

Information criteria for Firth's penalized partial likelihood approach in Cox regression models

Nagashima & Sato (*Stat Med* 2017)

統計数理研究所 リスク解析戦略研究センター
第9回 生物統計ネットワークシンポジウム 2018
2018/3/19

千葉大学大学院医学研究院
長島 健悟

本研究で扱う問題

- Cox回帰モデルの単調尤度問題 (Bryson & Johnson, 1981)
- 発生状況： β が大, 打ち切り多など
- 問題：点推定値&標準誤差が $-\infty$ or $+\infty$
- 対処法：Firthのバイアス補正法 (Firth, 1993; Heinze & Schemper, 2001, 2002)

妥当なモデル選択基準がない
⇒ 新たな手法を提案

Motivation example

- JLGK0901 study (Yamamoto et al., 2014)
 - 転移性脳腫瘍に対するガンマナイフ単独治療の有効性探索の前向き観察研究
 - 主目的：脳転移数で有効性に違いがないか評価
 - 副次目的：各イベントの予後因子探索
- 主要評価項目：全生存期間
- 副次評価項目：髄膜播種発生までの期間等
- 解析：Cox回帰モデル

データの説明と解析方針

- 髄膜播種発生までの期間を解析対象
- 解析計画：モデル選択 ⇒ 最終モデルの推定結果を評価の予定
- 単調尤度問題が . . .

<i>age</i>	年齢 (<65, ≥65)	<i>sex</i>	性別 (Female, Male)
<i>kps</i>	Karnofsky performance status (≥80, ≤70)	<i>diameter</i>	最大腫瘍径 (<1.6cm, ≥1.6cm)
<i>volume</i>	累積腫瘍体積 (<1.9mL, ≥ 1.9mL)	<i>ptumor</i>	原発腫瘍 (肺, 胸, 胃腸, 腎, その他)
<i>status</i>	脳外疾患 (なし, あり)	<i>neuro</i>	神経症状 (なし, あり)
<i>ntumor</i>	腫瘍個数 (1, 2-4, 5-10)		

Cox回帰による推定結果

共変量		HR	SE	95%CI	P (Wald)
転移個数	1 vs 2 - 4	0.72	0.20	0.49 to 1.06	0.09
	5 - 10 vs 2 - 4	1.58	0.22	1.03 to 2.41	0.04
原発腫瘍	胸 vs 肺	1.03	0.29	0.58 to 1.83	0.91
	胃腸	1.55	0.37	0.75 to 3.21	0.24
	腎	0.00	543.30	N.A.	0.98
	その他	0.77	0.59	0.24 to 2.45	0.66
神経症状	あり vs なし	1.50	0.23	0.96 to 2.36	0.07

※ 全因子の解析結果だがその他の因子は省略

腎の $\hat{\beta}$ は $-\infty$, 95%CIは計算不能

髄膜播種のデータ要約 ($n = 1073$)

共変量		髄膜播種発生数	打ち切り数	打ち切り%
転移個数	1	49	365	88.2%
	2 - 4	61	412	87.1%
	5 - 10	35	151	81.2%
原発腫瘍	肺	116	705	85.9%
	胸	17	95	84.8%
	胃腸	9	66	88.0%
	腎	0	32	100.0%
	その他	3	30	90.9%
神経症状	なし	105	272	87.2%
	あり	145	928	86.5%

