数理解析研究所講究録
第 635 巻 1987 年 59-79

Asymptotic behaviours of measures of small tubes: entropy, Liapunov's exponent and large deviation

Y.Takahashi 高橋陽一郎

Department of Pure and Applied Sciences
University of Tokyo
Komaba, Meguro, Tokyo 153, Japan

0. Introduction

The metrical entropy is the quantity which was introduced by Kolmogorov and Sinai as an invariant for the isomorphism problem among Bernoulli shifts and was proved to be a complete invariant among them by Ornstein. It has been known that it is related to various branches of mathematics, one of which is the theory of differentiable dynamical systems. What is our main concern in the present report is the Ruelle-Pesin inequality between the metrical entropy and the Liapunov's characteristic exponent viewed from the large deviation theory.

The large deviation problem is one of current topics in probability theory and we shall illustrate what it is in Section 3. A large deviation problem for compact dynamical systems was formulated in [20,19] in connection with the variational principle for one dimensional chaos. The formulation will be reproduced in Section 4. As our title suggests it, it turned out in [22] that a method of computation of large deviation rate functional is the asymptotic evaluation of small tubes, i.e., to estimate the logarithmic asymptotics for the long time of the measures of small tubes around orbits of dynamical systems. Such asymptotic evaluation also appeared for the metrical entropy as the local entropy theorem of Katok and Brin (cf. Section 1). Of course, Liapunov's characteristic exponent is, by its nature, the rate of logarithmic asymptotics of measures of such small tubes (cf. Section 2).

Thus our main object will be the asymptotics of measures of small tubes around orbits of compact dynamical systems and our goal of the present report is the following result:

Theorem 0. Let X be a compact Riemannian manifold and a diffeomorphism F of X. Take the Riemannian volume as the reference measure m of the large deviation problem stated in Sect.4. Then the lower rate functional $\underline{q}(\mu)$ satisfies the following estimate from below for every ergodic F-invariant measure μ:

$$\underline{q}(\mu) \geqq h_\mu(F) - \chi_\mu^+, \tag{0.1}$$

where $h_\mu(F)$ is the metrical entropy of the system (X,F,μ) and χ_μ^+ denotes the sum of positive characteristic exponents.

The proof of Theorem 0 will be given in the last section.

As a corollary we can obtain the following

Pesin-Ruelle inequality: For every F-invariant probability Borel mesure μ there holds the inequality

$$h_\mu(F) \leqq \chi_\mu^+. \tag{0.2}$$

In fact, (0.2) follows from the inequality $\underline{q}(\mu) \leqq 0$, which is trivial since $\underline{q}$ is defined as the logarithmic asymptotics of probability. The gap between "every F-invariant" and "every ergodic F-invariant" in (0.1) and (0.2) is filled by the Krein-Milman-Bishop theorem on the integration representation over extremal points (cf. [16]). Indeed, the ergodic measures are the extremal points among F-invariants measures, $\mu(F)$ is a convex function of μ and χ_μ^+ is an affine function of μ.

The title of the sections are as follows:

1. Local entropy and Shannon-McMillan theorem
2. Characteristic exponents and Pesin-Ruelle inequality
3. What is the large deviation?
4. Large deviation for compact dynamical systems
5. Proofs: Hamming distance, Fano's lemma and Katok's lemma

1. Local entropy and Shannon-McMillan theorem

Let (X,d) be a compact metric space and $F:X\to X$ a continuous transformation (or $F=(F^t)_{t\geq 0}$ a jointly continuous semiflow). For given positive T,δ and $x\in X$, put

$$(1.1) \qquad B_T(x,\delta) = \{ y\in X;\ d(F^t y, F^t x)\leq\delta,\ t\in[0,T) \}$$

and let us consider asymptotic behaviours of measures of "small tubes" $B_T(x,\delta)$. The first one is the local entropy theorem.

<u>Theorem 1</u>.(A.Katok [8] (ergodic case);M.Brin-A.Katok [3]) For an F-invariant probability Borel measure μ on X put

$$(1.2) \qquad \overline{h}_\mu(x) := \lim_{\delta\to 0}\ \limsup_{T\to\infty} -\frac{1}{T}\log \mu(B_T(x,\delta)) \qquad \text{and}$$

$$(1.3) \qquad \underline{h}_\mu(x) := \lim_{\delta\to 0}\ \liminf_{T\to\infty} -\frac{1}{T}\log \mu(B_T(x,\delta)).$$

Then the following statements are true:

(i) $\underline{h}_\mu(x)=\overline{h}_\mu(x):=h_\mu(x)$ for μ-almost every x.

(ii) $h_\mu(x)$ is F-invariant: $h_\mu(Fx)=h_\mu(x)$ for μ-almost every x.

(iii) $\int_X h_\mu(x)\mu(dx) = h_\mu(F)$

(iv) In particular, if μ is ergodic, $h_\mu(x)=h_\mu(F)$ for μ-almost every $x\in X$.

Let us explain the notations used above. The quantity $h_\mu(F)$ is the metrical entropy (or the Kolmogorov-Sinai entropy) of the measure preserving dynamical system (X,F,μ). Namely,

$$(1.4) \quad h_\mu(F) = \sup\{ h_\mu(F;\alpha);\ \alpha \text{ is a finite Borel partition of } X \}$$

$$(1.5) \qquad h_\mu(F;\alpha) = \lim_{n\to\infty} \int_X -\frac{1}{n}\log \mu(\alpha_n(x))\ \mu(dx)$$

where we denote by $\alpha_n(x)$ the cell of the refinement α_n of the partitions α, $F^{-1}\alpha$, $\cdots$, $F^{-(n-1)}\alpha$ to which x belongs:

$$\alpha_n(x)=\alpha(x)\cap F^{-1}\alpha(Fx)\cap\cdots\cap F^{-(n-1)}\alpha(F^{n-1}x).$$

Of course, the almost everywhere version of (1.5) is the well-known fundamental theorem of information theory:

<u>Theorem 2</u>.(Shannon-McMillan-Breiman) Let μ be F-invariant. Then,

(i) $I_\mu(x) := \lim_{n\to\infty} -\frac{1}{n} \log \mu(\alpha_n(x))$ exists for μ-almost every x.

