

Some Nonparametric Methods
for Multivariate Analysis

島根大 文理 田村 亮二

§ 1. 序

1955年に Hodges [7] が Bivariate sign test を提唱してから Blumen [4], Bennett [2] 等による貢献はあつたが nonparametric methods の関心は殆んど 1 変量の場合に限られていたといつてもよい。しかし 1964 年頃から多変量に対する nonparametric methods が Berkeley を中心として, Bickel [3], Puri-Sen [9] 等により系統的に研究されはじめた。我国では杉浦 [11], 田村 [12] がこの問題に関心を示した。彼等はいづれも Hotelling の T^2 -statistics の nonparametric 化をねらつたのであるが, Puri-Sen が permutation test を用いて“厳密に” nonparametric な test の構成に達したのは注目すべきである。一方 Anderson [1] は Wilks [14] によつて導入された statistically equivalent block の概念を用いて“厳密に” nonparametric な test を提唱したが, 検定力等

については何も得られていない。

さて本報告は下にのべるような多重比較に関する nonparametric procedure の研究であるが、1変量の場合はすでに Dunn [6], Steel [10], 田村 [13] 等の結果がある。多変量の場合では、正規分布の仮定の下で Krishnaiah - Rizvi [8] 等の研究があるが nonparametric な立場での検討はまだなされていない。

p -変量の処理母集団を π_i , その cdf を $F_i(x) = F(x - \theta_i)$, $\theta_i = (\theta_i^{(1)}, \dots, \theta_i^{(p)})$ $i = 0, 1, \dots, c$ とし, π_0 は対照母集団とする。 $F(x)$ の連続性は仮定されているが, 分布型は未知とする。我々の目的は π_0 より "better" な処理と "not better" な処理を分離することであるが, 処理の "良さ" の基準として何を採用すべきかは現実の問題に則して決めらるべきであろう。本稿では与えられた $Q = (a_1, \dots, a_p)$ に対して $Q' \Delta_i > 0$ ならば π_i は π_0 より better, $Q' \Delta_i \leq 0$ ならば π_0 の方が π_i より better という基準をとる。 $\Delta_i = \theta_i - \theta_0$ 。そして $\Delta_i = Q$, $i = 1, \dots, c$ のとき $(1-\alpha)$ の確率で π_0 が best として選ばれなければならないという条件を付す。したがって問題を形式化すれば次のようになる。制約条件 (1) を満足しながら次の判定のうちの一つを選択せよ;

D_0 : π_0 が best である

D_i : π_i が π_0 より better な唯一の処理である

D_{ij} : π_i と π_j が π_0 より better な唯2つの処理である

...

$D_{1\dots c}$: すべての π_i は π_0 より better である

$$(1) \quad P_{\pi} [\forall \Delta_i = 0 \text{ のとき } D_0 \text{ が採択される}] = 1 - \alpha$$

処理の“良さ”の基準として Q'_i を採用したので制約条件としては“ $\forall Q'_i = 0$ のとき $1 - \alpha$ の確率で D_0 を採択せよ”の方がより自然と思われるが、この制限の下では適当な procedure を見出すことができるか？

§ 2. Nonparametric を比較方式

$F_i(\pi)$ からの任意標本を $O_i : \{X_{i1}, \dots, X_{in_i}\}$ とする

$$X_{id} = (X_{id}^{(1)}, \dots, X_{id}^{(p)}) \quad i=0, 1, \dots, c, \quad \sum_{i=0}^c n_i = N$$

よって Chernoff-Savage type の統計量を定義すると、

$$(2) \quad T_{Ni}^{(k)} = n_i^{-1} \sum_{d=1}^N E_{Nd} Z_{Ni,d}^{(k)} \quad k=1, \dots, p$$

$$\tilde{T}_{Ni} = (T_{Ni}^{(1)}, \dots, T_{Ni}^{(p)}) \quad i=0, 1, \dots, c$$

ただし、 $Z_{Ni,d}^{(k)}$ は大きさ N の pooled sample において k -成分の k 番最小値が O_i からであれば 1 をとり、その他のときは 0 をとる確率変数、 E_{Nd} は与えられた定数。

