数理解析研究所講究録
231巻 1975年 88-95

星座グラフによる多次元データの表記とその応用

岡山大　教養部　　脇本和昌

[概要] 多次元データ解析の一つのねらいに、そのデータの特徴をできるだけ失れないように、一次元または二次元に縮約して数値化または図式化することにより、データの特性を見つけることがある。
与えられた一つの n 次元データは、n 次元空間の一つの点として考えられ、$n=1$、2 の場合を除いては視覚によりその点を直接見ることはできないのが普通である。そこで n $\geq$ 3 の場合に対して、その点を二次元平面にある方法で射影して視覚により n 次元での点の位置の異り方を見分ける方法が、いろいろ考えられている。その主なものとしては、ダイヤグラム，Andrewsの方法[1]，Chernoffの方法[2] などがあるが、(詳解については、後藤[3]を参照していただきたい) これらすべて個々の点の表現は上手に出来ても、多くの点の集まりの傾向を見るには、点の数が少し多くなるとほとん

ど不可能となる。そこで、ここでは、個々の点の性質もわかり、しかも同時に全体の点の様子もわかるという意味で有効な方法と思われる星座グラフによる図式化の方法と、その応用例について解説する。

1. 星座グラフの作り方

(1). 表1のようなデータが与えられているとする。これを適当な方法により、0度から180度までの角度に変換する。(表2参照)。この場合、もとの特性値は連続量でも、カテゴリカルなものでも、この変換は可能である。適切な変換の方法は、与えられたデータの性質や分析の目的等により異なるが、具体的な意味づけが行い易いように選ぶべきであろう。一つの変換の例としては、$x_{i1}, x_{i2}, \cdots\cdots, x_{in}$ の最大値を Z_u, 最小値を Z_l とすれば、

$$(1) \qquad \alpha_{ij} = \frac{x_{ij} - Z_l}{Z_u - Z_l} \times 180(\text{度})$$

$$(i = 1, 2, \cdots\cdots, n)$$

のようなものも考えられる。

表1 もとのデータ

	1	2	⋯	j	⋯	k
1	x_{11}	x_{12}	⋯	x_{1j}	⋯	x_{1k}
2	x_{21}	x_{22}	⋯	x_{2j}	⋯	x_{2k}
⋮	⋮	⋮		⋮		⋮
i	x_{i1}	x_{i2}	⋯	x_{ij}	⋯	x_{ik}
⋮	⋮	⋮		⋮		⋮
n	x_{n1}	x_{n2}	⋯	x_{nj}	⋯	x_{nk}

表2 0～180度に変換したデータ

	1	2	⋯	j	⋯	k
1	α_{11}	α_{12}	⋯	α_{1j}	⋯	α_{1k}
2	α_{21}	α_{22}	⋯	α_{2j}	⋯	α_{2k}
⋮	⋮	⋮		⋮		⋮
i	α_{i1}	α_{i2}	⋯	α_{ij}	⋯	α_{ik}
⋮	⋮	⋮		⋮		⋮
n	α_{n1}	α_{n2}	⋯	α_{nj}	⋯	α_{nk}

(ロ)、表2のデータに対して、次のような変換を考える。

$$(2)\quad \begin{cases} u_i = \sum_{j=1}^{k} w_j \cos\alpha_{ij} \\ u_i' = \sum_{j=1}^{k} w_j \sin\alpha_{ij} \end{cases} \qquad (i = 1, 2, \cdots, n)$$

ただし、w_j $(j = 1, 2, \cdots, k)$ は、各特性の重みと考え

$$(3)\quad \sum_{j=1}^{k} w_j = 1, \quad w_j \geqq 0 \quad (j = 1, 2, \cdots, k)$$

を満足するものとする。

この変換により、(u_i, u_i') $(i = 1, 2, \cdots, n)$ の n 個の点を半円の中にプロットしたものを、<u>星座グラフ</u>と呼ぶことにする（図1，参照）。