(ii) $I_\mu(Fx) = I_\mu(x)$ for μ-almost every x.

(iii) $\int_X I_\mu(x)\mu(dx) = h_\mu(F)$.

(iv) In particular, if μ is ergodic, then, $I_\mu(x) = h_\mu(F)$ μ-a.e.x.

As we have said that it is the almost everywhere version, Theorem 2 is proved from (1.5) by Birkhoff's individual ergodic theorem using Lebesgue's dominated convergence theorem. (Cf., e.g., [24] or [4].) The proof of Theorem 1 is finally reduced to Theorem 2. We shall give a short proof of Theorem 1 in the last section, which is based on a Katok's original idea and a large deviation result.

2. Characteristic exponent and the Pesin-Ruelle inequality

The following theorem is a version of Oseledec's multiplicative ergodic theorem. Let us say that a property holds for universally almost every x (abr., u.a.e.x) if it is true for μ-almost every x whenever μ is an invariant probability Borel measure.

<u>Theorem 3</u>. Let X be a compact Riemannian manifold and d the Riemannian distance. Denote the Riemannian volume measure by vol. For $x\in X$ let us denote

$$\overline{\chi}^+(x) = \lim_{\delta\to 0} \limsup_{T\to\infty} -\frac{1}{T} \log \mathrm{vol}(B_T(x,\delta)),$$

$$\underline{\chi}^+(x) = \lim_{\delta\to 0} \liminf_{T\to\infty} -\frac{1}{T} \log \mathrm{vol}(B_T(x,\delta)).$$

Then the following statements are true:

(i) The equality $\overline{\chi}^+(x) = \underline{\chi}^+(x)$ holds for u.a.e.$x\in X$.

Let us denote the common value by $\chi^+(x)$.

(ii) $\chi^+(Fx) = \chi^+(x)$ for u.a.e. $x\in X$.

(iii) $\chi^+(x)$ is the sum of all positive Liapunov exponents (taking into account of the multiplicities).

(iv) In particular, if μ is an ergodic F-invariant measure, then, $\chi^+(x)$ is constant μ-a.e. x. We shall denote the constant by $\chi_\mu^+(F)$.

As is stated in Introduction, we shall call the following inequality between the quantity $\chi_\mu^+(F)$ and the metrical entropy $h_\mu(F)$ the Pesin-Ruelle inequality.

Theorem 4. For every F-invariant probablity Borel measure μ there holds the inequality:

(2.1) $$h_\mu(F) \leqq \chi_\mu^+(F).$$

This inequality or the equality was obtained in particular cases by many people, for instance, for β-transformation the equality was obtained in [7]. But Pesin is the first who discussed it for general hyperbolic systems and D.Ruelle generalized it. In the recent text book [10], Mane calls the equality holding for absolutely continuous measures μ the Pesin equality and the general inequality the Ruelle inequality.

In [1] (cf. the references therein and in [14]) we can find a survey on some results concerning the relationship between the metrical entropy and the curvatures which are obtained from Pesin's equality. It may be regarded as a good example of the application of Theorem 1 and Theorem 3 or of the method of asymptotic evaluation of small tubes.

Let M be a compact manifold of negative curvature, $F=(F^t)$ the geodesic flow and μ the Lebesgue measure on the bundle SM of unit tangent vectors. Pesin showed from his equality that the metrical entropy $h_\mu(F)$ satisfies the following equality:

(2.2) $$h_\mu(F) = \int_{SM} - \operatorname{tr} U(v)\, \mu(dv)$$

where U(v) denotes the second fundamental form at v of the horosphere determined by v. (His proof may be interpreted as a computation based upon the local entropy theorem.) Using (2.2), Freire and Mane obtained the inequality:

(2.3) $$h_\mu(F) \leq (n-1)^{1/2}\{ \int_{SM} - \mathrm{Ric}(v)\mu(dv) \}^{1/2}$$

where n=dim M and Ric(v) stands for the Ricci curvature at v. Also Osserman and Sarnak obtained the following inequality:

(2.4) $$h_\mu(F) \geq \int_{SM} \mathrm{tr}(\sqrt{Q(v)})\mu(dv)$$

where $Q(v)X= - R(X,v)v$ and R is the curvature tensor of M.

