Random Walks in d-Dimensions: Do They Return to the Origin? (And various algebraic solutions)

Rick Bradford, revised 5/12/22

Contents

1. Definition of a Random Walk

1.1 One Dimension

A random walk in 1D is defined by taking a sequence of N steps, starting from the origin, each step being either +1 or -1.

1.2 d-Dimensions: AND versus OR Cases

A random walk in d-dimensions may be defined as a simultaneous and uncorrelated random walk in each Cartesian direction. Thus, in 2D, there are four possible steps (+1,+1), (+1, -1), (-1,+1), (-1,-1). This can be regarded as the "AND" case, because each step involves a step in x-direction and a step in the y-direction.

One can alternatively define, for dimensions greater than 1, an "OR" case in which every step consists of a step in the x-direction or a step in the y-direction or a step in the z-direction, etc. So, in 2D, the possible steps would be (1,0), (-1,0), (0,1), (0,-1).

Despite initial appearance there is no difference between the "AND" and "OR" cases in 2D. In 2D one merely has to rotate by 45 degrees and, other than the irrelevant change of scale by

$\sqrt{2}$, they transform into each other. However, in dimensions 3 and greater the "AND" and "OR" cases are different.

2. Polya's Random Walk Theorem

> Polya's Random Walk Theorem: In 1D and 2D the probability that the random path returns to the origin at some point tends asymptotically to unity for large N. In 3D and higher dimensions the probability of at least one return to the origin is less than 1.

(Redfield, 1927; Polya 1937).

This may be paraphrased by saying that, in 1D and 2D, the random walk "always returns to the origin", but in 3D and higher dimensions it does not.

Note that this does not mean there are no paths in 1D or 2D that fail to return to the origin. Trivially there are (many – infinitely many as $N \to \infty$). But, in 1D and 2D, the ratio of the number of such paths to the total number of possible paths tends to zero for $N \to \infty$.

3. Variant Problems

Two variant problems are of interest. The first is to consider two people (or birds) to execute stochastically equivalent random walks from the same origin, starting at the same time. What is the probability they will collide? It turns out to be the same problem in disguise as one random walker returning to the origin (see Appendix B). The second variant is to impose a "ground" so that one chosen Cartesian coordinate is not allowed to go negative. This simulates, for example, a bird in flight, the bird being unable to have a vertical altitude less than zero. This is dealt with in Appendix C and (I argue) also is the same problem in disguise – although this is disputed.

4. Proof of Polya's Theorem (for the "AND" case)

The proof of which cases return to the origin and which don't depends on two things,

[1] If the random walk "always returns to the origin", in the above sense that the probability of doing so tends to 1 for $N \to \infty$, then it does so an infinite number of times in an infinite number of steps.

[2] And so it follows that, if you can evaluate the expectation value, E, for the number of returns to the origin (in an infinite number of steps) then whether E is infinite (divergent) or finite determines whether it always returns, or not, respectively.

If [1] is not obvious, consider N_1 steps, followed by another N_1 steps, followed by another N_1 steps…N_2 times, so there are a total of $N_1 N_2$ steps. Letting $N_1 \to \infty$ means that every one of the N_2 sequences will have at least one return to the origin, so that there are at least N_2 returns to the origin overall. Now we can let $N_2 \to \infty$ so that there must be an infinite number of returns to the origin. (If that bothers you, let N_1 and N_2 tend to infinity together, in proportion).

An expression for the expectation value, E, of the number of returns to the origin can be found very simply because it is just the sum over all N of the probability of being at the origin at step N. In 1D, the probability of the random walk having r forward steps and hence

$N-r$ backwards steps is just the binomial coefficient divided by the total number of possible walks, i.e., 2^N,

$$\frac{N!}{2^N r!(N-r)!}$$

The walk is at the origin on step N if $r = N/2$ (and hence can only happen for an even number of steps). So the probability of being at the origin at step N is,

$$\frac{N!}{2^N \ \left((N/2)!\right)^2}$$

In d-dimensions, (and for the "AND" case) the probability of being at the origin on step N is the product of the probabilities of being at the x-origin and the y-origin and the z-origin, etc., simultaneously, as so is,

$$\left(\frac{N!}{2^N \ \left((N/2)!\right)^2}\right)^d$$

And so,

$$E = \sum_{N=0}^{\infty} \left(\frac{N!}{2^N \ \left((N/2)!\right)^2}\right)^d \qquad (1)$$

For the proof we only need to know if the sum in (1) converges or diverges. If it diverges then the walk "always returns to the origin"; if it converges it does not.

So we can use Stirling's approximation, $n! \approx \sqrt{2\pi n} \cdot n^n e^{-n}$, because it suffices to know how (1) behaves for very large N. Substituting in (1) gives,

$$\left(\frac{N!}{2^N \ \left((N/2)!\right)^2}\right)^d \to \left(\sqrt{\frac{2}{\pi N}}\right)^d \qquad (2)$$

Thus in 1D the terms in the sum (1) fall off only as $1/\sqrt{N}$ and so the sum diverges. Similarly, in 2D the terms reduce as $1/N$ which is also divergent. Consequently, the 1D and 2D cases "always return" to the origin.

