ACTA ARITHMETICA
120.2 (2005)

Random Liouville functions and normal sets

by

Alexander Fish (Jerusalem)

We define a random Liouville function λ_Q which depends on a random set Q of primes and prove that $A_Q = \{n \in \mathbb{N} \mid \lambda_Q(n) = -1\}$ is normal almost everywhere. This fact enables us to generate a family of normal sets such that the equation $xy = z$ is not solvable inside them. Additionally we prove that the equations $xy = z^2$, $x^2 + y^2 = \mathsf{square}$, $x^2 - y^2 = \mathsf{square}$ are solvable in any normal set, and for any equation $xy = cn^2$ ($c > 1$ is not a square) there exists a normal set A_c such that the equation is not solvable inside A_c.

1. Introduction. With the familiar notion of normal numbers in mind, we shall call an infinite binary sequence *normal* if any binary word ω of length $|\omega|$ occurs in the sequence with the right frequency: $2^{-|\omega|}$. We have the natural bijection between infinite $\{0,1\}$-sequences λ and the subsets of the natural numbers $A_\lambda = \{i \mid \lambda_i = 1\}$. We now have

Definition 1.1. A set $B \subset \mathbb{N}$ is called *normal* if the corresponding $\{0,1\}$-sequence is normal.

In this note we shall be interested in normal sets and the possibility of solving diophantine equations in integers from a given, but arbitrary, normal set. We expect that there are many diophantine equations (or systems of equations) which, if they are solvable at all in integers, are solvable in integers from a given normal set. We call such equations *N-regular*, and we denote by DSN the family of N-regular equations (or systems of equations).

An equation, or a system of equations, is called *partition-regular* if for any finite partition of the natural numbers, the system is solvable within one of the cells of the partition. One of the earliest examples of a partition-regular equation is Schur's equation: $x + y = z$. It is not hard to see that Schur's equation is also N-regular. Rado in [6] classified all systems of linear diophantine equations that are partition regular. Rado's theorem implies the

2000 *Mathematics Subject Classification*: 11N64, 05D10.

familiar van der Waerden theorem on existence of arbitrarily long monochromatic arithmetic progressions in any finite coloring of the natural numbers.

Using Furstenberg's theorem regarding Rado's systems in [4], one can obtain the analogous result for N-regularity: namely, any Rado system of linear equations is in DSN.

From the foregoing, we have many linear equations in DSN. But little is known in the non-linear case. For example, it is an open question whether the Pythagorean equation $x^2+y^2 = z^2$ is in DSN. The purpose of this note is to show that the equation $xy = z$ is not in DSN. This equation is called the *multiplicative Schur equation*. It is an easy consequence of Schur's additive theorem that his multiplicative equation is also partition-regular. In fact in any finite partition of $\mathbb{N}$ one can find solutions to both the additive and the multiplicative equations in the same cell ([1]). Thus partition regularity does not imply N-regularity. To show that $xy = z$ is not in DSN we will use a construction of random normal sets, based on a variant of the Liouville function $\lambda(n)$ from number theory. Recall

DEFINITION 1.2. *Liouville's function* $\lambda : \mathbb{N} \to \{-1, 1\}$ is defined by

$$\lambda(p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}) = (-1)^{e_1+e_2+\cdots+e_k}$$

where $p_1, \dots, p_k$ are primes.

It is a well known and very deep question whether the set $A = \{n \in \mathbb{N} \mid \lambda(n) = -1\}$ is normal (see [2] and [3]). It seems that at present we are far from resolving this outstanding problem. But just for clarity, if the answer to this question is positive, then the aforementioned set A gives us an example of a normal set with no solution to the equation $xy = z$.

In the following we will use a random Liouville function λ_Q which is defined by a random choice of a subset Q inside P (the prime numbers) as follows:

$$\lambda_Q(p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}) = \lambda_Q(p_1)^{e_1} \lambda_Q(p_2)^{e_2} \cdots \lambda_Q(p_k)^{e_k}$$

and

$$\lambda_Q(p) = \begin{cases} -1, & p \in Q, \\ 1, & p \notin Q. \end{cases}$$

By randomness of Q we mean that the choice of every prime number p is independent of the choice of any other prime numbers and $\Pr(p \in Q) = 0.5$ for any $p \in P$.