If we consider flows defined by a differential equation, the logarithmic asymptotics is governed by its variational equation. In particular, if we are concerned with a geodesic flow (F^t), it is described by the Jacobi fields. The Jacobi equation along geodesic γ is of the form

(2.5) $$\frac{D^2X}{dt^2} + R(\dot{\gamma}(t),X)\dot{\gamma}(t) = 0,$$

where $\frac{D}{dt}$ denotes the covariant derivative along γ. Note that Ric(v) is equal to the trace of the map $X \to -Q(v)X=R(v,X)v$. Now take a basis $Y_1,\cdots,Y_n$ parallel to the geodesic γ with $Y_n=\dot{\gamma}$ and put $X = \Sigma_{i=1}^{n-1} x_i Y^i$. Then, for $x=(x_i)_{1\leq i\leq n-1}$, the equation (2.3) takes the form

(2.6) $$\frac{d^2x}{dt^2} = A(t)x$$

where $A(t)=(a_{ij}(t))_{1\leq i,j\leq n-1}$ is the matrix representation of $Q(F^tv)$. Therefore, trace $A(t) = -\mathrm{Ric}(\dot{\gamma}(t))$ and $A(t)$ is a positive definite matrix since the sectional curvature of M is assumed to be negative. Thus, all the problems of asymptotic evaluation of the μ-measure of small tubes are reduced to the study of the second order linear equation (2.6) and, necessarily, of the Riccatti equation associated with it: the stable and the unstable subspaces of the tangent space TSM are characterized by the special solutions called limiting solutions, etc.

3. What is the large deviation?

Let us illustrate the large deviation problem in the case of coin-tossing. The large deviation problem is an old problem in probability theory (cf. Cramer[5]) but its current aspect was introduced by M.D.Donsker and R.S.Varadhan [6]. The probability theory is a branch of mathematics whose main objects are statistical laws. The most basic one is the law of large numbers. Let X_n, $n\geq 1$, be a trial of coin-tossing with $Pr(X_n=1)=p$ and $Pr(X_n=0)=1-p$ $(0<p<1)$. Here we identify the head with 1 and the tail with 0. Then the law of large numbers takes the form

$$\text{(3.1a)} \qquad \lim_{n\to\infty} \frac{1}{n}(X_1+X_2+\dots+X_n) = E[X_n] = p \qquad \text{w.p.1.}$$

where w.p.1 stands for "with probability one". In a higher level, denoting the Dirac measure at a point x by δ_x ($\delta_x(A)=1$ or 0 according as x belongs to A or does not), we have the following law of large numbers

$$\text{(3.2a)} \qquad \lim_{n\to\infty} \frac{1}{n}(\delta_{X_1}+\delta_{X_2}+\dots+\delta_{X_n}) = \delta_p \qquad \text{w.p.1,}$$

where we consider the weak topology on the measure space.

Furthermore we can go up to the space of infinite sequences $\{0,1\}^{\mathbf{N}}$ to obtain Birkhoff's individual ergodic theorem

$$\text{(3.3a)} \qquad \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \delta_{(X_i,X_{i+1},\dots)} = m \qquad \text{w.p.1,}$$

where m is the law of the coin-tossing X_n, i.e., m is the direct product measure of infinite copies of the probability measure p,1-p on $\{0,1\}$.

The second law is called the central limit theorem in probability theory concerning small fluctuations around the mean or the small deviation from the law of large numbers. For coin-tossing it takes the form

$$(3.1a)\qquad \lim_{n\to\infty} \Pr\{ \sqrt{n}[\tfrac{1}{n}(X_1+X_2+\cdots+X_n)-p]\in[a,b)\} = \int_a^b \frac{e^{-x^2/\sigma^2}}{\sqrt{2\pi}\sigma}\,dx$$

where σ^2 is the covariance $E[(X_n-E[X_n])^2]=p(1-p)$.

The third one is the problem of the large deviation from the law of large numbers. Here we mean $O(1)$ by "large" while we called $O(1/\sqrt{n})$ "small" in (3.1a). Let us consider the following probability for an interval I

$$\Pr\{ \tfrac{1}{n}(X_1+\cdots+X_n)\in I \}.$$

If the interval I does not contain the mean p, it must vanish as n goes to infinity. So we want to discuss the rate of its decay. Now using Stirling's formula it is immediate to see

$$\Pr\{ \tfrac{1}{n}(X_1+\cdots X_n)\in I \} = \sum_{k\in nI} \frac{n!}{k!(n-k)!}\, p^k(1-p)^{n-k}$$

$$\sim \sum_{k\in nI} \left(\frac{n}{2\pi k(n-k)}\right)^{1/2} \exp nH\left(\tfrac{k}{n},1-\tfrac{k}{n}\,\middle|\,p,1-p\right)$$

where

$$H(q_1,\cdots,q_N|p_1,\cdots,p_N) = -\sum_{i=1}^{N} q_i\log(q_i/p_i)$$

is the quantity called the relative entropy of probability distribution (q_i) relative to (p_i). Note that this quantity H is always nonpositive and $-H$ is the so-called relative information. Now taking the logarithm and the limit over n we obtain

$$(3.1c)\qquad \lim_{n\to\infty} \frac{1}{n} \log \Pr\{ \tfrac{1}{n}(X_1+\cdots+X_n)\in I \} = \max_{x\in I} H(x,1-x|p,1-p).$$

If we mount up to the sequence space, we cannot expect the equality any more but we obtain the following two inequalities:

$$\limsup_{n\to\infty} \frac{1}{n}\log \Pr\{ \frac{1}{n} \sum_{i=1}^{n} \delta_{(X_i,X_{i+1},\cdots)}\in C \} \leq \max_{\mu\in C} q(\mu)$$

(3.3c) if C is a weakly closed subset,

$$\liminf_{n\to\infty} \frac{1}{n}\log \Pr\{ \frac{1}{n} \sum_{i=0}^{n} \delta_{(X_i,X_{i+1},\cdots)}\in G \} \geq \sup_{\mu\in G} q(\mu)$$

if G is a weakly open subset.

where the functional $q(\mu)$ for a measure μ is given as follows. Let F be the shift transformation on the sequence space: $(Fx)_i = x_{i+1}$ for $x=(x_i)$. Then,

$$q(\mu) = \begin{cases} h_\mu(F) - \int_X U(x)\mu(dx) & \text{if } \mu \text{ is } F\text{-invariant} \\ -\infty & \text{otherwise.} \end{cases}$$

with

$$U(x) = \begin{cases} -\log p & \text{if } x_0=1 \\ -\log(1-p) & \text{if } x_0=0. \end{cases}$$

In particular, in case of the fair coin-tossing where $p=1/2$ the rate functional $q(\mu)$ governing the long time asymptotics (3.3c) takes the following form for an F-invariant μ:

$$q(\mu) = h_\mu(F) - \log 2 \ (\leq 0).$$

It is the metrical entropy renormalized so that the maximum is 0.

