Problems of Control and Information Theory, Vol. 18 (4), pp. 197—221 (1989)

CONTRIBUTIONS TO A THEORY OF ORDERING FOR SEQUENCE SPACES[1]

R. AHLSWEDE, Z. ZHANG

(Bielefeld)

(Received March 4, 1988)

We continue the investigations of [1] and obtain first results concerning optimal orderings of sequences in two simple, but basic, cases:

a non-probabilistic model with active memory and
a probabilistic model with unlimited passive memory.

1. Introduction

In [1] we have presented ideas about a theory of ordering and established some primary but basic results for a non-probabilistic model of ordering sequences under constraints on the operations and the knowledge about the sequences.

The basic model considered in [1] can be formulated as follows. Suppose we have a box that contains at time $t=0$ β balls. We assume that the balls are labelled with numbers from $\mathscr{X}=\{1, \ldots, \alpha\}$. For simplicity, we say a ball "i" instead of a ball labelled by "i". Thus, the content or "state" of the box can be described by a multi-set $s_0=(s_0(1), \ldots, s_0(\alpha))$, where $s_0(i)$ is the number of i's in the box and $\sum_{i=1}^{\alpha} s_0(i)=\beta$. We denote a sequence of balls just by its labelling. An arbitrary n-length sequence of balls, say

$$x^n=(x_1, \ldots, x_n)\in\mathscr{X}^n,$$

enters the box iteratively. At time t, x_t enters just after a person, called the organizer, has pulled out a ball "y_t" from the box. Consequently, the state s_{t-1} should be changed to s_t. We call x^n the input sequence and $y^n=(y_1, \ldots, y_n)$ the output sequence. The organizer's strategy obeys the following rules.

[1] Presented at the International Colloquium on Coding Theory, organized by Osaka University and the Armenian Academy of Sciences, Osaka, June 24–28, 1988.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

(1) The organizer can output only objects which he has in the box. At any time $t(0<t\leq n)$ he can and has to output exactly one ball.

(2) The organizer's strategy may depend on some of his knowledge such as knowledge concerning time, natural order of the balls in the box, the input sequence and the output sequence before the current time.

The subclass of problems, which have been investigated most intensively in [1], is characterized by a triple (π, β, φ). It specifies that β objects from $\mathscr{X}$ fit into the box and that at any time t the organizer $\mathscr{O}$ knows the state of the box s_t, that he can see the incoming letters $x_t, x_{t+1}, \ldots, x_{t+\varphi-1}$ and that he still remembers (or can see) the output letters $y_{t-\pi}, y_{t-\pi+1}, \ldots, y_{t-1}$ when he outputs y_t.

The goal of $\mathscr{O}$ is to minimize the number of output sequences for a specified blocklength n.

We continue here the studies of [1] in two directions.

I. *Active memory*

In addition to the knowledge described by the triple (π, β, φ), which we term "passive memory", the organizer may have storage space, in which he can (and has to) feed a number from $\{0, 1, \ldots, \gamma-1\}$. He uses this storage to remember a certain amount of any information relevant to him. We speak of an active memory. It can be realized by a switching board with γ states. The organizer can turn the state of the board to any one of the states labelled by the numbers $0, 1, \ldots, \gamma-1$ based on his current knowledge and try to "remember" something.

Formally the new model is described by a quadruple $(\pi, \beta, \varphi, \gamma)$. It involves passive and additional active memory. In this notation the case of passive memory only, described by the triple (π, β, φ), can equivalently be described by the quadruple $(\pi, \beta, \varphi, 1)$. Here we study another extremal case, namely that of "pure" active memory that is, the case $(0, \beta, 0, \gamma)$.

We denote the set of the organizer's strategies by $F_\alpha^n(\beta, \gamma)$. Note that here a strategy is a pair of functions (f, g), where for state of the box s and state of the switching board z $f(s, z)$ gives the output and $g(s, z)$ gives the next state of the board. For initial state of the box s_0, initial state of the board z_0 and strategy (f, g) an input sequence x^n determines as output sequence $y^n = y^n(z_0, s_0, x^n, f, g)$. The set of all n-length output sequences under (f, g) is therefore

$$\mathscr{Y}^n(f, g) = \{y^n(z_0, s_0, x^n, f, g) : 0 \leqq z_0 \leqq \gamma-1,\ s_0 \in \mathscr{S},\ x^n \in \mathscr{X}^n\}, \tag{1.1}$$

if $\mathscr{S}$ denotes the set of all possible states of the box.

In accordance with the terminology of [1], where $v_\alpha(\pi, \beta, \varphi)$ was defined, we introduce now

$$N_\alpha^n(0, \beta, 0, \gamma) = \min\{|\mathscr{Y}^n(f, g)| : (f, g) \in F_\alpha^n(\beta, \gamma)\}, \tag{1.2}$$

$$\nu_\alpha(0, \beta, 0, \gamma) = \lim_{n \to \infty} \frac{1}{n} \log N_\alpha^n(0, \beta, 0, \gamma). \tag{1.3}$$

It is convenient to use the abbreviations

$$\bar{N}_\alpha(\beta, \gamma) = N_\alpha^n(0, \beta, 0, \gamma),\ \bar{\nu}_\alpha(\beta, \gamma) = \nu_\alpha(0, \beta, 0, \gamma). \tag{1.4}$$

These quantities are studied in Section 2.

The key new observation is that in case $\gamma = 2$ our strategy of [1] in case $(1, \beta, 0)$ can be simulated by using instead of the one letter knowledge of the past the active memory and that this is optimal.

Section 3, the rest of the paper, is devoted to an analysis of the probabilistic model mentioned under C in Section 2 of [1]. The objects or letters are here produced by a stochastic process, which in the simplest case is a sequence $(X_t)_{t=1}^\infty$ of i.i.d RV's with values in $\mathscr{X} = \{0, 1, \ldots, \alpha - 1\}$ and generic distribution P_X. In Information Theory this is also called a discrete, memoryless source. For a strategy f_n, which depends on the triple (π, β, φ), let $Y^n = Y_1 \ldots Y_n$ be the output sequence corresponding to $X^n = X_1 \ldots X_n$. Let $F_\alpha^n(\pi, \beta, \varphi, P_X)$ be the set of strategies restricted to blocklength n.

We use the "per letter" entropy $\frac{1}{n} H(Y^n)$ as performance criterion and define

$$\eta_\alpha(\pi, \beta, \varphi, P_X) = \lim_{n \to \infty} \min_{f_n \in F_\alpha^n(\pi, \beta, \varphi, P_X)} \frac{1}{n} H(Y^n). \tag{1.5}$$

This is the smallest mean entropy of the output process, which can be achieved by $\mathcal{O}$ with strategies based on his knowledge. It corresponds to the optimal rate $\nu_\alpha(\pi, \beta, \varphi)$ in our non-probabilistic model. Our new quantity is much harder to analyse.

We consider here the simplest (non-trivial) source, that is the binary symmetric source defined by $P_X(0) = P_X(1) = 1/2$, in the first non-trivial case $\beta = 2$. Further it is assumed that $\mathcal{O}$ knows the ∞-past and has no knowledge about the future.

We give a nice formula for $\eta_2(\infty, 2, 0, P_X)$.

2. Ordering with active memory

This section is entirely devoted to the proof of the following result.

Theorem 1

(a) $\bar{\nu}_2(\beta, 2) = \nu_2(1, \beta, 0)$

(b) $\nu_2(1, \beta, 0) = \log \Psi_\beta$,
where Ψ_β is the positive root of $\lambda^\beta - \lambda^{\beta-1} - 1 = 0$.

