KYBERNETIKA — VOLUME 16 (1980), NUMBER 4

On the Error Exponent for Ergodic Markov Source

KAROL VAŠEK*

In this paper the exponential behaviour of the error probability for a finite ergodic Markov source is studied in the case of a fixed-length encoding scheme.

1. INTRODUCTION

Let $X = \{a_1, \ldots, a_r\}$ be a finite set, $\boldsymbol{p} = (p_1, \ldots, p_r)$ a probability distribution on X and $\boldsymbol{P} = \|p_{ij}\|_1^r$ an irreducible stochastic $(r \times r)$ matrix.

Let $(X^N, \mathscr{F}, \mathsf{P})$ be a Markov source with the alphabet X and the probability measure defined on the σ-algebra $\mathscr{F}$ (generated by all finite dimensional cylindres in X^N) by the initial probability distribution $\boldsymbol{p}$ and the stochastic matrix $\boldsymbol{P}$. Since the matrix $\boldsymbol{P}$ is supposed irreducible, the source $(X^N, \mathscr{F}, \mathsf{P})$ is ergodic.

We shall be interested in sequences $\boldsymbol{x}^n = (x_0, \ldots, x_{n-1})$ of length n generated by the source $(X^N, \mathscr{F}, \mathsf{P})$. The probability $\mathsf{P}(\boldsymbol{x}^n)$ of a sequence $\boldsymbol{x}^n$ is given then by:

$$\mathsf{P}(\boldsymbol{x}^n) = p(x_0) \,.\, p(x_1/x_0) \ldots p(x_{n-1}/x_{n-2}) \,, \tag{1.1}$$

where $p(x_0) = p_i$ if $x_0 = a_i$ and $p(x_j/x_i) = p_{ij}$ if $x_j = a_j$ and $x_i = a_i$.

We shall consider now the encoding problem for the source $(X^N, \mathscr{F}, \mathsf{P})$ as follows: Let us suppose that messages of length n are required to be encoded into N_n codewords where

$$\exp(nR) \leqq N_n < \exp(nR) + 1 \tag{1.2}$$

in such a way that the probability of erroneous decoding $P_e(n, R)$ was minimal. It will be achieved evidently if the N_n most probable sequences $\boldsymbol{x}^n$ are encoded into

* This work was done while the autor was at the Institute of Information Theory and Automation of the Czechoslovak Academy of Sciences in 1974.

distinct codewords. The coding of the other sequences can be then performed quite arbitrarily. For simplicity we assign them the same codeword. If we now denote by A_n the set of the most probable sequences $\boldsymbol{x}^n$, then

$$P_e(n, R) = \mathsf{P}(A_n^c) \tag{1.3}$$

(where A_n^c is the complement of the set A_n).

We shall be interested in limiting properties of the probability $P_e(n, R)$, for $n \to \infty$.

It is well known that the important characteristic of the source as regards behaviour of $P_e(n, R)$ is its entropy. In case of the Markov source $(X^N, \mathscr{F}, \mathsf{P})$ it can be shown (see [5]) that its entropy H is given by

$$H = -\sum_i \sum_j \hat{p}_i p_{ij} \ln p_{ij}\,, \tag{1.4}$$

where $\hat{\boldsymbol{p}} = (\hat{p}_1, \ldots, \hat{p}_r)$ is the stationary probability distribution of the stochastic matrix $\boldsymbol{P}$.

In general, the limit of the probability $P_e(n, R)$ when $n \to \infty$ tends to zero for ergodic sources. This can be deduced from Shannon-McMillan theorem (cf. [5]). The following assertion is true:

If the encoding rate R is greater than the entropy H then

$$\lim_{n \to \infty} P_e(n, R) = 0\,. \tag{1.5}$$

However for discrete memoryless sources it has been shown rather more. Jelinek [1] and also Csizár and Longo [2] have proved that the convergence in (1.5) is expotential and they have found the form of this exponent. The exponential convergence of $P_e(n, R)$ has been proved by Longo [3] also for finite ergodic Markov sources, but without being able to obtain the expression for the exponent. To give this expression will be our aim in the following sections of this paper.

