A two-parameter random walk with approximate exponential probability distribution

2006 J. Phys. A: Math. Gen. 39 7245

(http://iopscience.iop.org/0305-4470/39/23/005)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **39** (2006) 7245–7256 doi:10.1088/0305-4470/39/23/005

A two-parameter random walk with approximate exponential probability distribution

Erik Van der Straeten[1] and Jan Naudts

Department Fysica, Universiteit Antwerpen, Universiteitsplein 1, 2610 Antwerpen, Belgium

E-mail: Erik.VanderStraeten@ua.ac.be and Jan.Naudts@ua.ac.be

Received 17 January 2006
Published 23 May 2006
Online at stacks.iop.org/JPhysA/39/7245

Abstract

We study a non-Markovian random walk in dimension 1. It depends on two parameters ϵ_R and ϵ_L, the probabilities to go straight on when walking to the right, and respectively to the left. The position x of the walk after n steps and the number of reversals of direction k are used to estimate ϵ_R and ϵ_L. We calculate the joint probability distribution $p_n(x,k)$ in a closed form and show that, approximately, it belongs to the exponential family.

PACS numbers: 05.40.Fb, 02.50.−r

1. Introduction

Consider a random walk starting in the origin $x = 0$ of the lattice $\mathbb{Z}$. The probability that after n steps the walk is in x and changed its direction k times is denoted by $p_n(x,k)$. This paper investigates the question of how $p_n(x,k)$ depends on model parameters. We wonder whether it can be written in the form

$$p_n(x,k) = D_n(x,k)\exp(G + \beta k + Fx). \tag{1}$$

In this expression, G, β and F depend on model parameters. However, the prefactor $D_n(x,k)$ does not depend on model parameters. The function β has the interpretation of an inverse temperature (in dimensionless units), the function F is an external force, the function G, when divided by β, is a free energy and serves to normalize (1). A probability distribution $p_n(x,k)$ of the form (1) is said to belong to the *exponential family*. It has nice properties. In particular, averages of x and k can be calculated by taking derivatives of G with respect to the parameters.

Random walk models are omnipresent in statistical physics and have been studied extensively. Quite often results are obtained in the limit of large n. Here, the focus is on all n. Deviations from (1), found below, are negligible in the large-n limit. Standard techniques aim

[1] Research Assistant of the Research Foundation—Flanders (FWO—Vlaanderen).

at calculating correlation functions. It is rather seldom that exact expressions for probability distributions can be written down in a closed form. In the present model such closed-form expression exists for $p_n(x,k)$, but probably not for the marginals $p'_n(x) = \sum_k p_n(x,k)$ and $p''_n(k) = \sum_x p_n(x,k)$. Individual events have usually such a small probability that they cannot be evaluated numerically. In addition, in situations with a large number of degrees of freedom, knowledge of $p_n(x,k)$ is not sufficient to evaluate moments of the distribution in a closed form. However, if a closed-form expression of $p_n(x,k)$ is available then analytic relations can be used to evaluate relevant quantities.

The model, considered here, is that of a one-dimensional persistent random walk (see, e.g., [1]) with drift. Many generalizations of the persistent random walk can be found in the literature, e.g. for continuous time [2], or with a memory that goes back more than one step [3]. Planar persistent random walks have been studied in [4, 5].

Our model is a toy model that helps to understand features of more realistic models used in several branches of physics. One such application, well-known since the pioneering work of Flory [6], is the use of random walks to model the geometry of polymers. Persistent random walks play also a role in understanding the transition from ballistic to diffusive transport [7–9] and have been applied in financial physics, see e.g. [10].

Recent technological progress has made it possible to do experiments on single molecules and to measure the elongation of a single polymer $\langle x\rangle$ as a function of applied force F (see, e.g., [11–13]). The analysis of these experiments is based on the assumption that $p'_n(x)$ is proportional to $\exp(-\beta V(x) + Fx)$ for some potential $V(x)$. This relation follows from (1) with $\exp(-\beta V(x)) = \sum_k D_n(x,k)\,\mathrm{e}^{\beta k}$. The latter expression shows that $V(x)$ is indeed a free energy, as claimed in [13]. An in-depth discussion of these experiments based on the results of the present paper is found in [14].

In the next section the model is introduced. In section 3 average values for position x and number of reversals k are calculated using the method of generating functions. In section 4 the number of walks ending in x after n steps, and having a given number of reversals k, is calculated. These counting results are used in section 5 to write down the joint probability distribution $p_n(x,k)$. Section 6 considers the dependence of $p_n(x,k)$ on the parameters ϵ_{R} and ϵ_{L} and tries to answer the question of whether this two-parameter probability distribution function belongs to the exponential family. Section 7 shows how to calculate averages starting from the knowledge that the probability distribution function is exponential. The final section gives a short discussion of the results.

