Kybernetika

A Generalized Coding Problem for Discrete Information Sources

ŠTEFAN ŠUJAN

ACADEMIA
PRAHA

A general coding problem is formulated. The special cases are the coding problems of the distortionless coding theory, of source coding with a fidelity criterion, and of the source coding with side information, respectively. All sources examined in the paper are assumed to be discrete in time and stationary. Also the sources, the statistical properties of which are described by finitely additive probabilities, are admitted. The paper is devoted mainly to the problems of the distortionless coding. Further a generalization for the pairs of information sources is given. The corresponding coding theorems are established and the important properties of the resulting quantities are studied. Some applications are given concerning statistical problems of the random processes.

REFERENCES

[1] R. Ahlswede, P. Gács, J. Körner: Bounds on conditional probabilities with applications in multi-user communication. Z. Wahrscheinlichkeitstheorie verw. Gebiete *34* (1976), 157—177.

[2] R. Ash: Information theory. Interscience Publishers, New York—London—Sydney 1965.

[3] R. R. Bahadur: Some limit theorems in statistics. Regional Conference Series in Applied Mathematics 4, SIAM, Philadelphia 1971.

[4] R. R. Bahadur and M. Raghavachari: Some asymptotic properties of likelihood ratios on general sample spaces. Proc. Sixth Berkeley Symp. Math. Stat. Prob., Vol. 1 (1970), 129—152.

[5] J. R. Blum, D. L. Hanson: On invariant probability measures I. Pacific J. of Math. *10* (1960), 4, 1125—1130.

[6] D. L. Brown: Non-local optimality of appropriate LRT's. Ann. Math. Statistics *42*, (1971), 1206—1240.

[7] K. L. Chung: Markov chains with stationary transition probabilities. Second ed. Springer-Verlag, Berlin—Göttingen—Heidelberg 1967.

[8] N. Dunford, J. T. Schwartz: Linear operators Part I: General theory. Interscience Publishers, New York 1958.

[9] R. R. Farrell: Representation of invariant measures. Illionois J. of Math. *6* (1962), 447—467.

[10] A. Feinstein: Foundations of information theory. McGraw-Hill Book Co., New York 1958.

[11] R. Gray, L. Davisson: The ergodic decomposition of stationary discrete random processes. IEEE *IT-20* (1974), 5, 625—636.

[12] R. Gray, L. Davisson: Source coding theorem without the ergodic assumption. IEEE *IT-20* (1974), 4, 502—516.

[13] P. R. Halmos: Measure theory. D. van Nostrand, New York 1950.

[14] E. Hewitt, K. Yosida: Finitely additive measures. Transactions of the American Math. Society *72* (1952), 46—66.

[15] S. Horowitz: Transition probabilities and contractions of L_∞. Z. Wahrscheinlichkeitstheorie verw. Gebiete *24* (1972), 263—274.

[16] K. Jacobs: Lectures in ergodic theory, Vol. I, II. Aarhus Universitet, Mathematisk Institut 1962/1963.

[17] J. L. Kelley: General Topology. 9th ed. D. van Nostrand, Princeton N. J. 1968.

[18] D. F. Kerridge: Inaccuracy and inference. Journal of the Royal Stat. Society, Ser. B *23* (1961), 184—194.

[19] A. N. Kolmogorov: New metric invariant of transitive dynamical systems and automorphisms of Lebesgue spaces. (In Russian), DAN SSSR *119* (1958), 5, 861—864.
[20] N. Kryloff, N. Bogoliouboff: La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire. Ann. of Math. *38* (1937), 65—113.
[21] S. Kullback: Information theory and statistics. Wiley, New York 1958.
[22] K. Marton: Error exponent for source coding with a fidelity criterion. IEEE *IT-20* (1974), 2, 197—199.
[23] B. McMillan: The basic theorems of information theory. Ann. Math. Stat. *24* (1953), 196 to 219.
[24] J. F. Mertens: Intégration des mesures non dénombrablement additives: une généralisation du lemme de Fatou et du théorème de convergence de Lebesgue. Annales de la Société Scientifique de Bruxelles *84*, II (1970), 231—239.
[25] R. R. Olshen: Representing finitely additive invariant probabilities. Ann. Math. Stat. *39* (1968), 2131—2135.
[26] J. C. Oxtoby: Ergodic sets. Bull. Amer. Math. Society *58* (1952), 116—136.
[27] K. R. Parthasarathy: On the integral representation of the rate of transmission of a stationary channel. Illinois J. of Math. *5* (1961), 2, 299—305.
[28] K. R. Parthasarathy: A note on McMillan's theorem for countable alphabets. Transactions of the 3rd Prague Conf. on Inform. Theory etc., Prague 1964, 541—543.
[29] D. Pötschke: A statistical interpretation of the B-rate of information theory. Presented at the 1974 European Meeting of Statisticians and Seventh Prague conf. on Inform. Theory etc., Prague 1974.
[30] R. Phelps: Lectures on Chocquet's theorem. D. van Nostrand, Princeton, N. J., 1966.
[31] H. Rasiowa and R. Sikorski: Mathematics of metamathematics. PWN, Warszawa 1963.
[32] V. A. Rochlin: New progress in the theory of transformations with invariant measure. (In Rusian). Usp. Mat. Nauk *15* (1960), 3—26.
[33] C. P. Schnorr: Zufälligkeit und Wahrscheinlichkeit. Lect. Notes in Math. 218. Springer-Verlag, Berlin—Heidelberg—New York 1971.
[34] C. E. Shannon: A mathematical theory of communication. Bell Sys. Techn. J. *27* (1948), 379—432, 623—656.
[35] Ja. G. Sinaj: On the notion of entropy of a dynamical system. (In Russian). DAN SSSR *124* (1959), 4, 768—771.
[36] Ja. G. Sinaj: On flows with finite entropy. (In Russian). DAN SSSR *125* (1959), 6, 1200 to 1202.
[37] Š. Šujan: On the integral representation of the entropy rate. To appear in Studia Sci. Math. Hungar.
[38] Š. Šujan: On the asymptotic B-rate. Submitted to Studia Sci. Math. Hungar.
[39] G. Tusnády: On asymptotically optimal testss. To appear in Annals of Statistics.
[40] K. Winkelbauer: On discrete information sources. Transactions of the 3rd Prague Conf. on Inform. Theory etc., Prague 1964, 765—830.
[41] K. Winkelbauer: On the asymptotic rate of non-ergodic information sources. Kybernetika *6* (1970), 2, 127—148.
[42] J. Wolfowitz: Coding theorems of information theory. Second ed. Springer-Verlag, New York 1964.

INTRODUCTION

Practically all information-theoretical quantities can be derived from a properly chosen set of postulates. However, to obtain a reasonable interpretation of these quantities based on the set of the postulates needs, in general, very sophisticated arguments. A natural interpretation is provided by the coding theorems of information theory. The coding theorems constitute a bridge connecting these quantities with the practical problems concerning an optimal characterization of a sequence of letters randomly chosen from a given alphabet.

The paper consists of five parts. The first part deals with a general formulation of the coding problem. We are starting with four examples motivating our approach. The first two are taken as the distortionless coding problems (cf. [34] and [40]). The third example concerns the rate distortion function, i.e. the source coding with a fidelity criterion [22]. The last one deals with the source coding with side information [1]. All coding problems mentioned in the examples are shown to be the special cases of a general coding problem, which is formulated in Section 2.

In the first part the arguments are given for the existence of the information sources, the statistical properties of which are described by finitely additive probabilities. The ergodic theory of finitely additive invariant probabilities is developed in the second part of the paper. Actually, only the ergodic decomposition theorem will be used in the subsequent sections. But the structure of ergodic finitely additive measures seems to be of the separate interest, too.

The third part deals with the distortionless coding problem for discrete in time stationary information sources with a general alphabet. The results extend those obtained in [40] and [41].

The fourth part of the paper is devoted to the information-theoretical quantities defined for the pairs of information sources. The coding theorems are established. The resulting quantities generalize the notions of inaccuracy [18] and of I-divergence [21], respectively.

The methods used throughout the paper do not exceed the frame of the ergodic theory of invariant set functions. Therefore in the last part, devoted to the applications, we are dealing only with such problems, the solutions of which are obtainable within the framework used in the first four parts of the paper.

The paper is finished by an Appendix. In the appendix, we investigate another method for proving the main results concerning the asymptotic rate. The method provides another natural interpretation of the quantities introduced in the third part.

PART I: PRELIMINARIES AND THE GENERAL CODING PROBLEM

1. Basic Notations and Terminology

The following notations will be used throughout the whole paper. Let A denote an arbitrary set. The symbol $\mathfrak{P}(A)$ will be used to denote the family of all subsets of the set A. The symbol χ_A will designate the indicator function of the set A. For finite A only, card (A) will mean the number of elements in A. If B denotes another set, the symbol A^B will be used to designate the set of all mappings which map the set B into the set A.

The basic space of all possible messages will be represented by the set X^I of all doubly-infinite sequences $z = \{z_j\}_{j=-\infty}^{j=\infty}$ of letters (i.e. elements of) in X. The set X, called the alphabet, will be a separable metric space. The set I of all integers represents the discrete time. If X is a finite set and card $(X) = n$, we shall identify X with the set $\{1, 2, \ldots, n\}$. If X is countably infinite, it will be represented by the set N of all positive integers. The symbol $\mathscr{B}(X)$ will denote the σ-field of Borel sets in X. Let us note that $\mathscr{B}(X) = \mathfrak{P}(X)$ for at most a countable alphabet X.

The σ-field in X^I will be the usual product σ-field. This means it is generated by the field $\mathscr{A}_X$ of all finite-dimensional cylinders in X^I. A *finite-dimensional cylinder* is, by definition, any set of the form

$$[E]_J = \{z : z \in X^I, \{z_j\}_{j \in J} \in E\} \tag{1.1}$$

with $J \subset I$ and $0 < \text{card}\,(J) < \infty$. Here, the set E is a Borel measurable subset of the space X^J, in symbols $E \in \mathscr{B}(X^J)$. Clearly, for an at most countable set X, $\mathscr{B}(X^J) = \mathfrak{P}(X^J)$. In accordance with (1.1) we shall use the following notations:

$$[E]_J = [E]_{i,n} \quad \text{if} \quad J = \{i, \ldots, i + n - 1\}, \tag{1.2}$$

$$[E]_J = [E] \quad \text{if} \quad J = \{0, \ldots, n - 1\}, \tag{1.3}$$

$$[\{\bar{x}\}]_J = [\bar{x}]_J \quad \text{for} \quad \bar{x} \in X^J. \tag{1.4}$$

Further, let us set

$$\mathscr{P}_X^{i,n} = \{[\bar{x}]_{i,n} : \bar{x} \in X^n\}. \tag{1.5}$$

If X is at most countable, the family $\mathscr{P}_X^{i,n}$ is a countable partition of X^I for any $i \in I$ and $n \in N$. However, this fails to be true in the general case. The members of the family

$$\mathscr{P}_X = \cup \{\mathscr{P}_X^{i,n} : i \in I, n \in N\} \tag{1.6}$$

are said to be *elementary* cylinders. For a countable alphabet X the family $\mathscr{P}_X$ generates the product σ-field as does the field $\mathscr{A}_X$; in symbols

$$\sigma(\mathscr{P}_X) = \sigma(\mathscr{A}_X)\,. \tag{1.7}$$

In any case the σ-field $\sigma(\mathscr{A}_X)$ will be denoted by $\mathscr{F}_X$. To simplify the notations the following conventions are addopted. Let us have a mathematical object $\mathscr{O}$ related with the alphabet X. If $X = \{1, 2, \ldots, n\}$, we shall write $\mathscr{O}_X = \mathscr{O}_n$. If $X = N$, the subscripts are omitted. E.g. $\mathscr{A}_n$, $\mathscr{F}_n$, $\mathscr{A}$, etc.

All stationarity properties are defined relative to the *coordinate-shift transformation* T_X of the space X^I. The latter is defined by the property that

$$(T_X z)_i = z_{i+1} \quad \text{for} \quad z \in X^I,\ i \in I\,. \tag{1.8}$$

A measure μ defined on the σ-field $\mathscr{F}_X$ is called T_X-*invariant* (or shift-invariant) provided $\mu(T_X^{-1}E) = \mu(E)$ holds true for any set $E \in \mathscr{F}_X$. The latter fact will be symbolically denoted by $\mu = \mu T_X^{-1}$ (cf. e.g. [13]).

The product toplogy in the space X^I is always metrizable and yields a separable metric space [17]. For at most countable X, the corresponding distance function can be given by the formula

$$\varrho(z, z') = \begin{cases} \max\left\{\dfrac{1}{1+|i|} : z_i \neq z'_i,\quad i \in I\right\}, & \text{if} \quad z \neq z'\,, \\ 0, \quad \text{if} \quad z = z';\ z, z' \in X^I\,. \end{cases} \tag{1.9}$$

The equality $\mathscr{F}_X = \mathscr{B}(X^I)$ is an immediate corollary to the definition of the product topology. If $\operatorname{card}(X) < \infty$, the metric space (X^I, ϱ) is compact by Tichonov's theorem. In any case the transformation T_X is a homeomorphism of the space X^I onto itself. Some more information concerning the topological properties of the space X^I will be given in Section 9.

The ergodic theory of finitely additive probabilities will differ from the σ-additive theory. The basic notions of the σ-additive theory are the notions of quasiregularity and of regularity, respectively. As far as concerns the notion of the quasiregularity, we shall addopt for countable alphabets the definition of Winkelbauer [40]. A point $z \in X^I$ is said to be *quasiregular* provided there are the limits

$$\mu_z(A) = \lim_{n\to\infty} n^{-1} \sum_{j=0}^{n-1} \chi_A(T_X^j z)\,, \quad A \in \mathscr{P}_X\,. \tag{1.10}$$

A quasiregular point $z \in X^I$ is called *regular* if there is an ergodic probability measure μ_z on the σ-field $\mathscr{F}_X$ such that (1.10) takes place. It turns out that the measure μ_z is uniquely determined by the regular point z. Let us recall that a T_X-invariant probability measure μ is called *ergodic* if and only if

$$E = T_X^{-1}E \in \mathscr{F}_X \quad \text{and} \quad \mu(E) > 0 \quad \text{imply} \quad \mu(E) = 1\,. \tag{1.11}$$

We shall often use a seemingly weeker form:

$$\mu(E) \in \{0, 1\} \quad \text{for all} \quad E \in \mathscr{F}_X \quad \text{such that} \tag{1.12}$$
$$\mu(E \Delta T_X^{-1} E) = 0$$

In the context of the present paper the definitions are equivalent and the latter one is sometimes easier to work with. Here and in the sequel the symbol Δ will denote the symmetric difference:

$$E \Delta F = (E - F) \cup (F - E) .$$

Turning back to (1.10) we can easily conclude that μ_z is finitely additive and T_X-invariant on the family $\mathscr{P}_X$. It is σ-additive if and only if

$$\sum \mu_z(A) = 1 (A \in \mathscr{P}_X^{i,n}; \quad i \in I, n \in N) . \tag{1.13}$$

The conditions (1.13) imply that the formula

$$\mu_z[E]_{i,n} = \sum_{\bar{x} \in E} \mu_z[\bar{x}]_{i,n} , \quad E \in \mathfrak{B}(X^n)$$

uniquely determines the set function μ_z on the family $\sigma(\mathscr{P}_X^{i,n})$ for any $i \in I$ and $n \in N$. The Kolmogorov Extension Theorem (cf. e.g. [13], Chapter 9) then provides a unique σ-additive extension of μ_z to a shift-invariant probability measure (denoted by the same symbol μ_z) on the σ-field $\mathscr{F}_X$. Let us note that (1.13) is obviously satisfied when card $(X) < \infty$. As was pointed in [37], the conditions (1.13) fail to be true for $X = N$ (cf. also Section 3). This in turn implies that the quasiregular points do not determine, in general, a σ-additive probability measure.

If X is uncountable metric space, the family $\mathscr{P}_X$ does not generate the σ-field $\mathscr{F}_X$. Hence even the validity of (1.13) does not assure the existence of a probability measure on $\mathscr{F}_X$ extending the set function μ_z defined on the family $\mathscr{P}_X$. However, we can assume that the metric space X^I is complete (performing its completion, if necessary). Then the notion of the quasiregulatity can be redefined in the sense of Fomin (cf. e.g. [26], Sect. 7).

Summarizing, a *stationary discrete information source* (briefly a *stationary source*) will be identified with a finitely additive, not necessarily σ-additive, T_X-invariant probability defined on the field $\mathscr{A}_X$. In the paper we shall not distinguish the source and the corresponding set function.

2. The General Coding Problem

We shall start with some well-known illustrating examples. Then we shall formulate the general coding problem.

Example 1. Let $X = \{x_1, x_2, \ldots, x_N\}$. Let p be a probability N-vector $(p_1, p_2, \ldots, p_N)$, p_i being the probability of the outcome x_i; $i = 1, 2, \ldots, N$. A sequence

$x^{(1)}, \ldots, x^{(n)}$ of outcomes is obtained by means of the repeating the random experiment (X, p) n-times independently. We want to characterize this sequence by binary sequences, with the length of the binary sequence as small as possible. The asymptotic behaviour of the minimal length of such binary sequences is given in the following simple form of the source coding theorem.

Theorem 2.1. There exists a nonnegative real number H such that for arbitrary $\varepsilon > 0$ the sequence $x^{(1)}, \ldots, x^{(n)}$ of outcomes of n independent trials can be characterized by $0-1$ sequences of length $n(H + \varepsilon)$ with probability as close to unity as wanted, if n is large enough, but cannot be characterized, with any fixed positive probability, by $0-1$ sequences of length $n(H - \varepsilon)$ if n is large enough. Further we actually have

$$H = -\sum_{k=1}^{N} p_k \log_2 p_k \,. \tag{2.1}$$

The statement of the theorem can be shown to be valid for discrete memoryless source with the finite alphabet X. Let us recall that a discrete memoryless source produces the sequences of independent indentically distributed random variables. Hence, given the finite set X and a probability vector p as above, we define the measure on $\mathfrak{B}(X^n)$ simply as the product measure:

$$\mathsf{p}^n(\boldsymbol{x}) = \prod_{i=1}^{n} p(x_i) \quad \text{for} \quad \boldsymbol{x} = x_1, x_2, \ldots, x_n \,.$$

Example 2. Let X be a countable set. Assume we are given a σ-additive T_X-invariant probability measure μ on the σ-field $\mathscr{F}_X$ (i.e. a stationary source) and a positive number ε such that $\varepsilon < 1$. Since $\mu = \mu T_X^{-1}$, the formula

$$\mu_n(E) = \mu[E]_{i,n} \,, \quad E \in \mathfrak{B}(X^n) \tag{2.2}$$

defines a probability measure μ_n on the σ-field $\mathfrak{B}(X^n)$ independently of what $i \in I$ was chosen. Let us define

$$L_n(\varepsilon, \mu) = \min \{\operatorname{card}(E) : E \subset X^n, \sum_{\bar{x} \in E} \mu_n\{\bar{x}\} > 1 - \varepsilon\} \,. \tag{2.3}$$

Theorem 2.2. [41]. The limit

$$\lim_{n \to \infty} n^{-1} \log_2 L_n(\varepsilon, \mu) = V_\varepsilon(\mu) \tag{2.4}$$

exist except at most a countable set of numbers ε. The function $V_\varepsilon(\mu)$ monotonically increases for $\varepsilon \to 0$ to a limit, which will be denoted by $V(\mu)$ and called the *asymptotic rate* of the source μ.

A "coding theorem-like "form of Theorem 2.2 is the following one:

Theorem 2.2 [40]. On the space of all stationary discrete information sources with a given at most countable alphabet X there is one and only one non-negative extended real-valued function V such that

(1) $\forall \lambda > 0\ \forall 0 < \varepsilon < 1\ \forall i \in I\ \exists n_0 \in N\ \forall n \geqq n_0\ \exists \mathscr{E} \subset \mathscr{P}_X^{i,n} [\mu(\cup \mathscr{E}) > 1 - \varepsilon]$ et $[\operatorname{card}(\mathscr{E}) < 2^{n[V(\mu)+\lambda]}]$;

(2) $\forall \lambda > 0\ \exists 0 < \eta < 1\ \forall \varepsilon \leqq \eta\ \forall i \in I\ \exists n_0 \in N\ \forall n \geqq n_0\ \forall \mathscr{E} \subset \mathscr{P}_X^{i,n}$ with $\mu(\cup \mathscr{E}) > 1 - \varepsilon$,

(a) $\operatorname{card}(\mathscr{E}) > 2^{n[V(\mu)-\lambda]}$ if $V(\mu) < \infty$,

(b) $\operatorname{card}(\mathscr{E}) > 2^{n\lambda}$ if $V(\mu) = \infty$.

Example 3. Source coding with a side information (cf. [1]). We are given two finite sets X, Y and the transition probabilities $p(y \mid x)$ for $x \in X$, $y \in Y$. We set

$$p^n(\mathbf{y} \mid \mathbf{x}) = \prod_{i=1}^{n} p(y_i \mid x_i)$$

for $\mathbf{y} = y_1 y_2, \ldots, y_n$, $\mathbf{x} = x_1 x_2, \ldots, x_n$. The transition probability $p(\cdot \mid \cdot)$ represents the fact that the Y-outcomes are correlated with the X-outcomes, respectively. A set $B \subset Y^n$ is said to ε-decode the sequence $\mathbf{x} \in X^n$ if

$$p^n(B \mid \mathbf{x}) \geqq 1 - \varepsilon .$$

Let

$$\Psi_\varepsilon^{(n)}(B) = \{\mathbf{x} : \mathbf{x} \text{ is } \varepsilon\text{-decoded by } B\} .$$

We assume we are given a probability measure P on X and a probability measure Q on Y, i.e. we are given two memoryless sources. It is assumed further that P and Q never vanish.

Theorem 2.3. [1]. There is a function $T(c)$ of non-positive real numbers c such that the limit

$$\lim_{n \to \infty} n^{-1} \log S_n(c, \varepsilon) = T(c)$$

exists and is independent of ε for any choice of c. Here

$$S_n(c, \varepsilon) = \min \{Q^n(B) : B \subset Y^n,\ n^{-1} \log P^n(\Psi_\varepsilon^{(n)}(B)) \geqq c\} .$$

Example 4. Let us consider a discrete memoryless source with the finite alphabet $X = \{0, 1, \ldots, J-1\}$ and a finite reproducing alphabet $Y = \{0, 1, \ldots, K-1\}$. We assume that the source is determined by a probability J-vector $\mathsf{p} = (p(0), p(1), \ldots, p(J-1))$. Assume $p(i) > 0$. Let $\varrho : X \times Y \to R_1^+$ be a single-letter distortion measure satisfying the conditions:

$$\text{for any } j \in X,\ \min \{\varrho(j, k) : k \in Y\} = 0 .$$

Put

$$\varrho(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \varrho(x_i, y_i)$$

for $\mathbf{x} = x_1 x_2, \ldots, x_n \in X^n$, $\mathbf{y} = y_1 y_2, \ldots, y_n \in Y^n$. For fixed $d \geqq 0$ and for any set $B_n \subset Y^n$ we shall define a subset $H(B_n, d)$ of X^n by the relation

$$H(B_n, d) = \{\mathbf{x} \in X^n : \varrho(\mathbf{x}, B_n) \geqq nd\} .$$

For $R > 0$ let us set

$$P^n(R, d) = \min \{\mathrm{p}^n(H(B_n, d)) : B_n \subset Y^n, \operatorname{card}(B_n) \leqq e^{nR}\} .$$

It is natural to say that an error occurs when $\mathbf{x} \in H(B_n, d)$. Thus the problem formally described above can be interpreted in the following way. We are given the encoding rate R. Within all coding sets B_n with a fixed code-length we look for the B_n with the smallets probability of an erroneous decoding, the error being expressed by means of a single-letter distortion measure ϱ.

Theorem 2.4 [22]. Let Q be the set of all probability J-vectors q. Let $R(\mathsf{q}, \cdot)$ be the rate distortion function of the discrete memoryless source uniquely determined by vector q. Let

$$F_d(R) = \min \{I(\mathsf{q} : \mathsf{p}) : \mathsf{q} \in \mathsf{Q}, R(\mathsf{q}, d) \geqq R\} .$$

For any R, $R(\mathsf{p}, d) < R < \max_{\mathsf{q} \in \mathsf{Q}} R(\mathsf{q}, d)$,

$$0 < F_d(R - 0) \leqq \liminf_{n \to \infty} [-(1/n) \log p^n(R, d)] \leqq$$

$$\leqq \limsup_{n \to \infty} [-(1/n) \log P^n(R, d)] \leqq F_d(R + 0) < \infty .$$

If F_d is continuous in R, then

$$0 < \lim_{n \to \infty} [-(1/n) \log P^n(R, d)] = F_d(R) < \infty .$$

Now we shall formulate a general coding problem. All coding problems mentioned in the above examples will be shown to be the special cases of the general problem.

Let X and Y be two, not necessarilly distinct, separable metric spaces. Let $\mathscr{E}$ denote an abstract set (usually $\mathscr{E} \subset R_1$). For any fixed $n \in N$, the set $\mathscr{E}$ determines a one-parameter family $\{\psi_\varepsilon^{(n)} : \varepsilon \in \mathscr{E}\}$, where either $\psi_\varepsilon^{(n)} : Y^n \to X^n$ is Borel measurable or $\psi_\varepsilon^{(n)} : \mathscr{B}(Y^n) \to \mathscr{B}(X^n)$. Further we are given two set functions $K_n^{(X)} : \mathscr{B}(X^n) \to R_1$ and $K_n^{(Y)} : \mathscr{B}(Y^n) \to R_1$. Let $\varphi \subset R_1 \times R_1$ be a fixed binary relation, let $\{c_n\}_{n \in N}$ be a fixed sequence of real numbers.

We shall formulate now two versions of the general coding problem. They are in a sense dual.

General coding problem (I): The n-sequences $\boldsymbol{x} \in \psi_\varepsilon^{(n)}(B)$ for $B \in \mathscr{B}(Y^n)$ are said to be ε-decodable by means of the set B. We are interested in the minimum "size" of a B which satisfies a prescribed bound for the "size" of $\psi_\varepsilon^{(n)}(B)$. The "sizes" are measured using the functions $K_n^{(Y)}$ and $K_n^{(X)}$, respectively. Hence we are interested in the quantity:

$$S_n^{\mathrm{I}} = \min \{K_n^{(Y)}(B_n) : B_n \in \mathscr{B}(Y^n), (K_n^{(X)}(\psi_\varepsilon^{(n)}(B_n)), c_n) \in \varphi\} .$$

General coding problem (II): Given a bound for the "size" of the set B of the codewords itself, we are interested in the minimum "size" for $\psi_\varepsilon^{(n)}(B)$. The "sizes" are again measured using $K_n^{(X)}$ and $K_n^{(Y)}$. Hence we are interested in the quantity:

$$S_n^{\mathrm{II}} = \min \{K_n^{(X)}(\psi_\varepsilon^{(n)}(B_n)) : B_n \in \mathscr{B}(Y^n), (K_n^{(Y)}(B_n), c_n) \in \varphi\} .$$

Let us note that in all cases the dependence on n of the constants c_n can be removed using a modified version of the corresponding function K_n. Hence the corresponding coding theorem can be formulated in the following form:

General coding theorem: There is an extended real valued function $T^\alpha = T^\alpha(\varepsilon, c)$ $(\alpha = \mathrm{I}, \mathrm{II})$ such that

$$\lim_{n \to \infty} n^{-1} \log S_n^\alpha = T^\alpha (\alpha = \mathrm{I}, \mathrm{II}) .$$

The strong converse (cf. [42]) states that the function T is independent of ε for any choice of the constant c, hence it depends only on the functions $K_n^{(X)}$ and $K_n^{(Y)}$, respectively.

First of all we shall show that all coding problems met in the above examples are the special cases of the general coding problem.

1. Let $X = Y = \{1, 2, \ldots, N\}$, let $c_n = 1 - \delta$ for any $n \in N$. The mappings $\psi_\varepsilon^{(n)}$ are assumed to be the identity transformations. The relation φ is specified by $x \varphi y$ iff $x \geqq y$. Finally, let us choose

$$K_n^{(Y)}(B) = \mathbf{p}^n(B) , \quad B \in \mathfrak{B}(X^n) ,$$

$$K_n^{(X)}(B) = \operatorname{card}(B) , \quad B \in \mathfrak{B}(X^n) .$$

Then

$$S_n^{\mathrm{II}} = \min \{\operatorname{card}(B) : B \subset X^n, \mathbf{p}^n(B) \geqq 1 - \delta\} .$$

However, we can choose $K_n^{(X)}$ and $K_n^{(Y)}$ conversely, and then

$$S_n^{\mathrm{I}} = \min \{\operatorname{card}(B) : B \subset X^n, \mathbf{p}^n(B) \geqq 1 - \delta\} .$$

Hence the distortionless coding problem of the Example 1 can be viewed as the special case of both coding problems. The same argument can be used for the problem of the second example.

2. Let X, Y be finite sets as in the Example 3. For any ε, $0 < \varepsilon < 1$, we shall set

$$\psi_\varepsilon^{(n)}(B_n) = \{\mathbf{x} : \mathbf{x} \text{ is } \varepsilon\text{-decodable by } B_n\}$$

Let $x \varphi y$ iff $x \geqq y$. Finally, let us set

$$K_n^{(X)} = P^n , \quad K_n^{(Y)} = Q^n .$$

The constants c_n depend on a fixed nonpositive number c through the relation $c_n = e^{nc}$. Then

$$S_n^{\mathrm{I}} = \min \{Q^n(B_n) : B_n \subset Y^n,\ P^n(\psi_\varepsilon^{(n)}(B_n)) \geqq e^{nc}\} =$$
$$= \min \{Q^n(B_n) : B_n \subset Y^n,\ n^{-1} \log P^n(\psi_\varepsilon^{(n)}(B_n)) \geqq c\} .$$

Hence the source coding problem with side information is a special case of the general coding problem (I).

3. Let us consider the Example 4. Now the parameter set is R_1^+. Given $B_n \subset Y^n$ and $d \in R_1^+$ we define

$$\psi_d^{(n)}(B_n) = \{\mathbf{x} : \mathbf{x} \in x^n,\ \varrho(\mathbf{x}\, B_n) \geqq nd\} .$$

Let $R > 0$ be given. The constants c_n are then defined by $c_n = e^{nR}$. Choose $K_n^{(X)}$ as the n-th power of the given probability J-vector $\mathbf{p}$. Take $K_n^{(Y)}$ as the counting measure on the finite set Y^n. Then

$$S_n^{\mathrm{II}} = \min \{\mathbf{p}^n(\psi_d(B_n) : B_n \subset Y^n, \operatorname{card}(B_n) \leqq e^{nR}\} .$$

In the paper we shall deal mainly with the problems of the distortionless coding. Further we shall study the cases in which both criteria K_n arise from information sources. Let us note that none of the examples given above deals with such a situation. In Section 4 we shall show, however, that it is possible to prove the coding theorems in much more elementary setup.

