GAMBLER'S RUIN: THE LIMITING DISTRIBUTION OF POINTS VISITED EXACTLY ONCE OF A SIMPLE RANDOM WALK UP TO TIME OF EXIT AND THE INFLUENCER VOTER MODEL

by

Conner Hatton

A Thesis Submitted to the Faculty of the

DEPARTMENT OF MATHEMATICS

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2022

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Master's Committee, we certify that we have read the thesis prepared by: Conner Hatton
titled:

and recommend that it be accepted as fulfilling the thesis requirement for the Master's Degree.

_______________ Date: Jun 22, 2022
Sunder Sethuraman

_______________ Date: Jun 22, 2022
Thomas Kennedy

_______________ Date: Jun 22, 2022
Shankar Venkataramani

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to the Graduate College.

I hereby certify that I have read this thesis prepared under my direction and recommend that it be accepted as fulfilling the Master's requirement.

_______________ Date: Jun 22, 2022
Sunder Sethuraman
Thesis Committee Chair
Mathematics

Acknowledgments

First and foremost, I would like to express my gratitude to my advisor, Professor Sunder Sethuraman, who has been generous with his time and has provided copious amounts of feedback, support, and knowledge in preparation of this thesis. It has been a great pleasure working under his guidance.

I would also like to thank the committee members Professor Tom Kennedy and Professor Shankar Venkataramani for their time in reviewing this thesis. I would like further thank Sunder and Shankar for including me in the high school Modeling and Epidemics summer program in 2021 which was funded by the grant ARO-W911NF-18-1-0311. It was in this program that high school students Karl and Ted first simulated a variant of the influencer voter model as a capstone project.

Table of Contents

List of Figures **5**

Abstract **6**

Chapter 1. Introduction **7**
1.0.1. Scaling Limits of the Range and Multiple Range 7
1.0.2. Voter Model and the Influencer Voter Model 9

Chapter 2. The Limiting Distribution of Points Visited Exactly Once of a Simple Random Walk up to Time of Exit **14**
2.1. Introduction 14
2.2. The k^{th} Moment of R_N^1 15
2.2.1. Order of the Moments 16
2.2.2. Four Cases 16
2.3. Computing the Four Cases 17
2.3.1. Symmetry in the Cases 19
2.3.2. Cases 1 and 4 19
2.3.3. Bounding Cases 2 and 3 26
2.3.4. Concluding proof of Lemma 2.2.0.1 27
2.4. Proof of Proposition 2.1.1 28
2.5. Numerical Results 28
2.6. Further Considerations 31

Chapter 3. Influencer Voter Model **32**
3.1. Introduction 32
3.2. System Dynamics 33
3.3. Scaled Process and the Fluid Limit 34
3.4. Exit Probabilities 41
3.5. Time to Consensus 43
3.5.1. Bound on Time to Consensus 43
3.5.2. A Heuristic for Time to Consensus 44
3.6. Numerical Results 46
3.7. Further Considerations 47

References **49**

List of Figures

Figure 2.1. Illustration of the one-dimensional random walk with $N = 10$, $\alpha = \frac{1}{2}$. . . . 14
Figure 2.2. (a) In blue the histogram of points visited exactly once with 1000 simulations on the domain $[0, 10]$ and starting position of 5. In red density of Exp(2) scaled by the number of simulations.. (b) Empirical CDF of data in (a) plotted with the CDF of Exp(2). . . . 29
Figure 2.3. (a) In blue the histogram of points visited exactly once with 1000 simulations on the domain $[0, 100]$ and starting position of 50. In red density of Exp(2) scaled by the number of simulations.. (b) Empirical CDF of data in (a) plotted with the CDF of Exp(2). . . . 30
Figure 2.4. (a) In blue the histogram of points visited exactly once with 1000 simulations on the domain $[0, 1000]$ and starting position of 500. In red density of Exp(2) scaled by the number of simulations.. (b) Empirical CDF of data in (a) plotted with the CDF of Exp(2). . . . 30

Figure 3.1. Illustration of the Influencer Voter Model as a two-dimensional nearest neighbor random walk. Current state is highlighted in blue and absorbing states highlighted in red. . . . 34
Figure 3.2. (a), The two-dimensional sample path in terms of the densities ρ_R, ρ_I.(b) each sample path in terms of the densities ρ_R, ρ_I. Both plots are made continuous by linear interpolation. In this simulation we have $R = 700, I = 300, \rho_r = .57, \rho_I = .83$ and $p = 0.8$. Consensus was reached on (R, I) in around 5000 steps. . . . 35
Figure 3.3. Sample Path of Influencer Voter Model in blue, solution to system of ODEs given in (3.3.8) in yellow, and starting position given by the red asterisks. In this simulation we have $R = 700, I = 300, \rho_r = .57, \rho_I = .83$ and $p = 0.8$. . . . 40
Figure 3.4. Convergence time in terms of the number of nodes N. We have chosen the following parameters $p = .75, I = .3N, R = .7N, \rho_I(0) = 2/3$, and $\rho_R(0) = 5/7$. . . . 46
Figure 3.5. Numerically solved fluid limits for the ternary voter model. The three paths here in blue, yellow, and red exhibit the three general paths. For blue the initial conditions are $(u(0), v(0)) = (0.51, 0.4)$, for yellow $(u(0), v(0)) = (0.9, 0.9)$ and for red $(u(0), v(0)) = (0.3, 0.6)$. . . . 47

Abstract

In this thesis we analyze two stochastic processes. We first consider the simple symmetric random walk on the domain $[0, N] \subset \mathbb{Z}$ with a particular interest in the points visited exactly once by the random walk up to time of exit, denoted by R^1_N. Specifically, we are interested in determine a scaling limit of R^1_N as $N \uparrow \infty$. Making use of the simple geometry, we give a combinatorial argument and use the classic Gambler's ruin identities of the symmetric random walk to determine the moments of R^1_N. We then find that $R^1_N / \log(N)$ converges weakly to the exponential distribution Exp(2).

The second process we consider is a variant of the classic voter model on the complete graph with N nodes, which we call the influencer model. By allowing a subset of nodes, called the influencers, to be more likely to have their opinion adopted by a neighbor we generalize the classical voter model. We establish a weak law of large numbers via the fluid limit for the influencer model and find that the average trajectory of the process is deterministic and is the solution to a set of ordinary differential equations. Let $\rho_R(t)$ be the density of voter of opinion 1 in the regular nodes at time t and let $\rho_I(t)$ be the density of voters of opinion 1 in the influencer nodes at time t. We then determine the probability and time to consensus using martingale methods with the martingale $Y(t) = p\rho_I(t) + q\rho_R(t)$. We derive that the probability of reaching consensus on 1 is equal to $p\rho_I(0) + q\rho_R(0)$. For the expected time to consensus we arrive at a simple upper bound and then we also give a heuristic argument based on the results of the fluid limit section that gives an exact expectation. For our upper bound derived from using the martingale , we get that the expected time to consensus $E[\tau] = O(N)$ with upper bound

$$E[\tau] \leq \frac{N}{C}\mathbb{E}[Y_0^2].$$

In the heuristic section we find that

$$\mathbb{E}[\tau] = \left[\frac{4RI}{N}\left((1-\omega)\log\left(\frac{1}{1-\omega}\right) + \omega\log\left(\frac{1}{\omega}\right)\right)\right]$$

where $\omega = p\rho_R(0) + q\rho_I(0)$ and R, I are equal to the number of regular and influencer nodes, respectively.

Chapter 1

Introduction

In the classical Gambler's ruin problem we have a random walk and are interested in both the probability and expected time to either bust or make our fortune. Stated more generally, we are interested in the the statistics and behavior of random processes up to time of exit. With this perspective in mind, in this thesis we consider two different stochastic processes. The first is the simple random walk on the domain $[0, N]$ where we are interested in determining the limiting distribution of the points visited exactly once as $N \uparrow \infty$. The second process is a variant of the voter model on the complete graph which we call the influencer voter model. Here, we are interested in the macroscopic behavior of the model, which we determine by the fluid limit, and the probabilities of consensus and expected time to consensus. In this chapter we briefly introduce and discuss these processes that will be studied in the latter two chapters.

1.0.1 Scaling Limits of the Range and Multiple Range

For random walks, two classical objects of study are the range $\mathcal{R}_n$ and multiple range $\mathcal{R}_n^{(p)}$ where $\mathcal{R}_n$ is the set of sites that have been visited by a random walk at time n and $\mathcal{R}_n^{(p)}$ is the set of sites that have been visited exactly p times at time n. Both the range and multiple range and their statistics have been studied in various settings and for planar random walks see Bass et al., [3] for a historical account. Applications of the range and multiple range vary from astrophysics [12] to modeling disease transmission [18, 25]. For instance, assuming that an infection is modeled by a random walk, the distribution of the multiple range R_n^p gives insight into the probabilities of a person contracting the disease p times.

An interesting restriction to the random walks is considered by Athreya et al., [1] where, motivated by a stochastic version of the "locker" problem, they consider the range R_N, or the set of points visited up to time of exit from the domain $[0, \ldots, N] \subset \mathbb{Z}$. In particular, they derive scaling limits as $N \uparrow \infty$ and found $\frac{R_N}{N} \Rightarrow F_\alpha$ where F_α has density f_α where

$$f_\alpha(\beta) = \begin{cases} \frac{\alpha \wedge (1-\alpha)}{\beta^2} & \text{for } \alpha \wedge (1-\alpha) < \beta < \alpha \vee (1-\alpha) \\ \frac{1}{\beta^2} & \text{for } \alpha \vee (1-\alpha) \leq \beta \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Recently, the same problem was considered more generally by Doehrman et al., [9]. They derived scaling limits for both the range R_N and multiple range $R_N^{(p)}$ on some domain $D_N \subset \mathbb{R}^d$, where D is subject to some regularity conditions, $D_N = ND$, and $d \geq 2$. Let $\tau_{a,D}$ be the time of exit from D of a d-dimensional Brownian motion starting at $a \in D$, then they found that for $d = 2$ that $\log(N)R_N/N^2$ converges weakly to $\pi\tau_{a,D}$ and $\log(N)^2 R_N^{(p)}/N^2$ converges weakly to $2\pi^2\tau_{a,D}$. For $d \geq 3$ they found that R_N/N^2 converges weakly to $d/2(1-p_0)\tau_{a,D}$ and $R_N^{(p)}/N^2$ converges weakly to $d/2(1-p_0)^2(p_0)^{p-1}\tau_{a,D}$ where $p_0 < 1$ is the probability that a simple random walk on $\mathbb{Z}^d$ starting at the origin returns to the origin. For $d = 1$, however, the limiting distributions for the multiple range seems to remain open.

In chapter 2 we make progress on determining scaling limits for the multiple range by determining the scaling factor and limiting distribution for R_N^1, or the points visited exactly once on the domain $[0, N]$. From the simplicity of the geometry in $d = 1$, we are able to make a combinatorial argument in which we evaluate the moments of $\mathcal{R}_N^1$ by making use of the Gambler's ruin identities. The main proposition of this chapter says that $\log(N)^{-1}$ is an appropriate scaling factor for R_N^1 and that $R_N^1/\log(N)$ converges weakly to the exponential distribution Exp(2). Interestingly, in contrast to higher dimensions, the limiting distribution is not dependent on the starting position.

In Section 2.2 we give a description of the moments of R_N^1 in terms of the intersection of indicator functions. Let $A_{x_i} = \{T_{x_i} < \tau_N, T_{x_i}^{(2)} > \tau_N\}$ be the event that the point x_i is visited exactly once at the time of exit τ_N and where T_{x_1},$T_{x_i}^{(2)}$ are the first and second time x_i is visited,respectively. We find that we can write the expectation of the k^{th} moment of R_N^1 as

$$\mathbb{E}_{\alpha N}[(R_N^1)^k] = \mathbb{E}_{\alpha N}\left[\sum_{x_1,x_2,\ldots,x_k} 1_{A_{x_1}} 1_{A_{x_2}} \ldots 1_{A_{x_k}}\right].$$

We then prove lemma 2.2.0.1, which states that For $1 \leq j \leq k$ we have

$$\mathbb{E}_{\alpha N}\left[\sum_{x_1,x_2,\ldots,x_j} 1_{A_{x_1}} 1_{A_{x_2}} \ldots 1_{A_{x_j}}\right] = \frac{j!}{2^k}\log(N)^j + o(\log(N)^j)$$

The proof of lemma 2.2.0.1 is the bulk of the chapter and consists of four additional lemmas. We find that by ordering the x_i, counting permutations, and using Gambler's ruin identities on the A_{x_i} that there are only the four following cases:

Case 1:

$$\frac{1}{2^j}\sum_{x_1=1}^{\alpha N-j}\sum_{x_2=x_1+1}^{\alpha N-j+1}\cdots\sum_{x_j=x_{j-1}+1}^{\alpha N-1}\frac{N-\alpha N}{N-x_j}\cdot\frac{1}{x_j-x_{j-1}}\cdot\frac{1}{x_{j-1}-x_{j-2}}\cdots\frac{1}{x_2-x_1}\frac{1}{x_1}$$

Case 2:

$$\frac{1}{2^j}\sum_{x_1=1}^{\alpha N-j+1}\sum_{x_2=x_1+1}^{\alpha N-j+2}\cdots\sum_{x_j=\alpha N+1}^{N-1}\frac{\alpha N-x_{j-1}}{x_j-x_{j-1}}\cdot\frac{1}{x_j-x_{j-1}}\cdot\frac{1}{x_{j-1}-x_{j-2}}\cdots\frac{1}{x_2-x_1}\frac{1}{x_1}$$

Case 3:

$$\frac{1}{2^j}\sum_{x_1=1}^{\alpha N-1}\sum_{x_2=\alpha N+1}^{x_3-1}\cdots\sum_{x_j=\alpha N+j-1}^{N-1}\frac{x_2-\alpha N}{x_2-x_1}\cdot\frac{1}{x_2-x_1}\cdot\frac{1}{x_3-x_2}\cdots\frac{1}{x_j-x_{j-1}}\frac{1}{N-x_j}$$

Case 4:

$$\frac{1}{2^j}\sum_{x_1=\alpha N+1}^{x_2-1}\sum_{x_2=\alpha N+2}^{x_3-1}\cdots\sum_{x_j=\alpha N+j}^{N-1}\frac{\alpha N}{N-x_j}\cdot\frac{1}{x_j-x_{j-1}}\cdot\frac{1}{x_{j-1}-x_{j-2}}\cdots\frac{1}{x_2-x_1}\cdot\frac{1}{x_1}$$

In Section 2.3 we compute the four cases. Making use of a symmetry in the order of the cases, however, we find it is sufficient to compute only two cases. In particular, we find that case 1 and 4 are of the same order, but have multiplicative constants $1-\alpha$ and α, respectively. The same is true from cases 2 and 3, however, with an additional lemma we show that these two cases are of lower order and vanish in the limit. Thus it suffices to compute case 1.

