(16)「非正規制での統計理論とその応用」に関する研究報告

序 — ノンパラメトリック，ロバスト統計量，離散分布，B-スプライン，ウェーブレット

横浜市立大学大学院・総合理学研究科 白石 高章

1 はじめに

（1）ノンパラメトリック，ロバスト統計量 ——— 自分の考えと今後の研究
（2）離散分布 ——— 分割表、多項分布におけるパラメータの推定
（3）B-スプライン，ウェーブレット ——— 数学的解説

2 ノンパラメトリック，ロバスト統計量

○分布によらない検定法 ——— 順位検定、並べ替えM検定
○推定法 ——— 最小自乗法、順位推定法、分布の近傍での頑健推定法
◎並べ替えM検定と頑健推定法の数学的最良性は漸近理論によって証明されているが、計算機シミュレーションによって小標本の場合にもこれらの頑健手法が良いことが検証できる。
◎観測値の従っている連続分布を特定することは難しい。
　　理由：適合度検定は検出力が低い。
◎観測値の従っている分布がある特定の分布の近傍に入っていると判定し、そこでの頑健手法を選択し統計解析する．あらゆるデータに適用可能,外れ値にも考慮されたことになる．1次元に対してはかなり計算機による手法の良さの検証は可能

3 離散分布

分割表において検定統計量を使ってパラメータの縮小推定量 (shrinkage estimator) を提案することができ、通常の推定量を改良していることを漸近理論によって証明できる。もっと一般に多項分布のパラメータの縮小推定量の理論ができる。

4 B-スプライン

$N_1(x) = I_{[0,1)}$　　($[0,1)$ 上の特性関数)
m次のカーディナルB-スプライン関数

$$N_m(x) \equiv \int_{-\infty}^{\infty} N_{m-1}(x-t)N_1(t)dt = \int_0^1 N_{m-1}(x-t)dt$$

$N_m(x)$ は次の (i) から (v) の性質を持つ。

(i) supp $N(x) = [0, m]$

(ii) $0 < x < m$ なる $x \in R$ に対し $N_m(x) > 0$

(iii) $N_m(\frac{m}{2} + x) = N_m(\frac{m}{2} - x) \quad x \in \mathbb{R}$

(iv) $N_m(x) = \frac{1}{(m-1)!}\sum_{k=0}^{m}(-1)^k\binom{m}{k}[\max\{0, x-k\}]^{m-1}$

(v) $N_m(x) \in C^{m-2}$ かつ 任意の $k \in \mathbb{Z}$ にたいして $[k, k+1)$ 上では $N_m(x)$ は x の $m-1$ 次の多項式

$$V_j^m \equiv \overline{span\{2^{\frac{j}{2}}N_m(2^jx - k) : k \in Z\}} \qquad (\text{空間 } L^2(\mathbb{R}) \text{ 上の閉包})$$

とおくとき,

$$\cdots \subset V_{j-1}^m \subset V_j^m \subset V_{j+1}^m \subset \cdots$$

$$\bigcap_{j \in \mathbb{Z}} V_j^m = \{0\}, \quad \overline{\bigcup_{j \in \mathbb{Z}} V_j^m} = L^2(\mathbb{R})$$

B スプラインは $L^2(\mathbb{R})$ の基底ではあるが直交基底ではない。

5 スプラインウェーブレット

$$P_1(\xi) = 1,\ P_N(\xi) = (\cos\xi)P_{N-1}(\xi) - \frac{1}{N}(\sin\xi)P'_{N-1}(\xi)$$

$$\widehat{\varphi^n}(\xi) = \begin{cases} (\frac{\sin(\frac{1}{2}\xi)}{\frac{1}{2}\xi})^{n+1}\frac{1}{\sqrt{P_{2n+1}(\frac{1}{2}\xi)}} & (n \text{ が奇数}) \\ e^{-i\frac{\xi}{2}}(\frac{\sin(\frac{1}{2}\xi)}{\frac{1}{2}\xi})^{n+1}\frac{1}{\sqrt{P_{2n+1}(\frac{1}{2}\xi)}} & (n \text{ が偶数}) \end{cases}$$

$$\widehat{\psi^n}(2\xi) = \begin{cases} e^{i\xi}\frac{(\sin(\frac{1}{2}\xi))^{2n+2}}{(\frac{1}{2}\xi)^{n+1}}\sqrt{\frac{P_{2n+1}(\frac{1}{2}\xi+\frac{1}{2}\pi)}{P_{2n+1}(\xi)P_{2n+1}(\frac{1}{2}\xi)}} & (n \text{ が奇数}) \\ -ie^{i\xi}\frac{(\sin(\frac{1}{2}\xi))^{2n+2}}{(\frac{1}{2}\xi)^{n+1}}\sqrt{\frac{P_{2n+1}(\frac{1}{2}\xi+\frac{1}{2}\pi)}{P_{2n+1}(\xi)P_{2n+1}(\frac{1}{2}\xi)}} & (n \text{ が偶数}) \end{cases}$$

フーリェ逆変換の公式により, $\widehat{\varphi^n}$, $\widehat{\psi^n}$ から φ^n, ψ^n が求まる。

このとき, $\varphi^n(x) \in C^{n-1}$ かつ 任意の $k \in \mathbb{Z}$ にたいして $[k, k+1)$ 上では $\varphi^n(x)$ は x の n 次の多項式, $\psi^n(\frac{x}{2}) \in C^{n-1}$ かつ 任意の $k \in \mathbb{Z}$ にたいして $[k, k+1)$ 上では $\psi^n(\frac{x}{2})$ は x の n 次の多項式。 $\psi^n(\cdot)$ を $L^2(\boldsymbol{R})$ の n 次のスプラインウェーブレットと呼ばれている。

その他のウェーブレットとして, ハールウェーブレットやドベッシーのウェーブレットがあり、$L^2(\boldsymbol{R})$ の正規直交基底になっている。さらに, 無条件基底になっているウェーブレットが多い。

Proportional Reduction in Variation Measure for Contingency Tables

東京理科大学・理工　矢島僚太郎
東京理科大学・理工　宮本　暢子
東京理科大学・理工　富澤　貞男

説明変数 X と, 応答変数 Y からなる 2 元分割表において, X の値が与えられたときの Y の条件付分布に対する変動が, Y の周辺分布に対する変動よりもどの程度小さくなっているかを測る尺度 (proportional reduction in variation (PRV) 尺度) は, Agresti (1990, P.24) で述べられているように, 一般に

$$\frac{V(Y) - E[V(Y|X)]}{V(Y)}$$

のように与えられる. ここに $V(Y)$ は Y の周辺分布に対する変動を表し, $E[V(Y|X)]$ は X の分布に関して取られた条件付変動の平均である. Goodman and Kruskal (1954) は, 変動の指標 $V(\cdot)$ に Gini concentration を用いた concentration coefficient と呼ばれる PRV 尺度 τ を導入し, Theil (1970) は, 変動の指標に Shannon entropy を用いた uncertainty coefficient と呼ばれる PRV 尺度 U を導入した. また, Tomizawa, Seo and Ebi (1997) は, τ と U を含む一般化した PRV 尺度 $T^{(\lambda)}(\lambda > -1)$ を導入した. ここに尺度 τ, U, $T^{(\lambda)}$ は, X と Y に順序のないカテゴリをもつ分割表に適用される. 本講演の目的は, 説明変数 X に順序がなく, 応答変数 Y に順序のあるカテゴリをもつ 2 元分割表において, PRV 尺度を導入することである.

2 元 $R \times C$ 分割表において, $\Pr(X = i, Y = j) = p_{ij}$, $\Pr(X = i) = p_{i\cdot}$, $\Pr(Y = j) = p_{\cdot j}$ $(i = 1, 2, \ldots, R;\ j = 1, 2, \ldots, C)$ とする. すべての i と j に対して, $p_{i\cdot} > 0$, $p_{\cdot j} \neq 1$ を仮定して, PRV 尺度を次のように導入する: $\lambda > -1$ に対して,

$$\psi^{(\lambda)} = \frac{\sum_{j=1}^{C-1}[1 - \sum_{k=1}^{2}(F_{\cdot j}^{(k)})^{\lambda+1}] - \sum_{i=1}^{R}\sum_{j=1}^{C-1} p_{i\cdot}[1 - \sum_{k=1}^{2}(F_{ij}^{(k)}/p_{i\cdot})^{\lambda+1}]}{\sum_{j=1}^{C-1}[1 - \sum_{k=1}^{2}(F_{\cdot j}^{(k)})^{\lambda+1}]},$$

ただし,

$$\psi^{(0)} = \lim_{\lambda \to 0} \psi^{(\lambda)},$$

$$F_{ij}^{(1)} = \sum_{t=1}^{j} p_{it}, \quad F_{ij}^{(2)} = \sum_{t=j+1}^{C} p_{it}, \quad F_{.j}^{(1)} = \sum_{t=1}^{j} p_{.t}, \quad F_{.j}^{(2)} = \sum_{t=j+1}^{C} p_{.t}$$

である. また $\psi^{(\lambda)}$ は, Power-divergence を用いて表すこともできる. この尺度に用いた変動の指標は Patil and Taillie (1982) の diversity index の和である. 尺度は $0 \leq \psi^{(\lambda)} \leq 1$ であり, 任意の λ (> -1) に対して (i) $\psi^{(\lambda)} = 0$ であるための必要十分条件は, X と Y が独立であり, (ii) $\psi^{(\lambda)} = 1$ であるための必要十分条件は, 各 i に対して $p_{ij}/p_{i.} = 1$ となる j が存在することである. 尺度の信頼区間などの詳細は当日報告した.

参考文献

[1] Agresti (1990). *Categorical Data Analysis*, Wiley.

[2] Goodman and Kruskal (1954). *Journal of the American Statistical Association*, **49**, 732-764.

[3] Patil and Taillie (1982). *Journal of the American Statistical Association*, **77**, 548-561.

[4] Theil (1970). *American Journal of Sociology*, **76**, 103-154.

[5] Tomizawa, Seo and Ebi (1997). *Behaviormetrika*, **24**, 193-201.

球面対称分布のもとでのスタイン現象について

丸山 祐造
東京大学・空間情報科学研究センター

統計的決定理論の枠組みで推定問題を考えるとき，もっとも興味深いトピックーつはスタイン現象の解明である．スタイン現象とは自然な推定量 (MLE，UMVUE，MRE など) が非許容的である現象のことであり，Stein(1956) が p 次元正規分布の平均ベクトルの推定問題で，ミニマクス推定量でもある最尤推定量が $p \geq 3$ のとき非許容的であることを示したことに端を発している．さらに Brown(1966) はスタイン現象を非正規の場合にも拡張しており，位置母数の同時推定問題を考えるとき，最良共変推定量がパラメータの次元が 3 以上の場合に非許容的であることを示した．

統計的決定理論の立場からは，非許容的な推定量があればそれを改良することが重要なテーマである．また許容的な推定量，あるいは推定量のクラスを構成することもまた重要なテーマである．我々が興味があるのは，非許容的な推定量を改良していてかつ許容的な推定量を提案することである．このことは推定問題に対して，最後まで責任をとるという意味で非常に重要である．なぜならば非許容的な推定量を改良する推定量を提案しても，またその推定量が非許容的であれば再びその推定量を改良するべきで，その意味では数学的により難しい問題を作り出しただけだからである．

このような視点でスタイン問題を考えるとき，正規性のもとではミニマクスで (上記の問題においては，最良共変推定量を改良する，ということと同義) かつ許容的な推定量が得られているが，非正規のもとではほとんど考えられていない．もちろん一般的な設定のままでは難しいので，我々は特に球面対称分布の平均ベクトルの推定問題を考える．具体的には p 次元ベクトル X が密度関数 $f(\|x-\theta\|^2)$ をもつとき，平均ベクトル θ を自乗損失関数 $L(\theta, d) = \|\delta - \theta\|^2$ のもとで推定する問題を考える．

この問題に関する過去に得られた結果は，ミニマクスであるための十分条件に関するものがほとんどであり，それらは次のページの表のように整理することが出来る (縮小型推定量 $\delta_\phi(X) = (1 - \phi(\|X\|^2)/\|X\|^2)X$ の $\phi(w)$ に対する十分条件)．しかしその十分条件を満たす推定量を適当に一つ選んでも，ほとんど全ての場合非許容的であり，上でも述べたように我々は許容的な推定量を提案したい．

そのような推定量を提案するために最も手軽な方法は，適当な (一般化) ベイズ推定量のクラスのなかでミニマクスであるための十分条件を満たす推定量を探すことである．しかし原点に関して球面対称な事前分布 $g(\|\theta\|^2)$ に対するベイズ推定量は，そのままでは表のような十分条件をチェックすることは出来ず，そのことが研究が進んでいない最大の要因となっていた．しかし，我々は $g(\|\theta\|^2) = \|\theta\|^{2-p}$ の場合 (Brown(1979) によって，この一般化事前分布に関するベイズ推定量は

Author	p	$\phi(w)/w$	upper bound
general			
Berger(1975)	$p \geq 3$		$2(p-2)\inf_{s\in U} F(s)/f(s)$
Brandwein(1979)	$p \geq 4$	$\searrow$	$2(p-2)(pE_0(\|X\|^{-2}))^{-1}$
unimodal or f is nonincreasing			
Brandwein & Strawderman(1978)	$p \geq 4$	$\searrow$	$2p((p+2)E_0(\|X\|^{-2}))^{-1}$
Ralescu *et al.*(1992)	$p = 3$	$\searrow$	$0.93(E_0(\|X\|^{-2}))^{-1}$
$F(t)/f(t)$ is nondecreasing			
Bock(1985)	$p \geq 4$	$\searrow$	$2(E_0(\|X\|^{-2}))^{-1}$
scale mixtures of multivariate normal			
Strawderman(1974) & Berger(1975)	$p \geq 3$	$\searrow$	$2(E_0(\|X\|^{-2}))^{-1}$

表 1: ミニマクスであるための十分条件

許容的であることが示されている．）には，一般化ベイズ推定量が次のように表現されて，ミニマクスであるための十分条件が非常にチェックしやすいことがわかる．

$$\delta_*(X) = \left(1 - \frac{\int_0^1 t^{p/2-1} F(t\|X\|^2)dt}{\int_0^1 t^{p/2-2} F(t\|X\|^2)dt}\right) X$$

また

$$\phi_*(w) = w\frac{\int_0^1 t^{p/2-1} F(wt)dt}{\int_0^1 t^{p/2-2} F(wt)dt}$$

の挙動に関する性質は次のようにまとめられる．

1. $\lim_{w\to\infty} \phi_*(w) = (p-2)E_0(\|X\|^2)/p$
2. 任意の f に対し，$\phi_*(w)$ は w に関して単調非減少である．
3. $F(t)(tf(t))^{-1}$ が単調非増加ならば，$\phi_*(w)/w$ は単調非増加である．

この性質と表に挙げたような十分条件を組み合わせることにより，$\delta_*(X)$ がミニマクスでかつ許容的であるような十分条件が得られる．詳細は省略するが，f のかなり広いクラスに対してミニマクスでかつ許容的であることが証明できる．

SOONER AND LATER WAITING TIME PROBLEMS FOR PATTERNS IN MARKOV DEPENDENT TRIALS

The Institute of Statistical Mathematics Qing HAN and Katuomi HIRANO

The waiting time problems of patterns have been discussed by many authors (Robin and Daudin (1999), Fu (1996) and references therein). In this paper, we discuss the sooner and later waiting time problems for patterns S_0 and S_1 in mnltistate Markov dependent trials. The probability functions and the probability generating functions of the sooner and later waiting time random variables are obtained. Further the probability generating functions of the distributions of distances between successive occurances of S_0 and S_0, of S_0 and S_1 and of the waiting time until the r-th occurrance of S_0 are also given.

Let $X_1, X_2, \ldots$ be a first order Markov chain with the state space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$, the initial probabilities $p_{\omega_i} = Pr(X_1 = \omega_i)$, $(i = 1, \ldots, N)$ and the transition probabilities $p_{\omega_i,\omega_j} = Pr(X_{t+1} = \omega_j | X_t = \omega_i)$, $(t = 1, 2, \ldots;\ i, j = 1, \ldots, N)$. Let $\mathbf{P} = (p_{\omega_i,\omega_j})_{N\times N}$ be the transition probability matrix and $p^{(m)}_{\omega_i,\omega_j} = Pr(X_{t+m} = \omega_j | X_t = \omega_i)$, $(m = 1, 2, \ldots)$ be the m-step transition probabilities.

Let $S_0 = a_1 a_2 \cdots a_k$ and $S_1 = b_1 b_2 \cdots b_\ell$, $(a_i, b_j \in \Omega,\ 1 \le i \le k$ and $1 \le j \le \ell,\ k \le \ell)$, be any patterns of length k and ℓ, respectively. Let W_0 (resp. W_1) be the waiting time until the first coourrence of the pattern S_0 (resp. S_1) in $X_1, X_2, \ldots$. Let W_S be the waiting time until the first occurrence of S_0 or S_1, whichever comes sooner, (*i.e.* $W_S = \min\{W_0, W_1\}$, it be called the sooner waiting time). Let W_L be the waiting time until the first occurrence both S_0 and S_1, whichever comes later, (*i.e.* $W_L = \max\{W_0, W_1\}$, it be called the later waiting time).