2. THE AUXILIARY RESULTS

The purpose of this section is to introduce an auxiliary Markov source.

Consider, along with original stochastic matrix $\boldsymbol{P} = \|p_{ij}\|_1^r$ an matrix $\boldsymbol{P}(\alpha)$ depending on a real parameter α and defined as follows

$$\boldsymbol{P}(\alpha) = \begin{Vmatrix} p_{11}^\alpha & \ldots & p_{1r}^\alpha \\ \vdots & & \\ p_{r1}^\alpha & \ldots & p_{rr}^\alpha \end{Vmatrix} \tag{2.1}$$

where $p_{ij}^\alpha = 0$ if $p_{ij} = 0$.

The matrix $P(\alpha)$ is non-negative and irreducible, because the matrix $\boldsymbol{P}$ is supposed irreducible. Therefore we can apply Perron-Frobenius theorem (see [6]) that asserts that the matrix $\boldsymbol{P}(\alpha)$ possesses a positive eigenvalue $\lambda(\alpha)$ which is larger in modulus than any other eigenvalue of $\boldsymbol{P}(\alpha)$ (maximum eigenvalue of the matrix $\boldsymbol{P}(\alpha)$). Moreover, there exists a positive eigenvector $\boldsymbol{v}(\alpha) = (v_1(\alpha), \ldots, v_r(\alpha))$, $v_i(\alpha) > 0$, $i = 1, 2, \ldots, r$ such that

$$(2.2) \qquad \boldsymbol{P}(\alpha)\, \boldsymbol{v}^{\mathrm{T}}(\alpha) = \lambda(\alpha)\, \boldsymbol{v}^{\mathrm{T}}(\alpha)$$

where we denote by $\boldsymbol{v}^{\mathrm{T}}(\alpha)$ the transpose of vector $\boldsymbol{v}(\alpha)$.

The eigenvector $\boldsymbol{v}(\alpha)$ is determined uniquely except a multiplicative factor. To make it unique we can suppose that

$$(2.3) \qquad \sum_i v_i(\alpha) = 1 .$$

At first, we shall be interested in behaviour of $\lambda(\alpha)$ and $\boldsymbol{v}(\alpha)$ as functions of α.

Lemma 2.1.

(i) $\lambda(\alpha)$ and each component of the vector $\boldsymbol{v}(\alpha)$ are continuous and possess continuous first derivatives for $0 \leqq \alpha \leqq 1$,

(ii) $\ln \lambda(\alpha)$ is convex for $0 \leqq \alpha \leqq 1$.

Proof. For the continuity of $(\mathrm{d}/\mathrm{d}\alpha)\, \lambda(\alpha)$ see [4]. Then the continuity of the first derivative of $v_i(\alpha)$, $i = 1, 2, \ldots, r$ can be deduced from (2.2).

Using (2.2) we can now define for each $0 \leqq \alpha \leqq 1$ a stochastic $(r \times r)$ matrix $\boldsymbol{Q}(\alpha) = \|q_{ij}(\alpha)\|_1^r$ in this way:

$$(2.4) \qquad q_{ij}(\alpha) = \frac{p_{ij}^{\alpha}\, v_j(\alpha)}{\lambda(\alpha)\, v_i(\alpha)}$$

for $p_{ij} \neq 0$, and $q_{ij}(\alpha) = 0$ when $p_{ij} = 0$. The matrix $\boldsymbol{Q}(\alpha)$ is evidently also irreducible and $\boldsymbol{Q}(1) = \boldsymbol{P}$.

By means of the matrix $\boldsymbol{Q}(\alpha)$ and the initial probability distribution $\boldsymbol{p} = (p_1, \ldots, p_r)$ we shall further define a Markov source $(X^N, \mathscr{F}, \mathrm{Q}_\alpha)$ with alphabet X.

