JOURNAL OF COMBINATORIAL THEORY, Series B 61, 167–177 (1994)

The Asymptotic Behaviour of Diameters in the Average

RUDOLF AHLSWEDE AND INGO ALTHÖFER

Fakultät für Mathematik, Universität Bielefeld, Postfach 100 131, 33501 Bielefeld 1, Germany

Received October 21, 1991

In 1975 R. Ahlswede and G. Katona posed the following average distance problem (*Discrete Math.* 17 (1977), 10): For every cardinality $a \in \{1, ..., 2^n\}$ determine subsets A of $\{0, 1\}^n$ with $\# A = a$, which have *minimal* average *inner* Hamming distance. Recently I. Althöfer and T. Sillke (*J. Combin. Theory Ser. B* **56** (1992), 296–301) gave an exact solution of this problem for the central value $a = 2^{n-1}$. Here we present nearly optimal solutions for $a = 2^{\lambda n}$ with $0 < \lambda < 1$: Asymptotically it is not possible to do better than choosing $A_n = \{(x_1, ..., x_n) \mid \sum_{t=1}^n x_t = \lfloor \alpha n \rfloor\}$, where $\lambda = -\alpha \log \alpha - (1-\alpha) \log(1-\alpha)$. Next we investigate the following more general problem, which occurs, for instance, in the construction of good write-efficient-memories (WEMs). Given any finite set M with an arbitrary cost function $d: M \times M \to \mathbb{R}$, the corresponding sum type cost function $d_n: M^n \times M^n \to \mathbb{R}$ is defined by $d_n((x_1, ..., x_n), (y_1, ..., y_n)) = \sum_{t=1}^n d(x_t, y_t)$. The task is to find sets A_n of a given cardinality, which minimize the average inner cost $(1/(\# A_n)^2) \sum_{a \in A_n} \sum_{a' \in A_n} d_n(a, a')$. We prove that asymptotically optimal sets can be constructed by using "mixed typical sequences" with at most two different local configurations. As a non-trivial example we look at the Hamming distance for $M = \{1, ..., m\}$ with $m \geq 3$. © 1994 Academic Press, Inc.

1. $\{0, 1\}^n$ AND THE HAMMING DISTANCE

For two elements $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ in $\{0, 1\}^n$ the Hamming distance is defined by

$$d(x, y) = \# \{t \mid x_t \neq y_t\}.$$

For a set $A \subset \{0, 1\}^n$ the average inner distance is defined by

$$\bar{d}(A) = \frac{1}{(\# A)^2} \sum_{x \in A} \sum_{y \in A} d(x, y).$$

An important example is given by the set $A = \{x \mid w(x) = k\}$, where $w(x) = \# \{t \mid x_t = 1\}$. By symmetry of A its average inner distance is

$$\bar{d}(A) = \frac{1}{\binom{n}{k}} \sum_{i=0}^{k} \binom{n-k}{i} \binom{k}{k-i} 2i = \frac{2k(n-k)}{n},$$

0095-8956/94 $6.00
Copyright © 1994 by Academic Press, Inc.
All rights of reproduction in any form reserved.

so $\bar{d}(A)/n = 2k/n(1-k/n)$. Let

$$\bar{d}_n(a) = \min_{A \subset \{0,1\}^n:\ \#A = a} \bar{d}(A) \qquad \text{for all} \quad a \in \{0, 1, ..., 2^n\}.^1$$

We derive asymptotically tight bounds for $\bar{d}_n(a)$, when $a \approx \binom{n}{\alpha n}$ with a constant $\alpha \in (0, \frac{1}{2})$. In this section we show

THEOREM 1.1. *Let $(a_n)_{n=1}^{\infty}$ be a sequence of natural numbers with $0 \leqslant a_n \leqslant 2^n$ for all n and $\lim_{n \to \infty} \inf(a_n / (\binom{n}{\lfloor \alpha n \rfloor})) > 0$ for some constant α, $0 < \alpha < \frac{1}{2}$. Then*

$$\lim_{n \to \infty} \inf \frac{\bar{d}_n(a_n)}{n} \geqslant 2\alpha(1-\alpha).$$

The optimally of this bound follows from the example above. Note also that for $a = 2^{n-1}$ the subcube—and not the sphere—is the best configuration [5].