3. Existence of Finitely Additive Probabilities

This problem was discussed by the author in [37]. The problem was reduced to the problem of the existence of finitely additive probabilities on the σ-field $\mathfrak{B}(N)$ of all subsets of the set N of all positive integers. There were given simple examples of such set function in [37]. However, these set functions have no natural interpretation. The aim of this section is to give another justification for finitely additive probabilities in N, based on the notion of the quasiregularity. It follows from (1.10) that only one-sided sequences are to be considered. Let $z \in N^N$. We shall say that a set

$A \subset N$ possesses the *z-density* (and we shall write $A \in \mathscr{L}(z)$) provided there exists the limit

$$h_z(A) = \lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} \chi_A(z_j) . \tag{3.1}$$

If $z \in N^I$ is a quasiregular point, then the definition of the elementary cylinders together with the finite additivity of μ_z imply that $h_z(A)$ exists for all finite sets $A \subset N$. Thus we are restricted ourselves to the set $Q \subset N^N$ defined by the relation

$$Q \cap \{\{z : z \in N^N, \exists h_z(A)\} : \operatorname{card}(A) < \infty\} = \tag{3.2}$$

$$= \bigcap_{k=1}^{\infty} \{z : z \in N^N, \exists h_z\{k\}\} .$$

We shall classify the points in Q by partitioning it into the following three disjoint sets:

$$C_1 = \{z : z \in Q, \sup_{j\in N} z_j < \infty\} ,$$

$$C_2 = \{z : z \in Q, \sup z_j = \infty ,$$

$$\lim_{n\to\infty} \lim_{k\to\infty} k^{-1} \operatorname{card}\{j : 1 \leqq j \leqq k, z_j > n - 1\} = 0\} ,$$

$$C_3 = Q - (C_1 \cup C_2) .$$

Proposition 3.1. (1) Let $z \in C_1 \cup C_2$. Then $\mathscr{L}(z) = \mathfrak{B}(N)$ and h_z is a σ-additive probability measure on $\mathscr{L}(z)$. (2) Let $z \in C_3$. Then $\mathscr{L}(z)$ is a logic of subsets of N with respect to the set-theoretical operations of union and complementation. The set function h_z is a finitely additive non σ-additive probability on $\mathscr{L}(z)$.

Let us note that a family $\mathscr{L}$ of subsets of a given set X is called a *logic* if it has the following three properties:

(i) if $E \in \mathscr{L}$ then $E^c = X - E \in \mathscr{L}$,

(ii) if $E, F \in \mathscr{L}$ and $E \cap F = \emptyset$ thrn $F \cup E \in \mathscr{L}$,

(iii) if $E, F \in \mathscr{L}$ and $E \subset F$ then $F - E \in \mathscr{L}$.

Proof. The properties (i) and (ii) of logic obviously hold true for the family $\mathscr{L}(z)$ for any z, $z \in Q$.

(1a). Let $z \in C_1$. let $K = \sup\{z_j : j \in N\}$. Given a set $E \subset N$ there is a partition of it in the following form:

$$E = [E \cap \{1, 2, \ldots, K\}] \cup [E \cap \{K + 1, K + 2, \ldots\}] .$$

By the definition, $E \cap \{1, 2, \ldots, K\} \in \mathscr{L}(z)$. Since $z_j \leqq K$ for $j = 1, 2, \ldots, E \cap$ $\cap \{K + 1, K + 2, \ldots\} \in \mathscr{L}(z)$ and the z-density of this set equals 0. Hence $E \in \mathscr{L}(z)$ by the property (ii). The σ-additivity of h_z is obvious.

(1b). Let $z \in C_2$. When denoting by a_n the limit

$$\lim_{k\to\infty} k^{-1} \operatorname{card}\{j : 1 \leqq j \leqq k, z_j > n-1\}, \tag{3.3}$$

we have $h_z(\{n, n+1, \ldots\}) = a_n$. Since $z \in C_2$, $\lim a_n = 0$. Moreover $a_n \geqq a_{n+1}$ for any $n \in N$. Consequently, we have $h_z\{n\} = a_n - a_{n+1} \geqq 0$ and $\sum_{n=1}^{\infty} h_z\{n\} = 1$. The proof now follows because of the relation

$$h_z(E) = \sum_{n\in E} h_z\{n\}, \quad E \in \mathfrak{B}(N).$$

(2). Let $z \in C_3$. Then $h_z(\{n, n+1, \ldots\}) = a_n$ converges to a positive constant $L (0 < L \leqq 1)$. Since the decreasing sequence $\{\{n, n+1, \ldots\}\}_{m=1}^{\infty}$ has a void intersection, the set function h_z cannot be σ-additive. Finally, the property (iii) of $\mathscr{L}(z)$ can be proved easily using the mathematical induction. This finishes the proof of the proposition.

Remark 3.2. Let $z \in N^I$ be a quasiregular point with the coordinates

$$z_i = \begin{cases} 1 & \text{if } i \leqq 0, \\ i & \text{if } i > 0, \ i \in I. \end{cases}$$

The corresponding one-sided sequence is in Q with $\lim a_n$ positive and equal to unity. Hence $z \in C_3$. Let us define a set $A(z) \subset N$ recursively by the following list of instructions:

print the first positive integer (1),
do not print 2 subsequent (2,3),
print 3 subsequent (4, 5, 6),
do not print 6 subsequent (7—12),
print 3×4 subsequent (13—24),
do not print 6×4 subsequent (25—48),
print $3 \times 4 \times 4$ subsequent, etc.

Let

$$b_n = n^{-1} \sum_{j=1}^{n} \chi_{A(z)}(j), \quad n = 1, 2, \ldots$$

Assume $\{c_i\}_{i=1}^{\infty}$ be a recursive subsequence of the sequence $\{b_n\}_{n=1}^{\infty}$ defined inductively as follows

$$c_1 = b_3; \quad \text{if } c_i = b_n \text{ then } c_{i+1} = b_{2n}.$$

Then

$$c_i = \begin{cases} 1/3 & \text{if } i \text{ is odd}, \\ 2/3 & \text{if } i \text{ is even}. \end{cases}$$

Thus the sequence $\{(1/n) \sum_{j=1}^{n} \chi_{A(z)}(j)\}_{n=1}^{\infty}$ contains an oscillating subsequence, i.e. $A(z) \notin \mathscr{L}(z)$. Let us note that to any recursive sequence $z \in C_3$ it is possible to construct sets $A(z)$ which are not members of $\mathscr{L}(z)$.

The latter remark implies that the family of sets, to which a finitely additive probability can be reasonably assigned, forms merely a logic of sets. Generally speaking, an extension of a finitely additive set function defined on a logic to the σ-field generated by this logic is not possible. However, the following proposition is valid:

Proposition 3.3. Let $z \in C_3$. Then there is a finitely additive probability $\bar{h}_z$ on the σ-field $\mathfrak{B}(N)$ such that

$$\bar{h}_z(A) = h_z(A) \quad \text{for} \quad A \in \mathscr{L}(z) .$$

Proof. Let us consider the linear space $l^\infty(N)$ of all bounded sequences of real numbers. The norm will be the usual supremal norm. The space $\mathscr{K}(N)$ of all convergent sequences is a normed linear subspace of $l^\infty(N)$. By definition,

$$\mathscr{L}(z) = \{A : A \subset N, \{(1/n) \sum_{j=1}^{n} \chi_A(z_j)\}_{n=1}^{\infty} \in \mathscr{K}(N)\} .$$

The limit can be considered as a bounded linear functional on the space $\mathscr{K}(N)$. By Hahn-Banach Extension Theorem (cf. e.g. [8]) there is a bounded linear functional Lim on $l^\infty(N)$ such that

$$\operatorname{Lim} \{a_n\} = \lim a_n \quad \text{for} \quad \{a_n\} \in \mathscr{K}(N) .$$

If $E \subset N$ then $\{\chi_E(n)\}_{n=1}^{\infty} \in l^\infty(N)$. Let us set

$$\mu(E) = \operatorname{Lim} \{\chi_E(n)\} .$$

The set function μ is a finitely additive probability. Moreover, it can be easily shown that

$$\operatorname{Lim} \{a_n\} = \int_N a_n \, \mu(\mathrm{d}n) , \quad \{a_n\} \in l^\infty(N) .$$

The integral on the right-hand side of the latter relation is a finitely additive integral (cf. e.g. [8], Chapt. III). Given $E \subset N$, the sequence $\{n^{-1} \sum_{j=1}^{n} \chi_E(z_j)\}_{n=1}^{\infty}$ is in $l^\infty(N)$. The relation

$$\bar{h}_z(E) = \int_N n^{-1} \sum_{j=1}^{n} \chi_E(z_j) \, \mu(\mathrm{d}n)$$

determines a set function $\bar{h}_z$ on the σ-field $\mathfrak{B}(N)$ such that the proposition is valid. The proof is complete.

Remark 3.4. A completely different approach can be investigated within the frame of the constructive theory of probability (cf. e.g. [33]). In this approach it is assumed that the quasi-regular points are only the random sequences. The resulting set functions are always σ-additive on a σ-field, the notions of σ-additivity and of the σ-field being constructively redefined.

4. The Main Tools

The basic method used to prove the coding theorem 2.2 was investigated in [27] and further developed in [41]. We shall use a modification of this method as the main tool for proving our statements. In this section, unless explicitely stated the opposite, we shall confine ourselves to an at most countable alphabet X. The important results will be stated for the purpose of the later reference.

Let us start with the notion of the entropy rate for a stationary source μ with a countable alphabet (cf. e.g. [40]). The entropy rate $H(\mu)$ is defined as the limit

$$H(\mu) = -\lim_{n\to\infty} n^{-1} \int \log \mu[z_0, \ldots, z_{n-1}]\, \mu(\mathrm{d}z)\,. \tag{4.1}$$

Let us note that the quantity $H(\mu)$ is actually nothing but the entropy of the homeomorphism T_X as defined originally in [19]. The main idea consists in the introducing a measurable function g_μ on X^I such that

$$H(\mu) = -\int \log g_\mu(z)\, \mu(\mathrm{d}z)\,. \tag{4.2}$$

If we assume $H(\mu) < \infty$, the latter relation means that the function $g_\mu(z)$ is integrable, more precisely, that the function $\log g_\mu(z)$ is integrable. Hence, according to the individual ergodic theorem, there is the limit

$$h_\mu(z) = -\lim_{n\to\infty} (1/n) \sum_{j=0}^{n-1} \log g_\mu(T^j z) \text{ a.e.}[\mu] \tag{4.3}$$

satisfying the equality

$$\int h_\mu(z)\, \mu(\mathrm{d}z) = H(\mu)\,. \tag{4.4}$$

Theorem 4.1 [41]. If μ is a stationary source with $H(\mu) < \infty$ then the sequence $-(1/n) \log \mu[z_0, \ldots, z_{n-1}]$ converges in mean (w.r. to μ) to the function $h_\mu(z)$.

The following necessary facts from the ergodic theory are based on the notion of regularity as defined in the first section. For the original contribution see [20].

Lemma 4.2 [41, Lemma 2]. Let R_X denote the set of all regular points in X^I. Then $R_X \in \mathscr{F}_X$ and $\mu(R_X) = 1$ for every stationary source μ with the alphabet X.

Lemma 4.3 [41, Lemma 3]. For every ergodic source μ,

$$\mu\{z : z \in R_X,\ \mu_z = \mu\} = 1\,. \tag{4.5}$$

Lemma 4.4 [41, Lemma 4]. For any nonnegative measurable function f on the space $(X^I, \mathscr{F}_X)$ the integral $\int f \, d\mu_z$ is a measurable function of the variable z on R_X, and

$$\int f \, d\mu = \int_{R_X} \left[\int f \, d\mu_z \right] \mu(dz) \tag{4.6}$$

for all stationary sources μ.

Lemma 4.5 [41, Lemma 5]. If μ is a stationary source with finite entropy, i.e. if $H(\mu) < \infty$, then

$$\mu\{z : z \in R_X, h_\mu(z) = H(\mu_z)\} = 1 . \tag{4.7}$$

The following version of McMillan's theorem for countable alphabets is an immediate corollary both to Theorem 4.1 and the last lemma.

Theorem 4.6 [41, Theorem 2]. If μ is a stationary source with finite entropy, then the sequence

$$-(1/n) \log \mu[z_0, \ldots, z_{n-1}]$$

converges in mean (w.r. to μ) to the function $H(\mu_z)$, i.e. to the entropy rate of the ergodic component μ_z of the source given.

Let μ be a finite-alphabet source, i.e. there is a positive integer k such that

$$\mu(\{1, 2, \ldots, k\})^I = 1 . \tag{4.8}$$

For any stationary source satisfying (4.8) we have the inequality

$$H(\mu) \leqq \log_2 k \tag{4.9}$$

(cf. (4.1)). In this case the McMillan's theorem stated above enables to prove two basic lemmas desired for the proof of the coding theorem 2.2. However, if the alphabet is infinite, it may happen that $H(\mu) = \infty$. Hence Theorem 4.6 is not directly applicable. The proof is then performed using an approximation property of the entropy rate, as proved by Sinaj [36] and Parthasarathy [28]. Let τ_k be the mapping of N^I onto $\{1, 2, \ldots, k+1\}^I$ defined by the properties that

$$(\tau_k z)_i = \begin{cases} z_i & \text{if } z_i \leqq k , \\ k+1 & \text{if } z_i > k ; \end{cases} \quad z \in N^I , \quad i \in I . \tag{4.10}$$

Theorem 4.7 ([36], [28]). Let μ be a stationary source with the alphabet N. Let $\mu\tau_k^{-1}$ be the finite-alphabet source induced from μ by the mapping τ_k ($k = 1, 2, \ldots$ $\ldots$). Then

$$H(\mu\tau_k^{-1}) \leqq H(\mu\tau_{k+1}^{-1}) \quad (k = 1, 2, \ldots)$$

and

$$H(\mu) = \lim_{k \to \infty} H(\mu\tau_k^{-1}) . \tag{4.11}$$

To motivate the further discussion we shall give now the main idea of the proof. Let

$$\zeta = \{[k] : k = 1, 2, \ldots\}$$

be a countable partition of the space N^I. The partition is a generator [32]. Let us define, for $k = 1, 2, \ldots$, the partition ζ_k by

$$\zeta_k = \{[1], \ldots, [k], [\{k + 1, k + 2, \ldots\}]\} .$$

Then $\zeta_i < \ldots < \zeta_k < \ldots < \zeta$ (cf. [32]). But $H(\zeta) = H(\mu)$ and $H(\zeta_k) = H(\mu\tau_k^{-1})$. Hence a general result of [36] applies (cf. also [32], p. 15, Proposition (b)).

If X is an uncountable metric space, the partition $\{[x] : x \in X\}$ is uncountable, hence it cannot be a generator in the sense of Rochlin [32]. This in turn implies that the approximation property of Theorem 4.7 cannot be used in the general case. Hence, we are forced to consider the finite partitions of the alphabet, thus obtaining finite-alphabet sources. It will be proved in the third part of this paper that this approach gives the same results in the special case of a countably infinite alphabet as the approach of Winkelbauer [41] described above.

PART II: ERGODIC THEORY OF FINITELY ADDITIVE PROBABILITIES

5. Preliminary Discussion — The Markovian Case

In this section we shall study some questions connected with the denumerable Markov chains. The purpose of this investigation is to illustrate some problems of the ergodic theory of finitely additive probabilities. At the same time, the Markov chains will provide a lot of interesting examples concerning the information-theoretical quantities studied in the subsequent parts of the paper. An exhausting reference concerning Markov chains may be found in [7].

Let us start with the following special type of a finite Markov chain. The states will be identified with the elements of the finite set $\{1, 2, \ldots, k\}$. The one-step transition probabilities p_{ij} are assumed to form the following $k \times k$-type stochastic matrix:

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & 0 & \ldots & 0 & 0 \end{pmatrix} . \tag{5.1}$$

The symbol $p_{ij}^{(n)}$ will be used to denote the (i, j)-th entry of the matrix $\mathbf{P}^n$; to $p_{ij}^{(n)}$'s we are referred as to the n-step transition probabilities. The matrix $\mathbf{P}$ has the following property:

$$\mathbf{P}^k = \mathbf{I} \quad \text{(the identity matrix)}, \tag{5.2}$$

hence

$$\mathbf{P}^{k+1} = \mathbf{P}\,. \tag{5.3}$$

The classification of the states follows from the above properties of the matrix $\mathbf{P}$. Let (i, j) be any pair of states. Due to (5.3) there are positive integers n_1 and n_2 such that $p_{ij}^{(n_1)} > 0$ and $p_{ji}^{(n_2)} > 0$, respectively. Hence the corresponding Markov chain is irreducible, its state space consisting precisely of a single essential set of states. Moreover, the matrix $\mathbf{P}$ is obviously indecomposable. The relation (5.2) can be rewritten in the form

$$p_{jj}^{(k)} = 1\,, \quad j = 1, 2, \ldots, k\,.$$

Thus

$$\sum_{n=1}^{\infty} p_{jj}^{(n)} = \sum_{n=1}^{\infty} p_{jj}^{(kn)} = \infty\,,$$

i.e. all states are recurrent. From (5.3) it follows all states are periodic with the same period $d_j = k$, $j = 1, \ldots, k$. The ergodic theorem for Markov chains implies that

$$\lim_{n\to\infty} p_{jj}^{(n)} = d_j^{-1} = 1/k\,; \quad j = 1, 2, \ldots, k\,.$$

As well-known, for irreducible periodic Markov chains the above limit determines uniquely a probability k-vector $\mathbf{p} = (p_1, \ldots, p_k)$ such that

$$\mathbf{pP} = \mathbf{p}\,. \tag{5.4}$$

Any probability vector $\mathbf{p}$ satisfying (5.4) (if it exists) is said to be absolute stationary distribution of the Markov chain corresponding to the matrix $\mathbf{P}$. Hence, in our case, the absolute stationary distribution is the uniform distribution on the set $\{1, 2, \ldots, k\}$ of all states.

Recall now the usual construction of the Markov chain corresponding to the initial distribution $\mathbf{p}$ and the transition probability matrix $\mathbf{P}$. First of all we shall define the set function $P = P(\mathbf{p}, \mathbf{P})$ on the family of all elementary cylinders in the space $\{1, \in .., k\}^N$ by means of the formula

$$P([i_0, \ldots, i_{n-1}]_{0,n}) = p_{i_0} \prod_{t=1}^{n-1} p_{i_{t-1} i_t} \tag{5.5}$$

for $i_0, \ldots, i_{n-1} \in \{1, \ldots, k\}$. Using Kolmogorov Extension Theorem (cf. [13],

Chapter 9) the set function P can be extended to a σ-additive probability measure on the σ-field generated by the family of all elementary cylinders. Let

$$X_n(z) = z_n \quad \text{for} \quad z \in \{1, \ldots, k\}^N ; \quad n \in N$$

be the coordinate random variables. The sequence $\{X_n\}_{n \geqq 0}$ is the *Markov chain* corresponding to the initial distribution $\mathbf{p}$ and the matrix $\mathbf{P}$, respectively. The stationary behaviour of the transition probabilities itself does not assure that the Markov chain $\{X_n\}_{n \geqq 0}$ is stationary with respect to the shift-transformation T in the space $\{1, \ldots, k\}^N$. Let the matrix $\mathbf{P}$ be given by (5.1). Then we have the following criterium of stationarity.

Lemma 5.1. The Markov chain $\{X_n\}_{n \geqq 0}$ corresponding to the initial distribution $\mathbf{p} = (p_1, p_2, \ldots, p_k)$ and to the matrix $\mathbf{P}$ is stationary if and only if

$$p_1 = p_2 = \ldots = p_k (= 1/k) . \tag{5.6}$$

Proof. The "if" part follows from (5.5) and (5.6). Indeed, a random process is stationary if and only if the corresponding product measure is invariant.

Conversely, let $\{X_n\}_{n \geqq 0}$ be stationary. Especially

$$P([n, n+1]_{1,2}) = P(T_k^{-1}[n, n+1]_{1,2}) = P([n, n+1]_{2,2}) .$$

However,

$$P([n, n+1]_{1,2}] = p_{n-1}$$

and, on the other hand,

$$P([n, n+1]_{2,2}) = p_{n-2}$$

provided $3 \leqq n \leqq k - 1$. The remaining possibilities can be examined by use only of the minor changes. Consequently $p_1 = \ldots = p_k = 1/k$.

A natural infinite-dimensional analogue to the matrices $\mathbf{P}$ introduced in (5.1) is the infinite-dimensional matrix $\mathbf{P}$ defined entrywise by the properties that

$$p_{ij} = \begin{cases} 1 & \text{if } j = i + 1 , \\ 0 & \text{otherwise;} \quad i, j = 1, 2, \ldots . \end{cases} \tag{5.7}$$

The matrix $\mathbf{P}$ is again indecomposable. However, all states are now transient, non-recurrent and aperiodic. The idea used in the proof of the preceeding lemma implies that the corresponding Markov chain (see the construction below) is stationary if and only if

$$p_1 = p_2 = \ldots \tag{5.8}$$

Hence there is no σ-additive probability working as the absolute stationary distribution. Usually this difficulty is avoided by admitting also σ-finite measures as the

stationary distributions. An alternative approach admits finitely additive probabilities within the family of the absolute stationary distributions. However, there are infinitely many finitely additive probabilities on the σ-field $\mathfrak{P}(N)$ possessing the property (5.8) (cf. [37] and Section 3 of the present paper.). We should like to choose such one from which the finite-dimensional cases mentioned above could be derived. Let $\mathbf{p}$ on $\mathfrak{P}(N)$ be one of the extensions $\bar{h}_z$ of the z-density h_z (cf. (3.1) and Proposition 3.3) with z being the sequence of all positive integers. This choice is justified by the following statement.

Lemma 5.2. Let us consider the Markov chain with the stationary distribution $\bar{h}_z$ and the transition probability matrix $\mathbf{P}$ defined by (5.7). For any k, $k \in N$, there is a mapping $\varphi : N \to \{1, \ldots, k\}$ (called the collapse of the states) such that the collapsed process is a stationary finite Markov chain. The corresponding product measure P is the Markovian measure $P = P(\mathbf{p}, \mathbf{P})$ where $\mathbf{p} = (1/k, \ldots, 1/k)$ and the stochastic matrix $\mathbf{P}$ is given by (5.1).

Proof. Let us consider the partition of the state space N of the original Markov chain into the residue classes modulo k. The put

$$\varphi(l) = j + 1 \quad \text{if} \quad l \in \{kn + j : n = 1, 2, \ldots\}$$

for $j = 0, 1, \ldots, k - 1$. Then for any j, $0 \leqq j \leqq k - 1$,

$$p_{j+1} = h_z(\{kn + j : n = 1, 2, \ldots\}) = k^{-1},$$

hence $\mathbf{p} = (1/k, \ldots, 1/k)$. The one-step probability of remaining within a given collapsed state $\{kn + j : n = 1, 2, \ldots\} = j + 1$ is zero. If we are in the state labeled by j, the one-step probability of the transition into the state labeled by $j + 1 \pmod k$ equals unity. Hence the new transition matrix is that given by (5.1). The lemma is proved.

Now we shall give the idea of constructing a finitely additive Markov chain given a finitely additive initial distribution and the stochastic matrix $\mathbf{P}$ defined by (5.7). Here, the importance of the set function $\bar{h}_z$ will be manifested once more. Let us define the transition function $p(.,..) : N \times \mathfrak{P}(N) \to [0, 1]$ by the property that

$$p(n, A) = \chi_A(n + 1). \tag{5.9}$$

Clearly $p(n, \{n + 1\}) = 1$. Consequently, the transition function $p(.,..)$ induces the matrix $\mathbf{P}$. Moreover,

$$\int p(n, A)\, h_z(\mathrm{d}n) = h_z(A), \quad A \in \mathscr{L}(z).$$

If $A \in \mathfrak{P}(N) - \mathscr{L}(z)$, then

$$\bar{h}_z(A) = \int_N \frac{1}{n} \sum_{j=1}^{n} \chi_A(j)\, \mu(\mathrm{d}n) \tag{5.10}$$

(cf. the proof of the Proposition 3.3). Hence

$$\int p(n, A)\, \bar{h}_z(\mathrm{d}n) = \int \chi_A(n+1)\, \bar{h}_z(\mathrm{d}n) = \int \chi_{A\ominus 1}(n)\, \bar{h}_z(\mathrm{d}n)$$

where $A \ominus k = \{n : n + k \in A\}$ (if there is no such n, we shall set $A \ominus k = \emptyset$}. But the right-hand side of the latter relation equals $\bar{h}_z(A \ominus 1)$. The invariance property

$$\operatorname*{Lim}_n f(n) = \operatorname*{Lim}_n f(n+m), \quad m \geqq 0$$

of the Banach limits implies, by the definition of $\bar{h}_z$, the equality

$$\bar{h}_z(A \ominus 1) = \bar{h}_z(A)\,.$$

Therefore

$$\int p(n, A)\, \bar{h}_z(\mathrm{d}n) = \bar{h}_z(A)\,, \quad A \in \mathfrak{P}(N)\,. \tag{5.11}$$

The relation (5.11) means that $\bar{h}_z$ is invariant with respect to the transition function $p(.,..)$. Note that this property fails to hold, in general, for such initial distributions which are not resulting by the Banach limit procedure.

The relation (5.5) is meaningless for finitely additive initial distributions, because it reduces to the tautology $0 = 0$. Instead of (5.5) we shall use the well-known construction of a Markov process due to I. Tulcea. For the sake of simplicity, we shall confine ourselves to finite-dimensional cylinders with a low dimension. The ideas will, of course, work for arbitrary finite-dimensional cylinders. E.g. the measure of the two-dimensional cylinder $[A_1 \times A_2]_{1,2}$ is given by the formula

$$P([A_1 \times A_2]) = \int_N \left[\int_{A_1} p(n_1, A_2)\, p(n_0, \mathrm{d}n_1)\right] \bar{h}_z(\mathrm{d}n_0)\,. \tag{5.12}$$

Now

$$P([A_1 \times N]_{1,2}) = \int_N \left[\int_{A_1} p(n_1, N)\, p(n_0, \mathrm{d}n_1)\right] \bar{h}_z(\mathrm{d}n_0) =$$

$$= \int_N \left[\int_{A_1} p(n_0, \mathrm{d}n_1)\right] \bar{h}_z(\mathrm{d}n_0) = \int_N p(n_0, A_1)\, \bar{h}_z(\mathrm{d}n_0) = \bar{h}_z(A) = P([A_1]_{1,1})\,.$$

The latter equality is a consequence of (5.10). The equality of the first and of the last member in the latter relation implies that by means of the formula (5.12) (and of its analogues for larger dimensions) it is obtained a consistent family of finitely additive probabilities. Hence we are given a finitely-additive process. Let $\{X_n\}_{n \geqq 0}$ be the sequence of the coordinate variables. Then

$$\operatorname{Prob}(X_2 \in A_2 \mid X_0 \in A_0,\ X_1 \in A_1) =$$

$$= P(X_0 \in A_0,\ X_1 \in A_1,\ X_2 \in A_2)\,[P(X_0 \in A_0,\ X_1 \in A_1)]^{-1}$$

with 0/0 interpreted as 0. Using (5.12) the right-hand side of the latter relation equals

$$\int_{A_0}\left[\int_{A_1} p(n_1, A_2)\, p(n_0, \mathrm{d}n_1)\right]\bar{h}_z(\mathrm{d}n_0)\left\{\int_{A_0} p(n_0, A_1)\,\bar{h}_z(\mathrm{d}n_0)\right\}^{-1} =$$

$$= \int \chi_{A_0}(n_0)\left[\int_{A_1}\chi_{A_2}(n_2)\, p(n_0, \mathrm{d}n_1)\right]\bar{h}_z(\mathrm{d}n_0) \times \left\{\int \chi_{A_0}(n_0)\,\chi_{A_1}(n_1)\,\bar{h}_z(\mathrm{d}n_0)\right\}^{-1}.$$

Using the definition of the transition function one obtains for the last member the expression:

$$\int \chi_{A_0}(n_0)\,\chi_{A_1}(n_0+1)\,\chi_{A_2}(n_0+2)\,\bar{h}_z(\mathrm{d}n_0) \times \left\{\int \chi_{A_0}(n_0)\,\chi_{A_1}(n_0+1)\,\bar{h}_z(\mathrm{d}n_0)\right\}^{-1} =$$

$$= \int \chi_{A_0}(n_0)\,\chi_{A_1\ominus 1}(n_0)\,\chi_{A_2\ominus 2}(n_0)\,\bar{h}_z(\mathrm{d}n_0) \times \left\{\int \chi_{A_0}(n_0)\,\chi_{A_1\ominus 1}(n_0)\,\bar{h}_z(\mathrm{d}n_0)\right\}^{-1} =$$

$$= \int \chi_{A_0\cap(A_1\ominus 1)\cap(A_2\ominus 2)}(n_0)\,\bar{h}_z(\mathrm{d}n_0) \times \left\{\int \chi_{A_0\cap(A_1\ominus 1)}(n_0)\,\bar{h}_z(\mathrm{d}n_0)\right\}^{-1} =$$

$$= \bar{h}_z[A_0 \cap (A_1 \ominus 1) \cap (A_2 \ominus 2)]\,\{\bar{h}_z[A_0 \cap (A_1 \ominus 1]\}^{-1}.$$

Consider first the case $A_i \in \mathscr{L}(z)$, $i = 0, 1, 2$. Then we obtain the expression

$$h_z[A_0 \cap (A_1 \ominus 1) \cap (A_2 \ominus 2)]\,\{h_z[A_0 \cap (A_1 \ominus 1)]\}^{-1}$$

because $\bar{h}_z = h_z$ on $\mathscr{L}(z)$. The relative frequency of the transition from the set A_1 into the set A_2 does not depend on the relative frequency of the transition from A_0 into A_1. In symbols

$$\mathrm{card}\,\{j : 1 \leqq j \leqq n,\ j \in A_0,\ j+1 \in A_1,\ j+2 \in A_2\} \times$$

$$\times\,\{\mathrm{card}\,\{j : 1 \leqq j \leqq n,\ j \in A_0, j+1 \in A_2\}\}^{-1} =$$

$$= \mathrm{card}\,\{j : 1 \leqq j \leqq n,\ j+1 \in A_1,\ j+2 \in A_2\} \times$$

$$\times\,\{\mathrm{card}\,\{j : 1 \leqq j \leqq n,\ j+1 \in A_1\}\}^{-1}.$$

Consequently,

$$h_z[A_0 \cap (A_1 \ominus 1) \cap (A_2 \ominus 2)]\,\{h_z[A_0 \cap (A_1 \ominus 1)]\}^{-1} =$$

$$= h_z[(A_1 \ominus 1) \cap (A_2 \ominus 2)]\,\{h_z(A_1 \ominus 1)\}^{-1} =$$

$$= P(X_1 \in A_1, X_2 \in A_2)\,\{P(X_1 \in A_1)\}^{-1} = \mathrm{Prob}\,(X_2 \in A_2 \mid X_1 \in A_1).$$

Hence we obtained the equality

$$\mathrm{Prob}\,(X_2 \in A_2 \mid X_0 \in A_0,\ X_1 \in A_1) = \mathrm{Prob}\,(X_2 \in A_2 \mid X_1 \in A_1)$$

valid for all A_0, A_1, $A_2 \in \mathscr{L}(z)$. For general A_i's we have

$$\bar{h}_z[A_0 \cap (A_1 \ominus 1) \cap (A_2 \ominus 2)] \{\bar{h}_z[A_0 \cap (A_1 \ominus 1)]\}^{-1} =$$

$$= \int \frac{1}{n} \sum_{j=1}^{n} \chi_{A_0}(j)\, \chi_{A_1 \ominus 1}(j)\, \chi_{A_2 \ominus 2}(j)\, \mu(\mathrm{d}n) \times \left\{ \int \frac{1}{n} \sum_{j=1}^{n} \chi_{A_0}(j)\, \chi_{A_1 \ominus 1}(j)\, \mu(\mathrm{d}n) \right\}^{-1}.$$

Since $\int a(n)\, \mu(\mathrm{d}n) = \operatorname{Lim} a(n)$ for any $a(\cdot) \in l^{\infty}(N)$, we have $\int a(n+k)\, \mu(\mathrm{d}n) = \int a(n)\, \mu(\mathrm{d}n)$. Hence the same argument as before applies to obtain the equality

$$\frac{1}{n} \sum_{j=1}^{n} \chi_{A_0}(j)\, \chi_{A_1 \ominus 1}(j)\, \chi_{A_2 \ominus 2}(j) \times \left\{ \frac{1}{n} \sum_{j=1}^{n} \chi_{A_0}(j)\, \chi_{A_1 \ominus 1}(j) \right\}^{-1} =$$

$$= \frac{1}{n} \sum_{j=1}^{n} \chi_{A_1 \ominus 1}(j)\, \chi_{A_2 \ominus 2}(j) \left\{ \frac{1}{n} \sum_{j=1}^{n} \chi_{A_1 \ominus 1}(j) \right\}^{-1}.$$

It follows from the last equality that

$$\operatorname{Prob}(X_2 \in A_2 \mid X_0 \in A_0,\ X_1 \in A_1) = \operatorname{Prob}(X_2 \in A_2 \mid X_1 \in A_1)$$

for all A_0, A_1, $A_2 \in \mathfrak{B}(N)$. Note that the markovian property was established using the invariance properties of the Banach limits. On the other hand, if we should use finitely additive initial distributions which were not defined by means of the Banach limits, the markovian property could possibly fail to hold.

