

双対平坦空間の情報幾何を利用した統計的推定

東京大学・情報理工学系研究科 廣瀬 善大

東京大学・情報理工学系研究科, 理化学研究所・脳科学総合研究センター 駒木 文保

Yoshihiro Hirose

Graduate School of Information Science and Technology, University of Tokyo

Fumiyasu Komaki

Graduate School of Information Science and Technology, University of Tokyo

Brain Science Institute, RIKEN

概要

文献 [5], [6], [7] において提案された統計的推定手法 bisector regression について解説する。これらの論文で考えられている手法は, Efron らによる least angle regression の拡張を目指したものである。拡張にあたっては双対平坦空間の情報幾何を利用している。拡張により得られた bisector regression アルゴリズムは, least angle regression の厳密な拡張ではないが, least angle regression 同様に角の二等分線を利用したパラメータ推定手法となっている。一般化線形回帰, ガウシアングラフィカルモデル, 分割表の 3 つの問題を考え, 特にガウシアングラフィカルモデルについてのアルゴリズムを紹介する。また, bisector regression の適用例も紹介する。

1 導入

本稿では, 文献 [5], [6], [7] において提案された統計的推定手法 bisector regression について解説する。これらの論文で考えられている手法は, Efron らによる least angle regression (LARS, [3]) の拡張を目指したものである。拡張にあたっては双対平坦空間の情報幾何 ([1], [2], [8]) を利用している。拡張により得られた bisector regression アルゴリズムは, LARS の厳密な拡張ではないが, LARS 同様に角の二等分線を利用したパラメータ推定手法となっている。

LARS は線形回帰問題におけるパラメータ推定 (回帰係数の推定) のために提案されたアルゴリズムであり, 同時にモデル推定も行う。アルゴリズムの出力としてパラメータの推定値とモデルの系列が得られる。説明変数間の相関を内積 (角度) とみなし, 推定量の構成において反応変数との相関が大きな (角度の小さな) 説明変数を順次追加することで推定値の系列を構成する。LARS アルゴリズムはユークリッド幾何により記述され, 幾何学的に単純であり統計学的にも自然に解釈される。また, LARS は正則化の代表的な手法である lasso ([13], [14]) と関連がある。正則化とは, 最尤推定量の過学習を避けるために, 尤度に正則化項 (罰則化項) を加えたものを最大化する手法である ([4])。LARS 自

体は正則化ではないが、LARS アルゴリズムを修正することで lasso による推定値を出力できることが知られている。

双対平坦空間の情報幾何はユークリッド幾何を一般化したものである。確率分布族のなす空間を考え、計量や接続といった構造を入れる。統計学でよく現れる Fisher 情報を計量とし、 α -接続と呼ばれる接続の族を考える。 α -接続には互いに双対な構造をもつ組がある。本稿では、その中でも特に、1-接続と -1 -接続を利用する。1-接続と -1 -接続はそれぞれ e-接続、m-接続とも呼ばれる。リーマン接続に対応するのは 0-接続であり、自己双対な接続である。接続が与えられると測地線や平坦さを扱うことができ、e-接続と m-接続のそれぞれに関する測地線等があらわれる。また、e-接続に関して平坦な空間は m-接続の意味でも平坦であることが知られており、その逆も成り立つ。このことから互いに双対な接続に関して平坦な空間は双対平坦空間と呼ばれる。与えられた接続に関して平坦な空間には便利な座標系（アファイン座標系）が存在し、双対平坦空間には e-接続と m-接続というふたつの接続に関するアファイン座標系がそれぞれ存在する。指数型分布族は ± 1 -接続に関して双対平坦空間になることが知られている。自然パラメータが e-アファイン座標系になっており、期待値パラメータが m-アファイン座標系になっている。情報幾何は、統計学においては高次の漸近論などで有用であり、さらに制御論や最適化などの分野でも有用である。

文献 [5], [6], [7] において bisector regression による推定が扱われた問題は 3 つある。一般化線形回帰、ガウシアングラフィカルモデル、分割表である。一般化線形回帰は線形回帰を含むより広い問題のクラスであり、LARS の直接的な拡張を目指した問題である。3 つの問題の共通点は、いずれも指数型分布族の自然パラメータの推定問題であることである。本稿では特にガウシアングラフィカルモデルについての bisector regression を説明する。

本稿の構成は以下の通りである。第 2 節において bisector regression を適用する 3 つの問題を紹介する。それぞれの問題で推定するパラメータを定義し、後の節のための設定を確認する。第 3 節では bisector regression アルゴリズムの解説に必要な情報幾何の説明を行う。指数型分布族は双対平坦空間をなし、bisector regression アルゴリズムはその双対平坦空間内の入れ子構造を利用して推定を行う。アルゴリズムの記述とその解説は第 4 節において行う。ガウシアングラフィカルモデルをとりあげ、まずアルゴリズムの幾何学的な説明を行い、その後 bisector regression アルゴリズムを紹介する。第 5 節で本稿のまとめをする。

2 問題

提案手法 bisector regression が適用される問題について説明する. 文献 [5] にある通り, 3つの問題を考える. 扱う問題は, 一般化線形回帰, ガウシアングラフィカルモデル, 分割表である. いずれの場合も, bisector regression はパラメータ推定とモデル推定を同時に行う.

2.1 一般化線形回帰

一般化線形回帰におけるパラメータ推定は, 一般化線形モデル (generalized linear model, GLM) のパラメータ推定であり ([10]), 対応する指数型分布族の中から分布をひとつに決定する問題である. $\{y_a, x^a = (x_1^a, x_2^a, \dots, x_d^a)\}_{a=1,2,\dots,n}$ が観測されているものとする. 計画行列 X は $X = (x_i^a)_{1 \leq a \leq n, 1 \leq i \leq d} = (x_1, x_2, \dots, x_d)$ で与えられ, $n \times d$ 行列である. ただし, $x_i = (x_i^1, x_i^2, \dots, x_i^n)^\top$ ($i = 1, 2, \dots, d$) とおいた. 1 を n 個の 1 をもつベクトル, すなわち $1 = (1, 1, \dots, 1)^\top$ とし, $n \times (d+1)$ 行列 \tilde{X} を $\tilde{X} = (1|X)$ と定義する.

指数型分布族は

$$f(y|\xi) = \exp \left(\sum_{a=1}^n y_a \xi^a + \sum_{b=1}^r u_b(y) \xi^{b+n} - \psi(\xi) \right)$$

と表される. ただし, $\xi' := (\xi^1, \xi^2, \dots, \xi^n)^\top$, $\xi'' := (\xi^{n+1}, \xi^{n+2}, \dots, \xi^{r+n})^\top$, $\xi := (\xi^1, \xi^2, \dots, \xi^{n+r})^\top = (\xi'^\top, \xi''^\top)^\top$, $y = (y_1, y_2, \dots, y_n)^\top$ とおくことにする. $u(y) := \{y_1, \dots, y_n, u_1(y), u_2(y), \dots, u_r(y)\}$ は ξ に関する十分等計量であり, $\psi(\cdot)$ は正規化定数に対応する凸関数である. パラメータ ξ は自然パラメータと呼ばれ, $\psi(\cdot)$ は ξ のポテンシャル関数と呼ばれる. 一般化線形回帰問題においては, さらに

$$\xi' = \tilde{X} \theta', \quad \xi'' = \theta'', \quad \theta \in \Theta$$

として θ' を推定する. ただし, $\theta' := (\theta^0, \theta^1, \dots, \theta^d)^\top$, $\theta'' := (\theta^{d+1}, \theta^{d+2}, \dots, \theta^{r+d})^\top$, $\theta := (\theta^0, \theta^1, \dots, \theta^{d+r})^\top = (\theta'^\top, \theta''^\top)^\top$ とおく. $\Theta \subset \mathbb{R}^{d+r}$ はパラメータ空間である. 十分等計量 $u(y)$ の期待値 $\mu = (\mu_1, \dots, \mu_{n+r})^\top$ は $\mu_a = E[y_a]$ ($a = 1, \dots, n$), $\mu_{b+n} = E[u_b(y)]$ ($b = 1, \dots, r$) で与えられ, 期待値パラメータと呼ばれる.

