システム制御情報学会論文誌, Vol. 2, No. 8, pp. 274～282, 1989

論　文

区間データに対する判別問題の二つの定式化*

石渕久生**・田中英夫**

Two Formulations of Discriminant Problem for Interval Data*

Hisao ISHIBUCHI** and Hideo TANAKA**

In the conventional discriminant analysis, the attribute values of each sample are represented by a point in the multi-dimensional feature space. But they should be represented by an interval vector in the feature space when the observed values are not constant but fluctuating with time. Thus we intend to propose a discriminant method for multi-dimensional interval data. The discriminant method for the interval data is based on the interval linear discriminant function which maps an interval vector in the multi-dimensional feature space to an interval in the one-dimensional decision space. Therefore our discriminant problem is to determine the interval linear discriminant function and group intervals which represent the given groups in the one-dimensional decision space. In this paper, two mathematical programming problems are formulated for such a discriminant problem and computational algorithms using the linear programming are shown.

1. 緒　言

通常の判別分析の問題では，各々のサンプルの属性値に関するデータは，多次元属性空間上の点として表わされる点データである．しかし，実際には，われわれの脈拍や血圧，血糖値などを考えるとわかるように，属性値は一定ではなく，変動している場合も多い．このような場合では，属性値を単一の値で表わすのではなく，その属性値の変動する範囲を示す区間で表示すべきである．したがって，各々のサンプルの属性値に関するデータは，多次元属性空間上の区間ベクトルとして表わされる区間データとなる．このような区間データ表示は，属性値が変動している場合だけではなく，あいまいさを含むデータを表現するのにも適している．例えば，特定乳歯の萌出日時や眼底出血の時点などについては，事象生起の正確な時点を求めることは困難であるので，事象の生起する直前と直後の検診日時により算出される区間内で事象が生起したとするのが最も信頼できるデータである[1]．

そこで本研究では，各々のサンプルの属性値が区間として表わされているような区間データに対する判別分析の問題について考察する．そして，区間データに対する判別分析の問題を，数理計画問題に帰着させ，線形計画法を用いた解法を示す．線形計画法を用いて判別関数を求めるという手法は，通常の点データに対して，N. Freed ら[2~3]により提案されたものである．本研究で定式化する判別問題は，区間データを取り扱うことを目標に，N. Freed らの方法を区間概念に基づき拡張したものである．

区間データに対する判別分析の問題は，基本的な2群判別の場合では，次のような問題となる．すなわち，サンプルの属性値を表わす区間ベクトルを1次元上の区間に写像する区間線形判別関数と，1次元上で2群を表わす二つの区間を求める問題である．本研究では，このような問題を，1次元上に写像されたサンプル区間と2群を表わす区間との関係により，Intersection 問題と Inclusion 問題という二つの数理計画問題に帰着させる．簡単に言えば，Intersection 問題は，サンプル区間と交わりのあるような2群を表わす区間を求める問題であり，Inclusion 問題は，サンプル区間を包含するような2群を表わす区間を求める問題である．この二つの問題は，田中ら[4]によりファジィデータに対する回帰分析手法として提案された Conjunction 問題と Min 問題に対応する問題と考えることもできる．

本研究では，まず **2.** において，区間演算および区間

* 原稿受付　1988. 12. 15
** 大阪府立大学　工学部　University of Osaka Prefecture ; Sakai, Osaka 591, JAPAN
Key Words : discriminant analysis, interval data, interval model, mathematical programming, linear programming.

線形判別関数を定義する．つぎに，多次元区間データに対する判別分析の問題に対し，3. で Intersection 問題を，4. で Inclusion 問題を定式化し，線形計画法を用いた解法を示す．さらに 5. において，いくつかの定理により二つの判別問題の性質を示す．最後に 6. において，数値計算例により二つの判別問題の説明を行なう．

2. 区間モデル

2.1 区間演算

本研究では，区間を英字大文字で表わし，実数を英字小文字で表わすことにする．一般に閉区間は，その左端と右端により，次のように表現される．

$$A=[a_L, a_R]=\{t|a_L\leqq t\leqq a_R\} \quad (1)$$

ただし，t は実数であり，添字の L と R は，区間の左端 (Left limit) と右端 (Right limit) を表わしている．

また，区間の中心と幅を用いれば，次のように表わすこともできる．

$$A=\langle a_C, a_W\rangle=\{t|a_C-a_W\leqq t\leqq a_C+a_W\} \quad (2)$$

ただし，添字の C と W は，区間の中心 (Center) と幅 (Width) を表わしている．このとき，(1), (2) 式より，次の関係が得られる (Fig. 1 参照)．

$$a_C=(a_R+a_L)/2 \quad (3)$$

$$a_W=(a_R-a_L)/2 \quad (4)$$

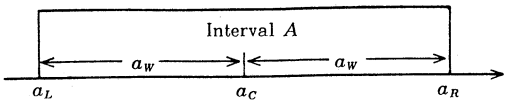


Fig. 1 Graphical explanation of interval

つぎに，本研究で用いる区間演算を次のように定義する．

[定義 1]

区間の和と区間の実数倍を次のように定義する．

$$A+B=[a_L, a_R]+[b_L, b_R]=[a_L+b_L, a_R+b_R] \quad (5)$$

$$A+B=\langle a_C, a_W\rangle+\langle b_C, b_W\rangle=\langle a_C+b_C, a_W+b_W\rangle \quad (6)$$

$$k\cdot A=k[a_L, a_R]=\begin{cases}[ka_L, ka_R] \text{ for } k\geqq 0\\ [ka_R, ka_L] \text{ for } k<0\end{cases} \quad (7)$$

$$k\cdot A=k\langle a_C, a_W\rangle=\langle ka_C, |k|a_W\rangle \quad (8)$$

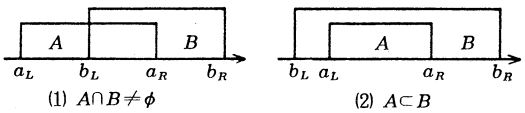
本研究では，区間演算の表記が簡単になるように，区間を中心と幅で表わし，(6) 式および (8) 式の演算を用いることにする．

