Mean field voter model on networks and multi-variate beta distribution

Shintaro Mori

Department of Mathematics and Physics, Faculty of Science and Technology, Hirosaki University
Bunkyo-cho 3, Hirosaki, Aomori 036-8561, Japan

Masato Hisakado

Nomura Holdings, Inc., Otemachi 2-2-2, Chiyoda-ku, Tokyo 100-8130, Japan

Kazuaki Nakayama

Department of Mathematical Sciences, Faculty of Science, Shinshu University, Asahi 3-1-1, Matsumoto, Nagano 390-8621, Japan

(Dated: October 16, 2018)
Abstract

The regional characteristics of elections show a strong spatial correlation and the logarithmic decay with distance suggests that a 2D noisy diffusive equation describes the system. Based on a study of U.S. presidential elections data, we find that the fluctuations of vote shares also exhibit a strong and long-range spatial correlation. It was previously thought that it was difficult to induce strong and long-ranged spatial correlation of the fluctuations without breaking the unimodality of the distribution. However, we demonstrate that a mean field voter model on networks shows such a behavior. In the model, voters in a node are affected by the agents in the node and by agents in the connected nodes. We derive a multivariate Wright-Fisher diffusion equation for the joint probability density of the vote shares. The stationary distribution is a multivariate generalization of the beta distribution. We estimate the equilibrium values and the covariance of the vote shares and obtain the correspondence with a multivariate normal distribution. This approach greatly simplifies the calibration of the parameters in the modeling of elections.

PACS numbers: 05.70.Fh,89.65.Gh
I. INTRODUCTION

Social physics has become an active research field and many studies have been devoted to attempting to understand social phenomena and interacting human behaviors. Opinion dynamics is a central research theme, and empirical studies based on election data have been extensively performed. In these investigations, the correlation of the voters’ decisions was clarified by studying the dependence of the variance of the turnout rate on the number of voters $N$. If the voters’ decisions are independent, the variance of the turnout rate should be proportional to $N^{-1}$. The empirical study of French election data showed that it was proportional to the power of $N^{-3/4}$. In addition, it was determined that the spatial correlation of the turnout rate in each election exhibited a logarithmic decay with a distance that suggested a description based on a 2D noisy diffusion equation.

On the other hand, the voter model and its noisy extension have been studied extensively in opinion dynamics. In particular, the validity of the voter model as a model for elections was tested in the U.S. presidential election. In this model, agents move between their living places and their workplaces. In both places, their decisions are affected by other voters. The model is called the social influence recurrent mobility model (SIRM model). Based on the diffusion approximation of the model, the noisy diffusion equation was derived. By balancing the strength of the noise with the voter models consensus mechanism or the force of conformity, it was concluded that the SIRM model can reproduce the statistical features of the vote-share in presidential elections, i.e. the stationarity of the variance of vote-share distributions and the long-range spatial correlation which decays logarithmically with distance. However, the model has a drawback in that under certain circumstances, the noise might break the range of vote shares. This was addressed by modifying the beta distributed noise. Furthermore, a generalization to the case of more than two political parties was also proposed in the same framework.

In this report, we study the correlation between the fluctuations of vote shares using theoretical and empirical methods. Based on U.S. presidential election data, we determine that the correlation of the fluctuation of the vote share of nearest neighbor counties exceed 80% and it is much higher than that of the regional characteristics. Furthermore, as with the regional characteristics, the fluctuation also shows long-ranged spatial correlation. As previously discussed, the threshold model is inappropriate for inducing such a strong cor-
relation, since the distribution of vote share becomes bimodal, which is inconsistent with empirical facts [21]. If the mean-field Ising model is utilized to induce a strong correlation among binary spin variables, the distribution of the magnetization becomes bimodal [31, 32]. Therefore, another model that can incorporate strong correlation without losing the unimodality of the vote share distribution should be introduced. We show that the mean field voter model on networks is a good candidate since the joint distribution of the vote shares becomes a multi-variate beta distribution which can incorporate strong correlation without losing the unimodality of the distribution of the vote shares. Furthermore, the distribution is similar to the multivariate normal distribution, and the calibration of the model parameters is easy.

The paper is organized into several sections as follows. In Sec. II, the U.S. presidential election data is studied and the vote shares are decomposed into the equilibrium values and the fluctuations around them. The regional characteristics are encoded in the former and both exhibit strong and long-ranged spatial correlation. It is shown that the fluctuations approximately obey a multivariate normal distribution. A mean field voter model on a networks is introduced in Sec. III. The multivariate Wright-Fisher diffusion equation is then derived for the joint probability density function of the vote shares. The stationary distribution is a multivariate beta distribution. We approximate the distribution using a multivariate normal distribution and estimate the covariance matrix of the vote shares. Sec. IV is devoted to the numerical analysis and verification of the theoretical results. Sec. V includes the conclusions and discussions of future problems.

