ON SOME APPLICATIONS OF PROBABILITY METHODS TO ADDITIVE NUMBER THEORETIC PROBLEMS

P. Erdös and A. Rényi

UNIVERSITY OF COLORADO AND MATHEMATICAL INSTITUTE
HUNGARIAN ACADEMY OF SCIENCES

Throughout this paper A and B will denote infinite sequences of integers, B_k denotes a sequence of integers having k terms. $A + B$ denotes the set of integers of the form $a_i + b_j$, $a_i \in A$, $b_j \in B$.

B is called a basis of order r if every sufficiently large integer is the sum of r or fewer b's, B is a basis if it is a basis of order r for some r.

$\overline{A}$ will denote the complementary sequence of A, in other words n is in $\overline{A}$ if and only if it is not in A.

Put $A(x) = \sum_{a_i \leq x} 1$, $A(u,v) = A(u) - A(v)$, $\lim_{x=\infty} \frac{A(x)}{x}$ if it exists is the density of A, $\liminf_{x = \infty} \frac{A(x)}{x}$ is the lower density.

R. Blum asked us the following question: Does there exist for every $0 < \alpha < 1$ a sequence A of density α so that for every B the density of $A + B$ is 1? We shall prove this by probabilistic methods, in fact we prove the following, (in the meantime Blum solved his original problem by different methods).

Theorem 1. To every α, $0 < \alpha < 1$ there is a sequence A of density α so that for every B_k, $k = 1,2,\cdots$ the density of $A + B_k$ is $1 - (1 - \alpha)^k$.

Theorem 1 clearly implies that for every B the density of $A + B$ is 1, thus the answer to Blum's question is affirmative.

Next we show that Theorem 1 is, in a certain sense, best possible. We prove

Theorem 2. Let A be any sequence of density α. Then to every $\epsilon > 0$ and to every k there is a B_k so that the lower density of $A + B_k$ is less than $1 - (1 - \alpha)^k + \epsilon$.

There is a slight gap between Theorems 1 and 2. It seems certain that

Theorem 1 can be slightly strengthened and that the following result holds:

To every α _there is a sequence_ A _of density_ α _so that for every_ B_k _the density of_ $A + B_k$ _is greater than_ $1 - (1 - \alpha)^k$.

We did not carry out the details of the construction of such a sequence A.

We observe that in Theorem 2 lower density cannot be replaced by density or upper density. To see this let $n_1 < n_2 < \cdots$ be a sequence of integers satisfying $n_{k+1}/n_k \to \infty$. For every j, $j = 1,2,\cdots$ and $k = 2^{j-1}(2r + 1)$, $r = 0,1,\cdots$, U is in A if $n_k < U \le n_{k+1}$ and $U \equiv \ell (\mathrm{mod}\ 2j)$, $\ell = 0,\cdots,j-1$. Clearly A has density 1/2, but for every B_2, $A + B_2$ has upper density 1 (to see this let b_1 and $b_1 + j$ be the elements of B_2 then for every $k = 2^{j-1}(2r + 1)$ all but $o(n_{k+1})$ of the integers not exceeding n_{k+1} are in $A + B_2$).

Finally we settle an old question of Stöhr. Stöhr [4] asked if there is a sequence A of density 0 so that for every basis B, A + B has density 1 ? He also asked if the primes have the above property? Erdős [1] proved that the answer to the latter is negative. We shall outline the proof of the following:

Theorem 3. Let $f(n)$ _be an increasing function tending to infinity as slowly as we please. There always is a sequence_ A _of density_ 0 _so that for every_ B _satisfying, for all sufficiently large_ n, $B(n) > f(n)$, $A + B$ _has density_ 1.

It is well known and easy to see that for every basis B of order r we have $B(n) > cn^{1/r}$, thus Theorem 3 affirmatively answers Stöhr's first question.

Before we prove our Theorems we make a few remarks and state some problems. First of all it is obvious that for every A of density 0 there is a B so that A + B also has density 0. On the other hand it is known [5] that there are sequences A of density 0 so that for every B of positive density A + B has density 1. It seems very likely that such a sequence A of density 0 cannot be too lacunary. We conjecture that if A is such that $n_{k+1}/n_k > c > 1$ holds for every k then there is a B of positive density so that the density of A + B is not 1.