2. Model

Consider a discrete-time random walk on the one-dimensional lattice $\mathbb{Z}$. The probability of the walk to step to the right (i.e., with increasing position) equals ϵ_{R} when coming from the left and $1-\epsilon_{\mathrm{L}}$ when coming from the right. This is not a Markov chain since the walk remembers the direction it comes from. Let x_n be the position of the walk after n steps. Let $\sigma_n = x_n - x_{n-1}$ be the direction of the n-th step. Then $x_{n+1} = x_n + 1$ with probability

$$\tfrac{1}{2}(1+\sigma_n)\epsilon_{\mathrm{R}} + \tfrac{1}{2}(1-\sigma_n)(1-\epsilon_{\mathrm{L}}), \tag{2}$$

$x_{n+1} = x_n - 1$ otherwise. The process of the increments σ_n is a two-state Markov chain with the transition matrix

$$P = \begin{pmatrix} \epsilon_{\mathrm{R}} & 1-\epsilon_{\mathrm{R}} \\ 1-\epsilon_{\mathrm{L}} & \epsilon_{\mathrm{L}} \end{pmatrix}. \tag{3}$$

In the stationary state σ_n equals ± 1 with probability $p_\pm^{(0)}$ given by

$$p_+^{(0)} = \frac{1-\epsilon_{\mathrm{L}}}{2-\epsilon_{\mathrm{R}}-\epsilon_{\mathrm{L}}}, \qquad p_-^{(0)} = \frac{1-\epsilon_{\mathrm{R}}}{2-\epsilon_{\mathrm{R}}-\epsilon_{\mathrm{L}}}. \tag{4}$$

Let k_n denote the number of reversals of the walk after n steps. By definition a reversal occurs at step n if $\sigma_{n-1}\sigma_n = -1$. Hence one has

$$x_n = \sum_{j=1}^{n} \sigma_j \tag{5}$$

$$k_n = \frac{1}{2}\sum_{j=1}^{n}(1-\sigma_{j-1}\sigma_j). \tag{6}$$

The quantity of interest in this paper is the joint probability of position x_n and number of reversals k_n. The appropriate initial conditions are $x_0 = 0$ and $\sigma_0 = \pm 1$ with probability $p_\pm^{(0)}$, because these are the stationary values. The choice of initial conditions is delicate because they introduce memory effects of a type well known in the renewal theory.

The physical interpretation of the model is twofold. The random walk is a simple model of a polymer with n units. Energy is proportional to minus the number of reversals k_n. The position of the end point x_n measures the effect of an external force applied to the end point. Alternatively, k_n is the number of domains (decreased by 1 if $\sigma_1 = \sigma_0$) of an Ising chain, and x_n is the total magnetization. Indeed, the variables σ_n describe Ising spins on a one-dimensional lattice. A domain is then a set of subsequent sites where the spins all have the same value, either up (+1) or down (-1). The boundary between two domains involves a reversal ($\sigma_{j-1}\sigma_j = -1$).

3. Generating functions

Let $p_n^\pm(x,k)$ denote the probability that $\sigma_n = \pm 1$, $x_n = x$ and $k_n = k$. The joint probability distribution, searched for, is then

$$p_n(x,k) = p_n^+(x,k) + p_n^-(x,k). \tag{7}$$

The following recursion relations hold:

$$p_n^+(x,k) = \epsilon_{\mathrm{R}} p_{n-1}^+(x-1,k) + (1-\epsilon_{\mathrm{L}}) p_{n-1}^-(x-1,k-1) \tag{8}$$

$$p_n^-(x,k) = (1-\epsilon_{\mathrm{R}}) p_{n-1}^+(x+1,k-1) + \epsilon_{\mathrm{L}} p_{n-1}^-(x+1,k). \tag{9}$$

Introduce generating functions

$$f_\pm^{(n)}(w,z) = \sum_{x=-n}^{n} w^x \sum_{k=0}^{n} z^k p_n^\pm(x,k) \tag{10}$$

and a similar expression for $f^{(n)}(w,z)$. They satisfy

$$\begin{pmatrix} f_+(n) \\ f_-(n) \end{pmatrix} = M(w,z) \begin{pmatrix} f_+(n-1) \\ f_-(n-1) \end{pmatrix} \tag{11}$$

with

$$M(w,z) = \begin{pmatrix} \epsilon_{\mathrm{R}} w & (1-\epsilon_{\mathrm{L}})wz \\ (1-\epsilon_{\mathrm{R}})z/w & \epsilon_{\mathrm{L}}/w \end{pmatrix}. \tag{12}$$

It is possible to calculate the n-th power of this matrix by first diagonalizing it. The result is

$$M^n(w,z) = \frac{1}{2}\left(\lambda_+^n + \lambda_-^n\right)\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} + \frac{1}{2\nu}\left(\lambda_+^n - \lambda_-^n\right)\begin{pmatrix} w\epsilon_R - \epsilon_L/w & 2(1-\epsilon_L)wz \\ 2(1-\epsilon_R)z/w & -w\epsilon_R + \epsilon_L/w\end{pmatrix}, \tag{13}$$

with

$$\lambda_\pm = \tfrac{1}{2}(\epsilon_R w + \epsilon_L/w \pm \nu) \tag{14}$$

and

$$\nu = \sqrt{(\epsilon_R w - \epsilon_L/w)^2 + 4(1-\epsilon_R)(1-\epsilon_L)z^2}. \tag{15}$$

