

(9) 「多変量解析～線形から非線形へ～」に関する研究報告

久保川達也 (東京大学経済学部) : Estimating the Covariance Matrix: A New Approach.....	349
萩原克幸 (三重大学工学部物理工学科) : 階層型ニューラルネットによる非線形回帰モデルのモデル選択規準について	351
辻谷将明 (大阪電気通信大学情報工学部) : ニューロ多群判別モデルにおけるブートストラップ法について	353
美馬芳江 (広島大・理)・藤越康祝 (広島大・理) : Selection of Variables and Non-Linearity in Multivariate Calibration	355
安楽和夫 (西南学院大学・文) : 順序制約下での変化点とモデル選択法について.....	357
布能英一郎 (関東学院大経済学部) : A dmissibility of the MLE for multinomial distribution with monotone missing data	359
杉山高一 (中央大学理工)・山田智哉 (中央大学理工) : 正準相関係数における permutation test と近似分布について	361
鈴川晶夫・種市信裕 (帯広畜産大学畜産学部)・打切りデータに基づく生存関数のセミパラメトリック推定	363
瀬尾 隆 (東京理科大学・理) : On the Modified Second Bonferroni's Approximation for the T_{\max}^2 Statistic in the Elliptical Distributions	365
綿森葉子 (大阪女子大学) : 球面上の統計解析	367
早川 毅 (一橋大学・経済) : Model misspecification のもとでの統計量の漸近展開について	369
伊藤雅明 (広島大学工学部) 佐藤 学 (広島大学工学部) : 完全な単純構造に近い場合におけるある因子数決定方式の適用条件について	371
長岡創志 (広島大・理)・藤越康祝 (広島大・理) : 次元縮小回帰モデルにおける次元の推定	373

Estimating the Covariance Matrix: A New Approach

東京大学 経済学部 久保川達也

独立な確率行列 \mathbf{X} ($p \times m$), \mathbf{S} ($p \times p$) が, それぞれ $\mathcal{N}(\boldsymbol{\Xi}, \boldsymbol{\Sigma}, \mathbf{I}_m)$, $\mathcal{W}_p(\boldsymbol{\Sigma}, n)$, $n \geq p$, に従うとする。このとき $\boldsymbol{\Sigma}$ を Stein 損失関数

$$L(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \text{tr } \widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1} - |\widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}| - p,$$

のもとで推定する問題を考える。

\mathbf{S} を正の対角成分をもった下三角行列 \mathbf{T} により $\mathbf{S} = \mathbf{T}\mathbf{T}^t$ のように分解したとき, James and Stein (1961) は, 不偏推定量 $\widehat{\boldsymbol{\Sigma}}^{UB} = n^{-1}\mathbf{S}$ が

$$\widehat{\boldsymbol{\Sigma}}^{JS} = \mathbf{T}\mathbf{D}\mathbf{T}^t, \quad \mathbf{D} = \text{diag}(d_1, \dots, d_p), \quad d_i = (n + p + 1 - 2i)^{-1}$$

なるミニマクス推定量によって改良されることを示した。以来 数多くの文献で研究されてきたが, \mathbf{X} に含まれる情報を用いて $\widehat{\boldsymbol{\Sigma}}^{JS}$ を改良する問題はなされてこなかった。ここでは, この問題を解決するための2つのアプローチを提案し, より有効な推定手法を与える。

[1] A New Approach.

$m \times p$ 行列 \mathbf{Y} を

$$\mathbf{Y} = (y_{ij}) = (\mathbf{y}_1, \dots, \mathbf{y}_p)$$

で定義し, $j = 1, \dots, p$ に対して, $m \times m$ 行列 \mathbf{C}_j を

$$\mathbf{C}_j = \mathbf{C}_{j-1} - (1 + \mathbf{y}_{j-1}^t \mathbf{C}_{j-1} \mathbf{y}_{j-1})^{-1} \mathbf{C}_{j-1} \mathbf{y}_{j-1} \mathbf{y}_{j-1}^t \mathbf{C}_{j-1}$$

$\mathbf{C}_1 = \mathbf{I}_m$ で定義する。このとき $|\mathbf{I}_p + \mathbf{Y}^t \mathbf{Y}| = \prod_{i=1}^p (1 + \mathbf{y}_i^t \mathbf{C}_i \mathbf{y}_i)$ が成り立つ。統計量 $\mathbf{y}_i^t \mathbf{C}_i \mathbf{y}_i$ を用いて

$$g_i = \min \left\{ \frac{1}{n + p + 1 - 2i}, \frac{1 + \mathbf{y}_i^t \mathbf{C}_i \mathbf{y}_i}{n + m + p + 1 - 2i} \right\}.$$

とおき, 新たな推定量

$$\widehat{\boldsymbol{\Sigma}}^{TR} = \mathbf{T} \text{diag}(g_1, \dots, g_p) \mathbf{T}^t,$$

を提案すると, 次の定理により $\widehat{\boldsymbol{\Sigma}}^{TR}$ の優越性が示される。

定理 1. 打ち切り推定量 $\widehat{\boldsymbol{\Sigma}}^{TR}$ は, James-Stein ミニマクス推定量 $\widehat{\boldsymbol{\Sigma}}^{JS}$ を改良する。

[2] An Approach Based on Srivastava-Kubokawa (1999).

ここでは $m \geq p$ の場合のみを考える。 $p \times p$ 直交行列 P を用いて対角化して

$$P^t(XX^t)^{-1/2}S(XX^t)^{-1/2}P = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$$

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, とおく。このとき, 非負の関数 $\psi_i(\Lambda)$ に対して

$$\Psi(\Lambda) = \text{diag}(\psi_1(\Lambda), \dots, \psi_p(\Lambda))$$

とおいて,

$$\widehat{\Sigma}(\Psi) = (XX^t)^{1/2}P\Psi(\Lambda)P^t(XX^t)^{1/2}$$

なる推定量を考える。これは, Srivastava and Kubokawa (1999) によって動機付けられた形をしている。 $\widehat{\Sigma}(\Psi)$ に対して次のような打ち切りルール $[\Psi(\Lambda)]^{TR}$ を提案する。

$$\begin{aligned} [\Psi(\Lambda)]^{TR} &= \text{diag}(\psi_1^{TR}(\Lambda), \dots, \psi_p^{TR}(\Lambda)), \\ \psi_i^{TR}(\Lambda) &= \min \left\{ \psi_i(\Lambda), \frac{\lambda_i + 1}{n + m} \right\}, \quad i = 1, \dots, p \end{aligned}$$

このルールに従うと

$$\widehat{\Sigma}([\Psi]^{TR}) = (XX^t)^{1/2}P\text{diag}(\psi_1^{TR}(\Lambda), \dots, \psi_p^{TR}(\Lambda))P^t(XX^t)^{1/2}$$

なる打ち切り推定量が得られる。

定理 2. 打ち切り推定量 $\widehat{\Sigma}([\Psi]^{TR})$ は $\widehat{\Sigma}(\Psi)$ を改良する。

この結果からいくつかの興味深いミニマクス推定量が導かれる。例えば, Stein 型ミニマクス推定量は

$$\Psi^S(\Lambda) = \text{diag}(d_1\lambda_1, \dots, d_p\lambda_p)$$

に対して, $\widehat{\Sigma}^S = \widehat{\Sigma}(\Psi^S)$ と表わされるが, 上述の打ち切りルールを適用することにより, $\widehat{\Sigma}([\Psi^S]^{TR})$ によってさらに改良されることがわかる。

以上述べてきたように, X に含まれる情報を用いて一層改良されたミニマクス推定量を導くことができる。また一般化分散の推定についても [1] の方法は有用で, 従来の結果の別証明を与えるとともに, 新たな推定量の導出を行うこともできる。

本研究成果の詳しい内容については下の文献を参照して下さい。

Kubokawa, T. and Srivastava, M.S. (1999). Estimating the covariance matrix: A new approach. Discussion Paper Series, CIRJE-F-52, Faculty of Economics, University of Tokyo.

1. はじめに

階層型ニューラルネットを用いた非線形回帰モデルでは、モデル選択規準 AIC(Akaike Information Criterion)[1] で仮定されているような、モデルが真の分布を含む状況において、真の結合重み(ネットワーク・パラメータ)が識別不能となり、モデル選択規準を導く際、通常の漸近展開が適用できないという問題が生じる。モデル選択規準は、学習誤差に基づく汎化誤差の不偏推定量であるため、階層型ニューラルネットのモデル選択の問題を解決するためには、真の結合重みが識別可能でない場合について、汎化誤差および学習誤差のデータの分布に関する期待値を知る必要がある。本研究では、階層型のニューラルネットの与える関数をパラメトライズされた基底関数の線形結合と見做し、その基底関数の学習における選択性に着目し、学習誤差の期待値の上界および汎化誤差の期待値の下界を導く。また、結合重みが識別可能な場合に漸近展開により導かれるそれぞれの期待値とこれらの結果を比較する。

2. 問題設定

確率密度 $q(x, y) = q(y|x)q(x)$ をもつ確率分布 $Q(x, y) = Q(y|x)Q(x)$ からの独立な T 個のデータ $(x_t, y_t), t = 1, \dots, T$ の集まりを $D_T = \{(x_t, y_t) : x_t \in \mathbf{R}^m, y_t \in \mathbf{R}, 1 \leq t \leq T\}$ と表し、 y_t は x_t に対し、 $y_t = h(x_t) + \xi_t, t = 1, \dots, T$ という規則により生成されるものとする。ただし、 ξ_t は平均が 0 で有限な分散 σ_*^2 をもつ確率分布からの独立なサンプルとする。一方、関数 $\psi_{b_i, \varepsilon_i} : \mathbf{R}^m \rightarrow \mathbf{R}, i = 1, \dots, n$ 達の線形結合 $f_w(x) = \sum_i c_i \psi_{b_i, \varepsilon_i}(x), x \in \mathbf{R}^m$ により D_T をフィッティングすることを考える。ただし、 $w = (c, b, \varepsilon) \in W$ はネットワークの結合重みすなわちパラメータであり、 $c = (c_1, \dots, c_n), c_i \in \mathbf{R}, b = (b_1, \dots, b_n), b_i \in \mathbf{B}, \varepsilon = (\varepsilon_1, \dots, \varepsilon_n), \varepsilon_i \in \mathbf{E}$ とする。ここで、 $W = \mathbf{R}^n \times \mathbf{B}^n \times \mathbf{E}^n$ は関数のパラメータ空間である。損失を二乗誤差とし、 $s_{D_T}(w) = s_{D_T}(c, b, \varepsilon) = \frac{1}{T} \sum_t (y_t - f_w(x_t))^2$ と定義する。これを経験二乗誤差と呼ぶ。 $b \in \mathbf{B}^n$ および $\varepsilon \in \mathbf{E}^n$ を任意に固定したとき、 $s_{D_T}(\hat{c}(b, \varepsilon), b, \varepsilon) = \min_c s_{D_T}(c, b, \varepsilon)$ と定義し、 $s_{D_T}(\hat{w}) = \inf_{b, \varepsilon} s_{D_T}(\hat{c}(b, \varepsilon), b, \varepsilon)$ と定義する。これを学習誤差と呼ぶ。 $\hat{w} = \hat{w}(D_T)$ は上の定義の下での最小二乗推定量である。学習誤差の D_T の確率分布に関する期待値を $E_{D_T} \{s_{D_T}(\hat{w})\} = \int \dots \int s_{D_T}(\hat{w}) \prod_{t=1}^T q(x_t, y_t) dy_t dx_t$ で定義する。これを学習誤差の (D_T の確率分布に関する) 期待値と呼ぶ。一方、 $D_T^+ = \{(u_t, z_t) : 1 \leq t \leq T\}$ を、互いに独立かつ各 (u_t, z_t) がどの $(x_t, y_t) \in D_T$ とも独立になるように Q からサンプルされたデータの集まりとする。このとき、汎化誤差を $E_{D_T^+} \{s_{D_T^+}(\hat{w})\} = \int \dots \int s_{D_T^+}(\hat{w}) \prod_{t=1}^T q(z_t, u_t) dz_t du_t$ と定義する。ただし、 $s_{D_T^+}(\hat{w}) = \frac{1}{T} \sum_t (z_t - f_{\hat{w}}(u_t))^2$ である。通常、モデル選択の枠組では、 $E_{D_T} \{E_{D_T^+} \{s_{D_T^+}(\hat{w})\}\}$ によるネットワークの評価を考える。これを汎化誤差の (D_T の確率分布に関する) 期待値と呼ぶ。

3. n 点のみをフィッティングする関数

いま、パラメータ $\beta \in \mathbf{R}^m$ により決まる指標関数を $\chi_\beta(x) = 1 (\beta = x); 0 (\beta \neq x), x \in \mathbf{R}^m$ と定義し、 $f_\theta(x) = \sum_i \alpha_i \chi_{\beta_i}(x)$ と定義する。ただし、 $\theta = (\alpha, \beta), \alpha = (\alpha_1, \dots, \alpha_n), \alpha_i \in \mathbf{R}, \beta = (\beta_1, \dots, \beta_n), \beta_i \in \mathbf{B} \subset \mathbf{R}^m, \mathbf{B} = \{x_t : x_t \in \mathbf{R}^m, 1 \leq t \leq T\}$ とする。 $s_{D_T}(\theta) := \frac{1}{T} \sum_t (y_t - f_\theta)^2$ とし、 $s_{D_T}(\hat{\theta}) = \min_\theta s_{D_T}(\theta)$ とする。このとき $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ はパラメータの最小二乗推定量である。いま、データについて次のことを仮定する。

仮定 1 $q(y|x) = 1/\sqrt{2\sigma_*^2} \exp\{-y^2/(2\sigma_*^2)\}$ とする。すなわち、 y_t の確率分布を正規分布 $N(0, \sigma_*^2)$ とする。これは、与えるデータが正規雑音のみ、すなわち、データの生成機構において、 $\xi_t \sim N(0, \sigma_*^2)$ であり、 $h(x) = 0, x$ a.e. であることを意味する。

この仮定の下では、パラメトライズされた基底関数の線形結合による非線形回帰モデルのパラメータ値は常に識別不能となり、通常の漸近展開により汎化誤差の期待値と学習誤差の期待値の間を導くことはできない。仮定 1 の下で、[2] の結果を使うと、次のことが言える。

補題 1 仮定 1 の下で、十分大きな $T \gg n$ に対し、 $E_{D_T} \{s_{D_T}(\hat{\theta})\} \sim \sigma_*^2 - \sigma_*^2 \frac{2n}{T} \log T$ 。

4. 主要な結果

$\psi_{\bar{b}_i, \varepsilon_i}, i = 1, \dots, n$ について次の仮定をおく.

