

平成 30 年 5 月 21 日現在

機関番号：10101

研究種目：若手研究(B)

研究期間：2014～2017

課題番号：26730014

研究課題名(和文) 予測の問題と条件付き正規化最尤分布

研究課題名(英文) Prediction and Conditional Normalized Maximum Likelihood Distribution

研究代表者

廣瀬 善大 (HIROSE, Yoshihiro)

北海道大学・情報科学研究科・准教授

研究者番号：10637391

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：本研究課題の興味の対象は「条件付き正規化最尤分布」と呼ばれる確率分布、および関連して現れる確率分布の評価尺度「条件付きリグレット」であった。過去の観測値を固定した場合、過去の観測値を平均化した場合とも、カルバック-ライブラー情報量リスクと呼ばれる評価尺度のもとでは、条件付き正規化最尤分布よりも性能のよい確率分布が存在することが分かった。その確率分布は、条件付きリグレットリスクについて、最悪の観測値が与えられた場合の性能を最大化するものであった。また、この確率分布は、条件付き正規化最尤分布をベイズ予測全体の空間に射影して得られる確率分布でもあった。

研究成果の概要(英文)：This research project treated the conditional normalized maximum likelihood distribution, which is a conditional probability distribution. We were also interested in a measure of probability distributions which is called the conditional regret. When we fix past observations and when we consider the average of the past observations, another probability distribution is superior to the conditional normalized maximum likelihood distribution if they are evaluated with the Kullback-Leibler information. That probability distribution attains the minimax of another measure of probability distributions, which is called the conditional regret risk. Furthermore, the probability distribution coincides with the projection of the conditional normalized maximum likelihood distribution on the space of the Bayesian predictions.

研究分野：統計科学

キーワード：統計的予測 統計的決定理論

1. 研究開始当初の背景

(1) 情報理論,特に符号理論の文脈において,リグレットと呼ばれる損失関数のミニマックスを達成するような符号として正規化最尤分布が知られている.リグレット損失は,将来の実現値 y を確率分布 $P(y)$ で符号化したときの符号長と「あとだし」符号化による符号長との差をとったものである.ただし,「あとだし」符号化とは,将来の観測値を仮想的に利用した符号である.正規化最尤分布とは,最悪の実現値 y でのこの符号長の差が最小になるような確率分布(ミニマックス分布)であり,きわめて自然な状況を考えてものである.この正規化最尤分布を予測の問題の観点から見ると,過去の観測値がない場合の予測の問題を扱っていると考えことができ,これは予測問題の特殊例になっている.条件付き正規化最尤分布は,正規化最尤分布の考え方を過去の観測値がある場合の予測に応用して得られるもので,予測の問題におけるひとつの予測分布に対応する.正規化最尤分布と同じく自然な状況を考えており,条件付き正規化最尤分布の性質を明らかにすることは有意義である.

(2) 条件付きリグレット(リグレット損失を観測ありの場合に応用した損失)と条件付き正規化最尤分布は,その定義の自然さにも関わらず,その性質について十分な結果が得られていなかった.たとえば,カルバック-ライブラー情報量リスクに関する条件付き正規化最尤分布の諸性質は不明な部分が多い.正規化最尤分布についても,カルバック-ライブラー情報量の下での評価については同様の状況で,関連する研究はいくつかあるが十分な結果が得られていない.カルバック-ライブラー情報量は符号長の観点からも自然な量であり,条件付き正規化最尤分布や正規化最尤分布がカルバック-ライブラー情報量に関してどのような性質をもつのかを調べることは重要である.また,条件付きリグレットは計算機科学におけるアイデアとの関連も指摘されており,諸分野の関連も興味深い.

