Equidistributed Consecutive Blocks in Binary Sequences

Shoni Gilboa∗ Romans Pinchasi†

June 30, 2013

Abstract

We show that for every k there is N(k) such that in every 0/1 sequence of length N(k) one can find k consecutive blocks (not necessarily of the same size) such that the percentage of 1’s in each block is the same for all blocks. We discuss related problems and extensions.

1 Introduction

A block in a sequence of elements is a finite subsequence consisting of consecutive elements in the sequence. Blocks are also sometimes called segments. It is an easy exercise to show that in every 0/1 sequence of length 4 there must be two consecutive equal blocks of some size. It is also well known that one can construct infinite such sequences that do not contain three consecutive equal blocks of any size. One such example (as shown in [6]) is the Thue-Morse sequence.

Erdős raised (see [2], p.240) the question of the maximum length of a sequence on r symbols such that no two consecutive blocks are obtained from each other by a permutation of their elements. For r ≤ 3 this maximum length is 2r − 1, but it turns out that for r large enough there are infinite such sequences. Evdokimov ([3]) first showed it for r = 25, then Pleasants ([7]) showed it for r = 5, and finally Keränen ([5]) showed it for r = 4.

The existence of an infinite sequence on 2 symbols such that no k consecutive blocks are obtained from each other by a permutation of their elements, was also considered. Justin ([4]) first proved that such a sequence exist for k = 5, and then Dekking ([1]) constructed such a sequences for k = 4 (while it is easy to check that for k = 3 such sequences do not exist).

One can rephrase the result in [1] as follows: There exist infinite sequences of 0’s and 1’s such that among every four consecutive blocks of the same size one can find two blocks with different sums of elements.

In this paper we show that although one can avoid a situation where k (even four) consecutive blocks have the same size and sum, one cannot avoid k consecutive blocks (of possibly different lengths) with pairwise equal averages.

For a finite sequence of numbers B we denote by s(B) the sum of all numbers in B and we denote by ℓ(B) the number of elements in B.

∗Mathematics Dept., The Open University of Israel, Raanana 43107, Israel. tipshoni@gmail.com, Tel: 972-77-7081316.
†Mathematics Dept., Technion—Israel Institute of Technology, Haifa 32000, Israel. room@math.technion.ac.il. Supported by ISF grant (grant No. 1357/12) and by BSF grant (grant No. 2008290).
Theorem 1. For every positive integer $k$ there exists $N(k)$ such that any sequence of 0’s and 1’s of length at least $N(k)$ must contain $k$ consecutive blocks $B_1, \ldots, B_k$ such that $\frac{s(B_1)}{\ell(B_1)} = \ldots = \frac{s(B_k)}{\ell(B_k)}$.

We note that one can construct infinite 0/1-sequences that do not contain an infinite sequence of consecutive blocks $\{B_i\}_{i=1}^\infty$ where $\frac{s(B_i)}{\ell(B_i)}$ is independent of $i$. One such simple explicit construction is the sequence $\{a_n\}_{n=1}^\infty$ defined by $a_{2^k} = 1$ and $a_n = 0$ for every $n$ that is not a square of an integer. Assume to the contrary that one can find in this sequence an infinite set of consecutive blocks $B_1, B_2, \ldots$ such that for all $i$ we have $\frac{s(B_i)}{\ell(B_i)} = c$, for some constant $c$. Then clearly $c > 0$ for otherwise the sequence $\{a_n\}_{n=1}^\infty$ cannot contain more than a finite number of nonzero elements. However, assuming $c > 0$ also leads to a contradiction because it implies $\limsup \frac{\sum_{i=1}^n a_i}{n} \geq c > 0$, which is clearly false.

2 A geometric interpretation

In this section we present a geometric interpretation of the result in Theorem 1 that will be helpful in the presentation of its proof. We also bring a slightly weaker result than Theorem 1, namely Theorem 3 below, whose statement is very natural in this geometric setting and whose proof is much shorter and more elegant.

Definition 1. A binary path on the two dimensional integer grid is a (finite or even infinite) sequence of integer grid points $\{(x_i, y_i)\}_{i=0}^L$, called the vertices, such that $x_{n+1} = x_n + 1$ and $y_n \leq y_{n+1} \leq y_n + 1$ for every $n \geq 1$. The edges of the path are the straight line segments connecting pairs of consecutive grid points in the binary path. The length of a binary path is the number of edges of the path (that is, $\{(x_i, y_i)\}_{i=0}^L$ has length $L$).