腎はリスクが低いのでは？

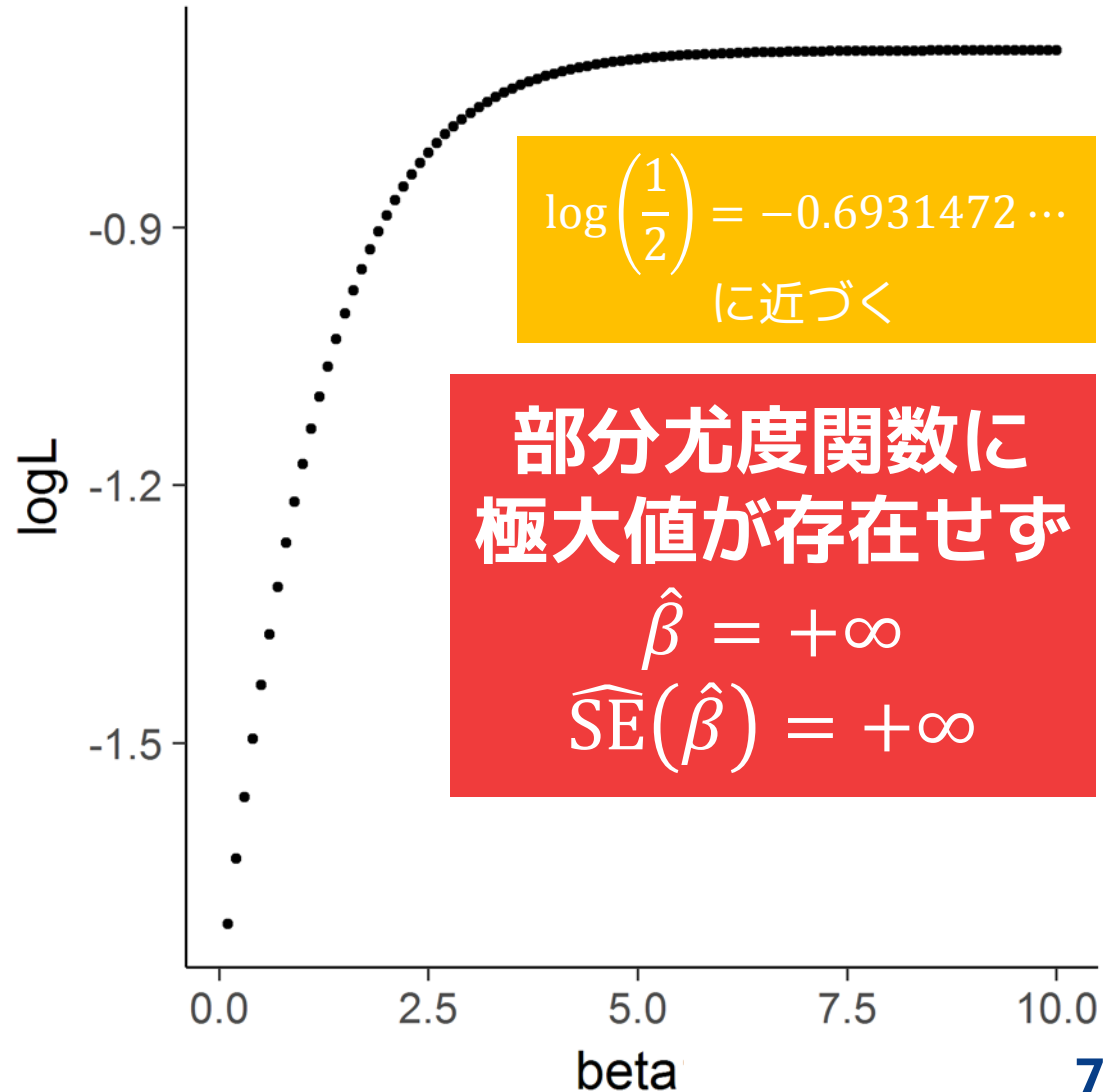
. . . (一部略) . . .

単調尤度 (monotone likelihood)

例：
時間, イベントの有無 (1/0), 共変量が
 $(t_1, t_2, t_3) = (1, 2, 3)$
 $(c_1, c_2, c_3) = (1, 1, 1)$
 $(x_1, x_2, x_3) = (1, 1, 0)$

部分尤度関数：

$$L(\beta) = \frac{e^\beta}{2e^\beta + 1} \cdot \frac{e^\beta}{e^\beta + 1} \cdot \frac{1}{1}$$



単調尤度 (monotone likelihood)

