

令和 6 年 6 月 7 日現在

機関番号：15401

研究種目：若手研究

研究期間：2020～2023

課題番号：20K14363

研究課題名（和文）説明変数・目的変数が高次元でも変数増減法の下で一致性をもつ変数選択規準の開発

研究課題名（英文）Development of consistent variable selection criteria with high-dimensional response and explanatory variables using forward-backward stepwise selection method

研究代表者

小田 凌也（Oda, Ryoya）

広島大学・先進理工系科学研究科（理）・助教

研究者番号：10853682

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：本研究の目的は、多変量モデルにおいて、変数の次元数が標本数を超えた場合も含んでいる高次元大標本データに対して良い性質をもつ変数選択規準を構築することである。特に、変数の個数が標本数を超えてもよい高次元下で、真の変数を選択する確率が漸近的に1となる性質である一致性をもつ規準を構築する。目的を達成するために、多変量モデルにおいて、標本数 $n$ は無限大だが変数の個数は無限大としてよい漸近理論により一致性を評価し、一致性をもつ変数選択法を提案した。

研究成果の学術的意義や社会的意義

近年では高次元データの使用は頻繁にされるため、そのような高次元データに対する統計分析手法の開発は重要である。本研究により提案された変数選択手法は高次元データに対しても良い性質をもちかつ計算も高速であるため、リーズナブルな手法であると考えられる。

研究成果の概要（英文）：The aim of this study is to propose a variable selection criterion with good properties for high-dimensional and large-sample data in multivariate models, including cases where the number of dimensions of the variables exceeds the sample size. In particular, we propose a criterion with consistency, which is the property that the probability of selecting the true variable is asymptotically one under a high-dimensional setting where the number of variables may exceed the sample size. In order to achieve the aim, we evaluate the consistency in multivariate models under the high-dimensional asymptotic theory, in which the sample size is infinite but the number of variables may be infinite, and propose a variable selection method with consistency.

研究分野：数理統計

キーワード：多変量モデル モデル選択 一致性 高次元

## 様式 C-19、F-19-1 (共通)

### 1. 研究開始当初の背景

複数の予測対象である変数(目的変数と呼ぶ)とそれらに影響を与えると考えられる変数(説明変数と呼ぶ)の関係を記述する多変量線形回帰モデルは、多変量モデルの中で最も基本となるモデルの1つである。このモデルにおいて、目的変数に影響を与えていない不要な説明変数を取り除き影響を与えている説明変数を特定することは、最適な説明変数の組み合わせを選択する変数選択問題として捉えられる。さらに、変数選択を実装する際、本当に目的変数に影響を与えている説明変数(真の説明変数と呼ぶ)をできるだけ高い確率で選択することがしばしば望まれる。この要望に応えるには、真の説明変数の組み合わせを選択する確率が漸的に1となる性質(一致性と呼ぶ)をもつ変数選択規準を用いればよい。なぜなら、その変数選択法は真の説明変数の組み合わせを選択する確率(選択確率と呼ぶ)が高くなる傾向をもつからである。一方、近年のハードウェアの発展に伴い、変数の個数が多い高次元データを解析する需要が高まっている。多変量線形回帰モデルにおいて、高次元データに対しても有効とされている変数選択法として以下のようなKick-one-out法(Bai *et al.*, 2018)やAdaptive elastic-net型罰則付き推定法(Xin *et al.*, 2017)が挙げられる。

I. Kick-one-out 法: すべての説明変数とすべての変数から1つ変数を除いた組み合わせに対して変数選択規準を計算し、その差の正負で変数選択を行う。

■ (長所) 標本数  $n$  だけでなく説明変数の個数  $k$  と目的変数の個数  $p$  を無限大としたときに一致性をもつので、高次元データに適した変数選択を実行できる。

■ (短所)  $k$  と  $p$  の両方ともに  $n$  未満でないとは実行できない。

II. Adaptive elastic-net 型罰則付き推定法: 重み付きグループ型罰則項を用いたスパース推定による変数選択を行う。

■ (長所)  $k$  と  $p$  の両方が  $n$  を超えても実行できる。

■ (短所)  $p$  を無限大としたとき一致性の保証がなく、使用されている重みが適切でない可能性がある。

しかし方法 I は  $k$  もしくは  $p$  が  $n$  を超えると実行できず、方法 II は  $p$  が無限大のときで的一致性の保証に問題を抱える。