In 2D the "AND" case is the same as the "OR" case so this establishes the result for both.

In contrast, in 3D or higher dimensions, the sum in (1) is convergent as the terms reduce as $1/N^{3/2}$ or faster. So these cases do not always return to the origin. However the above proof applies only for the "AND" case. The "OR" cases for dimensions 3 and above is treated below.

5. Algebraic Solution in 1D

Closed form expressions for key probabilities can be derived for the 1D case. Define the following quantities, and interpret in terms of "AND" steps,

$Q_{2N}(p)$ = the number of paths with $2N$ steps which have exactly p positive steps (and hence exactly $2N-p$ steps which are negative, $Q_{2N}(p) = \frac{(2N)!}{p!(2N-p)!}$.

P_{2N} = the number of paths which land on the origin at step $2N$ = $Q_{2N}(N) = \frac{(2N)!}{(N!)^2}$.

$\mathfrak{N}_{2N}$ = the number of paths which are at a positive position (strictly > 0) at all steps up to and including step $2N$.

Hence, the number of paths which do not return to the origin in $2N$ steps is $2\mathfrak{N}_{2N}$. It is far from obvious, but is proved below, that $2\mathfrak{N}_{2N} = P_{2N} = \frac{(2N)!}{(N!)^2}$.

Z_{2N} = the number of paths which have been at the origin one or more times up to and including step $2N$ (not counting the starting position).

2^{2N} = the total number of paths with $2N$ steps.

$R_{2N}(2r)$ = the number of paths which are at position $2r$ at step number $2N$ ***and*** have been at positive positions (strictly > 0) on every previous step.

S_{2N} = the number of paths which are at the origin on step $2N$ but at no point previously

5.1 Derivation of Expressions for Key Quantities and Proof that $2\mathfrak{N}_{2N} = P_{2N}$

The total number of paths comprises (i) all paths which are always positive, plus, (ii) all paths which are always negative, plus, (iii) all paths which hit the origin one or more times. Hence,

$$2\mathfrak{N}_{2N} + Z_{2N} = 2^{2N} \qquad \text{(Equ.1)}$$

The strategy is to set up a recursion relation in the quantities $R_{2N}(2r)$ and then to solve it.

Consider how the number of paths which are at position $2r$ at step $2(N+1)$ originate from the paths at step $2N$. Each path at step $2N$ which is at position $2(r \pm 1)$ gives rise to one path which is at position $2r$ at step $2(N+1)$. Each path at step $2N$ which is at position $2r$ gives rise to two paths which are at position $2r$ at step $2(N+1)$. Hence we have the recursion relation,

$$R_{2(N+1)}(2r) = R_{2N}\big(2(r-1)\big) + 2R_{2N}(2r) + R_{2N}\big(2(r+1)\big) \qquad \text{(Equ.2)}$$

When $r = 1$ the first term on the RHS is zero, i.e., $R_{2N}(0) = 0$ because we are considering paths which are never at zero.

Equ.2 gives us a "Pascal-like" triangle which solves for $R_{2N}(2r)$, see Table 1. The rule, which is just Equ.2 in words is "double the number above and add the numbers either side".

Table 1: Solution for $R_{2N}(2r)$: "Pascal-like" Triangle

	r							
N	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1							
2	2	1						
3	5	4	1					
4	14	14	6	1				
5	42	48	27	8	1			
6	132	165	110	44	10	1		
7	429	572	429	208	65	12	1	
8	1430	2002	1638	910	350	90	14	1

Just as the binomial coefficients solve Pascal's triangle, the above variant is solved by the following expression,

$$R_{2N}(2r) = \frac{r(2N)!}{N(N+r)!(N-r)!} \quad \text{(Equ.3)}$$

That this is the correct solution is proved by substituting into Equ.2 and showing it to be an identity. In what follows we shall only need,

$$R_{2N}(2) = \frac{(2N)!}{N!(N+1)!} \quad \text{(Equ.4)}$$

From (4) we immediately get the solution for S_{2N} as every path which has always been positive and is at position 2 on step $2(N-1)$ gives one path which hits the origin for the first time at step $2N$ (ditto for always-negative steps). So we have,

$$S_{2N} = \frac{2\big(2(N-1)\big)!}{N!(N-1)!} \quad \text{(Equ.5)}$$

The next step is to set up a recursion relation for Z_{2N}. To do so note that every path that has returned to the origin at least once by step $2N$ gives rise to four such paths by step $2(N+1)$. The total number of paths that have returned to the origin at least once by step $2(N+1)$ comprises these plus one further path for each path that was always positive up to step $2N$ and at position 2 at step $2N$ (plus the negative equivalents). Hence we have,

$$Z_{2(N+1)} = 4Z_{2N} + 2R_{2N}(2) \quad \text{(Equ.6)}$$

Given Equ.4 for $R_{2N}(2)$ it is readily checked that the solution for the recursion relation, Equ.6, subject to the initial condition $Z_2 = 2$, is,

$$Z_{2N} = 2^{2N} - \frac{(2N)!}{(N!)^2} \quad \text{(Equ.7)}$$

With this solution for Z_{2N} established, Equ.1 now gives,

$$2\mathfrak{R}_{2N} = 2^{2N} - Z_{2N} = 2^{2N} - \left(2^{2N} - \frac{(2N)!}{(N!)^2}\right) = \frac{(2N)!}{(N!)^2} = P_{2N} \quad \text{(Equ.8)}$$

QED

6. Calculation of Return Probability in ≥ 3 Dimensions, "OR" Case

I believe the proof along these lines was first given by Novak (2014). Any errors are mine.

