

タイトル	ジニの平均分類
著者	木村, 和範
引用	季刊北海学園大学経済論集, 58(3): 59-74
発行日	2010-12-31

## 《研究ノート》

## ジニの平均分類

木 村 和 範

もくじ

はじめに

1. 内部性の要請と平均
  2. 解析的平均と非解析的平均（位置上の平均）
  3. 単純解析的平均と加重解析的平均
  4. 指数形式から見た解析的平均  
— 底平均，指数平均，底-指数平均 —
  5. 計算式の一般性と個別性  
— 個別的平均と包括的平均 —
  6. 一価的平均と多価的平均
  7. 現実的平均（実質的平均）と虚構の平均（計算的平均）
  8. 2つの虚構の平均
- おわりに

## はじめに

ジニは『平均論』（ミラノ，1958年）<sup>1)</sup>において、さまざまな基準を設けて「平均」を分類している。しかしながら、その著書が「数学の観点」<sup>2)</sup>から執筆されたせいであろうか、平均の分類は並列的に述べられていて、分類基準の階層性にかんする言及がない。旧稿<sup>2)</sup>は（したがって、それらを再録した拙著<sup>3)</sup>もまた）ジニ理論の解説を旨としており、平均分類についてはその祖述にとどまっている。分類の階層性にかんする検討が、今後の課題として

残されていた。本稿は、この欠陥を少しでも補い、ジニの平均概念への理解を深めることを目的とする。

検討の目的を以上のように措定している関係上、以下の叙述には旧稿の内容と重なる部分があることをあらかじめ断っておく。

## 1. 内部性の要請と平均

実数の系列

$$x_1, x_2, \dots, x_n$$

$$\text{ただし, } x_1 \leq x_2 \leq \dots \leq x_n$$

からは、相加平均，相乗平均，調和平均などの他に、位置上の平均とも言われるメディアン（中央値），モード（最頻値）など、さまざまな平均が計算される。

たとえば、相加平均  $A$  は系列を構成する諸項の値を1つの値で代表（代替）している。この意味で相加平均は、元の系列にたいして

- 
- 2) ①木村和範「平均概念について——ジニ『平均論』（ミラノ，1958年）断章——」『経済論集』（北海学園大学）第56巻第3号，2008年12月（以下，木村（2008）と略記）；②同「比例関係と平均——ジニ『平均論』（ミラノ，1958年）序章を中心に——」同上，第57巻第1号，2009年6月（以下，木村（2009a）と略記）；③同「解析的平均と内部性の要請——ジニ『平均論』（ミラノ，1958年）によせて——」同上，第57巻第2号，2009年9月（以下，木村（2009b）と略記）。
- 3) 木村和範『ジニの統計理論』共同文化社，2010年（以下，木村（2010）と略記）。

1) Gini, Corrado, *Le Medie*, Milano 1958. とくに断らない限り，本文中の（ ）内数字は，この著書の頁を示す。

代表機能(代替機能)を果たしている。換言すれば、相加平均を得るための数学的操作によって、系列においてさまざまな値をとりうる各項を「ならす」(平す, 均す) 数値が得られるのである。このことを踏まえて、O. キズィーニは平均を次のように規定している<sup>4)</sup>。

「同質の量を表す任意の  $n$  個の従属変数

$x_1, x_2, \dots, x_n$  について関数

$$y=f(x_1, x_2, \dots, x_n)$$

があたえられるとき、関数  $f$  にかんする平均とは、 $x_1, x_2, \dots, x_n$  の代わりになって、その関数とまったく同一の値をあたえる数  $M$ 、すなわち

$$f(M, M, \dots, M)=f(x_1, x_2, \dots, x_n) \quad (1)$$

となるような数  $M$  のことを言う。」

このようなキズィーニの定義に整合的な平均の代表格は相加平均である。ところが、2つの実数  $a_1$  と  $a_2$  にかんする反調和平均  $AH$  の公式

$$AH=\frac{a_1^2+a_2^2}{a_1+a_2}$$

に  $a_1=+4$ ,  $a_2=-3$  を代入すると、 $AH$  の値は 25 となる。この値は、平均にかんするキズィーニの定義式 ((1)式) を満たしている。しかし、ジニはこれを平均とは見なしていない。ジニによれば、(1)式を満たす数は、一般に、「平衡数 (adeguati numerici)」(直訳すれば「数的平衡」と言われている。その平衡数のなかでとくに「内部性の要請 (il requisito della internalità)」を満たす値だけが平均と規定されている (57)。ここに、内部性の要請とは、平均が、系列を構成する諸項のなかで ①最小の値の項よりも小さくなく、②最大の値の項よりも大きくないという条件 (平均は系列の最小値と最大値の間にあるという条件) を満たしていなければならないとする制約のこと

4) Chisini, Oscar, "Sul concetto di media," *Periodico di Matematiche*, Serie IV, Volume IX, N. 2, 10 marzo 1929, p.108. ただし、式番号は引用者による。

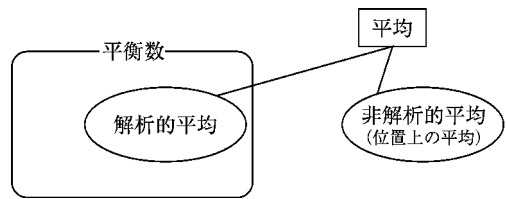


図1 解析的平均と非解析的平均

である。上の反調和平均を例にすれば、計算から得られる 25 は最大値  $+4 (=a_1)$  よりも大きく、内部性の要請を満たしていないので、平衡数ではあっても (反調和) 平均ではないことになる<sup>5)</sup>。

内部性の要請を満たす平衡数としての平均は、ジニの用語法に従えば、厳密には、解析的平均と言われる。この他に、ジニは、平衡数ではないが、内部性の要請を満たすという意味での平均として非解析的平均 (位置上の平均) があると述べている (図1)。これについては次項で取り上げる。

そして、ここでは、ひるがえって平均の概念規定に立ち返ることにする。キズィーニは、系列を構成する諸項の値を平し (均し) た、

5) ここでは、相加平均  $\bar{x}$  を例にして、 $\bar{x}$  が内部性の要請を満たしていることを示す。

実数の系列

$$x_1 \leq x_2 \leq \dots \leq x_n$$

においては

$$n \cdot x_1 \leq x_1 + x_2 + \dots + x_n \leq n \cdot x_n$$

が成立する。上式を  $n$  で割ると、次式を得る。

$$x_1 \leq \frac{x_1 + x_2 + \dots + x_n}{n} \leq x_n \quad (*)$$

ここに、

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

であるから、(\*) 式は

$$x_1 \leq \bar{x} \leq x_n$$

となり、相加平均  $\bar{x}$  は系列の最小値  $x_1$  と最大値  $x_n$  の間に落ちることが示される。よって、相加平均  $\bar{x}$  は、単なる平衡数ではなく、内部性の要請を満たす平均であることが証明される。なお、相加平均を含む 15 種類の解析的平均が内部性の要請を満たすことにかんするジニの証明については、木村 (2010), p.179 以下参照。