仮定 2 任意の $\beta \in \mathbf{R}^m$ に対し, $\bar{b}_i \in \mathbf{B}$ と $a_i \in \bar{\mathbf{E}}$ が存在して, 次式を満たすものとする.

$$\lim_{\varepsilon_i \rightarrow a_i} \psi_{\bar{b}_i, \varepsilon_i}(x) = \chi_{\bar{\beta}}(x), \quad x \in \mathbf{R}^m, \text{ a.e.}$$

この仮定の下で, $\beta = \hat{\beta}_i$ とおいて, $\lim_{\varepsilon \rightarrow a} s_{D_T}(\hat{c}, \bar{b}, \varepsilon) = s_{D_T}(\hat{\theta})$ が示される. ただし, $\bar{b} = (\bar{b}_1, \dots, \bar{b}_n)$ および $a = (a_1, \dots, a_n)$ である. これと補題 1 より, 学習誤差の期待値について次のことが言える.

定理 1 仮定 1 および 2 の下で, 十分大きな $T \gg n$ に対し,

$$E_{D_T} \{s_{D_T}(\hat{w})\} \leq \sigma_*^2 - \sigma_*^2 \frac{2n}{T} \log T.$$

次に汎化誤差の期待値を解析する. ここでは, 次のことを仮定する.

仮定 3 入力データの確率分布 $Q(x)$ について, $Q(x) = \frac{1}{T} \sum_t \delta(x - r_t)$ とする. ただし, $\delta(x - r_t), r_t \in \mathbf{R}^m$ は Dirac の δ 関数である. これは, 各入力データ x_t が確定的に r_t をとることを意味する.

仮定 3 の下では, 定理 1 の結果を使うと, 汎化誤差の期待値について, 直ちに次のことが知れる.

定理 2 仮定 1, 2 および 3 の下で, 十分大きな $T \gg n$ に対し,

$$E_{D_T} \left\{ E_{D_T^+} \left\{ s_{D_T^+}(\hat{w}) \right\} \right\} \geq \sigma_*^2 + \sigma_*^2 \frac{2n}{T} \log T.$$

定理 1 および定理 2 は, Gaussian Radial Basis Function や中間層に bell 型の活性化関数をもつ 3 層階層型ニューラルネットおよび中間層に sigmoid 型の活性化関数をもつ 3 層階層型ニューラルネットなどのニューラルネットのクラスに対して成り立つことが容易に示される.

一方, 仮定 1 の下で, モデルが識別可能な場合に漸近展開により導かれる学習誤差の期待値および汎化誤差の期待値は, それぞれ, $S_T(N, T) = \sigma_*^2 - \sigma_*^2 N/T$ および $S_G(N, T) = \sigma_*^2 + \sigma_*^2 N/T$ となる. ただし, N はネットワークの全パラメータ数である. したがって, この場合の学習誤差の期待値は定理 1 の上界より大きく, 汎化誤差の期待値は定理 2 の下界より小さい. これについては, 正規分布に従う確率変数の絶対値の列の最大値に関する [4] の結果を使うと, 次のような強い結果が得られる. 仮定 1, 2 および 3 の下で, $T \rightarrow \infty$ のとき, 確率 1 で,

$$s_{D_T}(\hat{w}) \leq \sigma_*^2 - \sigma_*^2 \frac{2}{T} \log T, \quad E_{D_T^+} \left\{ s_{D_T^+}(\hat{w}) \right\} \geq \sigma_*^2 + \sigma_*^2 \frac{2}{T} \log T.$$

がそれぞれ成り立つ. したがって, $T \rightarrow \infty$ のとき, $s_{D_T}(\hat{w}) \leq S_T(N, T)$ および $E_{D_T^+} \left\{ s_{D_T^+}(\hat{w}) \right\} \geq S_G(N, T)$ が確率 1 で成り立つ.

5. おわりに

本稿では, 結合重みが識別可能でなくなる場合の例として, データを正規雑音のみとした場合について, 定義した最小二乗誤差規範の下で, 任意の入力分布の下での学習誤差の期待値の上界および入力が決定的な場合の汎化誤差の期待値の下界を導き, これらが成り立つニューラルネットの例を示した. 今後は, 学習誤差の期待値の下界および汎化誤差の期待値の上界を導く必要がある.

参考文献

- [1] Akaike H. : In *2nd International Symposium on Information Theory*, B.N.Petrov and F.Csáki eds., Akadémia Kiado, Budapest, 267–281, (1973).
- [2] 早坂太一, 萩原克幸, 戸田尚宏, 臼井支朗 : 日本神経回路学会誌, 4, pp.18–26 (1997).
- [3] Murata N. et al. : *IEEE Trans. on Neural Networks*, 5, 6, pp.865–872 (1994).
- [4] Deo C. M. : *Sankhyā Series A*, 34, pp.289–292, (1972).

ニューロ多群判別モデルにおけるブートストラップ法について

大阪電気通信大学 情報工学部

辻谷 将明

1. はじめに

本報告では、階層型ニューラルネットワークを多群判別問題(Multiple-group discriminant problem)に適用する。これは多項型ロジスティック判別の拡張と考えられる。出力値の確率的解釈によってネットワーク尤度を構成し、尤度原理に基づく統計的推測を行う。ブートストラップ法を援用し、i)隠れユニット数の決定、ii)適合度検定およびiii)誤判別率のバイアス補正を試みる。更に、説明(予測)変数の個数が2あるいは3なら、それらの値を2次元平面や3次元空間にプロットし、データの特徴を視覚的に捉えることができる。しかし、それが4以上になるとグラフ表現はできない。この点を踏まえ、多次元データの視覚化の観点から、3層ニューラルネットの情報圧縮について考究する。すなわち、説明変数の1次結合から成る新しい変量(圧縮スコア)による次元縮約を行う。

多群判別問題の実際例として、3種類の検査項目(AST,ALT,GLDH)に基づいて、4つの疾患(急性ウイルス性肝炎、非活動性慢性肝炎、活動性慢性肝炎、肝硬変)に分類された肝臓病データを取上げ、階層型ニューラルネットワーク(図1)を適用する。

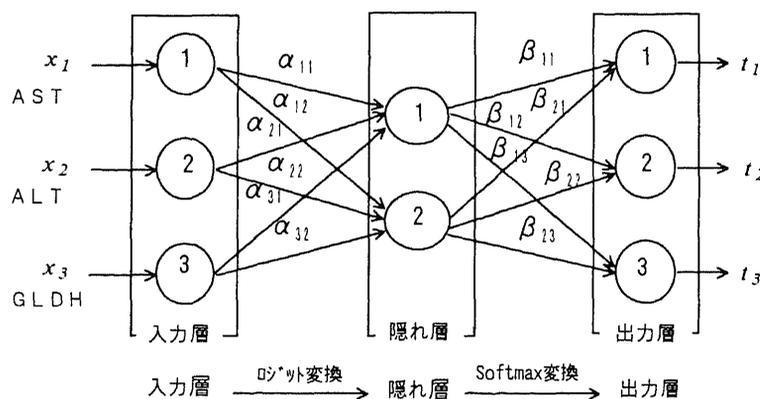


図1 肝臓病データのニューラルネットワーク表示

2. ブートストラップ法

K 個の群について、初期標本 $X = \{X_1, X_2, \dots, X_K\}$ が得られたとする。ここに、 $X_k = \{X_1^{(k)}, X_2^{(k)}, \dots, X_n^{(k)}\}$ の各要素 $X_i^{(k)}$ は、 I 個の入力値 $\{x_1, x_2, \dots, x_I\}$ で構成されている。

$d (= 1, 2, \dots, D; D = \sum_{k=1}^K n_k)$ 番目の出力値を $o_k^{<d>} = o_k^{<d>}(X; \theta)$, $\theta = \{\alpha, \beta\}$ すると、尤度は、

$$L(X|\theta) = \prod_{d=1}^D \prod_{k=1}^K \{o_k^{<d>}\}^{i_k^{<d>}}, \sum_{k=1}^K o_k^{<d>} = 1 \quad (1)$$

で与えられる。

(1) 隠れユニット数の決定

競合するモデルが複数個あるとき、対数尤度の比較によって、モデルを選択すると、自由なパラメータ数の大きいモデルほど選ばれやすい。EIC は、対数尤度のバイアスをブートストラップ法を用いて直接推定している。ブートストラップ法を用いると

$$C^* = E_X \{ \ln L(X^* | \hat{\theta}(X^*)) - \ln L(X^* | \hat{\theta}(X)) \} \quad (2)$$

によってバイアスの近似値が得られる。ただし、 $X^* = \{X_1^*, X_2^*, \dots, X_K^*\}$ はブートストラップ標本である。 $\hat{\theta}(X)$, $\hat{\theta}(X^*)$ はそれぞれ初期標本 X 、およびブートストラップ標本 X^* に基づく

θ の推定量とする。このとき、ブートストラップ法に基づく情報量規準

$$EIC = -2\ln L(X | \hat{\theta}(X)) + 2C^* \quad (3)$$

が得られる。 EIC が最小のモデルを最適モデルとして選択することができる。なお、本報告では、分離サンプリングを採用する。

(2) 適合度検定

モデルの適合度は、逸脱度(deviance)

$$Dev = 2[\ln L(\max) - \ln L(\{X | \theta\})] \sim \chi_{D-p}^2 \quad (4)$$

を用いて検証できる。ここに、 $L(\max) = \prod_{j=1}^p \prod_{k=1}^K \{t_i^{(k,j)}\}^{t_i^{(k,j)}}$ 、 p は未知パラメータの個数で $p = H(I+K-1) + H+K-1$ となる。しかし、本稿で取上げている二値データの場合、(4)式の漸近カイ二乗性は成立しない。そこで、ブートストラップ法に基づく棄却点の算出を試みる。

(3) 誤判別率の推定

判別分析では初期標本に基づいて何らかの判別ルールを構築するが、この初期標本に対する誤判別率は見かけ上の誤判別率と呼ばれる。実際の誤判別率は、初期標本に基づいて判別ルールが得られたという条件のもとで将来観測されるデータに対して予測を誤ってしまう確率であり、これを見かけ上の誤判別率で置き換えて推定することは、実際の誤判別率を過小に推定する傾向がある。ブートストラップ法を用い、見かけ上の誤判別率のバイアス補正を行う。

4. 情報圧縮

隠れ層の第 j ユニットの活性化値を用い、予測変数の線形結合として、圧縮スコアを

$$\hat{u}_j = \sum_{i=0}^I \hat{\alpha}_{ij} x_i \quad (5)$$

と定義する。これは、正準判別スコアと類似の考え方である。これは、正準判別スコアと類似の考え方である。図2は、隠れユニットが $h=3$ 個の場合の2次元プロットである。

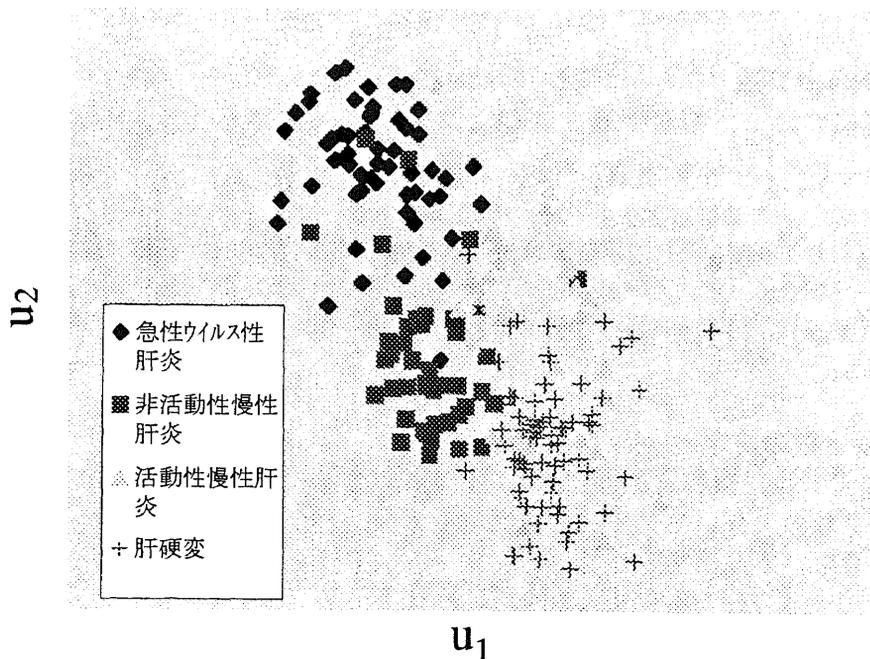


図2 圧縮スコア (\hat{u}_1, \hat{u}_2)

