The Annals of Mathematical Statistics
1972, Vol. 43, No. 5, 1692-1697

THE LAW OF THE ITERATED LOGARITHM FOR THE RANGE OF RANDOM WALK[1]

BY NARESH C. JAIN AND WILLIAM E. PRUITT

University of Minnesota

Let $\{X_n, n \geqq 1\}$ be a sequence of independent identically distributed random variables taking values in the d-dimensional integer lattice E_d, and let $S_0 = 0$, $S_n = X_1 + \cdots + X_n$. The range of the random walk $\{S_n, n \geqq 0\}$ up to time n, denoted by R_n, is the number of distinct lattice points visited by the random walk up to time n. Let p be the probability that the random walk never returns to the origin. It is known that $n^{-1}R_n \to p$ a.s. and that for $p < 1$ if the genuine dimension is $d \geqq 4$ or if the random walk is strongly transient then there is a positive constant σ^2 such that $\operatorname{Var} R_n \sim \sigma^2 n$. In the present note we shall prove that in these two cases

$$\limsup_{n\to\infty} \frac{R_n - np}{(2\sigma^2 n \log\log n)^{\frac{1}{2}}} = 1 \quad \text{a.s.}$$

and the lim inf of the same sequence is almost surely -1.

1. Introduction.

Let $\{X_n, n \geqq 1\}$ be a sequence of independent identically distributed random variables, defined on a probability space $(\Omega, \mathscr{F}, P)$, and taking values in the d-dimensional integer lattice E_d. The sequence $\{S_n, n \geqq 0\}$ defined by $S_0 = 0$, $S_n = \sum_{k=1}^n X_k$ is called a random walk. Let $p = P[S_1 \neq 0, S_2 \neq 0, \cdots]$. The random walk is called transient if $p > 0$ or, equivalently, if $\sum_{n=1}^\infty P[S_n = 0]$ converges. It is called strongly transient if $\sum_{n=1}^\infty \sum_{k=n}^\infty P[S_k = 0]$ converges. The range of the random walk up to time n, denoted by R_n, is the cardinality of the set $\{S_0, S_1, \cdots, S_n\}$. Kesten, Spitzer, and Whitman [9] proved that $n^{-1}R_n \to p$ a.s. for any random walk. Note that if $p = 1$, then $R_n = n + 1$ a.s. and this case is not very interesting. In [3] and [5] it was shown that for $p < 1$, if the random walk is strongly transient or if the random walk has genuine dimension $d \geqq 4$, then there is a positive constant σ^2 such that $\operatorname{Var} R_n \sim \sigma^2 n$. In the present note we obtain the law of the iterated logarithm for R_n in these two cases.

THEOREM. *If $p < 1$ and the random walk is either strongly transient or has genuine dimension $d \geqq 4$, then there is a positive constant σ^2 such that*

$$\limsup_{n\to\infty} \frac{R_n' - np}{(2\sigma^2 n \log\log n)^{\frac{1}{2}}} = 1 \quad \text{a.s.}$$

and the lim inf *of this sequence is* -1 a.s.

REMARK. If $d \geqq 4$, then it will be clear from our proof of the above theorem that actually we have proved more than just the law of the iterated logarithm. For example, referring to the terminology of Feller [2], we can say that

$$\varphi_n = \{2\log_2 n + 3\log_3 n + 2\log_4 n + \cdots + 2\log_{\rho-1} n + (2+\delta)\log_\rho n\}^{\frac{1}{2}}$$

Received July 19, 1971.

[1] This work was partially supported by the National Science Foundation.

belongs to the upper (lower) class if $\delta > 0$ ($\delta < 0$), where $\log_\rho n$ stands for the iterated logarithm of order ρ. However, we have not proved this stronger result in the strongly transient case with $d \leqq 3$.

The proof for the case $d \geqq 4$ is slightly easier and is given in Section 2. The proof for the strongly transient case is in Section 3.

The first work on the range of random walk was by Dvoretzky and Erdös [1]. Subsequent related work not mentioned above appears in [4], [6], and [7]. Some criteria for strong transience are also given in [6].

