数理解析研究所講究録
60巻 1968年 25-34

多変量層別について

阪大 基礎工　　石井 恵一
統計数理研　　多賀 保志

§1. 序

層別法や配分法の良い悪いの基準は，1次元分布の場合には推定量の分散の大小によるのが普通であるが，多変量分布のときには共分散行列の間に最初から全順序をつけてしまうのはやや不自然なので，何らかの自然な半順序に関して比較する必要がある．1次元変量を区間に層別するさいの最適層別についてはHayashi [1], Dalenius [2] により論じられた。多変量についてAoyama [3], Ghosh [4] 等の扱ったのは格子層別あるいはそれに類似のものの中での最適層別であり，一般化分散を比較の基準とするものである。区間層別や格子層別に限定せず，あらゆる層別法の中での最適性を考えるために，一般な層別を次のように定義する。

p次元確率ベクトル$X=(X_1,\ldots,X_p)$の母集団Πをℓ個の層$\Pi_1,\ldots,\Pi_\ell$に層別することは，Xの分布関数$F(x)$を

$F(x) = \sum_{i=1}^{l} F_i(x)$ と表現することにほかならない。ただし，$F_i(x)$ は（全測度 ≤ 1 の）p 次元分布関数である。$\{F_i\}$ を F の l-分解と呼ぶ。各 F_i は F に関し絶対連続だから密度 ϕ_i をもち，結局，Π の層別（$\Pi_1, \ldots, \Pi_l$）は

$$
(1) \qquad \sum_{i=1}^{l} \phi_i(x) = 1, \quad 0 \leq \phi_i(x) \leq 1 \qquad (\mathrm{a.e.}\ F)
$$

なる 1 の分解 $\phi = (\phi_1, \ldots, \phi_l)$ によって表現される。全空間 R^p を l 個の互いに素な領域に分けるのはもちろんその特別な場合である。上のような一般の層別法の中での最適条件は，1次元比例配分の場合に Taga [5] が論じ，多変量比例配分および多変量 Neyman 配分に対し，分散行列のトレースを基準にとった場合につき Isii [6] に論じてある。今回の報告では分散行列をある半順序に関して minimal にする層別法の条件を扱う。

さて，各層 Π_i から大きさ n_i の標本をとるとき，母平均 μ の不偏推定量 $\bar{X} = \sum_{i=1}^{l} w_i \bar{X}_i$ の分散行列は $\Sigma_{\bar{X}} = \sum_{i=1}^{l} \frac{w_i^2}{n_i} \Sigma_i$ となる。ただし，$w_i = n_i/n$，$n = \sum_{i=1}^{l} n_i$，$\bar{X}_i$ は第 i 層からの標本平均であり，Σ_i は Π_i の分散行列である。層別法と n を与えたとき，“最適な”配分 $\{n_i\}$ は1次元のときはよく知られた Neyman 配分，多次元では後述する最適配分であるが，層内分散を使わずにすむ“比例配分”も実用上よく用い

られる。一方，層別に対する配分のルールを最適配分なり比例配分なりに指定しておいて，そのルールのもとでどの層別が良いかを考えるのが最適層別の問題である。そこで，§2では多変量比例配分における最適層別の条件，§3では分散行列の実数値関数を比較の基準にとったときの最小問題を論ずる。§4で決定問題的な定式化をしたときのBayes解と§3との関係を，§5で1次元のNeyman配分に相当する“多変量最適配分”とそのもとでの最適層別を考える。

なお，Fに対し実際に最適層別を定める実用的な方法と，Fが現実には未知のため予備的な1次標本の情報に基づいて最適層別を“推定”する方法とが重要な実用上の問題であるが，ここではその基礎になる理論的な結果のみにとどめる。

§2. 多変量比例配分に対する層別

$X=(X_1, \ldots, X_p)'$.

F : Xの分布で，平均$\mu=(\mu_1, \ldots, \mu_p)'$，分散行列$\Sigma$.

$\phi=(\phi_1, \ldots, \phi_l)$: R^p上の1つのl-分解，$\Phi=\{\phi$の全体$\}$.

$(n_1, \ldots, n_l)$: 標本の大きさnの各層への配分 $(n=\sum_{i=1}^{l} n_i)$

$w_i=\int \phi_i(x)\,dF(x)$: 第i層のweight.

$\bar{X}=\sum_{i=1}^{l} w_i \bar{X}_i$: μの不偏推定量，$\bar{X}_i$: 第i層の標本平均.

比例配分 : $n_i=w_i n$.

