北海道大学工学部研究報告
第126号（昭和60年）

Bulletin of the Faculty of Engineering,
Hokkaido University, No. 126 (1985)

判別群が時間によって変化する場合の判別分析について

中　西　寛　子　　河　口　至　商
（昭和59年12月27日受理）

On the Discriminant Analysis when the Discriminant Groups Vary with Time

Hiroko NAKANISHI and Michiaki KAWAGUCHI
(Received December 27, 1984)

Abstract

We sometimes find in the discriminant analysis that a discriminant coefficient for each variable and hence a discriminant score for each sample are unstable; because of the difficulty to identify what sample come from which group. The purpose of this paper is to investigate the effect of the samples, that change the groups that they belong to from time to time, on the results in the discriminant analysis. In order to study this effect, a model with the multivariate normal distribution is considered. The condition in which such samples have no effect on the discriminant coefficients and hence the discriminant scores can be found. Even under this condition, however, the discriminant boundary values are changed. Without this condition, we can find that there exists a serious problem in this kind of the discriminant analysis.

1. はじめに

医学分野をはじめとして多くの分野で判別分析が用いられている。しかしながら，サンプルが事象Aに含まれるか否かということを明確に定義できないため，各変量に対する判別係数や各サンプルの判別得点が不安定となる場合がある。たとえば，健康者の集団検診後の追跡調査において次の様な状況が見られる。検診後，1年間に n_1 名が疾病Aを発症，さらに1年間に n_2 名，さらに……，と発症していくデータについて何年間内に発症した者を発症群からのサンプルとすべきかというような問題が生じている。この様な状況においては，検診年より x 年後までの発症者と非発症者に基づいて解析した判別分析の判別係数や判別得点が y 年後までの発症者と非発症者に基づいて解析した各々の値と非常に異なることがある。

本論文では x 年後から y 年後の間に非発症群から発症群に移行したサンプルが判別係数，判別得点，判別値にひきおこす影響についてモデルをたてて考察する。本論文で用いる判別分析の手法はFisherの線形判別関数法である。

情報工学専攻　情報数理工学第一講座

2. Fisher の線形判別関数

2つの母集団 π_i $(i=1, 2)$ における Fisher の線形判別関数[1] について説明する。k 変量ベクトル $\boldsymbol{x}=(x_1, \cdots, x_k)$ に対する線形モデル $z=a_1x_1+\cdots+a_kx_k$ のうち両母集団間の差を最も明確に説明する係数ベクトル $\boldsymbol{a}=(a_1,\cdots, a_k)$ は両母集団が各々，多変量正規分布 $\mathrm{N}(\boldsymbol{\mu}_i,\boldsymbol{\Sigma})$ $(i=1, 2)$ に従うとき

$$\boldsymbol{a}=(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\,\boldsymbol{\Sigma}^{-1} \tag{2. 1}$$

と表わされる。ここで，$\boldsymbol{\mu}_i(i=1, 2)$ は母集団 π_i の平均ベクトル，$\boldsymbol{\Sigma}$ は分散共分散行列である。Fisher の線形判別関数とは，ここで求められた係数ベクトル $\boldsymbol{a}$ によって作られる関数 $z(\boldsymbol{x})=\boldsymbol{a}\boldsymbol{x}'$ のことである。いったん係数ベクトル $\boldsymbol{a}$ が決定されると，各サンプル $\boldsymbol{x}_j(j=1, \cdots, N)$ に対して $z_j=\boldsymbol{a}\boldsymbol{x}'_j$ を求めることができる。この z_j をサンプル j の判別得点と呼ぶ。種々の目的によってある定数(判別値) h を決めることができ，$z_j \geqq h$ のときは，サンプル j を母集団 π_1 に含まれるものと判定，$z_j<h$ のとき母集団 π_2 に含まれるものと判定する。たとえば，両母集団の誤判別の確率を等しくする尤度法では判別値 $h_{\mathrm{M}}=(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)'/2$，誤判別の確率の合計を最小にするベイズ法では判別値 $h_{\mathrm{B}}=h_{\mathrm{M}}+\log(p_2/p_1)$ となる。ここで，p_i は母集団 π_i の事前確率である。また，係数ベクトルに関する重要なこととして“係数ベクトル $\boldsymbol{a}$ の定数倍は各サンプルの得点を定数倍するのみであって，各サンプルの得点の相対関係に何ら変化をおよぼさない”ということを注意しておく。その他，判別分析についての詳しい内容に関しては参考文献 2)，3) を参照されたい。

