COMPOSITIO MATHEMATICA

JOEL SPENCER

Sequences with small discrepancy relative to n events

Compositio Mathematica, tome 47, nº 3 (1982), p. 365-392

<http://www.numdam.org/item?id=CM_1982__47_3_365_0>

COMPOSITIO MATHEMATICA, Vol. 47, Fasc. 3, 1982, pag. 365–392

SEQUENCES WITH SMALL DISCREPANCY RELATIVE TO n EVENTS

Joel Spencer

§1. Statement

Let (Ω, μ) be a probability space, $X \subset \Omega$ measurable, and let $\omega = (\xi_1, \xi_2, \ldots)$ be an infinite sequence of points (not necessarily distinct) of Ω. We write $Z_X(t)$ for the number of $\xi_i \in X$, $1 \leq i \leq t$, and we define a discrepancy function

$$D_X(t) = Z_X(t) - t\mu(X) \tag{1.1}$$

Our object will be the following result, conjectured by R. Tijdeman and G. Wagner with an unspecified constant.

THEOREM A: *Let $A_1, \ldots, A_n \subset \Omega$, measurable. Then there exists a sequence ω so that*

$$|D_{A_i}(t)| \leq 1000\, n^{1/2} \ln(n+1) \tag{1.2}$$

for $1 \leq i \leq n$ and all positive integers t.

It has been shown (see §2) that Theorem A holds with the RHS of (1.2) replaced by $n + 1$. Thus Theorem A holds for all $n \leq 1000$ and we shall assume (often tacitly) $n > 1000$ throughout this paper. The constant "1000" in (1.2) might easily be improved but we do not attempt to do so in this paper.

The author wishes to thank Robert Tijdeman for bringing this problem to his attention, for correspondence and conversations, and for his aid in the preparation of this paper.

0010-437X/82090365–28$00.20/0

§2. Discussion and outline

We start with a trivial example. Let $\Omega = \{H, T\}$, $\mu(\{H\}) = \mu(\{T\}) = \frac{1}{2}$, $n = 2$, $A_1 = \{H\}$, $A_2 = \{T\}$. The "best" sequence is then $\omega = \{H, T, H, T, H, T, \ldots\}$, continuing periodically. For this ω $|D_{A_i}(t)| \leq \frac{1}{2}$ always. Note that if ω is a "ranndom sequence" then, with probability one, $D_{A_1}(t)$ will be unbounded.

In some special cases much better bounds than provided by Theorem A can be given.

(i) Suppose $A_1, \ldots, A_n$ are pairwise disjoint.

It was proven by Tijdeman [5] (see also Meijer [2] and Tijdeman [6]) that there exists a sequence ω so that $|D_{A_i}(t)| \leq 1$ for $1 \leq i \leq n$ and all positive integers t. Using this result and the reduction of §3 one can show immediately that Theorem A holds with the RHS replaced by $n + 1$.

(ii) Suppose $A_1 \subset A_2 \subset \ldots \subset A_n$.

This case corresponds to the classical situation of uniform distribution on a real interval $[0, 1)$ by identifying A_j with $[0, \mu(A_j))$. It follows from Theorem 2 of Tijdeman and Voorhoeve [7] that there exists a sequence ω so that $|D_{A_i}(t)| \leq \dfrac{\log(2n)}{2\log 2}$ for $1 \leq i \leq n$ and all positive integers t and from Corollary 2 of that paper that this result can be improved only by a constant factor.

(iii) Suppose each of the 2^n sets $B_1 B_2 \ldots B_n$ with $B_j \in \{A_j, A \backslash A_j\}$ is non-empty. For each $i \in \{1, \ldots, n\}$ define a sequence $\omega^{(i)} = \{\xi_k^{(i)}\}_{k=1}^{\infty}$ in A by $\xi_k^{(i)} \in A_i$ if and only if $[k\mu(A_i) + \frac{1}{2}] > [(k-1)\mu(A_i) + \frac{1}{2}]$. Define $\omega = \{\xi_k\}_{k=1}^{\infty}$ such that $\xi_k \in A_i$ if and only if $\xi_k^{(i)} \in A_i$ for all i and k. It is clear that such a sequence ω can be constructed and it is easy to see that $|D_{A_i}(t)| \leqslant \frac{1}{2}$ for $1 \leq i \leq n$ and all t.

Theorem A is proven through a series of reductions. Theorems 1 (§5), 2 (§7), 3 (§8) and 4 (§9) are successive reductions and, in a more formal presentation, would be presented in reverse order. In §4 a stronger version of Theorem A is proven with certain added assumptions. Here many of the basic ideas of the full proof are presented in simpler form. The paper is divided into thirteen sections as follows:

The proof of two inequalities is placed in an Appendix.

§3. Reduction to $n+1$ points

Fix $A_1, \ldots, A_n \subset \Omega$. For $x \in \Omega$ set

$$P(x) = (\varepsilon_1(x), \ldots, \varepsilon_n(x)) \in R^n \tag{3.1}$$

where

$$\varepsilon_i(x) = \begin{cases} 1 & \text{if } x \in A_i \\ 0 & \text{if } x \notin A_i \end{cases} \tag{3.2}$$

Set $\mu_i = \mu(A_i)$, $1 \leq i \leq n$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n) \in R^n$. Then $\boldsymbol{\mu}$ is the weighted average, or integral, of the $P(x)$ and so $\boldsymbol{\mu}$ is in the convex hull of the points $P(x)$. By Caratheodory's Theorem there exist $n+1$ of these points, which we may call $1, \ldots, n+1$ without loss of generality, and non-negative $p_1, \ldots, p_{n+1}$ with sum unity so that

$$\boldsymbol{\mu} = \sum_{i=1}^{n+1} p_i P(i) \tag{3.3}$$

If we set $\Omega^* = \{1, \ldots, n+1\}$ with $\mu^*(\{i\}) = p_i$ and $A_i^* = A_i \cap \Omega^*$ then $\mu^*(A_i^*) = \sum_{j=1}^{n+1} p_j \varepsilon_i(j) = \mu_i = \mu(A_i)$ for $1 \leq i \leq n$. A sequence ω^* in Ω^* that satisfies (1.2) for $A_1^*, \ldots, A_n^*, \Omega^* \mu^*$ will also satisfy it for $A_1, \ldots, A_n, \Omega, \mu$ since

$$D_{A_i^*}(t) = D_{A_i}(t), \quad 1 \leq i \leq n \tag{3.4}$$

This gives a reduction. Henceforth we will assume $\Omega = \{1, \ldots, n+1\}$ where $\mu(\{i\}) = p_i$.

§4. Proof when no p_i is extremely small

We here prove Theorem A under the additional hypothesis that all $p_i \geq 1/100n$. The proof of this result contains basic ideas that are unfortunately somewhat disguised in the full proof.

REMARK: If we are given, in the original formulation of §1, that $\mu_1, \ldots, \mu_n$ are not small – say even that $\mu_1 = \ldots = \mu_n = \frac{1}{2}$ – we cannot deduce, without further information, that the p_i defined in §3 will not be small.

Fix $p_1, \ldots, p_{n+1}$ with sum unity and $A_1, \ldots, A_n \subset \{1, \ldots, n+1\}$. Let t be an arbitrary, but fixed, positive integer. We first show that we can make all $|D_{A_i}(t)|$ small. For $1 \leq i \leq n+1$ set

$$\theta_i = \{p_i t\} \tag{4.1}$$

where $\{x\}$ represents the fractional part of x. (Observe that if all $\theta_i = 0$, i.e. all $p_i t$ are integral, we may simply let ω contain exactly $p_i t$ i's for each i and then every set would have zero discrepancy at t. When $\theta_i \neq 0$ we shall decide whether to place $[p_i t]$ or $[p_i t]$ i's.) For $1 \leq i \leq n+1$ let $\boldsymbol{\varepsilon}_i$ be independent random variables with

$$\Pr[\boldsymbol{\varepsilon}_i = +1] = \theta_i, \ \Pr[\boldsymbol{\varepsilon}_i = 0] = 1 - \theta_i \tag{4.2}$$

For any $X \subset \{1, \ldots, n+1\}$ set

$$\boldsymbol{S}_X = \sum_{i \in X} (\boldsymbol{\varepsilon}_i - \theta_i). \tag{4.3}$$

Then $\mathrm{E}[\boldsymbol{S}_X] = 0$. We require a result from probability theory (see Appendix). For every $\alpha > 0$,

$$\Pr[|\boldsymbol{S}_X| > \alpha] < 2\exp[-2\alpha^2/(n+1)] \tag{4.4}$$

We set

$$\alpha = 2^{-1/2}(n+1)^{1/2} \ln[2(n+1)]^{1/2} \tag{4.5}$$

so that this probability is less than $1/(n+1)$. The disjunction of the $n+1$ events "$|\boldsymbol{S}_X| > \alpha$" for $X = A_1, \ldots, A_n$ and $X = \Omega = \{1, \ldots, n+1\}$ has probability less than unity. With positive probability all these events are false. That is: there exist $\varepsilon_1, \ldots, \varepsilon_{n+1} \in \{0, 1\}$ such that

$$|\sum_{i\in X}(\varepsilon_i-\theta_i)|<\alpha \text{ for } X=A_1,\ldots,A_n,\Omega \tag{4.6}$$

Set

$$b=\sum_{i\in\Omega}\varepsilon_i-\sum_{i\in\Omega}\theta_i \tag{4.7}$$

so that $|b|<\alpha$, $b\in Z$. If $b>0$ change precisely b of the $\varepsilon_i=1$ to $\varepsilon_i=0$. If $b<0$ change precisely $|b|$ of the $\varepsilon_i=0$ to $\varepsilon_i=1$. Let ε_i^* represent the new ε_i. Then

$$\sum_{i\in\Omega}\varepsilon_i^*=\sum_{i\in\Omega}\theta_i \tag{4.8}$$

and for $X=A_1,\ldots,A_n$

$$|\sum_{i\in X}(\varepsilon_i^*-\theta_i)|\le|\sum_{i\in X}(\varepsilon_i-\theta_i)|+|\sum_{i\in X}(\varepsilon_i^*-\varepsilon_i)|\le 2\alpha. \tag{4.9}$$

Now we set

$$a_i=[p_it]+\varepsilon_i^* \tag{4.10}$$

for $1\le i\le n+1$.

