

経時測定データに対する一般化線型モデルに基づいた反応プロファイルの分類

○富金原悟¹⁾²⁾, 菅波秀規³⁾⁴⁾, 山田大志⁵⁾, 山口拓洋¹⁾

- 1) 東京大学医学部附属病院 臨床試験データ管理学
- 2) 小野薬品工業株式会社 臨床開発管理部統計解析課
- 3) 興和株式会社臨床解析部
- 4) 東京理科大学大学院工学研究科
- 5) アストラゼネカ株式会社 臨床統計・プログラミング部

1

発表内容

- 混合分布モデルの背景
- 一般化線型モデルの適用
- プログラムアルゴリズム
- 計算結果
- まとめ

2

発表内容

- 混合分布モデルの背景
- 一般化線型モデルの適用
- プログラムアルゴリズム
- 計算結果
- まとめ

3

本研究の対象となる試験の概要

— 試験デザイン —

対照薬に対する優越性を検証するための
第Ⅲ相 2重盲検比較試験.

- 目的** I 種の薬効に対して作用の比較を行なう.
- 対象患者** 各群 $i(i=1, \dots, I)$ に対して無作為に割付けられた患者 n_i とする.
- 測定方法** 治療前のベースラインと,それ治療以降に観察された T 個の測定時点で反応を観察する (合計 $T+1$ 回の測定)

4

経時測定データ

同じ治療でも被験者によって反応が大きく異なることが多い

治療に反応して評価指標が改善方向へ動く者もいれば、変化があまり見られない者、さらには、悪化の方向へ動く者もいる。

群毎の平均的な変動の大きさを比較する

「時点×個体」交互作用を検討する

「群×時点」の交互作用を検討するため、群毎の平均的なトレンドを比較する。

評価指標の変動パターンの類似度により個体を分類する方法

分類結果を基に群間比較を行う

- ・混合分布モデル (丹後 1989)
- ・潜在クラスモデル (Skene ら 1992)
- ・交互作用に関する多重比較法 (広津 1989)

混合分布モデル (丹後 1989)

特徴 「時点×個体」交互作用を検討するために、投与開始時点からの検査値の経時変動パターンが、「改善群」、「不変群」、「悪化群」などの分類群として各薬剤群に混在していると仮定した混合分布に基づき、個体を分類する。

個体の分類 (2剤比較, 3分類にて)

基本式

$$D_{ijt} = \mu(t) + \varepsilon_{ijt}$$

D : 経時測定データ(変化量)

i : 治療群 ($i=1, 2$)

j : 被験者 ($j=1, \dots, J$)

t : 測定時点 ($t=1, \dots, T$)

ε : 誤差

改善群 $\mu(t) = \mu_1(t) = \sum_{k=1}^{R_1} \beta_{1k} t^k \quad (< 0)$

不変群 $\mu(t) = \mu_0(t) = 0$

悪化群 $\mu(t) = \mu_2(t) = \sum_{k=1}^{R_2} \beta_{2k} t^k \quad (> 0)$

混合分布モデル

確率密度関数

$$h_i(d_{ij}|\theta) = p_{i0}f_0(d_{ij}) + p_{i1}f_1(d_{ij}) + p_{i2}f_2(d_{ij})$$

$$p_{i0}, p_{i1}, p_{i2} (p_{i0} + p_{i1} + p_{i2} = 1)$$

← 密度関数の和

「不変群」, 「改善群」, 「悪化群」での D_{ij} の同時確率密度関数

$$f_g(d_{ij}) = \left(\frac{1}{2\pi\sigma^2}\right)^{T/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^T (d_{ijt} - \mu_g(t))^2\right\}$$

g : 分類群

1=不変群, 2=改善群, 3=悪化群

← 密度関数の積

個体 j が分類群 g へ帰属する事後確率 (相対帰属確率)

$$Q_{ij}(g) = p_{ig} f_g(d_{ij}) / h_i(d_{ij}|\theta) \quad (i=1, 2; g=0, 1, 2; j=1, 2, \dots, n_i)$$

