

# エントロピー最大化法による不完全地理行列のデータ推定方法

矢野 桂 司

## I はじめに

今日、地理情報システムの発達に伴って、多くの地域情報がデータベース化されてきた。特に、地理学的研究において、その理論化や精緻化が進む中で、適切な地域情報の存在はその研究の成否を左右する場合も少なくない。近年、官庁統計は量・質ともに充実が図られてきたが、必要なデータが表章されているとは限らず、さらに個人情報の保護の観点から、多くの秘匿あるいは非表章項目が存在するのが現状である(矢野, 1989)。また、統計が未整備な国や地域のデータや、過去の歴史的データ(杉浦, 1990)に関してもその利用には大きな制約や限界が存在する。

そこで本研究では、既知のデータの情報を最大限に利用して、より非集計化されたデータを復元しようとするエントロピー最大化法(Chilton and Poet, 1973)を取り上げ、その地理学的データへの適用の有効性を検討する。そしてさらにその汎用プログラムを公開することによって、この方法の普及を試みるものとする。

## II エントロピー最大化法

ここで対象となる地域情報は、地理行列の形で集計されるものである。具体的には、行方向に地域、列方向に地域特性を配した2次元の地理行列(この中には、行方向に発地区、列方向に着地区を配したOD行列も含む)と、さらに時間や他の項目を加えた3次元地理行列(例えば、年齢階級別人口移動OD行列や産業別通勤OD行列など)である。そして、エントロピー最大化法とは、これらの行列の周辺和が既知で、当該行列のセル(要素)が未知の場合に、その個々のセルを復元する推定方法のことである。

### 1) エントロピー概念

エントロピー概念に関しては、これまで様々な紹介や説明がなされてきたが(Wilson, 1981; 杉浦, 1986; 矢野, 1992)、ここでは、エントロピー最大化法によって推定されたデータのもつ意味を

理解するために、エントロピー概念の基本的な考え方を示すことにする。

以下ではWilson(1981)のエントロピー最大化型空間的相互作用モデルの説明にしたがって、東京都区部の通勤流動システムを事例にみていくことにする。したがって、ここで対象となる地理行列は、行方向に常住地を列方向に従業地を配した、2次元の通勤OD行列となる。東京都区部に対して、23常住地×23従業地からなる閉じた通勤流動システムを想定すると、このシステムは3つの集計レベルで記述することができる。一番マクロな状態は、区部に常住・就業する就業者を各地区単位で示したものである。これは、OD行列の周辺和である行和・列和にそれぞれ対応するものである。そして、2番目の集計レベルは、各常住地から各従業地に何人の就業者が通勤しているかを示すもので、このスケールはメソ・スケールと呼ばれる。これはOD行列のセルに相当し、空間的相互作用モデルが問題とするスケールである。そしてこのスケールが本研究のエントロピー最大化法が推定しようとしているものである。

最も細かいマイクロ・スケールは、移動主体である就業者個人のレベルである。このスケールでは、全就業者について個人ごとにどこに常住し、どこに従業しているかを記述することになる。

これら3つの集計レベルを国勢調査と対応させると、マクロ・スケールは、市町村単位の常住地・従業地ごとの就業者数に当たり、メソ・スケールは、常住地による従業市町村(あるいはその逆の、従業地による常住地市町村)別就業者数に相当し、マイクロ・スケールは、常住地と従業地が記入された各就業者個人の個票に相当する。このことは、マイクロ・スケールでの情報を集計していけば、メソ・スケール、さらにマクロ・スケールの状態を記述していくことができることを意味する。しかし、その逆のマクロ・スケールの状態から、メソ・スケール、さらにはマイクロ・スケールの状態を記述することは、極めて困難なものとなる。しかし

この困難を克服しようとするものが、エントロピー最大化法であり、マクロ・スケールの情報から、メソ・スケールの状態を復元することに他ならない。すなわち、この事例では、東京都区部に常住かつ従業する就業者について、23区ごとの常住者数と従業者数が既知の場合に、23常住地×23従業地の通勤OD行列の個々のセルの値をいかにして推定するかが問題となる。

言い換えれば、この問題は、東京都区部に常住・従業する就業者を、どの様に23×23の行列のセルに配分するかという純粋に数学的な問題に帰着する。可能性としては、23×23×総就業者数のパターンの可能性が存在するわけだが、Wilsonに従えば現実に観察された通勤流動パターンは、この可能性の中の1つが起こったと理解されることになる。これらすべての可能性の中から、既知であるマクロ・スケールでの状態を制約条件として与えることによって、その流動パターンの可能性を大幅に絞りこむことができる。すなわち、千代田区には何人の就業者が常住するか、あるいは従業しているかという情報を得ているということは、推定した流動パターンについてもその周辺和が、既知の情報(すなわち実際の流動パターンの周辺和)と一致していなければならず、一致しない流動パターンは対象外として捨て去ることができるからである。

