Probab. Th. Rel. Fields 77, 117–128 (1988)

Probability Theory and Related Fields

On the Set Visited Once by a Random Walk *

Péter Major
Mathematical Institute of the Hungarian Academy of Sciences, Reáltanoda u. 13–15, H-1053 Budapest, Hungary

Summary. In this paper we prove the following statement. Given a random walk $S_n=\sum_{j=1}^{n} \varepsilon_j$, $n=1, 2, \ldots$ where $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. random variables, $P(\varepsilon_j=1)=P(\varepsilon_j=-1)=\frac{1}{2}$, let $\alpha(n)$ denote the number of points visited exactly once by this random walk up to time n. We show that there exists some constant C, $0<C<\infty$, such that $\limsup_{n\to\infty} \frac{\alpha(n)}{\log^2 n}=C$ with probability 1. The proof applies some arguments analogous to the techniques of the large deviation theory.

1. Introduction

The following problem was proposed by P. Erdös and P. Révész: Let us consider a random walk $S_0=0$, $S_n=\sum_{j=1}^{n} \varepsilon_j$, $n=1, 2, \ldots$ where $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. random variables, $P(\varepsilon_j=1)=P(\varepsilon_j=-1)=\frac{1}{2}$, and define the random set $\mathscr{A}_n$ consisting of the points visited by this random walk up to time n exactly once, i.e. let

$$\mathscr{A}_n=\{x, \exists k, 0\leqq k\leqq n, \text{ such that } S_k=x, \text{ and } S_l\neq x \text{ for } 0\leqq l\leqq n, l\neq k\}.$$

Let $\alpha(n)=|\mathscr{A}_n|$ denote the number of points in the set $\mathscr{A}_n$. Let $f(n)$ be an arbitrary function such that $f(n)\to\infty$ as $n\to\infty$. It is not difficult to see with the help of the zero-one law that there exists some constant C such that $\limsup_{n\to\infty} \frac{\alpha(n)}{f(n)}=C$ with probability 1, but $C=0$ or $C=\infty$ is also possible. Erdös and Révész asked with which choice of $f(n)$ is the above constant C such that $0<C<\infty$. This question is answered in the following

* Research supported by the Hungarian National Foundation for Scientific Research, Grant N° # 819/1

Theorem 1. *There exists some* $0<C<\infty$ *such that* $P\left(\limsup \frac{\alpha(n)}{\log^2 n}=C\right)=1$.

We can give an upper and a lower bound for the constant C in Theorem 1, but we cannot give its exact value. In the proof of Theorem 1 we need a good asymptotic for the probability $P(\alpha(n)>x_n)$ for certain values of x_n. Especially, we are interested in the question for which x_n has the above probability of order $1/n$. Hence we shall prove the following

Theorem 2. *For all* $K>0$ *and* $\varepsilon>0$ *there exists an* $n_0=n_0(K,\varepsilon)$ *threshold such that for* $n>n_0$

$$n^{-\frac{K}{L}-\varepsilon}<P(\alpha(n)>K\log^2 n)<n^{-\frac{K}{L}+\varepsilon}$$

with a constant $L>0$ *which is defined in the following Lemma* 3.

Lemma 3. *There exists some* L, $0<L<\infty$, *such that our random walk* S_n *satisfies the relation*

$$\lim_{n\to\infty} nP(S_j>0 \textit{ for all } 0<j\leqq n \textit{ and } S_j<S_n \textit{ for all } 0\leqq j<n)=L.$$