- パラメータ推定値や標準誤差が得られなくなる問題
- 起こりやすい状況：強力な因子, 打ち切り多い, 因子の分布のアンバランス; 詳細な条件は付録

**本当に強い効果を持つ因子などでも
結果を解釈できない状況が起こりうる**

Firthのバイアス補正法

- 元々は最尤推定量のバイアス補正の方法;
 $O(n^{-1})$ のバイアスを除去
- 罰則付き対数部分尤度関数の利用

$$l^*(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) + \frac{1}{2} \log |\mathbf{I}(\boldsymbol{\beta})|$$

- $\mathbf{I}(\boldsymbol{\beta})$ は観測情報行列

罰則項は正定値であれば
極大値を持つ

単調尤度でも副作用的に推定値が
得られる (Heinze & Schemper, 2001)

Firthの方法による推定結果

共変量		HR	SE	95%CI	P (Wald)	P (LR)
年齢	≥65 vs <65	1.00	0.17	0.72 to 1.40	0.98	0.98
性別	男性 vs 女性	1.19	0.19	0.83 to 1.72	0.35	0.34
KPS	≥80 vs ≤70	1.07	0.32	0.57 to 2.00	0.83	0.83
腫瘍径	≥1.6 vs <1.6	0.90	0.33	0.47 to 1.70	0.74	0.74
腫瘍体積	≥1.9 vs <1.9	1.11	0.32	0.59 to 2.08	0.76	0.76
転移個数	1 vs 2 - 4	0.72	0.20	0.49 to 1.06	0.09	0.09
	5 - 10 vs 2 - 4	1.59	0.22	1.04 to 2.43	0.03	0.04
原発腫瘍	胸 vs 肺	1.05	0.29	0.60 to 1.87	0.86	0.85
	胃腸	1.61	0.37	0.79 to 3.31	0.19	0.21
	腎	0.12	1.43	0.01 to 1.91	0.13	0.02
	その他	0.77	0.59	0.30 to 2.63	0.84	0.83
脳外疾患	あり vs なし	1.03	0.19	0.71 to 1.48	0.89	0.89
神経症状	あり vs なし	1.50	0.23	0.97 to 2.73	0.07	0.07

データの説明と解析方針（再掲）

- 髄膜播種発生までの期間を解析対象
- 解析計画：**モデル選択** ⇒ 最終モデルの推定結果を評価の予定
- 単調尤度問題は一応解決

<i>age</i>	年齢 (<65, ≥65)	<i>sex</i>	性別 (Female, Male)
<i>kps</i>	Karnofsky performance status (≥80, ≤70)	<i>diameter</i>	最大腫瘍径 (<1.6cm, ≥1.6cm)
<i>volume</i>	累積腫瘍体積 (<1.9mL, ≥ 1.9mL)	<i>ptumor</i>	原発腫瘍 (肺, 胸, 胃腸, 腎, その他)
<i>status</i>	脳外疾患 (なし, あり)	<i>neuro</i>	神経症状 (なし, あり)
<i>ntumor</i>	腫瘍個数 (1, 2-4, 5-10)		

Firthの方法下でのモデル選択

- 既存研究を発見できず
- RとStata：罰則付き最大対数部分尤度 $l^*(\hat{\beta})$ を出力可
- SAS：AIC*とBIC*を出力

p ：パラメータ数
 d ：イベント数

$$AIC^* = -2l^*(\hat{\beta}) + 2p$$

$$BIC^* = -2l^*(\hat{\beta}) + p \log d$$

**SASが計算するAIC*で
モデル選択を実施**

AIC*によるモデル選択の結果

Top 5 Model based on AIC*									AIC*
<i>age</i>	<i>sex</i>	<i>ntumor</i>	<i>kps</i>	<i>diameter</i>	<i>volume</i>	<i>ptumor</i>	<i>status</i>	<i>neuro</i>	1731.50
<i>age</i>	<i>sex</i>	<i>ntumor</i>		<i>diameter</i>	<i>volume</i>	<i>ptumor</i>	<i>status</i>	<i>neuro</i>	1731.80
<i>age</i>	<i>sex</i>	<i>ntumor</i>	<i>kps</i>	<i>diameter</i>		<i>ptumor</i>	<i>status</i>	<i>neuro</i>	1731.85
<i>age</i>	<i>sex</i>	<i>ntumor</i>	<i>kps</i>		<i>volume</i>	<i>ptumor</i>	<i>status</i>	<i>neuro</i>	1731.85
<i>age</i>	<i>sex</i>	<i>ntumor</i>			<i>volume</i>	<i>ptumor</i>	<i>status</i>	<i>neuro</i>	1732.16

