2I2-2

全状態探索による線形回帰のスパース変数選択

Exhaustive search for sparse variable selection in linear regression

五十嵐康彦 *1 Yasuhiko IGARASHI
竹中光 *1 Hikaru TAKENAKA
中西 (大野) 義典 *2 Yoshinori Nakanishi-Ohno
植村誠 *3 Makoto UEMURA
池田思朗 *4 Shiro IKEDA
岡田真人 *1 Masato OKADA

*1東京大学 大学院新領域創成科学研究科 Graduate School of Frontier Science, The University of Tokyo
*2東京大学大学院総合文化研究科 Graduate School of Arts and Sciences, The University of Tokyo
*3広島大学 宇宙科学センター Hiroshima Astrophysical Science Center, Hiroshima University
*4統計数理研究所 The Institute of Statistical Mathematicsy

We propose a K-sparse exhaustive search (ES-K) method and a K-sparse approximate exhaustive search method (AES-K) in which the optimal combination of explanatory variables is assumed to be K-sparse and the K-sparse combinations are exhaustively searched for sparse variable selection in linear regression. The density of states is thereby obtained and the solutions obtained by various approximate methods for sparse variable selection can be mapped to the obtained density of states. The ES-K method enables integration of previous sparse variable selection methods such as relaxation and sampling and evaluation of all approximate methods. In addition, for the problem of combinatorial explosion of the explanatory variables, the AES-K method enables effective reconstruction of the density of states by using the replica exchange Monte Carlo method and the multi-histogram method. In this study, we applyed the ES-K and AES-K methods to type Ia supernova data.

1. はじめに

変数選択を行うナイーブな手法は，その変数を使うか使わないかの全状態を網羅的に探索し，変数選択を行うことであるが [Ichikawa 14, Nagata 15]，この厳密なアルゴリズムの計算量は入力次元 N の指数オーダーとなってしまう [Cover 77]. これまで，指数オーダーの計算量を下げるために，スパース変数選択の問題を別の問題に変換して計算量を下げようとする，LASSO[Tibshirani 96] などの緩和法的な近似的アプローチや，マルコフ連鎖モンテカルロ (MCMC) 法や交換モンテカルロ (REMC) 法などのサンプリング手法による，サンプリングアプローチがとられてきた [George 93, Kim 06]. しかし，前者は，本来解くべき問題を別の問題に変換しているので，正しい解を求めている保証はない．また，後者は，指数オーダーの個数の状態の全貌を観察するという，サンプリング手法の優れた特性を利用していない.

そこで，最適な説明変数の組合せが K-スパースであると仮定し，その説明変数の組合せについて網羅的に探索する K-スパース全状態探索 (ES-K) 法を行う．本研究では，線形回帰の重みの事前分布として一様分布 [五十嵐ら 16] ではなく，データを用いて最適化を行い，解析的に各説明変数の組み合わせについてベイズ自由エネルギーを網羅的に求め，その状態密度を導出する．スパース変数選択の近似手法で得られた解をマッピングすることにより，近似的手法を評価する枠組みを提案する．本研究では，具体的に天文データへの適用結果について示す.

連絡先: 岡田真人，東京大学 大学院新領域創成科学研究科 〒 277-8561 千葉県柏市柏の葉 5-1-5，okada@k.u-tokyo.ac.jp

2. 手法

2.1 全状態探索 (Exhaustive-Search, ES) 法

本研究では，まず線形回帰問題における全状態探索 (Exhaustive-Search, ES) 法の導入を行う．目的変数 y_μ が N 個の説明変数 $\mathbf{x}_\mu = (x_{\mu 1}, x_{\mu 2}, \ldots, x_{\mu N})^{\mathrm{T}}$ によって生成され，その線形計測によって観測されていると仮定する．p 個のデータセット，$\mathbf{y} = (y_1, y_2, \ldots, y_p)^{\mathrm{T}}$ 及び $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p)^{\mathrm{T}}$，が与えられているとき，我々は $\mathbf{y} = \mathbf{1}\beta_0 + \mathbf{X}\boldsymbol{\beta}$ と書くことができる．ここで $\mathbf{1}$ は p 次元のすべての要素が 1 のベクトル，β_0 は切片をそれぞれ表し，$\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_N)^{\mathrm{T}}$ は，$\mathbf{X}$ の係数ベクトルを表す．また，本研究では各サンプルの観測ノイズの分散 σ_μ^2 $(\mu = 1, \ldots, p)$ が与えられているとする．$p \geq N$ の場合は，最小二乗法で $\boldsymbol{\beta}$ を決定できるが，$p < N$ であれば，解は一意に定まらず不定問題になる．本研究では，このような場合の線形問題を取り扱う.