Proof of Theorem 1.1. The key idea in the proof is to generalize the problem by studying probability distributions on $\{0, 1\}^n$ with a given entropy instead of sets with a given cardinality. Let $P = (P(x))_{x \in \{0,1\}^n}$ be a probability distribution on $\{0, 1\}^n$. The average inner distance of P is defined by

$$\bar{d}(P) = \sum_{x \in \{0,1\}^n} \sum_{y \in \{0,1\}^n} P(x)\, P(y)\, d(x, y).$$

The entropy of P is given by

$$H(P) = \sum_{x \in \{0,1\}^n} -P(x) \log P(x).$$

(In this note we take the logarithm with base 2.)

We have $0 \leqslant H(P) \leqslant n$ for every distribution P on $\{0, 1\}^n$. Let

$$\hat{d}_n(\lambda) = \min \bar{d}(P),$$

where the min is taken over all P with $H(P) \geqslant \lambda$. A local exchange argument shows that this minimum is assumed for a distribution P with $H(P) = \lambda$.

[1] In this paper d, $\bar{d}$, $\bar{d}_n$, etc. are functions related to distances or cost functions. When the same symbol is used for more than one function, their differences are made clear by the symbols used for the arguments (sets, integers, or probability distributions). As a benefit for this loose notation the reader is not burdened with too many symbols.

LEMMA 1.2. *Let* $(H_n)_{n=1}^{\infty}$ *be a sequence of real numbers with* $0 \leqslant H_n \leqslant n$ *for all* $n \in \mathbb{N}$ *and* $\lim_{n\to\infty} \inf(H_n/n) \geqslant \lambda$ *for some constant* $\lambda \in (0, 1]$. *Then*

$$\lim_{n\to\infty} \inf \frac{\bar{d}_n(H_n)}{n} \geqslant 2\alpha(1-\alpha),$$

where $\alpha \in (0, \frac{1}{2})$ *with* $\lambda = h(\alpha) := -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$.

The theorem can be derived from this lemma in the following way.

A set $A \subset \{0, 1\}^n$ corresponds in a natural way to the probability distribution P_A, given by

$$P_A(x) = \begin{cases} \frac{1}{\# A}, & \text{if } x \in A \\ 0, & \text{if } x \notin A. \end{cases}$$

We have

$$\bar{d}(A) = \bar{d}(P_A)$$

and

$$H(P_A) = \log \# A.$$

Theorem 1.1 follows from Lemma 1.2, as

$$\frac{1}{n+2} 2^{nh(\lfloor \alpha n \rfloor/n)} \leqslant \binom{n}{\lfloor \alpha n \rfloor} \leqslant 2^{nh(\lfloor \alpha n \rfloor/n)}$$

for all $0 \leqslant \alpha \leqslant 1$ and all n (see [6, pp. 284, 285] for an elementary proof). Hence

$$\lim_{n\to\infty} \inf \frac{a_n}{\binom{n}{\lfloor \alpha n \rfloor}} > 0$$

implies

$$\lim_{n\to\infty} \inf \frac{\log a_n}{n} \geqslant h(\alpha).$$

It remains to prove the lemma.

Proof of Lemma 1.2. For a probability distribution P on $\{0, 1\}^n$ we define marginal one-probabilities

$$p_t = \sum_{x \in \{0,1\}^n : x_t = 1} P(x) \qquad \text{for} \quad t = 1, ..., n.$$

From the properties of the entropy function [6, p. 28] it follows that

$$H(P) \leqslant \sum_{t=1}^{n} h(p_t) = \sum_{t=1}^{n} -p_t \log p_t - (1-p_t) \log (1-p_t), \tag{1.1}$$

where equality holds iff P is the product of n distributions $(1-p_t, p_t)$ on $\{0, 1\}$. For the average inner distance of P we have

$$\bar{d}(P) = \sum_{t=1}^{n} 2p_t(1-p_t); \tag{1.2}$$

hence it is completely determined by the p_t.