次に pooled sample の k -成分について $X_{id}^{(k)}$ の順位を $R_{id}^{(k)}$ で表し、rank matrix を R_N でかく。

$$(3) \quad R_N = \begin{bmatrix} R_{01}^{(1)} & \cdots & R_{0n_0}^{(1)} & \cdots & R_{cn_c}^{(1)} \\ \vdots & & \vdots & & \vdots \\ R_{01}^{(p)} & & R_{0n_0}^{(p)} & & R_{cn_c}^{(p)} \end{bmatrix}$$

$$(4) \quad v_{k,l} = N^{-1} \sum_{i=0}^c \sum_{d=1}^{n_i} E_{id}^{(k)} E_{id}^{(l)} - \bar{E}_N^2$$

$$\bar{E} = N^{-1} \sum_{d=1}^N E_{Nd}$$

$$\nabla(R_N) = [v_{k,l}] \quad k, l = 1, \dots, p$$

ただし, $E_{id}^{(k)}$ は $X_{id}^{(k)}$ に対応する E_{Ns} の値を表わす。

いま $E_{Nd} = \alpha$ にとれば $E_{id}^{(k)}$ は $X_{id}^{(k)}$ の rank に等しくなり

したがって $\nabla(R_N)$ は rank covariance matrix を表す。

最後に nonparametric procedure に用いる統計量を定義する

と,

$$(5) \quad W_{Ni} = T_{N(i)} / \left[\left(\frac{1}{\lambda_0} + \frac{1}{\lambda_i} \right) \underline{\alpha}' \nabla(R_N) \underline{\alpha} \right]^{\frac{1}{2}}$$

$$T_{N(i)} = \sqrt{N} \underline{\alpha}' (T_{Ni} - T_{N0})$$

$$0 < \lambda' \leq \lambda_i = n_i/N \leq 1 - \lambda' < 1, \quad \lambda' \leq 1/(c+1).$$

Nonparametric procedure w :

$W_{Ni} \leq t_i, \forall i$ ならば D_0 を採択せよ

$W_{Ni} > t_i, W_{Nj} \leq t_j, \forall j \neq i$ ならば D_i を採択せよ

.....

$W_{Ni} > t_i, \forall i$ ならば $D_{1 \dots c}$ を採択せよ

t_i は定数で, これらの決定については次の論ずる.

さて $X_{i\alpha}^{(n)}$ 等が観測されたとき R_N は定まるがこれを R_N^* と
 かし, R_N^* の各列を入れかえてできる $N!$ 個の rank matrix
 の全体を $S(R_N^*)$ で表わす. $\forall \alpha_i = 0$ のとき (これを H_0
 が真であると表現する) $S(R_N^*)$ の空間でのみ考えれば,
 この $N!$ 個の rank matrix は同じ確率をもつ. すなわち

$$P_r [R_N = r | H_0, S(R_N^*)] = 1/N! .$$

かつ W_{Ni} は R_N^* へのみ依存するので, 条件確率

$$P_r [(W_{N1}, \dots, W_{Nc}) \in w_d(R_N^*) | H_0] = 1-d$$

が満足するように各 $S(R_N^*)$ に対して $w_d(R_N^*)$ を決めるこ
 とができる. したがって

$$(6) P_r [W_{Ni} \leq t_i(R_N), \forall i | H_0] = 1-d$$

ならしめる $t_i(R_N)$ (確率変数であることに注意) を厳密に
 決定することができる. 実際には randomized technique
 が必要になる.

