

平成22年 5月 7日現在

研究種目：若手研究（B）  
 研究期間：2007～2009  
 課題番号：19700263  
 研究課題名（和文） 不完全観測を伴う計数過程情報に対する統計モデルの推測と3タイプ部分尤度の研究  
 研究課題名（英文） Inference in the statistical models for counting processes information with incomplete data and studies of three types of partial likelihoods  
 研究代表者  
 杉本 知之（SUGIMOTO TOMOYUKI）  
 大阪大学・医学系研究科・助教  
 研究者番号：70324829

研究成果の概要（和文）：ある母集団から標本抽出もしくは選択された複数の個体に対し、関心のあるイベントを経験するまでの時間を追跡し、その結果、収集されたデータから、研究対象としての科学的知見を、発見もしくは検証するための統計的推測法をとり上げる。このとき、観測された個体データが不確かな付随情報をもつ場合に、数理科学的流儀に基づいて、少なくとも3タイプの推測法が見つけられた。そして、それら3タイプが本当に科学的に妥当な方法であるかどうか、また、お互いにどのように関係し合う方法であるかを研究し、数理科学における新たな貢献や実質科学分野への応用についての有益な知見を得ることができた。

研究成果の概要（英文）：We consider methods for statistical inference in order to confirm or discovery a scientific knowledge under investigation from the data collected by following up times at which individuals sampled or selected from the population experience the event of interest. Then, if the observed data include uncertain concomitant information for individuals, it is founded that there are at least three types of statistical inference methods, based on outlooks in mathematical science. We studied whether the 3-types are truly valid methods in science, and how the 3-types are related to each other, so that we could obtain several useful findings on contributions in mathematical science and applications to substantial science fields.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	1,000,000	0	1,000,000
2008年度	800,000	240,000	1,040,000
2009年度	800,000	240,000	1,040,000
年度			
年度			
総計	2,600,000	480,000	3,080,000

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：統計的推測

## 1. 研究開始当初の背景

Cox の回帰モデルは、生存時間解析における予後因子探索ツールの代表であり、統計科学の一つのブレイクスルーとして知られている。Cox モデルでは、過去 20 数年間に、マルチンゲール理論が極めて重要な役割を果たすことが認識され(例えば、右側中途打ち切りデータのもとでの Cox モデルの諸種の推定量の漸近的性質は、前進計数過程と前進マルチンゲール理論によって構成できることが見つけられた)、計数過程から惹き出される前進マルチンゲール性質を用いて、90 年前後から Cox モデルの回帰診断法の多くが開発されてきている。

ただし、Cox モデルの適用拡大は、通常の間帰モデルほど広範囲に開拓されておらず、統計理論からみても依然として未開な部分も多い。例えば、右側中途打ち切りデータのもとで重要な役割を果たしたマルチンゲール接近法は、この適用拡大の領域では、それほど有益でないかもしれないとしてあまり整備されていない。また、仮にある有意義な拡張を与える Cox 型計数過程モデルを定式化できたとしても、何らかの不完全観測な情報を伴う形式で解析しなければならない問題をもつことも多い。とくに、不完全観測を伴う場合の Cox 型計数過程モデルの推測において、通常の統計的接近法には見られないこととして、異なる 3 タイプの部分尤度 (i) 周辺部分尤度, (ii) 双子の疑似部分尤度, (iii) EM 型を介した周辺プロフィール尤度) が出現することが挙げられる。この領域では、主として、周辺プロフィール尤度 (iii) を用いて、興味ある多くの Cox 型計数過程モデルの推測の研究が行われているが、この方法は、一般に、分散推定や診断法の開発に難があるように思え、しばしば、その計算アルゴリズムが容易に収束しない問題をもっている。また、この方法と周辺部分尤度 (i) や疑似部分尤度 (ii) との関係は、理論と方法の双方において、それほど明らかにされないままに、(iii) に基づく方法論と適用だけが進んでいるように思われた。

