

# 衆議院総選挙の Dirichlet 分布を用いた分散分析

## ANOVA of Election to the House of Representatives in Japan

佐野史昂 \*1      守真太郎 \*2      久門正人 \*3  
Fumiaki Sano      Shintaro Mori      Masato Hisakado

\*1\*2北里大学理学部物理学科      \*3Fintech Lab.  
Department of Physics, Kitasato University      Fintech Lab.

小選挙区導入後の衆議院議員選挙を平均場投票者モデルで記述することを提案する。投票者は固定支持層と浮動票層に分類されるとし、浮動票層の選択を投票者モデルの感染メカニズムで記述するモデルである。このとき、浮動票層の各政党への投票率はディリクレ分布に従い、各政党への投票率は政党、選挙回、値域の3要因とディリクレ分布に従うノイズに分解できることが示す。43回から47回の衆議院議員選挙のデータを用いて3つの要因と浮動票層の比率を推定した。また、固定支持層の支持率の地域差が得票率の時間・空間相関の物理であることを示した。

### 1. はじめに

近年、社会現象は経済物理学や社会物理学において主要な研究対象となり、多くの研究が行われている [Mantegna 2007, Pentland 2014, Ormerod 2012]。Opinion dynamics は中心的な研究テーマの1つであり、投票者モデルは典型的なモデルとして研究されてきた [Mori 2016, Castellano 2009, Mori 2010, Hisakado 2010, Mori 2012]。2014年、Fernandez-Graciaらは、“Is voter model a model for Voter?”というタイトルの論文を発表し、投票者モデルという単純な数理モデルで実際の選挙データを記述できるのかを検証した [Fernandez-Gracia 2014]。投票者の昼夜の仕事場と居住地間の移動と居住地・仕事場での意見の影響過程を投票者モデルでモデル化し、大統領選挙の投票データに見られる得票率の相関関数、分散などの統計的な振る舞いがモデルパラメータをうまく選ぶことで再現できることを示した。このモデルでは、すべての投票者の選択は投票者モデルの感染過程によって記述される。2人の有権者が無作為に選出され、1人の投票者の選択は他人の選択で置き換えられる。一方、投票者は通常、特定の政党を支持する固定支持層と特定の支持政党を持たず、政治状況によって投票する政党を変える浮動層、支持なし層が存在すると考えられている [Killan 2012]。ここでは、投票者モデルの投票者を固定支持層と浮動票層の2種類を考え、固定支持層は他人の選択の影響を受けないが、浮動票層は投票者モデルの感染メカニズムで影響されるとするモデルを考える [Sano 2017]。このモデルを用いて、小選挙区導入後の第43回から第47回衆議院議員選挙のデータ解析を行った。その結果を報告する。

### 2. 平均場投票者モデル

政党を  $K$ 、選挙区  $I$ 、選挙  $T$  とする。各回の選挙では、 $K$  の候補者は各選挙区の1つの議席を争う。政党を  $k \in \{1, 2, \dots, K\}$ 、選挙区を  $i \in \{1, 2, \dots, I\}$ 、選挙を  $t \in \{1, 2, \dots, T\}$  と定義する。選挙  $t$  で地区  $i$  に  $N(t, i)$  の投票があり、 $N(t, i, k)$  の投票者が政党  $k$  に投票した。  $N(t, i) = \sum_k N(t, i, k)$  が成り立つ。  $N(t, i, k)$  は政党  $k$  に投票した不動票層  $N_F(t, i, k)$  と固定支持者  $N_S(t, i, k)$  に分類される。  $N(t, i, k) = N_F(t, i, k) + N_S(t, i, k)$  となり、浮動票層と固定支持者の総数をそれぞれ  $N_F(t, i) \equiv \sum_k N_F(t, i, k)$ 、  $N_S(t, i) \equiv \sum_k N_S(t, i, k)$  とする。不動票層の比率を  $p(t, i) \equiv N_F(t, i)/N(t, i)$  と書く。政