Since we only care about the order, the computations for case 1 are made relatively straight forward by making repeated use of the Euler-Macluarin formula, partial fraction decomposition, and disregarding terms of lower order. However, care must be taken, for example, when the term $\log(\alpha N - x_{j-1} - 1)$ arises as a multiplicative factor in the sum above after using Euler-Macluarin. However, by truncating the sums to get a lower estimate and taking limits, we get the desired conclusion in lemma 2.2.0.1. The ingredients for our main proposition are in place and in Section 2.4 we give a brief proof of our main proposition. We then consider some numerical results in 2.5 and conclude chapter 2 with some remarks on future work.

1.0.2 Voter Model and the Influencer Voter Model

The voter model is one of the basic examples of an interacting particle system and has been studied extensively [10, 17, 24]. Each node on a graph is in one of two states $\{0, 1\}$, where we take the interpretation of the states as being opposing opinions. The voter model dynamics are such that at a given time of update, a node and one of its neighbors are randomly selected and the first node updates its opinion state to that of its neighbor's. This process continues indefinitely or until consensus is reached, that is, when all nodes are in the state 0 or 1. Some of the first seminal work by

Clifford [6], Holly and Liggett [13] on the voter model was on the infinite lattice $\mathbb{Z}^d$ with an interest in the structure and existence of invariant measures. It was found that the voter model on $\mathbb{Z}^d$ has two extremal invariant measures if $d \leq 2$ at point mass $0, 1$, representing consensus, but a one-parameter family of extremal invariant measures if $d \geq 3$.

In this thesis, however, we focus on finite graphs and specifically the complete graph. For finite graphs, since the voter model dynamics in non-consensus states result in an irreducible Markov chain, we almost surely reach consensus on 0 or 1, in a finite amount of time. In this setting, we are primarily interested in the probability of reaching consensus given the initial states of the voter model and the expected time to consensus. For the continuous time voter model on a complete graph with N nodes, where update times are exponentially distributed with mean time N^{-1}, we have the following results for the probability and expected time to consensus [22]. Let ρ be the density of voter of opinion 1. Then the probability of reaching consensus on 1 is equal to ρ and thus the probability of reaching consensus on 0 is $1-\rho$. Let $T_N(\rho)$ be the expected time to reach consensus where ρ is the initial density of voters of opinion 1. Then $T(N) = O(N)$ and more specifically,

$$T_N(\rho) = N\left[(1-\rho)\log\left(\frac{1}{1-\rho}\right) + \rho\log\left(\frac{1}{\rho}\right)\right].$$

In Chapter 3 we consider a novel continuous time weighted voter model on the complete graph which we call the influencer voter model. The influencer voter model, motivated by networks within the sphere of social media, accounts for certain nodes in the network having a disproportionate influence on the opinions of others. Voter models that include influencer type behavior have been considered by Bhat and Redner in [4] and a weighted voter model in Baronchelli et al., in [2]. However, the models are more complicated than the model given here. Here, the model is a simple generalization of the voter model in which the influencer voter model dynamics are determined by two fixed subsets of nodes which we call the influencers and regular nodes. The influencer and regular nodes are distinguished by the probabilities of a neighbor adopting their opinion. That is, at a time of update, a uniformly random chosen node will select a neighbor from the influencer set with probability p and a neighbor from the regular set with probability $q = 1-p$ and then adopt that neighbors opinion.[1] We can then think of the probability $0 \leq p \leq 1$ as the weight that influencers have on the opinions of the regular vertices. If p is greater than the density of influencer nodes, we expect that the influencers will have a larger effect on the

[1] Note that if the number of regular nodes is equal to the number of influencer nodes and $p = q$, you get the same formulation as the voter model on a complete graph, so that the influencer model is a generalization of the voter model.

probabilities of the system reaching consensus than the regular nodes.

In Section 3.1 we explicitly construct the continuous time influencer voter model and give the dynamics of the system. We find that it suffices to consider the process $(U(t), U^*(t))$, where $U(t)$(resp. $U^*(t)$) is the number of regular (resp. influencer) voters of opinion 1 at time t. The Markov process $(U(t), U^*(t))$ is then a two-dimensional random walk on a subset of the lattice $\mathbb{Z}^2$ with jump sizes of size 1, mean jump time N^{-1}, and absorbing states $(0,0), (R, I)$, where R is the number of regular voters, I is the number of influencer voters, and $R + I = N$.

In Section 3.3 we define a scaled process of the influencer voter model where we scale the process $(U(t), U^*(t))$. The scaling consists of speeding up the jump rate by scaling the number of voters N and decreasing the size of the jumps by N^{-1}. As $N \uparrow \infty$, we arrive at the Fluid Limit [8,11,16,21,26], which is given by a deterministic system of ordinary differential equations. To do this, we use Dynkin's formula to write the process $(U(t), U^*(t))$ as

$$f(U(t), U^*(t)) - f(U(0), U^*(0)) = \int_0^t \mathcal{L}f(U(t), U^*(t))ds + M^f(t)$$

where $\mathcal{L}$ is the generator and $M^f(t)$ is a martingale with respect to the natural filtration. Choosing f to be one the coordinate functions $f_1(x, y) = x$ and $f_2(x, y) = y$ and scaling by N^{-1} we get the scaled coordinates $u_N(t) = U(t)/N$ and $u^*_N(t) = U^*(t)/N$ as

$$u_N(t) = u_N(0) + N^{-1} \int_0^t \mathcal{L}U(s)ds + N^{-1}M^{f_1}(t)$$

and

$$u^*_N(t) = u^*_N(0) + N^{-1} \int_0^t \mathcal{L}U^*(s)ds + N^{-1}M^{f_1}(t).$$

To show that the scaled process converges to the fluid limit $(u(t), u^*(t))$ it suffices to show weak convergence of scaled coordinate sample paths to the solution of the ODE in $C[0, 1]$. To show convergence of the sample paths, we show that the families $\{u_N(t)\}, \{u^*_N(t)\}$ are tight and that their time derivatives converge to the desired system of ordinary differential equations. This result is summarized by proposition 3.3.1. We find that the scaled Markov Process $(u_N(t), u^*_N(t))$ converges weakly on any compact time interval to the continuous, Markov Process $(u(t), u^*(t))$ given by the system of ordinary differential equations

$$\frac{du(t)}{dt} = p \cdot r \left(\frac{u^*(t)}{i} - \frac{u(t)}{r} \right)$$
$$\frac{du^*(t)}{dt} = q \cdot i \left(\frac{u(t)}{r} - \frac{u^*(t)}{i} \right)$$

and solution, with initial conditions $u(0)$ and $u^*(0)$, is given by

$$u(t) = r\left(p\frac{u^*(0)}{i} + q\frac{u(0)}{r}\right) + r\cdot p\left(\frac{u(0)}{r} - \frac{u^*(0)}{i}\right)e^{-t}$$
$$u^*(t) = i\left(p\frac{u^*(0)}{i} + q\frac{u(0)}{r}\right) - i\cdot q\left(\frac{u(0)}{r} - \frac{u^*(0)}{i}\right)e^{-t}$$

where $r = \lim_{n\to\infty} r_N$ and $i = \lim_{n\to\infty} i_N$.

This establishes a weak law of large numbers result for the influencer voter model. From the solution to the set of ordinary differential equations we can then analyze the behavior of the influencer voter model. In particular, letting $\rho_I(t) = u^*(t)/i, \rho_R(t) = u^*(t)/r$ be the subgraph densities, we find that the average trajectory tends towards equal subgraph densities at the point $((p\rho_I(0) + q\rho_R(0), (p\rho_I(0) + q\rho_R(0))$. After reaching the point of equal subgraph densities, however, the dynamics of the trajectories characteristically change as the path then diffusively fluctuates along the diagonal $\rho_I = \rho_R$ until consensus is reached. This analysis leads us to the martingale $Y_t = p\rho_I(0) + q\rho_R(0)$ for which we use in section 3.4.1 to determine the exit probabilities and in section 3.5 to give an upper bound on the expected time to consensus using Doob's optional stopping theorem. We also give a heuristic given by Redner and Sood [22], that relies on the analysis of the fluid limit and this gives an exact expected time to consensus.
The results for the probability and expected time to consensus for the influencer voter model is as follows. The probability of reaching consensus on 1 is equal to $p\rho_I(0) + q\rho_R(0)$ and the probability of reaching consensus on 0 is $1 - p\rho_I(0) + q\rho_R(0)$ where $\rho_R(0), \rho_I(0)$ are the initial densities of voters of opinion 1 in the regular and influencer nodes, respectively. For the expected time to consensus, we have an upper bound derived from using the martingale $Y(t) = p\rho_I(t) + q\rho_R(t)$. We get that the expected time to consensus $E[\tau] = O(N)$ with upper bound

$$E[\tau] \leq \frac{N}{C}\mathbb{E}[Y_0^2].$$

In the general heuristic section we find that

$$\mathbb{E}[\tau] = \left[\frac{4RI}{N}\left((1-\omega)\log\left(\frac{1}{1-\omega}\right) + \omega\log\left(\frac{1}{\omega}\right)\right)\right]$$

where $\omega = \frac{\rho_R+\rho_I}{2}$ and R, I are equal to the number of regular and influencer nodes, respectively.

In Section 3.6 we verify our expected time to consensus results numerically. We then conclude the chapter with some further considerations section 3.7. Specifically, we discuss the addition of an undecided opinion e into the states set, so that $S = \{0, 1, e\}$. The addition of the undecided state for the voter model on the complete graph was considered by Perron et al, [19]. They found that by including undecided voter is desirable as they find that the probability of reaching consensus on the non-initial majority decays exponentially with the number of nodes N. Furthermore, they found that the convergence time is logarithmic in N. We briefly discuss the applications of including an undecided voter to the influencer voter model.

Chapter 2

The Limiting Distribution of Points Visited Exactly Once of a Simple Random Walk up to Time of Exit

2.1 Introduction

The goal of this chapter is to find the distributional limit of the number of points visited exactly once of a one-dimensional simple random walk up to time of exit on the domain $[0, N]$ as $N \uparrow \infty$. We first introduce the basic definitions and some notation.

Recall that a random walk is a stochastic process. Let $\{\xi_i : i \geq 1\}$ be a sequence of I.I.D random variables and

$$X_n = X_0 + \xi_1 +, \ldots, +\xi_n \ \{X_n : n \geq 1\}$$

where X_0 is the starting position. A one-dimensional simple random walk is a symmetric random walk so that

$$\{X_n : n \geq 0\} \text{ where } \mathbb{P}(X_{n+1} = x \pm 1 | X_n = x) = \frac{1}{2.}$$

We restrict the random walks to the domain $[0, N] = \{0, 1, 2, \ldots, N\}$ with starting point $X_0 = \alpha N \in \{1, 2, \ldots, N-1\}$, so that α is some scaling value such that $0 < \alpha < 1$ and for $\alpha N \notin \mathbb{N}$ we take $\lfloor \alpha N \rfloor$. We also only consider the random walk up to the time of exit τ_N where we define $T_x := \inf\{t : X_t = x\}$ and $\tau_N = \min\{T_0, T_N\}$.

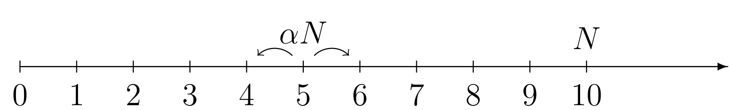


Figure 2.1: Illustration of the one-dimensional random walk with $N = 10$, $\alpha = \frac{1}{2}$.

We call the range $\mathcal{R}_n$ of a random walk the number of distinct points visited in the domain $[0, N]$ at time n. We let $\mathcal{R}_n^1$ be the number of points visited exactly once in the domain $[0, N]$ at time n. Since we are interested in the properties of the random walk at the time of exit, define $R_N = \mathcal{R}_{\tau_N}$ and $R_N^1 = \mathcal{R}_{\tau_N}^1$ so R_N is the number of distinct points visited at the time of exit from the domain $[0, N]$ and R_N^1 is the number of distinct points visited exactly once at the time of exit from the domain $[0, N]$.

With all of the main definitions stated, we now give the following proposition which is the main result for this chapter.

Proposition 2.1.1. *For a simple one dimensional random walk up to time of exit* τ_N *on the domain* $[0, N]$ *the kth moments of the scaled random variable* $R_N^1/\log(N)$ *converges as* $N \uparrow \infty$ *to the moments of the exponential distribution. That is,*

$$\lim_{N\to\infty} E_{\alpha N}\left[\left(\frac{R_N^1}{\log(N)}\right)^k\right] = \frac{k!}{2^k}$$

and therefore

$$\lim_{n\to\infty} \frac{R_N^1}{\log(N)} \sim \mathit{Exp}(2).$$

The rest of this chapter is devoted to proving this proposition. We begin with giving a compact description of the kth moment

2.2 The k^{th} Moment of R_N^1

In this section we will setup and prove all the necessary ingredients to show our main result, proposition 2.1.1. To do so we first consider the kth moment of the random variable R_N^1. We then prove five lemmas that together prove proposition 2.1.1.

Let $A_{x_i} = \{T_{x_i} < \tau_N, T_{x_i}^{(2)} > \tau_N\}$ be the event that the point x_i is visited exactly once. We can then write the expectation of the k^{th} moment of R_N^1 as

$$\mathbb{E}_{\alpha N}[(R_N^1)^k] = \mathbb{E}_{\alpha N}\left[\sum_{x_1,x_2,\ldots,x_k} 1_{A_{x_1}} 1_{A_{x_2}} \ldots 1_{A_{x_k}}\right]. \tag{2.2.1}$$

In (2.2.1) we write the expectation of the kth moment of points visited exactly once as the expectation of the sum of indicators where the points $x_i \neq x_j$ if $i \neq j$ are visited exactly once. The sum on the right hand side of (2.2.1) can be decomposed into sums which are determined by the intersections of the indicators. From counting, we get

$$\sum_{\substack{x_1,x_2,\ldots,x_k \\ \text{distinct}}} 1_{A_{x_1}} 1_{A_{x_2}} \ldots 1_{A_{x_k}} + \sum_{j=2}^{k} \binom{k}{j} \sum_{\substack{x_1,\ldots,x_{k-j+1} \\ \text{distinct}}} 1_{A_{x_1}} \ldots 1_{A_{k-j+1}} \tag{2.2.2}$$

where $x_i \neq x_j$ if $i \neq j$, and the combinatorial factor $\binom{k}{j}$ counts how many x_i are equal to one another out of the k terms.

2.2.1 Order of the Moments

In this section we determine the order of the moments which determines for us the appropriate scaling factor of $\log(N)^{-1}$ as seen in 2.1.1.The following lemma gives a precise statement of this.

Lemma 2.2.0.1. *For $1 \leq j \leq k$ we have*

$$\mathbb{E}_{\alpha N}\left[\sum_{\substack{x_1,x_2,\ldots,x_k\\ \textit{distinct}}} 1_{A_{x_1}} 1_{A_{x_2}} \ldots 1_{A_{x_j}}\right] = \frac{j!}{2^k}\log(N)^j + o(\log(N)^j)$$

Observe that following the proof of 2.2.0.1, the latter terms in (2.2.2) will be $o(\log(N)^k)$, so that we may disregard mentioning these terms moving forward.

