多変量解析における漸近展開: 微分作用素の使用の観点から

柿沢佳秀 (北海道大学大学院経済学研究科), 岩下登志也 (東京理科大学理工学部)

1. はじめに

Welch(1947; Biometrika)-James(1954; Biometrika) 法は 1950 年代から 1970 年代に多変量正規における諸統計量の漸近展開導出でしばしば利用され, 分布評価のための有力な手法の１つに位置づけられた. 典型例は正規 MANOVA の LH 統計量 (Siotani(1956,1971; AISM,AMS), Ito(1956,1960; AMS)) の非心分布である. Fujikoshi(1974; AISM) は GMANOVA 統計量の非心分布の漸近展開を求めたが, その導出法は正規 GMANOVA に対する正準型 (Gleser and Olkin(1970; In: *Essays in Probability and Statistics*)) から, $\mathrm{vec}(\mathbf{Z}') \sim N_{pN}(E[\mathrm{vec}(\mathbf{Z}')], \mathbf{I}_N \times \mathbf{\Sigma})$ なる正規行列 $\mathbf{Z}$ のパーティション $(\mathbf{Z}_{ab})_{a,b=1,2,3}$ において $\mathbf{Z}_{13}$ と $\mathbf{W}_{33} \equiv \mathbf{Z}'_{33}\mathbf{Z}_{33}$ を条件付けした分布論を MANOVA 統計量の非心ケースに帰着させ, Wishart 行列 $\mathbf{W}_{33}$ に関する期待値計算へ Welch-James 法を適用したことによる. 等分散な多変量正規の２群判別問題で, Okamoto(1963; AMS) は W-判別方式の誤判別確率の漸近展開導出のために標本平均ベクトルと標本分散行列の関数の期待値を微分作用素で公式化しており, それは Memon and Okamoto(1971; JMA) により Z-判別方式に対しても適用された. さらに, 多変量正規の分散行列に関する統計量では, 標本分散行列の関数の期待値の公式が杉浦・長尾・藤越による漸近展開導出の際に使用された. このような Welch-James 法を起源とするアプローチの背後には正規標本論での Cochran の定理があり, 独立な正規行列と Wishart 行列の関数あるいは複数の独立な Wishart 行列の関数として記述される場合には, 条件付き期待値を経由して Wishart 行列に関する期待値計算へ帰着させることができた.

一方, 非正規母集団ではこれらの事実は一般には成立せず, Wishart 分布が登場することも稀である. そこで, 報告者は最近の一連の研究において Welch-James 法の条件付き期待値をとるステップを外して考えることから, 非正規母集団における微分作用素アプローチを再考察し, １標本・２標本の Hotelling の T^2 統計量と１元配置モデルの LR,LH,BNP 統計量, 及び, James 統計量 (２標本の場合の BF 統計量も含む) の局所対立仮説の下での漸近展開導出に成功した (詳細は Kakizawa and Iwashita(2008a,b; To appear in JSPI,JMA), Kakizawa(2007; To appear in JJSS)). なお, 漸近展開の正当化については, 独立な確率ベクトルの和に関する Bhattacharya and Rao(1976; Wiley) を基礎として保証される.