II. EMPIRICAL STUDY

The U.S. presidential election data from 1980 to 2016 were studied. A total of ten elections occurred during this interval. They were labeled as \( t = 1, 2, \ldots, T = 10 \) where \( t = 1 \) corresponds to the election in 1980. The data of 3105 counties was studied and label as \( i = 1, 2, \ldots, I \). The data consist of the number of votes \( N(i, t) \) and the number of votes for the democratic party (DEM) \( k(i, t) \) in county \( i \) and election \( t \). The total number of votes cast in election \( t \) is calculated as \( N_T(t) = \sum_i N(i, t) \). The votes that were not cast to either the DEM or to the Republican party (GOP) were excluded, and the votes for the GOP is given by \( N(i, t) - k(i, t) \). We denote the vote share for the DEM as \( v(i, t) \equiv k(i, t)/N(i, t) \).
Initially, we detrend the vote share data. The weighted spatial average of \( v(i, t) \) is estimated as

\[
v(t) \equiv \sum_i k(i, t)/N_T(t).
\]

We estimate the temporal average of \( v(t) \) as \( v_{\text{avg}} \equiv \sum_t v(t)/T \) and obtain the detrended vote share for DEM as

\[
v_d(i, t) \equiv v(i, t) - (v(t) - v_{\text{avg}}).
\]

Based on this process, the weighted spatial average of \( v_d(i, t) \) becomes \( v_{\text{avg}} \) and it does not depend on \( t \). The temporal average of \( v_d(i, t) \) is defined as \( v_d(i) = \sum_i v_d(i, t)/T \), which is an estimate of the equilibrium values of \( v_d(i, t) \) in county \( i \). We denote the fluctuation (deviation) of \( v_d(i, t) \) around \( v_d(i) \) as \( \Delta v_d(i, t) = v_d(i, t) - v_d(i) \). Figure 1(a) shows the distribution of \( \Delta v_d(i, t) \). SD is approximately 8% and slightly left-skewed.

We then study the \( N \) dependence of the variance. The spatial average of \( v_d(i, t) \) is denoted as \( v_{d,\text{avg}} = \sum_i v_d(i)/I \). The fluctuation of \( v_d(i, t) \) around \( v_{d,\text{avg}} \) is decomposed as the sum of the fluctuation of \( \Delta v_d(i, t) \) and that of \( v_d(i) \).

\[
V(v_d(i, t)) = \frac{1}{IT} \sum_{i,t} (v_d(i, t) - v_{d,\text{avg}})^2 = \frac{1}{IT} \sum_{i,t} (v_d(i, t) - v_d(i) + v_d(i) - v_{d,\text{avg}})^2
\]

\[
= \frac{1}{IT} \sum_{i,t} \{\Delta v_d(i, t)^2 + (v_d(i) - v_{d,\text{avg}})^2\}
\]

\[
= V(\Delta v_d(i, t)) + V(v_d(i))
\]  

(1)

Here, we neglect the cross term in the last equality. The \( N \) dependence of the fluctuation of \( \Delta v_d(i, t) \) is then investigated. We bin \( v_d(i, t) \) according to \( N(i) \) into 31 classes and each class contains 100 counties, with almost the same number of average votes \( N(i) \). As previously discussed [21], if the voters choose independently, the variance of \( \Delta v_d(i, t) \) is proportional to the inverse of \( N(i) \) as \( v_d(i)(1 - v_d(i))/N(i) \). As we will evidentially demonstrate, \( V(v_d(i)) \) does not substantially depend on \( N(i) \). The \( N \) dependence of the variance of \( v_d(i, t) \) mainly related to that of \( \Delta v_d(i, t) \).

Figure 1(b) shows the \( N \) dependence of \( V(v_d(i, t)), V(v_d(i)) \) and \( V(\Delta v_d(i, t)) \). It is evident that, \( V(\Delta v_d(i, t)) \) is much larger than \( 1/4N \) in all the bins. \( V(\Delta v_d(i, t)) \) decreases with \( N \) and the dependence might obey a power law in the form \( N^{-1/4} \).

Next, we study the spatial correlation of \( v_d(i, t) \). We denote the bin of the county pairs
\((i, j)\) separated by their distance \(r\) as \(R(r)\). \(|R(r)|\) represents the number of pairs in the bin. The covariance of \(v_d(i, t)\) and \(v_d(j, t)\) of the pairs in \(R(r)\) is defined as

\[
\text{Cov}(v_d(i, t), v_d(j, t)|r) = \frac{1}{T} \sum_t \frac{1}{|R(r)|} \sum_{(i,j) \in R(r)} (v_d(i, t) - v_d(R(r)))(v_d(j, t) - v_d(R(r))).
\]

Here, \(v_d(R(r))\) is the average value of \(v_d(i, t)\), \(v_d(j, t)\) in \(R(r)\).

\[
v_d(R(r)) = \sum_t \sum_{(i,j) \in R(r)} v_d(i, t)/|R(r)|T = \sum_t \sum_{(i,j) \in R(r)} v_d(j, t)/|R(r)|T.
\]

The covariance is then decomposed using the following identity.

\[
(v_d(i, t) - v_d(R(r)))(v_d(j, t) - v_d(R(r))) = (v_d(i, t) - v_d(i) + v_d(i) - v_d(R(r)))(v_d(j, t) - v_d(j) + v_d(j) - v_d(R(r)))
\]

Here, \(v_d(i)\) and \(v_d(j)\) are the temporal averages of \(v_d(i, t)\) and \(v_d(j, t)\), respectively. We have
the next decomposition of the covariance as the cross term vanishes.

\[
\text{Cov}(v_d(i, t), v_d(j, t)|r) = \text{Cov}(\Delta v_d(i, t), \Delta v_d(j, t)|r) + \text{Cov}(v_d(i), v_d(j)|r).
\]

By normalizing the covariances with the variances, we estimate the correlation coefficients \(\rho\) for \(v_d(i, t), v_d(i)\) and \(\Delta v_d(i, t)\). Figure 2 represents the semi-logarithmic plot of the correlation vs \(r\).

