An invariance property of generalized Pearson random walks in bounded geometries

2009 J. Phys. A: Math. Theor. 42 105002

(http://iopscience.iop.org/1751-8121/42/10/105002)

IOP Publishing — Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **42** (2009) 105002 (8pp) doi:10.1088/1751-8113/42/10/105002

An invariance property of generalized Pearson random walks in bounded geometries

Alain Mazzolo

CEA-Centre d'Etudes de Saclay, DEN/DANS/DM2S/SERMA/LTSD, 91191 Gif-sur-Yvette, France

E-mail: alain.mazzolo@cea.fr

Received 29 October 2008, in final form 9 December 2008
Published 13 February 2009
Online at stacks.iop.org/JPhysA/42/105002

Abstract
Invariance properties of random walks in bounded domains are a topic of growing interest since they contribute to improving our understanding of diffusion in confined geometries. Recently, limited to Pearson random walks with exponentially distributed straight paths, it has been shown that under isotropic uniform incidence, the average length of the trajectories through the domain is independent of the random walk characteristic and depends only on the ratio of the volume's domain over its surface. In this paper, thanks to arguments of integral geometry, we generalize this property to any isotropic bounded stochastic process and we give the conditions of its validity for isotropic unbounded stochastic processes. The analytical form for the traveled distance from the boundary to the first scattering event that ensures the validity of the Cauchy formula is also derived. The generalization of the Cauchy formula is an analytical constraint that thus concerns a very wide range of stochastic processes, from the original Pearson random walk to a Rayleigh distribution of the displacements, covering many situations of physical importance.

PACS numbers: 02.50.Ey, 05.40.Fb

1. Introduction

We consider a random walker starting his first move on the surface of a bounded object and entering the domain. How long does it take for such a walker to reach the surface again? This stochastic quantity, related to first passage problems, is of interest in many fields of science from biology [1] to reactor physics [2]. Usually, the mean length of the trajectories inside the domain depends both on the geometry of the system and on the characteristic of the random walk. However, for a diffusive random walk (i.e. a random walk that has exponentially distributed straight paths between two successive scattering events), it has been established that the mean length of the trajectories is independent of the random walk characteristics and is given indeed by a very simple formula. More precisely, for such random walks starting uniformly and isotropically on the surface S of a three-dimensional bounded domain K (of volume V), Bardsley and Dubi [3] and more recently Blanco and Fournier [4] showed that the average length $\langle L\rangle$ of the trajectories through the domain is given by

$$\langle L\rangle = 4\frac{V}{S}. \tag{1}$$

This relation admits a generalization in n-dimensions [5],

$$\langle L\rangle = \frac{2\pi\, O_{n-1}}{O_n}\frac{V}{S}, \tag{2}$$

where O_n is the surface area of the n-dimensional unit sphere ($O_n = 2\pi^{(n+1)/2}/\Gamma\left[\frac{n+1}{2}\right]$, Γ denotes the Euler gamma function). Equation (2), sometimes called the extended Cauchy formula [6], can be viewed as a generalization of the Cauchy formula [7] which was originally established for straight paths (chords) in the field of integral geometry. However, until now, only trajectories with exponential jumps have exhibited such remarkable behavior, although several authors suggested that this relation remains valid for other stochastic processes, if the random walk enters the domain with a distribution compatible with equilibrium [5, 8]. Finding such distributions remains an open problem. Here, based on integral geometry arguments, we prove for any homogeneous stochastic processes that the mean length of the trajectories is still given by the Cauchy formula, if the random walker enters the domain with a distribution that is simply related to that of the random walk. The idea consists of starting the random walk outside the domain, and by considering realizations of the stochastic process as geometric objects (broken lines) that intercept the domain. Having shown that such random walks do verify the Cauchy formula, we will derive the analytical form for the traveled distance from the boundary to the first scattering event (i.e. the distribution of the first jump inside the domain) that ensures the validity of the Cauchy formula. During this two-step process, other properties of the random walk trajectories will be established. This approach takes advantage of the integral geometry apparatus where precise results concerning the measure of the random segments constituting the elementary pieces of the random walk are well established.

