

# Stochastic Processes and Applications

Report on Evolutionary Minority  
Games

**Haowei CHEN**

A MSDM5003 Final Project Report



February 18, 2025

# 1 Introduction

Minority Game is a model that is used to mimic characteristics of real-world markets in which the minority group may gain benefits. However, different from the setting in a basic minority game, in the real world, agents with the worst strategies are always eliminated and replaced by new agents. Hence, it would help the model fit the real-world market better if evolutionary features are equipped. In this project, we are going to study some kinds of evolutionary minority games.

# 2 Theory

The settings of a basic minority game are introduced by [1]. In order to simulate the learning process of agents in a minority game, Challet and Zhang extended the basic minority game to include the Darwinist selection: the worst players would be replaced by new players that are clones of the best player after certain time-steps  $\tau$  with a mutation. In this version of EMG (hereinafter referred to as EMG(V1)), The worst players are determined by ranking and poverty level. [1] Lo, Hui, and Johnson introduced another version of the evolutionary minority game, in which the worst players are eliminated if their scores fall down a certain threshold.(hereinafter referred to as EMG(V2)) [4] Combining a fixed threshold with an elimination per certain time-steps, we propose a version of the evolutionary minority game for comparisons. (here-

inafter referred to as EMG(V3)) Li, Riolo, and Savit put forward two versions of the evolutionary minority games based on EMG(V1). In this report, We take the version with a fixed strategy space into comparison. [2] All four versions of evolutionary minority games and the basic minority game are shown in the table below. For detailed model formalizations, please refer to the appendix.

Table 1: Information for Minority Games Studied

Version	Mutation	Threshold	Time to Eliminate	Elimination Rule
Basic	None	None	None	None
EMG(V1)	Best+Random	Dynamic	Fixed Time-steps	Worst
EMG(V2)	Random	Fixed	Anytime	Below Threshold
EMG(V3)	Best+Random	Fixed	Fixed Time-steps	Below Theshold
EMG(V4)	Best+Random	Dynamic	Fixed Time-steps	Poverty Level

### 3 Numerical Results

In this section, we focus on results obtained by numerical simulations and compare four versions of minority games above. For comparisons, we simulate each minority game with the same parameters. They are shown in the table below:

Table 2: Basic Parameters

Parameter	Value
Rounds	4000
Rounds Before Recording	200
Memory Size	5
Sample Size	10

### 3.1 Attendance Number

In this subsection, we compare the attendance numbers. For  $EMG(V1)$  and  $EMG(V4)$ , the poverty level is set as 0.2 and  $\tau = 50$ . For  $EMG(V2)$  and  $EMG(V3)$ , the threshold is set as  $-4$ . The obtained result is shown below:

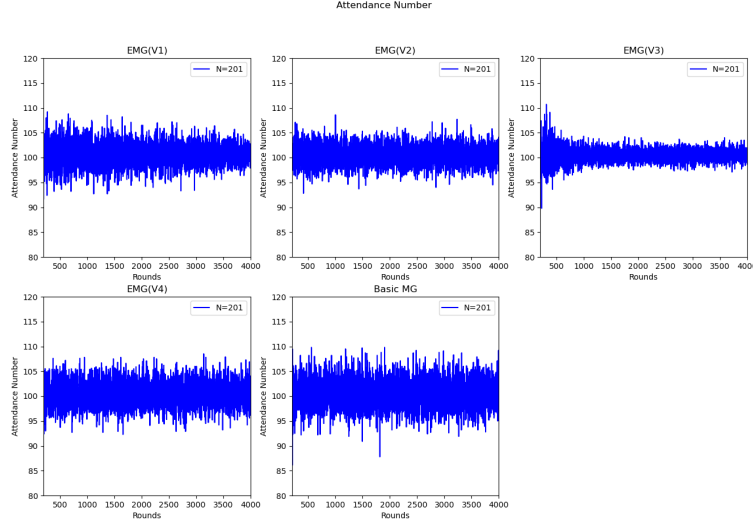


Figure 1: Attendance Number When  $N = 201$

It can be observed that the amplitudes of fluctuations are obviously smaller than the amplitude of the basic minority game. It appears to show that a fixed threshold can control the fluctuations better. For clearer observations, we simulate games for different values of  $N$ :

