Scale-Free First Return Probability: An Extensive Investigation

Afrina Sharmin, Md. Kamruzzaman

Abstract: - We flip a coin and it is head with a probability *p = 1/2* and tail with a probability *q = 1 - p = ½*. Consider if the coin land head we gain a dollar and if it is tail we lose a dollar. The game continues until a gambler who started with a finite, playing this fair game, go broke-known as the gamblers ruin problem. Also in the context of the random walk problem if one assumes that *t* denote the time at which the walker starting at the origin return to the origin. The purpose of this paper is to verify the theoretical prediction that the probability of the walker returns to the origin for the first time decay following the power law, $F(t) \sim t^{-3/2}$. In addition we show that it has fat tail which is reminiscent of the power law degree distribution of the scale free complex network as predicted by the Barabási-Albert Model.

Index Terms: - random walk, gambler's ruin, first return time, nodes, links, hubs, preferential attachment rule, power-law degree distribution, fat tail, cumulative distribution.

1 Introduction

Many natural phenomena, like Brownian motion, are modeled by random walk idea. Hence, random walks are traditionally explained in the context of some other social vice, such as the position of a drunkard who randomly staggers right or left or just vacillates in place during each time step. We consider the classical gambler's ruin problem which can be solved by total probability theory. We assume a game where a gambler, who start with respective bankroll, wins 1 unit each time with probability *p = 1/2* and losses with probability *q = 1 − p = 1/2* independently of the other moves. The game stops when the fortunes of the gambler becomes zero (Gambler's Ruin). This phenomenon is considered as the walker, initially at the origin, returns to the origin by the time defined as 'first return'. Hence, this first return time theoretically follows a power law with fat tailed distribution. Our aim is the verification of the theory. Moreover, we intend to observe the power law tail of first return probability which reminds us about the degree distribution of complex network. In 1999, Rèka Albert and Albert-Lászlo Barabási proposed a completely new model (BA model), which approaches the topological features of real-life networks and its mechanism completely explains the appearance of the power-law distribution.

2 Theoretical Models

2.1 Power Law: Heavy Tailed Distribution

Power law is a mathematical relationship which is **homogeneous** by nature and exhibits the **property of scale invariance** as well as the **universality**. One of the generalized forms of the power laws relates two variables such as

$$f(x) = ax^k + \Theta(x^k) \quad (1)$$

- *Afrina Sharmin, Scientific Officer, IFRD, Bangladesh Council of Scientific and Industrial Research (BCSIR), Dhaka - 1205, Bangladesh.*
- *E-mail: sharmin_afrina@yahoo.com*

- *Md. Kamruzzaman, Lecturer at Ahsanullah University of Science and Technology, Dhaka - 1208, Bangladesh.*
- *E-mail: m_kamruzzaman15@yahoo.com*

where, a and k are constant and $\Theta(x^k)$ is an asymptotically small function of x^k. Here k is typically called the **scaling exponent.** Power law tail is also known as the long tail, heavy tail, Pareto's tail or Fat tail was first coined by Chris Anderson in an article of 'wired magazine' in October 2004. The scientific study of the power law tail started in 1946. According to the Probability theory, heavy tailed distribution is probability distribution whose tails are not exponentially bounded i.e. they have heavier tails than the exponential distribution. In many applications it is the right tail of the distribution that is of interest.

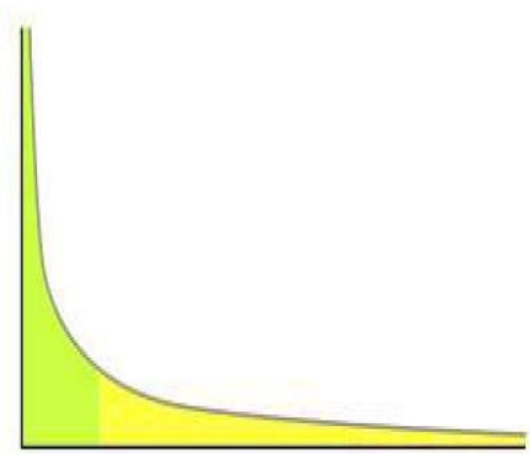

Fig.1: Long tail distribution

The distribution curve has a shape as Fig.1. The long-tailed curve shows a degree of inequality in the frequency distribution.