The functionals of the Markov processes are not, generally speaking, Markov processes. Lemma 5.1 showed that for a special type of the functional the induced process is again a Markov process. Now we shall prove a more general statement.

Proposition 5.3. Let $\{X_n\}_{n \geq 0}$ be the finitely additive Markov process with the distribution $P = P(\bar{h}_z, \mathbf{P})$. Let ξ be a finite partition of the state space N. If for any $C \in \zeta$, $C \in \mathscr{L}(z)$, then the induced probability measure P_ζ on the space $\{1, \ldots, \operatorname{card}(\zeta)\}^N$ corresponds to an ergodic finite Markov chain. The measure P_ζ is given explicitely by the formula $P_\zeta = P(\mathbf{q}, \mathbf{Q})$ with

$$q_i = h_z(C_i), \quad i = 1, 2, \ldots, \operatorname{card}(\zeta);$$

$$q_{ij} = [h_z(C_i)]^{-1} \lim_{n \to \infty} \frac{1}{n} \operatorname{card}\{k : 1 \leq k \leq n,\ (k, k+1) \in C_i \times C_j\}.$$

Proof. The way of inducing the measure P_ζ by means of the finite partitions of the state space is given in Section 10. From Lemma 10.2 it follows that the induced measure is ergodic. Hence we have to prove only the markovian property of the induced measure. Using the relations established above together with the invariance property (5.11) we compute

$$q_{ij} = \text{Prob}\,(X_1 \in C_j | X_0 \in C_i) = P(X_0 \in C_i,\ X_1 \in C_j)\,[\mathrm{P}(X_0 \in C_i)]^{-1} =$$

$$= [h_z(C_i)]^{-1} \lim_{n\to\infty} \frac{1}{n} \text{card}\,\{k : 1 \leqq k \leqq n, (k, k+1) \in C_i \times C_j\}\,.$$

Note that $h_z(C_i) = 0$ implies $q_{ij} = 0$ for all j. Thus, if there are some new states with h_z-probability vanishing, we have to reduce the state space by excluding these states. Let us denote by c_{ij} the limit on the right-hand side of the last relation.

Since the induced measure P_ζ is ergodic, the matrix $\mathbf{Q} = (q_{ij})$ is such that the linear system

$$\mathbf{q} = \mathbf{qQ}$$

has the unique nontrivial solution. The proof will be complete when showed that the vector $\mathbf{q}$ of the conclusion is a solution. Substituting for $\mathbf{q}$ into the above linear system, we obtain the following system of equalities:

$$h_z(C_j) = q_{1j} + q_{2j} + \ldots + q_{kj}\,, \quad j = 1, \ldots, k$$

with $k = \text{card}\,(\zeta)$. The latter relations are valid because of $(C_i \times C_j) \cap (C_i \times C_{j'}) = \emptyset$ for any $i, j \neq j'$, and because for any $1 \in \{1, \ldots, k\}$ there is certainly $j \in \{1, \ldots, k\}$ such that $l + 1 \in C_j$.

The last statement of this section will concern with an ergodic property of the Markov chain $\{X_n\}_{n\geqq 0}$ with $P = P(\bar{h}_z, \mathbf{P})$.

Proposition 5.4. For any $A \in \mathscr{L}(z)$ and for all $k \in N$,

$$\lim_{n\to\infty} \left(\frac{1}{n} \sum_{j=1}^{n} \mathbf{P}^j \chi_A\right)(k) = h_z(A)\,.$$

Proof. Clearly

$$\frac{1}{n} \sum_{j=1}^{n} \mathbf{P}^j = \begin{pmatrix} 0 & 1/n & 1/n & \ldots & 1/n & 0 & 0 & \ldots \\ 0 & 0 & 1/n & \ldots & 1/n & 1/n & 0 & \ldots \\ 0 & 0 & 0 & \ldots & & & & \\ & & \vdots & & & & & \end{pmatrix}.$$

Consequently

$$\frac{1}{n} \sum_{j=1}^{n} \mathbf{P}^j \chi_A(\cdot) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{P}^j \begin{pmatrix} \chi_A(1) \\ \chi_A(2) \\ \vdots \end{pmatrix} = \begin{pmatrix} 1/n\,(\chi_A(2) + \ldots + \chi_A(n+1)) \\ 1/n\,(\chi_A(3) + \ldots + \chi_A(n+2)) \\ \vdots \end{pmatrix}.$$

For $n \to \infty$, any component of the last column vector converges to $h_z(A)$ because of $A \in \mathscr{L}(z)$. Thus

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \mathbf{P}^j \chi_A(\cdot) = h_z(A)\,\chi_N(\cdot)\,.$$

The conclusion follows.

The above discussion motivates the following main problems:

(1) The ergodic decomposition of a finitely additive invariant probability.
(2) The concepts of the ergodicity in a finitely additive setting.
(3) The characterization of the structure of the ergodic set functions within the family of all additive invariant probabilities.

6. The Decomposition Problem

Let $(\Omega, \mathscr{S})$ be a measurable space, let $T: \Omega \to \Omega$ be a measurable transformation, i.e. $T^{-1}\mathscr{S} \subset \mathscr{S}$. Assume we are given a T-invariant probability measure on $\mathscr{S}$. The measure μ is said to be idecomposable if

$$\mu = \alpha\mu_1 + (1 - \alpha)\,\mu_2$$

with μ_1, μ_2 invariant implies $\mu_1 = \mu_2 = \mu$. In the frequently occuring situations the concepts of indecomposability and of ergodicity coincide. The set of all T-invariant probability measures is convex and the ergodic probabilities are indecomposable. Because of these facts the problem of the ergodic decomposition can be viewed as a special case of the following problem:

Let M denote an arbitrary convex set in a locally convex linear space. Let E denote the set of all extremal points in M.

(1) Find the sufficient conditions for M to be nonvoid.
(2) Find the conditions under which any element $m \in M$ can be represented by an integral with respect to a probability measure concentrated (in some certain sense) on the set E.

The classical solution is due to Krylov and Bogoljubov [20]. Their main results are stated in Lemmas 4·2 – 4·4 with a compact metric space X and regular Borel probability measures. Let us note that Lemma 4.4 gives as a special case the relation

$$\mu(E) = \int_{R_\Omega} \mu_z(E)\,\mu(\mathrm{d}z),\ E \in \mathscr{B}(\Omega)\,. \tag{6.1}$$

The proofs of these statements did not use convexity arguments (cf. e.g. [26]). For countably infinite alphabet the basic space N^I is no more a compact metric space. But any Borel probability measure defined in a complete separable metric space is regular. This in turn implies that the whole space can be approximated by a compact subspace, the probability of this subspace being as close to unity as wanted. The ideas of the extension of the theory to noncompact metric spaces are given in [26].

For the abstract measurable spaces the ergodic decomposition was obtained by Blum and Hanson [5]. Their proof did not use convexity arguments. However, the sufficient conditions assuring their proof is working involve convexity assumptions (cf. the condition (c) below).

Let $\mathscr{S}_0$ denote the subσ-field of $\mathscr{S}$ consisting of all T-invariant sets; in symbols

$$\mathscr{S}_0 = \{E : E \in \mathscr{S},\ T^{-1}E = E\}\,.$$

Let $M(\mathscr{S})$ denote the set of all T-invariant probability measures on $\mathscr{S}$. The usual Kolmogorov σ-field in $M(\mathscr{S})$ will be designated by the symbol $\mathscr{K}[M(\mathscr{S})]$, thus

$$\mathscr{K}[M(\mathscr{S})] = \sigma(\{\{\mu : \mu \in M(\mathscr{S}),\ \mu(A) \leqq t\} : 0 \leqq t \leqq 1,\ A \in \mathscr{S}\})\,. \tag{6.2}$$

Let $E(\mathscr{S})$ denote the set of all T-ergodic probabilities in $M(\mathscr{S})$. The aim of the ergodic decomposition is given $\mu \in M(\mathscr{S})$ to find a probability measure $\hat{\mu}$ on the σ-field

$$\mathscr{K}[E(\mathscr{S})] = E(\mathscr{S}) \cap \mathscr{K}[M(\mathscr{S})]$$

such that

$$\mu(E) = \int_{E(\mathscr{S})} \nu(E)\,\hat{\mu}(\mathrm{d}\nu)\,. \tag{6.3}$$

The sufficient conditions for the performability of the Blum-Hanson proof are the following ones (cf. [9]):

(a) let $\mu_1, \mu_2 \in M(\mathscr{S})$. The assumption that

$$\forall E \in \mathscr{S}_0\ \mu_1(E) = \mu_2(E)$$

implies the relations

$$\forall E \in \mathscr{S}\ \mu_1(E) = \mu_2(E)\,;$$

(b) $\forall A \in \mathscr{S}\ \forall 0 \leqq t \leqq 1\ \exists A_t \in \mathscr{S}_0$ such that

$$\{\nu : \nu \in E(\mathscr{S}),\ \nu(A) \leqq t\} = \{\nu : \nu \in E(\mathscr{S}),\ \nu(A_t) = 1\}\,;$$

(c) if $\forall \nu \in E(\mathscr{S})$, $\nu(A) = 0$ then $\mu(A) = 0$ for any μ, $\mu \in M(\mathscr{S})$.

The conditions (a) and (b) are proved in [5] to be the immediate corollaries to the individual ergodic theorem. The condition (c) depends, in general, on the topological properties of the space Ω.

Remark 6.1. The decomposition theorem in [5] was actually stated in a form seemingly different from the required (6.3). Indeed, the decomposing probability measure $\hat{\mu}$ was defined on the σ-field

$$\mathscr{R} = \{\{\nu : \nu \in E(\mathscr{S}),\ \nu(E) = 1\} : E \in \mathscr{S}_0\}\,.$$

On the other hand, using the condition (b) it can be easily shown that

$$\mathscr{R} = \mathscr{K}[E(\mathscr{S})]\,.$$

7. Stone Spaces and the Ergodic Decomposition

Let Ω be a compact Hausdorff totally disconnected space. This means the field $\mathscr{C}$ of all clopen sets in Ω constitutes the basis for its topology. The space Ω is called the Stone space. Actually, it is the Stone space of its own basis $\mathscr{C}$ and $\mathscr{C}$ is its own Stone field (cf. e.g. [31]). Let U denote an automorphism of the field $\mathscr{C}$. Then there is a homeomorphism h of the space Ω inducing U:

$$UC = h^{-1}C\,, \quad C \in \mathscr{C}$$

[31]. As well-known, $\mathscr{S} = \sigma(\mathscr{C})$ is the Baire σ-field in Ω (cf. e.g. [13]). Since both h and h^{-1} map Baire sets into Baire sets, the automorphism U can be extended to a σ-automorphism (denoted by the same symbol U) of the σ-field $\mathscr{S}$. Let $M(\mathscr{S})$, $\mathscr{K}[M(\mathscr{S})]$, $E(\mathscr{S})$ have the same meaning as in the preceeding section. In the present context, the set $E(\mathscr{S})$ of all ergodic measures coincides with the set of all indecomposable ones (cf. [30]).

Lemma 7.1. If $\mathscr{S}$ is the Baire σ-field of a Stone space Ω, if U is an σ-automorphism of the σ-field $\mathscr{S}$, then the condition (c) of Section 6 is satisfied.

The assumptions of the lemma imply that any $\mu \in M(\mathscr{S})$ is a convex linear combination of the ergodic measures. The lemma follows immediately. The proof of the statement follows the lines of the argument given by Chocquet as described e.g. in [30].

Theorem 7.2. Let Ω be a Stone space. For every continuous function f on Ω the integral $\int f \mathrm{d}\nu$ is $\mathscr{K}[E(\mathscr{S})]$-measurable function of the variable ν on $E(\mathscr{S})$. For every measure $\mu \in M(\mathscr{S})$ there is a unique probability measure $\hat{\mu}$ on the σ-field $\mathscr{K}[E(\mathscr{S})]$ such that

$$\int_{\Omega} f \,\mathrm{d}\mu = \int_{E(\mathscr{S})} \left[\int_{\Omega} f \,\mathrm{d}\nu\right] \hat{\mu}(\mathrm{d}\nu)\,. \tag{7.1}$$

Proof. 1. Let $C \in \mathscr{C}$. C is a clopen set, thus its indicator function is continuous. Using the method of Blum and Hanson (which is working according to the Lemma 7.1) there is a unique probability measure $\hat{\mu}$ on the σ-field $\mathscr{K}[E(\mathscr{S})]$ such that

$$\mu(C) = \int_{\Omega} \chi_C \,\mathrm{d}\mu = \int_{E(\mathscr{S})} \left[\int_{\Omega} \chi_C \,\mathrm{d}\nu\right] \hat{\mu}(\mathrm{d}\nu) = \int_{E(\mathscr{S})} \nu(C)\, \hat{\mu}(\mathrm{d}\nu)\,. \tag{7.2}$$

2. The rest of the proof consists of the extension of (7.2) to the space of all continuous functions on Ω. The proof involves only the standard arguments including the Stone's general version of the Weierstrass theorem and the limit theorems of the Lebesgue integral.

We shall use the theorem to find the ergodic decomposition of the finitely additive invariant probabilities. The result was obtained in [25]. Let Ω denote an abstract

nonvoid set, let $\mathscr{A}$ designate a field of subsets of Ω. The symbol T will be used to denote an automorphism of the space $(\Omega, \mathscr{A})$, i.e. a one-to-one mapping of Ω onto Ω such that $T^{-1}\mathscr{A} \subset \mathscr{A}$ and $T\mathscr{A} \subset \mathscr{A}$, respectively. An additive T-invariant probability μ on the field $\mathscr{A}$ is said to be ergodic, if there is no sequence $\{A_n\}_{n=1}^{\infty} \subset \mathscr{A}$ and no δ positive such that the following three conditions are satisfied:

$$(7.3) \qquad (1) \quad \lim_{m,n} \mu(A_n \Delta A_m) = 0 ,$$

$$(2) \quad \lim_{n} \mu(A_n \Delta T^{-1} A_n) = 0 .$$

$$(3) \quad \delta < \mu(A_n) < 1 - \delta , \quad n \in N .$$

Theorem 7.3 [25]. The set $M(\mathscr{A})$ of all T-invariant additive probabilities on $\mathscr{A}$ is nonvoid. To every $\mu \in M(\mathscr{A})$ there corresponds a unique σ-additive probability measure $\hat{\mu}$ on the Kolmogorov σ-field $\mathscr{K}[E(\mathscr{A})]$ such that

$$(7.4) \qquad \mu(A) = \int_{E(\mathscr{A})} \nu(A)\, \hat{\mu}(\mathrm{d}\nu) .$$

The main idea of the proof consists in the investigation of a Stone space such that the σ-additive Baire probabilities correspond to the finitely additive probabilities on the field $\mathscr{A}$. Olshen [25] used the Stone representation of a Boolean algebra. The proof contains an interesting equality $E(\mathscr{C}) = E(\mathscr{S})$, where $\mathscr{C}$ is the field of all clopen sets in the corresponding Stone space and $\mathscr{S} = \sigma(\mathscr{C})$, respectively. We shall prove this statement in a more general form.

Proposition 7.4. Let Ω be an abstract nonvoid set, let $\mathscr{A}$ be a field of subsets of Ω and let $\mathscr{S} = \sigma(\mathscr{A})$. Let T be a σ-automorphism of the σ-field $\mathscr{S}$. Then $E(\mathscr{A}) = E(\mathscr{S})$, i.e. for σ-additive measures, a measure is ergodic in the sense of (7.3) if and only if it is ergodic in the usual sense (cf. (1.11) and (1.12)).

Proof. 1. Let $\mu \notin E(\mathscr{S})$. Then there is an invariant set E_0 and a number c_0, $0 < c_0 < 1$ such that $\mu(E_0) = c_0$. Given $\varepsilon > 0$ there is a set $A \in \mathscr{A}$ possessing the property that $\mu(E_0 \Delta A) < \varepsilon$ (cf. [23], § 13, Theorem D). Consequently, there is a sequence $\{A_n\}_{n=1}^{\infty} \subset \mathscr{A}$ with the property

$$\mu(A_n \Delta E_0) < 1/n ,$$

i.e.

$$c_0 - 1/n < \mu(A_n) < c_0 + 1/n .$$

Assume $n_0 \in N$ is large enough to satisfy the inequalities

$$0 < c_0 - 1/n < c_0 + 1/n < 1 .$$

Let $B_n = A_{n_0+n-1} (n = 1, 2, \ldots)$. Then

$$0 < c_0 - 1/n < \mu(B_n) < c_0 + 1/n < 1 , \quad n = 1, 2, \ldots . \tag{7.5}$$

Since $\lim_n \mu(B_n \Delta E_0) = 0$, we have

$$\lim_{m,n} \mu(B_n \Delta B_m) = 0 . \tag{7.6}$$

Actually, if we define $\varrho(A, B) = \mu(A \Delta B)$, then ϱ is a pseudometric on $\mathscr{S}$, hence a convergent sequence is a Cauchy sequence. Further

$$\mu(TB_n \Delta E_0) = \mu(TB_n \Delta TE_0) = \mu(T(B_n \Delta E_0)) = \mu(B_n \Delta E_0) ,$$

therefore the triangle inequality yields

$$\lim_n \mu(TB_n \Delta B_n) = 0 . \tag{7.7}$$

The relations (7.5), (7.6) and (7.7) imply that $\mu \notin E(\mathscr{A})$. Hence $E(\mathscr{A}) \subset E(\mathscr{S})$.

2. Conversely, if $\mu \notin E(\mathscr{A})$, there are $\delta > 0$ and a sequence $\{A_n\}_{n=1}^{\infty} \subset \mathscr{A}$ such that

$$\lim_{m,n} \mu(A_n \Delta A_m) = 0 ,$$

$$\lim_n \mu(A_n \Delta TA_n) = 0 ,$$

$$\delta < \mu(A_n) < 1 - \delta \quad (n = 1, 2, \ldots) .$$

Let $A = \limsup_n T^n A_n$. Then $A \in \mathscr{S}$ and $\mu(A \Delta TA) = 0$. On the other hand, $0 < \delta \leqq \mu(A) \leqq 1 - \delta < 1$. Hence $\mu \notin E(\mathscr{S})$. The proof is finished.

8. On the Concept of Ergodicity

Theorem 7.3 can be proved also by representing additive not necessarily σ-additive measures as the elements of the space L_∞^* corresponding to some L_∞ space. This idea is due to Hewitt and Yosida [14]. This alternative approach can be used to prove Theorem 7.3 by means of the general Chocquet's representation theorem. For σ-additive measures this was done by Feldman (cf. [30]). We shall not give the alternative proof of the decomposition theorem, but we shall use the general method to obtain some interesting results concerning the ergodic finitely additive probabilities.

Let $(\Omega, \mathscr{F}, m)$ be a σ-finite measure space. The symbol $M(m)$ [$M^+(m)$] will denote the set of all bounded signed [positive] additive set functions on the σ-field $\mathscr{F}$ vanishing on every set $E \in \mathscr{F}$ for which $m(E) = 0$. Let us consider the linear space $L_\infty(m)$ of all equivalence classes of the essentially bounded (with respect to the measure m)

$\mathscr{F}$-measurable real functions. Let U denote the isometric isomorphism of the space $M(m)$ into the normed conjugate $L_\infty^*(m)$ of $L_\infty(m)$ (cf. [14], Theorem 2.3). Let

$$B = \{f^* : f^* \in L_\infty^*(m),\ \|f^*\| = 1, f^* \geqq 0, f^*(1) = 1\}\,. \tag{8.1}$$

Then B is a convex w^*-compact set. By Krein-Milman theorem (cf. e.g. [8]) B is the closed convex hull of the set ∂B of its extremal points, in symbols

$$B = \overline{\mathrm{co}}\,(\partial B)\,.$$

The set ∂B is closed in B, hence a compact subset of B. We shall denote it by the symbol S. Thus S is a compact Hausdorff space. Let us recall the well-known characterization of the space S:

$$f^* \in S \quad \text{iff} \quad \forall E \in \mathscr{F} \quad f^*\chi_E \in \{0, 1\}\,. \tag{8.2}$$

The mapping $V : L_\infty(m) \to C(S)$ defined by

$$(Vf)(f^*) = f^*(f)\,, \quad f \in L_\infty(m) \tag{8.3}$$

is an order preserving isometric isomorphism (cf. [14], Theorems 4.2 and 4.3). Clearly, by (8.3) and (8.2), to every set $E \in \mathscr{F}$ there corresponds a clopen set $\tilde{E} \subset S$ such that

$$V\chi_E = \chi_{\tilde{E}}\,.$$

Let

$$\mathscr{A} = \{\tilde{E} : \tilde{E} \subset S,\ \chi_E = V\chi_E \text{ for some } E \in \mathscr{F}\}\,.$$

The σ-field $\sigma(\mathscr{A})$ is the Baire σ-field in S [14]. The adjoint mapping $V^* : C^*(S) \to L_\infty^*(m)$ is an isometric isomorphism, too [15]. Consequently, the composition $U \circ V^{*-1}$ is an isometric isomorphism of the space $M(m)$ onto the set $C^*(S)$ of all Radon measures on S.

Let T denote an automorphism of the measurable space $(\Omega, \mathscr{F})$. The corresponding linear operator $\overline{T} : L_\infty(m) \to L_\infty(m)$ is defined by the property

$$(\overline{T}f)(\omega) = f(T\omega)\,, \quad f \in L_\infty(m)\,. \tag{8.4}$$

Let T_C be induced from $\overline{T}$ by the isomorphism V:

$$T_C(Vf) = V(\overline{T}f)\,, \quad f \in L_\infty(m)\,.$$

Since $VL_\infty(m) = C(S)$, T_C maps $C(S)$ into $C(S)$. By the symbol $\tilde{T}_1$ we shall denote the adjoint of T_C, i.e.

$$(\tilde{T}_1 h^*)(h) = h^*(T_C h)\,, \quad h^* \in C^*(S)\,.$$

A transformation of the space $C^*(S)$ corresponding to the automorphism T can be introduced also in another natural way. Let $\overline{T}^*$ be the adjoint of $\overline{T}$, i.e.

$$(\overline{T}^* f^*)(f) = f^*(\overline{T}f), \quad f^* \in L_\infty^*(m).$$

Let $\tilde{T}_2$ be induced from $\overline{T}^*$ by the isomorphism V^{*-1}:

$$\tilde{T}_2(V^{*-1}f^*) = V^{*-1}(\overline{T}^* f^*), \quad f^* \in L_\infty^*(m).$$

However, both constructions give the same result in the sense of the following lemma. Let

$$M_1^+(m, T) = \{\mu : \mu \in M^+(m),\ \mu(\Omega) = 1,\ \mu = \mu T^{-1}\}. \tag{8.5}$$

Lemma 8.1. The following three statements are equivalent:

(1) $\mu \in M_1^+(m, T)$,

(2) $V^{*-1}(U\mu) = \tilde{T}_1 V^{*-1}(U\mu)$,

(3) $V^{*-1}(U\mu) = \tilde{T}_2 V^{*-1}(U\mu)$.

The proof is elementary and therefore it is omitted.

Assume we are given an arbitrary Radon measure $h^* \in C^*(S)$ on S. Then there is a unique signed Baire measure Wh^* on S such that

$$h^*(h) = \int_S h(s)\,(Wh^*)(\mathrm{d}s), \quad h \in C(S) \tag{8.6}$$

(cf. [8]-Riesz representation theorem). Let Z denote the composition of U, V^{*-1} and W in this order, i.e.

$$Z\mu = W[V^{*-1}(U\mu)], \quad \mu \in M(m). \tag{8.7}$$

In [14] it is proved that each bounded positive finitely additive set function on a ring of sets uniquely decomposes into the σ-additive part and a pure charge. Let us recall that a finitely additive set function is said to be a pure charge provided that any σ-additive positive measure majorized by it vanishes everywhere. The correspondence between the Hewitt-Yosida and the Lebesgue decompositions is established in the following theorem.

Theorem 8.2. Let $\mu \in M^+(m)$. Let

$$Z\mu = \mu_1 + \mu_2 \tag{8.8}$$

be the Lebesgue decomposition of the measure $Z\mu$ into the Zm-absolutely continuous part μ_1 and a Zm-singular part μ_2, respectively. Then the formula

$$\mu = Z^{-1}\mu_1 + Z^{-1}\mu_2 \tag{8.9}$$

establishes the Hewitt-Yosida decomposition of the set function μ. $Z^{-1}\mu_1$ is the σ-additive part and $Z^{-1}\mu_2$ is the pure charge, respectively.

Proof. Let the formula (8.8) does not establish the Hewitt-Yosida decomposition. Then there are nontrivial Hewitt-Yosida decompositions

$$Z^{-1}\mu_i = \mu_i^c + \mu_i^t \, (i = 1, 2), \tag{8.10}$$

where μ_i^c is the σ-additive part and μ_i^t is the pure charge corresponding to $Z^{-1}\mu_i$ $(i = 1, 2)$. Now $Z\mu_i^c$ is Zm-absolutely continuous and $Z\mu_i^t$ is Zm-singular; in symbols

$$Z\mu_i^c \ll Zm\,, \quad Z\mu_i^t \perp Zm \tag{8.11}$$

(cf. [15]). Let us consider $i = 1$. By the assumption, $\mu_1 \ll Zm$. On the other hand, μ_1 contains a Zm-singular part $Z\mu_1^t$, a contradiction. Hence $\mu_1 = Z\mu_1^c$, i.e. $Z^{-1}\mu_1 = \mu_1^c$. This proves the σ-additivity of $Z^{-1}\mu_1$. A similar argument for $i = 2$ shows that the set function $Z^{-1}\mu_2$ is a pure charge. The theorem follows because of the uniqueness of the Lebesgue decomposition.

The following lemma makes sense only for σ-additive measures. Let $(\Omega, \mathscr{F})$ be a measurable space, let T be a measurable transformation of Ω. Assume μ and ν are two T-invariant probability measures on $\mathscr{F}$.

Lemma 8.3. Let $\nu = \nu_0 + \nu_1$ be the Lebesgue decomposition of the measure ν with respect to the measure μ. Then $\nu_0 = \nu_0 T^{-1}$ and $\nu = \nu_1 T^{-1}$, respectively.

The proof follows the lines of the usual proof of the Lebesgue decomposition theorem (cf. [13]). The only new fact is that the density f of ν with respect to $\mu + \nu$ is now invariant, i.e. $f = f \circ T[\nu]$ (cf. [30], the result is due to Feldman).

Theorem 8.4. Let $(\Omega, \mathscr{F}, m)$ be a probability space with $m = mT^{-1}$, where T is an automorphism of the measurable space $(\Omega, \mathscr{F})$. Let $\mu \in M_1^+(m, T)$ be indecomposable (i.e. ergodic). Then μ is either σ-additive or a pure charge.

Remark 8.5. Given two invariant pure charges, their linear combination is again a pure charge [14]. Since a linear combination of invariant set functions is again invariant we have, by the theorem:

$$\partial M_1^+(m, T) = E(\mathscr{F}) \cup \partial \,\text{P.Ch.},$$

where the symbol on the left-hand side designates the set of all indecomposable elements in $M_1^+(m, T)$ and the symbol on the very right of the relation denotes the set of all indecomposable pure charges.

Proof of the theorem. Let us define the mapping $\tau : S \to S$ by the property that

$$h^*(h \circ \tau) = (\tilde{T}h^*)(h)\,, \quad h^* \in C^*(S)\,.$$

Here, $\tilde{T}$ denotes one of the operators $\tilde{T}_1$ and $\tilde{T}_2$ introduced above. Clearly τ is continuous and $\tilde{T}h^* = h^*$ if and only if $(Wh^*)\tau^{-1} = Wh^*$. It can be easily shown that to the indecomposable elements of $M_1^+(m, T)$ there correspond the ergodic probability measures on $(S, \sigma(\tilde{\mathscr{A}}))$. Lemma 8.3 both with the definition of the ergodicity imply that the Lebesgue decomposition of the measure $Z\mu$ corresponding to an indecomposable μ must be trivial. Hence, by Theorem 8.2, the Hewitt-Yosida decomposition of μ must be trivial. The theorem is proved.

9. The Method of Maximal Compactification

In the special case of $\Omega = X^I$ with an at most countable set X we have still another possibility for the study of the properties of finitely additive probabilities by means of the corresponding σ-additive ones. The main fact used here is that the space X^I is a complete separable metric space, which is totally disconnected in the distance function ϱ (cf. (1.9)).

Lemma 9.1. Each totally disconnected T_1-space X possesses a Hausdorff compactification.

The only thing to prove is the complete regularity of the space X. Using the fact that the indicator functions of the sets belonging to the basis in X are continuous, one can easily check the complete regularity.

The space X is actually homeomorphic with a dense subset of its maximal compactification ([17], p. 226, [8], p. 300).

Lemma 9.2. The maximal compactification of a totally disconnected space X is a totally disconnected space.

Indeed, let $\mathscr{C}$ be the basis for X. The continuous extension of the continuous function χ_C on X is the indicator function $\chi_{\bar{C}}$ of the set $\bar{C}$, which is the closure of C in the topology of the maximal compactification. To show that the family $\{\bar{C} : C \in \mathscr{C}\}$ is the basis for the maximal compactification it suffices to prove that is a field. Since $\bar{\phi} = \phi \in \bar{\mathscr{C}}$ and $\bar{\mathscr{C}}$ is closed with respect to the finite intersections, it suffices to prove the following

Lemma 9.3. Let $C \in \mathscr{C}$, let $C^c = X - C$. Then $\bar{C} \cap \bar{C}^c = \phi$.

