

The Pearson random walk with unequal step sizes

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

1983 J. Phys. A: Math. Gen. 16 489

(<http://iopscience.iop.org/0305-4470/16/3/009>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 129.252.86.83

The article was downloaded on 31/05/2010 at 06:49

Please note that [terms and conditions apply](#).

The Pearson random walk with unequal step sizes

George H Weiss and James E Kiefer

Physical Sciences Laboratory, Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20205, USA

Received 3 August 1982

Abstract. There have been many analyses of Pearson random walks with equal step sizes. Several applications suggest the importance of these walks with unequal step sizes. Numerical comparisons are made of approximations obtained using the central limit theorem and the steepest descents method. These suggest that the latter generally, but not always, leads to more accurate results. We also derive an approximation that is extremely accurate at near-maximal extensions.

1. Introduction

The three-dimensional generalisation of the Pearson (1905) random walk has played a significant role in the analysis of polymer chain configurations (Volkenstein 1966, Flory 1968, Burchard and Kajiwara 1970, Yamakawa 1971). Its properties have been studied by many investigators (Lord Rayleigh 1919, Treloar 1946, Daniels 1952, Johnson 1966, Jernigan and Flory 1969, Dvořák 1972a, b, Barakat 1973). In contrast, there has been much less effort devoted to the two-dimensional random walk, although a formal solution for the probability distribution of the end-to-end distance was quickly furnished after Pearson's original inquiry (Kluyver 1906, Watson 1942). The two-dimensional Pearson walk has been used as a model for the locomotion of micro-organisms on surfaces and has application in crystallography (Srinivisan and Parthasarathy 1976). Our interest in this subject was stimulated by a recent query by Wilson (1980) relating to the distribution of the projection of a two-dimensional random walk on an arbitrary axis, in the limit in which the projection is approximately as long as the maximum extension. This problem arises in the context of crystallography, where the step sizes are generally unequal and are not randomly distributed.

The problem of approximating the probability density of a projection on an arbitrary axis appears in polymer physics, where Kuhn and Grün (1942) were the first to discuss the limit in which the projection is of the same order of magnitude as the maximal length. They used a heuristic method to find an approximate solution, later corrected by Jernigan and Flory (1969). Subsequently, Dvořák (1972a, b), Yamakawa (1971) and Daniels (1972) used the method of steepest descents to calculate the approximate densities. All of these analyses are valid when bond (or step) lengths are equal. There appears to be no comparable treatment available when step sizes differ, although similar problems arise in sound scattering in the ocean (Dyer 1970). In this paper we will do two things. The first is to compare the approximations produced by the central limit theorem and the steepest descents method (Daniels

1954) with accurate numerical solutions for the relevant probability densities. The second is to give a method for systematically approximating the probability density of the projection when it is close to the maximum extension.

2. Approximate solutions

It is well known that the central limit theorem leads to a Gaussian distribution for the sum of n random variables, when n is large and when the average step sizes are bounded. The Gaussian approximation is generally a good one when the end-to-end distance is $O(n^{1/2})$ for large n . Furthermore, the error made by using the Gaussian approximation is $O(n^{-1/2})$. It is less well known that Daniels (1954) has shown that the approximation obtained by using the method of steepest descents to evaluate the relevant integrals is valid to $O(n^{-1})$. It also appears to be true from examples that we have studied that, when step sizes are uniformly bounded, the steepest descents approximation is valid over the entire range in the variable x , whereas the Gaussian approximation can lead to serious errors at the edges of the range, i.e. when the end-to-end distance is approximately equal to the maximum possible extension. Domb and Offenbacher (1978) have applied the steepest descents method to a variety of simple random walk problems in which some of these points are illustrated numerically.

