SOME PROBLEMS CONCERNING THE STRUCTURE OF RANDOM WALK PATHS

By

P. ERDŐS (Budapest), corresponding member of the Academy,
and S. J. TAYLOR (Birmingham)

1. Introduction. We restrict our consideration to symmetric random walk, defined in the following way. Consider the lattice formed by the points of d-dimensional Euclidean space whose coordinates are integers, and let a point $S_d(n)$ perform a move randomly on this lattice according to the rules: at time zero it is at the origin and if at any time $n-1$ $(n=1,2,\ldots)$ it is at some point S of the lattice, then at time n it will be at one of the $2d$ lattice points nearest S, the probability of it being at any specified one of these being $\frac{1}{2d}$.

In the present note we examine in some detail the structure of the *path* formed by the points $S_d(0), S_d(1), \ldots, S_d(n), \ldots$. We will sometimes be interested in the first n points of the path, and at others in some property of the infinite path obtained as $n \to \infty$. Our results will depend to a large extent on those obtained in [2]; for convenience we shall use a notation which is consistent with that paper. In Section 2 we summarise the notations used and obtain some preliminary results which will be needed in the sequel.

The paper of DVORETZKY and ERDŐS [2] was only incidentally interested in the returns to the origin of a random walk, that is, the values of the integer n for which $S_d(n)=0$. We study these in detail in Sections 3 and 4. Since PÓLYA showed [8] as long ago as 1921 that a symmetric random walk will, with probability 1, return infinitely often to the origin if $d=1,2$, while if $d>2$, it will wander off to infinity with probability 1, the study of returns to the origin is only interesting for $d=1$ or 2. In the case of plane random walk we obtain the asymptotic distribution of the number of returns to the origin in n steps and use these to deduce strong laws analogous to the law of the iterated logarithm. The corresponding results for the case $d=1$ were previously obtained by CHUNG and HUNT [1]. In Section 4 we examine some properties of the sequence of successive returns to the origin.

In Section 5 we consider two problems related to the behaviour of $\varrho_d(n)$, the distance from the origin of $S_d(n)$. When $d \geqq 3$, the result of PÓLYA shows that $\varrho_d(n) \to \infty$ as $n \to \infty$ and DVORETZKY and ERDŐS obtained lower bounds for the rate at which $r_d(n)$ increases. Our first problem concerns the average rate of increase: this is of interest for any value of d and we obtain different results for the cases $d=1$, $d=2$, and $d \geqq 3$. The second problem concerns a modified form of the law of the iterated logarithm for $d \geqq 3$.

Finally, in Section 6, we consider briefly the multiplicity of points on the path. We are mainly interested in two questions: firstly, how many points of the path are entered a specified finite number of times; and secondly, how large is the maximum multiplicity occurring in a path of n steps.

We hope in a subsequent paper [5] to examine in detail the intersection properties of random walk paths.

2. Notation and preliminary results. For any fixed number of dimensions $d=1,2,\ldots$ we will be considering the space Ω_d of infinite random walks in d-space with a probability measure $\mathbf{P}(E)$ defined for measurable sets in Ω_d by extending the elementary definition of probabilities of single steps. (The measure can be defined by mapping the space of paths onto a q-adic ($q=2d$) representation of the real interval $0 \leqq x \leqq 1$, and using Lebesgue measure. Since measurability problems will not be important, we do not need to go into this.) $\mathbf{P}\{.\}$ will denote the probability that a path w in Ω_d satisfies the condition within the braces. If $E_1, E_2, \ldots, E_k, \ldots$ is a sequence of sets, then we will write

$$\mathbf{P}\{E_k \ i.\ o.\}$$

for the probability that a path w is in infinitely many of the sets E_k.

$c_1, c_2, \ldots, c_{16}$ will denote finite positive real constants. [x] will denote the largest integer not greater than the real number x.

$$l_1(x) = \log_e x, \quad l_k(x) = \log \ldots \log x \qquad (k=1,2,\ldots),$$

where the logarithm is iterated k times.

ε will always denote a positive number.

If X is a vector in d-space, $|X|$ denotes the distance from X to the origin.

For paths in Ω_d, we denote by $\gamma_d(n)$ the probability that in the first $n-1$ steps, the path does not return to the origin. Clearly

$$1 = \gamma_d(1) \geqq \gamma_d(2) \geqq \cdots \geqq \gamma_d(n) \geqq \cdots > 0. \tag{2.1}$$

In [1] it is proved that, for $d \geqq 3$,

$$\gamma_d(n) \to \gamma_d > 0 \tag{2.2}$$

as $n \to \infty$, and

$$\gamma_d < \gamma_d(n) < \gamma_d + O(n^{1-d/2}); \tag{2.3}$$

for $d=2$, $\gamma_d(n)\to 0$ and the estimate found is

$$\gamma_2(n)=\frac{\pi}{\log n}+O\left(\frac{\log\log n}{(\log n)^2}\right). \tag{2.4}$$

Let us first see that (2.4) can be improved slightly to

$$\gamma_2(n)=\frac{\pi}{\log n}+O\left(\frac{1}{(\log n)^2}\right). \tag{2.5}$$

Write $u_2(r)$ for the probability that $S_2(r)=0$. Then for odd integers r, $u_2(r)=0$, while

$$u_2(2r)=\frac{1}{\pi r}+O\left(\frac{1}{r^2}\right) \quad \text{as} \quad r\to\infty. \tag{2.6}$$

Counting the last return to the origin, we have

$$\sum_{k=0}^{[n/2]}\gamma_2(n-2k)u_2(2k)=1. \tag{2.7}$$

By (2.1) this gives $\gamma_2(n)\sum_{k=0}^{[n/2]}u_2(2k)\leqq 1$, which with (2.6) gives

$$\gamma_2(n)\leqq\frac{\pi}{\log c_1 n}+O\left(\frac{1}{(\log n)^2}\right). \tag{2.8}$$

Now if $1<k_1<k_2<[n/2]$, (2.1) and (2.7) give

$$\gamma_2(n-2k_1)\sum_{k=0}^{k_1}u_2(2k)+\gamma_2(n-2k_2)\sum_{k=k_1+1}^{k_2}u_2(2k)+\sum_{k=k_2+1}^{[n/2]}u_2(2k)\geqq 1.$$

Now take $k_1=\left[\frac{n}{4}\right]$, $k_2=\left[\frac{n}{2}-\frac{n}{\log n}\right]$ and apply (2.6) and (2.8) to obtain

$$\gamma_2([n/2])\geqq\frac{\pi}{\log n}-O\left(\frac{1}{\log n}\right)^2.$$

Replacing n by $2n$ gives

$$\gamma_2(n)\geqq\frac{\pi}{\log n}-O\left(\frac{1}{\log n}\right)^2$$

which, together with (2.8) completes the proof of (2.5).

Now suppose that P is a lattice point in the plane whose distance from the origin is ϱ. Let $u_2(P,n)$ be the probability that $S_2(n)=P$. According to the position of P, it can only be reached in either an even number of steps or an odd number of steps. If it can be reached in an even number of steps and

(i) $k > \varrho^2$, then

$$u_2(P, 2k) = \frac{1}{\pi k} + \varrho^2 \cdot O\left(\frac{1}{k^2}\right); \tag{2.9}$$

(ii) while if $k < \varrho^2$, then

$$u_2(P, 2k) \leqq \left[\frac{1}{\pi k} + O\left(\frac{1}{k^2}\right)\right] e^{-\frac{\varrho^2}{2k}}. \tag{2.10}$$

The formulae (2.9) and (2.10) can be obtained by counting the number of paths (out of 4^{2k}) which end at P and using Stirling's formula, or by using the central limit theorem.

Now let $\gamma_2(P, n)$ be the probability that in the first n steps the path does not pass through P. Again assuming that P can be reached in an even number of steps we have

$$\gamma_2(P, n) + \sum_{k=1}^{[n/2]} u_2(P, 2k)\gamma_2(n-2k) = 1 \tag{2.11}$$

on considering the last return to P.