つぎに，区間の交わりと区間の包含を次のように定義する．

$$A\cap B\neq\phi\Leftrightarrow a_L\leqq b_R,\ b_L\leqq a_R$$
$$\Leftrightarrow a_C-a_W\leqq b_C+b_W,\ b_C-b_W\leqq a_C+a_W \quad (9)$$

$$A\subset B\Leftrightarrow b_L\leqq a_L,\ a_R\leqq b_R$$
$$\Leftrightarrow b_C-b_W\leqq a_C-a_W,\ a_C+a_W\leqq b_C+b_W \quad (10)$$

この定義は，Fig. 2 からもわかるように，通常の集合論における交わりと包含の定義と同じものである．



(1) $A\cap B\neq\phi$　(2) $A\subset B$

Fig. 2 Intersection and inclusion between intervals

2.2 区間線形判別関数

いま，m 個の観測値が，n 次元属性空間上の区間ベクトル $\boldsymbol{X}_j=(X_{1j}, X_{2j}, \cdots, X_{nj})^T$, $j=1, 2, \cdots, m$ として与えられているとする．このとき，区間線形判別関数は，n 次元区間ベクトル $\boldsymbol{X}_j$ を 1 次元上の区間 Y_j に写像する次のような関数である．

$$Y_j=k_1X_{1j}+k_2X_{2j}+\cdots+k_nX_{nj} \quad (11)$$

ただし，$k_i (i=1, 2, \cdots, n)$ は実数係数であり，判別問題の決定変数となる．

区間ベクトル $\boldsymbol{X}_j$ の要素 X_{ij} を，中心 x_{ij} と幅 d_{ij} を用いて $X_{ij}=\langle x_{ij}, d_{ij}\rangle$ と表現すると，(11) 式は (6) 式と (8) 式を用いて次のように変形される．

$$Y_j=k_1\langle x_{1j}, d_{1j}\rangle+\cdots+k_n\langle x_{nj}, d_{nj}\rangle$$
$$=\langle\sum_{i=1}^{n}k_ix_{ij},\ \sum_{i=1}^{n}|k_i|d_{ij}\rangle \quad (12)$$

3. Intersection 問題

3.1 問題の定式化

まず，2 群判別の問題を考えよう．区間データに対する 2 群判別の問題は，(11) 式の区間線形判別関数と，1 次元上で 2 群 ($G1$ と $G2$) を表わす二つの区間を求める問題である．

2 群を表わす区間 Y_{G1} と Y_{G2} を，その中心 a_1, a_2 と幅 c_1, c_2 により，次のように表現する．

$$Y_{G1}=\langle a_1, c_1\rangle,\ Y_{G2}=\langle a_2, c_2\rangle \quad (13)$$

Intersection 問題では，2 群を表わす区間 Y_{G1}, Y_{G2} が，次の制約条件を満たすように求められる．

$$Y_j\cap Y_{Gh}\neq\phi \text{ for all } j \text{ in } Gh\ (h=1, 2) \quad (14)$$

ただし，Y_j は，(11) 式により写像された 1 次元上の区間

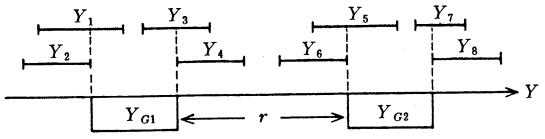


Fig. 3　Constraint of Intersection Problem

である．この条件を，$G1=\{1, 2, 3, 4\}$, $G2=\{5, 6, 7, 8\}$ の場合で図示すると，Fig. 3 のようになる．

区間 Y_{G1}, Y_{G2} に対し，区間 Y_{G2} は区間 Y_{G1} より右にあると仮定し，二つの区間の間の距離を表わす指標 r を次式で定義する．

$$r=(a_2-c_2)-(a_1+c_1) \tag{15}$$

Fig. 3 に示すように，上式における r の値が大きいほど，1次元上で2群を表わす区間 Y_{G1}, Y_{G2} が離れていることになる．また，区間の幅は，区間の位置のあいまいさを表わすと考えることができるので，Y_{G1} と Y_{G2} は，その幅が小さいほど明確に1次元上で2群の位置を表わしていると考えることができる．

以上のことから，Intersection 問題は，(14), (15) 式の制約のもとで，r を最大にし，c_1, c_2 の和を最小にする区間線形判別関数と二つの区間 Y_{G1}, Y_{G2} を求める問題となり，次のように定式化される．

[Intersection 問題]

目的関数：$z=w_1r-w_2(c_1+c_2)$　(最大化)　(16)

制約条件

$$\sum_{i=1}^{n} k_i x_{ij}-\sum_{i=1}^{n}|k_i|d_{ij}\leqq a_h+c_h, \quad \text{for all } j \text{ in } Gh\ (h=1, 2) \tag{17}$$

$$\sum_{i=1}^{n} k_i x_{ij}+\sum_{i=1}^{n}|k_i|d_{ij}\geqq a_h-c_h, \quad \text{for all } j \text{ in } Gh\ (h=1, 2) \tag{18}$$

$$r=(a_2-c_2)-(a_1+c_1) \tag{15}$$

$$\sum_{i=1}^{n}|k_i|=1 \tag{19}$$

$$c_1, c_2\geqq 0 \tag{20}$$

r, k_i, a_1, a_2 の符号は自由 $(i=1, 2, \cdots, n)$

決定変数：r, a_1, a_2, c_1, c_2, k_i $(i=1, 2, \cdots, n)$

判別分析の目的は与えられた2群を分離することであるので，目的関数のなかでは特に第1項が重要であり，重み w_1, w_2 は，$0<w_2\ll w_1$ を満たすように与えられる．したがって，この問題は，優先順位のある多目的計画問題と考えることもできる．また，第2項は，a_1, a_2, c_1, c_2 を一意的に決めるために必要なものである．一方，制約式 (17), (18) は，(14) 式を，(9) 式と (12) 式を用いて書き換えたものである．また，制約式 (19) は，係数 k_i の値がすべて0になったり，逆に無限大に発散したりするのを防ぐためのものである．さらに，非負条件 (20) は，区間の幅は負にはならないことを表わしている．