$r > 0$ の場合, 推定したいのは θ' であり θ'' は局外パラメータである. パラメータのうち bisector regression アルゴリズムにより推定されるのは θ' だけである. 局外パラメー

タの扱いについては、第4節においてガウシアングラフィカルモデルを例にとって説明する。一般化線形回帰においても同様な手続きが必要になる。

一般化線形回帰の例としてロジスティック回帰を考える。考える指数型分布族は

$$f(y|\xi) = \prod_{a=1}^n \frac{\exp(y_a \xi^a)}{1 + \exp \xi^a} = \exp \left(\sum_{a=1}^n y_a \xi^a - \psi(\xi) \right)$$

である。ただし、 $\xi = (\xi^1, \xi^2, \dots, \xi^n)^\top$ は自然パラメータであり、 $y \in \{0, 1\}^n$, $r = 0$ である。 ξ に関するポテンシャル関数は $\psi(\xi) = \sum_{a=1}^n \log(1 + \exp \xi^a)$ である。一般化線形回帰問題として $\xi = \tilde{X}\theta$, $\Theta = \mathbb{R}^d$, $\theta = (\theta^0, \theta^1, \dots, \theta^d)^\top$ を考える。期待値パラメータ μ は $\mu_a = \mathbb{E}[y_a] = \exp \xi^a / (1 + \exp \xi^a)$ ($a = 1, 2, \dots, n$) で与えられる。パラメータ θ を推定することが問題である。

2.2 ガウシアングラフィカルモデル

ガウシアングラフィカルモデル (Gaussian graphical model, GGM) を考える。問題は独立グラフと呼ばれるグラフにおける辺選択である。独立グラフとは、二変数の間の条件付き独立性に対応して辺が与えられたグラフである。多変量正規分布の精度行列 $\Sigma^{-1} = (\sigma^{ab})$ の非対角成分の推定により独立グラフにおける辺の判定を行う。本稿では無向グラフのみを考える。グラフィカルモデルについては、文献 [9] などを参照のこと。

p 変量確率変数 $X = (X_1, \dots, X_p)^\top$ を考え、 X は多変量正規分布にしたがうものとする。二変数の条件付き独立性は次のように定義される。すなわち、 X_a と X_b が他の変数が与えられた下で条件付き独立であるとは、 $\sigma^{ab} = 0$ が成り立つことである。ただし、 $\Sigma = (\sigma_{ab})$ は共分散行列であり、 $\Sigma^{-1} = (\sigma^{ab})$ は共分散行列 Σ の逆行列で精度行列と呼ばれる。独立グラフとは、 X の各成分に対してグラフの頂点をひとつずつ対応させたもので、二頂点 X_a と X_b の間の辺の有無は条件付き独立の場合には辺を与えないことにより決定される (図 1)。 p 変量正規分布にしたがう確率変数 X に対応する独立グラフは、 p 個の頂点をもち最大 $p(p-1)/2$ 本の辺をもつグラフである。確率分布に多変量正規分布 (ガウス分布) を仮定しているの、この独立グラフを利用して得られるモデルはガウシアングラフィカルモデルと呼ばれる。つまり、GGM とは、独立グラフにおいて辺のない二変数が条件付き独立であるような多変量正規分布の集合である。グラフにおいて辺がある二変数については、対応する GGM に含まれる分布に条件付き独立性に関する制約は課されない。

問題は GGM における辺選択である。条件付き独立性の定義から、精度行列 Σ^{-1} の非対角成分 σ^{ab} ($a < b$) の推定を行うことになる。観測として p 変量変数が n サンプル分与

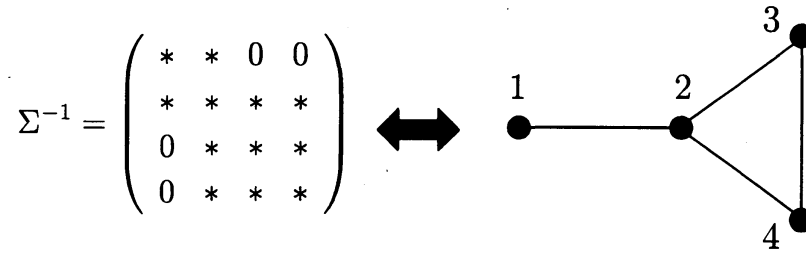


図1 確率変数が4変量の場合 ($p = 4$) における精度行列と独立グラフの例. * は任意の値を表す. 精度行列 Σ^{-1} において $\sigma^{13} = \sigma^{14} = 0$ となっている. Σ^{-1} に対応する独立グラフにおいては, 2つの辺 e_{13} と e_{14} が除去されている.

えられているものとする. あるいは観測の標本平均と標本共分散行列が与えられているものとする. 平均と Σ^{-1} の対角成分は局外パラメータである. 興味のあるパラメータと局外パラメータの扱いの違いについては, GGM を例にとり第4節で説明する.

2.3 分割表

分割表の例を表1に示す. ここでは二元分割表を例にして説明する. 因子を2つ以上もつ場合についても同様に扱うことができる.

2つの因子 X_1 と X_2 からなる分割表を考える. X_1 は $m+1$ 個の水準を, X_2 は $n+1$ 個の水準をもつものとする. すなわち, $X_1 = 0, 1, \dots, m$, $X_2 = 0, 1, \dots, n$. 考えるのは多項分布であり, その確率分布は

$$f(y|p) = \frac{N!}{y_{00}! y_{01}! \dots y_{mn}!} p_{00}^{y_{00}} p_{01}^{y_{01}} \dots p_{mn}^{y_{mn}}.$$

ただし, N は総観測数, y_{ij} はセル (i, j) の観測数であり $\sum_{i=0}^m \sum_{j=0}^n y_{ij} = N$ を満たす. また, p_{ij} はセル (i, j) の確率で $\sum_{i=0}^m \sum_{j=0}^n p_{ij} = 1$ を満たす.

確率分布の対数をとると,

$$\begin{aligned} \log f(y|p) &= \sum_{i=0}^m \sum_{j=0}^n y_{ij} \log p_{ij} + \log \frac{N!}{y_{00}! y_{01}! \dots y_{mn}!} \\ &= \sum_{i=1}^m \sum_{j=1}^n y_{ij} \log \frac{p_{ij} p_{00}}{p_{i0} p_{0j}} + \sum_{i=1}^m \left(\sum_{j=0}^n y_{ij} \right) \log \frac{p_{i0}}{p_{00}} \\ &\quad + \sum_{j=1}^n \left(\sum_{i=0}^m y_{ij} \right) \log \frac{p_{0j}}{p_{00}} + N \log p_{00} + \log \frac{N!}{y_{00}! y_{01}! \dots y_{mn}!} \end{aligned}$$

表 1 分割表の例. X_1 : $m + 1$ 個の水準をもつ因子, X_2 : $n + 1$ 個の水準をもつ因子, y_{ij} : セル (i, j) の観測数, N : 総観測数.

