8. L. Kanal, "Patterns in pattern recognition, 1968-1974," IEEE Trans. Inform. Theory, 20, No. 6, 697-722 (1974).
9. H. Cramer, Mathematical Methods of Statistics, Princeton University Press (1946).
10. G. H. Hardy, J. E. Littlewood, and G. Polya. Inequalities, 2nd ed., Cambridge University Press (1952).
11. A. N. Kolmogorov and S. V. Fomin, Elements of the Theory of Functions and Functional Analysis [in Russian], Nauka, Moscow (1972).

DATA TRANSMISSION OVER A DISCRETE CHANNEL WITH FEEDBACK. RANDOM TRANSMISSION TIME

M. V. Burnashev

UDC 621.391.16:519.2

Asymptotically optimal limits are obtained for the mean transmission time for specified error probability and number of possible messages. As a corollary of these results, the author gives an optimum reliability function for the channel in question (with noiseless feedback and a Markov instant of decision-making).

1. INTRODUCTION

There have been many studies of the optimum reliability function of a stationary memoryless channel (see, e.g., [1, 2]). If, however, the transmission time is fixed, it is not possible to compute this function even when there is noiseless feedback. For example, in the elementary case of a binary symmetrical channel with noiseless feedback, the author of [2] was able to obtain only certain upper and lower bounds for this function. If the transmission time is random, i.e., the receiver himself determines (on the basis of the signals he receives) whether to make a decision or to continue transmission, it turns out to be possible to solve this problem in the general case of a stationary discrete memoryless channel with noiseless feedback. In this paper we first obtain a lower limit for the mean decision-making time for any method of transmission with feedback (not necessarily noiseless). Then we set up a transmission method using noiseless feedback for which the mean transmission time coincides asymptotically with the limit in question for an error probability $P_e \to 0$. In particular, this makes it possible to compute the optimum reliability function of a discrete channel with noiseless feedback.

2. STATEMENT OF THE PROBLEM AND FORMULATION OF MAIN RESULTS

To simplify our subsequent exposition, we will briefly describe the overall process of message transmission over a stationary discrete memoryless channel involving feedback; it will be convenient to divide our description into three sections: 1) communications channel; 2) method of transmission with feedback (encoding); 3) method of decision-making (decoding).

A more detailed description of the first two sections can be found in [3, 4].

1. Communications Channel

A stationary discrete memoryless channel (DMLC) with a finite number of input symbols $E_1, \ldots, E_K$ and output symbols $\bar{E}_1, \ldots, \bar{E}_L$ is defined by the transition probability matrix $\mathscr{P} = \{p_{ij}\}$, $i = 1, \ldots, K$, $j = 1, \ldots, L$, $p_{ij} \geq 0$, $\sum_{j=1}^{L} p_{ij} = 1$, where p_{ij} specifies the conditional probability that input-alphabet symbol E_i will be converted to output-alphabet symbol $\bar{E}_j$ as a result of transmission. We will assume that matrix $\mathscr{P}$ contains

Translated from Problemy Peredachi Informatsii, Vol. 12, No. 4, pp. 10-30. October-December, 1976. Original article submitted March 7, 1975.

no columns that consist entirely of zeros and that all the rows of the matrix differ. The meaning of these assumptions is obvious, and we will not dwell on them.

2. Method of Transmission with Feedback

Assume that $\Theta_1, \ldots, \Theta_N$ is the set of possible messages at the transmitting end of the channel (N is the overall number of messages). We denote the signal at the DMLC input at the i-th instant by ξ_i and the signal at the DMLC output. by ξ_i'. Signals ξ_i and ξ_i' assume values from the alphabets $\{E_j\}$ and $\{\bar{E}_j\}$. Similarly, we denote the signal transmitted via the feedback channel at the i-th signal (after reception of ξ_i') by η_i' and the signal received at the transmitting end of the DMLC, by η_i. We will assume for simplicity that η_i and η_i' assume values from some large but finite alphabet Y^1. We will also use the notation

$$(\xi_1, \ldots, \xi_k) = \bar{\xi}^k \in \mathscr{E}^k, \quad (\xi_1', \ldots, \xi_k') = \bar{\xi}'^k \in \bar{\mathscr{E}}^k,$$
$$(\eta_0, \eta_1, \ldots, \eta_k) = \bar{\eta}^k \in Y^{k+1}, \quad (\eta_0', \eta_1', \ldots, \eta_k') = \bar{\eta}'^k \in \bar{Y}^{k+1}.$$

The feedback channel is defined by a system of conditional probabilities $P(\eta_i \mid \bar{\eta}^{i-1}, \bar{\eta}'^i)$ for all possible η_i, $\bar{\eta}^{i-1}$, $\bar{\eta}'^i$. The main part of our subsequent exposition will relate to noiseless feedback channels. In this case the system of conditional probabilities becomes particularly simple:

$$P(\eta_i \mid \bar{\eta}^{i-1}, \bar{\eta}'^i) = \begin{cases} 0, & \text{if} \quad \eta_i \neq \eta_i', \\ 1, & \text{if} \quad \eta_i = \eta_i'. \end{cases}$$

To specify a transmission method, we must specify the laws that the transmitter and receiver operate under. In the most general case, we should specify at the transmitting end a set of conditional probabilities $P(\xi_i = E_k \mid \Theta_j, \bar{\xi}^{i-1}, \bar{\eta}^{i-1})$ for all possible E_k, Θ_j, $\bar{\xi}^{i-1}$, and $\bar{\eta}^{i-1}$, defining the signal transmitted at the i-th instant, $i = 1, 2, \ldots$. Similarly, at the receiving end we should specify a set of conditional probabilities $P(\eta_i' \mid \bar{\xi}'^i, \bar{\eta}'^{i-1})$ for all possible $\bar{\xi}'^i$, η_i', and $\bar{\eta}'^{i-1}$, that defines the mode of utilization of the feedback channel.

The fact that the feedback channel has no memory means that the condition ($\bar{\xi}^n = (E_{i_1}, \ldots, E_{i_n})$, $\bar{\xi}'^n = (\bar{E}_{j_1}, \ldots, \bar{E}_{j_n})$): $P(\bar{\xi}'^n \mid \bar{\xi}^n, \Theta_k, \bar{\eta}'^{n-1}, \bar{\eta}^{n-1}) = P(\bar{\xi}'^n \mid \bar{\xi}^n) = p_{i_1 j_1} \cdot \ldots \cdot p_{i_n j_n}$ is valid for all possible $\bar{\xi}^n$, $\bar{\xi}'^n$, Θ_k, $\bar{\eta}^{n-1}$, and $\bar{\eta}'^{n-1}$.

3. Method of Decision-Making

The general decision-making procedure can be described as follows. After receiving a signal ξ_1', the receiver determines himself, on the basis of the information available to him (i.e., the signals ξ_1' and η_0'), whether to make a decision regarding the transmitted message or to continue transmission. The choice of decision can be conveniently described by using the random variable ζ_1 that assumes one of the N + 1 values 1, ..., N + 1. For this we specify at the receiving end of the DMLC a set of conditional probabilities $P(\zeta_1 = m \mid \xi_1', \eta_0')$ for $m = 1, \ldots, N + 1$ and all possible ξ_1', η_0'. If ζ_1 assumes the value m ($m \leq N$), then a decision is made in favor of the m-th message and transmission ceases. If ζ_1 assumes the value N + 1, then transmission continues and for decision-making we employ the random variable ζ_2 whose distribution depends on $\bar{\xi}'^2$ and $\bar{\eta}'^1$, etc. Thus, assume that $\zeta_1, \zeta_2, \ldots$ is the sequence of random variables at the receiving end of the DMLC. Each of the ζ_i assumes one of N + 1 values 1, ..., N + 1. At the receiving end of the DMLC we specify a set of conditional probabilities $P(\zeta_k = m \mid \bar{\xi}'^k, \bar{\eta}'^{k-1})$ for $m = 1, \ldots, N + 1$, $k = 1, 2, \ldots$ and all possible $\bar{\xi}'^k$, $\bar{\eta}'^{k-1}$. If ζ_k assumes the value m ($m \leq N$) and all the preceding ζ_i ($i < k$) were equal to N + 1, then a decision is made in favor of the m-th message. If ζ_k assumes the value N + 1 and all preceding ζ_i values were equal to N + 1, then transmission continues. In other words, after the next received symbol ξ_k' in sequence we randomly choose the value of ζ_k at the receiving end of the DMLC and, depending on the ζ_k value chosen, we make a decision regarding the transmitted message or continue the transmission. The distribution of ζ_k depends only on the received signals $\bar{\xi}'^k$ and signals $\bar{\eta}'^{k-1}$ transmitted over the feedback channel.