党  $k$  の投票率  $Z(t, i, k) \equiv N(t, i, k)/N(t, i)$  とする。  $Z(t, i, k)$  は下式として分解される。

$$Z(t, i, k) = (1 - p(t, i))N_S(t, i, k)/N_S(t, i) + p(t, i)N_F(t, i, k)/N_F(t, i)$$

$\mu(t, i, k) \equiv N_S(t, i, k)/N_S(t, i)$  は、  $N_S(t, i)$  の政党  $k$  の投票率を表す。  $N_F(t, i, k)/N_F(t, i)$  は、  $N_F(t, i)$  の有権者における政党  $k$  の投票率を表し  $X(t, i, k)$  で表される。

$$Z(t, i, k) = (1 - p(t, i))\mu(t, i, k) + p(t, i)X(t, i, k) \quad (1)$$

固定支持者の選択は変わらず、浮動票層の選択は投票者モデルによって記述されると仮定する。モデルの操作により、モデルにより選ばれた浮動票層が、無作為に選ばれた他の有権者の選択をコピーする。固定支持者は浮動票層の選択にのみ影響する。ここではまた、固定支持者  $N_S$  の中で  $0 \leq \theta \leq N_S$  が浮動票層の選択に影響を与えると仮定する。以下に示すように、  $\phi$  は  $\mu(t, i, k)$  から  $X(t, i, k)$  への変動を制御する。また、  $X(t, i, k)$  の確率密度は、Dirichlet 分布によって記述され、式 (2) で与えられる。

$$P(\vec{x}|\vec{\mu}(t, i), \phi) = \frac{1}{B(\phi\vec{\mu}(t, i))} \prod_k x_k^{\mu(t, i, k)\phi - 1} \quad (2)$$

ここでは Ferrari と Cribari-Neto による Dirichlet 分布のパラメータ化を採用する [Ferrari 2004]。  $\vec{\mu}(t, i) = (\mu(t, i, 1), \dots, \mu(t, i, K))$  とする。このパラメータ化は、  $X(t, i) \sim D_a(\vec{\mu}(t, i), \phi)$  によって表される。  $B((\phi\vec{\mu}(t, i)))$  は多項式のベータ関数であり、規格化定数として働く。これは

$$B((\phi\vec{\mu}(t, i))) = \frac{1}{\Gamma(\phi)} \prod_k \Gamma(\phi\mu(t, i, k))$$

と定義される。

連絡先: shintaro.mori@gmail.com\*1

$\Gamma(\cdot)$  は  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$  として定義されるガンマ関数。各変数  $X(t, i, k)$  は、 $\alpha = \phi\mu(t, i, k)$  及び  $\beta = \phi(1 - \mu(t, i, k))$  でベータ分布に従う。期待値は  $E(X(t, i, k)) = \mu(t, i, k)$  と定義され、分散は  $V(X(t, i, k)) = \mu(t, i, k)(1 - \mu(t, i, k))/(\phi + 1)$  であり、共分散は  $C(X(t, i, k), X(t, i, l)) = \mu(t, i, k)\mu(t, i, l)/(\phi + 1)$  となる。 $\phi$  は「精度」のパラメータであり、変数の分散をモデル化する。平均場投票者モデルの状態において、 $\phi$  が小さい場合は影響が弱く、 $X(t, i, k)$  の変動が大きいことを意味する。 $\phi$  が大きい場合影響が強く、 $\mu(t, i, k)$  からの  $X(t, i, k)$  の変動は抑えられ、 $X(t, i, k)$  は  $\mu(t, i, k)$  とほぼ一致する。

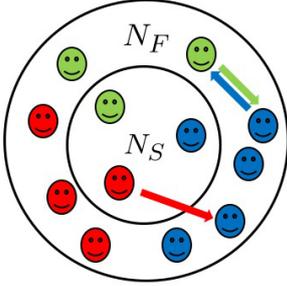


Figure 1: 投票者の構成。投票者は、固定支持者と浮動票層に分類され、それぞれの投票者を  $N_S(N_F)$  で表す。固定支持者は、自分の好みに基づいて決定し、他の投票者の選択に影響されない。不動票層の選択は、投票者モデルに従い他人の選択の影響を受ける。

