# Ratewise-optimal non-sequential search strategies under constraints on the tests ⋆

Rudolf Ahlswede

**Abstract**

Already in his Lectures on Search Renyi suggested to consider a search problem, where an unknown $x \in \mathcal{X} = \{1, 2, \ldots, n\}$ is to be found by asking for containment in a minimal number $m(n,k)$ of subsets $A_1, \ldots, A_m$ with the restrictions $|A_i| \leq k < \frac{n}{2}$ for $i = 1, 2, \ldots, m$.

Katona gave in 1966 the lower bound $m(n,k) \geq \frac{\log n}{h(\frac{k}{n})}$ in terms of binary entropy and the upper bound $m(n,k) \leq \left\lceil \frac{\log n+1}{\log n/k} \right\rceil \cdot \frac{n}{k}$, which was improved by Wegener in 1979 to $m(n,k) \leq \left\lceil \frac{\log n}{\log n/k} \right\rceil \left( \left\lceil \frac{n}{k} \right\rceil - 1 \right)$.

We prove here for $k = pn$ that $m(n,k) = \frac{\log n + o(\log n)}{h(p)}$, that is, ratewise optimality of the entropy bound: $\lim_{n\to\infty} \frac{m(n,pn)}{\log n} = \frac{1}{h(p)}$.

Actually this work was motivated by a more recent study of Karpovsky, Chakrabarty, Levitin and Avresky of a problem on fault diagnosis in hypercubes, which amounts to finding the minimal number $M(n,r)$ of Hamming balls of radius $r = \rho n$ with $\rho \leq \frac{1}{2}$ in the Hamming space $\mathcal{H}^n = \{0,1\}^n$, which separate the vertices. Their bounds on $M(n,r)$ are far from being optimal. We establish bounds implying

$$\lim_{n\to\infty} \frac{1}{n} \log M(n,r) = 1 - h(\rho).$$

However, it must be emphasized that the methods of prove for our two upper bounds are quite different.

## 1 Introduction

[1]

⋆ Presented at the Annual Meeting of the DFG Schwerpunkt Nr. 1126 "Algorithmik grosser und komplexer Netzwerke", March 26-28, 2003, University of Tübingen.

[1] The results were also presented at the meeting "General Theory of Information Transfer and Combinatorics" at the Zentrum für interdisziplinäre Forschung (ZiF)

Concepts and some basic results on search can be found in the books [4], [5]. Those needed in this paper are repeated in Sections 2 and 3. Basic is an information-theoretic idea to derive lower bounds on the number of tests.

**Quite surprisingly, eventhough this result is known for several decades, nobody proved that – or even seems to have wondered whether – it is essentially best possible** for instance for $k$-set tests "carrying $h\left(\frac{k}{n}\right)$ bit of information".

However, when we looked for a proof we realized an obstacle, which blocked the development even for people who may have believed in the entropy bound. The known proofs for upper bounds (Theorem KW in Section 3) are constructive and apparently hard to improve. In such a situation often a probabilistic argument helps. However, a standard approach by random choice is suboptimal even for the simple case of unrestricted tests as was noticed already by Renyi [10]. Using the uniform distribution for choosing a separating system (see Section 2) requires

$$m \geq 2 \log n + 6 \tag{1.1}$$

sets, where $\lceil \log n \rceil$ is optimal (see Lemma 1 in Section 2). So we are by a factor of 2 away from the optimum!

Our discovery is that – also in the restricted case – we can close the gap by advanced random choices used for code selections in information theory ([1], [7]), which we explain in Section 4 for error correcting codes.

After this preparation we turn to separating systems and present a dictionary, which explains how the methods for codes can be translated into methods for separating systems, when we focus on the **columns** of the incidence matrix. Thus we get in Section 5 in Theorem 1 (i), (ii) the desired entropy bound first for an average cardinality constraint. This is then improved in (iii) of Theorem 1 to a worst case constraint using a familiar large deviation argument.

Finally, in Section 6 we settle a separation problem with balls in Hamming space, which originated in the theory of diagnosis [8]. Here we interpret the problem as a covering problem and achieve the goal with the Covering Lemma of [2], **whereas the previous method used on the first problem fails and vice versa!**

**This leaves us with a challenging future task of analysing the interplay of separating systems and coverings.**

in Bielefeld, April 26-30, 2004, where discussions about them with G. Katona led us to add Appendices to the paper.

## 2 Nonsequential strategies and separating systems

Let the search domain be defined by $\mathcal{X} = \{1, \ldots, n\}$ $(n \in \mathbb{N})$. Every nonsequential strategy for the search problem presented in Section 1 can be described by a sequence $t_{A_1}, \ldots, t_{A_m}$ $(A_1, \ldots, A_m \subset \mathcal{X}, m \in \mathbb{N})$. The nonsequential strategy $s = (t_{A_1}, \ldots, t_{A_m})$ is said to be successful if and only if for every $x \in \mathcal{X}$, the sequence $t_{A_1}(x), \ldots, t_{A_m}(x)$ of results determines the object uniquely. Either the last test $t_{A_m}$ is superfluous or the strategy requires in the worst case $m$ tests in order to identify the object being sought. We can limit ourselves to the analysis of successful strategies for which the last test is not superfluous. From these strategies one should be chosen for which the worst search time, i.e., $m$, is minimal.