Subtracting (2.7) from (2.10) gives

$$\gamma_2(P, n) - \gamma_2(n) = \sum_{k=1}^{[n/2]} \{u_2(2k) - u_2(P, 2k)\}\gamma_2(n-2k). \tag{2.12}$$

Now suppose that

$$20 < \varrho < n^{1/3}. \tag{2.13}$$

We have, for $1 < k_1 < k_2 < [n/2]$,

$$\gamma_2(P, n) - \gamma_2(n) \leqq \gamma_2(n-k_1)\sum_{k=1}^{k_1} u_2(2k) + \gamma_2(n-k_2)\sum_{k=k_1+1}^{k_2}\{u_2(2k) - u_2(P, 2k)\} +$$
$$+ \sum_{k=k_2+1}^{[n/2]}\{u_2(2k) - u_2(P, 2k)\}.$$

Put $k_1 = \varrho^2$, $k_2 = n^{4/5}$ and use (2.1), (2.5), (2.6) and (2.10) to give

$$\gamma_2(P, n) - \gamma_2(n) \leqq \left[\frac{\pi}{\log n} + O\left(\frac{1}{\log^2 n}\right)\right]\frac{\{\log \varrho^2 + O(1)\}}{\pi} + O(e^{-n^{1/10}})$$

or

$$\gamma_2(P, n) - \gamma_2(n) \leqq \frac{\log \varrho^2 + O(1)}{\log n}. \tag{2.14}$$

Similarly, for $1 < k_3 < [n/2]$,

$$\gamma_2(P, n) - \gamma_2(n) \geqq \gamma_2(n)\sum_{k=1}^{k_3}\{u_2(2k) - u_2(P, 2k)\}.$$

On taking $k_3 = \varrho^2/\log\log\varrho$ this gives

$$(2.15)\qquad \gamma_2(P,n)-\gamma_2(n) \geqq \frac{\log\varrho^2}{\log n}\left[1-O\left(\frac{l_3(\varrho)}{l_1(\varrho)}\right)\right].$$

The results (2.5), (2.14) and (2.15) together show that under the conditions of (2.13)

$$(2.16)\qquad \gamma_2(P,n)=\frac{2\log\varrho}{\log n}\left[1+O\left(\frac{l_3(\varrho)}{l_1(\varrho)}\right)\right].$$

It is trivial to show that if $\varrho \leqq 20$,

$$(2.17)\qquad \gamma_2(P,n)=O\left(\frac{1}{\log n}\right).$$

Each of the results (2.16) and (2.17) can also be proved for points P which can only be reached in an odd number of steps from the origin: only obvious modifications to the proof are needed.

A calculation similar to the one we have carried out will show that if

$$\varrho = n^{1/2}/\psi \quad\text{and}\quad 20<\psi<n^{1/3},$$

then

$$(2.18)\qquad \gamma_2(P,n)=1-\frac{2\log\psi}{\log n}\left[1+O\left(\frac{l_2(\psi)}{l_1(\psi)}\right)\right].$$

We omit the proofs of (2.17) and (2.18) as these will not be needed in the sequel.

3. The number of returns to the origin. In the case $d \geqq 3$, the situation is very simple. Let γ_d be the probability that the random walk never returns to the origin. By (2.2), $0<\gamma_d<1$, and if R is the total number of returns to the origin for an infinite path in d-space, the random variable R must have the geometric distribution

$$(3.1)\qquad \mathbf{P}\{R=k\}=\gamma_d(1-\gamma_d)^k \qquad (k=0,1,2,\ldots).$$

Let us now consider in detail the case $d=2$ of a random walk in the plane. Let R_n be the number of returns to the origin in the first n steps. Let W_r denote the suffix of the r^{th} return to the origin. That is, there are $r-1$ returns to O among $S_2(1), S_2(2), \ldots, S_2(n-1)$ and $S_2(n)=0$ where $n=W_r$. It is clear that

$$(3.2)\qquad \gamma_2(n)=\mathbf{P}\{W_1\geqq n\}=\mathbf{P}\{R_{n-1}=0\}.$$

We shall see that R_n has the order of magnitude $\log n$, so let us define a new random variable T_n by

$$(3.3)\qquad T_n=\frac{R_n}{\log n} \qquad (n=3,4,\ldots).$$

Let $x>0$ be any real number; our aim now is to try to estimate $\mathbf{P}\{T_n<x\}$. Define an integer q by

$$q=[x\log n]+1. \tag{3.4}$$

Then if $W_q<n$, we certainly have $T_n\geqq x$. That is, we have

$$\mathbf{P}\{T_n\geqq x\}\geqq\mathbf{P}\{W_q<n\}\geqq\prod_{s=1}^{q}\mathbf{P}\left\{W_s-W_{s-1}<\frac{n}{q}\right\}=\left[\mathbf{P}\left\{W_1<\frac{n}{q}\right\}\right]^q,$$

since the variables W_s-W_{s-1} $(s=1,2,\ldots,q)$ are independent and all have the same distributions as W_1. If $p=\left[\frac{n}{q}\right]$, it follows from (3. 2) that

$$\mathbf{P}\{T_n\geqq x\}\geqq[1-\gamma_2(p)]^q.$$

Thus, provided $x<\frac{\log n}{(l_2(n))^{1+\varepsilon}}$, we have, on substituting the estimate (2. 5) for $\gamma_2(p)$,

$$\mathbf{P}\{T_n\geqq x\}\geqq e^{-\pi x}(1+o(1))\quad\text{as}\quad n\to\infty.$$

Note further that we have

$$\mathbf{P}\{T_n\geqq x\}\geqq e^{-\pi x}(1+o((\log n)^{-1/5})) \tag{3.5}$$

uniformly in x for $x<(\log n)^{3/4}$.

We will later also need an estimate for $\mathbf{P}\{T_n\geqq x\}$ in the case $x=k\log n$ where k is a constant. The method used above is not completely accurate in this case, but it is still sufficient to give

$$\mathbf{P}\{T_n\geqq k\log n\}\geqq\frac{e^{-k\pi\log n}}{(\log n)^{2k\pi(1+\varepsilon)}} \tag{3.6}$$

for any positive number ε.

Let us now try to obtain an upper bound corresponding to the lower bound (3. 5). Let E_k $(k=1,2,\ldots,q)$ be the event that precisely k of the variables W_s-W_{s-1} are greater than or equal to n, while $q-k$ of them are less than n. Then clearly, since the events E_k are mutually exclusive,

$$\mathbf{P}\{T_n<x\}\geqq\sum_{k=1}^{q}\mathbf{P}(E_k). \tag{3.7}$$

By (3. 2), we have that

$$\sum_{k=1}^{q}\mathbf{P}(E_k)=\sum_{k=1}^{q}\binom{q}{k}\{\gamma_2(n)\}^k(1-\gamma_2(n))^{q-k}=1-(1-\gamma_2(n))^q.$$

Now, using the estimate (2. 5), it follows that, for $x<(\log n)^{3/4}$,

$$\mathbf{P}\{T_n<x\}\geqq 1-e^{-\pi x}(1+O(\log n)^{-1/4}) \tag{3.8}$$

uniformly in x for that range.

Thus (3. 5) and (3. 8) together show that

$$\mathbf{P}\{T_n \geqq x\} = e^{-\pi x}[1 + o(\log n)^{-1/5}] \quad \text{as} \quad n \to \infty, \tag{3.9}$$

uniformly for $x < (\log n)^{3/4}$.

If x is restricted to a fixed range, say

$$c_2 < x < c_3,$$

then a better estimate can be obtained. We have, instead of (3. 8),

$$\mathbf{P}\{T_n < x\} \geqq \{1 - e^{-\pi x}\} \left\{1 + O\left(\frac{1}{\log n}\right)\right\}.$$

In this case the estimate (3. 5) can be improved to

$$\mathbf{P}\{T_n \geqq x\} \geqq e^{-\pi x} \left\{1 + O\left(\frac{\log \log n}{\log n}\right)\right\}.$$

Thus we have

$$\mathbf{P}\{T_n \geqq x\} = e^{-\pi x} \left\{1 + O\left(\frac{\log \log n}{\log n}\right)\right\} \tag{3.10}$$

uniformly for $c_2 < x < c_3$.

We need (3. 10) to allow us to obtain a satisfactory upper bound corresponding to (3. 6).