### 2. 研究の目的

本研究は上記の問題を解決するような変数選択法を開発することを目的とする。特に、標本数  $n$  は無限大、変数の個数は  $n$  を超えて無限大でもよい高次元漸近理論の下で一致性をもつ新たな変数選択規準を提案することにある。

### 3. 研究の方法

研究当初の予定では変数選択法として変数増減法の利用を考えていたが、Kick-one-out 法でもスクリーニング法を組み合わせることで上記の問題を解決可能なことを着想したため、Kick-one-out 法を用いた変数選択法に関する研究を展開する。具体的には、主に多変量線形回帰モデルにおける説明変数の選択問題を扱い、一致性を評価する変数選択規準として一般化情報量規準とリッジ型一般化  $C_p$  規準を用いる。一般化情報量規準の一致性の評価には、 $n$  は無限大だが  $p$  と  $k$  の和は  $n$  未満のもとで無限大でもよい高次元漸近理論を用いる。リッジ型一般化  $C_p$  規準の一致性の評価には、 $n$  は無限大だが  $p$  は  $n$  を超えて無限大でもよく  $k$  は  $n$  未満で無限大としてよい高次元漸近理論を用いる。以上の漸近理論のもとで一致性をもつための十分条件を導出し、その条件を満たすような新たな変数選択法を提案する。

### 4. 研究成果

#### 【一般化情報量規準の一致性】

変数の個数が標本数を超えない場合に高速に実行可能な変数選択法である Kick-one-out 法の下で一般化情報量規準が一致性をもつための条件を調べた。具体的には、真の分布は多変量正規分布に従うとし、標本数は無限大であるが、目的変数と説明変数の個数の和は標本数を超えない下で無限大でもそうでなくてもよい漸近理論により一般化情報量規準が一致性をもつための罰則項の条件を導出した。さらに、導出した条件を利用することで、標本数さえ大きく目的変数と説明変数の個数の和が標本数未満であれば、目的変数と説明変数が大きくても Kick-one-

out 法の下で真の変数を選択する確率が高い変数選択法を提案した。(雑誌論文: Oda & Yanagihara (2021))

#### 【リッジ型一般化 $C_p$ 規準の一致性】

目的変数の個数が標本数を超えてもよい高次元下を想定し, そのような状況に対応できるよう共分散行列の推定量をリッジ型に置き換えたリッジ型一般化  $C_p$  規準を用いた. このとき, 変数選択の性能を決めるパラメータは罰則の強さを表すパラメータ  $\alpha$  とリッジパラメータ  $\lambda$  の 2 つがある. 目的変数の個数が標本数に比べ圧倒的に速く無限大に発散してもよいだけでなく説明変数の個数も標本数未満だが無限大に発散してもよい漸近理論を用いて一致性をもつための  $\alpha$  と  $\lambda$  に関する十分条件を導出し, その条件を満たす変数選択法を提案した. これにより, 目的変数の個数が標本数を超えた場合でも高い確率で真の変数を選択することが可能な変数選択法を提案した.(雑誌論文: Oda (2023))

#### 【目的変数の選択問題における一般化情報量規準の一致性】

多変量線形回帰モデルにおける目的変数の選択問題に着目した. この選択問題は一部分の目的変数を条件付きしたときの残りの目的変数の分布が説明変数に依存しないときに条件付きした目的変数は冗長であるという設定であり, これは未知の説明変数を基準化された残差平方和最小化により推定する際の目的変数の選択問題と同等である. 本研究では標本数は無限大であるが, 目的変数と説明変数の個数の和は標本数を超えない下で無限大でもそうでなくてもよい漸近理論により一般化情報量規準が一致性をもつための罰則項の条件を導出し, その条件を満たす変数選択法を提案した.