Proof. Let $x \in \bar{C} \cap \bar{C}^c$. Since $x \in \bar{C}$, there is a net $x_\alpha \in C$ such that x_α converges to x in the topology of the maximal compactification. Similarly, since $x \in \bar{C}^c$, there is a net $y_\beta \in C^c$ such that $y_\beta \to x$. Let f denote the continuous extension of χ_C. Since $x_\alpha \to x$, $f(x_\alpha) \to f(x)$. But $f(x_\alpha) = \chi_C(x_\alpha) = 1$, hence $f(x) = 1$. Similarly, $f(y_\beta) \to f(x)$. But $f(y_\beta) = 0$, hence $f(x) = 0$, a contradiction yielding the desired conclusion.

Let us consider now the space $(N^I, \mathscr{A})$. Let βN^I denote as usually its maximal compactification. Let $\bar{\mathscr{A}}$ be the corresponding field of subsets of βN^I. For any $\bar{A} \in \bar{\mathscr{A}}$

let us define the σ-additive probability measure μ_β by the property that

$$\mu_\beta(\bar{A}) = \mu(A), \quad A \in \mathscr{A}.$$

For any $\mathscr{B} \subset \sigma(\bar{\mathscr{A}})$ let

$$\mathscr{U}(\mathscr{B}) = \{E : E \in \sigma(\bar{\mathscr{A}}), \quad \forall n\, \exists F_n \in \mathscr{B}\, \exists G_n \in \mathscr{B}, \; F_n \nearrow E, \; G_n \searrow E\}.$$

For any ordinal number α less than the first uncountable ordinal Ω we shall set

$$\mathscr{B}_\alpha = \bigcup_{\lambda<\alpha} \mathscr{U}(\mathscr{B}_\lambda), \quad \mathscr{B}_0 = \bar{\mathscr{A}}.$$

Then

$$\sigma(\bar{\mathscr{A}}) = \bigcup_{\alpha<\Omega} \Big(\bigcup_{\lambda<\alpha} \mathscr{U}(\mathscr{B}_\lambda) \Big).$$

Let us note that the latter statement represents the well-known transfinite construction of the σ-field generated by the field $\bar{\mathscr{A}}$. This way of introducing the σ-field $\sigma(\bar{\mathscr{A}})$ in βN^I allows us to define a set-transformation of the space βN^I induced by the shift T in N^I. Let

$$T_\beta \bar{A} = \overline{T^{-1}A}, \quad A \in \mathscr{A}.$$

Let

$$T^{(0)} = T_\beta \quad \text{for} \quad \mathscr{B}_0 = \bar{\mathscr{A}}.$$

If $\alpha < \Omega$ is an ordinal, we shall set

$$T^{(\alpha)}E = \begin{cases} T^{(\lambda)}E & \text{if} \quad E \in \mathscr{B}_\lambda, \lambda < \alpha; \\ \bigcup_{n=1}^{\infty} T^{(\alpha-1)}F_n & \text{if there is no } \lambda < \alpha \text{ with } E \in \mathscr{B}_\lambda; \end{cases}$$

with $F_n \in \mathscr{B}_{\alpha-1}$, $F_n \nearrow E$. Finally, let us define

$$\tilde{T}E = T^{(\alpha)}E, \quad E \in \mathscr{B}_\alpha, \quad \alpha < \Omega, \quad E \in \sigma(\bar{\mathscr{A}}).$$

Hence, by the transfinite induction, it is possible to obtain a simultaneous extension of both the measure μ_β and the transformation T_β to the (Baire) σ-field $\sigma(\bar{\mathscr{A}})$.

10. Elementary Properties of Finite Partitions

The finite partitions will be the main tool in the subsequent parts of the present paper. Therefore it is worthwile to mention some elementary facts concerning the finite partitions in advance.

Let A denote an arbitrary nonvoid set. A finite collection $\{C_1, \ldots, C_k\}$ of nonvoid pairwise disjoint subsets of the set A whose union is the whole of A is said to be a finite partition of the set A. If ζ and ξ are finite partitions of the set A, the symbol

$\zeta \vee \xi$ will be used to denote the finite partition of the same set A defined by

$$\zeta \vee \xi = \{C \cap D : C \in \zeta,\ D \in \xi\}. \tag{10.1}$$

Given two partitions ζ and ξ we shall write

$$\zeta \succ \xi \quad \text{iff} \quad \zeta \vee \xi = \zeta. \tag{10.2}$$

The relation $\succ$ partially orders the set of all finite partitions of the set A given. Let ζ be a finite partition of the set $A^k (k = 1, 2, \ldots)$. Then ζ^n will denote its n-th Cartesian power, i.e. a finite partition of the set A^{nk} consisting of all sets expressible in the form

$$C^1 \times \ldots \times C^n \ (C^i \in \zeta,\ i = 1, 2, \ldots, n).$$

The simplest properties of the finite partitions used throughout the paper are summed up in the following

Lemma 10.1. (1) Let $\zeta \subset \mathscr{A}_X$ be a finite partition of the space X^I. Then there are $i_0 \in I$, $n_0 \in N$ such that $\forall C \in \zeta\ \exists E(C) \in \mathscr{B}(X^{n_0})$ with the property that

$$C = [E(C)]_{i_0, n_0}.$$

(2) Let $\zeta \subset \mathscr{B}(X^n)$ be a finite partition of the space X^n. Let $[\zeta]_{i,n} = \{[C]_{i,n} : C \in \zeta\}$. The collection $[\zeta]_{i,n}$ is a finite partition of the space X^I for all $i \in I$.

(3) Let $\zeta \subset \mathscr{B}(X^n)$ be a finite partition of X^n. Then there is a finite partition ξ, $\xi \subset \subset \mathscr{B}(X)$ such that $\xi^n \succ \zeta$.

Let us recall that the symbol Z_X was introduced to denote the set of all finite partitions of the alphabet X into the Borel sets. If $\zeta \subset \mathscr{A}_X$ is a finite partition of the space X^I,

$$T_X^i \zeta = \{T_X^i C : C \in \zeta\}, \quad i \in I.$$

Clearly, for $\zeta \in Z_X$,

$$\bigvee_{j=0}^{n-1} T_X^{-j} [\zeta]_{0,1} = [\zeta^n]_{0,n}. \tag{10·3}$$

The following property of the partially ordered set $(Z_X, \succ)$ will be frequently used. Let $\zeta, \xi \in Z_X$. Then $\zeta \vee \xi \in Z_X$ and $\zeta \vee \xi \succ \zeta$, $\xi \vee \zeta \succ \xi$, respectively. Hence Z_X is a directed set by means of the relation $\succ$.

Now we shall introduce a general schema of the induction of the finite alphabet sources using the finite partitions of the alphabet. This way will be fixed throughout the whole paper.

Let $\zeta \in Z_X$. The notation $\zeta = \{C_1, \ldots, C_k\}$ will mean that

(1) $\operatorname{card}(\zeta) = k$,

(2) the elements of ζ are numbered in the following fixed one-to-one manner.

Let $\{x_1, x_2, \ldots\}$ be a countable dense set in X. Let $\zeta \in Z_X$ with card $(\zeta) = k$. Let

$$
\begin{aligned}
&C_1 \text{ be that } C \in \zeta \text{ for which } x_1 \in C\,,\\
&a_1 = \min\{k : x_k \in X - C_1\}\,,\\
&C_2 \text{ be that } C \in \zeta \text{ for which } x_{a_1} \in C\,,\\
&a_2 = \min\{k : x_k \in X - (C_1 \cup C_2)\}\,,\\
&\vdots\\
&C_k = X - \cup\{C_j : j = 1, \ldots, k-1\}\,.
\end{aligned}
\tag{10.4}
$$

Given $\zeta = \{C_1, \ldots, C_k\} \in Z_X$ we shall define the mapping $\tau_\zeta : N^I \to \{1, \ldots, k\}^I$ by the property that

$$
(\tau_\zeta z)_i = j \quad \text{if} \quad z_i \in C_j \quad \text{for} \quad z \in X^I\,, \quad i \in I\,. \tag{10.5}
$$

Since $\tau_\zeta^{-1}\mathscr{A}_k \subset \mathscr{A}_X$, the mapping is measurable. Hence given a stationary source, i.e. a finitely additive shift-invariant probability μ on the field $\mathscr{A}_X$ the notation $\mu\tau_\zeta^{-1}$ makes sense. Clearly

$$
\mu\tau_\zeta^{-1}(\{1, \ldots, k\}^I) = 1\,.
$$

Thus for any $\zeta \in Z_X$ the induced measure $\mu\tau_\zeta^{-1}$ is a σ-additive probability on $\mathscr{A}_k$. We shall denote by $\bar{\mu}_\zeta$ its unique extension to the σ-field $\mathscr{F}_k = \sigma(\mathscr{A}_k)$. Clearly $(\mu\tau_\zeta^{-1})\, T_k^{-1} = \mu\tau_\zeta^{-1}$. The ergodicity properties are related in the following

Lemma 10.2. The source $\mu \in M(\mathscr{A}_X)$ is indecomposable if and only if for every $\zeta \in Z_X$ the finite alphabet stationary source $\bar{\mu}_\zeta$ is ergodic with respect to the transformation $T_{\text{card}(\zeta)}$.

In one direction the proof is trivial because of the coincidence of the concepts of the indecomposability and of the ergodicity for σ-additive probabilities. The proof in the opposite direction follows using the method of the proof of Proposition 7.4.

PART III: RATES ASSOCIATED WITH A SOURCE

The coding theorem was established in [40] and [41] (cf. also Theorem 2.2′ or Theorem 2.2 of the present paper). The aim of this part is to study the quantities called the entropy and the asymptotic rates, respectively, in a more general setup, i.e. for finitely additive sources with an abstract alphabet X. This means that the symbol X will denote an uncountable separable metric space. The basic notions are modified. For the special case $X = N$ they will be shown to give the original quantites. Hence our approach constitues an alternative point of view concerning the entropy and the asymptotic rates, respectively.

11. The Notion of the Entropy Rate

Let $X = N$. Let μ denote a σ-additive T-invariant probability measure on the σ-field $\mathscr{F} = \sigma(\mathscr{A})$. The relation (4.1) can be rewritten in the form

$$H(\mu) = -\lim_{n\to\infty} \frac{1}{n} \sum_{\bar{x}\in N^n} \mu[\bar{x}] \log \mu[\bar{x}] \tag{11.1}$$

(cf. (1.3) and (1.4)). Here the symbol log means $\log_2$ and we shall adopt the usual convention $0 \cdot \log 0 = 0$. As mentioned earlier, the decisive rôle plays the family

$$\{[\bar{x}] : \bar{x} \in N^n\}$$

being the n-th Cartesian power of a generator. Since for an uncountable metric space this fact fails to hold, the relation (11.1) cannot be used to define the entropy rate. Even for $X = N$ it is impossible to use the relation (11.1) for finitely additive probabilities on $\mathscr{A}$, because the above family is not finite. Hence we are forced to modify the concept of the entropy rate.

Let $\zeta \in Z_X$ (cf. Section 10). Then define

$$H(\mu, \zeta) = -\lim_{n\to\infty} \frac{1}{n} \sum_{E\in\zeta^n} \mu[E]_{i,n} \log \mu[E]_{i,n} \tag{11.2}$$

where μ is any T_X-invariant finitely additive probability on the field $\mathscr{A}_X$ of all finite-dimensional cylinders in X^I. Since $\mu = \mu T_X^{-1}$, we have $\mu[E]_{i,n} = \mu[E]$ (cf. (1.3)), thus the right-hand side of (11.2) does not depend on $i \in I$. The *entropy rate* $H(\mu)$ is then defined as the supremum

$$H(\mu) = \sup \{H(\mu, \zeta) : \zeta \in Z_X\} . \tag{11.3}$$

Note that by (10.3) we have

$$H(\mu, \zeta) = -\lim_{n\to\infty} \frac{1}{n} \sum \mu(E) \log \mu(E) ,$$

where the summation is taken over all sets E belonging to the finite family $\bigvee_{j=0}^{n-1} T_X^{-j}[\zeta]_{0,1}$. This fact together with the elementary properties of the finite partitions as stated in Lemma 10.1 show that $H(\mu)$ is nothing but Sinaj's modification of the concept of entropy of the automorphism T_X (cf. [35]).

The entropy rate of a finitely additive source can assume any value from the interval $[0, \infty]$ (cf. also Remark 12.2 below). The following two examples illustrate the extremal cases.

Example 11.1. Let $\mathscr{F}$ denote the ultrafilter in $\mathfrak{P}(N)$ containing the filter of all complements of finite subsets of the set N. Let μ be the pure charge on $\mathfrak{P}(N)$ defined

by the properties that

$$\mu(E) = \begin{cases} 1 & \text{if} \quad E \in \mathscr{F}, \\ 0 & \text{if} \quad E \notin \mathscr{F}. \end{cases}$$

Let $\bar{\mu}$ denote the discrete memoryless source uniquely determined by the pure charge μ, i.e.

$$\bar{\mu} = \prod_{i \in I} \mu_i ; \quad \mu_i = \mu \quad \text{for} \quad i \in I .$$

(The infinite-dimensional product of finitely additive probabilities exists and it is again a finitely additive probability [8].) Clearly, $\bar{\mu}(E) \in \{0,1\}$ for all $E \in A$. Hence, for any ζ, $\zeta \in Z$,

$$H(\bar{\mu}, \zeta) = H(\zeta) = -\sum_{E \in \zeta} \mu(E) \log \mu(E) = 0 .$$

The left-hand side equality follows from the fact that $\bar{\mu}$ is a memoryless source. Finally,

$$H(\bar{\mu}) = \sup_{\zeta \in Z} H(\bar{\mu}, \zeta) = 0 .$$

Example 11.2. Let $\mu = P(\bar{h}_z, \mathbf{P})$ be a stationary Markov source (cf. Section 5). We shall show that $H(\mu) = \infty$ by showing that for any given positive integer k there is a finite partition ζ_k of the set N such that

$$H(\mu, \zeta_k) = \log k .$$

Let $k \in N$ be given. Consider the first k^2 positive integers and divide them into k disjoint classes constituted by the different rows of the following schema:

$$\begin{array}{llllll} 1 & 2 & 4 & 6 & \dots & 2k-2 \\ 3 & 2k & 2k+1 & 2k+3 & \dots & \\ 5 & 2k+2 & & & & \\ 7 & 2k+4 & & & \vdots & \vdots \\ \vdots & \vdots & & \dots & k^2-3 & k^2-2 \\ 2k-1 & & & \dots & k^2-1 & k^2 . \end{array}$$

Then continue in the same way with the subsequent k^2 positive integers. The juxtaposition of the (infinitely many) subsequent partitions yields an infinite schema with exactly k rows. The rows are identified with the new states, say $S_1, S_2, \dots, S_k$. Now

$$\lim_{n \to \infty} \frac{\operatorname{card} \{m : 1 \leqq m \leqq n, m \in S_i\}}{\operatorname{card} \{m : 1 \leqq m \leqq n, m \in S_j\}} = 1$$

for all pairs (i, j), hence

$$q_i = q(S_i) = 1/k\,; \quad i = 1, 2, \ldots, k$$

(cf. also Proposition 5.3). From the construction of the partition it follows that the one-step transition probability matrix $\mathbf{Q}$ of the collapsed process is

$$\mathbf{Q} = \begin{pmatrix} 1/k & 1/k & \ldots & 1/k \\ \vdots & \vdots & & \vdots \\ 1/k & 1/k & \ldots & 1/k \end{pmatrix}.$$

Therefore the collapsed process is an ergodic Markov chain (cf. Proposition 5.3 and Lemma 10.2). A well-known result concerning the entropy rate of the ergodic Markov chains gives

$$H(\mu, \zeta_k) = \sum_{i=1}^{k} q_i H(q_{i1}, \ldots, q_{ik}) = \log k$$

(cf. [2]). Consequently,

$$H(\mu) = \sup_{\zeta \in Z} H(\mu, \zeta) \geqq \sup_{k} H(\mu, \zeta_k) = \infty\,.$$

12. The Integral Representation of the Entropy Rate

The results of this Section are contained in the author's paper [37]. Here we shall give only the formulation of the main result and several comments concerning its proof. Note that the result and the proof in [37] were given in the case $X = N$. However, they can be transmitted to the general case of a separable metric space X without any effort.

Theorem 12.1. The entropy rate $H(\mu)$ of a finitely additive T_X-invariant probability μ defined on the field $\mathscr{A}_X$ can be represented by an integral in the form

$$H(\mu) = \int_{E(\mathscr{A}_X)} H(\nu)\, \hat{\mu}(\mathrm{d}\nu)\,. \tag{12.1}$$

(cf. Theorem 7.3).
In [37] it was proved the relation

$$H(\bar{\mu}_\zeta) = \int_{E(\mathscr{A}_X)} H(\bar{\nu}_\zeta)\, \hat{\mu}(\mathrm{d}\nu)\,, \quad \zeta \in Z_X \tag{12.2}$$

by means of Theorem 7.3. The remaining part of the proof was devoted to the justification of a general form of the limit theorem concerning with nets instead of the sequences. The direct method used in [37] can be replaced by the following considerations.

1. $H(\mu) = \infty$. Then the monotone net $\{H(\bar{\mu}_\zeta)\}_{\zeta \in Z_X}$ has no upper bound. By (12.2), the monotone net

$$\left\{\int_{E(\mathscr{A}_X)} H(\bar{\nu}_\zeta)\ \hat{\mu}(\mathrm{d}\nu)\right\}_{\zeta \in Z_X}$$

has no upper bound. Consequently, the monotone net $\{H(\bar{\nu}_\zeta)\}_{\zeta \in Z_X}$, with $\hat{\mu}$-probability positive, has no upper bound. Let the symbol $\lim_{Z_X}$ denote the limit of a net indexed by the elements of the directed set Z_X. Then

$$H(\mu) = \sup_{\zeta \in Z_X} H(\bar{\mu}_\zeta) = \lim_{Z_X} H(\bar{\mu}_\zeta)$$

and

$$\lim_{Z_X} H(\bar{\nu}_\zeta) = \sup_{\zeta \in Z_X} H(\bar{\nu}_\zeta) = H(\nu) = \infty$$

with positive $\hat{\mu}$-probability. Hence

$$\lim_{Z_X} \int_{E(\mathscr{A}_X)} H(\bar{\nu}_\zeta)\, \hat{\mu}(\mathrm{d}\nu) = \infty.$$

Consequently

$$\lim_{Z_X} H(\bar{\mu}_\zeta) = \lim_{Z_X} \int_{E(\mathscr{A}_X)} H(\bar{\nu}_\zeta)\, \hat{\mu}(\mathrm{d}\nu) = \int_{E(\mathscr{A}_X)} \lim_{Z_X} H(\bar{\nu}_\zeta)\, \hat{\mu}(\mathrm{d}\nu),$$

both sides being infinite. Thus the relation (12.1) follows in this case.

2. Let $H(\mu) = H < \infty$. Then the monotone increasing net $\{H(\bar{\mu}_\zeta)\}_{\zeta \in Z_X}$ is bounded from above by the constant H. Let $\mathscr{A}$ denote the field of subsets of Z_X consisting of all initial segments with respect to the relation $\succ$ and their complements. Let

$$p(B) = \begin{cases} 1 & \text{if } B \text{ is the complement of an initial segment,} \\ 0 & \text{if } B \text{ is an initial segment;} \quad B \in \mathscr{A}. \end{cases}$$

Then

$$\int_{Z_X} x_\zeta\, p(\mathrm{d}\zeta) = \lim_{Z_X} x_\zeta$$

for any convergent net $\{x_\zeta\}_{\zeta \in Z_X}$ or real numbers (cf. Proposition 3.3 and its proof). Hence, it suffices to prove the following Fubini theorem-like relation

$$\int_{Z_X} \left[\int_{E(\mathscr{A}_X)} H(\bar{\nu}_\zeta)\, \hat{\mu}(\mathrm{d}\bar{\nu})\right] p(\mathrm{d}\zeta) = \int_{E(\mathscr{A}_X)} \left[\int_{Z_X} H(\bar{\nu}_\zeta)\, p(\mathrm{d}_\zeta)\right] \hat{\mu}(\mathrm{d}\nu).$$

The last relation follows immediately from a general form of the Fubini theorem for finitely additive set functions [24].

Remark 12.2. Let $\{\alpha_i\}$ be any sequence such that

$$\text{(a)} \quad \alpha_i \geqq 0\,, \quad \sum_{i=1}^{\infty} \alpha_i = 1\,,$$

$$\text{(b)} \quad H(\{\alpha_i\}_{i=1}^{\infty}) = -\sum_{i=1}^{\infty} \alpha_i \log \alpha_i < \infty\,.$$

Let $\{\mu^{(i)}\}_{i=1}^{\infty}$ be any sequence of stationary σ-additive sources, all with the same at most countable alphabet X.

Let

$$\mu_{0,n}^{(i)}(E) = \mu^{(i)}[E]\,, \quad E \in \mathscr{B}(X^n)$$

(cf. (1.3)). Let $H(\mu_{0,n}^{(i)})$ denote the entropy of the probability distribution $\mu_{0,n}^{(i)}$ on $\mathscr{B}(X^n) = \mathfrak{P}(X^n)$. Then we have

$$H(\mu^{(i)}) = \lim_{n\to\infty} \frac{1}{n} H(\mu_{0,n}^{(i)})\,.$$

The assumption (b) together with the latter relation and the inequalities

$$\sum_{i=1}^{\infty} \alpha_i H(\mu_{0,n}^{(i)}) \leqq H(\sum_{i=1}^{\infty} \alpha_i \mu_{0,n}^{(i)}) \leqq \sum_{i=1}^{\infty} \alpha_i H(\mu_{0,n}^{(i)}) + H(\{\alpha_i\}_{i=1}^{\infty})$$

yield the relation

$$H(\sum_{i=1}^{\infty} \alpha_i \mu^{(i)}) = \sum_{i=1}^{\infty} \alpha_i H(\mu^{(i)})\,. \tag{12.3}$$

This fact is well-known. Now let $\{\mu^{(i)}\}_{i=1}^{\infty}$ be any sequence of the finitely additive sources, all defined on the same field $\mathcal{A}_X$. Let $\zeta \in Z_X$. We shall set

$$H_n(\mu, \zeta) = -\sum \mu(E) \log \mu(E)\,, \quad E \in \bigvee_{j=0}^{n-1} T_X^{-j}[\zeta]_{0,\cdot}\,.$$

Then we have the inequalities

$$\sum_{i=1}^{\infty} \alpha_i H_n(\mu^{(i)}, \zeta) \leqq H_n(\sum_{i=1}^{\infty} \alpha_i \mu^{(i)}, \zeta) \leqq \sum_{i=1}^{\infty} \alpha_i H_n(\mu^{(i)}, \zeta) + H(\{\alpha_i\}_{i=1}^{\infty})\,.$$

Hence

$$\sum_{i=1}^{\infty} \alpha_i H(\mu^{(i)}, \zeta) \leqq H(\sum_{i=1}^{\infty} \alpha_i \mu^{(i)}, \zeta) \leqq \sum_{i=1}^{\infty} \alpha_i H(\mu^{(i)}, \zeta)\,.$$

Consequently, for any $\zeta \in Z_X$,

$$\sum_{i=1}^{\infty} \alpha_i H(\mu^{(i)}, \zeta) = H(\sum_{i=1}^{\infty} \alpha_i \mu^{(i)}, \zeta)\,.$$

But this in turn implies that the relation (12.3) takes place even in this general setup. This fact can be used to obtain finitely additive sources with any nonnegative value of the entropy rate. Indeed, let us consider the pure charge of Example 11.1. Given any $\alpha \in (0, 1)$ and given any $h \in [0, \infty]$ let us take a discrete memoryless σ-additive source $\mu^{(1)}$ with $H(\mu^{(1)}) = h/\alpha$. Denoting the pure charge by $\mu^{(2)}$, we have

$$H(\alpha\mu^{(1)} + (1 - \alpha)\,\mu^{(2)}) = \alpha\, H(\mu^{(1)}) + (1 - \alpha)\, H(\mu^{(2)}) = h\,.$$

13. A Further Property of the Entropy

The construction of the associated σ-additive Baire probability measure to any given finitely additive probability, especially to any given finitely additive information source leads to the seemingly evident fact that there is no need to study finitely additive probabilities. This opinion is supported also by the result of this section.

We shall employ the notations used in Section 8. Let $\mu \in M_1^+(m, T)$. The entropy in the sense of Sinaj [35] is the supremum

$$H(\mu) = \sup_{\zeta} \left[- \lim_{n \to \infty} \frac{1}{n} \sum_{E} \mu(E) \log \mu(E) \right].$$

Here, the sum is taken over the elements E of the finite partition $\bigvee_{i=0}^{n-1} T^{-i}\,\zeta$ and the supremum is taken over all finite partitions ζ such that $\zeta \subset \mathscr{F}$. The corresponding Baire probability measure si denoted by $Z\mu$ and its entropy by $H(Z\mu)$.

Proposition 13.1. For any $\mu \in M_1^+(m, T)$ we have

$$H(\mu) = H(Z\mu)\,. \tag{13.1}$$

The σ-field $\mathscr{S}$ in the corresponding compact Hausdorff space S is generated by the field $\mathscr{A}$ of all clopen sets in S. As well-known,

$$\sup_{\zeta \subset \mathscr{A}} H(Z\mu, \zeta) = \sup_{\zeta \subset \mathscr{S}} H(Z\mu, \zeta)\,.$$

Using this equality and the construction of $Z\mu$ as given in Section 8 it is easy to obtain (13.1).

Remark 13.2. The representation of the finitely additive probabilities by means of the space $L_\infty^*(m)$ has one very serious disadvantage. The space $L_\infty(m)$ is not separable, hence the corresponding space $C(S)$ is not separable. This means that the compact space S itself cannot be metrizable (cf. e.g. [8]).

Thus using the $C^*(S)$ representation of $L_\infty^*(m)$ we obtain better properties of the set functions (namely the σ-additivity), but, on the other hand, we loose the good properties of the basic space. Especially, when considering the metric space X^I which is a Lebesgue space [32], the corresponding space S is not a Lebesgue space. Indeed, the σ-field $\mathscr{S}$ is not more countably generated, hence the approximation arguments (cf. Theorem 4.7) fail to hold.

14. The Asymptotic Rate

Let $\mu \in M(\mathscr{A}_X)$, let $\zeta \in Z_X$. Then we shall set

$$(14.1) \qquad (\bar{\mu}_\zeta)_n(E) = \bar{\mu}_\zeta[E]_{i,n} \quad \text{for} \quad E \subset \{1, \ldots, \operatorname{card}(\zeta)\}^n, \quad n = 1, 2, \ldots$$

(cf. Section 10 for the symbol $\bar{\mu}_\zeta$). Since $\mu = \mu T_X^{-1}$, the relation (14.1) holds independently of which $i \in I$ was chosen. In accordance with (2.3) we shall define the n-dimensional ε-length of the source $\bar{\mu}_\zeta$, in symbols $L_n(\varepsilon, \bar{\mu}_\zeta)$, by the relation

$$(14.2) \qquad L_n(\varepsilon, \bar{\mu}_\zeta) = \min \{\operatorname{card}(E) : E \subset \{1, \ldots, \operatorname{card}(\zeta)\}^n, \\ (\bar{\mu}_\zeta)_n(E) > 1 - \varepsilon\}; \quad 0 < \varepsilon < 1.$$

The coding Theorem 2.2 yields the quantity

$$(14.3) \qquad V(\bar{\mu}_\zeta) = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log L_n(\varepsilon, \bar{\mu}_\zeta).$$

In accordance with the remark finishing Section 4 we shall study the properties of the quantity

$$(14.4) \qquad V(\mu) = \sup_{\zeta \in Z_X} V(\bar{\mu}_\zeta),$$

which will be called the *asymptotic rate* of the source μ. Let us denote

$$\overline{V}(\bar{\mu}_\zeta) = \lim_{\varepsilon \to 0} \limsup_n \frac{1}{n} \log L_n(\varepsilon, \bar{\mu}_\zeta),$$

$$\underline{V}(\bar{\mu}_\zeta) = \lim_{\varepsilon \to 0} \liminf_n \frac{1}{n} \log L_n(\varepsilon, \bar{\mu}_\zeta).$$

We shall make use of the following auxiliary quantities:

$$(14.5) \qquad \overline{V}(\mu) = \sup_{\zeta \in Z_X} (\bar{\mu}_\zeta), \quad \underline{V}(\bar{\mu}_\zeta) = \sup_{\zeta \in Z_X} \underline{V}(\bar{\mu}_\zeta).$$

Let us note that for invariant μ we have $\overline{V}(\bar{\mu}_\zeta) = \underline{V}(\bar{\mu}_\zeta)$ for all $\zeta \in Z_X$ (cf. [40]), hence

$$(14.6) \qquad \overline{V}(\mu) = \underline{V}(\mu) = V(\mu), \quad \mu \in M(\mathscr{A}_X).$$

15. Basic Lemmas on the Asymptotic Rate

We proceed to the main results on the asymptotic rate by proving first two basic lemmas, which are the counterparts of Lemmas I and II in [41].

Lemma 15.1. If $\mu \in M(\mathscr{A}_X)$, if c is a finite real number, then the assumption that

$$\hat{\mu}\{v : v \in E(\mathscr{A}_X), \quad H(v) \leqq c\} = 1 \tag{15.1}$$

implies the inequality

$$\overline{V}(\mu) \leqq c\,. \tag{15.2}$$

Proof. 1. First of all we have to show that the set in (15.1) is measurable with respect to the σ-field $\mathscr{K}[E(\mathscr{A}_X)]$ (cf. (6.2)). It follows immediately from the definition of $H(v, \zeta)$ that it is a $\mathscr{K}[E(\mathscr{A}_X)]$-measurable function of the variable v on $E(\mathscr{A}_X)$. Now

$$H(v) = \sup_{Z_X} H(v, \zeta) = \lim_{Z_X} H(v, \zeta)\,.$$

If $H(v) = \infty$ then given $n \in N$ there is a partition $\zeta_n \in Z_X$ such that $H(v, \zeta_n) \geqq n$. Clearly, there are infinitely many such ζ_n's because of the inequality

$$H(v, \zeta) \leqq \log \operatorname{card}(\zeta)\,.$$

Hence $H(v) = \lim_n H(v, \zeta_n)$. If $H(v) < \infty$ then the set $\{H(v, \zeta) : \zeta \in Z_X\}$ is bounded from above by $H(v)$; $H(v)$ being exactly the least upper bound. Hence it is possible to find a sequence $\{h_n\} \subset \{H(v, \zeta) : \zeta \in Z_X\}$ converging to $H(v)$. Hence $H(.)$ is a $\mathscr{K}[E(\mathscr{A}_X)]$-measurable function of the variable v on $E(\mathscr{A}_X)$.