Erdös and Révész conjectured that in Theorem 1 one has to divide by $\log n$ instead of $\log^2 n$. This conjecture turned out to be incorrect, but it deserves some consideration. It is relatively simple to prove that with probability 1 there are infinitely many n and an appropriately small $C>0$ such that for these n there is a block $[k, k+c\log n]$, $k+c\log n<n$, with the following properties: a) $\varepsilon_k=\varepsilon_{k+1}=\ldots\varepsilon_{k+c\log n}=1$, b) $S_k>S_j$ for $j<k$ and $S_{k+c\log n}<S_j$ for $k+c\log n<j\leqq n$. These properties imply that $\{S_k, S_{k+1}, \ldots, S_{k+c\log n}\}\subset\mathscr{A}_n$ and $\alpha(n)>c\log n$ for these n. The idea behind the conjecture is the feeling that this is the typical way how exceptionally large sets $\mathscr{A}_n$ appear. Let us remark that if we have a dispersed set $0\leqq j_1<j_2<\ldots<j_m\leqq n$, where j_k-j_{k-1} is relatively large for $k=2,3,\ldots,m$ then the probability of the event $\{S_{j_1}, S_{j_2}, \ldots, S_{j_m}\}\subset\mathscr{A}_n$ is much less than the probability of the event $\{S_j, S_{j+1}, \ldots, S_{j+m}\}\subset\mathscr{A}_n$. But on the other hand there are much more dispersed sets $0<j_1<j_2<\ldots<j_m\leqq n$ than blocks $[j, j+m]$, $j\leqq n$. Hence the typical way how exceptionally large sets $\mathscr{A}_n$ appear is that $\mathscr{A}_n=\{S_{j_1}, \ldots, S_{j_m}\}$, and $\{j_1, \ldots, j_m\}$ is a dispersed set. This is the reason why Erdös and Révész' conjecture does not hold, but the proof requires a more refined analysis.

The paper consists of three sections. In Sect. 2 we prove Theorem 2 together with two lemmas, in Sect. 3 we prove Theorem 1, make some comments, and formulate some open problems.

2. The Proof of Theorem 2

We start this section with the

Proof of Lemma 3. Let us first consider the case $n=2m$. We have

$$I=P(S_j>0 \text{ for all } 0<j\leqq n,\ S_j<S_n \text{ for all } 0\leqq j<n)$$

$$=P(S_m>\sup_{m<j\leqq n}\{-((S_j-S_n)-(S_m-S_n))\},\ (S_n-S_m)>\sup_{0\leqq l<m} S_l-S_m \mid B_n)P(B_n) \quad (2.1)$$

with

$$B_n=\{S_j>0 \text{ for all } 0<j\leqq m,\ S_n-S_l>0 \text{ for all } m\leqq l<n\}.$$

Let us define the probability measures μ_n, $n=1, 2, \ldots$ on the plane R^2

$$\mu_n(dx, dy)=P\left(\frac{S_n}{\sqrt{n}}=x,\ \sup_{0\leqq k<n}\frac{S_k}{\sqrt{n}}=y \mid S_j>0 \text{ for all } 1\leqq j\leqq n\right).$$

Since the random vectors $\{S_1, \ldots, S_m\}$ and $\{S_n-S_{n-1}, S_n-S_{n-2}, \ldots, S_n-S_m\}$ are independent and identically distributed, hence it follows from (2.1) that

$$I=\mu_m\times\mu_m(A)\cdot P^2(S_j>0 \text{ for all } 0<j\leqq m) \tag{2.2}$$

with the following set $A\subset R^2\times R^2$:

$$A=\{(x_1, y_1, x_2\cdot y_2), x_1>y_2-x_2, x_2>y_1-x_1\}.$$

It is known (see, e.g. [3], Chap. III., Sect. 3) that

$$P(S_j>0 \text{ for all } 0<j\leqq m)\sim\frac{1}{\sqrt{n\pi}} \tag{2.3}$$

On the other hand it is proved (see, e.g. [1], Theorem 4.3 with $t=1$) that there is a probability measure μ^* such that $\mu_n\xrightarrow{w}\mu^*$. The measure μ^* is defined in [1] explicitly. We are not interested in its exact form, what is imported for us is the relation

$$\lim_{m\to\infty}\mu_m\times\mu_m(A)=\mu^*\times\mu^*(A)>0 \tag{2.4}$$

Relations (2.2), (2.3) and (2.4) imply Lemma 3 in the case $n=2m$. The case $n=2m+1$ is similar. Lemma 3 is proved.

Let $\alpha_k(n)$ denote the number of subsets of $\mathscr{A}_n$ with exactly k elements, i.e. let $\alpha_k(n)=\binom{|\mathscr{A}_n|}{k}$. In the proof of Theorem 2 the following Lemma 4 plays an essential role.

Lemma 4. *Let $k\sim\alpha\log n$ with some $\alpha>0$. For all $\eta>0$ there exists some $n_0=n_0(\alpha, \eta)$ threshold such that for $n>n_0$*

$$[(L-\eta)\log n]^k<E\alpha_k(n)<[(L+\eta)\log n]^k$$

where L is the same as in Lemma 3.

