

# 集中多様体と集中解析のシステム (I)\*

—— ジーニの統計方法論の幾何学的展開 ——

統計数理研究所 田 口 時 夫

(1988年5月 受付)

## 1. 序論 —— パレート・ジーニの統計業績とその継承 ——

かつて、V. カステラーノ (前ローマ大学統計学部長) は、C. ジーニ (同学部創設者) を追悼した一文において、ジーニを K. ピアソンに比肩する記述統計学者であると評価した。ジーニの統計学は、数理解析的な側面では未完成であり、ピアソンの統計学と比較すると、その対象の有機的な性格の故に遙かに未成熟の段階にあったが、彼自身が予言したように、将来を約束するものであった。

それは、以下の各節において逐次示されるが、その前にピアソンの統計解析と、ジーニの統計解析を明確に区別しなくてはならない。つまりピアソンの解析対象は、規模ないし度数分布表であり、ジーニの解析対象は、規模分布とシェア分布によってベクトルの的に構成された構造統計表であったのである。

ジーニの統計学の成立を歴史的に捉えると、V. パレートの果たした役割を見逃すことはできない。パレートは経験的な所得分布をピアソン系外に見出したのであり、それは後に、P. レヴィによって確認された。ジーニはこの分布に対して解析の道を開いたが、それを実現させたのは構造統計表を基礎とするローレンツ曲線を導入したことによるのである。

後者の統計表は、現に代表的な経済統計の形式を与えており、官庁統計を初めとして一般に普及していることから、この種の統計表をもとにした記述及び解析のシステムを整備し完成させることには大きな意義がある。

さて、ジーニはパレート分布を中心としてローレンツ曲線を解析し、集中係数に不平等尺度としての役割を見出し、更にその主たる構成要素として平均差を析出した。しかし彼の業績は概ね1変量の解析の範囲に留まったので、D.C. マハラノビスを初め、彼の後継者達は2変量の解析へと方法を拡張する努力を重ねてきた。しかしその成果は、ごく最近に至るまではたかだかローレンツ曲線の類似曲線を構成し、それを解析する段階に留まっていたのである\*\*。

2変量の解析系を完成させる為には、集中曲線の段階から集中曲面の構成へと次元を高めね

\* 本稿は次の諸口頭発表を理論的に統括したものである。

田口時夫 (1987). On the structure of multivariate concentration, First International Conference of Statistical Data Analysis Based on the  $L_1$  Norm and Related Methods, Neuchatel, Switzerland.

田口時夫 (1987). 多変量集中解析法の方法論的性格—非線形基準の統計解析法, 理論・計量経済学会.

田口時夫 (1988). 特異分布の解析システムについて, 日本数学会.

\*\* マハラノビスによれば、集中曲線はローレンツ曲線と区別されるべき、その類似曲線の1つであり、N.C. カワニはそれを踏襲している。しかし本稿では特に区別せず、ローレンツ曲線の意味でそれを用いている。

ばならない。それによって初めてピアソン系と並ぶ統計解析系の構成の可能性が生じ、またその性格が浮き彫りにされるのである。

## 2. 集中曲面の定義と性質

集中曲線から集中曲面へと推論を発展させる構想と試みは、単に著者のみでなく、イタリアの G. ルネッタ等を初めとして幾人かの統計学者によっても行われてきた。しかしこれまでのところ、以下に与える著者の形式が最も一般性をもつようであり、事実、それに沿った研究が近來行われている（後掲 Tranquilli (1985), Arnold (1986), Cifarelli and Regazzini (1986) 参照）。

集中曲面を定義するに先立って、経験的なローレンツ曲線が表1の構造統計表のグラフの表現であることを再認識する必要がある。そのグラフはベクトルのに、

$$(2.1) \quad L_i = \begin{pmatrix} \sum_{h=1}^i u_h \\ \sum_{h=1}^i v_h \end{pmatrix}$$

と表わされるものであって、解析的集中曲線は(2.1)式を連続的に表現したものに他ならない。つまり解析的な集中曲線は、規模分布関数及びシエア分布関数をそれぞれ

$$(2.2) \quad U(x) = \int_{-\infty}^x u(\xi) d\xi, \quad V(x) = \int_{-\infty}^x v(\xi) d\xi$$

の形で導入し、そのベクトル関数  $L(x)$  として定義されたものである。従って、 $L(x)$  は

$$(2.3) \quad L(x) \equiv \begin{pmatrix} U(x) \\ V(x) \end{pmatrix} = \int_{-\infty}^x \begin{pmatrix} 1 \\ \xi \\ \mu_x \end{pmatrix} u(\xi) d\xi$$

とすることができる。ここで、 $\mu_x$  は  $x$  の平均値を表わしている（詳細は、後掲 田口 (1984) 参照）。

2 標識  $x, y$  についての集中曲面の構想は、表1の代わりに表2の構造統計表を対象とすることによって得られる。この場合(2.2)式に準じて、規模分布及び  $x, y$  に関するシエア分布をそれぞれ

$$(2.4) \quad \begin{aligned} U(x, y) &= \int_{-\infty}^x \int_{-\infty}^y u(\xi, \eta) d\xi d\eta, \\ V(x, y) &= \int_{-\infty}^x \int_{-\infty}^y v(\xi, \eta) d\xi d\eta, \\ W(x, y) &= \int_{-\infty}^x \int_{-\infty}^y w(\xi, \eta) d\xi d\eta \end{aligned}$$