ES 法において，変数の数が N の場合，β_i が 0 かそれ以外かの，$2^N - 1 = {}_N\mathrm{C}_1 + {}_N\mathrm{C}_2 + \cdots + {}_N\mathrm{C}_K \cdots + {}_N\mathrm{C}_N$ 個の全状態を網羅的に探索し，変数選択を行う [Ichikawa 14, Nagata 15, Igarashi 16]. これはスパース変数選択の厳密なアルゴリズムであるが，この計算量は入力次元 N の指数オーダーである [Cover 77].

$2^N - 1$ 通りの説明変数の全組合せを表すベクトルとして，以下で定義するインディケーター $\boldsymbol{c} = (c_1, c_2, \ldots, c_N) \in \{0, 1\}^N$. を用いて定式化を行う．ここで，$i$ 番目の説明変数が含まれると $c_i = 1$，含まれないと $c_i = 0$ となる．このインディケーター $\boldsymbol{c}$ を用いて，線形回帰問題を次のように定式化できる.

$$\mathbf{y} = \mathbf{1}\beta_0 + \mathbf{X}(\boldsymbol{c} \circ \boldsymbol{\beta}), \tag{1}$$

ここで $\circ$ はアダマール積であり，$(\boldsymbol{c} \circ \boldsymbol{\beta})_i = c_i \beta_i$ となる．この定式化はスパース線形回帰問題の本質をより明確に抽出できており，モデル化や予測によって最も良い説明変数の組み合わ

せ $\boldsymbol{c}$ は，自由エネルギー (FE) や交差検証誤差 (CVE) の最小化によって探索される．

2.1.1 自由エネルギー

FE は，ベイズ推論の枠組みによってモデル選択するための指標である．スパース線形回帰においては，FE は解析的に計算でき，その最小化も可能である．まず，線形回帰における FE を導出する．ガウスノイズ $\boldsymbol{\epsilon}$ が各観測データに付加していると仮定すると，線形回帰の問題は，次のように定式化できる，$\mathbf{y} = \mathbf{1}\beta_0 + \mathbf{X}(\boldsymbol{c} \circ \boldsymbol{\beta}) + \boldsymbol{\epsilon}$. 説明変数の組み合わせ $\boldsymbol{c}$ を選択するために，事後確率 $P(\boldsymbol{c}|\mathbf{y})$ の最も高い組み合わせを最適であると考える．ベイズの定理より，事後確率は各説明変数の事前確率が一定であるとすると，$P(\boldsymbol{c}|\mathbf{y}) = \frac{P(\mathbf{y}|\boldsymbol{c})P(\boldsymbol{c})}{P(\mathbf{y})} \propto P(\mathbf{y}|\boldsymbol{c})$ となる．したがって，事後確率が次に定義される周辺尤度と比例することがわかる，$P(\mathbf{y}|\boldsymbol{c}) = \int P(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{c})P(\boldsymbol{\beta}|\boldsymbol{c})d\boldsymbol{\beta}$. 負の対数尤度が FE と呼ばれ，$\mathrm{FE}(\boldsymbol{c}) \equiv -\log P(\mathbf{y}|\boldsymbol{c})$, と定義され，FE の最小化が事後確率最大化に対応する．説明変数の組み合わせ $\boldsymbol{c}$ を決めると，β_0 と $\boldsymbol{\beta}$ の事前分布が次のように導入する．$P(\beta_0) = \text{const.}$, $P(\beta_i|c_i = 1) = \frac{1}{\sqrt{2\pi s^2}} \exp(\frac{\beta_i^2}{2s^2})$, $P(\beta_i|c_i = 0) = \delta(\beta_i)$ $(i = 1, 2, \ldots, N)$, として $c_i = 1$ のとき，説明変数として用いる i 番目の変数の係数 β_i の事前分布に平均 0，分散 s のガウス分布を仮定する．また，尤度は $P(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{c}) = \frac{1}{\det(2\pi\Sigma)^{1/2}} \exp\left(-\frac{1}{2}\Delta^{\mathrm{T}}\Sigma^{-1}\Delta\right)$, と与えられる．ここで，$\Delta = [\mathbf{y} - \{\mathbf{1}\beta_0 + \mathbf{X}(\boldsymbol{c} \circ \boldsymbol{\beta})\}]$ とした．また，ノイズの共分散 $\Sigma_{ii} = \sigma_i^2$ $(i = 1, \ldots, p)$ and $\Sigma_{ij} = 0$ $(i \neq j)$ を既知とし，とすると，自由エネルギーは，次のように計算することできる．