3. 具体例による問題の説明

本論文で取り扱う問題を具体的に説明する。ただし，説明を容易にするため，変量数，サンプル数は実際の調査より少なくしている。

ある年，健康者に対し4項目（肥満度，収縮期血圧，拡張期血圧，血清総コレステロール）の集団検診がなされた。そののち，疾病Aの発症についての追跡調査が行われた。検診より5年後，このデータをもとに疾病Aの予測をするために5年間に疾病Aを発症した者16名と非発症者，多数の中から対照群として32名を抽出し，線形判別関数を求めた。そのときの変量に対する係数と各個人の得点が表1（左），表2（左）に示されている。さらに，2年後（つまり，検診年より7年後），対照群の中から6名が新たに疾病Aを発症した。そこで，発症者22名と非発症者26名について線形判別関数を求めた。その結果が表1（右），表2（右）である。これら2組の結果からわかるように，非発症群の何名かが発症群に移行するような場合，移行したサンプルによって判別係数，判別得点，判別得点の順位が大きく左右される。この様な現象は，発症群の定義が困難

表1　5年後および7年後の発症群と非発症群に対する判別係数

項　　目	判別係数	
	5 年 後	7 年 後
x_1 肥満度×100	0.0878	0.0300
x_2 収縮期血圧	−0.0451	−0.0033
x_3 拡張期血圧	0.1585	0.0863
x_4 総コレステロール	0.0034	0.0042

表2　5年後および7年後の発症群と非発症群に対する判別得点とその順位

5年後				7年後			
判別得点		サンプル番号		判別得点		サンプル番号	
発症群	非発症群	発症群	非発症群	発症群	非発症群	発症群	非発症群
	2.82		37		4.43		37
	4.21		42		5.16		32
	4.30		34		5.16		34
	4.31		43		5.24		26
	4.53		36		5.43		43
	4.55		26		5.54		36
	4.78		22	5.57		22	
	4.85		32		5.58		31
	4.87		46		5.64		42
	4.97		40		5.87		40
	5.03		18		6.02		35
	5.15		41		6.14		41
5.24		15			6.21		46
	5.36		35	6.21		15	
	5.43		21	6.22		21	
	5.49		25	6.23		7	
	5.62		20	6.35		3	
	5.76		31		6.38		28
	5.78		45		6.42		47
5.78		13			6.46		38
	5.88		19	6.54		16	
	5.99		27		6.58		24
	6.04		30		6.70		27
	6.25		38		6.86		25
	6.27		47	6.92		18	
	6.35		39		6.93		30
	6.50		24	6.99		10	
	6.51		48		7.06		39
6.53		3		7.06		19	
	6.62		17		7.08		48
6.69		7		7.22		20	
7.10		16		7.25		17	
	7.24		28	7.34		13	
	7.31		44	7.42		9	
7.44		10			7.44		44
7.66		12			7.54		29
7.87		6		7.69		6	
7.93		9		7.79		8	
	8.29		29		7.79		45
8.33		14		7.85		12	
	8.38		23		7.89		23
8.46		8			7.92		33
	9.43		33	8.49		5	
9.55		5		8.53		11	
9.68		2		8.58		1	
9.80		11		8.93		14	
10.04		1		8.99		2	
12.88		4		10.72		4	

な医学データを始めとし少なからず見かけられる。本論文では，非発症群から発症群に移動したサンプルが判別分析の結果におよぼす影響をモデルをたてて考察する。

4. 多変量正規分布を用いた一般的なモデル

問題を数値的に示すため次の様なモデルを考察する。

ある時刻 $t'=t+\Delta t$ においても発症しなかった（$\bar{\mathrm{E}}$ と記す）母集団（π_1），ある時刻 t では発症しなかったが，時刻 t' で発症した（E と記す）母集団（π_2），時刻 t で発症した母集団（π_3）が各々，多変量正規分布 $\mathrm{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ $(i=1,\ 2,\ 3)$ に従うとする。また事前確率は p_i $(i=1,\ 2,\ 3)$ である。このとき時刻 t において非発症群，発症群の平均ベクトル $\boldsymbol{\mu}_{\bar{\mathrm{E}}}$, $\boldsymbol{\mu}_{\mathrm{E}}$, 分散共分散行列 $\boldsymbol{\Sigma}_{\bar{\mathrm{E}}}$, $\boldsymbol{\Sigma}_{\mathrm{E}}$ は各々，