2. Proof

In order to find such a distribution we consider a random walk starting uniformly and isotropically outside the domain, and to avoid technical difficulty regarding unbounded measure, at first only random walks with finite jumps (of maximal size a) are considered. It is thus sufficient to consider a volume of thickness a surrounding the domain since the random walks of interest are those that hit K at the first jump. Furthermore, we consider only random walks with isotropic reorientation corresponding to generalized Pearson random walks [9]. First of all, let us consider that all the jumps are constant and equal to l (the generalization to arbitrary bounded jumps will follow). We then distinguish two cases according to the fact that l is greater or smaller than the maximal size of the domain (Diam(K)). In the first case, formal proof involving only argument from integral geometry is given, while the second case is treated thanks to the equilibrium property mentioned before. For this purpose, we introduce the following measures (borrowed from De-Lin's book [10]) for a segment N of length l having an initial point O_1:

$$
\begin{aligned}
m_i(l) &= m\{N : N \cap \partial K \neq \varnothing, O_1 \in K\} \\
m_e(l) &= m\{N : N \cap \partial K \neq \varnothing, O_1 \notin K\} \\
m_e^{(1)}(l) &= m\{N : O_1 \notin K, N \cap \partial K \text{ has a single point}\} \\
m_e^{(2)}(l) &= m\{N : O_1 \notin K, N \cap \partial K \text{has exactly two points}\}.
\end{aligned} \tag{3}
$$

As noted by De-lin, clearly,

$$m_i(l) = m_e^{(1)}(l), \tag{4}$$

and,

$$m_e(l) = m_e^{(1)}(l) + m_e^{(2)}(l). \tag{5}$$

De-lin [10] also established that the kinematics measure $m(l)$ of a segment of length l entirely contained in K is given by

$$
\begin{aligned}
m(l) = \frac{1}{2} O_1 \ldots O_{n-1} V - \frac{O_{n-2}}{4(n-1)} O_0 O_1 \ldots O_{n-2} l S \\
+ \frac{1}{2} O_0 O_1 \ldots O_{n-2} \int_{\substack{M \cap K \neq \varnothing \\ (\sigma \leqslant l)}} (l - \sigma)\, \mathrm{d}M,
\end{aligned} \tag{6}
$$

where σ is the length of $M \cap K$ (i.e. the length of the chord intercepted by the line M), and where $\mathrm{d}M$ is the measure of random lines in $\mathbb{R}^n$.

2.1. *Case $l \geqslant Diam(K)$*

In this case, only two types of trajectories are possible. Either the process crosses the domain without collision, in which case the trajectory inside K is a chord, or the process enters K, undergoes a collision inside K, and necessarily exits the domain since the size of the jump is higher than the diameter of K. For this process, one defines P_0 as being the probability of having a chord, and thus $1 - P_0$ the probability of having a trajectory with a collision. The mean value of the trajectories $\langle L \rangle$ in K is then given by

$$\langle L \rangle = P_0 \langle \sigma_\eta \rangle + 2(1 - P_0) \langle r \rangle, \tag{7}$$

where $\langle \sigma_\eta \rangle$ indicates the mean value of a chord generated by the present distribution (labeled by the new index η) and $\langle r \rangle$ is the mean value of a ray, i.e. the average distance from a point in the domain to the frontier. The factor 2 in equation (7) comes from the symmetrical relationship equation (4). Then, the various terms of equation (7) remain to be determined. In the first place, according to the definitions of the previous measures in equation (3), P_0 is given by the ratio

$$P_0 = \frac{m_e^{(2)}(l)}{m_e(l)}, \tag{8}$$

whose numerator and denominator are worth respectively [10],

$$m_e^{(2)}(l) = 2 O_0 O_1 \ldots O_{n-2} \int_{\substack{M \cap K \neq \varnothing \\ (\sigma \leqslant l)}} (l - \sigma)\, \mathrm{d}M, \tag{9}$$