今、常住地  $i$  から従業地  $j$  への通勤流動を  $T_{ij}$  とすると、対象となる総就業者数は、 $T (= \sum_{ij} T_{ij}, i=1, \dots, 23; j=1, \dots, 23)$  であり、地区  $i$  の常住者数、地区  $j$  の従業者数の制約は以下の様与えられる。

$$\sum_j T_{ij} = O_i, \quad (1)$$

$$\sum_i T_{ij} = O_j. \quad (2)$$

これらの制約条件を満たす流動パターンはそれでもなお多数存在するわけだが、その中でどの流動パターンが最も現実に近いものかという基準が必要となる。それが次式のエントロピー最大化の基準である。

$$\text{MAX. } W(T_{ij}) = T / \prod_{ij} T_{ij}!. \quad (3)$$

すなわち、式(1)・(2)の制約条件を満たす流動パターンごとに統計力学的エントロピー(その流動パターンが生じる場合の数)を求め、そのエント

ロピーが最も大きい流動パターンを最も実現しやすいものとみなし、それを最良の推定値とする。これがエントロピー最大化法である。その場合、最大エントロピーを示した流動パターンの生起確率が最も高くなっている。

このエントロピー最大化の状態は、通勤流動を多項分布に従うと仮定したもとの最尤推定値に等しく、OD表を2次元分割表とみなした場合の、周辺確率から独立を仮定して求めた期待度数と一致する(矢野, 1992)。

空間的相互作用モデルの二重制約型エントロピー最大化モデルは、周辺和の制約に加え、総移動費用の制約が付け加えられている。これは、就業者の配分の際に距離減衰効果(常住地から就業地までの移動費用が少ないほど多くの就業者を配分する)を取り込むもので、従来の重力モデルとの整合性を図ったものである。したがって、周辺和のみを制約条件とするエントロピー最大化法は、距離関数を含めず(あるいは距離減衰効果が存在しないと仮定した場合)、発地区の放出性と着地区の吸引力だけから構成される重力モデルのようなものと解釈することもできる。

また近年、空間的相互作用モデルにみられる距離関数を、当該発地区・着地区間のペアーの関係を表わす指標とみなす見方もある(Yano, 1993)。例えば、対数線形モデルによって得られる一次の交互作用は、「真の」地区間の関係を表すもので、これを距離関数とみなして、エントロピー最大化モデルを適用すると、観察値とまったく一致する推定値を得ることができる。それゆえ、距離関数に替わって過去の流動パターンを用い、それを地区間の関係とみなして予測力を高めようとする試みもなされてきている(Snickars and Weibull, 1977)。したがって、周辺和のみを制約条件とするエントロピー最大化法は、かかる地区間の関係をすべて1.0(すべて等しい)とみなした場合に相当するものともいえる。

## 2) 2次元地理行列の場合

まず、 $n \times m$  の2次元の地理行列の要素  $E_{ij}$  ( $i=1, 2, \dots, n; j=1, 2, \dots, m$ ) をエントロピー最大化法で推定することを考えてみよう。いま当該行列の周辺和である、行和を  $R_i$ 、列和を  $C_j$  とする。そしてもし既にわかっている要素が存在する場合は、その既知の要素だけの、行和を  $P_i$ 、列和を  $Q_j$

とすると、 $i$ 行と $j$ 列のそれぞれの推計する要素の行和  $X_i$ と列和  $Y_j$ は、それぞれ以下になる（既知の要素が1つも存在しない場合は、 $P_i=Q_i=0$ ）。

$$X_i = R_i - P_i, \quad (4)$$

$$Y_j = C_j - Q_j. \quad (5)$$

そして推定しなければならない行列要素  $E_{ij}$ の( $i, j$ ) ペアアの集合を  $\alpha$  とすると、制約条件は

$$\sum_j E_{ij} = X_i, \quad ((i, j) \in \alpha), \quad (6)$$

$$\sum_i E_{ij} = Y_j, \quad ((i, j) \in \alpha), \quad (7)$$

となり、欠損している行列要素  $E_{ij}$  ( $(i, j) \in \alpha$ ) は、以下のように推定される。

$$E_{ij} = A_i X_i B_j Y_j F_{ij} \quad ((i, j) \in \alpha), \quad (8)$$

$$A_i = 1 / \sum_j B_j F_{ij}, \quad (9)$$

$$B_j = 1 / \sum_i A_i F_{ij}. \quad (10)$$

ただし、 $A_i$ 、 $B_j$ は均衡因子であり、 $F_{ij}$ は、下敷とする行と列の交互作用である。ここで下敷とは、行と列の関係を表し、OD 行列の場合は地区間の関係に対応する。この交互作用が存在しない ( $F_{ij}=1.0$ ) と仮定すると、周辺和のみを制約条件とするものとなる。その場合は、前述のように行と列が独立であると仮定した場合の期待値に一致する。

この推定方法は、 $F_{ij}$ を距離パラメータが特定された距離関数（例えば、指数関数）とみなすと、Wilsonの二重制約型エントロピー最大化モデルとまったく等価となる。また、これは地域産業連関表の推定を行うRAS法ともまったく同じもので、その場合は、対象とする地域の投入・産出の各産業間の関係が、より広域的な国家全体のもので一致しているという仮定のもとで、国家スケールの投入係数を  $F_{ij}$ とみなし、周辺和として当該地域の産業別の総投入量と総産出量を用いた、エントロピー最大化法に他ならない (Hewings, 1985; 矢野, 1991)。また、この方法を biproportional 法と呼ぶ場合もある (Macgill, 1981)。

### 3) 3次元地理行列の場合

実際、官庁統計などで、表章されないデータは、データ量が指数的に大きくなる3次元地理行列である。この場合、制約となる周辺和は、3次元のデータ立方体のそれぞれの側面にあたる2次元行