Put $s = [k(\log n)^2]$, $t = [k \log n]$. Then

$$\mathbf{P}\{T_n \geqq k \log n\} \leqq \mathbf{P}\{W_s \leqq n\} \leqq \mathbf{P}\left\{\bigcap_{r=1}^{[\log n]} \{W_{rt} - W_{(r-1)t} \leqq n\}\right\} = [\mathbf{P}\{W_t \leqq n\}]^{[\log n]},$$

since $W_{rt} - W_{(r-1)t}$ are independent and all have the same distribution as W_t.

Now $\mathbf{P}\{W_t \leqq n\} = \mathbf{P}\{R_n < t\}$. Using the estimate (3. 10), this shows that there exists a constant c_4, depending on k such that

$$\mathbf{P}\{T_n \geqq k \log n\} \leqq e^{-\pi k \log n}(\log n)^{c_4} \tag{3.11}$$

for large enough values of n.

The detailed results obtained in this section will be needed in the sequel. Let us summarise the picture in the following

THEOREM 1. *If R_n denotes the number of returns to the origin in the first n steps of a plane random walk, then*

$$\lim_{n \to \infty} \mathbf{P}\{R_n < x \log n\} = 1 - e^{-\pi x}$$

for $x < (\log n)^{3/4}$, and the limit is approached uniformly in this range.

Thus the asymptotic value of the mean of the random variable $T_n = \frac{R_n}{\log n}$ is $\frac{1}{\pi}$. This ties up with one of the main results of [2] where it is shown that the number of lattice points entered in n steps is $\frac{\pi n}{\log n}(1+o(1))$ with probability 1. Thus the average multiplicity of points entered must be $\frac{1}{\pi}\log n$.

REMARK. CHUNG and HUNT [1] found the result corresponding to Theorem 1 for random walk on the line. They showed that if N_n denotes the number of returns to the origin in the first n steps, then $\frac{N_n}{n^{1/2}}$ has a distribution which tends to that of $|Y|$ where Y is a normally distributed variable with mean 0 and variance 1.

We now go on to consider laws of the type of the iterated logarithm for the random variable T_n. Since the methods required are complicated and not essentially new, we will not always give complete proofs. First let us consider the small values of T_n.

THEOREM 2. *If $\varphi(x)$ decreases to zero, $\varphi(x)\log x$ increases to $+\infty$ as $x\to+\infty$, and R_n is the number of returns to the origin in the first n steps of a plane random walk, then*

$$\mathbf{P}\{R_n<\varphi(n)\log n \ \ i.\ o.\}=0 \quad or \quad 1,$$

according as $\int^{\infty}\frac{\varphi(x)}{x\log x}dx$ *converges or diverges.*

PROOF. The integral converges or diverges with the series $\sum_1^{\infty}\varphi_k$ where $\varphi_k=\varphi(n_k)$ and $n_k=2^{2^k}$ $(k=1,2,\ldots)$.

Suppose first that $\Sigma\varphi_k$ converges. We may assume that $\varphi(x)\geqq(\log x)^{-1/10}$, since otherwise we can replace φ by

$$\psi(x)=\max\{\varphi(x),(\log x)^{-1/10}\}$$

without upsetting the convergence of the integral. Hence by Theorem 1, for k large we have

$$\mathbf{P}\{R_{n_k}<2\varphi_k\log n_k\}<c_1\varphi_k.$$

By the Borel—Cantelli lemma, there exists with probability 1 an integer k_0 such that $R_{n_k}\geqq 2\varphi_k\log n_k$ for $k\geqq k_0$. Now if $n_{k+1}>n\geqq n_k$,

$$\frac{R_n}{\log n}\geqq\frac{1}{2}\frac{R_{n_k}}{\log n_k}\geqq\varphi_k.$$

Hence

$$\frac{R_n}{\log n}\geqq\varphi(n)\quad\text{for}\quad n\geqq n_{k_0}.$$

Now suppose that $\Sigma\varphi_k$ diverges. We need to be a little more careful as the events considered above are not independent. The conditions of Theorem 1 are satisfied so that

$$\mathbf{P}\left(W_{2^r}-W_{2^{r-1}}>e^{2^r/\varphi_r}\right)=\mathbf{P}\left(W_{2^{r-1}}>e^{2^r/\varphi_r}\right)\geqq\mathbf{P}(R_{t_r}<2^{r-1}) \tag{3.12}$$

where $t_r=[e^{2^r/\varphi_r}]$. Now

$$\mathbf{P}(R_{t_r}<2^{r-1})\geqq\mathbf{P}\left\{\frac{R_{t_r}}{\log t_r}<\frac{\varphi_r}{2}\right\}>c_3\varphi_r$$

for $r\geqq r_0$, by Theorem 1, since $t_r\to\infty$. The events $W_{2^r}-W_{2^{r-1}}>e^{2^r/\varphi_r}$ $(r=1,2,\ldots)$ are independent. Hence again appealing to the Borel—Cantelli lemma, there are infinitely many integers r for which

$$W_{2^r}>e^{2^r/\varphi_r}$$

and so

$$\frac{R_{t_r}}{\log t_r}<\varphi_r\leqq\varphi(t_r),$$

for infinitely many integers t_r. This completes the proof of the theorem.

Corollary. *There is probability* 1 *that, for any constant k, $R_n<\dfrac{\log n}{k\log\log n}$ infinitely often; but for any $\delta>0$, there is probability* 1 *that $R_n>\dfrac{\log n}{(\log\log n)^{1+\delta}}$ for all large enough n.*

Now let us go on to consider the unusually large values of R_n. For this purpose we shall find it useful to look at the sequence at the points

$$m_k=[e^{k/\log k}]\qquad(k=2,3,\ldots). \tag{3.13}$$

Suppose $\psi(x)$ is a monotonic function of x which increases to $+\infty$ as $x\to+\infty$.

Write

$$\psi_k=\psi(m_k)\qquad(k=2,3,\ldots), \tag{3.14}$$

$$\mathbf{P}\left\{\frac{\pi R_{m_k}}{\log m_k}>\psi_k\right\}=p_k.$$

Theorem 3. *If $\psi(x)$ is a monotonic function which increases to $+\infty$ as $x\to+\infty$ and $\sum_{k=2}^{\infty}e^{-\psi_k}$ converges, where ψ_k is given by* (3.13), (3.14), *then*

$$\mathbf{P}\left\{\frac{\pi R_n}{\log n}>\psi(n)\ i.\,o.\right\}=0.$$

PROOF. We may assume that $\psi(n) \leqq 2 \log_3 n$, since otherwise replacing $\psi(n)$ by $\psi_1(n) = \min\{\psi(n), 2\log_3 n\}$ will not effect the convergence of $\Sigma e^{-\psi_k}$ and will only strengthen the result of the theorem.

Then by (3. 13), (3. 14) and (3. 9), a simple computation shows that, for large enough k,

$$\mathbf{P}\left\{\frac{\pi R_{m_{k+1}}}{\log m_k} > \psi_k\right\} < e^3 \cdot e^{-\psi_k}.$$

Applying the Borel—Cantelli lemma we have, since $\Sigma e^{-\psi_k}$ converges, that there exists with probability 1, an integer k_0 such that

$$\frac{\pi R_{m_{k+1}}}{\log m_k} \leqq \psi_k \quad \text{for} \quad k \geqq k_0.$$

Then if $m_{k+1} > n \geqq m_k$ and $k \geqq k_0$,

$$\frac{\pi R_n}{\log n} \leqq \frac{\pi R_{m_{k+1}}}{\log m_k} \leqq \psi_k \leqq \psi(n);$$

and therefore there is probability 1 that

$$\frac{\pi R_n}{\log n} > \psi(n) \text{ only finitely often.}$$

To prove the converse of Theorem 3 requires a great deal more trouble due to the independence difficulties in the application of the Borel—Cantelli lemma. We state two forms of the theorem which are almost equivalent.

THEOREM 4A. *If $\psi(x)$ is a monotonic function satisfying $\psi(x) > c_5 l_3(x)$ for some $c_5 > 0$, then*

$$\mathbf{P}\left\{\frac{\pi R_n}{\log n} > \psi(n) \ i.\,o.\right\} = 0 \quad or \quad 1,$$

according as $\sum_{k=2}^{\infty} e^{-\psi_k}$ converges or diverges, where ψ_k is given by (3. 13), (3. 14).

