

Aitchison 距離を用いた将来の日本と現在の市区町村との類似度比較

吉田崇紘・堤盛人

A Similarity Comparison of Demographic Structure Using Aitchison Distance: Japan in The Future and its Municipalities in The Year 2010

Takahiro YOSHIDA and Morito TSUTSUMI

Abstract: This study compares the estimated future of Japan through 2010–2110 with Japanese municipalities in 2010 from the aspect of demographic structure based on percentages. For this purpose, this study employs Aitchison distance based on a methodology for the analysis of compositional data as the index of dissimilarity. The choropleth maps based on Aitchison distance show that municipalities located in the mountainous area of the Chugoku region represent the epitome of the future of Japan.

Keywords: 組成データ解析 (compositional data analysis), Aitchison 距離 (Aitchison distance), 人口構成比 (demographic structure), 日本 (Japan)

1. はじめに

我が国の社会問題として、人口減少・高齢化問題が取り上げられるようになって久しい。地方部の人口減少・高齢化も以前にまして深刻な問題となりつつあり、我が国全体が将来に直面する問題を地方が先んじて経験している可能性も指摘されている（増田, 2014, Ch. 5）。このような状況にあって、さまざまな観点から将来の日本の縮図ともいえる市区町村を探し、参考例として都市や地域の政策の長期的な検討を行うことの重要性はますます高まっている。本研究では、高齢化問題に関連する年齢別の人口構成に着目し、将来の日本の縮図を描写する市区町村や地域がどこなのかを、時点ごとの類似度比較と視覚化を通して明らかとすることを試みる。具体的には、比率か

らなる多次元データ（組成データ, compositional data）のための距離指標である Aitchison 距離を用いて、2010 年から 2110 年までの将来の日本と 2010 年の各市区町村における 5 歳階級別人口構成比の類似度分布図を作成し、類似度の高い市区町村を明らかとする。

2. 組成データと Aitchison 距離

2.1 組成データとその解析手法

組成データは比率からなる多次元データである (Aitchison, 1986)。組成データの主たる特徴として、各成分が正数制約と定数和制約を受けていることが挙げられる。組成データの標本空間は次式で表す単体空間上に限定される。

$$S_c^D = \left\{ \mathbf{x} = (x_1, \dots, x_D) \mid x_j > 0, j = 1, \dots, D; \sum_{j=1}^D x_j = c \right\} \quad (1)$$

ここで、 c は定数和、 D は次元数である。 c には 1 か 100 を与えることが多い。

組成データの解析においては、データの性質に

吉田崇紘 〒305-8573 茨城県つくば市天王台 1-1-1

筑波大学大学院システム情報工学研究科

Phone: 029-853-5572

E-mail: yoshida.takahiro@sk.tsukuba.ac.jp

応じた手法を用いることが望ましい。しかし、組成データの性質を無視した、通常用いられる実空間で定義された手法を用いても、一見して解釈可能な結果が得られてしまうことなどから、組成データの解析手法は、これまで関心が向けられることが少なかったと指摘されている（Pawlowsky-Glahn and Egozcue, 2006）。組成データに実空間で定義された手法を用いることの具体的な問題点は、Pawlowsky-Glahn and Egozcue (2006), Otero *et al.* (2005) などで詳細に述べられている。

Aitchison (1986) によって組成データの解析手法がはじめて体系化されて以降、地質学を中心に解析手法の理論・応用研究が進み、近年は人口学（*e.g.*, Lloyd, 2016）を含む他分野に広く適用されるようになっている。

2.2 Aitchison 距離

組成データの解析に用いる代表的な距離指標に Aitchison 距離（式 2）がある。

$$d_A(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^D \left(\ln \frac{x_j}{g(\mathbf{x})} - \ln \frac{y_j}{g(\mathbf{y})} \right)^2} \quad (2)$$

ここで、 $\mathbf{x}, \mathbf{y} \in S_c^D$ 、 g は幾何平均をとる演算子である。実空間上の距離指標である Euclidean 距離と Aitchison 距離の違いは、着目する差の情報が絶対的か相対的かである。二つの距離の違いの具体例は、次のように説明することできる（参考：Otero *et al.*, 2005）。