Now, before we consider this search problem, we provide a connection to a problem of combinatorics. The strategy $s = (t_{A_1}, \ldots, t_{A_m})$ is successful if and only if for $x \neq y$ $(x, y \in \mathcal{X})$ there is an $i \in \{1, \ldots, m\}$ such that $t_{A_i}(x) \neq t_{A_i}(y)$, i.e. $x \in A_i$ and $y \notin A_i$ or $x \notin A_i$ and $y \in A_i$. Such set systems are called separating systems.

**Definition 1.** $A_1, \ldots, A_m$ constitute a separating system in $\mathcal{X}$ if and only if the following condition is met:

$$\forall x, y \in X, x \neq y \quad \exists 1 \leq i \leq m : \quad x \in A_i, y \notin A_i \text{ or } x \notin A_i, y \in A_i.$$

These considerations can be summarized as follows:

**Remark 1:** The nonsequential strategy $s = (t_{A_1}, \ldots, t_{A_m})$ is successful if and only if the sets $A_1, \ldots, A_m$ constitute a separating system in $\mathcal{X}$.

In order to decide whether $s$ is a successful strategy, we present every test by its value table $(t_{A_i}(1), \ldots, t_{A_i}(n))$. Let $A = (a_{ix})$ be the following $m \times n$ matrix with values from $\{0, 1\}$:

$$a_{ix} = 1 :\leftrightarrow t_{A_i}(x) = 1.$$

$A$ is called the incidence matrix of the strategy $s$. We see that $s$ is successful if and only if all $n$ columns of $A$ are distinct.

The columns of $A$ have length $m$, and there are exactly $2^m$ distinct $0-1$ vectors of length $m$. Thus, for a successful strategy $s = (t_{A_1}, \ldots, t_{A_m})$, $2^m \geq n$ must hold and therefore $m \geq \lceil \log_2 n \rceil$. If $m = \lceil \log_2 n \rceil$ and thus $2^m \geq n$, we can select $n$ distinct $0-1$ vectors $a_1, \ldots, a_n$ of length $m$. The strategy whose incidence matrix is composed of the column vectors $a_1, \ldots, a_n$ is successful. In the following, log is always considered to be $\log_2$.

We summarize the solution to this search problem.

**Lemma 1.** *If all binary tests are admitted, there is a nonsequential strategy which identifies every object in a search domain with $n$ elements at the latest after $m = \lceil \log n \rceil$ tests. For all $m < \lceil \log n \rceil$ there is no successful strategy $t_{A_1}, \ldots, t_{A_m}$. (A minimal separating system for a set with $n$ elements consists of $\lceil \log n \rceil$ sets.)*

**Remark 2:** An alternative proof uses the representation of the numbers $1, 2, \ldots, n$ as binary sequences of length $\lceil \log n \rceil$. This gives $\mathcal{Y} = \{0,1\}^{\lceil \log n \rceil}$. Define $A_i = \{y \in \mathcal{Y} : y_i = 1\}$ and notice that $\{A_i : 1 \leq i \leq \lceil \log n \rceil\}$ is a separating system.

## 3 Separating systems of sets with at most $k$ elements

**Definition 2.** Let $m(n,k)$ be the maximum search time of an optimal nonsequential strategy for finding an object in a search domain of $n$ elements if only the binary tests $t_A$ with $|A| \leq k$ are admitted.

A nonsequential strategy $s = (t_{A_1}, \ldots, t_{A_m})$ is successful if and only if the sets $A_1, \ldots, A_m$ form a separating system. Therefore, $m(n,k)$ is also the number of sets which are contained in a minimal separating system on $\mathcal{X} = \{1, \ldots, n\}$ which consists of sets of at most $k$ elements.

For $n \leq 2k$, it follows from Lemma 1 that $m(n,k) = \lceil \log n \rceil$. In the following, we assume $n > 2k$. Katona [9] proved the following lower bound for $m(n,k)$.

**Theorem K.** *For $n > 2k$*

*(a) $m(n,k) \geq \frac{\log n}{h(k/n)}$, where $h$ is the binary entropy function $h(q) = -q \log q - (1-q)\log(1-q)$.*

*(b) $m(n,k) \geq \frac{\log n}{\log(en/k)} \frac{n}{k}$.*

Here (b) follows from (a) by elementary calculations. The proof of (a) is based on an information-theoretic result expressed in the following inequality. We repeat the original proof, because we shall later use it with an improvement based on convexity of $h$.

**Lemma 2.** *For the entropy of $m$ random variables $Y_1, \ldots, Y_m$, which assume*

*only a finite number of values, we have*

$$H(Y_1, \ldots, Y_m) \leq \sum_{1 \leq i \leq m} H(Y_i),$$

*with equality if and only if* $Y_1, \ldots, Y_m$ *are independent.*

**Proof of Theorem K:** (a) Let $A_1, \ldots, A_m$, $m = m(n,k)$ be a minimal separating system on $\mathcal{X} = \{1, \ldots, n\}$ of sets with at most $k$ elements. Let the uniform distribution be given on $\mathcal{X}$, and let $1_{A_i}$ be the indicator variable of $A_i$, i.e., $1_{A_i}$ assumes the value 1 or 0 depending on whether the object being sought is in $A_i$ or not. We have $Pr(1_{A_i} = 1) = |A_i|/n \leq k/n$. The entropy function $h$ increases monotonically in the domain $\left[0, \frac{1}{2}\right]$. For $n > 2k$, therefore, $H(1_{A_i}) = h(|A_i|/n) \leq h(k/n)$.