補助定理 1

$S(R_N^*)$ で考えたとき, H_0 が真であることを

$$E(T_{N(\alpha)}) = 0, \text{ cov. matrix of } T_{N(\alpha)} = \frac{N}{N-1} \{ \alpha' \nabla(R_N^*) \alpha \} \Lambda'$$

$$\Lambda' = [\lambda'_{ij}] \quad \lambda'_{ij} = \lambda_0^{-1} + \delta_{ij} \lambda_i^{-1} \quad i, j = 1, \dots, c$$

証明。

$$\text{Var } T_{Ni}^{(k)} = \frac{N-n_i}{(N-1)n_i} \left[N^{-1} \sum E_{Nd}^{(k)2} - \bar{E}_N^2 \right]$$

$$\text{Cov}(T_{Ni}^{(k)}, T_{Nj}^{(l)}) = \frac{N\delta_{ij} - n_i}{(N-1)n_i} \left[N^{-1} \sum_{l=0}^C \sum_{d=1}^N E_{id}^{(k)} E_{id}^{(l)} - \bar{E}_N^2 \right]$$

より容易に得られる。

上記は“厳密に” n_i を決定すること ε のべたが、漸近的立場においては次のような考察がなされる。まず Chernoff-Savage type [5] の仮定をおく。

$X_{id}^{(k)}$, $(X_{id}^{(k)}, X_{id}^{(l)})$ の経験周辺分布とそれぞれ $F_{Ni}^{(k)}(x)$, $F_{Ni}^{(k,l)}(x, y)$ とおき

$$H_N^{(k)}(x) = \sum_{i=0}^C \lambda_i F_{Ni}^{(k)}(x), \quad H^{(k)}(x) = \sum_{i=0}^C \lambda_i F_i^{(k)}(x)$$

$$H_N^{(k,l)}(x, y) = \sum_{i=0}^C \lambda_i F_{Ni}^{(k,l)}(x, y)$$

とする。また

$$E_{Nd} = J_N(d/N+1) \text{ とおく。}$$

$$(i) \quad J_N(u) \xrightarrow{N \rightarrow \infty} J(u), \quad 0 < u < 1$$

$$(ii) \quad \int_{-\infty}^{\infty} \left[J_N \left\{ \frac{N}{N+1} H_N^{(k)}(x) \right\} - J \left\{ \frac{N}{N+1} H_N^{(k)}(x) \right\} \right] dF_{Ni}^{(k)}(x) = o_p(1/\sqrt{N})$$

$k=1, \dots, P; \quad l=0, 1, \dots, C$

(iii) ある $\delta > 0$ に對して

$$\left| \frac{d^r}{du^r} J(u) \right| \leq K [u(1-u)]^{-r-\frac{1}{2}+\delta}, \quad r=0, 1$$

$$(iv) \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[J_N \left\{ \frac{N}{N+1} H_N^{(k)}(x) \right\} J_N \left\{ \frac{N}{N+1} H_N^{(l)}(y) \right\} - J \left\{ \frac{N}{N+1} H^{(k)}(x) \right\} J \left\{ \frac{N}{N+1} H^{(l)}(y) \right\} \right] \times dF_{Ni}^{(k,l)}(x, y) = o_p(1)$$

次の定理は Puri-Sen [9] および田村 [12] の結果から得られる。

定理2 Chernoff-Savage の条件を仮定する。 H_0 の下では

(i) $S(R_N^*)$ が与えられたとき, $W_{N\alpha}$ の条件付漸近分布は

$$N(Q, \Lambda), \quad \Lambda = [\lambda_{ij}]$$

$$\lambda_{ij} = \begin{cases} 1 & i=j \\ [\lambda_i \lambda_j / (\lambda_i + \lambda_j)(\lambda_i + \lambda_j)]^{\frac{1}{2}} & i \neq j \end{cases}$$

(ii) $t_i(R_N^*) = t(R_N^*)$ とすれば $t(R_N) \xrightarrow[N \rightarrow \infty]{P} t$ (定数)

ただし t は

$$(7) \quad \int_{-\infty}^t \cdots \int_{-\infty}^t n(Q, \Lambda) \prod_{i=1}^c dx_i = 1 - \alpha$$

