新統計集団論(1)

船津好明

はじめに

集団は統計学の基本概念であり、その認識が重要であることはいうまでもない。古くから様々な観点から論じられていて、理論展開の土台となっている点、今も変わりはない。

本稿は、集団を新しい観点からとらえ、理論を立て、新たな研究分野を拓こうとするものである。

本稿における目新しい事項は、集団内の負の構成単位の認識、集団の写像、それらに基づく母集団と標本、統計量と期待値などである。このうち前２者は、筆者が過去に公表した理論で、後２者は、前２者に基づくものとしては初めてのものである。負の分散が登場するなど、他に同類の論を知らない。

1．集団の認識

単純な例から入る。４人の各所得を、ある単位で７、８、９、10とする。この４人を１つのまとまりとするとき、大きさ４の集団が認識される。構成単位は各人で、各所得は、この例では各構成単位の値となっている。

集団の値を x で代表し、構成単位を i と表せば集団は一般に、

$$x = \{x_1,\ x_2,\ x_3, \cdots\} \quad ①$$

と表すことができる。構成単位の数が N で、各値が変数表示されていれば、①は

$$x = \{x_i\}_N \quad ②$$

と書くこともできる。上の例は具体的には

$$x = \{7,\ 8,\ 9,\ 10\} \quad ③$$

となる。

別の５人の各所得を７、８、８、９、10とする。これらは、もう１つの集団として

$$y = \{7,\ 8,\ 8,\ 9,\ 10\} \quad ④$$

と表すことができる。③と④は似ているが異なる。構成単位の数が違い、更に③には８が１つ、④には２つある点にも違いがある。言い換えれば、③と④は値の範囲は同じであるが、度数が違うということである。

集団の表示方法として、各値に注目し、同じ値の度数を付記して

$$y = \{7,\ 8(2),\ 9,\ 10\} \quad ⑤$$

と書けば、8(2) は８が２つあることを意味している。度数をすべて付記すれば

$$y = \{7(1),\ 8(2),\ 9(1),\ 10(1)\} \quad ⑥$$

となる。あるいは値と度数の対を明示して、これを同列並記すれば

$$y = \{(7,1),\ (8,2),\ (9,1),\ (10,1)\} \quad ⑦$$

と2元的表記になる。これを一般的に書けば

$$y = \{(v_1,\ f_1), (v_2,\ f_2), \cdots\cdots, (v_k,\ f_k)\} \\ = \{(v_i,\ f_i)\}_k \qquad ⑧$$

となる。この場合 v_i は変数の値、f_i はその度数であり、正の整数である。これは通常見られる度数分布の1つの表示形式に他ならない。

度数は相対表示などによって小数や0をとるようにすることができるので、上の f_i は実は非負の実数と考えてさしつかえない。更に後記の負の構成単位を考えれば、f_i は負数も含めた実数でよいことになる。

後の議論のため、値の表示を x に戻し、集団の構成単位を一つひとつ表示することにすれば、大きさ N の従来の集団は

$$x = \{(x_i,\ 1)\}_N \qquad ⑨$$

となる。あるいは1を省略して

$$x = \{x_i\}_N \qquad ⑩$$

と書けば、前記の②と同じものになる。

2．負の構成単位

負の構成単位の度数の単位は－1である。値が負であることを指すものではなく、単位そのものが負であるというものである。これに関して筆者は過去に詳しく論じた。念のため、負の構成単位を簡単に例示しておく。(a)統計の集計で誤って1単位を加えてしまったとき、この単位の度数を－1にして1単位加えると、先に誤って加えた単位が削除される。この場合、後で加えた1単位の度数が－1で、負の構成単位に当たる。この場合、その単位は集団に残らない。もし、負の単位を2単位加えると、負の単位が1つ集団に残る。(b)ある地域である期間に住民が1人転出または死亡すれば、その者はその地域に居なくなるが、これは別地域（他界）に1人出現したものと考えることができる。これを元の地域では－1人と数えることになる。

負の構成単位に対して、集団の中に実在する単位は正の構成単位であり、その度数の単位は1である、と考えるのは一応解り易い。しかし、これらは説明の便宜であって、数の正負は本来対等であり、意義に何らの差異はないから、正が実在で負が架空というのは理解のための方便に過ぎない。筆者の真意は、正も負も完全に対等な「存在」であって、よって負の構成単位は負としての存在なのである。よって度数の1と－1は意義において対等である。