We have found $a_1,\ldots,a_{n+1}$ so that

$$\text{each } a_i \text{ equals either } [p_it] \text{ or } [p_it]+1 \tag{4.11}$$

and such that if ω has precisely a_i i's in the first t symbols then

$$|D_{A_i}(t)|<2\alpha \text{ for } 1\leqslant i\le n+1. \tag{4.12}$$

Remark: Results very similar to those given above have been shown by Beck and Fiala [1], Spencer [4] and others.

We call the above procedure establishing a benchmark at t. It would be pleasant to simply say that we may establish a benchmark at every t and these determine the sequence. There is a problem of consistency. Let benchmarks at t, t' with $t<t'$ require a_i, a_i' appearances of symbol i. If $a_i>a_i'$ the benchmarks are inconsistent. Avoiding this problem is the heart of this paper.

Now assume all $p_i>1/100n$. (We have not yet used this fact.) We establish benchmarks at $t=0$, $100n$, $200n$, $300n,\ldots$ satisfying (4.11), (4.12). If t,t' are consecutive benchmarks then $t'-t=100n$ and hence

$$p_i t' - p_i t \geq 1 \tag{4.13}$$

This and (4.11) imply $a_i' \geq a_i$. Set

$$m_i = a_i' - a_i \tag{4.14}$$

At this point we introduce the Reduction Principle. It says, roughly, that having established two consecutive benchmarks at t, t' with numbers a_i, a_i', m_i as defined above it then suffices to find a sequence on $(t, t']$ which is "smooth" relative to the problem of placing m_i i's.

More precisely: Let a sequence ω be given. For $u \in (t, t']$ and $X \subset \Omega$ let

$$Z_X^*(u) \text{ be the number of } \xi_i \in X \text{ with } t < i \leq u, \tag{4.15}$$

$$p_i^* = m_i/(t' - t),$$

$$\mu^*(X) = \sum_{i \in X} p_i^*,$$

$$D_X^*(u) = Z_X^*(u) - (u - t)\mu^*(X)$$

We think of $D_X^*(u)$ as the relative discrepancy at u of event X, given the task of placing m_i i's in $(t, t']$. Write u as a convex combination of t, t'

$$u = \beta t + (1 - \beta)t' \tag{4.16}$$

An algebra calculation shows

$$D_X(u) = [\beta D_X(t) + (1 - \beta) D_X(t')] + D_X^*(u) \tag{4.17}$$

(Observe that the check (4.17) it suffices to consider $X = \{i\}$.) The Reduction Principle is (4.17). It says that the discrepancy may be found by interpolating the discrepancies at the benchmarks and adding the relative discrepancy. Hence, suppose at consecutive benchmarks t, t' all $|D_X(t)| \leq K$ and all $|D_X(t')| \leq K$. Suppose in the sequence in $(t, t']$ all relative discrepancies have $|D_X^*(u)| \leq L$. Then all $|D_X(u)| \leq K + L$. (By all we mean $X = A_1, \ldots, A_n$.)

The placement between benchmarks is given by the following result. Let a set of $n + 1$ symbols $\{1, \ldots, n + 1\}$, n subsets $A_1, \ldots, A_n \subset \{1, \ldots, n + 1\}$ and non-negative integers $m_1, \ldots, m_n$ with $m_1 + \ldots + m_n = M$ be given. Let $\omega = (\xi_1, \ldots, \xi_M)$ be a sequence of length M consisting of precisely m_i i's. Set

$$p_i = m_i/M, \qquad 1 \le i \le n$$
$$p_X = \sum_{i \in X} p_i, \qquad X \subset \{1,\ldots,n+1\}$$

For $1 \le u \le M$, $X \subset \{1,\ldots,n+1\}$ set

$$Z_X(u) = \text{the number of } \xi_i \in X,\ 1 \le i \le u \tag{4.19}$$
$$D_X(u) = Z_X(u) - up_X$$

Local Placement Theorem: *Under the above definitions there exists ω so that*

$$|D_X(u)| \le [\tfrac{1}{2}M \ln(2nM)]^{1/2} \tag{4.20}$$

for all u, $1 \le u \le M$ and for all $X = A_1,\ldots,A_n$.

Proof: Consider a random placement of the symbols. That is, take a deck of M cards consisting of m_i cards of each symbol i and perform a random shuffle to give ω. Then $Z_X(u)$ has a hypergeometric distribution – given a deck of M cards of which Mp_X are marked, $Z_X(u)$ is the number of marked cards picked when u cards are selected without replacement.

$$\Pr\{|D_X(u)| > \alpha\} < \Pr\{|B(u, p_X) - up_X| > \alpha\} \tag{4.21}$$

(B denotes the Binomial Distribution) since the RHS represents the equivalent selection with replacement. Applying (4.4) (with all $\theta_i = p_X$)

$$\begin{aligned}\Pr\{|D_X(u)| > \alpha\} &< 2\exp[-2\alpha^2/u] \\ &< 2\exp[-2\alpha^2/M]\end{aligned} \tag{4.22}$$

We set

$$\alpha = [\tfrac{1}{2}M \ln(2nM)]^{1/2} \tag{4.23}$$

so that

$$\Pr\{|D_X(u)| > \alpha\} < 1/nM \tag{4.24}$$

This implies that there exists an $\omega = (\xi_1,\ldots,\xi_M)$ consisting of precisely m_i i's such that $|D_{A_i}(u)| \le \alpha$ for $u = 1,\ldots,M$ and $i = 1,\ldots,n$. This proves the theorem.

We apply the Local Placement Theorem with $M = 100n$, (4.12) for establishing benchmarks, and the Reduction Principle to hold things together. The result is that there exists an ω such that all $|D_X(u)|$, $X = A_1, \ldots, A_n$, $1 \le u < \infty$ are bounded by

$$2 \cdot 2^{-1/2}(n+1)^{1/2} \ln[2(n+1)]^{1/2} + [50n \ln(200n^2)]^{1/2}$$

which is of the order $[n \ln n]^{1/2}$. Note this result is stronger than (1.2) by a factor of order $[\ln n]^{1/2}$.

§5. Reduction to history

There are certain technical difficulties in finding a sequence that we avoid by changing to a continuous problem. A history ω on a set $\Omega = \{1, \ldots, n+1\}$ is a family of pairs (i, t) where $i \in \Omega$, $t \in R^+$ and the t's are distinct. For $i \in \Omega$, $t \in R$ we let

$Z_i(t)$ be the number of $(i, s) \in \omega$ with $s \le t$. (5.1)

Let $p_1, \ldots, p_{n+1}$ be fixed non-negative reals. We set

$$D_i(t) = Z_i(t) - p_i t \tag{5.2}$$

NOTATION: $\|.\|$ represents the max norm. That is

$$\|(x_1, \ldots, x_{n+1})\| = \max_{1 \le i \le n+1} |x_i|$$

Let $\vec{\mathbf{a}}_1, \ldots, \vec{\mathbf{a}}_{n+1} \in R^{n+1}$ be fixed with $\|\vec{\mathbf{a}}_i\| \le 1$ for all i. We set

$$\mathbf{D}(t) = D_1(t)\vec{\mathbf{a}}_1 + \ldots + D_{n+1}(t)\vec{\mathbf{a}}_{n+1} \tag{5.3}$$

THEOREM 1: *Under the above conditions there exists a history ω so that*

$$\|\vec{\mathbf{D}}(t)\| < 500n^{1/2} \ln n \tag{5.4}$$

for all $t \in R^+$.

We now show that Theorem A follows from Theorem 1. By the reduction of §3 we may consider $\Omega = \{1, \ldots, n+1\}$ with $\mu(\{i\}) = p_i$. We are given $A_1, \ldots, A_n \subset \Omega$ and we set $A_{n+1} = \Omega = \{1, \ldots, n+1\}$. For $i \in \Omega$ set

$$\mathbf{a}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{i,n+1}) \in R^{n+1}$$

where

$$\varepsilon_{ij} = \begin{cases} 1 & \text{if } i \in A_j, \\ 0 & \text{if not.} \end{cases}$$

For these $p_1, \ldots, p_{n+1}, \vec{\mathbf{a}}_1, \ldots, \vec{\mathbf{a}}_{n+1}$ we apply Theorem 1 to find ω satisfying (5.4). Given a history ω we may define a sequence ω^* given by ordering the points $(\xi_i, t_i) \in \omega$ in increasing order of t and setting $\omega^* = (\xi_1, \xi_2, \ldots)$.