Let p_n be the probability of returning to the origin on step n but not before. This does not preclude the possibility of multiple returns to the origin after step n. The starting position does not count so $p_0 = 0$. These are probabilities of mutually exclusive events, so the probability of returning to the origin once or more is thus,

$$p = \sum_{n=0}^{\infty} p_n = \sum_{n=1}^{\infty} p_n \qquad \text{(Equ.9)}$$

The total number of paths after n steps in d-dimensions is $(2d)^n$, so if K_n is the number of paths which are at the origin for the first time on step n then,

$$p_n = \frac{K_n}{(2d)^n} \qquad \text{for } n > 0 \qquad \text{(Equ.10)}$$

Similarly, if L_n is the number of paths which are at the origin on step n, not necessarily for the first time, and q_n is the probability of being at the origin on step n, then,

$$q_n = \frac{L_n}{(2d)^n} \qquad \text{for } n > 0 \qquad \text{(Equ.11)}$$

For later convenience we set $q_0 = L_0 = 1$.

The L_n paths which are at the origin on step n can be considered as composed of paths which are at the origin for the first time on step $k < n$ followed by any path which returns to the origin after exactly $n - k$ further steps, and where k can be anywhere in the range $[0, n]$. We thus have,

$$L_n = \sum_{k=1}^{n} K_k L_{n-k} \qquad \text{(Equ.12)}$$

Dividing by $(2d)^n$ this becomes,

$$q_n = \sum_{k=1}^{n} p_k q_{n-k} = \sum_{k=0}^{n} p_k q_{n-k} \qquad \text{(Equ.13)}$$

Note the utility of setting $q_0 = L_0 = 1$ as the above expression is thus correct as regards inclusion of the K_n paths which reach the origin for the first time on step n.

We now define the generating functions,

$$P(z) = \sum_{n=0}^{\infty} p_n z^n = \sum_{n=1}^{\infty} p_n z^n \qquad \text{and} \qquad Q(z) = \sum_{n=0}^{\infty} q_n z^n = 1 + \sum_{n=1}^{\infty} q_n z^n \qquad \text{(Equ.14)}$$

The variable z has no physical meaning (as far as I am aware). It is merely a useful device. The identity (13) can now be written,

$$P(z)Q(z) = Q(z) - 1 \qquad \text{(Equ.15)}$$

Proof

$$P(z)Q(z) = \sum_{n=1}^{\infty} p_n z^n \sum_{m=0}^{\infty} q_m z^m = \sum_{N=1}^{\infty} \sum_{k=0}^{N} p_k q_{N-k} z^N = \sum_{N=1}^{\infty} q_N z^N = Q(z) - 1$$

QED.

Trivially rewriting (15),

$$P(z) = 1 - \frac{1}{Q(z)} \qquad \text{(Equ.16)}$$

But, from (14) and (9), the quantity we are attempting to calculate is,

$$p = P(1) = 1 - \frac{1}{Q(1)} \qquad \text{(Equ.17)}$$

This is quite a remarkable relationship which is not at all intuitively obvious (to me, anyway). Note that $Q(1) = 1 + \sum_{n=1}^{\infty} q_n$ and so clearly $Q(1) > 1$. In fact, because from (13) $q_n > p_n$ for all n it follows that $\sum_{n=1}^{\infty} q_n > p$ and so $Q(1) > 1 + p$, which also follows directly from (17).

In 1D and 2D, we shall see that $Q(1) \to \infty$ as the number of steps $N \to \infty$, so $p \to 1$. However, for $d \geq 3$ we shall find that $Q(1)$ tends to a finite asymptote, so $p < 1$.

Using (11) we can write $Q(z) = L\left(\frac{z}{2d}\right)$ where,

$$L(z) = \sum_{n=0}^{\infty} L_n z^n \qquad \text{(Equ.18)}$$

We now consider the so-called "exponential generating function", $E(z)$, defined by,

$$E(z) = \sum_{n=0}^{\infty} L_n \frac{z^n}{n!} \qquad \text{(Equ.19)}$$

The utility of this apparently arbitrary construction lies in the simplicity of its dependence upon dimension. Consider 2D, and take the "OR" interpretation of the random walk. Any "loop" of length n (i.e., any path which is at the origin on step n) must consist of k horizontal steps which add to zero (i.e., a 1D "loop" of length k), plus $n - k$ vertical steps which add to zero (i.e., a 1D "loop" of length $n - k$), where k is any number in the range $[0, n]$. Denote by $L_n^{(1)}$ the 1D value for L_n, i.e., the number of paths in 1D which are at the origin on step n, not necessarily for the first time. Hence, in 2D, the number of paths for a given k is $L_k^{(1)}$ times $L_{n-k}^{(1)}$ times the number of ways the k horizontal steps can be placed in the total of n steps, i.e., the binomial coefficient $n!/k!\,(n-k)!$. So we get, for 2D,

