

公開講演会要旨

集団構造の統計

——領域と方法との関連——

統計数理研究所 田 口 時 夫

(昭和61年11月5日, 統計数理研究所 講堂)

1. 序 論

領域と方法との関連を課題とすることは決して目新しいことではなく、又統計学に固有の問題でもない。

数学が「科学の言葉」であり「文法」であるとしても、言葉や文法は、それが適切に選択され行使されなければ、内容を正しく伝え又相手に理解させるという本来の機能を果たすことが出来ない。言葉や文法はその表現対象によって創造され得るし又改革を要求される場合もある。逆に言葉や文法が対象の創造を促す場合もあろう。領域と方法との関連は、正に表現対象に対する言葉や文法の関係にたとえられよう。両者は相互作用或いは相互依存の関係にあるといわねばならない。然し方法の選択や適用が対象領域に対して適切であるか否かは即座に決定される訳ではなく、時には激しい又長期的な論争を生むこともある。

統計学は勿論その例外ではなく、寧ろ典型的な実例を与えているときえ思われる。それは科学的性格を獲得して近代化する過程として多くの論争を経て来たし、又現に行われつつある。

実際に統計学は、これまでにその対象についての論争やその方法についての論争は勿論、そもそも統計学は実質科学であるのか、方法論であるのかといった最も一般的抽象的な論点から、統計調査において有意選出法をとるべきか、任意抽出法をとるべきかといった最も具体的実際の論点に至るまで、大小様々な課題を提供し、歴史的論争を経験して来たのである。そして高度成長期には或程度安定した支配的な統計的領域や方法論が確立されたかにみえたが、今日してみるとそれも一時的な小康状態であったという他はない。現に一般的に構造変化が認識されつつある社会情勢の下で、統計領域もその例外ではなく、統計調査員の質とかプライバシー問題とかといった形でその環境の変化が警告されその対策が急がれている一方、統計機構そのものが改革されつつある。現在における統計環境の変化はこうした消極的な側面のみを備えるものではない。それは技術進歩や革新によって多くの計算手段や測定機器が与えられ、又それに応じて統計対象領域が拡大しつつあるといった積極的側面を対極に置いている。こうした複雑な環境変化に呼応してか、一部の統計学者の間では過去の論争の見直しが可成り活発に行われつつあるように見受けられる。我々もこれに習って足下の統計数理研究所を振り返ってみよう。

戦中・戦後の混乱期を経て当研究所が漸く体制を建て直し、再発足を図ったと思われる時期に開巻し、現在の「統計数理」の前身となっている「統計数理研究所彙報」の第2巻第1号(創立十周年記念号, 1953年)の巻頭は、当時の所長であった佐々木達治郎博士の所感によって飾られているが、その冒頭は次のようなものであることが改めて注目される。「Andre G. Laurentの言を借りれば、数理統計学では数はそれが何を数えたものであるかに全く無関係に取り扱わ

れ、資料の物質的性質に関係なく作られた全く形式的な規約である。然し実際問題においては資料の性質を考え且つその資料の得られた世界の状態を充分に考慮してこの規約を適用しないと誤った結果に到達するであろう。統計理論はこのままの姿でよいであろうか？これに就て少しく論じて見たい」として博士は、今までの統計理論は「ベクトルの統計理論」であるが、経済学や物理学の領域では、「スカラーの統計理論」や「テンソルの統計理論」の所在を暗示するような量があると指摘される。更にそれらの統計を具体的に例示し乍らも、一般理論はまだ無いようだと言われている。又多変数の場合を捉えて「実証的に変数変換を行う」と主張されている。然し以上はあくまで所感であって十分に論理的に展開されたものではないから、筆者は充分に理解することが出来ない。ただそれらが領域と方法との関連の必要性を強調し、両者を接近させる方向で新たな方法が形成されると希望をこめて説かれている事は、誰の目にも明らかであろう。然しそれは決して一領域に固有の方法論が形成される事を意味するものでない事は、博士の引用した例によっても同様に明らかである。

既に述べたように今日領域は益々拡大され又細分されつつあるので、新たな統計方法が形成される機会が多くなると共に、それらの一般理論の形成は益々多くの領域に適用されるものとなり、益々抽象的な表現が取られることになるだろう。同時にそれらの一般理論は実証されるような具体化を可能とするものでなくてはならない。

然しこのような方法の一般理論の形成は、何も先端技術と大型コンピュータの開発した領域や難しい経済学や物理学の領域においてのみ生ずるものとは筆者には思われない。寧ろ一般に親しまれている統計図表や統計表及びそれらの作製手順や利用方法についての解説書を仔細に再吟味する方向において、これまでと違った一般的な方法理論を獲得することが出来るように思われる。そしてそのような地味な足固めが何よりも必要に思われる。

実際に素材を佐々木博士の例にもある経済量に求め、一般に普及している官庁統計報告書の統計表を取り上げてみよう。その場合は統計表の活用に直結する課題となる。

2. 構造統計表と解析的集団

まず官庁統計報告書に一般的に認められる経済統計表の形式を代表的に表わしている具体例を示すことから出発しよう。

次頁の二つの表は経済企画庁の昭和30年国富法人資産調査報告書でそれぞれ第12表(p.43)及び第15表(p.46)とされた統計表でありやや古い資料であるが、この形式は昭和45年国富調査報告書にも用いられている。又家計調査報告書や法人企業統計季報・年報、更に農家・漁家、統計報告書や工業・商業統計報告書に至るまで同様な形式の統計表を見出すのは難しくない。つまりそれらに共通した一般的形式を取り出すと表3のように抽象的に図式化されるということである。

ではこの種の統計表に対してこれまでどのような解説が与えられているであろうか。幾つかの統計実務にふれたテキストから関連した事項を拾うことにしよう。

まず統計解析の対象を筆者のように統計表の観察に置こうとする立場は、特に統計解析の発生を促した見地であることが、足利(1966)に述べられているのに驚く。つまり数理統計学の語源はウィットシュタインの「数理統計学とその経済学および保険学への適用」(1867)に見出されるというのである。

すなわち、「自然科学の立場からするならば、経験からその材料をとっている統計表は、本来の意味の科学(Wissenschaft)であることを要求する権利はない。これらの統計表は単なる材料にすぎないのであって、科学はかかる材料からはじめて形成されるのである。この統計表は、