The random vector $(1_{A_i}, \ldots, 1_{A_m})$ assumes, since $A_1, \ldots, A_m$ is a separating system, different values for different $x \in \mathcal{X}$. Therefore, the distribution of $(1_{A_1}, \ldots, 1_{A_m})$ is the uniform distribution on $n$ values and $H(1_{A_1}, \ldots, 1_{A_m}) = \log n$.

It follows from Lemma 2 that $\log n \leq mh(k/n)$ and thus (a).

We conclude with the familiar upper bounds.

**Theorem K,W.** *For* $n > 2k$,

*(a)* $m(n,k) \leq \left\lceil \frac{\log n + 1}{\log(n/k)} \right\rceil \frac{n}{k}$ *(Katona [9]).*

*(b)* $m(n,k) \leq \left\lceil \frac{\log n}{\log(n/k)} \right\rceil (\lceil n/k \rceil - 1)$ *(Wegener [12]).*

## 4 Basic methods of proving the Gilbert-type bounds on the cardinality of a code

Let us consider the following problem: we are given a code length $n$ and a value of $d \leq \frac{1}{2}n$. What is a lower bound on the cardinality of a binary code having the minimal distance not less than $d$?

### Maximal coding (Gilbert bound)

Since $d$ is the minimal distance of a code, we have an evident inequality

$$M \geq \frac{2^n}{S_d} \sim 2^{n(1-h(\delta))}, \tag{4.1}$$

where $\delta = d/n \leq \frac{1}{2}$ and $S_d$ is the cardinality of a Hamming ball of radius $d$ in $\{0,1\}^n$. It is well-known that $S_d \sim 2^{h(\delta)n}$.

**Selection of a random code**

Suppose, we want to find a code with $M$ codewords selecting the codewords at random. There are $2^{nM}$ codes. Let us fix the $m$-th codeword. The number of choices of all other codewords such that at least one of them is located at the Hamming distance less than $d$ from the $m$-th codeword is not greater than

$$(M-1)2^{n(M-2)}S_{d-1}.$$

Since $m$ can vary over $1,\ldots,M$ and the $m$-th codeword can take $2^n$ values, the number of 'bad' codes (the codes with the minimal distance less than $d$) is not greater than

$$M(M-1)2^{n(M-1)}S_{d-1}. \tag{4.2}$$

If this expression is less than the total number of codes, i.e.,

$$M(M-1)2^{n(M-1)}S_{d-1} < 2^{nM},$$

then there exists at least one code with the desired property. Direct calculations show that it is possible if

$$M^2 < \frac{2^n}{S_{d-1}}. \tag{4.3}$$

Hence, the exponent of our upper bound is **twice less** than the exponent we get in (4.1). The method that can be used to improve the result is known as **expurgation**. Note that the **probability** to select a bad $i$-th codeword is upper-bounded by

$$\frac{(M-1)S_{d-1}}{2^n}.$$

Thus, the average number of the bad words is upper-bounded by

$$M\frac{(M-1)S_{d-1}}{2^n}. \tag{4.4}$$

Let us require this to be smaller than $\frac{1}{2}(M-1)$ and let us expurgate bad words. Then, constructing a new code that contains only the remaining $M'$ codewords, we get the inequality

$$M' > \frac{1}{4}\frac{2^n}{S_{d-1}}, \tag{4.5}$$

which is only by a factor $\frac{1}{4}$ **less** than the Gilbert bound (the exponent of the bound is the same as the exponent of Gilbert's bound in the ratewise sense).

**Selection of clouds of random codes**

Suppose that we want to construct $M$ clouds such that **each cloud consists of $K$ codewords**. The minimal distance between some codeword of every cloud and all codewords belonging to the other clouds should be not less than $d$. A generalization of the previous counting leads to the following inequality, which upperbounds the number of bad cloud systems by the product of the number of bad first clouds, the number of messages $M$ and the total number of cloud systems for $M-1$ messages and requires that this be smaller than the total number of cloud systems for $M$ messages,

$$M(K(M-1)S_{d-1})^K 2^{nK(M-1)} < 2^{nKM} \tag{4.6}$$

or

$$M^{1/K}(K(M-1)S_{d-1}) < 2^n.$$

If we set $K=n$, then this inequality can be written as

$$M^{\frac{1}{n}}(M-1) < \frac{2^n}{n \cdot S_{d-1}}.$$

Sufficient for this is

$$M^{\frac{n+1}{n}} < \frac{2^n}{nS_{d-1}} \text{ or } M < \left(\frac{2^n}{nS_{d-1}}\right)^{\frac{n}{n+1}}.$$

Since $\frac{n^2}{n+1} \geq n-1$, again sufficient for this is

$$M < \frac{2^{n-1}}{nS_{d-1}}.$$

As a result we obtain

$$M \sim \frac{1}{2n}\frac{2^n}{S_{d-1}}, \tag{4.7}$$

i.e., the construction based on the clouds of codewords instead of one codeword assigned to each message leads to approximately the same result as expurgation.

## 5 Separating systems with an average cardinality constraint

Recall that to an $(m,n)$-separating system $(A_1,\dots,A_m)$ of subsets $A_i \subset \mathcal{X} = \{1,2,\dots,n\}$ corresponds the incidence matrix $A = (a_{ix})_{\substack{1\le i\le m\\ 1\le x\le n}}$ with $m$ rows and $n$ columns, where the columns are distinct. If every row contains at most $k$ 1's we speak of an $(m,n,k)$-separating system. For given $n,k$ $m(n,k)$ is the minimal $m$ for which an $(m,n,k)$-separating system exists.