$$FE(\boldsymbol{c}) = \frac{p}{2}\log(2\pi) + \frac{1}{2}\log\det(\Sigma^2) + K\log s + \frac{1}{2}\mathbf{y}^{-1}\Sigma^{-2}\mathbf{y} - \frac{1}{2}\boldsymbol{\mu}^{\mathrm{T}}\Lambda\boldsymbol{\mu} - \frac{1}{2}\det(\Lambda) \quad (2)$$

また，$\Lambda = (\boldsymbol{X}_{\mathrm{I}}^{\mathrm{T}}\Sigma^{-2}\boldsymbol{X}_{\mathrm{I}} + \frac{1}{s^2}\mathrm{I})^{-1}$，$\boldsymbol{\mu} = \Lambda\boldsymbol{X}_{\mathrm{I}}^{\mathrm{T}}\Sigma^{-2}\mathbf{y}$ とし，$\boldsymbol{X}_{\mathrm{I}}$ を $c_i = 1$ となる $\boldsymbol{X}$ の行をもつ行列とした．ES 法において，N 個の説明変数 $\boldsymbol{c}$ の全組合せ $2^N - 1$ 通りについて網羅的に FE を導出し，最小化することによって，最適な説明変数の組み合わせ $\boldsymbol{c}$ を導出する．

2.1.2 事前分布の推定

事前分布 $P(\beta_i|c_i = 1)$ における分散 s の推定を行う．$z = s^2$ とすると，自由エネルギーの z による微分は，$\frac{\partial FE(\boldsymbol{c},z)}{\partial z} = -\frac{K}{2z} + \frac{1}{2}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} + \frac{1}{2}\mathrm{Tr}(\Lambda)$ となり，$\frac{\partial FE(\boldsymbol{c},z)}{\partial z} = 0$ として固有値法と用いると，次のセルフコンシステント方程式が得られる．

$$z = K\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} + \sum_{k=1}^{K}\frac{1}{b_k + z}\right)^{-1} \quad (3)$$

ここで b_k は，$\boldsymbol{\mu}^{\mathrm{T}}\Lambda^{-1}\boldsymbol{\mu}$ の固有値を表す．式 (3) を反復して解くことにより，事前分布 $P(\beta_i|c_i = 1)$ における分散 $s(= \sqrt{z})$ の推定を行うことができる．

2.1.3 交差検証誤差 (Cross validation error, CVE)