$$L_n^{(2)} = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} L_k^{(1)} L_{n-k}^{(1)} \qquad \text{(Equ.20)}$$

where the superscript refers to the dimension. The utility of the exponential generating function, $E(z)$, is that,

$$E^{(2)}(z) = \left(E^{(1)}(z)\right)^2 \qquad \text{(Equ.21)}$$

Proof

$$\left(E^{(1)}(z)\right)^2 = \sum_{n,m=0}^{\infty} L_n^{(1)} \frac{z^n}{n!} L_m^{(1)} \frac{z^m}{m!} = \sum_{S=0}^{\infty} \sum_{n=0}^{S} L_n^{(1)} L_{S-n}^{(1)} \frac{z^S}{S!} \cdot \frac{S!}{n!(S-n)!} = \sum_{S=0}^{\infty} L_S^{(2)} \frac{z^S}{S!}$$

and the last expression is just $E^{(2)}(z)$, by (19). QED.

This extends to any number of dimensions, namely,

$$E^{(d)}(z) = \left(E^{(1)}(z)\right)^d \qquad \text{(Equ.22)}$$

For example in 3D we have,

$$L_n^{(3)} = \sum_{r=0}^{n-k} \sum_{k=0}^{n} \frac{n!}{k!r!(n-k-r)!} L_k^{(1)} L_r^{(1)} L_{n-k-r}^{(1)} \quad \text{(Equ.23)}$$

And so,

$$\left(E^{(1)}(z)\right)^3 = \sum_{n,m,k=0}^{\infty} L_n^{(1)} \frac{z^n}{n!} L_m^{(1)} \frac{z^m}{m!} L_k^{(1)} \frac{z^k}{k!}$$

$$= \sum_{S=0}^{\infty} \sum_{n=0}^{S-m} \sum_{m=0}^{S} L_n^{(1)} L_m^{(1)} L_{S-n-m}^{(1)} \frac{z^S}{S!} \cdot \frac{S!}{n!m!(S-n-m)!} = \sum_{S=0}^{\infty} L_S^{(3)} \frac{z^S}{S!} = E^{(3)}(z) \quad \text{(Equ.24)}$$

Identity (22) is what makes the exponential generating function so useful. However, there is a further property that makes it more useful still, namely the simple relationship between $E^{(1)}(z)$ and a (modified) Bessel function.

Recalling that $L_n^{(1)}$ is the number of paths in 1D which are at the origin on step n, but not necessarily for the first time, then as $n/2$ steps must be positive and the same number must be negative then,

$$L_n^{(1)} = \frac{n!}{\left(\left(\frac{n}{2}\right)!\right)^2} \quad \text{(Equ.25)}$$

(n must be even, of course). Hence, from (19),

$$E^{(1)}(z) = \sum_{k=0}^{\infty} \frac{z^{2k}}{(k!)^2} \quad \text{(Equ.26)}$$

The power series which defines the modified Bessel function of the first kind and order α is,

$$I_\alpha(z) = \sum_{k=0}^{\infty} \frac{\left(\frac{z}{2}\right)^{2k+\alpha}}{k!\Gamma(k+\alpha+1)} \quad \text{(Equ.27)}$$

So, using $\Gamma(k+1) = k!$, we find the exponential generating function in 1D in terms of the zeroth order modified Bessel function,

$$E^{(1)}(z) \equiv I_0(2z) \quad \text{(Equ.28)}$$

And so we immediately have from (22),

$$E^{(d)}(z) \equiv \left(I_0(2z)\right)^d \quad \text{(Equ.29)}$$

The final step is to appreciate that the exponential generating function, $E(z)$, (19), can be converted back to its parent generating function, $L(z)$, (18), using a Borel transformation, defined as

$$\tilde{E}(z) \equiv \int_0^{\infty} E(tz)e^{-t}dt \quad \text{(Equ.30)}$$

From which we find that: $L(z) \equiv \tilde{E}(z)$ (Equ.31)

Proof

$\tilde{E}(z) \equiv \int_0^{\infty} \sum_{n=0}^{\infty} L_n \frac{t^n z^n}{n!} e^{-t}dt = \sum_{n=0}^{\infty} L_n \frac{z^n}{n!} \cdot n! = \sum_{n=0}^{\infty} L_n z^n = L(z)$, QED,

where we have used $\int_0^{\infty} t^n e^{-t}dt = n!$ which is readily shown using integration by parts.