2. We shall show that for the set function satisfying the assumptions of the lemma the equalities

$$\bar{\mu}_\xi\{z : z \in R_\xi, \ H(\mu_z) \leqq c\} = 1\,, \quad \xi \in Z_X \tag{15.3}$$

take place. For the sake of simplicity in notations we have used R_ξ as an abbreviation for $R_{\operatorname{card}(\xi)}$, the set of all regular points in the space $\{1, \ldots, \operatorname{card}(\xi)\}^I$ (cf. Section 1). Using Lemma 4.4 we have

$$\bar{\mu}_\xi\{z : z \in R_\xi, \ H(\mu_z) \leqq c\} = \int_{R_\xi} \mu_x\{z : H(\mu_z) \leqq c\}\, \bar{\mu}_\xi(\mathrm{d}x) = \tag{15.4}$$

$$= \int_{R_\xi} \left[\int \chi_{\{z : H(\mu_z) \leqq c\}}(y)\, \mu_x(\mathrm{d}y) \right] \bar{\mu}_\xi(\mathrm{d}x)\,.$$

Using the idea developed in detail in [37] we conclude that the right-hand side of (15.4) equals the integral

$$\int_{E(\mathscr{A}_X)} \left[\int \chi_{\{z:H(\mu_z)\leqq c\}}(y)\, \bar{\nu}_\zeta(\mathrm{d}y) \right] \hat{\mu}(\mathrm{d}\nu) . \tag{15.5}$$

The assumption (15.1) implies

$$\hat{\mu}\{\nu : \nu \in E(\mathscr{A}_X),\ H(\bar{\nu}_\zeta) \leqq c\} = 1 , \quad \zeta \in Z_X . \tag{15.6}$$

Let $A = \{\nu : \nu \in E(\mathscr{A}_X),\ H(\bar{\nu}_\zeta) \leqq c\}$. The relation (15.6) implies that the integration domain in the outer integral in (15.5) can be replaced by the set A, thus the integral (15.5) equals the following one

$$\int_A \left[\int \chi_{\{z:H(\mu_z)\leqq c\}}(y)\, \bar{\nu}_\zeta(\mathrm{d}y) \right] \hat{\mu}(\mathrm{d}\nu) . \tag{15.7}$$

But for each $\nu \in A$, $\bar{\nu}_\zeta\{z : H(\mu_z) \leqq c\} = 1$ (cf. [37]). Combining (15.4)–(15.7) together with the latter equality we obtain (15.3). Henceforth, for any $\zeta \in Z_X$, Lemma I of [41] applies to the source $\bar{\mu}_\zeta$:

$$\limsup_n \frac{1}{n} \log L_n(\varepsilon, \bar{\mu}_\zeta) \leqq c \quad \text{for} \quad 0 < \varepsilon < 1 , \quad \zeta \in Z_X .$$

Thus

$$\lim_{\varepsilon \to 0} \limsup_n \frac{1}{n} \log L_n(\varepsilon, \bar{\nu}_\zeta) \leqq c \quad \text{for} \quad \xi \in Z_X ,$$

i.e.

$$\overline{V}(\bar{\mu}_\zeta) \leqq c , \quad \zeta \in Z_X .$$

The theorem follows using these inequalities both with (14.5).

Lemma 15.2. If $\mu \in M(\mathscr{A}_X)$ and if c is a finite real number, then the assumption that

$$\hat{\mu}\{\nu : \nu \in E(\mathscr{A}_X),\ H(\nu) \geqq c\} = 1 \tag{15.8}$$

implies the inequality

$$\underline{V}(\mu) \geqq c . \tag{15.9}$$

Proof. First we shall use the fact that

$$\sup_{\zeta \in Z_X} H(\mu, \zeta) = \lim_{Z_X} H(\mu, \zeta) \quad \text{(cf. [37])} .$$

Using the monotonicity of the net $\{H(\mu, \zeta)\}_{\zeta \in Z_X}$ we obtain the following statement:

$$\forall \delta > 0 \; \exists \zeta_0 \in Z_X \; \forall \zeta \succ \zeta_0 \,, \quad \zeta \in Z_X \,,$$
$$c - \delta < H(\bar{\nu}_\zeta) < H(\nu) \,.$$

It was proved in [37] that for all $\zeta \in Z_X$ we have

$$\bar{\nu}_\zeta\{z : z \in R_\zeta \,, \quad H(\mu_z) = H(\bar{\nu}_\zeta)\} = 1 \,.$$

Therefore for $\zeta \succ \zeta_0$ we have

$$\bar{\nu}_\zeta\{z : z \in R_\zeta, \; H(\mu_z) > c - \delta\} = 1 \,.$$

Now

$$\bar{\mu}_\zeta\{z : z \in R_\zeta, \; H(\mu_z) > c - \delta\} = \int_{E(\mathscr{A}_X)} \bar{\nu}_\zeta\{z : z \in R_\zeta, \; H(\mu_z) > c - \delta\} \, \hat{\mu}(\mathrm{d}\nu) =$$
$$= \int_A \nu_\zeta\{z : z \in R_\zeta, \; H(\mu_z) > c - \delta\} \, \hat{\mu}(\mathrm{d}\nu) \,,$$

where $A = \{\nu : \nu \in E(\mathscr{A}_X), H(\nu) \geqq c\}$. For each $\nu \in A$ we have $H(\bar{\nu}_\zeta) > c - \delta$ for any $\delta > 0$ and $\zeta \succ \zeta_0(\delta)$, thus

$$\forall \delta > 0 \; \exists \zeta_0 \in Z_X \; \forall \zeta \succ \zeta_0, \; \zeta \in Z_X$$
$$\bar{\mu}_\zeta\{z : z \in R_\zeta, \; H(\mu_z) > c - \delta\} = 1 \,.$$

Applying Lemma II of [41] with $\mu = \bar{\mu}_\zeta$, $c = c - \delta$ we conclude that

$$\underline{V}(\bar{\mu}_\zeta) > c - \delta \quad \text{for} \quad \zeta \succ \zeta_0(\delta) \,.$$

Let us assume, contrary to the conclusion of the lemma, that

$$\sup_{\zeta \in Z_X} \underline{V}(\bar{\mu}_\zeta) < c \,.$$

Then there is a finite real number K such that

$$\sup \underline{V}(\bar{\mu}_\zeta) < K < c \,.$$

Take $\delta = c - K < 0$. Then there is $\zeta_0(\delta) \in Z_X$ such that for all $\zeta \in Z_X$, $\zeta \succ \zeta_0(\delta)$,

$$\underline{V}(\bar{\mu}_\zeta) > c - \delta = K \,.$$

Hence we obtain

$$K < \underline{V}(\bar{\mu}_\zeta) \leqq \sup_{\zeta \in Z_X} \underline{V}(\bar{\mu}_\zeta) < K \,,$$

a contradiction. The proof of the lemma is complete.

The main theorem concerning the asymptotic rate deals with its connection with the entropy rates of the ergodic components of a stationary source given. For this purpose, let us recall the definition of the essential supremum. Let $(\Omega, \mathscr{F}, \mu)$ be a probability space. Given any measurable function f on Ω, we shall define its essential supremum as the number

$$\operatorname*{ess\,.\,sup}_{\omega\in\Omega[\mu]} f(\omega) = \inf\{t : \mu\{\omega : f(\omega) \leqq t\} = 1\}\,.$$

Theorem 15.3. The asymptotic rate of a source $\mu \in M(\mathscr{A}_X)$ equals the essential supremum of the entropy rates of its ergodic components; in symbols

$$V(\mu) = \operatorname*{ess\,sup}_{\nu\in E(\mathscr{A}_X)[\hat{\mu}]} H(\nu)\,.$$

16. The Proof: via Ergodic Theoy

The following proof of the Theorem 15.3 will use exclusively the tools within the ergodic theory of invariant set functions, as described in Section 4 and Part II. In the Appendix another proof will be given, which will provide an intuitive meaning for the asymptotic rate. This proof will use the methods of the information transmission theory.

1. Let us set $h = \text{ess . sup } H(\nu)$. Since $H(\nu)$ is a measurable function of the variable ν on $E(\mathscr{A}_X)$, the notion of the essential supremum makes sense. Let us assume $V(\mu) > h$. Then there is a finite real number c such that $V(\mu) > c > h$. This means that

$$\inf\{t : \hat{\mu}\{\nu : H(\nu) \leqq t\} = 1\} < c\,. \tag{16.1}$$

Indeed, if $t \leqq c$ then $\{\nu : H(\nu) \leqq t\} \subset \{\nu : H(\nu) \leqq c\}$, therefore

$$\hat{\mu}\{\nu : H(\nu) \leqq c\} \geqq \hat{\mu}\{\nu : H(\nu) \leqq t\}\,.$$

Taking $t < c$ such that $\hat{\mu}\{\nu : H(\nu) \leqq t\} = 1$ we conclude that

$$\hat{\mu}\{\nu : \nu \in E(\mathscr{A}_X),\ H(\nu) \leqq c\} = 1\,. \tag{16.2}$$

(Clearly, by (16.1), there is at least one t with the required properties.) Now Lemma 15.1 applies because of (16.2), consequently $\overline{V}(\mu) \leqq c$. Since $\overline{V}(\mu) = \underline{V}(\mu) = V(\mu)$, we have

$$\overline{V}(\mu) = V(\mu) \leqq c < V(\mu)$$

a contradiction. Hence the converse inequality $V(\mu) \leqq h$ must always take place.

2. Let us assume that the strict inequality $V(\mu) < h$ is valid. Then we can choose a finite real number c such that the inequalities $V(\mu) < c < h$ are valid. This means that

$$\inf\{t : \hat{\mu}\{v : H(v) \leqq t\} = 1\} > c\,. \tag{16.3}$$

Let us denote by $E(c)$ the set

$$\{v : v \in E(\mathscr{A}_X),\ H(v) \geqq c\}\,.$$

If $\hat{\mu}(E(c)) = 0$ then $\hat{\mu}\{v : H(v) < c\} = 1$. But the latter fact contradicts (16.3), since every t satisfying the relation $\hat{\mu}\{v : H(v) \leqq t\} = 1$ has to satisfy the inequality $t > c$. Consequently, $\hat{\mu}(E(c)) = \alpha > 0$.

Let us consider the case $\alpha = 1$. Then

$$\hat{\mu}\{v : H(v) \geqq c\} = 1\,,$$

hence by Lemma 15.2 we obtain the contradictory inequalities

$$\underline{V}(\mu) = V(\mu) \geqq c > V(\mu)\,.$$

The case $0 < \alpha < 1$ will be reduced to the former one. Let $0 < \alpha < 1$. Define the probability measures $\hat{\mu}'$, $\hat{\mu}''$ on $E(\mathscr{A}_X)$ by the properties that

$$\hat{\mu}'(E) = \frac{1}{\alpha}\,\hat{\mu}(E \cap E(c))\,,$$

$$\hat{\mu}''(E) = \frac{1}{1-\alpha}\,[\hat{\mu}(E) - \alpha\hat{\mu}'(E)],\ E \in \mathscr{K}[E(\mathscr{A}_X)]\,.$$

Then

$$\hat{\mu} = \alpha\hat{\mu}' + (1-\alpha)\,\hat{\mu}''\,.$$

Now by Theorem 7.3 there are stationary sources μ', μ'' such that $\mu = \alpha\mu' + (1-\alpha)\,\mu''$. Indeed, let

$$\mu'(A) = \int_{E(\mathscr{A}_X)} \nu(A)\,\hat{\mu}'(\mathrm{d}\nu)\,,\ A \in \mathscr{A}_X\,;$$

the source μ'' being defined analogously by means of the measure $\hat{\mu}''$. Using twice the Extension Theorem for measures we conclude that

$$\bar{\mu}_\zeta = \alpha\bar{\mu}'_\zeta + (1-\alpha)\,\bar{\mu}''_\zeta\,,\quad \zeta \in Z_X$$

Now $\hat{\mu}'\{v : H(v) \geqq c\} = \alpha^{-1}\,\hat{\mu}(E(c)) = 1$. Hence $\underline{V}(\mu') \geqq c$ by Lemma 15.2. Now

for every $\zeta \in Z_X$ we have the inequality

$$\liminf_n \frac{1}{n} \log L_n(\varepsilon, \bar{\mu}_\zeta) \geqq \liminf_n \frac{1}{n} \log L_n(\varepsilon/\alpha, \bar{\mu}'_\zeta)$$

valid for all $0 < \varepsilon < \alpha$ (cf. [41], p. 144). Hence

$$\underline{V}(\bar{\mu}_\zeta) \geqq \underline{V}(\bar{\mu}'_\zeta), \quad \zeta \in Z_X .$$

This finally gives the desired contradiction:

$$V(\mu) = \underline{V}(\mu) \geqq c > V(\mu) .$$

The theorem is proved.

17. The Basic Relations between the Rates

In order not to confuse the notations, we shall use the symbols $\mathscr{H}(\mu)$ and $\mathscr{V}(\mu)$ for the entropy rate and the asymptotic rate, respectively, as they were defined in [41].

Theorem 17.1. Let $\mu \in M(\mathscr{A}_X)$. Then $V(\mu) \geqq H(\mu)$. If, moreover, μ is an ergodic source, then $V(\mu) = H(\mu)$.

Proof. Let $\mu \in M(\mathscr{A}_X)$. Then the inequality stated in the theorem is a corollary both to Theorems 12.1 and 15.3. Indeed,

$$H(\mu) = \int_{E(\mathscr{A}_X)} H(\nu)\, \hat{\mu}(\mathrm{d}\nu) \leqq \operatorname{ess.sup}_{\nu \in E(\mathscr{A}_X)[\hat{\mu}]} H(\nu) = V(\mu) .$$

Let $\zeta \in Z_X$. Then $V(\bar{\mu}_\zeta) = \mathscr{V}(\bar{\mu}_\zeta)$, by the very definition of $V(\bar{\mu}_\zeta)$. If $\mu \in E(\mathscr{A}_X)$, then $\bar{\mu}_\zeta$ is an ergodic finite-alphabet source (cf. Lemma 10.2). Consequently, $\mathscr{V}(\bar{\mu}_\zeta) = \mathscr{H}(\bar{\mu}_\zeta)$ (cf. [40], Theorem 9.1). But for finite alphabet sources the concepts of the entropy rate coincide, hence $\mathscr{H}(\bar{\mu}_\zeta) = H(\bar{\mu}_\zeta)$. The theorem is proved.

Remark 17.2. Theorem 9.1 used in the proof of the preceeding theorem actually states more than was really used:

$$\lim_n \frac{1}{n} \log L_n(\varepsilon, \bar{\mu}_\zeta) = \mathscr{V}(\bar{\mu}_\zeta) = (\mathscr{H}\bar{\mu}_\zeta), \quad \zeta \in Z_X .$$

Hence we have

$$\sup_{\zeta \in Z_X} \lim_n \frac{1}{n} \log L_n(\varepsilon, \bar{\mu}_\zeta) = V(\mu) = H(\mu), \tag{17.1}$$

the relation being valid for any ergodic source μ. The relation (17.1) motivates the notion of the strong stability, introduced and examined in the next section.

Example 17.3. Let us consider the decomposable stochastic matrix

$$\mathbf{A} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/100 & 99/100 \\ 0 & 0 & 1/100 & 99/100 \end{pmatrix}$$

Denote by $\{Y_n\}$ the corresponding Markov chain. The indecomposable submatrices of the matrix $\mathbf{A}$ are denoted by the symbols $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$, respectively, i.e.

$$\mathbf{A}^{(1)} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}, \quad \mathbf{A}^{(2)} = \begin{pmatrix} 1/100 & 99/100 \\ 1/100 & 99/100 \end{pmatrix}.$$

The Markov chains $\{X_n^{(1)}\}$ and $\{X_n^{(2)}\}$ determined by the matrices $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ are ergodic. Denote the absolute stationary distribution by $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(2)}$, respectively. Then (cf. [2])

$$H(\{X_n^{(1)}\}) = p_1^{(1)} H(1/2, 1/2) + p_2^{(1)} H(1/2, 1/2) = 1,$$

$$H(\{X_n^{(2)}\}) = p_1^{(2)} H(1/100, 99/100) + p_2^{(2)} H(1/100, 99/100) =$$
$$= H(1/100, 99/100) \sim 0{\cdot}06.$$

Since $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(2)}$ are the absolute stationary distributions, any probability 4-vector

$$\mathbf{p} = (\alpha p_1^{(1)}, \alpha p_2^{(1)}, (1-\alpha) p_1^{(2)}, (1-\alpha) p_2^{(2)})$$

with $0 < \alpha < 1$ is the absolute stationary distribution of the Markov chain $\{Y_n\}$ corresponding to the original transition probability matrix $\mathbf{A}$. Now

$$H(\{Y_n\}) = \alpha H(\{X_n^{(1)}\}) + (1-\alpha) H(\{X_n^{(2)}\}).$$

Hence, for $0 < \alpha < 1$, $H(\{Y_n\})$ ranges within the interval

$$0{\cdot}06 < H(\{Y_n\}) < 1$$

(cf. (12.3)). On the other hand (cf. Theorem 15.3), we have

$$V(\{Y_n\}) = \max [H(\{X_n^{(1)}\}), \quad H(\{X_n^{(2)}\})] = 1$$

regardless of what value of α was chosen.

18. Strongly Stable Sources

The proof of the simplest form of the coding theorem (cf. Theorem 2.1) is based on the following statement:

Lemma 18.1. Let π_n be the (apriori) probability of the observed sequence $x^{(1)}, \ldots, x^{(n)}$ of independent trials. Then

$$\forall \eta > 0 \, \forall \delta > 0 \, \exists n_0 \, \forall n \geqq n_0$$

$$P\left\{\left| -\frac{1}{n} \log \pi_n - H\right| \leqq \eta\right\} > 1 - \delta . \tag{18.1}$$

A generalization of this theorem for ergodic sources is just the well-known McMillan's theorem [23]. These statements show that in a sequence of independent symbols (or in an ergodic sequence) the quantity of information per symbol is asymptotically stable. This means that the average quantity of information is, with probability as close to unity as wanted, nearly a constant, if n is large enough.

The following concept of stability makes sense for arbitrary, even nonstationary sources. Let μ be any finitely additive probability on the field $\mathscr{A}_X$. For $\zeta \in Z_X$, we shall set

$$L_{i,n}(\varepsilon, \mu, \zeta) = \min \{\operatorname{card}(\xi) : \xi \subset \zeta^n , \quad \sum_{D \in \xi} \mu[D]_{i,n} > 1 - \varepsilon\} \tag{18.2}$$

for $i \in I$, $n \in N$, $0 < \varepsilon < 1$, respectively. The source μ is said to be *strongly stable* provided there is a nonnegative (possibly infinite) real number H such that for all $i \in I$, $0 < \varepsilon < 1$,

$$\sup_{\zeta \in Z_X} \lim_{n \to \infty} \frac{1}{n} \log L_{i,n}(\varepsilon, \mu, \zeta) = H . \tag{18.3}$$

The relation (17.1) together with Theorem 17.1 imply that an ergodic source is strongly stable and the corresponding number H equals its entropy rate. For stationary sources, the following statement is valid:

Theorem 18.2. Let $\mu \in M(\mathscr{A}_X)$. Then the source μ is strongly stable if and only if $H(\mu) = V(\mu)$.

Proof. 1. Let μ be a strongly stable source. By Theorem 17.1, $V(\mu) \geqq H(\mu)$ for any stationary source μ. Let the strict inequality $V(\mu) > H(\mu)$ takes place with positive probability, i.e. let

$$\hat{\mu}\{\nu : \nu \in E(\mathscr{A}_X), \quad H(\nu) < V(\nu)\} > 0 .$$

Then there is a finite real number c such that $c < V(\mu)$ and

$$\hat{\mu}\{\nu : \nu \in E(\mathscr{A}_X), \quad H(\nu) < c\} = 1 - \alpha > 0 .$$

Moreover, $\alpha > 0$. Indeed, if $\alpha = 0$, then

$$\hat{\mu}\{\nu : \nu \in E(\mathscr{A}_X), \quad H(\nu) < c\} = 1 .$$

Hence we should have the contradictory inequality $V(\mu) \leqq c$ by Lemma 15.1. Repeating for $0 < \alpha < 1$ the argument used in the proof of Theorem 15.3 we find a source $\mu_1 \in M(\mathscr{A}_X)$ such that

$$\lim_n \frac{1}{n} \log L_n(\varepsilon, \mu, \zeta) \leqq V(\mu_1) \leqq c < V(\mu), \quad \zeta \in Z_X .$$

The latter inequalities are valid independently of what ε was chosen, hence $V(\mu) \leqq$ $\leqq V(\mu_1) \leqq c < V(\mu)$, a contradiction. Thus we have obtained the relation

$$\hat{\mu}\{\nu : \nu \in E(\mathscr{A}_X), \quad H(\nu) = V(\mu)\} = 1 . \tag{18.4}$$

On the other hand,

$$H(\mu) = \int_{E(\mathscr{A}_X)} H(\nu)\, \hat{\mu}(\mathrm{d}\nu) . \tag{18.5}$$

From (18.4) and (18.5) we conclude that

$$H(\mu) = \int_{E(\mathscr{A}_X)} H(\nu)\, \hat{\mu}(\mathrm{d}\nu) = \int_{E(\mathscr{A}_X)} V(\mu)\, \hat{\mu}(\mathrm{d}\nu) = V(\mu)\, \hat{\mu}(E(\mathscr{A}_X)) = V(\mu) .$$

2. Conversely, let $V(\mu) = H(\mu)$. Since $V(\nu) = H(\nu)$ for all ergodic sources ν, we obtain the following equality:

$$\operatorname{ess.sup}_{\nu \in E(\mathscr{A}_X)[\hat{\mu}]} V(\nu) = \int_{E(\mathscr{A}_X)} V(\nu)\, \hat{\mu}(\mathrm{d}\nu) .$$

This in turn implies

$$\hat{\mu}\{\nu : \nu \in E(\mathscr{A}_X), \ V(\nu) = V(\mu)\} = 1 . \tag{18.6}$$

Actually, if there was an $0 < \alpha < 1$ with the property $\hat{\mu}\{\nu : \nu \in E(\mathscr{A}_X), V(\nu) =$ $= V(\mu)\} = \alpha$, then we should have

$$H(\mu) = (1 - \alpha)\, H(\mu') + \alpha\, H(\mu'') < V(\mu)$$

(for the symbols μ' and μ'' cf. the proof of Theorem 15.3). Using Lemma 15.2 and (18.6) we obtain

$$\liminf_n \frac{1}{n} \log L_n(\varepsilon, \mu, \zeta) \geqq V(\mu), \quad \zeta \in Z_X .$$

On the other hand

$$V(\mu) \geqq \limsup_n \frac{1}{n} \log L_n(\varepsilon, \mu, \zeta)$$

for all $\zeta \in Z_X$ and $0 < \varepsilon < 1$, respectively. Hence the strong stability of the source μ follows with $H = H(\mu) = V(\mu)$.

There arises a natural question whether there are nontrivial stationary nonergodic sources possessing the property of strong stability. The affirmative answer is given by the following example:

Example 18.3. Let us consider the stochastic matrix

$$\mathbf{A} = \begin{pmatrix} 2/3 & 1/3 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 \\ 0 & 0 & 3/4 & 1/4 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}.$$

The indecomposable submatrices

$$\mathbf{A}^{(1)} = \begin{bmatrix} 2/3 & 1/3 \\ 1/4 & 3/4 \end{bmatrix}, \qquad \mathbf{A}^{(2)} = \begin{bmatrix} 3/4 & 1/4 \\ 1/3 & 2/3 \end{bmatrix}$$

determine the ergodic Markov chains. The absolute stationary distribution $\mathbf{p}^{(i)} (i = 1, 2)$ of the ergodic matrices $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ are given by

$$\mathbf{p}^{(1)} = (3/7, 4/7), \quad \mathbf{p}^{(2)} = (4/7, 3/7).$$

Note that for any α, $0 \leqq \alpha \leqq 1$, the probability 4-vector

$$\mathbf{p} = (3\alpha/7, 4\alpha/7, 4(1 - \alpha)/7, 3(1 - \alpha)/7)$$

is the absolute stationary distribution corresponding to the matrix $\mathbf{A}$. The entropy rates of the ergodic subchains are

$$H^{(1)} = 3/7 \, . \, H(2/3, 1/3) + 4/7 \, . \, H(1/4, 3/4),$$
$$H^{(2)} = 4/7 \, . \, H(1/4, 3/4) + 3/7 \, . \, H(2/3, 1/3) = H^{(1)} = H \, .$$

Hence the entropy rate of the Markov source corresponding to the original matrix $\mathbf{A}$, is given by the relation

$$H(\mathbf{A}) = \alpha H^{(1)} + (1 - \alpha) H^{(2)} = H \, .$$

On the other hand

$$V(\mathbf{A}) = \max (H^{(1)}, H^{(2)}) = H \, .$$

Consequently, $H(\mathbf{A}) = V(\mathbf{A}) = H$, i.e. the Markov source determined by the matrix $\mathbf{A}$ is strongly stable. On the other hand, the matrix $\mathbf{A}$ is decomposable, i.e. the state space consists of two essential sets of states. Thus the Markov chain cannot be ergodic.

19. On σ-Additive Sources with a Countable Alphabet.

In this section we shall return to the original setting of [40] and [41]. Thus we are given the alphabet $X = N$. The original method of obtaining the examined quantities was performed in two steps:

Step 1: The proof of a general form of McMillan's theorem for countably infinite alphabets (cf. Theorem 4.6)

Step 2: Using this form of McMillan's theorem the basic lemmas are proved yielding the necessary tools for the proof of the coding theorem.

Our method differs from the original one. It can be described also in two steps:

Step 1: Using finite partitions the problem of the convenient form of McMillan's theorem is reduced to the finite alphabet sources.

Step 2: Consists merely of the single definition by means of a supremalization process.

The second method seems to be far simpler. However, the proof of the statement that both methods provide the same quantities, needs a nontrivial statement we shall start with.

Let us consider the entropy rate $\mathscr{H}(\mu)$ as was defined by (4.1). A necessary and sufficient condition for the finiteness of $\mathscr{H}(\mu)$ is the finiteness of the alphabet entropy, i.e. the condition

$$-\sum_{k=1}^{\infty} \mu[k]_{0,1} \log \mu[k]_{0,1} < \infty \,. \tag{19.1}$$

As well-known, in this case $\mathscr{H}(\mu) = H(\mu)$ (cf. [32]). However, three are also well-known examples in which $\mathscr{H}(\mu) = \infty$ and $H(\mu)$ is finite [32]. The sources satisfying the condition $H(\mu) = \mathscr{H}(\mu)$ possess the following extended approximation property:

Theorem 19.1. Let μ be a σ-additive source satisfying the condition $\mathscr{H}(\mu) = H(\mu)$ Let the sequence $\{\tau_k\}_{k=1}^{\infty}$ of mappings be defined by the relation (4.10). Then the sequence $\mathscr{V}(\mu\tau_k^{-1})$ monotonically increases to the asymptotic rate $\mathscr{V}(\mu)$ of the source μ.

Proof. The monotonicity property of the sequence $\mathscr{V}(\mu\tau_k^{-1})$ can be easily verified. Hence, it suffices to prove that

$$\mathscr{V}(\mu) = \sup_k \mathscr{V}(\mu\tau_k^{-1}) \,. \tag{19.2}$$

Since μ is σ-additive, we have

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

(cf. Lemma 4.4). Hence ([41], Theorem II)

$$\mathscr{V}(\mu) = \operatorname*{ess.sup}_{z \in R[\mu]} \mathscr{H}(\mu_z) .$$

Now

$$\mathscr{H}(\mu_z) = \lim_k \mathscr{H}(\mu_z \tau_k^{-1})$$

(cf. Theorem 4.7). From the ergodic decomposition of the source μ we obtain especially

$$\mu\tau_k^{-1} = \int_R \mu_z \tau_k^{-1} \, \mu(\mathrm{d}z), \quad k = 1, 2, \ldots$$

thus

$$\mathscr{V}(\mu\tau_k^{-1}) = \operatorname*{ess.sup}_{z \in R[\mu]} \mathscr{H}(\mu_z \tau_k^{-1}) .$$

Consequently

$$\mathscr{V}(\mu) = \operatorname*{ess.sup}_{z \in R[\mu]} \mathscr{H}(\mu_z) = \operatorname*{ess.sup}_{z \in R[\mu]} \left[\sup_k \mathscr{H}(\mu_z \tau_k^{-1})\right] \geqq \operatorname*{ess.sup}_{z \in R[\mu]} \mathscr{H}(\mu_z \tau_k^{-1}) ,$$

i.e.

$$\mathscr{V}(\mu) \geqq \sup_k \left[\operatorname*{ess.sup}_{z \in R[\mu]} \mathscr{H}(\mu_z \tau_k^{-1})\right] = \sup_k \mathscr{V}(\mu\tau_k^{-1}) .$$

Let us assume contrary to (19.2) that the inequality

$$\mathscr{V}(\mu) > \sup_k \mathscr{V}(\mu\tau_k^{-1}) = V$$

takes place. From the definition of the essential supremum we conclude the existence of a positive δ such that

$$\mu\{z : \mathscr{H}(\mu_z) \leqq V + \delta\} < 1 ,$$

i.e.

$$\mu\Big(\bigcap_{k=1}^{\infty} \{z : \mathscr{H}(\mu_z \tau_k^{-1}) \leqq V + \delta\}\Big) < 1 .$$

Hence there is at least one k_0 such that

$$\mu\{z : \mathscr{H}(\mu_z \tau_{k_0}^{-1}) \leqq V + \delta\} < 1 . \tag{19.3}$$

On the other hand,

$$\operatorname*{ess.sup}_{z\in R[\mu]} \mathscr{H}(\mu_z\tau_k^{-1}) = \inf\{t_k : \mu\{z : \mathscr{H}(\mu_z\tau_k^{-1}) \leqq t_k\} = 1\}\,.$$

This in turn implies that

$$\mathscr{V}(\mu\tau_k^{-1}) \leqq \sup_k \mathscr{V}(\mu\tau_k^{-1}) = V < V + \delta (k = 1, 2, \ldots)\,.$$

Hence given k there is t_k (with $t_k < V + \delta$) such that

$$\mu\{z : \mathscr{H}(\mu_z\tau_k^{-1}) \leqq t_k\} = 1\,.$$

Especially, for $k = k_0$, there is $t_{k_0} < V + \delta$ such that

$$\mu\{z : \mathscr{H}(\mu_z\tau_{k_0}^{-1}) \leqq t_{k_0}\} = 1\,. \tag{19.4}$$

But the inequality $t_{k_0} < V + \delta$ yields a contradiction, by (19.3) and (19.4). The theorem is proved.

As an immediate corollary we obtain an affirmative answer to the question posed at the beginning of this section:

Theorem 19.2. Let μ be a stationary σ-additive source with at most countable alphabet. If either

(1) there is a natural number k such that

$$\mu(\{1, \ldots, k\}^I) = 1\,,$$

or, (2) there is no such k, but the source satisfies the condition $H(\mu) = \mathscr{H}(\mu)$ (especially the condition (19.1)),

then $V(\mu) = \mathscr{V}(\mu)$.

Remark 19.3. The absence of the pointwise partitions of the alphabet, which can serve as a generator causes that for the general alphabet X the approximation theorems 4.7 and 19.1 fail to hold. It would be interesting to find some simple partitions generating the σ-field F_X, if there are any.

PART IV. RATES ASSOCIATED WITH PAIRS OF SOURCES

20. Statement of the Coding Problems

Let us start with the general coding problem as introduced in Section 3. We shall consider the following special case. Let $X = Y$ be a given finite set. Let μ be a stationary source with the alphabet X. Then define

$$K_n^{(X)}(E_n) = \mu[E_n]\,, \quad E_n \subset X^n$$

(cf. (1.3)). The parameters ε will be independent on n, and $\varepsilon \in (0, 1)$. $\Psi_\varepsilon^{(n)}$ will be the identity mappings on X^n onto itself for any ε, $n = 1, 2, \ldots$. Finally, $(K_n^{(X)}(E_n), \varepsilon) \in \varphi$ if and only if $K_n^{(X)}(E_n) > 1 - \varepsilon$. This means that we shall deal with the n-dimensional ε-codes for fixed ε, $\varepsilon \in (0, 1)$. The criterion $K_n^{(Y)}(E_n)$ will be derived from another stationary source ν with the same alphabet X (the properties of the source ν will be specified later):

$$K_n^{(\nu)}(E_n) = \sum_{\bar{x} \in E_n} \frac{\mu[\bar{x}]}{\nu[\bar{x}]}, \quad E_n \subset X^n (= Y^n).$$

Our aim will be to derive the coding theorem for the quantity

$$S_n^I = \min \left\{ \sum_{\bar{x} \in E_n} \frac{\mu[\bar{x}]}{\nu[\bar{x}]} : E_n \subset X^n, \quad \mu[E_n] > 1 - \varepsilon \right\}. \tag{20.1}$$

A coding theorem together with its weak converse will be proved, i.e. the limit

$$\lim_n \frac{1}{n} \log S_n^I$$

will depend, in general, on ε. Under the additional constraint that μ is ergodic, we shall obtain also a strong converse, i.e. the above limit will not depend on ε.

