Fragment formation in biased random walks

Kabir Ramola
Department of Theoretical Physics, Tata Institute of Fundamental Research, Mumbai, India
E-mail: kabir@tifr.res.in and kabirramola@gmail.com

Received 14 August 2008
Accepted 30 August 2008
Published 1 October 2008

Abstract. We analyse a biased random walk on a 1D lattice with unequal step lengths. Such a walk was recently shown to undergo a phase transition from a state containing a single connected cluster of visited sites to one with several clusters of visited sites (fragments) separated by unvisited sites at a critical probability $p_c$ (Anteneodo and Morgado 2007 Phys. Rev. Lett. 99 180602). The behaviour of $\rho(l)$, the probability of formation of fragments of length $l$, is analysed. An exact expression for the generating function of $\rho(l)$ at the critical point is derived. We prove that the asymptotic behaviour is of the form $\rho(l) \simeq 3/[l(l \log l)^2].$

Keywords: Brownian motion, stochastic processes (theory)
1. Introduction

Anteneodo and Morgado recently discussed an interesting one-dimensional random walk model that exhibits a phase transition [1]. In this model, at each time step the random walker moves two lattice spacings to the right or one spacing to the left with probabilities \( p \) and \( q \) respectively. The sites visited by such a walk are, depending on \( p \), either part of a single connected cluster of visited sites or many clusters of visited sites (fragments) separated by single unvisited sites. A transition from one class to the other takes place at the critical probability \( p_c = 1/3 \). At large times, for \( p < p_c \), the cluster of visited sites contains no gaps; however, for \( p > p_c \) this cluster contains a finite density of unvisited sites. Using Monte Carlo simulation data, a power law dependence with an exponent \( \approx 1.15 \) was estimated for \( n(l) \), the fraction of fragments with length \( l \). Since the model is a simple one-dimensional random walk that is not known to contain any such anomalous exponents, such a behaviour seems unlikely. We prove below that the correct asymptotic behaviour of \( n(l) \) at the critical point is \( n(l) \sim l(\log l)^{-1} \) which is in fact hard to distinguish numerically from a power law dependence with exponent \( \approx 1.15 \).

At large times, it is easy to see (using the central limit theorem) that the probability distribution of the position of the walker is a Gaussian with mean \( \mu(t) = (2p - q)t \) and variance \( \sigma^2(t) = 9pqt \). When the walker is biased towards the left (\( p < 1/3 \)), the probability of the walker being on the left of the origin tends to 1 as \( \text{erf}(\sqrt{t}) \). Hence the visited sites are part of a single connected cluster. When the walker is biased towards the right, in general the clusters of visited sites are separated by single unvisited sites. The fraction of visited sites \( f_v \) is defined as \( S_n/L_n \) where \( S_n \) is the average number of distinct sites visited in an \( n \)-step walk and \( L_n \) is the average length of an \( n \)-step walk. The fraction of unvisited sites \( f_u \), is given by \( 1 - f_v \). For this case, \( f_u \) can be calculated exactly using methods outlined in [3], [1]. In the large \( n \) limit, as \( \Delta \to 0^+ \), \( f_u = \Delta + O(\Delta^2) \), where \( \Delta = (p - p_c) \). The total number of fragments in a walk is equal to the number of unvisited sites and hence \( N \), the average number of fragments, is equal to \( L_n \Delta \).

The single unvisited sites at the edge of each fragment ensure that the probability of formation of a fragment is independent of the previous history of the walker. The walk can therefore be considered as a discrete process of independent increments where at each step a new fragment of length \( l \) is added to the positive edge of the walk with probability \( \rho(l) \). These probabilities are multiplicative, i.e. the probability of formation of...
a fragment of length \( l_1 \) followed by one of length \( l_2 \) is \( \rho(l_1)\rho(l_2) \). The number of fragments of length \( l \) in a given walk is \( N\rho(l) \) where \( N \) is the total number of fragments. Therefore \( n(l) = \rho(l)/\sum_{i=1}^{\infty} \rho(l) \) and hence \( n(l) = \rho(l) \) since \( \sum_{i=1}^{\infty} \rho(l) = 1 \).

2. Determination of fragment formation probability

For a new fragment of length \( l \) to be formed the walker must (i) be at the positive edge of the walk (at site \(-1\) for convenience), (ii) hop over site 0 and reach site \( l \) by visiting each site between 1 and \( l \), without visiting sites 0 and \( l + 1 \). Thus at the end of each step the walker is once again at the positive edge of the walk. \( \rho(l) \) is therefore the probability of event (ii) occurring. Hence \( \rho(l) \) is the sum of probabilities of all paths that are consistent with (i).