表1. 単一標識  $x$  に関する構造統計表

分 布	$x$ の階層					計
	$x_0 \leq x < x_1$	...	$x_{i-1} \leq x < x_i$	...	$x_{n-1} \leq x$	
(I) 規 模 分 布	$u_1$	...	$u_i$	...	$u_n$	1
(II) $x$ に関するシエア	$v_1$	...	$v_i$	...	$v_n$	1

表 2. (x, y) 標識に関する構造統計表

$$u_{..} = \sum_{k=1}^m u_k = 1, v_{..} = \sum_{k=1}^m v_k = 1, w_{..} = \sum_{k=1}^m w_k = 1.$$

$y_0 \leq y < y_1$						
分 布	x の階層					計
	$x_0 \leq x < x_1$	...	$x_{i-1} \leq x < x_i$	...	$x_{n-1} \leq x$	
(I) 規 模 分 布	$u_{11}$	...	$u_{i1}$	...	$u_{n1}$	$u_{.1}$
(II) x に関するシェア	$v_{11}$	...	$v_{i1}$	...	$v_{n1}$	$v_{.1}$
(III) y に関するシェア	$w_{11}$	...	$w_{i1}$	...	$w_{n1}$	$w_{.1}$

$y_1 \leq y < y_2$						
分 布	x の階層					計
	$x_0 \leq x < x_1$	...	$x_{i-1} \leq x < x_i$	...	$x_{n-1} \leq x$	
(I) 規 模 分 布	$u_{12}$	...	$u_{i2}$	...	$u_{n2}$	$u_{.2}$
(II) x に関するシェア	$v_{12}$	...	$v_{i2}$	...	$v_{n2}$	$v_{.2}$
(III) y に関するシェア	$w_{12}$	...	$w_{i2}$	...	$w_{n2}$	$w_{.2}$

$y_2 \leq y < y_3$						
⋮						
$y_{m-1} \leq y$						

によって表わすと、我々は解析集中曲面  $L(x, y)$  を

$$(2.5) \quad L(x, y) \equiv \begin{pmatrix} U(x, y) \\ V(x, y) \\ W(x, y) \end{pmatrix} = \int_{-\infty}^x \int_{-\infty}^y \begin{pmatrix} 1 \\ \frac{\xi}{\mu_x} \\ \frac{\eta}{\mu_y} \end{pmatrix} u(\xi, \eta) d\xi d\eta$$

とすることができる。

従って、集中曲線が図 1 のように平面曲線として表現されるのに対して、集中曲面は一般に図 2 のように立体的な曲面として表現される。但し、図 2 は (2.4) 式の  $u(x, y)$  を相互独立なパレート型分布

$$(2.6) \quad u(x, y) = \frac{1}{4} \times 0.9^2 x^{-3/2} y^{-3/2}; \quad 1 \leq x, y \leq 100$$

とした特殊なものであり、一般的には図 3 が予想される。

図 2 及び 3 によって理解できるように、集中曲面は一般的には閉曲面として 2 つの空間的境界曲線をもっている。これらの曲線は (2.5) 式の記号を用いると、容易に  $L(x, \infty)$  及び  $L(\infty, y)$  であることが示されるので、それぞれ  $L_1(x)$  及び  $L_2(y)$  とすることができる。今  $L_1(x)$  及び  $L_2(y)$  の各座標軸方向の正射影曲線をそれぞれ

$$(2.7) \quad L_{10}(x), L_{11}(x), L_{12}(x); L_{20}(y), L_{21}(y), L_{22}(y)$$

とすると、それらは図 3 及び 4 に示されるような位置関係に置かれる。

諸曲線のうち特に、 $L_{12}(x)$  及び  $L_{21}(y)$  は、それぞれ  $x$  及び  $y$  に関するローレンツ曲線である。また、その他の曲線は集中類似曲線といえるもので、その機能に応じて従来から様々な呼

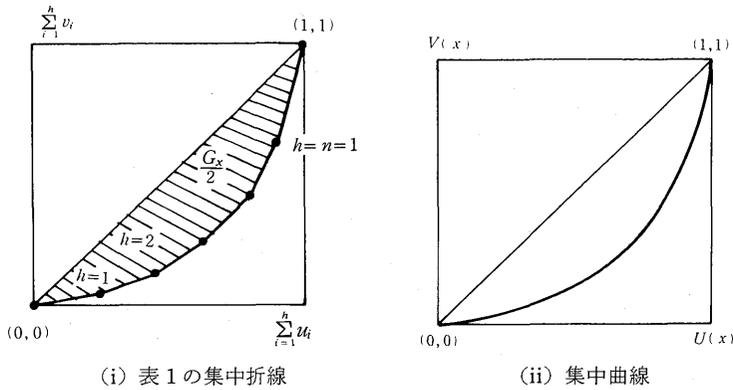


図1.