2.2.2 Four Cases

To prove the lemma we need to compute

$$\mathbb{E}\left[\sum_{\substack{x_1,x_2,\ldots,x_k\\ \text{distinct}}} 1_{A_{x_1}} 1_{A_{x_2}} \ldots 1_{A_{x_j}}\right].$$

To simplify the computation, we impose the ordering

$$0 < x_1 < x_2 < \cdots < x_{j-1} < x_j$$

and multiply by the number of permutations of the x_i terms, of which there are j terms, to get

$$\sum_{x_1,x_2,\ldots,x_j} 1_{A_{x_1}} 1_{A_{x_2}} \ldots 1_{A_{x_j}} = j! \sum_{x_1<x_2<\cdots<x_j} 1_{A_{x_1}} 1_{A_{x_2}} \ldots 1_{A_{x_j}}.$$

Observe that ordering results in only the following four cases since if αN were in any other positions at least one of the x_i must be visited at least twice.

Case 1: $0 < x_1 < x_2 < \cdots < x_j < \alpha N < N$

Case 2: $0 < x_1 < x_2 < \ldots x_{j-1} < \alpha N < x_j < N$

Case 3: $0 < x_1 < \alpha N < x_2 < \cdots < x_j < N$

Case 4: $0 < \alpha N < x_1 < x_2 < \cdots < x_j < N$
From equation (2.2.1) we can write each case in terms of the probabilities of the points being visited exactly once. We have

$$\mathbb{E}\left[\sum_{x_1<x_2<\cdots<x_j} 1_{A_{x_1}} 1_{A_{x_2}} \ldots 1_{A_{x_j}}\right] = \sum_{x_1<x_2<\cdots<x_j} \mathbb{P}\left[\bigcap_{i=1}^{j}(A_{x_i})\right].$$

Recall that $A_{x_i} = \{T_{x_i} < \tau_N, T^{(2)}_{x_i} > \tau_N\}$, so that the previous equation becomes

$$\sum_{x_1<x_2<\cdots<x_j} \mathbb{P}\left[\bigcap_{i=1}^{j}(\{T_{x_i} < \tau_N, T^{(2)}_{x_i} > \tau_N\})\right]. \tag{2.2.3}$$

2.3 Computing the Four Cases

The following proposition gives an explicit form of (2.2.3) for each case using the standard gambler's ruin identities.

Lemma 2.3.0.1. *For cases 1-4,* (2.2.3) *is explicitly given by*
Case 1:

$$\frac{1}{2^j}\sum_{x_1=1}^{\alpha N-j}\sum_{x_2=x_1+1}^{\alpha N-j+1}\cdots\sum_{x_j=x_{j-1}+1}^{\alpha N-1}\frac{N-\alpha N}{N-x_j}\cdot\frac{1}{x_j-x_{j-1}}\cdot\frac{1}{x_{j-1}-x_{j-2}}\cdots\frac{1}{x_2-x_1}\frac{1}{x_1} \tag{2.3.1}$$

Case 2:

$$\frac{1}{2^j}\sum_{x_1=1}^{\alpha N-j+1}\sum_{x_2=x_1+1}^{\alpha N-j+2}\cdots\sum_{x_j=\alpha N+1}^{N-1}\frac{\alpha N-x_{j-1}}{x_j-x_{j-1}}\cdot\frac{1}{x_j-x_{j-1}}\cdot\frac{1}{x_{j-1}-x_{j-2}}\cdots\frac{1}{x_2-x_1}\frac{1}{x_1} \tag{2.3.2}$$

Case 3:

$$\frac{1}{2^j}\sum_{x_1=1}^{\alpha N-1}\sum_{x_2=\alpha N+1}^{x_3-1}\cdots\sum_{x_j=\alpha N+j-1}^{N-1}\frac{x_2-\alpha N}{x_2-x_1}\cdot\frac{1}{x_2-x_1}\cdot\frac{1}{x_3-x_2}\cdots\frac{1}{x_j-x_{j-1}}\frac{1}{N-x_j} \tag{2.3.3}$$

Case 4:

$$\frac{1}{2^j}\sum_{x_1=\alpha N+1}^{x_2-1}\sum_{x_2=\alpha N+2}^{x_3-1}\cdots\sum_{x_j=\alpha N+j}^{N-1}\frac{\alpha N}{N-x_j}\cdot\frac{1}{x_j-x_{j-1}}\cdot\frac{1}{x_{j-1}-x_{j-2}}\cdots\frac{1}{x_2-x_1}\cdot\frac{1}{x_1} \tag{2.3.4}$$

Proof. These probabilities can be computed explicitly for each possible trajectory using the standard Gambler's ruin identities. That is, for the simple random walk at position b and $a < b < c$ we have

$$\mathbb{P}_b(T_a < T_c) = \frac{c-b}{c-a} \text{ and } \mathbb{P}_b(T_c < T_a) = \frac{b-a}{c-a} \tag{2.3.5}$$

The trajectories of the cases are similar, so let us show the result only for case 1.

We want to determine

$$\mathbb{P}_{\alpha N}\left(\bigcap_{i=1}^{j} A_{x_i}\right)$$

where $A_{x_i} = \{T_{x_i} < \tau_N, T^{(2)}_{x_i}\}$ and the intersection of the A_{x_i} is then the probability of the trajectory satisfying the conditions setup in case 1. Let

$$A^*_{x_i} = \{T_{x_i} < T^{(2)}_{x_{i+1}}, T_{x_i-1} = T_{x_i} + 1\}.$$

Observe, that

$$\bigcap_{i=1}^{j} A_{x_i} = \{T_{x_j} < T_N, T_0 < T^{(2)}_{x_1}, T_{x_j-1} = T_{x_j} + 1\} \cap \left(\bigcap_{i=1}^{j-1} A^*_{x_i}\right).$$

Furthermore, since the random walk is symmetric and from (2.3.5), we see that

$$\begin{aligned} \mathbb{P}_{x_i}(\{T_{x_i-1} = T_{x_i} + 1\}) &= \frac{1}{2} \\ \mathbb{P}_{x_{i+1}-1}(T_{x_i} < T^{(2)}_{x_{i+1}}) &= \frac{1}{x_{i+1} - x_i} \qquad (\ i \le j-1) \\ \mathbb{P}_{\alpha N}(T_{x_j} < T_N) &= \frac{N - \alpha N}{N - x_j} \\ \mathbb{P}_{x_1-1}(T_0 < T^{(2)}_{x_1}) &= \frac{1}{x_1} \end{aligned}$$

Therefore

$$\mathbb{P}_{\alpha N}\left(\bigcap_{i=1}^{j} A_{x_i}\right) = \frac{1}{2^j}\frac{N-\alpha N}{N-x_j} \cdot \frac{1}{x_j - x_{j-1}} \cdot \frac{1}{x_{j-1} - x_{j-2}} \cdots \frac{1}{x_2 - x_1}\frac{1}{x_1}$$

and summing over all possible values of x_i for $1 \le i \le j$ we get

$$\frac{1}{2^j}\sum_{x_1=1}^{\alpha N - j}\ \sum_{x_2=x_1+1}^{\alpha N - j+1} \cdots \sum_{x_j=x_{j-1}+1}^{\alpha N - 1} \frac{N-\alpha N}{N-x_j} \cdot \frac{1}{x_j - x_{j-1}} \cdot \frac{1}{x_{j-1} - x_{j-2}} \cdots \frac{1}{x_2 - x_1}\frac{1}{x_1}.$$

The other cases are similar. □

2.3.1 Symmetry in the Cases

Lemma 2.3.0.2. *Cases 1 and 4 have the same order but with constants* $(1-\alpha)$ *and* α*, respectively. Similarly, cases 2 and 3 have the same order.*

Proof. Let us show the result for cases 1 and 4. Given the domain $[0, N]$, consider the transformed domain determined by the mapping $x \mapsto N - x$. Case 1 under this transformation will then have the ordering

$$0 < N - \alpha N < N - x_j < \cdots < N - x_2 < N - x_1 < N$$

For $1 \leq i \leq j$, let $y_i = N - x_{j+1-i}$, so that the ordering becomes

$$0 < (1-\alpha)N < y_1 < y_2 < \cdots < y_j < N.$$

Observe that y_i traverses the same distance as x_{j+1-i} in case 1 and that distance between points is preserved by the transformation $x \mapsto N - x$. In particular, the gambler's ruin identities are invariant with respect to this transformation. That is, the sum determined by the above ordering and the gambler's ruin identities

$$\frac{1}{2^j} \sum_{y_1=(1-\alpha)N+1}^{y_2-1} \sum_{y_2=(1-\alpha)N+2}^{y_3-1} \cdots \sum_{y_j=(1-\alpha)N+j}^{N-1} \frac{(1-\alpha)N}{y_1} \cdot \frac{1}{y_2 - y_1} \cdots \frac{1}{y_j - y_{j-1}} \cdot \frac{1}{N - y_j} \tag{2.3.6}$$

is equal to the sum (2.3.1) for case 1. Furthermore, we can observe that (2.3.6) is in the same form as (2.3.4) for case 4, but differ only in the constant $(1-\alpha)$ and α. Therefore cases 1 and 4 have the same order but with constants $(1-\alpha)$ and α, respectively. A similar argument works for the order of cases 2 and 3. □

By lemma 2.3.0.2, it is sufficient to evaluate only cases 1 and 2 and account for the change in constants. We will now compute case 1 and therefore case 4 as well by the symmetry argument.

2.3.2 Cases 1 and 4

We compute case 1 by evaluating the sums iteratively. Let us establish a few of our main ingredients used in the calculation and introduce some notation. For $2 \leq j \leq k$ define S_j as

$$S_j = \sum_{x_1=1}^{\alpha N - j} \sum_{x_2=x_1+1}^{\alpha N - j+1} \cdots \sum_{x_j=x_{j-1}+1}^{\alpha N - 1} \frac{N - \alpha N}{N - x_j} \cdot \frac{1}{x_j - x_{j-1}} \cdot \frac{1}{x_{j-1} - x_{j-2}} \cdots \frac{1}{x_2 - x_1} \cdot \frac{1}{x_1}$$

We will start the iteration using the the partial fraction decomposition.

$$\frac{1}{N-x_i}\cdot\frac{1}{x_i-x_{i-1}}=\frac{1}{N-x_{i-1}}\left[\frac{1}{N-x_i}+\frac{1}{x_i-x_{i-1}}\right]. \tag{2.3.7}$$

After the partial fraction decomposition and distributing we will get two sums both of which can be approximated by an integral using the Euler-Maclaurin Formula. The difference between the sum and the integral is bounded by some constant, which we capture with $O(1)$.

$$\begin{aligned}\sum_{x_i=x_{i-1}+1}^{\alpha N-j-1+i}\frac{1}{N-x_i}&=\int_{x_{i-1}+1}^{\alpha N-j-1+i}\frac{1}{N-x_i}dx_i+O(1)\\&=-\log\left(\frac{N-\alpha N-(j-i+1)}{N-x_{j-1}-1}\right)+O(1)\\&=\log\left(\frac{N-x_{j-1}-1}{N-\alpha N-(j-i+1)}\right)+O(1)\end{aligned} \tag{2.3.8}$$

and

$$\begin{aligned}\sum_{x_i=x_{i-1}+1}^{\alpha N-j-1+i}\frac{1}{x_i-x_{i-1}}&=\int_{x_{i-1}+1}^{\alpha N-j-1+i}\frac{1}{x_i-x_{i-1}}dx_i+O(1)\\&=\log(\alpha N-x_{i-1}-(j-i+1))+O(1)\end{aligned} \tag{2.3.9}$$

The following lemma determines S_j up to $o(\log(N)^j)$ and therefore cases 1 and 4 up to $o(\log(N)^j)$.

Lemma 2.3.0.3. *Let $2\le j\le k$, then*

$$S_1=\sum_{x_1=1}^{\alpha N-j}\frac{N-\alpha N}{N-x_1}\cdot\frac{1}{x_1}=(1-\alpha)\log(N)+(1-\alpha)O(1)$$

and

$$S_j=(1-\alpha)\log(N)^j+o(\log(N)^j)$$

Proof. Let us first show the result for S_1. We have

$$S_1=\sum_{x_1=1}^{\alpha N-j}\frac{N-\alpha N}{N-x_1}\cdot\frac{1}{x_1} \tag{2.3.10}$$

We use the partial fraction decomposition (2.3.7) on (2.3.10) to get

$$\sum_{x_1=1}^{\alpha N-j}\frac{N-\alpha N}{N-x_1}\cdot\frac{1}{x_1}=\sigma_{(1,1)}+\sigma_{(2,1)} \tag{2.3.11}$$

where

$$\sigma_{(1,1)} = N - \alpha N \sum_{x_1=1}^{\alpha N - j} \frac{1}{N} \cdot \frac{1}{N - x_1} \tag{2.3.12}$$

and

$$\sigma_{(2,1)} = N - \alpha N \sum_{x_1=1}^{\alpha N - j} \frac{1}{N \cdot} \frac{1}{x_1}. \tag{2.3.13}$$

Applying (2.3.8) on $\sigma_{(1,1)}$ and simplifying we get

$$\sigma_{(1,1)} = (1 - \alpha) \left[\log \left(\frac{N - 1}{N - \alpha N - j} \right) + O(1) \right]$$

where

$$\log \left(\frac{N - 1}{N - \alpha N - j} \right) \xrightarrow[N \to \infty]{} \log(1 - \alpha) = O(1)$$

and therefore we get $\sigma_{(1,1)} = O(1)$.