2.2 The Concept of Random Walk from Brownian Motion

Brownian motion is a phenomena of small particles suspended in a liquid which tends to move in stochastic paths through the liquid. Einstein noticed that the motion is caused by random bombardment of water molecules on the pollen. Mathematical Brownian motion is related to the random walk idea where the displacement of particles is randomized. It's a stochastic process according to Modern Theory. An approximation of 2-dimensional Brownian motion can be described as *drunken man wandering around the square*. Consequently, it has the

Markov property which assures that the future state of the particles is determined entirely by its current state, not by any past state. Surprisingly enough, the story of the Brownian motion starts with a botanist [1]. The term Brownian motion derives its name from the famous scientist **Robert Brown**, a pioneer of Botany. In 1928, he began to make microscopic observations of suspensions of grains released from pollen sacks taken from a type of evening primrose called *Clarkia pulchella* [2]. And it was Perrin's microscope studies of Brownian particles that confirmed Einstein's theory and sealed the reality of the discontinuous, atomic nature of matter. In 1906, Smoluchowski first published an one-dimensional model to describe a particle undergoing Brownian motion [3]. In 1908, Perrin and his team of research students embarked on an exhaustive set of experiments [4]. Tragically, many of Perrin's team would lose their lives only a few years later in the First World War. However, the term random walk was first introduced by Karl Pearson in 1905 [5].

2.3 Random Walk In 1 D

We assume a particle suffering displacements in the form of a series of steps of equal length along a straight line (initially it is in 0 position), can move either in forward or in backward direction.

Now the equation of Brownian motion is derived as

$$\frac{\partial P(x,t)}{\partial t} = D\,\frac{\partial^2 P(x,t)}{\partial x^2} \qquad (2)$$

Now Eq-(2) is a differential equation for probability density *P(x,t)*, as well as, a well known Diffusion equation. So we can conclude that Brownian motion can be studied from the point view of diffusion.

and
$$P(x,t) = \frac{1}{\sqrt{4\pi Dt}}\exp[x^2/4DT] \qquad (3)$$

And hence it's the solution of diffusion equation.
Consequently, we obtain some interesting features from the solution as follows:

(a) Mean displacement, $<x> = 0$ (4)

(b) Root mean square displacement,

$$\overline{x^2} = < x^2 > = \frac{\int_{-\infty}^{\infty} x^2 P(x,t)dx}{\int_{-\infty}^{\infty} P(x,t)dx} = \int_{-\infty}^{\infty} x^2\, P(x,t)dx \qquad (5)$$

(c) Variance, $< x^2 > = 2Dt$ (6)

2.4 Gamblers' Ruin and First Return Time

A phenomenon called Gambler's Ruin says that a gambler raises his bet to a fixed fraction of bankroll when he wins, but does not reduce it when he losses. If he plays long enough, he will go bankrupt, or we say, he will eventually go broke. And the gambler with a finite wealth, playing a fair game will go broke against an opponent with infinite wealth, though each bet has expected value zero to both side. Consequently, a gambler playing a negative expected value game will eventually go broke. Both of the result above is the corollary of a general theorem by Christian Huygens [6]. In the study of stochastic processes, the first return time (or first hit time or first passage time) [7] is a particular instance of a stopping time, the first time at which a given process 'hits' a given subset of the state space. Exit times and return times are also examples of hitting times. A common example of a first-hitting-time model is a ruin problem, such as Gambler's ruin. In this example, a gambler has an amount of money which varies randomly with time. The model considers the event that the amount of money reaches 0, representing bankruptcy. The model can answer questions such as the probability that this occurs within finite time, or the mean time until it occurs.

2.5 Theory of first return time

By using the Markov chain property in the Random Walk problem, it is obtained that the probability distribution function *P(x,t)* of random walk problem obeys diffusion equation. Let us assume the probability of the walker returns to position *0* is *ut*, and *ft*, the probability that the first returns time is *t*. Now,

$$u_t = u_{2n} = \begin{cases} 1 & for\ n = 0 \\ \sum_{n=1}^{n} f_{2n}\, u_{2n-2m} & for\ n \geq 1 \end{cases} \qquad (7)$$

Here we define

$$U(z) = \sum_{n=0}^{\infty} u_{2n} z^2 \qquad (8)$$

$$F(z) = \sum_{n=1}^{\infty} f_{2n} z^n \qquad (9)$$

And *t = 2n* , where *n* is an integer.

Using the binomial theorem and Stirling formula, we obtain

$$F(z) = \sum_{n=1}^{\infty} \frac{\binom{2n}{n}}{(2n-1)2^{2n}}\, z^n$$

where,
$$f_{2n} = \frac{\binom{2n}{n}}{(2n-1)2^{2n}}$$

In the limit $n \to \infty$, $f_{2n} \sim n^{-3/2}$ or $f_t \sim t^{-3/2}$ (10)

Here *t* is called power law decay time.