7

混合分布モデル

パラメータの推定

$$\text{パラメータ } \theta = \{p_{ig}, \beta_g, \sigma^2\}$$

混合分布の混成割合

$$\hat{p}_{ig} = \sum_{j=1}^{n_i} Q_{ij}(g) / n_i$$

分散

$$\hat{\sigma}^2 = \sum_{i=1}^2 \sum_{j=1}^{n_i} \sum_{g=0}^2 Q_{ij}(g) \left\{ \sum_{t=1}^T (Y_{ijt} - \mu_g(t))^2 \right\} / (n_1 + n_2) T$$

回帰パラメータ (β)

尤度方程式

$$L = \prod_{i=1}^2 \prod_{j=1}^{N_i} h_i(d_{ij}|\theta)$$

$$\ln L = \sum_{i=1}^2 \sum_{j=1}^{N_i} h_i(d_{ij}|\theta)$$

$$\frac{\partial \ln L}{\partial \beta_g} = 0$$

$$\beta_g = \frac{1}{N_1 p_{1g} + N_2 p_{2g}} \left(\sum_{t=1}^T (t_g t'_g) \right)^{-1} \left(\sum_{d=1}^2 \sum_{j=1}^{N_i} Q_{ij}(g) \sum_{t=1}^T d_{ijt} \right) t_g$$

8

混合分布モデル

EMアルゴリズム

Step0 : 薬剤群 i ($=1, 2$) の個体 j ($=1, \dots, n_i$) が「不変群 $g=0$ 」, 「悪化群 $g=1$ 」, 「改善群 $g=2$ 」に属する 相対的帰属確率 $Q_{ij}(g)$ の初期値を与える.

Step1 [M-step] : $Q_{ij}(g)$ が与えられたとの条件でパラメータの最尤推定値を次式で計算する

$$\hat{p}_{ig} = \frac{\sum_{j=1}^{n_i} Q_{ij}(g)}{n_i} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^2 \sum_{j=1}^{n_i} \sum_{g=0}^2 Q_{ij}(g) \left\{ \sum_{t=1}^T (Y_{ijt} - \mu_g(t))^2 \right\}}{(n_1 + n_2)T}$$

$$\beta_g = \frac{1}{n_1 p_{1g} + n_2 p_{2g}} t_g' \left(\sum_{i=1}^2 \sum_j Q_{ij} \sum_{t=1}^T Y_{ijt} \right) t_g$$

$$t_g' = [t \ t^2 \ \dots \ t^{R_g}]$$

Step2 [E-step] :

パラメータ $\theta = \{p_{ig}, \beta_{gk}, \sigma^2\}$ の推定量を利用して相対的帰属確率 $Q_{ij}(g)$ を計算し、M-step と E-step が収束するまで繰り返す。

9

発表内容

- 混合分布モデルの背景
- 一般化線型モデルの適用
- プログラムアルゴリズム
- 計算結果
- まとめ

10

一般化線型モデル

従来の混合分布モデル

評価指標: 計量値
モデル: 低次の線型多項式

適用データに制限有り

臨床試験

評価指標: 2値応答(改善・非改善),
順序データ(1,2,3,4…), etc.

一般化線型モデル(Nelder, Wedderbur:1972)の適用

指数型分布族を用いることで, 正規分布の性質を活用できる.

説明変数と目的変数の関係が, 非線形のリンク関数を用いることで
単純な線型式に制限されなくなった.

11

一般化線型モデル

一般化線型モデルへの拡張

基本モデル $X_{ijt} = \mu(t) + \varepsilon_{ijt} = X_i^T \beta + \varepsilon$

X : 経時測定データ
 i : 治療群 ($i=1,2$)
 j : 被検者 ($j=1, \dots, J$)
 t : 測定時点 ($t=1 \dots, T$)
 ε : 誤差

リンク関数 $d(\mu_{ijt}) = X_i^T \beta + \varepsilon$,

確率密度関数

$$h_i(X_{ij}|\theta) = p_{i0}f_0(X_{ij}) + p_{i1}f_1(X_{ij}) + p_{i2}f_2(X_{ij})$$

$p_{i0}, p_{i1}, p_{i2}(p_{i0}+p_{i1}+p_{i2}=1)$: 相対帰属確率

「改善群」, 「不変群」, 「悪化群」での X_{ij} の同時確率密度関数 (指数分布族)

$$f_g(X_{ij}; \theta_{ij}) = \exp \left[\sum_{t=1}^T \{ a(X_{ijt}) b(\theta_{ijt}) + c(\theta_{ijt}) + d(X_{ijt}) \} \right]$$

