

# *n* 元度数分布の数量化について

京 極 純 一

(1967年5月受付)

## On a Method of Quantifying *N*-way Frequency Tables

Jun-ichi KYOGOKU

We have a well-known method of "quantifying" 2-way frequency tables. The method was developed by Dr. C. Hayashi and has found many applications in various research papers. This paper concerns about one of possible generalizations of Hayashi's method to *n*-way frequency tables.

A *n*-way frequency table is based on *n*-sets of categorical items. The purpose of "quantification" is to find and assign such numerical values to each categorical items throughout *n* sets that will represent, most efficiently in a certain operational sense, the mutual nearness or remoteness between and among these categorical items.

If we assign specific numerical values from without to each item, every set of these items is represented by a respective vector in a *N*-dimensional space, *N* being the number of total frequency. Also, we can set the length of each vector to be a standard deviation in each set by putting the origin in each set equal to zero.

Having *n* end points of *n* vectors scattered in a *N*-dimensional space, we turn to the amount of clustering of these *n* end points. As a measure of clustering, *D*, we take a total sum of squares of *n* distances between each end point and their centroid. In turn, *D* is able to be explicitly written in terms of mutual distances between *n* end points,  $d_{ij}$ ,

$$D = \frac{1}{2n} \sum_i \sum_{(i \neq j)} d_{ij}^2,$$

being a distance between the *i*-th and *j*-th end points. By the cosine law and the definition of the length of each vector,

$$d_{ij}^2 = \sigma_i^2 + \sigma_j^2 - 2\sigma_{ij},$$

where  $\sigma_i^2$  is the variance of the *i*-th item, and  $\sigma_{ij}$  is the covariance between the *i*-th and the *j*-th items. After substitution *D* is shown in terms of  $\sigma_i^2$ ,  $\sigma_j^2$  and  $\sigma_{ij}$  as follows,

$$D = \frac{n-1}{n} \sum_i \sigma_i^2 - \frac{n}{2} \sum_i \sum_{(i < j)} \sigma_{ij}.$$

Noting that

$$\frac{n-1}{n} \sum_i \sigma_i^2$$

represents only a scale factor to be determined freely from without, we now turn to the second measure  $\lambda$ , defined as

$$\lambda = \frac{2 \sum_{\substack{i < j \\ (i < j)}} \sigma_{ij}}{(n-1) \sum_i \sigma_i^2}.$$

Naturally, we get

$$D = \frac{n-1}{n} \sum_i \sigma_i^2 (1 - \lambda).$$

The purpose of "quantification" prescribes that

$$D \rightarrow \min. \text{ or, equivalently, } \lambda \rightarrow \max.$$

is to be solved. Now, in terms of matrix and quadratic form notation,

$$N\sigma_i^2 = x_i' F_i x_i, \quad N\sigma_j^2 = x_j' F_j x_j, \quad N\sigma_{ij} = x_i' A_{ij} x_j = x_j' A_{ji} x_i.$$

Definitions of vectors and matrices are as follows. Column vectors  $x_i$  and  $x_j$  give two sets of numbers to be assigned to the  $i$ -th and  $j$ -th sets of categorical items respectively.  $F_i$  is a diagonal matrix, the diagonal elements of which correspond to the so-called marginal frequencies in the  $i$ -th set of categorical items.  $F_j$  is similarly defined.  $A_{ij}$  is a matrix giving a 2-way frequency table between the  $i$ -th and the  $j$ -th sets of categorical items.

Then, in terms of such notation,

$$\lambda = \frac{2 \sum_{\substack{i < j \\ (i < j)}} x_i' A_{ij} x_j}{(n-1) \sum_i x_i' F_i x_i} \longrightarrow \max.$$

is the problem to be solved.

Setting the partial derivative of  $\lambda$  w.r.t  $x$ 's equal to zero  $n$  times, we get a matrix equation,

$$Bx = (n-1)\lambda x,$$

the definitions of matrix  $B$ , submatrices  $B_{\alpha\alpha}$ ,  $B_{\beta\beta}$ ,  $B_{\alpha\beta}$  and  $B_{\beta\alpha}$  being

$$\begin{aligned} [B_{\alpha\alpha}] &= [0], \quad [B_{\beta\beta}] = [0], \\ [B_{\alpha\beta}] &= [F_\alpha^{-1} A_{\alpha\beta}], \quad [B_{\beta\alpha}] = [F_\beta^{-1} A_{\beta\alpha}], \\ [B] &= \begin{bmatrix} [0] & \cdots & [B_{\alpha\beta}] \\ \vdots & & \vdots \\ [B_{\beta\alpha}] & \cdots & [0] \end{bmatrix} \end{aligned}$$

( $\alpha$  and  $\beta$  represent respectively the  $\alpha$ -th and  $\beta$ -th sets of categorical items.)

So, the problem of "quantifying" a  $n$ -way frequency table reduces to a task of numerical calculation of eigenvalues and eigenvectors of matrix  $B$ . It is also proved that the successive eigenvectors of matrix  $B$  give not only uni-dimensional numerical values but also multi-dimentional numerical values, with validity, to be assigned to each categorical items throughout  $n$ -sets.

## 1. まえがき

1.1) 社会事象に関するデータについていえば、すでに数量化された測定値の形で与えられているものは、必ずしも、多くない。定性的な分類区分を設定し、各区分項目について度数分布を示しただけのものも、少なくない。手近な例としては、たとえば、多くの世論調査の集計表や各種の選挙の開票結果一覧などがある。

こうした定性的な分類区分に基く度数分布表に対して、まず(1)分析目的を特定し、つぎに(2)分析目的に一致するよう、測度を操作的に定義し、(3)この測度に応じて算出された数値を、分類区分の各項目に、人工的に、わりあてる；この一連の方法が林知己夫氏に始まる数量化理論の基本的着想である。

1.2) 定性的な分類区分に基く度数分布表という形で与えられるデータは、1元(1-way)分布表(いわゆる周辺(marginal)度数表)あるいは2元(2-way)分布表(いわゆるクロス分布表)という形をとっている場合が多い。しかし、3元(3-way)およびそれ以上、一般にn元(n-way)にわたる分布表として示される場合も少なくない。

周知のように、2元度数分布表に対する数量化の方法は、林知己夫氏によって、数量化理論第Ⅲ類(外的基準のない場合の数量化)という形で与えられている。しかし、一般にn元にわたる度数分布に対する数量化の方法は、まだ、開発されていない。

本稿は、一般にn元にわたる度数分布を数量化するとき可能な方法のひとつについて述べることを、その内容としている。

1.3) 本稿については、林知己夫氏に何かと御教示を頂いた。厚くお礼申上げたい。

## 2. 測度の選択

2.1) 本稿においては、2元度数分布の数量化のために用いられている林氏の方法を、何らかの形で、拡張してn元度数分布を数量化する方法に到達することを目標とする。そのためには、2元度数分布を数量化するための方法を再検討することから出発しよう。

2.2) 左図のような2元度数分布表が与えられているとき、数量化の方法においては、分析目的として、行および列の各元における分類区分の各項目に、

各元の内部における共通性あるいは同質性を示す数値を、人工的に、わりあてる、ということが定められている。また、そのための測度として、外分散と全分散の比または相関比あるいは相関係数が採用されている。

(イ)  $f_i$  をウェイトにして  $x_i$  の平均がゼロになるという条件の下で、表2.2の行の元の  $i$  項目に  $x_i$  を与えるとすれば、列の元の  $j$  項目には  $\bar{x}_j$  が、 $x_i$  の線型和として、定められる。列の元における各項目の間の共通性ないし同質性の指標として  $\bar{x}_j$  の分散、いわゆる外分散を考え、この外分散のスケールを定めるものとして  $x_i$  の全分散をとり、両者の比を考えることから、 $x$  に関する相関比の自乗が測度として採用される。(分散の比を考える限り、平均ゼロという原点条件は解の一般性を損わない。) すなわち、

	..... $j$ .....	$\Sigma$
$i$	..... $f_{ij}$ .....	$f_i$
$\Sigma$	..... $f_j$ .....	$N$

$$\begin{aligned}\sum_i f_i x_i = 0, \quad \bar{x} = 0; \quad \sigma_x^2 = \frac{1}{N} \sum_i f_i x_i^2 \\ \bar{x}_j = \frac{1}{f_j} \sum_i f_{ij} x_i; \quad \sigma_{xj}^2 = \frac{1}{N} \sum_j f_j \bar{x}_j^2 \\ \eta_x^2 = \frac{\sigma_{xj}^2}{\sigma_x^2} \longrightarrow \max.\end{aligned}$$

