

## 2つの生存関数の差に対する同時信頼区間

平井 健太<sup>1</sup>・吉田祐樹<sup>2</sup>・田崎武信<sup>2</sup>

<sup>1</sup>住商情報システム(株) データサイエンス部ビジネスパートナー

<sup>2</sup>塩野義製薬(株) 解析センター

## Simultaneous confidence intervals for the difference of two survival functions using SAS

Kenta Hirai <sup>1</sup> / Yuuki Yoshida <sup>2</sup> / Takenobu Tasaki <sup>2</sup>

<sup>1</sup>Data Science Div. Business Partner, Sumisho Computer Systems Corp.

<sup>2</sup> Biostatistics Dept., Shionogi & Co.,Ltd.

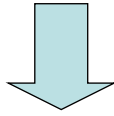
## 発表の流れ

- ・ はじめに
- ・ SAS評価版機能  
(ODS GRAPHICSステートメント)との比較
- ・ プログラムの検証
- ・ 検定との関係、発展、応用
- ・ プログラミングでの工夫
- ・ まとめ

はじめに

SAS V9の評価版機能として、ひとつの生存関数に対するポイントワイズ信頼区間や同時信頼区間を簡単に描くものがあることを知った。

しかし、2つの生存関数を比較したい場合には、生存時間の差に対する信頼区間を描くほうが妥当ではないかと感じた。

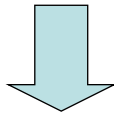


SASにはその機能を見つけられなかった

2

はじめに

2つの生存関数の差に対するノンパラメトリックな同時信頼区間がPaezen, Wei and Ying(1997)で提案されていたことを知った。



その手法を、SASでプログラム化し、実用的なマクロとして活用できないかと考えた

3

プログラミングを意識して Parzen et al.(1997)の方法を要約する

第  $i$  標本のサイズを  $n_i$  とする ( $i=1,2$ )。  
 第  $i$  群における第  $j$  患者の、死亡時間と中途打ち切り時間の小さいほうを  $X_{ij}$  とする。そして、 $X_{ij}$  が死亡時間のほうであれば  $\Delta_{ij}=1$ 、そうでなければ  $\Delta_{ij}=0$  とする。  
 ここに  $i=1,2$ 、そして  $j=1,\dots,n_i$  である。

$\hat{S}_i(t)$  を計算する ( $i=1,2$ )。

$se\{\hat{S}_i(t)\}$  を計算する ( $i=1,2$ )。

$\hat{v}(t) = \left\{ \sqrt{(se\{\hat{S}_1(t)\})^2 + (se\{\hat{S}_2(t)\})^2} \right\}^{-1}$  を計算する。

4

興味のある時間区間を  $[u_1, u_2]$

Default は  $u_1 = \inf\{t : \hat{S}_1(t) < 1 \text{ または } \hat{S}_2(t) < 1\}$

$u_2 = \sup\{t : \hat{S}_1(t) > 0 \text{ か } \hat{S}_2(t) > 0\}$

とりあえず  $N=1000$ 、 $\alpha=0.05$  とする。

標準正規分布に従う乱数  $\{Z_{ij}, i=1,2, j=1,\dots,n_i\}$  を抽出し

$\tilde{U}_i(t) = \hat{S}_i(t) \sum_{j=1}^{n_i} \left[ \left\{ \sum_{k=1}^{n_i} I(x_{ik} \geq x_{ij}) \right\}^{-1} I(x_{ij} \leq t) \delta_{ij} Z_{ij} \right]$  を計算する。

ここに、 $I(\cdot)$  は括弧内の条件を満たすとき値 1、そうでないとき値 0 をもつ指標関数である。

5

$\tilde{V}(t) = \hat{v}(t)\{\tilde{U}_2(t) - \tilde{U}_1(t)\}$  を計算する。

$\tilde{G} = \sup_{t \in [u_1, u_2]} |\tilde{V}(t)|$  を計算する。

上記のステップを  $N$  回反復し、 $\Pr\{\tilde{G} > c(\alpha)\} = \alpha$  となる  $c(\alpha)$  を求める。

