

Conditions for Robustness and Limitation on Bayesian Student- t Linear Regression Modeling

大阪公立大学 数学研究所 林由子

スチューデントの t 分布にもとづく頑健回帰モデルの構築を目的としたものである。これは、一般に **heavy-tail modeling** と呼ばれ、尤度に t 分布を用いることで外れ値を事前に識別することなく、自動的にモデルが外れ値を過小評価、あるいは棄却するモデルである。この研究では、極値理論に基づく手法を用いて、heavy tailed model を線形回帰分析に適用する場合の頑強モデルとなる十分条件の導出と頑強モデルが機能しない区間の提示を行っている。

孤立したデータは、必ずしも外れ値ではない。目的となる予測分布において、そのデータが生じる確率が極小さいときに外れ値といえる。つまり外れ値は予測分布に応じて定まる。しかし、予測モデルを定めた後にしか外れ値を認定できない。たとえば図1の観測値では、左図、右図ともに一つ観測値が他の観測値とは孤立している。2変数の相関とみた場合には、いずれも相関関係を損なう外れ値である。これに対して回帰分析を行う観測値としてみた場合には、右図では、孤立した観測値が回帰直線（直線）から大きく離れるのに対し、左図では、孤立した観測値は回帰直線の極めて近くに存在する。申請者は、この問題について、線形回帰分析における外れ値とは何かについての定義を行った。つまり、右図のように回帰直線から孤立したデータが離れていくものを外れ値であると定義し、この外れ値に対して事後分布において頑健となるための条件をてこ比を用いて示すことに成功した。これらの条件により、図1の右図のような外れ値が出現する存在範囲を明らかにした。

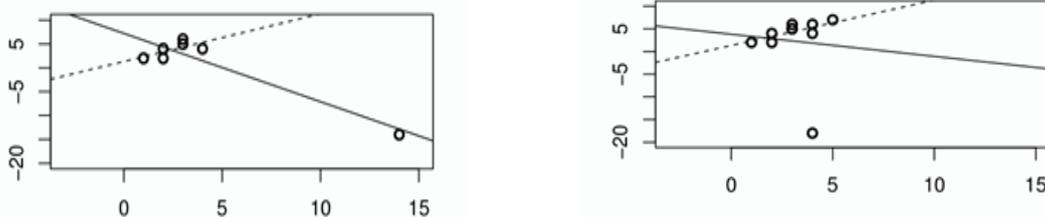


図1 回帰直線における外れ値の影響

さらに、上記の条件を満たす外れ値に対して、Andrade and O'Hagan の平均値に対する頑健モデルを応用することで、スチューデントの t 分布を用いた線形回帰モデルが頑健となる十分条件を示した。