（2）「非正規性での統計理論と応用の研究」に関する研究報告

Nonparametric Estimation of a Conditional Quantile for α-mixing Processes

筑波大学社会科学系 本田敏雄

本報告では，dependenceのある場合の条件付き分布のパーセント点の推定を扱った．$(X_1', Y_1)', (X_2', Y_2)', \ldots, (X_n', Y_n)'$ を，定常α-mixing過程とし($Y_i \in R$，$X_i \in R^d$)，$\theta(x)$ を $Y|X = x$ の $100 \times \alpha$ パーセント点とする．この$\theta(x)$ の推定を考える．25%点，メディアン，75%点などはデータの記述の強力な手段であり，条件付きの25%点，メディアン，75%点もまたデータ記述の強力な手段であることはいうまでもない(Chaudhuri et al. (1997) など)．またメディアンは裾の重い分布，外れ値にたいして頑強であることからも，本研究の意義は明らかである．

関連する文献としては，i.i.d. の場合にBhattacharya and Gangopadhyay(1990), Jones and Hall(1990), Mehra et al.(1991), Chaudhuri(1991a,b), Fan et al.(1994), Welsh (1996), Xiang(1996) などがある．本報告ではlocal polynomial fittingによる推定を扱うが，関連する文献は，Chaudhuri(1991a,b), Fan et al.(1994), Welsh(1996) である．ここでの結果は，最初の論文をdependenceのある場合へ拡張したとみることもできるが，その論文中の証明は不明確に思える．またFan et al.(1994)にはcriticalな誤りがあり，Welsh(1996)には証明がない．α-mixing過程の場合については，Truong and Stone(1992)が，local medianによるconditional median functionの推定を扱っている（本報告の $p = 1$ に対応）．本研究ではその結果を拡張，改善している．

Chaudhuri(1991a,b)は，損失関数を

$$H_\alpha(t) = |t| + (2\alpha - 1)t. \tag{1}$$

をとして，local polynomial fittingにより，$\theta(x)$ とその微分係数の推定量を定義している．もし2乗誤差をとれば，通常のlocal polynomial regressionになる．ここではまずBabu(1989)の方法により推定量のBahadur表現を導き，そして漸近正規性を示した．

推定量の具体的な定義を与える．bandwidth h は，$Cn^{-1/(2p+d)}$ とする．$\theta(x)$ の $(p-1)$ 階までのテイラー展開は，

$$\theta(v) = \sum_{\lambda \in \Lambda} \beta_\lambda^x h^{-|\lambda|}(v-x)^\lambda - r_x(v) = P_h(\beta^x, v-x) - r_x(v), \tag{2}$$

ここで $\Lambda = \{(\lambda_1, \ldots, \lambda_d) \mid \lambda_i$ は非負の整数で $\sum \lambda_i < p\}$，$|\lambda| = \sum \lambda_i$，$|\Lambda|$ は Λ の濃度，$v^\lambda = \prod v_i^{\lambda_i} (\lambda \in \Lambda$ および $v \in R^d)$．β_λ^x を λ について上昇順にならべ，$\beta^x \in R^{|\Lambda|}$ と書く（h に依存）．さらに有界非負で，コンパクトな台をもつ核関数 $K(\cdot)$ をとり，$K(\cdot/h)$ を $K_h(\cdot)$ と書く．そして最適化問題

$$\sum_{i=1}^{n} K_h(X_i - x) H_\alpha(Y_i - P_h(\beta, X_i - x)) \to \min.$$

を考え，解の一つを $\hat{\beta}^x$ とおく．そのとき $\hat{\beta}^x$ の第1要素により，$\theta(x)$ が推定される．

主な結果をまとめると，次のようになる．

1. x を固定したときの，$\theta(x)$ の推定量の Bahadur 表現をあたえ，さらに剰余項の詳しい評価をあたえた．

2. ついで漸近正規性をしめした．基準化のオーダーは，$h^{-p/(2p+d)}$ である．

3. x について一様な，$\theta(x)$ の推定量の Bahadur 表現をあたえた．この場合も剰余項の詳しい評価をあたえた．

4. x に関する，推定量の一様収束をしめした．収束のオーダーに関しては，$h \sim (n^{-1} \log n)^{1/(2p+d)}$ で，h^p となることにも言及した．

セミパラメトリック推測理論のおはなし — BKRW を読む

千葉大学理学部数学・情報数理学科　今野良彦[1]

0. はじめに．古典的な母数模型における推測理論を拡張し、確率分布の全体のより「大きな」部分集合をモデルとするセミパラメトリックモデルにおける推測理論に関する論文や著書が近年おおく出版されている．ここでは、ひとつのアプローチである BKRW 流の（i.i.d. 標本における）理論の骨子の紹介を行った．

1. パラメトリックモデルの話．モデル $\boldsymbol{P} = \{P_\theta;\ \theta \in \Theta \subset R^k\}$ における母数の推定問題において、不偏推定量の分散に対する下限をあたえる Cramér-Rao の定理の漸近理論版である Hájek=稲垣 のたたみこみ定理が知られている[2]．これは、$\boldsymbol{\nu} = q(\theta)$ の推定を考えたとき、(1) モデルがある種の「なめらかさ」[3] をもつ、(2) q が全微分可能, (3) 推定量が正則、という 3 つの仮定のもと、その漸近分布は正規分布（平均がゼロで、分散が $\boldsymbol{\nu}$ についての Fisher 情報行列の逆行列で与えられるもの）とそれとは独立のノイズの和の分布として表現でき、漸近的に観て推定がどこまで「よく」できるかということに対する限界を与えるものである．その限界に達する推定量を有効という．

2. セミパラメトリックモデルの話．$(\boldsymbol{X}, \boldsymbol{A})$ 上の確率測度の集まりを $\boldsymbol{P} = \{P_g \ll \mu; g \in \boldsymbol{G}\}$ と記す．ただし、$\boldsymbol{G}$ は関数自由度も許すとする．母数モデルにおける有効推定の議論を拡張するために、いくつかの概念が道具として必要である．基本的には、接空間、スコア作用素、情報作用素[4] 等である．真のモデルを P_0 と記す．P_0 におけるモデルの接空間（$\dot{\boldsymbol{P}}$ と書くことにする）は $L_2^0(P_0)$（P_0 のもとでの平均がゼロになる L_2 空間の元からなる部分空間）の部分空間とみなすことができるが、$\dot{\boldsymbol{P}}$ が有限次元ではなく、その上 $L_2^0(P_0)$ の真部分集合のとき、$\dot{\boldsymbol{P}}$ をセミパラメトリックモデルとよぶことにする．$\boldsymbol{\nu} = \boldsymbol{\nu}(P_g) = \psi(g)$ を推定するとき、(1) $\boldsymbol{P}$ のある種の 可微分性もしくは LAN 、(2) $\boldsymbol{\nu}$ の道ごとの可微分性、(3) 推定量が正則、であるときにたたみこみ定理の拡張が得られる．(2) の条件の成立を判定[5] するのが、Van der Vaart の可微分定理であり、(3) を示すために Hoffmann-Jørgensen=Dudley 流の経験過程理論[6] を援用する．また、Euclid パラメータの推定に対する M- 推定法を拡張したもの、ノンパラメトリック最尤推定法（NPMLE）、penalized 最尤推定

---

[1] email address: konno@math.s.chiba-u.ac.jp

[2] もうひとつの重要な定理である Hájek (1972) により局所漸近ミニマックス定理についてはここではふれないことにする．

[3] モデルの局所尤度比の局所漸近正規性 (LAN) に対する十分条件．

[4] 母数モデルにおけるスコア関数、情報行列に対応するもので、Euclid 空間における線形変換を有界線形作用素に置き換えた概念．

[5] パラメータの「なめらかさ」が失われると、情報作用素の逆作用素が有界でなくなり、推定量の収束のオーダーが $1/\sqrt{n}$ よりも遅くなり、たたみこみ定理の意味での有効性という概念が成立しなくなる．

[6] これを可分ではない距離空間における弱収束の理論に基づくものであり、 modern empirical theory とよんでいる．古典的な議論との（統計学者にとっての）違いは、確率や期待値の記号の右肩に " * " の印がつくことと entropy calculus という道具が必要になってくることである．この理論の統計学者向けのすぐれた成書として VdVW がある．証明抜きで必要最少限のモダンな経験過程理論をまとめたものとして vdL の第 1 章がある．

法、sieve 推定法等による有効推定量を構築するための一般的な議論が BKRW の第7章にあるが、個々のモデルにおいて有効推定量を導出するとき、一般議論における条件を確認するのは容易とはいいがたい．最後に、Interval censored data model, case I を例にしてこれらの概念がどのように働くを観た．

3. おわりに．BKRW 以降[7]、1990 年代の Annals of Statistics を中心にセミパラメトリックの理論研究[8] に関する論文がたくさんあり、「旬の話題」といっても過言ではない[9] ．

参考文献

[BKRW] Bickel, P.J., Klaassen, C.A., Ritov, Y. and Wellner, J. *Efficient and adaptive estimation for semiparametric models*, Springer, 1998.

[G] Groeneboom, P. *Lectures on inverse problems* in " Lectures on Probability Theory and Statistics", Lecture notes in Mathematics **1648**, 67–164, 1996.

[GW] Groeneboom, P. and Wellner, J. *Information bounds and nonparametric maximum likelihood estimation*, Birkhäuser, 1992.

[horowitz] Horowitz, J.L. *Semiparametric methods in Econometrics* Lecture notes in Statistics 131, Springer, 1998.

[huang] Huang, J. and Wellner, J. Asymptotic normality of the NPMLE of linear functionals for interval censored data, case 1, *Statistica Neerlandica* **49**, 153–163, 1995.

[stein] Stein, C. Efficient nonparametric testing and estimation, " *Proc. Third Berkeley Symp. Math. Statist. Prob.*" **1**, 187–195, Univ. California Press, Berkeley, 1956.

[vdL] van der Laan, M.J. Efficient and inefficient estimation in semiparametric models, CWI Tracts **114**, Center for Mathematics and Computer Science, 1995.

[VdV] Van der Vaart, A. *Statistical estimation in large parameter spaces*, CWI Tract **44**, Center for Mathematics and Computer Science, 1988.

[VdVW] van der Vaart, A. and Wellner, J. *Weak convergence and empirical preocess*, Springer, 1996.

[7] John Hopkins 版は 1991 年に刊行され、Springer 版と内容は同じ．

[8] 計量経済学の理論研究においてもこの話題に関する論文がおおく、最近の成書として Horowitz がある ..

[9] modern empirical theory の整備に伴いようやく端緒についたものであり、まだまだ始まったばかりのこれからの話題であるというような主旨の発言を Wellner 氏はしていたように思える．

比の統計量の漸近表現と平均二乗誤差

前園宜彦(九州大学経済学部)

1. はじめに

$X_1,\cdots,X_n$ を互いに独立で同じ分布 F に従う確率変数とする. $T_n=T_n(X_1,\cdots,X_n)$, $S_n=S_n(X_1,\cdots,X_n)$ を母数 t_n, s_n に関連する統計量とする. 標本相関係数, 高次のキュムラントの推定量, Pearson's coefficient of variation 等のいくつかの重要な統計量は二つの統計量の比 T_n/S_n で表わされる. 本報告では比の統計量の漸近表現を残差が $n^{-1/2}o_p^*(n^{-1})$ まで求め, それを利用して漸近平均二乗誤差を n^{-2} の項まで求める. ここで $o_p^*(n^{-1})$ は

$$P\{|o_p^*(n^{-1})|\geq n^{-1}(\log n)^{-1}\}=o(n^{-1})$$

を満たす. これらを標本相関係数へ適用した結果を報告する.

比の統計量の分母・分子について, 次の $ANOVA$-decomposition を仮定する.

[仮定]

$$\begin{aligned}T_n &= t_n+n^{-1}\delta_T+n^{-2}\sum_{i=1}^{n}\tau_0(X_i)+n^{-1}\sum_{i=1}^{n}\tau_1(X_i)+n^{-2}\sum_{C_{n,2}}\tau_2(X_i,X_j)\\ &\quad +n^{-3}\sum_{C_{n,3}}\tau_3(X_i,X_j,X_k)+n^{-1/2}o_p^*(n^{-1}),\\ S_n &= s_n+n^{-1}\delta_S+n^{-2}\sum_{i=1}^{n}\zeta_0(X_i)+n^{-1}\sum_{i=1}^{n}\zeta_1(X_i)+n^{-2}\sum_{C_{n,2}}\zeta_2(X_i,X_j)\\ &\quad +n^{-3}\sum_{C_{n,3}}\zeta_3(X_i,X_j,X_k)+n^{-1/2}o_p^*(n^{-1})\end{aligned}$$

と表わされる. ここで δ_T と δ_S は定数で, $\sum_{C_{n,k}}$ は $1\leq i_1<i_2<\cdots<i_r\leq n$ についての全ての和を表わす. また τ, ζ は $ANOVA$-decomposition されたものである.

2. 漸近表現と平均二乗誤差

$ANOVA$-decomposition を利用して, 各項のモーメントの評価を行うことにより, 次の比の統計量の漸近表現を求めることができる.