(ロ) 同様にして、表の 2.2 列の元の  $j$  項目に  $y_j$  を与えるとすれば、行の元の  $i$  項目に  $\bar{y}_i$  が定められ、(イ) と全く同様の立論によって、 $y$  に関する相関比の自乗という測度がえられる。

$$\begin{aligned}\sum_j f_j y_j = 0, \quad \bar{y} = 0; \quad \sigma_y^2 = \frac{1}{N} \sum_j f_j y_j^2 \\ \bar{y}_i = \frac{1}{f_i} \sum_j f_{ij} y_j; \quad \sigma_{yi}^2 = \frac{1}{N} \sum_i f_i y_i^2 \\ \eta_y^2 = \frac{\sigma_{yi}^2}{\sigma_y^2} \longrightarrow \max.\end{aligned}$$

(ハ) 同様にして、表 2.2 の行の元の  $i$  項目に  $x_i$ 、表 2.2 の列の元の  $j$  項目に  $y_j$  を与えるとすれば、行と列との両方の元にわたる同時的な共通性ないし同質性を示す測度として、 $x$  と  $y$  との間の相関係数が採用される。

$$\begin{aligned}\sum_i f_i x_i = 0, \quad \bar{x} = 0; \quad \sigma_x^2 = \frac{1}{N} \sum_i f_i x_i^2 \\ \sum_j f_j y_j = 0, \quad \bar{y} = 0; \quad \sigma_y^2 = \frac{1}{N} \sum_j f_j y_j^2 \\ \sigma_{xy} = \frac{1}{N} \sum_i \sum_j f_{ij} x_i y_j \\ \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \longrightarrow \max.\end{aligned}$$

(ニ) 周知のように、(イ)(ロ)(ハ)の3方法による解は、相関比あるいは相関係数についても、行に与えられる  $x$  および列に与えられる  $y$  のそれぞれのベクトルについても、すべて等値である。

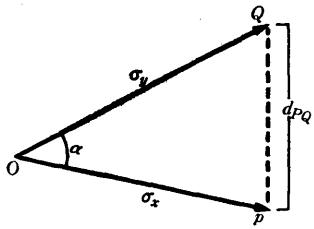
(ホ) 以上みたように、2元度数分布を数量化する方法は、分散成分の比の最大化という考え方を、測度に関する根本的な着想としている。

しかも、分散成分の比の最大化という考え方とは、行列算における2次形式の比の最大化という演算につながり、やがて対称行列の固有値および固有ベクトルの算出という数値計算につながっている。そこから、単に1次元に止らず多次元にわたる数量化が可能となる道が開けている。

2.3)  $n$  元度数分布を数量化する方法として、2元度数分布を数量化する方法を、直接に、拡張しようとすれば、当然に、 $n$  元にわたる分散成分の比の最大化ということを考えなければならない。しかし、 $n$  元にわたる分布度数の含む情報をそのままもれなく、分散成分といふ形で、利用するには、 $n$  次元行列の理論を応用しなければならないであろう。しかも、行と列とをもつといふ意味で2次元の行列、すなわち、ふつうの行列の理論の現状に比べると、 $n$  次元行列の理論の現状は、いまだ未開発である。すなわち、2次形式の拡張としての $n$  次形式、あるいは、 $n$  次元行列の固有値、固有ベクトルなどに関する理論は、残念ながら、まだ整備されていない。

したがって、 $n$  元度数分布を数量化する方法については、 $n$  元にわたる分布度数の含む情報を、直接にもれなく、利用することを断念し、着想の方向をかえ、別の解法ルートを探さなければならない。

2.4) ここで別の解法ルートを探す手掛りをえるために、2元度数分布の場合の特徴を、ベクトルという再度から、検討してみよう。



総度数を  $N$  として、 $N$  次元空間に(2.2)節にみた2元度数分布の行に対応するベクトル  $\vec{OP}$  と列に対応するベクトル  $\vec{OQ}$  と2本のベクトルを、図のように、引く。(2.2)節と同様に、 $f_i, f_j$  によってそれぞれウェイトをつけた  $x_i, y_j$  の平均をそれぞれゼロに定めておけば、 $\vec{OP}$  の長さは  $\sigma_x$ 、 $\vec{OQ}$  の長さは  $\sigma_y$  である。また、 $\angle POQ = \alpha$  とすれば、

$$\cos \alpha = \gamma_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad \dots \dots \dots \quad (2.4.1)$$

また、 $P$  と  $Q$  との間の距離を  $d_{PQ}$  とすれば、余弦法則から

$$d_{PQ}^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy} \quad \dots \dots \dots \quad (2.4.2)$$

である。

さて、2元度数分布を数量化する方法は、すでに検討したように、

$$\cos \alpha \rightarrow \max. \quad \dots \dots \dots \quad (2.4.3)$$

を測度として採用するものである。

ここで  $\triangle POQ$  に移して考えれば、 $\vec{OP}$  と  $\vec{OQ}$  という2つのベクトルの長さはスケール・ファクターとして外から人工的に定められるものであるから、 $\triangle POQ$  に関する限り、

$$\cos \alpha \rightarrow \max. \quad \dots \dots \dots \quad (2.4.3)$$

$$\alpha \rightarrow \min. \quad \dots \dots \dots \quad (2.4.4)$$

$$\triangle POQ \rightarrow \min. \quad \dots \dots \dots \quad (2.4.5)$$

$$d_{PQ} \rightarrow \min. \quad \dots \dots \dots \quad (2.4.6)$$

という4つの解法は、すべて同等である。

すなわち、2元度数分布を数量化する方法を、 $n$  元度数分布を数量化する方法に拡張する場合、少くとも、(2.4.4)式、(2.4.5)式、(2.4.6)式のそれぞれから着想する3通りの解法ルートが考えられる。

2.5) (イ) 角  $\alpha$  あるいは  $\angle POQ$  の最小化、というルートから着想すれば  $n$  元に対応する  $n$  本のベクトルが  $n$  次元球面において截りとる曲面が中心  $O$  においてもつ立体角をまず考え、ついでこの立体角を最小化するという形で、測度が考えられる。

(ロ)  $\triangle POQ$  の面積の最小化、というルートから着想すれば、 $n$  元に対応する  $n$  本のベクトルが作る多面体の体積をまず考え、ついでこの体積を最小化するという形で、測度が考えられる。

この場合、具体的には、分散・共分散行列の作る行列式、いわゆる一般化分散の値を最小化することを考えればよい。

(ハ)  $d_{PQ}$  の最小化、というルートから着想すれば、 $n$  元に対応する  $n$  本のベクトルの  $n$  個の終点の集中度 (clustering) をまず考え、何らかの形でこの集中度に関する測度を定義し、その測度を最小化することを考えればよい。

(ニ) 念のためつけ加えておけば、上の3ルートとも、2元分布度数が伝える情報を出発点として利用する、という基本特性を、共通に、もっている。 $n$  元にわたる分布度数が含む情報

を、直接にもれなく、利用する、という着想に照らせば、その限りで上の 3 ルートとも不十分な間に合わせのものである。

2.6) 上記 3 ルートにわたる着想のうち、とりあえずどれを選択するか、これが次の問題である。

本稿では、(イ) 定義式がなるべく単純であること、(ロ) 偏微分による解の演算がなるべく単純であること、(ハ) 2 次形式の比から固有値、固有ベクトルの算出に移行するものとなり、多次元の数量化が単純であること、という 3 点の単純性を選択の基準とした。

その結果、本稿においては、3 上記ルートのうち第 3 の着想、すなわち、 $n$  本のベクトルの  $n$  個の終点の集中度を考える、という着想を選択することにした。(1.2) 節で述べたように、これは  $n$  元にわたる度数分布を数量化するとき可能な方法のひとつであって、それ以上のものではない。

### 3. 測度の定義

3.1) まず、2 元度数分布を数量化する方法と同様に、 $n$  元度数分布を数量化するときも、 $n$  元のうち各元における分類区分の各項目に、 $n$  元にわたって同時に、共通性あるいは同質性を示す数値を、外から人工的に、わりあてる、という分析目的を設定する。

3.2) この分析目的に一致する考え方として、これまで検討してきた事情からして、 $n$  元に対応する  $n$  本のベクトルの  $n$  個の終点の集中度 (clustering) を、まず、手掛りとしてとりあげる。

つぎに、この  $n$  個の終点が、その重心のまわりでもつ距離の平方和  $D$  をこの集中度を数量的に表示するものとしてとりあげ、数量化するための第一次的な指標として考える。