そこで、正負の構成単位が混在する集団を、構成単位一つひとつを表示して（度数分布にまとめないで）

$$z = \{(x_i,\ u_i)\}_N \qquad ⑪$$

と表わすこととする。ここに、各 x_i のとる値に制限はないが、各 u_i のとる値は1または－1に限るものとする。u_i について更に詳しくいえば、u_i がとる1の個数は0から N までの間で、これを N_{+1} と表わし、－1の個数も0から N までの間で、これを N_{-1} と表わし、$N_{+1}+N_{-1}=N$ となるものである。特に u_i がすべて1をとるとき、即ち $N_{+1}=N$ のときは、$N_{-1}=0$ で従来の分布と同じ集団になるから、⑪は従来の集団より一般性に富む集団ということができる。u_i がこのように限定され、あるいは特定の研究場面で u_i をいちいち論じる必要がないときは、簡素のため u_i を省略し

$$z = \{x_i\}_N \qquad ⑫$$

と書けば、実質②と同じになり、従来の集団に帰着する。

3．集団の写像

ここでは、集団を⑫の形で、ただし記号を変

え、

$$q = \{q_i\}_N \quad ⑬$$

とおく。q_i は x_i と同じでも、$(x_i,\ u_i)$ であっても、更に多元であってもよい。

次に、q_i の関数を、例えば

$$r_j = f_j\{q_i\} \qquad j = 1,\ 2, \cdots,\ M \quad ⑭$$

とし、r_j からなる集団を

$$r = \{r_j\}_M \quad ⑮$$

とおく。この各 r_j を、集団 q の写像という。写像 r_j は集団 q の特性を示すものである。写像の表現形式はこのほかにも色々ある。⑭の形は1例に過ぎない。関数の形は無限に多くあるから、集団 q の写像は無限に多くあることになる。

関数 f_j は、集団 q に対して鏡の役割をもち、集団 q が鏡 f_j に当たって r_j として写し出されると考えるものである。r_j の集団 r が、q の写像集団である。

写像は本来 q と r の構成単位間の対応関係をいうものであるが、上のように考えると分かり易く、対応結果である r も含めていうこととする。r に対して q を原像ということがある。

実際に関数を立て、写像を作るには、何らかの意図を伴うものである。それによって写像は無限に多い中から特別のものが選択されることになる。写像は原像の特性値であるが、その特性値の意味は写像を設けた意図と同じである。

こうして、上の⑬～⑮によって、大きさ N の集団 q から大きさ M の集団 r が派生したことになる。

4. 写像集団の写像

今度は上で得た大きさ M の集団⑮に対し、先と同じ考えで各 r_j の関数による写像を、例えば

$$s_k = g_k\{r_j\} \qquad k = 1,\ 2,\ \cdots, L \quad ⑯$$

とし、s_k からなる集団を

$$s = \{s_k\}_L \quad ⑰$$

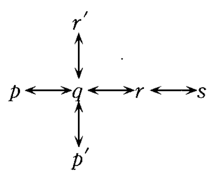
とする。この s は初めの集団 q からみれば、**写像集団の写像集団**である。以下同様にして、どの集団に対しても新たな写像集団を連鎖的に生成させることができる。これまでのことを整理すると、次の概念図ができ上がる。(図1)
$N,\ M,\ L$ に大小関係はない。

図1

$$\boxed{q = \{q_i\}_N} \rightarrow \boxed{r = \{r_j\}_M} \rightarrow \boxed{s = \{s_k\}_L}$$

この考え方を更に拡充すると、集団 q は別の集団の写像であってもよいし、q から r に行かず、別の r' に行ってもよいし、写像生成の方向も双方向的でよいから、集団 q を中心にすると（他を中心にしても同じ）次に示すような多元的集団網ができ上がる。(図2)

図2



5. 写像の小例

以上のことについて、簡単な例を示す。
［例1］初めに挙げた4人の所得をとり上げ、

$$q = \{q_i\}_4 = \{7,\ 8,\ 9,\ 10\} \qquad (N = 4) \quad ⑱$$

とする。次に r_j として

$$r_1 = f_1\{q_i\} = \sum_{i=1}^{4} q_i$$
$$= 7+8+9+10 = 34 \qquad ⑲$$
$$r_2 = f_2\{q_i\} = \sum_{i=1}^{4} q_i^2$$
$$= 7^2+8^2+9^2+10^2 = 294 \qquad ⑳$$