[定理 1]. モーメントに対する仮定の元で

$$\frac{T_n}{S_n}=U_n+n^{-1/2}o_p^*(n^{-1}).$$

ここで

$$\begin{aligned}U_n &= \frac{t_n}{s_n}+n^{-1}\delta+n^{-2}\sum_{i=1}^{n}\eta_0(X_i)+n^{-1}\sum_{i=1}^{n}\eta_1(X_i)\\ &+ n^{-2}\sum_{C_{n,2}}\eta_2(X_i,X_j)+n^{-3}\sum_{C_{n,3}}\eta_3(X_i,X_j,X_k)\end{aligned}$$

で δ, η は τ, ζ によって決まり，$ANOVA$-decomposition された形である．この表現を利用すると比の統計量のいろいろな漸近的性質を論じることができる．n^{-2} の項までの漸近平均二乗誤差 $AE(\cdot)$ は次で与えられる．

[定理 2]. モーメントに対する仮定の元で

$$AE(\frac{T_n}{S_n}) = E[U_n - \frac{t_n}{s_n}]^2 = n^{-1}E[\eta_1^2(X_1)]$$
$$+n^{-2}\{\delta^2 + 2E[\eta_0(X_1)\eta_1(X_1)] + \frac{1}{2}E[\eta_2^2(X_1, X_2)]\} + O(n^{-3}).$$

比の統計量は漸近U－統計量となるから，Lai and Wang (1993, Statistica Sinica) の結果を使って n^{-1} の項までのエッジワース展開を求めることができる．

3. 相関係数

$\{\mathbf{X}_i\}_{i\geq 1}$ を2次元の確率ベクトルとし，$\mathbf{X}_i^t = (Y_i, Z_i)$ とおく．このとき，母相関係数 ρ の推定について考える．

$$T_n = (n-1)^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y})(Z_i - \bar{Z}),$$
$$S_n = \{(n-1)^{-2}\sum_{i=1}^{n}(Y_i - \bar{Y})^2\sum_{i=1}^{n}(Z_i - \bar{Z})^2\}^{1/2}$$

とおく．ここで $\bar{Y} = \sum Y_i/n$, $\bar{Z} = \sum Z_i/n$ である．このとき $r_n = T_n/S_n$ とおくと，r_n は標本相関係数である．Knott and Frangos (1983, B.K.) は元のデータが2次元正規分布のときに，$nVar(r_n)$ の漸近表現を求めている．ここで先の [定理1] 及び [定理2] を利用して r_n の漸近表現を求め，漸近分散等を論じる．$\mathbf{X}_i$ が2次元正規分布のときの漸近平均二乗誤差は

$$AE(r_n) = n^{-1}(1-\rho^2)^2 + n^{-2}(1-\rho^2)^2(1 + \frac{23}{4}\rho^2)$$

となる．これを元にした漸近分散は，これまでに得られていた Knott and Frangos (1983, B.K.) の結果と一致している．

さらにバイアスの修正を考えてみよう．バイアス δ の推定量を代入したバイアス修正標本相関係数 $r_n^* = r_n - \hat{\delta}/n$ を考えると，再び比の統計量の漸近表現を利用して，r_n^* の漸近平均二乗誤差は，$\mathbf{X}_i$ が2次元正規分布のとき

$$AE(r_n^*) = n^{-1}(1-\rho^2)^2 + n^{-2}(1-\rho^2)^2(2 + \frac{5}{2}\rho^2)$$

で与えられる．よって正規分布の場合は，$AE(r_n) - AE(r_n^*) = (1-\rho^2)^2(13\rho^2 - 4)/4$ となるから $|\rho| \geq 2/\sqrt{13}(= 0.555)$ のとき，r_n^* は r_n より漸近平均二乗誤差の意味で優れていることが分かる．

本論で報告した比の統計量の漸近表現は，高次のキュムラント，Pearson' の Coefficient of Variation 等の推定量にも応用可能である．

Density Estimation on Special Manifolds and Related Problems

香川大学工学部　筑瀬　靖子

There exists a large literature on univariate density estimation by various methods, for example, the method of kernels first considered by Rosenblatt (1956) and the method of orthogonal series introduced by Čencov (1962). The methods were extended to vector-variate density estimation by, e.g., Cacoullos (1966), Epanechnikov (1969), and Scott (1992).

The problem of density estimation on the space S_m of all $m \times m$ symmetric matrices and on the space $R_{m,p}$ of all $m \times p$ rectangular matrices was considered by Chikuse (1997).

Hall, Watson and Cabrera (1987) considered kernel density estimation on the unit hypersphere $V_{1,m}$. This paper is concerned with estimating unknown density functions of distributions on the Stiefel manifold $V_{k,m}$ for general $k \geq 1$, extending the discussion of the paper cited.

The Stiefel manifold $V_{k,m}$ is the space whose points are k-frames in R^m, where a set of k orthonormal vectors in R^m is called a k-frame in R^m $(k \leq m)$. The Stiefel manifold $V_{k,m}$ is represented by the set of $m \times k$ matrices X such that $X'X = I_k$, where I_k is the $k \times k$ identity matrix. For $m = k$, $V_{k,m}$ is the orthogonal group $O(m)$. For the derivations of the results on $V_{k,m}$, we need to define the Grassmann manifold. The Grassmann manifold $G_{k,m-k}$ is the space whose points are k-planes $\mathcal{V}$, that is, k-dimensional hyperplanes in R^m containing the origin. To each k-plane $\mathcal{V}$ in $G_{k,m-k}$, corresponds a unique $m \times m$ orthogonal projection matrix P idempotent of rank k onto $\mathcal{V}$. Let $P_{k,m-k}$ denote the set of all $m \times m$ orthogonal projection matrices idempotent of rank k.

We develop decompositions (or transformations) of random matrices on the Stiefel manifold $V_{k,m}$ and on the manifold $P_{k,m-k}$ (or the Grassmann manifold $G_{k,m-k}$), which also lead to the corresponding decompositions of the invariant measures (or Jacobians of the transformations) on the manifolds. We present one-to-one transformations of $P_{k,m-k}$ onto the space $R_{m-k,k}$ and a subspace $R^{(1)}_{m-k,k}$ of $R_{m-k,k}$. Thus, apart from sets of measure zero, the manifold $P_{k,m-k}$ is analytically homeomorphic to the spaces $R_{m-k,k}$ or $R^{(1)}_{m-k,k}$; we note the

dimension of $P_{k,m-k}$ being $k(m-k)$. The discussion is followed by the one-to-one transformations of the manifold $V_{k,m}$ onto the product spaces $R_{m-k,k} \times O(k)$ or $R^{(1)}_{m-k,k} \times O(k)$. The results are not only of theoretical interest in themselves but also of practical use.

We are concerned with the density estimation by the method of kernels on $V_{k,m}$, and we propose two classes of kernel density estimators $\hat{f}_1(X;M)$ and $\hat{f}_2(X;M)$ for $X \in V_{k,m}$ which are based on two kinds of residuals, with small smoothing parameter (positive definite) matrix M, choosing a kernel function $K(T)$ of matrix argument. For small smoothing parameter matrix M and/or for large sample size n, we investigate asymptotic behavior of various statistical measures of the estimatiors $\hat{f}(X;M), j = 1,2$. The one-to-one stransformation of the manifold $V_{k,m}$ onto the product space $R^{(1)}_{m-k,k} \times O(k)$ plays a useful role for the asymptotic evaluation of the integrals over $V_{k,m}$ occurring in those statistical measures. The general discussion of kernel density estimation on $V_{k,m}$ is applied and examined for a special kernel function $K(T) = \mathrm{etr}(-T)$. The preceding discussion may indicate that, with small smoothing parameter matrix M, the kernel density estimation is independent of the choice of the kernel function.

There may often occur the case where estimating unknown density functions on the space $R_{m-k,k}$ and hence on the subspace $R^{(1)}_{m-k,k}$ (and on $O(k)$) is easier than that on the manifold $V_{k,m}$, while the problem of density estimation on $R_{m,p}$ was already discussed in Chikuse (1997). We propose methods to solve the density estimation problem for that case using some decompositions of $V_{k,m}$.

The theory of density estimation on the Grassmann manifold $G_{k,m-k}$ or, equivalently, the manifold $P_{k,m-k}$ is also discussed in this paper.

標本度数とGEM分布
（離散的なnonparametric Bayesの事前分布として）

大和　元（鹿児島大学・理），　野町　俊文（都城工業高専）

自然数の全体 $\{1,2,3,...\}$ の上のすべての離散的な確率分布（要素はすべて正）の集合を $\Delta=\{(z_1,z_2,...):z_j>0\ (j=1,2,...),\sum_{j=1}^{\infty}z_j=1\}$ で表す。このΔを supportとする事前分布μを想定するが、この事前分布μに従う（Δ の要素である）a random probability（infinite random proportions）を $\mathbf{Z}$ で表す。従って、確率1で $\mathbf{Z}\in\Delta$である。$(X_1,...,X_n)$ を $\mathbf{Z}(\in\Delta)$ からの大きさ n の標本とする。即ち、条件 $Z=(z_1,z_2,...)$ の下で, $(X_1,...,、X_n)$ は $P(X_1=j_1,X_2=j_2,...,X_n=j_n\mid\mathbf{Z}=(z_1,z_2,...))=z_{j_1}z_{j_2}\cdots z_{j_n}$ を満たしている。ここで、$j_1,j_2,...,j_n=1,2,...$。$\mathbf{Z}(\in\Delta)$ からの標本 $(X_1,...,X_n)$ に対して, $D_n=(D_{n1},D_{n2},...,D_{nk})$ は 順序統計量 $X_{(1)}<X_{(2)}<\cdots<X_{(k)}$ の頻度を表す, ここで k は異なる観測値の個数を表す。これについて、Donnelly and Joyce (1991) は次の命題を示している：

μ がパラメータ$\alpha(>0)$のGEM分布である(或いは $\mathbf{Z}$ がパラメータ$\alpha(>0)$のGEM分布に従う)ときに限り、D_n は次で与えられるDonnelly-Tavaré-Griffiths(DTG) formulaに従う、

$$P(D_n=(c_1,...,c_k))=\frac{\alpha^k}{\alpha^{[n]}}\cdot\frac{(n-1)!}{\prod_{j=1}^{k-1}(n-\sum_{i=1}^{j}c_i)}\,, \tag{1}$$

ただし、$1\le k\le n$, $c_1,...,c_k$ は $c_1+\cdots+c_k=n$ を満たす正の整数、また$\alpha^{[n]}=\alpha(\alpha+1)\cdots(\alpha+n-1)$。GEM分布とは次頁の式（3）で表される分布である。GEM分布及びDTG formulaの名称はEwens(1990)による。

上の命題をDonnelly and Joyce (1991)とは異なる方法で, 以下の論法により示す事ができる (Yamato and Nomachi (1997))。

正の整数 k, $c_1,...,c_k$ $(n=c_1+c_2+\cdots+c_k)$ について, 次が成り立つ事に注意する；

$$E(\sum_{i_1<\cdots<i_k}Z_{i_1}^{c_1}\cdots Z_{i_k}^{c_k})$$

$$=P(X_1=\cdots=X_{c_1}<X_{c_1+1}=\cdots=X_{c_1+c_2}<\cdots<X_{c_1+\cdots+c_{k-1}+1}=\cdots=X_n)$$

$$=P(D_{n1}=c_1,...,D_{nk}=c_k)/\left(\begin{array}{c}n\\c_1,c_2,...,c_k\end{array}\right).$$

Lemma 1 離散順序統計量の頻度 D_n, $n=1,2,...$ が (1) で与えられるパラメータ $\alpha(>0)$ の *DTG formula* に従うことと *infinite random proportions* $\mathbf{Z}$ が次の関係を満たすことは同値である；

$$E(\sum_{i_1<\cdots<i_k}Z_{i_1}^{c_1}\cdots Z_{i_k}^{c_k})=\frac{\alpha^k}{\alpha^{[c_1+\cdots+c_k]}}\prod_{j=1}^{k}\frac{c_j!}{\sum_{i=j}^{k}c_i} \tag{2}$$

或いは次の関係を満たすことは同値である；

$$P(X_1 = \cdots = X_{c_1} < X_{c_1+1} = \cdots = X_{c_1+c_2} < \cdots < X_{n-c_k+1} = \cdots = X_n)$$
$$= \frac{\alpha^k}{\alpha^{[c_1+\cdots+c_k]}} \prod_{j=1}^{k} \frac{c_j!}{\sum_{i=j}^{k} c_i},$$

ただし、α は正の定数、$k, c_1, ..., c_k\,(n = c_1 + \cdots + c_k)$ は正の整数であり，$\sum_{i_1<\cdots<i_k}$ は $i_1 < \cdots < i_k$を満足する全ての正の整数 $i_1, ..., i_k$についての和を表す。

Lemma 2

$$P(X_1 = \cdots = X_{c_1} < X_{c_1+1}, X_{c_1+2}, ..., X_n) = \frac{\alpha \cdot \alpha^{[c_2+\cdots+c_k]} c_1!}{(c_1 + \cdots + c_k)\alpha^{[c_1+\cdots+c_k]}},$$

