

研究ノート

## 地域間情報交流の分析手法について

中村 有一<sup>1</sup>

# Modeling Methodology for Inter-regional Information Flow

Yuichi Nakamura<sup>1</sup>

地域間の情報の流れを定量的に分析し、有効にモデル化することを目標として、これまでさまざまな手法を試みてきた。多次元尺度構成法の枠組みの中では、数量化4類をもとに情報圏の境界と核を求める手法を開発した。また、個人差モデルを通話や郵便などのメディアごとのODデータに適用して、全体的な構造を明らかにした。一方で、地域間の階層構造を捉えるために、Pレベル分析やISMのような構造モデリング手法を適用して分析を試みた。以上は一般的に使われている手法を応用したものであるが、さらに新たな手法として、区域内比率を用いた地域統合化手法や、フラクタル的な構造を意識した手法を開発しようと試みている。ここでは手法の視点から、これまでの研究を振り返るとともに、今後の方向性について検討していきたい。

Various methods have been tried aiming for quantitative analyses and effective models of inter-regional information flow. The technique for identifying the boundary and the nucleus of the information sphere based on quantification method of the fourth type was developed in the frame of multidimensional scaling method. Moreover, the individual difference

---

<sup>1</sup> 多摩大学経営情報学部, Tama University

model was applied to the OD data of each media such as telephone call and mail, and the overall structure was clarified.

On the other hand, the structural modeling technique such as P level analysis and ISM was applied to catch the layered structure of regions. It tries to develop a regional integration technique that uses the district internal ratio and the technique for considering a fractal structure as a newer technique. It wants to look back on a current research from the aspect of methodology, and to examine the direction of the further study.

キーワード：OD データ、多次元尺度構成法、数量化 4 類、結合度、個人差モデル、ALSCAL、P レベル分析、ISM、フラクタル、べき乗則  
OD(Origin-Destination) data, multidimensional scaling=MDS, quantification method of the fourth type, degree of connection, individual difference model, ALSCAL, P level analysis, interpretive structural modeling=ISM, fractal, power law

## 1. はじめに

この研究ノートでは、通話や郵便など地域間の情報の流れを定量的に分析しモデル化する手法として、これまで用いてきた各種の手法を振り返るとともに、その問題点や限界、今後の改良・発展の方向性について考察していきたい。

まず分析に先だって対象となるデータの特性を把握しておく必要がある。全国 47 都道府県のような地域区分は人為的に行われることが多いが、ほとんどの統計データは、これにもとづいて集計されている。人為的に決められた地域区分では、地域の規模によるばらつきは大きい。ただ逆に人為的であるがゆえに、極端な規模の違いはないようになっている。このばらつきは集計されたデータを使う以上、どうしても避けられない制約となる。

また地域間の関係を表すデータの特徴として、人口や面積などのような通常の統計データの場合と異なり、OD データと呼ばれる正方行列の形をしたデータとなる。たとえば  $i$  という地域から  $j$  という地域に  $f_{ij}$  という量の情報が流れたとすると、この  $f_{ij}$  を行列としてまとめたものが OD データとなる。

入力データを、その形式から分類する場合、「相」と「元」という用語を用いる。相とはその

潜在的には構造を取り出したいと考える対象の集合の種類のことであり、元とは用いるデータの次元に対応する。ただし統計的な意味での観測値の繰り返しについては、相や元とは呼ばない(斎藤,1980)。地域間の OD データの場合には、地域の集合は1種類しかないので単相となる。

(なお相が1つの場合は特別に1相ではなく「単相」という言葉が使われる。) また2次元の行列からなるので、2元データとなる。まとめて「単相2元データ」と呼ばれる。

また個人差モデルで分析されるメディアごとにまとめた OD データは、3次元の配列であり、対象となる集合は地域とメディアという2種類からなる。つまり、「2相3元データ」となる。ここではおもにこの2種類のデータが分析の対象となる。

OD データは正方行列の形をしているが、その対角要素は、自地域内で閉じた情報の流れに対応する。これは質的にそれ以外の要素とは異なり、量的にもそれ以外の要素に比べ格段に大きい値となる。この対角要素をどのように位置づけ、どのように扱うかが、分析手法の面からも重要になってくる。

