分類について — I. 二群への分類

藤 本 熙

(1963 年 6 月 受 付)

On the classification—I. The case of two populations

Hirosi HUDIMOTO

Summary: In the section 1, the situations and the procedures of the classification are generally discussed, and in the section 2, the results that have already known for the case of two multivariate normal distributions are summarized.

In the section 3, when $(X_1, \cdots, X_n)$ is a random sample obtained from either population Π_1 or population Π_2 with the one dimensional continuous distribution functions, respectively, the order statistics are used for the classification procedures. At the first step, if it is assumed that the distribution function $F_1(x^{(1)})$ of Π_1 and the distribution function $F_2(x^{(2)})$ of Π_2 are completely known and $F_1(t)>F_2(t)$ for all t, our decision rule in this classification is as follows:

The decision rule 1. Assign $(X_1, \cdots, X_n)$ to Π_1 or Π_2 according to

$$\sum_{k=1}^{n}\{F_1(X_k)+F_2(X_k)\}<n \text{ or } \sum_{k=1}^{n}\{F_1(X_k)+F_2(X_k)\}>n.$$

If $p-\dfrac{1}{2}=\Delta$, where $p=\int_{-\infty}^{\infty}F_1(t)dF_2(t)$, the success rate for this classification procedure is greater than $\left(1-2e^{-2n\left(\frac{\Delta}{2}\right)^2}\right)$.

If $F_1(x^{(1)})$ and $F_2(x^{(2)})$ are unknown, a straightforward generalization of this method is to consider the uniformly minimum variance unbiased estimate of p. Let $(X_1^{(1)}, \cdots, X_{n_1}^{(1)})$ and $(X_1^{(2)}, \cdots, X_{n_2}^{(2)})$ be the random samples obtained from Π_1 and Π_2, respectively. Denote by U_{n_1,n_2} the number of pairs $(X^{(1)}, X^{(2)})$ such that $X^{(1)}<X^{(2)}$ in the sample sizes n_1 and n_2, by U_{n_1+n,n_2} the number in the sample sizes n_1+n and n_2 when $(X_1, \cdots, X_n)$ is assumed to be a sample of Π_1, and by U_{n_1,n_2+n} the number in the sample sizes n_1 and n_2+n when $(X_1, \cdots, X_n)$ is assumed to be a sample of Π_2, respectively. In this case, our decision rule is as follows:

Decision rule 2. When $U_{n_1,n_2}/n_1n_2>\dfrac{1}{2}$, assign $(X_1, \cdots, X_n)$ to Π_1 or Π_2 accoring to $U_{n_1,n_2+n}/n_1(n_2+n)<U_{n_1+n,n_2}/(n_1+n)n_2$ or $U_{n_1,n_2+n}/n_1(n_2+n)>U_{n_1+n,n_2}/(n_1+n)n_2$. When $U_{n_1,n_2}/n_1n_2<\dfrac{1}{2}$, assign $(X_1, \cdots, X_n)$ to Π_2 or Π_1 according to the above described inequality sign relations.

In the section 4, we treat the case that observations are the dichotomous response vectors. If it is assumed that all of the correlations among the items more than two items are zero, the joint distribution of the response vector may be approximately normally distributed, but their covariance matrix is not generally expected to be equal. We consider the classification procedure in such a case.

Institute of Statistical Mathematics

§1. 二群への分類の一般形式

分類の手続きは，通常次のような一般的な形式で述べられる．議論の便から，はじめに二群への分類を取扱う．

測定値 $\boldsymbol{x}'=(x_1, \cdots, x_p)$ は確率分布 $F_1(\boldsymbol{x})$ あるいは，$F_2(\boldsymbol{x})$ で特徴づけられる集団 Π_1，あるいは，Π_2 のいずれかからの観測結果であるとする．このとき測定値を p-次元空間の一点と考え，この空間，即ち標本空間 R を重なり合わない二つの領域 R_1, R_2 に分離し，もし，$\boldsymbol{x}\in R_1$ なら Π_1 へ，$\boldsymbol{x}\in R_2$ なら Π_2 へ分類する．ただし $\boldsymbol{x}'$ は縦ベクトル $\boldsymbol{x}$ の転置を意味する．

いま，簡単のため Π_1 の密度関数を $f_1(\boldsymbol{x})$, Π_2 の密度関数を $f_2(\boldsymbol{x})$ とすると，$\boldsymbol{x}$ を誤らず分類する確率はそれぞれ，

$$P(\boldsymbol{x}\in R_1|\Pi_1)=\int_{R_1} f_1(\boldsymbol{x})d\boldsymbol{x}, \tag{1.1}$$

$$P(\boldsymbol{x}\in R_2|\Pi_2)=\int_{R_2} f_2(\boldsymbol{x})d\boldsymbol{x}, \tag{1.2}$$