表1. 資本金階級別資産項目別法人企業資産額 (単位 百万円)

資本金階級 資産項目	総 数	500万円 未 満	500～ 1,000万円	1,000～ 3,000万円	3,000～ 5,000万円	5,000万円 ～1億円	1～10 億 円	10億円 以 上
推 計 法 人 数	357,272	340,154	7,178	6,521	1,205	948	1,073	193
資 産 総 額	6,153,141	1,561,509	184,307	356,887	149,887	228,414	1,114,617	2,557,519
A 有形固定資産	4,132,230	828,638	96,629	190,831	88,351	153,467	760,950	2,013,360
1 建 物	1,513,307	542,389	40,899	76,004	37,176	71,733	300,784	444,319
2 構 築 物	669,151	6,234	8,577	6,214	2,212	8,085	93,599	544,227
3 機 械 及 び 装 置	1,246,534	147,963	30,004	62,793	32,114	47,758	227,756	698,142
4 船 舶	202,562	2,169	4,357	14,226	5,430	8,661	75,741	91,976
5 車 輜 及 び 運 搬 具	190,699	68,236	7,889	21,826	6,932	11,494	35,894	38,425
6 器 具 及 び 備 品	138,275	59,967	4,864	9,335	3,978	5,114	21,594	33,420
7 建 設 仮 勘 定	171,599	1,609	34	415	503	615	5,575	162,845
8 家 畜 家 禽	101	67	3	14	3	5	3	2
B 棚 卸 資 産	2,020,911	732,870	87,677	166,056	61,535	74,946	353,666	544,158

表2. 資産額階級別法人企業資産額 (単位 百万円)

資 産 額 階 級	推計法人数		資 産 額					
			総 額		有形固定資産		棚 卸 資 産	
	法 人 数	構 成 比	資 産 額	構 成 比	資 産 額	構 成 比	資 産 額	構 成 比
総 数	357,272	%	6,153,141	%	4,132,230	%	2,020,911	%
0 円	19,105	5.3	—	—	—	—	—	—
500 万 円 未 満	254,067	71.2	400,569	6.5	212,984	5.2	187,584	9.3
500 ～ 1,000 万円	35,802	10.0	254,784	4.1	114,960	2.8	139,824	6.9
1,000 ～ 3,000 万円	35,880	10.0	622,116	10.1	340,775	8.2	281,341	13.9
3,000 ～ 5,000 万円	3,138	0.9	116,745	1.9	69,881	1.7	46,863	2.3
5,000万円 ～ 1億円	5,389	1.6	418,114	6.8	169,530	4.1	248,583	12.3
1 ～ 10 億円	3,359	0.9	893,739	14.5	618,490	15.0	275,249	13.6
10 ～ 50 億円	379	0.1	829,640	13.5	556,340	13.5	273,300	13.5
50 ～ 100 億円	81	0.0	579,758	9.4	395,296	9.5	184,461	9.2
100 億 円 以 上	72	0.0	2,037,672	33.2	1,653,969	40.0	383,703	19.0

自然科学の意味においては観察(Beobachtung)結果の蒐集にすぎない。そもそも自然科学なるものは、観察から自然法則にと進み、この自然法則を明らかに認識するところにその本質がある。このことは統計学についてもいえる。しかし今までのところ、従来統計学といわれてきたものは、その定義に忠実であって、この方向に対して成し遂げたところはほとんどない。すなわち、数字をみたときに直接印象に残るわずかのものを取り上げ、この課題を明白に意識して新しいものを始めることはほとんどなかった。真に科学としての統計学は、従来の統計学がやめていたところではじまらなければならない。ところで、この新しい統計学の材料は本質的に数字であるから、数学がその補助手段でなければならず、数学によってその課題が解決される。従って新しい統計学は、数理統計学(mathematische Statistik)もしくは数学者が好んでいうように解析的統計学(analytische Statistik)とよぶことができよう(S.2)といっている。ここに我々は、ヴィットシュタインが数学を補助手段とし、自然法則に匹敵する法則の定立を目的とする統計学を、自然科学の意味における科学とし、それに数理統計学という名称を

与えていることに注意しなければならない。ケトレーにくらべて、彼がむしろ統計の解析的利用に重点を移している点がそれと異なる、といてよい。この考えはクナップにもはっきりと継承されている。足利 (1966, pp. 366-367) 参照。

然し乍らヴィットシュタインの意味する統計表は、本節の冒頭に示すような複雑な統計表ではなく、主として人口を対象とした通常の生命表のようなものと考えられる。つまり同書の「第3章、形式的人口論から数理統計学へ」によるとヴィットシュタインの立場はクナップ「死亡年表の研究」(1868)以降の形式人口論により継承されたからである。従ってこの段階ではレクシス学派に代表されるような確率論が、若干の批判的要素を含むが、統計解析においては主要な役割を演じたのである。逆にその時期の「解析的集団」はそのような性格のもので充分であったといえる。つまり彼等が関心をよせた統計表は今日の言葉でいう構造統計表の域を越えないものと考えてよいであろう。因みに構造統計とは「特定標識についての観察結果を同質的ないくつかのカテゴリーに分類し集計することにより統計集団の内部構造を明らかにするデータ」工藤 (1980, p. 40) であり、その例として「職業別人口、経営耕地規模別農家数等」が挙げられるのである。従って表1, 表2における推計法人数までの欄及び表3の対象数までの欄がこれに該当するものといえる。

ところで本稿で扱おうとする統計表は、このような構造統計表に資産額のような一般的には調査項目毎の数量分布が加えられて得られる。この数量分布は特に米沢教授の主張される「実体分布」に当たるものといえよう。「すなわち経済統計にあってはただふつうの度数分布の形状や特性を把握するだけでは対象の究明に十分とはいえない。たとえば人間の身長を観察ならば度数分布表を作成し、相加平均などを算出すればそれですべてであるが、賃金の分布となると、もちろん度数分布表の作成は必要であるが、ただ何千円以下の賃金を受け取るものの数が何人または何%というだけでは足りない。そこにいまひとつ分配の問題が介在し、一定の労働者の受け取る賃金額合計がいくらかで全労働者の受け取る賃金総額の何%に当たるかということまで知り尽くすことが必要となってくる。経済統計のばあいは、たんに分布の形状だけでなく、分配の不均等ということが問題になってくる。この点はふつうの度数分布のほか実体分布とでもいうべき関係を明らかにさせることによってはたされる。

