《論説》

ジーニの集中比

木　村　和　範

はじめに

任意の大きさ x' を越える強度の個体数 $N(x')$ が系列を構成する全個体数 $N(x)$ にしめる割合を P，$N(x')$ 個の個体の強度の合計 $A(x')$ が系列全体の強度の総和 $A(x)$ にしめる割合を Q とおいたとき，ジーニは，

$$P=Q^{\delta}$$

の右辺の「べき」δ が所得分布の集中度を統計的に計測するための指標になると考えた[1)]。

ここに，「強度」とはいわゆる集団性の方向とその強度と言うときの強度ではない。ジーニの「強度（intensitã）」とは，たとえば，ある人間の身長や体重，あるいは所得など，個体がもっている数量的規定性のことである。本稿では「強度」をジーニの意味で用いることとする。

ジーニは，上式の δ を「集中指数」と名づけたが，これを誘導するとき，彼はパレート分布を前提した。その意味では，δ は，パレート分布のひとつのパラメータと考えることもできる。いかなる所得分布といえども，このような関数関係 ($P=Q^{\delta}$) で一意的に特徴づけられるものであろうか。

ジーニは，集中指数 δ がパレート指数 α に較べて，所得分布の相異を鋭敏に検出するとして，その優位性を主張したものの，所得分布を含めて一般に，関数関係として把握することができない強度の分布について，その集中度を計測するには，集中係数ではその有効性を発揮しえないと考えた。そして，新たな計測指標を構想するにいたった。これが，後に「ジーニ係数」と通称された「集中比」を定式化するときに，ジーニの念頭にあった問題意識である。

1）①Gini, Corrado, “Indici di concentrazione e di dipendenza,” *Atti della Società Italiano per il Progresso delle Scienze, Terza Riunione, Padva, Settembre 1909*, Roma 1910; ②ditto, “Indici di concentrazione e di dipendenza,” *Biblioteca dell’Economista*, Serie 5, Vol.20, 1922.

この課題に取り組んだジーニは，1914年に，その研究成果を「特性の集中と変動性の計測について」というタイトルの論文（以下,［ジーニ（の）］1914年論文と略記）で公表した[2)]。

この論文において，はじめて「集中比」（いわゆるジーニ係数）が定義された。今日では，所得分布の集中度を計測するために，何らかの定義式が最初にあたえられ，そのもとでジーニ係数を計算して，所得分布の不平等度を計測することが珍しくない。ところが，計算の出発点に置かれるジーニ係数の定義式は単一ではなく，複数あって，そのどれを用いるかについて定型はない。ローレンツ曲線と所得均等直線とで囲まれた図形の面積とジーニ係数との間の数学的関係から，ジーニ係数（集中比）の数理的意味を視覚的に説明しようとする試みも古くから見られる。いずれにしても，ジーニ係数は，所得分布との関連で取り上げられ，換言すれば所得分布の統計的計測のための測度として果たす機能が強調されているように見受けられる。

このような理論状況のもとでは，ジーニの集中比（ジーニ係数）が，ひとり所得分布とだけ結びついた統計的な分布尺度ではないことを，1914年に公刊されたジーニの原典に即して確認することは，集中比が果たすとされる機能を解明するうえで必要なことと思われる。それとともに，ジーニ係数がさまざまな数式で表現される根拠を考察することも，ジーニ係数の意義を検討する際には，資するところがあるように思われる。

本稿では，1914年論文に即して，ジーニによる数式展開を跡づけるが，その論文では誘導と被誘導の関係にあるはずの数式と数式が，あたかも線で結ばれていない点と点のように見えて，数式展開が分かりやすい叙述になっているとは言いがたい。ジーニの所説を理解するにはこの点と点とを結ぶ作業が必要と考えられるゆえんである。以下では，1914年論文における数式展開にたいする解説を旨として，ジーニの所説を跡づける。

1．集中比の定義式

(1)　集中比の定義

個体のもつ数量的規定性（身長，体重，所得などのジーニのいわゆる強度）を一般にa_iで表す。強度をa_iとする個体はひとつあって，ひとつしかないものとする。そして，異なる個体の強度の間には，

$$a_i < a_j \tag{1}$$

$$\text{ただし，}\ i<j\,(i=1,\ 2,\ \cdots\cdots,\ n-1\ ;\ j=2,\ 3,\ \cdots\cdots,\ n)$$

という関係があるものとする[3)]。このような関係のもとにある強度の分布を数学的に取り扱う目的で，用いる記号の意味をあらかじめ表にまとめておく（表1）。

表1のように道具立てをしておいて，以下でジーニによる「集中比」（後のジーニ係数）の誘導を追跡する。

ジーニによれば，系列の構成要素としての個体をその強度の昇順に並べたときの，個体の個数にかんする累積百分率p_i（個体比率）とその強度にかんする累積百分率q_i（強度比率）の間には[4)]，強度が個体間で等分配さ

2）Gini, C., “Sulla misura della concentrazione e della variabilità dei caratteri,” *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Anno accademico 1913-1914, Tomo LXXIII, Parte seconda,［以下 Gini（1914）と略記］.

3）ジーニの1914年論文では$a_i \leqq a_j$となっているが，数学的取り扱いの簡便性からここでは，さしあたり$a_i < a_j$とした。しかし，以下の数式展開は$a_i \leqq a_j$の場合にも成り立つ。なお，特殊な場合の$a_i \leqq a_j$における取り扱いについては，後に2．と3．で言及する。

4）個体比率と強度比率は筆者の造語である。

表１　記号一覧表（その１）

個体番号	個体の強度	個体の累積度数	個体の累積百分率(個体比率)	累　積　強　度	強度の累積百分率(強度比率)
1	a_1	1	$\frac{1}{n}=p_1$	$\sum_{k=1}^{1} a_k=a_1 \quad =A_1$	$\frac{A_1}{A_n}=q_1$
2	a_2	2	$\frac{2}{n}=p_2$	$\sum_{k=1}^{2} a_k=a_1+a_2 \quad =A_2$	$\frac{A_2}{A_n}=q_2$
3	a_3	3	$\frac{3}{n}=p_3$	$\sum_{k=1}^{3} a_k=a_1+a_2+a_3 \quad =A_3$	$\frac{A_3}{A_n}=q_3$
⋮	⋮	⋮	⋮	⋮	⋮
i	a_i	i	$\frac{i}{n}=p_i$	$\sum_{k=1}^{i} a_k=a_1+\cdots\cdots\cdots+a_i \quad =A_i$	$\frac{A_i}{A_n}=q_i$
⋮	⋮	⋮	⋮	⋮	⋮
$n-1$	a_{n-1}	$n-1$	$\frac{n-1}{n}=p_{n-1}$	$\sum_{k=1}^{n-1} a_k=a_1+\cdots\cdots\cdots\cdots\cdots+a_{n-1} \quad =A_{n-1}$	$\frac{A_{n-1}}{A_n}=q_{n-1}$
n	a_n	n	$\frac{n}{n}=p_n=1$	$\sum_{k=1}^{n} a_k=a_1+\cdots\cdots\cdots\cdots\cdots+a_{n-1}+a_n=A_n$	$\frac{A_n}{A_n}=q_n=1$
	$\sum_{i=1}^{n} a_i=A_n$				

れていない限り，一般に，

$$p_i>q_i \tag{2}$$

という関係がある。不平等な所得分布において，劣位の所得階級の構成員 25%(p_i)が社会における総所得の 10%(q_i)を領有しているという事例を想定すれば，(2)式の意味は明らかである。ここで，任意の分布において

$$p_i-q_i$$

という差を考えてみる。この差は，分布が不平等な場合には，

$$p_i-q_i>0$$

となり，平等な場合には

$$p_i-q_i=0$$

となる。

ジーニが関心を向けたのは，不平等な分布である。そのような分布において，ジーニは差 p_i-q_i と p_i の比率

$$R_i=\frac{p_i-q_i}{p_i} \tag{3}$$

を考えた（p.1206[5)]）。これは，p_i を基準にして２つの累積百分率の乖離 p_i-q_i を相対化して計測したものであり，R_i の値が大きいほど，p_i に較べて相対的に乖離 p_i-q_i が大きいことを示す指標となりうる。たとえば，$R_i=0$ の場合には，$p_i-q_i=0$ となり，これは均等分布（equidistribuzione）を示している。他方で，$R_i=1$ は一般に $q_i=0$ を意味し，これは「集中が完全（perfetta）であること」を示している，とジーニは指摘している（p.1207）。