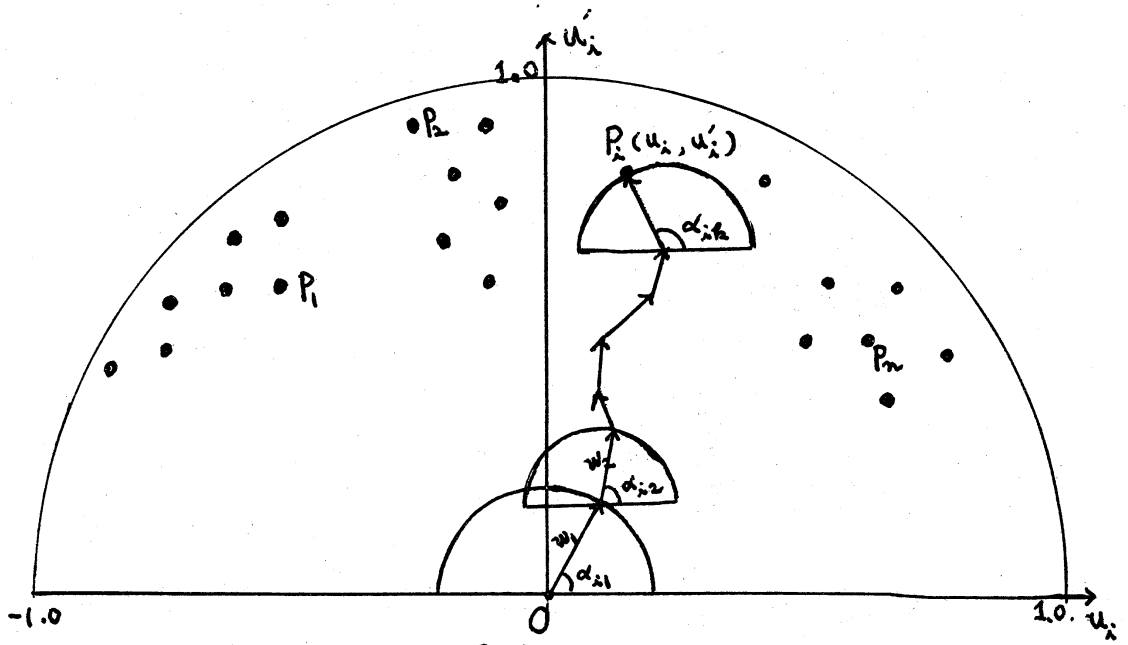

図 1. 星座グラフ

2、星座グラフの応用例

(1)、 判別分析への応用

脚本，田栗ヒサコにもその一部を示しているが、与えられたn個，m個の2つの群が、いずれも九次元のデータであるとする。すなわち、表2に相等する2つの群を

1群： $\alpha_{k1}, \alpha_{k2}, \cdots\cdots, \alpha_{km}$ $(k=1, 2, \cdots\cdots, n)$

2群： $\beta_{k1}, \beta_{k2}, \cdots\cdots, \beta_{km}$ $(k=1, 2, \cdots\cdots, m)$

とするとき、次のような変換を考える。

1群に対しては

(2) 式を用いる。

2群に対しては

$$(4)\quad \begin{cases} v_i = \sum_{j=1}^{n} w_j \cos \beta_{ij} \\ v_i' = \sum_{j=1}^{n} w_j \sin \beta_{ij} \end{cases} \qquad (i = 1, 2, \cdots, m)$$

ここで、(3) 式を満足する n 次元超平面上の第一象限の部分平面を S_n とする。S_n 上の一つの点 $(w_1, w_2, \cdots, w_n)$ に対して、(2) 式, (4) 式より星座の中に $n+m$ 個の点 $P_1, P_2, \cdots, P_n, Q_1, Q_2, \cdots, Q_m$ を作る (図2, 参照)。