ところがいままでに述べた扱い方ではこの2通りの分布のうち一方の単純な度数分布、このばあいでいえば人員分布だけを問題にした。だが、これでは経済現象に対処する仕方としては不十分であって、どうしてもここで2通りの分布の双方の関係を明らかにする手法を問題にせざるをえないことになる。」(米沢治文, 「経済統計計量分析」(1972), p. 60)。

これによって表1~3の統計表は実体分布を伴う構造統計表といえる。実体分布は一般にベクトル値をとるとみてよい。

では本稿の主題となるこの種の、今日一般的に適用されている経済統計表に対しては、どのような解析的集団を想定すべきであろうか。以下専ら抽象的表現をもった表3を対象として考察しよう。この表においては構造統計は、 μ_j ; $j=1, 2, \dots, m$ で表現される対象数の欄で表現され、実体分布は一般には行列 (D_{ij}) ; $i=1, 2, \dots, n, j=1, 2, \dots, m$ によって与えられる。 (D_{ij}) は又調査項目数を次数とするベクトル値分布と考えることにより

$$(2.1) \quad D(A_j) = (D_{1j}, D_{2j}, \dots, D_{nj}); \quad j=1, \dots, m$$

と書くことが出来る。 (D_{ij}) の各要素は、特定分類標識に基づく A_j 階層の I_i 調査標識に関する実測値又は推定値を表わす。

表3は任意特定した一つの分類標識としているが、一般の統計表は複数の分類標識をもつものであり、その場合、各階層記号 A_j は一般にベクトル値をとるものとして扱える。

表 3. 経済統計表の描象的一表現

客体の計量属性標識 ある分類標識による階級区分		対象の分類 (規模によるカテゴリー)					全 体	
		A_1	A_2		A_j			A_m
対 象 数		μ_1	μ_2		μ_j		μ_m	μ
調査項目の標識	I_1	D_{11}	D_{12}	...	D_{1j}	...	D_{1m}	$D_{1\cdot}$
	I_2	D_{21}	D_{22}	...	D_{2j}	...	D_{2m}	$D_{2\cdot}$
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
	I_i	D_{i1}	D_{i2}	...	D_{ij}	...	D_{im}	$D_{i\cdot}$
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
	I_n	D_{n1}	D_{n2}	...	D_{nj}	...	D_{nm}	$D_{n\cdot}$
総 計		$D_{\cdot 1}$	$D_{\cdot 2}$...	$D_{\cdot j}$...	$D_{\cdot m}$	$D_{\cdot \cdot}$

さて、以上で解説した表 3 をもとにして解析的集団を規定してみよう。この場合、表 3 の主要な性格を幾つかの公理として表現するのが適切であろう。その最も一般的な形式は次のようなものである。

定義 1. 集合 Ω における σ 加法族 \mathfrak{A} に属する任意の集合 A に対して、測度 $\mu(A)$ およびベクトル値集合関数 $\Delta_n(A)$ が与えられるものとする。このとき $\Delta_n(A)$ の各成分 $\Delta_i(A)$ が完全加法的であるとき $(\Omega, \mathfrak{A}, \mu, \Delta_n)$ を n 次元の統計解析的集団または、単に n 次元の解析的集団 H_n と定義する (田口 (1984), 第 2 章 統計解析の基礎概念, p. 30)。

ここで n が調査項目数に当たるものであり、 $\Delta_i(A)$ が i 項目の実体分布に当たることはいうまでもない。実体分布間や実体分布と構造分布間の関係は、統計比率や弾力性係数の算出等の形で経済分析に利用されて来たから、更に一般的な利用は第 4, 5 節で与えられる。

然し米沢教授の引用文中にも示される不平等等の分析方法を明確に把握する為には抽象的に与えた集合 Ω を具体的に規定する必要がある。そもそも Ω は分類標識によって構成される空間であるから、当面する標識が k 種類の計量標識であるとすれば Ω は k 次元のユークリッド空間と考えてよい。更に実際の場面においては、次のような事情を充分考慮すべきである。つまり「すべての統計調査はその結果を分類集計することを目的としている。調査表に取り上げられた調査標識 (具体的には調査項目) は分類・集計の段階では分類標識として改めて登場する。」(三渚・関 (1985), 分類標識と標準化, p. 68)。

従って定義 1 で与えた解析的集団の n 個の調査標識を n 重の分類標識に用いると、又集合 A を特徴づけるベクトル $S(A)$ を定義 1 における $\mu(A)$ と $\Delta_n(A)$ により構成されるベクトル $(\mu(A), \Delta_n(A))$ を意味するものとすれば、次のような n 次元の広義の集中構造を与えることが出来る。

定義 2. n 次元の解析的集団において、 Ω を n 次元ユークリッド空間 R_n とし、また \mathfrak{A} をボレル集合族 \mathfrak{B} (全ての区間から生ずる σ 加法族) とする。更に $\mu(R_n)=1, \Delta_i(R_n)=1; i=1, 2, \dots, n$ とする。このとき (1) ボレル集合 B (\mathfrak{B} の要素) に対して、 $S_n(B)$ の全体が特異点のない閉曲面の内点および面上の点となり、(2) その曲面は可付番個の点を除き、 $I_p=[-\infty, x_p]$ または $I_p^c=(x_p, \infty]$; $p=1, 2, \dots, n$ の決める区間 B_x に対して $S_n(B_x)$ の全体によって与えられ、

(3) かつすべての p に対して 2 次元ベクトル $(\mu(B), \Delta_p(B))$ が平面上の凸集合をなすとき、この集団を特に n 次元の集中構造 K_n という。またこの場合、 $S_n(B)$ を特に集中ベクトルまたは単に集中量 $C_n(B)$ という (田口, 上掲書, p. 33)。

