

論文審査の要旨

博士の専攻分野の名称	博士（理学）	氏名	伊 森 晋 平
学位授与の要件	学位規則第4条第①・2項該当		
論文題目 Model Selection Criteria in Generalized Linear Models and their Extensions (一般化線形モデル及びその拡張におけるモデル選択規準)			
論文審査担当者 主 査 教授 若木 宏文 審査委員 教授 井上 昭彦 審査委員 教授 松本 眞			
〔論文審査の要旨〕 <p>統計解析において、候補のモデルの集合から最も良いモデルを選択することは、事象の予測や要因を特定するために重要である。本論文では、Nelder and Wedderburn (1972) によって提案され、実データ解析で有用なモデルのクラスとして知られる一般化線形モデル、及びその拡張型モデルに対して、「モデルのリスク推定」と「真のモデルの選択」という2つの観点からモデル選択規準量を提案している。</p> <p>一般化線形モデルとは、各個体に対して、目的変数と呼ばれるひとつの変数と説明変数と呼ばれる複数個の変数が観測されるとき、目的変数に指数型分布族と呼ばれる確率分布を仮定し、目的変数の平均を説明変数の1次式の関数（リンク関数）で記述することによって、目的変数と説明変数群との因果関係を表現するモデルである。説明変数の1次式を線形予測子、一次式の係数を回帰係数と呼ぶ。説明変数群から、目的変数に真に影響を及ぼしている変数を選択して、その回帰係数を推定することで、説明変数の目的変数への影響の強さを評価したり、説明変数の値から目的変数の値を予測することができる。</p> <p>モデル選択とは、説明変数群から必要な変数を選択したり、リンク関数を選択したりすることであるが、赤池情報量規準（AIC）は、実データ解析において最も広く用いられているモデル選択規準のひとつである。AIC は、Kullback-Leibler 擬距離に基づくモデルのリスクの推定量として Akaike (1973) によって提案された規準量で、リスクの推定量として標本数 (n) が大きくなるときの漸近不偏性を有している。候補となるモデルの各々の AIC を計算し、AIC が最小であるモデルを選ぶことで、リスクの小さなモデルを選択することが期待できる。AIC の、リスクの推定量としてのバイアスは $O(n^{-1})$ であるため、標本数が十分大きくない場合には、バイアスはモデル選択結果に悪影響を及ぼすことがある。</p> <p>本論文では、第2節において一般化線形モデルのバイアス補正 AIC (Corrected AIC; CAIC) を導出した。CAIC のバイアスは $O(n^{-2})$ であり、数値実験により CAIC は AIC に比べバイアスが小さくなっているだけでなく、真のモデルを選択する確率も大きくなっていることを見出した。多項ロジスティックモデルは、一般化線形モデルのひとつである（2項）ロジスティックモデルを拡張したモデルであるが、第3節では</p>			

多項ロジスティックモデルに対する CAIC を導出しており、シミュレーション及び実データ解析例を通して、通常の AIC を用いたモデル選択法を改善していることを確認している。

第4節では Liang and Zeger (1986) によって提案された一般化推定方程式法 (Generalized Estimating Equations; GEE) におけるモデル選択問題を扱っている。GEE は、各個体に対して、複数時点で目的変数と説明変数が観測されたデータ (経時データ) に対して、予測や要因特定を行う手法のひとつである。個体ごとに異なる時点で観測される目的変数を成分とする目的変数ベクトルが得られるが、GEE では、目的変数ベクトルの各成分の周辺分布に対しては一般化線形モデルを仮定するが、同時分布の分布形は仮定しない。GEE は、個体内の相関を考慮した作業用相関行列を組み込んだ、「擬似尤度」と呼ばれる未知回帰係数の関数の最大化問題から得られる推定方程式である。GEE の利点のひとつは、作業用相関行列が目的変数ベクトルの真の相関行列と異なっても、回帰係数ベクトルの推定量が一致性や漸近正規性を持つことである。しかし、回帰係数ベクトルの漸近分散行列は、作業用相関行列によって大きくなってしまうため、精度の良い推定を行うためには、適切な作業相関行列を選択することが必要となる。

GEE における作業用相関行列を選択するための規準量は、これまでにいくつか提案されているが、これらの先行研究では、規準量の性質はシミュレーションを通してのみ確認され、それらの理論的性質の解明にはあまり関心が向けられていなかった。申請者は、真の作業用相関行列を選択するための規準量のクラスを提案し、その選択確率が 1 に収束するための条件を導出した。

一般のモデル選択において、AIC は、真のモデルを選択する確率が 1 に収束するという意味での一致性の性質を持たないことが知られている。Nishii (1984) は、AIC のバイアス補正項を一般化した規準量 (Generalized Information Criterion; GIC) を考え、説明変数の選択問題において、GIC が一致性を持つための条件を導出した。GEE においては、目的変数ベクトルの同時分布を仮定しないため、GIC を用いることができない。申請者は Stein のロス関数と呼ばれる、行列の差異を表す量を基に、作業用相関行列を選択するための情報量規準のクラスを提案し、その規準が一致性を持つための十分条件を示した。シミュレーションにより、従来の作業用相関行列を選択するための代表的規準量が一致性を持たないのに対し、提案手法は高確率で真の作業用相関行列を選択することが確認された。

一般化線形モデルや一般化推定方程式は実データ解析において広く用いられており、申請者が得た結果は有用である。

以上、審査の結果、本論文の著者は博士 (理学) の学位を授与される十分な資格があるものと認める。

公表論文

- (1) Consistent Selection of Working Correlation Structure in GEE Analysis Based on Stein's Loss Function.
S. Imori,
Hiroshima Math. J. (2014), to appear.
- (2) Simple Formula for Calculating Bias-Corrected AIC in Generalized Linear Models.
S. Imori, H. Yanagihara & H. Wakaki,
Scand. J. Stat. (2014), 41, 535–555.
- (3) Bias-Corrected AIC for Selecting Variables in Multinomial Logistic Regression Models.
H. Yanagihara, K. Kamo, S. Imori & K. Satoh,
Linear Algebra Appl. (2012), 436, 4329–4341.