$$\begin{aligned} t:\ & \boldsymbol{\mu}_{\bar{\mathrm{E}}}=p_1/(p_1+p_2)\boldsymbol{\mu}_1+p_2/(p_1+p_2)\boldsymbol{\mu}_2,\ \boldsymbol{\mu}_{\mathrm{E}}=\boldsymbol{\mu}_3, \\ & \boldsymbol{\Sigma}_{\bar{\mathrm{E}}}=p_1/(p_1+p_2)\boldsymbol{\Sigma}_1+p_2/(p_1+p_2)\boldsymbol{\Sigma}_2+p_1p_2/(p_1+p_2)^2(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)',\ \boldsymbol{\Sigma}_{\mathrm{E}}=\boldsymbol{\Sigma}_3 \end{aligned} \tag{4. 1}$$

である。また，時刻 t' における各々の値は次の様になる。

$$\begin{aligned} t':\ & \boldsymbol{\mu}_{\bar{\mathrm{E}}}=\boldsymbol{\mu}_1,\ \boldsymbol{\mu}_{\mathrm{E}}=p_2/(p_2+p_3)\boldsymbol{\mu}_2+p_3/(p_2+p_3)\boldsymbol{\mu}_3, \\ & \boldsymbol{\Sigma}_{\bar{\mathrm{E}}}=\boldsymbol{\Sigma}_1,\ \boldsymbol{\Sigma}_{\mathrm{E}}=p_2/(p_2+p_3)\boldsymbol{\Sigma}_2+p_3/(p_2+p_3)\boldsymbol{\Sigma}_3+p_2p_3/(p_2+p_3)^2(\boldsymbol{\mu}_2-\boldsymbol{\mu}_3)(\boldsymbol{\mu}_2-\boldsymbol{\mu}_3)'。 \end{aligned} \tag{4. 2}$$

これらより，時刻 t, t' における平均の差ベクトル $\boldsymbol{\delta}_t$, $\boldsymbol{\delta}_{t'}$ は，

$$t:\ \boldsymbol{\delta}_t=p_1/(p_1+p_2)\boldsymbol{\mu}_1+p_2/(p_1+p_2)\boldsymbol{\mu}_2-\boldsymbol{\mu}_3=\left(\sum_1^3 p_i\boldsymbol{\mu}_i-\boldsymbol{\mu}_3\right)/(p_1+p_2), \tag{4. 3}$$

$$t':\ \boldsymbol{\delta}_{t'}=\boldsymbol{\mu}_1-p_2/(p_2+p_3)\boldsymbol{\mu}_2-p_3/(p_2+p_3)\boldsymbol{\mu}_3=\left(\boldsymbol{\mu}_1-\sum_1^3 p_i\boldsymbol{\mu}_i\right)/(p_2+p_3) \tag{4. 4}$$

である。事前確率を考慮して時刻 t, t' における分散共分散行列を求めると次の様に示される。

$$t:\ \boldsymbol{\Sigma}_t=\sum_1^3 p_i\boldsymbol{\Sigma}_i+p_1p_2/(p_1+p_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)', \tag{4. 5}$$

$$t':\ \boldsymbol{\Sigma}_{t'}=\sum_1^3 p_i\boldsymbol{\Sigma}_i+p_2p_3/(p_2+p_3)(\boldsymbol{\mu}_2-\boldsymbol{\mu}_3)(\boldsymbol{\mu}_2-\boldsymbol{\mu}_3)'。 \tag{4. 6}$$

母集団 π_2 の平均ベクトル $\boldsymbol{\mu}_2$ が他の母集団の平均ベクトル $\boldsymbol{\mu}_1$ または $\boldsymbol{\mu}_3$ に等しい場合については P. A. Lachenbruch[4,5]，G. J. McLachlan[6] が考察している。

5. 条件を加えたモデルⅠの性質と結果

前章で示した一般的なモデルの全変数について調べることは不可能であるので，ここでは，次の様な条件Ⅰを加えたモデルⅠについて考察する。本モデルは母集団 π_2 の平均ベクトルが母集団 π_1 と母集団 π_3 の平均ベクトルで作られた直線上にあることを示している。

