数理解析研究所講究録
第538巻 1984年 96-120

無限に多くの撹乱母数を含む場合の推定理論
ーサンプルサイズが異なりうる場合の有効性

東大工学部　公文　雅之 (Masayuki Kumon)

1. はじめに

$x_1, x_2, \cdots, x_n$ を n 個の互いに独立な観測ベクトル、各 x_i は q_i 次元で確率密度関数 $p(x_i; \theta, \xi_i, q_i)$ を持つ分布に従うものとする。ここで各 x_i の分布に共通に含まれるスカラー母数 θ が、推定の対象となる未知母数で構造母数と呼ばれる。一方 ξ_i ; $i=1, \cdots, n$　は標本数 n に比例してその数が増大する撹乱母数で、これらは任意の値を取り得るものとする。我々は撹乱母数列 $\xi_1, \cdots, \xi_n$ の値が未知の情況の下で、構造母数 θ を推定する問題を考える。そして n が十分大きいときの θ の推定量の漸近的有効性について議論する。

この問題においては、独立、同一分布からの標本にもとづく正則な漸近理論において確立されている最尤推定量の良い性質が必ずしも保証されない。実際 Neyman & Scott [4] で指摘されたように、最尤推定量は必ずしも一致推定量とは

ならず、また一致性が成り立ってもその漸近分散は所謂Cramér-Raoの下限を達成しない。

本稿では2節でまず推定量のクラス$C_0 \supset C_1 \supset C_2$を設定し推定量の漸近分散及び有効性の定義を与える。クラスC_1は$x_1, \cdots, x_n$の置換に関して不変な一致推定量のクラスとして定義されるが、このクラスでは任意の撹乱母数列に対して有効な推定量は一般には存在しない。そこで推定量に対してさらに"information uniformity"なる要請を課し、この条件を満たす推定量のクラスをC_2とする。

3、4節でクラスC_2に属する推定量の漸近分散の表現、及びその新しい下限を与える。これはCramér-Raoの下限に微分幾何学的な考察にもとづくある非負の項が加わった形で表現される。新しい下限も必ずしも達成可能なものではないが、達成する推定量が存在する場合にはその具体形を与える。これらの主要結果はn回の観測における各サンプルサイズ、即ちx_iの次元q_iがi毎に異なりうる場合とq_iが全て等しい場合の有効推定量の相違をも明らかにする。

5、6節では主要結果の系を幾つか与え、これらをもとによく知られた例題での有効推定量を実際に構成する。そこでは最尤推定量の、有効性の観点からの評価もなされる。

2．推定関数および推定量のクラス

構造母数 θ の推定量 $\hat{\theta}=\hat{\theta}(x_1,\cdots,x_n)$ に対するクラスを幾つか設定する。$\bar{x}:=(x_1,\cdots,x_n)$, $\bar{\xi}:=(\xi_1,\cdots,\xi_n)$ とおき、$\tau_{i,j}$ を添字 i, j の置換とする。次のクラス C_0 はこの問題において広く用いられるものである。

クラス C_0: 推定量 $\hat{\theta}$ は、任意の i, j $(i, j=1,\cdots,n)$ に関して $\tau_{i,j}\Psi(\bar{x},\theta)=\Psi(\bar{x},\theta)$ を満たす $\bar{x}$ と θ の関数 $\Psi(\bar{x},\theta)$ に対する推定方程式 $\Psi(\bar{x},\hat{\theta})=0$ の解として得られるとき、クラス C_0 に属するという。このとき $\Psi(\bar{x},\theta)$ を $\hat{\theta}$ の推定関数と呼ぶ。条件 $\tau_{i,j}\Psi(\bar{x},\theta)=\Psi(\bar{x},\theta)$ は $\tau_{i,j}\hat{\theta}(\bar{x})=\hat{\theta}(\bar{x})$ と同値である。

推定関数 $\Psi(\bar{x},\theta)$ として特によく用いられるものに、$\sum_{i=1}^{n}\psi(x_i,\theta)$ なる n 個の関数 $\psi(x_i,\theta)$; $i=1,\cdots,n$ の和の形がある。例えば θ の最尤推定量は以下に示すように、この形の推定関数を持つ。

$x_1,\cdots,x_n$ の同時密度関数を $p(\bar{x};\theta,\bar{\xi})=\prod_{i=1}^{n}p(x_i;\theta,\xi_i)$ とおき、尤度関数 $p(\bar{x};\theta,\bar{\xi})$ の θ－score を $U(\bar{x};\theta,\bar{\xi})$ とする。即ち

$$U(\bar{x};\theta,\bar{\xi})=\partial_\theta \log p(\bar{x};\theta,\bar{\xi}) \ ; \ \partial_\theta=\partial/\partial\theta$$
$$=\sum_{i=1}^{n}u(x_i;\theta,\xi_i)=\sum_{i=1}^{n}\partial_\theta \log p(x_i;\theta,\xi_i)。$$

一方、ξ_i－score は

$$u(x_i;\theta,\xi_i)=\partial_{\xi_i}\log p(x_i;\theta,\xi_i)\ ;\ \partial_{\xi_i}=\partial/\partial\xi_i$$