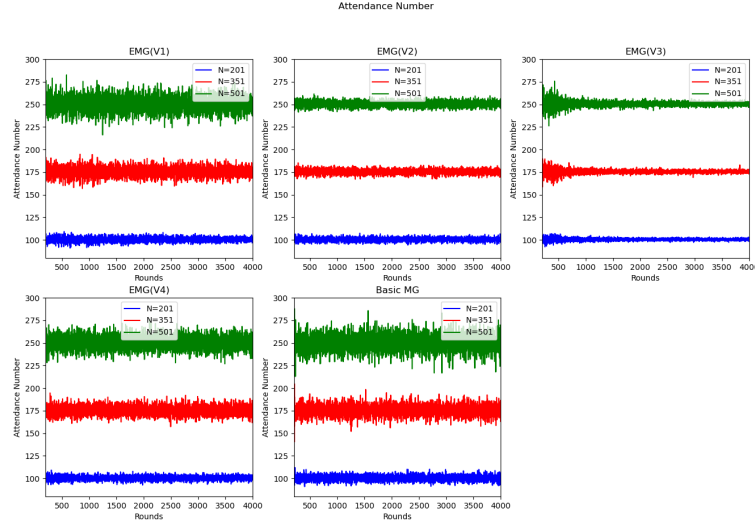


Figure 2: Attendance Number When  $N = 201, 351$  and  $501$

It can be observed that as  $N$  gets larger, the amplitude of the fluctuation is larger. Moreover, different versions of evolutionary minority games all appear to own smaller fluctuations than the basic version. All versions of EMG show a learning process, in which the fluctuations get smaller. As  $N$  gets larger, this learning process tends to become longer. Two versions with a fixed threshold still have smaller fluctuations than other versions, which matches our former observations.

### 3.2 Phase Transition

In this subsection, we plot the figure of phase transitions. The settings of basic parameters are the same as in the former subsection. The obtained results are shown below:

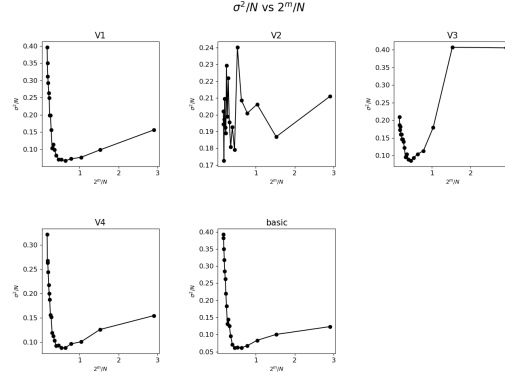


Figure 3:  $\frac{\sigma^2}{N}$  Versus  $\frac{2^m}{N}$

Similar phase transitions can be observed in both EMG and the basic minority game except EMG(V2). The figure for this model is somehow strange. Taking the different strategy space in EMG(V2) into consideration, this phenomenon is not difficult to comprehend.

### 3.3 Performances of Representative Agents

In this section, we select three representative agents (the best, the worst, and the median ones) to compare the overall system performances. We plot the wealth curves of them in the figure below:

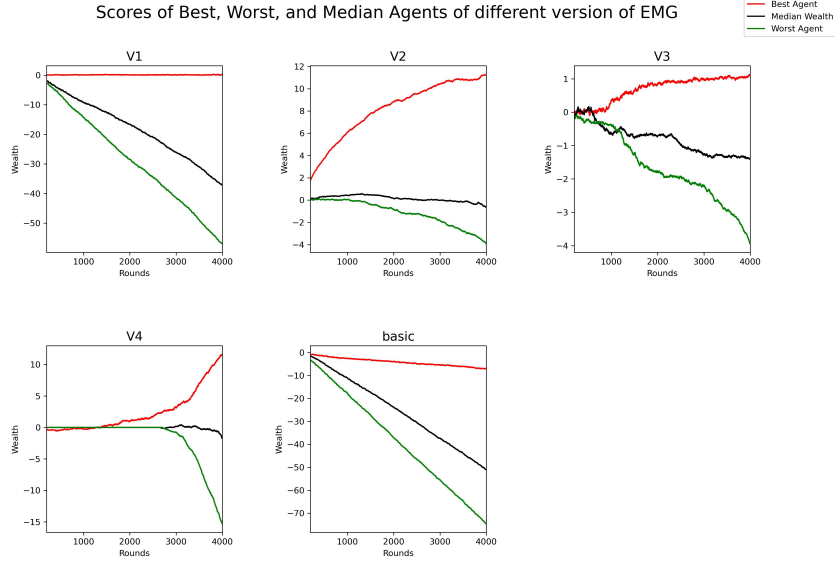


Figure 4: Wealth Curves of Representative Agents

Thanks to the eliminations, all representative agents from EMG perform better than those from the basic minority game. The best agents in EMG(V2) and EMG(V4) turn out to be the wealthiest ones among all representative agents selected from all versions of the minority game. In the meantime, the worst agent from EMG(V1) loses his wealth at a relatively high speed. Generally speaking, the median and the worst agents in the versions of EMG with fixed thresholds appear to be wealthier than those from EMGs with dynamic thresholds, since poorer agents are eliminated. Both with dynamic thresholds, the best agent in EMG(V4) earns more money than the one in EMG(V1). This is because, in EMG(V4), new agents don't clone the best agents' strategies. Instead, they randomly choose their initial strategies. This rule helps the best agents to keep advantages.

## 4 Conclusion

In the main text of this report, we focus on the overall performance of each system. According to our simulation results, evolutionary populations can significantly control the fluctuation and improve the overall performance on the basis of keeping the main characteristics of the minority game. By comparing different versions of EMG, we can find that a fixed threshold shows a better effect. For analysis and observations of single agents, we would refer you to the appendix.

## References

- [1] D. Challet & Y.-C. Zhang, Mergence of cooperation and organization in an evolutionary game, *Physica A*, **246**, 407-418, 1997.
- [2] Y. Li, R. Riolo, & R. Savit, Evolution in Minority Games(I): Games With a Fixed Strategy Space, *Physica A*, **276**, 234-264, 2000.
- [3] S. Sinha, A. Chatterjee, A. Chakraborti, & B. K. Chakrabarti, *Econophysics: An Introduction*, Wiley, 2011
- [4] T.S. Lo, P. M. Hui & N. F. Johnson, Theory of the evolutionary minority game, *Phys. Rev. E*, **62**(3), 4393-4396, 2000.
- [5] R. D. Hulst & G. J. Rodgers, The Hamming Distance in the Minority Game, *Physica A*, **270**, 514-525, 1999.



## Appendix

Since the main parameters are discussed and compared in the main text, we focus on some unique features of three versions of Evolutionary Minority Games (V1, V2 & V4) in this appendix. The plan of this appendix is as follows. In the first two subsections, we introduce parameters of numerical simulations and notations that will be used later. Then we discuss the unique features of each version of EMG alternatively.

### 1 Basic Parameters

For all numerical simulations in this appendix except for those with special instructions, values of basic parameters are shown in the table below:

Table 3: Basic Parameters	
Parameter	Value
Rounds(ite)	40000
Rounds Before Recording(iteq)	200
Memory Size(m)	5
Sample Size(nsam)	10
Number of Agents(N)	201
Way of Updating Scores	Linear Score

### 2 Notations

- $u_m$ : A string which stores the results of the former  $m$  rounds;
- $T_i^{[j]}(u_m)$ : The response to the string  $u_m$ ;
- $\Phi_i(j)$ : The probability that agent  $i$  uses strategy  $j$ ;
- $P(u_m)$ : the probability that the  $m$ -string  $u_m$  appears in the sequence of minority groups;
- $D_h$ : Hamming distance, the number of positions at which the corresponding symbols are different;
- $D_b$ : The average behavioral distance of the  $i_{th}$  agent from all other agents playing the game.  $D_b(i) = \sum_j \sum_{k,l} \sum_{u_m} P(u_m) \Phi_i(k) \Phi_j(l) |T_i^{[k]}(u_m) - T_j^{[l]}(u_m)|$ ;
- $F_N(n)$ : The probability of  $n$  agents choose to attend in a certain round of game; ( $N$  agents in total)