Now we unravel the meaning of (5.4). The i-th coordinate of $\vec{\mathbf{D}}(t)$ is the discrepancy of A_i at t. That is,

$$|Z_{A_i}(t) - t\mu(A_i)| < 500n^{1/2} \ln n \tag{5.6}$$

where $Z_X(t)$ is the number of appearances, in ω, of symbols $i \in X$ in the interval $(0, t]$. Let $t_1, t_2, \ldots$ be the times for appearances in ω, in ascending order. Applying (5.6) with the additional set $A_{n+1} = \Omega$ shows that

$$|m - t_m| < 500n^{1/2} \ln n \tag{5.7}$$

for all positive integers m – i.e. the History and Sequence are nearly together. Thus for the sequence ω^*

$$\begin{aligned} &|Z^*_{A_i}(m) - m\mu(A_i)| \\ &\quad = |Z_{A_i}(t_m) - m\mu(A_i)| \\ &\quad \le |Z_{A_i}(t_m) - t_m\mu(A_i)| + |(t_m - m)\mu(A_i)| \\ &\quad \le 500n^{1/2} \ln n + 500n^{1/2} \ln n \end{aligned} \tag{5.8}$$

for all i, $1 \le i \le n$ and all positive integers m and thus (1.2) is satisfied.

Henceforth, we restrict our attention to Theorem 1. Observe that we shall prove Theorem 1 for arbitrary non-negative $p_1, \ldots, p_{n+1}$ and arbitrary $\vec{\mathbf{a}}_1, \ldots, \vec{\mathbf{a}}_{n+1} \in R^{n+1}$ with $\|\vec{\mathbf{a}}_i\| \le 1$.

It is convenient to replace "$n + 1$" by "n". Thus: we are given $n > 999$, $p_1, \ldots, p_n > 0$, $\vec{\mathbf{a}}_1, \ldots, \vec{\mathbf{a}}_n \in R^n$ with $\|\vec{\mathbf{a}}_i\| \le 1$ for $i = 1, \ldots, n$ and we shall find a history $\omega = \{(\xi, t)\}$ with $\xi \in \Omega$ and distinct $t \in R^+$ such that

$$\|\vec{\mathbf{D}}(t)\| < 500n^{1/2} \ln n \tag{5.9}$$

for all t.

§6. A relevant game

In this section we discuss a game which is interesting by itself and will provide insight for our proof.

The game is zero-sum and has two players, Pointer and Pusher. Vectors $\vec{\mathbf{a}}_1,\ldots,\vec{\mathbf{a}}_r \in R^n$ with $\|\vec{\mathbf{a}}_i\| \le 1$ are predetermined and known to both players. There are r moves. There is a vector $\vec{\mathbf{s}} \in R^n$ which changes value during the game – we let $\vec{\mathbf{s}}_i$ denote its value after i moves. Initially $\mathbf{s}_0 = 0$. On the i-th move first Pointer selects $\varepsilon_i = +1$ or -1. Then Pusher selects w_i, $0 \le w_i \le 1$. Set $\vec{\mathbf{s}}_i = \vec{\mathbf{s}}_{i-1} + w_i \varepsilon_i \vec{\mathbf{a}}_i$. The payoff to Pusher is $\|\vec{\mathbf{s}}_r\|$, the value of $\|\vec{\mathbf{s}}\|$ at the end of the game.

REMARK: Similar games have been examined by this author [3].

We now give a strategy for Pointer that holds Pusher to a modest payoff. For $\vec{\mathbf{x}} = (x_1,\ldots,x_n) \in R^n$ set

$$F(\vec{\mathbf{x}}) = \sum_{i=1}^{n} \operatorname{Cosh}(\lambda x_i) \tag{6.1}$$

where λ is a constant to be determined later.

Let $x, a \in R$ with $|a| \le 1$. Then

$$\begin{aligned}\tfrac{1}{2}[\operatorname{Cosh}[\lambda(x+a)] + \operatorname{Cosh}[\lambda(x-a)]] &= \operatorname{Cosh}(\lambda x)\operatorname{Cosh}(\lambda a)\\ &\le \operatorname{Cosh}(\lambda x)\operatorname{Cosh}(\lambda)\\ &\le \operatorname{Cosh}(\lambda x)\exp(\lambda^2/2) \end{aligned}\tag{6.2}$$

Now let $\vec{\mathbf{x}}, \vec{\mathbf{a}} \in R^n$ with $\|\vec{\mathbf{a}}\| \le 1$. Summing (6.2) over the coordinates

$$\tfrac{1}{2}[F(\vec{\mathbf{x}} + \vec{\mathbf{a}}) + F(\vec{\mathbf{x}} - \vec{\mathbf{a}})] \le F(\vec{\mathbf{x}})\exp(\lambda^2/2) \tag{6.3}$$

$$\min[F(\vec{\mathbf{x}} + \vec{\mathbf{a}}), F(\overrightarrow{\mathbf{x} - \mathbf{a}})] \le F(\vec{\mathbf{x}})\exp(\lambda^2/2) \tag{6.4}$$

By a second derivative calculation, F is convex. Thus if $\mathbf{y}$ is on a line interval with endpoints $\vec{\mathbf{x}}, \vec{\mathbf{z}}$

$$F(\vec{\mathbf{y}}) \le \max[F(\vec{\mathbf{x}}), F(\vec{\mathbf{z}})] \tag{6.5}$$

Properties (6.4), (6.5) are critical in applications. We remark that the use of hyperbolic cosines is not new and is well known in the theory of suboptimal martingales. In our applications we think of $F(\vec{\mathbf{x}})$ as an analytic approximation to $\exp[\lambda\|\vec{\mathbf{x}}\|]/2$.

Pointer's strategy is to select ε_i so as to minimize $F(\mathbf{s}_{i-1} + \varepsilon_i \mathbf{a}_i)$. By

(6.4) he may select ε_i so that $F(\vec{s}_{i-1} + \varepsilon_i \vec{a}_i) \le F(\vec{s}_{i-1}) \exp(\lambda^2/2)$. By (6.5) regardless of Pusher's choice of w_i,

$$F(\vec{s}_i) \le F(\vec{s}_{i-1}) \exp(\lambda^2/2) \tag{6.6}$$

Since $F(\vec{0}) = n$,

$$F(\vec{s}) \le n \exp(\lambda^2 r/2) \tag{6.7}$$

at the end of the game.

We require one more property of F.

$$F(\vec{x}) \ge \operatorname{Cosh}[\lambda \|\vec{x}\|] \ge \exp(\lambda \|\vec{x}\|)/2 \tag{6.8}$$

or, in more convenient form

$$\|\vec{x}\| \le \lambda^{-1} \ln[2F(x)]. \tag{6.9}$$

Applying (6.9) to (6.7),

$$s \le \lambda^{-1} \ln(2n) + \lambda r/2. \tag{6.10}$$

Pointer's strategy depends on the choice of λ. It is now clear to select

$$\lambda = [2 \ln(2n)/r]^{1/2} \tag{6.11}$$

so as to minimize the RHS of (6.10). With this λ the payoff to Pusher is at most

$$[2r \ln(2n)]^{1/2} \tag{6.12}$$

What does this game have to do with the problem at hand? Suppose we are given r symbols $1, \dots, r$ with associated vectors $\vec{a}_1, \dots, \mathbf{a}_r \in R^n$, $\|\vec{a}_i\| \le 1$ and associated non-negative reals $p_1, \dots, p_r$. (This is the situation in §5, Theorem 1 except that now the number of symbols r might not equal the dimension n.) We wish to find a history ω so that the discrepancy vector.

$$\mathbf{D}(t) = \sum_{i=1}^{r} D_i(t) \vec{a}_i \tag{6.13}$$

(analogous to (5.3)) remains small in the $\|.\|$ metric. Assume further that each p_i is "much larger" than the preceding p_{i-1}. (For example, p_i

$= \exp(\exp(10i))$.) We will use our game to outline a construction of a history ω.