Let us now consider initial values. Note that $f_\pm^{(0)}(w,z)$ is not yet defined because k_0 involves σ_{-1}, which is undetermined. The starting point is therefore $f_\pm^{(1)}(w,z)$, which is found to be given by

$$f_\pm^{(1)}(w,z) = M(w,z)\begin{pmatrix}p_+^{(0)}\\ p_-^{(0)}\end{pmatrix}. \tag{16}$$

Hence, it is obvious to define

$$f_\pm^{(0)}(w,z) = \begin{pmatrix}p_+^{(0)}\\ p_-^{(0)}\end{pmatrix}. \tag{17}$$

The generating function $f^{(n)}(w,z)$ is now explicitly known as

$$\begin{aligned} f^{(n)}(w,z) = \frac{1}{2}\left(\lambda_+^n + \lambda_-^n\right) + \frac{1}{2\nu}\left(\lambda_+^n - \lambda_-^n\right)\big[(\epsilon_R w - \epsilon_L/w)\left(p_+^{(0)} - p_-^{(0)}\right) \\ + 2(1-\epsilon_R)p_+^{(0)}z/w + 2(1-\epsilon_L)p_-^{(0)}wz\big]. \end{aligned} \tag{18}$$

It can be used to calculate expectation values by taking derivatives. For example,

$$\begin{aligned} \langle k_n\rangle &= \frac{\partial}{\partial z}\bigg|_{w=z=1} f^{(n)}(w,z) \\ &= 2n\frac{(1-\epsilon_R)(1-\epsilon_L)}{2-\epsilon_R-\epsilon_L}, \end{aligned} \tag{19}$$

and

$$\begin{aligned} \langle x_n\rangle &= \frac{\partial}{\partial w}\bigg|_{w=z=1} f^{(n)}(w,z) \\ &= n\frac{\epsilon_R-\epsilon_L}{2-\epsilon_R-\epsilon_L}. \end{aligned} \tag{20}$$

4. Counting walks

The present section is temporarily limited to the special case $\epsilon_R = \epsilon_L = 1/2$. From the next section on the general model will be considered again. Indeed, we first determine the number of walks $c_\pm(n,x,s)$ which, starting in the origin in direction ± 1, end in x after n steps and have s segments. The result does not depend on the value of ϵ_R and ϵ_L. Hence the calculation can be done in the simplest case. In the next section the result will be used to calculate the joint probability distribution $p_n(x,k)$ for the general model.

Divide the walk into segments of constant σ_j. Number these segments from 1 to s_n. Note that $s_n = k_n + 1$ if $\sigma_1 = \sigma_0$, $s_n = k_n$ otherwise. This means that the number of segments equals 1 plus the number of reversals, not counting the initial reversal at $x = 0$, if present. Let τ_j

denote the length of the j-th segment. The probability that segment j has length l equals 2^{-l}. The probability of counting s segments in a walk of n steps satisfies

$$
\begin{aligned}
\mathcal{P}(s_n = s) &= \mathcal{P}\left(\sum_{j=1}^{s-1}\tau_j < n \leqslant \sum_{j=1}^{s}\tau_j\right) \\
&= \sum_{l_1=1}^{\infty}\cdots\sum_{l_s=1}^{\infty} 2^{-l_1-\cdots-l_s}\mathbb{I}\{l_1+\cdots+l_{s-1} < n \leqslant l_1+\cdots+l_s\} \\
&= \sum_{m=s-1}^{n-1} 2^{-m}\sum_{l_1=1}^{\infty}\cdots\sum_{l_{s-1}=1}^{\infty}\delta_{m,l_1+\cdots+l_{s-1}}\sum_{l_s=n-m}^{\infty}2^{-l_s} \\
&= 2^{-(n-1)}\sum_{m=s-1}^{n-1}\sum_{l_1=1}^{\infty}\cdots\sum_{l_{s-1}=1}^{\infty}\delta_{m,l_1+\cdots+l_{s-1}} \\
&= 2^{-(n-1)}\binom{n-1}{s-1}.
\end{aligned}
\tag{21}
$$

Hence the conditional probability given a certain number of segments equals

$$
\mathcal{P}(\tau_1 = l_1, \ldots, \tau_{s-1} = l_{s-1} \mid s_n = s) = \binom{n-1}{s-1}^{-1}. \tag{22}
$$

This means that the variables τ_j, after conditioning on a given number of segments, become uniformly distributed. This observation simplifies the following calculation.