条件Ⅰ） $\sum_1^3 p_i\boldsymbol{\Sigma}_i=\boldsymbol{I}_k, \boldsymbol{\mu}_i=(c_1\mu_i,\ c_2\mu_i,\ \cdots,\ c_k\mu_i)$ $(i=1, 2, 3)$,

ここで，$\boldsymbol{I}_k$ は単位行列である。この条件のもとで時刻 t に関する各値を求めると，結局，

$$\boldsymbol{\delta}_t=(\delta_{t_1},\ \cdots,\ \delta_{t_k}),\ \ \delta_{t_j}=c_j\left(\sum_1^3 p_i\mu_i-\mu_3\right)/(p_1+p_2)\ (j=1, \cdots, k), \tag{5. 1}$$

$$\boldsymbol{\Sigma}_t=\boldsymbol{I}_k+\alpha_t\begin{pmatrix} c_1c_1, & c_1c_2, & \cdots, & c_1c_k \\ \vdots & \vdots & & \vdots \\ c_kc_1, & c_kc_2, & \cdots, & c_kc_k \end{pmatrix},\ \ \alpha_t=p_1p_2(\mu_1-\mu_2)^2/(p_1+p_2), \tag{5. 2}$$

$$\boldsymbol{\Sigma}_t^{-1}=\frac{1}{1+\alpha_t\sum_1^k c_j^2}\begin{pmatrix} 1+\alpha_t\sum_{\hat{1}} c_j^2, & -c_1c_2\alpha_t, & \cdots, & -c_1c_k\alpha_t \\ -c_2c_1\alpha_t, & 1+\alpha_t\sum_{\hat{2}} c_j^2, & \cdots, & -c_2c_k\alpha_t \\ \vdots & & & \vdots \\ -c_kc_1\alpha_t, & -c_kc_2\alpha_t, & \cdots, & 1+\alpha_t\sum_{\hat{k}} c_j^2 \end{pmatrix}, \tag{5. 3}$$

ここで, $\sum_{\hat{j'}} c_j^2=c_1^2+\cdots+c^2{}_{j'-1}+c^2{}_{j'+1}+\cdots+c_k^2\ (j'=1, \cdots, k)$ となる。従って係数ベクトル $\boldsymbol{a}_t$ は次の様になる。

$$\boldsymbol{a}_t=\boldsymbol{\delta}_t\boldsymbol{\Sigma}_t^{-1}=(a_{t_1}, \cdots, a_{t_k}),\ a_{t_j}=\frac{c_j}{1+\alpha_t\sum_1^k c_{j'}^2}\cdot\left(\sum_1^3 p_i\mu_i-\mu_3\right)/(p_1+p_2)。 \tag{5. 4}$$

同様に, 時刻 t' について計算すると

$$\boldsymbol{a}_{t'}=\boldsymbol{\delta}_{t'}\boldsymbol{\Sigma}_{t'}^{-1}=(a_{t'_1}, \cdots, a_{t'_k}),\ a_{t'_j}=\frac{c_j}{1+\alpha_{t'}\sum_1^k c_{j'}^2}\cdot\left(\mu_1-\sum_1^3 p_i\mu_i\right)/(p_2+p_3), \tag{5. 5}$$

ここで, $\alpha_{t'}=p_2p_3(\mu_2-\mu_3)^2/(p_2+p_3)$ となる。両係数ベクトルを比較すると次の性質がわかる。

性質 1) 各件 I を満たす母集団において係数ベクトル $\boldsymbol{a}_{t'}$ は $\boldsymbol{a}_t$の定数倍となる。
つまり, どちらの係数ベクトルを用いてもサンプルの得点間の相対関係は変わらない。しかしながら, 種々の目的に対する判別値は変動する。