ここで、c_1,...,c_k は $n = c_1 + \cdots + c_k$を満たす正の整数。

Proposition 1 パラメータ$\alpha(> 0)$ の *DTG formula* に従う $\mathbf{Z} \in \Delta$ に対し、正の整数 m および c_1,...,c_k について

$$E[\sum_{i_1<\cdots<i_k} (\frac{Z_{m+i_1}}{1 - Z_1 - \cdots - Z_m})^{c_1} \cdots (\frac{Z_{m+i_k}}{1 - Z_1 - \cdots - Z_m})^{c_k}] = \frac{\alpha^k}{\alpha^{[c_1+\cdots+c_k]}} \prod_{j=1}^{k} \frac{c_j!}{\sum_{i=j}^{k} c_i}.$$

Proposition 2 式（2）を満足する $\mathbf{Z} = (Z_1, Z_2, ...) \in \Delta$ に対し、正の整数 m 及び c_1,...,c_k について

$$E[\sum_{m<i_1<\cdots<i_k} Z_{i_1}^{c_1} \cdots Z_{i_k}^{c_k}] = E(1 - Z_1 - \cdots - Z_m)^{c_1+\cdots+c_k} \frac{\alpha^k}{\alpha^{[c_1+\cdots+c_k]}} \prod_{j=1}^{k} \frac{c_j!}{\sum_{i=j}^{k} c_i}.$$

以上の結果を用いて次を示す事ができる。

Proposition 3 (Donnelly and Joyce (1991)) $\mathbf{Z} \in \Delta$とし、順序統計量 の頻度 D_n, $n = 1, 2, ...$は (1) で与えられるパラメータ$\alpha(> 0)$ の *DTG formula* に従うものとする。

このとき、Z_j, $j = 1, 2, ...,$ は次の様に表される。

$$Z_1 = V_1, \quad Z_j = (1 - V_1) \cdots (1 - V_{j-1}) V_j \quad (j = 2, 3, ...), \tag{3}$$

ただし、$V_1, V_2, ...$ は独立で同じベータ分布 $Be(1, \alpha)$ に従う。この $\mathbf{Z} = (Z_1, Z_2, ...)$ の分布のことをパラメータαの *GEM* 分布と言う。

上の命題の逆を GEM 分布が（3）で表現される事を用いて直接示す事ができる。

Proposition 4 (Donnelly and Joyce (1991)) *the infinite random proportions* $\mathbf{Z} = (Z_1, Z_2, ...)$ がパラメータ$\alpha(> 0)$ の GEM 分布に従うとき、$\mathbf{Z}$ は関係式 (2) を満たす。

[参考文献]

Donnelly, P. and Joyce, P. (1991), *Adv. Appl. Prob.* **23** , 229–258.

Ewens, W. (1990), Mathematical and Statistical Developments of Evolutionary Theory (ed. by S. Lessard), Kluwer Academic Publishers.

Yamato, H. and Nomachi, T. (1997), *J. Nonparametric Statist.*, **8**, 355–363.

Asymptotic approximation of the null distributions of one-way ANOVA test statistics under nonnormality

広島大・理　柳原 宏和

実験結果に影響する因子（実験結果に, 偶然的要因以外の何らかの影響を及ぼす原因）として一つの因子を考え, その因子は r 個の水準を持つとする. 第 i 水準の実験において得られた観測値 x_{ij} $(i=1,2,\cdots,q; j=1,2,\cdots,n_i)$ は, 平均 μ_i 分散 ${\sigma_i}^2$ をもつ母集団 Π_i $(i=1,2,\cdots,r)$ からなる標本とする. すなわち, x_{ij} を実現値とする確率変数 X_{ij} は, 一元配置モデル

$$X_{ij}=\mu_i+\varepsilon_i \qquad (j=1,2,\cdots,n_i; i=1,2,\cdots,q)$$

に従うとする. ここで, ε_i は互いに独立で平均 0, 分散 ${\sigma_i}^2$ を持つ確率変数で, 誤差と呼ばれる. この因子が実験結果に影響を及ぼしているかどうかを考察したい場合, 仮説検定問題,

$$H_0:\mu_1=\mu_2=\cdots=\mu_q \qquad \text{vs} \qquad H_1:\text{ある } i,\ j \text{ に対して } \mu_i\neq\mu_j,$$

を考える. この検定に対して, (1) ε_i の分散が等しい場合, (2) 分散が未知で等しくない場合に対して, それぞれ以下のような検定統計量 T_{01}, T_{02}, が提案されている.

$$(1) \qquad T_1=\frac{(n-q)\sum_{i=1}^{q} n_i(\bar{X}_{i.}-\bar{X}_{..})^2}{\sum_{i=1}^{q}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i.})^2},$$

$$(2) \qquad T_2=\sum_{i=1}^{q}\frac{n_i(n_i-1)(\bar{X}_{i.}-\tilde{X})^2}{\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i.})^2}.$$

ただし,

$$\bar{X}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij},\ \bar{X}_{..}=\frac{1}{n}\sum_{i=1}^{q}\sum_{j=1}^{n_i} X_{ij},\ n=\sum_{i=1}^{q} n_i,$$

$$\tilde{X} = \left(\sum_{j=1}^{n_i}\frac{n_i(n_i-1)}{(X_{ij}-\bar{X}_{i.})^2}\right)^{-1}\sum_{j=1}^{n_i}\frac{n_i(n_i-1)\bar{X}_{i.}}{(X_{ij}-\bar{X}_{i.})^2}.$$

誤差項が正規分布に従う場合, 検定統計量 T_1 は, 尤度比検定統計量であって, 仮説のもとで $(n-q)^{-1}T_1$ は F 分布に従うことが知られている. また分散が既知の場合の検定統計量 T_2 は, 標本分散 $\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i.})^2/(n_i-1)$ を真の分散 σ_i に置き換えたものであるが, この統計量は正規性の仮定において漸近的に χ^2 分布に従うことも良く知られている. 本報告では, 誤差項が非正規の場合に, これらの統計量の仮説のもとでの分布を取り扱う.

非正規の場合, T_1, T_2 は漸近的に χ^2 分布に従うことが知られている. しかし, 大標本でない場合, これらの検定統計量の分布としてこの漸近分布を使うこと非常に危険である. 特に小標本の場合, 名目上の検定のサイズと, 実際の検定のサイズの差は大きなものとなる. この名目上の検定のサイズと, 実際の検定のサイズの差との差異を小さくする方法をさぐることが本報告の目的である.

そのため, 以下の三つの方法を用いて検定のサイズを改良する.

1. 統計量の分布の漸近展開式を利用する方法.
2. 漸近近似を改良する単調変換統計量を用いる方法.
3. ブートストラップ法.

方法 1 に関して, T_1 の漸近展開の式は, Fujikoshi, Ohmae and Yanagihara (1998) で導出されているものを使う. T_2 に関しては, 適当な正則条件 (Bhattacharya and Ghosh (1978)) のもとで漸近展開を導出する. これらの場合, 漸近展開式の第二項まで考慮した分布で近似を行なうことは, 検定のサイズを改良することに関して有効な手段になっている. しかし, 非正規の場合, 一般に漸近展開式の導出に膨大な量の計算を必要とするので, 必ずしも有効な手段であるとはいえない. 次に改良変換として Bartlett 変換, 一般化 Bartlett 型変換, さらに Fujikoshi (1998) で提案された, 改良 Bartlett 変換を考える. Bartlett 変換は, 検定統計量の平均の摂動展開の形がわかっていれば用いることができ, 計算が比較的楽である. だが Bartlett 変換の場合は, 非正規の場合には誤差のオーダーは改良はされていない. 検定統計量 T_1 , T_2 のブートストラップ近似法は, Fisher and Hall (1990) によって提案されている. これらの 3 つの方法の比較, シミュレーションによる近似の精度について報告を行なった.

References

[1] Bhattacharya, R. N., and Ghosh, J.K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist* **6**, 434-451; Corrigendum, ibid. **8** (1980).

[2] Fisher, N. I., and Hall, P. (1990). On bootstrap hypothesis testing. *Austral. J. Statist.* **32**(2), 177-190.

[3] Fujikoshi, Y. (1998). Transformations with improved chi-squared approximations. Submitted for publications.

[4] Fujikoshi, Y., Ohmae, M., and Yanagihara, H. (1998). Asymptotic expansion for the null distribution of one-way ANOVA test statistic under nonnormality. Submitted for publications.

MANOVAモデルにおける次元に関する検定統計量の非正規性の影響について

東京理科大 理工 小林　正明
東京理科大 理工 瀬尾　隆

多変量解析の次元に関する検定問題における検定統計量は，通常，多変量正規母集団の仮定のもとで考えられており，正規性の仮定に対して頑健であるかどうかを調べることは重要な問題の１つである．本報告では，MANOVAモデルの次元に関する検定統計量の分布について漸近展開の形で議論し，特に，エリプティカル母集団の下での影響を調べ，より良い近似となる修正した検定統計量を与えた．この問題については，Seo, Kanda and Fujikoshi [2] の中ですでに議論されているが，より簡単化することによって尖度パラメータなどの κ や φ を含まない修正項をもつ検定統計量を与えることができた．

多変量分散分析モデルにおける平均ベクトルの検定問題において，第jグループからの観測ベクトルを$\boldsymbol{x}_\alpha^{(j)}$ $(\alpha = 1, \cdots, N_j)$とし，それに基づく群内積和行列を$\boldsymbol{S}_W$，群間積和行列を$\boldsymbol{S}_B$とすると，それらは，

$$\boldsymbol{S}_W = \sum_{j=1}^{q} \boldsymbol{S}_j, \quad \boldsymbol{S}_j = \sum_{\alpha=1}^{N_j} (\boldsymbol{x}_\alpha^{(j)} - \overline{\boldsymbol{x}}^{(j)})(\boldsymbol{x}_\alpha^{(j)} - \overline{\boldsymbol{x}}^{(j)})',$$

$$\boldsymbol{S}_B = \sum_{j=1}^{q} N_j (\overline{\boldsymbol{x}}^{(j)} - \overline{\boldsymbol{x}})(\overline{\boldsymbol{x}}^{(j)} - \overline{\boldsymbol{x}})'$$

であり，ここに，

$$\overline{\boldsymbol{x}} = \frac{1}{N} \sum_{j=1}^{q} N_j \overline{\boldsymbol{x}}^{(j)}, \quad \overline{\boldsymbol{x}}^{(j)} = \frac{1}{N_j} \sum_{\alpha=1}^{N_j} \boldsymbol{x}_\alpha^{(j)}, \quad N = \sum_{j=1}^{q} N_j.$$

である．このとき $\boldsymbol{S}_B \boldsymbol{S}_W^{-1}$ の固有根を $d_1 \geq \cdots \geq d_p$，対応する母集団の固有根を $\delta_1 \geq \cdots \geq \delta_p$ とすると，判別分析の次元に関する仮説

$$H_k : \delta_1 \geq \cdots \geq \delta_k > \delta_{k+1} = \cdots = \delta_p = 0$$

に対して，次の3つの検定統計量

(i) $T_1 = \log \prod_{j=k+1}^{p}(1+d_j)$,

(ii) $T_2 = \sum_{j=k+1}^{p} d_j$,

(iii) $T_3 = \sum_{j=k+1}^{p} d_j/(1+d_j)$

が考えられる ($k=0$ の場合は，平均ベクトルの有意性検定問題になっている). 正規性の仮定のもとでは，これらの検定統計量の仮説のもとでの χ^2近似および漸近展開が与えられている (Fujikoshi [1] を参照).

本報告では，これらの検定統計量の仮説のもとでの分布を非正規であるエリプティカル分布に対して，漸近展開の形で与え，修正項が κ や φ に依存しない修正検定統計量を導出した. この結果は，正規母集団の下で得られる修正項と一致しており，その意味でエリプティカル母集団の下でも漸近的に ロバスト であることがいえた. また，シミュレーション実験については，代表的なエリプティカル分布である multivariate normal, ε-contaminated normal について報告し，数値的な影響を調べた.

参考文献

[1] Fujikoshi, Y. (1977). Asymptotic expansions of the distributions of the latent roots in MANOVA and the canonical correlations, *J. Mult. Anal.* **7** 386–396.

[2] Seo, T., Kanda, T. and Fujikoshi, Y. (1995). The effects of nonnormality on tests for dimensionality in canonical correlation and MANOVA models, *J. Mult. Anal.* **52** 325–337.