Selection of Variables and Non-Linearity in Multivariate Calibration

美馬芳江 (広島大・理)
藤越康祝 (広島大・理)

1. はじめに

p 次元目的変数 \mathbf{y} と q 次元説明変数 \mathbf{x} との間に回帰モデル $\mathbf{y} = h(\mathbf{x}) + \mathbf{e}$ が想定され、 N 個の個体について、独立なデータが観測されているとするこのとき、 \mathbf{y} から \mathbf{x} を推定することは、校正 (Calibration) 問題と呼ばれる。本稿では、どの目的変数を用いるべきかという問題について考察する。AIC型リスクにおいて、リスクとその推定量との差 (バイアス) を修正することにより、改良された選択基準を提案する。また、モデルに線形性が仮定できない場合の非線形逆回帰推定問題についても考察する。

2. 変数選択基準の改良

N 個の観測値 $Y' = (\mathbf{y}_1, \dots, \mathbf{y}_N)$, $X' = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ に対して、候補のモデルのモデルとし、線形正規モデル $Y \sim N(\mathbf{1}_N \alpha' + X\beta, \Sigma \otimes I_N)$, $\Xi = (\alpha \beta')$ を考える。ここで、 $\text{rank}(X) = q$, $p > q$, $N > p + q$ とする。

\mathbf{x} の推定量に対し、モデル $M(k)$ として、 \mathbf{y}_1 のみが \mathbf{x} の推定に意味のある情報を与える、つまり $(p-k)$ 個の目的変数 $\mathbf{y}_2' = (\mathbf{y}_{k+1}, \dots, \mathbf{y}_N)$ が推定に対して追加情報を持たないとする条件付パラメトリックモデル $M(k)$

$$Y_1 \sim N(Z\Xi_1, \Sigma_{11} \otimes I_N)$$

$$Y_2|Y_1 \sim N(\mathbf{1}_N \tilde{\alpha}'_2 + Y_1\Gamma, \Sigma_{22.1} \otimes I_N)$$

を考える。ただし、 $\Xi = (\Xi_1 \Xi_2')$ 。

候補のモデル $M(k)$ に対して、AIC型リスク (予測対数尤度の (-2) 倍の期待値), R_k を考える。以下、候補のモデル $M(k)$ は真のモデル M^* を含むことを仮定すると、 R_k の推定として、 Y_F を Y で置き換えたものと、リスク R_k の差は

$$B = -Np + E_Y^* \left\{ N \text{tr}(\hat{\Sigma}^{-1} \Sigma) + \text{tr}(\hat{\Sigma}^{-1}(\hat{\Xi} - \Xi)' Z' Z(\hat{\Xi} - \Xi)) \right\}$$

$$= -Np + \frac{Nk(N+q+1)}{N-q-k-2} + \frac{N(N+1)(p-k)}{N-p-2}$$

$$+ \frac{N(N+1)(p-k)}{N-p-2} E^* \left(\text{tr}(\Sigma_{11} S_{11}^{-1}) \right) + \frac{N(p-k)}{N-p-2} E^* \left(\text{tr}(\Omega S_{11}^{-1}) \right)$$

と求められる。ここで $\Omega = \beta_1' S_{xx} \beta_1$ である。 S_{11} は、非心度 Ω のウィッシュャート分布に従うため、その逆行列の期待値を正確に求めることはできないが、 $\Omega = O(N)$ を仮定して、漸近的近似が求められる。この近似には、未知パラメーターが含まれているため、 Σ_{11} , Ω に対して次の不偏推定を代入する。バイアス補正した B を $\Omega = O(1)$, $\Omega = O(N)$ の場合についてそれぞれ、 \hat{B}_1 , \hat{B}_2 と書き、次の選択基準を提案する。

$$\text{AIC}_i = Np(\log 2\pi + 1) + N(\log |\hat{\Sigma}_{11}| + \log |\hat{\Sigma}_{22.1}|) + \hat{B}_i, \quad i = 1, 2$$

3. 基準の評価

ここでは、バイアス補正を行った AIC_1 AIC_2 が、 AIC に比べてどの程度改良しているかを調べる。Brown(1982)によって与えられた、水溶液中の成分比率に対する赤外線反射のデータに選択基準を適用する。このデータは、 x を水溶液中の水分とたんぱく質の比率とし、 y はその水溶液に対する4つの赤外線反射を測定したものであり、21個のデータセットが与えられている。ここで、選択基準の比較を行うため、 h 番目の個体からのデータを予測のためのものとし、残りの20個の個体からのデータを用いて予測を行う。ここでの問題は、説明変数 x を推定するために y_1, y_2, y_3, y_4 のうちの最もよい組み合わせはなんであるかということである。

AIC は、 $h = 2, 3$ の個体に対する推定に際して (y_1, y_2, y_4) の目的変数を用いることを最良とし、それ以外の個体に対する推定に際しては、すべての目的変数 (y_1, y_2, y_3, y_4) を用いることを最良としている。しかし、 $AIC_{(1,2,4)}$ と $AIC_{(1,2,3,4)}$ の値の差は非常に小さい。一方、 AIC_1, AIC_2 は、すべての個体に対する推定に際して、 (y_1, y_2, y_4) の目的変数を用いることを最良としている。この結果からもわかるように、 AIC_1, AIC_2 は、真のモデルよりも大きいモデルを選ぶ可能性がある AIC 選択基準を改良されているといえよう。

4. 線形から非線形へ

Brown(1991)は、説明変数と目的変数間の関係式 $h(x)$ が推定できる時、逆回帰推定を、逆変換によって x の推定を行った場合の以下の2つの問題点を指摘している。

- (1) 推定量 \hat{x} が、 $g_Y(x)$ の局所的な最小値をとってしまう。
- (2) x に対する信頼区間が、複数個に分解されてしまう。

そのためBrown(1993)は、上の問題がどの程度生じるかを調べるため、2重点の集合 $D^*(x_1, x_2) = \{Y : f'_Y(x_i) = 0, f''_Y(x_i) > 0, x_i \in I, i = 1, 2, x_1 \neq x_2\}$, $D^* = \bigcup D^*(x_1, x_2)$ について幾何学的に考察している。また、各水準 x に対して、 Y が D^* に含まれる確率 $p(x) = P(Y \in D^*)$ を求めることにより、推定された $\hat{h}(\cdot)$ の予測に対する有効性、候補となるモデルの比較をする際の基準として用いることができるとしている。

参考文献

- [1] Brown, P. J. (1982). Multivariate Calibration. *J.R.Statist.Soc., B*, **44**, 287-321.
- [2] Brown, P. J. (1991). Double points in nonlinear calibration. *Biometrika*, **78**, 33-43.
- [3] Brown, P. J. (1993). *Measurement, Regression, and Calibration*. Oxford University Press, Oxford.
- [4] Fujikoshi, Y. & Nishii, R. (1984). Selection of Variables in Multivariate Regression. *Hiroshima Math.J.*, **13**, 269-277.
- [5] Krutchkoff, R. G. (1967). Classical and Inverse Regression Methods of Calibration. *Technometrics*, **9**, 425-439.

順序制約下での変化点とモデル選択法について

西南学院大学・文 安楽 和夫

1 序

k 個の母集団分布があり, それらのスカラー母数 $\theta_1, \theta_2, \dots, \theta_k$ の間に単順序: $\theta_1 \leq \theta_2 \leq \dots \leq \theta_k$ のような制約が仮定できるとする. このとき真の母数の異なる値の組み合わせ

$$\theta_1 = \dots = \theta_{m_1} < \theta_{m_1+1} = \dots = \theta_{m_1+m_2} < \dots < \theta_{\sum_{j=1}^{h-1} m_j+1} = \dots = \theta_k$$

を同定すること, あるいは $\theta_j < \theta_{j+1}$ となるような変化点を見つけることがここでの問題である. 例えば, 薬物の容量-反応実験で, 投与量があるレベルに達すると, 副作用や毒性が発現するようになるとする. このような場合の最初の変化点あるいは閾値を見つけるという問題が一つの典型的な応用例である. これは, 薬の効果を評価する場合の変化点の問題とは本質的に異なるものである. 薬効では検定法を用いることの妥当性はあると思うが, このような毒性を問題にする場合, 通常の有意水準の下での検定法を使うと, 危険なレベルを見逃してしまう可能性が高くなってしまう. このような問題については, 通常の検定法よりもモデル選択法の方がむしろ妥当と思われる. ここでは, AIC と同様に Kullback-Leibler 情報量に基づいて, モデル選択の方法を考える.

2 順序制約の下での情報量規準

ここでは $x_{ij} \sim N(\theta_i, \sigma^2)$ (σ^2 は未知) ($i = 1, \dots, k, j = 1, \dots, n_i$) とする. このとき, $\theta_1 \leq \dots \leq \theta_k$ の下での θ_i, σ^2 の最尤推定量をそれぞれ $\hat{\theta}_i$ ($i = 1, \dots, k$); $\hat{\sigma}^2$ とする (Robertson 他 (1988), p.63). このとき, 期待対数尤度を対数尤度で推定するとすると, 次の意味でのバイアスを生じるであろう:

$$\begin{aligned} B_2(\boldsymbol{\theta}, \sigma^2) &= E \left\{ l(\tilde{\boldsymbol{\theta}}, \tilde{\sigma}^2) - \sum_{i=1}^k n_i \int f(t_i, \theta_i, \sigma^2) \log f(t_i, \tilde{\theta}_i, \tilde{\sigma}^2) dt_i \mid \boldsymbol{\theta}, \sigma^2 \right\} \\ &= E \left\{ \frac{N}{2} \left(\frac{\sigma^2}{\tilde{\sigma}^2} - 1 \right) + \frac{1}{2} \sum_{i=1}^k \frac{n_i (\tilde{\theta}_i - \theta_i)^2}{\tilde{\sigma}^2} \mid \boldsymbol{\theta}, \sigma^2 \right\}. \end{aligned} \quad (1)$$

ここで $B_2(\boldsymbol{\theta}, \sigma^2)$ の最小値は $\boldsymbol{\theta} \in H = \{\boldsymbol{\theta} \mid \theta_1 = \dots = \theta_k\}$ の時に得られる. 今, $w_j = n_j, \boldsymbol{w} = (w_1, \dots, w_k)$ とし, $P(i, k, \boldsymbol{w})$ を, $\boldsymbol{\theta} \in H$ の下で, $\tilde{\theta}_1, \dots, \tilde{\theta}_k$ の中に i 個の異なるものがある確率とする ($1 \leq i \leq k$) さらに, $\alpha = \sum_{i=1}^k P(i, k, \boldsymbol{w})i, \beta = \sum_{i=1}^k P(i, k, \boldsymbol{w})i^2$ とおく. これに対して Anraku(1999) は $B_2(\boldsymbol{\theta}, \sigma^2) = \sum_{i=1}^k P(i, k, \boldsymbol{w})i + 1 + O(N^{-1})$ よりモデル選択規準として

$$ORIC(\tilde{\boldsymbol{\theta}}, \tilde{\sigma}^2) = l(\tilde{\boldsymbol{\theta}}, \tilde{\sigma}^2) - \alpha - 1$$

を提案した. しかし, $\boldsymbol{\theta} \in H$ の下での $B_2(\boldsymbol{\theta}, \sigma^2)$ をより精密に評価すると

$$B_2(\boldsymbol{\theta}, \sigma^2) = \frac{(\alpha + 2)N^3 - (3\alpha^2 + 2\alpha - \beta)N^2 + \alpha^3 N}{2(N - \alpha)^3} + \frac{1}{2} \sum_{i=1}^k P(i, k, \boldsymbol{w}) \frac{Ni}{N - i - 2} + O(N^{-1})$$

となるので, これをペナルティ一項として用いた方法と先の ORIC との比較を行いたい.

なお, (4) の場合と同様, $B_2(\boldsymbol{\theta}, \sigma^2)$ は, $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}$ とすると, これは AIC のペナルティ項と同じもので, $k+1$ と近似されるが, 正確には $N(k+1)/(N-k-2)$ であり (Sugiura (1978)), これは制約領域での $B_2(\boldsymbol{\theta}, \sigma^2)$ の上限である (Anraku (1999) 参照).