The problem of minimizing $\bar{d}(P)$ for a fixed entropy level $H(P)$ is equivalent to maximizing $H(P)$ for a fixed distance level $\bar{d}(P)$. Thus by (1.1) and (1.2) it is sufficient to solve the following analytical problem. For $f(p_1, p_2, ..., p_n) = \sum_{t=1}^{n} h(p_t)$ find

$$\max_{0 \leqslant p_t \leqslant 1 \text{ for } t=1,...,n} f(p_1, ..., p_n) \qquad \text{under the constraint}$$

$$\sum_{t=1}^{n} 2p_t(1-p_t) = 2\alpha(1-\alpha)\, n. \tag{1.3}$$

By the symmetry of $h(p)$ and $p(1-p)$ in p and $(1-p)$ we may assume without loss of generality that $0 \leqslant p_t \leqslant \frac{1}{2}$ for all t.

The statement of the lemma suggests that the solution of (1.3) is to choose $p_t = \alpha$ for all t. This will be proved below by a simple exchange argument between only two coordinate:

$$\text{Find} \quad \max_{0 \leqslant p_1, p_2 \leqslant 1/2} f(p_1, p_2) \quad \text{under the constraint} \quad g(p_1, p_2)$$

$$= 2p_1(1-p_1) + 2p_2(1-p_2)$$

$$= c \qquad \text{for some constant } c \in [0, 1]. \tag{1.4}$$

Claim. For every constant $c \in [0, 1]$, (1.4) is solved by choosing $p_1 = p_2$.

Proof of the claim. A necessary condition for an inner point $(p_1, p_2) \in (0, \frac{1}{2})^2$ to be at least a local (maximum or minimum) solution of (1.4) is that

$$k_\kappa(p_t) := \log(1-p_t) - \log p_t - \kappa(1-2p^t) = 0 \qquad \text{for} \quad t = 1, 2,$$

where $\kappa \in \mathbb{R}$ is a Lagrange multiplier.

For $\kappa \leqslant 2$, $k_\kappa(\cdot)$ is strictly positive for all $p \in (0, \frac{1}{2})$. For every $\kappa > 2$ there exists some $p^*(\kappa) \in (0, \frac{1}{2})$ such that

$$k_\kappa(p) \begin{cases} <0, & \text{if } p^*(\kappa) < p < \frac{1}{2}, \\ =0, & \text{if } p^*(\kappa) = p, \\ >0, & \text{if } 0 \leqslant p < p^*(\kappa). \end{cases}$$

Hence the only candidates for local solutions of (1.4) are inner points (p_1, p_2) with $p_1 = p_2$ or boundary points which are of the form $(0, p)$ for $c \leqslant \frac{1}{2}$, or $(p, \frac{1}{2})$ for $c \geqslant \frac{1}{2}$.

$h' := dh/dp$ is continuous in p in the interval $(0, \frac{1}{2}]$. As $h'(0) = +\infty$ and $h'(p) < +\infty$ for all $p \in (0, \frac{1}{2}]$, (1.4) has a local minimum at $(0, p)$. Hence for $c \leqslant \frac{1}{2}$ (1.4) is solved by the point (p_1, p_2) with $p_1 = p_2$.

For $c \in [\frac{1}{2}, 1]$ let $p_c, q_c \in [0, \frac{1}{2}]$ be the real numbers satisfying $g(p_c, p_c) = c = g(q_c, \frac{1}{2})$.

We define

$$\tilde{f}(c) = f(p_c, p_c) - f(q_c, \tfrac{1}{2}).$$

As (p_c, p_c) and $(q_c, \frac{1}{2})$ are the only candidates for a solution of (1.4), we are done if $\tilde{f}(c)$ is non-negative for all $c \in [\frac{1}{2}, 1]$.

$\tilde{f}(\frac{1}{2}) > 0$, $\tilde{f}(1) = 0$, and $\tilde{f}$ is continuous in c. If there were some $c \in (\frac{1}{2}, 1)$ with $\tilde{f}(c) < 0$, there would have to be another parameter $c^* \in (\frac{1}{2}, c)$ with $\tilde{f}(c^*) = 0$. But it cannot be that (p_{c^*}, p_{c^*}) and $(q_{c^*}, \frac{1}{2})$, $(\frac{1}{2}, q_{c^*})$ are the only candidates for min or max solutions of (1.4), if $f(p_{c^*}, p_{c^*}) = f(q_{c^*}, \frac{1}{2})$.

