（9）「統計的推測の理論とその応用」に関する研究報告

高次元多変量２値データの判別における変数選択

広島大学大学院理学研究科　　薬真寺　裕
広島大学大学院理学研究科　　安部　友紀
広島大学大学院理学研究科　　藤越　康祝

本報告においては、p 個のベルヌーイ変量 $x_1, \ldots, x_p$ が q 個のグループ $G_1, \ldots, G_q$ において観測される場合の判別分析における変数選択問題を扱う。第 i グループ G_i における p 次元変量 $\boldsymbol{x} = (x_1, \ldots, x_p)$ の観測値を

$$G_i;\ \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i}, \quad i = 1, \ldots, q$$

$$\boldsymbol{x}_{ij} = (x_{ij}(1), \ldots, x_{ij}(p)), \quad j = 1, \ldots, n_i$$

とする。

ここでは、変数の次元 p が全標本数 n に近いか、あるいは、それより大きい高次元データの場合を想定している。より具体的には、Wilber et al. (Biometrics, 58(2002)) による DNA フィンガープリントデータ解析の再検討を目指しており、この場合

$$p = 84, \quad q = 4, \quad n_1 = 23, \quad n_2 = n_3 = n_4 = 22, \quad n = n_1 + n_2 + n_3 + n_4 = 89$$

である。そこでは、p 個の変量は互い独立で、$x_k | G_i \sim B(1, \theta_{ik})$ が仮定されており、ここでもこれを仮定する。従って

$$\boldsymbol{y}_i = \boldsymbol{x}_{i1} + \ldots + \boldsymbol{x}_{in_i}$$

とおくと、各 y_{ij} に対して

$$y_{ij} (j = 1, \ldots, p; i = 1, \ldots, q) \text{ は独立で、} \quad y_{ij} \sim B(n_i, \theta_{ij})$$

を想定していることになる。

変量間の独立性を仮定しているので、周辺データに基づく各変量の重要度尺度を用いて変量を選択することができる。実際、Wilber et al. (2002) は各変数毎に有意性検定を行い、有意な変数を選択する方法を提案している。

ここでは、データの確率構造に基づく変数選択法を提案する。pq 個のパラメータ θ_{ij} を行列 $\Theta = (\theta_{ij})$ を用いて表す。このとき、q 個のグループの比較に関するモデルは、一般に $p \times (2^q - 1)$ 個考えられる。ここでは、判別解析において、実際的かつ実行可能なモデル族を構成し、これらの中から適切なモデルを選ぶことによって有効な変量な組を決定する方法を提案する。このため、まず、次のような変量の順序付けを考える。ここに、もとのデータについての群間平方和積和行列を $S_B = (s_B(i,j))$、群内平方和積和行列を $S_W = (s_W(i,j))$、統計量 $D_k^2 = s_B(k,k)/s_W(k,k)$ とするとき、

1. 統計量 $D_k^2, (k = 1, \ldots, p)$ の値が大きい方から k 個の変量を選び、これらの値が大きい順に $x_{(1)}, \ldots, x_{(k)}$ であるとする。

2. 残りの $p-k$ 個の変量 $x_j, (j \neq (1), \ldots, (k))$ に関して、$\theta_{1j} = \ldots = \theta_{qj} = \theta_j$ としたときの最尤推定量 $\hat{\theta}_j$ を求め、これらの値の大きい順が $x_{(k+1)}, \ldots, x_{(p)}$ であるとする。

このような順序付けに対応して、判別解析のための変数選択モデル $M_{k\ell}, 1 \leq k \leq \ell \leq p$ は、記号簡単のため $(j) = j,\ j = 1, \ldots, p$ とすると、次のように表せる。

$$M_{k\ell};\ \Theta = \begin{pmatrix} \theta_{11} & \cdots & \theta_{1p} \\ & \cdot & \\ & \cdot & \\ \theta_{q1} & \cdots & \theta_{qp} \end{pmatrix} = \left(\begin{array}{c|ccc|ccc} & \theta_{k+1} & \cdots & \theta_l & \theta & \cdots & \theta \\ \theta_{ij} & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ & \theta_{k+1} & \cdots & \theta_l & \theta & \cdots & \theta \end{array} \right).$$

このようなモデル族 $\{M_{k\ell};\ 1 \leq k \leq \ell \leq p\}$ の中から、適切なモデルを選ぶ基準として AIC 基準を考える。モデル $M_{k\ell}$ に対す AIC 基準は

$$\begin{aligned} AIC(M_{k\ell}) \ = \ & -2\sum_{i=1}^{q}\sum_{j=1}^{k}\left\{y_{ij}\log\hat{\theta}_{ij} + (n_i - y_{ij})\log(1-\hat{\theta}_{ij})\right\} \\ & -2\sum_{j=k+1}^{\ell}\left\{y_{\cdot j}\log\hat{\theta}_j + (n - y_{\cdot j})\log(1-\hat{\theta}_j)\right\} \\ & -2\left\{\left(\sum_{j=\ell+1}^{p} y_{\cdot j}\right)\log\hat{\theta} + \left(n(p-\ell) - \sum_{j=\ell+1}^{p} y_{\cdot j}\right)\log(1-\hat{\theta})\right\} \\ & -2\sum_{i=1}^{q}\sum_{j=1}^{p}\log\begin{pmatrix} n_{ij} \\ y_{ij} \end{pmatrix} + 2\{qk + \ell - k + 1\} \end{aligned}$$

で与えられる。ここに、

$$\hat{\theta}_{ij} = \frac{n_{ij}}{n_i}, \quad \hat{\theta}_j = \frac{y_{\cdot j}}{n}, \qquad y_{\cdot j} = \sum_{i=1}^{q} y_{ij}, \quad \hat{\theta} = \frac{\sum_{j=\ell+1}^{p} y_{\cdot j}}{n(p-\ell)}.$$

AIC の修正基準 CAIC も同様に求められる。

Wilber et al. (2002) による変数選択法と本報告で提案した変数選択法を比べるための一方法として、実際の誤判別確率を算出することが考えられる。これに関連して、DNA フィンガープリントデータに対して、2 つの判別法　1. 正準判別関数による判別　2. 最大尤度法　を適用した場合の結果について報告を行った。

その結果として、基準量 D_k^2 を利用した変数選択モデルを導入し,AIC , CAIC 基準を適用し,さらに誤判別確率を総合して変数を選ぶことにより,より良い判別が得られた。また、正準判別法と最大尤度法での誤判別確率の違いは扱った観測値に 0 が多いため,後者の精度が悪くなったと考えられる。

さらに、Wilber et al. (2002) によると、選択された変数が増えると変数間の独立性の仮定が崩れていくことが述べられている。そして、今回我々が選択した変数の組においても、同様なことが期待される。また、独立性の検定・独立性の仮定をはずした時の変数選択法のアルゴリズムを紹介したが、その数値計算はまだである。その数値計算・変数選択・他モデルでの検証がこれからの課題である。

多変量正規母集団の平均ベクトルの多重比較法について

九州東海大学工学部　今田恒久

1　はじめに

複数個の正規母平均ベクトルに対する多重比較を考える．従来，多変量多重比較法では用いる検定統計量の正確な分布が求められない場合が多く，指定した有意水準を満たす棄却限界値を求めるためにBonferroniの不等式や，漸近分布が利用された．この発表では検定統計量の正確な分布を用いた多重比較法をいくつか提案する．最初に指定した1つの母平均ベクトルと他のそれぞれの母平均ベクトルを比較する多重比較法をDunnett (1955)のシングルステップ法に基づいて構築する．次に互いに独立に多変量正規分布に従う確率ベクトルの列において母平均の変化点を見つけるための多重比較を考える．従来，母平均ベクトルの変化点を見つけるための検定では尤度比検定法より導出された検定統計量の正確な分布は導出されておらず，Srivastava-Worsley (1986)は指定した有意水準を満たす棄却限界値を求めるために改良型Bonferroniの不等式を用いている．ここでは変化点を見つける検定を閉検定手順に基づくステップダウン法に基づいて構築する．

2　Dunnettの方法に基づく多変量多重比較法

いま，$K(\geq 3)$個のp変量確率ベクトル$\boldsymbol{X}_k$ $(k=1,2,\cdots,K)$は互いに独立にそれぞれp変量正規分布$N(\boldsymbol{\mu}_k,\boldsymbol{\Sigma})$に従うと仮定する．いま，各母集団$N(\boldsymbol{\mu}_k,\boldsymbol{\Sigma})$より$n$個の互いに独立な標本ベクトル$\boldsymbol{X}_{k1},\boldsymbol{X}_{k2},\cdots,\boldsymbol{X}_{kn}$が得られたとし，その平均を$\bar{\boldsymbol{X}}_k$として$\boldsymbol{Z}_k=(Z_{k1},Z_{k2},\cdots,Z_{kp})'=\sqrt{n}\bar{\boldsymbol{X}}_k$，$\boldsymbol{S}_k=\sum_{i=1}^{n}(\boldsymbol{X}_{ki}-\bar{\boldsymbol{X}}_k)(\boldsymbol{X}_{ki}-\bar{\boldsymbol{X}}_k)'$とおく．さらに$\boldsymbol{S}=\sum_{k=1}^{K}\boldsymbol{S}_k$とおく．このとき，$\boldsymbol{\mu}_1$と$\boldsymbol{\mu}_2,\cdots,\boldsymbol{\mu}_K$それぞれを比較する多重比較法をDunnett (1955)のシングルステップ法に基づいて構築する．

2.1　母平均の差についての多重比較

仮説H_{1k} $(k=2,3,\cdots,K)$を$H_{1k}:\boldsymbol{\mu}_1=\boldsymbol{\mu}_k$，その対立仮説を$H_{1k}^A:\boldsymbol{\mu}_1\neq\boldsymbol{\mu}_k$とする．$H_{12},H_{13},\cdots,H_{1K}$の同時検定を考える．仮説$H_{1k}$の検定のため統計量$Y_{1k}$を

$$Y_{1k}=\begin{cases}(\boldsymbol{Z}_k-\boldsymbol{Z}_1)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_k-\boldsymbol{Z}_1) & \boldsymbol{\Sigma}\text{が既知のとき}\\(\boldsymbol{Z}_k-\boldsymbol{Z}_1)'\boldsymbol{S}^{-1}(\boldsymbol{Z}_k-\boldsymbol{Z}_1) & \boldsymbol{\Sigma}\text{が未知のとき}\end{cases}$$

と定義する．このとき，限界値cに対し，$Y_{1k}<c$となるならば，H_{1k}を保留し，そうでないならば棄却する．

2.2　多変量片側検定についての多重比較

$\boldsymbol{\mu}_1=(\mu_{11},\mu_{12},\cdots,\mu_{1p})'$と他の$\boldsymbol{\mu}_k=(\mu_{k1},\mu_{k2},\cdots,\mu_{kp})'$ $(k=2,\cdots,K)$に対し，$\mu_{11}\geq\mu_{k1}$，$\mu_{12}\geq\mu_{k2}$，$\cdots$，$\mu_{1p}\geq\mu_{kp}$となると仮定する．いま，仮説H_{1k}を$H_{1k}:\mu_{11}\geq\mu_{k1}$，$\mu_{12}\geq\mu_{k2}$，$\cdots$，$\mu_{1p}\geq\mu_{kp}$のうち少なくとも1つは等式となる．と定義する．このとき，H_{1k}の対立仮説は$H_{1k}^A:\mu_{11}>\mu_{k1}$，$\mu_{12}>\mu_{k2}$，$\cdots$，$\mu_{1p}>\mu_{kp}$．いま，$H_{12},H_{13},\cdots,H_{1K}$の同時検定を$\boldsymbol{\Sigma}$が既知と仮定して考える．$\boldsymbol{\Sigma}=(\sigma_{ij})$とするとき，仮説$H_{1k}$の検定のためSasabuchi (1980)の尤度比検定法に基づき，統計量Y_{1k}を

$$Y_{1k}=\min\left\{(Z_{11}-Z_{k1})/\sqrt{\sigma_{11}},(Z_{12}-Z_{k2})/\sqrt{\sigma_{22}},\cdots,(Z_{1p}-Z_{kp})/\sqrt{\sigma_{pp}}\right\}$$

と定義する．そして限界値cに対し，$Y_{1k} < c$ となるならば，帰無仮説H_{1k}を保留し，そうでないならば棄却することにする．

2.3 限界値の決定

限界値cを指定した有意水準αを満たすように求める．いずれのH_{1k}も棄却されない確率は$P(Y_{1k} < c,\ k = 2, \cdots, K)$．限界値$c$はすべての$H_{1k}$が同時に成り立つと仮定したときこの確率が$1-\alpha$に等しくなるように定める．広津 (1992) を参考に限界値cを決定する．

3 母平均ベクトルの変化点を見つけるための多重比較

$K(\geq 3)$個の互いに独立なp変量正規分布$N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$に対し，

