KYBERNETIKA — VOLUME 19 (1983), NUMBER 1

ESTIMATION OF PROBABILISTIC NOISE MODELS BASED ON FILTRATION OF SAMPLE NOISE SEQUENCES

IGOR VAJDA

Optimum as well as suboptimum estimation of standard multistate noise models of information theory is introduced. The suboptimum estimation is based on a suitably filtered noise sequence only. It is also shown that a formula for probability of errorless transmission which has formerly been derived within the framework of a two-state model remains to be applicable independently on the number of states considered.

1. INTRODUCTION

In most communication situations, the information-theoietic concept of a channel reduces into simpler concept of a noise. This is connected with possibility to represent the channel output messages as sums $a + X$ of input messages $a = (a_1, \ldots, a_n)$ with some random sequences $X = (X_1, \ldots, X_n)$. In digital communication both a_i and X_i can be supposed to take on values from a binary alphabet $A = \{0, 1\}$ and the + can be interpreted as a coordinate mod 2 binary addition. Thus a piactically oriented information theory is concerned with binary random suquences $\{X_i : i = 0, \pm 1, \ldots\}$ representing probabilistic noise models (cf. [1]).

In this paper we consider a quite common class of probabilistic noise models. Models are simple and, on the other hand, realistic for a relatively wide variety of real communication channels.

Strictly speaking, each class of noise models is a set of probabilities P on the natuial σ-algebra of subsets of the set A^I of all binary sequences $\{x_i : i = 0, \pm 1, \ldots\}$, i.e. the set of all possible realizations of above considered binary random sequences. The aim of this paper is to propose statistical methods for estimation of probabilities P within the class considered. The estimation is based on samples $(x_1, \ldots, x_n) \in A^n$ from the noise random sequences, called sample noise sequences. As the title indicates, filtration of a sample sequence $(x_1, \ldots, x_n)$ is an important mid-step in our method of turning out this sequence into a concrete estimate P.

2. PROBABILISTIC NOISE MODELS

The class of noise models we shall deal with can be described as follows. Consider the stochastic automaton shown in Fig. 1 with a finite or infinite state space $S = \{g, b, g_1, \ldots, g_N\}$ (i.e. with $N \in \{0, 1, \ldots, \infty\}$ where the limit specifications for $N =$

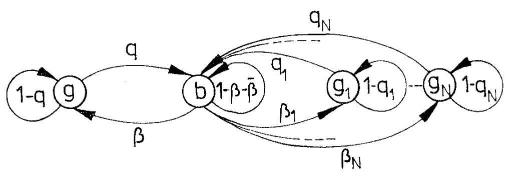

Fig. 1. ($\bar{\beta} = \beta_1 + \ldots + \beta_N$) .

$= 0$ or $N = \infty$ are perhaps clear from Fig. 1) and with an output function

(1) $$\Phi(s) = \begin{cases} 1 & \text{if} \quad s = b \\ 0 & \text{if} \quad s \in \{g, g_1, \ldots, g_N\} . \end{cases}$$

The $2N + 2$ parameters of the automaton are supposed to be arbitrary real numbers from the open interval $(0, 1)$ (state transition probabilities) satisfying condition $\beta + \bar{\beta} \leqq 1$.

Let $\{S_i : i = 0, \pm 1, \ldots\}$ be a stationary Markov chain with the state space S and with the transition probabilities matrix

$$\begin{array}{c} \\ b \\ g \\ g_1 \\ \vdots \\ g_N \end{array} \begin{bmatrix} b & g & g_1 & \ldots & g_N \\ 1-\beta-\bar{\beta} & \beta & \beta_1 & \ldots & \beta_N \\ q & 1-q & 0 & \ldots & 0 \\ q_1 & 0 & 1-q_1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ q_N & 0 & 0 & \ldots & 1-q_N \end{bmatrix} .$$

