【原著論文】

人口構成比の観点からみた将来日本の縮図 －組成データ解析の適用－

吉田崇紘*・堤盛人**

The epitome of the future Japan from the perspective of demographic composition - Application of compositional data analysis to population studies -

Takahiro Yoshida* and Morito Tsutsumi**

Abstract: Demographic composition, e.g., population composition by age-groups, is one of the most basic information for discussing regional issue or planning. The present study compares the future Japan through 2010–2060, estimated by National Institute of Population and Social Security Research, with all municipalities in 2010 from the perspective of demographic compositions classified by five-year age groups (0–4, 5–9, ..., 90+). The (dis-)similarity of compositional data can be measured with Aitchison distance, one of the core concepts of compositional data analysis. We create choropleth maps based on Aitchison distance. The maps, for example, indicate that the demographic composition of Sobetsu-cho of Hokkaido prefecture of 2010 is the most analogous of the future of Japan of 2060.

Keywords: 人口構成比 (demographic composition), 年齢階級別人口 (population by age-groups), 組成データ解析 (compositional data analysis), Aitchison 距離 (Aitchison distance), 縮図 (epitome)

1. はじめに

1.1. 「縮図」の意義

我が国の社会問題として、人口減少問題や少子高齢化問題が取り上げられるようになって久しい。いわゆる地方と呼ばれる市町村では、こうした問題が先行し既に様々な問題が生じている。一方で、我が国全体が将来に直面する問題を先んじて経験しているという意味で、将来の日本の課題先進地、あるいは将来の日本の縮図と指摘されることもある。たとえば増田 (2014, Ch. 5) では、人口数の減少率や、社会減・自然減に着目し、北海道を将来の日本の縮図として取り上げている。縮図を考えることは、実際の複雑な対象について何らかの観点から単純化し、その類推から対象への知見を得ることである。地域に係る政策を考える際において、様々な観点から自地域の将来の縮図を把握することは、ある課題

に対する先進あるいは参考となる事例を探す上で、有用と考えられる。

将来の日本 (以下、「将来日本」と呼称) の様相・縮図を測る指標は様々な観点があり得る。年齢階級別の人口構成は、そうした指標の中で広く用いられるものの一つである。人口構成の指標には男女別、教育程度別、職業別、産業別などもあるが、年齢階級別は、同時出生集団 (コーホート) に対して固有の属性であるため、将来推計にあたり扱いやすい特性を有している (河野, 2007)。実際、国立社会保障・人口問題研究所や国際連合などの公的機関を含め、年齢階級別の人口構成の将来推計値は様々な団体から公表されている。

人口構成を表すものとして、年齢階級別の人口構成比は、最も広く用いられる指標の一つである。人口高齢化という用語は、人口全体における高齢者の

* 学生会員 筑波大学大学院システム情報工学研究科 (University of Tsukuba)
〒305-8573 茨城県つくば市天王台 1-1-1 E-mail: yoshida.takahiro@sk.tsukuba.ac.jp

** 正会員 筑波大学システム情報系 (University of Tsukuba)

比率が増加していることを示すものであり、数よりも比率に関心を向けたものであると言えよう。年齢階級別の人口構成比は、上述のような特徴に加えて、日本全体と特定の市区町村との比較のように、規模が異なっても比較が容易であるという利点も有する。このようなことから、以下、本稿において、人口構成・人口構成比とは特に断りがない限り年齢階級別のものを指すものとする。

ところで、「縮図」という用語は、得てして概念的に用いられており、縮図とは何かということについては意外と明確に定義されていない。これまでに様々な観点から将来、あるいは現在の日本の縮図として特定の地域や市区町村を取り上げ、そこでの課題を指摘している論考（たとえば、増田, 2014, Ch. 5）は存在するものの、縮図の当てはまりを定量的に検証した研究論文は筆者らの知る限り皆無である。

日本創生会議・人口減少問題検討分科会(2014)は、人口減少問題に関して、次のように指摘をしている：「多くの国民は人口減少の深刻さを十分に認識していない。有効な対策を検討し、果敢に実施するためには、人口減少社会の実像と今後の対応のあり方に関し国民の基本認識の共有を図る必要がある。このため、人口減少の現状と将来の姿を身近な地域のレベルまで示すなど、国民に早急に情報提供する必要がある。」少子高齢化問題においても、その現状と将来の状況を身近な地域のレベルで認識する必要性は同様にあると考える。このような中で、本研究は、具体的にどの市区町村や地域が、人口構成比の観点から将来日本の縮図と考えられるのかを定量的な根拠に基づいて示唆することを目的とする。