On the other hand, simplifying (2.3.13) and using 2.3.9 we get

$$\begin{aligned} \sigma_{(2,1)} &= (1 - \alpha) \sum_{x_1=1}^{\alpha N - j} \frac{1}{x_1} \\ &= (1 - \alpha)[\log(\alpha N - j) + O(1)] \\ &= (1 - \alpha) \left[\log(N) + \log(\alpha) + \log \left(1 - \frac{j}{\alpha N} \right) + O(1) \right]. \end{aligned}$$

and therefore

$$\sum_{x_1=1}^{\alpha N - j} \frac{N - \alpha N}{N - x_1} \cdot \frac{1}{x_1} = (1 - \alpha) \log(N) + (1 - \alpha) O(1).$$

Let us now compute S_j. We have

$$S_j = \sum_{x_1=1}^{\alpha N - j} \sum_{x_2=x_1+1}^{\alpha N - j + 1} \cdots \sum_{x_j=x_{j-1}+1}^{\alpha N - 1} \frac{N - \alpha N}{N - x_j} \cdot \frac{1}{x_j - x_{j-1}} \cdot \frac{1}{x_{j-1} - x_{j-2}} \cdots \frac{1}{x_2 - x_1} \cdot \frac{1}{x_1}.$$

Using (2.3.7) we can write S_j as $\sigma_{(j,1)}$ and $\sigma_{(j,2)}$ where

$$\sigma_{(j,1)} = \sum_{x_1=1}^{\alpha N - j} \sum_{x_2=x_1+1}^{\alpha N - j + 1} \cdots \sum_{x_j=x_{j-1}+1}^{\alpha N - 1} \frac{N - \alpha N}{N - x_j} \cdot \frac{1}{N - x_{j-1}} \cdot \frac{1}{x_{j-1} - x_{j-2}} \cdots \frac{1}{x_2 - x_1} \cdot \frac{1}{x_1}$$

and

$$\sigma_{(j,2)} = \sum_{x_1=1}^{\alpha N-j} \sum_{x_2=x_1+1}^{\alpha N-j+1} \cdots \sum_{x_j=x_{j-1}+1}^{\alpha N-1} \frac{N-\alpha N}{x_j - x_{j-1}} \cdot \frac{1}{N-x_{j-1}} \cdot \frac{1}{x_{j-1}-x_{j-2}} \cdots \frac{1}{x_2-x_1} \cdot \frac{1}{x_1}.$$

Let us first consider $\sigma_{(j,1)}$. Using (2.3.8) we get

$$\sum_{x_1=1}^{\alpha N-j} \sum_{x_2=x_1+1}^{\alpha N-j+1} \cdots$$
$$\sum_{x_{j-1}=x_{j-2}+1}^{\alpha N-2} \frac{N-\alpha N}{N-x_{j-1}} \cdot \frac{1}{x_{j-1}-x_{j-2}} \cdots \frac{1}{x_2-x_1} \cdot \frac{1}{x_1} \left[\log\left(\frac{N-x_{j-1}-1}{N-\alpha N+1}\right) + O(1)\right]$$

Observe, using the bounds on x_{j-1}, that is $j-1 \leq x_{j-1} \leq \alpha N - 2$ we get that

$$\log\left(\frac{N-\alpha N+1}{N-\alpha N+1}\right) \leq \log\left(\frac{N-x_{j-1}-1}{N-\alpha N+1}\right) \leq \log\left(\frac{N-j}{N-\alpha N+1}\right)$$

so that

$$0 \leq \log\left(\frac{N-x_j-1}{N-\alpha N+1}\right) \leq \log\left(\frac{1}{1-\alpha}\right)$$

In particular,

$$\log\left(\frac{N-x_{j-1}-1}{N-\alpha N+1}\right) = O(1)$$

and therefore

$$\sum_{x_1=1}^{\alpha N-j} \sum_{x_2=x_1+1}^{\alpha N-j+1} \cdots \sum_{x_{j-1}=x_{j-2}+1}^{\alpha N-2} \frac{N-\alpha N}{N-x_{j-1}} \cdot \frac{1}{x_{j-1}-x_{j-2}} \cdots \frac{1}{x_2-x_1} \cdot \frac{1}{x_1} [O(1)+O(1)]$$
$$= S_{j-1}O(1)$$

so that $\sigma_{(j,1)} = S_{j-1}O(1)$. We will now disregard this term in further computations as once we show that $S_j = (1-\alpha)(\log(N)^j) + o(\log(N)^j)$ then $\sigma_{(j,1)}$ will be captured by the $o(\log(N)^j)$ term.

Let us now consider $\sigma_{(j,2)}$. Now, using (2.3.9), we see that

$$\sigma_{(j,2)} = \sum_{x_1=1}^{\alpha N-j} \sum_{x_2=x_1+1}^{\alpha N-j+1} \cdots$$
$$\sum_{x_{j-1}=x_{j-2}+1}^{\alpha N-2} \frac{N-\alpha N}{N-x_{j-1}} \cdot \frac{1}{x_{j-1}-x_{j-2}} \cdots \frac{1}{x_2-x_1} \cdot \frac{1}{x_1} [\log(\alpha N - x_{j-1} - 1) + O(1)]$$
$$= \sum_{x_1=1}^{\alpha N-j} \sum_{x_2=x_1+1}^{\alpha N-j+1} \cdots$$
$$\sum_{x_{j-1}=x_{j-2}+1}^{\alpha N-2} \frac{N-\alpha N}{N-x_{j-1}} \cdot \frac{1}{x_{j-1}-x_{j-2}} \cdots \frac{1}{x_2-x_1} \cdot \frac{1}{x_1} \log(\alpha N - x_{j-1} - 1) + S_{j-1}O(1) \tag{2.3.14}$$

To evaluate S_j further we now create upper and lower estimates which we will show converge to one another. Let us first establish an upper estimate using the bound $\log(\alpha N - x_{j-1} - 1) \leq \log(N)$. It follows immediately from (2.3.14) that

$$\sigma_{(j,2)} \leq \log(N)S_{j-1} + S_{j-1}O(1)$$

and therefore

$$S_j \leq \log(N)S_{j-1} + S_{j-1}O(1).$$

In particular, we can we can now iterate the above process as the only terms dependent on the constants is the $\log(\alpha N - x_i - k)$, but $\log(\alpha N - x_i - k) \leq \log(N)$. Thus we have a sequence of upper bounds where for each $2 \leq i \leq j$ we have

$$S_i \leq \log(N)S_{i-1} + S_{i-1}O(1).$$

Therefore, by the first part of this lemma evaluating S_1, it follows that

$$S_j \leq (1-\alpha)\log(N)^j + o(\log(N)^j).$$

We now create a lower bound on S_j. We iteratively truncate the sums by choosing a sequence of ϵ to allow lower bounding the sum over a logarithm term by $(1-\epsilon_i)\log(N)$.

Choose $\epsilon_j > 0$ such that $\log(\alpha N - x_{j-1} - 1) \geq (1-\epsilon_j)\log(N)$ so that $x_{j-1} \leq \alpha N - N^{1-\epsilon_j} - 1$. Consider then the lower estimate of given by the following truncated sum of (2.3.14) which we denote $\sigma^*_{(j,2)}$.

$$\sigma^*_{(j,2)} = \sum_{x_1=1}^{\alpha N - N^{1-\epsilon_j}-j} \sum_{x_2=x_1+1}^{\alpha N - N^{1-\epsilon_j}-j+1} \cdots$$
$$\sum_{x_{j-1}=x_{j-2}+1}^{\alpha N - N^{1-\epsilon_j}-2} \frac{N-\alpha N}{N-x_{j-1}} \cdot \frac{1}{x_{j-1}-x_{j-2}} \cdots \frac{1}{x_2-x_1} \cdot \frac{1}{x_1} \left[\log(\alpha N - x_{j-1} - 1) + S^{\epsilon_j}_{j-1} O(1)\right] \tag{2.3.15}$$

where for $2 \le i \le j$ we define

$$S^{\epsilon_i}_i = \sum_{x_1=1}^{\alpha N - N^{1-\epsilon_i}-j} \sum_{x_2=x_1+1}^{\alpha N - N^{1-\epsilon_i}-j+1} \cdots \sum_{x_i=x_{i-1}+1}^{\alpha N - N^{1-\epsilon_i}-j+i-1} \frac{N-\alpha N}{N-x_i} \cdot \frac{1}{x_i - x_{i-1}} \cdots \frac{1}{x_2-x_1} \cdot \frac{1}{x_1}.$$

Using our lower estimate $(1-\epsilon_j)\log(N) \le \log(\alpha N - x_{j-1} - 1)$ it follows immediately from (2.3.15) that

$$(1-\epsilon_j)\log(N) \le \sigma^*_{(j,2)} \le \sigma_{(j,2)}$$

and therefore

$$(1-\epsilon_j)\log(N) S^{\epsilon_{j-1}}_{j-1} + S^{\epsilon_{j-1}}_{j-1} O(1) \le S_j.$$

Let us now evaluate $S^{\epsilon_j}_{j-1}$ for a single iteration. Using (2.3.7) on $S^{\epsilon_j}_{j-1}$ we get $S^{\epsilon_j}_{j-1} = \sigma^*_{(j-1,1)} + \sigma^*_{(j-1,2)}$ where

$$\sigma^*_{(j-1,1)} = \sum_{x_1=1}^{\alpha N - N^{1-\epsilon_j}-j} \sum_{x_2=x_1+1}^{\alpha N - N^{1-\epsilon_j}-j+1} \cdots$$
$$\sum_{x_{j-1}=x_{j-2}+1}^{\alpha N - N^{1-\epsilon_j}-2} \frac{N-\alpha N}{N-x_{j-1}} \cdot \frac{1}{N-x_{j-2}} \cdot \frac{1}{x_{j-2}-x_{j-3}} \cdots \frac{1}{x_2-x_1} \cdot \frac{1}{x_1}$$

and

$$\sigma^*_{(j-1,2)} = \sum_{x_1=1}^{\alpha N - N^{1-\epsilon_j}-j} \sum_{x_2=x_1+1}^{\alpha N - N^{1-\epsilon_j}-j+1} \cdots$$
$$\sum_{x_{j-1}=x_{j-2}+1}^{\alpha N - N^{1-\epsilon_j}-2} \frac{N-\alpha N}{x_{j-1}-x_{j-2}} \cdot \frac{1}{N-x_{j-2}} \cdot \frac{1}{x_{j-2}-x_{j-3}} \cdots \frac{1}{x_2-x_1} \cdot \frac{1}{x_1}$$

Since $\sigma_{(1,1)} = O(S_{j-1})$, we can observe that the restricted $\sigma^*_{(j-1,1)} \le O(S_{j-1})$. In particular, we no longer need to continue with our calculations on this term. Continuing the calculation with $\sigma^*_{(i,2)}$, we have

$$\sigma^*_{(j-1,2)} = \sum_{x_1=1}^{\alpha N - N^{1-\epsilon_j}-j} \sum_{x_2=x_1+1}^{\alpha N - N^{1-\epsilon_j}-j+1} \cdots$$
$$\sum_{x_{j-1}=x_{j-2}+1}^{\alpha N - N^{1-\epsilon_j}-2} \frac{N-\alpha N}{x_{j-1}-x_{j-2}} \cdot \frac{1}{N-x_{j-2}} \cdot \frac{1}{x_{j-2}-x_{j-3}} \cdots \frac{1}{x_2-x_1} \cdot \frac{1}{x_1}$$

and using (2.3.9) we get

$$\begin{aligned}\sigma^*_{(j-1,2)} = & \sum_{x_1=1}^{\alpha N - N^{1-\epsilon_j}-j} \sum_{x_2=x_1+1}^{\alpha N - N^{1-\epsilon_j}-j+1} \cdots \\ & \sum_{x_{j-2}=x_{j-3}+1}^{\alpha N - N^{1-\epsilon_j}-3} \cdot \frac{N-\alpha N}{N-x_{j-2}} \cdot \frac{1}{x_{j-2}-x_{j-3}} \cdots \\ & \frac{1}{x_2-x_1} \cdot \frac{1}{x_1} \left[\log(\alpha N - N^{1-\epsilon_j} - x_{j-2} - 2) + O(1)\right]\end{aligned} \tag{2.3.16}$$

Therefore

$$\begin{aligned}\sigma^*_{(j-1,2)} = & \sum_{x_1=1}^{\alpha N - N^{1-\epsilon_j}-j} \sum_{x_2=x_1+1}^{\alpha N - N^{1-\epsilon_j}-j+1} \cdots \\ & \sum_{x_{j-2}=x_{j-3}+1}^{\alpha N - N^{1-\epsilon_j}-3} \cdot \frac{N-\alpha N}{N-x_{j-2}} \cdot \frac{1}{x_{j-2}-x_{j-3}} \cdots \\ & \frac{1}{x_2-x_1} \cdot \frac{1}{x_1} \log(\alpha N - N^{1-\epsilon_j} - x_{j-2} - 2) + O(1) S^{\epsilon_j}_{j-2}.\end{aligned}$$

We now truncate the sums further by choosing $\epsilon_{j-1} > 0$ such that $(1-\epsilon_{j-1})\log(N) \leq \log(\alpha N - N^{1-\epsilon_j} - x_{j-2} - 2)$. After truncating the sums, it follows immediately that

$$(1-\epsilon_{j-1})\log(N) S^{\epsilon_{j-1}}_{j-2} + O(1) S^{\epsilon_{j-1}}_{j-2} \leq \sigma^*_{(j-1,2)} \leq S^{\epsilon_j}_{j-1} \leq S_{j-1}.$$

We can continue iterate this process so that for $2 \leq i \leq j-1$ we have

$$(1-\epsilon_{i-1})\log(N) S^{\epsilon_{i-1}}_{i-1} + O(1) S^{\epsilon_{i-1}}_{i-1} \leq S^{\epsilon_i}_{i} \leq S_{i-1}.$$

It follows that

$$[(1-\alpha)\log(1-\epsilon_1)\ldots(1-\epsilon_j)\log(N)^{j-1} + o\log(N)^{j-1}] \leq S_j.$$

Let $\epsilon \leq \epsilon_1$, so that we have $(1-\epsilon)^{j-1} \leq \prod_{i=1}^{j}(1-\epsilon_i)$ and so we have the lower estimate

$$(1-\alpha)(1-\epsilon)^{j-1}\log(N)^j \leq S_j.$$

Combining the upper and lower estimates we have

$$(1-\alpha)(1-\epsilon)^j\log(N)^j \leq S_j \leq (1-\alpha)\log(N)^j.$$

Hence,

$$(1-\alpha)(1-\epsilon)^j \leq \frac{S_j}{\log(N)^j} \leq (1-\alpha)$$

and taking limits as $N \to \infty$ and then $\epsilon \to 0$ we get

$$\lim_{\epsilon\to 0}\lim_{N\to\infty}\frac{S_j}{\log(N)} = (1-\alpha).$$

The result follows.

