

# SASプロシジャにおける 非線形最適化法の実例

土居 主尚<sup>\*1</sup>、山口 拓洋<sup>\*2</sup>、吉永 信治<sup>\*1</sup>

<sup>\*1</sup> 放射線医学総合研究所 規制科学総合研究グループ

<sup>\*2</sup> 東京大学大学院医学系研究科 臨床試験データ管理学

## 背景 (最尤法と非線形最適化法)

- 統計解析モデルのパラメータ推定では最尤法は中心的な役割
- 最適化問題とは、与えられた条件の下で何らかの関数を最大化、もしくは最小化する問題
- オペレーションズ・リサーチ (OR) の分野で発達
- 尤度を最大化する最尤法のパラメータ推定に非線形最適化法が用いられている

## 背景 (SASにおける非線形最適化法)

- 個々の統計解析プロシジャでは、適用モデルに応じて安定性、速度にて適した非線形最適化法がデフォルト
- 複雑なモデルでは、しばしば反復計算によるパラメータ推定が収束しない
- 収束しない場合の対処に、非線形最適化法の変更

## 目的

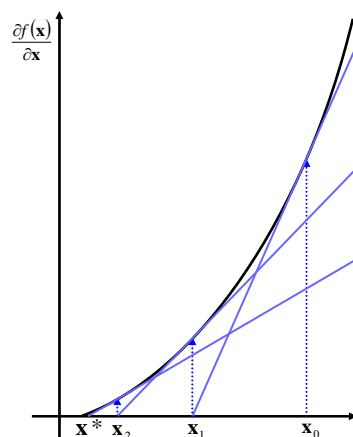
- SASで適用可能な非線形最適化法を説明
  - 統計解析プロシジャの内部処理をより理解
- 非線形最適化法を変更する必要がある解析例
- 複雑な対数尤度への非線形最適化法の適用例
  - 既存プロシジャでは実行できずIML を使う必要
  - 適切なプロシジャを主目的とは異なる使い方で実行可能

## 非線形最適化法

- 非線形の関数  $f(\mathbf{x})$  を最小化、最大化する問題
  - 初期値  $\mathbf{x}_0$
  - 反復式  $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k$
  - 最大値、最小値を与える解  $\mathbf{x}^*$
  - 無制約最適化問題 (制約付きは扱わない)
- 最尤法
  - 最大化の目的関数  $f(\mathbf{x})$  : 対数尤度関数
  - $f(\mathbf{x})$  を最大化する解  $\mathbf{x}^*$  : 最尤推定値

## ニュートン・ラフソン法

- 最小値を求める
  - 極値 ( $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = 0$ ) を求める
- 初期値  $\mathbf{x}_0$  を与える
- 目的関数  $f(\mathbf{x})$  を二次近似
  - $\mathbf{x}_0$  で  $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}$  を直線近似
- 極値に移動
  - 近似直線と0との交点に移動
  - この値を  $\mathbf{x}_1$  とする
- 以上を繰り返す
- SASでは2通りの改良した方法



## 準ニュートン法

- ニュートン・ラフソン法の改良
- ヘッセ行列(二次導関数)の計算
  - しばしば時間がかかる
  - 必ずしも正定値にならない
- ヘッセ行列を適切な正定値行列で近似
- NLMIXEDプロシジャのデフォルト
  - 多くの非線形混合効果モデルにて速度と安定性のバランスが良い

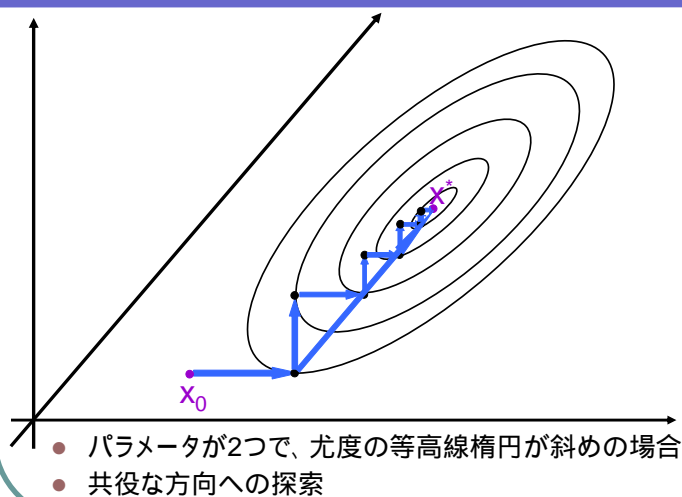
## 信頼領域法

- 目的関数の二次近似が適切な領域を決定
  - この領域を信頼領域と呼ぶ
- 目的関数を二次近似
  - 準ニュートン法などと同じ
- 信頼領域内にて二次近似モデルの最小化
  - 準ニュートン法などは一点で探索方向を決定
  - 信頼領域法は領域内で探索方向を決定するため安定
- アルゴリズムが複雑
  - 信頼領域の決定
  - 領域条件により制約付き最小化問題