Before its proof we make some comments about the role of Lemma 4 in the proof of Theorem 2. Theorem 2 is a large deviation type result. Indeed, as we shall later show in Remark 1, $\alpha(n)$ is less than a sufficiently large constant (which is independent of n) with probability almost one, and Theorem 2 gives an asymptotic for the probability of the event $\alpha(n)>K\log^2 n$. In classical large deviation theory one usually investigates the tail behaviour of a random variable Z_n with the help of its moment generating function $\varphi_n(t)=\mathsf{E}\exp(tZ_n)$. In our investigation the function $\mathsf{E}\alpha_k(n)$ plays the same role as the moment generating function in other problems. Here the good choice of the parameter k (like the good choice of the parameter t in $\varphi_n(t)$ in other cases) supplies good estimates for us. We can get an upper bound for $\alpha(n) > K \log^2 n$ with the help of the Markov inequality. In order to get a lower bound first we have to find the size of those random sets $\mathscr{A}_n$ whose subsets give the main contribution to the expectation $\mathsf{E}\alpha_k(n)$. Also the lower bound can be obtained with the help of Lemma 4. The method of the proof is similar to the way as the strong convexity of $\log \varphi_n(t)$ is exploited in the large deviation theory.

Proof of Lemma 4. Let us define the following events:

$$C_1(r,t) = \{\omega\colon\ S_r(\omega) < S_l(\omega) < S_t(\omega) \ \text{ for all } r < l < t\}$$

$$D_1(j) = \{\omega\colon\ S_l(\omega) < S_j(\omega), \ \text{ for all } 0 \leqq l < j\}$$

$$D_2(j) = D_2(j,n) = \{\omega\colon\ S_l(\omega) > S_j(\omega) \ \text{ for all } j < l \leqq n\}.$$

By Lemma 3

$$P(C(r,t)) = P(C(0,t-r)) = \frac{L}{t-r}(1+o(1)). \tag{2.5}$$

On the other hand (see, e.g. [2], Chap. III.)

$$P(D_1(j)) = \frac{1}{\sqrt{2\pi j}}(1+o(1)) \tag{2.6}$$

$$P(D_2(j)) = \frac{1}{\sqrt{2\pi(n-j)}}(1+o(1)). \tag{2.6'}$$

Put $\mathscr{A}_n^+ = \mathscr{A}_n \cap \{z, z \geqq 0\}$, $\alpha^+(n) = |\mathscr{A}_n^+|$ and $\alpha_k^+(n) = \binom{|\mathscr{A}_n^+|}{k}$. Then by exploiting that $\{S_{j_1}(\omega), S_{j_2}(\omega), \ldots, S_{j_k}(\omega)\} \subset \mathscr{A}_n^+$, $0 \leqq j_1 < j_2 < \ldots < j_k \leqq n$, if and only if $\omega \in D_1(j_1) \cap C(j_1,j_2) \cap \ldots \cap C(j_{k-1},j_k) \cap D_2(j_k)$ we get that

$$\begin{aligned}\mathsf{E}\alpha_k^+(n) &= \sum_{0 \leqq j_1 < j_2 < \ldots < j_k \leqq n} P(D_1(j_1) \cap C(j_1,j_2) \cap \ldots \cap C(j_{k-1},j_k) \cap D_2(j_k)) \\ &= \sum_{0 \leqq j_1 < j_2 < \ldots < j_k \leqq n} P(D(j_1))\, P(C(j_1,j_2)) \ldots P(C(j_{k-1},j_k))\, P(D_2(j_k)).\end{aligned} \tag{2.7}$$

Set

$$U(j,l) = \sum_{j = j_1 < j_2 < \ldots < j_k = l} P(C(j_1,j_2))\, P(C(j_2,j_3)) \ldots\ldots P(C(j_{k-1},j_k)).$$

Then

$$U(j,l) = U(0,l-j)$$

and

$$\mathsf{E}\alpha_k^+(n) = \sum_{0 \leqq j_1 < j_k \leqq n} U(j_1, j_k) P(D(j_1)) P(D(j_k))$$
$$= \sum_{r=0}^{n} U(0,r) \sum_{j=0}^{n-r} P(D_1(j)) P(D_2(r+j)).$$