で与えられるが、各θに対してξ_iに関する推定方程式 $u(x_i;\theta,\xi_i)=0$の解、即ちξ_iの最尤推定量$\hat{\xi}_i=\hat{\xi}_i(x_i,\theta)$が各$i$に対して一意的に存在すれば、$\theta$の最尤推定量は

$$\sum_{i=1}^{n}\hat{u}(x_i,\theta)=\sum_{i=1}^{n}u\{x_i;\theta,\hat{\xi}_i(x_i,\theta)\}=0$$

の解として得られる。従って最尤推定量はクラスC_0に属し、その推定関数は $y(\bar{x},\theta)=\sum_{i=1}^{n}\hat{u}(x_i,\theta)$ と、iに関する和の形で与えられる。

クラスC_0に属する推定量は必ずしも一致性を持たない。推定方程式 $y(\bar{x},\hat{\theta})=0$ を真の値θのまわりで一回展開し整理すると

$$\hat{\theta}-\theta\simeq-y(\bar{x},\theta)/\partial_\theta y(\bar{x},\theta)\ 。$$

これより適当な正則条件のもとで、$\hat{\theta}$の一致性は推定関数$y(\bar{x},\theta)$の zero-unbiasedness、即ち

$$\langle y(\bar{x},\theta)\rangle_\xi=0 \tag{1}$$

と同値であることがわかる。ここで$\langle\ \ \rangle_\xi$は$p(\bar{x};\theta,\xi)$に関する期待値を表わす記号。そこで一致推定量から成る、C_0のサブクラスC_1を以下のように定義する。

クラスC_1： C_0に属する推定量は、その推定関数が(1)を満たすとき、C_1に属するという。

C_1に属する推定量の漸近分散 $av(\hat{\theta},\xi)$、或いは

$\mathrm{av}(\Upsilon, \xi)$ を

$$\mathrm{av}(\Upsilon, \xi) = \langle \Upsilon^2(\tilde{x}, \theta) \rangle_\xi / \langle \partial_\theta \Upsilon(\tilde{x}, \theta) \rangle_\xi^2$$
$$= \langle \Upsilon^2 \rangle_\xi / \langle \Upsilon U(\xi) \rangle_\xi^2 \qquad (2)$$

で定義する。ここで $U(\xi)$ は $U(\tilde{x}; \theta, \xi)$ の略で、最後の等式は、恒等式 (1) を θ に関して微分することにより得られる。C_1 における推定量の有効性は、$\mathrm{av}(\Upsilon, \xi)$ を用いて次のように定式化される。

ある $\Upsilon^*(\tilde{x}, \theta) \in C_1$ が、任意の $\Upsilon(\tilde{x}, \theta) \in C_1$、および任意の撹乱母数列 ξ に対して

$$\mathrm{av}(\Upsilon^*, \xi) \le \mathrm{av}(\Upsilon, \xi) \qquad (3)$$

を満たすとき、$\Upsilon^*(\tilde{x}, \theta)$ をクラス C_1 における最良推定関数と呼ぶ。

このような条件を満たす $\Upsilon^*(\tilde{x}, \theta)$ が常に存在するとは必らない。実際にはクラス C_1 は広すぎて、任意の撹乱母数列に対して (3) 式を満たすような $\Upsilon^*(\tilde{x}, \theta)$ は存在しない方が一般的なのである。しかしながら C_1 における最良推定関数の存在を仮定した場合、その形について次のような結果が成り立つ。

定理1. C_1 における最良推定関数 $\Upsilon^*(\tilde{x}, \theta)$ は存在するとすれば、定数倍を除いて一意的に定まる。そして $\Upsilon^*(\tilde{x}, \theta)$ はある $y^*(x, \theta)$ を用いて

$$\Upsilon^*(\bar{x},\theta)=\sum_{i=1}^{n} y^*(x_i,\theta) \tag{4}$$

なる和の形で表わされる。

（略証） 前半の一意性に関しては、背理法を用いて簡単にいえる。以下、後半の(4)式に対する証明の手順のみを示す。まず次のような確率変数の空間R_ξ ; $R_{\xi_1},\cdots,R_{\xi_n}$を定義する。但し $\overline{}$ は閉包化記号。

$$R_\xi := \overline{\{r(\bar{x})\mid \langle r(\bar{x})\rangle_\xi=0,\ \langle r(\bar{x})^2\rangle_\xi<\infty\}},$$

$$R_{\xi_i} := \overline{\{r(x_i)\mid \langle r(x_i)\rangle_\xi=0,\ \langle r(x_i)^2\rangle_\xi<\infty\}},$$

$$i=1,\cdots,n\ .$$

$r_1, r_2\in R_\xi$に対して、内積$\langle r_1, r_2\rangle$を$\langle r_1, r_2\rangle=\langle r_1 r_2\rangle_\xi$で定義することにより$R_\xi$はHilbert空間となる。各$R_{\xi_i}$は$R_\xi$の閉部分空間であり、$r_i\in R_{\xi_i}$, $r_j\in R_{\xi_j}$ $(i\neq j)$に対して、R_{ξ_i}の定義およびx_i, x_jの独立性から

