

直線当てはめ問題の推定法について

東大・工 川鍋 一晃 (Motoaki kawanabe)

東大・工 甘利 俊一 (Shun-ichi Amari)

要旨

線形関係にある2次元量を測定して得られた誤差を含む2次元データからこの直線を推定する方法について議論する。この問題は観測ごとに真値の位置が異なり、観測とともに攪乱パラメータが増加するので、セミパラメトリックモデルとして扱う必要がある。情報幾何学の方法によりこのモデルの推定関数について調べた結果、最尤推定量はデータの真値の分布の仕方によらず一貫性を持つが、情報量の意味では有効でないことがわかる。特に、原点を通る直線を当てはめる場合にはデータの真値の分布の平均と分散の情報を用いることにより最尤推定法を改良する推定関数法が構成できる。

1 はじめに

線形関係にある多次元量を測定して得られた誤差を含む多次元データからこの線形関係を推定する問題は、変数誤差モデル (errors-in-variables model)、線形関係モデル、あるいは因子分析モデルという名前で数多くの文献の中で扱われている (Anderson[1984])。本研究で扱うのはその中でも最も簡単な2次元データに直線を当てはめる問題であり、以下のように表される n 個の独立な2次元データ

$$\begin{aligned} x_\alpha &= \xi_\alpha + \varepsilon_{1\alpha}, \\ y_\alpha &= \beta\xi_\alpha + \gamma + \varepsilon_{2\alpha}, \end{aligned} \quad \alpha = 1, \dots, n \quad (1.1)$$

が観測される。すなわち、この2次元データの真値は直線 $y = \beta x + \gamma$ 上についていて、この直線の傾き β 、 y 切片 γ が推定したいパラメータであり、 ξ_α は α 番目のデータの真値の位置を表している。この問題では観測とともに未知の値 ξ_α が増加していくため、 β, γ の推定も通常のパラメトリックモデルより難しくなる。実際、Neyman and Scott[1948]によってこのような状況のもとでは最尤推定量が一貫性を持たない、あるいは有効な推定量でない場合もあるということが指摘されている。

直線当てはめ問題においては、真値の位置を表す未知の値の系列 (ξ_1, \dots, ξ_n) の扱い方として以下の2通りが考えられる (Kendall and Stuart[1979])。

1. 非確率的な数列
2. 未知の分布 $k(\xi)$ からの i.i.d. 確率変数列

1 は線形関数関係 (linear functional relationship) あるいは層別モデル (stratified model) と呼ばれ、 ξ_α が観測とともに増加する攪乱パラメータとなる。2 は線形構造関係 (linear structural relationship) ある

いはミクスチャーモデル (mixture model) と呼ばれ, $k(\xi)$ が関数自由度を持つ攪乱パラメータである。いずれの場合も無限自由度の攪乱パラメータを含むため通常のパラメトリックモデルの推定理論の枠内で扱うことはできない。

測定誤差 $(\varepsilon_{1\alpha}, \varepsilon_{2\alpha})$ は平均 $\mathbf{0}$ の i.i.d. 確率ベクトルであり, 2 の場合は ξ_α とは独立であると仮定する。もし, 真値の分布 $k(\xi)$ や測定誤差 $(\varepsilon_{1\alpha}, \varepsilon_{2\alpha})$ の分布に何の仮定も設けなければパラメータ β, γ は識別可能ではない。そこで, 推定したいパラメータが識別可能になるようにこれらの分布に様々な仮定を設けたモデルが提案され, 研究されている (Reiersøl[1950], Bickel et al.[1987] etc.). 本研究では ε_1 と ε_2 は独立に正規分布 $N(0, \sigma^2)$ にしたがうと仮定する (制約正規誤差モデル)。さらに, ξ_α を i.i.d. 確率変数列とみる線形構造関係モデル (ミクスチャーモデル) を扱う。

以上で述べたモデルにおいて, 真値の分布 $k(\xi)$ のいかにかわらず β, γ の \sqrt{n} -一致推定量 (広域的ロバスト推定量) を与えるセミパラメトリック推定法について議論する。広域的ロバスト推定量を議論するための有用な道具として Godambe[1960, 1976] によって導入された推定関数という概念がある。

定義 1.1 推定関数

推定したいパラメータが d 次元ベクトル θ であるセミパラメトリックモデル

$$S = \{ p(x, \theta, k) \mid \theta \in \Theta \subset \mathbf{R}^d, k \in K \}$$

において, x と θ の d 次元ベクトル値関数

$$\mathbf{y}(x, \theta) = (y_1(x, \theta), \dots, y_d(x, \theta))^T$$

で以下の条件を満たすものを推定関数と呼ぶ。

$$E_{\theta, k}[\mathbf{y}(x, \theta)] = \mathbf{0}, \quad \forall \theta, \forall k \quad (1.2)$$

$$\det(E_{\theta, k}[\partial_\theta \mathbf{y}(x, \theta)]) \neq 0, \quad \forall \theta, \forall k \quad (1.3)$$

$$E_{\theta, k}[\|\mathbf{y}(x, \theta)\|^2] < \infty, \quad \forall \theta, \forall k \quad (1.4)$$

ただし, $E_{\theta, k}$ は分布 $p(x, \theta, k)$ に関する期待値, $\partial_\theta = \partial/\partial\theta$, $\|\cdot\|$ はユークリッドノルムとする。

注 (1.2) は広域的ロバスト性を保証し, (1.3), (1.4) は推定量の妥当性を保証する条件である。

もし推定関数が存在すれば, i.i.d. 観測値 x_1, \dots, x_n が得られたとき, 推定関数 $\mathbf{y}(x, \theta)$ からつくられる推定方程式

$$\sum_{i=1}^n \mathbf{y}(x_i, \hat{\theta}) = \mathbf{0} \quad (1.5)$$

の解 $\hat{\theta}$ は広域的ロバスト推定量になる。この推定量の漸近的性質は以下のようにまとめられる。

定理 1.1 推定関数 $\mathbf{y}(x, \theta)$ から得られる推定量 $\hat{\theta}$ は真の攪乱パラメータ k_0 のいかにかわらず \sqrt{n} -一致性を持ち, 漸近的に正規分布にしたがう。漸近分散行列は

$$\begin{aligned} \text{Av}[\mathbf{y}] &\equiv \lim_{n \rightarrow \infty} n E_{\boldsymbol{\theta}, k_0} [(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \\ &= A^{-1} V (A^T)^{-1} \end{aligned} \quad (1.6)$$

$$(A = E_{\boldsymbol{\theta}, k_0} [\partial \boldsymbol{\theta} \mathbf{y}], \quad V = E_{\boldsymbol{\theta}, k_0} [\mathbf{y} \mathbf{y}^T])$$

となる。

このように推定関数が存在すれば広域的ロバスト推定量が構成できるが、具体的なセミパラメトリックモデルに対して推定関数を具体的に求めることは容易ではない。それどころか、これまでは推定関数の存在条件や最適な推定関数の特定というような基本的な問題に関してもほとんどわかっていなかった。Amari and Kumon[1988]は観測とともに攪乱パラメータが増加するモデル(Neyman-Scott問題)において情報幾何学(Amari[1985])の方法を用いて推定関数が解析できることを示し、推定関数に関する諸問題に解答を与えた。この幾何学理論はAmari and Kawanabe[1994]によって一般のセミパラメトリックモデルに拡張されている。本研究ではこの幾何学理論を用いて直線当てはめ問題の推定関数とセミパラメトリック推定法について考察する。

本研究で扱っている2次元直線当てはめ問題をさらに一般化して、多次元の線形当てはめ問題を考えることができる。この問題は多次元変数誤差モデル、線形関係モデル、あるいは因子分析モデルという名前で呼ばれ、様々な研究がなされている(Anderson[1984], Gleser[1981] etc.)。p次元確率ベクトル $\mathbf{x}_{1\alpha}$ とr次元確率ベクトル $\mathbf{x}_{2\alpha}$ からなるn個の独立な(p+r)次元データ $\mathbf{x}_\alpha = (\mathbf{x}_{1\alpha}, \mathbf{x}_{2\alpha})$ を観測する。

$$\mathbf{x}_\alpha = \begin{pmatrix} \mathbf{x}_{1\alpha} \\ \mathbf{x}_{2\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{u}_{1\alpha} \\ \mathbf{u}_{2\alpha} \end{pmatrix} + \begin{pmatrix} \mathbf{e}_{1\alpha} \\ \mathbf{e}_{2\alpha} \end{pmatrix} = \mathbf{u}_\alpha + \mathbf{e}_\alpha \quad (1.7)$$

$$\mathbf{u}_{2\alpha} = \boldsymbol{\gamma} + B \mathbf{u}_{1\alpha}, \quad \alpha = 1, \dots, n \quad (1.8)$$