4. Large deviation problem for dynamical system

The large deviation problem is formulated for compact dynamical systems in [20,19] as follows. Let us denote by $\underline{M}(X)$ the totality of probability Borel measures on X and we consider the weak topology on $\underline{M}(X)$ as the dual space of the Banach space of all continuous functions on X with supremum norm. First of all let us fix a probability Borel measure m on X. For a subset G of $\underline{M}(X)$ put

(4.1) $$Q(G) = \limsup_{n\to\infty} \frac{1}{n} \log m\{ x\in X; \frac{1}{n}\sum_{i=0}^{n-1} \delta_{F^i x} \in G \}$$

(4.$\underline{1}$) $$\underline{Q}(X) = \liminf_{n\to\infty} \frac{1}{n} \log m\{ x\in X; \frac{1}{n}\sum_{i=0}^{n-1} \delta_{F^i x} \in G \}.$$

Then for $\mu\in\underline{M}(X)$ let us define the functionals

(4.2) $$q(\mu) = \inf\{ Q(G);\ G \text{ is open and } \mu\in G \},$$

(4.$\underline{2}$) $$\underline{q}(\mu) = \inf\{ \underline{Q}(G);\ G \text{ is open and } \mu\in G \}.$$

The folowing result is proved in [20]:

Theorem 5. (i) The functionals q and $\underline{q}$ on $\underline{M}(X)$ are lower semi-continuous :

$\limsup q(\mu_\alpha) \leqq q(\mu)$ and $\limsup \underline{q}(\mu_\alpha) \leqq \underline{q}(\mu)$ if $\mu = \lim \mu_\alpha$.
Also there hold the inequalities: $-\infty \leqq \underline{q}(\mu) \leqq q(\mu) \leqq 0$.

(ii) $Q(C) \leqq \max\{\ q(\mu);\ \mu\epsilon C\ \}$ for closed C.

(iii) $\underline{Q}(G) \geqq \sup\{\ \underline{q}(\mu);\ \mu\epsilon G\ \}$ for open G.

Following [25] let us say that the large deviation principle holds for the triplet (X,F,m) if there exists a functional q^*, called the rate functional, on $\underline{M}(X)$ with the following two properties:

(a) If $C\subset\underline{M}(X)$ is closed, then, $Q(C)\leqq\max\{q(\mu);\ \mu\epsilon C\ \}$.

(b) If $G\subset\underline{M}(X)$ is open, then, $\underline{Q}(G)\geqq\sup\{\underline{q}(\mu);\ \mu\epsilon G\ \}$.

Then it follows from Theorem 5 that the necessary and sufficient condition for the large deviation principle for the triplet (X,F,m) is the equality $q=\underline{q}$ and then the rate functional q^* is equal to q and $\underline{q}$.

Example 1. Let A be a finite alphabet set, $X=A^{\mathbf{N}}$, F the shift, $(Fx)_i=x_{i+1}$ for $x=(x_i)$, and m be a Bernoulli measure. Thus m is the product measure of infinite copies of a probability measure π on A. Then the large deviation principle holds and the rate functional q is given as follows for F-invariant μ:

$$q(\mu) = h_\mu(F) - \int_X \mu(dx)[-\log \pi(x_0)]. \tag{4.3}$$

A similar result holds for a class of measures μ including Markov measures. The proof is found in [20].

Example 2. Let (X,F) be an Anosov diffeomorphism and take the Riemannian volume measure as the reference measure m. Then the large deviation principle holds and the rate function q will take the following form for F-invariant μ:

$$q(\mu) = h_\mu(F) - \int_X \mu(dx)\log \phi^u(x), \tag{4.4}$$

where $\phi^u(x)$ is the expanding rate of the differential dF on the unstable subspace E^u of the tangent space T_xX at x. It seems to the author that the proof is eventually done among the works of Pesin et al. but he does not know the correct references.

Example 3([19]). Let X be a bounded closed interval, F a piecewise C^2-map of X to itself and m the Lebesgue measure. Assume that ess.inf $|F'|$ is positive. Then the large deviation principle holds and the rate functional for F-invariant μ is given as

$$(4.5) \qquad q(\mu) = \min\{\ 0,\ h_\mu(F) - \int_X \mu(dx)\log|F'(x)|\ \}.$$

Here note that if there is a strictly stable periodic orbit and μ is the uniform measure on it, then, $q(\mu)=0$ because $h_\mu(F)=0$ and the integral of the logarithm of the Jacobian, log $|F'|$, is negative.

