

令和 3 年 5 月 6 日現在

機関番号：14401

研究種目：基盤研究(C)（一般）

研究期間：2018～2020

課題番号：18K03413

研究課題名（和文）レニー・ダイバージェンスに対する多層融合化規準への拡張と議員定数問題への応用

研究課題名（英文）An extension to multi-fused criterion for Renyi divergence and its application to apportionment

研究代表者

濱田 悦生（Hamada, Etsuo）

大阪大学・基礎工学研究科・特任教授（常勤）

研究者番号：20273617

交付決定額（研究期間全体）：（直接経費） 2,300,000円

研究成果の概要（和文）：本研究課題に対して、モデル選択基準族に対する頑健な一般化によりAICを一般化したcriteriaのダイバージェンス族を導出し、多項式回帰モデルにおいて、この族が漸近的にAICと同等であることを示し、更に外れ値があっても十分なパフォーマンスが出ることを示すことが出来た結果を受けて、外れ値のあるデータを多項式回帰モデルで分析する時のロバストな情報量規準に関して、BICに対応する一致性とBHHJに対応する頑健性を同時に併せ持つ新しい情報量規準として、影響関数を基盤に持つRCCを提案することが出来た。

研究成果の学術的意義や社会的意義

モデル選択基準族において、外れ値を持つデータに対しても十分有効なパフォーマンスを発揮する情報量規準の提案が望まれていたが、この研究成果では多項式回帰モデルという制限下ではあるが、一致性と頑健性を併せ持つ新しい情報量規準RCCを提案することが出来たのは、大きな学術的意義を持つと同時に、データサイエンスの現場での応用可能という意味で社会的意義を持つといえる。

研究成果の概要（英文）：In response to this research, we derived a divergence family of criteria that generalizes AIC by a robust generalization of the model selection criterion family, showed that this family is asymptotically equivalent to AIC in polynomial regression models, and further showed that it performs well even in the presence of outliers. Based on these results, we proposed a new information criterion, RCC, which is based on the influence function, as a robust information criterion for analyzing data with outliers using polynomial regression models, and which has both the consistency corresponding to BIC and the robustness corresponding to BHHJ.

研究分野：数理統計学

キーワード：ダイバージェンス モデル評価基準 ロバストネス

1. 研究開始当初の背景

統計学的枠組みの1つに、観測される現象の未知なる「真の確率分布」を G とし、その現象に対して母数空間 Θ のパラメータ θ を含んだ統計モデルの分布を $F(\theta)$ としたとき、 G と $F(\theta)$ との差を、ある規準によって最小化するような分布 $F(\theta)$ を選択するモデル選択やその分布でのパラメータ θ の推定がある。分布 $G, F(\theta)$ の確率（密度）関数をそれぞれ $g(x), f(x|\theta)$ とおけば、その規準には、Kullback-Leibler 情報量 $D(G;F(\theta))$ をベースにしたダイバージェンスに基づくモデル評価規準もしくは情報量規準 (AIC や BIC)、その一般化である一般化情報量規準 (GIC)(Konishi and Kitagawa (1996)), minimum distance approach をベースにした f -ダイバージェンスの拡張規準 (Basu et al. (2011)), Minimum Description Length (MDL) 原理による MDL 規準 (Rissanen (1996)) などがあるが、本研究においては、ダイバージェンスに基づくモデル評価規準をベースとする。議員定数配分問題において、一森 (2012) により既存の定数配分方式との対応として指数 H_α を用いた Renyi のダイバージェンスが使用された。しかし、この規準自体だけではどの指数 H_α を用いれば定数配分として良いのかという、最適化を得ることが出来ないという問題点を本研究代表者が発見した。

2. 研究の目的

ダイバージェンスそれ自体で指数の最適解を与えることが出来ない場合、どのような指数の調整項を追加すれば最適解を得ることが出来るのか、またその一般化の定式化はどうなるのか、という「問い」を受けて、研究の目的は、Renyi ダイバージェンスを基にして指数の調整項を付与した最適な一般化拡張ダイバージェンスの多層融合化した定式化とその議員定数問題への適用、であった。

