

# SUGI25 at Indianapolis

## 参加報告

Advanced Tutorialのトピックスの1つ  
PROC NLMIXEDの紹介

国立がんセンター研究所  
がん情報研究部  
石塚直樹

## NLMIXEDプロシジャとは？

- 何のためのプロシジャ？  
「非線型混合モデル：nonlinear mixed models」  
をあてはめる
- 必要とされる状況は？  
薬物動態、  
Over dispersionのある二項分布のデータ

## 今までのプロシジャとは?

- 非線型回帰モデルの解析  
NLINプロシジャがあったけど、
- 混合モデルの解析  
MIXEDプロシジャ  
これを一般化するとNLMIXED
- GLIMMIX、NLINMIXマクロに近い  
推定方法が違う
- 非線型の最適化  
SAS/ORのNLPプロシジャに近い

## 統計的な特徴は?

- 変量効果について積分した尤度の近似  
を最大化する(最尤法)
- 積分の近似法  
adaptive Gaussian quadratureが使える  
first-order Taylor series approximation
- 最大化の方法  
デフォルト dual quasi-Newton algorithm.

## 統計的な特徴 つづき

- **標準誤差**  
対数尤度関数の2次導関数の行列に基づいたWaldタイプの近似、 $t$ 値は推定値 / 標準誤差
- **予測値の計算**  
変量効果に経験ベイズ推定値を用いる方法が可能
- **固定効果の任意の関数の推定**  
デルタ法により利用可能

## 数値例 その1

- オレンジの木 : Draper and Smith(1981)
- ロジスティック成長曲線の混合モデル  
: Lindstrom and Bates (1990), Pinheiro and Bates (1995)

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

$y_{ij}$  :  $i$ 番目の木の $j$ 番目の測定値  
 $d_{ij}$  :  $i$ 番目の日  
 $b_1, b_2, b_3$  : 固定効果  
 $u_{i1}$  : 変量効果  $N(0, \sigma_u^2)$   
 $e_{ij}$  : 誤差  $N(0, \sigma_e^2)$

## SASプログラム

```
proc nlmixed data=tree;
  parms b1=190 b2=700 b3=350 s2u=1000 s2e=60;
  num = b1+u1;
  ex = exp(-(day-b2)/b3);
  den = 1 + ex;
  model y ~ normal(num/den,s2e);
  random u1 ~ normal(0,s2u) subject=tree;
run;
```

未知パラメータの初期値の指定

モデルの左辺の分子

モデルの左辺の分母

変量効果の分散成分

誤差の分散成分

## SASアウトプット(1)

- **Specification**  
入力データセット、従属変数、変量効果の変数、それぞれの分布、個体(被験者)を示す変数、最適化の方法
- **Dimension**  
オブザベーション数、個体数、パラメータ数
- **Parameters**  
未知パラメータの初期値
- **Iteration Histroy**  
-log尤度 の最適化の経過

## SASアウトプット(2)

- **Fit Statistics**

あてはめの良さ（尤度）に関する情報

- **Parameter Estimates**

未知パラメータの最尤推定値

標準誤差：ヘシアン行列に基づく

自由度：個体数 - 変量効果のパラメータ数

分散成分( $s^2u$ ,  $s^2e$ )には解釈に注意が必要

この例では変量効果が線型なので積分の近似法は1次のテーラー展開でも同じ結果になる

## 数値例 その2

- 感染症治療薬の多施設共同試験(8施設、2剤、応答は2値)：Beitler and Landis(1985)
- ロジスティック混合モデル

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

$$h_{ij} = \log \frac{p_{ij}}{1 - p_{ij}} = \mathbf{b}_0 + \mathbf{b}_1 t_j + u_i$$