3.2　問題の解法

Intersection 問題は，変数 k_i $(i=1, 2, \cdots, n)$ に絶対値がついているため，このままでは簡単に解くことはできない．しかし，区間データの次元の数 n が小さい場合では，変数 k_i の符号のすべての組み合せを考慮することにより最適解を求めることができる．すなわち，変数 k_i の符号の組み合せに応じて Intersection 問題の絶対値をはずすことにより 2^n 個の線形計画問題を定式化し，そのなかで最も目的関数の値が大きくなるものを求めることにより，Intersection 問題の最適解を得ることができる．たとえば，5次元区間データに対する判別分析の問題では，$2^5=32$ 通りの線形計画問題を解くことにより，Intersection 問題の解が得られる．したがって，データの次元の数が2〜5程度であれば，このような総当たり的な方法で Intersection 問題の解を求めることは，十分に可能である．

一方，区間データの次元の数 n が大きく，上述の総当たり的な方法が現実的でない場合には，次のような方法により Intersection 問題の解を求めることができる．まず，変数 k_i を正の部分 k_i^+ と負の部分 k_i^- とに分割する．つまり，次のような変数の変換を行なう．

$$k_i=k_i^+-k_i^-, \quad i=1, 2, \cdots, n \tag{21}$$

ただし

$$k_i^+k_i^-=0, \quad i=1, 2, \cdots, n \tag{22}$$

$$k_i^+, k_i^-\geqq 0, \quad i=1, 2, \cdots, n \tag{23}$$

である．

上式を用いて Intersection 問題を書き換えると，次のようになる．

[問題 P 1]

目的関数：$z=w_1r-w_2(c_1+c_2)$　(最大化)　(16)

制約条件

$$\sum_{i=1}^{n}(k_i^+-k_i^-)x_{ij}-\sum_{i=1}^{n}(k_i^++k_i^-)d_{ij}\leqq a_h+c_h, \quad \text{for all } j \text{ in } Gh\ (h=1, 2) \tag{24}$$

$$\sum_{i=1}^{n}(k_i^+-k_i^-)x_{ij}+\sum_{i=1}^{n}(k_i^++k_i^-)d_{ij}\geqq a_h-c_h, \quad \text{for all } j \text{ in } Gh\ (h=1, 2) \tag{25}$$

$$r=(a_2-c_2)-(a_1+c_1) \tag{15}$$

$$\sum_{i=1}^{n}(k_i^++k_i^-)=1 \tag{26}$$

$$k_i^+k_i^-=0, \quad i=1, 2, \cdots, n \tag{22}$$

$$c_1, c_2, k_i^+, k_i^-\geqq 0, \quad i=1, 2, \cdots, n \tag{27}$$

r, a_1, a_2 の符号は自由

決定変数：r, a_1, a_2, c_1, c_2, k_i^+, k_i^- $(i=1, 2, \cdots, n)$

この問題の最適解は，Intersection 問題の最適解と一致する．しかし，この問題には相補条件 (22) があるので，線形計画問題として解くことはできない．そこで，**問題 P 1** から相補条件 (22) を取り除いた問題を**問題 P 2** とし，次のように定式化する．

[問題 P 2]

目的関数：(16) 式の最大化

制約条件：(15), (24), (25), (26), (27) 式

決定変数：$r, a_1, a_2, c_1, c_2, k_i^+, k_i^- (i=1, 2, \cdots, n)$

この問題は，線形計画問題であるので，簡単に解くことができる．このとき，**問題 P 2** の最適解が相補条件 (22) を満たせば，その解は，**問題 P 1** の最適解と一致し，Intersection 問題の最適解となる．したがって，このような場合には，**問題 P 2** を解くことより，Intersection 問題の最適解が得られる．一方，**問題 P 2** の最適解が相補条件 (22) を満たさないときには，**問題 P 2** の最適解と Intersection 問題の最適解は一致しないが，**問題 P 2** の最適解は次のような意味を持つ．すなわち，制約条件式 (24), (25) からわかるように，**問題 P 2** の最適解は，次のような区間線形判別関数と二つの区間 Y_{G1}, Y_{G2} を表わしている．

$$Y_j = \sum_{i=1}^{n} k_i^+ X_{ij} - \sum_{i=1}^{n} k_i^- X_{ij} = \langle \sum_{i=1}^{n} (k_i^+ - k_i^-) x_{ij}, \sum_{i=1}^{n} (k_i^+ + k_i^-) d_{ij} \rangle \quad (28)$$

$$Y_j \cap Y_{Gh} \neq \phi \text{ for all } j \text{ in } Gh \ (h=1, 2) \quad (14)$$

(29) 式の区間線形判別関数は，相補条件 (22) が成立しないときには，(11), (12) 式の区間線形判別関数とは一致しない．しかし，(28) 式が多次元区間データを 1 次元上の区間に写像する関数であることと，二つの区間 Y_{G1}, Y_{G2} が Intersection 問題の制約条件である (14) 式を満たしていることを考えれば，**問題 P 2** の最適解は，近似的には Intersection 問題の解になっているということがいえる．

3.3 多群判別問題への拡張

ここでは，**3.1** で定式化した Intersection 問題を，多群判別の問題に拡張する．

判別の対象となる群の数を p 個とし，各々の群を Gh $(h=1, 2, \cdots, p)$ で表わすことにする．また，各々の群を表わす 1 次元上の区間を Y_{Gh} とし，中心と幅により次のように表わすことにする．

$$Y_{Gh} = \langle a_h, c_h \rangle, \quad h=1, 2, \cdots, p \quad (29)$$

この p 個の区間が，Fig. 4 に示すように，1 次元上に $Y_{G1}, Y_{G2}, \cdots, Y_{Gp}$ の順に並ぶものとする．ただし，この順番が明らかでない場合には，すべての組み合せを考える必要がある．

1 次元上で各群を表わす p 個の区間は，**3.1** と同様に，次の制約条件を満たすように求められる．

$$Y_j \cap Y_{Gh} \neq \phi \text{ for all } j \text{ in } Gh \ (h=1, 2, \cdots, p) \quad (30)$$