2. 非正規 MANOVA, GMANOVA における漸近展開の進展

2.1. MANOVA の線形仮説

多変量線形回帰 (あるいは MANOVA) モデル $\mathbf{Y} = \mathbf{X\Theta} + \mathbf{U}$ において線形仮説 $\mathbf{B\Theta} = \mathbf{O}_{r\times p}$ を考える. ここに, 既知な $N \times q$ 計画行列 $\mathbf{X}$ のランクは q, 既知な線形制約の $r \times q$ 行列 $\mathbf{B}$ のランクは $r(\le q)$, そして $q \times p$ 係数行列 $\mathbf{\Theta}$ は未知母数とする. また, $\mathbf{U}'$ ($p \times N$ 行列) の i 列を $\mathbf{u}_i$ と書くとき $\mathbf{u}_1, \ldots, \mathbf{u}_N$ は平均ベクトル $\mathbf{0}$, 分散行列 $\mathbf{\Sigma}$, 高次キュムラント $\kappa_{j_1,\ldots,j_s}$ をもつ独立同一な p 次元確率ベクトルであるとする. 正規性の下での伝統的な LR,LH,BNP 統計量は仮説による平方和行列 $\mathbf{H}_Y = \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{B}'\{\mathbf{B}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{B}'\}^{-1}\mathbf{B}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ と残差平方和行列 $\mathbf{E}_Y = \mathbf{Y}'\{\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\}\mathbf{Y}$ を定義するとき, $\mathbf{H}_Y\mathbf{E}_Y^{-1}$ の固有値 $\lambda_{Y,j} \ge 0$ に基づいて統一的に議論される: $T_\psi = \nu\sum_{j=1}^{p}\psi(\lambda_{Y,j})$. 従って, 滑らかな関数 ψ に対して $\psi(0) = 0$, $\psi'(0) = 1$ (LR,LH,BNP は $\psi(x) = \log(1+x), x, 1-1/(1+x)$ の場合で, この条件を満足する) を仮定するときは特性関数について $E[\mathrm{e}^{\mathrm{i}tT_\psi}] \approx E[\mathrm{e}^{\mathrm{i}t\mathrm{tr}(\mathrm{H}_Y\widehat{\Sigma}_Y^{-1})}] + \frac{\mathrm{i}t\psi''(0)}{2N}E[\mathrm{tr}\{(\mathbf{H}_Y\widehat{\mathbf{\Sigma}}_Y^{-1})^2\}\mathrm{e}^{\mathrm{i}t\mathrm{tr}(\mathrm{H}_Y\widehat{\Sigma}_Y^{-1})}]$ となる. ここに, $\widehat{\mathbf{\Sigma}}_Y = \mathbf{E}_Y/(N-q)$ は $\mathbf{\Sigma}$ の不偏推定量である.

本報告の１つの貢献は Kakizawa and Iwashita(2008b; To appear in JMA) の拡張にある. すなわち, 位置不変性 $\widehat{\mathbf{\Sigma}}_Y = \widehat{\mathbf{\Sigma}}_U$ 及び, 局所対立仮説 $\mathbf{\Theta} = \mathbf{\Theta}_0 + (\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{\Theta}_\varepsilon$ (但し, $\mathbf{B\Theta}_0 = \mathbf{O}_{r\times p}$ と $\mathbf{B\Theta} \ne \mathbf{O}_{r\times p}$ を仮定する) の下での変形 $\mathbf{H}_Y = [\mathbf{U}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2} + \mathbf{\Theta}'_\varepsilon]\dot{\mathbf{M}}[\quad]'$ に注目し, $p \times r$ 確率行列 $\tilde{\mathbf{z}}_U^{(1:r)} \equiv \mathbf{U}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}\dot{\mathbf{V}}^{(1:r)}$ と $\widehat{\mathbf{\Sigma}}_U$ の多項式の関数の期待値公式を与えた. ここに,

$\dot{\mathbf{M}} = (\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{B}'\{\mathbf{B}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{B}'\}^{-1}\mathbf{B}(\mathbf{X}'\mathbf{X})^{-1/2}$ はランク r の巾等行列であり, スペクトル分解 $\dot{\mathbf{M}} = \dot{\mathbf{V}}^{(1:r)}(\dot{\mathbf{V}}^{(1:r)})'$ をもつ (なお, $\dot{\mathbf{V}} = [\dot{\mathbf{V}}^{(1:r)}, \dot{\mathbf{V}}^{(r+1,q)}]$ は $q \times q$ 直交行列である).