As we have mentioned, Kuhn and Grün (1942) used a heuristic method to find an approximation valid when the random walk is near full extension. In our formulation we will start by writing an exact expression for the probability density in terms of characteristic functions. In the appendix the equivalence of these two methods will be demonstrated. Let the random walk consist of steps of lengths L_1, L_2, \dots, L_n . The projection of a single step on an arbitrary axis is $L \cos \theta$, where θ is uniformly distributed in $(0, 2\pi)$. The characteristic function for a single projection is

$$\langle \exp(i\omega L \cos \theta) \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(i\omega L \cos \theta) d\theta = J_0(\omega L) \tag{1}$$

so that the probability density for the sum of n projections can be expressed as

$$\begin{aligned} p(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\omega x) \prod_{j=1}^n J_0(\omega L_j) d\omega \\ &= \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \exp(-vx) \prod_{j=1}^n I_0(vL_j) dv \end{aligned} \tag{2}$$

where the second integral is obtained from the first by the change of variables $\omega = -iv$. When n is large, and the step sizes are uniformly bounded, i.e. there is an L such that $\max_j L_j < L$, the method of steepest descents can be used to furnish an approximation to $p(x)$ (Daniels 1954). Let

$$L_T = \sum_{j=1}^n L_j; \tag{3}$$

then the (unique) real root of the equation

$$\sum_{j=1}^n \frac{L_j I_1(\bar{v}L_j)}{L_T I_0(\bar{v}L_j)} = \frac{x}{L_T} \tag{4}$$

is required for the approximation. This condition replaces the Langevin function that appears in the solution to the three-dimensional problem (Dvořák 1972a, b, Yamakawa 1971, Daniels 1972). In terms of \bar{v} , the steepest descents approximation can be written as

$$p_{SD}(x) \sim \left(2\pi \sum_j L_j^2 \psi(\bar{v}L_j) \right)^{-1/2} \exp(-\bar{v}x) \prod_{j=1}^n I_0(\bar{v}L_j) \tag{5}$$

where $\psi(x)$ is the function

$$\psi(x) = 1 - \frac{1}{x} \frac{I_1(x)}{I_0(x)} - \left(\frac{I_1(x)}{I_0(x)} \right)^2. \tag{6}$$

In all of our numerical calculations we have followed Daniels' (1954) suggestion and used the normalised form

$$p(x) \sim p_{SD}(x) \left(\int_{-L_T}^{L_T} p_{SD}(u) du \right)^{-1}. \tag{7}$$

The solution to equation (4) can be expanded in a power series in x when x/L_T is small. The first two terms of this series are

$$\bar{v} \sim \frac{2x}{\sum_j L_j^2} + \frac{\sum_j L_j^4 x^3}{(\sum_j L_j^2)^4} + \dots \tag{8}$$

When the lowest-order term is substituted into equation (5) and the Bessel function is replaced by its lowest-order approximation, the Gaussian or central limit approximation results. One would not expect the Gaussian approximation to be satisfactory at projections approaching maximum extension, whereas the steepest descents approximation does have the property of being exactly equal to zero when $x > L_T$. It is possible to derive an expansion for $p(x)$ valid near maximum extension, i.e. when x is close to L_T . For this analysis set $x = L_T - t$ and let the density function of t be denoted by $g(t)$. Equation (2) allows us to express $g(t)$ as

$$\begin{aligned} g(t) &= \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \exp(vt) \left(\prod_{j=1}^n \exp(-vL_j) I_0(vL_j) \right) dv \\ &= \mathcal{L}^{-1} \left(\prod_{j=1}^n \exp(-vL_j) I_0(vL_j) \right) \end{aligned} \tag{9}$$

where \mathcal{L}^{-1} denotes the inverse Laplace transform. We are interested in the behaviour of $g(t)$ for small t . By a Tauberian theorem for Laplace transforms (Doetsch 1943) we can infer an expansion for $g(t)$ from the large $|v|$ behaviour of the transform. For this purpose we make use of the asymptotic expansion

$$\exp(-vL) I_0(vL) \sim (2\pi Lv)^{-1/2} \left(1 + \frac{1}{8vL} + \frac{9}{128} \frac{1}{(vL)^2} + \dots \right) \tag{10}$$

to find for $g(t)$ the expansion

$$\begin{aligned} g(t) &\sim (2\pi)^{-n/2} \left(\prod_{j=1}^n L_j^{1/2} \right)^{-1} \frac{t^{n/2-1}}{\Gamma(n/2)} \left\{ 1 + \frac{1}{4} \left(\sum_j L_j^{-1} \right) \frac{t}{n} \right. \\ &\quad \left. + \frac{1}{32} \left[\left(\sum_j L_j^{-1} \right)^2 - 10 \sum_j L_j^{-2} \right] \frac{t^2}{n(n+2)} + \dots \right\}. \end{aligned} \tag{11}$$