$\langle r_i, r_j\rangle=0$ が成り立つ。つまりR_{ξ_i}とR_{ξ_j}は互いに直交している。

いま$\Upsilon^*(\bar{x},\theta)\in C_1$を$C_1$における最良推定関数とすると、$\Upsilon^*\in R_\xi$である。$\Upsilon^*$の$x_i$に関する条件付期待値を

$$y_i^*:=E[\Upsilon^*\mid x_i]\quad ;\ i=1,\cdots,n \tag{5}$$

とおくと、$y_i^*\in R_{\xi_i}$であり(5)は$\Upsilon^*\in R_\xi$の各R_{ξ_i}上への正射影を与える。そこで$F_\xi:=R_{\xi_1}\oplus\cdots\oplus R_{\xi_n}$, R_ξにおけるF_ξの直交補空間をG_ξとして、$\Upsilon_F^*:=\sum_{i=1}^{n} y_i^*$,

$\Upsilon_G^* := \Upsilon^* - \Upsilon_F^*$ とおくと、直和分解 $R_\xi = F_\xi \oplus G_\xi$ に応じて

$$\Upsilon^* = \Upsilon_F^* + \Upsilon_G^* \quad ; \quad \Upsilon_F^* \in F_\xi,\ \Upsilon_G^* \in G_\xi \tag{6}$$

と分解される。

以上の準備のもとに証明は次の2段階で行なう。

(i) ある ξ_0 を固定し、$\boldsymbol{\xi}_0 := (\xi_0, \cdots, \xi_0)$ とおく。Υ^*は任意の撹乱母数列に対して最良であるとしたから、$\boldsymbol{\xi}_0$ に対しても (3) の意味で最良でなければならない。この条件から分解 (6) において $\Upsilon_G^* = 0$ でなければならないことが導びかれる。

(ii) (i)の結果、$\Upsilon^* = \Upsilon_F^* = \sum_{i=1}^{n} y_i^*$ となるが、(i) で用いた ξ_0 の任意性より各 y_i^* が ξ に依らず $y_i^* = y^*(x_i, \theta)$ と書けることがわかる。従って Υ^*は (4) 式のように表現される。

この定理は C_1 において最良推定関数を探す場合には、始めから (4) の形のものに限ってよいことを示している。しかし既に述べたように C_1 での最良推定関数は一般には存在しないから (C_1 に属する推定量および最良推定量の存在条件に関しては Amari & Kumon [1]を参照)、我々は推定量にさらにある条件を加えた C_1 のサブクラス C_2 を設定することにする。

まず各 R_{ξ_i} の2次元部分空間 T_{ξ_i} を

$$T_{\xi_i} := \{\, a u(x_i; \theta, \xi_i) + b v(x_i; \theta, \xi_i) \}$$

で定義する。T_{ξ_i} は $p(x_i;\theta,\xi_i)$ に対する θ－score $u(x_i;\theta,\xi_i)$ と ξ_i－score $v(x_i;\theta,\xi_i)$ で張られる２次元確率変数空間で、幾何学的には (θ,ξ_i) を母数とする統計モデルの多様体 $M_i:=\{p(x_i;\theta,\xi_i)\}$ の接空間にあたる。$i\neq j$ のとき T_{ξ_i} と T_{ξ_j} は互いに直交することは容易にわかる。そこで $T_\xi:=T_{\xi_1}\oplus\cdots\oplus T_{\xi_n}$ とおき、R_ξ における T_ξ の直交補空間を N_ξ とおけば $R_\xi=T_\xi\oplus N_\xi$。任意の $r(\bar{x})\in R_\xi$ はこの直和分解に応じて

$$r(\bar{x})=t(\bar{x})+s(\bar{x})\ ;\quad t(\bar{x})\in T_\xi,\quad s(\bar{x})\in N_\xi$$

と一意的に分解される。t を r の接成分，s を法成分と呼ぶ。

任意の $y(\bar{x},\theta)\in C_1$ は R_ξ に属するから

$$y=y_T+y_N\ ;\quad y_T\in T_\xi,\ y_N\in N_\xi \tag{7}$$

と書け、接成分 y_T は T_ξ の直和分解に応じてさらに

$$y_T=\sum_{i=1}^{n}(a_i u_i+b_i v_i) \tag{8}$$

と分解される。但し $u_i:=u(x_i;\theta,\xi_i)$，$v_i:=v(x_i;\theta,\xi_i)$ と略した。(8) において係数 a_i が ξ_i に依らず、θ と x_i の次元 q_i のみに依存する、即ち $a_i=a(\theta,q_i)$ なる係数を持つ推定関数を uniformly informative な推定関数と呼ぶことにする。

クラス C_2： C_1 に属する推定関数で、uniformly informative な推定関数をクラス C_2 に属するという。

3. クラス C_2 における漸近分散

C_1 に属する推定関数 $\Psi = \sum_{i=1}^{n}(a_i u_i + b_i v_i) + \Psi_N$ に対し、恒等式 $\langle \Psi \rangle = 0$ を ξ_i に関して微分することにより、$\langle \Psi, v_i \rangle = 0$ が得られる。これより係数 b_i は

$$b_i = -(\langle u_i, v_i \rangle / \langle v_i, v_i \rangle)\, a_i$$

と定まり、Ψ の接成分 $\Psi_T = \sum_{i=1}^{n}(a_i u_i + b_i v_i)$ は

$$\Psi_T = \sum_{i=1}^{n} a_i w_i \tag{9}$$

とまとめられる。ここで

$$w_i := u_i - (\langle u_i, v_i \rangle / \langle v_i, v_i \rangle)\, v_i \tag{10}$$