さて，nとϕを与えたとき，比例配分による$\bar{X}$の分散行列

を $E[(\bar{x}-\mu)(\bar{x}-\mu)'|\phi]$ と書けば，$u_i=\int x\phi_i(x)\,dF(x)$ とおいて，

(2)　$$nE[(\bar{x}-\mu)(\bar{x}-\mu)'|\phi]=\int xx'dF(x)-\sum_{i=1}^{l}\frac{u_i u_i'}{w_i}$$

である。これを簡単に $\Sigma(\phi)$ としるす。

分散行列の大小の順序としては，$A-B$ が非負値行列のとき $A\geq B$ と定義する半順序を考えるのが最も自然で一般的であろう。これについて次の補題をあとで何度も用いる。

［補題 1］ $p\times p$ 実対称行列の全体を $\mathfrak{S}$ とし，普通の演算の意味でベクトル空間と考える。

(i) 要素ごとの収束で位相を入れると $\mathfrak{S}$ は線形位相空間として $R^{\frac{p(p+1)}{2}}$ と同型．

(ii) $\mathfrak{S}$ の中の非負値行列の全体 $\mathcal{P}$ は内点をもつ閉凸錐で，$A-B\in\mathcal{P}$ のとき $A\geq_{\mathcal{P}}B$ と定義すれば $\mathfrak{S}$ は半順序をもつ。

(iii) $\mathcal{P}$ の内点は正値行列と同値。

(iv)（非負線形汎関数の表現） $\forall T\in\mathcal{P}^*,\ \exists B\in\mathcal{P}$: $TZ=tr(BZ)$, $Z\in\mathfrak{S}$. ただし $\mathcal{P}^*$ は $\mathcal{P}$ 上で非負値をとる $\mathfrak{S}$ 上の線形汎関数の全体を意味する。

さて，ある $\phi^*\in\Phi$ に対し，$\Sigma(\phi^*)$ が上の順序の意味で集合 $\{\Sigma(\phi):\phi\in\Phi\}$ の極小元をとるとき，ϕ^* を "admissible" と呼ぶことにする（この用語の根拠は §4）。目的は admissibility の条件およびこれに関する complete class をみつ

けることである。分布Fが $l-1$ 個以下の点から成る集合上に集中しているときは，1点ずつを含む領域に層別することにより $\Sigma(\phi)$ を最小（0行列）にできることは明らかだから，次の仮定をおいても一般性を失うことはない：

[仮定（A）]．分布Fの台は少なくとも l 個の点を含む．

いま，オ i 層における平均ベクトルを μ_i とすれば $\mu_i=\dfrac{u_i}{w_i}$ である。これは ϕ の関数であるが，以下で特に ϕ^* としるす ϕ に対応する μ_i を μ_i^* と書く．

[定理 1]．比例配分の場合，admissible な ϕ^* の全体 Φ^* は最小完全類である。また，Fが仮定(A)を満足するならば，各 $\phi^*\in\Phi^*$ に対し $B\in\mathcal{P}$ が存在して，ϕ^* のオ i 層とオ j 層とは，$\dfrac{\mu_i^*+\mu_j^*}{2}$ を通り $B(\mu_i^*-\mu_j^*)$ に直交する超平面で分かたれる．従って超平面分割の全体は完全類である。

（証明の要点）$l(p+1)$ 次元ベクトル $(w_1,\cdots,w_l,u_1,\cdots,u_l)$ は Φ から $R^{l(p+1)}$ の中への線形写像として定まり（その像を $\mathcal{S}$ とする），$\Sigma(\phi)$ はこの写像と $\mathcal{S}$ から $\mathcal{Z}$ への連続凹（$\mathcal{P}$ 順序に関し）写像との合成として，Φ から $\mathcal{Z}$ への連続凹写像となる．Φ の位相は検定関数と同様なもの（Lehmann [7] 参照）を考えておく．Φ のコンパクト性に注意して Zorn の補題から容易に前半が導かれる．後半の証明には次の補題を用いる．この補題は以下の各節の結果の証明にも基本的である。

[補題2]．Φをベクトル空間の空でない凸部分集合，SをR^nの閉集合とし，$y(\phi)$をΦからSへの線形写像とする．また，$\mathcal{Z}$は線形位相空間で，内点をもつ閉凸錐$\mathcal{P}$によって半順序が導入されているものとする．さらに，$G(y)$をSから$\mathcal{Z}$への写像で次の意味で連続微分可能と仮定する：任意の$y=(y_1,\cdots,y_n)\in S$に対し，

$$\frac{\partial G(y)}{\partial y_i}=\lim_{\substack{y_i'\to y_i\\ y'\in S}}\frac{G(y')-G(y)}{y_i'-y_i}$$