The position x_n of the walk after n steps, assuming s_n segments, can be expressed into the segment lengths as

$$
x_n = \sigma_1\left(\tau_1 - \tau_2 + \cdots \pm \tau_{s_n-1} \mp \left(n - \sum_{j=1}^{s_n-1}\tau_j\right)\right). \tag{23}
$$

The $\pm$-sign depends on whether the number of segments s_n is even or odd and equals $(-1)^{s_n}$. One obtains

$$
x_n = \begin{cases} \sigma_1(2(\tau_1+\tau_3+\cdots+\tau_{s_n-1}) - n) & \text{if } s_n \text{ is even} \\ \sigma_1(n - 2(\tau_2+\tau_4+\cdots+\tau_{s_n-1})) & \text{if } s_n \text{ is odd.} \end{cases} \tag{24}
$$

For simplicity let us first consider the case of an even number of segments. Let $s > 0$ be even. Then one has

$$
\begin{aligned}
\mathcal{P}(x_n = x, s_n = s) &= \mathcal{P}(s_n = s)\mathcal{P}(\tau_1+\tau_3\cdots+\tau_{s-1} = (n+\sigma_1 x)/2 | s_n = s) \\
&= 2^{-(n-1)}\sum_{l_1=1}^{\infty}\cdots\sum_{l_{s-1}=1}^{\infty}\mathbb{I}\left\{\sum_{j=1}^{s-1}l_j < n, l_1+l_3+\cdots+l_{s-1} = (n+\sigma_1 x)/2\right\} \\
&= 2^{-(n-1)}\sum_{l_1=1}^{\infty}\sum_{l_3=1}^{\infty}\cdots\sum_{l_{s-1}=1}^{\infty}\mathbb{I}\{l_1+l_3+\cdots+l_{s-1} = (n+\sigma_1 x)/2\} \\
&\quad\times\sum_{l_2=1}^{\infty}\sum_{l_4=1}^{\infty}\cdots\sum_{l_{s-2}=1}^{\infty}\mathbb{I}\{l_2+l_4+\cdots+l_{s-2} < (n-\sigma_1 x)/2\} \\
&= 2^{-(n-1)}\Theta_n^{\sigma_1}(x,s)C(n+x-2,s-2)C(n-x-2,s-2)
\end{aligned}
\tag{25}
$$

with

$$C(n,m) = \frac{n!!}{m!!((n-m)!!} \tag{26}$$

and with $\Theta_n^{\pm}(x,s)$ equal 1 if there exists a walk of n steps, starting in the origin in direction ± 1, ending in x, and containing s segments, and zero otherwise. The definition of the double factorial is given by

$$n!! = \begin{cases} n \cdot (n-2) \ldots 5 \cdot 3 \cdot 1 & \text{if } n \text{ is odd,} \\ n \cdot (n-2) \ldots 6 \cdot 4 \cdot 2 & \text{if } n \text{ is even,} \end{cases} \tag{27}$$

($0!! = (-1)!! = 1$ by convention).

If s is odd, $s \geqslant 3$, then one has

$$\begin{aligned} \mathcal{P}(x_n = x, s_n = s) &= \mathcal{P}(s_n = s)\mathcal{P}(\tau_2 + \tau_4 \cdots + \tau_{s-1} = (n - \sigma_1 x)/2 | s_n = s) \\ &= 2^{-(n-1)} \sum_{l_1=1}^{\infty} \cdots \sum_{l_{s-1}=1}^{\infty} \mathbb{I}\left\{ \sum_{j=1}^{s-1} l_j < n, l_2 + l_4 + \cdots + l_{s-1} = (n - \sigma_1 x)/2 \right\} \\ &= 2^{-(n-1)} \sum_{l_1=1}^{\infty} \sum_{l_3=1}^{\infty} \cdots \sum_{l_{s-2}=1}^{\infty} \mathbb{I}\{l_1 + l_3 + \cdots + l_{s-2} < (n + \sigma_1 x)/2\} \\ &\quad \times \sum_{l_2=1}^{\infty} \sum_{l_4=1}^{\infty} \cdots \sum_{l_{s-1}=1}^{\infty} \mathbb{I}\{l_2 + l_4 + \cdots + l_{s-1} = (n - \sigma_1 x)/2\} \\ &= 2^{-(n-1)} \Theta_n^{\sigma_1}(x,s) C(n + \sigma_1 x - 2, s-1) C(n - \sigma_1 x - 2, s-3). \end{aligned} \tag{28}$$

Note that, in the case of an odd number of segments, the number of walks ending in x_n depends on whether the walk starts to the left or to the right.

Finally, if $s = 1$ then there is clearly only one walk ending in the point x_n.

5. The joint probability distribution

Let us return to the general case with arbitrary ϵ_R and ϵ_L. The probability of a given n-step walk depends only on σ_0, and on the final values x_n and k_n. To see this, note that a segment of length τ has probability $(1-\epsilon_R)\epsilon_R^{\tau-1}$ if the direction is positive, and $(1-\epsilon_L)\epsilon_L^{\tau-1}$ if the direction is negative. A factor ϵ_R can be associated with every step to the right, and ϵ_L with every step to the left. But then a factor $(1-\epsilon_L)/\epsilon_R$, respectively $(1-\epsilon_R)/\epsilon_L$, must be associated with every reversal of direction from leftgoing to rightgoing, respectively rightgoing to leftgoing.