説明変数の組み合わせの評価は，予測誤差の観点から CVE を用いても行うことができる．ここでは M-fold CV について説明する．まず，ランダムに p 個のデータのインデックスを M 個に分ける：$B_1, \ldots, B_M$. 次に，それぞれの m $(= 1, \ldots, M)$ に対して，トレーニングデータとして $y_\mu(\mu \notin B_m)$ を用いて係数 $\hat{\boldsymbol{\beta}}_m$ を推定する．最後に，評価データ y_μ $(\mu \in B_m)$ を用いて $\mathrm{CVE}(\boldsymbol{c}) = \frac{1}{M}\sum_{m=1}^{M}\mathrm{CVE}_m(\boldsymbol{c})$ を評価する．ここで，$\mathrm{CVE}_m(\boldsymbol{c}) = \frac{\sum_{\mu\in B_m}(y_\mu - \hat{y}_\mu)^2/\sigma_\mu^2}{\sum_{\mu\in B_m}1/\sigma_\mu^2}$，$\hat{y}_\mu = \mathbf{x}_\mu^{\mathrm{T}}(\boldsymbol{c} \circ \hat{\boldsymbol{\beta}}_m)$ である．

2.2 K-スパース全状態探索 (ES-K) 法

ES 法は，説明変数の次元 N に対して，指数関数的に計算量 $O(e^N)$ となり，説明変数の次元 N が多い場合には，組合せ爆発が生じて全状態探索を行うことが困難になる．そこで，最適な説明変数の組合せが K スパース，すなわち非ゼロ要素の数が K 個であると仮定して，その説明変数の組合せを網羅的に探索し，FE や CVE を計算することによって K-スパースな説明変数の組み合わせを評価する，K-スパース全状態探索 (ES-K) 法を提案する [五十嵐ら 16] .

2.3 近似的全状態探索 (Approximate Exhaustive Search, AES) 法

前節の K-スパース全状態探索法でも，次元 N が大きな場合は次元数 K に対して $O({}_N\mathrm{C}_K)$ の膨大な計算量が必要となる．そこで本研究では，マルコフ連鎖モンテカルロ法 [Metropolis 53] の一種である交換モンテカルロ法 (REMC) [Hukushima 96] に着目し，次元数 K を固定した場合の BIC，CVE の最小解を効率的に導出 [George 93, Kim 06] するだけでなく，それぞれの状態密度をマルチヒストグラム法 [Ferrenberg 89] によって再構成する．本手法を，K-スパース近似的全状態探索 (Approximate Exhaustive Search-K, AES-K) 法と呼ぶ．

2.3.1 交換モンテカルロ法

交換モンテカルロ法の目的は，ボルツマン分布，$P_\omega(\boldsymbol{c}|T_\omega) = \frac{1}{Z_\omega}\exp\left(-\frac{E(\boldsymbol{c})}{T_\omega}\right)$, から FE や CVE によって記述されるエネルギー E をもつ説明変数の組み合わせ $\boldsymbol{c}$ を効率的にサンプリングすることである．ここで，$T_\omega > 0$ は温度パラメータであり，Z_ω は規格化因子である．交換モンテカルロ法において，我々はいくつかの温度 $0 < T_1 < \cdots < T_\omega < \cdots < T_\Omega$ をもつボルツマン分布 P_ω のレプリカを準備する．交換モンテカルロ法は，次の同時確率分布から説明変数の組み合わせ $\boldsymbol{c}$ のサンプリングするために用いられる．$P(\boldsymbol{c}_1, ..., \boldsymbol{c}_\Omega) = \prod_{\omega=1}^{\Omega}P_\omega(\boldsymbol{c}_\omega|T_\omega)$. 具体的には，以下の 2 種類の状態遷移を交互に行うことによってサンプリングされる．

1. 各温度において，メトロポリス法 [Metropolis 53] によって $P_\omega(\boldsymbol{c}|T_\omega)$ からのサンプリングを並列に行う．説明変数の数は K 個に固定してサンプリングする [Nakanishi 16].