$X_1 \backslash X_2$	0	1	...	n	計
0	y_{00}	y_{01}	...	y_{0n}	y_{0+}
1	y_{10}	y_{11}	...	y_{1n}	y_{1+}
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
m	y_{m0}	y_{m1}	...	y_{mn}	y_{m+}
計	y_{+0}	y_{+1}	...	y_{+n}	N

と変形できる. $i = 1, 2, \dots, m, j = 1, 2, \dots, n$ に対して

$$\theta_{X_1}^i = \log \frac{p_{i0}}{p_{00}}, \theta_{X_2}^j = \log \frac{p_{0j}}{p_{00}}, \theta_{X_1 X_2}^{ij} = \log \frac{p_{ij} p_{00}}{p_{i0} p_{0j}}$$

とおき, 自然パラメータ

$$\begin{aligned} \theta &= (\theta_{X_1}; \theta_{X_2}; \theta_{X_1 X_2}) \\ &= (\theta_{X_1}^1, \theta_{X_1}^2, \dots, \theta_{X_1}^m; \theta_{X_2}^1, \theta_{X_2}^2, \dots, \theta_{X_2}^n; \theta_{X_1 X_2}^{11}, \theta_{X_1 X_2}^{12}, \dots, \theta_{X_1 X_2}^{mn}) \end{aligned}$$

を推定する. ここで, θ_{X_1} は m 次元ベクトル, θ_{X_2} は n 次元ベクトル, $\theta_{X_1 X_2}$ は mn 次元ベクトルである. bisector regression により推定したいパラメータは $\theta_{X_1 X_2}$ であり, θ_{X_1} と θ_{X_2} は局外パラメータである.

3 情報幾何

本節では bisector regression アルゴリズムに必要な情報幾何の説明を行う. 双対平坦空間の情報幾何はユークリッド幾何の一般化のひとつである. 双対平坦空間においては, ユークリッド空間と同様に直線や距離に対応する量を扱うことができる. そのためユークリッド幾何により記述されていた LARS アルゴリズムの拡張を目指すことができる. 本稿では指数型分布族のみを扱う. 情報幾何の詳細については, 文献 [1], [2], [8] などを参照のこと.

指数型分布族のなす空間 $S = \{f(x|\theta) | \theta \in \Theta\}$ を考える. 指数型分布族は

$$f(x|\theta) = \exp \left(\sum_{a=1}^n u_a(x) \theta^a - \psi(\theta) \right)$$

と表される。 θ は自然パラメータ、 ψ は θ に関するポテンシャル関数、 $u(x) = (u_\alpha(x))$ は十分統計量と呼ばれる。 指数型分布族のパラメータとして、十分統計量の期待値をとった期待値パラメータ $\eta = E[u(x)]$ も有用である。ここで、 $\partial_i = \partial/\partial\theta^i$, $\partial^i = \partial/\partial\eta_i$ と略記することにする。計量として Fisher 情報

$$g_{ij} = E[(\partial_i \log f(X|\theta))(\partial_j \log f(X|\theta))]$$

を用いる。 α -接続と呼ばれる接続の族を考え、特に $\alpha = \pm 1$ に注目する。1-接続と -1 -接続はそれぞれ e-接続、m-接続とも呼ばれ、互いに双対な構造をもつ。自然パラメータ θ に関する α -接続の係数 $\Gamma_{ijk}^{(\alpha)}$ は

$$\Gamma_{ijk}^{(\alpha)} = E\left[\left(\partial_i \partial_j \log f(X|\theta) + \frac{1-\alpha}{2} \partial_i \log f(X|\theta) \partial_j \log f(X|\theta)\right) \partial_k \log f(X|\theta)\right]$$

で与えられる。

一般に、あるパラメータに関して接続の係数がすべて 0 であるとき空間 S は α -平坦であると言われ、そのパラメータは α -アファイン座標系と呼ばれる。 S が α -平坦であれば $-\alpha$ -平坦であり、その逆も成り立つ。このことから、 $\pm\alpha$ -平坦な空間を双対平坦空間と呼ぶ。いま S は指数型分布族であり、 ± 1 -接続に関して双対平坦空間になることが知られている。自然パラメータが e-アファイン座標系であり、期待値パラメータが m-アファイン座標系である。

期待値パラメータ η に関するポテンシャル関数 ϕ を $\phi(\eta) = E[\log f(x|\eta)]$ で定義する。このとき、関係式

$$\begin{aligned} \partial_i \psi &= \eta_i, \quad \partial^i \phi = \theta^i, \quad \partial_i \partial_j \psi = g_{ij}, \quad \partial^i \partial^j \psi = g^{ij}, \\ \phi(\eta) + \psi(\theta) - \eta \cdot \theta &= 0 \end{aligned}$$

が成り立つ。ただし、 (g^{ij}) は Fisher 情報行列 (g_{ij}) の逆行列、 $\eta \cdot \theta = \sum \eta_i \theta^i$ である。

α -接続に関する測地線、 α -測地線を定義する。 S の曲線 $\gamma: t \mapsto \gamma(t)$ が α -測地線であるとは、任意の $t \in [0, 1]$, k に対して

$$\ddot{\gamma}^k(t) + \sum_{i,j} \dot{\gamma}^i(t) \dot{\gamma}^j(t) \left(\Gamma_{ij}^{(\alpha)k} \right)_{\gamma(t)} = 0$$

が成り立つことである。ただし、 $\Gamma_{ij}^{(\alpha)k}$ は $\Gamma_{ij}^{(\alpha)k} = \sum_l \Gamma_{ijl}^{(\alpha)} g^{lk}$ で与えられ、 $(\gamma^1(t), \gamma^2(t), \dots, \gamma^d(t))$ は曲線 $\gamma(t)$ の成分である。測地線は直線に対応するものである。 $\alpha = 1$ とすると、 $\theta_{(1)}$ と $\theta_{(2)}$ を結ぶ e-測地線は、 e-アファイン座標系である自然パラメータが

$$\theta_\gamma(t) = t\theta_{(1)} + (1-t)\theta_{(2)}, \quad t \in [0, 1]$$

と表される曲線である。同様に、 $\eta_{(1)}$ と $\eta_{(2)}$ を結ぶ m -測地線は、期待値パラメータが

$$\eta_\gamma(t) = t\eta_{(1)} + (1-t)\eta_{(2)}, \quad t \in [0, 1]$$

と表される曲線である。

点 $\eta_{(1)}$ から点 $\theta_{(2)}$ への Kullback-Leibler ダイバージェンス (KL ダイバージェンス) は

$$D(\eta_{(1)} | \theta_{(2)}) = E_{\eta_{(1)}} \left[\frac{\log f(X | \eta_{(1)})}{\log f(X | \theta_{(2)})} \right]$$

で定義される。KL ダイバージェンスはユークリッド距離の二乗に対応する量である。

KL ダイバージェンスはポテンシャル関数 ψ, ϕ を用いて

$$D(\eta_{(1)} | \theta_{(2)}) = \phi(\eta_{(1)}) + \psi(\theta_{(2)}) - \eta_{(1)} \cdot \theta_{(2)}$$

と表すことができる。

多変量正規分布 (μ_1, Σ_1) と (μ_2, Σ_2) との KL ダイバージェンス $D(\mu_1, \Sigma_1 | \mu_2, \Sigma_2)$ は

$$\begin{aligned} D(\mu_1, \Sigma_1 | \mu_2, \Sigma_2) = & -\frac{1}{2} \log |\Sigma_1| + \frac{1}{2} \log |\Sigma_2| + \frac{1}{2} \text{tr}(\Sigma_1 \Sigma_2^{-1}) - \frac{p}{2} \\ & + \frac{1}{2} (\mu_2 - \mu_1)^\top \Sigma_2^{-1} (\mu_2 - \mu_1) \end{aligned}$$

で与えられる。第 4 節で考える bisector regression では期待値 μ は固定される。期待値が同じ多変量正規分布については、 Σ_1 から Σ_2 への KL ダイバージェンスは

$$D(\Sigma_1 | \Sigma_2) = -\frac{1}{2} \log |\Sigma_1| + \frac{1}{2} \log |\Sigma_2| + \frac{1}{2} \text{tr}(\Sigma_1 \Sigma_2^{-1}) - \frac{p}{2}$$

