Proceedings of CONF-APMM 2025 Symposium: Simulation and Theory of Differential-Integral Equation in Applied Physics
DOI: 10.54254/2753-8818/2025.DL34030

From Stirling's Approximation to Random Walk Applications

Yifei Lu[1†], Xinyi Liang[2*†]

[1]Shanghai United International School Pudong-Sanlin Campus, Shanghai, China
[2]Shenzhen Senior High School Group International Division, Shenzhen, China
**Corresponding Author. Email: liangxy@cn-school.com*
†These authors contributed equally to this work and should be considered as co-first author

Abstract. In this paper, we demonstrate the proof of Stirling's Formula by employing Stolz Theorem along with Wallis' Formula. Stirling's Formula has a wide array of applications, and the random walk problem stands out as a highly popular research domain within it. So, we proceed to derive the fundamental probability formulas pertaining to the random walk model, while also examining its recurrence across one, two, and three dimensions. Moreover, we concentrate on the application of the Random Walk in connection with various fields such as Math, Economics, and Computer Science, presenting concepts like Random Walk on graphs, search algorithms, and financial models.

Keywords: Stirling's formula, Random Walk, Recurrence, Search algorithm

1. Introduction

The Stirling formula is employed to estimate combinatorial counts as well as to analyze probabilities and asymptotic behavior within the realm of statistics, particularly when dealing with large samples. At its core, the fundamental concept behind it involves incorporating the approximation of large factorials into the computation of probabilities and combinatorial quantities. When one attempts to compute large factorials through direct multiplication, there is a possibility for errors to occur or for the process to be inefficient. Hence, Stirling's formula serves as an effective means to tackle these issues. James Stirling was the first to discover Stirling's formula, which offers an efficient and precise estimation for large factorials. Polya put forth the random walk theorem, which stands as a key application of Stirling's formula. It suggests the recurrence of the random walk problem for different dimensions by utilizing Stirling's formula [1].

In 1994, László Lovász came up with the idea of employing the random walks on graphs for analyzing combinatorial structures and approximating complex quantities in an efficient manner, which set the stage for subsequent inventions [2]. Within the realm of computer science, contemporary researchers such as Larry Page and others put forward the extensively utilized PageRank search algorithm, which is a probability model based on random walks. This proposal completely transformed web search as well as network analysis [3]. In the sphere of Economics, Louis Bachelier in 1900 modeled stock prices as a continuous random walk, thereby being the first to introduce the notion that connects the random walk theory to the stock market, thus establishing the foundation for financial mathematics [4]. In 2001, Tommi Sottinen firmly linked the discrete

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.DL34030

random-walk-based market models with the FBM-based continuous-time financial models, enabling a more realistic prediction of stock prices and behaviors [5]. Over recent years, other researchers have employed random walk to evaluate the efficiency of various stock markets [6, 7].

Starting from the 1730s, which is when the formula was initially discovered, up until the present day, the research concerning this formula has steadily expanded in scope, and its practical applications have become increasingly diverse. Consequently, as Stirling's formula developed over time, it not only kindled scientists' keen interest in the precision of approximations, but also propelled technological progress, thereby bringing benefits to numerous aspects of our everyday lives.

In Section 2, we work on deriving the Stirling's formula. We do this by putting the Stolz Theorem to use as well as making use of Wallis' Formula. After that, in Section 3, this formula gets further developed for its application in the context of the Random Walk. Within this section, we go about proving its recurrence in one dimension. We accomplish this through calculations involving probability and combinatorics. Moreover, in Section 3.2.2, we carry out a simulation of the Random Walk across dimensions 1, 2, and 3. The purpose here is to create a model that illustrates its recurrence. Moving on to Sections 4 and 5, we compile the remarkable findings made by scientists in fields like Math, Computer Science, and Economics. These findings are related to Stirling's formula's application as well as the Random Walk.

2. Proof of Stirling's formula

In the derivation of Stirling's Formula, the Stolz's Theorem can first be employed to establish the relational connection among $n!$, n^n, and e^n ; subsequently, the Wallis' Formula is introduced to solve for the unknown parameters involved in the derivation process [8, 9].

Theorem 2.1

$$\lim_{n\to\infty} \frac{\ln(n!)}{n\ln n-n} = 1 \tag{2.1}$$

Proof. Let $x_n = \ln(n!)$ and $y_n = n\ln n - n$, then according to Stolz's Theorem:
If the sequence $\{y_n\}$ is strictly increasing and $\lim_{n\to\infty} y_n = \infty$,

$$\lim_{n\to\infty} \frac{x_n}{y_n} = \lim_{n\to\infty} \frac{x_{n+1}-x_n}{y_{n+1}-y_n} \tag{2.2}$$

Confirmed the existence of the right-hand limit.
Now we need to verify the applicable conditions of Stolz's Theorem:

