Random Walk Regression Problem in One and Two Dimensions

Chunyao Jia*

School of Mathematical Sciences, Nankai University, Tianjin, 300110, China

*Corresponding author: 2113425@mail.nankai.edu.cn

Abstract. This paper examines the previous discoveries and research on the random walk problem, considers the practical application of random walk. This paper starts from the basic questions and conducts a more detailed study of the basic regression problem with one-dimensional and two-dimensional situations. Markov chains and Markov property are introduced, and then the method and important properties of series convergence are properly explained and demonstrated. Through the introduction of Brownian motion, the key part of the article is introduced, and the explanation and proof of Stirling's formula are carried out, then the proof of one-dimensional cases is entered, and the regression is proved by the above knowledge and deduction, and then the same idea is extended to the two-dimensional case for proof. Through the research and proof of this paper, the basic problem of random walk will be solved and well interpreted, laying the foundation for further research and development.

Keywords: Random walk; series convergence; binomial theorem; Stirling's formula.

1. Introduction

Random walk is a classic and interesting problem, which is a fundamental process used to describe dynamic stochastic phenomena [1]. Since the development law of many things can be regarded as a dynamic random phenomenon, the idea of random walk problem is widely used in regression analysis, computer algorithm design, financial market, economic prediction, error of fiber optic gyro, medicine and many other fields [2-4].

The random walk focuses on a walker traversal a space through a random motion and tries to consider obtaining a probability distribution after each walk that characterizes the probability that each point in the graph is visited. Using this probability distribution as the input to the next walk and iterating the process, a stable probability distribution is achieved. Through learning and further in-depth study of the random walk problem, the development rules of many things could be mastered, so as to better promote the development of various industries and society.

Based on the study of random walk, Wu et al. have applied it to the field of neural network [5]. He and Li's algorithm research problem in social networks is based on the random walk [6]. Zhang et al. have applied it on the risk prediction of diabetes [7]. Zhou et al. expanded to the study of heterogeneous information network characterization [8]. Riju et al. have applied it for a community detection model [9]. And other important cutting-edge fields with dynamic random randomness are based on random walk, the importance of this basic problem research is self-evident.

Another extremely important and socially related is the application of random travel in the economic field. Zhang et al. have its application in macroeconomic forecast [10]. On the futures market forecast and research, Hou et al. have applied non-parametric method considering the Chinese apple market [11]. Zhang considered the efficiency of the EU coal related market [12]. Random walk model can also be used to test the effectiveness of the stock market [13]. It can be said that random walk can be analyzed and predicted in all aspects of the economic field, which shows that random walk is a very worthwhile and universal problem.

Initially, the Hungarian mathematician Polya researched on random walk problem. Polya has a very famous theorem, called Polya's recurrence theorem. Polya's recurrence theorem is a fundamental concept in probability theory and random walks. This theorem states that in one-dimensional and two-dimensional lattices, a random walk is recurrent, meaning the path will eventually return to its starting point with probability one. In more than two dimensions, however, the walk becomes transient, indicating that there's a nonzero probability that the path will never return to the starting point.

To understand this in simpler terms, imagine a random walk as the path of a drunk person trying to find their way home. In a one-dimensional world or a two-dimensional world, the theorem suggests that no matter how randomly the person walks, they will eventually make it back home. This is due to the limited directions they can move in these lower dimensions, making it inevitable for them to revisit points, including their starting point. In contrast, in a world with more than two dimensions, like a drunk bird flying through three-dimensional space, the theorem posits that the bird may never return to its starting point. This is because the increased number of dimensions provides more paths and directions to take, greatly reducing the probability of returning to any specific point.So, Polya's recurrence theorem is a mathematical way of saying that in low dimensions, a random walker will always find their way back to the start, but in higher dimensions, they might get lost forever.

In this paper, the Markov chains and the conditional probabilities will be introduced and used for the later proof, the Polya's recurrence theorem will be elaborated and explored, and the random walk regression problem in one and two dimensions will be investigated and proved.