This completes the proof of both the claim and the lemma. ∎

Next we extend the analytical method and generalize Theorem 1.1.

2. Arbitrary Sets M and Sum Type Cost Functions

In Section 1 we have investigated the problem of minimizing the average inner distance of subsets of $\{0, 1\}^n$ of a given cardinality. This is only a special case of the following more general problem:

Let $M = \{1, ..., m\}$ be a finite set and let $d: M \times M \to \mathbb{R}$ be an arbitrary real-valued cost function. For every $n \in \mathbb{N}$ the corresponding sum type cost function $d_n: M^n \times M^n \to \mathbb{R}$ is defined by

$$d_n(x, y) = \sum_{t=1}^{n} d(x_t, y_t)$$

for all $x = (x_1, ..., x_n)$, $y = (y_1, ..., y_n) \in M^n$.

For a set $A \subset M^n$ the average inner cost is defined by

$$\bar{d}_n(A) = \frac{1}{(\# A)^2} \sum_{x \in A} \sum_{y \in A} d_n(x, y),$$

and for every $a \in \{1, ..., m^n\}$ we define

$$\bar{d}_n(a) = \min_{A \subset M^n:\ \# A = a} \bar{d}_n(A).$$

We are interested in good bounds for the function $\bar{d}_n$.
This average inner cost plays an important role, for instance, in the design of good WEM-codes [4].

For the presentation of the general result on asymptotically optimal configurations of cardinality $\approx 2^{\lambda n}$, $0 < \lambda < \log m$, we need a notation of *typical sequences*. Let $P = (P(1), ..., P(m))$ be a probability distribution on M. A tuple $(x_1, ..., x_n) \in M^n$ is of *type* P, if $\# \{t \mid x_t = i\} = P(i)\, n$ for all $i \in \{1, ..., m\}$. Let $T_n(P) = \{x \in M^n \mid x$ has type $P\}$. Then $\# T_n(P) \approx 2^{H(P)\, n}$, if $T_n(P) \neq \emptyset$. See, for instance, Refs. [6, Section 12.1; 7] for a more detailed introduction and discussion of typical sequences.

Consider a constant v, $0 \leqslant v \leqslant 1$, and two probability distributions P and P' on M. The element $(x_1, ..., x_n) \in M^n$ is said to be of *mixed type* $(vP, (1-v)\, P')$, if

$$(x_1, ..., x_{\lfloor vn \rfloor}) \qquad \text{is of type } P$$

and

$$(x_{\lfloor vn \rfloor + 1}, ..., x_n) \qquad \text{is of type } P'.$$

As an example consider $M = \{0, 1\}$ and the level set $A = \{x \mid w(x) = k\}$. Let $P = ((n-k)/n, k/n)$ and let P' be any other distribution on $\{0, 1\}$. All elements of A are of the mixed type $(vP, (1-v)\, P')$ with $v = 1$.

Let $T_n(v, P, P') = \{x \in M^n \mid x$ is of the mixed type $(vP, (1-v)\, P')\}$. Then $\# T_n(v, P, P') \approx 2^{H(P)\, vn + H(P')(1-v)\, n}$.

THEOREM 2.1. *Fix* $M = \{1, 2, ..., m\}$ *and a cost function* $d\colon M \times M \to \mathbb{R}$. *For every* λ, $0 < \lambda \leqslant \log m$, *there exists a mixed type* (v, P, P') *with* $vH(P) + (1-v)\, H(P') = \lambda$, *such that*

$$\limsup_{n \to \infty} [\bar{d}_n(T_n(v, P, P')) - \bar{d}_n(2^{\lambda n})] < +\infty.$$

In case of $\lim_{n \to \infty} \bar{d}_n(2^{\lambda n}) \in \{\pm \infty\}$ *this means*

$$\lim_{n \to \infty} \frac{\bar{d}_n(T_n(v, P, P'))}{\bar{d}_n(2^{\lambda n})} = 1.$$

In other words, the set $T_n(\nu, P, P')$ have asymptotically minimal average inner cost.