3.3)  $n$  元の各元内の各区分項目に数値をわりあてれば、総度数を  $N$  として、 $N$  次元空間に、 $n$  元に対応する  $n$  本のベクトルを引くことができる。以下、一般に  $n$  元のうち  $i$  元目に対応するベクトルの終点を点  $i$  で表示するときめておく。また、以下の議論のため、(2.4) 節にならって、ベクトルの長さはすべて、その対応する元における標準偏差に等しい、ときめておく。

さて、 $n$  個の終点だけに注目すれば、これらの点は、たかだか、 $n$  次元空間におさまる。この  $n$  次元空間において、点  $i$  の座標を

$$(a_{1i}, a_{2i}, \dots, a_{li}, \dots, a_{ni})$$

$n$  個の終点の重心  $O$  の座標を

$$(\bar{a}_{1..}, \bar{a}_{2..}, \dots, \bar{a}_{l..}, \dots, \bar{a}_{n..})$$

で示す。 $(a_{li})$  は第  $l$  次元目における点  $i$  の座標であり、 $\bar{a}_{l..}$  は第  $l$  次元目における重心  $O$  の座標である。) 点  $i$  と重心  $O$  との間の距離を  $d_{i0}$  とすれば、(3.2) 節でとりあげた  $D$  について

$$D = \sum_i d_{i0}^2 = \sum_i \sum_l (a_{li} - \bar{a}_{l..})^2 = \sum_i \sum_l a_{li}^2 - n \sum_l \bar{a}_{l..}^2 \quad \dots \dots \dots \quad (3.3.1)$$

となる、さらに、点  $i$  と点  $j$  との間の距離を  $d_{ij}$  とすれば、 $D$  は、次のようにして、 $d_{ij}$  を使って表現される。

$$\begin{aligned} 2nD &= 2n \sum_i \sum_l a_{li}^2 - 2n^2 \sum_l \bar{a}_l^2 = \sum_i \sum_j (\sum_l a_{li}^2 + \sum_l a_{lj}^2 - 2 \sum_l a_{li} a_{lj}) \\ &= \sum_i \sum_j \{ \sum_l (a_{li} - a_{lj})^2 \} = \sum_i \sum_j d_{ij}^2 \end{aligned}$$

距離については、当然に、一般に  $d_{ii}=0$  であるから

$$2nD = \sum_i \sum_{\substack{j \\ (i \neq j)}} d_{ij}^2 \quad \dots \quad (3.3.2)$$

さて、ベクトル  $i$  の長さが一般に  $\sigma_i$  であるとすれば、すでに (2.4.2) 式でみたように

$$d_{ij}^2 = \sigma_i^2 + \sigma_j^2 - 2\sigma_{ij} \quad \dots \quad (3.3.3)$$

であるから

$$\sum_i \sum_{\substack{j \\ (i \neq j)}} d_{ij}^2 = \sum_i \sum_{\substack{j \\ (i \neq j)}} (\sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}) = 2 \{ \sum_i (n-1)\sigma_i^2 - 2 \sum_i \sum_{\substack{j \\ (i < j)}} \sigma_{ij} \} \quad \dots \quad (3.3.4)$$

である。

(3.3.4) 式を (3.3.2) 式に代入して整理すれば、

$$D = \frac{n-1}{n} \sum_i \sigma_i^2 - \frac{2}{n} \sum_i \sum_{\substack{j \\ (i < j)}} \sigma_{ij} \quad \dots \quad (3.3.5)$$

となる。

3.4) すでにみたように、 $D$  は  $n$  本のベクトルの  $n$  個の終点の集中度を表示する第一次的な指標である。ところで、 $D$  について、

$$\frac{n-1}{n} \sum_i \sigma_i^2$$

という前半の項は、明らかにスケール・ファクターであって、外から人工的にきめられる部分である。したがって、いまスケールに影響されない比  $\lambda$ 、

$$\lambda = \frac{2 \sum_i \sum_{\substack{j \\ (i < j)}} \sigma_{ij}}{(n-1) \sum_i \sigma_i^2} \quad \dots \quad (3.4.1)$$

という比、すなわち 2 次形式の比を考えると

$$D = \frac{n-1}{n} \sum_i \sigma_i^2 (1-\lambda) \quad \dots \quad (3.4.2)$$

という形になる。

すでにみてきたような数量化の着想からすれば、ここで、

$$D \rightarrow \min. \quad \dots \quad (3.4.3)$$

をみたす数量化を行えばよい。しかし、演算の方法としては

$$\lambda \rightarrow \max. \quad \dots \quad (3.4.4)$$

という解を求めて、スケール・ファクターを別とすれば、(3.4.3) 式の場合と解として明らかに同等であり。しかもより合目的的である。

したがって、本稿においては、 $n$  元度数分布を数量化するための測度として (3.4.1) 式の  $\lambda$  を採用し、(3.4.4) 式によって、数量化の方法を導き出すこととする。

#### 4. 数量化の方法——1 次元の場合——

4.1)  $n$  元について、一般に  $i$  元目に  $(x_i)$  という 1 次元の列ベクトルをわりあてる。すなわち、 $i$  元目  $s$  項目に  $x_{is}$  という数値をわりあてる。また、その度数  $f_{is}$  をウェイトとして

$$\sum_i f_{is} x_{is} = 0, \quad \bar{x}_{i\cdot} = 0,$$

あるいは、 $(x_i)$  の平均を予めゼロにきめておく。 $(\lambda)$  におけるように分散・共分散の比を問題にする限り、平均ゼロという原点条件は解の一般性に影響しない。)

4.2) 一般に,  $i$  元目  $s$  項目の 1 元分布度数  $f_{is}$  を  $s$  行  $s$  列の要素とする対角行列  $F_i$  を定義する.

同様に、一般に、 $i$  元目  $s$  項目と  $j$  元目  $t$  項目との 2 元分布度数  $f_{is, jt}$  を  $s$  行  $t$  列の要素とする行列  $A_{ij}$  を定義する。このとき

ただし  $A_{ji}'$  は  $A_{ji}$  の転置行列とする.

4.3) 2 次形式の表示法を系統的に利用すれば、一般に、 $i$  元目についての分散  $\sigma_i^2$  については、

また、同様に、一般に、 $i$  元目と  $j$  元目についての共分散  $\sigma_{ij}$  については、

となる。なお、この  $N$  は、すでにみたように、各元における総度数である。

4.4) さて(3.4)節において説明したように、

$$\lambda = \frac{2 \sum_i^{\langle i < j \rangle} \sigma_{ij}}{(n-1) \sum_i \sigma_i^2} \rightarrow \max. \quad \dots \dots \dots \quad (4.4.1)$$

を解くわけであるが、ここで、分子、分母に  $N$  を掛けて

$$\lambda = \frac{2 \sum_{\substack{i < j}} x_i' A_{ij} x_j}{(n-1) \sum_i x_i' F_i x_i} \rightarrow \max. \quad \dots \dots \dots \quad (4.4.2)$$

を解いてもよい。

(4.4.2) 式をまず  $x_i$  で偏微分してゼロとおけば

$$\frac{1}{2} \cdot \frac{\partial \lambda}{\partial x_i} = 0 \quad \dots \dots \dots \quad (4.4.3)$$

項を整理すれば

## あるいは

同様にして  $n$  回偏微分をくりかえせば、結果としてはこの形の式が  $n$  個連立することになる。

ここで一般に行列  $B$  およびその要素行列  $B_{\alpha\alpha}$ ,  $B_{\beta\beta}$ ,  $B_{\alpha\beta}$ ,  $B_{\beta\alpha}$  について次のように定義すると、( $\alpha, \beta$  は一般化した添字)

$$\left\{ \begin{array}{l} [\mathbf{B}_{\alpha\alpha}] = [\mathbf{0}], \quad [\mathbf{B}_{\beta\beta}] = [\mathbf{0}] \\ [\mathbf{B}_{\alpha\beta}] = [\mathbf{F}_\alpha^{-1} \mathbf{A}_{\alpha\beta}], \quad [\mathbf{B}_{\beta\alpha}] = [\mathbf{F}_\beta^{-1} \mathbf{A}_{\beta\alpha}] \\ [\mathbf{B}] = \begin{bmatrix} [\mathbf{0}] & \cdots & [\mathbf{B}_{\alpha\beta}] \\ \vdots & \ddots & \vdots \\ [\mathbf{B}_{\alpha\beta}] & \cdots & [\mathbf{0}] \end{bmatrix} \end{array} \right.$$