DOI: 10.54254/2753-8818/2025.DL34030

$$\begin{aligned} y_{n+1}-y_n &= [(n+1)\ln(n+1)-(n+1)]-[n\ln n-n] \\ &= (n+1)\ln(n+1)-n\ln n-1 \\ &= n[\ln(n+1)-\ln n]+\ln(n+1)-1 \\ &= n\ln\left(1+\tfrac{1}{n}\right)+\ln(n+1)-1 \\ &= n\ln\left(1+\tfrac{1}{n}\right)+\ln\left(1+\tfrac{1}{n}\right)+\ln n-1 \end{aligned} \tag{2.3}$$

By Taylor expansion, $\ln\left(1+\frac{1}{n}\right)=\frac{1}{n}-\frac{1}{2n^2}+O\left(\frac{1}{n^3}\right)$, Substituting this in $y_{n+1}-y_n$ gives

$$\begin{aligned} y_{n+1}-y_n &= n\left(\tfrac{1}{n}-\tfrac{1}{2n^2}+O\left(\tfrac{1}{n^3}\right)\right)+\left(\tfrac{1}{n}-\tfrac{1}{2n^2}+O\left(\tfrac{1}{n^3}\right)\right)+\ln n-1 \\ &= \left(1-\tfrac{1}{2n}+O\left(\tfrac{1}{n^2}\right)\right)+\left(\tfrac{1}{n}-\tfrac{1}{2n^2}+O\left(\tfrac{1}{n^3}\right)\right)+\ln n-1 \\ &= \ln n+\left(1+\tfrac{1}{n}-\tfrac{1}{2n}-1\right)+O\left(\tfrac{1}{n^2}\right) \\ &= \ln n+\tfrac{1}{2n}+O\left(\tfrac{1}{n^2}\right)>0 \end{aligned} \tag{2.4}$$

Thus, $\{y_n\}$ is strictly increasing.

As $n\to\infty$, $n\ln n$grows much faster than n, so $y_n=n\ln n-n\to\infty$.

In (2.4): If $\lim\limits_{t\to a}\frac{f(t)}{t}=0$, then it is said that “as $t\to a$, $f(t)=o(t)$ ”. Here, $o\left(\frac{1}{n^2}\right)$ denotes "a higher-order infinitesimal than $\frac{1}{n^2}$ ", proving the existence of function g(n) that satisfies $\lim\limits_{n\to\infty}\frac{g(n)}{\frac{1}{n^2}}=0$ and $o\left(\frac{1}{n^2}\right)$ is exactly $g(n)$.

$$x_{n+1}-x_n=ln((n+1)!)-ln(n!)=ln(n+1) \tag{2.5}$$

According to (2.3) and (2.5),

$$\begin{aligned} \frac{x_{n+1}-x_n}{y_{n+1}-y_n} &= \frac{\ln(n+1)}{\ln n+\frac{1}{2n}+O\left(\frac{1}{n^2}\right)} \\ &= \frac{\ln n+\frac{1}{n}+O\left(\frac{1}{n^2}\right)}{\ln n+\frac{1}{2n}+O\left(\frac{1}{n^2}\right)} \\ &= \frac{1+\frac{1}{n\ln n}+O\left(\frac{1}{n^2\ln n}\right)}{1+\frac{1}{2n\ln n}+O\left(\frac{1}{n^2\ln n}\right)} \end{aligned} \tag{2.6}$$

As $n\to\infty$, $\frac{1}{n\ln n}\to 0$:

$$\lim_{n\to\infty}\frac{x_{n+1}-x_n}{y_{n+1}-y_n}=\frac{1+0+0}{1+0+0}=1 \tag{2.7}$$

DOI: 10.54254/2753-8818/2025.DL34030

Therefore:

$$\lim_{n\to\infty}\frac{x_n}{y_n}=\lim_{n\to\infty}\frac{\ln(n!)}{n\ln n-n}=1 \tag{2.8}$$

which is

$$\ln(n!)=nlnn-n+o(n) \tag{2.9}$$

Take the exponential of both sides of $\ln(n!)=n\ln n-n+o(n)$:

$$n!=e^{n\ln n-n+o(n)}=e^{n\ln n}\cdot e^{-n}\cdot e^{o(n)}=n^n\cdot\frac{1}{e^n}\cdot e^{o(n)} \tag{2.10}$$

Let $a_n=\frac{n!\cdot e^n}{n^n}$, then $a_n=e^{o(n)}$.
Theorem 2.2 (Wallis' Formula)

$$\lim_{n\to\infty}\frac{[(2n)!!]^2}{(2n+1)\cdot[(2n-1)!!]^2}=\frac{\pi}{2} \tag{2.11}$$

In (2.11), the double factorial is defined as:

$$(2n)!!=2\cdot4\cdot6\cdot\cdots\cdot2n=2^n\cdot n! \tag{2.12}$$

$$(2n-1)!!=1\cdot3\cdot5\cdot\cdots\cdot(2n-1)=\frac{(2n)!}{2^n\cdot n!} \tag{2.13}$$