で与えられる。

最後に m -射影を定義する。 S' を S の部分空間とし、 $p \in S$ とする。点 p の S' への m -射影とは、 p からのダイバージェンスが最小であるような S' の点である。

4 推定手法

第 2 節において bisector regression を適用する問題として 3 つの問題を紹介した。アルゴリズムの基本的なアイデアはどの問題においても同じであるが、問題ごとに異なる点がある。本節で bisector regression のアルゴリズムを解説するに当たっては、3 つの問題のうち特にガウシアングラフィカルモデル (GGM) に注目することにする。線形回帰、一般化線形回帰では確率変数の期待値を扱っていたが、第 2.2 節で確認した通り GGM の辺選択においては確率変数の共分散行列を扱う。したがって、線形回帰と同じく正規分布

を扱うが、考える問題の構造は異なっている。多変量正規分布は指数型分布族であり双対平坦空間になっていた。GGM の辺選択問題は、情報幾何を用いることで、一般化線形回帰とは異なった形で自然に扱うことのできる問題である。

推定の手続きとしては、まず局外パラメータを固定し、残されたパラメータを bisector regression アルゴリズムで推定するという手続きをとる。局外パラメータの固定は、線形回帰における原点の設定に対応するものであり、計画行列の正規化作業を一般化したものと考えてよい。ただし、bisector regression においてはパラメータのスケーリングについては不変性があり、bisector regression を適用するモデルを設定することが局外パラメータの固定の幾何学的意味である。

多変量正規分布は平均ベクトル μ と共分散行列 Σ をパラメータとしてもつ。あるいは $u = \Sigma^{-1}\mu$ と Σ^{-1} をパラメータとすることもできる。前者が期待値パラメータであり、後者が自然パラメータである。2種類のパラメータの双対性を利用して局外パラメータを固定する。つまり、期待値パラメータである μ と Σ の対角成分を最尤推定値に固定し、自然パラメータである Σ^{-1} の非対角成分を推定する。パラメータの数は $d = p(p-1)/2$ である。これにより局外パラメータと推定するパラメータが直交し、bisector regression による推定を自然に行うことができるようになる。

以下では、まず局外パラメータの固定について説明を行い、その後 bisector regression アルゴリズムを紹介する。GGM に特化した bisector regression アルゴリズムなので、便宜上 BRGGM (Bisector Regression for Gaussian Graphical Models) と呼ぶことにする。

4.1 局外パラメータの固定

BRGGM アルゴリズムは、多変量正規分布モデルのうち、期待値と分散が固定されたサブモデルにおいて推定値を更新するアルゴリズムである。期待値ベクトル μ と共分散行列 Σ の対角成分 (分散 σ_{aa}) をそれぞれ $\mu = \hat{\mu}_{\text{MLE}}$ と $\sigma_{aa} = (\hat{\Sigma}_{\text{MLE}})_{aa}$ ($a = 1, \dots, p$) として固定する。ただし、 $\hat{\mu}_{\text{MLE}}$ と $(\hat{\Sigma}_{\text{MLE}})_{aa}$ はそれぞれ最尤推定量であり、実際には標本平均と標本分散である。標本共分散行列を $\hat{\Sigma}_{\text{MLE}}$ とおく。サブモデル M を $M = \{\Sigma \mid \sigma_{aa} = (\hat{\Sigma}_{\text{MLE}})_{aa} (a = 1, \dots, p)\}$ とおく (図 2)。なお、平均は固定されているので表記を省略する。サブモデル M は双対平坦空間であり、BRGGM アルゴリズムが動く空間である。表記上、 Σ と Σ^{-1} はともに共分散行列 Σ の多変量正規分布を表すものとする。BRGGM アルゴリズムは M における推定量の更新 (移動) として記述される。最初の推定量 $\hat{\Sigma}_{(0)}^{-1}$ はフルモデルの最尤推定量 $\hat{\Sigma}_{\text{MLE}}^{-1}$ であり、独立グラフ $\hat{G}_{(0)}$ としては完全グラフ \hat{G}_{MLE} に対応する。最後に出力される推定量 $\hat{\Sigma}_{(d)}^{-1}$ は各変数がそれぞれ独立で

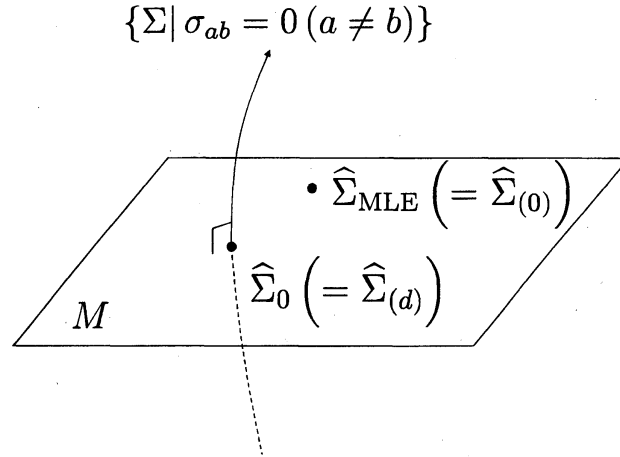


図 2 BRGGM アルゴリズムが適用されるサブモデル. $M = \{\Sigma | \sigma_{aa} = (\hat{\Sigma}_{MLE})_{aa} (a = 1, \dots, p)\}$. $\hat{\Sigma}_{MLE}$: 標本共分散, BRGGM の最初の推定量 $\hat{\Sigma}_{(0)}$ でもある. $\hat{\Sigma}_0$: 各変数が独立であるというモデル $\{\Sigma | \sigma_{ab} = 0 (a \neq b)\}$ での最尤推定量であり, BRGGM が最後に出力する推定量 $\hat{\Sigma}_{(d)}$ はこれと一致する. BRGGM アルゴリズムが適用されるサブモデル M と独立モデル $\{\Sigma | \sigma_{ab} = 0 (a \neq b)\}$ は直交する. $\hat{\Sigma}_{MLE}$ と $\hat{\Sigma}_0$ はともに M に含まれる.

あるというモデル $\{\Sigma | \sigma_{ab} = 0 (a \neq b)\}$ での最尤推定量 $\hat{\Sigma}_0^{-1}$ であり, 独立グラフ $\hat{G}_{(d)}$ は辺をひとつもたないグラフ \hat{G}_0 となる. なお, $\hat{\Sigma}_0$ の対角成分と $\hat{\Sigma}_{MLE}$ の対角成分の値は一致する. また, BRGGM アルゴリズムが適用されるサブモデル M と独立モデル $\{\Sigma | \sigma_{ab} = 0 (a \neq b)\}$ が直交することが知られている.

