

## NLMIXEDプロシジャの紹介

岸本 淳司(慶応義塾大学SFC)

## 非線形混合モデル

- ◆ 非線形モデル
  - パラメタが非線形に入っているモデル
- ◆ 混合効果モデル
  - 固定(母数)効果とランダム(変量)効果とが同時に入っているモデル
- ◆ 非線形混合モデル
  - 固定効果とランダム効果とが非線形に入っているモデル
  - たいへん複雑なモデルだ！
  - なんか新しそうだ！（発想自体は昔からある）

## 応用例1: 薬物動態

- ◆ 濃度の変化は非線形(コンパートメントモデル)
- ◆ 個人の効果をランダム効果とする
- ◆ 個人ごとにモデルを当てはめるのではなく、集団としてのモデルを推定する
- ◆ Population PK
- ◆ ブリッジングに使う!

3

## 応用例2: Overdispersion のある二項分布

- ◆ 二項分布では平均が決まると分散も決まる
- ◆ 実際には平均から期待されるより広い分散をもつデータが観測される(Overdispersion)
- ◆ 二項分布では、すべての個体は同一のパラメタを持つと仮定している
- ◆ 個体ごとにパラメタの違いを認めて、それが分布していると仮定すると、Overdispersionがうまく説明できる

4

## NLMIXED指定の基本方針

- ◆ 説明変数の値を与えたときのデータの分布(条件付分布)を記述する
  - DATAステップに類似したプログラミングステートメントを使う
- ◆ 条件付分布には、正規分布はもちろん、二項分布やポアソン分布が、さらには対数尤度関数を指定することによって任意の分布が指定できる
- ◆ ランダム効果について分布を記述する
  - ランダム効果に指定できるのは正規分布のみ

5

## NLMIXED推定の基本方針

- ◆ ランダム効果について積分した最尤推定を行う
  - 非線形モデルではこの積分計算が厄介
  - ここに **Adaptive Gaussian Quadrature** を採用
  - **NONMEM** の1次法もサポート(正規分布のみ)
- ◆ 最適化計算のテクニックはマニアックに揃っている
  - SAS/OR の NLP プロシジャからコードを借用
- ◆ 柔軟に指定できることを重視
  - 定型的使用にはテンプレートの開発が必要

6

## MIXEDプロシジャとの比較

- ◆ 線形混合モデルのREMLに相当する方法はない
- ◆ データの分布として正規分布以外の二項分布やポアソン分布も利用できる
- ◆ 誤差の分布形の指定(AR1とかCSなど)はできない
- ◆ 線形混合モデルの「ランダム係数モデル」の非線形への拡張である

7

## 正規分布の例(オレンジの木)

$$y_{ij} = [(b_1 + u_{i1}) / (1 + \exp[-(d_{ij} - b_2) / b_3])] + e_{ij}$$

- ◆ ロジスティック成長曲線の混合モデル
  - $y_{ij}$  : 測定値
  - $d_{ij}$  : 日付
  - $b_1, b_2, b_3$  : 固定効果
  - $u_{i1}$  : ランダム効果  $N(0, \sigma_u^2)$

8

## NLMIXEDの起動とパラメタ指定

```
proc nlmixed data=tree;
  parms b1=190 b2=700 b3=350
        s2u=1000 s2e=60;
```

- ◆ NLINプロシジャと同様
- ◆ 初期値はグリッド指定可

9

## モデルの指定

```
num = b1+u1;
ex = exp(-(day-b2)/b3);
den = 1 + ex;
```

- ◆ 複数ステートメントに分けて記述できる
- ◆ u1はランダム効果を示すパラメタ

10

## 分布の指定

```
model y ~ normal(num/den,s2e);
random u1 ~ normal(0,s2u) subject=tree;
run;
```

- ◆ 条件付分布をmodelステートメントで指定
  - ~記号を使うSASプロシジャは初めて
- ◆ ランダム効果の分布をrandomステートメントで指定
  - ランダム効果が複数ある場合は共分散を指定

11

## 出力(問題のサイズ)

### Dimensions

Observations Used	35
Observations Not Used	0
Total Observations	35
<u>Subjects</u>	<u>5</u>
Max Obs Per Subject	7
Parameters	5
<u>Quadrature Points</u>	<u>1</u>

12

## コメント

- ◆ Subjects の項目について
  - NLMIXEDにはCLASSステートメントがない
  - Subjectsでソートしていないといけない
  - ここで正しい数になっているかチェックすべし
- ◆ Quadrature Point について
  - ここでは u は線形関数なので、積分点は1でよい

## 出力(反復過程)

IterationHistory					
Iter	Calls	NegLogLike	Diff	MaxGrad	Slope
1	4	131.686742	0.805045	0.010269	-0.633
2	6	131.64466	0.042082	0.014783	-0.0182
3	8	131.614077	0.030583	0.009809	-0.02796
4	10	131.572522	0.041555	0.001186	-0.01344
5	11	131.571895	0.000627	0.0002	-0.00121
6	13	131.571889	5.549E-6	0.000092	-7.68E-6
7	15	131.571888	1.096E-6	6.097E-6	<u>-1.29E-6</u>

NOTE: GCONV convergence criterion satisfied.