## 2. 研究の目的

本研究目的は、Cox モデルの適用拡大を意図して、不完全観測を伴う Cox 型計数過程モデルの統計的推測の基盤を整備することである。この推測において現れる 3 タイプの部分尤度 (i) 周辺部分尤度, (ii) 双子の疑似部分尤度, (iii) EM 型を介した周辺プロフィール尤度) に対し、統一的な科学的見解を得ることを目指す。この目標のために、重要かつ具体的な現象を扱う統計モデルとして、主に、次のような三つのモデル: (a) Cox 型治癒混合モデル, (b) 諸種のランク検定の背後に現れるある Cox モデル (確率過程上の誤差を伴う時間共

変量をもつ Cox モデル), (c) 2 重中途打ち切りデータ (区間中途打ち切りデータ) を伴う Cox モデルをとり上げる。これら三つの具体的な Cox 型計数過程モデルにおいても、推測のための尤度が、本質的に 3 タイプの部分尤度からなるため、各モデルにおいて、3 タイプの部分尤度の相互関係および役割を明らかにしながら、各部分尤度の挙動・特徴を漸近論、および有限標本の側面から捉え、詳細に調査する。

諸種のモデルに対して、疑似部分尤度 (ii) と周辺プロフィール尤度 (iii) からの回帰パラメータに関する漸近的結果を、それぞれ確立する。そして、前進マルチンゲールと後退マルチンゲール定式化とその相関構造の数理的基盤を確立し、各推定量の性質を明らかにする。そして、これらの研究に基づいて、(ii) に基づく双子の推定量の良い重ね合わせ流儀を明らかにしたい。周辺部分尤度 (i) の漸近収束結果と極限形式を与え、それらがどのように有限標本挙動の結果に反映されるかを調査する。このような結果から、経路積分構造をもつ周辺部分尤度 (i) と、マルチンゲール構造をもつ双子の疑似部分尤度 (ii) に対し、熱方程式の解法としてよく知られる Feynman-Kac 公式に類似する知見を得ることを目指す。

また、上記の三つのモデルには固有の重要な適用と応用を含んでいるため、実際の使用に耐えることができるような計算アルゴリズムの研究に加え、診断法の開発とデータ解析の研究にも注力したい。このとき、通常の間帰モデルでの類推から、双子の疑似部分尤度から得られる前進・後退マルチンゲール残差の役割を研究する。

## 3. 研究の方法

三つの具体的な現象: (a) Cox 型治癒混合モデル, (b) 諸種のランク検定の背後に現れるある Cox 回帰モデル, (c) 2 重中途打ち切りデータを伴う Cox 回帰モデルの推測の研究を行いながら、3 タイプの部分尤度の性質や特徴を明らかにする。

他のモデルよりも基礎マルチンゲール性質を抽出しやすい、(c) 2 重中途打ち切りデータをとりあげ、疑似部分尤度 (ii) に内在するマルチンゲール構造を定式化し、前進マルチンゲールと後退マルチンゲールの相関構造を定式化するための数理的基盤を確立する。この性質を利用して、関心のあるパラメータ推測のための漸近理論を構築する。疑似部分尤度 (ii) と周辺プロフィール尤度 (iii) に伴う潜在ハザード関数のノンパラメトリック推定量に関する漸近結果を利用し、(ii) と (iii) からの回帰パラメータに関する漸近的結果を、それぞれ確立する。このとき、2 重中途打ち切りデータを伴う Cox モデルの推測では、(ii) と (iii) の関連を追及しやすいこ

とが予想できるため、ここで得られる知見を慎重に吟味する。また、この知見に基づいて、(ii)に基づく双子の推定量の良い重ね合わせ流儀を明らかにする研究を行う。これらの研究方法を、他のCox型計数過程モデルにも類似させ、他のモデルに対する研究の模範とすることを考える。