帰無仮説$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_K$

対立仮説H_1 :少なくとも1組の$i, j (i \neq j)$に対し，$\boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$

とした検定において，帰無仮説が棄却される場合，$\boldsymbol{\mu}_i \neq \boldsymbol{\mu}_{i+1}$ となる最小のiを見つける方法をステップダウン法による多重比較法を用いて構築する．そのため，各$2 \leq k \leq K$に対し，仮説H_{0k}を$H_{0k} : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_k$，対立仮説を

$H_{1k} : 1, 2, \cdots, k$の中の少なくとも1組の$i, j (i \neq j)$に対し，$\boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$

とする．この仮説に対し，$\boldsymbol{\Sigma}$が既知，未知のそれぞれの場合において尤度比検定法に基づいて検定統計量T_kを決定し，H_{0k}の下で$P(T_k > c_k) = \alpha$となるような限界値$c_k (> 0)$を決定する．このとき，

Step 1. $T_K \leq c_K$ならばH_{0K}の保留で，検定を終了し，そうでないときH_{0K}を棄却し，次の段階に進む．

Step 2. $T_{K-1} \leq c_{K-1}$ならばH_{0K-1}の保留で，検定を終了し，そうでないときH_{0K-1}を棄却し，次の段階に進む．

$\vdots$

Step $\boldsymbol{K-1}$. $T_2 \leq c_2$ならば，H_{02}を保留で検定を終了し，そうでないときH_{02}を棄却し，検定を終了する．

H_{0K}の保留で検定が終了するとき，変化点はないと判断する．$k < K$であるときH_{0k}の保留で検定が終了すれば，$\boldsymbol{\mu}_i \neq \boldsymbol{\mu}_{i+1}$ となる最小のiは$i = k$と判断する．H_{02}の棄却で検定が終了すれば，$\boldsymbol{\mu}_i \neq \boldsymbol{\mu}_{i+1}$ となる最小のiは$i = 1$と判断する．この多重比較法では最大タイプ I FWE (familywise error rate) がαとなる．

参考文献

[1] Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association 50, 1096-1121.

[2] 広津千尋 (1992). 実験データの解析．共立出版株式会社．

[3] Sasabuchi, S. (1980). A test of a multivariate normal mean with composite hypotheses determined by linear Inequalities. Biometrika 67, 429-439.

[4] Srivastava, M. S. and Worsley, K. J. (1986). Likelihood ratio tests for a change in the multivariate normal mean, Journal of the American Statistical Association 81, 199-204.

２つのポアソン母数に順序制約がある場合の線形関数の推定

目白大・人文　　張　元宗
慶応大・理工　篠崎　信雄

1．はじめに

X_1, X_2を互いに独立にポアソン分布$P_o(\lambda_i), i=1,2,$にしたがう確率変数とし、母数λ_iに順序制約$\lambda_1 \leq \lambda_2$があるとき、その条件を考慮した最尤推定量（ＭＬＥ）は

$$\hat{\lambda}_i(X_1,X_2) = X_i + (-1)^i \frac{(X_1 - X_2)^+}{2}, \qquad i=1,2,$$

である。ここで、$a^+ = \max(0,a)$であり、X_iはλ_iの不偏推定量（ＵＢ）である。その他に、制約条件を満たすつぎのような$\min-\max$型推定量も考えられる。

$$\tilde{\lambda}_1(X_1,X_2) = \min\{X_1,X_2\}, \qquad \tilde{\lambda}_2(X_1,X_2) = \max\{X_1,X_2\}.$$

平均二乗誤差（ＭＳＥ）を基準としたとき、Kushary、Cohen(1991)はつぎのような結果を与えた。
1) λ_1を推定するとき、$\hat{\lambda}_1(X_1,X_2)$または$\tilde{\lambda}_1(X_1,X_2)$は$X_1$より優れている。
2) λ_2を推定するとき、X_1が観測されていない場合にX_2のみに基づく推定量$\delta(X_2)$がλ_2の許容的な推定量ならば、X_1が観測されていても、$\delta(X_2)$は依然、λ_2の許容的な推定量である。

ここでは、２つの母数の線形関数を推定するとき、最尤推定量と不偏推定量との比較を考え、最尤推定量が不偏推定量より優れているための線形関数の係数に関する必要十分条件を与える。同様に、$\min-\max$型推定量が不偏推定量より優れているための線形関数の係数に関する必要十分条件を与える。また、$\lambda_i, i=1,2$に順序制約がある場合、$\lambda_i, i=1,2$の同時推定問題を考え、最尤推定量より優れている推定量を与える。

2．結果

$X_i \sim P_o(\lambda_i), i=1,2$とし、$\lambda_i, i=1,2$に順序制約$\lambda_1 \leq \lambda_2$があるとする。平均２乗誤差（ＭＳＥ）を基準に、$c_1\lambda_1 + c_2\lambda_2$の推定問題において、最尤推定量または$\min-\max$型推定量と不偏推定量との比較を考えるとき、以下の定理が得られる。

リスクの表現を得るためには、次のLemmaが有用である。

Lemma 1. $X \sim P_0(\lambda)$とするとき

$$E[Xg(X)] = \lambda E[g(X+1)]$$

が成立する。

定理1. すべての$\lambda_1 \leq \lambda_2$に対して、最尤推定量$c_1\hat{\lambda}_1 + c_2\hat{\lambda}_2$が不偏推定量$c_1X_1 + c_2X_2$より一様に優れているための必要十分条件は

$$c_2 \leq 0,\ 3c_1 + c_2 \geq 0 \quad \text{または} \quad c_2 \geq 0,\ 3c_1 + c_2 \leq 0$$

であることである。

定理2. すべての$\lambda_1 \leq \lambda_2$に対して、$\min - \max$型推定量$c_1\tilde{\lambda}_1 + c_2\tilde{\lambda}_2$が不偏推定量$c_1X_1 + c_2X_2$より一様に優れているための必要十分条件は

$$c_2 \leq 0,\ c_1 + c_2 \geq 0 \quad \text{または} \quad c_2 \geq 0,\ c_1 + c_2 \leq 0$$

であることである。

一方、$\lambda_i, i = 1, 2,$の同時推定問題において、損失関数を

$$\frac{(\delta_1 - \lambda_1)^2}{\lambda_1} + \frac{(\delta_2 - \lambda_2)^2}{\lambda_2}$$

と選んだとき、つぎの結果が得られる。

定理3. $\lambda_i, i = 1, 2,$の推定量

$$\hat{\lambda}_i^{CZ}(X_1, X_2) = X_i + (-1)^i \frac{(X_1 - X_2)^+}{2} - \frac{X_i \varphi(X_1 + X_2)}{X_1 + X_2 + 1}, \qquad i = 1, 2,$$

は、最尤推定量$\hat{\lambda}_i,\ i = 1, 2,$より一様に優れている。ここで、$\varphi : [0, \infty) \to [0, 1]$の非減少関数である。

参考文献

(1)Clevenson, M. and Zidek, J. (1975). Simultaneous estimation of the means of independent Poisson Laws. Jour. Amer. Statist. Assoc., Vol. 70, 698-705.
(2)Kushary, D. and Cohen, A. (1991). Estimation of ordered Poisson parameters. *Sankhya*, Vol. 53, Series A, 334-356.
(3)Shinozaki, N. and Chang, Y.-T. (1999). A comparison of maximum likelihood and best unbiased estimators in the estimation of linear combinations of positive normal means. *Statistics & Decisions*, 17, 125-136.
(4)Chang, Y.-T. and Shinozaki, N. (2002). A comparison of restricted and unrestricted estimators in estimating linear functions of order scale parameters of two gamma distributions. Ann. Inst. Statist. Math., Vol.54, No.4, 848-860.
(5) 張元宗、篠崎信雄 (2000) 順序制約がある２つのガンマ分布の尺度母数の線形関数の推定－形状母数が異なる場合－。第68回日本統計学会予稿集
(6) 張元宗、篠崎信雄 (2002) エントロピー損失の下での順序制約がある２つのガンマ分布の尺度母数の線形関数の推定。第70回日本統計学会予稿集

推定量の許容性の証明方法について

関東学院大学 経済学部　布能 英一郎

1. 推定量の許容性の証明方法

Bayes 推定量が、与えられた事前分布より一意に定まれば、許容的である。しかし、この方法によって示されない許容的推定量が数多く存在する。そこで、こうした推定量の許容性の証明方法として、次のようなものが考えられてきた。

- Cramer-Rao の不等式を用いた方法
- **Blyth's methods** すなわち、プロパーなベイズ推定量の極限操作を行う方法
- 一般化ベイズ法 (improper prior を用いる方法)
- Stepwize Bayes 法

2. SW 法の有用性

Stepwize Bayes 法 (以後、SB 法と略記) とは、次のようなものである： 標本空間、母数空間を直和分解し、直和分解した各標本部分空間、および母数部分空間上で **proper prior** を選ぶ。更に各部分空間上の **proper prior** は、互いに直交するように選ぶ。そうすると、各部分空間上でのベイズ推定量は、各部分空間上で許容的であるばかりか、全体、すなわち、元の標本空間、母数空間上でも許容的.

しかしながら、SB 法は標本空間 $\boldsymbol{\mathcal{X}}$ も母数空間 Θ も有限な場合から出発し、**Brown**(1981) によって $\boldsymbol{\mathcal{X}}$ が有限、Θ がコンパクトの場合に完備性定理が成り立つことまで示されたものの、そこから先の有用性については足踏み状態であった。

他方、一般化ベイズ法を用いた推定量の許容性証明法は次のように記述される

Theorem　δ が improper prior $d\pi(\theta)$ に対する一般化ベイズ推定量であるとき、一般化ベイズリスク $\gamma(\delta, d\pi) = E_{d\pi}E_{\theta}(L(\theta, \delta(X))$ の値が有限ならば、δ は許容的。

Cor.　$\boldsymbol{\mathcal{X}}$ が有限ならば、improper prior $d\tau(\theta)$ に対する一般化ベイズ推定量 $\delta(x)$ の一般化ベイズ $\gamma(d\pi, \delta)$ は有限。

上記 Cor によれば、$\boldsymbol{\mathcal{X}}$ の場合、SB 法を使わなくても Theorem によって推定量の許容性が示されることがわかる。では許容性を示せず、SB 法で許容性を示せるような例は、果

たしてあるのだろうか? Cor により、このような例があるのなら、$\mathcal{X}$ が可算無限以上であることが必要がある。このような視点から研究を進めたところ、自乗損失下における負の二項分布の期待値 $k(1-\theta)/\theta$ の推定量 $\delta(x)=x$ が、このような example となることがわかった。

Property 負の二項分布にて、$E_\theta(X)=k(1-\theta)/\theta$ の推定量 $\delta(x)=x$ の許容性は、a sequence of priors を

$$d\tau_1(\theta)=dI_{\{\theta=1\}}(\theta), \qquad d\tau_2(\theta)\propto\frac{restriction}{1-\theta}d\theta$$

に選ぶという SB 法で示すことができる。他方、$\delta(x)=x$ は、improper prior $d\pi(\theta)=d\theta/(1-\theta)$ に対する (一般化) ベイズ解であるが、Theorem を用いて許容性を示すことはできない。なぜなら

$$\gamma(d\pi,\delta)=\int R(\theta,\delta)d\pi(\theta)=\int_0^1\frac{k(1-\theta)}{\theta^2}\frac{d\theta}{1-\theta}=k\int_0^1\frac{d\theta}{\theta^2}=+\infty$$

4. Karlin の証明方法

Karlin(1958) は、指数形分布族の期待値パラメーターが、自乗損失の下で許容的であることを示した。ところが、Karlin の証明方法に SB 法のアイデアを入れることで、証明がシンプルになる場合があることが示された。

参考文献

[1] Berger, J.O. (1985). *Statistical decision theory and Bayesian analysis* (Second edition), Springer-Verlag.

[2] Brown, L.D. (1981). A complete class theorem for statistical problems with finite sample space, *Annals of Statistics*, **9**, 1289-1300.

[3] Hsuan, F. (1979). A stepwise Bayesian procedure, *Annals of Statistics*, **7**, 860-868.

[4] Karlin, S. (1958). Admissibility for estimation with quadratic loss, *Ann. Math. Statist.* **29**, 406-436.

[5] Meeden, G. and Ghosh, M. (1981). Admissibility in finite problems, *Annals of Statistics*, **9**, 846-852.

[6] Meeden, G. and Ghosh, M. (1997). *Bayesian Methods for Finite Population Sampling*, Chapman and Hall.

[7] Mikulski, P.W. and Smith, P.J. (1976). A variance bound for unbiased estimation in inverse sampling, *Biometrika*, **63**, 216-217.

[8] Zacks, S. (1971). *The Theory of Statistical Inference*, Wiley.