## 2. 多次元尺度構成法

多次元尺度構成法 (MDS=Multidimensional Scaling) とは、項目間の「距離」(または「親近度」) に相当するデータが行列の形で与えられたとき、近いものを近くに、遠いものを遠くに配置し、全体の構造を視覚的にわかりやすく表現するための手法である。情報流動量の分析においては、OD データを親近度データと解釈して、各地域を空間上に配置することになる。

### 2.1. 数量化IV類

MDS の中にもいろいろな手法があるが、情報流動量の分析に最初に用いたのが、「数量化IV類」と呼ばれる手法である。これはもともと行列の形で表される項目間の親近度のデータを使って、項目を空間上に配置する手法の一つである(斎藤,1980)。この手法の原理は、以下の式で表される  $Q$  の値を最大にするように項目を代表する点の空間座標を求めるものである。

項目  $i$  と項目  $j$  の親近度を  $e_{ij}$  とし、これを  $n \times n$  の行列の形にまとめたものを  $\mathbf{E}$  と表すことにする。ただし  $n$  は項目の個数とし、行列  $\mathbf{E}$  の対角要素である  $e_{ii}$  は定義されていないものとする。

$$Q = \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n (-e_{ij}) d_{ij}^2 \quad (1)$$

ここで  $d_{ij}$  は求めようとしている項目  $i$  と項目  $j$  の間のユークリッド距離である。すなわち項目  $i$  の配置に対応する座標を、 $\mathbf{x}_i$  というベクトルで表すとすると、

## 地域間情報交流の分析手法について

$$d_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j) \quad (2)$$

となる。ここで  $T$  の記号は、行列の転置を表す。つまり非親近度 ( $-e_{ij}$ ) と距離の二乗との積和を最大にする座標ベクトル  $\mathbf{x}_i$  を求めるという問題になる。数量化IV類では、このあと、計算をしやすいするために、行列を対称化する操作をし、さらに対角要素を他の要素から計算して求める。これを式で表すと以下の操作で対称行列  $\mathbf{G} = (g_{ij})$  を求めることに対応する。

$$h_{ij} = e_{ij} + e_{ji} \quad (i \neq j), \quad h_{ii} = 0 \quad (3)$$

$$g_{ij} = h_{ij} - \delta_{ij} \sum_{k=1}^n h_{ik} \quad (4)$$

ここで  $\delta_{ij}$  はクロネッカーのデルタと呼ばれるもので、 $i = j$  のときに 1 となり、 $i \neq j$  のときに 0 となる。

ラグランジュの未定乗数法を適用して  $Q$  の最大化を行い、 $\mathbf{x}_i$  をもとめる。このとき問題は以下の固有値問題に帰着される。

$$\mathbf{G}\mathbf{x} = \lambda\mathbf{x} \quad (5)$$

ここで、 $\lambda$  は固有値、 $\mathbf{x}$  は固有ベクトルである。

行列  $\mathbf{G}$  の固有値を求めると、固有値の値がゼロとなる自明の解が得られるが、これ以外の固有値を大きい順に並べ、大きいものから  $p$  個の固有値とそれに対応する固有ベクトルをとると、この  $p$  個の固有値の和が  $Q$  の最大値となり、それに対応する固有ベクトルは  $p$  次元空間への配置を表す座標値となる。

一方、固有値を小さいものから順に  $q$  個足し合わせたとすると、 $Q$  の値は最小にすることができる。この  $Q$  の値を最小にする配置は、「近いものを遠くに、遠いものを近くに」配置しようとする事に相当し、本来の数量化IV類の考え方と逆の発想であるが、これにも実際には意味のある解釈を与えることができる。

前者の  $Q$  を最大化する分析では、互いに緊密につながっている圏域を切りだすことができるのに対して、後者の  $Q$  を最小化する分析では、圏域内の地域を軸上に展開することにより、圏域の核にあたる地域を特定することができる。たとえば関東の地域が 1 つの圏域として取り出された場合、核としては東京が現れる。このように圏域の境界と核が裏表の関係にある分析手法により求められるということは、非常に興味深い点である。

現実のデータに数量化IV類を適用する場合、たとえば通話の都道府県間データを与えると、

ひずみが大きくなり、空間がねじれた感じに配置される。特にデータの値が小さい地域が大きな影響を及ぼすことになり、これが全体の配置を狂わせる要因と考えられる。この影響を取り除いて、全体のひずみを軽減する手法として、「結合度」と呼ばれる指標を導入することを試みた（中村・長谷川・出石,1983）。結合度としては、規模の大小関係を保存したまま規模のばらつきを平均化できるものであれば何でもよく、いろいろな可能性が考えられるが、実際の分析では、以下のような幾何平均の考え方を利用した式を用いて、結合度の定義とした（長谷川・中村・出石,1983）。

$$c_{ij} = \frac{e_{ij}}{\sqrt{\sum_{i=1}^n e_{ij}} \sqrt{\sum_{j=1}^n e_{ij}}} \quad (6)$$

