

経済統計における分布と 分配構造の解析法の研究

—ノルム p をもつ統計とベクトル値確率の基礎概念—

田 口 時 夫

(1974年7月 受付)

On the Size and Share Distributions in Economic Statistics

—Fundamental notions of statistics with norm p and vector valued probability—
Tokio Taguchi

In this paper, the author tries to explain the fundamental notions of his recent papers on the statistics with norm p and concentration curve and surface (see Taguchi, T [38], [31] and [34]) as scalar and vector probabilities.

The first paper [38] on the statistics with norm p treats the size distribution from the wider field of vision than the conventional mathematical statistical method (originated from Pearson's method of moment), which is restricted to the special case of norm 2, while the other papers [31] and [34] is concerned with share distribution (developed by Italian School) through the application of the method of vector analysis.

In this paper, the methods of analysis mentioned in the above papers are described more concretely through some real examples.

And in this case, the author introduces newly two fundamental notions, in order to give a unified form to each method. Namely the pan-expectation $E_h^{(p)}\{g(X)\}$ (to be exact, the expectation of $g(X)$ with respect to $h(X)$ with norm p) is introduced to the section 1 and the element of interaction $h(x_1, x_2) dx_1 dx_2$ (to be exact, the element of interaction defined between two infinitesimal classes (x_1, x_1+dx_1) and (x_2, x_2+dx_2)) to the section 2.

Lastly, the author arrived at the more fundamental notions of the scalar probability constituting the probability space with norm p (abbreviated $(\mathcal{Q}, B, P; p)$) and vector probability constituting the share vector space (abbreviated (\mathcal{Q}, B, T)) as extension of conventional probability for analysis of economic data.

The Institute of Statistical Mathematics

目 次

はしがき

第1章 規模分布の解析法について

第1節 規模分布の解析法の概要と問題点

第2節 汎期待値とモーメント解析法の一般化及び其等の解法

第3節 ノルム1の統計とピアソン第III型

第2章 特殊な分配構造の解析法について

第1節 特殊な分配構造の表現法とその概要

第2節 ローレンツ曲線の幾何学的諸性質

第3節 特殊な分配構造の記述的諸測定

第4節 集中解析法の理論的構成とジニ指数の意義

第3章 一般的な分配構造の解析法について

第1節 一般的な分配構造の具体的表現とその概要

第2節 一般的分配率ベクトルの理論的定義と依存度

第3節 ベクトル値確率とクラスター及びモデル解析について

むすび

は し が き

今世紀の初頭に、K. Pearson と C. Gini は、相異なる統計解析の方法を示した。(Castellano V [3] 参照) 特に前者における其の後の発達と普及は著しいものがあるが、反面後者に関する方法論的意義を看過する傾向を生んだ事を否定する事が出来ない。

本稿は、経済統計を中心にして、彼等のモメント法及び集中曲線法を一般化し発展させた筆者の最近の諸論文をより体系的に叙述したものであるが、上述の理由により後者の叙述には特に力を入れたつもりである。結論として筆者は、経済統計、特に所得分布に対して有効な解析法を与える為には、基礎となる確率諸概念の拡充が必要であると考え、それは具体的には、従来の確率空間 (Ω, B, P) を拡張したノルム p の確率空間 (Ω, B, P, p) 、及び P を分配ベクトル \mathbf{T} で置き換えた分配空間 (Ω, B, \mathbf{T}) 等の構成と展開として行われるべきものと思われる。其等の間関係は又スカラー値確率に対するベクトル値確率の関係ともいえる。猶、本稿の第1, 2章は主に筆者のこれまでの研究論文を集大成したものであるが、直接的には労働省官房政策課 [23] の委託に基づいて執筆した「所得・賃金統計のデータ解析の基礎」を母胎として加筆補正したものである。その際、新たに汎期待値 $E_h^{(p)}\{g(X)\}$ 及び微少クラス間に定義される相互作用素 $h(x_1, x_2) dx_1 dx_2$ を導入することにより各章の後半に於てそれぞれの方法に理論的に統一した形式を与えた。又第3章は、全く新たに起草した部分であり、第2章を更に包括的に発展させた解析法を与えるものである。猶、その方法論としての内容は、必ずしも経済統計に限定される必要はなく、事実、分布の歪みや、統計的順序法則を対象とする点で人口 (Zipf, G. K [48], 館稔 [28] 参照), 社会 (安田 [44] 参照) 言語 (Zipf, G. K [48], Heldon, G. [12], Mandelbrot, B [20] 参照), 生物 (Simon, H. A [27] 参照) 等に関連を見出せるが本稿に於ては、説明に具体性と一貫性を特たせる意味で、敢て経済部門に制限したものである。

第1章 規模分布の解析について

第1節 規模分布の解析法の概要と問題点

経済的規模分布に関する統計解析に於ては、通常 n 個の各規模階級別に、その代表値 x_i と相対的客体数 f_i とを対応させた、いわゆる経験的規模分布を対象として、累乗平均形式をもつモメント μ_h

$$(1) \mu_h = \sum_{i=1}^n f_i x_i^h : h = 1, 2, \dots$$

によるその特徴づけから出発する。その理論的基礎は、確率論によって歴史的に形成されて来たことはいうまでもない。

モメント又はその関数のうち、特に加法平均 μ_1

$$(2) \mu_1 = \sum_{i=1}^n f_i x_i$$

及び分散 σ^2 (又は標準偏差 σ)

$$(3) \sigma^2 = \mu_2 - \mu_1^2 = \sum_{i=1}^n f_i x_i^2 - \mu_1^2$$

は、主要な特性量であり、此等は分布の具体的関数型に拘わりのない普遍的形式をとるがこの方式は経済量に関しても主に資料 II のような比率分布にとつてはそれ程無理のない方式と思われる。然し、一方資料 I に示される数量分布に限定するとより基本的な問題が生ずる。例えば、代表的な所得分布、いわゆるパレート分布

$$(4) f(x) = \frac{\alpha \theta^\alpha}{x^{1+\alpha}}; x \geq \theta, \alpha > 1$$

を例にとれば、この分布は

$$(5) 1 < \alpha \leq 2$$

に於て、加法平均は $\mu_1 = \frac{\alpha \theta}{\alpha - 1}$ であるが、分散が無大となり、数理統計学的理論の上では特異分布として屢々除外の対象とされているのである。

更に附言するならば、人口密度の表現等に適用されるジップ分布

$$(6) f(x) = \frac{A}{x^k}; x \geq x_0, k > 1$$

については

$$(7) 1 < k \leq 2$$

に於て加法平均すらも無大となる事がよく知られている (Zipf, G. K [48] 及び拙稿 [33] 参照)

然し数理統計学的方法上から特異であっても、資料 I のような経験的データによると、(5) はよく適合する関数であり、事実パレート自体 $\alpha=1.5$ を標準値と考えているのである (Sinon H. A [27]).

このことは次章に述べるように、分布表現に代りローレンツ曲線法を登場させる有力な根拠であるが、他方加法平均一分散といった狭い基本形式の設定に問題を見出すことも出来る。

事実、経験資料の解析から出発した数理統計学の創始者達は、F. Galton を始め必ずしも加法平均や二乗偏差形式に固執してはいない (前掲 [7] 参照)。K. Pearson に於てすらその分布第 3 型との関連に於て所得及び生物諸現象を扱った場合、その歪みに異常性を認め、此の分野におけるモーメント法 (the method of moments) の適用結果に対し必ずしも満足していない。(Pearson, E. S. [22] 参照)。又数理統計学の古典的諸著作 (例えば Yule, G. U. [46] 及び [47], Kendall M. G. [16], Cramér, H. [4] 参照) に於ては、各種の代表値、或はそれに対応する各種の回帰形式の存在と可能性について簡単であっても一応の説明を加えており、決して省略はしていない。此の点に関する配慮は現在既に全く不要なのであろうか。たしかに現代の多くの数理統計学 (例えば Wilks, S. S [42] 等参照) や計量経済学 (例えば Johnston. J. [14] 参照) に関する教科書によれば、統計資料に対しては既に基本的には単一の解析体系と、その基準としての正規分布が存在するに過ぎないようである。此の平均二乗偏差への根強い執着が、統計解析を、依然として、期待値を軸とする確率論 (Kolmogoroff, A [17] 参照) に従属させ、その単なる応用 (例えば Feller, W [8] 参照) に固定させ、場合によっては具体的対象と抽象的方法の遊離を増大させる役割を果たしているようである。こうした問題意識は例えば佐和氏 ([25] 参照) にも強く認められる処である。

扱、Fechner, G. T. [7] によれば、加法平均を含めて代表値とは

$$(8) G^{(p)}(a) = \left\{ \sum_{i=1}^n |x_i - a|^p f_i \right\}^{\frac{1}{p}}; p \geq 1$$

を最小化する a として定義される。

$p > 1$ に対しては、こうした a は常に一義に定まり、又 $p = 1$ の場合に於ても、それに適合する理論形式 $f(x)$ が連続で凸領域で正の値をとると考えられる場合には略問題は生じない。

この一般化された定義に基づく代表値 $l^{(p)}$ は

$$(9) \sum_{x_i < l^{(p)}} (l^{(p)} - x_i)^{p-1} f_i = \sum_{x_i > l^{(p)}} (x_i - l^{(p)})^{p-1} f_i$$

を満たす、即ち $p-1$ 次の $l^{(p)}$ を中心にした絶対値の累乗和を左右に等分する点であることが見出される。(田口 [38])

一方、標準偏差を含めた散布度は一般に

$$(10) \quad d^{(p)} = \sqrt[p]{\sum_{i=1}^n |x_i - l^{(p)}|^p f_i}$$

と定義される、

此の定義は、特に $p=2$ の場合、従来通りの加法平均—標準偏差系を成立させるが、もし、 $p=1$ を採用すれば、メディアン—メディアンに関する平均偏差系を得ることはよく知られている。

更に

$$(11) \quad 0 < p \leq 1$$

の場合は (8) 式の代りに

$$(12) \quad G^{(p)}(a) = \left\{ \sum_{i=1}^n |x_i^{-1} - a^{-1}|^p f_i \right\}^{-\frac{1}{p}}$$

を用いることが出来、何れの場合に於ても、計算機の高度に発達した今日に於ては、その数値解析は困難ではない。

以上の Fechner の予想を論理的に展開すると後節に述べるように相関係数及び回帰係数も一般化されモメント法自体の拡張が可能となる。

特に当初に述べたパレート分布について言えば、 p を $p < a$ の範囲で選ぶと有限の散布度が得られ、数理統計学的データ解析法に準ずる解析を加えることが出来る。これは、 $p=1$ による解析が L. P. 法により現在普及しつつある事態に対応しているといえよう。

然し、一般的に p の解析上の最適値を与える具体的な選択基準が存在し得るか否かについては尚幾つかの検討が必要であり、その為には後節にふれるように分布の具体的構造に立入った分析も必要となるであろう。

然し何れにせよ、統計的分布に対する解析法は、代表値（屢々真値或いは中心と称される）及び散布度の存在が前提とされていることに注目せねばならない。勿論例えば、幾何平均は、(8) 式による定義の直接の産物ではないが、 x 及び a を対数変換する操作を加えることにより間接的に得られるから加法平均の一種とみて差支えないであろう。

同様に、例えば所得分布の一方の代表として知られるジブラ分布

$$(13) \quad f(x) = \frac{1}{\sqrt{2\pi} \sigma' x} e^{-\frac{1}{2} \frac{(\log x - \mu')^2}{\sigma'^2}}$$

に対しては、規模 x を対数変換した上で数理統計学的手法を適用することは、計量経済学等に於てよく行われるが、規模変換の意味が問われることを除けば、上記の分布に対する解析的観点の例外ではない。

このような分布観が、直接集団構造を正確に反映しているか否かは、勿論、具体的現象と経験科学的事実に照してその都度、究明さるべきであると共に、方法論的立場に於ても第2章の分配構造に対する観点との比較に於て吟味さるべきであり、事実それは本稿の最終課題の一つであるが、差当り本章では、次節で Fechner の代表値論 (Potenzmittelwerthen 前掲 [7] 参照) をもとにして、汎期待値論とでも言うべき広い視野の下で、モメント法を拡大再構成することにより、逆にその本質の所在を究明することにしよう。

第2節 汎期待値とモーメント解析法の一般化及び其等の解法

加法平均は、確率論的立場に於ては、期待値に当るが、本節のように平均値を一般的に代表

値の一つとして考察しようとする場合には、それに対応して期待値概念を拡張することが便利であり、見通しのよい結果を与えることになる。以下それを具体的に拙稿（前掲 [38] 参照）に即して解説することにしよう。

本質的な結果を効果的に得る為に、本節に於ては、前節で扱った相対的度数分布の代りに、専ら次の性質を備えた理論的密度関数を対象とする。

仮説 1 『確率変数 X の密度関数 $f(x)$ は、ある区間 I で定義され、可附番個の点を除いて連続且正值であるものとする。』

一方、此に対する基本的解析概念として、汎期待値を次のように定義する。

定義 1 『一般に x の二つの連続関数

$$(14) \quad h = h(x), \quad g = g(x)$$

及び任意の実数

$$(15) \quad p \geq 1$$

に対して、 $g(X)$ の h に関する norm p の期待値 $E_h^{(p)}\{g(X)\}$ を次のように定義する。即

$$(16) \quad E_h^{(p)}\{g(X)\} = E\{g(X)|h(X)|^{p-2}\}$$

である。』

扱、此の記法によると

$$(17) \quad E_{X-x}^{(p)}(X-x) = E\{[\text{sgn}(X, x)]|X-x|^{p-1}$$

但し

$$(18) \quad \text{sgn}(X, x) = \begin{cases} -1 & X < x \\ 0 & X = x \\ +1 & X > x \end{cases}$$

となるから、前節の Fechner の代表値は、次のように再定義される。

定義 2 『今 $G(x)$ を

$$(19) \quad G(x) = \sqrt[p]{E|X-x|^p} = \sqrt[p]{E_{X-x}^{(p)}(X-x)^2}$$

とし、 $G(x)$ の極小値を与える x が存在するとき、それを X の norm p の代表値 $l^{(p)}$ 又は $l^{(p)}(X)$ という。又その場合の極小値を norm p の散布度といい、 $D^{(p)}(X)$ で表わす』

この時 Minkowski の不等式 (Shisha, O [26] 参照) により容易に次の結果が得られる。

定理 1 『仮説 1 が成立するとき、 $p \geq 1$ で $l^{(p)}$ は、常に只一つ存在し、次の方程式をみたす。即

$$(20) \quad E_{X-l^{(p)}}^{(p)}\{X-l^{(p)}\} = 0$$

又此の時、散布度は

$$(21) \quad D^{(p)}(X) = \sqrt[p]{E_{X-l^{(p)}}^{(p)}\{X-l^{(p)}\}^2}$$

で表わされる。』

扱以上の結果を確率ベクトル (X, Y) に拡張することにする。この時、仮説 1 及び定義 1 に対応して

仮説 2 『 (X, Y) の密度関数 $f(x, y)$ は、ある凸領域 D で定義され、可附番個の点を除き連続且正值であるものとする。』

定義 3 『一般に (x, y) の二つの連続関数

$$(22) \quad h = h(x, y), \quad g = g(x, y)$$

及び (15) をみたす任意の実数に対して $g(X, Y)$ の h に関する norm p の期待値 $E_h^{(p)}\{g$

(X, Y) を次のように定義する. 即

$$(23) E_h^{(p)}\{g(X, Y)\} = E\{g(X, Y) | h(X, Y) |^{p-2}\}$$

である.]

今, (X, Y) の散布度及び X, Y の相関を求める為に更に次のような定義を加えることにする.

定義 4 $[X, Y$ の X 及び Y に関する norm p の共散布度 $\text{Codisp}_X^{(p)}(X, Y)$ 及び $\text{Codisp}_Y^{(p)}(X, Y)$ を次のように定義する.

$$(24) |\text{Codisp}_X^{(p)}(X, Y)|^{\frac{p}{2}} = |E_{X-l^{(p)}(X)}^{(p)}[\{X - l^{(p)}(X)\}\{Y - l^{(p)}(Y)\}]|$$

$$(25) \text{sign of } \text{Codisp}_X^{(p)}(X, Y) = \text{sign of } E_{X-l^{(p)}(X)}^{(p)}[\{X - l^{(p)}(X)\}\{Y - l^{(p)}(Y)\}]$$

更に (X, Y) の二次の散布度 $D^{(p)}(X, Y)$ 及び X, Y の共散布度 $\text{Codisp}_p^{(p)}(X, Y)$ 及び相関係数 $\rho_{XY}^{(p)}$ をそれぞれ次のように定義する. 即

$$(26) |\text{Codisp}_p^{(p)}(X, Y)| = \sqrt{|\text{Codisp}_X^{(p)}(X, Y) \text{Codisp}_Y^{(p)}(X, Y)|}$$

$$(27) \text{sign of } \text{Codisp}_p^{(p)}(X, Y) = \text{sign of } \{\text{Codisp}_X^{(p)}(X, Y) + \text{Codisp}_Y^{(p)}(X, Y)\}$$

$$(28) D^{(p)}(X, Y) = \sqrt[p]{\left| \det \begin{pmatrix} \{D^{(p)}(X)\}^{\frac{p}{2}} & |\text{Codisp}_p^{(p)}(X, Y)|^{\frac{p}{2}} \\ |\text{Codisp}_p^{(p)}(X, Y)|^{\frac{p}{2}} & \{D^{(p)}(Y)\}^{\frac{p}{2}} \end{pmatrix} \right|}$$

$$(29) \rho_{XY}^{(p)} = \frac{\text{Codisp}_p^{(p)}(X, Y)}{D^{(p)}(X) D^{(p)}(Y)}$$

とする.]

此の場合 Minkowski 及び Hölder の不等式 (前掲 [14] 参照) を交互に適用すれば容易に次の定理が得られる.

定理 2 『一般に任意の $p \geq 1$ に対して

$$(30) |\rho_{XY}^{(p)}| \leq 1$$

である.