2. $d \geqq 4$. We start by introducing a little notation. T_x will denote the first hitting time of the lattice point x, i.e.,

$$T_x = \min\{k \geqq 1 : S_k = x\} ;$$

if there are no integers k with $S_k = x$, then $T_x = \infty$. The transition probabilities are $P^n(x, y) = P_x[S_n = y]$ and the taboo probabilities are $P_z{}^n(x, y) = P_x[S_n = y, T_z \geqq n]$. We will use

$$u_n = P^n(0, 0) , \qquad f_n = P_0{}^n(0, 0) , \qquad r_n = \textstyle\sum_{k=n+1}^{\infty} f_k .$$

Note that $\sum_{k=1}^{\infty} f_k = P[T_0 < \infty] = 1 - p$. It is known [9] that $u_n = O(n^{-d/2})$ and it follows that for $d \geqq 3$, $r_n = O(n^{1-d/2})$.

Form a sequence $\{n_i\}$ of positive integers by taking for $k = 1, 2, \cdots$ all integers in the interval $[2^{2k}, 2^{2k+2})$ which are of the form $2^{2k} + j\{[k^{-1}2^k] + 1\}$ where j is a nonnegative integer. There will be at most $3k2^k$ members of the sequence in the interval $[2^{2k}, 2^{2k+2})$. Let $n_0 = 0$ and for $j \geqq 0$ define

U_j = number of distinct lattice points visited in $(n_j, n_{j+1}]$,

V_j = number of distinct lattice points visited in both $(n_j, n_{j+1}]$ and $[0, n_j]$.

We then have $R_{n_i} = 1 + \sum_{j=0}^{i-1} U_j - \sum_{j=0}^{i-1} V_j$ and so

$$\frac{R_{n_i} - ER_{n_i}}{(2\sigma^2 n_i \log\log n_i)^{\frac{1}{2}}} = \frac{\sum_{j=0}^{i-1} (U_j - EU_j)}{(2\sigma^2 n_i \log\log n_i)^{\frac{1}{2}}} - \frac{\sum_{j=0}^{i-1} (V_j - EV_j)}{(2\sigma^2 n_i \log\log n_i)^{\frac{1}{2}}} . \tag{2.1}$$

It will be enough to prove that the sequence in (2.1) has lim sup (lim inf) equal to 1 (-1) a.s. since if $n_i \leqq n < n_{i+1}$, then $n \sim n_i$ so the normalization is right, and

$$|R_n - ER_n - R_{n_i} + ER_{n_i}| \leqq n - n_i = o(n_i^{\frac{1}{2}}) .$$

We also need to see that ER_n can be replaced by np but this is valid since

$$ER_n = 1 + np + \textstyle\sum_{j=1}^{n} r_j = np + O(\log n) = np + o(n^{\frac{1}{2}}) .$$

To prove that the sequence in (2.1) has the right behavior, we will show that this is the case for the sequence involving the U's while the sequence involving the V's tends to zero a.s. The behavior of the U's follows immediately from Kolmogorov's law of the iterated logarithm ([8] page 260) since the U_j are independent,

$$\operatorname{Var} \textstyle\sum_{j=0}^{i-1} U_j \sim \sum_{j=0}^{i-1} \sigma^2(n_{j+1} - n_j) = \sigma^2 n_i ,$$

and for $2^{2k} \leqq n_i < 2^{2k+2}$,

$$|U_i - EU_i| \leqq n_{i+1} - n_i \leqq k^{-1}2^k + 1 = o((n_i/\log\log n_i)^{\frac{1}{2}}) .$$

The V's cause a little more difficulty and we need to estimate $\operatorname{Var}(\sum_{j=0}^{i-1} V_j)$. For $m \leqq n < \alpha$, let

$$W(\alpha; n, m) = I[S_\alpha \neq S_{\alpha-1}, \cdots, S_\alpha \neq S_{n+1}; S_\alpha = S_\gamma \text{ for some } \gamma \in [m, n]] ,$$

where $I(A)$ denotes the indicator function of A. Note that $V_j = \sum_{\alpha=n_j+1}^{n_{j+1}} W(\alpha; n_j, 0)$. Now suppose that $n_m < \beta \leqq n_{m+1}$, $n_j < \alpha \leqq n_{j+1}$ with $m < j$ and write

$$W(\alpha; n_j, 0) = W(\alpha; n_j, \beta) + W(\alpha; \beta - 1, 0) .$$

Then $W(\beta; n_m, 0)$ and $W(\alpha; n_j, \beta)$ are independent so that

$$\begin{aligned} \operatorname{Cov}(W(\beta; n_m, 0), W(\alpha; n_j, 0)) &= \operatorname{Cov}(W(\beta; n_m, 0), W(\alpha; \beta - 1, 0)) \\ &\leqq EW(\beta; n_m, 0) W(\alpha; \beta - 1, 0) \\ &\leqq \sum_x P^{\alpha-\beta}(0, x) P_x[\beta - n_m \leqq T_x \leqq \beta; T_0 \leqq \beta] , \end{aligned}$$

