H. G. MEIJER
H. NIEDERREITER
On a distribution problem in finite sets


<http://www.numdam.org/item?id=CM_1972__25_2_153_0>
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 \( \mu \) be a measure on \( X \) with \( \mu(x_i) = \lambda_i > 0 \) for \( 1 \leq i \leq k \) 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 \( \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k \). For an infinite sequence \( \omega \) in \( X \), let \( A(i; N; \omega) \) denote the number of occurrences of the element \( x_i \) among the first \( N \) terms of \( \omega \) 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 \( \omega \) and put \( C(\omega) = \sup_{M,N} |A(M; N; \omega) - \mu(M) N| \) (the supremum is taken over all subsets \( M \subseteq X \) 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 \( \omega \) in \( X \) can be found with

\[
D(\omega) \leq k - 1, \quad C(\omega) \leq (k-1) \left\lfloor \frac{k}{2} \right\rfloor.
\]

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

\[
D(\omega) \leq \frac{1}{2} + \frac{1}{2} \sum_{n=1}^{k-2} \frac{1}{n} \quad \text{and} \quad C(\omega) \leq \frac{1}{2}(k-1) \quad \text{for} \quad k \geq 2.
\]

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

For some special measures \( \mu \) 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 \( n \equiv i \mod k \) satisfies \( D(\omega) = 1 - 1/k \).
In section 3 we construct a sequence \( \eta \) in \( X \) which gives a better result than (1) if \( \lambda_k \) is sufficiently small and \( k \geq 3 \). In fact we prove

\[
D(\eta) \leq \begin{cases} 
\frac{1}{2} + \frac{1}{2} \lambda_k (k-2) & \text{if } k \text{ is even} \\
\frac{1}{2} + \frac{1}{2} \lambda_k (k-1) & \text{if } k \text{ is odd}
\end{cases}
\]

We remark that always \( \lambda_k \geq 1/k \).

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

\[
1 - \frac{1}{2(k-1)} \leq D_k \leq 1.
\]

Moreover he generalized the results to countable sets.

A refinement of this method gives

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

(see [4]).

\section{2.}

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

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

\[
|A(i; N; \omega) - \lambda_i N| \leq \frac{k-i}{2} + \frac{k-i}{n=1} \frac{1}{n}
\]

if \( 2 \leq i \leq k \)

\[
|A(1; N; \omega) - \lambda_1 N| \leq \frac{k-1}{2} \sum_{n=1}^{k-1} \frac{1}{n},
\]

therefore

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

\[
D(\omega) \leq \frac{1}{2} + \frac{k-2}{n=1} \frac{1}{n} \quad \text{if } k \geq 2;
\]

moreover

\[
C(\omega) \leq \frac{1}{2}(k-1).
\]
PROOF. We proceed by induction on $k$. Obviously the case $k = 1$ is trivial. Assuming the proposition to be true for an integer $k \geq 1$, we shall prove that it also holds for $k+1$.

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

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

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

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

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

$$(3) \quad \alpha_i \leq \frac{1}{k-i+1} \quad \text{for} \quad 1 \leq i \leq k.$$ 

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

$$|A(i; N; \tau) - \alpha_i N| \leq \frac{1}{2} + \frac{1}{2} \sum_{n=1}^{k-i} \frac{1}{n} \quad \text{if} \quad 2 \leq i \leq k,$$

$$(4) \quad |A(1; N; \tau) - \alpha_1 N| \leq \frac{1}{2} \sum_{n=1}^{k-1} \frac{1}{n}$$

for all $N \geq 1$, and with

$$(5) \quad C(\tau) \leq \frac{1}{2}(k-1).$$

We introduce the following notation: for a real number $a$ let $||a|| = [a + \frac{1}{2}]$, i.e. the integer nearest to $a$. For $n \geq 1$, put $R(n) = n - ||\lambda_{k+1} n||$.

We define a sequence $\omega = (z_n)_{n=1}^{\infty}$ in $X$ by setting

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

We get then

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

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

$$(6) \quad |A(k+1; N; \omega) - \lambda_{k+1} N| \leq \frac{1}{2}.$$ 

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

$$(7) \quad |A(i; N; \omega) - \lambda_i N| \leq |A(i; R(N); \tau) - \alpha_i R(N)| + |\alpha_i R(N) - \lambda_i N|.$$ 

