COMPOSITIO MATHEMATICA

H. G. MEIJER
H. NIEDERREITER

On a distribution problem in finite sets

Compositio Mathematica, tome 25, nº 2 (1972), p. 153-160
<http://www.numdam.org/item?id=CM_1972__25_2_153_0>

COMPOSITIO MATHEMATICA, Vol. 25, Fasc. 2, 1972, pag. 153-160
Wolters-Noordhoff Publishing
Printed in the Netherlands

ON A DISTRIBUTION PROBLEM IN FINITE SETS

by

H. G. Meijer and H. Niederreiter

1.

In [2] the following problem emerged which deserves some interest of its own. Let $X = \{x_1, \cdots, x_k\}$ be a nonvoid finite set and let μ be a measure on X with $\mu(x_i) = \lambda_i > 0$ for $1 \leqq i \leqq k$ and $\sum_{i=1}^{k} \lambda_i = 1$.

Without loss of generality we may suppose that the x_i are arranged in such a way that $\lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_k$. For an infinite sequence ω in X, let $A(i; N; \omega)$ denote the number of occurrences of the element x_i among the first N terms of ω and let $D(\omega) = \sup_{i,N}|A(i; N; \omega) - \lambda_i N|$ (the supremum is taken over $i = 1, 2, \cdots, k$; $N = 1, 2, \cdots$). We pose the problem: *how small can* $D(\omega)$ *be?*

Similarly, define $A(M; N; \omega)$ for a subset M of X to be the number of occurrences of elements from M among the first N terms of ω and put $C(\omega) = \sup_{M,N}|A(M; N; \omega) - \mu(M)N|$ (the supremum is taken over all subsets $M \subset X$ and $N = 1, 2, \cdots$). Then we may ask: *how small can* $C(\omega)$ *be?*

These problems are similar to the well-known problem of constructing a sequence with small discrepancy in the unit interval [0,1] (see e.g. v.d. Corput [1]).

It was shown in [2] that a 'very well' distributed sequence ω in X can be found with

$$D(\omega) \leqq k-1, \qquad C(\omega) \leqq (k-1)\left[\frac{k}{2}\right].$$

Those values, however, are far from being optimal. In section 2 of this paper we shall construct a sequence ω in X with

$$(1) \qquad D(\omega) \leqq \tfrac{1}{2} + \tfrac{1}{2}\sum_{n=1}^{k-2} \frac{1}{n} \quad \text{and} \quad C(\omega) \leqq \tfrac{1}{2}(k-1) \quad \text{for } k \geqq 2.$$

If $k = 1$ then, trivially, $C(\omega) = D(\omega) = 0$.

For some special measures μ on X better results can be obtained. If e.g. $\lambda_1 = \cdots = \lambda_k = 1/k$ then one easily verifies that the sequence $\omega = (y_n)_{n=1}^{\infty}$ defined by $y_n = x_i$ if $n \equiv i \ (\text{mod.}\, k)$ satisfies $D(\omega) = 1 - 1/k$.

In section 3 we construct a sequence η in X which gives a better result than (1) if λ_k is sufficiently small and $k \geqq 3$. In fact we prove

$$D(\eta) \leqq \begin{cases} \frac{1}{2}+\frac{1}{2}\lambda_k\,(k-2) \text{ if } k \text{ is even} \\ \frac{1}{2}+\frac{1}{2}\lambda_k\,(k-1) \text{ if } k \text{ is odd} \end{cases}$$

$$C(\eta) \begin{cases} = D(\eta) & \text{for } k = 2, 3 \\ \leqq \max\,(D(\eta), \frac{5}{4}) & \text{for } k = 4 \\ \leqq \max\,(D(\eta), \frac{25}{16}) & \text{for } k = 5 \\ \leqq \max\,(D(\eta), \frac{1}{2}(k-2)) & \text{for } k \geqq 6. \end{cases}$$

We remark that always $\lambda_k \geqq 1/k$.