この(4.4.5)式のような $n$ 個の式が連立する式は、次のように、行列を使った形で表示されることになる。

$$\mathbf{Bx} = (n-1)\lambda \mathbf{x} \quad \dots \quad (4.4.6)$$

この式について、固有値と固有ベクトルを算出すれば、求める解がえられるわけである。

4.5) (4.4.6)式の左辺の行列 $\mathbf{B}$ は、対角要素行列がすべてゼロ行列である非対称行列であって、通常の演算に必ずしも便利でない。ここで左辺の行列が対称行列になるようにするには、一般に(4.4.4)式の形の式を相互に代入して整理すればよい。その結果は、項をまとめ、 $x_i$ について

$$\left\{ \sum_{\substack{\alpha \\ (\alpha \neq i)}} (\mathbf{A}_{i\alpha} \mathbf{F}_\alpha^{-1} \mathbf{A}_{\alpha i}) \right\} x_i + \sum_{\substack{\alpha \\ (\alpha \neq i)}} \left[ \left\{ \sum_{\substack{\beta \\ (\beta \neq \alpha) \\ (\beta \neq i)}} (\mathbf{A}_{i\beta} \mathbf{F}_\beta^{-1} \mathbf{A}_{\beta\alpha}) \right\} x_\alpha \right] = (n-1)^2 \lambda^2 \mathbf{F}_i \mathbf{x}_i \quad \dots \quad (4.4.1)$$

という形の式が $n$ 個連立することになる。ここで、行列 $\mathbf{F}, \mathbf{C}$ 、およびそれらの要素行列 $\mathbf{F}_{\alpha\alpha}, \mathbf{F}_{\beta\beta}, \mathbf{F}_{\alpha\beta}, \mathbf{F}_{\beta\alpha}$ 、および $\mathbf{C}_{\alpha\alpha}, \mathbf{C}_{\beta\beta}, \mathbf{C}_{\alpha\beta}, \mathbf{C}_{\beta\alpha}$ を次のように定義すれば

$$\left\{ \begin{array}{l} [\mathbf{C}_{\alpha\alpha}] = \sum_{\gamma} (\mathbf{A}_{\alpha\gamma} \mathbf{F}_\gamma^{-1} \mathbf{A}_{\gamma\alpha}) \quad [\mathbf{C}_{\beta\beta}] = \sum_{\gamma} (\mathbf{A}_{\beta\gamma} \mathbf{F}_\gamma^{-1} \mathbf{A}_{\gamma\beta}) \\ [\mathbf{C}_{\alpha\beta}] = \sum_{\delta} (\mathbf{A}_{\alpha\delta} \mathbf{F}_\delta^{-1} \mathbf{A}_{\delta\beta}) \quad [\mathbf{C}_{\beta\alpha}] = \sum_{\delta} (\mathbf{A}_{\beta\delta} \mathbf{F}_\delta^{-1} \mathbf{A}_{\delta\alpha}) \\ [\mathbf{C}_{\alpha\beta}] = [\mathbf{C}_{\beta\alpha}'], \quad [\mathbf{C}_{\alpha\beta'}] = [\mathbf{C}_{\beta\alpha}], \\ [\mathbf{C}] = \begin{bmatrix} [\mathbf{C}_{\alpha\alpha}] & \cdots & [\mathbf{C}_{\alpha\beta}] \\ \vdots & \ddots & \vdots \\ [\mathbf{C}_{\beta\alpha}] & \cdots & [\mathbf{C}_{\beta\beta}] \end{bmatrix} \\ [\mathbf{F}_{\alpha\alpha}] = [\mathbf{F}_\alpha], \quad [\mathbf{F}_{\beta\beta}] = [\mathbf{F}_\beta] \\ [\mathbf{F}_{\alpha\beta}] = [\mathbf{0}], \quad [\mathbf{F}_{\beta\alpha}] = [\mathbf{0}] \\ [\mathbf{F}] = \begin{bmatrix} [\mathbf{F}_\alpha] & \cdots & [\mathbf{0}] \\ \vdots & \ddots & \vdots \\ [\mathbf{0}] & \cdots & [\mathbf{F}_\beta] \end{bmatrix} \end{array} \right.$$

要するに、 $\mathbf{F}$ は対角行列を対角要素行列とする対角行列である。

この(4.5.1)式のような $n$ 個の式が連立する形の式は次のように行列を使った形の式で表示されることになる。

$$\mathbf{Cx} = (n-1)^2 \lambda^2 \mathbf{Fx} \quad \dots \quad (4.5.2)$$

この式について、固有値、固有ベクトルを算出すれば、解がえられるわけである。

4.6) ここで、(4.5.2)式を、さらに、左辺にだけ対称行列のある最も便利な形にかえるためには、まず、一般に、 $i$ 元目 $s$ 項目に与えるべき $x_{is}$ について

$$x_{is} = \frac{z_{is}}{\sqrt{f_{is}}} \quad \dots \quad (4.6.1)$$

という $x_{is}$ をきめ、これに応じて $\mathbf{x}_i, \mathbf{x}_j$ などのベクトルを $\mathbf{z}_i, \mathbf{z}_j$ などのベクトルでおきかえればよい、ここで、行列 $\mathbf{G}$ 、およびその要素行列 $\mathbf{G}_{\alpha\alpha}, \mathbf{G}_{\beta\beta}, \mathbf{G}_{\alpha\beta}, \mathbf{G}_{\beta\alpha}$ を次のように定義すれば

$$\left\{ \begin{array}{l} [\mathbf{G}_{\alpha\alpha}] = \mathbf{F}_\alpha^{-\frac{1}{2}} \left\{ \sum_{\gamma} (\mathbf{A}_{\alpha\gamma} \mathbf{F}_\gamma^{-1} \mathbf{A}_{\gamma\alpha}) \right\} \mathbf{F}_\alpha^{-\frac{1}{2}} \\ [\mathbf{G}_{\beta\beta}] = \mathbf{F}_\beta^{-\frac{1}{2}} \left\{ \sum_{\gamma} (\mathbf{A}_{\beta\gamma} \mathbf{F}_\gamma^{-1} \mathbf{A}_{\gamma\beta}) \right\} \mathbf{F}_\beta^{-\frac{1}{2}} \\ [\mathbf{G}_{\alpha\beta}] = \mathbf{F}_\alpha^{-\frac{1}{2}} \left\{ \sum_{\delta} (\mathbf{A}_{\alpha\delta} \mathbf{F}_\delta^{-1} \mathbf{A}_{\delta\beta}) \right\} \mathbf{F}_\beta^{-\frac{1}{2}} \\ [\mathbf{G}_{\beta\alpha}] = \mathbf{F}_\beta^{-\frac{1}{2}} \left\{ \sum_{\delta} (\mathbf{A}_{\beta\delta} \mathbf{F}_\delta^{-1} \mathbf{A}_{\delta\alpha}) \right\} \mathbf{F}_\alpha^{-\frac{1}{2}} \\ [\mathbf{G}_{\alpha\beta}] = [\mathbf{G}_{\beta\alpha}'], \quad [\mathbf{G}_{\alpha\beta}'] = [\mathbf{G}_{\beta\alpha}], \\ [\mathbf{G}] = \begin{bmatrix} [\mathbf{G}_{\alpha\alpha}] & \cdots & [\mathbf{G}_{\alpha\beta}] \\ \vdots & \ddots & \vdots \\ [\mathbf{G}_{\beta\alpha}] & \cdots & [\mathbf{G}_{\beta\beta}] \end{bmatrix} \end{array} \right.$$

(4.5.2) 式は

$$\mathbf{Gz} = (n-1)^2 \lambda^2 \mathbf{z} \quad \dots \quad (4.6.2)$$

という最も単純で便利な形になる。ここでこの対称行列  $\mathbf{G}$  の固有値、固有ベクトルを算出すれば、求める解がえられる。

4.7) このようにして、(4.4.6) 式、(4.5.2) 式、(4.6.2) 式の何れの式を使っても、 $\lambda$  および  $\mathbf{x}$  ベクトルが算出される。n 元のうち、一般に、 $i$  元目の中に含まれている分類区分の項目数を  $r_i$ ,

$$\sum_j r_j = R$$

とすれば、上の演算は  $R \times R$  のサイズの行列について最大固有値とそれに対応する固有ベクトルを算出する演算である。

なお、3 式の右辺と左辺とをそれぞれ縦に加えれば明らかのように、これらの式は最大固有値としてつねに 1 をもち、これに対応する固有ベクトルはつねに、 $(\mathbf{x}) \equiv (1)$  となる。これは、形式的には  $\mathbf{x}_i$  の平均ゼロという始めの条件と矛盾し、実質的には、各元内の各項目の弁別が不可能となるから、求める解ではない。従って、実際の演算においては、 $\lambda$  については 2 番目に大きい固有値を、またベクトルについては、この  $\lambda$  に対応する固有ベクトルを採用しなければならない。