Remark 1. If the relation (4.3) or (4.4) holds, then the functional q is affine:

$$(4.6) \qquad q(t\mu_1+(1-t)\mu_2) = tq(\mu_1) + (1-t)q(\mu_2) \qquad (0<t<1)$$

because the metrical entropy is $h_\mu(F)$ is affine if F admits a generating partition. On the other hand, the relation (4.5) shows that q is not affine if F shows window phenomenon in the terminology of [YT's], namely, if F has a stable periodic orbit together with a topological chaotic repeller. As Example 4 below suggests it strongly, the affine property of the rate functional q appears to be closely related to the structural stability of the dynamical system (X,F). At least the following statement ([20] pp458-460) is true:

If the non-wandering set $\Omega(F)$ is finite, then the affine property of q is equivalent to the no-cycle condition on (X,F). (Here we take the Riemannian volume measure as the reference measure m on X.)

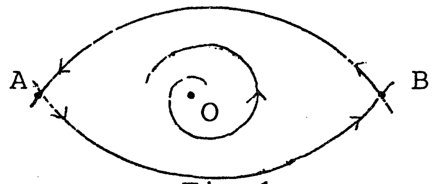

Fig.1

Example 4. Let X be an "eye-like" closed region of the plane with two saddles A and B at corners and one unstable focus O at the centre as in Fig.1 ([20]). There are several controllable parameters: the expanding and contracting rates at two foci, the approaching rate to the seperatrices on the boundary, etc. It can be shown that the following two cases may happen:

(a) $q(\mu) > \underline{q}(\mu)$ for some μ.

(b) $q(\mu) = \underline{q}(\mu)$ for all μ but the affine property fails.

Remark 2. The results in the present paragraph were obtained in the course of the study of the Gibbs type variational principle for one-dimensional chaos (cf. [12,18]). The Gibbs type variational principle for a compact dynamical system (X,F) with reference measure m is stated as follows. For a continuous function Φ on the measure space $\underline{M}(X)$ put

$$(4.7)\qquad P(\Phi) = \limsup_{n\to\infty} \frac{1}{n} \log \int_X m(dx)\exp\{ -n\Phi(\frac{1}{n} \sum_{i=0}^{n-1} \delta_{F^i x}) \}.$$

In particular, if $\Phi(\mu) = \Phi_U(\mu) = \int_X U(x)\mu(dx)$ with $U \in \underline{C}(X)$, then,

$$(4.8)\qquad P(\Phi_U) = \limsup_{n\to\infty} \frac{1}{n} \log \int_X m(dx)\exp\{ -\sum_{i=0}^{n-1} U(F^i x) \}.$$

Furthermore, let us denote the Legendre transform of $P(\Phi_U)$ by $P^\wedge$:

$$(4.9)\qquad P^\wedge(\mu) = \inf\{ P(\Phi_U) + \Phi_U(\mu) \ ;\ U \in \underline{C}(X) \}.$$

Theorem 6.([20,18]) The following statements are true:

(i) $-\infty \leq \underline{q}(\mu) \leq q(\mu) \leq P^\wedge(\mu) \leq 0$.

(ii) For every $\Phi\in\underline{C}(\underline{M}(X))$ the following variational principle of Gibbs type holds:

$$P(\Phi) = \max\{ q(\mu) - \Phi(\mu) \ ;\ \mu\in\underline{M}(X) \}.$$

(iii) In particular,

$$\max\{ q(\mu);\ \mu \in \underline{M}(X) \} = 0.$$

Remark 3. The following equality (the inverse Legendre transform theorem for convex functions on Banach spaces) is always true:

$$P(\Phi) = \max\{\ P^{\wedge}(\mu) - \Phi(\mu)\ ;\ \mu\in \underline{M}(X)\ \}.$$

(Cf. e.g., [17].)

5. Proofs: Hamming distance, Fano's lemma and Katok's lemma

First we shall prepare two lemmas and then proceed to the proofs of Theorem 0 and Theorem 1.

Let α be a finite (alphabet) set. The Hamming distance between words (finite sequences) $u=(a_0,\cdots,a_{n-1})$ and $v=(b_0,\cdots,b_{n-1})$ of length n ($a_i, b_i\in\alpha$, $0\leq i<n$) is defined as

$$(5.1)\qquad D_n(u,v) = \frac{1}{n}\ \#\{i;\ a_i\neq b_i\ \} = \frac{1}{n}\sum_{i=0}^{n-1} 1[a_i\neq b_i].$$

In what follows we shall define the Hamming distance between cells $\alpha_n(x)$ of the partition α_n by identifying the words $(a_0,\cdots,a_{n-1})$ and the intersections of a_0, $F^{-1}a_1,\cdots,F^{-(n-1)}a_{n-1}$ where a_i's are cells of a given finite partition α. For $\varepsilon>0$ and $x\in X$ let us denote the ε-neighbourhood of the cell $\alpha_n(x)$ with respect to the Hamming distance D_n by $\alpha_n^{\varepsilon}(x)$.

The following proposition is a version of Fano's lemma on the asymptotic error probability in information theory:

Lemma 1. Let α be a finite measurable partition of X and μ be an F-invariant probability Borel measure on X. Then,

$$(5.2)\qquad \limsup_{n\to\infty} \frac{1}{n} \log \mu(\alpha_n^{\varepsilon}(x)) \leq - I_{\mu}(x,\alpha) + J(\varepsilon,N) \qquad \mu\text{-a.e.}x,$$

where $N=\#\alpha:=$the number of cells of α and we set

$$(5.3)\qquad J(\varepsilon,N) = \max\{\ 0,\ -\varepsilon\log\varepsilon - (1-\varepsilon)\log(1-\varepsilon) + \varepsilon\log(N-1)\ \}.$$

Remark. The number of elements contained in the ε-neighbourhood of a word is independent of the choice of the word and is equal to

(5.4) $$L_n(\varepsilon,N) = \sum_{k\leqq\varepsilon n} \frac{n!}{k!(n-k)!}(N-1)^k \qquad \text{if} \quad \#\alpha=N.$$