If $\Omega = \{0, 1\}$, S_0 is a 0-run of length k ($S_0 = 0 \cdots 0$) and S_1 a 1-run of length ℓ ($S_1 = 1 \cdots 1$), then the waiting time problems of patterns reduce to the waiting time problens of run which have been discussed by many authors (e.g. Han and Aki (2000), Aki and Hirano (1999)).

The waiting time distributions for patterns depend on the structures of pattern (Blom and Thorburn (1982)). We introduce the overlapping indicator $\varepsilon_{i,j}(r)$ between S_i and S_j, $(i, j = 0, 1)$. $\varepsilon_{i,j}(r)$ is equal to 1 if the last r letters of S_i are the same as the first r letters of S_j, equal to 0 for otherwise, $(i, j = 0, 1)$.

Let $W_S^{(0)}$ (resp. $W_S^{(1)}$) be a random variable denoting the number of trials until S_0 (resp. S_1) occurs sooner than S_1 (resp. S_0). Note that

$$Pr(W_S = t) = Pr(W_S^{(0)} = t) + Pr(W_S^{(1)} = t).$$

We introduce some notations:

$$p_S^{(0)}(t) = Pr(W_S^{(0)} = t),\ \ p_S^{(1)}(t) = Pr(W_S^{(1)} = t),$$

$$A_0(i) = p_{a_i,a_{i+1}} p_{a_{i+1},a_{i+2}} \cdots p_{a_{k-1},a_k},\ (i = 1, \ldots, k-1),\ A_0(0) = p_{a_1} A_0(1),$$

$$A_1(j) = p_{b_j,b_{j+1}} p_{b_{j+1},b_{j+2}} \cdots p_{b_{\ell-1},b_\ell},\ (j = 1, \ldots, \ell-1),\ A_1(0) = p_{b_1} A_1(1).$$

Since the event $\{S_0$ *occurs at* $t\}$ can be divided in three disjoint events, we have

$$\begin{aligned}\sum_{i=1}^{N} p_{\omega_i} p^{(t-k)}_{\omega_i,a_1} A_0(1) &= Pr(W_S^{(0)} = t) \\ &+ \sum_{z=1}^{t-1} Pr(W_S^{(0)} = z) Pr(S_0 \text{ occurs at } t | W_S^{(0)} = z) \\ &+ \sum_{z=1}^{t-1} Pr(W_S^{(1)} = z) Pr(S_0 \text{ occurs at } t | W_S^{(1)} = z).\end{aligned} \tag{1}$$

From the relation (1), we can get the recurrence formulae of the probability function of W_S.

Let $\phi_S(x)$, $\phi_S^{(0)}(x)$, $\phi_S^{(1)}(x)$, $\phi_L(x)$, $\phi_0(x)$ and $\phi_1(x)$ be the probability generating functions of W_S, $W_S^{(0)}$, $W_S^{(1)}$, W_L, W_0 and W_1, respectively. (*i.e.* $\phi_S(x) = \sum_{t=1}^{\infty} p_S(t)x^t$). In both sides of formula (1), multiplying by x^t, and summing each side of these equations, we have

THEOREM 3.1. *The $\phi_S^{(0)}(x)$ and $\phi_S^{(1)}(x)$ satisfy:*

$$\Big(C_{00} + B_0(a_k)A_0(1)x^{k-1}\Big)\phi_S^{(0)}(x) + \Big(C_{10} + B_0(b_\ell)A_0(1)x^{k-1}\Big)\phi_S^{(1)}(x)$$
$$= \Big(p_{a_1}x + \sum_{i=1}^{N} p_{\omega_i} x B_0(\omega_i)\Big)A_0(1)x^{k-1},$$
$$\Big(C_{01} + B_1(a_k)A_1(1)x^{\ell-1}\Big)\phi_S^{(0)}(x) + \Big(C_{11} + B_1(b_\ell)A_1(1)x^{\ell-1}\Big)\phi_S^{(1)}(x)$$
$$= \Big(p_{b_1}x + \sum_{i=1}^{N} p_{\omega_i} x B_1(\omega_i)\Big)A_1(1)x^{\ell-1},$$

where
$$B_0(\omega_i) = \sum_{j=1}^{\infty}(p^{(j)}_{\omega_i,a_1}x^j), \quad B_1(\omega_i) = \sum_{j=1}^{\infty}(p^{(j)}_{\omega_i,b_1}x^j) \tag{2}$$

and
$$C_{00} = 1 + \sum_{r=1}^{k-1}\varepsilon_{0,0}(r)A_0(r)x^{k-r}, C_{10} = \sum_{r=1}^{k-1}\varepsilon_{1,0}(r)A_0(r)x^{k-r},$$
$$C_{01} = \sum_{r=1}^{k-1}\varepsilon_{0,1}(r)A_1(r)x^{\ell-r}, C_{11} = 1 + \sum_{r=1}^{\ell-1}\varepsilon_{1,1}(r)A_1(r)x^{\ell-r}.$$

And, the probability generating function $\phi_S(x)$ is

$$\phi_S(x) = \phi_S^{(0)}(x) + \phi_S^{(1)}(x).$$

Using Jordan type decomposition of matrix $\mathbf{P}$, we can compute $B_0(\omega_i)$ and $B_1(\omega_i)$.

By letting $\ell \longrightarrow \infty$, we can obtain the probability function and the probability generating function of W_0. From the relation $\{W_S = t\} \cup \{W_L = t\} = \{W_0 = t\} \cup \{W_1 = t\}$, we can also obtain the probability function and the probability generating function of the later waiting time W_L.

Similarly, we can obtain the probability generating functions of distances between successive occurances of S_0 and S_0, of S_0 and S_1 and of the waiting time until the r-th occurrance of S_0.

These methods and results can be extended to the waiting time problems of several patterns $S_1, \cdots, S_m$ and to a higher order Markov sequence (when the order $\le k$).

References

Aki, S. and Hirano, K. (1999). Sooner and later waiting time problems for runs in Markov dependent bivariate trials. *Ann. Inst. Statist. Math.* **51**, 17-29.

Blom, G. and Thorburn, D. (1982). How many random digits are required until given sequences are obtained ? *J. Appl. Prob.* **19**, 518-531.

Fu, J. C. (1996). Distribution theory of runs and patterns associated with a sequence of multistate trials. *Statist. Sinica* **6**, 957-974.

Han, Q. and Aki, S. (2000). Waiting time problems in a two-state Markov chain. *Ann. Inst. Statist. Math.* **52**, 778-789.

Robin, S. and Daudin, J. J. (1999). Exact distribution of word occurences in a random sequence of letters. *J. Appl. Prob.* **36**, 179-193.

パターンが起こるまでにサブパターンの起こる分布と幾何分布

統計数理研究所 平野 勝臣

大阪大学大学院 基礎工学研究科 安芸 重雄

1. はじめに

連やパターンに関する統計量の分布は、確率変数の系列が {0,1}-値で独立、同一の分布に従っていても普通は複雑である。しかしながら、最近、連についてのいくつかの統計量は幾何分布のような簡単な分布に従うことが報告されている (Hirano, Aki and Uchida (1997))。本報告では、パターンのあるクラスに対して「パターンがはじめて起こるまでに、指定された長さ以上のサブパターンの起こる回数の分布は、依存系列であっても、幾何分布に従う」ことを報告する。

確率生成母関数 $\phi(t) = pt/(1-(1-p)t)$ をもつ幾何分布を $G_1(p)$ とかく。

2. 結果

$X_1, X_2, \ldots$ を $\{1, 2, \ldots, \mu\}$-値 i.i.d. 確率変数列とする。ここに、$j = 1, 2, \ldots, \mu$ に対し $P(X_i = j) = p_j$、$p_1 + p_2 + \cdots + p_\mu = 1$ で、すべての $j = 1, 2, \ldots, \mu$ に対して $p_j > 0$ とする。また $(a_1, a_2, \ldots, a_k)$ を $\{1, 2, \ldots, \mu\}$-パターン、すなわち $\{1, 2, \ldots, \mu\}$ の要素の有限列、整数 ℓ $(1 \leq \ell < k)$ に対しパターンの左からのサブパターン $(a_1, a_2, \ldots, a_\ell)$ を考える。

もし $X_{i-\ell+1} = a_1, X_{i-\ell+2} = a_2, \ldots, X_i = a_\ell$ で $X_{i-\ell-h+1} = a_1, \ldots, X_i = a_{\ell+h}$ であるような $(k-\ell)$ 以下の正整数 h が存在しないならば、長さ ℓ 以上のサブパターンが i 番目の試行で起こったという。

定義 1. パターン $(a_1, a_2, \ldots, a_k)$ に対し、次の (i) か (ii) のいずれかであれば、サブパターン $(a_1, \ldots, a_\ell)$, $(\ell < k)$ は条件 A を満たすという：

(i) $a_1 = a_{i+1}, a_2 = a_{i+2}, \ldots, a_\ell = a_{i+\ell}$ であるような整数 i が存在しない。

(ii) もしそのような i が存在すれば、すべての $j = 1, 2, \ldots, (k-\ell-i)$ に対し $a_{\ell+i+j} = a_{\ell+j}$ を満たす。

本報告で採用するサブパターンの数え方は $(\ell-1)$-重複の数え方で、隣あったサブパターンの長さ $(\ell-1)$ 以下の重複を許す数え方である (Aki and Hirano (2000))。例えば、パターン $(1, 2, 3.1, 2, 1, 2, 4)$ がはじめて起こるまでに、長さ 3 以上のサブパターンの起こる回数を数える。系列の実現値が $(1, 2, 3, 1), 3, 2, (1, 2, 3), 4, (1, 2, 3, 1, 2, [1, 2), 3, 1, 2, 1, 2, 4]$ と

すると、パターンがはじめて起こるまでに、長さ 3 以上のサブパターンは 4 回起こっている (カッコを参照)。最後の 2 つのサブパターンは長さ 2 の重複部分をもっている。

定理 1. サブパターン $(a_1, a_2, \ldots, a_\ell)$ は条件 A を満たしているとする。そのとき、パターン $(a_1, a_2, \ldots, a_k)$ がはじめて起こるまでに、長さ ℓ 以上のサブパターンの起こる回数は、$(\ell-1)$-重複の数え方で数えると幾何分布 $G_1(p_{a_{\ell+1}} \cdots p_{a_k})$ に従う。

注意. 連のどんなサブパターンも条件 A を満たすので、定理 1 は Aki and Hirano (1995) の Corollary 3.1 の拡張である。パターンが連ならば、サブパターンの、$(\ell-1)$-重複で数えたときの、起こる回数は、non-overlapping で数えた生起数に等しい。

上の結果は依存系列に拡張できる。

$X_{-m+1}, X_{-m+2}, \ldots, X_0, X_1, X_2, \ldots$ を $\{1, 2, \ldots, \mu\}$-値 m-次マルコフ系列とする。ただし、初期分布と推移確率は、 $x_1, \ldots, x_m, x = 1, 2, \ldots, \mu$ と $i = 1, 2, \ldots$ に対して

$$\pi_{x_1,\ldots,x_m} = P(X_{-m+1} = x_1, X_{-m+2} = x_2, \ldots, X_0 = x_m),$$
$$p_{x_1,\ldots,x_m,x} = P(X_i = x | X_{i-m} = x_1, X_{i-m+1} = x_2, \ldots, X_{i-1} = x_m)$$

とする。また $x_1, x_2, \ldots, x_m, x$ に対して $0 < p_{x_1,\ldots,x_m,x} < 1$ と仮定する。

定理 2. サブパターン $(a_1, a_2, \ldots, a_\ell)$ は条件 A を満たし、系列は m-次マルコフ系列で $\ell \geq m$ とする。そのときパターン $(a_1, a_2, \ldots, a_k)$ がはじめて起こるまでに、長さ ℓ 以上のサブパターンの起こる回数は、$(\ell-1)$-重複の数え方で数えると、幾何分布 $G_1(p)$ に従う。ここに $p = p_{a_{\ell-m+1},\ldots,a_\ell,a_{\ell+1}} p_{a_{\ell-m+2},\ldots,a_{\ell+1},a_{\ell+2}} \cdots p_{a_{k-m},\ldots,a_{k-1},a_k}$ である。

なお、本報告は Hirano and Aki (2001) に基づいている.

参考文献

Aki, S. and Hirano, K. (1995). Joint distributions of numbers of success-runs and failures until the first consecutive k successes, *Annals of the Institute of Statistical Mathematics*, **47**, 225-235.

Aki, S. and Hirano, K. (2000). Numbers of success-runs of specified length until certain stopping time rules and generalized binomial distributions of order k, *Annals of the Institute of Statistical Mathematics*, **52**, 767-777.

Hirano, K. and Aki, S. (2001). Number of occurrences of subpattern until the first appearance of a pattern and geometric distribution, *Research Memorandum* No. 805, The Institute of Statistical Mathematics.

Hirano, K., Aki, S. and Uchida, M. (1997). Distributions of numbers of success-runs until the first consecutive k successes in higher order Markov dependent trials, *Advances in Combinatorial Methods and Applications to Probability and Statistics*, ed. by N. Balakrishnan, 401-410, Birkhäuser, Boston.

ASYMPTOTIC THEORY FOR THE MULTISCALE WAVELET DENSITY DERIVATIVE ESTIMATOR

島根大学大学院総合理工学研究科　落合 俊充
島根大学総合理工学部　内藤 貫太

1. はじめに ウェーブレット関数 $\psi \in \mathbb{L}_2(\mathbb{R})$ をもとに生成される関数系 $\{\psi_{\ell,m}(\cdot) = 2^{\ell/2}\psi(2^{\ell}\cdot -m) \mid \ell, m \in \mathbb{Z}\}$ が $\mathbb{L}_2(\mathbb{R})$ の正規直交基底であるため, 任意の関数 $f \in \mathbb{L}_2(\mathbb{R})$ は

$$f(x) = \sum_{\ell,m\in\mathbb{Z}} d_{\ell,m}\psi_{\ell,m}(x), \quad d_{\ell,m} = \langle f, \psi_{\ell,m}\rangle := \int_{\mathbb{R}} f(x)\psi_{\ell,m}(x)\,dx$$

と表される. 従って, ウェーブレットを用いて確率密度関数や回帰関数 (いずれも $\mathbb{L}_2$ 関数とする) を推定する場合, いかに各基底関数にかかる係数 $d_{\ell,m}$ を推定するかが問題となる. 密度関数の推定において, Donoho et al. (1996) は Wavelet Thresholding による非線形ウェーブレット推定量を提案している. これに対して, Kerkyacharian and Picard (1992) や Huang (1999) で議論されている線形ウェーブレット推定量はよりシンプルな形の推定量である. ここでは線形ウェーブレット推定量を扱うが, 推定対象は密度導関数である. 密度導関数を推定対象とした研究は Prakasa Rao (1996) にみられるが, 漸近正規性については触れられていない. また, Wu (1996) はマルチスケール法により密度関数の推定量を構成し, その漸近正規性について議論している. 本稿では, 密度導関数の推定量をマルチスケール法により構成し, その漸近的性質について議論する.