- $p_k$ : In a certain round of the game, the probability for agent  $k$  to choose to attend;
- $G_{N-1}^k(n-1)$ : The probability of the event in which  $n-1$  agents choose to attend,  $N-1$  agents in total except agent  $k$ ;
- $P(p)$ : The normalized frequency of gene value  $p$ . (probability distribution of gene values) we also use  $P(p, t)$  to denote the probability distribution of gene values  $p$  at time  $t$ ;
- $A(t)$ : Attendance number, the number of agents who choose to attend at the time point  $t$ ;
- $\tau(p_k)$ : The winning probability of agent  $k$  at the time his gene value is  $p_k$ ;
- $L(p)$ : The lifespan of the gene value  $p$ , which means the average living time for  $p$  between two modifications. We also use  $L(p, t)$  to denote the living time for the gene value  $p$  at time  $t$ ;
- $S(p, t)$ : The probability that an agent whose gene value is  $p$  would change his gene value at time  $t$ ;
- $D$ : The threshold for eliminating an agent. It is required that  $D < 0$ ;
- $R$ : The width of range while changing strategies.

### 3 EMG(V1)

In this section, we study the mutation setting and properties of individual agents by computing two distance measures in EMG(V1) put forward by Challet and Zhang. [1]

#### 3.1 Mutation

It is a very important setting in this version of EMG model to allow the new agent to change one of the strategies of the best agent, instead of simply cloning him. This is called 'mutation'. In order to observe the impact of mutations, we simulated and plotted the figure of the attendance number in both cases with and without mutations.

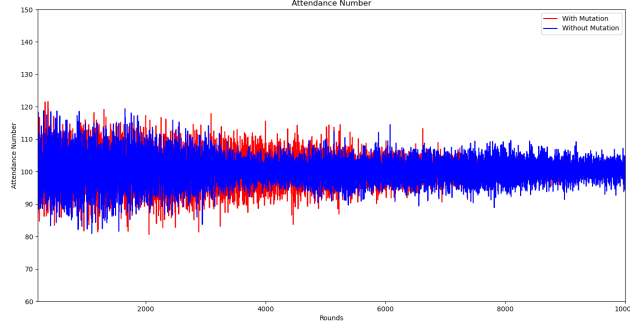


Figure 5: Attendance Number Comparison

As is shown in the figure above, as time goes by, mutations can reduce the fluctuations significantly. This tells us the importance of diversity in such an agent-based model.

### 3.2 Hamming Distance $D_h$

In this subsection, we study the Hamming distance between the two strategies of each agent. This intra-agent distance measures the difference between strategies held by the same agent, which helps us understand how agents' strategies change in this model.

The obtained figure of  $D_h$  versus the wealth is as shown below:

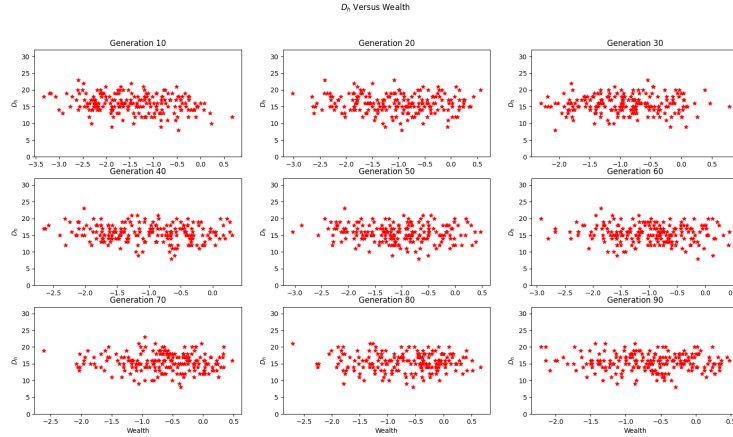


Figure 6: Hamming Distance Between Strategies of Each Agent

Notice that all data points are concentrated near  $D_h = 16$ , which is half the

number of all possible memory strings. Moreover, the overall position of the data points is moving down slowly, somehow faster at the beginning of the game. This observation reveals the fact that agents tend to show relative caution in this game. They would not choose too similar or totally different strategies. Instead, holding two strategies with  $D_h$  about 16 shows their preferences in gaining a balance.