the final bound being obtained by reversing the random walk. This bound follows trivially if $m = j$ with $\beta \leqq \alpha$. If we fix β and sum over all α in $[\beta, n_i]$, where j will vary with α, we obtain a bound of $c(\beta - n_m)^{-1} \log n_i$ by Lemma 4 of [5] since $d \geqq 4$. The β summation is now carried out by summing over $\beta \in (n_m, n_{m+1}]$ which gives a bound of $c \log^2 n_i$ and then multiplying by the number of possible values of m. This leads to

$$\operatorname{Var}(\sum_{j=0}^{i-1} V_j) \leqq ci \log^2 n_i .$$

Suppose now that $2^{2k} \leqq n_i < 2^{2k+2}$. Then

$$i \log^2 n_i \leqq (2k + 2)^2 \sum_{j=1}^{k} 3j2^j = O(k^3 2^k) ,$$

and so,

$$P[|\sum_{j=0}^{i-1} (V_j - EV_j)| \geqq \varepsilon n_i^{\frac{1}{2}}] = O(k^3 2^k n_i^{-1}) = O(k^3 2^{-k}) . \tag{2.2}$$

Now form a subsequence by taking every k^6th member of $\{n_i\}$ in $[2^{2k}, 2^{2k+2})$ and denote it by $\{n_{\nu_i}\}$. There will be at most $O(k^{-5}2^k)$ members of the subsequence in $[2^{2k}, 2^{2k+2})$ and multiplying the probability estimate in (2.2) by this number of terms still gives a convergent series. This means that we have the desired convergence to zero a.s. of the V part but only along the subsequence. However, we shall see that this implies convergence to zero along the original sequence. If $2^{2k} \leqq n_{\nu_m} \leqq n_i < n_{\nu_{m+1}} \leqq 2^{2k+2}$, then

$$\begin{aligned} \sum_{j=0}^{\nu_m} (V_j - EV_j) - \sum_{j=\nu_m+1}^{i} EV_j &\leqq \sum_{j=0}^{i} (V_j - EV_j) \\ &\leqq \sum_{j=0}^{\nu_{m+1}} (V_j - EV_j) + \sum_{j=i+1}^{\nu_{m+1}} EV_j \end{aligned}$$

so that we only need a bound for EV_j. But for $n_j \leqq 2^{2k+2}$,

$$EV_j \leqq \sum_{\alpha=0}^{n_{j+1}-n_j} r_\alpha = O(\log(n_{j+1} - n_j)) = O(k) ,$$

while both $\nu_{m+1} - i$ and $i - \nu_m$ can be no larger than k^6. Thus

$$\max(\sum_{j=\nu_m+1}^{i} EV_j, \sum_{j=i+1}^{\nu_{m+1}} EV_j) = O(k^7) = o(2^k)$$

and we have convergence along the original sequence as well.

3. Strongly transient random walk. In this case the sequence $\{n_i\}$ is formed by taking for $k = 3, 4, \cdots$ all integers in $[2^{2k}, 2^{2k+2})$ of the form $2^{2k} + j\xi_k$ with j a nonnegative integer and $\xi_k = [2^k(\log k)^{-\frac{3}{4}}] + 1$. The number of members of the sequence which are in $[2^{2k}, 2^{2k+2})$ will be $O((\log k)^{\frac{3}{4}}2^k)$. U_j and V_j are defined as in Section 2 and the Kolmogorov law of the iterated logarithm again applies to the sequence composed of the U's. It also suffices to look just at the behavior on the sequence $\{n_i\}$ as before. In this case,

$$ER_n = 1 + np + \sum_{j=1}^{n} r_j = np + O(1)$$

for strong transience implies the convergence of $\sum r_j$. Thus we only need to show that the sequence involving the V's in (2.1) converges to zero a.s. To do this we write $V_j = Y_j + Z_j$ where

$Y_j =$ number of distinct lattice points hit in both $(n_j, n_{j+1}]$ and $[n_j - \gamma_j, n_j]$,