THEOREM 4B. *If $\psi(x)$ is a monotonic function increasing to $+\infty$ such that $\psi(x)/\log x$ decreases to zero as $x \to +\infty$, then*

$$\mathbf{P}\left\{\frac{\pi R_n}{\log n} > \psi(n) \ i.\,o.\right\} = 0 \quad or \quad 1,$$

according as $\int^{\infty} \frac{\psi(x)}{x \log x} e^{-\psi(x)}$ converges or diverges.

COROLLARY. *If $r > 4$ is a positive integer, and*

$$\psi(x) = l_3(x) + 2l_4(x) + l_5(x) + \cdots + l_r(x) + t l_{r+1}(x),$$

then

$$\mathbf{P}\left\{\frac{\pi R_n}{\log n} > \psi(n) \ i.\,o.\right\} = 0 \quad or \quad 1,$$

according as $t > 1$ or $t \leq 1$.

It is clear that for functions $\psi(x)$ which satisfy the conditions of both theorems, the two Theorems 4A, 4B are equivalent. The corollary can be deduced from either. It seems likely that the condition $\psi(x) > c_6 l_3(x)$ of Theorem 4A could be relaxed, but some sort of lower bound to the rate of growth of $\psi(x)$ is necessary. A proof of Theorem 4A can be obtained by making suitable modifications to the proof of ERDŐS [4]; and a proof of Theorem 4B can be obtained by modifying the proof of CHUNG and HUNT [1]. By either method the details are extremely formidable, and we do not propose to write them down as there is only one idea needed which could be described as new. This idea will be illustrated if we give a proof for the first term only of the asymptotic expansion of the critical $\psi(x)$. This is given by

THEOREM 4C. *If $\psi(x) = c \log\log\log x$, and R_n is the number of returns to the origin in the first n steps of a plane random walk, then*

$$\mathbf{P}\left\{\frac{\pi R_n}{\log n} > \psi(n) \ i.\,o.\right\} = 0 \quad or \quad 1,$$

according as $c > 1$ or $c \leq 1$.

PROOF. With $c > 1$, and ψ_k defined by (3.13), (3.14) the series $\Sigma e^{-\psi_k}$ converges. By Theorem 3 it follows immediately that

$$\mathbf{P}\left\{\frac{\pi R_n}{\log n} > \psi(n) \ i.\,o.\right\} = 0.$$

For the case $c \leq 1$, it is sufficient to prove that

$$\mathbf{P}\left\{\frac{\pi R_n}{\log n} > l_3(n) \ i.\,o.\right\} = 1. \tag{3.15}$$

Put

$$s_k = [e^{k \log k}] \qquad (k = 2, 3, \ldots). \tag{3.16}$$

Let E_k be the event

$$E_k = \left\{\frac{\pi R_{s_k}}{\log s_k} > l_3(s_k)\right\}.$$

Then, by (3. 9), we have

$$\mathbf{P}(E_k)=\frac{1}{k\log k}(1+o(1)),$$

so certainly $\Sigma\mathbf{P}(E_k)$ diverges. The points s_k are sufficiently far apart to use the simplest form of argument for overcoming independence difficulties. We can show the existence of c_6 such that

$$\mathbf{P}(E_k|E_2'E_3'\ldots E_{k-1}')>c_6\mathbf{P}(E_k). \tag{3. 17}$$

At the end of s_{k-1} steps the random walk path is certainly at a distance $<s_{k-1}$ from O. If T_k is the event that there is at least one return to the origin[1] between s_{k-1} and $s_{k-1}^{\log k}$, it follows form (2. 16) that

$$\mathbf{P}(T_k|E_2'E_3'\ldots E_{k-1}')>1-\frac{2}{\log k} \tag{3. 18}$$

for large k. If T_k occurs, the path can be started from the first return to the origin after s_{k-1} and $t_k=(s_k-s_{k-1}^{\log k})$ steps taken. Hence

$$\mathbf{P}(E_k|T_k)\geqq\mathbf{P}(Q_k) \tag{3. 19}$$

where Q_k is the event that in t_k steps starting from O the number of returns is not less than $\frac{1}{\pi}l_3(s_k)\log s_k$.

Now it is clear that

$$t_k>s_k-s_k^{1/2},$$

so that $\log t_k>\log s_k(1-s_k^{-1/2})$.

It follows from (3. 9) and (3. 16) that

$$\mathbf{P}(Q_k)=\frac{1}{k\log k}(1+o(1)).$$

Combining this with (3. 18) and (3. 19) is sufficient to prove (3. 17). This shows that

$$\mathbf{P}(E_k|E_2'E_3'\ldots E_{k-1}')\text{ diverges},$$

so that, with probability 1, the event E_k occurs infinitely often. This completes the proof of the theorem.

4. The distribution of the returns to the origin. We have seen that $\frac{R_n}{n^{1/2}}, \frac{R_n}{\log n}$ in the cases $d=1,2$, respectively, each have an asymptotic distribution as $n\to\infty$. As a result these ratios do not approach a limit as

[1] This is where we introduce an idea not needed in [1] or [4].

$n \to \infty$. We first show that a suitable averaging process leads to a limit, and then show that if one only counts returns at a suitably sparse subsequence, then the number of returns has an asymptotic value.

THEOREM 5. *Suppose W_s denotes the suffix of the s^{th} return to the origin. Then there are constants c_7, c_8 such that*

(i) *if the random walk is on the line, then*

$$\lim_{n\to\infty} \left\{ \frac{1}{\log n} \sum_{s=1}^{n} \frac{1}{W_s^{1/2}} \right\} = c_7 \quad \textit{with probability } 1;$$

(ii) *if the random walk is in the plane, then*

$$\lim_{n\to\infty} \left\{ \frac{1}{\log n} \sum_{s=1}^{n} \frac{1}{\log W_s} \right\} = c_8 \quad \textit{with probability } 1.$$

PROOF OF (ii). For a random walk in the plane define a sequence of random variables by

$$\mu(k) = \begin{Bmatrix} \dfrac{1}{\log k} & \text{if} \quad S_2(k) = 0, \\ 0 & \text{otherwise} \end{Bmatrix} \qquad (k = 2, 3, \ldots).$$

Using (2. 6) we see that

$$\mathcal{E}\{\mu(2k)\} = \frac{1}{\pi k \log k} + O\left(\frac{1}{k^2}\right),$$

$$\mathcal{E}\{\mu(2k+1)\} = 0 \qquad (k = 1, 2, \ldots).$$

Hence, if we put $\nu(n) = \sum_{k=2}^{n} \mu(k)$, we obtain

$$\mathcal{E}\{\nu(n)\} = \frac{1}{\pi} \log\log n + O(1). \tag{4.1}$$

The variance $\sigma^2\{\nu(n)\}$ may be estimated since

$$\sigma^2 \leqq 2 \sum_{2 \leqq i \leqq j \leqq n} \frac{1}{\log i} \frac{1}{\log j} (\mathbf{P}\{S_2(i) = 0\}) \cdot$$

$$\cdot (\mathbf{P}\{S_2(j) = 0 \mid S_2(i) = 0\} - \mathbf{P}\{S_2(j) = 0\}).$$

Using (2. 6), a simple computation shows that

$$\sigma^2\{\nu(n)\} < c_9 \log\log n \tag{4.2}$$

for a suitable positive number c_9.

Let

$$q_k = [e^{e^{k^{10}}}] \qquad (k = 1, 2, \ldots); \tag{4.3}$$

then $\log\log q_k = k^{10}(1+o(1))$.