and

$$m_e(l) = O_0 O_1 \ldots O_{n-2} \frac{O_{n-2}}{n-1} l S. \tag{10}$$

By substituting the two expressions of $m_e(l)$ and $m_e^{(2)}(l)$ into equation (8), one obtains,

$$P_0 = \frac{2(n-1)}{l S\, O_{n-2}} \int_{\substack{M \cap K \neq \varnothing \\ (\sigma \leqslant l)}} (l - \sigma)\, \mathrm{d}M. \tag{11}$$

Moreover, since $l \geqslant \mathrm{Diam}(K)$, no segment of length l can entirely be contained in K and consequently $m(l) = 0$, thus from equation (6) we get

$$\int_{\substack{M\cap K\neq\varnothing \\ (\sigma\leqslant l)}} (l-\sigma)\,\mathrm{d}M = \frac{O_{n-2}}{2(n-1)} lS - \frac{O_{n-1}}{2} V. \tag{12}$$

Substituting this equality into equation (11) reduces the expression of P_0 to

$$P_0 = 1 - 2\pi \frac{O_{n-1}}{O_n} \frac{V}{Sl} = 1 - \frac{\langle\sigma\rangle}{l}, \tag{13}$$

where we used the fact that in $\mathbb{R}^n$, $2\pi O_{n-1} V/O_n S$ is precisely the mean value of the chord length distribution [7] (noted $\langle\sigma\rangle$). The two mean values in equation (7) remain to be determined. For the ray, its mean value is known, and is worth according to [2],

$$\langle r\rangle = \frac{\langle\sigma^2\rangle}{2\langle\sigma\rangle}. \tag{14}$$

To calculate the mean value of the chords generated by the present random distribution (recall that the index η labels this distribution) it is noted, following [11], that for a chord of length σ supported by the isotropic uniform random measure μ one has $f_\eta(\sigma) \propto (l-\sigma) f_\mu(\sigma)$ where the indices (μ, η) of $f_{\mu,\eta}(\sigma)$ label the various kinds of probability density functions (PDF) for the chords [12] (the notation $\langle\sigma\rangle$ corresponds to the usual average $\int_0^\infty \sigma f_\mu(\sigma)\,\mathrm{d}\sigma$). Normalizing the preceding equation leads to

$$f_\eta(\sigma) = \frac{(l-\sigma)}{l - \langle\sigma\rangle} f_\mu(\sigma), \tag{15}$$

and thanks to this expression, the mean value of the chords generated by the distribution η is calculated easily

$$\langle\sigma_\eta\rangle = \int_0^\infty \sigma f_\eta(\sigma)\,\mathrm{d}\sigma = \frac{l\langle\sigma\rangle - \langle\sigma^2\rangle}{l - \langle\sigma\rangle}. \tag{16}$$

Substituting this last result as well as the values of $\langle r\rangle$ and of P_0 into equation (7) one obtains

$$\langle L\rangle = \langle\sigma\rangle = \frac{2\pi O_{n-1}}{O_n} \frac{V}{S}, \tag{17}$$

and it proves the result in the case where $l \geqslant \mathrm{Diam}(K)$.

From the result relative to P_0 one can obtain the mean number of collisions. Indeed, since $l \geqslant \mathrm{Diam}(K)$ there cannot be more than one collision and the series giving the average number of shocks, $\langle N\rangle = \sum_{n=0}^{\infty} n P_n$ is reduced to $\langle N\rangle = 0 \times P_0 + 1 \times (1 - P_0)$ and thanks to equation (13) we immediately get the relation

$$\langle N\rangle = \frac{\langle\sigma\rangle}{l}, \tag{18}$$

similar to that established in the case of exponential jumps [13].