![Semi-logarithmic plots of the correlation of \(v_d(i, t), v_d(i)\) and \(\Delta v_d(i, t)\) vs. \(r\).](image)

FIG. 2. Semi-logarithmic plots of the correlation of \(v_d(i, t), v_d(i)\) and \(\Delta v_d(i, t)\) vs. \(r\).

In all the three cases, the correlation exhibits a logarithmic decay with \(r\). The interesting point is that the correlation of \(\Delta v_d\) is the highest and it is over 80% for the bin of the nearest neighbor county pairs. The correlation of \(\Delta v_d\) for nearest neighbor pairs is estimated to be
83.4%. The spatial correlation of \( v_d(i) \) implies that regional characteristics are similar when the two counties are near. The spatial correlation of \( \Delta v_d(i, t) \) represents the co-movement of the voters’ decisions in the two counties. These two correlations are completely different about their physical origins. For \( r \geq 10^3 [km] \), an unusual behavior is observed with an increase of the correlation with \( r \) about \( v_d(i, t), v_d(i) \), the correlation of \( \Delta v_d(k, t) \) shows a monotonically decreasing behavior up to \( 2000 [km] \).

The results will now be summarized. We decompose \( v_d(i, t) \) as \( v_d(i) + \Delta v_d(i, t) \) and \( v_d(i, t) \) fluctuates around \( v_d(i) \). \( v_d(i) \) is the regional characteristic of county \( i \) and its spatial correlation has been previously discussed\([20, 21]\). \( \Delta v_d(i, t) \) shows a stronger spatial correlation than \( v_d(i) \) and the distribution is unimodal. \( \Delta v_d(i, t) \) approximately obeys a multi-variate normal distribution.

III. SOCIAL INFLUENCE MODEL ON NETWORK

We now introduce a mean field voter model on a network. There are \( I \) nodes and they are labeled as \( i = 1, 2, \ldots, I \). The link set \( E = \{(i, j)\} \) consists of links which connect node \( i \) and \( j \). \( J(i) \equiv \{j | (i, j) \in E\} \) denotes the set of nodes which are linked with node \( i \). In each node there are \( N_i \) agents whose decisions obey the dynamics of the voter model\([20, 23]\). One agent is chosen at random from \( N_T = \sum_i N_i \) agents. If the agent is from node \( i \), another agent is chosen from node \( i \) or from nodes \( J(i) \) which is connected with node \( i \). \( k_i \) and \( v_i \equiv k_i/N_i \) denote the number of votes and the vote share of an option. We assume that the intrinsic tendency of the voter in the region \( i \) to vote for an option is determined by the parameters \( \mu_i \) and \( \theta_1 \). \( \mu_i \) is the probability that a voter votes for an option when there is no interaction between voters. \( \theta_1 \) is a parameter that controls the variance of the vote share. The strength of the influence of the voters in the linked node \( j \) is denoted as \( \theta_2 \).

The probability that the number of votes \( k_i \) for an option increases by 1 is written as the product of two probabilities. One probability is for the case where a voter who chooses another option in node \( i \) is chosen from all voters and this probability is given by \( (N_i - k_i)/N_T \). The other probability represents the likelihood that a voter who chooses the option is picked up from node \( i \) and the linked nodes \( j \in I(i) \). \( P(k_i \rightarrow k_i + 1) \) is then written as

\[
P(k_i \rightarrow k_i + 1) = \frac{N_i - k_i}{N_T} \cdot \frac{k_i + a_i + \sum_{j \in J(i)} \theta_2 k_j/N_j}{N_i - 1 + \theta + \sum_{j \in J(i)} \theta_2}.
\]
Here, \( a_i \) is defined as \( a_i = \theta_1 \mu_i \). Likewise, the probability that \( k_i \) decreases by 1 is written as
\[
P(k_i \rightarrow k_i - 1) = \frac{k_i}{N_T} \cdot \frac{N_i - k_i + b_i + \sum_{j \in J(i)} \theta_2 (N_j - k_j)/N_j}{N_i - 1 + \theta + \sum_{j \in J(i)} \theta_2}.
\]

Here, \( b_i \) is defined as \( b_i = \theta_1 (1 - \mu_i) \). There are two main differences in this model compared to the SIRM model \([20, 30]\). One is the intrinsic factor \( \mu_i \) of node \( i \) and and \( \theta_1 \). If the influence from other voters in the linked nodes \( j \in J(i) \) is turned off by setting \( \theta_2 = 0 \), the model reduces to Kirman’s ant colony model \([5, 33]\). The stationary probability distribution of \( k_i \) is the beta binomial distribution. The vote share \( v_i \) obeys a beta distribution in the limit \( N_i \rightarrow \infty \).
\[
v_i \sim \text{Beta}(a_i, b_i)
\]

The expectation value of \( v_i \) is \( \mu_i \) and the variance of \( v_i \) is \( \mu_i (1 - \mu_i) / (\theta_1 + 1) \). This variance originates from the interaction between the voters in node \( i \). The correlation of the voters choices is \( 1 / (\theta + 1) \) \([32]\). The second change is the normalization of \( k_j \) by \( N_j \). The mathematical reason for the modification is to avoid the ill-posedness in the original SIRM model \([30]\). As we shall show shortly, if we normalize as indicated above, the noise term becomes proportional to \( \sqrt{v_i (1 - v_i)} \) as in the Wright-Fisher diffusion equation and it does not break the condition \( v_i \ in (0, 1) \) even when \( v_i \) approaches 0 or 1.