Obviously, this chain is irreducible and aperiodic (i.e. ergodic) with

(2) $$\mathsf{P}(S_i = s) = \begin{cases} (1 + \beta/q + \bar{\beta}/\bar{q})^{-1} & \text{if} \quad s = b \\ \beta/q(1 + \beta/q + \bar{\beta}/\bar{q})^{-1} & \text{if} \quad s = g \quad \text{where} \quad \bar{\beta}/\bar{q} = \sum_{i=1}^{N} \beta_i/q_i . \\ \beta_i/q_i(1 + \beta/q + \bar{\beta}/\bar{q})^{-1} & \text{if} \quad s = g_i . \end{cases}$$

The random noise $\{X_i : i = 0, \rightarrow 1, \ldots\}$ is now defined as the output random sequence of the automaton, i.e.

(3) $$X_i = \Phi(S_i) , \quad i = 0, \rightarrow 1, \ldots .$$

Thus a general noise model which we consider is defined by a $(2N + 2)$-parameter automaton. The N is called the order of the model. It follows from (1) that $g, g_1 \ldots \ldots, g_N$ are good (errorless) states while b is a bad (error) state. Consequently the parameter $p = \mathsf{P}(X_i = 1) = \mathsf{P}(S_i = b)$ (see (2)) means a bit error rate of the noise under consideration. An obvious non-triviality condition on infinite-order models is thus

$$\frac{\bar{\beta}}{\bar{q}} = \sum_{i=1}^{\infty} \frac{\beta_i}{q_i} < \infty .$$

In what follows we restrict ourselves to models satisfying this condition. In fact, all practically interesting models are satisfying a stronger condition, namely

$$\frac{\bar{\beta}}{\bar{q}} \ll \frac{\beta}{q} < \infty \quad \text{(cf. Section 4)}. \tag{4}$$

Note that the zero-order models have first been considered by Gilbert [2]. These models are uniquely described by just two parameters $q, \beta \in (0, 1)$. It is obvious that the errors appear here in clusters (bursts of errors) of average length $1/\beta$. The average length $1/\beta$ is not, however, a proper measure of a burst memory since the actual scattering of errors over the noise sequence depends on the parameter q too. A proper measure of memory we obtain if we take the χ^2-divergence of two consecutive noise digits (see [3]), which is given in the zero-order models by $(1 - q - \beta)^2$. Thus (in a subclass of practically interesting models non-negative)

$$\chi = 1 - q - \beta \tag{5}$$

can be considered as a noise memory measure (if $1 - q - \beta < 0$ we replace (5) by $|1 - q - \beta|$). Since it follows from (2) and (5) that

$$q = p(1 - \chi), \quad \beta = (1 - p)(1 - \chi), \tag{6}$$

each zero-order model can be reparametrized by (p, χ) instead by the less intuitively appealing (q, β).

An obvious advantage of the zero-order over higher-order models is that they allow reconstruction of the sequence of states $(S_1, \ldots, S_k)$ from the observable noise sequence $(X_1, \ldots, X_k)$ (i.e. the statistic $(X_1, \ldots, X_k) = (\Phi(S_1), \ldots, \Phi(S_n))$ of the state sequence $(S_1, \ldots, S_k)$ is sufficient for the unknown parameters; such models are sometimes called unifilar, see [4]). A disadvantage is that, in real sample noise sequences, the burst of errors form block of mixtures of 1's and 0's inserted between long error-free gaps while, in zero-order models, the bursts are simply blocks of pure 1's. Consequently, the correlation properties of zero-order noise sequences are too far from the observed reality.

This disagreement can be supressed by passing to higher-order (multistate) models defined by Fig. 1. These models have been introduced into the literature by Fritchman [5]. According to [6, 7, 8], such models realistically describe HF radio channels and, according to [9], telephone channels too. A general discussion of practical applicability of these models can be found in [1]. Note that the infinite-order version of the Fritchman's model is just a theoretical abstraction introduced for the purpose of the present paper.

As soon as $N \geqq 1$, the noise random sequence is neither Markov chain nor the statistic $(X_1, \ldots, X_k)$ is sufficient for the parameters of the model. This means that to find out an asymptotically optimum solution of the estimation problem introduced above is a difficult task.