Proof. As in the proofs of Section 1 we start by generalizing the problem to probability distributions Q on M^n, defining average inner cost $\bar{d}(Q)$ and replacing the cardinality condition by a lower bound on the entropy $H(Q)$. Given $q_t(k) = \sum_{x: x_t = k} Q(x)$ for all $t \in \{1, ..., n\}$, $k \in M$, we have

$$\bar{d}(Q) = \sum_{t=1}^{n} \left[\sum_{k=1}^{m} \sum_{l=1}^{m} q_t(k)\, q_t(l)\, d(k, l) \right]$$

and

$$H(Q) \leqslant \sum_{t=1}^{n} H(q_t(1), ..., q_t(m)).$$

In the last line equality holds iff Q is the product of its n one-dimensional marginal distributions (see [6, p. 28] for a proof). For a fixed $\lambda \in [0, \log m]$, we want to solve the following analytical optimization problem: Find

$$\min \bar{d}(Q) \qquad \text{under the constraint} \quad \sum_{t=1}^{n} H(q_t(1), ..., q_t(m)) \geqslant \lambda n, \tag{2.1}$$

where each tuple $(q_t(1), ..., q_t(m))$ is a probability distribution on M.

Our goal is to show that (2.1) is solved approximately by a combination of at most two different distributions P and P' on M, taking P for the first $\lfloor \nu n \rfloor$ coordinates and P' for the other $n - \lfloor \nu n \rfloor$ coordinates. Of course P, P', and ν will depend on λ. We start with

LEMMA 2.2. *Consider real numbers $x_1, ..., x_n, y_1, ..., y_n$, and a probability distribution $(\mu_1, ..., \mu_n)$ on $N = \{1, ..., n\}$. Then there exist two elements $j, k \in \{1, ..., n\}$ and some $\mu \in [0, 1]$, such that*

$$\mu x_j + (1 - \mu)\, x_k \leqslant \bar{x} := \sum_{t=1}^{n} \mu_t x_t \tag{2.2}$$

and

$$\mu y_j + (1 - \mu)\, y_k \geqslant \bar{y} := \sum_{t=1}^{n} \mu_t\, y_t.$$

Proof of Lemma 2.2. We replace x_i by $x_i - \bar{x}$ and y_i by $y_i - \bar{y}$. So the desired inequalities are

$$\mu x_j + (1 - \mu)\, x_k \leqslant 0,$$

$$\mu y_j + (1 - \mu)\, y_k \geqslant 0.$$

Consider the set of points

$$S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

in the plane with the center of mass at the origin. The convex hull of this set is a convex polygon with the origin as an interior point. Some edge of the polygon must lie at least partially in the second quadrant and this give the result. ∎

We are grateful to an unknown referee for this concise proof. Our original proof can be found in [1].

For the next step consider a compact set $K \subset \mathbb{R}^m$, continuous functions $f, g: K \to \mathbb{R}$, and for all $n \in \mathbb{N}$ the optimization problem

$$\min_{(z_1, \ldots, z_n) \in K^n} \sum_{t=1}^{n} f(z_t) \qquad \text{under the constraint}$$

$$\sum_{t=1}^{n} g(z_t) \geq cn, \qquad \text{where } c \in \mathbb{R} \text{ is some fixed constant.} \tag{2.3}$$

The next lemma shows how a solution of this problem can be approximated by solutions of the simpler form $(\tilde{z}_1, \tilde{z}_1, \ldots, \tilde{z}_1, \tilde{z}_2, \tilde{z}_2, \ldots, \tilde{z}_2)$.