$i=n$ のときには $p_n=q_n=1$ となり（表１参照），(3)式は分布の不平等性を計測するための実質的意味を失うので，２つの累積百分率の乖離としては，最初から数えて$(n-1)$個の R_i だけが意味をもつ。

ジーニはこの$(n-1)$個の R_i について，

5）Gini（1914）［脚注２］のページを示す。以下同様。

その「平均（media）」が，

$$R=\frac{\sum_{i=1}^{n-1}(p_i-q_i)}{\sum_{i=1}^{n-1}p_i} \tag{4}$$

であると言い，このRを「集中比(il rapporto di concentrazione)」と名づけた（p.1207)。

1914 年論文では，(4)式が $(n-1)$ 個ある R_i の「平均」であることについて，説明のないままに，(4)式から集中比にかんするさまざまな計算式が誘導されている。一般に，平均と言う場合には，相加平均や相乗平均などの計算的平均やメディアンやモードのような位置の平均が考えられるが，(4)式であたえられる R_i の「平均」はそのいずれとも異なり，この「平均」という用語には独特の意味が込められていると考えられる。以下ではこのことを考察して，ジーニの言う「平均」としての集中比の意味を敷衍する。

この考察の手がかりとなるのは，「平均」が系列の各項を同一の値で代替的（代表的）に表現できるという，いわば平均の代替機能（代表機能）である。系列 x_1，x_2，……，x_n があり，その相加平均を $\bar{x}$ とする。このとき，系列の各項 x_i は単一の $\bar{x}$ で代替することができる[6)]。

このことから類推すれば，(3)式で表現される $(n-1)$ 個の比率 R_i について，その系列

R_1，R_2，……，R_{n-1}

の各項を単一の値 R で代替（代表）させるという考え方の成り立つ余地がある。すなわち，

$$\left.\begin{array}{l} R_1=\dfrac{p_1-q_1}{p_1} \\ R_2=\dfrac{p_2-q_2}{p_2} \\ \quad\cdot \\ \quad\cdot \\ R_{n-1}=\dfrac{p_{n-1}-q_{n-1}}{p_{n-1}} \end{array}\right\} \tag{5}$$

について

$$R_1=R_2=\cdots\cdots=R_{n-1}=R \tag{6}$$

となるような R を考えることができる。(6)式を(5)式に代入すれば，次のようになる。

6）この代替機能は，増加率（p_1，p_2，……，p_n）の平均としての相乗平均についても指摘できる。原系列の構成要素の一般項を x_t で表すと（$t=1$，2，……，n)，隣合う 2 つの項の間の増加率は，$\frac{x_2}{x_1}=p_1$，$\frac{x_3}{x_2}=p_2$，…，$\frac{x_n}{x_{n-1}}=p_{n-1}$ となる。この増加率の平均とは，$t=1$ から $t=n$ までの n 個の項にかんする $(n-1)$ 個の増加率（p_1，p_2，……，p_{n-1}）がどれも等しくなるような p のことである。すなわち，$(n-1)$ 個の増加率 p_i にかんして

$$p_1=p_2=\cdots\cdots=p_{n-1}=p \tag{$*$}$$

となるような p が増加率の平均である。p をもとめるには，次のようにすればよい。

$$p_1\cdot p_2\cdot\cdots\cdot p_{n-1}=\frac{x_2}{x_1}\cdot\frac{x_3}{x_2}\cdot\cdots\cdot\frac{x_n}{x_{n-1}} \tag{$\sharp$}$$

（♯）式に（＊）式を代入すれば，

$$p^{n-1}=p_1\cdot p_2\cdot\cdots\cdot p_{n-1}$$

ゆえに，増加率の平均 p は

$$p=\sqrt[n-1]{p_1\cdot p_2\cdot\cdots\cdot p_{n-1}}$$

となる。

一般に，$(n-1)$ 個の増加率 p_i（ただし，原系列の項数は n）の平均 p は，$(n-1)$ 個の増加率 p_i の積 $\prod_{i=1}^{n-1}p_i$ の $(n-1)$ 乗根（相乗平均）p であたえられ，p_i の系列の各項は p で代替（代表）される。

増加率の平均が相乗平均であたえられるのは以上のような事情からであるが，どの増加率も同一の p と想定して平均をもとめる仕方は，ジーニが「集中比」をもって「平均」と規定したことと，その基調において符合する。

$$\left.\begin{aligned} R&=\frac{p_1-q_1}{p_1}\\ R&=\frac{p_2-q_2}{p_2}\\ &\cdot\\ &\cdot\\ R&=\frac{p_{n-1}-q_{n-1}}{p_{n-1}} \end{aligned}\right\} \tag{7}$$

(7)式を変形すれば，次式を得る。

$$\left.\begin{aligned} R\cdot p_1&=p_1-q_1\\ R\cdot p_2&=p_2-q_2\\ &\cdot\\ &\cdot\\ R\cdot p_{n-1}&=p_{n-1}-q_{n-1} \end{aligned}\right\} \tag{7'}$$

(7′)式における $(n-1)$ 本の数式について辺々加えれば，

$$\sum_{i=1}^{n-1} R\cdot p_i=\sum_{i=1}^{n-1}(p_i-q_i)$$

$$\therefore \quad R\sum_{i=1}^{n-1}p_i=\sum_{i=1}^{n-1}(p_i-q_i) \tag{8}$$

となる。(8)式を整理すれば

$$R=\frac{\sum_{i=1}^{n-1}(p_i-q_i)}{\sum_{i=1}^{n-1}p_i} \qquad (4)［再掲］$$

となり，ジーニによる集中比の定義式［(4)式］を誘導することができる。

(2) 定義式からの誘導

1914 年論文でジーニは(4)式から次の 2 式が誘導されると述べている（p.1208）。

$$R=1-\frac{2}{(n-1)A_n}\sum_{i=1}^{n-1}A_i \tag{9}$$

$$R=\frac{2\sum_{i=1}^{n}(i-1)a_i}{(n-1)A_n}-1 \tag{10}$$

ただし，これらの誘導にかんする叙述は 1914 年論文では詳しくないので，以下では独自に誘導を試みる。

①$R=1-\frac{2}{(n-1)A_n}\sum_{i=1}^{n-1}A_i$ の誘導

表 1 より，

$$\frac{i}{n}=p_i$$

$$\begin{aligned} \therefore \quad \sum_{i=1}^{n-1}p_i&=\sum_{i=1}^{n-1}\frac{i}{n}\\ &=\frac{1}{n}\sum_{i=1}^{n-1}i\\ &=\frac{1}{n}\left[\frac{1}{2}\{1+(n-1)\}\times(n-1)\right]\\ &=\frac{1}{n}\cdot\frac{1}{2}\cdot n\cdot(n-1)\\ &=\frac{1}{2}(n-1) \end{aligned} \tag{11}$$

同様に，表 1 から

$$\frac{A_i}{A_n}=q_i$$

$$\begin{aligned} \therefore \quad \sum_{i=1}^{n-1}q_i&=\sum_{i=1}^{n-1}\frac{A_i}{A_n}\\ &=\frac{1}{A_n}\sum_{i=1}^{n-1}A_i \end{aligned} \tag{12}$$

(11)式と(12)式を(4)式に代入すると，次式を得る。

$$\begin{aligned} R&=\frac{\sum_{i=1}^{n-1}(p_i-q_i)}{\sum_{i=1}^{n-1}p_i}\\ &=\frac{\sum_{i=1}^{n-1}p_i}{\sum_{i=1}^{n-1}p_i}-\frac{\sum_{i=1}^{n-1}q_i}{\sum_{i=1}^{n-1}p_i}\\ &=1-\frac{\frac{1}{A_n}\sum_{i=1}^{n-1}A_i}{\frac{1}{2}(n-1)}\\ &=1-\left(\frac{1}{A_n}\sum_{i=1}^{n-1}A_i\right)\times\frac{2}{n-1}\\ &=1-\frac{2}{(n-1)A_n}\sum_{i=1}^{n-1}A_i \qquad (9)［再掲］ \end{aligned}$$