2. Methods

2.1. The Markov Chains

Definition 1. A Markov Chains is a random sequence $\{X_n, n \geq 0\}$fori_0, i_1,⋯,i_n,$i_{n+1} \in S$, n$\in \mathrm{N}_0$ and $\mathrm{P}\{\mathrm{X}_0 = i_0, \mathrm{X}_1 = i_1, \cdots., \mathrm{X}_\mathrm{n} = i_n\} > 0$, such that,

$$\mathrm{P}\{X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \cdots., X_n = i_n\} = \mathrm{P}\{X_{n+1} = i_{n+1} | X_n = i_n\}. \tag{1}$$

To intuitively understand the meaning of Markov property, imagine a particle making a random motion on an integer point of a straight line, where X_n indicates the position of the particle at time n, $X_n = i$ represents the random event that the particle is in the state i at time n. If time n is regarded as the present, $0,1,\cdots,n-1$ is the past, and $n+1$ as the future, then the above definition indicates that the conditional probability of the particle at the future moment $n+1$ depends only on the present event, but independently of what happened in the past. In short, given the known present, the future is independent of the past. The random walk that will be studied later will satisfy the Markov property.

2.2. Series Convergence

Definition 2. Let the function $f(x)$ is non-negative and decreasing on the interval $[1, \infty)$, $A_n = \int_1^n f(x)\,dx, n = 1,2,\cdots$, then the series $\sum_{n=1}^{\infty} f(n)$ converge or diverge simultaneously with the sequence $\{A_n\}$.

Lemma 1. $\sum_{n=1}^{\infty} \frac{1}{n^\alpha} = \begin{cases} \infty, 0 < \alpha \leq 1 \\ finite, 1 < \alpha \end{cases}.$

Proof of lemma 1: Let $f(x) = \frac{1}{x^\alpha}$, when $\alpha = 1$,

$$A_n = \int_1^n f(x)\,dx = \int_1^n \frac{dx}{x} = \ln n \to +\infty \ \ (n \to \infty). \tag{2}$$

Therefore, by definition 2, it can be known that the series $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$ is divergent when $\alpha = 1$. Moreover,it can be deduced that $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$ is divergent when $0 < \alpha \leq 1$.

When $\alpha > 1$,

$$A_n = \int_1^n f(x)\,dx = \int_1^n \frac{dx}{x^\alpha} = -\frac{1}{(\alpha-1)x^{\alpha-1}}\bigg|_1^n = -\frac{1}{(\alpha-1)n^{\alpha-1}} + \frac{1}{\alpha-1} \to \frac{1}{\alpha-1} \ \ (n \to \infty) \tag{3}$$

Therefore, by definition 2, it can be known that the series $\sum_{n=1}^{\infty}\frac{1}{n^{\alpha}}$ is convergent when $\alpha > 1$. Consequently, the series $\sum_{n=1}^{\infty}\frac{1}{n^{\alpha}}$ is divergent when $0 < \alpha \leq 1$, and convergent when $\alpha > 1$.

2.3. Binomial Theorem

Definition 3. In line with the binomial theorem, any non-negative exponent of the binomial $(x + y)$ can be expanded into a series represented by a sum in a specific form,

$$(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^k. \tag{4}$$

Lemma 2. $\binom{2n}{n} = \sum_{k=0}^{n}\left(\binom{n}{k}\right)^2$.

Proof of lemma 2: By definition 3, let $y = 1$, plugging in $\mathrm{n} = \mathrm{a},\ \mathrm{b},\ \mathrm{a}+\mathrm{b}$ respectively,

$$(x+1)^a = \sum_{i=0}^{a}\binom{a}{i}x^i, \tag{5}$$

$$(x+1)^b = \sum_{j=0}^{b}\binom{b}{j}x^j, \tag{6}$$

$$(x+1)^{a+b} = \sum_{k=0}^{a+b}\binom{a+b}{k}x^k. \tag{7}$$

Since the first equation times the second equation is equal to the third equation, so,

$$(x+1)^a(x+1)^b = \left(\sum_{i=0}^{a}\binom{a}{i}x^i\right)\left(\sum_{j=0}^{b}\binom{b}{j}x^j\right) = \sum_{n=0}^{a+b}\left(\sum_{k=0}^{n}\binom{a}{k}\binom{b}{n-k}\right)x^n = \sum_{k=0}^{a+b}\binom{a+b}{k}x^k = (x+1)^{a+b}. \tag{8}$$

Therefore,

$$\binom{a+b}{n} = \sum_{k=0}^{n}\binom{a}{k}\binom{b}{n-k}. \tag{9}$$

Now, plugging in $\mathrm{a} = \mathrm{b} = \mathrm{n}$,

$$\binom{2n}{n} = \sum_{k=0}^{n}\binom{n}{k}\binom{n}{n-k} = \sum_{k=0}^{n}\left(\binom{n}{k}\right)^2. \tag{10}$$