In practice, the coefficients of this series can be generated very quickly by recursion. If we return to equation (10) and write

$$\prod_{j=1}^n \left(1 + \frac{1}{8vL_j} + \frac{9}{128} \frac{1}{(vL_j)^2} + \dots \right) = 1 + \frac{a_1^{(n)}}{v} + \frac{a_2^{(n)}}{v^2} + \dots \tag{12}$$

then the a satisfy

$$\begin{aligned} a_1^{(n+1)} &= a_1^{(n)} + 1/8L_{n+1} \\ a_2^{(n+1)} &= a_2^{(n)} + \frac{a_1^{(n)}}{8L_{n+1}} - \frac{9}{128L_{n+1}^2} \end{aligned} \tag{13}$$

and so on. These were used in calculations made to compare exact and approximate results.

3. Numerical comparisons

Greenwood and Durand (1955) compared two-dimensional results obtained by numerical integration for the equal-step case against approximations based on an expansion in terms of Hermite functions. Daniels (1972) compared the steepest descents and Gaussian approximations and exact expressions for the three-dimensional random walk with equal bond lengths, and Lyon (1970) has compared the two approximations in the two-dimensional case, without calculating an exact result to provide a standard for the comparison. We have made a three-way comparison for three five-step random walks. The step lengths in the three cases are (i) (1, 1, 1, 1, 1), (ii) (1, 1, 1, 1, 5) and (iii) (1, 2, 3, 4, 5). The accurate probability densities are most readily found by using a formula first derived by Barakat (1973) and used by Barakat and Cole (1979). The probability density can be expressed as the Fourier series

$$p(x) = \frac{1}{2L_T} \left(1 + 2 \sum_{m=1}^{\infty} C(\pi m/L_T) \cos(\pi m x/L_T) \right) \tag{14}$$

where the coefficients $C(\pi m/L_T)$ are just

$$C(\pi m/L_T) = \prod_{j=1}^n J_0(\pi m L_j/L_T). \tag{15}$$

In figure 1 we show graphs of $p(x)$ for the three cases enumerated above. The symmetry of $p(x)$ is obvious so that only values for $x > 0$ are plotted. Notice that in case (ii) the resulting $p(x)$ is bimodal. This result is not surprising as can be seen by considering the case $n = 1$ where one easily shows that

$$p(x) = [\pi(L^2 - x^2)]^{-1} \quad -L \leq x \leq L \tag{16}$$

which has a singularity at the maximum extension. Since case (ii) has a single step of length 5 that dominates the remaining steps the resulting probability density should resemble that for the case of a single step.

In figure 2 we show curves of the relative error, calculated as

$$e(x) = 1 - \frac{p_{\text{appr}}(x)}{p_{\text{exact}}(x)} \tag{17}$$

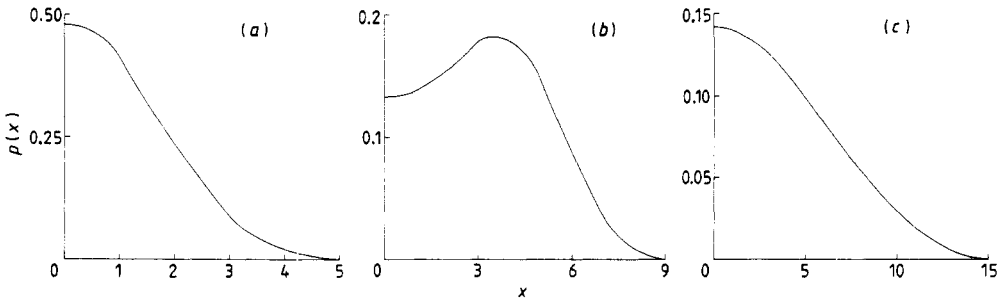


Figure 1. $p(x)$ for $x > 0$ ($p(-x) = p(x)$) for a five-step walk: (a) all step lengths equal to 1, (b) step lengths (1, 1, 1, 1, 5), (c) step lengths (1, 2, 3, 4, 5).