3. 研究の方法

研究の方法として、以下の5段階を計画していた：

(第1段階) Renyi ダイバージェンスにおける指数の調整項の決定理論的候補選定と、その最適な指数の決定による拡張ダイバージェンスの定式化。

(第2段階) 指数を持つ一般のダイバージェンスに対し、拡張ダイバージェンスを適用することによる一般化拡張ダイバージェンスとしての理論的定式化。

(第3段階) 一般化拡張ダイバージェンスでの指数に客観事前分布を適用して、Bayes 法を利用した多層融合化ダイバージェンスの定式化の提案。

(第4段階) 他のダイバージェンスに対する多層融合化ダイバージェンスの漸近的比較検討。

(第5段階) 多層融合化ダイバージェンスを用いた一票の格差問題における議員定数問題への適用及びその結果を踏まえた提言。

4. 研究成果

第1段階である Renyi ダイバージェンスにおける指数の調整項の決定理論的候補選定とその最適な指数の決定による拡張ダイバージェンスの定式化を行う予定であったが、調整項の候補リストに思いのほか時間がかかり、候補選定まで至らなかった。しかしながら、第2段階である指数を持つ一般のダイバージェンスに対し拡張ダイバージェンスを適用することによる一般化拡張ダイバージェンスとしての理論的定式化を行うために、Renyi ダイバージェンスだけでなく、他の指数を持つダイバージェンスへの適用可能性を探る事に関しては、モデル評価規準の一つとして、本研究に関連した外れ値に強いロバストなダイバージェンスに関する共同研究の結果、漸近的に選択の一致性と頑健性を両立する情報量規準を得た。

第3段階である「一般化拡張ダイバージェンスでの指数に客観事前分布を適用して Bayes 法を利用した多層融合化ダイバージェンスの定式化の提案」を行うことに関しては、Bayes 客観事前分布によるモデル化における事前パラメータの最適化に向けて、統計モデルのパラメータではなく事前パラメータに客観事前分布を想定した調整項を、1 番目の指数の調整項とリンクした形で検討を行った。

モデル選択基準族に対する頑健な一般化により AIC を一般化した criteria のダイバージェンス族を導出し、多項式回帰モデルにおいて、この族が漸近的に AIC と同等であることを示し、更に外れ値があっても十分なパフォーマンスが出ることを示すことが出来た結果を受けて、外れ値のあるデータを多項式回帰モデルで分析する時のロバストな情報量規準に関して、BIC に

対応する一致性と BHHJ に対応する頑健性を同時に併せ持つ新しい情報量規準として、影響関数を基盤に持つ RCC を提案することが出来た。また多値割り付けに対応する因果効果の推定法において、交絡変数の「バランス性」を考慮に入れた covariate balancing に基づいて、共変量の推定をすることなく不偏性と一致性を持つ因果効果の推定法を2つ提案することが出来た。

応用的な研究結果としては、厚生労働省のオープンデータを利用して、PCR 陽性者数と感染死亡者数の関係を Granger 因果性に関して検討し、令和2年6月18日付けの厚生労働省による感染死亡者の定義の変更による構造の変化を確認したが、残念ながら議員定数問題への適用までには至らなかった。

指数型分布族における十分統計量の誤用に関する指摘が行えたのは、意外な成果であり、今後研究を深めたいものとなった。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件／うち国際共著 0件／うちオープンアクセス 0件）

1. 著者名 Kurata, S. and Hamada, E	4. 巻 49
2. 論文標題 On the consistency and the robustness in model selection criteria	5. 発行年 2020年
3. 雑誌名 Communications in Statistics, Theory and Methods	6. 最初と最後の頁 5175-5195
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/03610926.2019.1615093	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 Orihara, S. and Hamada, E.	4. 巻 online
2. 論文標題 Determination of the Optimal Number of Strata for Propensity Score Subclassification	5. 発行年 2021年
3. 雑誌名 Statistics & Probability Letters	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.spl.2020.108951	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 濱田悦生	4. 巻 48
2. 論文標題 日本での COVID-19 禍における PCR 陽性者数と感染死亡者数の Granger 因果性の検討	5. 発行年 2021年
3. 雑誌名 行動計量学	6. 最初と最後の頁 39-48
掲載論文のDOI（デジタルオブジェクト識別子） 10.2333/jbhmk.48.39	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 -

〔学会発表〕 計1件（うち招待講演 0件／うち国際学会 0件）

1. 発表者名 濱田悦生
2. 発表標題 On a refinement of Hoeffding's inequality
3. 学会等名 統計関連学会連合大会
4. 発表年 2020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
---------	---------