- 寄与がほとんどなさそうな因子まで選択
 - *age* (HR = 1.00 [0.72, 1.40], P = 0.98), *status* (HR = 1.03 [0.71, 1.48], P = 0.89)
- 選択されるモデルはパラメータ数が多い

何かが誤っている . . . ?

本研究の目的

- Firthの方法適用下でのモデル選択基準について検討
- AIC*の性質の解明
- 妥当なモデル選択基準の提案
- Cox回帰のもとでのFirthのバイアス補正法の妥当性についての検討 (論文を参照)

記号の定義

- 主なものを記載 (その他は論文を参照)
- 個体 i ($= 1, \dots, n$), 生存時間 t_i ($[0, 1]$ にリスケール)と, 時点 t のイベントの有無 $N_i(t)$, 共変量 $\mathbf{Z}_i(t)$, リスク集合過程 $Y_i(t)$ が観測
- 対数部分尤度 : $l(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^1 \boldsymbol{\beta}^T \mathbf{Z}_i(x) - \log\{\sum_{j=1}^n Y_j(x) \exp\{\boldsymbol{\beta}^T \mathbf{Z}_j(x)\}\} dN_i(x)$
- スコア関数 : $\mathbf{U}(\boldsymbol{\beta}) = \partial l(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}$
- 観察情報行列 : $\mathbf{I}(\boldsymbol{\beta}) = \partial^2 l(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T$
- 罰則付き対数部分尤度 : $l^*(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) + 0.5 \log |\mathbf{I}(\boldsymbol{\beta})|$
- 最大部分尤度推定量 (通常) : $\hat{\boldsymbol{\beta}}_{cox} = \operatorname{argmax}_{\boldsymbol{\beta}} l(\boldsymbol{\beta})$
- 最大部分尤度推定量 (Firth) : $\hat{\boldsymbol{\beta}} = \operatorname{argmax}_{\boldsymbol{\beta}} l^*(\boldsymbol{\beta})$
- パラメータ真値とFisher情報行列 : $\boldsymbol{\beta}_0, \boldsymbol{\Sigma}(\boldsymbol{\beta})$

AIC*の性質

- 以下に変形できる

$$\text{AIC}^* = -2l(\hat{\beta}) + (2 - \log n)p - \log\{n^{-p} |\mathbf{I}(\hat{\beta})|\}$$

- $-2l(\hat{\beta})$: 部分尤度であり, p が大きいほど小さい値をとる
- $(2 - \log n)p$: $n \geq 8$ で負の値をとり, かつ p が大きいほど小さい値をとる
- $n^{-p} |\mathbf{I}(\hat{\beta})| \rightarrow_p |\Sigma(\beta_0)|$: $n \rightarrow \infty$ で定数に収束

p の大きいモデルを選びやすい (n に依存)

AIC*の性質

Top 5 Model based on AIC*									AIC*
<i>age</i>	<i>sex</i>	<i>ntumor</i>	<i>kps</i>	<i>diameter</i>	<i>volume</i>	<i>ptumor</i>	<i>status</i>	<i>neuro</i>	1731.50
<i>age</i>	<i>sex</i>	<i>ntumor</i>		<i>diameter</i>	<i>volume</i>	<i>ptumor</i>	<i>status</i>	<i>neuro</i>	1731.80
<i>age</i>	<i>sex</i>	<i>ntumor</i>	<i>kps</i>	<i>diameter</i>		<i>ptumor</i>	<i>status</i>	<i>neuro</i>	1731.85
<i>age</i>	<i>sex</i>	<i>ntumor</i>	<i>kps</i>		<i>volume</i>	<i>ptumor</i>	<i>status</i>	<i>neuro</i>	1731.85
<i>age</i>	<i>sex</i>	<i>ntumor</i>			<i>volume</i>	<i>ptumor</i>	<i>status</i>	<i>neuro</i>	1732.16