□

2.3.3 Bounding Cases 2 and 3

Lemma 2.3.0.4. *Cases 2 and 3 are bounded above by*

$$S_{j-1} = O(\log(N)^{j-1}) + o(\log(N)^{j-1}).$$

Proof. Recall that the summation for case 2 is given by

$$\frac{1}{2^j}\sum_{x_1=1}^{\alpha N-j+1}\sum_{x_2=x_1+1}^{\alpha N-j+2}\cdots\sum_{x_j=\alpha N+1}^{N-1}\frac{\alpha N-x_{j-1}}{x_j-x_{j-1}}\cdot\frac{1}{x_j-x_{j-1}}\cdot\frac{1}{x_{j-1}-x_{j-2}}\cdots\frac{1}{x_2-x_1}\frac{1}{x_1}. \tag{2.3.17}$$

Taking the inner sum and using Euler-Maclaurin formula, we get

$$\begin{aligned}\sum_{x_j=\alpha N+1}^{N-1}\frac{\alpha N-x_{j-1}}{(x_j-x_{j-1})^2} &= (\alpha N-x_{j-1})\left(\frac{1}{\alpha N-x_{j-1}+1}-\frac{1}{N-x_{j-1}-1}\right)+O(1)\\ &= (\alpha N-x_{j-1})\left(\frac{N-\alpha N-2}{(\alpha N-x_{j-1}+1)(N-x_{j-1}-1)}\right)+O(1)\\ &\leq \frac{N-\alpha N-2}{N-x_{j-1}-1}+O(1).\end{aligned}$$

Notice that for large N

$$\frac{N-\alpha N-2}{N-x_{j-1}-1} \leq \frac{N-\alpha N}{N-j-3} = \frac{1-\alpha}{1-\frac{j-3}{N}} = O(1)$$

so that the inner summation is $O(1)$. Therefore we can write (2.3.17) as

$$\frac{c}{2^j} \sum_{x_1=1}^{\alpha N-j+1} \sum_{x_2=x_1+1}^{\alpha N-j+2} \cdots \sum_{x_{j-1}=x_{j-2}+1}^{\alpha N-1} \frac{1}{x_{j-1}-x_{j-2}} \cdot \frac{1}{x_{j-2}-x_{j-3}} \cdots \frac{1}{x_2-x_1} \frac{1}{x_1}$$

where $0 < c < \infty$. Using (2.3.9) repeatedly we can see that the above summation is bounded above by

$$\frac{c}{2^j} \log^{j-1}(N).$$

Therefore in the scaling limit the terms contributed by case 2 will vanish. Furthermore, from proposition 2.3.0.2 we can conclude that case 2 has the same order as case 3, so that the sums from case 3 will vanish asymptotically when we divide by $\log(N)^j$ as well. □

2.3.4 Concluding proof of Lemma 2.2.0.1

From lemma 2.3.0.1, lemma 2.3.0.2, and lemma 2.3.0.4 we see that it suffices to compute case 1 given by (2.3.1) to determine

$$j!\mathbb{E}_{\alpha N}\left[\sum_{\substack{x_1,x_2,\ldots,x_k \\ \text{distinct}}} 1_{A_{x_1}} 1_{A_{x_2}} \ldots 1_{A_{x_j}}\right].$$

Furthermore, by lemma 2.3.0.3, we have that $S_j = (1-\alpha)\log(N)^j + o\log(N)^j$ and therefore case 1 is given by

$$\frac{j!}{2^j}(1-\alpha)\log(N)^j + o(\log(N)^j)$$

and by lemma 2.3.0.2 case 4 is given by

$$\frac{j!}{2^j}\alpha\log(N)^j + o(\log(N)^j).$$

Therefore

$$\begin{aligned}
\mathbb{E}_{\alpha N}\left[\sum_{\substack{x_1,x_2,\ldots,x_k \\ \text{distinct}}} 1_{A_{x_1}} 1_{A_{x_2}} \ldots 1_{A_{x_j}}\right] &= \frac{1}{2^j}(1-\alpha)\log(N)^j + \frac{j!}{2^j}\alpha\log(N)^j + o(\log(N)^j) \\
&= \frac{j!}{2^j}\log(N)^j + o(\log(N)^j)
\end{aligned} \tag{2.3.18}$$

proving lemma 2.2.0.1.

2.4 Proof of Proposition 2.1.1

We now prove proposition 2.1.1. From lemma 2.2.0.1

$$\mathbb{E}_{\alpha N}\left[(R_N^1)^k\right] = \frac{k!\log^k(N)}{2^k}.$$

It follows that

$$\mathbb{E}_{\alpha N}\left[\left(\frac{R_N^1}{\log(N)}\right)^k\right] = \frac{k!}{2^k}.$$

Now we need to show that the limiting distribution $\lim_{n\to\infty}\frac{R_N^1}{\log(N)} \sim \text{Exp}(2)$. We may either recognize that this is the known moment value for Exp(2), or we may show that the moment generating functions are the same. Let $\mathcal{R}_N^1 = \frac{R_N^1}{\log(N)}$. Hence,

$$\begin{aligned} M_{\mathcal{R}_N^1}(t) &= 1 + t\mathbb{E}[\mathcal{R}_N^1] + \frac{t^2\mathbb{E}[(\mathcal{R}_N^1)^2]}{2!} + \cdots + \frac{t^n\mathbb{E}[(\mathcal{R}_N^1)^n]}{n!} + \ldots \\ &= 1 + \frac{t}{2} + \frac{t^2}{2^2} + \cdots + \frac{t^n}{2^n} + \ldots \end{aligned}$$

which is a geometric series which for $|t| < 2$ converges, so that

$$M_{\mathcal{R}_N^1}(t) = \frac{2}{2-t} \text{ for } |t| < 2.$$

Furthermore, we have

$$M_{\text{Exp}(2)}(t) = \mathbb{E}[e^{t\exp(2)}] = \int_0^\infty e^{tx}2e^{-2x}dx = \frac{2}{2-t} = M_{\mathcal{R}_N^1}(t)$$

Therefore

$$\mathbb{E}_{\alpha N}\left[\left(\frac{R_N^1}{\log(N)}\right)^k\right] = \frac{k!}{2^k}$$

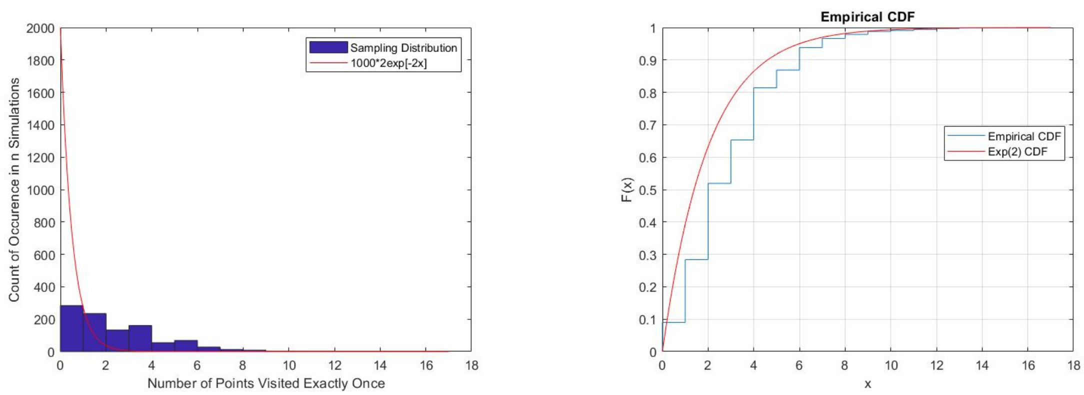
and

$$\lim_{n\to\infty}\frac{R_N^1}{\log(N)} \sim \text{Exp}(2)$$

so that the proposition is proven.

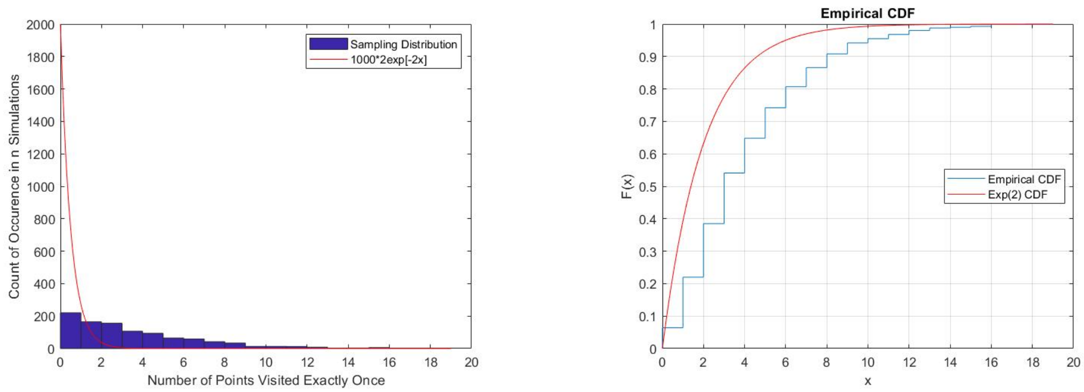
2.5 Numerical Results

In this section we discuss some numerical results for the empirical density and empirical cumulative distribution function of the points visited exactly once on the domain $[0, N]$. In Figures 2.2, 2.3, and 2.4, 1000 simulations are run on the domains $[0, 10], [0, 100]$, and $[0, 1000]$, respectively, each with starting position $N/2$. The graph on the left plots the histogram of points visited exactly once, resulting in an empirical density. We additionally plot the scaled density $1000 \cdot 2e^{-2x}$, scaling by the number of simulations, to compare the empirical density and the density of our limiting distribution. The graph on the right plots the empirical cumulative distribution versus the cumulative density function of Exp(2). We find, even for relatively small number of simulations and domain length N, that the empirical plots fit well with that of the limiting distribution Exp(2). Thus, we have a numerical indication that Proposition 2.1.1 will be numerically verified for large N.



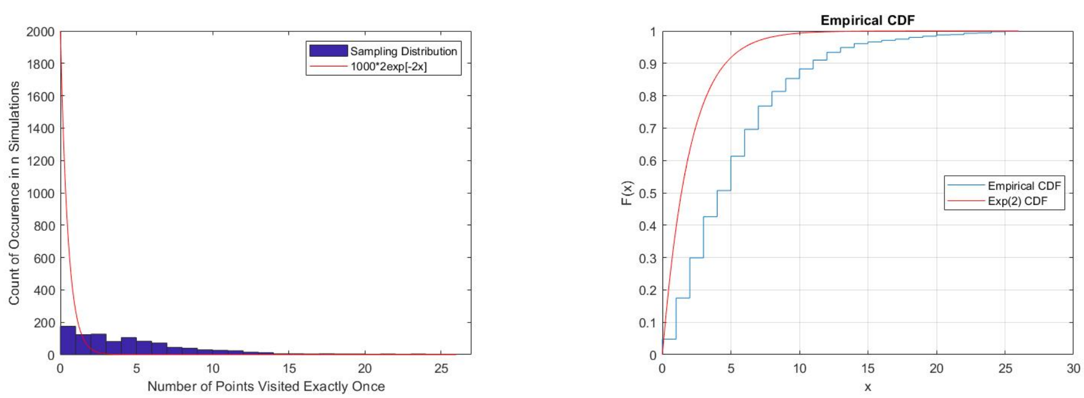
(a) (b)

Figure 2.2: (a) In blue the histogram of points visited exactly once with 1000 simulations on the domain $[0, 10]$ and starting position of 5. In red density of Exp(2) scaled by the number of simulations.. (b) Empirical CDF of data in (a) plotted with the CDF of Exp(2).



(a) (b)

Figure 2.3: (a) In blue the histogram of points visited exactly once with 1000 simulations on the domain $[0, 100]$ and starting position of 50. In red density of Exp(2) scaled by the number of simulations.. (b) Empirical CDF of data in (a) plotted with the CDF of Exp(2).



(a) (b)

Figure 2.4: (a) In blue the histogram of points visited exactly once with 1000 simulations on the domain $[0, 1000]$ and starting position of 500. In red density of Exp(2) scaled by the number of simulations.. (b) Empirical CDF of data in (a) plotted with the CDF of Exp(2).

2.6 Further Considerations

In concluding this chapter we make a few remarks on the future of the project. A first goal is to compute the scaling distributions for the multiple range $R_N^{(p)}$ at time of exit for $p > 1$ for the simple symmetric random walk. We are also interested in considering the case where the random walk is biased. Further work may also consider other domains such as the domain $[-N, N]$, graphs, and others. Another area of interest is due to recent work by Jego [14, 15] who discussed the connection of "thick points" with the Gaussian Free Field. Thick points are the points visited "often" by a continuous time random walk before exiting its domain. Specifically, the points need to be visited at least of an order $a(\log N)^2$ times in $d = 2$ or order $a \log N$ times in $d \geq 3$ where $a \geq 0$. In particular, we are interested in the random walk on domain $[-N, N]$ with starting position at zero and finding analogous definitions for thick points in $d = 1$ and study any potential connections to the Gaussian Free Field.

Chapter 3

Influencer Voter Model

3.1 Introduction

In this section we construct the influencer voter model on the complete graph and consider its dynamics. We find that it is sufficient to model the system by tracking the number of voters of opinion 1 in the influencer and regular sets of vertices. Hence, we can take the perspective of the influencer voter model as being a two-dimensional nearest neighbor random walk.

Let $K_N = (V, E)$ be the undirected complete graph with N nodes, or vertices. As with the voter model, each node is in either two states where the state space is $S = \{0, 1\}$. Let $\mathcal{I} \subset V$ be the set of influencer nodes and $\mathcal{R} \subset V$ be the set of regular nodes. We denote the number of influencers and regular nodes by I and R, respectively. That is, $|\mathcal{I}| = I$ and $|\mathcal{R}| = R$. We also require that $I + R = N$.

We use a continuous time model to update the states of the nodes. There are N independent unit rate Poisson Processes $T_v(t)$ associated with each node v, so that jump times are independent and given by Exp(1). As the minimum of of independent exponential random variables is an exponential random variable with a rate equal to the sum of the rates, this corresponds to system with a single clock ticking according to rate N Poisson processes at times Z_k, $k > 1$ where $\{Z_{k+1} - Z_k\}$ are i.i.d exponentials at rate N. At time Z_k, a node $i \in V$ is chosen uniformly at random and the node i will update its opinion to a chosen neighbor $j \in V \setminus i$ where the probability of j being chosen is determined by whether $j \in R$ or $j \in I$. We let p be the probability that we choose a neighbor uniformly from the set of influencer nodes and $q = 1 - p$ be the probability that we choose a neighbor uniformly from the set of regular nodes. We can summarize the process as follows:

1. At a time of update a node i is chosen uniformly at random from the set V.

2. A neighbor j is then chosen, such that with probability p, $j \in \mathcal{I}$ and, with probability q, $j \in \mathcal{R}$.

3. The node i then updates its opinions state to that of its randomly chosen neighbor j.

Repeat 1-3 until consensus is reached.

3.2 System Dynamics

We now describe the dynamics for the influencer voter model. Since we are only considering the influencer voter model on the complete graph, we do not need to account for neighbor relations and therefore the system at time t is given by the number of voters(nodes) of either opinion. For each vertex k let $U_k = 1$ if k is in state 1, $V_k = 1$ if k is in state 0. Then for any time $t > 0$ let

$$U(t) = \sum_{k\in\mathcal{R}} U_k,\ V(t) = \sum_{k\in\mathcal{R}} V_k,\ U^*(t) = \sum_{k\in\mathcal{I}} U_k,\ \text{and } V^*(t) = \sum_{k\in\mathcal{I}} V_k^*.$$

Let $P(i,j) \geq 0$ be the probability that node i chooses neighbor j to update its opinion from. Then

$$P(i,j) = \frac{pI}{N} \text{ if } j \in \mathcal{I} \text{ or } P(i,j) = \frac{qR}{N} \text{ if } j \in \mathcal{R}$$

and we can specify the transition rate $Q(i,j)$, or the rate at which node i changes to the opinion of node j by

$$Q(i,j) = P(i,j)\lambda_i.$$

Where $\lambda_i = N$ for all i, where N is the rate of the exponential timer giving the update time for our system. Therefore, temporarily suppressing the notation showing dependence on t, we can compactly express the dynamics of the system in terms of continuous-time Markov process (U, V, U^*, V^*) with transition rates given by

$$(U,V,U^*,V^*) \to \begin{cases} (U+1, V-1, U^*, V^*): & V\left(p\frac{U^*}{U^*+V^*} + q\frac{U}{U+V}\right) \\ (U-1, V+1, U^*, V^*): & U\left(p\frac{V^*}{U^*+V^*} + q\frac{V}{U+V}\right) \\ (U, V, U^*+1, V^*-1): & V^*\left(p\frac{U^*}{U^*+V^*} + q\frac{U}{U+V}\right) \\ (U, V, U^*-1, V^*+1): & U^*\left(p\frac{V^*}{U^*+V^*} + q\frac{V}{U+V}\right) \end{cases}$$

However, given the relations $U + V = R$ and $U^* + V^* = I$ the Markov process (U, U^*) gives a sufficient description with transition rates

$$(U,U^*) \to \begin{cases} (U+1, U^*): & (R-U)\left(p\frac{U^*}{I} + q\frac{U}{R}\right) \\ (U-1, U^*): & U\left(p\frac{I-U^*}{I} + q\frac{R-U}{R}\right) \\ (U, U^*+1): & (I-U^*)\left(p\frac{U^*}{I} + q\frac{U}{R}\right) \\ (U, U^*-1): & U^*\left(p\frac{I-U^*}{I} + q\frac{R-U}{R}\right) \end{cases} \tag{3.2.1}$$

In particular, the evolution of the influencer voter model can be represented as a two-dimensional nearest neighbor random walk, with transition rates given in (3.2.1).