また，p 個の区間 $Y_{G1}, Y_{G2}, \cdots, Y_{Gp}$ のなかで，隣接する区間 Y_{Gh} と $Y_{G,h+1}$ との間の距離を表わす指標 r_h を次のように定義する (Fig. 4 参照)．

$$r_h = (a_{h+1} - c_{h+1}) - (a_h + c_h), \quad h=1, 2, \cdots, p-1 \quad (31)$$

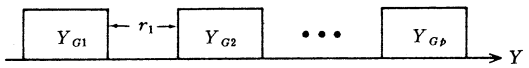


Fig. 4　Group intervals $Y_{G1}, Y_{G2}, \cdots, Y_{Gp}$

多群判別の目的は，p 個の群をできるだけ分離することにあるので，次式で定義される値 r の最大化を判別問題の第一目的とする．

$$r = \min\{r_1, r_2, \cdots, r_{p-1}\} \quad (32)$$

さらに，a_h, c_h を一意的に決定するため，c_h の和の最小化を判別問題の第二目的とする．したがって，**3.1** と同様に，重み w_1, w_2 を $0 < w_2 \ll w_1$ を満たすように与えることにより，多群に対する判別問題が次のように定式化される．

[多群 Intersection 問題]

目的関数：$z = w_1 r - w_2 \sum_{h=1}^{p} c_h$ （最大化）　(33)

制約条件

$$\sum_{i=1}^{n} k_i x_{ij} - \sum_{i=1}^{n} |k_i| d_{ij} \leqq a_h + c_h, \quad \text{for all } j \text{ in } Gh \ (h=1, 2, \cdots, p) \quad (34)$$

$$\sum_{i=1}^{n} k_i x_{ij} + \sum_{i=1}^{n} |k_i| d_{ij} \geqq a_h - c_h, \quad \text{for all } j \text{ in } Gh \ (h=1, 2, \cdots, p) \quad (35)$$

$$r_h = (a_{h+1} - c_{h+1}) - (a_h + c_h), \quad h=1, 2, \cdots, p-1 \quad (31)$$

$$r \leqq r_h, \quad h=1, 2, \cdots, p-1 \quad (36)$$

$$\sum_{i=1}^{n} |k_i| = 1 \quad (19)$$

$$c_h \geqq 0, \quad h=1, 2, \cdots, p \quad (37)$$

r, r_h の符号は自由 $(h=1, 2, \cdots, p-1)$

k_i, a_h の符号は自由 $(i=1, 2, \cdots, n; h=1, 2, \cdots, p)$

決定変数：$r, r_h (h=1, 2, \cdots, p-1)$

$k_i, a_h, c_h (i=1, 2, \cdots, n; h=1, 2, \cdots, p)$

制約式 (34), (35) は，(30) 式を，(9) 式と (12) 式を用いて書き換えたものである．また，目的関数が r の最大化であるため，(32) 式は制約式 (36) により表され

ている．

3.4　判別規則

ここでは，Intersection 問題を解くことにより得られた区間線形判別関数と各群を表わす区間 $Y_{G1}, Y_{G2}, \cdots, Y_{Gp}$ を用いた判別規則を示す．

いま，新たな観測値 $\boldsymbol{X}_j$ (n 次元区間ベクトル) が得られたとする．この区間ベクトルは，区間線形判別関数により，1次元上の区間 Y_j に写像される．この区間 Y_j を判別するため，Intersection 問題の制約条件を考慮し，次のような判別規則を提案する．

[Intersection 問題における判別規則]

1次元上に写像された区間 Y_j が次式を満たせば，観測値 $\boldsymbol{X}_j$ を Gh ($h=1, 2, \cdots, p$) と判別する．

$$Y_j \cap Y_{Gh} \neq \phi \tag{38}$$

4. Inclusion 問題

4.1　問題の定式化

前章と同様に，まず最も基本的な2群判別の問題を定式化する．Inclusion 問題では，2群を表わす区間 $Y_{G1}=\langle a_1, c_1\rangle$, $Y_{G2}=\langle a_2, c_2\rangle$ が，次の制約条件を満たすように求められる．

$$Y_j \subset Y_{Gh} \text{ for all } j \text{ in } Gh \ (h=1, 2) \tag{39}$$

ただし，Y_j は，Intersection 問題と同様に，区間線形判別関数 (11) により写像された1次元上の区間である．この条件を，$G1=\{1, 2, 3, 4\}$, $G2=\{5, 6, 7, 8\}$ の場合で図示すると，Fig. 5 のようになる．

また，二つの区間の間の距離を表わす指標 r を，3.1 と同様に次式で定義する (Fig. 5 参照)．

$$r=(a_2-c_2)-(a_1+c_1) \tag{15}$$

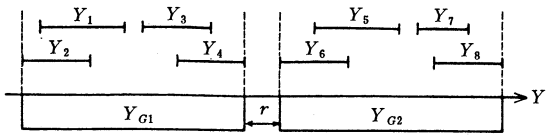


Fig. 5　Constraint of Inclusion Problem

Inclusion 問題は，(39) 式および (15) 式の制約のもとで，r を最大にし，c_1, c_2 の和を最小にする区間線形判別関数と二つの区間 Y_{G1}, Y_{G2} を求める問題であり，次のように定式化される．

[Inclusion 問題]

目的関数：$z=w_1 r-w_2(c_1+c_2)$　(最大化)　(16)

制約条件

$$\sum_{i=1}^{n} k_i x_{ij}-\sum_{i=1}^{n}|k_i| d_{ij} \geqq a_h-c_h, \quad \text{for all } j \text{ in } Gh \ (h=1, 2) \tag{40}$$

$$\sum_{i=1}^{n} k_i x_{ij}+\sum_{i=1}^{n}|k_i| d_{ij} \leqq a_h+c_h, \quad \text{for all } j \text{ in } Gh \ (h=1, 2) \tag{41}$$

$$r=(a_2-c_2)-(a_1+c_1) \tag{15}$$

$$\sum_{i=1}^{n}|k_i|=1 \tag{19}$$

$$c_1, c_2 \geqq 0 \tag{20}$$

r, k_i, a_1, a_2 の符号は自由 ($i=1, 2, \cdots, n$)

