Алгебра и анализ
Том. 17 (2005), вып. 3

St. Petersburg Math. J.
Vol. 17 (2006), No. 3, Pages 527–551
S 1061-0022(06)00919-8
Article electronically published on March 21, 2006

STATISTICAL ESTIMATION OF MEASURE INVARIANTS

E. A. TIMOFEEV

ABSTRACT. New invariants of measures, called the *β-statentropy*, are described. They are similar to the entropy and the HP-spectrum for dimensions. The β-statentropy admits construction of a statistical estimator calculated by n independent points distributed in accordance with a given measure. The accuracy of this estimator is $\mathcal{O}(n^{-c})$, where c is some constant, and the complexity of calculation is $\mathcal{O}(n^2)$.

It is shown that for an exact dimensional measure the 0-statentropy coincides with the Hausdorff dimension, and for a Markov measure the β-statentropy coincides with the HP-spectrum for dimensions.

An application of the β-statentropy to finding the entropy and dimensional characteristics of dynamical systems is described.

INTRODUCTION

Let Ω be a compact metric space with metric ρ, and let μ be a Borel probability measure on Ω. Suppose we are given n Ω-valued, independent, and identically distributed random variables $\xi_1, \xi_2, \dots, \xi_n$ for which the measure μ is their common distribution. In this paper, we propose a family of consistent statistical estimators for a new measure invariant that we call the *β-statentropy*.

The accuracy of these estimators (the variance and the bias) is equal to $\mathcal{O}(n^{-c})$ with some constant c, and the complexity of calculations is $\mathcal{O}(n^2)$.

It is shown that for an exact dimensional measure the 0-statentropy coincides with the Hausdorff dimension. In particular, for the invariant measure of the ergodic shift in a sequence space, the 0-statentropy is proportional to the entropy of the shift (with coefficient depending on the measure).

We show that, for a Markov measure, the β-statentropy with $\beta \neq 0$ is none other than the HP-spectrum for dimensions (see [Pes97]).

The basic distinction between the estimators we propose and the known estimators for other dimensions (such as the correlation dimensions, the Rényi dimensions, and the entropy; see §2) consists in the power order of accuracy (of the variance and the bias). For most of the estimators known before, their accuracy is only determined experimentally, and in the cases where the accuracy has been established it has turned out to be $\mathcal{O}(1/\ln n)$. It should be noted that the use of estimators for the β-statentropy instead of correlation dimensions lifts the fundamental restriction $d = \mathcal{O}(\log n)$ by Eckmann and Ruelle [ER92] on the dimension d of the dynamical system under consideration.

Computation of dimensional characteristics is an essential step in the analysis of experimental data, because such characteristics allow us to determine

- the dimension of the manifold (the minimal inertial manifold) where the points in question lie;
- the degree of chaoticity of the system.

2000 *Mathematics Subject Classification.* Primary 28A75, 62L20.

Key words and phrases. β-statentropy, statistical estimator, exact dimensional measure, Markov measure.

This explains the existence of a great number of estimators for dimensions, as well as the abundance of papers devoted to calculation of dimensional characteristics for various dynamical systems.

We note that in our setting several (short) trajectories are given, in contrast to the usual approach, which starts with one (long) trajectory. This is of little significance from the viewpoint of applications, but the proofs simplify. At the present time, application of the Monte-Carlo method to the study of dynamical systems is advocated by S. Smale (see, e.g., [CS01]).

The paper is organized as follows.

- In §1 we define the β-statentropy of a measure, describe the setting of the problem, present our statistical estimators, and formulate theorems showing that these estimators are consistent and unbiased. It should be noted that the definition itself of the β-statentropy was obtained as the mean value of these estimators.
- In §2, for comparison with the estimator proposed, we present a brief review of dimensions and their estimators known before.
- In §3 we prove that the β-statentropy is invariant under a bi-Lipschitz change of metrics and measures and establish a series of additional properties.
- In §4 we prove the consistency and unbiasedness of the statistical estimator for the β-statentropy. The accuracy (variance) is $\mathcal{O}(n^{-c})$, where c is some constant.
- In §5, for the Markov measures we show that the β-statentropy coincides with the entropy if $\beta = 0$ and with the HP-spectrum for dimensions if $\beta \neq 0$.
- In §6 we describe an application of the β-statentropy to finding the metric entropy and dimensional characteristics of dynamical systems.

§1. Statement of the problem, notation, and the main result

Let Ω be a compact metric space with metric ρ, and let μ be a Borel probability measure on Ω. For convenience, we assume that $\operatorname{diam}\Omega \leq 1$.

For an Ω-valued random variable ξ with distribution μ, and a real-valued function $f : \Omega \to \mathbb{R}$, we denote by $\mathsf{E}f(\xi)$ the mean value of $f(\xi)$ and by $\mathsf{D}f(\xi)$ the variance.

Let $\phi : \mathbb{R} \to \mathbb{R}$; we denote by

$$\Delta\phi(s) = \phi(s) - \phi(s-1)$$

the first-order finite difference of $\phi(s)$, and by

$$\Delta^{(k)}\phi(s) = \Delta^{(k-1)}\phi(s) - \Delta^{(k-1)}\phi(s-1)$$

the finite difference of kth order.

1.1. Definition of the β-statentropy of a measure. Before giving the definition, we introduce two auxiliary functions and a functional.

Let $B(x, r)$ denote the open ball of radius r centered at x. We introduce the function $r = \nu(t, x)$ inverse to the function $t = \mu(B(x, r))$ (for x fixed) by the relation

$$\nu(t, x) : \lambda(\{t : \nu(t, x) \leq r\}) = \mu(B(x, r)), \tag{1.1}$$

where λ is the Lebesgue measure on $[0, 1]$. Note that the function $t = \mu(B(x, r))$ may have discontinuities and regions of constancy.

Let β be a real parameter. For real-valued functions $u(x)$ on Ω, we introduce a functional $N_\beta(u)$ by putting

$$N_\beta(u) = \gamma^{-1}\left(\int_\Omega \gamma(|u(x)|)\, d\mu(x)\right), \tag{1.2}$$

where

$$\gamma(t) = \begin{cases} t^\beta & \text{if } \beta \neq 0; \\ -\ln t & \text{if } \beta = 0. \end{cases} \tag{1.3}$$

We note that only for the functions $\gamma(t)$ equal to t^β or to $-\ln t$ does the functional $N_\beta(u)$ satisfy the identity

$$N_\beta(Cu) = |C| N_\beta(u) \tag{1.4}$$

(see [HLP48, Chapter 3]). Also, observe that for arbitrary positive monotone functions we have the obvious inequality

$$N_\beta(u) \leq N_\beta(v) \quad \text{if } |u(x)| \leq |v(x)|. \tag{1.5}$$

Let $\chi(\beta, t)$ denote the function

$$\chi(\beta, t) = N_\beta(\nu(t, \cdot)). \tag{1.6}$$

By *the lower and the upper β-statentropy of the measure μ relative to the metric ρ* we shall mean the quantities

$$\underline{\eta}(\beta, \mu, \rho) = \varliminf_{t \to 0+} \frac{\ln t}{\ln \chi(\beta, t)}, \quad \overline{\eta}(\beta, \mu, \rho) = \varlimsup_{t \to 0+} \frac{\ln t}{\ln \chi(\beta, t)}. \tag{1.7}$$

If $\underline{\eta}(\beta, \mu, \rho) = \overline{\eta}(\beta, \mu, \rho) = \eta(\beta, \mu, \rho)$, then $\eta(\beta, \mu, \rho)$ will be called the *β-statentropy*.

Since the function $\chi(\beta, t)$, which characterizes the measure μ and the metric ρ, may fail to be smooth, we introduce yet another auxiliary function, which also characterizes the same objects, but is infinitely differentiable:

$$\phi_\beta(s) = \int_0^1 \gamma(\chi(\beta, t))(1 - t)^{s-1}\, dt. \tag{1.8}$$

Recall that the lower and the upper pointwise dimension of μ at x are defined as follows:

$$\underline{d}_\mu(x) = \varliminf_{u \to 0} \frac{\ln \mu(B(x, u))}{\ln u}, \quad \overline{d}_\mu(x) = \varlimsup_{u \to 0} \frac{\ln \mu(B(x, u))}{\ln u}. \tag{1.9}$$

If $\underline{d}_\mu(x) = \overline{d}_\mu(x) = d_\mu(x)$, then $d_\mu(x)$ is called the *pointwise dimension of μ at x*.

A measure μ is said to be *exact dimensional* if for μ-a.e. $x \in \Omega$ the pointwise dimension $d_\mu(x) = d$ exists and is constant (equal to d). For the exact dimensional measures, the number d coincides with the Hausdorff dimension of μ, which will be denoted by $\dim_H \mu$.

1.2. Statement of the problem. Suppose we have independent random variables $\xi_1, \xi_2, \ldots, \xi_n$ taking values in the metric space Ω and identically distributed with common distribution μ. We want to evaluate the β-statentropy $\eta(\beta, \mu, \rho)$.

For this, we suggest two statistical estimators, which will be described in the next subsection.

1.3. Statistical estimators for the β-statentropy. Let $\xi_1, \xi_2, \ldots, \xi_n$ be given points of the metric space Ω, and let k be an integer.

We construct statistical estimators $\eta_n^{(k)}(\beta, \rho)$ and $\zeta_n^{(k)}(\beta, \rho)$ by the following simple rule:

- we find the auxiliary random variable

$$r_n^{(k)}(\beta) = \frac{1}{n} \sum_{j=1}^{n} \gamma\Big(\min_{i: i \neq j}{}^{(k)} \rho(\xi_i, \xi_j) \Big), \tag{1.10}$$

  where $\min^{(k)}\{x_1, \ldots, x_N\} = x_k$ if $x_1 \leq x_2 \leq \cdots \leq x_N$;

- as the first estimator for the β-statentropy, we take the quantity

$$\zeta_n^{(k)}(\beta,\rho) = -\frac{\ln n}{\ln \gamma^{-1}(r_n^{(k)}(\beta))}; \tag{1.11}$$

- as the second estimator for the β-statentropy, we take

$$\eta_n^{(k)}(\beta,\rho) = \begin{cases} \frac{\beta r_n^{(k)}(\beta)}{k(r_n^{(k+1)}(\beta)-r_n^{(k)}(\beta))} & \text{if } \beta \neq 0; \\ \frac{1}{k(r_n^{(k)}(0)-r_n^{(k+1)}(0))} & \text{if } \beta = 0. \end{cases} \tag{1.12}$$

The integer k, $1 \le k \ll n$, is needed to ensure the existence of the mean values of each term in the sum (1.10) (for $\beta < 0$). Later on, we shall show that if k is sufficiently large, then both estimators do not depend on k and are consistent.

Thus, the construction of $\eta_n^{(k)}(\beta,\rho)$ and $\zeta_n^{(k)}(\beta,\rho)$ employs the parameter β and the metric ρ on Ω.

It should be noted that the measure μ is not used explicitly in the construction, being involved only as the distribution of the random variables $\xi_1, \xi_2, \dots, \xi_n$.