列となる。いま、 $n \times m \times l$ の3次元地理行列のセルを  $E_{ijk}$ とし ( $i=1, 2, \dots, n; j=1, 2, \dots, m; k=1, 2, \dots, l$ )、推定しなければならない行列要素  $E_{ijk}$ の( $i, j, k$ ) ペアアの集合を  $\alpha$  とすると、制約条件は、

$$\sum_k E_{ijk} = X_{ij}, \quad ((i, j, k) \in \alpha), \quad (11)$$

$$\sum_j E_{ijk} = Y_{ik}, \quad ((i, j, k) \in \alpha), \quad (12)$$

$$\sum_i E_{ijk} = Z_{jk}, \quad ((i, j, k) \in \alpha), \quad (13)$$

となり、欠損している行列要素  $E_{ijk}$  ( $(i, j, k) \in \alpha$ ) は、以下のように推定される。

$$E_{ijk} = A_{ij} X_{ij} B_{ik} Y_{ik} C_{jk} Z_{jk} F_{ijk}, \quad ((i, j, k) \in \alpha), \quad (14)$$

$$A_{ij} = 1 / \sum_k B_{ik} C_{jk} F_{ijk}, \quad (15)$$

$$B_{ik} = 1 / \sum_j A_{ij} C_{jk} F_{ijk}, \quad (16)$$

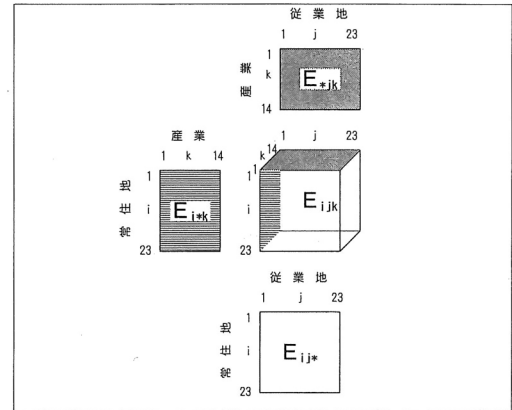
$$C_{jk} = 1 / \sum_i A_{ij} B_{ik} F_{ijk}. \quad (17)$$

ここで、2次元地理行列の場合と同様に、 $A_{ij}$ 、 $B_{ik}$ 、 $C_{jk}$ は均衡因子で、 $F_{ijk}$ は3変数の交互作用である。対数線形モデルでは、2変数の交互作用を1次の交互作用と呼び、3変数の交互作用を2次の交互作用と通常呼んでいる。

## III 3次元地理行列の推定

### 1) 東京都区部の産業別通勤流動

『平成2年国勢調査第6巻従業地・通学地別集計』からは、23区ごとの区部での14産業大分類別常住・就業者、と23区間の総通勤OD行列がわかっている（第1図）。ここで、産業大分類は、農業、林業、漁業、鉱業、建設業、製造業、電気・ガス・



第1図 産業別通勤流動の3次元地理行列

熱供給・水道業、運輸・通信業、卸売・小売業・飲食店、金融・保険業、不動産業、サービス業、公務（他に分類されないもの）の13産業に、分類不能の産業を含めた14分類である。

また『平成2年国勢調査第6巻従業地・通学地別集計』では、「人口30万人以上の市」（人口30万以上の市、人口30万未満の県庁所在市、12大都市の区）に関してのみ、常住地・従業地ごとに総数が50人以上の市区町村について表章されている（『平成2年国勢調査第6巻従業地・通学地別集計』の第6・7表）ている。したがって、観察された実際の産業別通勤OD行列と、エントロピー法で推計したものを比較することによって、推定値の適合を見ることが出来る。

ここでは、1) 周辺和以外の情報がすべて未知の場合(CASE1)、2) 対角要素にあたる、各産業ごとの区内就業業者数が既知の場合(CASE2)、について推定することにする。

OD行列の適合度に関しては、従来、様々な統計量が考案され、その長所・短所が指摘されているが、どの統計量が最も的確であるかは明らかではない(石川, 1988)。そこで、ここでは、決定係数( $R^2$ )、標準化平方根平均平方誤差(SRMSE)、カイ二乗統計値( $\chi^2$ )、情報獲得統計量に基づく、最小判別情報量(MDI)の、複数の統計量で適合の度合を比較することにする。なお、各統計量の取る範囲と、完全一致の場合の値は以下のようである。

- 1) 決定係数 ( $R^2$ )  
 $0.0 \leq R^2 \leq 1.0$ , 完全一致  $R^2 = 1.0$
- 2) 標準化平方根平均平方誤差 (SRMSE)  
 $0.0 \leq SRMSE \leq 1.0$ , 完全一致  $SRMSE = 0.0$
- 3) カイ二乗統計値 ( $\chi^2$ )  
 $0.0 \leq \chi^2 \leq +\infty$ , 完全一致  $\chi^2 = 0.0$
- 4) 最小判別情報量 (MDI)  
 $0.0 \leq MDI \leq +\infty$ , 完全一致  $MDI = 0.0$

エントロピー最大化法による推定結果の適合度は、第1表の様にまとめられる。CASE1、CASE2のいずれの推定結果も、概ね良好である。総数は若干の計算誤差が見られるがほぼ一致している。適合度の詳細をみると、対角要素が既知である制約条件を加えた、CASE2の方が、周辺和だけから推定したCASE1に比べ、かなり高い適合度を示していることがわかる。CASE1は観察値やCASE2に比べ標準偏差が若干小さく、平均化する