#### 【一般化情報量規準の漸近有効性】

一致性は変数選択の重要な性質の 1 つであるが, 予測の観点からの性質である漸近損失有効性(ロス関数を漸近的に最小にするモデルを選択する確率が 1 に収束する性質)も重要な性質の 1 つである. 本研究では多変量線形回帰モデルにおける説明変数の選択問題に対して, 一般化情報量規準が漸近損失有効性をもつための条件を導出した. その際,  $n$  は無限大だが  $p$  は  $n$  未満のもとで無限大でもよい高次元漸近理論を用いた. その結果,  $n$  のみを無限大とする場合及び  $p/n$  が 0 に収束する場合には一致性と漸近損失有効性の両方を持ち,  $p/n$  が 1 未満の正数に収束する場合には漸近損失有効性をもつような変数選択規準を提案した.(学会発表: 小田・柳原(2021), Oda (2022))

#### 【GMANOVA モデルにおける基底関数の選択問題に対する一般化 $C_p$ 規準の一致性及び漸近有効性】

他の多変量モデルへの拡張として, 一般化多変量分散分析 (GMANOVA) モデルにおける基底関数の選択問題に対して変数選択規準の一致性及び漸近損失有効性を扱った. 特に, 目的変数が標本数を超えても使用可能であるように二乗ロス関数を基にした一般化  $C_p$  規準を使用した. その際, 標本数  $n$  は無限大だが目的変数の個数  $p$  は標本数を超えて無限大としてよい高次元漸近理論により一致性を評価し, 一致性をもつ一般化  $C_p$  規準が二乗ロス関数に対する漸近損失有効性をもつことを示した. また, 候補モデルの数も発散する高次元漸近理論を用いたとき一致性をもつための十分条件を導出し, そのような場合でも一致性をもつ変数選択規準を提案した.(学会発表: 小田 (2022, 2023, 2024))

#### 【discrete first-order algorithm の拡張】

多変量モデルにおける変数選択では, グループ構造の選択に焦点が当てられることが多い.  $\ell_0$  罰則を伴うパラメータ推定アルゴリズムの 1 つである discrete first-order algorithm (Bertsimas *et al.*, 2016) をグループ構造のある変数選択が可能となるよう行列に関する  $\ell_{2,0}$  罰則を伴うパラメータ推定アルゴリズムに拡張しそのアルゴリズムの収束理論を調べた. この方法と変数選択規準を組み合わせることで変数の個数が標本数を超えても変数選択可能なアルゴリズムを提案した.(雑誌論文: Oda *et al.* (2023))

## 5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 7件/うち国際共著 1件/うちオープンアクセス 2件）

1. 著者名 Oda Ryoya, Ohishi Mineaki, Suzuki Yuya, Yanagihara Hirokazu	4. 巻 53
2. 論文標題 An $l_{2,0}$ -norm constrained matrix optimization via extended discrete first-order algorithms	5. 発行年 2023年
3. 雑誌名 Hiroshima Mathematical Journal	6. 最初と最後の頁 251-267
掲載論文のDOI（デジタルオブジェクト識別子） 10.32917/h2021058	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Oda Ryoya	4. 巻 352
2. 論文標題 Kick-One-Out-Based Variable Selection Method Using Ridge-Type Cp Criterion in High-Dimensional Multi-response Linear Regression Models	5. 発行年 2023年
3. 雑誌名 Intelligent Decision Technologies, Smart Innovation, Systems and Technologies	6. 最初と最後の頁 193-202
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-981-99-2969-6_17	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Oda Ryoya, Yanagihara Hirokazu	4. 巻 238
2. 論文標題 A Consistent Likelihood-Based Variable Selection Method in Normal Multivariate Linear Regression	5. 発行年 2021年
3. 雑誌名 Intelligent Decision Technologies, Smart Innovation, Systems and Technologies	6. 最初と最後の頁 391 ~ 401
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-981-16-2765-1_33	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Yanagihara Hirokazu, Oda Ryoya	4. 巻 238
2. 論文標題 Coordinate Descent Algorithm for Normal-Likelihood-Based Group Lasso in Multivariate Linear Regression	5. 発行年 2021年
3. 雑誌名 Intelligent Decision Technologies, Smart Innovation, Systems and Technologies	6. 最初と最後の頁 429 ~ 439
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-981-16-2765-1_36	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Oda Ryoya, Yanagihara Hirokazu, Fujikoshi Yasunori	4. 巻 21-07
2. 論文標題 On model selection consistency using a kick-one-out method for selecting response variables in high-dimensional multivariate linear regression	5. 発行年 2021年
3. 雑誌名 Hiroshima Statistical Research Group, Technical Report	6. 最初と最後の頁 1~15
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Oda Ryoya	4. 巻 50
2. 論文標題 Consistent variable selection criteria in multivariate linear regression even when dimension exceeds sample size	5. 発行年 2020年
3. 雑誌名 Hiroshima Mathematical Journal	6. 最初と最後の頁 339-374
掲載論文のDOI (デジタルオブジェクト識別子) 10.32917/hmj/1607396493	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Oda Ryoya, Yanagihara Hirokazu	4. 巻 238
2. 論文標題 A Consistent Likelihood-Based Variable Selection Method in Normal Multivariate Linear Regression	5. 発行年 2021年
3. 雑誌名 Intelligent Decision Technologies, Smart Innovation, Systems and Technologies	6. 最初と最後の頁 391-401
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-981-16-2765-1_33	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Shinpei Imori, Dietrich von Rosen, Ryoya Oda	4. 巻 84
2. 論文標題 Growth Curve Model with Bilinear Random Coefficients	5. 発行年 2021年
3. 雑誌名 Sankhya A	6. 最初と最後の頁 477-508
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s13171-020-00204-5	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕 計12件（うち招待講演 5件 / うち国際学会 5件）