Now, after we have described the transmission method and decision-making procedure, let us determine the corresponding error probability P_e. Let $\bar{\xi}^k \in \mathscr{E}^k$, $\bar{\xi}'^k \in \bar{\mathscr{E}}^k$, $\bar{\eta}^k \in Y^{k+1}$, $\bar{\eta}'^k \in \bar{Y}^{k+1}$, $\bar{\zeta}^k = (\zeta_1, \ldots, \zeta_k) \in Z^k$. We denote by $\mathscr{E}$ the ensemble of all possible infinite sequences of values of ξ. Similarly, we introduce the sets $\bar{\mathscr{E}}$, Y, $\bar{Y}$, and Z. Consider the product of spaces $\bar{\mathscr{E}}^n \times \bar{Y}^n \times Z^n = X^n$ and $\bar{\mathscr{E}} \times \bar{Y} \times Z = X$. We will denote points of space X by $x = (\bar{\xi}', \bar{\eta}', \bar{\zeta})$. It is clear that space X^n corresponds to observations at the receiving end of the DMLC over time n. Space X can be naturally partitioned into sets $X_1, \ldots, X_{N+1}$. For this we assign to set X_{N+1} all points x for which $\bar{\zeta} = (N + 1, N + 1, \ldots)$, i.e., $X_{N+1} = \{x + (\bar{\xi}', \bar{\eta}', \bar{\zeta}): \bar{\zeta} = (N + 1, N + 1, \ldots)\}$. Now we assign to set X_m ($1 \leq m \leq N$) those points x for which the first coordinate ζ that is different from N + 1 assumes the value m, i.e., $X_m = \{x = (\bar{\xi}', \bar{\eta}', \bar{\zeta}): \bar{\zeta} = (N+1, \ldots, N+1, m, \ldots)\}$. Obviously, $\bigcup_{i=1}^{N+1} X_i = X$ and $X_i \cap X_j = 0$ for $i \neq j$. Clearly, set X_m ($m \leq N$) corresponds to a decision in favor of the m-th message; this occurs when the coordinate ζ_i, $i = 1, 2, \ldots$, first assumes the value m. The set X_{N+1} corresponds to transmission being continued indefinitely. If a transmission method and decision-making procedure have been chosen, then for every message Θ_m we have defined the probabilities $P(X_j | \Theta_m)$, $j = 1, \ldots, N+1$, $m = 1, \ldots, N$. Let us define the error probability P_e of the transmission method by the expression

$$P_e = \frac{1}{N} \sum_{i=1}^{N} P(e | \Theta_i), \tag{2.1}$$

where

$$P(e|\Theta_i) = P(\bar{X}_i | \Theta_i) = 1 - P(X_i | \Theta_i), \quad i = 1, \ldots, N.$$

We will say that Markov instant τ determines the decision-making rule if the event $\{\tau = n\}$ is equivalent to the event $\{\zeta_n \neq N+1;\ \zeta_k = N+1,\ k = 1, \ldots, n-1\}$. We define the mean transmission time $\bar{\tau}$ by the expression

$$\bar{\tau} = \frac{1}{N} \sum_{i=1}^{N} M(\tau | \Theta_i). \tag{2.2}$$

Now we can formulate the following problems:

1) for specified number of messages N and error probability P_e, we are to construct a lower limit for the mean decision-making time for any transmission method over a DMLC employing feedback;

2) for any specified $\varepsilon > 0$ we are to set up a transmission method over a DMLC with noiseless feedback for which the error probability P_e does not exceed ε, while the mean decision-making time is asymptotically optimal as $P_e / N \to 0$.

We note that the presence of an a priori distribution for messages is not of importance to us in the statement of the problem and formulation of the main results. For convenience, however, we will assume in setting up a transmission method that messages Θ_i, $i = 1, \ldots, N$, have an equiprobable a priori distribution. It is easy to show that all the proofs remain in force if we introduce a fictitious uniform "a priori" distribution instead of the real a priori distribution.

As will be seen in what follows, the case in which matrix $\mathscr{P}$ has no zero elements is of greatest interest. Therefore, the major part of this paper (Secs. 3-5) will be devoted to this case. In Sec. 3 we obtain auxiliary results that are necessary in setting up the lower limit for the mean transmission time. Some of these results are of independent interest. In particular, we have evidently been the first to obtain a lemma that describes the change in the entropy of the a posteriori message distribution over one observation at sufficiently small entropy values (Lemma 3). In Sec. 4 we obtain the following lower limit for the mean decision-making time $\bar{\tau}$ that any transmission method over a DMLC with feedback must satisfy:

$$\bar{\tau} > \frac{\ln N}{C} - \frac{\ln P_e}{C_1} - \frac{\ln(\ln N - \ln P_e)}{C_1} - \frac{P_e \cdot \ln N}{C} + D(\mathscr{P}), \tag{2.3}$$

where C is the capacity of the DMLC, the constant C_1 is defined in (3.10) and $D(\mathscr{P})$ is bounded and depends only on $\mathscr{P}$.

In Sec. 5 we set up a transmission method involving noiseless feedback for which the mean decision-making time is asymptotically equivalent to the right side of (2.3) as $P_e \to 0$. We note that the transmission method constructed in Sec. 5 generalizes the transmission methods of [5, 6].

The results obtained in Secs. 4-5 permit us to compute the optimum reliability function $E(\bar{R})$ for a DMLC with noiseless feedback:

$$E(\bar{R}) = C_1(1 - \bar{R}/C), \quad 0 \leqslant \bar{R} \leqslant C. \tag{2.4}$$

In Sec. 6 we briefly consider the "degenerate" case in which matrix $\mathscr{P}$ contains zero elements. In this case the lower limit has the form

$$\bar{\tau} > C^{-1}[\ln N - P_e \cdot \ln(N-1)] + D(\mathscr{P}). \tag{2.5}$$

The mean decision-making time for the transmission method with zero error, set up in Sec. 6. is asymptotically equivalent to the right side of (2.5) as $N \to \infty$ and $P_e \to 0$.

Finally, the Appendix gives the proofs of some lemmas formulated in Secs. 3-4.

3. SOME AUXILIARY RESULTS

In this section we will prove a number of facts that will subsequently be required. After each observation we can compute the a posteriori probabilities $p_i(n)$ of messages Θ_i ($i = 1, \ldots, N$; n is the number of observations) and the corresponding entropy of the a posteriori distribution:

$$H_n = H(\bar{p}(n)) = -\sum_{i=1}^{N} p_i(n) \ln p_i(n),$$
$$\bar{p}(n) = (p_1(n), \ldots, p_N(n)). \tag{3.1}$$

Let τ be the Markov stopping instant (that determines the decision-making rule), where $P(\tau < \infty) = 1$, and let H_τ be the entropy of the a posteriori distribution at instant τ. Then for the error probability in decision-making P_e we have the following lemma.

LEMMA 1.

$$MH_\tau \leqslant h(P_e) + P_e \cdot \ln(N-1), \tag{3.2}$$

where

$$h(x) = -x \ln x - (1-x) \ln(1-x). \tag{3.3}$$

This lemma is an extension of a similar result proved in [1] (Theorem 4.3.1) for a fixed decision-making instant.

Proof. We will establish that the condition $P(\tau < \infty) = 1$ implies that the following limit exists:

$$\lim_{n \to \infty} MH_{n \wedge \tau} = MH_\tau. \tag{3.4}$$

Indeed,

$$MH_{n\wedge\tau} = \sum_{i=1}^{n} \mathrm{M}(H_i \mid \tau = i) P(\tau = i) + \mathrm{M}(H_n \mid \tau > n) P(\tau > n).$$

Using the fact that H_n is bounded, we have ($m < n$)

$$|MH_{n\wedge\tau} - MH_{m\wedge\tau}| \leqslant \mathrm{M}(H_n \mid \tau > n) P(\tau > n) + \sum_{i=m+1}^{n} \mathrm{M}(H_i \mid \tau = i) P(\tau = i) + \mathrm{M}(H_m \mid \tau > m) P(\tau > m)$$
$$\leqslant \ln N\left[P(\tau > n) + \sum_{i=m+1}^{n} P(\tau = i) + P(\tau > m)\right] = 2P(\tau > m)\ln N \to 0 \quad \text{as} \quad m \to \infty,$$

which yields that $\lim_{n\to\infty} MH_{n\wedge\tau}$ exists. The following inequality is essentially proved in [1]:

$$\sum_{j=1}^{N} H[\bar{p}(\Theta|X_j, \tau=n)] \cdot P(X_j|\tau=n) \leqslant h[P(e|\tau=n)] + P(e|\tau=n) \cdot \ln(N-1). \tag{3.5}$$

[Here $H(\bar{p}(\Theta))$ denotes $-\sum_{i=1}^{N} p(\Theta_i) \ln p(\Theta_i)$]. On the other hand, it is obvious that $H(\bar{p})$ is strictly $\cap$-convex.* Therefore, using the Jensen inequality, we have

$$H[\bar{p}(\Theta|X_j, \tau=n)] = H[\mathrm{M}(\bar{p}(\Theta|x)|X_j, \tau=n)] \geqslant \mathrm{M}[H(\bar{p}(\Theta|x))|X_j, \tau=n] = \mathrm{M}(H_n|X_j, \tau=n). \tag{3.6}$$

Expressions (3.5) and (3.6) now yield that

$$\mathrm{M}(H_n|\tau=n) = \sum_{j=1}^{N} \mathrm{M}(H_n|X_j, \tau=n) P(X_j|\tau=n) \leqslant h[P(e|\tau=n)] + P(e|\tau=n) \ln(N-1). \tag{3.7}$$

---

*The symbols $\cap$ and $\cup$ indicate the direction of convexity of a function (see [1]).

Multiplying both sides of (3.7) by $P(\tau = n)$, summing over n, and using the fact that $h(x)$ is $\cap$-convex, we obtain from (3.4) and (3.7)

$$\mathrm{M}H_\tau = \sum_{n=1}^{\infty} \mathrm{M}(H_n|\tau=n)P(\tau=n) \leqslant \sum_{n=1}^{\infty} h[P(e|\tau=n)]P(\tau=n) + P_e \ln(N-1) \leqslant h(P_e) + P_e \ln(N-1).$$

The lemma is thus proved.

Now let us consider in greater detail the process of variation of the entropy H_n of the a posteriori distribution. The following results appear almost obvious from the viewpoint of information theory. Let $\mathscr{F}_n$ be a σ-algebra generated by the random variables $\overline{\xi}'^n$, $\overline{\eta}'^n$. In other words, $\mathscr{F}_n$ corresponds to observations at the receiving end of the channel over time n. Then we have the following lemma.

LEMMA 2. For any $n \geq 0$ we have the inequality

$$\mathrm{M}(H_n - H_{n+1}|\mathscr{F}_n) \leqslant C, \tag{3.8}$$

where C is the channel capacity.