A visualization of the $(U(t), U^*(t))$ as a nearest neighbor random walk is seen in Fig. 3.2.

IVM as a Two-Dimensional Random Walk

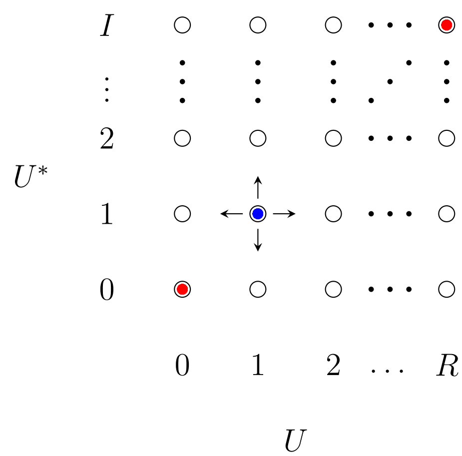


Figure 3.1: Illustration of the Influencer Voter Model as a two-dimensional nearest neighbor random walk. Current state is highlighted in blue and absorbing states highlighted in red.

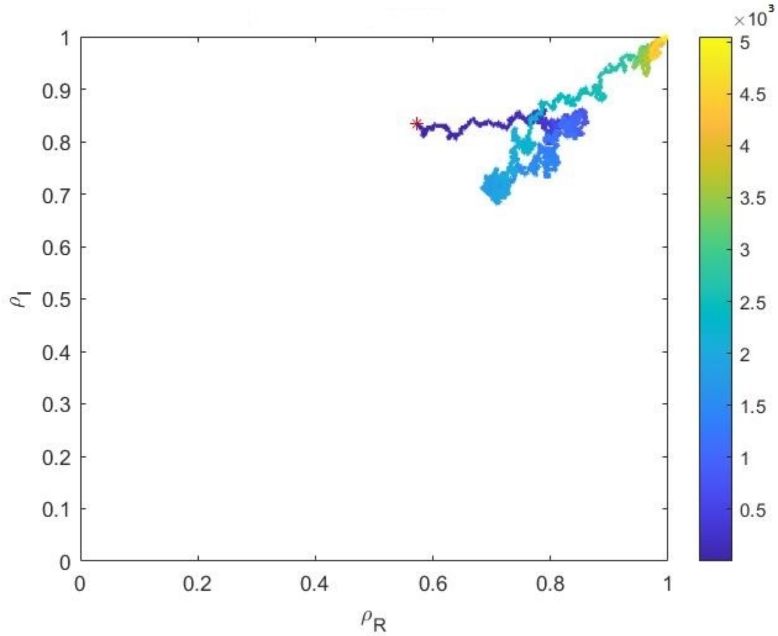
3.3 Scaled Process and the Fluid Limit

In this section we analyze the properties of the scaled Markov Process $(U(t)/N, U^*(t)/N) = (u_N(t), u_N^*(t))$. Specifically, interpreting the scaled Markov process as a random function taking values in the unit square $[0, 1]^2$, we get a law of large numbers type result where the sample paths converge weakly to a deterministic path, the fluid limit, which is the solution to a pair of ordinary differential equations. [1]

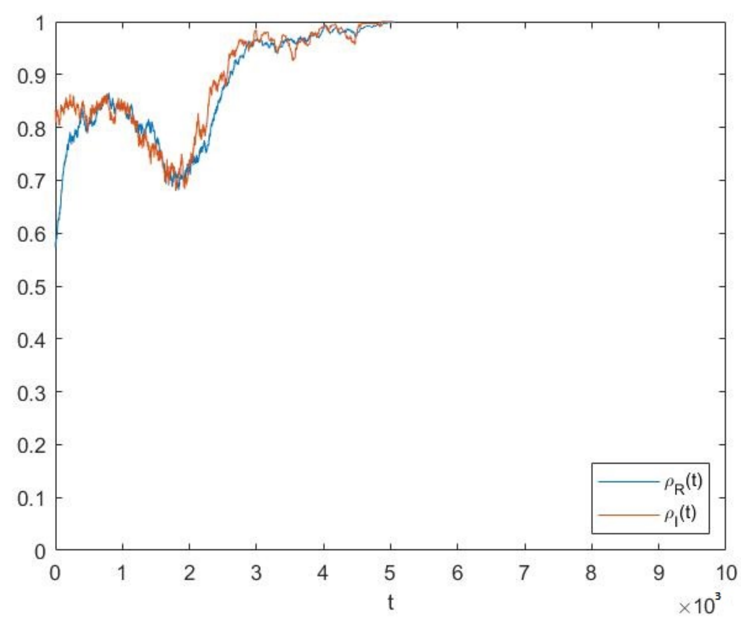
We define the scaled states $u_N(t) = U(t)/N$, $u_N^*(t) = U^*(t)/N$, the scaled constants $r_N = R/N$, $i_N = I/N$, and the scaled Markov process $(u_N(t), u_N^*(t))$. Notice for the scaled process that as N gets large the jumps become more frequent, with mean time $\frac{1}{N}$, and the sizes of the jumps, $\frac{1}{N}$ becomes smaller, not only suggesting that the scaled process can be approximated by a continuous path, but the law of large numbers result we are seeking to prove. In fact, we will show that$(u_N(t), u_N^*(t)) \Rightarrow (u(t), u^*(t))$ whose dynamics are given by a system ordinary differential equations.

[1] Other limits have been considered, for example, Cox et al., [7] found that under central limit scaling that two-dimensional voter models converge to super Brownian motion.

Scaled IVM Sample Paths



(a)



(b)

Figure 3.2: (a), The two-dimensional sample path in terms of the densities ρ_R, ρ_I.(b) each sample path in terms of the densities ρ_R, ρ_I. Both plots are made continuous by linear interpolation. In this simulation we have $R = 700, I = 300, \rho_r = .57, \rho_I = .83$ and $p = 0.8$. Consensus was reached on (R, I) in around 5000 steps.

Let $\mathcal{L}$ be the generator of $(U(t), U^*(t))$, which in this case is the transition matrix determined by the rates in (3.2.1). That is, for any $f \in \mathcal{D}(\mathcal{L})$ we have,

$$\begin{aligned}
\mathcal{L}f((U,U^*)) = {} & (f((U+1,U^*)) - f((U,U^*)) \left((R-U)\left(p\frac{U^*}{I} + q\frac{U}{R}\right)\right) \\
& + (f((U-1,U^*)) - f((U,U^*)) \left(U\left(p\frac{I-U^*}{I} + q\frac{R-U}{R}\right)\right) \\
& + (f((U,U^*+1)) - f((U,U^*)) \left((I-U^*)\left(p\frac{U^*}{I} + q\frac{U}{R}\right)\right) \\
& + (f((U,U^*-1)) - f((U,U^*)) \left(U^*\left(p\frac{I-U^*}{I} + q\frac{R-U}{R}\right)\right)
\end{aligned} \tag{3.3.1}$$

Key to our analysis will be the use of Dynkin's formula to represent our Markov Process as

$$f(U(t), U^*(t)) - f(U(0), U^*(0)) = \int_0^t \mathcal{L}f(U(t), U^*(t))ds + M^f(t) \tag{3.3.2}$$

where $M^f(t)$ is a Martingale with respect to the natural filtration for all $f \in D(\mathcal{L})$.

The following proposition states that the scaled process converges weakly to a system of ordinary differential equations, establishing a weak law of large numbers result. In particular, we see that the limiting process is deterministic and we determine the unique solution to the system of ordinary differential equations. Similar proofs and methods for a strong law of large numbers type result have been shown by Darling [8] and Ethier and Kurtz [11].

Proposition 3.3.1. *The scaled Markov Process* $(u_N(t), u_N^*(t))$ *converges weakly on any compact time interval to the continuous, Markov Process* $(u(t), u^*(t))$ *given by the system of ordinary differential equations*

$$\begin{aligned}
\frac{du(t)}{dt} &= p \cdot r\left(\frac{u^*(t)}{i} - \frac{u(t)}{r}\right) \\
\frac{du^*(t)}{dt} &= q \cdot i\left(\frac{u(t)}{r} - \frac{u^*(t)}{i}\right)
\end{aligned} \tag{3.3.3}$$

and solution, with initial conditions $u(0)$ *and* $u^*(0)$*, is given by*

$$\begin{aligned}
u(t) &= r\left(p\frac{u^*(0)}{i} + q\frac{u(0)}{r}\right) + r \cdot p\left(\frac{u(0)}{r} - \frac{u^*(0)}{i}\right)e^{-t} \\
u^*(t) &= i\left(p\frac{u^*(0)}{i} + q\frac{u(0)}{r}\right) - i \cdot q\left(\frac{u(0)}{r} - \frac{u^*(0)}{i}\right)e^{-t}
\end{aligned} \tag{3.3.4}$$

where $r = \lim_{n\to\infty} r_N$ *and* $i = \lim_{n\to\infty} i_N$

Proof. Since $(U(t), U^*(t))$ is a Markov process, from (3.3.2), we have the representation

$$f(U(t), U^*(t)) - f(U(0), U^*(0)) = \int_0^t \mathcal{L}f(U(t), U^*(t))ds + M^f(t).$$

We use the coordinate function $f_1(x, y) = x$ and scale by N^{-1} to get the one-dimensional scaled Markov process which counts the number of regular voters of opinion 1 as

$$u_N(t) - u_N(0) = \frac{1}{N}\int_0^t \mathcal{L}U(s)ds + \frac{1}{N}M^{f_1}(t). \tag{3.3.5}$$

We want to show that as $N \uparrow \infty$ that (3.3.5) converges to some deterministic equation. We first need to show that the family $\{u_N(t)\}$ is tight in the space $C[0, 1]$. Note, that although $u_N(t)$ is not continuous, for each N we can consider the linear interpolation of the points in the sample path of $u_N(t)$ as seen in Fig. 3.3. Of course, in the limit as $N \uparrow \infty$ the differences between the discrete model and the linear interpolation model vanish. To show tightness of $\{u_N(t)\}$ we show tightness of each term in (3.3.5). Notice that the initial values $\{u_N(0)\}$ are tight as they are uniformly bounded. For the tightness of $\{\frac{1}{N}\int_0^t \mathcal{L}U(s)ds\}$ we use Kolmogorov-Centsóv criterion for tightness. Since

$$0 \leq \frac{1}{N}\mathcal{L}U(s) = p \cdot \frac{R}{N}\left(\frac{U^*(t)}{I} - \frac{U(t)}{R}\right) \leq 1$$

we have

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{N}\int_0^s \mathcal{L}U(s)ds - \frac{1}{N}\int_s^t \mathcal{L}U(s)ds\right)^2\right] &\leq K\mathbb{E}\left[\left(\int_s^t 1ds\right)^2\right] \\ &= K(t-s)^2\end{aligned} \tag{3.3.6}$$

for some $K > 0$. Thus the Kolmogorov-Centsóv criterion for tightness is satisfied and since we are showing tightness on $C[0, 1]$, tightness holds for all $s, t \in [0, 1]$ with $s < t$.

Lastly, we show tightness on $\{\frac{M^{f_1}(t)}{N}\} = \{M_N^{f_1}(t)\}$. We use the criterion from Billinsgley [5, Ch2. Thm. 7.3] for tightness on $C[0, 1]$. We show that for each positive ϵ and η there exists a $\delta, 0 < \delta < 1$ and an integer n_0, such that

$$\frac{1}{\delta}\mathbb{P}\left\{\sup_{t \leq s \leq t+\delta} |u_N(s) - u_N(t)| \geq \epsilon\right\} \leq \eta.$$

From Doob's Maximal inequality, we have

$$\frac{1}{\delta}\mathbb{P}\left\{\sup_{t \leq s \leq t+\delta} \left|M_N^{f_1}(t) - M_N^{f_1}(s)\right| \geq \epsilon\right\} \leq \frac{c_4}{\delta}\frac{\mathbb{E}\left[(M_N^{f_1}(t+\delta) - M_N(t))^4\right]}{\epsilon^4}.$$

where c_4 is some positive constant. We can overestimate $\mathbb{E}\left[(M_N^{f_1}(t+s) - M_N(t))^4\right]$ by the Burkholder-Davis-Gundy inequality

$$\mathbb{E}\left[(M_N^{f_1}(t+s) - M_N(t))^4\right] \leq C_4 \mathbb{E}\left[\langle M_N^{f_1}(t+\delta) - M_N^{f_1}(t)\rangle^2\right]$$

where C_4 is some positive constant. Since $M_N^{f_1}(t)$ is continuous and has finite quadratic variation, $\langle M_N^{f_1}(t+\delta) - M_N^{f_1}(t)\rangle = \delta$, so that

$$\frac{1}{\delta}\mathbb{P}\left\{\sup_{t\leq s\leq t+\delta} |u_N(s) - u_N(t)| \geq \epsilon\right\} \leq \frac{K\delta}{\epsilon^4}$$

and

$$\frac{K\delta}{\epsilon^4} \xrightarrow[\delta\to 0]{} 0.$$

Therefore we have established that the family $\{u_N(t)\}$ is tight. We now what to determine what the limit of the process is. Let us first show that asymptotically the stochastic perturbations given by $M^{f_1}(t)$ vanish. Notice from (3.3.5) that $M^{f_1}(0) = 0$, so that $\mathbb{E}[M^{f_1}(t)] = 0$. It then suffices to show that the quadratic variation of the Martingale vanishes in the limit. We can explicitly write

$$\langle M^{f_1}(t)\rangle = \int_0^t \mathcal{L}U(s)^2 - 2U(s)\mathcal{L}U(s)ds,$$

and, observing that the integrand is $O(N)$ from (3.3.1), we have

$$\begin{aligned}\mathbb{E}\left[\left(\frac{M^{f_1}(t)}{N}\right)^2\right] &= \mathbb{E}\left[\frac{M^{f_1}(t)^2}{N^2}\right] \\ &= \frac{1}{N^2}\mathbb{E}\left[\int_0^t \mathcal{L}U(s)^2 - 2U(s)\mathcal{L}U(s)ds\right] \\ &\leq \frac{CN}{N^2}\end{aligned} \tag{3.3.7}$$

where C is some constant. In particular, in the limit the expectation of the square of $M^{f_1}(t)N^{-1}$ converges to the zero, so that $M^{f_1}(t)N^{-1}$ vanishes asymptotically. Define