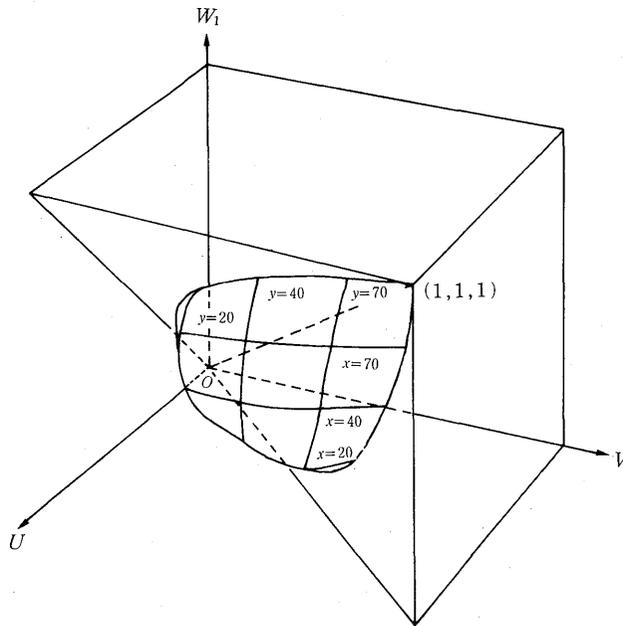


図2. 集中曲面

称を与えられてきた。しかし本稿では、それらの歴史的な由来や機能に拘泥せず、単に第1種及び第2種類似曲線と区別するに留める。

さて、集団の各個体への  $x$  及び  $y$  標識量の配分が均等化されるに従って、集中曲面は図4に示す空間的对角線  $OP$  へと収縮する。従って、線分  $OP$  を2変量  $x, y$  に関する均等線といふことができる。境界曲線の正射影が  $x$  及び  $y$  の集中曲線とその類似曲線を生じたように、空間的均等線の正射影は  $x$  及び  $y$  に関する平面的均等線とその類似直線を生じる。図3の  $OQ, OR, OS$  がそれである。このことは更に、境界曲線の正射影曲線と均等線の正射影との間に、 $x$  及び  $y$  に関する集中面積及びその類似面積を与えることになる。図4に示される  $A, B, C$  がその一例であって、特に面積  $C$  の2倍は、 $x$  に関するジーニ集中度係数を与えている。従って今、 $A, B, C$  の各々の2倍を集中係数及び集中類似係数として  $G_{10}, G_{11}, G_{12}$  で表わすと、それぞれ