Hence (2.6), (2.6)′ and (2.5) imply that

$$\mathsf{E}\alpha_k^+(n) \leqq \text{const} \sum_{r=0}^{n} U(0,r) \sum_{j=1}^{n-r-1} \frac{1}{\sqrt{j}} \frac{1}{\sqrt{n-j-r}}$$
$$\leqq \text{const}' \cdot \sum_{r=0}^{n} U(0,r) \leqq \text{const}' \left(\sum_{j=0}^{n} P(C(0,j)) \right)^{k-1}$$
$$\leqq [(L+\eta)\log n]^k. \tag{2.8}$$

On the other hand, since $j_1 \leqq \frac{n}{3}$, $j_l - j_{l-1} \geqq \frac{n}{3k}$, $l = 1, 2, \ldots, k$ imply that $j_l \leqq \frac{2}{3} n$, $P(D_2(j_l)) \geqq \frac{\text{const}}{\sqrt{n}}$, hence (2.7) implies that

$$\mathsf{E}\alpha_k^+(n) \geqq \sum_{\substack{0 \leqq j_1 \leqq \frac{n}{3} \\ \frac{n}{3k} \geqq j_l - j_{l-1} > 0 \\ l=1, \ldots, k}} P(D_1(j_1)) P(C(0, j_2 - j_1)) \ldots P(C(0, j_k - j_{k-1}) P(D_2(j_k)))$$
$$\geqq \frac{\text{const}}{\sqrt{n}} \sum_{j=1}^{n/3} P(D_1(j)) \left[\sum_{j=1}^{\frac{n}{3k}} P(C(0,j)) \right]^{k-1}$$
$$\geqq \text{const} \left[\left(L - \frac{\eta}{2} \right) \log \frac{n}{3k} \right]^{k-1} \geqq [(L-\eta)\log n]^k. \tag{2.9}$$

Put $\mathscr{A}_n^- = \mathscr{A}_n - \mathscr{A}_n^+$, $\alpha_n^- = |\mathscr{A}_n^-|$ and $\alpha_k^-(n) = \binom{|\mathscr{A}_n^-|}{k}$. The estimates (2.8) and (2.9) also hold for $\mathsf{E}\alpha_k^-(n)$. We claim that

$$\mathsf{E}\alpha_k^+(n) \leqq \mathsf{E}\alpha_k(n) \leqq \mathsf{E}\alpha_k^+(n) + \mathsf{E}\alpha_{k-1}^+(n) + \mathsf{E}\alpha_k^-(n) + \mathsf{E}\alpha_{k-1}^-(n). \tag{2.10}$$

Indeed, the left hand side of (2.10) is trivial, and one can see the right hand side with the help of the observation that $|\mathscr{A}_n^+| > 1$ implies $|\mathscr{A}_n^-| \leqq 1$ ($|\mathscr{A}_n^-|$ can contain only $\inf_{0 \leqq k \leqq n} S_k$ in this case) and $|\mathscr{A}_n^-| > 1$ implies $|\mathscr{A}_n^+| \leqq 1$. Lemma 4 follows from (2.8), (2.9) and (2.10).

Proof of Theorem 2. a) The proof of the upper bound: Set $k = \frac{K}{L} \log n$. We have by Lemma 4 and Stirling's formula

$$P\alpha(n)>K\log^2 n)=P\left(\alpha_k(n)>\binom{K\log^2 n}{k}\right)\leqq\frac{\mathsf{E}\alpha_k(n)}{\binom{K\log^2 n}{k}}$$

$$\leqq\frac{(L+\eta)^k\log n^k}{(K\log^2 n)^k}\cdot\left(\frac{1}{e}\frac{K}{L}\log n\right)^k=\left(\frac{1}{e}\frac{L+\eta}{L}\right)^k\leqq n^{-\frac{K}{L}+\varepsilon}$$

if η is chosen sufficiently small.

