On the Computational Statistics of Random Walks

University of BRISTOL

Author: Gilad Gur Harush
Supervisor: Luca Giuggioli

Fundamentals

Random walks are stochastic, memoryless processes. At each timestep, a walker randomly chooses whether to step left or right classically. Clearly, there are countless alterations and properties to investigate.

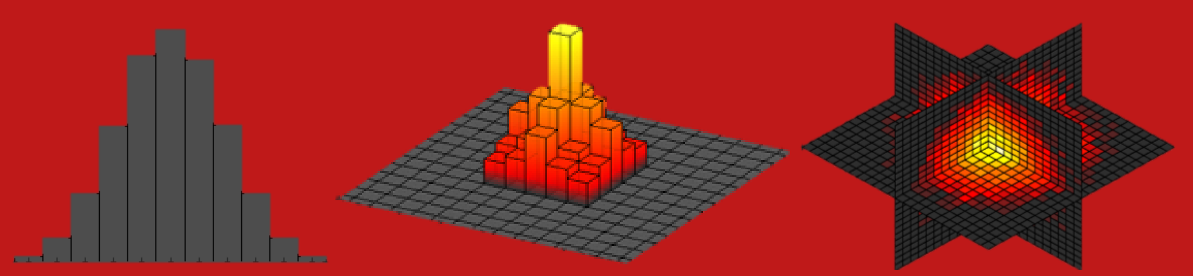
The first, classic model is called the 'symmetric' or 'isotropic' model, in which every direction has equal probability. Many variants have emerged since and many interesting dynamics have been discovered. Many arose out of pragmatism in other fields. This is the case with the correlated random walk, where each step favours the direction the walker previously stepped in; biologists and chemists have found great utility for this.

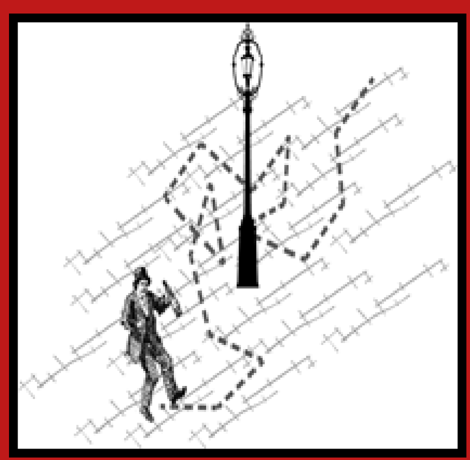
Many arose for the sake of mathematical curiosity. This is the case with some ballistic walks; one variant we demonstrate in this poster is the Muñoz walk which travels ballistically in sites already visited, and others have been put forward that only act this way under certain environmental conditions.

Random walks have a short but profound history. Initially stumbled upon by Blaise Pascal in the mid-1600's, they were left untouched for 350 years. Their first mention by name was then in 1900, when rediscovered by Einstein and Louis Bachelier independently. Their applications nowadays sprawl many contexts at many scales; from Twitter's recommendation service to meteorological mixing models.

We know some properties of random walks:

When symmetric and propagating freely, their distribution transitorily forms a Gaussian due to their symmetry:





1D and 2D walks are *recurrent*, meaning they return to the origin with 100% certainty, even in an infinite domain:

"[T]he most probable place to find a drunken man who is at all capable of keeping on his feet is somewhere near his starting point!"

But 3D (and higher) are not. This is known as *transience* and it plays a significant role in explaining many attributes and patterns we find in 1D and 2D, but not in higher dimensions:

"A drunk man will find his way home but a drunk bird may get lost forever!"

They are a *Markov process*. This helps us find their propagator function:

