Journal of Statistical Physics, Vol. 27, No. 2, 1982

On the Expected Number of Distinct Points in a Subset Visited by an N-Step Random Walk

George H. Weiss[1] **and Michael F. Shlesinger**[2]

Received March 10, 1981

Many investigators have calculated asymptotically valid expressions for the expected number of distinct points visited by an n-step random walk on a lattice. In this note we point out that the same formalism can be used to study the expected number of distinct points in a subset of lattice points. We also calculate the expected occupancy of the subset and give sufficient conditions for the ratio of the two calculated quantities to have the same asymptotic time dependence as for the full lattice. Specific examples are considered.

KEY WORDS: Random walks; occupancy; distinct site problems.

1. INTRODUCTION

Several investigators have calculated the expected number of distinct sites visited by an n-step random walk.[1–5] We will call the number of distinct sites visited the cardinality of the random walk. The expected cardinality will be denoted by $\langle S(n)\rangle$, where $S(n)$ is the random variable. The random walks of interest will be assumed to take place on a translationally invariant lattice. Results of these investigations have found application in different areas of metallurgy and chemical physics particularly relating to trapping phenomena.[6–18] The problem of finding the complete distribution of the cardinality involves considerably more sophisticated mathematics and results are not known in complete generality.[19–23] It seems not to have been remarked on that the formalism used to calculate $\langle S(n)\rangle$ for complete lattices can also be used for sublattices or for other sets of points. Random walks on binary lattices have been investigated by Brandt[24] and Argyrakis

[1] National Institutes of Health, Bethesda, Maryland 20205.

[2] Institute for Physical Sciences and Technology, University of Maryland, College Park, Maryland 20742, and La Jolla Institute, La Jolla, California 92038.

0022-4715/82/0200-0355$03.00/0

and Kopelman[12] in connection with energy migration in photosynthetic units. In this paper we examine the behavior of $S(n)$ for several different subsets of points.

2. FUNDAMENTAL QUANTITIES

Let the structure function of the random walk be denoted by

$$\lambda(\boldsymbol{\theta}) = \sum_{\mathbf{j}} p(\mathbf{j})\exp(i\mathbf{j}\cdot\boldsymbol{\theta}) \tag{1}$$

where $p(\mathbf{j})$ is the probability of a single step having displacement $\mathbf{j}$. Let T be the subset of points that is of interest, and Ω be the set of all lattice points. Then the expected cardinality on T of an n-step walk is equal to

$$\langle S_T(n)\rangle = \sum_{j=0}^{n} \sum_{\mathbf{s}\in T} F_j(\mathbf{s}) \tag{2}$$

where $F_j(\mathbf{s})$ is the probability that $\mathbf{s}$ is visited for the first time at step j. In order to find the asymptotic $(n\to\infty)$ behavior of $\langle S_T(n)\rangle$ it is convenient to introduce the generating functions

$$S(z;T) = \sum_{n=0}^{\infty} \langle S_T(n)\rangle z^n, \qquad F(\mathbf{s};z) = \sum_{n=0}^{\infty} F_n(\mathbf{s})z^n \tag{3}$$

If $P_n(\mathbf{s})$ is the probability of the random walker being at $\mathbf{s}$ at step n then it is known that

$$P(\mathbf{s};z) = \sum_{n=0}^{\infty} P_n(\mathbf{s})z^n = \frac{1}{(2\pi)^D}\int \overset{\pi}{\underset{-\pi}{\cdots}} \int \frac{\exp(-i\mathbf{s}\cdot\boldsymbol{\theta})}{1 - z\lambda(\boldsymbol{\theta})}\, d^D\boldsymbol{\theta} \tag{4}$$

where D is the number of dimensions. Furthermore, from the relation between $F(\mathbf{s};z)$ and $P(\mathbf{s};z)$ it follows that

$$S(z;T) = \sum_{\mathbf{s}\in T} P(\mathbf{s};z)/\left[(1-z)P(\mathbf{0};z)\right] \tag{5}$$