決定変数：$r, a_1, a_2, c_1, c_2, k_i$ ($i=1, 2, \cdots, n$)

Intersection 問題と同様に，重み w_1, w_2 は，$0<w_2 \leqslant w_1$ を満たすように与えられる．また，制約式 (40), (41) 式は，(39) 式を，(10) 式と (12) 式を用いて書き換えたものである．

4.2　問題の解法

Inclusion 問題も，基本的には，変数 k_i ($i=1, 2, \cdots, n$) の符号のすべての組み合せを考え，2^n 個の線形計画問題を定式化することにより，最適解を求めることができる．また，3.2 と同様に (21)～(23) 式の変数変換を用いて解くこともできる．まず，Inclusion 問題を次のように書き換える．

[問題 P 3]

目的関数：$z=w_1 r-w_2(c_1+c_2)$　(最大化)　(16)

制約条件

$$\sum_{i=1}^{n}(k_i^+-k_i^-) x_{ij}-\sum_{i=1}^{n}(k_i^++k_i^-) d_{ij} \geqq a_h-c_h, \quad \text{for all } j \text{ in } Gh \ (h=1, 2) \tag{42}$$

$$\sum_{i=1}^{n}(k_i^+-k_i^-) x_{ij}+\sum_{i=1}^{n}(k_i^++k_i^-) d_{ij} \leqq a_h+c_h, \quad \text{for all } j \text{ in } Gh \ (h=1, 2) \tag{43}$$

$$r=(a_2-c_2)-(a_1+c_1) \tag{15}$$

$$\sum_{i=1}^{n}(k_i^++k_i^-)=1 \tag{26}$$

$$k_i^+ k_i^-=0, \quad i=1, 2, \cdots, n \tag{22}$$

$$c_1, c_2, k_i^+, k_i^- \geqq 0, \quad i=1, 2, \cdots, n \tag{27}$$

r, a_1, a_2 の符号は自由

決定変数：$r, a_1, a_2, c_1, c_2, k_i^+, k_i^-$ ($i=1, 2, \cdots, n$)

つぎに，**問題 P 3** から相補条件 (22) を取り除いた問題を**問題 P 4** とし，次のように定式化する．

[問題 P 4]

目的関数：(16) 式の最大化

制約条件：(15), (26), (27), (42), (43) 式

決定変数：$r, a_1, a_2, c_1, c_2, k_i^+, k_i^-$ ($i=1, 2, \cdots, n$)

この問題は，線形計画問題であるので，簡単に解くことができる．このとき，**問題 P 4** の最適解が相補条件 (22) を満たせば，その解は，**問題 P 3** の最適解と一致し，Inclusion 問題の最適解となる．したがって，このような場合には，**問題 P 4** を解くことにより，Inclusion 問題の最適解が得られる．また，3.2 と同様に，**問題 P 4**

の最適解が相補条件(22)を満たさない場合では，**問題P4**の最適解はInclusion問題の近似解とみなすことができる．

4.3 多群判別問題への拡張

ここでは，**4.1**で定式化したInclusion問題を，判別の対象となる群の数がp個の場合に拡張する．**3.3**と同様に，各々の群をGh $(h=1,2,\cdots,p)$とし，群Ghを表わす1次元上の区間を$Y_{Gh}=\langle a_h, c_h\rangle$とする．

1次元上で各群を表わすp個の区間は，**4.1**と同様に，次の制約条件を満たすように求められる．

$$Y_j \subset Y_{Gh} \text{ for all } j \text{ in } Gh \ (h=1,2,\cdots,p) \tag{44}$$

さらに，**3.3**と同様に，このp個の区間が1次元上に，$Y_{G1}, Y_{G2}, \cdots, Y_{Gp}$の順に並ぶものとする．このとき，隣接する二つの区間Y_{Gh}, $Y_{G,h+1}$の間の距離を表わす指標r_hおよびそれらの最小値rを，**3.3**と同様に次のように定義する．

$$r_h=(a_{h+1}-c_{h+1})-(a_h+c_h),\quad h=1,2,\cdots,p-1 \tag{31}$$

$$r=\min\{r_1, r_2, \cdots, r_{p-1}\} \tag{32}$$

このとき，多群に対するInclusion問題は，rを最大にし，$c_h(h=1,2,\cdots,p)$を最小にする区間線形判別関数とp個の区間$Y_{G1}, Y_{G2}, \cdots, Y_{Gp}$を求める問題となり，次のように定式化される．

[多群Inclusion問題]

目的関数：$z=w_1 r-w_2\sum_{h=1}^{p} c_h$　(最大化)　(33)

制約条件

$$\sum_{i=1}^{n} k_i x_{ij}-\sum_{i=1}^{n}|k_i|d_{ij}\geqq a_h-c_h, \quad \text{for all } j \text{ jn } Gh \ (h=1,2,\cdots,p) \tag{45}$$

$$\sum_{i=1}^{n} k_i x_{ij}+\sum_{i=1}^{n}|k_i|d_{ij}\leqq a_h+c_h, \quad \text{for all } j \text{ in } Gh \ (h=1,2,\cdots,p) \tag{46}$$

$$r_h=(a_{h+1}-c_{h+1})-(a_h+c_h),\quad h=1,2,\cdots,p-1 \tag{31}$$

$$r\leqq r_h,\quad h=1,2,\cdots,p-1 \tag{36}$$

$$\sum_{i=1}^{n}|k_i|=1 \tag{19}$$

$$c_h\geqq 0,\quad h=1,2,\cdots,p \tag{37}$$

r, r_hの符号は自由 $(h=1,2,\cdots,p-1)$

k_i, a_hの符号は自由 $(i=1,2,\cdots,n;\ h=1,2,\cdots,p)$

決定変数：$r, r_h\ (h=1,2,\cdots,p-1)$

$k_i, a_h, c_h\ (i=1,2,\cdots,n;\ h=1,2,\cdots,p)$

この問題においても，重みw_1, w_2は，$0<w_2\ll w_1$を満たすように与えられる．また，制約式(45)，(46)は，(44)式を，(10)式と(12)式を用いて書き換えたものである．

