KYBERNETIKA — VOLUME 16 (1980), NUMBER 4

Output-Based Estimation of Communication Channels*

IGOR VAJDA

Maximum likelihood estimator of a channel state is obtained and studied provided linear coding and automatic repeat request decoding are employed and provided the channel noise is additive and between-block independent. The estimator is based only on the information represented by the repeat request statistic available at the output.

1. PROBABILITY SPACE OF REPEAT REQUESTS

Let us consider a communication channel with input and output alphabet $A = \{0, 1, \dots, a - 1\}$ and with additive noise. The channel is then described by a system of probability distributions $P_N(\cdot \mid s)$ on A^N, $N = 1, 2, \dots$. As we see, probabilities $P_N(x \mid s)$ of error sequences $x \in A^N$ are supposed to depend also on a collection $s = (s_1, \dots, s_r)$ of parameters describing states of the channel. Denote the set of all possible channel states by S and suppose $S \subset E_r$.

If a linear (N, L)-code $C_N \subset A^N$ is applied at the input of the channel, and if the noise words corresponding to any two codewords sent subsequently through the channel are statistically independent, then the channel is well described by the family of probability distributions $P_N(\cdot \mid s)$, $s \in S$, alone. In what follows we consider such codes and channels only, for which this independence assumption is satisfied. (This assumption is fulfilled even by channels with bursting errors provided proper acknowledgement backward signals follow all codewords sent forward – cf. the ARQ assumption below.)

We also suppose that, during subsequent transmission of codewords from C_N, the channel state $s \in S$ remains unaltered. Thus, if $(x_0, x_1, \dots)$ is a sequence of consecutive noise words added to respective codewords in the channel, then this sequence

* The problem of output-based estimation has first been outlined in [1].

can be interpreted as a sequence of mutually independent realizations of random sampling from sample probability space $(A^N, P_N(\cdot \mid s))$, $s \in S$ fixed. In other words, probability distribution of sequences of noise words on the infinite product sample space $(A^N)^\infty$ is supposed to be given as infinite product probability $P_N^\infty(\cdot \mid s)$ for some $s \in S$.

Suppose finally that decoding at the channel output consists in automatic repeat requests (ARQ, see [2, 3]) as long as no formal error is detected in the output word. Thus, if $x \in C_N$ is sent through the channel $P_N(\cdot \mid s)$ and if $(x_0, x_1, \ldots)$ is a respective sequence of noise words, then either

(i) $x + x_0 (\text{mod } a)$ is accepted by the decoder, which takes place iff $x + x_0 (\text{mod } a) \in C_N$, i.e. iff $x_0 \in C_N$ (error of such a decoding now depends on whether x_0 is a zero or non-zero word, but we are not interested in decoding errors) or

(ii) $x + x_0 (\text{mod } a)$ is rejected by the decoder (i.e. $x_0 \notin C_N$), in which case the x is retransmitted and the whole process is repeated with x_1 replacing x_0, etc.

From here we see that communication of a codeword x through the channel requires $t \in \{0, 1, \ldots, \infty\}$ retransmissions of the x iff

$$(x_0, x_1, \ldots) \in B(t) = \left\langle \begin{array}{ll} C_N \times (A^N)^\infty & \text{if } t = 0 \\ (A^N - C_N)^t \times C_N \times (A^N)^\infty & \text{if } 0 < t < \infty \\ (A^N - C_N)^\infty & \text{if } t = \infty , \end{array} \right.$$

where $\{B(t) : t = 0, 1, \ldots, \infty\}$ is a disjoint decomposition of the sample space $(A^N)^\infty$ of all noise sequences. We also see that the number of retransmissions is independent of the communicated codeword $x \in C_N$. This together with what has been said about noise sequence sample probability space implies that if T denotes the number of retransmissions of a codeword from C_N through the channel $P_N(\cdot \mid s)$, then T is a random variable attaining values $t = 0, 1, \ldots, \infty$ with probabilities

$$P_N^\infty(B(t) \mid s) = \left\langle \begin{array}{l} P_N(C_N \mid s) \,.\, 1 \\ P_N^t((A^N - C_N)^t \mid s) \,.\, P_N(C_N \mid s) \,.\, 1 \\ P_N^\infty((A^N - C_N)^\infty \mid s) \end{array} \right\rangle =$$

$$= \left\langle \begin{array}{l} P_N(C_N \mid s) \\ P_N(A^N - C_N \mid s)^t \, P_N(C_N \mid s) \\ P_N(A^N - C_N \mid s)^\infty \end{array} \right\rangle = \left\langle \begin{array}{ll} P_N(C_N \mid s) & \text{if } t = 0 \\ (1 - P_N(C_N \mid s))^t \, P_N(C_N \mid s) & \text{if } 0 < t < \infty \\ 0 & \text{if } t = \infty \,. \end{array} \right.$$