観測データの真値 \mathbf{u}_α は $\mathbf{R}^{(p+r)}$ の中のp次元部分空間に制限されていて、その位置を表すp次元ベクトル $\mathbf{u}_{1\alpha}$ が観測毎に異なっている。 \mathbf{e}_α は測定誤差であり、平均0、共分散行列 Σ のi.i.d.確率ベクトルである。観測データの真値がのっているp次元部分空間を表すパラメータ $B, \boldsymbol{\gamma}$ が推定したいパラメータである。

線形当てはめ問題は工学応用という側面からも重要な問題である。近年、コンピュータビジョンにおける2次元画像からの3次元環境の認識・理解の過程が線形当てはめ問題、あるいはさらに一般の非線形当てはめ問題であることが知られるようになり、数理工学的な研究が行なわれている(金谷[1994, 1995])。このコンピュータビジョンにおける研究が本研究を始める動機となった。

2 原点を通る直線当てはめ問題の推定法

まず最初に原点を通る直線(y切片 $\gamma = 0$)を当てはめる問題を考える。ある決められた1点を通る直線を当てはめる問題は一般性を失うことなくこの問題に帰着できる。観測されるのはn個の独立な2次

元データ

$$\begin{aligned} x_\alpha &= \xi_\alpha + \varepsilon_{1\alpha}, \\ y_\alpha &= \beta\xi_\alpha + \varepsilon_{2\alpha}, \end{aligned} \quad \alpha = 1, \dots, n \quad (2.1)$$

であり、推定したいパラメータは直線の傾き β である。真値の位置 ξ_α は未知の分布 $k(\xi)$ からの i.i.d. 確率変数とし、推定量の漸近的性質を調べるため $k(\xi)$ は有限の平均と分散を持つと仮定する。

$$\begin{aligned} \mu_\xi &= E_k[\xi] \\ \sigma_\xi^2 &= \text{Var}_k[\xi] \end{aligned}$$

測定誤差 $\varepsilon_{1\alpha}, \varepsilon_{2\alpha}$ は独立に正規分布 $N(0, \sigma^2)$ にしたがう、かつ ξ_α と独立であると仮定する。誤差分散 σ^2 は既知の場合と未知の場合の両方について考える。

ξ_α を i.i.d. 確率変数列とみる線形構造関係モデルにおいては (x, y) の密度関数は、

$$p(x, y, \beta, \sigma^2, k) = \int p(x, y, \beta, \sigma^2, \xi) k(\xi) d\xi \quad (2.2)$$

となる。ただし、

$$p(x, y, \beta, \sigma^2, \xi) = \frac{1}{2\pi\sigma^2} \exp \left\{ -\frac{1}{2\sigma^2}(x - \xi)^2 - \frac{1}{2\sigma^2}(y - \beta\xi)^2 \right\} \quad (2.3)$$

は真値 ξ を固定したときの (x, y) の (条件付) 密度関数である。このセミパラメトリックモデルを $S = \{p(x, y, \beta, \sigma^2, k)\}$ とおく。計算を簡単にするために、 β を固定したときの直線に沿う方向の座標を s 、直交する方向の座標を a とおく。

$$s(x, y, \beta) = x + \beta y \quad (2.4)$$

$$a(x, y, \beta) = y - \beta x \quad (2.5)$$

ξ を固定したときの条件付分布を考えると、 $s(x, y, \beta)$ は (β を固定したとき) ξ の十分統計量になり、正規分布 $N((1 + \beta^2)\xi, (1 + \beta^2)\sigma^2)$ にしたがう。一方、 $a(x, y, \beta)$ は補助統計量になり、正規分布 $N(0, (1 + \beta^2)\sigma^2)$ にしたがう。さらに、 s と a は独立である。

最初に、幾何学理論 (Amari and Kawanabe[1994]) を適用してこのモデルの推定関数について考察した結果を簡単に要約しておく (川鍋 [1995])。

1. $g(s, \beta)$ を適当な正則条件を満たす s の任意の関数とすると、

$$g(s, \beta) a = g(x + \beta y, \beta) (y - \beta x)$$

は推定関数である。

⇒ 最尤推定法の関数は $(x + \beta y)(y - \beta x)$ であるから最尤推定法は推定関数法である (後述)。

2. ξ の分布が $k(\xi)$ のときの最適な推定関数は

$$E_{\beta, \sigma^2, k}[\xi | s] a = E_{\beta, \sigma^2, k}[\xi | x + \beta y] (y - \beta x) \quad (2.6)$$

である (k に依存). この最適推定関数から得られる推定量はセミパラメトリックモデル S の k における情報量限界を達成する.

直線当てはめ問題 (変数誤差モデル) において最も基本的な推定法は最尤推定法である. この推定法は真値 ξ_α を非確率的なパラメータとみなす線形関数関係モデルにおいて, 尤度関数

$$\prod_{\alpha=1}^n p(x_\alpha, y_\alpha, \beta, \sigma^2, \xi_\alpha) = \frac{1}{(2\pi\sigma^2)^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{\alpha=1}^n (x_\alpha - \xi_\alpha)^2 - \frac{1}{2\sigma^2} \sum_{\alpha=1}^n (y_\alpha - \beta\xi_\alpha)^2 \right\} \quad (2.7)$$

を最大にする $\beta, \xi_\alpha, (\sigma^2)$ を推定量とする方法である. しかし, β の最尤推定量 $\hat{\beta}_{ML}$ は真値 ξ_α を確率変数とみなす線形構造関係モデルに対しても用いることができる. 尤度方程式は

$$\sum_{\alpha=1}^n \frac{\partial}{\partial \beta} \log p(x_\alpha, y_\alpha, \beta, \xi_\alpha) = \frac{1}{\sigma^2} \sum_{\alpha=1}^n \xi_\alpha (y_\alpha - \beta\xi_\alpha) = 0 \quad (2.8)$$

$$\frac{\partial}{\partial \xi_\alpha} \log p(x_\alpha, y_\alpha, \beta, \xi_\alpha) = \frac{1}{\sigma^2} \{x_\alpha + \beta y_\alpha - (1 + \beta^2)\xi_\alpha\} = 0 \quad (2.9)$$

となり, β, ξ_α の最尤推定量は σ^2 によらずに決めることができる. すなわち, 誤差分散 σ^2 が既知か未知かにかかわらず, 最尤推定法は同じ手続きになる. (2.9) を ξ_α について解けば β を固定したときの推定量が得られる.

$$\hat{\xi}_\alpha(\beta) = \frac{x_\alpha + \beta y_\alpha}{1 + \beta^2} \quad (2.10)$$

これを (2.8) に代入して整理すると β の推定方程式 (β に関する 2 次方程式) が得られる.

$$\sum_{\alpha=1}^n (x_\alpha + \beta y_\alpha)(y_\alpha - \beta x_\alpha) = 0 \quad (2.11)$$

$$\Leftrightarrow \overline{xy} \beta^2 - (\overline{y^2} - \overline{x^2}) \beta - \overline{xy} = 0 \quad (2.12)$$

ただし, 標本の 2 次モーメントを

$$\overline{x^2} = \frac{1}{n} \sum_{\alpha=1}^n x_\alpha^2$$

$$\overline{y^2} = \frac{1}{n} \sum_{\alpha=1}^n y_\alpha^2$$

$$\overline{xy} = \frac{1}{n} \sum_{\alpha=1}^n x_\alpha y_\alpha$$

とおいた. β の 2 次の係数 $\overline{xy} \neq 0$ ならば, 2 次方程式の解の公式より最尤推定量は

$$\hat{\beta}_{ML} = \frac{\overline{y^2} - \overline{x^2} + \sqrt{(\overline{y^2} - \overline{x^2})^2 + 4(\overline{xy})^2}}{2\overline{xy}} \quad (2.13)$$

となる. $\overline{xy} = 0$ のときには, $\overline{x^2} > \overline{y^2}$ ならば $\hat{\beta}_{ML} = 0$, $\overline{x^2} < \overline{y^2}$ ならば $\hat{\beta}_{ML} = \infty$ である.