2.2. *Case $l \leqslant$ Diam (K)*

If jumps have a length lower than the diameter of the domain, trajectories may have an infinite number of collisions and a proof of the relation equation (17) starting from the expression equation (9) encounters difficulties. In fact, equation (9) depends explicitly on the shape of the domain whereas the awaited end result relating to the average length of the trajectories depends only on the ratio V/S. To avoid this difficulty, we make use of the equilibrium argument described in [5, 8]. Indeed, as in the preceding case, these are two kinds of trajectories, either a chord or a trajectory with collisions as shown in figure 1. For the chords, trajectories are

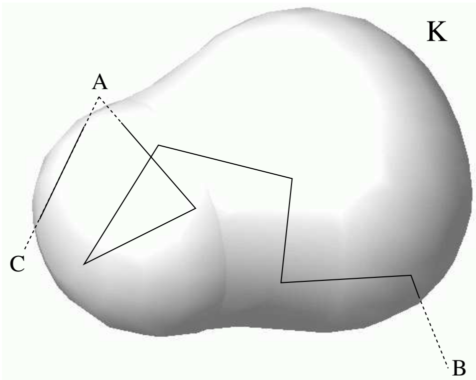

Figure 1. Examples of stochastic process realizations: AB leads to a complex trajectory inside the domain and AC to a chord.

clearly reversible because of the assumption that all initial directions are equally probable. For trajectories with collisions, the hypothesis of isotropic reorientation leads to the same conclusion. Trajectories drawn in figure 1, starting from A and ending in B or that starting from B and ending in A are equally probable. Thus all the trajectories are reversible, and since the whole process is compatible with equilibrium, we can conclude, following [8], that the mean length of the trajectories inside the domain is equal to

$$\langle L \rangle = \frac{2\pi\, O_{n-1}}{O_n} \frac{V}{S}. \tag{19}$$

Furthermore, the fact that all the trajectories are reversible involves the fact that the random walk enters and leaves the domain in the same way, therefore PDFs for the first jump inside K and for the last jump inside K are the same (however this PDF usually depends on the shape of K). This is indeed, the physical (or probabilistic) meaning of equation (4) upon which our reasoning is based.

Until now only segments of fixed length were considered, however this constraint can be relaxed without difficulty since the previous reasoning is only based on the isotropic reorientation hypothesis and on the fact that the random walk starts uniformly and isotropically. A move (segment) from a point to another now depends on the PDF for the jump but is still independent of the direction. Consequently, equation (19) is valid for any homogeneous bounded stochastic processes starting uniformly and isotropically outside the domain.

Next, we will find the corresponding probability density function $g(r)$ for the first jump through the domain. This is indeed a purely geometrical problem. Let us assume that the first jump outside the domain has a length l, of maximal size a (given by the PDF $p(l)$ on $[0; a]$). For a given point P on the surface, all the jump of length r through the surface and passing by P are those starting on the half-sphere of radius $R = l - r$ as indicated in figure 2. On the one hand, in the polar coordinate system centered in P, the uniform distribution of the initial points is proportional to $O_n R^{n-1}\, \mathrm{d}R$ and on the other hand the isotropic emission of jumps is proportional to $1/(O_n R^{n-1})$, and consequently the measure of initial points contributing to $g(r)$ is proportional to $\mathrm{d}R$ and independent of the dimension. In fact, there is a one-to-one correspondence between a point on the half-sphere of radius R and that of radius r and the problem is reduced to a one-dimensional problem. Finally, if x is the abscissa of the jump, only jumps having $x > r$ contribute to $g(r)$ as shown in figure 2 and consequently the conditional

