

令和 5 年 11 月 1 日現在

機関番号：12608

研究種目：研究活動スタート支援

研究期間：2019～2022

課題番号：19K24340

研究課題名（和文）非凸性に基づく統計モデリングと収束保証付き推定アルゴリズム

研究課題名（英文）Statistical modeling based on non-convexity with convergence guaranteed estimation algorithm

研究代表者

川島 孝行（Kawashima, Takayuki）

東京工業大学・情報理工学院・助教

研究者番号：60846210

交付決定額（研究期間全体）：（直接経費） 1,500,000円

研究成果の概要（和文）：実データ解析での問題に対処しようとする、自然と現れる非凸性に基づく統計モデリングを扱った。統計的な性質だけでなく、推定アルゴリズムの効率化も同時に達成する研究を目指した。

i) 歪正規分布を対象とした、推定アルゴリズムに関する研究では、従来の推定アルゴリズムとは異なり推定を加速させる慣性項が自然と入る更新式の導出に成功した。結果として、数値実験では、従来の推定アルゴリズムに比べて少ない計算時間で推定を行うことが可能になった。

ii) 超過死亡で用いられるFarringtonアルゴリズムに、地理的加重回帰モデルの枠組みを導入することで、少ないデータでも推論を行うことを可能にした。

研究成果の学術的意義や社会的意義

歪正規分布はそのモデリングの柔軟性から広い分野ですでに使われており、今回の研究により短時間での推定が可能になったため、より大規模なデータにも適用可能である。

超過死亡推定のために用いられているFarringtonアルゴリズムを少ないデータでも推定できるように拡張を行ったことで、データを大量に習得ができない状況や、対象の事象が初期の段階でも、本アルゴリズムを適用することで推定が可能になった。

研究成果の概要（英文）：We study statistical modeling based on the non-convexity that naturally arises when dealing with problems in real data analysis. The research aimed to simultaneously achieve not only statistical properties but also efficiency of the estimation algorithm.

i) In a study of estimation algorithms for skew-normal distributions, we succeeded in deriving an update formula that naturally includes an momentum term that generally accelerates estimation. Numerical experiments have shown that our estimation algorithm can perform in less computation time compared with conventional estimation algorithms.

ii) We have incorporated a geographically weighted regression model into the farrington algorithm used in excess mortality. It allows inferences to be made even with small amounts of data.

研究分野：統計科学

キーワード：統計モデリング 非凸性 推定アルゴリズム

### 1. 研究開始当初の背景

実データ解析の問題に対処しようとする自然と現れる非凸性に基づく統計モデリングを対象として、それに付随する推定アルゴリズムの効率性も求めた研究を行おうと考えていた。特に、外れ値、歪んだ分布を軸とした非凸性に基づく統計モデリングを扱おうと決めていた。すでに、外れ値では、回帰問題を対象としてある程度の成果を得ており、そこで得た知見を今回にも適用しようと考えていた。

歪んだ分布に関しては、広く使われている歪正規分布の既存の推定アルゴリズムとは異なる推定アルゴリズムの導出を考えていた。当初は、導出されたアルゴリズムに自然と推定を加速する項が含まれていることには気づいていなかった。

また、高次元回帰問題に対しても、射影勾配法の観点から別なアルゴリズムが導出できないかと、既存研究 (Duchi et. al., 2008) からの発展を考えていた。まずは、単純な拡張として LASSO の結果を Adaptive LASSO に拡張することを手がかりとした。

Farrington アルゴリズムに関しては、当初は全く想定しておらず、研究開始からしばらく経ったのちに着手したが、アルゴリズムの内部では実データ解析の問題に対処するために、上記で述べたような種々の統計モデリングが存在しており、複合的に解決することを目指した。

### 2. 研究の目的

歪正規分布の推定の効率をあげることは、今まで使われていた分野のデータだけではなく、より大規模なデータへの適用を可能にする研究である。得られた推定アルゴリズムを数値実験により、既存の推定アルゴリズムと比較することで計算時間の短縮が可能かを目的としていた。

高次元回帰問題の研究では、LASSO を対象とした射影勾配法に基づく推定アルゴリズムを、より一般の正則化項に拡張できないかを目的としていた。LASSO よりも統計的に良い性質をもつと知られている種々の非凸な正則化項に広げることが最終目標としていた。

Farrington アルゴリズムの研究では、すでに超過死亡推定の手法として確立していたものを、対象とする事象の初期の段階では小規模なデータしかない場合にも適用できるような手法へと拡張することを目的としていた。

### 3. 研究の方法

歪正規分布の研究では、まず従来とは異なる推定アルゴリズムとより相性の良い確率表現を考えた。これを用いて、導出したアルゴリズムには、一般に推定の効率を良くする慣性項が自然と入っていることがわかった。また、推定アルゴリズムの更新式が陽な形で得られるという利点もある。

