Štefan Šujan

On a characteristic property of the asymptotic rate

Kybernetika, Vol. 14 (1978), No. 4, (285)--291

Persistent URL: http://dml.cz/dmlcz/124442

KYBERNETIKA — VOLUME 14 (1978), NUMBER 4

On a Characteristic Property of the Asymptotic Rate

ŠTEFAN ŠUJAN

The asymptotic rate [4] is shown to be the only effective element of the family of all admissible measures of uncertainty for the set of all discrete stationary information sources.

1. INTRODUCTION

The concepts of sufficiency, regularity of conditional probabilities, and ergodicity are strongly related (cf. [1], and [3] in a more general setting). In the present note a new justification of the asymptotic rate [4] is given by means of these interrelations. The reasoning is independent on the information-theoretical concepts (such as communnication channels, transmission rate, capacity, etc.). Thus the results complete in a sense the program of [5]. According to [5] all the relevant properties of the asymptotic rate have to be established within the framework of the ergodic theory only, i.e. without calling attention to the information-theoretical concepts mentioned above.

2. SUFFICIENCY AND WEAK CONVERGENCE OF SAMPLE DISTRIBUTIONS

We shall follow the notations of [4] and [5]. Let $\mathcal{M}$ denote the family of all stationary sources discrete in time (the time being represented by the set I of all integers) and having a countably infinite alphabet (say, the alphabet $N = \{1, 2, \ldots\}$). Hence the space of all messages (or, the sample space) will be identified with the set

$$N^I = \{z: z = \{z_i\}, z_i \in N \quad \text{for} \quad i \in I\} . \tag{2.1}$$

Let $\mathbf{x} \in N^n$ ($n \in N$), let $i \in I$, respectively. The sets of the form

$$[\mathbf{x}]_{i,n} = \{z: z \in N^I, (z_i, \ldots, z_{i+n-1}) = \mathbf{x}\} \tag{2.2}$$

are said to be the elementary cylinders. Let $\mathscr{F}$ denote the product σ-algebra in N^I (we consider the family of all subsets as the σ-algebra in N). As well-known, the countably infinite family $\mathscr{P}$ of all elementary cylinders generates $\mathscr{F}$; in symbols $\mathscr{F} = \sigma(\mathscr{P})$. Let T denote the shift-transformation in N^I, i.e.

$$(2.3) \qquad (Tz)_i = z_{i+1} \quad \text{for} \quad z \in N^I,\ i \in I\,.$$

In what follows we shall identify the family $\mathscr{M}$ with the convex family of all probability measures μ on $(N^I, \mathscr{F})$ such that

$$(2.4) \qquad \mu(E) = \mu(T^{-1}E)\,, \quad E \in \mathscr{F}\,.$$

A measure μ satisfying (2.4) is called T-invariant. Let

$$(2.5) \qquad \mathscr{F}_0 = \{E\colon E \in \mathscr{F},\ T^{-1}E = E\}$$

be the subσ-algebra of $\mathscr{F}$ consisting of all T-invariant measurable sets. To finish the preliminaries let us introduce the following notations:

a. e. $[\mu]$ – for all $z \in N^I$ except a measurable set A with $\mu(A) = 0$,
a. e. $[\mathscr{M}]$ – for all $z \in N^I$ except a measurable set A with $\mu(A) = 0$ for all $\mu \in \mathscr{M}$,
$L^1(N^I, \mathscr{G}, \mathscr{M})$ – the space of all real-valued $\mathscr{G}$-measurable functions integrable with respect to every $\mu \in \mathscr{M}$,
$\mathscr{G} = \mathscr{H}[\mathscr{M}]$ – given any set $G \in \mathscr{G}$ there is a set $H \in \mathscr{H}$ such that for the symmetric difference $G \,\Delta\, H$ we have $\mu(G \,\Delta\, H) = 0$ for all $\mu \in \mathscr{M}$.

Proposition 1. The σ-algebra $\mathscr{F}_0$ (cf. (2.5)) is sufficient with respect to the family $\mathscr{M}$.

Proof. We have to prove that given any function $f \in L^1(N^I, \mathscr{F}, \mathscr{M})$ we can choose the versions $E_\mu\{f \,\|\, \mathscr{F}_0\}$ of the conditional expectations independently on $\mu \in \mathscr{M}$. The individual ergodic theorem gives

$$(2.6) \qquad E_\mu\{f \,\|\, \mathscr{F}_0\}\,(z) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j z) \quad \text{a. e. } [\mu]$$

for all $\mu \in \mathscr{M}$. For fixed f, the right-hand side of (2.6) does not depend on $\mu \in \mathscr{M}$. Hence the sufficiency follows.