$Z_j =$ number of distinct lattice points hit in both $(n_j, n_{j+1}]$ and $[0, n_j - \gamma_j)$ but not in $[n_j - \gamma_j, n_j]$,

and if $2^{2k} \leqq n_j < 2^{2k+2}$ then $\gamma_j = \eta_k = 2^k(\log k)^{-\frac{1}{2}}$. It will suffice to show that

$$\frac{\sum_{j=0}^{i-1}(Y_j - EY_j)}{(n_i \log\log n_i)^{\frac{1}{2}}} \to 0 \text{ a.s.} \quad \text{and} \quad \frac{\sum_{j=1}^{i-1}(Z_j - EZ_j)}{(n_i \log\log n_i)^{\frac{1}{2}}} \to 0 \text{ a.s.} \tag{3.1}$$

We consider the Z's first. Let

$$H_j = \sum_{\alpha,\beta\in D_j} I[S_\alpha = S_\beta]$$

where $D_j = \{(\alpha, \beta): 2^{2j} \leqq \beta \leqq 2^{2j+2}, \beta - \alpha \geqq \eta_j\}$. Then if $2^{2k} \leqq n_i < 2^{2k+2}$, it follows that $\sum_{j=1}^{i-1} Z_j \leqq \sum_{j=3}^{k} H_j$, and so

$$\begin{aligned} &P[\textstyle\sum_{j=1}^{i-1} Z_j \geqq \varepsilon 2^k(\log k)^{\frac{1}{2}} \text{ for some } i \text{ with } n_i \in [2^{2k}, 2^{2k+2})] \\ &\quad \leqq P[\textstyle\sum_{j=3}^{k} H_j \geqq \varepsilon 2^k(\log k)^{\frac{1}{2}}] \leqq \varepsilon^{-1}2^{-k}(\log k)^{-\frac{1}{2}} \sum_{j=3}^{k} EH_j\,. \end{aligned}$$

It will be sufficient now to show that this last bound is summable on k since it also dominates $2^{-k}(\log k)^{-\frac{1}{2}} \sum_{j=1}^{i-1} EZ_j$. Now

$$\begin{aligned} \sum_{k=3}^{\infty} 2^{-k}(\log k)^{-\frac{1}{2}} \sum_{j=3}^{k} EH_j &\leqq \sum_{k=3}^{\infty} 2^{-k}(\log k)^{-\frac{1}{2}} \sum_{j=3}^{k} 2^{2j+2} \sum_{\eta\geqq\eta_j} u_\eta \\ &\leqq 8 \sum_{j=3}^{\infty} 2^j(\log j)^{-\frac{1}{2}} \sum_{m=j}^{\infty} \sum_{\eta_m\leqq\eta<\eta_{m+1}} u_\eta \\ &\leqq c \sum_{m=3}^{\infty} 2^m(\log m)^{-\frac{1}{2}} \sum_{\eta_m\leqq\eta<\eta_{m+1}} u_\eta \\ &\leqq c \sum \eta u_\eta < \infty \end{aligned}$$

since strong transience is equivalent to the convergence of $\sum ju_j$. In order to prove the first statement in (3.1) we need to estimate $\mathrm{Var}(\sum_{j=0}^{i-1} Y_j)$. Note that $Y_j = \sum_{\alpha=n_j+1}^{n_{j+1}} W(\alpha; n_j, n_j - \gamma_j)$. Suppose that $n_m < \beta \leqq n_{m+1}$, $n_j < \alpha \leqq n_{j+1}$, with $m < j$. If $\beta \leqq n_j - \gamma_j$, then $W(\alpha; n_j, n_j - \gamma_j)$ and $W(\beta; n_m, n_m - \gamma_m)$ are independent. If $\beta > n_j - \gamma_j$, we write

$$W(\alpha; n_j, n_j - \gamma_j) = W(\alpha; n_j, \beta) + W(\alpha; \beta - 1, n_j - \gamma_j)$$

and then

$$\begin{aligned}\operatorname{Cov}(W(\beta; n_m, n_m - \gamma_m), W(\alpha; n_j, n_j - \gamma_j)) \\ &\leqq EW(\beta; n_m, n_m - \gamma_m)W(\alpha; \beta - 1, n_j - \gamma_j) \\ &\leqq EW(\alpha; \beta - 1, n_j - \gamma_j) \leqq r_{\alpha-\beta} .\end{aligned}$$