Thus, if we denote

$$\pi(s) = P_N(C_N \mid s) \quad \text{for} \quad s \in S \,, \quad \pi(s) \in (0, 1) \,, \tag{1}$$

then the sample probability space of the repeat request statistic T is $(\{0, 1, \ldots\}, p(\cdot \mid s))$, where $p(t \mid s) = \pi(s)\,(1 - \pi(s))^t$, $t = 0, 1, \ldots$. The independence assumption made above also implies that if $T_1, T_2, \ldots, T_n$ are repeat request statistics corresponding to arbitrary n codewords from C_N communicated through the channel, then they are mutually indepedent and each one is identically distributed with the T above. Thus we have proved the following theorem.

Theorem 1. If C_N is arbitrary linear (N, L)-code at the input of a chennel $P_N(\cdot \mid s)$ with states $s \in S$ and with between-block independent noise and ARQ decoding, and if $T_1, T_2, \ldots, T_n$ is a sequence of numbers of repeat requests corresponding to n subsequently communicated codewords from C_N, then

$$\Pr[T_1 = t_1, \ldots, T_n = t_n] = p^n(t_1, \ldots, t_n \mid s) = \pi(s)^n \, (1 - \pi(s))^{\Sigma t_i} \tag{2}$$

$$\Pr[T_1 + \ldots + T_n = t] = p^*(t \mid s) = \binom{n + t - 1}{t} \pi(s)^n (1 - \pi(s))^t \tag{3}$$

where $\pi(s)$ is given by (1) and $t, t_1, t_2, \ldots \in \{0, 1, \ldots\}$ are arbitrary. In other words, $T_1, T_2, \ldots, T_n$ are independent random variables, each one being distributed geometrically with parameter $1 - \pi(s)$, while their sum is negative-binomial random variable with parameters n and $1 - \pi(s)$.

The random vector $T_1, T_2, \ldots, T_n$ will be called repeat request statistic of size n. Prior entering analysis of statistical problems involving the repeat request statistics we introduce three concrete examples of this statistic.

Example 1. Suppose the channel is binary memoryless, $P_N(x \mid s) = s^{\Sigma x}(1 - s)^{N - \Sigma x}$ where $\Sigma x = (x)_1 + \ldots + (x)_N$ for $x \in A^N = \{0, 1\}^N$. In this case $S \subset [0, \frac{1}{2})$. Suppose that C_N is the iterated $(N_1 N_2, (N_1 - 1)(N_2 - 1))$-code, whose codewords x can be arranged into the following rectangular form

$$x = \begin{bmatrix} \xi_{11}, & \ldots, \xi_{1N_1-1}, & \xi_{1N_1} \\ \xi_{21}, & \ldots, \xi_{2N_1-1}, & \xi_{2N_1} \\ \ldots & \ldots\ldots & \ldots \\ \xi_{N_2-11}, & \ldots, \xi_{N_2-1N_1-1} & \xi_{N_2-1N_1} \\ \xi_{N_21}, & \ldots, \xi_{N_2N_1-1}, & \xi_{N_2N_1} \end{bmatrix} \quad N_1 N_2 = N,$$

where the right-hand digits check parity in rows and the bottom-line digits check parity in columns. The rest represents arbitrary information digits .

As a combinatorial analysis of this situation reveals, $\pi(s) \doteq 1 - N_1 N_2 s$ for s apriori very small and for the code parameters $N_1 N_2$ **such that** $N_1 N_2 s < 1$.

Example 2. If the situation is as above but there is in S a state s which is not a small number, we can calculate $\pi(s)$ at least for the simple iterated code with $N_1 = N_2 = 2$. Here the two possible codewords are of the form

$$x = \begin{bmatrix} 0, & 0 \\ 0, & 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 1, & 1 \\ 1, & 1 \end{bmatrix}$$

so that $\pi(s) = (1 - s)^4 + s^4$ for any $s \in S$.