を作る。r_1 は所得の合計、r_2 は所得の２乗の合計で、所得の個人差を測定する際の要素としての意味がある。こうして

$$r = \{r_j\}_2 = \{34,\ 294\} \qquad ㉑$$

が生成する。r は q の写像集団（特性値集団）である。次に s_k として

$$s_1 = g_1\{r_j\} = r_2 - \frac{r_1^2}{4} = 294 - 34^2/4 = 5 \qquad ㉒$$

とすると、s_1 は r により q の所得の個人間変動を計算したことになる。

$$s = \{s_k\}_1 = \{5\} \qquad ㉓$$

は r の写像であり、q の写像集団の写像である。

［例２］$q = \{q_i\}_N$ に対して、N 次の行列 A を用い

$$A\begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \end{bmatrix} = r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix} \qquad ㉔$$

とおくと、r は q の、A を介した写像となる。即ち $Aq = r$ の形である。次に r に対して、A の逆行列が存在するとして、$s = A^{-1}r$ を作れば、$s = q$ となる。これは写像から原像に戻る例である。例えば $N = 4$ で $q = \{7,\ 8,\ 9,\ 10\}$ とし、（q を４行１列の行列に直し）

$$A = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix},\ q = \begin{bmatrix} 7 \\ 8 \\ 9 \\ 10 \end{bmatrix} \qquad ㉕$$

として $Aq = r$ とおけば $r = \{41,\ 42,\ 43,\ 44\}$ が得られる。r は q の、行列 A を介した１次変換による写像集団（特性値集団）である。

一方、A の逆行列は

$$A^{-1} = \begin{bmatrix} 0.8 & -0.2 & -0.2 & -0.2 \\ -0.2 & 0.8 & -0.2 & -0.2 \\ -0.2 & -0.2 & 0.8 & -0.2 \\ -0.2 & -0.2 & -0.2 & 0.8 \end{bmatrix} \qquad ㉖$$

であるので、$A^{-1}r = s$ とおけば $s = \{7,\ 8,\ 9,\ 10\}$ となって $s = q$ となる。即ち s は q の写像の逆写像で、完全に q に戻っている。喩えれば、人の顔を歪んだ鏡に写し、それをもう１つの歪んだ鏡に写して元の顔に見せるようなものである。この例では、r は q の特性値集団であるので、$A^{-1}r = s = q$ は、特性値を知って元の集団を完全に復元することができる１例である。

［例３］$q = \{q_i\}_N$ において

$$r_j = \frac{1}{N}\sum_{i=1}^{N} q_i^{\,j} \qquad j = 1,\ 2,\ \cdots,\ M \qquad ㉗$$

とおくと、$r = \{r_j\}_M$ は q の原点の周りの積率群を作る。次にこれらの r_j に対して

$$\begin{aligned} s_1 &= r_1 \\ s_2 &= r_2 - r_1^2 \\ s_3 &= r_3 - 3r_2r_1 + 2r_1^3 \\ s_4 &= r_4 - 4r_3r_1 + 6r_2r_1^2 - 3r_1^4 \\ &\ \vdots \\ s_M &= r_M - \cdots\cdots \end{aligned} \qquad ㉘$$

を作れば、$s = \{s_k\}_M$ は、q からみれば、q の平均と、平均の周りの積率群となる。具体的に $N = M = 4,\ q = \{7,\ 8,\ 9,\ 10\}$ とおけば

$$\begin{aligned} r_1 &= (7+8+9+10)/4 = 8.5 \\ r_2 &= (7^2+8^2+9^2+10^2)/4 = 73.5 \\ r_3 &= (7^3+8^3+9^3+10^3)/4 = 646 \\ r_4 &= (7^4+8^4+9^4+10^4)/4 = 5764.5 \end{aligned} \qquad ㉙$$

となり、

$$r = \{8.5,\ 73.5,\ 646,\ 5764.5\} \qquad ㉚$$

が得られる。次にこれらの r_j に対して

$$s_1 = r_1 = 8.5$$
$$s_2 = r_2 - r_1^2 = 1.25$$
$$s_3 = r_3 - 3r_2r_1 + 2r_1^3 = 0 \quad ㉛$$
$$s_4 = r_4 - 4r_3r_1 + 6r_2r_1^2 - 3r_1^4 = 2.5625$$