次に, モデル I のもとで尤度法による判別値 h_M とベイズ法による判別値 h_B が時刻 t から時刻 t' の間でどの様に変化するかを調べる。ここでは, 定数倍の影響を除くため, 係数ベクトル $\boldsymbol{a}^1=(c_1, c_2, \cdots, c_k)$ を用いる。このとき時刻 t, t' における各判別値は各々,

$$h_{M(t)}=\frac{1}{2}\sum_1^k c_j^2(p_1/(p_1+p_2)\,\mu_1+p_2/(p_1+p_2)\,\mu_2+\mu_3), \tag{5. 6}$$

$$h_{B(t)}=h_{M(t)}+\log p_3/(p_1+p_2)\cdot\frac{\left(1+\alpha_t\sum_1^k c_j^2\right)(p_1+p_2)}{\left(\sum_1^3 p_i\mu_i-\mu_3\right)}, \tag{5. 7}$$

$$h_{M(t')}=\frac{1}{2}\sum_1^k c_j^2(\mu_1+p_2/(p_2+p_3)\mu_2+p_3/(p_2+p_3)\mu_3), \tag{5. 8}$$

$$h_{B(t')}=h_{M(t')}+\log(p_2+p_3)/p_1\cdot\frac{\left(1+\alpha_{t'}\sum_1^k c_j^2\right)(p_2+p_3)}{\left(\mu_1-\sum_1^3 p_i\mu_i\right)} \tag{5. 9}$$

となる。各判別値の変化率に対する指標を次の様に定義する。

$$\Delta h_M=(h_{M(t')}-h_{M(t)})/h_{M(t)}, \tag{5.10}$$

$$\Delta h_B=(h_{B(t')}-h_{B(t)})/h_{B(t)}, \tag{5.11}$$

表 3 は母集団 π_1, π_2 の平均ベクトルの変数 $\mu_1=0, \mu_3=1$, また, $\sum_1^k c_j^2=4$ とし, 他の変数 μ_2, p_1, p_3 を動かした結果を示している。ただし, $0<\mu_2\leqq 1/2$ とした。$1/2<\mu_2<1$ については表 3 より推測できる。その結果, 母集団 π_2 の事前確率 p_2 が小さいほど判別値の変化率が小さいことは当然のこ

表 3 モデル I における判別値の変化率
上段：尤 度 法 Δh_M
下段：ベイズ法 Δh_B

(a) $\mu_2=0$

p_3 \ p_1	1/6	1/3	1/2	2/3
1/6	−0.800 −4.308	−0.600 −3.392	−0.400 −2.814	−0.200 −2.136
1/3	−0.750 −2.014	−0.500 −1.487	−0.250 −0.943	
1/2	−0.667 −0.815	−0.333 −0.505		
2/3	−0.500 −0.211			

(b) $\mu_2=1/6$

p_3 \ p_1	1/6	1/3	1/2	2/3
1/6	−0.706 −2.433	−0.556 −2.312	−0.400 −2.090	−0.231 −1.698
1/3	−0.659 −1.432	−0.462 −1.195	−0.250 −0.809	
1/2	−0.583 −0.772	−0.307 −0.487		
2/3	−0.435 −0.300			

(c) $\mu_2=1/3$

p_3 \ p_1	1/6	1/3	1/2	2/3
1/6	−0.632 −1.686	−0.520 −1.701	−0.400 −1.617	−0.257 −1.383
1/3	−0.583 −1.151	−0.429 −1.004	−0.250 −0.707	
1/2	−0.510 −0.741	−0.282 −0.475		
2/3	−0.375 −0.362			

(d) $\mu_2=1/2$

p_3 \ p_1	1/6	1/3	1/2	2/3
1/6	−0.571 −1.315	−0.491 −1.340	−0.400 −1.305	−0.280 −1.151
1/3	−0.519 −1.000	−0.400 −0.882	−0.250 −0.632	
1/2	−0.444 −0.723	−0.259 −0.469		
2/3	−0.318 −0.408			

とと思われる。μ_2 の値が 1/2 に近いほど事前確率 p_1, p_3 に対する変化率の動きが小さい。また，ベイズ法における変化の割合の方が尤度法におけるそれよりも大きい。$\sum_1^k c_j^2$ に関しては，式より明らかな様に尤度法においては無関係である。ベイズ法の変化率 Δh_B は，一般に，$\sum_1^k c_j^2$ の値が大きいほど小さいことが他の結果から知ることができた。

6. 条件を加えたモデルIIの結果

本章では，4 章の一般的なモデルに次の様な条件IIを加えたモデルIIについて考察する。

条件II）$\sum_1^3 p_i \boldsymbol{\Sigma}_i=\boldsymbol{I}_k$，$\boldsymbol{\mu}_2=(c_1\mu_2, c_2\mu_2, \cdots, c_k\mu_2)$，ただし，$c_1=1$，