2.6 Scale free network: A study of BA model

The study of network began in 1959 by Paul Erdös and Albert Rènyi [8, 9, 10]. Their investigation then known as Graph theory or the Erdös Rènyi model [11, 12]. A complex network is explained by three basic idea-(a) Small world [13, 14], (b) Degree Distribution [15, 16, 17] and (c) Clustering. A new generalized model was proposed by Watts and Strogatz[18] in 1998. This new model was intermediate between the regular and random graph and even for a small probability P the graph behaves different just like ER model. For the first time in 1990, A. L. Barabási, R. Albert and H. Jeong proposed another model using preferential attachment mechanism [19]. For a large number of networks including the World Wide Web (Albert, Jeong,Barabási, 1999) [20] , Internet (Faloutsos el al. 1999) or metabolic networks (Jeong el al.2000) [21] the degree distribution has a power law tail , such networks are called scale free networks (Barabási and Albert 1999). The two ingredients, Growth and Preferential attachment, introduce the scale free (SF) network model that has a power law degree distribution. The algorithm [22] is as follows-

(i) Growth: Starting with a small number (m_0) of nodes at every time step, we add a new node with $m(m_0)$ edges that link the new node to m different nodes already present in the system.

(ii) Preferential Attachment: When choosing the nodes to which the new node connects, we assume that the probability π that a new node will be connected to node i depends on the degree k_i of node i, such that

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \quad (11)$$

After t time steps this algorithm results in a network with N = t + m_0 nodes and mt edges. Numerical simulations indicated that this network evolves into a scale invariant state with the probability that a new node has k edges following a power law with an exponent, $\gamma_{SF} = 3$. The scaling exponent is independent of m, the only parameter of the model. In the very new model the hubs tend to accumulate more links. And every new node prefers to attach with these hubs. However, in 1925 Yule [23] first uses preferential attachment mechanism, but it was not praised because of lacking of proper explanation. In 1965, Derek de Solla Price [24] noticed the fact that power law degree distribution in citation and proposed a model similar to BA model. Consequently, Hilbert Simon [25] enounced the rich get richer idea in 1955 which again follows the power law. The study of complex network is largely inspired by the empirical study of real world networks whose behavior cannot be described as purely regular or purely random. The BA model is one of the several proposed models that finally generates scale free network.

3 Methodology

We have written a program based on random walk idea. One of the most famous and efficient methods is the Monte Carlo method for producing random numbers. This method was first used in 1949. The American mathematicians John Von Neumann & Stanislaw Ulam are considered as its originator.
The program is written to carry out a trial and is repeated N times, each trial being independent of the rest. The simulation has been performed in a core 2quad PC with processor 2.93 GHz and memory 3 GB.

3.1 Algorithm

(i). At the beginning, the walker is at the origin of a *1-D* lattice.

(ii). Generate a random number *R*.

(iii). Check *R*. If $R \le p$ = *0.50*, go to step (iv); Otherwise go to step (vi).

(iv). The walker moves to the left with respect to its origin.

(v). Increase the time by one unit.

(vi). If *R > p = 0.50*, the walker moves to the right.

(vii). Increase the time by one unit and go to step (ii).

(viii). Repeat the steps (i)-(vii) ad infinitum.

(ix). Set counter-1 to count each time when the walker returns to the origin.

(x). Set counter-2 to count the time steps for each event of returning to origin by the walker.

(xi). After counting time step, T_1 = *2, 50, 000* by counter-1, random number generator is stopped which determines our sample size.

(xii). By counter-2, the *2, 50,000* events is counted, where in each event, the walker has returned to the origin by different number of steps.

(xiii). Make a table for different intervals of steps and count how many times the walker returned to origin for that particular interval and hence, calculate the probability for that interval as Table-1. Here, $\sum N$ = *2, 50, 000* and $\sum F(t)$ = *1*.

(xiv). Hence, *F(t) vs. t* is plotted.

4 Results and Discussion

Table 1
An imaginary table to illustrate how first return probability, *F(t)* is defined.

Interval of steps	Average, *t*	Times of return, *N*	Probability, *F(t)*
1-10,000	5,000	35,711	0.034
10,001-20,000	15,000	22,800	0.029
20,001-30,000	25,000	9,040	0.025
30,001-40,000	35,000	4,490	0.02
40,001-50,000	45,000	2,709	0.0004
..	..	..	..
..	..	..	..
..	..	..	..0.00008

4.1 rms Displacement

The root mean square displacement of Brownian motion is

$$< x^2 > = 2Dt \quad (12)$$

So, $$x_{rms} \sim t^{1/2} \quad (13)$$

So, *rms displacement Vs. time* curve shows the monotonically increases of the displacement. Hence, on logarithmic scale it shows a straight line with slope *1/2*. Consequently, *rms displacement Vs.* $\sqrt{t}$ will show a straight line.