第1表 産業別通勤流動の推定結果の適合度

	観察値	CASE 1	CASE 2
サンプル数	7406	7406	7406
総数	4115728.0	4115722.8	4115725.6
平均	555.7	555.7	555.7
標準偏差	2657.8	2624.1	2656.2
決定係数 $R^2$	—	0.9959	0.9994
SRMSE	—	0.3096	0.1135
カイ二乗( $\chi^2$ )	—	94306.1	25442.0
MDI	—	93442.3	27449.7

傾向にあるといえる。

また、実際の空間的パターンを、千代田区への産業別就業業者数でみると、各産業の就業業者数の数値は完全に一致しないが、順位などの全体的な傾向は、実際の観察値とほぼ一致していることがわかる(第2表)。さらに、すべての業種ではないが、両推定値とも、観察値よりもバラツキが少ない傾向が看取される。

この方法を東京大都市圏におけるすべての市区町村間の産業別通勤流動に拡張するには、表章されているデータを既知のものとし、さらに各都道府県や12大都市の合計などをうまく制約条件として用いればよい。その結果、かなり現実に適合した推定値を得ることが可能であるといえる。

#### 2) 東京都区部の産業別・形態別の事業所数

『平成3年事業所統計』からは、各区ごとの産業大分類別事業所数と形態別事業所数(民営についてのみ)が表章されているが(同様に就業業者数)、各地区ごとのそのクロス集計は存在しない(区部全体のものはある)。ここで、産業大分類は、先の通勤の14分類のうち、農業、林業、漁業を農林漁業を1まとめにし、公務と分類不能の産業を除いた10分類であり、事業所の形態は、民営に限られたもので、店舗・飲食店、事務所、営業所、工場・作業所・鉱業所、自家用倉庫・自家用油槽所、外見上一般の住居と区別しにくい事業所、その他(旅館、病院、神社、学校等)の7分類である1)(第2図)。

そこで、1) 各区ごとの2次元地理行列とした場合(CASE1)、2) 区部全体の産業×形態の下敷として、当該地区の周辺和から2次元地理行列へのエントロピー最大化法、3) 産業大分類×形態×23区の3次元地理行列へのエントロピー最大

第2表 千代田区への産業別就業者数の推定結果

1) 観察値

	総数	農業	林業	漁業	鉱業	建設業	製造業	電気・ガス 熱供給・ 水道業	運輸・ 通信業	卸売・ 小売業 飲食店	金融・ 保険業	不動産業	サービス業	公務	分類不能
区内全体	357161	36	69	597	313	13800	56225	2192	20102	81427	40434	8142	97803	33200	2821
1 千代田	15817	1	0	4	2	410	1580	5	318	6670	425	1398	3457	1504	43
2 中央区	2607	0	0	229	2	119	378	21	115	655	226	68	579	206	9
3 港区	7902	1	0	8	14	204	843	78	548	1420	1183	207	1777	1567	52
4 新宿区	15700	0	2	33	20	669	2162	117	961	3167	1851	345	4290	1995	88
5 文京区	12128	1	1	8	8	384	1828	47	502	3012	1344	275	3524	1111	83
6 台東区	4948	0	0	4	4	209	757	35	226	1510	451	107	1284	319	42
7 墨田区	6756	1	0	4	2	358	1103	19	309	1946	594	130	1867	359	64
8 江東区	16521	3	1	14	6	924	2615	119	944	4188	1713	306	4458	1160	70
9 品川区	11333	1	0	8	10	383	1703	55	1100	2453	1257	219	2916	1124	104
10 目黒区	11305	3	1	5	15	344	1488	65	552	2054	1694	238	2870	1935	41
11 大田区	17233	0	1	42	17	694	3132	126	1297	3618	2324	305	4447	1121	109
12 世田谷区	38306	9	39	60	75	1156	6544	251	2036	7131	5073	757	10551	4288	336
13 渋谷区	10268	1	2	7	10	321	1537	107	524	1850	1539	195	2919	1193	63
14 中野区	15916	0	3	13	17	493	2509	76	880	3143	1832	323	4628	1902	97
15 杉並区	30221	5	6	24	32	1034	5672	242	1571	5610	4128	561	8896	2161	279
16 豊島区	11995	0	0	8	3	413	1955	55	811	2606	1401	261	3568	789	125
17 北区	15859	1	0	14	6	589	2127	58	1072	3847	1351	292	4103	2272	127
18 荒川区	6483	0	1	3	2	330	925	58	373	1809	668	131	1731	376	76
19 板橋区	21880	4	8	28	16	860	3650	129	1418	4767	2285	422	6155	1924	214
20 練馬区	26076	1	1	35	19	1156	4631	192	1365	5148	3279	500	7705	1768	276
21 足立区	20542	1	1	10	9	1020	3145	159	1219	5432	1927	396	5512	1521	190
22 葛飾区	15109	1	0	17	19	667	2543	74	839	3892	1461	290	4210	987	109
23 江戸川区	22256	2	2	19	5	1063	3398	104	1122	5499	2428	416	6356	1618	224
平均	15528.7	1.6	3.0	26.0	13.6	600.0	2444.6	95.3	874.0	3540.3	1758.0	354.0	4252.3	1443.5	122.7
標準偏差	8193.6	2.1	7.9	45.5	15.1	321.7	1518.0	64.6	474.9	1735.8	1133.9	270.0	2383.1	850.8	86.0
レンジ	35699.0	9.0	39.0	226.0	73.0	1037.0	6166.0	246.0	1921.0	6476.0	4847.0	1330.0	9972.0	4082.0	327.0