In the next section we describe an asymptotically optimum solution based on a side information additional to the statistic $(X_1, \ldots, X_n)$. In Section 4 we describe an asymptotically suboptimum solution of this problem based on the statistic $(X_1, \ldots, X_n)$ only.

3. ASYMPTOTICALLY OPTIMUM ESTIMATION

From now on we consider an arbitrary fixed model of order $1 \leqq N \leqq \infty$ with a Markov chain of states $\{S_i : i = 0, \pm 1, \ldots\}$. Next definitions are based on the idea that the noise is "bursting" as long as S_i takes on values from the set $\{b, g_1, \ldots, g_N\}$. Thus $g \in S$ is the only truly good state of the model generating long blocks of 0's (of average size $1/q$) while $g_i \in S$ are semigood states generating short blocks of 0's (of average size $1/q_i$) frequently alternated by 1's (inside error bursting time intervals). Thus our basic intuition can formally be expressed by the assumption that $1/q$ is much (say two orders) greater than q_i for all $i = 1, \ldots, N$. Taking into account our experience with bursting error measurements both in HF and telephone channels, we can conclude that we shall stay well within limits of practical applicability of the theory if we simplify this assumption into the following form

$$q \lesssim 10^{-3}, \quad q_i \gtrsim 10^{-1} \quad \text{for all} \quad i = 1, \ldots, N. \tag{7}$$

It is to be noted, however, that neither (7) nor any similar assumption is formally needed for the theory developed below so that this theory applies to the arbitrary model under consideration.

The random sequence $\{Y_i : i = 0, \pm 1, \ldots\}$ defined by $Y_i = \Psi(S_i)$, where

$$\Psi(s) = \begin{cases} 1 & \text{if} \quad s \in \{b, g_1, \ldots, g_N\} \\ 0 & \text{if} \quad s = g, \end{cases} \tag{8}$$

is called burst indicator sequence.

Let us now consider a subsequence of the sequence $\{X_i : i = 0, \pm 1, \ldots\}$ of noise digits generated while S_i is continuing to stay inside the "bursting state" $\{b, g_1, \ldots, g_N)$ In other words, X_i belongs to the subsequence if $Y_i = 1$. Enumerate the elements X_i of the subsequence with $i > 0$ by $i_1, i_2, \ldots$ and the elements with $i \leqq 0$ by $i_0, i_{-1}, \ldots$ and denote $\{X_{i_j} : j = 0, \pm 1, \ldots\}$ by $\{Z_j : j = 0, \pm 1, \ldots\}$. It follows from this definition that the new sequence is generated by a reduced version of the automaton of Fig. 1 shown in Fig. 2 with the same output function (1) as that generating the noise sequence $\{X_i : i = 0, \pm 1, \ldots\}$. Thus if

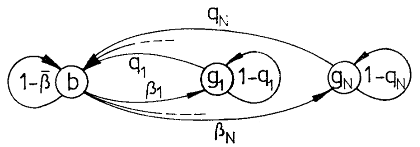

Fig. 2.

$\{S_i^* : i = 0, \pm 1, \ldots\}$ is a stationary Markov chain defined by the transition matrix of Fig. 2 then, obviously,

$$Z_i = \Phi(S_i^*) \quad \text{for} \quad i = 0, \pm 1, \ldots. \tag{9}$$

Let us call the $\{Z_i : i = 0, \pm 1, \ldots\}$ a burst noise sequence. Since, vaguely speaking, the burst noise sequence is the noise $\{X_i : i = 0, \pm 1, \ldots\}$ filtered by a random filter $\{Y_i : i = 0, \pm 1, \ldots\}$, we use a product notation

$$\{Z_i : i = 0, \pm 1, \ldots\} = \{X_i : i = 0, \pm 1, \ldots\} \,.\, \{Y_i : i = 0, \pm 1, \ldots\}\,.$$