Substitute (2.12) and (2.13) into Wallis' Formula:

$$\lim_{n\to\infty}\frac{(2^n\cdot n!)^2}{(2n+1)\cdot\left(\frac{(2n)!}{2^n\cdot n!}\right)^2}=\frac{\pi}{2} \tag{2.14}$$

Simplify left-hand side:

$$\frac{2^{2n}\cdot(n!)^2}{(2n+1)\cdot\frac{(2n)!^2}{2^{2n}\cdot(n!)^2}}=\frac{2^{4n}\cdot(n!)^4}{(2n+1)\cdot(2n)!^2} \tag{2.15}$$

we now have

DOI: 10.54254/2753-8818/2025.DL34030

$$n! = a_n \cdot \frac{n^n}{e^n} \tag{2.16}$$

similarly, for (2n)! :
$\left(2n\right)! = a_{2n} \cdot \frac{(2n)^{2n}}{e^{2n}}$
Substitute (2.16) and (2.17) into (2.15):

$$\lim_{n\to\infty} \frac{a_n^4}{(2n+1)\cdot a_{2n}^2} = \frac{\pi}{2} \tag{2.17}$$

Assume that a_n is of the same order as $\sqrt{n}$:

$$\lim_{n\to\infty} \frac{a_n}{\sqrt{n}} = C \tag{2.18}$$

where C is a constant
Then

$$a_n = C \cdot \sqrt{n} + o\left(\sqrt{n}\right) \tag{2.19}$$

As $n \to \infty$, $2n + 1 \sim 2n$. Then substitute (2.20) into (2.18):

$$\lim_{n\to\infty} \frac{C^4 \cdot n^2}{2n \cdot C^2 \cdot 2n} = \frac{\pi}{2} \Longrightarrow C^2 = 2\pi \Longrightarrow C = \sqrt{2\pi} \tag{2.20}$$

Now we obtain

$$\lim_{n\to\infty} \frac{a_n}{\sqrt{n}} = \lim_{n\to\infty} \frac{n! \cdot e^n}{n^n \cdot \sqrt{n}} = \sqrt{2\pi} \tag{2.21}$$

which is Stirling's Formula

$$n! \sim \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n (n \to \infty) \tag{2.22}$$

DOI: 10.54254/2753-8818/2025.DL34030

3. Random walk model and its recurrence

3.1. The model setup and distribution of probabilities for 1D random walk

3.1.1. Model setup

We define that the particle starts at the origin: $S_0 = 0$

Then, the location reached by the particle after n steps is $S_n = X_1 + X_2 + \cdots + X_n$

in which $X_1, X_2, \ldots, X_n$ are random variables that are "independently and identically distributed" (i.i.d)

Each$X_i$satisfies: $P(X_i = +1) = p = \frac{1}{2}, P(X_i = -1) = q = \frac{1}{2}$ [10]

3.1.2. Probability distribution

Theorem 3.1 (the distribution of probabilities of 1D random walk)

$$P(S_n = j) = \begin{cases} \binom{n}{\frac{n+j}{2}} \left(\frac{1}{2}\right)^n, & \text{if } n+j \text{ is even and } \left|j\right| \le n \\ 0, & \text{if } n+j \text{ is odd or } |j| > n \end{cases} \tag{3.1}$$

proof. We define that the number of "rightward" steps is k and the number of "leftward" steps is $n-k$.

After n steps, the result of the position is $S_n = j$, in which

$$j = k - (n-k) = 2k - n \tag{3.2}$$

Therefore, for $S_n = j$, $k = \frac{n+j}{2}$.

Apparently, if $n+j$ is odd or $|j| > n$,

$$P(S_n = j) = 0 \tag{3.3}$$

The probability of making k rightward moves is given by the binomial distribution:

$$P = \binom{n}{k} p^k q^{n-k} = \binom{n}{k} \left(\frac{1}{2}\right)^n \tag{3.4}$$

considering each step is independent.

Substituting $k = \frac{n+j}{2}$:

$$P(S_n = j) = \binom{n}{\frac{n+j}{2}} \left(\frac{1}{2}\right)^n \tag{3.5}$$

DOI: 10.54254/2753-8818/2025.DL34030

3.2. The model setup and probability distribution of the 2D random walk

3.2.1. Model setup

We define that the particle starts at the origin: $S_0 = (0,0)$

After$n$ steps, the position of the particle is $S_n = (X_1 + X_2 + \ldots + X_n, Y_1 + Y_2 + \ldots + Y_n)$, where $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are independent and identically distributed (i.i.d.) random vectors

Each step satisfies that the particle moves one unit in one of the four directions (up, down, left, right) with equal probability: $P((X_i, Y_i) = (1,0)) = P((X_i, Y_i) = (-1,0)) = P((X_i, Y_i) = (0,1)) = P((X_i, Y_i) = (0,-1)) = \frac{1}{4}$
[10]