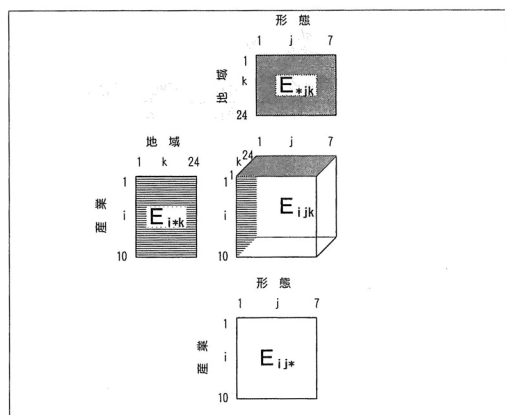
2) 周辺和以外の情報がすべて未知の場合(CASE 1)の推定値

	総数	農業	林業	漁業	鉱業	建設業	製造業	電気・ガス 熱供給・ 水道業	運輸・ 通信業	卸売・ 小売業 飲食店	金融・ 保険業	不動産業	サービス業	公務	分類不能
区内全体	357161	36	69	597	313	13800	56225	2192	20102	81427	40434	8142	97803	33200	2821
1 千代田区	15817	1	2	4	4	382	1690	12	539	5739	786	1044	3744	1820	51
2 中央区	2607	0	0	121	0	79	345	15	153	810	81	72	586	332	13
3 港区	7902	1	2	14	4	202	821	41	403	2021	703	318	1975	1336	62
4 新宿区	15700	2	3	32	15	568	2179	118	829	3704	1443	364	4135	2207	101
5 文京区	12128	1	1	11	7	387	1791	54	546	3064	1358	340	3342	1138	88
6 台東区	4948	0	0	8	2	249	981	22	216	1350	318	144	1295	307	56
7 墨田区	6756	1	0	5	2	289	1342	47	382	1755	590	132	1730	426	54
8 江東区	16521	1	1	22	7	812	2793	111	1026	4120	1640	322	4382	1206	80
9 品川区	11333	1	1	14	6	445	1829	41	692	2802	1130	254	3039	935	143
10 目黒区	11305	1	2	5	23	367	1658	59	499	2478	1262	271	3137	1484	60
11 大田区	17233	2	1	43	17	733	3111	97	878	3935	2206	372	4772	928	139
12 世田谷区	38306	6	41	60	65	1281	5954	191	1971	7751	4988	819	11165	3697	319
13 渋谷区	10268	1	2	15	9	372	1452	80	509	2353	1182	296	2870	1042	86
14 中野区	15916	1	2	17	16	553	2342	112	855	3496	1904	305	4473	1764	77
15 杉並区	30221	3	4	35	41	997	4968	184	1457	6010	4217	624	9245	2188	249
16 豊島区	11995	1	0	18	23	519	2014	74	749	2790	1149	269	3398	847	144
17 北区	15859	1	0	16	10	615	2350	115	1018	3499	1614	259	4165	2059	139
18 荒川区	6483	0	0	9	1	319	1128	52	394	1601	723	124	1660	414	59
19 板橋区	21880	2	3	33	12	973	3476	142	1463	4479	2703	390	5803	2213	187
20 練馬区	26076	4	3	44	20	1063	4281	136	1572	4816	3801	467	7439	2195	236
21 足立区	20542	2	2	14	9	937	3330	222	1424	4584	2231	319	5490	1791	187
22 葛飾区	15109	2	0	22	10	620	2625	102	1004	3381	1670	260	4092	1228	94
23 江戸川区	22256	3	0	37	10	1039	3767	165	1523	4891	2736	377	5866	1643	199
平均	15528.7	1.6	3.0	26.0	13.6	600.0	2444.6	95.3	874.0	3540.3	1758.0	354.0	4252.3	1443.5	122.7
標準偏差	8193.6	1.4	8.2	24.7	14.2	316.6	1345.4	57.4	482.5	1615.4	1215.2	214.4	2437.8	781.1	74.8
レンジ	35699.0	5.8	41.0	116.6	65.0	1201.5	5608.2	210.1	1818.6	6940.8	4906.2	972.2	10578.6	3390.0	306.3