定義 2 を広義の集中構造を与えるものとしたのは、それが相互に相異なる幾つかのローレンツ曲線の定式とその拡張形式に共通した性格を、その全部又は一部の条件として包括しているからである。条件 (1), (2), (3) はその意味で集中に関する公理と解してよい。実際に、先の引用文に記されるような調査標識と分類標識が一致する実践的な手続きに対しては、(1) の公理のみで充分なのである。(2) と (3) はこの場合は (1) のみから論理的に誘導される。(2) と (3) の公理は調査標識と分類標識が異なる一般の場合を対象として、その間にどのような関係があればローレンツ曲線の一般化とみられるような性質が見出されるかを規定したにすぎない。田口 (上掲書, 2.1~3, pp. 38~42) における定理は一変量の場合であるが、 μ や Δ_1 に微分可能性を導入して上述の解説事項の証明を試みたものである。

以上の事柄を更に具体的に確認する為に、これまでローレンツ曲線を解析する為に用いられて来た方式を振り返ってみよう。その表現は一、二に止まらないが、それらはストカスチックな立場に従属させられていたといえよう。その代表的な形式は、連続な確率密度関数 $f(x)$ を媒介させ、次式のような分布関数 $F(x)$ と相対的不完全 1 次モメント関数 $M_1(x)$ との対応を示すグラフとすることである。つまりこの場合ローレンツ曲線は

$$(2.2) \quad \begin{aligned} F(x) &= \int_{-\infty}^x f(\xi) d\xi \\ M_1(x) &= \frac{1}{E(X)} \int_{-\infty}^x \xi f(\xi) d\xi \end{aligned}$$

で示されるものとされる。ここで $E(X)$ は確率的期待値を表現している。この形式においては、もはや統計対象にふれる事も、また、いささかの統計実務の手続きに対応する事もない。従って、統計表の利用者は、この形式を通しては具体的に対象の集団の性格を把握するのが困難であり、却って確率密度をめぐるあれこれの観念的解釈の迷路に落ち込むことにもなるのである。それに対して、測度論的立場で与えた上述の集団の構造規定は、 Ω を個体の全体で構成する集合、 A, B をその部分としてのグループ、 Δ_n や S_n はそれらについての測定値集合等と解釈することによって調査と集計・解析についての理解を一体化するように思われる。

さて、以上のような経験的な統計表の作製対象となる統計集団の抽象的一般的規定に対応して、ローレンツ曲線、集中度等の経験的な統計図式や指標についても次のような一般的・抽象的再規定が可能である。

定義 3. n 次元の集中構造は、定義 2 により集中量 $C_n(B)$ の全体に対して特異点のない境界曲面をもつが、この曲面を n 次元完全集中曲面、またその曲面で囲まれる体積を n 次元集中度という。更に $C_n(B)$ の全体を含む $n+1$ 次元ユークリッド空間 R_{n+1} 内の領域を集中空間という。

特に $n=1$ のとき $C_n(B)$ の境界曲線を完全集中曲線、またその内部の面積を集中度、集中面積を含む領域を集中平面という (田口, 上掲書, p. 33)。

この規定も経験的なローレンツ曲線の作製手続きに即した把握ということが出来そうである。つまりデータからのローレンツ曲線の作製は、通常、測定値の大なるものから逐次累積して得られるから、その他の任意の累積順位に従って得られる同種の曲線 (折線) は、常にこの

二種類のローレンツ曲線に挟まれた状態の曲線になるという事情に即しているからである。

累積の方向を小から大へとするか、大から小へとするかによって得られる二つの曲線の何れを正しいとするかという課題が屢々提起されるが、この定義は全ての累積曲線の境界曲線として両者を一体として捉える事に意義を見出している。つまりこの領域は任意の累積曲線の存在範囲を示すものであり、ジーニ係数は直接その大きさを示すものとして新たに意義を見出すからである。因みに(2.2)式によった場合のジーニ係数 G_x は

$$(2.3) \quad G_x = \frac{\Delta_x}{2E(x)}$$

$$\Delta_x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x_1 - x_2| f(x_1) f(x_2) dx_1 dx_2$$

で与えられることは、よく知られているところである。

本節の結びに集中構造の例として、歴史的に著名なパレート分布を取り上げてみよう。その最も簡単な形式は

$$(2.4) \quad F(x) = 1 - \left(\frac{a}{x}\right)^p; \quad p > 1, a > 0, x \geq a$$

である。この分布に対して(2.3)式のジーニ係数 G_x 及び平均差 Δ_x はそれぞれ

$$(2.5) \quad G_x = \frac{1}{2p-1}, \quad \Delta_x = \frac{2ap}{(p-1)(2p-1)}$$

となる。又定義3で与えられる完全集中曲線は(2.2)式の x を消去し、 $F(x)$ 、 $M_1(x)$ をそれぞれ記号 U 、 V で表わすと次のような二つの曲線の結合として与えられる。すなわち

$$(2.6) \quad 1 - V = (1 - U)^{(p-1)/p}, \quad V = U^{(p-1)/p}$$

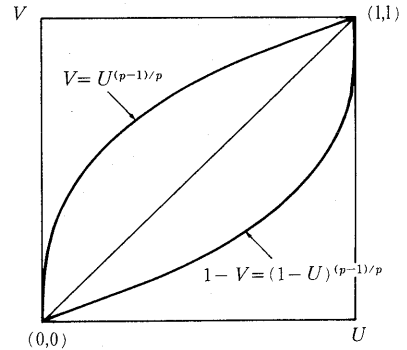


図1. パレート分布((2.2)式)の完全集中曲線 ($p=1.5$ の場合)

3. 2次元の集中構造

前節の定義1~3で与えた解析的集団は、本来は、多種標識によって捉えられることを前提としている。これに対応するようなストカスチックな定義は、これまでのところ、ジーニの影響を受けたイタリア学派に二変量の場合について若干の試みがあるにすぎない。

前節においてはローレンツ曲線と比較して本稿の測度論的定義の利点を解説したが、この定義の利点は寧ろ多変量を対象とした時に明確にされる。つまりストカスチックな表現の複雑さに比べて、直観的で簡潔である事が浮彫りにされるからである。因みに筆者や Lunetta の与えたストカスチックな形式の二変量の集中曲面は

$$(3.1) \quad U(x, y) = P(X \leq x, Y \leq y)$$

$$V(x, y) = \frac{E(X | X \leq x, Y \leq y)}{E(X)} P(X \leq x, Y \leq y)$$

$$W(x, y) = \frac{E(Y | X \leq x, Y \leq y)}{E(Y)} P(X \leq x, Y \leq y)$$

によって表現されるものである。定義2~3で示される完全集中曲面は、(3.1)式のより一般的な形式であり、その最も単純な場合、つまり調査標識と分類標識が一致した場合、その一部分