The entropy $H(\alpha)$ of this source is then given by:

$$(2.5) \qquad H(\alpha) = - \sum_i \sum_j \hat{q}_i(\alpha)\, q_{ij}(\alpha) \ln q_{ij}(\alpha)$$

(cf. (1.4)), where $\hat{\boldsymbol{q}}(\alpha) = (\hat{q}_1(\alpha), \ldots, \hat{q}_r(\alpha))$ is the stationary probability distribution of the stochastic matrix $\boldsymbol{Q}(\alpha)$. Let us consider the matrix $\boldsymbol{A} = \|a_{ij}\|_1^r$, which has arising from the matrix $\boldsymbol{P}$ as follows: $a_{ij} = 1$ when $p_{ij} > 0$ and $a_{ij} = 0$ when $p_{ij} = 0$. Let λ_0 be the maximum eigenvalue of $\boldsymbol{A}$.

Lemma 2.2.

(i) $H(\alpha)$ is continuous function of α for $0 \leqq \alpha \leqq 1$;

(ii) $H(1) = H$;

(iii) $H(0) = \ln \lambda_0$.

Proof. The continuity of each component of the matrix $\boldsymbol{Q}(\alpha)$ follows from Lemma 2.1. To prove the continuity of the components of the stationary distribution $\hat{\boldsymbol{q}}(\alpha)$, it is sufficient to take into account that

$$\boldsymbol{Q}^{\mathrm{T}}(\alpha)\,\hat{\boldsymbol{q}}^{\mathrm{T}}(\alpha) = \hat{\boldsymbol{q}}^{\mathrm{T}}(\alpha)\,. \tag{2.8}$$

Then (i) immediately results from (2.5). Part (ii) is evident, since $\boldsymbol{Q}(1) = \boldsymbol{P}$. From (2.4) it can be seen that $q_{ij}(0) = 0$ if $p_{ij} = 0$ and

$$q_{ij}(0) = \frac{v_j(0)}{\lambda_0 v_i(0)}$$

if $p_{ij} > 0$. Then we can write

$$H(0) = -\sum_i \sum_j \hat{q}_i(0)\,\frac{v_j(0)}{\lambda_0 v_i(0)} \ln \frac{v_j(0)}{\lambda_0 v_i(0)}\,, \tag{2.9}$$

where $\hat{q}_i(0)$, $i = 1, 2, \ldots, r$ are the components of the stationary probability distribution of the matrix $\boldsymbol{Q}(0)$. Now, when we shall use the fact that

$$\sum_i \hat{q}_i(0)\,\frac{v_j(0)}{\lambda_0 v_i(0)} = \hat{q}_j(0) \tag{2.10}$$

holds for each $j = 1, 2, \ldots, r$, then (iii) will follow from (2.9).

The following lemma establishes a relation between $\lambda(\alpha)$ and $H(\alpha)$.

Lemma 2.3.

$$H(\alpha) = \ln \lambda(\alpha) - \frac{\mathrm{d}}{\mathrm{d}\alpha} \ln \lambda(\alpha)\,. \tag{2.11}$$

Proof. Let us consider, for fixed $x_0 \in X$ and $\alpha \in (0, 1)$, the sequence of functions $f_n(\alpha, x_0)$ defined as follows:

$$f_n(\alpha, x_0) = \frac{1}{n} \ln \sum_{\boldsymbol{x}^n} \mathsf{P}^{\alpha}(\boldsymbol{x}^n/x_0)\, v_{x_n}(\alpha)\,, \tag{2.12}$$

where $\boldsymbol{x}^n = (x_1, \ldots, x_n)$, $v_{x_n}(\alpha) = v_i(\alpha)$ if $x_n = a_i$ and

$$\mathsf{P}(\boldsymbol{x}^n/x_0) = p(x_1/x_0) \ldots p(x_n/x_{n-1})\,.$$

Applying (2.2) in (2.12), we obtain

$$f_n(\alpha, x_0) = \ln \lambda(\alpha) + \frac{1}{n}\, v_{x_0}(\alpha)\,. \tag{2.13}$$

Thus

$$\lim_{n\to\infty} f_n(\alpha, x_0) = \ln \lambda(\alpha) . \tag{2.14}$$

From (2.13) we also see that

$$\lim_{n\to\infty} \frac{d}{d\alpha} f_n(\alpha, x_0) = \frac{d}{d\alpha} \ln \lambda(\alpha) . \tag{2.15}$$