Suppose symbols $1, \ldots, i-1$ have already been placed such that $\|\mathbf{s}_{i-1}(t)\|$ is small for all t, where

$$\vec{\mathbf{s}}_{i-1}(t) = \sum_{j=1}^{i-1} D_j(t)\vec{\mathbf{a}}_j \tag{6.14}$$

Now we wish to place symbols i such that $|D_i(t)| \leq 1$ and $\|\vec{\mathbf{s}}_i(t)\|$ is small for all t. We use the strategy of Pointer. Since p_i is much larger than p_{i-1}, the function $\vec{\mathbf{s}}_{i-1}(t)$ is nearly constant on intervals of lengths much larger than $1/p_i$. Suppose $\vec{\mathbf{s}}_{i-1}(t)$ is nearly constant around t_0 and we have to place a symbol i near t_0 in order to keep $\|\vec{\mathbf{s}}_i(t)\|$ small around t_0. We have for such t

$$\mathbf{s}_i(t) = \sum_{j=1}^{i} D_j(t)\mathbf{a}_j \sim \mathbf{s}_{i-1}(t_0) + D_i(t)\mathbf{a}_i \tag{6.15}$$

Our strategy is to select $\varepsilon_i \in \{-1, +1\}$ so as to minimize $F(\mathbf{s}_{i-1}(t_0) + \varepsilon_i\vec{\mathbf{a}}_i)$. If $\varepsilon_i = +1$, then we keep $D_i(t)$ non-negative by waiting until $D_i(t)$ becomes 1 before placing symbol i. If $\varepsilon_i = -1$, then we keep $D_i(t)$ non-positive by placing symbol i as soon as $D_i(t)$ becomes 0. (Of course, one has to meld these local strategies into a global history for i, but we do not claim a full proof here.) Formula (6.12) suggests that we can obtain

$$\|\vec{\mathbf{D}}(t)\| \leq [2r\ln(2n)]^{1/2} \tag{6.16}$$

at all times t. When, as in §5, $r = n$, we have

$$\|\mathbf{D}(t)\| \leq c(n\ln n)^{1/2}, \tag{6.17}$$

which is better than (1.2) by a factor of order $(\ln n)^{1/2}$.

§7. Bunched sets

In §4 and §6 we have proven (albeit in outline form) that Theorem 1 holds if the p_i are either fairly close together or quite spread apart. In both cases a result of $c(n\ln n)^{1/2}$ was achieved, as opposed to the $cn^{1/2}\ln n$ claimed in (1.2). The additional factor of $(\ln n)^{1/2}$ will enter in splitting an arbitrary set of p_i into sets of the above two manageable types.

We split R^+ into $[500 \ln n]$ sets by setting

$$C_i = \{x \in R^+ : [(\ln x)/(\ln 1.1)] \equiv i \bmod [500 \ln n]\} \tag{7.1}$$

for $0 \le i < [500 \ln n]$. Each C_i is the union of intervals of the form $[p, 1.1p)$ where consecutive intervals $[p, 1.1p)$, $[q, 1.1q)$ have

$$q/p = 1.1^{[500 \ln n]} > n^{40}. \tag{7.2}$$

A set $p_1, \ldots, p_r$ is called bunched if all p's lie in a common C_i. A bunched set is split into sets $p_1, \ldots, p_s$ lying in a common interval $[p, 1.1p)$ – these sets we call bunches. Any arbitrary set of p's splits into $[500 \ln n]$ bunched sets.

Let symbols $1, \ldots, r$ be given, $r \le n$, with corresponding $\vec{\mathbf{a}}_1, \ldots, \vec{\mathbf{a}}_r \in R^n$ with $\|\vec{\mathbf{a}}_i\| \le 1$ and corresponding positive $p_1, \ldots, p_r$. Given a history ω we set (in the notation of §5)

$$\vec{\mathbf{D}}(t) = \sum_{i=1}^{r} D_i(t)\mathbf{a}_i \tag{7.3}$$

THEOREM 2: *Using the notation of the above paragraph, assume that $p_1, \ldots, p_r$ is bunched. Then there exists a history ω so that*

$$\|\mathbf{D}(t)\| < 20 r^{1/2} (\ln n)^{1/2} \tag{7.4}$$

for all t.

Assume Theorem 2 is known. In Theorem 1 we are given arbitrary $p_1, \ldots, p_n$. We split these into v bunched sets of sizes $r_1, \ldots, r_v$ where $\sum_{i=1}^{v} r_i = n$. For each bunched set Theorem 2 gives a history ω_i on these symbols. Superimposing these histories gives a history ω with $\|\vec{\mathbf{D}}(t)\|$ bounded by

$$\sum_{i=1}^{v} 20 r_i^{1/2} (\ln n)^{1/2} \tag{7.5}$$

(We required, in §5, that the values t in a history ω be distinct and superimposing histories ω_i may cause ω to violate that condition. The Pushing Lemma of §12, or a continuity argument, shows that given ω_i satisfying (7.4) we may prescribe intervals around each t with $(A, t) \in \omega_i$ so that if the symbols A are allowed to move arbitrarily in their respective intervals the inequality (7.4) remains valid. We adjust ω_i to ω_i^*, still satisfying (7.4) and with no common values of t, and superimpose the ω_i^*.) We have

$$\sum_{i=1}^{v} r_i^{1/2} \leq v(n/v)^{1/2} = (nv)^{1/2} \tag{7.6}$$

the maximum being achieved when $r_1 = \ldots = r_v = n/v$. Thus

$$\begin{aligned} \|\vec{\mathbf{D}}(t)\| &\leq 20(\ln n)^{1/2}[n(500 \ln n)]^{1/2} \\ &\leq 500 n^{1/2} \ln n \end{aligned} \tag{7.7}$$

yielding Theorem 1.

Hence forth we concentrate on proving Theorem 2.

REMARK: The splitting of the p's into v bunched sets has a "cost" of $v^{1/2}$. Our attempts to remove this factor of $(\ln n)^{1/2}$ have not been successful.

§8. Hyperbolic cosine

Motivated by the remarks of §6 we reintroduce the hyperbolic cosine function. For $\vec{\mathbf{x}} = (x_1, \ldots, x_n) \in R^n$ we set

$$F(\vec{\mathbf{x}}) = \sum_{i=1}^{n} \operatorname{Cosh}(\lambda x_i) \tag{8.1}$$

We shall use the following four properties of F:

If $\vec{\mathbf{y}}$ is on a line segment with endpoints $\vec{\mathbf{x}}, \vec{\mathbf{z}}$ then

$$F(\vec{\mathbf{y}}) \leq \max[F(\vec{\mathbf{x}}), F(\vec{\mathbf{z}})] \tag{8.2}$$

If $\vec{\mathbf{v}}, \mathbf{a} \in R^n$, $\|\vec{\mathbf{a}}\| \leq 1$, $0 \leq \theta < 1$ then

$$\min F(\vec{\mathbf{v}} + (1-\theta)\vec{\mathbf{a}}), F(\vec{\mathbf{v}} - \theta\vec{\mathbf{a}}) \leq F(\mathbf{v})\, e^{\lambda^2}/8 \tag{8.3}$$

For all $\vec{\mathbf{x}} \in R^n$

$$\|\mathbf{x}\| \leq \lambda^{-1} \ln[2F(\vec{\mathbf{x}})] \tag{8.4}$$

If $\mathbf{v}, \vec{\mathbf{a}} \in R^n$, $\|\vec{\mathbf{a}}\| \leq \delta$ then

$$F(\vec{\mathbf{v}} + \vec{\mathbf{a}}) \leq F(\vec{\mathbf{v}})\, e^{\lambda\delta} \tag{8.5}$$

Properties (8.2) and (8.4) were previously given as (6.5) and (6.9). Property (8.5) follows by summing over coordinates the inequality

$$\mathrm{Cosh}[\lambda(v_i + a_i)] \leq \mathrm{Cosh}(\lambda v_i)\, e^{\lambda |a_i|} \tag{8.6}$$

To show the central property, (8.3), we begin with the inequality (see Appendix)

$$\theta\,\mathrm{Cosh}[\lambda(v + (1-\theta)a)] + (1-\theta)\mathrm{Cosh}[\lambda(v - \theta a)] \leq \mathrm{Cosh}(\lambda v) e^{\lambda^2/8} \tag{8.7}$$

valid for $0 \leq \theta \leq 1$ and $|a| \leq 1$. Summing over coordinates we derive

$$\theta F[\vec{\mathbf{v}} + (1-\theta)\vec{\mathbf{a}}] + (1-\theta)F[\vec{\mathbf{v}} - \theta\vec{\mathbf{a}}] \leq F[\vec{\mathbf{v}}]\, e^{\lambda^2/8} \tag{8.8}$$

from which (8.3) follows. Observe that (8.3) serves as a generalization of (6.4). At one point in our proof we shall need a slight generalization of (8.3). If $\vec{\mathbf{v}}, \vec{\mathbf{a}} \in R^n$, $0 \leq \theta \leq 1$, $\|\vec{\mathbf{a}}\| \leq X$ then

$$\min[F(\vec{\mathbf{v}} + (1-\theta)\vec{\mathbf{a}}), F(\vec{\mathbf{v}} - \theta\vec{\mathbf{a}})] \leq F(\mathbf{v})\, e^{\lambda^2 X^2/8} \tag{8.9}$$

Now we make a further reduction:

THEOREM 3: *Under the conditions of Theorem* 2 *there exists* ω *so that for all* $t > 0$

$$F(\vec{\mathbf{D}}(t)) \leq n\, e^{14\lambda^2 r} \tag{8.10}$$

where F *is given by* (8.1) *and*

$$\lambda = [\ln(2n)/14r]^{1/2} \tag{8.11}$$

Given this result, Theorem 2 follows immediately by applying (8.4) to (8.10). In fact, we shall show Theorem 3 for any $\lambda \geq n^{-4}$ and in all but one step of the proof λ may be an arbitrary positive real. Henceforth, we concentrate on proving Theorem 3.