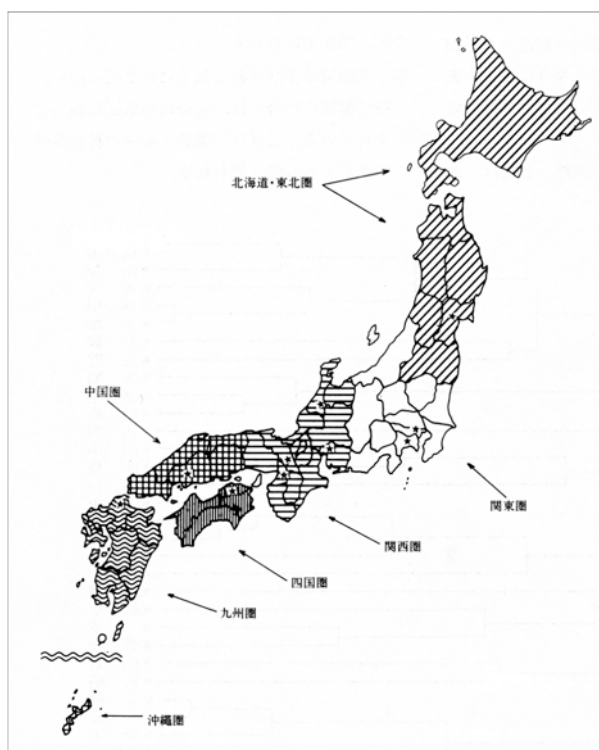


図1 数量化IV類とクラスターリングにより情報圏とその核を求めた例  
（都道府県間通話交流データの場合）

## 地域間情報交流の分析手法について

図1から、情報圏はほぼ地域ブロックと一致し、圏域の核もその中心都市と等しい傾向がわかる。

データの特徴にも依存するが、都道府県間のODデータのように、大都市圏と地方で大きな格差があるようなデータについては、結合度のような一種の正規化の操作が有効であることが明らかになった。

結合度の式としてはいろいろな可能性があり、データの特徴によっても何が有効かわ変わってくるものと思われる。結合度の理論的な意味づけ、およびデータの特徴との関連を求めることは今後の課題である。

数量化IV類について、ODデータなどの地域間データの分析を通じて得られた利点や問題点を整理してみよう。まず利点を挙げるとすると、以下のようなになる。

(1) 対称行列の固有値・固有ベクトルを求めるだけでよく、手法が簡便である。つまり線形代数の範囲で解が求められるので、非線形の問題に付きまとうさまざまな困難に煩わされることがない。

(2) 裏返しの手法が利用できる。最大化問題を最小化問題に読み替えることにより、MDSの本来の目的とは異なるが、意味のある分析が可能となる。

次に問題点を整理してみると、以下のようなになる。

(1) 最適化基準が、入力データに対する距離の適合度を直接評価するものではなく、厳密な意味でのMDSではないという批判もある。

(2) 対称データしか取り扱えない。

(3) 対角要素が扱えない。対角要素のあるデータでは、対角要素の値を無視している。そもそも対角要素はどのように位置づけられるか、合理的な解釈が難しい。対角要素の意味づけを明確にしたうえで、対角要素のあるデータの分析手法についても検討する必要がある。

今後の課題としては、線形代数の範囲で手法を進化させていくことが考えられる。数量化IV類は、行列の固有値問題を解くことに帰着されるが、同様の手法は主成分分析などの多変量解析の分野では多く使われている。また特異値分解による特徴抽出とも関連性があると思われる。さらにデータの特徴による違いの検討や、情報理論的な考察なども必要となるだろう。

## 2.2. 個人差モデル

郵便、通話、旅客など、いくつかの種類 OD データがある場合には、これを一度に分析できる手法はないだろうか。MDS の手法の中に、「個人差モデル」と呼ばれる手法があるが、これはもともと個人ごとに親近度行列のようなデータが与えられて、全体に共通する構造と個人の特性とに分けて分析する手法である。この種のデータは、MDS の分野では、2相3元データと呼ばれている。これは単相2元データが、個人の数だけ積み重なったものと考えられることもできる。

