

無限に多くの攪乱母数を含む場合の推定理論 — 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, \dots, 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 $\xi_i; i = 1, \dots, n$ は標本数 n に比例してその数が増大する攪乱母数で、これらは任意の値を取り得るものとする。我々は攪乱母数列 ξ_1, \dots, ξ_n の値が未知の状況の下で、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、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, \dots, 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攪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を加わ、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。こ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 $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ に対するクラスを幾つか設定する。 $\bar{x} = (x_1, \dots, x_n)$, $\bar{\xi} = (\xi_1, \dots, \xi_n)$ 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 ($i, j = 1, \dots, n$) 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、 $\sum_{i=1}^n y(x_i, \theta)$ なる n 個の関数 $y(x_i, \theta)$; $i = 1, \dots, n$ 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, \dots, 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

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

一方、 ξ_i -scoreは

$v(x_i; \theta, \xi_i) = \partial_{\xi_i} \log p(x_i; \theta, \xi_i)$; $\partial_{\xi_i} = \partial / \partial \xi_i$
 で与えられるが、各 θ に対して ξ_i に関する推定方程式 $v(x_i; \theta, \xi_i) = 0$ 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、 θ 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貫性を持たない。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

$$\hat{\theta} - \theta \approx -Y(\bar{x}, \theta) / \partial_{\theta} Y(\bar{x}, \theta) .$$

これより適当な正則条件のもとで、 $\hat{\theta}$ の一貫性は推定関数 $Y(\bar{x}, \theta)$ の zero-unbiasedness、即ち

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

と同値であることがわかる。ここで $\langle \cdot \rangle_{\xi}$ 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

$av(Y, \xi)$ を

$$\begin{aligned} av(Y, \xi) &= \langle Y^2(\bar{x}, \theta) \rangle_{\xi} / \langle \partial_{\theta} Y(\bar{x}, \theta) \rangle_{\xi}^2 \\ &= \langle Y^2 \rangle_{\xi} / \langle Y U(\xi) \rangle_{\xi}^2 \end{aligned} \quad (2)$$

で定義する。ここで $U(\xi)$ は $U(\bar{x}; \theta, \xi)$ 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、 $av(Y, \xi)$ を用いて次のように定式化される。

ある $Y^*(\bar{x}, \theta) \in C_1$ が、任意の $Y(\bar{x}, \theta) \in C_1$ 、および任意の擾乱母数列 ξ に対して

$$av(Y^*, \xi) \leq av(Y, \xi) \quad (3)$$

を満たすとき、 $Y^*(\bar{x}, \theta)$ 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 $Y^*(\bar{x}, \theta)$ が常に存在するとは必ずしもない。実際にはクラス C_1 は広すぎて、任意の擾乱母数列に対して (3) 式を満たすような $Y^*(\bar{x}, \theta)$ 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 $Y^*(\bar{x}, \theta)$ 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 $Y^*(\bar{x}, \theta)$ はある $y^*(x, \theta)$ を用いて

$$Y^*(\bar{x}, \theta) = \sum_{i=1}^n y_i^*(x_i, \theta) \quad (4)$$

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 $R_{\xi}; R_{\xi_1}, \dots, R_{\xi_n}$ を定義する。但し $\overline{\quad}$ は閉包化記号。

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

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

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

$r_1, r_2 \in R_{\xi}$ 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_{ξ} は Hilbert 空間となる。各 R_{ξ_i} は R_{ξ} 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 $Y^*(\bar{x}, \theta) \in C_1$ を C_1 における最良推定関数とすると、 $Y^* \in R_{\xi}$ である。 Y^* の x_i に関する条件付期待値を

$$y_i^* := E[Y^* \mid x_i] \quad ; \quad i = 1, \dots, n \quad (5)$$

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

$Y_G^* := Y^* - Y_F^*$ とおくと、直和分解 $R_{\xi} = F_{\xi} \oplus G_{\xi}$ に応じて

$$Y^* = Y_F^* + Y_G^* \quad ; \quad Y_F^* \in F_{\xi}, \quad Y_G^* \in G_{\xi} \quad (6)$$