であり，$\boldsymbol{x}$ の分類を誤る確率はそれぞれ

$$P(\boldsymbol{x}\in R_1|\Pi_2)=\int_{R_1} f_2(\boldsymbol{x})d\boldsymbol{x}, \tag{1.3}$$

$$P(\boldsymbol{x}\in R_2|\Pi_1)=\int_{R_2} f_1(\boldsymbol{x})d\boldsymbol{x} \tag{1.4}$$

であるが，ここでこの二種類の誤り，即ち Π_1 からの $\boldsymbol{x}$ を Π_2 に，Π_2 からの $\boldsymbol{x}$ を Π_1 に分類し誤る損失 $l(2|1)$, $l(1|2)$（非負の実数を仮定）と，$\boldsymbol{x}$ が Π_1 あるいは Π_2 からのものである事前確率 ω_1, ω_2 $(=1-\omega_1)$ がそれぞれ知られるならば，

$$l(2|1)\omega_1\int_{R_2} f_1(\boldsymbol{x})d\boldsymbol{x}+l(1|2)\omega_2\int_{R_1} f_2(\boldsymbol{x})d\boldsymbol{x} \tag{1.5}$$

を可能な限り小さくするように R_1 および R_2 を分離するのが最適な方法となる（Bayes の決定法則）．$l(2|1)=l(1|2)=1$ の場合，これは分類を誤る確率 $\omega_1\int_{R_2} f_1(\boldsymbol{x})d\boldsymbol{x}+\omega_2\int_{R_1} f_2(\boldsymbol{x})d\boldsymbol{x}$ である．

また与えられた測定値 $\boldsymbol{x}$ に対して，それが Π_i $(i=1, 2)$ からのものである事後確率は

$$\frac{\omega_i f_i(\boldsymbol{x})}{\omega_1 f_1(\boldsymbol{x})+\omega_2 f_2(\boldsymbol{x})} \tag{1.6}$$

であることから，測定値 $\boldsymbol{x}$ に対して高い事後確率をもつ Π_i へ $\boldsymbol{x}$ を割当てることによって分類を誤る確率を小さくすることができる．従ってこの場合の決定法則（$l(2|1)=l(1|2)$ の場合）は，$\omega_1 f_1(\boldsymbol{x})>\omega_2 f_2(\boldsymbol{x})$ なる領域を R_1, $\omega_1 f_1(\boldsymbol{x})<\omega_2 f_2(\boldsymbol{x})$ なる領域を R_2 とすることである．同様に一般の $l(2|1)$, $l(1|2)$ に対しては，

$l(2|1)\omega_1 f_1(\boldsymbol{x})>l(1|2)\omega_2 f_2(\boldsymbol{x})$ なる領域を R_1,

$l(2|1)\omega_1 f_1(\boldsymbol{x})<l(1|2)\omega_2 f_2(\boldsymbol{x})$ なる領域を R_2

とすることによって (1.5) を可能な限り小さくできることになる．例えば [1], [2] 参照．

§2. 共分散行列の等しい正規分布の場合

共分散行列 Σ の等しい正規分布 $N(\boldsymbol{\mu}^{(1)}, \Sigma)$, $N(\boldsymbol{\mu}^{(2)}, \Sigma)$ の場合の議論については，Wald [13] をはじめいろいろよく知られているが，その概要は次のようになろう．

$$f_i(\boldsymbol{x})=\frac{1}{(2\pi)^{\frac{1}{2}p}|\Sigma|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(i)})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(i)})\right] \quad (i=1,\ 2), \tag{2.1}$$

ただし $\boldsymbol{x}'=(x_1, \cdots, x_p)$, $\boldsymbol{\mu}^{(i)\prime}=(\mu_1^{(i)}, \cdots, \mu_p^{(i)})$ において その密度関数の比 $f_1(\boldsymbol{x})/f_2(\boldsymbol{x})$ の対数をとると,

$$\boldsymbol{x}'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})-\frac{1}{2}(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) \tag{2.2}$$

であるから，この場合の分類の最適な領域は，$\boldsymbol{x}'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})-\frac{1}{2}(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})>\log k$ を R_1, $\boldsymbol{x}'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})-\frac{1}{2}(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})<\log k$ を R_2 に割当てることとなる．また k は $\dfrac{\omega_2 l(1|2)}{\omega_1 l(2|1)}$ で与えられる．