以上により，(4)式から(9)式が誘導された。よって，

$$R = 1 - \frac{2}{(n-1)A_n}\sum_{i=1}^{n-1} A_i \qquad (9)［再掲］$$

$$②R = \frac{2\sum_{i=1}^{n}(i-1)a_i}{(n-1)A_n} - 1 \text{ の誘導}$$

(4)式から(10)式への誘導についても，ジーニはその結果だけを記載して，多くを語らない(p.1208)。そこで，(10)式を誘導するために，次の恒等式を措定する。

$$A_n \equiv A_n \tag{13}$$

(13)式の両辺に n を掛けても(13)式は恒等的に成立するので，

$$nA_n = nA_n$$

したがって，

$$nA_n = (n-1+1)A_n$$

これを変形すれば，

$$nA_n - A_n = (n-1)A_n \tag{14}$$

を得る。

$$A_n = \sum_{i=1}^{n} a_i \tag{15}$$

であるから（表 1 参照），(14)式は

$$n\sum_{i=1}^{n} a_i - \sum_{i=1}^{n} a_i = (n-1)A_n$$

$$\sum_{i=1}^{n} na_i - \sum_{i=1}^{n} a_i = (n-1)A_n$$

$$\sum_{i=1}^{n-1} na_i + na_n - \sum_{i=1}^{n} a_i = (n-1)A_n \tag{16}$$

と変形される。

ここで，以後の計算の便宜のために，$\sum_{i=1}^{n-1} ia_i$ と $\sum_{i=1}^{n} ia_i$ をもとめておく。

$$\sum_{i=1}^{n-1} ia_i = 1a_1 + 2a_2 + \cdots\cdots + (n-1)a_{n-1}$$

$$\sum_{i=1}^{n} ia_i = 1a_1 + 2a_2 + \cdots\cdots + (n-1)a_{n-1} + na_n$$

したがって，

$$-\sum_{i=1}^{n-1} ia_i + \sum_{i=1}^{n} ia_i = na_n \tag{17}$$

(17)式を(16)式の左辺に代入すると，

$$\sum_{i=1}^{n-1} na_i - \sum_{i=1}^{n-1} ia_i + \sum_{i=1}^{n} ia_i - \sum_{i=1}^{n} a_i = (n-1)A_n$$

となるので，結局，(16)式は次式となる。

$$\sum_{i=1}^{n-1}(n-i)a_i + \sum_{i=1}^{n}(i-1)a_i = (n-1)A_n \tag{18}$$

(18)式を整理する目的で $\sum_{i=1}^{n-1}(n-i)a_i$ をもとめる（表 1 参照）。

$$\begin{aligned}
&\sum_{i=1}^{n-1}(n-i)a_i \\
&= (n-1)a_1 + (n-2)a_2 + \cdots \\
&\quad + \{n-(n-2)\}a_{n-2} + \{n-(n-1)\}a_{n-1} \\
&= \quad a_1 \\
&\quad + (a_1 + a_2) \\
&\quad \cdot \\
&\quad \cdot \\
&\quad + (a_1 + a_2 + \cdots + a_{n-2}) \\
&\quad + (a_1 + a_2 + \cdots + a_{n-2} + a_{n-1}) \\
&= A_1 + A_2 + \cdots\cdots + A_{n-2} + A_{n-1} \\
&= \sum_{i=1}^{n-1} A_i
\end{aligned} \tag{19}$$

(19)式を(18)式に代入すると

$$\sum_{i=1}^{n-1} A_i + \sum_{i=1}^{n}(i-1)a_i = (n-1)A_n \tag{20}$$

(20)式の両辺に $\frac{2}{(n-1)A_n}$ を掛けて，整理すれば，次のようになる。

$$\left\{\sum_{i=1}^{n-1} A_i + \sum_{i=1}^{n}(i-1)a_i\right\} \times \frac{2}{(n-1)A_n} = (n-1)A_n \times \frac{2}{(n-1)A_n}$$

$$\frac{2}{(n-1)A_n}\left\{\sum_{i=1}^{n-1} A_i + \sum_{i=1}^{n}(i-1)a_i\right\} = 2$$

$$\frac{2}{(n-1)A_n}\sum_{i=1}^{n-1}A_i+\frac{2}{(n-1)A_n}\sum_{i=1}^{n}(i-1)a_i=1+1$$

$$\therefore\ \frac{2}{(n-1)A_n}\sum_{i=1}^{n-1}A_i-1=1-\frac{2}{(n-1)A_n}\sum_{i=1}^{n}(i-1)a_i \quad (21)$$

(21)式の両辺に－１を掛けると

$$1-\frac{2}{(n-1)A_n}\sum_{i=1}^{n-1}A_i=\frac{2}{(n-1)A_n}\sum_{i=1}^{n}(i-1)a_i-1 \quad (22)$$

を得る。(22)式の左辺は，すでに誘導した

$$R=1-\frac{2}{(n-1)A_n}\sum_{i=1}^{n-1}A_i \quad (9)\ [再掲]$$

である。したがって，(22)式の右辺は集中比 R に等しい。よって

$$R=\frac{2\sum_{i=1}^{n}(i-1)a_i}{(n-1)A_n}-1 \quad (10)\ [再掲]$$

が誘導された。

＊　　＊　　＊

以上から $R=\dfrac{\sum_{i=1}^{n-1}(p_i-q_i)}{\sum_{i=1}^{n-1}p_i}$ で定義される集中比 R は，データが表１のようにまとめられるときには，

$$1-\frac{2}{(n-1)A_n}\sum_{i=1}^{n-1}A_i=R=\frac{2}{(n-1)A_n}\sum_{i=1}^{n}(i-1)a_i-1$$

であることが証明される（(9)式・(10)式の証明終り）。

２．強度を同じくする個体が複数ある場合の集中比

(1)　計算式の誘導

強度を同じくする個体が複数個ある場合（$a_i \leqq a_j$，ただし，$i<j$ ［$i=1$，２，……，$n-1$；$j=2$，３，……，n］）には，集中比 R は別の計算式であたえられる。ただし，誘導される計算式それ自体は，定義式［(4)式］（より厳密にはそこから誘導された(10)式）のバリエーションである。前と同様に，用いる記号の意味を表としてあたえておく（表２）。それとともに，とくに次の点を確認する。

(a)　原系列を構成する個体の総数は n 個である。

(b)　個体の強度を一般に a で表し，その大小関係は $a_i \leqq a_j$ とする（$i<j$；$i_{\min}=1$，$j_{\max}=n$）。

(c)　強度を同じくする個体は，強度ごとに固有の度数 f をもつ。

(d)　強度ごとに個体を s 個のグループに分け，その強度を一般に x で表す。

(e)　グループ別の強度の増分を＋１とする。この想定は非現実的と考えられるかもしれないが，たとえば身長，体重，所得などの計測単位をそれぞれセンチメートル，キログラム，リラとすれば，とりうる値はいずれも増分を＋１とする整数であるから，強度の増分を＋１とすることは荒唐無稽ではない。

ジーニによれば，このような場合には，集中比 R は

$$R=\frac{1}{(n-1)A_n}\sum_{l=1}^{s}(i_{l-1}+i_l-1)f_l x_l-1 \quad (23)$$

であたえられる（p.1209）。

(23)式の誘導についてもジーニの叙述は簡潔なので，以下，表２を参照してこれを誘導したい。そのために，すでに誘導した

$$R=\frac{2\sum_{i=1}^{n}(i-1)a_i}{(n-1)A_n}-1 \quad (10)\ [再掲]$$

の右辺第１項の分子における $\sum_{i=1}^{n}(i-1)a_i$ に

表 2　記号一覧表（その 2）

強度番号	強度*	原系列**	同一の強度をもつ個体の数（度数）	累積度数	（簡略表記）
1	x_1	$a_1 \;=a_2=\cdots\cdots\cdots\cdots=a_{i_1}$	f_1	f_1	i_1
2	x_2	$a_{i_1+1}=\cdots\cdots\cdots\cdots\cdots=a_{i_2}$	f_2	f_1+f_2	i_2
・ ・	・ ・	・ ・	・ ・	・ ・	・ ・
$l-1$	x_{l-1}	$\cdots\cdots\cdots\cdots\cdots\cdots\cdots=a_{i_{l-1}}$	f_{l-1}	$f_1+f_2+\cdots+f_{l-1}$	i_{l-1}
l	x_l	$a_{i_{l-1}+1}=\cdots\cdots\cdots\cdots\cdots=a_{i_l}$	f_l	$f_1+f_2+\cdots+f_{l-1}+f_l$	i_l
$l+1$	x_{l+1}	$\cdots\cdots\cdots\cdots\cdots\cdots\cdots=a_{i_{l+1}}$	f_{l+1}	$f_1+f_2+\cdots\cdots\cdots+f_l+f_{l+1}$	i_{l+1}
・ ・	・ ・	・ ・	・ ・	・ ・	・ ・
s	x_s	$\cdots\cdots\cdots\cdots\cdots=a_{n-1}=a_n$	f_s	$f_1+f_2+\cdots\cdots\cdots+f_l+\cdots\cdots+f_s$	i_s