$n_{ij}$ ：施設*i*で治療法*j*の患者数、 $x_{ij}$ ：治癒した患者数

$t_j$ ：治療法*j*、 $u_i$ ：施設の変量効果 $N(0, \sigma_u^2)$

# SASプログラム

```

proc nlmixed data=infection;
  parms beta0=-1 beta1=1 s2u=2;
  eta = beta0 + beta1*t + u;
  expeta = exp(eta);
  p = expeta/(1+expeta);
  model x ~ binomial(n,p);
  random u ~ normal(0,s2u) subject=clinic;
  predict eta out=eta;
  estimate '1/beta1' 1/beta1;
run;

```

変量効果の分散成分

二項分布の  
モデルの指定

1/β<sub>1</sub>の推定

η<sub>ij</sub>の経験ベイズ推定値を含むデータセットetaの作成

# SASアウトプット

## ? Additional Estimate

ESTIMATEステートメントで指定した結果

Parameter	Estimate	Standard Error	DF	t Value	Pr >  t
beta0	-1.1974	0.5561	7	-2.15	0.0683
beta1	0.7385	0.3004	7	2.46	0.0436

推定値 $1/0.7385=1.3542$ , 標準誤差 $0.3004/0.7385^2=0.5509$

Label	Estimate	Standard Error	DF	t Value	Pr >  t
1/beta1	1.3542	0.5509	7	2.46	0.0436

## SASアウトプットつづき

### 信頼区間

Parameter	Al pha	Lower	Upper
beta0	0.05	-2.5123	0.1175
beta1	0.05	0.02806	1.4488

線型の変換ではない場合に、  
信頼限界から直接変換した  
ほうが良いかもしれない

デルタ法では、  
1/0.02806=12.4069 2.6569  
1/1.4488=0.6923 0.05146

Label	Al pha	Lower	Upper
1/beta1	0.05	0.05146	2.6569

## 数値例 その3

- 薬物動態の解析(テオフィリン経口投与)  
: Pinheiro and Bates(1995)
- 1コンパートメント・モデル

$$C_{it} = \frac{Dk_{e_i}k_{a_i}}{Cl_i} \left[ \exp(-k_{e_i}t) - \exp(-k_{a_i}t) \right] + e_{it}$$

$C_{it}$  : 被験者*i*の時間*t*の血中濃度

$D$  : 用量

$k_{e_i}$  : 被験者*i*の消失速度定数

$k_{a_i}$  : 被験者*i*の吸収速度定数

$Cl_i$  : 被験者*i*のクリアランス

$e_{it}$  : 被験者*i*の時間*t*の正規分布に従う誤差

## 数値例 その3 つづき

- 個体差を表す変量効果 $b_{i1}$ ,  $b_{i2}$ の導入

$$C_{it} = \exp(\mathbf{b}_1 + b_{i1})$$

$$k_{a_i} = \exp(\mathbf{b}_2 + b_{i2})$$

$$k_{e_i} = \exp(\mathbf{b}_3)$$

$$\begin{pmatrix} b_{i1} \\ b_{i2} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} s2b1 & cb12 \\ cb12 & s2b2 \end{pmatrix} \right)$$

## SASプログラム

```
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;
  c1 = exp(beta1 + b1);
  ka = exp(beta2 + b2);
  ke = exp(beta3);
  pred = dose*ke*ka*(exp(-ke*time)-exp(ka*time))
        /c1/(ka-ke);
  model conc ~ normal(pred,s2);
  random b1 b2 ~ normal([0,0],[s2b1,cb12,s2b2])
        subject=subject;
run;
```

変量効果の分散成分

変量効果



## MODELステートメントで 指定可能な分布

**normal(m,v)** : 平均 $m$ 、分散 $v$ の正規分布

**binary(p)** : 確率 $p$ の二値(ベルヌーイ)分布

**binomial(n,p)** : カウント $n$ 、確率 $p$ の二項分布

**gamma(a,b)** : 形状パラメータ $a$ 、尺度パラメータ $b$   
のガンマ分布

**negbin(n,p)** : カウント $n$ 、確率 $p$ の負の二項分布

**poisson(m)** : 平均 $m$ のポアソン分布

**general(l)** : PROGRAMMINGステートメントで  
指定した対数尤度関数 $l$ に従う分布

## 開発担当者Dr.Wolfinger に質問してきました

- **NONMEMと同じことができるか?**  
積分の近似方法をMETHOD=FIROとすることで、1次のテーラー展開を用いれば可能
- **GENMOD(GEE)と同じ結果が得られるか?**  
Subject-specific(NLMIXED)に対して、  
Population-avaraged(GEE)という違いがある
- **MIXEDプロシジャと同様にrepeatedステートメントは  
利用可能か?**  
使えない  
: 非線型混合モデルでは未だ一般的ではない