§9. Background function

We find the history ω desired in Theorem 3 in stages. Split $p_1, \ldots, p_r$ (which we recall is bunched) into bunches, in increasing order of p (i.e. rarest symbols first). At each stage we determine the entire history of all symbols in that bunch. (In §6 each bunch consisted of a single symbol.) The symbols previously placed (which we will call the old symbols) have a history ω_0 and a discrepancy function $\vec{\mathbf{D}}_0(t)$. Now we need find a history ω of the next bunch (which we call the new symbols) and a discrepancy function $\vec{\mathbf{D}}(t)$. The combined history, the union of ω_0 and ω, has discrepancy function $\vec{\mathbf{D}}_0(t) + \vec{\mathbf{D}}(t)$. We shall show the following:

THEOREM 4: *Let $s \leq n$. Let s symbols $1,\ldots,s$ be given with associated $p_1,\ldots,p_s$ and $\vec{\mathbf{a}}_1,\ldots,\vec{\mathbf{a}}_s \in R^n$, $\|\vec{\mathbf{a}}_i\| \leq 1$. Assume $p_1,\ldots,p_s \in [p, 1.1p)$. Assume for a disjoint set (possibly empty) of symbols that a history ω_0 and a corresponding discrepancy function $\vec{\mathbf{D}}_0(t)$ are given. Let K be such that*

$$F(\vec{\mathbf{D}}_0(t)) \leq K \tag{9.1}$$

for all t. Assume that the appearances of these old symbols are at least $5/p$ apart. Then there exists a history ω on symbols $1,\ldots,s$ so that

$$F(\vec{\mathbf{D}}_0(t) + \vec{\mathbf{D}}(t)) \leq K\, e^{14\lambda^2 s} \tag{9.2}$$

for all t and the symbols in the combined history are at least n^{-5}/p apart.

Assume Theorem 4 and let us be given a bunched set of symbols. We initially have $\vec{\mathbf{D}}_0(t) = \vec{\mathbf{0}}$ so that $K = F(\vec{\mathbf{0}}) = n$. The adding of s symbols to the history increases the upper bound on $F(\vec{\mathbf{D}}_0(t))$ by a factor of at most $\exp(14\lambda^2 s)$. When all r symbols have been placed $F(\vec{\mathbf{D}}_0(t))$ will be bounded by $n\exp(14\lambda^2 r)$ as desired. We must also check the distances between appearances. A history ω_0 will have appearances at least $n^{-5}/p_{\max}$ apart where $p_{\max}$ is the maximal p_x for the old symbols. In the next bunch of symbols all $p_i \in [p, 1.1p)$ where $p \geq p_{\max} n^{40}$ by (7.2). Thus certainly the appearances in ω_0 are at least $5/p$ apart so that we may apply Theorem 4 and so Theorem 3 follows. Henceforth we concentrate on proving Theorem 4.

REMARK: The considerations of distance between appearances are technically bothersome and an intuitive explanation may help. Let us say a symbol with corresponding p has time span p^{-1}. That is, the symbol should occur once in time p^{-1}. The old symbols have enormous time spans relative to the time spans of the new symbols. In an interval with length the new time span (or five times that) we require that one never has two appearances of old symbols.

In proving Theorem 4 we need certain facts about $\vec{\mathbf{D}}_0(t)$, which we will refer to as the Background Function. The function $\vec{\mathbf{D}}_0(t)$ is right-continuous with discontinuities only at appearances of symbols. If $(i,t)\in\omega_0$ then $\vec{\mathbf{D}}_0(t)=\vec{\mathbf{D}}_0(t^-)+\vec{\mathbf{a}}_i$. (Here $f(t^-)=\lim_{x\uparrow t}f(x)$). If no $(i,t)\in\omega_0$ for $t'<t<t''$ then $\vec{\mathbf{D}}_0(t)$ is linear on $[t',t'')$. In particular, if $t=\alpha t'+(1-\alpha)t''$ with $0\le\alpha<1$ then

$$\vec{\mathbf{D}}_0(t)=\alpha\vec{\mathbf{v}}+(1-\alpha)\vec{\mathbf{w}} \tag{9.3}$$

where $\vec{\mathbf{v}}=\vec{\mathbf{D}}_0(t')$ and $\vec{\mathbf{w}}=\vec{\mathbf{D}}_0(t''^-)$.

§10. Benchmark lemmas

To prove Theorem 4 we are guided by the arguments of §4, with $\|.\|$ replaced by F. We begin by establishing benchmarks. In the next lemma we think of $\vec{\mathbf{v}}=\vec{\mathbf{D}}_0(t)$ and θ_i as the fractional part of p_it.

BENCHMARK LEMMA: *Let* $\vec{\mathbf{v}}\in R^n$, $\vec{\mathbf{a}}_1,\dots,\vec{\mathbf{a}}_s\in R^n$, $\|\vec{\mathbf{a}}_i\|\le 1$. *Let* $0\le\theta_1,\dots,\theta_s<1$. *There exist* $\varepsilon_1,\dots,\varepsilon_s\in\{0,1\}$ *so that setting*

$$\vec{\mathbf{w}}=\vec{\mathbf{v}}+(\varepsilon_1-\theta_1)\vec{\mathbf{a}}_1+\dots+(\varepsilon_s-\theta_s)\vec{\mathbf{a}}_s \tag{10.1}$$

we have

$$F(\vec{\mathbf{w}})\le F(\vec{\mathbf{v}})\,e^{\lambda^2 s/8} \tag{10.2}$$

PROOF: For $s=1$ this is precisely (8.3). The proof for all s follows by induction.

REMARK: We may think of the Benchmark Lemma probabilistically. If the ε_i are independent random variables with values 0, 1 and expectation θ_i then $F(\vec{\mathbf{w}})$ is a random variable whose expectation is at most $F(\mathbf{v})\exp(\lambda^2 s/8)$.

When t is a discontinuity of $\vec{\mathbf{D}}_0$, an appearance of an old symbol, there is a special problem. The solution is given in the next lemma where we think of $\vec{\mathbf{v}}=\vec{\mathbf{D}}_0(t^-)$ and $\vec{\mathbf{v}}+\vec{\mathbf{b}}=\vec{\mathbf{D}}_0(t)$.

DOUBLE BENCHMARK LEMMA: *Let* $\vec{\mathbf{v}}\in R^n$, $\mathbf{b}\in R^n$, $\|\vec{\mathbf{b}}\|\le 1$, $F(\vec{\mathbf{v}})\le K$, $F(\vec{\mathbf{v}}+\vec{\mathbf{b}})\le K$. *Let* $\vec{\mathbf{a}}_1,\dots,\vec{\mathbf{a}}_s\in R^n$, $\|\vec{\mathbf{a}}_i\|\le 1$. *Let* $0\le\theta_1,\dots,\theta_s<1$. *Then there exist* $\varepsilon_1,\dots,\varepsilon_s\in\{0,1\}$ *so that,*

$$F(\vec{\mathbf{w}})\le K\,e^{\lambda^2 s/2} \text{ and } F(\vec{\mathbf{w}}+\vec{\mathbf{b}})\le K\,e^{\lambda^2 s/2} \tag{10.3}$$

where $\vec{\mathbf{w}}$ *is given by* (10.1).

PROOF: By induction it suffices to prove the Double Benchmark Lemma for $s = 1$, with a single θ, ε, $\vec{\mathbf{a}}$.

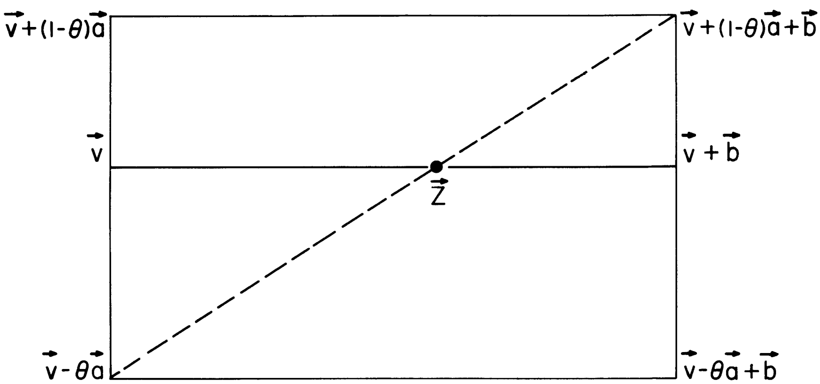

Figure: Double Benchmark Lemma

Let us say $\vec{\mathbf{x}}$ is small if $F(\vec{\mathbf{x}}) \leq K \exp(\lambda^2/2)$. By (8.3) either $\vec{\mathbf{v}} - \theta\vec{\mathbf{a}}$ or $\vec{\mathbf{v}} + (1-\theta)\vec{\mathbf{a}}$ is small and either $\vec{\mathbf{v}} - \theta\vec{\mathbf{a}} + \vec{\mathbf{b}}$ or $\vec{\mathbf{v}} + (1-\theta)\vec{\mathbf{a}} + \vec{\mathbf{b}}$ is small. Our only problem would be if two diagonal points, say $\vec{\mathbf{v}} - \theta\vec{\mathbf{a}}$ and $\vec{\mathbf{v}} + (1-\theta)\vec{\mathbf{a}} + \vec{\mathbf{b}}$ were not small. Set $\vec{\mathbf{z}} = \vec{\mathbf{v}} + \theta\vec{\mathbf{b}}$. By (8.2), $F(\vec{\mathbf{z}}) \leq K$. These two points are the values of $\vec{\mathbf{z}} + (\varepsilon - \theta)(\vec{\mathbf{a}} + \vec{\mathbf{b}})$ with $\varepsilon = 0, 1$. Since $\|\vec{\mathbf{a}} + \vec{\mathbf{b}}\| \leq \|\vec{\mathbf{a}}\| + \|\vec{\mathbf{b}}\| \leq 2$ we may apply (8.9) to show that one of the two points is indeed small.