## ダブルドッグレッグ法

- 準ニュートン法と信頼領域法の組み合わせ
- 信頼領域法と似た手順
  - 信頼領域を求める
  - 目的関数を二次近似
  - 信頼領域内にて二次近似モデルの最小化
- 二次近似モデルの最小化
  - 信頼領域の幅によって探索方向が変化
    - 信頼領域が広い場合、準ニュートン法
    - 信頼領域が狭い場合、信頼領域法

## 共役勾配法



## ロジスティックモデル

### ● 最尤法の実例

#### ● ロジスティックモデル

結果変数(二値) :  $Y_i$





説明変数(連続量) :  $Z_i$

$$\Pr(Y_i = 1) = \frac{\exp(\beta_0 + \beta_1 Z_i)}{1 + \exp(\beta_0 + \beta_1 Z_i)}$$

#### ● 対数尤度 ( $\pi_i = \Pr(Y_i = 1)$ )

$$l = \sum_{i=1}^N \left[ Y_i \log \left( \frac{\pi_i}{1 - \pi_i} \right) + n_i \log(1 - \pi_i) + \log \binom{n_i}{Y_i} \right]$$
$$= \sum_{i=1}^N \left[ Y_i (\beta_0 + \beta_1 Z_i) - n_i \log[1 + \exp(\beta_0 + \beta_1 Z_i)] + \log \binom{n_i}{Y_i} \right]$$

## IMLプロシジャのプログラム

```
proc iml;  対数尤度関数を行列表記にて定義
start logistic(beta) global(y, x, n);
  logL= sum(y # (beta[1] + beta[2] # x) -
           n # log(1 + exp(beta[1] + beta[2] # x)));
  return(logL);
finish;
init={1 1};  .....パラメータ初期値
optn={1 1};  .....最大化、計算の途中経過の表示
call nlpqn(rc, result, "LOGISTIC", init, optn);
 .....準ニュートン法のサブルーチン
```

## NLPプロシジャのプログラム

```

proc nlp data=beetle vardef=n cov=h
  tech=quanew; ←……準ニュートン法
  max logL; ←……下で記述するlogLの最大化
  parms beta1=1, beta2=1; ←……パラメータ初期値
  logL = y1 * (beta1 + beta2 * x) -
    n * log(1 + exp(beta1 + beta2 * x));
run;

```

$$l = \sum_{i=1}^n \left[ Y_i (\beta_0 + \beta_1 Z_i) - n_i \log[1 + \exp(\beta_0 + \beta_1 Z_i)] + \log \binom{n_i}{Y_i} \right]$$

## NLMIXEDプロシジャのプログラム

```

proc nlmixed data=beetle tech=quanew;
  parms beta1=1 beta2=1; ← NLPと同じ
  logL = y1 * (beta1 + beta2 * x) -
    n * log(1 + exp(beta1 + beta2 * x));
  model y1 ~ general(logL); ←……対数尤度を指定
run;

```

## 広島・長崎の原爆被爆者がんリスク解析

- 広島、長崎の原爆被ばく者を対象としたコホート研究
- 1958 年にがんを未発症、1998 年まで追跡した 105,427 名 (男性46%、女性54%)
- 総イベント数17,448、総人年2,764,734
- がんの発症率と放射線線量の用量反応関係
- 固形癌の発症をイベントとしたポアソン回帰

Preston DL et al. Radiation Research 2007; 168: 1-64.