$\lim_{N\to\infty} u_N(t) = u(t)$ and $\lim_{N\to\infty} u_N(0) = u(0)$. It follows that

$$\begin{aligned}
u(t) - u(0) &= \lim_{N\to\infty} \frac{1}{N} \int_0^t \mathcal{L}U(s)ds \\
&= \lim_{N\to\infty} \frac{1}{N} \int_0^t \left((R-U)\left(p\frac{U^*(s)}{I} + q\frac{U(s)}{R}\right)\right) \\
&\qquad - \left(U(s)\left(p\frac{I-U^*(s)}{I} + q\frac{R-U(s)}{R}\right)\right) ds \\
&= \lim_{N\to\infty} \int_0^t p\cdot r_N \left(\frac{u_N^*(s)}{i_N} - \frac{u_N(s)}{r_N}\right) ds \\
&= \int_0^t p\cdot r\left(\frac{u^*(s)}{i} - \frac{u(s)}{r}\right).
\end{aligned}$$

and therefore

$$\frac{du(t)}{dt} = p\cdot r\left(\frac{u^*(t)}{i} - \frac{u(t)}{r}\right).$$

Thus, from tightness of the family, for any subsequence $u_{N_k}(t)$ that $u_{N_k}(t) \to u(t)$. Since we can verify that the solution to $\frac{du(t)}{dt}$ with initial conditions $u(0), u^*(0)$, is

$$u(t) = r\left(p\frac{u^*(0)}{i} + q\frac{u(0)}{r}\right) + r\cdot p\left(\frac{u(0)}{r} - \frac{u^*(0)}{i}\right)e^{-t}$$

we get uniqueness of a limit and therefore convergence of the sequence $u_N(t) \to u(t)$. The argument is similar for $u^*(t)$ taking the coordinate function $f_2(x,y) = y$ and we get (3.3.3)

$$\begin{aligned}
\frac{du(t)}{dt} &= p\cdot r\left(\frac{u^*(t)}{i} - \frac{u(t)}{r}\right) \\
\frac{du^*(t)}{dt} &= q\cdot i\left(\frac{u(t)}{r} - \frac{u^*(t)}{i}\right)
\end{aligned}$$

with solution (3.3.4)

$$\begin{aligned}
u(t) &= r\left(p\frac{u^*(0)}{i} + q\frac{u(0)}{r}\right) + r\cdot p\left(\frac{u(0)}{r} - \frac{u^*(0)}{i}\right)e^{-t} \\
u^*(t) &= i\left(p\frac{u^*(0)}{i} + q\frac{u(0)}{r}\right) - i\cdot q\left(\frac{u(0)}{r} - \frac{u^*(0)}{i}\right)e^{-t}.
\end{aligned}$$

□

Notice that we can write the (3.3.4) in terms of the subgraph densities ρ_I, ρ_R as

$$\begin{aligned} \rho_R(t) &= (p\rho_I(0) + q\rho_R(0)) + p\,(\rho_R(0) - \rho_I(0))\,e^{-t} \\ \rho_I(t) &= (p\rho_I(0) + q\rho_R(0)) - q\,(\rho_R(0) - \rho_I(0))\,e^{-t}. \end{aligned} \tag{3.3.8}$$

In particular, we can see that the path of the average trajectory follows the solution in (3.3.8) and tends towards equal subgraph densities at the point $((p\rho_I(0) + q\rho_R(0), (p\rho_I(0) + q\rho_R(0))$. After reaching the point of equal subgraph densities, however, the dynamics of the trajectories characteristically change as the path then diffusively fluctuates along the diagonal $\rho_I = \rho_R$ until consensus is reached. This behavior can be seen in Fig.3.3.

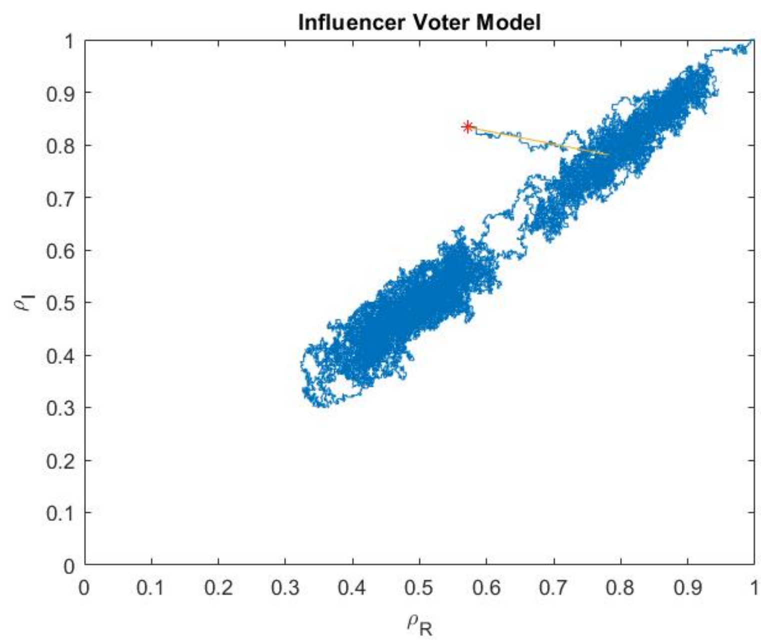


Figure 3.3: Sample Path of Influencer Voter Model in blue, solution to system of ODEs given in (3.3.8) in yellow, and starting position given by the red asterisks. In this simulation we have $R = 700, I = 300, \rho_r = .57, \rho_I = .83$ and $p = 0.8$.

We can therefore describe the trajectories of the influencer voter model into two parts. The first part is determined by (3.3.8), which is given by the bias in the transition probabilities which we call the drift. The drift here is the product of the average jump size by the total rate. We have derived these as the differential equations in (3.3.3). Rewriting (3.3.3) in terms of the densities, we get

$$\begin{aligned} \frac{d\rho_R(t)}{dt} &= p(\rho_I(t) - \rho_R(t)) \\ \frac{d\rho_I(t)}{dt} &= q(\rho_R(t) - \rho_I(t)). \end{aligned} \tag{3.3.9}$$

For large N, the drift pushes the trajectory along the solution (3.3.8) until equal subgraph densities are reached [2]. It is clear from (3.3.9), that once equal subgraph densities are reached, then there is no bias in the transition probabilities, so any movement about the diagonal $\rho_r = \rho_I$, is due to diffusive fluctuations. This characterizes the second part of the trajectory as diffusive fluctuations about the diagonal push the system to consensus.

The analysis provided by the fluid limit yields a few takeaways. First, we get a macroscopic view of the trajectories. Another takeaway is that a particularly useful Martingale $Y(t) = p\rho_I(t) + q\rho_R(t)$, if not already seen, has made itself evident by the presence of $p\rho_I(0) + q\rho_R(0)$ on the diagonal where the drift vanishes. In the next two sections we find that $Y(t)$ aids in the computation of the exit probabilities and the expected time to consensus using Doob's optional stopping theorem.

3.4 Exit Probabilities

Let $\mathcal{E}(u_N(t), u_N^*(t))$ be the exit probability of reaching consensus on opinion 1 given the system at time t.

Proposition 3.4.1. *The exit probability $\mathcal{E}(u_N(t), u_N^*(t))$ is determined the densities ρ_R, ρ_I where*

$$\mathcal{E}(u_N(0), u_N^*(0)) = p\rho_I(0) + q\rho_R(0) \tag{3.4.1}$$

Proof. Consider the random variable $Y_t = p\rho_I(t) + q\rho_R(t)$ where $\rho_I(t) = \frac{U^*(t)}{I}$ and $\rho_R(t) = \frac{U(t)}{R}$ are the density of regular and influencer voters of opinion 1 at time t. To determine the exit probabilities we show that Y_t is a Martingale and use Doob's optional stopping theorem.

To show that Y_t is a martingale, we first observe that $0 \leq Y_t \leq 1$, so that $\mathbb{E}(|Y_t|) < \infty$. Furthermore, recalling the system dynamics given in (3.2.1), we have the difference, Δ, of the transition rates in each coordinate as

[2]In fact, for Markov processes that have bounded transition rates, or unbounded transition rates with some additional conditions, Darling [8] gives both L^2 and exponential probabilistic bounds on how far the trajectories can deviate from the solution to the system of ordinary differential equations with convergence in probability as $N \uparrow \infty$.

$$\Delta U(t) = \left((R-U)\left(p\frac{U^*}{I} + q\frac{U}{R}\right) - \left[U\left(p\frac{I-U^*}{I} + q\frac{R-U}{R}\right)\right]\right)$$
$$= \left[\frac{pRU^*}{I} - pU\right]$$
$$= pR\left[\frac{U^*}{I} - \frac{U}{R}\right].$$

and

$$\Delta U^*(t) = \left((I-U^*)\left(p\frac{U^*}{I} + q\frac{U}{R}\right) - \left[U^*\left(p\frac{I-U^*}{I} + q\frac{R-U}{R}\right)\right]\right)$$
$$= \left[\frac{pRU^*}{I} - pU\right]$$
$$= qI\left[\frac{U}{R} - \frac{U^*}{I}\right].$$

Thus the difference in transition rates for Y_t is given by

$$\begin{aligned}\Delta Y_t &= p\Delta\rho_I + q\Delta\rho_R \\ &= p\frac{\Delta U^*(t)}{I} + q\frac{\Delta U(t)}{R} \\ &= pq\left[\frac{U}{R} - \frac{U^*}{I}\right] + qp\left[\frac{U^*}{I} - \frac{U}{R}\right] \\ &= 0\end{aligned} \tag{3.4.2}$$

and from (3.4.2), we can conclude that

$$\mathbb{E}\left(Y_t|\{(U(T), U^*(T), T \le s)\}\right) = Y_s.$$

Therefore Y_t is a martingale with respect to $(U(t), U^*(t))$. Consider now the random variable τ where

$$\tau = \inf\{t \ge 0 : (U(t), U^*(t)) = (0,0) \wedge (U(t), U^*(t)) = (R, I)\}.$$

Since $N < \infty$, it follows that $E[\tau] < \infty$ a.s., so that τ is a stopping time with respect to the natural filtration. By Doob's optional stopping theorem it follows that

$$E[Y_\tau] = E[Y_0] = p\rho_I(0) + q\rho_R(0).$$

Hence,

$$E[Y_\tau] = \mathbb{P}((U(\tau), U^*(\tau)) = (R, I)) \cdot (p+q) + \mathbb{P}((U(\tau), U^*(\tau)) = (0,0)) \cdot 0$$

and therefore

$$\mathbb{P}((U(\tau), U^*(\tau)) = (R, I)) = p\rho_I(0) + q\rho_R(0)$$

and

$$\mathbb{P}((U(\tau), U^*(\tau)) = (0,0)) = 1 - p\rho_I(0) - q\rho_R(0).$$

□

3.5 Time to Consensus

3.5.1 Bound on Time to Consensus

Consider the Martingale $Y(t) = p\frac{u^*(t)}{I} + q\frac{u(t)}{R}$ which we used to derive the time to consensus. To get a bound on the time to consensus we use the fact that for any martingale $M(t)$ we get that $M^2(t) - \langle M(t) \rangle$ is also a Martingale. Here, we have explicitly that,

$$\langle Y(t) \rangle = \int_0^t \mathcal{L}Y^2(s) - 2Y(s)\mathcal{L}Y(s)ds.$$

Since Y_t is a martingale it follows that $\mathcal{L}Y(s) = 0$, so that

$$\langle Y(t) \rangle = \int_0^t \mathcal{L}Y^2(s)ds$$

and therefore

$$Y^2(t) - \int_0^s \mathcal{L}Y^2(s)ds$$

is a martingale. Observe,

$$\begin{aligned} \mathcal{L}Y^2(s) &= \left(\frac{2pq}{IR} + \frac{q^2}{R^2}\right)(R - U(s))Y(s) + \left(-\frac{2pq}{IR} - \frac{q^2}{R^2}\right)U(s)(1 - Y(s)) \\ &\quad + \left(\frac{2pq}{IR} + \frac{p^2}{R^2}\right)(I - U^*(s))Y(s) + \left(\frac{2pq}{IR} + \frac{p^2}{I^2}\right)U^*(s)(1 - Y(s)) \\ &\leq \frac{C}{N} \end{aligned} \tag{3.5.1}$$

for some constant C, as $0 \leq Y(s) \leq 1$, $0 \leq U \leq R, 0 \leq U^* \leq I$ and $I, R \sim O(N)$. Let τ be the time to consensus, which is almost surely finite, so it is a stopping time. Since $Y^2(t) - \langle Y(t) \rangle$ is a Martingale we have that

$$\mathbb{E}[Y_0^2] = \mathbb{E}[Y_\tau^2] = \mathbb{E}\left[\int_0^\tau \mathcal{L}Y^2(s)ds\right].$$

Hence, using the bound from (3.5.1),

$$\begin{aligned}\mathbb{E}[Y_0^2] &= \mathbb{E}\left[\int_0^\tau \mathcal{L}Y^2(s)ds\right] \\ &\leq \mathbb{E}\left[\int_0^\tau \frac{C}{N}ds\right] \\ &= \frac{C}{N}\cdot\mathbb{E}[\tau]\end{aligned}$$

and therefore

$$\mathbb{E}[\tau] \geq \frac{N}{C}\mathbb{E}[Y_0^2]. \tag{3.5.2}$$

From (3.5.2) we see that the expected time to consensus is $O(N)$.