Inequality (3.8) becomes trivial for small H_n. In reality, a stronger result is valid in this case.

LEMMA 3. If $H_n \leq H^*(\mathscr{P})$, then we have the inequality

$$\mathrm{M}(\ln H_n - \ln H_{n+1}|\mathscr{F}_n) \leqslant C_1, \tag{3.9}$$

where

$$C_1 = \max_{i,k} \sum_{l=1}^{L} p_{il} \ln \frac{p_{il}}{p_{kl}}, \qquad C_1 > 0, \tag{3.10}$$

and the quantity $H^*(\mathscr{P})$ depends only on the transition matrix $\mathscr{P}$.

The proofs of this lemma and of the one that follows can be found in the Appendix.

We should note that inequality (3.9) can become an asymptotic equality in the case of two messages. Indeed, let us denote by E_{i_0} and E_{k_0} input symbols for which

$$C_1 = \sum_{l=1}^{L} p_{i_0 l} \ln \frac{p_{i_0 l}}{p_{k_0 l}}.$$

Let $p_2 = 1 - p_1$ and $p_1 \to 1$. We choose a coding as follows:

$$P(\xi_{n+1} = E_{i_0}|\Theta_1, \mathscr{F}_n) = 1 \text{ and } P(\xi_{n+1} = E_{k_0}|\Theta_2, \mathscr{F}_n) = 1.$$

Then we obtain

$$\mathrm{M}(\ln H_n - \ln H_{n+1}|\mathscr{F}_n) = C_1 + \frac{C_1}{\ln p_2} + o\left(\frac{1}{\ln p_2}\right).$$

It is easy to show that $C_1 \geq C$. Indeed,

$$C - C_1 = \max_{\overline{f}} \left[\sum_{i=1}^{K} \sum_{j=1}^{L} f_i p_{ij} \ln\left(p_{ij} \Big/ \sum_{k=1}^{K} f_k p_{kj}\right)\right] - \max_{i,l} \sum_{j=1}^{L} p_{ij} \ln \frac{p_{ij}}{p_{lj}} \leqslant \max_{\overline{f}} \left[\sum_{i=1}^{K} \sum_{j=1}^{L} f_i p_{ij} \ln\left(p_{ij} \Big/ \sum_{i=1}^{K} f_k p_{kj}\right) - \sum_{i=1}^{K} \sum_{j=f}^{L} f_i p_{ij} \ln \frac{p_{ij}}{p_{lj}}\right]$$

$$= \max_{\overline{f}} \left[\sum_{j=1}^{L} \sum_{i=1}^{K} f_i p_{ij} \ln\left(p_{lj} \Big/ \sum_{k=1}^{K} f_k p_{kj}\right)\right] \leqslant \max_{\overline{f}} \left[\sum_{j=1}^{L} \sum_{i=1}^{K} f_i p_{ij} \left(\left(p_{lj} \Big/ \sum_{k=1}^{K} f_k p_{kj}\right) - 1\right)\right] = 0.$$

We note that the sum from the right side of (3.10) is the Kuhlback information $I(i : k)$ between measures corresponding to input signals E_i and E_k (see [7]). Then C_1 can be regarded as the maximum from among the Kuhlback informations $I(i : k)$. A quantity analogous to C_1 also arises in some problems of sequential experimental design [8].

LEMMA 4. For any $n \geq 0$ and $l = 1, \ldots, L$ we have the inequality

$$(\ln H_n - \ln H_{n+1}|\mathscr{F}_n, \xi'_{n+1} = \overline{E}_l) \leqslant \max_{i,k} \ln \frac{p_{kl}}{p_{il}}. \tag{3.11}$$

COROLLARY. From (3.11) we have

$$(\ln H_n - \ln H_{n+1} | \mathscr{F}_n) \leqslant \max_{i,k,l} \ln (p_{kl}/p_{il}) = C_2. \tag{3.12}$$

We note that it suffices to prove the assertions of Lemmas 2-4 for channels with noiseless feedback. Indeed, every transmission method involving "noisy" feedback defines an equivalent transmission method involving noiseless feedback. Noise in the feedback channel essentially means that additional randomization is introduced at the transmitting end into some transmission method involving noiseless feedback.

In concluding this section, let us consider one specific decision-making method that is frequently employed in problems with a random number of observations. Let $\bar{\xi}'^n = (\xi'_1, \ldots, \xi'_n)$ be the sequence of signals obtained at the receiving end of the channel over time n ($\bar{\xi}'^n \in Y^n$), and let $\bar{\eta}'^{n-1} = (\eta'_0, \ldots, \eta'_{n-1})$ be the sequence of signals sent via the feedback channel up to the n-th instant ($\bar{\eta}'^{n-1} \in Z^n$). Consider the product of spaces $Y^n \times Z^n = X^n$ and $Y \times Z = X$. We will employ $x_n = (\bar{\xi}'^n, \bar{\eta}'^{n-1})$ to denote points of space X_n, ($x = (\bar{\xi}', \bar{\eta}') \in X$). In other words, x_n constitutes all the information about the transmitted message that is available at the receiving end of the channel after n instants. The proposed decision-making method consists in the following. After each observation x_n we compute the values of N likelihood functions $\ln \frac{P_j(x_n)}{1-P_j(x_n)}$ ($j = 1, \ldots, N$), where $P_j(x_n)$ is the a posteriori probability of message Θ_j. As soon as one of the likelihood functions exceeds the $\ln(1/\varepsilon)$ level after the next observation in sequence, we make a decision in favor of the corresponding message. Let us also assume that this decision-making procedure terminates with probability 1 for any message Θ_j. Then we can readily obtain

$$P_e = \frac{1}{N} \sum_{j=1}^{N} P(\bar{X}_j | \Theta_j) = \frac{1}{N} \sum_{j=1}^{N} P(e | \Theta_j) \leqslant \frac{\varepsilon}{1+\varepsilon} < \varepsilon, \quad \varepsilon > 0.$$

Thus, the error probability P_e for the proposed decision-making method is less than ε. In Sec. 5 of this paper we will use this decision-making procedure in setting up a transmission method.

We should note that the expression $P_{e\,\max} = \max_{1 \leqslant j \leqslant N} P(e|\Theta_j) < \varepsilon$ does not hold for the proposed decision-making procedure. Indeed, let us write $P_{ij} = P(X_i | \Theta_j)$ and consider matrix $\{P_{ij}\}_{i,j=1}^{N}$ of the following form:

$$\begin{aligned} &P_{11} = P_1, \quad P_{1j} = \alpha, \quad j = 2, \ldots, N; \\ &P_{i1} = \beta, \quad i = 2, \ldots, N; \quad P_{ii} = P, \quad i = 2, \ldots, N; \\ &P_{ij} = 0 \text{ — otherwise.} \end{aligned}$$

We choose parameters α and β so that the following equations are satisfied:

$$\begin{aligned} &P = 1 - \beta; \quad P_1 = 1 - \alpha(N-1); \\ &\varepsilon P = \alpha; \quad \varepsilon P_1 = \beta(N-1). \end{aligned} \tag{3.13}$$

If $\varepsilon(N-1) < 1$, then (3.13) yields

$$\beta > \frac{\varepsilon}{N-1}[1 - \varepsilon(N-1)], \quad \alpha > \varepsilon\left(1 - \frac{\varepsilon}{N-1}\right).$$

But this means that $P(e|\Theta_1) = \alpha(N-1) > \varepsilon(N-1) - \varepsilon^2$. To eliminate the possibility of such examples, we would have to investigate the distribution of the P_{ij} in greater detail. However, it is fairly complex to do this. If, however, we know from the outset that the conditional error probability $P(e|\Theta_j)$ is the same for all hypotheses Θ_j, then the equivalence of P_e and $P_{e\,\max}$ is obvious.

4. LOWER LIMIT FOR THE MEAN TRANSMISSION TIME

The main result of this section is the following theorem.

THEOREM 1. For any transmission method over a discrete stationary channel with feedback the mean number of observations $\bar{\tau}$ satisfies the inequality

$$\bar{\tau} > \frac{\ln N}{C} - \frac{\ln P_e}{C_1} - \frac{\ln(\ln N - \ln P_e)}{C_1} - \frac{P_e \cdot \ln N}{C} + D(\mathscr{P}), \tag{4.1}$$

where

$$C = \max_{\bar{f}} \sum_{k=1}^{K} \sum_{l=1}^{L} f_k p_{kl} \ln \left(p_{kl} \Big/ \sum_{i=1}^{K} f_i p_{il} \right) \tag{4.2}$$

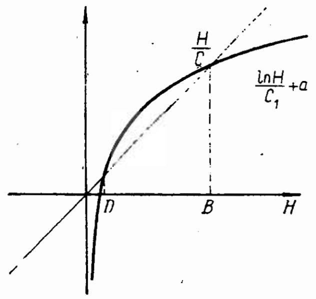

Fig. 1

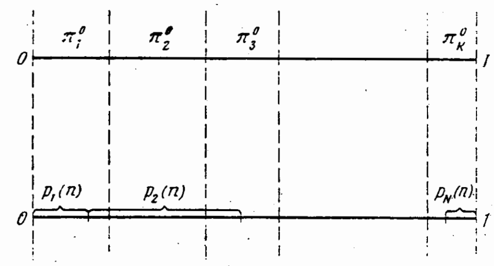

Fig. 2

is the channel capacity,

$$C_1 = \max_{i,k} \sum_{l=1}^{L} p_{il} \ln \frac{p_{il}}{p_{kl}}, \tag{4.3}$$

and D is finite and depends only on the transition matrix $\mathcal{P}$.