4.2 Bisector regression アルゴリズム

GGM の辺選択に対して bisector regression アルゴリズムを説明する. 本節の説明はサブモデル M におけるものであり, 平均と分散は固定されているので省略する. 精度行列 Σ^{-1} の非対角成分 $\sigma^{ab} (a \neq b)$ を推定し, 対応する独立グラフ G を推定することが目的である. $(a, b) \in I \subseteq \{(a', b') | 1 \leq a' < b' \leq p\}$, $s \in \mathbf{R}$ に対して, ふたつのサブモデル $M(I)$ と $M((a, b), s, I)$ を

$$M(I) = \left\{ \Sigma \mid \sigma^{a'b'} = 0 \left((a', b') \notin I \right) \right\},$$

$$M((a, b), s, I) = \left\{ \Sigma \mid \sigma^{ab} = s, \sigma^{a'b'} = 0 \left((a', b') \notin I \right) \right\}$$

により定義する. サブモデル $M(I)$ は $(a', b') \notin I$ に対して $X_{a'}$ と $X_{b'}$ の間に辺をもたない独立グラフに対応するモデルである. サブモデル $M((a, b), 0, I)$ は $(a', b') \notin I$ に対して $X_{a'}$ と $X_{b'}$ が辺をもたず, さらに X_a と X_b も辺をもたないような独立グラフに対応す

るモデルである。

まず BRGGM アルゴリズムで何をしようとしているのかについて幾何学的な説明をし、その後 BRGGM アルゴリズムを紹介する。BRGGM アルゴリズム、あるいはガウシアングラフィカルモデル以外の問題における bisector regression アルゴリズムについても、線形回帰問題におけるパラメータ推定・モデル推定の手法である least angle regression (LARS) の拡張を目指したものである。アルゴリズムの反復ごとにパラメータの推定値を出力し、推定値の入るモデルが入れ子状に順に小さくなっていくのが特徴である。これによりアルゴリズムの反復はパラメータの次元と同じ回数だけであることが分かり、考えるモデルの候補数が大幅に削減される。

LARS では、説明変数と反応変数の相関を内積（角度）として扱い、反応変数となす相関の大きい（角度の小さい）説明変数を推定量の構成に順次取り込んでいた。すでに推定量を構成している説明変数同士は反応変数となす相関（角度）が互いに一致しており、幾何学的には推定量が角の二等分線上を動いているものとみなせる。双対平坦空間の情報幾何を利用して LARS の拡張を行うにあたって、角の二等分線を距離を用いた形で一般化し、bisector regression はその曲線上を推定量が動くようなアルゴリズムになっている。ただし、bisector regression は LARS の厳密な拡張ではない。LARS では原点を出発点にしてアルゴリズムが始まり、フルモデルの最尤推定量を推定量の最終的な到達点にしていた。しかし、bisector regression では、推定量はフルモデルの最尤推定量を出発し、最終的に原点にたどりつくようになっている。これはユークリッド空間を双対平坦空間に一般化する際に生じる困難を避けるためであった。推定量の移動（更新）によりパラメータの推定値とモデルの組の系列を出力する点は同じである。

4.1 節において局外パラメータの固定について説明した。BRGGM アルゴリズムはサブモデル M に対して働くアルゴリズムである。 M 内を動く推定量の点を考え、 $M(I)$ の形で表される M のサブモデルに推定量が順次動いていく手続きを記述する。図 3 は推定量の点が角の二等分線に対応する曲線に沿って移動する様子を表したものであり、bisector regression アルゴリズムではこの手続きを反復することになる。推定量の点が k 回目の反復で得られた推定値 $\hat{\Sigma}_{(k)}$ にいるものとする。アルゴリズム開始時には $\hat{\Sigma}_{(0)} = \hat{\Sigma}_{\text{MLE}}$ である。精度行列 $\hat{\Sigma}_{(k)}^{-1}$ は $d - k$ 個の非零成分をもち、 $|I| = d - k$ を満たすある $I \subseteq \{(a, b) \mid 1 \leq a < b \leq p\}$ に対して $\hat{\Sigma}_{(k)}^{-1} \in M(I)$ となっている。各 $(a, b) \in I$ に対して、サブモデル $M(I \setminus (a, b))$ を考える。サブモデル $M(I \setminus (a, b))$ は、 $M(I)$ よりも一次元だけ次元の小さなサブモデルで $M(I)$ に含まれている。現在の推定値 $\hat{\Sigma}_{(k)}$ から各サブモデル $M(I \setminus (a, b))$ への KL ダイバージェンスを測り、その最小値を t^* とおくことにする。 $\hat{\Sigma}_{(k)}$ からの KL ダイバージェンスが t^* になるように各サブモデル $M(I \setminus (a, b))$

を平行移動したものは、それぞれ $M((a, b), s_{ab}^*, I)$ の形で表される。 $M(I)$ における各 $M((a, b), s_{ab}^*, I)$ の交点を $\hat{\Sigma}_{(k+1)}$ と決める。このとき、 t^* を達成した $(a^*, b^*) \in I$ に対して $\hat{\Sigma}_{(k+1)} \in M(I \setminus (a^*, b^*))$ であるから、推定値を含むサブモデルの次元がひとつ小さくなったことが分かる。なお、上述の説明では角の二等分線に対応する曲線を陽に用いていないが、曲線の両端がそれぞれ推定値 $\hat{\Sigma}_{(k)}$ と $\hat{\Sigma}_{(k+1)}$ に対応している。

角の二等分線に対応する曲線について図 3 にしたがって簡単に説明する。現在の推定値 $\hat{\Sigma}_{(k)}$ から各サブモデル $M(I \setminus (a, b))$ への m -射影を $\bar{\Sigma}_{(k)}^{-ab}$ とおく。 $\hat{\Sigma}_{(k)}$ から各 $\bar{\Sigma}_{(k)}^{-ab}$ への KL ダイバージェンスのうち最小値が t^* であり、最小値を達成するのは $(a^*, b^*) \in I$ である。 $\hat{\Sigma}_{(k)}$ から $\tilde{\Sigma}_{(k)}^{-a'b'}$ への KL ダイバージェンスが $\hat{\Sigma}_{(k)}$ から $\bar{\Sigma}_{(k)}^{-a^*b^*}$ への KL ダイバージェンスと一致するように、サブモデル $M(I \setminus \{(a', b')\})$ を平行移動して $M((a', b'), s_{a'b'}^*, I)$ とする。ただし、 $\tilde{\Sigma}_{(k)}^{-a'b'}$ は $\hat{\Sigma}_{(k)}$ の $M((a', b'), s_{a'b'}^*, I)$ への m -射影であり、 $s_{a'b'}^*$ は $D(\hat{\Sigma}_{(k)} | \tilde{\Sigma}_{(k)}^{-a'b'}) = t^*$ を満たす値として決まる。 (a^*, b^*) に対して同様のことを考えると、 $s_{a^*b^*}^* = 0$ 、 $\tilde{\Sigma}_{(k)}^{-a^*b^*} = \bar{\Sigma}_{(k)}^{-a^*b^*}$ であることが分かる。新しい推定値 $\hat{\Sigma}_{(k+1)}$ は $M((a', b'), s_{a'b'}^*(t), I)$ と $M((a^*, b^*), s_{a^*b^*}^*(t), I) = M(I \setminus \{(a^*, b^*)\})$ との交点であった。ここで、 $0 \leq t \leq t^*$ なる t に対して、 t^* の代わりに t を使って $\tilde{\Sigma}_{(k)}^{-a'b'}$ と $\tilde{\Sigma}_{(k)}^{-a^*b^*}$ に対応する点を作ることができる。そして、それらをそれぞれ含むようなサブモデル $M((a', b'), s_{a'b'}^*(t), I)$ と $M((a^*, b^*), s_{a^*b^*}^*(t), I)$ の交点 $\hat{\Sigma}(t)$ を考えることができる (図 3 における点線上の点になる)。特に、 $t = 0$ のとき $\hat{\Sigma}(t) = \hat{\Sigma}_{(k)}$ であり、 $t = t^*$ のとき $\hat{\Sigma}(t) = \hat{\Sigma}_{(k+1)}$ である。拡張ピタゴラスの定理 ([1], [2]) より、

$$\begin{aligned} D(\hat{\Sigma}_{(k)} | \tilde{\Sigma}_{(k)}^{-a'b'}(t)) &= D(\hat{\Sigma}_{(k)} | \tilde{\Sigma}_{(k)}^{-a^*b^*}(t)), \\ D(\tilde{\Sigma}_{(k)}^{-a'b'}(t) | \hat{\Sigma}(t)) &= D(\tilde{\Sigma}_{(k)}^{-a^*b^*}(t) | \hat{\Sigma}(t)) \end{aligned}$$

が成り立つ。このことから $\{\hat{\Sigma}(t) | 0 \leq t \leq t^*\}$ が角の二等分線に対応する曲線になっていることが分かる。