Remark. It is possible to prove a stronger result. Actually, a. e. $[\mu]$ can be replaced in (2.6) by a. e. $[\mathscr{M}]$ and, moreover, by a. e. $[\mathscr{M}]$ simultaneously for all $f \in L^1(N^I, \mathscr{F}, \mathscr{M})$. The proof depends on the topological properties of the space N^I ([4], p. 810) and the resulting topological properties of the space $\mathscr{M}$ (cf. e.g. [2], especially Section II. 6).

The family $\mathscr{M}$ contains, in general, an uncountable infinity of pairwise singular probability measures. Consequently, it is not dominated. It follows that there is no single sufficient statistics $Y: N^I \to R^1$ such that

$$Y^{-1}\mathscr{B}_1 = \mathscr{F}_0[\mathscr{M}] .$$

Here, $\mathscr{B}_1$ designates the Borel σ-algebra in R^1. On the other hand, we shall obtain another useful characterization of $\mathscr{F}_0$.

A measure $\mu \in \mathscr{M}$ is said to be ergodic (in symbols, $\mu \in \mathscr{M}_*$) provided $\mu(E) \in \{0, 1\}$ for all $E \in \mathscr{F}_0$. The family $\mathscr{M}_*$ can be parametrized by the set R of all regular points N^I ([4], p. 808). Recall that $z \in R$ iff there is $\mu_z \in \mathscr{M}_*$ uniquely determined by the point z via the relations

$$\mu_z(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_A(T^j z)\,, \quad A \in \mathscr{P} \tag{2.7}$$

(χ_A denotes the indicator function of the set A). Then $R \in \mathscr{F}_0$ and $\mu(R) = 1$ for all $\mu \in \mathscr{M}$ ([4], p. 809). Consequently,

$$\mathscr{M}_* = \{\mu_z\colon z \in R\} . \tag{2.8}$$

According to (2.7) the one-parameter family $\mathscr{M}_*$ is $\mathscr{F}_0$-measurable, i.e.

(a) for every fixed $z \in R$, $\mu_z(\cdot)$ is a probability measure on $(N^I, \mathscr{F})$;

(b) for every fixed $E \in \mathscr{F}$, $\mu_{(\cdot)}(E)$ is $\mathscr{F}_0$-measurable (only on R, but since $\mu(R) = 1$ for all $\mu \in \mathscr{M}$, we have $\mathscr{F}_0 = R \cap \mathscr{F}_0[\mathscr{M}]$).

Note that the parametrization by means of the set R is not identifiable, i.e. there may be many points $z \in R$ yielding the same measure $\mu \in \mathscr{M}_*$. It is possible (of course, only by means of the axiom of choice) to get an identifiable set R', however, we need not this property.

Proposition 2. The σ-algebra $\mathscr{F}_0$ is the least σ-algebra such that the one-parameter family $\{\mu_z\colon z \in R\}$ is measurable.

This means that if $\{\mu_z\colon z \in R\}$ is $\mathscr{G}$-measurable then either $\mathscr{G} \supset \mathscr{F}_0$ or $\mathscr{G} = \mathscr{F}_0[\mathscr{M}]$.

Proof. Assume $\mathscr{G}$ is such that there are a set $F_0 \in \mathscr{F}_0 - \mathscr{G}$ and a measure $\mu \in \mathscr{M}$ with $\mu(F_0) > 0$. Let $\mathscr{B}[0, 1]$ denote the Borel σ-algebra of the unit interval $[0, 1]$. Given F_0, the function $f_0(z) = \mu_z(F_0)$ maps R into $[0, 1]$. By the assumption $f_0^{-1}\mathscr{B}[0, 1] \subset \mathscr{G}$. Let us consider the set $\{1\} \in \mathscr{B}[0, 1]$. Then $f_0^{-1}\{1\} = R \cap F_0 = F_0[\mathscr{M}]$. Consequently, $f_0^{-1}\{1\} \notin \mathscr{G}$, a contradiction.