The increasing rate of $L_n(\varepsilon,N)$ is the quantity $J(\varepsilon,N)$:

(5.5) $$J(\varepsilon,N) = \limsup_{n\to\infty} \frac{1}{n} \log L_n(\varepsilon,N).$$

Remark to Remark. The relation (5.3) is a large deviation result. In fact, the quantity

$$N^{-n}L_n(\varepsilon,N) = \sum_{k\leqq\varepsilon n} \frac{n!}{k!(n-k)!}(1/N)^{n-k}(1-1/N)^k$$

is equal to the probability $\Pr\{\frac{1}{n}(X_1+\cdots+X_n)\leqq\varepsilon\}$ for the coin tossing with $p=1-1/N$. Hence, using the results in Sect.3,

$$\text{the RHS of (5.5)} = \max_{0\leqq t\leqq\varepsilon} H(t,1-t\,|\,1-1/N,1/N) + \log N$$

$$= \text{the RHS of (5.3)}.$$

Proof of Lemma 1. For given n, ε and a real s, set

$$p_n(\varepsilon,s) = \mu\{\, x\in X;\ \mu(\alpha_n^\varepsilon(x)) \geqq 2^{ns}\mu(\alpha_n(x))\,\}.$$

Then, appealing to the Chebyshev inequality, we obtain

$$p_n(\varepsilon,s) \leqq (2^{ns})^{-1} \int_X \mu(dx) \frac{\mu(\alpha_n^\varepsilon(x))}{\mu(\alpha_n(x))} 1[\mu(\alpha_n(x))>0]$$

$$= 2^{-ns} \sum_{c\in\alpha_n}^{*} \mu(c) \frac{1}{\mu(c)} \sum_{D_n(c',c)\leqq\varepsilon} \mu(c')$$

$$\leqq 2^{-ns}L_n(\varepsilon,N).$$

where * indicates the sum over such c that $\mu(c)>0$. Now if we take $s>J(\varepsilon,N)$, then, $\sum_n p_n(\varepsilon,s)<\infty$. Therefore the Borel-Cantelli lemma is applicable and we obtain for μ-a.e.x

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{\mu(\alpha_n^\varepsilon(x))}{\mu(\alpha_n(x))} \leqq s \qquad \text{for } s>J(\varepsilon,N).$$

On the other hand, from Theorem 2 we know the existence of the limit

$$I_\mu(x) = \lim_{n\to\infty} -\frac{1}{n}\log \mu(\alpha_n(x)).$$

Consequently, we obtain the assertion of Lemma 1.

Next let us state a trick due to A.Katok [8]. For a partition α and a positive number δ let us denote

(5.6) $\partial\alpha$ = the union of boundaries ∂c of $c\in\alpha$ and

(5.7) $\delta\text{-}\partial\alpha$ = the δ-neighbourhood of $\partial\alpha$ in X.

Lemma 2. Let $x\in X$. If

(5.6) $$\frac{1}{n}\sum_{i=0}^{n-1} 1_{\delta-\partial\alpha}(F^i x) < \varepsilon,$$

then,

(5.7) $$B_n(x,\delta) \subset \alpha_n^\varepsilon(x).$$

Proof. Take $y\in B_n(x,\delta)$. If $\alpha(F^i y)\neq\alpha(F^i x)$, then we must have $d(F^i x,\partial\alpha)<\delta$. Hence,

$$D_n(\alpha_n(y),\alpha_n(x)) = \frac{1}{n}\sum_{i=0}^{n-1} 1[\alpha(F^i y)\neq\alpha(F^i x)] \leq \frac{1}{n}\sum_{i=0}^{n-1} 1_{\delta-\partial\alpha}(F^i x) < \varepsilon.$$

Consequently, $y\in\alpha_n^\varepsilon(x)$. Hence, (5.7) holds.

The following proof of Theorem 1 is much shorter than the proof in [3] and the ideas are helpful for the proof of Theorem 0.

Proof of Theorem 1.

Let μ be an F-invariant probability measure on X and put

(5.8) $$I_\mu^* = \sup\{\, I_\mu(x,\alpha);\ \alpha \text{ is a finite Borel partition, } \mu(\partial\alpha)=0\,\},$$

where $I_\mu(x,\alpha)$ is the quantity appeared in the Shannon-McMillan-Breiman theorem (Theorem 2). It suffices to prove the following two:

Part I (Upper estimate): $\overline{h}_\mu(x) \leq I_\mu^*(x)$.

Part II (Lower estimate): $\underline{h}_\mu(x) \geq I_\mu^*(x)$.

In fact, the partition into points can be approximated by the partitions α satisfying the condition $\mu(\partial\alpha)=0$. Hence it follows from Parts I and II that

$$h_\mu(F) = \int_X \underline{h}_\mu(x)\mu(dx) = \int_X \overline{h}_\mu(x)\mu(dx).$$

Then the other assertions of Theorem 1 are immediate.

Proof of Part I. For a given $\delta>0$, take a finite Borel partition α such that $\mu(\partial\alpha)=0$ and $\text{mesh}(\alpha):=\max\{\text{diam}(c);\ c\in\alpha\} < \delta$. Then, $\alpha_n(x)$ is contained in $B_n(x,\delta)$. Thus, $\mu(\alpha_n(x))\leq\mu(B_n(x,\delta))$. Therefore,

$$I_\mu(x,\alpha) \geq \limsup_{n\to\infty} -\frac{1}{n}\log\mu(B_n(x,\delta)). \quad \text{Hence, } I^*_\mu(x)\geq\overline{h}_\mu(x).$$

Remark. The compactness assumption on X is necessary only for the existence of finite partitions α with arbitrary small mesh and is absolutely unnecessary for Part II.