We calculate \( q(l) \), the probability of condition (ii) with the constraint of ‘visiting every site between 1 and \( l \)’ relaxed. Hence \( q(l) \) is the probability of starting at site \(-1\) and reaching site \( l \) without visiting 0 and \( l + 1 \). \( q(l) \) thus includes the probability of formation of smaller fragments within the segment \([1, l]\). For example \( q(3) = \rho(3) + \rho(1)\rho(1) \). In general

\[
q(l) = \rho(l) + \sum_{x_1 + x_2 + 1 = l} \rho(x_1)\rho(x_2) + \sum_{x_1 + x_2 + x_3 + 2 = l} \rho(x_1)\rho(x_2)\rho(x_3) + \cdots.
\]

This translates to the following equation involving generating functions:

\[
\overline{q}(x) = \frac{\overline{p}(x)}{1 - x\overline{p}(x)}, \quad \text{i.e. } \overline{p}(x) = \frac{\overline{q}(x)}{1 + x\overline{q}(x)} \tag{2}
\]

where \( \overline{p}(x) = \sum_{l=1}^{\infty} \rho(l)x^l \) and \( \overline{q}(x) = \sum_{l=1}^{\infty} q(l)x^l \).

2.1. Exact expression for \( q(l) \)

We calculate \( q(l) \) by summing over probabilities of all paths that are consistent with the above definition of \( q(l) \). For example, \( q(1) \) simply involves the walker reaching site 1 from \(-1\) by hopping over site 0. Thus \( q(1) = p \). It is also easy to see that \( q(2) \) is 0 since unvisited sites in this walk cannot be separated by two lattice sites. In the calculation of \( q(3) \) we sum over paths in which the walker, starting from \(-1\), jumps over site 0 to reach 1 and then over 2 to reach site 3 with probability \( p^2 \). Once at site 3, the walker can move two leftward steps followed by a rightward step with probability \( pq^2 \), any number of times, to return to site 3. Thus \( q(3) = p^2(1 + pq^2 + (pq^2)^2 + \cdots) = p^2/(1 - pq^2) \). Generalizing this procedure of summing over all relevant walks, we derive below a closed form expression of \( q(l) \) in two equivalent ways.

The walk can be considered as a Markov process where at each time step the walker moves either two lattice spacings to the right or one to the left with probabilities \( p \) and \( q \) respectively, independent of the previous step. The probability \( q(l) \) can be calculated from \( P_l \), the matrix of transition probabilities [2] with no transitions from the sites 0, \( l + 1 \),
Monte Carlo plots of log($S(l)$) where $S(l) = \sum_{n \geq l} n(l)$ for different values of $\Delta = p - p_c$, along with the theoretical curve calculated using equations (8) and (2).

and $l + 2$. For example

$$P_3 = \begin{bmatrix} 0 & q & 0 & 0 & 0 \\ 0 & 0 & q & 0 & 0 \\ 0 & 0 & 0 & q & 0 \\ 0 & p & 0 & 0 & 0 \\ 0 & 0 & p & 0 & 0 \\ 0 & 0 & 0 & p & 0 \end{bmatrix}$$

(3)

where the rows correspond to sites $[0\ 1\ 2\ 3\ 4\ 5]$. Consider the column vector $|p(t)\rangle$, the element in row $s$, is the probability of the walker starting at site 1 being at site $s$ at time step $t$. Now $|p(t)\rangle$ evolves as $|p(t)\rangle = P_t^l|p(0)\rangle$ which ensures that $|p(t)\rangle$ only contains probabilities of paths that never visit sites 0 and $l + 1$. Since the walker starts at site 1, $\langle s|p(0)\rangle = \delta_{1,s}$. We calculate $|P\rangle$ where $\langle s|P\rangle$ is the probability of the walker starting from site 1 being at site $s$, at any time step. $|P\rangle$ is thus the sum of $|p(t)\rangle$ at every time step. Hence

$$|P\rangle = (I + P_1 + P_1^2 + \cdots)|p(0)\rangle.$$ 

(4)

All walks that start at 1 and end at $l$ (at any time step) contribute to $q(l)$; therefore

$$q(l) = p\langle l|(I - P_l)^{-1}|1\rangle$$

(5)

where $\langle s|1\rangle = \delta_{1,s}$ and $\langle s|l\rangle = \delta_{l,s}$. Thus $q(l)$ is the $(l, 1)$ matrix element of $p(I - P_l)^{-1}$. The factor $p$ is present because the walker must hop over site 0 to reach 1 as in condition (ii). We thus obtain $q(1) = p$, $q(2) = 0$, $q(3) = (p^2/(1 - pq^2))$, $q(4) = (p^3q/(1 - 2pq^2))$ and so on. Substituting these into equation (2) we obtain $\rho(1) = p$, $\rho(2) = 0$, $\rho(3) = (p^3q^2/(1 - pq^2))$, $\rho(4) = (p^4q^3 + 3pq^5q^4)/(1 - pq^2)(1 - 3pq^2)$ and so on.