On the other hand, computing the first derivative of $f_n(\alpha, x_0)$ in (2.12) we see

$$\frac{d}{d\alpha} f_n(\alpha, x_0) = \frac{1}{\alpha} \ln \lambda(\alpha) - \frac{1}{\alpha} H_n(\alpha, x_0) + o(1) , \tag{2.16}$$

where

$$H_n(\alpha, x_0) = -\frac{1}{n} \sum_{x^n} Q(x^n/x_0) \ln Q(x^n/x_0) \tag{2.17}$$

and where

$$Q(x^n/x_0) = q(x_1/x_0) \dots q(x_n/x_{n-1}) .$$

However, from the definition of the entropy (see [5]) it follows

$$\lim_{n\to\infty} H_n(\alpha, x_0) = H(\alpha) \tag{2.18}$$

for each $x_0 \in X$ and $\alpha \in (0, 1)$. Then (2.16) along with (2.15) and (2.18) imply (2.11).

Lemma 2.4. $H(\alpha)$ is a decreasing function of α for $\alpha \in [0, 1]$.

Proof. From (2.11) we have

$$\frac{d}{d\alpha} H(\alpha) = -\alpha \frac{d^2}{d\alpha^2} \ln \lambda(\alpha) .$$

To prove the lemma, it is sufficient to remind that the function $\ln \lambda(\alpha)$ is convex for $\alpha \in [0, 1]$.

To obtain some insight on the meaning of λ_0, let us consider the set $\{x^n : P(x^n) > 0\}$. Let $N_0(n)$ be the number of the elements of this set. Then the following theorem holds:

Theorem 2.1.

$$\lim_{n\to\infty} \frac{1}{n} \ln N_0(n) = \ln \lambda_0 \tag{2.19}$$

Proof. It can be easily seen that

$$\lim_{\alpha\to 0+} \sum_{x^n} P^\alpha(x^n) = N_0(n) .$$

But in [4] it has been shown that

(2.20) $$\lim_{n\to\infty} \frac{1}{n} \ln \sum_{x^n} \mathsf{P}^\alpha(x^n) = \ln \lambda(\alpha)$$

uniformly for $\alpha \in [0, 1]$. Now if we pass in (2.20) to the limit $\alpha \to 0^+$, (2.19) will be proved.

3. AN UPPER BOUND ON THE ERROR PROBABILITY

Now we go back to the problem of the estimation of the error probability $P_e(n, R)$. At first we shall find an upper bound on the $P_e(n, R)$.

Theorem 3.1. If

(3.1) $$\exp(nR) \leqq N_n < \exp(nR) + 1$$

where $H < R < \ln \lambda_0$, then

(3.2) $$P_e(n, R) \leqq K(\alpha) \exp\left(-n\left[\frac{1-\alpha}{\alpha} R - \frac{1}{\alpha} \ln \lambda(\alpha)\right]\right),$$

for each $n > 1$ and $0 < \alpha < 1$.

Proof. Using the initial probability distribution $\boldsymbol{p} = (p_1, \ldots, p_r)$ of the source $(X^N, \mathscr{F}, \mathsf{P})$ we can define for each real α a probability distribution $\boldsymbol{q}(\alpha) = (q_1(\alpha), \ldots \ldots, q_r(\alpha))$ as follows:

(3.3) $$q_i(\alpha) = \frac{p_i^\alpha}{\sum_i p_i^\alpha}$$

if $p_i \neq 0$ and $q_i(\alpha) = 0$ if $p_i = 0$.

Now let us consider an auxiliary function $f_\alpha(x^n)$ defined for each $x^n = (x_0, \ldots, x_{n-1}) \in X^n$ and each positive α by:

(3.4) $$f_\alpha(x^n) = q_\alpha(x_0) \frac{p(x_1/x_0)^\alpha \ldots p(x_{n-1}/x_{n-2})^\alpha \cdot v_m(\alpha)}{\lambda^{n-1}(\alpha)\, v_M(\alpha)},$$

where $v_M(\alpha) = \max_i (v_i(\alpha))$, $v_m(\alpha) = \min_i (v_i(\alpha))$. The function $f_\alpha(x^n)$ is non-negative. Moreover using (2.4) we can see that