3) 各産業の自区内就業者数が既知の場合(CASE 2)の推定値

	総数	農業	林業	漁業	鉱業	建設業	製造業	電気・ガス 熱供給・ 水道業	運輸・ 通信業	卸売・ 小売業 飲食店	金融・ 保険業	不動産業	サービス業	公務	分類不能
区内全体	357161	36	69	597	313	13800	56225	2192	20102	81427	40434	8142	97803	33200	2821
1 千代田区	15817	1	0	4	2	410	1580	5	318	6670	425	1398	3457	1504	43
2 中央区	2607	0	0	208	1	80	370	18	142	610	252	59	599	258	11
3 港区	7902	1	0	6	8	187	917	68	538	1554	877	202	2014	1483	47
4 新宿区	15700	2	2	26	15	589	2198	127	850	3315	1926	348	4385	1829	88
5 文京区	12128	1	1	9	6	363	1879	49	538	2751	1409	257	3584	1196	86
6 台東区	4948	1	0	6	2	212	760	28	252	1377	517	111	1306	333	44
7 墨田区	6756	1	0	4	2	275	1202	38	403	1762	660	131	1737	490	52
8 江東区	16521	1	2	16	7	815	2683	144	1195	4011	1697	297	4284	1295	75
9 品川区	11333	1	1	9	7	406	1838	48	740	2695	1226	231	3013	996	123
10 目黒区	11305	1	1	5	22	339	1651	52	469	2438	1382	241	3155	1499	51
11 大田区	17233	2	0	36	16	707	2959	118	1184	3786	2141	344	4670	1149	121
12 世田谷区	38306	6	46	51	67	1311	6040	206	1858	7853	4762	822	11220	3730	334
13 渋谷区	10268	1	1	12	7	318	1483	80	505	2277	1374	238	2982	910	79
14 中野区	15916	2	1	15	16	547	2383	87	863	3563	1833	313	4556	1652	85
15 杉並区	30221	4	4	32	40	1025	5122	188	1474	6123	4103	613	9153	2085	256
16 豊島区	11995	1	0	16	24	468	1903	85	715	2728	1396	241	3508	779	132
17 北区	15859	1	0	13	9	632	2482	89	983	3499	1502	258	4048	2205	138
18 荒川区	6483	1	0	7	1	330	1073	85	419	1539	704	120	1694	450	59
19 板橋区	21880	2	3	30	13	959	3634	129	1437	4535	2514	394	5807	2218	205
20 練馬区	26076	4	3	38	20	1127	4376	140	1449	5037	3456	503	7307	2355	263
21 足立区	20542	2	2	12	9	1006	3288	174	1412	4742	2100	343	5400	1849	205
22 葛飾区	15109	1	0	18	10	616	2639	96	937	3529	1603	283	4047	1222	108
23 江戸川区	22256	2	0	26	10	1079	3766	138	1421	5035	2576	395	5877	1716	216
平均	15528.7	1.6	3.0	26.0	13.6	600.0	2444.6	95.3	874.0	3540.3	1758.0	354.0	4252.3	1443.5	122.7
標準偏差	8193.6	1.2	9.3	40.7	14.4	337.1	1388.0	53.3	466.8	1763.1	1113.9	276.5	2418.6	773.7	82.9
レンジ	35699.0	5.4	46.3	203.8	65.8	1230.6	5670.4	201.0	1716.5	7242.6	4510.3	1338.8	10620.9	3471.3	323.8

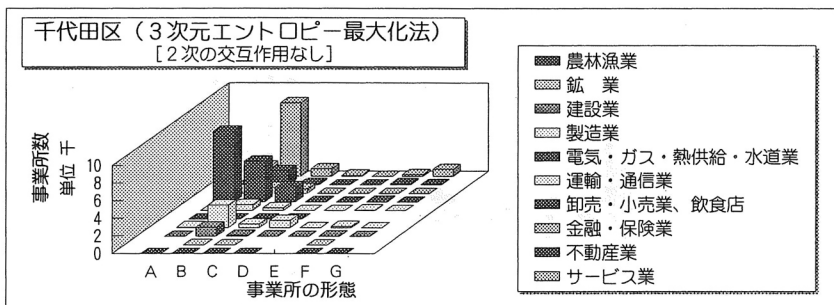
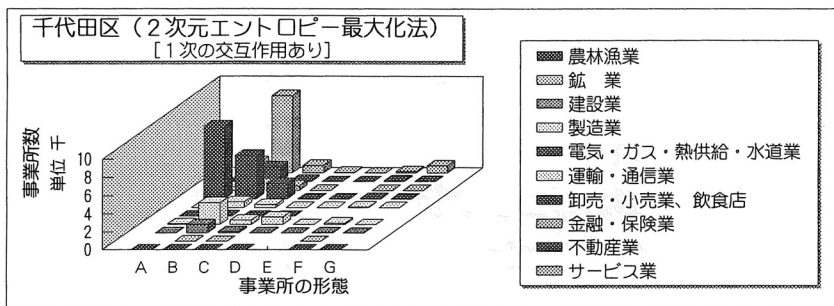
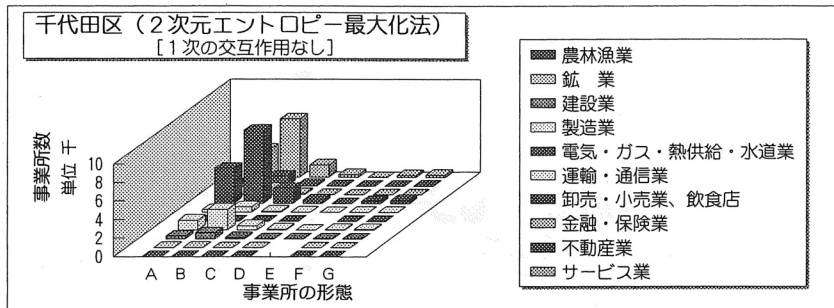
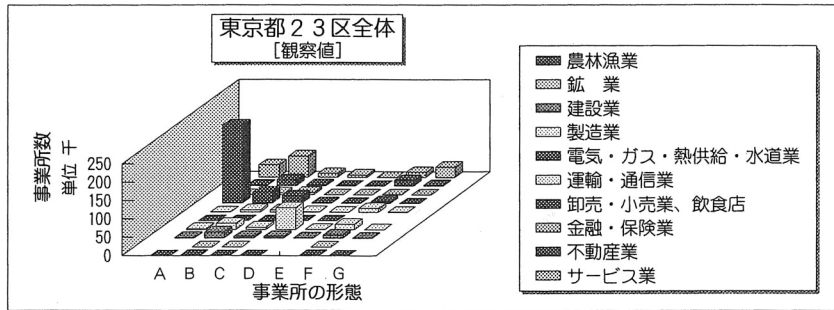
注) 陰影部分は既知のデータである。