直線当てはめ問題は観測とともに攪乱パラメータが増加するモデル (Neyman-Scott 問題) である. しかし, このモデルの推定関数について考察すると, 幸運なことに最尤推定法は推定関数法であり, 最尤推定量は広域的ロバスト推定量になることが示される (川鍋 [1995]).

- 最尤推定法は推定関数法である。すなわち、適当な正則条件のもとで最尤推定量 $\widehat{\beta}_{ML}$ は ξ の分布 $k(\xi)$ によらずに \sqrt{n} - 一致性を持ち、漸的に正規分布にしたがう。漸近分散は

$$\text{Av}[\widehat{\beta}_{ML}] = \frac{(1 + \beta^2)\sigma^2\{\sigma^2 + (1 + \beta^2)(\mu_\xi^2 + \sigma_\xi^2)\}}{(1 + \beta^2)(\mu_\xi^2 + \sigma_\xi^2)^2} \quad (2.14)$$

となる (ただし, $E_k[\xi] = \mu_\xi$, $\text{Var}_k[\xi] = \sigma_\xi^2$)。

本稿のように観測誤差 $(\varepsilon_{\alpha 1}, \varepsilon_{\alpha 2})$ が正規分布にしたがうと仮定すると、最尤推定法は実は 2 次元空間の最小自乗法になる。

- 最尤推定法は観測データ (x_α, y_α) , $\alpha = 1, \dots, n$ と直線 $y = \beta x$ の距離の自乗和

$$\sum_{\alpha=1}^n \frac{(y_\alpha - \beta x_\alpha)^2}{1 + \beta^2} \quad (2.15)$$

を最小にする β を推定量とする最小自乗法に一致する。

以上で述べたように最尤推定法は推定関数法であるが、推定関数法の中で最適な方法ではない。また、最適な推定関数は理論的には求めることができるが、 ξ の分布 $k(\xi)$ に依存するので実際には使うことができない。そこで、本稿では ξ の平均 μ_ξ と分散 σ_ξ^2 の情報のみを利用して最尤推定法を改良する推定関数法を提案する。

まず、最尤推定法の推定関数 $(x + \beta y)(y - \beta x)$ を含む推定関数の集合

$$\{(x + \beta y + c)(y - \beta x); c \in \mathbf{R}\}$$

を考える。この集合に属する推定関数を L - 推定関数と呼ぶことにする。 $c = 0$ のとき最尤推定法の推定関数になるから、 L - 推定関数の中で漸近分散を最小にするものを用いた推定関数法は漸近分散 (あるいは情報量) という規準でみて最尤推定法を改良する。 L - 推定関数の推定方程式は

$$\sum_{\alpha=1}^n (x_\alpha + \beta y_\alpha + c)(y_\alpha - \beta x_\alpha) = 0$$

となる ($c = 0$ のとき最尤推定法)。この推定方程式から得られる推定量 $\widehat{\beta}_c$ の漸近分散は

$$\text{Av}[\widehat{\beta}_c] = \frac{(1 + \beta^2)\sigma^2 \left[\{c + (1 + \beta^2)\mu_\xi\}^2 + (1 + \beta^2)^2\sigma_\xi^2 + (1 + \beta^2)\sigma^2 \right]}{\{c\mu_\xi + (1 + \beta^2)(\mu_\xi^2 + \sigma_\xi^2)\}^2}$$

であり、

$$c^* = \frac{\sigma^2 \mu_\xi}{\sigma_\xi^2} \quad (2.16)$$

のとき最小値をとることがわかる (図 1)。 L - 推定関数の中で漸近分散を最小にする推定関数

$$(x + \beta y + c^*)(y - \beta x) = \left(x + \beta y + \frac{\sigma^2 \mu_\xi^2}{\sigma_\xi^2} \right) (y - \beta x) \quad (2.17)$$

を最適 L - 推定関数 (L^* - 推定関数) と呼ぶことにする。 L^* - 推定関数には ξ の分布 $k(\xi)$ の平均 μ_ξ と分散 σ_ξ^2 という未知の量に依存する定数が含まれるので、推定法を構成するときには μ_ξ と σ_ξ^2 を推定量で置き換えたものを用いる。このようにして構成される推定関数法を最適 L - 推定関数法 (L^* - 推定関数法) と呼ぶ。まず、誤差分散 σ^2 が既知の場合には以下のような推定手順で $\widehat{\beta}_{L^*}$ を求める。

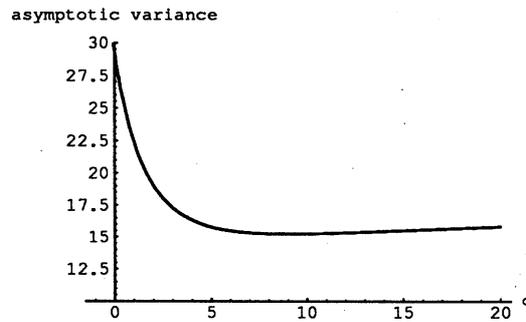
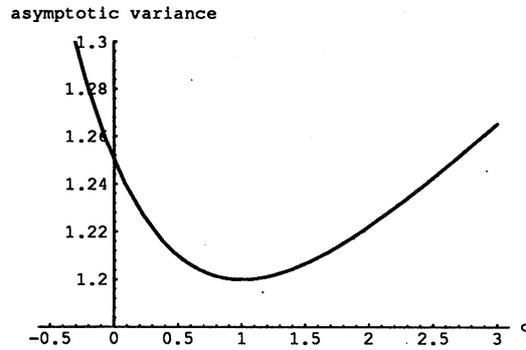


図 1: L -推定関数法の漸近分散

上: $\beta = 1, \sigma = 1, \mu_\xi = 1, \sigma_\xi = 1, c^* = 1$

下: $\beta = 1, \sigma = 3, \mu_\xi = 1, \sigma_\xi = 1, c^* = 9$

1. 未知パラメータ μ_ξ, σ_ξ^2 の \sqrt{n} -一致推定量を用いて, 係数 c^* の推定量を計算する.

$$\begin{cases} \hat{\mu}_\xi = \frac{1}{n} \sum_{\alpha=1}^n x_\alpha \\ \hat{\sigma}_\xi^2 = \frac{1}{n} \sum_{\alpha=1}^n x_\alpha^2 - (\hat{\mu}_\xi)^2 - \sigma^2 \end{cases} \quad (2.18)$$

$$\hat{c}^* = \frac{\sigma^2 \hat{\mu}_\xi}{\hat{\sigma}_\xi^2} \quad (2.19)$$

2. 推定方程式 (2 次方程式) を解いて推定量 $\hat{\beta}_L$ を求める.

$$\sum_{\alpha=1}^n \left(x_\alpha + \beta y_\alpha + \frac{\sigma^2 \hat{\mu}_\xi}{\hat{\sigma}_\xi^2} \right) (y_\alpha - \beta x_\alpha) = 0 \quad (2.20)$$

$$\Leftrightarrow \bar{xy} \beta^2 - (\bar{y^2} - \bar{x^2} - \hat{c}^* \bar{x}) \beta - \bar{xy} - \hat{c}^* \bar{y} = 0 \quad (2.21)$$

ただし, $\bar{x} = \frac{1}{n} \sum_{\alpha=1}^n x_{\alpha}$, $\bar{y} = \frac{1}{n} \sum_{\alpha=1}^n y_{\alpha}$ は標本平均, $\overline{x^2}, \overline{y^2}, \overline{xy}$ は標本 2 次モーメントである.

$$\hat{\beta}_{L^*} = \frac{\overline{y^2} - \overline{x^2} - \hat{c}^* \bar{x} + \sqrt{(\overline{y^2} - \overline{x^2})^2 + 4\overline{xy}(\overline{xy} + \hat{c}^* \bar{y})}}{2\overline{xy}} \quad (2.22)$$

誤差分散 σ^2 が未知の場合には推定量 $\hat{\sigma}^2$ を求めなければならないが, これを計算するためには β の適当な初期推定量 (一致推定量) が必要になる.