が(3.1)式と一致する。この最も単純な場合においても定義3で与えた完全集中曲面を表現しようとすれば、更に(3.1)式に類した三つの曲面を与えねばならない。逆に(3.1)式を部分とする最も単純な場合、定義2は単に公理(1)のみで充分なのである。

然し乍ら(3.1)式で示すようなストカスチックな形式は、定義1~3で示される一般的な規定に到達する為の貴重な手掛りであった。この形式は筆者自身が探索的に与えたものであり、この曲面を補完する三つの曲面を加えて得られた完全集中曲面は閉曲面であり、完全集中曲面上の任意の点における法線ベクトルは常に面の外側に向かうことが確認された。それは原点がこの面上にあることから、法線ベクトルを U, V, W の各軸ベクトルに分解することによって把握されるのであり、これによって完全集中曲面が体積を確定し、その結果は

$$(3.2) \quad V = \frac{1}{4E(X)E(Y)} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \{\text{sgn}(x_1 - x_3)\} \{\text{sgn}(y_2 - y_3)\} \\ \times \begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} \prod_{i=1}^3 f(x_i, y_i) dx_i dy_i$$

と算定されるのである。これによると、 V は(2.3)式との対応において二変量のジュー=係数 G_{xy} を規定するのに適した形式をもつといえよう。

連続な密度関数 $f(x, y)$ に対して、その完全集中曲面は法線ベクトルの性質から凸曲面であり、これを定義2に対応させると、集中ベクトル $S_2(B)$ の全体は凸集合であるといえる。この考察を更に帰納的に推し進めると、一般に n 次元の完全集中曲面は凸な多様体であり、 $S_n(B)$ の全体が凸集合を形成している公算が大きい。実体分布を伴うだけに集団・集中構造は一般に豊富な内容をもつものであり、その全体及び部分についての体積量やその結合は、多くの統計的指標を与えてくれる。話を単に二変量に関する集中構造に限定してみても、(3.2)式で与えられる二変量のジュー=係数のみならず、左辺の積分形式は二変量の平均差と考えられる。又この場合の完全集中曲面の各構成曲面の各座標面への正射影は、その正射影面積として一変量のジュー=係数の他に相関係数や線形回帰係数を決定する基本量を与える。曲線回帰の場合は、その基本量は x, y 間の弾力性係数に依存する。これらの基本量は、集中構造に固有の外積量(階層間の相互作用素)の平均形式をとるもので、確率論的なモメント形式をとるものではない。次節に述べるように、それらは異なる観点に立つものであるから、正確には相関係数や回帰係数と区別して、相互依存度や傾向係数等とした方がよいかも知れない。又均等線方向への正射影は二変量の平均偏差を決定する。その詳細は田口(上掲書)に示されているのでここではこれ以上触れないことにする。ただ前節と同様に二変量のパレート分布について K.V. Mardia の与えた形式をもとにして、その二変量ジュー=係数と集中曲面を解析的実例として示すことにする。

次の(3.3)式で示す独立の場合を除いて、Mardia は二種のパレート分布を提案したが、それは線形回帰をもつ第I型と、定弾性値回帰をもつ第II型である。それらはそれぞれ(3.4)、(3.5)式で示される。

$$(3.3) \quad f(x, y; a, b, p, q) = \begin{cases} \frac{pqa^pb^q}{x^{p+1}y^{q+1}}; & x \geq a > 0, y \geq b > 0; p > 1, q > 1 \\ 0; & x < a, y < b; p > 1, q > 1 \end{cases}$$

$$(3.4) \quad f(x, y; a, b, p) = \begin{cases} \frac{p(p+1)(ab)^{p+1}}{(bx+ay-ab)^{p+2}}; & x \geq a > 0, y \geq b > 0; p > 1 \\ 0; & x < a, y < b; p > 1 \quad (\text{第I型}) \end{cases}$$

ただしこの場合、周辺分布に関して $p=q>1$ に限られる。

$$(3.5) \quad f(x, y; a, b, p, q, \alpha) = \begin{cases} \frac{pq}{(1-\alpha)xy} \left\{ \left(\frac{a}{x}\right)^p \left(\frac{b}{y}\right)^q \right\}^{1/(1-\alpha)} I_0 \left[\frac{2 \{ \alpha pq \log(x/a) \log(y/b) \}^{1/2}}{1-\alpha} \right]; & x \geq a > 0, y \geq b > 0; p > 1, q > 1, 0 < \alpha < 1 \\ 0; & x < a \text{ または } y < b; p > 1, q > 1, 0 < \alpha < 1 \quad (\text{第 II 型}) \end{cases}$$

ただし $I_k(z)$ は k 次の第 1 種変形ベッセル関数を表わす (田口, 上掲書, p. 150)。

既に述べたように (3.3) 式は X, Y が確率的に独立な場合を意味する最も平凡な場合であるが、この分布の (3.1) 式で表わされる集中曲面は図 2 で示される。

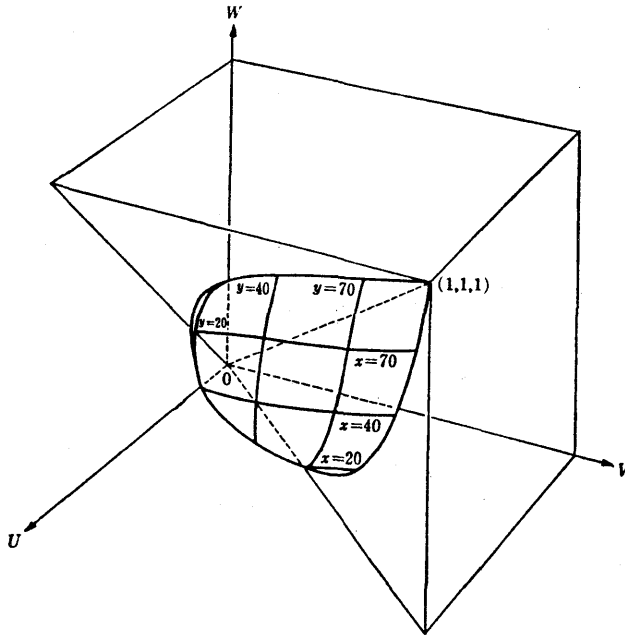


図 2. 相互独立な二変量パレート分布の集中曲面の一部 (田口 上掲書, p. 112)