ある値を平均とみなしていることは、すでに述べた。この意味では、彼の平均概念にふさわしい英語は average であり、それは傷物を意味するアラビア語 awāriya に由来する。語源的な説明によれば、傷物という言葉から、海難により傷物が生じたときの、その責任に応じて損害を割当てるという意味が派生した。さらに、そこからその損害賠償を割り当てるための計算操作(平均操作)という意味が派生したと考えられている。なお、損害保険の分野では、航海で生じた損害を海損と言い、average はその英語である。

これにたいして、内部性の要請を満たす数値をもって平均とみなすジニの平均概念は、古代ギリシア(とりわけピュタゴラス学派)の数学理論に淵源する。それにもとづくと、ジニの平均に該当する英語は、ラテン語 medianus(中間にあるもの)を語源とする mean になる。『平均論』が執筆された頃にも、依然として影響力があったコーシーの考え方をジニは踏襲して、平均概念を規定した。ただし、無批判的にコーシーを踏襲したのではなく、ジニはみずからピュタゴラス学派による比例関係の分析を研究し、それを拡充するなかで、コーシーの見解の妥当性を確信した。そして、平均は内部性の要請を満たすと規定した。

ジニの本格的な研究生活は出生性比の統計分析に始まった<sup>6)</sup>。その後、分布尺度に研究の重点を移し、その成果を標本調査法の有効性をめぐる検討に応用した。一連の研究は、広く社会研究に属す。このような研究歴を勘案すれば、平均概念にかんする上記の規定には、彼の社会研究が何らかの影響をあたえていると考えることも可能である。しかし、『平均論』にはその内容を推測させる陽表的

な叙述を見出すことはできない。

## 2. 解析的平均と非解析的平均 (位置上の平均)

「数学の観点」からの平均にかんする考察を目的としたジニ『平均論』の力点は、解析的平均の数学的性質の解明に置かれている。このために、平均を、解析的平均と非解析的平均(位置上の平均)とに、対照的に二分する考え方がジニの立論の根底にあると考えても、それは合理的である。

ジニの解析的平均はいわゆる計算的平均に該当する。しかし、後に言及するように、『平均論』では「計算的平均」という用語は通常の用法とは異った独特の使われ方をしている。そこで、叙述の曖昧さを避けるために、以後は、この国の統計学界で慣れ親しんだ意味では計算的平均という用語を使用しないことにする。そして、本稿ではもっぱらジニの意味で用いる。

ジニによれば、解析的平均は、基本的には

$$M = f(x_1, x_2, \dots, x_n) \quad (2)$$

という関数関係を満たす  $M$  である(64)。この点だけを見れば、キズィーニの定義と変わるところがない。しかし、あらゆる平均は内部性の要請を満たすことを要し、単に(2)式を満たすだけでは、 $M$  は平衡数にすぎないとジニは考えた。このことは、すでに述べた。換言すれば、解析的平均は、(2)式において内部性の要請を満たす平衡数  $M$  のことである。ジニによれば、解析的平均は、この制約のもとで(2)式のような関数関係によって表現されることをその本旨とする。

これにたいして、非解析的平均は、系列を構成するすべての項の値にかんする解析的関数(funzioni analitiche)にはなじまない、とジニは言っている。そして、それは、系列における項の位置に依存し、その項は、考察の対象とした諸項の一部(もしくは全部)につい

6) Gini, C., *Il sesso dal punto di vista statistico*, Milano, Palermo e Napoli 1908. これはジニの学位論文である。

表1 さまざまな非解析的平均(位置上の平均)

非解析的平均	意味
①メディアン	総度数を1/2に分割する変量の値
②モード	度数が最大となる変量の値
③反モード	度数が最小となる変量の値
④優先値	(変量の値)×(その度数)が最大となる変量の値
⑤中央項	変量の最大値と最小値の中央に位置する変量の値(両端項の総和の半分とも言う)
⑥四分位数	総度数を1/4に分割する変量の値(3個)
⑦十分位数	総度数を1/10に分割する変量の値(9個)
⑧百分位数	総度数を1/100に分割する変量の値(99個)
⑨多分位数	総度数を $m$ 個に等分する変量の値( $(m-1)$ 個)
⑩2等分値	ある変量を境にして、その左右についてもとめられる(変量の値)×(その度数)を等しくする変量の値
⑪任意区分値	第1の部分の(変量の値)×(その度数)の和を、それに後続する第2の部分の(変量の値)×(その度数)の和の $k$ 倍にする変量の値

注記) 変量は昇順に配列されているものとする。

出所) Gini, C., *Le Medie*, Milano 1958, p. 102ff. の叙述から作成。

ての位置関係にもとづいて定まるとされている(102)。この規定は分かりにくい。そこで、ジニが掲げた非解析的平均の具体例から、その特性を考察する。

ジニは非解析的平均を位置上の平均とも言っている。ここに言う位置上の平均は、現在もわれわれが使用している用語と同一の意味内容をもっている。ジニがその例として挙げているのは、①メディアン(中央値)(*mediana*)、②モード(最頻値)(*moda*)、③反モード(*antimodo*)、④優先値(*valore poziore*)、⑤中央項(*termine centrale*)、⑥四分位数(*quartili*)、⑦十分位数(*decili*)、⑧百分位数(*centili*)、⑨多分位数(*quantili*)、⑩2等分値(*valore divisorio*)、⑪任意区分値(*tantili*)である(102ff.)<sup>(補註)</sup>。これらの平均とその数理的意味をまとめた表を上に掲げる(表1)。

この一覧表を概観すると、非解析的平均(位置上の平均)には、①変量の値を度数(階級区分された度数を含む)と関係づけることによって得られる非解析的平均と②対をなす変

量と度数のうち、変量だけに着目し、その変量の範囲から特定される非解析的平均との2種類があることが分かる。ただし、後者の非解析的平均(変量の範囲だけから特定される非解析的平均)は、⑤中央項(表1)のみである<sup>7)</sup>。

本稿では、中央項を除く10種類の非解析的平均(位置上の平均)を第1カテゴリーに分類する。そして、中央項を第2カテゴリーの非解析的平均(位置上の平均)と名づけることにする。なお、これらの非解析的平均は変量の領域内に落ちるのであるから、すべてが内部性の要請を満たしていることは明らかである。

この項の最後に、非解析的平均を関数関係で表現することはできないとジニが言うことについて述べておく。たとえば、メディアンやモードの一般的な性質はそのようなものであろうか。(相加)平均がゼロ、分散

7) 中央項は「両端項の総計の半分(*semisomma degli estremi*)」とも言われている(63)。

が1の正規分布  $N(0, 1)$  のメディアンとモードを考へてみる。この密度関数  $f(x)$  は

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

である。このような分布にあっては、そのメディアンはゼロであり、モードも同じくゼロである。これは、関数関係で表現される分布の1つのパラメータである相加平均がメディアンと一致し、しかもモードとも一致することを示している。このことから、非解析的平均と関数表現とを切り離すジニの考へ方について、さらに、その本意を検討する必要性が示唆される。