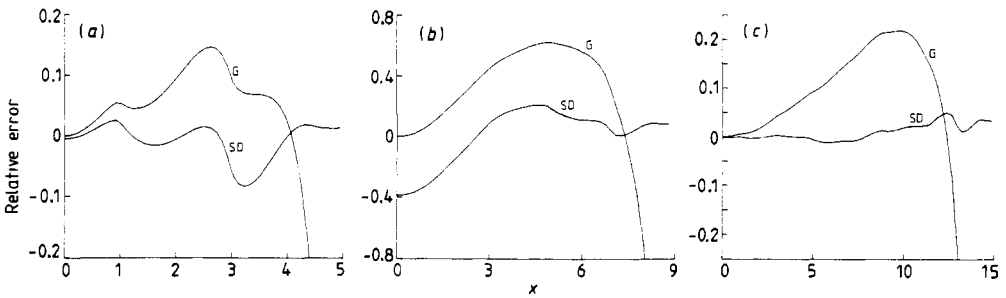


Figure 2. (a) Curves of the relative error for two approximations (SD, steepest descents; G, Gaussian) to the exact $p(x)$ for the case of an equal-length five-step walk. Notice that the relative error for the SD approximation remains bounded over the entire interval in x while the Gaussian approximation is badly in error in the maximal extension region. (b) Curves of the relative error for the bimodal density shown in figure 1(b). In this case the SD approximation is poor near the origin, but is seen to be an improvement over the Gaussian approximation at moderate and large projections. (c) Curves of the relative error for the five-step walks with step lengths (1, 2, 3, 4, 5). The SD approximation is seen to be useful over the entire range of x .

for the three cases. In each case the Gaussian approximation leads to large relative errors at the maximum extension while the steepest descents approximation appears to yield useful results over the entire interval, with the exception in one case of a region near $x = 0$. Examination of a much larger set of data shows this to be true quite generally. The endpoint approximation derived in equation (11) yields extremely accurate representations of $p(x)$ in the neighbourhood of the maximum extension. For example, with six terms of the series in equation (11) for the case (1, 1, 1, 1, 5) at $x = 7$ one finds a relative error of 1×10^{-2} using the endpoint approximation compared with a relative error equal to 4×10^{-2} obtained with the steepest descents approximation. Further, the relative error using the endpoint approximation rapidly goes to zero as x approaches 9, the maximum extension, while it approaches a nearly constant value of 8.5×10^{-2} at the endpoint in the steepest descents case. Similar results are obtained for other combinations of lengths.

One can develop a theory along parallel lines for three-dimensional freely jointed chains. However, there do not appear to be any significant problems relating to random walks with differing step sizes. It is also possible to develop an endpoint

approximation for the original Kuhn–Grün problem parallel to that derived in two dimensions. However, the Laplace transform procedure leads rather trivially to Treloar’s (1946) exact solution and is therefore uninteresting.

Appendix. Equivalence of the Kuhn–Grün approximation and the characteristic function method

In this appendix we will deal only with the case where all step lengths are equal since this corresponds to the problem dealt with by Kuhn and Grün (1942). The Kuhn–Grün calculation starts by dividing the angular interval $(-\pi, \pi)$ into sub-intervals:

$$-\pi = \theta_0 < \theta_1 < \theta_2 < \dots < \theta_m = \pi \tag{A1}$$

and $\Delta\theta_j = \theta_j - \theta_{j-1}$ so that

$$\sum_{j=1}^m \Delta\theta_j = 2\pi. \tag{A2}$$

Assume now that the $\Delta\theta$ are sufficiently small so that the projection of a single step can be adequately represented by $x = L \cos \bar{\theta}_j$ where $\theta_j \leq \bar{\theta}_j < \theta_{j+1}$. Since, by assumption, the angle of any step with respect to an arbitrary line is uniformly distributed in $(-\pi, \pi)$, it follows that the probability that its angle falls in (θ_{j-1}, θ_j) is $\Delta\theta_j/2\pi$. Supposing that the random walk has n steps, the joint probability that there are n_j angles in the interval (θ_{j-1}, θ_j) , $j = 1, 2, \dots, m$, is

$$q(\mathbf{n}) = n! \prod_{j=1}^m \left[\frac{1}{n_j!} \left(\frac{\Delta\theta_j}{2\pi} \right)^{n_j} \right]. \tag{A3}$$