*強度 x の増分は＋１である（$x_l+1=x_{l+1}$）。
**系列を構成する項 a_1，a_2，……，a_n のうち，a_1 から a_{i_1} までは，その強度は x_1 に等しく，その項数（度数）は f_1 である。以下同様。

着目する。そして，これを表 2 のデータと整合させる。そうすると，次のようになる。

$$
\begin{aligned}
&\sum_{i=1}^{n}(i-1)a_i\\
&=(1-1)a_1+(2-1)a_2+\cdots\cdots\cdots\cdots\cdots\cdots+(i_1-1)a_{i_1}\\
&\quad+\{(i_1+1)-1\}a_{i_1+1}+\cdots\cdots\cdots\cdots\cdots\cdots+(i_2-1)a_{i_2}\\
&\quad+\cdots\cdots\cdots\\
&\quad+\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots+(i_{l-1}-1)a_{i_{l-1}}\\
&\quad+\{(i_{l-1}+1)-1\}a_{i_{l-1}+1}+\cdots\cdots\cdots\cdots\cdots+(i_l-1)a_{i_l}\\
&\quad+\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots+(i_{l+1}-1)a_{i_{l+1}}\\
&\quad+\cdots\cdots\cdots\\
&\quad+\cdots\cdots\cdots\cdots\cdots\cdots+\{(i_s-1)-1\}a_{n-1}+(i_s-1)a_n \qquad (24)
\end{aligned}
$$

表 2 から関連データを拾えば，次のようになる。

$a_1=a_2=\cdots\cdots=a_{i_1}=x_1$，その度数は f_1
・
・
$a_{i_{l-1}+1}=\cdots\cdots=a_{i_l}=x_l$，その度数は f_l
・
・
$\cdots\cdots\cdots=a_{n-1}=a_n=x_s$，その度数は f_s

したがって，(24)式右辺の一般項は次のように整理できる。

$$
\begin{aligned}
&\{(i_{l-1}+1)-1\}a_{i_{l-1}+1}+\cdots\cdots+(i_l-1)a_{i_l}\\
&=\{(i_{l-1}+1)-1\}x_l+\cdots\cdots\cdots+(i_l-1)x_l\\
&=\sum_{i=i_{l-1}+1}^{i_l}(i-1)x_l\\
&=x_l\times\sum_{i=i_{l-1}+1}^{i_l}(i-1)
\end{aligned}
$$

この一般項から，(24)式は次のように変形できる。

$$\sum_{i=1}^{n}(i-1)\,a_i=\sum_{l=1}^{s}x_l\left\{\sum_{i=i_{l-1}+1}^{i_l}(i-1)\right\} \qquad (24')$$

ただし，s は階級区分の個数（表 2 参照）。

この(24′)式を(10)式に代入すれば，次式を得る（p.1208）。

$$R=\frac{2\sum_{i=1}^{n}(i-1)a_i}{(n-1)A_n}-1 \qquad (10)［再掲］$$

$$=\frac{2}{(n-1)A_n}\sum_{l=1}^{s}x_l\left\{\sum_{i=i_{l-1}+1}^{i_l}(i-1)\right\}-1 \qquad (25)$$

ジーニはこの式を掲げた後に，

$$R=\frac{1}{(n-1)A_n}\sum_{l=1}^{s}(i_{l-1}+i_l-1)x_l f_l-1 \qquad (23)\ [再掲]$$

を誘導しているが（p.1208 f.），その叙述は簡潔すぎて分かりにくい。そこで，(25)式から(23)式を導くことにする。そのために，(25)式の $\sum_{i=i_{l-1}+1}^{i_l}(i-1)$ に着目する。

$$\sum_{i=i_{l-1}+1}^{i_l}(i-1)$$
$$=\{(i_{l-1}+1)-1\}+\{(i_{l-1}+2)-1\}+\cdots\cdots+(i_l-1) \qquad (26)$$

これは，初項を $\{(i_{l-1}+1)-1\}$，末項を (i_l-1) とする等差数列の和である。i は整数なので，この数列の公差は1である。ここで，表2において第 l 番目の強度 x_l をもつ原系列は $a_{i_{l-1}+1}$ から a_{i_l} までであり，その個数は度数 f_l であたえられていることを想起する。すなわち，$(i_{l-1}+1)$ 番目の項から i_l 番目の項までの個数は f_l である。したがって，(26)式の右辺であたえられる数列の和は

$$\sum_{i=i_{l-1}+1}^{i_l}(i-1)=\frac{1}{2}\ [\{(i_{l-1}+1)-1\}+(i_l-1)]\,f_l$$
$$=\frac{1}{2}\ (i_{l-1}+i_l-1)\,f_l \qquad (26')$$

となる。

この(26′)式を(25)式に代入して整理すれば，次のようになり，(23)式が誘導される。

$$R=\frac{2}{(n-1)A_n}\sum_{l=1}^{s}x_l\left\{\sum_{i=i_{l-1}+1}^{i_l}(i-1)\right\}-1$$
$$=\frac{2}{(n-1)A_n}\sum_{l=1}^{s}x_l\left\{\frac{1}{2}(i_{l-1}+i_l-1)f_l\right\}-1$$
$$=\frac{1}{(n-1)A_n}\sum_{l=1}^{s}(i_{l-1}+i_l-1)f_l x_l-1 \qquad (23)[再掲]$$

(2) 数値例

ジーニは誘導された(23)式に関連数値を代入して，アメリカ先住民（成人男子）の心拍数にかんする集中比 R を次のように5.88％と計算した（表3参照）。

$$R=\frac{4,533,688}{262\times16,343}-1=0.0588[=5.88\%]$$

ただし，　$n-1=262$

$A_n=16,343$

$$\sum_{i=1}^{s}(i_{l-1}+i_l-1)f_l x_l=4,533,688$$

そして，この結果をハルガ・オアシス（当時のエジプトの一地域）における原住民の心拍数にかんする集中比6.73％と比較している。心拍数にかんする集中比の比較が何を明らかにするかについて，ジーニは何事も語ってはいない。したがって，本稿ではその比較によって解明される医学的な特質について言及することは避ける。ここでは，集中比（いわゆるジーニ係数）が，今日でこそ，所得分布の集中度を計測するための統計的測度として不動の位置を占めてはいるが，この集中比は，所得分布に限らず，度数分布であたえられるデータにおける集中を計測するための統計的測度として構想されたことに注目する。

集中比が必ずしも所得分布に固有の測度ではないということについては，ジーニが（ガルバーニとともに），上に誘導した(23)式を活用して，後に1921年イタリア人口センサスにかんする全数集計結果と標本集計結果とを対照したこと[7]からも明らかである。この比較にもとづいて，系列の総平均のような概括的な測度によって代表性が確認された標本と言えども，集中比を基準にして対比してみれ

7）Gini, C. e Galvani, L., “Di una applicazione del metodo rappresentativo all’ultimo censimento italiano della popolazione(1° dicembre 1921),” *Annali di Statistica*, Serie VI, Vol.IV, 1929, p.15.

表 3　アメリカ先住民の心拍数にかんする集中比を計算するための基礎データ

x_l	f_l	$f_l x_l$	i_l	$i_{l-1}-1$	$i_l+i_{l-1}-1$	$(i_l+i_{l-1}-1)f_l x_l$
44	1	44	1	−1*	0	0
45	1	45	2	0	2	90
46						
47						
48	3	144	5	1	6	864
49	1	49	6	4	10	490
50	4	200	10	5	15	3,000
51	3	153	13	9	22	1,166
52	5	260	18	12	30	7,800
53	2	106	20	17	37	3,922
54	12	648	32	19	51	33,048
55	4	220	36	31	67	14,740
56	19	1,064	55	35	90	95,760
57	7	399	62	54	116	46,284
58	24	1,392	86	61	147	204,624
59	7	413	93	85	178	73,514
60	23	1,380	116	92	208	287,040
61	2	122	118	115	233	28,426
62	19	1,178	137	117	254	299,212
63	11	693	148	136	284	196,812
64	19	1,216	167	147	314	381,824
65	3	195	170	166	336	65,520
66	32	2,112	202	169	371	783,552
67	5	335	207	201	408	136,680
68	18	1,224	225	206	431	527,544
69	1	69	226	224	450	31,050
70	12	840	238	225	463	388,920
71	2	142	240	237	477	67,734
72	12	864	252	239	491	424,224
73	1	73	253	251	504	36,792
74	3	222	256	252	508	112,776
75	1	75	257	255	512	38,400
76	2	152	259	256	515	78,280
77						
78	3	234	262	258	520	121,680
79						
80	1	80	263	261	524	41,920
合計	263	16,343	−	−	−	4,533,688