Let now $(Z_1, \ldots, Z_{m_n})$ with a random m_n be the burst noise subsequence of the noise sequence $(X_1, \ldots, X_n)$ and denote this subsequence symbolically by $(X_1, \ldots, X_n) \,.\, (Y_1, \ldots, Y_n)$. Under our assumptions it obviously holds

$$\lim_{n \to \infty} m_n = \infty \quad \text{a.s.} \tag{10}$$

For each $k, l \in \{0, 1\}$ we shall consider statistics $k(Y_1, \ldots, Y_n)$, $kl(Y_1, \ldots, Y_n)$ or $k((X_1, \ldots, X_n) \,.\, (Y_1, \ldots, Y_n))$, $kl((X_1, \ldots, X_n) \,.\, (Y_1, \ldots, Y_n))$ denoting absolute frequencies of digits k and absolute frequencies of consecutive digits (k, l) in the sequence $(Y_1, \ldots, Y_n)$ or $(Z_1, \ldots, Z_{m_n}) = (X_1, \ldots, X_n) \,.\, (Y_1, \ldots, Y_n)$ respectively.

Define now estimators

$$q_n = 1 - \frac{00(Y_1, \ldots, Y_n)}{0(Y_1, \ldots, Y_n)}\,, \quad \beta_n = 1 - \frac{11(Y_1, \ldots, Y_n)}{1(Y_1, \ldots, Y_n)}\,, \tag{11}$$

$$\bar{q}_n = 1 - \frac{00((X_1, \ldots, X_n) \,.\, (Y_1, \ldots, Y_n))}{0((X_1, \ldots, X_n) \,.\, (Y_1, \ldots, Y_n))}\,,$$

$$\bar{\beta}_n = 1 - \frac{11((X_1, \ldots, X_n) \,.\, (Y_1, \ldots, Y_n))}{1((X_1, \ldots, X_n) \,.\, (Y_1, \ldots, Y_n))}\,,$$

of unknown parameters q, β or parametric functions $\bar{q}$, $\bar{\beta}$ of the model under consideration.

Theorem 1. The estimators $q_n, \beta_n, \bar{q}_n, \bar{\beta}_n$ defined in (11) are strongly consistent, i.e.

$$\lim_{n\to\infty} q_n = q\,,\quad \lim_{n\to\infty} \beta_n = \beta \quad \text{a.s.} \tag{12}$$

$$\lim_{n\to\infty} \bar{q}_n = \bar{q}\,,\quad \lim_{n\to\infty} \bar{\beta}_n = \bar{\beta} \quad \text{a.s.} \tag{13}$$

Proof. (i) By definition (11) and by definition of the burst indicator

$$\lim_{n\to\infty} q_n = 1 - \lim_{n\to\infty} \frac{00(Y_1, \ldots, Y_n)}{0(Y_1, \ldots, Y_n)} =$$

$$= 1 - \lim_{n\to\infty} \frac{\dfrac{00(\Psi(S_1), \ldots, \Psi(S_n))}{n}}{\dfrac{0(\Psi(S_1), \ldots, \Psi(S_n))}{n}} =$$

$$= 1 - \frac{\lim\limits_{n\to\infty} \dfrac{1}{n} \sum\limits_{i=1}^{n-1} I_{(g,g)}(S_i, S_{i+1})}{\lim\limits_{n\to\infty} \dfrac{1}{n} \sum\limits_{i=1}^{n} I_g(S_i)}\,.$$

Since $\{S_i : i = 1, 2, \ldots\}$ is an ergodic chain, the limits in numerator and denominator exist and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n-1} I_{(g,g)}(S_i, S_{i+1}) = \mathsf{P}(S_1 = g, S_2 = g) \quad \text{a.s.}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I_g(S_i) = \mathsf{P}(S_1 = g) \quad \text{a.s.}$$

Since

$$\frac{\mathsf{P}(S_1 = g, S_2 = g)}{\mathsf{P}(S_1 = g)} = \mathsf{P}(S_2 = g \mid S_1 = g) = 1 - q \quad \text{(see Fig. 1)}\,,$$

the first relation in (12) is proved. The second relation in (12) can be proved analogically.