2. Mutiscale Wavelet Density Derivative Estimator (MWDDE) f を確率密度関数とし, $X_1, X_2, \cdots, X_n \overset{i.i.d.}{\sim} f$ とするとき, f の m 階導関数 $f^{(m)} \in \mathbb{L}_2(\mathbb{R})$ の推定量を考える. multiresolution approximation (multiresolution analysis) [Daubechies (1992), Vidakovic (1999)] を用いれば, 関数系 $\{\varphi_{j,k}(\cdot) = 2^{j/2}\varphi(2^j\cdot -k) \mid k \in \mathbb{Z}\}$ を空間 V_j ($V_j \subset V_{j+1}$, $j \in \mathbb{Z}$) の正規直交基底とする $\mathbb{L}_2(\mathbb{R})$ の閉部分空間の列が得られる. また, $f^{(m)}$ を空間 V_j に直交射影した関数

$$f_j^{(m)}(x) = \sum_{k\in\mathbb{Z}} \alpha_{j,k}\varphi_{j,k}(x), \quad \alpha_{j,k} = \langle f^{(m)}, \varphi_{j,k}\rangle$$

は, $f^{(m)}$ に $\mathbb{L}_2$ の意味で収束する ($\lim_{j\to\infty}\|f^{(m)} - f_j^{(m)}\|_{\mathbb{L}_2} = 0$ が成り立つ). 従って, $j \to \infty$ のもとでは関数 $f_j^{(m)}$ の推定量を関数 $f^{(m)}$ の推定量と見なすことができる. Prakasa Rao (1996) はひとつの空間 V_j において推定量を構成しているが, ここでは複数の空間での特徴を含む推定量を考える. まず, $\boldsymbol{X} = (X_1, X_2, \cdots, X_n) = (Y_1, Y_2, \cdots, Y_{n_1}, Z_1, Z_2, \cdots, Z_{n_2}) = (\boldsymbol{Y}, \boldsymbol{Z})$ とする. ただし, $n_1 + n_2 = n, n_1 = O(n_2)$ である. L 個の各空間 V_{j_ℓ} $(0 < j_1 < j_2 < \cdots < j_L)$ での $f^{(m)}$ の推定量を $\hat{\alpha}_{j_\ell,k} = (-1)^m n_2^{-1}\sum_{i=1}^{n_2}\varphi_{j_\ell,k}^{(m)}(Z_i)$ として,

$$\hat{f}_\ell^{(m)}(x, Z) = \frac{(-1)^m}{n_2}\sum_{i=1}^{n_2}\mathbb{K}_{j_\ell}^{(m)}(x, Z_i)$$

とする. ここで, $\mathbb{K}_{j_\ell}^{(m)}(x,y) = 2^{(m+1)j_\ell}\sum_{k\in\mathbb{Z}}\varphi_\ell(2^{j_\ell}x-k)\varphi_\ell^{(m)}(2^{j_\ell}y-k)$ であり, φ_ℓ は空間 V_{j_ℓ} の正規直交基底を生成する関数 (スケーリング関数) である. この $\hat{f}_\ell^{(m)}$ $(\ell = 1,2,\cdots,L)$ を sub-estimator とし, $f^{(m)}$ の MWDDE を次のように定義する.

$$\hat{f}_M^{(m)}(x,\boldsymbol{X}) = \sum_{\ell=1}^{L}\lambda_\ell(x,\boldsymbol{Y})\hat{f}_\ell^{(m)}(x,\boldsymbol{Z}).$$

ただし, $\lambda_\ell(x,\boldsymbol{Y})$ $(\ell = 1,2,\cdots,L)$ は任意の $x\in\mathbb{R}$ に対して, $0<\lambda_\ell(x,\boldsymbol{Y})<1$, $\sum_{\ell=1}^{L}\lambda_\ell(x,\boldsymbol{Y}) = 1$ を満たす $\boldsymbol{Y}$ に基づく重みである.

3. MWDDE の漸近的性質 適当な条件のもとで MWDDE のバイアスと分散はそれぞれ,

$$\begin{aligned}
Bias\left[\hat{f}_M^{(m)}(x,\boldsymbol{X})\right] &= -\frac{2^{-sj_1}}{s!}\bar{\lambda}_1(x)f^{(m+s)}(x)b_s(2^{j_1}x)+O\left(2^{-(s+\alpha)j_1}\right)+O\left(2^{-sj_2}\right).\\
Var\left[\hat{f}_M^{(m)}(x,\boldsymbol{X})\right] &= \frac{2^{(2m+1)j_L}}{n_2}\bar{\lambda}_L(x)^2f(x)V_m(2^{j_L}x)\\
&\quad +O\left(\frac{2^{2mj_L}}{n_2}\right)+O\left(\frac{2^{(2m+1)(j_L+j_{L-1})/2}}{n_2}\right)
\end{aligned}$$

となる. ただし, $b_s(x) = x^s - \int_{\mathbb{R}}\mathbb{K}_0^{(0)}(x,y)y^s\,dy$, $V_m(x) = \int_{\mathbb{R}}\mathbb{K}_0^{(m)}(x,y)^2\,dy$, $\bar{\lambda}_\ell(x) = E_{\boldsymbol{Y}}\left[\lambda_\ell(x,\boldsymbol{Y})\right]$ $(\ell = 1,2,\cdots,L)$ である. また, $\lim_{j_L\to\infty}V_m(2^{j_L}x) = \sigma_{m,L}(x)^2$ とし, バイアスと分散を求めたときより条件を少し強めることで次が成り立つ.

$$\frac{\sqrt{n_2 2^{-(2m+1)j_L}}\left(\hat{f}_M^{(m)}(x,\boldsymbol{X})-E\left[\hat{f}_M^{(m)}(x,\boldsymbol{X})\right]\right)}{\bar{\lambda}_L(x)\sigma_{m,L}(x)\sqrt{f(x)}} \xrightarrow{\mathcal{D}} N(0,1),\quad as\ \ n\longrightarrow\infty. \tag{3.1}$$

(3.1) と同じ条件のもとで sub-estimator $\hat{f}_L^{(m)}$, つまりひとつの空間 V_{j_L} に対して構成された推定量についても漸近正規性が成り立つ. しかし, MWDDE の方が漸近分散の値が小さい. また, これらの結果は特に $m=0$ とするとき, Wu (1996) の結果に一致する.

参考文献

[1] Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM. Philadelphia.

[2] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1996). Density estimataion by wavelet thresholding. *Ann. Statist.* **24**, 508-539.

[3] Huang, S. Y. (1999). Density estimation by wavelet-based reproducing kernel. *Statistical Sinica*, **9**, 137-151.

[4] Kerkyacharian, G. and Picard, D. (1992). Density estimation in Besov spaces. *Statist. Probab. Letters.* **13**, 15-24.

[5] Prakasa Rao, B. L. S. (1996). Nonparametoric estimation of the derivatives of a density by the method of wavelets. *Bull. Inform. Cybernet.* **28**, 91-100.

[6] Vidakovic, B. (1999). *Statistical Modeling by Wavelets.* Wiley.

[7] Wu, D. (1996). Asymptotic normarity of the multiscale wavelet density estimator. *Comm. Statist. Theory Methods*, **25**(9), 1957-1970.

B-Spline Non-parametric Regression Model における過剰適合の回避について

統計数理研究所　柳原 宏和
広島大・原医研　大瀧 慈

1. Introduction

誤差を伴いかつ複雑なトレンドを持つデータに関して平滑化をおこなう場合, B-スプラインノンパラメトリック回帰モデルを用いた手法は, その他のノンパラメトリックな手法に比べて, その手法が従来の線形モデルにおける未知パラメーター推定問題に帰着できるために, 単純でかつ短い計算時間で結果を得ることができるといった利点を持つ. しかしながらこのモデルは, 本来柔軟なモデルであるため, ときに柔らか過ぎる適合, つまり過剰な適合を起こすことがある. 本発表ではこのような**過剰適合**に着目し, それを回避するような手法を提案する.

2. B-Spline Nonparametric Regression Model

説明変数と目的変数の組 (X, Y) に対する n 個の観測値を $\{(x_i, y_i) \mid i = 1, 2, \dots, n\}$ とする. B-スプラインノンパラメトリックモデルではデータの平均を以下のような既知の基底関数の和として捉える.

$$y_i = \sum_{j=1}^{m} a_j B_j(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

ただし ε_i はそれぞれ独立な確率変数で, 平均 $\mathrm{E}(\varepsilon_i) = 0$, 分散 $\mathrm{Var}(\varepsilon_i) = \sigma^2$ をもつ分布に従うとする. 多くの場合, この誤差項に正規性を仮定して議論を進めていくが, より広いクラスへの適応のため, 誤差項の具体的な分布の仮定を外す. 今,

$$B = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)' = \begin{pmatrix} B_1(x_1) & \cdots & B_m(x_1) \\ \vdots & \ddots & \vdots \\ B_1(x_n) & \cdots & B_m(x_n) \end{pmatrix},$$

とし, $\boldsymbol{y} = (y_1, y_2, \dots, y_n)'$, $\boldsymbol{a} = (a_1, a_2, \dots, a_m)'$ とおくと, モデルは $\boldsymbol{y} = B'\boldsymbol{a}$, と書きかえることができ, B-スプラインを用いた平滑化は, 線形モデルにおける $\boldsymbol{a}$, σ^2 の推定問題に帰着されることがわかる. このモデルの $\boldsymbol{a}$ は, 残差平方和に曲線の局所変動の程度を考慮した罰則付き残差平方和 (Penalized Residual Sum of Squares) の最小化に基づいて推定される. この局所変動を制御する平滑化パラメータ λ と基底関数の個数 m に関する最適化には情報量基準が用いられ, 情報量基準を一番小さくする λ と m を最適な値とされる.

3. New Criterion and Robust B-Spline Smoothing Method

従来の最適化では, 基底関数の個数の増加と伴いモデルが柔軟になり過剰適合を起こすことがある. これは最適化を行う情報量基準のバイアス項に較べ, 推定分散が小さくなりすぎていることに原因があると思われる. このような状況を回避するために分散の推定値に下限を与え, それよりも小さい推定分散は下限と置き換える基準量を提案する. この下限として局所線形フィットに基づいた推定量 (Gasser, Sroka and Jennen-Steinmetz, 1986) を用いた. 今, 観測値 x_i は $x_1 < x_2 < \cdots < x_n$ といった大小関係が成立しているとする. このとき局所線形フィットに基づいた分散の推定量は,

$$S_\varepsilon^2 = \frac{1}{n-2} \sum_{i=2}^{n-1} c_i^2 \tilde{\varepsilon}_i^2,$$

である．ただし，

$$\tilde{\varepsilon}_i = \frac{x_{i+1}-x_i}{x_{i+1}-x_{i-1}}y_{i-1} + \frac{x_i - x_{i-1}}{x_{i+1}-x_{i-1}}y_{i+1} - y_i = a_i y_{i-1} + b_i y_{i+1} - y_i, \quad (2 \le i \le n-1),$$

であり，$c_i^2 = (a_i^2 + b_i^2 + 1)^{-1}$．その下限を用いて以下のような新しい C_p 基準を提案する．

$$C_p(\lambda, S_\varepsilon^2) = \frac{\widetilde{RSS}(\hat{\boldsymbol{a}})}{S_\varepsilon^2} + 2\mathrm{tr}(H_\lambda) - n.$$

ただし，

$$\widetilde{RSS}(\hat{\boldsymbol{a}}) = \begin{cases} \sum_{i=1}^{n}(y_i - \boldsymbol{b}_i'\hat{\boldsymbol{a}})^2 & (\tilde{\sigma}^2 > S_\varepsilon^2) \\ nS_\varepsilon^2 & (\tilde{\sigma}^2 \le S_\varepsilon^2) \end{cases}, \quad \left(\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{b}_i'\hat{\boldsymbol{a}})^2\right).$$

またこのモデルは標本数が少ないときに外れ値に敏感になり，過剰適合を起こすこともある．次に未知パラメーターの推定に関して罰則付き残差平方和ではなく，Silverman (1985) 等で紹介されている罰則付き重み付き残差平方和を最小にすることで推定量を構成するで頑健な推定法を提案する．この重みの決め方には Andrews (1974) の手法を適応する．実際の最適化アルゴリズムは以下のようになる．

1. 係数 c を決定する (奨励は $c = 1.5$).
2. 基底関数の個数 (節点の個数) m を決める．
3. 平滑化パラメータ λ をあらかじめ与えておく．
4. 初期値 $\hat{\boldsymbol{a}}_{(0)} = (B'B)^{-1}B'\boldsymbol{y}$ を計算し，それをもとに $\hat{\sigma}^2_{(0)}$ を求める．
5. $\hat{r}_i^{(l-1)} = y_i - \boldsymbol{b}_i'\hat{\boldsymbol{a}}_{(l-1)}$ $(1 \le i \le n,\ l = 1, 2, \cdots)$, $\hat{r}_m^{(l-1)} = \mathrm{median}_{i=1,\ldots,n}\left(\left|\hat{r}_i^{(l-1)}\right|\right)$, として，重み
$$\hat{w}_i^{(l-1)} = \begin{cases} \sin\left(\dfrac{\hat{r}_i^{(l-1)}}{c\hat{r}_m^{(l-1)}}\right)\Big/\hat{r}_i^{(l-1)} & \left(\left|\dfrac{\hat{r}_i^{(l-1)}}{\hat{r}_m^{(l-1)}}\right| \le c\pi\right) \\ 0 & (\text{otherwise}) \end{cases},$$
を更新する．
6. $\widehat{W}_{(l-1)} = \mathrm{daig}(\hat{w}_1^{(l-1)}, \hat{w}_2^{(l-1)}, \ldots, \hat{w}_n^{(l-1)})$ とし，係数 $\hat{\boldsymbol{a}}_{(l)}$ を
$$\hat{\boldsymbol{a}}_{(l)} = (B'\widehat{W}_{(l-1)}B + \lambda D_k'D_k)^{-1}B'\widehat{W}_{(l-1)}\boldsymbol{y},$$
で更新し，$\hat{\sigma}^2_{(l)}$ を計算する．
7. $|\hat{\sigma}^2_{(l)} - \hat{\sigma}^2_{(l-1)}| < \delta$ で評価し，この条件を満たすまで 5, 6 を反復する．
8. $\hat{\sigma}^2_{(l)}$ または $\hat{\boldsymbol{a}}_{(l)}$ をもちいて情報量基準 $IC(m, \lambda)$ を計算しする．
9. m と λ を変えて 2 ～ 8 を反復させ $IC(m, \lambda)$ が最小になる m, λ を最適な値とする．

References

1. Andrews, D. F. (1974). A robust method for multiple linear regression. *Technometrics*, **16**, 523–531.
2. Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, **73**, 625–633.
3. Silverman, B. W. (1985). Some aspects of spline smoothing approach to non-parametric regression curve fitting. *J. Roy Statist. Soc. Ser.* **B**, **47**, 1–52.

Generalized binomial and negative binomial distributions of order k by the ℓ-overlapping enumeration scheme

大阪大学 大学院基礎工学研究科 井上 潔司
大阪大学 大学院基礎工学研究科 安芸 重雄

1 はじめに

$X_0, X_1, X_2, \ldots$ を time homogeneous $\{0,1\}$-valued Markov chain とし，その推移確率は，$p_{ij} = P(X_t = j | X_{t-1} = i)$，$(t \geq 1,\ i, j = 0, 1)$，初期確率は，$P(X_0 = 0) = p_0$，$P(X_0 = 1) = p_1$ とする．(今後は，"1"，"0" をそれぞれ，成功，失敗と呼ぶことにする.) 長さが k の成功連が r 回現れるまでの waiting time の分布はオーダー k の負の二項分布と呼ばれている．この分布に関しては，他の従属試行列 (例えば，binary sequence of order k, sampling from urn models など) のもとでの考察も行われている．また，成功連の数え方も，重複しないで数える方法，長さ k 以上の連を数える方法，長さ k の連を重複して数える方法，ちょうど長さ k の連を数える方法，が用いられるなど，様々な設定の下で今までに多くの研究がなされてきた．本報告では，最近 Aki and Hirano (2000) によって提案された ℓ-overlapping count 法を用いて，$\{0,1\}$-valued 独立試行列において，長さ k の 成功連が r 回起こるまでの waiting time の分布について考察を行う．ただし，$0 < \ell \leq k-1$ とする．

2 Generating functions

T_r, $(r \geq 1)$ を 長さ k の 成功連が r 回起こるまでの waiting time とする．このとき，T_r の probability generating function $H_r(z)$, double generating function $H(z, w)$ をそれぞれ次のように定義する．

$$
\begin{aligned}
H_r(z) &= E[z^{T_r}] = \sum_{n=0}^{\infty} \Pr[T_r = n] z^n, \\
H(z, w) &= \sum_{r=0}^{\infty} H_r(z) w^r = \sum_{r=0}^{\infty} \sum_{n=0}^{\infty} \Pr[T_r = n] z^n w^r.
\end{aligned}
$$

Koutras and Alexandrou (1995) によれば，$H(z, w)$ は，適当な $s \times s$ 行列 A, B を用いて，

$$
H(z, w) = wz\pi_0 \sum_{i=1}^{s} \beta_i \left[I - z(A + wB)\right]^{-1} e_i'.
$$

と表されることが知られている．ただし，$\beta_i = e_i B 1'$，$1 \leq i \leq s$，$e_i = (0, \ldots, 1, \ldots, 0) \in R^s$，$\pi_0 = (p_0, p_1, 0, \ldots, 0)$．さて，$P(z)$, $Q(z)$, $R_r(z)$ を次のように定義する:

$$
P(z) = p_1 + (p_0 p_{01} - p_1 p_{00}) z,
$$

$$Q(z) = 1 - p_{00}z - p_{01}p_{10}z^2 \sum_{i=2}^{k} (p_{11}z)^{i-2},$$

$$R_x(z) = \begin{cases} 1 + z + z^2 + \cdots + z^x, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise.} \end{cases}$$

このとき $H(z, w)$ は, 次のように表される.

$$H(z, w) = \frac{wP(z)(p_{11}z)^{k-1}}{[1 - w(p_{11}z)^{k-\ell}]Q(z) - wp_{01}p_{10}p_{11}^{k-1}z^{k+1}R_{k-\ell-1}(p_{11}z)}$$

3 Negative binomial distribution of order k

$H(z, w)$ を展開し, w^r の係数を拾うと, $H_r(z)$ を求めることが出来る.