### 3.3 Average Behavioral Distance $D_b$

In this subsection, we study the average behavioral distance  $D_b$ . This parameter measures how different agents would behave according to their strategies. The obtained results are shown below:

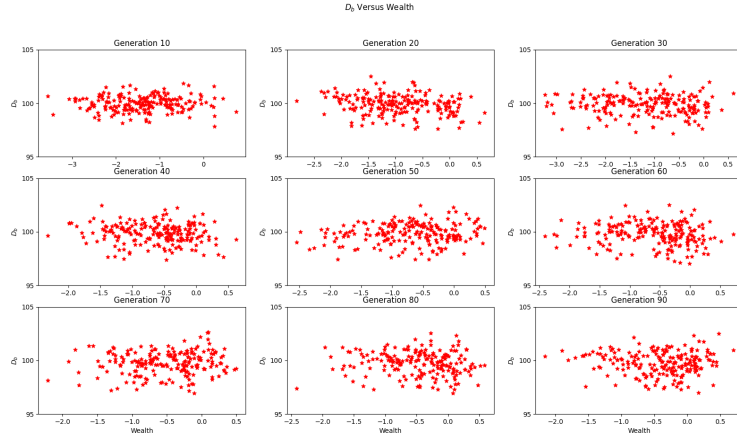


Figure 7:  $D_b$  Versus Wealth

As is shown in the figure above, the distribution of  $D_b$  is similar to the distribution of  $D_h$  before. However, the trend of the drop in  $\langle D_b \rangle$  is not obvious. All data points are concentrated near  $D_b = 100$ .

## 4 EMG(V2)

In this section, we study two unique features of EMG(V2) model put forward by Lo, Hui, & Johnson: the gene value distribution and the lifespan. [4]

## 4.1 Theoretical Derivations

### 4.1.1 Attendance Number

According to the Central Limit Theorem, after a sufficiently long time,  $A(t)$  follows an approximately normal distribution with a mean  $N\bar{p} = N \int_0^1 pP(p)dp$  and a variance  $N \int_0^1 p(1-p)P(p)dp$ .

### 4.1.2 Winning Probability

By definition, for  $F_N(n)$  we have:

$$F_N(n) = p_k G_{N-1}^k(n-1) + (1-p_k) G_{N-1}^k(n) \quad (1)$$

For  $\tau(p_k)$ , we have the following equality according to the definition:

$$\tau(p_k) = p_k \sum_{n=0}^{(N-3)/2} G_{N-1}^k(n) + (1-p_k) \sum_{n=(N+1)/2}^{N-1} G_{N-1}^k(n) \quad (2)$$

By summing up the first  $(N-3)/2$  equalities of 1, we obtain the formula as below:

$$\begin{aligned} \sum_{n=1}^{(N-3)/2} F_N(n) &= p_k \sum_{n=1}^{(N-3)/2} (G_{N-1}^k(n-1) - G_{N-1}^k(n)) + \sum_{n=1}^{(N-3)/2} G_{N-1}^k(n) \\ &= p_k (G_{N-1}^k(0) - G_{N-1}^k(\frac{N-3}{2})) + \sum_{n=1}^{(N-3)/2} G_{N-1}^k(n) \\ &= \sum_{n=1}^{(N-3)/2} G_{N-1}^k(n) + p_k G_{N-1}^k(0) - p_k G_{N-1}^k(\frac{N-3}{2}) \end{aligned} \quad (3)$$

Similarly, by summing up the equalities 1 from  $n = (N+1)/2$  to  $n = N-1$ , the following equality is obtained:

$$\begin{aligned} \sum_{n=(N+1)/2}^{N-1} F_N(n) &= p_k \sum_{n=(N+1)/2}^{N-1} G_{N-1}^k(n-1) + (1-p_k) \sum_{n=(N+1)/2}^{N-1} G_{N-1}^k(n) \\ &= p_k \sum_{n=(N+1)/2}^{N-1} (G_{N-1}^k(n-1) - G_{N-1}^k(n)) + \sum_{n=(N+1)/2}^{N-1} G_{N-1}^k(n) \\ &= \sum_{n=(N+1)/2}^{N-1} G_{N-1}^k(n) + p_k G_{N-1}^k(\frac{N-1}{2}) - p_k G_{N-1}^k(N-1) \end{aligned} \quad (4)$$

For the boundary cases, in which  $n = 1$  or  $n = 0$ , it's not difficult to get the

following two equalities:

$$\begin{cases} F_N(0) = (1 - p_k)G_{N-1}^k(0) \\ F_N(N) = p_k G_{N-1}^k(N-1) \end{cases} \quad (5)$$