We assume the hypotheses of Theorem 4. We first establish, using the Double Benchmark Lemma, a benchmark at each occurrence t of an old point. If $\varepsilon_i = 1$ then we set $Z_i(t) = [p_i t] + 1$ and if $\varepsilon_i = 0$ then we set $Z_i(t) = [p_i t]$. Then

$$F(\vec{\mathbf{D}}_0(t) + \vec{\mathbf{D}}(t)) \leq K e^{\lambda^2 s/2}$$
$$F(\vec{\mathbf{D}}_0(t^-) + \vec{\mathbf{D}}(t)) \leq K e^{\lambda^2 s/2} \tag{10.4}$$

and

$$|Z_i(t) - p_i t| \leq 1 \tag{10.5}$$

for every new symbol i. Furthermore, if two consecutive occurrences t', t'' are more than $10/p$ apart we use the Benchmark Lemma to place benchmarks at points $t' + i(5/p)$, i positive integral and $t' + (i + 1)(5/p) < t''$, so that the distance between consecutive benchmarks is not smaller than $5/p$ and not larger than $10/p$. (If there are no old symbols then benchmarks are placed at $i(5/p)$, $i = 1, 2, 3, \ldots$) Each new benchmark t also satisfies (10.4) and (10.5). Since consecutive benchmarks t', t'' are at least $5/p$ apart the benchmarks are consistent, for if $t'' \geq t' + 5/p$ and $p \leq p'$ then, by (10.5)

$$Z_i(t'') - Z_i(t') \geq (p_i t'' - p_i t') - 2 \geq 5 - 2 \geq 0 \tag{10.6}$$

On the other hand, if t' and t'' are consecutive benchmarks then, since $t'' < t' + 10/p$, $p_i \leq 1.1p$ and (10.5),

$$Z_i(t'') - Z_i(t') \leq (p_i t'' - p_i t') + 2 < 11 + 2 = 13 \tag{10.7}$$

§11. Local placement

In this section we show how to find "local" histories ω such that the discrepancy vector $\vec{\mathbf{D}}(t)$ is small in the sense that $F(\vec{\mathbf{D}}(t) + \vec{\mathbf{D}}_0(t))$ is small. We begin with the simplest case.

Let a single symbol A be given with associated vector $\vec{\mathbf{a}} \in R^n$, $\|\vec{\mathbf{a}}\| \leq 1$. Consider a history ω on $[0, 1]$ in which symbol A appears exactly once. We set

$$\vec{\mathbf{D}}(t) = \begin{cases} -t\vec{\mathbf{a}} & \text{if } A \text{ does not appear in } (0, t] \\ (1-t)\vec{\mathbf{a}} & \text{if it does} \end{cases} \tag{11.1}$$

Let $\vec{\mathbf{v}}, \vec{\mathbf{w}} \in R^n$ with $F(\vec{\mathbf{v}}) \leq K$ and $F(\vec{\mathbf{w}}) \leq K$ and let

$$\vec{\mathbf{D}}_0(t) = (1-t)\vec{\mathbf{v}} + t\vec{\mathbf{w}} \tag{11.2}$$

LEMMA 11.1: *With the above conditions there exists a history ω so that*

$$F(\vec{\mathbf{D}}(t) + \vec{\mathbf{D}}_0(t)) < K e^{\lambda^2/8} \tag{11.3}$$

for all $t \in [0, 1]$ and A appears in the open interval $(0, 1)$.

PROOF: Set

$$\begin{aligned} F^-(t) &= F[-t\vec{\mathbf{a}} + \vec{\mathbf{D}}_0(t)] \\ F^+(t) &= F[(1-t)\vec{\mathbf{a}} + \vec{\mathbf{D}}_0(t)] \end{aligned} \tag{11.4}$$

so that $F(t) = F^-(t)$ if A does not appear in $[0, t]$ and otherwise $F(t) = F^+(t)$. Let $t \in [0, 1]$. By (8.2), $F[\vec{\mathbf{D}}_0(t)] \leq K$. By (8.3), either

$F^-(t) < K e^{\lambda^2/8}$ or $F^+(t) < K e^{\lambda^2/8}$. At $t = 0$ F^- has this property, at $t = 1$ F^+ has this property. Since both F^- and F^+ are continuous there exists $t_0 \in (0, 1)$ so that both $F^-(t_0) < K e^{\lambda^2/8}$ and $F^+(t_0) < K e^{\lambda^2/8}$. Define ω by placing A at t_0. By (8.2), for $t \in [0, t_0]$

$$F(t) = F^-(t) \le \max[F^-(0), F^-(t_0)] < K e^{\lambda^2/8} \tag{11.5}$$

and for $t \in [t_0, 1]$

$$F(t) = F^+(t) \le \max[F^+(t_0), F^+(1)] < K e^{\lambda^2/8} \tag{11.6}$$

so the Lemma is proven.

Let $q = 2^d$ symbols $1, \dots, q$ be given with associated vectors $\vec{\mathbf{a}}_1, \dots, \vec{\mathbf{a}}_q \in R^n$, $\|\vec{\mathbf{a}}_i\| \le 1$. Consider a history ω on $[0, 1]$ in which each symbol appears exactly once. Set, for $1 \le i \le q$

$$D_i(t) = \begin{cases} -t & \text{if symbol } i \text{ does not appear in } [0, t] \\ 1 - t & \text{if it does} \end{cases} \tag{11.7}$$

and

$$\vec{\mathbf{D}}(t) = \sum_{i=1}^{q} D_i(t)\vec{\mathbf{a}}_i \tag{11.8}$$

Let $\vec{\mathbf{v}}, \vec{\mathbf{w}} \in R^n$ with $F(\vec{\mathbf{v}}) \le K$, $F(\vec{\mathbf{w}}) \le K$ and let

$$\vec{\mathbf{D}}_0(t) = (1 - t)\vec{\mathbf{v}} + t\vec{\mathbf{w}} \tag{11.9}$$

LEMMA 11.2: *With the above conditions there exists a history ω so that*

$$F[\vec{\mathbf{D}}(t) + \vec{\mathbf{D}}_0(t)] < K e^{\lambda^2 q/2} \tag{11.10}$$

for all $t \in [0, 1]$ and each interval $(i/q, (i + 1)/q)$, $i = 0, 1, \dots, q - 1$ contains precisely one symbol.

PROOF: We use induction on d. The case $d = 0$ was shown (in a strengthened form) in the preceding Lemma.

By the Benchmark Lemma we may find $\varepsilon_1, \dots, \varepsilon_{q/2} \in \{0, 1\}$ so that, setting

$$\vec{\mathbf{z}} = (\vec{\mathbf{v}} + \vec{\mathbf{w}})/2 + \sum_{i=1}^{q/2} (\varepsilon_i - \tfrac{1}{2})(\vec{\mathbf{a}}_i - \vec{\mathbf{a}}_{i+q/2}) \tag{11.11}$$

we have

$$F(\mathbf{z}) \leq K\, e^{\lambda^2 q/4} \tag{11.12}$$

(As $\|\mathbf{a}_i - \vec{\mathbf{a}}_{i+q/2}\| \leq 2$ we apply (8.9).) We split the symbol set $\{1, \ldots, q\}$ into two sets L, R as follows. If $\varepsilon_i = 1$ place $i \in L$, $i + q/2 \in R$; if $\varepsilon_i = 0$ place $i \in R$, $i + q/2 \in L$. (We have established a benchmark at $t = \frac{1}{2}$. The elements $i, i + q/2$ were paired to insure $|L| = |R| = q/2$.)

For all $t \in [0, \frac{1}{2}]$,

$$\vec{\mathbf{D}}_0(t) + \vec{\mathbf{D}}(t) = (1 - 2t)\vec{\mathbf{v}} + (2t)\vec{\mathbf{z}} + \sum_{i \in L} D_i^*(t)\vec{\mathbf{a}}_i$$

where

$$D_i^*(t) = \begin{cases} -2t & \text{if } i \text{ has not appeared in } [0, t] \\ 1 - 2t & \text{if it has} \end{cases}$$

(This is the Reduction Principle.) By induction there is a history ω on $[0, \frac{1}{2}]$ with symbol set L such that for all $t \in [0, \frac{1}{2}]$

$$\begin{aligned} F[\vec{\mathbf{D}}_0(t) + \vec{\mathbf{D}}(t)] &\leq \max[F(\vec{\mathbf{v}}), F(\vec{\mathbf{z}})]\, e^{\lambda^2 q/4} \\ &\leq K\, e^{\lambda^2 q/4}\, e^{\lambda^2 q/4} \leq K\, e^{\lambda^2 q/2} \end{aligned} \tag{11.15}$$

and for every integer i with $0 \leq i < q/2$ the interval $(i/q, (i+1)/q)$ contains precisely one element. There is a similar history for symbols R on $[\frac{1}{2}, 1]$, completing the proof of the Lemma.

