TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 213, 1975

# BINARY DIGIT DISTRIBUTION OVER NATURALLY DEFINED SEQUENCES

BY

D. J. NEWMAN AND MORTON SLATER

ABSTRACT. In a previous paper the first author showed that multiples of 3 prefer to have an even number of ones in their binary digit expansion. In this paper it is shown that in some general classes of naturally defined sequences, the probability that a member of a particular sequence has an even number of ones in its binary expansion is ½.

**I. Binary digits contained in arithmetic progressions.** In [4] Newman proved that in counting the number of ones in the binary expansion of a multiple of 3 that count will come out even more often than odd. This is a striking result in that superfically there seems nothing either special or distinguishable about multiples of 3 in the binary system. Certainly a random sequence would be expected to oscillate in "odds" or "evens" being more numerable.

In the case of multiples of 3, the probability is ½ that the multiples have an even number of digits. That is, the preponderance of evens over odds is $o(n)$, in fact about $n^{\log 3/\log 4}$.

A natural question here would be: Are there other natural sequences which exhibit this odd behavior and further is there a natural sequence in which the preponderance of evens over odds or vice versa is a positive proportion of the number of elements in the sequence? If $E(N)$ gives the preponderance of evens over odds and $A(N)$ is the counting function of the sequence does $\lim_{N\to\infty}(E(N)/A(N)) = 0$ ?

In this paper we shall prove that the above limit is zero over various classes of natural sequences.

**Theorem.** *Let $A$ be the sequence $\{M \cdot n\}_{n=1}^{\infty}$. Let $A_M(N)$ be the counting function of this sequence. Then*

$$\lim_{N\to\infty} \frac{E_M(N)}{A_M(N)} = 0 \quad \text{and} \quad |E_M(N) - \tfrac{1}{2}A_M(N)| < CN^{\log 3/\log 4}.$$

---

Received by the editors June 4, 1974.
*AMS (MOS) subject classifications* (1970). Primary 10L10, 10H20, 10H30.
*Key words and phrases.* Binary digits, arithmetic progressions, sieving, abundant numbers.

**Proof.** Let $D(n)$ be the number of ones in the binary expansion for $n$. Let

$$S_M(N) = \sum_{0 \le j \le N/M} (-1)^{D(N-Mj)}.$$

Thus we want to prove that when $N$ is a multiple of $M$, $|S_M(N)| < CN^{\ln 3/\ln 4}$.

As in [4] it follows that, if $N = 2^{K_1} + 2^{K_2} + \cdots + 2^{K_j}$, $K_1 > K_2 > \cdots > K_j \ge 0$, then

$$S_M(N) = S(2^{K_1} - \rho_1) - S(2^{K_2} - \rho_2) + \cdots + (-1)^{j-1} S(2^{K_j} - \rho_j) + (-1)^j$$

where $\rho_i$ is the small positive integer so that $\rho_i + 2^{K_{i+1}} + 2^{K_{i+2}} + \cdots + 2^{K_j}$ is divisible by $M$.

If it can be shown that $S(2^{K_i} - \rho_i) < C2^{K_i(\ln 3/\ln 4)}$ then the result

$$S_M(N) < C_1 \sum_{i=1}^{j} 2^{K_i(\ln 3/\ln 4)} < C_1 \left( \sum_{i=1}^{j} 2^{K_i} \right)^{\ln 3/\ln 4} = C_1 N^{\ln 3/\ln 4}$$

follows.

$$S(2^K - \rho) = \int_{|x|=1/2} \frac{\prod_{n=0}^{k-1} (1 - x^{2^n})}{(1 - x^M) x^{2^k - p + 1}} \, dx.$$

This can be evaluated by the method of residues and is equal to

$$\frac{1}{M} \sum_{\omega_i \text{ an } M\text{th root of unity}} \frac{\prod_{n=0}^{k-1} (1 - \omega_i^{2^n})}{\omega_i^{2^k - \rho + 1}} \le \max_{\omega_i} \prod_{n=0}^{k-1} (1 - \omega_i^{2^n}).$$

In the lemma which follows we shall show that

$$\max_{|z|=1} \prod_{n=0}^{k-1} (1 - z^{2^n}) | < C_1 2^{k \ln 3/\ln 4}.$$

Thus pending proof of the lemma, the theorem is established.