The raising operator \( R_i = P(k_i \rightarrow k_i + 1) \) is rewritten using vote shares \( \vec{v} = (v_1, v_2, \cdots, v_I) \) as:
\[
R_i(\vec{v}) = \frac{N_i}{N_T} \cdot (1 - v_i) \cdot \frac{v_i + a_i / N_i + \sum_{j \in J(i)} v_j \theta_2 / N_i}{1 - 1 / N_i + \theta_1 / N_i + \sum_{j \in J(i)} \theta_2 / N_i}.
\]

The lowering operator \( L_i = P(k_i \rightarrow k_i - 1) \) is also rewritten as
\[
L_i(\vec{v}) = \frac{N_i}{N_T} \cdot v_i \cdot \frac{(1 - v_i) + b_i / N_i + \sum_{j \in J(i)} (1 - v_j) \theta_2 / N_i}{1 - 1 / N_i + \theta_1 / N_i + \sum_{j \in J(i)} \theta_2 / N_i}.
\]

We write \( v_i^c \) for the average value of the vote shares \( \{v_j\} \) of the linked nodes \( j \in J(i) \).
\[
v_i^c = \sum_{j \in J(i)} v_j / |J(i)|.
\]
Here, $|J(i)|$ is the number of linked nodes with node $i$. $R_i, L_i$ are then rewritten as

\[
R_i(\vec{v}) = \frac{N_i}{N_T} \cdot (1 - v_i) \cdot \frac{v_i + a_i/N_i + v_i^c \theta_2 |J(i)|/N_i}{1 - 1/N_i + \theta_1/N_i + \theta_2 |J(i)|/N_i},
\]

\[
L_i(\vec{v}) = \frac{N_i}{N_T} \cdot v_i \cdot \frac{(1 - v_i) + b_i/N_i + (1 - v_i^c) \theta_2 |J(i)|/N_i}{1 - 1/N_i + \theta_1/N_i + \theta_2 |J(i)|/N_i}.
\]

The stochastic differential equation \[34\] for $v_i$ is written with drift $d_i$ and diffusion $D_i$ as:

\[
dv_i = d_i dt + \sqrt{D_i} dW_i(t).
\]

Here, $dW_i(t)$ is iid white noise, or Brownian motion. The drift term $d_i$ is estimated as:

\[
d_i = \frac{\delta v_i}{\delta t} (R_i - L_i) = \frac{\delta v_i}{\delta t} \frac{1}{N_T} (a_i - \theta_1 v_i + \theta_2 |J(i)| (v_i^c - v_i)).
\]

Here, we take the limit $N_i \to \infty$ in the second equality. The diffusion term $D_i$ is estimated as:

\[
D_i = \frac{\delta v_i^2}{\delta t} (R_i + L_i) = \frac{\delta v_i^2}{\delta t} \frac{N_i}{N_T} 2v_i(1 - v_i).
\]

If we set $N_i = N$ and $N_T = IN$, we have $\delta v_i = 1/N$ and $\delta t = 1/IN^2$. $d_i$ and $D_i$ are written as:

\[
d_i = (a_i - \theta_1 v_i + \theta_2 |J(i)| (v_i^c - v_i))
\]

\[
D_i = 2v_i(1 - v_i). \tag{2}
\]

The Fokker-Plank equation for the time evolution of the joint probability density function $f(\vec{v}, t)$ is give as:

\[
\partial_t f(\vec{v}, t) = - \sum_i \left( \partial_i d_i - \frac{1}{2} \partial^2_i D_i \right) f.
\]

Here, we write the derivative by $v_i$ as $\partial_i$. This is a multi-variate Wright-Fisher diffusion process \[35\].

Since the drift and diffusion terms in Eq.(2) does not explicitly depend on $t$, the stochastic system is a statistically stationary process and the solution of the Fokker-Plank equation converges to a stationary distribution \[34\].

\[
f_{st}(\vec{v}) = \lim_{t \to \infty} f(\vec{v}, t).
\]
A potential solution \( f_{st}(\vec{v}) = e^{-\phi(\vec{v})} \) of Eq. (3) exists if \( \phi \) satisfy

\[
-\partial_i \phi = \partial_i \ln f = \frac{2d_i - \partial_i D_i}{D_i} \equiv Z_i
\]

This puts a constraint on \( Z_i \) as:

\[
\partial_j Z_i = \partial_i Z_j.
\]

We obtain

\[
\theta_2 \frac{v_i(1-v_i)}{v_j(1-v_j)} = \frac{\theta_2}{v_j(1-v_j)}.
\]