1.4. Unbiasedness and consistency of estimators.

Theorem 1. *Suppose that, for the measure μ and the metric ρ, the β-statentropy $\eta(\beta,\mu,\rho) > 0$ exists, and that the number k satisfies the inequality*

$$k\eta(\beta,\mu,\rho) + \beta > 0. \tag{1.13}$$

Then

$$\lim_{n\to\infty} \mathsf{E}\left[\zeta_n^{(k)}(\beta,\rho)\right]^{-1} = \eta(\beta,\mu,\rho)^{-1}. \tag{1.14}$$

In particular, this theorem shows that the estimator $\zeta_n(\beta,\rho)$ is independent of k provided k is sufficiently large.

Theorem 2. *Suppose that, for the measure μ and the metric ρ, the quantities $\eta(\beta) = \eta(\beta,\mu,\rho) > 0$ and $\eta(2\beta) = \eta(2\beta,\mu,\rho) > 0$ exist, and that the following is true: as $r \to 0$, we have*

$$\exists d > 0 : \int_\Omega \mu(B(x,r))\, d\mu(x) = \mathcal{O}(r^d), \tag{1.15}$$

$$\exists \delta > 0 : \forall m > 0 \int_\Omega \exp\big(-r^{-\delta}\mu(B(x,r))\big)\, d\mu(x) = \mathcal{O}(r^m). \tag{1.16}$$

Also, suppose that

$$\frac{\beta}{\eta(2\beta)} - \frac{\beta}{\eta(\beta)} \ge \frac{b-1}{2} \tag{1.17}$$

for some $b > 0$. If k satisfies

$$([(k+1)/2] - 2)\,\eta(\beta) + \beta > 0, \tag{1.18}$$

then

$$\mathsf{D} r_n^{(k)}(\beta) = \mathcal{O}(n^{-c-2\beta/\eta(\beta)}),$$

where $c < \min\{1, b, d/\delta\}$.

Conditions (1.15) and (1.16) are awkward, but they are of integral nature, and they simplify substantially upon localization. The following statement is easy to verify.

Remark 1. If the condition

$$\exists \overline{d} > \underline{d} > 0, \quad C > 0 : C^{-1} r^{\overline{d}} \le \mu(B(x,r)) \le C r^{\underline{d}} \tag{1.19}$$

is fulfilled for all $r > 0$ and for μ-a.e. $x \in \Omega$, then condition (1.15) is satisfied with $d = \underline{d}$, and condition (1.16) is satisfied with $\delta > \overline{d}$.

So, the estimator $\zeta_n^{(k)}(\beta, \rho)$ is consistent.

The presence of the factor $\ln n$ in (1.11) makes it impossible to estimate the accuracy of $\zeta_n^{(k)}(\beta, \rho)$ better than $\mathcal{O}(1/\ln n)$. For $\eta_n^{(k)}(\beta, \rho)$, the accuracy $\mathcal{O}(n^{-c})$ will be proved under stronger restrictions.

Theorem 3. *Suppose that the β-statentropy $\eta(\beta, \mu, \rho)$ exists, conditions* (1.15)–(1.17) *are satisfied, and there exist constants $\eta > 0$, $a > 0$, and A (depending on β) such that, for some k satisfying* (1.18), *as $s \to +\infty$ we have*

$$|\Delta^{(k)} \phi_\beta(s)| = \begin{cases} \frac{\Gamma(k+\beta/\eta+1)\Gamma(s-k)}{\Gamma(s+\beta/\eta+1)} \left(e^{\beta A} + \mathcal{O}(s^{-a})\right) & \text{if } \beta \ne 0, \\ \frac{k!\Gamma(s-k)}{\Gamma(s+1)} \left(\frac{\psi(s+1)-\psi(k+1)}{\eta} - A + \mathcal{O}(s^{-a})\right) & \text{if } \beta = 0. \end{cases} \tag{1.20}$$

Then $\eta(\beta, \mu, \rho) = \eta$ and

$$\mathsf{E}\left[\eta_n^{(k)}(\beta, \rho)^{-1} - \eta^{-1}\right]^2 = \mathcal{O}(n^{-c}), \tag{1.21}$$

where $c < \min\{1 - b, d/\delta, 2a\}$.

By $\Gamma(x)$ we denote the gamma function, and $\psi(x)$ is the psi function

$$\psi(x) = \frac{d}{dx} \ln \Gamma(x).$$

Thus, the estimator $\eta_n^{(k)}(\rho)$ is asymptotically unbiased and consistent, and its accuracy is of power order of decay.

Condition (1.20), which ensures the power order of convergence for the estimator, is the most complicated restriction imposed on the measure and the metric. The next two remarks show that the factor and the first summand in (1.20) are precisely as they should be.

Remark 2. If

$$\ln \chi(\beta, t) = \frac{\ln t}{\eta} + A + \mathcal{O}(t^a), \quad t \to 0+, \tag{1.22}$$

for some constants $\eta > 0$, $a > 0$, and A, then condition (1.20) is fulfilled with the same constants for any $k > -\beta/\eta$.

Remark 3. If condition (1.20) is fulfilled for some $k > -\beta/\eta$, $k > 0$, then so it is for $k - 1 > -\beta/\eta$.

These two statements will be proved after the proof of Theorem 3 in Subsection 4.4.

§2. Dimensional characteristics

In this section we present descriptions of a series of estimates for dimensions, together with some theoretical results. A detailed survey devoted to earlier theorems can be found in [Cu93].

We pay special attention to the entropy, because this quantity is most essential for describing chaoticity, and it characterizes the dynamical system rather than the measure, unlike the other dimensions do.

2.1. Correlation dimension. Most popular is the correlation dimension, introduced by Grassberger and Procaccia in [GP83].

In a metric space Ω with metric $\rho = \rho(x, y)$ and a Borel probability measure μ, the *correlation integral* is defined as follows:

$$K(r, \mu) = \int_{\Omega} \mu(B(x, r))\, d\mu(x). \tag{2.1}$$

If the limit

$$\lim_{r \to 0} \frac{\ln K(r, \mu)}{\ln r} = D_2 \tag{2.2}$$

exists, then D_2 is called the *correlation dimension.*

In experimental calculations, when points $X_1, \dots, X_n$ are given in Ω (usually, Ω is the Euclidean space), the estimator

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta(X_i) \tag{2.3}$$

is used instead of the unknown measure μ (here $\delta(X)$ is the measure with unit mass concentrated at the point X).

We note that the consistency of the estimator $\frac{\ln K(r, \mu_n)}{\ln r}$ for the correlation dimension is proved only under the sharp scaling condition (see [Cu93, Ser96]), i.e.,

$$K(r, \mu) = \text{const} \cdot r^{D_2}, \quad r \le r_0.$$

In applications, to estimate the limit (2.2) as $r \to 0$, one usually chooses a region on which $\ln K(r, \mu_n)$ is as close to a linear function of $\ln r$ as possible, and then finds the slope on that region. This principle served as the basis for construction of the refined estimators obtained in [Tak85, TL93], which are more efficient. The consistency of those estimators was studied in [Ser96, BBD99].

2.2. The HP-spectrum for dimensions. Applying other averagings in (2.1), namely, taking

$$K_q(r, \mu) = \left(\int_{\Omega} \mu(B(x, r))^{q-1}\, d\mu(x) \right)^{1/(q-1)}, \tag{2.4}$$

we obtain other dimensions. If the limit

$$\lim_{r \to 0} \frac{\ln K_q(r, \mu)}{\ln r} = D_q \tag{2.5}$$

exists, then D_q is called the HP-*spectrum for dimensions*, or *the generalized dimension*, or *the spectrum of correlation dimensions.* The names for various dimensional invariants of a measure are not universally adopted; we follow the terminology of [Pes97].

Recall that, in accordance with [Pes97, p. 182], the HP-*spectrum for dimensions of a measure* μ (Hentschel, Procaccia) is the family of pairs

$$\begin{aligned} \underline{HP}_q(\mu) &= \tfrac{1}{q-1} \varliminf_{r \to 0+} \tfrac{1}{\ln r} \ln \textstyle\int_{\Omega} \mu(B(x, r))^{q-1}\, d\mu(x), \\ \overline{HP}_q(\mu) &= \tfrac{1}{q-1} \varlimsup_{r \to 0+} \tfrac{1}{\ln r} \ln \textstyle\int_{\Omega} \mu(B(x, r))^{q-1}\, d\mu(x). \end{aligned} \tag{2.6}$$

Comparison between the HP-spectrum for dimensions and the β-statentropy shows that for the former the function $\gamma(t) = t^{q-1}$ is used in a similar way. A difference is in averaging over μ: for the β-statentropy we average the inverse function, and then take the inverse quantity.

An essential variation of the HP-spectrum D_q for different values of q tells us about a complicated (multifractal) nature of the system in question.

Moreover, the heuristic arguments presented in [HJKPS86] make it possible to characterize the local singularities of the measure. This approach, called multifractal analysis, allows one (by taking the Legendre transformation) to find also the Hausdorff dimensions of the sets X_a of points x at which μ has a given point dimension $d_\mu(x) = a$ (see, e.g., [Pes97]). However, this approach is justified only for a few systems, and in [BS00] it was shown that the set of all other points (at which $d_\mu(x)$ fails to exist) has the same Hausdorff dimension as the entire set.

2.3. The Rényi dimensions. The definition of the Rényi dimensions resembles that of the entropy and is based on partitions. Let $\mathcal{A} = \{A_1, \dots, A_s\}$ be a partition of the space Ω into cubes with side r; the *Rényi dimensions* are defined as the quantities

$$R_q = \frac{1}{q-1} \lim_{r\to 0} \frac{1}{\log r} \log \sum_{i:\mu(A_i)>0} \mu(A_i)^q \tag{2.7}$$

(if the limit exists). Note that this expression can be made meaningful also for $q = 1$, by assuming that it is equal to minus the formal derivative of the numerator with respect to q. This derivative coincides with the entropy relative to the partition $\mathcal{A}$; for this reason R_1 is called the *information dimension*.

As in the case of the HP-spectrum for dimensions, an estimator for the Rényi dimensions is obtained by replacing μ with μ_n (see (2.3)).

The fact that the Rényi dimensions and the HP-spectrum for dimensions coincide in the case of the diameter regular measures (see [Fed87]) was proved in [Pes97].

The definition of the Rényi dimensions can be extended to the case of negative q if we replace partitions by coverings with intersecting cubes (see [Ri95]).

We note that the replacement of partitions into cubes with arbitrary partitions leads to different results. For example, in [TV98] it was shown that, for $q > 1$, the supremum over all partitions does not depend on $q > 1$ and coincides with the entropy of the dynamical system.

2.4. The nearest neighborhood method. So, the main difficulty in the numerical finding of dimensions is the evaluation of the limit as the auxiliary parameter r tends to 0, while the required values are only given on some interval of values of r. In order to estimate this parameter, we can average the distances to nearest points.

