

氏 名 尾崎 凌斗

学位(専攻分野) 博士(統計科学)

学位記番号 総研大甲第 2451 号

学位授与の日付 2023 年 9 月 28 日

学位授与の要件 複合科学研究科 統計科学専攻
学位規則第6条第1項該当

学位論文題目 Cox 比例ハザードモデルにおける変化点検出のための情報
量規準

論文審査委員 主 査 船渡川 伊久子
統計科学コース 准教授
鎌谷 研吾
統計科学コース 教授
二宮 嘉行
統計科学コース 教授
杉本 知之
滋賀大学 データサイエンス学系 教授

博士論文の要旨

氏名：尾崎 凌斗

論文題目：Cox 比例ハザードモデルにおける変化点検出のための情報量規準

はじめに

Cox 比例ハザードモデルは、time-to-event の評価項目を設定した臨床研究における生存時間解析の手法として広く用いられている。そのモデルには、比例ハザード性や対数線形性といった潜在的な仮定が含まれている。実際の臨床試験におけるデータの構造は、しばしばこれらの仮定から逸脱していることがある。比例ハザード性が成立しない状況では、それを前提とした解析手法をそのまま適用することに理論的妥当性はなく、共変量の検出力が下がることや推定されたハザード比の解釈が困難となることが指摘されている。また、対数線形性の仮定からの逸脱がある場合も、推定量は一致性等の良い性質をもたないため、ハザード比の解釈が困難となる。免疫チェックポイント阻害剤やがんワクチン等の治療薬とプラセボ等の対照薬を比較する臨床試験では、その作用機序を考慮すると治療開始から効果発現までに時間を要すると考えられている。このような臨床試験では、観察期間全体で比例ハザード性は成立せず、ハザード比関数に 1 つ以上の急激な変化が生じていると考えられることから、その時点を検出する変化点解析が生存時間解析において重要視されている。また、特に観察研究や標本サイズに対して共変量数が多い場合等では、共変量間に多重共線性が存在することが多く、それは推定が不安定になる原因となる。そのため、Cox 比例ハザードモデルを用いた生存時間解析では、代替手法としてリッジ回帰が用いられている。そこで本論文では、Cox 比例ハザードモデルと変化点モデルを組み合わせたものを考え、推定に用いる損失関数にリッジタイプの正則化項を加え、予測の観点でモデル評価を行う赤池情報量規準 (AIC) 元来の導出に基づくという意味で合理的な、すなわち、正則化項を加えた損失関数に対して Kullback-Leibler ダイバージェンスに基づくリスク関数に対して漸近不偏推定量が得られる情報量規準を開発した。更に、対数線形性が成立しない状況や、観察研究等でハザード比推定に必要なすべての共変量を測定できていない状況では、推定に用いるモデルは誤特定している可能性がある。そこで、モデル誤特定下での AIC、いわゆる竹内情報量規準 (TIC) への拡張も行った。

また、変化点解析は計量生物の領域だけでなく、様々な応用に利用されているが、特に計量経済の分野でも需要が大きい。不均一分散性を持つ時系列データに用いられる自己回帰条件付き不均一分散モデル (ARCH モデル) は、特に計量経済の分野で頻用される標準的なモデルの一つである。そこで本論文では、ARCH モデルを用いた変化点解析のための合理的な情報量規準も開発した。ARCH モデルが計量経済学の分野で広く用いられている一方、ARCH モデルを適用する実際の時系列データが忠実にそのモデルに従っている状況はほとんどないであろう。つまり、モデル誤特定は頻繁に発生しているはずである。そこで、変化点をもつ ARCH モデルに対する TIC の導出を行った。