は T_{ξ_i} に属し、v_i と直交する。

一方 Ψ の法成分 Ψ_N は直和分解 $R_\xi = F_\xi \oplus G_\xi$ に応じて

$$\Psi_N = \sum_{i=1}^{n} n_i + \Psi_G \tag{11}$$

と分解される。ここで $\Psi_G \in G_\xi$ であり、$\sum_{i=1}^{n} n_i = \sum_{i=1}^{n} n(x_i; \theta, \xi_i)$ は F_ξ における T_ξ の直交補空間に属する。

そこで n_i / a_i を改めて n_i とおくと、(9)、(11) より $\Psi \in C_1$ は

$$\Psi = \Psi_F + \Psi_G\ ; \quad \Psi_F = \sum_{i=1}^{n} a_i (w_i + n_i) \tag{12}$$

と表わされる。$\Psi \in C_2$ ならば、(12) において $a_i = a(\theta, q_i)$ である。

我々は C_2 における最良推定関数を求めようとするが、ここでも C_1 での定理1と同じことが成り立つ。即ち C_2 にお

ける最良推定関数は存在するならば定数倍を除いて一意であり、(4)のようにF_ξの元だけで表わされる。そこでC_2に属する推定関数としては以下(12)のΨ_F部分、即ち

$$\Psi(\bar{x}, \theta) = \sum_{i=1}^{n} y(x_i, \theta) \tag{13}$$
$$= \sum_{i=1}^{n} a(\theta, q_i)\{w(x_i; \theta, \xi_i) + n(x_i; \theta, \xi_i)\}$$

なる形のものに限ることにする。ここで係数aがξ_iに依らないから

$$\Psi_0(\bar{x}, \theta) = \sum_{i=1}^{n} y_0(x_i, \theta) \tag{14}$$
$$= \sum_{i=1}^{n} \{w(x_i; \theta, \xi_i) + n(x_i; \theta, \xi_i)\}$$

もC_2に属する推定関数であることに注意する。ΨとΨ_0の与える推定量は一般には異なるが、$q_i = q$、即ちx_iの次元が全て等しい場合には、

$$\sum_{i=1}^{n} y(x_i, \theta) = a(\theta, q) \sum_{i=1}^{n} y_0(x_i, \theta)$$

からわかるように同じ推定量を与える。またこのことから、$q_i = q$の場合一般性を失うことなく、(13)の分解において任意の$\Psi \in C_2$に対して$a = 1$としてよい。

C_2に属する推定量の漸近分散を見やすい形で与えるため、幾つかの簡略化記号を定義する。まず$p(x_i; \theta, \xi_i)$の添字iを省いた$p(x; \theta, \xi)$で構成される2次元統計モデルを$M := \{p(x; \theta, \xi)\}$とおく。$p(x; \theta, \xi)$に対する$\theta$－score, ξ－scoreをそれぞれ単にu, vとおき、wも(10)で添字i

を省いたもので定義する。$p(x;\theta,\xi)$ に関する期待値を

$\langle\ \rangle$ で表わし、wの２乗期待値を

$$\bar{g}:=\bar{g}(\theta,\xi)=\langle w^2\rangle=\langle u^2\rangle-\langle uv\rangle^2/\langle v^2\rangle \quad (15)$$

とおく。MのFisher情報量行列を

$$\begin{bmatrix} g_{\theta\theta} & g_{\theta\xi} \\ g_{\xi\theta} & g_{\xi\xi} \end{bmatrix} = \begin{bmatrix} \langle u^2\rangle & \langle uv\rangle \\ \langle uv\rangle & \langle v^2\rangle \end{bmatrix}$$

とすれば、$\bar{g}=g_{\theta\theta}-(g_{\theta\xi}^2/g_{\xi\xi})$ とも書ける。$\bar{g}$はθに関する部分情報量と呼ばれる。明らに $\bar{g}\leq g_{\theta\theta}$ であり、等号が成り立つのは $g_{\theta\xi}=0$、即ちuとvが直交するときに限る。

次に $A(\theta,\xi,q)$ なる量に関する平均化量 $[A]$ を

$$[A]:=n^{-1}\sum_{i=1}^{n}A(\theta,\xi_i,q_i)$$

で定義する。(13)の形を持つ推定関数から得られる推定量の漸近分散を $av(y;\xi)$ と書くことにすれば、

$$av(y,\xi)=[\langle y^2\rangle]/[\langle\partial_\theta y\rangle]^2 \quad (16)$$

と表わされるが、(13) の $n(x;\theta,\xi)$ に対して

$$g^n:=\langle n(x;\theta,\xi)^2\rangle \quad (17)$$

と定義すると、(16) はさらに

$$av(y,\xi)=[a^2(\bar{g}+g^n)]/[a\bar{g}]^2 \quad (18)$$

となる。

(18) 式で $\bar{g}$は統計モデルMに固有な量、aおよびg^n が推定関数に依存する量である。さらに $q_i=q$ の場合には前に

も述べたように任意の$y \in C_2$に対して$a=1$とおけるから(18)は

$$av(y, \xi) = [\bar{g} + g^n] / [\bar{g}]^2 \qquad (19)$$

となって漸近分散は$g^n = \langle n^2 \rangle$のみを通して推定関数に依存する。$g^n \geq 0$であるから(19)より、$av(y, \xi) \geq 1/[\bar{g}]$。右辺は平均化部分情報量$[\bar{g}]$の逆数、即ち$\xi$が未知の場合のCramér-Raoの下限にあたる。次節において、最適な$a(\theta, q)$と$n(x;\theta,\xi)$を求めることにより、C_2に属する推定量の漸近分散の下限を与える。