When $T = \Omega$ one can use the relation

$$\sum_{\mathbf{s}} P(\mathbf{s};z) = 1/(1-z) \tag{6}$$

to infer that $S(z;\Omega) = 1/[(1-z)^2 P(\mathbf{0};z)]$ as given by Montroll and Weiss.[4] A second interesting relation results from the observation that the expected number of visits to T by an n-step walk is

$$\langle V_T(n)\rangle = \sum_{\mathbf{s}\in T} \sum_{j=0}^{n} P_j(\mathbf{s}) \tag{7}$$

The quantity $V_T(n)$ will be termed the occupancy of T by an n-step random walk. The generating function of the $\langle V_T(n)\rangle$ is

$$V(z;T) = \sum_{\mathbf{s}\in T} P(\mathbf{s};z)/(1-z) \tag{8}$$

It follows from this that

$$S(z;T) = V(z;T)/P(\mathbf{0};z) \tag{9}$$

3. PERIODIC SETS

It is instructive to consider the special case of sets T whose elements are contained in a unit cell, i.e., the smallest translationally invariant unit of the lattice. For simplicity and without loss of generality, we carry out the calculation in one dimension. Let the period be denoted by N, and let $T = \{r_1, r_2, \ldots, r_k\}$ in the fundamental cell. To calculate either $V(z;T)$ or $S(z;T)$ we need the sum

$$\begin{aligned}\sum_{l=1}^{k}\sum_{j=-\infty}^{\infty} P(r_l + jN; z) &= \frac{1}{2\pi}\sum_{l=1}^{k}\sum_{j=-\infty}^{\infty}\int_{-\pi}^{\pi}\frac{\exp[-i(r_l+jN)\theta]}{1-z\lambda(\theta)}\,d\theta \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{B_T(\theta)\sum_{j=-\infty}^{\infty}\exp(-ijN\theta)}{1-z\lambda(\theta)}\,d\theta \end{aligned}\tag{10}$$

where

$$B_T(\theta) = \sum_{l=1}^{k}\exp(-ir_l\theta) \tag{11}$$

The sum in the last term of Eq. (10) is easily evaluating using the identity

$$\sum_{j=-\infty}^{\infty}\exp(-ijN\theta) = \sum_{j=-\infty}^{\infty}\delta\left(\frac{N\theta}{2\pi} - j\right) \tag{12}$$

The only term that contributes to the integral in Eq. (10) is the $j=0$ term so that the sum is equal to

$$\frac{k}{N(1-z)} = \frac{\rho}{1-z} \tag{13}$$

where ρ is the density of T points in the unit cell. If we let $N\to\infty$ in such a way that $\lim_{N\to\infty} k/N = \rho = \text{const}$, it is clear that

$$\lim_{N\to\infty}\left\langle \sum_{s\in T_N} P(s;z)\right\rangle = \rho\sum_{s} P(s;z) \tag{14}$$

This is valid for a periodic array of points. The opposite extreme is a random set T, i.e., a point belongs to T with probability c. The strong law of large numbers ensures that in the limit $N \to \infty$ k/N approaches c with probability 1 so that Eq. (14) remains valid with $\rho = c$. One can conceive of many sets T that are intermediate between strict periodicity and purely random for which Eq. (14) is valid. In addition Eq. (14) is also valid in any number of dimensions by a similar set of calculations.

4. APERIODIC SETS

Interesting results emerge when the set T has a nonuniform density. Consider, for example, the set T in one dimension consisting of all points of the form $\pm j^m$, $j = 0, 1, 2, \ldots$, $m = 1, 2, \ldots$, and a symmetric random walk [so that $\lambda(\theta) = \lambda(-\theta)$]. The sum appearing in Eq. (10) can then be written

$$\Gamma_T(z) = \sum_{s \in T} P(s; z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(1 + 2 \sum_{j=1}^{\infty} \cos j^m \theta \right) \frac{d\theta}{1 - z\lambda(\theta)} \tag{15}$$

where $\Gamma_T(z)$ is the indicated sum. The asymptotic evaluation of either the expected occupancy or the expected cardinality on T requires that we determine the analytic behavior of $\Gamma_T(z)$ near $z = 1$. We will assume that $\theta = 0$ is the only value of θ for which $\lambda(\theta) = 1$. For $\theta \sim 0$ we have

$$\begin{aligned} 1 + 2 \sum_{j=1}^{\infty} \cos j^m \theta &\sim 2 \int_0^{\infty} \cos(x^m \theta)\, dx \\ &= 2\Gamma\left(1 + \frac{1}{m}\right) \cos\left(\frac{\pi}{2m}\right) \Big/ |\theta|^{1/m} \end{aligned} \tag{16}$$