Proof of Part II. Take $\delta>0$ and a finite measurable partition α such that $\mu(\partial\alpha)=0$. Then it follows from the individual ergodic theorem that the limit

$$f_\delta(x) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} 1_{\delta-\partial\alpha}(F^i x)$$

exists for μ-a.e.x. Its integral $\int_X f_\delta(x)\mu(dx)=\mu(\delta-\partial\alpha)$ tends to 0 as δ goes to 0 since $\mu(\partial\alpha)=0$. Consequently, for $\varepsilon>0$ and $\delta>0$, the measure $\mu(E)$ of the set E of points x such that

$$(5.9) \qquad \frac{1}{n}\sum_{i=0}^{n-1} 1_{\delta-\partial\alpha}(F^i x) < \varepsilon \quad \text{for every sufficiently large } n$$

can be arbitrarily close to 1 if δ is chosen so small.

Now for the point x satisfying (5.9), $B_n(x,\delta)$ is contained in $\alpha_n^\varepsilon(x)$ by virtue of Lemma 2. Thus, using Lemma 1, we obtain

$$\begin{aligned}&\liminf_{n\to\infty} -\frac{1}{n}\log\mu(B_n(x,\delta))\\ &\geq \liminf_{n\to\infty} -\frac{1}{n}\log\mu(\alpha_n^\varepsilon(x)) \geq I_m(x,\alpha) - J(\varepsilon,\#\alpha)\end{aligned}$$

on the set E. Hence, there holds the inequality independently of δ:

$$\underline{h}_\mu(x) \geqq I_\mu(x,\alpha) - J(\varepsilon,\#\alpha).$$

And now it is true for μ-almost every x. Since ε is arbitrary and $\lim_{\varepsilon\to 0} J(\varepsilon,\#\alpha) = 0$, it follows that $\underline{h}_\mu(x) \geqq I_\mu(x,\alpha)$ provided that $\mu(\partial\alpha)=0$. Consequently, $\underline{h}_\mu(x) \geqq I^*_\mu(x)$. The proof is completed.

Proof of Theorem 0.

For the sake of simplicity of the notation let us write for $x\in X$,

$$\xi_n = \xi_n^x = \frac{1}{n}\sum_{i=0}^{n-1} \delta_{F^i x}. \tag{5.10}$$

Take an arbitrary ergodic F-invariant measure μ. By the definition of the weak topology on the measure space $\underline{M}(X)$ we may take the sets of the following form as a fundamental system of neighbourhoods of the measure μ:

$$\{\ \nu\in\underline{M}(X);\ a_c<\nu(c)<b_c \quad (c\in\alpha)\ \}$$

where α is a finite Borel partition with $\mu(\partial\alpha)=0$ and a_c, b_c are non-negative real numbers such that $a_c<\mu(c)<b_c$ $(c\in\alpha)$. Now for a given neighbourhood G of μ take a finite Borel partition α, a positive number ε and nonnegative numbers a_c, b_c $(c\in\alpha)$ such that

$$\mu(\partial\alpha)=m(\partial\alpha)=0 \quad \text{and}$$

$$G \supset G(s) := \{\ \nu\in\underline{M}(X);\ a_c-s<\nu(c)<b_c+s\ \} \qquad (0\leqq s\leqq\varepsilon). \tag{5.11}$$

Note that the positive number ε may be chosen arbitrarily small. Put

$$A_n(s) = \{\ x\in X;\ \xi_n\in G(s)\ \} \qquad (0\leqq s\leqq\varepsilon). \tag{5.12}$$

Claim: if $x\in A_n(0)$, then, $\alpha_n^\varepsilon(x) \quad A_n(\varepsilon)$.

In fact, take $y\in\alpha_n^\varepsilon(x)$. Then,

$$\xi_n^y(c) = \frac{1}{n}\sum_{i=0}^{n-1} 1_c[F^i y]$$

$$\geqq \frac{1}{n}\sum_{i=0}^{n-1} 1_c[F^i x] - \frac{1}{n}\sum_{i=0}^{n-1} 1[a(F^i y)\neq\alpha(F^i x)] > a_c - \varepsilon.$$

Similarly, $\xi_n^y(c) < b_c + \varepsilon$. Consequently, $y \in A_n(\varepsilon)$.

Now it follows from (5.11), (5.12) and the claim above that

$$\mathrm{vol}\{\, x \in X;\ \xi_n \in G \,\} \geq \mathrm{vol}(A_n(\varepsilon))$$

$$\geq \sum_{c \in \alpha_n,\ c \subset A_n(\varepsilon)} \mathrm{vol}(c^{\varepsilon})/L_n(\varepsilon,\#\alpha) \tag{5.13}$$

$$\geq L_n(\varepsilon,\#\alpha)^{-1} \int_{\alpha_n(x)>0} \mu(dx) \frac{\mathrm{vol}(\alpha_n^{\varepsilon}(x))}{\mu(\alpha_n(x))} .$$

Now, using Theorem 2 (Shannon-McMillan-Breiman) and Theorem 3 (Oseledec), we obtain the following inequalities on a set E with $\mu(E) > 1-\eta$ for any given $\eta>0$ and $\theta>0$: for any sufficiently large n,

$$\log \mathrm{vol}(\alpha_n^{\varepsilon}(x)) \geq -n(\chi_\mu^+ + \theta) \quad \text{and} \tag{5.14}$$

$$-\infty < \log \mu(\alpha_n(x)) \leq n(-h_\mu(F,\alpha) + \theta). \tag{5.15}$$

Here we used the ergodicity of μ.