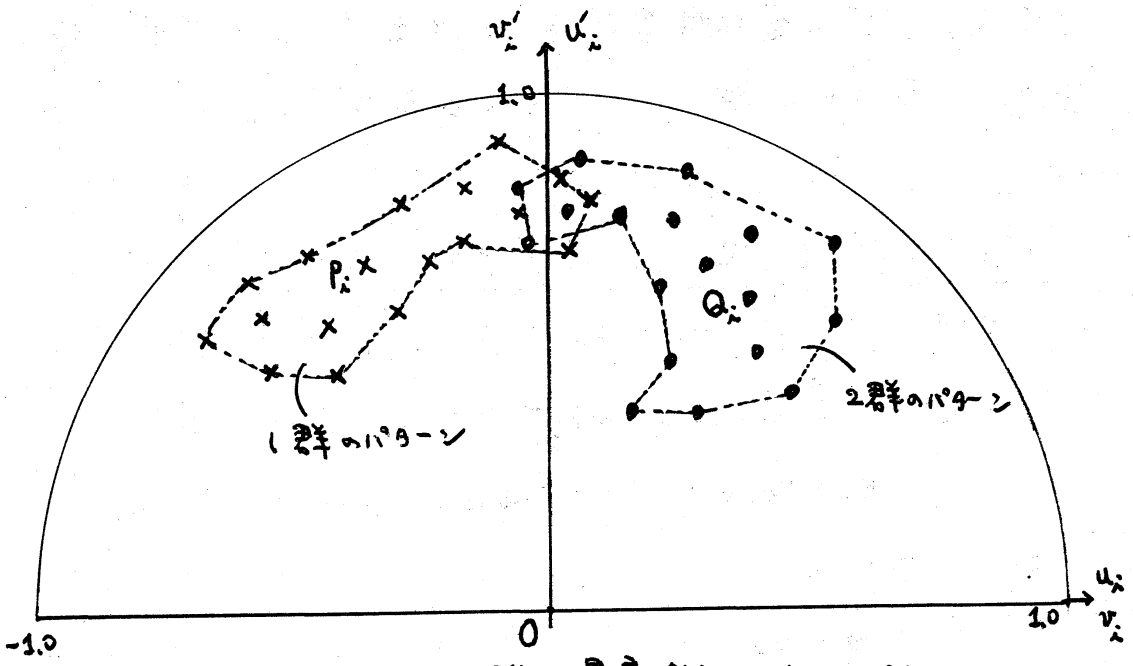

図2. 2つの群の星座グラフの中でのパターン

この図において、視覚により2つの群が最も判別できる点 $(w_1, w_2, \cdots, w_n)$ を探索する。このために、まず、S_n 上に l_1 個のランダムな点列を発生し、そのうちで、最もよく判別できる点の値を

$$\mathbf{w}^{(1)*} = (w_1^{(1)*}, w_2^{(1)*}, \cdots\cdots, w_n^{(1)*})$$

とする。

次に、点 $\mathbf{w}^{(1)*}$ の近傍でランダムな l_2 個の点列を発生し、そのうちで最もよく判別できる点の値を

$$\mathbf{w}^{(2)*} = (w_1^{(2)*}, w_2^{(2)*}, \cdots\cdots, w_n^{(2)*})$$

とする。以下必要な精度が得られるまで、この手順を繰り返す。具体的例題については当日報告する。

ランダムな L_1 個，L_2 個の点列の作り方

(イ)、区間（0，1）上の $k-1$ 個の一様乱数を $u_1, u_2, \cdots\cdots, u_{k-1}$ とする。これを大きさの順に並べかえて、

$$u_{(1)} \leqq u_{(2)} \leqq \cdots\cdots \leqq u_{(k-1)}$$

とする。これより

$$w_1 = u_{(1)}$$

$$w_j = u_{(j)} - u_{(j-1)} \quad (j = 2, 3, \cdots, k-1)$$

$$w_k = 1 - u_{(k-1)}$$

とおけば、$(w_1, w_2, \cdots\cdots, w_k)$ が S_k 上の一つのランダム点となる。これを L_1 個つくればよい。

(ロ)、

$$\begin{cases} \sum_{j=1}^{k} w_j' = a, & w_j' \geqq 0 \quad (j = 1, 2, \cdots, k) \\ \sum_{j=1}^{k} w_j'' = a, & w_j'' \geqq 0 \quad (j = 1, 2, \cdots, k) \end{cases} \qquad 0 < a < 1$$

を満足する k 次元超平面 S_k' 上の独立な2つのランダム点を作る。この作り方は、(イ) と同様にすればよい。

次に、

$$w_j^{(2)} = w_j^{(1)*} + w_j' - w_j'' \quad (j = 1, 2, \cdots, k)$$

により、$w^{(2)}$ を作る。$w_j^{(2)}$ が負になれば採用しない。

このようにして出来た点が、(3)式を満足していることは明らかであり、このような点をL2個作れば、これは $w^{(0)*}$ の近傍をランダムに探していることになり、その中で最も判別度のよいものを $w^{(0)*}$ とすればよい。a の値は、どの程度の近傍を探すかによって適当に与えればよい。

(2)、星座式シンオメトリー，その他への応用についても報告する。

参考文献

[1] Andrews, D.F. (1972). Plots of high-dimensional data. Biometrics, vol.28, pp.125-36.

[2] Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically. J. Amer. Statist. Ass., vol.68, No.342, pp.361-68.

[3] 後藤昌司（1974）．多変量データの解析における図的表現。日科技連、第4回官能検査シンポジウム講演録

[4] 脇本和昌，田栗正章（1974）．2次元図式パターンを用いる判別分析。応用統計学 3巻3号