Whereas (b) repeats an earlier result, Theorem 6 of [1], (a) establishes a new and interesting connection. Since the organizer $\mathscr{C}$ can use the active memory to remember the last letter send, clearly

$$\bar{v}_2(\beta, 2) \leqq v_2(1, \beta, 0). \tag{2.1}$$

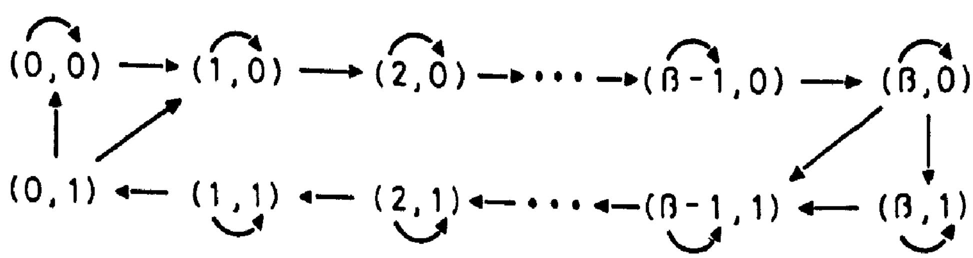
The issue is that in the present situation there is no better way to use the active memory. Actually we show that the strategy which we found in [1] for the case $(1, \beta, 0)$, is also optimal here. Of course now we have to exclude more possibilities and the proof is therefore somewhat more complicated than the one in [1]. However, it is based on similar ideas. We denote by $\bar{s}_t$ the pair (s_t, z_t), where s_t is the state of the box and z_t the state of the board at time t. z_t takes values in $\{0, 1\}$ and s_t can be assumed to count the number of 1's in the box.

We claim that the following strategy (f^*, g^*) is optimal:

$$f^*(s, 0) = 0 \quad \text{for} \quad s < \beta \quad \text{and} \quad f^*(s, 1) = 1 \quad \text{for} \quad s > 0;$$

$$f^*(0, z) = 0, \quad f^*(\beta, z) = 1, \quad \text{and} \quad g^* \equiv f^*. \tag{2.2}$$

Notice that it simply repeats the previous action, if this is possible. We can draw a state transition chart for this strategy. The 2 outgoing arrows correspond to the 2 possible inputs. Loops are included.



Denote by $\mathscr{M}^t(s, z)$ the set of all possible output sequences of length t achievable with $\bar{s}_t = (s, z)$ and some initial state. Since $\mathscr{M}^{t-1}(s, 0) \subset \mathscr{M}^{t-1}(s-1, 0)$ and $\mathscr{M}^{t-1}(s, 1) \subset \subset \mathscr{M}^{t-1}(s+1, 1)$, one readily verifies the following relations for $M^t(s, z) = |\mathscr{M}^t(s, z)|$

(i) $M^t(s, 0) = M^{t-1}(s-1, 0)$ for $2 \leqq s \leqq \beta$
(ii) $M^t(s, 1) = M^{t-1}(s+1, 1)$ for $0 \leqq s \leqq \beta - 2$
(iii) $M^t(0, 0) = M^{t-1}(0, 1) + M^{t-1}(0, 0)$;
$M^t(\beta, 1) = M^{t-1}(\beta, 0) + M^{t-1}(\beta, 1)$
(iv) $M^t(1, 0) = M^t(0, 0)$; $M^t(\beta - 1, 1) = M^t(\beta, 1)$
(v) $M^t(\beta - 1, 1) = M^t(1, 0)$.

Therefore we can conclude that

$$M^t(\beta - 1, 1) = M^{t-1}(\beta - 1, 1) + M^{t-\beta}(\beta - 1, 1) \tag{2.3}$$

and that $M^t(\beta - 1, 1) = \max\{M^t(s, z): 0 \leqq s \leqq \beta;\ z = 0, 1\}$.

In [1] a similar relationship was used to derive (b).

The equation $\lambda^\beta - \lambda^{\beta-1} - 1 = 0$ shows that for smaller β there is less compression of the sequence spaces.

We are going to prove that (f^*, g^*) is an optimal strategy. We begin with a simple observation.

Lemma 1. For every optimal strategy (f, g)

$$f(s, 0) \neq f(s, 1) \quad \text{for} \quad 1 \leqq s \leqq \beta - 1 .$$

Proof. Assume to the opposite that for some s in the specified range $f(s, 0) = f(s, 1)$. Then either $\{s' \mid 0 \leqq s' \leqq s\}$ or $\{s' \mid s \leqq s' \leqq \beta\}$ is a closed set of states, that is, starting from a state inside this set, we can never reach a state outside this set. This can be viewed as having a smaller box of size either s or $\beta - s$ and in any case of a size smaller than β. Proceeding inductively in β, we see that (f, g) cannot be optimal.

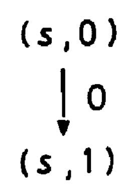
Lemma 2. For an optimal strategy (f, g) the condition

$$(f, g)(s, 0) = (0, x) \quad \text{for any} \quad s \geqq 1 \quad \text{and any} \quad x \in \{0, 1\} \tag{2.4}$$

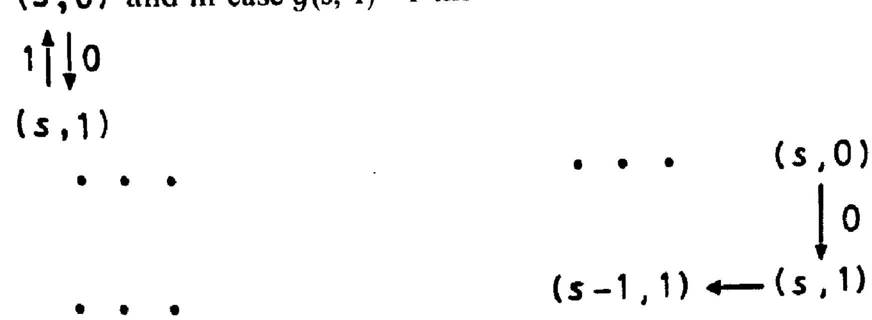
implies that there exists a directed path in the state transition chart from $(s, 1)$ to $(s, 0)$ of a length not exceeding $2s$.

Proof

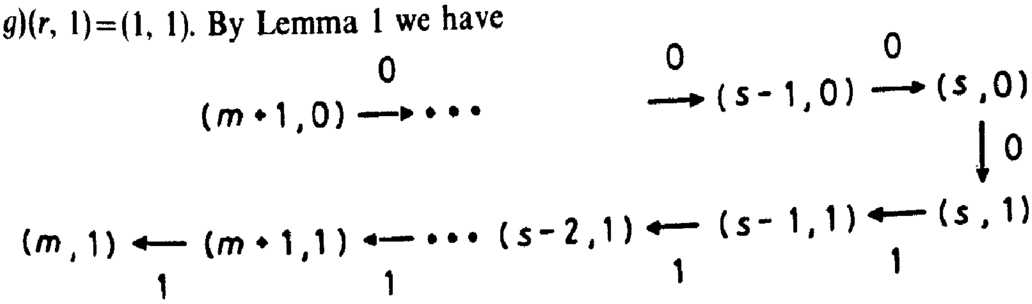
Case $x = 1$: Since $f(s, 0) = 0$ we have



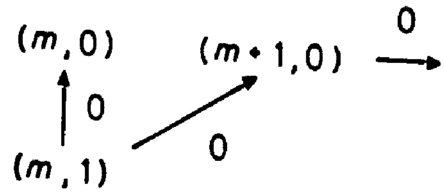
Further, since by Lemma 1 $f(s, 1) = 1$ we have in case $g(s, 1) = 0$ the desired path (s,0) and in case $g(s, 1) = 1$ the chart



Let now m be the smallest number such that for all r with $m + 1 \leqq r \leqq s$ $(f, g)(r, 1) = (1, 1)$. By Lemma 1 we have

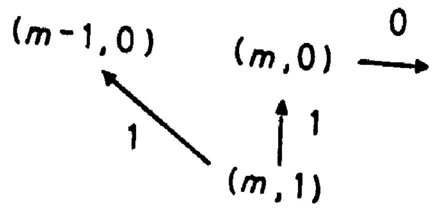


If $g(m, 1)=0$, then we get either



in case $f(m, 1)=0$

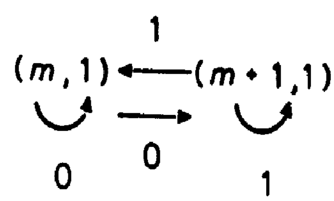
or



in case $f(m, 1)=1$.

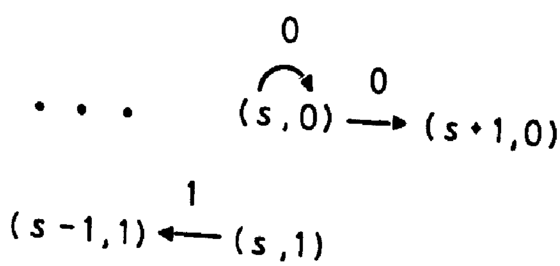
In both cases there is a path from $(s, 1)$ to $(s, 0)$ of a length not exceeding $2s$.