1. 最尤推定量 $\hat{\beta}_{ML}$ を求める.
2. 誤差分散 σ^2 の一致推定量 $\hat{\sigma}^2$ を求める.

$$\hat{\sigma}^2 = \frac{1}{\{1 + (\hat{\beta}_{ML})^2\}n} \sum_{\alpha=1}^n (y_{\alpha} - \hat{\beta}_{ML}x_{\alpha})^2 \quad (2.23)$$

3. 未知パラメータ $\mu_{\xi}, \sigma_{\xi}^2$ の \sqrt{n} -一致推定量を用いて, 係数 c^* の推定量を計算する.

$$\begin{cases} \hat{\mu}_{\xi} = \frac{1}{n} \sum_{\alpha=1}^n x_{\alpha} \\ \hat{\sigma}_{\xi}^2 = \frac{1}{n} \sum_{\alpha=1}^n x_{\alpha}^2 - (\hat{\mu}_{\xi})^2 - \hat{\sigma}^2 \end{cases} \quad (2.24)$$

$$\hat{c}^* = \frac{\hat{\sigma}^2 \hat{\mu}_{\xi}}{\hat{\sigma}_{\xi}^2} \quad (2.25)$$

4. 推定方程式 (2 次方程式) を解いて推定量 $\hat{\beta}_{L^*}$ を求める.

$\mu_{\xi}, \sigma_{\xi}^2$ の一致推定量 $\hat{\mu}_{\xi}, \hat{\sigma}_{\xi}^2$ および σ^2 の一致推定量 $\hat{\sigma}^2$ を用いれば, 推定量 $\hat{\beta}_{L^*}$ の漸近的な性質は真値 $\mu_{\xi}, \sigma_{\xi}^2, \sigma^2$ を用いた L^* -推定関数から得られる推定量 $\hat{\beta}_{L^*}$ ($c^* = \sigma^2 \mu_{\xi} / \sigma_{\xi}^2$) と同じになる.

- 適当な正則条件のもとで推定量 $\hat{\beta}_{L^*}$ は ξ の分布 $k(\xi)$ によらずに \sqrt{n} -一致性を持ち, 漸的に正規分布にしたがう. 漸近分散は

$$\text{Av}[\hat{\beta}_{L^*}] = \frac{(1 + \beta^2)\sigma^2\{\sigma^2 + (1 + \beta^2)\sigma_{\xi}^2\}}{(1 + \beta^2)(\mu_{\xi}^2 + \sigma_{\xi}^2)\sigma_{\xi}^2 + \sigma^2\mu_{\xi}^2} \quad (2.26)$$

となる.

以上では情報量の意味で最尤推定法を改良するセミパラメトリック推定法として L^* -推定関数法を構成した. 次に L^* -推定関数法の持つ性質について議論する. L^* -推定関数法は正規分布モデルと密接な関係があり, またある意味でミニマックス推定法であるという特徴を持っている.

まず正規分布モデルとの関係について述べる. ξ の分布 $k(\xi)$ のときの最適な推定関数は

$$E_{\beta, \sigma^2, k}[\xi | x + \beta y] (y - \beta x)$$

と表される (k に依存する). $k(\xi)$ が正規分布 $N(\mu_{\xi}, \sigma_{\xi}^2)$ のときの最適な推定関数を計算すると L^* -推定関数に一致する.

補題 2.1 L^* -推定関数は ξ の分布が正規分布 $N(\mu_\xi, \sigma_\xi^2)$ の場合の最適推定関数 (有効スコア関数) である。

証明 ξ の分布が正規分布 $N(\mu_\xi, \sigma_\xi^2)$ のとき, $s = x + \beta y$ を固定したときの ξ の条件付分布は正規分布 $N(\nu(s), \tau^2)$ となる。

$$\nu(s) \equiv \frac{\sigma_\xi^2 s + \sigma^2 \mu_\xi}{\sigma^2 + (1 + \beta^2) \sigma_\xi^2} \quad (2.27)$$

$$\tau^2 \equiv \frac{\sigma^2 \sigma_\xi^2}{\sigma^2 + (1 + \beta^2) \sigma_\xi^2} \quad (2.28)$$

定数倍の自由度は無視できるから, 正規分布 $N(\mu_\xi, \sigma_\xi^2)$ における最適推定関数は

$$E[\xi|s] a = \nu(s) a \simeq \left(s + \frac{\sigma^2 \mu_\xi}{\sigma_\xi^2} \right) a \quad (2.29)$$

となり, L^* -推定関数に一致する。 \square

直線当てはめ問題の場合, k における最適推定関数はセミパラメトリックモデル S の k における情報量限界を達成するから補題 2.1 より以下の定理が示される。

定理 2.1 ξ が正規分布にしたがっているとき, L^* -推定関数法によって求められる推定量 $\hat{\beta}_L$ はセミパラメトリックモデル $S = \{p(x, y, \beta, \sigma^2, k)\}$ の情報量限界を達成する。

注 正規分布 $N(\mu_\xi, \sigma_\xi^2)$ を仮定した場合, β を固定したときの ξ_α の推定量は

$$\tilde{\xi}_\alpha(\beta) = \nu(s) = \frac{(1 + \beta^2) \sigma_\xi^2}{\sigma^2 + (1 + \beta^2) \sigma_\xi^2} \frac{s_\alpha}{1 + \beta^2} + \frac{\sigma^2}{\sigma^2 + (1 + \beta^2) \sigma_\xi^2} \mu_\xi \quad (2.30)$$

であるとみなすことができる ($s_\alpha = x_\alpha + \beta y_\alpha$)。これは, 最尤推定量 $\hat{\xi}_\alpha(\beta) = s_\alpha / (1 + \beta^2)$ を ξ の平均 μ_ξ の方向へ一定の割合だけ縮小したものとみることができる。しかし, L^* -推定関数から得られる β の推定量に対しては直観的な意味はよくわからない。

注 L -推定関数のクラス $\{(x + \beta y + c)(y - \beta x); c \in \mathbf{R}\}$ は ξ の分布 $k(\xi)$ に正規分布モデル

$K_0 = \{N(\mu_\xi, \sigma_\xi^2) | \mu_\xi, \sigma_\xi^2\}$ を仮定したときに最適な推定関数がつくるクラスである。 ξ の分布 $k(\xi)$ に異なるパラメトリックモデルを仮定すれば異なる推定関数のクラスをつくることができる。

ξ の分布に μ_ξ, σ_ξ^2 をパラメータとする正規分布モデル $K_0 = \{N(\mu_\xi, \sigma_\xi^2) | \mu_\xi, \sigma_\xi^2\}$ を仮定する。このとき, (x, y) の分布は2次元正規分布モデルになり, 正規分布の平均と分散はパラメータ $\beta, \sigma^2, \mu_\xi, \sigma_\xi^2$ によって

$$\begin{bmatrix} \mu_\xi \\ \beta \mu_\xi \end{bmatrix}, \quad \begin{bmatrix} \sigma_\xi^2 + \sigma^2 & \beta \sigma_\xi^2 \\ \beta \sigma_\xi^2 & \beta^2 \sigma_\xi^2 + \sigma^2 \end{bmatrix}.$$

と表される。このようにしてできる S のパラメトリック部分モデルを $S_0 = \{p(x, y, \beta, \mu_\xi, \sigma_\xi^2)\}$ とおく。 S と S_0 の情報量限界に対する考察から, ξ が正規分布にしたがっているときには L^* -推定関数法は正規分布モデル S_0 の最尤推定法と漸近的に等価であることがわかる。

補題 2.2 正規分布モデル S_0 上では, セミパラメトリックモデル $S = \{p(x, y, \beta, \sigma^2, k)\}$ における β の情報量限界はパラメトリックモデル S_0 における β の情報量限界と一致する.