One can generalize this concept by requiring

$$\frac{1}{m}\sum_{i=1}^{m}|A_i| \le k. \tag{5.1}$$

Here $k$ is an average cardinality constraint. Correspondingly we consider $(m,n)$-separating systems meeting constraint (5.1) and denote the minimal $m$ for which for given $n,k$ such a system exists by $\underline{m}(n,k)$.

Our main result is the

**Theorem 1.** *For $n > 2k$ and $p = \frac{k}{n}$,*

(i) $\underline{m}(n,k) \ge \frac{\log n}{h(p)}$.
(ii) $\underline{m}(n,k) \le \frac{\log n + o(\log n)}{h(p)}$,
(iii) $\underline{m}(n,k) \le m(n,k) = \frac{\log n + o(\log n)}{h(p)}$.

**Proof:** (i) Obviously $\underline{m}(n,k) \le m(n,k)$. Therefore (i) improves (a) in Theorem K. The proof follows again Lemma 2. Thus

$$\log n \le \sum_{i=1}^{m} h\left(\frac{|A_i|}{n}\right)$$

and now we proceed differently with the convexity of entropy

$$= m\sum_{i=1}^{m}\frac{1}{m}h\left(\frac{|A_i|}{n}\right) \le m\ h\left(\frac{1}{m}\sum_{i=1}^{m}\frac{|A_i|}{n}\right) \le m\ h\left(\frac{k}{n}\right) = m(h(p)),$$

because $h$ is monotone increasing for $p \le \frac{1}{2}$.

(ii) We translate the proof for codes based on clouds, which is presented in Section 4, into the present situation.

**Construction**

For an $m \times n$-matrix with $pm$ 1's in every column we use the following **dictionary** relating to code concepts.

| | | | | |
|---|---|---|---|---|
| codewords | | $\leftrightarrow$ | columns | |
| $M$ | | $\leftrightarrow$ | $n$ | (want to separate many columns) |
| $2^n$ | (number of possible codewords) | $\leftrightarrow$ | $\binom{m}{pm}$ | (number of possible columns) |
| $S_{d-1}$ | (bad codewords for one codeword) | $\leftrightarrow$ | 1 | (bad column is an identical column) |

To make a first observation recall that in the random choice of codes the number of bad codes in (5.2) is bounded by

$$M(M-1)2^{n(M-1)}S_{d-1} \overset{!}{<} 2^{nM}.$$

**Translation by dictionary**

$$\text{number of bad matrices } \leq n(n-1)\binom{m}{pm}^{n-1} \overset{!}{<} \binom{m}{pm}^{n} \tag{5.2}$$

Sufficient for (5.2) is

$$n^2 < \binom{m}{pm} \sim 2^{h(p)m} \tag{5.3}$$

$$m \sim \frac{2\log n}{h(p)}. \tag{5.4}$$

The factor "2" occurs again as in (1.1).

Now we make a random choice of clouds of matrices.

Recall (4.6), where the number of bad cloud systems for coding is upper bounded by

$$M(K(M-1)S_{d-1})^K 2^{nK(M-1)} \overset{!}{<} 2^{nKM}. \tag{5.5}$$

**Translation by dictionary**

$$n(K(n-1)\cdot 1)^K\binom{m}{pm}^{K(n-1)} \overset{!}{<} \binom{m}{pm}^{Kn} \tag{5.6}$$

or

$$n^{1/K}(K(n-1)) \overset{!}{<} \binom{m}{pm} \sim 2^{h(p)m} \tag{5.7}$$

or

$$\frac{1}{K}\log n + \log K + \log(n-1) \overset{!}{<} h(p)m. \tag{5.8}$$

Choose $K = (\log n)^2$ and obtain

$$\left(\frac{1}{\log n} + 2\log\log n\right) + \log n = o(\log n) + \log n \stackrel{!}{<} h(p)m \tag{5.9}$$

and thus (ii).

(iii) We have to get now the worst case constraint for the rows and not as previously for columns. We achieve this by a kind of expurgation with the following procedure:

1. We follow the cloud construction as before in (ii). Clearly the number of bad cloud systems with average size constraint
   $\geq$ number of bad cloud systems with worst case constraint.
2. Choose columns as before with probability $\frac{1}{\binom{m}{pm}}$. Thus for the $i$-th row of matrix $X = (X_{ij})_{\substack{1\leq i\leq m\\ 1\leq j\leq n}}$ we have $\mathbb{E}\ X_{i1} = \cdots = \mathbb{E}\ X_{in} = p$ and for $i = 1, 2, \ldots, m$

$$Prob\left(\sum_{j=1}^{n} X_{ij} > (p+\varepsilon)n\right) \leq e^{-E(p,\varepsilon)n}. \tag{5.10}$$

   Therefore $Prob(X$ does not meet $n(p+\varepsilon)$ constraint$) \leq m\ e^{-E(p,\varepsilon)n}$, $E(p,\varepsilon) > 0$.

