

平成 26 年 6 月 11 日現在

機関番号：32612

研究種目：若手研究(B)

研究期間：2012～2013

課題番号：24700281

研究課題名(和文)積分と最適化を伴う統計的推測

研究課題名(英文)Statistical inference involving integration and optimization

研究代表者

清 智也 (Sei, Tomonari)

慶應義塾大学・理工学部・講師

研究者番号：20401242

交付決定額(研究期間全体)：(直接経費) 1,000,000円、(間接経費) 300,000円

研究成果の概要(和文)：統計学における最尤推定の計算において、確率分布族の正規化定数を表す重積分が陽に求まらない場合、多くの計算時間が必要となることがある。ホロノミック勾配法は、そのような問題に対処するために開発された数値計算技法であり、近年研究が進んでいる。本研究では、Bingham 分布族と呼ばれる確率分布族について、ホロノミック勾配法が適用できることを示した。また、不均衡なデータの判別問題における二項回帰モデルの挙動に付いても考察した。ここで、不均衡なデータとは、判別すべき2値ラベルの頻度が偏ったデータのことである。本研究では、このようなデータに対しては、モデルの極限が特殊なポアソン点過程となることを示した。

研究成果の概要(英文)：In statistics, the maximum likelihood estimation sometimes needs heavy computation due to the normalizing constant. The holonomic gradient methods are developed to avoid such computation and progress in recent years. In this research, the holonomic gradient methods are shown to be available for the Bingham distributions. Behavior of the binomial regression model for an imbalanced dataset is also investigated. The imbalanced data means that the frequency of the binary label that should be classified is imbalanced. In this research, the limit of the model for such datasets is a special class of Poisson point processes.

研究分野：統計科学

科研費の分科・細目：情報学・統計科学

キーワード：統計数学 最適化 ホロノミック勾配法 不均衡データ 情報幾何

1. 研究開始当初の背景

本研究の背景となる問題は「積分を含む目的関数の最適化」であった。例として以下の2つが挙げられる。

(1) ベイズ法における周辺尤度の最大化

時空間統計解析や生命情報処理、画像処理などにはベイズ法が利用されている。ベイズ法における困難な問題の一つとして、周辺尤度の最大化がある。この問題に対してはマルコフ連鎖モンテカルロ法 (MCMC) は強力な方法の一つであるが、それでも計算時間は多く要する。

(2) 指数型分布族に対する最尤法

最尤法は統計的推測理論における根幹をなす手法の一つであり、広く用いられている。そして指数型分布族は、エントロピー最大化の原理から自然に導かれる分布族であるが、そこには正規化定数が含まれ、その計算がネックとなる場合がある。

2. 研究の目的

本研究の目的は、計算機代数を用いて「積分を含む目的関数の最大化」を効率化することであった。明らかにしたいことは以下の3つであった。

(1) 「積分を含む目的関数の最適化」が必要となる統計的問題の整理

まず、解決すべき問題をある程度確定しておく必要がある。大まかには背景で紹介した2つの例、つまり「ベイズ法における周辺尤度の最大化」、「指数型分布族に対する最尤法」と分類できる。その他に適用可能な例もあり得る。また、分布族の定義されるサポート集合や母数空間の次元など、個別のモデルが持つ特性に応じて分類するなどの系統的な整理も必要となる。

(2) 「積分を含む目的関数の最適化」が効率化できるようなクラスの導出

本研究のアプローチ、つまり計算機代数による方法、が有効となるようなモデルを見極める必要がある。適用可能なクラスを予め見積もっておくことにより、不必要な試行錯誤を避けることができる。計算機代数の専門家の意見も適宜取り入れながら、クラスを導出する。

(3) 実装と比較

計算手法がある程度確立された時点でソフトウェアの実装も行なう。実装には2つのフェーズがある。1つは計算機代数による記号処理であり、もう1つは数値計算であ

る。両者を実装して、さらに既存の手法と比較することによって初めて、本手法が有効であることの証明となる。また、実装したプログラムが機能しなかった場合は、方法論・アルゴリズムの両面から適宜修正する。

3. 研究の方法

本研究の目的を達成するため、以下の方法をとった。

(1) ホロノミック勾配法が適用可能な統計的問題の整理

ホロノミック勾配法と呼ばれる、主に積分計算のための数値計算手法を、現実的な統計問題に応用する。そのために、まずホロノミック勾配法が適用可能なモデルを探すことから始めた。そしてその拡張を試みた。

(2) 実装と比較

Bingham 分布族に大してはホロノミック勾配法が適用可能であることを確かめ、それをソフトウェアとして実装することを試みた。ここでは汎用性なども考慮しつつ実装を行った。言語としては統計学においてよく用いられている R を利用した。

(3) 他の統計的問題へのアプローチ

別の観点として、ホロノミック勾配法という方法にこだわることなく、実データから数学的に興味深い部分を抽出するというアプローチも試みた。特に、不均衡データ (2 値ラベルの頻度が極端に偏っているようなデータ) に対して、ロジスティック回帰などの二項回帰モデルを適用した場合に、どのような性質が現れるかを詳細にした考察した。