定理 2.2 ξ が正規分布にしたがっているとき, L^* -推定関数から得られる推定量 $\hat{\beta}_{L^*}$ はパラメトリックモデル S_0 の情報量限界を達成する (正規分布モデル S_0 の最尤推定法と漸近的に同等).

次にミニマックス性について述べる. ξ の分布のセミパラメトリックモデルを以下のように位置-尺度モデル (location-scale model) の形に書きなおす.

$$K = \left\{ \frac{1}{\sigma_\xi} f \left(\frac{\xi - \mu_\xi}{\sigma_\xi} \right) \mid \mu_\xi, \sigma_\xi, f \in \mathcal{F} \right\} \quad (2.31)$$

ただし, \mathcal{F} は平均 0, 分散 1 の確率密度関数の集合とする. これに対応して (x, y) のセミパラメトリックモデルも

$$S = \{p(x, y, \beta, \sigma^2, \eta, f) \mid \beta, \sigma^2, \eta = (\mu_\xi, \sigma_\xi), f \in \mathcal{F}\} \quad (2.32)$$

のように書き換えられる.

推定関数 $g(x, y, \beta)$ の推定方程式

$$\sum_{\alpha=1}^n g(x_\alpha, y_\alpha, \hat{\beta}_g) = 0 \quad (2.33)$$

から得られる推定量を $\hat{\beta}_g$ と置く. 推定量に対するリスク関数として漸近分散

$$Av_{\beta, \sigma^2, \eta, f} [\hat{\beta}_g] = \frac{E_{\beta, \sigma^2, \eta, f} [g(x, y, \beta)]^2}{\{E_{\beta, \sigma^2, \eta, f} [\partial_\beta g(x, y, \beta)]\}^2} \quad (2.34)$$

を用いることにする. 推定量の良さを評価する際に ξ の分布の分布形 f に関しては \mathcal{F} 全体で最大値をとることにし, この最大値を最小ににする推定量をミニマックス推定量と呼ぶことにする.

定義 2.1 任意の $\beta, \sigma^2, \eta = (\mu_\xi, \sigma_\xi)$ に対して漸近分散の最大値

$$\max_{f \in \mathcal{F}} Av_{\beta, \sigma^2, \eta, f} [\hat{\beta}_g]$$

を最小にする推定量 $\hat{\beta}_g$ をミニマックス推定量と呼ぶ.

定理 2.3 L^* -推定関数から得られる推定量 $\hat{\beta}_{L^*}$ はミニマックス推定量である. 最も不利な分布形は標準正規分布の密度関数 ϕ になる.

証明 $\hat{\beta}_{L^*}$ の漸近分散が f によらないこと, 正規分布 ϕ においては L^* -推定関数法が最適推定関数法であることをあわせると,

$$\max_{f \in \mathcal{F}} Av_{\beta, \sigma^2, \eta, f} [\hat{\beta}_{L^*}] = Av_{\beta, \sigma^2, \eta, \phi} [\hat{\beta}_{L^*}] \leq Av_{\beta, \sigma^2, \eta, \phi} [\hat{\beta}_g] \leq \max_{f \in \mathcal{F}} Av_{\beta, \sigma^2, \eta, f} [\hat{\beta}_g]$$

が導かれる. □

最後に ξ の分布 $k(\xi)$ としていくつかの代表的な分布をとって計算機実験を行なった結果について述べる. 誤差分散 σ^2 が未知の場合を考える. 各々の分布について 1000 個のデータを生成し, 最尤推定法お

よび L^* -推定関数法 (σ^2 が未知の場合) を用いて推定量を計算する. これを 1000 回行なって 1000 個の推定量 $\hat{\beta}_i, i = 1, \dots, 1000$ を求め, この 1000 個の推定量の標本 2 乗誤差を漸近分散の推定値とする.

$$\widehat{Av} = \sum_{i=1}^{1000} (\hat{\beta}_i - \beta)^2$$

表中の比の数値は $(\hat{\beta}_{L^*}$ の漸近分散) / $(\hat{\beta}_{ML}$ の漸近分散) である.

1. 正規分布

(i) $\sigma = 1$

(μ_ξ, σ_ξ)		最尤推定法	L^* -推定関数法	比
(1, 1)	バイアス	2.12×10^{-3}	1.72×10^{-3}	
	分散	1.28×10^{-3}	1.23×10^{-3}	
	平均 2 乗誤差	1.29×10^{-3}	1.23×10^{-3}	
	漸近分散 (推定値)	1.29	1.23	0.961
	漸近分散 (理論値)	1.25	1.2	0.96
(1, 1/3)	バイアス	1.17×10^{-3}	-4.34×10^{-5}	
	分散	2.59×10^{-3}	1.95×10^{-3}	
	平均 2 乗誤差	2.59×10^{-3}	1.95×10^{-3}	
	漸近分散 (推定値)	2.59	1.95	0.751
	漸近分散 (理論値)	2.61	1.96	0.751
(1, 3)	バイアス	-3.13×10^{-4}	-3.13×10^{-4}	
	分散	2.10×10^{-4}	2.10×10^{-4}	
	平均 2 乗誤差	2.10×10^{-4}	2.10×10^{-4}	
	漸近分散 (推定値)	0.210	0.210	1.00
	漸近分散 (理論値)	0.21	0.210	1.00
(1/3, 1)	バイアス	2.81×10^{-3}	1.96×10^{-3}	
	分散	2.70×10^{-3}	2.64×10^{-3}	
	平均 2 乗誤差	2.71×10^{-3}	2.64×10^{-3}	
	漸近分散 (推定値)	2.71	2.64	0.978
	漸近分散 (理論値)	2.61	2.57	0.985
(3, 1)	バイアス	9.19×10^{-4}	8.07×10^{-4}	
	分散	2.02×10^{-4}	2.00×10^{-4}	
	平均 2 乗誤差	2.03×10^{-4}	2.00×10^{-4}	
	漸近分散 (推定値)	0.203	0.200	0.988
	漸近分散 (理論値)	0.21	0.207	0.985

(ii) $\sigma = 3$

(μ_ξ, σ_ξ)		最尤推定法	L^* -推定関数法	比
(1,1)	バイアス	1.03×10^{-2}	-3.50×10^{-3}	
	分散	3.31×10^{-2}	1.54×10^{-2}	
	平均2乗誤差	3.32×10^{-2}	1.54×10^{-2}	
	漸近分散(推定値)	33.2	15.4	0.463
	漸近分散(理論値)	29.3	15.2	0.521
(1,1/3)	バイアス	5.39×10^{-2}	-3.75×10^{-3}	
	分散	0.118	1.87×10^{-2}	
	平均2乗誤差	0.121	1.87×10^{-2}	
	漸近分散(推定値)	121	18.7	0.155
	漸近分散(理論値)	81.8	18.0	0.219
(1,3)	バイアス	-1.05×10^{-3}	-1.79×10^{-3}	
	分散	2.54×10^{-3}	2.50×10^{-3}	
	平均2乗誤差	2.54×10^{-3}	2.51×10^{-3}	
	漸近分散(推定値)	2.54	2.51	0.986
	漸近分散(理論値)	2.61	2.57	0.985
(1/3,1)	バイアス	6.44×10^{-2}	7.01×10^{-3}	
	分散	0.124	7.11×10^{-2}	
	平均2乗誤差	0.128	7.11×10^{-2}	
	漸近分散(推定値)	128	71.1	0.556
	漸近分散(理論値)	81.8	61.4	0.751
(3,1)	バイアス	2.26×10^{-3}	3.29×10^{-4}	
	分散	2.54×10^{-3}	1.94×10^{-3}	
	平均2乗誤差	2.55×10^{-3}	1.94×10^{-3}	
	漸近分散(推定値)	2.55	1.94	0.761
	漸近分散(理論値)	2.61	1.96	0.751

2. 一様分布 ; $(\mu_{\xi}, \sigma_{\xi}) = (1, 1)$

σ		最尤推定法	L^{\sim} -推定関数法	比
1	バイアス	1.58×10^{-3}	8.68×10^{-4}	
	分散	1.25×10^{-3}	1.20×10^{-3}	
	平均 2 乗誤差	1.25×10^{-3}	1.20×10^{-3}	
	漸近分散 (推定値)	1.25	1.20	0.963
	漸近分散 (理論値)	1.25	1.20	0.96
3	バイアス	1.97×10^{-2}	-1.12×10^{-3}	
	分散	3.33×10^{-2}	1.69×10^{-2}	
	平均 2 乗誤差	3.37×10^{-2}	1.69×10^{-2}	
	漸近分散 (推定値)	33.7	16.9	0.501
	漸近分散 (理論値)	29.3	15.2	0.521