Proof. Lemma 1 yields that the mean entropy at the decision-making instant must be small in order for error probability P_e to be achieved. Lemmas 2-4 describe the limits of variation of the entropy over one observation. The entropy H_n decreases with each observation by not more than C, on average. Beginning with sufficiently small H_n values, the nature of the change in the entropy is altered and it becomes more convenient to consider the variation in $\ln H_n$. In this case $\ln H_n$ decreases with each observation by not more than C_1 on average. Assume that $\mathcal{F}_n$ is a σ-algebra corresponding to observations at the receiving end of the channel over time n. Obviously, $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots$. Let us assume that we have succeeded in setting up a sequence of random variables ξ_n that form a submartingale relative to $\{\mathcal{F}_n\}$, i.e.,

$$\mathbf{M}(\xi_{n+1} | \mathcal{F}_n) \geqslant \xi_n, \; n=0, 1, \ldots; \tag{4.4}$$

the ξ_n having the following form:

$$\xi_n = \begin{cases} C^{-1} H_n + n, & \text{if} \quad H_n \geqslant B, \\ C_1^{-1} \ln H_n + g(H_n) + n, & \text{if} \quad H_n \leqslant B, \end{cases} \tag{4.5}$$

where B depends only on $\mathcal{P}$ and $g(H_n)$ is a function that is bounded for all H_n by a constant that also depends only on $\mathcal{P}$. Assume that τ is an arbitrary Markov stopping instant. We will assume that $M\tau < \infty$ (otherwise, the assertion of the theorem is obvious). On the basis of the theorem regarding stopping of a submartingale ([9], p. 62), the sequence $(\xi_{n \wedge \tau}, \mathcal{F}_n)$, n = 0, 1, . . . also forms a submartingale. Therefore,

$$\xi_0 \leqslant \mathbf{M}(\xi_{n\wedge\tau} | \mathcal{F}_0) \leqslant \lim_{n\to\infty} \mathbf{M}(\xi_{n\wedge\tau} | \mathcal{F}_0). \tag{4.6}$$

Taking account of representation (4.5) for ξ_n, we obtain from (4.6) $(H_0 > B)$:

$$\begin{aligned} \xi_0 = C^{-1} H_0 &\leqslant \lim_{n\to\infty} \mathbf{M}(n \wedge \tau | \mathcal{F}_0) + \lim_{n\to\infty} \{C^{-1} \mathbf{M}(H_{n\wedge\tau} \chi_{\{H_{n\wedge\tau} > B\}} | \mathcal{F}_0) \\ &+ C_1^{-1} \mathbf{M}(\ln H_{n\wedge\tau} \chi_{\{H_{n\wedge\tau} \leqslant B\}} | \mathcal{F}_0) + \mathbf{M}(g(H_{n\wedge\tau}) \chi_{\{H_{n\wedge\tau} \leqslant B\}} | \mathcal{F}_0)\} \\ &\leqslant \mathbf{M}(\tau | \mathcal{F}_0) + \lim_{n\to\infty} [C^{-1} \mathbf{M}(H_{n\wedge\tau} | \mathcal{F}_0) + C_1^{-1} \mathbf{M}(\ln H_{n\wedge\tau} | \mathcal{F}_0)] + D(\mathcal{P}). \end{aligned} \tag{4.7}$$

Using the inequality $\mathbf{M}(\ln \xi) \leq \ln(\mathbf{M}\xi)$, we obtain from (4.7) that

$$\begin{aligned} \mathbf{M}(\tau | \mathcal{F}_0) &\geqslant C^{-1} H_0 - C^{-1} \lim_{n\to\infty} \mathbf{M}(H_{n\wedge\tau} | \mathcal{F}_0) \\ &- C_1^{-1} \lim_{n\to\infty} [\ln \mathbf{M}(H_{n\wedge\tau} | \mathcal{F}_0)] + D(\mathcal{P}) = C^{-1} H_0 - C^{-1} \mathbf{M}(H_\tau | \mathcal{F}_0) - C_1^{-1} \ln \mathbf{M}(H_\tau | \mathcal{F}_0) + D(\mathcal{P}). \end{aligned} \tag{4.8}$$

Recalling that $H_0 = \ln N$ and using inequality (3.2) for $\mathbf{M}(H_\tau | \mathcal{F}_0)$, we obtain the necessary inequality (4.1) from (4.8). Thus, the theorem will be proved if we can succeed in setting up a sequence of random variables ξ_n satisfying conditions (4.4) and (4.5). It is fairly simple to do this.

To clarify the idea of our construct, let us imagine that Lemma 3 is valid for all H_n, and let us consider two sequences of random variables:

$$\begin{aligned} \xi_n^1 &= C^{-1} H_n + n, \\ \xi_n^2 &= C_1^{-1} \ln H_n + n + a, \; a > 0. \end{aligned} \tag{4.9}$$

Each of them forms a submartingale relative to $\{\mathcal{F}_n\}$. Figure 1 shows the functions $C^{-1}H$ and $C_1^{-1}\ln H + a$. It is clear that for large a these graphs have two common points B and D; with increasing a, point D shifts leftward toward the coordinate origin, while point B moves to the right. The quantities H_n and H_{n+1} at two successive instants satisfy the inequality [see (3.12)] $H_n - H_{n+1} \le (1 - \min_{i,k,l} (p_{kl}/p_{il}))H_n$. This yields that if the entropy H_n is greater than B, then for sufficiently large a the entropy H_{n+1} cannot be less than D. Therefore, if we create a sequence of random variables ξ_n of the form

$$\xi_n = \begin{cases} C^{-1}H_n+n, & \text{if} \quad H_n \geqslant B, \\ C_1^{-1}\ln H_n+a+n, & \text{if} \quad H_n \leqslant B, \end{cases} \tag{4.10}$$

it is easy to establish that it forms a submartingale relative to $\{\mathcal{F}_n\}$, n = 0, 1, ... Our reasoning will not work, however, because of the fact that Lemma 3 was proved only for $H_n < H^*(\mathcal{P})$. The following lemma makes our reasoning rigorous.

LEMMA 5. Assume that for the sequence $(\eta_n, \mathcal{F}_n)$, n = 0, 1, . . . , we have the inequalities

$$\mathbf{M}(\eta_{n+1}|\mathcal{F}_n) \geqslant \eta_n - C_1, \quad \text{if} \quad \eta_n \geqslant 0, \qquad |\eta_{n+1}-\eta_n| \leqslant C_2,\ C_1>0. \tag{4.11}$$

For parameter α satisfying the condition

$$\alpha \geqslant C_2+C_2^2C_1^{-1}, \tag{4.12}$$

we define the sequence $\{\zeta_n\}$ as follows:

$$\zeta_n = \begin{cases} \dfrac{\eta_n}{C_1}+\dfrac{\alpha^2}{C_1(\eta_n-\alpha)}+n, & \text{if} \quad \eta_n \leqslant 0, \\ -\alpha C_1^{-1}+n, & \text{if} \quad \eta_n \geqslant 0. \end{cases} \tag{4.13}$$

Then the sequence $(\zeta_n, \mathcal{F}_n)$ forms a submartingale, i.e.,

$$\mathbf{M}(\zeta_{n+1}|\mathcal{F}_n) \geqslant \zeta_n, \quad n=0, 1, \ldots. \tag{4.14}$$

Proof. For $\eta_n \le 0$, using conditions (4.11) and (4.12), we have

$$\begin{aligned} \mathbf{M}\left[\frac{\eta_{n+1}}{C_1}+\frac{\alpha^2}{C_1(\eta_{n+1}-\alpha)}+n+1\Big|\mathcal{F}_n\right] &= \eta_n+1+\frac{1}{C_1}\mathbf{M}(\eta_{n+1}-\eta_n|\mathcal{F}_n) \\ +\frac{\alpha^2}{C_1}\mathbf{M}\left(\frac{1}{\eta_{n+1}-\alpha}-\frac{1}{\eta_n-\alpha}\Big|\mathcal{F}_n\right) &= \eta_n+1+\frac{1}{C_1}\left[1-\frac{\alpha^2}{(\eta_n-\alpha)^2}\right]\mathbf{M}(\eta_{n+1}-\eta_n|\mathcal{F}_n) \\ +\frac{\alpha^2}{C_1(\eta_n-\alpha)^2}\mathbf{M}\left[\frac{(\eta_{n+1}-\eta_n)^2}{\eta_{n+1}-\alpha}\Big|\mathcal{F}_n\right] &\geqslant \eta_n+\frac{\alpha^2}{(\eta_n-\alpha)^2}\left[1-\frac{C_2^2}{C_1(\alpha-C_2)}\right] \geqslant \eta_n. \end{aligned} \tag{4.15}$$

Since $\alpha \ge C_2$, for $0 \le x \le C_2$ we have

$$\frac{x}{C_1}+\frac{\alpha^2}{C_1(x-\alpha)} \leqslant -\frac{\alpha}{C_1}. \tag{4.16}$$

Therefore, taking account of the fact that $\eta_{n+1} \le C_2$, we have from (4.15) and (4.16) that

$$\mathbf{M}(\zeta_{n+1}|\mathcal{F}_n) \geqslant \mathbf{M}\left[\frac{\eta_{n+1}}{C_1}+\frac{\alpha^2}{C_1(\eta_{n+1}-\alpha)}+n+1\Big|\mathcal{F}_n\right] \geqslant \zeta_n, \quad \eta_n \leqslant 0. \tag{4.17}$$

If $\eta_n > 0$, we obtain, similarly to (4.15)-(4.17),

$$\begin{aligned} \mathbf{M}(\zeta_{n+1}|\mathcal{F}_n) &= \zeta_n+1+\mathbf{M}\left[\left(\frac{\alpha}{C_1}+\frac{\eta_{n+1}}{C_1}+\frac{\alpha^2}{C_1(\eta_{n+1}-\alpha)}\right)\chi_{(\eta_{n+1}\leqslant 0)}\Big|\mathcal{F}_n\right] \\ &\geqslant \zeta_n+1+\frac{\alpha}{C_1}-\frac{C_2}{C_1}-\frac{\alpha^2}{C_1(\alpha+C_2)} \geqslant \zeta_n, \quad \eta_n>0. \end{aligned} \tag{4.18}$$

Expressions (4.17) and (4.18) yield inequality (4.14).