なお、解析的平均と非解析的平均(位置上の平均)の数学的特性にかんして、ジニは

解析的平均=固定的、

非解析的平均=弛緩的

と規定している箇所がある(64)。ここに固定的とは、系列を構成する項の値の変化にたいして、もとめられる平均の数値が鋭敏に反応する性質を言う。たとえば、系列(1, 2, 3, 4, 5)が(2, 2, 3, 4, 5)に変化すると、相加平均は3.0から3.2になる。これにたいして、この2つの系列にかんするメディアンは2つの系列のどれをとっても3.0であり、項の値の変化にたいして不感的である。この性質を弛緩的と言う。

ただし、上述した第1の系列(1, 2, 3, 4, 5)が(2, 2, 3, 4, 4)に変化しても、相加平均は3.0のままである。このように、相加平均にも項の値の変化にたいして不感的な性質があって、そこには弛緩的性質を検出することができる。この事例から明らかなように、解析的平均の特性を固定的と規定し、他方で非解析的平均の特性を弛緩的と規定して、両者の平均を区別することは「厳密ではない」とジニは明言している(69)。

### 3. 単純解析的平均と加重解析的平均

ジニは値が同一となる項についてその項数が度数としてまとめられている系列を「重複的総体 (insieme con ripetizione)」(66)と名づけ、この系列にかんする解析的平均を加重解析的平均と言っている。これにたいして、同一の値の項数が度数としてまとめられていない場合の解析的平均を単純解析的平均と言っている。

ジニは、さまざまな解析的平均が内部性の要請を満たすかどうかについて、数学的考察を試みている<sup>8)</sup>。このときに加重解析的平均の計算式を用いた。それは、重複的総体を前提すれば、単純解析的平均を含めた解析的平均について一般的な検討が可能と考えられているからである。ただし、加重解析的平均をもとめるべき系列は単純解析的平均のための系列に編成替えすることができるので、単純解析的平均のほうが、加重解析的平均よりも根源的である。

### 4. 指数形式から見た解析的平均

#### — 底平均, 指数平均, 底-指数平均 —

ここで取り上げる分類は、解析的平均(単純解析的平均か加重解析的平均かを問わない)の数理形式が指数表現とどのような関係にあるかということに基づいている。すなわち、系列を構成する各項の値が、指数形式における①底の位置に置かれるか(底平均)、②指数の位置に置かれるか(指数平均)、③底と指数の両方の位置に置かれるか(底-指数平均)によって、ジニは解析的平均の表現形式を3つに分類している(表2)。

たとえば、相加平均  $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$

8) 木村(2010)、第5章参照。

表2 底平均, 指数平均, 底-指数平均 (例)

	単純解析的平均		加重解析的平均
底平均	相加平均	$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$	$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_k \cdot f_k}{f_1 + f_2 + \dots + f_k}$
	相乗平均	$G = (x_1 \cdot x_2 \cdot \dots \cdot x_n)^{\frac{1}{n}}$	$G = (x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_k^{f_k})^{\frac{1}{f_1 + f_2 + \dots + f_k}}$
指数平均	$c^E = \frac{c^{x_1} + c^{x_2} + \dots + c^{x_n}}{n}$		$c^E = \frac{c^{x_1} \cdot f_1 + c^{x_2} \cdot f_2 + \dots + c^{x_n} \cdot f_n}{f_1 + f_2 + \dots + f_n}$
底-指数平均	$y^y = \frac{x_1^{x_1} + x_2^{x_2} + \dots + x_n^{x_n}}{n}$		$y^y = \frac{x_1^{x_1} \cdot f_1 + x_2^{x_2} \cdot f_2 + \dots + x_n^{x_n} \cdot f_n}{f_1 + f_2 + \dots + f_n}$

は、変数  $x$  が指数形式の底の位置に置かれているので、①底平均に属すと言われる。②指数平均と③底-指数平均については、それぞれ以下の平均が、その例として挙げられている (64f.)。

$$\text{指数平均: } c^E = \frac{c^{x_1} + c^{x_2} + \dots + c^{x_n}}{n}$$

$$\text{底-指数平均: } y^y = \frac{x_1^{x_1} + x_2^{x_2} + \dots + x_n^{x_n}}{n}$$

## 5. 計算式の一般性と個別性

### — 個別的平均と包括的平均 —

特定の解析的平均をそれに特有の関数関係で表現したとき、その平均を個別的平均という。たとえば、

$${}^1A = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (3)$$

は相加平均である。

また、

$${}^2A = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}} \quad (4)$$

は平方平均である。

そして、

$${}^3A = \sqrt[3]{\frac{x_1^3 + x_2^3 + \dots + x_n^3}{n}} \quad (5)$$

は立方平均である。これらはいずれも具体的に解析的平均を特定していることから、個別的平均と言う。

ここで、累乗平均 ( $m$  次) と名づけられた

$${}^mA = \sqrt[m]{\frac{x_1^m + x_2^m + \dots + x_n^m}{n}} \quad (6)$$

を見ると、この(6)式は、 $m=1$  のとき相加平均 ((3)式)、 $m=2$  のとき平方平均 ((4)式)、 $m=3$  のとき立方平均 ((5)式) を表現していることが分る。(6)式のように一般型で表現した解析的平均を包括的平均と言う (65)。

次頁には、解析的平均の一般型としての包括的平均を、累乗平均に限定して、さまざまなバリエーション (個別的平均) とともにまとめた表を掲げてある (表3)。ここでとくに累乗平均を取り上げるのは、それがジニ『平均論』では包括的平均の例とされていることによる (65)。

これまでの考察では、分類基準の階層性としては、前項で取り上げた平均分類 (変数が指数形式のどの位置に置かれるかを基準にした平均分類) のほうが、この項で取り上げた平均分類 (一般性と個別性からの平均分類) よりも基底的存在であるということを、暗黙のうちに前提した。しかし、ジニがそのように考えているとは断定できない。2つの分類基準の階層性にかんする叙述は私見の域を出ないことを、ここに述べておく。

ただし、上で取り上げた累乗平均だけでなく、表4 (次頁) に掲げる累乗和平均 (139f.) もまた底平均の1種であり、そのなかに包括的平均と個別的平均があるから、指数形式と関連させた数式表現による分類 (①底平均, ②指数平均, ③底-指数平均) のほうが、より基底

表3 底平均としての累乗平均(包括的平均と個別的平均)

	$m$	数式	平均の名称
包括的平均	整数	${}^m A = \sqrt[m]{\frac{x_1^m + x_2^m + \dots + x_n^m}{n}}$	累乗平均