(ii) In view of (10) it holds

$$\lim_{n\to\infty} q_n = 1 - \lim_{n\to\infty} \frac{\dfrac{00(Z_1, \ldots, Z_{m_n})}{m_n}}{\dfrac{0(Z_1, \ldots, Z_{m_n})}{m_n}}$$

so that, by (9), it holds for $\bar{g} = \{g_1, \ldots, g_N\}$

$$\lim_{n\to\infty} \bar{q}_n = 1 - \frac{\lim\limits_{n\to\infty} \frac{1}{m_n} \sum\limits_{i=1}^{m_n} I_{\bar{g}\times\bar{g}}(S_i^*, S_{i+1}^*)}{\lim\limits_{n\to\infty} \frac{1}{m_n} \sum\limits_{i=1}^{m_n} I_{\bar{g}}(S_i^*)}.$$

Since the chain $\{S_i^* : i = 1, 2, \ldots\}$ is ergodic, the limits in numerator and denominator exist and

$$\lim_{n\to\infty} \frac{1}{m_n} \sum_{i=1}^{m_n} I_{\bar{g}\times\bar{g}}(S_i^*, S_{i+1}^*) = \mathsf{P}(S_1^* \in \bar{g}, S_2^* \in \bar{g}) \text{ a.s.}$$

$$\lim_{n\to\infty} \frac{1}{m_n} \sum_{i=1}^{m_n} I_{\bar{g}}(S_i^*) = \mathsf{P}(S_1^* \in \bar{g}) \quad \text{a.s.}$$

Since

$$\mathsf{P}(S_1^* \in \bar{g}, S_2^* \in \bar{g}) = \sum_{i=1}^{N} \mathsf{P}(S_2^* \in \bar{g} \mid S_1^* = g_i)\, \mathsf{P}(S_1^* = g_i) =$$

$$= \sum_{i=1}^{N} (1 - q_i)\, \mathsf{P}(S_1^* = g_i) \quad \text{(see Fig. 1)}$$

and

$$\mathsf{P}(S_1^* \in \bar{g}) = \sum_{i=1}^{N} \mathsf{P}(S_i^* = g_i),$$

we have proved that

$$\lim_{n\to\infty} \bar{q}_n = \frac{\sum\limits_{i=1}^{N} q_i\, \mathsf{P}(S_1^* = g_i)}{\sum\limits_{i=1}^{N} \mathsf{P}(S_i^* = g_i)} \quad \text{a.s.}$$

Since, further

$$\mathsf{P}(S_1^* = g_i) = \frac{\beta_i}{q_i}\left(1 + \frac{\bar{\beta}}{\bar{q}}\right)^{-1} \quad \text{(cf. (2))},$$

it holds by (2)

$$\lim_{n\to\infty} q_n = \frac{\bar{\beta}}{\bar{\beta}/\bar{q}} = \bar{q}$$

and the first relation in (13) is proved. The second relation in (13) can be proved analogically. □

Since the estimators (11) are in fact maximum likelihood estimators of the respective parameters, they are asymptotically optimum not only in the stated sense, but also in the sense of efficiency. Details are omitted here.

4. ASYMPTOTICALLY SUBOPTIMUM ESTIMATION

In the preceding section we have supposed an additional side information $(Y_1, \ldots, Y_n)$. Here we consider the problem of estimation of unknown parameters $\beta, q, \bar{\beta}, \bar{q}$ on the basis of noise $(X_1, \ldots, X_n)$ only.