The lemma has thus been proved.

Now it clear from Lemmas 3 and 4 that the sequence

$$\eta_n = \ln H_n - \ln H^*(\mathcal{P}), \quad n=0, 1, \ldots \tag{4.19}$$

satisfies the conditions of Lemma 5. Therefore, each of the following sequences forms a submartingale relative to the system $\{\mathscr{F}_n\}$, n = 0, 1, . . . :

$$\xi_n^1 = C^{-1}H_n + n,$$

$$\xi_n^2 = \begin{cases} \dfrac{\ln H_n - \ln H^*(\mathscr{P})}{C_1} + \dfrac{\alpha^2}{C_1(\ln H_n - \ln H^*(\mathscr{P}) - \alpha)} + a + n, \\ \text{if} \quad H_n \leqslant H^*(\mathscr{P}); \\ -\alpha C_1^{-1} + a + n, \quad \text{if} \quad H_n \geqslant H^*(\mathscr{P}), \quad a > 0, \end{cases} \tag{4.20}$$

where C and C_1 are defined in (4.2) and (4.3), while parameter α satisfies condition (4.12) with C_2 from (3.12). It is clear that all the reasoning made using Fig. 1 is valid for sequences $\{\xi_n^1\}$ and $\{\xi_n^2\}$; for all H_n the sequence $\{\xi_n^2\}$ behaves like $C_1^{-1} \ln H_n + a + n$. Thus, we have proved that Theorem 1 is valid.

We should note that paper [10] proposed a method for obtaining results similar to those of Theorem 1 that differs from the one employed here. But the proof in [10] is valid for a restricted class of transmission methods. It turns out that the authors of [10] accepted without proof the fact that the optimum transmission strategy cannot terminate before the a posteriori probability of one of the hypotheses attains the $1 - P_e$ level. But this fact is hardly obvious for an arbitrary number of hypotheses. We know only a fairly cumbersome proof of it for the case of two hypotheses (see, e.g., [11]).

5. ASYMPTOTICALLY OPTIMAL TRANSMISSION METHOD

In this section we will create, for the discrete channel under consideration, a transmission method involving noiseless feedback for which the mean decision-making time coincides in the principal term with lower limit (4.1) as $P_e \to 0$.

Assume that the maximum in the definition (3.10) of C_1 is attained on input symbols (E_{i_0}, E_{k_0}), i.e.,

$$C_1 = \sum_{l=1}^{L} p_{i_0 l} \ln \frac{p_{i_0 l}}{p_{k_0 l}}. \tag{5.1}$$

We define the quantity C_1^* as follows:

$$C_1^* = \sum_{l=1}^{L} p_{k_0 l} \ln \frac{p_{k_0 l}}{p_{i_0 l}}. \tag{5.2}$$

If the maximum in (3.10) is achieved on several pairs of input symbols, then we take (E_{i_0}, E_{k_0}) to be the pair for which C_1^* is at a maximum. In setting up the transmission method we need to consider two cases: a) $C_1^* > C$ and b) $C_1^* \leq C$. In both cases the transmission methods are conceptually similar, but case b) calls for a more detailed investigation. Let us begin with the first case, which is simpler.

a) $C_1^* > C$. We denote by $(\pi_1^0, \ldots, \pi_K^0)$ the probability distribution for the input symbols on which the channel capacity C is achieved; without loss of generality we can assume that $\pi_i^0 > 0$, i = 1, . . . , K. Then we have ([1], Theorem 4.5.1) the equality

$$\mathscr{L}_k = \sum_{l=1}^{L} p_{kl} \ln \left(p_{kl} \Big/ \sum_{i=1}^{K} \pi_i^0 p_{il} \right) = C, \quad k = 1, \ldots, K. \tag{5.3}$$

We denote by $p_j(n)$ the a posteriori probability of message Θ_j after n instants [the a posteriori probability of a true message $p_{tr}(n) = p(n)$]. We introduce the quantities

$$Z_j(n) = \ln \frac{p_j(n)}{1 - p_j(n)}, \quad (Z_{tr}(n) = Z(n)), \quad j = 1, \ldots, N. \tag{5.4}$$

We also introduce the Markov instants $\tau(\varepsilon)$ and $\tau_{tr}(\varepsilon)$:

$$\begin{aligned} \tau(\varepsilon) &= \min\{n : \max_j Z_j(n) \geqslant \ln(1/\varepsilon)\}, \\ \tau_{tr}(\varepsilon) &= \min\{n : Z(n) \geqslant \ln(1/\varepsilon)\}. \end{aligned} \tag{5.5}$$

Markov instant $\tau(\varepsilon)$ defines the decision-making rule. As shown at the end of Sec. 3, the error probability P_e for this rule does not exceed ε. If message Θ_j is transmitted, then obviously we have

$$M(\tau(\varepsilon)\mid\Theta_j)\leqslant M(\tau_{tr}(\varepsilon)\mid\Theta_j),\ j=1,\ldots,N. \tag{5.6}$$

Bounding the right side of (5.6) from above, we obtain the necessary upper bound for the mean decision-making time.

Let us now describe the encoding method employed. The upper segment in Fig. 2 successively gives the probabilities $\pi_1^0, \ldots, \pi_K^0$; similarly, the lower segment shows the a posteriori probabilities $p_j(n)$, $j = 1,\ldots,N$. We denote by $\{\pi_j\}$, $\{p_j(n)\}$ the corresponding segments on Fig. 2. We place each segment $\{p_j(n)\}$ in correspondence with a vector $\bar{\alpha}_j = (\alpha_j^1, \ldots, \alpha_j^K)$. The coordinate α_j^i of this vector is equal to the length of the part of $\{p_j(n)\}$ under segment $\{\pi_i^0\}$, $i = 1, \ldots, K$. For example, in Fig. 2 we have $\bar{\alpha}_1 = (p_1(n), 0, \ldots, 0)$, $\bar{\alpha}_2 = (\pi_1^0 - p_1(n), \pi_2^0, p_1(n) + p_2(n) - \pi_1^0 - \pi_2^0, 0, \ldots, 0)$, and so forth. Obviously.

$$\sum_{i=1}^{K}\alpha_j^i=p_j(n),\quad \sum_{j=1}^{N}\alpha_j^i=\pi_i^0,\quad j=1,\ldots,N,\ i=1,\ldots,K. \tag{5.7}$$

Before transmission at the $(n + 1)$-th instant at the receiving end of the channel all messages Θ_j are randomly partitioned into groups $\{\pi_1\}, \ldots, \{\pi_K\}$. Here the probability that message Θ_j will be assigned to group $\{\pi_1\}$ is $\alpha_j^i/p_j(n)$, and randomization of each message is done independently of the other ones. Because of the use of noiseless feedback, the results of randomization become known at the transmitting end, and symbol E_i is transmitted at the $(n + 1)$-th instant if the true message Θ_{tr} was assigned to group $\{\pi_i\}$. Obviously, for the length π_i of random segments $\{\pi_i\}$ we have

$$M(\pi_i|\mathcal{F}_n)=\sum_{j=1}^{N}\alpha_j^i=\pi_i^0,\quad i=1,\ldots,K.$$

If true message Θ_{tr} was assigned to group $\{\pi_k\}$, while a signal $\xi'_{n+1} = \bar{E}_l$ was obtained at the channel output, then $p(n+1)=p(n)p_{kl}\left(\sum_{i=1}^{K}\pi_i p_{il}\right)^{-1}$. As a result we obtain the following equation for Θ_{tr}:

$$M[Z(n+1)-Z(n)\,|\,\mathcal{F}_n, \Theta_{tr}]=\ln(1-p(n))+\sum_{k=1}^{K}\frac{\alpha_{tr}^k}{p(n)}\sum_{l=1}^{L}p_{kl}M\ln\left[p_{kl}\left(\sum_{i=1}^{K}\pi_i p_{il}-p(n)p_{kl}\right)^{-1}\,\middle|\,\mathcal{F}_n,\Theta_{tr}\in\{\pi_k\}\right]. \tag{5.8}$$

Now we note that

$$M[\pi_i|\mathcal{F}_n,\Theta_{tr}\in\{\pi_k\}]=\sum_{\substack{j=1\\ \Theta_j\neq\Theta_{tr}}}^{N}\alpha_j^i+\delta_{ik}\cdot p(n)=\pi_i^0+\delta_{ik}p(n)-\alpha_{tr}^i, \tag{5.9}$$

(δ_{ik} is the Kronecker delta), $i = 1, \ldots, K$. If for the $\cup$-convex function $\ln(1/(\xi - a))$ we now employ the Jensen inequality and (5.9), and then condition (5.3), we obtain from the right side of (5.8) that

$$M[Z(n+1)-Z(n)|\mathcal{F}_n]\geqslant\sum_{k=1}^{K}\frac{\alpha_{tr}^k}{p(n)}\sum_{l=1}^{L}p_{kl}\ln\left[p_{kl}\left(\sum_{i=1}^{K}\pi_i^0 p_{il}\right.\right.$$

$$\left.\left.-\sum_{i=1}^{K}\alpha_{tr}^i p_{il}\right)^{-1}\right]+\ln(1-p(n))=C+\ln(1-p(n))-\frac{1}{p(n)}\sum_{l=1}^{L}\sum_{k=1}^{K}\alpha_{tr}^k p_{kl}\ln\left[1-\left(\sum_{i=1}^{K}\alpha_{tr}^i p_{il}\Big/\sum_{i=1}^{K}\pi_i^0 p_{il}\right)\right]. \tag{5.10}$$