Exact or Whittle's approximate likelihood estimation for Gaussian time series regression models and its plug-in approach

北海道大学大学院経済学研究科　　柿沢　佳秀

1. はじめに

統計モデルに含まれる母数 θ (ここでは 1 次元とする) を点推定する M-推定法 (ある規準関数の最大化) がしばしば議論されてきた. IID 設定ならば典型的には $M_n(\widehat{\theta}_n) = \sup_{\theta\in\Theta} M_n(\theta)$, $M_n(\theta) \equiv n^{-1}\sum_{i=1}^n g(X_i;\theta)$ なる M-推定量 $\widehat{\theta}_n$ の性質を調べることである. 解の大域的な存在性, 及び, (弱) 一致性を仮定すれば, 局所的には一階条件 $M_n'(\widehat{\theta}_n) = 0$ に帰着され, $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ の漸近正規性は多くの場合にスコア $\sqrt{n}M_n'(\theta_0)$ の CLT, 及び, ヘシアン $M_n''(\theta_0)$ の LLN と stochastic equicontinuity から示される.

母数ベクトルの場合でも同様の議論を行うが, 要素毎に異なる normalizing を必要とするかもしれない. 実際, 平均 0 の定常時系列モデルは $n^{1/2}$-漸近理論であるが, 時系列回帰モデルの場合デザイン行列によっては normalizing matrix を導入しなければいけない. 特に, 長期記憶過程の回帰モデルならば normalizing matrix はデザイン行列の型に依存することが知られている.

2. 時系列回帰モデル

$\{U_t\}$ は平均 0 の正規定常過程で, そのスペクトル密度が有限次元な母数ベクトル $\theta = (\theta^1,\ldots,\theta^p)' \in \Theta \subset \mathrm{R}^p$ で規定されているとする. 時系列回帰モデルは $Y_t = x_{t1}\beta^1 + x_{t2}\beta^2 + \cdots + x_{tr}\beta^r + U_t = \mathrm{x}_t'\beta + U_t$, $\beta = (\beta^1,\ldots,\beta^r)' \in \mathrm{R}^r$ で定義される. このような n 観測 $\mathbf{Y} \sim N_n(\ \mathbf{X}\beta, T(f_\theta)\ )$ の対数尤度関数は

$$L(\theta,\beta) = -\frac{1}{2}\log\det T(f_\theta) - \frac{1}{2}(\mathbf{Y}-\mathbf{X}\beta)'T(f_\theta)^{-1}(\mathbf{Y}-\mathbf{X}\beta)\,.$$

また, $n\times n$ Toeplitz 行列 $T(f_\theta)$ の行列式と逆行列を避けた Whittle's approximate (quasi) log-likelihood は

$$\begin{aligned} W(\theta,\beta) &= -\frac{n}{4\pi}\int_{-\pi}^{\pi}\left\{\log f_\theta(\lambda) + \frac{1}{2\pi n}\Big|\sum_{t=1}^{n}(Y_t - \mathrm{x}_t'\beta)\mathrm{e}^{it\lambda}\Big|^2\frac{1}{f_\theta(\lambda)}\right\}d\lambda \\ &= -\frac{n}{4\pi}\int_{-\pi}^{\pi}\log f_\theta(\lambda)\,d\lambda - \frac{1}{2}(\mathbf{Y}-\mathbf{X}\beta)'T((4\pi^2 f_\theta)^{-1})(\mathbf{Y}-\mathbf{X}\beta)\,. \end{aligned}$$

推定問題は $\eta_0 = (\theta_0', \beta_0')'$ の同時推定, あるいは, LSE $\widehat{\beta}_{LS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ を plug-in した場合の θ_0 の推定であるが, 計量経済の文脈では LSE を plug-in することが多い. その理由の 1 つは同時最適化では BLUE $\widehat{\beta}_{BLU} = [\mathbf{X}'T(f_\theta)^{-1}\mathbf{X}]^{-1}\mathbf{X}'T(f_\theta)^{-1}\mathbf{Y}$ が登場し, θ を決める際 profile likelihood を扱う必要があるからであろう. また Grenander (1954;AMS) と Adenstedt (1974;AS) を拡張した Yajima (1988,1991;AS) による LSE の BLUE に対する漸近相対効率の研究もあり, $\{U_t\}$ のスペクトル密度 f_θ が既知である場合, Grenander 型条件を満足するデザイン行列に対し LSE が漸近有効である

ための必要十分条件が議論された．したがって，性質のよいデザインに対して LSE を plug-in する理論的根拠が明らかにされたことになり，この意味において漸近有効になり得ないデザイン行列が与えられたとき LSE を plug-in することは躊躇せねばいけないように思えるけれども，長期記憶過程な定数回帰において Dahlhaus (1989;AS) は inefficent な算術平均を plug-in して得られるスペクトル母数の推定量が full MLE と漸近的に同等であるという結果を与え，さらに Giraitis&Koul (1997;SPA) は多項式回帰モデルへ拡張した．本報告は途中経過報告であるが，このような plug-in 推定量の性質を Yajima (1991;AS) の回帰モデルで考察する．

3&4. 短期記憶過程/長期記憶過程な回帰モデルにおける同時推定

$n \times r$ デザイン行列 $\mathbf{X}$ の第 j 列を $\mathbf{X}_j = (x_{1j}, x_{2j}, \ldots, x_{nj})'$ と書く．短期記憶過程では Grenander 条件

(G1) $||\mathbf{X}_j|| = (x_{1j}^2 + x_{2j}^2 + \cdots + x_{nj}^2)^{1/2} \to \infty$,

(G2) $x_{n+1,j}/||\mathbf{X}_j|| \to 0$,

(G3) $\sum_{t=1}^{n} x_{t+h,j} x_{tk}/(||\mathbf{X}_j|| ||\mathbf{X}_k||) \to \rho_{jk}(h)\ \forall h \in \mathrm{N}_0$

を仮定するが，長期記憶過程の場合はスペクトルの原点での発散のために Yajima 条件 (Y1)–(Y3) も課す．Toeplitz 行列とその逆行列の積の trace に関する近似公式 (e.g. Taniguchi (1991;Springer), Fox&Taqqu (1987;PTRF), Dahlhaus (1989;AS))，スペクトルノルム $SN(f,g) = ||T(f)^{-1/2}T(g)T(f)^{-1/2}||_{sp}$ の評価 (e.g. Dahlhaus (1989;AS)) と BLUE $\widehat{\beta}_{BLU} = [\mathbf{X}'T(f_\theta)^{-1}\mathbf{X}]^{-1}\mathbf{X}'T(f_\theta)^{-1}\mathbf{Y}$ の分散行列の逆行列 $\mathbf{X}'T(f_\theta)^{-1}\mathbf{X}$ の近似公式 (e.g. Grenander (1954;AMS), Yajima (1991;AS)) を用いて，スコアベクトルとヘシアン行列の性質を調べた．特に，推定量の漸近正規性の証明で鍵を握るヘシアン行列の stochastic equicontinuity について $||\mathbf{X}_j|| \to \infty$ の速さによっては不都合が生じること，及び，その対処を考察した．

5. Plug-in 推定量について

スペクトル母数の推定にのみ関心があるときは簡便な LSE $\widehat{\beta}_{LS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ を plug-in した目的関数 $L^*(\theta) = L(\theta, \widehat{\beta}_{LS})$ を最大化することが多い．3,4 節に与えた仮定に加えて，$(\mathbf{X}'\mathbf{X})^{-1}$ を保証するために $[\rho_{jk}(0)]$ が正則行列であるとし，長期記憶過程の場合はさらに Yajima 条件 (Y4) も課すとき，LSE $\widehat{\beta}_{LS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ の分散行列 $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'T(f_\theta)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ の近似公式 (e.g. Grenander (1954;AMS), Yajima (1991;AS)) も利用できる．Full MLE と漸近的に同等であることをみるため log-likelihood と plug-in log-likelihood について

$$
\begin{aligned}
&\partial_{j_1}\cdots\partial_{j_k}\{L(\theta,\widehat{\beta}_{LS}) - L(\theta,\beta)\} \\
&= \sum_{\kappa=1}^{r}(\widehat{\beta}_{LS}^{\kappa} - \beta^{\kappa})\mathbf{X}_\kappa'[\partial_{j_1}\cdots\partial_{j_k}T^{-1}(f_\theta)](\mathbf{Y} - \mathbf{X}\beta) \\
&\quad -\frac{1}{2}\sum_{\kappa,\mu=1}^{r}(\widehat{\beta}_{LS}^{\kappa} - \beta^{\kappa})(\widehat{\beta}_{LS}^{\mu} - \beta^{\mu})\mathbf{X}_\kappa'[\partial_{j_1}\cdots\partial_{j_k}T^{-1}(f_\theta)]\mathbf{X}_\mu
\end{aligned}
$$

が $o_p(n^{1/2})$ であることを示した．ここに $\partial_j = \partial/\partial\theta^j$.

LAN Theorem for Non-Gaussian Locally Stationary Processes and Its Applications

Junichi HIRUKAWA, Masanobu TANIGUCHI*

Department of Mathematical Sciences, School of Science and Engineering, Waseda University, Tokyo, 169-8555, Japan

Abstract

For a class of locally stationary processes introduced by Dahlhaus, we derive the LAN theorem under non-Gaussianity and apply the results to asymptotically optimal estimation and testing problems. For a class $\mathcal{F}$ of statistics which includes important statistics, we derive the asymptotic distributions of statistics in $\mathcal{F}$ under contiguous alternatives of unknown parameter. Because the asymptotics depend on the non-Gaussianity of the process, we discuss the non-Gaussian robustness. An interesting feature of effect of non-Gaussianity is elucidated in terms of LAN. Furthermore, the LAN theorem is applied to adaptive estimation when the innovation density is unknown.

keywords: Locally stationary process; Local asymptotic normality; Asymptotically efficient estimator; Optimal test; Non-Gaussian robustness; Adaptive estimation

1. Introduction.

Time series analysis under stationary assumption is well established. The assumption of stationarity guarantees that the increase of sample size provides more and more information of the same kind which is basic for an asymptotic theory to make sense. However it is not sufficient for stationary time series models to describe the real world. Many empirical studies show that real time series data are generally non-stationary. Therefore the assumption of non-stationarity is natural in time series analysis. When we deal with non-stationary processes, one of the difficult problems to solve is how to set up an adequate asymptotic theory. Recently, to overcome this problem, Dahlhaus (1996a, 1996b, 1997) proposed an important class of locally stationary processes and elucidated some fundamental results of the statistical inference.

In statistical asymptotic theory local asymptotic normality (LAN) (see e.g. LeCam (1986), Strasser (1985)) is one of the most fundamental concept and

*Corresponding author, +81-3-5286-8095

E-mail address: hirukawa@ruri.waseda.jp (J. Hirukawa), taniguchi@waseda.jp (M. Taniguchi).

describes the optimal solution of virtually all asymptotic inference and testing problems. The LAN approach has been introduced in time series analysis. Swensen (1985) established LAN for autoregressive models of finite order with a regression trend and applied it in the derivation of the local power of the Durbin-Watson test. Kreiss (1987, 1990) also proved the LAN property for ARMA and AR(∞) models, and constructed locally asymptotically minimax LAM adaptive estimators, as well as locally asymptotically maximin tests. For time series regression models with long memory disturbance, Hallin et al. (1999) showed LAN theorem and discussed an adaptive estimation. Taniguchi and Kakizawa (2000) gave an extensive review for LAN approach in time series analysis.

For non-stationary case, Dahlhaus (1996b) developed asymptotic theory for Gaussian locally stationary processes and derived the LAN property. In this paper we drop the Gaussian assumption, i.e., we prove the LAN theorem for non-Gaussian locally stationary processes, and apply the results to the asymptotic estimation and testing theory. Our LAN theorem elucidates the non-Gaussian asymptotics and the proof and method contain a lot of different parts from Gaussian case. We discuss a non-Gaussian robustness for a class of satatistics which have quadratic forms. The class includes the standardized of QMLE, tests and discriminant statistics etc. We are often interested in the local asymptotics under $\theta_T = \theta + \frac{h}{\sqrt{T}}$. In view of LeCam' third lemma, together with the LAN results we give the limit distributions of the quadratic forms under θ_T, which depend on non-Gaussian quantities. If they are independent of the non-Gaussian quantities, we say that they are non-Gaussian robust. Then we give a set of sufficient conditions for the non-Gaussian robustness, which illuminate new scope of non-Gaussian approach. Also some numerical studies are given, and show unexpected features. Finally, we apply the LAN theorem to construct adaptive estimators. When the innovation density is unknown, the asymptotically efficient estimator is given, as well as when the innovation density is known.