The first statistical estimator based on this idea was suggested in Dobrushin's paper [Do58] for estimating the entropy (per symbol) of sequences of discrete random variables. For $k = 1$, the estimator (1.11) of the present paper can be regarded as a generalization of Dobrushin's estimator. Application of (1.11) with $k = 1$ to finding dimensions was proposed in [BP88]. The consistency of this estimator was proved in [MT99]. In [TA83], the idea of considering the distance to the kth nearest point was put forward, and some heuristic arguments were described for estimation of dimensions by a series of estimators (for different k). In [MT02], the consistency of (1.11) was proved, and it was also shown that (1.11) serves as an estimator for the Rényi dimensions of self-similar fractals.

2.5. Entropy. The entropy we consider is that of a discrete stationary random process. Recall that (see, e.g., [Si73]), for a discrete stationary random process $\xi = (\omega_1, \omega_2, \dots, \omega_t, \dots)$, the entropy (per symbol) is defined as

$$h = \lim_{t\to\infty} -\frac{1}{t} \mathsf{E} \ln p(z_1, \dots, z_t),$$

where $p(z_1, \dots, z_t) = \mathsf{P}\{\omega_1 = z_1, \dots, \omega_t = z_t\}$.

The *entropy of a dynamical system* $T : X \to X$ with invariant measure P is the number

$$h_P(T) = \sup_{\mathcal{A}} h(T, \mathcal{A}),$$

where the supremum is taken over all finite measurable partitions $\mathcal{A}$ of the space X, and $h(T, \mathcal{A})$ denotes the entropy per symbol calculated for the sequence of the symbolic dynamical system that corresponds to the partition $\mathcal{A}$.

In experimental calculations, the evaluation of $h_P(T)$ is restricted usually to that of the quantity $h(T, \mathcal{A})$ for a sufficiently fine partition $\mathcal{A}$. This is justified by the Kolmogorov theorem on the existence of a *generating* partition $\mathcal{A}$ for which $h_P(T) = h(T, \mathcal{A})$. The simplest case of this theorem can be found in [Bo79, Proposition 2.5], where it was shown that, for the expansive homeomorphisms, any partition $\mathcal{A}$ with $\operatorname{diam} \mathcal{A} \le \varepsilon$ is generating, where ε is the separating constant.

We recall that a homeomorphism $T : X \to X$ of a compact metric space (X, d) is said to be *expansive* if a *separating constant* $\varepsilon > 0$ can be found such that $x = y$ whenever $d(T^n(x), T^n(y)) < \varepsilon$ for all $n \in \mathbb{Z}$.

It should be emphasized that no such theorem about generating partitions is valid for the Rényi dimensions considered above. Moreover, the limit value as $r \to 0$ may even fail to exist (see, e.g., [Pes97]). This is due to the absence of the subadditivity property for dimensions, which is fundamental for the proof of the fact that the limit in question exists in the case of the entropy.

For the first time, a statistical estimator for the entropy (per symbol) of sequences of discrete random variables was proposed in Dobrushin's paper [Do58]. This estimator coincides with the mean value of the logarithms of some distances, so that the estimator (1.11) with $k = 1$ can be viewed as a generalization of Dobrushin's estimator. The bias and the variance for Dobrushin's estimator were found in [VM95].

In [Gr89], Grassberger suggested the following estimator for the entropy of a sequence $\xi = (x_1, x_2, \dots, x_t, \dots)$:

$$\hat{H}_n = \frac{n \log n}{\sum_{i=1}^{n} L_i^n},$$

where L_i^n is the length of the smallest prefix of the sequence $x_i, x_{i+1}, \dots$ that is not a prefix of any other sequence x_j, $x_{j+1}, \dots$ with $j \le n$. In [Shi92] it was shown that this estimator is not consistent for general ergodic processes, but is consistent in the case of the Markov processes. In [KS94], consistency was proved for a wider class of processes; a series of consistent modifications of this estimator was suggested in [KASW98].

§3. Properties of the β-statentropy

In this section we prove a series of properties of the β-statentropy $\eta(\beta, \mu, \rho)$, including invariance under a bi-Lipschitz change of measures and metrics.

Proposition 1. *The function $\beta/\overline{\eta}(\beta, \mu, \rho)$ is concave in β.*

Proof. For $\beta \ne 0$ and a given t, consider the function

$$f(\beta) = \ln \gamma(\chi(\beta, t)).$$

We put $f(0) = 0$.

The definition (1.6), (1.2) of the function χ shows that

$$\gamma(\chi(\beta, t)) = \mathsf{E} R^{\beta},$$

where $R = R(\xi)$ is a random variable (equal to the radius of the ball centered at ξ and having measure t, and ξ is a random variable with distribution μ). Therefore, the function $f(\beta)$ is convex (see [Fel84, Subsection 5.8]).

Since $f(\beta) = \beta \ln \chi(\beta, t)$ and $\ln t < 0$, the function $\frac{\beta \ln \chi(\beta,t)}{\ln t}$ is concave. Consequently, so is the function

$$\varliminf_{t\to 0+} \frac{\beta \ln \chi(\beta, t)}{\ln t} = \frac{\beta}{\overline{\eta}(\beta, \mu, \rho)}. \qquad \square$$

Proposition 2. *Let Ω_1 and Ω_2 be compact metric spaces with metrics ρ_1 and ρ_2, let μ_1 be a Borel probability measure on Ω_1, and let $F : \Omega_1 \to \Omega_2$ be a bi-Lipschitz homeomorphism. Suppose that the β-statentropy $\eta(\beta, \mu_1, \rho_1)$ exists for the triple $(\Omega_1, \rho_1, \mu_1)$. Then for the triple $(\Omega_2, \rho_2, \mu_2 = \mu_1 \circ F^{-1})$ the β-statentropy also exists and is equal to the same function $\eta(\beta, \mu_1, \rho_1)$.*

Proof. In the space Ω_2, we have

$$\mu_2(B(F(x), u)) = \mu_1(\{y \in \Omega_1 : \rho_2(F(y), F(x)) \le u\}), \quad u \ge 0, x \in \Omega_1.$$

Since $F(x)$ is a bi-Lipschitz homeomorphism of compact spaces, there is a constant $C > 1$ such that

$$C^{-1}\rho_1(x, y) \le \rho_2(F(x), F(y)) \le C\rho_1(x, y), \quad x, y \in \Omega_1.$$

Therefore,

$$\mu_1(B(x, C^{-1}u)) \le \mu_2(B(F(x), u)) \le \mu_1(B(x, Cu)).$$

For the inverse functions, we have

$$C^{-1}\nu_1(t, x) \le \nu_2(t, F(x)) \le C\nu_1(t, x).$$

Plugging these estimates in (1.6) and using (1.2), (1.4), and (1.5), we obtain

$$C^{-1}N_\beta(\nu_1(t, \cdot)) \le \chi_2(\beta, t) \le CN_\beta(\nu_1(t, \cdot)).$$

Recalling (1.7), we see that

$$\lim_{t\to 0+} \frac{\ln t}{\ln \chi_2(\beta, t)} = \lim_{t\to 0+} \frac{\ln t}{\ln \chi_1(\beta, t)} = \eta(\beta, \mu_1, \rho_1).$$

Consequently, the β-statentropy for the triple $(\Omega_2, \rho_2, \mu_2 = \mu_1 \circ F^{-1})$ exists and is equal to $\eta(\beta, \mu_1, \rho_1)$. $\square$

The statement proved above not only carries the β-statentropy over to other spaces, but also shows its invariance under a smooth change of the metric.

Corollary 1. *Let ρ_1 and ρ_2 be two metrics on Ω such that*

$$C^{-1}\rho_1(x, y) \le \rho_2(x, y) \le C\rho_1(x, y), \quad x, y \in \Omega,$$

for some constant C. If the β-statentropy exists for the triple (Ω, ρ, μ), then for the triple $\omega(\beta, \rho, \tilde{\mu})$ it also exists and is equal to the same function $\eta(\beta, \mu, \rho)$.

The next proposition shows that, for $\beta \ne 0$, the β-statentropy is invariant under a smooth change of the measure.

Proposition 3. *Let μ and $\tilde{\mu}$ be two Borel probability measures on Ω such that*

$$C^{-1}\mu(S) \le \tilde{\mu}(S) \le C\mu(S), \quad S \subset \Omega,$$

for some constant $C > 1$, and let $\beta \ne 0$. If the β-statentropy $\eta(\beta, \mu, \rho)$ exists for the triple (Ω, ρ, μ), then for the triple $(\Omega, \rho, \tilde{\mu})$ it also exists and is equal to the same function $\eta(\beta, \mu, \rho)$.

Proof. All auxiliary functions corresponding to the measure $\tilde{\mu}$ will be denoted by adding a tilde.

By the condition on $\tilde{\mu}$, for the integral

$$\widetilde{N}_\beta(\tilde{\nu}(t,\cdot)) = \left(\int_\Omega \tilde{\nu}(t,x)^\beta \, d\tilde{\mu}(x)\right)^{1/\beta}$$

we have

$$C^{-1/|\beta|} N_\beta(\tilde{\nu}(t,\cdot)) \le \widetilde{N}_\beta(\tilde{\nu}(t,\cdot)) \le C^{1/|\beta|} N_\beta(\tilde{\nu}(t,\cdot)).$$

The measure of a ball satisfies the inequality

$$C^{-1}\mu(B(x,u)) \le \tilde{\mu}(B(x,u)) \le C\mu(B(x,u)).$$

Thus, for the inverse functions we have

$$\nu(C^{-1}t,x) \le \tilde{\nu}(t,x) \le \nu(Ct,x).$$

Substituting this in (1.6) and using (1.2), (1.4), and (1.5), we obtain

$$C^{-1}\chi(\beta, C^{-1}t) \le \tilde{\chi}(\beta,t) \le C\chi(\beta,Ct).$$

Thus, by (1.7),

$$\lim_{t\to 0+} \frac{\ln t}{\ln \tilde{\chi}(\beta,t)} = \lim_{t\to 0+} \frac{\ln t}{\ln \chi(\beta,t)} = \eta(\beta,\mu,\rho).$$

Consequently, for the triple $(\Omega,\rho,\tilde{\mu})$, the β-statentropy exists and is equal to the same function $\eta(\beta,\mu,\rho)$. □

Observe that Proposition 3 reveals a distinction between the β-statentropy with $\beta \ne 0$, which is similar to the HP-spectrum for dimensions and survives under a smooth change of the measure, and the β-statentropy with $\beta = 0$, similar to the metric entropy, which is affected by such a change.

Proposition 4. *If condition* (1.19) *is fulfilled, then*

$$\underline{d} \le \underline{\eta}(\beta,\mu,\rho) \le \overline{\eta}(\beta,\mu,\rho) \le \overline{d}.$$

Proof. Condition (1.19) implies that the function inverse to $t = \mu(B(x,u))$ satisfies the estimate

$$\left(\frac{t}{C}\right)^{1/\underline{d}} \le \nu(t,x) \le (Ct)^{1/\overline{d}} \quad \text{for } \mu\text{-a.e. } x \in \Omega.$$

We substitute this in (1.6) and apply (1.5) and the Lebesgue dominated convergence theorem (see [Fed87, Subsection 2.4.9]) to obtain

$$\left(\frac{t}{C}\right)^{1/\underline{d}} \le \chi(\beta,t) \le (Ct)^{1/\overline{d}}.$$

Recalling (1.7), we arrive at the required inequalities. □

The next statement shows that the 0-statentropy $\eta(0,\mu,\rho)$ possesses a stronger property.