Let $\lambda_n(\cdot, z)$ assign the mass $1/n$ to every of the points $z, Tz, \ldots, T^{n-1}z$. If $\mu \in \mathscr{M}$ is the product measure (i.e. a memoryless source) then the weak limit of the sequence $\{\lambda_n(\cdot, z)\}_{n\in N}$ of the sample distributions a.e. $[\mu]$ equals the "true" measure μ; in symbols

$$\mu\{z\colon z \in N^I,\ \lambda_n(\cdot, z) \Rightarrow \mu(\cdot)\} = 1$$

(cf. [2], Sect. II.7). In the general case start with (2.6). Since N^I is a complete separable metric space in its product topology ([4], p. 810) there are regular versions of the conditional probabilities $P_\mu\{\cdot \parallel \mathscr{F}_0\}$ (cf. [2], Sect. V. 8), especially

$$E_\mu\{f \parallel \mathscr{F}_0\}(z) = \int f(y)\, P_\mu\{dy \parallel \mathscr{F}_0\}(z) \quad \text{a. e. } [\mu] \tag{2.9}$$

for all $\mu \in \mathscr{M}$, $f \in L^1(N^I, \mathscr{F}, \mathscr{M})$. According to the definition of the sample distributions, both (2.6) and (2.8) imply

$$\lim_{n\to\infty} \int f(y)\, \lambda_n(dy, z) = \int f(y)\, P_\mu\{dy \parallel \mathscr{F}_0\}(z) \quad \text{a. e. } [\mu]\,. \tag{2.10}$$

Using (2.7), (2.6), and (2.9) one easily concludes

$$\mu\{z\colon z \in R,\, P_\mu\{\cdot \parallel \mathscr{F}_0\}(z) = \mu_z(\cdot)\} = 1\,, \quad \mu \in \mathscr{M}\,;$$

hence

$$\mu\{z\colon z \in R,\, \lambda_n(\cdot, z) \Rightarrow \mu_z(\cdot)\} = 1 \tag{2.11}$$

for all $\mu \in \mathscr{M}$ (cf. also Theorem VI. 9.1 in [2]). Rewriting (2.11) we obtain the following

Proposition 3. The family of all weak limits of the sample distributions coincides with the $\mathscr{F}_0$-measurable one-parameter family $\{\mu_z\colon z \in R\}$ of all ergodic measures.

Remark. It follows from the above proposition that it is unpossible to recover by means of a sample path the measure $\mu \in \mathscr{M}$ unless μ is ergodic. Hence in order to construct some reasonable testing and parameter estimation procedures for stationary processes, one is forced to implement also some additional constraints upon the underlying stationary process. The author hopes to deal with the related problems in a separate paper.

3. ADMISSIBLE AND EFFECTIVE MEASURES OF UNCERTAINTY

Let $V(\cdot)\colon \mathscr{M} \to (0, \infty]$ be assumed as a candidate for a measure of uncertainty. Given $\mu \in \mathscr{M}$, let $I(z; \mu)$ denote the amount of information provided by the sample path z. Let $P(\mu, z)$ denote the posterior distribution given μ and the sample path z. According to the usual interpretation of uncertainty and information we shall require that the information provided by a sample path equals the difference between the prior and the posterior uncertainties, respectively. In symbols,

$$I(z; \mu) = V(\mu) - V(P(\mu, z))\,. \tag{3.1}$$

In spite of (2.11),

$$I(z; \mu) = V(\mu) - V(\mu_z)\,[\mu] \tag{3.2}$$

for all $\mu \in \mathcal{M}$. In information theory, the statistical properties of the information sources are assumed to be known. Since they are described by the corresponding probability measure μ on the sample space N^I, all the relevant information before sampling is given by the prior distribution μ. If $\mu \in \mathcal{M}_*$, then

$$\mu\{z: z \in R, \mu_z = \mu\} = 1 \tag{3.3}$$

([4], p. 810). In this case, the posterior distribution coincides with the prior one almost everywhere. Consequently, sampling cannot provide us with an additional information. This fact is conform with the intuitive meaning of sufficiency, since for $\mu \in \mathcal{M}_*$, the prior information is measurable with respect to a sufficient σ-algebra. On the other hand, it is reasonable to take as a measure of uncertainty in the ergodic case the "least possible" one. This means that a reasonable measure of uncertainty should be $\mathcal{F}_0$-measurable. The entropy rate $H(\cdot)$ (cf. [5], Lemma 5 and (1.8)) has the required property. Thus, our first requirement is

$$V(\mu_z) = H(\mu_z) \quad \text{for all} \quad z \in R\,. \tag{3.4}$$

On the other hand, if the prior information is not $\mathcal{F}_0$-measurable then the sampling can provide us with some additional information. Thus, we require

$$\mu\{z: z \in R, I(z; \mu) \geqq 0\} = 1\,; \quad \mu \in \mathcal{M}\,,$$

i.e.

$$\mu\{z: z \in R, V(\mu) - H(\mu_z) \geqq 0\} = 1\,; \quad \mu \in \mathcal{M}\,. \tag{3.5}$$

Definition 1. A measure $V(\cdot)$: $\mathcal{M} \to [0, \infty]$ of uncertainty is said to be *admissible for the family* $\mathcal{M}$ provided (3.4) and (3.5) take place. In (3.5), we addopt the convention $\infty - \infty = 0$.