4. 主要定理

3節での議論にもとづき、(13)での$y(x,\theta) = a(\theta, q)\{w(x;\theta,\xi) + n(x;\theta,\xi)\}$自身を$C_2$に属する推定関数と呼ぶことにする。いま$s = s(x;\theta,\xi,q)$を$C_2$に属する任意の推定関数$y = a(w+n)$の法成分$n$に対して、

$$\partial_\xi \langle s, n \rangle = 0 \quad および \quad \langle s \rangle = 0 \qquad (20)$$

を満たす確率変数とし、$\mathcal{S} := \{s(x;\theta,\xi,q)\}$をそのような$s$の全体とする。Mの接空間の元、即ち$s = au + bv$に対しては$\langle s, n \rangle = 0$が恒等的に成り立つから集合$\mathcal{S}$は空ではない。

確率変数の空間R_ξ，T_ξを2節で定義したR_{ξ_i}，T_{ξ_i}の

の添字 i を省いたものとすると、$R_\xi = T_\xi \oplus N_\xi$ と直和分解される。任意の $s \in S$ は仮定 (20) より R_ξ に属するが、

$$Ds := \dot{s} + s v \quad , \text{ただし } \dot{s} := \partial_\xi s \tag{21}$$

で定義される Ds も R_ξ に属する。(即ち $\langle Ds \rangle = 0$)
そこで上記の直和分解に従って $Ds = D^t s + D^n s$; $D^t s \in T_\xi$, $D^n s \in N_\xi$ と分解する。次に $s \in S$ の関数 $f(s)$ を

$$f(s) := \langle \dot{w}, s \rangle^2 / \langle (D^n s)^2 \rangle \tag{22}$$

と定義し、さらに

$$\bar{f} := \bar{f}(\theta, \xi, q) = \sup_{s \in S} f(s) \tag{23}$$

とおく。

以上の準備のもとに次の定理が得られる。

定理2. C_2 に属する任意の推定関数の漸近分散 $av(y, \xi)$ に対して $av(y, \xi) \geq av^*$, 但し

$$av^* = [\bar{g}^2 / (\bar{g} + \bar{f})]^{-1} \tag{24}$$

が成り立つ。

(証明) $y = a(w + n) \in C_2$, $s \in S$ に対して、Cauchy-Schwarz の不等式

$$\langle n^2 \rangle \langle (D^n s)^2 \rangle \geq \langle D^n s, n \rangle^2$$

を考える。右辺は (20) を用いて $\langle D^n s, n \rangle = \langle \dot{w}, s \rangle$ と書けるから $\langle n^2 \rangle \geq f(s)$。これより $g^n \geq \bar{f}$ となり、(18) の $av(y, \xi)$ は

$$av(y, \xi) \geq [a^2(\bar{g} + \bar{f})] / [a\bar{g}]^2$$

と評価される。右辺においてさらに $[a\bar{g}]^2 \leq [a^2(\bar{g}+\bar{f})][\bar{g}^2/(\bar{g}+\bar{f})]$ が成り立つから、結局 $av(y, \xi)$ は (24) の av^* で下から押えられる。

定理2は C_2 に属する推定量の漸近分散に関して一つの下限 av^* を与えるが、これが達成可能なものかどうか、即ち av^* に一致する漸近分散を持つ推定量が存在するかどうかが問題となる。次の定理はこの最良推定関数の存在に関する十分条件を与えるものである。（クラス C_2 に属する推定関数および最良推定関数の存在条件に関しては Amari & Kumon [1] を参照）

定理3．ある $s^* \in S$ を用いて作った

$$y_0^* = w + \{\langle \dot{w}, s^* \rangle / \langle (D^n s^*)^2 \rangle\} D^n s^* \qquad (25)$$

と

$$a^* = \bar{g} / \{\bar{g} + f(s^*)\} \qquad (26)$$

が共に ξ に依らないとする。このとき $y^* := a^* y_0^*$ は定理2の下限 av^* を達成する最良推定関数である。

（証明）(18) 式の右辺に $a = a^*$, $g^n = f(s^*)$ を代入すればよい。

次に各 x_i の次元 q_i が全て等しい場合の結果を系として与える。

系1．(25) の y_0^* が ξ に依らず、(26) の a^* は ξ に依るとする。このとき y_0^* は q_i の値が全て等しい場合の C_2 における最良推定関数である。そして推定量の漸近分散は $av(y, \xi) \geq av^{**}$、但し

$$av^{**} := [\bar{g} + \mathfrak{f}] / [\bar{g}]^2 \tag{27}$$

で押えられる。

（証明）$q_i = q$ の場合の漸近分散の表現 (19) 式に、$g^n \geq \mathfrak{f}$ を適用すればよい。y_0^* が av^{**} を達成することも容易。

定理3と系1は、最良推定関数 y^* 或いは y_0^* の法成分が (21) 式で定義した Ds なる確率変数の法成分 $D^n s$ で与えられることを示している。D は ξ－座標軸に沿う mixture－共変微分と呼ばれる微分幾何学的な演算である。共変微分は本来平行移動という幾何学的概念から導びかれるものであり、Amari & Kumon [1] では mixture－平行移動とそれに双対な exponential－平行移動を用いて、確率変数の空間を θ の推定に関しての情報を持つ部分空間に直和分解することにより、C_1，C_2 に属する推定量および最良推定量の存在条件