LEMMA 2.3. *There exist $\tilde{z}_1, \tilde{z}_2 \in K$ and $v \in [0, 1]$, all depending on c, such that*

$$\lfloor vn \rfloor g(\tilde{z}_1) + (n - \lfloor vn \rfloor) g(\tilde{z}_2) \leq cn$$

and

$$\lfloor vn \rfloor f(\tilde{z}_1) + (n - \lfloor vn \rfloor) f(\tilde{z}_2) - \sum_{t=1}^{n} f(z^*_{t,n}) \leq |f(\tilde{z}_1) - f(\tilde{z}_2)|$$

for all $n \in \mathbb{N}$, where $(z^*_{1,n}, \ldots, z^*_{n,n})$ is an optimal solution of (2.3).

Proof. The optimization problem

$$\min_{(z_1, z_2) \in K^2,\ v \in [0,1]} [vf(z_1) + (1 - v) f(z_2)]$$

$$\text{under the constraint} \quad vg(z_1) + (1 - v) g(z_2) \geq c \tag{2.4}$$

has a solution, say $(\tilde{z}_1, \tilde{z}_2, v)$. (2.5)

For the existence of this solution the continuity of f and g is needed. Without loss of generality assume that $g(z_1) \leq g(z_2)$. Now fix $n \in \mathbb{N}$.

Putting $x_t = f(z^*_{t,n})$, $y_t = g(z^*_{t,n})$, and $\lambda_t = 1/n$ for $t = 1, ..., n$, we can apply Lemma 2.2 and see that there are $j, k \in \{1, ..., n\}$ and $\mu \in [0, 1]$, such that

$$\mu n f(z^*_{j,n}) + (1-\mu)\, n f(z^*_{k,n}) \leqslant \sum_{t=1}^{n} f(z^*_{t,n})$$

and

$$\mu n g(z^*_{j,n}) + (1-\mu)\, n g(z^*_{k,n}) \geqslant cn.$$

Thus by (2.4) and (2.5) we also have

$$\nu n f(\tilde{z}_1) + (1-\nu)\, n f(\tilde{z}_2) \leqslant \sum_{t=1}^{n} f(z^*_{t,n})$$

and

$$\nu n g(\tilde{z}_1) + (1-\nu)\, n g(\tilde{z}_2) \geqslant cn.$$

This completes the proof of Lemma 2.3. ▌

Let $\hat{d}_n(\lambda) = \min \bar{d}(Q)$, where the min is taken over all probability distributions Q on M^n with $H(Q) \geqslant n\lambda$.

For a mixed type (ν, P, P') we define the corresponding product distribution Q_n on M^n by the marginal probabilities

$$q_t(k) = \begin{cases} P(k), & \text{if } \ 1 \leqslant t \leqslant \lfloor \nu n \rfloor, \\ P'(k), & \text{if } \ \lfloor \nu n \rfloor < t \leqslant n, \end{cases}$$

for all $k \in M$.

LEMMA 2.4. *For every $\lambda \in [0, \log m]$ there exists a mixed type (ν, P, P'), such that*

$$\limsup_{n \to \infty} [\bar{d}(Q_n) - \hat{d}_n(\lambda)] \leqslant \max_{j,k \in M} d(j,k) - \min_{j,k \in M} d(j,k) < \infty$$

and

$$\lfloor \nu n \rfloor H(P) + (n - \lfloor \nu n \rfloor)\, H(P') \geqslant \lambda n.$$

Proof of Lemma 2.4. For probability distributions P on M we define two functions,

$$f(P) = \sum_{j=1}^{m} \sum_{k=1}^{m} P(j)\, P(k)\, d(j,k)$$

and

$$g(P) = H(P),$$

and apply Lemma 2.3. Obviously $|f(P)-f(P')| \leqslant \max d(j,k) - \min d(j,k)$ for all P, P'. This completes the proof of Lemma 2.4. ▮

Theorem 2.1 follows immediately, as $-c \leqslant \bar{d}_n(T_n(v, P, P')) - \bar{d}(Q_n) \leqslant c$ for all $n \in \mathbb{N}$, where the finite bound c depends only on m and $d: M \times M \to \mathbb{R}$. ▮

Theorem 1.1 shows that the special case of a degenerate optimal mixed type (v, P, P') with $v = 1$ occasionally occurs. Let us now apply Theorem 2.1 to a non-trivial example.