If $g(m, 1)=1$, then by definition of m necessarily $f(m, 1)=0$. This results in a chart



which cannot occur for an optimal strategy.

Case $x=0$: Either $g(s, 1)=0$ and we have a path of length 1 or $g(s, 1)=1$ and the left part of the chart



can be analyzed as in the previous case.

Our next result holds again for all optimal strategies.

Lemma 3. For $\beta \geqq 4$, any optimal strategy (f, g) satisfies in *case* $s=1$

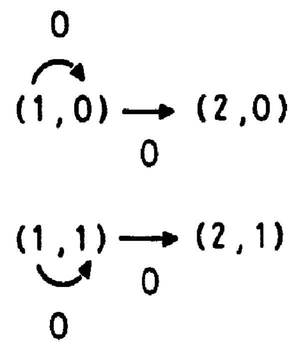
$$\begin{aligned} &\text{either} \quad (f, g)(1, 0)=(0, 0) \\ &\text{or} \quad (f, g)(1, 1)=(0, 1) \end{aligned} \tag{2.5}$$

and in *case* $s=\beta-1$

$$\begin{aligned} &\text{either} \quad (f, g)(\beta-1, 0)=(1, 0) \\ &\text{or} \quad (f, g)(\beta-1, 1)=(1, 1). \end{aligned} \tag{2.6}$$

Proof. By symmetry only one of the two cases has to be established. In case

$s=1$ our statement can be visualized by the following state transition chart:



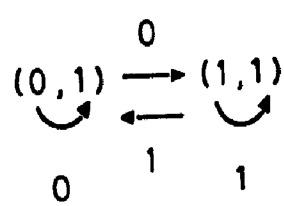
The number at an arrow, in this case the number 0, indicates the output object.

Suppose now that for an optimal strategy our claim for $s=1$ is false.

Case $(f, g)(1, 0)=(0, 1)$: By Lemma 1 thus $f(1, 1)=1$ and we are left with the subcases

(a) $(f, g)(1, 1)=(1, 1)$
(b) $(f, g)(1, 1)=(1, 0)$.

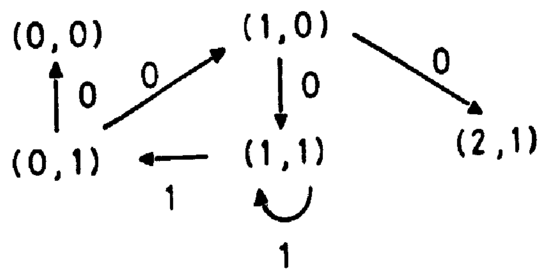
Ad (a): Since always $f(0, 1)=0$, necessarily $(f, g)(0, 1)=(0, x)$. However, if $x=1$, then a part of the chart is of the form



This leads to the relations

$$\begin{aligned} M^t(1, 1) &= M^{t-1}(1, 1)+M^{t-1}(0, 1) \\ &= M^{t-1}(1, 1)+M^{t-2}(1, 1)+M^{t-2}(0, 1) \\ &\geqq M^{t-1}(1, 1)+M^{t-2}(1, 1). \end{aligned}$$

Since the biggest root Ψ_β of $\lambda^\beta-\lambda^{\beta-1}-1=0$ is strictly decreasing in β and since $\beta\geqq 4\geqq 2$, $M^t(1, 1)$ grows to fast. This means that we must have $(f, g)(0, 1)=(0, 0)$ and a chart



From the cycle in this chart we derive the inequality

$$M^t(1, 1)\geqq M^{t-1}(1, 1)+M^{t-3}(1, 1),$$

which results in a rate of growth not smaller than $\log \Psi_3$. Case (a) cannot occur.

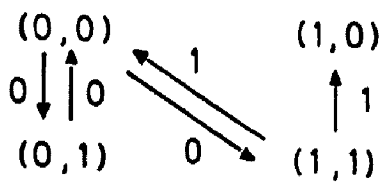
In case (b) we have the following 3 possibilities

(b1) $(f, g)(0, 0)=(0, 0)$
(b2) $(f, g)(0, 0)=(0, 1)$ and $(f, g)(0, 1)=(0, 1)$
(b3) $(f, g)(0, 0)=(0, 1)$ and $(f, g)(0, 1)=(0, 0)$.

One readily checks with the state transition charts that in cases (b1) and (b2) inequalities of the form

$$M^t(.,\ .)\geqq M^{t-1}(.,\ .)+M^{t-3}(.,\ .)$$

hold, whereas in case (b3) we have



Therefore
$$M^t(0, 0)=M^{t-1}(0, 1)+M^{t-1}(1, 1),$$
$$M^{t-1}(0, 1)=M^{t-2}(0, 0)$$
and
$$M^{t-1}(1, 1)\geqq M^{t-2}(0, 0).$$

In all these cases the rate of growth exceeds $\log \Psi_4$ and (b) cannot occur for an optimal strategy.

We are left with

Case $f(1, 0)=1$: By Lemma 1 $f(1, 1)=0$ and by our supposition necessarily $g(1, 1)=0$. Now just notice that the previous case $(f, g)(1, 0)=(0, 1)$ and the present case $(f, g)(1, 1)=(0, 0)$ differ only in the labelling of states in the active memory.

Lemma 4. For $\beta=2, 3$ there are optimal strategies for which (2.5) and (2.6) hold. (f^*, g^*) is optimal for $\beta=2$.

Proof. Inspection of the previous proof shows that in case $\beta=3$ a strategy violating (2.5) or (2.6) cannot be better than (f^*, g^*), for which (2.5) and (2.6) hold. The case $\beta=2$ requires a more refined analysis. Here by Lemma 1 the optimal f is up to the labelling of the states in the active memory unique, namely, $f=f^*$.

We go again through the cases of the proof of Lemma 3. If (2.5) does not hold, then necessarily

$$(f, g)(1, 0)=(f^*, g)(1, 0)=(0, 1) \tag{2.7}$$

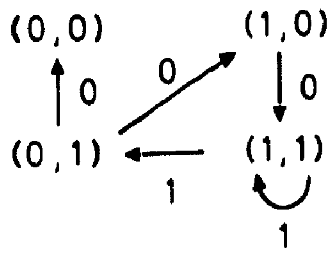
and we are left with the alternatives

(a) $(f, g)(1, 1)=(1, 1)$
(b) $(f, g)(1, 1)=(1, 0)$.

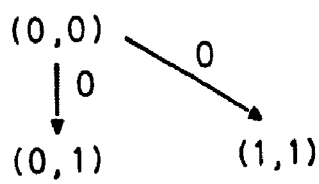
Ad (a): We have seen that in the case $(f, g)(0, 1)=(0, 1)$

$$M^t(1, 1) \geqq M^{t-1}(1, 1)+M^{t-2}(1, 1)$$

and therefore (f^*, g^*) is not superseded. We are left with the case $(f, g)(0, 1)=(0, 0)$ and the chart



Subcase $g(0, 0)=1$:

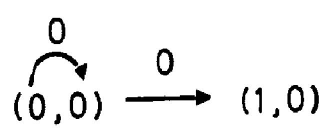


The output space of n-sequences contains all sequences with 0-strings of an even length. The number $E(n)$ of these sequences satisfies the recursion

$$E(n)=E(n-1)+E(n-2). \tag{2.8}$$

This is the familiar relation for (f^*, g^*), which is therefore again not defeated.

Subcase $g(0, 0)=0$:

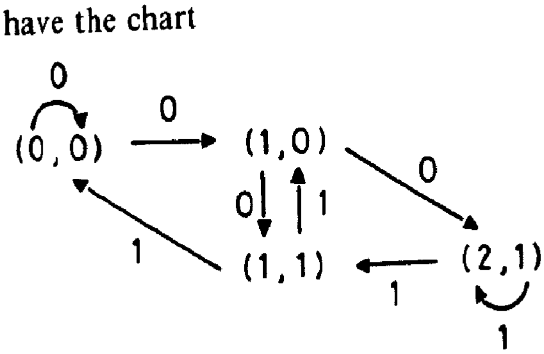


Here all output sequences with 0-strings of length $\geqq 2$ and arbitrary 1-strings occur. Their number is bigger than $E(n)$.