We note that to every binary path $\{(x_i, y_i)\}_{i=0}^L$ of length $L$ corresponds in a natural and straightforward way a 0/1-sequences $\{a_i\}_{i=1}^L$ of length $L$. This correspondence is as follows: For every $1 \leq i \leq n$ we define $a_i = y_i - y_{i-1}$. This correspondence is one to one if we restrict ourselves to binary paths $\{(x_i, y_i)\}_{i=0}^L$ where $(x_0, y_0) = (0, 0)$. That is, the 0/1 sequence $\{a_i\}_{i=1}^L$ corresponds to the binary path $\{(x_i, y_i)\}_{i=0}^L$ defined by $(x_0, y_0) = (0, 0)$ and for every $1 \leq n \leq L$ we define $x_n = n$ and $y_n = \sum_{i=1}^n a_i$.

Suppose now that $A$ is a 0/1 sequence that contains $k$ consecutive blocks $B_1, \ldots, B_k$ with $\frac{s(B_1)}{\ell(B_1)} = \ldots = \frac{s(B_k)}{\ell(B_k)}$. We claim that this happens if and only if the binary path that corresponds to $A$ has $k+1$ collinear vertices. We leave this easy fact as an exercise to the reader.

In view of the above observation, we can reformulate Theorem 1 as follows:

Theorem 2. For every natural number $k$ there exists $N(k)$ such that any binary path of length at least $N(k)$ contains $k$ collinear vertices.

We will now state and prove a weaker theorem that is very close in spirit to Theorem 2 but whose proof is much shorter and of independent interest.

Theorem 3. $\{(x_i, y_i)\}_{i=0}^\infty$ be a binary path. Then for every $k$ there is a line meeting at least $k$ edges of the path.

We note that it is not always possible to find a line that meets infinitely many edges of the path in Theorem 3. This can be seen for example by considering the path defined by $(x_n, y_n) = (n, \lfloor \sqrt{n} \rfloor)$.
for every $n \geq 0$ (which is the binary path corresponding to the 0/1 sequence discussed after Theorem 1). Notice also that by a standard compactness argument it follows from Theorem 3 that for every $k$ there exists $N(k)$ such that for every binary path of length at least $N(k)$ there is a line meeting at least $k$ edges of the path.

**Proof of Theorem 3.** Without loss of generality assume that $(x_0, y_0) = (0, 0)$. Observe that $0 \leq y_n \leq n$ and $x_n = n$ for every $n \geq 0$. It is easy to see that if $\lim \frac{y_n}{n}$ does not exist, then there is a line through the origin that meets infinitely many edges of the path $P$. Indeed, any line whose slope is between $\lim \inf e$ or equal to $\frac{a}{n}$ the upper and lower bounds for the area of $\Delta$ is equal to $1$, that is, $(1 + a)/2$ with $\alpha(e_n)$. To have an estimate on $\alpha(e)$ we consider the triangle $\Delta_n$ with vertices $O, A, B$ and $C$. The area of $\Delta_n$ is equal to $\frac{1}{2} \sin(\alpha(e_n)) ||OA||OB||$ which in turn is less than or equal to $\alpha(e_n)n^2$. To get a lower bound in the area of $\Delta_n$ we consider two cases:

**Case 1.** $e_n$ is horizontal, that is $y_{n+1} = y_n$. In this case the area of $\Delta_n$ equals $y_n/2$. However, $\frac{y_n}{n} > a/2$ and therefore the area of $\Delta_n$ is at least $na/4$.

**Case 2.** The slope of $e_n$ is equal to 1, that is, $y_{n+1} = y_n + 1$. In this case the area of $\Delta_n$ is equal to $(n - y_n)/2$. However, $\frac{y_n}{n} < (1 + a)/2$ and therefore $(n - y_n)/2 \geq n\frac{1-a}{4}$.

Setting $b = \min(a/4, (1-a)/4) > 0$ we conclude that the area of $\Delta_n$ is at least $nb$. Comparing the upper and lower bounds for the area of $\Delta_n$ we find $\alpha(e_n) \geq \frac{b}{n^2}$. It follows that $\sum_{n>N} \alpha(e_n)$ diverges. A simple application of the geometric pigeon hole principle now implies that for every $k$ there exists a line through the origin that cuts through at least $k$ triangles $\Delta_n$ and therefore meets at least $k$ edges $e_n$ of $P$. \]

3 Covering right horns with lines and the proof of Theorem 1.

In this section we prove Theorem 1. We first introduce one more geometric object:

**Definition 2.** A right horn $T$ is a triangle with vertices $A, B, C$ such that $B$ and $C$ have the same $x$-coordinate say $z_2$ and $A$ has smaller $x$-coordinate $z_1$. The length of $T$ is defined to be $z_2 - z_1$. We denote by $\alpha(T)$ the measure of the internal angle at $A$ in triangle $\Delta ABC$. The vertex $A$ is called the apex of $T$ (see Figure 1).
Figure 1: A right horn and a binary path.