	$m$	数式	平均の名称
個別的平均	$m=1$	${}^1 A = \frac{x_1 + x_2 + \dots + x_n}{n}$	相加平均
	$m \rightarrow 0$	${}^{m \rightarrow 0} A = \lim_{m \rightarrow 0} \sqrt[m]{\frac{x_1^m + x_2^m + \dots + x_n^m}{n}}$ $= \sqrt[n]{\prod_{i=1}^n x_i}$	相乗平均 <sup>1)</sup>
	$m=-1$	${}^{-1} A = \sqrt[n]{\frac{x_1^{-1} + x_2^{-1} + \dots + x_n^{-1}}{n}}$ $= \left( \frac{1}{n \sum_{i=1}^n \frac{1}{x_i}} \right)^{-1}$	調和平均
	$m=2$	${}^2 A = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$	平方平均
	$m=2$	${}^2 A = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$	標準偏差 ただし, $\begin{cases} x_i = y_i - \bar{y} \\ \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \end{cases}$
	$m=3$	${}^3 A = \sqrt[3]{\frac{x_1^3 + x_2^3 + \dots + x_n^3}{n}}$	立方平均
	$m=4$	${}^4 A = \sqrt[4]{\frac{x_1^4 + x_2^4 + \dots + x_n^4}{n}}$	4乗平均 <sup>2)</sup>

1) 証明は, Dunkel, O., "Generalized Geometric Means and Algebraic Equations," *Annals of Mathematics*, Vol.11, N.2, 1938 参照。相乗平均  $G$  が底平均の一種であることは,  $G = (x_1 \cdot x_2 \cdot \dots \cdot x_n)^{\frac{1}{n}}$  と書き直せば, 明確になる(表2参照)。

2) さらに  $k$  乗平均にまで拡張することができる ( $m=k$ )。

表4 底平均としての累乗和平均(包括的平均と個別的平均) —単純平均—

	$m$	数式	平均の名称
包括的平均	整数(ただし, 0を除く)	${}^m T = \frac{x_1^m + x_2^m + \dots + x_n^m}{x_1^{m-1} + x_2^{m-1} + \dots + x_n^{m-1}}$	累乗和平均

	$m$	数式	平均の名称
個別的平均	$m=2$	${}^2 T = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{x_1 + x_2 + \dots + x_n}$	反調和平均

注記)  $m=1$  のとき, 累乗和平均は相加平均に等しい。



的な階層に位置すると考えることができる。

しかし、底平均以外の2種類の平均(指数平均と底-指数平均)についても包括的平均と個別的平均の両方が提示されているかと言えば、そうではない(80f., 101f.)。ジニが単純解析的平均と加重解析的平均の例として掲げた上記2種類の平均については、それが包括的平均であるか、あるいは個別的平均であるかにかんする判断は読者に委ねられているように見受けられる(表2参照)。ここでは、指数による数式表現の形式にもとづいて分類された3種類の解析的平均のすべてが、包括的平均と個別的平均に分類されるとするのは、あくまでも1つのジニ解釈にすぎないことを付言しておく。

## 6. 一価的平均と多価的平均

非解析的平均(位置上の平均)の場合、系列から一意的に平均値が特定される。解析的平均の1種である相加平均もまた、任意の系列から一意的に平均値が算出される。このように特定の1つの系列から単一の平均値が計算されるとき、それを一価的平均と言う。

ところが、(-9, -4, +4, +9)という系列から計算される相乗平均 $G$ は、

$$G = \sqrt[4]{(-9) \cdot (-4) \cdot (+4) \cdot (+9)} \\ = \pm 6$$

となり、相乗平均としては-6と+6の複数の値があたえられる。このような平均を多価的平均と言う(65)。

なお、相乗平均 $G$ がつねに多価的平均であるとは言い難い。たとえば(+4, +9)の系列について $G$ は+6だけであり、系列(+4, +9)は一価的平均をあたえる。平均が一価的か多価的かは、①系列と②平均の計算式によって定められる。

## 7. 現実的平均(実質的平均)と虚構の平均(計算的平均)

前項で取り上げた平均分類は、同一の系列にたいして計算式が何個の平均をあたえるかという形式的な基準にもとづいている。これにたいして、ここで取り上げる分類基準は、計算結果の実質の意味から平均を分類することにその目的が置かれている。この分類基準においては、平均が解析的か、あるいは非解析的かは問題にならない。要は、原系列を構成する項のなかに結果数字と同一の数値が存在するかどうかにかかっている。たとえば、家計所得(年収)の平均(相加平均でもメディアンでもよい)が450万円となったとき、この値を原系列のなかに見出しうるならば、それは現実的平均となる。それによつて、原系列のなかに見出すことができなければ、それは虚構の平均と言われる。ジニは、この虚構の平均を計算的平均とも言っている(63)。この言い回しは、この国で一般的に使用される「計算的平均」の用法とは異なっている。ここでは、このことを強調しておく。

## 8. 2つの虚構の平均

ジニは虚構の平均(計算的平均)を2つに分類している。第1の虚構の平均は、算出された平均と同一の値をもつ項をたとえ原系列のなかに見出すことができなくても、そのような値をとる項の存在が可能性として認められるときの平均である。たとえば、年収の相加平均が450万円と算出されたとき、そのような年収の世帯が元の系列のなか存在しなければ、それは虚構の平均(計算的平均)である。しかし、年収450万円の世帯の存在が合理的に推認される場合には、それは単なる虚構の平均(計算的平均)ではなくて、可能な虚構の平均(可能な計算的平均)となる。これにたいして、そのような年収をもつ世帯が存在する

ことに合理性を見出せない場合には、不可能な虚構の平均となる。

ジニが不可能な虚構の平均として例示しているのは、4.14人という平均世帯人員である。「1世帯が4.14人で構成されるということは絶対にあり得ないからである」とジニは述べている(445)。

このように指摘されてはいるが、平均計算のときの桁数の小数第1位を四捨五入して、平均世帯人員を4人とした場合、この構成の世帯は、たとえ元の系列のなかには存在しない場合であっても、実際に存在すると考えられるので、可能な虚構の平均が得られたことになる。また、元の系列のなかに4人家族の世帯が存在していれば、この平均世帯人員4人は、現実的平均(実質的平均)に分類される。

### おわりに

本稿では、さまざまな次元で平均を分類したジニの見解を取り上げて、その分類基準の階層性を考察した。図を用いて、その結果を要約する(図2~図4)。

まず、平均は解析的平均と非解析的平均とに分類される(図2)。ここに、解析的平均とは、内部性の要請を満たす特殊な平衡数であ

る。これにたいして、非解析的平均は同じく内部性の要請を満たしてはいるが、平衡数とは関係がないと考えられている。それは、度数分布と直接的あるいは間接的に関係づけられて定まる、変量の値である。この意味で、非解析的平均は位置上の平均とも言われている。

解析的平均は次のように分類される。すなわち、変量  $x$  が度数  $f$  とともにあたえられるかいなかで、加重解析的平均と単純解析的平均とに分類される。加重解析的平均が算出される系列は、単純解析的平均をあたえる系列に帰着させることができる。このため、解析的平均を計算するとき、変量の度数があらかじめあたえられている必要はない。この点で、度数分布を前提する非解析的平均とは異なっている。

解析的平均はその数理形式が指数関数の表現形式とどのような関係にあるかによって、それぞれ、底平均、指数平均、底-指数平均に分類される。さらに、これらの解析的平均の計算式が一般型であたえられるか、個別的な特定の計算式として表現されるかによって、包括的平均と個別的平均とに分類される。