$\boldsymbol{\mu}_i=(\mu_i, \cdots, \mu_i)\ (i=1,3)$。

この条件のもとで時刻 t に関する各値を求めると次の様になる。

$$\boldsymbol{\delta}_t=(\delta_{t_1}, \cdots, \delta_{t_k}),\quad \delta_{t_j}=(p_1\mu_1+c_jp_2\mu_2+p_3\mu_3-\mu_3)/(p_1+p_2), \tag{6. 1}$$

$$\boldsymbol{\Sigma}_t=\boldsymbol{I}+\alpha_t\begin{pmatrix}\sigma_1\sigma_1, & \sigma_1\sigma_2, & \cdots, & \sigma_1\sigma_k\\ \vdots & \vdots & & \vdots\\ \sigma_k\sigma_1, & \sigma_k\sigma_2, & \cdots, & \sigma_k\sigma_k\end{pmatrix},\ \alpha_t=p_1p_2/(p_1+p_2),\ \sigma_j=(\mu_1-c_j\mu_2)。 \tag{6. 2}$$

分散共分散行列の逆行列 Σ_t^{-1} も式（5.3）と同様に計算され，係数ベクトル $\boldsymbol{a}_t$ は次の様になる。

$$\boldsymbol{a}_t=\boldsymbol{\delta}_t\boldsymbol{\Sigma}_t^{-1}=(a_{t_1}, \cdots, a_{t_k}),$$

$$a_{t_j}=\{-(p_1\mu_1+p_2\mu_2+p_3\mu_3-\mu_3)\sigma_j\sigma_1\alpha_t$$

$$-(p_1\mu_1+c_2p_2\mu_2+p_3\mu_3-\mu_3)\sigma_j\sigma_2\alpha_t$$

$$\cdots \tag{6. 3}$$
$$+(p_1\mu_1+c_jp_2\mu_2+p_3\mu_3-\mu_3)(1+\alpha_t\sum_{\hat{j}}\sigma_{j'}{}^2)$$
$$\cdots$$
$$-(p_1\mu_1+c_kp_2\mu_2+p_3\mu_3-\mu_3)\sigma_j\sigma_k\alpha_t\}/(p_1+p_2)。$$

同様に時刻 t' について計算すると

$$\boldsymbol{a}_{t'}=\boldsymbol{\delta}_{t'}\boldsymbol{\Sigma}_{t'}^{-1}=(a_{t'_1},\cdots,a_{t'_k}),$$
$$a_{t'_j}=\{-(\mu_1-p_1\mu_1-p_2\mu_2-p_3\mu_3)\sigma'_j\sigma'_1\alpha_{t'}$$
$$-(\mu_1-p_1\mu_1-c_2p_2\mu_2-p_3\mu_3)\sigma'_j\sigma'_2\alpha_{t'}$$
$$\cdots \tag{6. 4}$$
$$+(\mu_1-p_1\mu_1-c_jp_2\mu_2-p_3\mu_3)(1+\alpha_{t'}\sum_{\hat{j}}\sigma'^2_{j'})$$
$$\cdots$$
$$-(\mu_1-p_1\mu_1-c_kp_2\mu_2-p_3\mu_3)\sigma'_j\sigma'_k\alpha_{t'}\}/(p_2+p_3),$$

ここで，$\alpha_{t'}=p_2p_3/(p_2+p_3)$，$\sigma'_j=(c_j\mu_2-\mu_3)$ となる。３つの母集団の平均ベクトルの変数 $\mu_1=0$，$\mu_2=1/k$，$\mu_3=1$ と仮定し，$k=4$，$c_j=j$ $(j=1,\cdots,k)$ としたときについて時刻 t と t' に対する判別係数を求めると表４(a) に示した様になる。これは判別係数の大きさの順が時刻 t と t' とでは全く逆になった例である。また，$\mu_1=0$，$\mu_2=1$，$\mu_3=1$，$k=4$，$\boldsymbol{c}=(1,1,1,0)$ のときの結果が表４(b) である。この例において，第１変量，第２変量，第３変量に対する係数はほぼ同じ値である。しかしながら，第１，２，３変量と第４変量を比較するとその大きさの逆転が見られる。このように，