4.4 判別規則

いま，新たな観測値$\boldsymbol{X}_j$(n次元区間ベクトル)が得られ，Inclusion問題を解くことにより得られた区間線形判別関数により，1次元上の区間Y_jに写像されたとする．この区間Y_jを判別するため，Inclusion問題の制約条件を考慮し，次のような判別規則を提案する．

[Inclusion問題における判別規則]

1次元上に写像された区間Y_jが次式を満たせば，観測値$\boldsymbol{X}_j$を$Gh(h=1,2,\cdots,p)$と判別する．

$$Y_j\subset Y_{Gh} \tag{47}$$

5. 判別問題の性質

これまで，区間データに対する判別分析の問題に対し，Intersection問題とInclusion問題という二つの判別問題の定式化を行なった．ここでは，この二つの定式化に関する性質について考察する．

まず，区間データの特別な場合として，すべての区間データの幅が0の場合，すなわち与えられたデータが点データの場合を考えると，次の定理が得られる．

[定理1]

与えられたデータが点データであるときは，Intersection問題の最適解と，Inclusion問題の最適解は一致する．

(証明)

与えられたデータが点データであるので，すべてのデータの幅は0である．すなわち，

$$d_{ij}=0 \text{ for all } i,j \tag{48}$$

である．上式を，Intersection問題とInclusion問題の制約条件に代入すると，それぞれの問題の制約条件は一致し，その結果，二つの問題は全く同じ問題となる．したがって，最適解も一致する．　(証明終り)

この定理と全く同様にして，次の定理を証明することができる．

[定理2]

与えられたデータが点データであるときは，多群Intersection問題の最適解と，多群Inclusion問題の最適解は一致する．

この二つの定理は，区間Y_jの幅が0であるときには次の二つの制約条件が同じ条件になることから導かれる定理である．

$$Y_j\cap Y_{Gh}\neq\phi \text{ for all } j \text{ in } Gh \ (h=1,2,\cdots,p)$$

$$Y_j\subset Y_{Gh} \text{ for all } j \text{ in } Gh \ (h=1,2,\cdots,p)$$

つぎに，データが区間データの場合について，二つの

判別問題の比較を行なうことにする．データが区間の場合では，Fig. 3 と Fig. 5 から明らかなように，次の定理が成り立つ．

[定理 3]

Intersection 問題の最適解を r^*_{Int}，Inclusion 問題の最適解を r^*_{Inc} とする．このとき，$r^*_{\mathrm{Int}} \geqq r^*_{\mathrm{Inc}}$ が成り立つ．

(証明)

二つの判別問題において，重み w_1, w_2 は，$0 < w_2 \ll w_1$ を満たすように与えられる．したがって，二つの判別問題の目的関数は共に r の最大化とみなすことができる．また，制約条件は，区間の交わりと区間の包含の定義からわかるように，Intersection 問題の方が Inclusion 問題より緩やかな条件である．したがって，$r^*_{\mathrm{Int}} \geqq r^*_{\mathrm{Inc}}$ が成り立つ．　(証明終り)

同様に，多群判別の場合にも次の定理が成り立つ．

[定理 4]

多群 Intersection 問題の最適解を r^*_{Int}，多群 Inclusion 問題の最適解を r^*_{Inc} とする．このとき，$r^*_{\mathrm{Int}} \geqq r^*_{\mathrm{Inc}}$ が成り立つ．

これらの定理から，Inclusion 問題により得られた区間 Y_{Gh} に交わりがない場合，すなわち $r^*_{\mathrm{Inc}} > 0$ の場合では，Intersection 問題により得られた区間 Y_{Gh} にも交わりがないことがわかる．また，Inclusion 問題により得られた区間 Y_{Gh} に交わりがあるような場合，すなわち $r^*_{\mathrm{Inc}} \leqq 0$ の場合でも，Intersection 問題から得られた区間 Y_{Gh} には交わりがないような場合もある．

Intersection 問題から得られた区間 Y_{Gh} には交わりがなく，Inclusion 問題による区間 Y_{Gh} に交わりがあるような場合では，区間データの判別に Intersection 問題を用いるべきであると思われる．しかし，二つの判別問題のどちらを区間データの判別に用いるべきかという問題に関する一般的な議論は困難であり，この問題は対象となる個々のデータに応じて使用者により決定されるべき問題である．

6. 数値計算例

ここでは，2 次元区間データを用いた 2 群判別の問題に関する簡単な数値計算例により，前章までに述べた判別問題の二つの定式化の説明を行なうことにする．

[例題 1]

$G1$ と $G2$ という 2 群に分けられた 2 次元区間データが Table 1 のように与えられているとする．このときの判別分析の問題は，次のような区間判別関数と 2 群を表わす区間 Y_{G1}, Y_{G2} を求める問題である．

$$Y_j = k_1 X_{1j} + k_2 X_{2j} = k_1\langle x_{1j}, d_{1j}\rangle + k_2\langle x_{2j}, d_{2j}\rangle \quad (49)$$

$$Y_{G1} = \langle a_1, c_1\rangle,\ Y_{G2} = \langle a_2, c_2\rangle \quad (50)$$

Table 1　Interval data of example 1

	G 1		G 2	
j	X_{1j} $\langle x_{1j}, d_{1j}\rangle$	X_{2j} $\langle x_{2j}, d_{2j}\rangle$	X_{1j} $\langle x_{1j}, d_{1j}\rangle$	X_{2j} $\langle x_{2j}, d_{2j}\rangle$
1	⟨1, 0.5⟩	⟨8, 1⟩	5 ⟨6, 1⟩	⟨9, 0.5⟩
2	⟨3, 0.5⟩	⟨8, 0.5⟩	6 ⟨7, 1⟩	⟨7, 0.5⟩
3	⟨2, 1⟩	⟨4, 1⟩	7 ⟨9, 0.5⟩	⟨8, 0.5⟩
4	⟨4, 1⟩	⟨6, 0.5⟩	8 ⟨10, 0.5⟩	⟨6, 0.5⟩