が存在して連続 $(i=1,\cdots,n)$．ただし $y'=(y_1,\cdots,y_i',\cdots,y_n)$．

このとき，

(i) ϕ^*が$\{G(y(\phi)):\phi\in\Phi\}$の極小点をとるための必要条件は，$\phi^*$が$\{\sum_{i=1}^n y_i(\phi)\alpha_i^*:\phi\in\Phi\}$の境界点をとることである．ただし，

(3) $$\alpha_i^*=\frac{\partial G}{\partial y_i}(y(\phi))\Big|_{\phi=\phi^*},\quad i=1,\cdots,n.$$

(ii) さらにGが凹（$\mathcal{P}$順序の意味で）写像なら，ϕ^*は$\{\sum_{i=1}^n y_i(\phi)\alpha_i^*:\phi\in\Phi\}$の極小点をとる．また，$G(y(\phi))$の極小をとる$\phi^*$を固定し，$\alpha_i^*$を(3)で定義するとき，

$$\sum_{i=1}^n y_i(\phi)\alpha_i^*=\sum_{i=1}^n y_i(\phi^*)\alpha_i^*$$

を満足する任意のϕ'は$G(y(\phi'))=G(y(\phi^*))$を満足し，従ってまた$G(y(\phi))$の極小をとる．

この補題の証明は省略するが，これを $p \times p$ 対称行列の空間 $\mathfrak{X}$ と $y = (w_1, \cdots, w_\ell, u_1, \cdots, u_\ell)$ および (2) の右辺で定義される $G(y)$ に適用して，実際に α_i^* を計算すれば，$\Sigma(\phi)$ に関する非線形極小問題を $\{ \int \sum_{i=1}^{\ell} (x-\mu_i^*)(x-\mu_i^*)' \phi_i(x) dF(x) : \phi \in \Phi \}$ に関する線形極小問題に帰着することができる．これをさらに実数値関数の最小問題に帰着するために次の補題を用いる。

[補題3]. Φ, $\mathfrak{X}$, $\mathcal{P}$ は補題2と同様とし，ψ を Φ から $\mathfrak{X}$ への凹写像とする．このとき，$\phi^* \in \Phi$ に対し $\psi(\phi^*)$ が $\psi(\Phi) - \mathcal{P}$ の境界点をとるための必要十分条件は，$\mathfrak{X}$ 上の連続線形汎関数 T が存在して次の条件を満足することである：

$$T\psi(\phi) \leq T\psi(\phi^*), \quad \phi \in \Phi,$$
$$Tz \geq 0, \quad z \in \mathcal{P}.$$

この補題は基本的には，汎関数の拡張定理に基づく Mazur-Bourgin の定理から導かれる。この補題と補題1の表現を用いれば定理1の後半が証明されるわけである。

§3. 実数値関数による criteria.

実用上の目的に応じて，ある実数値関数 $f(\Sigma(\phi))$ を最小にする層別を求める問題も起こる．この f としては，前節の半順序に関して order-preserving であるべきであろう．f が

凹ならば，前節の $\mathfrak{X}$ を単に R^1 でおきかえて殆んど平行に議論できて，より簡単である。ここでは一般の f についての話はやめて，その典型的な例：ある正値行列 C によって $f(A)=tr(CA)$ と定義される場合について具体的な結果を述べる．

[定理2]． C を与えられた $p\times p$ 正値行列とする。比例配分の場合，$tr(C\Sigma(\phi))$ を最小にする ϕ^* が存在し，仮定(A)のもとに，ϕ^* の第 i 層と第 j 層とは，$\frac{\mu_i^*+\mu_j^*}{2}$ を通り $C(\mu_i^*-\mu_j^*)$ に直交する超平面で分かたれる。しかも，境界上に F の正の質量はない（狭義超平面分割）．

（注意） C が非負値で $rank(C)<p$ のときは，全空間を C の固有空間に射影して考えればよい。

§4．決定問題としての定式化．

§2の半順序に関して ϕ^* が極小をとることは，“任意の $\theta\in R^p$，$\|\theta\|=1$，に対して $\theta'\Sigma(\phi)\theta \le \theta'\Sigma(\phi^*)\theta$ かつ少なくとも1つの θ に対して不等号” となる ϕ が存在しないこと」と同値だから，$\{\|\theta\|=1,\ \theta\in R^p\}$ を母数空間とみなすことにより，決定問題の1つの型である．この意味で“admissible”とか“完全類”という語を用いた．しかし，decision ϕ について risk $\theta'\Sigma(\phi)\theta$ が線形構造をもたない点で，普通の統計的決定問題とは構造が異なる。ここで，$\{\theta : \|\theta\|=1\}$ に a

priori distribution $\mathcal{P}$ を入れて Bayes 解を考えてみよう.