If we set \( \theta_2 = 0 \), the stationary solution \( f_{st}^0(\vec{v}) \) becomes the direct product of the beta distribution \( f_{Beta}(v_i|a_i, b_i) \).

\[
f_{st}^0(\vec{v}) = \prod_i f_{Beta}(v_i|a_i, b_i).
\]

The expectation value of \( \vec{v} \) is \( \vec{\mu} = (\mu_1, \mu_2, \cdots, \mu_I) \) and the covariance matrix \( \Sigma^0 \) of \( \vec{v} \) is given by

\[
\Sigma^0_{i,j} = \text{Cov}(\vec{v})_{i,j} = \delta_{i,j}v_i(1-v_j) \frac{1}{\theta^2 + 1}.
\]

When \( \theta_1 \gg 1 \), the joint probability function \( f_{st}^0(\vec{v}) \) can be approximated by the multivariate normal distribution as:

\[
\vec{v} \sim N_I(\vec{\mu}, \Sigma^0).
\]

When \( \theta_2 \neq 0 \), the potential solution does not exist. \( \vec{v} \) fluctuates around their equilibrium values \( \vec{v}^* \) and \( v_i^* \) is determined by the condition that \( d_i = 0 \).

\[
v_i^* = \mu_i + (\theta_2 |J(i)|/\theta_1)(v_i^{c*} - v_i^*) = 0.
\]

Here, \( v_i^{c*} \) is the average values of \( v_j^*, j \in J(i) \). We assume that the fluctuation of \( \vec{v} \) is small and approximates \( D_i = 2v_i(1-v_i) \) as \( D_i^* = 2v_i^*(1-v_i^*) = D^* = 2v^*(1-v^*), \forall i \in \{1, 2, \cdots, I\} \). \( v^* \) is the mean value of \( v_i^* \). In this case, the potential condition is satisfied. We obtain \( \ln f \) as:

\[
\ln f = \int \prod_i Z_i d\vec{v} = \frac{1}{v^*(1-v^*)} \left( \vec{a} \cdot \vec{v} - \frac{1}{2} \vec{v}^\Sigma^{-1} \vec{v} \right)
\]
The inverse of the covariance matrix $\Sigma^{-1}$ is

$$(\Sigma^{-1})_{i,j} = \begin{cases} 
(\theta_1 + \theta_2|J(i)|) & i = j \\
-\theta_2 & i \neq j, j \in J(i) \\
0 & i \neq j, j \notin J(i) 
\end{cases}$$

The multi-variate normal approximation of $\bar{v}$ is

$$\bar{v} \sim N(\bar{v}^*, v^*(1 - v^*) \Sigma).$$ (4)

IV. NUMERICAL STUDY

We verify the normal distribution approximation numerically. The conditional probability density function for $v_i$ with $v^*_i$ fixed is a beta distribution with the shape parameters $a_i(v^*_i) = \theta_1 \mu_i + \theta_2|J(i)|v^*_i, b_i(v^*_i) = \theta_1(1 - \mu_i) + \theta_2|J(i)|(1 - v^*_i)$.

$$v_i \sim \text{Beta}(a_i(v^*_i), b_i(v^*_i)).$$

We set the initial values for $\bar{v}$ as $v_i \sim \text{Beta}(a_i, b_i)$. Afterwards, we choose a node $i$ at random and calculate the shape parameters $a_i(v^*_i), b_i(v^*_i)$ and generate new $v_i$ according to $v_i \sim \text{Beta}(a_i(v^*_i), b_i(v^*_i))$. The process is repeated for $I$ times (1 MCS) and we obtain a sample $\bar{v}(1)$. The procedure is repeated with the initial condition $\bar{v}(t), t = 1, \cdots$ and we obtain a sample $\bar{v}(t+1)$. The length (MCS) of the sample sequence $T$ is set as $10^6$. (In 2D system, we set $T = 10^5$.)

A. Two nodes ($I = 2$) case

At first, we consider the $I = 2$ case. We adopt $\mu_1, \mu_2$ which satisfies $v^*_1(1 - v^*_1) = v^*_2(1 - v^*_2)$. We then set $\theta_1 = 10$ and $a_i = b_i = 5$ in case I. In case II, we set $\theta_1 = 10$ and $a_1 = b_2 = 7, a_2 = b_1 = 3$. Based on the symmetry of the system, $v^*_1(1 - v^*_1) = v^*_2(1 - v^*_2)$ holds in both cases. $\bar{v}^*$ is estimated as:

$$v^*_{1,2} = \frac{1}{2} \frac{a_1 + a_2}{\theta_1} \pm \frac{a_1 - a_2}{\theta_1 + 2\theta_2}$$ (5)
The variance of $v_i$ is:

$$V(v_i) = v^*(1-v^*) \frac{\theta_1 + \theta_2}{\theta_1^2 + 2\theta_1\theta_2}$$ (6)

The correlation coefficient $\rho$ of $v_1$ and $v_2$ is:

$$\rho \equiv \frac{\text{Cov}(v_1, v_2)}{\sqrt{V(v_1)V(v_2)}} = \frac{\theta_2}{\theta_1 + \theta_2}. \quad (7)$$

Figure 3 shows the results of the MC studies. The numerical data are plotted with symbols.
and the Gaussian approximation results are presented with lines. Figure 3(a) shows $E(v_i)$ and $v_1^*$ vs. $\theta_2$ for case II. Figure 3(b) shows $V(v_i)$ and Eq.(6) vs. $\theta_2$. Figure 3(c) shows the correlation coefficient $\rho$ and Eq.(7) vs. $\theta_2$. There is some discrepancy in the estimation of the variance, which originates from the diffusion approximation. We see that the Gaussian approximation works well.