3. Results and Discussion

Imagine a minuscule particle floating in a uniform fluid. This particle experiences random impacts with the fluid's molecules, leading to its continuous and erratic movement. This is what we refer to as Brownian motion. To model Brownian motion, we use a simplified concept known as a "discrete random walk" in a one-dimensional space:

The particle's movement is confined to the x-axis. It moves only at specific intervals, $t = n\Delta t$, where n can be 0, 1, 2, and so on. Assume the particle is at position x at a certain time t. Then, irrespective of its past movements, the particle has an equal chance to shift to either $x + \Delta x$ or $x - \Delta x$ $(\Delta x > 0)$ with probability $\frac{1}{2}$. This means that at each interval, the particle randomly moves a distance Δx either to the right or to the left, with equal probability.

3.1. Stirling's Formula

Theorem 1. Stirling's formula provides an estimated calculation for the factorial function $\mathrm{n}!$ for $\mathrm{n} \gg 1$, it can be shown that

$$\mathrm{n}! \sim \sqrt{2\pi\mathrm{n}}\,\mathrm{n}^{\mathrm{n}}\,\mathrm{e}^{-\mathrm{n}}. \tag{11}$$

Proof of Theorem 1: The equation can be obtained as well by utilizing the integral interpretation of the factorial,

$$\mathrm{n}! = \int_0^{\infty}\mathrm{e}^{-\mathrm{x}}\,\mathrm{x}^{\mathrm{n}}\,dx. \tag{12}$$

Be aware that the derivative of the integrand's logarithm is expressible in a certain way,

$$\frac{d}{dx}\ln(e^{-x}\,x^{n})=\frac{d}{dx}(n\ln x-x)=\frac{n}{x}-1 \tag{13}$$

The integrand exhibits a pronounced peak, with significant contributions occurring primarily around $\mathrm{x}=\mathrm{n}$. Therefore, define x as $n+\varepsilon$ where $\varepsilon \ll n$,and write

$$\ln(e^{-x}\,x^{n})=n\ln x-x=n\ln(\mathrm{n}+\varepsilon)-(\mathrm{n}+\varepsilon). \tag{14}$$

Now,

$$\ln(\mathrm{n}+\varepsilon)=\ln n+\ln\left(1+\frac{\varepsilon}{n}\right)=\ln n+\frac{\varepsilon}{n}-\frac{1}{2}\frac{\varepsilon^2}{n^2}+\cdots. \tag{15}$$

So,

$$\ln(e^{-x}\,x^{n})=n\ln(\mathrm{n}+\varepsilon)-(\mathrm{n}+\varepsilon)=n\ln n+\varepsilon-\frac{1}{2}\frac{\varepsilon^2}{n}-n-\varepsilon+\cdots=n\ln n-n-\frac{\varepsilon^2}{2n}+\cdots \tag{16}$$

Transforming each side by taking the exponential results in

$$e^{-x}\,x^{n}\approx e^{n\ln n}e^{-n}e^{-\varepsilon^2/2n}=n^{n}e^{-n}e^{-\varepsilon^2/2n}. \tag{17}$$

Substituting this into the integral formula for $n!$ yields

$$\mathrm{n}!\approx\int_{-\mathrm{n}}^{\infty}n^{n}e^{-n}e^{-\varepsilon^2/2n}\,d\varepsilon\approx n^{n}e^{-n}\int_{-\infty}^{\infty}e^{-\varepsilon^2/2n}\,d\varepsilon. \tag{18}$$

Evaluating the integral gives

$$\mathrm{n}!\approx n^{n}e^{-n}\sqrt{2\pi n}. \tag{19}$$

3.2. One-Dimensional Random Walk

Similar to the discrete random walk described above, to facilitate subsequent studies and proof, assume that the particle is at original point in the beginning and $\Delta t=1,\Delta x=1$. For an arbitrary number of steps $n\gg 1$, if the probability of the particle ultimately returning to the starting point is 1, it is called recurrent. In this part, the thing that the particle is recurrent in dimension one will be proved.

Definition 4. Note that u_n is the probability of returning back to the origin with n steps, $n=0,1,2,\cdots$, with $u_0=1$.Note that f_n is the probability of returning back to the origin with n steps for the first time, $n=0,1,2,\cdots$, with $f_0=0$.

In this paper, say that the state is recurrent if $\sum_{n=1}^{\infty}f_n=1$,otherwise say that the state is transient if $\sum_{n=1}^{\infty}f_n<1$.