One defines $A_Q = \{n \in \mathbb{N} \mid \lambda_Q(n) = -1\}$. In Section 2 we prove

THEOREM 1.1. *For almost every Q the set A_Q is normal.*

This theorem gives us an infinite family of normal sets such that the multiplicative Schur equation is not solvable in these sets.

In Section 3 we prove that the equations $xy = z^2$, $x^2 + y^2 = \mathsf{square}$ and $u^2 - v^2 = \mathsf{square}$ are in DSN.

Acknowledgments. The author would like to thank Prof. Hillel Furstenberg without whose support this work would not have been done, Prof. Vitaly Bergelson for valuable discussions, Prof. Alex Samorodnitsky for remarkable suggestions, and Michael Fish for his help in proving Theorem 3.2.

2. A_Q is normal for a.e. Q. We start from an obvious claim about normality of A_Q.

LEMMA 2.1. *Let $Q \subset P$ be given. Then A_Q is normal $\Leftrightarrow$ for any $k \in \mathbb{N} \cup \{0\}$ and any $i_1 < \cdots < i_k$ we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \lambda_Q(n)\lambda_Q(n+i_1)\cdots\lambda_Q(n+i_k) = 0. \tag{1}$$

We proceed with the following statement which is readily proved:

LEMMA 2.2. *Let $\{a_n\}$ be a bounded sequence. Define $T_N = N^{-1}\sum_{n=1}^{N} a_n$. Then T_N converges to a limit t $\Leftrightarrow$ there exists an increasing sequence $\{N_i\}$ of indices such that $N_i/N_{i+1} \to 1$ and $T_{N_i} \to t$ as $i \to \infty$.*

The next step is to show

$$\sum_{N=1}^{\infty} E\biggl(\biggl(\frac{1}{N^{40}} \sum_{n=1}^{N^{40}} \lambda_Q(n)\lambda_Q(n+i_1)\cdots\lambda_Q(n+i_k)\biggr)^2\biggr) < \infty.$$

LEMMA 2.3. *Let T_N be as above. Then $E(T_N^2) \le O(1/N^{0.05})$.*

Proof. By linearity of expectation we get

$$E(T_N^2) = \frac{1}{N^2} \sum_{x,y=1}^{N} E(\lambda_Q(x)\lambda_Q(x+i_1)\cdots\lambda_Q(x+i_k)\lambda_Q(y)\lambda_Q(y+i_1)\cdots\lambda_Q(y+i_k)).$$

Note that for any $m \in \mathbb{N}$, $E(\lambda_Q(m)) = 0$ unless m is a square in which case $E(\lambda_Q(m)) = 1$.

Set

$$\phi(x) = \lambda_Q(x)\lambda_Q(x+i_1)\cdots\lambda_Q(x+i_k), \qquad \xi(x) = x(x+i_1)\cdots(x+i_k).$$

By distribution of Q we get

$$E(\phi(x)\phi(y)) = 1 \iff \xi(x)\xi(y) = m^2.$$

Otherwise

$$E(\phi(x)\phi(y)) = 0.$$

Therefore, to obtain an upper bound on $E(T_N^2)$, we give an upper bound on the number of pairs $(x, y) \in [1, N] \times [1, N]$ which satisfy $\xi(x)\xi(y) = \text{square}$.

For a given $x \in [1, N]$ assume that $\xi(x) = c_x m^2$, where c_x is a square-free number, say with prime factorization $c_x = p_{j_1} \cdots p_{j_l}$. Then we define

$h(x) = l$ (thus $h(x)$ is the number of primes in the prime factorization of the maximal square-free number which divides x). Denote by D the set of all possible common divisors of the numbers $x, x+i_1, \ldots, x+i_k$ (i.e. positive integers which divide at least two of them). For a finite non-empty set S of positive numbers we denote by $m(S)$ the product of all elements of S; for the empty set, we set $m(\emptyset) = 1$.

Note that $\xi(x)\xi(y) = \mathsf{square} \Rightarrow$ there exist $S_1 \subset D$ and $S_2 \subset \{p_{j_1}, \ldots, p_{j_l}\}$ such that $y = m(S_1)m(S_2)\,\mathsf{square}$.