### 3. 総選挙データのデータ解析

Table 1: それぞれの回の日本の代表選挙（総選挙）データ。第43回から47回までの5つの選挙のデータを使用している。[Mizusaki 2015]

t	date	Discripts	Regions	Voters	Voters
1	2003/11/9	300	3345	$1.02 \times 10^8$	$6.12 \times 10^7$
2	2005/9/11	300	2534	$1.03 \times 10^8$	$6.95 \times 10^7$
3	2009/8/30	300	2037	$1.04 \times 10^8$	$7.20 \times 10^7$
4	2012/12/16	300	1994	$1.04 \times 10^8$	$6.17 \times 10^7$
5	2014/12/14	295	1983	$1.04 \times 10^8$	$5.47 \times 10^7$

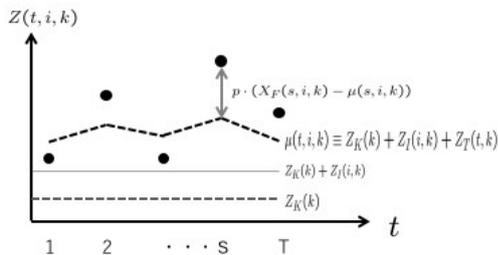


Figure 2:  $Z(t, i, k)$  の分解。 $Z(t, i, k)$  は、式 (1) のように分解できる。 $Z_k$  は、政党  $k$  の実力であり、 $Z_k + Z_l$  は、地域  $i$  の中の政党  $k$  の選挙区内での適性を表している。 $Z_T(t, k)$  は、選挙  $t$  における政党  $k$  の傾向、強さを表している。

日本の代表議院（総選挙）の選挙データを解析する。複数回の選挙でそれぞれに約 300 の選挙区があり、複数の候補者が 1 つの議席を争う。データは table.1 の第 43 回（2003 年）から第 47 回（2014 年）（ $t = 1, 2, 3, 4, 5$ ,  $T \equiv 5$  と置いた）の選挙のデータを用いている [Mizusaki 2015]。日本の自民党（LDP）民主党（DPJ）は、それぞれ  $t=1, 2, 3, 5$  回、 $t=4$  回の選挙で与党として選挙に臨んでいる。自民党の候補者を  $k=1$ 、民主党を  $k=2$ 、他の政党を  $k=3$  とする。立候補者の所属する政党を特定し、各地区には 3 人の候補者がいる場合を用いる。各選挙区は、選挙の結果が記録されたいくつかの地域に分かれており。地域数は 2000 から 3000 程存在する。

3 人の候補者が 1 つの議席に対して 5 回連続して争った地域のデータを使用する。この制限の理由は、平均場投票者モデルにおいて  $\mu(t, i, k)$  を推論する必要があるためである。この制限において 488 の領域があり、それらを  $i = 1, 2, \dots, 488$ ,  $I = 488$  と呼ぶ。選挙時の領域  $i$  における候補  $k$  の投票率を  $Z(t, i, k)$  とする。 $\sum_k Z(t, i, k) = 1$  は、選挙  $t$ 、地域  $i$  のそれぞれで成り立つ。式 (1) のように  $Z(t, i, k)$  を分解するために、 $Z(t, i, k)$  を

$$\begin{aligned} Z(t, i, k) &= \mu(t, i, k) + (Z(t, i, k) - \mu(t, i, k)) \\ \mu(t, i, k) &\equiv Z_k(k) + Z_T(t, k) + Z_l(i, k) \\ Z_k(k) &\equiv \frac{1}{IT} \sum_t \sum_i z_i(t, i, k) \\ Z_T(t, k) &\equiv \frac{1}{I} \sum_i z(t, i, k) - Z_k(k) \\ Z_l &\equiv \frac{1}{I} \sum_i z(t, i, k) - Z_k(k) \end{aligned}$$