Alternatively, consider $G(s)$, the probability of a walker starting at site $-1$ being at site $s$ while never visiting 0 and $l + 1$. Now $G(s)$ satisfies the difference equation

$$G(s) = pG(s - 2) + qG(s + 1)$$

(6)
with the boundary conditions $G(0) = 0$ and $G(l + 1) = 0$. A general solution of $G(s)$ is $Az_1^s + Bz_2^s + Cz_3^s$ which on substitution into equation (6) yields the cubic equation $qz^3 - z^2 + p = 0$. The solutions to this equation are $z = 1, z_\pm$ where $z_\pm = (p \pm \sqrt{p^2 + 4pq})/2q$. By satisfying boundary conditions we obtain

$$G(s) = A \left[ 1 - \frac{(1 - z_{l+1}^{(l+1)})}{z_{l+1}^{(l+1)} - z_{l+1}^{(l+1)}}z_+^s - \frac{(z_+^{(l+1)} - 1)}{z_+^{(l+1)} - z_{l+1}^{(l+1)}}z_-^s \right]$$

(7)

where $A$ is a constant which is determined from the normalization condition $G(-1) = 1$. Alternatively from the definition of $q(l)$ as the connecting probability between $-1$ and $l$ we have

$$q(l) = G(l)/G(-1).$$

(8)

We thus have a closed form expression for $q(l)$. We verify that the two expressions for $q(l)$ are equal as the matrix elements of the column corresponding to site 1 in $p(I - R)^{-1}$ satisfy the same difference equation and, with an appropriate change of variables, the same boundary conditions as $G(s)$.

3. Asymptotic behaviour

Away from the critical point ($p > p_c$), for large $l$,

$$q(l) \approx \frac{1 - (1/z_+)}{1 - (1/z_-)} - \frac{z_- - z_+ (z_+ - 1)}{(1 - z_-)^2 z_+^s} \exp[-l/\xi]$$

(9)

where $\xi = 1/(\log[z_+] \approx 1/(3(p - p_c))$. $\xi$ thus defines a correlation length for $q(l)$ and hence for $\rho(l)$.

At the critical point, $q(l)$ simplifies to

$$q(l) = \frac{2[1 - (-1/2)^l] - 3l(-1/2)^l}{8 + (-1/2)^l + 6l} \approx \frac{1}{3l}$$

(10)

for large $l$. 

Figure 2. Theoretical plots of $1/S(l)$ versus log$(l)$ for different values of $\Delta$ showing a limiting slope of $\approx 0.33$ in accordance with equation (12).
Therefore as $x \to 1$, the generating function $\mathcal{P}(x)$ diverges as $(-1/3) \log(1 - x)$. Hence from equation (2),

$$\mathcal{P}(x) = \frac{(-1/3) \log(1 - x)}{1 - (x/3) \log(1 - x)} \quad \text{for} \quad (x \to 1, p = p_c). \quad (11)$$

Having found the generating function, we can extract the behaviour of $\rho(l)$ for large $l$. From equation (11), $\mathcal{P}(x) \simeq 1 + 3/\log \epsilon$ where $\epsilon = (1 - x)$. We can approximate the summation in the generating function by an integral $\mathcal{P}(x) \simeq \int_{1}^{\infty} \exp[-\epsilon l] \rho(l) \, dl$. The singular part of this integral can be evaluated by equating $\mathcal{P}(x) \simeq \int_{1}^{1/\epsilon} \rho(l) \, dl$, from which we see that $\int_{1/\epsilon}^{\infty} \rho(l) \, dl \simeq -3/\log \epsilon$. Alternatively, from the fact that the integral $\lim_{\epsilon \to 0} \int_{2}^{\infty} \exp[-\epsilon l]/(\log l)^2 \, dl$ diverges as $[\epsilon(\log \epsilon)^2]^{-1}(1 + O[1/\log \epsilon])$, we obtain

$$\rho(l) \simeq \frac{3}{l(\log l)^2} \quad \text{for large} \quad l. \quad (12)$$

We thus have an expression for the asymptotic behaviour of $\rho(l)$.

Cases in which the rightward steps are of length $k > 2$ can be treated in a similar way. In this case the independent elements are clusters of visited and unvisited sites separated by $k - 1$ zeros.

**Acknowledgments**

I thank Professor Deepak Dhar for providing the central ideas that led to these results and R Loganayagam for helpful discussions on the series expansions involved.

**References**

[1] Anteneodo C and Morgado W A M, 2007 *Phys. Rev. Lett.* 99 180602
[2] Feller W, 1968 *Introduction to Probability Theory and Its Applications* (New York: Wiley)
[3] Montroll E W and Weiss G H, 1965 *J. Math. Phys.* 6 167