**Lemma.**

$$\max_{|z|=1} \prod_{n=0}^{k-1} (1 - z^{2^n}) | < C_1 2^{k \ln 3/\ln 4}.$$

The proof of this lemma is quite interesting. Although we have no idea of where the maximum occurs or what the value is, we can show that the maximum is quite close to the value at $z = e^{2\pi i/3}$.

**Proof.** If $z = e^{i\theta}$, $|1 - z^{2^n}| = 2|\sin 2^{n-1}\theta|$, so

$$\max_{|z|=1} \left| \prod_{n=0}^{k-1} (1 - z^{2^n}) \right| = \max_{\theta} 2^k \left| \prod_{n=0}^{k-1} \sin 2^{n-1}\theta \right|.$$

Substituting $2\theta$ for $\theta$ to simplify notation we rewrite the above product as follows:

$$\left(|\sin^{1/3}\theta| \prod_{n=0}^{k-2} |\sin^{2/3} 2^n\theta \sin^{1/3} 2^{n+1}\theta|\right) |\sin^{2/3} 2^{k-1}\theta| .$$

Since the max of the product $\leq$ the product of the maximums, each term may be treated separately: $|\sin^{1/3}\theta|\,|\sin^{2/3}2^{k-1}\theta| \leq 1$.

If

$$f(\theta) = \sin^{2/3} 2^n\theta \sin^{1/3} 2^{n+1}\theta,$$

$$f'(\theta) = \frac{2^{n+1}}{3} \sin^{-1/3} 2^n\theta \cos 2^n\theta \sin^{1/3} 2^{n+1}\theta$$

$$+ \frac{2^{n+1}}{3} \sin^{-2/3} 2^{n+1}\theta \cos 2^{n+1}\theta \sin^{2/3} 2^n\theta$$

at a max, we have $\tan 2^{n+1}\theta = -\tan 2^n\theta$. The only solutions of this are $2^n\theta = \pi/3, 2\pi/3, 4\pi/3, 5\pi/3, 0, \pi$. A quick check shows then that $\max_\theta |\sin^{2/3} 2^n\theta \sin^{1/3} 2^{n+1}\theta| = \sqrt{3}/2$. Thus

$$\max_{|z|=1} \left|\prod_{n=0}^{k-1} (1 - z^{2^n})\right| \leq 2^k \left(\frac{\sqrt{3}}{2}\right)^{k-1} = \frac{2}{3}\sqrt{3}\,(2^{\ln 3/\ln 4})$$

and the lemma is established.

II. Suppose that we had a general sequence of integers $a_1, a_2, \cdots, a_n \cdots$ and we wanted to determine the relative frequency that $a_i$ has an even number of ones in its binary expansion. Let $A(n)$ be the counting function for the sequence.

If the following integral is $o(A(n))$ then $\frac{1}{2}$ of the $a_i$'s have an even number of ones. Choose $k$ so that $2^k \leq N < 2^{k+1}$,

$$P(N) = \int_{|z|=1} \frac{\prod_{n=0}^{k} (1 - z^{2^n})}{z} \left(\sum_{a_i<N} z^{-a_i}\right) dz.$$

We will establish that the square free numbers divide evenly.

**Theorem.** *Let $\{a_i\}$ be an arithmetic progression. Then $P(N) < CN^{.8-}$ log $N$.*

**Proof.** $|P(N)| < \int_{|z|=1} |\prod_{n=0}^{k} 1 - z^{2^n}|\,|\sum_{a_i<N} z^{-a_i}|\,d\theta$.

In §I we showed that on the unit circle the max of

$$\left|\prod_{n=0}^{k} (1 - z^{2^n})\right| < CN^{\log 3/\log 4}, \quad \frac{\log 3}{\log 4} < .8- .$$

So $|P(N)| < CN^{.8-} \int |\Sigma z^{-a_i}|\,d\theta$.