を作れば、$s = \{8.5,\ 1.25,\ 0,\ 2.5625\}$ が得られる。s は r の写像であるが、q の平均と、平均の周りの4次までの積率群からなっている。一般に大きさ N の未知の集団に対して、N 個の積率を与えれば、それらを特性値とするような元の集団が存在すれば（複数通りありうる）、それを再現することができる。

6．集団の同一性

集団を前記⑪で表すこととし、2つの集団の等、不等について、例を通して考える。集団 a と b を⑪の具体例として、次のようにおく。

$$a = \{(7,\ 1),\ (8,\ 1),\ (8,\ -1),\ (9,\ 1),\ (10,\ 1)\} \quad ㉜$$

$$b = \{(7,\ 1),\ (9,\ 1),\ (10,\ 1)\} \quad ㉝$$

a と b を比較するに、構成単位は a が5、b が3、変数 x のとりうる値でみると a は7，8，9，10、b は7，9，10で、これらのことから、a と b は異なる集団である。

a の中で、変数 x が同じである単位の u を統合して、即ち（8，1）と（8，－1）を合わせて、

$$c = \{(7,\ 1),\ (8,\ 0),\ (9,\ 1),\ (10,\ 1)\} \quad ㉞$$

を作り a と c の関係を考えてみる。まず c は要素として（8，0）を含むので、⑪の形には該当しない。よって c は a と異なる集団である。c は実は a の写像の1つであって、㉞では（8，0）は意味のある要素である。一方、ある研究場面で（8，0）を無意味と定義する場合は、その研究場面において c から（8，0）が除かれ、b と同じものになる。

次に a に似た形として

$$d = \{(7,\ 1),\ (8,\ 1),\ (-8,\ 1),\ (9,\ 1),\ (10,\ 1)\} \quad ㉟$$

を考える。構成単位はすべて正であるが、8の1つが負となっている。d は a と異なる集団である。

更に b に似た形として

$$e = \{(9,\ 1),\ (7,\ 1),\ (10,\ 1)\} \quad ㊱$$

と b を比べてみる。e は b の要素の配列順を変えたものである。e と b が同じ集団であるかどうかは、研究の立場によって異なる。即ち配列順を問題としない研究においては $e = b$ としてよいが、配列順が問題になる研究においては、e と b は別の集団となる。

集団の同一性を考えるに、1つの集団と同じ集団は原則としてそれ自身のみであるが、特別の研究場面で一見異なる2つの集団を等しいと定義することはさしつかえない。

7．母集団と関連集団

ここでは負の構成単位を含む集団⑪に戻る。z は x と u からなる2元の同時分布であって、周辺分布は、x が一般で特段の制約なく、u は1と－1をとる2項分布である。（1と0ではない。）

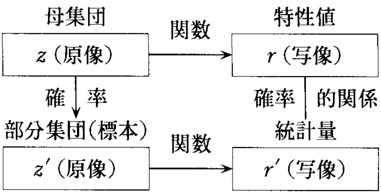
集団 z から、写像集団 r を作る。r は z の特性値の集団である。

集団 z から幾つかの単位を確率抽出し、z の部分集団で z' を作る。z' は z の標本である。

z' からは写像集団 r' を作る。r' は統計量の集団である。こうして4つの集団 z，z'，r，r' ができる。これらのうち、z と r、z と z'、z' と r' の関係はもはや説明を要しないが、r と

r' の関係は特に留意に値する。即ち、r と r' の関係には、z と z' の間の確率的関係を反映して、確率的関係がある。z と r を未知、z' と r' を既知とし、r と r' の確率的関係を記述することを、z に対する統計的推測という。(図3)

図3



8．母集団と写像の例——特性値

次に、上記について、数値計算例を含め、具体的展開を試みる。$z=\{(x_i,\ u_i)\}_N$ とおく。数値例として、$N=19$、$z=\{(3.01,\ 1), (4.15,\ 1), (4.22,\ 1), (4.84,\ 1), (5.09,\ 1), (5.83,\ -1), (6.07,\ 1), (6.46,\ 1), (6.61,\ 1), (6.96,\ 1), (7.11,\ 1), (7.45,\ 1), (7.79,\ 1), (8.09,\ 1), (8.14,\ 1), (8.72,\ 1), (9.37,\ 1), (9.55,\ 1), (11.34,\ 1)\}$ とする。これを基にして z の写像を幾つか設ける。まず、z の中の正負の単位の構成を明らかにする。N 個（数値例19個）の単位のうち、u_i が1となる単位数を N_{+1}（数値例18個）、-1 となる単位数を N_{-1}（数値例1個）とする。当然 $N_{+1}+N_{-1}=N$ で、$0\leqq N_{+1}\leqq N$、$0\leqq N_{-1}\leqq N$ である。ここで、便宜 N_{+1} の N に対する割合を