- 選択されるモデルはパラメータ数が多い
- モデルの適切性とは無関係にパラメータ数が多いものを選択していた

補足：BIC*の考察

- 以下に変形できる

$$\text{BIC}^* = -2l(\hat{\beta}) + p \log(d/n) - \log\{n^{-p} |\mathbf{I}(\hat{\beta})|\}$$

- $p \log(d/n)$ ：打ち切りが多くなるほど $\log(d/n)$ が負の小さい値をとる, p が大きいほど小さい値をとる

**p の大きいモデルを選びやすい
(d/n に依存)**

提案法

- Cox回帰のもとでのAICの妥当性 (Xu *et al.*, 2009)
- Kullback–Leibler情報量を最小にするような基準を導出

$$\text{AICF} = -2l(\hat{\beta}) + 2p$$

- 罰則付き尤度関数ではなく通常の尤度関数を用いる
- 同様に

$$\text{BICF} = -2l(\hat{\beta}) + p \log d$$

実データへの適用

Top 5 Model based on AICF								AICF
		<i>ntumor</i>				<i>ptumor</i>	<i>neuro</i>	1753.84
		<i>ntumor</i>					<i>neuro</i>	1754.72
	<i>sex</i>	<i>ntumor</i>				<i>ptumor</i>	<i>neuro</i>	1754.90
		<i>ntumor</i>		<i>diameter</i>		<i>ptumor</i>	<i>neuro</i>	1755.83
		<i>ntumor</i>				<i>ptumor</i>	<i>status</i> <i>neuro</i>	1755.83

- AIC*最適のモデルと異なり, AICFで上位のモデルのパラメータ数は少ない

AICFで選択されたモデル

共変量		HR	SE	95%CI	P (Wald)	P (LR)
転移個数	1 vs 2 - 4	0.71	0.19	0.49 to 1.05	0.08	0.08
	5 - 10 vs 2 - 4	1.57	0.21	1.03 to 2.38	0.04	0.04
原発腫瘍	胸 vs 肺	0.94	0.26	0.57 to 1.57	0.82	0.82
	胃腸	1.60	0.36	0.79 to 3.26	0.19	0.21
	腎	0.12	1.43	0.01 to 1.99	0.14	0.02
	その他	0.87	0.55	0.30 to 2.56	0.80	0.79
神経症状	あり vs なし	1.53	0.20	1.04 to 2.24	0.03	0.04

イベントへの影響が大きい因子がきちんと選択されていた

シミュレーションによる評価

- 目的：AICF, BICF, AIC*, BIC* の性能比較
- シミュレーション条件
 - 指数分布に従う生存時間を生成
 - ハザード関数： $h_i(t) = \exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_i\}$
 - $\boldsymbol{\beta}_0^T = (\log \theta, \log \theta, \log \theta, 0, 0)$, $\mathbf{Z}_i = \{Z_{ij}\}$,
 $Z_{ij} \sim \text{Bernoulli}(0.5)$, $\theta = 1.3, 2, 4, 16$
 - $n = 100, 200, 1000$, 打切割合 $c = 0, 50, 90\%$
 - 2万回反復