L-統計量の漸近分布

九州大学大学院経済学研究院　　前園宜彦

統計量のクラスとして有用な L-統計量の部分族について，ジャックナイフ分散推定量の漸近表現を求め，スチューデント化 L-統計量の漸近表現と $n^{-1/2}$ のオーダーまでのエッジワース展開を求めた．

$X_1, \cdots, X_n$ を i.i.d. $F(x)$ とし，$J(u)$ をスコア関数，$F_n(u)$ を経験分布関数，すなわち $F_n(u) = n^{-1}\sum_{i=1}^{n} I(X_i \le u)$ とする．ただし $I(\cdot)$ は定義関数である．このとき L-統計量の部分族

$$T(F_n) = \int_0^1 F_n^{-1}(u)J(u)du$$

について考察する．ここで $F_n^{-1}(u) = \inf\{x; F_n(x) \ge u\}$ である．この $T(F_n)$ は L-統計量の部分族 (see Serfling (1980)) を成す．ここで $T(F) = \int_0^1 F^{-1}(u)J(u)du$ とおくと，$\sqrt{n}(T(F_n) - T(F))$ の漸近分散は

$$\begin{aligned} &\sigma^2(J, F) \\ = &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J(F(u))J(F(v))[F(\min(u,v)) - F(u)F(v)]dudv \end{aligned}$$

となる．標準化 L-統計量のエッジワース展開は Helmers (1982), Alberink, Pap and van Zuijlen (2001) 等により議論されている．$\sqrt{n}(T(F_n) - T(F))$ のジャックナイフ分散推定量は

$$\hat{\sigma}^2(J, F) = (n-1)\sum_{i=1}^{n}[T(F_{n;i}) - T(F_n)]^2$$

で与えられる．ここで $F_{n;i}$ は X_i を除いた $n-1$ 個の標本に基づく経験分布関数である．この $\hat{\sigma}^2(J, F)$ の一致性 は Parr & Schucany (1982) や Shao (1991) により示されている．

本報告では $\hat{\sigma}^2(J, F)$ の漸近表現を求め，スチューデント化 L-統計量 $\sqrt{n}\{T(F_n) - T(F)\}/\hat{\sigma}(J, F)$ の $n^{-1/2}$ の項までのエッジワース展開を求めた．

[定理 1]　スコア関数が $J^{(1)}(u)$ は 有界で，$J^{(2)}(u)$ はオーダー $s > 0$ の Lipshitz 条件を満たすとする．もし

$$\int_{-\infty}^{\infty}\{F(u)(1 - F(u))\}^{1/4}du < \infty$$

ならば

$$\hat{\sigma}^2(J, F) = \sigma^2(J, F) + n^{-1}\sum_{i=1}^{n}\alpha_1(X_i) + o_L(n^{-1/2})$$

および

$$\frac{\sqrt{n}(T(F_n)-T(F))}{\hat{\sigma}(J,F)}$$
$$= \quad n^{-1/2}\tau + n^{-1/2}\sum_{i=1}^{n}\nu_1(X_i) + n^{-3/2}\sum_{C_{n,2}}\nu_2(X_i,X_j) + o_L(n^{-1/2})$$

が成り立つ．ここで $E[\nu_2(X_1,X_1)|X_1]=0$ $a.s.$ で，これらの表現は Hoeffding 分解（H-分解）されたものになっている．

Lai and Wang (1993) による漸近 U-統計量に対するエッジワース展開を使うと，スチューデント化 L-統計量のエッジワース展開を求めることができる．次の条件を考える．

(C_1) $E\{|\nu_1(X_1)|^3+|\nu_2(X_1,X_2)|^3\}<\infty$, $\sigma^2(J,F)>0$.

(C_2) $\limsup_{|t|\to\infty}|E[\exp\{it\nu_1(X_1)\}]|<1$.

[定理 2] （定理 1）の条件及び C_1, C_2 が成り立つとき

$$\sup_x\Big|P\Big\{\frac{\sqrt{n}\{T(F_n)-T(F)\}}{\hat{\sigma}(J,F)}\le x\Big\}-Q_n(x)\Big|=o(n^{-1/2})$$

が成り立つ．ここで

$$\begin{aligned}\kappa_3 &= E[\nu_1^3(X_1)]+3E[\nu_1(X_1)\nu_1(X_2)\nu_2(X_1,X_2)],\\ P_1(x) &= \frac{\kappa_3(x^2-1)}{6},\\ Q_n(x) &= \Phi(x)-n^{-1/2}\phi(x)\{P_1(x)+\tau\}\end{aligned}$$

で，τ は（定理 1）で求めた H-分解された項を使って定義されるものである．

これらの成果は順序統計量の線形和からなる基準統計量について広く適用できるものである．

[参考文献]

Alberink, I.V., Pap, G. and van Zuijlen, M.C.A. (2001). Edgworth expansions for L-statistics. *Prob. Math. Statist.* **21**, 277-302.

Helmers, R. (1982). Edgeworth expansions for linear combinations of order statistics. *Math. Centre Tracts 105*, Amsterdam.

Lai, T.L. and Wang, J.Q. (1993). Edgeworth expansion for symmetric statistics with applications to bootstrap methods. *Statistica Sinica* **3**, 517-542.

Parr, W.C. and Schucany, W.R. (1982). Jackknifing L-statistics with smooth weight functions. *Jour. Amer. Statist. Assoc.* **77**, 629-638.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Shao, J. (1991). Jackknife variance estimators for generalized L-statistics. *Statist. Prob. Lett.* **11**, 27-32.

Link function elicitation through grobal orthogonality of parameters in GLM

統数研・領域系　大西俊郎
統数研・領域系　柳本武美

1. 問題設定

一般化線形モデルにおけるリンク関数の選択について，モデルビルディングの観点から議論する．確率変数ベクトル $\boldsymbol{x} = (x_1, \ldots, x_n)^T$ の各成分が指数型分布族

$$p(x_i;\, \mu_i) = \exp\bigl\{c(\mu_i)\, x_i - M(\mu_i)\bigr\}\, a(x_i)$$

に従っているとし，平均ベクトル $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$ がリンク関数 $f(\mu)$ によって共変量行列 $Z = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)$ と関係づけられているとする．

$$f(\boldsymbol{\mu}) = \alpha \mathbf{1}_n + Z^T \boldsymbol{\beta}.$$

適当な条件を満たす関数 $g(\mu)$ を用いて新しいパラメータ γ を導入する．

$$\gamma = \sum_{i=1}^{n} g(\mu_i).$$

スロープ $\boldsymbol{\beta}$ が μ_i の $\boldsymbol{z}_i$ に対する依存性を表しているのに対し、γ は $(\mu_1, \ldots, \mu_n)$ の全体的な傾向を示す．モデルを $(\boldsymbol{\beta}, \gamma)$ でパラメトライズし，同時密度関数 $\prod p(x_i;\, \mu_i)$ を $p(\boldsymbol{x};\, \boldsymbol{\beta}, \gamma)$ と表示する．次の 2 つのケースに場合分けして考察する．

(i) $\boldsymbol{\beta}$ が興味あるパラメータであり，γ には関心がない，

(ii) γ が興味あるパラメータであり，$\boldsymbol{\beta}$ には関心がない．

本発表の第 1 の目的は，上記の 2 つのケースそれぞれに適した関数の組 $\bigl(f(\mu), g(\mu)\bigr)$ を導出することである．第 2 の目的は，導かれたリンク関数の下で成立する性質を調べることである．

2. スコア関数の強不偏性

リンク関数の選択に際して指針となる要請を定義として与える．それは関心のないパラメータのミススペシフィケーションに対するある種の頑健性である．密度関数 $p(\boldsymbol{x};\, \boldsymbol{\beta}, \gamma)$ に対応するスコア関数を $l_{\boldsymbol{\beta}}(\boldsymbol{x};\, \boldsymbol{\beta}, \gamma)$ および $l_\gamma(\boldsymbol{x};\, \boldsymbol{\beta}, \gamma)$ と書くことにする．

定義 2.1.　共変量行列 Z を任意に固定する．

(i) スコア関数 $l_{\boldsymbol{\beta}}(\boldsymbol{x};\, \boldsymbol{\beta}, \gamma)$ が強不偏であるとは，次の等式が成り立つことである．

$$\mathrm{E}\bigl[\, l_{\boldsymbol{\beta}}(\boldsymbol{x};\, \boldsymbol{\beta}, \gamma^*) \bigm| p(\boldsymbol{x};\, \boldsymbol{\beta}, \gamma) \bigr] = 0 \quad \text{for any } (\boldsymbol{\beta}, \gamma, \gamma^*). \tag{2.1}$$

(ii) スコア関数 $l_\gamma(\boldsymbol{x};\, \boldsymbol{\beta}, \gamma)$ が強不偏であるとは，次の等式が成り立つことである．

$$\mathrm{E}\bigl[\, l_\gamma(\boldsymbol{x};\, \boldsymbol{\beta}^*, \gamma) \bigm| p(\boldsymbol{x};\, \boldsymbol{\beta}, \gamma) \bigr] = 0 \quad \text{for any } (\boldsymbol{\beta}, \boldsymbol{\beta}^*, \gamma). \tag{2.2}$$

次の Fisher 直交条件は，要請 (2.1) の必要条件であり，かつ，要請 (2.2) の必要条件でもある．

$$\mathrm{E}\big[\, l_{\boldsymbol{\beta}\gamma}(\boldsymbol{x};\, \boldsymbol{\beta},\, \gamma) \;\big|\; p(\boldsymbol{x};\, \boldsymbol{\beta},\, \gamma) \big] = \mathbf{0} \quad \text{for any } (\boldsymbol{\beta}, \gamma).$$

Fisher 直交条件について次の定理が得られる．

定理 2.1. 任意の共変量行列 Z に対して $(\beta,\, \gamma)$ の Fisher 情報行列がブロック対角になるための必要十分条件は，次の等式が成り立つことである．

$$\frac{c'(\mu)}{f'(\mu)g'(\mu)} = Const.$$

3. 対数リンク、恒等リンクおよび正準リンクの特徴づけ

スコア関数 $l_{\boldsymbol{\beta}}(\boldsymbol{x};\, \boldsymbol{\beta},\, \gamma)$ の強不偏性からは対数リンクと恒等リンクが特徴づけられる．特に後者は前者の特殊ケースである．一方，スコア関数 $l_\gamma(\boldsymbol{x};\, \boldsymbol{\beta},\, \gamma)$ の強不偏性によって特徴づけられるのは正準リンクである．

定理 3.1. 任意の共変量行列 Z に対してスコア関数 $l_{\boldsymbol{\beta}}(\boldsymbol{x};\, \boldsymbol{\beta},\, \gamma)$ が強不偏であるのは

$$f(\mu) = \frac{a_1}{a_2}\log\left(1 + \frac{a_2}{a_1}\frac{\mu - a_0}{a_1}\right),$$
$$g(\mu) = \frac{a_2 b_1}{a_1}M(\mu) + \left(a_1 - \frac{a_0 a_2}{a_1}\right) b_1 c(\mu) + b_0,$$

のときに限る．ただし，a_0, a_1, a_2, b_0 および b_1 は定数であり，$a_1 b_1 \neq 0$ である．

定理 3.2. 任意の共変量行列 Z に対してスコア関数 $l_\gamma(\boldsymbol{x};\, \boldsymbol{\beta},\, \gamma)$ が強不偏であるのは

$$f(\mu) = \frac{c(\mu)}{a_1} + a_0,$$
$$g(\mu) = a_1 b_1 \mu + b_0,$$

のときに限る．ただし，a_0, a_1, b_0 および b_1 は定数であり，$a_1 b_1 \neq 0$ である．

4. 強不偏性が含意するもの

第 2 節で定義した強不偏性 (2.1) および (2.2) は，半大域的な直交性である．実際，対数リンクでは変形 Pythagoras 関係が成立し，恒等リンクと正準リンクでは Pythagoras 関係が成り立つ．

5. 例（ガンマ分布の対数リンクモデル）

ガンマ分布の対数リンクモデルにおけるスロープパラメータの推定問題を数値的に扱った．対数正規分布の対数リンクモデルとの比較を行い，ミニマックスの意味でガンマ分布の方がよいことを示した．

多変量線形校正問題における古典的推定量の改良について

統計数理研究所　津熊 久幸

本報告では, 多変量線形モデル

$$\underset{p\times1}{\boldsymbol{y}_i} = \underset{p\times1}{\boldsymbol{\alpha}} + \underset{p\times q}{\boldsymbol{\Theta}'}\,\underset{q\times1}{\boldsymbol{x}_i} + \underset{p\times1}{\epsilon_i}, \qquad i=1,\ldots,n, \tag{1}$$

$$\underset{p\times1}{\boldsymbol{y}_{0j}} = \underset{p\times1}{\boldsymbol{\alpha}} + \underset{p\times q}{\boldsymbol{\Theta}'}\,\underset{q\times1}{\boldsymbol{x}_0} + \underset{p\times1}{\boldsymbol{\epsilon}_{0j}}, \qquad j=1,\ldots,m, \tag{2}$$

における校正問題を扱った. ただし, $\boldsymbol{y}_i$, $\boldsymbol{y}_{0j}$, $\boldsymbol{x}_i$ は既知であり, $\boldsymbol{\alpha}$, $\boldsymbol{\Theta}$, $\boldsymbol{x}_0$ は未知である. また, ϵ_i と ϵ_{0j} はそれぞれ, 誤差ベクトルであり, 互いに独立に平均零ベクトル, 未知の共分散行列 $\boldsymbol{\Sigma}$ の多変量正規分布に従っているとする. ここで, $p\geq q$, $n+m-q-2\geq p$, $\sum_i \boldsymbol{x}_i = \boldsymbol{0}_{q\times1}$, 行列 $(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)$ はフルランクとする. 上記のモデルにおいて目的は $\{(\boldsymbol{y}_i, \boldsymbol{x}_i)\}_{i=1}^n$ と $\{\boldsymbol{y}_{0j}\}_{j=1}^m$ に基づいて $\boldsymbol{x}_0$ の推定することである.