1. 2. 本研究の目的と論文の構成

以上のような背景を踏まえ、本研究は、人口構成比に着目し、将来日本の縮図を現在時点で表す市区町村がどこかを、類似度／非類似度を算出して定量的に示すことを目的とする。その際、ここで用いるデータが比率で構成されるデータであることを留意し、そのようなデータの特性を考慮した「組成データ解析 (CoDA: compositional data analysis)」の手法を用いる。また、後述するように、社会経済データ

へのCoDAの適用が始まったばかりであり、その適否についてはこれまで十分な議論がなされていないことから、比率で構成されるデータの特性を考慮するか否かによって分析結果にどのような影響があるのかについても報告する。

以降の各章では、本研究で用いる将来、及び現時点の日本の人口構成比データ及びその加工方法について（第2章）、縮図の作成に際して用いるCoDAの概要と距離指標について（第3章）、そして分析結果について（第4章）、説明する。

2. 人口構成比データ

本研究では、2010、2020、2030、2040、2050、2060年各時点の日本と、2010年時点の市区町村、それぞれの5歳階級別人口構成比の比較を行う（図1）。本章では、本研究で用いた日本と市区町村の人口構成比データについてそれぞれ説明する。

2. 1 将来日本の人口構成比データ

本研究では、将来日本の人口構成比データは、国立社会保障・人口問題研究所（2012）で推計された人口構成データを利用して作成を行った。同データで用いられている推計方法は、人口変動の三要素である出生、死亡、移動（転入・転出）について男女別・年齢別（106階級：0歳，1歳，…，104歳，105歳以上）に仮定を設定するコーホート要因法である。公表されているデータは、将来の男女別・年齢別の人口構成データであるが、本研究では、男女別・年齢別を5歳階級別に集計した上で人口構成比データへの変換を行った。男女別については、年齢階級別の人口構成比に焦点を当てるため、また、年齢別については、次節にて後述する国立社会保障・人口問題研究

図1：本研究における比較の概要

所(2013)が公表する市区町村の人口構成データの階級に合わせるため、年齢別を5歳階級別(19階級: 0-4歳, 5-9歳, ..., 85-89歳, 90歳以上)へと集計した。なお、本研究は、人口構成比の観点から定量的に将来日本の縮図を探る最初の研究と位置づけ、まずは5歳階級別の人口構成比に焦点を当てている。このため、男女の差や年齢階級の区分を考慮した分析等は、今後の課題とし、機会を改めて報告したい。国立社会保障・人口問題研究所(2012)における人口変動要素の仮定は、出生・死亡がそれぞれ低位、中位、高位の3パターン、移動(転入・転出)が1パターンである。本研究では出生、死亡ともに一般的に利用頻度の高い中位の仮定を用いた。

このデータの推計期間は2010年を起点とし2060年を終点としていることから、本研究における将来時点は2060年までとした。なお、このデータでは、人口変動要素の仮定を2060年の値を用いて以降一定とした参考推計が2110年までの期間で示されている。本研究においても2110年までを対象に分析することも可能であり、筆者らは実際に参考推計に基づき計算を行ったが、第4章で示す2060年の結果と2061年以降の結果がほぼ同様となることを確認した上で、紙面の都合により、ここではその詳細については省略をする。

2. 2. 2010年時点の市区町村の人口構成比データ

上述の国立社会保障・人口問題研究所(2012)は、日本全国のみについてのデータであるため、市区町村のデータについては、国立社会保障・人口問題研究所(2013)の2010年の基準人口の人口構成データ(男女別, 5歳階級別(19階級: 0-4歳, 5-9歳, ..., 85-89歳, 90歳以上))を利用して作成を行った。具体的には、将来日本との類似度/非類似度を算出する対象は、国立社会保障・人口問題研究所(2013)において基準人口が公表されている、2013年3月1日現在の福島県内の市区町村を除く1,799市区町村である。ただし、国立社会保障・人口問題研究所(2013)では、同日政令指定都市であった19市の内、さいたま市、相模原市、新潟市、静岡市、浜松市、堺市、岡山市の7市は区単位ではなく市単位となっているため、本研究でも、

この7市を市単位として扱うこととした。

総務省統計局は、日本全国と都道府県の人口について、国勢調査人口(人口等基本集計結果)に含まれる国籍及び年齢不詳人口を按分した基準人口を公表している<URL: <http://www.stat.go.jp/data/jinsui/9.htm>>。しかし、市区町村について、その基準人口は公表されていない。基準人口の算出方法は公開されており、日本全国と都道府県と同様に、国際調査人口から市区町村の基準人口を作成することは可能であるが、本稿では簡便を期して、あらかじめ国籍及び年齢不詳人口が按分されており、国立社会保障・人口問題研究所(2012)の基準人口と整合的な国立社会保障・人口問題研究所(2013)のデータを用いることとした。このデータの5歳階級別への集計と構成比データへの変換は、将来日本と同様の方法で行った。