Let us consider an arbitrary fixed natural number k. Define a filter $F: A^I \to A^I$ of the noise by $F(\{x_i : i = 0, \pm 1, \ldots\}) = (\{\hat{y}_i : i = 0, \pm 1, \ldots,\})$ where

$$\hat{y}_i = I_{(0,\infty)}(x_i + x_{i+1} + \ldots + x_{i+k}) \quad \text{(ordinary addition)}. \tag{14}$$

The filtered noise sequence

$$\{\hat{Y}_i : i = 0, \pm 1, \ldots\} = F(\{X_i : i = 0, \pm 1, \ldots\})$$

will be considered as an estimator of the unknown burst indicator $\{Y_i : i = 0, \pm 1, \ldots\}$

We shall say that a noise model for $\varepsilon < 0$ is ε-admissible if

$$q \leqq 1 - (1 - \varepsilon)^{1/k}, \quad q_i \geqq 1 - \varepsilon^{1/k} \quad \text{for} \quad i = 1, \ldots, N. \tag{15}$$

We see that for $10 \leqq k \leqq 50$ and for ε of order 10^{-2} the class of ε-admissible models contains all models satisfying (7) i.e. all practically interesting models.

Theorem 2. If the model is ε-admissible then $\mathsf{P}(\hat{Y}_i \neq Y_i) \leqq \varepsilon$ for all $i = 0, \pm 1, \ldots$

Proof. It holds

$$\mathsf{P}(\hat{Y}_i \neq Y_i) = \sum_{s \in S} \mathsf{P}(\hat{Y}_i \neq Y_i \mid S_i = s)\, \mathsf{P}(S_i = s) =$$

$$= \sum_{s \in \{b, g_1, \ldots, g_N\}} \mathsf{P}(\hat{Y}_i \neq 1 \mid S_i = s)\, \mathsf{P}(S_i = s) + \mathsf{P}(\hat{Y}_i \neq 0 \mid S_i = g)\, \mathsf{P}(S_i = g)$$

$$= \sum_{s \in \{b, g_1, \ldots, g_N\}} \mathsf{P}(\hat{Y}_i = 0 \mid S_i = s)\, \mathsf{P}(S_i = s) + (1 - \mathsf{P}(\hat{Y}_i = 0 \mid S_i = g))\, \mathsf{P}(S_i = g).$$

Since $S_i = b$ implies $X_i = \Phi(S_i) = 1$, it follows from (14) that $\mathsf{P}(\hat{Y}_i = 0 \mid S_i = b) = 0$. Further by (14)

$$\begin{aligned} \mathsf{P}(\hat{Y}_i = 0 \mid S_i = g_i) &= \mathsf{P}(X_i = \ldots = X_{i+k} = 0 \mid S_i = g_i) \\ &= \mathsf{P}(S_i = \ldots = S_{i+k} = g_i \mid S_i = g_i) \\ &= \mathsf{P}(S_1 = \ldots = S_k = g_i \mid S_0 = g_i) \\ &= (1 - q_i)^k \quad \text{(cf. Fig. 1)} \end{aligned}$$

and, analogically,

$$\mathsf{P}(\hat{Y}_i = 0 \mid S_i = g) = (1 - q)^k.$$

Thus we have proved

$$\mathsf{P}(\hat{Y}_i \neq Y_i) = \sum_{i=1}^{N} (1 - q_i)^k\, \mathsf{P}(S_i = g_i) + (1 - (1 - q)^k)\, \mathsf{P}(S_i = g).$$

Therefore

$$\mathsf{P}(\hat{Y}_i \neq Y_i) \leqq \max \{\max_i (1 - q_i)^k , 1 - 1 - q)^k\} .$$

The desired inequality now follows from the fact that (15) implies

$$\max \{\max_i (1 - q_i)^k, 1 - (1 - q)^k\} \leqq \varepsilon .$$

This result implies that if $10 \leqq k \leqq 50$, then the filtered noise estimates the unknown burst indicator in all practically interesting situations with an error frequency of order at most 10^{-2}. This order is essentially less if k is properly adapted to concrete noise sample sequences.