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

$$(8) \quad |\alpha_i R(N) - \lambda_i N| = |\alpha_i (N - \lambda_{k+1} N + \varepsilon) - \lambda_i N| = |\alpha_i \varepsilon| \leq \frac{1}{2(k-i+1)}.$$
Hence by (7), (4) and (8) we get
\[
|A(i; N; \omega) - \lambda_i N| \leq \frac{1}{2} + \frac{1}{2} \sum_{n=1}^{k-i+1} \frac{1}{n} \quad \text{if } 2 \leq i \leq k,
\]
\[
|A(1; N; \omega) - \lambda_1 N| \leq \frac{1}{2} \sum_{n=1}^{k} \frac{1}{n}.
\]
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 \(\omega\) satisfies \(C(\omega) \leq k/2\). If \(M\) is a subset of \(X\) and \(M^c\) denotes its complement in \(X\), then
\[
|A(M^c; N; \omega) - \mu(M^c) N| = |A(M; N; \omega) - \mu(M) N|.
\]
Consequently, it suffices to consider subsets \(M\) of \(Y\). Using (5) and the same type of arguments as above, we arrive at
\[
|A(M; N; \omega) - \mu(M) N| \leq |A(M; R(N); \tau) - v(M) R(N)| + |v(M) R(N) - \mu(M) N| \leq \frac{1}{2}(k-1) + |v(M)\omega| \leq \frac{1}{2}k.
\]

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 \(k \geq 3\). Since the case \(k = 1\) is trivial we restrict ourselves to \(k \geq 2\). For a real number \(a\) we denote as above \(|a| = [a + \frac{1}{2}]\); moreover we define
\[
\{\{a\}\} = a - |a|.
\]
Hence
\[
-\frac{1}{2} \leq \{\{a\}\} < \frac{1}{2}.
\]
We consider the following scheme consisting of an infinite number of rows and \(k\) columns.

| \(x_1\) | \(x_2\) | \cdots | \(x_k\) |
| \(\lambda_1\) | \(\lambda_2\) | \cdots | \(\lambda_k\) |
| \(\|\lambda_1\|\) | \(\|\lambda_2\|\) | \cdots | \(\|\lambda_k\|\) |
| \(\|2\lambda_1\|\) | \(\|2\lambda_2\|\) | \cdots | \(\|2\lambda_k\|\) |
| \(\|n\lambda_1\|\) | \(\|n\lambda_2\|\) | \cdots | \(\|n\lambda_k\|\) |

The \(i^{th}\) column consists of \(\|\lambda_i\| \leq \|2\lambda_i\| \leq \cdots \leq \|n\lambda_i\| \leq \cdots\), 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

\[
\{\{n\lambda_i\}\} \geq \frac{1}{2} - \lambda_i,
\]

\[
\{\{(n+1)\lambda_i\}\} < -\frac{1}{2} + \lambda_i.
\]

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 \(\eta\). For this sequence \(\eta\) we will prove the following result.

**Theorem 2.** For the sequence \(\eta\) we have

\[
|A(i; N; \eta) - \lambda_i N| \leq \frac{1}{2} + \frac{1}{2}\lambda_i(k-d),
\]

where \(d = 1\) if \(k\) is odd and \(d = 2\) if \(k\) is even. Therefore

\[
D(\eta) \leq \frac{1}{2} + \frac{1}{2}\lambda_k(k-d).
\]

Moreover

\[
C(\eta) \begin{cases} 
= D(\eta) & \text{for } k = 2, 3 \\
\leq \max(D(\eta), \frac{k}{4}) & \text{for } k = 4 \\
\leq \max(D(\eta), \frac{2k}{5}) & \text{for } k = 5 \\
\leq \max(D(\eta), (k-2)/2) & \text{for } k \geq 6.
\end{cases}
\]

**Proof.** Since there is no risk of ambiguity we omit the \(\eta\) 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)

\[
|\sum_{h=1}^k \{\{n\lambda_h\}\}| \leq \frac{1}{2}(k-d),
\]

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

\[
A(i; T(n)) - \lambda_i T(n) = |n\lambda_i| - \lambda_i \sum_{h=1}^k |n\lambda_h|\]