Assume $|D| = r$ (r depends only on the set $\{i_1, \ldots, i_k\}$ and does not depend on x). Then we obtain $\xi(x)\xi(y) = \mathsf{square}$ for at most $2^r 2^{h(x)}\sqrt{N}$ y's inside $[1, N]$. Thus

$$E(T_N^2) \le \frac{1}{N^2}\Big(\sum_{n=1}^{N} 2^r 2^{h(n)}\sqrt{N}\Big) \le \frac{c}{N^{1.5}} \sum_{n=1}^{N} 2^{h(n)}.$$

Therefore it remains to bound the expression $\sum_{n=1}^{N} 2^{h(n)}$.

Let $p = p_i$ be the smallest prime number such that $(k+1)/\log_2 p \le 0.45$. If $\xi(n)$ is not divisible by any of the primes $2, 3, \ldots, p$ then

$$h(n) \le \log_p (n+i_k)^{k+1} = (k+1)\,\frac{\log_2 (n+i_k)}{\log_2 p}.$$

This gives us

$$2^{h(n)} \le (n+i_k)^{(k+1)/\log_2 p} \le (n+i_k)^{0.45}.$$

But if $\xi(n)$ is arbitrary then $h(n)$ can increase by at most i, which means $2^{h(n)} \le 2^i (n+i_k)^{0.45}$. Thus $\sum_{n=1}^{N} 2^{h(n)} \le C_1 (N+i_k)^{1.45}$ and therefore we get

$$E(T_N^2) \le C_2\,\frac{1}{N^{0.05}}. \ \blacksquare$$

Proof of Theorem 1.1. From the last lemma we conclude that $\sum_{N=1}^{\infty} E(T_{N^{40}}^2) < \infty$. Thus $T_{N^{40}} \to 0$ almost surely. Lemma 2.2 implies that $T_N \to 0$ almost surely. And from Lemma 2.1 (with countably many conditions for A_Q to be normal) it follows that for almost all $Q \subset P$ the sets A_Q are normal. $\blacksquare$

We can now demonstrate the main result of this note.

THEOREM 2.1. *There exists a normal set $A \subset \mathbb{N}$ such that the multiplicative Schur equation is not solvable inside A.*

Proof. We have already shown the existence of many Q ($Q \subset P$) such that A_Q is normal. By definition of A_Q, we have $xy \notin A_Q$ for any $x, y \in A_Q$. $\blacksquare$

COROLLARY 2.1. *For any equation $xy = cn^k$ (where c, k are natural numbers, c is not a square and k is even) we can find a normal set $A_{c,k} \subset \mathbb{N}$ such that for any $x, y \in A$ we have $xy \ne cn^k$ for every natural n.*

Proof. We take A_Q normal and such that $\lambda_Q(c) = -1$ (this happens with probability $1/2$, and thus such sets exist). Then obviously we cannot solve the above equation inside A_Q. ∎

3. Solvability of the equation $xy = z^2$ and related problems

THEOREM 3.1. *Let $A \subset \mathbb{N}$ be a normal set. Then there exist $x, y, z \in A$ ($x \neq y$) such that $xy = z^2$.*

Proof. For a set $S \subset \mathbb{N}$ and $a \in \mathbb{N}$ define $S_a = \{n \in \mathbb{N} \mid an \in S\}$. It is easily seen that if S is normal then so is each S_a (see [5]). We denote by $d(S)$ the density of a set S, if it exists.

Let A be a normal set. Define $R_n = A_{2^n}$. For any n we have $d(R_n) = 1/2$. Set

$$\mu_N(S) = \frac{|S \cap \{1, \ldots, N\}|}{N}$$

for any $S \subset \mathbb{N}$ and any $N \in \mathbb{N}$.

By Szemerédi's theorem (finite version), for any $\delta > 0$ and $l \in \mathbb{N}$ there exists $N(l, \delta)$ such that for any $N \geq N(l, \delta)$ and $F \subset \{1, \ldots, N\}$ such that $|F|/N \geq \delta$ the set F contains an arithmetic progression of length l (see [7]).

One chooses $K \geq N(3, 1/3)$. Then there exists N_K such that $\mu_{N_K}(R_i) \geq 1/3$ for every $1 \leq i \leq K$.