と書く。

$Z_k, Z_T, Z_l$  は、政党  $k$  の候補者のいくつかの種類の実力を表している。 $Z_k$  は政党ごとの全体の適応度を示し、 $Z_T$  および  $Z_l$  は  $Z_k$  からの時間的および地域的な差を示す。 $\mu(t, i, k)$  は選挙  $t$  の地域  $i$  における政党  $k$  の候補者の実力を示す。 $Z(t, i, k) - \mu(t, i, k)$  は、前節の平均場投票者モデルによって記述されると仮定する。上記の分解を

$$\begin{aligned} Z(t, i, k) &\equiv (1 - p(t, i))\mu(t, i, k) + p(t, i)X(t, i, k) \\ X(t, i, k) &\equiv \frac{Z(t, i, k) - (1 - p(t, i))\mu(t, i, k)}{p(t, i)} \end{aligned} \quad (3)$$

と書ける。

次のセクションでは、モデルパラメータを推定するために最尤法を使用する。この結果に基づいて、平均場投票者モデルの妥当性を調べる。

### 4. 結果

浮動票層の比  $p(t, i)$  は地域  $i$  に依存せず、それを  $p(t)$  と書くとする。平均場投票者モデルでは、 $Z(t, i)$  は

$$\begin{aligned} \vec{Z}(i, t) &= (1 - p(t))\vec{\mu}(t, i) + p(t)\vec{X}(t, i) \\ \vec{X}(t, i) &\sim D_a(\vec{\mu}(t, i), \phi) \end{aligned} \quad (4)$$

として分解される。

モデルには  $T+1$  個のパラメータ  $(\theta, p_1, p_2, \dots, p_T)$  が存在する。最尤法でそれらを推定する。最尤法の原理的に、いくつかの  $X(i, t, k)$  は小さい  $p$  に対して負になる。この場合、 $p=1$  と仮定する。 $\theta$  は  $10.8 \pm 0.2$  (SE) と推定される。Table2 は、1 行目の  $p_1, \dots, p_T$  の結果を示す。2 行目に  $p=1$  を採用したケースの数も示す。I = 488 の場合の約 3 分の 1 では、各  $t, i$  についての  $X(t, i, k)$  の少なくとも 1 つが負になる。これは  $\mu(t, i, k)$  の推定値が適切でないか、地域  $i$  に対して大きすぎるためである。

Table 2:  $p(t)$  の最尤推定。1 行目は  $p(t)$  の結果、2 行目は I = 488 のうち  $p=1$  になった数を示している。3 行目は日本放送協会 (NHK) の調査結果を示している。

t	1	2	3
p(t)	0.367(1)	0.323(1)	0.320(1)
p=1	146	124	132
NHK poll	0.38(1)	0.25(1)	0.29(1)
t	4	5	
p(t)	0.355(1)	0.331(1)	
p=1	78	142	
NHK poll	0.33(1)	0.26(1)	

日本放送協会 (NHK) の表の 3 行目の推定値は、1000 人以上の世論調査に基づいている。これによると、浮動票層の割合は 25 % から 39 % と推定される。投票者モデルによる推定値は近い値を示した。

#### 4.1 $\mu, Z$ と標準化された $X_R$

式 (1) の分解が正しければ、 $Z(t, i, k)$  は  $\mu(t, i, k)$  の周りに分布する。Figure.3 から分かるように、 $k=1$  および  $k=2$  についての  $Z(t, i, k)$  は、ほぼ対角線の周りに対称的に分布している。一方、 $Z(t, i, 3)$  の分布は対角を中心に対称ではない。これにより後者は、分解が  $k=3$  に対してあまり適切ではないことを示している。

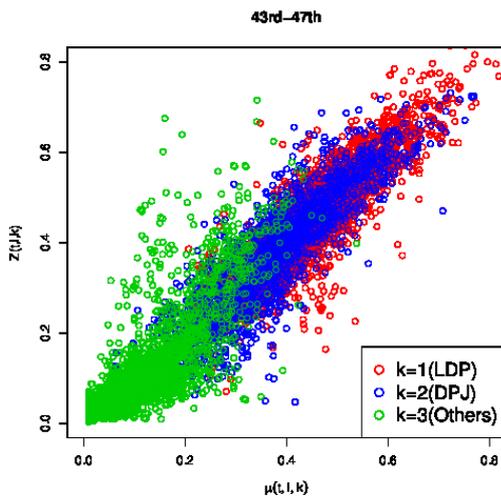


Figure 3: 全てのデータ  $t=1, 2, 3, 4, 5$  に対する  $\mu(t, i)$  と  $Z(t, i)$  の散布図。赤、青、緑のプロットはそれぞれ  $K=1, 2, 3$  を示している。