2. サンプル $\boldsymbol{c}_\omega$ と $\boldsymbol{c}_{\omega+1}$ を確率 $\min\{1, r'\}$ で交換する．ここで r' は以下のように与えられる．

$$r' = \frac{P_\omega(\boldsymbol{c}_{\omega+1}|T_\omega)P_{\omega+1}(\boldsymbol{c}_\omega|T_{\omega+1})}{P_\omega(\boldsymbol{c}_\omega|T_\omega)P_{\omega+1}(\boldsymbol{c}_{\omega+1}|T_{\omega+1})} = \exp\{(1/T_{\omega+1} - 1/T_\omega)[E(\boldsymbol{c}_{\omega+1}) - E(\boldsymbol{c}_\omega)]\} \quad (4)$$

これらの二つのステップを十分に反復して，説明変数の組み合わせ $\boldsymbol{c}$ の分布が同時確率分布 $\prod_{\omega=1}^{\Omega}P_\omega(\boldsymbol{c}_\omega|T_\omega)$ に収束する．以上のような温度交換を伴うアルゴリズムにより，サンプリングの緩和を促進するだけでなく，局所安定状態から脱却しやすくなるという特徴がある．

2.3.2 マルチヒストグラム法

マルチヒストグラム法 [Ferrenberg 89] では，交換モンテカルロ法によって得られた各温度のサンプルを組合せることによって，FE や CVE の全状態密度を近似的に求めることができる．交換モンテカルロ法により，各温度 T_ω ごとにエネルギー E をとる状態が $H_\omega(E)$ 個のヒストグラムとして得られたとすると，状態密度 $g(E)$ は $g(E) = \frac{\sum_{\omega=1}^{\Omega}H_\omega(E)}{\sum_{\omega=1}^{\Omega}n_\omega\exp(f_\omega - E/T_\omega)}$, によって与えられる．ここで n_ω は温度 T_ω における総サン

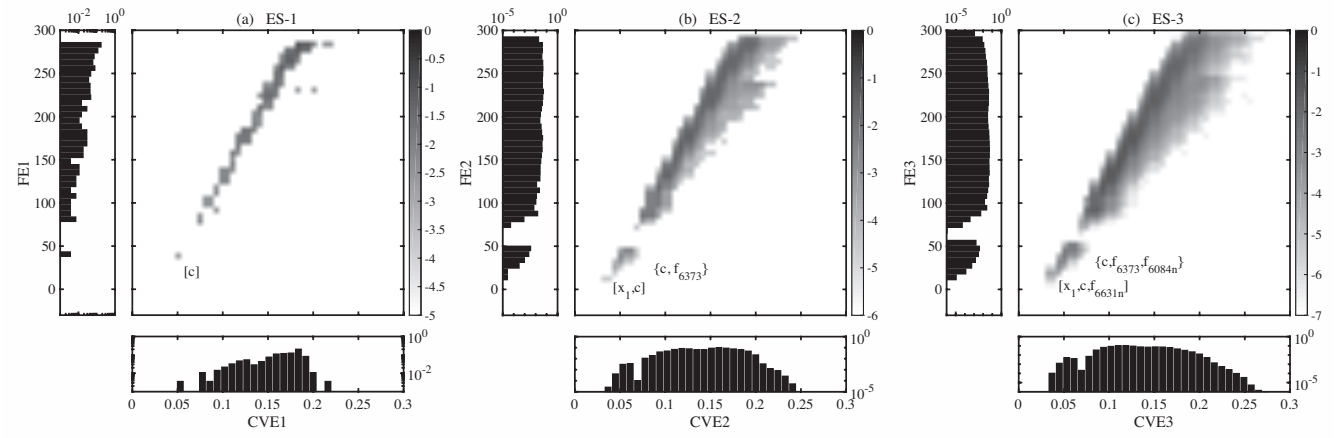

図 1: 天文データに対する ES-1,2,3 法の状態密度．各データの横軸，縦軸は，それぞれ CVE と BIC の状態密度の常用対数を表し，中央の図は CVE と BIC の 2 次元状態密度の常用対数を表す．[·]，{·} で囲まれた文字は，それぞれ ES-K 法と λ-スキャン法で得られた説明変数を表す．

プル数であり，f_ω は自由エネルギーと呼ばれる．また，$f_\omega = -\log\sum_E g(E)\exp(-E/T_\omega)$. となる．この二式を交互に繰り返し解くことによって全状態密度 $g(E)$ を近似的に再構成することができる．