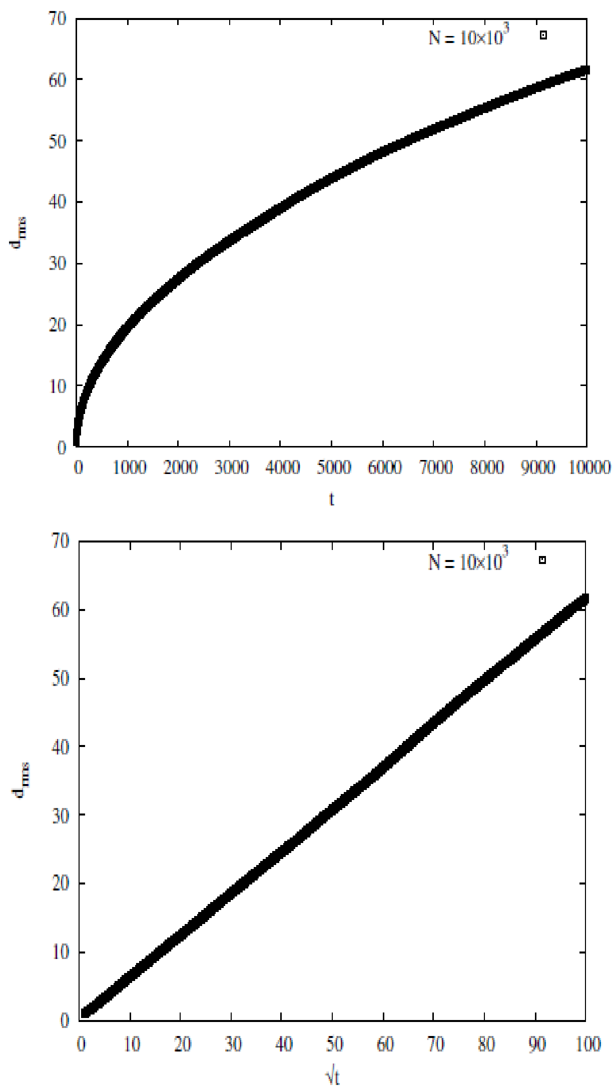

Fig.2: d_{rms} *Vs. t* (upper curve) and d_{rms} *Vs.* $\sqrt{t}$ curves (lower curve).

For *N = 10,000*, d_{rms} *Vs. t* curve behaves same with the theoretical approach and shows the monotonically increase in displacement. Consequently, on the logarithmic scale, it is a straight line with slope *1/2* as shown in Fig.2. Hence, d_{rms} *Vs.* $\sqrt{t}$ curve is a straight line as well. And here the slope is *0.62*.

4.2 Clock Randomness

Here, in Fig.3, *lnT Vs. N* is plotted for a single realization. We have shown the steps, *N* in x-axis needed for each event to return to origin. And in *y*-axis the respective time, *T*. The *y*-axis is chosen to be in log scale in order to compensate the extreme values of return time, *T*. From Fig.3, it is observed that for the maximum of the events, the value of *lnT* varies from 5-8 i.e. the number of steps varies from *148.413159103* to *2980.957987042*. Consequently, there are few steps where the value of *lnT* crosses *20* i.e. they needed steps over *485165195.409790278* to return to origin. These characteristics surely exhibit power law nature.

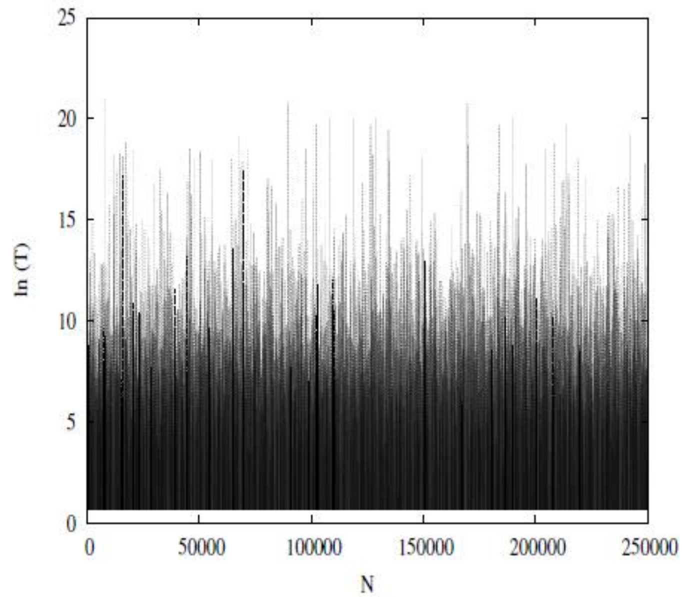

Fig 3. *ln(t) Vs N* curve

4.3 The first five events to return to origin

Though the program is written for a trail where the gambler goes to be bankrupted *2,50,000* times and it is repeated *25* times, we record here only first five events of going broke. The Table-2 contents the distance of the random walker from the origin after each time steps. And in Fig.4, we have plotted walking *distance Vs. time* for first *24* steps while the random walker returned to origin for five times by five different events.