(訳注) *$l=1$ のとき i_{l-1} は i_0 である。i_0 は実在しない強度までの累積度数なので，$i_0=0$。よって $l=1$ のとき $i_0-1=-1$。

(出所) Hrdlicka, A., *Physiological and Medical Observations among the Indians of Southwerstern United States and Northern Mexico*, Smithsonian Institution Bureau of American Ethnology, Washington, Government Printing Office, 1908, pp.348-371. ただし，引用は Gini, C., "Sulla misura della concentrazione e della variabilità dei caratteri," *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Anno accademico 1913-1914, Tomo LXXIII, Parte seconda, p.1210 による。

ば，代表標本とは見なしがたいと述べた。そして，標本の代表性という概念は，相対的であって，対照のためのハードルを網羅的にクリアするという，絶対的な意味での代表標本は得難いと結論した[8]。

3．強度別の度数と強度別の強度の総和があたえられている場合の集中比とその簡便式

(1)　計算式の誘導

①　計算式

(23)式にもとづいて，アメリカ先住民の心拍数にかんする集中比を上のように5.88%と計算した後，ジーニは，強度 a_i の度数分布が r 個に階級区分され，かつ階級別の度数 f_k と階級内の強度の合計 S_k があたえられている場合を取り上げた（系列を構成する個体の強度が個別にはあたえられていない点で表2とは異なる）。そして，そのようなときの集中比 R の計算式（とその簡便式）を示した（p.1212 f.）。しかし，この計算式の誘導にかんしてもその叙述は簡潔である。そこで，以下に私見を交えて計算式を誘導する。そのために，ここでも，関連する記号を以下で一覧する（表4）。なお，表内の記号を含めて，以下における表記の一部がジーニとは異なっていることを，あらかじめ断っておく。

表4では，第 k 階級の度数を f_k，またその階級に落ちている個体の強度の合計を S_k で表した。このとき，第 k 階級に属す個体の強度の相加平均は

$$\frac{S_k}{f_k}$$

である。以下では，今後の数式展開の必要か

表4　記号一覧表（その3）

階級番号	階級の限界値*（上段：下限）（下段：上限）	度数		階級別の強度の合計
		階級内度数	階級の上限(l_k)までの強度をもつ個体の数（累積度数）	
1	l_1' l_1	f_1	$i_1 \;= f_1$	S_1
2	l_2' l_2	f_2	$i_2 \;= f_1+f_2$	S_2
・ ・	・ ・	・ ・	・ ・	・ ・
$k-1$	l_{k-1}' l_{k-1}	f_{k-1}	$i_{k-1} = f_1+\cdots+f_{k-1}$	S_{k-1}
k	l_k' l_k	f_k	$i_k \;= f_1+\cdots+f_{k-1}+f_k$	S_k
$k+1$	l_{k+1}' l_{k+1}	f_{k+1}	$i_{k+1} = f_1+\cdots\cdots\cdots\cdots+f_k+f_{k+1}$	S_{k+1}
・ ・	・ ・	・ ・	・ ・	・ ・
r	l_r' l_r	f_r	$i_r \;= f_1+\cdots\cdots\cdots\cdots+f_k+\cdots\cdots\cdots\cdots+f_r$	S_r

（注記）*第1階級の下限 l_1' と第 r 階級の上限 l_r があたえられていなくても，集中比を計算することができる。

8）Gini, C., “Une application de la méthode représentative aux matériaux du dernier recensement de la population italienne (1er décembre 1921),” *Bulletin of the International Statistical Institute*, Vol.23, No.2, 1929. なお，このことについては，①木村和範『標本調査法の生成と展開』北海道大学図書刊行会，2001年，第6章；②同「ネイマンの標本調査理論とその周辺（上）」『経済論集』（北海学園大学）第50巻第3号，2002年。

ら，第 k 階級におけるこの相加平均 $\frac{S_k}{f_k}$ と個体の強度 ${}^k a_i$ との偏差 ${}^k\delta_i$ を考える（p. 1211)。この偏差の一般項は

$$ {}^k\delta_i = {}^k a_i - \frac{S_k}{f_k} \tag{27} $$

ただし，$i=1, 2, \cdots\cdots, f_k$

である。一般に，平均からの偏差は正負の値をとり得るが，正負の偏差は相殺されるために，その合計はゼロになる。したがって，第 k 階級における偏差 ${}^k\delta_i$ の総和は次のようになる。

$$ \sum_{i=1}^{f_k} {}^k\delta_i = 0 \tag{28} $$

次に，同じく今後の数式展開の必要から，以前に

$$ \sum_{i=i_{l-1}+1}^{i_l} (i-1) = \frac{1}{2}\ (i_{l-1}+i_l-1)\, f_l \quad (26')\,[再掲] $$

を導出したことを想起する。そして，この(26′)式を表 4 にあてはめる。そうすると，第 k 階級については次式を得る。

$$ \sum_{i=i_{k-1}+1}^{i_k} (i-1) = \frac{1}{2}\ (i_{k-1}+i_k-1)\, f_k $$

この両辺を，第 k 階級に属す個体の個数（度数）f_k で割る。このとき，

$$ \frac{\sum_{i=i_{k-1}+1}^{i_k} (i-1)}{f_k} = \frac{1}{2}\ (i_{k-1}+i_k-1) $$

を得る。上式の左辺は第 k 階級における f_k 個の（$i-1$）の合計をその個数で割ったときの商であるから，上式の右辺 $\frac{1}{2}(i_{k-1}+i_k-1)$ は，第 k 階級における $(i-1)$ の相加平均であると言うことができる。

この平均 $\frac{1}{2}(i_{k-1}+i_k-1)$ と当該階級にかんする $(i-1)$ との偏差を ${}^k\varepsilon_i$ とおく。第 k 階級における偏差 ${}^k\varepsilon_i$ は全部で f_k 個あり，その一般項は

$$ {}^k\varepsilon_i = (i-1) - \frac{1}{2}(i_{k-1}+i_k-1) \tag{29} $$

となる（p.1211)。

平均偏差の和はゼロなので，

$$ \sum_{i=1}^{f_k} {}^k\varepsilon_i = 0 \tag{30} $$

である。

${}^k\delta_i$ と ${}^k\varepsilon_i$ とについて，以上の準備をしておいて，データが表 4 となる場合の集中比 R をもとめることにする。そのために，第 k 階級における（$i-1$）と a_i の積をもとめる。

(27)式より ${}^k a_i = {}^k\delta_i + \frac{S_k}{f_k}$，また，(29)式より $i-1 = {}^k\varepsilon_i + \frac{1}{2}(i_{k-1}+i_k-1)$ を得るので，次のようになる。

$$ \begin{aligned} (i-1)a_i &= \left\{ {}^k\varepsilon_i + \frac{1}{2}(i_{k-1}+i_k-1) \right\} \times \left({}^k\delta_i + \frac{S_k}{f_k} \right) \\ &= {}^k\varepsilon_i \cdot {}^k\delta_i + {}^k\varepsilon_i \cdot \frac{S_k}{f_k} \\ &\quad + \frac{i_{k-1}+i_k-1}{2} \cdot {}^k\delta_i + \frac{i_{k-1}+i_k-1}{2} \cdot \frac{S_k}{f_k} \end{aligned} \tag{31} $$

この(31)式は，第 k 階級における積 $(i-1)\,a_i$ の一般項であるから，当該階級にかんする積の合計（積和）は次のようになる。

$$ \begin{aligned} \sum_{i=i_{k-1}+1}^{i_k} (i-1)a_i &= \sum_{i=1}^{f_k} {}^k\varepsilon_i \cdot {}^k\delta_i + \sum_{i=1}^{f_k} {}^k\varepsilon_i \cdot \frac{S_k}{f_k} \\ &\quad + \sum_{i=1}^{f_k} \frac{i_{k-1}+i_k-1}{2} \cdot {}^k\delta_i \\ &\quad + \sum_{i=1}^{f_k} \frac{i_{k-1}+i_k-1}{2} \cdot \frac{S_k}{f_k} \\ &= \sum_{i=1}^{f_k} {}^k\varepsilon_i \cdot {}^k\delta_i + f_k \left(\frac{S_k}{f_k} \sum_{i=1}^{f_k} {}^k\varepsilon_i \right) \\ &\quad + \frac{i_{k-1}+i_k-1}{2} \sum_{i=1}^{f_k} {}^k\delta_i \\ &\quad + f_k \left(\frac{i_{k-1}+i_k-1}{2} \cdot \frac{S_k}{f_k} \right) \end{aligned} $$