Added in proof: Recently Tijdeman [3] found by an entirely different method: if $D_k = \sup_\mu \inf_\omega D(\omega)$, then it holds

$$1-\frac{1}{2(k-1)} \leqq D_k \leqq 1.$$

Moreover he generalized the results to countable sets.

A refinement of this method gives

$$D_k = 1-\frac{1}{2(k-1)}$$

(see [4]).

2.

By using some refinements of the method employed in [2], we can prove the following result.

THEOREM 1. *For any nonvoid finite set $X = \{x_1, \cdots, x_k\}$ and every measure μ* on *X* with *$\mu(x_i) = \lambda_i > 0$ $(i = 1, \cdots, k)$, $\lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_k$ and $\sum_{i=1}^k \lambda_i = 1$, there is a sequence ω in X such that*

$$(2)\qquad \begin{aligned} |A(i; N; \omega)-\lambda_i N| &\leqq \tfrac{1}{2}+\tfrac{1}{2}\sum_{n=1}^{k-i} \frac{1}{n} \quad \text{if} \quad 2 \leqq i \leqq k \\ |A(1; N; \omega)-\lambda_1 N| &\leqq \tfrac{1}{2}\sum_{n=1}^{k-1} \frac{1}{n}, \end{aligned}$$

therefore

$$D(\omega) = 0 \quad \text{if } k = 1,$$

$$D(\omega) \leqq \tfrac{1}{2}+\tfrac{1}{2}\sum_{n=1}^{k-2} \frac{1}{n} \quad \text{if} \quad k \geqq 2;$$

moreover

$$C(\omega) \leqq \tfrac{1}{2}(k-1).$$

PROOF. We proceed by induction on k. Obviously the case $k = 1$ is trivial. Assuming the proposition to be true for an integer $k \geqq 1$, we shall prove that it also holds for $k+1$.

We consider the set $X = \{x_1, \cdots, x_{k+1}\}$ and a measure μ on X with

$$\mu(x_i) = \lambda_i > 0, \qquad \lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_{k+1}, \sum_{i=1}^{k+1} \lambda_i = 1.$$

On the subset $Y = \{x_1, \cdots, x_k\}$ of X, introduce a measure ν by

$$\nu(x_i) = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_k} = \alpha_i.$$

Since $\lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_k$ it follows that

$$\alpha_i \leqq \frac{1}{k-i+1} \quad \text{for} \quad 1 \leqq i \leqq k. \tag{3}$$

By induction hypothesis, there exists a sequence $\tau = (y_n)_{n=1}^{\infty}$ in Y with

$$\begin{aligned} |A(i; N; \tau) - \alpha_i N| &\leqq \tfrac{1}{2} + \tfrac{1}{2} \sum_{n=1}^{k-i} \frac{1}{n} \quad \text{if } 2 \leqq i \leqq k, \\ |A(1; N; \tau) - \alpha_1 N| &\leqq \tfrac{1}{2} \sum_{n=1}^{k-1} \frac{1}{n} \end{aligned} \tag{4}$$

for all $N \geqq 1$, and with

$$C(\tau) \leqq \tfrac{1}{2}(k-1). \tag{5}$$

We introduce the following notation: for a real number a let $||a|| = [a+\frac{1}{2}]$, i.e. the integer nearest to a. For $n \geqq 1$, put $R(n) = n - ||\lambda_{k+1} n||$. We define a sequence $\omega = (z_n)_{n=1}^{\infty}$ in X by setting

$$\begin{aligned} z_n &= x_{k+1} \quad \text{if} \quad ||\lambda_{k+1} n|| > ||\lambda_{k+1}(n-1)||, \\ z_n &= y_{R(n)} \quad \text{if} \quad ||\lambda_{k+1} n|| = ||\lambda_{k+1}(n-1)||. \end{aligned} \qquad (n = 1, 2, \cdots)$$