1. 発表者名 小田凌也
2. 発表標題 Condition of GIC to the model minimizing KL-loss function in high-dimensional multivariate linear regression
3. 学会等名 5th International Conference on Econometrics and Statistics (EcoSta 2022) (招待講演) (国際学会)
4. 発表年 2022年

1. 発表者名 小田凌也
2. 発表標題 高次元 GMANOVA モデルにおける予測のための一般化 Cp 規準の漸近性質
3. 学会等名 2022年度統計関連学会連合大会
4. 発表年 2022年

1. 発表者名 小田凌也
2. 発表標題 多変量モデルにおける複合型高次元漸近理論を用いたモデル選択規準の漸近損失有効性
3. 学会等名 多変量統計学・統計的モデル選択の新展開
4. 発表年 2023年

1. 発表者名 Oda Ryoya, Yanagihara Hirokazu
2. 発表標題 A consistent likelihood-based variable selection method in normal multivariate linear regression
3. 学会等名 The 13th KES International Conference on Intelligent Decision Technologies (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 Oda Ryoya, Yanagihara Hirokazu
2. 発表標題 A consistent variable selection method with GIC in multivariate linear regression even when dimensions are large
3. 学会等名 4th International Conference on Econometrics and Statistics (EcoSta 2021) (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 Oda Ryoya, Yanagihara Hirokazu
2. 発表標題 Asymptotically KL-loss efficiency of GIC in normal multivariate linear regression models under the high-dimensional asymptotic framework
3. 学会等名 2021年度統計関連学会連合大会
4. 発表年 2021年

1. 発表者名 Oda Ryoya, Yanagihara Hirokazu
2. 発表標題 高次元多変量線形回帰における KL ロス最小化に基づくモデルの一致性
3. 学会等名 2021年度広島大学金曜セミナー
4. 発表年 2021年

1. 発表者名 Yanagihara Hirokazu, Oda Ryoya
2. 発表標題 Coordinate descent algorithm for normal-likelihood-based group Lasso in multivariate linear regression
3. 学会等名 The 13th KES International Conference on Intelligent Decision Technologies (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 小田凌也
2. 発表標題 多変量線形回帰における正規尤度に基づく簡便なモデル選択法をその一致性の評価について
3. 学会等名 広島大学金曜セミナー
4. 発表年 2020年

1. 発表者名 Oda Ryoya
2. 発表標題 Kick-One-Out-Based Variable Selection Method Using Ridge-Type Cp Criterion in High-Dimensional Multi-response Linear Regression Models
3. 学会等名 The 15th KES International Conference on Intelligent Decision Technologies (招待講演) (国際学会)
4. 発表年 2023年

1. 発表者名 Oda Ryoya
2. 発表標題 Asymptotic loss efficiency of a model selection criterion in a high-dimensional GMANOVA model
3. 学会等名 統計数理研究所 共同利用 2023年度 重点型研究 研究集会 「高次元データ解析・スパース推定法・モデル選択法の開発と融合」
4. 発表年 2024年

1. 発表者名 小田凌也
2. 発表標題 GMANOVAモデルとモデル選択規準の高次元漸近性質
3. 学会等名 岡山統計研究会 第182回研究会全体レクチャー
4. 発表年 2024年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------