Hence we have,

$$L(z) \equiv \int_0^\infty E(tz)e^{-t}dt = \int_0^\infty \left(I_0(2tz)\right)^d e^{-t}dt \qquad \text{(Equ.32)}$$

Recalling that $Q(z) = L\left(\frac{z}{2d}\right)$ we thus have,

$$Q(z) = \int_0^\infty \left(I_0\left(\frac{tz}{d}\right)\right)^d e^{-t}dt \qquad \text{(Equ.33)}$$

Recall that the quantity we need, when it is finite, is $Q(1)$ which is thus,

$$Q(1) = \int_0^\infty \left(I_0\left(\frac{t}{d}\right)\right)^d e^{-t}dt \qquad \text{(Equ.34)}$$

The advantage of this formulation is that the asymptotic behaviour of the modified Bessel function, $I_0(z)$, is well known, namely,

$$I_0(z) \to 0.39894\frac{e^z}{\sqrt{z}} \qquad \text{(Equ.35)}$$

Hence the integrand of (34) becomes, for sufficiently large t, $0.39894^d\left(\frac{d}{t}\right)^{d/2}$. Hence, for $d = 1$ or $d = 2$, (34) is divergent and $Q(1) \to \infty$ and so, from (17), $p \to 1$, confirming our earlier proof. For $d \geq 3$ the integral (34) is convergent and $p < 1$. This form is convenient for rapid, accurate numerical evaluation as computing platforms, such as Python, include the modified Bessel functions as standard functions.

In 3D the result can be expressed as a product of four gamma functions,

$$Q^{(3)}(1) = \frac{\sqrt{6}}{32\pi^3}\Gamma\left(\frac{1}{24}\right)\Gamma\left(\frac{5}{24}\right)\Gamma\left(\frac{7}{24}\right)\Gamma\left(\frac{11}{24}\right) = 1.5163860592 \qquad \text{(Equ.36)}$$

Which then gives $p = 1 - \frac{1}{Q(1)} = 0.34053733$.

Numerical evaluation of (34) using Python-Scipy.Special gives the results up to dimension 15 given Table 2 below.

Note that the integral (34) from some large $t = T$ to infinity, such that (35) is accurate, is given by $\frac{\left(0.39894\sqrt{d}\right)^d}{\left(\frac{d}{2}-1\right)T^{\left(\frac{d}{2}-1\right)}}$ and this was used after a million integration steps up to $T = 100d$, which produced agreement with (36) to six decimal places in the 3D case.

6.1 References

J. Novak. Pólya's Random Walk Theorem. American Mathematical Monthly, 121:711-716, October 2014.

7. Calculation of Return Probability in ≥ 3 Dimensions, "AND" Case

An efficient algorithm for this calculation is provided by Parsons' recursion formula. Derivation of Equs.37 & 38 below are given in Appendix A.

Denoting the number of paths at the origin at step $2n$ by $P(2n)$, and by $S(2n)$ the number of paths at the origin for the first time at step $2n$, it follows immediately that,

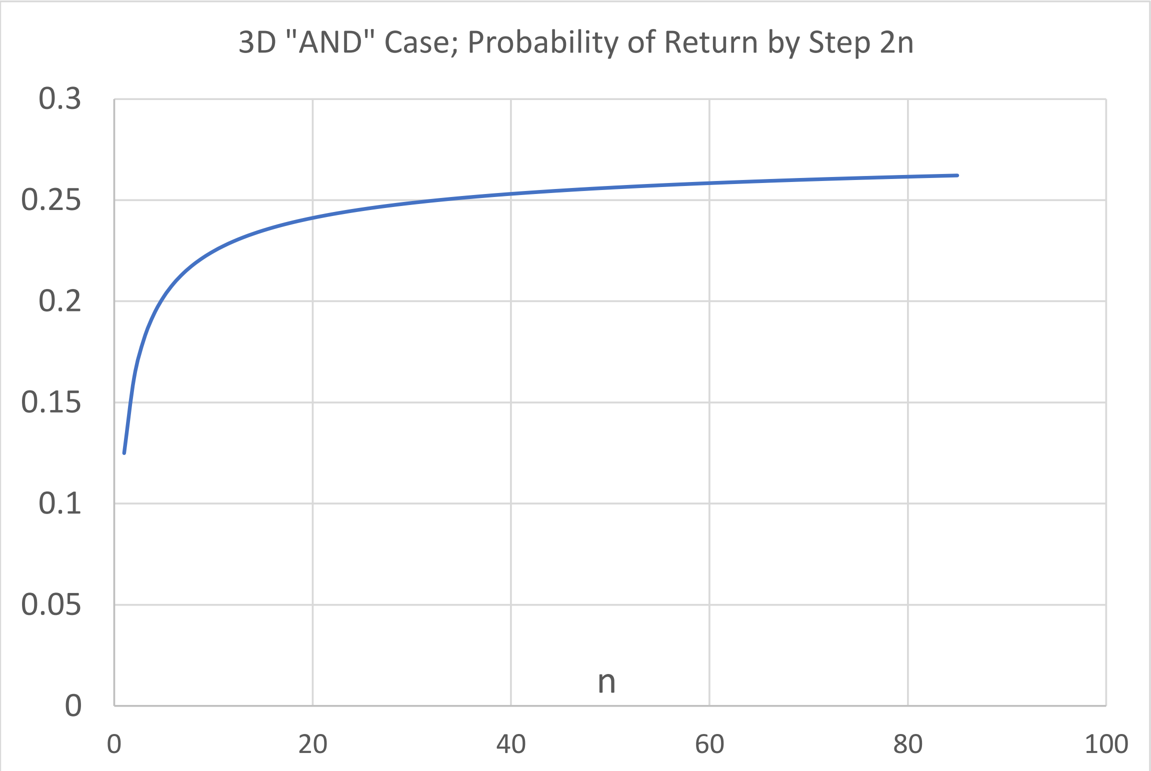
$$S(2n) = P(2n) - \sum_{r=1}^{n-1} P(2r)S\big(2(n-r)\big) \qquad \text{(Equ.37)}$$

But $P(2n) = \left(\frac{(2n)!}{n!^2}\right)^d$ and hence (37) allows $S(2n)$ to be found recursively. The probability of being at the origin for the first time is then $\Delta p(n) = S(2n)/T(2n)$ where $T(2n) = 2^{2dn}$ is the total number of paths of length $2n$. The cumulative sum of these probabilities is then the probability of returning to the origin one or more times by step $2n$.