Proposition 5. *We have*

$$\left(\int_\Omega \frac{d\mu(x)}{\underline{d}(x)}\right)^{-1} \le \underline{\eta}(0,\mu,\rho) \le \overline{\eta}(0,\mu,\rho) \le \left(\int_\Omega \frac{d\mu(x)}{\overline{d}(x)}\right)^{-1}. \tag{3.1}$$

Proof. By (1.7),

$$\underline{\eta}(0,\mu,\rho)^{-1} = \varlimsup_{t\to 0+} \int_\Omega \frac{\ln(\nu(t,x))}{\ln t}\, d\mu(x).$$

Applying the Fatou lemma (see [Fed87, Subsection 2.4.9]), we get

$$\underline{\eta}(0,\mu,\rho)^{-1} \le \int_\Omega \varlimsup_{t\to 0+} \frac{\ln(\nu(t,x))}{\ln t}\, d\mu(x). \tag{3.2}$$

Definition (1.9) of the pointwise dimensions implies that

$$\varlimsup_{t\to 0+} \frac{\ln(\nu(t,x))}{\ln t} = \left(\varliminf_{t\to 0+} \frac{\ln t}{\ln(\nu(t,x))}\right)^{-1} = \left(\varliminf_{r\to 0+} \frac{\ln \mu(B(x,r))}{\ln r}\right)^{-1} = \underline{d}(x)^{-1}.$$

Plugging this in (3.2), we obtain the first inequality in (3.1).

The upper estimate is proved in a similar way. □

Now, applying the Young theorem (see, e.g., [Pes97, p. 42]), we arrive at the following statement.

Corollary 2. *If μ is an exact dimensional measure, then the* 0*-statentropy $\eta(0,\mu,\rho)$ exists and is equal to* $\dim_H \mu$.

We note that the Markov measures (to be considered below) are exact dimensional, and the values of $\eta(\beta,\mu,\rho)$ for various β fill an interval.

§4. Properties of statistical estimates

In this section we prove Theorems 1–3. Let $\xi_1,\xi_2,\dots,\xi_n$ be independent random variables identically distributed with distribution μ and taking values in Ω.

4.1. Distribution of the distances to the kth nearest point. Let $R^{(k)}_{n,i}$ denote the random variable defined by

$$R^{(k)}_{n,i} = \min_{j:j\ne i}{}^{(k)} \rho(\xi_j,\xi_i), \quad i=1,2,\dots,n. \tag{4.1}$$

We introduce two auxiliary conditional probabilities:

$$Q_{n,k}(u,x) = \mathsf{P}\{R^{(k)}_{n,i} = u \mid \xi_i = x, \rho(\xi_j,\xi_i) = u\} \tag{4.2}$$

is the conditional probability of the event that the points $\xi_1,\xi_2,\dots,\xi_n$ are located so that precisely $k-1$ points are in the ball $B(y,u)$, $\xi_i = y$, and a point lies on the boundary of $B(y,u)$; and

(4.3)

$$Q^{(2)}_{n,k}(u,v,x,y) = \mathsf{P}\{R^{(k)}_{n,i} = u, R^{(k)}_{n,j} = v \mid \xi_i = x, \xi_j = y, \rho(\xi_l,\xi_i) = u, \rho(\xi_m,\xi_j) = v\}$$

is the conditional probability of the event that the points $\xi_1,\xi_2,\dots,\xi_n$ are located so that each of the balls $B(x,u)$, $B(y,v)$ contains precisely $k-1$ points, $\xi_i = x$, $\xi_j = y$, a point lies on the boundary of $B(x,u)$, and a point lies on the boundary of $B(y,v)$.

The probability $Q_{n,k}(u,y)$ is calculated easily:

$$Q_{n,k}(u,x) = \binom{n-2}{k-1} \mu(B(x,u))^{k-1}(1-\mu(B(x,u)))^{n-k-1}. \tag{4.4}$$

The probability $Q^{(2)}_{n,k}(u,v,x,y)$ admits no explicit formula, but some estimates are available.

Lemma 1. *We have*

$$Q^{(2)}_{n,k}(u,v,x,y) \le Q_{m,k_1}(u,x)Q_{m,k_2}(v,y), \tag{4.5}$$

where $m = [n/2]-k-1$, $k_1 = \max\{[k/2]-1,1\}$, *and* $k_2 = \max\{[(k+1)/2]-2,1\}$. *Moreover,*

$$Q^{(2)}_{n,k}(u,v,x,y) \le (1+C_0n^{-1})Q_{n-k-1,k}(u,y)Q_{n-k-1,k}(v,y) \tag{4.6}$$

whenever $B(x,u)\cap B(y,v)=\varnothing$, *where* C_0 *is a constant depending only on* k.

Proof. In the definition (4.3) of the conditional probability $Q^{(2)}_{n,k}(u,v,x,y)$, four points are fixed: $\xi_i = x$, $\xi_j = y$, and two points ξ_l and ξ_m lying on the boundaries of the balls $B(x,u)$ and $B(y,v)$. We put

$$M = \{1,2,\dots,n\}\setminus\{i,j,l,m\}.$$

Then each of the balls $B(x,u)$, $B(y,v)$ contains at least $k-3$ points of M, and there are at least $n-2k-2$ points of M that lie outside these balls.

To prove (4.5), we argue as follows. We extract two subsets from M, the first of which contains $[k/2]-1$ points and lies inside $B(x,u)$, and the second contains $[n/2]-k-1$ points and lies outside the balls $B(x,u)$ and $B(y,v)$. Then the remaining part of M also contains two subsets, the first of which has at least $[(k+1)/2]-2$ points and lies inside $B(y,v)$, and the second has at least $[n/2]-k-1$ points and lies outside the balls $B(x,u)$ and $B(y,v)$.

We prove inequality (4.6). If the balls $B(x,u)$ and $B(y,v)$ are disjoint, the probability $Q^{(2)}_{n,k}(u,v,x,y)$ is evaluated easily:

$$Q^{(2)}_{n,k}(u,v,x,y) = \binom{n-4}{k-1}\binom{n-k-3}{k-1}\mu(B(x,u))^{k-1}\mu(B(y,v))^{k-1} \times (1-\mu(B(x,u))-\mu(B(y,v)))^{n-2k-2}.$$

Applying the inequality $1-s-t \le (1-s)(1-t)$ $Q_{n-k-1,k}(u,x)$, and using (4.4), we obtain

$$Q^{(2)}_{n,k}(u,v,x,y) \le \frac{(n-4)!(n-2k-2)!}{(n-k-3)!^2}Q_{n-k-1,k}(u,x)Q_{n-k-1,k}(v,y).$$

Thus, (4.6) is valid with a constant C_0 such that

$$\frac{(n-4)!(n-2k-2)!}{(n-k-3)!^2} \le 1+\frac{C_0}{n}. \qquad \square$$

4.2. Proof of Theorem 1. We find the conditional mean value of $\gamma(R^{(k)}_{n,i})$ under the condition $\xi_i = x$.

By (4.2), we have

$$\mathsf{E}\left[\gamma(R^{(k)}_{n,i})|\xi_i = x\right] = \int_0^1 (n-1)\gamma(u)Q_{n,k}(u,x)\,d_u\mu(B(x,u)).$$

The definition (1.1) of the function $\nu(t,x)$ shows that the change $t=\mu(B(u,x))$ reduces the above integral to the integral

$$\mathsf{E}\left[\gamma(R^{(k)}_{n,i})|\xi_i = x\right] = \int_0^1 (n-1)\gamma(\nu(t,x))Q_{n,k}(\nu(t,x),x)\,dt.$$

Recalling the expression (4.4) for $Q_{n,k}(u,y)$, we obtain

$$\mathsf{E}\left[\gamma(R^{(k)}_{n,i})|\xi_i = x\right] = (n-1)\binom{n-2}{k-1}\int_0^1 \gamma(\nu(t,x))t^{k-1}(1-t)^{n-k-1}\,dt. \tag{4.7}$$

Now we integrate (4.7) with respect to μ and substitute $\chi(\beta,t)$ (see (1.6)). This yields

$$\mathsf{E}r_n^{(k)}(\beta) = \mathsf{E}\gamma(R_{n,i}^{(k)}) = (n-1)\binom{n-2}{k-1}\int_0^1 \gamma(\chi(\beta,t))t^{k-1}(1-t)^{n-k-1}\,dt. \tag{4.8}$$

Now, we prove the existence of the limit in (1.14).

Since the β-statentropy $\eta = \eta(\beta,\mu,\rho)$ exists by assumption, for any $\varepsilon > 0$ there exists t_0 such that for $t < t_0$ we have

$$t^{1/\eta+\varepsilon} \le \chi(\beta,t) \le t^{1/\eta-\varepsilon}.$$

For $t \ge t_0$ we use the estimate $(1-t)^{n-k-1} \le (1-t_0)^{n-k-1}$; for n sufficiently large this yields the inequality

$$\begin{aligned}(n-1)&\binom{n-2}{k-1}\int_{t_0}^1 \gamma(\chi(\beta,t))t^{k-1}(1-t)^{n-k-1}\,dt \\ &< n^k(1-t_0)^{n-k-1}\int_{t_0}^1 t^{k-1}\gamma(\chi(\beta,t))\,dt < C_0\theta^n,\end{aligned} \tag{4.9}$$

where $1-t_0 < \theta < 1$, and C_0 is a constant independent of n.

Plugging these estimates into (4.8) and calculating the corresponding beta functions, we see that, for $\beta \ne 0$,

$$\frac{\Gamma(k+\beta/\eta+\varepsilon|\beta|)\Gamma(n)}{\Gamma(n+\beta/\eta+\varepsilon|\beta|)\Gamma(k)} - C_0\theta^n \le \mathsf{E}r_n^{(k)}(\beta) \le \frac{\Gamma(k+\beta/\eta-\varepsilon|\beta|)\Gamma(n)}{\Gamma(n+\beta/\eta-\varepsilon|\beta|)\Gamma(k)} + C_0\theta^n.$$

Inequality (1.13) shows that all the integrals involved are finite provided ε is sufficiently small.

Invoking the asymptotics of the gamma function, we conclude that

$$C_1 n^{-1/\eta-\varepsilon} \le \gamma^{-1}(\mathsf{E}r_n^{(k)}(\beta)) \le C_2 n^{-1/\eta+\varepsilon}$$

with some constants C_1 and C_2 independent of n. This implies (1.14) for $\beta \ne 0$.

If $\beta = 0$, then after the substitution of estimates in (4.8), the integrals can also be calculated (see [GR71, 4.253.1]), and

$$\frac{1}{\eta+\varepsilon}\sum_{i=k}^{n-1}\frac{1}{i} \le \mathsf{E}r_n^k(0) \le \frac{1}{\eta-\varepsilon}\sum_{i=k}^{n-1}\frac{1}{i}.$$

Since $\sum_{i=k}^{n-1}\frac{1}{i} \sim \ln n$, we obtain (1.14) for $\beta = 0$.