   Furthermore, we thus have $\binom{m}{pm}^{nK}(1 - m\ e^{-E(p,\varepsilon)n})^K \geq \binom{m}{pm}^{nK}(1 - Kme^{-E(p,\varepsilon)n}) \triangleq T$ cloud systems with an $n(p+\varepsilon)$ worst case constraint. On the other hand the number of bad cloud systems (see (5.6)) is bounded by

$$n(K(n-1))^K \binom{m}{pm}^{K(n-1)} \stackrel{!}{<} T \tag{5.11}$$

   and therefore

$$n^{1/K}(K(n-1)) \stackrel{!}{<} \binom{m}{pm}(1 - Kme^{-E(p,\varepsilon)n})^{1/K}. \tag{5.12}$$

   Again, with the choice $K = (\log n)^2$ we obtain

$$\left(\frac{1}{\log n} + 2\log\log n\right) + \log n \stackrel{!}{<} h(p)m - me^{-E(p,\varepsilon)n}. \tag{5.13}$$

   To get the constraint $np$ we replace $p$ by $p - \varepsilon$ in this derivation and thus get

$$o(\log n) + \log n < h(p-\varepsilon)m - \log(1 - me^{-E(p-\varepsilon)n}).$$

Make now $\varepsilon = \varepsilon(n)$ dependent on $n$ such that

$$(h(p) - h(p-\varepsilon))m + \log(1 - me^{-E(p-\varepsilon,\varepsilon)n}) = o(\log n)$$

and thus we obtain (iii).

**Remark 3:** Choosing rows instead of columns with a constant number of 1's gives immediately the desired worst case constraint. However, there seems to be no way to get the desired entropy bound this way. We just got $m < \frac{f(p)\log n}{h(p)}$ for some $f(p) > 1$.

## 6 A search problem arising with a problem on fault diagnosis

Let $M(n,r)$ be the minimum number of balls of radius $r \leq \frac{1}{2}n$ in the Hamming space $\mathcal{H}^n = \{0,1\}^n$, which separate the vertices. This means that there is a system $\mathcal{B}(n,r)$ of balls, whose members are contained in $\mathcal{H}^n$, of cardinality $|\mathcal{B}(n,r)| = M(n,r)$ such that for every $x, y \in \mathcal{H}^n$ for some ball $B \in \mathcal{B}(n,r)$ we have

$$x \in B, y \notin B \text{ or } x \notin B, y \in B. \tag{6.1}$$

We see that $\mathcal{B}(n,r)$ is a separating system of sets, which possess a geometrical property, namely, they are balls of radius $r$.

We don't see how to extend the constructions described so far to derive an upper bound on $M(n,r)$.

If we select columns as in Section 6, it is difficult to get rows which constitute balls. We follow now another idea, namely, hypergraph covering (see [2]).

As edge-regular hypergraph $(\mathcal{V}, \mathcal{E})$ we choose as vertex set $\mathcal{V} = \{(x,y) : x, y \in \mathcal{H}^n, x \neq y\}$ and as edge set $\mathcal{E} = \{E_r(z) : z \in \mathcal{H}^n\}$ where

$$E_r(z) = \{(x,y) : x \in B_r(z), y \notin B_r(z)\} \tag{6.2}$$

for the ball $B_r(z)$ with center $z$ and radius $r = \rho n$ with $\rho \leq \frac{1}{2}$.

Now obviously for all edges the equal cardinalities are

$$e = |E_r(z)| = S_r(2^n - S_r) \sim 2^{h(\rho)n}(2^n - 2^{h(\rho)n}). \tag{6.3}$$

It is also readily seen that

$$d_{\min} = \min_{v \in \mathcal{V}} \deg(v) = \binom{n-1}{r}. \tag{6.4}$$

The inequality

$$\binom{n}{\rho n} \geq \frac{2^{h(\rho)n}}{\sqrt{8n\rho(1-\rho)}}$$

implies now

$$\begin{aligned} d_{\min} = \binom{n-1}{r} &= \frac{n-r}{n} \cdot \binom{n}{r} \\ &\geq \frac{2^{h(\rho)n}}{2\sqrt{2n}} \end{aligned}$$

(On the other hand we have for the average vertex degree

$$\bar{d} = \frac{1}{|\mathcal{V}|}|\mathcal{E}|e \sim \frac{2^n}{2^{2n}} 2^{h(\rho)n}(2^n - 2^{h(\rho)n}) = 2^{h(\rho)n}(1 - 2^{(h(\rho)-1)n}). \tag{6.5}$$

Therefore $\frac{\bar{d}}{d\max} = \frac{\bar{d}(\rho,n)}{d_{\max}(\rho,n)} \to 1$ as $n \to \infty$.)

By the Covering Lemma of [2] there exists a covering $\mathcal{C}$ which is a separating system such that

$$\begin{aligned} M(n,r) \leq |\mathcal{C}| &\leq \frac{|\mathcal{E}|}{d_{\min}} \log |\mathcal{V}| + 1 \\ &\leq \frac{2\sqrt{2n} \cdot 2^n}{2^{h(\rho)n}} \cdot \log 2^n(2^n - 1) + 1 \\ &\leq 4\sqrt{2} \cdot n^{3/2} \cdot 2^{(1-h(\rho))n} + 1 \end{aligned}$$

Since $|\mathcal{V}| = 2^n(2^n - 1)$, it follows from (6.3) that

$$M(n,r) \geq \frac{2^n(2^n-1)}{e} \sim \frac{1}{1 - 2^{(h(\rho)-1)n}} 2^{(1-h(\rho))n}. \tag{6.7}$$

Consequently

$$\lim_{n\to\infty} \frac{1}{n} \log M(n,r) = 1 - h(\rho). \tag{6.8}$$

Using the entropy argument with Lemma 2 we obtain

$$\log 2^n \leq M(n,r) h\left(\frac{S_r}{2^n}\right) \tag{6.9}$$

and thus with $S_r \sim 2^{h(\rho)n}$ and the inequality $\log x \leq x - 1$ for all $x > 0$, we

obtain

$$\begin{aligned} M(n,r) &\geq \frac{n}{S_r/2^n \log 2^n/S_r + (1-S_r/2^n)\log \frac{1}{1-S_r/2^n}} \\ &\geq \frac{n}{n(1-h(\rho))2^{(h(\rho)-1)n} + 2^{(h(\rho)-1)n}} \\ &\geq \frac{2^{(1-h(\rho))n}}{1-h(\rho)+\frac{1}{n}} \end{aligned} \tag{6.10}$$

which is much better than (6.7).