2. 研究の目的

(1) 条件付き正規化最尤分布の基本的な性質を明らかにする

(2) 条件付きリグレットリスクに関する基本的な性質を明らかにする

(3) 正規化最尤分布のカルバック-ライブラー情報量に関する許容性を調べる

(4) 条件付き正規化最尤分布の拡張や修正を検討する

3. 研究の方法

(1) 具体的な統計モデルの下で条件付き正規化最尤分布の性質を調べる: 多変量正規

分布モデルや二項分布モデルなどの比較的単純なモデルにおいて,条件付き正規化最尤分布の性質を調べる.カルバック-ライブラー情報量リスクや条件付きリグレットリスクに関する許容性やミニマックス性について調べる.予測の問題におけるカルバック-ライブラー情報量リスクの下での評価については既存の研究があり,既存研究の議論を参照しながら条件付き正規化最尤分布の許容性を調べる.また,条件付きリグレットリスクに関する許容性はカルバック-ライブラー情報量リスクでの許容性と同様の議論ができることが分かったので,カルバック-ライブラー情報量リスクに関する結果から条件付き正規化最尤分布の許容性を調べることができる.条件付きリグレットリスクに関するミニマックス性については,カルバック-ライブラー情報量リスクに関するミニマックス性と異なる議論が必要になるため,次の項目(2)の研究が必要となる.上記の研究においては,一般に解析的な評価が難しいことも想定される.そのようなときには,計算機による数値計算により近似的な評価を行い,理論的評価の助けとする.ところで,条件付き正規化最尤分布の性質として,他の条件付き分布(予測分布)との関係も調べる予定である.特に,ベイズ統計における事前分布のうち条件付き正規化最尤分布に対応するものが存在するかどうかに興味深い問題である.事前分布の存在は扱う統計モデルごとに結果が変わることも考えられる.

(2) 条件付きリグレット損失,条件付きリグレットリスクの性質を調べる: 上記項目(1)にも書いたように,条件付きリグレットリスクに関する許容性はカルバック-ライブラー情報量リスクでの許容性と同様の議論ができる.本項目で興味があるのは,主に条件付きリグレットリスクに関するミニマックス性である.一般にミニマックスを達成する条件付き分布を見つけることは難しいが,まずは具体的な統計モデルを仮定して調べる.項目(1)と並行して作業を進める.カルバック-ライブラー情報量リスクに関するミニマックス性の研究の結果の一部,あるいはそこで用いられている方法が参考になるものと期待される.

(3) 正規化最尤分布のカルバック-ライブラー情報量の下での評価を与える: 正規化最尤分布がカルバック-ライブラー情報量の下で許容性をもつかどうかを調べる.ある条件の下で,カルバック-ライブラー情報量に関して正規化最尤分布を優越するか同等な確率分布の存在が知られている.ただし,正規化最尤分布が許容的であるような場合にはその確率分布は正規化最尤分布と一致するので,その事実から正規化最尤分布の非許容性を結論することはできない.正規化最尤分布を優越する確率分布と正規化最尤分布と

が一致するかどうかは正規化最尤分布に対応する事前分布の存在の有無で判定されるので、そのような事前分布の存在・非存在を証明することを目指す。正規化最尤分布に対応する事前分布についてもまずは具体的な統計モデルの下で考える。

(4) 一般的な設定で条件付き正規化最尤分布の性質を調べる：特定の統計モデルを仮定せずに、条件付き正規化最尤分布の性質を調べる。項目(1)で具体的なモデルにおいて条件付き正規化最尤分布の性質を調べることにより得られた知見をもとに、得られた条件付き正規化最尤分布の性質が統計モデルのどのような性質によるものなのかを整理する。また、項目(2)と(3)に対応する内容を一般的な状況で考える。

4. 研究成果

(1) まず基本的な統計モデルにおいて条件付き正規化最尤分布を確認した。たとえば、正規分布や二項分布などである。特に、条件付き正規化最尤分布の定義には複数のものが提案されているため、それらをすべて確認した。さらに、ベイズ統計における代表的な事前分布による予測分布との比較した。統計モデルによっては、一部の条件付き正規化最尤分布がベイズ予測分布と一致することを確認した。

(2) 過去の観測値 x を固定した状況で、カルバック-ライブラー情報量を評価尺度として条件付き正規化最尤分布を評価した。条件付き正規化最尤分布は条件付きリグレットを最小化する確率分布であったが、カルバック-ライブラー情報量リスクで評価する場合には、条件付き正規化最尤分布よりもよい性能を示す確率分布が存在する。その確率分布は、条件付きリグレットに基づくリスクのミニマックスを達成する確率分布であった。幾何学的には射影を考えていることに相当する。