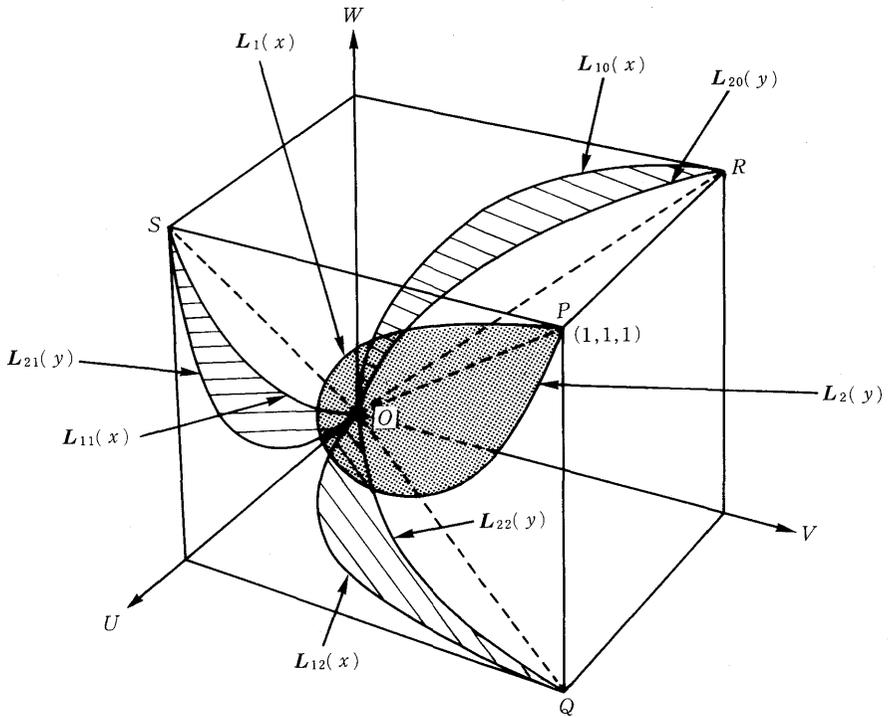


図3. 一般的集中曲面

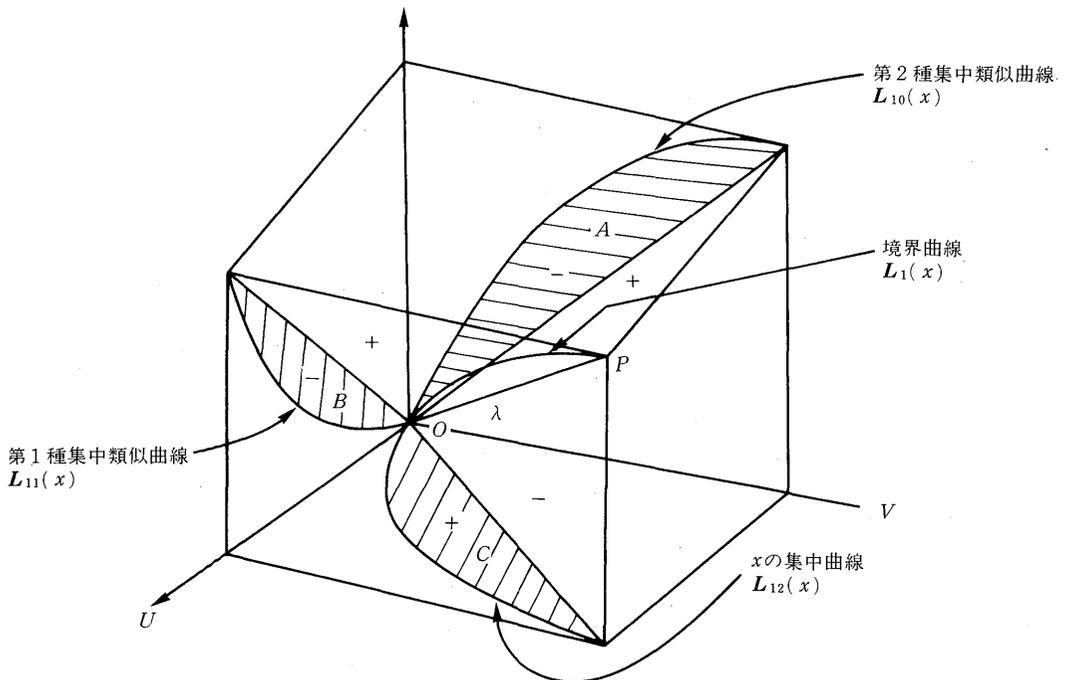


図4.  $L_1(x)$ 境界曲線の性質

$$\begin{aligned}
 (2.8) \quad G_{10} &= 2A = \frac{\Delta_{10}}{2\mu_x\mu_y} && \text{(第2種集中類似係数),} \\
 G_{11} &= 2B = \frac{\Delta_{11}}{2\mu_y} && \text{(第1種集中類似係数),} \\
 G_{12} &= 2C = \frac{\Delta_{12}}{2\mu_x} && \text{(集中係数)}
 \end{aligned}$$

と表わすことができる。ここで  $\mu_x, \mu_y$  はそれぞれ  $x$  及び  $y$  に関する平均値を意味する。また、 $\Delta_{10}, \Delta_{11}, \Delta_{12}$  は平均差及び類似平均差といえるもので、それぞれ

$$\begin{aligned}
 \Delta_{10} &= E_{(X_1, Y_1)} E_{(X_2, Y_2)} \left[ \left\{ \text{sgn}(X_2 - X_1) \right\} \begin{vmatrix} X_1 & X_2 \\ Y_1 & Y_2 \end{vmatrix} \right] \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \text{sgn}(x_2 - x_1) \right\} \begin{vmatrix} x_1 & x_2 \\ \mu_{y|x_1} & \mu_{y|x_2} \end{vmatrix} f_1(x_1) f_1(x_2) dx_1 dx_2 && \text{(第2種類似平均差),} \\
 (2.9) \quad \Delta_{11} &= E_{(X_1, Y_1)} E_{(X_2, Y_2)} \left[ \left\{ \text{sgn}(X_2 - X_1) \right\} \begin{vmatrix} Y_1 & Y_2 \\ 1 & 1 \end{vmatrix} \right] \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \text{sgn}(x_2 - x_1) \right\} (\mu_{y|x_1} - \mu_{y|x_2}) f_1(x_1) f_1(x_2) dx_1 dx_2 && \text{(第1種類似平均差),} \\
 \Delta_{12} &= E_{(X_1, Y_1)} E_{(X_2, Y_2)} \left[ \left\{ \text{sgn}(X_2 - X_1) \right\} \begin{vmatrix} 1 & 1 \\ X_1 & X_2 \end{vmatrix} \right] \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x_2 - x_1| f_1(x_1) f_1(x_2) dx_1 dx_2 = \Delta_x && \text{(平均差).}
 \end{aligned}$$

ここで

$$(2.10) \quad \mu_{y|x} = E(Y | X = x)$$

である。

今、ベクトル  $(G_{10}, G_{11}, G_{12})$  及び  $(\Delta_{10}, \Delta_{11}, \Delta_{12})$  をそれぞれベクトルジーニ係数  $\mathbf{G}_1$  及びベクトル平均差  $\mathbf{\Delta}_1$  とすると、これらのベクトルとその成分は、後述するように、集中解析系の構成上重要な役割を演ずることになる。上述のように、 $\mathbf{G}_1$  は直接集中曲面の幾何学的特性量を与えるが、 $\mathbf{\Delta}_1$  は間接的にその役割を演じている。 $\mathbf{\Delta}_1$  に直接幾何学的表現を与え、その算出を可能とするものは、表1及び2のパーセント表の原形となる構造統計表である。

### 3. 集中多様体の構成とその意義

第2節において集中曲面は、統計表の構成と対応させて、集中曲線の形式的拡張として定義され、解析された。この曲面に、より実質的な意味を与えるものは、図5に示す集中多様体である。一般に与えられた統計的集団は、その任意の部分集団に対してその規模と標識量の配分比率の算定を可能としている。それは(2.4)式の積分領域を部分集団の存在領域  $D$  で置き換えることによって与えられる。その場合(2.5)式によって構成されるベクトル  $\mathbf{L} \{(x, y) \in D\}$  は  $UVW$  空間に対応する点として示される。一定の条件のもとで任意の部分集団の全体は  $\mathbf{L}$  を介して  $UVW$  空間内に凸集合を形成することがいえる。その境界曲面が集中多様体であり、それは集中曲面と、その同種曲面によって合成されることが示される。ここで同種曲面とは、(2.4)式における集中曲面成分  $x$  及び  $y$  に関する積分区間  $(-\infty, x)$  及び  $(-\infty, y)$  の一方または両