On the basis of Theorems 1 and 2 we propose the following estimators q_n, β_n, $\bar{q}_n$, $\bar{\beta}_n$ of the unknown parameters q, β, $\bar{q}$, $\bar{\beta}$:

(16) $$q_n = 1 - \frac{00(\hat{Y}_1, \ldots, \hat{Y}_n)}{0(\hat{Y}_1, \ldots, \hat{Y}_n)} , \quad \beta_n = 1 - \frac{11(\hat{Y}_1, \ldots, \hat{Y}_n)}{1(\hat{Y}_1, \ldots, \hat{Y}_n)}$$

$$\bar{q}_n = 1 - \frac{00((X_1, \ldots, X_n) \cdot (\hat{Y}_1, \ldots, \hat{Y}_n))}{0((X_1, \ldots, X_n) \cdot (\hat{Y}_1, \ldots, \hat{Y}_n))} ,$$

$$\bar{\beta}_r = 1 - \frac{11((X_1, \ldots, X_n) \cdot (\hat{Y}_1, \ldots, \hat{Y}_n))}{1((X_1, \ldots, X_n) \cdot (\hat{Y}_1, \ldots, \hat{Y}_n))}$$

with

(17) $$\hat{Y}_i = \begin{cases} I_{(0,\infty)}(X_i + \ldots + X_{i+k}) & \text{for } i = 1, \ldots, n - k \\ I_{(0,\infty)}(X_i + \ldots + X_n) & \text{for } i = n - k + 1, \ldots, n . \end{cases}$$

5. APPLICATION OF THE REDUCED MODEL KNOWLEDGE

The above described statistical procedure provides us with a limited knowledge of the model (only two of $2N + 2$ unknown parameters with two more parametric functions). But analytical treatment of some communication problems can be carried out with this limited knowledge equally well as if a complete knowledge was available. As an example mention here theoretical problems around the transmission rate in ARQ communication. All the theory is based here on the function

$$P_n(0) = \mathsf{P}((X_1, \ldots, X_n) = (0, \ldots, 0))$$

approximating $\mathsf{P}((X_1, \ldots, X_n) \in \mathscr{K})$ for linear codes $\mathscr{K} \subset A^n$ (see [10] and [11]). This function can be decomposed as follows

(18) $$P_n(0) = \sum_{s \in S} P_n(0 \mid s) \, \mathsf{P}(S_0 = s) ,$$

where

$$P_n(0 \mid s) = \mathsf{P}((X_1, \ldots, X_n) = (0, \ldots, 0) \mid S_0 = s)$$

and $\mathsf{P}(S_0 = s)$ is given by (2). It follows from Fig. 1 and from (3) that

$$P_n(0 \mid g) = (1 - q)^n$$

$$P_n(0 \mid b) = (1 - q)^{n-1} + \sum_{i=1}^{N} \beta_i (1 - q_i)^{n-1}$$

$$P_n(0 \mid g_i) = (1 - q_i)^n .$$

This together with (2) and (18) yields

$$\left(1 + \frac{\beta}{q} + \frac{\bar{\beta}}{\bar{q}}\right) P_n(0) = \frac{\beta}{q}(1 - q)^n + \beta(1 - q)^{n-1} + \sum_{i=1}^{N} \beta_i (1 - q_i)^{n-1} + $$

$$+ \sum_{i=1}^{N} \frac{\beta_i}{q_i}(1 - q_i)^n = \frac{\beta}{q}(1 - q)^{n-1} + \sum_{i=1}^{N} \frac{\beta_i}{q}(1 - q_i)^{n-1}$$

Hence, in view of (4),

$$\frac{\frac{\beta}{q}}{1 + \frac{\beta}{q}}(1 - q)^{n-1} \leqq P_n(0) \leqq \frac{\frac{\beta}{q}}{1 + \frac{\beta}{q}}(1 - q)^{n-1} + $$

$$+ \frac{\frac{\bar{\beta}}{\bar{q}}}{1 + \frac{\beta}{q}}(1 - \min_i q_i)^{n-1}$$

Applying again (4) and (7) we obtain that in all practically interesting models $P_n(0) \doteq$ $\doteq (1 - q)^{n-1}$, i.e.

$$P_n(0) \doteq e^{-qn} . \tag{19}$$

Therefore $P_n(0)$ is essentially independent of the $2N - 1$ remaining model parameters as well as of the model order N.