(3.5) $$\sum_{x^n} f_\alpha(x^n) \leqq 1 .$$

Going further, let

(3.6) $$A_{n,\alpha} = \{x^n \in X^n : f_\alpha(x^n) > \exp(-nR)\}$$

and let $|A_{n,\alpha}|$ be the number of the elements of the set $A_{n,\alpha}$. Then from (3.1), (3.5) and (3.6) it follows easily that

$$|A_{n,\alpha}| < \exp(nR) \leqq N_n . \tag{3.7}$$

Now, if we regard to the definitions of the sets A_n, $A_{n,\alpha}$ and if we use that the function $f_\alpha(x^n)$ is strictly increasing with the probability $\mathsf{P}(x^n)$, we immediately obtain from inequality (3.7) that $A_{n,\alpha} \subset A_n$ or $A_n^c \subset A_{n,\alpha}^c$. However, we know that $P_e(n, R) = \mathsf{P}(A_n^c)$. Therefore we can write

$$P_e(n, R) \leqq \mathsf{P}(A_{n,\alpha}^c) . \tag{3.8}$$

To obtain an upper bound of the $P_e(n, R)$ we shall try to bound the probability $\mathsf{P}(A_{n,\alpha}^c)$.

It holds

$$\mathsf{P}(A_{n,\alpha}^c) = \sum_{A_{n,\alpha}^c} \mathsf{P}(x^n) = \sum_{A_{n,\alpha}^c} p(x_0)\, p(x_1/x_0) \ldots p(x_{n-1}/x_{n-2}) . \tag{3.9}$$

Using (2.4) we get from (3.9) that

$$\mathsf{P}(A_{n,\alpha}^c) = \sum_{A_{n,\alpha}^c} \mathsf{Q}_\alpha(x^n) \frac{\lambda^{n-1}(\alpha)\, v(x_0)}{p(x_1/x_0)^{\alpha-1} \ldots p(x_{n-1}/x_{n-2})^{\alpha-1}\, v(x_{n-1})} . \tag{3.10}$$

But for each $x^n \in A_{n,\alpha}^c$ we have

$$\ln f_\alpha(x^n) \leqq -nR . \tag{3.11}$$

This inequality then implies

$$\ln \frac{v_m(\alpha)}{v_M(\alpha)} \lambda(\alpha) + \alpha \ln p(x_1/x_0) \ldots p(x_{n-1}/x_{n-2}) + \\ + \ln q_\alpha(x_0) - n \ln \lambda(\alpha) \leqq -nR . \tag{3.12}$$

Since we suppose that $0 < \alpha < 1$, from (3.12) after an elementary calculation we obtain that

$$\ln \frac{v_\alpha(x_0)\, \lambda^{-1}(\alpha)}{p(x_1/x_0)^{\alpha-1} \ldots p(x_{n-1}/x_{n-2})^{\alpha-1}\, v_\alpha(x_{n-1})} \leqq \\ \leqq -n \frac{1-\alpha}{\alpha} [R - \ln \lambda(\alpha)] - \frac{1-\alpha}{\alpha} \ln q_\alpha(x_0) - \frac{1}{\alpha} \ln \frac{v_m(\alpha)}{v_M(\alpha)} \lambda(\alpha) . \tag{3.13}$$

By means of (3.13) then from (3.10) we get the following inequality

$$\mathsf{P}(A_{n,\alpha}^c) \leqq \exp\left(-n\left[\frac{1-\alpha}{\alpha} R - \frac{1}{\alpha} \ln \lambda(\alpha)\right] - \frac{1}{\alpha} \ln \frac{v_m(\alpha)}{v_M(\alpha)} \lambda(\alpha)\right) . \tag{3.14}$$

$$\cdot \sum_{A_{n,\alpha}^c} \exp\left(\frac{1-\alpha}{\alpha} \ln q_\alpha(x_0)\right) . \mathsf{Q}_\alpha(x^n) .$$

At last it can be easily shown that

$$\sum_{A_{n,\alpha}^c} \exp\left(-\frac{1-\alpha}{\alpha} \ln q_\alpha(x_0)\right) \mathsf{Q}_\alpha(x^n) \leqq r^{1/\alpha} . \tag{3.15}$$