表４　モデルIIにおける判別係数の変化

上段：時刻 t，下段：時刻 t'

(a) $\mu_1=0$，$\mu_2=1/4$，$\mu_3=1$，$k=4$，$c_j=j$ $(j=1,\cdots,4)$

p_1	p_3	1	2	3	4
1/6	1/6	−0.967 −0.387	−0.615 −0.580	−0.366 −0.800	−0.097 −1.024
	1/3	−0.969 −0.523	−0.650 −0.643	−0.407 −0.829	−0.152 −1.027
	1/2	−0.972 −0.665	−0.697 −0.713	−0.475 −0.866	−0.245 −1.021
	2/3	−0.979 −0.822	−0.780 −0.822	−0.607 −0.918	−0.433 −1.010
1/3	1/6	−1.100 −0.424	−0.684 −0.609	−0.461 −0.818	−0.201 −1.039
	1/3	−1.083 −0.599	−0.755 −0.706	−0.553 −0.870	−0.332 −1.044
	1/2	−1.056 −0.787	−0.849 −0.826	−0.704 −0.925	−0.554 −1.029
1/2	1/6	−1.150 −0.486	−0.754 −0.655	−0.574 −0.846	−0.356 −1.049
	1/3	−1.094 −0.729	−0.866 −0.806	−0.737 −0.923	−0.599 −1.044
2/3	1/6	−1.117 −0.612	−0.861 −0.743	−0.746 −0.892	−0.615 −1.048

(b) $\mu_1=0$，$\mu_2=1$，$\mu_3=1$，$k=4$，$\boldsymbol{c}=(1,1,1,0)$

p_1	p_3	1	2	3	4
1/6	1/6	−0.200 −1.133	−0.187 −1.007	−0.186 −1.021	−1.373 −0.177
	1/3	−0.250 −1.200	−0.236 −1.000	−0.235 −1.033	−1.350 −0.331
	1/2	−0.333 −1.200	−0.318 −1.000	−0.317 −1.033	−1.311 −0.497
	2/3	−0.500 −1.133	−0.485 −1.007	−0.485 −1.021	−1.234 −0.709
1/3	1/6	−0.400 −1.125	−0.331 −1.031	−0.327 −1.039	−1.513 −0.228
	1/3	−0.500 −1.167	−0.435 −1.037	−0.433 −1.051	−1.430 −0.442
	1/2	−0.667 −1.125	−0.620 −1.031	−0.619 −1.039	−1.290 −0.685
1/2	1/6	−0.600 −1.111	−0.468 −1.049	−0.467 −1.053	−1.473 −0.314
	1/3	−0.750 −1.111	−0.668 −1.049	−0.669 −1.053	−1.302 −0.628
2/3	1/6	−0.800 −1.083	−0.682 −1.053	−0.686 −1.054	−1.297 −0.485

非発症群から発症群に移行するサンプルの母集団によっては係数ベクトルが大きく変わる様な場合が生じるので注意が必要である。

7. おわりに

本論文では，ある年の非発症群から発症群に移行するサンプルが判別係数，判別得点，判別値におよぼす影響を多変量正規分布を用いたモデルをたてて考察した。その結果，判別係数・判別得点には何の影響も与えない条件が判明した。しかしながら，判別値についてはこの条件のもとでもその値を大きく変化させる。このことは，新たに検診を受けた人が後に発症するか否かを決定する判別値が不安定であることを示している。他の条件のもとでは，時間によって判別群が変化する場合，明らかに困難な問題が生じることが示された。つまり，判別係数の値の大小が逆転する様な現象である。判別係数の各値は時としてその項目（変量）の重み（重要性）を示す尺度として用いられることがあるためこの様な現象は何らかの基準や手法を用いて解決する必要がある。

参考文献

1) Fisher, R. A.: Ann. Eugen., 8 (1936), p. 376
2) 河口至商：多変量解析入門 I （昭 48），p. 79，森北出版
3) ラッヘンブルック：判別分析（昭 54），現代数学社
4) Lachenbruch, P. A.: Technometrics, 8 (1966), 4, p. 657
5) Lachenbruch, P. A.: Technometrics, **16** (1974), 3, p. 419
6) McLachlan, G. J.: Technometrics, **14** (1972), 2, p. 415