2.2. GMANOVA の一般化線形仮説

多変量一般化線形回帰モデル (GMANOVA の他, Potthoff and Roy(1964; Biometrika) モデル, 成長曲線モデル (GCM) など別名は多い) $\mathbf{Y} = \mathbf{X\Xi A} + \mathbf{U}$ において一般化線形仮説 $\mathbf{B\Xi C} = \mathbf{O}_{r\times s}$ を考える. ここに, 既知な $N \times q$ 計画行列 $\mathbf{X}$ のランクは q, $m \times p$ 計画行列 $\mathbf{A}$ のランクは $m(\le p)$, 既知な線形制約の $r \times q$ 行列 $\mathbf{B}$ と $m \times s$ 行列 $\mathbf{C}$ のランクはそれぞれ $r(\le q)$ と $s(\le m)$, そして $q \times m$ 係数行列 $\mathbf{\Xi}$ は未知母数とする. 正規性の下での LR,LH,BNP 統計量と Kleinbaum(1973; JMA) の Wald 型統計量を含む検定統計量のクラス $T_{(\psi,d)} = (N-q)\sum_{j=1}^{s}\psi(\lambda_{Y,j}^{(d)})$ の非正規かつ局所対立仮説 $\mathbf{\Xi} = \mathbf{\Xi}_0 + (\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{\Xi}_\varepsilon$ (但し, $\mathbf{B\Xi}_0\mathbf{C} = \mathbf{O}_{r\times s}$ と $\mathbf{B\Xi C} \ne \mathbf{O}_{r\times s}$ を仮定する) での検出力比較に興味があり, 特に, 非負の固定された実数 d と $\psi(0)=0$, $\psi'(0)=1$ を満たす滑らかな関数 $\psi(x)$ 及び非正規性の影響を調べることは重要である. ここに, $\mathbf{E}_Y^\circ = \mathbf{C}'(\mathbf{A}\mathbf{E}_Y^{-1}\mathbf{A}')^{-1}\mathbf{C}$,

$$\mathbf{R}_Y^{(d)} \equiv (\mathbf{X}'\mathbf{X})^{-1} + d(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\{\mathbf{E}_Y^{-1} - \mathbf{E}_Y^{-1}\mathbf{A}'(\mathbf{A}\mathbf{E}_Y^{-1}\mathbf{A}')^{-1}\mathbf{A}\mathbf{E}_Y^{-1}\}\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

$$\mathbf{H}_Y^{(d)} = (\mathbf{B}\widehat{\mathbf{\Xi}}_Y\mathbf{C})'(\mathbf{B}\mathbf{R}_Y^{(d)}\mathbf{B}')^{-1}(\mathbf{B}\widehat{\mathbf{\Xi}}_Y\mathbf{C}), \quad \widehat{\mathbf{\Xi}}_Y = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{E}_Y^{-1}\mathbf{A}'(\mathbf{A}\mathbf{E}_Y^{-1}\mathbf{A}')^{-1}$$

を定義したとき, $\lambda_{Y,j}^{(d)} \ge 0$ は $\mathbf{H}_Y^{(d)}(\mathbf{E}_Y^\circ)^{-1}$ の固有値である. 位置不変性 $\mathbf{E}_Y^\circ = \mathbf{E}_U^\circ$, $\mathbf{R}_Y^{(d)} = \mathbf{R}_U^{(d)}$ と局所対立仮説の下での変形 $\mathbf{B}\widehat{\mathbf{\Xi}}_Y\mathbf{C} = \mathbf{B}(\mathbf{X}'\mathbf{X})^{-1/2}\{(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}'\mathbf{U} + \mathbf{\Xi}_\varepsilon\mathbf{A}\}\widehat{\mathbf{\Sigma}}_U^{-1}\mathbf{A}'(\mathbf{A}\widehat{\mathbf{\Sigma}}_U^{-1}\mathbf{A}')^{-1}\mathbf{C}$ に注意し, $T_{(\psi,d)}$ を (かなり煩雑ではあるが) 確率展開することから, $\tilde{\mathbf{z}}_U^{(1:r)}$ と $\widehat{\mathbf{\Sigma}}_U$ の多項式の特性関数計算に帰着させた (MANOVA ケースの公式を適用する; 但し, 計算は極めて煩雑である).