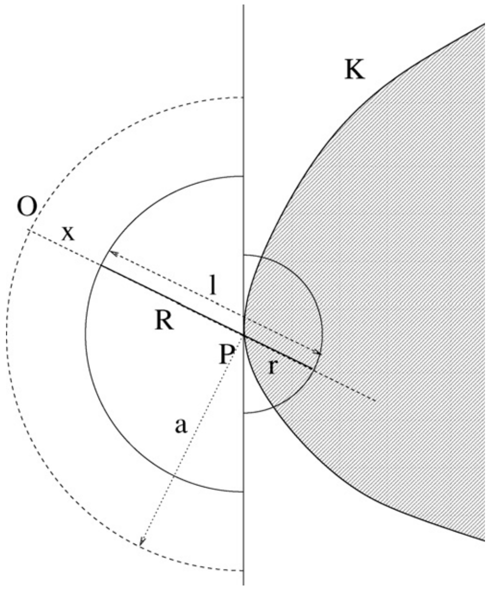

Figure 2. Initial points contributing to a first jump of length r starting on a point P on the surface: they are located on the half-sphere of radius $R = l - r$.

probability density function $g(r)$ of having a first jump of length r through K knowing that a jump through K has occurred is

$$g(r) = \frac{\int_0^a \mathrm{d}x\ p(a + r - x)\Theta(x - r)}{\int_0^a \mathrm{d}r \int_0^a \mathrm{d}x\ p(a + r - x)\Theta(x - r)}, \tag{20}$$

where $\Theta(x)$ is the Heaviside step function, or,

$$g(r) = \frac{\int_r^a \mathrm{d}x\ p(a + r - x)}{\int_0^a \mathrm{d}r \int_r^a \mathrm{d}x\ p(a + r - x)}, \tag{21}$$

which is independent of the geometry of the system. Therefore, we have established the following result: for isotropic random walks with displacements given by a probability density function $p(l)$ and entering isotropically a bounded domain with the PDF given by equation (21) then the mean length of the trajectories inside the domain is independent of the random walk characteristic and is given by the Cauchy formula (19). In particular, in two dimensions $\langle L \rangle = \pi S/P$ (with P being the perimeter of the domain) and $\langle L \rangle = 4V/S$ in three dimensions.

Since integral geometry deals with the whole nature of the object, possible extension of the techniques used in this proof to the case when the walkers are incident only on a restricted part of the boundary is not straightforward. However, when the object presents symmetries, it is sufficient for the walkers to enter the object (with the appropriate distribution given by equation (21)) on the part that is left invariant under such symmetries. For instance, for a square the walkers may enter by a side of the square. The situation is even more interesting for a sphere or a circle since in these cases, it is sufficient for the walkers to enter by a point on the surface. This last situation corresponds to the experiments realized on cockroaches entering a circular arena and described in [14].

3. Examples

(i) Constant jumps of length a, $p(l) = \delta[l - a]$.
This important case corresponds to the freely joint chain and has applications to polymer physics [15]. Substituting the expression of $p(l)$ into equation (21) leads immediately to

$$g(r) = \frac{1}{a}. \tag{22}$$

Consequently, a freely joint chain entering the domain with a uniform law has its average length inside the domain given by the Cauchy formula.

(ii) Power law on the interval $[0; a]$: $p(l) = \frac{\alpha+1}{a^{\alpha+1}} l^{\alpha}$ with $\alpha \geqslant 0$.
From equation (21) we get

$$g(r) = \frac{1}{a}\left(\frac{\alpha+2}{\alpha+1}\right)\left[1 - \left(\frac{r}{a}\right)^{\alpha+1}\right]. \tag{23}$$

In particular, when $\alpha = 0$, uniform jumps on $[0; a]$ entering the domain with the distribution $g(r) = 2/a(1 - r/a)$ have the mean value of their trajectories in the domain given by the Cauchy formula.

Those examples concern bounded stochastic processes. However, equation (21) is not limited to these cases and is still valid assuming that the integral at the denominator of this equation exists (the integral at the numerator is bounded by 1 and always converges).