12

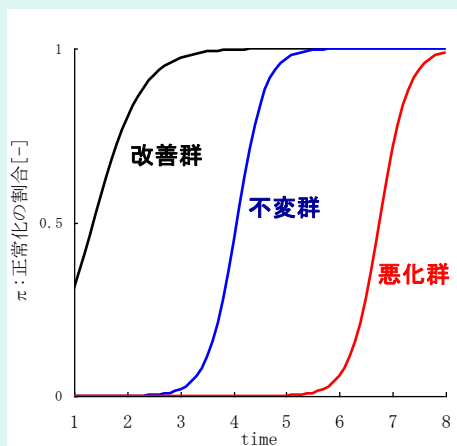
一般化線型モデル

指数分布族としてのポアソン分布, 正規分布, 2項分布

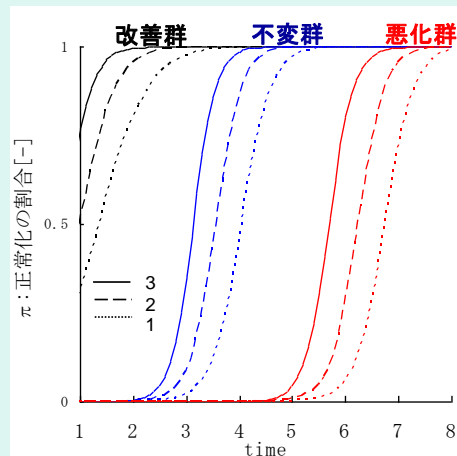
分布	b	c	s
ポアソン分布	$\log(\theta)$	$-\theta$	$-\log y!$
正規分布	$\frac{\mu}{\sigma^2}$	$-\frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)$	$\frac{y^2}{2\sigma^2}$
2項分布	$\log\left(\frac{\pi}{1-\pi}\right)$	$n\log(1-\pi)$	$n\log\binom{n}{y}$

13

分類群毎の反応パターン



2値応答(正常・異常)に対する分類群
(改善群, 不変群, 悪化群)の反応パターン



スコア(1,2,3)に対する分類群
(改善群, 不変群, 悪化群)の反応パターン

一般化線型モデル

例) logistic回帰

経時的 ($t=1, \dots, T$) に測定された, 正常・異常の2値応答データ (M)

$$\pi_{ig}(t) = \frac{\exp(\beta_{ig0} + \beta_{ig1}t)}{1 + \exp(\beta_{ig0} + \beta_{ig1}t)}$$

リンク関数 $\mu_g(t) = \log\left(\frac{\pi_{ig}}{1 - \pi_{ig}}\right) = \beta_{0g} + \beta_{1g}t = t_i^T \beta \quad (g = 0, 1, \dots, G)$

初期条件 $\sum_{t=1}^T M_{t2} \leq \sum_{t=1}^T M_{t0} \leq \sum_{t=1}^T M_{t1} \quad (t = 1, \dots, T)$

M : 経時測定データ (1=正常, 0=異常)

確率密度関数

$$h_i(M_{ij}|\theta) = p_{i0}f_0(M_{ij}|\theta) + p_{i1}f_1(M_{ij}|\theta) + p_{i2}f_2(M_{ij}|\theta)$$

「改善群」, 「不変群」, 「悪化群」での同時確率密度関数

$$f_g(M_{ij}) = \exp\left[\sum_{t=1}^T \left\{ M_{ijt} \log \pi_{ijt} + (n_{ijt} - M_{ijt}) \log(1 - \pi_{ijt}) + \log\left(\frac{n_{ijt}}{M_{ijt}}\right) \right\}\right]$$

n : 対象例数

15

一般化線型モデル

パラメータ推定 (β)