Example 3. Let the code be as in Example 2 but we admit a more wide class of channels with Markov binary noise (see Fig. 1). If the initial probabilities of the noise are defined as $P(0) = s_2/(s_1 + s_2)$, $P(1) = s_1/(s_1 + s_2)$ with $0 < s_1 \leqq s_2$, then this channel reduces to the previous one iff $s_2 = 1 - s_1$. The state space S of this channel is a subset of E_2. Obviously,

$$\pi(s) = P_4((0, 0, 0, 0) \mid s) + P_4((1, 1, 1, 1) \mid s) = \frac{s_2(1 - s_1)^3}{s_1 + s_2} + \frac{s_1(1 - s_2)^3}{s_1 + s_2}$$

for every $s = (s_1, s_2) \in S$.

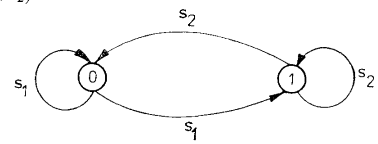

Fig. 1.

2. MAXIMUM LIKELIHOOD ESTIMATORS

The repeat request statistic of size n is available at the output of any channel with ARQ decoding as soon as n subsequent codewords have been received and it always provides certain evidence concerning the channel state. However, the inference can be successful only if the state remains unaltered during transmission of a relatively large number of codewords. (In our idealized model we suppose $s \in S$ to be constant during all the transmission period but this obviously can be avoided in practical application.) Moreover, in order to specify the degree of accuracy attainable in estimating the channel state or in order to impose reasonable optimality criteria, it is necessary to know what is the stochastic dependence between channel states and the statistic. As one can see from Theorem 1, our assumptions result into very simple probabilistic description of stochastic relations between states and the statistic. Due to this, in solving the statistical problem of estimation the unknown channel state $s \in S$ on the basis of data represented by a repeat request statistic of size n, we can afford to start with a family of probability distributions $(p^n(\cdot \mid s) : s \in S)$ given in (2).

We see from (2) that the repeat request count statistic T, whose family is explicitely given in (3), is sufficient for any statistical inference concerning the channel state. This fact is to be respected in reading this paper.

We shall say that an estimator $\sigma : \{0, 1, \ldots\} \to S$ of the unknown channel state $s \in S$ based on the repeat request count is maximum likelihood, if

$$p^*(t \mid \sigma(t)) = \max_{s \in S} p^*(t \mid s) .$$

In this paper we restrict ourselves to maximum likelihood estimation of communication channel states due to certain universality of this estimation. Indeed, according to invariance principle (theorem of Zehna, see e.g. [5], p. 223), the above considered maximum likelihood estimator determines maximum likelihood estimators of all communication parameters reasonably functionally related to the channel state.

We introduce the notation

$$I(p, p') = p \log \frac{p}{p'} + (1 - p) \log \frac{1 - p}{1 - p'}, \qquad p, p' \in [0, 1]$$

for the well-known measure of divergence between distributions $(p, 1 - p)$ and $(p', 1 - p')$ on a dichotomy and we shall consider $I(p, p') \geqq 0$ with equality iff $p = p'$ as a known fact.

Theorem 2. An estimator $\sigma : \{0, 1, \ldots\} \to S$ is maximum likelihood iff

$$I\left(\frac{n}{t + n}, \pi(\sigma(t))\right) = \min_{s \in S} I\left(\frac{n}{t + n}, \pi(s)\right) \tag{4}$$

so that the maximum likelihood estimator is defined as a solution $s = \sigma(t)$ to the equation

$$\pi(s) = \frac{n}{t + n} \tag{5}$$

as soon as the solution exists in S.

Proof. Since the logarithmic function is increasing, σ is maximum likelihood iff

$$\log p^*(t \mid \sigma(t)) = \max_{s \in S} \log p^*(t \mid s),$$

i.e. iff

$$\log \binom{n + t - 1}{t} + n \log \pi(\sigma(t)) + t \log (1 - \pi(\sigma(t))) =$$

$$= \max_{s \in S} \left[\log \binom{n + t - 1}{t} + n \log \pi(s) + t \log (1 - \pi(s)\right],$$

i.e. iff

$$\frac{n}{t + n} \log \pi(\sigma(t)) + \frac{t}{t + n} \log (1 - \pi(\sigma(t))) =$$

$$= \max_{s \in S} \left[\frac{n}{t + n} \log \pi(s) + \frac{t}{t + n} \log (1 - \pi(s))\right],$$

i.e. iff

$$\frac{n}{t + n} \log \frac{1}{\pi(\sigma(t))} + \frac{t}{t + n} \log \frac{1}{1 - \pi(\sigma(t))} =$$

$$= \min_{s \in S} \left[\frac{n}{t + n} \log \frac{1}{\pi(s)} + \frac{t}{t + n} \log \frac{1}{1 - \pi(s)}\right].$$