3. 用いる距離指標

本章では、縮図の作成に際して将来日本と各市区町村の人口構成比の類似度/非類似度を測る距離指標について説明を行う。

本研究では、人口構成比の観点から将来日本の縮図を探る。具体的には、人口構成比のデータを用い、複数の将来時点における日本全体の人口構成比と似通った人口構成比を持った現在時点における市区町村がどこなのかを、各時点の組み合わせに対して、多次元データである人口構成比(ベクトル)間の距離を測ることで探る。当然のことながら、その際にはどのような距離指標を用いるのかが重要となる。

したがって本章では、縮図の作成に際して将来日本と各市区町村の人口構成比の類似度/非類似度を測る距離指標について説明を行う。

3. 1. Euclid距離

人口構成比の類似度/非類似度を測る際に、誰もが思いつく指標は、式(1)で表されるEuclid距離(Euclidean distance)である。

Euclid距離:

$$d_{\text{Euc}}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{d=1}^D (x_d - y_d)^2}. \quad (1)$$

本研究では、 $\mathbf{x}$, $\mathbf{y}$ は、将来時点での日本及び現在時点である市区町村の人口構成比ベクトルであり、前章で説明した通り、5歳階級別（19階級）としていることから、 $D = 19$ である。

3. 2. Aitchison 距離

3. 2. 1. 組成データと組成データ解析

本研究で用いる人口構成比のような比率のデータを扱う統計分野として、「組成データ解析 (CoDA)」と呼ばれるものがある。その方法論は、岩石の化学組成データを扱う地質学を中心に発展してきた (e.g., Aitchison, 1986; Pawlowsky-Glahn and Egozcue, 2006)。組成データ (CoDa: compositional data) という名称自体は必ずしも一般的に知られていないものの、実際には幅広い分野で扱われているデータの種類の。このため、CoDAの方法論は多くの分野への適用可能性があるものの、現状ではいまだ多くの研究余地を残している。その発展の経緯については、Pawlowsky-Glahn et al. (2015) が詳しい。

なお、組成データという用語を含め、本稿で用いる訳語は、この分野に関する日本語で書かれた総説論文である太田・新井 (2006) に倣っている。

CoDaは、比率からなる多次元データである (Aitchison, 1986)。CoDaの最も基本的な特徴として、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。この二つの制約のため、CoDaの標本空間は、 D を次元数とすると、実数空間 $\mathbf{R}^D$ とは異なり、式 (2) に表す単体空間 $\mathbf{S}^D \subset \mathbf{R}^D$ に限定される。

$$\mathbf{S}^D = \left\{ \mathbf{x} = [x_1, \dots, x_D]^T \mid x_d > 0 (d = 1, \dots, D); \sum_{d=1}^D x_d = \kappa \right\}. \quad (2)$$

ここで、「 T 」はベクトルの転置を表す演算子である。また、 κ は定数であり、たとえば、成分を割合で表す場合は $\kappa = 1$ 、百分率で表す場合は $\kappa = 100$ となる。

CoDaと実数空間を標本空間とするデータでは標本空間が異なるため、CoDaに対してそのようなデータのために構築されてきた従来の統計解析手法を実行することに対して、様々な問題が生じうる (例えば、Aitchison, 1986; 太田・新井, 2006)。CoDAではこうした認識に基づいて、CoDaの特性を考慮した方法論

の開発が進められてきた (e.g., Scaaly and Welsh, 2014; Pawlowsky-Glahn et al., 2015; Barceló-Vidal and Martín-Fernández, 2016)。Aitchison (1986) や Pawlowsky-Glahn et al. (2015) は、CoDaの特性を考慮しないことに起因した、従来の統計解析手法を用いた分析結果への影響を詳細に述べている。しかし、そうした警告がある一方で、いまだ多くの分野で、それが無視、あるいは放置されている、と Aitchison and Egozcue (2005) は指摘している。

3. 2. 2. 写像変換を利用した組成データ解析

CoDaを分析するためのアプローチの一つに、写像変換を利用するものがある。これは、CoDaに従来の統計解析手法を適用したい場合は、まず、式 (2) で表したCoDaの標本空間から実数空間へ写像変換を行えばよい、とするものである。その変換は、操作の特徴から対数比変換 (log-ratio transformation) と呼ばれる (Aitchison, 1986)。この変換操作の例として、本稿では、単体空間における距離指標に深く関連した代表的な変換操作である有心対数比 (clr: centred log-ratio) 変換 (式3) を取り上げる。

$$\text{clr}(\mathbf{x}) = \left[\ln \frac{x_1}{g(\mathbf{x})}, \ln \frac{x_2}{g(\mathbf{x})}, \dots, \ln \frac{x_D}{g(\mathbf{x})} \right]^T \quad (3)$$