Now, when we put

$$K(\alpha) = \exp\left(-\frac{1}{\alpha} \ln \frac{\lambda(\alpha)\, v_m(\alpha)}{r\, v_m(\alpha)}\right), \tag{3.16}$$

the inequality (3.2) follows from (3.8), (3.14) and (3.15). The upper estimation (3.2) of the probability of error $P_e(n, R)$ is depended on a parameter α. In the following section we shall show that the parameter α can be chosen so that the inequality will yield the asymptotic optimal estimation of the probability $P_e(n, R)$.

4. THE LIMITING RATE OF CONVERGENCE FOR $P_e(n, R)$

Our following considerations will be based on the ergodicity of the auxiliary Markov source $(X^N, \mathscr{F}, \mathsf{Q}_\alpha)$.

Let A_n, $n = 1, 2, \ldots$ be a sequence of non-empty sets such that $A_n \subset \{x^n : \mathsf{P}(x^n) > 0\}$ for each n. Let further $|A_n|$ be the number of the elements of set A_n.

Lemma 4.1. If

$$\lim_{n\to\infty} \frac{1}{n} \ln |A_n| = R , \tag{4.1}$$

where $H < R < \ln \lambda_0$, then for every $\varepsilon > 0$ there exist $\alpha \in (0, 1)$ and an integer $n(\varepsilon, \alpha)$ such that for $n > n(\varepsilon, \alpha)$ it holds

$$\mathsf{P}(A_n^c) \geqq K^*(\alpha) \exp\left(-n\left[\frac{1-\alpha}{\alpha} R - \frac{1}{\alpha} \ln \lambda(\alpha) + \varepsilon\right]\right), \tag{4.2}$$

with

$$K^*(\alpha) = \tfrac{1}{4} \exp\left(\frac{1}{\alpha} \ln \frac{v_m(\alpha)}{\lambda(\alpha)\, v_M(\alpha)}\right).$$

Proof. Let R be such that $H < R < \ln \lambda_0$. Let further ε_1 be an arbitrary positive number satisfying the condition

$$R + 2\varepsilon_1 < \frac{R + \ln \lambda_0}{2} . \tag{4.3}$$

Then from Lemma 2.2 we can deduce that there exists $\alpha_1 \in (0, 1)$ such that

$$R + 2\varepsilon_1 = H(\alpha_1)\,. \tag{4.4}$$

For α_1 chosen in this manner, let us consider the Markov source $(X^N, \mathscr{F}, \mathsf{Q}_{\alpha_1})$ which is defined by the stochastic matrix $\boldsymbol{Q}(\alpha_1) = \|q_{ij}(\alpha_1)\|_1^r$ and the initial probability distribution $\boldsymbol{p} = (p_1, \ldots, p_r)$. Since the source $(X^N, \mathscr{F}, \mathsf{Q}_{\alpha_1})$ is ergodic, by Shannon-McMillan theorem it can be easily shown that

$$\lim_{n\to\infty} \frac{1}{n} \ln q_{\alpha_1}(x_1/x_0) \ldots q_{\alpha_1}(x_{n-1}/x_{n-2}) = -H(\alpha_1) \tag{4.5}$$

almost everywhere with respect to Q_{α_1}.

Let

$$B_n = \left\{x^n : \left| -\frac{1}{n} \ln q_{\alpha_1}(x_1/x_0) \ldots q_{\alpha_1}(x_{n-1}/x_{n-2}) - H(\alpha_1)\right| < \varepsilon_1\right\}. \tag{4.6}$$