By Chebyshev's inequality, using (4.1) and (4.2), for any $\varepsilon > 0$,

$$\mathbf{P}\left\{\left|\frac{\nu(q_k)}{\log\log q_k} - \frac{1}{\pi}\right| > \varepsilon\right\} \leqq \frac{c_{10}}{k^5}$$

for a suitable constant c_{10}. Using the Borel—Cantelli lemma, we see that

$$\lim_{k\to\infty} \frac{\nu(q_k)}{\log\log q_k} = \frac{1}{\pi} \quad \text{with probability 1.}$$

Now if $q_{k+1} > n \geqq q_k$,

$$\frac{\nu(q_{k+1})}{\log\log q_k} \geqq \frac{\nu(n)}{\log\log n} \geqq \frac{\nu(q_k)}{\log\log q_{k+1}}.$$

By (4.3) $\dfrac{\log\log q_{k+1}}{\log\log q_k} \to 1$ as $k \to \infty$. It follows immediately that, with probability 1,

$$\lim_{n\to\infty} \frac{\nu(n)}{\log\log n} = \frac{1}{\pi}. \tag{4.4}$$

We now show that the result (4.4) is equivalent to (ii). By Theorem 2, we know that, with probability 1, $R_n > \log n(\log\log n)^{-8/2}$ for large n. Hence

$$W_n < e^{n(\log n)^2} \quad \text{for large } n. \tag{4.5}$$

Similarly from Theorem 3 we can deduce that

$$W_n > e^{n/\log n} \quad \text{for large } n. \tag{4.6}$$

Now $\sum_{s=1}^{n} \dfrac{1}{\log W_s} = \nu(W_n)$, so that, by (4.5) and (4.6),

$$\frac{1}{\log n}\,\nu(e^{n/\log n}) < \frac{1}{\log n}\sum_{s=1}^{n}\frac{1}{\log W_s} < \frac{1}{\log n}\,\nu(e^{n(\log n)^2}) \tag{4.7}$$

for large n, with probability 1.

Since, with probability 1, both sides of (4.7) approach the limit $1/\pi$, this completes the proof that

$$\mathbf{P}\left\{\lim_{n\to\infty}\frac{1}{\log n}\sum_{1}^{n}\frac{1}{\log W_s} = \frac{1}{\pi}\right\} = 1.$$

PROOF OF (i). Precisely the same method will work in this case, using the results of CHUNG and HUNT [1] instead of Theorems 2 and 3.

THEOREM 6. (i) *For a random walk on the line, let* $R_n(2k^2)$ *be the number of integers* k *for which* $S_1(2k^2)=0$ $(1 \leq k \leq n)$. *Then there is a positive number* c_{11} *such that*

$$\mathbf{P}\{\lim_{n\to\infty} R_n(2k^2)/\log n = c_{11}\} = 1.$$

(ii) *For a random walk in the plane, let* $R_n(2[k \log k])$ *be the number of integers* k *for which* $S_2(2[k\log k])=0$ $(1 \leq k \leq n)$. *Then there is a positive number* c_{12} *such that*

$$\mathbf{P}\left\{\lim_{n\to\infty} \frac{R_n(2[k\log k])}{\log\log n} = c_{12}\right\} = 1.$$

The proof of this theorem is very similar to that of Theorem 5, so we omit it.

Let us now ask the following question about random walks in the plane. We know that the walk returns to the origin infinitely often. However, there will be some long "gaps" when the walk does not return. How long can these gaps be? To make this precise, if $g(n)$ is a monotonic function, let us ask whether or not there are only finitely many integers n for which the path $(n, n+g(n))$ does not contain *at least* one return to the origin. We have succeeded in answering this question in the following form:

THEOREM 7. *Suppose* $f(n)$ *is a monotonic function which increases to* $+\infty$ *as* $x\to\infty$, *and let* E_n *be the event that the plane random walk path does not return to the origin between* n *and* $n^{f(n)}$: $\mathbf{P}\{E_n \text{ i.o.}\}=0$ *or* 1, *according as the series* $\sum_{k=1}^{\infty}\frac{1}{f(2^{2^k})}$ *converges or diverges.*

PROOF. Let F_n be the event $|S_2(n)| > n^{1/4}$. Then $\mathbf{P}(F_n) > 1 - n^{-1/4}$, and an application of (2. 16) will give

$$\mathbf{P}(F_n \cap E_n) > (1-n^{-1/4})\frac{2\log n^{1/4}}{f(n)\log n}(1+o(1)),$$

since the behaviour of the random walk after n depends only on its position at the n^{th} step. Hence, for large n,

$$\mathbf{P}(E_n) > \frac{1}{3}\cdot\frac{1}{f(n)}. \tag{4. 8}$$

Similarly, since $|S_2(n)| \leq n$, we can apply (2. 16) to give, again for large n,

$$\mathbf{P}(E_n) < \frac{3}{f(n)}. \tag{4. 9}$$

Now suppose first that $\Sigma \frac{1}{f(n_k)}$ converges where $n_k = 2^{2^k}$. Put $f_1(n) = \frac{1}{2} f(n)$, and let Q_k be the event that there is no return to the origin between n_k and $n_k^{f_1(n_{k+1})}$. By (4. 9), $\Sigma \mathbf{P}(Q_k)$ converges, so, with probability 1, there is an integer K such that there is a return between n_k and $n_k^{f_1(n_{k+1})}$ for all $k \geqq K$. Since $n^{f(n)} \geqq n_k^{f(n_k)} = n_{k+1}^{f_1(n_k)}$, this will imply that for $n_k \leqq n < n_{k+1}$, $k \geqq K$ there is at least one return between n and $n^{f(n)}$.

Conversely suppose that $\Sigma \frac{1}{f(n_k)}$ diverges. Because of the law of zero or one, it is sufficient to show that there exists an $\eta > 0$ such that for every integer k_1 there is an integer k_2 with

$$\mathbf{P}\left\{\bigcup_{k=k_1}^{k_2} E_{n_k}\right\} > \eta. \tag{4.10}$$

Since $f(n)$ is monotonic, the series $\sum_{k=1}^{\infty} \frac{1}{f(n_{5k})}$ must also diverge. Thus if k_1 is given and sufficiently large, we can certainly find k_2 such that

$$\frac{1}{20} < \sum_{k_1 \leqq 5k \leqq k_2} \frac{1}{f(n_{5k})} < \frac{1}{10}. \tag{4.11}$$

We will show that this choice of k_2 satisfies (4. 10) with $\eta = \frac{1}{40}$. Consider the events

$$D_k = E_{n_{5k}} - E_{n_{5k}} \cap \left[\bigcup_{5k < 5r \leqq k_2} E_{n_{5r}}\right]. \tag{4.12}$$

The sets D_k, for k an integer satisfying $k_1 \leqq 5k \leqq k_2$, are clearly disjoint and

$$\bigcup_{k=k_1}^{k_2} E_{n_k} \supset \bigcup_{k_1 \leqq 5k \leqq k_2} D_k,$$

so that (4. 10) will immediately follow if we can prove that, for $k_1 \leqq 5k \leqq k_2$,

$$\mathbf{P}(D_k) > \frac{1}{2} \mathbf{P}(E_{n_{5k}}). \tag{4.13}$$

There are two cases to consider in estimating $\mathbf{P}(D_k)$: (i) $r-k$ small, (ii) $r-k$ large.

(i) If r is such that $n_{5k}^{f(n_{5k})} > n_{5r}$, then the probability of no return between $n_{5k}^{f(n_{5k})}$ and $n_{5r}^{f(n_{5r})}$ is, on using (2. 16),

$$< (2+\varepsilon) \frac{\log n_{5k}}{\log n_{5r}} \frac{f(n_{5k})}{f(n_{5r})},$$

since we know that $|S_k(n_{5k}^{\zeta(n_{5k})})| < n_{5k}^{\zeta(n_{5k})}$. Thus we have in this case that

$$\mathbf{P}(E_{n_{5k}} \cap E_{n_{5r}}) < \frac{3}{2^{5(r-k)}} \mathbf{P}(E_{n_{5k}}). \tag{4.14}$$

(ii) On the other hand, if $n_{5k}^{\zeta(n_{5k})} \leqq n_{5r}$, we have by (4.9) that

$$\mathbf{P}(E_{n_{5k}} \cap E_{n_{5r}}) < \frac{3}{f(n_{5r})} \mathbf{P}(E_{n_{5k}}). \tag{4.15}$$

The two cases (4.14), (4.15) together show that

$$\mathbf{P}(E_{n_{5k}} \cap E_{n_{5r}}) < 3\left(\frac{1}{f(n_{5r})} + \frac{1}{2^{5(r-k)}}\right) \mathbf{P}(E_{n_{5k}})$$

for any $r > k$. This result applied to (4.11) and (4.12) immediately gives (4.13). This completes the proof of the theorem.