Since the parameter q is estimated (through the procedure proposed in Section 4 or any other procedure) less comfortably than the noise error rate $p = \mathsf{P}(X_i = 1) =$ $= \mathsf{P}(S_i = b)$, (p is easily estimated through $p_n = 1(X_1, \ldots, X_n)/n$) it would be interesting to try to replace the parameter q by the parameter p and other additional parameters. This procedure is, however, meaningful only if the estimation of the additional parameters is simpler than the estimation of q itself. Fortunately, our experience with noise sequences from various concrete channels indicates that the parameter β is more stable than the parameter q when outer conditions of a given real channel vary at random. This fact together with the formula

$$p = \frac{1}{1 + \frac{\beta}{q} + \frac{\bar{\beta}}{q}} \doteq \frac{q}{\beta} \quad (\text{cf. } (2), (4), (7))$$

offer possibility to replace the parameter q in (19) by the product* $p\beta$. Moreover, it follows from (5) that for q satisfying (7) it holds $\beta \doteq 1 - \chi$ where χ is a burst memory introduced in Section 2. Thus (19) can be replaced by

$$P_n(0) \doteq e^{-p\beta n} \doteq e^{-p(1-\chi)n} . \tag{20}$$

Notice that, placing ourselves within the framework of the zero-order model, (20) would be obtained directly from (6) and (19). Since the interpretation of both parameters p and χ is intuitively appealing, the right-hand formula in (20) seems to be quite convenient for theoretical analysis of ARQ communication problems. This formula had in fact been already used in [10] but with a aless general reasoning than that presented in this paper.

A typical application of the formula (20) is as follows. If a noise source of a communication channel slowly fluctuates within a class of multistate sources with a stable χ, then the frequency of errorless transmission $P_n(0)$ fluctuates monotonically with average error frequencies $1(X_1, \ldots, X_n)$ or $p_n = 1(X_1, \ldots, X_n)/n$ and

$$P_n(0) \doteq e^{-(1-\chi).1(X_1,\ldots,X_n)}, \quad p_n \doteq \frac{\ln(1/P_n(0))}{n(1-\chi)} . \tag{21}$$

(Received April 22, 1982.)

REFERENCES

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

[2] E. N. Gilbert: Capacity of a burst-noise channel. Bell System Tech. J. *39* (1960), 1253—1366.

[3] I. Vajda: Theory of Information and Statistical Decision (in Slovak). Alfa, Bratislava 1982.

[4] T. Ericson: Unifilac sources and the source approximation problem. Ericson Techn. *28* (1972), 175—290.

[5] B. D. Fritchman: A binary channel characterization using partitioned Markov chains. IEEE Trans. Inform. Theory *IT-13* (1967), 221—227.

[6] S. Tsai: Evaluation of burst error correcting codes using a simple partitioned Markov chain model. IEEE Trans. Comm. *COM-21* (1973), 1031—1034.

[7] P. McManamon: HF Markov chain models and measured error averages. IEEE Trans. Comm. COM-18 (1970), 201—208.

* In view of the formula $p \doteq q/\beta$ we can say that above stated hypothesis means that the frequency q of transitions from the good state g to the bad state b of the channel is much more responsible for fluctuations of the average noise error rate p than the frequency β of transitions from the bad state b back so the good state g. The frequency β is usually between 0·7—0·9, i.e. of constant order while q is frequently fluctuating over two decadic orders.

[8] J. A. Picardi and P. I. Trafton: Error behavior of tactical HF data transmissions. Trans. IEEE Conf. on Commun. 4020—4025, Seatle 1973.
[9] K. Brayer: Characterization of the digital high speed AUTOVON channel. Final Report of the MITRE Corporation, MTR-2968, 1976.
[10] I. Vajda: Theoretical aspects of high-speed data transmission (in Czech). Přenos dat, ČSVTS, Brno 1981.
[11] I. Vajda: Output based estimation of communication channels. Kybernetika *16*(1980), 330—340.

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.