- Euclidean 距離：1% と 2% の差と、10% と 11% の差を同一とみなす。
- Aitchison 距離：1% と 2% の差と、10% と 20% の差を同一とみなす。

3. 将来の日本と 2010 年の市区町村の類似度比較

3.1 分析の概要と用いるデータ

本研究では、2010 年から 2110 年までの将来の日本と、2010 年時点の各市区町村の 5 歳階級別人口構成比を比較する（図 1）。5 歳階級別人口構成比データは、国立社会保障・人口問題研究所（2012, 2013）が推計・公表しているデータを用いる。なお、将来推計（2010 年から 2060 年まで）・参考推計（2061 年から 2110 年まで）に用いた出生と死亡の仮定は、ともに中位とした。類似度を算出する市区町村の数は福島県内を除く全国 1,799 市区町村、次元の数 (D) は 0–4 歳、5–9 歳、…、85–89 歳、90 歳以上の 19 階級、定数和 (c) は 100（百分率）とした。

口構成比を比較する（図 1）。5 歳階級別人口構成比データは、国立社会保障・人口問題研究所（2012, 2013）が推計・公表しているデータを用いる。なお、将来推計（2010 年から 2060 年まで）・参考推計（2061 年から 2110 年まで）に用いた出生と死亡の仮定は、ともに中位とした。類似度を算出する市区町村の数は福島県内を除く全国 1,799 市区町村、次元の数 (D) は 0–4 歳、5–9 歳、…、85–89 歳、90 歳以上の 19 階級、定数和 (c) は 100（百分率）とした。

3.2 比較分析の結果

2010 年から 2110 年までの代表的な 6 時点について、Aitchison 距離で順位付けした分布図を図 2、各時点上位 10 市区町村とその 2010 年時点の人口規模を表 1 に示す。なお、参考推計では生残率、出生率、出生性比、国際人口移動率（数）を 2061 年以降一定としていることから、参考推計の期間内における Aitchison 距離の変化は小さい。このため、参考推計の期間内における代表時点として 2110 年時点の分布図を図示している。図 2 と表 1 より、2010 年時点では都道府県庁所在地などの中心的な市区町村が高い類似度を示すが、時点が進むにつれて中核的な市区町村と離れた市区町村で高い類似度を示す傾向にあることがわかる。中国地方の山間部は、2030 年以降、高い類似度を示す市区町村が集塊した地域に挙げができる。この結果は、2010 年時点において、中国地方の山間部が、2030 年以降の将来の日本の縮図を表す地域ということができよう。

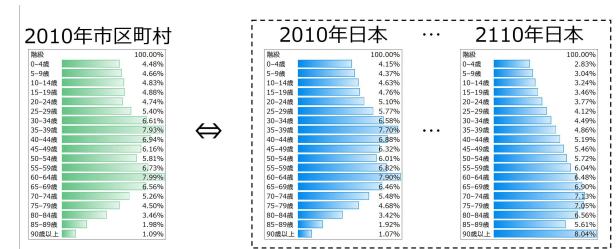


図-1 分析のイメージ

表-1 将来の日本と人口構成比が類似する上位 10 市区町村（人口は 2010 年時点）

順位	2010年		2020年		2030年	
	市区町村	人口 [人]	市区町村	人口 [人]	市区町村	人口 [人]
1	群馬県 高崎市	371,302	山梨県 富士川町	16,307	熊本県 甲佐町	11,181
2	神奈川県 小田原市	198,327	北海道 南富良野町	2,814	北海道 豊浦町	4,528
3	静岡県 浜松市	800,866	三重県 多気町	15,438	長崎県 東彼杵町	8,903
4	群馬県 前橋市	340,291	熊本県 玉東町	5,554	宮城県 丸森町	15,501
5	岐阜県 大垣市	161,160	新潟県 柏崎市	91,451	熊本県 山鹿市	55,391
6	茨城県 土浦市	143,839	岡山県 笠岡市	54,225	兵庫県 淡路市	46,459
7	福岡県 八幡西区	257,097	熊本県 玉名市	69,541	宮城県 川崎町	9,978
8	茨城県 水戸市	268,750	山梨県 山梨市	36,832	山形県 大江町	9,227
9	愛媛県 松山市	517,231	兵庫県 篠山市	43,263	長野県 小海町	5,180
10	岐阜県 岐阜市	413,136	熊本県 宇城市	61,878	山梨県 市川三郷町	17,111