The singular behavior of $\Gamma_T(z)$ is determined by the behavior of the integrand in Eq. (15) in a neighborhood of $\theta = 0$, hence we want to represent the integrand accurately in this region. If we assume that the variance of the single step transition probability is finite,

$$\sum_{j=-\infty}^{\infty} j^2 p(j) = \sigma^2 < \infty \tag{17}$$

then, near $\theta = 0$,

$$\lambda(\theta) = 1 - \frac{\sigma^2 \theta^2}{2} + o(\theta^2) \tag{18}$$

for a symmetric random walk. It therefore follows that for z approximately equal to 1 the singular behavior of $\Gamma_T(z)$ is that of

$$\begin{aligned}\frac{2}{\pi}\Gamma\left(1+\frac{1}{m}\right)\cos\left(\frac{\pi}{2m}\right)\int_0^\infty \frac{1}{\theta^{1/m}}\,\frac{1}{1-z+\sigma^2\theta^2/2}\,d\theta \\ = \frac{2^{(1-1/m)/2}}{\sigma^{1-1/m}}\,\frac{\Gamma(1+1/m)}{(1-z)^{(1+1/m)/2}} = \frac{K}{(1-z)^{(1+1/m)/2}}\end{aligned} \tag{19}$$

where K lumps all of the constants shown. For $z \sim 1$ it is known that $P(0;z) \sim \{\sigma[2(1-z)]^{1/2}\}^{-1}$ so that

$$\begin{aligned} V(z;T) &\sim K(1-z)^{-(3+1/m)/2} \\ S(z;T) &\sim K\sigma\sqrt{2}\,(1-z)^{-(1+1/(2m))} \end{aligned} \tag{20}$$

A Tauberian theorem for power series[25] together with the fact that $\langle S_T(n)\rangle$ and $\langle V_T(n)\rangle$ are monotonic in n allows us to conclude that

$$\begin{aligned} \langle S_T(n)\rangle &\sim \frac{K\sigma\sqrt{2}}{\Gamma(1+1/(2m))}\,n^{1/(2m)} \\ \langle V_T(n)\rangle &\sim \frac{K}{\Gamma(3/2+1/(2m))}\,n^{(1+1/m)/2} \end{aligned} \tag{21}$$

If we relax the assumption of a finite variance for single step jump probabilities then other forms of n dependence can be obtained for both the expected occupancy and the expected cardinality.

5. EFFICIENCY IN SAMPLING OF DISTINCT POINTS

A function that gives some information on the efficiency with which the random walk samples the set T is

$$U_T(n) = S_T(n)/V_T(n) \tag{22}$$

Equation (21) allows us to calculate the quantity

$$\frac{\langle S_T(n)\rangle}{\langle V_T(n)\rangle} \sim \frac{\Gamma(3/2+1/(2m))}{\Gamma(1+1/(2m))}\left(\frac{2\sigma^2}{n}\right)^{1/2} \tag{23}$$

for the set defined, above which can be compared to the result for $T = \Omega$:

$$\frac{\langle S_\Omega(n)\rangle}{\langle V_\Omega(n)\rangle} \sim \left(\frac{8\sigma^2}{\pi n}\right)^{1/2} \tag{24}$$