We get then

$$A(k+1; N; \omega) = ||\lambda_{k+1} N|| = \lambda_{k+1} N + \varepsilon$$

with $|\varepsilon| \leqq \frac{1}{2}$, and therefore

$$|A(k+1; N; \omega) - \lambda_{k+1} N| \leqq \tfrac{1}{2}. \tag{6}$$

For $1 \leqq i \leqq k$, we have $A(i; N; \omega) = A(i; R(N); \tau)$ for all $N \geqq 1$ (if $R(N) = 0$, we had to read $A(i; R(N); \tau) = 0$). Now we write

$$|A(i; N; \omega) - \lambda_i N| \leqq |A(i; R(N); \tau) - \alpha_i R(N)| + |\alpha_i R(N) - \lambda_i N|. \tag{7}$$

Using the definitions of $R(N)$ of α_i and (3), we obtain

$$|\alpha_i R(N) - \lambda_i N| = |\alpha_i(N - \lambda_{k+1} N - \varepsilon) - \lambda_i N| = |\alpha_i \varepsilon| \leqq \frac{1}{2(k-i+1)}. \tag{8}$$

Hence by (7), (4) and (8) we get

$$|A(i;N;\omega)-\lambda_i N| \leqq \tfrac{1}{2}+\tfrac{1}{2}\sum_{n=1}^{k-i+1}\frac{1}{n} \quad \text{if } 2 \leqq i \leqq k,$$

$$|A(1;N;\omega)-\lambda_1 N| \leqq \tfrac{1}{2}\sum_{n=1}^{k}\frac{1}{n}.$$

Moreover (6) implies that the first inequality also holds for $i = k+1$. Therefore the relations (2) have been proved for $k+1$.

Furthermore we have to show that ω satisfies $C(\omega) \leqq k/2$. If M is a subset of X and M^c denotes its complement in X, then

$$(9) \qquad |A(M^c;N;\omega)-\mu(M^c)N| = |A(M;N;\omega)-\mu(M)N|.$$

Consequently, it suffices to consider subsets M of Y. Using (5) and the same type of arguments as above, we arrive at

$$\begin{aligned}|A(M;N;\omega)-\mu(M)N| &\leqq |A(M;R(N);\tau)-\nu(M)R(N)| \\ &\quad +|\nu(M)R(N)-\mu(M)N| \leqq \tfrac{1}{2}(k-1)+|\nu(M)\varepsilon| \leqq \tfrac{1}{2}k.\end{aligned}$$

3.

In this section we exhibit another construction principle which gives better results than the sequence of section 2 if $\lambda_k = \max \lambda_i$ is small and $k \geqq 3$. Since the case $k = 1$ is trivial we restrict ourselves to $k \geqq 2$. For a real number a we denote as above $||a|| = [a+\frac{1}{2}]$; moreover we define

$$(10) \qquad \{\{a\}\} = a-||a||.$$

Hence

$$(11) \qquad -\tfrac{1}{2} \leqq \{\{a\}\} < \tfrac{1}{2}.$$

We consider the following scheme consisting of an infinite number of rows and k columns.

x_1 λ_1	x_2 λ_2	$\cdots$ $\cdots$	x_k λ_k	
$\|\|\lambda_1\|\|$	$\|\|\lambda_2\|\|$	$\cdots$	$\|\|\lambda_k\|\|$	1st row
$\|\|2\lambda_1\|\|$	$\|\|2\lambda_2\|\|$	$\cdots$	$\|\|2\lambda_k\|\|$	2nd row
$\|\|n\lambda_1\|\|$	$\|\|n\lambda_2\|\|$	$\cdots$	$\|\|n\lambda_k\|\|$	nth row

The i^{th} column consists of $||\lambda_i|| \leqq ||2\lambda_i|| \leqq \cdots \leqq ||n\lambda_i|| \leqq \cdots$, where $||(n+1)\lambda_i|| = ||n\lambda_i||$ or $||(n+1)\lambda_i|| = ||n\lambda_i||+1$. Now we change this column in the following way.

If $||(n+1)\lambda_i|| = ||n\lambda_i||$ $(n = 0, 1, 2, \cdots)$ we omit $||(n+1)\lambda_i||$ such that we get a void place in the scheme.