Three simple steps significantly strengthen Lemma 11.2 and place us in position for our final assault. First, given an arbitrary number q of symbols we may add "dummy" symbols, each with associated vector $\vec{\mathbf{0}}$, until reaching $q' = 2^d$ symbols where $q' < 2q$. Second, if a symbol i is supposed to appear n_i times we may replace it with n_i symbols, each with the same associated vectors as the original symbol, each to appear once. Third, the interval $[0, 1]$ may be replaced by an arbitrary interval $[t', t'']$ by a linear transformation. Combining these gives the following result.

Let s symbols $1, \ldots, s$ be given with associated vectors $\vec{\mathbf{a}}_1, \ldots, \vec{\mathbf{a}}_s \in R^n$, $\|\vec{\mathbf{a}}_i\| \leq 1$ and associated "multiplicities" $n_1, \ldots, n_s$. Set $N = \sum_{i=1}^{s} n_i$. Given a history ω on these symbols on $[t', t'']$ in which symbol i appears exactly n_i times set

$$\begin{aligned} Z_i(t) &= \text{the number of appearances of symbol } i \text{ in } [t', t''] \\ D_i(t) &= Z_i(t) - n_i(t - t')/(t'' - t') \\ \vec{\mathbf{D}}(t) &= \sum_{i=1}^{s} D_i(t)\mathbf{a}_i \end{aligned} \tag{11.16}$$

Let $\vec{\mathbf{v}}, \vec{\mathbf{w}} \in R^n$ with $F(\mathbf{v}) \leq K$, $F(\vec{\mathbf{w}}) \leq K$ and let

$$\vec{\mathbf{D}}_0(t) = \left[1 - \frac{t - t'}{t'' - t'}\right]\vec{\mathbf{v}} + \left[\frac{t - t'}{t'' - t'}\right]\vec{\mathbf{w}} \tag{11.17}$$

LEMMA 11.3: *With the above conditions there exists a history ω so that*

$$F[\vec{\mathbf{D}}(t) + \vec{\mathbf{D}}_0(t)] \leq K\, e^{\lambda^2 N} \tag{11.18}$$

for all $t \in [t', t'']$ and we may split $[t', t'']$ into disjoint intervals, each of length at least $(t'' - t')/2N$ so that each interval has at most one appearance of a symbol.

§12. Final steps

At the end of §10 we had constructed a system of consistent benchmarks, at most $10/p$ apart, with

$$\begin{aligned} &F[\vec{\mathbf{D}}_0(t_0) + \vec{\mathbf{D}}(t_0)] \leq K\, e^{\lambda^2 s/2} \\ &F[\vec{\mathbf{D}}_0(t_0^-) + \vec{\mathbf{D}}(t_0)] \leq K\, e^{\lambda^2 s/2} \end{aligned} \tag{12.1}$$

for each benchmark t_0. Let t', t'' be consecutive benchmarks. Set

$$\begin{aligned} &n_i = Z_i(t'') - Z_i(t') \\ &N = \sum_{i=1}^{s} n_i \end{aligned} \tag{12.2}$$

By (10.7), all $n_i \leq 12$ hence $N \leq 12s$. Let

$$t = (1 - \alpha)t' + \alpha t'' \in [t', t'') \tag{12.3}$$

Then (and this is the Reduction Principle)

$$\vec{\mathbf{D}}_0(t) + \vec{\mathbf{D}}(t) = (1 - \alpha)\vec{\mathbf{v}} + \alpha\vec{\mathbf{w}} + \vec{\mathbf{D}}^*(t) \tag{12.4}$$

where

$$\begin{aligned} &\vec{\mathbf{v}} = \vec{\mathbf{D}}_0(t') + \vec{\mathbf{D}}(t') \\ &\vec{\mathbf{w}} = \vec{\mathbf{D}}_0(t''^-) + \vec{\mathbf{D}}(t'') \end{aligned} \tag{12.5}$$

and $\vec{\mathbf{D}}^*(t)$ is the discrepancy function defined (as $\vec{\mathbf{D}}(t)$) by (11.16). Hence, by (12.1) and Lemma 11.3, there exists a history ω on $[t', t'')$ so that

$$F[\vec{\mathbf{D}}_0(t) + \vec{\mathbf{D}}(t)] \le K\, e^{\lambda^2 s/2}\, e^{\lambda^2 N} \le K\, e^{13\lambda^2 s} \tag{12.6}$$

for all $t \in [t', t'')$. As t', t'' were arbitrary consecutive benchmarks, (12.6) holds for all $t \ge 0$.

To complete the proof we modify ω so that all symbols are at least n^{-5}/p apart. From Lemma 11.3 each new symbol lies in an interval of length at least $(t'' - t')/2N$ not containing any other symbols. Since $t'' - t' \ge 5/p$ and $N \le 12s \le 12n$ these intervals are at least (as $n \ge 999$) $4n^{-5}/p$ apart. Note, however, that two symbols may be arbitrarily close together if they lay on the extreme right and extreme left sides of adjacent intervals. The following Lemma enables us to push such "kissing" pairs apart by brute force.

Pushing Lemma: *Let a history ω on s symbols, $s \le n$ be given. Let p be an upper bound for the associated p_i. Suppose K is an upper bound on $F(\vec{\mathbf{D}}(t))$ for all t. Let a symbol appear at t_0 and no symbol appear in $(t_0, t_0 + \varepsilon]$. Let ω^* be the altered history given by moving this appearance from t_0 to $t_0 + \varepsilon$ and let $\vec{\mathbf{D}}^*(t)$ denote the altered $\vec{\mathbf{D}}(t)$. Then*

$$F[\vec{\mathbf{D}}^*(t)] \le K\, e^{\lambda p \varepsilon n} \tag{12.7}$$

for all $t_0 \le t < t_0 + \varepsilon$ and $\vec{\mathbf{D}}^*(t) = \vec{\mathbf{D}}(t)$ for all other $t \in R^+$.

Proof: The second part is clear. Let $t = t_0 + \delta$ where $0 \le \delta < \varepsilon$. Then

$$Z_i^*(t_0 + \delta) = Z_i(t_0^-) - \delta p_i \tag{12.8}$$

so

$$|Z_i^*(t_0 + \delta) - Z_i(t_0^-)| \le \delta p_i \le \delta p \tag{12.9}$$

and so

$$\|\vec{\mathbf{D}}^*(t_0 + \delta) - \vec{\mathbf{D}}(t_0^-)\| \le \delta p s \le \delta p n \le \varepsilon p n \tag{12.10}$$

We use, for the first and only time, property (8.5) to complete the proof.

The Pushing Lemma also holds if the symbol is moved from t_0 to $t_0 - \varepsilon$. Returning to our history ω we move any new symbol that is within n^{-5}/p of the edge of its interval a distance $\varepsilon = n^{-5}/p$ toward the center of its interval. In the new history ω^* all symbols (new and old) are at least n^{-5}/p apart. We have increased F by at most a factor of

$\exp(\lambda p \in n) = \exp(\lambda n^{-4})$. For the first and only time we use the actual value of λ given by (8.11). Actually, we need only that λ is not "extremely small" – more precisely, $\lambda \geq n^{-4}$. Then $\lambda^2 \geq \lambda n^{-4}$ and so the "non-kissing" ω^* has

$$F[\vec{\mathbf{D}}_0(t) + \mathbf{D}(t)] \leq K\, e^{13\lambda^2 s} e^{\lambda^2} \leq K\, e^{14\lambda^2 s} \tag{12.11}$$

completing the proof of Theorem 4 and hence completing the proof of Theorem A.

§13. Lower bounds

In this section we return to the original formulation of §1 and show that the result (1.2) is best possible up to a factor of $(\ln n)$. The techniques are well known.

A square matrix $H = (h_{ij})$ of order n is called a Hadamard matrix if all entries $h_{ij} = \pm 1$ and if the row vectors $\mathbf{r}_i = (h_{i1}, \ldots, h_{in}) \in R^n$ are mutually orthogonal. We call H normalized if $\mathbf{r}_1 = \mathbf{1} = (1, \ldots, 1)$.

Let a normalized Hadamard matrix H of even order n be given. Let $\Omega = \{1, \ldots, n\}$ and set $\mu(\{j\}) = n^{-1}$ for all $j \in \Omega$. For $2 \leq i \leq n$ set

$$A_i = \{j \in \Omega : h_{ij} = +1\} \tag{13.1}$$

Let $\omega = (\xi_1, \xi_2, \ldots)$ be an arbitrary sequence of points from Ω. We shall show that

$$|D_{A_i}(n/2)| \geq n/4(n-1)^{1/2} \tag{13.2}$$

for some i, $2 \leq i \leq n$.

