A primer on random walks, first passages/returns and broad distributions

Some important properties of random walks can turn unintuitive, such as the distribution of first and last return times. We will study later first passage properties, introducing generating functions and martingales. Here, we supply elementary derivations of the main results. We will show in particular that the first return time is broadly distributed with Lévy index 1/2, with thus an infinite mean (moments of all orders do diverge).

We consider a 1D random walk, defined by a succession of N unit length steps to the right or to the left, with the same probabilities 1/2 in both cases. Time is discrete: a random step is made every time τ_0. The total time elapsed is thus $t = N\tau_0$. All walks start at time 0 from the origin O (point $x = 0$). The first return time is the smallest non zero time at which the walker is back to $x = 0$. The last return time is the time of the last visit before time t, see Fig. 1.

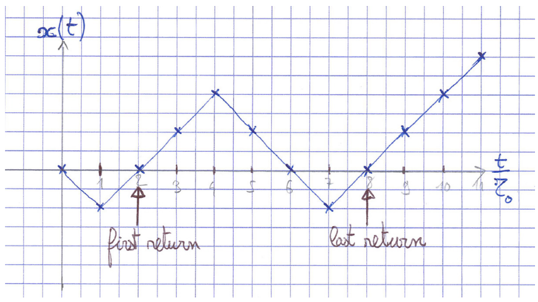

Figure 1: One dimensional random walk $x(t)$ with 11 steps, starting from the origin. The first return time is $t/\tau_0 = 2$, while the last return time is $t/\tau_0 = 8$.

The first return time distribution is denoted $F(t)$. Its cumulative $\int_0^t F(t')dt'$ is the probability to return to the origin before time t. Hence, the survival probability $S(t)$ defined as the probability of not returning to 0, is

$$S(t) = 1 - \int_0^t F(t')dt' \qquad \Longrightarrow \qquad F(t) = -\frac{dS(t)}{dt}. \tag{1}$$

We proceed to compute $F(t)$ from the enumeration of "survival" paths, those not returning to the origin, for a given time $t = N\tau_0$. The probability of such a path is given by the ballot theorem[1], that we show making use of a cyclic representation of paths, as illustrated in Fig. 2. We denote N_+ and N_- the number of steps to the right and the left respectively, with $N = N_+ + N_-$. We will also consider the other paths represented, that start not from O on the right panel, but from any of the N points shown, going from a site to the next in a clockwise fashion. To see if a path is of the surviving type, we note that any +1 followed by a -1 can be eliminated from the track. Repeating the elimination procedure, starting again from the origin, we are left with the number of steps in excess (say to the right if $N_+ > N_-$). If the origin has not been eliminated in the process, we have a survival path.

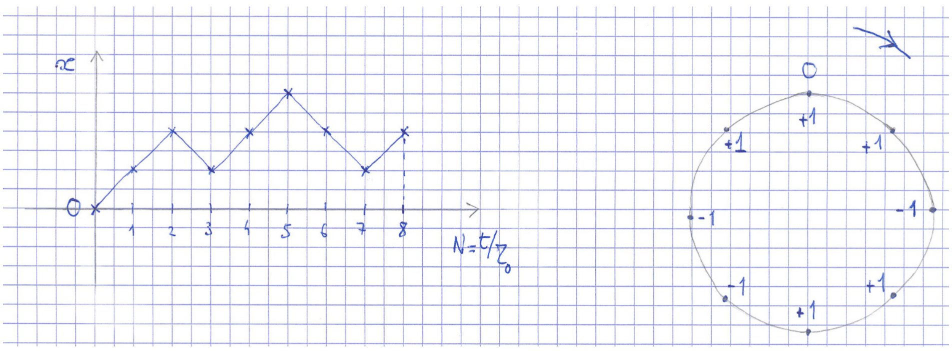

Figure 2: Mapping of a random walk to a cyclic representation (+1 for a step to the right, -1 for a step to the left). Here, $N_+ = 5$, $N_- = 3$, $N = 8$. Such a track shows N possible clockwise paths, starting from each of the N points represented. Among these 8 paths, $N_+ - N_- = 2$ are of the survival type, *ie* such that the corresponding random walk is always strictly to the right of the origin.

[1] in a ballot where candidates A and B have respectively a and b votes, with $a > b$, the probability that A is ahead of B throughout the whole counting is $(a - b)/(a + b)$.