非正規母集団の下での漸近展開公式とそのホテリング T^2 型統計量への応用

北海道大学経済　　柿沢　佳秀
明星大学　　岩下　登志也

1　はじめに

$\boldsymbol{X}_1^{(a)},\ldots,\boldsymbol{X}_{N_a}^{(a)}$ $(N_a > p)$ を第 a 母集団からの p 次元の i.i.d. 連続確率ベクトルとし、$E(\boldsymbol{X}_t^{(a)}) = \boldsymbol{\mu}^{(a)}$、$Var(\boldsymbol{X}_t^{(a)}) = \Sigma^{(a)}$ とする。さらに、必要に応じて任意の高次のキュムラント $Cum(X_{t,j_1}^{(a)},\ldots,X_{t,j_s}^{(a)}) = \kappa_{j_1\ldots j_s}^{(a)}$ $(s \geq 3)$ の存在を仮定する。ただし、$X_{t,j}^{(a)}$ は $\boldsymbol{X}_t^{(a)}$ の第 j 成分とする。また、$\boldsymbol{X}_t^{(a)}$ と $\boldsymbol{X}_s^{(b)}$ は $a \neq b$ で独立と仮定する。標本平均ベクトル、標本分散共分散行列を

$$\overline{\boldsymbol{X}}^{(a)} = N_a^{-1}\sum_{t=1}^{N_a}\boldsymbol{X}_t^{(a)}, \quad S_X^{(a)} = (N_a-1)^{-1}\sum_{t=1}^{N_a}(\boldsymbol{X}_t^{(a)} - \overline{\boldsymbol{X}}^{(a)})(\boldsymbol{X}_t^{(a)} - \overline{\boldsymbol{X}}^{(a)})'$$

で定義する。なお、1標本問題に対しては上付の $^{(1)}$ 等を省略する。

仮説 $H : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ を検定する（あるいは、$\boldsymbol{\mu}$ の信頼領域を構成する）ための（1標本）ホテリングの T^2 統計量は

$$T_{[1]}^2 = N(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)'S_X^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)$$

で定義され、正規性の下で帰無分布について $(N-p)T_{[1]}^2/\{p(N-1)\} \sim F_{N-p}^p$ が知られている。また、等分散の仮定の下で仮説 $H : \boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} = \boldsymbol{\Delta}_0$ を検定する（あるいは、$\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}$ の信頼領域を構成する）ための（2標本）ホテリングの T^2 統計量は

$$T_{[2]}^2 = \frac{N_1N_2}{N_1+N_2}(\overline{\boldsymbol{X}}^{(1)} - \overline{\boldsymbol{X}}^{(2)} - \boldsymbol{\Delta}_0)'S_{X,pool}^{-1}(\overline{\boldsymbol{X}}^{(1)} - \overline{\boldsymbol{X}}^{(2)} - \boldsymbol{\Delta}_0)$$

で定義され、正規性の下で帰無分布について $(N_1+N_2-p-1)T_{[2]}^2/\{p(N_1+N_2-2)\} \sim F_{N_1+N_2-p-1}^p$ が知られている。ただし、

$$S_{X,pool} = \frac{(N_1-1)S_X^{(1)} + (N_2-1)S_X^{(2)}}{N_1+N_2-2}$$

とする。

非正規性の $T_{[1]}^2$ あるいは $T_{[2]}^2$ の帰無分布への影響に関して 1960 年代、1970 年代の Monte Carlo simulation による研究報告があり、特に Everitt (1979) は (I) 2標本問題の方が1標本問題よりも非正規性の影響がないこと、(II) 1標本問題は歪度にひどく影響されることを指摘している。混合正規分布に関して、$T_{[1]}^2$ の精密な帰無分布を扱った文献も見られるが、最近、$T_{[1]}^2$ の非正規母集団の下での漸近展開が明らかにされている。Iwashita (1997) が楕円母集団に対し局所対立仮説の下での $T_{[1]}^2$ の分布の漸近展開を与え、Kano (1995) 及び Fujikoshi (1997) は非正規母集団に対し $T_{[1]}^2$ の帰無分布の漸近展開を与えた。Kano は Iwashita の導出を行列演算を用いて整理し、一方、Fujikoshi は多次元 t-統計量 $\sqrt{N}S_X^{-1/2}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})$ の漸近展開を経由して導出した。

本報告では岩下 (1996, 統計学会) で考察された微分オペレータによるアプローチを非正規母集団へ拡張し、T^2-type 統計量の分布の漸近展開へ応用する。典型的な応用例は $T_{[1]}^2$ あるいは $T_{[2]}^2$ であり、局所対立仮説の下での漸近展開を導出する。Everitt (1979) の Monte Carlo simulation による結論が漸近展開として説明された。また、関連する話題についても当日紹介する。

2　期待値の漸近展開公式とその応用

$\boldsymbol{Y}_t^{(a)} = \boldsymbol{X}_t^{(a)} - E(\boldsymbol{X}_t^{(a)}) = (Y_{t1}^{(a)}, \ldots, Y_{tp}^{(a)})'$ とおき、

$\sigma_{ij}^{(a)} = E(Y_{ti}^{(a)} Y_{tj}^{(a)})$, $\kappa_{ijk}^{(a)} = E(Y_{ti}^{(a)} Y_{tj}^{(a)} Y_{tk}^{(a)})$, $\kappa_{ijk\ell}^{(a)} = E(Y_{ti}^{(a)} Y_{tj}^{(a)} Y_{tk}^{(a)} Y_{t\ell}^{(a)}) - [3]\sigma_{ij}^{(a)}\sigma_{k\ell}^{(a)}$, etc.

とおく。ただし、$[3] = (ij|k\ell) + (ik|j\ell) + (i\ell|jk)$ は3通りの和を表す。

[定理]　$h(\boldsymbol{z}^{(a)}, \Gamma^{(a)})$ を $p \times 1$ ベクトル $\boldsymbol{z}^{(a)} = (z_i^{(a)})$ と $p \times p$ 正定値行列 $\Gamma^{(a)} = (\gamma_{ij}^{(a)})$ の正則関数とするとき、

$$E[h(N_a^{1/2}\overline{\boldsymbol{Y}}^{(a)}, S_Y^{(a)})] = \Theta^{(a)}(\boldsymbol{\partial}^{(a)}, \partial^{(a)}) h(\boldsymbol{z}^{(a)}, \Gamma^{(a)})\Big|_{\boldsymbol{z}^{(a)}=0, \Gamma^{(a)}=\Sigma^{(a)}} + \mathrm{o}(N_a^{-1}) .$$

ただし、

$$\boldsymbol{\partial}^{(a)} = (\partial_1^{(a)}, \ldots, \partial_p^{(a)})' , \ \partial_i^{(a)} = \frac{\partial}{\partial z_i^{(a)}}$$

及び

$$\partial^{(a)} = (\partial_{ij}^{(a)}; i, j = 1, \ldots, p) , \ \partial_{ij}^{(a)} = \partial_{ji}^{(a)} = \frac{1}{2}(1 + \delta_{ij}) \frac{\partial}{\partial \gamma_{ij}^{(a)}}$$

とおくと

$$\begin{aligned}
\Theta^{(a)}(\boldsymbol{\partial}^{(a)}, \partial^{(a)}) = \exp\Big(\frac{1}{2}\boldsymbol{\partial}^{(a)\prime}\Sigma^{(a)}\boldsymbol{\partial}^{(a)}\Big)\Bigg[1 &+ N_a^{-1/2}\sum_{ijk}\kappa_{ijk}^{(a)}\Big(\partial_{ij}^{(a)}\partial_k^{(a)} + \frac{1}{6}\partial_i^{(a)}\partial_j^{(a)}\partial_k^{(a)}\Big) \\
&+ N_a^{-1}\Big\{\mathrm{tr}(\Sigma^{(a)}\partial^{(a)})^2 + \sum_{ijk\ell}\kappa_{ijk\ell}^{(a)}\Big(\frac{1}{2}\partial_{ij}^{(a)}\partial_{k\ell}^{(a)} + \frac{1}{2}\partial_{ij}^{(a)}\partial_k^{(a)}\partial_\ell^{(a)} + \frac{1}{24}\partial_i^{(a)}\partial_j^{(a)}\partial_k^{(a)}\partial_\ell^{(a)}\Big)\Big\} \\
&+ N_a^{-1}\frac{1}{2}\Big\{\sum_{ijk}\kappa_{ijk}^{(a)}\Big(\partial_{ij}^{(a)}\partial_k^{(a)} + \frac{1}{6}\partial_i^{(a)}\partial_j^{(a)}\partial_k^{(a)}\Big)\Big\}^2\Bigg] .
\end{aligned}$$

定理の応用として、局所対立仮説 $A_N : E(\boldsymbol{X}_t) = \boldsymbol{\mu}_0 + N^{-1/2}\boldsymbol{\varepsilon}$ の下での $T_{[1]}^2$ の MGF は

$$E[\exp(tT_{[1]}^2)] = \Theta(\boldsymbol{\partial}, \partial)\exp\{t(\boldsymbol{z}+\boldsymbol{\varepsilon})'\Gamma^{-1}(\boldsymbol{z}+\boldsymbol{\varepsilon})\}\Big|_{\boldsymbol{z}=0, \Gamma=\Sigma} + \mathrm{o}(N^{-1})$$

と展開されるが、微分作用素の計算は正規分布の高次モーメントを通して実行できる。注目すべき点として、T^2 型統計量に対しては上記と同様な計算に至ることである。たとえば、$T_{[2]}^2$ の帰無分布を考えよう。$N_1 = cN, N_2 = (1-c)N$ 及び、$\Sigma^{(1)} = \Sigma^{(2)} = \Sigma$ を仮定すると、帰無仮説の下で

$$E[\exp(tT_{[2]}^2)] = \Theta^{(1)}\{(1-c)^{1/2}\boldsymbol{\partial}, c\partial\}\Theta^{(2)}\{-c^{1/2}\boldsymbol{\partial}, (1-c)\partial\}\exp(t\boldsymbol{z}'\Gamma^{-1}\boldsymbol{z})\Big|_{\boldsymbol{z}=0, \Gamma=\Sigma} + \mathrm{o}(N^{-1})$$

と展開される。

参考文献

[1] Everitt, B.S. (1979). *J. Amer. Statist. Ass.* **74** 48–51.

[2] Fujikoshi, Y. (1997). *J. Mult. Anal.* **61** 187–193.

[3] Iwashita, T. (1997). *J. Statist. Plan. Inf.* **61** 85–104.

[4] Kano, Y. (1995). *Amer. J. Math. Management Sciences* **15** 317–341.

経時測定データに対する一般化推定方程式に基づく推定量の効率

広島大・理　佐藤　由佳
広島大・理　藤越　康祝

Liang and Zeger (*Biometrika,* **73** (1986), 13-22) は経時データの解析に対して, 次のモデル, および, 一般化推定方程式(GEE) に基づく推定法を提案した. 個体 i $(i=1,\ldots,n)$ の時点 t $(t=1,\ldots,p_i)$ における反応変数を y_{it} , $(k+1)$ 次元共変量ベクトル(説明変数ベクトル) を $\boldsymbol{x}_{it}$ とする. また, 個体 i の p_i 次元反応ベクトルを $\boldsymbol{y}_i=(y_{i1},\ldots,y_{ip_i})'$, $p_i\times(k+1)$ 個体内計画行列を $X_i=(\boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{ip_i})'$ とする. このとき, まず, $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n$ は互いに独立で, かつ, $\boldsymbol{y}_i$ の各成分に一般化線形モデルを仮定する. すなわち,

1. $\boldsymbol{y}_i$ の第 t 成分 y_{it} $(t=1,\ldots,p_i)$ は指数分布族に従う. すなわち,

$$f(y_{it}) \quad = \quad \exp[\ \{y_{it}\theta_{it}-b(\theta_{it})\}/\phi+c(y_{it},\phi)\ ]$$

2. $\mathrm{E}(y_{it})=\mu_{it}=h^{-1}(\boldsymbol{x}_{it}'\boldsymbol{\beta})$

3. $\mathrm{Var}(y_{it})=\phi g(\mu_{it})$

ただし, b,c,h,g は既知の関数, $\boldsymbol{\beta}$ は $(k+1)$ 次元未知パラメータベクトル, ϕ はスケールパラメータ, $\theta_{it}=\theta_{it}(\mu_{it})$ は自然パラメータである. 関数 h は連結関数, 関数 g は分散関数と呼ばれる. 次に, $\boldsymbol{y}_i$ の相関構造に対して, 仮の相関行列 $R_i(\boldsymbol{\alpha})$ をもつと仮定する. ここに, $\boldsymbol{\alpha}$ は q 次元未知パラメータベクトルである. $R_i(\boldsymbol{\alpha})$ は作業(Working) 相関行列と呼ばれ, すべての個体に対して同じ構造を仮定し,

$$V_i=\phi\Delta_i^{\frac{1}{2}}R_i(\boldsymbol{\alpha})\Delta_i^{\frac{1}{2}},\quad \Delta_i=\mathrm{diag}(g(\mu_{i1}),\cdots,g(\mu_{ip_i}))$$

とおく. V_i は作業(Working) 共分散行列と呼ばれ, もし $R_i(\boldsymbol{\alpha})$ が $\boldsymbol{y}_i$ の真の相関行列 $\mathrm{Corr}(\boldsymbol{y}_i)$ に等しければ $V_i=\mathrm{Cov}(\boldsymbol{y}_i)$ となる. このとき, $\boldsymbol{\beta}$ は一般化推定方程式(GEE)

$$U_{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\alpha}) \quad = \quad \sum_{i=1}^{n}D_i'V_i^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_i) \quad = \quad 0$$

の解として推定される. ただし, $D_i=\partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}'$, $\boldsymbol{\mu}_i=(\mu_{i1},\cdots,\mu_{ip_i})'$ である.

推定量 $\hat{\boldsymbol{\beta}}_G$, より正確には $\sqrt{n}(\hat{\boldsymbol{\beta}}_G-\boldsymbol{\beta})$ の漸近共分散行列 $V_G(\boldsymbol{\beta})$ は,

$$V_G(\boldsymbol{\beta})=\lim_{n\to\infty}n\{H_1(\boldsymbol{\beta})\}^{-1}H_2(\boldsymbol{\beta})\{H_1(\boldsymbol{\beta})\}^{-1}$$

で与えられる. ただし,

$$H_1(\boldsymbol{\beta}) = \sum_{i=1}^{n} D_i' V_i^{-1} D_i, \quad H_2(\boldsymbol{\beta}) = \sum_{i=1}^{n} D_i' V_i^{-1} \mathrm{Cov}(\boldsymbol{y}_i) V_i^{-1} D_i.$$

作業相関構造 $R_i(\boldsymbol{\alpha})$ が p_i 次単位行列, つまり, 個体内の観測が独立である場合の推定方程式を $U_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = U_I(\boldsymbol{\beta})$ と表す. この推定方程式 $U_I(\boldsymbol{\beta})$ は独立推定方程式(Independent Estimating Equations; IEE) と呼ばれる. IEEの解を $\hat{\boldsymbol{\beta}}_I$ と表し, $\sqrt{n}(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta})$ の漸近共分散行列を $V_{\mathrm{I}}(\boldsymbol{\beta})$ と表す.