**Theorem 4.** For every $\alpha > 0$ and every $M$ there exists $L = L(\alpha, M)$ with the following property. For every binary path $P = \{p_i = (x_i, y_i)\}_{i=0}^L$ of length $L$ there exists an integer $N > M$ and a right horn $T$ of length $N$ and with an apex that is a vertex of $P$ such that $\alpha(T) \leq \alpha$ and $T$ contains at least $0.1N$ of the vertices of $P$.

**Proof.** For a binary path $P = \{(x_i, y_i)\}_{i=0}^L$ we denote by $\theta(P)$ the measure of the acute angle created between the horizontal line through $(x_0, y_0)$ and the line passing through $(x_0, y_0)$ and $(x_K, y_K)$. Notice that for every binary path $P$ we have $0 \leq \theta(P) \leq \frac{\pi}{4}$.

Let $\alpha$ and $M$ be given. Let $P = \{p_i = (x_i, y_i)\}_{i=0}^L$ be a binary path of length $L$ and assume that $L$ is large enough (in terms of $\alpha$ and $M$ in a way that will be specified later). We will define a finite sequence of binary paths $P_0, P_1, P_2, \ldots, P_t$. These paths will have the property that $P_0 = P$ and for every $i \geq 1$ the binary path $P_i$ is a block of $P_{i-1}$ (hence also a block of $P$) and $\theta(P_i) \geq \theta(P_{i-1}) + \frac{\alpha}{4}$. In addition, the length of $P_i$ will be at least one quarter of the length of $P_{i-1}$. At each step when we define $P_i$ we will also define a right horn $T$ with $\alpha(T) = \alpha$. The apex of $T$ is always the first vertex of $P_{i-1}$ and the length of $T$ is the length of $P_{i-1}$. We stop only when $T$ contains enough vertices of $P_{i-1}$.

We will define the $P_i$’s inductively. Assume we have already defined $P_{i-1}$ and that it satisfies the required properties. We define $P_i$ in the following way: Let $P_{i-1} = \{b_0, \ldots, b_\ell\}$. We (re)define $T$ to be the right horn with apex $b_0$ and length $\ell$ such that $\alpha(T) = \alpha$ and the line through $b_0$ and $b_\ell$ bisects the internal angle of $T$ at $b_0$. Let $B$ and $C$ denote the two vertices of $T$ other than $b_0$ and assume without loss of generality that $B$ lies above $C$ (see Figure 2(a)).

If none of the vertices $b_{\ell+1}, \ldots, b_\ell$ lies outside $T$, then we stop and notice that at least half of the vertices of $P_{i-1}$ lie in the horn $T$.

Assume that there is an element $b_k$ of $P_{i-1}$ such that $\frac{\ell}{2} \leq k < \ell$ and $b_k$ lies outside $T$. We assume that $k$ is the minimum such index. If $k \geq \frac{3}{4}\ell$ we stop and notice that at least one quarter of the vertices of $P_{i-1}$, namely $b_{\ell+1}, \ldots, b_{k-1}$, lie in the right horn $T$. Therefore, we assume that $\frac{\ell}{2} \leq k < \frac{3}{4}\ell$.

**Case 1.** $b_k$ lies above the line through $b_1$ and $B$. In this case we define the binary path $P_i$ as
$P_i = \{b_0, \ldots , b_k\}$. Observe that the length of $P_i$ is at least $\frac{k}{2}$ and $\theta(P_i) \geq \theta(P_{i-1}) + \frac{\alpha}{2}$ (see Figure 2(b)).

**Case 2.** $b_k$ lies below the line through $b_1$ and $C$. Recall that that $\frac{k}{2} \leq k < \frac{3}{4} \ell$. In this case we define $P_i = \{b_k, b_{k+1}, \ldots , b_\ell\}$. As $k \leq \frac{3}{4} \ell$, the length of $P_i$ is at least $\frac{1}{4} \ell$ (see Figure 2(c)).