尤度方程式 $L = \prod_{i=1}^2 \prod_{j=1}^{N_i} h_i(\pi_{ij}|\theta) \quad \ln L = \sum_{i=1}^2 \sum_{j=1}^{N_i} h_i(\pi_{ij}|\theta)$

$$\frac{\partial \ln L}{\partial \beta_g} = 0 \Rightarrow \left\{ \sum_{i=1}^2 \sum_{j=1}^{N_i} Q_{ij}(g) \right\} \sum_{g=0}^2 M_{ijt_g} = \left\{ \sum_{d=1}^2 \sum_{j=1}^{N_d} Q_{ij}(g) \right\} \left\{ \sum_{g=0}^2 n_{ij} \mu_g t_g \right\}$$

$$\sum_{t=1}^T \mu_g t_g = \frac{1}{n_{ij}(N_0 p_{0g} + N_1 p_{1g} + N_2 p_{2g})} \left(\sum_{d=1}^2 \sum_{j=1}^{N_d} Q_{ij}(g) \sum_{t=1}^T M_{ijt} \right) t_g$$

各分類群の混成割合の係数

上記式の反復演算により β を算出する.

16

発表内容

- 混合分布モデルの背景
- 一般化線型モデルの適用
- プログラムアルゴリズム
- 計算結果
- まとめ

17

IMLによるEM Algorithm

$$y = X\beta + \varepsilon$$

value	: X_{ijt}
treatment	: $i=1, 2$
subject	: $j=1, \dots, n_i$
time point	: $t=0, 1, \dots, T$
divided group	: $g=0, g=1, g=2$
change	: $d_{ijt} = X_{ijt} - X_{ij0}$

①初期値

$$S_{ij} = \sum d_{ijt} \quad \begin{cases} \text{if } S_{ij} \leq -100 & Q_{ij}(1)=1 \\ \text{if } S_{ij} \geq 100 & Q_{ij}(2)=1 \\ \text{if } -100 < S_{ij} < 100 & Q_{ij}(0)=1 \end{cases}$$

②Maximization Step

$$\theta = \{p_{ig}, \beta_{gk}, \sigma^2\}$$

$$g_i(d_{ij}|\theta) = p_{i0}f_0(d_{ij}) + p_{i1}f_1(d_{ij}) + p_{i2}f_2(d_{ij})$$

$$\bar{p}_{ig} = \sum_{j=1}^{N_i} Q_{ij}(g) / N_i \quad \log L / \partial \beta = 0 \Rightarrow \beta$$

$$\bar{\sigma}^2 = \sum_{i=1}^2 \sum_{j=1}^{N_i} \sum_{g=0}^2 Q_{ij}(g) \left\{ \sum_{t=1}^T (d_{ijt} - \mu(t))^2 \right\} / (N_1 + N_2)T$$

③Expectation Step

$$Q_{ij}(g)$$

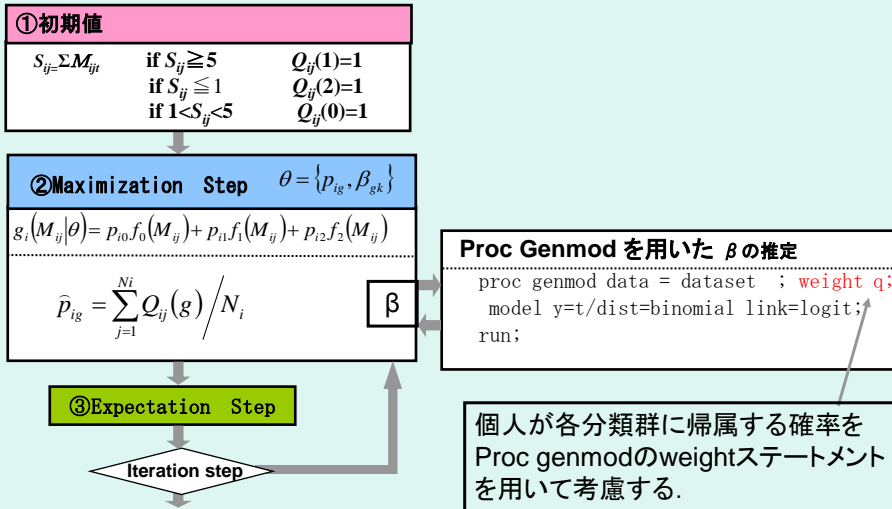
$$Q_{ij}(g) = p_{ig} f_g(d_{ij}) / g_i(d_{ij}|\theta) \quad (i=1, 2; g=0, 1, 2; j=1, \dots, N_i)$$