B. Lattice case

Next, we investigate the 1D and 2D lattice cases. We are interested in the $r$ dependence of the correlation. We consider $L$ sites for a 1D lattice and $L \times L$ sites for a 2D lattice. The periodic boundary condition is imposed in both cases. The nodes are indexed by $i \in \{1, \cdots, L\}$ for a 1D lattice and $(i, j), i, j \in \{1, \cdots, L\}$ for the 2D lattice, respectively. Nodes are linked with their nearest neighbors and $|J(i)| = 2(4)$ For a 1D (2D) lattice. We set $\mu_i = \mu_{(i,j)} = 1/2$ and $\theta = 10$.

The fluctuation of $v_i$ obeys a multivariate normal distribution for 1D lattice case. The inverse of the covariance matrix $\Sigma^{-1}$ for a 1D lattice is

$$(\Sigma^{-1})_{i,j} = (\theta + 2\theta_2)\delta_{i,j} + \theta_2(\delta_{i+1,j} + \delta_{i-1,j}).$$

For a 2D lattice, the inverse of the covariance matrix $\Sigma^{-1}$ is

$$(\Sigma^{-1})_{(i,j),k,l} = (\theta + 4\theta_2)\delta_{i,k}\delta_{j,l} + \theta_2(\delta_{i+1,k}\delta_{j,l} + \delta_{i-1,k}\delta_{j,l} + \delta_{i,k}\delta_{j+1,l} + \delta_{i,k}\delta_{j-1,l}).$$

The variance of $v_i$ is given by:

$$V(v_i) = v_i^*(1 - v_i^*)\Sigma_{i,i}. \quad (8)$$

The correlation between $v_i$ and $v_{1+r}$ is:

$$\rho(r) = \Sigma_{1,1+r}/\Sigma_{1,1}. \quad (9)$$
FIG. 4. (a) Plots of $V(v_i)$ vs. $\theta_2$ for a 1D lattice with $L = 10^2$. (b) Plots of $\rho(r)$ vs. $r$ for a 1D lattice with $L = 10^2$. (c) Plots of $V(v_{(i,j)})$ vs. $\theta_2$ for a 2D lattice with $L = 32$. (d) Plots of $\rho(r)$ vs. $r$ for a 2D with $L = 32$. We adopt $\theta = 10$ and $\mu_i = \mu_{(i,j)} = 1/2$. In (a) and (c), solid line plots Eq. (8) vs. $\theta_2$ for a 1D and Eq. (10) vs. $\theta_2$ for a 2D. We adopt $\theta_2 = 10^2$ and $10^3$ for a 1D lattice in (b) and $\theta_2 = 10^4/4$ and $\theta_2 = 10^5/4$ for a 2D lattice in (d). In (b) and (d), we plot theoretical results from Eq. (9) for 1D and results from Eq. (11) for the 2D by solid curves.

For a 2D lattice case, we obtain similar equations by replacing $\Sigma_{i,j}$ with $\Sigma_{(i,j),(k,l)}$.

$$V(v_{(i,j)}) = v_{(i,j)}^* (1 - v_{(i,j)}^*) \Sigma_{(i,j),(k,l)}. \quad (10)$$

The correlation between $v_{(i,j)}$ and $v_{(i+r,j)}$ is

$$\rho(r) = \Sigma_{(1,1),(1+r,1)}/\Sigma_{(1,1),(1,1)}. \quad (11)$$
Figure 4 shows a comparison of the MC data with the results of the Gaussian approximation. It can be seen that the multivariate normal distribution describes the joint probability function of $\vec{v}$ quite well. Furthermore, we can confirm that the $r$ dependence of the correlation decays exponentially with $r$ for the 1D lattice case. For the 2D case, the $r$ dependence does not obey an exponential decay. The exponential decay is a generic feature of the model for a regular lattice case. In order to induce a strong correlation, we adopt a large $\theta_2 >> \theta_1$. The correlation length becomes comparable with the system size $L$ and the exponential decay is not observed for the limited system size $L = 32$. If one sets a small $\theta_2$, an exponential decay is observed.

C. U.S. county network case

Here, we calibrate the model parameters $\theta_1, \theta_2$ using the U.S. presidential election data in Section II and the Gaussian approximation of the model. We construct an artificial county network where 3105 counties constitute nodes of the network and the counties with their nearest $z$ neighbors are connected as links. We adopt $z \in \{3, 4, 5\}$ and all the nodes are included in the largest components. We then set $\theta_1, \theta_2$ so that SD of $v_i$ is approximately 8% and the correlation $\rho$ between the nearest neighbor counties becomes approximately 83%, which are the empirical values in Section II. We adopt $\theta_1 = \{0.044, 0.034, 0.03\}$ and $\theta_2 = \{73, 50, 40\}$ for $z = \{3, 4, 5\}$, respectively. As $\theta_2 >> \theta_1$, the equilibrium values $v_i^*$ are almost the same with each other, which suggests that the regional characteristic cannot be encoded in the model parameters $\mu_i$. This point is discussed in the last section. Here, we adopt $\mu_i = 1/2$.