3.2.2. Probability distribution

Theorem 3.2 (the distribution of probabilities of 2D Random Walk)

$$P(S_n = (j,k)) = \begin{cases} \frac{1}{4^n} \sum_{\substack{a+b+c+d=n \\ a-b=j \\ c-d=k}} \frac{n!}{a!b!c!d!}, & \text{if } |j| + |k| \le n \text{and } |j| + |k| \text{has the same parity as } n \\ 0, & \text{otherwise} \end{cases} \tag{3.6}$$

proof. We define a as the number of rightward steps, b as the number of leftward steps, c as the number of upward steps, and d as the number of downward steps.

$$a + b + c + d = n \tag{3.7}$$

The horizontal coordinate of the position is $j = a - b$, and the vertical coordinate is $k = c - d$. Therefore,

$$a = \frac{|j|+j}{2} + t, b = \frac{|j|-j}{2} + t, c = \frac{|k|+k}{2} + s, d = \frac{|k|-k}{2} + s \tag{3.8}$$

where$t$, s are non-negative integers, and $|j| + |k| + 2(t + s) = n$ (let $m = t + s = \frac{n-|j|-|k|}{2}$, which requires $n - |j| - |k|$ to be a non-negative even number)

The total number of paths is the permutation number of (a, b, c, d) that satisfy the conditions $\frac{n!}{a!b!c!d!}$. The probability of each step is $\left(\frac{1}{4}\right)^n$, so the total probability can be calculated by the product of the number of paths and the probability of single step:

DOI: 10.54254/2753-8818/2025.DL34030

$$P\left(S_n=(j,k)\right)=\frac{1}{4^n}\sum_{\substack{a+b+c+d=n\\a-b=j\\c-d=k}}\frac{n!}{a!b!c!d!} \tag{3.9}$$

If $|j|+|k|>n$ or $|j|+|k|$ has a different parity from n, there are no valid (a,b,c,d) that satisfy the conditions, so the probability is 0.

3.3. First return to the origin problem and Pólya's random walk theorem

3.3.1. First return to the origin problem of the 1D random walk

According to Chapter III of Feller's work [10]:

Define p_n as the probability of the particle first returning to the origin at the $n-th$ step (n must be even), then

$$p_{2n}=\frac{(2n-1)}{\binom{2n}{n}}\cdot 2^{-2n} \tag{3.10}$$

Substitute Stirling's Formula $n!\sim\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$, then

$$p_{2n}\sim\frac{1}{n^{3/2}\sqrt{\pi}} \tag{3.11}$$

as $n\to\infty$

This indicates that in a 1d symmetric random walk, the particle will eventually return with probability 1—a property known as Recurrence.

3.3.2. Pólya's random walk theorem

Theorem 3.2

For simple random walks on the integer lattice $\mathbb{Z}^d$ with $d=1$ and $d=2$ (one-dimensional and two-dimensional), they are recurrent: the particle returns to the starting point with probability 1.

For $d\geq 3$ (three-dimensional and higher-dimensional) integer lattices, the random walk is transient: the probability that the particle returns to the starting point is less than 1. (When $d=3$, the probability that the particle returns to the starting point is approximately 0.34) [11, 12]

DOI: 10.54254/2753-8818/2025.DL34030

3.3.3. Python simulation for random walk when d=1, 2, 3

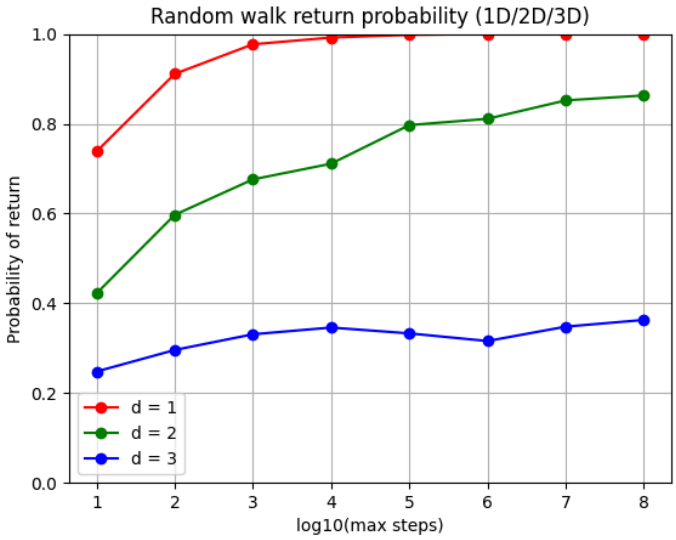

Figure 1. Random walk simulation for three different dimensions

As shown in figure 1, we construct a python program to stimulate the behavior of random walk for dimension 1, 2, and 3. For the case of d=1, our simulation clearly that the returning probability of the random walk converges rapidly to 1, which is correct according to the convergence property analyzed using the equation. Furthermore, the probability equation, presented previously in this paper, suggests that the random walk should converge to 1 for d=2. However, due to the limitation of computing ability, this simulation merely presents the trend of convergence as we increase the steps but does not eventually converge to 1. For d=3, the probability remains stable and is approximately 0.34, which conforms to the value presented previously in the paper.