We once considered sequences A which have the property P that for every B $A + B$ contains all sufficiently large integers [2]. We observed that then there is a subsequence B_k of B so that $A + B_k$ also contains all sufficiently large integers (k depends on B).

It is easy to see that the necessary and sufficient condition that A does not have property P is that there is an infinite sequence $t_1 < t_2 < \cdots$ so that for infinitely many n and for every $t_i < n$

$$\bar{A}(n - t_i, n) \geq i \ . \tag{1}$$

(1) easily implies that if A has property P then the density of A is 1 (the converse is of course false).

It is not difficult to construct a sequence A which has property P and for which there is an increasing sequence $t_1 < t_2 < \cdots$ so that for every i there are infinitely many values of n for which

$$\bar{A}(n - t_i, n) > i \ . \tag{2}$$

(2) of course does not imply (1). Also we can construct a sequence A having property P so that for every k there is a $B^{(k)}$ so that for every subsequence $B_k^{(k)}$ of $B^{(k)}$ infinitely many integers should not be of the form $A + B_k^{(k)}$.

Now we prove our Theorems. The proof of Theorem 1 will use the method used in [3]; thus it will be sufficient to outline it. Define a measure in the space of all sequences of integers. The measure of the set of sequences which contain n is α and the measure of the set of sequences of n which does not contain n is $1 - \alpha$. It easily follows from the law of large numbers that in this measure almost all sequences have density α. We now show that almost all of them satisfy the requirement of our theorem.

For the sake of simplicity assume $\alpha = 1/2$. Then our measure is simply the Lebesgue measure in (0,1) (we make correspond to the sequence $A = \{a_1 < \cdots\}$ the real number $\sum_{i=1}^{\infty} \frac{1}{2^{a_i}}$). Our theorem is then an immediate consequence of the following theorem (which is just a restatement of the classical theorem of Borel that almost all real numbers are normal). Almost all real numbers $X = \sum_{i=1}^{\infty} \frac{1}{2^{a_i}}$ have the following property: Let $b_1 < \cdots < b_k$ be any k integers. Then the density of integers n for which $n - b_j$ is one of the a's for some $j = 1, \cdots, k$ is $1 - \frac{1}{2^k}$. For $\alpha \neq \frac{1}{2}$ the proof is the same.

Next we prove Theorem 2. Here we give all the details. Let $T = T(k, \epsilon)$ be sufficiently large, we shall show that there is a sequence B_k in $(1,T)$ (i.e. $1 \leq b_1 < \cdots < b_k \leq T$) so that the lower density of $A + B_k$ is less than $1 - \frac{1}{2^k} + \epsilon$.

First we show

$$\sum_{n=T}^{x} \overline{A}(n - T, n) = (1 + o(1))\frac{Tx}{2}. \tag{3}$$

Let $\overline{a}_1 < \overline{a}_2 < \cdots$ be the elements of $\overline{A}$. To prove (3) observe that with a number (at most T) of exceptions, independent of x, every $\overline{a}_i \leq x - T$ occurs in exactly T of the intervals $(n - T, n)$, $T \leq n \leq x$ and each a_i satisfying $x - T < \overline{a}_i \leq x$ occurs in fewer than T of these intervals. Thus the $a_i \leq x - T$ each contribute T to the sum on the left of (3). Hence

$$o(x) + T\overline{A}(x - T) \leq \sum_{n=T}^{x} \overline{A}(n - T, n) \leq T\,\overline{A}(x)$$

which by $\overline{A}(x) = (1 + o(1))\frac{x}{2}$ proves (3).

Let now $T \leq n \leq x$. Clearly we can choose in

$$\binom{\overline{A}(n - t, n)}{k}$$

ways k integers $1 \leq b_1 < \cdots < b_k \leq T$ so that $A + B_k$ should not contain n. Thus by a simple averaging argument there is a choice of a B_k in $(1,T)$ so that there are at least

$$\frac{1}{\binom{T}{k}} \sum_{n=T}^{x} \binom{\overline{A}(n - T, n)}{k} \tag{4}$$

values of $n \leq x$ not in $A + B_k$. Now it follows from (3) that

$$(5) \qquad \sum_{n=T}^{x} \binom{\overline{A}(n - T, n)}{k} \geq (1 + o(1))\, x \binom{[\frac{T}{2}]}{k}$$

since it is well known and easy to see that if $\Sigma\, w_i$ is given then $\Sigma\binom{w_i}{k}$ is a minimum if the w_i's are as equal as possible. Finally observe that for $T > T(k, \epsilon)$

$$(6) \qquad \binom{[\frac{T}{2}]}{k} > \left(1 - \frac{\epsilon}{2}\right)2^{-k}\binom{T}{k} .$$