Substracting from the both sides the entropy of the probability distribution $(n/(t+n), t/(t+n))$ we see that σ is maximum likelihood iff

$$\frac{n}{t+n}\log\frac{\frac{n}{t+n}}{\pi(\sigma(t))}+\frac{t}{t+n}\log\frac{\frac{t}{t+n}}{1-\pi(\sigma(t))}=$$

$$=\min_{s\in S}\left[\frac{n}{t+n}\log\frac{\frac{n}{t+n}}{\pi(s)}+\frac{t}{t+n}\log\frac{\frac{t}{t+n}}{1-\pi(s)}\right],$$

which coincides with (4).

Suppose now that S is finite and let $s_1, s_2, \ldots, s_k$ be elements chosen arbitrarily from subsets of states possessing common value of π, where $\pi(s_1) < \pi(s_2) < \ldots$ $\ldots < \pi(s_k)$. This will be called monotone enumeration of elements in S.

Lemma 1. Let S be finite and let $s_1, s_2, \ldots, s_k$ be a monotone enumeration of elements in S. Then there exist real numbers $0 = y_0 < \pi(s_1) < y_1 < \ldots < y_{k-1} <$ $< \pi(s_k) < y_k = 1$ such that the maximum likely value $\sigma(t) = s_i$ iff $n/(t+n) \in$ $\in (y_{i-1}, y_i]$.

Proof. Let us consider the functions

$$\varphi_i(y) = I(y, \pi(s_i)), \quad y \in [0, 1], \quad i = 1, 2, \ldots, k.$$

In view of (1) these functions are well-defined and continuous in $[0, 1]$. Obviously, there exists a disjoint decomposition $[0, 1] = D_1 + D_2 + \ldots + D_k + \Delta$, where

$$\varphi_i(y) - \varphi_j(y) < 0 \quad \text{for all} \quad j \neq i \quad \text{on} \quad D_i, \quad i = 1, 2, \ldots, k,$$

and where $\varphi_i(y) = \varphi_j(y)$ for some $i \neq j$ on Δ. Moreover, D_i are open in the closed interval $[0, 1]$. Since

$$\varphi_i(y) - \varphi_j(y) = y \log\frac{\pi(s_j)(1-\pi(s_i))}{\pi(s_i)(1-\pi(s_j))} + \log\frac{1-\pi(s_j)}{1-\pi(s_i)}$$

is linear, Δ must be a finite subset of $[0, 1]$ and each D_i is an open subinterval of the interval $[0, 1]$. Now, since

$$\varphi_i(\pi(s_i)) - \varphi_j(\pi(s_i)) = -I(\pi(s_i), \pi(s_j)) < 0 \quad \text{for all} \quad j \neq i,$$

we see that each interval D_i is non-empty and $\pi(s_i) \in D_i$.

Next we prove $\sup D_i \leqq \inf D_l$ for $i < l \leqq k$. If $\inf D_l < \sup D_i$ then either there exists $y_0 \in D_i \cap D_l$ for which

$$\varphi_l(y_0) - \varphi_j(y_0) < 0 \quad \text{for all} \quad j \neq l$$

$$\varphi_i(y_0) - \varphi_j(y_0) < 0 \quad \text{for all} \quad j \neq i$$

or $\sup D_l < \inf D_i$. In the first case one easily obtains $\varphi_l(y_0) < \varphi_i(y_0) < \varphi_l(y_0)$ while in the second case $\pi(s_l) < \pi(s_i)$. Neither of these consequences is possible.