特に X, Y が独立ならば

$$(31) \rho_{XY}^{(p)} = 0$$

が成立する. 又, 完全正又は負相関関係にあるとき,

$$(32) \rho_{XY}^{(p)} = 1 \quad (\text{又は } -1)$$

となる.]

では norm p の回帰直線は approximate form ではどのように定義出来るであろうか.

定義 5 『今

$$(33) G(a, \beta) = E\{|Y - aX - \beta|^p\} = E_{Y-aX-\beta}^{(p)}(Y - aX - \beta)^2$$

とするとき, (33) の極小値が存在し, そのとき $(a, \beta) = (a^{(p)}, b^{(p)})$ ならば,

$$(34) y = a^{(p)}x + b^{(p)}$$

を y の x に関する最少 p 乗回帰直線とする.]

代表値, 相関に関するこれまでの考察と同様に我々は次の定理を容易に得ることが出来る.

定理 3 『仮説 2 の下で, 任意の $p \geq 1$ に対して, y の x に関する最少 p 乗回帰直線は

常に只一つ存在する。その場合回帰係数は、次の一般化 normal equation を満足する。即

$$(35) \quad E_{Y-a^{(\phi)} X-b^{(\phi)}}^{(\phi)} (Y - a^{(\phi)} X - b^{(\phi)}) = 0$$

及び

$$(36) \quad E_{Y-a^{(\phi)} X-b^{(\phi)}}^{(\phi)} \{X (Y - a^{(\phi)} X - b^{(\phi)})\} = 0$$

である。』

更に、残差 $\zeta^{(\phi)}$ を

$$(37) \quad \zeta^{(\phi)} = Y - a^{(\phi)} X - b^{(\phi)}$$

とするとき

定理 4 『 y の x に関する最少 ϕ 乗回帰直線の残差 $\zeta^{(\phi)}$ は常に次の方程式を満足する。

$$(38) \quad E_{\zeta}^{(\phi)} (\zeta^{(\phi)}) = 0$$

即 $\zeta^{(\phi)}$ の norm ϕ に関する代表値は 0 である。』

一般のモーメント及び確率ベクトルに対しては、以上の所論から容易にその拡張定義を与えることが出来る。

例えば (s, t) 次の原点の周りの moment は X 及び Y に関してそれぞれ

$$(39) \quad \mu_{st;X}^{(\phi)} = E_X^{(\phi)} (X^s Y^t)$$

$$(40) \quad \mu_{st;Y}^{(\phi)} = E_Y^{(\phi)} (X^s Y^t)$$

とすることが出来る。

Fechner は更に norm ϕ に対して典型的分布の存在を想定している (前掲 [7] 参照)。

正規分布を $\phi=2$ の場合の典型とすれば、我々は一般に norm ϕ に関して

$$(41) \quad f(x) = \frac{\phi \left\{ \frac{|x-\alpha|}{\beta} \right\}^{\frac{\phi}{2}-1}}{4\pi\beta} e^{-\frac{1}{2} \left\{ \frac{|x-\alpha|}{\beta} \right\}^{\phi}}$$

を一つの典型分布とする根拠がある。すなわち、此の場合

$$(42) \quad \alpha = l^{(\phi)}(X), \quad \beta = D^{(\phi)}(X)$$

が容易に得られるからである。

以上を集約すると、確率空間 (Ω, B, P) の拡張として $(\Omega, B, P; \phi)$ を考察する必要性があり且可能であるということである。

では具体的に与えられた分布に対して、 $\phi, l^{(\phi)}, D^{(\phi)}(X)$ は如何なる関係を示すであろうか。

今、例えば $D^{(\phi)}(X)$ の極値を考える為に、Lagrangian

$$(43) \quad L(l^{(\phi)}, \phi) = D^{(\phi)}(X) + \lambda E \{ \text{sgn}(X, l^{(\phi)}) |X - l^{(\phi)}|^{\phi-1} \} \\ = \sqrt[\phi]{E |X - l^{(\phi)}|^{\phi}} + \lambda E \{ \text{sgn}(X, l^{(\phi)}) |X - l^{(\phi)}|^{\phi-1} \}$$

を構成すると、

$$(44) \quad \frac{\partial L}{\partial l^{(\phi)}} = - \{D^{(\phi)}(X)\}^{\frac{1}{\phi}-1} E \{ \text{sgn}(X, l^{(\phi)}) |X - l^{(\phi)}|^{\phi-1} \} \\ - \lambda (\phi - 1) E |X - l^{(\phi)}|^{\phi-2}$$

となる。

従って $D^{(\phi)}(X)$ が極値を持つ為には分布の如何に拘らず $\phi=1$ でなければならない。

一方、 $\phi=2$ については特に極立つ特徴は認められない。然し、分布によっては例えば $D^{(\phi)}(X)$ の飽和値 (曲率の極値) 等を与える可能性はある。

何れにせよ $\phi=1$ について特に改めて吟味する必要がある。

勿論、 $p=2$ に関する統計は、屢々ユークリッド距離概念や、物理学的回転、慣性率或は重心や、エネルギー概念とのアナロジーにより、自然科学者に対し説得力を示している。然し単一の貨幣単位で表現される、或は換算されて利用される事の多い経済量の諸統計に於ては、そのような類推の余地はなく寧ろ $p=1$ により諸計算が抵抗なく適用され且説得力をもっていることも事実である。

処で以上の諸事実によって、もし $p=2$ が経済統計学上、何等の構造的特性を持ち得ないとしても、その理論上の先駆的役割と共に、計算上の実用的価値は、決して消滅するものではない。すなわち、もしある $p \neq 1, 2$ が採択されたとして (20) 式をデータにより解く事を考えると、例えば Newton の $f(x) = 0$ の解法に準ずる方法が存在する。この場合

$$(45) \quad \frac{d}{dl^{(p)}} = - (p-1) E_{X-l^{(p)}}^{(p-1)(p)} \{X - l^{(p)}\}$$

を考慮すれば、それが0でない限り $l^{(2)}$ を初期値 $l_0^{(p)}$ とし逐次

$$(46) \quad l_{i+1}^{(p)} = l_i^{(p)} + \frac{E_{X-l_i^{(p)}}^{(p)} \{X - l_i^{(p)}\}}{(p-1) E_{X-l_i^{(p)}}^{(p-1)(p)} \{X - l_i^{(p)}\}} \quad i = 0, 1, 2, \dots$$

による近似を行うことが出来るからである。

特に $p=1$ の場合は別個に L.P 法により直接算出するか、又は (46) の代りに反復法

$$(47) \quad l_{i+1}^{(1)} = l_i^{(1)} + \frac{E|X - l_i^{(1)}|}{2f(l_i^{(1)})}; \quad i = 0, 1, 2, \dots$$

を用いることが出来るが、(47) は実際的ではないし、解が一意であるともいえない。

同様の事は、(36) 及び (37) の normal 方程式をデータにより解く場合についても言える。すなわち、 $p \neq 1, 2$ の場合

$$(48) \quad \frac{\partial E_{Y-a_i^{(p)} X-b_i^{(p)}}^{(p)} (Y - a_i^{(p)} X - b_i^{(p)})}{\partial a_i^{(p)}} = - (p-1) E_{Y-a_i^{(p)} X-b_i^{(p)}}^{(p-1)(p)} \{X (Y - a_i^{(p)} X - b_i^{(p)})\}$$

等の結果を考慮して、最少二乗回帰係数 $a^{(2)}, b^{(2)}$ を $a_0^{(p)}, b_0^{(p)}$ として反復法

$$(49) \quad a_{i+1}^{(p)} = a_i^{(p)} + h_i^{(p)}, \quad b_{i+1}^{(p)} = b_i^{(p)} + k_i^{(p)}; \quad i = 0, 1, 2, \dots$$

を適用する事になる。但し

$$(50) \quad h_i^{(p)} = \frac{\begin{vmatrix} E_{Y-a_i^{(p)} X-b_i^{(p)}}^{(p)} (Y - a_i^{(p)} X - b_i^{(p)}) & E_{Y-a_i^{(p)} X-b_i^{(p)}}^{(p-1)(p)} (Y - a_i^{(p)} X - b_i^{(p)}) \\ E_{Y-a_i^{(p)} X-b_i^{(p)}}^{(p)} \{X (Y - a_i^{(p)} X - b_i^{(p)})\} & E_{Y-a_i^{(p)} X-b_i^{(p)}}^{(p-1)(p)} \{X (Y - a_i^{(p)} X - b_i^{(p)})\} \end{vmatrix}}{(p-1) \Delta_i^{(p)}}$$

$$(51) \quad k_i^{(p)} = \frac{\begin{vmatrix} E_{Y-a_i^{(p)} X-b_i^{(p)}}^{(p-1)(p)} \{X (Y - a_i^{(p)} X - b_i^{(p)})\} & E_{Y-a_i^{(p)} X-b_i^{(p)}}^{(p)} (Y - a_i^{(p)} X - b_i^{(p)}) \\ E_{Y-a_i^{(p)} X-b_i^{(p)}}^{(p-1)(p)} \{X^2 (Y - a_i^{(p)} X - b_i^{(p)})\} & E_{Y-a_i^{(p)} X-b_i^{(p)}}^{(p)} \{X (Y - a_i^{(p)} X - b_i^{(p)})\} \end{vmatrix}}{(p-1) \Delta_i^{(p)}}$$

$$(52) \quad \Delta_i^{(p)} = \begin{vmatrix} E_{Y-a_i^{(p)} X-b_i^{(p)}}^{(p-1)(p)} \{X (Y - a_i^{(p)} X - b_i^{(p)})\} & E_{Y-a_i^{(p)} X-b_i^{(p)}}^{(p-1)(p)} (Y - a_i^{(p)} X - b_i^{(p)}) \\ E_{Y-a_i^{(p)} X-b_i^{(p)}}^{(p-1)(p)} \{X^2 (Y - a_i^{(p)} X - b_i^{(p)})\} & E_{Y-a_i^{(p)} X-b_i^{(p)}}^{(p-1)(p)} \{X (Y - a_i^{(p)} X - b_i^{(p)})\} \end{vmatrix}$$

である。

特に $h=1$ の場合は, (50), (51), (52) の代りに

$$(53) \quad h_i^{(1)} = \frac{\begin{vmatrix} P(S_1) - P(S_2) & \int_{-\infty}^{\infty} f(x, a_i^{(1)}x + b_i^{(1)}) dx \\ E(X|S_1)P(S_1) - E(X|S_2)P(S_2) & \int_{-\infty}^{\infty} xf(x, a_i^{(1)}x + b_i^{(1)}) dx \end{vmatrix}}{\Delta_i^{(1)}}$$

$$(54) \quad \hat{h}_i^{(1)} = \frac{\begin{vmatrix} \int_{-\infty}^{\infty} xf(x, a_i^{(1)}x + b_i^{(1)}) dx & P(S_1) - P(S_2) \\ \int_{-\infty}^{\infty} x^2f(x, a_i^{(1)}x + b_i^{(1)}) dx & E(X|S_1)P(S_1) - E(X|S_2)P(S_2) \end{vmatrix}}{\Delta_i^{(1)}}$$

$$(55) \quad \Delta_i^{(1)} = \begin{vmatrix} \int_{-\infty}^{\infty} xf(x, a_i^{(1)}x + b_i^{(1)}) dx & \int_{-\infty}^{\infty} f(x, a_i^{(1)}x + b_i^{(1)}) dx \\ \int_{-\infty}^{\infty} x^2f(x, a_i^{(1)}x + b_i^{(1)}) dx & \int_{-\infty}^{\infty} xf(x, a_i^{(1)}x + b_i^{(1)}) dx \end{vmatrix}$$

但し

$$(56) \quad S_1 = \{(X, Y); Y - a_i^{(1)}X - b_i^{(1)} > 0\}$$

$$S_2 = \{(X, Y); Y - a_i^{(1)}X - b_i^{(1)} < 0\}$$

を用いることになるからデータに適用する場合実際的にも、一義的でもない。従って此の場合も、当初 W. D. Fisher 等により定式化された L. P 法が有効とならう。

従って $p=1$ は此の意味でも $p=2$ に対して相対的に独立しており、別個に考察される余地が十分に存在する。

norm p の線型回帰方程式に関しては、以上の観点とは別に、例えば最尤法との比較といった専ら推定上の観点から佐和氏が分析しておられる（後掲 [25] 参照）。そこでは $p=1$ の線型回帰は、例えば誤差項の分布が両側指数分布の場合最尤線型回帰となることが指摘されている。更に、両側指数分布の位置母数の最尤推定量は、サンプルメディアンである事が、系統推定量についての竹内氏の研究（後掲 [41] 参照）に於て指摘されている。両氏の研究によっても、 p の撰択は、分布の裾の広狭に影響されることが容易に理解されるのである。猶ほ標本代表値が母集団の同じ norm をもつ代表値の不偏推定量であるか否かについて筆者の試論（後掲 [38] 参照）があるが、標本についての一般的な検討は今後の課題といえよう。

処で特定の p を撰定する場合は、例えば単峰型分布に対しては、その p が $\text{mode} = l^{(p)}$ を成立させるか否かを問うべきであり、他方標本代表値については、それが位置母数のどのような推定量であるかを明確にすべきであろう。

第3節 ノルム1の統計とピアソン第III型

前節の体系は $p=2$ の場合、全く従来の確率論の体系に合致することは明らかである。

それ以外の norm に関しては、歴史的にも又現在に於ても $p=1$ の場合に殆んど関心が絞られていると言っても過言ではないであろう。

仮説 1, 2 の下で、此の場合 $l^{(p)}(X)$, $D^{(p)}(X)$, $\rho_{xy}^{(p)}$ 及び normal equation (35), (36) は、どのように表現されるであろうか。

定理 5 『仮説 1, 2 の下で次の関係が成立する。

$$(57) \quad l^{(1)}(X) = \text{med. } x \text{ (median)}$$

$$(58) \quad D^{(1)}(X) = E|x - \text{med. } x| = \delta_{x;\text{med.}} \text{ (メディアン周りの平均偏差)}$$

又此の場合共散布度を共平均偏差 $\text{codev}(X, Y; \text{med.})$

とすれば

$$(59) \quad |\text{codev}(X, Y; \text{med})| = |E\{\text{sgn}(X, \text{med. } x)\}(Y - \text{med. } y)| \\ \times |E\{\text{sgn}(Y, \text{med. } y)\}(X - \text{med. } x)|$$

及び

$$(60) \quad \text{sign fo codev}(X, Y; \text{med.}) \\ = \text{sign of } [E\{\text{sgn}(X, \text{med. } x)\}(Y - \text{med. } y) \\ + E\{\text{sgn}(Y, \text{med. } y)\}(X - \text{med. } x)]$$

であり、相関係数 $\rho_{xy}^{(1)}$ は

$$(61) \quad \rho_{xy}^{(1)} = \frac{\text{codev}(X, Y; \text{med})}{\delta_{x;\text{med}} \delta_{y;\text{med}}}$$

と表わされる.

更に normal 方程式は、此の場合

$$(62) \quad P(Y > \alpha X + \beta) = P(Y \leq \alpha X + \beta) = \frac{1}{2}$$

及び

$$(63) \quad E(X|Y > \alpha X + \beta) = E(X|Y \leq \alpha X + \beta) = E(X)$$

と変形される.]

扱それでは此の場合の典型的分布は何であろうか。(41)式によればそれは

$$(64) \quad f(x) = \frac{\left\{ \frac{|x - \alpha|}{\beta} \right\}^{-\frac{1}{2}}}{4 \pi \beta} e^{-\frac{1}{2} \frac{|x - \alpha|}{\beta}}$$

$$(65) \quad \alpha = \text{med } x, \quad \beta = \delta_{x;\text{med}}$$

である.

今 (64) 式に対して $\alpha=0$ として $x>0$ の部分のみを考察することは、本質的に大きな変更とは考えられない. 此の場合の分布型は Γ 分布に属し,

$$(66) \quad f(x) = K x^{-\frac{1}{2}} e^{-2\lambda x}, \quad x > 0, K, \lambda > 0$$

である.

(66) 式に対して我々はピアソン (前掲 [22] 参照) が、甚しい 'extreme case' として Goshen の Royal Statistical Society に於ける 1887 年の Presidential Address (Journal vol. 50 Appendix II. pp.610-2) から引用した家計資産の評価額分布に対する取り扱いを想起することが出来る.

此の場合ピアソンは、彼の第 III 型に属する分布曲線

$$(67) \quad y = y_0 x^p e^{-\gamma x}$$

を想定し、当初

$$(68) \quad p = -0.65448 \quad \gamma = 0.2003$$

を算出し、特に

$$(69) \quad -1 \leq p \leq 0$$

であることに確信を得ている.