- バイアス項の推定とブートストラップ法について

(1) のバイアス項を直接評価するとする際,

$$\frac{N}{2} \left(\frac{\sigma^2}{\tilde{\sigma}^2} - 1 \right) + \frac{1}{2} \sum_{i=1}^k \frac{n_i (\tilde{\theta}_i - \theta_i)^2}{\tilde{\sigma}^2}$$

の分散が N のオーダーより小さくはないことに注意しなければならない. つまり, $B_2(\boldsymbol{\theta}, \sigma^2)$ をブートストラップ法等により直接推定しようとしても, 分散が大きく, 推定は不安定なものになる. 推定量の漸近正規性が仮定できるような問題に対して Konishi and Kitagawa (1996), Ishiguro, Sakamoto and Kitagawa (1997) はブートストラップ推定で分散を低減する方法 (variance reduction technique) を提案している. \mathbf{x}_{ij} ($j = 1, \dots, n_i$) の経験分布関数を \hat{G}_i とし, これからのブートストラップ標本, それに基づく θ_i ($i = 1, \dots, k$), σ^2 の推定値をそれぞれ, X_{ij}^* , ($j = 1, \dots, n_i$), $\tilde{\theta}_i^*$, $\tilde{\sigma}_*^2$ とする. 分散低減法を用いたバイアス項のブートストラップ推定法をここでの問題に対して形式的に適応すると次のようになる.

$$E_{(\hat{G}_i)} \left[\sum_{i=1}^k \sum_{j=1}^{n_i} \frac{(x_{ij}^* - \tilde{\theta}_i)^2}{2\tilde{\sigma}^2} + \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{(x_{ij} - \tilde{\theta}_i^*)^2}{2\tilde{\sigma}_*^2} - N \right]$$

ただし, 順序制約のあるような最尤推定量は漸近正規性が成り立たず, ブートストラップ法の直接の適用には問題が残る. 真の $\boldsymbol{\theta}$ が領域 $M = \{\boldsymbol{\theta} \in \mathbf{R}^k \mid \theta_1 \leq \dots \leq \theta_k\}$ の境界, 特に, $\boldsymbol{\theta} \in H$ に近い場合, その期待値は $\tilde{\boldsymbol{\theta}}$ のそれとは異なることが確かめられた.

3 数値評価とまとめ

本報告では, 上に挙げたモデル選択法と関連するいくつかの方法をシミュレーションにより比較した. その結果から, おおざっぱに評価すると, 真の $\boldsymbol{\theta}$ が領域の境界に近いときは ORIC が, 境界から離れている場合にはブートストラップ法による方法がそれぞれ良好な結果を与えているように思われる.

ここで調べたブートストラップ法については, 真の $\boldsymbol{\theta}$ が境界の近くにある場合に十分な働きをしないようだが, ブートストラップのサンプリングのやり方などを工夫することにより, 改善の余地はあるかも知れない. 今後の課題としたい.

参考文献

- Anraku, K. (1999). *Biometrika* **86**, 141-152.
- Ishiguro, M., Sakamoto, Y. and Kitagawa, G. (1997). *Ann. Inst. Statist. Math.*, **49**, 411-434.
- Konishi, S. and Kitagawa, G. (1996). *Biometrika* **83**, 875-90.
- Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). New York: Wiley.
- Sugiura, N. (1978). *Comm. Statist.*, **A 7**, 13-26.

Admissibility of the MLE for multinomial distribution with monotone missing data

関東学院大経済学部 布能英一郎

Summary Asano(1965) は、monotone missing であるような多項分布の列を考察し、MLE が exact に解けることを示した。本稿は、この MLE が、自乗損失下で許容的であることを示すものである。更に、この問題の多少の拡張も行う。

主要結果 主要結果を定理として述べる前に、次の例を考察する。

Example 1. $(X_0, X_1, X_2, X_3) \sim Multinomial(N_1, \theta_0, \theta_1, \theta_2, \theta_3)$, $(Y_0, Y_1, Y_2) \sim Multinomial(N_2, \theta'_0, \theta'_1, \theta'_2)$, $(Z_0, Z_1) \sim Binomial(N_3, \theta''_0, \theta''_1)$ 但し $\theta'_i = \theta_i / (\theta_0 + \theta_1 + \theta_2)$, $\theta''_j = \theta'_j / (\theta'_0 + \theta'_1)$. この状況から、第一回目の観測 (X_0, X_1, X_2, X_3) で $X_0 = X_1 = X_2 = 0$ の場合、第二回目の観測 (Y_0, Y_1, Y_2) は行われないものとする。同様に、第二回目までの観測で $X_0 = X_1 = Y_0 = Y_1 = 0$ の場合、第三回目の観測 (Z_0, Z_1) は行われないものとする。 θ_i の MLE を $\hat{\theta}_i$ で表記する。計算により

$$\begin{aligned}\hat{\theta}_0 &= \frac{x_0 + y_0 + z_0}{x_0 + x_1 + y_0 + y_1 + z_0 + z_1} \frac{x_0 + x_1 + y_0 + y_1}{x_0 + x_1 + x_2 + y_0 + y_1 + y_2} \frac{x_0 + x_1 + x_2}{x_0 + x_1 + x_2 + x_3}, \\ \hat{\theta}_1 &= \frac{x_1 + y_1 + z_1}{x_0 + x_1 + y_0 + y_1 + z_0 + z_1} \frac{x_0 + x_1 + y_0 + y_1}{x_0 + x_1 + x_2 + y_0 + y_1 + y_2} \frac{x_0 + x_1 + x_2}{x_0 + x_1 + x_2 + x_3}, \\ \hat{\theta}_2 &= \frac{x_2 + y_2}{x_0 + x_1 + x_2 + N_2} \frac{x_0 + x_1 + x_2}{x_0 + x_1 + x_2 + x_3}, \quad \hat{\theta}_3 = \frac{x_3}{x_0 + x_1 + x_2 + x_3}.\end{aligned}$$

である。この推定量の自乗損失下での許容性を次のようにして示せる

最初に、標本空間 \mathcal{X} は、 $\mathcal{X} = \mathcal{X}(N_1) \cup \mathcal{X}(N_1, N_2) \cup \mathcal{X}(N_1, N_2, N_3)$ であることに注意する。但し $\mathcal{X}(N_1) = \{(0, 0, 0, N_1)\}$, $\mathcal{X}(N_1, N_2) = \{(0, 0, x_2, x_3, 0, 0, N_2) \mid x_3 \leq N_1 - 1, x_2 + x_3 = N_1\}$, $\mathcal{X}(N_1, N_2, N_3) = \{(x_0, x_1, x_2, x_3, y_0, y_1, y_2, z_0, z_1) \mid x_i = 0, 1, \dots, N_1, x_0 + x_1 + x_2 + x_3 = N_1, x_3 \leq N_1 - 1, y_i = 0, 1, \dots, N_2, y_0 + y_1 + y_2 = N_2, z_i = 0, 1, \dots, N_3, z_0 + z_1 = N_3, x_0 + x_1 + y_0 + y_1 + z_0 + z_1 \geq 1\}$.

母数空間 $\Theta = \{(\theta_0, \theta_1, \theta_2, \theta_3) \mid \theta_0 + \theta_1 + \theta_2 + \theta_3 = 1, 0 \leq \theta_i \leq 1\}$ 上に the sequence priors $\{d\tau_i(\theta) \mid i = 1, 2, 3, 4\}$ を

$d\tau_1(\theta) : \theta_i = 1 (i=0,1,2,3)$ にのみ集中する事前確率

$d\tau_2(\theta) : \theta_i + \theta_j = 1, 0 < \theta_i, \theta_j < 1, i \neq j$ にのみ集中する事前確率で、かつ

$$d\tau_2(\theta) \propto (\text{restriction}) \sum_{i \neq j} \frac{d\theta_i}{\theta_i \theta_j}$$

$d\tau_3(\theta) : \theta_i + \theta_j + \theta_k = 1, 0 < \theta_i, \theta_j, \theta_k < 1, i \neq j \neq k$ にのみ集中する事前確率で、かつ

$$d\tau_3(\theta) \propto (\text{restriction}) \sum_{i \neq j \neq k} \frac{d\theta_i d\theta_j}{\theta_i \theta_j \theta_k}$$

$$d\tau_1(\theta) \propto (\text{restriction}) \frac{d\theta_1 d\theta_2 d\theta_3}{\theta_0 \theta_1 \theta_2 \theta_3}$$

を導入する。そうすると、上記の事前確率の列から stepwise Bayes 法によって計算されたベイズ解は、MLE に一致する。stepwise Bayes 法の基本定理 (Meeden and Ghosh, Ann. Stat. 1981) により、許容性が保証される。

この一般化は容易に行え、多項分布が monotone missing の場合、母数の MLE が、自乗損失下で許容的であることが示される。すなわち、次の定理を得る。

定理 1.

$$\text{第 1 回目の観測 } (X_{10}, X_{11}, \dots, X_{1k-1}, X_{1k}) \sim \text{Multinomial}(N_1, \theta_0, \theta_1, \dots, \theta_{k-1}, \theta_k)$$

$$\text{第 2 回目の観測 } (X_{20}, X_{21}, \dots, X_{2k-1}) \sim \text{Multinomial}(N_2, \frac{\theta_0}{\theta_0 + \dots + \theta_{k-1}}, \dots, \frac{\theta_{k-1}}{\theta_0 + \dots + \theta_{k-1}})$$

⋮

$$\text{第 } k-1 \text{ 回目の観測 } (X_{k-10}, X_{k-11}, X_{k-12}) \sim \text{Multinomial}(N_{k-1}, \frac{\theta_0}{\theta_0 + \theta_1 + \theta_2}, \dots, \frac{\theta_2}{\theta_0 + \theta_1 + \theta_2})$$

$$\text{第 } k \text{ 回目の観測 } (X_{k0}, X_{k1}) \sim \text{Binomial}(N_k, \frac{\theta_0}{\theta_0 + \theta_1}, \frac{\theta_1}{\theta_0 + \theta_1})$$

第 1 回目の観測で $X_{10} = X_{11} = \dots = X_{1k-1} = 0$ の場合、第 2 回目の観測は行わない。第 2 回目までの観測で $X_{10} = X_{20} = \dots = X_{1k-2} = X_{2k-2} = 0$ の場合、第 3 回目の観測は行わない。以下、同様の状況を仮定する。そうすると、 $\theta_i, i=0,1,2,\dots,k$ の MLE は、自乗損失下で許容的。

発展課題 定理 1 では、各観測に多項分布を仮定した。しかし、多項分布でなくても定理 1 と同様の結果が得られるような例がある。

Example 2

$$P(x_1, \dots, x_k | \theta_0, \theta_1, \dots, \theta_k)$$

$$= \begin{cases} \frac{(x_1 + \dots + x_k)!}{x_1! \dots x_k!} \theta_0^{\theta_0} \theta_1^{x_1} \theta_2^{x_2} \dots \theta_k^{x_k} & \text{if } x_1 + x_2 + \dots + x_k < n \\ & x_i = 0, 1, \dots, n-1 \text{ for all } i=1,2,\dots,k, \\ \frac{n!}{x_1! \dots x_k!} \theta_1^{x_1} \theta_2^{x_2} \dots \theta_k^{x_k} & \text{if } x_1 + x_2 + \dots + x_k = n \\ & x_i = 0, 1, \dots, n \text{ for all } i=1,2,\dots,k, \end{cases}$$

なる確率分布を $\mathcal{D}(n, \theta_0, \theta_1, \dots, \theta_k)$ と記す。

定理 2. 第 1 回目の観測 $(X_{11}, X_{12}, \dots, X_{1k-1}, X_{1k})$, 第 2 回目の観測 $(X_{21}, X_{22}, \dots, X_{2k-1}), \dots$, 第 $k-1$ 回目の観測 (X_{k-11}, X_{k-12}) , 第 k 回目の観測 X_{k1} に関して、第 1 回目の観測で $X_{11} = \dots = X_{1k-1} = 0$ の場合、第 2 回目の観測は行わない。第 2 回目までの観測で $X_{11} = X_{21} = \dots = X_{1k-2} = X_{2k-2} = 0$ の場合、第 3 回目の観測は行わない。以下、同様の状況を仮定する。更に、確率分布に関して、第 1 回目の観測が $\mathcal{D}(n, \theta_0, \theta_1, \dots, \theta_k)$ に従い、第 j 回目の観測 ($j=2,3,\dots,k$) が $\mathcal{D}(n, \theta_0/(\theta_0 + \dots + \theta_{k-j+1}), \dots, \theta_{k-j+1}/(\theta_0 + \dots + \theta_{k-j+1}))$ に従うならば、 θ_i ($i=0,1,2,\dots,k$) の MLE は、自乗損失下で許容的。

正準相関係数における permutation test と 近似分布について

杉山 高一 (中央大学理工)
山田 智哉 (中央大学理工)

1. はじめに

$x(p \times 1)$, $y(q \times 1)$ を母分散 Σ をもつ確率変数とする. ここで $p \leq q$ とする. x, y の関連性をみる方法の1つとして正準相関分析法がある. しかし正準相関分析を行う場合には変数 p が大きい時には重みベクトルの解釈が難しいという問題点があるように思われる. よって我々は変数 x, y それぞれに対して主成分分析を行い, 主成分スコアを用いた正準相関分析を考える.