## コメント

- ◆ slopeがゼロに近くなったら収束
- ◆  $10^{-6}$  がデフォルトの基準

15

## 出力(当てはまりの情報)

### Fit Statistics

-2 Log Likelihood	263.1
AIC (smaller is better)	273.1
BIC (smaller is better)	271.2
Log Likelihood	-131.6
AIC (larger is better)	-136.6
BIC (larger is better)	-135.6

16



## コメント

- ◆ この情報は、同じデータに対して複数のモデルを当てはめてみたとき、そのモデル間の比較に有用である

17

## 出力(パラメタ推定値)

Parameter	Parameter Estimates		DF	Value	Pr >  t	Alpha	Lower	Upper	Gradient
	Estimate	Standard Error							
B1	192.05	15.6473	4	12.27	0.0003	0.05	148.61	235.50	1.154E-6
B2	727.90	35.2472	4	20.65	<.0001	0.05	630.04	825.76	5.289E-6
B3	348.07	27.0790	4	12.85	0.0002	0.05	272.88	423.25	-6.1E-6
S2u	999.88	647.44	4	1.54	0.1974	0.05	-797.70	2797.45	<u>-3.84E-6</u>
S2e	61.5139	15.8831	4	3.87	0.0179	0.05	17.4153	105.61	2.892E-6

18

## コメント

- ◆ 標準誤差は、2次導関数を使ったWald流
- ◆ 自由度は4
  - オレンジの木5本 – ランダム効果の次元1
  - t分布とみなす
  - 正規分布とみなすより保守的
- ◆ 分散推定値の下限が負になっている
  - データの対数をとるとか平方根をとるともってあてはまりがよくなるはず

19

ロジスティックモデル・二項分布  
・Overdispersionのモデル

$$x_{ij} | u_i \sim \text{Binomial}(n_{ij}, p_{ij})$$

$$\eta_{ij} = \log \frac{p_{ij}}{1 - p_{ij}} = \beta_0 + \beta_1 t_j + u_i$$

- ◆ 固定効果: 治療法 j
- ◆ ランダム効果: 施設 i

20

## NLMIXEDの起動と初期パラメタの指定

---

```
proc nlmixed data=infection;  
  parms beta0=-1 beta1=1 s2u=2;
```

- ◆ 二項分布だから残差のパラメタはない

21

## モデルの指定

---

```
eta = beta0 + beta1*t + u;  
expeta = exp(eta);  
p = expeta/(1+expeta);
```

22

## 分布の指定

```
model x ~ binomial(n,p);  
random u ~ normal(0,s2u)  
subject=clinic;
```

- ◆ データの分布は二項分布を指定
- ◆ ランダム効果については、正規分布しか用意されていない

23

## オプションな計算

```
predict eta out=eta;  
estimate '1/beta1' 1/beta1;  
run;
```

- ◆ 予測値の計算は Empirical Bayes による
- ◆ パラメタの関数については、あらゆる表現が可能

24

## 出力(問題のサイズ)

### Dimensions

Observations Used	16
Observations Not Used	0
Total Observations	16
Subjects	8
Max Obs Per Subject	2
Parameters	3
<u>Quadrature Points</u>	<u>5</u>

25

## 出力(パラメタ推定値)

### Parameter Estimates

Parameter	Estimate	Standard Error	DF	t	Value	Pr >  t	Alpha	Lower	Upper	Gradient
Beta0	-1.1974	0.5561	7	-2.15	0.0683	0.05	-2.5123	0.1175	-3.1E-7	
Beta1	0.7385	0.3004	7	2.46	0.0436	0.05	0.02806	1.4488	-2.08E-6	
S2u	1.9591	<u>1.1903</u>	7	1.65	0.1438	0.05	-0.8554	4.7736	-2.48E-7	

- ◆ ランダム効果S2uの標準誤差は0overdispersionを考慮している
  - 普通のLogistic回帰分析ならもっと狭いはず

26

## ESTIMATEステートメントの効果

Additional Estimates								
Label	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper
1/beta	11.3542	<u>0.5509</u>	7	2.46	0.0436	0.05	0.05146	2.6569

- ◆ この推定値の標準誤差は 法で計算されている

27

## コンパートメントモデルの例

$$C_{it} = [(D k_{ei} k_{ai}) / (Cl_i (k_{ai} - k_{ei}))] [\exp(-k_{ei} t) - \exp(-k_{ai} t)] + e_{it}$$

$$Cl_i = \exp(\beta_1 + b_{i1})$$

$$k_{ai} = \exp(\beta_2 + b_{i2})$$

$$k_{ei} = \exp(\beta_3)$$

C:濃度 D:用量 k:定数 Cl:クリアランス

$\beta$  :固定効果 b:ランダム効果

28

## プログラム例

```
proc nlmixed data=theoph;
  parms beta1=-3.22 beta2=0.47 beta3=-2.45
        s2b1=0.03 cb12=0 s2b2=0.4 s2=0.5;
  cl = exp(beta1 + b1);
  ka = exp(beta2 + b2);
  ke = exp(beta3);
  pred = dose*ke*ka*(exp(-ke*time)
    -exp(-ka*time))/cl/(ka-ke);
  model conc ~ normal(pred,s2);
  random b1 b2 ~ normal([0,0],[s2b1,cb12,s2b2])
    subject=subject;
run;
```

29

## 結論

- ◆ 時代の要求に応える新しい道具が用意された
- ◆ モデルの扱いが柔軟で、応用範囲が広い
  - そのため、定型的作業のためにはテンプレートの開発が必要
- ◆ 使いこなすためにはかなり勉強する必要がある

30



*The Power to Know™*