Hence the probability that the projection of the random walk is exactly equal to x is given by

$$p(x) = \sum_{\{\mathbf{n}\}} q(\mathbf{n}) \tag{A4}$$

where the sum is over all \mathbf{n} subject to the constraints

$$\sum_{j=1}^m n_j = n \quad \sum_{j=1}^m n_j L \cos \bar{\theta}_j = x. \tag{A5}$$

At this point Kuhn and Grün (1942) approximate $p(x)$ by replacing the sum in equation (A4) by its largest term, and then maximise the result subject to the constraints of equations (A5). An alternative method that is less crude is to deal with the exact expression for $p(x)$ given in equation (A4) and include the second constraints of equation (A5) by means of a delta function, i.e.,

$$\begin{aligned} p(x) &= \sum_{\{\mathbf{n}\}} q(\mathbf{n}) \delta\left(\sum_j n_j L \cos \bar{\theta}_j - x\right) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\omega x) \sum_{\{\mathbf{n}\}} q(\mathbf{n}) \exp\left(i\omega \sum_j n_j L \cos \bar{\theta}_j\right) d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\omega x) \left(\sum_j \frac{\Delta\theta_j}{2\pi} \exp(i\omega L \cos \theta_j)\right)^n d\omega \end{aligned} \tag{A6}$$

where we have used an integral representation for the delta function. In the limit $m \rightarrow \infty$, $\Delta\theta_j \rightarrow 0$, it follows that $\bar{\theta}_j \rightarrow \theta_j$ and

$$\begin{aligned} \frac{1}{2\pi} \sum_j \exp(i\omega L \cos \bar{\theta}_j) \Delta\theta_j &\rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(i\omega L \cos \theta) d\theta \\ &= J_0(\omega L) \end{aligned} \quad (\text{A7})$$

which is just the characteristic function for a single projection so that equation (2) gives the exact expression for $p(x)$ in terms of characteristic functions.

References

- Barakat R 1973 *J. Phys. A: Math., Nucl. Gen.* **6** 796–804
 Barakat J and Cole J E 1979 *J. Sound Vib.* **62** 365–77
 Burchard W and Kajiwara K 1970 *Proc. R. Soc. A* **316** 185–99
 Daniels H E 1952 *Proc. R. Soc. Edinburgh A* **63** 290–311
 ——— 1954 *Ann. Math. Stat.* **25** 631–50
 ——— 1972 *Proc. 6th Berkeley Symp.* vol III (Berkeley: University of California) p 533–53
 Doetsch G 1943 *Theorie und Anwendungen der Laplace-Transformation* (New York: Dover)
 Domb C and Offenbacher E L 1978 *Am. J. Phys.* **46** 49–56
 Dvořák S 1972a *J. Phys. A: Gen. Phys.* **5** 78–84
 ——— 1972b *J. Phys. A: Gen. Phys.* **5** 85–94
 Dyer I 1970 *J. Acoust. Soc. Am.* **48** 337–45
 Flory P J 1968 *Statistical Mechanics of Chain Molecules* (New York: Wiley)
 Greenwood A J and Durand D 1955 *Ann. Math. Stat.* **26** 233–46
 Jernigan R L and Flory P J 1969 *J. Chem. Phys.* **50** 4185–200
 Johnson N L 1966 *Technometrics* **8** 303–17
 Kluyver J C 1906 *Kon. Akad. Wet. Amst.* **8** 341–50
 Kuhn W and Grün F 1942 *Kolloidzeit.* **101** 248–71
 Lyon R H 1970 *J. Acoust. Soc. Am.* **48** 145–9
 Pearson K 1905 *Nature* **72** 294
 Rayleigh J W S (Lord) 1919 *Phil. Mag.* **37** 321–47
 Srinivisan R and Parthasarathy S 1976 *Some Statistical Applications in X-ray Crystallography* (London: Pergamon)
 Treloar L R G 1946 *Trans. Faraday Soc.* **42** 77–82
 Volkenstein M V 1966 *Configurational Statistics of Polymeric Chains* (New York: Interscience)
 Watson G N 1942 *A Treatise on the Theory of Bessel Functions* 2nd edn (Cambridge: Cambridge University Press)
 Wilson A J C 1980 *Technometrics* **22** 629–30
 Yamakawa H 1971 *Modern Theory of Polymer Solutions* (San Francisco: Harper and Row)