**Claim 1.** $\theta(P_i) \geq \theta(P_{i-1}) + \frac{\alpha}{4}$.

**Proof.** Let $X$ and $Y$ be the intersection points of the vertical line through $b_k$ with the lines $b_0C$ and $b_0b_\ell$, respectively. Let $\beta' = \angle b_kb_\ell b_0$ and let $\beta = \angle Xb_\ell b_0$. We have $\beta < \beta'$. Observe that $\theta(P_i) = \theta(P_{i-1}) + \beta' \geq \theta(P_{i-1}) + \beta$ (see Figure 3). We will now bound from below $\beta$ in terms of $\alpha$ and show that $\beta \geq \frac{\alpha}{4}$. Because it is enough to prove Theorem 4 for small values of $\alpha$ we will assume that $\alpha$ is small enough so that the following is satisfied:

$$\sqrt{2}\sin(\alpha/2) \leq \frac{1}{10},$$

$$\cos(\alpha/2) \geq \frac{9}{10}.$$  

We can also assume that $\beta \leq \frac{\alpha}{4}$, or else the claim is proved. Recall that for every angle $0 < x < \frac{\pi}{2}$ we have $\sin x < x < \frac{\sin x}{\cos x}$.

Considering the triangle $\Delta b_0Xb_\ell$, the law of sines gives:

$$\frac{\sin \beta}{\sin(\alpha/2)} = \frac{|b_0X|}{|Xb_\ell|}.$$
Therefore,
\[
\frac{\beta}{\alpha/2} \geq \frac{\sin \beta \cos(\alpha/2)}{\sin(\alpha/2)} \geq \frac{9}{10} \frac{\sin \beta}{\sin(\alpha/2)} = \frac{9}{10} \frac{|b_0X|}{|Xb_\ell|}.
\] (3)

Because \( k \geq \frac{\ell}{2} \) we have \(|b_0Y| \geq |b_\ell Y|\). Let \( \gamma = \angle b_0YX \) and notice that \( \frac{\pi}{4} \leq \gamma \leq \frac{\pi}{2} \). Observe that
\[
|XY| \sin(\alpha/2) = |b_0X| \sin \gamma \leq |b_0Y| \sin \gamma \leq \sqrt{2}|b_0Y|.
\]

Therefore, \(|XY| \leq \sqrt{2} \sin(\alpha/2)|b_0Y| \leq \frac{1}{10} |b_0Y|\) and hence:
\[
\frac{|b_0X|}{|Xb_\ell|} \geq \frac{|b_0Y| - |XY|}{|b_\ell Y| + |XY|} \geq \frac{|b_0Y| - |XY|}{|b_0Y| + |XY|} \geq \frac{9}{10} \frac{|b_0Y|}{|b_0Y|} = \frac{9}{10} \geq \frac{1}{2}.
\]

It follows now from (3) that \( \frac{\beta}{\alpha/2} \geq \frac{9}{10} \frac{9}{11} > \frac{1}{2} \). \( \blacksquare \)

This completes the induction step of our construction of the sequence \( P = P_0, P_1, P_2, \ldots \). Notice that for every binary path \( Q \) we have \( \theta(Q) \leq \frac{\pi}{4} \). One property of our construction is that \( \theta(P_i) \geq \theta(P_{i-1}) + \frac{\pi}{4} \) for every \( i \geq 1 \). It follows that the sequence \( P_0, P_1, P_2, \ldots, P_t \) may consist of at most \( \frac{\pi}{4}/\frac{\pi}{4} \) elements. Hence when we stop \( t \leq \frac{\pi}{\alpha} \). The length \( N \) of binary path \( P_t \) is at least \( \frac{L}{4\pi} \). Moreover, because we stopped at \( P_t \), there is a horn \( T \) with \( \alpha(T) = \alpha \) whose apex is a vertex in \( P \) (in fact it is the first element of \( P_t \)), and \( T \) contains at least one quarter of the vertices of \( P_t \). The length of \( T \) is equal to the length of \( P_t \).