を調べている。

5．主要定理の系1

統計モデルMの接ベクトル、即ちθ－score uとξ－score vの線形結合で表わされる確率変数の全体をTとする。 節でも述べたようにTは集合Sの部分集合であり、(23)式の極大値$\bar{f}$をS上で求めるのは一般には難しいが、部分集合T上での極大値は簡単に求めることができる。この極大値を用いて表わされる漸近分散の下限も、もちろん必ずしも達成可能なものではないが、その利点は与えられた統計モデルから陽に計算可能なことにあり、後で示すように幾つかの典型的な例題で実際に達成可能なものとなる。

wとvのmixture-共変微分の法成分を成分とする2次元の確率変数を$h := (D^n w,\ D^n v)^T$とする。ここで^{T}は転置を示す。hの分散－共分散行列を

$$H := \begin{bmatrix} h_{\theta\theta} & h_{\theta\xi} \\ h_{\theta\xi} & h_{\xi\xi} \end{bmatrix} = \begin{bmatrix} \langle (D^n w)^2 \rangle & \langle D^n w,\ D^n v \rangle \\ \langle D^n w,\ D^n v \rangle & \langle (D^n v)^2 \rangle \end{bmatrix}$$

とおく。HはMのmixture-曲率行列と呼ばれる。

次に2次元列ベクトルcを

$$c := (c_\theta,\ c_\xi)^T = (\langle \dot{w},\ w \rangle,\ \langle \dot{w},\ v \rangle)^T$$

により定義する。これらを用いて、Hが正則の場合は　$k =$

$c^T H^{-1} c$、Hが非正則の場合は、$h_{\theta\theta} \neq 0$ ならば$k = c_\theta^2 h_{\theta\theta}^{-1}$、$h_{\xi\xi} \neq 0$ ならば$k = c_\xi^2 h_{\xi\xi}^{-1}$ とおく。

定理4．(22)の $f(s)$ の集合Ｔ上の最大値はkで与えられる。従って

$$av_T^* := [\bar{g}^2/(\bar{g}+k)]^{-1} \tag{28}$$

は、C_2 における漸近分散の下限を与える。

定理5．確率変数 y_0^* をHが正則な場合

$$y_0^* := w + c^T H^{-1} h \tag{29}$$

で、またHが非正則な場合は

$$y_0^* := \begin{cases} w + c_\theta h_{\theta\theta}^{-1} h & (h_{\theta\theta} \neq 0) \\ w + c_\xi h_{\xi\xi}^{-1} h & (h_{\xi\xi} \neq 0) \end{cases} \tag{30}$$

で定義する。

$y^* := \{\bar{g}/(\bar{g}+k)\} y_0^*$ とおいたとき、右辺がξに依らないならば、y^* はC_2 における最良推定関数である。そして(28)の下限 av_T^* はこの y^* によって達成される。

（定理4，5の証明）$s \in T$に対する $f(s)$ の最大値を求めるため $s = a_1 w + a_2 v$ とおく。ここでa_i はθとξに依存する係数。このとき$D^n s = a^T h$ となる。但し　$a = ($

$a_1, a_2)^T$。従って $f(S) = (a^T c)^2 / (a^T H a)$ となり、Hが正則な場合は、Cauchy-Schwartz の不等式より $(a^T c)^2 \leq (a^T H a)(c^T H^{-1} c)$。故に $f(S) \leq k$ となって定理4がいえる。またここで等号が成り立つのは $a = H^{-1}c$ のときで、これより定理5がいえる。Hが非正則のときも簡単にいえる。

系2. (29)、或いは(30)で定義した y_0^* が ξ に依らず、$\bar{g}/(\bar{g}+k)$ が ξ に依るとする。このとき y_0^* は $g_i = g$ の場合の C_2 における最良推定関数であり、漸近分散は $av_T^{**} := [\bar{g}+k]/[\bar{g}]^2$ で下から押えられる。

（証明）容易。

定理5と系2を用いて最良推定関数が得られる例を幾つか示す。

（例1） $x = (x_1, \cdots, x_q)$ とし、x_j $(j = 1, \cdots, q)$ は互いに独立に正規分布 $N(\xi, \theta)$ に従うとする。x の確率密度関数は

$$p(x; \theta, \xi) = (2\pi)^{-\frac{1}{2}q} \exp\left\{-\frac{1}{2\theta}\sum_{j=1}^{q}(x_j - \xi)^2\right\}.$$

score 関数はそれぞれ

$$w = u = \frac{1}{2}q\{z^2 + (x_{\cdot} - \xi)^2 - \theta\}/\theta^2,$$

$$v = q\,(x_{.} - \xi)\,/\,\theta$$

で与えられる。ここで

$$x_{.} = \left(\sum_{j=1}^{q} x_j\right)/\,q \quad , \quad z^2 = \left\{\sum_{j=1}^{q} (x_j - x_{.})^2\right\}/\,q\text{。}$$

定理4、5に関する諸量は、

$$D^n v = q\,\{(q-1)(x_{.}-\xi)^2 - z^2\}\,/\,\theta^2 \ , \quad C_\xi = -q\,/\,\xi^2,$$

$$C_\theta = 0,\ h_{\xi\xi} = 2q\,(q-1)\,/\,\theta^2,\ h_{\theta\xi} = 0,\ h_{\theta\theta} \neq 0,$$

$$\bar{g} = \tfrac{1}{2}q\,/\,\theta^2,\quad k = q\,/\,\{2(q-1)\,\theta^2\}\text{。}$$