b) The proof of the lower bound: Put $q_m=q_m(n)=P(\alpha(n)=m)$, $k=\frac{K}{L}\log n$, $k'=k(1+\delta)$ with

$$\delta=\varepsilon\frac{L}{100K}\quad\text{and}\quad\bar{K}=K(1+2\delta).$$

We claim that

$$\sum_{m<K\log^2 n} q_m\binom{m}{k'}\leqq\tfrac{1}{3}\mathsf{E}\alpha_{k'}(n)\tag{2.11}$$

and

$$\sum_{m>\bar{K}\log^2 n} q_m\binom{m}{k'}\leqq\tfrac{1}{3}\mathsf{E}\alpha_{k'}(n).\tag{2.12}$$

First we show that (2.11) and (2.12) imply Part b of Theorem 2. Indeed, since

$$\mathsf{E}\alpha_{k'}(n)=\Sigma q_m\binom{m}{k'}$$

they imply that

$$\sum_{K\log^2 n<m<\bar{K}\log^2 n} q_m\binom{m}{k'}\geqq\tfrac{1}{3}\mathsf{E}\alpha_{k'}(n),$$

hence

$$P(\alpha(n)\geqq K\log^2 n)\geqq\sum_{K\log^2 n<m<\bar{K}\log^2 n} q_m$$

$$\geqq\frac{1}{\binom{\bar{K}\log^2 n}{k'}}\sum_{K\log^2 n<m<\bar{K}\log^2 n} q_m\binom{m}{k'}\geqq\frac{1}{3}\frac{\alpha_{k'}(n)}{\binom{\bar{K}\log^2 n}{k'}}.$$

Now, by Lemma 4 and the Stirling formula

$$P(\alpha(n)>K\log^2 n)\geqq\frac{1}{3}\frac{[(L-\eta)\log n]^{k'}}{(\bar{K}\log^2 n)^{k'}}\cdot\left(\frac{K(1+\delta)}{eL}\log n\right)^{k'}$$

$$=\frac{1}{3}\left(\frac{1}{e}\frac{L-\eta}{L}\frac{1+\delta}{1+2\delta}\right)^{k'}\geqq n^{-\frac{K}{L}-\varepsilon}$$

if η is chosen sufficiently small.

In order to prove (2.12) introduce $k''=\frac{\bar{K}}{L}\log n=k(1+2\delta)$. Since $k''>k'$ we can write

$$I_1=\sum_{m>\bar{K}\log^2 n} q_m\binom{m}{k'}=\sum_{m>\bar{K}\log^2 n} q_m\frac{\binom{m}{k'}}{\binom{m}{k''}}\binom{m}{k''}$$

$$\leqq\frac{\binom{\bar{K}\log^2 n}{k'}}{\binom{\bar{K}\log^2 n}{k''}}\sum_{m>\bar{K}\log^2 n} q_m\binom{m}{k''}\leqq\frac{\binom{\bar{K}\log^2 n}{k'}}{\binom{\bar{K}\log^2 n}{k''}}\mathsf{E}\,\alpha_{k''}(n).$$

Now Lemma 4 and the Stirling formula imply that

$$I_1\leqq\frac{(\bar{K}\log^2 n)^{k'-k''}}{\left(\frac{K}{L}(1+\delta)\log n\right)^{k'}}\left(\frac{\bar{K}}{L}\log n\right)^{k''}e^{k'-k''}[L(1+\delta^3)\log n]^{k''}$$

$$=(L\log n)^{k'}\left(\frac{1+2\delta}{1+\delta}\right)^{k'}e^{k'-k''}(1+\delta^3)^{k''}.$$

A simple Taylor expansion up to the second term implies that

$$e^{k'-k''}\left(\frac{1+2\delta}{1+\delta}\right)^{k'}=\exp\{k[-\delta+(1+\delta)(\delta-\tfrac{3}{2}\delta^2+O(\delta^3))]\}$$

$$=\exp\left\{-k\left(\frac{\delta^2}{2}+O(\delta^3)\right)\right\}=\exp\left\{-\log n\left(\frac{K}{2L}\delta^2+O(\delta^3)\right)\right\},$$

and

$$I_1\leqq(L\log n)^{k'}n^{-\frac{K}{2L}\delta^2+O(\delta^3)}\leqq\tfrac{1}{3}\mathsf{E}\,\alpha_{k'}(n).$$