$$E_P[\theta'\Sigma(\phi)\theta] = \int \theta'\Sigma(\phi)\theta\, dP(\theta) = tr\left[\int \theta\theta' dP(\theta)\cdot\Sigma(\phi)\right]$$

が Bayes risk である。$C=\int\theta\theta' dP(\theta)$ とおくと, C は非負値かつ $tr(C)=1$. 逆にそのような C はある $\mathcal{P}$ によって $C=\int\theta\theta' dP(\theta)$ と書けるから, 前節の $tr(C\Sigma(\phi))$ の最小問題は Bayes 解を求めることにほかならない。以上のことから,

[定理 3] 比例配分において, Bayes 解の全体は狭義超平面分割の全体に含まれるが, これは一般には完全類ではない。$p=1$ のときは完全類である (§2 と §3 は一致).

§5. 多変量における最適配分

固定した層別に対する最適配分法は 1 次元ではよく知られた Neyman 配分であるが, 多変量では分散行列の半順序に関して極小をとる配分法を最適と考えれば, 一般には一意的でない。層別 ϕ を固定すると, $\gamma_i = n_i/n$ とおいて,

$$(4)\qquad E[(\bar{x}-\mu)(\bar{x}-\mu)'\mid\phi] = \sum_{i=1}^{l} w_i^2\frac{\Sigma_i}{n_i} = \frac{1}{n}\sum_{i=1}^{l}\frac{w_i^2}{\gamma_i}\Sigma_i$$

と書ける。これを $\gamma=\{\gamma_i\}$ の関数と考えて補題 1, 3 を適用することにより, (4) の極小をとる $\gamma=\gamma^*$ は, ある $A\in\mathcal{P}$ に対して次の式を満足するものとして与えられる:

$$(5)\qquad \gamma_i^* = \frac{w_i\sqrt{tr(A\Sigma_i)}}{\sum_{j=1}^{l} w_j\sqrt{tr(A\Sigma_j)}},\qquad i=1,\cdots,l.$$

逆に正値の A を与えたとき，(5) により定まる γ^* は1つの最適配分になる．これは，$\|x\|_A^2 = x'Ax$ で R^p にノルム $\| \ \|_A$ を入れたとき，$E[\|\bar{x}-\mu\|_A^2]$ を最小にする配分にほかならない。

そこで次に，A を固定し ϕ と γ の関係を (5) で指定しておいて，$E[(\bar{x}-\mu)'A(\bar{x}-\mu)]$ を最小にする ϕ を定める場合を考えれば，ϕ^* がこれの最小値をとる必要条件として次のことがわかる：

$w_i^* = \int \phi_i^* dF$，$\mu_i^* = \frac{1}{w_i^*}\int x\phi_i^*(x)dF$，$\tau_i^{*2} = \frac{1}{w_i^*}\int (x-\mu_i^*)'A(x-\mu_i^*)\phi_i^*(x)dF(x)$

とおくとき，ϕ^* は ϕ の関数

$$H(\phi) = \int \sum_{i=1}^{\ell}\left[\frac{1}{\tau_i^*}(x-\mu_i^*)'A(x-\mu_i^*) + \tau_i^*\right]\phi_i(x)\,dF(x)$$

の最小値をとる．従って，ϕ^* は各層が高々 $\ell-1$ 個の2次超曲面で囲まれる（一般に連結でない）集合になるような分割で与えられる。

文　献

[1] C. Hayashi and F. Maruyama, "On some method for stratification" (in Japanese), Research Memoir, Inst. Statist. Math. 4 (1948), 399−411.

[2] T. Dalenius, "The problem of optimum stratification", Skand. Aktuar. 3-4 (1950), 203−213.

[3] H. Aoyama, "Stratified random sampling with optimum allocation for multivariate population", Ann. Inst. Statist. Math., 14 (1963) 251−258.

[4] S. P. Ghosh, "Optimum stratification with two characteristics," Ann. Math. Statist., 34 (1963), 866−872

[5] Y. Taga, "On optimum stratification for the objective variable based on concominant variables using prior informations". Ann. Inst. Statist. Math., 19 (1967), 101−129.

[6] K. Isii, "Mathematical programming and statistical inference", (in Japanese), Sūgaku, 18 (1966), 85−95.

[7] E. L. Lehmann, Testing Statistical Hypotheses, John Wiley, 1959.