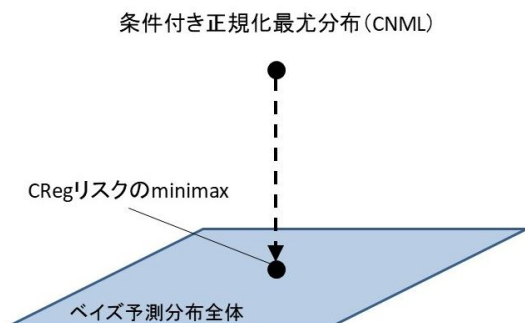


図: 条件付き正規化最尤分布と条件付きリグレットリスクのミニマックスを達成する確率分布

(3) 過去の観測値 x を固定せず、観測値 x についても期待値をとって条件付き正規化最尤分布を評価した。カルバック-ライブラー情報量で評価を行う場合、条件付きリグレットリスクのミニマックスを達成する確率分布が条件付き正規化最尤分布よりもよい性能をもつことを示した。この事実の証明はベイズ予測の既存研究を組み合わせたものであった。

(4) カルバック-ライブラー情報量と条件付きリグレットの関係から、それぞれを評価尺度として利用した際の違いについて考察した。

(5) 本課題の最終年度に、本研究課題とは別の問題意識から、局外パラメータが存在する場合の条件付き正規化最尤分布の扱いがあらたに問題として浮上した。この問題はまだ解決しておらず、今後の課題として、本研究課題の終了後も継続して取り組む予定である。

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 4 件)

Toyoto Tanaka, Yoshihiro Hirose and Fumiyasu Komaki (2018). Second-Order Matching Prior Family Parametrized by Sample Size and Matching Probability. Accepted for publication in Statistical Papers. 査読有り.
DOI: 10.1007/s00362-018-1001-5

Kenji Araki, Yoshihiro Hirose and Fumiyasu Komaki (2018). Paired Comparison Models with Age Effects Modeled as Piecewise Quadratic Splines. Accepted for publication in International Journal of Forecasting. 査読有り.
DOI: 10.1016/j.ijforecast.2018.02.006

Yoshihiro Hirose (2015). An Information-Geometrical Path Algorithm for Poisson Regression. Proceedings of the 60th ISI World Statistics Congress, 2329-2334. 査読有り.
<http://www.isi2015.ibge.gov.br/>

Yoshihiro Hirose and Fumiyasu Komaki (2015). An Estimation Procedure for Contingency Table Models Based on the Nested Geometry. Journal of the Japan Statistical Society, 45, 57-75. 査読

有り.
DOI: 10.14490/jjss.45.57

〔学会発表〕(計 6 件)

Yoshihiro Hirose. A Decision-theoretic Property of Conditional Normalized Maximum Likelihood Distribution. CMStatistics 2017, 2017, London, UK.
廣瀬善大. 条件付き正規化最尤分布とその評価尺度について. 2017 年度統計関連学会連合大会, 2017, 南山大学.
廣瀬善大. 入れ子構造の情報幾何を利用した統計的推定(小川研究奨励賞受賞講演). 2016 年度統計関連学会連合大会, 2016, 金沢大学.
廣瀬善大. 条件付き正規化最尤分布とその決定理論的性質について. 2016 年度統計関連学会連合大会, 2016, 金沢大学.
Yoshihiro Hirose. An Estimation Procedure Based on the Information Geometry of Dually Flat Spaces. RSS 2015 Annual Conference, 2015, Exeter, UK.
廣瀬善大. 条件付き正規化最尤分布の許容性に関する性質. 2014 年度統計関連学会連合大会, 2014, 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:
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
<https://sites.google.com/view/yoshihirohirose/japanese>

6. 研究組織
(1) 研究代表者

廣瀬 善大 (HIROSE, Yoshihiro)
北海道大学・大学院情報科学研究科・准教授
研究者番号: 10637391

(2) 研究分担者
()

研究者番号:

(3) 連携研究者
()

研究者番号:

(4) 研究協力者
()