GEE は, 個体内の観測が一般化線形構造に加え何らかの相関構造をもつ場合の推定法であるが, その相関構造は仮のものを利用する場合が多い. このため, 誤った相関を用いた時に回帰パラメータの推定にどの程度影響を及ぼすかが問題になる. とりわけ, 推定法が簡単になる独立相関構造を用いた時の影響に関心がある. 逆のいい方をすれば, ある種の相関構造を考慮することによって, どれだけ推定の効率が上げられるかの問題でもある.

本報告では, (i) 真の相関が Exchangeable 型 であるのに Working 相関を Independent 型 と誤って特定化した場合, (ii) 真の相関が Autoregressive 型(AR(1)) であるのに Working 相関を Independent 型 と誤って特定化した場合 において, 回帰パラメータ $\boldsymbol{\beta}$ に及ぼす影響を漸近相対効率を通して検討した. 回帰パラメータ β_j $(j = 0, 1, \ldots, k)$ の二つの推定量 $\hat{\beta}_{Ij}$ と $\hat{\beta}_{Gj}$ の漸近分散 $V_{Ij}(\boldsymbol{\beta})$, $V_{Gj}(\boldsymbol{\beta})$ はそれぞれ $V_I(\boldsymbol{\beta})$, $V_G(\boldsymbol{\beta})$ の対角成分で与えられる. このとき, $\hat{\beta}_{Gj}$ に対する $\hat{\beta}_{Ij}$ の漸近相対効率(ARE) は

$$\mathrm{ARE}_j \;=\; \frac{V_{Gj}(\boldsymbol{\beta})}{V_{Ij}(\boldsymbol{\beta})}$$

と表される. 漸近相対効率 ARE_j は説明変数のタイプに依存しているが, この依存性を解析的に表すため, $\boldsymbol{x}_{it}$ に対してある種の相関構造を導入し, その挙動に対する一般的表示を与えた. この一般的結果の特別な場合として, 正規分布, 二項分布, ポアソン分布 の場合について,

(i) 時点に依存していない説明変数の漸近相対効率は α の大きさにあまり影響を受けない.

(ii) 時点に依存する説明変数の漸近相対効率は α の大きさとともに減少する.

などを指摘した.

Proving admissibility by the stepwise generalized Bayesian procedure

関東学院大学経済学部　布能英一郎

定理 1. 与えられた事前確率分布に対し、ベイズ推定量が一意的に決定されるとき、それは許容的である。

ところが、自乗損失下でMVUEが許容的だとしても、定理1.1.によって許容性を証明できない。ではMVUEが自乗損失下で許容的だとすれば、その許容性を（更に一般化して、通常のベイズ法では許容性を示せないような許容的推定量の許容性を）どのようにして示せるであろうか? 従来から (1) Limit Bayes (2) Extended Bayes(ϵ -Bayes) (3) generalized Bayes(improper priorを用いたBayes的手法) が考えられてきた。もちろん, このいずれの方法によって求められる推定量も一般には許容性を保証しない。そのため、たとえば、improper priorを用いた場合には、許容性を保証するには仮定条件を強くする必要がある。

定理 2. generalized Bayes推定量が一意的に決定され、かつリスクが有限のとき、許容的である。

こうした方法に対し、Hsuan(1979), Meeden and Ghosh(1981), Brown(1981) らによってステップワイズベイズ法が考え出された。この方法は「母数空間、標本空間を適度直和分割し、各分割上でベイズ解が事前分布から一意に定まれば、全体でも許容的」というものである。ステップワイズベイズ法は、標本空間が有限で母数空間がコンパクトであるような離散型分布において最小分散不偏推定量の自乗損失下での許容性を調べるのに大変強力な手段である。なぜかというと：(a) 完備性定理が成立している。(b) 分布論の立場からは、通常ならimproper priorとなってpriorとなりえないimproper prior distributionに対しrestricted problemを考えることで、これがpriorとしての意味を持つことができる根拠を与えている。(c) 応用面から言えば単純明解に許容性が示せ、多くの統計学の分野に適用可能である。しかし、この方法といえども万能ではなく、一般の母数空間・標本空間では、許容的推定量をステップワイズベイズ法では示せない例がいくつかある。だが、母数空間・標本空間を適度直和分割して考察するとの考え方は、何らかのメリットのあるものと思える。「適度の直和分割」にLimit BayesなりExtended BayesなりGeneralized Bayesを用いることが考えられる。本報告では、直和分割し、各分割上でGeneralized Bayesを用いた場合に、推定量の許容性を証明することを考察する。

定義、記号、仮定条件、準備：標本空間を $\boldsymbol{\mathcal{X}}$, 母数空間をΘ で表記する。以下、離散型確率分布 $P(x|\theta)$ のみを考える。

$\boldsymbol{\mathcal{X}}$ の空でない部分集合 $\boldsymbol{\mathcal{X}}(i)$ に対し、$\Theta(\boldsymbol{\mathcal{X}}(i)) = \{\theta \in \Theta | g_i(\theta) = \sum_{x\in\boldsymbol{\mathcal{X}}} P(x|\theta) > 0\}$ と定める。そうすると、標本空間を $\boldsymbol{\mathcal{X}}(i)$, 母数空間を $\Theta(\boldsymbol{\mathcal{X}}(i))$ とするrestricted probability distribution $P_{\boldsymbol{\mathcal{X}}(i)}(x|\theta)/g_i(\theta)$ がwell-definedである。Θ の空でない部分集合 Θ^* に対して、Θ^* 上で定義されている事前測度 $d\pi(\theta)$ に対し、$\Theta - \Theta^*$ 上でzero massを持つと定める。これにより、$d\pi(\theta)$ は Θ 上で定義された事前測度となる。

定理 3. (The generalized stepwise Bayes procedure)

$\boldsymbol{\mathcal{X}}$ の空でない部分集合の列 $\{\boldsymbol{\mathcal{X}}(i)|i \in I\}$ と，事前測度の列 $\{d\pi(i)|i \in I\}$ が

(i) $\boldsymbol{\mathcal{X}}(1) = \{x \in \boldsymbol{\mathcal{X}} | g(x : \pi_1) > 0\}$, $\boldsymbol{\mathcal{X}}(j) = \{x \in \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{X}}(2) - ... - \boldsymbol{\mathcal{X}}(j-1) | g(x : \pi_j) > 0\}$ for j=2,3,... と置くと、各$i = 1, 2, ...$は $\boldsymbol{\mathcal{X}}$(i) $\neq \emptyset$ で、$\boldsymbol{\mathcal{X}} = \bigcup_{i\in I} \boldsymbol{\mathcal{X}}(i)$

(ii) $\Theta(i)=\{\theta\in\Theta(\boldsymbol{\mathcal{X}}(i)) : d\pi_i$ は positive mass を持つ $\}$ と置くと、$\{\Theta(i)|i\in I\}$ は disjoint.

(iii) 推定量 $\delta(x)$ が各 $(\Theta(i),\boldsymbol{\mathcal{X}}(i))$ 上で事前測度 $d\pi_i(\theta)$ より一意に定まる generalized Bayes で、かつリスクが有限

ならば、$\delta(x)$ は $(\Theta,\boldsymbol{\mathcal{X}})$ で許容的。

ところが、この定理は、有用でない。

例 1. k を任意の自然数とする。$X\sim NegativeBinomial(k,\theta)$, すなわち $P(x|\theta)=\binom{x+k-1}{x}\theta^k(1-\theta)^x$ にて 標本空間 $\boldsymbol{\mathcal{X}}$ を $\boldsymbol{\mathcal{X}}(1)=\{0\}$, $\boldsymbol{\mathcal{X}}(2)=\{1,2,...\}$ に分解する。そして、a sequence of improper priors を $d\pi_1(\theta)=dI_{\{\theta=0\}}(\theta)$, $d\pi_2(\theta)\propto((1-\theta^k)/((1-\theta)\theta)))d\theta$ に選ぶ。定理 3 に従って、各 $(\Theta(i),\boldsymbol{\mathcal{X}}(i))$ 上で generalized Bayes 推定量を計算すると、この推定量 $\delta(x)$ は一意に定まり、$\delta(x)=k/(x+k)$ である。さて、リスクを計算すると、

$$R_{\boldsymbol{\mathcal{X}}(2)}(\theta,\delta)=\sum_{x=1}^{+\infty}(\theta-k/(x+k))^2P_{\boldsymbol{\mathcal{X}}(2)}(x|\theta)=R(\theta,\delta)/(1-\theta^k)-(\theta-1)^2\theta^k/(1-\theta^k),$$

$$\gamma_{(\Theta(2),\boldsymbol{\mathcal{X}}(2))}(d\pi_2,\delta)=\int R_{\boldsymbol{\mathcal{X}}(2)}(\theta,\delta)d\pi_2(\theta)=\gamma(d\pi,\delta)-B(k,2),$$

B(k,2) は有限の値であるから

$$\gamma(d\pi,\delta)<+\infty \quad\Leftrightarrow\quad \gamma_{(\Theta(2),\boldsymbol{\mathcal{X}}(2))}(d\pi_2,\delta)<+\infty \qquad (*)$$

なる関係が示された。よって、もし $\delta(x)=k/(x+k)$ の許容性を「improper prior を使ってリスクが有限」という視点から証明したいのであれば、わざわざ定理 3 を用いる必要がなく、定理 2 で示せる。

例 2. 例 1 の続き。$X\sim NegativeBinomial(k,\theta)$ で、$k\geq 2$ の場合に、$\delta(x)=(k-1)/(x+k-1)$ の自乗損失下での許容性を考察する。標本空間 $\boldsymbol{\mathcal{X}}$ を $\boldsymbol{\mathcal{X}}(1)=\{0\}$, $\boldsymbol{\mathcal{X}}(2)=\{1,2,...\}$ に分解する。そして、a sequence of improper priors を $d\pi_1(\theta)=dI_{\{\theta=0\}}(\theta)$, $d\pi_2(\theta)\propto((1-\theta^k)/((1-\theta)\theta^2)))d\theta$ に選ぶ。定理 3 に従って、各 $(\Theta(i),\boldsymbol{\mathcal{X}}(i))$ 上で generalized Bayes 推定量を計算すると、この推定量 $\delta(x)$ は一意に定まり、$\delta(x)=(k-1)/(x+k-1)$ である。リスクを計算すると、例 1 同様、同値関係 $(*)$ が成り立つ

このことは、例 1, 例 2. に限らない。以下、話を単純にするために Θ を実数直線の区間に限ったが、Θ に対するこの仮定は、本質的なものではない。

定理 4. Θ を実数直線 $\mathbf{R}$ の区間で、少なくとも 1 つの boundary point を持つもとする。すなわち、Θ は $[\theta_0,\theta_1]$, $[\theta_0,\theta_1)$, $(\theta_0,\theta_1]$, $(-\infty,\theta_1]$, $[\theta_0,+\infty)$ のいずれかとする。$d\pi(\theta)$ を与えられた nonnegative measure on $\Theta\subset\mathbf{R}$ とする。$\boldsymbol{\mathcal{X}}(1)=\{x\in\boldsymbol{\mathcal{X}}|\int P(x|\theta)d\pi(\theta)=+\infty\}$, $\boldsymbol{\mathcal{X}}(2)=\boldsymbol{\mathcal{X}}-\boldsymbol{\mathcal{X}}(1)$, $d\pi_1(\theta):\Theta$ の boundary point(s) のみに concentrate した measure, $d\pi_2(\theta)=\{1-\sum_{x\in\boldsymbol{\mathcal{X}}}P(x|\theta)\}d(\theta)$ と定める。そして、各 $x\in\boldsymbol{\mathcal{X}}(1)$ に対して $d\pi(\theta)$ より定まる Generalized Bayes 推定量 $\delta(x)$ が well-defined で、$\boldsymbol{\mathcal{X}}(1)$ が有限集合のとき $(*)$ が成り立つ。

補正関数によるセミパラメトリック回帰

島根大・総合理工 内藤 貫太

1．はじめに 平滑化は数理統計学において現在最も盛んに議論され、その数理科学全般への応用から最も重要な分野の１つであろう。ここでは特に回帰関数の推定に焦点を当てる。回帰問題において広く用いられている手法として多項式回帰があり、その係数は例えば最小２乗法で推定される。これはいわゆるパラメトリックアプローチで、回帰関数が多項式であれば正当化される。一方、そのような構造的仮定を伴わないアプローチとしてノンパラメトリック回帰があり、こちらも広く用いられている。しかしながらノンパラメトリックに推定された構造は解釈が容易ではない。このようにパラメトリック、ノンパラメトリック共に広く用いられている反面それぞれが困難さを抱えているという事実はその両方のアプローチを用いた手法の開発を動機づける。本報告で提案される回帰推定はパラメトリック推定を初期推定と位置づけ、これをノンパラメトリックに補正するというアプローチである。パラメトリック推定を“粗い”推定と見なし、“残差”に対応する部分がノンパラメトリックに推定される。その補正項（関数）の推定は局所的な適合を通して実行される。