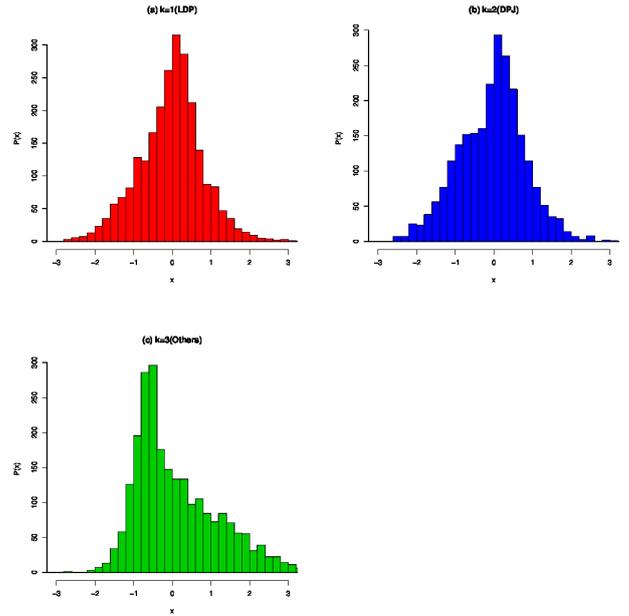


Figure 4: 式 (5) の規格化した  $X_R(t, i, k)$  に対する確率密度関数。分散は、それぞれ  $k=1, 2, 3$  について、0.68, 0.74, 1.59 と推定される。

モデルの有効性を確認するために、規格化された  $X(t, i, k)$  を推定する。モデルが正しく、 $\bar{X}(t, i) \sim D_a(\bar{\mu}(t, i), \phi)$  ならば  $V(X(t, i, k)) = \mu(1 - \mu(t, i, k)) / (\phi + 1)$  であり、規格化された  $X_R(t, i, k)$  は

$$X_R(t, i, k) \equiv \sqrt{\phi + 1} \frac{(X(t, i, k) - \mu(t, i, k))}{\sqrt{\mu(t, i, k)(1 - \mu(t, i, k))}} \quad (5)$$

として定義される。

Figure.4 は、 $k=1, 2, 3$  に対する  $X_R(t, i, k)$  の確率密度をそれぞれ示している。 $X_R(t, i, k)$  の分散は、 $k=1, 2, 3$  についてそれぞれ 0.68, 0.74, 1.59 と推定される。Figure.4 から分かるように、 $k=3$  の分布は  $k=1, 2$  の分布よりも広い。しかし  $X_R$  は正規分布に従わないため、分布の非対称性は重要ではない。しかし、 $k=3$  の分散は、 $k=1, 2$  の場合と比較して約 2 倍であり、式 (1) の分解は  $k=3$  に対してあまり適切でないを示唆している。最後に、結論としてこのモデルの改善について議論する。

#### 4.2 空間的および時間的相関

空間的および時間的相関を推定する。空間に対しての相関関数は

$$C(r, t) = (\langle Z(t, i, k)Z(t, j, k) \rangle_{|r(i,j)=r} - Z(t, k)^2 / V(Z(t, i, k)))$$

$$C(r) = \frac{1}{T} \sum_{t=1}^T C(r, t, k) \quad (6)$$

として計算される。ここで、 $Z(t, k)$  はすべての地域での平均投票率、 $V(Z(t, i, k))$  はその分散、 $\langle Z(t, i, k)Z(t, j, k) \rangle_{|r(i,j)=r}$  は、距離  $r$  離れた地域同士で平均される。これまでの研究で報告されているように、空間的相関は地理的距離と対数的に乖離