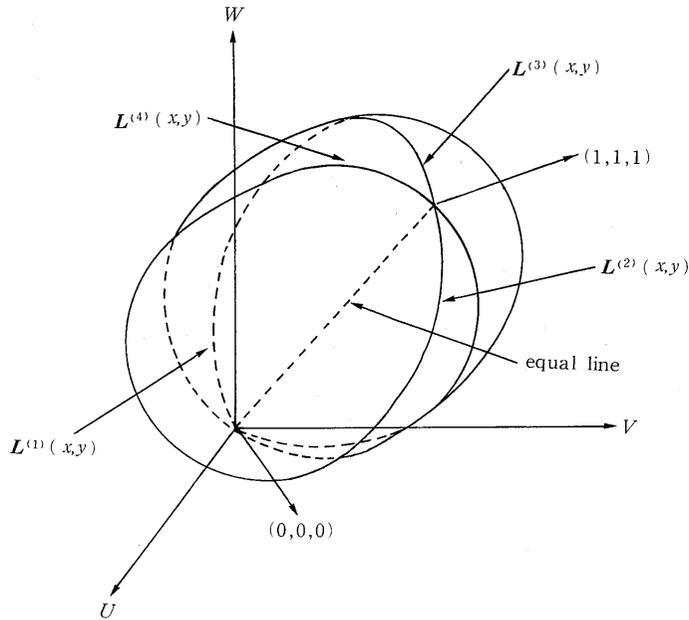


図5. 集中多様体

者を区間  $(x, \infty)$  及び  $(y, \infty)$  によって置き換えることによって生じた曲面であり、3種類の曲面として実現する。

以上によって集中多様体は確定した体積をもつことになり、それを集中体積といえることができる。この体積  $V$  は  $G_1, G_2; \Delta_1, \Delta_2$  と共に集中解析上重要な役割をもつことになるが、それはその構成から  $x, y$  に関する集中度係数  $G_{xy}$  とすることができる。

計算によると

$$(3.1) \quad G_{xy} = V = \frac{\Delta_{xy}}{4\mu_x\mu_y}$$

である。ここで  $\Delta_{xy}$  は

$$(3.2) \quad \Delta_{xy} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \{ \text{sgn}(x_1 - x_3) \} \{ \text{sgn}(y_2 - y_3) \} \begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} \cdot \prod_{i=1}^3 f(x_i, y_i) dx_i dy_i$$

で表現されるから (2.8) 式や (3.1) 式の関係からみて  $x, y$  に関する平均差といえることができる。

$x, y$  標識量の各個体への配分が均等化されると、集中多様体も空間的均等線に接近し、体積が、従って、 $G_{xy}$  及び  $\Delta_{xy}$  が 0 に近づくことはいうまでもない。

さて、 $G_1, G_2$  は集中解析上の基本統計量となるものであるが、これらは集中多様体上の微小面分を大きさとする法線ベクトル  $\mathbf{n}^{(i)}(x, y) dx dy$ ;  $i=1, 2, 3, 4$  の合成によって得られる。また  $G_{xy}$  は内積  $\mathbf{n}^{(i)}(x, y) \cdot \mathbf{L}^{(i)}(x, y) dx dy$  の総和で与えられる。ここで  $\mathbf{L}^{(1)}(x, y)$  及び  $\mathbf{n}^{(1)}(x, y)$  は集中曲面とその面上の法線ベクトルを表わし、他の  $\mathbf{L}^{(i)}(x, y)$  及び  $\mathbf{n}^{(i)}(x, y)$  は  $\mathbf{L}^{(1)}(x, y)$

の同種曲面とその面上の法線ベクトルを表わす。従って、 $\mathbf{n}^{(i)}(x, y)$ はこの解析の基礎となるベクトルであるが、それは

$$(3.3) \quad \mathbf{e}_1^{(i)}(x, y) = \frac{\partial}{\partial x} \mathbf{L}^{(i)}(x, y), \quad \mathbf{e}_2^{(i)}(x, y) = \frac{\partial}{\partial y} \mathbf{L}^{(i)}(x, y)$$

とすると

$$(3.4) \quad \mathbf{n}^{(i)}(x, y) = \mathbf{e}_1^{(i)}(x, y) \times \mathbf{e}_2^{(i)}(x, y); \quad i=1, 2, 3, 4$$

で表わされる。

$\mathbf{n}^{(i)}(x, y)$ を更に仔細に点検すると、集団内の任意特定の個体  $\mathbf{I}_1, \mathbf{I}_2$  間に働く相互作用素  $\mathbf{M}(\mathbf{I}_1, \mathbf{I}_2)$  を根源的要因とすることができる。つまり、 $\mathbf{I}_1, \mathbf{I}_2$  はそれぞれベクトル  $\left(1, \frac{x_1}{\mu_x}, \frac{y_1}{\mu_y}\right)$ ,  $\left(1, \frac{x_2}{\mu_x}, \frac{y_2}{\mu_y}\right)$  を表わすものとする、 $\mathbf{M}$  は

$$(3.5) \quad \mathbf{M} = \{S(\mathbf{I}_1, \mathbf{I}_2)\} \mathbf{I}_1 \times \mathbf{I}_2$$

で示される符号付外積量である。ここで、 $S(\mathbf{I}_1, \mathbf{I}_2)$  は  $(x_1, y_1), (x_1, y_2), (x_2, y_2)$  をそれぞれ  $xy$  平面上の3点  $P_{11}, P_{12}, P_{22}$  を表わすものとする、 $P_{ij}$  のこの順での回転が右廻りのとき +1, 左廻りのとき -1 の値をとる符号関数を意味する。