またその標本分布 ($\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ および Σ 即知) について,

$$U=\boldsymbol{X}'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})-\frac{1}{2}(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) \tag{2.3}$$

の確率分布は $N(\boldsymbol{\mu}^{(1)}, \Sigma)$ のもとでは，期待値

$$\mathcal{E}(U|\Pi_1)=\frac{1}{2}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}), \tag{2.4}$$

分散

$$\mathrm{Var}\,(U|\Pi_1)=(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) \tag{2.5}$$

であるから，$N(\boldsymbol{\mu}^{(1)}, \Sigma)$, $N(\boldsymbol{\mu}^{(2)}, \Sigma)$ の間の Mahalanobis の距離を

$$(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})=d \tag{2.6}$$

で示すと，U の分布は $N\left(\frac{1}{2}d,\ d\right)$ となる．また $N(\boldsymbol{\mu}^{(2)}, \Sigma)$ に $\boldsymbol{X}$ が従うならば,

$$\mathcal{E}(U|\Pi_2)=-\frac{1}{2}d,\ \mathrm{Var}\,(U|\Pi_2)=d \tag{2.7}$$

で，U は $N\left(-\frac{1}{2}d,\ d\right)$ に従う．

また (2.2) 式の第一項 $\boldsymbol{x}'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})$ は

$$\frac{[\mathcal{E}(\boldsymbol{X}'\boldsymbol{a}|\Pi_1)-\mathcal{E}(\boldsymbol{X}'\boldsymbol{a}|\Pi_2)]^2}{\mathrm{Var}\,(\boldsymbol{X}'\boldsymbol{a})} \tag{2.8}$$

を $\boldsymbol{a}'=(a_1, \cdots, a_p)$ について最大ならしめるような一次関数であり，これはよく知られる一次判別関数である．

$N(\boldsymbol{\mu}^{(1)}, \Sigma)$ からの標本 $\boldsymbol{x}_1^{(1)}, \cdots, \boldsymbol{x}_{n_1}^{(1)}$, $N(\boldsymbol{\mu}^{(2)}, \Sigma)$ からの標本 $\boldsymbol{x}_1^{(2)}, \cdots, \boldsymbol{x}_{n_2}^{(2)}$ により $\boldsymbol{x}$ を Π_1 あるいは Π_2 に分類しようとする場合に，$\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$ の推定値として $\overline{\boldsymbol{x}}^{(1)}=\sum_{j=1}^{n_1}\boldsymbol{x}_j^{(1)}/n_1$, $\overline{\boldsymbol{x}}^{(2)}=\sum_{j=1}^{n_2}\boldsymbol{x}_j^{(2)}/n_2$, Σ の推定値として $(n_1+n_2-2)\boldsymbol{S}=\sum_{j=1}^{n_1}(\boldsymbol{x}_j^{(1)}-\overline{\boldsymbol{x}}^{(1)})(\boldsymbol{x}_j^{(1)}-\overline{\boldsymbol{x}}^{(1)})'+\sum_{j=1}^{n_2}(\boldsymbol{x}_j^{(2)}-\overline{\boldsymbol{x}}^{(2)})(\boldsymbol{x}_j^{(2)}-\overline{\boldsymbol{x}}^{(2)})'$ で定義される $\boldsymbol{S}$ を用いようとするのは自然であろう．(2.2) 式のパラメター $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ および Σ をこの推定値でおきかえると

$$V=\boldsymbol{x}'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}^{(1)}-\overline{\boldsymbol{x}}^{(2)})-\frac{1}{2}(\overline{\boldsymbol{x}}^{(1)}+\overline{\boldsymbol{x}}^{(2)})'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}^{(1)}-\overline{\boldsymbol{x}}^{(2)}) \tag{2.9}$$

であるが，(2.9) 式の第1項ははじめ Fisher [5] によって議論された標本にもとずく一次判別関数である．即ち“級内分散”に対して“級間分散”を最大にするような一次関数である．

V の極限分布が U の分布であることは，はじめ Wald [13] によって示された．なおまた，Wald はそこで $\boldsymbol{x}'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}^{(1)}-\overline{\boldsymbol{x}}^{(2)})$ の exact な標本分布をも扱った．その後 Sitgreaves [10], Harter [6], T. W. Anderson [2] 等が矢張りこの問題を考察しているが，いずれもなかなかに面倒なものになる．最近では S. John [8] が Σ 既知を仮定して同じ問題を扱ってはいるが，これでも余り簡単な形のものにはならない．