3. ロジスティック分布 ; $(\mu_{\xi}, \sigma_{\xi}) = (1, 1)$

σ		最尤推定法	L^{\sim} -推定関数法	比
1	バイアス	2.47×10^{-4}	7.85×10^{-5}	
	分散	1.27×10^{-3}	1.20×10^{-3}	
	平均 2 乗誤差	1.27×10^{-3}	1.20×10^{-3}	
	漸近分散 (推定値)	1.27	1.20	0.948
	漸近分散 (理論値)	1.25	1.20	0.96
3	バイアス	1.34×10^{-4}	-1.22×10^{-3}	
	分散	3.09×10^{-2}	1.71×10^{-2}	
	平均 2 乗誤差	3.11×10^{-2}	1.71×10^{-2}	
	漸近分散 (推定値)	31.1	17.1	0.549
	漸近分散 (理論値)	29.3	15.2	0.521

4.t 分布 (自由度 4); $(\mu_\xi, \sigma_\xi) = (1, 1)$

σ		最尤推定法	L^* -推定関数法	比
1	バイアス	2.28×10^{-3}	1.81×10^{-3}	
	分散	1.15×10^{-3}	1.09×10^{-3}	
	平均 2 乗誤差	1.15×10^{-3}	1.09×10^{-3}	
	漸近分散 (推定値)	1.15	1.09	0.949
	漸近分散 (理論値)	1.25	1.20	0.96
3	バイアス	2.00×10^{-3}	-4.02×10^{-3}	
	分散	3.17×10^{-2}	1.58×10^{-2}	
	平均 2 乗誤差	3.21×10^{-2}	1.58×10^{-2}	
	漸近分散 (推定値)	32.1	15.8	0.491
	漸近分散 (理論値)	29.3	15.2	0.521

5. 両側指数分布; $(\mu_\xi, \sigma_\xi) = (1, 1)$

指数分布の尺度パラメータが負の側と正の側で $1 : r$ になっている連続な両側指数分布で実験を行なう。平均 0, 分散 1 に規格化したときの密度関数は

$$f(\xi) = \begin{cases} \frac{\sqrt{r^2+1}}{r+1} \exp\left(-\frac{\sqrt{r^2+1}\xi+r-1}{r}\right) & \xi \geq -\frac{r-1}{\sqrt{r^2+1}} \\ \frac{\sqrt{r^2+1}}{r+1} \exp\left(-\sqrt{r^2+1}\xi-r+1\right) & \xi < -\frac{r-1}{\sqrt{r^2+1}} \end{cases}$$

となる。 $r=1$ のときは対称な分布であり, $r=3, r=1/3$ のときは非対称な分布になる。

r	σ		最尤推定法	L^* -推定関数法	比
1	1	バイアス	9.91×10^{-4}	7.17×10^{-4}	
		分散	1.24×10^{-3}	1.19×10^{-3}	
		平均2乗誤差	1.24×10^{-3}	1.19×10^{-3}	
		漸近分散(推定値)	1.24	1.19	0.962
		漸近分散(理論値)	1.25	1.20	0.96
	3	バイアス	7.24×10^{-3}	-2.69×10^{-2}	
		分散	3.04×10^{-2}	1.65×10^{-2}	
		平均2乗誤差	3.05×10^{-2}	1.65×10^{-2}	
		漸近分散(推定値)	30.5	16.5	0.543
		漸近分散(理論値)	29.3	15.2	0.521
1/3	1	バイアス	-2.08×10^{-3}	-2.54×10^{-3}	
		分散	1.26×10^{-3}	1.20×10^{-3}	
		平均2乗誤差	1.26×10^{-3}	1.21×10^{-3}	
		漸近分散(推定値)	1.26	1.21	0.958
		漸近分散(理論値)	1.25	1.20	0.96
	3	バイアス	7.04×10^{-3}	-9.50×10^{-3}	
		分散	3.15×10^{-2}	1.48×10^{-2}	
		平均2乗誤差	3.16×10^{-2}	1.49×10^{-2}	
		漸近分散(推定値)	31.6	14.9	0.472
		漸近分散(理論値)	29.3	15.2	0.521
3	1	バイアス	3.64×10^{-4}	5.15×10^{-6}	
		分散	1.31×10^{-3}	1.26×10^{-3}	
		平均2乗誤差	1.31×10^{-3}	1.26×10^{-3}	
		漸近分散(推定値)	1.31	1.26	0.958
		漸近分散(理論値)	1.25	1.20	0.96
	3	バイアス	1.17×10^{-2}	-5.04×10^{-3}	
		分散	3.26×10^{-2}	1.55×10^{-2}	
		平均2乗誤差	3.28×10^{-2}	1.56×10^{-2}	
		漸近分散(推定値)	32.8	15.6	0.475
		漸近分散(理論値)	29.3	15.2	0.521

計算機実験の結論は以下のようにまとめることができる。

- L^* -推定関数法は誤差分散 σ^2 が大きい場合には最尤推定法を改良する。
- L^* -推定関数法を改良する推定方法は理論的には考えられるが、その改良度はわずかである。

注 漸近的にはバイアスは分散に比べて無視できることから漸近分散の推定量として平均2乗誤差を用いたが、 L^* -推定関数法が最尤推定法を大きく改良している所では最尤推定量のバイアスも分散に比べて無視できない大きさになっている。

注 2次元直線当てはめ問題の最尤推定法は経済学においては同時方程式モデルの制限情報最尤推定法として知られている。この推定量の厳密な分布が計算されており、特にモーメントを持たないことが示されている (Mariano and Sawa[1972], Anderson[1976])。

3 一般の直線当てはめ問題の推定法

ここでは一般の直線当てはめ問題について考える。観測されるのは n 個の独立な2次元データ

$$\begin{aligned} x_\alpha &= \xi_\alpha + \varepsilon_{1\alpha}, \\ y_\alpha &= \beta\xi_\alpha + \gamma + \varepsilon_{2\alpha}, \end{aligned} \quad \alpha = 1, \dots, n \quad (3.1)$$

であり、推定したいパラメータは直線の傾き β と y 切片 γ である。真値の位置 ξ_α は未知の分布 $k(\xi)$ からの i.i.d. 確率変数とし、推定量の漸近的性質を調べるため $k(\xi)$ は有限の平均と分散を持つと仮定する。

$$\begin{aligned} \mu_\xi &= E_k[\xi] \\ \sigma_\xi^2 &= \text{Var}_k[\xi] \end{aligned}$$

測定誤差 $\varepsilon_{1\alpha}, \varepsilon_{2\alpha}$ は独立に正規分布 $N(0, \sigma^2)$ にしたがう、かつ ξ_α と独立であると仮定する。誤差分散 σ^2 は既知の場合と未知の場合の両方について考える。

ξ_α を i.i.d. 確率変数列とみるミクスチャーモデルにおいては (x, y) の密度関数は、

$$p(x, y, \beta, \gamma, \sigma^2, k) = \int p(x, y, \beta, \gamma, \xi) k(\xi) d\xi \quad (3.2)$$

となる。ただし、

$$p(x, y, \beta, \gamma, \sigma^2, \xi) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}(x-\xi)^2 - \frac{1}{2\sigma^2}(y-\beta\xi-\gamma)^2\right\} \quad (3.3)$$

は真値 ξ を固定したときの (x, y) の (条件付) 密度関数である。このセミパラメトリックモデルを $S = \{p(x, y, \beta, \gamma, \sigma^2, k)\}$ とおく。一般の直線当てはめ問題の推定関数とセミパラメトリック推定法は原点を通る直線当てはめ問題と同様に議論することができる。