今, 変数 X, Y についてそれぞれ別々に主成分分析を行ない, その i 成分を $l_i^{(1)}, l_i^{(2)}$ とし, 固有ベクトルを $h_i^{(1)}, h_i^{(2)}$ とすると

$$\begin{aligned} H_x' S_{xx} H_x &= \text{diag}(l_1^{(1)}, l_2^{(1)} \dots, l_p^{(1)}) \\ H_y' S_{yy} H_y &= \text{diag}(l_1^{(2)}, l_2^{(2)} \dots, l_q^{(2)}) \end{aligned}$$

を満たすここで $H_x = (h_1^{(1)}, \dots, h_p^{(1)})$, $H_y = (h_1^{(2)}, \dots, h_q^{(2)})$ とし, $l_1^{(1)} > l_2^{(1)} > \dots > l_p^{(1)} > 0$, $l_1^{(2)} > l_2^{(2)} > \dots > l_q^{(2)} > 0$ とする. 主成分分析により得た主成分スコアのうち r 成分までの要約した変数を用いて正準相関分析を行うことを考える. 主成分分析法では固有値の大きい順に第一, 第二主成分と得られるので, そのデータの情報が第一, 第二主成分に大きく依存している場合が多い. また2つの集合に対して主成分同士の関係を解析することはそのデータを知る上で重要な問題となる. これらのことから最初の r 成分の主成分スコアを新たにデータとして正準相関分析を行うことを考える. 簡単のため, 2つの集合の最初の2つの主成分の相関を考える. この場合正準相関係数は次のようにして求められる:

$$O' \Lambda_x^{-\frac{1}{2}} H_1' S_{xy} H_2 \Lambda_y^{-\frac{1}{2}} (\Lambda_x^{-\frac{1}{2}} H_1' S_{xy} H_2 \Lambda_y^{-\frac{1}{2}})' O = \begin{pmatrix} r_1^2 & 0 \\ 0 & r_2^2 \end{pmatrix}$$

ここで O はある直交行列とし, $H_1 = (h_1^{(1)} h_2^{(1)})$, $H_2 = (h_1^{(2)} h_2^{(2)})$ とする. もし $r = p$ の時は通常の正準相関分析と変わらないが, $r \leq p$ の時は $r + 1$ 番目以降の主成分が影響されないため, 情報損失が起こり通常の結果よりも精度が悪くなる. しかし第1, 第2主成分の寄与率が高い場合にはこれらの成分により標本が要約されることから意味のあるものとなってくる. そこでどの程度の情報が損失するかを調べるため, 次のような仮説検定を考える.

$$\begin{aligned} H_{10}; & \rho_1^2 = \rho_{10}^{*2} \text{ and } \rho_2^2 = \rho_{20}^{*2} \\ H_{20}; & \rho_1^2 = \rho_{10}^{*2} \\ H_{30}; & \rho_2^2 = \rho_{20}^{*2} \\ H_{40}; & \rho_1^2 + \rho_2^2 = \rho_{10}^{*2} + \rho_{20}^{*2} \end{aligned}$$

これらの帰無仮説の中で特に H_{40} については統計量 $r_1^2 + r_2^2$ は

$$r = \text{tr} \left(\Lambda_x^{-\frac{1}{2}} H_1' S_{xy} H_2 \Lambda_y^{-\frac{1}{2}} (\Lambda_x^{-\frac{1}{2}} H_1' S_{xy} H_2 \Lambda_y^{-\frac{1}{2}})' \right).$$

と表記できることから, この統計量の近似分布を与えることができる. その他の仮説における分布はその統計量自体が複雑であることから, 困難を要することが予想される. そこでこれらの仮説に対してノンパラメトリックな検定方法として Permutation Test を試みる.

2. r の近似分布について

仮説 H_{40} に対する統計量として $r = r_1^2 + r_2^2$ を考え、 r についての極限分布を摂導法を用いて導出する。

Theorem

$r = r_1^2 + r_2^2$ の極限分布は

$$\frac{\sqrt{n}(r - \bar{\rho})}{\sigma^{*2}} \rightarrow \Phi(x)$$

となる。ここで

$$\rho = \sum_{ij} \sigma_{ij}^{xy} \rho_{ij} = \sum_{ij} \frac{\sigma_{ij}^{xy2}}{\lambda_j^{(1)} \lambda_j^{(2)}}$$

$$\sigma^{*2} = f' A_1 f + g' A_2 g + 2f' A_3 g - 4p' A_4 f - 4p' A_5 g + 4p' A_6 p$$

3. Permutation Test

正準相関係数の分布を考える場合、その和の分布の統計量でさえ複雑であることから、極限分布に関しても前節のような複雑な分布になってしまう。正準相関係数 r_1^2 はさらに複雑な統計量であることから近似分布を求めていくことは有効ではないように思われる。そこでこれらの検定問題に対してノンパラ検定の1つである Permutation Test (並べ替え検定) を行う。Permutation Test は対称な分布に対してデータを並べ替え、その新しいデータに対して繰り返し検定を考えていく方法である。相関係数に対する検定も帰無仮説が無相関であるという場合には並べ替えたデータも無相関になることから有用である。本節ではこの Permutation Test について考えるが、我々の考えている問題は正準相関係数がある値であるということに興味があるため、その使い方にも検討を要する。今回我々が提案する手法は得られたデータに対して無相関になるよう変数変換を行い、その後 Permutation Test を行うという手法である。

データとして正準相関スコア u_1, u_1 が得られたとする。例えば第1正準相関係数に対する検定

$$H_{10} : \rho_1 = \rho_{10}$$

を考える場合、帰無仮説の下では u_1 と u_2 との相関係数が ρ_{10} であることから変数変換として

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 1 & \rho_{10} \\ \rho_{10} & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

を考える。このとき新しくできたデータは帰無仮説の下では無相関となることから、これらに対して並べ替えを行いケンドールの順位相関係数などの統計量を用いて検定を行えばよい。ここでは統計量としてはケンドールの順位相関係数を考えているが、場合によってはその他の統計量を考える必要もある。これらの手法をまとめると以下のような手順を行えばよい。

1. 正準相関分析を行い、重みベクトルを与え、それを用いて正準相関スコア u_1, u_2 を計算する。
2. 仮説検定 H_{10} を行う場合、変数変換として $v = Ru$ を行い、新しいスコアを v_1, v_2 とする。
3. v_1, v_2 を並び替えることにより得られたデータ v_1^*, v_2^* を用いてケンドールの τ を計算する。
4. 計算された τ が棄却される出現回数の割合を求めることにより p 値を計算する。

打ち切りデータに基づく生存関数のセミパラメトリック推定

帯広畜産大学畜産学部 鈴木 晶夫 種市 信裕

1. はじめに

ランダム右側打ち切りデータに基づく生存関数の推定問題について考える. n 個の個体の生存時間を表す確率変数 X_1, \dots, X_n が互いに独立で, 未知の分布 $F(x)$ (生存関数 $\bar{F} = 1 - F$) に従うとする. Y_i を第 i 個体の打ち切りまでの時間を表す確率変数とし, Y_1, \dots, Y_n は互いに独立に未知の分布 $G(y)$ (生存関数 $\bar{G} = 1 - G$) に従い, $X_1, \dots, X_n, Y_1, \dots, Y_n$ は互いに独立であると仮定する. ランダム右側打ち切りデータとは, $(X_i, Y_i) \ i = 1, \dots, n$ は直接観測されず, $(Z_i, \delta_i) = (\min(X_i, Y_i), I(X_i \leq Y_i)), \ i = 1, \dots, n$ が観測されたデータである. ただし, $I(A)$ を集合 A の定義関数とする.

観測された $Z_i, \ i = 1, \dots, n$ の異なる値を $z_{(1)} < z_{(2)} < \dots < z_{(k)}$ とし, $j = 1, \dots, k$ に対して, $d_{1j} = \sum_{i=1}^n I(Z_i = z_{(j)}, \delta_i = 1), \ n_j = \sum_{i=1}^n I(Z_i \geq z_{(j)})$ とおく. このとき, F に対するノンパラメトリック最尤推定量は $\bar{F}_n(t) = \prod_{j: z_{(j)} \leq t} (1 - d_{1j}/n_j)$ によって与えられ (Kaplan and Meier (1958)), \bar{F}_n は Kaplan-Meier 推定量 (KM推定量) と呼ばれている. 一方, $\bar{F}(x)$ に対してパラメトリックな分布型 $\bar{F}(x; \beta)$ (密度関数 $f(x; \beta)$) をもつを仮定できる場合には, 尤度関数 $\prod_{i=1}^n \{f(Z_i; \beta)\}^{\delta_i} \{\bar{F}(Z_i; \beta)\}^{1-\delta_i}$ に基づいて未知パラメータ β に対する最尤推定量 (MLE) を求めることができる.

本稿では, 推定すべき生存関数 \bar{F} に対してモデルを仮定するのではなく, 関数

$$m(z) = E[\delta_i | Z_i = z] = \text{pr}(\delta_i = 1 | Z_i = z).$$

に対してパラメトリックモデルを仮定する. 関数 $1 - m(z)$ は観測可能な確率変数 Z を $Z = z$ と固定したもとの打ち切りが起こる条件付き確率であり, $m(z)$ は z に対する打ち切りのパターンを表す関数であると考えられる. そこで, 打ち切りパターン関数 $m(z)$ に対して, 未知の p 次元パラメータ θ をもつパラメトリックモデル $m(z) = m(z; \theta)$ を仮定したうえで, 生存関数 $\bar{F}(z)$ を推定することを考える. 打ち切りデータ (Z_i, δ_i) の確率分布は, 打ち切りパターン関数 $m(z)$ と Z_i の生存関数 $\bar{H}(z)$ によって完全に決定される. したがって, 打ち切りパターン関数に対してのみパラメトリックモデルを仮定したモデル ($\bar{H}(z)$ に対してはモデルを仮定しない) は, セミパラメトリックなモデルである.

2. 生存関数と累積ハザード関数の推定

打ち切りデータ $(Z_i, \delta_i), \ i = 1, \dots, n$ に基づく尤度関数は

$$\prod_{j=1}^k \{m(z_{(j)}; \theta)\}^{d_{1j}} \{1 - m(z_{(j)}; \theta)\}^{d_{0j}} \times \prod_{j=1}^k \{\bar{H}(z_{(j)}-) - \bar{H}(z_{(j)})\}^{d_j} = \mathcal{L}_m(\theta) \times \mathcal{L}_H$$

と分解される. ただし, $d_{0j} = \sum_{i=1}^n I(Z_i = z_{(j)}, \delta_i = 0), \ d_j = d_{0j} + d_{1j}$. したがって, $\mathcal{L}_m(\theta)$ の最大化により θ の MLE $\hat{\theta}$ が得られ, \mathcal{L}_H の最大化により, 経験生存関数 $\bar{H}_n(z) = \prod_{j: z_{(j)} \leq z} (1 - d_j/n_j)$ が得られる. その結果, $\bar{F}(t)$ と \bar{F} に対応する累積ハザード関数 $\Lambda^F(t) = \int_0^t dF(u)/\bar{F}(u-)$ の推定量として,

$$\bar{F}_n^*(t) = \prod_{j: z_{(j)} \leq t} \{1 - m(z_{(j)}; \hat{\theta})d_j/n_j\}, \quad \Lambda_n^{F*}(t) = \sum_{j=1}^k I(z_{(j)} \leq t)m(z_{(j)}; \hat{\theta})d_j/n_j$$

が導出される.

特に, $p = k$, $\theta = (\theta_1, \dots, \theta_k)'$, $\theta_j = m(z_{(j)}; \theta)$, $j = 1, \dots, k$ とすれば, $\hat{\theta} = (d_{11}/d_1, \dots, d_{1k}/d_k)'$ であり, したがって, 推定量 $\bar{F}_n^*(t)$ は KM 推定量 $\bar{F}_n(t)$ に一致する. また, このとき, 推定量 $\Lambda_n^{F^*}(t)$ は, Nelson-Aalen 推定量 (NA 推定量) (Nelson (1972), Aalen (1978)) $\Lambda_n^F(t) = \sum_{j=1}^k I(z_{(j)} \leq t) d_{1j}/n_j$ に一致する. この意味において, 推定量 $\bar{F}_n^*(t)$ と $\Lambda_n^{F^*}(t)$ は, それぞれ KM 推定量と NA 推定量の一般化である.