又二変量の集中係数は (3.3) 式に対しては

$$(3.6) \quad G_{xy} = \frac{3}{4} G_x G_y; \quad G_x = \frac{1}{2p-1}, \quad G_y = \frac{1}{2q-1}$$

であるが、(3.4), (3.5) 形式の分布に対しては、それぞれ

$$(3.7) \quad G_{xy} = \frac{(p-1)^2}{(2p-1)^2} \left\{ \frac{p+1}{p-1} {}_3F_2(p-1, 1, 1; 2p, 2p; 1) - \frac{(p+1)^2}{p^2} {}_3F_2(p, 1, 1; 2p, 2p; 1) \right\}$$

$$= \frac{1-1/p^2}{(2p-1)^2} \{p^2 {}_3F_2(p-1, 1, 1; 2p, 2p; 1) - (p^2-1) {}_3F_2(p, 1, 1; 2p, 2p; 1)\} \quad (\text{第 I 型})$$

及び

$$(3.8) \quad G_{xy} = -\frac{(p-1+\alpha)(q-1+\alpha)}{\{2(p-1)+\alpha\}\{2(q-1)+\alpha\}-\alpha(p-1)(q-1)} + \frac{(p-1)(q-1+\alpha)}{2p(q-1+\alpha)+(2-\alpha)\{(p-1)(q-1)-\alpha\}} + \frac{(q-1)(p-1+\alpha)}{2q(p-1+\alpha)+(2-\alpha)\{(p-1)(q-1)-\alpha\}} + \frac{-pq+p+q}{(2p-1)(2q-1)-\alpha(p-1)(q-1)} - \frac{1}{2} \left\{ \frac{(p-1)\alpha-(q-1)}{(2p-1)(q-1)+(p-1)\alpha} - \frac{\alpha}{2(q-1)+\alpha} + \frac{1}{2p-1} + \frac{(q-1)\alpha-(p-1)}{(2q-1)(p-1)+(q-1)\alpha} - \frac{\alpha}{2(p-1)+\alpha} + \frac{1}{2q-1} \right\} \quad (\text{第 II 型})$$

となる。ジブラ所得分布を初めとするその他の分布の解析や実際のデータ解析についても、田口（上掲書）に詳しいのでここでは触れない。

4. 一般の解析的集団と相互依存関係

実体分布数が増えることによって集中構造が豊富な統計解析の指標を与えることは既に述べたが、定義1の解析的集団は、一般的には集中構造と異なり、分類標識は量的なものに制限されず又その数も制限されないから、豊富な実例を示すことが出来るし、応用範囲も広い。それは単に表1~3のような統計表の全体及び部分の抽象的表現であるに止まらず、図3~5のような諸統計グラフの解析的表現でもある。