Theorem 1 is proved. □

The estimates of $\mathsf{E}r_n^{(k)}(\beta) = \mathsf{E}\gamma(R_{n,i}^{(k)})$ obtained in the proof of Theorem 1 can be written in a unified form.

Corollary 3. *Under the conditions of Theorem* 1, *we have*

$$\mathsf{E}\gamma(R_{n,i}^{(k)}) = \mathcal{O}(n^{\varepsilon-\beta/\eta}) \tag{4.10}$$

for every $\varepsilon > 0$.

4.3. Proof of Theorem 2. We are going to estimate the covariance and the variance of the random variables $\gamma(R_{n,i}^{(k)})$.

Lemma 2. *Under the conditions of Theorem* 2, *we have*

$$\operatorname{Cov}\left[\gamma(R_{n,i}^{(k)}), \gamma(R_{n,j}^{(k)})\right] = \mathcal{O}(n^{-c-2\beta/\eta}), \tag{4.11}$$

where $c < \min\{1, d/\delta\}$.

Proof. We estimate $\mathsf{E}\left[\gamma(R_{n,i}^{(k)})\gamma(R_{n,j}^{(k)})\right]$.

Let $a > 0$ be a parameter, and let S denote the event that $\rho(\xi_j, \xi_i) < 2n^{-a}$. Then

$$\mathsf{P}\{S\} = \int_\Omega \mu(B(x, 2n^{-a}))\, d\mu(x).$$

By inequalities (1.15), we obtain

$$\mathsf{P}\{S\} = \mathcal{O}(n^{-ad}).$$

The total probability formula yields

$$\mathsf{E}\left[\gamma(R_{n,i}^{(k)})\gamma(R_{n,j}^{(k)})\right] = \mathsf{E}\left[\gamma(R_{n,i}^{(k)})\gamma(R_{n,j}^{(k)})|S\right]\mathsf{P}\{S\} + \mathsf{E}\left[\gamma(R_{n,i}^{(k)})\gamma(R_{n,j}^{(k)})|\overline{S}\right](1 - \mathsf{P}\{S\}). \tag{4.12}$$

We estimate both terms on the right in (4.12). To estimate the conditional probability $\mathsf{E}\big[\gamma(R_{n,i}^{(k)})\gamma(R_{n,j}^{(k)})|S\big]$, we apply Lemma 1. Since the proof of inequality (4.5) is independent of the mutual location of the points x and y, this proof remains valid under the assumption that the event S happens.

Multiplying (4.5) by

$$(n-1)(n-2)\gamma(u)\gamma(v)d_u\mu(B(u,x))d_v\mu(B(v,y))d\mu(x)d\mu(y)$$

and integrating, we obtain

$$\mathsf{E}\left[\gamma(R_{n,i}^{(k)})\gamma(R_{n,j}^{(k)})|S\right] \le \frac{(n-1)(n-2)}{([n/2]-1)([n/2]-1)}\mathsf{E}\gamma\left(R_{[n/2]-k-1,i}^{([k/2]-1)}\right)\mathsf{E}\gamma\left(R_{[n/2]-k-1,i}^{([(k+1)/2]-2)}\right). \tag{4.13}$$

We take $\varepsilon > 0$ and apply (4.10). All the conditions of Theorem 1 are satisfied: $\eta(\beta,\mu,\rho)$ exists by assumption, and (1.13) is implied by (1.18) because $[(k+1)/2]-2 \le [k/2]-1$.

Since the right-hand side of (4.10) does not depend on k provided (1.18) is fulfilled, and does not change if we divide n by a constant, we may assume that inequalities (4.10) are true (with one and the same ε) for all $R_{N,i}^{(l)}$ with $n/2-k-1 \le N \le n$ and all l with $(k+1)/2-2 \le l \le k$. Using (4.10), from (4.13) we deduce that

$$\mathsf{E}\left[\gamma(R_{n,i}^{(k)})\gamma(R_{n,j}^{(k)})|S\right] = \mathcal{O}(n^{2\varepsilon-2\beta/\eta}).$$

So, the first term in (4.12) satisfies

$$\mathsf{E}\left[\gamma(R_{n,i}^{(k)})\gamma(R_{n,j}^{(k)})|S\right]\mathsf{P}\{S\} = \mathcal{O}(n^{-ad+2\varepsilon-2\beta/\eta}). \tag{4.14}$$

We estimate the second term in (4.12). Consider the case where $R_{n,i}^{(k)}, R_{n,j}^{(k)} < n^{-a}$. Multiplying (4.6) by

$$(n-1)(n-2)\gamma(u)\gamma(v)d_u\mu(B(u,x))d_v\mu(B(v,y))d\mu(x)d\mu(y)$$

and integrating, we get

$$\mathsf{E}\left[\gamma(R_{n,i}^{(k)})\gamma(R_{n,j}^{(k)})|\overline{S}\right] = (1+\mathcal{O}(n^{-1}))\mathsf{E}\gamma(R_{n-k-1,i}^{(k)})\mathsf{E}\gamma(R_{n-k-1,j}^{(k)}). \tag{4.15}$$

Now, suppose that $R_{n,i}^{(k)} > n^{-a}$ or $R_{n,j}^{(k)} > n^{-a}$. Inequalities (4.5) show that in this case it suffices to estimate the quantity

$$T_n(r) = \int_\Omega\int_r^1 (n-1)\gamma(u)\binom{n-2}{k-1} \times \mu(B(x,u))^{k-1}(1-\mu(B(x,u)))^{n-k-1}\, d_u\mu(B(x,u))\, d\mu(x).$$

Clearly,

$$T_n(r) \le n^k \max\{\gamma(r), 1\} \int_\Omega \exp\left(-(n-k-1)\mu(B(x,r))\right) d\mu(x).$$

Since $\gamma(r) \le r^{-|\beta|-1}$, for $r = n^{-a}$ we obtain

$$T_n(n^{-a}) \le n^{k+a|\beta|+a} \int_\Omega \exp\left(-(n-k-1)\mu(B(x,n^{-a}))\right) d\mu(x).$$

Applying condition (1.16) with $m = (k+1+a|\beta|+a)/a$ for $a < 1/\delta$, we see that

$$T_n(n^{-a}) = \mathcal{O}(n^{-1}).$$

We plug this estimate and (4.14), (4.15) into (4.12), obtaining

$$\begin{aligned} &\operatorname{Cov}\left[\gamma(R_{n,i}^{(k)}), \gamma(R_{n,j}^{(k)})\right] \\ &\quad = \mathcal{O}(n^{-ad+2\varepsilon-2\beta/\eta}) + \mathsf{E}\gamma(R_{n-k-1,i}^{(k)})\mathsf{E}\gamma(R_{n-k-1,j}^{(k)}) - \mathsf{E}\gamma(R_{n,i}^{(k)})\mathsf{E}\gamma(R_{n,j}^{(k)}). \end{aligned} \tag{4.16}$$

Let

$$S_n = \mathsf{E}\gamma(R_{n-k-1,j}^{(k)}) - \mathsf{E}\gamma(R_{n,i}^{(k)}).$$

By (4.8), we have

$$S_n = (n-k-2)\binom{n-k-3}{k-1} \int_0^1 \gamma(\chi(\beta,t))t^{k-1}(1-t)^{n-2k-2} f(t)\, dt, \tag{4.17}$$

where

$$f(t) = 1 - \frac{(n-1)!(n-2k-2)!}{(n-k-1)!(n-k-2)!}(1-t)^{k+1} \le 1 - (1-t)^{k+1} \le (k+1)t.$$

Inequality (4.9), obtained in the proof of Theorem 1, shows that for estimating the integral in (4.17) it suffices to consider a neighborhood of 0.

The existence of the β-entropy $\eta = \eta(\beta, \mu, \rho)$ implies that for any $\varepsilon > 0$ there exists t_0 such that for $t < t_0$ we have

$$t^{1/\eta+\varepsilon} \le \chi(\beta, t) \le t^{1/\eta-\varepsilon}.$$

We take the same ε as in (4.14). Using the above estimate in (4.17) and evaluating the integrals, we obtain

$$S_n = \mathcal{O}(n^{-1-\beta/\eta+|\beta|\varepsilon}).$$

Returning to (4.16), we see that, for any $\varepsilon > 0$ and any $a < 1/\delta$,

$$\operatorname{Cov}\left[\gamma(R_{n,i}^{(k)}), \gamma(R_{n,j}^{(k)})\right] = \mathcal{O}(n^{-ad+2\varepsilon-2\beta/\eta}) + \mathcal{O}(n^{-1-2\beta/\eta+|\beta|\varepsilon+\varepsilon}).$$

Consequently, (4.11) is valid with an arbitrary constant $c < \min\{1, d/\delta\}$. □

We pass to the proof of Theorem 2.

For the variance of the random variable $r_n^{(k)}(\beta)$, we have

$$\mathsf{D}(r_n^{(k)}(\beta)) = \frac{1}{n}\mathsf{D}\gamma(R_{n,i}^{(k)}) + \frac{n-1}{n}\operatorname{Cov}\left(\gamma(R_{n,i}^{(k)}), \gamma(R_{n,j}^{(k)})\right). \tag{4.18}$$

The quantity $\mathsf{D}\gamma(R_{n,i}^{(k)})$ is estimated as follows:

$$\mathsf{D}\gamma(R_{n,i}^{(k)}) \le \mathsf{E}\gamma^2(R_{n,i}^{(k)}).$$

We fix $\varepsilon > 0$ and apply inequalities (4.10) with the parameter 2β. Since all the conditions of Theorem 1 are satisfied, we can apply Corollary 3:

$$\mathsf{E}\gamma^2(R_{n,i}^{(k)}) = \mathcal{O}(n^{\varepsilon-2\beta/\eta(2\beta)}).$$

Using (1.17), for $\beta \neq 0$ we obtain

$$\mathsf{D}\gamma(R_{n,i}^{(k)}) = \mathcal{O}(n^{-b+1+\varepsilon-2\beta/\eta(\beta)}). \tag{4.19}$$

Obviously, this estimate remains valid for $\beta = 0$, because $b = 1$ in this case.

Substituting (4.19) and (4.11) into (4.18), we get

$$\mathsf{D}r_n^{(k)}(\beta) = (\mathcal{O}(n^{-b+\varepsilon}) + \mathcal{O}(n^{-c}))n^{-2\beta/\eta(\beta)},$$

where $c < \min\{1, d/\delta\}$.