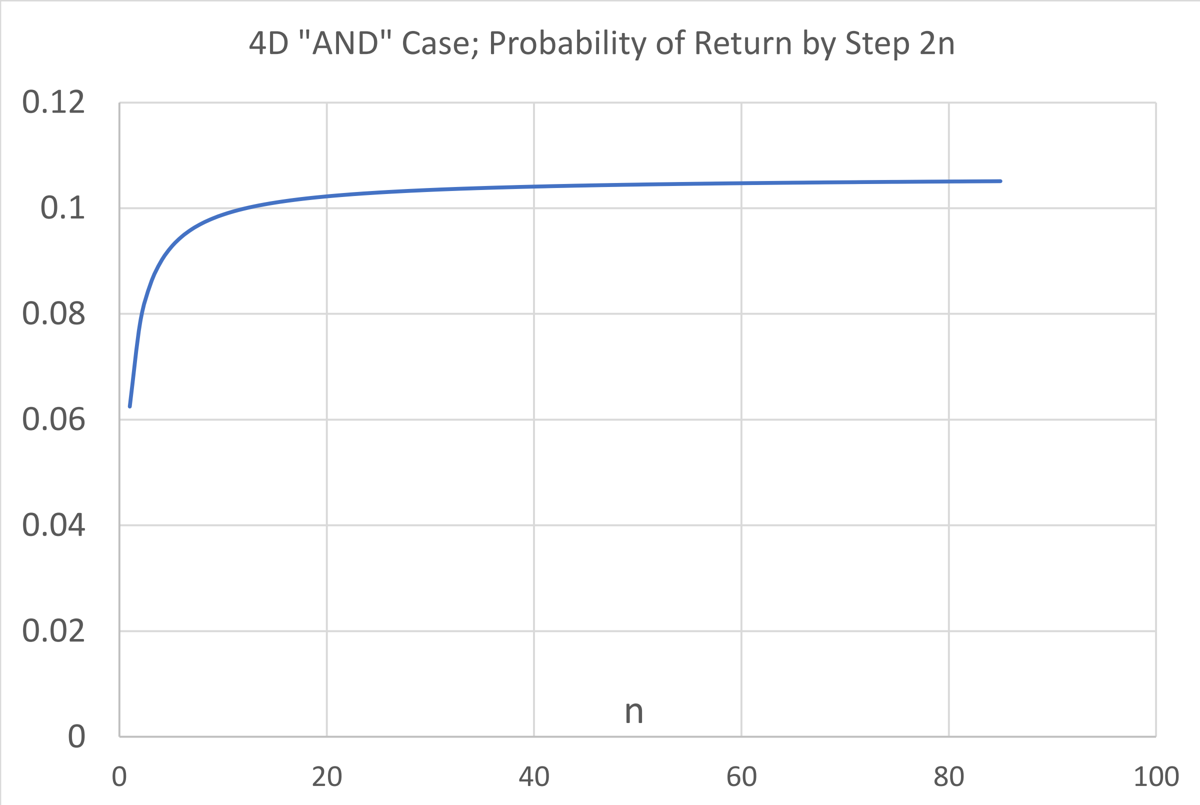
In practice this sum can be found up to some maximum n and the rest of the sum to infinity found by assuming the increments of probability reduce in proportion to $r^{-d/2}$. Integration, as an approximation of the rest of the sum, gives a correction to be added of,

$$\frac{n}{\frac{d}{2}-1}\Delta p(n) \qquad \text{(Equ.38)}$$

where $\Delta p(n)$ is the last evaluated increment. The results of using this efficient algorithm for the "AND" case are given in Table 2. The probability of return is smaller for the "AND" case than for the "OR" case of the same dimensionality. The build-up of the probability with increasing number of steps is shown below, for the 3D case. The correction integral after 2n = 170 is as large as 0.020, so quite significant.