Iteration step

$$\theta = \{p_{ig}, \beta_{gk}, \sigma^2\}$$

18

Genmodを用いたEM Algorithm



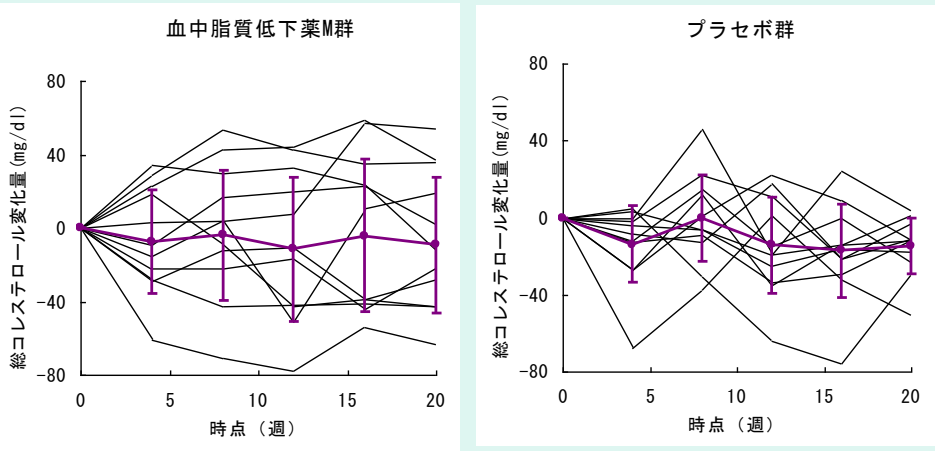
19

発表内容

- 混合分布モデルの背景
- 一般化線型モデルの適用
- プログラムアルゴリズム
- 計算結果
- まとめ

20

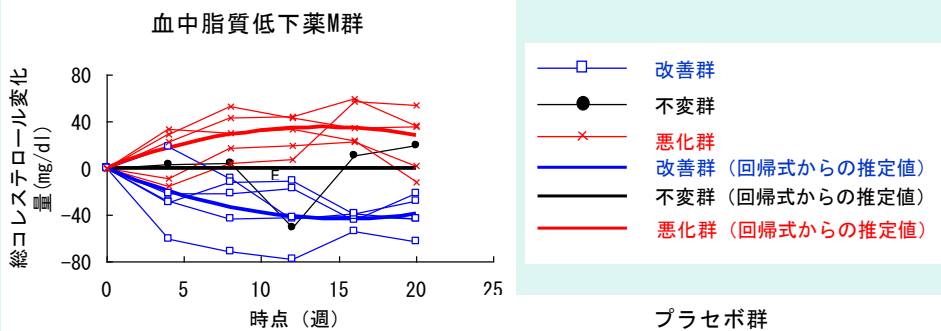
データ解析(血中脂質低下薬M: 広津 1989)



被験者分類毎の総コレステロールの推移

21

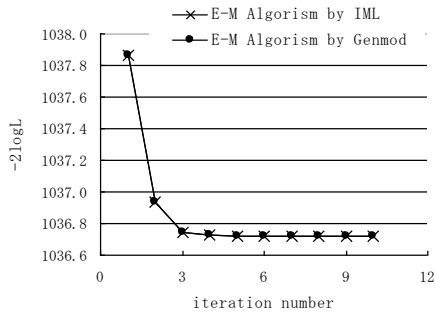
データ解析(血中脂質低下薬M)



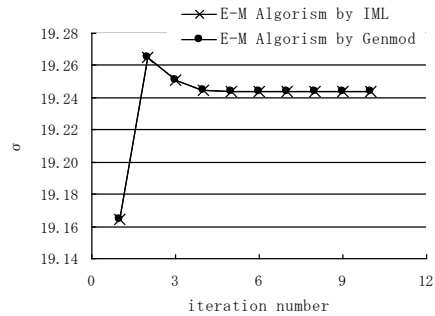
薬剤	血中脂質低下薬M	プラセボ
改善群	6 (50.0%)	2 (18.2%)
不変群	1 (8.3%)	9 (81.8%)
悪化群	5 (41.7%)	0 (0.0%)
合計	12	11