(iii) $W_{N\alpha}$ の漸近分布は $N(Q, \Lambda)$

この定理より大標本の場合には, (7) から t を求めて, non-parametric な comparison procedure を行うことができる。

§ 3. 漸近相対効率

前節で $\forall \Delta_i = Q$ のとき, π_0 が選出される確率が厳密に又は漸近的に $(1-\alpha)$ になるような方式を与えたが, $\forall Q'_i \Delta_i \leq 0$ のときは, もちろん π_0 が best であるので, π_0 が選出される確率は $(1-\alpha)$ 以上になることが望まれるのは当然である。実は一般的にはこのことは証明されず, $F(x)$ による条

件を付したときに成立つことが示される。

定理3 Chernoff-Savage の条件を仮定する。

$$\forall \Delta_i = O(1/\sqrt{n}), \quad \forall \alpha' \Delta_i = \delta_i/\sqrt{n}, \quad \delta_i \leq 0 \quad \frac{N}{n} \xrightarrow{n \rightarrow \infty} \lambda \text{ (定数)}$$

今 $F_i^{(k)}(\alpha) = F(\alpha)$ なるば

$$Pr [D_0 \text{ が採択} \mid \forall \alpha' \Delta_i \leq 0] \geq 1 - \alpha.$$

証明

まず W_{Ni} の漸近同時分布は $N(\underline{\mu}, \Lambda)$ に収束することは

田村 [12] と次のことから示される。 $\underline{\mu} = (\mu_1, \dots, \mu_c)$

W_{Ni} の $\nabla(R_N)$ に $\nu(F)$ を代入したものを W'_{Ni} とおく。ただし

$$\nu(F) = [\nu_{kl}]$$

$$(8) \quad \nu_{kl} = \begin{cases} \int_0^1 J^2(u) du - \left(\int_0^1 J(u) du \right)^2 & k=l \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} J[F^{(k)}(\alpha)] J[F^{(l)}(\beta)] dF^{(k,l)}(\alpha, \beta) - \left(\int_0^1 J(u) du \right)^2 & k \neq l. \end{cases}$$

(P) $W_{Ni} \rightarrow W'_{Ni}$ として W'_{Ni} は正規分布に法則収束する。

また mean vector については

$$\begin{aligned} E [T_{Ni}^{(k)} - T_{N0}^{(k)}] &= \int J \left[\sum_j \lambda_j F^{(k)}(x - \theta_j^{(k)}) \right] dF^{(k)}(x - \theta_0^{(k)}) \\ &\quad - \int J \left[\sum_j \lambda_j F^{(k)}(x - \theta_j^{(k)}) \right] dF^{(k)}(x - \theta_0) \\ &\sim \Delta_i^{(k)} \int_{-\infty}^{\infty} f^{(k)}(x) \frac{d}{dF^{(k)}(\omega)} J[F^{(k)}(\alpha)] dF^{(k)}(\alpha) \end{aligned}$$

より

$$\begin{aligned} \mu_i &= \sqrt{\lambda} \delta_i \int_{-\infty}^{\infty} \frac{d}{d\alpha} J[F(\omega)] dF(\alpha) / \left[\left(\frac{1}{\lambda_0} + \frac{1}{\lambda_i} \right) \alpha' \nu(F) \alpha \right]^{\frac{1}{2}} \\ &\leq 0 \quad (\text{ } J(u) \text{ を増加関数にとる}) \end{aligned}$$

かくて

$$\Pr[D_0 \text{ が採択} \mid \forall \alpha'_i \Delta_i \leq 0] \\ \sim \int_{-\infty}^{t_1 - \mu_1} \cdots \int_{-\infty}^{t_c - \mu_c} n(\underline{Q}, \Lambda) \prod_i d\alpha_i \geq 1 - \delta.$$