4. Application of Stirling's formula - the central extended binomial coefficient

In this paper, Steffen Eger [13] generalizes a formula for the "central extended binomial coefficient".

The "central binomial coefficient" is the coefficient of x in $(1+x)^k$, and is represented as $\binom{k}{k/2}$. We then approximate it using Stirling's formula to get:

$$\binom{k}{k/2} \sim \frac{2^k}{\sqrt{\pi k/2}} \tag{4.1}$$

To study the coefficient of the term x in the expression " $\left(1+x+x^2+\ldots+x^l\right)^k$ ", the central extended binomial coefficient is introduced. This refers to the extended binomial coefficient situated in the middle of the range, which is denoted as $\binom{k}{\frac{kl}{2}}_{l+1}$.

It is noticed that the coefficient $\left(1+x+x^2+\ldots+x^l\right)^k$ exhibits a pattern resembling a bell-shaped curve, akin to the normal distribution. Given that the coefficient can be computed through

DOI: 10.54254/2753-8818/2025.DL34030

combinatorial means, probabilistic methods become applicable. The variance and mean of this curve are calculated, and it turns out that it aligns with the normal distribution as per the CLT (Central Limit Theorem). The CLT posits that the sum of numerous independent variables tends to behave in a manner like a bell-shaped curve (normal distribution). The central index is determined with the primary focus on the extended binomial coefficient around the middle area. This is because CLT and LLT (Local Limit Theorem) are more precise in the middle portion of the range and have a dominant influence over the overall behavior, making them the most crucial coefficients. Subsequently, LLT is utilized to ascertain the exact asymptotic value for the extended binomial coefficients. LLT states that the actual probability at each point can be closely estimated by the value of the normal density at that specific point. In the end, by employing Stirling's formula, the author approximates large factorials and ultimately derives the simplified expression for the central extended binomial coefficient:

$$\binom{k}{\frac{kl}{2}}_{l+1} \sim \frac{(l+1)^k}{\sqrt{2\pi k\frac{(l+1)^2-1}{12}}} \tag{4.2}$$

5. Application of random walk

5.1. Random walks on graphs

László Lovász [2] mainly talks about finite and undirected random walks that take place on graphs, as well as how a graph's structure impacts the random walks' behaviors.

In the first section, László Lovász creates a graph consisting of *n* nodes and *m* edges where the random walk travels between nodes through edges. He introduces properties of theorems of this graph governing the access time and cover times. "Access time" refers to the expected number of steps needed to reach a particular node from a given starting node, whereas "cover time" represents the expected number of steps required to visit every node in the graph at least once.

The part about 'the eigenvalue connection' stops us from looking at individual nodes. It shows that the transition matrix P, the eigenvector v_k, and the eigenvalue λ_k together decide how random walks behave. The transition matrix P is decided by the graph's structure and properties. It stands for the probability of moving between specific nodes. Every eigenvector v_k has entries for each node, which show how each node affects the overall behavior of the random walk. The eigenvalueλ_k mirrors the speed and manner of the random walk. Here, the largest eigenvalue ($\lambda_1 = 1$) indicates the steady rate, while the second-largest eigenvalue λ_2 shows the speed at which things converge. These mathematical structures fit into an equation expressed as:

$$P\,v_k = \lambda_k\,v_k \tag{5.1}$$

Moreover, the author presents a formula that is based on eigenvalues, which can be used to calculate the access time between two nodes in the network. The formula is as follows:

$$H(s,t) = 2m\sum_{k=2}^{n}\frac{1}{1-\lambda_k}\left(\frac{v_{k,t}^2}{d_t} - \frac{v_{k,s}\cdot v_{k,t}}{\sqrt{d_s\cdot d_t}}\right) \tag{5.2}$$

DOI: 10.54254/2753-8818/2025.DL34030

Here, $H(s,t)$ is the access time between node s and node t, and d represents the degree of the nodes, in other words, the number of edges connected to a specific node.

This concept is then applied briefly to electrical circuits in "the electrical connection" section. In this case, the random walk is the electricity flows through wires with resistance, represented by edges. The electrical version of the access time $K(s,t)$ can be expressed as:

$$\frac{1}{R_{s,t}} = \frac{2m}{K(s,t)} \tag{5.3}$$

$R_{s,t}$ denotes the resistance between node s and node t.

Eventually, the portion labeled "mixing rate" delves into the way eigenvalues impact the convergence pace of the random walk attaining the stationary state. This convergence speed is chiefly dictated by the second-largest eigenvalue λ_2 .

In conclusion, this paper stands out as the very first to delve into the exploration of random walks within the realm of general graphs. As a pioneering piece of work in the practical application of random walk methodologies, it firmly establishes the foundational underpinnings that are crucial for subsequent advancements across various fields such as mathematics and data science. Furthermore, it plays a significant role in propelling forward innovative concepts like PageRank, which will be elaborated upon in later sections of our paper.