(iii) Rayleigh distribution $p(l) = \frac{l}{\sigma^2} \exp\left(\frac{-l^2}{2\sigma^2}\right), l \geqslant 0$.
This distribution serves for modelizing the movement of soil insects [16] and also has applications in acoustics [17], substituting the expression of $p(l)$ into equation (21) leads to

$$g(r) = \sqrt{\frac{2}{\sigma^2 \pi}} \exp\left(\frac{-r^2}{2\sigma^2}\right). \tag{24}$$

(iv) Diffusive random walks, $p(l) = (1/\lambda)\exp(-l/\lambda), l \geqslant 0$.
This case corresponds to that originally studied by Blanco and Fournier [4], substituting the expression of $p(l)$ into equation (21) and taking the limit $a \to \infty$ gives,

$$g(r) = \frac{1}{\lambda} \exp\left(\frac{-r}{\lambda}\right), \tag{25}$$

and we recover this peculiarity of diffusive random walks, that the average length of the trajectories satisfies the Cauchy formula when they enter the domain with the same law as that inside [2, 4, 5]. This reflects the well-known memoryless property of exponential distributions (diffusive random walks have no memory of where they started).

4. Conclusion

To conclude, we have addressed a bounded version of the century old Pearson random walk problem [9]. A walker enters a finite domain under isotropic uniform incidence in a straight line given by the probabilistic law equation (21). He then turns through any angle whatever and walks a distance given by the probabilistic law $p(l)$ in a second straight line. He repeats this process until he crosses the boundary of the domain for the first time. Then, the mean length of the trajectories in the domain is independent of the random walk characteristics. This average quantity depends only on the geometry of the system and is given by the Cauchy formula, in particular it is worth $4V/S$ in three dimensions. Valid for all stochastic processes with finite jumps (thus including a very large variety of physical processes), this exact relation is a strong constraint that links short and long trajectories in confined geometries. Our guess is that equation (21) is the only way to start the random walk that leads to a mean length of the trajectories inside the domain given by the Cauchy formula.

Acknowledgments

It is a pleasure to thank Zoë Ascoli for reading the paper and Cathy Dubois for her enlightening comments. This work was partially supported by EDF (Electricité de France) and AREVA-NP in the framework of the MACOE project.

References

[1] Condamin S, Benichou O, Tejedor V, Voituriez R and Klafter J 2007 *Nature* **450** 77
[2] Mazzolo A 2004 *Europhys. Lett.* **68** 350
[3] Bardsley J N and Dubi A 1981 *SIAM J. Appl. Math.* **40** 71
[4] Blanco S and Fournier R 2003 *Europhys. Lett.* **61** 168
[5] Benichou O, Coppey M, Moreau M, Suet P H and Voituriez R 2005 *Europhys. Lett.* **70** 42
[6] Mazzolo A, Roesslinger B and Gille W 2003 *J. Math. Phys.* **44** 6195
[7] Santaló L A 1976 *Integral Geometry and Geometric Probability* (Reading, MA: Addison-Wesley)
[8] Blanco S and Fournier R 2006 *Phys. Rev. Lett.* **97** 230604
[9] Pearson K 1905 *Nature* **72** 294
[10] De-lin R 1994 *Topics in Integral Geometry* (Singapore: World Scientific)
[11] Mazzolo A 2004 *J. Phys. A: Math. Gen.* **37** 7095
[12] Coleman R 1969 *J. Appl. Probab.* **6** 430
[13] Mazzolo A 2005 *Ann. Nucl. Energy* **32** 549
[14] Jeanson R, Blanco S, Fournier R, Deneubourg J L, Fourcassie V and Theraulaz G 2003 *J. Theor. Biol.* **225** 443
[15] Flory P J 1969 *Statistical Mechanics of Chain Molecules* (New York: Wiley)
[16] Wiktorsson M, Rydén T, Nilsson E and Bengtsson G 2004 *J. Theor. Biol.* **231** 497
[17] Gaines S and Denny M 2002 *Chance in Biology: Using Probability to Explore Nature* (Princeton, NJ: Princeton University Press)