ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 1, Number 4, Fall 1971

SOME PROBABILISTIC REMARKS ON FERMAT'S LAST THEOREM

P. ERDÖS AND S. ULAM

Let $a_1 < a_2 < \cdots$ be an infinite sequence of integers satisfying $a_n = (c + o(1))n^\alpha$ for some $\alpha > 1$. One can ask: Is it likely that $a_i + a_j = a_r$ or, more generally, $a_{i_1} + \cdots + a_{i_k} = a_{i_r}$ has infinitely many solutions. We will formulate this problem precisely and show that if $\alpha > 3$ then with probability 1, $a_i + a_j = a_r$ has only finitely many solutions, but for $\alpha \leqq 3$, $a_i + a_j = a_r$ has with probability 1 infinitely many solutions. Several related questions will also be discussed.

Following [1] we define a measure in the space of sequences of integers. Let $\alpha > 1$ be any real number. The measure of the set of sequences containing n has measure $c_1 n^{1/\alpha - 1}$ and the measure of the set of sequences not containing n has measure $1 - c_1 n^{1/\alpha - 1}$. It easily follows from the law of large numbers (see [1]) that for almost all sequences $A = \{a_1 < a_2 < \cdots\}$ ("almost all" of course, means that we neglect a set of sequences which has measure 0 in our measure) we have

$$(1) \qquad A(x) = (1 + o(1))c_1 \sum_{n=1}^{x} \frac{1}{n^{1/\alpha - 1}} = (1 + o(1))c_1 \alpha x^{1/\alpha}$$

where $A(x) = \sum_{a_i < x} 1$. (1) implies that for almost all sequences A

$$(2) \qquad a_n = (1 + o(1))(n/c_1\alpha)^\alpha.$$

Now we prove the following

THEOREM. *Let $\alpha > 3$. Then for almost all* A

$$(3) \qquad a_i + a_j = a_r$$

has only a finite number of solutions. If $\alpha \leqq 3$, then for almost all A, (3) has infinitely many solutions.

It is well known that $x^3 + y^3 = z^3$ has no solutions, thus the sequence $\{n^3\}$ belongs to the exceptional set of measure 0.

Assume $\alpha > 3$. Denote by E_α the expected number of solutions of $a_i + a_j = a_r$. We show that E_α is finite and this will immediately

Received by the editors April 28, 1970.
AMS 1970 *subject classifications*. Primary 10K99, 10L10.

imply that for almost all sequences A, $a_i + a_j = a_r$ has only a finite number of solutions. Denote by $P(u)$ the probability (or measure) that u is in A. We evidently have

$$E_\alpha = \sum_{n=1}^{\infty} P(n) \sum_{u+v=n} P(u)P(v)$$

$$= c_1^3 \sum_{n=1}^{\infty} \frac{1}{n^{1-1/\alpha}} \sum_{u+v=n} \frac{1}{u^{1-1/\alpha}v^{1-1/\alpha}}$$

$$< c_2 \sum_{n=1}^{\infty} \frac{1}{n^{1-1/\alpha}} \frac{1}{n^{1-2/\alpha}} = c_2 \sum_{n=1}^{\infty} \frac{1}{n^{2-3/\alpha}} < c_3$$

which proves our theorem for $\alpha > 3$. One could calculate the probability that (3) has exactly r solutions ($r = 0, 1, \cdots$).

Let now $\alpha \leqq 3$. The case $\alpha = 3$ is the most interesting; the case $\alpha < 3$ can be dealt with similarly. Denote by $E_\alpha(x)$ the expected number of solutions of (3) if a_i, a_j and a_r are $\leqq x$. We have

$$E_3(x) = \sum_{n=1}^{x} P(n) \sum_{u+v=n} P(u)P(v) = c_1^3 \sum_{n=1}^{x} \frac{1}{n^{2/3}} \sum_{u+v=n} \frac{1}{(uv)^{2/3}}$$

$$(4) \qquad = (1 + o(1))c_1^3 \sum_{n=1}^{x} \frac{1}{n^{2/3}} \frac{c_2}{n^{1/3}} = (1 + o(1))c_1^3 c_2 \log x.$$

By a little calculation, it would be easy to determine c_2 explicitly. Now we prove by a simple second moment argument that for almost all A the number of solutions $f_3(A, x)$ of $a_i + a_j = a_r$, $a_r \leqq x$ satisfies

$$(5) \qquad f_3(A, x) = (1 + o(1))c_1^3 c_2 \log x, \quad \text{that is} \quad f_3(A, x)/E_3(x) \to 1.$$

To prove (5) we first compute the expected value of $f_3(A, x)^2$.

