

スパースな疫学研究データ解析におけるベイズ的回帰分析手法の比較
Comparison of Bayesian Regression Methods for Sparse Epidemiologic Data Analysis

田島 里華
Rika Tajima

指導教員：松山 裕 教授
Tutor: Professor Yutaka Matsuyama

東京大学大学院医学系研究科公共健康医学専攻生物統計学分野
Department of Biostatistics, School of Public Health,
Graduate School of Medicine, the University of Tokyo

抄録

日本人における循環器疾患発症に対するリスク因子の影響の定量化を目的とした大規模疫学研究である JALS-ECC (Japan Arteriosclerosis Longitudinal Study Existing Cohort Combine, 日本動脈硬化縦断研究 0 次統合研究) にて Tanabe et al. らは男女を合わせて心筋梗塞 (MI) 発症に対する総コレステロール (TC) の影響の大きさを調べたが、女性のみ限定して同様の解析を試みると、女性の MI 発症が少ない、喫煙という確立したリスク因子について曝露人数が少なく分布に偏りがある、調整変数の数が多い等複合的な理由により、ロジスティック回帰によるオッズ比の推定値が喫煙に関しては 0 という非現実的な値となり、他の変数でも信頼区間幅が非常に広くほぼ情報量をもたなかった。イベント数、調整変数の数や分布、影響の強さ等の兼ね合いにより、調整変数で層別した際人数の少ないセルが生じるデータはスパースデータと呼ばれ、それを回帰分析した結果は信頼性が低いと問題視されている。医学研究では、対象疾患やリスク因子による影響に関する見地が研究実施時点で既存の場合が多く、これを解析に反映させたベイズ流回帰分析を用いることでスパースデータ解析の性能が改善されるかを検討することが本研究の目的である。シミュレーション実験にて、事前分布の情報量および事後分布からパラメータ推定を行う際の積分計算法の異なる 4 つのベイズ的ロジスティック回帰のスパースデータ解析における性能を、通常の方法と比較した。その結果、事前分布の持つ情報量が増えるほど結果のバイアスは増加したが MSE が小さくなり、信頼幅の平均長も短縮された。つづいて JALS 0 次研究の女性対象者について、異なる事前情報を事前分布に取り入れたベイズ的ロジスティック回帰を行い、MI 発症に対する TC の影響の大きさを調べた。結果として、ベイズ的ロジスティック回帰を行うことで喫煙のオッズ比の推定値は現実的な値となり、その他変数についても信頼区間幅が縮まり臨床的に解釈可能な程度まで結果が補正され、その安定化の度合いは事前情報の量と共に増すことが確認された。