Theorem 3.1 T_r の確率母関数 $H_r(z)$ は, 次のように表される;

$$H_r^{(+)}(z) = \frac{(p_{11}z)^{k-1}P(z)}{Q(z)} \left[(p_{11}z)^{k-\ell} + \frac{p_{01}p_{10}p_{11}^{k-1}z^{k+1}R_{k-\ell-1}(p_{11}z)}{Q(z)} \right]^{r-1} \quad r \geq 1.$$

このことから, 次が導かれる.

Theorem 3.2 $r \geq 2$ のとき, T_r は, 独立な r 個の確率変数の和によって表される;

$$T_r \stackrel{d}{=} T + \sum_{j=1}^{r-1} W_j^*.$$

4 補足

当日は, $\ell \leq 0$ の場合に拡張した ℓ-overlaping counting 法も紹介する. また, $r \to \infty$ としたときの, オーダー k の負の二項分布の漸近的挙動についても取り上げる. さらには, 長さが n (fixed integer) である $\{0, 1\}$-sequence における長さ k の 成功連の数の分布についても考察する.

参考文献

Aki, S. and Hirano, K. (2000). Numbers of success-runs of specified length until certain stopping time rules and generalized binomial distributions of order k. *Ann. Inst. Statist. Math.*, **52**, 767-777.

Han, S. and Aki, S. (2000). A unified approach to binomial-type distributions of order k. *Communications in Statistics- Theory and Methods*, **29(8)**, 1929-1943.

Inoue, K. and Aki, S. (2001). Generalized binomial and negative binomial distributions of order k by the ℓ-overlapping enumeration scheme, *Research Report on Statistics*, **57**, Osaka University.

Koutras, M. V. and Alexandrou, V. A. (1995). Runs, scans and urn model distributions: A unified Markov chain approach. *Ann. Inst. Statist. Math.*, **47**, 743-766.

二次元パターンの待ち時間問題

大阪大学大学院 基礎工学研究科 安芸 重雄
統計数理研究所 平野 勝臣

この報告では，次のような二次元パターンの待ち時間問題を扱う．$\boldsymbol{X}_i$, $i = 1, 2, \ldots$ を m-次元の独立で同じ分布に従う確率列ベクトル (random column vectors) とし，その各成分は 独立同分布の { 0,1 }-値確率変数で，

$$P(X_{i,j} = 1) = p(= 1 - q)$$

を満たすものとする．ここで，$X_{i,j}$ は $\boldsymbol{X}_i$ の j-番目の成分を表す．D を有限の幅をもつ高さが高々m の 1 のパターンであるとする．

また，問題を正確に述べるため，パターン D を覆うような長方形の scanning window R を用意する．ただし，R は長方形の scanning window で，その幅は D の幅と同じとし，R の高さは D の高さ以上であるとする．この scanning window は m-次元確率列ベクトルを横に並べた列の中をスキャンし，window の中で，パターン D が起こっているかどうかをチェックする．そして，R の高さが m よりも小さいときは，R は左右に動くだけでなく，上下にも $m-$(R の高さ)-ステップだけ動くことができる．

この問題では，今まで行われてきたような直感的な条件付けが困難になるので，ここでは，可能なすべての状態を列挙し，各状態が与えられたときの待ち時間分布の条件付確率母関数が満たす線形方程式系を生成するアルゴリズムを与えるような系統的なアイデアを出し，厳密分布の確率母関数を導く．

Scanning window の高さが確率ベクトルの次元に一致するときには，scanning window は左右にしか動くことができないので，待ち時間問題はかなり簡単になる．そこで，まずこの場合において，二次元パターンの待ち時間分布を導出する基本的な考え方を説明する．

二次元パターン D とそれを覆うような scanning window R が与えられたとする．すると，$R = [\boldsymbol{r}_1 \ \ \boldsymbol{r}_2 \ \ldots \ \boldsymbol{r}_\ell]$ は，“1” または “*” を成分にもつ $m \times \ell$ 行列である．ここで，“*” は 1 または 0 のどちらでもよいことを意味する．

二次元パターン D がはじめて起こるまでの待ち時間を扱うときには，有限マルコフ連鎖埋め込み法における状態空間や条件付確率母関数法における「どこかで起こり得る状態の全体」は，各時点において各部分行列 $R(i)(\equiv [\boldsymbol{r}_1 \ \ \boldsymbol{r}_2 \ \ldots \ \boldsymbol{r}_i])$ が成り立っているかどうかを考慮することによって構成することができる．このことは，1 次元のパターンの待ち時間問題で，各時点において起こっている最大の長さの部分パターンだけを考慮して構成することができたことと対照的である．なぜこのように複雑な条件を考慮しなければならないかというと，*-成分があるために，ある時点で，R のある部分行列が起こっているときに，それだけの情報ではその時点でそれより小さい部分行列が起こっているかどうかを必ずしも判断できるとは限らないからである．そこで，長さ ℓ の { 0,1 }-ベクトル $\boldsymbol{a} = (a_1, a_2, \ldots, a_\ell)$ によって，各時点の状態を表すことにする．ここで，

$$a_i = \begin{cases} 0, & (\text{その時点で } R(i) \text{ が起こっていなければ}) \\ 1, & (\text{その時点で } R(i) \text{ が起こっていれば}) \end{cases}$$

である．もちろん，$\boldsymbol{a} \in \{0,1\}^\ell$ であるが，$\{0,1\}^\ell$ の要素がすべて「どこかの時点で起こり得る状態」になるとは限らない．そこで,「どこかの時点で起こり得る状態」の全体を $S(D,R)$ と表す．また，$S_1(D,R) = \{\boldsymbol{a} \in S(D,R) | a_\ell = 1\}$ と $S_0(D,R) = S(D,R) \setminus S_1(D,R)$ を定義しておく．

そこで，条件付確率母関数の満たす線形方程式系を構成する一般的な方法を与える．$V(m)(=\{0,1\}^m)$ を成分が 0 か 1 からなる m-次元列ベクトルの全体とする．各 $i=1,\ldots,\ell$ に対し，$\boldsymbol{u}_i$ を $\boldsymbol{r}_i$ の $*$ に 0 を代入して得られる列ベクトルとする．ここで，$\boldsymbol{r}_i$ は scanning window R の第 i 番目の列ベクトルである．次に写像 $f_D : S_0(D,R) \times V(m) \to S(D,R)$ を $f_D(\boldsymbol{a},\boldsymbol{e}) = \boldsymbol{b}(=(b_1,b_2,\ldots,b_\ell))$ によって定義する．ここで,

$$b_1 = \begin{cases} 1 & \text{if } \boldsymbol{e} - \boldsymbol{u}_1 \geq \boldsymbol{0} \\ 0 & \text{otherwise} \end{cases}$$

そして $i=2,\ldots,\ell$ に対しては,

$$b_i = \begin{cases} 1 & \text{if } \boldsymbol{e} - \boldsymbol{u}_i \geq \boldsymbol{0} \text{ and } a_{i-1} = 1 \\ 0 & \text{otherwise} \end{cases}$$

である．このとき，次の結果が得られる．

定理 1. 与えられたパターン D と scanning window R に対して，パターン D が，独立で同じ分布に従う $V(m)$-値確率ベクトルを横に並べた列の中ではじめて現れるまでの条件付確率母関数は次の線形方程式系を満たす．各 $\boldsymbol{a} \in S_0(D,R)$ に対し，

$$\phi(\boldsymbol{a};t) = \sum_{\boldsymbol{e} \in V(m)} p^{N_1(\boldsymbol{e})}(1-p)^{m-N_1(\boldsymbol{e})} t \phi(f_D(\boldsymbol{a},\boldsymbol{e});t)$$

が成り立つ．ここで，$N_1(\boldsymbol{e})$ は $\boldsymbol{e}$ の成分中の 1 の数を意味する．また，各 $\boldsymbol{a} \in S_1(D,R)$ に対しては，$\phi(\boldsymbol{a};t)=1$ が成り立つ．

注意．定理 1 は二次元パターン D の待ち時間の条件付確率母関数の満たす線形方程式系を明示的に与えているので，これに基づいて，その線形方程式系を生成するアルゴリズムを与えるのは容易である．また，このアルゴリズムと数式処理システムを用いれば，パターンと scanning window を入力すれば待ち時間分布の確率母関数を出力するプログラムを書くことができる．また，scanning window R の高さ h が m より小さい場合にも，やや複雑ではあるが，定理 1 と同様な一般的な結果を得ることができる．詳細は下記文献参照．

Aki, S. and Hirano, K. (2001). Waiting time problem for a tw-dimensional pattern, *Research Memorandum* No. 804, the Institute of Statistical Mathematics.

LINEX 損失関数のもとでのポアソン分布の平均のベイズ逐次推定

熊本大学　工学部　　　　高田佳和

$X_1, \cdots, X_n$ は互いに独立で、同一分布 $P_\omega, \omega \in \Omega$, に従う確率変数列とし、$\theta = \theta(\omega)$ の逐次推定問題を考える。ただし、θ の値を $\tilde{\theta}$ で推定したときの損失を $L(\theta, \tilde{\theta})$ とし、標本 1 個あたりのサンプリングコストを $c > 0$ とする。逐次推定方式を

$$\delta = (S, \tilde{\theta}_S),$$

とする。ここで、S は停止則、$\tilde{\theta}_S$ は推定量を表す。δ のリスクは

$$R(\omega, \delta) = E_\omega \left\{ L(\theta, \tilde{\theta}_S) + cS \right\}$$

で与えられる。Π を ω の事前分布としたとき、

$$R_\Pi(\delta) = \int R(\omega, \delta)\Pi(d\omega)$$

を δ のベイズリスクといい、このベイズリスクを最小にする δ_Π を (事前分布 Φ に関する) ベイズ逐次推定方式という。

$\hat{\theta}_n$ を n 個の標本 $X_1, \ldots, X_n$ にもとづく、損失関数 $L(\theta, \tilde{\theta})$ に対するベイズ推定量とうる。任意の逐次推定方式　$\delta = (S, \tilde{\theta}_S)$　に対して、推定量としてベイズ推定量を用いた逐次推定方式　$\hat{\delta} = (S, \hat{\theta}_S)$ を考えると

$$R_\Pi(\delta) \geq R_\Pi(\hat{\delta})$$

が成立する。従って、ベイズ逐次推定方式を求める問題は最適な停止則（ベイズ停止則）を求める問題に帰着する。

$$\begin{aligned} Z_n &= nE\left\{ L(\theta, \hat{\theta}_n) | X_1, \ldots, X_n \right\}, \\ L_n(c) &= \frac{Z_n}{n} + cn \end{aligned}$$

とおくと、

$$R_\Pi(\hat{\delta}) = E\left\{ L_S(c) \right\}$$

と表され、

$$E\left\{ L_{\tau_c}(c) \right\} = \min_S E\left\{ L_S(c) \right\} (= \rho(c))$$

を満たす停止則　τ_c がベイズ停止則である。

ベイズ停止則は多くの問題においてその存在は保証されているが (Chow, et al., 1970)、その構成は困難な場合が多い。そのため漸近的 $(c \to 0)$ にベイズ停止則と同等となる停止則の構成が研究されてきた。ここでは、Bickel and Yahav (1967, 1968) による APO(asymptotically pointwise optimal) 方式について考える。

$$Z_n \to V(>0) \quad a.s. \quad as \quad n \to \infty$$

を仮定する。このとき、停止則

$$T_c = \inf\left\{ n \geq 1, Z_n < cn^2 \right\}$$

が APO である。更に

$$\sup_n E(Z_n) < \infty$$

が成立すれば、T_c は AO (asymptotically optimal)、すなわち、

$$\lim_{c\to 0} \frac{E\{L_T(c)\}}{\rho(c)} = 1$$

が成立する。 次なる問題は、

$$E\{L_T(c)\} = \rho(c) + o(c) \qquad as \quad c \to 0$$

が成立するかどうかである。これが成立するとき、T_c は asymptotically non-deficient であるという。

X_i の分布はポアッソン分布 $P_o(\theta)$ の平均 θ の推定を LINEX (Linear Exponential) 損失関数

$$L(\theta,\hat{\theta}) = \exp[a(\hat{\theta}-\theta)] - a(\hat{\theta}-\theta) - 1 \quad (a \neq 0)$$

のもとで考察する。θ の事前分布は次の密度関数を持つ Γ 分布とする。

$$\frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\Gamma(\alpha)\beta^{\alpha}}, \quad \theta > 0 \quad (\alpha > 0, \beta > 0)$$

このとき $X_1,\ldots,X_n$ が与えられたときの θ のベイズ推定量は

$$\hat{\theta}_n = \left(\frac{\alpha+S_n}{a}\right)\log\left(1+\frac{a\beta}{n\beta+1}\right)$$

となり、

$$Z_n = (\alpha+S_n)A_n$$

ただし、

$$\begin{aligned} A_n &= n\left\{\log\left(\frac{n\beta+1}{(a+n)\beta+1}\right)+\frac{a\beta}{n\beta+1}\right\} \\ S_n &= X_1+\ldots+X_n \end{aligned}$$

従って、

$$Z_n \to \frac{a^2\theta}{2} \quad a.s. \quad as \quad n\to\infty$$

このことから停止則

$$T_c = \inf\left\{n \geq 1, Z_n < cn^2\right\}$$

は APO である。また、

$$E(Z_n) = (\alpha + n\alpha\beta)A_n$$

より、$\sup_n E(Z_n) < \infty$ が成立し、停止則 T_c は AO である。更に次の結果 (Takada, 2001) から asymptotically non-deficieny である。

定理 $a\beta+1>0, \alpha>1$ ならば、

$$E\{L_T(c)\} = \Gamma_0\sqrt{c} + \Gamma_1 c + o(c) \quad as \quad c\to 0$$

ここで、

$$\Gamma_0 = \frac{\sqrt{2a^2}\Gamma\left(\frac{1}{2}+\alpha\right)\sqrt{\beta}}{\Gamma(\alpha)}, \quad \Gamma_1 = -\frac{1}{\beta}+\frac{2a}{3}+\frac{1}{4(\alpha-1)\beta}$$

更に、

$$E\{L_T(c)\} = \rho(c) + o(c) \quad as \quad c\to 0$$

順序制約がある２つのガンマ分布の尺度母数の線形関数の推定

目白大・人文　張　元宗
慶応大・理工　篠崎　信雄

１．はじめに

$X_i \sim \Gamma(\alpha_i, \lambda_i), i = 1, 2$ とし、その密度関数を

$$f_{\lambda_i}(x_i) = x_i^{\alpha_i - 1} \lambda_i^{-\alpha_i} e^{-\frac{x_i}{\lambda_i}} / \Gamma(\alpha_i), \qquad 0 < x_i < \infty$$

とする。尺度母数λ_iに順序制約$\lambda_1 \leq \lambda_2$があるとき、その条件を考慮した最尤推定量（ＭＬＥ）は

$$\hat{\lambda}_i = \frac{X_i}{\alpha_i} + (-1)^i \frac{(\alpha_2 X_1 - \alpha_1 X_2)^+}{\alpha_i(\alpha_1 + \alpha_2)}, \qquad i = 1, 2,$$

である。ここで、$a^+ = max(0, a)$ であり、X_i/α_iはλ_iの不偏推定量（ＵＢ）である。一方、制約条件を無視した尺度母数の、個別には許容的な推定量 $X_i/(\alpha_i + 1)$ に基づいて、制約条件を満たす推定量を次のように与える。

$$\tilde{\lambda}_i = \frac{X_i}{\alpha_i + 1} + (-1)^i \frac{((\alpha_2 + 1)X_1 - (\alpha_1 + 1)X_2)^+}{(\alpha_i + 1)(\alpha_1 + \alpha_2 + 2)}, \qquad i = 1, 2$$

$\tilde{\lambda}_i$はＭＬＥにおいて、α_iを$\alpha_i + 1$ に置き換えた形になっている。

ここでは、平均２乗誤差を基準に、２つの尺度母数の線形関数の推定問題を考える。ＭＬＥが不偏推定量,$\frac{X_i}{\alpha_i}$, よりも優れている、さらに、$\tilde{\lambda}_i$が推定量,$\frac{X_i}{\alpha_i + 1}$, よりも優れているための、線形関数の係数に対する必要十分条件をそれぞれについて与える。

２．準備

リスクの表現を得るためには、次のLemmaが有用である。

Lemma 1. $X \sim \Gamma(\alpha, \lambda)$ とし、$g(x)$ を絶対連続関数とし、その微分を $g'(x)$ とする。いま、gに対して、

$(i) E[|Xg'(X)|] < \infty$ かつ $E[|g(X)|] < \infty$

(ii) すべての$\lambda > 0$に対して、$\lim_{x \to 0} g(x) x^\alpha e^{-\frac{x}{\lambda}} = \lim_{x \to \infty} g(x) x^\alpha e^{-\frac{x}{\lambda}} = 0$

