This note is about a classic random walk problem that has been well studied. I like the problem for both it's droll interpretation as a drunkard's walk and for its many surprising conclusions.

Question: Consider the integer lattice $\mathbb{Z}^d$ for some dimension d. A random walk starts at the origin and from any lattice point, it moves to any of the neighbouring lattice points with equal probability. What can we say about this random walk's return to the origin?

The question is left specifically vague. We can be interested in many things about the return time of this walk - expected time/distribution of first return to origin, expected number of returns to origin, probability of never returning to the origin to name a few.

Let us begin by restricting ourselves to the simple case of a 1-D lattice and concern ourselves with the first return to origin. In this simple case, let p denote the probability of a hop to the right and $q = 1 - p$ the probability of a hop to the left. Let X_k denote the location of the walk after k hops. The probability that the walk returns to the origin after $2n$ steps is

$$P(X_{2n} = 0) = \binom{2n}{n}(pq)^n$$

. Note that this probability does not preclude the fact that there is some $k < n$ for which $X_{2k} = 0$. In other words, the probability above is *not* the probability of first return. Instead, the first return probability is embedded inside $P(X_{2n} = 0)$. To tease it out, let us start by defining

$$f_n \triangleq P(X_{2n} = 0 \text{ and } X_{2k} \neq 0 \text{ for all } k < n) \quad n = 0, 1, 2, \ldots \tag{1}$$

$$p_n \triangleq P(X_{2n} = 0) \quad n = 0, 1, 2, \ldots \tag{2}$$

We also define the initial conditions $f_0 = 0$ and $p_0 = 0$. Since a walk that returns to the origin at time $2n$ should have visited the origin at some time between 2 and $2n$ for the first time, we have the relation

$$p_n = \sum_{k=1}^{n} f_k p_{n-k} \tag{3}$$

The convolution-like product suggests that the next step is to go to the generating functions of p_n and f_n. Define the formal power series $F(x) = \sum_{n=0}^{\infty} f_n x^n$ and similarly for $P(x)$. Then, the above convolution can be written in this domain as

$$P(x) - 1 = P(x)F(x)$$

. Note that the -1 in the LHS is a consequence of the initial conditions f_0 and p_0. Thus, we have the generating function

$$F(x) = 1 - \frac{1}{P(x)}$$

.

The next step is to evaluate the function $P(x)$. Since $p_n = \binom{2n}{n}(pq)^n$, we have [1]

$$\sum_{n=0}^{\infty} \binom{2n}{n} x^n = \tfrac{1}{\sqrt{1-4x}} \tag{4}$$

$$\sum_{n=0}^{\infty} p_n x^n = \tbinom{2n}{n}(pqx)^n \tag{5}$$

$$= \tfrac{1}{\sqrt{1-4pqx}} \tag{6}$$

The generating function $F(x)$ then simplifies to

$$F(x) = 1 - \sqrt{1 - 4pqx} \tag{7}$$

We can use the binomial theorem now to expand $F(x)$ to calculate f_n. The (generalized) binomial theorem gives us

$$\sqrt{1-x} = \sum_{m=0}^{\infty} \frac{-1}{2m-1} \binom{2m}{m} \frac{x^m}{4^m} \tag{8}$$

Substituting this into $F(x)$ gives us

$$F(x) = \textstyle\sum_{n=0}^{\infty} f_n x^n \tag{9}$$

$$= \textstyle\sum_{n=1}^{\infty} \frac{1}{2n-1} \binom{2n}{n} (pq)^n x^n \tag{10}$$

giving us $f_n = \frac{1}{2n-1}\binom{2n}{n}(pq)^n$. The $2n-1$ in the denominator signals that there might be a combinatorial interpretation of the answer in terms of Catalan numbers. I haven't looked into it yet.

Manipulating $F(x)$ gives us more insights. First, setting $x = 1$ we find that $F(1) = \sum_{n=1}^{\infty} f_n = 1 - \sqrt{1-4pq} = 1 - |p-q|$. The sum of all the f_n is simply the probability that the random walk ever returns to the origin [2]. Thus, with probability $|p-q|$, the walk will never return to the origin. The expected time of return to origin is again easy to calculate from $F(x)$. The expectation is just $F^{'}(1)$ and equals $\frac{2pq}{|p-q|}$. Note that this is to be interpreted as the expected time to return to the origin *given* that the walk does indeed return to the origin.

Simplifying to the case of the symmetric random walk ($p = q = \frac{1}{2}$), we find that the probability of returning to the origin is 1 while the expected time of return to the origin is ∞. So a symmetric walk will almost surely return to the origin but will take forever to do so. Contrast this with an asymmetric walk that has a chance to wander away forever but returns in a finite time when it indeed does return.

A natural generalization is to consider higher dimensional lattices $\mathbb{Z}^d$. From now on, we focus solely on symmetric walks and ask ourselves the probability that a random walk on $\mathbb{Z}^d$ ever returns to the origin. Before proceeding further, a rough heuristic can motivate the problem. Even in the 1-D walk above, return to origin proved quite delicate and only $p = q = \frac{1}{2}$ ensured that it happened almost surely. As the dimension

[1] this is a classic result in combinatorics and can be "proved" using the generalized binomial theorem. However, I lack any intuition into why we would expect something like this to be true.

[2] I initially expected $F(1) = 1$ and was looking for errors in the derivation when it didn't turn out to be the case. One question I have about generating functions - when do we attribute $1 - F(1)$ to the probability mass at infinity and when is it just a mistake?

d of the lattice grows, the walk has more room to spread out and we can expect the probability of eventual return to go down with d. More precisely, a large deviation analysis shows that a walk of length N is concentranted within a ball of radius $\sqrt{N}$. Now the number of lattice points inside this ball is $N^{d/2}$. For the 1-D case, the number of points is much smaller than the number of steps which means all the points (including the origin) are visited frequently. If $d \geq 3$, the number of distinct points is much larger than the number of points visited and the probability of revisiting a point goes down. $d = 2$ is a borderline case that this heuristic is too weak to handle.