3. 検出力比較

以上, 非正規母集団における平均推測の N^{-1} の漸近展開の最近の進展を微分作用素の使用の観点から報告した. ところで, 帰無分布 ($d=1$ は Yanagihara(2001,2007; HMJ,JJSS; Wakaki et al.(2002; HMJ) の GMANOVA 版)) について $\Pr[T_{(\psi,d)} \le CF(x)|H] = G_{rs}(x) + \mathrm{o}(N^{-1})$ となる Cornish-Fisher 型展開 $CF(x) = x\{1 + (2/N)\sum_{j=1}^{3} c_j x^{j-1}\}$, 及び, $\Pr[B(T_{(\psi,d)}) \le x|H] = G_{rs}(x) + \mathrm{o}(N^{-1})$ となる Bartlett 型変換 $B(x) = x\{1 - (2/N)\sum_{j=1}^{3} c_j x^{j-1}\}$ (なお c_j は $\psi''(0), d$ と母集団分布の分散行列 $\mathbf{\Sigma}$, 3次・4次キュムラントに依存する) が得られるから, これらの N^{-1} の漸近展開に基づく次の2つの (ψ,d)-検定を考える. $B_{(\psi,d)}$-検定では $\widehat{B}(T_{(\psi,d)}) > \chi^2_{rs,\alpha}$ のとき帰無仮説を棄却し, $CF_{(\psi,d)}$-検定では $T_{(\psi,d)} > \widehat{CF}(\chi^2_{rs,\alpha})$ のとき帰無仮説を棄却する. このとき本報告の主要結果として, 局所対立仮説の下での一般化線形仮説の検定の検出力に関する以下の結論を得た.

定理 [1 元配置モデル (Kakizawa and Iwashita(2008b; To appear in JMA)) の GMANOVA 版]
(i) $B_{(\psi,d)}$-検定の局所検出力 $\beta_N^{\mathrm{B}}(\alpha:\psi,d)$ は $CF_{(\psi,d)}$-検定の局所検出力 $\beta_N^{\mathrm{CF}}(\alpha:\psi,d)$ と N^{-1} まで一致する. (ii) (ψ,d_1)-検定の局所検出力は (ψ,d_2)-検定の局所検出力と N^{-1} まで一致する. 従って, N^{-1} の局所検出力は $d \ge 0$ に依存しない. (iii) N^{-1} の局所検出力の差は分布に無関係で,

$$\lim_{N\to\infty} N\{\beta_N^{\mathrm{CF}}(\alpha:\psi_2,d) - \beta_N^{\mathrm{CF}}(\alpha:\psi_1,d)\} = \lim_{N\to\infty} N\{\beta_N^{\mathrm{B}}(\alpha:\psi_2,d) - \beta_N^{\mathrm{B}}(\alpha:\psi_1,d)\}$$

$$= \frac{\psi_2''(0) - \psi_1''(0)}{2}\left[\mathrm{tr}(\mathbf{\Omega}_\circ^2) - \frac{r+s+1}{rs+2}\{\mathrm{tr}(\mathbf{\Omega}_\circ)\}^2\right] g_{rs+8}\{\chi^2_{rs,\alpha};\mathrm{tr}(\mathbf{\Omega}_\circ)\}.$$

ここに, $\mathbf{Q}^{\Sigma} = \mathbf{\Sigma}^{-1}\mathbf{A}'(\mathbf{A}\mathbf{\Sigma}^{-1}\mathbf{A}')^{-1}\mathbf{C}\{\mathbf{C}'(\mathbf{A}\mathbf{\Sigma}^{-1}\mathbf{A}')^{-1}\mathbf{C}\}^{-1}\mathbf{C}'(\mathbf{A}\mathbf{\Sigma}^{-1}\mathbf{A}')^{-1}\mathbf{A}\mathbf{\Sigma}^{-1}$ を定義するとき, 非心行列は $\mathbf{\Omega}_\circ = \mathbf{Q}^{\Sigma}(\mathbf{\Xi}_\varepsilon\mathbf{A})'(\lim_{N\to\infty}\dot{\mathbf{M}})(\mathbf{\Xi}_\varepsilon\mathbf{A})\mathbf{Q}^{\Sigma}\mathbf{\Sigma}$.

これにより, 藤越 (2003; 日本統計学会誌) の 2.4 節の "個人的チャレンジ" の一部が解決された. また, 誤差項を対数正規とする MANOVA の数値実験から, T_{LH} については多項式 $B(x)$ の非単調性が検出力損失の原因となる (Cordeiro-Ferrari(1991; Biometrika) による Bartlett 型統計量は非単調性のために帰無仮説から離れるほど負になる傾向が強い) ことが分かり, 検出力回復の策として単調な Bartlett 型調整 (Kakizawa(1996; Biometrika)) の重要性を指摘した.