Hence, the equalities 3 and 4 can be expanded to the following versions:

$$\begin{cases} \sum_{n=0}^{(N-3)/2} F_N(n) = \sum_{n=0}^{(N-3)/2} G_{N-1}^k(n) - p_k G_{N-1}^k(\frac{N-3}{2}) \\ \sum_{n=(N+1)/2}^N F_N(n) = \sum_{n=(N+1)/2}^{N-1} G_{(N+1)/2}^{N-1} G_{N-1}^k(n) + p_k G_{N-1}^k(\frac{N-1}{2}) \end{cases} \quad (6)$$

By substituting the above two equalities into the expression of  $\tau(p_k)$ , we then obtain:

$$\begin{aligned} \tau(p_k) &= p_k \left[ \sum_{n=0}^{(N-3)/2} F_N(n) + p_k G_{N-1}^k(\frac{N-3}{2}) \right] + (1 - p_k) \left[ \sum_{n=(N+1)/2}^N F_N(n) - p_k G_{N-1}^k(\frac{N-1}{2}) \right] \\ &= p_k \left[ \sum_{n=0}^{(N-1)/2} F_N(n) - (1 - p_k) G_{N-1}^k(\frac{N-1}{2}) \right] + (1 - p_k) \left[ \sum_{n=(N+1)/2}^N F_N(n) - p_k G_{N-1}^k(\frac{N-1}{2}) \right] \\ &= p_k \sum_{n=0}^{(N-1)/2} F_N(n) + (1 - p_k) \sum_{n=(N+1)/2}^N F_N(n) - 2p_k(1 - p_k) G_{N-1}^k(\frac{N-1}{2}) \end{aligned} \quad (7)$$

From 1, we have:

$$\begin{cases} F_N(n) = p_k G_{N-1}^k(n-1) + (1 - p_k) G_{N-1}^k(n) \\ F_N(n-1) = p_k G_{N-1}^k(n-2) + (1 - p_k) G_{N-1}^k(n-1) \end{cases} \quad (8)$$

By multiplying the first formula with  $1 - p_k$  and the second formula with  $p_k$ , we can eliminate  $G_{N-1}^k(n-1)$  item:

$$(1 - p_k)F_N(n) - p_k F_N(n-1) = (1 - p_k)^2 G_{N-1}^k(n) - p_k^2 G_{N-1}^k(n-2) \quad (9)$$

Repeat this step, we can obtain:

$$\sum_{j=0}^n (-1)^{n-j} F_N(j) \left( \frac{p_k}{1 - p_k} \right)^{n-j} = (1 - p_k) G_{N-1}^k(n) \quad (10)$$

By substituting 10 into 7, we can express  $\tau(p_k)$  entirely with  $F_N(k)$ :

$$\tau(p_k) = p_k \sum_{n=0}^{(N-1)/2} F_N(n) + (1-p_k) \sum_{n=(N+1)/2}^N F_N(n) - 2p_k \sum_{n=0}^{(N-1)/2} (-1)^{\frac{N-1}{2}-n} F_N(n) \left(\frac{p_k}{1-p_k}\right)^{\frac{N-1}{2}-n} \quad (11)$$

In fact, the winning probability in each round is independent of which agent we discuss about. It only depends on the gene values. Hence, we shall just change the notation as below:

$$\tau(p) = p \sum_{n=0}^{(N-1)/2} F_N(n) + (1-p) \sum_{n=(N+1)/2}^N F_N(n) - 2p \sum_{n=0}^{(N-1)/2} (-1)^{\frac{N-1}{2}-n} F_N(n) \left(\frac{p}{1-p}\right)^{\frac{N-1}{2}-n} \quad (12)$$

where  $p$  is the gene value. In the steady state, the winning rounds of 0 and 1 are the same on average. Hence,  $p \sum_{n=0}^{(N-1)/2} F_N(n) + (1-p) \sum_{n=(N+1)/2}^N F_N(n)$  equals  $\frac{1}{2}$  in the steady state. In this case, we have

$$\tau(p) = \frac{1}{2} - 2p \sum_{n=0}^{(N-1)/2} (-1)^{\frac{N-1}{2}-n} F_N(n) \left(\frac{p}{1-p}\right)^{\frac{N-1}{2}-n} \quad (13)$$

#### 4.1.3 Gene Value Distribution and Lifespan

The time evolution of  $P(p)$  is given by the following equation according to the definition of the evolutionary minority game.

$$P(p, t + dt) = P(p, t) - S(p, t)dt + \frac{dt}{2} [S(p + \Delta p, t) + S(p - \Delta p, t)] \quad (14)$$

This can be rewritten as:

$$\frac{\partial P}{\partial t} = \frac{\Delta p^2}{2} \frac{\partial^2 S}{\partial p^2} \quad (15)$$

where  $\Delta(p) = \frac{R}{4}$ .