Since $\varepsilon > 0$ is arbitrary, for $c < \min\{1, b, d/\delta\}$ we have

$$\mathsf{D}r_n^{(k)}(\beta) = \mathcal{O}(n^{-c-2\beta/\eta(\beta)}).$$

Theorem 2 is proved. □

4.4. Proof of Theorem 3. We shall show that the β-statentropy $\eta(\beta, \mu, \rho)$ coincides with the parameter η occurring in (1.20). It is easily seen that

$$\Delta^{(m)}\phi_\beta(s) = (-1)^m \int_0^1 \gamma(\chi(t))t^m(1-t)^{s-m-1}\,dt. \tag{4.20}$$

Plugging this into (4.20), we get the identity

$$\mathsf{E}r_n^{(m)}(\beta) = (-1)^{m-1}\frac{\Gamma(n)}{\Gamma(m)\Gamma(n-m)}\Delta^{(m-1)}\phi_\beta(n-1). \tag{4.21}$$

Condition (1.20) and Remark 3 imply that if $m \le k+1$ and $m + \beta/\eta > 0$, then

$$\mathsf{E}r_n^{(m)}(\beta) = \begin{cases} \frac{\Gamma(k+\beta/\eta)\Gamma(n)}{\Gamma(n+\beta/\eta)\Gamma(k)}\left(e^{\beta A} + \mathcal{O}(n^{-a})\right) & \text{if } \beta \neq 0; \\ \frac{\psi(n)-\psi(m)}{\eta} + A + \mathcal{O}(n^{-a}) & \text{if } \beta = 0. \end{cases} \tag{4.22}$$

Consequently,

$$\ln(\gamma^{-1}(\mathsf{E}r_n^{(m)}(\beta))) = -\frac{\ln n}{\eta}(1 + o(1)).$$

Applying Theorem 1, we conclude that $\eta(\mu, \rho) = \eta$.

Now, we prove (1.21). We have

$$\mathsf{E}\left[\eta_n^{(k)}(\beta,\rho)^{-1} - \eta^{-1}\right]^2 = \mathsf{D}\left[\eta_n^{(k)}(\beta,\rho)^{-1}\right] + \left(\mathsf{E}\left[\eta_n^{(k)}(\beta,\rho)^{-1}\right] - \eta^{-1}\right)^2.$$

Theorem 2 shows that

$$\mathsf{E}\left[\eta_n^{(k)}(\beta,\rho)^{-1} - \eta^{-1}\right]^2 = \mathcal{O}(n^{-c}) + \left(\mathsf{E}\left[\eta_n^{(k)}(\beta,\rho)^{-1}\right] - \eta^{-1}\right)^2, \tag{4.23}$$

where $c < \min\{1-b, d/\delta\}$.

We pass to estimation of the bias (the second term in (4.23)). Suppose $\beta \neq 0$. We introduce the auxiliary random variable

$$\tilde{r}_n^{(k)} = r_n^{(k)}(\beta)\frac{\Gamma(n+\beta/\eta)}{\Gamma(n)}. \tag{4.24}$$

Then, by (4.22),

$$\mathsf{E}\tilde{r}_n^{(k)} = e^{\beta A}\frac{\Gamma(k+\beta/\eta)}{\Gamma(k)} + \mathcal{O}(n^{-\alpha}). \tag{4.25}$$

Substituting $\tilde{r}_n^{(k)}$ in (1.12), we obtain

$$\mathsf{E}\left[\eta_n^{(k)}(\beta,\rho)^{-1} - \eta^{-1}\right] = \mathsf{E}\left[\frac{k\tilde{r}_n^{(k+1)} - (k+\beta/\eta)\tilde{r}_n^{(k)}}{\beta\tilde{r}_n^{(k)}}\right]. \tag{4.26}$$

To estimate the right-hand side, we use the Schwarz inequality:

$$\begin{aligned} &\mathsf{E}\left[\frac{k\tilde r_n^{(k+1)} - (k+\beta/\eta)\tilde r_n^{(k)}}{\beta \tilde r_n^{(k)}}\right] \\ &\quad \le \frac{1}{|\beta|}\left(\mathsf{E}[\tilde r_n^{(k)}]^{-2}\right)^{1/2}\left(\mathsf{E}\left[k\tilde r_n^{(k+1)} - (k+\beta/\eta)\tilde r_n^{(k)}\right]^2\right)^{1/2}. \end{aligned}$$

Relation (4.25) shows that $\mathsf{E}[\tilde r_n^{(k)}]^{-2} = \mathcal{O}(1)$, so that we only need to estimate the second factor. We have

$$\begin{aligned} &\mathsf{E}\left[k\tilde r_n^{(k+1)} - (k+\beta/\eta)\tilde r_n^{(k)}\right]^2 \\ &\quad = \mathsf{D}\left[k\tilde r_n^{(k+1)} - (k+\beta/\eta)\tilde r_n^{(k)}\right] + \left(k\mathsf{E}\tilde r_n^{(k+1)} - (k+\beta/\eta)\mathsf{E}\tilde r_n^{(k)}\right)^2. \end{aligned}$$

Applying the inequality

$$\mathsf{D}(X+Y) \le \left(\sqrt{\mathsf{D}X} + \sqrt{\mathsf{D}Y}\right)^2,$$

which is valid for any two random variables X and Y, and also identity (4.25) and the estimate for the variance given in Theorem 2, we obtain

$$\mathsf{E}\left[k\tilde r_n^{(k+1)} - (k+\beta/\eta)\tilde r_n^{(k)}\right]^2 = \mathcal{O}(n^{-c}) + \mathcal{O}(n^{-2a}). \tag{4.27}$$

Now, (1.21) with $c < \min\{1-b, d/\delta, 2a\}$ is a consequence of (4.27), (4.26), and (4.23).

If $\beta = 0$, then

$$\mathsf{E} r_n^{(k)}(0) = \frac{1}{\eta}\sum_{i=k}^{n-1}\frac{1}{i} - A + \mathcal{O}(n^{-a}) \tag{4.28}$$

by (4.22). Consequently, for the bias of the estimator (4.12) we have

$$\mathsf{E}\left[\eta_n^{(k)}(0,\rho)^{-1} - \eta^{-1}\right] = \mathcal{O}(n^{-a}). \tag{4.29}$$

Theorem 3 is proved. □

Proof of Remark 2. We substitute (1.22) into (4.20). For $\beta \ne 0$ this leads to an integral that reduces to the beta function and yields (1.20).

For $\beta = 0$, we obtain the formula

$$\Delta^{(k)}\phi_0(s) = (-1)^k\int_0^1 \Big[-\frac{\ln t}{\eta} - A + \mathcal{O}(t^a)\Big] t^k(1-t)^{s-k-1}\,dt.$$

Computing the integrals (see [GR71, 4.253.1]), we see that

$$\begin{aligned} &\Delta^{(k)}\phi_0(s)(-1)^k \\ &\quad = \frac{k!\Gamma(s-k)}{\Gamma(s+1)}\left[\frac{\psi(s+1)-\psi(k+1)}{\eta} - A + \mathcal{O}\left(\frac{\Gamma(k+a+1)\Gamma(s+1)}{\Gamma(k+1)\Gamma(s+a+1)}\right)\right]. \end{aligned}$$

Now, to get (1.20) it remains to use the asymptotics of the gamma function as $s \to +\infty$. □

Proof of Remark 3. We start with two identities ($x, y \in \mathbb{R}, k \in \mathbb{N}$):

$$
\begin{aligned}
&y \sum_{i=1}^{\infty} \frac{\Gamma(x-y+i)}{\Gamma(x+i+1)} = \frac{\Gamma(x-y+1)}{\Gamma(x+1)}, \\
&k \sum_{i=1}^{\infty} \frac{\Gamma(x-k+i)}{\Gamma(x+i+1)} \left(\psi(x+i+1) - \psi(k+1)\right) \\
&\qquad\qquad = \frac{\Gamma(x-k+1)}{\Gamma(x+1)} \left(\psi(x+1) - \psi(k)\right),
\end{aligned}
\tag{4.30}
$$

which can easily be deduced from the simple relation

$$
y\frac{\Gamma(x-y+i)}{\Gamma(x+i+1)} = \frac{\Gamma(x-y+i)}{\Gamma(x+i)} - \frac{\Gamma(x-y+i+1)}{\Gamma(x+i+1)}.
$$

Suppose condition (1.20) is satisfied. We have

$$
\Delta^{(k-1)}\phi_\beta(s) = -\sum_{i=1}^{\infty} \Delta^{(k)}\phi_\beta(s+i).
$$

Let $\beta = 0$. Using (1.20), we obtain

$$
\begin{aligned}
\Delta^{(k-1)}\phi_\beta(s) &= (-1)^{m-1}\Gamma(k+1) \\
&\quad\times \sum_{i=1}^{\infty} \frac{\Gamma(s+i-k)}{\Gamma(s+i+1)} \Big(\frac{\psi(s+i+1) - \psi(k+1)}{\eta} - A + \mathcal{O}((s+i)^{-a}) \Big).
\end{aligned}
$$

Recalling identities (4.30) and estimating the terms with $\mathcal{O}(s^{-a})$, we arrive at condition (1.20) with $k-1$.

The case where $\beta \neq 0$ is treated similarly. □

§5. Markov measures

Here we show that for the Markov measures (on sequence spaces) the β-statentropy exists and all the conditions of Theorems 1 and 2 are fulfilled.

Suppose we are given a Markov chain with a set of states $S = \{0, 1, \dots, s-1\}$, with transition probabilities $\|a_{ij}\|_0^{s-1}$, and with a stationary distribution $\{p_i : i \in S\}$ on S, where

$$
p_j = \sum_{i=0}^{s-1} p_i a_{ij}, \quad j \in S.
$$

For simplicity, we assume that $0 < a_{ij} < 1$.

Consider the sequence space $\Omega = S^{\mathbb{N}}$, where $S = \{0, 1, 2, \dots, s-1\}$, and $\mathbb{N} = \{0, 1, 2, \dots\}$. The points of $S^{\mathbb{N}}$ will be denoted by

$$
x = (x_0, x_1, \dots, x_n, \dots), \quad x_i \in S.
$$

We introduce a metric on $S^{\mathbb{N}}$ by putting

$$
\rho(x, y) = \theta^{-n}, \tag{5.1}
$$

where n is such that $x_n \neq y_n$, but $x_0 = y_0, \dots, x_{n-1} = y_{n-1}$ (here $\theta > 1$ is some parameter).

Let σ denote the shift

$$
\sigma(x_0, x_1, \dots, x_n, \dots) = (x_1, x_2, \dots, x_{n+1}, \dots).
$$

On the space $\Omega = S^{\mathbb{N}}$, we define a measure μ as the Markov measure with the initial distribution p_i and the transition probability matrix a_{ij}.

Clearly, the shift σ is a continuous transformation in the metric ρ given by (5.1), $\rho(\sigma(x), \sigma(y)) = \min\{1, \theta\rho(x, y)\}$, and μ is shift invariant.

Proposition 6. *For the Markov measure μ and the metric ρ, the 0-statentropy exists and is given by the formula*

$$\eta(0, \mu, \rho) = \frac{h(\sigma)}{\ln \theta}, \tag{5.2}$$

where

$$h(\sigma) = -\sum_{i=0}^{s-1}\sum_{j=0}^{s-1} p_i a_{ij} \ln a_{ij} \tag{5.3}$$

is the entropy of the shift σ, and θ is the parameter of the metric (5.1).