Now if $\{a_i\}$ is an arithmetic progression, with difference $a$, containing $n$ terms $< N$,

$$\int \left| \sum_{a_i<N} z^{-a_i} \right| d\theta = \int_{-\Pi}^{\Pi} \left| \frac{1 - z^{na}}{1 - z^a} \right| d\theta.$$

Set

$$\theta = \frac{\theta}{a} = \int_{-\Pi}^{\Pi} \left| \frac{1 - z^n}{1 - z} \right| d\theta \leq C\left( \int_0^k n\, d\theta + \int_k \frac{d\theta}{\theta} \right) \leq C(nk - \log k).$$

Differentiating, we find the expression minimized for $k = 1/n$ so $\int |\Sigma_{a_i<N} z^{-a_i}|\, d\theta < C \log N$ and for $\{a_i\}$ an a.p. $|P(N)| < CN^{\cdot 8-} \log N$. This, of course shows that $P(N) = o(A(N))$.

We might remark that the above argument is quite general. We have $P(N) < N^{\cdot 8-} L^1(A_N)$. For equal frequency we need $P(N) = o(A(N))$ so $L^1(A_N) = o(A(N)/N^{\cdot 8-})$ is sufficient.

Returning to the question of arithmetic progressions we see from the above that the $L^1$ norm of an a.p. is only $\log N$. Thus a set which is a combination of many arithmetic progressions would have to be equidistributed with respect to the digits in the binary expansions. By a combination of a.p.'s, we mean the set of integers formed by using the inclusion-exclusion principle on the collection of a.p.'s. Under these conditions we can combine $N^{\cdot 2}$ progressions and our set will be equidistributed.

If our set $\{C_i\}$, $C_i \neq C_j$, $i \neq j$, is made up of a.p.'s in this way, then

$$\sum z^{C_i} = \sum z^{a_i} + \sum z^{b_i} + \cdots - \sum z^{x_i} - \sum z^{y_i} \cdots,$$

$$\int \left| \sum z^{C_i} \right| \leq \int \left| \sum z^{a_i} \right| + \cdots + \int \left| \sum z^{x_i} \right| + \cdots \leq N^{\cdot 2} \log N.$$

We note in passing that if a set is a complement of one of the above sets and has positive density, then it also has the equidistribution property. The square-free numbers are such a set.

We first form the set of numbers containing a square factor as follows: combine multiples of $2^2$ and $3^2$ then subtract off multiples of $6^2$ etc., or if $S_d^2$ is the set of multiples of $d^2 < N$ we form $-\Sigma_{1<d<\sqrt{N}} \mu(d) S_d^2$.

Now while each integer containing a square factor occurs exactly once, we have too many arithmetic progressions, namely $N^{\cdot 5}$, to use our estimate. We observe however that our set of progressions can be divided into two sets of progressions: one set containing fewer than $N^{\cdot 2}$ progressions to which our theorem applies and the other set of progressions which contains $o(N)$ integers all together.

We note that this approach is valid because instead of using the integral formula for evaluating the variance in distribution, we can simply use as an

upper bound the number of integers in a set. With this in mind we divide the progressions $S_d^2$ as follows: $S_{d'}^2$ $d < k$, and $S_{d'}^2$ $k \le d \le \sqrt{N}$. This gives $k$ progressions plus the remaining integers which total

$$\left(\sum_{k\le d\le\sqrt{N}} \left[\frac{N}{d^2}\right]\right) < N/k + \sqrt{N}.$$

Thus the total estimate for the variance is $N^{.8}k \log N + N/k + \sqrt{N}$ and this is minimized for $k = N^{.1}/\sqrt{\log N}$. Thus we find that the variance in distribution is $< N^{.9}\sqrt{\log N}$. This establishes the equidistribution for square-free numbers provided that numbers containing a square factor have density $< 1$.

In order to establish this, we must evaluate the sum

$$-\sum_{1<d<\sqrt{N}} \mu(d)d^{[N/2]} = -N \sum_{1<d<\sqrt{N}} \frac{\mu(d)}{d^2} + O(\sqrt{N})$$

$$= -N - \frac{N}{\zeta(2)} + N \sum_{d>\sqrt{N}} \frac{\mu(d)}{d^2} + O(\sqrt{N})$$

and since $\zeta(2) = \pi^2/6$ the number of numbers $< N$ containing a square factor is $(\pi^2 - 6)N/\pi^2$ and this is of course a set of density $< 1$.