On average, we have:

$$S(p, t) = \frac{P(p, t)}{L(p, t)} \quad (16)$$

Hence, we can conclude that:

$$\frac{P(p)}{L(p)} = C \quad (17)$$

where  $C$  is a constant.

For agents with a gene value  $p$ , on average he wins  $C(p, t)$  scores per round:

$$C(p, t) = C(p, t_0) - \left(\frac{1}{2} - \tau(p)\right)(t - t_0) = C(p, t_0) - \left(\frac{1}{2} - \tau(p)\right)L(p, t) \quad (18)$$

Hence,

$$L(p, t) = \frac{C(p, t) - C(p, t_0)}{-\frac{1}{2} + \tau(p)} = \frac{D}{\tau(p) - \frac{1}{2}} \quad (19)$$

By substituting 19 into 15, we obtain:

$$\frac{\partial P}{\partial t} = \frac{R^2}{32} \frac{\partial^2}{\partial p^2} \left[ \left( \tau(p) - \frac{1}{2} \right) P(p) \right] \quad (20)$$

By integrating this formula with respect to  $p$ , we obtain the expression of  $P(p)$ :

$$P(p) = \frac{N_0}{1/2 - \tau(p)} \quad (21)$$

where  $N_0$  is constant given by the normalization of  $P(p)$ :  $\int_0^1 P(p) dp = 1$ . Hence,  $L(p)$  can be expressed in the same form with only the constant different.

Note that this result tells us that the behaviours of  $P(p)$  and  $L(p)$  depend on the difference between  $\tau(p)$  and  $\frac{1}{2} - \tau(p)$ . For convenience, we denote this function by  $G(p)$ . By 13, we have:

$$G(p) = 2p \sum_{n=0}^{(N-1)/2} (-1)^{\frac{N-1}{2}-n} F_N(n) \left( \frac{p}{1-p} \right)^{\frac{N-1}{2}-n} \quad (22)$$

## 4.2 Numerical Results

### 4.2.1 Distribution of Gene Values

The obtained simulation results of the distribution of gene values are as shown below:



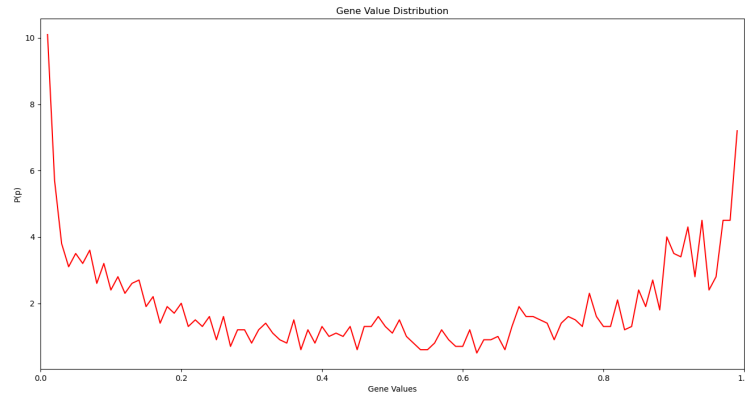


Figure 8: Frequency Distribution of Gene Values

This figure is in good agreement with the theoretical result. It can be observed that  $P(p)$  has peaks around  $p = 0$  and  $p = 1$ , which implies a very interesting fact that in this model agents prefer to perform extreme strategies.