Fix ω and for each $j \in \Omega$ let (suppressing the value $t = n/2$)

$$Z_j \text{ be the number of } i \text{ with } 1 \leq i \leq n/2 \text{ and } \xi_i = j \tag{13.3}$$

$$D_j = Z_j - \tfrac{1}{2}$$

$$\mathbf{D} = (D_1, \ldots, D_n) \in R^n$$

Since Z_j is integral, $|D_j| \geq \frac{1}{2}$ and therefore

$$|\vec{\mathbf{D}}|^2 \geq n/4 \tag{13.4}$$

Since the D_i are discrepancies $\vec{\mathbf{D}} \cdot \mathbf{1} = \mathbf{0}$. The vectors $\vec{\mathbf{r}}_2, \ldots, \vec{\mathbf{r}}_n$ form an orthogonal basis for the hyperspace of R^n perpendicular to **1**. Hence

$$|\vec{\mathbf{D}}|^2 = \sum_{i=2}^{n} (\vec{\mathbf{D}} \cdot \mathbf{r}_i)^2 / |\vec{\mathbf{r}}_i|^2 \tag{13.5}$$

The entries of all $\mathbf{r}_i$ are all ± 1 so all $|\mathbf{r}_i|^2 = n$. Thus

$$\sum_{i=2}^{n} (\mathbf{D} \cdot \bar{\mathbf{r}}_i)^2 \geq n^2/4 \tag{13.6}$$

and therefore

$$|\vec{\mathbf{D}} \cdot \vec{\mathbf{r}}_i| \geq n/2(n-1)^{1/2} \tag{13.7}$$

for some i, $2 \leq i \leq n$. We observe that

$$D_{A_i}(n/2) = \sum_{x \in A_i} D_x = \sum_{x \in A_i} D_x - \tfrac{1}{2} \sum_{x \in X} D_x = \tfrac{1}{2} \vec{\mathbf{D}} \cdot \vec{\mathbf{r}}_i \tag{13.8}$$

and (13.2) immediately follows.

Let $TIJ(n)$ be the minimal value so that Theorem A holds with this value as the RHS of (1.2). We have shown that if a normalized Hadamard matrix of order n exists then

$$TIJ(n) \geq n/4(n-1)^{1/2} \tag{13.9}$$

Such matrices are known to exist for many n, in particular all n of the form $4^s 12^t$. These orders n are thus asymptotically dense in the weak sense that for all $\varepsilon > 0$ if m is sufficiently large there exists such an order between m and $m(1+\varepsilon)$. We combine Theorem A and the Lower Bound with a single asymptotic statement:

$$c_1 n^{1/2} \leq TIJ(n) \leq c_2 n^{1/2} \ln n \tag{13.10}$$

Appendix

We give here proofs of inequalities (8.7) and (4.4).

LEMMA: *For all reals α, β with $|\alpha| \leq 1$*

$$\operatorname{Cosh} \beta + \alpha \operatorname{Sinh} \beta \leq \exp(\beta^2/2 + \alpha\beta) \tag{A1}$$

PROOF: This is immediate if $\alpha = +1$ or $\alpha = -1$ or if $|\beta| \geq 100$. If the Lemma were false the function

$$f(\alpha, \beta) = \operatorname{Cosh}\beta + \alpha \operatorname{Sinh}\beta - \exp(\beta^2/2 + \alpha\beta) \tag{A2}$$

would assume a negative global minimum in the interior of the rectangle

$$R = \{(\alpha, \beta) : |\alpha| \leq 1, |\beta| \leq 100\} \tag{A3}$$

Setting partial derivatives equal zero we find

$$\begin{aligned} \operatorname{Sinh}\beta + \alpha \operatorname{Cosh}\beta &= (\alpha + \beta)\exp(\beta^2/2 + \alpha\beta) \\ \operatorname{Sinh}\beta &= \beta \exp(\beta^2/2 + \alpha\beta) \end{aligned} \tag{A4}$$

and thus $\operatorname{Tanh}\beta = \beta$ which implies $\beta = 0$. But $f(\alpha, 0) = 0$ for all α, hence the Lemma is true.

THEOREM: *For all reals θ, a, λ, v with $0 \leq \theta \leq 1$ and $|a| \leq 1$*

$$\theta \operatorname{Cosh}[\lambda(v + (1 - \theta)a)] + \\ + (1 - \theta)\operatorname{Cosh}[\lambda(v - \theta a)] \leq \operatorname{Cosh}(\lambda v)\, e^{\lambda^2/8} \tag{A5}$$

PROOF: Expressing $\operatorname{Cosh}\alpha = (e^{\alpha} + e^{-\alpha})/2$ the LHS becomes

$$\tfrac{1}{2}e^{\lambda v}[\theta\, e^{(1-\theta)\lambda a} + (1 - \theta)\, e^{-\theta\lambda a}] + \\ + \tfrac{1}{2}e^{-\lambda v}[\theta\, e^{-(1-\theta)\lambda a} + (1 - \theta)\, e^{\theta\lambda a}] \tag{A6}$$

and so it suffices to show

$$\begin{aligned} \theta\, e^{(1-\theta)\lambda a} + (1 - \theta)\, e^{-\theta\lambda a} &\leq e^{\lambda^2/8} \\ \theta\, e^{-(1-\theta)\lambda a} + (1 - \theta)\, e^{\theta\lambda a} &\leq e^{\lambda^2/8} \end{aligned} \tag{A7}$$

Transforming θ to $1 - \theta$ switches the above equations, so it suffices to show the first one. It suffices to show this equation for $a = 1$, i.e.,

$$\theta\, e^{(1-\theta)\lambda} + (1 - \theta)\, e^{-\theta\lambda} \leq e^{\lambda^2/8} \tag{A8}$$

as we may then set $\lambda = \lambda' a$. Setting $\theta = (1 + \alpha)/2$ and $\lambda/2 = \beta$ this reduces to the Lemma above.

For $1 \leq i \leq s$ let ε_i be independent random variables with

$$\Pr(\boldsymbol{\varepsilon}_i = +1) = \theta_i, \Pr(\boldsymbol{\varepsilon}_i = 0) = 1 - \theta_i, \tag{A9}$$

where $0 \le \theta_i \le 1$. Set $\boldsymbol{Y}_i = \boldsymbol{\varepsilon}_i - \theta_i$ and that

$$\boldsymbol{S} = \sum_{i=1}^{t} \boldsymbol{Y}_i \tag{A10}$$

For all reals λ (here E represents Expected Value)

$$\mathrm{E}[e^{\lambda \boldsymbol{Y}_i}] = \theta_i e^{\lambda(1-\theta_i)} + (1-\theta_i) e^{-\lambda\theta_i} \le e^{\lambda^2/8} \tag{A11}$$

as shown above. Since the $\boldsymbol{Y}_i$ are independent,

$$\mathrm{E}[e^{\lambda \boldsymbol{S}}] = \prod_{i=1}^{s} \mathrm{E}[e^{\lambda \boldsymbol{Y}_i}] \le e^{\lambda^2 s/8} \tag{A12}$$

and thus, for every $\alpha > 0$

$$\begin{aligned} \Pr[\boldsymbol{S} > \alpha] &= \Pr[e^{\lambda \boldsymbol{S}} > e^{\lambda\alpha}] \\ &< \mathrm{E}[e^{\lambda \boldsymbol{S}}]\, e^{-\lambda\alpha} \\ &\le e^{\lambda^2 s/8 - \lambda\alpha}. \end{aligned} \tag{A13}$$

Similarly

$$\Pr[\boldsymbol{S} < -\alpha] = \Pr[e^{-\lambda \boldsymbol{S}} > e^{\lambda\alpha}] \le e^{\lambda^2 s/8 - \lambda\alpha} \tag{A14}$$

so that

$$\Pr[|\boldsymbol{S}| > \alpha] < 2\, e^{\lambda^2 s/8 - \lambda\alpha}. \tag{A15}$$

Setting $\lambda = 4\alpha/s$ (so as to minimize the above expression)

$$\Pr[|\boldsymbol{S}| > \alpha] < 2\, e^{-2\alpha^2/s}. \tag{A16}$$

When $s \le n + 1$, as in (4.4),

$$\Pr[|\boldsymbol{S}| > \alpha] < 2\, e^{-2\alpha^2/(n+1)}. \tag{A17}$$

REFERENCES

[1] J. BECK and T. FIALA: Integer-making theorems. *Discr. Appl. Math. 3* (1981) 1–8.
[2] H.G. MEIJER: On a distribution problem in finite sets. *Nederl. Akad. Wetensch. Indag. Math. 35* (1973) 9–17.
[3] J. SPENCER: Balancing games. *J. Combinatorial Th. B 23* (1977) 68–74.
[4] J. SPENCER: Discrete ham sandwich theorems. *European J. Combinatorics 2* (1981) 291–298.
[5] R. TIJDEMAN: On a distribution problem in finite and countable sets. *J. Combinatorial Th. 15* (1973) 129–137.
[6] R. TIJDEMAN: The chairman assignment problem. *Discr. Math. 32* (1980) 323–330.
[7] R. TIJDEMAN and M. VOORHOEVE: Bounded discrepancy sets. *Compositio Math. 42* (1981) 375–389.

(Oblatum 6-I-1982)

Mathematics Dept.
SUNY at Stony Brook
Stony Brook, NY 11794
U.S.A.