第2図 産業別・形態別事業所数の3次元地理行列

化法を適用し、その推定値を検討する。ただし、産業別通勤OD行列のように実際のデータが存在しないので、観察値との適合を見ることはできない。3つの場合を比較すると、制約の大きいCASE3, CASE2, CASE1の順に、現実にもっと近い

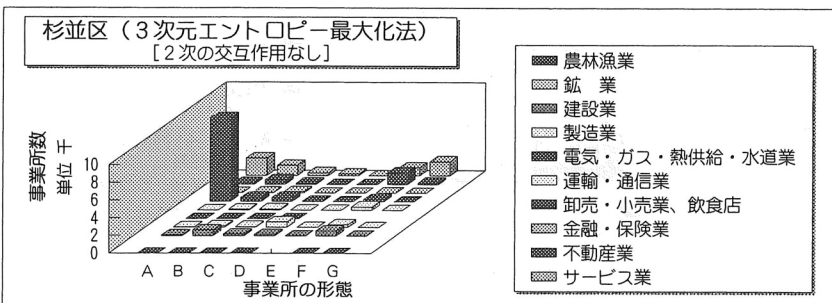
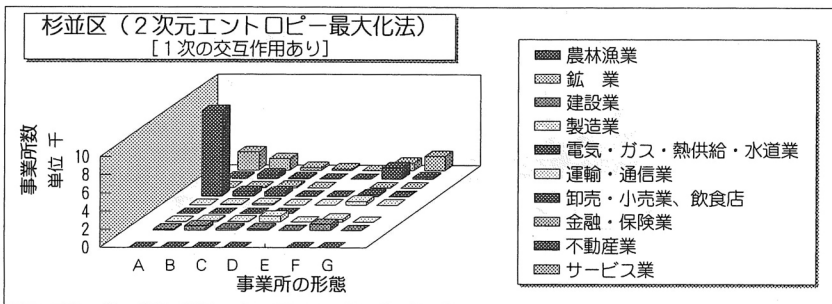
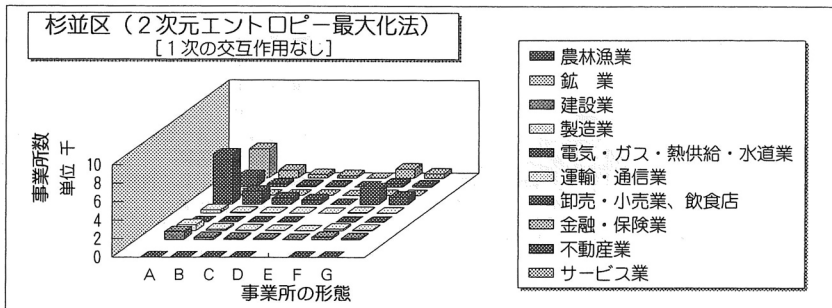
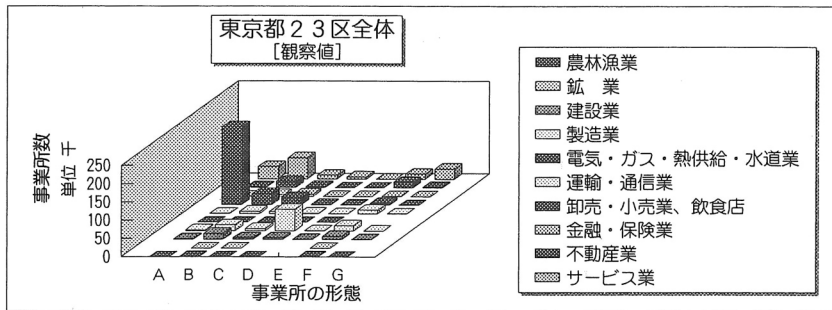
推定値であると考えられる。CASE1は、産業と形態に関係がまったく見られないと仮定した場合であり、頻度分布が均等化する傾向にあることがわかる。3つの場合で推定された、産業別・形態別事業所数を都心の業務地区である千代田区と、山の手の住宅地区である杉並区を例に比較すると、23区全体で見た、産業と形態の関係が大きく異なることがわかる(第3・4図)。CASE3の場合が現実にもっと近いものであるとすると、卸売・小売業、飲食店が、千代田区では店舗・飲食店が最も多いが、事務所や営業所の事業所も多くみられるのに対し、杉並区では店舗・飲食店がほとんどである。さらにサービス産業をみると、千代田区では営業所の事業所が卓越するのに対し、杉並区では店舗・飲食店あるいはその他の事業所(このカテゴリーには前述のように、病院や学校などが含まれる)が多い。このような相違は両区の業務地区・住宅地区といった違いに対応するが、同じ産業であっても形態が地域によって大きく異なることがデータの的に示されることが明らかになった。さら