The number of steps to the right respectively to the left is $(n + x_n)/2$, respectively $(n - x_n)/2$. The number of reversals from leftgoing to rightgoing is denoted by k_n^-, from rightgoing to leftgoing k_n^+. They depend on whether the number of reversals is even or odd. If k_n is odd then

$$k_n^{\pm} = (k_n \pm \sigma_0)/2. \tag{29}$$

Obviously is $k_n^- + k_n^+ = k_n$ and $k_n^+ - k_n^- = \sigma_0$. On the other hand, if k_n is even then the number of reversals is $k_n^- = k_n^+ = k_n/2$, independent of the direction σ_0. In both cases, the probability of the n-step walk, given that it ends in x, has k^+ reversals when going right and k^- when going left, is

$$\gamma_n(x, k^+, k^-) \equiv \epsilon_R^{\frac{n+x}{2}} \epsilon_L^{\frac{n-x}{2}} \left(\frac{1-\epsilon_R}{\epsilon_L}\right)^{k^+} \left(\frac{1-\epsilon_L}{\epsilon_R}\right)^{k^-}. \tag{30}$$

The number of such walks is denoted by $D_n^{\sigma_0}(x, k^+, k^-)$ and equals

$$D_n^{\sigma_0}(x, k^+, k^-) = \Theta_{n+1}^{\sigma_0}(x + \sigma_0, k^+ + k^- + 1)\,\Xi^{\sigma_0}(k^+, k^-)\,C(n - 1 + x + \sigma_0, 2k^- + \sigma_0 - 1) \\ \times C(n - 1 - x - \sigma_0, 2k^+ - \sigma_0 - 1). \tag{31}$$

The function $\Xi^{\sigma_0}(k^+, k^-)$ equals 1 if $k^+ = k^-$ or $k^+ - k^- = \sigma_0$ and zero otherwise. To see from where (31) follows, consider a walk of $n + 1$ steps starting at position $-\sigma_0$ and apply the results of the previous section. Note that the number of segments of this walk is $k^+ + k^- + 1$.

The final result for the joint probability distribution $p_n(x, k)$ is then

$$p_n(x, k) = p_+^{(0)} q_n^+(x, k) + p_-^{(0)} q_n^-(x, k) \tag{32}$$

with the probability distribution $q_n^\pm(x, k)$ given by

$$\begin{aligned} q_n^\pm(x, k) &= D_n^\pm\left(x, \frac{k}{2}, \frac{k}{2}\right)\gamma_n\left(x, \frac{k}{2}, \frac{k}{2}\right), && k \text{ even} \\ &= D_n^\pm\left(x, \frac{k \pm 1}{2}, \frac{k \mp 1}{2}\right)\gamma_n\left(x, \frac{k \pm 1}{2}, \frac{k \mp 1}{2}\right), && k \text{ odd.} \end{aligned} \tag{33}$$

As an example let us calculate

$$p_4(2, 3) = p_+^{(0)} D_4^+(2, 2, 1)\gamma_4(2, 2, 1) + p_-^{(0)} D_4^-(2, 1, 2)\gamma_4(2, 1, 2). \tag{34}$$

One has

$$D_4^+(2, 2, 1) = \Theta_5^+(3, 4)\,\Xi^+(2, 1)C(6, 2)C(0, 2) \tag{35}$$

$$D_4^-(2, 1, 2) = \Theta_5^-(1, 4)\,\Xi^-(1, 2)C(4, 2)C(2, 2). \tag{36}$$

Clearly, $\Theta_5^+(3, 4) = 0$ and $\Theta_5^-(1, 4) = 1$. Using

$$\gamma_4(2, 1, 2) = \epsilon_R(1 - \epsilon_R)(1 - \epsilon_L)^2 \tag{37}$$

one obtains

$$p_4(2, 3) = 2\epsilon_R \frac{(1 - \epsilon_R)^2(1 - \epsilon_L)^2}{2 - \epsilon_R - \epsilon_L}. \tag{38}$$

This example shows that sometimes only one of the two terms contributing to the right-hand side of (32) does not vanish.

