

一般化線形混合モデルにおけるランダム効果の分布の誤特定による 推定への影響

塩野義製薬（株）・解析センター 坂口 弘樹
大阪大・基礎工 狩野 裕

経時測定データとは、同一個体に対して複数時点で時間を追って測定されるデータである。臨床データや、人間や動物の成長を記録したデータなどがそれに当てはまる。経時測定データに対するモデルとして、線形混合モデル (Linear mixed model(LMM), [4]) や一般化線形混合モデル (Generalized linear mixed model(GLMM), [2]) が用いられる。LMM や GLMM は、時間経過による測定値の変化が処置群と対照群の間で差異があるかという経時測定データに対する興味に応えることができるモデルである。さらに、モデルに組み込まれたランダム効果と呼ばれる潜在変数によって個体間におけるばらつきを表現することで、経時測定データに見られる個体差を柔軟にとらえることも可能である。

これらのモデルによってパラメータの最尤推定を行う際、通常はランダム効果は正規分布に従うという仮定が置かれることが多い。しかし、ランダム効果は実際には観測されない変数であるため真の分布は未知である。そのため、正規分布の仮定は必ずしも正しいとは限らず、分布を誤って特定してしまう恐れがある。

LMM や GLMM におけるランダム効果の分布の誤特定に関する問題に対して、Verbeke and Lesaffre (1997)[3] は、LMM においてはランダム効果の分布を正規分布と誤特定した場合でも最尤推定量 (MLE) に一致性があることを証明している。また、Litière et al. (2008)[1] はロジスティック回帰モデルを用いた GLMM においてランダム効果の分布を誤特定した場合の MLE の挙動を数値実験で確認し、誤特定の下で MLE に一致性が無くなることを指摘している。その他にもランダム効果の分布の誤特定に触れた文献はあるが、GLMM における MLE の性質については数値実験のみの考察にとどまっており、理論面の考察も含めたより詳しい言及はされていない。

そこで、本報告では GLMM、特にロジスティック回帰モデルにおけるランダム効果の分布を誤特定した場合の MLE の性質に関する理論的考察を行う。具体的には、特定の状況下ではランダム効果の分布を誤特定した場合でも、MLE によってロバストな推定が可能となるパラメータと一致推定が可能となるパラメータが存在することを示す。

参考文献

- [1] Litière, S., Alonso, A. and Molenberghs, G. (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine*, **27**, 3125-3144.
- [2] McCulloch, C., Searle, S. and Neuhaus, J. (2008). *Generalized, Linear, and Mixed Models (second editon)*. John Wiley & Sons, New Jersey.
- [3] Verbeke, G. and Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis*, **23**, 541-556.
- [4] Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer.
- [5] 坂口弘樹 (2015). ランダムでない欠測と shared-parameter model に関する研究. 大阪大学修士論文.