We summarize our findings

**Theorem 2.** *For the separation problem with Hamming balls for* $0 < \rho \leq 1/2$ *and* $r = \rho n$

(i) $M(n,r) \leq 1 + 4\sqrt{2}n^{3/2} \cdot 2^{(1-h(\rho))n}$
(ii) $M(n,r) \geq \frac{1}{1-h(\rho)+\frac{1}{n}} \cdot 2^{(1-h(\rho))n}$.

**Remark 4:**

(a) If we use the Covering Lemma for the separation problem in Section 5 for the hypergraph $(\mathcal{V}, \mathcal{E})$, where

$$\mathcal{V} = \{(x,y) : x,y \in \mathcal{X} = \{1,2,\ldots,n\}, x \neq y\}$$
$$\mathcal{E} = \left\{E : E \subset \mathcal{V}, E = E_A = \{(x,y) : |\{x,y\} \cap A| = 1\} \text{ for some } A \in \binom{[n]}{k}\right\},$$

$|E| = k \cdot (n-k)$, $\deg(v) = 2\binom{n-2}{k-1}$, $|\mathcal{V}| = 2\binom{n}{2}$, $|\mathcal{E}| = \binom{n}{k}$
then we obtain

$$m(n,k) \leq \frac{\binom{n}{k}}{2\binom{n-2}{k-1}} \log 2\binom{n}{2} \leq \frac{n^2}{k(n-k)} \log n,$$

which is by $\frac{n}{n-k}$ worse than the old results.

(b) On the other hand we don't know how to handle the second problem in Hamming space by clouds as in our first approach. Choosing columns at random how do we get **balls into the rows**?

These two observations show that there is more to be understood about the **interplay of covering and search**, eventually giving better results in one of the two areas by coming from the other!

## 7 Other directions: sequential search, guessing, inspections

In the model considered here the search space $\mathcal{X}$ carries no probability distribution $P$. It would not make any difference for the task, anyhow. However, sources $(\mathcal{X}, P)$ are considered in noiseless coding or, what is equivalent, sequential (also called adaptive) search. We propose to study this also **under cardinality constraints $k$ on the tests** with the expected search time as performance criterion.

Actually in his unpublished "Guessing and exponential entropy", (Nov. 15, 1993 according to E. Arikan [6]) James Massey considered the problem of guessing the value of a realization of a random variable $X$ by asking questions of the form "Is $X$ equal to its $i$th possible value?" until the answer is yes.

Notice that except for the wording this is just our problem for $k = 1$!

Other constraints on tests have been discussed in [4], [5]. We draw here especially attention to linear search problems (or alphabetical noiseless coding). They also should be studied under an additional cardinality constraint on the tests.

Perhaps an even more important observation is that **guessing** is also a special case of what has been called in Part 4 of [4], [5] search problems with inspections or **inspections** in short. The model has the following ingredients:

(1) a search space $(\mathcal{X}, P)$
(2) $c(j,k) \in \mathbb{R}^+$, $(j \in \mathbb{N}, 1 \leq k \leq n)$ the costs of the $j$th inspection of object $k$
(3) $q(j,k) \in [0,1]$ $(j \in \mathbb{N}, 1 \leq k \leq n)$ the probability that the object $k$, the true one, is found as such exactly in its $j$th inspection.

This model covers a wide range of practical problems. Some of them are mentioned in [4], [5], where also references to the pioneering works can be found.

Now just notice that a very special case ($c(1,k) = q(1,k) = 1$ for $1 \leq k \leq n$) corresponds to guessing!

Thus guessing comes up as special case of two models. There are by now also many results on guessing (also with variations of the original task like the incorporation of distortion criteria). Obviously, interactions between the areas described should be very challenging and fruitful.

## 8 Appendix: Improvements

One can consider $(m, n)$-separating systems $(A_1, \ldots, A_m)$ with the stronger constraint

$$|A_i| = k \quad \text{for} \quad i = 1, 2, \ldots, m. \tag{8.1}$$

We speak here also about a $k$-uniform separating system.

For given $n, k$ let $\bar{m}(n, k)$ be the minimal $m$ for which a $k$-uniform $(m, n)$-separating system exists. Clearly,

$$\bar{m}(n, k) \geq m(n, k) \geq \underline{m}(n, k). \tag{8.2}$$

Gyula Katona proved in [9] two remarkable theorems, which we now present.

Clearly, $k$-uniform $(m, n)$-separating systems correspond to $M_{mn}$ matrices with the properties

(a) the elements are 0 or 1
(b) each row contains $k$ ones
(c) no two columns are identical.