The binary path \( P_t \) itself is nothing else but a block of size \( N \) of the original binary path \( P = P_0 \). Hence \( P_t = \{p_{N_0}, p_{N_0+1}, \ldots, p_{N_0+N-1}\} \), for some index \( N_0 \). Taking \( L \) to be large enough so that \( \frac{L}{4\pi} \geq M \) we conclude that \( N \geq M \) and this completes the proof of the theorem. \( \blacksquare \)

We will combine the result in Theorem 4 and the next lemma in order to prove Theorem 1.
Lemma 1. Let $T$ be right horn of length $N$ whose apex is the origin and assume it is contained in the portion of the plane $\{(x, y) \mid 0 \leq y \leq x\}$. Then one can find a set of no more than $1 + 4\sqrt{\alpha(T)} N$ (parallel) lines that together cover all integer lattice points that lie in $T$.

Proof. Let $b = \tan \beta$ and $c = \tan \gamma$ be the slopes of the rays from $O$, the origin, to the other vertices of $T$. Set $Q = \lfloor \sqrt{\frac{2}{|b-c|}} \rfloor$. By a standard use of the pigeonhole principle there is some integer $1 \leq q \leq Q$ such that the distance of $q \frac{b+c}{2}$ to the closest integer $p$ is no more then $\frac{1}{Q}$.

Let now $\hat{T}$ be the right horn with apex $O$, length $N$, and such that the rays from $O$ to its two other vertices $\hat{B}, \hat{C}$ have slopes $\min\{b, c, \frac{p}{q} \}$ and $\max\{b, c, \frac{p}{q} \}$. $\hat{T}$ is in fact the minimal extension of $T$ to a right horn of length $N$ and apex $O$ that contains some portion of the line with slope $\frac{p}{q}$ through the origin $O$. Let $S$ be area in the plane bounded between the two parallel lines with slope $\frac{p}{q}$ through $\hat{B}$ and $\hat{C}$. Obviously $T \subseteq \hat{T} \subseteq S$. The set of all lattice points in $S$ are surely covered by the set of lines with slope $\frac{p}{q}$ through all the points in the vertical segment $\hat{B}\hat{C}$ whose $y$-coordinate is an integer multiple of $\frac{1}{q}$. The number of such points is at most $1 + q \left( \max\{b, c, \frac{p}{q}\} - \min\{b, c, \frac{p}{q}\} \right) N$. Now we are done because

$$\max\{b, c, \frac{p}{q}\} - \min\{b, c, \frac{p}{q}\} = \max\{|b - c|, \frac{|b - c|}{2} + \frac{b + c - \frac{p}{q}}{2}\}$$

and hence

$$1 + q \left( \max\{b, c, \frac{p}{q}\} - \min\{b, c, \frac{p}{q}\} \right) N \leq 1 + \max\{q|b - c|, \frac{q|b - c|}{2} + \frac{|b + c - \frac{p}{q}|}{2}\} N$$

$$\leq 1 + \max\{Q|b - c|, \frac{Q|b - c|}{2} + \frac{1}{Q}\} N$$

$$\leq 1 + (\sqrt{2|b - c|} + |b - c|) N$$

$$= 1 + \left(\sqrt{\frac{2\sin \alpha(T)}{\cos \beta \cos \gamma}} + \frac{\sin \alpha(T)}{\cos \beta \cos \gamma}\right) N$$

$$\leq 1 + 4\sqrt{\sin \alpha(T)} N \leq 1 + 4\sqrt{\alpha(T)} N.$$  

We used the fact that $\cos \beta \cos \gamma \geq \frac{1}{2}$ because $T$ is contained in the region $\{(x, y) \mid 0 \leq y \leq x\}$ and consequently both $\beta$ and $\gamma$ are between 0 and $\frac{\pi}{4}$.

We are now ready to prove Theorem 1. Given an integer $k$, let $\alpha = \alpha(k) > 0$ be small enough ($\alpha = \frac{1}{20000k^2}$ is enough here) so that

$$\frac{0.1}{\alpha} > k.$$  

(4)

By Theorem 4, there exists $L = L(k)$ such that for any binary path $P$ of length $L$ there is a right horn $T$ of length $N > 100k$ with apex in $P$ such that $T$ contains at least $0.1 N$ of the points of $P$. Therefore, given any binary path of length $L$, we use Lemma 1 and find $1 + 4\sqrt{\alpha} N$ lines that cover all the integer grid points in $T$ and in particular all the points of $P$ in $T$ whose number is at least $0.1 N$. By the pigeonhole principle there must be one line that contains at least $\frac{0.1 N}{1 + 4\sqrt{\alpha} N} = \frac{0.1}{1 + 4\sqrt{\alpha}}$ points of $P$. It follows from (4) and from $N > 100k$ that

$$\frac{0.1}{1/N + 4\sqrt{\alpha}} > k.$$  

1
References