これらより、

$$y_0^* = q\,\{\,q z^2 - (q-1)\,\theta\,\}\,/\,\{2\theta^2(q-1)\},$$

$$\bar{g}\,/\,(\bar{g} + k) = (q-1)\,/\,q\text{。}$$

故に

$$y^* = \{\,q z^2 - (q-1)\theta\,\}\,/\,(2\theta^2)$$

となって y^* は ξ によらないから C_2 における最良推定関数である。漸近分散の下限は $av_I^* = 2\theta^2/([q]-1)$ となる。

（例2）$x = (x_1,\ x_2)$, $x_1 = (x_{11}, \cdots, x_{1q})$, $x_2 = (x_{21}, \cdots, x_{2q})$ として、x_{1j} は $N(\xi,\ 1)$, x_{2j} は $N(\theta\xi,\ 1)$ に従うものとし、互いに独立とする。x の密度関数は

$$p(x;\theta,\xi) = (2\pi)^{-q} \exp\left[-\frac{1}{2}\sum_{j=1}^{q}\{(x_{1j}-\xi)^2 + (x_{2j}-\theta\xi)^2\}\right]$$

これより

$w = q\xi(x_{2.} - \theta x_{1.})/\rho$, $v = q(x_{1.} + \theta x_{2.} - \xi\rho)$,
但し $\rho = \theta^2 + 1 > 0$。従って定理4、5の諸量は

$D^n w = wv$, $c_\theta = q\xi/\rho$, $c_\xi = 0$, $h_{\theta\theta} = q^2\xi^2$,
$h_{\theta\xi} = 0$, $h_{\xi\xi} \neq 0$, $\bar{g} = q\xi^2/\rho$, $k = 1/\rho^2$。

故に

$$y_0^* = q(x_{2.} - \theta x_{1.})(x_{1.} + \theta x_{2.})/\rho^2,$$
$$\bar{g}/(\bar{g} + k) = q\xi^2\rho/(q\xi^2\rho + 1)$$

となって、y_0^* だけがξに依らないから y_0^* が $q_i = q$ の場合の C_2 における最良推定関数となる。この場合の漸近分散の下限は、$av_T^{**} = (1 + \rho q[\xi^2])/(q[\xi^2])^2$ 。

ここで例1では $y^* = \hat{u} - \langle \hat{u} \rangle$, 例2では $y_0^* = \hat{u}$ であることを注意しておく。但し $\hat{u}$ は最尤推定量の推定関数。つまりこれらの例では有効推定量は実質的には最尤推定量で与えられる。また例2は、q_i の値が異なる場合には C_2 での有効推定量が存在しない例である。

6. 主要定理の系2

前節では $f(s)$ の極大値を S の部分集合 T に限って求めたが、本節では T に属さない $s \in S$ を用いて4節で与えた下限を達成する最良推定関数が存在する場合について説明する。

wおよびvが次の形で与えられる統計モデルを考える。

$$w = c w^*, \quad v = d v^* + e, \tag{31}$$

ここでc, d, e は(θ, ξ)に依存した係数で、w^*, v^* はξに依らないxとθの関数とする。ここで

$$y_0^* = w^* / v^* \tag{32}$$

とおくと、これはξに依らない確率変数であるが、y_0^* が

$$\langle y_0^* \rangle = 0, \quad \langle y_0^* w \rangle = \bar{g} \tag{33}$$

なる2条件を満たすとする。次に

$$y^* = (\bar{g} / \langle y_0^{*2} \rangle)\, y_0^* \tag{34}$$

とおいて、y^* がξに依らないとする。このとき

定理6．(31)の形で表わされる統計モデルにおいて、C_2における最良推定関数が存在するならば、それは(34)のy^*で与えられる。

（証明）Kumon& Amari［2］参照。

系3．(32)のy_0^* が(33)を満たし、$\bar{g} / \langle y_0^{*2} \rangle$が$\xi$に依るとする。このとき$y_0^*$ は$q_i = q$の場合のC_2における最良推定関数である。

（証明）定理6の証明より明らか。

定理6，系3の最良推定関数y^*, y_0^* は定理3の形で与え

ることができる。実際y_0^*の直和分解を$y_0^* = w + n^*$として、$s^* = y_0^* / \langle n^{*2} \rangle$とおく。このとき任意の$y = a(w + n) \in C_2$に対して

$$\langle s^*, n \rangle = \langle n^*, n \rangle / \langle n^{*2} \rangle = 1 ,$$

つまり$s^* \in S$であることがいえる。そしてこのs^*を用いた(25)式の右辺が(32)で定義したy_0^*に一致することが計算される。ここで用いたs^*は明らかにTには属さない元である。

C_2における最良推定関数が定理6で与えられる例を示す。

（例3）$x = (x_1, \cdots, x_q)$とし、x_j $(j = 1, \cdots, q)$は互いに独立に$N(\theta, \xi)$に従うとする。xの密度関数は

$$p(x; \theta, \xi) = (2\pi)^{-\frac{1}{2}q} \exp\left\{-\frac{1}{2\xi}\sum_{j=1}^{q}(x_j - \theta)^2\right\}。$$