Ad (b): We distinguish the cases

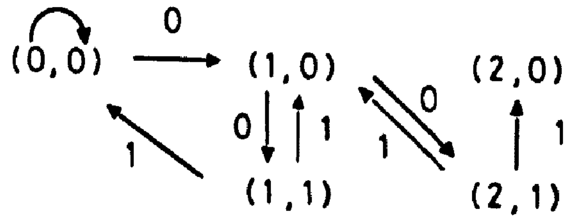
(A)	$(f, g)(0, 0)=(0, 0)$	and	$(f, g)(2, 1)=(1, 1)$
(B)	$(f, g)(0, 0)=(0, 0)$	and	$(f, g)(2, 1)=(1, 0)$
(C)	$(f, g)(0, 0)=(0, 1)$	and	$(f, g)(0, 1)=(0, 1)$
(D)	$(f, g)(0, 0)=(0, 1)$	and	$(f, g)(0, 1)=(0, 0)$.

Case (A): We have the chart

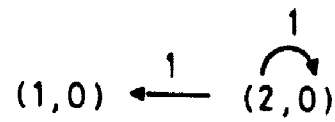


where the set $\mathscr{A}(n)$ of n-sequences with arbitrary 1-strings and 0-strings of length ≥ 2 can be produced. Clearly $|\mathscr{A}(n)| > E(n)$ and this case is excluded.

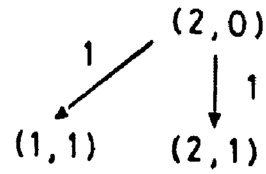
Case (B): We have the chart



and subcases (B_1):



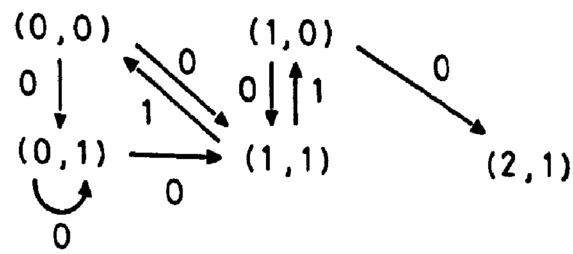
and (B_2):



In the first subcase one can produce $\mathscr{B}_1(n)$, the set of sequences with arbitrary 0-strings and arbitrary 1-strings interrupted by 01-strings of length ≥ 1, called "gates".

In the second subcase one can produce $\mathscr{B}_2(n)$, the set of n-sequences with arbitrary 0-strings, 1-strings of even length, 1-strings of even (≥ 2) length followed by a 0, and with 01-strings as gates between these 3 types of strings.

Case (C): We have the chart



Subcase (C_1): $(f, g)(2, 1) = (1, 1)$.

The output space contains $\mathscr{C}_1(n)$, the set of n-sequences with 10-strings between 1-strings and 0-strings of arbitrary length.

Subcase (C_2): $(f, g)(2, 1) = (1, 0)$.

Here we have two more cases.

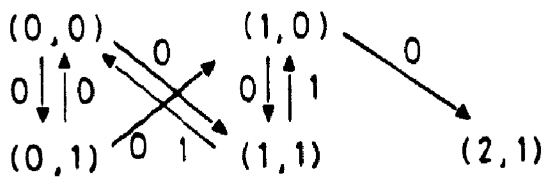
(C_{21}): $(f, g)(2, 0) = (1, 0)$.

$\mathscr{C}_{21}(n)$ = set of n-sequences with arbitrary 0-strings, 1-strings of length ≥ 2 and 10-strings as gates.

(C_{22}): $(f, g)(2, 0) = (1, 1)$.

$\mathscr{C}_{22}(n)$ = set of n-sequences with arbitrary 0-strings, 1-strings of even length, 1-strings of odd length followed by one 0, and 10-strings as gates.

Case (D): We have the chart



Subcase (D$_1$): $(f, g)(2, 1) = (1, 1)$.

Subcase (D$_2$): $(f, g)(2, 1) = (1, 0)$.

Finally we give now the bounds on cardinalities of sets. We observe that

$$|\mathscr{B}_1(n)| = |\mathscr{C}_1(n)| > |\mathscr{C}_{21}(n)|.$$

Further, considering in $\mathscr{C}_{21}(n)$ only 01-strings of length 1 as gates and replacing them by a 11-string of length 1 we get all n-sequences with arbitrary 0-strings and 1-strings of even length. Therefore $|\mathscr{C}_{21}(n)| \geqq E(n)$. Clearly, the same map gives also $|\mathscr{B}_2(n)| \geqq$ $\geqq E(n)$.

Similarly, replacing 10 by 11 we get $|\mathscr{C}_{22}(n)| \geqq E(n)$.

Finally, for subcase (D$_1$) (and similarly for (D$_2$)) there are transitions between strings 10, $1x$, $1y$, 01 and $0z$, where x is a 0-string of odd length, y is a 0-string of even length and z is an arbitrary 1-string. It can be shown that $|D_1(n)|, |D_2(n)| \geqq E(n)$.

Our next and main result describes a class of strategies, which includes an optimal strategy. We shall see that (f^*, g^*) is best within this class and therefore an optimal strategy.

Lemma 5. For $\beta \geqq 4$ there is an optimal strategy, say (f, g), satisfying either

$$(f, g)(s, 0) = (0, 0) \quad \text{and} \quad (f, g)(s, 1) = (1, 1) \tag{2.9}$$

for all $2 \leqq s \leqq \beta - 2$ or

$$(f, g)(s, 0) = (1, 0) \quad \text{and} \quad (f, g)(s, 1) = (0, 1) \tag{2.10}$$

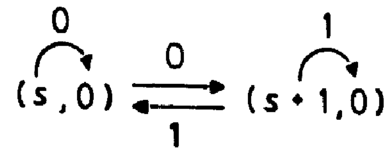
for all $2 \leqq s \leqq \beta - 2$.

This result says that the state transition chart of some optimal strategy has the form:

For $2 \leqq s \leqq \beta - 2$, either

$$\begin{matrix} (s,0)\rightarrow \\ \leftarrow(s,1) \end{matrix} \quad \text{or} \quad \begin{matrix} \leftarrow(s,0) \\ (s,1)\rightarrow \end{matrix}.$$

Proof. First notice that validity of (2.9) or (2.10) for $2 \leqq s \leqq \beta - 2$ implies validity of (2.9) for all $2 \leqq s \leqq \beta$ or validity of (2.10) for all $2 \leqq s \leqq \beta$, because we have situations like



and a rate log 2.

Now, by symmetry it suffices to consider the case $f(s, 0) = 0$. This and Lemma 1 leave us with the cases

(a) $(f, g)(s, 0) = (0, 0)$, $(f, g)(s, 1) = (1, 1)$.
(b) $(f, g)(s, 0) = (0, 0)$, $(f, g)(s, 1) = (1, 0)$.
(c) $(f, g)(s, 0) = (0, 1)$, $(f, g)(s, 1) = (1, 1)$.
(d) $(f, g)(s, 0) = (0, 1)$, $(f, g)(s, 1) = (1, 0)$.

Clearly (f^*, g^*) has property (2.9).

We want to prove that an optimal strategy satisfies (a). We need to check only that in other cases the strategies cannot be better than (f^*, g^*).

We consider (b) first. There are the following 2 subcases:

(b1) $(f, g)(s-1, 0) = (0, 0)$
(b2) $(f, g)(s-1, 0) = (0, 1)$.

For (b1) the following set of states is a closed set:

$$\Delta = \{(s-1, 0), \quad (u, z) \quad \text{with} \quad s \leqq u \leqq \beta\}.$$

From the induction hypothesis, this leads to the conclusion that for one of the $M^t(\,.\,,\,.\,)$ assigned to a state from this set

$$M^t(\,.\,,\,.\,) = M^{t-1}(\,.\,,\,.\,) + M^{t-\beta-s+1}(\,.\,,\,.\,).$$

The strategy cannot be optimal.

In case (b2) by Lemma 2 there exists a path from $(s-1, 0)$ to $(s-1, 1)$ of length not greater than $2(s-1)$. Therefore there exists a cycle from $(s-1, 1)$ to itself of length not greater than $2s \leqq \beta$. It is easy to prove that

$$M^t(s-1, 1) = M^{t-1}(s-1, 1) + M^{t-2s}(s-1, 1),$$

which shows that this strategy cannot be optimal.