2.4 LASSO

$\boldsymbol{\beta}$ が十分にスパースであるとき，緩和アプローチの代表的な手法の一つである $L1$ 正則化による LASSO(Least Absolute Shrinkage and Selection Operator) が線形回帰におけるスパース変数選択にうまく働くことが知られている [Tibshirani 96]. LASSO によって得られた解を ES 法による解と比較する事を考える．$L1$ 正則化による LASSO 回帰では，最適な説明変数の係数 $\hat{\boldsymbol{\beta}}$ が，以下のように $\boldsymbol{\beta}$ の $L1$ ノルムを正則化項として含めた関数を最小化するように決定される．

$$\hat{\boldsymbol{\beta}}(\lambda) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\Sigma^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \lambda||\boldsymbol{\beta}||_1 \right\}, \quad (5)$$

ここで $|\cdot|_1$ は $L1$-ノルムと呼ばれ，$\|\boldsymbol{\beta}\|_1 = \sum_i |\beta_i|$ と定義される．また，係数 λ は正則化パラメータと呼ばれる．LASSO によって得られている解を評価しやすくするため，λ が与えられている元での LASSO によって選ばれた説明変数の組み合わせを，インディケーター $\boldsymbol{c}$ を用いて，$\boldsymbol{c}(\lambda)$: $c_i(\lambda) = 1\ \hat{\beta}_i(\lambda) \neq 0$ とする．このとき，正則化パラメータ λ の決め方について二つの方法を次に説明する．

2.4.1 λ-最適化法

一般的には，正則化パラメータ λ は，CVE を用いて最適化される．本研究では，Subsubsection 2.1.3 を CVE として用いた．λ の最適化に際しては，CVE(λ) の最小にする $\lambda_{\min}$ があるが，選ばれる変数の数が多くなる傾向があるため，ヒューリスティックな方法として"one-standard-error rule"(1SE) がある．この方法は，$\lambda_{\min}$ から標準偏差分だけ CVE が大きいモデルは許容し，その中で最大の正則化パラメータ $\lambda_{1\mathrm{SE}}$ を，最適なパラメータとして決めることで，よりスパースな 変数を抽出することができる．

2.4.2 λ-スキャン法

λ-スキャン法は ES 法に着想を得ている．λ を最適化する代わりに，λ-スキャン法では λ の値によらず LASSO が与える説明変数の組み合わせ $\boldsymbol{c}(\lambda)$ を網羅的に探索する手法である．我々は，すべての $\boldsymbol{c}(\lambda)$ に対して FE や CVE を計算し，どの説明変数の組み合わせが最適化を評価する．FE($\boldsymbol{c}(\lambda)$) や CVE($\boldsymbol{c}(\lambda)$) は ES 法と同様に計算する．LASSO は，ES 法で探索する，説明変数の組み合わせを削減させる役割となっている．

3. 実データとその解析結果

本章では，ES-K 法，AES-K 法によるスパース解の網羅的探索法を，Berkeley Supernova Database から得た Ia 型超新星の極大等級の説明変数抽出に適用する [Silverman 12]．データサンプルとして，$p = 78$ 個の天体について考える [Uemura 15]．説明変数として，光度曲線の幅 (x_1)，色 (c)，3500−8500Å の総フラックスで規格化したスペクトル (134 個の説明変数)，連続光レベルで規格化したスペクトル (134 個の説明変数)，先行研究で提案されてきたフラックス比 (6 個の説明変数)[Silverman 12] の計 $N = 276$ 個を用いた．

3.1 Results of ES-K and AES-K methods

まず，天文データにおける ES-1,2,3 法の状態密度を図 1 に示す．交差検証では $M = 10$ として 10 分割交差検証を行った．まず，$K = 1$ の場合，最も FE 及び CVE が低い変数は (c) となった．次に，$K = 2$ の場合，従来から慣用的に用いられている変数の組み合わせ，光度曲線の幅 (x_1) と色 (c) をもつ解の組合せによる FE と CVE が 2 変数の組合せの中で低いことが示された．$K = 3$ において，光度曲線の幅 (x_1) と色 (c) をもつ解の組合せが，他の変数の組合せに比べて BIC，CVE ともに小さくなる傾向にあり，この 2 変数を中心にして上位のモデルがクラスターを形成していることが確認された (図 1(c))．