ここで、 $\mathbf{x} \in \mathbf{S}^D$ であり、 $g(\mathbf{x}) = \left(\prod_{d=1}^D x_d \right)^{1/D}$ ($\mathbf{x}$ の幾

何平均) である。clr変換により、単体空間 $\mathbf{S}^D$ は、 $\mathbf{S}^D \rightarrow \mathbf{U}^{D-1} \subset \mathbf{R}^D$ である実数空間 $\mathbf{R}^D$ の部分空間 $\mathbf{U}^{D-1}$ に写像される。clr変換は同型写像 (isomorphism) かつ等長変換 (isometry) であり、この意味で、単体空間におけるCoDaの特性を実数空間に反映可能な変換方法の一つである。

CoDaを分析するための別のアプローチには、単体空間 $\mathbf{S}^D$ を距離空間 (metric space) として再定義し、実数空間に写像することなく分析を行うアプローチである (Aitchison, 1986)。距離空間と定義することで、たとえば、単体空間における平均 (距離中心, metric center) や分散 (距離分散, metric variance) を定義することができる。これまで、手法の蓄積の多い前者のアプローチが用いられることが多かった

が、近年は、後者のアプローチに基づく、距離中心や距離分散などの統計量を基礎とした単体空間における統計解析手法の開発が盛んに行われている(e.g., Pawlowsky-Glahn et al., 2015)。

3.2.3. Aitchison距離の定義

距離関数を定義するための距離指標の選定は、Aitchison(1992)で検討が行われた。CoDAでは、式(4)で表すAitchison距離(Aitchison's distance)が広く用いられている。

Aitchison距離：

$$\begin{aligned} d_{\text{Ait}}(\mathbf{x}, \mathbf{y}) &= d_{\text{Euc}}(\text{clr}(\mathbf{x}), \text{clr}(\mathbf{y})) \\ &= \sqrt{\sum_{d=1}^D \left(\ln \frac{x_d}{g(\mathbf{x})} - \ln \frac{y_d}{g(\mathbf{y})} \right)^2} \\ &= \sqrt{\sum_{d=1}^D \left(\ln \left(\frac{x_d}{y_d} \cdot \frac{g(\mathbf{y})}{g(\mathbf{x})} \right) \right)^2}. \end{aligned} \quad (4)$$

ここで、 $\mathbf{x}, \mathbf{y} \in S^D$ である。Euclid距離が成分間の値の差に着目した指標であるのに対し、Aitchison距離は、次節でもう少し詳しく説明するとおり、成分間の相対的な値の差、あるいは別の言い方をすれば、何倍の差があるのかに着目した指標といえる。

3.3. Euclid距離とAitchison距離の相違点

Euclid距離とAitchison距離の違いは、CoDaに対して異なる見方をとる。式(2)で示した単体空間は、 $D=3$ のとき、図2(a)において斜線を引いた領域となる。無制約の3次元データが3次元上で任意の点を取りうるのに対して、正数制約と定数和制約の掛かるCoDaは、この2次元平面である正三角形上に限定される。図2(b); (c)は、この単体空間上のある3点からのそれぞれ、Euclid距離、Aitchison距離に基づく等距離線を示したものである。Euclid距離に基づく等距離線は同心円上に引かれ、歪みを持たないのに対して、Aitchison距離に基づく等距離線は、各頂点に近づく(ある成分の値が0に近づく)ほど間隔が狭くなっており、歪みが大きくなる。この歪みは、成分の値が0に近いほど、比の差が大きくなりやすいためである。

Euclid距離とAitchison距離の違いをAitchison et al. (2000), Lovell et al. (2011), Pawlowsky-Glahn et al. (2015)を参考に、具体的に説明する。ここでは、次のようなCoDaを考えることとする(図3)： $\mathbf{a} = [5, 65, 30]^t$ ； $\mathbf{b} = [10, 60, 30]^t$ ； $\mathbf{c} = [50, 20, 30]^t$ ； $\mathbf{d} = [55, 15, 30]^t$ ； $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in S^3$ 。このとき、 $d_{\text{Euc}}(\mathbf{a}, \mathbf{b}) = d_{\text{Euc}}(\mathbf{c}, \mathbf{d}) = 7.07$ であり、二組のペア $(\mathbf{a}, \mathbf{b})$ 、 $(\mathbf{c}, \mathbf{d})$ 間のEuclid距離は等しくなる。一方、このペア間のAitchison距離は、 $d_{\text{Ait}}(\mathbf{a}, \mathbf{b}) = 0.60$ 、 $d_{\text{Ait}}(\mathbf{c}, \mathbf{d}) = 0.28$ となり、異なる。ここで、CoDaの第一成分に着目すると、ペア $(\mathbf{a}, \mathbf{b})$ では2倍の差($2 = 10/5$)である一方、ペア $(\mathbf{c}, \mathbf{d})$ では1.1倍の差($1.1 = 55/50$)である。従って、Euclid距離が成分間の値の差を測る指標であるのに対して、Aitchison距離は成分間に何倍の差があるのかを測る指標と考えることができる。ここでの例を、3区分(0-14, 15-64, 65+)の人口構成比データであると考えると、0-14歳の人口がペア間で同じ5%の差があるとするか、2倍あるいは1.1倍の差があるとするかの違い、という風に考えることができる。ある市区町村の0-14歳の人口構成比が他と比較して2倍程度異なる市区町村では、地域に関わる施策の力点が大きく異なると考えられる。Aitchison距離は、このような点に着目した指標であるといえよう。無論、そのような差異を強調するのが適切か否かは、実際の課題やそれを解決する方策によっても異なると考えら