## 5. 数量化の方法——多次元の場合——

5.1) 以上のような方法によって、n 元度数分布における各元内の各項目に、1 次元の数値を、同時に、わりあてることが可能となる。さらにまた、2 元度数分布を数量化する場合と全く同様に、n 元の各元各項目に、多次元の数値をわりあてるこども可能である。以下、その方法を説明する。

5.2) まず、2 元度数分布の数量化のとき用いられる方法をみてみよう。そこでは、上記(2.2)節の説明に戻って、同じ術語を使用すると、(イ)の場合、行に  ${}^1\mathbf{x}, {}^2\mathbf{x}$  という 2 つのベクトルをわりあててえられる 2 つの  ${}^1\eta_x, {}^2\eta_x$  の積  ${}^1\eta_x {}^2\eta_x$ ; (ロ)の場合、列に  ${}^1\mathbf{y}, {}^2\mathbf{y}$  という 2 つのベクトルをわりあててえられる 2 つの  ${}^1\eta_y, {}^2\eta_y$  の積  ${}^1\eta_y {}^2\eta_y$ ; (ハ)の場合、行に  ${}^1\mathbf{x}, {}^2\mathbf{x}$ 、列に  ${}^1\mathbf{y}, {}^2\mathbf{y}$  というそれぞれ 2 つのベクトルをわりあててえられる 2 つの  ${}^1\rho_{xy}, {}^2\rho_{xy}$  の積、 ${}^1\rho_{xy} {}^2\rho_{xy}$ ;

これら表見的には3通りで実は等値の測度を最大にすることによって、2次元の数量化を考えている\*。

この考え方を2元度数分布の数量化から $n$ 元度数分布の数量化の場合に拡張するすれば、次のようにすればよい。

$n$ 元について、一般に、 $i$ 元目に ${}^1\mathbf{x}_i, {}^2\mathbf{x}_i$ という2次元のベクトルをわりあてるとすれば、 ${}^1\mathbf{x}$ のベクトルに対応して ${}^1\lambda, {}^2\lambda$ のベクトルに対応して ${}^2\lambda$ が、(3.4.1)式からきまる。ここで、

$$\lambda = {}^1\lambda \cdot {}^2\lambda \quad \dots \dots \dots \dots \dots \dots \dots \quad (5.2.1)$$

という測度を定義して

$$\lambda \rightarrow \max. \quad \dots \dots \dots \dots \dots \dots \dots \quad (5.2.2)$$

を解けば、求める2次元の数値がえられることになる。

5.3) 上記の(4.1)節から(4.3)節までと同様に、2次元のそれぞれの元について、ベクトル、 ${}^1\mathbf{x}, {}^2\mathbf{x}$ および分散、共分散を定義する。なお、全体として ${}^1\mathbf{x}$ と ${}^2\mathbf{x}$ が同一のベクトルに帰着しては解として役に立たないから、 ${}^1\mathbf{x}$ と ${}^2\mathbf{x}$ の直交性を解の制限条件としてつけておく。

$$\sum {}^1\mathbf{x}_i' F_i {}^2\mathbf{x}_i = 0 \quad \dots \dots \dots \dots \dots \dots \dots \quad (5.3.1)$$

あるいは

$${}^1\mathbf{x}' F^2 \mathbf{x} = 0 \quad \dots \dots \dots \dots \dots \dots \dots \quad (5.3.2)$$

ただし、この $F$ は(4.5.2)式に出てくる $F$ である。

なお、これは $n$ 元のすべての元について、一般に、 ${}^1\mathbf{x}_i' F_i {}^2\mathbf{x}_i = 0$ ではないことに注意する必要がある。

したがって解としては

$$\lambda - \mu({}^1\mathbf{x}' F^2 \mathbf{x}) \rightarrow \max. \quad \dots \dots \dots \dots \dots \dots \dots \quad (5.3.3)$$

を $\mathbf{x}_i$ で偏微分して解けばよい。なお、 $\mu$ はLagrangeの未定乗数である。

5.4) 予め、便宜上

$${}^1\mathbf{x}' F^1 \mathbf{x} = p, \quad (p > 0) \quad \dots \dots \dots \dots \dots \dots \dots \quad (5.4.1)$$

$${}^2\mathbf{x}' F^2 \mathbf{x} = q, \quad (q > 0) \quad \dots \dots \dots \dots \dots \dots \dots \quad (5.4.2)$$

ときめておく。まず、 ${}^1\mathbf{x}_i$ で偏微分して、

\* まず、分散成分の比を考えるとき、1次元分散の代りに一般化分散を考え、その成分の比について考える。いま便宜上上記(2.2)節の(イ)の場合についていえば、次のような比 $s^2$ が成り立つ。

$$s^2 = \frac{\begin{vmatrix} {}^1\sigma_W^2 & c_{12}W \\ c_{21}W & {}^2\sigma_W^2 \end{vmatrix}}{\begin{vmatrix} {}^1\sigma_T^2 & c_{12} \\ c_{21} & {}^2\sigma_T^2 \end{vmatrix}}$$

${}^1\sigma_T^2$  は1次元目の分散  
 ${}^2\sigma_T^2$  は2次元目の分散  
 $c_{12}$  は共分散  
 ${}^1\sigma_W^2$  は1次元目の内分散  
 ${}^2\sigma_W^2$  は2次元目の内分散  
 $c_{12}W$  は内分散の共分散

ここで直交条件に $c_{12}W=0$ を入れると、 $s^2$ は $c_{12}=0$ のとき最小となる。そして解の直交条件から、 $c_{12}W=0, c_{12}=0$ とも充足される。このとき

$$s^2 = \frac{{}^1\sigma_W^2}{{}^1\sigma_T^2} \cdot \frac{{}^2\sigma_W^2}{{}^2\sigma_T^2} = (1-\eta_1^2)(1-\eta_2^2)$$

但し  $\eta_1^2 = 1 - \frac{{}^1\sigma_W^2}{{}^1\sigma_T^2}$   
 $\eta_2^2 = 1 - \frac{{}^2\sigma_W^2}{{}^2\sigma_T^2}$

となる。 $s^2$ の最小は $(1-s^2)$ の最大であり、これは $1-(1-\eta_1^2)(1-\eta_2^2)=\eta'^2$ の最大と同値となる。そして、 $\eta'^2$ の最大は ${}^2\eta_1^2\eta_2=\eta^2$ を最大にすることと同等である。その意味で、2次元の数量化において ${}^1\eta_x {}^2\eta_x$ の積を最大にすることは1次元の数量化の場合の、すなはな拡張である。

項を整理して

$$\sum_{\alpha} A_{i\alpha}^{-1} x_{\alpha} = (n-1)!\lambda F_i^1 x_i + \frac{p}{2\lambda} \mu F_i^2 x_i \quad \dots \dots \dots (5.4.4)$$

この式に  $x_i^2$  を左から掛け、 $i$  にわたって両辺をそれぞれ加えると、添字を一般化して表示しなおして

同様にして、 ${}^2x_i$  で偏微分して、

$$\frac{\partial}{\partial^2 x_i} \{ \lambda - \mu(\mathbf{x}' \mathbf{F}^2 \mathbf{x}) \} = 0 \quad \dots \dots \dots \quad (5.4.6)$$

上と同様な演算を加えると、

したがって (5.4.5) 式と (5.4.7) 式とから

あるいは

となる. 一般に  ${}^1\lambda \neq {}^2\lambda$ ,  $p > 0$ ,  $q > 0$  であるから, 当然に

となる。

すなわち、(5.4.4) 式の形の式は、一般に、(4.4.4) 式の形の式に帰着する。したがって、(4.4.4) 式から由来した (4.4.6) 式、(4.5.2) 式、(4.6.2) 式のいずれかの式から、 ${}^1\lambda$ ,  ${}^2\lambda$ ,  ${}^1x$ ,  ${}^2x$  を算出すればよいことになる。具体的には、これら 3 式の固有値のうち、1 の次に大きい固有値を  ${}^1\lambda$ 、その次に大きい固有値を  ${}^2\lambda$  として採用し、また  ${}^1\lambda$  に対応する固有ベクトルを  ${}^1x$ ,  ${}^2x$  に対応する固有ベクトルを  ${}^2x$  として採用すればよい。