一般の直線当てはめ問題の場合の最尤推定法は、 ξ_α を非確率的なパラメータとみなす線形関係モデルの尤度関数を最大にする $\beta, \gamma, \xi_\alpha, (\sigma^2)$ を推定量とする方法であり、尤度方程式は

$$\sum_{\alpha=1}^n \xi_\alpha (y_\alpha - \gamma - \beta\xi_\alpha) = 0 \quad (3.4)$$

$$\sum_{\alpha=1}^n (y_{\alpha} - \gamma - \beta \xi_{\alpha}) = 0 \quad (3.5)$$

$$x_{\alpha} - \xi_{\alpha} + \beta(y_{\alpha} - \gamma - \beta \xi_{\alpha}) = 0, \quad \alpha = 1, \dots, n \quad (3.6)$$

である。(3.6)より, β, γ を固定したときの ξ_{α} の推定量

$$\hat{\xi}_{\alpha}(\beta, \gamma) = \frac{x_{\alpha} + \beta(y_{\alpha} - \gamma)}{1 + \beta^2} \quad (3.7)$$

が得られ, これを(3.4), (3.5)に代入して整理すると

$$(\bar{x}y - \gamma\bar{x})\beta^2 - (\bar{y}^2 - 2\gamma\bar{y} + \gamma^2)\beta - \bar{x}y + \gamma\bar{x} = 0 \quad (3.8)$$

$$\gamma = \bar{y} - \beta\bar{x} \quad (3.9)$$

となる。ただし, \bar{x}, \bar{y} は標本平均, $\bar{x}^2, \bar{x}\bar{y}, \bar{y}^2$ は標本2次モーメントである。さらに, (3.9)を(3.8)に代入すれば β のみの推定方程式(2次方程式)が得られる。

$$S_{xy}\beta^2 - (S_{yy} - S_{xx})\beta - S_{xy} = 0 \quad (3.10)$$

ただし, S_{xx}, S_{xy}, S_{yy} は標本共分散である。 β の2次の係数 $S_{xy} \neq 0$ ならば, 2次方程式の解の公式より最尤推定量は

$$\hat{\beta}_{ML} = \frac{S_{yy} - S_{xx} + \sqrt{(S_{yy} - S_{xx})^2 + 4(S_{xy})^2}}{2S_{xy}} \quad (3.11)$$

$$\hat{\gamma}_{ML} = \bar{y} - \hat{\beta}_{ML}\bar{x} \quad (3.12)$$

となる。 $S_{xy} = 0$ のときには, $S_{xx} > S_{yy}$ ならば $\hat{\beta}_{ML} = 0$, $S_{xx} < S_{yy}$ ならば $\hat{\beta}_{ML} = \infty$ である。

最尤推定法の性質は原点を通る直線当てはめ問題の場合と同じである。第1に, Neyman-Scott問題であるにもかかわらず, 最尤推定法は推定関数法であり, 最尤推定量は広域的ロバスト推定量になる(川鍋[1995])。

- 最尤推定法は推定関数法である。すなわち, 適当な正則条件のもとで最尤推定量 $(\hat{\beta}_{ML}, \hat{\gamma}_{ML})$ は ξ の分布 $k(\xi)$ によらずに \sqrt{n} -一致性を持ち, 漸近的に正規分布にしたがう。漸近分散は

$$\begin{aligned} & \text{Av}_{\theta, k} \left[\hat{\beta}_{ML}, \hat{\gamma}_{ML} \right] \\ &= \frac{\sigma^2 \{ \sigma^2 + (1 + \beta^2) \sigma_{\xi}^2 \}}{\sigma_{\xi}^4} \begin{bmatrix} 1 & -\mu_{\xi} \\ -\mu_{\xi} & \frac{\sigma^2 \mu_{\xi}^2 + (1 + \beta^2) \sigma_{\xi}^2 (\mu_{\xi}^2 + \sigma_{\xi}^2)}{\sigma^2 + (1 + \beta^2) \sigma_{\xi}^2} \end{bmatrix} \end{aligned} \quad (3.13)$$

となる。

第2に, 最尤推定法は2次元空間の最小自乗法になっている。

- 最尤推定法は観測データ (x_{α}, y_{α}) , $\alpha = 1, \dots, n$ と直線 $y = \beta x + \gamma$ の距離の自乗和

$$\sum_{\alpha=1}^n \frac{(y_{\alpha} - \beta x_{\alpha} - \gamma)^2}{1 + \beta^2} \quad (3.14)$$

を最小にする β, γ を推定量とする最小自乗法に一致する。

原点を通る直線当てはめ問題の場合と同様に、最尤推定法は推定関数法の中で最適な方法ではない。また、最適な推定関数は ξ の分布 $k(\xi)$ に依存するため実際には使うことができないということも同じである。原点を通る直線当てはめ問題の場合には最尤推定法を改良する簡単な L^* -推定関数法を構成した。ところが、一般の直線当てはめ問題の場合には L -推定関数

$$\{x + \beta(y - \gamma) + c\} (y - \gamma - \beta x) \quad (3.15)$$

$$y - \gamma - \beta x \quad (3.16)$$

($c \in \mathbf{R}$) はすべて最尤推定法の推定関数 ($c = 0$) と同値になる。すなわち、データ (x, y) に依存しない正規関数行列をかけることにより互いに移り合う。したがって、任意の $c \in \mathbf{R}$ に対して推定方程式

$$\sum_{\alpha=1}^n \{x_{\alpha} + \beta(y_{\alpha} - \gamma) + c\} (y_{\alpha} - \gamma - \beta x_{\alpha}) = 0 \quad (3.17)$$

$$\sum_{\alpha=1}^n (y_{\alpha} - \gamma - \beta x_{\alpha}) = 0 \quad (3.18)$$

の解 $\hat{\beta}_c, \hat{\gamma}_c$ は最尤推定量 $\hat{\beta}_{ML}, \hat{\gamma}_{ML}$ に一致する。この L -推定関数の同値性より、一般の直線当てはめ問題における最尤推定法が原点を通る直線当てはめ問題における L^* -推定関数法と類似の性質を持つことが示される。

まず正規分布モデルとの関係について述べる。

補題 3.1 最尤推定法の推定関数は ξ の分布が正規分布 $N(\mu_{\xi}, \sigma_{\xi}^2)$ である場合の有効スコア関数(最適推定関数)と同値である。

定理 3.1 ξ が正規分布にしたがっているとき、最尤推定量 $\hat{\beta}_{ML}, \hat{\gamma}_{ML}$ はセミパラメトリックモデル $S = \{p(x, y, \beta, \gamma, \sigma^2, k)\}$ の情報量限界を達成する。

ξ の分布に $\mu_{\xi}, \sigma_{\xi}^2$ をパラメータとする正規分布モデル $K_0 = \{N(\mu_{\xi}, \sigma_{\xi}^2) | \mu_{\xi}, \sigma_{\xi}^2\}$ を仮定する。このとき、 (x, y) の分布は正規分布モデルになり、正規分布の平均と分散はパラメータ $\beta, \gamma, \sigma^2, \mu_{\xi}, \sigma_{\xi}^2$ によって

$$\begin{bmatrix} \mu_{\xi} \\ \beta\mu_{\xi} + \gamma \end{bmatrix}, \quad \begin{bmatrix} \sigma_{\xi}^2 + \sigma^2 & \beta\sigma_{\xi}^2 \\ \beta\sigma_{\xi}^2 & \beta^2\sigma_{\xi}^2 + \sigma^2 \end{bmatrix}.$$

と表される。このようにしてできる S のパラメトリック部分モデルを $S_0 = \{p(x, y, \beta, \gamma, \sigma^2, \mu_{\xi}, \sigma_{\xi}^2)\}$ とおく。

補題 3.2 正規分布モデル S_0 上では、セミパラメトリックモデル $S = \{p(x, y, \beta, \gamma, \sigma^2, k)\}$ における β, γ の情報量限界はパラメトリックモデル S_0 における β, γ の情報量限界と一致する。

定理 3.2 ξ が正規分布にしたがっているとき、線形関係モデルの最尤推定量 $\hat{\beta}_{ML}, \hat{\gamma}_{ML}$ はモデル S_0 の情報量限界を達成する。さらに、線形関係モデルの最尤推定法は正規分布モデル S_0 の最尤推定法と一致する。