3.5.2 A Heuristic for Time to Consensus

We present a heuristic method that is shown by Sood and Redner [23] to determine the order of the time to consensus via the backward Kolmogorov equation. Recall the transition probabilities in (3.2.1). Rewriting in terms of the densities ρ_R and ρ_I, we define

$$\begin{aligned}\mathbf{R}_R(\rho_R) &= \mathbb{P}[\rho_R \to \rho_R + \delta\rho_R] = \frac{R}{N}(1-\rho_R)[p\rho_I + q\rho_R] \\ \mathbf{L}_R(\rho_R) &= \mathbb{P}[\rho_R \to \rho_R - \delta\rho_R] = \frac{R}{N}\rho_R[p(1-\rho_I) + q(1-\rho_R)] \\ \mathbf{R}_I(\rho_I) &= \mathbb{P}[\rho_I \to \rho_I + \delta\rho_I] = \frac{I}{N}(1-\rho_I)[p\rho_I + q\rho_R] \\ \mathbf{L}_I(\rho_I) &= \mathbb{P}[\rho_I \to \rho_I - \delta\rho_I] = \frac{I}{N}\rho_I[p(1-\rho_I) + q(1-\rho_R)]\end{aligned} \tag{3.5.3}$$

Let $T_N(\rho_R, \rho_I)$ be the mean time to consensus. Then from first step analysis [20, Ch.5] we have

$$\begin{aligned}T_N(\rho_R,\rho_I) = E[\delta t] &+ \mathbf{R}_R(\rho_R)T(\rho_R+\delta\rho_R,\rho_I) + \mathbf{L}_R(\rho_R)T(\rho_R-\delta\rho_R,\rho_I) \\ &+ \mathbf{R}_I(\rho_I)T(\rho_R,\rho_I+\delta\rho_I) + \mathbf{L}_I(\rho_I)T(\rho_R,\rho_I-\delta\rho_I) \\ &+ [1-\mathbf{R}_R(\rho_R)-\mathbf{L}_R(\rho_R)-\mathbf{R}_I(\rho_I)-\mathbf{L}_I(\rho_I)]\,T(\rho_R,\rho_I)\end{aligned} \tag{3.5.4}$$

Expanding (3.5.4) to second order in ρ_I, ρ_R we get the backwards Kolmogorov equation

$$
\begin{aligned}
-E[\delta t] = \mathbf{R}_R(\rho_R) & \left[\delta\rho_R \frac{\partial}{\partial \rho_R} T(\rho_R, \rho_I) + \frac{1}{2}(\delta\rho_R)^2 \frac{\partial^2}{\partial {\rho_R}^2} T(\rho_R, \rho_I)\right] \\
& + \mathbf{L}_R(\rho_R) \left[-\delta\rho_R \frac{\partial}{\partial \rho_R} T(\rho_R, \rho_I) + \frac{1}{2}(\delta\rho_R)^2 \frac{\partial^2}{\partial {\rho_R}^2} T(\rho_R, \rho_I)\right] \\
& + \mathbf{R}_I(\rho_I) \left[\delta\rho_I \frac{\partial}{\partial \rho_I} T(\rho_R, \rho_I) + \frac{1}{2}(\delta\rho_I)^2 \frac{\partial^2}{\partial {\rho_I}^2} T(\rho_R, \rho_I)\right] \\
& + \mathbf{L}_I(\rho_R) \left[-\delta\rho_I \frac{\partial}{\partial \rho_I} T(\rho_R, \rho_I) + \frac{1}{2}(\delta\rho_I)^2 \frac{\partial^2}{\partial {\rho_I}^2} T(\rho_R, \rho_I)\right].
\end{aligned}
$$

Since $E[\delta t] = \frac{1}{N}$, $\delta\rho_R = \frac{1}{R}$, and $\delta\rho_I = \frac{1}{I}$ we get

$$
\begin{aligned}
-1 = & \; p(\rho_I - \rho_R) \frac{\partial}{\partial \rho_R} T(\rho_R, \rho_I) + q(\rho_R - \rho_I) \frac{\partial}{\partial \rho_I} T(\rho_R, \rho_I) \\
& + \frac{1}{2R} \left(p(\rho_I - 2\rho_r\rho_I + \rho_R) + q(2\rho_R - 2{\rho_R}^2)\right) \frac{\partial^2}{\partial {\rho_R}^2} T(\rho_R, \rho_I) \\
& + \frac{1}{2I} \left(p(2\rho_I - 2{\rho_I}^2) + q(\rho_R - 2\rho_R\rho_I + \rho_I)\right) \frac{\partial^2}{\partial {\rho_I}^2} T(\rho_R, \rho_I).
\end{aligned} \tag{3.5.5}
$$

Since the subgraph densities ρ_R, ρ_I asymptotically approach each other, being pushed to the diagonal from the bias in the transition probabilities, we can set $\omega = \frac{\rho_R+\rho_I}{2}$ and thus $\frac{\partial}{\partial \rho_R} = \frac{\partial}{\partial \rho_I} = \frac{1}{2}\frac{\partial}{\partial \omega}$ to get

$$
\begin{aligned}
-1 = & \; \frac{1}{2R} \left(p(\omega(1-\omega)) + q(\omega(1-\omega))\right) \frac{1}{4} \frac{\partial^2}{\partial \omega^2} T(\omega) \\
& + \frac{1}{2I} \left(p(\omega(1-\omega)) + q(\omega(1-\omega))\right) \frac{1}{4} \frac{\partial^2}{\partial \omega^2} T(\omega)
\end{aligned}
$$

and therefore we arrive at the ordinary differential equation

$$
\frac{-4RI}{R+I} = (\omega(1-\omega)) \frac{\partial^2}{\partial \omega^2} T(\omega). \tag{3.5.6}
$$

Using the boundary conditions $T(0) = T(1) = 0$ we get the solution to (3.5.6) given by

$$
T_N(\omega) = \frac{4RI}{N} \left((1-\omega)\log\left(\frac{1}{1-\omega}\right) + \omega \log\left(\frac{1}{\omega}\right)\right) \tag{3.5.7}
$$

Interestingly, this is exactly the solution to the time to consensus of the complete bipartite graph as seen in [23]. The commonality of the two models being the role of the diffusion terms and the vanishing of the drift terms once the equal subdensities are reached. From the solution to the ODE in, notice that we can rewrite (3.5.7) in terms of the initial conditions $\rho_R(0)$ and $\rho_I(0)$ as $\omega = p\rho_R(0) + q\rho_I(0)$. Thus the total time to consensus for Influencer voter model is

$$T_N(\omega) = \frac{4RI}{N}\left((1-\omega)\log\left(\frac{1}{1-\omega}\right) + \omega\log\left(\frac{1}{\omega}\right)\right) + T'(\rho_R(0), \rho_I(0)) \qquad (3.5.8)$$

where $T'(\rho_R, \rho_I) \sim O(1)$ is the mean time to reach equal subgraph densities, $\omega = p\rho_R(0) + q\rho_I(0)$. Notice we may verify that we get the same result for the classical voter model on the complete graph by setting $I = R = N/2$, $p = q$, and $\rho = \rho_I(0) = \rho_R(0)$. Substituting these values into (3.5.8), we get

$$T_N(\rho) = N\left[(1-\rho)\log\left(\frac{1}{1-\rho}\right) + \rho\log\left(\frac{1}{\rho}\right)\right]$$

which is the time to consensus for the voter model as discussed in chapter 1. We can then see that p and q play their role in the time to consensus in the initial $O(1)$ time approach to the diagonal of initial densities and in terms of the constant $\left((1-\omega)\log\left(\frac{1}{1-\omega}\right) + \omega\log\left(\frac{1}{\omega}\right)\right)$.

3.6 Numerical Results

In this section we plot simulated times to consensus and compare to the analytical expected time for the voter and influencer voter model.

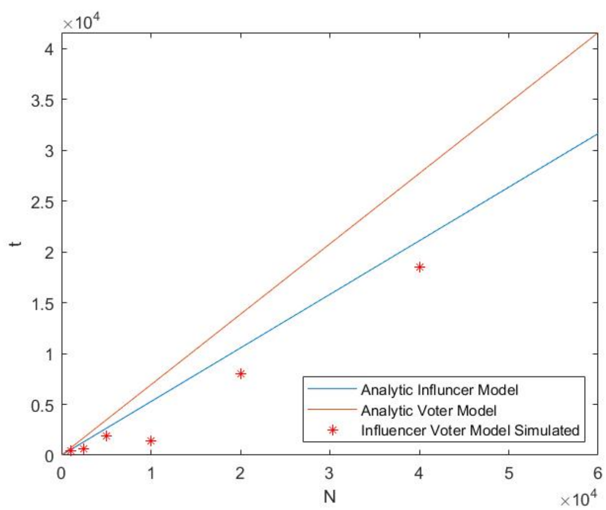


Figure 3.4: Convergence time in terms of the number of nodes N. We have chosen the following parameters $p = .75, I = .3N, R = .7N, \rho_I(0) = 2/3$, and $\rho_R(0) = 5/7$.

Fig. 3.6 shows the convergence time for the voter model and influencer voter model. In particular, the simulations confirm that the expected time to consensus is linear in N.

3.7 Further Considerations

One of the motivating papers for the influencer voter model was [19] in which they found that by including an additional voter undecided state, with certain dynamics, is desirable in solving the binary consensus problem. The binary consensus problem is where, given a network where each node initially observes one of the states, 0 or 1, how to construct a robust, distributed protocol which ensures that the nodes reach the right consensus on the initial majority observation. They find that the probability of reaching consensus on the non-initial majority decays exponentially with the number of nodes N. Furthermore, they found that the convergence time is logarithmic in N. In Fig. 3.5 we see how incorporating the undecided voter makes the fluid limit behave quite nicely as the drift pushes to consensus on the initial majority. In the case where the number of opinions of 0 and 1 are the same, we see that the drift pushes the system to equal subgraph densities between the three opinion states.

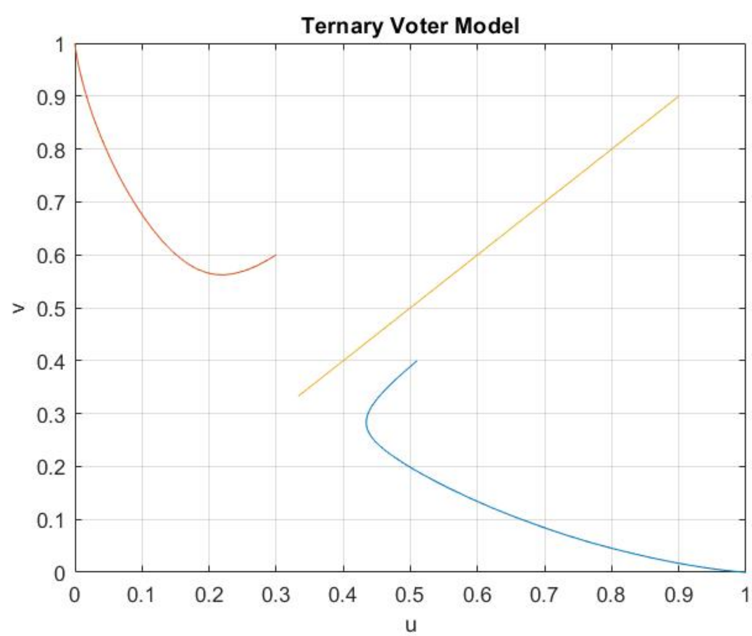


Figure 3.5: Numerically solved fluid limits for the ternary voter model. The three paths here in blue, yellow, and red exhibit the three general paths. For blue the initial conditions are $(u(0), v(0)) = (0.51, 0.4)$, for yellow $(u(0), v(0)) = (0.9, 0.9)$ and for red $(u(0), v(0)) = (0.3, 0.6)$.

In terms of the influencer voter model and the binary consensus problem, we can take the interpretation of the influencers as nodes that are "trusted" as they have been identified to be more likely to have the "correct" initial opinion. Future work would be interested in exploring the potential speed up and the probability of reaching consensus by identifying such trusted nodes.

Another route for future work would be considering the influencer voter model on heterogeneous networks as Redner and Sood [22] did for the voter model.

References

[1] Siva Athreya, Sunder Sethuraman, and Bálint Tóth, *On the range, local times and periodicity of random walk on an interval*, arXiv preprint arXiv:1009.3999 (2010).

[2] Andrea Baronchelli, Claudio Castellano, and Romualdo Pastor-Satorras, *Voter models on weighted networks*, Physical Review E **83** (2011), no. 6, 066117.

[3] Richard F Bass, Xia Chen, and Jay Rosen, *Moderate deviations for the range of planar random walks*, American Mathematical Soc., 2009.

[4] D Bhat and S Redner, *Reputation-driven voting dynamics*, Journal of Statistical Mechanics: Theory and Experiment **2019** (2019), no. 6, 063208.

[5] Patrick Billingsley, *Convergence of probability measures*, John Wiley & Sons, 2013.

[6] Peter Clifford and Aidan Sudbury, *A model for spatial conflict*, Biometrika **60** (1973), no. 3, 581–588.

[7] J Theodore Cox, Richard Durrett, and Edwin A Perkins, *Rescaled voter models converge to super-brownian motion*, The Annals of Probability **28** (2000), no. 1, 185–234.

[8] Richard WR Darling and James R Norris, *Differential equation approximations for markov chains*, Probability surveys **5** (2008), 37–79.

[9] Thomas Doehrman, Sunder Sethuraman, and Shankar C Venkataramani, *Remarks on the range and multiple range of a random walk up to the time of exit*, Rocky Mountain Journal of Mathematics **51** (2021), no. 5, 1603–1614.

[10] Richard Durrett, *Lecture notes on particle systems and percolation*, Wadsworth Publishing Company, 1988.

[11] Stewart N Ethier and Thomas G Kurtz, *Markov processes: characterization and convergence*, John Wiley & Sons, 2009.

[12] Mario Ferraro and Lorenzo Zaninetti, *Statistics of visits to sites in random walks*, Physica A: Statistical Mechanics and its Applications **338** (2004), no. 3-4, 307–318.

[13] Richard A Holley and Thomas M Liggett, *Ergodic theorems for weakly interacting infinite systems and the voter model*, The annals of probability (1975), 643–663.

[14] Antoine Jego, *Characterisation of planar brownian multiplicative chaos*, arXiv preprint arXiv:1909.05067 (2019).

[15] ______, *Thick points of random walk and the gaussian free field*, Electronic Journal of Probability **25** (2020), 1–39.

[16] Claude Kipnis and Claudio Landim, *Scaling limits of interacting particle systems*, vol. 320, Springer Science & Business Media, 1998.

[17] Thomas Milton Liggett and Thomas M Liggett, *Interacting particle systems*, vol. 2, Springer, 1985.

[18] Akira Okubo and Simon A Levin, *Diffusion and ecological problems: modern perspectives*, vol. 14, Springer, 2001.

[19] Etienne Perron, Dinkar Vasudevan, and Milan Vojnovic, *Using three states for binary consensus on complete graphs*, IEEE INFOCOM 2009, IEEE, 2009, pp. 2527–2535.

[20] Nicolas Privault, *Understanding markov chains*, Examples and Applications, Publisher Springer-Verlag Singapore **357** (2013), 358.

[21] Philippe Robert, *Stochastic networks and queues*, vol. 52, Springer Science & Business Media, 2013.

[22] Vishal Sood, Tibor Antal, and Sidney Redner, *Voter models on heterogeneous networks*, Physical Review E **77** (2008), no. 4, 041121.

[23] Vishal Sood and Sidney Redner, *Voter model on heterogeneous graphs*, Physical review letters **94** (2005), no. 17, 178701.

[24] Frank Spitzer, *Principles of random walk*, vol. 34, Springer Science & Business Media, 2001.

[25] Tzai-Hung Wen, Min-Hau Lin, and Chi-Tai Fang, *Population movement and vector-borne disease transmission: differentiating spatial–temporal diffusion patterns of commuting and noncommuting dengue cases*, Annals of the Association of American Geographers **102** (2012), no. 5, 1026–1037.

[26] Ward Whitt, *Stochastic-process limits: an introduction to stochastic-process limits and their application to queues*, Space **500** (2002), 391–426.