5.5) 上記の(5.2)節から(5.4)節までの立論は、 $n$ 元度数分布の各元内の各項目に、3次元目、4次元目などなどの数値を、次々と、わりあてるときにも、逐次成立する。理論的には、(4.4.6)式、(4.5.2)式、(4.6.2)式の行列のサイズを $R \times R$ として、( $R-1$ )次元目にいたるまで、次々と多次元の数値を $n$ 元分布の各元内の各項目にわりあてることが可能である。しかし、具体的なデータの処理としては、固有値が小さくなるなり方、あるいは行列のrankの落ち方をみながら、適当な $S$ 次元まで数量化し( $S < R-1$ )、その残りの( $R-1-S$ )次元については数量化を打切ることが、より実際的である。

## 6. 補註の 1—ウェイトつきの場合—

6.1) これまでの説明で明らかなように、本稿における  $n$  元度数分布を数量化する方法は、  
 $\binom{n}{2}$  通り成立つ 2 元度数分布が含む情報を出発点としている。しかも、この  $\binom{n}{2}$  通りの 2 元  
 度数分布が含む情報を、すべて等しいウェイトで扱っている。すなわち、(3.5.5) 式の  $D$  を  
 考えるとき、 $n$  本のベクトルの  $n$  個の終点を同じウェイトをもつものとして扱い、また

(3.4.1) 式の  $\lambda$  を考えるとき、 $\binom{n}{2}$  通りの共分散を等しいウェイトをもつものとして扱っている。

しかし、与えられたデータにおいて、一般に、 $i$  元と  $j$  元との間の共分散  $\sigma_{ij}$  が 2 元度数分布の数量化によってとりうる  $\binom{n}{2}$  通りの値の間には、かなりのバラツキのある場合の方が、むしろ、通例である。この事情を考慮する場合、一般に  $\sigma_{ij}$  がとりうる値のバラツキを考慮に入れ何らかのウェイトをつけて  $\sigma_{ij}$  を扱いたいという考え方も成り立つであろう。そして、この場合において、すべての  $\sigma_{ij}$  を同じウェイトで扱うよりも、 $\lambda$  の値がより大きくなることも考えられる。あるいは、与えられたデータの事情により合致した数量化がなされているとも考えられるであろう。このような考え方をとる場合を考慮して、以下に、 $\sigma_{ij}$  にウェイトをつけたときの数量化の方法を、簡単に補足する。

6.2) まず、一般に、 $i$  元と  $j$  元との間のペアについて、左表のようなウェイトのシステムが与えられていると考える。

元	$i$	$\dots$	$j$	$\dots$	$n$	$\Sigma$
$i$	$\dots(\bar{w}_i)$	$\dots$	$\bar{w}_{ij}$	$\dots$	$\bar{w}_{in}$	$W_i = n \cdot \bar{w}_i$
$j$	$\dots \bar{w}_{ji}$	$\dots(\bar{w}_j)$	$\dots$	$\bar{w}_{jn}$	$\dots$	$W_j = n \cdot \bar{w}_j$
$n$	$\dots \bar{w}_{ni}$	$\dots \bar{w}_{nj}$	$\dots(\bar{w}_n)$	$\dots$	$\dots$	$W_n = n \cdot \bar{w}_n$
$\Sigma$	$\dots W_i$	$\dots W_j$	$\dots W_n$	$\dots$	$\dots$	$W_T = n^2 \cdot \bar{w}$

一般に、

$$\bar{w}_{ij} = \bar{w}_{ji}$$

$$\sum_j \bar{w}_{ij} = W_i = n \cdot \bar{w}_i$$

$$\sum_j \sum_i \bar{w}_{ij} = W_T = n^2 \cdot \bar{w}$$

このとき、(3.3.1) 式については、ウェイトつきの  $D$  を  $D_w$  で表示して、

$$D_w = \sum_i \bar{w}_i d_{ii}^2 = \sum_i \sum_j \bar{w}_{ij} d_{ij}^2 - n \bar{w} \sum_i \bar{x}_i^2 \quad \dots \dots \dots (6.2.1)$$

となる。また (3.3.2) 式については

$$2n D_w = \sum_i \sum_{\substack{j \\ (i \neq j)}} \bar{w}_{ij} d_{ij}^2 \quad \dots \dots \dots (6.2.2)$$

となり、(3.3.4) 式については

$$2n D_w = 2 \left( \sum_i W_i \sigma_i^2 - 2 \sum_i \sum_{\substack{j \\ (i < j)}} \bar{w}_{ij} \sigma_{ij} \right) \quad \dots \dots \dots (6.2.3)$$

となる。したがって、(3.3.5) 式は

$$D_w = \frac{1}{n} \sum_i W_i \sigma_i^2 - \frac{2}{n} \sum_i \sum_{\substack{j \\ (i < j)}} \bar{w}_{ij} \sigma_{ij} \quad \dots \dots \dots (6.2.4)$$

となる。このときの  $\lambda$  を  $\lambda_w$  で表示すれば、(3.4.1) 式は、

$$\lambda_w = \frac{2 \sum_i \sum_{\substack{j \\ (i < j)}} \bar{w}_{ij} \sigma_{ij}}{\sum_i W_i \sigma_i^2} \quad \dots \dots \dots (6.2.5)$$

という形になる。

6.3) ここで、

$$\lambda_\alpha \rightarrow \max. \quad \dots \dots \dots (6.3.1)$$

を解けば、(4.4.4) 式は

$$\sum_{\substack{\alpha \\ (\alpha \neq i)}} \frac{\bar{w}_{i\alpha}}{W_i} A_{i\alpha} x_\alpha = \lambda F_i x_i \quad \dots \dots \dots (6.3.2)$$

の形になる。行列で表示すれば、行列  $B$  およびその要素行列  $B_{\alpha\alpha}$ ,  $B_{\beta\beta}$ ,  $B_{\alpha\beta}$ ,  $B_{\beta\alpha}$  を次のように定義して、

$$\left\{ \begin{array}{l} [B_{\alpha\alpha}] = [\mathbf{0}], \quad B_{\beta\beta} = [\mathbf{0}] \\ [B_{\alpha\beta}] = \left[ \frac{\bar{w}_{\alpha\beta}}{W_\alpha} F_\alpha^{-1} A_{\alpha\beta} \right], \quad [B_{\beta\alpha}] = \left[ \frac{\bar{w}_{\beta\alpha}}{W_\beta} F_\beta^{-1} A_{\beta\alpha} \right] \\ [B] = \left[ \begin{array}{c} [\mathbf{0}] \cdots [B_{\alpha\beta}] \\ \vdots \quad \ddots \quad \vdots \\ [B_{\beta\alpha}] \cdots [\mathbf{0}] \end{array} \right] \end{array} \right.$$

(4.4.6) 式が

$$Bx = \lambda x \quad \dots \dots \dots \quad (6.3.3)$$

の形になる。

この (6.3.3) 式から出発して、(4.5.2) 式、(4.6.2) 式も、それぞれ書き直されるが、その書き直しは簡単なので、ここでは省略する。

なお、説明を省略するが、このウェイトつきの数量化の方法においても、すでにみた場合と全く同様に、多次元の数値を、次々に、わりあてることが可能である。

## 7. 補註の 2 ——元の追加——

7.1)  $n$  元度数分布を数量化するこの方法は、(4.5.2) 式から明らかなように、 $n=2$  のとき、2 元度数分布を数量化するこれまでの方法に帰着する。このとき、(2.2) 節の定義に戻って

$$\lambda = \rho_{xy} = \eta_x = \eta_y \quad \dots \dots \dots \quad (7.1.1)$$

である。

以上の意味で、本稿の方法は 2 元度数分布を数量化する方法を、特殊ケースとして、その内側に含む方法であり、その限りで、本稿の方法は、2 元度数分布を数量化する方法を  $n$  元の場合に拡張したものである。

しかし、すでに (2.5), (2.6) 節で述べたように、本稿の方法は、 $n$  元にわたる分布度数が伝える情報を、直接にもれなく、利用した方法ではない。分散分析になぞらえたいい方をすれば、主効果と 2 次の交互作用の効果とだけを利用し、3 次およびそれ以上の交互作用の効果を除外しているものである。その限りで、本稿の方法は、2 元度数分布を数量化する方法の根本的着想を  $n$  元度数分布の場合にまで拡張したものではない。

7.2) 本稿の方法は、具体的な演算としては、 $R \times R$  のサイズの行列の固有値、固有ベクトルの演算に帰着する。その限りで、本稿の方法の実用性はコンピューターの容量と研究費とによって決定される。したがって、ある元が、数千、数万の項目（たとえば、一連番号をつけた標本個体）を含むとき、本稿の方法を直接に利用することは、必ずしも、実用的といえない。この場合には、次のような便宜的な処理が可能である。