**Theorem $\mathbf{K}_1$.** *Let $m, n$, $1 \leq k \leq n/2$, $s_0, s_1, \ldots, s_m$ be fixed non-negative integers. Then there is an $M_{mn}$ matrix with the properties (a), (b) and (c), in which $s_i$ is the number of columns containing $i$ ones, if and only if*

*(1)* $mk = \sum_{i=1}^{m} i s_i$
*(2)* $n = \sum_{i=0}^{m} s_i$
*(3)* $s_i = \binom{m}{i}$ *for* $i = 0, 1, \ldots, m$

The point of this theorem is that the (obviously) necessary conditions (1)-(3) are also sufficient.

**Corollary $\mathbf{K}_1$.** *$\bar{m}(n, k)$ is equal to the least number $m$ for which there exists a system of non-negative integers, $s_0, s_1, \ldots, s_m$ satisfying conditions (1)-(3).*

**Theorem $\mathbf{K}_2$.** *If for $k < n/2$ $(A'_1, \ldots, A'_m)$ is an $(m, n, k)$-separating system, then there exists an $(m, n)$-separating system $(A_1, \ldots, A_m)$, which is $k$-uniform.*

**Corollary $\mathbf{K}_2$.** *$\bar{m}(n, k) = m(n, k)$.*

Actually, we observed that an even stronger result holds.

**Theorem 3.** *If for $k < n/2$ $(A''_1, \ldots, A''_m)$ is an $(m, n)$-separating system with*

$$\sum_{i=1}^{m} |A''_i| \leq mk$$

*then there exists an $(m, n)$-separating system $(A_1, \ldots, A_m)$ which is $k$-uniform.*

Consequently we have also

**Corollary.** $\underline{m}(n, k) = \bar{m}(n, k) = m(n, k)$.

**Proof:** This is exactly what is shown – but not stated, because the concept of an average constraint was not present – in the proof of Theorem $K_2$!

Furthermore, this result also can be proved by our expurgation technique by not only guaranteeing $|A_i| \leq k = pn$ but **simultaneously**, also $n - |A_i| \leq n - k = (1 - p)n$ and thus $|A_i| = k$.

For this just replace the probability $Kme^{-E(p,\varepsilon)n}$ by $Kme^{-E(p,\varepsilon)n} + Kme^{-E(1-p,\varepsilon)n}$.

**Remark 5:** Notice that the proof of Theorem 1 can be altered. The expurgation used to derive (iii) from (ii) can be replaced by the Corollary.

**Problem:** We are now curious whether our entropy upper bound (ii) in Theorem 1 can also be derived by Katona's characterisation of $m(n, k)$ in terms of a system of inequalities and intend to return to this question as soon as time permits.

**Remark 6:** From Theorems $K_1$, $K_2$ Katona derives the upper bound $m(n, k) \leq \left\lceil \frac{\log 2n}{\log n/k} \right\rceil \frac{n}{k}$ and compares it with the lower bound $\frac{\log n}{h(\frac{k}{n})} \leq m(n, k)$. He notices the similarity of these formulas and gives estimates on their ratios. He addresses on page 193 the dependence $k = cn$ as the **most important case**. However, unfortunately – apparently due to some error in calculation –, he concludes "... it is not difficult to show that the lower estimation is not even asymptotically the best ...".

In the paper "Search with small sets in presence of a liar", Katona returns in Section 2 "Improvements for the case of zero lies" to the issue of estimating $m(n, k)$, which is $f(n, k)$ in his terminology.

**He considers the ranges $k = \kappa n^{\alpha}, \alpha < 1$.** To be specific we quote his results.

**Theorem K 2.3.** *Let the integer $2 \leq R$ and the real number*

$$\kappa \geq \frac{R}{R!^{1/R}}$$

*be fixed. Then*

$$f(n, \kappa n^{1-\frac{1}{R}}) = \gamma n^{\frac{1}{R}} + O(1)$$

*where $\gamma$ is the only real solution of the equation*

$$\kappa\gamma + \frac{\gamma^R}{R!} = R + 1$$

*and $O(1)$ does not depend on $n$, but may depend on $R$ and $\kappa$. On the other hand, if*

$$\kappa < \frac{R}{R!^{1/R}}$$

*holds, then*

$$f(n, \kappa n^{1-\frac{1}{R}}) = \frac{R}{n} n^{\frac{1}{R}} + O(1)$$

*is the approximate solution.*

**Another perspective**

In the same paper Katona is led to study 1-error correcting codes $(u_1, \ldots, u_n)$ for which the matrix $U$ with columns $u_i = \binom{u_{i1}}{\cdots}{u_{im}}$ has rows with constraints on the number of letters (0 and 1 in his case). Notice that for codes of fixed composition the constraints are on the columns.

Notice also that frequency counts in rows arise in the 1-dimensional marginal distributions of the uniform distribution on the set of codewords ("Fano-sources") and for instance also in the derivation of Plotkin's bound.

The row constraints can be imposed – if practically feasable or useful – not only on $t$-error correcting codes, but also in Shannon's probabilistic theory of transmission over noisy channels.

Furthermore, they can be imposed also on codes in multi-user transmission theory and even in our general theory of information transfer, especially, for the theory of identification.

Many coding theorems can be improved to meet such constraints using double-exponentially large deviational estimates like we gave for codes generated via permutations.