この報告では 2 乗損失関数 $L_1(\tilde{\boldsymbol{x}}_0; \boldsymbol{x}_0) = (1/c_{n,m})(\tilde{\boldsymbol{x}}_0-\boldsymbol{x}_0)^t\boldsymbol{\Theta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Theta}^t(\tilde{\boldsymbol{x}}_0-\boldsymbol{x}_0)$ ($c_{n,m} = 1/n+1/m$, $\tilde{\boldsymbol{x}}_0$ は $\boldsymbol{x}_0$ の推定量) のもとで $\boldsymbol{x}_0$ の推定問題を扱い, 古典的推定量の改良を目標とした. しかし 2 乗損失関数のもとではリスク評価が困難であるため, まずその損失関数に似た擬似損失関数 $L(\tilde{\boldsymbol{x}}_0; \boldsymbol{x}_0) = (1/c_{n,m})(\widehat{\boldsymbol{\Theta}}^t\tilde{\boldsymbol{x}}_0 - \boldsymbol{\Theta}^t\boldsymbol{x}_0)^t\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Theta}}^t\tilde{\boldsymbol{x}}_0 - \boldsymbol{\Theta}^t\boldsymbol{x}_0)$ ($\widehat{\boldsymbol{\Theta}}$ は $\boldsymbol{\Theta}$ の最小 2 乗推定量)のもとで古典的推定量の代替推定量の候補を解析的に導出した. そして数値実験より 2 乗損失関数のもとでの古典的推定量と, その代替推定量のリスクの比較検討をおこなった. 以下, その詳細について報告する.

今, モデル (1), (2) に適当な変換を施すことにより, 次のような正準形を得る:

$$\boldsymbol{B}\sim\mathcal{N}_{q\times p}(\boldsymbol{\beta}, \boldsymbol{I}_q\otimes\boldsymbol{\Sigma}), \quad \boldsymbol{S}\sim\mathcal{W}_p(\boldsymbol{\Sigma}, l), \quad \boldsymbol{z}\sim\mathcal{N}_p(\boldsymbol{\beta}^t\boldsymbol{\xi}, \boldsymbol{\Sigma}). \tag{3}$$

ただし $l=n+m-q-2$ であり, $(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\xi})$ は未知で推定目標は $\boldsymbol{\xi}$ となる. このとき, 古典的推定量は

$$\hat{\boldsymbol{\xi}} = (\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z} \tag{4}$$

であり, 上記の 2 乗損失関数と擬似損失関数はそれぞれ

$$L_1(\tilde{\boldsymbol{\xi}}; \boldsymbol{\xi}) = (\tilde{\boldsymbol{\xi}}-\boldsymbol{\xi})^t\boldsymbol{\beta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}^t(\tilde{\boldsymbol{\xi}}-\boldsymbol{\xi}), \tag{5}$$

$$L(\tilde{\boldsymbol{\xi}}; \boldsymbol{\xi}) = (\boldsymbol{B}^t\tilde{\boldsymbol{\xi}}-\boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{B}^t\tilde{\boldsymbol{\xi}}-\boldsymbol{\beta}^t\boldsymbol{\xi}) \tag{6}$$

となる.

2 乗損失関数 (5) のもとでの古典的推定量 (4) の改良を考えるために, まず擬似損失関数 (6) のもとで代替推定量を探すことにし, (3) の $\boldsymbol{\xi}$ の推定量として次のような推定量の族を考えた:

$$\hat{\boldsymbol{\xi}}(\phi, \boldsymbol{G}) = \phi\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}. \tag{7}$$

ただし, ϕ は $\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z}$ の関数, $\boldsymbol{G}$ は $q\times q$ 対称行列で各成分は $\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t$ の関数とする. この推定量 (7) は古典的推定量 (4) の拡張である.

次に推定量 (7) のリスクの評価, いわゆるリスクの不偏推定量を求める. 部分積分などから擬似損失関数 (6) に関する推定量 (7) のリスクは次のようになる:

定理 1 $\boldsymbol{F}=\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t$, $t=\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z}$, $D_F=((1/2)(1+\delta_{ij})\partial/\partial F_{ij})$, $\phi'=d\phi/dt$ とする. 擬似損失関数 (6) に関する推定量 $\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G})$ のリスクは次のように書ける:

$$
\begin{aligned}
R(\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}),\boldsymbol{\xi}) &= E\left[L(\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G});\boldsymbol{\xi})\right]\\
&= E\Big[-p+4\phi'\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}+2\phi\,\mathrm{tr}\,(\boldsymbol{F}\boldsymbol{G})\\
&\qquad +(l-p-1)\,\mathrm{tr}\,(\boldsymbol{S}^{-1}(\boldsymbol{z}-\phi\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z})(\boldsymbol{z}-\phi\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z})^t)\\
&\qquad +4\phi'\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}(\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z}-\phi\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z})\\
&\qquad +4\phi\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t(\boldsymbol{I}_q-\phi\boldsymbol{G}\boldsymbol{F})\{(\boldsymbol{F}D_F)^t\boldsymbol{G}\}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}\\
&\qquad +2\phi\{\,\mathrm{tr}\,(\boldsymbol{F}\boldsymbol{G})\}(\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z}-\phi\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z})\\
&\qquad +2\phi(\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}-\phi\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{F}\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z})\Big].
\end{aligned}
$$

ただし, '$\{(\boldsymbol{F}D_F)^t\boldsymbol{G}\}$' の表記は D_F が $\boldsymbol{G}$ にのみ作用することを表し, $\{(\boldsymbol{F}D_F)^t\boldsymbol{G}\}$ の (i,j) 成分は $\{(\boldsymbol{F}D_F)^t\boldsymbol{G}\}_{ij}=\sum_{a,b}F_{ab}\left[\{D_F\}_{ia}G_{bj}\right]$ である. □

擬似損失関数 (6) に関する古典的推定量 $\hat{\boldsymbol{\xi}}$ の代替推定量を具体的に得るために, 次のような縮小型推定量を考えた:

$$\hat{\boldsymbol{\xi}}(\psi)=(1-\psi/t)(\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}. \tag{8}$$

ただし, $\psi\equiv\psi(t)$, $t=\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z}$ である. 定理 1 より次の結果が得られた:

定理 2 $q\geq 3$ とする. ψ が非減少, $0\leq\psi\leq 2(q-2)/(l-p+3)$ であれば, 擬似損失関数 (6) のもとで縮小型推定量 (8) は古典的推定量 (4) を改良する. □

数値実験

定理 2 において, 擬似損失関数 (6) のもとで古典的推定量と縮小型推定量のリスクに関する解析的な結果が得られたが, 2 乗損失関数 (5) のもとで同様な結果を与えるのは困難であるため数値実験を通じて 2 乗損失関数 (5) のもとでのリスクの比較をおこなった.

実験の回数は 10,000 回とし, モデル (3) のもとで実験をおこなった. 実験の仮定として, $\psi=(q-2)/(l-p+3)$, $(n,m,p,q)=(30,20,7,5)$, $\boldsymbol{\beta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}^t$ は対角行列とした. また, $\boldsymbol{\xi}=(1,1,1,1,1)^t$ と $\boldsymbol{\xi}=(2,2,2,2,2)^t$ を仮定した. この数値実験より以下のことがわかった.

1. $\boldsymbol{\beta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}^t$ の対角成分が小さい場合, 縮小型推定量による古典的推定量の改良率は大きかった.

2. 逆に $\boldsymbol{\beta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}^t$ の対角成分が大きい場合, あるいは対角成分の最大値と最小値の比が大きいときにはその改良率は小さくなるが負になることはなかった.

限られたパラメータ設定ではあるが, 上記の実験結果から擬似損失関数 (6) のもとで解析的に得られた縮小型推定量 (8) は, 2 乗損失関数 (5) のもとでも古典的推定量 (4) を改良すると予想される. この実験結果の解析的な証明方法については今後の研究課題としたい.

On Fisher's information loss

Osaka University Etsuo Kumagai
Osaka Univeristy Nobuo Inagaki

In this paper, we propose a generalized hyperbolic model as an extension of the Nile problem by Fisher [2], which showed the exact information loss with a hyperbolic curve as a parameter in the two dimensional exponential distribution, and we investigate the exact information loss with respect to MLE in the generalized hyperbolic model. In this calculation, we use a projection matrix directly, which is partially based on the projection with respect to the conditional expectation given MLE in Inagaki [3]. Also we define Efron's information curvature as an extension of the information loss with Efron's statistical curvature [1].

Let k be a positive integer and let random variables $\{X_i\}_{i=1}^k$ be independent mutually and be each distributed with the gamma distribution $\{G_A(q_i, \alpha_i^{-1})\}_{i=1}^k$, where $\{q_i\}_{i=1}^k$ and $\{\alpha_i\}_{i=1}^k$ are constants and parameters respectively which are all positive and finite. We consider the k-dimensional gamma distribution $(X_1, \dots, X_k)$ whose probability density function is

$$\begin{aligned} p(x_1, \dots, x_k) &= \exp\{\langle \boldsymbol{\alpha}, \boldsymbol{x}\rangle - \psi(\boldsymbol{\alpha})\}\, h(\boldsymbol{x}) \\ &= \exp\left\{ -\sum_{i=1}^{k} \alpha_i\, x_i + \sum_{i=1}^{k} \log \alpha_i^{q_i} \right\} \prod_{i=1}^{k} \frac{x_i^{q_1 - 1}}{\Gamma(q_i)} . \end{aligned}$$

We consider the k-dimensional gamma distribution whose parameters satisfy the condition $\psi(\boldsymbol{\alpha}(\boldsymbol{\theta})) = 0$, that is, parameters have the relationships as follows:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \begin{pmatrix} \alpha_1(\boldsymbol{\theta}) \\ \vdots \\ \alpha_{k-1}(\boldsymbol{\theta}) \\ \alpha_k(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_{k-1} \\ \vartheta^{-1} \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_{k-1} \end{pmatrix},$$

where

$$\vartheta = \prod_{i=1}^{k-1} \theta_i^{r_i}, \quad 0 < \theta_1, \dots, \theta_{k-1} < \infty, \quad r_1 = \frac{q_1}{q_k}, \ \dots, \ r_{k-1} = \frac{q_{k-1}}{q_k}.$$

Note that $r_1, \dots, r_{k-1}$ are positive constants. We call this *a generalized hyperbolic model.* The following theorem holds with respect to this generalized hyperbolic model:

Theorem 1 *In the k-dimensional generalized hyperbolic model, let $\boldsymbol{P}^*(\boldsymbol{\theta})$ be the projection matrix into the subspace spanned by $\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{\theta})\dot{\boldsymbol{\alpha}}(\boldsymbol{\theta})$, that is,*

$$\boldsymbol{P}^*(\boldsymbol{\theta}) \;=\; \boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{\theta})\dot{\boldsymbol{\alpha}}(\boldsymbol{\theta})\left({}^t\dot{\boldsymbol{\alpha}}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta})\dot{\boldsymbol{\alpha}}(\boldsymbol{\theta})\right)^{-1}{}^t\dot{\boldsymbol{\alpha}}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{\theta}),$$

so that the generalized statistical curvature $\boldsymbol{\Gamma}(\boldsymbol{\theta})^2$ is represented as follows:

$$\begin{aligned}
&\boldsymbol{\Gamma}(\boldsymbol{\theta})^2 \\
&= \left({}^t\dot{\boldsymbol{\alpha}}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta})\dot{\boldsymbol{\alpha}}(\boldsymbol{\theta})\right)^{-2}\left\{{}^t\ddot{\boldsymbol{\alpha}}(\boldsymbol{\theta})\left[\mathbf{I}_{k-1}\otimes\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{\theta})(\mathbf{I}_k-\boldsymbol{P}^*(\boldsymbol{\theta}))\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{\theta})\right]\ddot{\boldsymbol{\alpha}}(\boldsymbol{\theta})\right\} \\
&= \frac{1}{q_1+\cdots+q_k}\,\mathbf{I}_{k-1},
\end{aligned}$$

where $\mathbf{I}_{k-1}$ and $\mathbf{I}_k$ are the $k-1$ and k dimensional identity matrices, respectively, and the notation $\otimes$ means the Kronecker product. □

Theorem 2 *The exact Fisher information loss of k-dimensional hyperbolic model is represented by*

$$\boldsymbol{I}_n(\boldsymbol{\theta})-\boldsymbol{I}_{\boldsymbol{T}}(\boldsymbol{\theta}) \;=\; \frac{n}{n(q_1+\cdots+q_k)+1}\,\boldsymbol{I}(\boldsymbol{\theta})$$

and this converges to EIC as follows:

$$\lim_{n\to\infty}\left\{\boldsymbol{I}_n(\boldsymbol{\theta})-\boldsymbol{I}_{\boldsymbol{T}}(\boldsymbol{\theta})\right\} \;=\; \frac{1}{q_1+\cdots+q_k}\boldsymbol{I}(\boldsymbol{\theta}) \;=\; \mathrm{EIC}(\boldsymbol{\theta}).$$

□

References

[1] B.Efron, Defining the curvature of a statistical problem (with applications to second order efficiency), Ann. Statist. 6 (1975) 1189–1242.