Proof. The shift σ is ergodic (see, e.g., [MI88]). Therefore, by the Breiman theorem (see [Si73]), the limit

$$\lim_{n\to\infty} \frac{1}{n} \ln \mu(C_n(x)) = -h(\sigma)$$

exists for μ-a.e. $x \in \Omega$; here the sets

$$C_n(x) = \{y \in S^{\mathbb{N}} : y_i = x_i, i = 0, 1, \ldots, n\} \tag{5.4}$$

are cylinders in the sequence space $S^{\mathbb{N}}$.

For the metric (5.1) we have

$$C_n(x) = B(x, \theta^{-n}) = B(x, r), \quad \theta^{-n-1} < r \le \theta^{-n}. \tag{5.5}$$

Consequently, the pointwise dimension $d_\mu(x) = \frac{h(\sigma)}{\ln \theta}$ exists and is constant for μ-a.e. $x \in \Omega$.

Thus, we can apply Corollary 2, which yields the required statement. □

Now we find the β-statentropy for $\beta \neq 0$.

Theorem 4. *For the Markov measure μ and the metric ρ, the β-statentropy exists and is given by the formula*

$$\eta(\beta, \mu, \rho) = \frac{\beta}{1 - q(\beta)}, \tag{5.6}$$

where $q(\beta)$ stands for the root of the equation

$$\Phi(q) = \theta^\beta, \tag{5.7}$$

and $\Phi(q)$ is the spectral radius of the matrix $\|a_{ij}^q\|_0^{s-1}$.

Proof. The definition (1.1) of $\nu(t, x)$ implies that

$$\nu(t, x) = \theta^{-n} \quad \text{if } \mu(C_n(x)) < t \le \mu(C_{n-1}(x)),\ n \ge 0,\ x \in \Omega. \tag{5.8}$$

(In the definition (5.4) we assume that $C_{-1}(x) = \Omega$.)

The function

$$\chi(\beta, t)^\beta = \int_\Omega \nu(t, x)^\beta \, d\mu(x)$$

can be represented as a sum:

$$\chi(\beta, t)^\beta = \sum_{i=0}^{s-1} f_i(t), \tag{5.9}$$

where

$$f_i(t) = \int_{T_i(\Omega)} \nu(t, x)^\beta \, d\mu(x) = \int_\Omega \nu(t, T_i(x))^\beta \, d\mu(T_i(x)), \tag{5.10}$$

and the maps T_i are defined by

$$T_i(x_0, x_1, \dots, x_n, \dots) = (i, x_0, x_1, \dots, x_n, \dots).$$

Similarly,

$$f_i(t) = \sum_{j=0}^{s-1} \int_\Omega \nu(t, T_i \circ T_j(x))^\beta \, d\mu(T_i \circ T_j(x)). \tag{5.11}$$

The definition of a Markov measure shows that the measure of the cylinders $C_n(T_i(x)) = T_i(C_{n-1}(x))$ can be found as follows:

$$\mu(C_n(T_i \circ T_j(x))) = \frac{a_{ij} p_i}{p_j} \mu(C_{n-1}(T_j(x))).$$

For the metric (5.1) the balls coincide with cylinders, the diameter of a cylinder coincides with the radius of the corresponding ball (see (5.5)), and

$$\operatorname{diam}(C_n(T_i(x))) = \theta^{-1} \operatorname{diam}(C_{n-1}(x)).$$

Therefore,

$$\nu(t, T_i \circ T_j(x)) = \theta^{-1} \nu\Big(\frac{t p_j}{a_{ij} p_i}, T_j(x)\Big), \quad d\mu(T_i \circ T_j(x)) = \frac{a_{ij} p_i}{p_j} d\mu(T_j(x)).$$

Substituting this into (5.11), we obtain the following recurrence equations for the functions $f_i(t)$:

$$f_i(t) = \sum_{j=0}^{s-1} \frac{a_{ij} p_i}{p_j} \theta^{-\beta} f_j\Big(\frac{t p_j}{a_{ij} p_i}\Big). \tag{5.12}$$

We show that the solution of (5.12) is of the form $f_i(t) \approx t^{1-q(\beta)}$, i.e., for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$t^{1-q(\beta)+\varepsilon} \le f_i(t) \le t^{1-q(\beta)-\varepsilon}, \quad 0 < t \le \delta. \tag{5.13}$$

By the Perron theorem (see [Gan88]), the positive matrix $\|b_{ij}^q\|$ possesses a positive eigenvector $(e_0(q), \dots, e_{s-1}(q))$, where

$$b_{ij} = \frac{a_{ij} p_i}{p_j}, \quad i, j \in S.$$

Since the spectra of the matrices $\|a_{ij}^q\|_0^{s-1}$ and $\|b_{ij}^q\|_0^{s-1}$ coincide, this eigenvector corresponds to the eigenvalue $\Phi(q)$. We assume that this eigenvector is normalized so that

$$\sum_{i=0}^{s-1} e_i(q) = 1.$$

Suppose that for some constant b we have

$$f_i(t) \le t^b e_i(1-b), \quad 0 < t \le t_0, \ \ i \in S. \tag{5.14}$$

We show that for any $b < 1 - q(\beta)$ this upper estimate can be improved. For such b we have $\Phi(1-b)\theta^{-\beta} < 1$ (because $\Phi(q)$ is a monotone decreasing function).

We put $\alpha = \min_i p_i$ and $t_1 = \alpha t_0$, and choose $z > 0$ so that

$$g(z) = t_1^{-z} C(z) \Phi(1-b-z) \theta^{-\beta} < 1,$$

where

$$C(z) = \max_i \frac{e_i(1-b)}{e_i(1-b-z)}.$$

Such a number z exists indeed, because $g(0) = \Phi(1-b)\theta^{-\beta} < 1$. We prove that

$$f_i(t) \le t^{b+z} e_i(1-b-z)$$

for $t \le t_1$.

Indeed, for $t_1 \le t \le t_0$ we have $f_i(t) \le t_1^{-z} t^{b+z} e_i(1-b)$. Therefore,

$$f_i(t) \le C(z) t_1^{-z} e_i(1-b-z) t^{b+z}, \quad t_1 \le t \le t_0.$$

Substituting this in the recurrence equation (5.12), we obtain

$$\begin{aligned} f_i(t) &\le t_1^{-z} C(z) t^{b+z} \sum_{j=0}^{s-1} \Big(\frac{a_{ij} p_i}{p_j}\Big)^{1-b-z} \theta^{-q} e_j(1-b-z) \\ &= g(z) t^{b+z} e_i(1-b-z) < t^{b+z} e_i(1-b-z). \end{aligned}$$

So, the estimate (5.14) of the function $f_i(t)$ on the interval $\alpha t_0 = t_1 \le t \le t_0$ carries over to the interval $\alpha^2 t_0 \le t \le \alpha t_0$ with the replacement of b by the larger value $b+z$. Therefore, (5.14) with $b+z$ in place of b is true on the entire interval $(0, \alpha t_0)$.

This proves inequalities (5.14).

In a similar way we check that for $a > 1 - q(\beta)$ the lower estimate

$$f_i(t) \ge t^a e_i(1-b), \quad 0 < t \le t_0, \ i \in S,$$

can be improved.

Thus, we have verified (5.13). By (5.13) and (5.9), for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$t^{1-q(\beta)+\varepsilon} \le \chi(\beta, t)^\beta \le t^{1-q(\beta)-\varepsilon}, \quad 0 < t \le \delta.$$

These estimates imply that the limit in (1.7) exists and that the β-statentropy is given by (5.6). □

Corollary 4. *For the Bernoulli* ($a_{ij} = p_j$) *measure* μ *and the metric* ρ, *the* β*-statentropy exists and is given by* (5.6), *where* $q(\beta)$ *is the root of the equation*

$$\sum_{i=0}^{s-1} p_i^q = \theta^\beta. \tag{5.15}$$

The definition of $q(\beta)$ shows that this function is monotone decreasing from $+\infty$ to $-\infty$. Therefore, the β-statentropy coincides with the HP-spectrum for dimensions with the parameter $q+1$.

So, for a Markov measure the β-statentropy exists. We show that the other conditions imposed in Theorem 2 on the measure and the metric are fulfilled.

Condition (1.19) in Remark 1 is satisfied for any point x, because for the metric (5.1) the balls are cylinders, and

$$\min_i p_i \big(\min_{i,j} a_{ij}\big)^m \le \mu(C_m(x)) \le \max_i p_i \big(\max_{i,j} a_{ij}\big)^m.$$

Referring to Remark 1, we see that conditions (1.15) and (1.16) are fulfilled in the case of a Markov measure.

To prove (1.17) in this case, we substitute the value (5.6) of the β-statentropy into inequality (1.17), obtaining an equivalent inequality:

$$q(2\beta) < 2q(\beta). \tag{5.16}$$

This is equivalent to

$$\theta^{2\beta} = \Phi(q(2\beta)) = \Phi(q(\beta))^2 > \Phi(2q(\beta)).$$

We prove that $\Phi(q)^2 > \Phi(2q)$.

Let $(e_0(q), \dots, e_{s-1}(q))$ be an eigenvector corresponding to the eigenvalue $\Phi(q)$ of the matrix $\|b_{ij}^q\|_0^{s-1}$. Then for the vector $(e_0(q)^2, \dots, e_{s-1}(q)^2)$ we have

$$\sum_{j=0}^{s-1} b_{ij}^{2q} e_j(q)^2 < \left(\sum_{j=0}^{s-1} b_{ij}^q e_j(q)\right)^2 = \Phi(q)^2 e_i(q)^2.$$

By the Perron theorem (see [Gan88, Chapter XIII, Remark 5]), it follows that $\Phi(2q) < \Phi(q)^2$. Thus, inequality (5.16) is fulfilled.

§6. Dynamical systems

We describe how the estimators (1.10)–(1.12) can be applied to finding dimensional characteristics of a dynamical system.

Let X be a compact metric space with metric d, and let $T : X \to X$ be a continuous map. We denote by P the (unknown) invariant measure of T and assume that P is a Borel probability measure.

First of all, to apply (1.10)–(1.12) we need n independent and identically distributed (with distribution P) initial points $x_1, \dots, x_n$.

For constructing $x_1, \dots, x_n$, we choose some measure P_0 on X and n independent initial points $x_1^0, \dots, x_n^0$ with distribution P_0. Then we consider $x_i = T^m(x_i^0)$, where the parameter m is chosen sufficiently large. The measure P_0 should be sufficiently smooth, so as to admit modeling of a sequence of independent and identically distributed initial points, and P_0 must be supported in the vicinity of the attractor. The parameter m should ensure that a point approach the attractor within a satisfactory accuracy.

Second, we must choose a compact metric space Ω and a metric ρ. We consider the following two standard cases in more detail:

- $\Omega = X$;
- Ω is a symbolic dynamical system.