このような訳で分類の問題に現われる統計量の分布が，共分散行列一定の正規分布を仮定しても面倒（あるいは，その標本分布に未知パラメターを含む）なために，分析の小標本的取扱いを困難にすると考えられなくもないが，Solomon [11] の記述をかりれば，“未知ではあるが等しい共分散行列の多次元正規分布という事情は the most sophisticated treatment で，正確にはほとんど非現実的である.”——ことになるかもしれない．確かに心理学的あるいは社会学的な問題を扱っている彼について，それは正当な不満であるに違いない．だがしかし，測定値 $\boldsymbol{x}$ の成分の一次式 $\boldsymbol{x}'\boldsymbol{a}$ の漸近的正規性はその近づき方に問題があるとしても自然な事柄である．と同時に分類の基準を標本 $\boldsymbol{x}_1^{(1)}, \cdots, \boldsymbol{x}_{n_1}^{(1)}$ および $\boldsymbol{x}_1^{(2)}, \cdots, \boldsymbol{x}_{n_2}^{(1)}$ から定めようとする場合には，あらたな $\boldsymbol{x}$ を Π_1, Π_2 に割当てた後に，それを以後の分析にどう利用するかがさらに考察されねばならぬはずである．§3 では，この点に留意して幾分異った観点から二群への分類の方法を考えてみることにする．

§3. 順序統計量による分類

$F_1(t)=P(X^{(1)}\leqq t)$，$F_2(t)=P(X^{(2)}\leqq t)$ は一次元の確率変数 $X^{(1)}$，$X^{(2)}$ の分布関数で，簡単のためそれは連続であるとしよう．また分布の型についての仮定はおかないが，すべての t に対して $F_1(t)>F_2(t)$ をはじめ仮定する．大きさ n の標本 $(X_1, \cdots, X_n)$ が $F_1(t)$ あるいは，$F_2(t)$ いずれかからの標本であることが，あらかじめ知られているとき，これを $F_1(t)$ を分布関数とする Π_1，あるいは $F_2(t)$ を分布関数とする Π_2 へ割当てることを今までとは幾分異る観点から扱ってみる．

今 $(X_1, \cdots, X_n)$ の順序統計量 $X_{(1)}<\cdots<X_{(n)}$ により，経験分布関数

$$\hat{F}_n(t)=\begin{cases} 0 & (t<X_{(1)} \text{ に対して}), \\ \dfrac{k}{n} & (X_{(k)}\leqq t<X_{(k+1)} \text{ に対して}). \\ 1 & (X_{(n)}\leqq t \text{ に対して}). \end{cases} \tag{3.1}$$

をつくり，

$$P(X^{(1)}<X^{(2)})=\int_{-\infty}^{\infty}F_1(t)dF_2(t)=p \tag{3.2}$$

との比較にこの経験分布 $\hat{F}_n(t)$ を利用する．$F_1(t)>F_2(t)$ を仮定したから，$p>\dfrac{1}{2}$ である．勿論 $F_1(t)<F_2(t)$ を仮定すれば，$p<\dfrac{1}{2}$ であるから，議論の簡便化の点以外で不等号の向きは本質的ではないが，この比較を次のように行なう．

$$\hat{p}_1=\text{est.}\,P(X^{(1)}<X)=\frac{1}{n}\sum_{k=1}^{n}F_1(X_k), \tag{3.3}$$

ただし est. $P(\cdot\cdot)$ は $P(\cdot\cdot)$ の推定量を意味する．また

$$\hat{p}_2=\text{est.}\,P(X<X^{(2)})=1-\frac{1}{n}\sum_{k=1}^{n}F_2(X_k) \tag{3.4}$$

そこでこの期待値を考えると

$$\mathcal{E}(\hat{p}_1|\Pi_1)=\int_{-\infty}^{\infty}F_1(t)dF_1(t)=\frac{1}{2},$$

$$\mathcal{E}(\hat{p}_1|\Pi_2)=\int_{-\infty}^{\infty}F_1(t)dF_2(t)=p.$$

また

$$\mathcal{E}(\hat{p}_2|\Pi_1)=1-\int_{-\infty}^{\infty}F_2(t)dF_1(t)=\int_{-\infty}^{\infty}F_1(t)dF_2(t)=p$$

$$\mathcal{E}(\hat{p}_2|\Pi_2)=\frac{1}{2}.$$

そこでこの $(X_1, \cdots, X_n)$ を Π_1 あるいは Π_2 へ割当てる決定法則として，

決定法則 1. すべての t に対して $F_1(t)>F_2(t)$ であるならば，$\hat{p}_1 \lesseqgtr \hat{p}_2$ に従って $(X_1, \cdots, X_n)$ を Π_1 あるいは Π_2 に割当てる．