同時信頼帯

$B = \left\{ \hat{S}_2(t) - \hat{S}_1(t) \pm c(\alpha)\hat{v}^{-1}(t); u_1 \leq t \leq u_2 \right\}$   
を計算する。

## 発表の流れ

- ・ はじめに
- ・ SAS評価版機能  
(ODS GRAPHICSステートメント)との比較
- ・ プログラムの検証
- ・ 検定との関係、発展、応用
- ・ プログラミングでの工夫
- ・ まとめ

## SAS評価版の機能を紹介する

SAS V9以降では、評価版の機能として提供されているODS GRAPHICSステートメントを使用して、これまででない形態のグラフをHTML、PDFおよびRTF形式などのファイルとして出力できるようになりました。

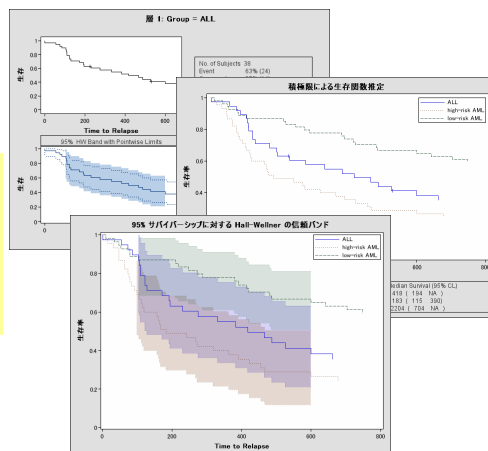
(中略)

この機能を利用するには、「ODS GRAPHICS ON;」と「ODS GRAPHICS OFF;」を、プロシジャの前後に挿入します。多くのプロシジャでは、この指定だけでグラフが自動的に描画されます。

SAS Technical News 2005 Winterより抜粋

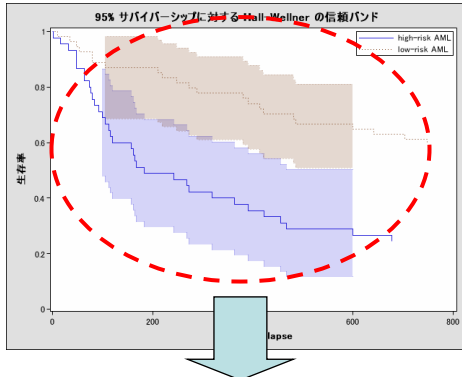
## ODS GRAPHICSステートメント(PROC LIFETEST)

```
ods html;
ods graphics on;
proc lifetest data = BMT noprint;
time T * Status(0);
survival out=Out1
  confband = all
  bandmin = 100
  bandmax = 600
  maxtime = 800
  conftype = asinsqrt
  plots = ( stratum, survival, hwb);
strata Group;
run;
ods graphics off;
ods html close;
```



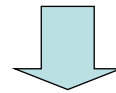
SASヘルプ Example 40.2:より抜粋

SAS V9.1.3 Example 40.2から、Low-risk群とHigh-risk群に注目して、SAS評価版機能のODS GRAPHICSステートメントで、Hall-Wellnerの同時信頼区間を描いた。



検定 Pr > Chi-Square

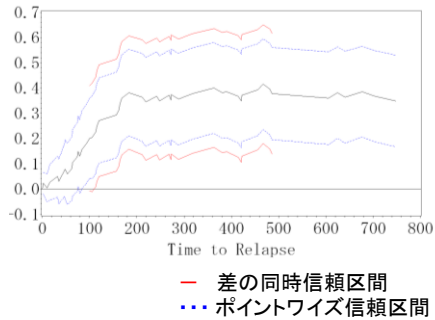
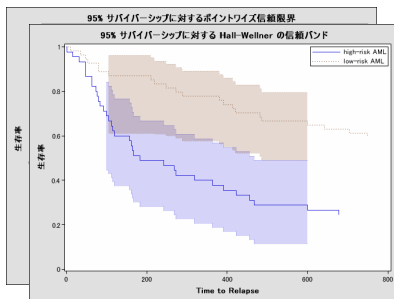
ログランク 0.0002  
Wilcoxon <.0001



検定の結果では、有意な差がある

有意な差がありそうだが、本当にそうだろうか？

2つの生存関数には大きな差が見られるが、この差の統計的有意性を解釈するためには、差の信頼区間を見るほうが、より直接的と考える。Parzen et al.(1997)にもとづき、今回開発したプログラムから出力した2つの生存関数の差の同時信頼区間を示す



個々の生存関数と信頼区間を眺めるよりは、差の信頼区間を眺めるほうが、わかりやすい

## 発表の流れ

- ・ はじめに
- ・ SAS評価版機能  
(ODS GRAPHICSステートメント)との比較
- ・ プログラムの検証
- ・ 検定との関係、発展、応用
- ・ プログラミングでの工夫
- ・ まとめ

12

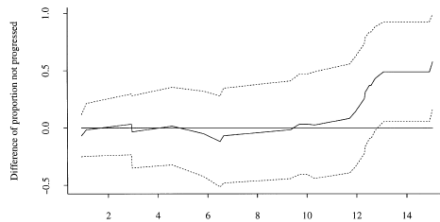
## プログラムの検証

開発したプログラムが正しく動作することを、2つの方法で検証した。

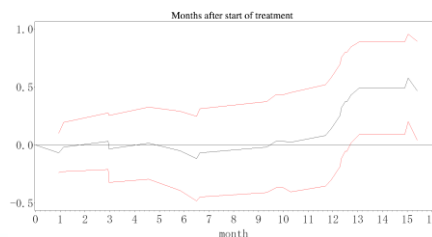
- ① Parzen et al.(1997)の2つ目の例で示された図の再現
- ② 小規模シミュレーション研究で、名目95%の同時信頼区間のカバー率を確認した。

## プログラムの検証①

開発したプログラムのバリデーションとして、Parzen et al.(1997)の2つ目の例に掲載されていた図を再現した。



Parzen et al.(1997)



開発したSASプログラムからの出力

14

## プログラムの検証②

小規模シミュレーション研究で、名目95%の同時信頼区間のカバー率が94.8%になることを確かめた。

## シミュレーション条件

- 生存時間をエンドポイントとする1群100例の2群並行群間比較試験を想定した。
- 応答変数である生存時間は指数分布を仮定し、群1の5年生存率を50%、群2の5年生存率を33.5%として発生させた。
- なお、8年以上生存の場合は打ち切りと扱った。
- このシミュレーション条件での検出力は約80%である。

15



## 発表の流れ

- ・ はじめに
- ・ SAS評価版機能  
(ODS GRAPHICSステートメント)との比較
- ・ プログラムの検証
- ・ 検定との関係、発展、応用
- ・ プログラミングでの工夫
- ・ まとめ

16

## 検定との関係

2つの生存曲線の差の検定では、ログランク検定と一般化Wilcoxon検定が多用されている。

ふつう、検定と信頼区間は表裏の関係にあるが、いまの場合はそうではない。

2つの生存関数の差の同時信頼区間の下限がゼロを超える(または同時信頼区間の上限がゼロを超えない)ことと、検定との関係を確かめた。

17

## 検定との関係

### 同時信頼区間と検定の関係

	検定で有意かつ信頼区間下限がゼロを超える	検定で有意かつ信頼区間下限がゼロを超えない	検定で有意でないかつ信頼区間下限がゼロを超える	検定で有意でないかつ信頼区間下限がゼロを超えない
同時信頼区間	65.1% 63.5%	15.3% 11.8%	1.6% 3.2%	18.0% 21.5%
ポイントワイズ信頼区間	80.4% 75.3%	0.0% 0.0%	16.6% 21.3%	3.0% 3.0%

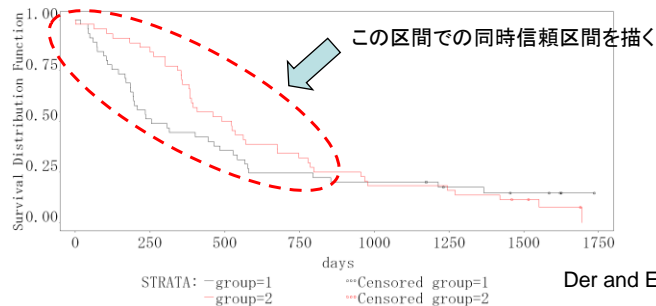
上段がログランク検定、下段が一般化Wilcoxon検定

## 応用①

今回開発したSASマクロでは、同時信頼区間を求める時間区間を指定できる。

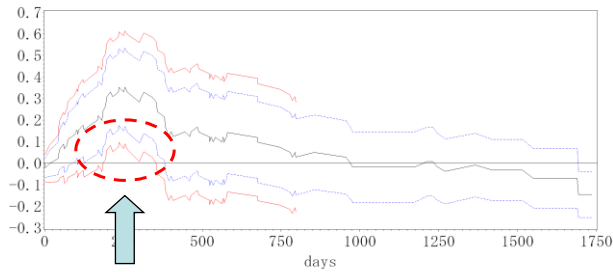
Der and Everitt(2001)で紹介されている胃がんデータでは、前半(およそ800日目まで)の差が注目される。

前半に焦点を当てた同時信頼区間を描く



Der and Everitt(2001)

前半(およそ800日目まで)の同時信頼区間



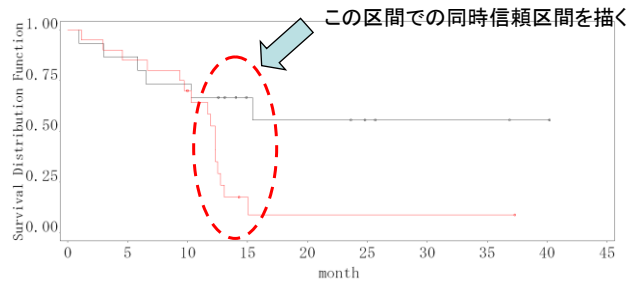
この部分に注目

Test	Chi-Square	DF	Pr > Chi-Square
ログランク	5.7629	1	0.0164
Wilcoxon	8.5779	1	0.0034
-2Log(LR)	3.5794	1	0.0585

応用②

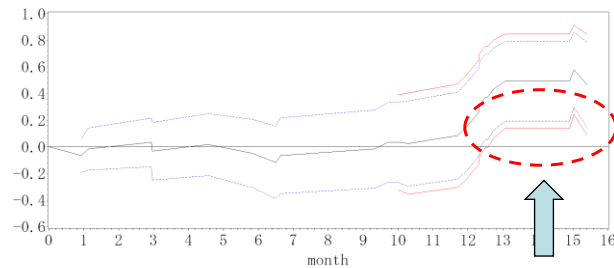
Parzen et al.(1997)の2つめの例では、およそ10ヶ月~16ヶ月目までの差が注目される。

この区間に焦点を当てた同時信頼区間を描く



Parzen et al.(1997)

## およそ10ヶ月～16ヶ月目までの同時信頼区間



この部分に注目

Test	Chi-Square	DF	Pr > Chi-Square
ログランク	5.3300	1	0.0210
Wilcoxon	3.0738	1	0.0796
-2Log(LR)	2.0201	1	0.1552

22

## 発表の流れ

- ・ はじめに
- ・ SAS評価版機能  
(ODS GRAPHICSステートメント)との比較
- ・ プログラムの検証
- ・ 検定との関係、発展、応用
- ・ プログラミングでの工夫
- ・ まとめ

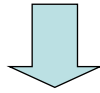
23

## プログラミングでの工夫

Parzen et al.(1997)で提案されている方法をSASプログラム化  
する際の最大の課題は、実行時間を短縮させることであった。

$$\tilde{U}_i(t) = \hat{S}_i(t) \sum_{j=1}^{n_i} \left[ \left\{ \sum_{k=1}^{n_i} I(x_{ik} \geq x_{ij}) \right\}^{-1} I(x_{ij} \leq t) \delta_{ij} Z_{ij} \right]$$

この式を計算する際に、繰り返し処理が多く発生する。

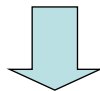


繰り返し処理を毎回データセットとして作成した場合、実行時間  
が大きく増えてしまった。

24

## 対策として

- 繰り返し処理は1つのデータステップ内で、DOステートメントを  
使用し、データセットの書き込みと読み込みの回数を減らした。
- データセット中の変数を必要最小限に絞ることで、データセット  
のサイズを小さくするように注意した。



対策を加えた結果、実行時間が大幅に短縮され、100例程度の  
データであれば10～15秒程度で実行が完了するようになった。

25

別の対策として、RAM DISKの導入も検討した。

RAM DISKは、PCの主記憶(メモリ)の一部をハードディスクのように扱う仕組み。物理的な動作がないため、ハードディスクよりも高速にデータを読み書きできる。

SASプログラムの実行時間の内訳で大部分を占めるのはデータセットの読み書きである。この速度を向上させると、実行時間が大幅に短縮されると期待できる。

PCにRAM DISKのシステムをインストールし、メモリ2GBのうち512MBを割り当てた場合の実行時間を調べた。

26

## RAM DISKの導入の結果

RAM DISKの導入の有無で、実行時間にほとんど差はなかった。



検証したプログラムでは、小さいサイズのデータセットを、繰り返し処理ごとに読み書きしていた。

個々のデータセットの読み書きの時間は短縮されたが、読み書きの回数が多いため、実効時間は短縮されなかったと考えられる。

27

RAM DISKはサイズの大きいデータを頻繁に読み書きする際に、実行時間に大きく影響する。

(例えば、500MBのSASデータセットを作成する時間は、RAM DISK導入ありで1秒、なしで30秒程度だった)

しかし、RAM DISKの導入で無条件に実行時間が短縮されるわけではない。プログラムの構成による。



使用するPCの環境に合わせ、実行時間の短縮を意識したプログラミングが重要である。

## 発表の流れ

- ・ はじめに
- ・ SAS評価版機能  
(ODS GRAPHICSステートメント)との比較
- ・ プログラムの検証
- ・ 検定との関係、発展、応用
- ・ プログラミングでの工夫
- ・ まとめ

## まとめ

Parzen et al.(1997)の方法にもとづいて2つの生存関数の差に対する同時信頼区間を計算するSASマクロプログラムを開発した。

同時信頼区間は差を直截解釈・評価するのに有用である。さらに、同時信頼区間とポイントワイズ信頼区間の特徴を理解したうえで、それらを一緒に表示することはより有意義であると考えられる。

## 参考文献

- 1) 吉田祐樹・平井健太・田崎武信(2009). ブートストラップサンプリングによる生存関数の差の同時信頼区間の構築。(日本計量生物学会年会に提出したが、新型インフルエンザの影響でこの年会の開催は中止された.)
- 2) Parzen, M.I., Wei, L.J. and Ying, Z. (1997). Simultaneous confidence intervals for the difference of two survival functions. *Scand. J. Statist.*, 24, 309-314.
- 3) Der, G and Everitt, S. (2001). *Handbook of Statistical Analyses Using SAS, Second Edition*, CHAPMAN & HALL.



End of File