If on the other hand $||(n+1)\lambda_i|| = ||n\lambda_i||+1$ $(n = 0, 1, 2, \cdots)$ we replace $||(n+1)\lambda_i||$ by x_i. We remark that in the last case

$$\{\{n\lambda_i\}\} \geqq \tfrac{1}{2}-\lambda_i, \tag{12}$$

$$\{\{(n+1)\lambda_i\}\} < -\tfrac{1}{2}+\lambda_i. \tag{13}$$

The i^{th} column now consists of places with x_i and void places. Up till the n^{th} row there are exactly $||n\lambda_i||$ places with x_i. We do so for $i = 1, 2, \cdots, k$. The sequence $\eta = (\eta_n)_{n=1}^{\infty}$ is the sequence which we get if we read the consecutive rows from the left to the right. After we have passed through the n^{th} row we have had $||n\lambda_i||$ times the element x_i and altogether $T(n) = \sum_{i=1}^{k}||n\lambda_i||$ elements of η. For this sequence η we will prove the following result.

THEOREM 2. *For the sequence* η *we have*

$$|A(i; N; \eta)-\lambda_i N| \leqq \tfrac{1}{2}+\tfrac{1}{2}\lambda_i(k-d), \tag{14}$$

where $d = 1$ *if* k *is odd and* $d = 2$ *if* k *is even. Therefore*

$$D(\eta) \leqq \tfrac{1}{2}+\tfrac{1}{2}\lambda_k(k-d).$$

Moreover

$$C(\eta)\begin{cases} = D(\eta) & \text{for } k = 2, 3 \\ \leqq \max\,(D(\eta), \tfrac{5}{4}) & \text{for } k = 4 \\ \leqq \max\,(D(\eta), \tfrac{25}{16}) & \text{for } k = 5 \\ \leqq \max\,(D(\eta), (k-2)/2) & \text{for } k \geqq 6. \end{cases}$$

PROOF. Since there is no risk of ambiguity we omit the η in $A(i; N; \eta)$ and $A(M, N; \eta)$.

First we remark that by (10)

$$\sum_{h=1}^{k}\{\{n\lambda_h\}\} = n-\sum_{h=1}^{k}||n\lambda_h||,$$

which implies that $\Sigma\{\{n\lambda_h\}\}$ has to be an integer. If we exclude the case k even, $(\{\{n\lambda_1\}\}, \cdots, \{\{n\lambda_k\}\}) = (-\frac{1}{2}, \cdots, -\frac{1}{2})$ we may conclude from (11)

$$\Big|\sum_{h=1}^{k}\{\{n\lambda_h\}\}\Big| \leqq \tfrac{1}{2}(k-d), \tag{15}$$

where $d = 1$ if k is odd, $d = 2$ if k is even. Using again (10) we get

$$\begin{aligned} A(i; T(n))-\lambda_i T(n) &= ||n\lambda_i||-\lambda_i\sum_{h=1}^{k}||n\lambda_h|| \\ &= -\{\{n\lambda_i\}\}+\lambda_i\sum_{h=1}^{k}\{\{n\lambda_h\}\}. \end{aligned} \tag{16}$$

Let N be an integer with $T(n) \leqq N \leqq T(n+1)$. Then $A(i;N) = A(i;T(n))$ or $A(i;N) = A(i;T(n))+1$. In the first case we have by (16), (11) and (15)

$$A(i;N)-\lambda_i N \leqq A(i;T(n))-\lambda_i T(n) = -\{\{n\lambda_i\}\}+\lambda_i \sum_{h=1}^{k} \{\{n\lambda_h\}\}$$
$$\leqq \tfrac{1}{2}+\tfrac{1}{2}\lambda_i(k-d).$$

In the second case x_i is an element of the $(n+1)^{\text{th}}$ row. Then by (12)

$$\{\{n\lambda_i\}\} \geqq \tfrac{1}{2}-\lambda_i. \tag{17}$$

Moreover $N \geqq T(n)+1$. Therefore using (16), (17) and (15) we arrive at

$$\begin{aligned} A(i;N)-\lambda_i N &\leqq A(i;T(n))-\lambda_i T(n)+1-\lambda_i \\ &= -\{\{n\lambda_i\}\}+\lambda_i \sum_{h=1}^{k} \{\{n\lambda_h\}\}+1-\lambda_i \leqq -\tfrac{1}{2}+\lambda_i+\tfrac{1}{2}\lambda_i(k-d)+1-\lambda_i \\ &= \tfrac{1}{2}+\tfrac{1}{2}\lambda_i(k-d). \end{aligned}$$

This upper bound trivially holds as well with $d = 2$ in the exceptional case excluded above.