まず、 $n$  元に対応する  $n$  本のベクトルが、(3.4.4) 式の

$$\lambda \rightarrow \max.$$

および、(3.4.3) 式の

$$D \rightarrow \min.$$

をへて、上述の演算によって、すでに予め定められたと仮定する。

ここに、 $(n+1)$  元目を追加して導入し、これに対応する  $(n+1)$  本目のベクトルを引き、 $(n+1)$  番目の終点を定めるとすれば、 $n$  元のときの  $D_n$  が  $(n+1)$  元に 1 元だけふえたとき

の  $D_{n+1}$  に増加するわけであるが、 $D_n$  から  $D_{n+1}$  への増分が最小になるように定めることができ、統計数理的にみて、最も合理的な方法である。

(n+1) 元目の  $p$  項目に数値  $y_p$  を与えるとする。また、 $f_{n+1,p}$  をその度数として、比較の便宜上これまでの  $n$  元と同様に、

$$\sum_p f_{n+1,p} y_p = 0, \quad \bar{y} = 0 \quad \dots \dots \dots \quad (7.2.1)$$

という原点条件をきめておくと

$$y_p = \frac{1}{n} \cdot \frac{1}{f_{n+1,p}} \sum_i \sum_s f_{is,n+1,p} x_{is} \quad \dots \dots \dots \quad (7.2.2)$$

となり、この(7.2.2)式から  $y_p$  が算出される。おな、この式は  $\lambda$  という係数を別にすれば、(4.4.5)式と同じ形の式である。

このようにして、予め  $n$  元について数量化しておく、(n+1) 元目を追加することも一つの便法である。しかも、この場合、 $n$  元についてすでに与えられている多次元の数値のそれぞれについて、(n+1) 元目を追加して導入すること、あるいは(n+1) 元目にも多次元の数値を与えることが可能である。

7.3) また、(7.2) 節の便法を利用すれば、まず特定の2元度数分布に注目してこれを数量化し、この結果を前提として、第3元目を追加して(7.2.2)式により数量化し、以上3元の数値を前提としてさらに第4元目を追加する、という工合に順次ひとつずつ元を追加し積み上げることも可能である。逆にいえば、 $n$  元度数分布の数量化を、最後的には、2元度数分布の数量化にまず還元し、2元から順次積み上げることが可能である。

ただし、念のため繰返していえば、(n+1) 元目を追加導入したとき、(n+1) 元目について(7.2.2)式によって算出される数値は、(n+1) 個のすべての元を本稿の方法によって一度に数量化したとき、(n+1) 番目の元について与えられる数値の近似値であっても、同一のものではない。とくに、元をひとつずつ順次追加する場合、上の意味の誤差が累積していくことに注意する必要がある。

## 8. 数値例

8.1) 最後に、数値例をひとつあげておく。

データは、神奈川三区における衆議院選挙の開票結果である。第1元として、第27回(昭和30年)、第28回(昭和33年)、第29回(昭和35年)という3項目をえらび、第2元として、多くの候補者の中から、河野一郎、小金義照、森島守人、片山哲という4人の候補者(4項目)をえらび、第3元として選挙区内に31ある全市町村(31項目)をえらんである。すなわち、これは3元度数分布のデータであり、3元を通じて項目数の総計38はである。

各回の開票結果、あるいは3元にわたる分布度数、すなわち、もともとのデータは表8.1.1、8.1.2、表8.1.3、に示されている。

表 8.1.1 第27回(昭和30年)神奈川三区 衆議院議員選挙得票数

市町村名	河野一郎	小金義照	森島守人	片山哲
平塚	15,065	3,393	6,001	6,278
藤沢	8,374	1,443	6,540	12,123
小田原	4,945	10,240	6,196	4,539
茅ヶ崎	5,129	1,169	3,619	4,456
相模原	3,813	150	4,117	6,606
秦野	5,200	3,403	2,886	3,527
厚木	2,073	222	1,134	1,683
大和	1,805	94	1,925	3,459
寒川	1,036	134	682	930
海老名	845	49	613	885
座間	495	38	485	1,053
綾瀬	641	23	240	696
大磯	2,422	1,352	1,298	1,688
二宮	1,051	1,055	775	812
伊勢原	4,565	636	803	1,666
中井	907	851	193	271
大井	547	1,112	317	189
松田	582	1,807	700	482
山北	1,446	3,590	1,434	975
南足柄	1,147	2,989	1,073	863
開成	179	662	241	117
橋	676	510	406	364
箱根	970	2,205	512	708
真鶴	388	1,657	316	296
湯河原	1,315	1,846	897	798
愛川	794	231	270	597
清川	89	6	44	102
城山	417	67	304	521
津久井	1,244	147	724	1,302
相模湖	694	55	563	508
藤野	579	20	476	529
合計	69,433	41,156	45,784	58,923

表 8.1.2 第28回(昭和33年)神奈川三区 衆議院議員選挙得票数

市町村名	河野一郎	小金義照	森島守人	片山哲
平塚	13,456	6,351	9,322	7,619
藤沢	7,870	3,435	8,979	13,439
小田原	10,353	16,385	10,124	6,777
茅ヶ崎	5,452	3,159	5,127	5,299
相模原	1,404	398	4,108	4,775
秦野	2,833	2,956	3,079	2,222
厚木	764	468	1,708	1,580
大和	1,450	328	2,321	3,135
寒川	832	338	935	725
海老名	366	201	892	911
座間	278	90	648	910
綾瀬	704	60	374	621
大磯	2,337	1,869	1,779	1,751
二宮	1,244	1,379	1,261	1,063
伊勢原	3,002	1,161	1,515	1,521
中井	404	803	310	229
大井	487	1,460	556	325
松田	573	1,884	1,094	692
山北	1,287	3,252	1,785	967
南足柄	1,189	3,500	2,107	1,252
開成	136	1,582	363	224
橋	681	775	577	386
箱根	2,035	2,470	819	753
真鶴	951	1,025	597	398
湯河原	1,474	2,448	1,315	954
愛川	300	644	340	301
清川	29	28	74	66
城山	101	36	349	294
津久井	583	189	982	719
相模湖	556	45	604	362
藤野	134	32	632	319
合計	63,265	58,751	64,676	60,589

表 8.1.3 第 29 回(昭和 35 年)神奈川三区 衆議院議員選挙得票数

市町村名	河野一郎	小金義照	森島守人	片山哲
平 塚	13,278	5,717	11,054	6,739
藤 沢	6,422	4,085	9,624	11,369
小 田 原	11,946	15,801	11,326	5,927
茅 ケ 崎	5,167	2,923	6,430	4,410
相 模 原	3,412	633	7,507	3,663
秦 野	2,562	3,086	3,428	1,510
厚 木	797	389	1,687	1,206
大 和	1,879	505	3,230	2,089
寒 川	836	355	907	495
海 老 名	805	247	1,053	652
座 間	704	125	983	564
綾 瀬	755	78	409	358
大 磯	2,255	1,726	2,041	1,452
二 宮	1,352	1,406	1,411	861
伊 势 原	3,452	1,345	1,601	1,199
中 井	317	931	312	153
大 井	548	1,417	559	183
松 田	559	1,864	1,258	411
山 北	1,286	3,226	1,883	626
南 足 柄	1,207	3,258	2,478	981
開 成	140	1,615	415	140
橋	784	761	679	241
箱 根	1,824	2,634	1,079	687
真 鶴	1,179	1,133	700	412
湯 河 原	1,523	2,320	1,343	765
愛 川	330	385	279	243
清 川	79	19	74	37
城 山	96	30	443	213
津 久 井	519	150	784	521
相 模 湖	386	40	644	254
藤 野	92	68	634	180
合 計	66,491	58,272	76,255	48,541