Finally $\sup D_i = \inf D_{i+1}$ for $i < k$ since in the opposite case $\sup D_i < \inf D_{i+1}$, which implies that Δ contains infinite subset $(\sup D_i, \inf D_{i+1})$. Analogical reason yields $\inf D_1 = 0$, $\sup D_k = 1$. Thus we have proved that $D_i = (y_{i-1}, y_i)$ and $\Delta \subset \{y_0, y_1, \dots, y_k\}$, where $y_i = \sup D_i = \inf D_{i+1}$, $i = 1, 2, \dots, k-1$, $y_0 = \inf D_1 = 0$, $y_k = \sup D_k = 1$. Thus we have proved

$$\varphi_i(y) = \min_j \varphi_j(y) \quad \text{for} \quad y \in (y_{i-1}, y_i]\,, \quad i = 1, 2, \dots, k\,.$$

In view of (4), this and the fact that for every $s \in S$ there exists $s_j \in S$ with $\pi(s) = \pi(s_j)$ complete the proof of Lemma 1.

Lemma 2. Let us consider independent sampling from a geometrical family of probability distributions with parameter $1 - \theta$ and let $t = t_1 + \dots + t_n$ be the corresponding sufficient statistics. Then the Bayes test of a priori equiprobable simple hypothesis $\theta = \pi(s_i)$ against the simple alternative $\theta = \pi(s_{i+1})$ is given by

$$\text{accept } \theta = \pi(s_i) \ \text{ iff } \ \frac{n}{t+n} \in [0, y_i] \ \text{ and accept } \ \theta = \pi(s_{i+1}) \ \text{ iff } \ \frac{n}{t+n} \in (y_i, 1]\,.$$

Proof. Repeating the proof of Theorem 2 with $S = \{s_i, s_{i+1}\}$ we find that the Bayes test is given by

$$\text{accept } \theta = \pi(s_i) \ \text{ iff } \ \varphi_i\left(\frac{n}{t+n}\right) \leqq \varphi_{i+1}\left(\frac{n}{t+n}\right),$$

$$\text{accept } \theta = \pi(s_{i+1}) \ \text{ iff } \ \varphi_i\left(\frac{n}{t+n}\right) > \varphi_i\left(\frac{n}{t+n}\right).$$

Since it follows from what has been said in the proof of Lemma 1 that $\varphi_i(y) \leqq \varphi_{i+1}(y)$ holds iff $y \lesseqgtr y_i$ respectively, Lemma 2 is proved.

Lemma 3. Let Bayes test of a hypothesis $\theta = \theta_1$ against alternative $\theta = \theta_2$ based on independent observations $t_1, t_2, \dots, t_n$ each of which is supposed to be drawn from a family $p(\cdot \mid \theta)$ be of the form

$$\text{accept } \theta = \theta_1 \ \text{ iff } \ t \leqq y \ \text{ and accept } \ \theta = \theta_2 \ \text{ iff } \ t > y\,,$$

where $t = t(t_1, \dots, t_n)$ is a statistic. Let us further consider probability distribution $p^*(t \mid \theta)$ induced by the statistic t and joint probability distribution $p^n(t_1, \dots, t_n(\theta) = p(t_1 \mid \theta) \dots p(t_n \mid \theta)$. Then

$$\lim_{n\to\infty} \left(\sum_{t \leqq y} p^*(t \mid \theta_2) \right)^{1/n} = \lim_{n\to\infty} \left(\sum_{t > y} p^*(t \mid \theta_1) \right)^{1/n} = \lambda(p(\cdot \mid \theta_1), p(\cdot \mid \theta_2)),$$

where

$$\lambda(P, P') = \inf_{0<\alpha<1} \sum_t \left(\frac{P(t)}{P'(t)} \right)^\alpha P'(t) = \lambda(P', P). \tag{6}$$

Proof. This Lemma has been proved in [4] for the particular statistic

$$t = \sum_{i=1}^{n} \log \frac{p(t_i \mid \theta_1)}{p(t_i \mid \theta_2)}.$$

Since two Bayes tests of a fixed pair of hypotheses in a fixed statistical model differ only on sample sets of zero probability, the above mentioned proof easily extends to any statistic.

It is seen from Theorem 2 and Lemma 1 that the maximum likelihood estimator $\sigma : \{0, 1, \ldots\} \to S$ may depend on the parameter n of the family of distributions (3). Hence we write σ_n instead of σ when convenient. A basic question connected with maximum likelihood estimator σ_n is whether or under what conditions concerning S or mapping (1) it is consistent, i.e. $\sigma_n(t_1 + \ldots + t_n) \to s$ a.s. $[p^\infty(\cdot \mid s)]$ and what is the rate of convergence of probability $p^*(\{t : |\sigma_n(t) - s| > \varepsilon\} | s)$ to zero.