In order to get a lower bound we proceed in a similar way. We have $A(i;N) = A(i;T(n+1))$ or $A(i;N) = A(i;T(n+1))-1$.

For the calculations we first exclude the case k even,

$$(\{\{(n+1)\lambda_1\}\}, \cdots, \{\{(n+1)\lambda_k\}\}) = (-\tfrac{1}{2}, \cdots, -\tfrac{1}{2}).$$

Then we obtain in the first case

$$\begin{aligned} A(i;N)-\lambda_i N &\geqq A(i;T(n+1))-\lambda_i T(n+1) \\ &= -\{\{(n+1)\lambda_i\}\}+\lambda_i \sum_{h=1}^{k} \{\{(n+1)\lambda_h\}\} \geqq -\tfrac{1}{2}-\tfrac{1}{2}\lambda_i(k-d). \end{aligned}$$

In the second case we have $N \leqq T(n+1)-1$. Moreover x_i occurs in the $(n+1)^{\text{th}}$ row and (13) gives

$$\{\{(n+1)\lambda_i\}\} < -\tfrac{1}{2}+\lambda_i.$$

Therefore

$$\begin{aligned} A(i;N)-\lambda_i N &\geqq A(i;T(n+1))-\lambda_i T(n+1)-1+\lambda_i \\ &= -\{\{(n+1)\lambda_i\}\}+\lambda_i \sum_{h=1}^{k} \{\{(n+1)\lambda_h\}\}-1+\lambda_i \\ &\geqq \tfrac{1}{2}-\lambda_i-\tfrac{1}{2}\lambda_i(k-d)-1+\lambda_i = -\tfrac{1}{2}-\tfrac{1}{2}\lambda_i(k-d). \end{aligned}$$

One easily verifies that these lower bounds also hold with $d = 2$ for the case k even,

$$(\{\{(n+1)\lambda_1\}\}, \cdots, \{\{(n+1)\lambda_k\}\}) = (-\tfrac{1}{2}, \cdots, -\tfrac{1}{2}).$$

Hence (14) has been proved.

In order to get an estimate for $C(\eta)$ we consider a nonvoid subset M of X. Put

$$M = \{x_{i_1}, \cdots, x_{i_j}\}, \quad \mu M = \lambda_{i_1} + \cdots + \lambda_{i_j} = \Lambda,$$
$$X \setminus M = \{x_{i_{j+1}}, \cdots, x_{i_k}\}.$$

Then

$$\begin{aligned} A(M; T(n)) - \Lambda T(n) &= \sum_{\nu=1}^{j} \|n\lambda_{i_\nu}\| - \Lambda \sum_{h=1}^{k} \|n\lambda_h\| \\ &= -\sum_{\nu=1}^{j} \{\{n\lambda_{i_\nu}\}\} + \Lambda \sum_{h=1}^{k} \{\{n\lambda_h\}\} \\ &= -(1-\Lambda) \sum_{\nu=1}^{j} \{\{n\lambda_{i_\nu}\}\} + \Lambda \sum_{\nu=j+1}^{k} \{\{n\lambda_{i_\nu}\}\}. \end{aligned}$$

Let N be an integer with $T(n) \leqq N \leqq T(n+1)$ and suppose

$$A(M; N) = A(M; T(n)) + t \text{ with } 0 \leqq t \leqq j.$$

Then $N \geqq T(n)+t$ and

$$\begin{aligned} A(M; N) - \Lambda N &\leqq A(M; T(n)) - \Lambda T(N) + t - \Lambda t \\ &= -(1-\Lambda) \sum_{\nu=1}^{j} \{\{n\lambda_{i_\nu}\}\} + \Lambda \sum_{\nu=j+1}^{k} \{\{n\lambda_{i_\nu}\}\} + t - \Lambda t. \end{aligned}$$

Suppose that $x_{u_1}, \cdots x_{u_t}$ are the elements of the $(n+1)^{\text{th}}$ row which are counted in $A(M; N)$ and not in $A(M; T(n))$.