6. Exponential family

One can write

$$\gamma_n\left(x, \frac{k}{2}, \frac{k}{2}\right) = \exp(G + \beta k + Fx) \tag{39}$$

with

$$F = \frac{1}{2}\ln\frac{\epsilon_R}{\epsilon_L} \tag{40}$$

$$\beta = -\frac{1}{2}\ln\frac{\epsilon_R}{\epsilon_L}(1 - \epsilon_R)(1 - \epsilon_L) \tag{41}$$

$$G = \frac{n}{2}\ln\epsilon_R\epsilon_L. \tag{42}$$

This reparametrization allows us to write $q_n^{\pm}(x,k)$, appearing in our main result (32), as

$$q_n^{\pm}(x,k) = D_n^{\pm}\left(x, \frac{k\pm\Delta}{2}, \frac{k\mp\Delta}{2}\right)\exp(G+\beta k+Fx\pm\gamma\Delta), \tag{43}$$

where

$$\gamma = \frac{1}{2}\ln\frac{(1-\epsilon_R)\epsilon_R}{(1-\epsilon_L)\epsilon_L}, \tag{44}$$

and with $\Delta = 1$ if k is odd, and zero if k is even. Hence the probability distributions $q_n^{\pm}(x,k)$ belong to the exponential family, however not with two but with three parameters β, F and γ. The third parameter γ controls boundary effects. Hence, $p_n(x,k)$ is a superposition of two distributions $q_n^{\pm}(x,k)$, both belonging to the exponential family. However, the domains on which these two probability distribution functions differ from zero are not identical.

If n is large then the variable Δ can usually be neglected, being small compared to typical values of k and x. One obtains the approximate result that, for those values of x and k for which $p_n(x,k)\neq 0$,

$$p_n(x,k) \simeq \binom{\frac{n+x}{2}}{\frac{k}{2}}\binom{\frac{n-x}{2}}{\frac{k}{2}}\exp(G+\beta k+Fx). \tag{45}$$

This shows that $p_n(x,k)$ approximately belongs to the exponential family with two parameters β and F. Deviations between the left-hand side and the right-hand side of (45) occur for two reasons: there is a subtle difference in expressions for even k and for odd k, and there is a small dependence on the initial conditions.

Simple random walk corresponds with the choice $\epsilon_R = \epsilon_L = 1/2$. This implies infinite temperature (i.e. vanishing β) and absence of drift ($F = 0$). The third parameter γ vanishes as well. Also random walk with drift is a special case, corresponding with $\epsilon_R + \epsilon_L = 1$. Again, $\beta = 0$ and $\gamma = 0$ follow. A persistent random walk is obtained when $\epsilon_R = \epsilon_L$. This implies $F = 0$, but non-vanishing β and γ.

7. Calculating averages

Let us now see what exponential expressions are good for. First consider the approximate expression (45). From $\sum_{x,k} p_n(x,k) = 1$ follows

$$0 \simeq \sum_n \frac{\partial}{\partial\beta}p_n(x,k) = \frac{\partial G}{\partial\beta} + \langle k_n\rangle \tag{46}$$

$$0 \simeq \sum_n \frac{\partial}{\partial F}p_n(x,k) = \frac{\partial G}{\partial F} + \langle x_n\rangle. \tag{47}$$

Using (42) there follows

$$0 = \frac{n}{2\epsilon_R} - \frac{1}{2\epsilon_R(1-\epsilon_R)}\langle k_n\rangle + \frac{1}{2\epsilon_R}\langle x_n\rangle \tag{48}$$

$$0 = \frac{n}{2\epsilon_L} - \frac{1}{2\epsilon_L(1-\epsilon_L)}\langle k_n\rangle - \frac{1}{2\epsilon_L}\langle x_n\rangle. \tag{49}$$

When solving these equations for $\langle k_n\rangle$ and $\langle x_n\rangle$ one recovers (19), (20). Hence, from the approximate result (45), which one can guess without hard work, one obtains immediately exact results for the averages $\langle k_n\rangle$ and $\langle x_n\rangle$.

Let us now try to do the same starting from the exact expressions (32), (43). From the normalization of $q_n^{\pm}(x,k)$ follows the set of equations

$$0 = \frac{n}{2\epsilon_R} - \frac{1}{2\epsilon_R(1-\epsilon_R)}\langle k_n\rangle^{\pm} + \frac{1}{2\epsilon_R}\langle x_n\rangle^{\pm} \pm \frac{1-2\epsilon_R}{2\epsilon_R(1-\epsilon_R)}\langle \Delta_n\rangle^{\pm} \tag{50}$$

$$0 = \frac{n}{2\epsilon_L} - \frac{1}{2\epsilon_L(1-\epsilon_L)}\langle k_n\rangle^{\pm} - \frac{1}{2\epsilon_L}\langle x_n\rangle^{\pm} \mp \frac{1-2\epsilon_L}{2\epsilon_L(1-\epsilon_L)}\langle \Delta_n\rangle^{\pm}. \tag{51}$$

They can be written as

$$\langle x_n\rangle^{\pm} = n\frac{\epsilon_R-\epsilon_L}{2-\epsilon_R-\epsilon_L} \mp 2\frac{1-\epsilon_R-\epsilon_L}{2-\epsilon_R-\epsilon_L}\langle \Delta_n\rangle^{\pm} \tag{52}$$

$$\langle k_n\rangle^{\pm} = 2n\frac{(1-\epsilon_R)(1-\epsilon_L)}{2-\epsilon_R-\epsilon_L} \mp \frac{\epsilon_R-\epsilon_L}{2-\epsilon_R-\epsilon_L}\langle \Delta_n\rangle^{\pm}. \tag{53}$$

These averages are calculated with boundary conditions $\sigma_0 = +1$ or $\sigma_0 = -1$. The expressions are the sum of a part independent of the boundary condition and a small contribution which depends on the boundary condition and on the probability $\langle \Delta_n\rangle$ that the number of reversals k_n is odd. For large n, the effect of the terms in Δ_n is small, as can be seen from these equations. However for small n these terms cannot be ignored and we are left with only four equations for six variables. This is an annoying consequence of the fact that the probabilities $q_n^{\pm}(x,k)$ belong to the exponential family with three parameters instead of two.