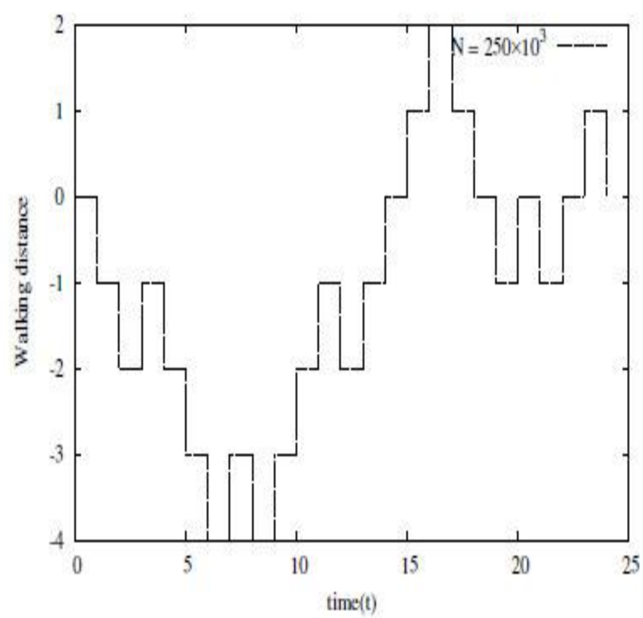

Fig. 4: The events of the random walker returning to origin for first five times

TABLE 2
TABLE FOR THE STEPS OF RETURNING TO ORIGIN

Time steps	Walking distance
0	0
1	-1
2	-2
3	-1
4	-2
5	-3
6	-4
7	-3
8	-4
9	-3
10	-2
11	-1
12	-2
13	-1
14	0
15	1
16	2
17	1
18	0
19	-1
20	0
21	-1
22	0
23	1
24	0

4.4 SCALE-FREE FIRST RETURN PROBABILITY (FRP)

We have performed the Monte Carlo simulation and collected data for $N = 2,50,000$. Our first goal is to quantify the degree of randomness in data having in mind the question: what is the first return probability, $F(t)$ that the walker returns to zero in time t? We have calculated the probabilities for each interval. Then the result is normalized and plotted $F(t)$ *Vs.* t curves for a single event in Fig.5, on normal scale and Fig.6 on logarithmic scale. Hence, on logarithmic scale, Fig.6 shows a remarkable feature that it exhibits power law tail with a head on the left and a long tail on the right. The left portion represents the few that dominates and the right portion to the majority with the long tail. In our problem, it indicates that very few events take infinitely large number of steps to go broke for a gambler. Consequently, maximum of the events need a few or few more steps. From the Fig.5, we have observed that for about *92%* events, it takes few steps to go broke. Or, we may say, the random walker returns to the origin with a very small number of steps in *92%* cases.

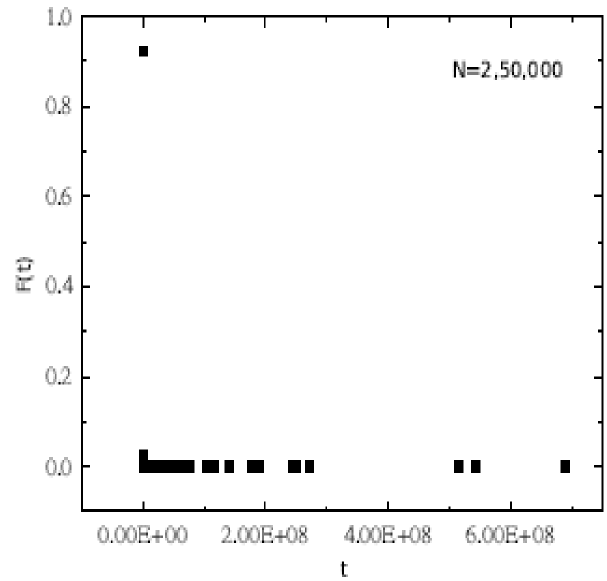

Fig.5: First return probability distribution on normal scale.

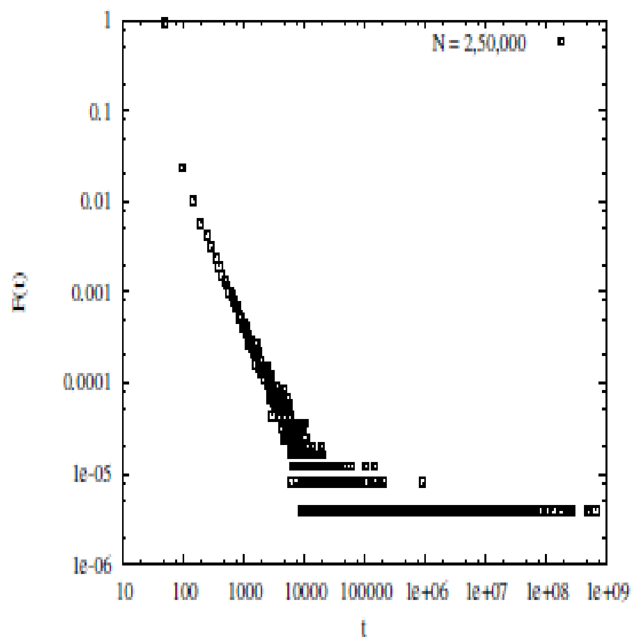

Fig.6: First return probability distribution on logarithmic scale.