If $m = j$ and $\beta \leqq \alpha$ we use the bound

$$\operatorname{Cov}(W(\beta; n_j, n_j - \gamma_j), W(\alpha; n_j, n_j - \gamma_j)) \leqq EW(\alpha; n_j, n_j - \gamma_j) \leqq r_{\alpha - n_j - 1} .$$

Now fix α and sum these covariances over all $\beta \leqq \alpha$ to obtain a bound of

$$\sum_{\gamma=\alpha-n_j}^{\alpha-n_j+\gamma_j} r_\gamma + (\alpha - n_j) r_{\alpha-n_j-1} .$$

Next we sum over all $\alpha \in (n_j, n_{j+1}]$ which leads to a bound of

$$\sum_{\gamma=1}^{2\gamma_j} \gamma r_\gamma + \sum_{\gamma=1}^{n_{j+1}-n_j} \gamma r_{\gamma-1} = O(\sum_{\gamma=1}^{2\gamma_j} \gamma r_\gamma) ,$$

and multiplying by the number of possible values of j, we obtain the estimate

$$\operatorname{Var}(\sum_{j=0}^{i-1} Y_j) = O(i \sum_{\gamma=1}^{2\gamma_i} \gamma r_\gamma) .$$

Then for $2^{2k} \leqq n_i < 2^{2k+2}$,

$$P[|\sum_{j=0}^{i-1} (Y_j - EY_j)| \geqq \varepsilon 2^k (\log k)^{\frac{1}{2}}] \leqq c \frac{2^k (\log k)^{\frac{3}{4}} \sum_{\gamma=1}^{2\eta_k} \gamma r_\gamma}{2^{2k} \log k} . \tag{3.2}$$

As in Section 2 we must now consider a subsequence of the $\{n_i\}$ sequence by taking every $[2^k(\log k)^{\frac{1}{4}}]$th member of $\{n_i\}$ in $[2^{2k}, 2^{2k+2})$. There will be $O((\log k)^{\frac{1}{4}})$ members of the subsequence in $[2^{2k}, 2^{2k+2})$. We multiply this number of terms by the probability estimate in (3.2) and we want to show that the resulting series converges. To see this,

$$\begin{aligned}\sum_{k=3}^\infty 2^{-k} (\log k)^{\frac{1}{4}} \sum_{\gamma=1}^{2\eta_k} \gamma r_\gamma &= \sum_{k=3}^\infty 2^{-k} (\log k)^{\frac{1}{4}} \sum_{m=3}^k \sum_{2\eta_{m-1} < \gamma \leqq 2\eta_m} \gamma r_\gamma \\ &\leqq c \sum_{m=3}^\infty 2^{-m} (\log m)^{\frac{1}{4}} \eta_m \sum_{2\eta_{m-1} < \gamma \leqq 2\eta_m} r_\gamma \\ &\leqq c \sum r_\gamma < \infty .\end{aligned}$$

This means that the first sequence in (3.1) tends to zero a.s. along the subsequence. The filling in can be done as in Section 2 since $EY_j \leqq \sum r_m = O(1)$ and the number of j between two successive members of the subsequence in $[2^{2k}, 2^{2k+2})$ is $[2^k(\log k)^{\frac{1}{4}}] = o(2^k(\log k)^{\frac{1}{2}})$.

REFERENCES

[1] DVORETZKY, A. and ERDÖS, P. (1951). Some problems on random walk in space. *Proc. Second Berkeley Symp. Math. Statist. Prob.* University of California Press 353–367.

[2] FELLER, W. (1943). The general form of the so-called law of the iterated logarithm. *Trans. Amer. Math. Soc.* **54** 373–402.

[3] JAIN, N. C. and OREY, S. (1968). On the range of random walk. *Israel J. Math.* **6** 373–380.

[4] JAIN, N. C. and PRUITT, W. E. (1970). The range of recurrent random walk in the plane. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **16** 279–292.

[5] JAIN, N. C. and PRUITT, W. E. (1971). The range of transient random walk. *J. Analyse Math.* **24** 369–393.

[6] JAIN, N. C. and PRUITT, W. E. (1972). The range of random walk. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* Univ. of California Press.

[7] JAIN, N. C. and PRUITT, W. E. (1972). Further limit theorems for the range of random walk. To appear in *J. Analyse Math.*

[8] LOÈVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton.

[9] SPITZER, F. (1964). *Principles of Random Walk*. Van Nostrand, Princeton.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455