図2：組成データの標本空間(a)と等距離線(b: Euclid距離；c: Aitchison距離)(b, cはPawlowsky-Glahn et al. (2015, Figure 4.3)より筆者作成)

a	5	65	30
	↓×2		
b	10	60	30
c	50	20	30
	↓×1.1		
d	55	15	30

図3：3次元の組成データの例

れ (Otero et al., 2005), 地質学を中心に発展してきた CoDA を, そうした検討を行わずにそのまま用いることの適否については今後の議論が待たれる。

3. 4. 本研究で用いる距離指標

日本における CoDA の適用例は, 筆者らの知る限り, 地質学 (太田・新井, 2006; 新井・太田, 2006a; 2006b), 政治学 (中西, 2003), 統計学 (中西, 2009) においてでしかみられない。また, 国際的にみてもその対象は広がりを見せているものの, その適用はまだ限定的な状況にあり, 人口研究や地域分析への適用例は, C.D. Lloyd やその共同研究者らによる一連の研究 (Lloyd et al., 2012; Lloyd, 2015; 2016) があるのみである。前節でも述べたとおり, CoDA の手法そのものにも課題が残されており, 社会経済データへの適用についてもこれまで十分な議論がなされていないことから, 本研究では, 必ずしも CoDA の手法を用いることが正しいという立場を取らず, 距離指標として, 通常の Euclid 距離と CoDA の枠組みに基づく Aitchison 距離の両方を併用することとした。

4. 分析結果

本章では, 人口構成比の観点からみたときの将来日本の縮図を表す市区町村を, Aitchison 距離を用いて探し出す。その際, CoDa の特性の考慮の有無によって分析結果にどのような影響があるのかを, 第 3.3 節で説明した Aitchison 距離と Euclid 距離との比較を通じて実証的に報告する。

4. 1. Aitchison 距離の空間分布とクラスタリング

図 4 は, 2010, 2020, 2030, 2040, 2050, 2060 年のそれぞれの時点の日本と, 2010 年の各市区町村における Aitchison 距離の分布を示す。また, 各年について Aitchison 距離が小さい上位 10 市区町村を表 1 に示す。同一時点間の比較である 2010 年では, 太平洋ベルト上の多くの市区町村が日本全体との高い類似度 (以下, 単に「類似度」というときは日本全体との類似度を指すものとする) を示す傾向にある。類似度が特に高い市区町村は, 高崎市や小田原市といった地域の中核的な市区町村である。2020

年は, 三大都市圏に含まれる市区町村の類似度が 2010 年時点から小さくなり, この傾向は 2030 年以降時点が進むに従って顕著になる。2040 年以降に類似度が高い市区町村は, 北海道地方や, 中国地方の山間部に属する市区町村となった。2050, 2060 年では, 2010 年人口が 5,000 人以下の市区町村が特に高い類似度を示している。

図 5 は 2010 年における日本と市区町村間の Aitchison 距離に基づいた階層的クラスタリングを行った結果である。クラスタ化の手法には, クラスターの重心まわりの偏差平方和を最小にするように他クラスターを順次併合していく Ward 法を用いている。市区町村間の距離指標は Aitchison 距離, Ward 法の計算アルゴリズムは Murtagh and Legendre (2014) の方法を用いている。図 4, 表 1 の結果も参照すると, 図 5 は, 各クラスターが以下に示す時点の日本の縮図をよく表す市区町村を類型化した結果と解釈できる。

- ・クラスター A : 2010 年
- ・クラスター B : 2020 年
- ・クラスター C : 2030 年, 2040 年
- ・クラスター D : 2050 年, 2060 年