The n dependence is the same in both cases, although the coefficients differ. It is tempting to conjecture that the n dependence of $\langle S_T(n)\rangle / \langle V_T(n)\rangle$ is independent of the set T. We have not succeeded in showing this to be the case but it is possible to state sufficient conditions for its validity. As an example, let $P(\mathbf{0}; z)$ and $\Gamma_T(z)$ have the form

$$P(\mathbf{0}; z) \sim \frac{1}{(1-z)^{\alpha}} L_1\left(\frac{1}{1-z}\right), \qquad \Gamma_T(z) \sim \frac{1}{(1-z)^{\beta}} L_2\left(\frac{1}{1-z}\right) \tag{25}$$

in a neighborhood of $z = 1$, where $L_1(x)$ and $L_2(x)$ are slowly varying functions at $x = \infty$ [i.e., $\lim_{x\to\infty} L_i(cx)/L(x) = 1$ for all $c > 0$]. Further, let the $L_i(x)$ satisfy

$$\frac{d \ln L_i(x)}{dx} \leqslant \frac{\text{const}}{x} \tag{26}$$

Under these conditions one can assert that the ratio in Eq. (23) does not depend, as a function of n, on the set T. Any multiplicative constants may, however, depend on T. It would be interesting to furnish conditions under which this last constraint does not itself apply, that is,

$$\langle S_T(n)\rangle / \langle V_T(n)\rangle = \langle S_\Omega(n)\rangle / \langle V_\Omega(n)\rangle \tag{27}$$

for all sets T. We note that it is possible for more complicated forms for $P(\mathbf{0}; z)$ and $\Gamma_T(z)$ to arise as shown by some recent work by Hughes, Shlesinger, and Montroll,[26] but the sufficient conditions in Eqs. (25) and (26) cover a variety of interesting cases. A second, and apparently much harder problem suggested by our analysis is to determine the relation between the asymptotics of $\langle U_T(n)\rangle$ and those of $\langle S_T(n)\rangle / \langle V_T(n)\rangle$. We conjecture that the asymptotic n dependence of these quantities is the same, but we have no evidence either for or against the conjecture.

6. EXAMPLES IN HIGHER DIMENSIONS

Several trivial generalizations of Eqs. (10) and (19) can be made. As an example, if we consider a Cartesian lattice in two dimensions, a symmetric random walk with the properties

$$\begin{gathered} p(j,k) = p(j,-k) = p(-j,k) \\ \sum_{j,k} j^2 p(j,k) = \sigma_1^2, \qquad \sum_{j,k} k^2 p(j,k) = \sigma_2^2 \end{gathered} \tag{28}$$

and a set T consisting of all points of the form $(\pm j^2, \pm k^2)$, where $j, k = 0, 1, 2, \ldots$ then

$$\langle S_T(n)\rangle \sim \left(\frac{\sigma_1 \sigma_2}{2}\right)^{1/2} \Gamma^2\left(\frac{1}{4}\right) \frac{n^{1/2}}{\ln n} \tag{29}$$

and

$$\frac{\langle S_T(n)\rangle}{\langle V_T(n)\rangle} \sim \frac{2\pi\sigma_1\sigma_2}{\ln n} \tag{30}$$

This coincides with the result for $T = \Omega$. A second generalization is the continuous time random walk.[4] Let the random walks be separable in the sense that the joint probability that a random walker makes a jump $\mathbf{r}$ after having remained on the lattice point for a time between t and $t + dt$ can be expressed as

$$p(\mathbf{r}, t)\,dt = p(\mathbf{r})\psi(t)\,dt \tag{31}$$

where $\psi(t)$ is the probability density of pausing times. Then the Laplace transforms for the expected number of visits to T in time t and the expected number of distinct points visited in T during time t, can be written as $V(\psi^*(u), T)$ and $S(\psi^*(u), T)$ respectively, where

$$\psi^*(u) = \int_0^\infty e^{-ut}\psi(t)\,dt \tag{32}$$

If the mean time between successive jumps is finite, i.e.,

$$\langle t\rangle = \int_0^\infty t\psi(t)\,dt < \infty \tag{33}$$

we can expand $\psi^*(u)$ around $u = 0$ as $\psi^*(u) \sim 1 - u\langle t\rangle + o(u)$ and it can be shown that this implies that formulas for asymptotic estimates derived so far are correct in continuous time provided that we replace n by $t/\langle t\rangle$. Likewise if the mean time between successive jumps is infinite and at long times $\psi(t) \sim t^{-1-\alpha}$, $0 < \alpha < 1$, then $\psi^*(u)$ can be expanded around $u = 0$ as $\psi^*(u) \sim 1 - cu^\alpha + o(u^\alpha)$ where c is a constant. Our asymptotic results will again be correct if we replace n by $t^\alpha/[c\Gamma(1 + \alpha)]$.[27]