更に 'trial and error' による (68) の修正値は

$$(70) \quad y_0 = 1388.32 \\ p = -0.690077 \\ \gamma = -0.3057256$$

であった。

この例によると (66) は (67) に対する一つの典型とみることも出来るように思われる。

こうしてみると、当時於て、もしピアソンが、優生学に対する関心よりも経済学に対する関心がより強く、且 norm に対する理解があったならば、 $p=2$ に対する資産分布の、特に尾部における歪みとそれによる適合度の悪さに苦慮することなく、 $p=1$ を採択したかも知れないと推測することはそれ程大きな飛躍とも思われないのである。

ピアソンの分布型の母胎となる方程式は現在からみると極めて狭隘且つ経験的なものであるが当時もし彼が (4) 及び (5) 式に示したより高度の歪みをもつ分布に接していたならば $p=2$ よりは $p=1$ を採択する可能性は充分考えられることである。何故ならば此の場合に於てすら $\mu_0^{(1)}, \mu_1^{(1)}, \mu_2^{(1)}$ が存在するからである。或は Zipf 分布の場合に於てすら $\mu_0^{(1)}, \mu_1^{(1)}$ は存在するからである。

然しそれにしても歪みをもつ分布に対して代表値の存在を第一義的に想定することは、特殊な変換、特殊な類型化によって成功することはあっても、一般的には猶多くの問題を残すものといわねばならない。

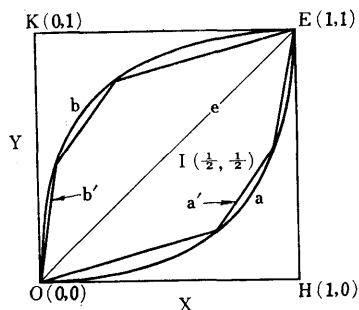
第2章 特殊な分配構造の解析法について

第1節 特殊な分配構造の表現法とその概要

経済統計に於ては、その多くの統計表は、必ずしも規模分布を示すものではなく、寧ろ各規模階層別に、客体数と金額計との対応関係によって分配構造を表わしている。それは一つには、集計、製表上の制約や慣習によるものでそれ自体問題を含むものであるが、他面競争を伴なうシステムの統計として分布に示されない現実性を含んでいる事も事実である。すなわち競争の条件は、一般に類似した階層内にあっては、その成員数と、その階層の占有量の大小に依存するからである。本章に於て扱う階層は専ら測定値 x による規模階層であるが、その分析手法は更に次章に於て一般階層の場合に拡張される。

扱、上記の客体数及び金額計をそれぞれ相対化 (% 又は単なる構成比率) すると前章の記法に対応して、第 i 階層に対し f_i 及び $u_i = x_i f_i$; $i = 1, 2, \dots, n$ と表わされるが、この場合 f_i に対する u_i の関数関係がシェア分布を与える。

更に f_i 及び u_i の順次の累積額 $X_k = \sum_{i=1}^k f_i$ 及び $Y_k = \sum_{i=1}^k u_i$; $k = 1, 2, \dots, n$ を X, Y 両軸に関連させて幾何学的に捉えたものガローレンツ曲線 (集中曲線) である。この累積の順序は、 x の低額階級から行った図の a 曲線 (実際のデータからは折線 a' が得られる) と、 x の高額階級から行った図の b 曲線とが代表的なものであるが、a, b 両曲線は



第 1 図

常に原点 $O(0, 0)$ を始点とし、点 $E(1, 1)$ を終点とするもので、もし x_i が常に正ならば、正方形 OHEK 内に存在する曲線である。又 a 曲線は X 軸に対して凸であり、b 曲線は Y 軸に対して凸である。実は、a, b 両曲線は点 $I(\frac{1}{2}, \frac{1}{2})$ に関して点対称の関係にあるので a 曲線の性質を調べれば、b 曲線の性質も判明するのである。つまり今 a 曲線を

$$(71) \quad Y = A(X)$$

で表わすと b 曲線は

$$(72) \quad 1 - Y = A(1 - X)$$

を満足する曲線となるからである。従って以下に於ては数学的に扱い易い a 曲線を主な対象としてローレンツ曲線の特性を示すことにする。然し以上は一般論であって、実務上の理由から屢々 b 曲線が用いられるし、又理論上からも例えば (4) のパレート分布は b 曲線で扱った方が数学的に扱い易いのである。因みに (4) 式に対する a 曲線及び b 曲線はそれぞれ

$$(73) \quad 1 - Y = (1 - X)^{\frac{\alpha - 1}{\alpha}} \quad (\text{a 曲線})$$

及び

$$(74) \quad Y = X^{\frac{\alpha - 1}{\alpha}} \quad (\text{b 曲線})$$

となるからである。

従って此の両式によって α を異にするパレート曲線は $X=0$ 及び 1 以外では決して交わらないことが分る。

扱、Kendall, M. G. [16] によると、ローレンツ曲線は単に規模分布を扱う上での特殊なグラフ的方法に過ぎないことになるが、それは前章の確率論的観点に偏った聯か皮相な観察と思われる。つまり以下個々に説明するように、この曲線の特性は、ベクトル解析法の適用によって、明快に表現されるからである。

他方、米沢氏 [45] は、此の曲線に深い関心と可成りの理解を持ち乍ら、適当な解析法を見出し得ない為に不十分な結論に終始している観がある。例えば此の方法が比率分析に通ずるものであることを人口密度の集中曲線という特殊な形で予感しておられるが、その場合の数学的制約についての理解が不足しているようである。ローレンツ曲線法の発展形態が、將にシェアの分析と比率分析を兼ねたユニークな効果的分析法を与えることは、実は後節で触れるように多次元的観点に立ってローレンツ曲面を構成した場合に初めて明確な事実となるからである。

更に此の方法は、前節に述べた確率論的立場と異なり、代表値の存在に一義的意味を認める必要がない点に於てもユニークな方法である事が理解さるべきである。勿論、理論的に密度関数 $f(x)$ のローレンツ曲線を考慮する場合加法平均 μ_1 の存在を前提とし、且 μ_1 自身ローレンツ曲線の一特性点に対応するが、データ処理に照し合わせると、この条件は代表値の存在というよりは、各階層の x に関する構成比率が有限確定であることの一表現にすぎないことが理解される。(田口 [30])

又此の曲線が X 又は Y 軸に対して凸であるという特性は、データの規模階級分割の恣意性に起因する、記述の際の客観性の喪失を最少限に留める条件を与えることになることに着目すべきである。つまり規模階級構成によって得られた集中折線 (第1図の a' 又は b') は、常に階級を構成しない場合の曲線 (a 又は b) に接し且常にその一方の側にのみ位置するからである。

この特性からみて米沢氏 ([45] p. 82) の記述はやや曖昧である。

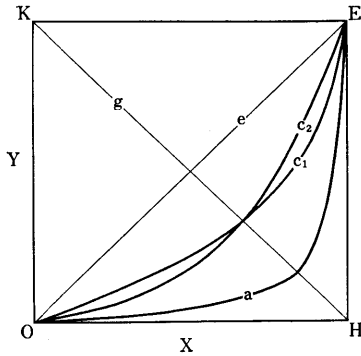
扱、本章に於て以下更に具体的に a 曲線の特性を示すことにするが、その際屢々米沢氏の著書を引用するのは、ローレンツ曲線に関する専門書が乏に事情に加えて、氏がその利用に最も積極的であり、且経験資料に照して克明に此の方法に触れ、卒直に問題を提起しておられるからである。

第2節 ローレンツ曲線の幾何学的諸性質

(1) ローレンツ曲線の対称性

ローレンツ曲線 a の対称性とは、第2図に示すように、通常均等線と称される正方形 OHEK の一つの対角線 OE と異なる他の対角線 HK に対する関係の一表現である。

一つの興味のある事実は、ジブラ所得分布 (13) のローレンツ曲線は常に OE に関して対称



第 2 図

$$(75) \quad \frac{f\left(\frac{\mu_1^2}{x}\right)}{f(x)} = \left(\frac{x}{\mu_1}\right)^3$$

が成立することが必要である。(田口 [31]) 勿論, (75) 式で μ_1 は加法平均を表わすものである。

且更に興味のあることは, この対称条件はジブラ分布のみならず, (6) 式で与えたジップ分布に於ても $k=1.5$ の場合に充足されていることである。但しこの場合 x の存在領域は

$$(76) \quad 0 < x_0 \leq x \leq x_1 \quad (x_1 \text{ は } x_0 < x_1 \text{ を満たす任意の実数を表わす})$$

である。

又, (76) の条件を満たす, つまり尾部を除いた (lower truncated) 二つのジップ分布が, それぞれ $k=1.5+h$ 及び $k=1.5-h$ (h は任意の実数) をパラメータとするとき其等のローレンツ曲線は第 2 図に示す C_1 及び C_2 曲線のように HK に対して相互に対称となることである。

且 $h>0$ とした場合 $k=1.5+h$ をもつ曲線は, 常に C_1 の如く右に歪み, 逆に $k=1.5-h$ をもつ曲線は常に C_2 のように左に歪むことが証明される。(田口 [31]) このローレンツ曲線における歪みについては後に更に詳論するが, 以上の考察を進めると, パレート分布 (4) は常に右に歪みをもつものであり, その歪みの強度は α が大である程著しいことが判明する**。このことはパレート常数 α についての理解を深めることである。

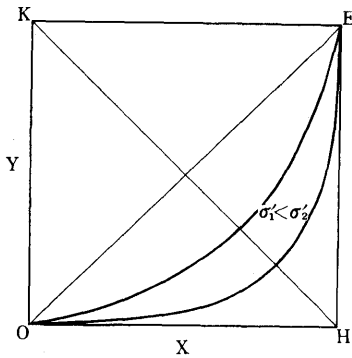
しかもパレート分布の集中曲線は下限 θ の決め方に関係なく一定の曲線形を保持するのである。

一方 (13) 式にジブラ分布は, 単に HK に対して自己対称であるばかりか, 第 3 図のように散布度を示す σ' が大である程 X 軸に対する凸の度合を増大させる。

かくして集中曲線は, 同じ所得分布の異なる代表的形態とされるジブラ分布とパレート分布をグラフ上で截然と区別するのみならず, それぞれの分布の主要パラメータについての比較判断の基準を与えているのである。その結果米沢氏の ([45] p. 85) の

* 「このように, ローレンツ曲線の曲線としての形状に注目することはきわめて意義があり, とくにその左右対称からのはずし方は, じつはログ・ノーマル分布を基準とした, それに対するはずれ方をあらわすことが特筆されてよいであろう」 ([45] p. 81)

** 曲線の歪みの意義については既に第 2 図の C_1 及び C_2 のように交錯する関係にある場合を想定して最近多くの論評がみられるが Atkinson, A. B [2] の指摘が嚆矢であるという。特にパレート曲線の分析の詳細については田口 [29] 及び Taguchi, T [31] を参照されたい。



第 3 図

『ローレンツ曲線のばあいもそうした結果になるのだが、その場合の扱い方が、ローレンツ曲線では、他の場合に比して恣意的で、与えられたデータの内容に応じて対処するといったやり方になっているのである。それはいわばゴムの物差しを使うような仕方である。一方ではこれがローレンツ曲線の一種の融通無礙な便利な性格を生み出していることは否定出来ないが、やはりそのために表示内容が限定され、強調すべき点がむしろ稀釈されてしまうといった欠陥を免れないことになるのである。この点はローレンツ曲線の利用において忘れることのできないところである。』

と述べているあたりは、ローレンツ曲線が相対的測定単位を用いることによる利点をもつという説明を除くと、専ら此の曲線の幾何学的特徴を把握する方法が不備であることを訴えておられるものと解釈される。彼って次節に於て、更にこのような不備を補う前に、既存の概念について一通りの解説を加えることにしよう。

(2) ローレンツ曲線の局所的性質

第1図について e 直線と a 曲線又は e 直線と b 曲線によって囲まれる部分の面積 A は集中面積 (concentration area) と言われている。図から明らかなようにこの面積の2倍は常に正方形 OHEK の面積 1 より小さい。従って

$$(77) \quad 0 \leq 2A \leq 1$$

が成立する。(正確に言うと、この関係は x が常に正である分布についてのみ成立するのであるが所得分布等に於ては常にこの条件が満たされているので特に断わらないことにする)。

此の面積は計算によると、

$$(78) \quad 2A = \frac{A}{2\mu_1}$$

である。 A は相対度数を用いると、

$$(79) \quad A = \sum_{i,j=1,2,\dots,n} \sum | \det \begin{pmatrix} f_i & f_j \\ u_i & u_j \end{pmatrix} | = \sum_{i,j=1,2,\dots,n} |x_i - x_j| f_i f_j$$

によって測定される。

この (79) 式はジニの平均差 (Gini, C. [9]) といわれるものであるが、一見してデータのちらばりの大小を表現していることが理解出来る。特に

$$(80) \quad A = 0$$

とは、凡ての測定値が完全に等しいことを意味している。但し第1章の (8) 及び (12) 式で表わされた散布度と本質的に異なる点は、 A は決して代表値を基準としたものではなく、個々の測定値間の比較から構成されていることである。

扱、今

$$(81) \quad G = 2A$$

で定義される G を考えると (77), (78), (79) によって、 G は一種の相対的散布度を表わすものであることが理解される。尠で此の係数が特にジニの集中係数 (或は集中度) として特に経済統計分析に於て有用視されるゆえんは、標準偏差 $\frac{\sigma}{\mu_1}$ 等と異なり、一点分布を基準にしていることと共に、將に (77) で示されるように限られた値をとるからに他ならない。

実は此の特性は、相対平均偏差 $\frac{\delta_{\mu}}{\mu_1}$ 又は $\frac{\delta_{med}}{\mu_1}$ (但し δ_{μ} 及び δ_{med} はそれぞれ平均値及びメデイアの周りの平均偏差 $\delta_{\mu} = \sum_{i=1}^n f_i |x_i - \mu_1|$ 及び $\delta_{med} = \sum_{i=1}^n f_i |x_i - med|$ を表わす。) に於ても認められるが、内容的には G の方がはるかに豊富である。

扱、此の係数 G 及び A については米沢氏の見解 ([45] p. 78 全頁) は、(77) 式の関係について稍過小評価しているように思われるが、基本的には当を得たものであろう。つまり『……ローレンツ曲線そのものについていえば、その表示内容は決して、 $2\lambda^*$ にのみ尽きるものではない。 2λ として等しい等があらわされるそれぞれの場合でも、曲線の形状は決して単純ではなく、種々な彎曲の工合を示すであらう。そしてこのような差こそ、主として原度数分布の歪度の差を反映するものとして理解することが出来るであらう。この点に想到すると、ローレンツ曲線はその図形において、 2λ よりもちがった面にむしろ着目し、たとえば曲線が図の中心にたいして対称であるか否かを測ることによって、 2λ とはべつにこれを補うような指標が与えられるのではないかという構想が生じてくる。これについてはどうしてもローレンツ曲線と原度数分布との関係を基本的に再検討することが必要になってくる』

ところでこの『構想』を実現する前に特に前記の諸所得分布について次の著しい特徴が所在することが明記されねばならない。

パレート分布 (4) の集中係数は

$$(82) \quad G = \frac{1}{2a - 1}$$

である。一方、ジブラ分布 (3) の集中係数は、 $\Phi(x)$ を規準正規分布、即

$$(83) \quad \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi$$

とした場合

$$(84) \quad G = 1 - 2\Phi\left(-\frac{\sigma'}{\sqrt{2}}\right) = 2\Phi\left(\frac{\sigma'}{\sqrt{2}}\right) - 1$$

である。[Iyengar, N. S. [13]]

パレート係数 a 又はジブラ係数 σ' は、此の関係を利用して測定することができる。

(3) ローレンツ曲線の局所的性質

以上 (1) 及び (2) においては、対称性とか集中面積とかの大域的性質を考察したが、ここでは寧ろ曲線の局所的性質を検討することにしよう。この場合に於ても正值領域で定義された連続な密度関数 $f(x)$ の集中曲線を考察する事が見通しのよい結果を与える。

この仮定に於ては、第 1 図の a 曲線に相当する集中曲線は、 x をパラメータとして

$$(85) \quad \begin{cases} X = \int_0^x f(\xi) d\xi \\ Y = \frac{1}{\mu_1} \int_0^x \xi f(\xi) d\xi \end{cases}$$

と表わされる。(例えば Kendall, M. G. [15])

従って、任意の x に対応する曲線上の一点における接線の勾配及びその変化率はそれぞれ

$$(86) \quad \frac{dY}{dX} = \frac{x}{\mu_1}$$

* ジニ係数 G を意味する。

及び

$$(87) \quad \frac{d^2 Y}{dX^2} = \frac{1}{\mu_1 f(x)} > 0$$

となる。

(86) 及び (87) の関係は、(85) と逆にローレンツ曲線 $Y=A(X)$ が与えられた時に、規模分布 $f(x)$ を導くのに直接役立つ関係である。既に本章の第1節にローレンツ曲線は常に X 軸に対し凸であるから、規模分類された客体数と x の合計額についての経験資料から理論的ローレンツ曲線を導くのが容易であると述べた。今更にそのようにして予想した理論的曲線から X を等分割して第1次、第2次の階差 ΔY , $\Delta^2 Y$ を算出し (86), (87) の関係をもとに其等に対応づけるならば、極めて容易に且正確に理論的規模分布 $f(x)$ を再現することが出来るのである。

この事のメリットを米沢氏 ([45] p.81) は、『……、われわれは、ローレンツ曲線と原度数分布との可逆的な関係を利用して、一方から他方を推量することにとりぞん想到するであろう。その場合、度数分布からローレンツ曲線を類推することは、それほど効用が大きいことが、その逆に、与えられたローレンツ曲線から原度数分布を推量することはじつに絶大の効用をもたらすであろう。というのは、いままでのところでは、度数分布を変形してローレンツ曲線を導くような局面のみを扱ってきたが、これが度数分布よりさきにローレンツ曲線が与えられるというばあいがいじさきに多くあるからである』と述べておられる。

そして経済統計表は多くは金額のシェア分布を表わしているから、此の叙述は説得力をもつであろう。

処て米沢氏はここで大きな見落しをされたように思われる。それは米沢氏はその著 [45] の71 ページ後半から72 ページにかけて、ローレンツ曲線を本稿の第1図と反対に軸を定め、本稿の記法によると、

$$(88) \quad X = A^{-1}(Y)$$

によって曲線を定義されている事である。その結果氏は

$$(89) \quad \frac{dX}{dY} = \frac{\mu_1}{x} > 0 \quad ([28] \text{ の (5.45) 式に当る})$$

$$(90) \quad \frac{d^2 X}{dY^2} = \frac{-\mu_1^2}{x^3 f(x)} < 0 \quad ([28] \text{ の (5.46) 式に当る})$$

をローレンツ曲線と密度関数との変換式であると「想到」されたのである。(89) 及び (90) 式が (86) 及び (87) 式に比べて見通しが悪いことは、一見して明らかであろう。しかもこの事が筆者にいわせると脚註に引用するような誤った結論に氏を導いたと思われるのである。聊さか長文に互るが、此のような簡単な事実は見落され勝ちであり且データ処理の上で一考に値すると思われるので、83 頁から85 頁に亘って表及び図を含めて脚註に引用させて載くことにする。(文中の番号及びかぎ括弧は筆者によるもので以上の検討により問題が含まれる箇所を示したものである。)