事業所の形態：A. 店舗・飲食店；B. 事務所；C. 営業所；D. 工場・作業所  
E. 倉庫・油槽所；F. 区別難事業所；G. その他

第3図 千代田区の産業別・形態別事業所数の推定結果

注) 事業所の産業分類は、手前から奥に向かって凡例の順で並んでいる。



事業所の形態：A. 店舗・飲食店；B. 事務所；C. 営業所；D. 工場・作業所  
E. 倉庫・油槽所；F. 区別難事業所；G. その他

第4図 杉並区の産業別・形態別事業所数の推定結果

注) 事業所の産業分類は、手前から奥に向かって凡例の順で並んでいる。



にいえば、いわゆるサービス産業が、事業所の形態によって、消費者を対象とするものなのか事業所を対象とするものなのかをある程度見分けることが可能となることを示唆するものといえる。このような地域差は、従来の官庁統計のみに基づいたマクロな分析では明らかにすることができなかったものであり、データの非集計化によって初めて把握できるものであるといえる。

#### IV おわりに

本研究は、実際の統計データを用いて、エントロピー最大化法による欠損データの復元の有効性を検討した。

産業別通勤 OD 行列の復元に関しては、対角要素を既知としない場合と既知とする場合について、推定値と実際の観察値との比較を行なった。その結果、対角要素を既知とした場合に適合度が高くなることが明らかとなった。また、産業別・形態別事業所数の推定では、2次元地理行列として、産業と形態の関係が独立の場合、既知の場合、3次元地理行列として推定した場合、の3つを比較したが、3次元地理行列として推定した場合が最も制約が大きく、現実にも近いものであると考えられた。

さらに、制約条件の違いからエントロピー最大化法による推定結果の特徴を理解することもできる。すなわち、2次元行列よりも3次元行列の方が制約が厳しいこと、対角要素など既知のデータの存在によって、推定結果が飛躍的に向上すること、などが明らかとなった。

本研究で用いた、『平成2年国勢調査』及び『平成3年事業所統計』のデータは、立命館大学文学部地理学教室の磁気テープを利用した。また、『平成3年事業所統計』のデータに関しては、立命館大学の古賀慎二氏にご教示頂いた、記して感謝する次第である。

(立命館大学・文学部)

#### 注

- 1) 事業所統計の産業分類に関しては、非収録データから大分類よりもさらに細かい中分類で同様の推計を行なうことができる。

#### 文献

- 石川義孝(1988):『空間的相互作用モデル—その系譜と体系—』地人書房, 254ページ。
- エヴェリット著, 山内光哉監訳(1980):『質的データの解析』新曜社, 138ページ。
- 杉浦芳夫(1986):空間的相互作用モデルの近年の展開—重力モデルからエントロピー最大化型モデルへ—。野上道男・杉浦芳夫著『パソコンによる数理地理学演習』古今書院, 138-185。
- 杉浦芳夫(1989):電報流動からみた昭和初期のわが国の機能地域。理論地理学ノート6, 35-49。
- 松田紀之(1988):『質的情報の多変量解析』朝倉書店, 214ページ。
- 矢野桂司(1991):Lowryモデル族と地域産業連関分析の統合—豊田市の都市内部システムへの適用—。地理学評論, 61A, 1-24。
- 矢野桂司(1991):東京大都市圏内部の年齢階級人口移動パターン。地理学評論, 62A, 269-288。
- 矢野桂司(1991):一般線形モデルによる空間的相互作用モデルの統合。地理学評論, 64A, 367-387。
- Chilton, R. and Poet, R. R. W. (1973): An entropy maximising approach to the recovery of detailed migration patterns from aggregated census data. *Environment and Planning*, 5, 135-146。
- Hewings, G. J. D. (1985): *Regional Input-output Analysis*(Scientific Geography Series 6). Sage, 96p.
- Macgill, S. (1977): Theoretical properties of bi-proportional matrix adjustments. *Environment and Planning A*, 9, 687-701。
- Snickars, F. and Weibull, J. W. (1977): A minimum information principle: theory and practice. *Regional Science and Urban Economics*, 7, 137-168。
- Wilson, A. G. (1981): *Geography and the Environment: Systems Analytical Methods*. John Wiley & Sons, 297p.

補遺: エントロピー最大化法の FORTRAN プログラム

本研究で用いたプログラムは、1) ENT2.F(2次元地理行列)と、2) ENT3.F(3次元地理行列)、である。

入力データは、1) 周辺和、2) 既知データ、3) 下敷となる交互作用、のそれぞれのベクトルあるいは行列である。2)と3)はオプションである。2)の既知のデータが存在する場合は、未

知のセルには-1を入れたもの (IOPT1), OD 行列を想定し, 対角要素だけが既知の場合 (IOPT2) に分けられる。既知のデータが無い場合は, IOPT1=0である。また, 交互作用に関しては, 2次元行列の場合は2次元の1次の交互作用を, 3次元行列の場合は3次元の2次の交互作用を用意し, IOPT2=1とする。無い場合は, IOPT2=0とする。

希望者には, 上記のエントロピー最大化法のプログラムと本研究で用いたサンプル・データは, 実費で配布します。

連絡先 〒603 京都市北区等持院北町56-1  
立命館大学文学部地理学教室 矢野桂司  
TEL 075-465-1111 FAX 075-465-8296