One can ask different kinds of questions about higher-dimensional random walks. For example, in two dimensions we can ask for the average number of distinct points visited to a line at height m. That is, the set T consists of all points of the form (j, m), $j = 0, \pm 1, \pm 2, \ldots$. It is then easily verified that

$$\Gamma_T(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{-im\theta}\,d\theta}{1 - z\lambda(0, \theta)} \tag{34}$$

If we make the assumption of symmetry in the form given in Eq. (28) then Eq. (34) implies that

$$\begin{aligned}\Gamma_T(z) &\sim \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\cos m\theta\,d\theta}{1 - z + \sigma_2^2\theta^2/2} \\ &= \frac{1}{\sigma_2\left[2(1 - z)\right]^{1/2}} \exp\left\{-\frac{|m|}{\sigma_2}\left[2(1 - z)\right]^{1/2}\right\}\end{aligned} \tag{35}$$

and

$$S(z;T_m)=\frac{\sqrt{2}\,\pi\sigma_1}{(1-z)^{3/2}\ln[1/(1-z)]}\exp\left\{-\frac{|m|}{\sigma_2}[2(1-z)]^{1/2}\right\} \quad (36)$$

Notice that in the limit $z=1$

$$\sum_{m=-\infty}^{\infty}S(z;T_m)=\frac{\sqrt{2}\,\pi\sigma_1}{(1-z)^{3/2}\ln[1/(1-z)]}\coth\left[\frac{1}{\sigma_2}\left(\frac{1-z}{2}\right)^{1/2}\right]$$
$$\sim\frac{2\pi\sigma_1\sigma_2}{(1-z)^2\ln[1/(1-z)]}=S(z;\Omega) \quad (37)$$

the known result for two dimensions. It is interesting to observe that although the exponential term in Eq. (36) is a slowly varying function, the hyperbolic cotangent in Eq. (37), which is the sum of an infinite number of slowly varying functions, is not itself of this class. It follows from Eq. (36) that for m fixed, and $n\to\infty$

$$\langle S_{T_m}(n)\rangle\sim\frac{\sigma_1(8\pi n)^{1/2}}{\ln n} \quad (38)$$

and

$$\langle V_{T_m}(n)\rangle\sim\frac{1}{\sigma_2}\left(\frac{2n}{\pi}\right)^{1/2} \quad (39)$$

so that $\langle S_{T_m}(n)\rangle/\langle V_{T_m}(n)\rangle\sim 2\pi\sigma_1\sigma_2/\ln n$ as one would expect since the conditions in Eqs. (25) and (26) are fulfilled. Analogous problems can be treated in three dimensions. Let the set T consist of the plane $z=m$ on a simple cubic lattice. Following the steps described above we find that

$$\begin{gathered}\langle S_T(n)\rangle\sim\frac{1}{\sigma_3}\left(\frac{2n}{\pi}\right)^{1/2}\frac{1}{P(\mathbf{0};1)}\\ \langle S_T(n)\rangle/\langle V_T(n)\rangle\sim[P(\mathbf{0};1)]^{-1}\end{gathered} \quad (40)$$

where[3,4] $P(\mathbf{0};1)=1.5164$ for a simple cubic lattice. The corresponding result for a line $y=k$, $z=m$ is

$$\langle S_T(n)\rangle\sim\frac{\ln n}{\sigma_2\sigma_3P(\mathbf{0};1)},\qquad \langle V_T(n)\rangle/\langle S_T(n)\rangle\sim P(\mathbf{0};1) \quad (41)$$

All of the results cited so far are for $\langle S_T(n)\rangle$; the more interesting problems related to the higher moments of cardinality or to the distribution of the cardinality do not yield to techniques based on generating functions and require the sophisticated methods pioneered by Jain and his collaborators[20–23] as well as by Donsker and Varadhan.[24] The statistics of occupancy of a set T are considerably simpler and can be handled in a straightforward manner by the generating function technique suggested by Darling and Kac,[28] Spitzer,[29] and Rubin and Weiss.[30]

ACKNOWLEDGMENTS

We are grateful to Professor Kurt Shuler for a useful discussion. MFS thanks the Xerox Corporation for financial support.