One could try to proceed by using the results of section 3. Comparison of (52), (53) with (19), (20) gives

$$p_+^{(0)}\langle \Delta_n\rangle^{+} = p_-^{(0)}\langle \Delta_n\rangle^{-}. \tag{54}$$

Hence one can write

$$\langle \Delta_n\rangle^{+} = \frac{2-\epsilon_R-\epsilon_L}{2(1-\epsilon_L)}\langle \Delta_n\rangle \tag{55}$$

$$\langle \Delta_n\rangle^{-} = \frac{2-\epsilon_R-\epsilon_L}{2(1-\epsilon_R)}\langle \Delta_n\rangle. \tag{56}$$

However, one cannot obtain a closed-form expression for $\langle \Delta_n\rangle$, which is the probability that k_n is odd. So one is forced to calculate expressions for $\langle \Delta_n\rangle^{\pm}$ explicitly. In the appendix the results

$$\langle \Delta_n\rangle^{\pm} = (1-[\epsilon_R+\epsilon_L-1]^n)p_{\mp}^{(0)} \tag{57}$$

$$\langle \Delta_n\rangle = 2p_+^{(0)}p_-^{(0)}(1-[\epsilon_R+\epsilon_L-1]^n) \tag{58}$$

are obtained by deriving recursion relations for $\langle \Delta_n\rangle^{\pm}$. Note that equality (54) indeed holds. Relations (57) together with (52), (53) form a closed set of equations allowing us to obtain expression for $\langle k_n\rangle$, $\langle x_n\rangle$ and $\langle \Delta_n\rangle$.

8. Discussion

We have studied a simple model of random walk depending on two parameters ϵ_R and ϵ_L. The parameters are estimated using the position x_n of the walk after n steps, and the number of reversals of direction k_n. The technique of generating functions is used to calculate averages

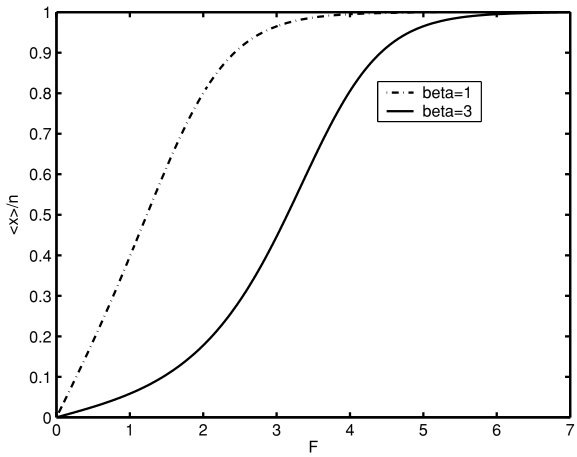

Figure 1. Average position $\langle x\rangle$, divided by n, as a function of external force F for two different values of β; based on equations (20), (40), (41).

$\langle x_n\rangle$ and $\langle k_n\rangle$. Next, explicit expressions are obtained for the number of walks that end in the same position x_n and have the same number of reversals k_n. These counts are used to write an explicit result (32) for the joint probability distribution $p_n(x, k)$. In the final part of the paper we try to write this joint probability distribution in the form of an exponential family. This succeeds only in an approximate manner. The distribution $p_n(x, k)$ is a superposition of two probability distribution functions $q_n^{\pm}(x, k)$, both belonging to the exponential family, but with three parameters instead of two. The third parameter controls the probability that the number of reversals is odd. The difference between walks with even or odd number of reversals is negligible in the limit of large n. The consequence of the third parameter is that a closed set of equations in the averages of k, x and Δ cannot be obtained, by taking derivatives of the partition function with respect to the parameters. An explicit calculation of the average of Δ is needed to close the set of equations.

Some of the main results of the paper are explicit expressions (40), (41) for thermodynamic parameters β and F in terms of the model parameters ϵ_{R} and ϵ_{L}. They are used in figure 1 to plot average position as a function of external force F. The latter quantity can be measured experimentally. Our result shows a typical sigmoidal curve, in qualitative agreement with the measurements of [11]. A more complete discussion of the application of the present work is found in [14].

We did not succeed to obtain a closed expression for the marginal distribution $p'_n(x) = \sum_k p_n(x, k)$ of the position of the walker. Note that the result of section 4, counting walks with given x and k, is not needed in later sections to derive expressions for average position and average number of reversals. This is a positive consequence of knowing that the parameter dependence of the probability distributions of (43) is exponential. There is good hope to find distributions belonging to the exponential family also in more general models, because exact relations can be derived, even in the cases where counting walks would raise an unsurmountable problem.