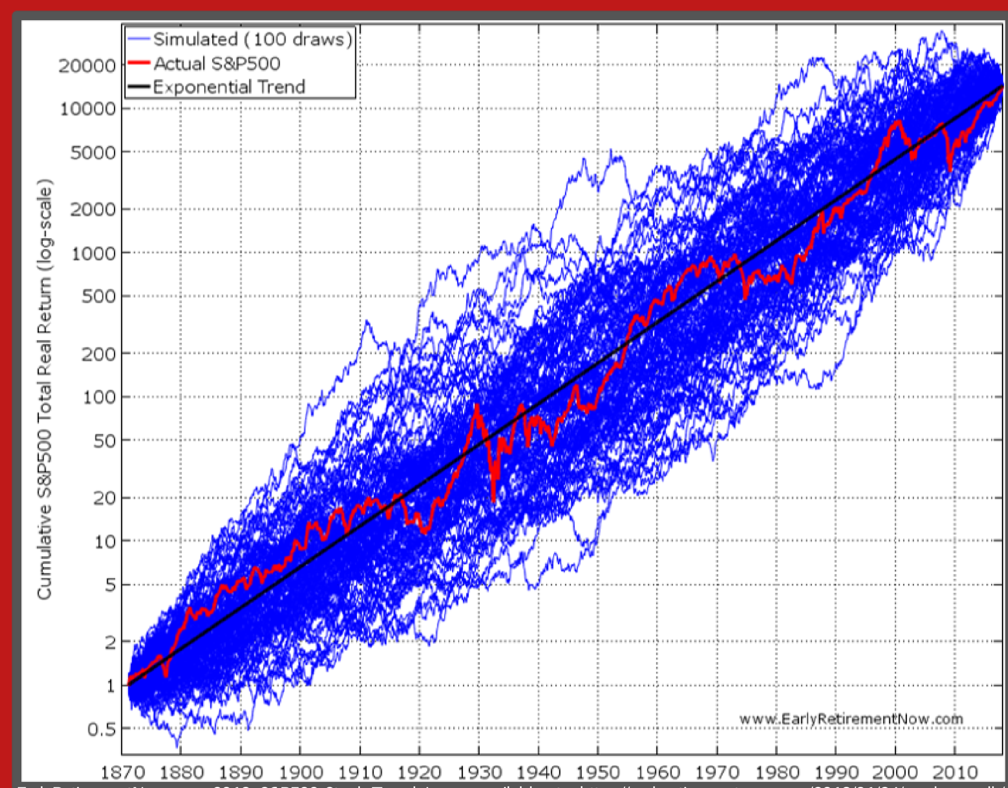
$$\mathbf{P}^t = \mathbf{V}\mathbf{J}^t\mathbf{V}^{-1} = [\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_N]\begin{bmatrix}\lambda_1^t & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_N^t\end{bmatrix}[\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_N]^{-1}$$

$$P_{n_0}(N,n,p,t) = \frac{1}{N}\sum_{k=1}^{N}\alpha_k \cos\left[\frac{(2n-1)(k-1)\pi}{2N}\right]\cos\left[\frac{(2n_0-1)(k-1)\pi}{2N}\right]\lambda_k^t$$

$$P_{n_0,m_0}(N,M,n,m,p,t) = \frac{1}{NM}\sum_{k=1}^{N}\sum_{k'=1}^{M}\alpha_k\alpha_{k'}\cos\left[(2n-1)\frac{k-1}{2N}\pi\right]\cos\left[(2m-1)\frac{k'-1}{2M}\pi\right]$$
$$\cos\left[(2n_0-1)\tfrac{k-1}{2N}\pi\right]\cos\left[(2m_0-1)\tfrac{k'-1}{2M}\pi\right]\ (\lambda_k\lambda_{k'})^t$$

$$\lambda_k = 1 - p(1-\cos\left[\frac{(k-1)\pi}{N}\right]) \qquad \alpha_k = \begin{cases}1 & \text{if } k=1\\ 2 & \text{otherwise}\end{cases}$$

Due to their stochastic memoryless nature, we know real-world processes that resemble them are effectively unpredictable. This is exemplified extremely well by the S&P500 stock over the last century. Accounting for a slight drift and major global events, we see exactly why most aspiring brokers don't make it very far.



EarlyRetirementNow.com, 2018. S&P500 Stock Trend. Image available at: <https://earlyretirementnow.com/2018/01/24/random-walk/> [Accessed 6 May 2020].

Path Analysis

As above, we know the symmetric walk's propagator function. This is sometimes called its *occupational probability* as it yields the probability of occupying a given node at a given time. We use this function in the derivation of a *first-passage probability* function; similarly but distinctly, this tells us the likelihood of occupying that node for the first time instead.

This metric becomes incredibly useful when looking at stochastic first-hit models. These include scenarios such as DNA Helicase enzymes attempting to unzip the double helix, models of foraging animals used by researchers to model ecological dynamics, and gambling financial models used by casinos worldwide. These are just three of the most impactful applications of the past 50 years; there are also a plethora of random-walk-based stock market models, pandemic models, and countless more.