2．パラメトリックとノンパラメトリックーその役割分担 数理科学全般において、モデルとは有限個のパラメターでの考察対象の記述であると言えるだろう。統計学において用いられるノンパラメトリック法は一致性を持つが、得られる結果の解釈は容易ではない。我々はパラメターの推定値を通して構造を理解するのであり、ノンパラメトリックのような無限次元では理解しえない。このような観点からは、あくまで解析の主段階はパラメトリックで行い、それを補完する役割がノンパラメトリックに与えられることになる。本報告のセミパラメトリック回帰手法は、このような観点より導かれるものである。

3．回帰推定量 データ $(x_1,y_1),...,(x_n,y_n)$ が密度関数 $f(x,y)=f(x)g(y|x)$ から得られたとする。我々の興味は条件付き平均関数 $m(x)=E[Y|X=x]$ の推定である。まず初期推定量として

$$m(x,\hat{\boldsymbol{\beta}})=\hat{\beta}_0+\hat{\beta}_1x+\cdots+\hat{\beta}_px^p$$

を用意する。ここで、$\hat{\boldsymbol{\beta}}=(\hat{\beta}_0,\hat{\beta}_1,...,\hat{\beta}_p)'$ は$\boldsymbol{\beta}$の推定量である。$m(x)$ の初期近似として $m(x,\hat{\boldsymbol{\beta}})$ を想定し、その近似を補正項ξを用いて $m(x,\hat{\boldsymbol{\beta}})\xi$の型で改良する事を試みる。$\xi$の推定については局所 L_2適合基準

$$Q_n(x,\xi|\alpha)=\sum_{i=1}^{n}K_h(x_i-x)\left\{\frac{y_i}{m(x_i,\hat{\boldsymbol{\beta}})}-\xi\right\}^2 m(x_i,\hat{\boldsymbol{\beta}})^{2-\alpha}$$

のξ−最小化で行う。ここで、$\alpha\geq 0$、$K_h(z)=(1/h)K(z/h)$ で、$K(z)$ は原点対称な密度関数である。局所的な適合は、改良を試みる点xの遠くにあるデータは改良するための情報に乏しいだろうという直感の反映である。簡単な計算から

$$\hat{\xi}=\hat{\xi}(x)=arg-\min_{\xi}Q_n(x,\xi|\alpha)=\frac{\sum_{i=1}^{n}K_h(x_i-x)y_im(x_i,\hat{\boldsymbol{\beta}})^{1-\alpha}}{\sum_{i=1}^{n}K_h(x_i-x)m(x_i,\hat{\boldsymbol{\beta}})^{2-\alpha}}$$

が得られ、$\hat{m}(x) = m(x, \hat{\boldsymbol{\beta}})\hat{\xi}$を$m(x)$の推定量とする。

4．**推定量の挙動** $v(x)^2 = Var[Y|X = x]$とする。$\hat{\boldsymbol{\beta}}$の$\boldsymbol{\beta}$の周りでの展開を用いると、

$$\begin{aligned} E[\hat{m}(x)|X_1, X_2, ..., X_n] &= m(x) + \frac{h^2}{2}\mu_2(K)\left[B_1(x) - \alpha B_2(x)\right] + O_p\left(h^4 + \frac{h}{n} + \frac{1}{n^2}\right) \\ Var[\hat{m}(x)|X_1, X_2, ..., X_n] &= \frac{R(K)v(x)^2}{nhf(x)} + O_p\left(\frac{h}{n}\right) \end{aligned}$$

が得られる。ここに、$\mu_2(K) = \int z^2 K(z)dz$、$R(K) = \int K(z)^2 dz$、

$$\begin{aligned} B_1(x) &= \left[m'' - m_0''\left(\frac{m}{m_0}\right) + 2\left(m' - m_0'\frac{m}{m_0}\right)\left(\frac{m_0'}{m_0} + \frac{f'}{f}\right)\right](x) \\ B_2(x) &= 2\left(m' - m_0'\frac{m}{m_0}\right)\left(\frac{m_0'}{m_0}\right)(x) \\ m_0(x) &= m(x, \boldsymbol{\beta}) \end{aligned}$$

である。バイアスの$O(h^2)$項がαの線型式である事に注目されたい。また$m(x) = m_0(x)$ならば、即ち我々の想定したパラメトリックモデルが正しければ、バイアスの$O(h^2)$項は消えることがわかるだろう。分散の主項はαには依存せず、例えばNadaraya-Watson推定量(cf, Wand and Jones (1995))、局所線型推定量(Fan (1992))といったノンパラメトリック回帰において代表的な推定量と同じである事に注意する。更に$\alpha = 2$のケースがHjort and Glad (1995)で議論された推定量に対応している。

バイアスおよび分散の評価式を用いることにより、回帰推定量の良さの尺度としてのAMISEは

$$\frac{h^4}{4}C_1\left\{d_1\alpha^2 - 2d_2\alpha + d_3\right\} + \frac{C_2}{nh}$$

という骨格を持つ事がわかる。$d_1 > 0$だからAMISEはαについて最小化できて、ただちにHjort and Glad (1995)で議論された推定量より良い推定量が構成される。

5．**考察** ノンパラメトリック回帰推定量、特に、Nadaraya-Watson推定量、局所線型推定量との比較を考察する。

参考文献

FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, 998-1004.

HJORT, N. L. AND GLAD, I. K. (1995). Nonparametric density estimation with a parametric start. *Ann. Statist.*, **23**, 882-904.

WAND, M. P. AND JONES, M. C. (1995). *Kernel Smoothing.*, Chapman & Hall, London.

共分散関数を利用した fractional Gaussian noise および fractional Brownian motion のパラメータ推定

慶大理工　加藤 剛

1 序

$H \in (0,1)$ とし，$\{B_H(t) : t \in \boldsymbol{R}\}$ を，次の 3 つの条件を満たすものとして定義される fractional Brownian motion (fBm) とする．

(i) $B_H(t)$ は実数値 Gauss 過程．

(ii) $B_H(t)$ は a.s. で $t \in \boldsymbol{R}$ の連続関数．

(iii) 任意の $t,\ \tau \in \boldsymbol{R}$ について，増分 $B_H(t+\tau) - B_H(t)$ は，平均 0，分散 $\sigma^2|\tau|^{2H}$ の正規分布に従う．ここで，$\sigma > 0$ は定数である．

パラメータ H は，fractional differencing parameter または Hurst index と呼ばれる．$H = 1/2$ のときが，通常の Brownian motion にあたる．この過程は Hurst (1951) によるナイル川の流量データの解析に誕生の経緯を持ち，最近では，信号解析，画像処理，金融工学といった幅広い分野で応用されている．

fBm は，H と σ^2 の 2 つのパラメータでその挙動が決定される．これらのパラメータの推定については，これまでに，ペリオドグラムを利用した回帰モデルによる方法，最尤法，Allan-variance を用いた方法などが提案されてきた．本報告では，共分散関数の推定を利用して，これらより計算が容易な推定量を構成し，その漸近的な性質を提示する．

2 強一致推定量の構成

定数 $p > 0$ を 1 つ固定し，fBm $\{B_H(t)\}$ をもとに新たな確率過程 $\{X_{H,p}(t) : t \in \boldsymbol{R}\}$ を

$$X_{H,p}(t) := B_H(t+p) - B_H(t) \tag{1}$$

によって定める．特に $p = 1$ で $t \in \boldsymbol{Z}$ であるときは，$\{X_{H,1}(t) : t \in \boldsymbol{Z}\}$ は fractional Gaussian noise (fGn) と呼ばれる．

$\{X_{H,p}(t)\}$ は定常で，その相関関数 $R_{H,p}(\tau)$ について，

$$R_{H,p}(\tau) \sim p^2\sigma^2 H(2H-1)\,\tau^{2H-2} \quad (\tau \to +\infty)$$

という性質を持つ．また，fBm は Gauss 過程なので，$\{X_{H,p}(t)\}$ も Gauss 過程である．この 2 つのことから，共分散関数についての大数の強法則が適用でき，次の結果を導くことができる．

定理 1 $\{X_{H,p}(t) : t \in \boldsymbol{R}\}$ を (1) で定義される確率過程とする．このとき，

$$\widehat{H}_{p,n} := \frac{1}{2\log p} \log \frac{\sum_{k=1}^{n}\{X_{H,p}(k)\}^2}{\sum_{k=1}^{n}\{X_{H,1}(k)\}^2}, \qquad \widehat{\sigma^2_n} := \frac{1}{n}\sum_{k=1}^{n}\{X_{H,1}(k)\}^2$$

は，それぞれ H と σ^2 の強一致推定量になる．

特に $p>0$ が整数のとき，定理 1 を次の形に書き換えることができる．

系 $\{X_{H,1}(t):t\in \boldsymbol{Z}\}$ を，平均 0，分散 σ^2，fractional differencing parameter $H\in(0,1)$ を持つ fGn とする．このとき，

$$\widehat{H}_{p,n}:=\frac{1}{2\log p}\log\frac{\sum_{k=1}^{n}\left\{\sum_{j=k}^{k+p-1}X_{H,1}(j)\right\}^2}{\sum_{k=1}^{n}\{X_{H,1}(k)\}^2},\qquad \widehat{\sigma_n^2}:=\frac{1}{n}\sum_{k=1}^{n}\{X_{H,1}(k)\}^2 \tag{2}$$

は，それぞれ H と σ^2 の強一致推定量になる．

定理 1 および系で提示された推定量 $\widehat{H}_{p,n}$，$\widehat{\sigma_n^2}$ の利点は，データから容易に計算できるということである．実際，計算の本質的な部分は，データの積と和のみである．このため，新たにもう一つのデータが得られたとき，簡単に，素早く推定値を更新することができる．しかも，これらの推定量は，強一致性を持っている．計算機による数値実験によっても，推定量が強一致性を持っていることが確かめられる．また，次の定理 2 で述べるように，ある場合には漸近分布を求めることもできるので，真の値への収束の速さを評価することも可能である．

3 漸近分布

(2) で定義された H と σ^2 の推定量 $\widehat{H}_{2,n}$，$\widehat{\sigma_n^2}$ に対しては，強一致性を示すだけでなく，その漸近分布を求めることもできる．実際，推定量を構成する一部分が $(X_{H,1}(1),X_{H,1}(2),\ldots,X_{H,1}(n+1))$ の 2 次形式で書けることを利用して，2 次形式についての中心極限定理から，次の結果を示すことができる．

定理 2 $f(x)$ を fGn のスペクトル密度

$$f(\lambda)=\frac{\sigma^2H\,\Gamma(2H)\sin(\pi H)}{\pi}|e^{i\lambda}-1|^2\sum_{k=-\infty}^{+\infty}|\lambda+2\pi k|^{-2H-1},\quad \lambda\in[-\pi,\pi]\setminus\{0\},$$

とし，

$$s_1^2=4\pi\int_{-\pi}^{\pi}\{f(\lambda)(2-2^{2H}+2\cos\lambda)\}^2d\lambda,\qquad s_2^2=4\pi\int_{-\pi}^{\pi}\{f(\lambda)\}^2d\lambda$$

とおく．このとき，$H\in(0,3/4)$ ならば，(2) で $p=2$ とした $\widehat{H}_{2,n}$ と $\widehat{\sigma_n^2}$ について，

$$\sqrt{n}(\widehat{H}_{2,n}-H)\xrightarrow{d}N\left(0,\ \left(\frac{s_1}{2^{2H+1}\sigma^2\log 2}\right)^2\right)\quad(n\to+\infty), \tag{3}$$

$$\sqrt{n}(\widehat{\sigma_n^2}-\sigma^2)\xrightarrow{d}N(0,\ s_2^2)\quad(n\to+\infty) \tag{4}$$

が成り立つ．

この定理では $H\in(0,3/4)$，$p=2$ の場合しか述べていないが，$H\in[3/4,1)$ と一般の $p\in\boldsymbol{Z}\setminus\{1\}$ についても，漸近分布が求まることがほぼ分かっている．ただし，$H\in[3/4,1)$ のときは，漸近分布は正規分布にはならない．

Statistical estimation of the proportions of subpopulations through complaining calls

Kumamoto University, Toshio Sakata
Australian National University, Jeferry Xu Yu,

1 Introduction

1.1 Problems

Let all audience of FM radio in an area be grouped into k subpopulations $\Pi_i, i = 1, 2, ..., k$ by their demand for the contents of broadcasting. Let $p_1, p_2, ..., p_k$ with $\sum_{i=1}^{k} p_i = 1$ be the proportions of the subpopulations Π_i. We assume that their demands are mutually exclusive and denoted by $S_i, i = 1, 2, ..., k$. Each day we broadcast $S_i, i = 1, 2, ..., k$ with the ratios of broadcasting time length $q_1, ..., q_k$ where $\sum_{i=1}^{k} q_i = 1$. Now we assume that if a listener switch on the radio and at the moment can not find the content that he want he necessarily makes an complaining call to the broadcasting station. Let $n_i, i = 1, 2, ..., k$ be the number of access from each subpopulation Π_i in a day and so $N = \sum_{i=1}^{k} n_i$ denotes the total number of accesses in a day. Let $Y_i, i = 1, 2, ..., k$ denote the number of complaining calls in a day from each subpopulation Π_i. We have a need to adjust the proportions of broadcasting time lengths to that of the subpopulations. This is because then the mean number of complaining calls becomes smallest. So we have a need to estimate the proportion of subpopulations, (p_i) based on observables.