Theorem 3. Let S be open subset of a compact set in E_r and let $\pi : S \to (0, 1)$ be one-to-one and continuous. Then the maximum likelihood estimator σ_n is consistent.

Proof. If the conditions given above are fulfiled, LeCam's general conditions of consistency, as stated in Theorem 5.3.1 of [5], hold quite evidently except, perhaps, the condition that for each $s \neq s^* \in S$ there exists a neighbourhood $U(s^*) \subset S$ such that

$$\sum_{t=0}^{\infty} p(t \mid s) \inf_{s' \in U(s^*)} \left[\log \frac{p(t \mid s)}{p(t \mid s')} \right] > -\infty.$$

Since, however, this sum is for any $U(s^*)$ greater than or equal

$$\sum_{t=0}^{\infty} p(t \mid s) \log p(t \mid s) = \sum_{t=0}^{\infty} \pi(s)(1 - \pi(s))^t \log \pi(s)(1 - \pi(s))^t =$$

$$= \log \pi(s) + \frac{\pi(s)}{1 - \pi(s)} \log(1 - \pi(s)) > -\infty,$$

this condition is satisfied too.

The law-of-large-numbers based proof of consistency used below applies also to the continuous case considered in Theorem 3 above, provided continuity of $\pi^{-1}: :(0,1) \to S$ is supposed.

Theorem 4. If S is finite, then necessary and sufficient condition for consistency of the maximum likelihood estimator is the one-to-one property of the mapping $\pi : S \to (0, 1)$. If this is satisfied, then $p^*(\{t : \sigma_n(t) \neq s\} \mid s) = \lambda(s)^{n+o(n)}$ for all $s \in S$, where $\lambda(s) \in (0, 1)$ is defined as follows

$$\lambda(s) = \begin{cases} \lambda(p(\cdot \mid s_1), p(\cdot \mid s_2)) & \text{if } s = s_1 \\ \max\left[\lambda(p(\cdot \mid s_i), p(\cdot \mid s_{i+1})), \lambda(p(\cdot \mid s_i), p(\cdot \mid s_{i-1}))\right] & \text{if } s = s_i, \ 1 < i < k \\ \lambda(p(\cdot \mid s_k), p(\cdot \mid s_{k-1})) & \text{if } s = s_k, \end{cases}$$

where $s_1, s_2, \ldots, s_k$ is the monotone enumeration of (all) elements in S.

Proof. Since the expectation of the geometric probability distribution $p(\cdot \mid s)$ is $\pi(s)/(1 - \pi(s))$, the strong law of large numbers implies $t/n \to \pi(s)/(1 - \pi(s))$ a.s. $[p^\infty(\cdot \mid s)]$ so that $n/(t + n) \to \pi(s)$ a.s. $[p^\infty(\cdot \mid s)]$. This together with Lemma 1 implies consistency of σ_n provided π is one-to-one mapping. If $\pi(s) = \pi(s^*)$ for some $s \neq s^* \in S$, then it is easy to construct σ_n in such a way that s^* is not contained in the target space $\sigma_n(\{0, 1, \ldots\}) \subset n = 1, 2, \ldots$, at all.

To prove the second part of Theorem 4, observe first that, due to Lemma 1,

$$p^*(\{t : \sigma_n(t) \neq s_1\} \mid s_1) = \sum_{n/(t+n) > y_1} p^*(t \mid s_1),$$

$$p^*(\{t : \sigma_n(t) \neq s_i\} \mid s_i) = \sum_{n/(t+n) > y_i} p^*(t \mid s_i) + \sum_{n/(t+n) \leqq y_{i-1}} p^*(t \mid s_i),$$

$$p^*(\{t : \sigma_n(t) \neq s_k\} \mid s_k) = \sum_{n/(t+n) \leqq y_{k-1}} p^*(t \mid s_k).$$