1変量の集中解析においては、2個体の標識量をそれぞれ  $x_1, x_2$  としたとき、計量経済学上その間に作用する相対的支配度  $D$  (degree of dominance) として

$$(3.6) \quad D = \frac{|x_1 - x_2|}{\mu_x}$$

が用いられる。(3.5)式の  $\mathbf{M}(\mathbf{I}_1, \mathbf{I}_2)$  はこれを展開すると

$$(3.7) \quad \mathbf{M}(\mathbf{I}_1, \mathbf{I}_2) = \text{sgn}(x_2 - x_1) \text{sgn}(y_2 - y_1) \\ \times \left( \frac{1}{\mu_x \mu_y} \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}, \frac{y_1 - y_2}{\mu_y}, \frac{x_2 - x_1}{\mu_x} \right)$$

となるから、 $D$  をその1つの成分として含んでいる。その意味で  $\mathbf{M}(\mathbf{I}_1, \mathbf{I}_2)$  を2変量の相対的ベクトル支配度とすることができる。また  $\mathbf{n}^{(i)}(x, y)$  は、特定の部分集団の内部での  $\mathbf{M}(\mathbf{I}_1, \mathbf{I}_2)$  の総和を意味することになる。

同様に、 $\mathbf{I}_1, \mathbf{I}_2$  のベクトルをそれぞれ  $(1, x_1, y_1), (1, x_2, y_2)$  とすると、絶対的相互作用素が得られ、それは2変量の絶対的ベクトル支配度を与える。またこのとき、 $\mathbf{A}_1, \mathbf{A}_2$  は、特定の部分集団の内部での絶対的相互作用素の総和をもとにして直接合成されることになる。

なお、2変量の支配度としてスカラー量が要請される場合は(3.2)式を構成する体積要素を基準にして考察する必要がある。従って、そのとき支配度は集団内の任意特定の3個体  $\mathbf{I}_1, \mathbf{I}_2, \mathbf{I}_3$  の間に成立する尺度概念であり、ベクトル3重積を用いて表現されることになる。

#### 4. ベクトル積率と記述統計量

これまで我々は度々、 $\mathbf{G}_1, \mathbf{G}_2; \mathbf{A}_1, \mathbf{A}_2$  の各成分及び  $G_{xy}; \Delta_{xy}$  は集中解析系で基本的な役割を果たすと述べてきた。実際にこれらは表3に示すようにそれらの間の積・商等の簡単な演算によって各種の記述統計量を与えることができるのであって、その役割はピアソン系におけるモメントに例えることができる。

表 3. 線形構造に対するベクトル積率の記述統計量

集中統計量の種類	集団構造に対する機能	
	一般的性格	事例的性格
(1) $\Delta_{12}, \Delta_{21}$	散布度を与える	正規分布に対しては $\Delta_{12} = 2\sigma_x/\sqrt{\pi} = \Delta_x, \Delta_{21} = 2\sigma_y/\sqrt{\pi} = \Delta_y$ .
(2) $-\Delta_{11}/\Delta_{21}, -\Delta_{22}/\Delta_{12}$	線形相関係数を与える	$E(Y X=x), E(X Y=y)$ が exact linear ならば (2) はモメント相関とそれが存在する限り一致する.
(3) $-\Delta_{11}/\Delta_{12}$	線形回帰係数を与える	$E(Y X=x) = \alpha_1 + \beta_1 x$
(4) $-\Delta_{10}/\Delta_{12}$	線形回帰係数を与える	ならば (3), (4) はそれぞれ $\beta_1, \alpha_1$ を与える.
(5) $(4/3)\Delta_{xy}$	二変量散布度を与える	正規分布に対しては $4\sigma_x^2\sigma_y^2(1-\rho^2)/(\pi\sqrt{1-\rho^2}/4)$ $= 4\Delta_x\Delta_y(1-\rho^2)$ $= 4/\pi V(x, y)$ . ここで $V(x, y)$ は $x, y$ に関する 2 次元分散を表わす.
(6) $(4/3)(\Delta_{21}/\Delta_{12})\Delta_{xy}$	(線形回帰残差) <sup>2</sup> を与える	正規分布に対して (6) は $4\sigma_y^2/(1-\rho^2)(\pi\sqrt{1-\rho^2}/4)$ $= \frac{\Delta_y^2(1-\rho^2)}{\sqrt{1-\rho^2}/4}$ $= \Delta_y^2(1-\rho^2)$ となる. また第 1 種パレート分布に対して (6) は $\Delta_y^2(1-1/p^2)\{1+1/8p^2+O(1/p^3)\}$ . この場合 $p$ は $X$ 及び $Y$ に関するパレート係数であり, $1/p$ は $XY$ 間の線形相関係数を与える.
(7) $(4/3)\Delta_{xy}/\Delta_x\Delta_y$	多様性又は不確実性係数を与える (多様性又は不確実性係数)	
(8) $\Delta_{10} + \mu_x\Delta_{11} + \mu_y\Delta_{12} = 0$	線形回帰条件	

また第 3 節では, これらが外積形式をもとにした相対的及び絶対的相互作用素の合成成分であることを示した. これらの意味において, 従来のモメントをスカラー積率(または内積率)とするならば, これらはベクトル積率(または外積率)といえよう.