$$p=\frac{N_{+1}}{N}\quad(0\leqq p\leqq 1)\qquad ㊲$$

（数値例18/19＝0.9473 6842）

とおく。次に各 x_i, u_i を用いて関数を立て、集団 r を構成する写像を作る。例えば

周辺分布 x の平均

$$\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i\quad(\text{数値例 }6.8842\ 1053)\qquad ㊳$$

周辺分布 x^2 の平均

$$\overline{x^2}=\frac{1}{N}\sum_{i=1}^{N}x_i^2\quad(\text{数値例 }51.5896\ 6316)\qquad ㊴$$

周辺分布 x の分散

$$\sigma_x^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2\qquad ㊵$$

（分母は N、数値例 4.1973 0859）

周辺分布 u の平均

$$\bar{u}=\frac{1}{N}\sum_{i=1}^{N}u_i=\frac{1}{N}(N_{+1}-N_{-1})$$
$$=\frac{1}{N}\{N_{+1}-(N-N_{+1}\}=\frac{1}{N}(2N_{+1}-N)$$
$$=2p-1\qquad ㊶$$

（数値例 0.8947 3684）

周辺分布 u^2 の平均

$$\overline{u^2}=\frac{1}{N}\sum_{i=1}^{N}u_i^2=1\quad(\text{数値例 }1)\qquad ㊷$$

周辺分布 u の分散

$$\sigma_u^2=\frac{1}{N}\sum_{i=1}^{N}(u_i-\bar{u})^2=\frac{1}{N}\sum_{i=1}^{N}u_i^2-\overline{u}^2$$
$$=1-(2p-1)^2=4p(1-p)\qquad ㊸$$

（分母は N、数値例 0.1994 4598）

同時分布 xu の平均

$$\overline{x}_u=\frac{1}{N}\sum_{i=1}^{N}x_iu_i\quad(\text{数値例 }6.2705\ 2632)\qquad ㊹$$

x の u 加重平均（同時分布）

$$\mu=\frac{\sum_{i=1}^{N}x_iu_i}{\sum_{i=1}^{N}u_i}=\frac{\sum_{i=1}^{N}x_iu_i}{N(2p-1)}\qquad ㊺$$

（数値例 7.0082 3529）

x の u 加重積率（同時分布）

$$\sigma_j=\frac{\sum_{i=1}^{N}(x_i-\mu)^ju_i}{\sum_{i=1}^{N}u_i}=\frac{\sum_{i=1}^{N}(x_i-\mu)^ju_i}{N(2p-1)}\qquad ㊻$$

$$j=2,\ 3,\ \cdots,\ N$$

（数値例 $\sigma_2 = 4.5449\ 7924$,
$\sigma_3 = -0.3787\ 5805$,
$\sigma_4 = 50.2552\ 0206$）

これらのうち、㊳－㊸は周辺分布に関するもの、㊹－㊻は同時分布に関するもので、特に㊺、㊻の2式はu_iがx_iのウェイトの位置にあり、u_iがすべて1またはすべて－1のとき、普通の平均や積率の式に戻る。

9．負の分散など

u_iが1と－1の両方をとるときは、㊺、㊻は変わった振舞をする。まず、分母が0になることがあり、そのときは値がなくなる。

特に㊻で$j=2$のときσ_jは

$$\sigma_2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2 u_i}{\sum_{i=1}^{N} u_i} \qquad ㊼$$

となる。これはxの分散の形である。σ_2は負になることがある。例えば$\{(x_i, u_i)\}_5 = \{(20, 1), (21, -1), (21, -1), (21, -1), (22, 1)\}$とおくと、

$$\mu = \{20\times1+21\times(-1)+21\times(-1)+21\times(-1)+22\times1\}/\{1+(-1)+(-1)+(-1)+1\} = -21/(-1) = 21 \qquad ㊽$$

$$\sigma_2 = \{(20-21)^2\times1+(21-21)^2\times(-1)+(21-21)^2\times(-1)+(21-21)^2\times(-1)+(22-21)^2\times1\}/\{1+(-1)+(-1)+(-1)+1\} = 2/(-1) = -2 \qquad ㊾$$