逆に、これらのグラフは解析的な集団の部分的図示といえる。すなわち図3を例にとると、比

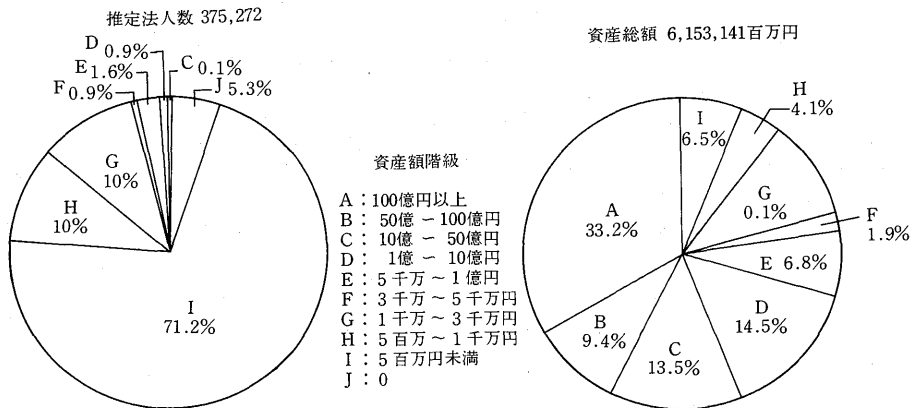


図3. 法人企業資産額階級別構成（昭和30年12月31日現在）
経済企画庁「昭和30年国富調査年報」

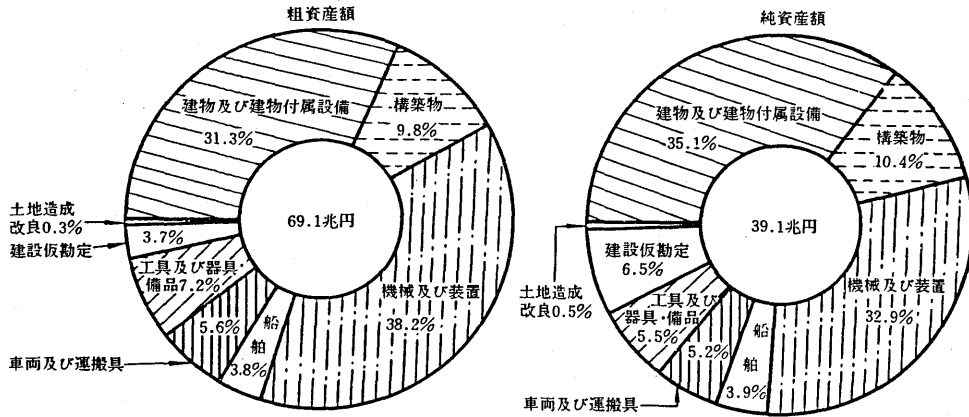


図4. 法人企業有形固定資産額の資産項目別構成
経済企画庁「昭和45年国富調査法人資産調査報告」

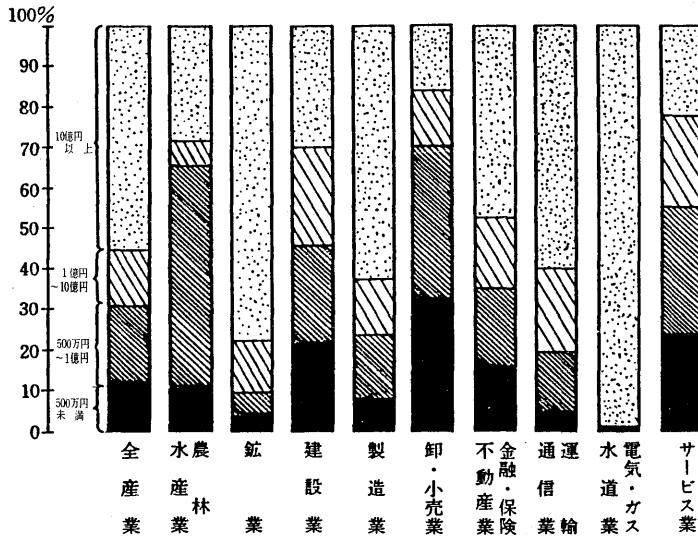


図5. 法人企業の産業別、資本金階級別所有粗資産総額構成比
(単位：%) (同上, p. 25)

較対象とされる推定法人数と資産総額とをそれぞれ横軸，縦軸とし，各扇形の測定規準点から階層分点までの角度を 2π で割り標準化して軸座標とすれば，階層順位が一定している限り，それがローレンツ曲線（折線）そのものを与えることが分かる。

図4については粗資産額を横軸とし，純資産額を縦軸として図3と同様な操作を施すと，これは二次の解析的集団における実体分布の一つの構造表現とみることが出来る。つまり表3の調査項目標識は，それが集計項目標識であったとしても，抽象的な統計表の構成には何等影響を与えないからである。又図4は図3と異なり分類標識は計量的な性格ではない資産の種類であるが，これも定義1の抽象的な集団構造の実体分布として捉えることが出来る。

図5は扇状グラフと異なる柱状グラフであるが，このうちの何れか二つを選びそれぞれの階

層分点の%尺度を縦、横の座標として読み取れば、それが図4の場合と同様な集団構造が得られるのである。

一方においてこのグラフは各産業を調査項目標識として捉え、資本金額を分類標識として受け取れば、この全体のグラフが即、多変量の実体分布を構成していることになる。これらの諸図は最もポピュラーな統計グラフであり、一般の理解を容易にするものであるから、それらが表1~2の統計表と共に集団構造と直結しているということは、この抽象化が統計的な意味で普遍的妥当性をもつものであり、この抽象概念の論理的解析的展開は一般的に統計的認識を深めることになるといえる。

さて、これまで説明したことから図3~5の各々が図6に示すようなグラフに変換出来ることが理解されよう。このグラフを一般的に相互依存曲線とすれば、それは規模分布と実体分布或いは二つの実体分布間の相互依存をローレンツ曲線に似たグラフで表現したものであり、ある分類標識に基づくカテゴリーの規模と実体量或いは実体量間の相互依存関係を示すものといえる。

ところで実用的な諸々の統計グラフに対してこの相互依存曲線を基準とすることは、一つの解析的意味をもっている。それは、ローレンツ曲線に対する均等線やジニ係数の果たす役割が、この図式に対しても導入出来るからである。すなわち図6において均等線に当たるものは対角線 OB であり、集中面積に当たるものは OB と相互依存曲線とで囲まれる面積である(但しこの場合の面積は、一般にはそれが OB に対して上下何れかの側にあるかによって符号上で区別されるものとする)。この時、各階層がしめる二つの軸成分をそれぞれ正の調査標識 x, y に関する実体分布 v_i, w_i とすれば、面積 A_{vw} は折線で考えると

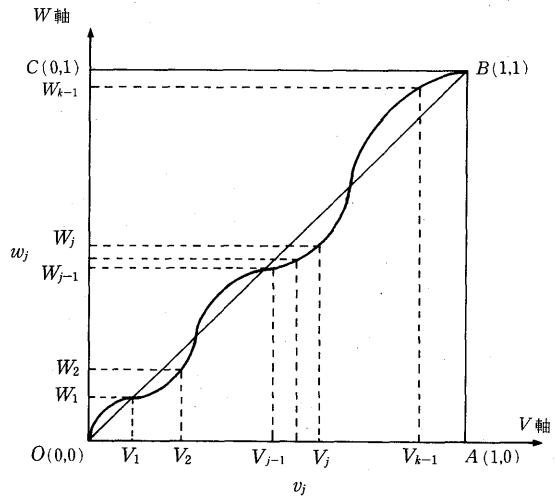


図6. 一般的な相互依存曲線

$$A_{vw} = \frac{\Delta_{vw}}{4}, \tag{4.1}$$

$$\Delta_{vw} = \frac{1}{k^2} \sum_{i=1}^k \sum_{j=1}^k \text{sgn}(j-i) \begin{vmatrix} v_i & v_j \\ w_i & w_j \end{vmatrix}$$

となる。一般に A_{vw} はカテゴリーの累積順位に依存するもので、もし $\frac{w}{v}$ の小なるものから逐次大なるものへと累積して曲線を構成している場合は、 A_{vw} は全ての順位のうちで最大の値 \bar{A}_{vw} をとり、この時曲線は常に U 軸に対して凸の曲線となる。又この順位を逆に大なるものから累積した場合、 A_{vw} は最小の値 \underline{A}_{vw} をとり、曲線は U 軸に対して凹となる。更に(4.1)式の関係によって \bar{A}_{vw} と \underline{A}_{vw} に対応する Δ_{vw} をそれぞれ $\bar{\Delta}_{vw}$ と $\underline{\Delta}_{vw}$ とすれば

$$\bar{\Delta}_{vw} = \frac{1}{k^2} \sum_{i=1}^k \sum_{j=1}^k \begin{vmatrix} v_i & v_j \\ w_i & w_j \end{vmatrix}, \tag{4.2}$$

$$\underline{\Delta}_{vw} = -\bar{\Delta}_{vw}$$

となる。更に一般に

$$(4.3) \quad -1 \leq 2A_{vw} \leq 2\bar{A}_{vw} \leq 1$$

が成立するから、 $2A_{vw}$ を集中度に準じて x, y 間の相互依存度 D_{xy} とすることが出来る。この時 D_{xy} は、もし階層の順位が合理的な基準に従って固定されている場合は、二つの実体分布間の総合的比較指標を与える。すなわち、もし両分布の構造の差が大きければ D_{xy} は 1 に近い値をとり、逆に両分布の構造が均等的になれば D_{xy} は 0 に近くなるからである。