表 8.2.1 項目別の数値—数量化の結果—

元	項目	$^1X$	$^2X$	$^3X$
I	〈回数〉 27回（昭和30年）	.06472	-.03828	-.04160
	28回（昭和33年）	-.02476	.01578	-.00828
	29回（昭和35年）	-.03130	.01739	.04409
II	〈候補者〉 河野一郎	.04064	-.09675	.01535
	小金義照	-.23964	.00872	-.02948
	森島守人	.03402	.05426	.05511
	片山哲	.13959	.04618	-.05168
III	〈市町村〉 平塚	.04179	-.07892	.04521
	藤沢	.13151	.05598	-.04227
	小田原	-.14601	.02469	.02317
	茅ヶ崎	.06184	.01025	.03111
	相模原	.18400	.08138	.03010
	秦野	-.00937	-.04745	-.04679
	厚木	.14180	.02289	-.01551
	大和	.17355	.07180	-.00294
	寒川	.10203	-.02362	.01904
	海老名	.13688	.04055	.03399
	座間	.17546	.07254	.00471
	綾瀬	.16819	-.08351	-.02061
	大磯	-.00610	-.03540	-.00537
	二宮	-.05884	.00739	.01045
	伊勢原	.05944	-.16595	.00759
	中井	-.17487	-.08632	-.09677
	大井	-.25315	.00031	-.04452
	松田	-.20210	.05728	-.03519
	山北	-.19277	.01187	-.05207
	南足柄	-.18859	.05656	-.01780
	開成	-.37761	.09338	-.05639
	橋本	-.08205	-.03831	.02277
	箱根	-.18955	-.04577	-.03608
	真鶴	-.17318	-.04113	-.02231
	湯河原	-.14522	-.01146	-.02437
	愛川	-.00548	-.06097	-.10216
	清川	.13724	-.01026	-.01181
	城山	.19183	.04738	-.03662
	津久井	.16842	-.01955	-.04704
	相模原	.15729	-.03395	.05296
	藤野	.17514	.05758	.03875
$\lambda$		.19501	.09004	.06493

8.2) このデータから、ウェイトを一切つけないで、(4.4.4)式によって数量化した結果が、表 8.2.1 である。ここでは、固有値の動きを考慮して第3次元まで数量化した結果の各項目ごとの数値が示されている。

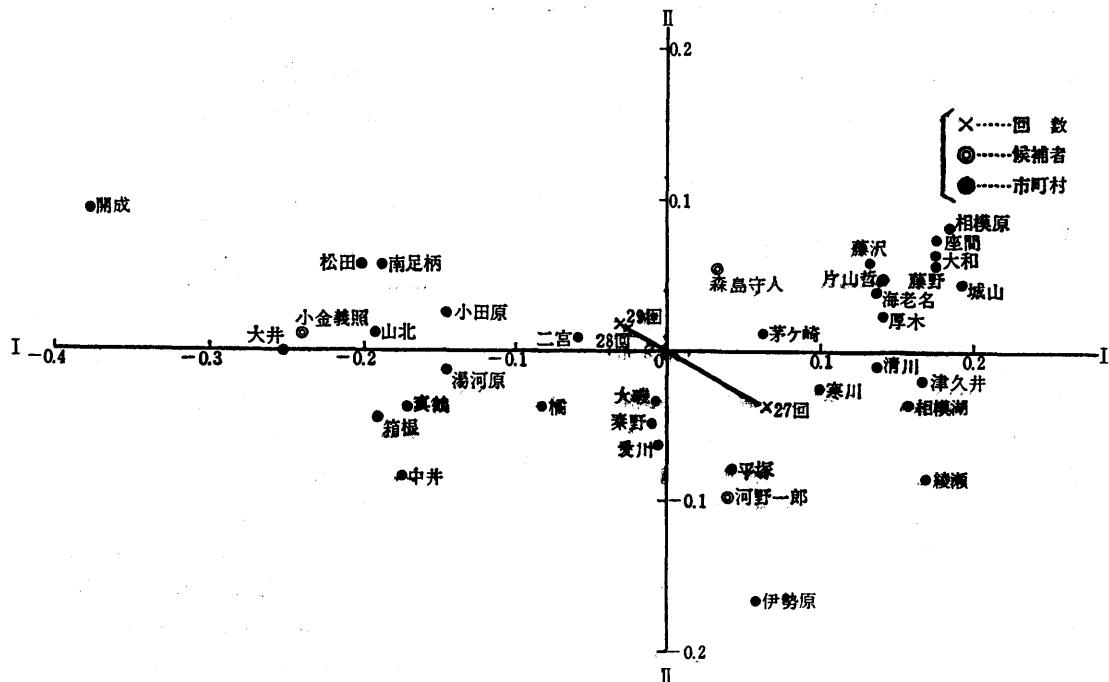


図 8.2.1 回数、候補者および市町村の数値—数量化の結果—(I × II)

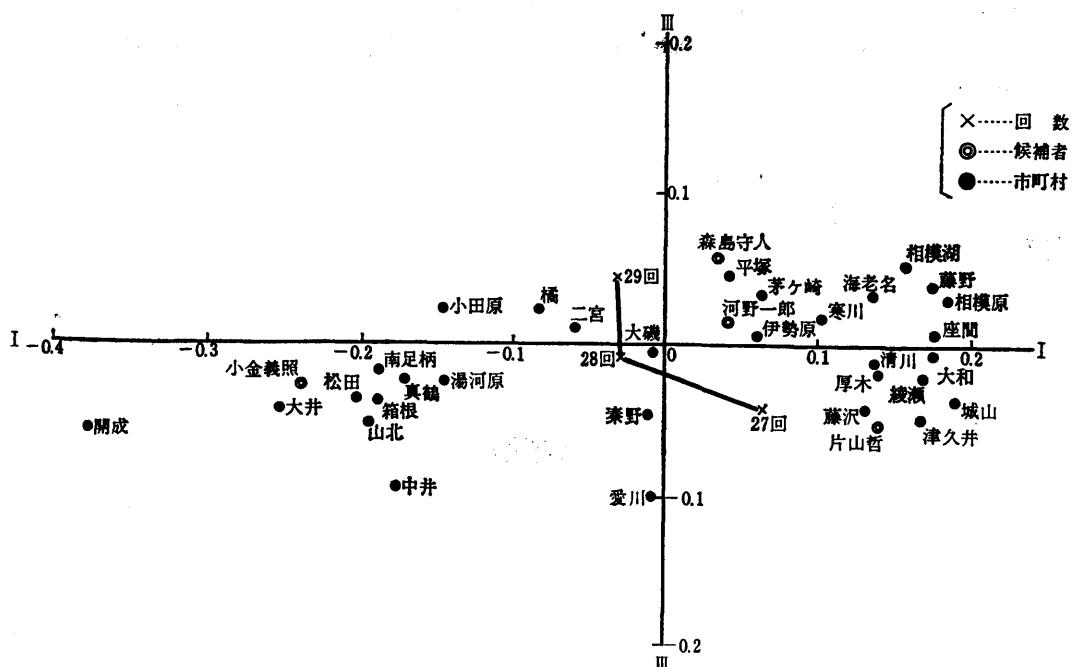
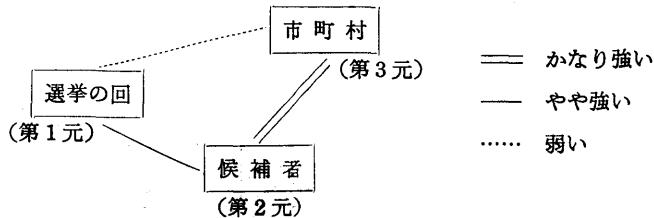


図 8.2.2 回数、候補者および市町村の数値—数量化の結果—(I × III)

なお、この数量化の結果を図示したものが、第1次元と第2次元との組合せについては図 8.2.1、第1次元と第3次元との組合せについては図 8.2.2 である。これらの図は、選挙の回、候補者、市町村という3要因を同時におりこんだ3元地盤図であり、また、この意味での地盤関係を、かなりよく、表示しているといってよい。

8.3) 表 8.2.1 から明らかなように、この計算においては  $\lambda$  が非常に小さい。実は、ここでは、数量化した元のうち、第2元の候補者と第3元の市町村との間の結びつきがかなり強く、



第2元の候補者と第1元の選挙の回との間の結びつきがやや強く、第3元の市町村と第1元の選挙の回との間の結びつきが当然のことながら弱い、というデータの特性がある。この事情に対して、この3通りの結びつきを同等に扱った結果、この3つの結びつきのいわば一種の平均として  $\lambda$  が算出されているのである。これはとりあげた3元を同等のウェイトで扱う、という作業目的に由来している。これに対して、かりにこの3通りの結びつきのそれぞれにウェイトをかけて数量化したとすれば、 $\lambda$  は当然に、かなり大きくなりうると考えてよい。そして、 $n$  元の数量化においてウェイトなしとウェイトつきとどちらを選ぶかは、研究目的から、研究者の責任において、判断すべきことである。一般に、反覆の回、たとえば選挙の回とか世論調査の年度とかを独立の元にたてて数量化するとき、この数値例のような弱い結びつき——2元度数分布からいわゆる期待度数分布を引き去った残りの分布に偏りがあまり見られない場合——が、しばしば、登場する。上にみたウェイトの問題が、研究上の方法として、大きな問題となる次第である。しかし、この問題の肯定あるいは否定については、別の機会に、譲ることにしたい。

東京大学教養学部