候補モデル

- Model1 : $\log h_i(t) = \beta_1 Z_{i1}$
- Model2 : $\log h_i(t) = \beta_4 Z_{i4}$
- Model3 : $\log h_i(t) = \beta_1 Z_{i1} + \beta_2 Z_{i2}$
- Model4 : $\log h_i(t) = \beta_1 Z_{i1} + \beta_4 Z_{i4}$
- Model5 : $\log h_i(t) = \beta_4 Z_{i4} + \beta_5 Z_{i5}$
- Model6 : $\log h_i(t) = \beta_1 Z_{i1} + \beta_2 Z_{i2} + \beta_3 Z_{i3}$ **(true)**
- Model7 : $\log h_i(t) = \beta_1 Z_{i1} + \beta_2 Z_{i2} + \beta_4 Z_{i4}$
- Model8 : $\log h_i(t) = \beta_1 Z_{i1} + \beta_4 Z_{i4} + \beta_5 Z_{i5}$
- Model9 : $\log h_i(t) = \beta_1 Z_{i1} + \beta_2 Z_{i2} + \beta_3 Z_{i3} + \beta_4 Z_{i4}$
- Model10 : $\log h_i(t) = \beta_1 Z_{i1} + \beta_2 Z_{i2} + \beta_4 Z_{i4} + \beta_5 Z_{i5}$
- Model11 : $\log h_i(t) = \boldsymbol{\beta}^T \mathbf{Z}_i$ **(full)**

評価指標

- 平均リスクの推定値とAICF or AIC*の差
 - AICFはリスクの推定量であり, この値が0に近いほどバイアスが少ない
- 各モデルの選択割合
 - BICFはモデル選択の一致性をもつ
 - AIC*は n が大きいたときにModel11 (full)を高い割合で選択することが期待される

バイアスの評価結果 (抜粋)

c (%)	θ	n	true model AICF	true model AIC*	full model AICF	full model AIC*
90	1	100	-0.4	-2.7	-1.1	-4.8
90	1	200	0.1	-4.4	-0.1	-7.7
90	1	1000	1.6	-8.0	1.6	-14.4
90	2	100	-0.4	-2.7	-1.0	-4.7
90	2	200	-0.6	-5.2	-0.9	-8.5
90	2	1000	1.3	-8.3	1.3	-14.8
90	4	100	-0.7	-3.0	-1.4	-5.1
90	4	200	0.0	-4.6	-0.3	-7.9
90	4	1000	-1.8	-11.4	-1.9	-17.9
90	16	100	-0.7	-3.0	-1.4	-5.1
90	16	200	-0.6	-5.2	-0.9	-8.5
90	16	1000	0.1	-9.5	0.1	-15.9

モデル選択確率の評価結果 (抜粋)

c (%)	θ	n	true model AICF	true model AIC*	true model BICF	true model BIC*
0	1	100	0.270	0.000	0.087	0.304
0	1	200	0.468	0.000	0.188	0.468
0	1	1000	0.778	0.000	0.919	0.658
0	2	100	0.745	0.000	0.820	0.662
0	2	200	0.778	0.000	0.968	0.670
0	2	1000	0.786	0.000	0.991	0.670
0	4	100	0.772	0.000	0.959	0.676
0	4	200	0.773	0.000	0.971	0.664
0	4	1000	0.786	0.000	0.990	0.670
0	16	100	0.774	0.000	0.958	0.690
0	16	200	0.779	0.000	0.974	0.679
0	16	1000	0.788	0.000	0.991	0.668

モデル選択確率の評価結果 (抜粋)

c (%)	θ	n	full model AICF	full model AIC*	full model BICF	full model BIC*
0	1	100	0.033	1.000	0.001	0.067
0	1	200	0.054	1.000	0.001	0.108
0	1	1000	0.085	1.000	0.001	0.155
0	2	100	0.091	1.000	0.007	0.145
0	2	200	0.086	1.000	0.003	0.150
0	2	1000	0.081	1.000	0.001	0.149
0	4	100	0.089	1.000	0.007	0.144
0	4	200	0.085	1.000	0.004	0.150
0	4	1000	0.081	1.000	0.001	0.149
0	16	100	0.088	1.000	0.007	0.137
0	16	200	0.085	1.000	0.003	0.144
0	16	1000	0.081	1.000	0.001	0.151

結果のまとめ

- バイアス
 - AIC*に比べてAICFは圧倒的に小さい
 - AIC*は妥当でない
 - AICFはリスクの推定量を構成
- モデル選択確率
 - BICFのモデル選択一貫性
 - AIC*は確かに n が大きいと p が大きいモデルを選択する傾向