3. 漸近的性質

定理 1 (Uniform consistency) $\tau = \sup\{z; \bar{H}(z) > 0\}$ とおく.

$$\sup_{0 \leq t < \tau} |\Lambda_n^{F^*}(t) - \Lambda^F(t)| \xrightarrow{P} 0, \quad \sup_{0 \leq t < \tau} |F_n^*(t) - F(t)| \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty$$

定理 2 (Asymptotic normality) 分布 F と G は共通のジャンプをもたないと仮定する. このとき, $t \in [0, \tau)$ に対して, $\sqrt{n}\{\Lambda_n^{F^*}(t) - \Lambda^F(t)\}$ と $\sqrt{n}\{\bar{F}_n^*(t) - \bar{F}(t)\}$ の漸近分布はそれぞれ $N(0, \sigma^{*2}(t))$, $N(0, v^{*2}(t))$ で与えられる. ただし,

$$\begin{aligned} \sigma^{*2}(t) &= \mathbf{W}'(t; \theta_0) \mathbf{I}^{-1}(\theta_0) \mathbf{W}(t; \theta_0) + \int_0^t \frac{\{m(u; \theta_0)\}^2 \{1 - \Delta\Lambda(u)\}}{\bar{H}(u-)} d\Lambda(u), \\ v^{*2}(t) &= \{\bar{F}(t)\}^2 \left[\mathbf{U}'(t; \theta_0) \mathbf{I}^{-1}(\theta_0) \mathbf{U}(t; \theta_0) + \int_0^t \frac{\{m(u; \theta_0)\}^2 \{1 - \Delta\Lambda(u)\}}{\bar{H}(u-)\{1 - m(u; \theta_0)\Delta\Lambda(u)\}^2} d\Lambda(u) \right], \\ \mathbf{I}(\theta) &= \int \{m(u; \theta)\}^{-1} \{1 - m(u; \theta)\}^{-1} \frac{\partial m(u; \theta)}{\partial \theta} \frac{\partial m(u; \theta)}{\partial \theta'} \bar{H}(u-) d\Lambda(u), \\ \mathbf{W}(t; \theta) &= \int_0^t \frac{\partial m(u; \theta)}{\partial \theta} d\Lambda(u), \quad \mathbf{U}(t; \theta) = \int_0^t \frac{\partial m(u; \theta)}{\partial \theta} \{1 - m(u; \theta_0)\Delta\Lambda(u)\}^{-1} d\Lambda(u), \end{aligned}$$

$\Lambda(u) = \int_0^u dH(s)/\bar{H}(s-)$, $\Delta\Lambda(u) = \Lambda(u) - \Lambda(u-)$ であり, θ_0 を打切りパターン関数 $m(u; \theta)$ における真のパラメータベクトルとする.

定理 3 (Asymptotic relative efficiency) 定理 2 と同じ仮定のもとで, 推定量 $\bar{F}_n^*(t)$ の KM 推定量 $\bar{F}_n(t)$ に対する漸近相対効率と, 推定量 $\Lambda_n^{F^*}(t)$ の NA 推定量 $\Lambda_n^F(t)$ に対する漸近相対効率は, いずれも 1 以上である.

参考文献

- Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.*, **6**, 701-726.
- Kaplan, E.L. and Meier, P. (1958). Non-parametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457-481.
- Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, **14**, 945-965.

On the Modified Second Bonferroni's Approximation for the T_{\max}^2 Statistic in the Elliptical Distributions

東京理科大学・理 瀬尾 隆

複数個の楕円母集団の平均ベクトル間の多重比較における対比較について、同時信頼区間を構成する際に必要となる T_{\max}^2 の上側パーセント点の非正規性の影響を考える。

$\mathbf{x}_1^{(i)}, \dots, \mathbf{x}_N^{(i)} (i = 1, \dots, k)$ を母平均ベクトル $\boldsymbol{\mu}^{(i)}$, 母分散共分散行列 $\boldsymbol{\Sigma}^{(i)} = \boldsymbol{\Sigma}$ の楕円分布に従う第 i 母集団からの互いに独立な N 個の観測ベクトルとする。すなわち各母集団の分散共分散行列は等しく、各母集団の標本の大きさも同じと仮定する。

このとき、 k 個の平均ベクトル間の対比較に対する同時信頼区間は、次のように書くことができる。

$$\mathbf{a}'(\boldsymbol{\mu}^{(\ell)} - \boldsymbol{\mu}^{(m)}) \in \left[\mathbf{a}'(\bar{\mathbf{x}}^{(\ell)} - \bar{\mathbf{x}}^{(m)}) \pm t\sqrt{2\mathbf{a}'\mathbf{S}\mathbf{a}/N} \right], \forall \mathbf{a} \in \mathbb{R}^p - \{\mathbf{0}\}, 1 \leq \ell < m \leq k.$$

ここに、 $\bar{\mathbf{x}}^{(i)} = (1/N) \sum_{j=1}^N \mathbf{x}_j^{(i)}$, $\mathbf{S} = (1/k) \sum_{i=1}^k \mathbf{S}^{(i)}$, $\mathbf{S}^{(i)} = \sum_{j=1}^N (\mathbf{x}_j^{(i)} - \bar{\mathbf{x}}^{(i)})(\mathbf{x}_j^{(i)} - \bar{\mathbf{x}}^{(i)})' / (N - 1)$. 実際にこの同時信頼区間を求めるためには、 $t > 0$ を与える必要があり、この t は次のような最大値統計量の上側パーセント点である。

$$\Pr\left\{ \max_{1 \leq \ell < m \leq k} T_{\ell m}^2 > t^2 \right\} = \alpha.$$

ただし、 $T_{\ell m}^2 = \frac{1}{2}(\mathbf{z}^{(\ell)} - \mathbf{z}^{(m)})' \mathbf{S}^{-1}(\mathbf{z}^{(\ell)} - \mathbf{z}^{(m)})$, $\mathbf{z}^{(\ell)} = \sqrt{N}(\bar{\mathbf{x}}^{(\ell)} - \boldsymbol{\mu}^{(\ell)})$.

一般にこの t の正確な値を与えることは、正規母集団の下でさえ非常に困難であり、正規母集団の下では、ボンフェロニ不等式による近似法（いわゆるボンフェロニ不等式の第1項を利用した方法、すなわち F 分布のパーセント点を用いる方法）とさらに第2項までを考慮した修正2次近似法が提案され、漸近展開近似による精度のよい近似値が具体的に与えられ、近似値の評価がモンテカルロシミュレーションにより調べられている (Siotani(1959), Seo and Siotani(1992), Seo(1995) 参照).

本報告では、母集団が非正規分布の一種である楕円分布に従うという仮定の下で、修正2次近似法による近似値を摂動展開法を用いて漸近展開の形で導出し、非正規性の影響について考えている。

$\mathbf{y}_1 = \mathbf{z}^{(1)} - \mathbf{z}^{(2)}, \mathbf{y}_2 = \mathbf{z}^{(1)} - \mathbf{z}^{(3)}, \dots, \mathbf{y}_M = \mathbf{z}^{(k-1)} - \mathbf{z}^{(k)}$ (ただし, $M = k(k-1)/2$) とすると、ボンフェロニの不等式は、

$$\sum_{i=1}^M \Pr\left\{\frac{1}{2}\mathbf{y}'_i \mathbf{S}^{-1} \mathbf{y}_i > t^2\right\} - \beta(t^2) < \Pr\left\{\max_{1 \leq \ell < m \leq k} T_{\ell m}^2 > t^2\right\} < \sum_{i=1}^M \Pr\left\{\frac{1}{2}\mathbf{y}'_i \mathbf{S}^{-1} \mathbf{y}_i > t^2\right\},$$

ここに、

$$\beta(t^2) = \sum_{i < j} \Pr\left\{\frac{1}{2}\mathbf{y}'_i \mathbf{S}^{-1} \mathbf{y}_i > t^2, \frac{1}{2}\mathbf{y}'_j \mathbf{S}^{-1} \mathbf{y}_j > t^2\right\}$$

であり、修正2次近似値は、

$$\sum_{i=1}^M \Pr\left\{\frac{1}{2}\mathbf{y}'_i \mathbf{S}^{-1} \mathbf{y}_i > t_2^{*2}\right\} - \beta(t_1^2) = \alpha$$

を満たす t_2^{*2} のことである。ここに、 t_1^2 は、 $\sum_{i=1}^M \Pr\left\{\frac{1}{2}\mathbf{y}'_i \mathbf{S}^{-1} \mathbf{y}_i > t_1^2\right\} = \alpha$ を満たすものである。 $\beta(t_1^2)$ については、 $\gamma_1(t_1^2) = \Pr\{T_{ij}^2 > t_1^2, T_{kl}^2 > t_1^2\}$ ($i \neq j \neq k \neq l$), $\gamma_2(t_1^2) = \Pr\{T_{ij}^2 > t_1^2, T_{ik}^2 > t_1^2\}$ ($i \neq j \neq k$) の2つのケースの同時確率を求めることが必要であり、この同時確率の漸近展開を摂動法により求め、非正規性の影響を調べている。さらに、いくつかのケースについて、モンテカルロシミュレーションを行い、得られた近似値との比較を行っている。

参考文献

- [1] Seo, T. (1995), "Simultaneous Confidence Procedures for Multiple Comparisons of Mean Vectors in Multivariate Normal Populations," *Hiroshima Mathematical Journal*, **25**, 387-422.
- [2] Seo, T. and Siotani, M. (1992), "The Multivariate Studentized Range and Its Upper Percentiles," *Journal of the Japan Statistical Society*, **22**, 123-137.
- [3] Siotani, M. (1959), "The extreme value of the generalized distances of the individual points in the multivariate normal sample," *Annals of the Institute of Statistical Mathematics*, **10**, 183-208.

球面上の統計解析

大阪女子大学・理 綿森 葉子

1 はじめに

標本空間が球面（ここで扱う球面とは、特に断らない限り円周も含む一般 p 次元の超球面 S^p のことをさす。また、通常の2次元球面のことを S^2 、円周のことを S^1 であらわすことにする。）の場合、通常の統計的手法を適用したのでは問題がおこる場合がある。あるいは、これまでに用いてきた統計手法をそのままでは適用できない場合もある。ここでは、通常の p 次元ユークリッド空間 (\mathbb{R}^p) での手法と対比しながら球面の性質を考察してみる。

2 座標

\mathbb{R}^p には、自然に座標が入っていて、それ以外に特に統計上考慮の対象となるような座標系はない。しかし、 S^p にもっとも自然と思われる座標は角度であろうが、これはいろいろな点でやっかいである。角度を極座標変換により $\mathbb{R}^{(p+1)}$ の元とみなせば線形性は保てるが、ノルムの制約が入り成分間の独立性が最初から崩れてしまう。また、次元が本来の次元より1つ上がってしまうことも問題である。ここで極座標表示は

$$x_j = \cos \theta_j \prod_{k=0}^{j-1} \sin \theta_k, \quad j = 1, \dots, p, \quad \sin \theta_0 = \cos \theta_p = 1,$$

$$0 \leq \theta_j \leq \pi, \quad j = 1, \dots, p-2, \quad 0 \leq \theta_{p-1} \leq 2\pi$$

であり、ヤコビアンは $J_p = \prod_{j=2}^{p-1} \sin^{p-1} \theta_{j-1}$, $J_2 = 1$ である。ノルムの制約から導かれるモーメント制約は、 Σ を分散共分散行列、 μ を平均ベクトルとするとき

$$\text{trace} \Sigma + \|\mu\| = 1$$

となる。ここでは、角度を対象とするときは θ を用い、 \mathbb{R}^p の元とみるときは \mathbf{x} を用いることにする。

3 分布の特性量

いったん確率変数と確率測度が定義されれば、母集団平均、分散を含むモーメントは Ω が球面如何に関わらず定義される。それに対応する座標系の下で得られたデータに対する標本平均や分散も同様である。しかし、 Ω が \mathbb{R}^p のときと、 Ω が

S^p のときでは、その意味、解釈が異なってくる。例えば、 \mathbb{R}^p での平均は位置の中心であり重心であった。 S^p の極座標をもとに得た平均は、一般に S^p の外にありもはや重心の意味を持ち得ない。つまり量的な中心と位置的な中心は別の概念として捉えなければならない。ここに至って「中心」のさすところを再度考え直さなければならなくなる。位置的な中心の1つとして

$$\min_{\nu \in S^p} E[\rho(\Theta, \nu)^2]$$

をみたく ν を提案し、特に S^1 のときの性質を調べた。ここで、 $\rho(\Theta, \nu)$ は適当な距離関数である。

4 相関と回帰

極座標を用いて \mathbb{R}^p の特異な分布として捉えた場合、相関係数はそのまま定義できるが、回帰関数はそうはいかない。特異性の \mathbf{x} のノルム制約が大きく影響する。では、 Θ をもとにしてはどうか？今度は新しい概念を導入する必要に迫られる。 S^1 上では例えば Fisher&Lee(1983)

$$\Phi = \Theta + \theta_0 \pmod{2\pi}$$

または

$$\Phi = -\Theta + \theta_0 \pmod{2\pi}$$

なる T-linear association が定義されている。これに基づいて

$$\rho_T = \frac{E[\{\sin(\Theta_1 - \Theta_2) \sin(\Phi_1 - \Phi_2)\}]}{[E\{\sin^2(\Theta_1 - \Theta_2)\} E\{\sin^2(\Phi_1 - \Phi_2)\}]^{1/2}}$$

という T-相関係数を定義し研究している。これは一般の球面上では

$$\rho_{FL} = \frac{\det E[\mathbf{xy}']}{\det E[\mathbf{xx}'] \det E[\mathbf{yy}]}$$

に拡張される。この他にもいくつか提案されているが、それぞれ長所短所があつて決めてになるものは今のところない。

回帰に関しては、 S^1 では例えば

$$\mu_i = \mu + g(\beta x_i), \quad x_i \in [-1, 1]$$

とおいて関数 g の選び方や β について研究されているが、統一的な扱いはなされていないようである。

Model misspecification のもとでの統計量の漸近展開について

一橋大学・経済 早川 毅

1. 序

$X = (x_1, \dots, x_n)$ はある未知の確率密度関数 $h(x)$ を持つ母集団からの random sample とする。しかし、これらを parametric model $\{f(x; \underline{\theta}); \underline{\theta} \in R^p\}$ からの sample として取り扱うとする。