Denoting the double sum on the right side of (5.10) by F and introducing the variables

$$A_l=\sum_{k=1}^{K}\alpha_{tr}^k p_{kl},\quad B_l=\sum_{i=1}^{K}\pi_i^0 p_{il},\quad l=1,\ldots,L,$$

we obtain

$$F=\sum_{l=1}^{L}A_l\ln\left(1-\frac{A_l}{B_l}\right),$$

$$\sum_{l=1}^{L}A_l=p(n),\quad \sum_{l=1}^{L}B_l=1,\quad 0\leqslant A_l\leqslant B_l.$$

Obviously, the function $x \ln(1-x)$ is $\cap$-convex. Therefore, using Jensen's inequality, we obtain

$$F \leqslant \left(\sum_{l=1}^{L} B_l \frac{A_l}{B_l}\right) \ln\left[1 - \left(\sum_{j=1}^{L} B_j \frac{A_j}{B_j}\right)\right] = p(n)\ln(1-p(n)). \tag{5.11}$$

Finally, we have from (5.10) and (5.11) that

$$\mathrm{M}[Z(n+1)-Z(n) \mid \mathcal{F}_n] \geqslant C. \tag{5.12}$$

Thus, for the proposed encoding method the quantity $Z_{tr}(n)$ increases with each observation by not less than C on average. When the a posteriori probability of one of the hypotheses becomes sufficiently large, we alter the encoding method, adapting it to the problem of discriminating two hypotheses. More precisely, we will define the quantity $1/2 \le p_0 < 1$ below in such a way that if $p_j(n) \ge p_0$, then at the $(n+1)$-th instant we examine the two hypotheses $\{\Theta_j$ is a true message$\}$ and $\{\Theta_j$ is a false message$\}$, and encoding will consist in the following: 1) symbol E_{i_0} is transmitted if $\Theta_j = \Theta_{tr}$; 2) symbol E_{k_0} is transmitted if $\Theta_j \neq \Theta_{tr}$ (E_{i_0} and E_{k_0} being defined at the beginning of this section). In this case, if $p_{tr}(n) \ge p_0$, we have

$$\mathrm{M}[Z(n+1)-Z(n) \mid \mathcal{F}_n] = \sum_{l=1}^{L} p_{i_0 l} \ln \frac{p_{i_0 l}}{p_{k_0 l}} = C_1. \tag{5.13}$$

If $p_j(n) \ge p_0$ and $\Theta_j \neq \Theta_{tr}$, i.e., the level of p_0 exceeds the a posteriori probability of a false message, then

$$\mathrm{M}[Z(n+1)-Z(n) \mid \mathcal{F}_n] = \sum_{l=1}^{L} p_{k_0 l} \ln \frac{p_{k_0 l}(1-p(n))}{p_{k_0 l}(1-p(n)) + p_j(n)(p_{i_0 l} - p_{k_0 l})}, \quad p(n) \leqslant p_j(n). \tag{5.14}$$

As $p_j(n) \to 1$ the right side of (5.14) tends to C_1^*. Since, by supposition, $C_1^* > C$, there exists a $p_0 < 1$ such that for $p_j(n) \ge p_0$ the minimum of the right side of (5.14) is equal to C. More precisely, we can determine p_0 from the formula

$$p_0 = \min\left\{x : \min_{\substack{1/2 \leqslant x \leqslant 1 \\ 0 \leqslant y \leqslant 1-x}} \left[\sum_{l=1}^{L} p_{k_0 l} \ln \frac{p_{k_0 l}(1-y)}{p_{k_0 l}(1-y) + x(p_{i_0 l} - p_{k_0 l})}\right] = C\right\}. \tag{5.15}$$

Now the transmission method has been fully described. For the $Z_{tr}(n) = Z(n)$ we have from (5.12)-(5.15) that

$$\begin{aligned} &\mathrm{M}[Z(n+1)-Z(n) \mid \mathcal{F}_n] \geqslant C, \quad \text{if} \quad Z(n) < \ln \frac{p_0}{1-p_0}; \\ &\mathrm{M}[Z(n+1)-Z(n) \mid \mathcal{F}_n] = C_1, \quad \text{if} \quad Z(n) \geqslant \ln \frac{p_0}{1-p_0}. \end{aligned} \tag{5.16}$$

Moreover, we can readily obtain the formulas $|Z(n+1) - Z(n)| \le C_2 = \max_{i,j,l} \ln(p_{jl}/p_{il})$. The following lemma was proved in [6].

LEMMA 6. Assume that the sequence $(\xi_k, \mathcal{F}_k)$, $k = 0, 1, \ldots$ forms a submartingale, where

$$\begin{aligned} &\mathrm{M}(\xi_{k+1} \mid \mathcal{F}_k) \geqslant \xi_k + C, \quad \text{if} \quad \xi_k < 0, \; C > 0, \\ &\mathrm{M}(\xi_{k+1} \mid \mathcal{F}_k) \geqslant \xi_k + C_1, \quad \text{if} \quad \xi_k \geqslant 0, \; C_1 > C, \\ &|\xi_{k+1} - \xi_k| \leqslant C_2, \quad \xi_0 < 0, \end{aligned}$$

and the Markov instant τ is given by the condition $\tau = \min\{n: \xi_n \ge B\}$, $B > 0$. Then we have the inequality $\mathrm{M}(\tau \mid \mathcal{F}_0) \le C_1^{-1}B + C^{-1}|\xi_0| + D(C, C_1, C_2)$, where function D depends only on C, C_1, and C_2.

Obviously, the conditions of Lemma 6 hold for the sequence $\left(Z(n) - \ln \frac{p_0}{1-p_0}, \mathcal{F}_n\right)$, $n = 0, 1, \ldots$. Therefore, we have

$$\mathrm{M}[\tau_{tr}(P_e) \mid \Theta_j] < C^{-1} \ln N + C_1^{-1} \ln(1/P_e) + D(\mathcal{P}). \tag{5.17}$$

b) $C_1^* \le C$. In this case the first inequality in (5.16) is violated at those instants at which the a posteriori probability of one of the false messages exceeds the p_0 level and the true message in the encoding process is relegated to an "unfavorable" group corresponding to input symbol E_{k_0}. It is clear, however, that if we choose p_0 in accordance with P_e and N to be sufficiently close to unity, this event will have a low probability and will not significantly affect the mean decision-making time. To describe the transmission method more precisely, we specify p_0, the level $Z_0 = \ln(p_0/(1-p_0))$ associated with it, and yet another level $A < 0$. All these quantities, which depend on P_e and N, will be determined below. Our proposed method consists in the following.

1) If prior to instant n the a posteriori probabilities of all messages Θ_j were less than p_0. the transmission method at the $(n+1)$-th instant coincides with that described in case a).

2) Assume that after the n-th instant the a posteriori probability of one of the messages first exceeds the p_0 level, e.g., $p_j(n) \geq p_0$. Then we subsequently solve a problem of discriminating two hypotheses: $G_0 = \{\Theta_j$ is a true message$\}$ and $G_1 = \{\Theta_j$ is a false message$\}$. Here G_0 is placed in correspondence with input symbol E_{i_0}, while G_1 is placed in correspondence with E_{k_0}. If $Z_j(n)$ corresponding to G_0 attains the $\ln(1/P_e)$ level before it attains the A level, we make a decision in favor of message Θ_j. If the A level is reached first, then hypothesis G_0 is rejected and transmission begins from the very beginning, i.e., with equiprobable a priori distribution for all messages. If $p_j(n) \geq p_0$ and $\Theta_j \neq \Theta_{tr}$, then the following relationships are valid during the time that G_0 and G_1 are discriminated:

$$M[Z_j(k+1)-Z_j(k)\,|\,\mathcal{F}_k]=-C_1^*,\quad k\geqslant n,$$
$$|Z_j(k+1)-Z_j(k)|\leqslant\max|\ln(p_{i_0l}/p_{k_0l})|\leqslant C_2. \tag{5.18}$$

We denote by P_1 the probability that $Z_{tr}(n)$ will attain the A level before $\ln(1/P_e)$ in the problem of discriminating two hypotheses. Then for the mean decision-making time τ we readily obtain the equation

$$\bar{\tau}<\frac{Z_0-Z(0)}{C}+p_0\left[\frac{\ln(1/P_e)-Z_0}{C_1}+P_1\bar{\tau}\right]+(1-p_0)\left[\frac{Z_0-A}{C_1^*}+\bar{\tau}\right]+D(\mathscr{P}). \tag{5.19}$$

To estimate P_1, we note that $M[\exp\{-(Z(n+1)-Z(n))\}\,|\,\mathcal{F}_n]=1$. This means that the sequence $(e^{-Z(n)}, \mathcal{F}_n)$, $n=0,1,\ldots,$ forms a martingale. Therefore, if we take τ to be the instant that either A or $\ln(1/P_e)$ is first reached, we readily obtain that

$$e^{-Z_0}\geqslant e^{-Z(0)}=M[e^{-Z(\tau)}\,|\,\mathcal{F}_0]\geqslant P_1e^{-A}+(1-P_1)\exp\{-(\ln(1/P_e)+C_2)\}. \tag{5.20}$$

(Here the zero instant corresponds to the moment that Z_{tr} first intersects the Z_0 level.)

From this we have

$$P_1\leqslant(e^{-Z_0}-P_ee^{-C_2})/(e^{-A}-P_ee^{-C_2}). \tag{5.21}$$

Now let us set

$$\varepsilon=1-p_0=\left[\ln\left(\frac{\ln N}{C}-\frac{\ln P_e}{C_1}\right)\right]^{-1},\quad A=-1/\varepsilon. \tag{5.22}$$

Then we obtain from (5.19)-(5.22), for sufficiently small P_e,

$$\bar{\tau}<\left(\frac{\ln N}{C}-\frac{\ln P_e}{C_1}+\frac{C_1-C}{CC_1}\ln\frac{1}{\varepsilon}\right)\Big/(1-\varepsilon)+D(\mathscr{P}). \tag{5.23}$$

The results of this section can be formulated in the following theorem.