The proof of (2.11) is similar. Since $k'>k$

$$I_2<\sum_{m<K\log^2 n} q_m\binom{m}{k'}\leqq\frac{\binom{K\log^2 n}{k'}}{\binom{K\log^2 n}{k}}\sum_{m<K\log^2 n} q_m\binom{m}{k}\leqq\frac{\binom{K\log^2 n}{k'}}{\binom{K\log^2 n}{k}}\mathsf{E}\,\alpha_k(n).$$

Hence by Lemma 4

$$I_2\leqq\frac{(K\log^2 n)^{k'-k}}{\left(\frac{K}{L}(1+\delta)\log n\right)^{k'}}\left(\frac{K}{L}\log n\right)^{k}e^{k'-k}(L(1+\delta^3)\log n)^k$$

$$\leqq(L\log n)^{k'}(1+\delta)^{-k'}e^{k'-k}(1+\delta^3)^k,$$

and since

$$(1+\delta)^{-k'} e^{k'-k} = \exp\left\{\log n\left[\frac{K}{L}\,\delta - \frac{K}{L}\,(1+\delta)\left(\delta - \frac{\delta^2}{2} + O(\delta^3)\right)\right]\right\}$$

$$= \exp\left\{\log n\left(-\frac{K}{2L}\,\delta^2 + O(\delta^3)\right)\right\}$$

$$I_2 \leqq (L\log n)^{k'} \cdot n^{-\frac{K}{2L}\delta^2 + O(\delta^3)} \leqq \tfrac{1}{3}\,\mathsf{E}\,\alpha_{k'}(n).$$

Theorem 2 is proved.

3. The Proof of Theorem 1 and Comments

Proof of Theorem 1. By Theorem 2

$$\sum_{n=1}^{\infty} P(\alpha(n) > (L+\varepsilon)\log^2 n) < \infty$$

for arbitrary $\varepsilon > 0$. Hence the Borel-Cantelli lemma implies that

$$\limsup \frac{\alpha(n)}{\log^2 n} \leqq L \quad \text{with probability 1.}$$

In order to make an estimate from below introduce the stopping times M_n, $n = 1, 2, \ldots$

$$M_n = \{\inf j,\ S_j \geqq n\}.$$

Then

$$\lim_{n\to\infty} \frac{M_n}{n^4} = 0 \quad \text{with probability 1,} \tag{3.1}$$

and

$$M_{(n+1)^2} - M_{n^2} \geqq 2n+1.$$

Let us define the event $A_k = A_k(\varepsilon)$, $k = 1, 2, \ldots$

$A_k = \{$the random walk $0, S_{M_{k^2}+1} - S_{M_{k^2}}, \ldots, S_{M_{k^2}+k} - S_{M_{k^2}}$ visits at least $(L-\varepsilon)\log^2 k$ points with positive coordinates exactly once$\}$.

The events A_k are independent, and it follows from Theorem 2 that for large $k\, P_k(A_k) > k^{-1+\frac{\varepsilon}{2}}$. Hence the Borel-Cantelli lemma implies that infinitely many events A_k occur with probability one. This implies that

$$\alpha(M_{k^2} + k) > (L-\varepsilon)\log^2 k \quad \text{for infinitely many } k \quad \text{with probability 1} \tag{3.2}$$

On the other hand it follows from (3.1) that $M_k^2 < k^8$ for all sufficiently large k. Thus (3.2) implies that

$$\frac{\alpha(n)}{\log^2 n} > \frac{L-\varepsilon}{64} \quad \text{for infinitely many } n \quad \text{with probability 1.}$$

We have shown that

$$P\left(\frac{L}{64}\leqq \limsup \frac{\alpha(n)}{\log^2 n}\leqq L\right)=1 \tag{3.3}$$

Let $\alpha'(n)$ denote the number of points visited exactly once by the random walk $S_{\log n}, S_{1+\log n}, \ldots, S_n$. Then $|\alpha(n)-\alpha'(n)|\leqq \log n$, hence

$$P\left(\limsup \frac{\alpha(n)}{\log^2 n}=\limsup \frac{\alpha'(n)}{\log^2 n}\right)=1.$$

On the other hand the zero- one law can be applied for $\frac{\alpha'(n)}{\log^2 n}$, and this implies that

$$P\left(\limsup \frac{\alpha'(n)}{\log^2 n}=C\right)=P\left(\limsup \frac{\alpha(n)}{\log^2 n}=C\right)=1$$

with some $0\leqq C\leqq\infty$. But it follows from (3.3) that $\frac{L}{64}\leqq C\leqq L$. Theorem 1 is proved.