と負になる。

また、集団によっては0にもなる。例えば$\{(x_i, u_i)\}_3 = \{(20, 1), (20, -1), (23, 1)\}$とおくと

$$\mu = \{20\times1+20\times(-1)+23\times1\}/\{1+(-1)+1\} = 23 \qquad ㊿$$

$$\sigma_2 = \{(20-23)^2\times1+(20-23)^2\times(-1)+(23-23)^2\times1\}/\{1+(-1)+1\} = 0 \qquad ⑤①$$

と0になる。

これらについて、読者に疑問が湧くかもしれない。第1に、分散が負になるのは不自然だ、その意味も不可解だ、というものである。これに対しては次のように答えることができる。即ち各$u_i > 0$のときは確かにσ_2は負になることはない。しかし、u_iが1と－1をとり、写像として㊼を採用する限り、σ_2は正負に亙って生じうる。σ_2の意味は、この写像を設けた意味と同じである。なお、負の分散は、推測統計学において、分散の不偏推定量の実現値となることがあり、以前から知られているものである。第2の疑問は、x_iに変動があるのに、その分散が0になるのは不合理ではないか、というものである。これに対しては次のように答えることができる。この場合の分散0は、やはり前の疑問と同様、集団内の負の構成単位と写像の設定仕様により生じるものである。分散は正負零のいずれにもなり得て、すべて等価等質で、特に構成単位が正ばかり、あるいは負ばかりのときは、従来のようにxのばらつきの尺度として分かり易いものになるものである。第3の疑問は、σ_2が負になったときは、分散と呼ばず別の名で呼ぶべきではないか、というものである。これに対しては次のように答えることができる。呼び名は自由であるから、別の名で呼ぶのは差しつかえない。また、負の分散を避けたいなら、そうなるような写像を設定すればよい。それは別の研究課題である。負になった分散を別の名で呼ぶとすると、現に推測統計学で分散と呼ばれている分散の負の値も、呼び方を改めなければならないことになる。筆者はそれには反対で、㊼が負の値となっても分散と呼ぶこととしている。

10. 部分集団と写像の例――標本と統計量

次に図3における集団z'とr'の具体例を示す。

集団$z=\{(x_i,\ u_i)\}_N$を大きさNの母集団として、大きさnの標本を、等確率、非復元方式で抽出し、

$$z'=\{(X_i,\ U_i)\}_n \quad ⑫$$

を作る。z'はzの部分集団（標本）である。非復元抽出は、大きさNの母集団から1つずつ、元に戻さずn個を抽出するもので、標本は$_NP_n$通りできる。順序を考慮するから$_NC_n$通りではない。

次にz'から幾つかの写像を作り、集団r'を構成する。例えば

周辺分布Xの標本平均

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \quad ⑬$$

周辺分布Xの2乗の標本平均

$$\overline{X^2}=\frac{1}{n}\sum_{i=1}^{n}X_i^2 \quad ⑭$$

周辺分布Xの標本分散（分母n）

$$S_X^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 \quad ⑮$$

周辺分布Uの標本平均

$$\overline{U}=\frac{1}{n}\sum_{i=1}^{n}U_i \quad ⑯$$

周辺分布Uの2乗の標本平均

$$\overline{U^2}=\frac{1}{n}\sum_{i=1}^{n}U_i^2=1 \quad ⑰$$

周辺分布Uの標本分散（分母n）

$$S_U^2=\frac{1}{n}\sum_{i=1}^{n}(U_i-\overline{U})^2 \quad ⑱$$

同時分布XUの標本平均

$$\overline{X_U}=\frac{1}{n}\sum_{i=1}^{n}X_iU_i \quad ⑲$$

XのU加重標本平均（同時分布）

$$m=\frac{\sum_{i=1}^{n}X_iU_i}{\sum_{i=1}^{n}U_i} \quad ⑳$$

Xの、標本平均の周りのU加重標本積率（同時分布）

$$S_j=\frac{\sum_{i=1}^{n}(X_i-m)^jU_i}{\sum_{i=1}^{n}U_i} \qquad j=2,\ 3,\ \cdots,\ N \quad ㉑$$