1.2 Statistical Formulation of the Problem

We assume that the total number of access in a day, N, be a random variable with the Poisson distribution $Po(\lambda)$. For each access number n_i from subpopulation, we assume that given N, $(n_1, n_2, ..., n_k)$ be a random vector with the multinomial distribution $\mathrm{M}(N, p_1, p_2, ..., p_k)$ with $\sum_{i=1}^{k} n_i = N$. For the number of complaining calls from each subpopulation we assume that given $(n_1, n_2, ..., n_k)$, Y_i has the binomial distribution $\mathrm{B}(n_i, 1 - q_i)$. Note that the binomial parameter $1 - q_i$ is reasonable. Further it is assumed that they are conditionally independent.
Under the above model we consider the following estimation problem.
(I-1). Assume that $(n_1, ..., n_k)$ is unobservable and N and $(Y_1, Y_2, \ldots, Y_k)$ are observable. Then estimate $(p_1, p_2, ..., p_k)$ by the maximum likelihood method.
(I-2). Assume that both N and $(n_1, ..., n_k)$ are unobservable and only $(Y_1, Y_2, ..., Y_k)$ is observable. Then estimate $(p_1, p_2, ..., p_k)$ by the maximum likelihood method.
The likelihood equations are derived and solved for the problem I-1 and I-2 respectively.

2 The Problem I-1

2.1 Likelihood equation

$$\frac{Y_i}{p_i} + \frac{\Delta}{\Gamma} q_i = N \quad for \quad all \quad i = 1, 2, ..., k. \tag{1}$$

2.2 Solution of the likelihood equation

Theorem 2.1 *Let Γ be the largest zero of a some polynomial $g(x)$. Then the mle of p_i of Problem I-1 is given by*

$$\hat{p}_i = \frac{Y_i}{\Gamma - (\Delta/N) q_i}, i = 1, 2, ..., k. \tag{2}$$

3 The problem I-2

3.1 Likelihood equation

The likelihood equation for (p_i) becomes

$$\frac{Y_i}{p_i} = \frac{(1 - q_i)\Delta_Y}{(1 - \Gamma)}, \quad i = 1, 2, ..., k. \tag{3}$$

3.2 Solution of the likelihood equation

We obtain that

Theorem 3.1 *The mle of p_i of the problem I-2 is given by*

$$\hat{p}_i = \frac{Y_i/(1 - q_i)}{\sum_{i=1}^{k} Y_i/(1 - q_i)}. \tag{4}$$

3.3 Consistency

An estimator $\hat{\theta}$ is said to be τ-consistent if $\hat{\theta}$ converges to θ in probability when τ goes to ∞. That is, for any $\epsilon > 0$ $\lim_{\tau \to \infty} P\{|\hat{\theta} - \theta| > \epsilon\} = 0$. Let m denote the number of days and λ be a mean access number per day. For the problem I-1 and I-2 we can show both λ-consistency and m-consistency of the mle.

Reference

[1] Kain-Lee Tan and Xu.Y.Jeferry(1996). Energy efficient filtering of nonuniform broadcast. In *Proceedings of the 16th IEEE International Conference on Distributed Computing Systems.*

密度推定による対称性の検定

大阪府立大・工　　水嶋　高正

$X_1, X_2, \ldots$が互いに独立に同じ連続分布に従うとし，その確率密度関数$f(x)$は連続であるとする．このとき

帰無仮説H_0：$f(x) \equiv f(-x)$

対立仮説H_1：$f(x) \not\equiv f(-x)$

を検定する問題について考える．この問題に対して，$f(x)$の$x=0$についての対称性の尺度

$$I_f = \int \{f(x) - f(-x)\}^2 dx$$

を考える．$f(x) \equiv f(-x)$と$I_f = 0$とは同値であるので，I_fの推定に基づいた検定法を考える．

I_fの推定には，$f(x)$のkernel estimator

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - X_i}{h})$$

を用いる．ただし，推定に必要なbandwidth hと核関数$K(x)$は適当な条件を満たすものとする．I_fは

$$I_f = 2 \int \{f(x) - f(-x)\} f(x) dx = 2\{\mathrm{E}f(X) - \mathrm{E}f(-X)\}$$

と表されるので，$\hat{f}(x)$を用いると

$$\tilde{I}_f = \frac{2}{n} \sum_{i=1}^{n} \{\hat{f}(X_i) - \hat{f}(-X_i)\} = \frac{2}{n^2 h} \sum_{i=1}^{n} \sum_{j=1}^{n} \{K(\frac{X_i - X_j}{h}) - K(\frac{-X_i - X_j}{h})\}$$

がI_fの推定量である．また，$i=j$の項を取り除いて

$$\hat{I}_f = \frac{2}{n(n-1)h} \sum_{i=1}^{n} \sum_{j \neq i} \{K(\frac{X_i - X_j}{h}) - K(\frac{-X_i - X_j}{h})\}$$

という推定量も考えられる．

H_0のもとで，$\hat{I}_f$と$\tilde{I}_f$の分布は，Hall(1984)の定理を用いると漸近的に正規分布であるとわかる（[5]参照）．従って，$z_{1-\alpha}$を標準正規分布の上側確率α点とすると，有意水準αの棄却域は

$$\frac{n\sqrt{h}\hat{I}_f}{4\sqrt{R(f)R(K)}} \geq z_{1-\alpha} \quad \text{または} \quad \frac{n\sqrt{h}\{\tilde{I}_f - \frac{2}{nh}R(K)\}}{4\sqrt{R(f)R(K)}} \geq z_{1-\alpha}$$

となる．

同様の検定に対して，signed rank testを用いることができる（[1]参照）．そこで，$\hat{I}_f$や$\tilde{I}_f$を用いた検定とsigned rank testの検出力を比較するため，反復回数1万回のモンテカルロシミュレーションを行った．真の分布が正規分布及び両側指数分布の場合，signed rank testが優れているが，コーシー分布の場合は$\hat{I}_f$による検定が優れているという結果を得た．また，対立仮説$\mathrm{H}_1 : f(x) \neq f(-x)$に属する分布の中で，signed rank testが検出力がαとほとんど変わらないものがあることがわかった．例えば，非負の実数c_0に対して

$$B(x,4) = \begin{cases} 140x^3(1-x)^3 & 0 \leq x \leq 1 \\ 0 & \text{その他} \end{cases}$$

$$f_{B0}(x) = B(x,4)/2 + B(-x,4)/2$$

$$f_{B1}(x) = c_0\{3B(3x,4)/2 + 3B(3x-2,4)/2 - 3B(3x-1,4)\}$$

とする．$0 < c_0 \leq 1/6$のとき，$f_B(x) = f_{B0}(x) + f_{B1}(x)$は$\mathrm{H}_1$に属している．$f_B(x)$の場合，signed rank testに用いられる検定統計量のH_1のもとでの期待値がH_0のもとでの期待値と一致し，検出力はαに近い値である．しかしながら，同じ分布に対して，$\hat{I}_f$や$\tilde{I}_f$を用いた検定では，αより大きな検出力を与えるので，signed rank testの欠点を克服していると言える．

また，nに依存する対立仮説について漸近的な検出力に基づく漸近効率を比較する．標本の大きさnを無限大に近づけたとき，$\hat{I}_f$や$\tilde{I}_f$による検定の検出力は1に収束するが，signed rank testの検出力は1より小さい値に収束することがある．よって，この場合において漸近効率は無限大となり，$\hat{I}_f$や$\tilde{I}_f$による検定が優れているといえる．さらに，検出力が共に1に収束する場合でも，漸近効率において$\hat{I}_f$や$\tilde{I}_f$による検定が優れている状況がある．

参考文献

[1] Gibbons,J.D. and Chakraborti,S.(1992). Nonparametric statistical inference, 3rd ed., Dekker, New York.

[2] Hall,P.(1984). Central limit theorem for integrated squared error of mutivariate nonparametric density estimators, Journal of Multivariate Analysis, 14, 1-16.

[3] Hall,P. and Marron,J.S.(1987). Estimation of integrated squared density derivatives, Statistics & Probability Letters 6, 109-115.

[4] Jones,M.C. and Sheather,S.J.(1991). Using non-stochastic terms to advantage in kernel-based estimation of integrated squared derivatives, Statistics & Probability Letters, 11, 511-514.

[5] Mizushima,T. and Nagao,H.(1998). A test for symmetry based on density estimates, Journal of Japan Statistical Society, 28, 125-145.

[6] Scott,D.W.(1992). Mutivariate Density Estimation, Wiley, New York.

[7] Silverman,B.W.(1986). Density Estimation for Statistics and Data Analysis, Chapman and Hall, London.

レーマン対立仮説を検定するランクの対数尤度比の展開

長崎大学・経　永井　圭二

1　序

Woodroofe(1983) は、Savage-Sethuraman(1966) による二標本問題でレーマン対立仮説を検定する順位による逐次確率比検定に対し、非線形更新定理が用いられるかどうかという問題を提起した。この報告では今までの研究者とは異なる方法で順位の対数尤度比を Chernoff-Savage 的に高次に展開し、その問題を解決する。

2　Lehmann 対立仮説の順位による尤度比検定

独立な確率変数列 $(X_1,\ldots,X_n,Y_1,\ldots,Y_n)$ が観測されるものとする。ここで X_i は分布 F に従い、Y_i は分布 G に従うものとする。ここでレーマン対立仮説の検定 $H_0: G=F$ vs $H_1: G=F^{\Delta}, \Delta>0$ を考える。Savage (1956) によれば、順位の対数尤度比はつぎのように書ける。

$$l_n = n\log\Delta + n\int_{-\infty}^{\infty}\log\left(\frac{F_n+G_n}{F_n+\Delta G_n}\right)d(F_n+G_n). \tag{2.1}$$

Lai(1975) は l_n を Chernoff-Savage 統計量と見なして、ランダムウォーク（独立同一な確率変数の和）と残余項の和に書いた。すなわち、

$$\begin{aligned} l_n &= S_n+\xi_n, \\ S_n &= nS(\Delta,F,G)+n(1-\Delta)\int_{-\infty}^{\infty}\frac{G_n-G}{F+\Delta G}dF+n(\Delta-1)\int_{-\infty}^{\infty}\frac{F_n-F}{F+\Delta G}dG, \end{aligned} \tag{2.2}$$

ここで $S(\Delta,F,G)$ は Kulback-Leibler 情報量で

$$S(\Delta,F,G)=\log(\Delta)+\int_{-\infty}^{\infty}\log\left(\frac{F+G}{F+\Delta G}\right)d(F+G).$$

残余項に対して Lai はどのような小さな $\mu>0$ に対しても $n^{-\mu}\xi_n\to 0$ という漸近的な結果を得た。これに対し、U 統計量の理論を用いることにより、次の結果が得られる。

レーマン対立仮説 $G=F^{\Delta}$ を検定する順位の対数尤度比 l_n はランダムウォークと残余項の和に書ける；$l_n=S_n+\xi_n$. ここで S_n は (2.2) で定義されたものと同じである。残余項 ξ_n は次の関係を満足する。

$$E\Big[\max_{n\le k\le 2n}\Big(\xi_k-c^*(k)\Big)^2\Big]=O(\log n), \tag{2.3}$$

$c^*(n)$ は次のような数列である。$H=\frac{F+G}{2}$ として、

$$\begin{aligned} c^*(n) &= \int_{H>n^{-1}}\left(\frac{1}{F+G}-\frac{1}{F+\Delta G}\right)(1-F)dF+\int_{H>n^{-1}}\left(\frac{1}{F+G}-\frac{\Delta}{F+\Delta G}\right)(1-G)dG \\ &\quad+\int_{H>n^{-1}}\left\{\left(\frac{-1}{(F+G)^2}+\frac{1}{(F+\Delta G)^2}\right)F(1-F)+\left(\frac{-1}{(F+G)^2}+\frac{\Delta^2}{(F+\Delta G)^2}\right)G(1-G)\right\}dH. \end{aligned}$$

さらに G が実際にレーマン対立仮説であるとすると、すなわち $A \neq 1$ に対して $G = F^A$ とすると、

$$\{\max_{0\leq j\leq n} |\xi_{n+j}|, n \geq 1\} \quad \text{は一様可積分、、} \quad \xi_n \to V \quad \text{弱収束、} (\ n \to \infty) \ . \tag{2.4}$$

ここで V はある可積分な確率変数である。

3 順位による逐次確率比検定の期待標本数の漸近展開

Savage and Sethuraman (1966) は順位による逐次確率比検定 $N = \inf\{n \geq 1 : l_n < a \text{ or } l_n > b\}$ $(a < 0 < b)$ を定義した。Berk (1973) によれば、その期待標本数の 1 次の漸近近似は

$$\begin{aligned} EN &= \frac{b}{S(\Delta, F, G)}(1 + o(1)) \quad \text{if } S(\Delta, F, G) > 0 \ , \\ EN &= \frac{|a|}{|S(\Delta, F, G)|}(1 + o(1)) \quad \text{if } S(\Delta, F, G) < 0, \end{aligned}$$

によって与えられる。$(\min(|a|, b) \to \infty)$ Woodroofe (1983) は、この順位による逐次確率比検定で第一種と第二種の誤りを犯す確率の 2 次の近似を求めたが、同時に非線形更新定理を期待標本数の 2 次の漸近展開に使えるかどうかという問題を提起した。その解答として次の結果を得る。

もし $S(\Delta, F, G) > 0$ ならば、

$$E(N) = \frac{b - c^*(b)}{S(\Delta, F, G)} + o(\log b) \qquad (\ |a|, b \to \infty). \tag{3.1}$$

$G = F$ の場合、

$$EN = \frac{|a|}{|S(\Delta, F, F)|} + \frac{(c^{**} + o(1))}{|S(\Delta, F, F)|} \log |a| \quad (|a|, b \to \infty). \tag{3.2}$$

ここで $c^{**} = \frac{1}{2}(1 - \Delta)^2/(1 + \Delta)^2$。さらに $G = F^A$、$A \neq 1$、かつ $S(\Delta, F, F^A) > 0$ とすると、

$$E(N) = \frac{b + r - EV}{S(\Delta, F, F^A)} + o(1) \quad (\ |a|, b \to \infty). \tag{3.3}$$

ここで、$r = ES^2_{\tau_+}/(2ES_{\tau_+})$、$\tau_+ = \inf\{n;\ S_n > 0\}$ である。

参考文献

[1] Robert H. Berk. Some asymptotic aspects of sequential analysis. *Ann. Statist.*, 1:1126–1138, 1973.