THEOREM 2. If $C_1^* > C$, the mean decision-making time of our proposed transmission method satisfies the inequality

$$M(\tau\,|\,\Theta_{tr})<C^{-1}\ln N-C_1^{-1}\ln P_e+D(\mathscr{P}).$$

If $C_1^* \leq C$, then inequality (5.23) holds, where ε is defined in (5.22) and D is finite and depends only on $\mathscr{P}$.

The results of Theorems 1 and 2 permit us to readily compute the reliability function $E(\bar{R})$ of the channel with noiseless feedback under consideration. Indeed, let us define the mean rate $\bar{R}$ and reliability function $E(\bar{R})$ of the transmission method as follows:

$$\bar{R}=(\ln N)/\bar{\tau},$$
$$E(\bar{R})=\lim_{P_e\to 0}(-(\ln P_e)/\bar{\tau}). \tag{5.24}$$

Then we have the following theorem.

THEOREM 3. The optimum reliability function $E_{opt}(\bar{R})$ of the discrete channel with noiseless feedback under consideration is $E_{opt}(\bar{R}) = C_1(1-(\bar{R}/C))$, $0 \leq \bar{R} \leq C$.

In concluding this section, we offer two examples.

Example 1. For a DSC with channel error probability p we have

$$C_1 = C_1^* = (q-p)\ln\frac{q}{p} < C = p\ln 2p + q\ln 2q, \qquad q+p=1.$$

Example 2. Consider a channel with transition matrix $\mathscr{P}$ of the following form:

$$\mathscr{P} = \begin{Vmatrix} \varepsilon & (1-\varepsilon)/n \ldots\ldots & (1-\varepsilon)/n \\ 1-\alpha & \alpha/n \quad\ldots\ldots & \alpha/n \\ \ldots & \ldots\ldots\ldots & \ldots \\ \alpha/n & \alpha/n \ldots \quad\ldots & 1-\alpha \end{Vmatrix}, \quad 0<\alpha<1.$$

For it for $\varepsilon < 1/(n+1)$ we have

$$C_1 = (1-\alpha)\ln(1/\varepsilon) - H(\alpha) - \alpha\ln(1-\varepsilon),$$
$$C_1^* = -(1-\varepsilon)\ln\alpha - \varepsilon\ln(1-\alpha) - H(\varepsilon),$$
$$C > (1-\alpha)\ln(n+1) - H(\alpha) + \alpha\ln((n+1)/n).$$

Obviously, in this case, we have $C > C_1^*$ for sufficiently large n.

6. TRANSMISSION WITH ZERO ERROR PROBABILITY

If transition matrix $\mathscr{P}$ contains zero elements, we formally cannot employ the results of Theorems 1 and 2, since in this case $H_j(n)$, $Z_j(n)$, etc., can have jumps of infinite size. But we can readily obtain new analogs of these theorems. Indeed, Lemmas 1 and 2 remain valid. Therefore, if we take account of the fact that the sequence $(C^{-1}H_n + n, \mathscr{F}_n)$, $n = 0, 1, \ldots$ forms a submartingale, we have for an arbitrary Markov stopping instant τ

$$C^{-1}\mathbf{M}(H_\tau|\mathscr{F}_0) + \mathbf{M}(\tau|\mathscr{F}_0) \geqslant C^{-1}H_0.$$

From this we have

$$\mathbf{M}(\tau|\mathscr{F}_0) \geqslant C^{-1}[\ln N - h(P_e) - P_e\ln(N-1)], \tag{6.1}$$

and this yields a lower limit for the mean decision-making time.

In this "degenerate" case we were unable to investigate a transmission method analogous to those set up in Sec. 5, because of the fact that $Z_{tr}(n)$ can have jumps of infinite size with positive probability in this case. Therefore, we will limit ourselves to proving the existence of a transmission method with zero error probability for which the mean decision-making time is equivalent to the right side of (6.1) as $N \to \infty$ and $P_e \to 0$.

We denote by $P(N, n)$ the minimum error probability achieved by means of a code of volume N and length n without employing feedback. It is known [1] that

$$P(N,n) \leqslant \exp\left\{-\frac{n}{\alpha}\left(C - \frac{\ln N}{n}\right)^2\right\}, \tag{6.2}$$

where α is some quantity that depends only on $\mathscr{P}$. [It is known, for example, that $\alpha \le 8e^{-2} + 4(\ln L)^2$.] We select some code satisfying condition (6.2). The quantity n will be determined subsequently.

Assume that after transmission with the code in question message Θ_j has maximum a posteriori probability. We will assume that $p_{i_0 1} = 0$ and $p_{k_0 1} \geq p_{k_1}$, $k = 1, \ldots, K$. Now let us consider the problem of discriminating the two hypotheses $H_0 = \{\Theta_j$ is a true message$\}$ and $H_1 = \{\Theta_j$ is a false message$\}$, placing H_0 in correspondence with input symbol E_{i_0} and H_1 in correspondence with E_{k_0}. As soon as output symbol $\bar{E}_1$ is received at the receiving end, transmission terminates and a decision is made in favor of hypothesis H_0. If symbol $\bar{E}_1$ is not received at all over time n_1, then transmission begins anew with uniform a priori distribution. We can readily obtain the following expression for the mean decision-making time of the transmission method in question:

$$\bar{\tau} < n + P(N,n)(\bar{\tau} + n_1) + (1 - P(N,n))[p_{k_0 1}^{-1} + (1-p_{k_0 1})^{n_1}(n_1 + \bar{\tau})]. \tag{6.3}$$

Now let us set

$$n = \frac{\ln N}{C(1-\varepsilon)}, \qquad n_1 = -\frac{\varepsilon^2 C\ln N}{\alpha(1-\varepsilon)\ln(1-p_{k_0 1})},$$

where $\varepsilon = \sqrt{(\alpha\ln\ln N)/(C\ln N)}$.

Now we obtain from (6.2) and (6.3) that

$$\bar{\tau} < \frac{\ln N}{C}\left[1+\sqrt{\frac{\alpha \ln\ln N}{C\ln N}}+\frac{\alpha \ln\ln N}{2C\ln N}\right]+D(\mathscr{P}). \tag{6.4}$$

It is obvious that the right side of (6.4) is equivalent to lower limit (6.1) as $N \to \infty$ and $P_e \to 0$.

APPENDIX

Taking account of the remarks made in Sec. 3, we will prove Lemmas 3 and 4 for channels with noiseless feedback. We will require the following additional lemma.

LEMMA 7. For arbitrary nonnegative p_l, f_i, β_{il}, $l = 1, \ldots, L$; $i = 1, \ldots, N$ we have the inequality

$$\sum_{l=1}^{L} p_l \ln\left(\sum_{i=1}^{N} f_i \Big/ \sum_{i=1}^{N} \beta_{il}\right) \leqslant \max_i \sum_{l=1}^{L} p_l \ln \frac{f_i}{\beta_{il}}. \tag{A1}$$

Proof. Obviously, it suffices to establish that (A1) is valid for N = 2, since the more general case can readily be reduced to this one. Without loss of generality, we will assume that $f_1 + f_2 = 1$ and $\sum_{l=1}^{L} p_l = 1$. Let us find the minimum value of the right side of (A1) relative to $\bar{f} = (f_1, f_2)$. It is easy to understand that at the minimum point we have the condition

$$\sum_{l=1}^{L} p_l \ln \frac{f_1}{\beta_{1l}} = \sum_{l=1}^{L} p_l \ln \frac{f_2}{\beta_{2l}}, \qquad f_1+f_2=1. \tag{A2}$$

Solving (A2) with respect to f_1 and substituting the resultant values of f_1 and f_2 into the right side of (A1), we obtain

$$\min_{f_1+f_2=1} \left[\max_i \sum_{l=1}^{L} p_l \ln \frac{f_1+f_2}{\beta_{1l}+\beta_{2l}}\right] = -\ln\left[\prod_{l=1}^{L} (\beta_{1l})^{p_l} + \prod_{l=1}^{L} (\beta_{2l})^{p_l}\right]. \tag{A3}$$

Since the left side of (A1) is equal to $-\ln\left[\prod_{l=1}^{L} (\beta_{1l}+\beta_{2l})^{p_l}\right]$, it remains to establish that the following inequality is valid:

$$\prod_{l=1}^{L} (\beta_{1l}+\beta_{2l})^{p_l} \geqslant \prod_{l=1}^{L} (\beta_{1l})^{p_l} + \prod_{l=1}^{L} (\beta_{2l})^{p_l}, \qquad \sum_{l=1}^{L} p_l = 1. \tag{A4}$$

The validity of this last relationship is implied by inequality (2.7.2) of [12].

