

Mixed effect models for nonnormal responses

飯島勇人^{*†§}

2007年7月4日

目次

この章で扱う内容	1
10.1 Generalized linear mixed models	2
10.2 Generalized estimating equations (一般化推定方程式)	2

この章で扱う内容

一般化線形混合モデル (Generalized linear mixed model: GLMM) の実行方法、原理の紹介、一般化推定方程式について。要所要所で訳者の小うるさいツッコミ付き^^;)。

定式化

繰り返しになるが、前章までで扱ってきた Random effect (γ) は平均 0、分散 D の正規分布と仮定され、 V の関数として表される。そして、Fixed effect が入った行列を x 、Random effect が入った行列を z とすると、 $\theta_i = x_i\beta + z_i^T\gamma$ で表現される (θ は GLM において予測値にリンク関数をかけたもの。6 章参照)。

Random effect を含むモデルの尤度の一般形は、以下のように表現される。

$$L(\beta, \phi, V|y) = \prod_{i=1}^n \int f(y_i|\beta, \phi, \gamma)h(\gamma|V)d\gamma$$

..... なのだが、この解を求める (最尤推定する) のは、すごい大変。あの手この手は開発されているが、完全なコンセンサスが得られているものはまだないだろう。

* 北海道大学大学院農学院専門研究員

† 連絡先: hayato-i@for.agr.hokudai.ac.jp または http://www.geocities.jp/iiijima_web/index.html

‡ 本ゼミのサポートページ: <http://www7.atwiki.jp/hayatoiiijima/pages/33.html>

§ この文章は L^AT_EX で作成しました。わーどなどではけっしてありません。

10.1 Generalized linear mixed models

GLMM は Random effect を持った GLM である。通常 Random effect は平均 0、分散 D である正規分布をした変数として扱われる。

*この後はデータを使った例。読んでもあんまりいいことはないので省略。

R における GLMM の実行方法

いろいろ。以下にまとめた。

- glmmPQL: 罰則つき擬似尤度による方法。モデル選択できんし推定値の精度もあまりよろしくないの、lmer() が使える今はあまりお勧めではない。
- lmer: Laplace 近似による方法。2 つ以上の Random effect をモデルに入れるならこれしかない。
- (この本では紹介していないが) glmmML: family は binomial と poisson に限定。Random effect も 1 個だけ。ただしモデル選択が簡単なので、これらの条件が適合するならお勧め。
- (この本では紹介していないが) aod: ベータ分布や負の二項分布の場合に使える。Random effect は 1 つだけ。

10.2 Generalized estimating equations (一般化推定方程式)

GEE とは何か

- 応答変数の分布が仮定できないときに使うモデル (リンク関数と分散だけ与えればよい)。
- つまり、擬似尤度を使うモデリング方法。
- 繰り返し測定や時系列データにおける、単一処理内での連続測定による相関を変数として組み込むことができるので、これらのデータのモデリングに有効。
- Random effect は指定できない。
- 推定される係数の値は比較的ロバストである (ほんまかいな)。
- R の中では、gee パッケージ内の gee() で実行できる。
- GLMM はあくまで各個体に対して係数推定を行うが、GEE では反復測定となるある一つの単位 (ある 1 処理で、その中には複数の測定値がある) に対して算出される。
- (訳者から一言) 尤度が構築できんのは致命的だと思うが。

gee() の使い方

- 反復して測定する単位は、gee() 内で id=以降に指定する。
- 反復測定値の相関のパターンは corstr=" "内で指定する必要がある。どの測定値間でも同程度の相関が予想されるなら exchangeable を指定する。

解析例: 薬の投与による発作回数の変化

- 問題設定: 痙攣症を持つ患者に対し、薬を与える群と偽の薬(プラセボ)を与える群を設け、薬を与える前後での発作回数を測定。薬投与前の発作回数は8週間分のデータ、投与後の発作回数は、2週間ごとに4回(質問によって?)データを得て算出。
- 解析前のデータ眺め 1: 薬投与前後での発作回数の違いの平均値を見る(p. 206 下部)。薬を与えても発作回数は変わらず、プラセボ群では増加していたことから、薬以外の要因が発作回数を増加させている疑い。
- 解析前のデータ眺め 2: 1を踏まえ、薬投与後の時間と発作回数の関係、投与前と投与後の発作回数の関係を図示化(p. 207 上部)。結局薬群とプラセボ群間で傾向の違いは認められず。
- 解析: `gee()`を使い、薬そのものの影響と、医者処方という行為、それらの交互作用の影響を検討。繰り返し測定の単位は個人を示す `id` を使用。
- `corstr`: 最も近い測定値間ほど相関が強くなると仮定し、"AR-M"を指定。
- 解釈: 薬そのものは大して効いていない。

以下、訳者の勝手な解釈(まとめ?)

反復、繰り返し測定結果のモデリング手法として GEE は使えるかも。でも今なら `lmer()` とか `glmmML()` で反復、繰り返し単位となるものを Random effect に投入する方がいい気がする。ただし、モデルがすでに完全に固まっている場合は、相関のパターンを決めて指定できる GEE は注目に値する。