地域間交流データに応用する場合には、全体に共通する構造として地域を空間に配置し、メディアの種類を個性に当たるものとして重み空間に配置する。メディアごとの地域の配置は、共通の配置にメディアごとの重みを掛け合わせたものとなる。

具体的な計算手法としては、INDSCAL, ALSCAL などがあるが、ALSCAL を用いて実際に分析を行った。分析の結果として、共通する地域の配置が求められ、メディアごとの特性も明らかになった(中村,1987)。

この種の問題では、非線形の最適化を行う必要があり、いろいろ難しい問題をはらんでいるが、ALSCAL は、統計パッケージなどにも多く採用されているだけでなく、さまざまなデータに適用され、頑健なアルゴリズムであることが知られている。このモデルでは、全体のデータが、各メディアに共通する成分と、メディアごとの個性に相当する成分に分解できることが前提となっている。この前提は多くの場合成り立っていると考えられるが、データの質が大きく異なる場合には、適用できない。

## 3. 階層構造の抽出手法

MDS では、データに内在する構造を可視化するためにユークリッド空間に項目を配置するという方法をとったが、ここでは項目間に関係があるかないかという2値のデータに置き換え、地域間の階層構造を抽出する手法について考えてみよう。

### 3.1. P レベル分析

P レベル分析は、OD データをもとに地域間の階層構造を明らかにするための手法である。まず、発信地ごとに総発信量（内々も含む）を100%とした着信地のシェアをもとめる。次に、着信地ごとに上で求めたシェアがある一定の閾値以上のものの数を数える。この数をその着信地のその閾値での「P レベル」と呼ぶ。このP レベルは、着信地が影響を及ぼしている発信地の数を表していると考えられるので、この値が高ければ、その着信地は階層構造の上位に位置

## 地域間情報交流の分析手法について

することになる。閾値を変えることにより、どの程度の影響力まで考慮するか変えることができる。なお、発信地と着信地の立場を入れ替えて同様の分析を行うものとして、「Q レベル分析」がある。対称性が高い OD データの場合には、どちらの手法を使ってもほぼ同じような結果がえられる。

P レベル分析で使われる閾値としては、通勤圏の分析の場合、通常 5% が使われる。この値は、データの特性や分析の目的によって変えるべきであるが、閾値を小さくしすぎると細かな影響力まで考慮することになり、かえって大まかな構造がわかりにくくなる。また大きく取りすぎると、細かな関係が捨てられてしまい、粗すぎる分析になってしまう可能性がある。

この分析手法の利点と問題点を整理してみよう。まず、利点としては以下のような点が指摘できる。

- (1) 分析の作業が簡単である。足し算と割り算および閾値との比較を行うだけでよく、表計算ソフトのレベルで可能である。
- (2) 作業が単純なわりに、階層構造がうまく抽出できる。

問題点あるいは課題としては、以下のような点がある。

- (1) 経験的に有効な手法であることが知られているが、理論的な裏付けはあいまいである。
- (2) 最終的にデジタルなデータに還元しているが、その閾値の設定基準は、データから合理的に導き出すべきではないか。つまり情報理論的に最適な解を導き出すことができないか。

この手法により階層の上位にある地域が特定され、その支配下にある地域がいくつあるかで、その圏域の核としての強さが把握できる。P レベルの値を縦軸にとって、地域を配置し、影響のある地域同士を線で結べば、全体の階層構造を図に表すこともできる。

実際に関東地方の都市間の階層構造を MA (Message Area: 単位料金区域) のレベルで分析してみたが、旅客と通話を比較することで、大まかな構造を特定することができた(中村・南部・長谷川文雄,1992)。結果を図 2 に示す。