Proof of Lemma 3. We introduce the following notation (which apply to the entire Appendix):

$$\begin{aligned} &f_i = P(\Theta_i|\mathscr{F}_n), \qquad i=1,\ldots,N;\\ &f_i(l) = P(\Theta_i|\mathscr{F}_n, \xi'_{n+1}=E_l), \qquad l=1,\ldots,L;\\ &p(l|\Theta_i) = P(\xi'_{n+1}=E_l|\mathscr{F}_n, \Theta_i),\\ &p(l) = P(\xi'_{n+1}=E_l|\mathscr{F}_n). \end{aligned} \tag{A5}$$

Using (A5) and Lemma 6, we have

$$\mathrm{M}(\ln H_n - \ln H_{n+1}|\mathscr{F}_n) = \sum_{l=1}^{L} p(l)\ln\left[\frac{-\Sigma f_i \ln f_i}{-\Sigma f_i(l)\ln f_i(l)}\right] \leqslant \max_i\left\{\sum_{l=1}^{L} p(l)\ln\left[\frac{-f_i\ln f_i}{-f_i(l)\ln f_i(l)}\right]\right\} = \max_i\{F_i\}. \tag{A6}$$

For the quantities $p(l)$ and $f_i(l)$ from the transmission method we have

$$\begin{aligned} &p(l) = f_i p(l|\Theta_i) + (1-f_i)p(l|\bar{\Theta}_i),\\ &f_i(l) = f_i p(l|\Theta_i)/p(l), \end{aligned} \tag{A7}$$

where

$$p(l|\Theta_i) = \sum_{k=1}^{K} p(\xi_{n+1}=E_k|\Theta_i)p_{kl} = \sum_{k=1}^{K} \alpha_{ik}p_{kl},$$

$$p(l|\overline{\Theta}_i)=p(\xi'_{n+1}=E_l|\Theta_{\text{tr}}\neq\Theta_i)=\sum_{k=1}^{K}p(\xi_{n+1}=E_k|\overline{\Theta}_i)p_{kl}=\sum_{k=1}^{K}\beta_{ik}p_{kl}. \tag{A8}$$

Obviously, for fixed i the variables f_i, $\{\alpha_{ik}\}$ and $\{\beta_{ik}\}$ are mutually independent, and

$$\sum_{k=1}^{K}\alpha_{ik}=1,\quad \sum_{k=1}^{K}\beta_{ik}=1,\quad \alpha_{ik},\beta_{ik}\geqslant 0.$$

Now we will show that the maximum of F_i relative to the variables $\{\alpha_{ik}\}$ and $\{\beta_{ik}\}$ is attained on degenerate distributions $\{\alpha_{ik}\}$ and $\{\beta_{ik}\}$, i.e., either $\alpha_{ik}=0$ or $\alpha_{ik}=1$, and similarly for β_{ik}, $k=1,\ldots,K$. First let us consider maximization relative to $\{\beta_{ik}\}$. We can assume that $\beta_{i1},\ldots,\beta_{iK-1}$ are independent variables, while $\beta_{iK}=1-\sum_{k=1}^{K-1}\beta_{ik}$. Then for the derivatives of F_i we readily obtain

$$\frac{d^2F_i}{d\beta_{ik}^2}=\frac{\partial^2F_i}{\partial\beta_{ik}^2}+\frac{\partial^2F_i}{\partial\beta_{iK}^2}-2\frac{\partial^2F_i}{\partial\beta_{ik}\,\partial\beta_{iK}},$$

$$\frac{\partial^2F_i}{\partial\beta_{ik}\,\partial\beta_{iK}}=(1-f_i)^2\sum_{l=1}^{L}\frac{\partial^2F_i}{\partial p^2(l)}p_{kl}p_{Kl},$$

$$\frac{\partial^2F_i}{\partial p^2(l)}=\frac{1}{p(l)}\left[1-\left(\ln\frac{p(l)}{f_ip(l|\Theta_i)}\right)^{-1}+\left(\ln\frac{p(l)}{f_ip(l|\Theta_i)}\right)^{-2}\right]\geqslant 0. \tag{A9}$$

Expressions (A9) yield

$$\frac{d^2F_i}{d\beta_{ik}^2}=(1-f_i)^2\sum_{l=1}^{L}\frac{\partial^2F_i}{\partial p^2(l)}(p_{kl}-p_{Kl})^2>0,\quad k=1,\ldots,K-1. \tag{A10}$$

Consequently, for fixed i all the β_{ik} except for one are 0. Similarly, we can obtain

$$\frac{d^2F_i}{d\alpha_{ik}^2}=\sum_{l=1}^{L}(p_{kl}-p_{Kl})^2\frac{[p(l)-f_ip(l|\Theta_i)]^2}{p(l)p^2(l|\Theta_i)}\left[1-\left(\ln\frac{p(l)}{f_ip(l|\Theta_i)}\right)^{-1}+\left(\ln\frac{p(l)}{f_ip(l|\Theta_i)}\right)^{-2}\right]>0. \tag{A11}$$

Consequently, either $\alpha_{ik}=0$ or $\alpha_{ik}=1$, $k=1,\ldots,K$. Now from (A6) and (A7) we have

$$\mathrm{M}(\ln H_n-\ln H_{n+1}|\mathscr{F}_n)\leqslant\max_{j,k}\max_{f}\left\{\sum_{l=1}^{L}p(l)\ln\frac{f\ln f}{f(l)\ln f(l)}\right\}, \tag{A12}$$

where

$$p(l)=fp_{kl}+(1-f)p_{jl},\quad 0\leqslant f\leqslant 1,$$
$$f(l)=fp_{kl}/p(l). \tag{A13}$$

We were not able to maximize in the right side of (A12) for all possible f values. To establish that Lemma 3 is valid, however, it suffices to maximize the right side of (A12) as $f\to 0$ and $f\to 1$. We can readily show that the sum in braces on the right side of (A12) is equal to

$$\sum_{l=1}^{L}p_{jl}\ln\frac{p_{jl}}{p_{kl}}\left[1+\frac{1}{\ln f}+o\left(\frac{1}{\ln f}\right)\right]\quad\text{as}\quad f\to 0,$$
$$\sum_{l=1}^{L}p_{kl}\ln\frac{p_{kl}}{p_{jl}}-(1-f)\sum_{l=1}^{L}(p_{kl}-p_{jl})\ln\frac{p_{kl}}{p_{jl}}+o(1-f)\quad\text{as}\quad f\to 1. \tag{A14}$$

Using the inequality $\sum(\alpha_l-\beta_l)\ln\frac{\alpha_l}{\beta_l}\geqslant 0$, $\sum\alpha_l=\sum\beta_l=1$, we can readily establish that expressions (A14) imply that Lemma 3 is valid.

Proof of Lemma 4. Using the obvious inequality

$$\sum_{i=1}^{K}\alpha_i\Big/\sum_{l=1}^{K}\beta_l\geqslant\min_i\frac{\alpha_i}{\beta_i},\quad \alpha_i,\beta_i\geqslant 0,$$

the notation from the preceding proof, and the fact that the function $-x\ln x$ is $\cap$ -convex, we have

$$X(l)=\frac{(H_{n+1}|\mathcal{F}_n, \xi'_{n+1}=E_l)}{(H_n|\mathcal{F}_n)}=\frac{-\sum_{i=1}^{N} f_i(l)\ln f_i(l)}{-\sum_{i=1}^{N} f_i\ln f_i}\geqslant \min_i\left[\frac{f_i(l)\ln f_i(l)}{f_i\ln f_i}\right]$$

$$\geqslant \min_{j,k}\min_{0\leqslant f\leqslant 1}\left[\left(\frac{p_{jl}f}{p_{jl}f+(1-f)p_{kl}}\ln\frac{p_{jl}f}{p_{jl}f+(1-f)p_{kl}}\right)\Big/ f\ln f\right]. \tag{A15}$$

First assume that $p_{kl} \geq p_{jl}$. Then the expression in brackets on the right side of (A15) is not less than

$$\min_f\left[\frac{p_{jl}}{p_{jl}f+(1-f)p_{kl}}\right]=\frac{p_{jl}}{p_{kl}}. \tag{A16}$$

Now let $p_{jl} \geq p_{kl}$. Then it is easy to establish that for any $0 \leq x \leq 1$, $1 \leq a \leq 1/x$ we have

$$\frac{ax\ln ax}{x\ln x}\geqslant\frac{1-ax}{1-x}.$$

Using this inequality, we obtain that the expression in braces on the right side of (A15) is not less than

$$\min_f\left[\left(1-\frac{fp_{jl}}{fp_{jl}+(1-f)p_{kl}}\right)\Big/(1-f)\right]=\min_f\left[\frac{p_{kl}}{fp_{jl}+(1-f)p_{kl}}\right]=\frac{p_{kl}}{p_{jl}}. \tag{A17}$$

Now the necessary inequality (3.11) follows from (A15)-(A17), and the lemma is thus proved.

LITERATURE CITED

1. R. G. Gallager, Information Theory and Reliable Communication, Wiley (1968).
2. E. R. Berlekamp, "Block coding with noiseless feedback," Doctoral Dissertation, Massachusetts Institute of Technology (1964).
3. R. L. Dobrushin, "Data transmission in a channel with feedback," Teor. Veroyatn. Ee Primen., 3, No. 4, 395-412 (1958).
4. R. L. Dobrushin, "Mathematical aspects of Shannon's theory of optimal information encoding," in: Problems of Information Transmission [in Russian], No. 10, Izd. Akad. Nauk SSSR, Moscow (1961), pp. 63-107.
5. M. V. Burnashev and K. Sh. Zigangirov, "One problem of interval estimation in observation control," Probl. Peredachi Inform., 10, No. 3, 51-61 (1974).
6. M. V. Burnashev and K. Sh. Zigangirov, "One problem of observation control," Probl. Peredachi Inform., 11, No. 3, 44-52 (1975).
7. S. Kuhlback, Information Theory and Statistics, Peter Smith.
8. H. Chernoff, "Sequential analysis and optimal design," Soc. Industr. Appl. Math. (1972).
9. R. Sh. Liptser and A. N. Shiryaev, Statistics of Stochastic Processes [in Russian], Nauka, Moscow (1974).
10. J. P. M. Schalkwijk and M. Barron, "Sequential signaling under a peak power constraint," IEEE Trans. Inform. Theory, 17, No. 3, 278-282 (1971).
11. E. Lehmann, Testing Statistical Hypotheses, Wiley (1959).
12. G. H. Hardy, J. E. Littlewood, and G. Polya, Inequalities, 2nd ed., Cambridge University Press (1952).