順位	2040年		2050年		2110年	
	市区町村	人口 [人]	市区町村	人口 [人]	市区町村	人口 [人]
1	北海道 豊浦町	4,528	新潟県 出雲崎町	4,907	北海道 壮瞥町	3,232
2	広島県 安芸高田市	31,487	北海道 豊浦町	4,528	長野県 麻績村	2,970
3	山口県 平生町	13,491	北海道 壮瞥町	3,232	新潟県 出雲崎町	4,907
4	熊本県 甲佐町	11,181	広島県 安芸高田市	31,487	群馬県 川湯村	3,898
5	北海道 沿田町	3,612	長野県 小海町	5,180	高知県 越知町	6,374
6	北海道 壮瞥町	3,232	長野県 麻績村	2,970	島根県 邑南町	11,959
7	新潟県 出雲崎町	4,907	北海道 黒松内町	3,250	島根県 川本町	3,900
8	高知県 芽西村	4,048	高知県 本山村	4,103	北海道 豊浦町	4,528
9	長野県 小海町	5,180	熊本県 和水町	11,247	高知県 土佐町	4,358
10	岡山県 矢掛町	15,092	広島県 庄原市	40,244	広島県 安芸高田市	31,487

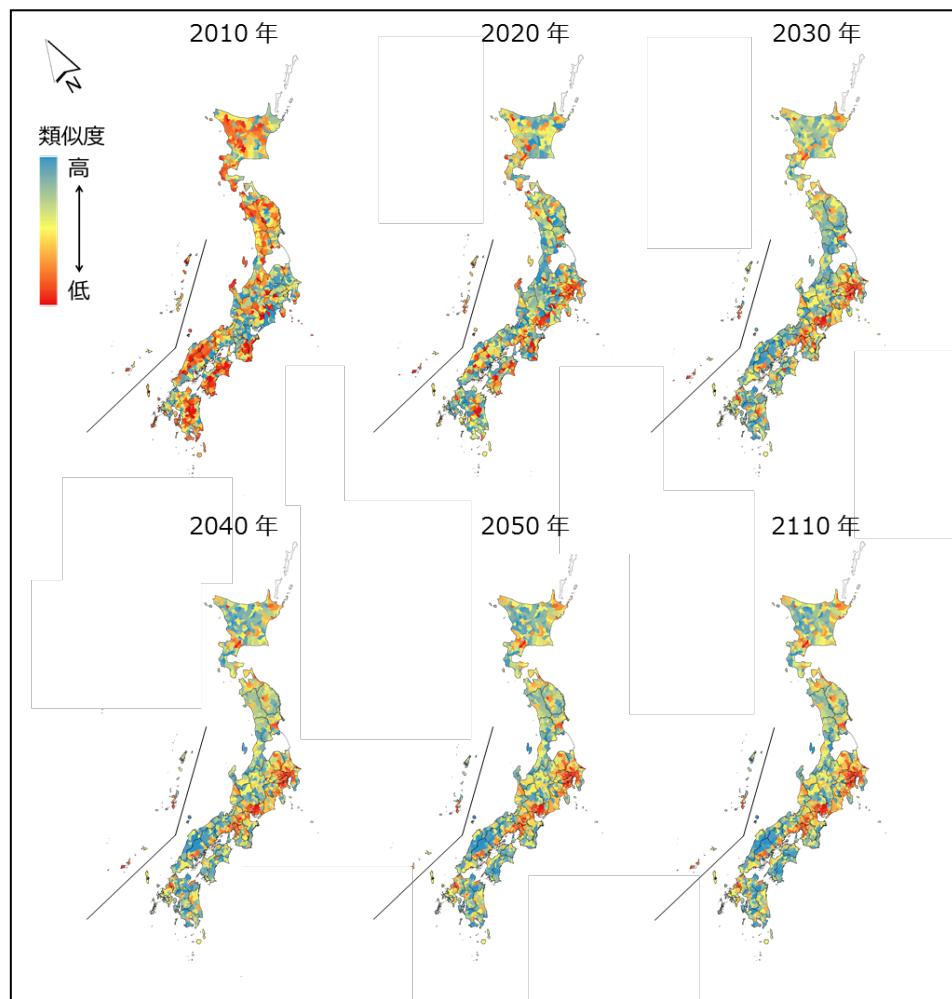


図-2 Aitchison 距離の分布図

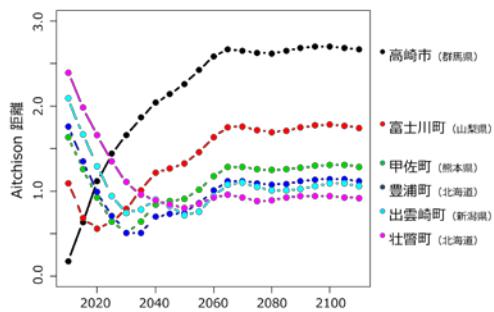


図-3 将来の日本と人口構成比が最も類似する市区町村の Aitchison 距離の推移

上述の 6 時点において、将来の日本の人口構成比と最も類似した 2010 年時点の市区町村（表 1 において 1 番目の順位に記載のある 6 つ）の Aitchison 距離の推移を図 3 に示す。将来の日本の人口構成比は、いずれも各時点で類似する市区町村に近づいていく傾向にあることがみてとれる。高い類似度を示す市区町村は、将来の日本の縮図を 2010 年時点で先んじて表しているということができるよう。

4. 結論と今後の課題

本研究では、国立社会保障・人口問題研究所が公表する 5 歳階級別人口構成比データを用いて、将来の日本と 2010 年の市区町村の類似度を視覚化した。類似度の指標には、比率で構成されたデータに適した距離指標である Aitchison 距離を用いた。分析の結果、将来日本の縮図を 2010 年時点で表している市区町村が中国地方の山間部に集塊していることを示した。また、最も高い類似度を示す市区町村の Aitchison 距離の推移から、将来の日本がそれらの市区町村の人口構成比に近づく傾向にあることを示した。

本研究の手法は、市区町村間の類似度比較にも応用可能である。自市区町村の将来の縮図を、現時点での描寫している他市区町村の姿を通じて知る際には、人口規模を人口構成比に合わせて考慮した手法がより有用と考えられる。そのような手法の検討を今後の課題としたい。