[2] R.A.Fisher, *Statistical Methods and Scientific Inference*, Hafner Press, New York, 1973.

[3] N.Inagaki, The asymptotic distribution of a conditional likelihood ratio function given a maximum likelihood estimator in *Statistical Theory and Data Analysis (K.Matusita(Editor))*, North-Holland, Amsterdam, 1985.

[4] N.Inagaki and E.Kumagai, Exact information loss in Fisher's circle model, Mathematica Japonica, 44 (1996) 455–467.

[5] K.Inoue, N.Inagaki and E.Kumagai, The exact information loss of the maximum likelihood estimator in the k-dimensional sphere model, Mathematica Japonica, 52 (2000) 289–310.

[6] E.Kumagai and N.Inagaki, Exact information loss in the multivariate gamma distribution (to be submitted).

Automatic Recognition of Estrangelo through Invariants

Kyushu University T. Sakata.
Kyushu Institute of Design(Doctor course) S.C.Tan.

We discussed about automatic recognition of Estrangelo, an old Syriac language, through invariants. The motivation of our study is that such an automatic recognition engine might help to mutual understanding between peoples with different languages and cultures. Such classical languages are very important through many historically precious classical texts. Our purpose of this experiment was to reconsider the Cloksin and Fernando's result of the recognition of Estrangelo by using different features and a different classification technique. One major difference between the works of Cloksin et. al. and us is that they used word segmentation into strokes and we treated a word as a whole. In general, invariant character recognition methods are plausible and as the features we used Flusser moment invariants, Zernike moment invariants and some feature of a polar transformed image. Flusser invariant moments defined through complex moments

$$c_{pq} = \int (x+iy)^p (x-iy)^q dxdy$$

and given by

$$J = \prod c_{p_i q_i}^{k_i} \quad with \sum k_i(p_i - q_i) = 0.$$

On the other hand the Zernike moment invariants are defined through the Zernike moments which are calculated for images on the unit circle. The Zernike moment of order p is defined as

$$Z_{pq} = \frac{(p+1)}{\pi} \int_0^{2\pi} V_{pq}^*(r,\theta) f(r,\theta) r dr d\theta,$$

where the function $V_{pq}(r,\theta)$ is the Zernike polynomials of order p with repetition q and "$*$" denotes complex conjugate, and the real Zernike moment invariants are given by

$$Z_{pq}^* Z_{rs}^m + (Z_{pq}^* Z_{rs}^m)^*.$$

Character recognition rates by Zernike Moments

ALAP	BEITH	GAMAL	DALATH	HE	WAU
90.00	80.00	90.00	90.00	80.00	100.00
ZAIN	KHEITH	TEITH	YUDH	KAP	LAMADH
90.00	90.00	60.00	100.00	90.00	100.00
MIM	NUN	SIMKATH	E	PE	SADHE
100.00	100.00	90.00	90.00	100.00	100.00
QOP	RESH	SHEEN	TAU	AVERAGE	
100.00	80.00	80.00	30.00	87.73	

Recognition rates of words by using Zernike Moments

Data Type	Percentage					
word	accuser	altar	angel	brother	city	east
1(TS)	60.00	70.00	50.00	50.00	50.00	60.00
2(TS)	100.00	100.00	60.00	80.00	100.00	60.00
3(HW)	80.00	20.00	40.00	60.00	40.00	40.00
4(HW)	100.00	80.00	40.00	100.00	60.00	80.00

The first table represents the recognition rates of 22 alphabets where 10 shapes were handwritten carefully for each alphabet. Each shape of the total 220 shapes was assigned to the word with the nearest mean feature vectors. The second table represents the part of the recognition rates of 50 basic words by typeset(TS) and hand(HW). The second and 4-th row are the recognition rates for 5 chosen rather similar shapes for each word, total 250 and the 1st and third row are the recognition rates for 10 more diversely varaiated shapes for each word, total 500 shapes. Note that in this paper our experiment has an aim to check to waht extent such statistical moment invariants perform well in recognition of this language and did not pursue the high recognition rates by using possible more distinctive features, such as shape information. Note that we reported only the part of the recognition rates by using the Zernike moment invariants due to the lack of space. The recognition rate of characters is comparatively high and the recognition of words is too low. In Estrangelo we have the big diversity of shapes among the same word and this makes the recognition of the words more difficult. It may be concluded that we need to treat the shapes of the same word with a big difference as the different one by preprocessing.

極値データとその解析法

神戸大学　海事科学部　高橋倫也

1982年から2001年までの20年間の1万円以上の火災による損害額データがある．このデータから，年の最大または年での非常に大きな損害額の分布を求めたい．

一般にこのような問題では極値理論が用いられる．極値理論によると最大データに対しては，次の一般極値分布：

$$G_\xi\left(\frac{z-\mu}{\sigma}\right)=\begin{cases}\exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, & 1+\xi(z-\mu)/\sigma>0,\ \ (\xi\neq 0),\\ \exp\left\{-\exp\left[-\left(\frac{z-\mu}{\sigma}\right)\right]\right\}, & -\infty<z<\infty,\ \ (\xi=0),\end{cases}$$

が，また 上位 r 個のデータに対しては，次の密度関数を持つ同時一般極値分布：$z^{(1)}\geq z^{(2)}\geq\cdots\geq z^{(r)}$ として，$\xi\neq 0$ のとき，

$$\exp\left\{-\left[1+\xi\left(\frac{z^{(r)}-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}\times\prod_{k=1}^{r}\frac{1}{\sigma}\left[1+\xi\left(\frac{z^{(k)}-\mu}{\sigma}\right)\right]^{-1/\xi-1},$$

(ただし，$1+\xi(z^{(k)}-\mu)/\sigma>0$，$k=1,2,\ldots,r$．)
$\xi=0$ のとき，

$$\exp\left\{-\exp\left[-\left(\frac{z^{(r)}-\mu}{\sigma}\right)\right]\right\}\times\prod_{k=1}^{r}\frac{1}{\sigma}\exp\left[-\left(\frac{z^{(k)}-\mu}{\sigma}\right)\right],$$

が用いられる．そして，十分大きな閾値以上の超過データに対しては，次の一般パレート分布：

$$H_\xi\left(\frac{y}{\sigma}\right)=\begin{cases}1-\left(1+\xi\,\frac{y}{\sigma}\right)^{-1/\xi}, & \xi\neq 0,\\ 1-\exp\left(-\frac{y}{\sigma}\right), & \xi=0,\end{cases}$$

が用いられる．

すなわち，古典的な極値理論によるデータ解析では，毎年の最大値（極値データ），毎年の上位 r 個のデータ，そしてある閾値以上のデータにそれぞれ一般極値分布，同時一般極値分布，そして一般パレート分布を適合して解析する．また，パラメータの推定は最尤法で行う．

データがどの様な分布に従うか調べた．すべての損害額データ（1535個）と上位100個のデータを，正の値を取る分布の指数分布，ワイブル分布，対数正規分布そしてパレート分布に従うか試してみた．各確率紙にデータをプロットしその形状を調べた．

指数分布とワイブル分布ではプロットの形状は下に凸であった．これはデータはこれらの分布よりも裾の重たい分布に従う可能性を示唆している．対数正規確率紙での全データのプロットは小さな値を除いてほぼ直線を示している．データは左打ち切り対数正規分布に従う可能性がある．より裾の重たいパレート確率紙の全データのプロットを見ると上に凸の形状を示している．しかし，上位100個のデータでは上側の一部を除き直線性を示している．

まず，データに対して極値理論による古典的な解析を試みた．20個の年最大値データに一般極値分布を適合し解析した．結果は思わしくない．最大値データは一般極値分布に適合しない．また，このことから毎年の上位 r 個のデータを用いる解析は意味がない．

そこで，全データに対して閾値を決め，超過データに一般パレート分布を適合し解析を試みた．パレート確率紙を用いて閾値を決めた．すなわち，その値以上ではプロットが直線をしていると見なせる点を閾値とした．決めた閾値に対して最尤法で一般パレート分布の尺度と形状パラメータを推定した．閾値以上のデータに対して推定した一般パレート分布は適合しているように見える．

次に，すべてのデータに対して左打ち切り対数正規分布を適合することを試みた．確率紙から打ち切り点を決め，最尤法でパラメータを推定した．推定した左打ちきり対数正規分布は打ち切り点以上のデータに適合しているように見える．また，標本平均対数超過のプロットを見ると，データは対数正規分布に従っているように見える．

推定した一般パレート分布と左打ちきり対数正規分布を，各年での閾値以上のデータと打ち切り点以上のデータに適合させてみた．半分くらいの年で適合は悪かった．

データの大きな値に，一般パレート分布と左打ち切り対数正規分布を適合して解析した．どちらの分布を用いてもデータへの適合度はほぼ同じであった．しかし，一般に対数正規分布に従うデータを古典的な極値理論で扱うと最大値の予測は不安定になる．

二段階法による母平均ベクトルの信頼領域について

広島大学・理　若木 宏文

1 はじめに

$\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots i.i.d. N_p(\boldsymbol{\mu}, \Sigma)$ とし，標本平均ベクトル $\overline{\boldsymbol{X}}_n = \frac{1}{n}\sum_{j=1}^n \boldsymbol{X}_j$ を用いた，$\boldsymbol{\mu}$ の球形の信頼領域 $\{\boldsymbol{\mu} : \|\boldsymbol{\mu} - \overline{\boldsymbol{X}}_n\| \leq d\}$ を構成する問題を扱った．半径が d であり，かつ，信頼係数が $1-\alpha$ 以上となるような標本数を Σ に関する情報なしに決めることは不可能であることが知られている．

Healy (1956) は，二段階標本抽出法によって半径と信頼係数を実現する方法を与えた．適当に決めた m に対し，$\boldsymbol{X}_1, \cdots, \boldsymbol{X}_m$ を抽出し，

$$\boxed{N = \max\left\{m, \left[\frac{a_0 l_m}{d^2}\right] + 1\right\}} \tag{1}$$

とする．ただし，$[r]$ は r を越えない最大の整数値，$a_0 = \frac{p(m-1)}{m-p}F_{p,m-p}(\alpha)$，$F_{p,m-p}(\alpha)$ は自由度 $p, m-p$ の F 分布の上側 $100\alpha\%$ 点，l_m は，不偏分散共分散行列 $S_m = \frac{1}{m-1}\sum_{j=1}^m (\boldsymbol{X}_j - \overline{\boldsymbol{X}}_m)(\boldsymbol{X}_j - \overline{\boldsymbol{X}}_m)'$ の最大固有値である．

$N > m$ のとき，$\boldsymbol{X}_{m+1}, \cdots, \boldsymbol{X}_N$ を抽出し，$\boldsymbol{X}_1, \cdots, \boldsymbol{X}_N$ の標本平均を用いると，

$$\Pr\{\|\boldsymbol{\mu} - \overline{\boldsymbol{X}}_N\| \leq d\} \geq 1 - \alpha \tag{2}$$

となる．

2 有効性

二段階法においては，期待標本数が有効性の尺度に用いられる．本報告では，$0 < c < \lambda_{\min}$ を満たす c が既知であって，$m = \left[\frac{\chi_p(\alpha)c}{d^2}\right] + 1$ ととる場合において，$d \to 0$ としたときの期待標本数の漸近展開を導出した．ただし，λ_{min} は，Σ の最小固有値である．容易にわかるように，十分小さな d に対して (1) で定義された N は確率 1 で $N = \left[\frac{a_0 l_m}{d^2}\right] + 1$ となる．