## 9 Appendix: on the q-ary case

Instead of partitions $(A_i, A_i^c)$ considered are now partitions of $\mathcal{X} = \{1, 2, \ldots, n\}$ into $q$ sets $\vec{A}_i = (A_{i1}, \ldots, A_{iq})$ with associated test function $T_i : \mathcal{X}^n \to Q = \{0, 1, 2, \ldots, q-1\}$, where

$$T_i(x) = t - 1 \quad \text{iff} \quad x \in A_{it}. \tag{9.1}$$

$(\vec{A}_1, \ldots, \vec{A}_m)$ is an $(m, n)$-separating system, if the associated matrix $A = (a_{ix})_{\substack{1 \le i \le m \\ 1 \le x \le n}}$ defined by

$$a_{ix} = t - 1 \quad \text{iff} \quad x \in A_{it} \tag{9.2}$$

has distinct columns.

As in the previous case ($q = 2$)

$$\log n \le \sum_{i=1}^{m} H(T_i). \tag{9.3}$$

With $P_{it} = \frac{|A_{it}|}{n}$ and $P_i = (P_{i1}, \ldots, P_{iq})$, the entropy $H(T_i)$ equals

$$-\sum_{t=1}^{q} P_{it} \log P_{it} = H(P_i)$$

(with the usual abuse of notation).

By convexity of entropy $\log n \le m\ H(\tilde{P})$, where $\tilde{P} = \frac{1}{m} \sum_{i=1}^{m} P_i$, and $\frac{\log n}{H(\tilde{P})} \le m_q(n, c) \triangleq \min\{m : \exists\ (m, n)\text{-separating system with } H(\tilde{P}) \le c\}$.

Thus we have the **entropy bound**

$$m_q(n, c) \ge \frac{\log n}{c}. \tag{9.4}$$

In the binary case for $c = h\left(\frac{k}{n}\right)$,

$$H(\tilde{P}) \le c \Leftrightarrow \frac{1}{m} \sum_{i=1}^{m} |A_i| \le k \quad \text{or} \quad \frac{1}{m} \sum_{i=1}^{m} |A_i^c| \ge n - k.$$

Now obviously we can derive

**Proposition.** $\frac{\log n}{c} \le m_q(n, c) \le \frac{\log n + o(n)}{c}$.

Indeed choose columns by a cloud random selection with $P_t m$ many $t$'s and $H(P) \leq c$. This results in a matrix with $P_t\ mn$ many $t$'s and $H(\tilde{P}) \leq c$.

Technically, the number of possible columns is now $\binom{m}{P_0 m, P_1 m, \ldots, P_{q-1} m} \sim q^{H(P)m+o(m)}$ and this quantity takes the role of $2^{h(p)m+o(n)}$ in the proof of (ii) in Theorem 1.

Now we go for $(m, n)$-separating systems which are $(k_0, k_1, \ldots, k_{q-1})$-uniform, meaning that every column contains $k_t$ many $t$'s $(0 \leq t \leq q-1)$.

Finally, by the expurgation techniques applied simultaneously for $k_t = P_t n$ $(0 \leq t \leq q-1)$ as earlier for $k = pn$ and $n - k = (1-p)n$, we get now (in obvious notation)

**Theorem 4.** $\bar{m}_q(n; k_0, k_1, \ldots, k_{q-1}) = \frac{\log n + o(\log n)}{H(P)}$, *if* $P_t = \frac{k_t}{n}$ *and* $P = (P_0, \ldots, P_{q-1})$.

## References


[1] R. Ahlswede, Elimination of correlation in random codes for arbitrarily varying channels, Z. Wahrscheinlichkeitstheorie und verw. Geb. 44, 159-175, 1978.

[2] R. Ahlswede, Coloring hypergraphs: A new approach to multi-user source coding, Part I, Journ. of Combinatorics, Information and System Sciences, Vol. 4, No. 1, 76-115, 1979.

[3] R. Ahlswede, Coloring hypergraphs: A new approach to multi-user source coding, Part II, Journ. of Combinatorics, Information and System Sciences, Vol. 5, No. 3, 220-268, 1980.

[4] R. Ahlswede and I. Wegener, Suchprobleme, Teubner Verlag, Stuttgart (Russian Edition with Appendix by Maljutov 1981), 1979.

[5] R. Ahlswede and I. Wegener, Search Problems, Engl. Edition of [4] with Supplement of recent Literature, Wiley-Interscience Series in Discrete Mathematics and Optimization, R.L. Graham, J.K. Leenstra, R.E. Tarjan, edit., 1987.

[6] E. Arikan, On the average number of guesses required to determine the value of a random variable, Proc. 12th Prague Conf. on Information Theory, Statistical Decision Functions and Rendom Processes, 20-23, 1994.

[7] L.A. Bassalygo, S.I. Gelfand, and M.S. Pinsker, Coding for channels with localized errors, Proc. 4th Joint Swedish-Soviet Workshop Inf. Theory, Sweden, 95-99, 1989.

[8] M.G. Karpovsky, K. Chakrabarty, L.B. Levitin, D.R. Avresky, On the covering of vertices for fault diagnosis in hypercubes, Information Processing Letters 69, 99-103, 1999.

[9] G. Katona, On separating systems of a finite set, Journal of Combinatorial Theory 1, 174-194, 1966.

[10] A. Rényi, On the theory of random search, Bull. Americal Math. Soc. 71, 809-828, 1965.

[11] A. Rényi, Lectures on the theory of search, University of North Carolina, Chapel Hill, Institute of Statistics, Mimeo Ser. No. 6007, 1969.

[12] I. Wegener, On separating systems whose elements are sets of at most $k$ elements, Discrete Mathematics 28, 219-222, 1979.