For case (c) there are two possibilities:

(c1) $(f, g)(s-1, 0) = (0, 0)$
(c2) $(f, g)(s-1, 1) = (0, 1)$.

(c2) is obviously poor. We study only subcase (c1). If $\beta=2s$ we consider the part of the states (s', z) with $s'\geqq s$. Then case (c2) is just case (b). We have already proved that such a strategy cannot be optimal. If $\beta\geqq 2s+1$, then using Lemma 2 we can prove that an inequality

$$M^t(\,.\,,.\,)\geqq M^{t-1}(\,.\,,.\,)+M^{t-\beta}(\,.\,,.\,)$$

will hold. This is just what we need.

Case (d). There are two possibilities:

case (d1) $(f,g)(s-1,0)=(0,0)$
case (d2) $(f,g)(s-1,1)=(0,1)$.

For subcase (d1), the strategy is definitely poor, because of the existence of a cycle of length 3 from $(s-1, 0)$ to itself. Subcase (d2) is similar to subcase (c1). An inequality

$$M^t(\,.\,,.\,)\geqq M^{t-1}(\,.\,,.\,)+M^{t-\beta}(\,.\,,.\,)$$

can be derived, and the strategy cannot be better than (f^*, g^*).

Finally, to prove the claimed optimality of (f^*, g^*) the only thing left is to determine the form of the optimal strategy in Lemmas 3, 4 and 5 for $s=0, 1, \beta-1$ and β.

It is easy to check that in case $(f,g)(1,1)=(1,0)$ or $(f,g)(\beta-1,0)=(0,1)$, we have inequalities

$$M^t(1,0)\geqq M^{t-1}(1,0)+M^{t-\beta+1}(\beta-1,1)$$

and

$$M^t(\beta-1,1)\geqq M^{t-1}(\beta-1,1)+M^{t-\beta}(1,0).$$

This shows that such a strategy is poorer than (f^*, g^*). Therefore we must have $(f,g)(1,1)=(1,1)$ and $(f,g)(\beta-1,0)=(0,0)$. It is also readily seen that the form of such an optimal strategy at $s=0$ and $s=\beta$ is the same at that of (f^*, g^*). The proof of the theorem is complete.

3. A first result in a probabilistic model

Here we determine $\eta_\alpha(\pi, \beta, \varphi, P_X)$, which is defined in (1.5), in one of the simplest non-trivial cases, that is,

$$\alpha=2,\quad \pi=\infty,\quad \beta=2,\quad \varphi=0\quad\text{and}\quad P_X(0)=P_X(1)=\frac{1}{2}. \tag{3.1}$$

Since the main results, Theorem 2 and Theorem 3, are stated in terms of concepts which arise during the analysis, we state them in their natural contexts.

We need more notation.

The set of all strategies for blocklength n is simply denoted by F^n.

We will use two kinds of states of the box. The first one is the number of 1's in the box after the last ball X_t has entered the box. We denote this state of the box by S_t. It takes its values in $\{0, 1, 2\}$. Another kind of state of the box is the number of 1's in the box after Y_{t-1} left and before X_t entered the box. This quantity will be denoted by $\hat{S}_t$. It takes values in $\{0, 1\}$. Both, S_t and $\hat{S}_t$, are random variables. Clearly

$$S_t = \hat{S}_t + X_t. \tag{3.1'}$$

After a strategy $f = (f_1, f_2, \ldots)$ has been fixed and $Y^{t-1} = Y_1 \ldots Y_{t-1}$ is the output sequence before time t, then

$$Y_t = f_t(S_t, Y^{t-1}) \tag{3.2}$$

is the output at time t.

After X_{t+1} has entered the box it moves to the state S_{t+1}, where

$$S_{t+1} = S_t - Y_t + X_{t+1}, \quad \hat{S}_{t+1} = S_t - Y_t. \tag{3.3}$$

Conditional on $Y^{t-1} = y^{t-1}$ $\hat{S}_t$ has a distribution

$$P(\hat{S}_t = \varepsilon \mid Y^{t-1} = y^{t-1}) = P_\varepsilon; \quad \varepsilon = 0, 1. \tag{3.4}$$

Since X_t is independent of $(\hat{S}_t, Y^{t-1})$ by (3.1)

$$P(S_t = 0 \mid Y^{t-1} = y^{t-1}) = \frac{P_0}{2}$$

$$P(S_t = 1 \mid Y^{t-1} = y^{t-1}) = \frac{1}{2}$$

$$P(S_t = 2 \mid Y^{t-1} = y^{t-1}) = \frac{P_1}{2}. \tag{3.5}$$

If $S_t = 0$ (resp. 2), then $\mathcal{O}$ has to send a 0 (resp. 1). Therefore at time t strategies can only differ on the domain $\{(1, y^{t-1}) : y^{t-1} \in \mathcal{Y}^t\}$. How do we find a good strategy? We are guided by the idea to minimize the conditional entropy $H(Y_t \mid Y^{t-1} = y^{t-1})$ and we define therefore:

$f = (f_1, f_2, \ldots)$ is locally optimal at t, if in the notation of (3.4)

$$f_t(1, y^{t-1}) = \begin{cases} 0 & \text{if} \quad P_0 \geq p_1 \\ 1 & \text{if} \quad p_0 < p_1 \end{cases}. \tag{3.6}$$

For $t = 1$ we use the letters p and $q = \bar{p}$ instead of p_0 and p_1, that is,

$$P(\hat{S}_1 = 0) = p, \quad P(\hat{S}_1 = 1) = q. \tag{3.7}$$

We allow all initial distributions (p, q), but by symmetry there is no loss of generality, if we always assume

$$p \leqq q. \tag{3.8}$$

We call a strategy $f = (f_1, f_2, \ldots)$ *normal*, if it is locally optimal for every $t = 1, 2, \ldots$.

Our first result can now be stated.

Theorem 2. Under assumptions (3.1) for all p the normal strategy is optimal.

The proof is based on 5 lemmas. They concern estimates on entropies of random variables involving also mixed 1 step strategies $g(r)$, $0 \leqq r \leqq 1$, for which 1 is send with probability r and 0 is send with probability $1-r$. For the initial value p and any r, $0 \leqq r \leqq 1$, let us consider now those strategies which use $g(r)$ at the first step and subsequently, for $t \geqq 2$, follow the normal strategy.

The entropy of the thus produced output process Y^n depends on n, r and p. This justifies the notation

$$H_n^*(p; r) = H(Y^n). \tag{3.9}$$

Notice that by (3.8) $g(1)$ is locally optimal in the first step.

We present and prove now the 5 lemmas, which make statements about $H_n^*(p; r)$. The last lemma says that we shall follow the locally optimal strategy also in the first step, if we use it subsequently. Thus Theorem 2 follows inductively.

In the sequel we frequently use the notation

$$\bar{\mu} = 1 - \mu \quad \text{for} \quad 0 \leqq \mu \leqq 1. \tag{3.10}$$

Lemma 1. $H_n^*(p; r)$ is convex ($\cap$) in p.

Proof. By our definitions the conditional probabilities $\Pr(Y^n = y^n | \hat{S}_1 = \varepsilon)$ do not depend on p. The unconditional probabilities

$$\Pr(Y^n = y^n; P) = \Pr(Y^n = y^n | \hat{S}_1 = 0)p + \Pr(Y^n = y^n | \hat{S}_1 = 1)\bar{p} \tag{3.11}$$

are therefore linear in p and thus

$$\Pr(Y^n = y^n; \lambda p + \bar{\lambda} p') = \lambda \Pr(Y^n = y^n; p) + \bar{\lambda} \Pr(Y^n = y^n; p'). \tag{3.12}$$

Convexity of $H_n^*(p; r)$ then follows from the convexity of the entropy function.

Lemma 2. $\lambda H_n^*(p;\ r) + \bar{\lambda} H_n^*(p';\ r) \geqq H_n^*(\lambda p + \bar{\lambda} p'; r) + \lambda h(p) + \bar{\lambda} h(p') - h(\lambda p + \bar{\lambda} p')$, where h is the binary entropy function.

