

階層尤度に基づく経時測定データと二値アウトカムデータに対する ジョイントモデリング

横浜市立大学 医学部 三角 俊裕

1. 概要

経時測定データと二値アウトカムデータの関係性を調べることを目的として、階層尤度 (Lee and Nelder, 1996) に基づくジョイントモデリング手法を提案する。これまでのジョイントモデルの文脈では、経時測定データおよび二値アウトカムデータに対して、共通の変量効果を含むサブモデルをそれぞれ仮定した上で、同時尤度に基づいて推定を行う方法が提案されている。本報告では、階層尤度を用いることで推定アルゴリズムにおける数値積分が不要となることを示し、数値実験を通して提案手法によるパラメータの推定精度の検討を行う。

2. 同時尤度に基づくジョイントモデリング

ジョイントモデルは、経時測定データの挙動と二値アウトカムの関係を、 n 組の観測データ $\{x_{ij}, y_i, z_i; i = 1, \dots, n, j = 1, \dots, J_i\}$ に基づいてモデル化する。ここで、 x_{ij} は第 i 個体の j 番目の時点 t_{ij} における観測値、 y_i は二値アウトカム、 z_i はベースライン共変量である。このとき、経時測定データに対しては線形混合効果モデル、二値アウトカムデータに対してはロジスティックモデルを、それぞれ単純なサブモデルとして、

$$x_{ij} = \gamma_1 z_i + \gamma_2 t_{ij} + b_i + \varepsilon_{ij}, \quad (1)$$

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \alpha z_i + \beta b_i, \quad (2)$$

と仮定する。ここで、(1) 式のサブモデルにおいて、 γ_1 と γ_2 はそれぞれ z_i と t_{ij} に対する固定効果、 b_i は平均 0、分散 σ_b^2 の正規分布に従う変量効果、 ε_{ij} は互いに独立に平均 0、分散 σ_ε^2 の正規分布に従う測定誤差である。(2) 式のサブモデルにおいて、 $\pi_i = P(y_i = 1)$ 、 α は z_i に対する固定効果である。また、 β は関連パラメータと呼ばれ、経時測定データの挙動が二値アウトカムに与える影響の強さを表す。これらのサブモデルに対して、以下の同時尤度に基づいてモデル推定を行う。

$$\prod_{i=1}^n \left[\int_{-\infty}^{\infty} \left\{ \prod_{j=1}^{J_i} f_{\gamma_1, \gamma_2, \sigma_\varepsilon^2}(x_{ij} | b_i) \right\} f_{\alpha, \beta}(y_i | b_i) f_{\sigma_b^2}(b_i) db_i \right] \quad (3)$$

本推定法は、経時測定データの測定誤差が二値アウトカムに影響しないという仮定の下での同時推定であるため、データ毎に二段階でモデル推定を行う方法に比べて、関連パラメータ β を高い精度で推定できることが知られている。一方で、推定アルゴリズムにおいて数値積分が必要となるため、仮定するサブモデルの複雑化に伴い計算コストの問題が生じる。

3. 階層尤度に基づくジョイントモデリング

上記のサブモデルに対して、階層尤度は以下で与えられる。

$$\prod_{i=1}^n \left[\left\{ \prod_{j=1}^{J_i} f_{\gamma_1, \gamma_2, \sigma_\varepsilon^2}(x_{ij} | b_i) \right\} f_{\alpha, \beta}(y_i | b_i) f_{\sigma_b^2}(b_i) \right] \quad (4)$$

この階層尤度は、(3) 式の同時尤度のように周辺化を行わないため、推定アルゴリズムにおける数値積分を必要としない。より複雑なモデリングの詳細と数値実験の結果は当日報告する。

参考文献

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society. Series B* 58:619–678.