## 広島・長崎の原爆被爆者がんリスク解析

- ポアソン回帰モデル  $E\left(\frac{Y}{T}\right) = \lambda_0(c, s, a, e, l) (1 + \beta_s d \exp(\gamma_1 e + \gamma_2 \log(a)))$ 
  - 放射線線量 ( $d$ )
  - 性別ごとの線量あたりのリスク増加:  $\beta_s$
- ベースラインハザード  $\lambda_0(c, s, a, e, l) = \exp(\gamma_3 s + \gamma_4 c + \gamma_5 c \times l + S_s(a) + S_s(e))$ 
  - 都市 ( $c$ )、性別 ( $s$ )、年齢 ( $a$ )、被ばく時年齢 ( $e$ )、被ばく時の所在地 ( $l$ )
  - $S_s(a)$ : 性別ごとの年齢の二次スプライン関数
  - $S_s(e)$ : 性別ごとの被ばく時年齢の二次スプライン関数
- パラメータの初期値
  - 全て0、全て1、 $\beta_s$ のみ100(残り0)



## がんリスク解析結果

### 各方法による収束

初期値設定	1	2	3
ニュートン-ラフソン			
準ニュートン	×		×
信頼領域			
ダブルドッグ法			×
共役勾配法			

設定1: 全て初期値0 (デフォルト)

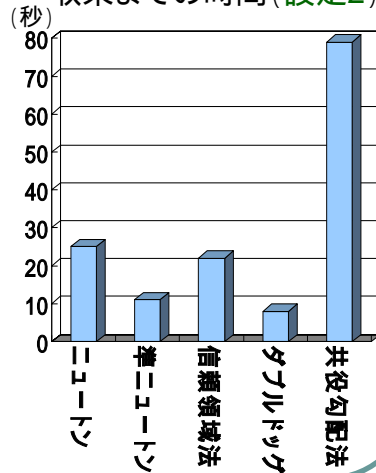
設定2: 全て初期値1

設定3:  $\beta_0$  初期値100、残り0

: 元論文と同じの値に収束

×: 違う値に収束

### 収束までの時間 (設定2)



## 表在性膀胱がんの再発データ

- 表在性膀胱がん患者を対象とした臨床試験
- 2群比較 (術後補助化学療法)
  - 抗がん剤の膀胱内注入 (N=205)
  - 生理食塩水の膀胱内注入 (N=215)
- 主要エンドポイント: 初回再発までの時間
- 2つの部位からの再発
  - 原発部位と同一部位
  - 他部位
- 2回目以降の再発の時間と部位も測定
- 再発パターンを考慮したモデリング

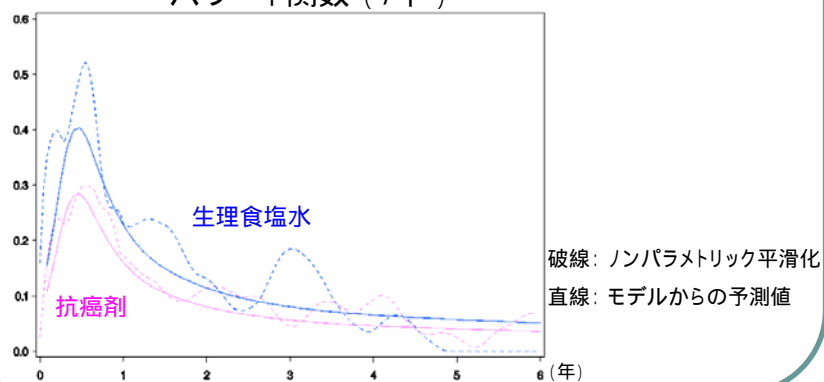
## 表在性膀胱がんの再発データ

- NLMIXEDにて尤度を指定して推定
  - (尤度) = (生存関数) × (ハザード)
    - (ハザード) = (同一部位の再発) + (他部位の再発)
- 再発部位ごとに様々なモデル化
  - 指数分布
  - ワイブル分布
  - 対数正規分布
  - スプライン関数
    - ハザード
    - 累積ハザード

## 表在性膀胱がんの再発データ

- 累積ハザードのスプライン関数(他部位)と対数正規分布(同一部位)がAICの観点より最もあてはまりが良い

ハザード関数 ( /年 )



## 考察

- 原爆被ばく者のリスク解析
  - 初期値、非線形最適化法のいずれかの変更で収束
  - 非線形最適化法の変更は収束しない場合の対処法
  - ダブルドッグレッグ法が速いが、初期値に依存
  - さまざまな非線形最適化法を試み結果の比較
- 表在性膀胱がんの再発データ
  - NLMIXEDプロシジャにて対数尤度が記述可能
  - IMLと比べてプログラムの記述が簡便
  - 既存プロシジャの出力が容易に利用可能

## 結論

- SASで適用可能な非線形最適化法を説明
  - IML、NLP、NLMIXEDで実行するプログラムの提示
- 原爆被ばく者のリスク解析
  - 非線形最適化の方法間に収束の有無、実行時間に差
- 表在性膀胱がんの再発データ
  - NLMIXEDプロシジャによる柔軟なモデル化