Remark 1. There is some constant $C>0$ (indepent of n) such that

$$P(\alpha(n)\geqq k)\leqq\frac{C}{k} \quad \text{for all } k=1,2,\ldots \tag{3.4}$$

Relation (3.4) immediately follows from the inequality

$$\mathsf{E}\alpha(n)\leqq C.$$

But

$$\mathsf{E}\alpha(n)=\sum_{j=1}^{n} P(S_j\in\mathscr{A}_n)=4\sum_{j=1}^{n} P(D_1(j))\,P(D_2(j))\leqq C$$

by relations (2.6) and (2.6)'. (Some calculation would show that for arbitrary $\varepsilon>0$ $C=2+\varepsilon$ can be chosen in (3.4) if $n>n(\varepsilon)$.)

Remark 2. It is not difficult to see that

$$\lim_{n\to\infty} P(\sup_{k\leqq n} S_k\in\mathscr{A}_n)=\tfrac{1}{2} \tag{3.5}$$

and

$$\lim_{n\to\infty} P(\inf_{k\leqq n} S_k\in\mathscr{A}_n)=\tfrac{1}{2}. \tag{3.5'}$$

Indeed, let $j=j(n)$ denote the first place of maximum of the random walk $S_1, S_2, \ldots, S_n$. Then we have

$$\begin{aligned}&P(\sup_{k\leqq n} S_k\in\mathscr{A}_n\,|\,\sup_{k\leqq n} S_k=S_j,\ S_j>S_l \text{ for } l<j)\\&=P(S_{j+1}-S_j<0,\ldots,S_n-S_j<0\,|\,S_{j+1}-S_j\leqq 0,\ldots,S_n-S_j\leqq 0)\to\tfrac{1}{2}\end{aligned} \tag{3.6}$$

if $n-j\to\infty$. (See, e.g. formula (3.5) in Chap. III of [3]. We cannot write identity at the end of formula (3.6), because $n-j$ can be odd.) Since $\lim_{n\to\infty} P(n-j(n)\to\infty)=1$, hence (3.6) implies (3.5). The proof of (3.5)′ is the same. Relations (3.5) and (3.5)′ mean that $\sup_{k\leq n} S_k$ and $\inf_{k\leq n} S_k$ belong to $\mathscr{A}_n$ with positive probability. On the other hand we cannot decide whether the probability of the event that $\mathscr{A}_n$ contains points different from the maximum and minimum tends to zero or not.

Remark 3. It would be interesting to know whether our results hold (with possibly different constants) for more general random walks. We think that the answer is in the affirmative, but our proofs exploited the simple geometric structure of this random walk.

Remark 4. Let $\alpha(p, n)$ denote the number of points visited exactly p times by the random walk $S_0, S_1, \ldots, S_n$. Erdös and Révész also posed the following question: Let $p(n)$ be some "nice" function. For which function $f(n)$ is $\limsup \frac{\alpha(p(n), n)}{f(n)}=C$ with a non-trivial constant C, $0<C<\infty$? In the case $p(n)=1$ $f(n)=\log^2 n$. The analogue of Theorems 1 and 2 can be proved for $p(n)=\text{const.}$ with some extrawork, but actually in the same way as for $p(n)=1$. The solution of the problem for $p(n)\to\infty$ requires some new ideas. The natural approach would be again to give a good estimate on the probability $P(\alpha(p(n), n)>x_n)$ for certain x_n, but this seems to be a rather hard problem. We expect that the cases $p(n)<\text{const}\sqrt{n}$ and $\frac{p(n)}{\sqrt{n}}\to\infty$ are essentially different, because typically a point is visited up to time n less than const. $\sqrt{n}$ times.