4. 研究成果

主に2つの研究成果を得た。

(1) Bingham 分布族に対するホロノミック勾配法

Bingham 分布族は方向データと呼ばれる特殊なデータに用いられる分布族の一つである。この分布族の正規化定数は陽には求まらないことが知られていた。これに対し、Bingham 分布族に対してホロノミック勾配法は適用可能であることが示された。また、Fisher-Bingham 分布族の正規化定数に対する既存の結果も見直し、理解しやすい結果を得た。

(2) 不均衡データに対する二項回帰モデル

の挙動

不均衡データとは、判別分析において、判別すべき2値ラベル(0か1)の頻度が偏ったデータのことであり、例えばサーバーへの不正アクセスのデータなどが該当する。このような不均衡性の数学的極限を考えると、ロジスティック回帰モデルがポアソン点過程モデルに収束するという結果が知られていた。この既存結果に対し、ロジスティック回帰以外の二項回帰モデルに対しても同様の極限定理が成り立つことを示し、さらに収束先のポアソン点過程の強度関数がべき乗の形に限られることも示した。この結果は、情報幾何学や極値理論への新たな視点を与える結果となった。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 3 件)

□ SEI, Tomonari (2014). Infinitely imbalanced binomial regression and deformed exponential families, *Journal of Statistical Planning and Inference*, Vol.149, pp.116—124, 査読有.

□ SEI, Tomonari and Alfred KUME. (2013). Calculating the normalising constant of the Bingham distribution on the sphere using the holonomic gradient method, *Statistics and Computing*, Online First, 査読有.

□ Sei, Tomonari (2013). A Jacobian inequality for gradient maps on the sphere and its application to directional statistics, *Communications in Statistics – Theory and Methods*, 42 (14), 2525—2542, 査読有.

〔学会発表〕(計 16 件)

1 清 智也 (2014). 不均衡データに対する二値回帰モデルの挙動, RIMS 共同研究「Asymptotic Statistics and Its Related Topics」, 2014年3月5日, 京都大学.

2 清 智也 (2014). 不均衡データと変形指数型分布族, RIMS 共同研究「統計多様体の幾何学の新展開」, 2014年2月21日, 京都大学.

3 清 智也 (2013). 不均衡データに対する二項回帰モデルの挙動, 極値理論の工学への応用, 2013年12月7日, 統計数理研究所.

4 清 智也, A. Kume (2013). Bingham 分布族に対するホロノミック勾配法, ホロノミック勾配法に関する研究会, 2013年12月4日, 東京大学.

5 清 智也 (2013). 二項回帰モデルの不均衡極限と変形指数型分布族, 一般化

線形モデルの最新の展開とその周辺, 2013年11月9日, 千葉大学.

6 Sei, Tomonari (2013). Infinitely imbalanced binomial regression and deformed exponential families, Boston-Keio Summer Workshop, Sep. 20, 2013, Boston University, USA.

7 清 智也 (2013). 二項回帰モデルの不均衡極限と変形指数型分布族, 統計関連学会連合大会, 2013年9月9日, 大阪大学.

8 Sei, Tomonari (2013). Holonomic gradient descent in directional statistics, SIAM Conference on Applied Algebraic Geometry 2013, Aug. 2, 2013, Colorado State University, USA.

9 清 智也, A. Kume (2013). ホロノミック勾配法によるBingham分布族の計算, 統計数学セミナー, 2013年5月17日, 東京大学.

10 清 智也 (2012). ホロノミック勾配法の今後, 研究集会「数理統計学の沃野」, 2012年11月23日, 慶應義塾大学.

11 清 智也 (2012). ホロノミック勾配法の適用例, RIMS 共同研究「トーリックイデアルに付随する統計モデルの探究」, 2012年9月13日, 京都大学.

12 清 智也 (2012). 構造的勾配モデルの統計的曲率, 統計関連学会連合大会, 2012年9月10日, 北海道大学.

13 清 智也 (2012). ホロノミック勾配法による正規化定数の計算法, 統計サマーセミナー, 2012年8月5日, 熱海.

14 清 智也 (2012). コピュラモデルの統計的曲率, JST CREST「現代の産業社会とグレブナー基底の調和」若手研究会, 2012年7月16日, 慶應義塾大学.

15 Sei, Tomonari (2012). Construction of directional distributions using gradient maps, The 2nd Institute of Mathematics Asia Pacific Rim Meeting (IMS-APRM), July 4, 2012, Tsukuba, Japan.

16 清 智也 (2012). 尤度計算のためのホロノミック勾配法, 統計数学セミナー, 2012年5月31日, 東京大学.

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

名称:

発明者:

権利者:

様 式 C - 1 9、F - 1 9、Z - 1 9、C K - 1 9 (共 通)

種類：
番号：
出願年月日：
国内外の別：

取得状況 (計 0 件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

[その他]
ホームページ等
<http://www.math.keio.ac.jp/~sei/>

6 . 研究組織

(1) 研究代表者

清 智也 (SEI, Tomonari)

研究者番号 : 20401242

(2) 研究分担者

()

研究者番号 :

(3) 連携研究者

()

研究者番号 :