REFERENCES

1. A. Dvoretzky and E. Erdös, *Proc. Sec. Berkeley Symp.* (University of California Press, 1951), p. 33.
2. G. H. Vineyard, *J. Math. Phys.* **4**:1191 (1963).
3. E. W. Montroll, *Proc. Symp. Appl. Math., AMS* **16**:193 (1964).
4. E. W. Montroll and G. H. Weiss, *J. Math. Phys.* **6**:167 (1965).
5. J. E. Gillis and G. H. Weiss, *J. Math. Phys.* **11**:1307 (1970).
6. R. J. Beeler and J. A. Delaney, *Phys. Rev.* **130**:962 (1963).
7. R. J. Beeler, *Phys. Rev.* **134A**:1396 (1964).
8. H. B. Rosenstock, *Phys. Rev.* **187**:1166 (1969).
9. H. B. Rosenstock, *J. Math. Phys.* **11**:487 (1970).
10. D. Haarer and H. Mohwald, *Phys. Rev. Lett.* **34**:1447 (1975).
11. D. D. Smith, R. D. Mead, and A. Zewail, *Chem. Phys. Lett.* **50**:358 (1977).
12. P. Argyrakis and R. Kopelman, *J. Theoret. Biol.* **78**:205 (1978).
13. D. D. Dlott, M. D. Fayer, and R. D. Weiting, *J. Chem. Phys.* **69**:2752 (1978).
14. H. Bouchricha, V. Ern, J. L. Fave, C. Guthmann, and M. Schott, *Chem. Phys. Lett.* **53**:288 (1978).
15. M. Nechtschein, F. Devreux, R. L. Greene, T. C. Clark, and G. B. Street, *Phys. Rev. Lett.* **44**:356 (1980).
16. H. Scher, S. Alexander, and E. W. Montroll, *Proc. Nat. Acad. Sci.* **77**:3758 (1980).
17. D. J. Pocker and S. J. Hruska, *J. Vac. Sci. Technol.* **8**:700 (1971).
18. D. Shalitin, *J. Vac. Sci. Technol.* **10**:405 (1973).
19. N. C. Jain and S. Orey, *Isr. J. Math.* **6**:373 (1968).
20. N. C. Jain and W. E. Pruitt, *Z. Wahrsch. Verwend. Geb.* **16**:279 (1970).
21. N. C. Jain and W. E. Pruitt, *J. Analyse Math.* **24**:369 (1971).
22. N. C. Jain and W. E. Pruitt, *Proc. Sixth Berkeley Symp.* (University of California Press, 1971) Vol. III, p. 31.
23. M. D. Donsker and S. R. S. Varadhan, *Commun. Pure Appl. Math.* **32**:721 (1979).
24. W. W. Brandt, *J. Chem. Phys.* **63**:5162 (1975).
25. G. H. Hardy, *Divergent Series* (Oxford University Press, Oxford, 1949), p. 166.
26. B. D. Hughes, M. F. Shlesinger, and E. W. Montroll, *Proc. Nat. Acad. Sci.* **78**:3287 (1981).
27. M. F. Shlesinger, *J. Stat. Phys.* **10**:421 (1974).
28. D. A. Darling and M. Kac, *Trans. Am. Math. Soc.* **84**:444 (1957).
29. F. Spitzer, *Principles of Random Walk* (Springer Verlag, New York, 2nd ed., 1976).
30. R. J. Rubin and G. H. Weiss, *J. Math. Phys.* (to appear).