The first-passage time of a random walk has been historically elusive for a combination of reasons. However, coordination between different mathematical disciplines and extensive contributions by famous mathematicians have made its attainment a reality.

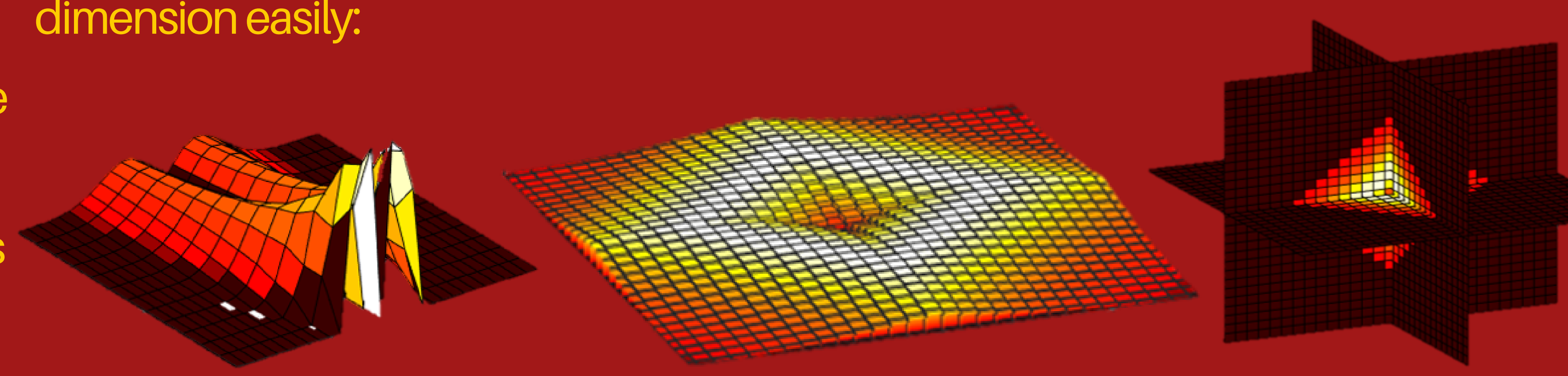
In this research we obtain the FPT distribution using a different technique. Often introduced in tandem with the Laplace transform, the *Z-transform* is nowadays a staple in random walk theory papers for its interdisciplinary utility. It effectively performs the same transformation, converting time-series data into complex frequency-domain representations, but it's especially useful for discrete distributions. We use this to uncover the distribution of first-passage times and probabilities.

Applying this transform to the propagator function is easy enough:

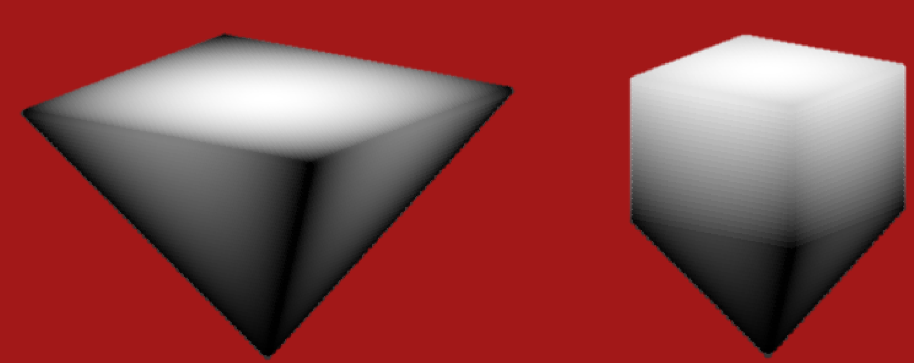
$$F_{n_0} = \frac{1}{2\pi j}\oint_{\Phi} \mathrm{d}z\, z^{-(t+1)}\frac{\sum_{k=1}^{N}\dfrac{\alpha_k\cos\left[\frac{(2n-1)(k-1)\pi}{2N}\right]\cos\left[\frac{(2n_0-1)(k-1)\pi}{2N}\right]}{1-z\lambda_k}}{\sum_{k=1}^{N}\dfrac{\alpha_k\cos^2\left[\frac{(2n-1)(k-1)\pi}{2N}\right]}{1-z\lambda_k}}$$

Its inversion isn't so easy though. Thankfully, recent works by well-known signal processing researchers allow us to do this computationally.