BRGGM アルゴリズムは以下の通りである。ステップ 2 から 6 が反復され、反復ごとにパラメータの推定値とモデルが出力される。独立グラフを $G = (V, E)$ とおき、 V は頂点集合を、 E は辺集合を表すものとする。 V は固定されており、辺集合 E の要素が推定される。ステップ 4 の $l_{(k)}^{-ab}$ は、 $\hat{\Sigma}_{(k)}$ と $\bar{\Sigma}_{(k)}^{-ab}$ を結ぶ m -測地線である。

BRGGM

入力： 標本平均 $\mu_{\text{sample}} = \hat{\mu}_{\text{MLE}}$ 、 標本共分散 $\Sigma_{\text{sample}} = \hat{\Sigma}_{\text{MLE}}$

出力： 推定値の系列 $\hat{\Sigma}_{(0)}^{-1}, \hat{\Sigma}_{(1)}^{-1}, \dots, \hat{\Sigma}_{(d)}^{-1}$ 、 独立グラフの系列 $\hat{G}_{(0)}, \hat{G}_{(1)}, \dots, \hat{G}_{(d)}$

1. $\hat{\Sigma}_{(0)} := \hat{\Sigma}_{\text{MLE}}$ 、 $\hat{G}_{(0)} := \hat{G}_{\text{MLE}}$ 、 $I := \{(a, b) | 1 \leq a < b \leq p\}$ 、 $k := 0$ とおく。

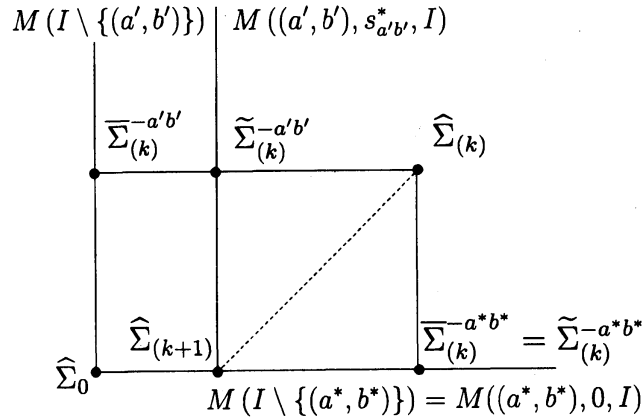


図3 推定量の更新. $I \subseteq \{(a, b) \mid 1 \leq a < b \leq p\}$: Σ^{-1} の非対角成分のうち推定されるべきものを指定するインデックス. $M((a, b), s, I) = \{\Sigma \mid \sigma^{ab} = s, \sigma^{a''b''} = 0 ((a'', b'') \notin I)\}$. $\hat{\Sigma}_0$: 独立モデルの最尤推定量. $\hat{\Sigma}_{(k)}$: k 番目の推定量. $\hat{\Sigma}_{(k+1)}$: $k+1$ 番目の推定量. $\bar{\Sigma}_{(k)}^{-ab}$: サブモデル $M(I \setminus \{(a, b)\}) = M((a, b), 0, I)$ における最尤推定量, あるいは $\hat{\Sigma}_{(k)}$ の $M(I \setminus \{(a, b)\})$ への m -射影でも同じ. $D(\hat{\Sigma}_{(k)} \mid \bar{\Sigma}_{(k)}^{-a'b'}) > D(\hat{\Sigma}_{(k)} \mid \bar{\Sigma}_{(k)}^{-a*b^*})$ である. $\tilde{\Sigma}_{(k)}^{-a'b'}$: $\hat{\Sigma}_{(k)}$ の $M((a', b'), s_{a'b'}^*, I)$ への m -射影. s_{ab}^* : 条件 $D(\hat{\Sigma}_{(k)} \mid \tilde{\Sigma}_{(k)}^{-a'b'}) = D(\hat{\Sigma}_{(k)} \mid \bar{\Sigma}_{(k)}^{-a*b^*})$ より決まる値. $\tilde{\Sigma}_{(k)}$ は $M(I \setminus \{(a', b')\})$ よりも $M(I \setminus \{(a^*, b^*)\})$ に近いので $\tilde{\Sigma}_{(k)}^{-a*b^*} = \bar{\Sigma}_{(k)}^{-a*b^*}$ である. $\hat{\Sigma}_{(k+1)}$ は $M((a^*, b^*), 0, I)$ と $M((a', b'), s_{a'b'}^*, I)$ の交点として定義される. 点線は角の二等分線に対応する曲線である. BRGGM アルゴリズムでは, 推定量がこの曲線に沿って動いていると見ることができる.

2. 各 $(a, b) \in I$ に対してモデル $M(I \setminus \{(a, b)\})$ の最尤推定量 $\bar{\Sigma}_{(k)}^{-ab}$ を計算する.
3. $t^* := \min_{(a,b) \in I} D(\hat{\Sigma}_{(k)} \mid \bar{\Sigma}_{(k)}^{-ab})$, $(a^*, b^*) := \arg \min_{(a,b) \in I} D(\hat{\Sigma}_{(k)} \mid \bar{\Sigma}_{(k)}^{-ab})$ とおく.
4. 各 $(a, b) \in I$ に対して, 二条件 $D(\hat{\Sigma}_{(k)} \mid \tilde{\Sigma}_{(k)}^{-ab}) = t^*$, $\tilde{\Sigma}_{(k)}^{-ab} \in M((a, b), s_{ab}^*, I)$ を満たすような s_{ab}^* と $\tilde{\Sigma}_{(k)}^{-ab} \in l_{(k)}^{-ab}$ を求める.
5. $(a, b) \in I$ に対して $\hat{\sigma}_{(k+1)}^{ab} := s_{ab}^*$, $(a, b) \notin I$ に対して $\hat{\sigma}_{(k+1)}^{ab} := 0$ とする. $\hat{E}_{(k+1)} := \hat{E}_{(k)} \setminus \{e_{a^*b^*}\}$ とする.
6. $k+1 < d-1$ の場合, $k := k+1, I := I \setminus \{(a^*, b^*)\}$ としてステップ2へ戻る. $k+1 = d-1$ の場合, ステップ7へ進む.
7. $\hat{\Sigma}_{(d)} := \hat{\Sigma}_0$, $\hat{E}_{(d)} := \emptyset$ とし, アルゴリズム終了.

ステップ5では, $(a^*, b^*) \in I$ であるが $\hat{\sigma}_{(k+1)}^{a^*b^*} = 0$ であることに注意が必要である. このことにより, 1回の反復において Σ^{-1} の非対角成分がひとつ0になり, モデルが入れ

子状に小さくなっていくことになる。

4.3 例

一般化線形回帰とガウシアングラフィカルモデルについての例を紹介する。計算には統計解析のフリーソフトウェア R を利用した ([12])。

4.3.1 一般化線形回帰の例

一般化線形回帰のひとつとしてロジスティック回帰を考える。ロジスティック回帰については 2.1 節で簡単に説明した。一般化線形回帰におけるリンク関数は $g(\mu) = \log\{\mu/(1-\mu)\}$ で与えられ、考える確率分布は二項分布のサブモデルである。ロジスティック回帰は、線形回帰問題に対する正則化の代表的な手法である lasso を形式的に適用することのできる問題である ([11])。しかし、lasso の適用は最適化問題として定式化することはできるものの、ロジスティック回帰問題の構造を取り入れたものではないことに注意が必要である。一方、bisector regression は確率分布のなす双対平坦空間の情報幾何に基づいており、問題の構造を利用して推定を行っている。一般化線形回帰に対する bisector regression アルゴリズムを BRGLM (Bisector Regression for Generalized Linear Models) と呼ぶことにする。