高次元回帰問題の研究では、LASSO の場合の厳密な射影勾配法の更新式を、まずは同じ考えで Adaptive LASSO にも適用を行った。多くの必要な式の導出は全く同じように行えたが、重要な更新式を導出する際に必要な式変形では、少し改良した考えが必要になった。また、複数の凸集合からなる集合への射影である Dykstra の方法をこれに用いることでより複雑な正則化項にも拡張しようと考えた。非凸な正則化項へも似たような考えで拡張できいかと試みたが、これは今後の課題となった。

Farrington アルゴリズムの研究では、多くの場合が市区町村のように地理的なつながりを暗にもつデータを対象することがほとんどで、そのつながりを利用することで、一つ一つの区分(市単位、区単位など)では少数のデータしか集まっていない場合にも、周辺のデータとまとめて推定することを可能にできないかと考えた。具体的には、地理的加重回帰の手法を導入することで、これを可能にした。また、推定ののちに行う推論ではどのように予測区間を構成するかの問題が存在していて、そこを既存の可変係数モデルの理論を用いて構成することを行った。

### 4. 研究成果

歪正規分布に対する新たな効率の良い陽な推定アルゴリズムを導出することができた。これにより、より多くのユーザーが歪んだ分布に基づく統計的推論を簡便に行えると期待できる。また、実際にデータ解析を行うための R 言語によるパッケージの作成・公開も行った。

LASSO の拡張である Adaptive LASSO に対して、射影勾配法に基づく厳密な射影に基づく更新式の導出を行った。これにより、実問題において推定されるパラメータが厳密に制約を満たしてほしい場合にも、推定を行うことが可能になった。近似ではなく厳密に満たすことが可能な我々の推定アルゴリズムは、実データ解析において広く使われることが期待される。

地理的回帰の枠組みを用いて、Farrington アルゴリズムの一般化を行った。従来、ある程度のサイズの時系列データを推定のために必要としていた Farrington アルゴリズムを地理的な方向でのデータも協調的に使用する地理的回帰の考えを導入することでより、少数のデータでも正確な推定を行うことが可能になった。本手法を用いてデータ解析を行うための R 言語によるコードの作成・公開を行った。これにより、すでに広く使われている Farrington アルゴリズムがより広範な分野で使用されることが期待できる。

[ 雑誌論文 ]

i) Yoneoka, D., Kawashima, T., Makiyama, K., Tanoue, Y., Nomura, S., Eguchi, A., "Geographically weighted generalized Farrington algorithm for rapid outbreak detection over short data accumulation periods" [Statistics in Medicine](#)

ii) Abe, T., Fujisawa, H., Kawashima, T. and Lay, C., "EM algorithm using overparametrization for multivariate skew-normal distribution" [Econometrics and Statistics](#)

[学会発表]

i) 川島 孝行 “射影勾配法に基づく高次元回帰” 2020 年度科研費シンポジウム 「多様な分野のデータに対する統計科学・機械学習的アプローチ」

ii) 川島 孝行 “ロバストかつスパースな回帰モデリング” 第 4 回統計・機械学習若手シンポジウム

iii) 川島 孝行 “ガンマ・ダイバージェンス最小化に基づくロバストかつスパースな回帰” 統計学と機械学習の数理と展開

5. 主な発表論文等

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

1. 著者名 Yoneoka Daisuke, Kawashima Takayuki, Makiyama Koji, Tanoue Yuta, Nomura Shuhei, Eguchi Akifumi	4. 巻 40
2. 論文標題 Geographically weighted generalized Farrington algorithm for rapid outbreak detection over short data accumulation periods	5. 発行年 2021年
3. 雑誌名 Statistics in Medicine	6. 最初と最後の頁 6277 ~ 6294
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/sim.9182	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 Abe Toshihiro, Fujisawa Hironori, Kawashima Takayuki, Ley Christophe	4. 巻 19
2. 論文標題 EM algorithm using overparameterization for the multivariate skew-normal distribution	5. 発行年 2021年
3. 雑誌名 Econometrics and Statistics	6. 最初と最後の頁 151-168
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.ecosta.2021.03.003	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3件（うち招待講演 3件/うち国際学会 0件）

1. 発表者名 川島 孝行
2. 発表標題 射影勾配法による高次元回帰モデリング
3. 学会等名 2020年度科研費シンポジウム 「多様な分野のデータに対する統計科学・機械学習的アプローチ」 (招待講演)
4. 発表年 2020年

1. 発表者名 川島 孝行
2. 発表標題 ロバストかつスパースな回帰モデリング
3. 学会等名 第4回統計・機械学習若手シンポジウム (招待講演)
4. 発表年 2019年

1. 発表者名 川島 孝行
2. 発表標題 ガンマ・ダイバージェンス最小化に基づくロバストかつスパースな回帰
3. 学会等名 統計学と機械学習の数理と展開（招待講演）
4. 発表年 2019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