又もし v, w が同一性質の分布であるが異なる時間において得られるものとすれば、 D_{xy} は分布の時間に関する相互依存度として、その大小は構造変化の大小を示す統計指標となる。

さて、上例では二つの分布を全て実体分布としたのであるが、実は規模分布と実体分布或いは規模分布間の相互依存関係を同様に構成し得るのである。又その分布の種類も決して図 3~5 で示したような経済量である必要はない。それは例えば二つの比較時点における年齢階層別の人口規模分布をもとにしても構成されるのである。この場合に得られる年齢別人口の時間に関する相互依存曲線は必ずしも常に凸又は凹とはならない。つまり世代間で人口の増減がまちまちに変化するからである。この世代間の変化の大きさは D_{xy} と \bar{D}_{xy} との比較によって測定することが出来る。その差が大である時 D_{xy} の変化は大きく、世代間の複雑な変化を示すことになる。この比較は又

$$(4.4) \quad M_{xy} = 1 - \frac{|D_{xy}|}{\bar{D}_{xy}}$$

の形で指標化すれば

$$(4.5) \quad 0 < M_{xy} < 1$$

で捉えられることになる。

さて、相互依存関係は必ずしも確率論を前提とするものではないが、もし階層標識 A と調査標識 x, y の間に同時生起の確率が想定されるならば、(4.1) の式 Δ_{vw} は

$$(4.6) \quad \Delta_{vw} = E_{(A_i, A_j)} \left[\{ \text{sgn}(j-i) \} \left| \frac{E(X | A \in A_i) E(X | A \in A_j)}{E(Y | A \in A_i) E(Y | A \in A_j)} \right| \times P(A_i) P(A_j) \right] / E(X) E(Y)$$

と表わすことが出来る。又二つの規模分布に対しては

$$(4.7) \quad \Delta_{vw} = E_{(A_i, A_j)} \left[\{ \text{sgn}(j-i) \} \left| \frac{P(X | A \in A_i) P(X | A \in A_j)}{P(Y | A \in A_i) P(Y | A \in A_j)} \right| \times P(A_i) P(A_j) \right]$$

と表現することが出来る。

(4.6) 式や (4.7) 式は全く形式的に導いたものであるが、これは確率事象の解析に何等かの構造を持ち込む可能性を示すかも知れない。

以上の例で示した各種の相互依存度とそれらの間の関係は、経験的な統計グラフの理解を深め、対象とする集団についての分析を進める上で役立つことが期待される。

現に、横軸で中心地からの同心円の面積を、縦軸でその面積内の人口を表わすことによって得られる相互依存関係図は人口集中の分析に利用されているし、分類標識を職業や産業にとった就業人口分布と賃金分布による相互依存関係は、産業構造や賃金構造の変化の基準指標を与え得るであろう。又年齢を分類基準とした男女別の人口の比較図は、一つの社会指標を与えよう。

更に任意特定の二つの国を各軸とした相互依存曲線・相互依存度は多面的な尺度の下で国際比較を可能とするであろうし、それをもとに各国の類別や順位づけを試みてみることを可能としよう。又同一種類の分布の時間に関する相互依存度は寄与度・寄与率の総合指標でもある。構造解析の対象は、その他投入算出表や商品種別の輸出入表等を数えることが出来よう。

5. む す び —— 集団構造の方法論的利用 ——

集団構造や相互依存関係は単に統計的記述や解析にとって有用であるばかりではない。それは現にQC活動においてパレート図として実用的方式を生んでいる。又統計的推論においても二標本法の形で検定に用いられている。更に単純任意抽出法と規模比例任意抽出法の併用を前提として、集団的構造を与え、集中度の不偏推定量を与えることが出来る (Taguchi (1978))。そればかりでなく、集団構造を適当に想定することによって、我々はそれを標本調査論そのものにも利用することが出来る。それは有意選出法の見直しにかかわるものであって、exactな回帰関数と近似的な回帰関数の二つをもとに構成した相互依存曲線の性質をもとにして展開された平均値の近似的推定方式である。従って、この場合の相互依存曲線は回帰適合曲線ともいえる性格を所有している。パレート母集団を想定すると、この方式により、実務的なロバスト推定量が与えられることが立証出来る。パレート母集団に対する確率抽出法がこれまで充分満足の結果を得ていない実情からすれば、この方式は実用化の希望を抱かせるに充分な例題を与えているといえよう。然し、その内容の詳細は本誌の別稿に詳しいのでここでは特にこれ以上触れないことにする。

参 考 文 献

- 足利末男 (1966). 社会統計学史, 三一書房, 366-367.
 工藤弘安 (1980). 統計学, 地方公務員新研修選書9, 学陽社, 40.
 Mardia, K.V. (1962). Multivariate Pareto distributions, *Ann. Math. Statist.*, **33**, 1008-1015.
 三瀧信邦・関弥三郎 (1985). 経済統計論, 有斐閣ブックス, 68.
 佐々木達治郎 (1954). 統計数理に関する所感, 統計数理研究所彙報, **2**, 第1号, 1.
 Taguchi, T. (1978). On an unbiased consistent and asymptotically efficient estimation of Gini's concentration coefficient, *Metron*, **XXXVI**, No. 3-4, 57-72.
 田口時夫 (1984). 経済分析と多次元解析, 東洋経済新報社, 30, 33, 38-42, 112, 150.
 米沢治文 (1972). 経済統計計量分析, 日本評論社, 60.

参 考 資 料

- 「法人資産調査報告」昭和30年国富調査, 第3巻, 経済企画庁, 1957.
 「法人資産調査報告」昭和45年国富調査, 第4巻, 経済企画庁, 1974.