ところが，(30)式$[\sum_{i=1}^{f_k}{}^k\varepsilon_i=0]$と(28)式$[\sum_{i=1}^{f_k}{}^k\delta_i=0]$により，第$k$段階における積$(i-1)a_i$の一般項は

$$\sum_{i=i_{k-1}+1}^{i_k}(i-1)a_i=\sum_{i=1}^{f_k}{}^k\varepsilon_i\cdot{}^k\delta_i+f_k\left(\frac{i_{k-1}+i_k-1}{2}\cdot\frac{S_k}{f_k}\right)$$
$$=\sum_{i=1}^{f_k}{}^k\varepsilon_i{}^k\delta_i+\frac{i_{k-1}+i_k-1}{2}\cdot S_k \quad (32)$$

と整理される。

すでに，集中比Rとして，

$$R=\frac{2\sum_{i=1}^{n}(i-1)a_i}{(n-1)A_n}-1 \quad (10)［再掲］$$

を誘導した。表4の場合には，強度a_iがr個の階級に区分されているので，一般項(32)式を(10)式に代入すると，設例にたいする集中比Rの計算式として，次式を得る（p.1212）。

$$R=\frac{2\sum_{k=1}^{r}\sum_{i=1}^{f_k}{}^k\varepsilon_i{}^k\delta_i+\sum_{k=1}^{r}(i_{k-1}+i_k-1)S_k}{(n-1)A_n}-1 \quad (33)$$

② 簡便式

ジーニは，上述のようにして(33)式を誘導した。しかし，これにもとづいて集中比を計算するには，$\sum_{k=1}^{r}\sum_{i=1}^{f_k}{}^k\varepsilon_i{}^k\delta_i$をもとめる必要がある。この計算の煩を避ける目的で，ジーニは(33)式から$2\sum_{k=1}^{r}\sum_{i=1}^{f_k}{}^k\varepsilon_i{}^k\delta_i$を除去し，

$$R'=\frac{\sum_{k=1}^{r}(i_{k-1}+i_k-1)S_k}{(n-1)A_n}-1 \quad (34)$$

をもって，(33)式の簡便式とした（p.1213）。

さらにまた，強度の個数nが十分に大きい場合の簡便式のバリエーションとして

$$R'_2=\frac{\sum_{k=1}^{r}(i_{k-1}+i_k-1)S_k}{nA_n}-1 \quad (34')$$

をあたえている（p.1213）。十分に大きなnについては

$$\frac{1}{n-1}\fallingdotseq\frac{1}{n}$$

が成立するからである。

③ 簡便式による集中比R'の計算例（数値例）

以上で，データが表4であたえられる場合にたいする集中比Rの計算式［(33)式］とその簡便式［(34)式］を誘導した。1914年論文でジーニは簡便式による計算を例解しているので，以下でこのことについて述べる。なお，それに先だって，ジーニは上に記した(33)式と(34)式の分子に共通して置かれた$(i_{k-1}+i_k-1)$におけるiの順序を入れ替えていることを指摘しておく。計算例のためにあたえられているデータ（表5）はそのような簡便式に整合的な体裁となっているので，ジーニの所説を再現するという本稿の目的のためにも，以下にそれを掲げることとする。

$$R'=\frac{\sum_{k=1}^{r}(i_k+i_{k-1}-1)S_k}{(n-1)A_n}-1 \quad (34)［再掲］$$

以下に，表5のデータを(34)式にあてはめたときの計算結果を掲げる（p.1214）。

$$R'=\frac{4,525,631}{262\times16,343}-1=0.0569[=5.69\%]$$

(2) 簡便式の近似性

(33)式とその簡便式として誘導された(34)式の大小関係を比較する目的で，以下に(33)式を再掲するとともに，その変形式も掲げておく。

$$R=\frac{2\sum_{k=1}^{r}\sum_{i=1}^{f_k}{}^k\varepsilon_i{}^k\delta_i+\sum_{k=1}^{r}(i_{k-1}+i_k-1)S_k}{(n-1)A_n}-1 \quad (33)［再掲］$$
$$=\frac{\sum_{k=1}^{r}(i_{k-1}+i_k-1)S_k}{(n-1)A_n}+\frac{2\sum_{k=1}^{r}\sum_{i=1}^{f_k}{}^k\varepsilon_i{}^k\delta_i}{(n-1)A_n}-1$$

表 5　簡便式のためのデータ（アメリカ先住民の心拍数）

l_k’	l_k	f_k	S_k	i_k	$i_{k-1}-1$	$i_k+i_{k-1}-1$	$(i_k+i_{k-1}-1)S_k$
44	49	6	282	6	−1*	5	1,410
50	54	26	1,367	32	5	37	50,579
55	59	61	3,488	93	31	124	432,512
60	64	74	4,589	167	92	259	1,188,551
65	69	59	3,935	226	166	392	1,542,520
70	74	30	2,141	256	225	481	1,029,821
75	80	7	541	263	255	518	280,238
合計		263	16,343**	−	−	−	4,525,631

（訳注）*$k=1$ のとき i_{k-1} は i_0 となる。i_0 は実在しない累積度数なので，$i_0=0$ である。したがって，$i_0-1=-1$ となる。

**$\sum_{k=1}^{7} S_k=A_n$

（出所）表 3 に同じ。ただし，引用は，Gini, C., “Sulla misura della concentrazione e della variabilità dei caratteri,” *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Anno accademico 1913-1914, Tomo LXXIII, Parte seconda, 1914, p.1214 による。

ここに

$$\sum_{k=1}^{r}\sum_{i=1}^{f_k} {}^k\varepsilon_i {}^k\delta_i>0 \tag{35}$$

であり（p.1212），かつ $(n-1)A_n>0$ なので，(34)式と(33)式の大小関係は

$R'<R$

となる。したがって，簡便式は本来の集中比 R よりも小さい値をあたえることが分かる。(33)式と(34)式を対照させれば明らかなように，この 2 つの集中比の乖離 $R-R'$ は 2 つの平均偏差の積 ${}^k\varepsilon_i {}^k\delta_i$ に依存する。このことから，ジーニは次のように述べている（p.1213）（引用文の［　］内は本文における表記と対応する)。

差 $R-R'$ は $\frac{\varepsilon_{kl}}{n-1}\left[\frac{{}^k\varepsilon_i}{n-1}\right]$ および $\frac{\delta_{kl}}{A_n}\left[\frac{{}^k\delta_i}{A_n}\right]$ とともに大きくなる。$\frac{\varepsilon_{kl}}{n-1}\left[\frac{{}^k\varepsilon_i}{n-1}\right]$ の値は ── そして（他の条件が等しければ）$\frac{\delta_{kl}}{A_n}\left[\frac{{}^k\delta_i}{A_n}\right]$ の値もまた ── $\frac{f_k}{n}$ が大きくなるにつれて大きくなり，さらに $\frac{\delta_{kl}}{A_n}\left[\frac{{}^k\delta_i}{A_n}\right]$ の値は $\frac{f_k}{n}$ の値が等しいときには，a_i の値の差が大きくなるとともに大きくなる。したがって，差 $R-R'$ は，特性の集中の強化ならびに特性の強度をグループに分けるときの，階級の包括性（comprensività delle classi）の強化［階級区分を大まかにすること］とともに大きくなる，と言うことができる。

すなわち，ジーニによれば，乖離 $R-R'$ の大きさに影響をあたえるのは，①階級区分の個数と②系列における集中の度合いである。

乖離 $R-R'$ の大きさと①階級区分の個数との関連について，ジーニは，階級区分の個数 r を変えて，簡便式によってオーストラリア・ビクトリア州における土地所有の集中比 R' を計算し，r が大きくなるにつれて，R' が大きくなることを確認し，間接的にではあるが，r の増大による乖離 $R-R'$ の減衰[9] を検証した（表 6）。

次に，乖離 $R-R'$ の大きさと②系列にお

9）アメリカ先住民の心拍数にかんする 2 つの集中比を比較してみても同様のことが指摘できる。強

表 6　階級区分の個数 r と集中比 R'

土地面積の階級区分数 r	簡便式による集中比 R' (%)
30	69.0
25	68.9
20	68.8
15	68.4
10	67.8
8	66.5
6	65.8
5	64.6
4	58.1