が満たされているとすると

$$E[Xg(X)] = \lambda \{\alpha E[g(X)] + E[Xg'(X)]\}.$$

が成立する。

Lemma 2. $X_i \sim \Gamma(\alpha_i, \lambda_i), i = 1, 2$ とし、Ｚ $\sim Beta(\alpha_1, \alpha_2)$ とする。任意の定数 $b \geq 0$ に対して、

$$(i) \frac{E[X_2 | X_1 \geq bX_2]}{E[X_1 | X_1 \geq bX_2]} \geq \frac{E_0[X_2 | X_1 \geq bX_2]}{E_0[X_1 | X_1 \geq bX_2]} = \frac{1}{E[Z | Z \geq \rho]} - 1$$

が成立する。ここで、$\rho = b/(b+1)$ で、$E_0[\cdot]$ は$\lambda_1 = \lambda_2$のときの期待値である。$E_0[X_2 | X_1 \geq bX_2]/E_0[X_1 | X_1 \geq bX_2]$ は$\lambda_1 = \lambda_2$の値に依存しない。

$$\begin{aligned}
(ii) E[Z | Z \geq \rho] &= \frac{\alpha_1}{\alpha_1 + \alpha_2} \frac{1 - I_\rho(\alpha_1 + 1, \alpha_2)}{1 - I_\rho(\alpha_1, \alpha_2)} \\
&= \frac{\alpha_1}{\alpha_1 + \alpha_2} \left\{ 1 + \frac{1}{1 - I_\rho(\alpha_1, \alpha_2)} \frac{\rho^{\alpha_1}(1 - \rho)^{\alpha_2}}{(\alpha_1 + \alpha_2) B(\alpha_1 + 1, \alpha_2)} \right\}
\end{aligned}$$

ここで、$I_x(\alpha,\beta)=\int_0^x u^{\alpha-1}(1-u)^{\beta-1}du/B(\alpha,\beta)$, であり、$B(\alpha,\beta)$ はベータ関数である。

３．結果

$X_i \sim \Gamma(\alpha_i,\lambda_i), i=1,2$ とし、λ_iに順序制約$\lambda_1 \leq \lambda_2$があるとする。平均２乗誤差（ＭＳＥ）を基準に、$c_1\lambda_1+c_2\lambda_2$の推定問題を考えるとき、下記の定理が得られる。

（１）ＭＬＥ$(\hat{\lambda}_1,\hat{\lambda}_2)$と不偏推定量$(\frac{X_1}{\alpha_1},\frac{X_2}{\alpha_2})$との比較

定理１．すべての$\lambda_1 \leq \lambda_2$に対して、$MSE(\sum_{i=1}^2 c_iX_i/\alpha_i) \geq MSE(\sum_{i=1}^2 c_i\hat{\lambda}_i)$が成立するための必要十分条件は

$$\frac{c_1}{c_2} \leq \frac{\rho}{1-\rho}\frac{2-\rho-R}{R-\rho} \qquad (c_2=0\text{のケースを含む})$$

である。ここで、$\rho=\alpha_1/(\alpha_1+\alpha_2)$、$R=(1-I_\rho(\alpha_1+1,\alpha_2))/(1-I_\rho(\alpha_1,\alpha_2))$である。

系１．すべての$\lambda_1 \leq \lambda_2$、すべての$c_1 \geq 0, c_2 \leq 0$(したがってすべての$c_1 \leq 0, c_2 \geq 0$)に対して、$MSE(\sum_{i=1}^2 c_iX_i/\alpha_i) \geq MSE(\sum_{i=1}^2 c_i\hat{\lambda}_i)$が成立するための必要十分条件は$\alpha_1 \geq 1$である。

（２）ＭＬＥにおいて、α_iをα_i+1に置き換えた推定量$(\tilde{\lambda}_1,\tilde{\lambda}_2)$と$(\frac{X_1}{\alpha_1+1},\frac{X_2}{\alpha_2+1})$との比較

定理２．すべての$\lambda_1 \leq \lambda_2$に対して、$MSE(\sum_{i=1}^2 c_iX_i/(\alpha_i+1)) \geq MSE(\sum_{i=1}^2 c_i\tilde{\lambda}_i)$が成立するための必要十分条件は、$\rho'=(\alpha_1+1)/(\alpha_1+\alpha_2+2)$、$R'=(1-I_{\rho'}(\alpha_1+1,\alpha_2))/(1-I_{\rho'}(\alpha_1,\alpha_2))$とするとき

$$2\rho R' > \rho'+1\text{のとき、}\quad \frac{\rho'-2\rho R'}{2\rho R'-(\rho'+1)} \leq \frac{c_1}{c_2} \leq \frac{\alpha_1+1}{\alpha_2+1}$$

であり、

$$2\rho R' \leq \rho'+1\text{のとき、}\quad -\infty < \frac{c_1}{c_2} \leq \frac{\alpha_1+1}{\alpha_2+1} \qquad (c_2=0\text{のケースを含む})$$

である。

系２．すべての$\lambda_1 \leq \lambda_2$に対して、$MSE(X_1/(\alpha_1+1)) \geq MSE(\tilde{\lambda}_1)$が成立するための十分条件は$\alpha_2 \geq \alpha_1, \alpha_2 \geq 1$である。
数値計算の結果については、当日示す。

参考文献

(1)Berger, J. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. *Ann. Statist.*, 8, 545-571. (2)Kaur, A. and Singh, H. (1991). On the estimation of ordered means of two exponential populations. *Ann. Inst. Statist. Math.*, 43, 347-356.
(3)Shinozaki, N. and Chang, Y.-T. (1999). A comparison of maximum likelihood and best unbiased estimators in the estimation of linear combinations of positive normal means. *Statistics & Decisions*, 17, 125-136.
(4) 張元宗、篠崎信雄(2000)順序制約がある２つのガンマ分布の尺度母数の線形関数の推定－形状母数が異なる場合－。第６８回日本統計学会予稿集

正規混合分布におけるロバスト推定

統計数理研究所　藤澤洋徳

1 INTRODUCTION

Suppose that there are k different populations whose underlying distributions are normal. When the observation is sampled with mixing proportions ω_j's from the populations and the label of the population is missing, the probability density function is given by

$$f(x;\theta) = \sum_{j=1}^{k} \omega_j \phi(x;\mu_j,\sigma_j^2),$$

where θ is the symbol of the parameter, $\sum_{j=1}^{k}\omega_j = 1$, and $\phi(x;\mu_j,\sigma_j^2)$ is the normal density with mean μ_j and variance σ_j^2. This is called normal mixture model.

Let $x_1,\ldots,x_n$ be the random data sampled from the normal mixture model. A feature of the normal mixture model is flexibility of density shape, but its flexibility causes some difficulties for inference. One of the difficulties is that the likelihood function is unbounded, more precisely, $\sup_\theta \sum_{i=1}^{n} \log f(x_i;\theta) = \infty$, which brings about spurious local maxima. See Section 3.10 of McLachlan and Peel (2000) for detailed discussion. These make the maximum likelihood inference unreliable.

This paper adopts a modified likelihood for robust inference, given by

$$l_\beta(\theta) = \frac{1}{n\beta}\sum_{i=1}^{n} f(x_i;\theta)^\beta - b_\beta(\theta),$$

where $\beta > 0$ and $b_\beta(\theta) = \int f(x;\theta)^{1+\beta}dx/(1+\beta)$, which was suggested in Basu et al. (1998). It may be noted that the case $\beta = 0$ corresponds to the usual log-likelihood. This modified likelihood and its maximizer will be called β-likelihood and β-estimator throughout this paper. Results obtained can also be applied to the multivariate normal mixture model.

2 β-LIKELIHOOD METHOD

The β-likelihood and its properties were discussed in Basu et al. (1998). This section reviews some facts, which will help our understanding for this paper.

The estimating equation is expressed as $\partial l_\beta/\partial\theta = \sum_{i=1}^{n} h_\beta(x_i;\theta)/n = 0$, where $h_\beta(x;\theta) = f(x;\theta)^\beta\{\partial \log f(x;\theta)/\partial\theta\} - \partial b_\beta(\theta)/\partial\theta$, which can be viewed as M-estimation and unbiased as estimating function. The weight $f(x;\theta)^\beta$ of the score function would weaken the contribution of an outlier.

Let $\widehat{\theta}_\beta$ be the maximizer of the β-likelihood. The standard theory of the M-estimation will show the asymptotic normality. It might be noted that the case $\beta > 1$ causes much loss of efficiency for some basic models, as described in Basu et al. (1998). Hereafter we focus on the case $0 < \beta \le 1$ because of high efficiency.

3 BOUNDEDNESS

The likelihood function is unbounded under the normal mixture model, as discribed in Introduction. It can be shown that the β-likelihod function is bounded above under the mild condition:

$$\min_j \omega_j \ge \psi(\beta)/n,$$

where $\psi(\beta) = (1+\beta)^{3/2}/\beta$. It should be noted that the likelihood function is not bounded even under the above condition and it may be noted that the above condition is changeless even for the multivariate normal mixture model.

4 ROBUSTNESS AND EFFICIENCY

The influence function, say $IF(x;\theta)$, is known as a criterion of robustness (See Huber 1981). Let $GES = \sup_x \|IF(x;\theta)\|$, which is called gross-error-sensitivity. If the gross-error-sensitivity is finite, then we will expect that the estimator is robust to outliers. It is well known that the gross-error-sensitivity of the maximum likelihood estimator is infinity for $k = 1$. However, we can show that the gross-error-sensitivity of the β-estimator is finite under the normal mixture model.

Focus on the neighborhood of $\beta = 0$. Let us investigate a behavior of the gross-error-sensitivity as well as the asymptotic relative efficiency. The gross-error-sensitivity goes to infinity and the asymptotic relative efficiency to one, which means fully asymptotic efficiency, as β limits to zero. We can show that the bound of the gross-error-sensitivity is the order $O(1/\beta)$ and the convergence of the asymptotic relative efficiency is the order $O(\beta^2)$. These imply that we can expect both the robustness and the high efficiency of the β-estimator with a small β.

5 METHOD FOR SELECTING β

The β-likelihood method has been suggested already and explored as robust version of the usual likelihood. If we hope more strong robustness, we will use a larger β. However, the large β would reduce the efficiency of the β-estimator because the case $\beta = 0$ corresponds to the maximum likelihood estimator with full asymptotic efficiency. We would like to select a reasonable β. Two methods are suggested in this section.

One can be suggested in view of the condition $\min_j \omega_j \geq \psi(\beta)/n$ and the asymptotic relative efficiency. The values around $\beta = 0.2$ would be appropriate candidates.

Another is to take $\check{\beta} = \arg\min_\beta D(g(x), f(x;\widehat{\theta}_\beta))$, where $D(g,f)$ means a divergence, if we know the true distribution $g(x)$. As a candidate of the divergence, we can take the Cramer-von Mises type, given by $\int (G(x) - F(x;\theta))^2 \, dG(x)$. Using the empirical version and taking into consideration the idea of the cross-validation, we can propose a method for selecting the optimum β by

$$\hat{\beta} = \arg\min_\beta \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{i-0.5}{n} - F(x_i;\widehat{\theta}_\beta^{(-i)}) \right\}^2,$$

where $\widehat{\theta}_\beta^{(-i)}$ is the β-estimator obtained by leaving the i-th observation out.

REFERENCES

Basu, A., Harris, I. R., Hjort, N. L., and Jones, M. C. (1998), "Robust and Efficient Estimation by Minimising a Density Power Divergence," *Biometrika*, 85, 549–559.

Huber, P. J. (1981), *Robust Statistics*, New York: Wiley.

McLachlan, G., and Peel, D. (2000), *Finite Mixture Models*, New York: Wiley.

ある容量によって定義された近傍の特徴づけとロバスト推定への応用

南山大学・経営・院　安藤 雅和
南山大学・数理情報　木村 美善

1 はじめに

推定量のロバストネス (頑健性) をはかる測度として，最も重要なもののひとつは，モデル分布の近傍上での推定量の最大バイアスである．これは推定量の大域的なロバストネスを考察するうえで有用な情報を多くもっており,「ずれ」に対処することのできる範囲の限界をあらわす破綻点やモデル分布からの微小の「ずれ」に対する影響の大きさを表す GES に関する情報も含んでいる．この最大バイアスをロバストネスの主要な測度とし，これを最小にする推定量を求めるというバイアスーロバストネス アプローチは Huber (1964) により提唱されたが，その重要性が近年再認識されつつある．

本報告では，標本の分布のモデル分布からの「ずれ」を表現する近傍として，特別容量 (special capacity) により定義される近傍を提案し，この近傍の特徴づけ定理と関連するいくつかの基本定理を与える．そして，これらの結果を代表的な推定量のバイアス・ロバストネス研究に応用し，最大バイアスを評価する．

2 定義と基本的結果

$\mathcal{X}$ を完備可分距離空間, $\mathcal{B}$ を $\mathcal{X}$ の部分集合からなる Borel 集合族, $\mathcal{M}$ を $(\mathcal{X},\mathcal{B})$ 上の確率測度の全体からなる集合とする．

いま $F\in\mathcal{M}$ の近傍として

$$\mathcal{P}_F(c,\gamma)=\{F\in\mathcal{M}\ :\ G(B)\le c\,F(B)+\gamma,\ ^\forall B\in\mathcal{B}\},\tag{1}$$

を考える．ここで $0\le\gamma<1,\ 1-\gamma\le c<\infty$．この近傍は次のように表すことができる (Ando and Kimura 2001)．

定理 2.1

$$\mathcal{P}_F(c,\ \gamma)=\{F=c(F-W)+\gamma K\mid W\in\mathcal{W}_{F,\lambda},\ K\in\mathcal{M}\}.\tag{2}$$

ただし $\mathcal{W}_{F,\lambda}$ は $^\forall B\in\mathcal{B}$ に対して $W(B)\le F(B)$ であり，$W(\mathcal{X})=\lambda=(c+\gamma-1)/c$ となるようなすべての測度 W からなる集合である．

実数直線 R 上の確率測度 F_0 はルベーグ測度に関して単峰で原点対称な確率密度関数 f_0 をもつとし，a を F_0 の上側 $\frac{100(c+\gamma-1)}{2c}$%点とする．また $\hat{W}$ を

$$\hat{W}(B)=F_0(B\cap[-a,\ a]^c),\qquad ^\forall B\in\mathcal{B}$$

により定義し，Δ_0 は原点 0 において確率 1 をとる 1 点分布を表すとする．

定理 2.2 X,Y を確率変数とするとき，$F \times F(F \in \mathcal{P}_{F_0}(c,\gamma))$ のもとでの $|X-Y|$ の分布は $F = \hat{F} = c(F_0 - \hat{W}) + \gamma\Delta_0$ のとき確率的に最小になる．すなわち $0 \leq {}^{\forall}t < \infty$ に対して

$$\sup_{F \in \mathcal{P}_{F_0}(c,\gamma)} P(|X-Y| \leq t) = P_{\hat{F}\times\hat{F}}(|X-Y| \leq t) \tag{3}$$

が成り立つ．

3 ロバスト推定への応用

1. 位置母数の推定問題では，すべての位置共変推定量のクラスにおいて ε-contamination 近傍を含み，その一般化となる (2) の近傍上での最大バイアスを最小にする推定量はメディアンであることを示した．
2. 尺度母数の推定問題では，代表的なロバスト尺度推定量である MAD (median absolute deviation) と Rousseeuw and Croux (1993) が提案した S と Q の近傍 (2) 上での内破バイアスを導出した．
3. 確率分布族 $\{F_\theta\}$ における母数 θ のロバスト推定量に関しては，(2) の近傍上での推定量の最大バイアスの下界を導出し，He and Simpson (1993) の結果の拡張となることを示した．
4. ロバスト回帰への応用としては，Rousseeuw and Yohai (1984) により提案された S 推定量を取り上げ，近傍 (2) の上での S 推定量の最大バイアスの上界と下界を導出し，$c < 1$(Rieder(1977) の近傍) の場合には上界と下界が一致して最大バイアスが得られることを示した．

参考文献

[1] Ando,M. and Kimura,M.(2001). A characterization of the neighborhoods defined by certain special capacities and their applications to bias-robustness of estimates. submitted.

[2] Ando,M. and Kimura,M.(2001). The maximum bias of S-estimates for regression over the neighborhoods defined by certain special capacities. unpublished manuscript.

[3] He,X. and Simpson,D.G.(1993). Lower bounds for contamination bias: Globally minimax versus locally linear estimation. *Ann. Statist.*. **21**, 314-337.

[4] Huber,P.J.(1964). Robust estimation of a location parameter. *Ann. Math. Statist.*, **35**, 73-101.

[5] Martin,R.D., Yohai,V.J., and Zamar,R.H.(1989). Min-max bias robust regression, *Ann. Statist.*, **17**, 1608-1630.

[6] Rieder,H.(1977). Least favorable pairs for special capacities. *Ann. Statist.*, **6**, 1080-1094.

[7] Rousseeuw, P.J.(1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79**, 871-880.

[8] Rousseeuw, P.J., and Yohai,V.J.(1984). Robust regression by means of S-estimators, *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26**. Springer, New York, 256-272.

[9] Rousseeuw, P.J., and Croux, C.(1993). Alternatives to the median absolute deviation, *J. Amer. Statist. Assoc.* **88**, 1273-1283.