データとして South African Heart Disease (SAHD) データを使用した ([4])。正則化手法 lasso をロジスティック回帰に適用した文献 [11] においても SAHD データを解析している ([11], Fig. 2)。使用したデータは R の ElemStatLearn パッケージに含まれている。SAHD データは 9 個の説明変数 x_1, x_2, \dots, x_9 と反応変数 y からなり、 $n = 462$ 人分のデータが集められている。反応変数 y は chd である。説明変数はそれぞれ x_1 : sbp, x_2 : tobacco, x_3 : ldl, x_4 : adiposity, x_5 : family history, x_6 : typeA, x_7 : obesity, x_8 : alcohol, x_9 : age である。BRGLM により出力される推定値は 10 個であるが、説明変数の組合せは全部で $2^9 = 512$ 通りある。BRGLM により考えるモデル数が劇的に減少していることが分かる。

SAHD データに対する BRGLM の結果を図 4 に示した。横軸は BRGLM が出力した各推定値から原点までの KL ダイバージェンスの平方根を示している。原点については 4.1 節において説明した。縦軸は説明変数 x_i ($i = 1, 2, \dots, 9$) の回帰係数 θ^i ($i = 1, 2, \dots, 9$) の値を表している。BRGLM の出発点であるフルモデルの最尤推定値が図 4 の右端に対応する。BRGLM アルゴリズムは図の右端から左端に向かって進む。図に入っている縦線はそれぞれ BRGLM アルゴリズムが出力した推定値に対応している。左に進むにしたがって非零成分の個数がひとつずつ減っており、縦線の上の数字は非零成分の個数を示し

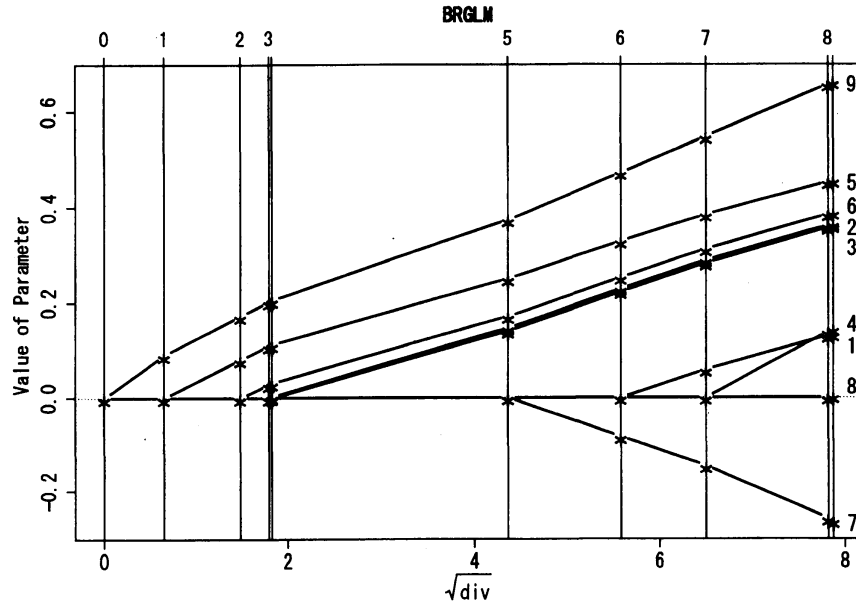


図4 SAHDデータのロジスティック回帰に対するBRGLMの結果。横軸は各推定値から原点までのKLダイバージェンスの平方根を示しており、ユークリッド距離に対応するものである。縦軸は説明変数 x_i ($i = 1, 2, \dots, 9$)の回帰係数 θ^i ($i = 1, 2, \dots, 9$)の値を表す。図の右端はフルモデルの最尤推定値であり、BRGLMアルゴリズムの出発点を表している。左端はBRGLMアルゴリズムの出力した最後の推定値であり、推定したいパラメータがすべて0になっている。図に入っている縦線はそれぞれBRGLMアルゴリズムが出力した推定値に対応している。左に進むにしたがって非零成分の個数がひとつずつ減っていく。縦線の上の数字は非零成分の個数を示している。パラメータが0になる順番は $\theta^8, \theta^4, \theta^1, \theta^7, \theta^3, \theta^2, \theta^6, \theta^5, \theta^9$ であった。

ている。図4より、 $\hat{\theta}$ の成分が0になった順番は $\theta^8, \theta^4, \theta^1, \theta^7, \theta^3, \theta^2, \theta^6, \theta^5, \theta^9$ である。

Park and Hastieの方法([11])をSAHDデータに適用した結果を図5に示す。Park and Hastieの方法はlassoを一般化線形回帰に拡張した正則化の方法である。Rの`glm`パッケージを利用して計算を行った。図5の横軸は推定値の l_1 -ノルムを、縦軸はパラメータ θ^i ($i = 1, 2, \dots, 9$)の値を示している。縦線の上の数字は非零成分の個数を示している。図5の右端から左端に見ていくと、パラメータが0になる順番は $\theta^8, \theta^4, \theta^7, \theta^1, \theta^6, \theta^3, \theta^2, \theta^5, \theta^9$ であったことが分かる。

BRGLMとPark and Hastieによる出力は似ているが、以下のような違いがある。BRGLMにより出力されるパラメータの推定値は単調性をもつことが分かる。Park and Hastieの方法は単調性をもたない。このことは θ^9 の推定値の変化から明らかである。また、BRGLMによる推定値は符号が変化しない。Park and Hastieの方法においては推定

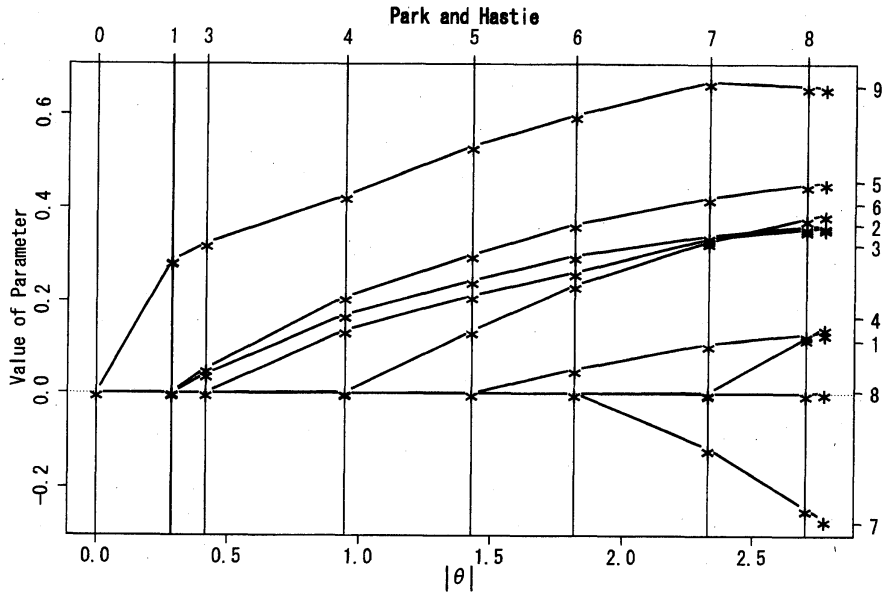


図5 SAHD データに対する Park and Hastie の方法の結果. 横軸は推定値の l_1 -ノルムを, 縦軸はパラメータ θ^i の値を示している. 縦線の上の数字は非零成分の個数を示している. 図の右端から左端に見ていくと, パラメータが 0 になる順番は $\theta^8, \theta^4, \theta^7, \theta^1, \theta^6, \theta^3, \theta^2, \theta^5, \theta^9$ であった.

値の符号は変化する.