Thus from (4), (5) and (6) it follows that there is a B_k in $(1,T)$ so that more than $x\left(\frac{1}{2^k} - \frac{\epsilon}{2}\right)$ integers $n \leq x$ are not in $A + B_k$. This B_k may depend on x, but there are at most $\binom{T}{k}$ possible choices of B_k and infinitely many values of x. Thus the same B_k occurs for infinitely many different choices of the integer X.

In other words for this B_k the lower density of $A + B_k$ is less than $1 - \frac{1}{2^k} + \epsilon$ as stated.

It is easy to see that Theorem 2 remains true for all sequences A of lower density α. The only change in the proof is the remark that (3) does not hold for all X but only for the subsequence x_i, $x_i \to \infty$ for which $\lim\limits_{x_i = \infty} A(X_i)/X_i = \alpha$.

Now we outline the proof of Theorem 3. The proof is similar but more complicated than the proof of Theorem 1. We can assume without loss of generality that $f(x) = o(x^{\eta})$ for every $\eta > 0$, but $g(x) = [f(\log x)^{1/2}]$. Define a measure in the space of sequences of integers so that the set of sequences containing n has measure $\frac{1}{g(n)}$ and the measure of the set of sequences not containing n has measure $1 - \frac{1}{g(n)}$. It easily follows from the law of large numbers that for almost all sequences

$$A(x) = (1 + o(1))\frac{x}{g(x)} .$$

We outline the proof that for almost all sequences A, $A + B$ has density 1 for all B satisfying $B(x) > f(x)$ for all sufficiently large x. In fact we prove the following statement:

For every $\epsilon > 0$ there is an $n_o(\epsilon)$ so that for every $n > n_o(\epsilon)$ the measure of the set of sequences A for which there is a sequence B_k, $k > [f(\log n)]$ in $(1, \log n)$ so that the number of integers $m \leq n$ not of the form $A + B_k$ is greater than ϵn, is less than $\frac{1}{n^2}$.

Theorem 3 easily follows from our statement by the Borel-Cantelli lemma.

Thus we only have to prove our statement. Let $1 \leq b_1 < \cdots < b_k < \log n$ be one of our sequences B_k. If m is not in $A + B_k$ then none of the numbers $m - b_i$, $i = 1,\cdots,k$, $k \geq f(\log n)$, are in A. Thus the measure of the set of sequences for which $A + B_k$ does not contain m equals

$$(7) \qquad \prod_{i=1}^{k}\left(1 - \frac{1}{g(m - b_i)}\right) < \left(1 - \frac{1}{g(n)}\right)^k = \left(1 - \frac{1}{\sqrt{k}}\right)^k < \frac{\epsilon}{4} .$$

Let now $m_1,\cdots,m_r$ be any r integers which are pairwise congruent mod $[\log n]$. A simple argument shows that the r events: m_i does not belong to $A + B_k$ are independent. Then by a well known argument it follows from (7) that the measure of the set of sequences A for which these are more than $\frac{\epsilon n}{2}$ integers $m \equiv u(\text{mod } [\log n])$, $m < n$ which are not in $A + B_k$ is less than $(\exp 2 = e^2)$

$$(8) \qquad \exp(-c_\epsilon n/\log n) < \exp(-n^{1/2}) .$$

From (8) and from the fact that there are only log n choices for u it follows that the measure of the set of sequences A so that for a given B_k there should be more than ϵn integers $m \leq n$ not in $A + B_k$ in less than

$$(9) \qquad \log n. \quad \exp(-n^{1/2}).$$

There are clearly fewer than $2^{\log n} < n$ possible choices for B_k, thus by (9) the measure of the set of sequences A for which there is a B_k in $(1,\log n)$ so that there should be more than ϵn integers not in $A + B_k$ is less than

$$n \ \log n \ \exp(-n^{1/2}) < 1/n^2$$

for $n > n_o$, which proves our statement, and also Theorem 3.