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、 $\xi_0 := (\xi_0, \dots, \xi_0)$ とおく。 Y^* は任意の攪乱母数列に対して最良であるとしたから、 ξ_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 $Y_G^* = 0$ でなければならないことが導ひかれる。

(ii) (i) の結果、 $Y^* = Y_F^* = \sum_{i=1}^n y_i^*$ となるが、(i) で用いた ξ_0 の任意性より各 y_i^* が ξ_i に依らず $y_i^* = y^*(x_i, \theta)$ と書けることかわかる。従って Y^* 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 $T_{\xi} := T_{\xi_1} \oplus \dots \oplus T_{\xi_n}$ 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 $r(x) \in R_{\xi}$ はこの直和分解に応じて

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

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 $Y(x, \theta) \in C_1$ は R_{ξ} に属するから

$$Y = Y_T + Y_N ; \quad Y_T \in T_{\xi}, \quad Y_N \in N_{\xi} \quad (7)$$

と書け、接成分 Y_T は T_{ξ} の直和分解に応じてさらに

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

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 $Y = \sum_{i=1}^n (a_i u_i + b_i v_i) + Y_N$ に対し、恒等式 $\langle Y \rangle = 0$ を ξ_i に関して微分することにより、 $\langle Y, v_i \rangle = 0$ 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、 Y の接成分 $Y_T = \sum_{i=1}^n (a_i u_i + b_i v_i)$ は

$$Y_T = \sum_{i=1}^n a_i w_i \quad (9)$$

とまとめられる。ここで

$$w_i = u_i - (\langle u_i, v_i \rangle / \langle v_i, v_i \rangle) v_i \quad (10)$$

は T_{ξ_i} に属し、 v_i と直交する。

一方 Y の法成分 Y_N は直和分解 $R_{\xi} = F_{\xi} \oplus G_{\xi}$ に応じて

$$Y_N = \sum_{i=1}^n n_i + Y_G \quad (11)$$

と分解される。ここで $Y_G \in G_{\xi}$ 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 $Y \in C_1$ は

$$Y = Y_F + Y_G ; \quad Y_F = \sum_{i=1}^n a_i (w_i + n_i) \quad (12)$$

と表わされる。 $Y \in C_2$ 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 Y_F 部分、即ち

$$\begin{aligned} Y(\bar{x}, \theta) &= \sum_{i=1}^n y(x_i, \theta) \\ &= \sum_{i=1}^n \alpha(\theta, q_i) \{w(x_i; \theta, \xi_i) + n(x_i; \theta, \xi_i)\} \end{aligned} \quad (13)$$

なる形のものに限ることにする。ここで係数 α が ξ_i に依らないから

$$\begin{aligned} Y_0(\bar{x}, \theta) &= \sum_{i=1}^n y_0(x_i, \theta) \\ &= \sum_{i=1}^n \{w(x_i; \theta, \xi_i) + n(x_i; \theta, \xi_i)\} \end{aligned} \quad (14)$$

も C_2 に属する推定関数であることに注意する。 Y と Y_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、

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

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 $Y \in C_2$ に対して $\alpha = 1$ 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 θ -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、 $\bar{g} = g_{\theta\theta} - (g_{\theta\xi}^2 / g_{\xi\xi})$ と書ける。 \bar{g} は θ に関する部分情報量と呼ばれる。明らかに $\bar{g} \leq g_{\theta\theta}$ 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 $[A]$ と

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

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

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

となつて漸近分散は $g^n = \langle n^2 \rangle$ 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、 $av(y, \xi) \geq 1 / [\bar{q}]$ 。右辺は平均化部分情報量 $[\bar{q}]$ の逆数、即ち ξ 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、

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

を満たす確率変数とし、 $S = \{s(x; \theta, \xi, q)\}$ 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 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。 $R_\xi = T_\xi \oplus N_\xi$ と直和分解される。任意の $s \in S$ は仮定 (20) より R_ξ に属するから、

$$Ds := \dot{s} + s\nu, \quad \text{ただし } \dot{s} := \partial_\xi s \quad (21)$$

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

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

と定義し、さらに

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

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

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

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

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

を考へる。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
 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