謝辞

データ整備について、筑波大学大学院の安達修平氏、金祥生氏、柳澤直哉氏にご協力頂いた。本研究は JSPS 科研費 16J02219 の助成を受けたものである。ここに記して謝意を表する。

参考文献

- 国立社会保障・人口問題研究所 (2012) 日本の将来推計人口 (平成 24 年 1 月推計) . URL: <<http://www.ipss.go.jp/syoushika/tohkei/newest04/sh2401top.html>> .
- 国立社会保障・人口問題研究所 (2013) 日本地域別将来推計人口 (平成 25 年 3 月推計) . URL: <<http://www.ipss.go.jp/pp-shicyoson/j/shicyoson13/t-page.asp>> .
- 増田寛也 (2014) 地方消滅：東京一極集中が招く人口急減. 中公新書.
- Aitchison, J. (1986) *The Statistical Analysis of Compositional Data*, London: Chapman and Hall.
- Lloyd, C. D. (2016) Spatial scale and small area population statistics for England and Wales. *International Journal of Geographical Information Science*, **30** (6), 1187–1206.
- Otero, N., Tolosana-Delgado, R., Soler, A., Pawlowsky-Glahn, V. and Canals, A. (2005) Relative vs. absolute statistical analysis of compositions: A comparative study of surface waters of a Mediterranean river. *Water Research*, **39**, 1404–1414.
- Pawlowsky-Glahn, V. and Egozcue, J. J. (2006) Compositional data and their analysis: an introduction. In Buccianti, A, Mateu-Figueras, G., and Pawlowsky-Glahn, V. (eds.) *Compositional Data Analysis in the Geoscience: From Theory to Practice*, London: Geological Society, 1–10.