することが分かった [Figure.5(a)]。この図では、 $Z_I(i, k)$  の相関関数をプロットした。これに見られるように、 $Z(t, i, k)$  に対して  $C(t)$  とほぼ一致する。これは、空間相関に  $Z_I$  の空間的な不均質性があることを示唆している。どの地域にもいくつかの特徴、地域のタイプ、歴史的な性質などがある。これらの特性は、たとえ地域間に何人かの人々の交流があったとしても急速に変わらない。すべての投票者が浮動票層であるならば [Fernandez-Gracia 2014]、票の拡散とノイズは  $Z_I$  の傾向を維持するように制御されなければならない。今回のモデルでは、投票者は固定支持者と浮動票層の分類のみで、異なる地域間での人の動きは考慮していない。

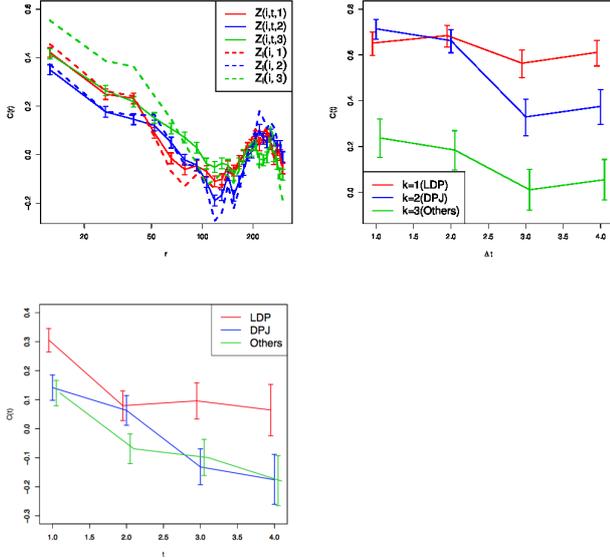


Figure 5: (a) 距離の関数としての空間的な投票率の相関と  $\mu$  の相関。実線は  $Z(t, i, k)$  の空間相関を示し、点線は  $Z_I(i, k)$  の空間相関を示す。(b)  $Z(t, i, k)$  の時間的得票率の相関。(c) は時間距離の関数としての  $X(t, i, k)$  のものである。

Figure.5 (b) は、下式で計算された時間相関  $C(\Delta t)$  の結果を示す。

$$C(\Delta t) = (Cov(Z(1, i, k), Z(1 + \Delta t, i, k))) / \sqrt{V(Z(1, i, k))V(Z(1 + \Delta t, i, k))}$$

ここで、 $Cov(A(i), B(i))$  は、 $\sum_i A(i)B(i)/I - \sum_i A(i)/I \cdot \sum_j B(j)/I$  として定義される。 $C(\Delta t)$  は  $\Delta t$  によって減衰しない。これは式 (1) の分解で説明できる。 $Z(1, i, k)$  と  $Z(1 + \Delta t, i, k)$  の共分散は、 $\mu(t, i, k)$  と  $\mu(t + \Delta t, i, k)$  の共分散と  $X(t, i, k)$  と  $X(t + \Delta t)$  の共分散に分解される。

$$\begin{aligned} Cov(Z(t, i, k), Z(t + \Delta t, i, k)) = \\ (1 - p(1))(1 - p(1 + \Delta t))Cov(\mu(1, i, k), \mu(1 + \Delta t, i, k)) \\ + p(1)p(1 + \Delta t)Cov(X(t, i, k), X(t + \Delta t, i, k)) \end{aligned}$$

ここで、 $\mu(t, i, k)$  と  $X(t, i, k)$  は  $X(1, i, k)$  と  $X\Delta t$  で減衰する。また、

$$Cov(\mu(1, i, k), \mu(1 + \Delta t, i, k)) \simeq V(Z_I(i, k))$$

と仮定する。次に、大きな  $\Delta t$  に対する時間的相関  $C(t)$  は近似的に

$$C(\Delta t) = (1 - p(1))(1 - p(1 + \Delta t)) \frac{V(Z_I(i, k))}{\sqrt{V(Z(1, i, k))V(Z(1 + \Delta t, i, k))}} \quad (7)$$