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

と

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

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 α^* 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

$$av^{**} := [\bar{g} + \bar{f}] / [\bar{g}]^2 \quad (27)$$

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、 $q^n \geq \bar{f}$ 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 D_s 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かれるものであり、

Amari & Kumon [1] では mixture-平行移動とそれに対応した 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

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

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

$$y_0^* := w + C^T H^{-1} h \quad (29)$$

で、また H が非正則な場合は

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

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、
 $\bar{q} / (\bar{q} + k)$ が ξ 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り、漸近分散は
 $av_T^{**} = [\bar{q} + k] / [\bar{q}]^2$ 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) $x = (x_1, \dots, x_q)$ とし、 x_j ($j = 1, \dots,$
 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

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

$$v = q(x - \xi) / \theta$$

で与えられる。ここで

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

定理4.5に関する諸量は、

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

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

$$\bar{g} = \frac{1}{2} q / \theta^2, \quad k = q / \{ 2(q-1) \theta^2 \}.$$

これらより、

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

$$\bar{g} / (\bar{g} + k) = (q-1) / q.$$

故に

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

となり、 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 $av_T^* = 2 \theta^2 / ([q] - 1)$

となる。

(例2) $x = (x_1, x_2)$, $x_1 = (x_{11}, \dots, x_{1q})$, $x_2 = (x_{21}, \dots, x_{2q})$ 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

$$w = q\xi(x_2 - \theta x_1) / \rho, \quad v = q(x_1 + \theta x_2 - \xi\rho),$$

但し $\rho = \theta^2 + 1 > 0$ 。従って定理4.5の諸量は

$$D^n w = wv, \quad c_\theta = q\xi / \rho, \quad c_\xi = 0, \quad h_{\theta\theta} = q^2 \xi^2,$$

$$h_{\theta\xi} = 0, \quad h_{\xi\xi} \neq 0, \quad \bar{q} = q\xi^2 / \rho, \quad k = 1 / \rho^2.$$

故に

$$y_0^* = q(x_2 - \theta x_1)(x_1 + \theta x_2) / \rho^2,$$

$$\bar{q} / (\bar{q} + k) = q\xi^2 \rho / (q\xi^2 \rho + 1)$$

となつて、 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 に限つて求めたが、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。

$$w = cw^*, \quad v = dv^* + e, \quad (31)$$

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

$$y_0^* = w^* / v^* \quad (32)$$

とおくと、これは ξ に依らない確率変数であるが、 y_0^* が

$$\langle y_0^* \rangle = 0, \quad \langle y_0^* w \rangle = \bar{q} \quad (33)$$

なる 2 条件を満たすとする。次に

$$y^* = (\bar{q} / \langle y_0^{*2} \rangle) y_0^* \quad (34)$$

において、 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、 $\bar{q} / \langle y_0^{*2} \rangle$ が ξ 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

ることかできる。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 $s^* \in S$ であることかいえる。そしてこの s^* を用いた
(25) 式の右辺か (32) で定義した y_0^* に一致することか計算さ
れる。ここで用いた s^* は明らかに \mathbb{T} には属さない元である。

C_2 における最良推定関数か定理 6 で与えられる例を示す。

(例 3) $x = (x_1, \dots, x_q)$ とし、 x_j ($j = 1, \dots, q$) 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

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

となつて

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

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

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

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

となつてこれは 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。($\xi_i, i=1, \dots, n$ 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。従つて $q_i = q$ 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 ξ_i ; $i = 1, \dots, n$ 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 ξ_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 -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 q_i ; $i = 1, \dots, n$ の値に応じた推定関数に対する最適な重み $a^*(\theta, q) = \bar{g} / (\bar{g} + k)$ の持つ情報といえる。未知の ξ_i の持つ情報を段階的に ξ -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^{**} に到達する。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

えたか、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 でなければならぬのである。

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.