ここで $(p<\hat{p}+\epsilon)$ なる事象を E_1, $\left(\frac{1}{2}>\hat{p}-\epsilon\right)$ なる事象を E_2 とするならば，

$$P(E_1 \cap E_2)=1-P(\bar{E}_1 \cup \bar{E}_2) \geqq 1-\sum_{i=1}^{2} P(\bar{E}_i), \tag{3.5}$$

ただし $\epsilon>0$, $\bar{E}_i$ は E_i の余事象である．また

$$P\{p<\hat{p}+\epsilon\}>1-e^{-2n\epsilon^2} \text{ (例えば [3])} \tag{3.6}$$

を利用して，$p-\frac{1}{2}=\Delta$ とすると，この決定法則の成功率は $>1-2e^{-2n\left(\frac{\Delta}{2}\right)^2}$ となる．

ここでもし，Π_1, Π_2 の確率分布が知られないならば，Π_1 からの標本 $(X_1^{(1)}, \cdots, X_{n_1}^{(1)})$ と Π_2 からの標本 $(X_1^{(2)}, \cdots, X_{n_2}^{(2)})$ の順序統計量 $X_{(1)}^{(1)}< \cdots <X_{(n_1)}^{(1)}$ および $X_{(1)}^{(2)}< \cdots <X_{(n_2)}^{(2)}$ から経験分布関数 $\hat{F}_{n_1}^{(1)}$, $\hat{F}_{n_2}^{(2)}$ をつくり，これを用いて $p=\int F_1(t)dF_2(t)$ の推定を実行するのが自然であるであろう．しかしながら，これは

$$U_{n_1,n_2}=\{X_i^{(1)}<X_j^{(2)} \text{ なる組 } (X_i^{(1)}, X_j^{(2)}) \text{ の数}\} \tag{3.7}$$

をかぞえて，

$$\hat{p}=\frac{U_{n_1n_2}}{n_1n_2} \tag{3.8}$$

を p の推定量にすることである．$\hat{p}$ は p の一様最小分散不偏推定量で，かつ p の一致推定量でもある [4], [9].

ここで est. $P(X^{(1)}<X)$ に対応するものとして，

$$U_{n_1,n}=\{X_i^{(1)}<X_j \text{ なる組 } (X_i^{(1)}, X_j) \text{ の数}\} \tag{3.9}$$

$$\hat{\hat{p}}_1=\frac{U_{n_1,n}}{n_1n} \tag{3.10}$$

を考え，est. $P(X<X^{(2)})$ として

$$U_{n,n_2}=\{X_i<X_j^{(2)} \text{ なる組 } (X_i, X_j^{(2)}) \text{ の数}\} \tag{3.11}$$

$$\hat{\hat{p}}_2=\frac{U_{n,n_2}}{nn_2} \tag{3.12}$$

を考えると，前述の決定法則は，すべての t に対して $F_1(t)>F_2(t)$ を前提して，$\hat{\hat{p}}_1 \lesseqgtr \hat{\hat{p}}_2$ に従って $(X_1, \cdots, X_n)$ を Π_1 あるいは，Π_2 に割当てる．——ことになるが，今度は，$F_1(t)$, $F_2(t)$ は正確に知られている訳ではないので，一般に，$F(t)>F_2(t)$ を前提するのは無理である．しかしながら，仮りに，$(X_1, \cdots, X_n)$ を $(X_1^{(1)}, \cdots, X_{n_1}^{(1)})$ に合併して $(X_1^{(1)}, \cdots, X_{n_1}^{(1)}, X_1, \cdots, X_n)$ をあらためて Π_1 からの標本と見做し，U_{n_1+n,n_2} をつくると，

$$\frac{U_{n_1+n,n_2}}{(n_1+n)n_2}=\frac{U_{n_1,n_2}}{n_1n_2}+\frac{n_1U_{n,n_2}-nU_{n_1,n_2}}{(n_1+n)n_1n_2} \tag{3.13}$$

となる．また $(X_1, \cdots, X_n)$ を $(X_1^{(2)}, \cdots, X_{n_2}^{(2)})$ へ合併して，U_{n_1,n_2+n} をつくると，

$$\frac{U_{n_1,n_2+n}}{n_1(n_2+n)}=\frac{U_{n_1n_2}}{n_1n_2}+\frac{n_2U_{n_1n}-nU_{n_1n_2}}{(n_2+n)n_1n_2} \tag{3.14}$$