We state without proof the result for a one-dimensional walk which corresponds to Theorem 7. It can be proved by similar methods.

THEOREM 7A. *Suppose $f(n)$ is a monotonic function which increases to $+\infty$ as $x \to \infty$, and E_n is the event that a random walk path on the line does not return to the origin between n and $n\{f(n)\}^2$, then*

$$\mathbf{P}(E_n \ i.o.) = 0 \quad or \quad 1,$$

according as the series $\Sigma \frac{1}{f(2^k)}$ converges or diverges.

We end this section by mentioning a related problem which we have been unable to solve completely. Clearly the lattice points in any given square will eventually all be entered by a plane random walk. How quickly does this happen? More precisely we have

PROBLEM. How quickly does the function $f(n)$ need to increase so that in an infinite plane random walk, with probability 1, all the lattice points within a distance n of the origin will be entered by the walk before $f(n)$ steps except for finitely many values of n?

We can show using the methods we have discussed above that $f(n) = n^{(\log n)^{1+\varepsilon}}$ is large enough, but we have failed to get a satisfactory lower estimate and have no plausible conjecture regarding a necessary and sufficient condition for the rate of increase of $f(n)$.

5. Behaviour of the distance from the origin. For a random walk in d-space we put $\varrho_d(n) = |S_d(n)|$. Then for $d = 1$, the celebrated law of the iterated logarithm gives an upper bound to $\varrho_d(n)$ for large n. The corresponding theorem in d-space is

THEOREM 8. *For random walk in d-space*

$$\mathbf{P}\left\{\varrho_d(n) > c\sqrt{\frac{2}{d}\, n \log\log n}\ \ i.o.\right\} = 0 \quad or \quad 1,$$

according as $c > 1$ *or* $c \leqq 1$.

This result must be well-known though we have not found it stated explicitly in the literature. It can be proved by modifying the proof for the case $d = 1$.

This form of the theorem deals with the unusually large values of $\varrho_d(n)$. We may ask: how large can a sphere be for $S_d(n)$ to remain outside it for $n \geqq n_0$? This is equivalent to obtaining an upper bound for the rate of escape of $S_d(n)$. The lower bound was obtained in [2]. We will need to use

LEMMA 1. *If* $d \geqq 3$, *then for a random walk in d-space*

$$\mathbf{P}\{\varrho_d(n) < n^{1/2}(\log n)^{-2}\ i.o.\} = 0.$$

This is a special case of the rate of escape result of [2].

LEMMA 2. *If* $d \geqq 3$ *and we start a random walk in d-space at a distance* R *from* O, *then it will enter a sphere of centre* O *and radius* λR $(0 < \lambda < 1)$ *with probability*

$$p = (\lambda)^{d-2}(1 + o(1))$$

as $R \to \infty$.

This is proved for Brownian motion by DVORETZKY [3]. The random walk case follows immediately from the relationship connecting them (see [7]).

Because of the result of PÓLYA, the problem of rate of escape is only meaningful for $d \geqq 3$.

THEOREM 9. *Suppose* $c < 1$, $\varrho_d(n)$ *is the distance from* O *at the* n^{th} *step of a random walk in d-space,* $d \geqq 3$, *and* $\tau_d(n) = \inf_{m \geqq n} \varrho_d(m)$, *then*

$$\mathbf{P}\left\{\tau_d(n) > c\sqrt{\frac{2}{d}\, n \log\log n}\ \ i.o.\right\} = 1.$$

REMARK. Since $\tau_d(n) \leqq \varrho_d(n)$, it follows from Theorem 8 that for $c > 1$ we must have

$$\mathbf{P}\left\{\tau_d(n) > c\sqrt{\frac{2}{d}\, n \log\log n}\ \ i.o.\right\} = 0.$$

The case $c = 1$ can also be decided: in fact, by taking a great deal more care one can prove that, for any $a < 1$,

$$\mathbf{P}\left\{\tau_d(n) > \left[\frac{n}{d}(2\log\log n + a\, l_3(n))\right]^{1/2}\ \ i.o.\right\} = 1.$$

We have been unable to obtain necessary and sufficient conditions for upper bounds to the function $\tau_d(n)$ which would correspond to the results of ERDŐS [4] and FELLER [6].

PROOF OF THEOREM 9. Let c_{13}, c_{14} satisfy

$$0 < c < c_{13} < c_{14} < 1. \tag{5.1}$$

Consider a single axis in d-space and let $q(n)$ be the number of steps which are taken in the direction of this axis. Let

$$Q_n = \left\{ q(n) \geqq \left[\frac{n}{d}\right] \right\}.$$

Then, since steps along the direction of each of the d axes are equally likely, we have

$$\mathbf{P}(Q_n) \geqq \frac{1}{2}. \tag{5.2}$$

Now, by considering only the distance from the origin in the direction of this one specified axis, we have

$$\mathbf{P}\left\{ \varrho_d(n) > c_{14}\left(\frac{2}{d}\, n \log\log n\right)^{1/2} \middle| Q_n \right\} \geqq \frac{1}{(\log n)^{c_{14}}}. \tag{5.3}$$

Further, by Lemma 2,

$$\mathbf{P}\left\{ \tau_d(n) > c_{13}\left(\frac{2}{d}\, n \log\log n\right)^{1/2} \middle| \varrho_d(n) > c_{14}\left(\frac{2}{d}\, n \log\log n\right)^{1/2} \right\} > \tag{5.4}$$

$$> 1 - \left(\frac{c_{13}}{c_{14}}\right)^{d-2} (1 + o(1));$$

since the required probability is that of not entering a sphere of radius $c_{13}x$ if you start from a distance $c_{14}x$ from its centre. By (5.1), (5.2), (5.3) and (5.4) it follows that for a suitable c_{15} we have

$$\mathbf{P}\left\{ \tau_d(n) > c_{13}\left(\frac{2}{d}\, n \log\log n\right)^{1/2} \right\} > \frac{c_{15}}{(\log n)^{c_{14}}}. \tag{5.5}$$

In order to apply the Borel—Cantelli lemma we must replace the events in (5.5) by suitable independent ones. Let

$$n_k = [e^{k^{1+\delta}}] \qquad (k = 1, 2, \ldots) \tag{5.6}$$

where

$$1 < 1 + \delta < c_{14}^{-1}. \tag{5.7}$$

Now put

(5. 8) $$\mu_k = \inf_{n_{k+1} \leqq n \leqq n_{k+2}} |S_d(n) - S_d(n_k)|.$$

Putting $n_{k+1} - n_k = t_k$, we clearly have

$$\mathbf{P}\{\mu_k > \lambda\} \geqq \mathbf{P}\{\tau_d(t_k) > \lambda\} \quad \text{for any} \quad \lambda \geqq 0.$$

Hence if we put

$$E_k = \left\{ \mu_k > c_{15} \left(\frac{2}{d} t_k \log \log t_k \right)^{1/2} \right\},$$

we have by (5. 5) and (5. 6) that

$$\mathbf{P}(E_k) > \frac{c_{15}}{k^{c_{16}}} \quad \text{where} \quad 0 < c_{16} < 1.$$

But, by (5. 8), the events $E_2, E_4, \ldots, E_{2k}, \ldots$ are independent. Hence, by Borel—Cantelli,

(5. 9) $$\mathbf{P}\{E_k \ i.o.\} = 1.$$

Now, by Theorem 8, there exists with probability 1 an integer k_0 such that for $k \geqq k_0$

$$\varrho_d(n_k) < 2 \left(\frac{n_k}{d} \log \log n_k \right)^{1/2}.$$

Using (5. 6) this shows that

(5. 10) $$\varrho_d(n_k) < n_{k+1}^{1/2}.$$

Finally, by Lemma 1, there exists with probability 1 an integer k_1 such that if $k \geqq k_1$,

$$\tau_d(n_{k+2}) \geqq \frac{n_{k+2}^{1/2}}{(\log n_{k+2})^2}.$$

Again using (5. 6), this shows that

(5. 11) $$\tau_d(n_{k+2}) \geqq (n_{k+1} \log \log n_{k+1})^{1/2}.$$

Now suppose $k \geqq \mathrm{Max}(k_0, k_1)$ is such that E_k occurs. Then, by (5. 8) and (5. 10),

$$\inf_{n_k+1 \leqq n \leqq n_{k+2}} |S_d(n)| > c_{15} \left(\frac{2}{d} t_k \log \log t_k \right)^{1/2} - n_{k+1}^{1/2} > c \left(\frac{2}{d} n_{k+1} \log \log n_{k+1} \right)^{1/2}$$

for large enough k, by (5. 1), (5. 5). Now using (5. 11) we see that for such values of k we have

$$\tau_d(n_{k+1}) > c \left(\frac{2}{d} n_{k+1} \log \log n_{k+1} \right)^{1/2}.$$

This completes the proof of the theorem.