Σ の最大固有値が単根の場合には，l_m を摂動展開し，項別に期待値をとることで，

$$\mathrm{E}(N) = \frac{a_0 \lambda_1}{d^2}\left(1 + \frac{1}{m}\mathrm{tr}(\Xi)\right) + \mathrm{E}(U) + o(1) \tag{3}$$

を得る．ただし，

$$U = \frac{a_0 l_m}{d^2} - \left[\frac{a_0 l_m}{d^2}\right], \tag{4}$$

Σ の固有値を $\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_p$ として，$\Xi = \mathrm{diag}(0, \frac{\lambda_2}{\lambda_1 - \lambda_2}, \cdots, \frac{\lambda_p}{\lambda_1 - \lambda_p})$ である．

Σ の最大固有値が重根でない場合，一般の次元に対して $\mathrm{E}[l_m]$ の漸近展開の導出は困難である．$p=2$ の場合には，l_m の厳密な期待値を導出することができ，それを利用すると

$$\mathrm{E}(N) = \frac{a_0\lambda}{d^2}\left(1+\frac{\pi}{\sqrt{2m}}\right) + \mathrm{E}(U) + o(1). \tag{5}$$

が導かれる．

3　U の漸近一様性

期待標本数の漸近展開のためには，(4) の U の漸近期待値が必要となる．Σ の最大固有値が単根である場合には，より一般的な形で，U の極限分布が区間 $(0,1)$ 上の一様分布であることが示された．

$\boldsymbol{Z}, \boldsymbol{Z}_1, \boldsymbol{Z}_2, \cdots$ を独立に同じ p–次元分布に従う確率ベクトルの列，H を $\boldsymbol{R}^p$ で定義された実数値関数，$\overline{\boldsymbol{Z}}_n = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{Z}_j$ とするとき，

$$U_n = nH(\overline{\boldsymbol{Z}}_n) - [nH(\overline{\boldsymbol{Z}}_n)] \tag{6}$$

の分布に関して，次の定理を得た．

定理 1．$\boldsymbol{Z}$ の分布と関数 H に関して

$$\begin{aligned} &\text{(i)}\quad \mathrm{E}(\|\boldsymbol{Z}\|^3) < \infty, \\ &\text{(ii)}\quad H \text{ は } \boldsymbol{\mu} \text{ の近傍で } C^3\text{-級}, \\ &\text{(iii)} \limsup_{\|t\|\to\infty} |\mathrm{E}[\exp(i\boldsymbol{t}'\boldsymbol{Z})]| < 1 \end{aligned} \tag{7}$$

が成り立つならば，

$$\sup_{0\le a<b\le 1} |\Pr\{a < U_n \le b\} - (b-a)| = O(n^{-1/2}). \tag{8}$$

最大固有値が重根である場合には，標本最大固有値の極限分布の確率密度関数の形状が鍵となる．極限確率密度関数が，有界，連続で極値が有限個であること，および，裾が指数オーダーで 0 に近いことが示されれば，U の漸近一様性が証明できる．$p=2$ の場合には，標準化した最大固有値の極限密度は，

$$\frac{1}{2}\phi(y_1) + \sqrt{\frac{\pi}{2}}\,y_1\phi\left(\frac{y_1}{\sqrt{2}}\right)\Phi\left(\frac{y_1}{\sqrt{2}}\right) \tag{9}$$

であることから，U の漸近一様性が示された．$p=3$ の場合にも，同様な極限密度関数の表現が与えられているので漸近一様性の証明が可能であるが，一般の p に対する証明は今後の課題である．

Regression Depth and Computation

大阪大学大学院基礎工学研究科　　藤木 美江
大阪大学大学院基礎工学研究科　　白旗 慎吾

1　はじめに

Regression Depth とは, Rousseeuw and Hubert(1999) によって提案された, ロバスト回帰の分野における新しい概念である. これが, どのように回帰分析に応用されているかを理論と応用の両面から考察し, 他の回帰推定量との比較を通して, その違いについての研究を行なった. 破綻点 (breakdown point) は, 大域的な信頼性を測る尺度で, 影響関数 (Influence function) は, 局所的なロバストネスを測る尺度である. この 2 つを用いて, 推定量のロバストネスについて調べた結果, 今までのロバスト推定法よりも, Regression Depth をもとに導き出された推定量の方が, 高い破綻点を保ちながら高い漸近効率を持つことがわかった.

本報告では, Regresssion Depth に基づく推定量の Catline (Rousseeuw and Hubert, 1998) と Deepest Regression (Rousseeuw and Aelst, 2002) についての説明を行い, 比較研究の結果を示す.

2　Regression Depth の定義

まずはじめに、回帰 (regression)depth の概念を示すために不適合 (nonfit) の定義を与える. 単回帰において, データ集合 $Z_n = \{(x_i, y_i) : i = 1, \cdots, n\}$ に対して, 直線 $y = \theta_1 x + \theta_2$ をあてはめる. θ_1 は傾き, θ_2 は切片であり, $\boldsymbol{\theta} = (\theta_1, \theta_2)$ とする. また, 残差は $r_i(\boldsymbol{\theta}) = r_i = y_i - (\theta_1 x_i + \theta_2)$ である.

定義 1　どの x_i とも一致しない実数 v が存在し, 次の *(i)* または *(ii)* が成り立つとき, データ集合 Z_n に対して, $\boldsymbol{\theta} = (\theta_1, \theta_2)$ は不適合 **(nonfit)** という.

$$(i)\quad r_i(\boldsymbol{\theta}) < 0, \forall x_i < v \quad \text{かつ} \quad r_i(\boldsymbol{\theta}) > 0, \forall x_i > v$$

$$(ii)\quad r_i(\boldsymbol{\theta}) > 0, \forall x_i < v \quad \text{かつ} \quad r_i(\boldsymbol{\theta}) < 0, \forall x_i > v$$

そこで, 回帰 depth は次のように定義することができる.

定義 2　データ集合 Z_n に対して $\boldsymbol{\theta} = (\theta_1, \theta_2)$ の $rdepth(\boldsymbol{\theta}, Z_n)$ は, $\boldsymbol{\theta}$を不適合にするために取り除かれる必要のある観測値の最小数である. 言い換えると, $rdepth(\boldsymbol{\theta}, Z_n)$ は, $\boldsymbol{\theta}$を不適合にするために符号を変える必要のある残差の最小数である.

定義 1, 2 は x_i に同点がある場合も有効であり, 分布に関する仮定もしていない.

3　Catline と Deepst Regression

Catline とは, 回帰 depth の概念をもとにした新しい単回帰分析の方法である. x の値に従ってデータ集合 $Z_n = \{(x_i, y_i) : i = 1, \cdots, n\}$ を, $x_1 \leq x_2 \leq \cdots \leq x_n$ に並べかえる. もし, x_i に同順位がある場合には y_i の小さい方から大きい方へと並べることとする. データ集合 Z_n に対して, 次のようにデータを左のグループ L, 中央のグループ M, 右のグループ R の 3 つのグループにわける. n が 3 の倍数 $(n = 3m)$ のとき, $L = \{(x_1, y_1), \cdots, (x_m, y_m)\}, M = \{(x_{m+1}, y_{m+1}), \cdots, (x_{2m}, y_{2m})\}, R = \{(x_{2m+1}, y_{2m+1}), \cdots, (x_{3m}, y_{3m})\}$ となり, $n = 3m + 1$ のとき ♯M$=m+1$, $n = 3m + 2$ のとき ♯L=♯R$=m+1$ とする.

定義 3　***Catline*** $\boldsymbol{\theta_{cat}} = (\theta_{cat(1)}, \theta_{cat(2)})$ は $L \cup M$ と $M \cup R$ を同時に 2 等分する直線である.

Catline は単回帰分析の方法であったが, 最深回帰 (Deepest Regression) は重回帰分析に拡張することができる.

定義 4　p 次元における最深回帰推定量 $T^*(Z_n)$ は, $rdepth(\boldsymbol{\theta}, Z_n)$ を最大にする $\boldsymbol{\theta}$ と定義する.

$$T^*(Z_n) = arg \max_{\boldsymbol{\theta}} \; rdepth(\boldsymbol{\theta}, Z_n)$$

4　ロバストネス

4.1　破綻点

任意の推定量 T_n の有限標本の破綻点は,

$$\varepsilon_n^*(T_n, Z_n) = \min\{ \; \frac{k}{n} \; ; \; \sup_{Z_n'} \|T_n(Z_n') - T_n(Z_n)\| = \infty\}$$

によって定義される. ここで, データ集合 Z_n の任意の k 個の観測値を, 任意の値に置き換えたときのデータ集合を Z_n' とする. 破綻点は, 推定量 T_n が $T_n(Z_n)$ から任意に離れた値をとるときの汚染の最小の割合である. この汚染により Z_n' は y_i に関する外れ値だけでなく, x_i に関する外れ値を含む. Catline と 最深回帰の漸近的な破綻点は, ともに 1/3 となる.

4.2　影響関数

分布 H における推定量 T の影響関数は, $\boldsymbol{z} = (\boldsymbol{x}^t, y)$ に小さい確率を加えることによる, T への影響を測るものである. $\Delta_{\boldsymbol{z}}$ によって, $\boldsymbol{z}$で確率 1 をもつ確率分布を表し, $H_\varepsilon = (1-\varepsilon)H + \varepsilon\Delta_{\boldsymbol{z}}$ とするとき, 影響関数は

$$\begin{aligned} IF(\boldsymbol{z}, T, H) &= \lim_{\varepsilon \downarrow 0} \frac{T((1-\varepsilon)H + \varepsilon\Delta_{\boldsymbol{z}}) - T(H)}{\varepsilon} \\ &= \lim_{\varepsilon \downarrow 0} \frac{T(H_\varepsilon) - T(H)}{\varepsilon} \\ &= \frac{\partial}{\partial \varepsilon} T(H_\varepsilon)|_{\varepsilon=0} \end{aligned}$$

により定義される. Catline と 最深回帰の傾きと切片の影響関数は, ともに区分的に滑らかで有界であるので, 外れ値にそれほど影響を受けないことがわかる.

5　Computation

2 次元において, 最深回帰は $O(n \log^2 n)$ 回 で計算することができる. n を大きくすると (または次元 p を高くすると), $O(n^{2p-1} \log n)$ 回となり, とても遅くなる. そこで, 計算時間を速くするために近似アルゴリズムが構築された. 2 次元以上でも計算時間を速く保つために, MEDSWEEP というアルゴリズムが構築された. このアルゴリズムではシミュレーションの結果より, 偏り (bias) と 平均 2 乗誤差 (MSE) が小さく, 次元 p と共に値が増加しないことが示されている (Rousseeuw and Aelst, 2002).

参考文献

[1] Aelst,S.V.,Rousseeuw,P.J.,Hubert.M.,and Struyf.A.(2002). The Deepest Regression Method, *Journal of Multivariate Analysis*, 81,138-166.

[2] Rousseeuw,P.J. and Aelst,S.V.(2000). Robustness of Deepest Regression, *Journal of Multivariate Analysis*,73,82-106.

[3] Rousseeuw,P.J. and Hubert,M.(1998). The Catline for Deep Regression, *Journal of Multivariate Analysis*,66,270-296.

[4] Rousseeuw,P.J. and Hubert,M.(1999). Regression Depth, *Journal of the American Statistical Association*,94,388-402.