4. 2. Aitchison 距離の推移

図 6 は, 10 年ごとの日本との Aitchison 距離が最も小さい市区町村である, 高崎市 (2010 年), 富士川町 (2020 年), 甲佐町 (2030 年), 豊浦町 (2040 年), 出雲崎町 (2050 年), 壮瞥町 (2060 年), の Aitchison 距離の推移を表す。これより, 2010 年以降の高崎市の Aitchison 距離は次第に大きくなっていくが, その他の五つの市区町村は, 最も類似する時点を最小に, 概ね U 字型の推移となっていることが分かる。Aitchison 距離が小さくなる局面では, 図 6 の各市町村は, 将来日本の人口構成比を先行して描写していると解釈することができる。

4. 3. Aitchison 距離と Euclid 距離の比較

距離指標の違いを調べるため, 2010 年の日本と 2010 年の市区町村を例に二つの距離を比較する。図 7 には, 距離の値に基づいた順位の散布図を示す。二つの距離の順位の関係は, 同順位となる 45 度線

付近に散布する傾向が確認できる。ここで、どういった市区町村で大きな差が生じたのかを検証するため、45度線から最も外れた市区町村である美浜区（千葉県千葉市）と利根町（茨城県）、そして基準となっている2010年の日本について、組成データとこれをclr変換したデータを図8に示す。美浜区は85-89歳、90歳以上の比率が1.00%以下であるため、0に近い変数の影響を強く受ける。一方、利根町では60-64歳の比率が13.02%と突出して高く、値の大きい変数の影響を強く受けるEuclid距離が大きく評価されたと考えられる。なお、利根町は、1970年代のニュータウン開発によってこの時期に20代人口が増加しており、その結果、2010年時点で60-64歳の比率が突出して高くなったものと考えられる。

図9には、二つの距離の順位の差の分布を示す。東京大都市圏に着目すると、千葉県北西部などのい

図6：各年における順位1位の市区町村のAitchison距離の推移

図7：Aitchison距離とEuclid距離に基づく順位の比較

階級	$x_{\text{Japan, 2010}}$	$\text{clr}(x_{\text{Japan, 2010}})$
90歳以上	1.07%	-0.65
85-89歳	1.92%	-0.40
80-84歳	3.42%	-0.15
75-79歳	4.68%	-0.01
70-74歳	5.48%	0.06
65-69歳	6.46%	0.13
60-64歳	7.90%	0.21
55-59歳	6.82%	0.15
50-54歳	6.01%	0.10
45-49歳	6.32%	0.12
40-44歳	6.88%	0.15
35-39歳	7.70%	0.20
30-34歳	6.58%	0.13
25-29歳	5.77%	0.08
20-24歳	5.10%	0.02
15-19歳	4.76%	-0.01
10-14歳	4.63%	-0.02
5-9歳	4.37%	-0.04
0-4歳	4.15%	-0.07

階級	$y_{\text{美浜区, 2010}}$	$\text{clr}(y_{\text{美浜区, 2010}})$
90歳以上	0.32%	-1.12
85-89歳	0.67%	-0.79
80-84歳	1.52%	-0.44
75-79歳	3.26%	-0.11
70-74歳	5.54%	0.12
65-69歳	7.56%	0.26
60-64歳	8.67%	0.32
55-59歳	6.23%	0.18
50-54歳	5.42%	0.12
45-49歳	7.00%	0.23
40-44歳	9.01%	0.34
35-39歳	9.57%	0.36
30-34歳	7.08%	0.23
25-29歳	5.04%	0.08
20-24歳	4.17%	0.00
15-19歳	4.24%	0.01
10-14歳	5.10%	0.09
5-9歳	5.00%	0.08
0-4歳	4.61%	0.04

階級	$y_{\text{利根町, 2010}}$	$\text{clr}(y_{\text{利根町, 2010}})$
90歳以上	1.10%	-0.62
85-89歳	2.00%	-0.36
80-84歳	3.38%	-0.13
75-79歳	4.60%	0.00
70-74歳	6.25%	0.14
65-69歳	10.80%	0.38
60-64歳	13.02%	0.46
55-59歳	9.27%	0.31
50-54歳	5.36%	0.07
45-49歳	4.37%	-0.02
40-44歳	4.73%	0.02
35-39歳	6.88%	0.18
30-34歳	5.64%	0.09
25-29歳	4.79%	0.02
20-24歳	4.25%	-0.03
15-19歳	3.71%	-0.09
10-14歳	3.55%	-0.11
5-9歳	3.59%	-0.10
0-4歳	2.71%	-0.23

図8：組成データとclr変換したデータ

いわゆるベッドタウンと言われる市区町村で両距離の差が大きいことがわかる。これらの市区町村は、美浜区と同様に、2010年時点では高齢層の比率が0に近い、日本全体との比の差が大きく表れたと考えられる。