[2] Herman Chernoff and I. Richard Savage. Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.*, 29:972–994, 1958.

[3] I. Richard Savage. Contributions to the theory of rank order statistics—the two-sample case. *Ann. Math. Statist.*, 27:590–615, 1956.

[4] I. Richard Savage and J. Sethuraman. Stopping time of a rank-order sequential probability ratio test based on lehmann alternatives. *Ann. Math. Statist.*, 37:1154–1160, 1966.

[5] Michael Woodroofe. On sequential rank tests. In *Recent advances in statistics*, pages 115–140. Academic Press, New York, 1983.

Two Classes of Transformation Models and Rank Estimation

成城大学経済学部　塚原英敦

Abstract

In the literature the following two classes of models have been studied under the name of "transformation model". One is expressed in terms of distribution function(df):

$$X \sim G_\theta = D(F(\cdot);\theta), \tag{0.1}$$

where $D(\cdot\,;\theta)$ is a known continuous df on $(0,1)$ and F is an arbitrary baseline df. θ is a parameter whose values are in some parameter space $\Theta \subset \mathbb{R}$. This model for the two-sample problem is studied in Dabrowska, Doksum and Miura (1989) and, with censored data, in Tsukahara (1991). The above two papers considered semiparametric estimation of θ based on ranks.

The other is expressed in terms of random variable(rv):

$$h(X) = \nu(\theta) + \epsilon, \tag{0.2}$$

where h is an unknown strictly monotone increasing function and ϵ is distributed according to Ψ which is a known df. $\nu(\theta)$ is a function of θ such as $\log\theta$ and is often connected by the linear model $\nu(\theta) = \boldsymbol{\beta}'\boldsymbol{z}$. With this linear form, various methods of estimation of regression parameters $\boldsymbol{\beta}$ are suggested.

We call (0.1) *Lehmann alternative transformation* (LAT) model and (0.2) *location transformation*(LOT) model. The same word "transformation" is used, but in LOT model, transformation acts on the sample space, while in LAT model it acts on the space of probability distributions. One can immediately see that a LOT model is always rewritten as a LAT model. This class of the models includes several important models such as the proportional hazards model. For a given LAT model to be rewritable as a LOT model, the following condition is sufficient (and essentially necessary):

$$D[D(t;\theta_1);\theta_2] = D(t;\theta_1\theta_2), \quad t \in (0,1), \quad \theta_1,\theta_2 \in (0,\infty). \tag{0.3}$$

We consider the following model: the X_i, $i = 1,\ldots,n$ are independent random variables, each of which has df $G_i(x) = D(F(x);\lambda(\boldsymbol{\beta},\boldsymbol{z}_i))$ where $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)'$ is a regression parameter and $\boldsymbol{z}_i = (z_{i1},\ldots,z_{ip})'$ is a vector of covariates. The most popular form of λ is $\lambda(\boldsymbol{\beta},\boldsymbol{z}) = \exp\{\boldsymbol{\beta}'\boldsymbol{z}\}$, which gives a LOT model $h(X) = \boldsymbol{\beta}'\boldsymbol{z} + \epsilon$ if D satisfies (0.3). We then give a short review of previous works: Pettitt (1982)'s quadratic approximation of partial likelihood, Doksum (1987)'s likelihood sampler, and Pettitt (1987)'s least rank mean square. And we extend those methods to LAT model framework to derive estimators of regression parameter $\boldsymbol{\beta}$. The estimator we propose is a generalization of the rank approximate M(RAM) estimator obtained in Dabrowska, Doksum and Miura(1989) for two-sample problem and in Cuzick(1988) for LOT model. It is defined as follows: When F is known, the full log-likelihood turns out to be $l(\boldsymbol{\beta}) \triangleq \log L(\boldsymbol{\beta}) = \sum_{i=1}^n \log d(F(x_i);\lambda(\boldsymbol{\beta},\boldsymbol{z}_i))f(x_i)$, where f is the density of F. The likelihood equation is then given by

$$\sum_{i=1}^n \dot{\lambda}(\boldsymbol{\beta},\boldsymbol{z}_i)\frac{\dot{d}(F(x_i);\lambda(\boldsymbol{\beta},\boldsymbol{z}_i))}{d(F(x_i);\lambda(\boldsymbol{\beta},\boldsymbol{z}_i))} = \mathbf{0}.$$

Replacing $\dot{d}/d$ by any estimating function ϕ satisfying $\mathbf{E}_\beta[\phi(V_i; \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i))] = 0$ where $V_i \sim D(\cdot\,; \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i))$, an M-estimate is defined to be the solution to

$$\sum_{i=1}^{n} \dot{\lambda}(\boldsymbol{\beta}, \boldsymbol{z}_i)\phi(F(x_i); \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i)) = \mathbf{0}.$$

In the case of unknown F, we shall replace F by its estimate $\mathbb{F}_n^\beta$ given as follows: define

$$\bar{\mathbb{G}}_n(x) \triangleq \frac{1}{n+1}\sum_{i=1}^{n} I_{[X_i \le x]}, \quad \bar{G}_\beta(x) \triangleq \frac{1}{n}\sum_{i=1}^{n} D(F(x); \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i)),$$

$$\bar{D}_\beta(t) \triangleq \frac{1}{n}\sum_{i=1}^{n} D(t; \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i)).$$

Then we have $\bar{G}_\beta(x) = \bar{D}_\beta(F(x))$. This indicates that F may be estimated by

$$\mathbb{F}_n^\beta(x) \triangleq \bar{D}_\beta^{-1}(\bar{\mathbb{G}}_n(x)).$$

Set $\widehat{V}_i \triangleq \mathbb{F}_n^\beta(X_i) = \bar{D}_\beta^{-1}(R_i/(n+1))$. Then a RAM estimate $\widehat{\boldsymbol{\beta}}_{RAM}$ is defined to be the solution to

$$\sum_{i=1}^{n} \dot{\lambda}(\boldsymbol{\beta}, \boldsymbol{z}_i)\phi(\widehat{V}_i; \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i)) = \mathbf{0}.$$

Note that $\bar{\mathbb{G}}_n(X_i) = 1/(n+1)R_i$, so that the estimate may be viewed as an M-estimate based on ranks.

Monte Carlo simulation is performed to compare the empirical properties of several estimators for the Cox model. When β is near zero, all estimates have similar nice performance. But for large values of β, $\widehat{\beta}_Q$, $\widehat{\beta}_{RAM}$ and $\widehat{\beta}_M$ have large negative bias. It seems natural for β, $\widehat{\beta}_Q$, $\widehat{\beta}_M$ because these methods are based on local approximation to $L_\tau(\beta)$ near $\beta = 0$. On the other hand, although the RAM estimate is not derived from local approximation to the rank likelihood, its behavior is not good for large values of β in this experiment.

References

[1] Cuzick, J.(1988). Rank regression, *Ann. Statist.*, **16**, 1369-1389.

[2] Dabrowska, D. M., Doksum, K. A. and Miura, R.(1989). Rank estimates in a class of semiparametric two-sample models, *Ann. Inst. Statist. Math.*, **41**, 63-79.

[3] Doksum, K. A.(1987). An extension of partial likelihood methods for proportional hazard models to general transformation models, *Ann. Statist.*, **15**, 325-345.

[4] Pettitt, A. N.(1982). Inference for the linear model using a likelihood based on ranks, *J. Roy. Statist. Soc. Ser. B*, **44**, 234-243.

[5] Pettitt, A. N.(1987). Estimation for a regression parameter using ranks, *J. Roy. Statist. Soc. Ser. B*, **49**, 58-67.

[6] Tsukahara, H.(1992). A rank estimator in the two-sample transformation model with randomly censored data, *Ann. Inst. Statist. Math.*, **44**, 313-333.

楕円分布モデルの推定とEMアルゴリズム

立教大学社会学部　山口和範

1　はじめに

Dempster *et al.*(1977)により「EMアルゴリズム」という名前による統一が行われ、その命名以来、EMアルゴリズムは欠測値の処理をはじめとして様々な分野で幅広く使用されるようになった(Meng & van Dyk 1997の1.2節参照)。また一方で、その収束の遅さが指摘され、80年代にはLouis(1982)等による加速化も提唱された。さらに、90年代に入りRubinを中心としたグループが様々な拡張を行っている。本報告では、90年代に行われたEMアルゴリズムの拡張と最近の成果を、楕円分布モデルにおける推定との関連で紹介し、2段階推定法の推定効率や収束スピードをシミュレーションにより調べた。

2　シミュレーション

ここでは、簡単な共分散構造モデルを使ったシミュレーションにより、2段階法とOptimal EMアルゴリズムの効果を調べてみる。いま、$\boldsymbol{Y}$を4次の確率ベクトルで、その平均と$\boldsymbol{\mu}$、共分散行列$\boldsymbol{\Sigma}$が$\mu = \mathbf{0}$, $\boldsymbol{\Sigma} = \sigma_1\mathbf{1} + \sigma_2\boldsymbol{I}$であるとする。ここで、$\mathbf{0}$はゼロベクトルで, $\mathbf{1}$はすべての成分が1のこのシミュレーションではσ_1とσ_2を共に0.5とし、母集団分布として正規分布と自由度4のt分布を用いる。標本の大きさは100とし、10000個の標本をそれぞれの分布で作成し、次の4つの推定計算を行う。

- 通常のEMによるMLE(MLE),
- optimal EMによるMLE(OEM),
- 通常のEMによる2段階推定(TSM),
- optimal EMによる2段階推定(OTS).

表1と表2がシミュレーション結果である。推定値の特性にはほとんど差はみられない。反復数は、先の例題同様Optimal EMを使ったほうがかなり少なくなっている。とくに、MLEの場合は半分以下の反復数になっている。また、普通にMLEを求める場合に比べ、2段階推定の方が若干反復回数が多くなっており、optimal EMを使用する効果も若干薄れているようだ。

表1: 推定値の基本統計量

Dist.	Normal distribution				t distribution			
Method	MLE	OEM	TSM	OTS	MLE	OEM	TSM	OTS
Mean	0.497	0.497	0.496	0.496	0.496	0.496	0.496	0.496
Std.DEV.	0.0545	0.0545	0.0545	0.0545	0.0568	0.0568	0.0570	0.0570
Skewness	-0.226	-0.226	-0.225	-0.225	-0.186	-0.186	-0.187	-0.187
Kutorsis	0.051	0.051	0.051	0.051	-0.027	-0.027	-0.026	-0.026
MIN	0.259	0.259	0.263	0.263	0.256	0.256	0.250	0.250
25%	0.460	0.460	0.460	0.460	0.458	0.458	0.458	0.458
Median	0.499	0.499	0.498	0.498	0.499	0.499	0.498	0.498
75%	0.534	0.534	0.533	0.533	0.536	0.536	0.536	0.536
MAX	0.667	0.667	0.670	0.670	0.692	0.692	0.694	0.694

表2: 反復回数

Dist.	Normal distribution				t distribution			
Method	MLE	OEM	TSM	OTS	MLE	OEM	TSM	OTS
Mean	17.46	7.69	18.89	10.93	16.89	7.87	18.59	11.07
MIN	9	5	11	8	6	5	9	8
MAX	20	10	34	19	22	11	36	19

参考文献

Dempster, A.P., Laird, N.M. and Rubin, D.B.(1977), Maximum likelihood from incomplete data via the EM algorithm(with Discussion), *Journal of The Royal Statistical Society* **B 39**, 1-38.

Louis, T.A.(1982).Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society* **B 44**, 226-233.

Meng, X.L. and van Dyk, D.(1997). The EM algorithm - an old folk-song sung to a fast new tune (with discussion), *Journal of the Royal Statistical Society* **B 59**, 511-567.