これまで，ES-1,2,3 法の結果について述べた．FE や CVE が最も低くなる説明変数の次元数を求めるため，$K = 1, 2, \ldots, 7$ までの説明変数の次元数に対して，FE 及び CVE の最小値を求め，各次元での最小値を比べた (図 2)．$K = 1, 2, 3, 4$ まで ES-K 法により計算し，計算量の多い $K = 4, 5, 6, 7$ については交換モンテカルロ法による AES-K 法を適用し，サンプリング数は 100,000 とした．$K = 1, \ldots, 4$ について ES-K 法を行った結果と AES-K 法との結果が一致することを確認している．この結果，次元数に対して，CVE は単調減少し，$K = 7$ のときに最小となった．また，FE は $K = 6$ のときに最小となった．この結果から，データから非ゼロ要素の個数 K を推

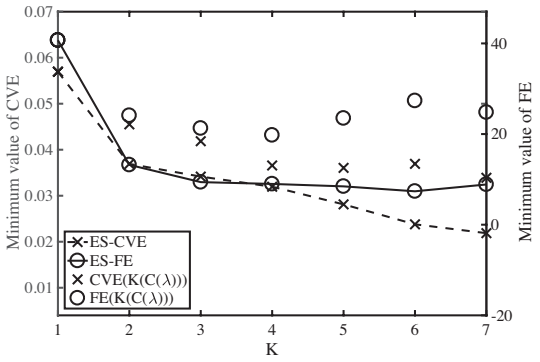

図 2: 天文データにおける説明変数の次元数 K に対する BIC, CVE の最小値．$K = 1, 2, \ldots, 4$ まで ES-K 法，$K = 5, 6, 7$ については AES-K 法を適用した．

定しようとすると，ES-K 法では，天文学において慣用的に用いられてきた $K = 2$ の説明変数の組み合わせではない組み合わせが抽出されることがわかる．FE については説明変数の係数に対する事前分布 $P(\beta_i|c_i = 1)$ を一様と仮定した先行研究に比べて [五十嵐ら 16]，事前分布をガウス分布として最適化した場合，次元数が大きくなることにより FE が大きくなるため，よりスパースな説明変数の選択を行えることを確認した．ただ，その場合においても，先行研究 [五十嵐ら 16] と同様に，$K \geq 2$ の説明変数の組み合わせが抽出されることがわかる．

3.2 LASSO 解の評価

この章では，実データ解析に用いた LASSO 解の評価を行う．先行研究では，1SE ルールによる λ-最適化法を用いることで，6 つの説明変数が抽出され，天文学において慣用的に用いられている説明変数の組み合わせとは違った組み合わせが抽出されることが報告されている [Uemura 15]．我々は λ-スキャン法を同じ実データに対して適用した．ES-K 法と λ-スキャン法を比較するために，すべての $\boldsymbol{c}(\lambda)$ を，それぞれ K 個の非ゼロ要素をもつグループにクラス分けし，各グループ内で最小の FE，CVE をそれぞれ FE($\boldsymbol{c}(\lambda_K)$)，CVE($\boldsymbol{c}(\lambda_K)$) と記述し，図 2 に $K = 1, 2, \ldots, 7$ までプロットした．その結果，$K \geq 2$ において ES-K 法は，図. 2 で示すように，FE や CVE どちらにおいても λ-スキャン法に比べて性能が優れていた．このことから，LASSO は，ES-K や AES-K 法で得られる最小の FE や CVE をもつ K-スパースな説明変数を抽出することが出来ていないことが分かる．特に，図 1(a)-(c) のように，LASSO で得られた解を ES-K 法で得られた状態密度にマッピングすると，$K = 2$ では，天文学において重要だと考えられている説明変数の組み合わせ $\{x_1, c\}$ が ES-K 法では抽出されているが LASSO では抽出されていないことが分かる．このように，様々な解法によって得られた解の関係を理解するのに，FE や CVE の状態密度を知ることが重要である．また，ES-K 法はスパース変数選択において LASSO よりも有効であることが分かる．