Figure 5 shows the results. The correlation $\rho$ is plotted as a function of $r$ for the three cases $z = 3$ (solid black), $z = 4$ (solid, gray) and $z = 5$ (broken black). They start from the same value of 83% of the nearest neighbor correlation and decays monotonically with $r$. As $z$ increases, the decay rate becomes small and the model shows a longer spatial correlation. We also plot the empirical results of the correlation of $\Delta v_d$ as the function of $r$ using the symbols o. The $z = 5$ case best fits the empirical behavior of the correlation of $\Delta v_d$ for the three cases.
FIG. 5. Semi-logarithmic plots of the correlation of \( v_i \) (model) vs. \( r \). We adopt \( \theta_1 = \{0.044, 0.034, 0.03\} \) and \( \theta_2 = \{73, 50, 40\} \) for \( z = \{3, 4, 5\} \), respectively. The Gaussian approximation is used to estimate the correlation of the models. The symbols (○) is used to plot the correlation of \( \Delta v_d \) vs. \( r \).

V. CONCLUSIONS

In this paper, we study the fluctuation of the vote share in the US presidential election data. Compared with the regional characteristics, which are given by the temporal average of the vote shares in each county, the fluctuation shows a stronger and long-ranged correlation. In order to describe the behavior, we propose a mean field voter model on networks. There are many voters in each node and they choose another voter at random and copy the choice as in the case of the voter model. Another voter is chosen from the same node where the
voter lives or from the nodes that are linked to the living node. Each node has intrinsic parameter $\vec{\mu}$ that determine the preference to an option. In addition, $\theta_1$ and $\theta_2$ incorporate the influence from the voters in the living nodes and from the voters in the linked nodes, respectively. We derive the multi-variate Wright-Fisher diffusion equation for the joint probability density function of the vote shares. The pdf is a multivariate generalization of the beta distribution. We approximate the pdf using the multivariate normal distribution and estimate the variance and the correlation coefficient of the vote shares. The results were then checked numerically.

There are a few unresolved problems for future study. For example, the statistical modeling of elections and the estimations of the model parameters, $\vec{\mu}, \theta_1$ and $\theta_2$ that can realize the empirical nature of the election data is a concern. The estimation should be compatible with the long-ranged nature of the correlation with distance $r$ [20]. We think it is necessary to generalize the model by the incorporation of several types of voters. As the equilibrium values $v_i^*$ becomes approximately equal when $\theta_2 >> \theta_1$, the regional characteristics $\{v_d(i)\}$ cannot be encoded in $\{v_i^*\}$. In order to realize the characteristics, the assumption that all voters are voter model voters is too simple. Some voters do not change their choices even if they interact with many other voters of different choices. This possibility was previously considered in the modeling of Japan’s parliament election with three political parties [36]. We assume two types of voters, the fixed supporter of each political party and the floating voter. The probability function then becomes the combination of the multinomial distribution of the fixed supporter and the Dirichlet distribution of the floating voters. If we take into account the network structure of the social influence, we have a combination of the multinomial distribution and a multivariate generalization of the Dirichlet distribution. Using this idea, it is possible to incorporate the regional characteristics based on the model. It is also worthwhile to solve the Wright-Fisher diffusion equation for the multi-variate beta distribution (Eq.3). Another type of multi-variate beta distribution has been derived for the inference in a statistical control process [37]. The multi-variate beta distribution for the mean field voter model on networks should be derived since it is the natural multi-variate extension of a beta distribution based on the similarity with the multivariate normal distribution.
ACKNOWLEDGMENTS

This work is supported by JPSJ KAKENHI[Grant No. 17K00347].

[1] S. Galam, Int. J. Mod. Phys. C 19, 409 (2008).
[2] C. Castellano, S. Fortunato, and V. Loreto, Rev.Mod.Phys. 81, 591 (2009).
[3] P. Ormerod, Positive Linking (Faber & Faber, 2012).
[4] A. Pentland, Social Physics: How good ideas spread (Penguin Press, 2014).
[5] A. Kirman, Q. J. Econ. 108, 137 (1993).
[6] T. Lux, Econ. J. 105, 881 (1995).
[7] R. Cont and J. Bouchaud, Macroecon. Dynam. 4, 170 (2000).
[8] M. J. Salganik, P. S. Dodds, and D. Watts, Science 311, 854 (2006).
[9] D. J. Watts, J. Consumer Research 34, 441 (2007).
[10] L. Conradt and C. Lisst, Phil. Trans. R. Soc. B364, 719 (2009).
[11] L. Rendell, R. Boyd, D. Cownden, M. Enquist, K. Eriksson, M. W. Feldman, L. Fogarty, S. Ghirlanda, T. Lillicrap, and K. N. Laland, Science 328, 208 (2010).
[12] R. A. Bentley, P. Ormerod, and M. Batty, Behav. Ecol. Sociobiol. 65, 537 (2011).
[13] R. M. Bond, C. J. Fariss, J. Jones, A. Kramer, C. Marlow, J.E.Settle, and J. Fowler, Nature 489, 295 (2012).
[14] S. Mori and M. Hisakado, J. Phys. Soc. Jpn. 79, 034001 (2010).
[15] S. Mori, M. Hisakado, and T. Takahashi, Phys. Rev. E 86, 026109 (2012).
[16] S. Mori, K. Nakayama, and M. Hisakado, Phys.Rev. E 94, 052301 (2016).
[17] K. Nakayama, M. Hisakado, and S. Mori, Sci.Rep. 7, 1937 (2017).
[18] M. Hisakado, F. Sano, and S. Mori, J. Phys. Soc. Jpn. 87, 024002 (2018).
[19] N. A. Araújo, J. S. A. Jr., and H. J. Herrmann, PLoS One 5, e12446 (2010).
[20] J. Fernandez-Gracia, K. Suchecki, J. J. Ramasco, M. S. Miguel, and V. M. Eguíluz, Phys.Rev.Lett. 112, 158701 (2014).
[21] C. Borghesi and J.-P. Bouchaud, Eur. Phys. J.B. 75, 395 (2010).
[22] C. Borghesi, J.-C. Raynal, and J.-P. Bouchaud, PloS One 7, e36289 (2012).
[23] T. M. Liggett, Interacting Particle Systems (Springer-Verlag Berlin Heidelberg, 2005).
Appendix A: Wright-Fisher diffusion for SIRM model

