ケースの分類

判別分析におけるケースの分類法としては 2 通りある。

1　マハラノビス汎距離による方法

この方法では，ケースと各グループの重心までのマハラノビス汎距離を計算し，もっとも近いグループに判別する。このとき使われるものが分類関数 classification function である。

k 個の群の母集団の平均を $\boldsymbol{\mu}_j = (\mu_{1j},\ \mu_{2j},\ \ldots,\ \mu_{pj})'$（$j = 1,\ 2,\ \ldots,\ k$），観測値を $\boldsymbol{X} = (X_1,\ X_2,\ \ldots,\ X_p)'$ とする。

各群の分散共分散行列を $\boldsymbol{\Sigma}_j$，その逆行列を $\boldsymbol{\Sigma}_j^{-1}$ とするき，(1) 式による各群までの**マハラノビス距離**を計算し，最も近い群に属すると判定する。

$$d_j^2 = (\boldsymbol{X} - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{X} - \boldsymbol{\mu}_j) \tag{1}$$

もし，各群の分散共分散行列が等しい，すなわち $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \cdots = \boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}$ が仮定できれば (2) 式のようになる。

$$\begin{aligned} d_j^2 &= (\boldsymbol{X} - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\mu}_j) \\ &= \boldsymbol{X}' \boldsymbol{\Sigma}^{-1} \boldsymbol{X} - 2\, \boldsymbol{X}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j + \boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j \end{aligned} \tag{2}$$

第 1 項は各群に共通，第 3 項は各群ごとに異なる定数（これを c_j とする）である。各ケースごとに異なるのは第 2 項のみであるから，(3) 式の計算を行えばよい。

$$2\, \boldsymbol{X}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j = a_{1j} X_1 + a_{2j} X_2 + \ldots + a_{pj} X_p \tag{3}$$

係数 $a_{1j},\ a_{2j},\ \ldots,\ a_{pj}$ は $\boldsymbol{\Sigma}^{-1}$ の要素を σ^{ij} とすれば (4) 式で求められる。

$$a_{ij} = 2\,(\sigma^{i1} \mu_{1j} + \sigma^{i2} \mu_{2j} + \ldots + \sigma^{ip} \mu_{pj}), \qquad i = 1, 2, \ldots, p \tag{4}$$

(2) 式の第 1 項は群に関係ないため無視できるので，(5) 式の数値が最も小さい群に属すると判定すればよい。(5) 式は，**分類関数**と呼ばれる。

$$f_j = a_{1j} X_1 + a_{2j} X_2 + \ldots + a_{pj} X_p + c_j \tag{5}$$

また，マハラノビス距離の大小を比較する代わりにあらゆる 2 群の組合わせに対して，(6) 式で表される ${}_kC_2$ 個の**判別関数**を定義しておくこともできる。例えば，第 1 群と第 2 群の判別関数は，

$$\begin{aligned} Z_{1\cdot 2} &= d_1^2 - d_2^2 = f_1 - f_2 \\ &= (a_{11} - a_{12}) X_1 + (a_{21} - a_{22}) X_2 + \cdots + (a_{p1} - a_{p2}) X_p + (c_1 - c_2) \end{aligned} \tag{6}$$

(6) 式の判別関数は群の数が 2 群のときは 1 個の判別関数を計算すればよいので便利であるが，群の数が 3 群以上のときには，分類関数を使う方がわかりやすい。

2　ベイズの定理による方法

各グループの事前確率が与えられていれば，ベイズの定理を適用することにより，ケースがどのグループに属するかという事後確率によってケースの分類を行うことができる。

グループ G_g の事前確率を P_g，G_g の判別得点の平均ベクトルを $\bar{\boldsymbol{y}}_g$ とする。また，全体での平均ベクトルを $\bar{\boldsymbol{y}}$，プールされた判別得点の分散・共分散行列を $\hat{\boldsymbol{\Sigma}}_y$ とする。このとき，任意のケース s の判別得点が $\boldsymbol{y}_s$ であるならば，元のデータが多変量正規分布に従うことと，各グループの分散・共分散行列が同等であることを仮定すれば，ケース s がグループ G_g に属する確率 $\Pr(G_g|\boldsymbol{y}_s)$ は，

$$\Pr(G_g|\boldsymbol{y}_s) = \alpha P_g \exp[-(\boldsymbol{y}_s - \bar{\boldsymbol{y}}_g)'\hat{\boldsymbol{\Sigma}}_y^{-1}(\boldsymbol{y}_s - \bar{\boldsymbol{y}}_g)/2] \tag{7}$$

となり，定数 α を $\sum_{j=1}^{k} \Pr(G_j|\boldsymbol{y}_s) = 1$ となるように定義すると，$\Pr(G_g|\boldsymbol{y}_s)$ は，

$$\Pr(G_g|\boldsymbol{y}_s) = \frac{P_g \exp[-(\boldsymbol{y}_s - \bar{\boldsymbol{y}}_g)'\hat{\boldsymbol{\Sigma}}_y^{-1}(\boldsymbol{y}_s - \bar{\boldsymbol{y}}_g)/2]}{\sum_{j=1}^{k} P_j \exp[-(\boldsymbol{y}_s - \bar{\boldsymbol{y}}_j)'\hat{\boldsymbol{\Sigma}}_y^{-1}(\boldsymbol{y}_s - \bar{\boldsymbol{y}}_j)/2]} \tag{8}$$

となる。この確率がもっとも高いグループに判別すればよい。

なお，ケース s がグループ G_g に属するとき，判別得点 $\boldsymbol{y}_s$ をとる確率 $\Pr(\boldsymbol{y}_s|G_g)$ は，自由度 p のカイ二乗分布に従い，

$$\Pr(\boldsymbol{y}_s|G_g) = \Pr\{\chi^2 \leqq (\boldsymbol{y}_s - \bar{\boldsymbol{y}}_g)'\hat{\boldsymbol{\Sigma}}_y^{-1}(\boldsymbol{y}_s - \bar{\boldsymbol{y}}_g)\} \tag{9}$$

となる。