#### 4.2.2 Lifespans

The obtained simulation results of lifespans of gene values are shown in the figure below:

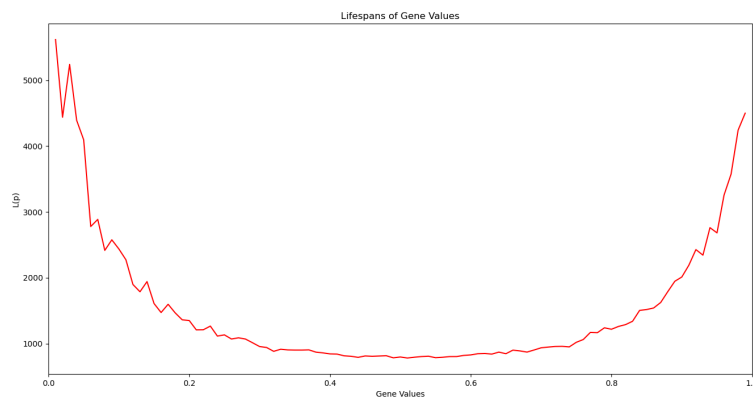


Figure 9: Lifespans of Gene Values

This curve appears to be similar to the curve of gene value distributions, which matches the theoretical results. Gene values around 0 and 1 have longer lifespans. This result reveals the fact that extreme strategies perform better than others in

this model. This also explains the phenomenon that agents prefer extreme strategies we observed before.

## 5 EMG(V4)

In this section, we study two distance measures  $D_h$  and  $D_b$  to comprehend the behaviors of agents in EMG(V4) put forward by Li, Riolo, and Savit. [2]

### 5.1 Hamming Distance $D_h$

In this subsection, we focus on the Hamming distance. We obtain the following figure by numerical simulations:

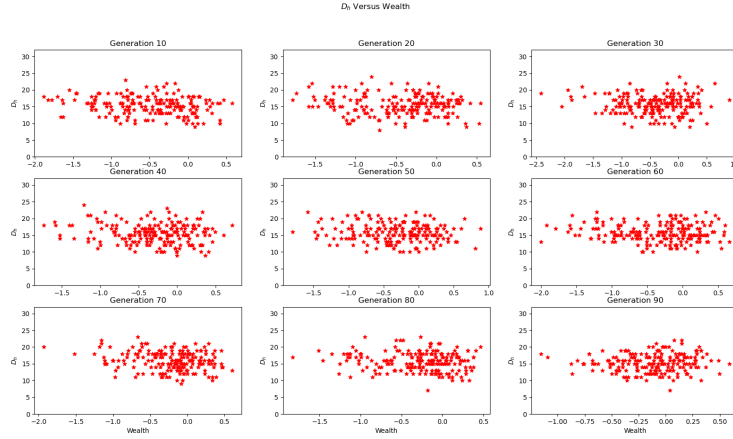
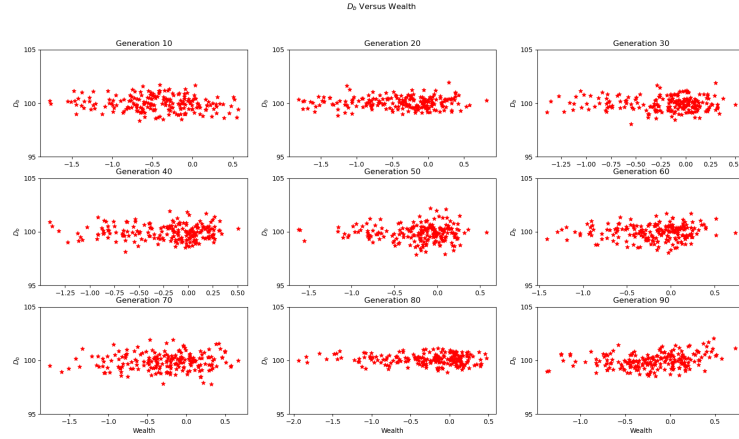


Figure 10: Hamming Distances Versus Wealth

The data points appear to gather as time goes by. Like the phenomenon we observed before in Figure 6, all data points are concentrated near  $D_h = 16$ , which is half the number of all possible memory strings. This shows us the similarity between EMG(V1) and EMG(V4).

### 5.2 Average Behavioral Distance $D_b$

In this subsection, we focus on  $D_b$ . We obtain the following figure by numerical simulations:

Figure 11:  $D_b$  Versus Wealth

Similar to the figure of EMG(V1) we observed before, all data points are concentrated near  $D_b = 100$ , which is half the number of agents in the system. Every time a spread is observed, the distribution has narrowed in the next 10 generations rapidly and considerably. As time goes by, the distribution tends to form a center of gravity on the right side with a tail on the left.