

# 学位論文要旨

$C_p$  type criterion for model selection in the generalized estimating equation method  
(一般化推定方程式における  $C_p$  型モデル選択規準)

氏名 佐藤 倫治

統計解析において、予測精度の向上のため複数のモデルの中から最適なモデルを選択する「モデル選択」が重要な問題となっている。観測データが相関を持つこともあり、医学や経済の分野でしばしば観測される相関を持つ「経時データ」と呼ばれる、同一観測対象における繰り返し測定データの解析に焦点を当てる。経時データを解析する手法の代表的なものとして Liang and Zeger (1986) で提案された一般化推定方程式 (GEE) がある。ほかの統計的枠組みと同様に GEE 法においてもモデル選択は重要な問題であり、本論文は GEE 法におけるモデル選択問題を扱う。

GEE 法では、観測データの周辺密度のみを仮定し、同時密度を仮定することなく作業用相関行列を用いて推定を行う。作業用相関行列は解析者が自由に選ぶことのできる行列で、真の構造と異なる誤った相関構造を用いて推定を行っても回帰係数の推定量は一致性や漸近正規性などの良い性質を持つことが知られている (Xie and Yang, 2003)。推定に用いた相関構造によって推定量の漸近分散が変わり、特に真の相関構造を用いた時の漸近分散が最小となるため、推定に用いる相関構造として最適なものを選択することも重要である。

GEE 法におけるモデル選択規準は様々提案されてきており、Pan (2001) で提案された QIC や Cantoni *et al.* (2005) で提案された  $GC_p$  規準は、リスクに相関を考慮しておらず、相関構造の選択までは行われていなかった。一方、Hin and Wang (2009) や Gosho *et al.* (2011) で提案された CIC は相関構造を選択するモデル選択規準であり、説明変数の選択が行われていなかった。Inatsu and Imori (2018) では、説明変数の選択のため  $C_p$  型のモデル選択規準 PMSEG (prediction mean squared error in the GEE) が提案されたが、モデルに含まれる尺度パラメータと相関行列に含まれる相関パラメータが既知である状況で導出されたものである。

本論文の目的は、Inatsu and Imori (2018) で提案された PMSEG を修正するため、尺度パラメータと相関パラメータを未知にして推定した影響を考慮したモデル選択規準の提案を行うこと、さらに、標本数  $n$  とともに観測時点数  $m$  も大きくなる場合の GEE 推定量の漸近性質を用いて、PMSEG の漸近性質を導出することである。

本論文の第 2 章では、GEE 法の枠組みを紹介し、尺度パラメータと相関パラメータが共に未知の場合のすべてのパラメータの推定方法を提案する。ここで

提案する推定アルゴリズムとモーメント推定量を用いることで、経験的にこのアルゴリズムは収束する。

本論文の第 3 章では、Inatsu and Imori (2018) で提案された  $C_p$  型のモデル選択規準 PMSEG の導出時の仮定を変更し、尺度パラメータと相関パラメータを未知として同様の  $C_p$  型モデル選択規準の導出を行った。未知である尺度パラメータと相関パラメータは局外パラメータであるが、推定したい回帰係数の推定をするためにはこれらの推定を行う必要がある。未知パラメータの推定の影響をモデル選択に反映させるため、Inatsu and Imori (2018) で提案されたモデル選択規準における正則条件とは別に、いくつかの正則条件を付加することで、尺度パラメータと相関パラメータの推定の影響を無視できることを示した。さらに、パラメータの推定量としてモーメント推定量を用いることで付加的な正則条件を満たすことも示した。数値実験では統計解析ソフト R を用いて、作業用相関行列として真の相関行列を用いた場合、誤った相関行列を用いた場合のモデルの選択確率と予測誤差について比較を行った。真の相関行列を用いたほうが誤った相関行列を用いた時より予測誤差が小さくなった。さらに、説明変数と同時に作業用相関行列も選択した場合の選択確率と予測誤差もシミュレーションした。説明変数と同時に作業用相関行列も選択した場合のほうが、作業用相関行列として真の相関行列を用いて変数選択した場合より、予測誤差が小さくなることも示した。この数値実験により、作業用相関行列を決め打ちで用いるより、複数の相関構造を用いて、説明変数と作業用相関行列を選択することで予測誤差を小さくでき、予測精度を向上させることが期待できることがわかった。さらに、一般的な統計解析ソフト R や SAS などにおいて GEE 法ではモーメント推定量が用いられており、実用面においても有用であることを主張した。

第 4 章では、経時データにおける個体数と同時に観測時点数も大きくなる場合の GEE 推定量の漸近性質および PMSEG の漸近バイアスについていくつかの証明を行った。GEE 推定量の漸近性質の証明のためにいくつかの正則条件を追加し、漸近オーダーを示した。GEE 法において作業用相関行列が正しいか否かによらず  $m/\sqrt{n} \rightarrow 0$  となる状況で GEE 推定量の一致性が示せた。一方で、真の相関行列を用いた場合は標本数が大きくなれば、観測時点数の発散速度によらず GEE 推定量の一致性が示せた。さらに GEE 推定量の漸近オーダーを用いて PMSEG の漸近オーダーも導出した。作業用相関行列が正しいか否かによらず、 $m^4/n \rightarrow 0$  となる状況で候補モデルに依存しないバイアスを除き漸近オーダーが 0 となることを示した。さらに真の相関行列を用いた場合は  $m^2/n \rightarrow 0$  となる状況で、候補モデルに依存しないバイアスを除き漸近オーダーが 0 となることが示せた。数値実験では、作業用相関行列として真のものを用いたときのほうが誤った相関行列を用いたときより予測誤差が小さいことが分かった。そして作業用相関行列が正しいか否かによらず漸近バイアスが 0 に収束することが見て取れた。