他方で、非解析的平均(位置上の平均)は、①変量とその度数とを関連させて、変量の範

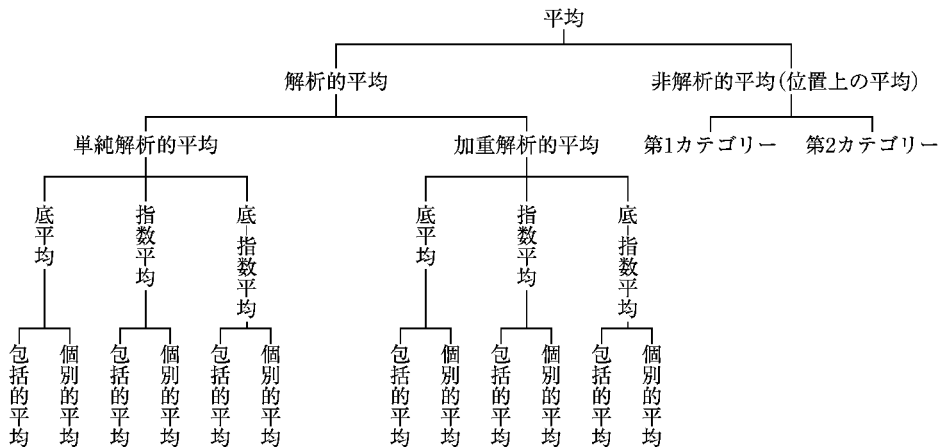


図2 数理形式から見た平均分類

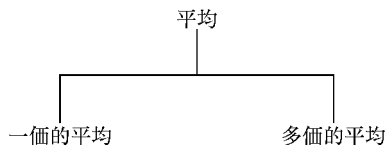


図3 結果数字の個数による平均分類

囲内で特定される変量の値であるか、あるいは②度数分布を前提とするが、変量の範囲だけで特定される値であるかによって、2つに分類される。本稿では、①の非解析的平均を第1カテゴリーに、また②を第2カテゴリーに分類した。

なお、図2に記載した平均がすべて、『平均論』のなかで具体的に示されているわけではない。たとえば、ジニが例示した指数平均と底-指数平均が個別的平均か、あるいは包括的平均かについては、判然としない(表2参照)。ここでは、数理形式からジニの平均を分類すれば、図2のように要約されるという私見の提示にとどまることを指摘しておく。図2に分類した平均の具体化は、今後に残されている。

これまでの、数理形式的な観点からの平均分類を述べた。ジニの平均分類は、計算式の形式的な特徴からだけでなされているのではない。結果数字から見た平均分類もなされている。

図3は、計算式のあたえる平均値が単一か複数かを基準にした平均分類を図示している。

図4は、平均としてあたえられる数値が、原系列のなかに見出さるかどうかを基準にした分類を示している。算出された平均値が原系列のなか存在していれば、それは現実

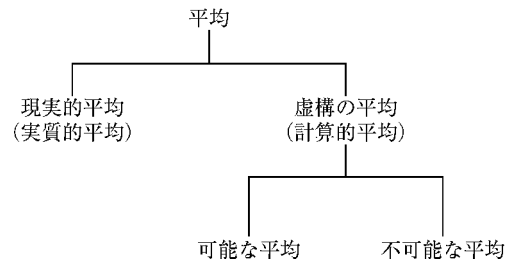


図4 結果数字の現実性から見た平均分類

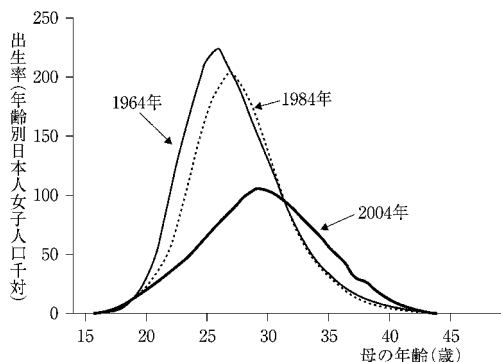
的平均(実質的平均)である。これにたいして、平均値と同一の値をもつ項が原系列のなか存在していなければ、その平均値は虚構の平均(計算的平均)である。

虚構の平均(計算的平均)は、その数値に現実性があるかそうでないかによって、可能な平均と不可能な平均とに分類される。本稿で取り上げた不可能な虚構の平均の例は、4.14人という平均世帯人員であった。ここで、この平均世帯人員を4人とする。これが原系列に存在しない場合には、可能な虚構の平均と見なされる。そして、原系列に世帯人員4人の世帯が存在していれば、この平均世帯人員は、現実的平均(実質的平均)と見なされる。

ジニの平均分類は以上のように要約される。その所説は、社会分析における平均(および平均操作)の意義と限界にかんして示唆的である。しかしながら、「数学の観点」から平均概念を考察するという『平均論』の執筆方針に規定されて、そこにおける叙述から、社会科学的研究のための指針を読み取ることは、必ずしも容易なことではない。この点の検討は今後の課題である。

#### (補注)

旧稿(木村(2008))の末尾に数学注「メディアン(中央値)  $Me$  とモード(最頻値)  $Mo$ 」を付した。その叙述(3. メディアン, モード, 相加平均の数学的關係——ドゥードソンの近似式——)に誤りがあったので(144頁以下)、訂正部分を含めた数学注の全文を以下に掲載する。



付図1 母の年齢別にみた出生率の年次比較  
 (出所)『国民衛生の動向・厚生指標』(臨時増刊) 第53巻第9号 2006年, p.41。

(数学注) メディアン(中央値)  $Me$  と  
 モード(最頻値)  $Mo$

付図1の縦軸は出生率(年齢別日本人女子人口千対)を示している。この図から母の出産年齢は上昇傾向にあることが分かる。付図1では、縦軸に相対数がとられているが、出生数(絶対数)をとることも可能である(付図2参照)。そのような度数分布図において、出産した母の総人数を半分に分かつ年齢をメディアン(中央値)  $Me$  という。また、もっとも出産数が多い母の年齢をモード(最頻値)  $Mo$  という。

この国では相加平均や相乗平均が計算的平均と言われるのにたいして、 $Me$  や  $Mo$  は位置上の平均と言われている。記述統計学が比較的

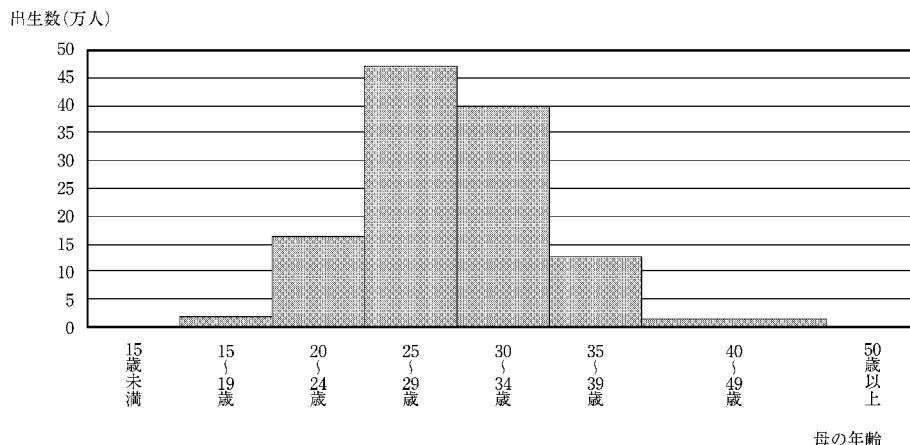
視されていた戦前・戦中期には、統計学の教科書では「計算的」と「位置上」という2種類の平均が取り上げられていたが、戦後になってからは確率論基調の統計理論(推測統計学)が主流になり、いつの間にか位置上の平均は忘れ去られたかの感を呈していた。