5.2. PageRank searching algorithm

In 1998, Larry Page and Sergey Brin [3] came up with the PageRank searching algorithm, which forms the basis of the Google search engine. PageRank is a search algorithm that ranks the significance of web pages to enhance search precision. This is accomplished by predicting the importance of web pages via an analysis of the web graph's structure, considering both the quantity and quality of backlinks. PageRank simulates a random walk across the global network, which moves through the links between web pages, and the likelihood of the surfer ending up on a particular page indicates its importance. The "importance score" of a web page is equal to the sum of the scores of its backlinks, and if it has several outgoing links, its score is distributed among them.

The score of a web page's importance calculating using R', satisfies:

$$R'(u) = c\sum_{v\in B_u} \frac{R'(v)}{N_v} + cE(u) \tag{5.4}$$

such that c is the damping factor maximized and $\| R' \|_1 = 1$ *(* $\| R \|_1$ *denotes the* L_1 *norm of* R'*).* $E(u)$*is the set of pages that page u links to,* B_u denotes the set of items that link to u *and* N_v *is the total outgoing connections from page u.*

The initial value of R' serves as the very first estimation of the importance in question. This formula is employed in a recursive manner. What this means is that with each iteration that takes place, the score gets updated in reference to the one from the preceding iteration. As such, the accuracy of the score tends to improve gradually over a period. Once enough repetitions have been carried out, the importance score eventually converges and stabilizes at a fixed value.

DOI: 10.54254/2753-8818/2025.DL34030

The PageRank has rules for special cases as well. Dangling links are web pages without outgoing links, so they're equally linked to all web pages to avoid getting stuck. The variable E stops a walk from being trapped in a cycle among pages, letting it teleport randomly to another page if needed. Teleportation lets users make personalized PageRank, favoring their favorite pages. The random walk's convergence property keeps the probability distribution stable when landing on web pages, with teleportation and dangling link handling speeding up convergence for efficiency and accuracy. In total, the PageRank algorithm calculates the importance score, considering both the quantity and quality of web pages' backlinks, predicting influential web pages. It's used in web search, personalized search, and citation analysis. Most crucially, it's the foundation of Google, which is widely used and helps our daily lives.

5.3. ItemRank scoring algorithm

A recommender system refers to an algorithm that can automatically provide personalized product recommendations according to the user's past choices and actions. A. Pucci, M. Gori, as well as M. Maggini [14] put forward a random walk-based scoring algorithm named "ItemRank" with reference to the "PageRank" algorithm. They conduct tests using the MovieLens dataset and draw a conclusion that this algorithm outperforms the previously proposed algorithm in aspects such as memory usage, computational cost, and recommendation accuracy. "ItemRank" creates nodes on a plane, with each node representing an item. Lines are drawn to connect every pair of nodes. Depending on the general preferences of other users, some lines possess stronger connections and are more probable to be "walked" upon.

The scoring process is mathematically expressed as:

$$I\mathbf{R}_{u_i} = \alpha \bullet C \bullet I\mathbf{R}_{u_i} + (1 - \alpha) \bullet \mathbf{d}_{u_i} \tag{5.5}$$

Here, IR stands as a score vector pertaining to useru_i. The variableα serves to regulate the equilibrium existing between the preferences of users and the score that gets calculated through the algorithm. C denotes the correlation matrix, which illustrates the connections between various items. Lastly, $\mathbf{d}_{u_i}$ signifies the preference of the user, referring to those items that have already been marked as favorites by the user in question.

Furthermore, there's a greater likelihood of the walk making a jump towards the item that the user has marked as liked, which underscores the significance of the items that users like. The degree of this probability hinges on the extent to which the user liked a particular item. The application of random walk enables the walk to traverse the plane via lines, with each line having varying probabilities of being passed through. The recommended items are then ranked according to the frequency with which they are passed through. The recurrence of the classical random walk problem indicates that the number of times each item is passed through converges to a fixed figure, thereby bolstering the precision of the "ItemRank" algorithm.

5.4. Analysis of phase errors in the "SKW quantum search algorithm"

Zhang Yu-Chao et al. [15] primarily focus on analyzing the influence of systematic phase errors on the performance as well as robustness of the optimized SKW (Shenvi-Kempe-Whaley) quantum search algorithm and compare it with the Grover algorithm. Both the SKW algorithm and the Grover algorithm belong to search algorithms that enhance the probability of locating an item. However, the

DOI: 10.54254/2753-8818/2025.DL34030

Grover algorithm employs amplitude amplification by means of reflection around the mean, while the SKW algorithm adopts amplitude amplification via phase marking and interference. The authors design an algorithm model with introduced phase errors for study, and the outcomes reveal that the success rate and the quantity of iterations required are closely related to the database size and the degree of phase errors.