By Lemma 2, the criterion below the first sum represents an acceptance criterion in the Bayes test of hypothesis $\theta = \pi(s_1)$ against alternative $\theta = \pi(s_2)$, where $1 - \theta$ is a parameter of geometrical distribution. Thus, by Lemma 3, $p^*(\{t : \sigma_n(t) \neq s_1\} \mid s_1) = \lambda(s_1)^{n+o(n)}$, where $\lambda(s_1) = \lambda(p(\cdot \mid s_1), p(\cdot \mid s_2))$. Analogically, the criteria below sums in the second row represent acceptance criteria in testing $\theta = \pi(s_i)$ against $\theta = \pi(s_{i+1})$ or $\theta = \pi(s_{i-1})$ against $\theta = \pi(s_i)$ respectively. Thus, again by Lemma 3,

$$p^*(\{t : \sigma_n(t) \neq s_i\} \mid s_i) =$$
$$= \lambda(p(\cdot \mid s_i), p(\cdot \mid s_{i+1}))^{n+o(n)} + \lambda(p(\cdot \mid s_i), p(\cdot \mid s_{i-1}))^{n+o(n)} =$$
$$= (\max\left[\lambda(p(\cdot \mid s_i), p(\cdot \mid s_{i+1})), \lambda(p(\cdot \mid s_i), p(\cdot \mid s_{i-1}))\right])^{n+o(n)} = \lambda(s_i)^{n+o(n)}.$$

The same argument applies also to the third sum and the Theorem is proved.

Note that, using (6), the parameters of exponential convergence $\{\lambda(s) : s \in S\}$ can be found as explicit functions of parameters $\pi(s_1), \ldots, \pi(s_k)$, but the formulas are too heavy to be stated here.

3. EXAMPLES

Let us consider the situation described in Example 1, i.e., let us consider an arbitrary binary memoryless symmetric channel with a per-bit error probability $s \in S \subset \subset (0, 10^{-3})$ and arbitrary iterated code with parameters N_1, N_2 such that $N_1 N_2 < < 10^3$. Then $\pi(s)$ can well be approximated on S by $\pi(s) = 1 - N_1 N_2 s$.

If S is open set then, by Theorem 2,

$$(7 \qquad \sigma_n(t) = \frac{1}{N_1 N_2} \cdot \frac{t}{t + n} = \frac{1}{N_1 N_2} \cdot \frac{\bar{t}}{\bar{t} + 1}\,, \quad \text{where} \quad \bar{t} = \frac{t}{n} = \frac{t_1 + \ldots + t_n}{n}\,,$$

is the maximum likelihood estimator of the unknown channel state. Since in this case the conditions of Theorem 3 are satisfied too, (7) is consistent estimator.

If $S = \{s_1, \ldots, s_k\}$, equations $I(y, 1 - N_1 N_2 s_i) = I(y, 1 - N_1 N_2 s_{i+1})$, yield points

$$y_i = \frac{\log \dfrac{s_i}{s_{i+1}}}{\log \dfrac{(1 - N_1 N_2 s_{i+1})\, s_i}{(1 - N_1 N_2 s_i)\, s_{i+1}}}\,, \quad i = 1, 2, \ldots, k - 1$$

defined in Lemma 1. Thus, by Lemma 1, $\sigma_n(t) = s_i$ iff $n/(t + n) \in (y_{i-1}, y_i]$, i.e.

$$(8) \qquad \sigma_n(t) = s_i \quad \text{iff} \quad \frac{1}{y_i} - 1 \leqq \bar{t} < \frac{1}{y_{i-1}} - 1$$

is the maximum likelihood estimator of $s \in S$. Since the conditions of Theorem 4 are satisfied, (8) is consistent estimator for which the probability of incorrect identification of a true state tends to zero exponentially with increasing n.

Contrary to this situation, if we consider the channel of Example 3, we see that, for general $S \subset E_2$, $\pi(s) = \pi(s_1, s_2)$ is not one-to-one mapping. Multiplicity of solutions to equation (5) is evident, but neither of the too many existing maximum likelihood estimators is consistent. If however one of the parameters is known in advance (or a functional dependence between them is supposed), then our theory can be successfully applied.

(Received November 5, 1979.)

[1] I. Vajda: Estimation of parameters of a transmission line based on a repeat request statistics (in Czech). Dálkový přenos dat. ČSVTS, Praha 1979.

[2] J. Pužman: Coding for decision feedback communication systems. Trans. 6th Prague Conf. on Inf. Theory, ..., Academia, Praha 1973.

[3] L. N. Kanal, A. R. K. Sastry: Models for channels with memory and their application to error control. Proc. of IEEE *66* (1978), 724—744.

[4] H. Chernoff: A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann. Math. Stat. *23* (1952), 493—501.

[5] S. Zacks: The Theory of Statistical Inference. J. Wiley, New York 1971.

Ing. Igor Vajda, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.