We now state a theorem regarding the average rate of growth of $\varrho_d(n)$. This is relevant for $d=1,2$ as well as $d\geqq 3$.

THEOREM 10. *For a random walk in d-space, if $\varrho_d(n)$ denotes the distance from the origin of $S_d(n)$, then there are constants λ_d such that*

(i) $$\frac{1}{\log N}\sum_{n=1}^{N}\frac{n^{-1/2}}{1+\varrho_1(n)}\to\lambda_1 \quad \textit{with probability } 1,$$

(ii) $$\frac{1}{(\log N)^2}\sum_{n=1}^{N}\frac{1}{1+\{\varrho_2(n)\}^2}\to\lambda_2 \quad \textit{with probability } 1,$$

(iii) $$\frac{1}{\log N}\sum_{n=1}^{N}\frac{1}{1+\{\varrho_d(n)\}^2}\to\lambda_d \quad \textit{with probability } 1, \quad \textit{for} \quad d=3,4,\ldots .$$

PROOF. The method of proof is very similar to that used in Theorem 5. The three cases are also very similar, so we consider only the plane case (ii).

Suppose P is a lattice point in the plane. Let $\mathbf{Y}(P,n)$ be the probability that $S_2(n)=P$. Then

(a) If $|P|<n^{1/2}/\log n$,

$$\mathbf{P}\{\mathbf{Y}(P,n)\}=0 \quad \text{or} \quad \frac{2}{\pi n}(1+o(1)), \tag{5.12}$$

according as P cannot or can be reached in n steps. Note that for fixed n half the points can be reached and the other half cannot.

(b) If $|P|>n^{1/2}\log n$,

$$\mathbf{P}\{\mathbf{Y}(P,n)\}=O\left(\frac{1}{n^2}\right). \tag{5.13}$$

Thus $\mathcal{E}\left(\frac{1}{1+\{\varrho_d(n)\}^2}\right)=\sum_P\frac{\mathbf{P}\{\mathbf{Y}(P,n)\}}{1+|P|^2}$. Using (5.12) and (5.13) we have

$$\sum_{|P|<\frac{n^{1/2}}{\log n}}\frac{\mathbf{P}\{\mathbf{Y}(P,n)\}}{1+|P|^2}<\mathcal{E}\left(\frac{1}{1+\{\varrho_d(n)\}^2}\right)<\sum_{|P|<n^{1/2}\log n}\frac{\mathbf{P}\{\mathbf{Y}(P,n)\}}{1+P^2}+O(1)$$

which gives

$$\frac{2}{n}(\log n-\log\log n)<\mathcal{E}\left(\frac{1}{1+\{\varrho_d(n)\}^2}\right)(1+o(1))<\frac{2}{n}(\log n+\log\log n)$$

or

$$\mathcal{E}\left(\frac{1}{1+\{\varrho_d(n)\}^2}\right)=\frac{2\log n}{n}(1+o(1)). \tag{5.14}$$

Hence we have

$$\mathcal{E}\left\{\frac{1}{(\log N)^2}\sum_{n=1}^{N}\frac{1}{1+\{\varrho_4(n)\}^2}\right\}=\frac{2}{(\log N)^2}\sum_{n=1}^{N}\frac{\log n}{n}(1+o(1))=1+o(1)$$
$$\text{as}\quad N\to\infty.$$

A similar computation will show that the variance is small. It can be shown that

$$\sigma^2\left\{\frac{1}{(\log N)^2}\sum_{n=1}^{N}\frac{1}{1+\{\varrho_4(n)\}^2}\right\}=O\left(\frac{1}{\log N}\right).$$

The argument of Theorem 5 can be applied, proving first that the limit exists as $n\to\infty$ through the sequence $r_k=[e^{k^4}]$ and then deducing the general result.

6. Multiplicity of points on a random walk. A point P of the lattice is of multiplicity $m(P,n)$ if the random walk of n steps is at P precisely $m(P,n)$ times in the first n steps. Let us first consider how many points there are which are entered once and only once. For $d=1$, there will be 0, 1, or 2 of these, while for $d\geqq 2$ there will clearly be many. Let us consider the case $d=2$ in some detail. In a plane random walk of n steps how many points have multiplicity one?

In [2] it is shown that the probability that the k^{th} step of a random walk brings it to a point not previously entered is $\gamma_2(k)$, the same is the probability of no return to the origin in the first $k-1$ steps. It is clear therefore that the probability that at the k^{th} step a plane random walk enters a new point to which it does not return before the n^{th} step is $\gamma_2(k)\gamma_2(n-k)$.

Let $M_1(n)$ be the number of points of multiplicity 1 on a plane random walk of n steps. Then clearly

$$\mathcal{E}\{M_1(n)\}=\sum_{k=0}^{n}\gamma_2(k)\gamma_2(n-k). \tag{6.1}$$

By (2.1), $\mathcal{E}\{M_1(n)\}\geqq(n+1)\{\gamma_2(n)\}^2$.

Using the estimate (2.5), we have

$$\mathcal{E}\{M_1(n)\}\geqq n\left[\frac{\pi^2}{(\log n)^2}+O\left(\frac{1}{(\log n)^3}\right)\right]. \tag{6.2}$$

Again using (2.1) and (2.5) we have, if $k_1=\left[\frac{n}{(\log n)^2}\right]$, that

$$\mathcal{E}\{M_1(n)\}\leqq 2\sum_{k=0}^{k_1}\gamma_2(n-k)+\gamma_2(k_1)\sum_{k=k_1}^{n-k_1}\gamma_2(k)\leqq$$
$$\leqq O\left(\frac{n}{(\log n)^3}\right)+n\left[\frac{\pi^2}{(\log n)^2}+O\left(\frac{\log\log n}{(\log n)^3}\right)\right].$$

This together with (6. 2) shows that

$$\mathscr{E}\{M_1(n)\} = n\left[\frac{\pi^2}{(\log n)^2} + O\left(\frac{\log\log n}{(\log n)^3}\right)\right]. \tag{6.3}$$

In order to estimate the variance we need

LEMMA 3. *Let $\nu(n)$ be the probability that a plane random walk path* (i) *does not return to the origin in the first n steps and* (ii) *enters a new point at the n^{th} step. Then*

$$\nu(n) \leq \left\{\gamma_2\left(\left[\frac{n}{2}\right]\right)\right\}^2.$$

REMARK. It is clear that $\nu(n) \sim \{\gamma_2(n)\}^2$ as $n \to \infty$, but we need only an upper bound.

PROOF. Let $q = \left[\frac{n}{2}\right]$; then the probability that a random walk path has not returned to the origin in the first $q-1$ steps is $\gamma_2(q)$. Start a path from $S_2(q)$ of length $n-q+1$ steps. The probability that the last of these steps brings the path to a point not entered since $S_2(q)$ is $\gamma_2(n-q+1) \leq \gamma_2(q)$, by (2. 1). Now if the path is to satisfy both conditions (i) and (ii) it clearly must not return to the origin in first $q-1$ steps, and the n^{th} point must certainly be a point not entered since the q^{th} step. Thus $\nu(n) \leq \{\gamma_2(q)\}^2$, as required.