方法・結果

まずは、Cox 比例ハザードモデルにおける変化点モデルの定義を与え、変化点解析における漸近論のために通常置かれる条件を仮定する。AIC 元来の導出に基づき、正則化最大対数部分尤度を漸近的にバイアス補正したものとして、AIC タイプの情報量規準を定義した。そして、その罰則項となる正則化最大対数部分尤度の漸近バイアスを評価した。その結果、変化点によって生じる漸近バイアスは、負のドリフト項をもつ両側ランダムウォークに関する期待値で表せることを示した。更に、変化点解析の漸近論においてしばしば課される自然な条件を追加で仮定してその期待値を解析評価すると、漸近バイアスがシンプルかつ陽な形で評価できることを示した。それにより、正則化部分尤度法で推定した際の AIC が造作なく与えられ、正則化項がない場合には、変化点パラメータによる漸近バイアスは更に容易に 3 (通常のパラメータならば 1) と書けることを確認した。また、モデル誤特定下での推定量の漸近的性質を与え、同様に AIC 元来の導出に基づいてモデル誤特定を許容する情報量規準として TIC も導出した。数値実験では、本論文で評価した漸近バイアスが精度よく近似できていることを、まず示した。更に、真の構造に近い推定を与えるという AIC 本来の目的に照らすと、形式的な情報量規準や逐次検定アプローチを適用した場合と比べ、提案の規準は明らかに真の構造に近い Kullback-Leibler ダイバージェンスを与えることから、それは既存の規準をはっきりと優越することを確認した。実データ解析では、形式的な AIC が漸近バイアスを過小評価しているために変化しているとは考えにくい時点を変化点として検出してしまっている一方、提案の規準は変化していると想定される時点を適切に検出しているであろうことを確認した。

また、通常 ARCH モデルでは、時系列データの観測時点までの情報集合を与えた下での条件付き分布に正規分布を仮定するが、モデルが誤特定された状況として、その条件付き分布が正規分布とならない状況を考えた。そして、Cox 比例ハザードモデルに対する情報量規準の開発と同様に、AIC 元来の導出方法に基づいて漸近バイアスを再評価し、変化点をもつ ARCH モデルに対する情報量規準として TIC を導出した。そして、既報の独立系列に対する変化点モデルと同様、モデル誤特定がない下では、AIC における変化点パラメータの罰則は通常のパラメータの罰則の三倍となるものの、TIC ではそこから少し乖離することを確認した。数値実験では、形式的に AIC を適用することに比べ、変化点モデルの非正則性を考慮して厳密に導いた AIC を用いる方が、明らかに真の構造に近い推定を与えることを示した。また、モデル誤特定があるときは、それよりもさらに TIC が妥当な結果を返す、つまり上述の乖離には意味があることを確認した。

考察

変化点をもつ Cox 比例ハザードモデルは、通常の変化点モデルとは異なり、時間が結果変数の取り得る値となっている。これは、新たな漸近論の必要性を予想させるが、Cox の部分尤度法ではイベント発生時点のハザードを直接モデル化していることで、実は変化点解析と同じ漸近論を用いるので十分となることを本論文で明らかにした。また、分散変動をモデル化している ARCH モデルでは変化点推定量の分散が通常の変化点モデルと同一か自明でなかったものの、速い収束レートをもつことを明らかにした。その結果、変化点をもつ ARCH モデルに対する TIC の罰則項はそのモデル固有のものになることを示した。

博士論文審査結果

Name in Full
氏名 尾崎 凌斗

Title
論文題目 Cox 比例ハザードモデルにおける変化点検出のための情報量規準

2023年7月24日15時から約2時間にわたり尾崎凌斗氏の博士論文審査委員会を本人および4名の審査委員全員の出席のもとに行った。出願者による1時間の公開発表による概要説明と質疑応答、さらに約1時間の審査委員のみによる審査を行った結果、審査委員会は本論文が学位（統計科学）の授与に値すると判断した。

[論文の概要]