ところが、近年、OECDがメディアンに注目するようになった。OECDの調査研究報告書では、所得分布のメディアンに該当する所得の2分の1以下の所得層を「貧困層」と規定し、全体に占めるその割合(相対的貧困率)をもとめ、それによる国際比較が試みられている。経済格差が社会問題になるにつれて位置上の平均は、その分析手法として復位したかに見える。メディアンとモードにかんするこの「数学注」は、内容的には旧聞に属し、しかも初等数学的な叙述に終始するので、屋上屋を架するとの「そしり」を受けるかもしれない。しかし、このような現状にあっては、幾ばくかの有用性があるのではないかと考えて、先学の叙述を参考にして、数値例を付けた数学注をおくことにした。

1. メディアン(中央値)

付表1は2000年における母親の年齢階級別出生数(日本)を表章している。付表1から度数分布図を描くと付図2のようになる。

メディアン  $Me$  とは、度数分布の総度数を半



付図2 母の年齢階級別出生数の度数分布図

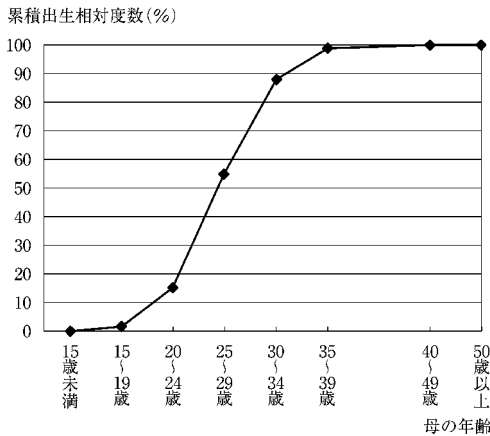
付表1 母の年齢階級別出生数\* (2000年)

15歳未満	15～19歳	20～24歳	25～29歳	30～34歳	35～39歳	40～49歳	50歳以上
43	19,729	161,361	470,833	396,901	126,409	14,848	6

\*総数(年齢不詳を含む) 1,190,547人

\*総数(年齢不詳を除く) 1,190,130人

(出所) 総務省統計局『日本の統計 2006年版』日本統計協会, 2007年, p.26より抜粋。



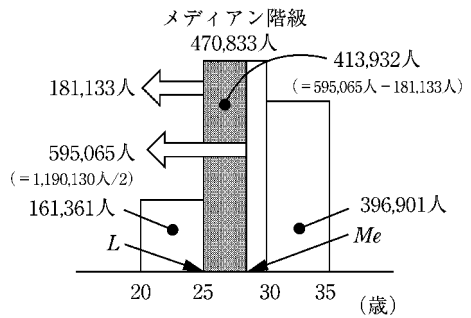
付図3 母の年齢階級別累積出生相対度数 (2000年)

分にする横軸の座標であたえられる。 $Me$ がどのあたりにあるかは、累積相対度数を折れ線グラフに書いて、縦軸の50%に該当する階級の年齢を読み取れば、おおよその見当がつく。そのために付表1から累積相対度数のグラフを描くことにする(付図3)。

付図3から、出生総数の1/2(累積相対度数が50%)に当たる母の年齢( $Me$ )は、「25～29歳」階級に落ちていることが分かる。この階級をメディアン階級と言う。ところが、メディアン階級が分かっても、何歳の母を境にして、出生数が出生総数の半分になっているかは分からない。

そこで、 $Me$ の値を特定するために、付図2から、メディアン階級を真ん中にして、前後の階級(全部で3つの階級)を抜き出す(付図4)。この付図4には、メディアン階級(階級間隔(=階級幅)は5歳、これを $c$ とおく)の人数(度数) $f$ が47万833人であると記載されている。

ここで、メディアンを想起する。それによれば、年齢が $Me$ までの母の人数は、すべての母の人数(総度数、ただし年齢不詳を除く)



付図4 メディアン階級(25～29歳)を中心とする度数分布図(部分)

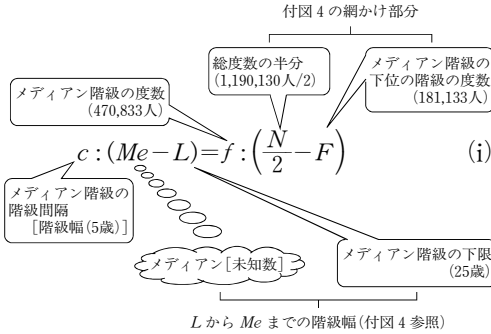
119万130人( $N$ )の半分、すなわち $N/2$ 人(=59万5,065人)である。この $N/2$ は、①一番若い母の年齢からメディアン階級の下限 $L$ までの母の人数 $F$ と②メディアン階級の下限 $L$ からメディアン $Me$ までの人数に分解される(付図4参照)。

このことをメディアン階級に着目して考えてみると次のようになる。すなわち、①総度数の半分 $N/2$ 人(=59万5,065人)から②25歳未満までの階級(メディアン階級の左側にあるすべての階級)に属す母の度数 $F$ (18万1,133人)を引けば( $N/2-F$ )、メディアン階級に属す母親のうち、付図4の $L$ から $Me$ までの間にいる人数が得られる。その人数は41万3,932人(=59万5,065人-18万1,133人)である。メディアン階級の下限(25歳)を $L$ で表すと(付図4)、( $Me-L$ )で示される範囲には( $N/2-F$ )人の母親がいることになる。このことから次のことが分かる。

- ① メディアン階級の階級間隔 $c$ には、メディアン階級の度数 $f$ が対応していること。
- ② メディアン階級の下限 $L$ からメディアン $Me$ までの間隔( $Me-L$ )には、総度数の半分( $N/2$ )に足りない度数をメディアン階級から補う度数( $N/2-F$ )が対応していること。こ

ここに、 $F$  はメディアン階級よりも下位の全階級の度数である。

①と②で述べた関係を数式で表現すれば、次のようになる。



(i) 式を整理すると、次式を得る。

$$Me = L + \frac{\frac{N}{2} - F}{f} \times c \quad (ii)$$

これに関連数値を代入すれば、

$$Me = 25 + \frac{1,190,130}{2} - 181,133}{470,833} \times 5 = 29.40(\text{歳})$$

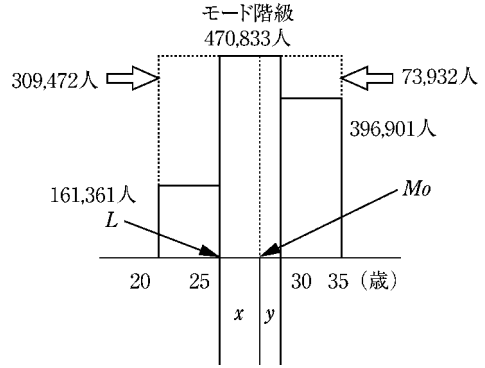
となり、メディアンは 29.4 歳である。

以上から、メディアンは、それが存在する階級の幅(メディアン階級の階級間隔： $c$ )をその階級の下限  $L$  からメディアン  $Me$  までの度数で按分していることが分かる。