Combining (5.13), (5.14) and (5.15) we obtain

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathrm{vol}\{\, x \in X;\ \xi_n \in G \,\}$$

$$\geq \liminf_{n\to\infty} \{ -\frac{1}{n} \log L_n(\varepsilon,\#\alpha) + h_\mu(F,\alpha) - \chi_\mu^+ - 2\theta + \frac{1}{n} \log \mu(E) \}$$

$$= h_\mu(F,\alpha) - \chi_\mu^+ - 2\theta - J(\varepsilon,\#\alpha).$$

Consequently, taking the limits as $\theta\to 0$ and then as $\varepsilon \to 0$ and the supremum with respect to partition α, we obtain the inequality

$$\underline{Q}(G) \geq h_\mu(F) - \chi_\mu^+$$

for every neighbourhood G of an ergodic F-invariant measure μ. Hence follows the desired inequality $\underline{q}(\mu) \geq h_\mu(F) - \chi_\mu^+$ and the proof of Theorem 0 is completed.

18

References

[1] W.Ballmann, Structure of manifolds of nonpositive sectional curvature, Proc. Taniguchi Symp. on Curvature and Topology of Riemannian Manifolds, Katata 1985, 1-13, Springer Lecture Notes in Math. 1201 (1985).

[2] R.Bowen, Equilibrium States and the Ergodic Theory of Anosov diffeomorphisms, Springer Lecture Notes in Math. 470(1975).

[3] M.Brin and A.Katok, On local entropy, Geometric Dynamics, Springer Lecture Notes in Math. 1007(1983), 30-38.

[4] I.P.Cornfeld, S.V.Fomin and Ya.G.Sinai, Ergodic Theory, Nauka; English translation, Springer, 1980.

[5] H.Cramer, On a new limit theorem in the theory of probability, Colloquium on the Theory of Probability, Herman, Paris, 1937.

[6] M.Donsker and S.R.S.Varadhan, Asymptotic evaluation of certain Markov process expectations for large time I, IV, Comm. Pure Appl. Math. 28(1975), 1-45, 36(1983), 183-212.

[7] S.Ito and Y.Takahashi, Markov shifts and realization of β-transformations, J.Math.Soc.Japan 26(1974),33-55.

[8] A.Katok, Lyapunov exponents, entropy and periodic points for diffeomorphisms, I.H.E.S. Publ. Math. 51(1980), 137-174.

[9] F.Ledrappier and L.C.Young, The metric entropy of diffeomorphisms, Bull. A.M.S. 11(1984), 343-346.

[10] R.Mane, Ergodic Theory and Dynamical Systems, Springer 1987.

[11] A.Manning, A relation between Lyapunov exponents, Hausdorff dimension and entropy, Erg. Theory and Dynam. Sys. 1(1981),431-459.

[12] Y.Oono and Y.Takahashi, Chaos, external noise and Fredholm theory, Prog. Theor. Physics 63(1980), 1804-1807.

[13] V.I.Oseledec, Multiplicative ergodic theorem. Liapunov numbers for dynamucal systems, Trudy Moskov. Obsc. 19(1968), 179-

210.

[14] R.Osserman and P.Sarnack, A new curvature invariant and entropy of geodesic flows, Invent. math. 77(1984), 455-462.

[15] J.Pesin, Characteristic exponents and smooth ergodic theory, Uspehi Mat. Nauk 22(1977), 55-112.

[16] R.Phelps, Lectures on Choquet's Theorem, van Nostrand, 1966.

[17] A.W.Roberts and D.E.Varberg, Convex Functions, Academic Press, 1973.

[18] Y.Takahashi, An ergodic-theoretical approach to the chaotic behaviours of dynamical systems, Publ. R.I.M.S. Kyoto Univ. 19 (1983), 1256-1282

[19] Y.Takahashi, Observable chaos and variational formalism for one dimensional maps, Proc. 4th Japan-USSR Symp. on Probability Theory and Mathematical Statistics, Springer Lecture Notes in Math. 1021 (1983), 676-686.

[20] Y.Takahashi, Entropy functional (free energy) for dynamical systems and their random perturbations, Proc.Taniguchi Intern. Symp. on Stochastic Analysis, Katata and Kyoto 1982, 437-467, Kinokuniya/North Holland, 1984.

[21] Y.Takahashi, One dimensional maps and power spectrum, Recent Studies on Turbulent Phenomena, 1985, eds. T.Tatsumi et al., 99-116, Assoc. Sci. Documents Inform. (学術文献普及会), Tokyo, 1985.

[22] Y.Takahashi, Two aspects of large deviation theory for large time, Proc. Taniguchi Symp. on Probabilistic Methods in Mathematical Physics, Katata and Kyoto 1985, 363-384, Kinokuniya/Academic Press, 1987.

[23] Y.Takahashi, Long time asymptotics of the ratio of measures of small tubes and a large deviation result, to appear in Proc. 5th Japan-USSR Symp. on Probability Theory, Springer Lecture Notes in Math.

[24] H.Totoki, Introduction to Ergodic Theory, Kyoritsu, Tokyo 1972 (in Japanese).

[25] S.R.S.Varadhan, Large Deviations and Applications, CBMS-NSF Regional Conference Series in Applied Mathematics 46, Society for Industrial and Applied Mathematics, 1984.

[26] L.C.Young, Dimension, entropy and Lyapunov exponents, Erg. Theory Dynam. Sys. 2(1982), 109-124.