Random sample X にもとづく $\underline{\theta}$ の対数尤度関数を $l(\underline{\theta}) = \sum_{\alpha=1}^n \log f(x_\alpha; \underline{\theta})$ とする。

Fraser の情報量は

$$F(\underline{\theta}) = \int \log f(x; \underline{\theta}) h(x) dx$$

として定義され、 $F(\underline{\theta})$ を最大にする $\underline{\theta}$ を $\underline{\theta}_0 = \underline{\theta}_0(h)$ と記す。このとき

$$\left. \frac{\partial F(\underline{\theta})}{\partial \underline{\theta}} \right|_{\underline{\theta}_0} = \underline{0}$$

とする。

Kent (1982), Viraswami and Reid (1996) は仮説 $H : \underline{\theta} = \underline{\theta}_0$ について尤度比規準、Rao の Score 統計量、Wald 統計量を考察した。そして、Model misspecification のもとで各統計量の極限分布が χ^2 -分布となる robust 統計量を提案し、後者はその分布の漸近展開を与えている。

2. 問題

k 個の母集団の確率密度関数を $h_i(x)$, $i = 1, 2, \dots, k$ とし、これらを parametric model $\{f(x; \underline{\theta}^{(i)}), \underline{\theta}^{(i)} \in R^p\}$, $i = 1, 2, \dots, k$ であるとして取り扱う。

取り扱いたい検定問題は

$$H : h_1(x) = \dots = h_k(x) (\equiv h(x)), \quad K : \exists i, j, h_i(x) \neq h_j(x)$$

であるが、この問題を次の形で取り扱う。

$$H_0 : \underline{\theta}^{(1)} = \dots = \underline{\theta}^{(k)} (\equiv \underline{\theta}), \quad K_0 : \exists i, j, \underline{\theta}^{(i)} \neq \underline{\theta}^{(j)}.$$

この問題に対する Robust Rao Score 統計量は

$$R = \sum_{i=1}^k y_i(\underline{\theta})' M_{\cdot\cdot}^{-1}(\underline{\theta}) y_i(\underline{\theta})$$

が提案される。

ここで、

$$y_i(\underline{\theta}^{(i)}) = \frac{1}{\sqrt{n_i}} \sum_{\alpha=1}^{n_i} \frac{\partial \log f(x_\alpha^{(i)}; \underline{\theta}^{(i)})}{\partial \underline{\theta}^{(i)}}, \quad i = 1, 2, \dots, k,$$

$$M_{\cdot,\cdot}(\underline{\theta}) = (m_{\alpha,\beta}(\underline{\theta})),$$

$$m_{\alpha,\beta}(\underline{\theta}) = \int \frac{\partial \log f(x; \underline{\theta})}{\partial \theta_\alpha} \frac{\partial \log f(x; \underline{\theta})}{\partial \theta_\beta} h(x) dx,$$

$\underline{\theta} : H_0$ のもとでの $n = \sum_{i=1}^k n_i$ 個の random sample による $\underline{\theta}$ の最尤推定量である。

3. Rao の分布の漸近展開

Robust Rao Score 統計量 R の分布は $\theta = \theta_0$ のもとで Hayakawa (1977, 1987) と同様にし
て次の様に展開される。

$$P\{R \leq x\} = P_f + \frac{1}{n} \sum_{\alpha=0}^3 a_\alpha P_{f+2\alpha} + o\left(\frac{1}{n}\right),$$

$$a_3 = A_3, \quad a_2 = -3A_3 + A_2, \quad a_1 = 3A_3 - 2A_2 + A_1,$$

$$a_0 = -A_3 + A_2 - A_1, \quad f = p(k-1),$$

$$A_3 = \frac{1}{24} \sum_{i=1}^k \frac{1}{\rho_i} (1 - \rho_i)^2 \left\{ 3M_{\cdot,\cdot,\cdot,\cdot} \circ M_{\cdot,\cdot}^{-1} \circ M_{\cdot,\cdot}^{-1} (M_{\cdot,\cdot,\cdot,\cdot} \circ M_{\cdot,\cdot}^{-1}) \right. \\ \left. + 2M_{\cdot,\cdot,\cdot,\cdot} (*M_{\cdot,\cdot}^{-1})^3 * M_{\cdot,\cdot,\cdot,\cdot} \right\}$$

$$- \frac{1}{8} (k-1)^2 M_{\cdot,\cdot,\cdot,\cdot} \circ M_{\cdot,\cdot}^{-1} \circ M_{\cdot,\cdot}^{-1} (M_{\cdot,\cdot,\cdot,\cdot} \circ M_{\cdot,\cdot}^{-1})$$

$$- \frac{1}{12} (k-1) M_{\cdot,\cdot,\cdot,\cdot} (*M_{\cdot,\cdot}^{-1})^3 * M_{\cdot,\cdot,\cdot,\cdot},$$

$$A_2 = \frac{1}{8} \sum_{i=1}^k \frac{1}{\rho_i} (1 - \rho_i)^2 \left\{ M_{\cdot,\cdot,\cdot,\cdot} (\circ M_{\cdot,\cdot}^{-1})^2 - p(p+2) \right\},$$

$$A_1 = \frac{1}{2} (k-1) \left\{ 2M_{\cdot,\cdot,\cdot,\cdot} * M_{\cdot,\cdot}^{-1} * (M_{\cdot,\cdot}^{-1} M_{\cdot,\cdot} M_{\cdot,\cdot}^{-1}) * M_{\cdot,\cdot}^{-1} * M_{\cdot,\cdot,\cdot,\cdot} \right. \\ - 2M_{\cdot,\cdot,\cdot,\cdot} * M_{\cdot,\cdot} * (M_{\cdot,\cdot}^{-1} M_{\cdot,\cdot} M_{\cdot,\cdot}^{-1}) * M_{\cdot,\cdot}^{-1} * M_{\cdot,\cdot,\cdot,\cdot} \\ \left. - 2M_{\cdot,\cdot,\cdot,\cdot} (*M_{\cdot,\cdot}^{-1})^2 * M_{\cdot,\cdot}^{-1} * M_{\cdot,\cdot,\cdot,\cdot} + p - 2M_{\cdot,\cdot,\cdot,\cdot} \otimes M_{\cdot,\cdot}^{-1} \otimes M_{\cdot,\cdot}^{-1} \right\}.$$

References

- [1] Fraser, D.A.S. (1965). *Ann. Math. Statist.*, **36**, 890-896.
- [2] Hayakawa, T. (1977). *Ann. Inst. Statist. Math.*, **29**, 359-378.
(Correction: (1987). *Ann. Inst. Statist. Math.*, **39**, 681.)
- [3] Kent, J.T. (1982). *Biometrika*, **69**, 19-27.
- [4] Virsawami, K. and Reid, N. (1996). *The Canadian J. Statist.*, **24**, 263-278.
- [5] Virsawami, K. and Reid, N. (1997). To appear in *Special Issue of J. Statist. Plan. Inf.*

完全な単純構造に近い場合における ある因子数決定方式の適用条件について

広島大学 工学部 伊藤 雅明
広島大学 工学部 佐藤 学

1 はじめに

主成分分析は、本来は因子分析とは異なる目的のための手法であるが、因子分析的に用いられることが多い。そして相関行列を解析するとき、「1より大きい固有根の個数」を“因子数”とする方法を採用することが多い。そこで、この方法の妥当性を議論したい。

このため、 p 次元データ \mathbf{x} が因子分析の模型 $\mathbf{x} = \boldsymbol{\mu} + \Lambda \mathbf{f} + \mathbf{u}$ に従うと仮定する。ただし Λ は p 行 k 列の因子負荷行列である。 \mathbf{x} の母分散共分散行列 Σ は、 $E\{\mathbf{f}\} = \mathbf{0}$, $E\{\mathbf{u}\} = \mathbf{0}$, $E\{\mathbf{f}\mathbf{u}'\} = \mathbf{0}$, $E\{\mathbf{u}\mathbf{u}'\} = \Psi$ (対角行列) という仮定のもとで、 $\Sigma = \Lambda\Lambda' + \Psi$ となる。 $D = \text{diag } \Sigma$ とおき、 $D^{-1/2}\Lambda$, $D^{-1/2}\Psi D^{-1/2}$ を改めて Λ , Ψ とそれぞれ書くことにすると、母相関行列 P は $P = \Lambda\Lambda' + \Psi$ と表される。以下の議論では、 Λ は一意性条件を満たし $\text{diag}(I - \Lambda\Lambda')$ が正定値になるように与え、 $\Psi = \text{diag}(I - \Lambda\Lambda')$ として P を定める。 P の固有根を $\theta_1 \geq \dots \geq \theta_p$ とする。

実際のデータ解析で潜在構造を検証するときは、完全な単純構造 (各変量がただ1つの因子にしか影響しない形) を想定する場合が多い。その場合には、 P の1より大きい固有根の個数は k 個であることが示されている (Sato(1990))。しかし現実には各変量はその他の因子にも若干は影響しているので、完全な単純構造にはならない。そこで、ここでは完全な単純構造では0であると想定していた要素が0とは限らないときを扱う。

2 $k = 2$ で完全な単純構造に近い形

$k = 2$ で完全な単純構造に近い形における結果を得るために必要な Lemmas を初めに述べよう。

Lemma 1 (1より大きい固有根の個数の上限; Sato(1992) の Theorem 6.1)

$k = 2$ のとき、 P の1より大きい固有根の個数はたかだか2個である。 ■

Lemma 2 (P の第2固有根が1より大きいための必要十分条件)

$k = 2$ のとき、 P の固有多項式を $f(\theta)$ とする。 $\theta_2 > 1$ であるための必要十分条件は、 p が偶数のとき $f(1) > 0$, p が奇数のとき $f(1) < 0$ である。 ■

Lemma 3

$$\Lambda^0 = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{p_1 1} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \lambda_{p_1+1 2} & \cdots & \lambda_{p_1+p_2 2} \end{pmatrix}', \quad P^0 = \Lambda^0 \Lambda^{0'} + \text{diag}(I_p - \Lambda^0 \Lambda^{0'}) = \begin{pmatrix} P_{11}^0 & P_{12}^0 \\ P_{21}^0 & P_{22}^0 \end{pmatrix},$$

$$\Lambda = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{p_1 1} & \delta_{p_1+1 1} & \cdots & \delta_{p_1+p_2 1} \\ \delta_{12} & \cdots & \delta_{p_1 2} & \lambda_{p_1+1 2} & \cdots & \lambda_{p_1+p_2 2} \end{pmatrix}', \quad P = \Lambda\Lambda' + \text{diag}(I - \Lambda\Lambda')$$

$p_1 \geq 3$, $p_2 \geq 3$, $p = p_1 + p_2$, $\lambda_{i1}^2 + \delta_{i2}^2 < 1$ ($i = 1, \dots, p_1$), $\delta_{\ell 1}^2 + \lambda_{\ell 2}^2 < 1$ ($\ell = p_1 + 1, \dots, p_1 + p_2$) を仮定し、さらに一般性を失うことなく $\lambda_{i1} > 0$ ($i = 1, \dots, p_1$), $\lambda_{\ell 2} > 0$ ($\ell = p_1 + 1, \dots, p_1 + p_2$) が成立つと仮定する。

$$f(\theta) = \det|P - \theta I_p|, \quad \epsilon = \max_{i, \ell} \left\{ \frac{|\delta_{i2}|}{\lambda_{i1}}, \frac{|\delta_{\ell 1}|}{\lambda_{\ell 2}} \right\}, \quad I_q \text{ を } q \text{ 次の単位行列とおくと,}$$

$$f(1) = \det|P_{11}^0 - I_{p_1}| \det|P_{22}^0 - I_{p_2}| (1 + O(\epsilon^2)) \text{ である。} \quad \blacksquare$$

Lemmas 1, 2, 3 を使うと次の Theorem が得られる。

Theorem 1 ($k=2$ で完全な単純構造に近い場合)

Lemma 3における P の 1 より大きい固有根の個数は, $\epsilon \ll 1$ のとき 2 個である. ■

3 $k \geq 2$ で実際上しばしば出会う構造

$k \geq 2$ に対して, 実際の場合でしばしば出会う次の 2 つの構造を考察する.

初めに, $p = 4k$ で完全な単純構造に近くて要素に特別な関係がある場合を扱う.

Theorem 2 ($p = 4k$ で完全な単純構造に近くて要素に特別な関係がある場合)

$$\Lambda = \begin{pmatrix} \lambda & \lambda & \lambda & \lambda & \delta & \delta & -\delta & -\delta & \cdots & \delta & \delta & -\delta & -\delta \\ \delta & \delta & -\delta & -\delta & \lambda & \lambda & \lambda & \lambda & \cdots & \delta & \delta & -\delta & -\delta \\ \vdots & & & & & & & & \ddots & \ddots & & & \vdots \\ \delta & \delta & -\delta & -\delta & \delta & \delta & -\delta & -\delta & \cdots & \lambda & \lambda & \lambda & \lambda \end{pmatrix}'$$

$p = 4k, k \geq 2, \lambda^2 + (k-1)\delta^2 < 1, 0 \leq \delta < \lambda < 1, P = \Lambda\Lambda' + \text{diag}(I - \Lambda\Lambda')$ とする. この前提のもとで, P の 1 より大きい固有根の個数が k 個であるのは以下のときに限る.