「さてローレンツ曲線については、その欠点として指摘されるものも少なくない。そのうちで分布の下限からの累積比率に基づいて画かれるものであるために、(1) 「分布の上限に近づくに従って、比率の感度が弱くなり、そのへんの表現力がとぼしい」という批判がある。これはたしかにそういえるところで、つぎの具体的な例について検討を加えよう。

たとえばここにつぎに示すような、異常な二極集中型分布がある。これは仮説例でなく、1939年の日本の工業統計調査による綿糸紡績業の経営規模(従業者数による)別の工場(げんざいの事業所にあたる)分布をあらわす。

表 5.10 綿糸紡績業の経営規模別構成 (1939 年)

(1) 経営規模別	(2) 工場数	(3) 従業者数	(4) (2) の%	(5) (3) の%	(6) (4) の累計	(7) (5) の累計
5- 9	444	2970	59.7	1.8	59.7	1.8
10- 14	75	887	10.1	0.5	69.8	2.3
15- 29	40	791	5.4	0.5	75.2	2.8
30- 49	8	355	1.1	0.2	76.3	3.0
50- 99	9	751	1.2	0.5	77.5	3.5
100-199	6	608	0.8	0.4	78.3	3.9
200-499	39	16,255	5.2	10.0	83.5	13.9
500-999	74	58,167	9.9	35.6	93.4	49.5
1000以上	49	82,316	6.6	50.5	100.0	100.0
計	744	163,100	100.0	100.0	—	—

商工大臣官房調査課 昭和 14 年工業統計表中巻による。

ローレンツ曲線は表 5.10 の (6) 列と (7) 列の数字, すなわち累積比率 Y, Z にもとづいて作図される。(図 5.18) この図形はたしかにやや異常な様相を呈しているが, それでもそれは集中度がかなり大であることを示しているが, (2) 「はたして原度数分布に含まれているような二極集中型という異常な性格をよく反映しているといえるであろうか. その点, どうも疑問を抱かざるをえないのが卒直な感じである。

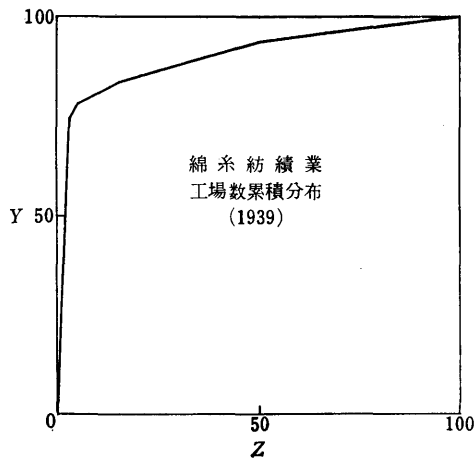


図 5.18

要するにローレンツ曲線では, このような原度数分布の特異性はひじょうに稀釈化されてあらわされることを否定できない. 現実に存在する二極集中型の生産集積の状況はこれでよく把握されるといえた義理ではない. そしてその点をいまい少し突っこんで明らかにすると, たとえば (3) 「ローレンツ曲線の上で Z を等間隔 (ここではとりあえず 9 区分) に分ち, その点の Y を挿入法によって算出し, その累積形を非累積形に民して y を求めると, つぎのような一種の度数分布となって再現される。

じつはローレンツ曲線によって表現されるのは, このような形の度数分布にほかならない. これでは, たんに小規模段階に度数 (ここでは工場数) の大部分が吸収されていることを物語るだけで, 原度数分布の特徴である二極集中のうち, 大の極への集中がいちじるしく弱められて示されるにすぎない. ローレンツ曲線が原分布の変量 x から遊離された形での表示に基づいていること, そして累積比状としてもつ特有の性格から, とくに (4) 「高い階層の部分の特性がまったくくい小化されてしまうことは否定できない. だから少なくともこの例にあげたような型の分布を表現する仕方としては, はなはだ不十分なものなのである」。

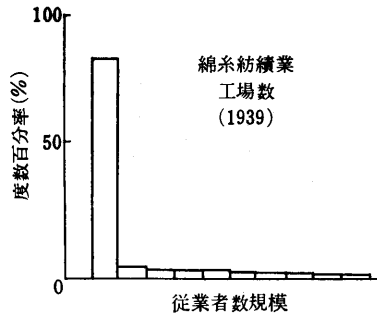


図 5.19

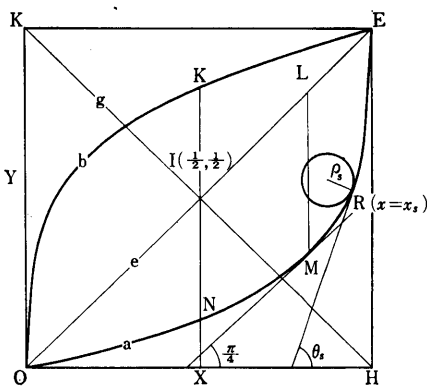
これまでに指摘した欠陥により脚註における引用の括弧内の箇所の誤謬は明らかと思われるが、特に括弧 (1) 及び (4) の箇所については軸の設定を本節のように逆転すれば結論も又逆転するのである。又括弧 (3) の箇所に見える「挿間法」については詳かでないが、もし Z の分割を Y の分割に改め本節に述べた第1次・第2次階差を用いたならば、果して図 5.19 が得られるであろうか。

最後に括弧 (2) の箇所について、図 5.18 の Z 軸と Y 軸を逆転した時、2次の導関数即接線の勾配が著しく変化する点の所在に気付かないであろうか。(29) 式によれば將にその点は密度が最小となる点に他ならない。そしてその点の Y 座標が 70%~80% 付近にあるということは正しく「二極集中型分布」の特性を端的に物語っていないであろうか。逆に又もしある一点で2次の導関数が著しく緩慢となるならばそれは最頻値(モード)の所在を示すことも忘れてはならない。

然し次節に於て此の点を更に明確に追跡することにしよう。

第3節 特殊な分配構造の記述的測度

本章の初めに於て筆者は分配構造の解析に於ては、代表値は、規模分布に対する関係程本質的ではないと述べた。然しこの事は、代表値が構造解析の上で意義をもたないということではない。例えば第4図に於て、 e, g 両対角線の交わる中点 $I(\frac{1}{2}, \frac{1}{2})$ より X 軸におろした垂線が a 曲線と交わる点 N は、メディアンに対応する曲線上の点であり、又、曲線の接線が e



第 4 図

線と平行する、即ち勾配が $\frac{\pi}{4}$ であるような接点 M は、加法平均 μ_1 に対応する点を示している。更に、二次の微係数が最大となる、即ち接線の勾配の変化の著しい点は、規模分布に於ける最頻値に対応するのである。そればかりでなく、 I から Y 軸に下した垂線が曲線と交わる点は、やはり曲線の形状を特徴づけるので、それに対応する x の値をメディアンと称する(米沢、前掲書)が、この例は集中曲線が代表値を創造する能力をもつことを示している。

一方、第4図の二曲線が囲む集中面積が集中度に等しく、相対的散布度を表わすことは既に第3節で述べた通りである。又図の ML は、

同様に相対的平均偏差 $\frac{\delta_{\mu_1}}{\mu_1}$ の $\frac{1}{2}$ に等しく、 $X = \frac{1}{2}$ が a, b 両曲線と交わる二点 N, K の長さ $\frac{\delta_{\text{med.}}}{\mu_1}$ に等しい。

又、集団を二つの階層に分割する際に階層間の平均差を最大に、逆に階層内の平均差を最少にする分割点は加法平均値である。このことは代表値に対して前章と異なる意義を与えることであり興味深い事実である。(田口 [32])

しかも、規模分布をもとにした代表値、散布度と比較して、ローレンツ曲線における其等の著しい特徴は、その明快な幾何学的特性にある。

では、規模分布における代表値・散布度以外の記述的諸測度、歪度、尖度等に対応する尺度をローレンツ曲線に於ても見出すことが出来るであろうか。

筆者の研究(田口 [31])では此の点に関して肯定的な回答を与えることが出来る。

一般に微分幾何学的曲線論に於ては、接線法線と並んで接触円、曲率の分析が大きな役割を果たしている。此の一般的立場に対して、ローレンツ曲線を除外する根拠はない。事実此の曲線の曲率の分析は、ローレンツ曲線自体の対称性からの歪みや、曲線の尖りを表わすのみならず、其等が規模分布の歪み(此の場合の分布の歪みは、分布の対称性からの歪みではなく、分散の大小或は存否にかゝる分布の尾部の収斂速度に関連した歪みを意味している)や尖度も表現しているのである。

扱、ローレンツ曲線 $Y = A(X)$ の曲率 C は一般に

$$(91) \quad C = \frac{A''(X)}{\{1 + A'(X)^2\}^{3/2}}$$

であるが、(86) 及び (87) の関係により密度関数 $f(x)$ を用いると、 x に対応する点に於ては

$$(92) \quad C = \frac{1}{\mu_1 f(x) \left\{1 + \frac{x^2}{\mu_1^2}\right\}^{3/2}}$$

が得られる。

第4図をみて分るように今この曲率の最大の点がどこに位置するかによって、 a 曲線の g 線に対称性からの歪みの度合いを判断することが出来る。この点 R の位置は種々の形式で表現されるが筆者の計算結果では、図の接線の勾配 θ_s を用いるのが適切であり、 θ_s は S に対応する x の値を飽和値 x_s と定義すると、

$$(93) \quad \tan \theta_s = \frac{x_s}{\mu_1}$$

の関係を保つことになる。筆者の見解では、ローレンツ曲線の歪みの測度 S としては、

$$(94) \quad S = \cos 2\theta_s = \frac{\mu_1^2 - x_s^2}{\mu_1^2 + x_s^2}$$

が最も適した基準を与えると思われる。

その理由として

$$(95) \quad -1 \leq S \leq 1$$

であり、 θ_s が只一つの場合 $0 \leq \theta_s \leq \frac{\pi}{2}$ をみたと θ_s が大である程、右歪みとして負の値をとり、逆に θ_s が小であれば、左歪みとして正の値をとることが図より判断されるからである。更

に, $S=0$ は $\theta_s = \frac{\pi}{4}$, 従って $x_s = \mu_1$ となって, 屢々対称曲線の一条件となる.

又, 右又は左への歪みは $x_s > \mu_1$ 又は $x_s < \mu_1$ に対応するから, 規模分布との対応関係の理解にも役立つのである.

事実, 第2節で分析した, 対称性に関して極めて鋭敏なパレート分布について S を計算すると, 最も経験分布に適合する (5) 式の α の範囲 $1 \leq \alpha \leq 2$ に於て, x_s は只一つ存在し, 且

$$(96) \quad S = \frac{1-2\alpha}{3}$$

が得られる. この結果は資料 III に掲げた殆んどが右に歪んだ曲線群の例によって明瞭に裏書きされているといえよう. 又, $x_0 \leq x \leq x_1$ の範囲で考察したジップ分布 (6) に対しては

$$(97) \quad S = 1 - \frac{2}{3}k$$

が成立する結果, 第2節で述べた $k=1.5+h$ と $k=1.5-h$ との相互対称性が, S にそのまま反映されていることが明らかになるのである.

処で (5) の条件を考慮せずに (96) 式を機械的に拡大すると, $\alpha > 2$ の場合, S は (95) の制約をおかす結果になる. 処がこの α の範囲に於ては実は飽和値が存在しないのであり, 同時にその範囲は分散の存在範囲である. 従って密度関数が意味をもつ範囲にも一致するのである. 密度関数と, ローレンツ曲線の相反的性格が曲率を介して, この点に於て明確に示されているといえよう. 一方, ジブラ分布に於ては曲率は O, E に於て最大となるのであってその飽和値は O 以外に存在しないのである.

以上によって, ローレンツ曲線から曲率最大点を見出すことは, 所得分布の形態の予想及び, パラメータの推定に役立つばかりでなく, 解析手段の選択に関する基準を与えるのである.

扱, 再び (92) の曲率の式を考察しよう. この式の逆数 ρ は接触円の半径, 即ち曲率半径を与える. 即ち,

$$(98) \quad \rho = \mu_1 f(x) \left[1 + \frac{x^2}{\mu_1^2} \right]^{3/2}$$

である.

今, 前節末に引用した米沢氏の二極集中型分布, 即ち x の中位の規模 x_a に於て異常に密度の低い分布が与えられたとしよう. その場合のローレンツ曲線は, (98) 式によって, x_a に対応する点に至るまでは比較的緩慢な曲率の変化が続き, x_a を越えると加速的に曲率が減少する即直線状に近づく.

図 5.18 は, そのような点を将に Z の 5% 点近くで示していると言えないであろうか.

(98) 式は又, 両端を切捨てたコーシー分布のように, 尖りの著しい規模分布に対しては, 逆に直線状から曲線状に転ずる集中曲線の出現を予想させるのである. こうした意味を含めて今 C の最大値, 従って第4図の R に示される飽和値 x_s に対応する曲率 C_s をローレンツ曲線の尖りの測度 K とするならば, 極めて強力な情報をもたらすものといえるであろう. 事実パレート分布のような連続且微分可能な歪み型分布の C_s は常に有限であるのに対し, $x \rightarrow \infty$ にまでのびる裾をもつ歪みのない連続且微分可能な分布の集中曲線は $X \rightarrow 1$ のとき, 限りなく大となるのである. 因みに, 自由度2の t -分布

$$(99) \quad f(x) = \frac{A}{\left(1 + \frac{t^2}{m}\right)^{3/2}} \quad -\infty < t < \infty$$

のローレンツ曲線は (98) 式から明かに円弧を形成する。

第4節 集中解析法の理論的構成とジニ指数の意義

(1) 集中曲線法の理論的構成

モメント解析法が、いわゆる確率論に従ってフーリエ解析法と結びついた理論的展開を行うのに対し、第2章に述べた集中解析法はどのような理論的構成と展開を可能にするであろうか。

今、規模 x の周辺の微少階層 $(x, x + dx)$ に対する個体構成比率と x の構成比率によって得られるベクトル $\begin{pmatrix} X(x + dx) - X(x) \\ Y(x + dx) - Y(x) \end{pmatrix}$ をその階層の分配率ベクトルと定義する。その際特に $X'(x) = f(x)$ 及び $Y'(x) = u(x)$ が存在するならば

$$(100) \quad \mathbf{s}(x) = \begin{pmatrix} f(x) \\ u(x) \end{pmatrix} = \begin{pmatrix} 1 \\ x \\ \mu_x \end{pmatrix} f(x)$$

を限界分配率ベクトルといふことが出来る。以下 x は常に正值であると仮定し、 $\mathbf{s}(x)$ を x の分配率ベクトルと略称する。

その際、先ず二つの微少規模階級 $(x_1, x_1 + dx_1)$ 及び $(x_2, x_2 + dx_2)$ の間の相互作用素 $\mathbf{h}(x_1, x_2) dx_1 dx_2$ を $\mathbf{s}(x_1) dx_1$ と $\mathbf{s}(x_2) dx_2$ の外積によって定義することにする。此の場合 $\mathbf{h}(x_1, x_2)$ の大きさは

$$(101) \quad |\mathbf{h}(x_1, x_2) dx_1 dx_2| = |\mathbf{s}(x_1) \times \mathbf{s}(x_2)| dx_1 dx_2 \\ = \frac{|x_1 - x_2|}{\mu_x} f(x_1) f(x_2) dx_1 dx_2$$

である。すなわち、相互作用素の大きさは、両階層間の相対的支配度 (degree of dominance, [11] 参照) に他ならない。

従って又前節で述べたジニの集中度 G は、

$$(102) \quad G = \frac{1}{2} \int_0^\infty \int_0^\infty |\mathbf{h}(x_1, x_2)| dx_1 dx_2$$

で表わされる。すなわち G は相互作用素の平均的大きさ (上掲 [11] 参照) と再定義することが出来る。

更に前節に述べた記述的諸特性量は、凡てベクトルとしての $\mathbf{s}(x)$ の諸演算量として簡潔に表現されるのである。

例えば接線の勾配は $\frac{\mathbf{s}(x)}{|\mathbf{s}(x)|}$ で与えられ、曲率は、

$$(103) \quad C = \frac{|\mathbf{s}(x) \times \frac{d}{dx} \mathbf{s}(x)|}{|\mathbf{s}(x)|^3} = \frac{\frac{1}{dx} |\mathbf{h}(x, x + dx)|}{|\mathbf{s}(x)|^3}$$

である。ローレンツ曲線自体 $\mathbf{s}(x)$ の単純な積分形式 $\int_0^x \mathbf{s}(\xi) d\xi$ である事はいうまでもない。

扱、以上の諸考察からすると集中解析法は、確率空間 (Ω, B, P) の P を $\mathbf{S}(A) = \int_{x \in A} \mathbf{s}(x) dx$; $\forall A \in B$ で置で置き換えた分配率空間 (Ω, B, \mathbf{S}) を構成することによって理論的に基礎づけられよう。

事実 $\mathbf{s}(x)$ の各成分 $f(x)$, $u(x)$ はそれぞれ確率密度関数の特性を有しているから、 \mathbf{S} 自体ベクトル値確率と言える量なのである。この観点によると集中曲線は、将にベクトル値確率 \mathbf{S}

の分布関数といえるであろう。

こうした理論的観点は、第3章に於て一層明確となるが、その場合集中解析法は、特殊な分配率ベクトル或いは数学的にみて、特殊なベクトル値確率を扱う方法といえるのである。因みにベクトル値確率とは、その各成分が確率公理群をみたす（但し確率とか事象とかの歴史的概念や用語にこだわる必要はなく、其等を取除いた純粹に量的な関係を保有するという意味である。確率や確率事象概念は、それぞれ分配率及び階層或いはカテゴリー概念により置きかえればよい。）ものとして定義される。

そしてその場合 \mathbf{h} は \mathbf{s} の特性ベクトルということが出来るのである。

(2) 集中曲面の概要*

例えば所得形態をより詳細に分析しようとする場合、当然財産所得 (x) と給与所得 (y) に分割し、それぞれの分配率と共に、両者間の配分比率を吟味することになるであろう。ここにローレンツ曲線の多変量に対応する形態として集中曲面又は集中超曲面の概念が発生する。