正準判別分析における高次元漸近展開

姫野 哲人　　広島大・理
若木 宏文　　広島大・理
藤越 康祝　　広島大・理

この研究では、正準判別分析において重要な役割を果たす固有値の漸近分布や次元の検定に関する検定統計量の漸近分布を高次元という枠組の中で導いた。

まず、データを

$$x_1^{(i)},\cdots,x_{n_i}^{(i)} \sim N_p(\mu^{(i)},\Sigma) \quad (i=1,\cdots,q+1,\ N=n_1+\cdots+n_{q+1})$$

とする。ここで、$p \geq q$ を仮定する。そして、S_h, S_e をそれぞれ群間平方和積和行列、群内平方和積和行列とする。つまり、

$$S_h=\sum_{i=1}^{q+1} n_i(\bar{x}^{(i)}-\bar{x})(\bar{x}^{(i)}-\bar{x})', \qquad S_e=\sum_{i=1}^{q+1}\sum_{j=1}^{n_i}(x_j^{(i)}-\bar{x}_i)(x_j^{(i)}-\bar{x}_i)',$$

$$\bar{x}^{(i)}=\frac{1}{n_i}\sum_{j=1}^{n_i}x_j^{(i)}, \qquad \bar{x}=\frac{1}{N}\sum_{i=1}^{q+1}\sum_{j=1}^{n_i}x_j^{(i)}$$

とする。そして、$S_e^{-1}S_h$ の固有値を

$$l_1 \geq \cdots \geq l_q > l_{q+1}=\cdots=l_p=0$$

とおき、この漸近分布を考える。次に、

$$\tilde{\Omega}=\sum_{i=1}^{q+1} n_i(\bar{\mu}^{(i)}-\bar{\mu})(\bar{\mu}^{(i)}-\bar{\mu})'$$

(ただし、$\bar{\mu}=\frac{1}{n}(n_1\bar{\mu}^{(1)}+\cdots+n_{q+1}\bar{\mu}^{(q+1)}), \zeta : p\times q$) とする。そして、$\Sigma^{-1}\tilde{\Omega}$ の固有値を

$$\omega_1>\cdots>\omega_k>\omega_{k+1}=\cdots=\omega_p=0$$

とおく。ここで、k は $k\leq q$ とする。この時、$\tilde{\Omega}$ の次元の検定として、

$$H_0 : rank(\tilde{\Omega})=k\ (\Leftrightarrow \omega_1\geq\cdots\geq\omega_k>\omega_{k+1}=\cdots=\omega_p=0)$$

を考える。

正準判別分析において固有値の大きさは対応する正準判別変量の有効性を表す。特に固有値が 0 であることは、正準判別変量が群の識別に寄与しないことを意味する。したがって、固有値の推定や検定問題において上記の l_i の分布を導出することが重要となる。次元 p を固定して $n\to\infty$ の場合の漸近論については Siotani et al.(1985.10 章) 等の結果があるが、本研究では $p\to\infty, n\to\infty, \frac{p}{n}\in(0,1)$ の設定での基本的な統計量の分布の漸近展開を導出し、次元 p を固定したときの漸近展開と比較することが目的である。

それで、固有値の分布や次元の検定を考えていくが、次元の検定に関し代表的な検定統計量は、$l_{k+1},\cdots,l_q$ の対称関数として与えられる。そして、

$$B\sim W_q(p,I_q;\Omega), W\sim W_q(m,I_q), B \amalg W$$

とすると (だたし、$\Omega = diag(\omega_1, \cdots, \omega_q), m = n - p + q, n = N - q - 1$)、$BW^{-1}$ の固有値の分布と $l_1, \cdots, l_q$ の分布は等しくなる。よって、BW^{-1} の固有値について考えればいいことになる。あと、これらの漸近展開を求めていく時に、

$$\omega_1 > \cdots > \omega_k > 0,\ \Omega = O(p)$$

は仮定しておく。

固有値の漸近分布に関しては、まず U, V を

$$U = \frac{1}{\sqrt{p}}(B - pI_q - \Omega), \qquad V = \frac{1}{\sqrt{m}}(W - mI_q)$$

とおく。すると、U, V は漸近的に正規分布に従い、

$$B = p(I_q + \frac{1}{p}\Omega) + \sqrt{q}U, \qquad W = m(I_q + \frac{1}{\sqrt{m}}V)$$

と変形できる。ここで、$D = W^{-\frac{1}{2}}BW^{-\frac{1}{2}}$ とおくと、

$$\begin{aligned} D &= \Delta + \frac{1}{\sqrt{m}}(\sqrt{r}U - \frac{1}{2}V\Delta - \frac{1}{2}\Delta V) \\ &\quad + \frac{1}{m}(\frac{3}{8}V^2\Delta + \frac{3}{8}\Delta V^2 + \frac{1}{4}V\Delta V - \frac{\sqrt{r}}{2}VU - \frac{\sqrt{r}}{2}UV) + O(m^{-\frac{3}{2}}) \end{aligned}$$

(ただし、$\Delta = \mathrm{diag}(\delta_1, \cdots, \delta_q),\ \delta_i = r + \frac{r}{p}\omega_i,\ r = \frac{p}{m}$) と摂動展開できて、これを下に $l_1, \cdots, l_k$ のそれぞれの漸近周辺分布や、$l_{k+1}, \cdots, l_q$ の極限同時分布が求められた。

次に、次元の検定統計量としては、

$$(i) -\log \prod_{i=k+1}^{q} (1 + l_i)^{-1},\ (ii) \sum_{i=k+1}^{q} l_i,\ (iii) \sum_{i=k+1}^{q} \frac{l_i}{1 + l_i}$$

の 3 つが考えられる。帰無仮説の下でこれら 3 つの統計量はそれぞれ n 倍して、次元 p を固定して $n \to \infty$ とすると、漸近分布は全て $\chi^2_{(p-k)(q-k)}$ になることが分かっているので、
$p \to \infty, n \to \infty, \frac{p}{n} \in (0, 1)$ の状況では、

$$\begin{aligned} T_{LR} &= -\sqrt{p}(1 + \frac{m}{p})\left\{\log \prod_{i=k+1}^{q} (1 + l_i)^{-1} + (q - k)\log(1 + \frac{p}{m})\right\} \\ T_{LH} &= \sqrt{p}\left(\frac{m}{p}\sum_{i=k+1}^{q} l_i - (q - k)\right) \\ T_{BNP} &= \sqrt{p}(1 + \frac{p}{m})\left\{(1 + \frac{m}{p})\sum_{i=k+1}^{q} \frac{l_i}{1 + l_i} - (q - k)\right\} \end{aligned}$$

とおき、$l_{k+1}, \cdots, l_q$ に関する摂動展開を下に、これらの漸近分布を求めて、その漸近展開を下に Peter Hall の変形をしたものを $\chi^2_{(p-k)(q-k)}$ と数値シュミレーションで比較し、近似の良さを調べた。また、次元の検定に関する検出力の極限も求めた。

Prediction Problem in Multivariate Mixed Linear Models with Unequal Replications

University of Tokyo
Hitoshi Koyano

Mixed linear models have been extensively used in applied statistics. For example, in estimation of small area means, they have been used as a method of pooling observations to strengthen the accuracy of the estimators. For the univariate mixed linear models many theoretical studies and applications have been made by Fay and Herriot[3], Battese, Harter and Fuller[2], Prasad and Rao[9] and others. For the multivariate mixed linear models a predictor of multivariate small area means was proposed by Fuller and Harter[4] and the results were extended to more general models by Amemiya[1]. In predicting multivariate small area means, It is desired to improve upon ordinary predictors through the Stein effects. The problem of predicting multivariate small area means in the multivariate mixed linear models with equal replications was considered and a series of the results in the framework of decision theory was given by Kubokawa and Srivastava[7]. But in application it is not necessarily possible to replicate observation by the same time in all small areas. So in this paper we deal with multivariate mixed linear models with unequal replications. We consider the problem of predicting the sum of the regression mean and the random effect and gives some minimaxity results.

We deal with the multivariate mixed linear model with unequal replications

$$\boldsymbol{y}_{ij} = \boldsymbol{\beta}\boldsymbol{b}_{ij} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij}, \quad i = 1, \cdots, k, \quad j = 1, \cdots, r_i,$$

where $\boldsymbol{y}_{ij}$'s are p-variate observation vectors, $\boldsymbol{\beta}$ is a common unknown $p \times q$ regression coefficient, $\boldsymbol{b}_{ij}$'s are $q \times 1$ covariates, $\boldsymbol{\alpha}_i$'s are $p \times 1$ random effects having $N_p(\boldsymbol{0}, \Sigma_A)$ and $\boldsymbol{\epsilon}_{ij}$'s are $p \times 1$ error terms having $N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. It is supposed that $\boldsymbol{\alpha}_i$'s and $\boldsymbol{\epsilon}_{ij}$'s are independent and that $\boldsymbol{\Sigma}_A$ and $\boldsymbol{\Sigma}$ are unknown. Letting $\boldsymbol{\theta}_i = \boldsymbol{\beta}\bar{\boldsymbol{b}}_{i\cdot} + \boldsymbol{\alpha}_i$ for $\bar{\boldsymbol{b}}_{i\cdot} = r_i^{-1}\sum_{j=1}^{r_i} \boldsymbol{b}_{ij}$ and $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1 \cdots \boldsymbol{\theta}_k)$, we consider the problem of predicting $\boldsymbol{\Theta}$ under the loss

$$\mathrm{tr}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}),$$

where $\hat{\boldsymbol{\Theta}}$ is an estimator of $\boldsymbol{\Theta}$. Letting

$$\hat{\boldsymbol{\beta}} = \sum_{i=1}^{k} r_i \bar{\boldsymbol{y}}_{i\cdot}\bar{\boldsymbol{b}}_{i\cdot}' \Big(\sum_{i=1}^{k} r_i \bar{\boldsymbol{b}}_{i\cdot}\bar{\boldsymbol{b}}_{i\cdot}'\Big)^{-1}, \quad \boldsymbol{S} = \sum_{i=1}^{k}\sum_{j=1}^{r_i}(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_{i\cdot})(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_{i\cdot})'$$

for $\bar{\boldsymbol{y}}_{i\cdot} = r_i^{-1}\sum_{j=1}^{r_i} \boldsymbol{y}_{ij}$ and putting $n = \sum_{i=1}^{k} r_i - k$ and $r_* = \{r_1, \cdots, r_k\}$, we show that the shrinkage emprical Bayes estimator of the form

$$\hat{\boldsymbol{\Theta}} = (\hat{\boldsymbol{\theta}}_1 \cdots \hat{\boldsymbol{\theta}}_k),\ \hat{\boldsymbol{\theta}}_i = \bar{\boldsymbol{y}}_{i\cdot} - \hat{\boldsymbol{\Delta}}_i(\bar{\boldsymbol{y}}_{i\cdot} - \hat{\boldsymbol{\beta}}\bar{\boldsymbol{b}}_{i\cdot}),$$

where $\hat{\boldsymbol{\Delta}}_i = (\boldsymbol{Q}')^{-1}\boldsymbol{\Phi}_i(\boldsymbol{F})\boldsymbol{Q}'$, $\boldsymbol{Q}$ is a $p \times p$ nonsingular matrix such that

$$\boldsymbol{Q}'\boldsymbol{S}\boldsymbol{Q} = \boldsymbol{I}_p,\quad \boldsymbol{Q}'\sum_{i=1}^{k}(\bar{\boldsymbol{y}}_{i\cdot} - \hat{\boldsymbol{\beta}}\bar{\boldsymbol{b}}_{i\cdot})(\bar{\boldsymbol{y}}_{i\cdot} - \hat{\boldsymbol{\beta}}\bar{\boldsymbol{b}}_{i\cdot})'\boldsymbol{Q} = \boldsymbol{F}$$

for $\boldsymbol{F} = \mathrm{diag}(f_1, \cdots, f_p)$ with $f_1 \geqq \cdots \geqq f_p \geqq 0$ and $\boldsymbol{\Phi}_i(\boldsymbol{F}) = \mathrm{diag}(\phi_1^{(i)}(\boldsymbol{F}), \cdots, \phi_p^{(i)}(\boldsymbol{F}))$, is minimax estimator which dominates the sample mean $\bar{\boldsymbol{Y}} = (\bar{\boldsymbol{y}}_{1\cdot} \cdots \bar{\boldsymbol{y}}_{k\cdot})$, if

$$0 < a \leqq \frac{2}{n+p+1}\Big\{\sum_{i=1}^{k}\frac{r_*^2(1-c_i)}{r_i^2} - (p+1)(1-c_*)\Big\}$$

when we set

$$\phi_j^{(i)}(\boldsymbol{F}) = \frac{1}{1 + r_i\hat{\lambda}_j(\boldsymbol{F})},\ \hat{\lambda}_j(\boldsymbol{F}) = \frac{f_j}{a}$$

for the constant a and if

$$\begin{aligned} 0 \ <\ a \ &\leqq\ \frac{2}{n+p+1}\Big\{\sum_{i=1}^{k}\frac{r_*^2(1-c_i)}{r_i^2} - (p+1)(1-c_*)\Big\}, \\ 0 \ <\ 2a+b \ &\leqq\ \frac{2}{n+p+1}\Big\{\sum_{i=1}^{k}\frac{r_*^2(1-c_i)}{r_i^2} - (p+1)(1-c_*) + (p-1)(1-c_*)\Big\} \end{aligned}$$

when we set $\boldsymbol{\Psi}_i(\boldsymbol{F}) = \boldsymbol{F}\boldsymbol{\Phi}_i(\boldsymbol{F})$ and

$$\psi_j^{(i)}(\boldsymbol{F}) = \frac{f_j}{1 + r_i f_j/a} + \frac{f_j}{1 + r_i\sum_{c=1}^{p} f_c/b}$$

for the constants a and b, where $c_i = r_i\bar{\boldsymbol{b}}_{i\cdot}'(\sum_{l=1}^{k} r_l\bar{\boldsymbol{b}}_{l\cdot}\bar{\boldsymbol{b}}_{l\cdot}')^{-1}\bar{\boldsymbol{b}}_{i\cdot}$ and $c_* = \{c_1, \cdots, c_k\}$. From the results of the simulations we recommend the second minimax empirical Bayes estimator.