さて次は各判定 $D_{i_1 \dots i_r}$ が正しい確率を $P_{i_1 \dots i_r}$ で表せば、前定理から容易に次の表示をうることができる。

$$P_i = \Pr[W_{N_i} > t_i, W_{N_j} \leq t_j \quad \forall j \neq i \mid \sigma_i > 0, \sigma_j \leq 0, \forall j \neq i]$$

$$\sim \int_{t_i - \mu_i}^{\infty} d\alpha_i \int_{-\infty}^{t_1 - \mu_1} \cdots \int_{-\infty}^{t_c - \mu_c} n(\underline{Q}, \Lambda) \prod_{j \neq i} d\alpha_j$$

$$P_{ij} \sim \int_{t_i - \mu_i}^{\infty} \int_{t_j - \mu_j}^{\infty} d\alpha_i d\alpha_j \int_{-\infty}^{t_1 - \mu_1} \cdots \int_{-\infty}^{t_c - \mu_c} n(\underline{Q}, \Lambda) \prod_{k \neq i, j} d\alpha_k$$

$$\dots \dots \dots \\ P_{1 \dots c} \sim \int_{t_1 - \mu_1}^{\infty} \cdots \int_{t_c - \mu_c}^{\infty} n(\underline{Q}, \Lambda) d\alpha.$$

今 $F(\alpha_i)$ が normal と仮定すれば次のような方式が用いられる。記法を今迄と多少変更して、各標本の大きさを $n'_i(n)$, $\sum_{i=0}^c n'_i = N'$, $n'_i/N' = p_i$, $N'/n \xrightarrow{n \rightarrow \infty} p$ とする。

$$(9) \quad \nabla_{N'_i} = \sqrt{N'} \alpha'_i (\bar{X}_i - \bar{X}_0) / \left[\left(\frac{1}{p_0} + \frac{1}{p_i} \right) \alpha'_i \Sigma \alpha \right]^{\frac{1}{2}} \\ i = 1, \dots, c$$

ただし Σ は $F(x)$ の Covariance matrix で \bar{X}_i は O_i の sample mean vector である。

Normal procedure V_i :

$V_{N_i} \leq t_i$, V_i ならば D_0 を採択せよ。

$V_{N_i} > t_i$, $V_{N_j} \leq t_j$ $V_{j \neq i}$ ならば D_i を採択せよ。

.....

$V_{N_i} > t_i$ $\forall i$ ならば $D_{1,2,\dots,c}$ を採択せよ。

定理4 $n \rightarrow \infty$ のとき $n_0/n_i \sim n'_0/n'_i$ にちよよに標本の大きさを大きくしていけば, 漸進的に (1) が成立する。

証明 H_0 が真のとき V_{N_i} の漸進分布は $N(Q, \Pi)$ であること, および $\forall \alpha_i = O(1/\sqrt{n})$, $\forall \alpha'_i = \delta_i/\sqrt{n}$ のときの漸進分布は $N(\lambda, \Pi)$ にちよよは容易に知られる, ただし

$$\lambda = (\lambda_1, \dots, \lambda_c), \quad \lambda_i = \sqrt{p} \delta_i / \left[\left(\frac{1}{p_0} + \frac{1}{p_i} \right) \alpha'_i \Sigma \alpha_i \right]^{\frac{1}{2}}$$

$$\Pi = [\pi_{ij}] \quad \pi_{ij} = \begin{cases} 1 & i=j \\ \left[p_i p_j / (p_0 + p_i)(p_0 + p_j) \right]^{\frac{1}{2}} & i \neq j \end{cases}$$

定理の条件より $\delta_i \sim \lambda_i$, (したがって $\Pi \sim \Lambda$) かくて

$$P_n [D_i \text{ が採択} \mid \forall \alpha_i = \underline{0}] \sim \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_c} n(Q, \Lambda) d\alpha_n \\ = 1 - \alpha$$