The success probability of the SKW search algorithm is given by:

$$p = \frac{19}{c^2(19+5\delta^2 \bullet 2^n)} \sin^2\left(\sqrt{\frac{4}{2^n} + \delta^2} \bullet t\right) \tag{5.6}$$

Here, the constant$c$ is specified by the way the algorithm is implemented, δ is the phase error, n is the size of the database, and t is the number of iterations required.

When the phase errors go up, the highest success rate and the quantity of iterations dwindle, which in turn brings down the performance. With sizable errors, the peak probability of success becomes lesser, thus diminishing its resilience. Moreover, should the database be sufficiently small, the search algorithm hardly gets impacted by the phase errors at all. In general, when put side by side with the Grover search algorithm, the SKW algorithm boasts superior robustness.

5.5. Binary market model

Tommi Sottinen [5] was the first one to utilize Donsker's theorem and propose that the random walk converges weakly to the standard Brownian motion.

“Standard Brownian motion” refers to a continuous-time stochastic process which begins from zero. This process features continuous paths, and its increments are independent, following a normal distribution with a mean of 0. Moreover, the variance of these increments is in proportion to the length of the time interval. To put it differently, the motion’s position at any given time t conforms to a normal distribution characterized by a mean of 0 and a variance equal to t. It is defined using the Wiener measure μ_W , which is employed to analyze the probability associated with all potential paths. At time t, this measure can be represented as a cumulative distribution function (CDF).

$$\mu_{\mathrm{W}}\left\{ X\left(\mathrm{t}\right) \leq a \right\} = \frac{1}{\sqrt{2\pi \mathrm{t}}} \int_{-\infty}^{\mathrm{a}} e^{-\frac{u^2}{2t}} \, du \tag{5.7}$$

Here, a is a specific position on the number line.

“Donsker’s theorem” asserts that at any given time t, the value of a suitably scaled random walk will have a distribution which aligns with the cumulative distribution function (CDF) of standard Brownian motion.

$\{X_n \mid n \geq 1\}$ *is a sequence of random variables with* $E[X_1] = 0$, $Var(X_1) = \sigma^2$, $0 < \sigma^2 < \infty$. The partial sum $S_n = X_1 + X_2 + \ldots + X_n$. The scaled random walk is expressed as $W_n(t) = \frac{S_{\lfloor nt \rfloor}}{\sigma\sqrt{n}}$, $t \in \left[0,\ 1\right]$. Then, as $n \to \infty$, the process $W_n(t)$ converges to the standard Brownian motion $W(t)$ with Wiener measure μW , expressing as

DOI: 10.54254/2753-8818/2025.DL34030

$$W_n \to^d W \; in \; C[0,1] \tag{5.8}$$

Then, to take past memory and trends into consideration, the author introduces fractional Brownian motion, to which the scaled random walk converges. “Fractional Brownian motion” is the generalization of “normal Brownian motion” where increments can be correlated, introducing trend memory controlled by the Hurst parameter$H$. The relationship between the normal Brownian motion and the fractional Brownian motion is given by:

$$Z_t = \int_0^t z(t,s)\, dW_s \tag{5.9}$$

Z_t represents the “fractional Brownian motion” with $H > \frac{1}{2}$, while W_s is the “standard Brownian motion”. $z(t,s)$ denotes the kernel function analyzing how effective is the past time s on the fractional Brownian motion at the current time$t$. This deterministic kernel function is expressed as:

$$z(t,s) = c_H \left(H - \tfrac{1}{2}\right) s^{\frac{1}{2}-H} \int_s^t u^{H-\frac{1}{2}} (u-s)^{H-\frac{3}{2}}\, du \tag{5.10}$$

Herec_H is the normalizing constant calculated as:

$$c_H = \sqrt{\frac{2H\,\Gamma\left(\frac{3}{2}-H\right)}{\Gamma\left(H+\frac{1}{2}\right)\Gamma(2-2H)}} \tag{5.11}$$

After applying the geometric transformation, the author gets the geometric fractional Brownian motion to ensure the model remains positive because stock prices can never be smaller than zero.

Next, the “fractional Black–Scholes model” is introduced based on geometric fractional Brownian motion. S_t represents the risky asset, stock, and can be expressed as:

$$dS_t = S_t\left(a_t dt + \sigma dZ_t\right) \tag{5.12}$$

Here, t is a time variable,$a_t$is the drift term, representing the expected return rate at time t, σ *controls how randomness affects the asset, and* Z_t denotes a “fractional Brownian motion” with the Hurst index $H > \frac{1}{2}$.

An adapted version of Itô calculus is utilized to make the fractional Black-Scholes SDE (stochastic differential equation) mathematically traceable. Finally, the binary market model is constructed by approximating this SDE, providing a discrete-time model that shows a better approximation than the random-walk-based model in predicting stock behavior.