In order to observe the distribution of the probability, we separate the Fig.5, which is plotted on normal scale for a single realization, in two portions to have a closer look in Fig.7. Here, we observe that the left portion (head in Fig.7(a)) has an L-shaped distribution. Consequently, on the right portion (tail in Fig.7(b)), we observe that there are some points of scarce density. On the logarithmic scale, these points show the fat tail distribution. This fat tail distribution reminds us about the power law nature of wealth condensation.

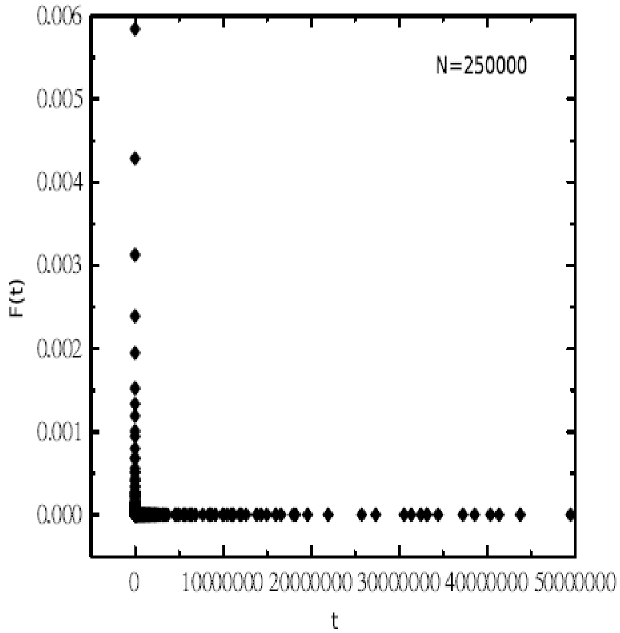

Fig. 7(a): First Return Probability Distribution for head

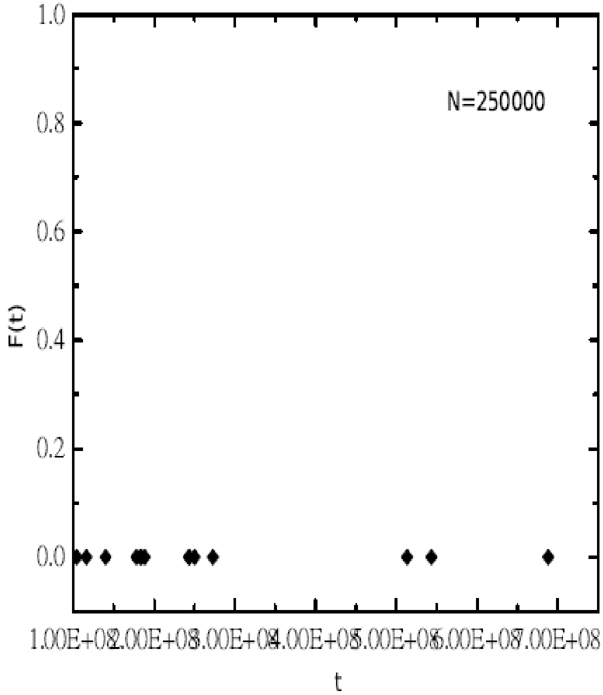

Fig. 7(b): First Return Probability Distribution for tail.

J. P. Boucheaud & M. Mezard in their wealth condensation model stated that about *80%* of the total wealth is owned by the riches who are about *20%* of the world population. And *20%* of the wealth is distributed among the *80%* of the world population. Similar phenomena have been occurred in the most social, biological, and technological networks. Again, data for *25* individual realizations are superimposed in Fig.8 and Fig.9 on normal and logarithmic scale respectively.

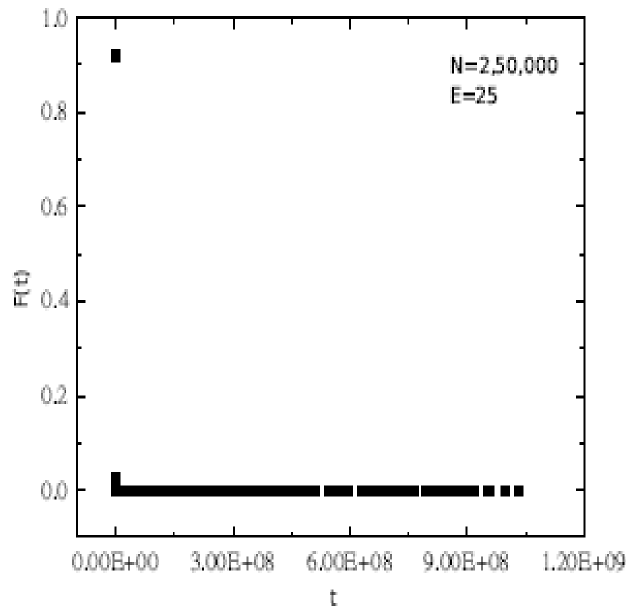

Fig.8: First Return Probability Distribution: data for *25* independent realization are superimposed (normal scale)