である．また，$\hat{\hat{p}}_1 \lesseqgtr \hat{\hat{p}}_2$ は $n_2U_{n_1,n} \lesseqgtr n_1U_{n,n_2}$ とかけ，こうであれば近似的に $U_{n_1,n_2+n}/n_1(n_2+n) \lesseqgtr U_{n_1+n,n_2}/(n_1+n)n_2$ であるから，

決定法則 2. $U_{n_1n_2}/n_1n_2>\frac{1}{2}$ のときには，

$$\frac{U_{n_1,n_2+n}}{n_1(n_2+n)} \lesseqgtr \frac{U_{n_1+n,n_2}}{(n_1+n)n_2} \tag{3.15}$$

に従って Π_1 あるいは Π_2 へ割当て，$U_{n_1,n_2}/n_1n_2<\frac{1}{2}$ のときには，(3.15) 式に従って Π_2 あるいは Π_1 へ割当てる．ただしこのときには n_1，n_2 の大きさについての配慮が更に必要である．

という決定法則が考えられる．

ここでこの決定法則 2. について少しばかり考察を続けることにする．今 $p=\int F_1(t)dF_2(t)>\frac{1}{2}$ としてみる．このとき確かに分類が正しく実行されたとすると，

$$\mathcal{E}\left(\frac{U_{n_1+n,n_2}}{(n_1+n)n_2}\middle|(x_1,\ \cdots,\ x_n)\in\Pi_1\right)$$
$$=\mathcal{E}\left(\frac{U_{n_1,n_2+n}}{n_1(n_2+n)}\middle|(x_1,\ \cdots,\ x_n)\in\Pi_2\right)=p.$$

また誤った分類を実行したとすると

$$\mathcal{E}\left(\frac{U_{n_1+n,n_2}}{(n_1+n)n_2}\middle|(x_1,\ \cdots,\ x_n)\in\Pi_2\right)<p.$$
$$\mathcal{E}\left(\frac{U_{n_1,n_2+n}}{n_1(n_2+n)}\middle|(x_1,\ \cdots,\ x_n)\in\Pi_1\right)<p$$

となる．即ち $p>\frac{1}{2}$ の場合には，その期待値をなるべく小さくしないように分類し，同様に $p<\frac{1}{2}$ の場合には期待値を増加させぬように分類しようとするのである．

§4. 二者択一型の場合の取扱い

例えばある事柄への適，不適を各個体の数種類の特性から判別しようとする場合，それが質的に直接にはその特性の有無だけしか得られないとすれば，その反応は二者択一型 (dichotomous type) である．今 p 種類の特性のそれぞれに適応の反応を 1，不適応の反応を 0 で示し，j 番目の特性で反応 1，0 を実現値とする確率変数を X_j とすると，$\boldsymbol{X}'=(X_1,\ \cdots,\ X_p)$ はその反応ベクトルである．

一般にこの反応ベクトルの実現値は 2^p 個の値が可能であるが，Π_i $(i=1,\ 2)$ のもとでのその確率を反応ベクトル (0, 0, …, 0) からはじめて (1, …, 1) に至る任意の順序でその k 番目を $f_k^{(i)}=f(x_1,\ \cdots,\ x_p|\Pi_i)$ で示すことにする．またその周辺分布の確率を $\lambda_j^{(i)}=P(X_j=1|\Pi_i)$ とする．Bahadur の表現に従えば，

$$U_j^{(i)}=(X_j-\lambda_j^{(i)})/\sqrt{\lambda_j^{(i)}(1-\lambda_j^{(i)})}$$

とおいて

$$\rho_{j_1,j_2}^{(i)}=\mathcal{E}(U_{j_1}^{(i)}U_{j_2}^{(i)}),$$
$$\rho_{j_1,j_2,j_3}^{(i)}=\mathcal{E}(U_{j_1}^{(i)}U_{j_2}^{(i)}U_{j_3}^{(i)}),$$
$$\vdots$$
$$\rho_{1,2,\cdots,p}^{(i)}=\mathcal{E}(U_1^{(i)}U_2^{(i)},\ \cdots,\ U_p^{(i)}),$$

および

$$f^*(x_1,\ \cdots,\ x_p|\Pi_i)=\prod_{j=1}^{p}\lambda_j^{(i)x_j}\,(1-\lambda_j^{(i)})^{1-x_j}$$

とすると

$$f(x_1,\ \cdots,\ x_p|\Pi_i)=f^*(x_1,\ \cdots,\ x_p|\Pi_i)\times(1+\sum_{j_1<j_2}\rho_{j_1,j_2}^{(i)}u_{j_1}^{(i)}u_{j_2}^{(i)}+\cdots+\rho_{12\cdots p}^{(i)}u_1^{(i)}u_2^{(i)}\cdots u_p^{(i)}) \tag{4.1}$$