Correction terms become rapidly smaller in higher dimensions, which converge more rapidly, e.g., see d = 4 and d = 5 below,



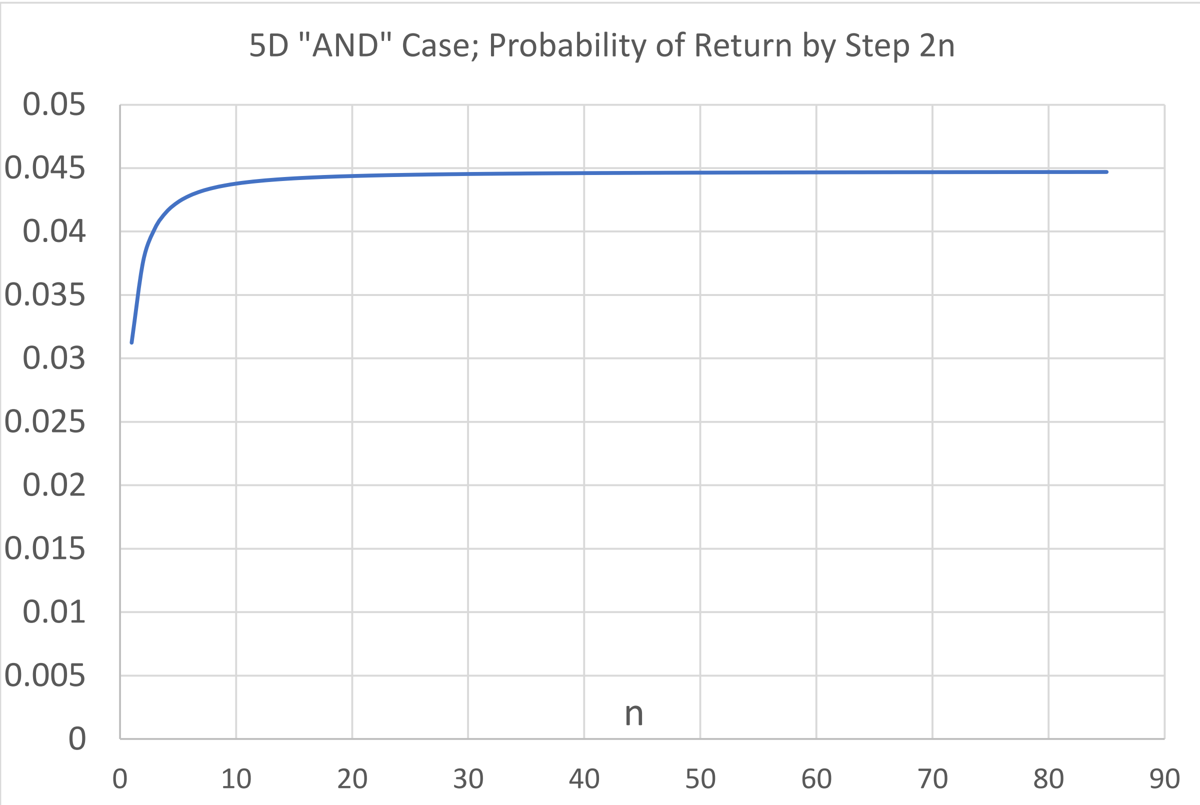


Table 2: Numerical Evaluation of Return Probabilities Using Equs (34) or (37)

d	p	
	"OR" [(1)]	**"AND"** [(2)]
3	0.340537	0.2823
4	0.1932	0.10605
5	0.1352	0.04473
6	0.1047	0.02005
7	0.08584	0.009318
8	0.07291	0.004431
9	0.06345	0.002139
10	0.05620	0.001043
11	0.05046	0.000512
12	0.04579	0.000253
13	0.04192	0.000125
14	0.03866	0.000062
15	0.03587	0.000031

[(1)]Using Novak integral-over-Bessel function method, Equ.(34)
[(2)]Using Parsons' recursion formula, Equ.(37) plus (Equ.38)

Appendix A: Derivation of Equations (37) & (38)

$P(2n)$ = the number of paths at the origin at step $2n$

$S(2n)$ = the number of paths at the origin for the first time at step $2n$

Any path which is at the origin at step $2n$ is either at the origin for the first time on that step or there is an integer r where $1 \leq r < n$ such that the path first returned to the origin on step 2r and subsequently returned to the origin again after a further $2(n-r)$ steps. Hence,

$$P(2n) = S(2n) + \sum_{r=1}^{n-1} P\big(2(n-r)\big)S(2r)$$

This applies in any number of dimensions as the dimension plays no part in the argument.

Changing the summation variable to $\tilde{r} = n - r$ this becomes,

$$P(2n) = S(2n) + \sum_{\tilde{r}=1}^{n-1} P(2\tilde{r})S\big(2(n-\tilde{r})\big)$$

Re-arranging gives,

$$S(2n) = P(2n) - \sum_{r=1}^{n-1} P(2r)S\big(2(n-r)\big)$$

which is Equ.(37). **<u>QED</u>**.

For the "AND" case we have $P(2n) = \left(\frac{(2n)!}{n!^2}\right)^d$ and hence Equ.(37) allows $S(2n)$ to be found recursively.

The probability of being at the origin for the first time is then $\Delta p(n) = S(2n)/T(2n)$ where $T(2n) = 2^{2dn}$ is the total number of paths of length $2n$ in d dimensions.