The set of square-free integers is the complement of the above set. It follows that the set of square-free integers has positive density and also

**Theorem.** *The probability that a square-free integer has an even number of one's in its binary expansion is* ½.

**III. Some general sieved sets.** In the previous section we formed the set of square-free integers by sieving out all multiples of the perfect squares. In this section we pose the following more general question. If we sieve out from the set of all integers those integers which are multiples of some arbitrary set, are the remaining integers equidistributed in the binary system? Of course the answer to this is no if no restriction is made on the sieving set. For example we might sieve with the set of all odd primes. In that case the remaining set would be only powers of 2.

We will prove the following

**Theorem.** *Given a sequence of integers* $a_0, a_1, \cdots$ *such that* $\sum_{i=0}^{\infty} 1/a_i < \infty$ *then the set of all integers that are not multiples of any of the* $a_i$*'s is equidistributed in the binary system.*

We will prove the theorem in two steps. Step 1 will establish that our sieved set has positive lower density. Step 2 will show that the arithmetic progression estimates of §II apply.

**Lemma 1.** *Let $A(n)$ be the counting function of the sieved set.*

$$\underline{\lim} \frac{A(N)}{N} \geq \prod_{k=0}^{\infty} (1 - 1/a_i).$$

We note that $\Pi_{i=0}^{\infty} (1 - 1/a_i) > 0$ since the infinite product converges if and only if $\Sigma_{i=0}^{\infty} 1/a_i$ converges and this was given. Erdös [2] has shown that the density of such a set actually exists.

**Proof of Lemma 1.** The lemma will follow if we can show that the density of a set sieved by the first $M$ of the $a_i$'s is $\geq \Pi_{i=0}^{M-1}(1 - 1/a_i)$; for by the convergence of $\Sigma a_i$, sieving by the remaining $a_i$'s removes at most $N\Sigma_{i>M} 1/a_i < \epsilon N$ integers by proper choice of $M$.

The proof is by induction. If we sieve with only $a_0$ then $D_0 \geq (1 - 1/a_0)$ and we need to show that $D_k \geq (1 - 1/a_k)D_{k-1}$ where $D_k$ is the density after sieving with $k + 1$ integers.

We note that there exists an exact probability or density associated with a number not being divisible by any of $k$ integers namely

$$P = 1 - \sum_{i=0}^{k-1} \frac{1}{a_i} + \sum_{i,j<k-1} \frac{1}{\{a_i, a_j\}} - \cdots (-1)^k \frac{1}{\{a_0 \cdots a_{k-1}\}}$$

where $\{\ \}$ denotes the least common multiple. Thus we can phrase the argument in terms of probability.

We want to establish that for $n < N$,

$$P(n \text{ not divisible by any } a_0, \cdots, a_k) \geq \left(1 - \frac{1}{a_k}\right) P(n \text{ not divisible by any } a_0, \cdots, a_{k-1})$$

or

$$\begin{aligned}
\frac{1}{a_k} P(n \text{ not divisible by any } a_0, \cdots, a_{k-1}) &\geq P(n \text{ not divisible by } a_0, \cdots, a_{k-1}) \\
&\quad - P(n \text{ not divisible by } a_0, \cdots, a_k) \\
&\geq P(n \text{ divisible by } a_k \text{ but not } a_0, \cdots, a_{k-1}) \\
&\geq P(ma_k \text{ not divisible by } a_0, \cdots, a_{k-1}) \\
&= \frac{1}{a_k} P\left(n \text{ not divisible by } \frac{a_0}{(a_0, a_k)}, \cdots, \frac{a_{k-1}}{(a_{k-1}, a_k)}\right)
\end{aligned}$$

where $(a_i, a_j)$ is the g.c.d. The last step follows since $a_i | ma_k$ if and only if $a_i | m(a_i, a_k) \rightarrow a_i/(a_i, a_k) | m$ and also the number of numbers $< N$ is $a_k$ times as much as the number of numbers $< N/a_k$ and the probability that a

number $< N/a_k$ is divisible by a member of a certain set is the same as the probability of a number $< N$ being divisible by a member of the set for $N$ sufficiently large. But our last inequality

$$P(n \text{ not divisible by any } a_0, \cdots, a_k)$$

$$\geq P\left(n \text{ not divisible by } \frac{a_0}{(a_0, a_k)}, \cdots, \frac{a_{k-1}}{(a_{k-1}, a_k)}\right)$$

is true by inclusion. Thus reversing the logic we arrive at the desired lemma.