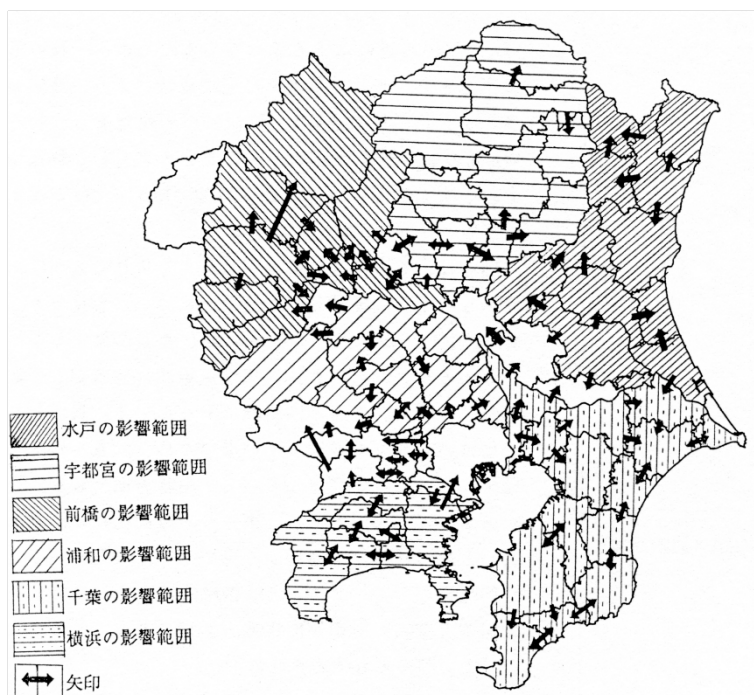


図2 Pレベル分析による影響力関係図（通話回数データ）

図2からは、長野原を除くすべてのMAが東京の影響範囲である。行政区画とほぼ一致した階層構造が読み取れる。

### 3.2. グラフ理論的手法

構造モデリングは、大規模な構造をもった問題群（問題複合体）を解きほぐして、その構造を目に見える形に表現するために考案された手法である。これが一般のシステムについても適用され、さまざまな手法が開発されている。この中にISM(Interpretive Structural Modeling)と呼ばれる手法がある。ISMでは、項目間の関係を2値（ゼロイチ）データで表す。ある特定の意味で関係があれば1で、なければ0とする。これをグラフ理論で使われる隣接行列の形にまとめて入力データとし、結果として項目間の階層構造をDAG(Directed Acyclic Graph=有向無サイクルグラフ)と呼ばれるグラフの形で表現する。DAGは、各辺に向きがあり、サイクルがない。つまり矢印に沿ってどのように動いても元の出発点に戻ってこないようなグラフである。このグラフの最大の特徴は、頂点を順位づけることにより階層構造を抽出することができるという点である。ISMは、この性質を、階層構造をもつシステムのモデル

## 地域間情報交流の分析手法について

化に利用しようというものである(樫木・河村,1981)。

OD データ対して、この手法を適用するためには、元のデータを2値(ゼロイチ)の隣接行列の形に変換する必要がある。このための前処理としては、以下のような方法が考えられる。

- (1)そのまま閾値を設けてそれより大きい要素は1に、小さい要素はゼロに変換する。
- (2)閾値を使うことは上と同様であるが、結合度のような変換式を使って、ある程度小さな構造も切り捨ててしまわないようにする。

いずれにしても閾値をどのように決めるかが重要になる。また、最終的にデータをゼロイチの2値データに変換しているため、データのもつ多くの情報量を捨てていることになる。

ISMについては、実際に都道府県間のODデータに適用し、階層構造を抽出できることが確認できた。ただし、専用の計算プログラムが必要な点や、最終的に2値の行列からDAGの形のグラフを手書きで描く必要があり、単に機械的に描くだけでは見やすいグラフにならない点など、いろいろな問題点がある。グラフ理論に基づく手法は、ISM以外にもさまざまな可能性がある。この方面は今後の発展が期待できる分野でもある。

### 4. MA統合のためのクラスタリング

単位料金区域(MA)とは、固定電話が市内通話料金で通話できる区域のことである。従来、MAは社会生活圏と一致するように設定されていたが、大都市近郊のように、社会生活圏が拡大するのに伴い、両者のずれが大きい場所も存在する。特に問題となるのは、社会生活圏がMAを超えて広がってしまった場合である。これはデータとしては区域内比率が低いことに表れてくる。(ここで区域内比率とは、MAから発信された通話が同一MAに着信する割合のことである。内々率ともいう。)

このずれを解消するためにMAを統合する手法を考案した。この手法のベースになっているのは、クラスタリングの手法であるが、通常のクラスタリングの手法と異なり、統合した時の区域内比率を高める方向に統合を行う。また統合が可能なのは地理的に隣接している区域のみという制限を設けている。