表 3 は線形回帰を与える分布に対するベクトル積率の記述形式と, その性格を示すものである. 同表中の (6) 式はパレート分布に対して幅が充分大ならば正規分布の場合と同じ結果になることを示している. この場合  $p$  がそれ程大きくない場合が気になるが, (7) 式をもとにした検討結果は図 7 に示すように

$$(4.1) \quad 1 - \frac{1}{p^2} < \frac{1 - \frac{1}{p^2}}{\sqrt{1 - \frac{1}{4p^2}}} < \frac{4}{3} \frac{\Delta_{xy}}{\Delta_x\Delta_y}$$

となり, 3 式とも同じ傾向を示している. そのうえ, 集中統計量は 3 者のうち最大値をとるから (6) 式の残差の評価においても安全性の高い尺度を与えるものといえる. その結果は図 6 に示される.

因みに第 1 種パレート分布に対しては田口 (1984) に示すように

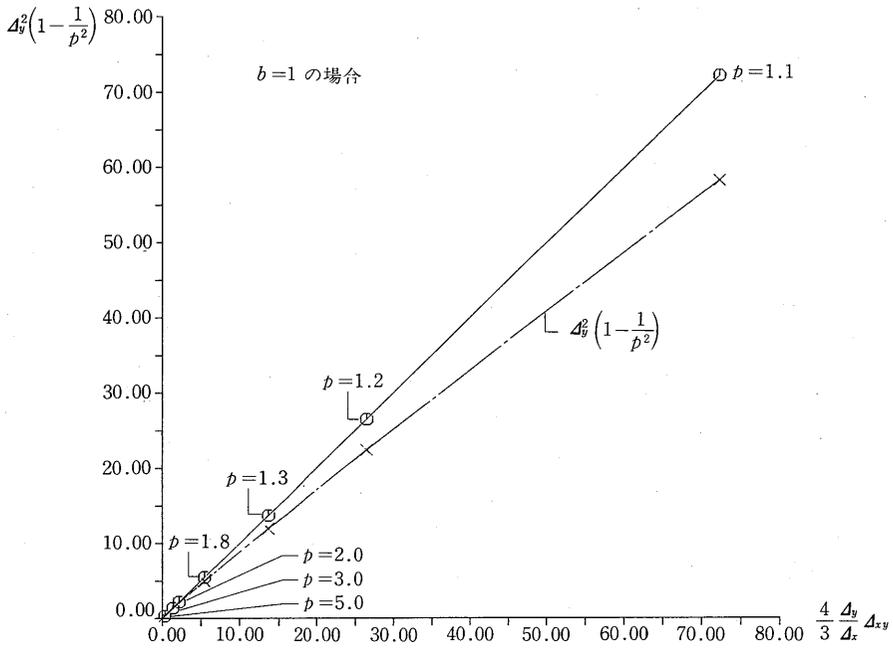


図6. 第1種パレート分布の集中線形回帰残差.

ここで、 $\Delta_x = \frac{2ap}{(p-1)(2p-1)}$ ,  $\Delta_y = \frac{2bp}{(p-1)(2p-1)}$

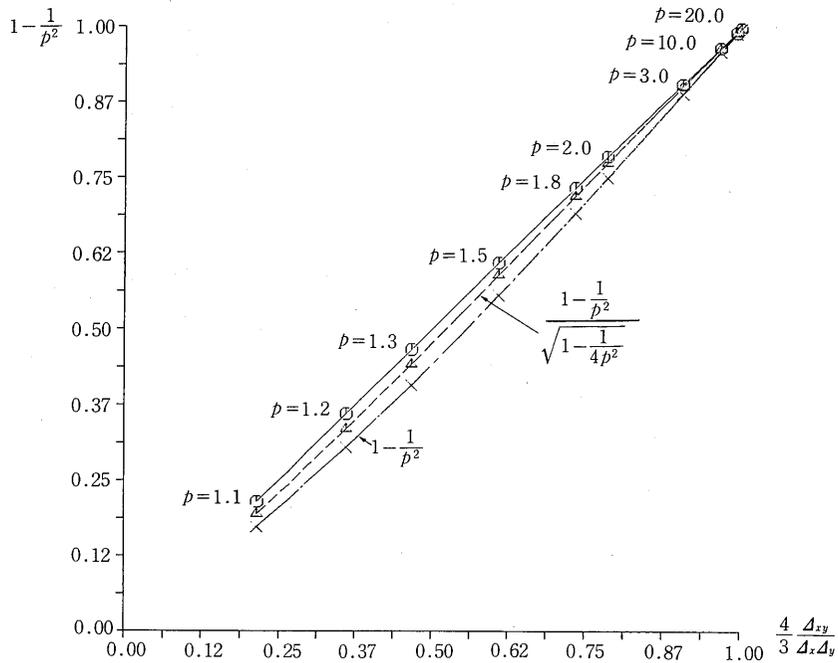


図7. 第1種パレート分布の多様性係数

$$(4.2) \quad \frac{4}{3} \frac{\Delta_{xy}}{\Delta_x \Delta_y} = \frac{4}{3} \left(1 - \frac{1}{p^2}\right) \{p^2 {}_3F_2(p-1, 1, 1; 2p, 2p; 1) - (p^2-1) {}_3F_2(p, 1, 1; 2p, 2p; 1)\}$$