（出所）*Statistical Register of the State of Victoria for Year 1911*, Melbourne, Part III, Production, p.10. ただし，引用は，Gini, C., “Sulla misura della concentrazione e della variabilità dei caratteri,” *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Anno accademico 1913-1914, Tomo LXXIII, Parte seconda, p.1215 による。

ける集中の度合いの関係については，次のようにすればジーニの考えを理解することができる。${}^k\delta_i$ は定義により，

$$ {}^k\delta_i = {}^k a_i - \frac{S_k}{f_k} \qquad (27)［再掲］$$

であるから，${}^k\delta_i$ の絶対値は，強度 ${}^k a_i$ がその相加平均 $\frac{S_k}{f_k}$ に近い値となるような個体であるほど小さく，またそのような個体の個数が多ければ多いほど，$\sum_{k=1}^{r}\sum_{i=1}^{f_k}{}^k\varepsilon_i{}^k\delta_i$ の値は小さくなる。逆に，相加平均 $\frac{S_k}{f_k}$ から強度が大きく乖離した個体ほど，${}^k\delta_i$ の絶対値は大きくなり，そのような個体の増加に応じて $\sum_{k=1}^{r}\sum_{i=1}^{f_k}{}^k\varepsilon_i{}^k\delta_i$ の値が大きくなる。したがって，系列を構成する各項の強度がその相加平均に近づくほど，

$$ \frac{2\sum_{k=1}^{r}\sum_{i=1}^{f_k}{}^k\varepsilon_i{}^k\delta_i}{(n-1)A_n} $$

は小さくなり，これが小さくなった分だけ(33)式と(34)式の差は小さくなる。すなわち，乖離 $R-R'$ は小さくなる。

4．強度 $\boldsymbol{a_i}$ がゼロとなる個体の個数を系列の項数に算入させる場合の集中比（$\boldsymbol{R_t}$）と算入させない場合の集中比（$\boldsymbol{R_p}$）

ジーニは，その強度が正の値であたえられる個体を「正の事例（casi positivi）」と名づけた。他方で，強度がゼロとなる個体を「ゼロの事例（casi nulli）」と名づけた。そして，これら 2 種類の個体が，系列の全体を構成することから，そのすべての項を「全体事例（casi totali）」と言っている（p.1227）。ここで取り扱う 2 種類の集中比をサフィックス（p[ositivo]，t[otale]）で識別して，R_p（「正の事例」だけの集中比）と R_t（「全体事例」の集中比）としたのは，そのような意味からであろう。

(1)　強度 a_i がゼロとなる個体の個数を系列の項数に算入させない場合（「正の事例」だけ）の集中比（R_p）

「正の事例」のみで系列が構成されるときの集中比 R_p は，表 2 の場合に誘導した集中比と同一である（p.1228）。すなわち，

$$ R_p = \frac{1}{(n-1)A_n}\sum_{l=1}^{s}(i_{l-1}+i_l-1)f_l x_l - 1 \qquad (23)［再掲］$$

である。

度を 7 つに階級区分した表 5 では集中比が 0.0569 であったのにたいして，原系列の強度そのものを用いた表 3 では集中比は 0.0588 となった。階級区分が大まかなほど（ジーニのいわゆる「包括性」が強まるほど），小さな値の集中比が得られる。

⑵　強度 a_i がゼロとなる個体の個数を系列の項数に算入させる場合（「全体事例」）の集中比（R_t）

ジーニは㉓式を

$$R_p=\frac{\sum_{l=1}^{s}(i_{l-1}+i_l-1)f_lx_l-(n-1)A_n}{(n-1)A_n} \tag{36}$$

と変形し，この式にもとづいて R_t を

$$R_t=\frac{\sum_{l=1}^{s}(i_{l-1}+v+i_l+v-1)f_lx_l-(n+v-1)A_n}{(n+v-1)A_n} \tag{37}$$

ただし，n は「正の事例」の個数，
v は「ゼロの事例」の個数

と規定している（p.1228)。表 2 において，強度番号の如何にかかわらず，「ゼロの事例」の個数 v を定数と見なせば，累積度数の簡略表記 i_{l-1} は $i_{l-1}+v$ となり，また i_l は i_l+v となるからである。

㊲式を整理すれば，次のようになる。

$$\begin{aligned}
R_t&=\frac{\sum_{l=1}^{s}(i_{l-1}+v+i_l+v-1)f_lx_l-(n+v-1)A_n}{(n+v-1)A_n}\\
&=\frac{\sum_{l=1}^{s}\{(i_{l-1}+v)+(i_l+v)-1\}f_lx_l-\{(n+v)-1\}A_n}{\{(n+v)-1\}A_n}\\
&=\frac{\sum_{l=1}^{s}\{(i_{l-1}+i_l-1)+2v\}f_lx_l-\{(n+v)-1\}A_n}{\{(n+v)-1\}A_n}\\
&=\frac{\left\{\sum_{l=1}^{s}(i_{l-1}+i_l-1)f_lx_l+2v\sum_{l=1}^{s}f_lx_l\right\}-\{(n-1)+v\}A_n}{\{(n+v)-1\}A_n}\\
&=\frac{\sum_{l=1}^{s}(i_{l-1}+i_l-1)f_lx_l-(n-1)A_n+2v\sum_{l=1}^{s}f_lx_l-vA_n}{\{(n+v)-1\}A_n}\\
&=\frac{\sum_{l=1}^{s}(i_{l-1}+i_l-1)f_lx_l-(n-1)A_n}{(n+v-1)A_n}\times\frac{n-1}{n-1}+\frac{2v\sum_{l=1}^{s}f_lx_l-vA_n}{(n+v-1)A_n}\\
&=\frac{n-1}{n+v-1}\times\frac{\sum_{l=1}^{s}(i_{l-1}+i_l-1)f_lx_l-(n-1)A_n}{(n-1)A_n}+\frac{2v\sum_{l=1}^{s}f_lx_l-vA_n}{(n+v-1)A_n}
\end{aligned} \tag{38}$$

A_n はすべての個体の強度の総計であるから（表 1 参照），

$$A_n=\sum_{l=1}^{s}f_lx_l \tag{39}$$

である。また，

$$R_p=\frac{\sum_{l=1}^{s}\{(i_{l-1}+i_l-1)f_lx_l-(n-1)A_n}{(n-1)A_n} \tag*{(36)[再掲]}$$

である。

㊴式と㊱式を㊳式に代入すると，

$$\begin{aligned}
R_t&=\frac{n-1}{n+v-1}\times R_p+\frac{2vA_n-vA_n}{(n+v-1)A_n}\\
&=\frac{n-1}{n+v-1}R_p+\frac{v}{n+v-1}
\end{aligned} \tag{38'}$$

を得る。

ここで

$$m=\frac{n-1}{n+v-1} \tag{40}$$

とおくと，

$$\begin{aligned}
1-m&=1-\frac{n-1}{n+v-1}\\
&=\frac{(n+v-1)-(n-1)}{n+v-1}\\
&=\frac{v}{n+v-1}
\end{aligned} \tag{41}$$

となる。㊵式と㊶式を (38′) 式に代入すると，ジーニのいわゆる「ゼロの事例」を算入した「全体事例」の集中比 R_t は，次のようになる（p.1227)。

$$R_t=mR_p+(1-m) \tag{42}$$

⑶　R_t と R_p の乖離にかんする数学的関係

ジーニは R_t と R_p を掲げた後，「つねに $R_t>R_p$ であり，m と R_p の減少とともに，この差は増大する」と述べている（p.1228)。以下ではこのことを敷衍する。そのために 2

つの集中比の差をとることにする。(42)式からR_pを引けば，

$$\begin{aligned}R_t-R_p&=\{mR_p+(1-m)\}-R_p\\&=R_p(m-1)+(1-m)\\&=-R_p(1-m)+(1-m)\\&=(1-m)(1-R_p)\end{aligned} \tag{43}$$

ここに，

$1-m>0$　かつ　$1-R_p>0$

であるから，(43)式は

$R_t-R_p>0$

となる。

R_pは強度a_iをゼロとする個体（いわゆる「ゼロの事例」）を除いて計算される（すなわち「正の事例」だけから計算される）集中比であった。このR_pが計算された元の任意の系列にとって，その値は所与と見なすことができるので，２つの集中比の差R_t-R_pについては，さらに次のように指摘することができる。$(1-R_p)$をさしあたり系列に固有の定数cと見なせば，(43)式は

$$R_t-R_p=c(1-m) \tag{44}$$

とおくことができる。このようにすると，上に引用したように，ジーニが$m(>0)$の減少とともに，乖離R_t-R_pが増大すると述べたことの意味がさらに明瞭となる。