4. 議論とまとめ

本研究では，ES 法と AES 法を線形回帰問題のスパース変数選択に用いて，全状態を網羅的に探索する ES-K 法と AES-K 法を行う枠組みを提案した．この枠組みは，Nagata らと同様に，スパース変数選択の緩和アプローチとサンプリングアプローチを統合することが可能であり [Nagata 15, Igarashi 16]，すべての近似的手法を評価できる．提案手法を，Ia 型超新星の極大等級データ [Silverman 12, Uemura 15] に適用した結果，先行研究 [Uemura 15] とは，異なる結果が得られた．先行研究では CVE を評価することで，a 型超新星の極大等級データの説明変数 (x_1, c) であると結論している．しかしながら，我々の手法では，CVE だけでなく BIC においても，(x_1, c) の組み合わせより，性能の高い組み合わせが存在した．これは先行研究で用いられた近似的探索手法が不完全であること意味する．

今後は，今回の実データ解析で得られた結果が，どのようなデータ構造によって得られたかを議論するために，実データの生成モデルを考え，ここから人工的に生成した仮想データに同様の手法を適用して解析する．この枠組みを VMA(Virtual Measurement and Analysis, 仮想計測解析) と呼び [五十嵐ら 16]，実データ解析の結果と比較することにより，実データ解析の結果を間接的に評価を行う．

謝辞

論文中で扱ったデータはカリフォルニア大学バークレー校の超新星研究グループが公開しているものを使用した [Silverman 12]．また，本研究は JSPS 科研費 新学術領域研究 (JP25120001, JP25120007, JP25120008, JP25120009) の助成を受けた．

参考文献

[Ichikawa 14] Ichikawa, H., Kitazono, J., Nagata, K., Manda, A., Shimamura, K., Sakuta, R., Okada, M., Yamaguchi, K. Y., Kanazawa, S., and Kakigi, R. (2014), Front. Human Neurosci., 8, 480.

[Nagata 15] Nagata, K., Kitazono, J., Nakajima, S., Eifuku, S., Tamura, R., and Okada, M. (2015), IPSJ Trans., 8, 25-32.

[Igarashi 16] Igarashi, Y., Nagata, K., Kuwatani, T., Omori, T., Nakanishi-Ohno, Y., and Okada, M. (2016), J. Phys. Conf. Ser. 699(1), 012001.

[五十嵐ら 16] 五十嵐康彦，竹中光，中西（大野）義典，植村誠，池田思朗，岡田真人 (2016), 信学技報, 116(300), 313-320.

[Cover 77] Cover, T. M., and Van Campenhout, J. M. (1977), IEEE Trans. Sys., Cyber., 7(9), 657-661.

[Tibshirani 96] Tibshirani, R. (1996), J. Royal Stat. Soc. Ser. B, 267-288.

[Metropolis 53] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), J. Chem. Phys., 21(6), 1087-1092.

[George 93] George, E. I., and McCulloch, R. E. (1993), Journal of the American Statistical Association, 88(423), 881-889.

[Hukushima 96] Hukushima, K., and Nemoto, K. (1996), J. Phys. Soc. Japan, 65(6), 1604-1608.

[Kim 06] Kim, S., Tadesse, M. G., and Vannucci, M. (2006), Biometrika, 93(4), 877-893.

[Ferrenberg 89] Ferrenberg, A. M., and Swendsen, R. H. (1989), Phys. Rev. Lett., 63(12), 1195.

[Silverman 12] Silverman, J. M., Ganeshalingam, M., Li, W., and Filippenko, A. V. (2012), Monthly Notices of the Royal Astronomical Society, 425(3), 1889-1916.

[Uemura 15] Uemura, M., Kawabata, K. S., Ikeda, S., and Maeda, K. (2015), Pub. Astro. Soc. of Japan, 67(3), 55.

[Nakanishi 16] Nakanishi-Ohno, Y., Obuchi, T., Okada, M., and Kabashima, Y. (2016), J. Stat. Mech.: Theory and Experiment, 2016(6), 063302.