クラスタリングの手法の違いから、統合の様子はさまざまな様相を呈する。最初に試みた「区域内比率最大化法(MAX法)」では、区域内比率が最小のMAを選び出し、隣接するMAを統合の候補として、統合後の区域内比率が最大になるように統合相手を決める。この方法では先に統合したクラスタが後から統合する単独のMAに比べて、相対的に統合対象となりやすいと考えられる。特に大都市圏では、大きなクラスタが次々に周りのMAを統合していき、巨大な

クラスタに成長する。これは「チェイニング」と呼ばれる現象である。

この点を改良するために、「MIN-MAX 法」という手法を開発した。これはクラスタリングの用語では、最長距離法あるいは完全連結法と呼ばれる方法に対応する。この方法では、クラスタ間の親近度を、そのクラスタに属するMA間の親近度の最小値と定義する(鷺尾・大橋,1989)。

MAX-MIN 法では、チェイニングにより大都市周辺地域で巨大なクラスタが形成されることはないが、クラスタ間の融合よりも単独のMA間の融合が起りやすいために、クラスタの成長が2~3個のMAで止まってしまう、単独のMA間の融合ができなくなると、クラスタ間だけの融合になり、今度は急激にクラスタのサイズが大きくなる現象がみられる。

以上のように、MAの統合化は、用いる手法によりその結果が大きく異なる。今後さらにデータの持っている特性を考慮して、自然なクラスタリングのできる手法を開発することが課題である(中村・高嶋裕,1993)。図3に適用した結果を示す。

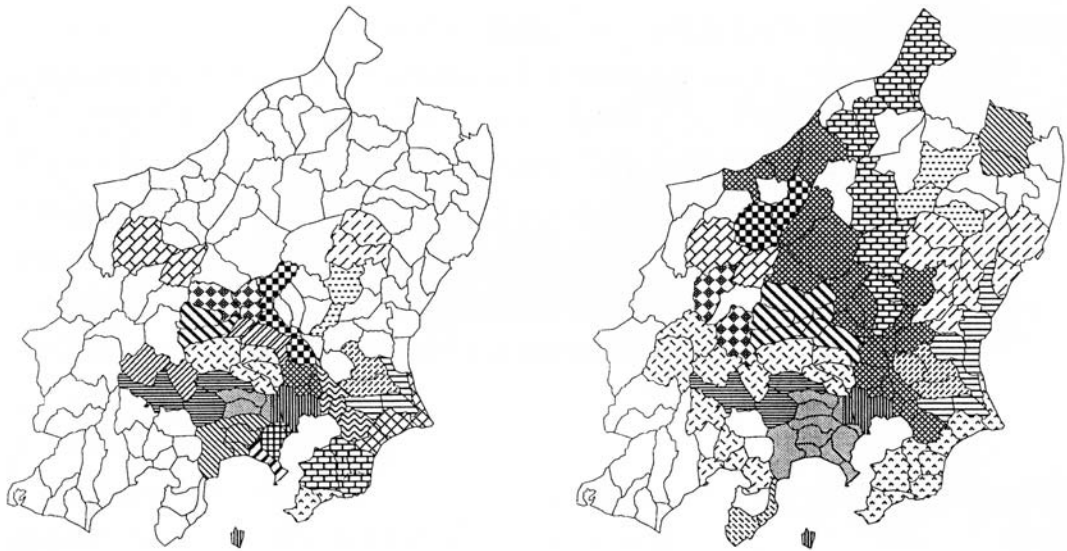


図3 地域統合化手法（MAX-MIN法）を関東地方に適用した例

図3において、左図が閾値60%、右図が閾値70%の場合であり、閾値が上がると急激に統合が進むことが分かる。

## 5. フラクタル構造を意識したモデル化の試み

自然界のデータは、フラクタル的なふるまいをすることが多い。これは単純な原理の積み重ねにより複雑な構造を生み出すことができるという複雑系の仕組みが隠されているからである。この原理をうまく利用すれば、OD データのような階層的な構造をもったデータの分析ができるのではないかと考え、フラクタル画像圧縮の原理を応用した分析手法の開発を目指している。

都市の人口規模とその順位（ランク）を両対数グラフにプロットすると、ほぼ直線状に分布することが知られている。これはランクサイズルールと呼ばれているものであるが、同様の現象は、さまざまな場面で観測される。このような関係を一般に「べき乗則」と呼んでいる。べき乗則は、フラクタル性（自己相似性）と深い関係にあり、自律的な発展過程の結果として得られたある種の安定状態を記述する法則である。したがって自律的な都市成長や自己組織化が行われれば、ある段階から階層的な構造をもつようになり、それがべき乗則として発現するものと考えられる。