We claim that there exists $F \subset \{1, \ldots, K\}$ such that $|F|/K \geq 1/3$ and $\mu_{N_K}(\bigcap_{j \in F} R_j) > 0$. If not, let 1_{R_i} be the indicator function of the set R_i inside $\{1, \ldots, N_K\}$. Then on the one hand,

$$\int\limits_{[1, N_K]} (1_{R_1} + \cdots + 1_{R_K})\, d\mu_{N_K} = \sum_{j=1}^{K} \int\limits_{[1, N_K]} 1_{R_j}\, d\mu_{N_K} \geq \frac{K}{3}.$$

But on the other hand,

$$\int\limits_{[1, N_K]} (1_{R_1} + \cdots + 1_{R_K})\, d\mu_{N_K} < \frac{K}{3}$$

because $1_{R_1} + \cdots + 1_{R_K} < K/3$.

Let F be as above. Then by the choice of K it follows that F necessarily contains an arithmetic progression of length 3. This means there exist $a, b, c \in F$ such that $a + c = 2b$. We have $R_a \cap R_b \cap R_c \neq \emptyset$ and so there exists $n \in \mathbb{N}$ such that $x := n2^a \in A$, $z := n2^b \in A$ and $y := n2^c \in A$. Then

$$xy = z^2. \ ∎$$

QUESTION. Are the equations $xy = c^2 z^2$, where $c > 0$ is a natural number, always solvable inside an arbitrary normal set?

THEOREM 3.2. *Let $A \subset \mathbb{N}$ be an arbitrary normal set. Then there exist $x, y, u, v \in A$ such that $x^2 + y^2 = \mathsf{square}$ and $u^2 - v^2 = \mathsf{square}$.*

Proof. Note that there exist $a, b, c \in \mathbb{N}$ such that $a^2 + b^2 = \mathsf{square}$ and $a^2 + c^2 = \mathsf{square}$ and $b^2 + c^2 = \mathsf{square}$, for example $a = 44$, $b = 117$, $c = 240$.

Let $A \subset \mathbb{N}$ be an arbitrary normal set. We look at the triple of sets A_a, A_b, A_c defined as in the proof of Theorem 3.1. Then $d(A_a) = d(A_b) = d(A_c) = 1/2$ and thus it cannot be true that the intersection of each pair from the triple is empty.

Without loss of generality, assume that $A_a \cap A_b \neq \emptyset$. Thus there exists $z \in A_a \cap A_b$ or equivalently $za, zb \in A$. But $a^2 + b^2 = \mathsf{square}$ and therefore $(za)^2 + (zb)^2 = \mathsf{square}$.

The proof that the equation $u^2 - v^2 = \mathsf{square}$ is solvable in any normal set is similar. We use the fact that there exist $a, b, c \in \mathbb{N}$ such that $a < b < c$ and $c^2 - b^2 = \mathsf{square}$, $c^2 - a^2 = \mathsf{square}$ and $b^2 - a^2 = \mathsf{square}$, for example $a = 153, b = 185, c = 697$. ∎

Questions. 1) For an arbitrary normal set A do there exist $x, y, z \in A$ such that $x^2 + y^2 = z^2$?

2) For an arbitrary normal set A do there exist $x, y, z \in A$ such that $x^2 - y^2 = z^2$?

References

[1] V. Bergelson and N. Hindman, *Additive and multiplicative Ramsey theorems in N—some elementary results*, Combin. Probab. Comput. 2 (1993), 221–241.

[2] J. Cassaigne, S. Ferenczi, C. Mauduit, J. Rivat and A. Sárközy, *On finite pseudorandom binary sequences III: The Liouville function, I*, Acta Arith. 87 (1999), 367–390.

[3] —, —, —, —, —, *On finite pseudorandom binary sequences IV: The Liouville function, II*, ibid. 95 (2000), 343–359.

[4] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton Univ. Press, 1981.

[5] —, *Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation*, Math. Systems Theory 1 (1967), 1–49.

[6] R. Rado, *Note on combinatorial analysis*, Proc. London Math. Soc. 48 (1943), 122–160.

[7] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. 27 (1975), 199–245.

Einstein Institute of Mathematics
The Hebrew University of Jerusalem
Jerusalem, 91904, Israel
E-mail: afish@math.huji.ac.il

Received on 28.1.2005
and in revised form on 27.6.2005 (4929)