Incidentally we may rewrite the statement of Lemma 1 as

$$1 - \sum_i \frac{1}{a_i} + \sum_{i,j} \frac{1}{a_k a_j} - \cdots (-1)^k \frac{1}{a_0 a_1 \cdots a_k}$$

$$\geq 1 - \sum_i \frac{1}{a_i} + \sum_{i,j;i<j} \frac{1}{\{a_i, a_j\}} \cdots (-1)^k \frac{1}{\{a_0, a_1, \cdots, a_k\}},$$

an inequality which would seem quite difficult to establish directly.

**Proof of theorem.** We have now shown that our sieved set has positive lower density. Thus if we can show that its complement has the equidistribution property the theorem will follow.

The complement may be formed in two parts as follows. First we sieve with the first $k$-integers in our sieving set. This puts $k$ progressions into our complement. However, some integers have been placed in the complement more than once so we subtract off the multiples of $\{a_i, a_j\}$. Then add back multiples of $\{a_i, a_j, a_k\}$ since we subtracted off some numbers too many times etc. Altogether we have $2^k$ progressions. The remaining sieving adds $N\Sigma_{i>k} 1/a_i$ integers to the complement. Thus the equidistribution estimate is

$$2^k N^{.8} \log N + N \sum_{i>k} \frac{1}{a_i}$$

and we want to show that this is $o(N)$.

If we pick $k = (2/11) \log N$ we need only pick $N$ so large that $\Sigma_{i>k} 1/a_i < \epsilon$ by convergence, and the proof of the theorem is complete.

One application of the above theorem is of course that the square-free integers are equidistributed.

A second application would be that both the abundant numbers and the deficient numbers have the equidistribution property. An abundant number is a number which is exceeded by the sum of its proper divisors. If a number is abundant any multiple is also abundant for if $A$ is abundant and $d|A$, $Bd|BA$ and $BA$ will be abundant. A number is said to be primitive abundant if it is abundant but not a multiple of another abundant number.

Erdös [2] has proven that the sum of the reciprocals of the primitive abundant numbers converges. Thus the primitive abundants can serve in our theorem as the sieving set. Thus the abundants are equidistributed and the sieved set composed of deficient and perfect numbers is also equidistributed. Since the perfect numbers are rare (Euler proved that all even perfect numbers are of the form $2^{2k-1} - 2^{k-1}$ and all odds have the form $PN^2$) we have equidistribution for the deficient numbers also.

A further example would be that the set of all numbers not divisible by any twin prime has the equidistribution property. This follows at once from Brun's result that the sum of the reciprocals of the twin primes converges [3].

## BIBLIOGRAPHY

1. L. Dickson, *History of the theory of numbers*. Vols. 1, 2, 3, Publ. no. 256, Camegie Inst., Washington, D.C., 1919, 1920, 1923; reprint, Stechert, New York.

2. P. Erdös, *On the density of abundant numbers*, J. London Math. Soc. 9 (1934), 278–282.

3. E. Landau, *Elementare Zahlentheorie*, Teubner, Leipzig, 1927; English transl., Chelsea, New York, 1958. MR 19, 1159.

4. D. J. Newman, *On the number of binary digits in a multiple of three*, Proc. Amer. Math. Soc. 21 (1969), 719–721.

DEPARTMENT OF MATHEMATICS, BELFER GRADUATE SCHOOL OF SCIENCE, YESHIVA UNIVERSITY, NEW YORK, NEW YORK 10033

DEPARTMENT OF MATHEMATICS, CITY COLLEGE (CUNY), NEW YORK, NEW YORK 10031