Then by (12)

$$\{\{n\lambda_{u_\tau}\}\} \geqq \tfrac{1}{2} - \lambda_{u_\tau}. \qquad (\tau = 1, \cdots, t)$$

Therefore

$$\sum_{\nu=1}^{j} \{\{n\lambda_{i_\nu}\}\} \geqq \tfrac{1}{2}t - (\lambda_{u_1} + \cdots + \lambda_{u_t}) - \tfrac{1}{2}(j-t) \geqq t - \tfrac{1}{2}j - \Lambda.$$

Hence

$$\begin{aligned} A(M; N) - \Lambda N &\leqq -(1-\Lambda)(t - \tfrac{1}{2}j - \Lambda) + \tfrac{1}{2}\Lambda(k-j) + t - \Lambda t \\ &= \frac{j}{2} + \Lambda \left(\frac{k}{2} - j + 1 \right) - \Lambda^2. \end{aligned}$$

In a similar way we find a lower bound for $A(M; N) - \Lambda N$ which has the same absolute value.

Hence

$$|A(M; N) - \Lambda N| \leqq \frac{j}{2} + \Lambda \left(\frac{k}{2} - j + 1 \right) - \Lambda^2.$$

Since for $k = 2$, trivially, $C(\eta) = D(\eta)$, we suppose $k \geqq 3$. We observe that we can restrict ourselves to $\frac{1}{2}k \leqq j \leqq k-1$ (compare (9)). If $j = k-1$, the complement of M is a singleton which was dealt with in $D(\eta)$. In particular this implies $C(\eta) = D(\eta)$ for $k = 3$. If $\frac{1}{2}k+1 \leqq j \leqq k-2$, then clearly

$$\frac{j}{2} + \Lambda\left(\frac{k}{2} - j + 1\right) - \Lambda^2 \leqq \tfrac{1}{2}(k-2).$$

If $\frac{1}{2}k \leqq j \leqq \frac{1}{2}k+\frac{1}{2}$, then

$$\frac{j}{2} + \Lambda\left(\frac{k}{2} - j + 1\right) - \Lambda^2 \leqq \tfrac{1}{2}\left(\frac{k}{2} + \tfrac{1}{2}\right) + \Lambda - \Lambda^2 \leqq \frac{k}{4} + \tfrac{1}{2}.$$

For $k \geqq 6$, we have $\frac{1}{4}k+\frac{1}{2} \leqq (k-2)/2$ and so $C(\eta) \leqq \max(D(\eta), (k-2)/2)$. For $k = 4, 5$ one finds by separate discussion of the permissible values for $j : C(\eta) \leqq \max(D(\eta), \frac{5}{4})$ for $k = 4$, $C(\eta) \leqq \max(D(\eta), \frac{25}{16})$ for $k = 5$. This completes the proof.

REFERENCES

J. G. v. d. Corput

[1] Verteilungsfunktionen II, Nederl. Akad. Wetensch. Proc. Ser. A. 38 (1935), 1058–1066.

H. Niederreiter

[2] On the existence of uniformly distributed sequences in compact spaces, submitted for publication.

R. Tijdeman

[3] On a distribution problem in finite and countable sets, submitted for publication.

H. G. Meijer

[4] On a distribution problem in finite sets, to appear.

H. Niederreiter

[5] A distribution problem in finite sets, Proc. Symp. on Applications of Number Theory to Numerical Analysis (Montréal, 1971) Academic Press, New York, to appear.

(Oblatum 28–VI–1971 and 7–X–1971)

University of Technology
Department of Mathematics
Delft, Netherlands
and
Southern Illinois University
Department of Mathematics
Carbondale, Illinois 62901