Then there exists an integer $n_1(\varepsilon_1, \alpha_1)$ such that for $n > n_1(\varepsilon_1, \alpha_1)$ we have

$$\mathsf{Q}_{\alpha_1}(B_n) \geqq \tfrac{1}{2}\,. \tag{4.7}$$

On the other hand from (4.1) it follows that there exists an integer $n_2(\varepsilon_1)$ such that for $n > n_2(\varepsilon_1)$ it holds

$$|A_n| \leqq \tfrac{1}{4} \exp\left(n[R + \varepsilon_1]\right). \tag{4.8}$$

However we can write also

$$|A_n| = \sum_{A_n} \mathsf{Q}_{\alpha_1}(x^n)\,[\mathsf{Q}_{\alpha_1}(x^n)]^{-1} \tag{4.9}$$

where

$$\mathsf{Q}_{\alpha_1}(x^n) = p(x_0)\, q_{\alpha_1}(x_1/x_0) \ldots q_{\alpha_1}(x_{n-1}/x_{m-2})\,.$$

From (4.6) we see that

$$\frac{1}{n} \ln\,[\mathsf{Q}_{\alpha_1}(x^n)]^{-1} > H(\alpha_1) - \varepsilon_1 \tag{4.10}$$

for each $x^n \in B_n$, so that (4.9) immediately yields

$$|A_n| > \exp\left(n[H(\alpha_1) - \varepsilon_1]\right) . \sum_{A_n \cap B_n} \mathsf{Q}_{\alpha_1}(x^n)\,. \tag{4.11}$$

(4.4) and the inequalities (4.8), (4.11) imply

$$\sum_{A_n \cap B_n} \mathsf{Q}_{\alpha_1}(x^n) < \tfrac{1}{4}\,. \tag{4.12}$$

From (4.12) in view of (4.7) we then have

$$\sum_{A_n^c \cap B_n} \mathsf{Q}_{\alpha_1}(\boldsymbol{x}^n) \geqq \tfrac{1}{4}. \tag{4.13}$$

The probability $\mathsf{P}(A_n^c)$ can be bounded for $n > \max(n_1(\varepsilon_1, \alpha_1), n_2(\varepsilon_1))$ in the following way

$$\mathsf{P}(A_n^c) \geqq \sum_{A_n^c \cap B_n} p(x_0) \,.\, p(x_1/x_0) \ldots p(x_{n-1}/x_{n-2}) = \tag{4.14}$$

$$= \sum_{A_n^c \cap B_n} \mathsf{Q}_{\alpha_1}(\boldsymbol{x}^n) \frac{\lambda^{n-1}(\alpha_1)\, v_{\alpha_1}(x_0)}{p(x_1/x_0)^{\alpha_1 - 1} \ldots p(x_{n-1}/x_{n-2})^{\alpha_1 - 1}\, v_{\alpha_1}(x_{n-2})}.$$

From the definition of the set B_n it is seen that

$$\frac{1}{n} \ln \frac{1}{q_{\alpha_1}(x_1/x_0) \ldots q_{\alpha_1}(x_{n-1}/x_{n-2})} < H(\alpha_1) + \varepsilon_1 \tag{4.15}$$

for each $\boldsymbol{x}^n \in A_n^c \cap B_n$. Since $0 < \alpha_1 < 1$, the following inequality can easily be obtained from (4.15)

$$\frac{1}{n} \ln \frac{\lambda^{-1}(\alpha_1)\, v_{\alpha_1}(x_0)}{p(x_1/x_0)^{\alpha_1 - 1} \ldots p(x_{n-1}/x_{n-2})^{\alpha_1 - 1}\, v_{\alpha_1}(x_{n-1})} > \tag{4.16}$$

$$> \frac{1 - \alpha_1}{\alpha_1} \left(H(\alpha_1) + \varepsilon - \ln \lambda(\alpha_1)\right) + \frac{1}{\alpha_1} \,.\, \frac{1}{n} \ln \frac{v_m(\alpha_1)}{\lambda(\alpha_1)\, v_M(\alpha_1)}.$$

Then (4.16) along with (4.13) and (4.14) provide that

$$\mathsf{P}(A_n^c) > K^*(\alpha_1) \exp\left(-n\left[\frac{1 - \alpha_1}{\alpha_1} H(\alpha_1) - \frac{1}{\alpha_1} \ln \lambda(\alpha_1) + \frac{1 - \alpha_1}{\alpha_1} \varepsilon_1\right]\right) \tag{4.17}$$

where

$$K^*(\alpha_1) = \tfrac{1}{4} \exp\left(\frac{1}{\alpha_1} \ln \frac{v_m(\alpha_1)}{\lambda(\alpha_1)\, v_M(\alpha_1)}\right). \tag{4.18}$$