被験者分類毎の総コレステロールの推移

IMLとGenmodによるE-Mアルゴリズムの推定精度



-2 Log Lと収束回数の関係



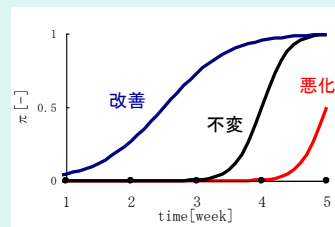
σと収束回数の関係

23

2値応答の解析

シミュレーションデータ

観測データ : 1=改善, 0=非改善
 測定時点 : 1, ..., 5
 治療群 : 1=実薬, 2=プラセボ
 例数 : 各群50例(合計100例)
 Logit関数の次数: 1次回帰

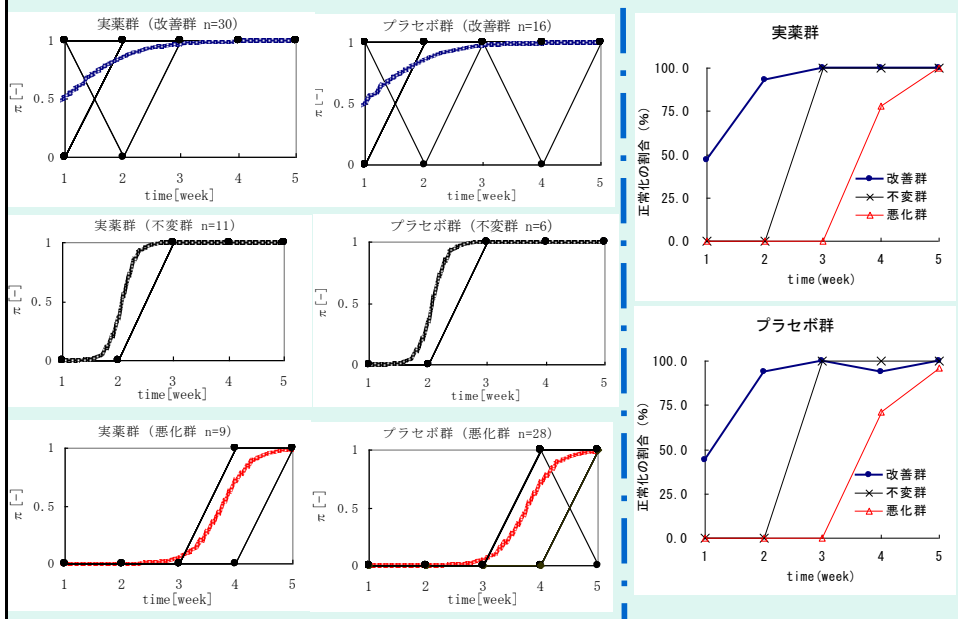


シミュレーションパラメータ

分類群毎のパラメータ		改善群	不変群	悪化群
係数	β_0	-25.0	-5.0	-20.0
	β_1	2.5	2.0	5.0
例数	実薬群	25	20	5
	プラセボ群	10	15	25
存在割合	実薬群	0.50	0.40	0.10
	プラセボ群	0.20	0.30	0.50

24

2項分布の経時測定データに対する分類結果



2項分布の経時測定データに対する分類結果

分類後のパラメータおよび相対帰属確率

事後分類群		改善群	不変群	悪化群
係数	β_0	-1.7	-16.2	-14.2
	β_1	1.8	7.8	3.8
例数	実薬群	30	11	9
	プラセボ群	16	6	28
相対帰属確率	実薬群	0.54	0.27	0.19
	プラセボ群	0.30	0.12	0.58
$-2 \log L$		338.3		

分類後の ΣM (正常化した時点数)の分布

		ΣM (正常化した時点数)				
		1	2	3	4	5
実薬群	改善群				18	12
	不変群			11		
	悪化群	2	7			
プラセボ群	改善群			1	9	6
	不変群			6		
	悪化群	9	19			

26

ま と め

<まとめ>

- 混合分布モデルに一般化線型モデルを適用した.
- EMアルゴリズムによるパラメータ推定を簡便に行うため Proc Genmodを使用した.
- 一般化線型モデルでの結果を適用可能性を確認した.

<今後の課題>

- イベントデータ, 順序カテゴリカルスコア等の対応.
- モデル分散でなく, ロバスト分散の検討を行う.
- 実データを用いた解析.