次にミニマックス性について述べる。ξの分布のセミパラメトリックモデルを以下のように位置-尺度モデル (location-scale model) の形に書きなおす。

$$K = \left\{ \frac{1}{\sigma_\xi} f \left(\frac{\xi - \mu_\xi}{\sigma_\xi} \right) \mid \mu_\xi, \sigma_\xi, f \in \mathcal{F} \right\} \quad (3.19)$$

ただし、 \mathcal{F} は平均0、分散1の確率密度関数の集合とする。これに対応して (x, y) のセミパラメトリックモデルも

$$S = \{ p(x, y, \beta, \gamma, \sigma^2, \eta, f) \mid \beta, \gamma, \sigma^2, \eta = (\mu_\xi, \sigma_\xi), f \in \mathcal{F} \} \quad (3.20)$$

のように書き換えられる ($\theta = (\beta, \gamma)$ とおく)。

推定関数 $g(x, y, \beta, \gamma)$ の推定方程式

$$\sum_{\alpha=1}^n g(x_\alpha, y_\alpha, \hat{\beta}_g, \hat{\gamma}_g) = 0 \quad (3.21)$$

から得られる推定量を $\hat{\beta}_g, \hat{\gamma}_g$ と置く。任意の2次元ベクトル w を固定し、推定量 $\hat{\beta}_g, \hat{\gamma}_g$ のリスク関数を漸近分散行列 (定理 1.1参照) による2次形式

$$w^T \left(A_{v_{\theta, \sigma^2, \eta, f}} \left[\hat{\beta}_g, \hat{\gamma}_g \right] \right) w$$

とする。

定義 3.1 任意の $\beta, \gamma, \sigma^2, \eta = (\mu_\xi, \sigma_\xi)$ に対して

$$\max_{f \in \mathcal{F}} w^T \left(A_{v_{\theta, \sigma^2, \eta, f}} \left[\hat{\beta}_g, \hat{\gamma}_g \right] \right) w$$

を最小にする推定量 $\hat{\beta}_g, \hat{\gamma}_g$ をミニマックス推定量と呼ぶ。

定理 3.3 任意の2次元ベクトル w に対して最尤推定量 $\hat{\beta}_{ML}, \hat{\gamma}_{ML}$ はミニマックス推定量である。最も不利な分布形は標準正規分布の密度関数 ϕ になる。

最後に一般の直線当てはめ問題の推定法と原点を通る直線当てはめ問題の推定法の類似点と相違点をまとめておく。

類似点

1. 最尤推定法は推定関数法である。

⇒ 最尤推定量はξの分布 $k(\xi)$ がどのようなものでも \sqrt{n} -一致推定量になる。

2. 最尤推定法は最小自乗推定法と一致する。

相違点

1. L -推定関数法は最尤推定法と同値になる (同じ推定量が得られる)。最尤推定法は最適 L -推定関数法では改良できない。
2. ξの分布が正規分布 $N(\mu_\xi, \sigma_\xi^2)$ のとき最尤推定法は最適な推定関数法である (情報量限界を達成)。

4 結論と課題

本稿では昔から数多くの研究がなされている線形関係の推定問題の最も簡単な場合である2次元直線当てはめ問題の推定法について議論した。この問題では観測とともに攪乱パラメータ(データの真値の位置)が増加するので、セミパラメトリックモデルとして取り扱わなければならない(Neyman-Scott[1948])。本研究では情報幾何学の方法を用いてこのセミパラメトリックモデルの推定関数といくつかのセミパラメトリック推定法について考察した。その結果、通常用いられている最尤推定法は幸運なことに推定関数法であり、最尤推定量は真値の分布の仕方によらず一致推定量であることが示される。また、情報量限界および最適な推定関数法を求めることができ、最尤推定法は推定関数法の中では情報量の意味で最適な方法ではないことがわかる。しかし、最適な推定関数は真値の分布に依存するため実際にこれを用いて推定を行なうことはできない。そこで、原点を通る直線を当てはめ問題において、真値の分布の平均と分散の情報を用いることにより最尤推定法を改良する L^* -推定関数法を提案した。一般の直線当てはめ問題の場合には、最尤推定法の関数と L -推定関数がすべて同値になるため、 L^* -推定関数法のような簡単な方法で最尤推定法を改良することはできない。

コンピュータビジョンなどの工学の問題への応用を考えると、今後本稿で扱ったモデルをさらに複雑にして研究を進めていくことが必要になる。

1. 誤差分布が正規分布以外の場合、さらに誤差分布が未知の場合の議論。
2. 誤差分散が異なる未知の値である場合 (Reiersøl モデル) の議論。
3. 多次元線形当てはめ問題への拡張。非線形当てはめ問題の考察。

また、現実のデータに本稿で考察した推定法を適用して実際にどの程度役に立つのかを示すことも今後の課題である。

参考文献

- [1] Amari, S. *Differential-Geometrical Method in Statistics, Lecture Note in Statistics*, **28**, Springer-Verlag, New York, 1985.
- [2] Amari, S. and Kawanabe, M. Information geometry of estimating functions in semiparametric statistical models. 1994. submitted to *Bernoulli*.
- [3] Amari, S. and Kumon, M. Estimation in the presence of infinitely many nuisance parameters — geometry of estimating functions. *Ann.Statist.*, **16**, 1044–1068, 1988.
- [4] Anderson, T.W. Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics. *J. Roy. Statist. Soc. Ser. B*, **38**, 1–36, 1976.

- [5] Anderson, T. W. Estimating linear statistical relationships. *Ann. Statist.*, **12**, 1–45, 1984.
- [6] Begun, J. M., Hall, W. J., Huang, W. M., and Wellner, J. A. Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.*, **11**, 432 – 452, 1983.
- [7] Bickel, P. J. On adaptive estimation. *Ann. Statist.*, **10**, 647–671, 1982. The 1980 Wald memorial lectures.
- [8] Bickel, P. J. and Ritov, Y. Efficient estimation in the errors in variables model. *Ann. Statist.*, **15**, 513–540, 1987.
- [9] Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore, 1993.
- [10] Gleser, L. J. Estimation in a multivariate errors in variables regression model: large sample results. *Ann. Statist.*, **9**, 24–44, 1981.
- [11] Godambe, V. P. Conditional likelihood and unconditional optimum estimating equations. *Biometrika*, **63**, 277 – 284, 1976.
- [12] Godambe, V. P., editor. *Estimating Functions*, New York, 1991. Oxford University Press.
- [13] Godambe, V. P. An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.*, **31**, 1208–1212, 1960.
- [14] Huber, P. J. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 221–233, Berkeley, 1967. University of California Press.
- [15] Kendall, M. G. and Stuart, A. *The Advanced Theory of Statistics*, **2**, Griffin, London, 4th edition, 1979.
- [16] Kumon, M. and Amari, S. Estimation of a structural parameter in the presence of a large number of nuisance parameters. *Biometrika*, **71**, 445 – 459, 1984.
- [17] Lindsay, B. G. Conditional score functions : Some optimality results. *Biometrika*, **69**, 503 – 512, 1982.
- [18] Lindsay, B. G. Using empirical partially bayes inference for increased efficiency. *Ann. Statist.*, **13**, 914 – 931, 1985.
- [19] Mariano, R. S. and Sawa, T. The exact finite-sample distribution of the limited-information maximum likelihood estimator in the case of two included endogenous variables. *J. Amer. Statist. Assoc.*, **67**, 159–163, 1972.

- [20] Nagaoka, H. and Amari, S. Differential geometry of smooth families of probability distributions. Technical Report 82 - 7, Univ. Tokyo, 1982.
- [21] Neyman, J. and Scott, E. L. Consistent estimates based on partially consistent observations. *Econometrica*, **32**, 1 - 32, 1948.
- [22] Reiersøl, O. Identifiability of a linear relation between variables which are subject to error. *Econometrica*, **18**, 375-389, 1950.
- [23] 金谷 健一. コンピュータビジョンの数理的方法. 応用数理, 4, 259-268, 1994.
- [24] 金谷 健一. 空間データの数理. 朝倉書店, 1995.
- [25] 川鍋 元明. 推定関数の幾何学とその応用. 東京大学工学系研究科博士論文, 1995.