In general, the criterion K_n^Y is not necessarily arising from a stochastic process. This means that instead of a consistent family $\{\nu_n; n \in N\}$ of finite-dimensional distribution of a process we can consider a family $\sigma = (\sigma_n)_{n \in N}$ of finite measures (each σ_n defined on the σ-field $\mathfrak{B}(X^n)$, $n = 1, 2, \ldots$). We shall consider the following special family $\sigma = (\sigma_n)$ with

$$\sigma_n\{\bar{x}\} = \frac{\mu[\bar{x}]}{\nu[\bar{x}]}, \quad \bar{x} \in X^n.$$

The resulting quantity will be

$$S_n^I = \min \{\nu_n(E_n) : E_n \subset X^n, \ \mu_n(E_n) > 1 - \varepsilon\}. \tag{20.2}$$

It is intuitively clear that the quantity just obtained can serve as a measure of discrimination between the processes μ and ν, respectively. The problem will be studied more in detail in the last part of the present paper.

To motivate the choice of the two above given coding problems, let us recall some facts from [18]. Let X be a finite set, let $\mathbf{p}, \mathbf{q}$ be two probability vectors on X (to avoid complications it is assumed that $\mathbf{q}$ is strictly positive on X). In [18] there was introduced the notion of inaccuracy following the formula

$$H(\mathbf{p}, \mathbf{q}) = -\sum_{x \in X} p(x) \log q(x). \tag{20.3}$$

Note that

$$H(\mathbf{p}, \mathbf{q}) = H(\mathbf{p}) + I(\mathbf{p}, \mathbf{q}) \tag{20.4}$$

where $H(\mathbf{p})$ denotes the usual entropy of the probability vector $\mathbf{p}$ and $I(\mathbf{p}, \mathbf{q})$ is the well-known I-divergence [21]:

$$I(\mathbf{p}, \mathbf{q}) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}. \tag{20.5}$$

These facts remain valid without changes for the discrete memoryless sources. The coding theorems for the quantities (20.1) and (20.2) will provide a generalization of the notions of inaccuracy and I-divergence to stochastic processes.

21. The K-Entropy and the K-Rate

We shall impose the following conditions upon the possible pairs (μ, ν) of sources:

(I) μ is a stationary source, i.e. $\mu \in M(\mathscr{F}_X)$;

(II) ν is a stationary k-Markov source, positive on all elementary cylinders in the space X^I, in symbols $\nu \in M(\mathscr{F}^X; k)$;

X being a common finite alphabet. The quantities S_n^I defined by (20.1) and (20.2) were studied by Pötschke [29] in case μ is ergodic. Note that the positivity of ν can be replaced by the condition that μ_n's are absolutely continuous with respect to the corresponding ν_n's. However, it is only an unessential difference. Therefore we shall prefer the presented form of the condition to avoid more complicated a.e. considerations.

The markovian property is fairly more stringent. It is imposed due to the fact that a corresponding form of McMillan's theorem is needed for the proof of the coding theorems. A general form of McMillan's theorem under the markovian assumption was obtained by Shu-Teh C. Moy (Generalizations of Shannon - McMillan Theorem. Pacific J. Math. *11* (1961), 705–714). Theorem 21.2 below is a special case of this general result. Recently Perez introduced a condition (actually a very strong condition) assuring the validity of McMillan's theorem for the generalized entropy without the markovian assumption (cf. e.g. Generalization of Chernoff's result on the asymptotic discernibility of two random processes. In: Progress in Statistics (J. Gani, K. Sarkadi, and I. Vincze, Eds.) Vol. II., North-Holland, Amsterdam –London 1972, 619–632, and Asymptotic Discernability of Random Processes. In: Proceedings of the Prague Symposium on Asymptotic Statistics (J. Hájek, ed.), Prague 1973, Vol. II, 311–322). The related problems will be studied by the author in a separate paper.

We define the *K-entropy* (called the B-entropy in [29] and [38]) of the pair (μ, ν) by the formula

$$K(\mu, \nu) = -\lim_n \frac{1}{n} \int \log \nu[z_1, \ldots, z_n] \, \mu(\mathrm{d}z) \tag{21.1}$$

(cf. (1.5) and (1.6)). Because of the invariance of the measures μ and ν the limit in (21.1) always exists and condition (II) implies it is always finite. If we had used the formulation of (II) by means of the absolute continuity, the finiteness of the limit in (21.1) could be assured by means of some regularity conditions similar to that given by Bahadur (cf. [3] and the papers cited therein). The main properties of the K-entropy were studied in the author's paper [38]. Here we shall describe only the main ideas, in many aspects similar to that of [27] and [41].

Theorem 21.1. [29]. If the pair (μ, ν) of sources satisfies the conditions (I) and (II) and if, moreover, μ is an ergodic source, then

$$\mu\left\{z : z \in X^I, -\lim_n \frac{1}{n} \log \nu[z_1, \ldots, z_n] = K(\mu, \nu)\right\} = 1 . \tag{21.2}$$

Theorem 21.1 is nothing but a version of McMillan's theorem. The next limit theorem concerns with the *I-entropy* $d(\mu, \nu)$ defined by the relation

$$\mathrm{d}(\mu, \nu) = K(\mu, \nu) - H(\mu) . \tag{21.3}$$

(cf. (4.1)).

Theorem 21.2. [29]. If the conditions of Theorem 21.1 are satisfied then the sequence

$$\frac{1}{n} \log \frac{\mu[z_1, \ldots, z_n]}{\nu[z_1, \ldots, z_n]}$$

converges in probability (with respect to μ) to the I-entropy $\mathrm{d}(\mu, \nu)$.

Theorem 21.2 is weaker that Theorem 21.1, because it states only stochastic convergence. However, only this type of convergence is needed for the proof of the coding theorems (cf. e.g. the proofs of lemmas I and II in [41]).

Since ν is a k-Markov source, we can define an $\mathscr{F}_X$-measurable function g on X^I by the relation

$$g(z) = -\log \nu([z_{k+1}] \mid [z_1, \ldots, z_k]) .$$

The condition (II) implies that the function g is bouded, hence μ-integrable. Consequently, the individual ergodic theorem of Birkhoff applies for μ and g. Accordingly, there is a μ-integrable T_X-invariant function $\hat{g}_\nu$ such that

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} g(T_X^j z) = \hat{g}_\nu(z) \quad \text{a.e.} \quad z[\mu] , \tag{21.4}$$

$$K(\mu, \nu) = \int g \, \mathrm{d}\mu = \int \hat{g}_\nu \, \mathrm{d}\mu .$$

The first equality in (21.4) follows from the definitions of $K(\mu, \nu)$ and g making use of the markovian property of ν (cf. [29]). Actually a stronger result was obtained in [29]:

$$-\lim_n \frac{1}{n} \log \nu[z_1, \ldots, z_n] = \hat{g}_\nu(z) \quad \text{a.e.} \quad z[\mu] . \tag{21.5}$$

In [38] it was proved the following important lemma:

Lemma 21.3. The function $\hat{g}_\nu(z)$ equals a.e. $[\mu]$ the K-entropy $K(\mu_z, \nu)$, where μ_z is the ergodic component of any stationary source μ; in symbols

$$\mu\{z : z \in R_X,\ K(\mu_z, \nu) = \hat{g}_\nu(z)\} = 1 ; \quad \mu \in M(\mathscr{F}_X) .$$

The first immediate corollary to the lemma is the theorem on the integral representation of the K-entropy.

Theorem 21.4 [38]. Let (μ, ν) be a pair of sources satisfying the conditions (I) and (II). Then

$$K(\mu, \nu) = \int_{R_X} K(\mu_z, \nu)\, \mu(\mathrm{d}z) . \tag{21.6}$$

Lemma 21.3 was the key step in deriving the desired generalizations of Theorem 21.1 for stationary non-ergodic sources.

Theorem 21.5 [38]. Let (μ, ν) be a pair of sources satisfying the conditions (I) and (II). Then the sequence $(-1/n) \log \nu[z_1, \ldots, z_n]$ converges a.e. μ to the K-entropy $K(\mu_z, \nu)$; μ_z being the ergodic component of the stationary source μ; in symbols

$$\mu\left\{z : z \in R_X ,\ -\lim_n \frac{1}{n} \log \nu[z_1, \ldots, z_n] = K(\mu_z, \nu)\right\} = 1 .$$

In accordance with (14.2) we shall denote the quantity S_n^I defined in (20.1) by the symbol $L_n(\varepsilon, \mu, \nu)$. Note that if $\mu = \nu$ then

$$L_n(\varepsilon, \mu, \nu) = \min\{\mathrm{card}\,(E_n) : E_n \subset X^n,\ \mu[E_n] > 1 - \varepsilon\} = L_n(\varepsilon, \mu) .$$

The coding theorem together with its strong converse was obtained in [29] under the additional constraint that μ be ergodic:

Theorem 21.6 [29]. Let (μ, ν) be a pair of sources satisfying the conditions (I) and (II). Let μ be an ergodic source. Then

$$\lim_n \frac{1}{n} \log L_n(\varepsilon, \mu, \nu) = K(\mu, \nu); \quad 0 < \varepsilon < 1 . \tag{21.7}$$

Now we shall prove a version of the coding theorem without the ergodicity assumption. The method will be similar to that given in [41] (cf. also [38] and the third part of the present paper).

Lemma 21.7 [38]. Let (μ, ν) be a pair of sources satisfying the conditions (I) and (II). If c is a finite real number, then the assumption that

$$\mu\{z : z \in R_X,\ K(\mu_z, \nu) \leqq c\} = 1$$

implies the inequality

$$\limsup_n \frac{1}{n} \log L_n(\varepsilon, \mu, \nu) \leqq c \quad \text{for} \quad 0 < \varepsilon < 1 .$$

The dual version of Lemma 21.7 is the following

Lemma 21.8 [38]. Under the assumptions of the preceeding lemma, the relation

$$\mu\{z : z \in R_X,\ K(\mu_z, \nu) \geqq c\} = 1$$

implies the inequality

$$\liminf_n \frac{1}{n} \log L_n(\varepsilon, \mu, \nu) \geqq c \quad \text{for} \quad 0 < \varepsilon < 1 .$$

Using these two lemmas together with some elementary properties of $L_n(\varepsilon, \mu, \nu)$ derived in [38], the following theorem was proved:

Theorem 21.9 [38]. Let the pair (μ, ν) of sources satisfy the conditions (I) and (II). Then the inequality

$$\limsup_n \frac{1}{n} \log L_n(\varepsilon_1, \mu, \nu) \leqq \liminf_n \frac{1}{n} \log L_n(\varepsilon_2, \mu, \nu)$$

holds for $0 < \varepsilon_2 < \varepsilon_1 < 1$; consequently, the limit

$$\lim_n \frac{1}{n} \log L_n(\varepsilon, \mu, \nu) = V_\varepsilon(\mu, \nu)$$

exists except at most a countable set of numbers ε. The function $V_\varepsilon(\mu, \nu)$ monotonically increases for $\varepsilon \to 0$ to a limit, which will be denoted by the symbol $V(\mu, \nu)$ and called the *asymptotic K-rate* of the pair (μ, ν) of sources.

The equivalent form of Theorem 21.9 is the following:

Theorem 21.10. There exists one and only one nonnegative real-valued function

V on the set $M(\mathscr{F}_X) \times M(\mathscr{F}_X; k)$ such that

(1) $$\forall \lambda > 0\, \forall 0 < \varepsilon < 1\, \exists n_0\, \forall n \geqq n_0\, \exists E_n \subset X^n$$
$$[\mu_n(E_n) > 1 - \varepsilon] \text{ et } \left[\sum_{\bar{x} \in E_n} \frac{\mu_n\{\bar{x}\}}{\nu_n\{\bar{x}\}} < 2^{n[V(\mu,\nu)+\lambda]}\right];$$

(2) $$\forall \lambda > 0\, \exists 0 < \eta < 1\, \forall 0 < \varepsilon \leqq \eta\, \exists n_0\, \forall n \geqq n_0\, \forall E_n \subset X^n$$
$$\mu_n(E_n) > 1 - \varepsilon \quad \text{implies}$$
$$\sum_{\bar{x} \in E_n} \frac{\mu_n\{\bar{x}\}}{\nu_n\{\bar{x}\}} > 2^{n[V(\mu,\nu)-\lambda]}.$$

This means that given any λ there is an arbitrarily good n-dimensional code (i.e. a code with the probability of the erroneous decoding less than any ε) with the property

$$\frac{1}{n} \log \sum_{\bar{x} \in E_n} \frac{\mu_n\{\bar{x}\}}{\nu_n\{\bar{x}\}} < V(\mu, \nu) + \lambda$$

provided n is sufficiently large, but on the other hand, there are no good n-dimensional codes for which

$$\frac{1}{n} \log \sum_{\bar{x} \in E_n} \frac{\mu_n\{\bar{x}\}}{\nu_n\{\bar{x}\}} < V(\mu, \nu)$$

provided n is sufficiently large. Thus Theorem 21.10 is a coding theorem (statement (1)) together with its weak converse (statement (2)). If μ is ergodic, we obtain also the strong converse. This theorem states actually a little bit more than the original theorem (cf. (21.7)):

Corollary 21.11. If the conditions of Theorem 21.9 are satisfied and if, moreover, μ is an ergodic source, then

$$\lim_n \frac{1}{n} \log L_n(\varepsilon, \mu, \nu) = V(\mu, \nu) = K(\mu, \nu); \quad 0 < \varepsilon < 1.$$

Using Lemmas 21.7 and 21.8 we can obtain similarly as in [41] the following theorem, connecting the quantities $V(\mu, \nu)$ and $K(\mu, \nu)$ in the stationary non-ergodic case:

Theorem 21.12. The asymptotic K-rate $V(\mu, \nu)$ equals the essential supremum of the K-entropies $K(\mu_z, \nu)$; μ_z being the ergodic component of the stationary source μ; in symbols

$$V(\mu, \nu) = \operatorname*{ess\,sup}_{z \in R_X[\mu]} K(\mu_z, \nu).$$

22. The Asymptotic *I*-Rate

Throughout this section we shall assume that the pairs (μ, ν) of sources satisfy the conditions (I) and (II) of Section 21. The aim is to prove a coding theorem for the quantity S_n^I defined by (20.2). Hence the family $\sigma = (\sigma_n)_{n\in N}$ will be defined by the properties

$$\sigma_n\{\bar{x}\} = \mu_n\{\bar{x}\}/\nu_n\{\bar{x}\}\ ; \quad \bar{x} \in X^n\ ;$$

$$\sigma_n(E) = \sum_{\bar{x}\in E} \sigma_n\{\bar{x}\}\ ; \quad E \in \mathfrak{B}(X^n)\ .$$

Let $E_n \subset X^n$ be an n-dimensional ε-code. If for some $\bar{x} \in E_n$ we would have $\mu_n\{\bar{x}\} = 0$, the corresponding term $0/0$ in the sum

$$\sum_{\bar{x}\in E_n} \mu_n\{\bar{x}\}/\sigma_n\{\bar{x}\}$$

could be interpreted as 1, because we are interested only in the minimum of such sums. Moreover, if E_n is an ε-code and $\bar{x} \in E_n$ with $\mu_n\{\bar{x}\} = 0$, then $E_n - \{\bar{x}\}$ will remain an ε-code. The quantity S_n^I will be denoted by the symbol $I_n(\varepsilon, \mu, \nu)$, i.e.

$$I_n(\varepsilon, \mu, \nu) = \min\{\nu_n(E_n) : E_n \subset X^n,\ \mu_n(E_n) > 1 - \varepsilon\} \tag{22.1}$$

and called the *n-dimensional I-divergence (at level* ε) for the pair (μ, ν) of sources. The coding theorem and its strong converse were proved in [29] for μ ergodic

Theorem 22.1 [29]. Let the pair (μ, ν) of sources satisfy the conditions (I) and (II). Let μ be an ergodic source. Then

$$-\lim_n \frac{1}{n} \log I_n(\varepsilon, \mu, \nu) = d(\mu, \nu)\ , \quad 0 < \varepsilon < 1\ . \tag{22.2}$$

(cf. 21.3)).

First of all we shall generalize Theorem 21.2 and then we shall proceed to the proof of the coding theorem along the lines of the preceding section.

Theorem 22.2. Let the pair (μ, ν) of sources satisfy the conditions (I) and (II). Then

$$\mu\left\{z : z \in R_X,\ \lim_n \frac{1}{n} \log \frac{\mu[z_1, \ldots, z_n]}{\nu[z_1, \ldots, z_n]} = d(\mu_z, \nu)\right\} = 1\ . \tag{22.3}$$

Indeed

$$\left|\frac{1}{n} \log \frac{\mu[z_1, \ldots, z_n]}{\nu[z_1, \ldots, z_n]} - d(\mu_z, \nu)\right| = \left|\frac{1}{n} \log \frac{\mu[z_1, \ldots, z_n]}{\nu[z_1, \ldots, z_n]} - K(\mu_z, \nu) + H(\mu_z)\right| \leqq$$

$$\leqq \left|\frac{1}{n} \log \mu[z_1, \ldots, z_n] + H(\mu_z)\right| + \left|\frac{1}{n} \log \nu[z_1, \ldots, z_n] - K(\mu_z, \nu)\right|.$$

Theorem 4.6 applies to the first term on the right-hand side of the latter inequality. Hence, this term can be made arbitrarily small with probability 1 when choosing n large enough. Similarly, theorem 21.5 applies to the remaining term and this proves the theorem.

Note that if μ is ergodic then the above theorem yields an improvement of Theorem 21.2. Namely, the stochastic convergence is replaced by the a.e. convergence.

We can repeat word by word the proof of Theorem 22.1 as given in [29] just using now Theorem 22.2 instead of Theorem 21.2 and obtain

Theorem 22.3. Let the pair (μ, ν) of sources satisfy the conditions (I) and (II). Then

$$\mu\left\{z : z \in R_X, -\lim_n \frac{1}{n} \log I_n(\varepsilon, \mu, \nu) = \mathrm{d}(\mu_z, \nu)\right\} = 1 . \tag{22.4}$$

Now we shall prove the basic lemmas needed for the coding theorem.

Lemma 22.4. Let the pair (μ, ν) of sources satisfy the conditions (I) and (II). Let c be a finite real number. The assumption that

$$\mu\{z : z \in R_X, \ \mathrm{d}(\mu_z, \nu) \leqq c\} = 1$$

implies the inequality

$$\limsup_n \left[-\frac{1}{n} \log I_n(\varepsilon, \mu, \nu)\right] \leqq c , \quad 0 < \varepsilon < 1.$$

Proof. From the definition of $I_n(\varepsilon, \mu, \nu)$ we conclude there is a set $F_n \subset X^n$ such that

$$\mu_n(F_n) > 1 - \varepsilon, \ \nu_n(F_n) = I_n(\varepsilon, \mu, \nu)$$

(because there are only finitely many subsets of X^n at all). Theorem 22.2 gives

$$\forall 0 < \varepsilon < 1 \ \forall \delta > 0 \ \exists n_0 \ \forall_n \geqq n_0$$

$$\mu\left\{z : z \in R_X, \frac{1}{n} \log \frac{\mu[z_1, \ldots, z_n]}{\nu[z_1, \ldots, z_n]} < \mathrm{d}(\mu_z, \nu) + \delta\right\} > 1 - \varepsilon .$$

Let

$$E_n^{(1)} = \left\{\bar{x} : \bar{x} \in E_n, \frac{\mu_n\{\bar{x}\}}{\nu_n\{\bar{x}\}} < 2^{n(C+\delta)}\right\} .$$

By the assumption we have

$$\mu_n(E_n^{(1)}) > 1 - \varepsilon .$$

Now

$$\nu_n(F_n) \geqq \nu_n(F_n \cap E_n^{(1)}) = \sum_{\bar{x} \in F_n \cap E_n^{(1)}} \nu_n\{\bar{x}\} > 2^{-n(c+\delta)} \sum_{\bar{x} \in F_n \cap E_n^{(1)}} \mu_n\{\bar{x}\} > 2^{-n(c+\delta)}(1 - 2\varepsilon) .$$

When choosing n_0 sufficiently large, we obtain the following statement:

$$\forall \delta > 0 \; \exists n_0 \; \forall n \geqq n_0 \; - \frac{1}{n} \log I_n(\varepsilon, \mu, \nu) < c + \delta .$$

But this implies the desired inequality because of the arbitrariness of δ.

Lemma 22.5. Let the pair (μ, ν) of sources satisfy the assumptions of the preceeding lemma. Let c be a finite real number. Then the relation

$$\mu\{z : z \in R_X, \; \mathrm{d}(\mu_z, \nu) \geqq c\} = 1$$

implies the inequality

$$\liminf_n \left[- \frac{1}{n} \log I_n(\varepsilon, \mu, \nu) \right] \geqq c , \quad 0 < \varepsilon < 1 .$$

Proof. By Theorem 22.2,

$$\forall 0 < \varepsilon < 1 \; \forall \delta > 0 \; \exists n_0 \; \forall n \geqq n_0$$

$$\mu \left\{ z : z \in R_X, \frac{1}{n} \log \frac{\mu[z_1, \ldots, z_n]}{\nu[z_1, \ldots, z_n]} > \mathrm{d}(\mu_z, \nu) - \delta \right\} > 1 - \varepsilon .$$

Let us set

$$E_n^{(2)} = \left\{ \bar{x} : \bar{x} \in X^n, \frac{\mu_n\{\bar{x}\}}{\nu_n\{\bar{x}\}} > 2^{n(c-\delta)} \right\} .$$

The assumption of the lemma yields again the inequality

$$\mu_n(E_n^{(2)}) > 1 - \varepsilon .$$

Consequently,

$$I_n(\varepsilon, \mu, \nu) \leqq \nu_n(E_n^{(2)}) = \sum_{\bar{x} \in E_n^{(2)}} \nu_n\{\bar{x}\} < 2^{-n(c-\delta)} \mu_n(E_n^{(2)}) < 2^{-n(c-\delta)} ,$$

i.e.

$$- \frac{1}{n} \log I_n(\varepsilon, \mu, \nu) > c - \delta .$$

Since δ was chosen arbitrarily, the desired inequality follows.

Now let us state some elementary properties of the quantities I_n similar to the properties of the quantities V_n established in [38]. The proofs are simple and therefore omitted.

Lemma 22.6. $0 < \varepsilon_2 < \varepsilon_1 < 1$ and for any $n \in N$ we have the inequalities

$$- \frac{1}{n} \log I_n(\varepsilon_1, \mu, \nu) \geqq - \frac{1}{n} \log I_n(\varepsilon_2, \mu, \nu) .$$

Lemma 22.7. Let $0 \leqq \zeta < 1$, let μ_1, μ_2, μ be stationary sources such that

$$\mu = (1 - \zeta)\,\mu_1 + \zeta\mu_2\,.$$

Then for all ε, $\zeta < \varepsilon < 1$, we have the inequality

$$\limsup_n \left[-\frac{1}{n} \log I_n(\varepsilon, \mu_1, \nu) \right] \leqq \limsup_n \left[-\frac{1}{n} \log I_n(\varepsilon - \zeta, \mu, \nu) \right].$$

Lemma 22.8. Let $0 < \zeta' \leqq 1$, let μ, μ_1, μ_2 be stationary sources such that

$$\mu = \zeta'\mu_1 + (1 - \zeta')\,\mu_2\,.$$

Then for all ε, $0 < \varepsilon < \zeta'$, we have the inequality

$$\liminf_n \left[-\frac{1}{n} \log I_n(\varepsilon, \mu_1, \nu) \right] \geqq \liminf_n \left[-\frac{1}{n} \log I_n(\varepsilon/\zeta', \mu, \nu) \right].$$

Repeating the proof of theorem I in [41] (cf. also the proof of Theorem 15.3 and [38] for analogous idea) we obtain the main theorem.

Theorem 22.9. Let pair (μ, ν) of sources satisfy the conditions (I) and (II). Then the inequality

$$\limsup_n \left[-\frac{1}{n} \log I_n(\varepsilon_1, \mu, \nu) \right] \leqq \liminf_n \left[-\frac{1}{n} \log I_n(\varepsilon_2, \mu, \nu) \right]$$

holds for $0 < \varepsilon_2 < \varepsilon_1 < 1$; consequently, the limit

$$-\lim_n \frac{1}{n} \log I_n(\varepsilon, \mu, \nu) = I_\varepsilon(\mu, \nu)$$

exists for all except at most a countable set of numbers ε. The function $I_\varepsilon(\mu, \nu)$ monotonically increases for $\varepsilon \to 0$ to a limit, which will be denoted by $I(\mu, \nu)$ and called the *asymptotic I-rate* of the pair (μ, ν) of sources.

Let us give again an equivalent statement.

Theorem 22.10. There exists one and only one nonnegative real-valued function I on the set $M(\mathscr{F}_X) \times M(\mathscr{F}_X; k)$ such that

(1) $$\forall \lambda > 0\ \forall 0 < \varepsilon < 1\ \exists n_0\ \forall n \geqq n_0\ \exists E_n \subset X^n\,,$$
$$\left[\mu_n(E_n) > 1 - \varepsilon\right] \text{ et } \left[\nu_n(E_n) < 2^{-n[I(\mu,\nu)-\lambda]}\right];$$

(2) $$\forall \lambda > 0\ \exists 0 < \eta < 1\ \forall 0 < \varepsilon \leqq \eta\ \exists n_0\ \forall n \geqq n_0\ \forall E_n \subset X^n\,.$$
$$.\left[\mu_n(E_n) > 1 - \varepsilon\right] \text{ implies } \left[\nu_n(E_n) > 2^{-n[I(\mu,\nu)+\lambda]}\right].$$

This means that the quantity $I(\mu, \nu)$ determines the rate of the exponential convergence of $I_n(\varepsilon, \mu, \nu)$ to 0. This fact will be used in the last part of the present paper in connection with the problems of the asymptotic optimality.

Corollary 22.11. If μ is an ergodic source, then

$$-\lim_n \frac{1}{n} \log I_n(\varepsilon, \mu, \nu) = I(\mu, \nu) = \mathrm{d}(\mu, \nu); \quad 0 < \varepsilon < 1.$$

It is again an improvement of the coding theorem given in [29] because it states also the equality

$$I(\mu, \nu) = d(\mu, \nu).$$

Finally, we have the following analogue of Theorem 21.12.

Theorem 22.12. The asymptotic I-rate of any pair (μ, ν) of sources satisfying the conditions (I) and (II) equals the essential supremum of the I-entropies $d(\mu_z, \nu)$; μ_z being the ergodic component of the stationary source μ; in symbols

$$I(\mu, \nu) = \operatorname*{ess\,sup}_{z \in R_X[\mu]} d(\mu_z, \nu). \tag{22.5}$$

At glance, it would seem to be possible to define the I-rate in the stationary non-ergodic case by means of the formula

$$I^*(\mu, \nu) = V(\mu, \nu) - V(\mu). \tag{22.6}$$

The necessary and sufficient conditions for the equalities $V(\mu, \nu) = K(\mu, \nu)$ and $I(\mu, \nu) = d(\mu, \nu)$ will not be studied in detail. We shall confine ourselves to the following statement.

Proposition 22.13. Let μ be a strongly stable source (cf. (18.3)). Then

$$I(\mu, \nu) = I^*(\mu, \nu).$$

Proof. Since μ is strongly stable, we have

$$V(\mu) = H(\mu)$$

(cf. Theorem 18.2), i.e.

$$\operatorname*{ess\,sup}_{z \in R_X[\mu]} H(\mu_z) = \int_{R_X} H(\mu_z)\, \mu(\mathrm{d}z).$$

Consequently,

$$H(\mu_z) = \int_{R_X} H(\mu_z)\, \mu(\mathrm{d}z) \quad \text{a.e.} \quad z[\mu],$$

i.e. $H(\mu_z)$ is almost everywhere a constant, namely $V(\mu)$. This in turn implies

$$I^*(\mu, \nu) = \operatorname*{ess\,sup}_{z \in R_X[\mu]} K(\mu_z, \nu) - H(\mu) = \operatorname*{ess\,sup}_{z \in R_X[\mu]} [K(\mu_z, \nu) - H(\mu_z)] =$$

$$= \operatorname*{ess\,sup}_{z \in R_X[\mu]} d(\mu_z, \nu) = I(\mu, \nu) .$$

The proposition is proved.

Other statements of a similar character can be proved following the same manner (cf. also Section 18 for a detailed discussion in the special case $\mu = \nu$).

23. The General Case

Here we shall turn back to the notations used in the third part of the present paper. The necessary notations concerning finite partitions are to be found in Section 10. The symbol X will denote a separable metric space. The conditions (I) and (II) are to be replaced by the following ones:

(I′) $\mu \in M(\mathscr{A}_X)$;

(II′) the source $\nu \in M(\mathscr{A}_X)$ is such that for any $\zeta \in Z_X$, $\bar{\nu}_\zeta$ satisfies the condition (II) of Section 21.

The condition (II′) is clearly satisfied if ν arises from a sequence of independent identically distributed random variables, the common distribution being possibly only finitely additive. As the Markov source examined in Section 5 shows, it is expected that the Markov sources satisfying the condition (II′) will be of a very special type.

Nevertheless, the ideas developed in the third part of the paper are working when confined to the pairs (μ, ν) satisfying the above conditions. Here we shall only summarize the results omitting their proofs. Only the important differences will be stated explicitely. Otherwise, the proofs may be obtained simply using the ideas developed in [37] and the present paper. Let

$$R^1(\mu, \nu) = \sup_{\zeta \in Z_X} K(\bar{\mu}_\zeta, \bar{\nu}_\zeta) = K(\mu, \nu) ; \tag{23.1}$$

$$R^2(\mu, \nu) = \sup_{\zeta \in Z_X} d(\bar{\mu}_\zeta, \bar{\nu}_\zeta) = d(\mu, \nu);$$

$$S^1(\mu, \nu) = \sup_{\zeta \in Z_X} V(\bar{\mu}_\zeta, \bar{\nu}_\zeta) = V(\mu, \nu) ;$$

$$S^2(\mu, \nu) = \sup_{\zeta \in Z_X} I(\bar{\mu}_\zeta, \bar{\nu}_\zeta) = I(\mu, \nu) .$$

Let $S_n^1 = L_n$, $S_n^2 = I_n$. We shall make use of the quantities

$$\bar{S}^i(\mu, \nu) = \sup_{\zeta \in Z_X} \bar{S}^i(\bar{\mu}_\zeta, \bar{\nu}_\zeta) ,$$

where

$$\bar{S}^i(\bar{\mu}_\zeta, \bar{\nu}_\zeta) = \lim_{\varepsilon \to 0} \limsup_n \frac{1}{n} \log S_n^i(\varepsilon, \bar{\mu}_\zeta, \bar{\nu}_\zeta) ; \quad i = 1, 2 ;$$

and of the quantities $\underline{S}^i(\mu, \nu)$ defined similarly by means of lim inf.