各判定 $D_{i_1 \dots i_r}$ の正しい確率を $P'_{i_1 \dots i_r}$ で表わせば,

$$P_i \sim \int_{t_i - \nu_i}^{\infty} dx_i \int_{-\infty}^{t_i - \nu_i} \dots \int_{-\infty}^{t_i - \nu_i} n(\underline{Q}, \Lambda) \prod_{j \neq i} dx_j$$

$$P_{ij} \sim \int_{t_i - \nu_i}^{\infty} \int_{t_j - \nu_j}^{\infty} dx_i dx_j \int_{-\infty}^{t_i - \nu_i} \dots \int_{-\infty}^{t_i - \nu_i} n(\underline{Q}, \Lambda) \prod_{k \neq i, j} dx_k$$

.....

$$P_{1 \dots r} \sim \int_{t_1 - \nu_1}^{\infty} \dots \int_{t_r - \nu_r}^{\infty} n(\underline{Q}, \Lambda) d\underline{x}$$

は前と同様にして得られる。

さて Procedure W の Procedure V に対する漸近相対効率とは、
 ある $P_{i_1 \dots i_r}$ と $P'_{i_1 \dots i_r}$ を等しくする sample size の
 逆比の極限によって定義すれば

定理 5. 定理 3, 4 の仮定の下で, W の V に対する漸近相対
 効率 $e_{W, V}$ は次式で与えられる

$$(10) \quad e_{W, V} = \left[\int_{-\infty}^{\infty} \frac{d}{d\underline{x}} J[F(\underline{x})] dF(\underline{x}) \right]^2 \underline{a}' \underline{\Sigma} \underline{a} / \underline{a}' \nu(F) \underline{a}.$$

証明 $\forall \mu_i = \nu_i$ にするからには $\forall P_{i_1 \dots i_r} = P'_{i_1 \dots i_r}$

にするから

$$\int_{-\infty}^{\infty} \frac{d}{d\underline{x}} J[F(\underline{x})] dF(\underline{x}) / \left[\left(\frac{1}{\lambda_0} + \frac{1}{\lambda_i} \right) \underline{a}' \nu(F) \underline{a} \right]^{\frac{1}{2}}$$

$$= 1 / \left[\left(\frac{1}{\lambda_0} + \frac{1}{\lambda_i} \right) \underline{a}' \underline{\Sigma} \underline{a} \right]^{\frac{1}{2}}$$

より (10) をうる。

文 献

- [1] Anderson, T.W. (1965) : Some nonparametric multivariate procedures based on statistically equivalent blocks. *Multivariate Analysis*. 5-27, Acad. Press.
- [2] Bennett, B.M. (1962) : On multivariate sign tests. *JRSS* 24, 159-161
- [3] Bickel, P.F. (1965) : On some asymptotically non-parametric competitors of Hotelling's T^2 . *AMS* 36, 160-173.
- [4] Blumen, I. (1958) : A new bivariate sign test. *JASS*, 53, 448-456
- [5] Chernoff, H. and Savage, I. R. (1958) : Asymptotic normality of certain nonparametric test statistics. *AMS* 29, 972-994
- [6] Dunn, O. J. (1964) : Multiple comparisons using rank sums. *Technometrics* 6, 241-252
- [7] Hodges, J. L. Jr (1955) : A bivariate sign test. *AMS* 26, 523-527
- [8] Krishnaiah, P.R. and Rizvi, M.H. (1966) : Some procedures for selection of multivariate normal populations better than a control. *Multivariate Analysis*, 477-490 Acad. Press
- [9] Puri, M.L. and Sen, P.K. (1966) : On a class of multivariate multi-sample rank-order tests. *Sankhya*, 28, 353-376.
- [10] Steel, R.G.D. (1961) : Some rank sum multiple comparison tests, *Biometrics*, 17, 539-552.

- [11] Sugiura, N. (1965) : Multisample and multivariate nonparametric tests based on Γ -statistics and their asymptotic efficiencies. Osaka J. of Math. Vol. 2 (1965) 385-426
- [12] Tamura, R. (1966) : Multivariate nonparametric several-sample tests. AMS 37, 611-618
- [13] Tamura, R. : Distribution-free multiple comparison procedures. (appear in Shimane Univ. Bull.)
- [14] Wilks, S. S. (1962) : Mathematical Statistics. Wiley