A PROBLEM ABOUT PRIME NUMBERS AND THE RANDOM WALK II

BY

P. Erdős

I am going to prove \( \gamma = 1 \). Denote by \( u(a, b) \) the probability of the random walk passing through \( a \) if it starts at \( b \). It is known and easy to prove that

\[
(1) \quad u(a, b) \sim c_1 |b - a|^{-1}
\]

(see K. Itô and H. P. McKean, Jr., Potentials and the random walk, Illinois J. Math., vol. 4 (1960), pp. 119–132; also a paper of Murdoch cited therein where a sharper result is obtained). In the sequel, the letters \( p \) and \( q \) denote primes and \( u(p, q) = u(a, b) \) in case \( a = (p, 0, 0) \) and \( b = (q, 0, 0) \).

Consider the number \( e(n) \) of points \( (p, 0, 0) (p \leq n) \) that the path hits. We have to prove that for almost all paths \( e(n) \uparrow \infty \) as \( n \uparrow \infty \).

By (1) and Mertens' estimate \( \sum_{p \leq n} p^{-1} \sim \lg^2 n \) \( (\lg n = \lg \lg) \), we evidently have

\[
(2) \quad E[e(n)] = \sum_{p \leq n} u(0, p) \sim c_1 \sum_{p \leq n} p^{-1} \sim c_1 \lg^2 n.
\]

Next, we prove by a customary argument

\[
(3) \quad E[(e - c_1 \lg^2 n)^2] = o(\lg^4 n),
\]

which establishes the weak law of large numbers for \( e(n) \), i.e., it shows that \( e(n) = c_1 \lg^2 n + o(\lg^2 n) \) except for a set of small measure, and this is enough for our purpose.

Clearly by (2)

\[
(4) \quad E[(e - c_1 \lg^2 n)^2] = E(e^2) - c_1^2(\lg^2 n)^2 + o(\lg^2 n)^2.
\]

Further we evidently have

\[
(5) \quad E(e^2) = \sum_{p \leq n} u(0, p) + \sum_{q < p \leq n} [u(0, p)u(p, q) + u(0, q)u(q, p)]
\]

\[= 2c_1^2 \sum_{p \leq n} [1/p(p - q) + 1/q(p - q)] + o(\lg^2 n)^2.\]

Mertens' estimate cited above gives \( \sum_{p \leq n} 1/(pq) = \frac{1}{2}(\lg^2 n)^2 + O(\lg n) \), and so

\[
(6) \quad \sum_{q < p \leq n} 1/q(p - q) = \sum_{q < p \leq n} 1/(qp) + \sum_{q < p \leq n} [1/q(p - q) - 1/qp]
\]

\[= \frac{1}{2}(\lg^2 n)^2 + \sum_{q < p \leq n} 1/p(p - q) + O(\lg^2 n).\]

Thus we have only to estimate \( \sum_{q < p \leq n} 1/p(p - q) \).

Received May 19, 1960. See the preceding paper for a statement of the problem. This paper is from a letter sent by P. Erdős to the Illinois Journal of Mathematics. It was edited for publication by H. P. McKean, Jr., who then wrote the preceding paper which is a treatment of the same problem.
Put $\varepsilon_k = 0$ if $k$ is not prime and $\varepsilon_k = \sum_{q < p} 1/(p - q)$. We have

$$\sum_{q < p \leq n} 1/p(p - q) = \sum_{k = 1}^{n} \varepsilon_k / k = \sum_{k = 1}^{n} a_k / k(k + 1) + O(1)$$

by partial summation ($s_k = \sum_{i = 1}^{k} \varepsilon_i$). A well-known theorem of Schnirelmann states that the number of solutions of $p - q = a$ ($p \leq k$) is less than $c_2 k(\log k)^{-2} \prod_{p \mid a} (1 + p^{-1})$ where $c_2$ is an absolute constant. Thus

$$s_k < c_2 k(\log k)^{-2} \sum_{a = 1}^{k} a^{-1} \prod_{p \mid a} (1 + p^{-1}) < e_1 k / \log k$$

since by interchanging the order of summation we have the well-known

$$\sum_{a = 1}^{k} a^{-1} \prod_{p \mid a} (1 + p^{-1}) = \sum_{d = 1}^{k} d^{-1} \sum_{a \equiv 0(\mod d), a \leq k} \nu_a$$

$$< e_1 \sum_{d = 1}^{k} \log k / d^2 < c_3 \log k.$$

Thus from (7) and (8)

$$\sum_{q < p \leq n} 1/p(p - q) < c_8 \log n.$$

From (9), (6), and (5), we finally obtain $E(\varepsilon^n) = c_1^2 (\log n)^2 + o(\log n)^2$ which proves (3), and thus the proof of our theorem is complete.

By using a sharper estimate than (1), it is easy to show that for almost all paths

$$\lim_{n \to \infty} e(n) / c_1 \log n = 1.$$

By the same method one can prove that if the integers $q = q_n < q_{n+1} < \cdots$ satisfy

$$q_n - q_{n-1} > c_7 \log n \quad (n \leq 2), \sum 1/q = \infty,$$

then almost all paths pass through infinitely many points $(q, 0, 0)$. The primes probably do not satisfy (11) since probably there are an infinite number of prime twins, but one can prove by Brun's method that one can select a subsequence that does satisfy (11).

Australian National University
Canberra, Australia