此の場合 $\mathbf{s}(x)$ に対応して二変量の限界分配率ベクトル $\mathbf{s}(x, y)$ が考えられる。すなわち

$$(104) \quad \mathbf{s}(x, y) = \begin{pmatrix} f(x, y) \\ u(x, y) \\ v(x, y) \end{pmatrix}$$

である。ここで $u(x, y)$ は点 (x, y) における x に関する限界分配率 $\frac{x}{\mu_x} f(x, y)$ を、又

$v(x, y)$ は同じく y に関する限界分配率 $\frac{y}{\mu_y} f(x, y)$ を表わす。

以下、 x, y は共に正であるとするが、集中曲面は

$$(105) \quad F(x, y) = \int_0^x \int_0^y \mathbf{s}(\xi, \eta) d\xi d\eta$$

によって定義することが出来る。此の時 $F(x, y)$ の $X-Y$ 及び $Z-X$ 座標面への正射影の境界線はそれぞれ x 及び y についての集中曲線を与えるのである。

処でローレンツ曲線は $\int_0^x \mathbf{s}(\xi) d\xi$ 及び $\int_0^y \mathbf{s}(\xi) d\xi$ に応じて二種の曲線、すなわち第1図の a, b 両曲線を得たが、集中曲面については、四方向からの累積に応じて $\int_0^x \int_0^y \mathbf{s}(\xi, \eta) d\xi d\eta$, $\int_x^\infty \int_0^y \mathbf{s}(\xi, \eta) d\xi d\eta$, $\int_0^x \int_y^\infty \mathbf{s}(\xi, \eta) d\xi d\eta$ 及び $\int_0^x \int_y^\infty \mathbf{s}(\xi, \eta) d\xi d\eta$ が得られ、それぞれ $\Pi^{(1)}$, $\Pi^{(2)}$, $\Pi^{(3)}$, $\Pi^{(4)}$ 曲面を形成する。今、此等の組合せ曲面をもって完全集中曲面 Π を定義すれば、 Π は単位立方体の内部の閉じた曲面を形成する。

一方、 (x_1, y_1) を含む微小階層と、 (x_2, y_2) を含む微小階層との間に作用する相互作用素 $\mathbf{h}(x_1, y_1; x_2, y_2) dx_1 dy_1 dx_2 dy_2$ は

$$(106) \quad \mathbf{h}(x_1, y_1; x_2, y_2) = \mathbf{s}(x_1, y_1) \times \mathbf{s}(x_2, y_2)$$

で定義される。

この時、各 $\Pi^{(i)}$ 面に於ける有向面分 $\mathbf{h}^{(i)}(x, y)$ は、

$$(107) \quad \mathbf{h}^{(i)}(x, y) = \int_0^x \int_0^y \mathbf{h}(x, \eta; \xi, y) d\xi d\eta$$

等で表わされる。従って、今 $\mathbf{h}^{(i)}(x, y)$ をその各成分を用いて

$$(108) \quad \mathbf{h}^{(i)}(x, y) = \{h_x^{(i)}(x, y), h_y^{(i)}(x, y), h_z^{(i)}(x, y)\}$$

* 本節の全体に亘って拙稿 [34] 及び [36] を参照されたい。

によって表わすと、曲面 Π の各座標面への正射影によって得られる閉領域の面積の大きさ A_x, A_y 及び A_z は

$$(109) \quad A_x = \frac{1}{2} \int_0^\infty \int_0^\infty \sum_{i=1}^4 |h_x^{(i)}(x, y)| dx dy$$

等で示されることになる。

このうち、 A_y 及び A_z は、それぞれ x 及び y に関する集中度、すなわち G_x 及び G_y に等しいことが容易に算定される。(Taguchi, T [34] 参照)

一方 A_x は

$$(110) \quad A_x = \frac{A_x}{2 \mu_x \mu_y}$$

但し

$$(111) \quad \begin{aligned} A_x = & \int_0^\infty \int_0^\infty \left| \int_0^x \int_0^y \frac{x}{\eta} \frac{\xi}{y} |f(x, \eta) f(\xi, y) d\xi d\eta| dx dy \\ & + \int_0^\infty \int_0^\infty \left| \int_x^\infty \int_0^y \frac{x}{\eta} \frac{\xi}{y} |f(x, \eta) f(\xi, y) d\xi d\eta| dx dy \\ & + \int_0^\infty \int_0^\infty \left| \int_0^x \int_y^\infty \frac{x}{\eta} \frac{\xi}{y} |f(x, \eta) f(\xi, y) d\xi d\eta| dx dy \\ & + \int_0^\infty \int_0^\infty \left| \int_0^x \int_y^\infty \frac{x}{\eta} \frac{\xi}{y} |f(x, \eta) f(\xi, h) d\xi d\eta| dx dy \end{aligned}$$

で表わされる結果、 x, y 間の比率に関する集中度を示すものということが出来る。つまり A_x は比率が一定ならば 0 となり、不安定となる程増大する傾向をもつ。 A_x は又一般に A_y, A_z と共に

$$(112) \quad 0 \leq A_x \leq 1$$

をみたすものである。

以上により、今 π の囲む体積 U を集中体積と定義すると、それは、 x, y のそれぞれに関する集中度及び x, y 間の比率に関する集中度を含めた総合集中度であるといえる。(拙稿 [39] 参照) 集中体積は、計算の結果

$$(113) \quad U = \frac{A_{xy}^{(1)}}{4 \mu_x \mu_y}$$

但し

$$(114) \quad \begin{aligned} A_{xy}^{(1)} = & \int_0^\infty \dots \int_0^\infty \operatorname{sgn}(x_1, x_3) \operatorname{sgn}(y_2, y_3) h(x_1, y_1; x_2, y_2) s(x_3, y_3) \prod_{i=1}^3 dx_i dy_i \\ = & \int_0^\infty \dots \int_0^\infty \operatorname{sgn}(x_1, x_3) \operatorname{sgn}(y_2, y_3) \left| \frac{x_1 - x_3}{y_1 - y_3} \frac{x_2 - y_3}{y_2 - y_3} \right| \prod_{i=1}^3 f(x_i, y_i) dx_i dy_i \end{aligned}$$

となる。(Taguchi, T 同上)

此の U も又

$$(115) \quad 0 \leq U \leq 1$$

の性質をもつから、我々は第 1 種二次元集中度 $G_{xy}^{(1)}$ を

$$(116) \quad G_{xy}^{(1)} = \sqrt{U}$$

によって、又第 1 種二次元平均差を (114) 式の $A_{xy}^{(1)}$ によって定義することが出来る。

体積 U と共に、 π の表面積 H を集中表面積と定義すると、 H も又 U と同じ意味で総合集中度を構成することが出来る。すなわち

$$(117) \quad 0 \leq H \leq 6$$

であるから

$$(118) \quad G_{xy}^{(s)} = \frac{H}{3!} = \frac{\Delta_{xy}^{(s)}}{3! \mu_x \mu_y}$$

をもって第2種集中度を定義することが出来る。其の時

$$(119) \quad \Delta_{xy}^{(s)} = \mu_x \mu_y \int_0^\infty \int_0^\infty \sum_{i=1}^4 |h^{(i)}(x, y)| dx dy$$

を第2種平均差ということが出来る。

実際に (107) 式の関係によって明らかなようにこの表現の方が一変量の集中度の定義に近いものであり、且 $G_{xy}^{(s)} = 0$ は完全に一点分布の場合に限られるから、この意味では $G_{xy}^{(1)}$ より有効な指標である。 $G_{xy}^{(1)}$ は $G_{xy}^{(1)} = 0$ の状態を考えると明らかなように、相関係数としての機能を強くもっているのである。

実際上の問題として、具体的に今企業を対象として、生産量 (x) と売上高 (y) に関する二次元集中曲面を構成したものとする。この時、複雑な諸要因の影響を無視すると、各正射影面積 A_x, A_y, A_z は、それぞれ価格、売上高及び生産量に関する集中度、すなわち個別独占度を示すことになる、従って $G_{xy}^{(1)}, G_{xy}^{(s)}$ は総合独占度の一種を与える結果になるが、このような総合指標形式は、経験的観点からも無理のないように思われる。(独占及び集中度については、宮崎 [21] 等を参考にした。)

二次元集中曲面は、上述の比率集中度 A_x 、二次元集中度 $G_{xy}^{(1)}, G_{xy}^{(s)}$ 、二次元平均差 $\Delta_{xy}^{(1)}, \Delta_{xy}^{(s)}$ 等一次元集中曲線で得られない諸指標をもたらしたが、更に二次元第3種平均差 $\Delta_{xy}^{(s)}$ やそれにもとづく xy に関する相関係数 $\rho_{xy; \Delta}$ を与えることが出来る。例えば二次元第3種平均差 $\Delta_{xy}^{(s)}$ は、(114) 式において $\Delta_{xy}^{(1)}$ から比率格差を控除した、 x, y に関する正味の総合散布度として、 $H_Y^{(i)}$ と $H_Z^{(i)}$ の性質のみを利用して得られるものである。(Taguchi, T 上掲論文 [34] の II 及び III の $\Delta_{xy}^{(s)}$ 参照)

その結果は

$$(120) \quad \Delta_{xy}^{(s)} = \begin{vmatrix} \Delta x & |\text{codif}^{(s)}(y, x)| \\ |\text{codif}^{(s)}(x, y)| & \Delta y \end{vmatrix}$$

但し

$$(121) \quad \text{codif}^{(s)}(x, y) = \int_0^\infty \cdots \int_0^\infty \text{sgn}(y_1, y_2) (x_1 - x_2) \prod_{i=1}^2 f(x_i, y_i) dx_i dy_i$$

$$(122) \quad \text{codif}^{(s)}(y, x) = \int_0^\infty \cdots \int_0^\infty \text{sgn}(x_1, x_2) (y_1 - y_2) \prod_{i=1}^2 f(x_i, y_i) dx_i dy_i$$

で表わされる。

又、平均差を用いた曲線相関係数の一つは

$$(123) \quad \rho_{xy; \Delta} = \pm \sqrt{\frac{|\text{codif}^{(s)}(x, y) \text{codif}^{(s)}(y, x)|}{\Delta_x \Delta_y}}$$

で示される。但し符号は

$$(124) \quad \Theta_{xy; \Delta} = \sum_{i=1}^4 H_Y^{(i)} H_Z^{(i)} = \frac{1}{2} \{ \Delta_y \text{codif}^{(s)}(x, y) + \Delta_x \text{codif}^{(s)}(y, x) \}$$

の符号によって決定される。

つまり $\rho_{xy;\Delta}$ は、常に

$$(125) \quad -1 \leq \rho_{xy;\Delta} \leq 1$$

となり、 X, Y が独立の場合に 0 を示す他、 Y が X の単調増大或は減少関数の場合にそれぞれ +1 及び -1 の値をとるからである。

以上の $\Delta_{xy}^{(s)}$ 及び $\rho_{xy;\Delta}$ は単にモメント解析法に対応する概念に止まらず、独自の本来的意義を有している。すなわち相互作用素 $h(x_1, y_1; x_2, y_2)$ が任意の二点 (x_1, y_1) (x_2, y_2) に関して常に一定方向を有する時、 $\rho_{xy;\Delta} = \pm 1$ 、つまり「完全な相関々係」が成立するのであり、多方向性を示す程 0、つまり「独立な状態」に近づくからである。

扱、集中曲面概念は、勿論単に独占度の表現のみに止まらない。例えば人口の地理的集中に関する諸指標を構成する場合にも有効である。此の場合、例えば地方自治体を単位とし、人口 (x) の所有地積 (y) に関する集中曲面を基礎にすると、上述の A_x, A_y, A_z はそれぞれ人口密度、人口及び所有地積の集中度を示すことになり、 G_{xy} は其等の総合指数となるからである。

然し以上の指数と別に、地理的条件に関しては、既に非類似係数、特化係数等の経験諸指標が存在する。此等の係数は、一次元の集中曲線及び二次元の集中曲面の幾何学的性質と直ちに結びつくものではない。其等とは、次項に述べる s の構成する場に於て規定され、 h の構成する場に於て発展的に再規定される諸量である。

然しながら、 s 及び h の与える場の考察に移る前に、以上の問題に関して再び米沢氏の前掲書を取上げる事は、問題を一層具体化する意味で適切な事と思われる。

氏は、ローレンツ曲線の応用として将に上述の人口の地理的集中に関する諸係数（立地係数、地理的結合係数、非類似係数、局地化係数）を考察しておられる。この場合、氏及び氏の引用された安田氏（[44] 参照）の発想は基本的に一次元的な観点に立つものである。つまり上例に於て x の累積と y の累積を両軸にすると、ローレンツ曲線に似た曲線が得られるから、その幾何学的性質から総合指標を構成し或いは総合指標との対応関係を見出そうと意図されているからである。

この場合の問題は、 x 及び y の累積順位を何によって行うかであるが、一般に x 又は y 又はその比 $\frac{x}{y}$ の大小よるか或は第三の量（例えば中心部からの距離）の大小を用いることが行われており、事実氏等は x, y 及び $\frac{x}{y}$ による場合を中心に考察されておられる。

もし累積の順位を x 又は y の大小の順によるとすると、これは将に Blitz, R. C 及び Brittain J. A. の相関曲線（田口 [32] 系列曲線参照）そのものであり、それは又集中曲面に於る $H_z^{(1)}$ 又は $H_y^{(1)}$ の一方の境界線を与えるものである。従ってこの曲線から相関を得ることは既述のように可能であり、事実 Blitz 及び Brittain はその簡単な定義を与えているが、氏等の主な関心は局地化係数

$$(126) \quad L = \frac{\sum_i \left| \frac{x_i}{\sum_k x_k} - \frac{y_i}{\sum_k y_k} \right|}{2}$$

にある。（前掲 [45] pp 90~93）

そこで問題になるのは、 $\frac{x}{y}$ の大小順による y, x の累積曲線であるが、それは氏の指摘されているように、将に $\frac{x}{y}$ の集中曲線そのものに一致するのである。ところで米沢氏は、そのこと

から本稿第2章第2節(2)で述べた $\frac{\delta_{\mu_1}}{\mu_1}$ の幾何学的性質を想起されて、(126)の L は、原点を通る対角線に平行な接線をもつ曲線上の点を通る縦軸に平行な直線が、曲線及び対角線と交わることによって形成する線分に等しいと判断されたのである。これはいかにも紛らわしい関係によって生じた錯覚であって、この線分は、実は

$$(127) \quad L' = \frac{1}{2} \sum_i \left| \frac{x_i}{y_i} / \sum_k \frac{x_k}{y_k} - 1 \right|$$

を示すものであり、(126)式と(127)式は一般には一致しない。

集中曲線、曲面は直接 L に対応する幾何学的特性量を所有しないようである。その事情は分布曲線がモメントを幾何学的特性量として直接具現するものではない事と同一であり、それにも拘らず、集団の直観的具体的表現形式として分布曲線も集中曲線も大きな意味をもっているのである。

同時に、直観的具体的形式の制約として、分布関数や、集中曲線はどうしても累積の順序による規制を受けるのである。特に集中曲線、曲面は、パラメーターとして規模 x 或は x 及び y を必要とするのである。こうした制約による分配率についての不完全な情報を補う為には、どうしても以下のように、直接ベクトルの場を考察する必要が生ずるのである。

(3) 分配率ベクトル s の場の性質

(A) ノルム1の統計との関係

今、(100)式で示された限界分配率ベクトル s の直交座標系 Σ を

$$(128) \quad \Sigma = (0; i, j)$$

としよう。

此の時 s は

$$(129) \quad s = s_u i + s_v j = f(x) \left(i + \frac{x}{\mu_x} j \right)$$

で表わすことが出来る。

扱 s の $i-j$ 方向成分の絶対値の平均を $\overline{|s_{i-j}|}$ とすると

$$(130) \quad \overline{|s_{i-j}|} = \frac{1}{\sqrt{2}} \int_0^{\infty} |s_u - s_v| dx = \frac{1}{\sqrt{2} \mu_x} \int_0^{\infty} |\mu_x - x| f(x) dx = \frac{\delta_x}{\sqrt{2} \mu_x}$$

である。従って $\sqrt{2} \overline{|s_{i-j}|}$ は、第2章第2節(2)で述べた相対的平均偏差を表わすものといえる。

更に θ を $0 \leq \theta \leq \frac{\pi}{2}$ をみたく自由変数として、 s の $\sin \theta i - \cos \theta j$ 方向成分の絶対値の平均を $\overline{|s_\theta|}$ を考えると、

$$(131) \quad |s_\theta| = \int_0^{\infty} |s_u \sin \theta - s_v \cos \theta| dx = \frac{1}{\mu_x} \int_0^{\infty} |\mu_x \sin \theta - x \cos \theta| f(x) dx$$

となる。従って $\frac{\overline{|s_\theta|}}{\cos \theta}$ を極小化する $\alpha = \tan \theta_0$ と、それに対応する $\frac{\overline{|s_{\theta_0}|}}{\cos \theta_0}$ を求める問題は、

第1章で述べた Fechner の累乗中位数とその偏差に関する $p=1$ の場合に他ならない。故に Fechner の課題は s の代表的方向 $\cos \theta i + \sin \theta j$ を決定することでもある。

より興味のある問題は (104) 式で与えられた二変量に関する分配率ベクトル s の構成する場について生ずる。今此の場合の直交座標系 Σ を

$$(132) \quad \Sigma = (0; i, j, k)$$

とする。その時 \mathbf{s} は

$$(133) \quad \mathbf{s} = s_u \mathbf{i} + s_v \mathbf{j} + s_w \mathbf{k} = f(x, y) \left(\mathbf{i} + \frac{x}{\mu_x} \mathbf{j} + \frac{y}{\mu_y} \mathbf{k} \right)$$

で表わされる。此の場合、自由変数 $\varphi, \theta; 0 \leq \varphi, \theta \leq \frac{\pi}{2}$ を導入して \mathbf{s} の $\cos \varphi \mathbf{k} - \sin \varphi$