Since the random walk satisfies the Markov property, there is a relation between u_n and f_n,

$$u_n=u_0f_n+u_1f_{n-1}+u_2f_{n-2}+\cdots+u_{n-1}f_1+u_nf_0, \tag{20}$$

Multiply u_n by x^n, $n=0,1,2,3,\cdots$, and note that

$$U_n(x)=\sum_{k=0}^{n}u_kx^k \tag{21}$$

and

$$\lim_{n\to\infty}U_n(x)=U(x). \tag{22}$$

Do the same with f_n, then

$$F_n(x)=\sum_{k=0}^{n}f_kx^k \tag{23}$$

and

$$\lim_{n\to\infty}F_n(x)=F(x). \tag{24}$$

Now,

$$U_n(x)F_n(x) = \left(\sum_{k=0}^{n} u_k x^k\right)\left(\sum_{k=0}^{n} f_k x^k\right)$$
$$= \sum_{k=0}^{n}\left(\sum_{i=0}^{k} u_i f_{n-i}\right) x^k$$
$$= \sum_{k=0}^{n}(u_k)\, x^k$$
$$= U_n(x) - 1. \tag{25}$$

This shows that,

$$F_n(x) = 1 - \frac{1}{U_n(x)}. \tag{26}$$

Plugging in $x = 1$ gives

$$F_n(1) = 1 - \frac{1}{U_n(1)}. \tag{27}$$

On the other hand, $u_{2n} = \binom{2n}{n}\left(\frac{1}{2}\right)^n$, using the corner marker as $2n$ since that it can not return back to the origin if it moves odd steps.

Using Stirling's formula, by Theorem 1 gives,

$$u_{2n} = \binom{2n}{n}\left(\frac{1}{2}\right)^n = \frac{(2n)!}{n!n!}\left(\frac{1}{2}\right)^n \sim \frac{\sqrt{2\pi 2n}\,(2n)^{2n}\,e^{-2n}}{\sqrt{2\pi n}\,n^n\,e^{-n}\,\sqrt{2\pi n}\,n^n\,e^{-n}}\left(\frac{1}{2}\right)^n = \frac{1}{\sqrt{\pi}}\frac{1}{\sqrt{n}}. \tag{28}$$

Based on the facts mentioned in the Lemma 1, indeed, $\sum_{n=o}^{\infty}\frac{1}{\sqrt{n}} = \infty$. So,

$$U(1) = \lim_{n\to\infty} U_n(1) = \sum_{n=0}^{\infty} u_n = \sum_{n=0}^{\infty} u_{2n} = \frac{1}{\sqrt{\pi}}\sum_{n=o}^{\infty}\frac{1}{\sqrt{n}} = \infty. \tag{29}$$

Then,

$$\sum_{n=1}^{\infty} f_n = F(1) = \lim_{n\to\infty} F_n(1) = \lim_{n\to\infty}\left(1 - \frac{1}{U_n(1)}\right) = 1. \tag{30}$$

This means that the particle is recurrent in dimension one.

3.3. Two-Dimensional Random Walk

Imagine the case in which the particle moves in a two-dimensional plane as mentioned in the one-dimensional case. As previously defined, assume that the particle starts from the origin and moves by a distance of 1 at each step. For an arbitrary number of steps $n \gg 1$, if the probability of the particle ultimately returning to the starting point is 1, it is called recurrent. In this part, the thing that the particle is recurrent in dimension two will be proved.

Note that k_1 means stepup, k_2 means stepdown, k_3 means stepleft and k_4 means stepright with $k_1 + k_2 + k_3 + k_4 = 2n$, $k_1 = k_2 = k$, $k_3 = k_4 = n - k$,so that the particle could return back to origin point.

The definition of u_n and f_n is the same as definition 4. So formula (20) to formula (27) still holds in this two-dimensional case.