## 質問のつづき

- 初期値の設定方法は？

指定のない時はすべて初期値は 1

gridサーチも可能 `ex. parms -10 to 10 by 5;`

注：Dr.Wolfingerは初期値によらず収束すると言うが、パラメータ推定値が異なることが多々ある!!

収束条件を変えて確認？ (`GCOV`, `XCOV`オプション)

最適化の方法にオプションを加えるか？

(`LINESEARCH`=オプション, `UPD`=オプション)

- 分散成分の自由度の計算は？

個体(被験者)数 - 変数効果のパラメータ数、  
あるいはオブザベーション数で固定効果と同じ  
ただし、任意の自由度を`DF=d`として指定可能

## 質問のつづき

- パラメータに非負の制約があるときは？

`BOUNDS`ステートメントで範囲の指定が可能

`ex. bounds 0 <= a1-a9 x <=1, b1-b10 y>=0;`

- 3つ以上の変数効果の指定方法は？

3つの変数効果  $b_1, b_2, b_3$  の場合

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{pmatrix} \right)$$

ただし、 $s_{21}=s_{12}, s_{23}=s_{32}, s_{31}=s_{13}, s_{32}=s_{23}$

`random b1 b2 b3 ~ normal([0,0,0],`

`[s11,s21,s22,s31,s32,s33]) subject=person`

## 質問のつづき

- 施設と被験者のような二重に枝別れした指定は？  
指定できない。ただし、新しいNLINMIXマクロ (nlmm801.sas)で可能
- 2群のパラメータの比較方法は？  
PROGRAMMINGステートメントで、群毎に別の変数を指定し(CLASSステートメントは使えない!!)、ESTIMATEステートメントで差を指定する  
例えば、

```
if (group=1) then eta = alpha1 + beta1*logtstd + e;  
                else eta = alpha2 + beta2*logtstd + e;  
estimate 'alpha1-alpha2' alpha1-alpha2;  
estimate 'beta1-beta2' beta1-beta2;
```

## 質問のつづき

- 変数効果の正規性のチェック方法は？  
PREDICTステートメントで経験ベイズ推定値を出力して確認することが可能
- パラメータが多すぎないかチェックする方法は？  
極端な場合にはヘシアン行列がランク落ちする  
パラメータ間の共分散、相関係数が大きくなる  
パラメータ推定値の分散共分散行列を出力  
(COVオプション)  
パラメータ推定値の相関行列を出力  
(CORRオプション)  
ヘシアン行列の出力(HESSオプション)

## 最後に教科書は？

- Davidian, M. and Giltinan, D.M. (1995)  
*Nonlinear Model for repeated Measurement Data*,  
New York: Chapman & Hall.
- Vonesh, E.F. and Chinchihill, V.M. (1997)  
*Linear and Nonlinear Models for the Analysis of  
Repeated Measurements*,  
New York: Marcel Dekker.
- Agresti, Booth, Hobert, Caffoのカテゴリカルデータ  
についての論文  
<http://www.stat.ufl.edu/~aa/>

## やっぱりSASは凄かった!!

- **NLMIXED** プロシジャの開発者のDr.  
**Wolfinger**は**MULTTEST** プロシジャ、**MIXED**  
プロシジャの開発も担当
- Nonlinear Mixed Modelについて**Biometrika**な  
どに彼自身も論文を書いている
- アカデミアでの実績だけではなく、研究成果  
を広く一般に、多くの人が使えるプロシジャ  
にして提供するってただ者ではない