These are mutually exclusive possibilities, and so because every path that returns to the origin must do so for the first time exactly once it follows that the cumulative sum of these probabilities is the probability of returning to the origin one or more times by step $2n$. So the return probability is,

$$Prob = \sum_{n=1}^{\infty} \frac{S(2n)}{T(2n)} = \sum_{n=1}^{\infty} \Delta p(n) \qquad \text{(X)}$$

In practice this sum can be found up to some maximum n and the rest of the sum to infinity found by assuming the increments of probability reduce in proportion to $r^{-d/2}$ (a result which follows from the preceding analyses). Hence if the sum, (X), is evaluated up to some $n = N$, and with $\Delta p(N) = S(2N)/T(2N)$, the remainder can be approximated by,

$$\sum_{n=N+1}^{\infty} \frac{S(2n)}{T(2n)} \approx \int_N^{\infty} \Delta p(N) \left(\frac{N}{r}\right)^{\frac{d}{2}} dr = \left(\frac{N}{\frac{d}{2} - 1}\right) \Delta p(N)$$

which is Equ.38 (corrected). **<u>QED</u>**.

Appendix B: Probability That Two Random Walkers Collide

In the limit of a very large number of steps, the probability that two random walkers, starting from the same place at the same time, subsequently collide is the same as the probability of one random walker returning to the origin. Hence, in 1D and 2D this probability is asymptotic to 1, whilst in higher dimensions it is given by Table 2, above.

For any finite number of steps, the probability that two random walkers, starting from the same place at the same time, collide on step N is the same as the probability of one random walker returning to the origin on step $2N$.

The most elegant means of showing this is to consider some position vector $\bar{r}$. Now consider all the paths that person 1 could take to reach that position on step N. Combine each of those paths with every path that person 2 could take to reach the same point on step N. That gives the total number of cases which involve a collision at position $\bar{r}$ on step N, which, on dividing by 2^{2Nd} gives the corresponding probability.

But by reversing the second person's path and adding to the first person's path one gets a path of $2N$ steps that takes a single walker back to the origin on step $2N$, having passed through position $\bar{r}$ on step N. Moreover, every such path can be constructed exactly once in this way. As the number of single walker paths over $2N$ steps is again 2^{2Nd}, the corresponding probability is the same. This applies for every arbitrary $\bar{r}$ and hence also applies when one sums over $\bar{r}$ to get the total probabilities. **QED**.

Another means of demonstrating this result is as follows.

For a single person, from any position in two moves he will go 2 or -2 or 0 with probabilities 0.25, 0.25 and 0.5 respectively. For two people, the movement of one with respect to the other on every step is also 2, -2, 0 with probabilities 0.25, 0.25, 0.5. Hence the probability of collision of the two people after n steps equals the probability of a single person returning to the origin in 2n steps. Because a single person cannot return to the origin in an odd number of steps, it follows that in the limit of an infinite number of steps the two problems have the same answer. Indeed they have the same answer for small n also, if n steps of two people is compared with 2n steps of one person. As long as only the AND case is considered, this same argument works in any number of dimensions because the AND case in d dimensions is just d 1D cases superimposed. However, it is perhaps not so clear that it works for the "OR" case, though the first proof does apply to both.

Appendix C: The Ground Makes No Difference

My contention is that restricting one Cartesian coordinate to never be negative makes no difference to the probabilities of returns to the origin. More generally, if it is the z coordinate that is constrained, i.e., $z \geq 0$, then the probability after N steps of being at any given position is the same as that for the case without ground if negative z is identified with abs(z).

There has been some dispute over this. However, my view, enunciated here, is that this variant problem illustrates a type of fallacy that is familiar in probability calculations. To illustrate the matter, consider 1D and just the first three steps. Only three paths are possible, as follows,

	Step Number		
	1	2	3
Path 1	1	2	3
Probability of each step	*1*	*0.5*	*0.5*
Path 2	1	2	1
Probability of each step	*1*	*0.5*	*0.5*
Path 3	1	0	1
Probability of each step	*1*	*0.5*	*1*

If one assumes all paths are equally probable, then the probability of Path 1 is 1-in-3, as is the probability of Paths 2 and 3.

However, it is clear that the probability of Path 1 is actually 1-in-4 because it relies on two consecutive, independent, events each of probability 0.5.

The same is true for Path 2, whist Path 3 has a probability of 0.5 because it relies only on a single 0.5 probability event.

Hence, the paths are NOT equally probable.

One cannot simply make all paths equally probable ***by fiat*** because this would involve the paradoxical result that the probability of Path 1 was 1-in-3 which is incompatible with it being the outcome of two consecutive, independent, events each of probability 0.5.

Any Monte Carlo simulation that gives equal weighting to all paths is therefore incorrect.

To give the correct weighting to paths which reach the ground, it is necessary to double-up the subsequent step in the counting. This is illustrated up to step 4 by Table 3, below, by comparing with and without ground.

Table 3

With Ground					No Ground			
$n = 1$	$n = 2$	$n = 3$	$n = 4$		$n = 1$	$n = 2$	$n = 3$	$n = 4$
1	0	1	0		-1	-2	-3	-4
1	2	1	2		1	0	-1	-2
	0	1	0			0	-1	-2
	2	3	2			2	1	0
		1	0				-1	-2
		1	2				1	0
		1	2				1	0
		3	4				3	2
			0					-2
			2					0
			0					0
			2					2
			0					0
			2					2
			2					2
			4					4

At every step the possible z values "with ground" are just the absolute values "without ground". In particular, the number of occurrences of $z = 0$ is the same for the two cases at every step. Both these enumerations now have the same probabilities for returns to the origin, and the same probabilities for any position if compared on the basis of $abs(z)$. **QED**.