1042.

3. Antonevich, A.B.; Bakhtin, V.I.; Lebedev, A.V. A road to the spectral radius of transfer operators. *Dyn. Syst. Group Actions* **2012**, *567*, 17–51.
4. Falconer, K.J. Random fractals. *Math. Proc. Camb. Philos. Soc.* **1986**, *100*, 559–582.
5. Mauldin, R.D.; Williams, S.C. Random recursive constructions: Asymptotic geometric and topological properties. *Trans. Am. Math. Soc.* **1986**, *295*, 325–346.
6. Graf, S. Statistically self-similar fractals. *Probab. Theory Relat. Fields* **1987**, *74*, 357–392.
7. Falconer, K.; Miao, J. Random subsets of self-affine fractals. *Mathematica* **2010**, *56*, 61–76.
8. Koivusalo, H. Dimension of uniformly random self-similar fractals. *Real Anal. Exch.* **2014**, *39*, 73–90.
9. Järvenpää, E.; Järvenpää, M.; Käenmäki, A.; Koivusalo, H.; Stenflo, Ö.; Suomala, V. Dimensions of random affine code tree fractals. *Ergod. Theor. Dyn. Syst.* **2014**, *34*, 854–875.
10. Billingsley, P. Hausdorff dimension in probability theory. *Ill. J. Math.* **1960**, *4*, 187–209.
11. Billingsley, P. Hausdorff dimension in probability theory II. *Ill. J. Math.* **1961**, *5*, 291–298.
12. Young, L.-S. Dimension, entropy and Lyapunov exponents. *Ergod. Theor. Dyn. Syst.* **1982**, *2*, 109–124.
13. Pesin, Y.B. *Dimension Theory in Dynamical Systems: Contemporary Views and Applications*; Chicago University Press: Chicago, IL, USA, 1998.
14. Falconer, K. *Fractal Geometry: Mathematical Foundations and Applications*, 2nd ed.; Wiley: Chichester, UK, 2003.
15. Deuschel, J.-D.; Stroock D.W. *Large Deviations: Pure and Applied Mathematics*; Academic Press: Boston, MA, USA, 1989; Volume 137.
16. Varadhan, S.R.S. Large deviations. *Ann. Probab.* **2008**, *36*, 397–419.
17. Eggleston, H.G. The fractional dimension of a set defined by decimal properties. *Q. J. Math.* **1949**, *20*, 31–36.
18. Athreya, K.B.; Ney, P.E. *Branching Processes*; Springer: Berlin/Heidelberg, Germany, 1972.
19. Harris, T.E. *The Theory of Branching Processes*; Springer: Berlin/Göttingen/Heidelberg, Germany, 1963.