まず，Intersection 問題の解を求めるために，重み w_1, w_2 を $(w_1, w_2)^T = (100, 1)^T$ として**問題 P 2** を解いた結果，次のような解が得られた．ただし，重みを $(1000, 1)^T$ としても同じ解が得られた．

$$k_1^+ = 0.8,\ k_1^- = 0,\ k_2^+ = 0.2,\ k_2^- = 0$$
$$a_1 = 3.25,\ c_1 = 0.25,\ a_2 = 8.1,\ c_2 = 0.6$$

この解は，(22) 式の相補条件を満たしているので，Intersection 問題の最適解と一致する．そこで，(21) 式の変数変換により，区間線形判別関数と 2 群を表わす区間 Y_{G1}, Y_{G2} が次のように得られる．

$$Y_j = 0.8\,X_{1j} + 0.2\,X_{2j} \quad (51)$$

$$Y_{G1} = \langle 3.25, 0.25\rangle,\ Y_{G2} = \langle 8.1, 0.6\rangle \quad (52)$$

Table 1 の区間データを図示すると Fig. 6 のようになり，2 直線 g_1, g_2 の間の空間が，$Y_j \cap Y_{G1} \neq \phi$ を満たす $G1$ の領域であり，この領域と交わりのある区間データが $G1$ に判別される．また，2 直線 g_3, g_4 の間の空間が，$Y_j \cap Y_{G2} \neq \phi$ を満たす $G2$ の領域であり，この領域と交わりのある区間データが $G2$ に判別される．

つぎに，Table 1 の区間データに対して**問題 P 4** を

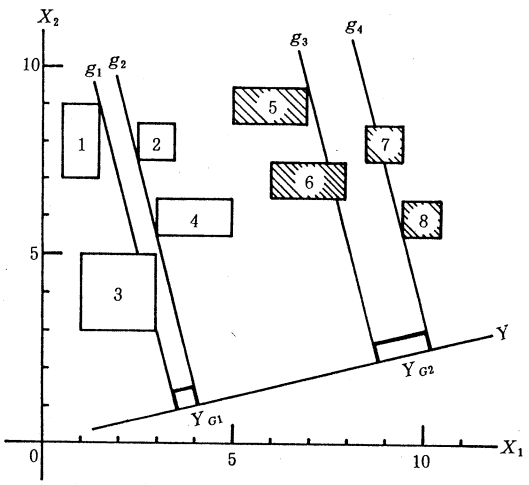


Fig. 6　Interval data and two group intervals by Intersection Problem

解くことにより，Inclusion 問題の解が次のように得られた．ただし，重み w_1, w_2 は，$(w_1, w_2)^T = (100, 1)^T$ および $(w_1, w_2)^T = (1000, 1)^T$ としたが，同じ解が得られた．

$$Y_j = 0.667\, X_{1j} + 0.333\, X_{2j} \qquad (53)$$

$$Y_{G1} = \langle 3.583, 1.917\rangle,\ Y_{G2} = \langle 7.667, 1.5\rangle \qquad (54)$$

このとき，Fig. 7 に示す 2 直線 g_1, g_2 の間の空間が，$Y_j \subset Y_{G1}$ を満たす $G1$ の領域であり，この領域に包含される区間データが $G1$ に判別される．また，2 直線 g_3, g_4 の間の空間が $Y_j \subset Y_{G2}$ を満たす $G2$ の領域であり，この領域に包含される区間データが $G2$ に判別される．

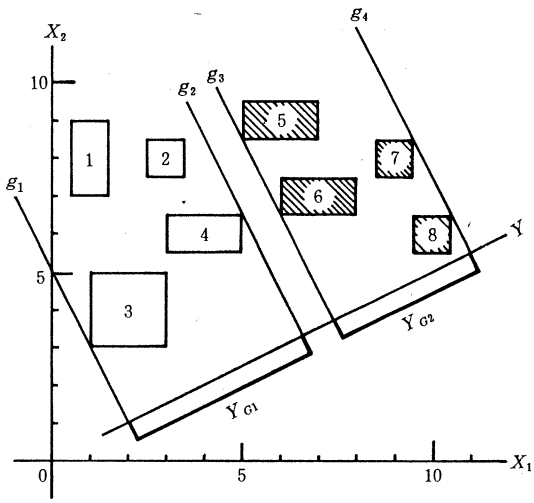


Fig. 7　Interval data and two group intervals by Inclusion Problem

Fig. 6 および Fig. 7 からわかるように，本研究で定式化した判別問題は，各群に属するサンプルが存在する領域を求める問題である．これは，属性空間を 2 分割する超平面を求める問題となる通常の判別分析における 2 群判別の問題と大きく異なる点である．

[例題 2]

例題 1 と同様に，2 群に分けられた 2 次元区間データが Table 2 のように与えられているとする．この区間データを Fig. 8 に示す．

まず，Table 2 のデータに対して**問題 P 2** を解くと，相補条件 (22) を満たす解が得られ，Intersection 問題の最適解が次のように得られた．

$$Y_j = 0.8\, X_{1j} + 0.2\, X_{2j} \qquad (55)$$

$$Y_{G1} = \langle 4.75, 0.75\rangle,\ Y_{G2} = \langle 8.45, 0.75\rangle \qquad (56)$$

このとき，(56) 式からわかるように，二つの区間 Y_{G1}, Y_{G2} に交わりはなく，$r^*_{\mathrm{Int}} = 2.2$ となる．

Table 2　Interval data of example 2

G 1		G 2	
j	X_{1j} $\langle x_{1j}, d_{1j}\rangle$	X_{2j} $\langle x_{2j}, d_{2j}\rangle$	j	X_{1j} $\langle x_{1j}, d_{1j}\rangle$	X_{2j} $\langle x_{2j}, d_{2j}\rangle$
1	$\langle 2, 0.5\rangle$	$\langle 9, 1\rangle$	5	$\langle 7, 1\rangle$	$\langle 6, 0.5\rangle$
2	$\langle 5, 1\rangle$	$\langle 8, 1\rangle$	6	$\langle 9, 1\rangle$	$\langle 9, 1\rangle$
3	$\langle 4, 1\rangle$	$\langle 3, 1\rangle$	7	$\langle 8, 1\rangle$	$\langle 2, 0.5\rangle$
4	$\langle 7, 1\rangle$	$\langle 4, 0.5\rangle$	8	$\langle 11, 0.5\rangle$	$\langle 5, 1\rangle$