## 2. モード(最頻値)

モード  $Mo$  は、度数分布においてもっとも度数が大きい値である。付表1(付図2)から明らかのように、最大度数の階級(モード階級)は「25~29歳」階級である。モード階級の下限(25歳)とその右隣の階級の下限(30歳)の相加平均(27.5歳)をモードの値  $Mo$  と見なすことがある。近似的にはこれでもよいが、これは幾分、厳密さに欠ける。

モードの数値を特定する目的で、付図2からモード階級とその前後の階級にかんする度数分布図を抜き出すことにする(付図5)。モード階級を中心とする3つの階級分布が付図5のよう



付図5 モードは大きい度数の階級(30~34歳階級)に引き寄せられる

になるとき、モード階級の右隣に位置する階級(「30~34歳」階級)の度数(39万6,901人)は、「20~24歳」階級(モード階級の左隣の階級)の度数(16万1,361人)よりも多い。この場合、モード  $Mo$  は度数が大きい右隣の階級に引き寄せられていると考えるのが自然である。すなわち、付図5に示すように  $x > y$  である(これとは逆に、モード階級の左隣にある階級の度数が右隣の階級の度数よりも大きい場合には、モードは左隣の階級に引き寄せられる)。

$Mo$  の数値を特定するためには、モード階級の下限  $L$  (25歳) から  $Mo$  がどれだけ乖離しているか、換言すれば  $Mo$  の位置を規定する  $x$  がどれだけ大きいかを特定しなければならない。 $x$  の値がもとめられれば、モード階級の下限  $L$  にこの  $x$  を足すことによって、 $Mo$  が分かる( $Mo = L + x$ )。

さて、 $x$  によって  $Mo$  の値を特定するには、付図5において  $x > y$  という大小関係にある  $x$  (と  $y$ ) が、モード階級とその両隣の階級との度数差によって規定されればよい(米澤治文・一条勝夫『講要 統計学』日本評論社、1958年、p.50、および足利末男『社会統計学の基礎』晃洋書房、1982年、p.142以下参照)。

付図5にもとづいてモード階級の両隣の階級について関連数値を付表2に表章する。この度数差の割合の違いが  $x$  の値を規定し、その  $x$  によってモードの位置が定まると考えることが、

付表2 モード階級の左右の階級別度数差とその割合

階 級	左側の階級 (20~24 歳)	右側の階級 (30~34 歳)
モード階級との度数差	309,472 人 (=470,833-161,361)	73,932 人 (=470,833-396,901)
度 数 差 の 割 合	$\frac{309,472}{383,404} \left( = \frac{309,472}{309,472+73,932} \right)$	$\frac{73,932}{383,404} \left( = \frac{73,932}{309,472+73,932} \right)$

モードの数値的特定におけるポイントである。度数差の割合が小さい階級 (実際の度数がモード階級の度数により近い階級) ほど, モードをその階級に引き寄せる。そして, 逆に, 度数差の割合が大きい階級 (実際の度数がモード階級の度数からより大きく乖離している階級) ほど, モードをその階級から遠ざけている。このことは付図5から直観的に理解できる。

設例では, 度数差の割合がより大きい階級は, モード階級の左隣に位置する階級 (「20~24 歳」階級) である。この階級に着目すれば, モード  $Mo$  の値は, この階級に  $x$  だけ近づいていると言うよりは, 度数差の割合がより小さい階級 (「30~34 歳」階級) の方へと (モード階級の下限  $L$  (25 歳) から距離にして  $x$  だけ右方に) 「押しやる」と見るべきであろう。

以上の考察によって, モード階級の階級間隔 (5 歳)  $(=x+y)$  —これを  $c$  とおく— を度数差の割合で按分すれば, モード  $Mo$  が定まることになる (付図5, 付表2 参照)。すなわち

$$x:(x+y) = \frac{309,472}{309,472+73,932} : \left( \frac{309,472}{309,472+73,932} + \frac{73,932}{309,472+73,932} \right) \quad (\text{iii})$$

を満たす  $x$  をもとめれば,  $Mo$  の位置を定めることができる。上で述べたように  $x+y$  はモード階級の階級間隔  $c$  であるから,  $x+y=5 (=c)$  になる。これを (iii) 式に代入して整理すれば,

$$x:5 = \frac{309,472}{383,404} : 1$$

$$\therefore x = 5 \times \frac{309,472}{383,404} = 4.04 \quad (\text{iv})$$

を得る。これにより付図5の  $x$  は  $x=4.0$  になる。

モード階級 (「25~29 歳」階級) の下限 ( $L$ ) は 25 であるから,

$$Mo = L + x = 25 + 4.0 = 29.0 \quad (\text{v})$$

となり, モードは 29.0 歳である。

(iv) 式と (v) 式を参考にすれば, モード  $Mo$  をもとめる一般式は次のようになる。

$$Mo = L + c \times \frac{D_l}{D_l + D_r} \quad (\text{vi})$$

ただし,  $L$  はモード階級の下限。

$c$  はモード階級の階級間隔。

$D_l$  はモード階級の左側の階級にかんする度数差。

$D_r$  はモード階級の右側の階級にかんする度数差。

ここで, モード階級の度数を  $f$ , その左右に隣接する階級の度数をそれぞれ  $f_l$  および  $f_r$  とすれば, それぞれの階級の度数差は

$$D_l = f - f_l$$

$$D_r = f - f_r$$

である。この関係を (vi) 式に代入すれば, モード  $Mo$  をもとめる一般式は次のようにも表現することができる。

$$Mo = L + c \times \frac{f - f_l}{(f - f_l) + (f - f_r)} \quad (\text{vii})$$

$$= L + c \times \frac{f - f_l}{2f - (f_l + f_r)} \quad (\text{vii}')$$

この (vii) 式 (または (vii)' 式) を用いれば, モード階級およびそれを挟む階級の度数から  $Mo$  をもとめることができる。

3. メディアン, モード, 相加平均の間の  
数学的関係 — ドゥードソンの近似式 —

中程度 (a moderate degree) の非対称分布にあってはメディアン  $Me$ , モード  $Mo$ , 相加平均  $\bar{x}$  の間に

$$|Mo - Me| = \frac{2}{3} \times |Mo - \bar{x}| \quad (\text{viii})$$

という数学的関係があると考えられていた時期があった。カール・ピアソンはこの関係を特殊な分布について証明した (Pearson, Karl, "Contribution to the Mathematical Theory of Evolution, II. Skew Variation in Homogeneous Material," *Philosophical Transactions of the Royal Society of London*, Ser. A., Vol. 187, 1895; also in *Karl Pearson's Early Statistical Papers*, Cambridge 1948.)。