上述の点はアルゴリズムの定義から分かる点であるが, 結果には次の違いもある. ふたつの手法で出力されるモデル (説明変数の組合せ) は異なっている. このことは, 例えば, 6 個の説明変数をもつモデルを比較すると分かる. BRGLM によるモデルには x_7 が含まれているが, Park and Hastie の方法によるモデルには x_7 が含まれていない. 実際, 図4と図5において6と書かれた縦線に対応する推定値を確認すると, BRGLM においては $\hat{\theta}_{(3)}^7 \neq 0$ であるが Park and Hastie の方法では $\hat{\theta}_{PH}^7 = 0$ となっていることが分かる.

4.3.2 ガウシアングラフィカルモデルの例

GGM の辺選択の例を紹介する. 精度行列の非対角成分を推定し, 対応する独立グラフを確認する. アルゴリズムの出力として, 独立グラフの系列を紹介する. 比較のため graphical lasso ([15]) と呼ばれる手法の出力も載せている.

データは数学成績データを利用した. R の SIN パッケージに含まれる mathmarks データである. このデータは 88 人の学生の 5 科目の試験成績を集めたものである. 5 科目とは, 1: 力学, 2: 線形代数, 3: 代数, 4: 解析, 5: 統計である. 辺の総数は $d = 5 \cdot 4 / 2 = 10$

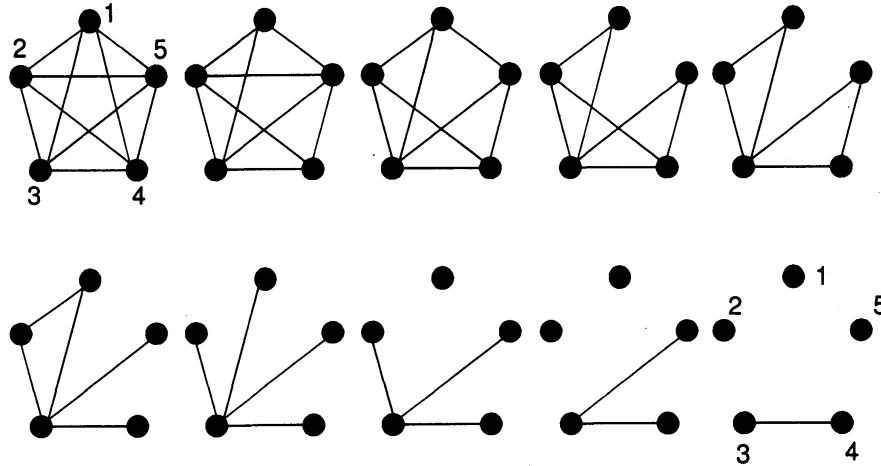


図6 数学成績データに対する BRGGM の結果. 5つの科目は, 1: 力学, 2: 線形代数, 3: 代数, 4: 解析, 5: 統計. BRGGM により出力される系列では, 左上から右下まで辺がひとつずつ少なくなっていく.

である.

数学成績データに対する BRGGM アルゴリズムの出力を図6に示す. 出力としては11個のグラフが得られるが, 最後のグラフは辺をもたないグラフであり図6では省略した. BRGGM アルゴリズムは左上から右下に進む. 同じデータに対する graphical LASSO の出力 (の一部) を図7に示す. ふたつの手法による出力は, 4本の辺をもつグラフが異なっておりまったく同じではないが, 似た系列になっている. 線形回帰の LARS は lasso と関連があることは第1節で述べた. この結果は, LARS の拡張を目指した bisector regression の lasso との関連を示唆するものであると考えられる.

5 まとめ

本稿では, least angle regression (LARS) を双対平坦空間の情報幾何にもとづいて拡張した結果 ([5], [6], [7]) についての解説を行った. 扱った問題は, 一般化線形回帰, ガウシアングラフィカルモデル, 分割表の3つであった. 推定手法である bisector regression アルゴリズムの説明は, 特にガウシアングラフィカルモデルの場合について行った. 推定手続きは, 局外パラメータの固定と bisector regression アルゴリズムによる推定の二段階であった. LARS では角の二等分線が重要な働きをしたが, bisector regression においても双対平坦空間において対応する曲線を考え, その曲線を利用して推定を行った. アルゴリズムの紹介とその幾何学的説明を行った. また, bisector regression アルゴリズムの適

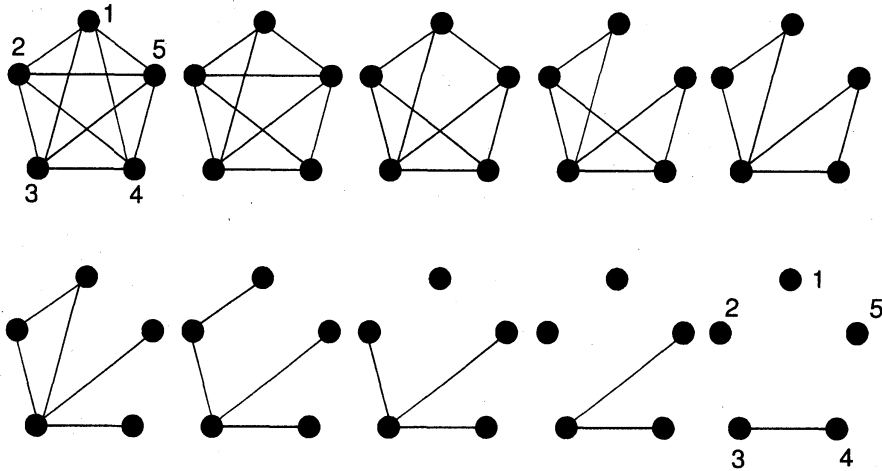


図7 数学成績データに対する graphical lasso の結果. 1: 力学, 2: 線形代数, 3: 代数, 4: 解析, 5: 統計. Lagrange 乗数が大きくなるにしたがって, graphical LASSO による出力は左上から右下に進む.

用例を紹介した.

参考文献

- [1] S. Amari (1985). *Differential-Geometrical Methods in Statistics*. Springer Lecture Notes in Statistics, vol. 28, Springer.
- [2] S. Amari and H. Nagaoka (2000). *Methods of Information Geometry*. Translations of Mathematical Monographs, Vol. 191, Oxford University Press.
- [3] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least Angle Regression (with discussion), *The Annals of Statistics*, vol. 32, 407–499.
- [4] T. Hastie, R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition. Springer, New York.
- [5] Y. Hirose (2012). Dimension Reduction Based on the Geometry of Dually Flat Spaces. Doctoral thesis, University of Tokyo.
- [6] Y. Hirose and F. Komaki (2010). An Extension of Least Angle Regression Based on the Information Geometry of Dually Flat Spaces, *Journal of Computational and Graphical Statistics*, vol. 19, 1007–1023.
- [7] Y. Hirose and F. Komaki (2012). Edge Selection Based on the Geometry of Dually

- Flat Spaces for Gaussian Graphical Models, *Statistics and Computing*, .
- [8] R. Kass and P. Vos (1997). *Geometrical Foundations of Asymptotic Inference*. John Wiley, New York.
 - [9] S. L. Lauritzen (1996). *Graphical Models*. Clarendon Press, Oxford.
 - [10] P. McCullagh and J. A. Nelder (1989). *Generalized Linear Models*, 2nd Edition. Chapman and Hall/CRC, Boca Raton.
 - [11] M. Y. Park and T. Hastie (2007). L_1 -Regularization Path Algorithm for Generalized Linear Models, *Journal of the Royal Statistical Society, Series B*, vol. 69, 659–677.
 - [12] R Development Core Team (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org/>.
 - [13] R. Tibshirani (1996). Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society, Series B*, vol. 58, 267–288.
 - [14] R. Tibshirani (2011). Regression Shrinkage and Selection via the Lasso: a Retrospective, *Journal of the Royal Statistical Society, Series B*, vol. 73, 273–282.
 - [15] M. Yuan and Y. Lin (2007). Model Selection and Estimation in the Gaussian Graphical Model. *Biometrika*, vol. 94, 19–35.