Here, we are very familiar with the first return probability distribution curves shown both in Fig.6 and Fig.9. In the logarithmic scale, they have the same fat-tailed distribution like the degree distribution curve of the complex network. The scale free degree distribution of the BA model also exhibits this power law nature giving the idea of some nodes maintaining a large number of connections. And again maximum of the nodes have just one or few connections or links. Consequently, it will be an easy job to target the high degree nodes, called 'hubs', if someone try to affect a network system. We have obtained the theoretical value of the exponent of first return probability as *-3/2* or *-1.5*. And in our work it is *-1.5*, which finely matches with the theoretical value of first return probability.

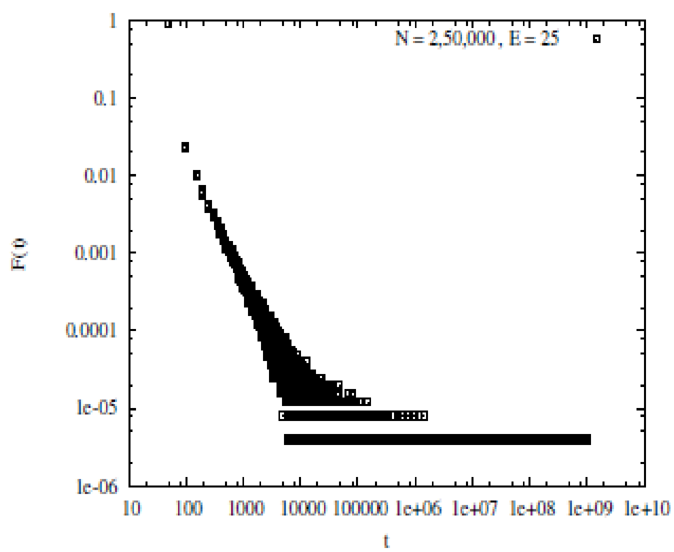

Fig.9: First Return Probability Distribution: data for *25* independent realization are superimposed (logarithmic scale)

4.5 Cumulative Distribution

In probability theory and statistics, the cumulative distribution completely describes the probability distribution of a real-valued random variable. It is also used to specify the distribution of multivariate random variables

The cumulative distribution is mathematically described as

$$F(t' > t) = \sum_{t}^{t_{max}} F(t)$$

Hence, $$\frac{\partial F(t' > t)}{\partial t} = F(t)$$

Where, $F(t' > t)$ is the cumulative distribution and $F(t)$ is the probability distribution.

If $$F(t' > t) \sim t^{-\gamma}$$

Then, $$F(t) \sim -\gamma t^{-(\gamma+1)}$$

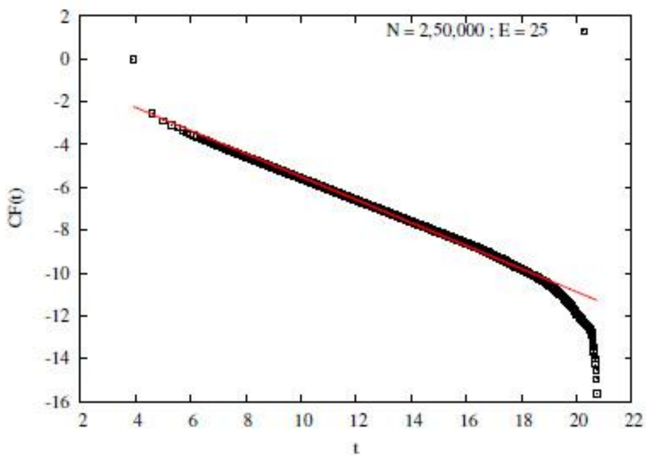

Fig.10: Cumulative distribution curve

Here the simulation has been performed with ensemble size, *E = 25* and in each event the random walker has returned to the origin *2,50,000* times which is a pretty long record. Hence the cumulative distribution is plotted in Fig.10. The red line here is drawn to guide our eyes for a straight line. Again, for the cumulative distribution, the theoretical value of the exponent is *-0.50*. And from Fig.10, we obtain the slope-*0.51*.