5. おわりに

本研究では、5歳階級別の人口構成比の観点から将来時点における日本全体の縮図を探った。具体的には、複数の将来時点における日本全体の人口構成比と似通った人口構成比を持った現在時点における市区町村がどこなのかを、各時点の組み合わせに対して、人口構成比間の類似度を測ることで探った。分析方法には、人口構成比という比率で構成されたデータを扱うにあたり、その特性を考慮した、地質学分野で用いられてきた組成データ解析 (CoDA) という手法を適用した。その結果、2010年の日本の縮図を表す市区町村は太平洋ベルト上に多いが、2040年の日本の縮図を表す市区町村は北海道地方や中国地方の山間部に位置する傾向を明らかとした。また、組成データ (CoDa) の特性の考慮の有無による影響を Aitchison 距離と Euclid 距離の違いを例に比較分析を行った。その結果、成分の値が0に近いほど比の差が大きく出やすいため、そうした成分を持つ市区町村において、二つの距離に基づく類似度／非類似度の順位が大きく変動すること、特に、いわゆるベッドタウンと言われるような市区町村においては、高齢層の比率が0に近い、その傾向が顕著であることなどを実証的に確認した。

本研究では、人口構成比データの特性を踏まえて、CoDAにおける基本的な距離指標である Aitchison 距離を用いて分析を行った。しかしながら、実際には、Otero et al. (2005) が指摘するように、距離指標の選択は、分析者が課題に応じて何を強調・捨象するかを見極めて判断するものである。現時点では、人口研究や地域分析においては、Aitchison 距離の適用例は限定的であり、その適用の適否についてはさらなる議論が必要であろう。本研究で用いた市区町村別の5歳階級別人口構成比データに0は含まれていなかったが、Aitchison 距離は組成データの成分の値に

0がある場合、Aitchison 距離は計算することができないという適用上の大きな課題を有しているため、集計単位を市区町村単位から町丁目単位や1km・500mメッシュ単位などへと細かくした場合、男女別のデータとした場合、あるいは、年齢階級別のデータとした場合では、そうした問題に直面する可能性が増えよう。組成データ解析の分野において、0の対処は、基礎的な問題として多くの研究が蓄積されている (Pawlowsky-Glahn et al., 2015)。データを詳細化した場合の分析に向けて、そうした方法論の適用も重要な研究課題である。

謝辞

本研究はJSPS 科研費 16J02219 及び 17K18554 の助成を受けた。本研究の分析に用いたデータの整備について、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。さらに、本稿に対し、匿名の査読者並びに編集委員から貴重な助言を頂いた。ここに記して謝意を表する。

参考文献

- 新井宏嘉・太田亨 (2006a) 組成データ解析における 0 値および欠損値の扱いについて。「地質学雑誌」, **112**(7), 439–451.
- 新井宏嘉・太田亨 (2006b) 対数比法における規格化成分選定プログラム: Woronow-Love-Schedl 法の自動化。「地質学雑誌」, **112**(6), 430–435.
- 太田亨・新井宏嘉 (2006) 組成データ解析の問題点とその解決方法。「地質学雑誌」, **112**(3), 173–187.
- 河野稠果 (2007) 『人口学への招待: 少子・高齢化はどこまで解明されたか』。中央公論新社。
- 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)。<<http://www.ipss.go.jp/syoushika/tohkei/newest04/sh2401top.html>>.
- 国立社会保障・人口問題研究所 (2013) 日本の地域別将来推計人口 (平成 25 年 3 月推計)。<<http://www.ipss.go.jp/pp-shicyoson/j/shicyoson13/t-page.asp>>.
- 中西寛子 (2003) 都道府県別選挙得票率からわかること: 統計的データ分析の例として。「オペレーションズ・リサーチ」, **48**(1), 17–22.

図4：2010年時点の市区町村と2010～2060年の各時点の日本とのAitchison距離

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]	クラス	市区町村	人口 [人]	クラス	市区町村	人口 [人]	クラス
1	群馬県 高崎市	371,302	A	山梨県 富士川町	16,307	B	熊本県 甲佐町	11,181	C
2	神奈川県 小田原市	198,327	A	北海道 南富良野町	2,814	B	北海道 豊浦町	4,528	C
3	静岡県 浜松市	800,866	A	三重県 多気町	15,438	B	長崎県 東彼杵町	8,903	C
4	群馬県 前橋市	340,291	A	熊本県 玉東町	5,554	B	宮城県 丸森町	15,501	C
5	岐阜県 大垣市	161,160	A	新潟県 柏崎市	91,451	B	熊本県 山鹿市	55,391	C
6	茨城県 土浦市	143,839	A	岡山県 笠岡市	54,225	B	兵庫県 淡路市	46,459	C
7	福岡県 北九州市八幡西区	257,097	A	熊本県 玉名市	69,541	B	宮城県 川崎町	9,978	C
8	茨城県 水戸市	268,750	A	山梨県 山梨市	36,832	B	山形県 大江町	9,227	C
9	愛媛県 松山市	517,231	A	兵庫県 篠山市	43,263	B	長野県 小海町	5,180	C
10	岐阜県 岐阜市	413,136	A	熊本県 宇城市	61,878	B	山梨県 市川三郷町	17,111	C