Remark 5. Let $L(t, x)$ denote the local time of a Wiener process at the point x and at the time t. It would be interesting to know whether the results analogous to Theorems 1 and 2 are also valid for the local time of a Wiener process, i.e. whether the relations

$$P\left(\limsup_{t} \frac{\lambda\{x;\alpha\leqq L(t, x)\leqq 1\}}{\log^2 t}=C\right)=1$$

and

$$t^{-\frac{K}{L}-\varepsilon}\leqq P(\lambda\{x, \alpha\leqq L(t, x)\leqq 1\}>K\log^2 t)\leqq t^{-\frac{K}{L}+\varepsilon}$$

hold for sufficiently large t for all $0<\alpha<1$ with some appropriate $0<C<\infty$ and $0<L<\infty$, where $\lambda\{\cdot\}$ denotes the Lebesque measure.

Remark 6. One would like to know the explicit value of the constant L appearing in Theorem 2 and Lemma 3. It can be expressed with the help of the proof of Lemma 3 as a complicated integral which we cannot handle. On the other hand E. Csáki explained an argument to the author which shows that $L=\frac{1}{4}$. Define

$$p_{2k,2n}=P(S_0=0,\ S_{2n}=2k,\ 0<S_j<2k \text{ for } 1\leqq j\leqq 2n-1)$$

and

$$p_{2n}=\sum_{k=0}^{\infty} p_{2k,2n}$$

Then, by Lemma 3

$$p_{2n}=\frac{L}{2n}(1+o(1)) \tag{3.7}$$

On the other hand it is proved (see, e.g. formulas (3.6) and (3.7 in [2])) that for all $k\geqq 1$

$$\sum_{n=1}^{\infty} p_{2k,2n} z^k=\frac{1-w}{1+w}\frac{w^k}{1-w^{2k}} \quad \text{for } 0\leqq z<1 \text{ with } w=\frac{1-\sqrt{1-z^2}}{1+\sqrt{1-z^2}}.$$

Summing up these identities for all $k\geqq 1$ we get that for

$$f(z)=\sum_{n=1}^{\infty} p_{2n} z^n, \quad f(z)=\frac{1-w}{1+w}\sum_{k=1}^{\infty}\frac{w^k}{1-w^{2k}}, \quad 0\leqq z<1 \tag{3.8}$$

It follows from (3.7) that

$$\lim_{z\to 1}\frac{f(z)}{\log(1-z^2)}=-\frac{L}{2} \tag{3.9}$$

On the other hand we show with the help of (3.8) that

$$\lim_{z\to 1}\frac{f(z)}{\log(1-z^2)}=-\frac{1}{8} \tag{3.9$'$}$$

A comparison of (3.9) and (3.9)′ yields that $L=\frac{1}{4}$. We can write

$$f(z)=\frac{1}{1+w}\sum_{k=1}^{\infty}\frac{w^k}{2k}+\frac{1}{1+w}\sum_{k=1}^{\infty}\left(\frac{1-w}{1-w^{2k}}-\frac{1}{2k}\right)w^k=f_1(z)+f_2(z)$$

Then

$$f_1(z)=-\frac{1}{2(1+w)}\log(1-w)=-\frac{1}{2(1+w)}\log\frac{2\sqrt{1-z^2}}{1+\sqrt{1+z^2}}$$

$$=-\frac{1}{2(1+w)}\left(\tfrac{1}{2}\log(1-z^2)-\log(1+\sqrt{1+z^2})\right),$$

hence

$$\lim_{z\to 1}\frac{f_1(z)}{\log(1-z^2)}=-\frac{1}{8}.$$

On the other hand for $z<1$

$$\left|\frac{1-w}{1-w^{2k}}-\frac{1}{2k}\right|=\left|\frac{(w^{2k-1}-1)+(w^{2k-2}-1)+\ldots+(w-1)}{2k(1+w+\ldots+w^{2k-1})}\right|\leqq 1-w,$$

hence

$$|f_2(z)| \leqq (1-w) \sum_{k=1}^{\infty} w^k < 1,$$

and

$$\lim_{z \to 1} \frac{f_2(z)}{\log(1-z^2)} = 0.$$

These relations imply (3.9).

References

1. Csáki, E., Mohanty, S.G.: Excursion and meander in random walk. Can. J. Stat. **9**, 57–70 (1981)
2. Csáki, E., Mohanty, S.G.: Some joint distributions for conditional random walks. Can. J. Stat. **14**, 19–28 (1986)
3. Feller, W.: An introduction to probability theory and its applications, vol. 1. New York Chichester Brisbane Toronto: Wiley 1970

Received October 1, 1986; received in revised form August 31, 1987