$(\cos \theta \mathbf{j} + \sin \theta \mathbf{i})$ 方向成分の絶対値に関する平均値を $\overline{|s_{\theta, \varphi}|}$ とすれば、 $\frac{\overline{|s_{\theta, \varphi}|}}{\cos \theta \cos \varphi}$ の極少

を与える $\alpha = \tan \varphi_0, \beta = \tan \varphi_0 \tan \theta_0$ を求める問題は、前章に述べた $p=1$ の回帰係数 $a^{(1)}, b^{(1)}$ の決定の問題に帰着する。従って、此の限りに於て、norm 1 の回帰係数を決定することは \mathbf{s} の代表的方向を $\sin \varphi \mathbf{k} + \cos \varphi (\sin \theta \mathbf{j} - \cos \theta \mathbf{i})$ の形で決定する一方法ということが出来る。

結局 $p=1$ の場合の統計は、 \mathbf{s} の構成する場の特性を記述するものであると言うことも出来る。そして $p=1$ 以外の一般の norm p の統計は \mathbf{s} の特性を間接的に記述しているのである。

(B) 局地化係数の意義

(133) 式の \mathbf{s} について、その $\mathbf{j}-\mathbf{k}$ 方向成分の絶対値に関する平均値を $\overline{|s_{j-k}|}$ とすると

$$(134) \quad \overline{|s_{j-k}|} = \frac{1}{\sqrt{2}} \int_0^\infty \int_0^\infty |s_v - s_w| dx dy = \frac{1}{\sqrt{2}} \int_0^\infty \int_0^\infty \left| \frac{x}{\mu_x} - \frac{y}{\mu_y} \right| f(x, y) dx dy$$

となる。今 n 個の観測値ベクトル

$$(135) \quad \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \dots, \begin{pmatrix} x_n \\ y_n \end{pmatrix}; x_i, y_i \geq 0$$

を対象にして、(131) に該当する関係を求めると

$$(136) \quad |s_{j-k}| = \frac{1}{\sqrt{2}} \left| \frac{\sum_{i=1}^n x_i}{\sum_{k=1}^n x_k} - \frac{\sum_{i=1}^n y_i}{\sum_{k=1}^n y_k} \right|$$

従って (126) 式と比較すると

$$(131) \quad L = \frac{\overline{|s_{j-k}|}}{\sqrt{2}}$$

が成立する。つまり局地化係数は、一変量の場合の相対的平均偏差的な役割を二変量の比に関して果すものである。

(C) \mathbf{s} のその他の特性量とその意義

(133) 式で与えられた \mathbf{s} について、その $\sin \theta \mathbf{j} - \cos \theta \mathbf{k}$ 方向成分の絶対値を $\overline{|s_{i, \theta}|}$ とすると

$$(138) \quad \begin{aligned} |s_{i, \theta}| &= \int_0^\infty \int_0^\infty |s_v \sin \theta - s_w \cos \theta| dx dy \\ &= \int_0^\infty \int_0^\infty \left| \frac{x}{\mu_x} \sin \theta - \frac{y}{\mu_y} \cos \theta \right| f(x, y) dx dy \end{aligned}$$

である。

今、 θ を $0 \leq \theta \leq \frac{\pi}{2}$ をみたす自由変数として、その $\frac{\overline{|s_{i, \theta}|}}{\cos \theta}$ の極少 θ_0 及び極少値を考える

と、容易に

$$(139) \int_0^{\infty} \operatorname{sgn}\left(\frac{y}{\mu_y} / \frac{x}{\mu_x}, \tan \theta_0\right) \frac{y}{\mu_y} f(x, y) dx dy = 0$$

となる。 $\alpha = \tan \theta_0$ 及び $\frac{|s_{i,\theta}|}{\cos \theta_0}$ は、それぞれ $\frac{\mu_x y}{\mu_y x}$ の比率の化表値及び比率の散布度ということができる。その役割は、一変量の場合の中位数及び平均偏差のそれに該当する。何故ならば、 α を誘導する条件は、 $\alpha' = \frac{\mu_y}{\mu_x} \alpha$ とおけば

$$(140) E|Y - \alpha'X| = E\left|X \left|\frac{Y}{X} - \alpha'\right|\right|$$

を最少にする条件だからである。

(4) 相互作用素 h の場の性質

S の構成する場における諸関係に対して、 h の構成する場における諸関係は、より本質的なものである。以下の諸考察によって示されるように、少なくとも Gini の思考に従うとそのように言えるであろう。事実 h の具有する関係が与えられると、それに応じて密度関数 $f(x)$, $f(x, y)$ 等が決定され、従って更に s が決定されることは拙稿の示す通りである。(一変量の場合については [31] pp. 116-117 Theorem 4 及び Theorem 5 参照。又二変量の場合については [34] No. 1 pp. 377-378 Lemma 3 及び Proposition 14 参照)

(A) 修正ジニ係数の定義とその関係

第4節(1)に於て述べたように、一変量の場合、 h の大きさの平均は(102)式のように $2G_x$ であった。

扱、二変量に関する(106)式の h を前項のように Σ 系によって表わすと

$$(141) \quad h = h_u i + h_v j + h_w k \\ = f(x, y) \left(\frac{1}{\mu_x \mu_y} \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} i + \frac{1}{\mu_y} \begin{vmatrix} y_1 & y_2 \\ 1 & 1 \end{vmatrix} j + \begin{vmatrix} 1 & 1 \\ x_1 & x_2 \end{vmatrix} k \right)$$

が得られる。

従って、 $|\overline{h_u}|$, $|\overline{h_v}|$, $|\overline{h_w}|$ をそれぞれ h_u , h_v , h_w の絶対値についての平均値を表わすものとすれば、其等は又、 x, y 間の比率の相対的偏差、 y に関する相対的偏差、及び x に関するそれを表わしている。特に

$$(142) \quad \frac{|\overline{h_v}|}{2} = G_y, \quad \frac{|\overline{h_w}|}{2} = G_x$$

が成立するから

$$(143) \quad G_{x|y} = \frac{|\overline{h_u}|}{2} = \frac{\Delta x | y}{2 \mu_x \mu_y}$$

によって与えられる $G_{x|y}$ を(2)における A_x に対して x, y 間の比率に関する修正集中度と定義することが出来るよう。

但し

$$(144) \quad \Delta_{x|y} = \int_0^{\infty} \dots \int_0^{\infty} |\det \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}| \prod_{i=1,2} f(x_i, y_i) dx_i dy_i$$

であり、この $\Delta_{x|y}$ を(2)における Δ_x に対して x, y の比率に関する修正平均差と定義しよ

う。この時明らかに

$$(145) \quad \Delta_{x_1y} \geq \Delta_x$$

であるが、従って Δ_x の場合のように、 x, y の正值分布に対して一般的に

$$(145) \quad 0 \leq G_{x_1y} \leq 1$$

が成立するのである。

他方 $h(x_1, y_1; x_2, y_2)$ と (133) に示した $s(x_3, y_3)$ との内積の絶対値の平均を $\frac{\Delta'_{xy}}{\mu_x \mu_y}$ とすると、

$$(146) \quad \begin{aligned} \Delta'_{xy} &= \mu_x \mu_y \int_0^\infty \dots \int_0^\infty |h s| \prod_{i=1,2,3} dx_i dy_i \\ &= \int_0^\infty \dots \int_0^\infty \left| \det \begin{pmatrix} x_1 - x_3 & x_2 - x_3 \\ y_1 - y_3 & y_2 - y_3 \end{pmatrix} \right| \prod_{i=1,2,3} f(x_i, y_i) dx_i dy_i \end{aligned}$$

となり、 Δ'_{xy} は (2) に示した Δ_{xy} と同様に xy の総合的偏差を表現するから、これを二次元修正第1種平均差と言うことが出来る。同様に

$$(147) \quad G_{xy}^{(1)'} = \sqrt{\frac{\Delta_{xy}^{(1)'}}{3! \mu_x \mu_y}}$$

を修正第1種集中度とすることが出来る。その際、 x, y の正值分布に対して一般的に

$$(148) \quad 0 \leq G_{xy}^{(1)'} \leq 1$$

が成立する。

更に、 H に対応して、修正第2種平均差 $\Delta_{xy}^{(2)'}$ 及び修正第2種集中度 $G_{xy}^{(2)'}$ は、それぞれ

$$(149) \quad \Delta_{xy}^{(2)'} = \mu_x \mu_y \int_0^\infty \dots \int_0^\infty |h| \prod_{i=1,2} dx_i dy_i$$

$$(150) \quad G_{xy}^{(2)'} = \frac{\Delta_{xy}^{(2)'}}{3! \mu_x \mu_y}$$

とすることが出来る。 $\Delta_{xy}^{(2)'}$ は

$$(151) \quad 0 \leq \Delta_{xy}^{(2)'} \leq \mu_y \Delta_x + \mu_x \Delta_y + \Delta_{x_1y}$$

であり、従って x, y の正值分布に対して一般に

$$(152) \quad 0 \leq G_{xy}^{(2)'} \leq 1$$

が成立する。

以上の修正諸係数をベクトルの本来の性質から考えると、 G_x, G_y, G_{x_1y} はそれぞれ h の軸方向成分の格差に関する総合指標であり、 $G_{xy}^{(1)'}$ は h と s との間の方向のズレに関する総合指標である。又 $G_{xy}^{(2)'}$ は直接 h の大きさに関する格差の総合指標といえる。

(B) h における代表比率と回帰係数

(3) に於て s の代表的方向を考察したが、 h の構成する場に関して同様の考察を加えることが出来る。

まず、 h の $\sin \theta j + \cos \theta k$ 方向成分を $h_{i,\theta}$ とし、その絶対値に関する平均 $\overline{|h_{i,\theta}|}$ を考える。今、 θ を自由変数として $\frac{\overline{|h_{i,\theta}|}}{\cos \theta}$ を極小化する θ を求めると、それは $E|Y_1 - Y_2 - a$

$(X_1 - X_2) | ; \alpha = \frac{\mu_x}{\mu_y} \tan \theta$ を極小化する α を求める事に帰着する。

従って此の α は一種の x, y 間の比率というよりは、変分商の代表値を与えるものであり、 X, Y の定義域を微小階層に限定すれば、 y の x に関する微係数の代表値に該当する

$$(153) \quad E \left[\left\{ \operatorname{sgn} \left(\alpha, \frac{Y_1 - Y_2}{X_1 - X_2} \right) \right\} (X_1 - X_2) \right] = 0$$

の解である。

一方、単に $|\overline{h_{i,\theta}}|$ の極小条件を考えると、これは $E \left| \frac{(Y_1 - Y_2)}{\mu_y} \cos \theta' + \frac{(X_1 - X_2)}{\mu_x} \sin \theta' \right| ; \theta = -\theta'$ を極小化する θ' を求める事に帰着する。従って、此の場合は、

$$(154) \quad \alpha' = \tan \theta, \quad \beta' = \mu_y - \alpha' \mu_x$$

として

$$(155) \quad \frac{y}{\mu_y} = \alpha' \frac{x}{\mu_x} + \beta'$$

を考えると、それは $\frac{x}{\mu_x}$ と $\frac{y}{\mu_y}$ に関する一種の直交回帰直線であると言えるが、相対的観念を含むものである。

同様にして \mathbf{h} の $\cos \theta \mathbf{i} + \sin \theta \mathbf{j}$ 方向成分の絶対値に関する平均 $|\overline{h_{k,\theta}}|$ に対して $\frac{|\overline{h_{k,\theta}}|}{\cos \theta}$ の極小条件を考えてみよう。其の条件をみたす θ をとくことは $E |\det \begin{pmatrix} X_1 - \beta & X_2 - \beta \\ Y_1 & Y_2 \end{pmatrix}| ; \beta = \mu_x \tan \theta$ を極小化する β を解く事に帰着する。従って β は

$$(156) \quad E \left\{ \left(\operatorname{sgn} \frac{\begin{vmatrix} X_1 & X_2 \\ Y_1 & Y_2 \end{vmatrix}}{Y_1 - Y_2}, \beta \right) (Y_1 - Y_2) \right\} = 0$$

を満たす

$$(157) \quad y = \alpha (x - \beta) ; \alpha = \frac{\mu_y}{\mu_x - \beta}$$

を構成すると、これは一種の y の x に関する相対的回帰直線といえる。その意味で (157) 式における α 及び $-\alpha\beta$ を y の x に関する相対的回帰係数と定義する。従来回帰直線は観測された y と回帰直線によって決定される \hat{y} との差の累乗平均に関する極小条件で決定されたが相対的回帰係数は y と \hat{y} の比の加重平均差を極小化する条件により決定されるといえよう。

\mathbf{h} の $\cos \theta \mathbf{i} + \sin \theta \mathbf{k}$ への方向成分 $h_{j,\theta}$ についての同様な考察は x の y に関する回帰係数を与えることになる。

ジブラ分布に対して、直線回帰する場合は y と \hat{y} との比率の対数の極小化条件を与えた方が妥当するように、上記の相対的回帰直線の決定方式は、歪みをもつ分布に対しては寧ろ妥当な形式を与えるであろう。

何れにせよ、本節で扱った、回帰係数は \mathbf{s} 又は \mathbf{h} に関する代表的方向を決定する係数であるといえる。

扱、以上の考察を通して、Gini の与えた諸統計指標は、主に \mathbf{h} の特性或いは \mathbf{s} の外積及び三重積に関連して構成されたものと言えるであろう。Gini の統計学の特徴は norm 1 の統計と共に本質的にベクトルのであり、スカラー的な Pearson 流の統計学と質的に異なるもので

ある。

因みに前章で述べた norm 1 の統計は、本章第 4 節 (3) で述べたように、ベクトル \mathbf{s} の場の特性を表現する方法でもあった。その両義的な或は両方法の接点的な性格で \mathbf{h} の場の特性を表現するのに役立つのである。反面 norm 1 の統計はその代表値が一般の norm の代表値の特異点であったように、他の norm に対して特異な形態を示すのであるが、分配率ベクトルの記述表現に対しては、極めて自然な形態を示すことが注目される。

(5) 時系列解析についての一考察

観点を変えて今二時点 $t-1$ 及び t における状態 (例えば所得、資産或は企業利潤) x_t 及び x_{t-1} に関するシェア・ベクトル $\mathbf{s}(x_{t-1}, x_t)$ を考察しよう。その場合の完全集中曲面の与える集中体積 $U_{t-1,t}(x)$, 集中表面積 $H_{t-1,t}(x)$ 又は集中度 $G_{t-1,t}^{(1)}(x)$ $G_{t-1,t}^{(2)}(x)$ は、各時点における x の集中と共に、 x_{t-1} と x_t の比即成長率或は増加率の集中を含む総合指標を与えることになる。

一方、 $t-1$ 及び t 時点に於て $(x_{t-1}^{(1)}, x_{t-1}^{(1)} + dx_{t-1}^{(1)})$ 及び $(x_t^{(1)}, x_t^{(1)} + dx_t^{(1)})$ の区間内の状態を示す微小階層と $(x_{t-1}^{(2)}, x_{t-1}^{(2)} + dx_{t-1}^{(2)})$ 及び $(x_t^{(2)}, x_t^{(2)} + dx_t^{(2)})$ の状態にある微小階層の間に定義される相互作用素は

$$(158) \quad \mathbf{h}(x_{t-1}^{(1)}, x_t^{(1)}; x_{t-1}^{(2)}, x_t^{(2)}) dx_{t-1}^{(1)} dx_t^{(1)} dx_{t-1}^{(2)} dx_t^{(2)} \\ = \mathbf{s}(x_{t-1}^{(1)}, x_t^{(1)}) \times \mathbf{s}(x_{t-1}^{(2)}, x_t^{(2)}) dx_{t-1}^{(1)} dx_t^{(1)} dx_{t-1}^{(2)} dx_t^{(2)}$$

とすることが出来るが、此の場合の \mathbf{h} は各時点の状態と、その間の成長率によって規定される。この \mathbf{s} 及び \mathbf{h} の構成する場合は、それぞれの自己回帰直線を与えることが出来る。

勿論 x 以外の要素或は $t-2, t-3, \dots$ 等遡った時点の効果を検討する必要がある場合は、それに対応する集中超曲面を構成することにより、動的状態空間とすることが出来るのである。処で前節の思考方法によると時系列の解析に於ても \mathbf{s} 及び \mathbf{h} の各々について考える必要がある。勿論従来の方法は \mathbf{s} の段階で適用可能である。

扱、複雑な経済量の動態モデルを取上げるのに先立って、まず \mathbf{s}, \mathbf{h} の何れの段階に於ける相関も一致して $\rho=1$ を与える最も単純な例から考える。この場合は、既に述べたように $\mathbf{h}_{t-1,t}$ は定方向を保ち従って Π は平面となる。更に $\Pi^{(1)}$ は Π の縁を与える空間曲線となり、その特性は $X-Y_{t-1}$ 及び Y_t-X 面に対する正射影曲線 $Y_{t-1}=A_{t-1}(X)$ 及び $Y_t=A_t(X)$ によって決定される。且、 A_{t-1} 及び A_t はそれぞれ $t-1$ 及び t 時点における x の集中曲線に他ならないのである。

例題 1. (A) 今対象とする各人の所得額 x_{t-1} を一様に定額 a だけアップしたとする。その前後に於て、他の諸条件が変化しないものとすれば、 A_{t-1} 及び A_t はどのような関係を示すであろうか。その場合

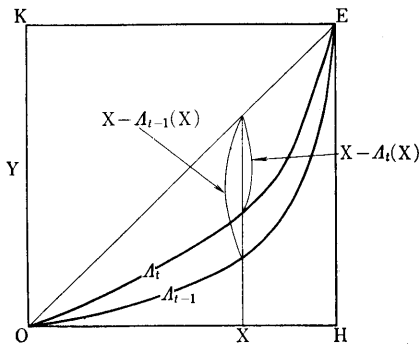
$$(159) \quad \frac{X - A_t(X)}{X - A_{t-1}(X)} = \frac{\mu_t}{\mu_{t-1}} = \frac{G_t}{G_{t-1}}$$