DOI: 10.54254/2753-8818/2025.DL34030

5.6. Efficiency analysis on Brazilian stock market

Benjamin Miranda Tabak [6] mainly employs the IBOVESPA index for the purpose of analyzing the degree to which the Brazilian stock market, known as BOVESPA, adheres to the RWH, or the random walk hypothesis. The RWH posits that the market price is entirely random and cannot be predicted, thus suggesting that the market operates in an efficient manner.

The IBOVESPA index is calculated by:

$IBOV_t = \Sigma_{i=1}^{n} P_{i,t} Q_{i,t} \#$ (5.13) Here, calculates the price of stock i and time t, and represents the shares number of stock i considered in the index at time t.

The results concluded by Benjamin Miranda Tabak indicate that the RWH is rejected in the case of the entire sample. Upon splitting the sample, it was observed that this rejection was mainly due to the pre-1994 data. This suggests that the market post-1994 is moving towards being more efficient. The accuracy of the result is safeguarded by statistical tests that are resilient against both homoscedasticity and heteroscedasticity. This is as the inflow of capital is cointegrated with the stock market (BOVESPA). Moreover, there exists a bicausality relationship between them, which means that the inflow enhances the market, and the improvement of the market strengthens the inflow, rendering it more efficient. These conclusions are derived from the statistical analysis of IBOVESPA returns, employing tests like autocorrelation and variance ratios to evaluate the random walk hypothesis.

6. Conclusion

As computer technology advances at a rapid pace, the efficiency of search engines as well as the accuracy of stock predictions have become increasingly crucial. The progress in search algorithms and stock models is closely associated with Stirling's formula and the Random Walk that were discussed in this paper. There is no doubt that traditional search engines and models possess certain advantages. However, these algorithms frequently yield an overwhelming volume of information, and the financial model often fails to make a precise evaluation of the market's efficiency. In such situations, the Random Walk not only provides a knowledge foundation for advancements but also offers inspiration for innovations. It incorporates algorithms into financial market models, and it introduces the concept of connecting web pages or items on a graph to assess their relevance – both of which bring convenience and advancement to daily life. Nevertheless, the Random Walk still holds considerable potential for future development. To begin with, the exact probability and fine properties of the Random Walk-in higher dimensions remain unknown. Moreover, although the mixing rate of the walk can be analyzed accurately for some special graphs, determining precise bounds on arbitrary graphs can still be challenging. Furthermore, even though mathematicians can often calculate the upper bounds and the lower bounds of cover time on certain graphs, this cannot be generalized to all types of graphs. These challenges present opportunities for future breakthroughs and innovations. Overcoming these issues enables the network and algorithm to become faster, more accurate, and more efficient.

Acknowledgement

Yifei Lu and Xinyi Liang contributed equally to this work and should be considered co-first authors

DOI: 10.54254/2753-8818/2025.DL34030

References

[1] Novak J. Polya's random walk theorem [J]. The American Mathematical Monthly, 2014, 121(8): 711-716.
[2] Lovász L. Random walks on graphs [J]. Combinatorics, Paul erdos is eighty, 1993, 2(1-46): 4.
[3] Page L, Brin S, Motwani R, et al. The PageRank citation ranking: Bringing order to the web [R]. Stanford infolab, 1999.
[4] Bachelier L. Théorie de la spéculation [C]//Annales scientifiques de l'École normale supérieure. 1900, 17: 21-86.
[5] Sottinen T. Fractional Brownian motion, random walks and binary market models [J]. Finance and Stochastics, 2001, 5(3): 343-355.
[6] Miranda Tabak B. The random walk hypothesis and the behaviour of foreign capital portfolio flows: the Brazilian stock market case [J]. Applied Financial Economics, 2003, 13(5): 369-378.
[7] Chaibi L F. Testing the random walk: the case of Hong Kong stock exchange [J]. Journal of Empirical Studies, 2014, 1(2): 54-61.
[8] Conrad, K. (n.d.). Stirling's formula. [Unpublished manuscript].
[9] Zui Mei Shu Xue Xi Lie. (2023, October 15). Zui jian dan de Sitelin Stirling gong shi de tui dao fang fa [Video]. BiliBili. Retrieved November 5, 2023, from https: //b23.tv/cRaqEs6
[10] Feller, William. An Introduction to Probability Theory and Its Applications. Vol. 1, 3rd ed., Wiley, 1968.
[11] Novak, J. (n.d.). Polya's random walk theorem. [Unpublished manuscript].
[12] Pólya, G. (1921). Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz. Mathematische Annalen, 84(1–2), 149–160.
[13] Konvalina J. A unified interpretation of the binomial coefficients, the Stirling numbers, and the Gaussian coefficients [J]. The American Mathematical Monthly, 2000, 107(10): 901-910.
[14] Pucci A, Gori M, Maggini M. A random-walk based scoring algorithm applied to recommender engines [C]//International Workshop on Knowledge Discovery on the Web. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006: 127-146.
[15] Zhang Y C, Bao W S, Wang X, et al. Effects of systematic phase errors on optimized quantum random-walk search algorithm [J]. Chinese Physics B, 2015, 24(6): 060304.