これより

$$w = u = q(x_. - \theta)/\xi , \quad v = \tfrac{1}{2}q\{z^2 + (x_. - \theta)^2 - \xi\}/\xi^2$$

となって

$$c = 1/\xi, \quad d = \tfrac{1}{2}q/\xi^2 , \quad e = -\tfrac{1}{2}q/\xi ,$$

$$w^* = q(x_. - \theta) , \quad v^* = z^2 + (x_. - \theta)^2$$

によって(31)の形で表わされる。そこで$y_0^* = w^*/v^*$とおくと$\langle y_0^* \rangle = 0$, $\langle y_0^* w \rangle = \bar{g} = q/\xi$, 即ち$y_0^*$が(33)式を満たすことが確かめられる。さらに$\langle y_0^{*2} \rangle = q^2/\{\xi(q-2)\}$より$\bar{g}/\langle y_0^{*2} \rangle = (q-2)/q$であるから、

$$y^* = (\bar{g}/\langle y_0^{*2}\rangle)\, y_0^* = (q-2)(x_{.}-\theta)/\{z^2+(x_{.}-\theta)^2\}$$

となってこれはξに依らないからC_2における最良推定関数である。漸近分散の下限は$av^* = [(q-2)/\xi]^{-1}$である。

ついでながら、この例題においてよく用いられる推定量として $\hat{\theta} = (\sum_{i=1}^{n} x_{i.}/c_i z_i^2)/(\sum_{i=1}^{n} 1/c_i z_i^2)$ なる形のものがある。ここで係数c_iは$\hat{\theta}$の漸近分散の値が最小になるように選ばれる。(ξ_i, $i=1,\cdots,n$の値が既知の場合、上式で$c_i z_i^2$の部分をξ_iで置きかえたものが所謂BLUEとなる。) 最適な重みc_i^*は$c_i^* = (q_i-5)^{-1}$であることがわかり、c_i^*を上式に代入したものを$\hat{\theta}^*$とおくと、$\hat{\theta}^*$の推定関数は$\tilde{y} = (x_{.}-\theta)(q-5)/z^2$である。$\tilde{y}$は$C_2$に属することがわかり、$\tilde{y} \neq y^*$であるから、これは有効ではない。

またこの例題での最尤推定量の推定関数はy_0^*に一致する。従って$q_i = q$の場合には最尤推定量は有効であるが、q_iが異なる場合には最尤推定量は有効ではない。ここで求めた最良推定関数y^*はNeyman & Scott[4]が、最尤推定量よりも有効な推定量として提案したものである。

7. おわりに

構造母数θの推定量$\hat{\theta}$の漸近分散の下限としては、情況に

応じて様々なものが考えられる。最も単純なものはθに関するFisher情報量$g_{\theta\theta}$を用いて$v_1=[g_{\theta\theta}]^{-1}$で与えられるもので、これは$\xi_i$; $i=1,\cdots,n$の値が全て等しく$\xi_i=\xi$、ξの値が既知の時には達成される。次は部分情報量$\bar{g}$を用いて$v_2=[\bar{g}]^{-1}$で与えられる。これは$\xi_i=\xi$ではあるがξの値が未知の時に達成されるものである。従って$v_2-v_1\geq 0$は、共通の値ξに関する知識の持つ情報量であるといえる。第3、第4の下限は、定理4および系2で与えた$v_3=[\bar{g}^2/(\bar{g}+k)]^{-1}$と$v_4=[\bar{g}+k]/[\bar{g}]^2$である。これらはもちろん$\xi_i$の値が任意で未知の場合の下限であるが、集合Sを接ベクトルの集合Tに制限して与えたものであった。この制限は未知のξ_iの持つ情報を統計モデルMのξー座標軸に沿う局所的な曲率、即ちMのmixture-曲率行列Hに反映させることに対応している。また差$v_4-v_3\geq 0$は、異なりうるg_i ; $i=1,\cdots,n$の値に応じた推定関数に対する最適な重み$a^*(\theta,g)=\bar{g}/(\bar{g}+k)$の持つ情報といえる。未知の$\xi_i$の持つ情報を段階的に$\xi$ー座標軸に沿った曲率量に反映させてゆくことにより、次々とより高い下限が得られ、最終的に定理2と系1で与えたav^*、av^{**}に致達する。ここで差$av^{**}-av^*\geq 0$は、$v_4-v_3\geq 0$と同じ意味を持つ。

我々は定理3でC_2における最良推定関数$y^*=a^*y_0^*$を与

えたが、この y^* はLindsay[3]の考えた推定関数に対する information unbiasednessなる条件、即ち $\langle y^2 \rangle + \langle \partial_\theta y \rangle = 0$ を満たしていることを指摘しておく。つまり C_2 における最良推定関数が存在するならば、それは information unbiased でなければならないのである。

参考文献

[1] Amari, S. & Kumon, M. (1983). Geometrical theory on estimation of structural parameter in the presence of infinitely many nuisance parameters. 京都大学数理解析研究所講究録。

[2] Kumon, M. & Amari, S. (1984). Estimation of a structural parameter in the presence of a large number of nuisance parameters. Biometrika, 71, 3, to appear.

[3] Lindsay, B. G. (1982). Conditional score functions: Some optimality results. Biometrika, 69, 503-512.

[4] Neyman, J. & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. Econometrica, 16, 1-32.