Unlike in the previous discussion, now,

$$u_{2n} = \sum_{k=0}^{n}\binom{2n}{k}\binom{2n-k}{k}\binom{2n-2k}{n-k}\binom{n-k}{n-k}\left(\frac{1}{4}\right)^{2n} = \sum_{k=0}^{n}\frac{(2n)!}{k!k!\,(n-k)!\,(n-k)!}\left(\frac{1}{4}\right)^{2n} = \sum_{k=0}^{n}\frac{n!}{k!\,(n-k)!}\,\frac{n!}{k!\,(n-k)!}\frac{(2n)!}{n!\,n!}\left(\frac{1}{4}\right)^{2n} = \left(\frac{1}{4}\right)^{2n}\binom{2n}{n}\left[\sum_{k=0}^{n}\left(\binom{n}{k}\right)^2\right]. \tag{31}$$

By lemma 2 and formula (28),

$$u_{2n} = \left(\frac{1}{4}\right)^{2n} \binom{2n}{n} \left[\sum_{k=0}^{n} \left(\binom{n}{k}\right)^2\right] = \left(\frac{1}{4}\right)^{2n} \binom{2n}{n}\binom{2n}{n} = \left[\left(\frac{1}{2}\right)^n \binom{2n}{n}\right]^2 \sim \frac{1}{\pi}\frac{1}{n}. \tag{32}$$

Based on the facts mentioned in the Lemma 1, indeed, $\sum_{n=o}^{\infty} \frac{1}{n} = \infty$. So,

$$U(1) = \lim_{n\to\infty} U_n(1) = \sum_{n=0}^{\infty} u_n = \sum_{n=0}^{\infty} u_{2n} = \frac{1}{\pi}\sum_{n=o}^{\infty} \frac{1}{n} = \infty. \tag{33}$$

Then,

$$\sum_{n=1}^{\infty} f_n = F(1) = \lim_{n\to\infty} F_n(1) = \lim_{n\to\infty} \left(1 - \frac{1}{U_n(1)}\right) = 1. \tag{34}$$

This means that the particle is recurrent in dimension two.

4. Conclusion

In this article, the definition of a Markov chain is given. The method of judging series convergence and the content of the binomial theorem are demonstrated, and two important lemmas in these two contents are proved in detail. The focus of this paper is the random walk elicited by Brownian motion, some definitions and important Stirling's Formula with proof are given. Then, after some normalization operations that are convenient for proof, based on the idea of applying the limit and the related knowledge of series, with using Stirling's Formula to estimate, the particle is recurrent in dimension one and dimension two are clearly and concisely proved. Therefore, the problem of random walk regression in both one-dimensional and two-dimensional cases is well proven.

References

[1] Gao Hong. Concept error and correction of the random walk definition. Mathematics Learning and Research, 2021, 28: 136-139.

[2] Zheng Wei, Wang Chaokun, Liu Zhang, et al. A multi-label classification algorithm based on a random walk model. Journal of Computer Science, 2010, 33(08): 1418-1426.

[3] Zhang Bing, Li Xiaoming. Study on the incremental effectiveness of China Stock market. Economic Research, 2003, 1: 54-61 + 87-94.

[4] Han Junliang, Ge Shengmin, Shen Yi. Modeling and Simulation study of a digital closed-loop fiber-optic gyro. Journal of System Simulation, 2008, 4: 833-836 + 881.

[5] Wu Bo, Liang Xun, Zhang Shusen, et al. Figure Frontier progress and application of neural network. Journal of Computer Science, 2022, 45, 1: 35-68.

[6] He Jianfeng, Li Hongyu. Cluster random walk sampling algorithm based on social networks in the context of big data. Statistical study, 2021, 38(04): 131-144.

[7] Zhang Chunfu, Wang Song, Wu Yadong, et al. Risk prediction of diabetes mellitus based on the GA-Xgboost model. Computer Engineering, 2020, 46(03): 315-320.

[8] Zhou Lihua, Wang Jialong, Wang Lizhen, et al. Summary of heterogeneous information network representation learning. Journal of Computer Science, 2022, 45(01): 160-189.

[9] Riju B, Kumar N N, Sarsij T. A community detection model using node embedding approach and graph convolutional network with clustering technique. Decision Analytics Journal, 2023, 9.

[10] Zhang Jinfan, Gang Jianhua, Qian Zongxin, et al. Macroeconomic prediction based on a hybrid vector autoregression model. Financial Research, 2018, 7: 34-48.

[11] Hou Xiaokang, Liu Tianjun, Yuan Xuepei. Is the Chinese Apple futures market valid -Test based on non-parametric methods. Agricultural economic problems, 2022, 5: 98-109.

[12] Zhang Yuan-kun. Research on the Efficiency of the EU carbon Futures market. Jilin University, 2023.

[13] Sun Lixu, Yang Xiaoying. Research on testing the effectiveness of the stock market based on the random walk model. Cooperative Economy and Technology, 2019, 20: 74-75.