Theorem 23.1. Let (μ, ν) be a pair of sources satisfying the conditions (I′) and (II′). Then

$$R^i(\mu, \nu) = \int_{E(\mathscr{A}_X)} R^i(\varkappa, \nu)\, \hat{\mu}(\mathrm{d}\varkappa) ; \quad i = 1, 2 .$$

The relations

$$R^i(\bar{\mu}_\zeta, \bar{\nu}_\zeta) = \int_{E(\mathscr{A}_X)} R^i(\bar{\varkappa}_\zeta, \bar{\nu}_\zeta)\, \hat{\mu}(\mathrm{d}\varkappa) , \quad \zeta \in Z_X ; \quad i = 1, 2 ;$$

follow as in the proof of the main theorem in [37]. However, the fundamental lemma of Feinstein ([8], Lemma I.3) has to be replaced. For $i = 1$, we shall use

Lemma 23.2. Let $\zeta = \{C_1, \ldots, C_k\} \subset \mathscr{B}(X)$. Let

$$\xi = \{D_1, D_2, C_2, \ldots, C_k\}$$

with $D_1 \cap D_2 = \emptyset$, $D_1 \cup D_2 = C_1$. Then

$$\sum_{D \in \xi} \mu_1(D) \log \nu_1(D) \leqq \sum_{C \in \zeta} \mu_1(C) \log \nu_1(C) .$$

For $i = 2$, we shall use

Lemma 23.3. Let ζ, ξ be as in Lemma 23.2. Then

$$\log I_1(\varepsilon, \bar{\mu}_\zeta, \bar{\nu}_\zeta) \geqq \log I_1(\varepsilon, \bar{\mu}_\xi, \bar{\nu}_\xi) .$$

Lemma 23.4. Let the pair (μ, ν) of sources satisfy the conditions (I′) and (II′). If c is a finite real number then the assumption that

$$\hat{\mu}\{\varkappa : \varkappa \in E(\mathscr{A}_X),\ R^i(\varkappa, \nu) \leqq c\} = 1$$

implies the inequality

$$\bar{S}^i(\mu, \nu) \leqq c ; \quad i = 1, 2 .$$

Lemma 23.5. Under the assumptions of the preceeding lemma, the assumption that

$$\hat{\mu}\{\varkappa : \varkappa \in E(\mathscr{A}_X),\ R^i(\varkappa, \nu) \geqq c\} = 1$$

implies the inequality

$$\underline{S}^i(\mu, \nu) \geqq c ; \quad i = 1, 2 .$$

Finally, we have again the formula

$$S^i(\mu, \nu) = \operatorname*{ess\,sup}_{\varkappa \in E(\mathscr{A}_X)[\hat{\mu}]} R^i(\varkappa, \nu) . \tag{23.2}$$

24. The Interpretation of the Coding Theorems

Let μ be a stationary (possibly finitely additive) source on the space $(X^I, \mathscr{A}_X)$, X being a separable metric space. The sequence of the coordinate variables defined by the property that $X_n(z) = z_n$, $n \in I$, is the stationary process with the state space X and the distribution μ. Let $\zeta \in Z_X$. Any finite partition ζ of X can be interpreted as a measurement performed on the process $\{X_n\}$. The partition $\bigvee_{i=0}^{n-1} T^{-i}\zeta$ corresponds to the subsequent repetitions of the measurement ζ in times $0, 1, \ldots, n-1$. Hence $H(\mu, \zeta)$ can be interpreted as the uncertainty of the process $\{X_n\}$ discovered when performing the given measurement ζ. To compute the actual uncertainty of the process $\{X_n\}$ it is natural to consider the quantity

$$\sup_{\zeta \in Z_X} H(\mu, \zeta) . \tag{24.1}$$

But this is exactly the entropy rate of the process $\{X_n\}$. A similar natural interpretation can be given also to any of the quantities (23.1).

PART V: A STATISTICAL INTERPRETATION

25. The Notion of the Asymptotic Optimality

The methods proposed for the investigation of the asymptotic optimality of the sequences of tests fall into two categories.

There are "local" methods dealing with the asymptotic efficiency mainly. A sequence of alternatives is chosen in such a way that the probability of type II error is bounded away from 0 and 1, and the speed of convergence of this sequence of alternatives to the null-set is measured somehow. The speed is taken as the optimality criterion.

On the other hand, there are "nonlocal" methods. A fixed alternative is chosen. The rate of exponential convergence of the probability of type II error is considered. The size is either held fixed and bounded away from 1 or it is allowed to approach 0 exponentially with a prescribed rate.

The choice of the exponential convergence is justified both by the reasons of inference [6] and by its computational simplicity.

In this part we shall make use of the natural logarithms. Note that there are no problems when reformulating the results of the preceding part in this fashion.

Throughout this section we are given the finite set $\{1, 2, \ldots, k\}$ of all possible

experimental outcomes and the family $\mathscr{P}$ of all stationary Markov sources with the alphabet $\{1, \ldots, k\}$ and with entrywise positive transition probability matrices, respectively. The set of all such matrices will be denoted by Θ, hence

$$\mathscr{P} = \{P_\theta : \theta \in \Theta\} .$$

Note that the assumptions imply that for any $\theta^1, \theta^2 \in \Theta$ the finite-dimensional distributions $P_{\theta^1,n}$ and $P_{\theta^2,n}$ of P_{θ^1} and P_{θ^2} are mutually absolutely continuous even the same fails to hold for P_{θ^1} and P_{θ^2} themselves. This in turn implies that the quantities $I_n(\varepsilon, P_{\theta^1}, P_{\theta^2})$ as well as the quantities $d(P_{\theta^1}, P_{\theta^2})$ (cf. Theorem 21.2 and (22.1)) are always finite. The unique stationary distribution of the matrix θ will be denoted by p^θ (if $\theta = \theta^i$, we shall write $p = p^{(i)}$; for the existence and uniqueness cf. [2], Theorems 6.3.1 and 6.3.2). Moreover, any Markov source determined by a matrix $\theta \in \Theta$ is regular ([2], p. 185), hence ergodic ([2], Theorem 6.6.2). As already mentioned in Section 5,

$$H(P_\theta) = \sum_{i=1}^{k} p_i^\theta H(\theta_i) = -\sum_{i=1}^{k} p_i^\theta \sum_{j=1}^{k} \theta_{ij} \log \theta_{ij} . \tag{25.1}$$

Let $\theta^1, \theta^2 \in \Theta$, $\theta^1 \neq \theta^2$. Then P_{θ^1} and P_{θ^2} satisfy the assumptions (I) and (II) of Section 21. Hence we can make use of the asymptotic properties of the quantity

$$I_n(\varepsilon, P_{\theta^1}, P_{\theta^2}) = \min \{P_{\theta^2,n}(E_n) : E_n \subset \{1, 2, \ldots, k\}^n,\ P_{\theta^1,n}(E_n) > 1 - \varepsilon\} .$$

To simplify the notations we shall write P_θ for $P_{\theta,n}$ if there is no danger of confusion. Let us interpret the set E_n^c as the critical region of a test φ_n for testing the problem $P_{\theta^1,n} : P_{\theta^2,n}$. Then $I_n(\varepsilon, P_{\theta^1}, P_{\theta^2})$ represents the minimum probability of type II error subject to the constraint that the level does not exceed the value ε. Consequently, Corollary 22.11 represents a generalization of Stein's lemma (cf. [3], Section 6, and [29]).

Let $\Theta_0 \neq \emptyset$, $\Theta_0 \subset \Theta$. Let φ_n be a sequence of tests for the problem $\Theta_0 : \theta^1$ with $\theta^1 \in \Theta_1 = \Theta - \Theta_0$. More exactly, given any $n \in N$, the test φ_n is proposed for the n-dimensional problem

$$\Theta_0^{(n)} = \{P_{\theta^0,n} : \theta^0 \in \Theta_0\} : P_{\theta^1,n}$$

for $\theta^1 \in \Theta_1$. The probability of type I error will be denoted by $\alpha_n^\varphi(\theta)$, i.e.

$$\alpha_n^\varphi(\theta) = P_\theta\{\varphi_n \text{ rejects } \theta\} = \tag{25.2}$$

$$= P_{\theta,n}\{(z_1, \ldots, z_n) : \varphi_n(z_1, \ldots, z_n) \text{ rejects } P_{\theta,n}\} .$$

For any $\theta \in \Theta$,

$$\beta_n^\varphi(\theta) = 1 - \alpha_n^\varphi(\theta) . \tag{25.3}$$

The function function $\beta_n^\varphi(\cdot)$ defined on Θ_0 is called the power function of the test φ_n. For any $\theta^1 \in \Theta_1$, $\beta_n^\varphi(\theta^1)$ is the probability of type II error when testing the problem

$\Theta_0 : \theta^1$. The size of the test φ_n will be denoted by α_n^φ, i.e.

$$\alpha_n^\varphi = \sup \{\alpha_n^\varphi(\theta^0) : \theta^0 \in \Theta_0\} . \tag{25.4}$$

The following notions were introduced in [39] for the independent identically distributed case (i.e. for probability measures of the product type).

A sequence φ_n of tests is said to be *of rate A*, $0 \leqq A \leqq \infty$, provided

$$\limsup_n \alpha_n^\varphi < 1 \qquad \text{in case} \quad A = 0 , \tag{25.5}$$

$$\limsup_n \frac{1}{n} \log \alpha_n^\varphi \leqq -A \text{ in case } \quad A > 0 . \tag{25.6}$$

According to this definition, a sequence φ_n is of rate 0 provided that the size is bounded away from 1. Otherwise A is the rate of the exponential convergence $\alpha_n^\varphi \to 0$. The second situation is typical in the problems of nonlocal asymptotic optimality (cf. [3], [4], [6] and [39]). Let Φ_A denote the set of all sequences φ_n of rate A. A sequence $\varphi_n \in \Phi_A$ is said to be *exponential rate optimal* (ERO) at an alternative θ^1 if

$$\lim_n \frac{1}{n} \log \beta_n^\varphi(\theta^1) = -B \tag{25.7}$$

holds with the best possible constant B, i.e. with B such that for any sequence $\psi_n \in \Phi_A$,

$$\liminf_n \frac{1}{n} \log \beta_n^\varphi(\theta^1) \geqq -B . \tag{25.8}$$

Hence

$$B = B_A(\theta^1, \Theta_0) = \inf_{\varphi_n \in \Phi_A} \left\{ -\liminf_n \frac{1}{n} \log \beta_n^\varphi(\theta^1) \right\} . \tag{25.9}$$

Corollary 22.11 deals with the simplest testing problem $\theta^0 : \theta^1$, i.e. $\Theta_0 = \{\theta^0\}$, and $\varphi_n \in \Phi_0$:

$$\lim_n \frac{1}{n} \log \beta_n^\psi(\theta^1) = -d(P_{\theta^0}, P_{\theta^1}) .$$

Using a well-known explicit representation of $d(P_{\theta^0}, P_{\theta^1})$ (similar to (25.1)) we obtain also an explicit expression for the K-entropy $K(P_{\theta^0}, P_{\theta^1})$. Let $I(\theta_i^0, \theta_i^1)$ denote the I-divergence of the probability vectors θ_i^0, θ_i^1; $i = 1, 2, \ldots, k$, i.e.

$$I(\theta_i^0, \theta_i^1) = \sum_{j=1}^k \theta_{ij}^0 \log (\theta_{ij}^0 / \theta_{ij}^1) .$$

Then

$$d(P_{\theta^0}, P_{\theta^1}) = \sum_{i=1}^k p_i^{(0)} I(\theta_i^0, \theta_i^1) = \sum_{i=1}^k p_i^{(0)} \sum_{j=1}^k \theta_{ij}^0 \log (\theta_{ij}^0 / \theta_{ij}^1) . \tag{25.10}$$

Now since P_{θ^0} is ergodic, we have the equality

$$K(P_{\theta^0}, P_{\theta^1}) = d(P_{\theta^0}, P_{\theta^1}) + H(P_{\theta^0}) . \tag{25.11}$$

From (25.11), (25.1) and (25.10) it follows that

$$K(P_{\theta^0}, P_{\theta^1}) = \sum_{i=1}^{k} p_i^{(0)} H(\theta_i^0, \theta_i^1) = -\sum_{i=1}^{k} p_i^{(0)} \sum_{j=1}^{k} \theta_{ij}^0 \log \theta_{ij}^1 . \tag{25.12}$$

For any $\Theta' \subset \Theta$ let us set

$$d(\Theta', \theta) = \inf_{\theta' \in \Theta'} d(P_{\theta'}, P_\theta) ,$$

$$d(\theta, \Theta') = \inf_{\theta' \in \Theta'} d(P_\theta, P_{\theta'}) .$$

We shall say that an alternative θ^1 cannot be discriminated from the set Θ_0 if for any sequence φ_n,

$$\alpha_n^\varphi(\theta^1) \leqq \alpha_n^\varphi .$$

Let $\bar{\Theta}_0$ be the set of all such θ^1's. Clearly $\bar{\Theta}_0 \supset \Theta_0$, and if $\Theta_0 = \{\theta^0\}$ then $\bar{\Theta}_0 = \{\theta^0\}$. For any A, $0 \leqq A \leqq \infty$, let

$$\Theta_A = \{\theta : \theta \in \Theta,\ d(\theta, \Theta_0) < A\} , \tag{24.13}$$

$$\bar{\Theta}_A = \{\theta : \theta \in \Theta,\ d(\theta, \bar{\Theta}_0) < A\} .$$

Note that any sequence φ_n of rate A for Θ_0 is of the same rate A also for the extended hypothesis $\bar{\Theta}_0$. A more general result for $A > 0$ is established in the following

Lemma 25.1. Let φ_n be a sequence of rate A, $0 \leqq A \leqq \infty$. Then for every $\theta^1 \in \bar{\Theta}_A$ we have

$$\limsup_n \alpha_n^\varphi(\theta^1) < 1 \quad \text{in case} \quad A = 0 ;$$

$$\lim_n \alpha_n^\varphi(\theta^1) = 0 \quad \text{in case} \quad A > 0 . \tag{25.14}$$

For every $\theta^1 \in \Theta$

$$\liminf_n \frac{1}{n} \log \beta_n^\varphi(\theta^1) \geqq -d(\bar{\Theta}_A, \theta^1) . \tag{25.15}$$

Proof. (25.14) for $A = 0$ is trivial by the very definition of $\bar{\Theta}_0$. Let $A > 0$. Assume we are given $\theta^1 \in \bar{\Theta}_A$ and a subsequence $\{n_k\} \subset \{n\}$ such that

$$\limsup_k \beta_{n_k}^\varphi(\theta^1) < 1 .$$

Consider the problem $\theta^1 : \theta^0$ for $\theta^0 \in \bar{\Theta}_0$. Then

$$\liminf_k \frac{1}{n_k} \log \alpha_{n_k}^\varphi(\theta^0) \geqq -d(P_{\theta^1}, P_{\theta^0}) .$$

Since $\theta^1 \in \bar{\Theta}_A$, we have $d(\theta^1, \bar{\Theta}_0) < A$. Hence there is $\theta^0 \in \bar{\Theta}_0$ such that $d(P_{\theta^1}, P_{\theta^0}) < A$, i.e.

$$\liminf_k \frac{1}{n_k} \log \alpha^{\varphi}_{n_k}(\theta^0) > -A$$

contradictory to the assumption that $\varphi_n \in \Phi_A$. Hence

$$\limsup_k \beta^{\varphi}_{n_k}(\theta^1) = 1 ,$$

i.e.

$$\limsup_k (1 - \alpha^{\varphi}_{n_k}(\theta^1)) = 1 .$$

Since $\alpha^{\varphi}_{n_k}(\theta^1) \geqq 0$, it follows that

$$\lim_n \alpha^{\varphi}_n(\theta^1) = 0 .$$

Now let $B > d(\bar{\Theta}_A, \theta^1)$ be arbitrary. Then there is $\theta \in \bar{\Theta}_A$ such that $d(P_\theta, P_{\theta^1}) < B$. For $\theta \in \bar{\Theta}_A$ there is $\theta^0 \in \bar{\Theta}_0$ such that $d(P_\theta, P_{\theta^0}) < A$. The corollary 22.11 for the problem $\theta : \theta^1$ gives

$$\liminf_n \frac{1}{n} \log \beta^{\varphi}_n(\theta^1) \geqq -d(P_\theta, P_{\theta^1}) > -B .$$

The lemma is proved.

26. The Generalized Likelihood Ratio Test

Let $z_1, \ldots, z_n$ be a finite strip of a sample path of the length n. The trasition count matrix $\mathbf{A}^{(n)}(z)$ is defined entrywise as follows:

$$a^{(n)}_{ij}(z) = \operatorname{card}\{l : 1 \leqq l \leqq n-1,\ z_l = i,\ z_{l+1} = j\} ,$$

$i, j = 1, \ldots, k$. Let

$$b^{(n)}_i(z) = \sum_{j=1}^{k} a^{(n)}_{ij}(z) .$$

Then $\sum_{i=1}^{k} b^{(n)}_i(z) = n$ and we shall denote by $\mathbf{c}^{(n)}(z)$ the probability k-vector

$$(b^{(n)}_1(z)/n, \ldots, b^{(n)}_k(z)/n) .$$

Let

$$c^{(n)}_{ij}(z) = \begin{cases} a^{(n)}_{ij}(z)/b^{(n)}_i(z), & \text{if } b^{(n)}_i(z) > 0 , \\ 1/k \text{ otherwise} . \end{cases}$$

Then the matrix $\mathbf{C}^{(n)}(z)$ is a maximum likelihood estimate (MLE) of the true transition probability matrix θ and the vector $\mathbf{c}^{(n)}(z)$ is a MLE of its stationary distribu-

tion p^θ, respectively. Especially,

$$\lim_n \mathbf{C}^{(n)}(z) = \theta \quad \text{a.e.} \quad z[P_\theta] ;$$

$$\lim_n \mathbf{c}^{(n)}(z) = p^\theta \quad \text{a.e.} \quad z[P_\theta] .$$

Let

$$U_n(z, \theta) = \sum_{i=1}^{k} c_i^{(n)}(z)\, I(\mathbf{C}_i^{(n)}(z), \theta_i) = \tag{26.1}$$

$$= \sum_{i=1}^{k} c_i^{(n)}(z) \sum_{j=1}^{k} c_{ij}^{(n)}(z) \log \left(c_{ij}^{(n)}(z)/\theta_{ij}\right) .$$

If $g(\theta)$ is a function of the parameter θ, if $\hat{\theta}$ is a MLE of θ, then $g(\hat{\theta})$ is a MLE of $g(\theta)$. From this well-known property of MLE it follows that $U_n(z, \theta)$ has the following properties:

$$\lim_n U_n(z, \theta_0) = 0 \quad \text{a.e.} \quad z[P_{\theta^0}] , \tag{26.2}$$

$$\lim_n U_n(z, \theta^0) = d(P_{\theta^1}, P_{\theta^0}) \quad \text{a.e.} \quad z[P_{\theta^1}] .$$

The generalized likelihood ratio statistics is the function $T_n(z)$ defined by the formula

$$T_n(z) = \inf \{U_n(z, \theta^0) : \theta^0 \in \Theta_0\} . \tag{26.3}$$

The sequence T_n is asymptotically optimal in the sense of the exact slope [4]. Hence

$$\lim_n T_n(z) = \inf_{\theta^0 \in \Theta_0} d(P_{\theta^1}, P_{\theta^0}) = d(\theta^1, \Theta_0)\, [P_{\theta^1}] . \tag{26.4}$$

Theorem 26.1. Let Θ_0 be a convex subset of Θ. For any $0 < A \leqq \infty$ the sequence φ_n of LRT's defined by the property that

$$\varphi_n(z) = 1 \quad \text{iff} \quad T_n(z) > A$$

is of rate A. Moreover, it is ERO at any alternative θ^1, i.e.

$$\lim_n \frac{1}{n} \log \beta_n^{\varphi}(\theta^1) = -B_A(\theta^1, \Theta_0) .$$

To obtain a reasonable statistical inference, it is necessary to have $B_A(\theta^1, \Theta_0) > 0$. The sufficient condition is given by the following (cf. (26.5))

Lemma 26.2. If $0 < A < d(\theta^1, \theta_0)$ then $d(\overline{\Theta}_A, \theta^1) > 0$.

The lemma is obvious, so the proof is omitted. Concerning the exact slope of the sequence T_n the lemma shows that the exact slope gives the upper bound for the rate

of a sequence of tests such that the rate of the exponential convergence $\beta_n^{\varphi}(\theta^1) \to 0$ is still positive, for an arbitrary but fixed alternative θ^1.

Remark 26.3. Since

$$P_\theta\{\varphi_n \text{ rejects } \Theta_0\} = \alpha_n^{\varphi}(\theta) = P_\theta\{z : T_n(z) > A\}$$

the set $\{z : T_n(z) > A\}$ is the critical region of the test φ_n. The case $A = 0$ is excluded. In this case the size would approach 1, thus the sequence φ_n would become asymptotically useless for testing.

Remark 26.4. The first part of the theorem remains valid also without the assumption on the convexity of Θ_0.

Proof of the theorem. Let $\theta^0 \in \Theta_0$. Then

$$P_{\theta^0}\{z : U_n(z, \theta^0) \geqq A\} \leqq n^{k^2} e^{-nA}$$

(cf. [4]). Now

$$P_{\theta^0}\{z : T_n(z) \geqq A\} \leqq P_{\theta^0}\{z : U_n(z, \theta^0) \geqq A\}$$

for any $\theta^0 \in \Theta_0$ and any $A > 0$. Using the definitions of φ_n and α_n^{φ} one obtains immediately the inequality (25.6), thus proving the first part of the theorem.

To prove the second part of the theorem, let us note that

$$B_A(\theta^1, \Theta_0) = d(\bar{\Theta}_A, \theta^1). \tag{26.5}$$

Actually, for any sequence φ_n of rate A and for any $\theta^1 \in \Theta$ we have

$$\liminf_n \frac{1}{n} \log \beta_n^{\varphi}(\theta^1) \geqq -d(\bar{\Theta}_A, \theta^1),$$

i.e.

$$\inf\left\{-\liminf_n \frac{1}{n} \log \beta_n^{\varphi}(\theta^1)\right\} \leqq d(\bar{\Theta}_A, \theta^1).$$

$$\varphi_n \in \Phi_A.$$

This means that

$$B_A(\theta^1, \Theta_0) \leqq d(\bar{\Theta}_A, \theta^1).$$

If the inequality $B_A(\theta^1, \Theta_0) < d(\bar{\Theta}_A, \theta^1)$ would take place then there would be a number, say C, such that

$$B_A(\theta^1, \Theta_0) < C < d(\bar{\Theta}_A, \theta^1).$$

Then for any $\varphi_n \in \Phi_A$

$$\liminf_n \frac{1}{n} \log \beta_n^{\varphi}(\theta^1) \geqq -C$$

and, consequently, it would be impossible for $B_A(\theta^1, \Theta_0)$ to satisfy the relation (25.9). The relation (26.5) yields a definite geometrical meaning to Lemma 26.2. Because of Lemma 25.1 to prove the second part of the theorem it suffices to establish the inequality

$$\limsup_n \frac{1}{n} \log \beta_n^{\varphi}(\theta^1) \leqq -d(\bar{\Theta}_A, \theta^1) . \tag{26.6}$$

Let $\theta^0 \in \bar{\Theta}_0$, let $\theta^1 \in \Theta_1$ be arbitrary. Consider the testing problem $\theta^0 : \theta^1$. For any sequence ψ_n of rate 0 we have

$$\limsup_n \frac{1}{n} \log \beta_n^{\psi}(\theta^1) \leqq -d(P_{\theta^0}, P_{\theta^1}) .$$

Consequently, for any sequence ψ_n of rate 0 for the problem $\bar{\Theta}_0 : \theta^1$ we obtain the inequality

$$\limsup_n \frac{1}{n} \log \beta_n^{\psi}(\theta^1) \leqq -d(\bar{\Theta}_0, \theta^1) = -B_0(\theta^1, \Theta_0) .$$

The value $B_0(\theta^1, \Theta_0)$ is optimal for $A = 0$. Hence, when increasing the rate to some $A > 0$, the relation

$$\limsup_n \frac{1}{n} \log \beta_n^{\psi}(\theta^1) \leqq -B_0(\theta^1, \Theta_0)$$

takes place for any sequence $\psi_n \in \Phi_A$. Especially

$$\limsup_n \frac{1}{n} \log \beta_n^{\varphi}(\theta^1) = \limsup_n \frac{1}{n} \log P_{\theta^1}\{\bar{\mu}z : T_n(z) \leqq A\} \leqq -B_0(\theta^1, \Theta_0) . \tag{26.7}$$

Let $\theta^0 \in \bar{\Theta}_0$ be such that

$$d(P_{\theta^0}, P_{\theta^1}) = B_0(\theta^1, \Theta_0) = d(\bar{\Theta}_0, \theta^1) .$$

Because of Corollary 22.11 there is at least one such point θ^0 (otherwise the optimum rate would not be attainable) and due to the convexity of Θ_0, the point θ^0 is unique.

Since Θ_0 is a convex set, the set $\bar{\Theta}_A$ is convex as well. Therefore the point $\theta^* \in \bar{\Theta}_A$ minimizing the "distance" $d(P_\theta, P_{\theta^1})$ on $\bar{\Theta}_A$ lies on the "segment" connecting the points θ^0 and θ^1. Hence

$$d(P_{\theta^0}, P_{\theta^1}) = d(P_{\theta^*}, P_\theta) + A$$

i.e.

$$B_0(\theta^1, \Theta_0) = A + B_A(\theta^1, \Theta_0) . \tag{26.8}$$

Using the evident inequality

$$-B_0(\theta^1, \Theta_0) \leqq -B_0(\theta^1, \Theta_0) + A$$

together with (26.7) and (26.8) one obtains the desired inequality

$$\limsup_n \frac{1}{n} \log \beta_n^a(\theta^1) \geqq -B_A(\theta^1, \Theta_0) = -d(\bar{\Theta}_A, \theta^1) .$$

The theorem is proved.

To finish this section we shall derive a simple formula for the optimum rate provided Θ_0 is simple, i.e. $\Theta_0 = \{\theta^0\}$.

Proposition 26.5. Let $\theta^0, \theta^1 \in \Theta$; $\theta^1 \neq \theta^0$. Then for any A, $0 \leqq A \leqq \infty$, we have

$$B_{A+\varepsilon}(\theta^1, \{\theta^0\}) = B_A(\theta^1, \{\theta^0\}) - \varepsilon \tag{26.9}$$

provided $0 \leqq \varepsilon \leqq B_A(\theta^1, \{\theta^0\})$.

The proof follows from the fact that for any $\theta^0 \in \Theta$, the set $\bar{\Theta}_A$ is convex. Using (26.9) we obtain the following formula

$$B_A(\theta^1, \{\theta^0\}) = \sum_{i=1}^k p_i^{(0)} \sum_{j=1}^k \theta_{ij}^0 \log (\theta_{ij}^0/\theta_{ij}^1) - A . \tag{26.10}$$

The formula avoids a cumbersome computation of the optimum rate using the methods of the convex minimization.

27. Uncertainty of the Null Set and the Optimality

Let us start with the relation (25.11):

$$K(P_{\theta^0}, P_{\theta^1}) = d(P_{\theta^0}, P_{\theta^1}) + H(P_{\theta^0}) .$$

Given the sample path $z_1, \ldots, z_n$, we can define

$$\tilde{U}_n(z, \theta^0) = \sum_{i=1}^k c_i^{(n)}(z) \sum_{j=1}^k c_{ij}^{(n)}(z) \log (1/\theta_{ij}^0) , \tag{27.1}$$

where

$$-\sum_{j=1}^k c_{ij}^{(n)}(z) \log \theta_{ij}^0 = H(\mathbf{C}_i^{(n)}(z), \theta_i^0)$$

is the inaccuracy as introduced by Kerridge [18]. Then

$$\tilde{U}_n(z, \theta^0) = \sum_{i=1}^k c_i^{(n)}(z) \sum_{j=1}^k c_{ij}^{(n)}(z) \log (c_{ij}^{(n)}(z)/\theta_{ij}^0) - \tag{27.2}$$

$$- \sum_{i=1}^k c_i^{(n)}(z) \sum_{j=1}^k c_{ij}^{(n)}(z) \log c_{ij}^{(n)}(z) = U_n(z, \theta^0) + H(\mathbf{C}^{(n)}(z)) .$$

Here, $H(\mathbf{C}^{(n)}(z))$ is a MLE of $H(P_\theta)$ provided θ is the true parameter. Hence the function $\tilde{U}_n(z, \theta^0)$ consists of the term $U_n(z, \theta^0)$ appropriate for the discrimination

(or testing) and of a MLE of the entropy of the true parameter, respectively. Especially,

$$\lim_n \tilde{U}_n(z, \theta^0) = H(P_{\theta^0}) \quad [P_{\theta^0}] ;$$

$$\lim_n \tilde{U}_n(z, \theta^0) = d(P_{\theta^1}, P_{\theta^0}) + H(P_{\theta^1}) \quad [P_{\theta^1}] ;$$

i.e.

$$\lim_n \tilde{U}_n(z, \theta^0) = K(P_{\theta^1}, P_{\theta^0}) \quad [P_{\theta^1}] .$$

From the latter relations it is clear that if

$$\sup \{H(P_{\theta^0}) : \theta^0 \in \Theta_0\}$$

is small enough (i.e. the null-hypothesis consists of almost deterministic sources) the asymptotic behaviour of the statistics

$$\tilde{T}_n(z) = \inf \{\tilde{U}_n(z, \theta^0) : \theta^0 \in \Theta_0\}$$

is nearly the same as the asymptotic behaviour of the optimal statistics $T_n(z)$. Otherwise, the sequence $\tilde{T}_n(z)$ may be far from being ERO at many alternatives θ^1. Hence, the uncertainty of the null set can be considered as a nuissance parameter. It is an open problem, whether this reasoning, in general, remains true. More precisely, there is the following problem:

Suppose $\varphi_n \in \Phi_A$ is not an ERO sequence. Does the corresponding sequence T_n of statistics necessarily involve an estimate (or a function of it) of the uncertainty of the null set?

28. Concerning the Stationary Non-Ergodic Case

Let $\mathscr{P}$ be the set of all stationary probability measures on the measurable space $(\{1, \ldots, k\}^N, \mathscr{F}_k)$, where $\mathscr{F}_k$ is the σ-field generated by the family of all finite-dimensional cylinder sets. Let $\mathscr{P}_0$ consists of all mixtures of the ergodic sources $P_\theta : \theta \in \Theta$, i.e. $P_0 \in \mathscr{P}_0$ provided there is a probability measure ξ on the space $(\Theta, \mathscr{B}(\Theta))$ such that

$$P_0 = \int_\Theta P_\theta \, \xi(\mathrm{d}\theta) ;$$

$\mathscr{B}(\Theta)$ being the Borel σ-field in Θ. The alternatives will be chosen from the set $\mathscr{P}_1 = \{P_\theta : \theta \in \Theta\}$.

Let us consider the testing problem $P_0 : P_1$ with $P_i \in \mathscr{P}_i$ $(i = 0, 1)$. To dicuss the matter it is worthwile to reformulate the coding theorem 22.10 in statistical terms within the framework adopted in this part.