が成立するのである。(田口時夫 [32] 参照) 但し μ_{t-1}, G_{t-1} 及び μ_t, G_t は、それぞれ $t-1$ 時点の平均と集中度及び t 時点における其等を表わすものである。従って又両時点における平均差 Δ_{t-1} 及び Δ_t が等しいことになり

$$(160) \quad \Delta_{t-1} = \Delta_t$$

成立するのである。

A_{t-1} 及び A_t を同一平面で促えて、グラフ的に此の関係を説明すると、第 6 図のように二つの線分の比が、 X に無関係に一定であることを示すのである。



第 6 図

(B) 今所得 x を一率に $\lambda \times 100\%$ だけアップしたとする。或は賃金変化率が一様に入っていたとする。その場合は A_{t-1} と A_t は一致する。即

$$(161) \quad A_{t-1}(X) = A_t(X)$$

が成立する。此の場合には特に $F_X(x, y) = 0$ が凡ての (x, y) に対して成立する同様な関係として、物価の一率の上昇によっては売上高の集中曲線は変化しない。つまり集中曲線の変化は階層間の実質的な関係の変化によるものである。

(C) では一定累進率 p をもって、所得額 x を x^p に増大させた場合は、どのような変化を示すであろうか。この場合は、既に第2章第2, 3の各節で述べたように各曲線の最大曲率の大きさに変化をもたらす。それを一般的に表現することは難しいが、 $t-1$ 時点に於てパレート分布をしていたと仮定した場合には、次の簡単な変化が生ずる。即両時点の歪みの測度 (80) 及び (83) 式参照) を S_{t-1} 及び S_t とすれば

$$(162) \quad S_t - S_{t-1} = \frac{2}{3} (1 - p)$$

が成立するのである。

所得を増大させる、即 $p > 1$ の場合、 $S_t < S_{t-1}$ つまりカーブの歪みは右の方向に増大すると共に最大曲率即尖度 K も増大する。然し両曲線は交わらない。

勿論、一様或は一率方式を避けた階層別所得政策が行われる場合は、全体として複雑であるが、各階層に対応した曲線部分については、上記の性質が認められるのである。(田口 [31] 及び [32] 参照)

扱、以上のような一時的な政策効果の計測を目的とすることではなく、市場メカニズムを知る目的で、ある商品の売上高を異時点で行う場合を考えてみよう。この場合は一般に極めて安定した市場であっても $\rho_{i-1, i; \Delta} < 1$ であり、安定の完全に失われた極端な場合に於ては次の例題のように $\rho = 0$ と考えてよいであろう。

例題 2. X_t が前時点の状態 X_{t-1} によって何らの影響も受けない場合は、 $\Pi^{(1)}$ の $X - Y$ 面及び $Z - X$ 面への正射影の面積は、それぞれ x_{t-1} 及び x_t の集中面積に等しい。

では上記の市場構造に或程度の安定性が認められる場合はどうであろうか。この場合は一般に $0 < \rho_{i-1, i; \Delta} < 1$ である。

例題 3. 今 X_t が X_{t-1} のみと相関をもつ、すなわちマルコフ過程 (Rosenblatt, M [24] Dynkin, E. B. [6] 参照) を想定すると、もし ρ が一定、即ち定常過程ならば、 $\rho = 1$ の場合が X_t 及び X_{t-1} の集中曲線の比較の問題に還元されたのに対して、この場合は一般に (X_t, X_{t-1}) によって構成される二次元集中曲面の特徴を把握する問題に帰着する。更に $\rho_{i-1, i; \Delta}$ が t によって変化する非定常過程ならば、それは幾つかの集中曲面の比較或は $h_{i, t-1}$ の時間的変化を追求する問題へと発展するのである。(集中曲面の曲率等についての局所的性質については田口 [35] 参照)

逆に、上記の過程が更に autoregressive process や harmonic process のような一般的には非マルコフ的定常過程 (Wold, H [43], Cramer, H [5] 参照) であると仮定するならば、自己相関系列や集中超曲面 $(X_{t-r}, X_{t-r+1}, \dots, X_t)$ を考察することが必要である。

以上は、主に数理統計学の公式的時系列解析法の思考の筋書に沿った考察であるが、定常・

非定常の何れの場合についても、推移確率が、相互作用素 $h_{i,i+1}$ によって決定されるというのが、此の解析法の独自の立場である。又、事実 $h_{i,i+1}$ を与えると、推移確率が決定される。かくして推移確率を所与のものとする事なく、更に遡って、それを形成する相互作用要因を追求する事が、主要な課題となるのである。

扱、経済時系列の解析に於ては、合併、提携、分離や転業、兼業及び発生、消滅等の多様な形態が認められ、其等が屢々単純な規模変動よりも重要な指標を提供する。その際此等、諸現象に対しても、単に表面的にそれぞれの生起確率としてでなく、生産、市場の拡大要因と、その抑制要因の相互作用に於て捉える事が必要であり、其の為には、上記の諸事態の裏面の動きとして資本の増減や形態の変化を把握せねばならない。

此の場合に有効な解析方法は、次章のように分配率ベクトルの概念を更に一般化することによって得られるであろう。

第3章 一般的な分配構造の解析法について

第1節 一般的な分配構造の具体的表現とその概要

前章で扱った解析法は、集中曲線、曲面解析法を初めとして、 x の規模階層への客体と x 量の分配を扱うものであった。

その意味で、此の解析法は、極めて制限された特殊階層に関する分配構造に関する方法といえる。

然し、一般的に分配構造を扱おうとする場合は、かかる特殊な階層或は分配量から脱却せねばならない。事実、国富、国民所得の分配を論ずる場合、部門別要素別の構成比率を扱わねばならない。又問題を法人部門に限定した場合に於ても、産業別諸構成、地域別諸構成が分析上重要となる。勿論、部門間、業種間或いは企業系列間等の分配と並んで、例えば法人統計のように、資本金額等の規模階層間の分配又賃金統計のように年令別賃金構成、更に国富統計のように取得時期別資産構成等を分析する必要があるが、此の場合の規模の規定量は前章の場合と異なり、一般的には、分配比率規定量 x と異なる種類に属するとみなければならぬ。(参考資料 IV 及び V 参照)

然し、こうした一般的な分配構造を扱う方法論的な準備は、既に前章に述べた分配率ベクトルの概念と、その場の分析法に於て行われているのである。つまり一旦確立された s 及び h 等のベクトル概念は、特定の規定量 x とは無関係な性質をも示すからである。

今、一般的階層基準を α とする。勿論 α は数量基準に限定される必要はない。

この場合、対象とする集団は階層基準 α によって C_1, C_2, \dots, C_n の各クラスに分類されるものとしよう。集団全体 Q 及び各クラス C_i は各個体を要素とする数学的集合と考えると、一般に

$$(163) \quad \bigcup_{i=1}^n C_i = Q$$

及び

$$(164) \quad C_i \cap C_j = \phi$$

をみたすものと規定出来る。

つまり、集団を形成する各個体は何れかのクラスに分類されるが、同時に他のクラスに分類されることはないものとする。

この c_i 各クラスに対しては、それに所属する個体数の構成比率 u_i と、 x の総量の構成比率 w_i が対応する。即、ベクトル値集合関数

$$(165) \quad T_i = \begin{pmatrix} v_i \\ w_i \end{pmatrix} \quad i = 1, 2, \dots, n$$

が C_i に対し定義されるのである。この T_i を階層 C_i の分配率ベクトルと定義しよう。もし α が x そのものならば $T_i = S_i$ 即前章の場合に帰着する。

扱、この場合に於ても前章と同様に C_i, C_j 間に作用する支配力 (相互作用力) $G_{i,j}$ を

$$(166) \quad G_{i,j} = T_i \times T_j$$

によって定義することが出来る。これにより又必然的に一般化 β = 係数及び一般化平均差を

$$(167) \quad G_{x;c} = \frac{1}{2} \sum |G_{ij}|$$

及び

$$(168) \quad A_{x;c} = 2 \mu_x G_{x;c} = \sum_{i,j=1,2,\dots,n} \left| \det \begin{pmatrix} v_i & v_j \\ w_i & w_j \end{pmatrix} \right|$$

によって定義することが出来る。

この時

$$(169) \quad A_{x;c} \leq A_x$$

が一般に成立する関係である (拙稿 [31] 参照)。実は $A_{x;\alpha}$ 及び A_x は、単に x と α によって決定されるのみでなく、経験的には、クラス数 n 及び各クラスのサイズ f_i によって変化するが、 f_i 間にそれ程大きな変動がなければその誤差は $\frac{1}{n^2}$ の order に留まる。特に左辺の A_x は階層分類を経ずに直接計算によって得られる平均差を意味するものとすれば、(169) 式が常に成立するのである。

この場合、新たに、分類 C の x に関する依存度 $D_x(C)$ を

$$(170) \quad D_x(C) = \frac{A_{x;c}}{A_x}$$

によって定義すると、(167) 及び (169) 式の関係によって

$$(171) \quad 0 \leq D_x(C) \leq 1$$

が成立する。

x が α と一致する場合は

$$(172) \quad D_x(C) \rightarrow 1 \quad (n \rightarrow \infty, v_i \rightarrow 0)$$

が容易に予想されるであろう。

第2節 一般的分配率の理論的定義と依存度

前節に於ては、専ら経験的な分類を基準にして、一般的分配率を扱ったが、これを理論的に定義する場合は、分類標識の全体の構成する空間を Ω とし、 $T_x(C) = \begin{pmatrix} V(C) \\ W_x(C) \end{pmatrix}$ を Ω に属する任意のルベーグ可測集合 (Lebesgue, M [19] 及び Kolmogorov, A. N [18] 参照) に対して定義されるベクトル値確率 (拙稿 [37] 参照) とすることが有効である。其の場合、 $T_x(C)$ の各成分 $V(C)$ 及び $W_x(C)$ は C に対して定義される二種類の確率を表わすものである。但し此の段階に於ては $W_x(C)$ の確率は x によって規定される。

今、 α を Ω の任意の要素とし、もし

$$(173) \quad T_x(C) = \int_{\alpha \in C} t_x(\alpha) dm$$

を満足する $t_x(\alpha)$ が存在するならば、 $t_x(\alpha)$ を α の限界ベクトル値密度関数と定義しよう。此の場合、勿論 $t_x(\alpha)$ の各成分 $v(\alpha)$ 及び $w_x(\alpha)$ は α の二種類の確率密度関数となる。

更に、此の場合、前節の (166), (167) 及び (168) に対応して、 x に関する相互作用素 $g_x(a_i, a_j) dm_i dm_j$, x の a に関するシニ係数 $G_{x;\alpha}$ 及び x の a に関する平均差 $\Delta_{x;\alpha}$ をそれぞれ

$$(174) \quad g_x(a_i, a_j) dm_i dm_j = t_x(a_i) \times t_x(a_j) dm_i \cdot dm_j$$

$$(175) \quad G_{x;\alpha} = \frac{1}{2} \int_{\Omega \times \Omega} |g_x(a_i, a_j)| dm_i dm_j$$

及び

$$(176) \quad \Delta_{x;\alpha} = 2 \mu_x G_{x;\alpha}$$

によって定義することが出来る。

更に (170) 式に対応して a の x に関する依存度 $D_x(a)$ を

$$(177) \quad D_x(a) = \frac{\Delta_{x;\alpha}}{\Delta_x}$$

によって定義すると、一般に

$$(178) \quad 0 \leq D_x(a) \leq 1$$

が成立するのである。

此の依存度の性質の考察は、 x 及び a に関する同時密度関数 $f(x, a)$ を導入することによって容易となる。つまり、此の場合、(175) 式は、

$$(179) \quad G_{x;\alpha} = \frac{1}{2} \int \left| \frac{1}{\mu_x} f(a_i) \int_0^\infty x f(x, a_i) dx \right. \\ \left. - \frac{1}{\mu_x} f(a_j) \int_0^\infty x f(x, a_j) dx \right| dm_i dm_j \\ = \frac{1}{2\mu_x} \int_{\Omega \times \Omega} |\bar{x}_i - \bar{x}_j| f(a_i) f(a_j) dm_i dm_j$$

と表現されるからである。但し \bar{x}_i 及び \bar{x}_j はそれぞれ a_i 及び a_j における x の条件付平均を表わすものとする。

(179) 式によって、又容易に x と a が独立の場合 $D_x(a)=0$ であり、一方 x と a が完全曲線相関々係にあるならば、 $D_x(a)=1$ であることを知る事が出来る。

特に a が正の実数と一対一に対応するならば簡潔な形式

$$(180) \quad \rho_{x;\alpha;D} = \sqrt{D_\alpha(x) D_x(a)}$$

によって、新たに x と a 間の曲線相関係数を与えることが出来るのである。

勿論このことは (170) の形式そのものが、直接 a と x との関係の規定する有効な指標であり、又 $G_{x;\alpha}$ 自体は、分類基準 α_1 と α_2 の比較及び選択を与えることを否定するものではない。

第3節 ベクトル値確率とクラスター及びモデル解析について

前節までの結果は、専ら階層区分を所与のものとして、分配量 x の構成比率を中心に考察して得られたものである。

然し乍ら、実際上屢々統計調査は、寧ろ分類の基準 a が果して妥当であるかを検討することが目的であったり、又集団内に幾つかの類型が認められる場合、それを表現する a を見いだすことを目標とするものである。

その場合は仮説 α 或は推定さるべき a が主要指標であり、 x は単なるパラメータの役割を果たすものとなる。つまり分析の立場に於ては、 x の測定そのものよりは、パターン決定要素 a を見出すことが重要なのである。

一方、前節の理論を注意深く検討するならば、 $w_x(\alpha)$ 従って $t_x(\alpha)$ は、特定の x によって規定されることなく、単に α に関する任意の確率密度関数及び任意の確率密度関数を成分とするベクトル値確率密度関数として一般的に理論を展開し得ることを見出すであろう。

ここに於て初めて一般的ベクトル値確率の概念が確立され、それが一般的分配構造の分析に対して基本的役割を果たすものと主張することが出来るのである。

一方、此の立場に於ては、第2章に述べた x の規模階層に対する分配率ベクトルは、 x と α が一意に対応する特殊なケースとして扱われるにすぎないのである。

又 x がベクトル \mathbf{x} である場合或は更に \mathbf{x} の各成分がタイム・パラメータを伴う場合の解析法は、第2章における多変量解析及び時系列解析に準じて、発展させる事が出来る。

ベクトル値確率に於て、より重要な場合は、 α がベクトル \mathbf{y} 或はタイム、パラメータを伴う y_i に一意に対応づけられる場合であり、その考察が本来的なる多量解析或は時系列解析の対象となる。

扱、計量経済学に於て用いられる同時方程式型モデル (Johnoton [14] 参照) のように \mathbf{x} と \mathbf{y} 或いは \mathbf{x}_i' と y_i の何れかを観測値ベクトル、他方を説明変数としてその間に蓋然的関数関係を設定すると、又以上の x 及び α の拡張に於いて (177) 式及び (181) 式をテンソル概念等によって拡張して $D_{\mathbf{y}}(\mathbf{x})$ 及び $\rho_{\mathbf{x}\mathbf{y};D}$ を構成し得るならば、此等の依存度及び相関係数が \mathbf{x} , \mathbf{y} に関する計量モデルの決定に役立つ事が予想されるのであり、今後の主要課題といえる。

因みに、この場合の計量モデルは本質的に非線型であることを前提としている。

む す び

第1章に於て、モメント法は中心の存在を認めるものであると述べた。それは本来的に回帰の思想によるものであり、不規則的な循環過程を想定するものである。それは、屢安定した重心をもつ質点系に喩えられるように集団内部の各個体は、相互に不干渉であり比較的独立性を保ち、且不規則的運動をし乍らも、中心そのものは失なわれない、いはば代表値を核とする構造を設定することである。その場合中心からの距離の表現がノルム ρ に関係するといえよう。

一方第2章で扱った集中曲線法のように、安定した中心の存在を否定することは、回帰の否定であり、曲率の変化に具現されるような飛躍と停滞の交錯する消長の激しい発展過程であり、それは又個体相互が乱立し規制し合い、複雑に絡み合う場を形成する競合的集団を予想させる。より正確にいうと、Gini の統計は中心を持たないのではなく、個体のそれぞれが中心なのである。従って、多極的乃至多層的構造をもつ集団に関する統計であり、相対論的観点に基づく統計といえる。

事実、遺伝、優生学や人類学に関心をよせた Galton, Pearson に於て認められるのは回帰と予定調和であり、(Pearson, E. S [22], Walkar, H. M [9] 参照) 適当な変動巾をもった不規則循環過程であるのに対し、経済及び社会現象に主に関心を示した Gini の mutability に代表される思想には、既により高次の流動的關係を予想させるものがある。(Gini, B [10], Castellano, V [3] 参照) 此等の集団観念の差がスカラの変量とベクトルの変量の異なる認識を齎らしたのであろう。

然し、何れの認識に従うとしても、本稿の末尾に附したような経済統計資料を直接解析の対象とする場合は、これまでの各章で述べたように、分布の表現或は確率概念そのものの拡張が必然的に先行する条件となるのである。かくして拡張された確率空間は多彩であり、表現力に富むものであるから、より具体的且つ複雑な現象を把握するのに有効な方法を呈供しうるものと期待出来ないであらうか。