フラクタル性は地域の区分自体にも表れる。たとえば、都道府県単位で人口のデータをとると、べき乗則が現れる。これに対して面積のデータではべき乗則は成り立たない。このようにまずデータの持つ特性を把握することが重要になる。

フラクタル性を意識した分析あるいはシミュレーションに先立って、データの基本的な性質をとらえようとしていくつかの分析を行った。まず関東圏における MA 単位の通話交流データを用い、べき乗則が成り立つかどうか実際に検討したが、おおむね成り立っていることが明らかになった。また OD データそのものに対して、その分布形を調べた結果、対数正規分布に従っていることが確認できた。図 4 にそのプロット図を示す。

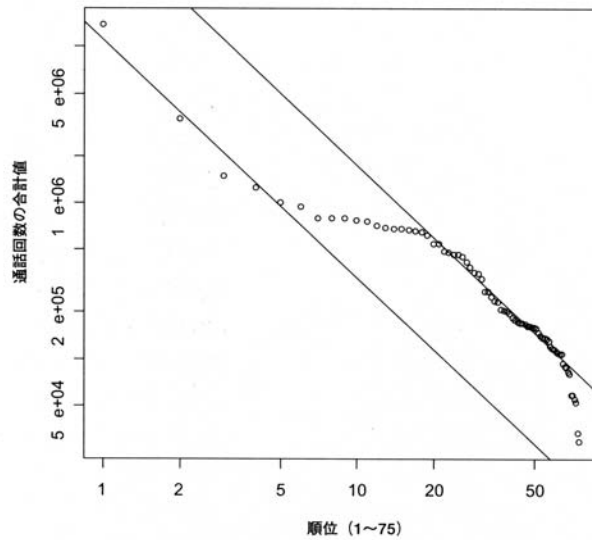


図4 通話回数の合計値とその順位の両対数プロット（関東地方全般）

図4から大都市圏と地方で、それぞれ地域が、べき乗則に従っていることが読み取れる。

これによりフラクタル的なふるまいが明らかになったので、さらに定量的に分析する手法について検討していきたい。またそれにともない、べき乗則をもつデータの正規化手法やフラクタル性を測る指標についても開発していきたい(中村,2007)。

## 6. おわりに

従来、地域間の旅客のデータなどは、グラビティモデルやエントロピーモデルを使って、距離あるいはコストによって説明されることが多かった。情報のデータについては、距離の概念は必ずしも明確ではなく、距離だけでは説明できないものと考え、さまざまなモデルを検討してきた。単に既存のモデルにデータをあてはめるだけでなく、新しいモデルを提案し、現実のデータにあてはめることを通じて、モデル自体の改良を進めてきた。この過程で得られた知見をもとに理論的な意味づけを明確にし、さらにモデルを使いやすく広範囲に応用できるものにしていきたい。

## 地域間情報交流の分析手法について

### 参考文献

- 齊藤堯幸(1980). 多次元尺度構成法,朝倉書店.
- 榎木義一・河村和彦(1981). 参加型システムズ・アプローチ —手法と応用—. 日刊工業新聞社.
- 中村有一・長谷川文雄・出石宏彦(1983). 通話からみた地域間の結合に関する研究, 第18回日本都市計画学会学術研究発表会論文集,73-78.
- 中村有一 (1987). 地域間情報流動に関する研究, 第22回日本都市計画学会学術研究発表会論文集,463-468.
- 中村有一・南部世紀夫・長谷川文雄 (1992). 関東地方における通話の都市間交流に関する研究,1992年度日本都市計画学会論文集,325-330.
- 中村有一・高嶋裕一(1993). 通話トラヒックデータに基づく単位料金区域の統合化手法に関する考察. *TIMIS JOURNAL*,32.
- 中村有一 (2007). 大都市圏における情報交流の分析手法. 多摩大学研究紀要,11, 15-25.
- 長谷川文雄・中村有一・出石宏彦(1983). 通話と郵便による情報圏の比較分析,昭和58年度情報通信学会年,21-41.
- 鷺尾泰俊・大橋靖雄(1989). 多次元データの解析. 岩波書店.

---

受理日 2010年1月10日

採択日 2010年2月10日