が成立する。従って、

$$(4.3) \quad {}_3F_2(\alpha_1, \alpha_2, \alpha_3; \beta_1, \beta_2; 1) = 1 + \sum_{n=1}^{\infty} \frac{\Gamma(\alpha_1+n)}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_2+n)}{\Gamma(\alpha_2)} \frac{\Gamma(\alpha_3+n)}{\Gamma(\alpha_3)} \times \frac{\Gamma(\beta_1)}{\Gamma(\beta_1+n)} \frac{\Gamma(\beta_2)}{\Gamma(\beta_2+n)} \frac{1}{n!}$$

を級数展開することにより、容易に

$$(4.4) \quad \frac{4}{3} \frac{\Delta_{xy}}{\Delta_x \Delta_y} = \left(1 - \frac{1}{p^2}\right) \left\{1 + \frac{1}{8p^2} + O\left(\frac{1}{p^3}\right)\right\}$$

が得られる。

非線形回帰、特に定弾力性をもつ回帰曲線を与える分布に対しては、 $\Delta$ 形式の代わりに  $G$ 形式を用いて、同種の記述体系を構成することができる。モメント形式においては、この種の分布の解析の為に、対数変換を加えた統計量が用いられるが、 $G$ 形式は、しばしば、それより良い結果を与える。しかしその詳論は次の機会に譲ることにしよう。

## 5. 結 び

集中解析のシステムは、本来非対称で非線形回帰を与える複雑な分布の解析を目標として発達してきた。しかし、表3にみる結果はこの解析系がピアソン系の解析機能を十分に果たすのみでなく、経済統計に認められる特異分布に対してまで効果を及ぼす極めて適用範囲の広いものであることを物語っている。集中統計量は単に1次の絶対モメントの存在のみを前提として得られるからである。

因みに集中曲面の原像と各座標面への正射影との関係によって

$$(5.1) \quad E(|X|) < \infty, \quad E(|Y|) < \infty$$

のとき容易に

$$(5.2) \quad |\Delta_{11}| \leq \Delta_{21} < \infty, \quad |\Delta_{22}| \leq \Delta_{12} < \infty,$$

$$(5.3) \quad G_{10} \leq -G_{11}, \quad G_{20} \leq -G_{22},$$

等が成立する。更に、 $x, y$  が正ならば、

$$(5.4) \quad 0 \leq G_{xy} \leq \min(G_{12}, G_{21})$$

が得られる。

これらの事実は非線形構造についてもいえるのであって、結局、経済統計の記述等においては、官庁統計によく用いられる構造統計表が妥当であることを立証している。

## 謝 辞

本稿の全般にわたって山田晴美氏の協力を得た。ここに深く感謝する。

## 参 考 文 献

- Arnold, B.C. (1986). *Majorization and the Lorenz Order ; A Brief Introduction*, Lecture Notes in Statistics, 43, (eds. D. Brillinger, S. Fienber, J. Gani, J. Hartigan and K. Krickeberg), Springer-Verlag.
- Cifarelli, D.M. and Regazzini, E. (1986). On a general definition of concentration function, Quaderno 18/1986, Dipartimento di Matematica, Universita degli Studi di Milano.
- Taguchi, T. (1972a). On the two-dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to the two dimensional case I, *Ann. Inst. Statist. Math.*, **24**, 355-381.
- Taguchi, T. (1972b). On the two-dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to the two dimensional case II, *Ann. Inst. Statist. Math.*, **24**, 599-619.
- Taguchi, T. (1973). On the two-dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to the two dimensional case III, *Ann. Inst. Statist. Math.*, **25**, 215-237.
- Taguchi, T. (1981). On a multiple Gini's coefficient and some concentrative regressions, *Metron*, **XXXIX**-N. 1-2, 69-98.
- 田口時夫 (1983). 二変量パレート分布の集中解析 —— 構造統計量の提案 ——, 統計数理研究所彙報, **31**, 1-32.
- 田口時夫 (1984). 経済分析と多次元解析 —— 新しい計量空間の形成と展望 ——, 東洋経済新報社.
- Taguchi, T. (1987). On the structure of multivariate concentration—Some relationships among the concentration surface and two variate mean difference and regressions—, *Comput. Statist. Data Anal.*, **6**, North-Holland, 307-334.
- Tranquilli, G.B. (1985). Concentrazione Semplice tra Due Variabili Ponderate Congiunte Non Negative, Caso Assolutamente continuo, Dipartimento di Statistica, Probabilità e Statistiche Applicate, Serie A-Ricerche, N. 16, Universita Degli Studi di Roma "la Sapienza".

On the Concentration Manifold and the System  
of Concentration Analysis (I)  
— A Geometrical Development of Gini's Methodology —

Tokio Taguchi

(The Institute of Statistical Mathematics)

The author had already extended the concentration curve (Lorenz curve) to the concentration surface in his previous papers (Taguchi (1972a, 1972b, 1973)). In these papers, he found out some statistical measures similar to Gini's concentration coefficient and mean difference in the geometrical characteristics of this surface. Furthermore, he derived the two-dimensional concentration coefficient and mean difference from the volume of the concentration manifold.

In this paper, he points out that these measures have the same roles as the traditional product moments in statistical description, as shown in Table 3.

He suggests to call these new fundamental statistics 'vector product moments' and the traditional moment 'scalar product moments'.

Both systems of description are compared with each other. The new system is distinguished in some points of view.

For example, the new system has a wider scope of application than the traditional one.

On the whole, he verified the validity of such a kind of structural statistical table as Tables 1 and 2.