まとめ

- Firthの方法適用下での妥当なモデル選択基準を提案
 - 既存法の問題点の検証
 - AICF, BICFの提案
- ロジスティック回帰モデルの完全分離の場合にも全く同様の問題が起こる

謝辞

- 本研究においてJLGK0901 studyのデータ使用をご快諾いただいた, 勝田病院 水戸ガンマハウスの山本昌昭教授, および築地神経科クリニックの芹澤徹先生に感謝申し上げます
- 本研究はJSPS科研費 26870099, 16K16014 の助成を受けたものです

Rererences

1. Nagashima K, Sato Y. Information criteria for Firth's penalized partial likelihood approach in Cox regression models. *Statistics in Medicine* 2017;**36**(21):3422–3436. DOI: [10.1002/sim.7368](https://doi.org/10.1002/sim.7368).
2. Bryson MC, Johnson ME. The incidence of monotone likelihood in the Cox model. *Technometrics* 1981; **23**(4):381–383. DOI: [10.1080/00401706.1981.10487683](https://doi.org/10.1080/00401706.1981.10487683)
3. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika* 1993; **80**(1): 27–38. DOI: [10.1093/biomet/80.1.27](https://doi.org/10.1093/biomet/80.1.27).
4. Heinze G, Schemper M. A solution to the problem of monotone likelihood in Cox regression. *Biometrics* 2001; **57**(1): 114–119. DOI: [10.1111/j.0006-341X.2001.00114.x](https://doi.org/10.1111/j.0006-341X.2001.00114.x).
5. Yamamoto M, Serizawa T, Shuto T, et al. Stereotactic radiosurgery for patients with multiple brain metastases (JLGK0901): a multi-institutional prospective observational study. *The Lancet Oncology* 2014; **15**(4): 387–395. DOI: [10.1016/S1470-2045\(14\)70061-0](https://doi.org/10.1016/S1470-2045(14)70061-0).
6. Xu R, Vaida F, Harrington DP. Using profile likelihood for semiparametric model selection with application to proportional hazards mixed models. *Statistica Sinica* 2009; **19**(2):819–842.

付録

単調尤度の発生条件

- イベントあり ($c_i = 1$) のすべての時点について $x_k - x_i \geq 0$ (or $x_k - x_i \leq 0$)
- $k \in R(t_i)$: 時点 t_i のリスク集合の添字
- 例 : 特定の層内のイベント数が0の時
 - 全ての i について $x_k = x_i$ が成立
 - パラメータの値が非常に大きい (群間差や効果が大きい) ときに起こりうる

SASプログラム

```
data sample;
  h0 = 1;
  beta = 1;
  do id = 1 to 50;
    z = rand('Bernolli', 0.5);
    time = rand('Exponential')/
      (h0*exp(z*beta));
    event = rand('Bernolli', 0.9);
    output;
  end;
  keep id time event z;
run;
```

```
proc phreg data = sample outest = est;
  class z / param = ref ref = first;
  model time*event(0) = z / firth;
run;
ods listing close; ods results off;
proc phreg data = sample inest = est;
  class z / param = ref ref = first;
  model time*event (0) = z /
    maxiter = 0;
  ods output FitStatistics = ABICF;
run;
ods results on; ods listing;
proc print data = ABICF; run;
```

Rプログラム

```
library("survival")
library("coxphf")
n <- 50
h0 <- 1
beta <- 1
z <- rbinom(n, 1, 0.5)
time <- rexp(n) / (h0*exp(z*beta))
event <- rbinom(n, 1, 0.9)
sim <- data.frame(time, event, z)
est <- coxphf(sim,
  formula = Surv (time , event) ~ z)
```

```
abicf <- coxph(sim,
  formula = Surv(time, event) ~ z,
  init = est$coef, iter.max = 0)
est
extractAIC(abicf)
-2*(abicf$loglik[length(abicf$loglik)]+
  sum(!is.na(abicf$coefficients))*log(abicf$n))
```