Deviations from exponential distribution, as in expression (45), are due to memory effects. The walker remembers initial conditions, even if these are carefully chosen. The reason here is that the process is non-Markovian. Of course, these effects are negligible when the number of steps n is large. In many realistic models long-range interactions produce memory effects which remain important for large n. For example, in polymers the excluded volume effect causes long-range interactions. Such models are less suited for rigorous analysis. We expect that deviations from exponential dependence, found for finite n in the present model, will occur in models with long-range interactions, even in the limit of large system size.

Acknowledgments

We thank Frank den Hollander for suggesting the techniques used in section 4, and for his interest in the present work. We thank an anonymous referee for suggesting the proof of the appendix.

Appendix. Probability of odd number of reversals

$\langle \Delta_n \rangle$ is the probability that the number of reversals is odd. In other words, it is the probability that the n-th step is in the opposite direction of the initial step. Let us write

$$\langle \Delta_n \rangle = p_+^{(0)} \langle \Delta_n \rangle^+ + p_-^{(0)} \langle \Delta_n \rangle^-, \tag{A.1}$$

with $\langle \Delta_n \rangle^\pm$ the conditional probability that the number of reversals is odd under the constraint that $\sigma_0 = \pm 1$. To calculate an expression for $\langle \Delta_n \rangle^+$, look to the following recursion relation:

$$\begin{aligned} \langle \Delta_n \rangle^+ &= \epsilon_{\mathrm{L}} \langle \Delta_{n-1} \rangle^+ + (1 - \epsilon_{\mathrm{R}})(1 - \langle \Delta_{n-1} \rangle^+) \\ &= (\epsilon_{\mathrm{R}} + \epsilon_{\mathrm{L}} - 1) \langle \Delta_{n-1} \rangle^+ + 1 - \epsilon_{\mathrm{R}}. \end{aligned} \tag{A.2}$$

The solution of this equation is

$$\langle \Delta_n \rangle^+ = (1 - [\epsilon_{\mathrm{R}} + \epsilon_{\mathrm{L}} - 1]^n) p_-^{(0)}. \tag{A.3}$$

An expression for $\langle \Delta_n \rangle^-$ can be obtained analogously

$$\langle \Delta_n \rangle^- = (1 - [\epsilon_{\mathrm{R}} + \epsilon_{\mathrm{L}} - 1]^n) p_+^{(0)}. \tag{A.4}$$

Expressions (A.3) and (A.4) allow us to write $\langle \Delta_n \rangle$ as

$$\langle \Delta_n \rangle = 2 p_+^{(0)} p_-^{(0)} (1 - [\epsilon_{\mathrm{R}} + \epsilon_{\mathrm{L}} - 1]^n). \tag{A.5}$$

References

[1] Weiss G H 2002 Some applications of persistent random walks and the telegraphers equation *Physica* A **311** 381, 410

[2] Masoliver J, Lindenberg K and Weiss G H 1989 A continuous-time generalization of the persistent random walk *Physica* A **157** 891–8

[3] Berrones A and Larralde H 2001 Simple model of a random walk with arbitrarily long memory *Phys. Rev.* E **63** 031109

[4] Weiss G H and Shmueli U 1987 Joint densities for random walks in the plane *Physica* A **146** 641–9

[5] Bracher Ch 2004 Eigenfunction approach to the persistent random walk in two dimensions *Physica* A **331** 448, 466

[6] Flory P J 1969 *Statistical Mechanics of Chain Molecules* (New York: Interscience)

[7] Boguñá M, Porrà J M and Masoliver J 1999 Persistent random walk model for transport through thin slabs *Phys. Rev.* E **59** 6517–26

[8] Cwilich G A 2002 Modelling the propagation of a signal through a layered nanostructure: connections between the statistical properties of waves and random walks *Nanotechnology* **13** 274–9

[9] Miri M F and Stark H 2005 Modelling light transport in dry foams by a coarse-grained persistent random walk *J. Phys. A: Math. Gen.* **38** 3743–9

[10] Kullmann L, Kertész J and Kaski K 2002 Time-dependent cross-correlations between different stock returns: a directed network of influence *Phys. Rev.* E **66** 026125

[11] Smith S B, Finzi L and Bustamante C 1992 Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads *Science* **258** 1122–6

[12] Bustamante C, Marko J F and Siggia E D 1994 Entropic elasticity of λ-phage DNA *Science* **265** 1599–600

[13] Keller D, Swigon D and Bustamante C 2003 Relating single-molecule measurements to thermodynamics *Biophys. J.* **84** 733–8

[14] Van der Straeten E and Naudts J 2006 A one-dimensional model for theoretical analysis of single molecule experiments *J. Phys. A: Math. Gen.* **39** at press (*Preprint* math-ph/0601263)