定義により，

$$m=\frac{n-1}{n+v-1} \quad (40)［再掲］$$

なので，「ゼロの事例」の個数vが多いほど，mが小さくなり，その結果$(1-m)$が大きくなって，結局，(44)式の値は大きくなるからである。

ジーニは，離婚した夫婦の子ども数（ブダペスト，1903年－08年）について子どもがいない夫婦（子ども数という強度がゼロの夫婦）を算入したときとそうでないときの２つの場合について，集中比を計算した。その他にも，資産をもたない個人（資産という強度がゼロの個人）を勘案した場合とそうでない場合などについての集中比を比較して，その所説の妥当性を検証している（ｐ.1228）。

む　す　び

ジーニの1914年論文に掲載されている数式を点に見立てて，その点と点を線で結ぶべく，数式展開を敷衍してきた。その結果，次のように要約することができる。

(1)　さしあたり，どの個体の強度も異なる場合を想定することにするが，その場合，強度を最小とする個体から順に数えてi個の強度の合計$\sum_{k=1}^{i}a_k=A_i$をもとめ，それが系列全体の強度の総計A_nのなかにしめる割合（強度比率）$\frac{A_i}{A_n}$をq_iとする。他方で，強度を最小とする個体から順に数えたときの，第i番目までの個体数（i）が個体総数（n）のなかにしめる割合（個体比率）$\frac{i}{n}$をp_iとする。

このとき，$p_i-q_i=0$は，強度比率と個体比率が等しいことを意味する。たとえば，個体を世帯，強度を所得とすると，$p_i-q_i=0$（すなわち$p_i=q_i$）は世帯の割合と所得の割合が等しいこと，換言すれば，所得分布が完全に均等であることを意味する。他方で，所得分布について一般に見られるように$p_i>q_i$の場合には，所得分布が不均等であることを意味する。そして，乖離p_i-q_iが大きいほど，所得分布の集中度が強化されていることになる。ジーニはこの乖離p_i-q_iをp_iで除し（$\frac{p_i-q_i}{p_i}$），p_iを基準にして，p_i-q_iを相対的に計測しよう

とした。

(2)　ジーニはこの（$n-1$）個の $\frac{p_i-q_i}{p_i}$ について，その「平均」を $\frac{\sum_{i=1}^{n-1}(p_i-q_i)}{\sum_{i=1}^{n-1}p_i}$ とし，これを「集中比」R と名づけた。このとき，ジーニの「平均」という言葉には特殊な意味が込められたと考えられる。

(3)　ジーニは

$$R=\frac{\sum_{i=1}^{n-1}(p_i-q_i)}{\sum_{i=1}^{n-1}p_i}$$

から，

$$R=1-\frac{2}{(n-1)A_n}\sum_{i=1}^{n-1}A_i$$

および

$$R=\frac{2\sum_{i=1}^{n}(i-1)a_i}{(n-1)A_n}-1$$
$$=\frac{2}{(n-1)A_n}\sum_{i=1}^{n-1}(i-1)a_i-1$$

の 2 式を誘導した（記号の意味については本文参照。以下同様)。

(4)　強度 a_i が増分を 1 とする整数 x_l であたえられ，かつ s 個に階級区分された強度ごとの度数が f_l であるときの集中比 R は

$$R=\frac{1}{(n-1)A_n}\sum_{l=1}^{s}(i_{l-1}+i_l-1)f_lx_l-1$$

であたえられる。ジーニは（ガルバーニとともに）代表標本の代表性を検討したときに，この計算式を活用した。

(5)　強度 a_i の個々の値はあたえられていないが，r 個の階級ごとにその度数 f_k と強度の合計 S_k があたえられているときの集中比 R とその簡便式があたえる近似値 R' は，それぞれ

$$R=\frac{2\sum_{k=1}^{r}\sum_{i=1}^{f_k}{}^k\varepsilon_i{}^k\delta_i+\sum_{k=1}^{r}(i_{k-1}+i_k-1)S_k}{(n-1)A_n}-1$$

$$R'=\frac{\sum_{k=1}^{r}(i_{k-1}+i_k-1)S_k}{(n-1)A_n}-1$$
$$=\frac{1}{(n-1)A_n}\sum_{k=1}^{r}(i_{k-1}+i_k-1)S_k-1$$

である。

R と R' には $R'<R$（近似値が実際よりも小さい値となる）の関係がある。より R に近い近似値 R' が得られるための条件は，次の 2 つが同時に，もしくはいずれか一方が成立していることである。すなわち，

①　系列を構成する項の強度がより均一であること。

②　階級区分の数がより多いこと。

この 2 つのうち，①は系列のもつ性質によって規定されるので，作為の余地はない。しかし，階級区分の数を増やすことについては工夫の可能性があり，簡便式による集中比の近似度の向上を期待することができる。

(6)　強度がゼロとなる v 個の個体（「ゼロの事例」）を項数に入れて計算したときの集中比 R_t とそうではなく n 個の「正の事例」だけで計算した集中比 R_p は次のようになる（s は階級区分の数)。

$$R_t=\frac{1}{(n-1)A_n}\sum_{l=1}^{s}(i_{l-1}+i_l-1)f_lx_l-1$$

$$R_p=\frac{1}{\{(n+v)-1\}A_n}\sum_{l=1}^{s}\{(i_{l-1}+v)+(i_l+v)-1\}f_lx_l-1$$

(7)　R_t と R_p の間には，

$$R_t=mR_p+(1-m)$$
$$\text{ただし，}m=\frac{n-1}{n+v+1}$$

という関係がある。また，

$$R_t - R_p = (1 - m)(1 - R_p) > 0$$

である。このことから，強度をゼロとする個体が多いほど，２つの集中比の乖離は大きくなることが分る。

以上に要約されるジーニの所説は，1914年論文における論点をおおまかに３つに絞り込んだときの第１論点である。残された２つの論点（後述）についての考察は別稿で行うこととして，さしあたり，ここでは次のことを確認する。

ジーニは，分布の特性を統計的に測定する目的で，「集中指数（いわゆるジーニ指数）」を考案した（1909年）。ところが，この「指数」は，$P = Q^{\delta}$ という関数で表される分布のパラメータ δ のことであり，δ の計算はこの関数関係を前提とする。

関数関係を前提としなければ，分布における集中度を計測できないとなると，その尺度として集中指数の応用は限定されてしまう。このことのうちに，集中指数に付帯する制約から逃れて集中度の計測を可能ならしめる測度を考案させるにいたった根拠を見ることができる。そして，ジーニは，「特性の分布曲線（curva di distribuzione del carattere）から独立していること」という要件を満たして，「特性の［分布の］集中にかんする比較が可能であること」という目的を達成する測度として「集中比」を定式化した（1914年）。集中比 R は，$P = Q^{\delta}$ はもとより，分布にたいして一切の関数関係を要請しない。こうして，集中比によって，強度の分布が何らかの関数型で表現されない場合でも，集中度は計測可能となった。ガエターノ・ピエトラは関数関係を前提とすることなく，集中度を計測できるということのうちに，集中比の特質を見出している[10)]。

この集中比は，後に「ジーニ係数」として所得分布の不平等度を計測する指標となり，今日に至っている。この計測にあっては，一般に，世帯と所得のそれぞれにかんする累積百分率が用いられている。このことから，第１に，ジーニ係数（集中比）の計算には２つの累積百分率（個体比率と強度比率）があたえられていること，そして第２には，ジーニ係数は所得分布の計測指標として考案されたことが，いわば「常識」となっているかの観を否むことができない。しかし，集中比が初めて定式化されたジーニの1914年論文を見る限り，次のように述べることができる。すなわち，第１に，集中比は個体数と強度にかんする２つの累積百分率があたえられているときはもとより，そうでない場合にも計算可能であること，第２には，集中比は所得分布の集中度の計測のみを目的とするものではなく，その他の度数分布にたいしても応用可能であること，がそれである。

以上に述べたほかに，ジーニは1914年論文で

①　集中比とローレンツ曲線との関係

②　集中比と平均差の関係

が考察されている。これらの検討は今後の課題である。

10）Pietra, Gaetano, “Delle relazioni tra gli indici di variabilità,” *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Anno accademico 1914‑15, Tomo LXXIV, Parte seconda, p.776.