となる．ただし $u_j^{(i)}=(x_j-\lambda_j^{(i)})/\sqrt{\lambda_j^{(i)}(1-\lambda_j^{(i)})}$ である．

ここで特別な場合として $f(x_1,\ \cdots,\ x_p|\Pi_i)=f^*(x_1,\ \cdots,\ x_n|\Pi_i)$ は p 種類の特性の無相関な場合となるが，$\rho_{j_1,j_2,j_3}^{(i)}=\cdots=\rho_{1,2,\ldots,p}^{(i)}=0$ $(i=1,\ 2)$ の場合にはこれを正規分布で近似することが考え

られる．しかしこの場合に p 個の特性が真に判別に役立つためには，$\lambda_j^{(1)}\neq\lambda_j^{(2)}$ $(j=1,2,\cdots,p)$ であるべきだから，一般に共分散行列の等しいことは期待できない．だがこの場合にも §2 で触れた一次判別関数の利用をある観点から正当化する理由が考えられるならば便利である．というのはそれを判別に役立てるというだけでなく，p 種類の特性の判別に寄与する程度をも知りたいのが通例であるからで，時としては分析の手間をはぶく目的から，寄与の少ない特性を落して改めて考え直す場合もあろう．従ってこのような場合の第一近似としてでも，一次判別関数の利用が考えられはしまいか——という訳である．

このような問題への考察に [7] がある．また §3 で記述した方法に沿って考えれば，期待値 $\mathcal{E}(X^{(i)})=\mu_i$，分散 Var $(X^{(i)})=\sigma_i^2$ の一次元正規分布に従う確率変数 $X^{(i)}$ $(i=1,2)$ について $P(X^{(1)}<X^{(2)})$ を求めてみると

$$P(X^{(1)}<X^{(2)})=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\mu_2-\mu_1}{\sqrt{\sigma_1^2+\sigma_2^2}}} e^{-\frac{t^2}{2}}dt \tag{4.2}$$

であり，また期待値 $\mathcal{E}(\boldsymbol{X}^{(i)})=\boldsymbol{\mu}^{(i)}$，共分散行列 $\sum^{(i)}$ の p 次元正規分布に従う確率変数 $\boldsymbol{X}^{(i)}$ の各成分の一次式 $\boldsymbol{X}^{(i)\prime}\boldsymbol{a}$ の確率分布は期待値 $\boldsymbol{\mu}^{(i)\prime}\boldsymbol{a}$，分散 $\boldsymbol{a}'\sum^{(i)}\boldsymbol{a}$ の一次元正規分布であるから，$P(\boldsymbol{X}^{(1)\prime}\boldsymbol{a}<\boldsymbol{X}^{(2)\prime}\boldsymbol{a})$ を最大にするような $\boldsymbol{a}$ は

$$\max_{\boldsymbol{a}} \frac{[\mathcal{E}(\boldsymbol{X}^{(1)\prime}\boldsymbol{a})-\mathcal{E}(\boldsymbol{X}^{(2)\prime}\boldsymbol{a})]^2}{\mathrm{Var}\ (\boldsymbol{X}^{(1)\prime}\boldsymbol{a})+\mathrm{Var}\ (\boldsymbol{X}^{(2)\prime}\boldsymbol{a})} \tag{4.3}$$

から求まることになる．

Solomon [12] はこの種（二者択一型）の問題についての実例を分析して，判別の方法を比較している．彼の方法は実験結果を母集団と見做し，(4.1) 式に対応する相対頻度を実験から求め，それから尤度比 $L=f_k^{(1)}/f_k^{(2)}$ をつくり，$\alpha_c=\sum_{L\leqq c}f_k^{(1)}$，$\beta_c=\sum_{L>c}f_k^{(2)}$ なる (α_c,β_c) を c の値を変えて図表示したものを最良の基準に，他の方法と比較してみるのである．特に特性の数が $4(p=4)$ の場合の図（[12] の 423 頁，Fig. 3）には，一次判別関数利用の場合，$\rho_{j_1,j_2}=0$ の場合，$\rho_{j_1j_2j_3}=\rho_{1,2,3,4}=0$ の場合，単純和の場合との比較が行なわれるが，単純和の場合特に悪いということ以外には，余り顕著な傾向は認められそうにみえない．