1) Note that the removal of pairs does not affect any other path; besides, if a site disappears by elimination, it cannot be a valid starting point for a survival path. Among the N possible different paths starting from one of the N sites of the track, how many do survive?

2) What is the probability to choose a survival path in a track having given N_+ and N_-?

3) Making use of the fact that a random permutation of the steps leaves the above result unchanged, show that for a given N_-, the number of paths remaining to the right of the origin is

$$\frac{N-2N_-}{N}\binom{N}{N_-}. \tag{2}$$

4) Take N even for simplicity. Summing over all eligible values of N_-, show that the total number of paths remaining to the right of the origin is $\binom{N-1}{N/2-1}$. One can make use of

$$\binom{n}{p} = \binom{n-1}{p-1} + \binom{n-1}{p}. \tag{3}$$

5) Taking into account the paths that always stay to the left of the origin, show that the survival probability reads

$$S(N) = \frac{1}{2^{N-1}}\binom{N-1}{\frac{N}{2}-1} = \frac{1}{2^N}\binom{N}{\frac{N}{2}}. \tag{4}$$

6) What does this expression become for large N (*ie* large $t = N\tau_0$)? Show that $S(t) \to 0$, for $t \to \infty$, which is in agreement with Polya's theorem: on an infinite lattice of dimension $d < 2$, symmetric random walks return to the origin. The walk is then said to be *recurrent*.

7) For large t, how does the distribution of return times $F(t)$ behave? What can we say about the first moment of this distribution?

8) The previous results shed light on the last return time distribution. We denote by $2n_\ell$ the position of the walker at the last return time (note that this quantity need be even). What is the probability, for n_ℓ given, that the walker is at the origin after $2n_\ell$ steps exactly?

9) Multiplying the above result by the adequate survival probability, show that the probability that the last return occurs at step $2n_\ell$ is, for large n_ℓ

$$\frac{1}{\pi\sqrt{n_\ell(n-n_\ell)}}. \tag{5}$$

where $2n$ is the total number of steps. Denoting $\tau = 2n_\ell/(2n) = n_\ell/n$ the normalized time for the last visit, this means that

$$P(0 \le \tau \le a) = \int_0^a \frac{1}{\pi}\frac{d\tau}{\sqrt{\tau(1-\tau)}} = \frac{2}{\pi}\arcsin(\sqrt{a}). \tag{6}$$

This is the so-called arcsine law (also referred to as Lévy law). Remarkably, the same arcsine law also holds for

- the total time that $x(t) > 0$ in a given time interval (same result with the total time $x(t) < 0$);
- the time at which the motion achieves its maximum (same result for the minimum).

For these latter two quantities, the distribution is universal, independent from the jump distribution, as long as it is symmetric (Sparre-Andersen theorem[2]).

10) You can now ponder on the rather unexpected nature of these results. As put by Feller[3]: *We are now prepared for a closer analysis of the nature of chance fluctuations in random walks. The results are startling. According to widespread beliefs a so-called law of averages should ensure that in a long coin-tossing game each player will be on the winning side for about half the time, and that the lead will pass not infrequently from one player to the other. Imagine then a huge sample of records of ideal coin-tossing games, each consisting of exactly 2n trials. We pick one at random and observe the epoch of the last tie (in other words, the number of the last trial at which the accumulated numbers of heads and tails were equal). This number is even, and we denote it by 2k (so that $0 \le k \le n$). Frequent changes of the lead would imply that k is likely to be relatively close to n, but this is not so. Indeed, [an] amazing fact [is] that the distribution of k is symmetric in the sense that any value k has exactly the same probability as $n-k$. This symmetry implies in particular that the inequalities $k > n/2$ and $k < n/2$ are equally likely. With probability 1/2 no equalization occurred in the second half of the game, regardless of the length of the game. Furthermore, the probabilities near the end points are greatest; the most probable values for k are the extremes 0 and n. These results show that intuition leads to an erroneous picture of the probable effects of chance fluctuations.*

Reference: S. Kostinski and A. Amir, *Am. J. Phys.* **84**, 57 (2016).

[2] see S.N. Majumdar, `arXiv:0912:2586`, Physica A **389**, 4299 (2010).
[3] in the classic *An Introduction to Probability Theory and Its Applications*, ch III.