Side-stepping Cauchy's inversion integral (the standard inverse *Z*-transform method), we are able to invert this using a numeric inversion algorithm devised by Whitt et al. The results can be directly visualised and analysed for any dimension easily:



Additionally, we study the *enumeration* of random walk paths. For any dimension, the number of paths a random walk can take to a given position follows certain predictable patterns. For example, a 2D random walk can be enumerated using a 3D square-based right pyramid. We also show the effects of confinement on this enumeration:



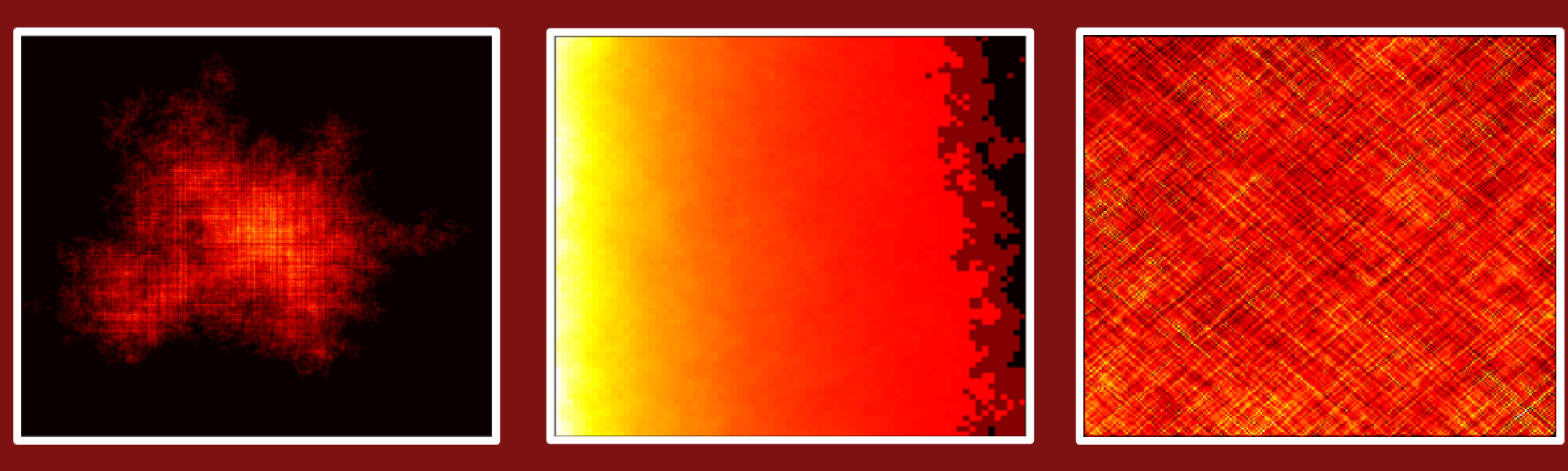
Coverage Analysis

Suppose you have a robot aiming to vacuum your entire floorspace, or you're modelling the spread of an epidemic within a social structure, or you want to know the time it takes your satellite to search an entire area for a target. These tasks can all be reduced to the same question: how long does it take for a random walk to cover a given domain?

Amongst other names, this time is known as the cover time and has been extensively studied since the 1990's. While the parameters and particulars in all these tasks can vary, they all share the same underlying dynamics, which is what allows results from research to be applied to an abundance of applications.

In this research we tackle this problem for a few types of random walks. We also look deeper than just at trends in the coverage time itself; we put a special emphasis on propagation metrics such as the mean-square displacement and rate of distinct site visits. These help us find the main and subsidiary influences on the elongation or shortening of coverage times. We do this for many different walks, looking at different scales in multiple dimensions.

First studied in a paper from the 90's, one of the walks we study is the Muñoz non-Markov walk. As the name suggests it exhibits non-Markovian properties; it remembers the direction it previously stepped in until it lands on a new site. We also study walks with a constant drift and those with a correlation, both parametrically controlled:



One of the unique additions we make to the literature is in our inclusion of a standstill probability. We control this probability with a parameter, $\mathbf{p}$. If we let this standstill probability go to 0, we observe a grid-like distribution. This can have adverse effects on coverage as it restricts a walker to a single diagonal on the lattice: