大規模統計処理に対する期待と現実、そしてSAS　SASユーザー総会　2015

クラスター生存時間データの解析法に関する研究

加藤 雄一郎 ・ 佐野 雅隆 ・ 寒水 孝司 ・ 浜田 知久馬
東京理科大学大学院 工学研究科 経営工学専攻

背景・目的

➢ 生存時間解析[3]

- 試験終了時の患者の生存や追跡不可能の打切りを扱う
- 生存時間解析において標準的なモデルとして，Coxの比例ハザードモデル(以下，Coxモデル)がある
  - 生存時間分布に特定の分布を仮定することなく，共変量の効果を推定可能

$$h(t) = h_0(t)\exp(\boldsymbol{\beta}^T \boldsymbol{z}_i)$$

t: 時点
h_0 : 基準ハザード関数
$\boldsymbol{\beta}$: パラメータベクトル
$\boldsymbol{z}$: 共変量ベクトル

➢ クラスター生存時間データ[3]

- 複数の生存時間が独立だと仮定できない場合の多変量データ
- 一個体に対して2つの生存時間データ
  - 同じ生活習慣や遺伝的背景を有するため，非常に高い相関をもつ
- 個人に異なる治療を割り付け
  - 個体差や交絡因子の影響を抑えることが可能

➢ 糖尿病性網膜症[5]

- 代表的な糖尿病合併症
  - 視野の欠損や著しい視力低下，失明に繋がる
  - 成人の失明原因の第一位
- 重症化しない限り自覚症状を生じない

→個人に対して，左右眼の生存時間データが得られる
  - 情報の損失をすることなく解析する必要がある

目的

- クラスター生存時間データに対する適切な解析手法の評価し，推奨される手法をシミュレーションによって検討する
  - 従来の解析手法，周辺Coxモデル，frailty model

シミュレーション条件

●クラスターサイズ2の生存時間データを想定

➢シミュレーションによる検討

想定1:1クラスターに異なる2つの治療法(点眼薬など)
独立と仮定したCoxモデル，層別Coxモデル，
周辺Coxモデル，frailtyモデル

想定2:1クラスターに同一の治療法（経口剤など）
片側のみCoxモデル，独立と仮定したCoxモデル，
周辺Coxモデル，frailtyモデル

- シミュレーション回数：10000回
- クラスターサイズ:2
- クラスター数: 200
- 有意水準:5%
- 観察期間:60ヶ月(*60ヶ月以上生存は観察打切り)
- 生存時間分布：ワイブル分布
- 途中打切り:約20%
- クラスター内の相関係数：R = 0, 0.1, 0.2, ⋯ , 0.9

評価指標：αエラー，検出力，ハザード比のMSEとbias

方法

➢ 周辺Coxモデル[1]

- サンドイッチ分散を用いたCoxモデル
- Coxモデルの共変量の推定値と一致

●相関を考慮しない場合の分散

$$\widehat{\boldsymbol{V}} = \boldsymbol{I}^{-1}$$

●サンドイッチ分散：誤差項に共分散が存在する際，有効な分散共分散推定量

$$\widehat{\boldsymbol{V}} = \boldsymbol{I}^{-1}\boldsymbol{U}(\widehat{\boldsymbol{\beta}})\boldsymbol{U}^T(\widehat{\boldsymbol{\beta}})\,\boldsymbol{I}^{-1}$$

$\boldsymbol{I}$: Fisherの情報行列
$\boldsymbol{U}(\widehat{\boldsymbol{\beta}})$:スコアベクトル

```
proc phreg data=data covs(aggregate);
class 分類変数;
model 時間変数*打切り変数(0)=説明変数;
id 変数;
run;
```

➢ frailty model[2]

- Coxモデルにランダム効果を考慮
  →クラスター内の非独立性に起因する相関をモデル化
- ハザード関数

$$h_{ij}(t|w_i) = h_0(t)\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{z}_{ij} + w_i)$$
$$= h_0(t)u_i\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{z}_{ij})$$

i: クラスター
j: クラスターiに属する観測値
w_i: クラスターiのランダム効果
u_i : frailty項（=exp(w_i)）

- frailty項(u_i)は一般的にガンマ分布や対数正規分布に従うと仮定

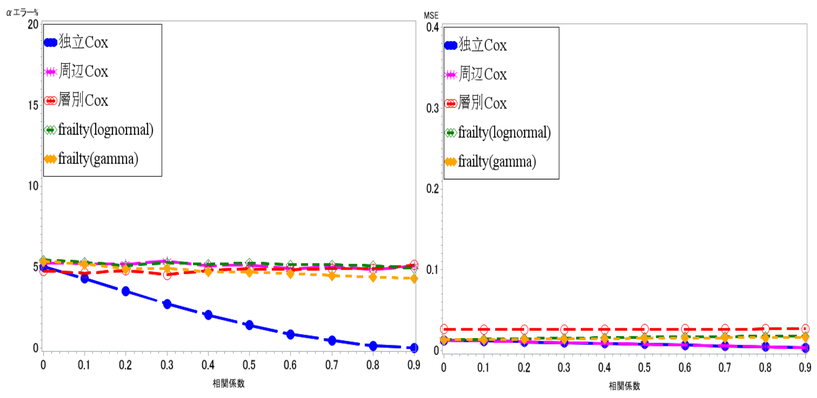
```
proc phreg data=data;
class 分類変数;
model 時間変数*打切り変数(0)=説明変数;
random 変数 / dist=gamma;
run;
```

➢ コピュラ関数

- 一次元周辺分布を結合し，多次元同時分布を構成する結合関数
  - 正規コピュラ，tコピュラ，アルキメディアンコピュラなどがある
- 本シミュレーションでは，コピュラ関数を用いてクラスター生存時間データを生成させた

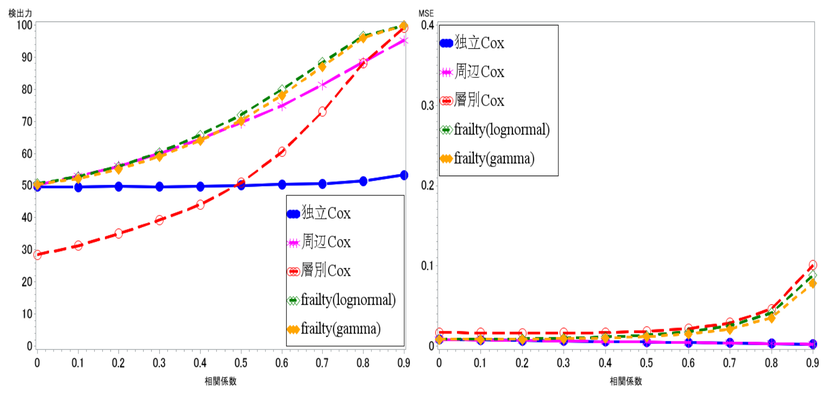
結果・考察

想定1



αエラー　MSE(帰無仮説)

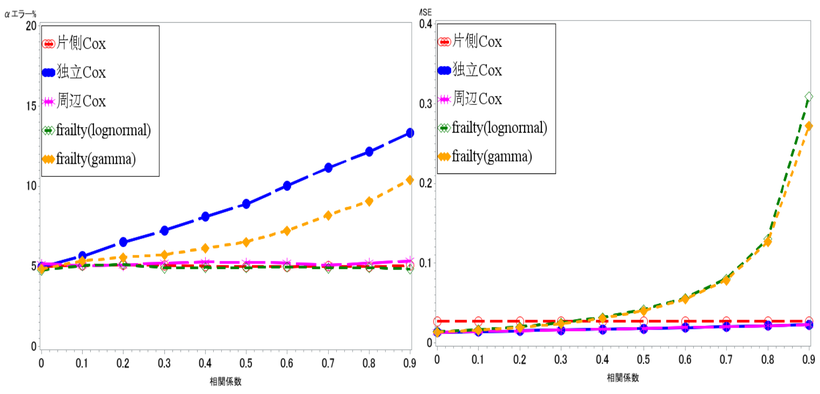
- 独立と仮定したCoxモデルでは，αエラーが小さい
  →個人間変動を考慮できず，ばらつきを過大評価



検出力　MSE(対立仮説)

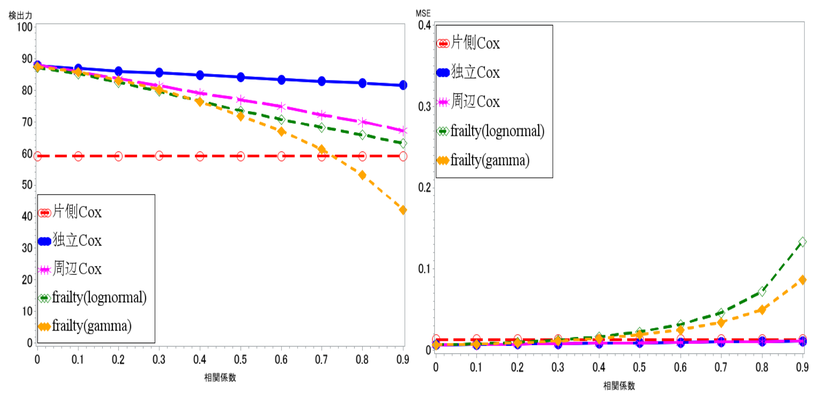
- frailty modelは，MSEが大きいためばらつきが大きい
  →frailty項の分散が大きくなり，比例ハザード性が成り立たなくなるため

想定2



αエラー　MSE(帰無仮説)

- 周辺Coxモデルと対数正規frailtyモデルのαエラーが保たれる



検出力　MSE(対立仮説)

- 片側のみを用いたCoxモデルでは，検出力が小さい
  →実質的なサンプルサイズが小さいため

まとめ

- 研究の複雑化に伴いクラスター生存時間データが得られる機会が増加
- クラスター生存時間データに対して提案されている手法の評価を行った

●独立と仮定したCoxモデルはαエラーが有意水準に保たれない
●frailtyモデルはバラツキが大きく精度が非常に悪い

推奨する手法
周辺Coxモデル

参考文献

[1] David A.F (2012), ‘ On the so-called “Huber Sandwich Estimator” and “Robust Standard Errors” ’ , *The American Statistician*, **60**:*299-302*
[2] Luise C.K (2012),Statistical Analysis of Familial Aggregation of Adverse Outcomes, Technical University of Denmark, *DTU Informatics*,37-43
[3]大橋靖雄・浜田知久馬(1995), 生存時間解析 SASによる生物統計 東京大学出版会 , 第4刷
[4]影山裕基(2012), 糖尿病性網膜症データを用いた区間打切りを伴うクラスター生存時間データの解析手法に関する研究, 東京理科大学大学院工学研究科経営工学専攻
[5] 日本眼科学会：目の病気 糖尿病網膜症 http://www.nichigan.or.jp/public/disease/momaku_tonyo.jsp, 最終閲覧日:2015.07.08