Proof. By (3.11) $H_n^*(p; r)$ can be written in the form $H(pP + \bar{p}Q)$ for two distribution P and Q. The inequality can therefore be restated as

$$h(\lambda p + \bar{\lambda} p') - \lambda h(p) - \bar{\lambda} h(p') \geqq H(\lambda(pP + \bar{p}Q) + \bar{\lambda}(p'P + \bar{p}'Q)) - \\ - \lambda H(pP + \bar{p}Q) - \bar{\lambda} H(p'P + \bar{p}'Q).$$

The expression to the left can be interpreted as a mutual information $I(J \wedge K)$, where $\Pr(J=1)=\lambda$, $\Pr(J=2)=\bar{\lambda}$ and

$$\Pr(K=1|J=1)=p, \quad \Pr(K=2|J=1)=\bar{p},$$
$$\Pr(K=1|J=2)=p', \quad \Pr(K=2|J=2)=\bar{p}'.$$

Postposing to the "channel" $\begin{pmatrix} p & \bar{p} \\ p' & \bar{p}' \end{pmatrix}$ the "channel" $\begin{pmatrix} P \\ Q \end{pmatrix}$ results in the "channel" $\begin{pmatrix} pP+\bar{p}Q \\ p'P+\bar{p}'Q \end{pmatrix}$, which for the input variable J induces the output variable L. Now $I(J \wedge L)$ is the expression to the right of the inequality, which is thus shown to be a special case of the data-processing inequality.

Lemma 3. $H_n^*(p;1)$ is monotone increasing in $[0, 1/2]$.

Proof. We proceed by induction in n. Since $H_1^*(p;1)=h\left(\frac{p}{2}\right)$ and since $\frac{d}{dr}h(r)=\log\frac{1-r}{r}$, we get

$$\frac{dH_1^*(p;1)}{dp}=\frac{1}{2}\log\frac{2-p}{p}\geqq 0 \quad \text{for} \quad p\in[0, 1/2]$$

and the case $n=1$ is established.

By Lemma 1 we know that $H_n^*(p;1)$ is convex. It suffices therefore to show that

$$\frac{d}{dp}H_n^*(p;1)|_{p=0.5}\geqq 0. \tag{3.13}$$

We use now and also later a basic recursion. For the underlying strategy

$$\begin{aligned} H_n^*(p;1) &= H(Y_1)+H(Y_n,\ldots,Y_2|Y_1) \\ &= h\left(\frac{p}{2}\right)+H(Y_n,\ldots,Y_2|Y_1=0)\Pr(Y_1=0) \\ &\quad + H(Y_n,\ldots,Y_2|Y_1=1)\Pr(Y_1=1) \\ &= h\left(\frac{p}{2}\right)+H_{n-1}^*(0;1)\frac{p}{2}+H_{n-1}^*\left(\frac{1}{2-p};0\right)\left(1-\frac{p}{2}\right) \end{aligned}$$

and therefore

$$H_n^*(p;1)=h\left(\frac{p}{2}\right)+\frac{p}{2}H_{n-1}^*(0;1)+\left(1-\frac{p}{2}\right)H_{n-1}^*\left(\frac{1-p}{2-p};1\right). \tag{3.14}$$

Hence

$$\frac{d}{dp}H_n^*(p;1)=\frac{1}{2}\log 3+\frac{1}{2}H_{n-1}^*(0;1)-\frac{1}{2}H_{n-1}^*\left(\frac{1}{3};1\right)$$

$$+\frac{3}{4}\frac{d}{dp}H_{n-1}^*\left(\frac{1-p}{2-p};1\right)_{|p=0.5} \qquad (3.15)$$

Using (3.14) again and the induction hypothesis we obtain

$$\frac{d}{dp}H_{n-1}^*\left(\frac{1-p}{2-p};1\right)_{|p=0.5}$$

$$=-\frac{1}{(1.5)^2}\left[\frac{1}{2}H_{n-2}^*(0;1)-\frac{1}{2}H_{n-2}^*\left(\frac{2}{5};1\right)+\frac{1}{2}\log 5\right.$$

$$\left.-\frac{1}{2}\frac{d}{dp}\left(H_{n-2}^*\left(\frac{1}{3-p};1\right)\right)_{|p=0.5}\right]\geqq$$

$$\geqq-\frac{2}{9}H_{n-2}^*(0;1)+\frac{2}{9}H_{n-2}^*\left(\frac{2}{5};1\right)-\frac{2}{9}\log 5.$$

Substituting this in (3.15) we get

$$\frac{d}{dp}H_n^*(p;1)_{|p=0.5}\geqq\frac{1}{2}\log 3+\frac{1}{2}H_{n-1}^*(0;1)-\frac{1}{2}H_{n-1}^*\left(\frac{1}{3};1\right)$$

$$-\frac{1}{6}H_{n-2}^*(0,1)+\frac{1}{6}H_{n-2}^*\left(\frac{2}{5};1\right)-\frac{1}{6}\log 5.$$

Applying the basic recursion (3.14) to $H_{n-1}^*(0;1)$ and $H_{n-1}^*\left(\frac{1}{3};1\right)$ we continue the derivation with

$$\frac{d}{dp}(H_n^*(p;1)_{|p=0.5}\geqq\frac{1}{2}\log 3+\frac{1}{2}H_{n-2}^*\left(\frac{1}{2};1\right)-\frac{1}{2}h\left(\frac{1}{6}\right)$$

$$-\frac{1}{12}H_{n-2}^*(0;1)-\frac{3}{8}H_{n-2}^*\left(\frac{2}{5};1\right)$$

$$-\frac{1}{6}H_{n-2}^*(0;1)+\frac{1}{6}H_{n-2}^*\left(\frac{2}{5};1\right)-\frac{1}{6}\log 5$$

$$\geqq\frac{1}{2}\log 3-\frac{1}{2}h\left(\frac{1}{6}\right)-\frac{1}{6}\log 5$$

by induction hypothesis.

Since $\frac{1}{2}\log 3-\frac{1}{2}h\left(\frac{1}{6}\right)-\frac{1}{6}\log 5=\frac{1}{4}\log\frac{5}{4}\geq 0$, the proof is complete.

Lemma 4. For $p\leq\frac{1}{2}$ we have

$$H_n^*(p;1)\leq H_n^*(p;0).$$

Proof. Since $H_n^*\left(\frac{1}{2};0\right)=H_n^*\left(\frac{1}{2};1\right)$, it suffices to show that

$$\frac{d}{dp}(H_n^*(p;0)-H_n^*(p;1))\leq 0 \quad\text{for}\quad 0\leq p\leq 0.5. \tag{3.16}$$

The proof of this inequality repeatedly makes use of (3.14) and needs formidable calculations.

Since $H_n^*(p;0)=H_n^*(1-p;1)$ we get from (3.14)

$$H_n^*(p;0)=h\left(\frac{1-p}{2}\right)+\frac{1-p}{2}H_{n-1}^*(0;1)+\frac{1+p}{2}H_{n-1}^*\left(\frac{p}{1+p};1\right). \tag{3.17}$$

Also from (3.14)

$$H_{n-1}^*\left(\frac{p}{1+p};1\right)=h\left(\frac{1}{2+p}\right)+\frac{p}{2(1+p)}H_{n-2}^*(0;1)+\frac{2+p}{2(1+p)}H_{n-2}^*\left(\frac{1}{2+p};1\right). \tag{3.18}$$

From these two equations we deduce

$$\frac{d}{dp}H_n(p;0)=\frac{1}{2}\log\frac{1-p}{1+p}-\frac{1}{2}H_{n-1}^*(0;1) \tag{3.19}$$
$$+\frac{1}{2}H_{n-1}^*\left(\frac{p}{1+p};1\right)+\frac{1+p}{2}\frac{d}{dp}H_{n-1}^*\left(\frac{p}{1+p};1\right)$$

and

$$\frac{d}{dp}H_{n-1}^*\left(\frac{p}{1+p};1\right)=\frac{2+p}{2(1+p)}\frac{d}{dp}H_{n-2}^*\left(\frac{1}{2+p};1\right) \tag{3.20}$$
$$+\frac{1}{2}\left(\frac{1}{1+p}\right)^2\left[\log\frac{2+p}{p}+H_{n-2}^*(0;1)-H_{n-2}^*\left(\frac{1}{2+p};1\right)\right].$$