V-, LB-統計量を含む U-統計量の線形結合の Berry-Esseen bound

鹿児島大理工研 D2　戸田 光一郎
鹿児島大理学部　大和 元

1. 序

$X_1,...,X_n$ を分布 F からの大きさ n の標本とする. 次数 k の対称な kernel $g(x_1,...,x_k)$ をもつ estimable parameter $\theta(F)$ の推定量として, U-統計量 U_n と V-統計量 V_n ;

$$U_n = \binom{n}{k}^{-1} \sum_{1\le j_1<\cdots<j_k\le n} g(X_{j_1},...,X_{j_k}),$$
$$V_n = \frac{1}{n^k}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n} g(X_{j_1},...,X_{j_k})$$

が良く知られている (例えば, Lee (1990) を参照). 一方, Yamato (1977) により, $\theta(F)$ の推定量として極限 Bayes 推定量 (LB-統計量) B_n ;

$$B_n = \binom{n+k-1}{k}^{-1} \sum_{r_1+\cdots+r_n=k} g(\underbrace{X_1,...,X_1}_{r_1},...,\underbrace{X_n,...,X_n}_{r_n})$$

が得られた. 但し, $\sum_{r_1+\cdots+r_n=k}$ は $r_1+\cdots+r_n=k$ を満たすすべての非負整数 $r_1,...,r_n$ 上でとられる和を表す.

Toda and Yamato (2001) は, V-統計量, LB-統計量を含む U-統計量の線形結合 Y_n を提案した. Y_n は以下のようにして与えられる.

$j=1,...,k$ に対して, $w(r_1,...,r_j;k)$ を $r_1+\cdots+r_j=k$ を満たす正整数 $r_1,...,r_j$ の対称な非負関数とする. 但し, k は kernel g の次数である. $j=1,...,k$ に対して, $w(r_1,...,r_j;k)$ の少なくとも 1 つは正であるとし,

$$d(k,j) = \sum\nolimits^{+}_{r_1+\cdots+r_j=k} w(r_1,...,r_j;k)$$

とおく. 但し, $\sum^{+}_{r_1+\cdots+r_j=k}$ は $r_1+\cdots+r_j=k$ を満たすすべての正整数 $r_1,...,r_j$ 上でとられる和を表す. $j=1,...,k$ に対して, $g_{(j)}(x_1,...,x_j)$ を

$$g_{(j)}(x_1,...,x_j) = \frac{1}{d(k,j)}\sum\nolimits^{+}_{r_1+\cdots+r_j=k} w(r_1,...,r_j;k) g(\underbrace{x_1,...,x_1}_{r_1},...,\underbrace{x_j,...,x_j}_{r_j})$$

により与えられる kernel とし, $g_{(j)}(x_1,...,x_j)$ に対応する U-統計量を $U_n^{(j)}$ とする. このとき, $\theta(F)$ の推定量として

$$Y_n = \frac{1}{D(n,k)}\sum_{j=1}^{k} d(k,j)\binom{n}{j} U_n^{(j)}$$

を提案した. 但し, $D(n,k)=\sum_{j=1}^{k} d(k,j)\binom{n}{j}$ である. 特に, $r_1+\cdots+r_j=k\,(j=1,...,k)$ となる正整数 $r_1,...,r_j$ に対して, $w(r_1,...,r_j;k)=1$ のとき Y_n は LB-統計量 B_n となり, $w(r_1,...,r_j;k)=k!/(r_1!\cdots r_j!)$ のとき Y_n は V-統計量 V_n となる.

本報告では, Y_n に対する正規分布への収束の速さと法則収束の下限の評価について紹介する. 以下では, $r\,(\ge 1)$ に対して $\mu_r, \nu_r, \xi, \gamma, C, C_r$ を分布 F に依存しない正定数とし, $\Phi(x)$ を標準正規分布の分布関数とおく. また, $j=1,...,k$ に対して, $\psi_j(x_1,...,x_j)=E\{g(X_1,...,X_k)\mid X_1=x_1...X_j=x_j\}$, $\sigma_j^2=Var\{\psi_j(X_1,...,X_j)\}$ とおく.

2. Y_n に対する正規分布への収束の速さ

Toda and Yamato (2001) は, Y_n に対する Berry-Esseen bound を得た. $\sigma_1^2>0$ とし,

$$1-\frac{d(k,k)}{D(n,k)}\binom{n}{k} = \frac{1}{D(n,k)}\sum_{j=1}^{k-1} d(k,j)\binom{n}{j} \le \frac{\beta_1}{n}$$

を満たす正定数 β_1 が存在すると仮定する.

Theorem 1　$E\mid g(X_1,...,X_k)-\theta\mid^3\le\xi<\infty$ とし, 任意の $j_1,...,j_k\ (1\le j_1\le\cdots\le j_k\le k)$ に対して, $E\mid g(X_{j_1},...,X_{j_k})-Eg(X_{j_1},...,X_{j_k})\mid^2\le\nu_2$ とする. 但し, $k\sigma_1^2\le\nu_2\le\xi^{2/3}$. このとき, 任意の $n\ (\ge k)$ に対して

$$\sup_x \left| \Pr\left[\frac{\sqrt{n}(Y_n-EY_n)}{k\sigma_1}\le x\right]-\Phi(x)\right| \le \frac{C\xi}{k^3\sigma_1^3}n^{-1/2}.$$

更に,

$$\frac{1}{D(n,k)}\sum_{j=1}^{k-2} d(k,j)\binom{n}{j} \le \frac{\beta_2}{n^2}$$

を満たす正定数 β_2 が存在すると仮定する. このとき, Theorem 1 よりさらに弱い条件の下で Y_n に対する Berry-Esseen bound を得る.

Theorem 2 $E\mid g(X_1,...,X_k)-\theta\mid^3 \le \xi < \infty$, $E\mid g(X_1,...,X_k)-\theta\mid^{3/2} \le \gamma$, $E\mid g(X_1,X_1,X_2,...,X_{k-1})-Eg(X_1,X_1,X_2,...,X_{k-1})\mid^{3/2} \le \gamma$ とし, 任意の $j_1,...,j_k$ $(1 \le j_1 \le \cdots \le j_k \le k)$ に対して, $E\mid g(X_{j_1},...,X_{j_k})\mid \le \mu_1$ とする. 但し, $0 < 2\mu_1 \le \gamma^{2/3}$, $\gamma^2 \le \xi$. このとき, 任意の $n\ (\ge k)$ に対して

$$\sup_x \mid \Pr\left[\frac{\sqrt{n}(Y_n-\theta)}{k\sigma_1} \le x\right] - \Phi(x) \mid \le \left(\frac{C_1\xi}{k^3\sigma_1^3} + \frac{C_2\gamma}{k^{3/2}\sigma_1^{3/2}} + \frac{C_3\mu_1}{k\sigma_1}\right)n^{-1/2}.$$

Borovskikh (1996) は, 次数 2 の U-統計量に対して Berry-Esseen bound とは異なる形の正規分布への収束の早さについて示した. Borovskikh (1996) の結果を用いて, 次数 2 の Y_n に対する上で述べた結果とは異なる形の結果を得る.

Theorem 3 $E\mid g(X_1,X_2)-\theta\mid^3 < \infty$ とし, 任意の整数 j_1,j_2 $(1 \le j_1 \le j_2 \le 2)$ に対して $E\mid g(X_{j_1},X_{j_2})-Eg(X_{j_1},X_{j_2})\mid^3 \le \nu_3$ とする. このとき, 任意の $x\ (\in \mathcal{R})$ と $n\ (\ge 8)$ に対して

$$\mid \Pr\left[\frac{\sqrt{n}(Y_n-EY_n)}{2\sigma_1} \le x\right] - \Phi(x) \mid \le \frac{C}{(1+\mid x\mid)^3}n^{-1/2}.$$

3. Y_n に対する法則収束の下限の評価

$\mathcal{R}^2$ 上の標本を考え, 次数 2 の kernel と対応する Y_n について考える. $w(2;2)=w_2$, $w(1,1;2)=w_{1,1}$ とおく. $h(x_i,x_j)=g(x_i,x_j)-\psi_1(x_i)-\psi_1(x_j)$ とおくと, 次数 2 の Y_n は

$$Y_n = \frac{(n-1)w_{1,1}+2w_2}{D(n,2)}\sum_{i=1}^n \psi_1(X_i) + \frac{1}{D(n,2)}\{w_2\sum_{i=1}^n h(X_i,X_i) + w_{1,1}\sum_{1\le i<j\le n} h(X_i,X_j)\} \quad (1)$$

と表すことができる. $Eg(X_1,X_2)=0$, $E\psi_1^2(X_1)=1$ とし, (1) の右辺の第 1 項の分散を τ_n^2 とおく.

$\eta_1,\eta_2,...$ を, 独立で同一の標準正規分布に従う確率変数の列とし, $Z_1, Z_2, ...$ を, 独立で同一の分布に従う確率変数の列とする. さらに, η_i と Z_j は独立とし, $\mathcal{C}$ によりすべての確率変数 $X_i=(\eta_i,Z_i)\in\mathcal{R}^2$ の族を表す. Bentkus, Götze and Zitikis (1994) は, 関数 g ;

$$g(x,y)=x_1+y_1+x_2y_2;\ x=(x_1,x_2),\ y=(y_1,y_2),\ x,y\in\mathcal{R}^2;\ \mathbf{X}=\mathcal{R}^2 \quad (2)$$

を用いて, $X_i\in\mathcal{C}$, $E\mid Z_i\mid^\alpha=1$ のとき, 次数 2 の U-統計量に対する法則収束の下限の評価をした.

$\alpha>0$, 可測空間 $\mathbf{X}$, 対称な関数 $g:\mathbf{X}^2\to\mathcal{R}$ に対して, $E\psi_1(X_1)=0$, $E\psi_1^2(X_1)=1$, $E\mid\psi_1(X_1)\mid^3=4$, $Eh(x_1,X_2)\equiv 0$, $E\mid h(X_1,X_2)\mid^\alpha=1$, $E\mid h(X_1,X_1)\mid^{\alpha/2}\le 1$ を満たすすべての確率変数の族を $\mathcal{B}_\alpha(\mathbf{X},g)$ とする. 族 $\mathcal{C}$ における確率変数 $X_i=(\eta_i,Z_i)\in\mathcal{R}^2$ $(i=1,2,...)$ に対して, $\mathcal{B}_\alpha(\mathbf{X},g)$ が空ではない (2) により表される関数 g を考える (即ち, $g(X_i,X_j)=\eta_i+\eta_j+Z_iZ_j$). このとき, (1) より

$$\tau_n^{-1}Y_n \stackrel{d}{=} \eta_1 + \frac{1}{\tau_n D(n,2)}\{w_2\sum_{i=1}^n Z_i^2 + w_{1,1}\sum_{1\le i<j\le n} Z_iZ_j\}$$

となり, $T_n=(w_2\sum_{i=1}^n Z_i^2+w_{1,1}\sum_{1\le i<j\le n}Z_iZ_j)/\{\tau_nD(n,2)\}$ とおくと

$$\sup_x \mid \Pr[\tau_n^{-1}Y_n \le x]-\Phi(x)\mid = \sup_x \mid E\Phi(x-T_n)-\Phi(x)\mid \quad (3)$$

が成り立つ. (3) の下限の評価を, $E\mid Z_i\mid^\alpha=1$ を満たし高々 3 つの値をとるすべての対称な確率変数 Z_i 上で考えることにより, 次数 2 の Y_n に対する法則収束の下限の評価を得る.

Theorem 4 $\varepsilon \le 2/3$ とする. 任意の $n\ (\ge 1)$ に対して,

$$\sup\sup_x \mid \Pr[\tau_n^{-1}Y_n \le x]-\Phi(x)\mid \ge C(\varepsilon)n^{3\varepsilon/4-1/2}$$

となる ε のみに依存する定数 $C(\varepsilon)\ (>0)$ と可測空間 $\mathbf{X}$ と対称な関数 g が存在する. 但し, 最初の sup は, すべての確率変数 $X_i=(\eta_i,Z_i)\in\mathcal{R}^2$ $(i=1,2,...)$ 上でとられる. ここで, η_i は標準正規分布に従う確率変数であり, Z_i は $E\mid Z_i\mid^{2-\varepsilon}=1$ $(2-\varepsilon>0)$ をみたす対称な離散型確率変数である.

参考文献

[1] Bentkus, V., Götze, F. and Zitikis, R. (1994), *Ann. Probab.* **22**, 1707–1714.
[2] Borovskikh, Yu. V. (1996), *U-statistics in Banach spaces.* VSP, Utrecht.
[3] Lee, A.J. (1990), *U-statistics.* Marcel Dekker, New York.
[4] Toda, K. and Yamato, H. (2001), *J. Japan Statist. Soc.* **31**(2), (to appear).
[5] Yamato, H. (1977), *J. Japan Statist. Soc.* **7**, 57–66.

Estimation of Moments Parameters in Elliptical Distributions

東京理科大学大学院理学研究科　丸山 芳人
東京理科大学理学部　瀬尾　隆

1　はじめに

平均，分散がそれぞれ $E(\boldsymbol{X}) = \boldsymbol{\mu}$, $\Sigma \equiv Cov(\boldsymbol{X}) = -2\psi'(0)\Lambda$ で与えられる p 変量楕円分布 $E_p(\boldsymbol{\mu}, \Lambda)$ に従う確率ベクトル $\boldsymbol{X}$ のモーメントパラメータと，その推定について議論する．

2　楕円分布

p 変量楕円分布 $E_p(\boldsymbol{\mu}, \Lambda)$ に従う $\boldsymbol{X}$ の密度関数と特性関数は，次のように定義される (Muirhead(1982) 参照).

$$f(\boldsymbol{x}) = c_p|\Lambda|^{-\frac{1}{2}} g\left((\boldsymbol{x}-\boldsymbol{\mu})'\Lambda^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right), \quad c_p\text{は正定数}, g\text{ はある非負関数} \tag{2.1}$$

$$\phi(\boldsymbol{t}) = \exp[i\boldsymbol{t}'\boldsymbol{\mu}]\psi(\boldsymbol{t}'\Lambda\boldsymbol{t}), \quad \psi\text{はある関数} \tag{2.2}$$

3　モーメント

$\boldsymbol{Y} = A^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$, $AA' = \Lambda$ とおき，$R^2 = \boldsymbol{Y}'\boldsymbol{Y}$, $\boldsymbol{U} = \boldsymbol{Y}/R$ とする．ここで，$\boldsymbol{U}$ は $\boldsymbol{U}'\boldsymbol{U} = 1$ の一様分布に従い，$R(\geq 0)$ と $\boldsymbol{U}$ は独立である．R と $\boldsymbol{U}$ のモーメントを求めることにより，次の定理が成り立つ．

【定理 2.1】 $\boldsymbol{X} - \boldsymbol{\mu}$ の奇数次モーメントは 0，$2m$ 次モーメントは，次のように表される．

$$\mu_{ij\cdots uv} = (\kappa_{(m)} + 1)\sum_{(d_m)} \sigma_{ij}\cdots\sigma_{uv}$$

ここで，

$$\kappa_{(m)} + 1 = \frac{\psi^{(m)}(0)}{\{\psi'(0)\}^m}, \qquad d_m = 2^m\left(\frac{1}{2}\right)_m = 2^m\frac{\Gamma(\frac{1}{2}+m)}{\Gamma(\frac{1}{2})}$$

4　推定

モーメントパラメータ $\kappa_{(m)}$ の推定について，Anderson(1993) が議論している尖度パラメータ $\kappa \equiv \kappa_{(2)}$ の一致推定量の修正と一般形を考えた．Anderson(1993) による結果は，

$$E\left[\{(\boldsymbol{X}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\}^2\right] = p(3\kappa + p + 2)$$

これを修正すると，

$$E\left[\{(\boldsymbol{X}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\}^2\right] = (\kappa+1)p(p+2)$$

であり，κ の一致推定量は次のように与えられる．

$$\hat{\kappa} = \frac{1}{p(p+2)}\hat{\beta}_{2,p} - 1, \quad \hat{\beta}_{2,p} \equiv \frac{1}{N}\sum_{i=1}^{N}\left\{(\boldsymbol{X}_i - \overline{\boldsymbol{X}})'S^{-1}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})\right\}^2$$

これより，一般形は次の結果となる．

【定理 4.1】 $\boldsymbol{X}-\mu$ の $2m$ 次モーメントパラメータ $\kappa_{(m)}$ の一致推定量 $\hat{\kappa}_{(m)}$ は，次のように与えられる．

$$\hat{\kappa}_{(m)}=\frac{\Gamma(\frac{p}{2})}{2^m\Gamma(\frac{p}{2}+m)}\hat{\beta}_{m,p}-1,\quad \hat{\beta}_{m,p}=\frac{1}{N}\sum_{i=1}^{N}\left\{(\boldsymbol{X}_i-\overline{\boldsymbol{X}})'S^{-1}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})\right\}^m \tag{4.3}$$