The expected value of $f_3(A, x)$ was $E_3(x)$ which we computed in (4). Denote by $E_3^2(x)$ the expected value of $f_3(A, x)^2$. We evidently have

$$(6) \quad E_3^2(x) = \sum_{1 \leqq n_1 \leqq x;\ 1 \leqq n_2 \leqq x} P(n_1)P(n_2) \sum_{u_1+v_1=n_1;\ u_2+v_2=n_2} P(u_1, u_2, v_1, v_2)$$

where $P(u_1, v_1, u_2, v_2)$ is the probability that u_1, v_1, u_2, v_2 occurs in our sequence. If these four numbers are distinct, then clearly $P(u_1, u_2, v_1, v_2) = P(u_1)P(u_2)P(v_1)P(v_2)$, but if say $u_1 = u_2$, the probability is larger. Hence $E_3^2(x) > (E_3(x))^2$ and to get the opposite inequality we have to add a term which takes into account that the four terms do not have to be distinct, or $n_1 < n_2$, $u_1 = u_2$.

$$E_3^2(x) < (E_3(x))^2 + c \sum_{n_1=1}^{x} P(n_1)P(n_1 + v_2 - v_1) \sum_{u_1+v_1=n_1;\, v_2<x} P(u_1)P(v_1)P(v_2)$$

$$(7) \qquad < (E_3(x))^2 + \sum_{n_1=1}^{x} \frac{c_1}{n_1} \sum_{v_2=1}^{x} P(v_2)P(n_1 + v_2 - v_1)$$

$$< (E_3(x))^2 + \sum_{n_1=1}^{x} \frac{c_1}{n_1} \sum_{v_2=1}^{\infty} P(v_2)^2 < (E_3(x))^2 + \sum_{n=1}^{x} \frac{c_2}{n}$$

$$< (E_3(x)^2) + c_3 \log x.$$

Thus

$$(8) \qquad (E_3(x^2)) < E_3^2(x) < (E_3(x))^2 + c_3 \log x.$$

(8) implies by the Tchebycheff inequality that the measure of the set A for which

$$(9) \qquad |f_3(A, x) - E_3(x)| > \epsilon \log x$$

is less than $c/\epsilon^2 \log x$. This easily implies that for almost all A

$$(10) \qquad \lim_{x=\infty} f_3(A, x)/E_3(x) = 1.$$

To show (10) let $x_k = 2^{k(\log k)^2}$. From (9) and the Borel-Cantelli Lemma it follows that

$$(11) \qquad \lim_{k=\infty} f_3(A, x)/E_3(x_k) = 1.$$

(11) now easily implies (10), $f_3(A, x)$ is a nondecreasing function of x, thus if $x_k < x < x_{k+1}$, $f_3(A, x_k) \leqq f_3(A, x) \leqq f_3(A, x_{k+1})$. Thus (11) follows from $E_3(x_n)/E_3(x_{k+1}) \to 1$.

By the same method we can prove that for $\alpha < 3$

$$\lim_{x=\infty} \frac{f_\alpha(A, x)}{E_\alpha(x)} \to 1.$$

Similarly we can investigate the equation

$$(12) \qquad a_{c_1} = a_{c_1} + a_{c_2} + \cdots + a_{c_k}.$$

Here by the same method we can prove that for $\alpha > k + 1$ with probability 1, (12) has only a finite number of solutions and for $\alpha \leqq k + 1$ it has infinitely many solutions.

Euler conjectured that the sum of $k - 1$ (kth) powers is never a kth power. This is true for $k = 3$, unknown for $k = 4$ and has been recently disproved for $k = 5$ [2]. As far as we know it is possible that for every $k \geqq 3$ there are only a finite number of kth powers which are the sum of $k - 1$ or fewer kth powers.

Let $\beta > 1$ be a rational number. One can ask whether $[n^\beta] + [m^\beta] = [l^\beta]$, has solutions in integers n, m, l. One would guess that for $\beta < 3$ the equation always has infinitely many solutions but that the measure of the set in β, $\beta > 3$, for which it has infinitely many solutions has measure 0, but it is not hard to prove that the β's for which it has infinitely many solutions is everywhere dense.

References

1. P. Erdös and A. Rényi, *Additive properties of random sequences of positive integers*, Acta Arith. **6** (1960), 83–110. MR **22** #10970; See also: H. Halberstam and K. F. Roth, *Sequences*, Vol. 1, Clarendon Press, Oxford, 1966. MR **35** #1565.

2. L. J. Lander and T. Parkin, *A counterexample to Euler's sum of powers conjecture*, Math. Comp. **21** (1967), 101–103. MR **36** #3721.

University of Colorado, Boulder, Colorado 80302