Since by Lemma 3 the first term on the right side of (3.20) is negative we conclude that

$$\frac{d}{dp}H_{n-1}^*\left(\frac{p}{1+p};1\right)\leq\frac{1}{2}\frac{1}{(1+p)^2}\left[\log\frac{2+p}{p}+H_{n-2}^*(0;1)-H_{n-2}^*\left(\frac{1}{2+p};1\right)\right]. \tag{3.21}$$

This and (3.19) imply

$$\frac{d}{dp}H_n^*(p;0)\leqq\frac{1}{2}\Big[\log\frac{1-p}{2+p}-H_{n-1}^*(0;1)+H_{n-1}^*\left(\frac{p}{1+p};1\right)+$$
$$+\frac{1}{2(1+p)}\left(\log\frac{2+p}{p}+H_{n-2}^*(0;1)-H_{n-2}^*\left(\frac{1}{2+p};1\right)\right).\quad(3.22)$$

Similarly, we can prove

$$\frac{d}{dp}H_n^*(p;1)\geqq\frac{1}{2}\Big[H_{n-1}^*(0;1)+\log\frac{2-p}{p}-H_{n-1}^*\left(\frac{1}{3-p};1\right)-$$
$$-\frac{1}{2(2-p)}\left(H_{n-2}^*(0;1)-H_{n-2}^*\left(\frac{1}{3-p};1\right)+\log\frac{3-p}{1-p}\right)\Big].$$

Using these two inequalities, we deduce

$$\frac{d}{dp}[H_n^*(p;0)-H_n^*(p;1)]\leqq-H_{n-1}^*(0;1)-\frac{1}{2}\log\frac{(2+p)(1+p)}{p(1-p)}+$$
$$+\frac{1}{2}H_{n-1}^*\left(\frac{1-p}{2-p};1\right)+\frac{1}{2}H_{n-1}^*\left(\frac{p}{1+p};1\right)+$$
$$+\frac{1}{4(1+p)}\left[H_{n-2}^*(0;1)-H_{n-2}^*\left(\left(\frac{1}{2+p}\right);1\right)+\log\frac{2+p}{p}\right]+$$
$$+\frac{1}{4(2-p)}\left[H_{n-2}^*(0;1)-H_{n-2}^*\left(\frac{1}{3-p};1\right)+\log\frac{3-p}{1-p}\right].$$

Using (3.14) for all the terms with parameter $n-1$, we obtain

$$\frac{d}{dp}[H_n^*(p;0)-H_n^*(p;1)]\leqq-H_{n-2}^*\left(\frac{1}{2};1\right)+\frac{1-p}{4(2-p)}H_{n-2}^*(0;1)$$
$$+\frac{3-p}{4(2-p)}H_{n-2}^*\left(\frac{1}{3-p};1\right)+\frac{1}{2}h\left(\frac{1-p}{2(2-p)}\right)+\frac{p}{4(1+p)}H_{n-2}^*(0;1)$$
$$+\frac{2+p}{4(1+p)}H_{n-2}^*\left(\frac{1}{2+p};1\right)+\frac{1}{2}h\left(\frac{p}{2(1+p)}\right)-\frac{1}{2}\log\frac{(2-p)(1+p)}{p(1-p)}$$
$$+\frac{1}{4(1+p)}\left[H_{n-2}^*(0;1)-H_{n-2}^*\left(\frac{1}{2+p};1\right)+\log\frac{2+p}{p}\right]$$
$$+\frac{1}{4(2-p)}\left[H_{n-2}^*(0;1)-H_{n-2}^*\left(\frac{1}{3-p};1\right)+\log\frac{3-p}{1-p}\right].$$

Noting that by Lemma 3 $H^*_{n-2}(p;1)$ is monotone increasing in p and in particular $H^*_{n-2}\left(\frac{1}{2};1\right) \geqq H^*_{n-2}(p;1)$ for all $p \leqq \frac{1}{2}$ we obtain

$$\frac{d}{dp}[H^*_n(p;0) - H^*_n(p;1)] \leqq -\frac{1}{2}\log\frac{(2-p)(1+p)}{p(1-p)} + \frac{1}{2}h\left(\frac{1-p}{2(2-p)}\right) + \frac{1}{2}h\left(\frac{p}{2(1+p)}\right)$$

$$+\frac{1}{4(1+p)}\log\frac{2+p}{p} + \frac{1}{4(2-p)}\log\frac{3-p}{1-p}$$

$$= \frac{1}{4}\log\frac{16p(1-p)}{(3-p)(2+p)} \leqq \frac{1}{4}\log\frac{4}{(3-p)(2+p)}$$

$$\leqq \frac{1}{4}\log\frac{4}{6+p-p^2} \leqq \frac{1}{4}\log\frac{4}{6} \leqq 0.$$

Lemma 4 is proved.

Lemma 5. For $0 \leqq p \leqq \frac{1}{2}$ and $0 \leqq r \leqq 1$ we have

$$H^*_n(p;1) \leqq H^*_n(p;r).$$

Proof. To save notation we introduce $t = 1-r$. We distinguish between the following 3 cases:

$$(1)\quad t \leqq p; \qquad (2)\quad p < t \leqq 1-p; \qquad (3)\quad t > 1-p.$$

Notice that case (3) can be proved in the same way as case (1) by changing r to t and p to $1-p$.

We use the general form of the basic recursion in case (1).

$$H^*_n(p;r) = \frac{p+t}{2}H^*_{n-1}\left(\frac{t}{p+t};1\right) + \left(1-\frac{p+t}{2}\right)H^*_{n-1}\left(\frac{1-p}{2-p-t};1\right) + h\left(\frac{p+t}{2}\right). \tag{3.23}$$

Lower bounding the first summand with Lemma 1, we get

$$H^*_n(p;r) \geqq \frac{p}{2}H^*_{n-1}(0;1) + \frac{t}{2}H^*_{n-1}(1;1) +$$

$$+\left(1-\frac{p+t}{2}\right)H^*_{n-1}\left(\frac{1-p}{2-p-t};1\right) + h\left(\frac{p+t}{2}\right). \tag{3.24}$$

Since by Lemma 4 $H^*_{n-1}(1;1)=H^*_{n-1}(0;0)\geqq H^*_{n-1}(0;1)$ we conclude that

$$H^*_n(p;r)\geqq \frac{p}{2}H^*_{n-1}(0;1)+\frac{t}{2}H^*_{n-1}(0;1)+$$
$$+\left(1-\frac{p+1}{2}\right)H^*_{n-1}\left(\frac{1-p}{2-p-t};\ 1\right)+h\left(\frac{p+t}{2}\right). \quad (3.25)$$

Application of Lemma 2 to the 2 central summands gives

$$H^*_n(p;1)\geqq \frac{p}{2}H^*_{n-1}(0;1)+\left(1-\frac{p}{2}\right)H^*_{n-1}\left(\frac{1-p}{2-p};\ 1\right)+\frac{t}{2}h(0)+$$
$$+\left(1-\frac{p+t}{2}\right)h\left(\frac{1-p}{2-p-t}\right)-\left(1-\frac{p}{2}\right)h\left(\frac{1-p}{2-p}\right)+h\left(\frac{t+p}{2}\right). \quad (3.26)$$

The first two summands can be rewritten via (3.14) and by some manipulations

$$H^*_n(p;r)\geqq H^*_n(p;1)-h\left(\frac{p}{2}\right)+\left(1-\frac{p+t}{2}\right)h\left(\frac{1-p}{2-p-t}\right)-$$
$$-\left(1-\frac{p}{2}\right)h\left(\frac{1-p}{2-p}\right)+h\left(\frac{t+p}{2}\right)=$$
$$=H^*_n(p;1)-\frac{p+t}{2}h\left(\frac{p}{p+t}\right)+\frac{1}{2}h(t). \quad (3.27)$$

Since $t\leqq p\leqq \frac{1}{2}$, we have $t\leqq \frac{t}{p+t}$ and $h(t):h\left(\frac{t}{p+t}\right)\geqq t:\frac{t}{p+t}$.

Therefore $h(t)\geqq (p+t)h\left(\frac{p}{p+t}\right)$ and finally $H^*_n(p;r)\geqq H^*_n(p;1)$ in this case.

We now prove the result in case (2).