本報告では，$T=\sqrt{N}(\hat{\kappa}_{(m)}-\kappa_{(m)})$ の漸近分散について，(1) Σ が既知のとき，(2) Σ が未知のときに分けて漸近展開の形で議論する．一般性を失うことなく $\Sigma=I_p$ とする．まず (1) の場合は次の結果を得る．

【定理 4.2】 Σ が既知のとき，$2m$ 次モーメントパラメータ $\kappa_{(m)}$ の一致推定量を

$$\hat{\kappa}^*_{(m)}=\frac{1}{2^m(\frac{p}{2})_m}\hat{\beta}^*_{m,p}-1\ ,\quad \hat{\beta}^*_{m,p}=\frac{1}{N}\sum_{i=1}^{N}\left\{(\boldsymbol{X}_i-\overline{\boldsymbol{X}})'\Sigma^{-1}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})\right\}^m$$

とする．このとき，$T=\sqrt{N}(\hat{\kappa}^*_{(m)}-\kappa_{(m)})$ の漸近分散は，

$$\sigma^2_{\hat{\kappa}_{(m)}}=\frac{(\frac{p}{2}+m)_m}{(\frac{p}{2})_m}(\kappa_{(2m)}+1)-(\kappa_{(m)}+1)^2 \tag{4.4}$$

次に，(2) の場合を考える．しかし大変に複雑な計算を伴うので，$m=2$ の場合，つまり尖度パラメータ κ の一致推定量 $\hat{\kappa}$ について漸近的性質を考えることにする．これは，Seo and Toyama(1996) の結果と一致する．

【定理 4.3】 Σ が未知のとき，4 次モーメントパラメータ (尖度パラメータ) κ の一致推定量を

$$\hat{\kappa}=\frac{1}{p(p+2)}\hat{\beta}_{2,p}-1\ ,\quad \hat{\beta}_{2,p}=\frac{1}{N}\sum_{i=1}^{N}\left\{(\boldsymbol{X}_i-\overline{\boldsymbol{X}})'S^{-1}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})\right\}^2$$

とする．このとき，$T=\sqrt{N}(\hat{\kappa}-\kappa)$ の漸近分散は，

$$\sigma^2_{\hat{\kappa}}=\frac{(p+4)(p+6)}{p(p+2)}(\kappa_{(4)}+1)-\frac{4(p+4)}{p}(\kappa_{(3)}+1)(\kappa+1)+\frac{4(p+2)}{p}(\kappa+1)^3-(\kappa+1)^2 \tag{4.5}$$

参考文献

[1] Anderson, T. W. (1993). Nonnormal multivariate distributions: Inference based on elliptically contoured distributions. *Multivariate Analysis. Future Directions*, (C.R.Rao ed.), Elsevier Science Publishers, pp.1-24.

[2] Kellker, D. (1970). Distribution theory of spherical distributions and a location scale parameter generalization, *Sankhya* **A32**, 419-430.

[3] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519-530.

[4] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, NewYork.

[5] Seo, T. and Toyama, T. (1996). On the estimation of kurtosis parameter in elliptical distributions. *J. Japan Statist. Soc* **26**, 59-68.

共分散構造に関する尤度比検定統計量の帰無分布の非正規性に対する影響

広島大学大学院理学研究科　冨田 哲治
広島大学大学院理学研究科　松本 智恵子
統計数理研究所　柳原 宏和

本研究では，非正規性のもとで共分散行列のある構造に関する仮説検定の尤度比統計量の帰無分布を取り扱う．考える共分散構造は，(i) $H_{01}:\Sigma=\Sigma_0$，(ii) $H_{02}:\Sigma=\sigma^2I_p$，(iii) $H_{03}:\Sigma=\mathrm{diag}(\Sigma_{11},\cdots,\Sigma_{kk})$，(iv) $H_{04}:\Sigma_1=\cdots=\Sigma_k$ の4つ，つまり　(i) は共分散行列がある値かどうか，(ii) は単位行列の定数倍，(iv) は k 個の母集団の共分散行列の同等性の仮説であり，(iii) は p 次元ベクトル $\boldsymbol{x}=(x_1,\cdots,x_p)'$ を k セット $\boldsymbol{x}=(\boldsymbol{x}_1',\cdots,\boldsymbol{x}_k')'$, $\boldsymbol{x}_i:p_i\times 1$, に分割し，それに対応した Σ の分割を

$$\Sigma=\begin{bmatrix}\Sigma_{11} & \cdots & \Sigma_{1k}\\ \vdots & \cdots & \vdots\\ \Sigma_{k1} & \cdots & \Sigma_{kk}\end{bmatrix},\quad \Sigma_{ab}\ :\ p_i\times p_j \text{ sub matrix of } \Sigma$$

としたとき，共分散行列が対角ブロック行列であるいう仮説である．これら4つの仮説に対する尤度比統計量を L_i, $(i=1,\cdots,4)$ とする．正規性の下では，$T_i=-2\log L_i$ の帰無分布は漸近的に χ^2 分布に従うことが知られており，それらの漸近展開も導出されている (see, e.g., Anderson, 1984; Siotani, Hayakawa and Fujikoshi, 1985)．これら正規性に基づいた研究結果は数々得られているが，この正規性の仮定を外したときの研究も進められている．Ito (1969) では，非正規性の下で帰無分布の漸近分布が母集団の4次キュムラントに依存することを報告している．一般に分散に関する検定は非正規性に関して頑健ではないと言われる理由がこの漸近分布が非正規と正規の場合に異なるということであり，これが平均に関する検定と決定的に異なる点である．このように分布のずれに関して敏感に反応してしまう検定において，非正規性の影響を調べることは非常に重要な問題であると考える．しかしながら Ito (1969) では漸近分布の具体的な形には触れていない．具体的な形としては，Muirhead and Waternaux (1980) は楕円分布の下での帰無分布の漸近分布を求めている．これから，κ を kurtosis parameter とすると，T_2, T_3 は漸近的に χ^2 分布を $(1+\kappa)$ 倍した分布と同値であり，統計量を $(1+\kappa)$ で割って修正したものの漸近分布は χ^2 分布となる．また，Tyler (1983) は楕円分布の下で $(1+\kappa)$ で割って尤度比統計量を修正することが可能であるような一般的な条件を与えている．それは，仮説のタイプに関する条件であり condition h と呼ばれる．楕円分布の下で $H_0:h(\Sigma)=0$ のタイプの仮説に対してその尤度比等軽量を T とすると，condition h を満たすとき : $T\longrightarrow(1+\kappa)\chi^2_r$ であり，condition h を満たさ

ないとき : $T \longrightarrow (1+\kappa)\chi^2_{r-1} + \{(1+\kappa)+\kappa\delta_h(\Sigma,\Sigma)\}\chi^2_1$ である．$\Gamma=\Sigma$ の近傍でその導関数が連続な $p\times p$ 正定値行列上の r 次元関数 $h(\Gamma)$ に対して，$h(\alpha\Gamma)=h(\Gamma)$ が任意の $\alpha>0$ で正定値行列 Γ に対して成り立つとき，condition h が成り立つという．Hayakawa (1999) ではこの条件を満たす具体的な仮説の例をあげている．このように修正すれば，楕円分布のクラスでも漸近分布は kurtosis parameter の影響を受けないのである．しかし，このような修正はいつも可能というわけではなく，例えば，T_1 と T_4 に関しては修正を行うことはできない．なぜならば，T_1, T_4 の帰無分布は楕円分布の下で漸近的に異なる重みの 2 つの独立な χ^2 変量の重み付き和の分布と同値となるからである．本研究の目的は，分布の条件を楕円分布よりもさらに弱めた具体的な分布を仮定しない非正規性の下で，帰無分布の非正規性の影響を調べることである．

4 つの仮説 (i)～(iv) に対する尤度比統計量 L_i は，線形変換 $\boldsymbol{x}\rightarrow\Sigma^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu})$ に関して不変なので，帰無仮説の下で考えるとき，一般性を失うことなく $\boldsymbol{\mu}=\mathbf{0}$, $\Sigma=I_p$ とすることができる．以下の議論では $\boldsymbol{\mu}=\mathbf{0}$, $\Sigma=I_p$ ととして考える．仮説 (i)$\Sigma=\Sigma_0$ の検定統計量 $T_1=-2\log L_1$ は標本共分散行列に基づく統計量 $U=\sqrt{n}(S-I_p)$ で表現され，対称行列 U の上三角成分を並べたベクトル

$$\boldsymbol{u}_s=\mathrm{vecs}(U)=\left(u_{11}/\sqrt{2},\cdots,u_{pp}/\sqrt{2},u_{12},\cdots,u_{1,p},u_{23},\cdots,u_{p-1p}\right)'.$$

と既知行列 D_1 を用いて，$T_1=\boldsymbol{u}_sD_1\boldsymbol{u}_s+\mathrm{O}_p(n^{-1/2})$ と表現される．$\boldsymbol{u}_s$ は漸近正規性をもち，漸近的に平均 $\mathbf{0}$, 共分散行列 Ω_s の $q=p(p+1)/2$ 次元多変量正規分布に従うことから，T_1 の漸近分布の特性関数は以下のように書ける．

$$C_1(t)=|I_q-2it\Omega_s|^{-1/2}=\prod_{j=1}^{q}(1-2it\lambda_j^{(1)})^{-1/2}$$

ただし $\lambda_j^{(1)}$ $(j=1,\cdots,q)$ を Ω_s の固有値とする．これより，T_1 の漸近分布は独立な自由度 1 の χ^2 変量の重み (Ω_s の固有値) 付き和の分布と同値であることがわかる．他の 3 つの統計量に関しても同様な形で表現できる．これらの結果から次の定理が成り立つ．

THEOREM 1. U は標本共分散に基づいた，漸近正規性をもつ統計量とする．つまり $\boldsymbol{u}=\mathrm{vecs}(U)\rightarrow\mathrm{N}_{\alpha}(\mathrm{vecs}(\Sigma),\Omega)$, ただし $\Omega=\mathrm{Cov}(\boldsymbol{u})$. このとき統計量が $T=\boldsymbol{u}'D\boldsymbol{u}+\mathrm{O}_p(n^{-1/2})$ と展開できたとすると，$(\Omega D)^2=c(\Omega D)$, (c は定数) であれば $T/c\longrightarrow\chi^2_d$, $(n\rightarrow\infty, d=\mathrm{rank}(D))$.

参考文献

[1] Muirhead, R. J. and Waternaux, C. M. (1980). *Biometrika*, **67**, 31-43.

[2] Tyler, D. E. (1983). *Biometrika*, **70**, 411-420.

他省略.

HIGHER ORDER EFFICIENCY OF LINEAR COMBINATIONS OF U-STATISTICS AS ESTIMATERS OF ESTIMABLE PARAMETERS

野町俊文 (都城高専)、近藤正男 (鹿児島大学・理)、大和　元 (鹿児島大学・理)

$\theta(F)$ は対称なカーネル $g(x_1,\ldots,x_k)$ を持つ分布 Fの汎関数または推定可能な母数とする。また、$\{X_1,\ldots,X_n\}$ を分布関数 Fからの大きさ n の標本とするとき、Toda and Yamato (2001) は$\theta(F)$ の推定量として、U-統計量の線形結合 Y_nを次のように導入している。

２つの 正の整数 $k=1,2\ldots$ と $j=1,\ldots,k$に対して $w(r_1,\ldots,r_j;k)$ は $r_1+\cdots+r_j=k$を満たす任意の正の整数の組 $(r_1,\ldots,r_j)$ に対して非負の値を取り、その中の少なくとも一つは正の値を取る対称な関数とする。また、$j=1,\ldots,k$に対して、関数 $g_{(j)}(x_1,\ldots,x_j)$ を $d(k,j)\neq 0$ のとき、

$$g_{(j)}(x_1,\ldots,x_j)=\frac{1}{d(k,j)}\sum\nolimits^{+}_{r_1+\cdots+r_j=k}w(r_1,\ldots,r_j;k)g(\underbrace{x_1,\ldots,x_1}_{r_1\text{個}},\cdots,\underbrace{x_j,\ldots,x_j}_{r_j\text{個}})$$

と置く。ただし、記号$\sum^{+}$は $r_1+\cdots+r_j=k$ を満たす全ての正の整数の組に対して和を取り、$j=1,\ldots,k$に対して、$d(k,j)=\sum^{+}_{r_1+\cdots+r_j=k}w(r_1,\ldots,r_j;k)$ とする。また、$U_n^{(i)}$はカーネル $g_{(i)}$に対応する U-統計量を表すとする。このとき、統計量 Y_nを次のように与える。

$$Y_n=\frac{1}{D(n,k)}\sum_{j=1}^{k}d(k,j)\begin{pmatrix} n \\ j \end{pmatrix}U_n^{(j)} \tag{1}$$

ただし、 $D(n,k)=\sum_{j=1}^{k}d(k,j)\begin{pmatrix} n \\ j \end{pmatrix}$とする。
wを $j=1,\ldots,k$、$r_1+\cdots+r_j=k$ を満たす、正の整数の組 $r_1,\cdots,r_j$に対して、$w(r_1,\ldots,r_j;k)=k!/(r_1\cdots r_j)$ によって与えられる関数とする。新たな統計量 S_nを次のように考える。カーネルは

$$g_{(j)}(x_1,\ldots,x_j)=\frac{1}{j!|s(k,j)|}\sum\nolimits^{+}_{r_1+\cdots+r_j=k}\frac{k!}{r_1\cdots r_j}g(\underbrace{x_1,\ldots,x_1}_{r_1\text{個}},\cdots,\underbrace{x_j,\ldots,x_j}_{r_j\text{個}})$$

である。ただし、$D(n,k)=\sum_{j=1}^{k}|s(k,j)|(n)_j$ であり、$s(k,j)$ は第１種のスターリング数を表す。これらのカーネルに対応する U-統計量 $U_n^{(j)}$によって、S_nは

$$S_n=\frac{1}{D(n,k)}\sum_{j=1}^{k}|s(k,j)|(n)_jU_n^{(j)}$$

によって表される。この S_nで表される統計量をS-統計量と呼ぶ。また、w を $j=1,\ldots,k$、$r_1+\cdots+r_j=k$ を満たす、正の整数の組 $r_1,\cdots,r_j$に対して、$w(r_1,\ldots,r_j;k)=k!/(r_1!\cdots r_j!)$ によって与えられる関数とする。このとき、統計量 Y_nは、V-統計量 V_nである (Yamato and Toda (2001), Lee (1990);p.183-184 and Koroljuk and Borovskich (1994); p.40) 。
次数 $k\geq 3$ のカーネルを考える。カーネル $g_{(j)}(x_1,\ldots,x_j)$、$j=1,\ldots,k$に対して、

$$\theta_j=E[g_{(j)}(X_1,\ldots,X_j)]$$

$$\psi_{(j),c}(x_1,\ldots,x_c)=E[g_{(j)}(X_1,\ldots,X_j)|X_1,\ldots,X_c]\quad c=1,\ldots,j$$

と置く。さらに $j=1,\ldots,k$、$c=2,3,\ldots,k$に対して、次のように置く。

$$\begin{aligned} h_{(j)}^{(1)}(x_1) &= \psi_{(j),1}(x_1)-\theta_j \\ h_{(j)}^{(c)}(x_1,\ldots,x_c) &= \psi_{(j),c}(x_1,\ldots,x_c)-\sum_{i=1}^{c-1}\sum_{1\leq l_1<\ldots<l_i\leq c}h_{(j)}^{(i)}(x_{l_1},\ldots,x_{l_i})-\theta_j \end{aligned}$$

$D(n,k), d(k,k), d(k,k-1), d(k,k-2)$ に対して、$d(k,j) \neq 0$, $(j = k, k-1, k-2)$ のとき、$\beta_2 + \beta_{21} + \beta_{22} = 0$, $\beta_1 > 0, \beta_2 > 0, \beta_{21} > 0$ を満たす定数$\beta_1, \beta_2, \beta_{21}, \beta_{22}$を用いて、展開することができる。

$$\frac{d(k,k)}{D(n,k)}\begin{pmatrix} n \\ k \end{pmatrix} = 1 - \frac{\beta_1}{n} + \frac{\beta_{21}}{n^2} + o(\frac{1}{n^2}), \qquad \frac{d(k,k-2)}{D(n,k)}\begin{pmatrix} n \\ k-2 \end{pmatrix} = \frac{\beta_2}{n^2} + o(\frac{1}{n^2})$$

$$\frac{d(k,k-1)}{D(n,k)}\begin{pmatrix} n \\ k-1 \end{pmatrix} = \frac{\beta_1}{n} + \frac{\beta_{22}}{n^2} + o(\frac{1}{n^2})$$

ノンパラメトリックな分布のクラスの Efficiency を考え、Forth Order Efficiency を定義する。

定義　2つの統計量 $Y_{i,n}$, $(i = 1,2)$ の平均2乗誤差は $\mathrm{MSE}(Y_{i,n})$, $(i = 1,2)$ であるとする。もし、有限な極限値

$$\lim_{n\to\infty} n^2 \Big\{1 - \frac{MSE[Y_{1,n}]}{MSE[Y_{2,n}]}\Big\}$$

存在するとき、この極限値を $Y_{2,n}$の、$Y_{1,n}$に関する Forth Order Efficiency(FOE) といい、$FOE(Y_{2,n}, Y_{1,n})$ で表す。