これにたいして, A.T.ドゥードソンは, 分布型や分布関数のパラメータの値によって近似度が異なるが, 非対称分布の  $Me$ ,  $Mo$ ,  $\bar{x}$  については近似的に

$$|Me - \bar{x}| = \frac{1}{3} \times |Mo - \bar{x}| \quad (\text{ix})$$

が成り立つことを説明したと言われている (Doodson, A.T., "Relation of the Mode, Median and Mean in Frequency Curves," *Biometrika*, Vol. XI, 1915-17, pp. 425ff.; 近藤次郎・守岡隆「度数分布」中山伊知郎編『統計学辞典(増補版)』東洋経済新報社, 1951年, p. 169参照)。この数学注の1.と2.で取り上げた母の年齢分布にかんする相加平均  $\bar{x}$  の近似値は, この (ix) 式によって, もとめることができる。1.から  $Me = 29.4$  歳, また2.から  $Mo = 29.0$  歳となった。この数値を (ix) 式に代入すれば,

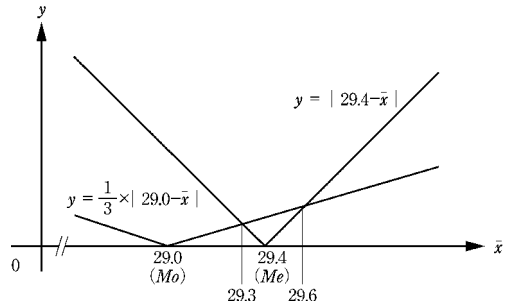
$$|29.4 - \bar{x}| = \frac{1}{3} \times |29.0 - \bar{x}|$$

となる。したがって, もとめる  $\bar{x}$  の値は

$$\begin{cases} y = |29.4 - \bar{x}| \\ y = \frac{1}{3} \times |29.0 - \bar{x}| \end{cases}$$

を満たす。この方程式をグラフで表示すれば, 付図6のようになる。

この付図6から, もとめる解は次の連立方程



付図6  $Me$ ,  $Mo$ ,  $\bar{x}$  の近似的関係

式の解としてあたえられることが分かる。すなわち,

$$\begin{cases} y = 29.4 - \bar{x} \\ y = \frac{1}{3} \bar{x} - \frac{29}{3} \end{cases} \quad \begin{cases} y = \bar{x} - 29.4 \\ y = \frac{1}{3} \bar{x} - \frac{29}{3} \end{cases}$$

これを解けば

$$29.4 - \bar{x} = \frac{1}{3} \bar{x} - \frac{29}{3} \quad \bar{x} - 29.4 = \frac{1}{3} \bar{x} - \frac{29}{3}$$

$$\frac{4}{3} \bar{x} = 39.1 \quad \frac{2}{3} \bar{x} = 19.7$$

$$\therefore \bar{x} = 29.3 \text{ (歳)} \quad \therefore \bar{x} = 29.6 \text{ (歳)}$$

となって, それぞれの方程式は母の平均年齢(相加平均)の近似値として29.3歳と29.6歳をあたえる。

それでは付図7のように, 年収の総平均(相加平均, 厳密には加重相加平均)  $\bar{x}$  とメディアン  $Me$  があたえられているときは, モード  $Mo$  の近似値はどうなるであろうか。

$Me = 458$  万円,  $\bar{x} = 563.8$  万円であるから,  $Me < \bar{x}$  である。また, 付図7から明らかのようにモード階級は総平均が存在する階級よりも小さい階級にあるので,  $Mo < \bar{x}$  である。したがって, このような場合には, ドゥードソンの近似式((ix)式)は,

$$\bar{x} - Me = \frac{1}{3} \times (\bar{x} - Mo)$$

である。この式に関連数値を代入すれば,

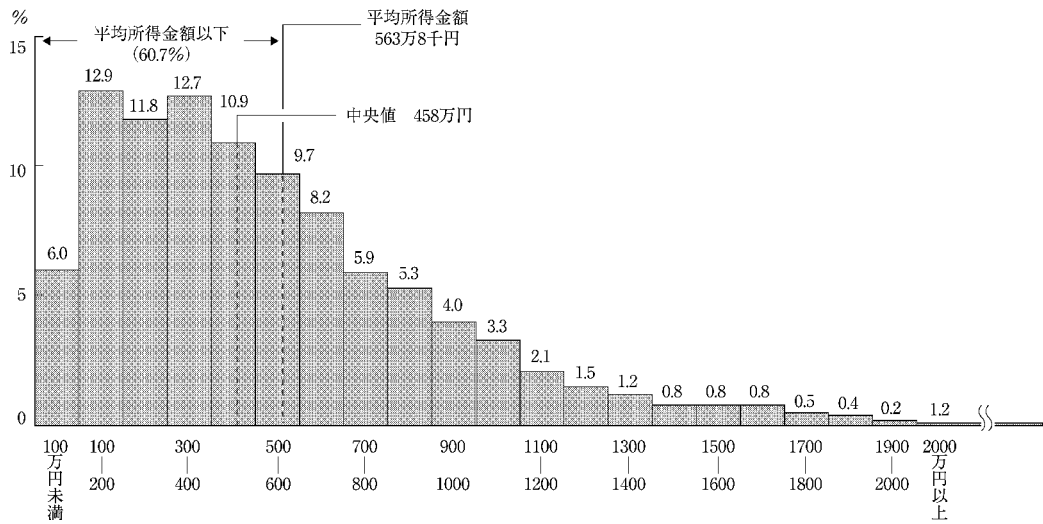
$$563.8 - 458 = \frac{1}{3} \times (563.8 - Mo)$$

$$\frac{1}{3} Mo = -(563.8 - 458) + \frac{1}{3} \times 563.8$$

$$\therefore Mo = 246.4 \text{ (万円)}$$

となり, モード  $Mo$  の近似値は246万円となる。





付図7 所得分布(2006年調査)

(出所) 厚生労働省『国民生活基礎調査』(2006年)

<http://www.mhlw.go.jp/toukei/saikin/hw/k-tyosa/k-tyosa06/2-1.html>, accessed on March 3, 2007.

ここで、ドゥードソンの近似式の適用結果と比較する目的から、ピアソンの近似式(viii)式に関連数値 ( $Me=458$  万円,  $\bar{x}=563.8$  万円) を代入すると、

$$|Mo-458| = \frac{2}{3} \times |Mo-563.8|$$

を得る。上式の辺々を2乗すると、

$$(Mo-458)^2 = \left(\frac{2}{3}\right)^2 \times (Mo-563.8)^2$$

になる。これを整理すると

$$0.6Mo^2 - 414.8Mo + 68488.2 = 0$$

であり、その解は  $Mo = 272.6, 418.7$  である。よって、モードの近似値は273万円または419万円である。この値とドゥードソンの近似式があたえる値(246万円)を比較すると、後者の方が良好な近似値をあたえていることが分かる。

〔付記〕 本稿は、北海学園学術研究助成金(2010年度)の交付を受けた共同研究の成果の一部である。