Now let p_i be the probability that in a random walk of n steps, the i^{th} step leads to a new point of multiplicity one, and let p_{ij} be the probability that at both the i^{th} and j^{th} steps points of multiplicity one are entered. By splitting the path into three parts it follows that

$$p_{ij} \leq \gamma_2(i)\nu(j-i)\gamma_2(n-j) \tag{6.4}$$

for $0 \leq i < j \leq n$. Now

$$\sigma^2\{M_1(n)\} = \sum_{i,j}(p_{ij} - p_i p_j) \leq 2 \sum_{1 \leq i \leq j \leq n}(p_{ij} - p_i p_j).$$

Using Lemma 3 and (6. 4), this gives

$$\sigma^2\{M_1(n)\} \leq 2 \sum_{1 \leq i \leq j \leq n} \gamma_2(i)\gamma_2(n-j)\left\{\left[\gamma_2\left(\frac{j-i}{2}\right)\right]^2 - \gamma_2(j)\gamma_2(n-i)\right\}. \tag{6.5}$$

The double sum in (6. 5) can be estimated by splitting it into 4 parts. Let $k_1 = \left[\frac{n}{(\log n)^5}\right]$. Since all the terms are positive and less than 1,

$$\sum_{1 \leq i \leq j \leq n} \leq \sum_{1 \leq i \leq k_1} \sum_{1 \leq j \leq n} + \sum_{1 \leq i \leq n} \sum_{n-k_1 \leq j \leq n} + \sum_{1 \leq i \leq n} \sum_{i \leq j \leq i+k_1} + \sum_{k_1 \leq i \leq n-k_1} \sum_{i+k_1 \leq j \leq n-k_1}.$$

The first 3 terms are $O\left(\frac{n^2}{(\log n)^6}\right)$, and the fourth, by using (2.5), is $O\left(\frac{n^2 \log\log n}{(\log n)^5}\right)$. Thus

$$\sigma^2\{M_1(n)\} = O\left(\frac{n^2 \log\log n}{(\log n)^5}\right).$$

This variance is not quite small enough for a straightforward application of Chebyshev's inequality. However, the method used in Section 5 of [2] can be applied here with only minor modifications to show that

$$\mathbf{P}\left\{\left|M_1(n) - \frac{\pi^2 n}{(\log n)^2}\right| > \frac{\varepsilon n}{(\log n)^2}\right\} = O\left(\frac{1}{(\log n)^{1+\delta}}\right) \qquad (\delta > 0),$$

and the strong law can be deduced, as in [2], by using the sequence $t_k = [e^{k^\theta}]$ for

$$\frac{1}{1+\delta} < \theta < 1.$$

For details of the method the reader is referred to [2]. Thus we have proved

THEOREM 11. *If $M_1(n)$ is the number of points of the lattice entered once and only once in the first n steps of a plane random walk, then*

$$\mathbf{P}\left\{\lim_{n\to\infty} \frac{M_1(n)(\log n)^2}{\pi^2 n} = 1\right\} = 1.$$

REMARK. For a fixed positive integer t, a modified version of the above proof will show that the number of points of multiplicity t in the first n steps of a plane random walk is given asymptotically by the same formula

$$\frac{\pi^2 n}{(\log n)^2}.$$

A much simplified version of the same argument suffices to prove

THEOREM 12. *If t is a positive integer, and $d \geqq 3$, then the number $Q_d(t, n)$ of points which are entered by a random walk in d-space precisely t times in the first n steps is such that*

$$\mathbf{P}\left\{\lim_{n\to\infty} \frac{Q_d(t, n)}{n} = \gamma_d(1-\gamma_d)^{t-1}\right\} = 1,$$

where γ_d is the probability that the path will never return to the origin.

REMARK. This means that in $d \geqq 3$ dimensions the proportion of points entered by a random walk of n steps which have a given multiplicity agrees with the distribution (3. 1) for the number of returns to the origin. We feel sure that this result must also be true for the plane random walk, though we have not attempted to prove it.

The result of Theorem 12 shows that, for $d \geqq 3$, most of the points entered will have small multiplicity. Let us now ask what is the largest multiplicity occurring in the first n steps of a random walk.

THEOREM 13. *Let $T_d(n)$ be the upper bound of the multiplicity of points entered in the first n steps of a d-dimensional random walk ($d \geqq 3$). Then*

$$\mathbf{P}\left\{\lim_{n\to\infty}\frac{T_d(n)}{\log n}=\lambda_d\right\}=1$$

where

$$\lambda_d=-\frac{1}{\log(1-\gamma_d)} \qquad (d=3, 4, \ldots).$$

PROOF. Suppose first that $\lambda>\lambda_d$; then by (3. 1),

$$\mathbf{P}\{R_d(n)>\lambda\log n\}<(1-\gamma_d)^{\lambda\log n}.$$

There are at most n points entered: hence

$$\mathbf{P}\{T_d(n)>\lambda\log n\}<n(1-\gamma_d)^{\log n}=n^{1-\lambda/\lambda_d}.$$

Using Borel—Cantelli it follows that the event $\{T_d(n)>(\lambda_d+\varepsilon)\log n\}$ happens only finitely often for the sequence $n_k=2^k$ $(k=1,2,\ldots)$ and as a result happens for only finitely many integers n with probability 1.

There are independence difficulties in obtaining the result in the opposite direction. This time we avoid these by splitting the path into a large number of small pieces.

Let

$$u=[\log n], \qquad v=\left[\frac{n}{(\log n)^2}\right].$$

Consider a piece of the path containing u^2 steps. The probability that the first point of this piece is returned to in the first u steps is

$$1-\gamma_d+O\left(\frac{1}{(\log n)^{1/2}}\right)$$

by (2. 3). Hence the probability that this first point is entered $\lambda\log n$ times in the u^2 steps is at least

$$\mu(n)=\left\{1-\gamma_d+O\left(\frac{1}{(\log n)^{1/2}}\right)\right\}^{[\lambda\log n]}.$$

There are at least v such pieces which are now independent. It follows that

$$\mathbf{P}\{T_d(n) > \lambda \log n\} \geqq 1 - \{1-\mu(n)\}^v,$$

so that

$$\mathbf{P}(T_d(n) \leqq \lambda \log n\} < \{1-\mu(n)\}^v < e^{-n^\delta}$$

for a suitable $\delta > 0$, provided $\lambda < \lambda_d$. Hence, if $\lambda = \lambda_d$, by Borel—Cantelli, there are, with probability 1, only finitely many n for which

$$T_d(n) \leqq \lambda \log n.$$

This completes the proof of the theorem.

The problem of maximum multiplicity also has a meaning in the case $d=2$. The method used in the proof of Theorem 13, using pieces of length $[n^{1/2}]$ and the estimates (3. 6) and (3. 11) is good enough to prove that

$$\mathbf{P}\left\{\frac{1}{4\pi} \leqq \liminf_{n\to\infty} \frac{T_2(n)}{(\log n)^2} \leqq \limsup_{n\to\infty} \frac{T_2(n)}{(\log n)^2} \leqq \frac{1}{\pi}\right\} = 1.$$

We think it likely that in fact

$$\mathbf{P}\left\{\lim_{n\to\infty} \frac{T_2(n)}{(\log n)^2} = \frac{1}{\pi}\right\} = 1,$$

though we have not succeeded in proving this.

(Received 6 October 1959)

References

[1] K. L. Chung and C. A. Hunt, On the zeros of $\sum_1^n \pm 1$, *Annals of Math.*, **50** (1949), pp. 385—400.

[2] A. Dvoretzky and P. Erdős, Some problems on random walk in space, *Proceedings of second Berkeley Symposium on Stochastic Processes*, (1950), pp. 353—367.

[3] A. Dvoretzky, Brownian motion in space and subharmonic functions (under press).

[4] P. Erdős, On the law of the iterated logarithm, *Annals of Math.*, **43** (1942), pp. 419—436.

[5] P. Erdős and S. J. Taylor, Some intersection properties of random walk paths, *Acta Math. Acad. Sci. Hung.*, **11** (1960) (under press).

[6] W. Feller, The general form of the so-called law of the iterated logarithm, *Trans. Amer. Math. Soc.*, **54** (1943), pp. 373—402.

[7] M. Kac, Random walk and Brownian motion, *Amer. Math. Monthly*, **54** (1947), pp. 369—391.

[8] G. Pólya, Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend der Irrfahrt im Straßennetz, *Math. Annalen*, **84** (1921), pp. 149—160.