In the SIRM model, voters who live in node $i$ move to node $j$ for work. They interact with other agents of node $i$ and those of node $j$ in addition to the agents on the link $(i, j)$. We denote the number of voters on link $(i, j)$ as $N_{i,j}$ and the number of votes for an option as $k_{i,j}$. As in the social influence model on a network in the main text, we introduce parameter $\vec{\mu}$ which determine the intrinsic preference for an option of node $i$. We also introduce $\theta_1$ and $\theta_2$ which control the variance of the vote share and correlation of the votes shares between nodes, respectively. In SIRM model, there is a parameter $\alpha$ which controls the strength of social influence on the living node and the working node. $N_i \equiv \sum_j N_{i,j}$ and $k_i \equiv \sum_j k_{i,j}$ indicates the number of voters who live in node $i$ and the number of votes for an option due to them. $N_j \equiv \sum_i N_{i,j}$ and $k_j \equiv \sum_i k_{i,j}$ indicates the number of voters who work in node $j$ and the number of votes for an option due to them. Likewise, we also write
\( v_i, v_j \) which represents the vote shares of the voters who lives in node \( i \) and those who work in node \( j \). We then write \( N_T = \sum_{i,j} N_{i,j} \) as the total number of voters.

We write the probabilities for \( k_{i,j} \rightarrow k_{i,j} + 1 \) and \( k_{i,j} \rightarrow k_{i,j} - 1 \) as:

\[
P(k_{i,j} \rightarrow k_{i,j} + 1) = \frac{N_{i,j} - k_{i,j}}{N_T} \cdot \frac{k_{i,j} + \alpha(a_i + \theta_2 v_i) + (1 - \alpha)(a_j + \theta_2 v_j)}{N_{i,j} - 1 + \theta_1 + \theta_2}
\]

\[
P(k_{i,j} \rightarrow k_{i,j} - 1) = \frac{k_{i,j}}{N_T} \cdot \frac{N_{i,j} - k_{i,j} + \alpha(b_i + \theta_2(1 - v_i)) + (1 - \alpha)(b_j + \theta_2(1 - v_j))}{N_{i,j} - 1 + \theta_1 + \theta_2}
\]

Here, \( a_i \equiv \theta_1 \mu_i, b_i = \theta_1(1 - \mu_i) \) are defined as before. The raising operator \( R_{i,j}(\vec{v}) \) and the lowering operator \( L_{i,j}(\vec{v}) \) are then defined as:

\[
R_{i,j}(\vec{v}) = \frac{N_{i,j}(1 - v_{i,j})}{N_T} \cdot \frac{v_{i,j} + \alpha(a_i + \theta_2 v_i)/N_{i,j} + (1 - \alpha)(a_j + \theta_2 v_j)/N_{i,j}}{1 - 1/N_{i,j} + \theta_1/N_{i,j} + \theta_2/N_{i,j}}
\]

\[
L_{i,j}(\vec{v}) = \frac{N_{i,j}}{N_T} v_{i,j} \cdot \frac{v_{i,j} + \alpha(a_i + \theta_2 v_i)/N_{i,j} + (1 - \alpha)(a_j + \theta_2 v_j)/N_{i,j}}{1 - 1/N_{i,j} + \theta_1/N_{i,j} + \theta_2/N_{i,j}}
\]

The stochastic differential equation for \( v_{i,j} \) is written as:

\[
dv_{i,j} = d_{i,j}dt + \sqrt{D_{i,j}}W_{i,j}(t)
\]

Here \( W_{i,j}(t) \) is an iid Wiener process. The drift term \( d_{i,j} \) and diffusion terms \( D_{i,j} \) are written as

\[
d_{i,j} = \alpha a_i + (1 - \alpha)a_j - \theta v_{i,j} + \theta_2(\alpha v_i + (1 - \alpha)v_j), \quad (A1)
\]

\[
D_{i,j} = 2v_{i,j}(1 - v_{i,j}). \quad (A2)
\]