5 Conclusions

In this paper, we set a problem of the random walk idea in the context of gambling. There is well established theory for Gambler's Ruin problem. However, to the best of our knowledge there do not exist experimental or numerical verification. We for the first time verified it numerically and to our surprise we found fat tail which was not predicted by the theory. Our goal is to observe the randomness i.e. the probability of first return time. Through this work, we have tried to study the power law behavior of the scale-free first return probability of a gambler's ruin problem which follows the power-law, $F(t) \sim t^{-3/2}$. Consequently, it is highly instructive to note that this first return probability distribution curve is a recollection of the fat tailed degree distribution of scale free complex network predicted by A.-L. Barabási and R. Albert. In addition to this, we have also studied some other characteristics such as cumulative distribution and *rms displacement Vs. time*. For cumulative distribution, from the *CF(t) Vs. t* graph, we obtain the slope of the value of *-0.51*. And d_{rms} *Vs. t* curve shows the monotonically increase in displacement and as a result, on the logarithmic scale, it is a straight line with slope *1/2*. Again, d_{rms} *Vs.* $\sqrt{t}$ curve is a straight line as well with the slope *0.62*. For rms displacement the simulation has been performed for *N = 10,000*. For larger *N*, the curves become smoother. Hence, *lnT* Vs. *N* is plotted for a single realization to show the randomness in data. We also plot walking *distance vs. time* for first *24* steps while the random walker returns to origin for five times by five different events. We intended to compare the theoretical and experimental results. And we have obtained some considerable results from our work as well.

Acknowledgments

The authors would like to convey their sincere gratefulness to Professor Dr. M. Kamrul Hassan, Department of Physics, University of Dhaka, for supervision of this work.

References

[1] M Haw, Colloidal suspensions, Brownian motion, molecular reality: a short history, J. Phys.: Condens. Matter, vol. 14, pp. 7769-7779, 2002.

[2] Brown, Robert, A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. Phil. Mag., vol. 4, pp. 161-173, 1828.

[3] Smoluchowski, M., Zur kinetischen Theorie der Brownschen Moleku-larbewegung und der Suspensionen, Annalen der Physik 326 (14): 756-780, doi:10.1002/andp.19063261405, 1906.

[4] J. Perrin, Mouvement brownien et ralit molculaire. Ann. Chim. Phys.8ime srie 185 , 114 (1909). See also Perrin's book "Les Atomes",1914.

[5] Pearson, K. , The problem of the Random Walk. Nature. Vol. 72, pp. 294, 1905.

[6] E. Shoesmith, Hygens' solution to the gambler's ruin problem, Historia Math, vol. 13(2), pp. 157-164, 1986.

[7] Whitmore, G. A., First passage time models for duration data regression structures and competing risks, The Statistician, vol. 35, pp. 207-219, 1986.

[8] P. Erdös and A. Rènyi, Publications Mathematics, vol. 6, pp. 290,1959; Publ. Math. Inst. Hung. Acad. Sci. vol. 5, pp. 17, 1960.

[9] B. Bollobas, Modern Graph Theory, Springer-Verlag, New York, 1998.

[10] Erdoös, P. and Rènyi, A., On the Evolution of Random Graphs, Publications of the Mathematical Institute of the Hungarian Academy of Sciences, vol. 5, pp. 17-61, 1960.

[11] P. Erdös and A. Rènyi, On random graphs, Publications Mathematicae, vol 6, pp. 290-297, 1959.

[12] P. Erdös and A. Rènyi, On the strength of connectedness of a random graph, Acta Mathematica Scientia Hungary, vol. 12, pp. 261-267, 1961.

[13] S. Milgram, The small-world problem, Psychology Today, vol. 2, pp. 6067, 1967.

[14] M. Kochen, The Small World, 1989.

[15] R. Albert, H.jeong, and A.-L. Barab´asi, Diameter of the world-wide
web, Nature, vol. 401, pp. 130-131, Sept. 1999.

[16] M. Faloutsos, P. Faloutsos, and C. Faloutsos, On power-law relationship of the internet topology, Computer Communications Review, vol. 29, pp. 251-262, 1999.

[17] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, The large-scale organization of metabolic networks, Nature, vol. 407, pp. 651-654, 2000.

[18] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small worlds'
networks, Nature, vol. 393, pp. 440 - 442, 1998.

[19] Albert R. and Barabási A.-L., Statistical mechanics of complex net-works. Review of Modern Physics, vol. 74, pp. 47-94, 2002.

[20] Jeong, H., Albert R. and Barabasi, A.-L., ''Diameter of the World Wide Web", Nature, vol. 401, pp. 130, 1999.

[21] Albert, R., Jeong, H., and Barabási, A.-L., Attack and error tolerance of complex networks, Nature, vol. 406, pp. 378, 2000.

[22] Reka Albert and Albert -Laszlo Barabási, Statistical Mechanics of Complex Networks, arxiv:cond-mat/0106096v1[cond-mat.stat-mech] 6 Jun, 2001.

[23] Udny Yule, A Mathematical Theory of Evolution Based on the Con-
clusions of Dr. J. C. Willis, F.R.S., Journal of the Royal Statistical Society, vol. 88, pp. 433-436, 1925

[24] D. J. de S. Price, Networks of scientific papers , Science, vol. 149, pp. 510-515, 1965.

[25] H. A. Simon, On a class of skew distributions functions, Biometrika, vol. 42, pp. 425-440, 1955.