Here the basic recurrence takes the form

$$H^*_n(p;r)=\frac{p+t}{2}H^*_{n-1}\left(\frac{p}{p+t};\ 1\right)+\left(1-\frac{p+t}{2}\right)H^*_{n-1}\left(\frac{1-t}{2-p-t};\ 1\right)+h\left(\frac{p+t}{2}\right). \quad (3.28)$$

This differs from (3.23) only insofar as at the right side p and $\pm$ are exchanged. Instead of (3.27) we get therefore now

$$H^*_n(p;r)\geqq H^*_n(t;1)-\frac{p+t}{2}h\left(\frac{t}{p+t}\right)+\frac{1}{2}h(p). \quad (3.29)$$

Since $p<t\leqq 1-p$, we have $p\leqq \frac{p}{p+t}$ and $h(p):h\left(\frac{p}{p+t}\right)\geqq p:\frac{p}{p+t}$. Therefore we get in this case

$$H_n^*(p;r)\geqq H_n^*(t;1). \tag{3.30}$$

For $t\leqq \frac{1}{2}$, since $p<t$ by Lemma 3

$$H_n^*(t;r)\geqq H_n^*(p;1)$$

and thus the desired result

$$H_n^*(p;r)\geqq H_n^*(p;1), \tag{3.31}$$

and for $t>\frac{1}{2}$ by Lemma 4

$$H_n^*(t;1)\geqq H_n^*(t;0)=H_n^*(r;1),$$

by Lemma 3 $H_n^*(r;1)\geqq H_n^*(p;1)$, and again the inequality (3.31).

We derive now a formula for the limiting entropy rate $\eta_2(\infty,2,0,P_X)$. Suppose that (w.l.o.g.) $\hat{S}_0=0$ and that we follow the optimal strategy. For its analysis we introduce the events $E_k=\{Y^k=01010\ldots\}$ and $D_k=E_k\setminus E_{k+1}$. The D_k's are disjoint and $q(k)\triangleq \text{Prob.}(D_k)$ satisfies

$$\sum_{n=1}^{\infty} q(k)=1. \tag{3.32}$$

Theorem 3. For $P_X(0)=P_X(1)=1/2$

$$\eta_2(\infty,2,0,P_X)=\frac{H(q)}{\sum_{k=1}^{\infty} kq(k)}.$$

Proof. By the grouping axiom for entropy

$$\begin{aligned} H(Y^i|\hat{S}_0=0)=&-\sum_{t=1}^{i=1} q(t)\log(t)\\ &-(1-q(1)-\cdot-q(i-1))\log(1-q(1)-\cdot-q(i-1))\\ &+\sum_{t=1}^{i-1} q(t)H(Y^{i-t}|\hat{S}_0=0). \end{aligned}$$

Therefore we get

$$\sum_{i=1}^{n} H(Y^i|\hat{S}_0=0) = -\sum_{i=1}^{n}\sum_{t=1}^{i-1} q(t)\log q(t)$$

$$-\sum_{i=1}^{n}(1-q(1)-\cdots-q(i-1))\log(1-q(1)-\cdots-q(i-1))$$

$$+\sum_{i=1}^{n}\sum_{t=1}^{i-1} q(t)H(Y^{i-t}|\hat{S}_0=0) = -\sum_{t=1}^{n-1}(n-t)q(t)\log q(t)$$

$$-\sum_{i=1}^{n}\left(1-\sum_{j=1}^{i-1}q(j)\right)\log\left(1-\sum_{j=1}^{i-1}q(j)\right)+\sum_{i=1}^{n-1}\sum_{t=1}^{n-i}q(t)H(Y^i|\hat{S}_0=0)$$

and consequently

$$H(Y^n|\hat{S}_0=0)+\sum_{i=1}^{n-1}\left(1-\sum_{t=1}^{n-i}q(t)\right)H(Y^i|\hat{S}_0=0)$$

$$\geqq -\sum_{t=1}^{n-i}(n-t)q(t)\log q(t). \tag{3.33}$$

Notice that $\displaystyle\sum_{i=1}^{n-1}\left(1-\sum_{t=1}^{n-i}q(t)\right)=\sum_{i=1}^{n-1}\sum_{t=n-i+1}^{\infty}q(t)$

and

$$\sum_{t=1}^{\infty}q(t)+\sum_{i=1}^{n-1}\sum_{t=n-i+1}^{\infty}q(1)=\sum_{i=1}^{n}\sum_{t=n-i+1}^{\infty}q(t)$$

$$=\sum_{s=1}^{\infty}\sum_{i=n+1-s}^{n}q(s)=\sum_{s=1}^{\infty}sq(s). \tag{3.34}$$

Since $H(Y^n|\hat{S}_0=0)\geqq H(Y^i|\hat{S}_0=0)$ for $i\leqq n$ from (3.33) and (3.34) we can derive

$$\sum_{s=1}^{\infty}sq(s)H(Y^n|\hat{S}_0=0)\geqq -\sum_{t=1}^{n-1}(n-t)q(t)\log q(t). \tag{3.35}$$

Since for fixed k $\displaystyle -\sum_{t=1}^{n-1}(n-t)q(t)\log q(t)\geqq(n-k)\sum_{t=1}^{k}-q(t)\log q(t)$ we continue with

$$\frac{1}{n}H(Y^n|\hat{S}_0=0)\geq\frac{(n-k)\sum_{t=1}^{k}-q(t)\log q(t)}{n\sum_{s=1}^{\infty}sq(s)}.$$

Let now first n tend to infinity and then k. We obtain

$$\lim_{n\to\infty}\frac{1}{n}H(Y^n|\hat{S}_0=0)\geq\frac{H(q)}{\sum_{s=1}^{\infty}sq(s)}. \tag{3.36}$$

Instead of (3.33) we derive first the upper bound

$$H(Y^n|\hat{S}_0=0)+\sum_{i=1}^{n-1}\left(1-\sum_{t=1}^{n-i}q(t)\right)H(Y^i|\hat{S}_0=0)\leqq$$

$$\leqq -n\sum_{t=1}^{n-1}q(t)\log q(t)-\sum_{i=1}^{n}\left(1-\sum_{j=1}^{i-1}q(j)\right)\log\left(1-\sum_{j=1}^{i-1}q(j)\right)\leqq nH(q). \tag{3.37}$$

Therefore

$$H(Y^n|\hat{S}_0=0)+\sum_{i=n-k+1}^{n-1}\sum_{t=n-i+1}^{\infty}q(t)H(Y^i|\hat{S}_0=0)\leqq nH(q)$$

and a fortiori

$$H(Y^{n-k+1}|\hat{S}_0=0)\sum_{i=1}^{k}iq(i)\leqq nH(q). \tag{3.38}$$

Since $H(Y^n|\hat{S}_0=0)\leqq H(Y^{n-k+1}|\hat{S}_0=0)+k$ we derive from (3.38)

$$\overline{\lim_{n\to\infty}}\frac{1}{n}H(Y^n|\hat{S}_0=0)\leqq\frac{H(q)}{\sum_{i=1}^{\infty}iq(i)}$$

for all k and therefore also

$$\overline{\lim_{n\to\infty}}\frac{1}{n}H(Y^n|\hat{S}_0=0)\leqq\frac{H(q)}{\sum_{i=1}^{\infty}iq(i)}.$$

This and (3.36) complete the proof.

Remark. The sequence $q(1)$, $q(2)$, obeys a simple rule, so that we can calculate

$$\frac{H(q)}{\sum_{i=1}^{\infty}iq(i)}=0.5989\ldots.$$

The formula in Theorem 3 has a nice structure. It suggests a general principle for arbitrary sources.

Reference

1. *Ahlswede, R., Ye, J. P., Zhang, Z.*, Creating order under constraints on mind and matter. Submitted to Information and Computation.

Вклад в теорию упорядочения пространств последовательностей

Р. АЛСЬЕДЕ, З. ЗАНГ

(Билефелд)

Настоящая работа продолжает исследования, начатые в [1]. Получены первые результаты об оптимальном упорядочении последовательностей в двух простых, но существенных случаях: невероятностная модель с активной памятью и вероятностная модель с неограниченной пассивной памятью.

R. Ahlswede
Universität Bielefeld
Fakultät für Mathematik
Universitätsstraße
4800 Bielefeld 1
Federal Republic of Germany