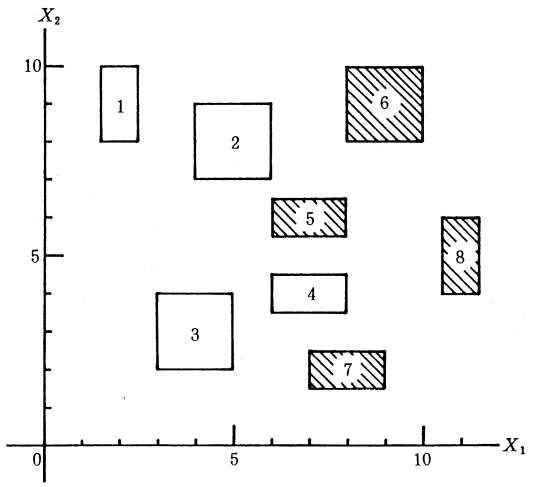


Fig. 8　Interval data of example 2

つぎに，Table 2 のデータに対して**問題 P 4** を解くと，次のような相補条件 (22) を満たさない解が得られた．

$$k_1^+ = 0.3333,\ k_1^- = 0,$$
$$k_2^+ = 0.375,\ k_2^- = 0.2917$$
$$a_1 = 1.9583,\ c_1 = 1.375,\ a_2 = 3.5417,\ c_2 = 1.375$$

(28) 式を用いれば，**問題 P 4** の最適解から Inclusion 問題の近似解が次のように得られる．

$$Y_j = \langle 0.3333\, x_{1j} + 0.0833\, x_{2j}$$
$$0.3333\, d_{1j} + 0.6667\, d_{2j}\rangle \qquad (57)$$

$$Y_{G1} = \langle 1.9583, 1.375\rangle,\ Y_{G2} = \langle 3.5417, 1.375\rangle \qquad (58)$$

(57) 式により 1 次元上に写像された区間と 2 群を表わす区間 Y_{G1}, Y_{G2} を Fig. 9 に示す．この図から，(57) 式による区間線形判別関数と (58) 式による区間 Y_{G1}, Y_{G2} が Inclusion 問題の制約条件である次式を満たし，Inclusion 問題の近似解となっていることがわかる．

$$Y_j \subset Y_{Gh} \text{ for all } j \text{ in } Gh\ (h = 1, 2) \qquad (39)$$

また，Inclusion 問題の最適解は，係数 k_1, k_2 の符号

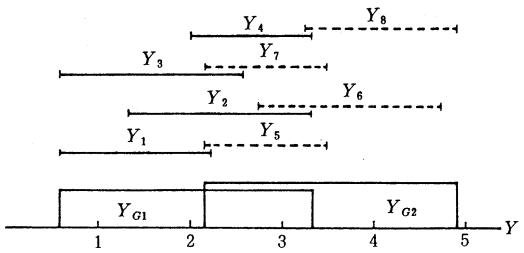


Fig. 9 Approximate solution of Inclusion Problem

のすべての組み合せ（4通り）に対してInclusion問題を解くことにより，次のように得られる．

$$Y_j = 0.8\,X_{1j} + 0.2\,X_{2j} \tag{59}$$

$$Y_{G1} = \langle 5.05, 2.25\rangle,\ Y_{G2} = \langle 8.15, 2.25\rangle \tag{60}$$

これをFig. 10に示す．Fig. 9とFig. 10から，Fig. 9に示すInclusion問題の近似解がFig. 10の最適解の良好な近似解になっていることがわかる．なお，Inclusion問題の最適解では，$r^*_{\mathrm{Inc}} = -1.4$となり，**定理3**で示した$r^*_{\mathrm{Int}} \geqq r^*_{\mathrm{Inc}}$が成り立っている．

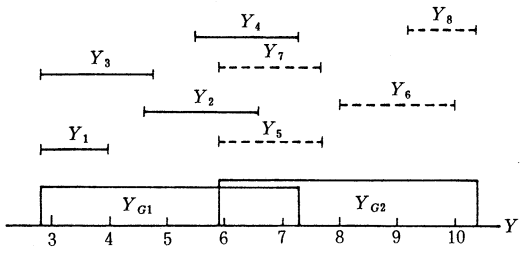


Fig. 10 Optimal solution of Inclusion Problem

7. 結 言

本研究では，多次元区間データに対する判別分析の手法を提案することを目的に，次のことを行なった．

(1) 多次元区間データに対する2群判別の問題を，Intersection問題とInclusion問題という二つの数理計画問題に帰着させ，線形計画法を用いた解法を示した．さらに，2群判別の問題に対して定式化した二つの数理計画問題を，多群判別の問題に拡張した．

(2) 与えられたデータが通常の点データの場合では，本研究で定式化した二つの数理計画問題の最適解が一致することを示した．

(3) 数値計算例により，二つの数理計画問題の説明を行なった．

参 考 文 献

1) 仮谷：対応のある2標本の区間データに基づく一般化符号検定；応用統計学，16巻，2号，pp. 77～88 (1987)
2) N. Freed & F. Glover : Simple but Powerful Goal Programming Models for Discriminant Problem ; European Journal of Operational Research, Vol. 7, No. 1, pp. 44～60 (1981)
3) N. Freed & F. Glover : A Linear Programming Approach to the Discriminant Problem ; Decision Sciences, Vol. 12, No 1, pp. 68～74 (1981)
4) 田中，和多田，林：ファジィ線形回帰分析の三つの定式化；計測自動制御学会論文集，22巻，10号，pp. 1052～1057 (1986)