として表される。

図 5(c) は、 $X(1, i, k)$  と  $X(1 + \Delta t, i, k)$  の  $C(\Delta t)$  を示す。 $C(\Delta t)$  は  $\Delta t$  により減衰し、 $\Delta t = 2$  ではほぼ相関は消滅し、 $C(\Delta t)$  は  $\Delta t \geq 3, k = 2, 3$  で負となる。これはより微妙な分解を行うべきであることを示している。

## 5. 結論

複数の選挙における平均場投票者モデルを提案する。有権者は固定支持者と浮動票層に分類され、後者の行動は投票者モデルの感染メカニズムによって記述されると仮定する。投票率  $\vec{Z}$  は前者の投票  $\mu$  と後者の  $\vec{X}$  から  $\vec{Z} = (1 - p) \cdot \vec{\mu} + p \cdot \vec{X}$  として分解される。 $X$  はパラメータ  $\theta, \mu$  を用いて Dirichlet 分布に従う。 $p$  が地域に依存せず、2003~2014 年の日本の総選挙の選挙データを用いて、モデルパラメータ  $p, \theta, \mu$  を推定する。推定されたパラメータを用いて、 $\vec{Z}$  を分解し、提案したモデルの妥当性を調べる。

- 標準化された  $\vec{X}_R$  の分散は、 $k = 1, 2, 3$  (図 4) について 0.68, 0.74, 1.59 である。

- $\vec{Z}$  の空間相関関数は、 $Z_I(i, k)$  のものによって記述される (図 5(a))。

- $\vec{Z}$  の時間相関関数は、 $Z_I(i, k)$  の分散によって説明される (図 5(b) 及び式 (7))。

これらの結果は、 $\vec{Z} = (1 - p)\vec{\mu} + p\vec{X}$  の分解が選挙データの統計的性質をシミュレートするのに有効であることを示している。しかし、分解は平均場投票者モデルの妥当性を検証するために、より慎重に行わなければならない。 $\mu$  をいくつかの  $\vec{Z}$  平均値で推定ことは最尤推定で行うべきである。さらに、定数  $p(t)$  の仮定も粗すぎるため、それらの領域依存性を考慮する必要がある。これらの点は今後の研究で改善すべきである。

## References

[Mantegna 2007] R. N. Mantegna and H. E. Stanley, "Introduction to Econophysics: Correlations and Complexity in Finance", Cambridge University Press, Cambridge, 2007.

[Pentland 2014] A. Pentland, "Social Physics: How good ideas spread", Penguin Press, 2014.

[Ormerod 2012] P. Ormerod, "Positive Linking", Faber & Faber, 2012.

[Mori 2016] S. Mori, K. Nakayama and M. Hisakado, Phys. Rev. E.94(2016)052301.

[Castellano 2009] C. Castellano, S. Fortunato, and V. Loreto, Rev. Mod. Phys.81(2009)591.

[Mori 2010] S. Mori and M. Hisakado, J. Phys. Soc. Jpn.79(2010)034001.

[Hisakado 2010] [7] M. Hisakado and S. Mori, J. Phys. A43(2010)315207.

- 
- [Mori 2012] S. Mori, M. Hisakado, and T. Takahashi, Phys. Rev. E.86(2012)026109.
- [Fernandez-Gracia 2014] J. Fernandez-Gracia, K. Suchecki, J. J. Ramasco, M. San Miguel, and V. M. Eguíluz, Phys. Rev. Lett, 112(2014)158701.
- [Killan 2012] L. Killan, "The Swing Vote: The Untapped Power of Independents", St. Martin's Press, 2012.
- [Sano 2017] F. Sano, M. Hisakado and S.Mori, accepted in the proceeding of APEC-SSS2016 arXiv:1702.03603.
- [Ferrari 2004] S. P. L. Ferrari and F. Cribari-Neto, Journal of Statistical Software 34(2004)799.
- [Mizusaki 2015] S. Mizusaki and Y. Mori, "JED-M Ver 3.2 28th-47th general elections regional data", LDP press, 2015.