\[
= -\{\{n\lambda_i\}\} + \lambda_i \sum_{h=1}^k \{\{n\lambda_h\}\}.
\]
Let $N$ be an integer with $T(n) \leq N \leq T(n+1)$. 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)
\[
A(i; N) - \lambda_i N \leq A(i; T(n)) - \lambda_i T(n) = -\{\{n\lambda_i\}\} + \lambda_i \sum_{h=1}^{k} \{\{n\lambda_h\}\} \\
\leq -\frac{1}{2} + \frac{1}{2}\lambda_i (k-d).
\]
In the second case $x_i$ is an element of the $(n+1)^{\text{th}}$ row. Then by (12)
\[
\{\{n\lambda_i\}\} \geq \frac{1}{2} - \lambda_i.
\]
Moreover $N \geq T(n) + 1$. Therefore using (16), (17) and (15) we arrive at
\[
A(i; N) - \lambda_i N \leq A(i; T(n)) - \lambda_i T(n) + 1 - \lambda_i \\
= -\{\{n\lambda_i\}\} + \lambda_i \sum_{h=1}^{k} \{\{n\lambda_h\}\} + 1 - \lambda_i \leq -\frac{1}{2} + \frac{1}{2}\lambda_i (k-d) + 1 - \lambda_i \\
= \frac{1}{2} + \frac{1}{2}\lambda_i (k-d).
\]
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,
\[
(\{\{(n+1)\lambda_i\}\}, \ldots, \{\{(n+1)\lambda_k\}\}) = (-\frac{1}{2}, \ldots, -\frac{1}{2}).
\]
Then we obtain in the first case
\[
A(i; N) - \lambda_i N \geq A(i; T(n+1)) - \lambda_i T(n+1) \\
= -\{\{n+1\lambda_i\}\} + \lambda_i \sum_{h=1}^{k} \{\{(n+1)\lambda_h\}\} \geq -\frac{1}{2} - \frac{1}{2}\lambda_i (k-d).
\]
In the second case we have $N \leq T(n+1) - 1$. Moreover $x_i$ occurs in the $(n+1)^{\text{th}}$ row and (13) gives
\[
\{\{(n+1)\lambda_i\}\} < -\frac{1}{2} + \lambda_i.
\]
Therefore
\[
A(i; N) - \lambda_i N \geq A(i; T(n+1)) - \lambda_i T(n+1) - 1 + \lambda_i \\
= -\{\{n+1\lambda_i\}\} + \lambda_i \sum_{h=1}^{k} \{\{(n+1)\lambda_h\}\} + 1 + \lambda_i \\
\geq \frac{1}{2} - \lambda_i - \frac{1}{2}\lambda_i (k-d) + 1 + \lambda_i = -\frac{1}{2} - \frac{1}{2}\lambda_i (k-d).
\]
One easily verifies that these lower bounds also hold with $d = 2$ for the case $k$ even,
\[
(\{\{(n+1)\lambda_i\}\}, \ldots, \{\{(n+1)\lambda_k\}\}) = (-\frac{1}{2}, \ldots, -\frac{1}{2}).
\]
Hence (14) has been proved.
In order to get an estimate for $C(\eta)$ we consider a nonvoid subset $M$ of $X$. Put

$$M = \{x_{i_1}, \ldots, x_{i_j}\}, \quad \mu M = \lambda_{i_1} + \cdots + \lambda_{i_j} = \Lambda,$$

$$X \setminus M = \{x_{i_{j+1}}, \ldots, x_{i_k}\}.$$  

Then

$$A(M; T(n)) - \Delta T(n) = \sum_{v=1}^{j} ||n\lambda_{i_v}|| - \Delta \sum_{h=1}^{k} ||n\lambda_h||$$

$$= -\sum_{v=1}^{j} \{\{n\lambda_{i_v}\}\} + \Delta \sum_{h=1}^{k} \{\{n\lambda_h\}\}$$

$$= -(1-\Delta) \sum_{v=1}^{j} \{\{n\lambda_{i_v}\}\} + \Delta \sum_{v=j+1}^{k} \{\{n\lambda_{i_v}\}\}.$$  

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

$$A(M; N) = A(M; T(n)) + t$$

with $0 \leq t \leq j$. Therefore

$$A(M; N) - \Delta N \leq A(M; T(n)) - \Delta (N + t) - \Delta t$$

$$= -(1-\Delta) \sum_{v=1}^{j} \{\{n\lambda_{i_v}\}\} + \Delta \sum_{v=j+1}^{k} \{\{n\lambda_{i_v}\}\} + t - \Delta t.$$  

Suppose that $x_{u_1}, \ldots, x_{u_t}$ are the elements of the $(n+1)$th row which are counted in $A(M; N)$ and not in $A(M; T(n))$. Then by (12)

$$\{\{n\lambda_{u_v}\}\} \geq \frac{1}{2} - \lambda_{u_v}, \quad (\tau = 1, \ldots, t)$$

Therefore

$$\sum_{v=1}^{j} \{\{n\lambda_{u_v}\}\} \geq \frac{1}{2} t - (\lambda_{u_1} + \cdots + \lambda_{u_t}) - \frac{1}{2} (j-t) \geq t - \frac{1}{2} j - \Delta.$$

Hence

$$A(M; N) - \Delta N \leq -(1-\Delta)(t - \frac{1}{2} j - \Delta) + \frac{1}{2} \Delta (k - j) + t - \Delta t$$

$$= \frac{j}{2} + \Delta \left(\frac{k}{2} - j + 1\right) - \Delta^2.$$  

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

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

Since for $k = 2$, trivially, $C(\eta) = D(\eta)$, we suppose $k \geq 3$. We observe
that we can restrict ourselves to $\frac{1}{2}k \leq j \leq k - 1$ (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 $\frac{1}{2}k + 1 \leq j \leq k - 2$, then clearly

$$\frac{j}{2} + A \left( \frac{k}{2} - j + 1 \right) - A^2 \leq \frac{1}{2}(k - 2).$$

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

$$\frac{j}{2} + A \left( \frac{k}{2} - j + 1 \right) - A^2 \leq \frac{j}{2} \left( \frac{k}{2} + \frac{1}{2} \right) + A - A^2 \leq \frac{k}{4} + \frac{1}{2}.$$

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

REFERENCES

J. G. V. d. CORPUT

H. NIEDERREITER

R. TIDEKAM

H. G. MEIJER

H. NIEDERREITER

(Oblatum 28–VI–1971 and 7–X–1971)