周辺部分尤度(i)の研究は、数値的および解析的に扱いやすく、他のモデルの模範となり得るために、Cox型治癒混合モデルの周辺部分尤度の研究から出発する。周辺部分尤度の全体和には、大数の法則を直裁に使用できないが、不完全観測に伴う $2^m$ 個からなる全経路の各部分尤度内の要素に対し、 $\{0, 1\}^m$ 上の部分和過程の理論を用いることができると考えられる。その結果、周辺部分尤度(i)の漸近収束結果を証明する。この漸近収束結果をつける極限形式が、どのような収束率をもって、どのように有限標本挙動の結果に反映されるかを確認・調査する。ここでの周辺部分尤度(i)からの推定量の挙動をみるためのシミュレーション研究の方法では、アンサンブル方式それ自身に対するEMアルゴリズム的な解法などを行い、計算上の工夫を行う。また、周辺部分尤度から導かれる潜在ハザード関数の推定値がどのように構築でき、これが疑似部分尤度(ii)や周辺プロフィール尤度(iii)から得られる推定潜在ハザード関数とどのように異なるかを研究する。これらの研究に基づき、さらに、周辺部分尤度を構築するための高次元の多重積分に対する、あるLaplace近似の方法を研究することで、周辺部分尤度(i)からの回帰パラメータと潜在ハザード関数の推定量の大標本理論を構築する。このような結果から、経路積分構造をもつ周辺部分尤度(i)と、マルチンゲール構造をもつ双子の疑似部分尤度(ii)に対し、熱方程式の解法のFeynman-Kac公式に類似する知見が得られるかどうかを研究する。

#### 4. 研究成果

(1): 双子の疑似部分尤度(ii)の特色をより抽出しやすく、周辺プロフィール尤度(iii)の結果と比較しやすい2重中途打ち切りデータを伴う帰無モデルとしてのCoxモデルをとり上げ、(ii)、(iii)に関するいくつかの研究成果を得た。まず、(iii)から(ii)を橋渡しするために、前進マルチンゲールと後退マルチンゲールの構造を定式化し、重ね合わせフィルトレーションを用いることによって、前進と後退のマルチンゲールの相関を定式化できた。これにより、それらのマルチンゲール構造から惹き出される前進型推定と後退型推定の重ね合わせ推定値の数理構造を定式化できる見通しが得られた。さらに、(iii)から得られるノンパラメトリック最尤推定量(NPMLE)の分散公式を提案し、その分

散公式の漸近一致性の証明と、有限標本上の性質を調べるためのシミュレーション研究を行った。とくに、このNPMLEの分散推定量の研究では、過去において、ブートストラップ法、セミパラメトリック・プロフィール尤度比、Fredholm積分方程式を解くことなど、計算上の大きな負荷をもって分散推定値を求めることが必要であったが、本研究によって、より簡便に反復解法なしに求めることができること、そのような分散公式を通じてNPMLEに内在する前進・後退マルチンゲールの構造を解明するためのヒント、および、この定式化の妥当性確認を得ることができた。

(2): (1)の研究に基づいて、2重中途打ち切りデータを伴うCoxモデルの帰無形式に対して、周辺プロフィール尤度に対する有効スコア関数に、二三のTaylor近似を適用することで、周辺プロフィール尤度(iii)の最尤推定値であるNPMLEにマルチンゲール構造を導入することに成功した。その結果、NPMLEが前進マルチンゲールGaussian過程と後退マルチンゲールGaussian過程の重ね合わせ分布に弱収束することを、すなわち、過去の研究よりも見通しのよい漸近分布の定式化を達成することができた。このような結果は、諸種の不完全観測を伴うCox型計数過程モデルにおいて共通にみられる事実であることが示唆され、本研究目的における一つの目標である、疑似部分尤度(ii)と周辺プロフィール尤度(iii)の統合にむけて、それらの関連性の研究を達成することができた。

(3): (1)と(2)の研究において、高次元3重対角化行列の逆行列に関する二三の性質が重要なキーであったため、この逆行列公式とそれらの性質の数理構造を、独立した研究として纏めた。

(4): (3)の研究で定式化した3重対角化行列の逆行列の性質を用いて、Cox型治癒混合モデルにおける推定生存時間関数の分散公式および信頼区間構成法を与える方法として、従来の計算不可能になる無限次元行列形式を避け、大標本でも計算可能な微分形式に基づく分散公式を研究した。これにより、Cox型治癒混合モデルにおける前進と後退のマルチンゲール接近法が(2)の研究に類似する手順で行うことができる見通しが得られた。

(5): 治癒混合モデルをとり上げ、周辺部分尤度(i)の漸近収束を研究した。この収束結果を得るための理論は、2値欠測共変量、2重中途打ち切りデータのCoxモデルにもほとんど直截に適用できることがわかり、これらの成果を一つの研究成果として纏めた。ここで得られた周辺部分尤度(i)の極限形式が有限標本においてどのように振舞うかを確認するためのシミュレーション研究を行った。