順位	2040年			2050年			2060年		
	市区町村	人口 [人]	クラス	市区町村	人口 [人]	クラス	市区町村	人口 [人]	クラス
1	北海道 豊浦町	4,528	C	新潟県 出雲崎町	4,907	D	北海道 壮瞥町	3,232	D
2	広島県 安芸高田市	31,487	C	北海道 豊浦町	4,528	D	新潟県 出雲崎町	4,907	D
3	山口県 平生町	13,491	C	北海道 壮瞥町	3,232	D	長野県 麻績村	2,970	D
4	熊本県 甲佐町	11,181	C	広島県 安芸高田市	31,487	D	群馬県 川場村	3,898	D
5	北海道 沼田町	3,612	C	長野県 小海町	5,180	D	北海道 豊浦町	4,528	D
6	北海道 壮瞥町	3,232	C	長野県 麻績村	2,970	D	長野県 阿南町	5,455	D
7	新潟県 出雲崎町	4,907	C	北海道 黒松内町	3,250	D	長野県 小川村	3,041	D
8	高知県 芸西村	4,048	C	高知県 本山町	4,103	D	広島県 安芸高田市	31,487	D
9	長野県 小海町	5,180	C	熊本県 和水町	11,247	D	島根県 川本町	3,900	D
10	岡山県 矢掛町	15,092	C	広島県 庄原市	40,244	D	高知県 越知町	6,374	D

図 5：2010 年における日本と各市区町村との Aitchison 距離に基づく Ward 法による階層的クラスタ分析の結果

図 9：Aitchison 距離，Euclid 距離に基づく順位の差の分布（2010 年）

- 中西寛子 (2009) Compositional data に関する変数選択. 「行動計量学」, **36**(2), 67–80.
- 日本創生会議・人口減少問題検討分科会 (2014) 『成長を続ける 21 世紀のために：ストップ少子化・地方元気戦略』.
- < <http://www.policycouncil.jp/pdf/prop03/prop03.pdf>>.
- 増田寛也 (2014) 『地方消滅：東京一極集中が招く人口急減』. 中央公論新社.
- Aitchison, J. (1986) *The Statistical Analysis of Compositional Data*. London: Chapman and Hall.
- Aitchison, J. (1992) On criteria for measures of compositional difference. *Mathematical Geology*, **24** (4), 365–379.
- Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J.A., and Pawlowsky-Glahn, V. (2000) Logratio Analysis and Compositional Distance. *Mathematical Geology*, **32** (3), 271–275.
- Aitchison, J. and Egozcue, J.J. (2005) Compositional data analysis: Where are we and where should we be heading? *Mathematical Geology*, **37** (7), 829–850.
- Barceló-Vidal, C. and Martín-Fernández, J.A. (2016) The mathematics of compositional analysis. *Austrian Journal of Statistics*, **45**, 57–71.
- Lloyd, C.D., Pawlowsky-Glahn, V., and Egozcue, J.J. (2012) Compositional data analysis in population studies. *Annals of the Association of American Geographers*, **102** (6), 1251–1266.
- Lloyd, C.D. (2015) Assessing the spatial structure of population variables in England and Wales. *Transactions of the Institute of British Geographers*, **40** (1), 28–43.
- Lloyd, C.D. (2016) Are spatial inequalities growing? The scale of population concentrations in England and Wales. *Environment and Planning A*, **48** (7), 1318–1336.
- Lovell, D., Müller, W., Taylor, J., Zwart, A. and Helliwell, C. (2011) Proportions, percentages, ppm: do the molecular biosciences treat compositional data right? In Pawlowsky-Glahn, V. and Buccianti, A. eds. *Compositional data analysis: Theory and applications*, Chichester: John Wiley & Sons, 193–207.
- Murtagh, F. and Legendre, P. (2014) Ward’s hierarchical agglomerative clustering method: Which algorithms implement Ward’s criterion? *Journal of Classification*, **31** (3), 274–295.
- 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.
- Pawlowsky-Glahn, V., Egozcue, J.J., and Tolosana-Delgado, R. (2015) *Modeling and Analysis of Compositional Data*. Chichester: John Wiley & Sons.
- Scealy, J.L. and Welsh, A.H. (2014) Colours and cocktails: Compositional data analysis 2013 Lancaster lecture. *Australian and New Zealand Journal of Statistics*, **56** (2), 145–169.
- (2017年2月2日原稿受理, 2017年9月20日採用決定, 2017年10月30日デジタルライブラリ掲載)