Let φ_n be a sequence of tests for the problem $P_0 : P_1$ with the probability of type I error bounded from above by ε, $\varepsilon \in (0, 1)$. Let $E_n^c \subset \{1, \ldots, k\}^n$ be the critical region of the test φ_n, $n = 1, 2, \ldots$

(1) Given any $\lambda > 0$ there is a test φ_n such that

$$\alpha_n^{\varphi}(P_0) = P_0(E_n^c) < \varepsilon$$

and

$$\beta_n^{\varphi}(P_1) = P_1(E_n) < e^{-n[I_\varepsilon(P_0, P_1) - \lambda]},$$

provided n is sufficiently large.

This means that

$$\limsup_n \alpha_n^{\varphi}(P_0) < 1$$

and

$$\limsup_n \frac{1}{n} \log \beta_n^{\varphi}(P_1) \leqq -I_\varepsilon(P_0, P_1).$$

(2) Given any $\lambda > 0$, there is η, $0 < \eta < 1$ such that for all ε, $0 < \varepsilon \leqq \eta$ the probability of type II error for any sequence φ_n of tests with the size bounded above by ε is

$$\beta_n^{\varphi}(P_1) > e^{-n[I_\varepsilon(P_0, P_1) + \lambda]},$$

provided n is sufficiently large.

This means that

$$\liminf_n \frac{1}{n} \log \beta_n^{\varphi}(P_1) \geqq -I_\varepsilon(P_0, P_1)$$

(cf. Theorem 22.9). Hence in the stationary non-ergodic case the optimum rate B depends on ε for $A = 0$. According to the Theorem 22.9

$$\lim_{\varepsilon \to 0} I_\varepsilon(P_0, P_1) = I(P_0, P_1)$$

and

$$I(P_0, P_1) \geqq I_\varepsilon(P_0, P_1) \quad \text{for any} \quad \varepsilon > 0.$$

Therefore it is impossible to use the reasoning of the preceding sections to solve the problems of the asymptotic optimality. In what follows we shall reduce the problem $P_0 : P_1$ to the problem $P_0' : P_1$ with P_0' ergodic in such a way that the two problems become asymptotically equivalent. The main idea is due to Gray and Davisson [11], [12]. According to the ergodic decomposition theorem (cf. Section 6)

$$P_0(E) = \int_{R_k} \mu_z(E)\, P_0(\mathrm{d}z). \tag{28.1}$$

Given any $z \in R_k$, let

$$\Omega_z = \{y : y \in R_k,\ \mu_y = \mu_z\}\,.$$

Then the family $\{\Omega_z\}$ constitutes (a possibly uncountable) partition of the set R_k of all regular points. Let us choose a representative sequence z from each Ω_z. The set of all chosen representative sequences will be denoted by $\mathfrak{Z}$; it is the new parameter set. Let

$$Z(\omega) = z \quad \text{if} \quad \omega \in \Omega_z\,.$$

Then $Z : R_k \to \mathfrak{Z}$. If

$$\mathscr{F}\mathfrak{Z} = \{A : A \subset \mathfrak{Z},\ Z^{-1}A \in \mathscr{F}_k \cap R_k\}$$

then Z is a random variable. Let $W(A) = P_0(Z^{-1}A)$. Then

$$P_0(E \cap Z^{-1}A) = \int_A P_z(E)\, W(\mathrm{d}z)\,,$$

where P_z is the ergodic measure uniquely determined by the sequence z. Now $P_z \in \mathscr{P}$ (cf. [16] for the ergodic decomposition of Markov sources). Note that

$$P_z(E) = P_0(E \mid \omega \in \Omega_z)\ [W]\,,$$

or

$$P_{Z(\omega)}(E) = P_0(E \mid \omega \in \Omega_{z(\omega)})\ [P_0]\,.$$

The detailed construction of the conditional probability within our setting was given by Rochlin [32]. The latter relations are the precise formulation of a well-known fact that stationarity can be replaced by ergodicity simply using the conditional probabilities. Since the ergodic sets Ω_z are disjoint, it is natural to think of a stationary source as of the result of nature randomly chosing a particular ergodic source at time minus infinity, and then sending it forever [11]. This in turn implies that it is natural to search for an estimate of the "true" ergodic source.

Let P_z, P_y be two different ergodic sources. Let

$$\varrho_n(P_z, P_y) = \sum_{i=1}^{n} \frac{1}{2^i} \sum_{\mathbf{x} \in \{1,\ldots,k\}^i} \left|P_z[\mathbf{x}] - P_y[\mathbf{x}]\right|\,.$$

Let $RF_{\omega,n}$ denote the empirical probability obtained by means of a sample $\omega_1, \omega_2, \ldots$ $\ldots, \omega_n$. The estimate $\hat{P}_{Z(\omega,n)}$ of the true ergodic source will be defined in terms of the metric ϱ_n:

$$\hat{P}_{Z(\omega,n)} = P_z$$

for which

$$\varrho_n(RF_{\omega,n}, P_z) \leqq \varrho_n(RF_{\omega,n}, P_y) + \varepsilon_n\,, \quad y \neq z\,.$$

Let

$$\varrho(P_z, P_y) = \lim_n \varrho_n(P_z, P_y) = \sup_n \varrho_n(P_z, P_y) .$$

If $\varepsilon_n \to 0$ and if $\omega \in \Omega_z$, i.e. if P_z is the true ergodic source, we have

$$\lim_n \varrho(\hat{P}_{Z(\omega,n)}, P_z) = 0 \tag{28.2}$$

(cf. [11]). This means that we can replace asymptotically the original testing problem $P_0 : P_1$ by the problem $P_z : P_1$, where P_z is the "true" ergodic source. Since $P_z \in \mathscr{P}_1 = \{P_\theta : \theta \in \Theta\}$, there is a matrix $\theta(z)$ such that $P_z = P_{\theta(z)}$. Clearly $P_1 = P_{\theta^1}$ for some matrix $\theta^1 \in \Theta$. Hence the new problem $P_{\theta(z)} : P_{\theta^1}$ is of the type solved in the preceding sections.

An alternative approach is given by Bahadur and Raghavachari [4]. The main idea is to consider the conditional tests. This means that instead of the original sequence T_n of statistics we shall use the sequence $T_n(\cdot \mid \omega \in \Omega_z)$. Some regularity conditions are necessary to obtain some reasonable results. But we shall not go into details in this paper.

The replacement of the original problem $P_0 : P_1$ by the problem $P_{\theta(z)} : P_{\theta^1}$ avoids one serious gap. We have actually replaced the testing problem $P_0 : P_1$ in the preceding sections by the sequence $P_{0,n} : P_{1,n}$ of finite-dimensional testing problems. It is not clear whether a sequence φ_n of tests for the problems $P_{0,n} : P_{1,n}$ converges (in some sense) to a test φ for the original problem $P_0 : P_1$. Moreover, we do not know whether certain optimality properties of every φ_n imply the same optimality property for the test φ provided φ exists. Of course, this problem does not arise when the probabilities P_0 and P_1 correspond to sequences of independent identically distributed random variables. If $P_{\theta(z)}$ and P_{θ^1} are obtained as above (and, consequently, they are ergodic) then the replacement of the testing problem by the sequence of the finite-dimensional problems is correct, too. Actually, let $I_n(z_n; Z \mid z_1, \ldots, z_{n-1})$ denote the amount of information contained in z_n about the unknown parameter Z given the first $n-1$ observations. Then

$$\lim_n I_n(z_n; Z \mid z_1, \ldots, z_{n-1}) = 0 \tag{28.3}$$

(cf. [11], Theorem 5.1). So, if n is large enough, the supplementary information provided by the subsequent observations becomes negligible. Thus a correct decision in the problem $P_{0,n} : P_{1,n}$ can be considered as a correct decision in the problem $P_0 : P_1$.

APPENDIX A: THE METHOD OF THE INFORMATION TRANSMISSION

Throughout this part we are given a countably infinite alphabet, represented by the set N of all positive integers. A sligthly more general result will be proved than that given in Remark 12.2 (cf. (12.3)).

Let us start with some basic notions. A one-parameter family $\nu = \{\nu(\cdot \mid y) : y \in N^I\}$ of σ-additive probability measures on the σ-field $\mathscr{F}$ is called a *channel* provided the following measurability condition is fullfilled: $\forall E \in \mathscr{F}$ $\nu(E \mid \cdot)$ is $\mathscr{F}$-measurable function on N^I. A channel ν is called *stationary* if

$$\nu(TA \mid Ty) = \nu(A \mid y), \quad A \in \mathscr{F}, \quad y \in N^I. \tag{A.1}$$

A stationary channel ν is called *historyless* if

$$\forall A, B \subset N^n \ \forall y, y' \in [B]_{0,n}.$$
$$\nu([A] \mid y) = \nu([A] \mid y'). \tag{A.2}$$

If the parameter space Y as well as the set X on which the probability measures $\nu(\cdot \mid y)$ are defined are countable, we call ν a (Y, X)-channel. The condition (A.2) implies that the relation

$$\nu_n(E \mid (y_0, \ldots, y_{n-1})) = \nu([E] \mid y) \tag{A.3}$$

determines a (N^n, N^n)-channel ν_n. For every $n \in N$ the channel ν can be characterized be the n-dimensional ε-size – the maximum number of the input signals (of the given length n) which are distinguishable by means of the output signals with the probability larger than $1 - \varepsilon$. In symbols,

$$S_n(\varepsilon, \nu) = \sup_{\psi} S_n(\psi, \varepsilon, \nu). \tag{A.4}$$

Here the supremum is taken over the family of all mappings $\psi : N^n \to N^n$, and

$$S_n(\psi, \varepsilon, \nu) = \operatorname{card}\{\mathbf{y} : \mathbf{y} \in N^n, \ \nu_n(\psi^{-1}\{\mathbf{y}\} \mid \mathbf{y}) > 1 - \varepsilon\}.$$

Remark A.1. Let us recall that for a given $\psi : N^n \to N^n$ the family

$$\{(\mathbf{y}, \psi^{-1}\{\mathbf{y}\}) : \mathbf{y} \in N^n, \ \nu_n(\psi^{-1}\{\mathbf{y}\} \mid \mathbf{y}) > 1 - \varepsilon\}$$

is nothing but the n-dimensional ε-code of length $S_n(\psi, \varepsilon, \nu)$ in the sense of Wolfowitz [42]. The coding theorem deals with the asymptotic behaviour of the sequence

$$\frac{1}{n} \log S_n(\varepsilon, \nu).$$

In the special case of the historyless channels the coding theorem can be formulated as follows (cf. [40], Lemma 6.1):

Theorem A.2. Let ν be a historyless channel. Then there is a real number $C(\nu)$ such that

(a) $$\forall 0 < \varepsilon < 1 \quad C(\nu) \leqq \liminf_n \frac{1}{n} \log S_n(\varepsilon, \nu) ;$$

(b) $$\forall t > C(\nu)\ \exists 0 < \lambda < 1\ \forall 0 < \varepsilon \leqq \lambda \quad \limsup_n \frac{1}{n} \log S_n(\varepsilon, \nu) < t .$$

As a consequence

$$C(\nu) = \lim_{\varepsilon \to 0} \limsup_n \frac{1}{n} \log S_n(\varepsilon, \nu) = \lim_{\varepsilon \to 0} \liminf_n \frac{1}{n} \log S_n(\varepsilon, \nu) .$$

The number $C(\nu)$ is said to be the *capacity* of the historyless channel ν.

Now let $n \in N$, let $\zeta \in Z_N$. The mappings $\varkappa : \zeta^n \to N^n$ and $\delta : N^n \to \zeta^n$ are called the coding and the decoding transformations, respectively. Note that these notions have nothing common with the notion of the ε-code defined above. The n-dimensional error probability (given ζ) is the number

$$e_n(\mu, \nu, \varkappa, \delta, \zeta) = 1 - \sum_{D \in \zeta^n} \nu_n(\delta^{-1} D \mid \varkappa D)\, \mu[D] .$$

The minimal n-dimensional error probability is the number

$$e_n(\mu, \nu, \zeta) = \inf_{\varkappa, \delta} e_n(\mu, \nu, \varkappa, \delta, \zeta) .$$

Let $\mu \in M(\mathscr{A})$. The source μ is said to be *representable* if there is a probability space $(S, \mathscr{S}, \lambda)$ and an $\mathscr{S}$-measurable family

$$\{\mu_s : s \in S\} \subset M(\mathscr{A})$$

such that

(A.5) $$\mu(A) = \int_S \mu_s(A)\, \lambda(\mathrm{d}s) .$$

Let $(S, \mathscr{S}) = (E(\mathscr{A}), \mathscr{K}[E(\mathscr{A})])$, let $\{\mu_s\} = E(\mathscr{A})$. Then clearly each $\mu \in M(\mathscr{A})$ is representable (cf. Theorem 7.3).

Lemma A.3. Let μ be representable. Let

$$\underline{V}(\mu_s) \leqq c < \infty , \quad s \in S .$$

Then

$$\overline{V}(\mu) \leqq c .$$

Proof. Let $\underline{V}(\mu_s) \leqq c < \infty$, $s \in S$. This implies

$$\forall s \in S\ \forall \zeta \in Z_N \quad \underline{V}(\mu_s, \zeta) \leqq c .$$

Since $\underline{V}(\mu, \zeta) = \underline{V}(\mu\tau_\zeta^{-1})$, we obtain

$$\underline{V}(\mu_s\tau_\zeta^{-1}) \leqq c\,; \quad s \in S\,, \quad \zeta \in Z_N\,.$$

Using Lemma 7.3 [40] we obtain

$$\overline{V}(\mu\tau_\zeta^{-1}) \leqq c\,, \quad \zeta \in Z_N\,,$$

i.e.

$$\overline{V}(\mu) = \sup_{\zeta \in Z_N} \overline{V}(\mu\tau_\zeta^{-1}) \leqq c\,.$$

The lemma is proved.

The generalization of the Theorem 16.3 is given in

Theorem A.4. Let μ be a representable source with a countable alphabet (i.e. (A.5) takes place). Then

$$V(\mu) = \operatorname{ess.sup}_{s \in S[\lambda]} V(\mu_s)\,. \tag{A.6}$$

Proof. We have

$$\{\mu : L_n(\varepsilon, \bar{\mu}_\xi) \leqq k\} =$$

$$= \{\mu : \min\{\operatorname{card}(\xi) : \xi \subset \zeta^n,\ \sum_{D \in \xi} \mu[D] > 1 - \varepsilon\} \leqq k\} =$$

$$= \bigcup \{\{\mu : \sum_{D \in \xi} \mu[D] > 1 - \varepsilon\} : \xi \subset \zeta^n,\ \operatorname{card}(\xi) \leqq k\}\,.$$

Consequently, $L_n(\varepsilon, \mu_s\tau_\zeta^{-1})$ is an $\mathscr{S}$-measurable function, hence $V(\mu_s)$ is an $\mathscr{S}$-measurable function on S. Therefore the number h is well-defined by the relation

$$h = \operatorname{ess.sup}_{s \in S[\lambda]} V(\mu_s)\,.$$

1. Let $V(\mu) > h$. Let c be a finite real number such that $V(\mu) > c > h$. Let us denote

$$A(c) = \{s : s \in S,\ V(\mu_s) < c\}\,,$$

then $A(c) \in \mathscr{S}$ and $\lambda(A(c)) = 1$. Indeed, let for all $c \in (h, V(\mu))$ we would have $\lambda(A(c)) < 1$. Then the definition of the essential supremum would give the contradictory inequality

$$\operatorname{ess.sup}_{s \in S[\lambda]} V(\mu_s) = h \geqq V(\mu)\,.$$

Since $\lambda(A(c)) = 1$, it suffices to apply Lemma A.3 to the measurable space $(A(c), \mathscr{S} \cap A(c))$. Then the contradictory inequalities

$$V(\mu) \leqq c < V(\mu)$$

follow. Hence we have $V(\mu) \leqq h$.

2. The proof will be finished by showing that the strict inequality $V(\mu) < h$ yields a contradiction. Let us assume its validity. Then there is a finite real number c such that $V(\mu) < c < h$. Let us denote

$$\boldsymbol{P} = \{\boldsymbol{p} : \boldsymbol{p} = \{p_n\}_{n=1}^{\infty},\ p_n \geqq 0,\ \sum_{1}^{\infty} p_n = 1,\ \exists n_0 = n_0(\boldsymbol{p})\ \forall n \geqq n_0,\ p_0 = 0\}\ .$$

Let

$$k[\boldsymbol{p}] = \min\{k : \sum_{i=1}^{k} p_i = 1\}\ ,\quad \boldsymbol{p} \in \boldsymbol{P}\ .$$

Let $\boldsymbol{p} \in \boldsymbol{P}$ and $k \in N$, $k > k[\boldsymbol{p}]$. We shall define a stationary memoryless channel ν by setting

$$\nu_1(m \mid s) = p_{k+m-s+1},\ m < s\ ;$$
$$\nu_1(m \mid s) = p_{m-s+1},\ m \geqq s\ (m = 1, 2, \ldots, k)\ ;$$
$$\nu_1(m \mid kt + s) = \nu_1(m \mid s),\ s = 1, 2, \ldots, k;\ t \in N\ .$$

Since a memoryless channel ν possesses the property

$$\nu_n(\boldsymbol{x} \mid \boldsymbol{y}) = \prod_{i=1}^{n} \nu_1(x_i \mid y_i)$$

for $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$, the knowledge of ν_1 uniquely determines the channel ν (cf. also [40], p. 795).

The channels of this type will be denoted by the symbol $\nu[\boldsymbol{p}, k]$ ($\boldsymbol{p} \in \boldsymbol{P}$, $k \in N$, $k > k[\boldsymbol{p}]$) and called the *circulant* channels. Following Lemma 6.3 [40] for any c, $0 \leqq c \leqq \infty$, there is a circulant channel ν such that $C(\nu) = c$. Let us note that the circulant channels form a subfamily of the family of all stationary historyless channels, hence the capacity $C(\nu)$ can be defined by means of Theorem A.2. Therefore we have

(A.7) $$V(\mu) < C(\nu) < h\ ,$$

i.e.

$$\overline{V}(\bar{\mu}_\zeta) < C(\nu) < h\ ,\quad \zeta \in Z_N\ .$$

A glance at the definitions of $\overline{V}(\bar{\mu}_\zeta)$ and $C(\nu)$, respectively, gives

(A.8) $$\forall \varepsilon > 0\ \exists n_0 \in N\ \forall n \geqq n_0\ L_n(\varepsilon, \bar{\mu}_\zeta) \leqq S_n(\varepsilon, \nu)\ .$$

Consequently, there is a mapping $\psi : N^n \to N^n$ such that

(A.9) $$L = L_n(\varepsilon, \bar{\mu}_\zeta) \leqq S_n(\psi, \varepsilon, \nu) = S\ .$$

This means that there are the points $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^S \in N^n$ such that

(A.10) $$\nu_n(\psi^{-1}\{\boldsymbol{y}^i\} \mid \boldsymbol{y}^i) > 1 - \varepsilon\ ,\quad i = 1, 2, \ldots, S\ .$$

Let us denote by $C^1, \ldots, C^L, C^{L+1}, \ldots, C^k$ the elements of the finite partition ζ^n. Since $L \leqq S$, we can define

$$\varkappa C^i = \mathbf{y}^i, \quad i = 1, 2, \ldots, L;$$
$$\varkappa C = \mathbf{y}^S \quad \text{for} \quad C \notin \{C^1, \ldots, C^L\}$$

(if $L = S$, we can use only the set $\{C^1, \ldots, C^{L-1}\}$). Further we shall put

$$\delta^{-1}C^i = \psi^{-1}(\varkappa C^i); \quad i = 1, 2, \ldots, L, \tag{A.11}$$
$$\delta^{-1}C = \emptyset \quad \text{otherwise}.$$

Then

$$1 - e_n(\mu, \nu, \varkappa, \delta, \zeta) = \sum_{i=1}^{L} \nu_n(\delta^{-1}C^i \mid \varkappa C^i)\, \mu[C^i]_{0n} + \sum_{C \neq C^i} \nu_n(\delta^{-1}C \mid \varkappa C)\, \mu[C]_{0n}.$$

(A.11) implies that the right-most term vanishes. Using this fact and (A.10), the inequality

$$1 - e_n(\mu, \nu, \varkappa, \delta, \zeta) \geqq (1 - \varepsilon) \sum_{i=1}^{L} \mu[C^i]_{0n} > 1 - 2\varepsilon$$

can be derived. This means that

$$\forall \varepsilon > 0 \ \exists n_0 \ \forall n \geqq n_0 \ \ e_n(\mu, \nu, \zeta) < 2\varepsilon. \tag{A.12}$$

Let us choose

$$\varepsilon = \lambda\{s : s \in S, \ V(\mu_s) > c\}$$

The inequality $c < h$ implies $\varepsilon > 0$ and (A.12) shows that there is increasing sequence $\{n(k)\}_{k=1}^{\infty}$ of positive integers such that

$$e_{n(k)}(\mu, \nu, \varkappa_k, \delta_k, \zeta) < (\varepsilon/2^k)^2 \tag{A.13}$$

for a convenient pair $\varkappa_k, \delta_k$. Now clearly

$$e_{n(k)}(\mu, \nu, \varkappa_k, \delta_k, \zeta) = \int_S e_{n(k)}(\mu_s, \nu, \varkappa_k, \delta_k, \zeta)\, \lambda(\mathrm{d}s).$$

Let

$$S_{n(k)} = \{s : s \in S, \ e_{n(k)}(\mu_s, \nu, \varkappa_k, \delta_k, \zeta) < \varepsilon/2^k\}.$$

By the definition of ε and of $S_{n(k)}$, respectively, there is an element $s \in S$ possessing the following two properties

(a) $$\overline{V}(\mu_s, \zeta) > C(\nu);$$

(b) $$e_{n(k)}(\mu_s, \nu, \varkappa_k, \delta_k, \zeta) < 1/2^k, \quad k \in N.$$

Now it is intuitively clear that the properties (a) and (b) will give the desired contradiction. Indeed, if the amount of information on the input of a channel exceeds the capacity of the channel (cf. (a)), it is impossible to transmit the information with the error probability as close to zero as wanted (cf. (b)). Now we shall give a formal deduction of this contradiction.

Let us assume that (a) holds. Since $\mu_s = \mu_s T^{-1}$, $\overline{V}(\mu_s, \zeta) = \underline{V}(\mu_s, \zeta) > C(\nu)$. Choose t such that

$$\underline{V}(\mu_s, \zeta) > t > C(\nu) .$$

This means that

$$\text{(A.14)} \qquad \forall \eta \in (0, 1)\ \exists\, n_0(\eta)\ \forall n \geqq n_0(\eta) \quad S_n(\eta, \nu) < L_n(\mu_s, \eta, \zeta)$$

Let $\varkappa_n : \zeta^n \to N^n$ and $\delta_n : N^n \to \zeta^n$ be arbitrary. Then define the mapping $\varphi : N^n \to N^n$ satisfying the relations

$$\mu_s([\varphi \mathbf{y}]) = \max \{\mu_s[E] : \varkappa_n E = \mathbf{y}\} , \quad \mathbf{y} \in \varkappa_n(\zeta^n) .$$

If $k = \operatorname{card}(\zeta)$ then $\operatorname{card}(\zeta^n) = k^n$. From the definition of the mapping φ it follows that

$$\varphi^{-1}\{\mathbf{y}\} = \bigcup \{\delta_n^{-1} E : \varkappa_n E = \mathbf{y}\} , \quad \mathbf{y} \in N^n .$$

The inequality (A.14) gives

$$L_n(\eta, \mu_s, \zeta) > S_n(\varphi, \eta, \nu) = S_n .$$

Now there are the points $\mathbf{y}^1, \ldots, \mathbf{y}^{S_n} \in N^n$ such that

$$\nu_n(\varphi^{-1}\{\mathbf{y}^i\} \mid \mathbf{y}^i) > 1 - \eta , \quad i = 1, 2, \ldots, S_n .$$

Let us denote further

$$q = \sum_{k=1}^{S_n} \mu_s[\varphi \mathbf{y}^k] .$$

Then $q \leqq 1 - \varepsilon$. Now let us compute

$$1 - e_n(\mu_s, \nu, \varkappa_n, \delta_n, \zeta) = \sum_{j=1}^{k^n} \sum_{E \in \varkappa_n^{-1} \mathbf{y}^j} \nu_n(\delta_n^{-1} E \mid \varkappa_n E)\, \mu_s[E] \leqq$$

$$\leqq \sum_{j=1}^{k^n} \nu_n(\varphi^{-1}\{\mathbf{y}^j\} \mid \mathbf{y}^j)\, \mu_s[\varphi \mathbf{y}^j] \leqq q + (1 - \eta)(1 - q) \leqq 1 - \eta^2 ,$$

a contradiction. The theorem is proved.

Intuitively speaking, a numerical characteristic of the amount of information produced by an information source has to fullfill the following two conditions in order to be an effective expression of the information quantity:

(1) If the rate of the source does not exceed the capacity of the channel, the information given by the source can be transmitted by the channel with the error probability as small as wanted.

(2) If the rate exceeds the capacity, the error probability when transmitting the information is necessarily strictly positive.

These statements are implicity contained in the proof of the preceding theorem. Let us formulate them exactly:

Theorem A.5. Let $\mu \in M(\mathscr{A})$, let ν be a historyless channel. Then the assumption that

$$V(\mu) < C(\nu)$$

implies the relation

$$\sup_{\zeta \in Z_N} \limsup_n e_n(\mu, \nu, \zeta) = 0 .$$

Proof. The assumption gives

$$V(\bar{\mu}_\zeta) < C(\nu) , \ \zeta \in Z_N ,$$

hence $\limsup_n e_n(\bar{\mu}_\zeta, \nu) = 0$ (cf. [40], Theorem 6.1). But, by the definition of the mapping τ_ζ, $e_n(\bar{\mu}_\zeta, \nu) = e_n(\mu, \nu, \zeta)$. This means that

$$\limsup_n e_n(\mu, \nu, \zeta) = 0 , \quad \zeta \in Z_N ,$$

hence the theorem follows.

Theorem A.6. Let $\mu \in M(\mathscr{A})$, let ν be a historyless channel. If

$$V(\mu) > C(\nu)$$

then there is $\zeta_0 \in Z_N$ and there is a number $c_0 > 0$, respectively, such that

$$\limsup_n e_n(\mu, \nu, \zeta_0) \geqq c_0 .$$

The proof is similar to the proof of the preceding theorem. The analogous statements can be proved also for $\liminf_n e_n(\mu, \nu, \zeta)$.

APPENDIX B: PURE CHARGES ARE ENTROPY DENSE

As we have shown on examples there are pure charges (i.e. purely finitely additive stationary sources) with the entropy rate zero as well as plus infinity (cf. Examples 11.1 and 11.2) The Hewitt-Yosida decomposition tells us that every finitely additive

stationary source μ uniquely decomposes into its σ-additive part μ^c and a pure charge μ^t according to the formula

$$\mu = \alpha\mu^c + (1 - \alpha)\,\mu^t \quad (0 \leqq \alpha \leqq 1)$$

(cf. Section 8). By Remark 12.2, the entropy rate $H(\mu)$ can be then expressed in the form

$$H(\mu) = \alpha\, H(\mu^c) + (1 - \alpha)\, H(\mu^t)\,.$$

There arises a natural problem whether there are pure charges μ satisfying the relation $H(\mu) = h$ for any in advance given real number h, $0 < h < \infty$. Otherwise, the whole theory should reduce to rather trivial cases: the σ-additive case (if $h = 0$), and the case of infinite entropy. It is clear that we can confine ourselves to product pure charges. In this case the problem reduces to a much simpler one. Before stating the corresponding proposition let us introduce some necessary notions. A nonnegative set function μ on $(N, \mathfrak{P}(N))$ is called a *normalized pure charge* if μ is finitely additive, μ assigns unit mass to N, and μ assigns zero mass to each one-point set $\{n\}$, $n \in N$. Let $\zeta \in Z$. Then define

$$h(\mu, \zeta) = -\sum_{C \in \zeta} \mu(C) \log \mu(C)\,.$$

The *entropy* $h(\mu)$ is defined as the supremum

$$h(\mu) = \sup \{h(\mu, \zeta) : \zeta \in Z\}\,.$$

Recall that if $\tilde{\mu}$ is the product pure charge on $(N^I, \mathscr{A}^I)$ determined by μ then $H(\tilde{\mu}) = h(\mu)$.

Proposition B.1. To every finite positive real number h there exists a normalized pure charge μ on $(N, \mathfrak{P}(N))$ satisfying the relation $h(\mu) = h$.

Proof. First let us consider the case $0 < h \leqq \log 2 = 1$. Let $A_1 = \{2n : n \in N\}$; $A_2 = \{2n - 1 : n \in N\}$. Since $A_1 \cap A_2 = \emptyset$, the images under canonical injections $A_1 \to N$, $A_2 \to N$ of the Fréchet filter yield two different, so called elementary filters (cf. N. Bourbaki: Éléments de Mathématique, Livre III, Topologie générale. Hermann, Paris 1961). Let $\mathscr{U}_i$ $(i = 1, 2)$ be the corresponding ultrafilters. The symbol ν_i will denote the normalized pure charge defined by

$$\nu_i(E) = \begin{cases} 1 & \text{if } E \in \mathscr{U}_i\,; \\ 0 & \text{otherwise} \end{cases} \quad (E \subset N;\ i = 1, 2)\,.$$

Given h, $0 < h \leqq 1$, we can choose α, $0 < \alpha < 1$, such that

$$-\alpha \log \alpha - (1 - \alpha) \log (1 - \alpha) = h\,.$$

Now put $\mu = \alpha\nu_1 + (1 - \alpha)\nu_2$. Then μ is again a normalized pure charge. If $\zeta \in Z$, there are a unique set $C \in \zeta$ with $C \in \mathscr{U}_1$ and a unique set $D \in \zeta$ with $D \in \mathscr{U}_2$.

(i) Let $C = D$. Then $\mu(C) = \mu(D) = 1$ and $\mu(E) = 0$ for all $E \in \zeta$, $E \neq C$. Hence $h(\mu, \zeta) = 0$.

(ii) Let $C \neq D$. Then $C \cap D = \emptyset$ and $\mu(C) = \alpha$, $\mu(D) = 1 - \alpha$, $\mu(E) = 0$ for all other $E \in \zeta$. Thus $h(\mu, \zeta) = h$.

The family of partitions fitting (ii) is nonvoid, because e.g. the partition $\{A_1, A_2\}$ is such one. Consequently, $h(\mu) = h$. If $h > 1$ then we can repeat the above reasoning by making use of an appropriate finite number of different elementary filters.

CONCLUSION

Our aim was to show the possible applications of the ergodic theory when solving some problems connected with the source coding. The choice of the problems was strongly affected by the choice of the methods. Therefore we propose to consider related problems the solutions of which do not fall within the frame of the present paper in separate papers.

The third part of the paper represents the main results of the author's dissertation made under the guidance of K. Winkelbauer. The author is gratefully indebted to him for a current support and many invaluable discussions. Other results of the paper were partly obtained during the research done in the period 1971–1975 at the Institute for Measurement and Measuring Technic of the Slovak Akademy of Sciences. The author wishes to express sincere thanks to L. Kubáček for his systematic encouragement. At last but not at least thanks are due to I. Csiszár and to his collaborators from the Mathematical Institute of the Hungarian Academy of Sciences in Budapest for many helpful conversations concerning the subject of the present paper.