If $V(\cdot)$ is admissible, define $V'(\cdot)$ by the properties

$$V'(\mu) = V(\mu) \qquad \text{for all} \quad \mu \in \mathcal{M}_*\,,$$
$$V'(\mu) = V(\mu) + 5 \quad \text{for all} \quad \mu \in \mathcal{M} - \mathcal{M}_*\,.$$

Clearly, $V'(\cdot)$ is again admissible. In order to avoid such pathological situations, we introduce the following

Definition 2. A measure $V_0(\cdot)$: $\mathcal{M} \to [0, \infty]$ of uncertainty is said to be *effective* provided

(1) $V_0(\cdot)$ is admissible for the family $\mathcal{M}$;

(2) for any other admissible $V(\cdot)$ we have the inequalities

$$V_0(\mu) \leqq V(\mu), \quad \mu \in \mathscr{M}.$$

Proposition 4. The entropy rate $H(\cdot)$ is not an admissible measure of uncertainty for the family $\mathscr{M}$ of all stationary information sources.

Theorem. The only effective measure of uncertainty is the measure $V_0(\cdot)$ defined by the relations

$$V_0(\mu) = \operatorname*{ess.\,sup}_{z \in R[\mu]} H(\mu_z); \quad \mu \in \mathscr{M},$$

i.e. the asymptotic rate.

Due to (3.5) the above statements are almost self-explanatory. For the sake of completeness the proofs are presented below.

Proof of Proposition 4. Let $\mu \in \mathscr{M}$. Then

$$H(\mu) = \int_R H(\mu_z)\, \mu(\mathrm{d}z)$$

(cf. e.g. [5], Lemma 5 and (1.9)). Choose $\mu \in \mathscr{M}$ such that

$$\mu\{z\colon z \in R,\ H(\mu) \neq H(\mu_z)\} > 0.$$

It is well-known that such sources μ always exist. The exceptional ones are the so-called strongly stable sources (cf. [4], esp. Sect. 11). Then we can find a set $R_1 \subset R$ such that

$$\mu(R_1) > 0,$$
$$H(\mu_z) > H(\mu), \quad z \in R_1,$$

a contradiction with (3.5).

Proof of the Theorem. Let $\mu \in \mathscr{M}_*$. Then

$$\mu\{z\colon z \in R,\ \mu_z = \mu\} = 1,$$

hence $V_0(\mu) = H(\mu)$. Clearly

$$\mu\{z\colon z \in R,\ V_0(\mu) - H(\mu_z) \geqq 0\} = 1.$$

Let $V(\cdot)$ be any other admissible measure of uncertainty. Let $\mu \in \mathscr{M}$ be such that

$$V_0(\mu) > V(\mu).$$

Since

$$\operatorname*{ess.\,sup}_{z \in R[\mu]} H(\mu_z) = \inf\{t\colon \mu\{z\colon z \in R,\ H(\mu_z) \leqq t\} = 1\},$$

it follows that

$$\mu\{z: z \in R,\ H(\mu_z) > V(\mu)\} > 0 .$$

The latter relation contradicts (3.5) for $V(\cdot)$.

(Received March 1, 1977.)

REFERENCES

[1] R. H. Farrell: Representation of invariant measures. Illinois J. of Math. *6* (1962), 447–467.

[2] K. R. Parthasarathy: Probability Measures on Metric Spaces. Academic Press, New York 1967.

[3] Š. Šujan: Integral representations of probability measures. Submitted to Z. Wahrscheinlichkeitstheorie verw. Gebiete.

[4] K. Winkelbauer: On discrete information sources. In: Transactions of the 3rd Prague Conf. on Inf. Theory etc., Prague 1964, 765–830.

[5] K. Winkelbauer: On the asymptotic rate of non-ergodic information sources. Kybernetika *6* (1970), 127–147.

RNDr. Štefan Šujan, Ústav merania a meracej techniky SAV (Institute of Measurement and Measuring Technic — Slovak Academy of Sciences), 885 27 Bratislava. Czechoslovakia.