6.1. $\Omega = X$**.** In this case the sequence $\xi_1, \xi_2, \dots, \xi_n$ with which we start the construction of our estimators coincides with $x_1, \dots, x_n$, and $\mu = P$. The role of ρ is played by the metric d, or by the family of metrics

$$d_m(x, y) = \max_{0 \le i \le m} \{d(T^i(x), T^i(y))\}, x, y \in X\}. \tag{6.1}$$

We recall (see, e.g., [Pes97]) the Brin–Katok formula for the entropy of an ergodic transformation T:

$$h_\mu(T) = \lim_{m \to \infty, r \to 0} \frac{-\ln \mu(B_m(x, r))}{m} \quad \text{for } \mu\text{-a.e. } x \in X, \tag{6.2}$$

where $h_\mu(T)$ is the metric entropy of T, and $B_m(x, r)$ is a ball in the metric d_m.

Applying this formula and Corollary 2, we obtain the following statement.

Proposition 7. *If the transformation T is ergodic, then*

$$\lim_{m \to \infty} \eta(0, \mu, d_m) = h_\mu(T).$$

6.2. Symbolic dynamical systems. We recall the definition of a symbolic dynamical system (see, e.g., [Pes97]).

Suppose we have a finite measurable partition of the space X into s subsets; this partition can be given via a function $\alpha : X \to S$. Consider the sequence space $S^{\mathbb{N}}$, where $S = \{0, 1, 2, \dots, s - 1\}$.

We endow $S^{\mathbb{N}}$ with the metric (5.1) with parameter $\theta > 1$. Note that this choice of the metric ρ is explained by the following properties:

1) in this metric, the balls are cylinders;

2) for evaluating $\rho(x,y)$, it suffices to know only finitely many elements of the sequences x and y.

The first property simplifies the proofs and calculations substantially. The second property makes it possible to find the exact values of $\rho(x,y)$ in experimental calculations.

The invariant measure P of the transformation T gives rise to the invariant measure μ of the shift σ on Ω (see, e.g., [Si73]). By construction, μ is a Borel probability measure.

The symbolic dynamical system (Ω,σ,μ) (σ is the shift) corresponding to the dynamical system (X,T,P) and the function $\alpha : X \to S$ is introduced as follows:

$$\Omega = \{\xi : \xi = (\alpha(x), \alpha(T(x)), \dots, \alpha(T^n(x)), \dots), x \in X\}. \tag{6.3}$$

Proposition 8. *Suppose (X,T,P) is a dynamical system with an ergodic transformation T. Then*

$$\underline{\eta}(0,\mu,\rho) = \overline{\eta}(0,\mu,\rho) = \frac{h(T,\alpha)}{\ln\theta},$$

where $h(T,\alpha)$ is the entropy of T relative to the partition α, and θ is the parameter of the metric (5.1).

Proof. Since T is ergodic, so is the shift σ. Therefore, by the Breiman theorem (see [Si73]), the limit

$$\lim_{n\to\infty} \frac{1}{n} \ln \mu(C_n(\xi)) = -h(T,\alpha)$$

exists for μ-a.e. $\xi \in \Omega$.

Since the balls are cylinders (for the metric (5.1) under consideration), the point dimension $d_\mu(\xi) = \frac{h(T,\alpha)}{\ln\theta}$ exists and is constant for μ-a.e. $\xi \in \Omega$. Applying Corollary 2, we get the required statement. □

So, in order to apply (1.10)–(1.12), we choose an auxiliary parameter m and, starting with a given sequence $x_1, \dots, x_n$, construct a sequence of points $\xi_1, \xi_2, \dots, \xi_n \in \Omega$ by putting

$$\xi_{ij} = \alpha(T^j(x_i)), \quad j = 0, 1, \dots, m, \ \ i = 1, 2, \dots, n.$$

If, when finding the estimators (1.10)–(1.12), all distances $\rho(\xi_i, \xi_j)$ turn out to be greater than θ^{-m}, then they are the distances between infinite sequences ξ_i (though they are determined by the first m elements of the sequences). If one of the distances $\rho(\xi_i, \xi_j)$ is equal to θ^{-m}, then the parameter m should be increased.

In conclusion, the author expresses his gratitude to professor V. L. Dol′nikov for attention and friendly advice.

References

[BP88] R. Badii and A. Politi, *Numeric investigation of nonhomogeneous fractals*, In: Fractals in Physics (Proc. of the Sixth Trieste International Symposium on Fractals in Physics, Trieste, 1985), North-Holland, Amsterdam, 1986.

[Bi69] P. Billingsley, *Ergodic theory and information*, Wiley, New York, 1965. MR0192027 (33:254)

[Bo79] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Math., vol. 470, Springer, Berlin, 1975.

[VM95] V. A. Vatutin and V. G. Mikhaĭlov, *Statistical estimation of the entropy of discrete random variables with a large number of outcomes*, Uspekhi Mat. Nauk **50** (1995), no. 5, 121–134; English transl., Russian Math. Surveys **50** (1995), no. 5, 963–976. MR1365051 (97b:62066)

[Gan88] F. R. Gantmacher, *Theory of matrices*, 4th ed., "Nauka", Moscow, 1988; German transl., *Matrizentheorie*, VEB Deutscher Verlag Wiss., Berlin, 1986. MR0986246 (90a:15001); MR0863127 (87i:15001)

[GR71] I. S. Gradshteĭn and I. M. Ryzhik, *Table of integrals, sums, series, and products*, "Nauka", Moscow, 1971; English transl., Acad. Press, Inc., San Diego, CA, 2000. MR1773820 (2001c:00002)

[Do58] R. L. Dobrushin, *A simplified method of experimentally evaluating the entropy of a stationary sequence*, Teor. Veroyatnost. i Primenen. **3** (1958), no. 4, 462–464. (Russian) MR0103134 (21:1917)

[MT99] V. V. Maĭorov and E. A. Timofeev, *A consistent estimate for the dimension of manifolds and self-similar fractals*, Zh. Vychisl. Mat. Mat. Fiz. **39** (1999), no. 10, 1721–1729; English transl., Comput. Math. Phys. **39** (1999), no. 10, 1651–1659. MR1726499

[MT02] ———, *Statistical estimation of generalized dimensions*, Mat. Zametki **71** (2002), no. 5, 697–712; English transl., Math. Notes **71** (2002), no. 5–6, 634–648. MR1936194 (2003h:62033)

[MP00] G. G. Malinetskiĭ and A. B. Potapov, *Modern problems of nonlinear dynamics*, "URSS", Moscow, 2002. (Russian)

[MI88] N. F. G. Martin and J. W. England, *Mathematical theory of entropy*, Encyclopedia Math. Appl., vol. 12, Addison–Wesley Publishing Co., Reading, MA, 1981. MR0612318 (83k:28019)

[Si73] Ya. G. Sinaĭ, *Introduction to ergodic theory*, Erevan Univ., Erevan, 1973; English transl., Math. Notes, vol. 18, Princeton Univ. Press, Princeton, NJ, 1976. MR0584788 (58:28437)

[Fed87] H. Federer, *Geometric measure theory*, Grundlehren Math. Wiss., vol. 153, Springer-Verlag New York, Inc., New York, 1969. MR0257325 (41:1976)

[Fel84] W. Feller, *An introduction to probability theory and its applications*. Vol. II, John Wiley and Sons, Inc., New York, 1971. MR0270403 (42:5292)

[HLP48] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge Univ. Press, Cambridge, 1988. MR0944909 (89d:26016)

[BS00] L. Barreira and J. Schmeling, *Sets of "non-typical" points have full topological entropy and full Hausdorff dimension*, Israel J. Math. **116** (2000), 29–70. MR1759398 (2002d:37040)

[BBD99] S. Borovkova, R. Burton, and H. Dehling, *Consistency of the Takens estimator for the correlation dimension*, Ann. Appl. Probab. **9** (1999), 376–390. MR1687339 (2000g:60035)

[CS01] F. Cucker and S. Smale, *On the mathematical foundation of learning*, Bull. Amer. Math. Soc. (N.S.) **39** (2002), no. 1, 1–49. MR1864085 (2003a:68118)

[Cu93] C. Cutler, *A review of the theory and estimation of fractal dimension*, Dimension Estimation and Models, Nonlinear Time Ser. Chaos, vol. 1, World Sci. Publishing, River Edge, NJ, 1993, pp. 1–107. MR1307656 (95i:58106)

[ER92] J.-P. Eckmann and D. Ruelle, *Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems*, Phys. D **56** (1992), no. 2–3, 185–187. MR1164396 (93c:58122)

[GP83] P. Grassberger and I. Procaccia, *Characterization of strange attractors*, Phys. Rev. Lett. **50** (1983), no. 5, 346–349. MR0689681 (84k:58141)

[Gr89] P. Grassberger, *Estimating the information content of symbol sequences and efficient codes*, IEEE Trans. Inform. Theory **35** (1989), 669–675. MR1022090 (90h:94030)

[HJKPS86] Th. C. Halsey, M. N. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman, *Fractal measures and their singularities: The characterization of strange sets*, Phys. Rev. A (3) **33** (1986), no. 2, 1141–1151. MR0823474 (87h:58125a)

[KS94] I. Kontoyiannis and Yu. M. Suhov, *Prefixes and the entropy rate for long-range sources*, Probability, Statistics and Optimization (F. P. Kelly, ed.), Wiley, Chichester, 1994. MR1320744 (96f:94004)

[KASW98] I. Kontoyiannis, P. H. Algoet, Yu. M. Suhov, and A. J. Wyner, *Nonparametric entropy estimation for stationary processes and random fields, with applications to English text*, IEEE Trans. Inform. Theory **44** (1998), 1319–1327. MR1616653 (99c:94015)

[Pes97] Ya. Pesin, *Dimension theory in dynamical systems: Contemporary views and applications*, Univ. Chicago Press, Chicago, IL, 1997. MR1489237 (99b:58003)

[Ri95] R. H. Riedi, *An improved multifractal formalism and self-similar measures*, J. Math. Anal. Appl. **189** (1995), 462–490. MR1312056 (95k:28021)

[Ser96] R. Serinko, *Ergodic theorems arising in correlation dimension estimation*, J. Statist. Phys. **85** (1996), 25–40. MR1413235 (97j:62060)

[Shi92] P. C. Shields, *Entropy and prefixes*, Ann. Probab. **20** (1992), 403–409. MR1143428 (93b:28048)

[Tak85] F. Takens, *On the numerical determination of the dimension of the attractor*, Dynamical Systems and Bifurcations (Groningen, 1984), Lecture Notes in Math., vol. 1125, Springer, Berlin, 1985, pp. 99–106. MR0798084 (87c:58063)

[TV98] F. Takens and E. Verbitski, *Generalized entropies: Rényi and correlation integral approach*, Nonlinearity **11** (1998), 771–782. MR1632625 (99b:28023)

[TA83] Y. Termonia and Z. Alexandrowicz, *Fractal dimension of strange attractors from radius versus size of arbitrary clusters*, Phys. Rev. Lett. **51** (1983), no. 14, 1265–1268. MR0718464 (84k:58155)

[TL93] J. Theiler and T. Lookman, *Statistical error in a chord estimator of correlation dimension: "the rule of five"*, Internat. J. Bifur. Chaos **3** (1993), no. 3, 765–771.

Yaroslavl′ State University, Sovetskaya Street 14, Yaroslavl′ 150000, Russia
E-mail address: `tim@uniyar.ac.ru`

Received 20/FEB/2004

Translated by A. PLOTKIN