Choose $M = \{1, 2, 3\}$ and

$$d(x, y) = \begin{cases} 0, & \text{if } x = y, \\ 1, & \text{if } x \neq y. \end{cases}$$

Hence d_n is the Hamming distance again, but now for alphabet size 3 instead of size 2.

The results mentioned below have been found by computer runs. We omit the theoretical proofs which are due to J. Johnen. In the first step we have to understand the case $n = 1$.

Fact 2.5. Fix some $\lambda \in [0, \log 3]$. Among all distributions P on M with $H(P) = \lambda$ the distribution with minimal average inner cost is of the form $(q, (1-q)/2, (1-q)/2)$ with $q \geqslant (1-q)/2$.

Minimizing $\bar{d}_n$ for a given cardinality 2^{λ_n} is equivalent to maximizing the cardinality under the condition $\bar{d}_n \leqslant cn$. The computer results give

Fact 2.6. Among all subsets of $\{1, 2, 3\}^n$ with average inner cost $\leqslant cn$ the following ones have asymptotically maximal cardinality:

(i) $T_n(P)$, where $P = (q, (1-q)/2, (1-q)/2))$ with $q \geqslant (1-q)/2$ and $\bar{d}(P) = c$, if $0 \leqslant c \leqslant \frac{1}{2}$.

(ii) $T_n(v, P, P')$, where $P = (\frac{2}{3}, \frac{1}{6}, \frac{1}{6})$, $P' = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, and $v\bar{d}(P) + (1-v)\,\bar{d}(P') = c$, if $\frac{1}{2} \leqslant c \leqslant \frac{2}{3}$.

In the more general case with $M = \{1, ..., M\}$, $m \geqslant 3$, and

$$d(x, y) = \begin{cases} 0, & \text{if } x = y \\ 1, & \text{if } x \neq y, \end{cases}$$

our computer results indicate that the optimal solutions have the following structure.

Observation 2.7. Fix some $\lambda \in [0, \log m]$. Among all distributions P on M with $H(P) = \lambda$ the one with minimal average inner cost is of the form $(q, (1-q)/(m-1), ..., (1-q)/(m-1))$ with $q \geqslant (1-q)/(m-1)$.

Observation 2.8. For every $m \geq 3$ there is some threshold $c_m^* = (2m-3)/m(m-1)$ such that among all subsets of M^n with average inner cost $\leq cn$ the following ones have asymptotically maximal cardinality:

(i) $T_n(P)$, where $P = (q, (1-q)/(m-1), ..., (1-q)/(m-1))$ with $q \geq (1-q)/(m-1)$ and $\bar{d}(P) = c$, if $0 \leq c \leq c_m^*$.

(ii) $T_n(v, P, P')$, where $P = (q_m^*, (1-q_m^*)/(m-1), ..., (1-q_m^*)/(m-1))$ with $\bar{d}(P) = c_m^*$, $P' = (1/m, ..., 1/m)$, and $v\bar{d}(P) + (1-v)\,\bar{d}(P') = c$, if $c_m^* \leq c(m-1)/m$.

Acknowledgments

We are gratefully indebted to two anonymous referees for several helpful comments and to Professor Johannes Johnen for his proofs of Observation 2.7 and 2.8.

References

1. R. Ahlswede and I. Althöfer, The asymptotic behaviour of diameters in the average, Preprint 91-099 SFB 343, Bielefeld.
2. R. Ahlswede, N. Cai, and Z. Zhang, Diametric theorems in sequence spaces, *Combinatorica* **12**, No. 1 (1992), 1–17.
3. R. Ahlswede and G. Katona, Contributions to the geometry of Hamming spaces, *Discrete Math.* **17** (1977), 1–22.
4. R. Ahlswede and Z. Zhang, On write-efficient memories, *Inform. and Comput.* **83** (1989), 80–97.
5. I. Althöfer and T. Sillke, An "average distance" inequality for large subsets of the cube, *J. Combin. Theory Ser. B* **56** (1992), 296–301.
6. T. M. Cover and J. A. Thomas, "Elements of Information Theory," Wiley, New York, 1991.
7. I. Csiszár and J. Körner, "Information Theory: Coding Theorems for Discrete Memoryless Systems," Academic Press, New York, 1982.