次に蛇足ではあるが，判別のためだけなら尺度化（例えば一次判別関数を用いる）の無意味な場合を付加しよう．またそれは明らかなことであるかもしれないが，周辺分布の確率 $\lambda_j^{(i)}$ $(i=1,2,j=1,\cdots,p)$ から同時分布の確率（反応ベクトルの確率）が完全に規定される場合である．だがこの型はこの種の問題では時折おこる可能性がある．例えばある事柄に好意的か非好意的かの態度評価の目的で，p 種類の質問に賛成か反対かを問うものとする．そのときこの質問群の作成者は，主観的ではあるが，その事柄についてのある一般的な性格を念頭に，比較的賛成しやすいものから賛成しにくいものへ順位がつきやすく出題するであろう．換言すると，既に質問の提案者には主観的な質問の一次元化ができている場合である．しかもその結果が確かにそうなったとして，特に $\lambda_1^{(i)}>\lambda_2^{(i)}>\cdots>\lambda_p^{(i)}$ だったとすると，反応ベクトルは $(0,0,\cdots,0)$，$(1,0,\cdots,0)$，$\cdots$，$(1,\cdots,1,0)$，$(1,\cdots,1,1)$ とベクトルの左端から次第に 0を 1 に変えた $p+1$ 個の反応ベクトルからなるパタンが現われる．

今賛成者の群 Π_1 の事前確率 ω_1 と反対者の群 Π_2 の事前確率 ω_2 が知られるならば，

$$\min_j\,[\omega_1(1-\lambda_j^{(1)})+\omega_2\lambda_j^{(2)}]$$

はこのときの分類の誤りの可能な最小値である．しかしながら，一般の 2^p 個の反応ベクトルをもつパタンで

$$\begin{aligned}&\omega_1(1-\lambda_j^{(1)})+\omega_2\lambda_j^{(2)}\\ &=\omega_1\left(1-\sum_k^{2^p}x_jf_k^{(1)}\right)+\omega_2\sum_k^{2^p}x_jf_k^{(2)}\end{aligned}$$

であるから，$f_k^{(1)} \geqq f_k^{(2)}$ のとき $x_i=1$，$f_k^{(1)} \leqq f_k^{(2)}$ のとき $x_j=0$ となるのでなければ，周辺分布 $\lambda_1^{(i)}, \cdots, \lambda_p^{(i)}$ を与えたときの一般のパタンでの最適な分類の誤りの確率にはならない．換言すれば，周辺分布の確率を与えたときの可能なすべての反応パタンについての最適な分類の誤りの確率の最大値を与える．また前述の反応パタンの場合，その判別は反応ベクトルの単純和から行えることになるが，むしろ p 個の特性はその中で最も判別に役立つ特性を探すのに役立てられるだけである．

（統計数理研究所）

［参 考 文 献］

［1］ Anderson, T. W. (1958), Introduction to multivariate statistical analysis, New York, John Wiley and Sons., pp. 126–142.

［2］ ——— (1951), Classification by multivariate analysis, Psychometrika, Vol. 16, pp. 31–50.

［3］ Birnbaum, Z. and Tingey (1951), One-sided confidence contours for probability distribution functions, Ann. Math. Stat., Vol. 22, pp. 592–596.

［4］ Birnbaum, Z. W. (1956), On a use of the Mann-Whitney statistic, Proceeding of the Third Berkeley Symposium, Vol. I, pp. 13–17.

［5］ Fisher, R. A. (1936), The use of multiple measurements in taxonomix problems, Ann. Eugenics, Vol. 7, pp. 179–188.

［6］ Harter, H. L. (1951), On the distribution of Wald's classification statistic, Ann. Math. Stat., Vol. 22, pp. 58–67.

［7］ 林知己夫 (1954) 数量化理論の応用例——予測の判断的中率と相関比との関係についての一つの考察と共に——，統計数理研究所彙報 Vol. 2, pp. 11–30.

［8］ John, S. (1960), On some classification statistics, Sankhya, Vol. 22, pp. 309–316.

［9］ Lehmann, E. L. (1951), Consistency and unbiasedness of certain nonparametric tests, Ann. Math. Stat., Vol. 22, pp. 165–179.

［10］ Sitgreaves, R. (1952), On the distribution of two random matrices used in classification procedures, Ann. Math. Stat., Vol. 23, pp. 263–270.

［11］ Solomon, H. (1956), Probability and statistics in psychometric research: item analysis and classification techniques, Proceeding of the Third Berkeley Symposium, Vol. V, pp. 169–184.

［12］ ——— (1960), Classification procedures based on dichotomous response vectors, Contribution to probability and statistics, pp. 414–423.

［13］ Wald, A. (1944), On a statistical problem arising in the classification of an individual into one of two groups, Ann. Math. Stat., Vol. 15, pp. 145–163.