Let us take now α_0 such that

$$\frac{K + \ln \lambda_0}{2} = H(\alpha_0). \tag{4.19}$$

Since the function $H(\alpha)$ is decreasing for $0 < \alpha < 1$, clearly $\alpha_0 < \alpha_1$ or

$$\frac{1 - \alpha_1}{\alpha_1} < \frac{1 - \alpha_0}{\alpha_0}. \tag{4.20}$$

 Using (4.4) and (4.20) we finally obtain that

$$(4.21) \quad \mathsf{P}(A_n^c) > K^*(\alpha_1) \exp\left(-n\left[\frac{1-\alpha_1}{\alpha_1} R - \frac{1}{\alpha_1} \ln \lambda(\alpha_1) + 3\frac{1-\alpha_0}{\alpha_0} \varepsilon_1\right]\right)$$

for all $n > \max(n_1(\varepsilon_1, \alpha_1), n_2(\varepsilon_1))$. If we take for an arbitrary $\varepsilon > 0$ such ε_1 in (4.3) that it holds $3(1 - \alpha_0)\,\varepsilon_1 < \alpha_0 \varepsilon$, lemma will be proved.

Now we can state the following result

Theorem 4.1. If

$$\exp(nR) \leqq N_n < \exp(nR) + 1$$

where $H < R < \ln \lambda_0$, then

$$(4.22) \qquad \lim_{n\to\infty} \frac{1}{n} \ln P_e(n, R) = \frac{1}{\alpha^*} \ln \lambda(\alpha^*) - \frac{1 - \alpha^*}{\alpha^*} R$$

where α^* satisfies the equation

$$(4.23) \qquad H(\alpha^*) = R\,.$$

Proof. From Theorem 3.1 and Lemma 4.1 we can deduce that

$$(4.24) \qquad \lim_{n\to\infty} \frac{1}{n} \ln P_e(n, R) = -\max_{0<\alpha\leqq 1} \left(\frac{1-\alpha}{\alpha} R - \frac{1}{\alpha} \ln \lambda(\alpha)\right).$$

Using further (2.11) we obtain that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\left(\frac{1-\alpha}{\alpha} R - \frac{1}{\alpha} \ln \lambda(\alpha)\right) = -\frac{1}{\alpha^2}(R - H(\alpha))\,.$$

If now take into account the convexity of the function $\ln \lambda(\alpha)$, the proof is complete.

In the end we note some properties of the exponent

$$E(R) = \max_{0<\alpha\leqq 1} \left(\frac{1-\alpha}{\alpha} R - \frac{1}{\alpha} \ln \lambda(\alpha)\right).$$

It can be easily shown that $E(R)$ is increasing and convex for $H \leqq R \leqq \ln \lambda_0$. In addition to this, $E(H) = 0$ and

$$E(\ln \lambda_0) = -\left[\frac{\mathrm{d}}{\mathrm{d}\alpha} \ln \lambda(\alpha) + \ln \lambda(\alpha)\right]_{\alpha=0}.$$

(Received October 15, 1979.)

[1] F. Jelínek: Probabilistic Information Theory. Mc Graw-Hill, New York 1968.

[2] I. Csizár, G. Longo: On the Exponent for Source Coding etc. Studia Sc. Math. Hung. *6* (1971), 181—191.

[3] G. Longo: On the error exponent for Markov sources. Transactions of 6-th Prague Conference on Information Theory ... Praha 1971, Academia Praha 1973.

[4] L. Koopmans: Asymptotic Rate of Discrimination for Markov Processes. Ann. Math. Stat. *31* (1960), 982—994.

[5] P. Billingsley: Ergodic Theory and Information. John Wiley and Sons., Inc., New York 1965.

[6] M. Marcus, H. Minc: A Survey of Matrix Theory and Matrix Inequalities. Allyn and Bacon, Inc., Boston, 1964.

Ing. Karol Vašek, CSc., Výskumný ústav dopravný (Institute of Transportation Research), Veľký Diel, 011 80 Žilina. Czechoslovakia.