提出された論文は、全7章と引用文献で計78頁からなり、日本語で書かれている。

第1章は、序章である。Coxモデルは、生存時間解析における標準的なモデルであるが、比例ハザード性や対数線形性という強い仮定を課しており、実際の臨床試験におけるデータの構造はこれらをしばしば満たさない。例えば免疫チェックポイント阻害剤やがんワクチン等の治療薬とプラセボ等の対照薬を比較する臨床試験では、その作用機序を考慮すると治療開始から効果発現までに時間を要すると考えられ、観察期間全体で比例ハザード性は成立せず、ハザード比関数に急激な変化が生じていると考えられる。そして、このようなケースでは変化点解析を組み込むことが重要視されてきている。また、標本サイズに対して共変量数が多いケースでは、共変量間の多重共線性による推定の不安定性を回避するため、推定に用いる損失関数にリッジ正則化項を加えることが多々ある。モデル選択が不可欠であることは、これらの設定においても変わらない。そこで、本論文ではCoxモデルと変化点モデルを組み合わせたモデルを扱い、部分尤度版のKullback-Leiblerダイバージェンスという適切なリスクに対する漸近不偏推定量としての情報量規準を導いている。これは、赤池情報量規準(AIC)元来の導き方と同じであり、変化点やリッジ正則化がなければ情報量規準の罰則項は二倍のパラメータ数となる。

第2章では、準備として、モデルの定義、仮定する条件、正則化最大対数尤度を漸近的にバイアス補正した場合のAICタイプの情報量規準の定義を与えている。

第3章では、漸近バイアスの導出とその解析評価を行っている。導出の結果、変化点によって生じる罰則項は、負のドリフト項をもつ両側ランダムウォークの関数の期待値で表せることを示している。更に、変化点解析の漸近論においてしばしば課される自然な条件を追加で仮定してその期待値を解析評価すると、シンプルかつ陽な形になることを示している。それにより、AICは造作なく与えられ、正則化がない場合には、変化点パラメータによる罰則項は6（通常のパラメータによる罰則項は2）と書けることを確認している。

第4章では、数値実験と実データ解析を行っている。数値実験では、本論文で導いた情報量規準が、部分尤度版のKullback-Leiblerダイバージェンスを精度よく近似できていることを、まず示している。そして、形式的なAICや形式的なBIC、逐次検定手法と比べ、

真の構造に近い推定を与えるという AIC 本来の目的を、提案の規準が達成することを確認している。実データ解析では、提案の規準と比べ、形式的な AIC が罰則項を過小評価していることから過剰適合して多くの変化点を検出していること、形式的な BIC が罰則項を過大評価していることから過少適合して少ない変化点しか検出していないこと、つまり提案の規準と既存の規準は大きく異なる結果を返すことを確認している。

第 5 章では、観察研究等でハザード比推定に必要なすべての共変量を測定できていない状況ではモデル誤特定が起こることを鑑み、モデル誤特定下での AIC、いわゆる竹内情報量規準 (TIC) への拡張を行っている。

第 6 章では、計量経済分野での変化点解析での情報量規準の開発を行っている。変化点解析は、計量生物の分野以上に計量経済の分野での需要が大きい。分散不均一性を有する時系列データに用いられる自己回帰条件付き分散不均一モデル (ARCH モデル) は、そこで重用される標準モデルの一つである。本論文では、ARCH モデルを用いた変化点解析のための合理的な情報量規準である AIC と TIC も開発している。ARCH モデルは推定が不安定である傾向があり、複雑な罰則項を有する TIC の性能が憂慮されたが、数値実験では、サンプルサイズが大きければ他を優越することが示されている。

第 7 章は、本論文のまとめである。

[論文の評価]

本論文で組み合わされて用いられているのは Cox モデルのための部分尤度に基づく漸近理論と変化点解析のための漸近理論であり、どちらも専門性が高い。変化点モデルの情報量規準において正則化推定もモデル誤特定もそれらがどう影響するかは自明でなく、実際、情報量規準の式は予想できないものとなっており、その新規性は十分に評価できる。また、Cox モデルの強い条件を緩和するという計量生物分野の近年の課題にアプローチする有力な候補モデルに対し、統計解析に不可欠なモデル選択手法を開発しており、今後重用される可能性を有し、実用面での貢献も期待できる。以上より、本論文は統計科学分野の博士論文として十分に高い水準に達していると判断する。

[その他]

Cox モデルに対する本論文の主結果は、査読付国際学術雑誌 **Biometrics** に掲載されている。第 6 章の ARCH モデルに対する結果は、現在投稿中である。