×

×

×

終りに当り、本稿の全般に亙り多くの助言を戴いた林知己夫所長に深く感謝するものである。
統計数理研究所

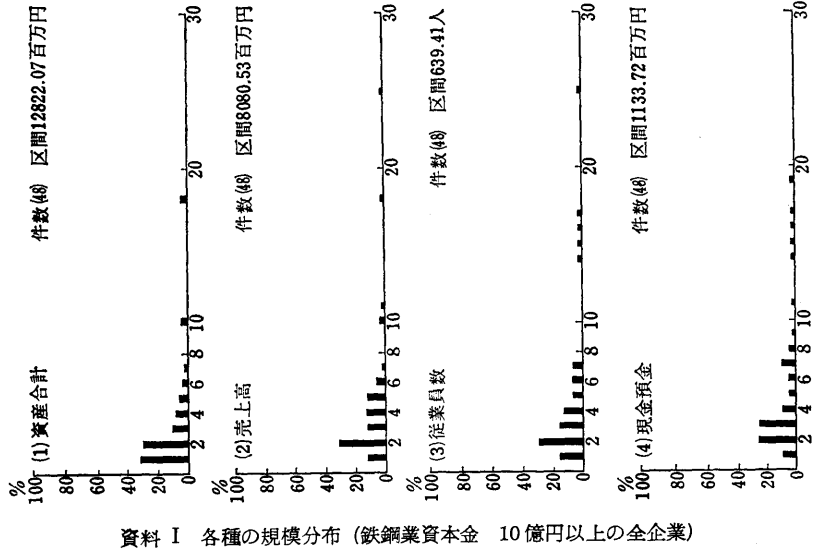
参 考 文 献

- [1] Aitchison, J. and Brown, J.A.C. "*The Lognormal Distribution*," Cambridge Univ. Press, (1958)
- [2] Atkinson, A.B. "On the measurement of inequality" with non mathematical summary in A.B. Atkinson, ed. *Wealth, Income and Inequality* Harmondsworth (1973)
- [3] Castellano, V. "Corrad Gini; a memoir," *Metron* Vol. XXIV N. 1-4 (1965)
- [4] Cramér, H. "*Mathematical Methods of Statistics*," Princeton University Press (1946)
- [5] Cramér, H. & Landbetter, M.R. "*Stationary and related stochastic processes*" John Wiley & Sons Inc., New York, London, Sydney, (1967)
- [6] Dynkin, E.B. "*Markov Processes*," vol. 1, Springer-Verlag, Berlin, Göttingen, Heidelberg, (1965)
- [7] Fechner, G.T. "Ueber den Ausgangswerth der Kleinsten Abweichungssumme, dessen Bestimmung, Verwendung und Verallgemeinerung," *Abhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften Mathematisch-physische Klasse*, X (1878) introduced by Walker, H.M. "*Studies in the History of Statistical Method*," Baltimore, The Williams & Wilkins Company, (1929)
日本語訳 足利末男, 辻 博「統計方法論史」, 京都, 高城書店 (1959)
- [8] Feller, W. "*An Introduction to Probability Theory and its Applications*" Vol. 1 and 2, New York, London, Sydney, Johns Wiley & Sons Inc., (1957 and 1966)
- [9] Gini, C. "Sulla misura della concentrazione e della variabile dei caratteri," *Atti del R. Ist veneto di Scienze, Lettere ed Art*, tomo LXXIV, Parte, Venezia, (1914)
- [10] Gini, C. "On the characteristics of Italian Statistics," *J.R.S.S.*, Vol. 128, (1965)
- [11] Hart, P.E. and Prais, S.J. "The analysis of business concentration," *J.R.S.S. Ser. A*. 119 [1956]
- [12] Heldon, G. "The relation between dictionary distribution and the occurrence distribution of word length and its importance for the study of quantitative linguistics," *Biometrika*, Vol. 45 (1958)
- [13] Iyengar, N.S. "On the standard error of the Lorenz concentration ratio," *Sankya*, Vol. 22, (1960)
- [14] Johnston, J. "*Econometric Methods*," McGraw-Hill, (1963)
- [15] Kendall, M.G. and Stuart, A. "*The Advanced Theory of statistics*," Vol. 1 (third edition) London, Charles Griffin & Company Ltd (1961)
- [16] Kendall, M.G. and Buckland, W.R., "*A Dictionary of Statistical Terms*," London, Oliver and Boyd, (1957)
- [17] Kolmogoroff, A.N. "*Grund begriffe der Wahrscheinlichkeitsrechnung*," Berlin Verlag von Julius Springer, (1933)
- [18] Kolmogorov, A.N. and Fomin, S.V. "*Measure Lebesgue Integrals, and Hilbert Space*," Academic Press, New York and London, (1961)
- [19] Lebesgue, H. "*Measure and the integral*" editted with a biographical essay by Kenneth O. May, Holden — Day Inc. San Francisco, London, Amsterdam, (1966)
- [20] Mandelbrot, B. "A note on a class of skew distribution function analysis and critique of a paper by H.S. Simon "Information and Control," Vol. 2 (1950)
- [21] 宮崎義一著「独占, 寡占, 独占的競争」近代経済学講座 III, 有斐閣, (1961)
- [22] Pearson, E.S. "*Early Statistical Papers*," Cambridge at the university press, (1948)
- [23] 労働省大臣官房政策課システム分析室「所得分配と資産形成に関するシステム分析」(1973)
- [24] Rosenblatt, M. "*Markov Processes*" Structure and asymptotic behavior, Springer Verlag, Berlin, Meidelberg, New York (1971)
- [25] 佐和隆光「計量経済学の基礎」東洋経済報社, 昭和45年
- [26] Shisha, O. "*Inequalities*," New York and London, Academic Press, (1967)
- [27] Simon, M.A. "On a class of skew distribution function," *Biometrika*, Vol. 42, (1955)
- [28] 館 稔「人口分析の方法」古今書院 (1963)
- [29] 田口時夫「パレート分布とパレート曲線の分析」統計数理研究所彙報, 第12巻, 第1号, (1964)
- [30] Taguchi, T. "On some properties of concentration curve and its applications," *Metron*, Vol. XXVI N. 3-4 (1967)

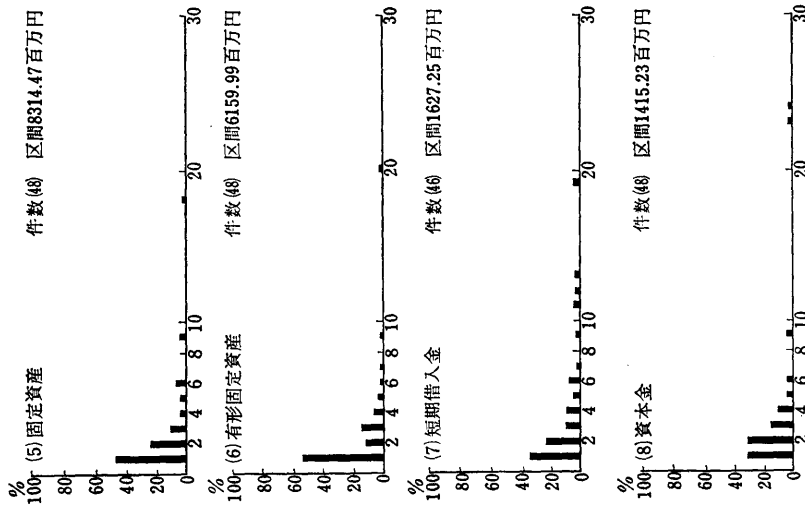
- [31] Taguchi, T. "Concentration curve methods and structures of skew populations," Ann. Inst. Statist. Math., Vol. 20, No. 1, (1958)
- [32] 田口時夫「競争システムにおける統計」, 統計数理研究所彙報, 第16巻, 第1号, (1968)
- [33] 田口時夫「Zipt 法則についての覚え書き」, 統計数理研究所彙報, 第17巻, 第2号, (1969)
- [34] Taguchi, T. "On the two dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to the two dimensional case I, II and III," Ann. Inst. Statist. Math., Vol. 24, No. 2, (1972), Vol. 24, No. 3 (1972) and Vol. 25, No. 1, (1973).
- [35] 田口時夫「二次元集中曲面の局所的性質」, 統計数理研究所彙報, 第19巻, 第1号, (1971)
- [36] 田口時夫「二次元離散型分布の集中多面体, 集中係数及び新たな各種相関係数について」, 統計数理研究所彙報, 第20巻, 第2号, (1973)
- [37] 田口時夫「J字型分布の標本抽出法及び分析法の一般化」, 昭和47年度年度末研究発表会予講集, 統計数理研究所, (1973)
- [38] Taguchi, T. "On Fechners thesis and statistics with norm p," Ann. Inst. Statist. Math. Vol. 26, No. 2 (1974)
- [39] 田口時夫「多次元集中曲面の研究」, 林 知己夫編「応用多変量解析の研究」(1974)
- [40] 竹内 啓「数理統計学」—データ解析の方法—, 東洋経済新報社, 昭和38年
- [41] 竹内 啓「計量経済学の研究」, 東洋経済新報社 (1972)
- [42] Wilks, S.S. "Mathematical Statistics," Princeton, New Jersey, Princeton University Press (1950)
- [43] Wold, H. "Demand Analysis," Johns Wiley and Sons Inc., New York, Almquist & Wiksell, Stockholm.
- [44] 安田三郎「社会統計学」, 丸善株式会社, (1969)
- [45] 米沢治文「経済統計量分析」, 日本評論社, (1972)
- [46] Yule, G.U. "An Introduction to the Theory of Statistics," London, Charles Griffin and Company, Limited, (1929)
- [47] Yule, G.U. and Kendall, M.G. "An Introduction to the Theory of Statistics," London, Charles Griffin & Company Limited, (1950)
- [48] Zipf, G.K. "Human behavior and the principle of least effort," Cambridge, (1949)

参 考 資 料

- I, II 大蔵省法人統計季報及び年報, 財政金融統計月報及びその基礎資料より作製 (昭和48年10月13日付行管承統計第339号, 昭和48年10月23日付蔵証第2894号)
- III. 総理府統計局昭和44年事業所統計調査報告, 第4巻企業篇第2表, 株式会社欄より作製
- IV, V 経済企画庁昭和45年国富調査第4巻, 法人資産調査報告, 第4-1表及び第4-2表より作製

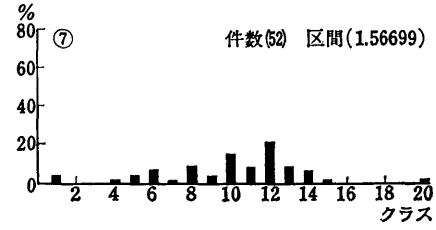
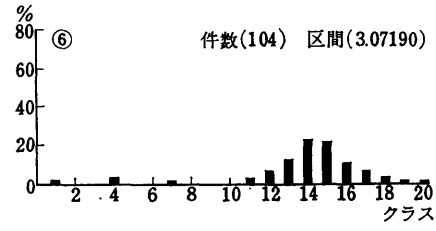
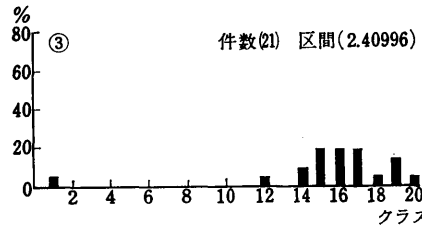
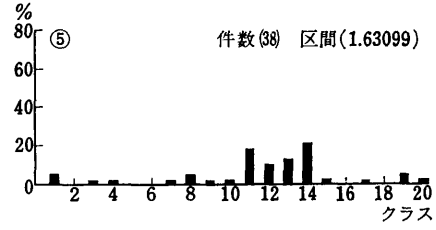
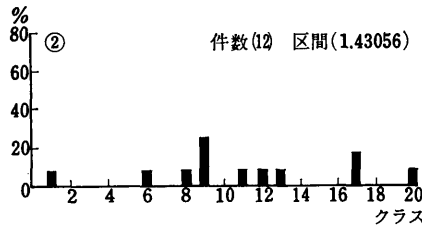
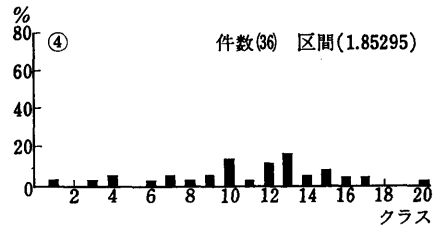
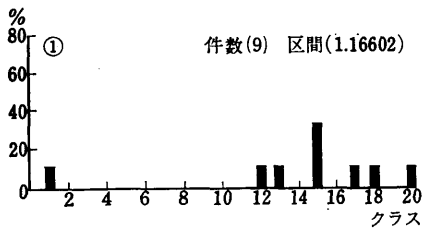


資料 I 各種の規模分布 (鉄鋼業資本金 10 億円以上の全企業)



46 年度大蔵省法人企業統計年報による

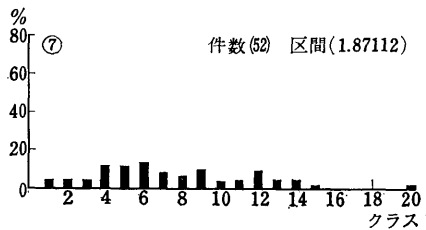
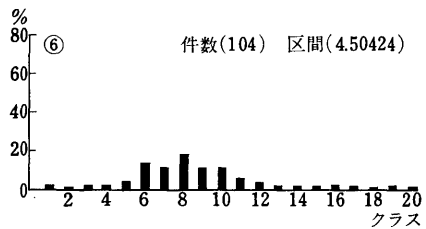
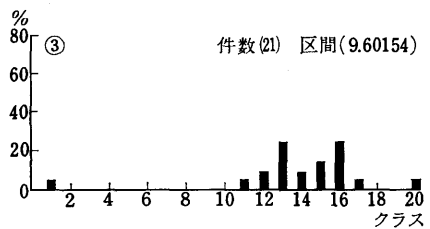
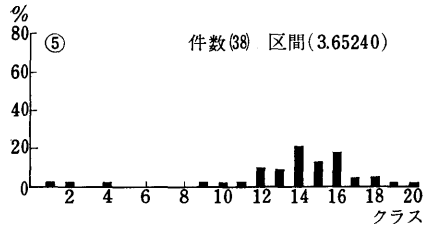
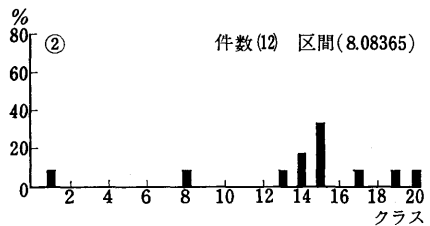
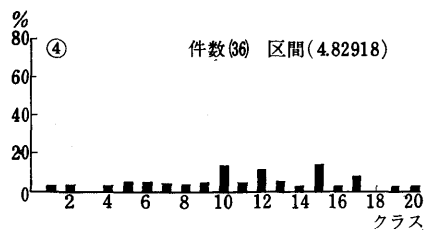
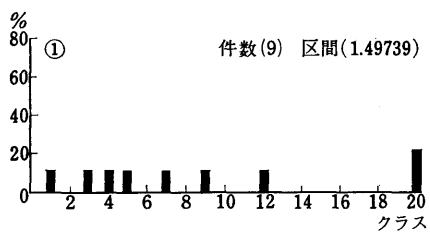
(1) 売上高営業利益率



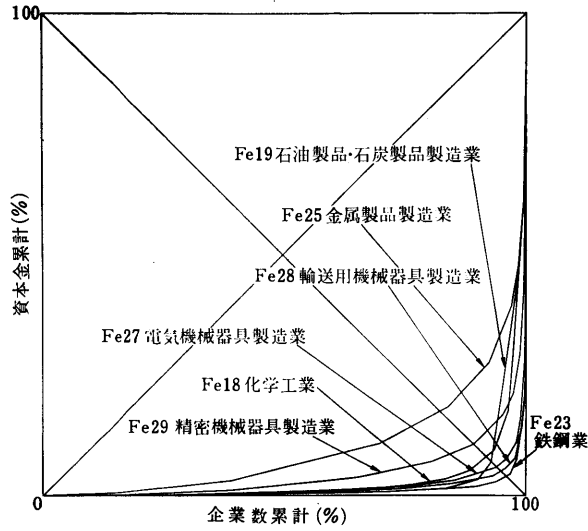
資料 II 各種の規模 (資本金) 階層別財務比率の分布 (鉄鋼業)

規模コード ① 2百万円未満 ② 2百万~5百万未満 ③ 5百万~1千万未満 ④ 1千万~5千万未満
 ⑤ 5千万~1億未満 ⑥ 1億~10億未満 ⑦ 10億以上

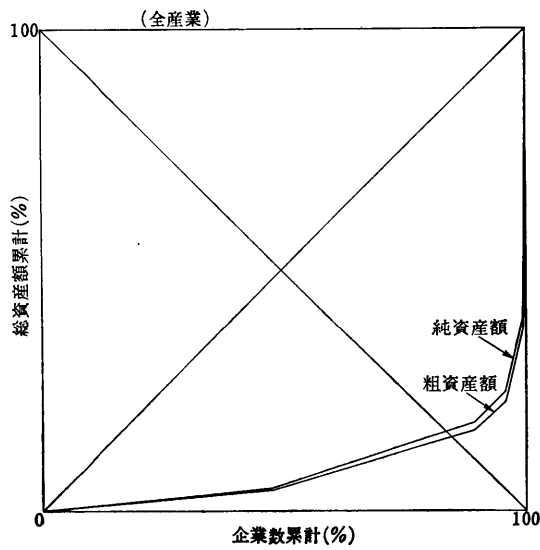
(2) 自己資本比率



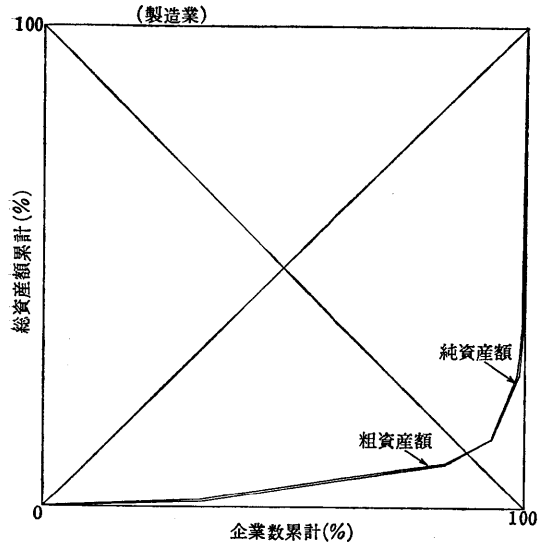
46年度大蔵省法人企業統計年報による



資料 III 資本金の株式会社への集中



資料 IV 従業者総数階級区分による所有資産の法人企業への集中



資料 V 従業者総数階級区分による所有資産の法人企業への集中