(1) $k = 2, \dots, 8$ に対して, 前提条件を満たす任意の λ, δ .

(2) $k \geq 9$ に対して

(2-1) $\lambda < \frac{1}{2}\sqrt{1 - \frac{3}{k-2}}$ ($\equiv \lambda_1$) のとき, $\delta < \sqrt{\frac{3}{k-5}}\lambda$.

(2-2) $\lambda_1 \leq \lambda$ のとき, 前提条件を満たす任意の δ . ■

次に, $p = 3k$ で完全な単純構造に近くて要素に特別な関係がある場合を扱う.

Theorem 3 ($p = 3k$ で完全な単純構造に近くて要素に特別な関係がある場合
—主要な因子以外の負荷がすべて正のとき—)

$$\Lambda = \begin{pmatrix} \lambda & \lambda & \lambda & \delta & \delta & \delta & \cdots & \cdots & \delta & \delta & \delta \\ \delta & \delta & \delta & \lambda & \lambda & \lambda & \delta & \delta & \delta & \cdots & \delta & \delta & \delta \\ \vdots & & & & & & & \ddots & \ddots & & & & \vdots \\ \delta & \delta & \delta & \delta & \delta & \delta & \cdots & \cdots & \lambda & \lambda & \lambda \end{pmatrix}'$$

$p = 3k, k \geq 2, \lambda^2 + (k-1)\delta^2 < 1, 0 \leq \delta < \lambda < 1, P = \Lambda\Lambda' + \text{diag}(I - \Lambda\Lambda')$ とする. この前提のもとで, P の 1 より大きい固有根の個数が k 個であるのは以下のときに限る.

(1) $\lambda < \left(1 + \frac{4(k-1)}{(3 + \sqrt{1+2k})^2}\right)^{-1/2}$ ($\equiv \lambda_2$) のとき, $\delta < \frac{2}{3 + \sqrt{1+2k}}\lambda$.

(2) $\lambda_2 \leq \lambda$ のとき, 前提条件を満たす任意の λ, δ . ■

Theorem 4 ($p = 3k$ で完全な単純構造に近くて要素に特別な関係がある場合
—主要な因子以外の負荷がすべて負のとき—)

Theorem 3 と同じ形の Λ に対して, $p = 3k, k \geq 2, \lambda^2 + (k-1)\delta^2 < 1, -1 < -\lambda < \delta \leq 0, P = \Lambda\Lambda' + \text{diag}(I - \Lambda\Lambda')$ とする. この前提のもとで, P の 1 より大きい固有根の個数が k 個である必要十分条件を与えることができる. ■

結果の詳細は当日発表した.

次元縮小回帰モデルにおける次元の推定

長岡 創志 (広島大・理)
藤越 康祝 (広島大・理)

1. はじめに

目的変数 y と説明変数 $\mathbf{x} = (x_1, \dots, x_p)'$ との間の非線形構造を探るための次元縮小回帰モデル (Li[3]) は

$$y = f(\beta_1' \mathbf{x}, \dots, \beta_K' \mathbf{x}, \varepsilon) \quad (1)$$

で定義される。ここに、 f は未知関数、 β_1, \dots, β_K は独立な未知ベクトル、 ε は誤差項で \mathbf{x} と独立である。

モデル (1) の β_1, \dots, β_K は有効次元縮小方向と呼ばれ、また、これらによって張られる空間 $B_K (= \mathcal{R}[\beta_1, \dots, \beta_K])$ を有効次元縮小空間という。本報告では、有効次元縮小空間の次元 K の推定法に関して、正準判別分析との類似性に基づいた予測的観点からの基準を提案する。

2. 逆回帰部分空間

有効次元縮小方向は、中心化逆回帰曲線 $\eta(y)$ とその共分散行列 $\Sigma_{\eta(y)}$ 、すなわち $\eta(y) = E(\mathbf{x}|y) - E(\mathbf{x})$ 、 $\Sigma_{\eta(y)} = \text{Var}(\eta(y))$ とに関係しているが、特に逆回帰部分空間 $S_E(\mathbf{x}|y) = \mathcal{R}[\eta(y)|y \in \mathcal{A}]$ が重要となる。ここに、 \mathcal{A} は目的変数 y の標本空間である。以下においては、説明変数 \mathbf{x} の分布は楕円分布とし、かつ線形計画条件と呼ばれる次の条件を満たしていると仮定する。

任意の p 次元定数ベクトル \mathbf{b} に対して

$$E(\mathbf{b}' \mathbf{x} | \beta_1' \mathbf{x}, \dots, \beta_K' \mathbf{x}) = c_0 + c_1 \beta_1' \mathbf{x} + \dots + c_K \beta_K' \mathbf{x} \quad (2)$$

を満たす定数 c_0, c_1, \dots, c_K が存在する。

定理 1 (1), (2) の下で、

$$S_E(\mathbf{x}|y) \subset \mathcal{R}[\Sigma_x \beta_1, \dots, \Sigma_x \beta_K]$$

が成り立つ。ここで、 Σ_x は \mathbf{x} の共分散行列である。

この結果より、次のことが言える。 $\Sigma_{\eta(y)}$ の Σ_x に関する一般固有値、固有ベクトルを $\lambda_1, \dots, \lambda_p, \gamma_1, \dots, \gamma_p$ とする。すなわち、

$$\Sigma_{\eta(y)} \gamma_i = \lambda_i \Sigma_x \gamma_i, \quad \gamma_i \Sigma_x \gamma_j = \delta_{ij} \quad (3)$$

とする ($\lambda_1 \geq \dots \geq \lambda_p \geq 0$)。このとき、定理 1 より、0 でない固有値 λ_j に対応する固有ベクトルに対して $\gamma_j \in \mathcal{R}[\beta_1, \dots, \beta_K]$ となる。(3) の固有値について、 $\lambda_1 \geq \dots \geq \lambda_K > \lambda_{K+1} = \dots = \lambda_p = 0$ の場合には、有効次元縮小空間の次元は一般固有値問題 (3) のゼロでない固有値の数、有効次元縮小空間はゼロでない固有値に対応する固有ベクトルの張る空間となる。

3. 逆回帰推定法 (SIR 法)

有効次元縮小空間の基底として、固有ベクトル $\gamma_1, \dots, \gamma_K$ をとることができることから、 Σ_x および $\Sigma_{\eta(y)}$ の推定を考えることによって固有ベクトルの推定量が構成される。以下の推定法は、Li[3] によって提案された。

変数 (y, \mathbf{x}) について大きさ n のデータ $(y_i, \mathbf{x}_i), (i = 1, \dots, n)$ が与えられているとする。 y の値を H 個のスライス I_1, \dots, I_H に分け、スライス I_h に属する観測値 \mathbf{x}_i の平均を求める。すなわち、 $\bar{\mathbf{x}}^{(h)} = \frac{1}{n_h} \sum_{y_i \in I_h} \mathbf{x}_i$ である。ただし、 n_h はスライス I_h に属する標本数である。このとき、 Σ_x は通常の標本分散行列 $\hat{\Sigma}_x = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$ で推定することができ、 $\Sigma_{\eta(y)}$ は $\hat{\Sigma}_{\eta(y)} = \sum_{h=1}^H \frac{n_h}{n} (\bar{\mathbf{x}}^{(h)} - \bar{\mathbf{x}})(\bar{\mathbf{x}}^{(h)} - \bar{\mathbf{x}})'$ で推定することができる。よって、 $\lambda_i, \gamma_i (i = 1, \dots, K)$ は (3) の $\Sigma_x, \Sigma_{\eta(y)}$ をそれぞれ $\hat{\Sigma}_x, \hat{\Sigma}_{\eta(y)}$ で置き換えた一般固有値問題： $\hat{\Sigma}_{\eta(y)} \hat{\gamma}_i = \hat{\lambda}_i \hat{\Sigma}_x \hat{\gamma}_i, \quad \hat{\gamma}_i \hat{\Sigma}_x \hat{\gamma}_j = \delta_{ij}$ の解として推定できる。

4. 次元 K の推定

4.1 推定問題

有効次元縮小空間の次元のとりうる値は, $0, 1, \dots, p$ の $p+1$ 通りなので, これと対応させて $p+1$ 個のモデル $M_K: y = f(\beta'_1 x, \dots, \beta'_K x, \varepsilon)$ を考える. 一般に, モデル M_K が正しいとき, 0 でない固有値 λ_j の数は K 個以下であって, モデル M_K と 0 でない固有値 λ_j の数がちょうど K 個であるというモデル $\tilde{M}_K: \lambda_1 \geq \dots \geq \lambda_K > \lambda_{K+1} = \dots = \lambda_p = 0$ とは必ずしも同値ではない. 簡単のため, 後者のモデル選択問題として考察されているが, M_K と \tilde{M}_K との違いも考慮する必要がある.

4.2 逐次検定法による推定法

Li[3] では, x が正規分布に従うとの仮定の下で, 仮説 \tilde{M}_K を逐次検定して次元を推定する方法が提案されている. 検定量としては漸近的に自由度 $(p-K)(H-K-1)$ の χ^2 分布に従う $n(\hat{\lambda}_{K+1} + \dots + \hat{\lambda}_p)$ を用いる. この方法は, まず仮説 $\tilde{M}_0: \lambda_1 = \dots = \lambda_p = 0$ を検定し, それが受容されれば $K=0$ と推定し, 棄却されれば次に仮説 \tilde{M}_1 を検定する. 以下, 受容されるまで逐次的に検定を続け, 仮説 \tilde{M}_K が受容されれば, 次元が K であると推定する. この方法には, 予測的観点からの推定法ではなく, また, 有意水準の設定をどのようにすべきかという逐次検定方式の問題点がある. Schott[5] は x が楕円分布に従う場合を考察している.

4.3 有効次元縮小方向の推定の良さに基づく推定法

Ferré[1] では, 有効次元縮小空間とその推定量の空間との近さを測る, という観点から得られた基準量を提案している. モデル (1) のもとで, 大きい方の $q (\leq K)$ 個の固有値に基づく有効次元縮小空間およびその推定空間の基底行列を $B_q = [\gamma_1, \dots, \gamma_q]$, $\hat{B}_q = [\hat{\gamma}_1, \dots, \hat{\gamma}_q]$ とする. これら 2 つの空間の近さを変数 $B'_q x$ と変数 $\hat{B}'_q x$ との正準相関係数の 2 乗和の平均 $r(q) = \frac{1}{q} \text{tr}\{(B'_q \Sigma B_q)^{-1} (B'_q \Sigma \hat{B}_q) (\hat{B}'_q \Sigma \hat{B}_q)^{-1}\} (\hat{B}'_q \Sigma B_q)$ で測る. Ferré[1] は, 楕円分布のもとで, $R(q) = E[r(q)]$ を漸近的に評価し, その推定量 $\hat{R}(q)$ を次元の推定量として提案している. 実際には, 各 K について, $\hat{R}(1), \dots, \hat{R}(K)$ を求め, 最大を与える q が有効次元縮小空間の次元として推定される.

4.4 正準判別分析との類似性に基づく推定法

有効次元縮小方向の推定においては, x についての標本が y がどのスライスに属するかによって H 個の群に分けられる. このとき, $\Sigma_{\eta(y)}$ は群間平方和積和行列であり, $(n-1)\Sigma_x$ は全平方和積和行列である. 従って, 正準判別分析における次元の推定との類似性 (Fujikoshi and Veitch[2]) より,

$$C_K = n(\hat{\lambda}_{K+1} + \dots + \hat{\lambda}_p) - 2(p-K)(H-K-1)$$

が提案される. この基準量を用いて, $\min\{C_0, C_1, \dots, C_p\} = C_K$ ならば有効次元縮小空間の次元は K であると推定する. 上記基準の導出についての妥当性について言及する必要があるが, これには目的変数 y の離散版 $\tilde{y}, \tilde{y} = h$ ($y \in I_h$) に関連した次の量が関係している.

$$E(x|\tilde{y} = h) = \mu_h, \quad \text{Var}(x|\tilde{y} = h) = \Omega_h, \quad p(\tilde{y} = h) = p_h \\ \sum_{h=1}^H p_h \mu_h = \mu, \quad \sum_{h=1}^H p_h \Omega_h = \Omega, \quad \sum_{h=1}^H p_h (\mu_h - \mu)(\mu_h - \mu)' = Q$$

参考文献

- [1] Ferré, L. (1998). Determining the Dimension in Sliced Inverse Regression and Related Methods, *J. Amer. Stat. Assoc.*, **93**, 132-140.
- [2] Fujikoshi, Y. and Veitch, L.G. (1979). Estimation of Dimension in Canonical Correlation Analysis. *Biometrika.*, **66**, 345-351.
- [3] Li, K.C. (1991). Sliced Inverse Regression for Dimension Reduction, (with discussion). *J. Amer. Stat. Assoc.*, **86**, 316-342.
- [4] Li, K.C. (1992). On Principal Hessian Direction for Data Visualization and Dimension Reduction: Another Application of Stein's Lemma. *J. Amer. Stat. Assoc.*, **87**, 1025-1039.
- [5] Schott, J.R. (1994). Determining the Dimensionality in Sliced Inverse Regression. *J. Amer. Stat. Assoc.*, **89**, 141-1408.