2つの統計量 $Y_{i,n}$, $(i = 1,2)$ は異なる関数 wによって (1) で与えられる統計量とし、β_1は同じ値を取るものとする。$Y_{1,n}$に対して、$\beta_{2,1}$, $\beta_{21,1}$, $\beta_{22,1}$, $\zeta^{(1)}_{k-2,1}$, $\theta_{k-2,1}$によって表し、$Y_{2,n}$に対して、$\beta_{2,2}$, $\beta_{21,2}$, $\beta_{22,2}$, $\zeta^{(1)}_{k-2,2}$, $\theta_{k-2,2}$によって表す。$Y_{1,n}$と $Y_{2,n}$に対する $g_{(k)}$と $g_{(k-1)}$は、それぞれ一致するので、対応する$\zeta^{(1)}_{k-1}$は同じ値になる。ただし $i = 1,2,3$, $j = 1,2$, $c = k, k-1$ に対して

$$\delta^2_{c,(i)} = Var[h^{(i)}_{(c)}(X_1,\ldots,X_i)],\ \zeta^{(i)}_{k-j} = Cov\big[h^{(i)}_{(k)}(X_1,\ldots,X_i), h^{(i)}_{(k-j)}(X_1,\ldots,X_i)\big],\ i \leq k-j$$

定理　$1 \leq i_1 \leq \cdots \leq i_k \leq k$に対して、$E[g(X_{i_1},\ldots,X_{i_k})]^2 < \infty$ を仮定する。$Y_{1,n}$と $Y_{2,n}$は (1) で、β_1のみ同じ値を取る異なる関数 wによって与えられる統計量とする。$\delta^2_{k,(1)} \neq 0$ のとき、$Y_{2,n}$の $Y_{1,n}$に関する FOE は

$$FOE(Y_{2,n}, Y_{1,n}) = \frac{A}{k^2\delta^2_{k,(1)}}$$

によって与えられる。ただし、

$$A = 2\Big\{k^2(\beta_{21,2} - \beta_{21,1})\delta^2_{k,(1)} + k(k-1)(\beta_{22,2} - \beta_{22,1})\zeta^{(1)}_{k-1} + k(k-2)((\beta_{2,2}\zeta^{(1)}_{k-2,2} - \beta_{2,1}\zeta^{(1)}_{k-2,1})$$
$$+\beta_1(\beta_{22,2} - \beta_{22,1})(\theta_{k-1} - \theta)^2 + \beta_1(\theta_{k-1} - \theta)\Big[\beta_{2,2}(\theta_{k-2,2} - \theta) - \beta_{2,1}(\theta_{k-2,1} - \theta)\Big]\Big\}$$

命題　次数 $k \geq 3$ を持つカーネルに対して、$E[g(X_{i_1},\ldots,X_{i_3})]^2 < \infty$, $\delta^2_{3,(1)} \neq 0$ のとき、

$$\begin{aligned} FOE(S_n, V_n) &= \frac{(k-1)(k-2)}{24\delta^2_{k,(1)}}\Big\{2(k-2)\Big[(3k-1)\zeta^{(1)}_{k-2,1} - (3k-5)\zeta^{(1)}_{k-2,2}\Big] \\ &\quad -8k\delta^2_{k,(1)} + (k-1)(\theta_{k-1} - \theta)\Big[(3k-1)\theta_{k-2,1}(3k-5)\theta_{k-2,2} - 4\theta\Big]\Big\} \end{aligned}$$

を得る。ただし、$\zeta^{(1)}_{k-2,1}$と$\theta_{k-2,1}$は V_nに関する値であり、$\zeta^{(1)}_{k-2,2}$と$\theta_{k-2,2}$は S_nに関する値である。

References

[1] Koroljuk, V. S. and Borovskich, Yu. V. (1994), *Theory of U-Statistics*, Kluwer Academic Publishers, Dordrecht, The netherlands.

[2] Lee, A. J. (1990), *U-statistics: theory and practice* . Marcel Dekker, Inc. New York.

[3] Nomachi, T. and Yamato, H. (2001), "Asymptotic Comparisons of U-Statistics, V-Statistics and Limits of Bayes Estimates by Deficiencies", *Journal of Japan Statistical Society* **31**, 85–98.

[4] Toda K. and Yamato H. (2001)Berry Esseen bounds for some statistics including LB-statistic and V-statistic. J. Japan statis. Soc. **31**, No.2 (To appear.)

ノンパラメトリックな検定統計量に基づく信頼区間

九州大学・経済学研究院　　前園宜彦

$X_1, X_2, \cdots, X_n$ を分布 $F(x-\theta)$ からの無作為標本とし，分布関数 $F(x)$ は $F(-x) = 1-F(x)$ を満たす (密度関数が $x=\theta$ に関して対称な分布) とする．このとき

$$\text{帰無仮説}\, H_0 : \theta = 0 \quad \text{v.s.} \quad H_1 : \theta > 0$$

の検定問題に対して，ノンパラメトリックな検定統計量として符号検定，ウィルコクソンの符号付順位検定が良く使われ，その性質が研究されてきた．同順位を避けるために $F(x)$ は連続型の分布とし，

$$\psi(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

とおくと，同等な検定統計量は

$$S = \sum_{i=1}^{n} \psi(X_i), \qquad W^+ = \sum_{1 \leq i \leq j \leq n} \psi(X_i + X_j)$$

で与えられる．これらの統計量の分布は H_0 の下で分布 $F(x)$ に依存しない (distribution-free).

他方これらの検定統計量に基づく θ の点推定も提案されている．S に基づく点推定はメディアンで，W^+ に基づく点推定は Hodges Lehmann 推定量として知られている．本報告ではこれらのノンパラメトリックな検定統計量に基づく θ の信頼区間の構成を議論する．

帰無仮説 H_0 の下で，それぞれの統計量の α-点に一番近い値を $s(\alpha)$, $w(\alpha)$

$$s(\alpha) = \min_s \{s \mid P_0(S \leq s) \geq \alpha\}, \qquad w(\alpha) = \min_w \{w \mid P_0(W^+ \leq w) \geq \alpha\}$$

とおく．このとき

$$P_0\Big(s(\alpha/2) \leq S \leq s(1-\alpha/2)\Big) \approx 1-\alpha,$$
$$P_0\Big(w(\alpha/2) \leq W^+ \leq w(1-\alpha/2)\Big) \approx 1-\alpha$$

となる．また対立仮説の下では

$$\tilde{S} = \sum_{i=1}^{n} \psi(X_i - \theta), \qquad \tilde{W}^+ = \sum_{1 \leq i \leq j \leq n} \psi(X_i + X_j - 2\theta)$$

の分布は H_1 の下で S, W^+ の帰無仮説の下での分布と一致する．ここで順序統計量を

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

とし，Walsh 平均と呼ばれる $\{(X_i + X_j)/2, 1 \leq i \leq j \leq n\}$ を小さい方から並べたものを

$$W_{(1)} \leq W_{(2)} \leq \cdots \leq W_{(n(n+1)/2)}$$

とおく．これらを利用すると

$$P_\theta\Big(X_{([s(\alpha/2)])} \le \theta \le X_{([s(1-\alpha/2)])}\Big) \approx 1-\alpha,$$
$$P_\theta\Big(W_{([w(\alpha/2)])} \le \theta \le W_{([w(1-\alpha/2)])}\Big) \approx 1-\alpha$$

の信頼区間が構成でき，これらはデータの従う分布に依存しない．ただし $[\cdot]$ はガウス記号とする．また漸近分布による近似も可能である．$z(\alpha)$ を標準正規分布 $N(0,1)$ の α-点とすると，それぞれの α-点の近似は

$$s(\alpha) \approx E_0(S) + z(\alpha)\sqrt{V_0(S)}, \qquad w(\alpha) \approx E_0(W^+) + z(\alpha)\sqrt{V_0(W^+)}$$

となる．ここで期待値及び分散は帰無仮説 H_0 の下で計算されるものである．

他方 Maesono (1987) では approximate Bahadur A.R.E. の意味で一様に W^+ より優れている検定統計量

$$\begin{aligned} M &= \sum_{1\le i\le j\le k\le n} \Big\{\psi(X_i+X_j)\psi(X_i+X_k) + \psi(X_i+X_j)\psi(X_j+X_k) \\ &\qquad + \psi(X_i+X_k)\psi(X_j+X_k)\Big\} \end{aligned}$$

を議論している．この M の帰無仮説の下での分布もやはり $F(x)$ に依存しない．ここで α-点を

$$m^*(\alpha) = \min_{m^*}\{m^* \mid P_0(M^* \le m^*) \ge \alpha\}$$

とおく．また $W_{[\ell]}$ を一つ固定して $i = 1.2\cdots,n$ に対して

$$u_i^\ell = \#\Big\{j \mid X_i + X_j \ge 2W_{[\ell]},\ 1 \le j \le n\Big\}$$

とおき

$$m^\ell = \sum_{i=1}^n (u_i^\ell)^2, \qquad m^{\ell^*(\alpha)} = m^*(\alpha)$$

なる自然数 $\ell^*(\alpha)$ を求める．このとき M に基づく信頼区間は

$$P_\theta\Big(W_{([\ell(\alpha/2)])} \le \theta \le W_{([\ell(1-\alpha/2)])}\Big) \approx 1-\alpha$$

で与えられる．

ウィルコクソンの符号付順位検定統計量 W^+ と統計量 M については，エッジワース展開が求められており，これを適用すると α-点の近似の改善が得られ，推測の精度を上げることができる．またこの他にもノンパラメトリックな推測として，ブートストラップ法を利用することができるが，シュミレーションの結果では，W^+ 及び M に基づく推測のほうが良くなっている．

[1] Hollander, M and Wolfe. D.A. (1999) Nonparametoric Statistical Methods. 2nd ed.
[2] Maesono, Y. (1987) Annals of the Institute of Statistical Mathematics, Vol.39, No.2, Part A, pp.363-375.

無限次元空間上の極限定理とその対称統計量への応用

金沢大学工学部　　金川秀也

対称統計量

$\{\xi_j, j \geq 1\}$ を分布 μ に従う実数値確率変数列とする．また実対称関数 $u(x_1, x_2, \cdots, x_k)$ を核関数とする統計量を対称統計量と呼ぶ． $k=2$ の対称関数 $u(x,y)$ に対して、対称統計量

$$U_n := \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} u(\xi_i, \xi_j),\ V_n := \frac{1}{n^2} \sum_{1 \leq i, j \leq n} u(\xi_i, \xi_j)$$

をそれぞれ degree 2 の U-統計量、V-統計量とよぶ．退化型核関数を持つU-またはV-統計量に対して、無限次元空間上に値を持つ確率変数列の単純和に関する極限定理を応用して、各種の漸近的な性質を調べることが出来る．

証明のアイデア

$\{\xi_j, j \geq 1\}$ を i.i.d. 確率変数列でその分布を μ とする. $u(x,y)$ $\mathbf{R} \times \mathbf{R}$上の実数値対称関数とする. また u は $\mu \times \mu$ に関し 2 乗可積分で退化している、i.e., for any real x

$$E(u(\xi_1, x)) = 0$$

とする．Serfling (1980) に従って, kernel u から有界線形作用素 $T_u : L^2 \to L^2$ (trace class) を $T_u f(x) := E[u(\xi_1, x) f(\xi_1)]$, $f \in L^2$ とすると、その固有関数 $\{g_i\}$ と固有値は $\{\lambda_i\}$ は以下の条件を満たすと仮定する．任意の $i \geq 1$ に対して

$$\begin{cases} E(g_i(\xi_1)) = 0, & E(g_i^2(\xi_1)) = 1 \\ E(g_i(\xi_1) g_j(\xi_1)) = 0\ (i \neq j), & E(h(\xi_1, x) g_i(\xi_1)) = \lambda_i g_i(x) \end{cases}.$$

このとき u は $u(\xi_i, \xi_j) = \sum_{k=1}^{\infty} \lambda_k g_k(\xi_i) g_k(\xi_j)$ と表現される．

また可分ヒルベルト空間 H とその内積$\langle \cdot,\cdot \rangle$、ノルムを$\|\cdot\|$ 次のように定義する．

$$H := \left\{ \boldsymbol{x} = (x_1, x_2, \cdots) \in \mathbf{R}^{\infty} : \sum_{k=1}^{\infty} |\lambda_k| \xi_k^2 < \infty \right\},\ \langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{k=1}^{\infty} |\lambda_k| x_k y_k,\ \|\boldsymbol{x}\| = \left(\sum_{k=1}^{\infty} |\lambda_k| x_k^2 \right)^{1/2}.$$

$\sum_{k=1}^{\infty} |\lambda_k| < \infty$ を仮定する $E\left(\sum_{k=1}^{\infty} |\lambda_k| g_k^2(\xi_i) \right) = \sum_{k=1}^{\infty} |\lambda_k| E(g_k^2(\xi_i)) = \sum_{k=1}^{\infty} |\lambda_k| < \infty$ より H-値確率変数 $G_i := (g_1(\xi_i), g_2(\xi_i), g_3(\xi_i), \cdots) \in H$ $i \geq 1$ を用いてdegree 2のU-statistics $\{U_n, n \geq 1\}$、V-statistics $\{V_n, n \geq 1\}$ を次のように表すことができる．

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} G_i = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_1(\xi_i), \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_2(\xi_i), \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_3(\xi_i), \cdots \right) \in H$$

より次の仮定 $\lambda_i \geq 0,\quad i \geq 1$ の下で

$$nV_n = \frac{1}{n}\sum_{1\leq i,j\leq n} u(\xi_i,\xi_j) = \frac{1}{n}\sum_{1\leq i,j\leq n}\sum_{k=1}^{\infty}\lambda_k g_k(\xi_i)g_k(\xi_j) = \frac{1}{n}\sum_{k=1}^{\infty}\lambda_k\sum_{1\leq i,j\leq n} g_k(\xi_i)g_k(\xi_j) = \frac{1}{n}\sum_{k=1}^{\infty}\lambda_k\left\{\sum_{i=1}^{n} g_k(\xi_i)\right\}^2 = \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}G_i\right|$$

$$\sqrt{n(n-1)}U_n = \frac{n}{\sqrt{n(n-1)}}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}G_i\right\| - \frac{1}{\sqrt{n(n-1)}}\sum_{k=1}^{\infty}\lambda_k\sum_{i=1}^{n}\frac{g_k^2(\xi_i)}{n}.$$

即ち $\{U_n, n\geq 1\}$ 及び $\{V_n, n\geq 1\}$ が H-値確率変数列 $\{G_i,\ i\geq 1\}$ の単純和で表すことができる.

大偏差原理への応用

Donsker-Varadhan (1976) 及び Bolthausen (1986) から $\{U_n, n\geq 1\}$、$\{V_n, n\geq 1\}$ に対する大偏差原理が成り立つ.

Kanagawa (1999)

$E\left(\exp\left(t\|G_1\|\right)\right)<\infty$ を仮定する. また $\Phi(x) := \left|\sum_{k=1}^{\infty}\lambda_k x_k^2\right|^{1/2}$, $x\in H$ に対して G_1 の entropy function h を $h(\mathbf{x}) := \sup_{\varphi\in H^*}\left(\varphi(\mathbf{x}) - \log M(\varphi)\right)$ とおく. ただし、H^* は H の双対空間、$M(\varphi) := E(\exp(\varphi(G_1)))$ とする. このとき

$$\lim_{n\to\infty}\frac{1}{n}\log E\left(\exp\left(n|V_n|^{1/2}\right)\right) = \sup_{\mathbf{x}\in H}\left(\Phi(\mathbf{x}) - h(x)\right).$$

$$\lim_{n\to\infty}\frac{1}{n}\log E\left(\exp\left(\left|n(n-1)U_n + \sum_{i=1}^{n}u(\xi_i,\xi_i)\right|^{1/2}\right)\right) = \sup_{\mathbf{x}\in H}\left(\Phi(\mathbf{x}) - h(\mathbf{x})\right).$$

参考文献

E. Bolthausen, Laplace approximations for sums of independent random vectors, Probab. Th. Rel. Fields, 72 (1986), 305-318.

M. D. Donsker and S. R. S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time III, Comm. Pure Appl. Math. 29 (1976), 89-461.

S. Kanagawa and K. Yoshihara, The almost sure invariance principles of degenerate U-statistics of degree two for stationary random variables, Stochastic Processes and their Applications, 49, (1994) 347 - 356.

S. Kanagawa, A representation of the rate functions in large deviation principles for U-statistics with degenerate kernels, Trends in Probability and Related Analysis (Proceedings of Symposium on Analysis and Probability 1998, Tiwan University, Taipei), 254-264, (1999).

R. J. Serfling, Approximation Theorems of Mathematical Statistics, John Wiley and Sons, New York, 1980.