その結果、周辺部分尤度の厳密計算で必要とされる  $2^m$  規模の反復数よりも相当に小さな、標本サイズの大きさに比例する反復数を用いたモンテカルロ法によって、十分な近似精度が得られることを検証することができた。

(6): 治癒混合モデルに対して、周辺部分尤度(i)に基づく回帰パラメータの大標本理論を研究した。このとき、(5)の研究結果から周辺部分尤度の漸近収束が保証されるため、周辺部分尤度を構築するための高次元の多重積分に対する、ある Laplace 近似の方法を研究した。このとき、これまでの研究[(4)や Sugimoto & Hamasaki(2006)]を駆使して、Laplace 近似法を得るための重要な結果、全尤度のハザード関数に関する 2 次微分行列の行列式の漸近収束を得た。Laplace 近似の方法を介して、周辺部分尤度(i)に基づく回帰パラメータの大標本理論を研究した。さらに、Laplace 近似の結果に基づいて、最終的に得られる周辺部分尤度の極限形式は、周辺プロフィール尤度(iii)に比例するものとして得られることがわかった。また、この Laplace 近似法を用いて、周辺部分尤度の条件付き平均(経路積分型平均)として導かれる潜在基礎ハザードの推定量が、周辺プロフィール尤度(iii)から Fredholm 積分方程式を解くことにより得られる潜在基礎ハザード関数の推定量と漸近同等になるという知見も得ることができた。

(7): 治癒混合モデルにおいて、実際の使用に耐えることができるような計算アルゴリズムの研究を行い、データ解析の研究(うつ病のある臨床試験データへの適用)を行った。

(8): ランク検定の背後に現れるある Cox モデルの研究において、潜在次元が様々に変化する潜在モデルをとり上げ、BIC 規準に基づいて実行するデータ適応型のログランク検定法を提案した。この方法の漸近理論と研究し、シミュレーションを用いて有限標本挙動を調査した。

(9): 中央生存時間の比較の問題において、マルチンゲール接近法と Laplace 近似法を用いて、中央順序統計量の性質を研究し、2 標本中央生存時間の検定法を提案した。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 5 件)

- ① Sugimoto, T., A Wald-type variance estimation for the nonparametric distribution estimators for doubly censored

data, Annals of the Institute of Statistical Mathematics, 査読有, (印刷中)

- ② Arano, I., Sugimoto, T., Hamasaki, T. and Yuko Ohno, Detailed analyses of a sertraline randomized withdrawal study in patients with major depressive disorder, BMC Medical Research Methodology, 査読有, 2010, 10:33
- ③ 尼ヶ崎太郎・杉本知之・後藤昌司, 2 標本における中央生存時間の推測、日本統計学会誌、査読有、39 巻(J-1)、2009、pp.95-119
- ④ 杉本知之、中途打ち切りデータに対するあるデータ適応型重み付けログランク検定方式、日本統計学会誌、査読有、38 巻(J-2)、2009、pp.165-187
- ⑤ 元垣内広毅、杉本知之、後藤昌司、MARSにおける非負圧縮推定量とその性能、応用統計学、査読有、36 巻、2007、pp.99-118

[学会発表] (計 3 件)

- ① Nakatsuji, Y., Moriya, M., Okuno, T., Kinoshita, M., Sugimoto, T., Nakano, M., Kikutani, H., Sakoda, S., Kumanogoh, A., Sema4A as a diagnostic marker and its pathogenic roles in multiple sclerosis, 25th Congress of the European Committee for the Treatment and Research in Multiple Sclerosis, September 9-12, 2009, Germany
- ② Sozu, T., Sugimoto, T., Hamasaki, T., Sample size determination in clinical trials with multiple co-primary binary endpoints, the 2009 Joint Statistical Meetings, August 1-6, 2009, Washington DC, USA
- ③ 朝倉こう子、上坂浩之、杉本知之、濱崎俊光、相関をもつ二つの変数の比に対する解析、応用統計学会 2009 年度年会、8 月 1 日、2009、東京

## 6. 研究組織

### (1) 研究代表者

杉本 知之 (SUGIMOTO TOMOYUKI)

大阪大学・医学系研究科・助教

研究者番号：70324829