これらのうち、⑬－⑱は周辺分布に関するもの、⑲－㉑は同時分布に関するもので、特に⑳は標本平均、㉑は2次以上の標本積率に相当し、特別の注意を向けたい。これら2つの式は、U_iがX_iのウェイトの位置にあり、U_iがすべて1またはすべて－1のとき、普通の標本平均や標本積率の式に戻る。

11. 期待値

こうして作った写像は集団r'を作る。各写像は統計量であり確率変数である。rとr'の関係は、r'の確率変動の特性であるr'の期待値や分散（これもr'の写像）を求めることによってさぐることができる。よってまず、上記の⑬以降の式について期待値を求め、㊳以降のいくつかの式と比べてみる。期待値は標本の実現値と実現確率の総積和であり、普通の方法で求めることができる。数値例は前記8で設定した大きさ$N=19$の集団を母集団とし、これより等確率、非復元抽出で大きさ$n=4$の標本を抽出し、期待値を数値計算して、母集団の特性値との関係をみようとするものである。標本が作る集団は$_{19}P_4=93042$個の実現値からなる。構成単位はすべて正である。1つの標本値が実現する確率は1/93042であり、期待値は標本の実現値の個々に1/93042をかけ、93042個に亙って合算することによって得られる。以下㉒－㉘の「数値例」とあるのは、こうして求めた値であ

る。

㊸の期待値

$$E(\overline{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n}(\sum_{i=1}^{N} x_i \cdot \frac{1}{N}) = \frac{1}{n}\sum_{i=1}^{n}\overline{x} = \overline{x}$$

㊳に一致、数値例 6.8842 1053 ㊷

㊹の期待値

$$E(\overline{X^2}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i^2) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^2)$$

$$= \frac{1}{n}\sum_{i=1}^{n}(\sum_{i=1}^{N} x_i^2 \cdot \frac{1}{N})$$

$$= \frac{1}{n}\sum_{i=1}^{n}\overline{x^2} = \overline{x^2} \qquad ㊸$$

㊴に一致、数値例 51.5896 6316

㊺の期待値

$$E(S_X^2) = E\{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\} = \text{(途中略)}$$

$$= \frac{N(n-1)\sigma_x^2}{(N-1)n} \qquad ㊹$$

㊵に関係、数値例 3.3328 6930

(=4.1973 0859×19×(4−1)/(19−1)/4)

㊻の期待値

$$E(\overline{U}) = E(\frac{1}{n}\sum_{i=1}^{n} U_i) = \frac{1}{n}\sum_{i=1}^{n} E(U_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\{1 \times p + (-1) \times (1-p)\}$$

$$= 2p - 1 \qquad ㊺$$

㊶に一致、数値例 0.8947 3684

㊼の期待値

$$E(\overline{U^2}) = E(\frac{1}{n}\sum_{i=1}^{n} U_i^2) = E(\frac{1}{n}\sum_{i=1}^{n} 1) = 1$$

㊷に一致、元々確率変動なし ㊻

㊽の期待値

$$E(S_U^2) = E(\frac{1}{n}\sum_{i=1}^{n}(U_i - \overline{U})^2 = \text{(途中略)}$$

$$= \frac{4N(n-1)p(1-p)}{(N-1)n} \qquad ㊼$$

㊸に関係、数値例 0.1578 9474

(=0.1994 4598×19×(4−1)/(19−1)/4)

㊾の期待値

$$E(\overline{X_U}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i U_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n}(\sum_{i=1}^{N} x_i u_i \cdot \frac{1}{N})$$

$$= \frac{1}{n}\sum_{i=1}^{n}\overline{x_u} = \overline{x_u} \qquad ㊽$$

㊹に一致、数値例 6.2705 2632

次は㊿の期待値を示す順番であるが、㊿では分母の変数が確率変動をもつから、これまでのように簡単でない。㊿や51の期待値の表現は前例がないので、結果に辿りつくまでの手順を、字数を惜しまず披露したい。そのため、及び紙数の事情もあり、残りの課題も含めて次号につなげることとする。

上の62−68の式は、それぞれ㊳−㊹の式に関連している。これは集団 r と r' の関係の一端を表している。r' から r を統計的に推測するには、これらの関係の状況を確率によって説明する必要がある。具体的には関係する統計量の確率分布を用いて推定や検定などを行うこととなるが、その方法は従来知られているとおりである。(続)

参考文献

(1)舩津好明 「負の分散について」ESTRELA 2000年9月号(No.78) 統計情報研究開発センター発行