

博士論文

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

尾崎 凌斗

博士（統計科学）

総合研究大学院大学
複合科学研究科
統計科学専攻

2023年9月

目次

第1章 序章	4
第2章 準備	10
2.1 モデルと仮定	10
2.2 部分尤度法に基づく推定量の漸近的性質	12
2.3 AIC タイプの情報量規準	13
2.4 部分尤度に基づく AIC タイプの情報量規準	14
第3章 主結果	16
3.1 漸近バイアスの導出	16
3.2 漸近バイアスの解析評価	18
第4章 数値実験	22
4.1 シミュレーション実験	22
4.1.1 形式的な情報量規準や既存手法との比較	22
4.1.2 交差検証法との比較	24
4.2 実データ解析	29
第5章 モデル誤特定を許容する情報量規準への拡張	34
第6章 ARCH モデルへの拡張	41
第7章 結論	62
付録 A (3.1) の導出過程	65
付録 B (3.3) の導出過程	67

	3
付録 C (3.9) の導出過程	69
付録 D 条件 (3.11) の下での変化点推定量の漸近挙動	71
付録 E 定理 2 の証明	72
謝辞	76
参考文献	77

第1章 序章

Cox (1972) により提案された 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。比例ハザード性とは時間に依らずハザード比が一定であるという性質であるが、実際の臨床研究におけるデータでは、しばしばこ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（例えば Uno et al. 2014 参照）。比例ハザード性からの逸脱に関する問題は古くからあり、Gill and Schumacher (1987) や Hess (1995) 等は、その仮定からの逸脱を検出する手法や代替手法を提案している報告である。

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研究では、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 (FDA 2011)。このような臨床研究では生存曲線が治療開始から一定期間重なったまま推移しその後曲線に差が生まれることが想定される。効果発現に遅発性がある状況だけでなく、離れていった生存曲線がしばらくして平行になる状況や、平行からさらに2つの生存曲線が重なり合う状況等でも、比例ハザード性が成立しない (Ananthakrishnan et al. 2021)。これらでは、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。このようにベースライン共変量のイベント発現に対する影響が追跡期間中一定ではない状況をモデルに反映するために、通常の Cox 比例ハザードモデルに時間依存係数を組み込んだモデルが用いられる（例えば Therneau and Grambsch 2013 参照）。具体的には、回帰係数を定数とするのではなく、時間に関する関数と考えるわけだが、その関数は時間に関して連続な関数又は一定区間ごとに定数となる階段関数に大別される。臨床研究におけるハザード比関数の変化を表現しているのは後者の関数と考えるのが自然であり、更に後者の関数を用いる方が解釈が容易であることも指摘されている (Thomas and Reyes 2014)。実

際, Liu et al. (2008) や He et al. (2013) は, 区分的にはハザード比が一定であると仮定した下でその区分の変化点を検出するための手法として, それぞれ最大スコア検定をベースとした手法や逐次検定アプローチを提案している等, 様々な研究報告が存在している. しかし, 今なお情報量規準による変化点検出の手法は報告されていない. 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状況等では, 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 (例えば Houwelingen et al. 2006 や Witten and Tibshirani 2010 参照).

比例ハザード性の成立が疑われる状況では, 例えば重み付きログランク検定 (例えば Fleming and Harrington 2013 参照) や制限付き平均生存時間 (Uno et al. 2014) といった比例ハザード性を仮定しない手法が選択され得る. 比例ハザード性の仮定は必要としないが, これらの手法を用いるには, 重みを大きく与える期間や平均生存時間を推定する期間を規定する必要がある, そのためには変化点を特定することが重要となる. そして, 早期相の臨床試験や同一の機序をもつ薬剤の臨床研究等の情報を基にそれらを研究開始前に事前に規定する必要がある. つまり, 今後得られるデータにおける変化点の数とその位置を, 過去のデータから予測することが求められる.

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, 予測の観点でモデル評価を行う赤池情報量規準 (AIC; Akaike Information Criterion, Akaike 1973) 元来の導出に基づくという意味で合理的な, すなわち適切に定義された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する. 具体的には, 部分尤度ベースで議論を展開し, 推定に用いる損失関数としては正則化項を加えたものも考える. そして, 通常 AIC タイプの情報量規準を鑑み, m 個の変化点をもつモデルの AIC を開発する際には m 個の変化点の存在を仮定する. これにより, ベースとする理論としては, 部分尤度法に基づく回帰パラメータの推定量の漸近的性質を導いた Tsiatis (1981) と変化点モデルにおける漸近理論を評価した Pons (2002) を用いる. 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を定義し, その漸近不偏推定量を導出する.

変化点モデルは固有の漸近理論を必要とするため (例えば Csörgő and Horváth 1997 参照), 検定ベースによるアプローチでも様々な理論が開発されており, そのことは情報量規準の開発においても固有の理論を必要とすることを示唆する. 実際に変化点モデル固有の理論を用い, 情報量

規準元来の定義に基づき、つまり形式的なものではなく漸近論に基づいた妥当なものとして、初めて情報量規準を導いたのは Siegmund (2004) である。そこでは、ベイズファクターからいわゆるベイズ情報量規準 (BIC; Bayesian Information Criterion, Schwarz 1978) が導出され、さらに Zhang and Siegmund (2007) と Hannart and Naveau (2012) によって一般化されている。本論文では、予測の観点で良い変化点モデル選択を行いたいため AIC タイプの情報量規準の導出を目的とする。時点ごとにデータが独立に観測されているような基本的な変化点モデルに対し、その導出をおこなったのは Ninomiya (2015) である。そこでは、AIC 元来の定義に基づき、真の分布と推定された分布の Kullback-Leibler ダイバージェンスの漸近不偏推定量として、規準を与えている。その導出によれば、通常のパラメータによる漸近バイアスが 1 であるのに対し、変化点パラメータによる漸近バイアスは 3、つまり 3 倍となる。つまり、変化点モデル固有の理論を AIC 導出において考慮することは、解析結果に大きな違いを生み出すということを示唆する。

本論文でも同様に、変化点をもつ Cox 比例ハザードモデルに対する固有の漸近理論を構築し、それを用いてリスク関数の漸近不偏推定量を評価する。まずは、正則化最大対数部分尤度のリスク関数に対する漸近バイアスを評価し、特に変化点パラメータ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す。そして、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下では、その期待値は両側ブラウン運動に関する期待値で表せ、結果、シンプルかつ陽な形で漸近バイアスを評価できることを示す。それにより、正則化項を加えた部分尤度を用いて推定した際の AIC が造作なく与えられることを確認する。また、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 3 と、更に容易に書けることを確認する。シミュレーション実験では、評価した漸近バイアスが精度良く近似できていることをまずは確認する。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、既存規準と比して提案する規準が明らかに真の構造に近い K-L を与えることから、予測の意味で良いモデル選択ができることを確認する。また、交差検証法と比した性能評価も行い、予測の観点でのモデル選択を目的としている場合には提案規準を用いることが妥当であることを示す。実データ解析では、形式的な AIC 又は形式的な BIC は、それぞれ漸近バイアスを過小評価又は過大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たり、変化していると考え得る時点を変化点として検出しなかったりする一方、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得ることを確認する。

本論文で扱う変化点をもつ Cox 比例ハザードモデルは、観察期間全体では比例ハザード性が成立しないが、区分的には比例ハザード性が成立しているモデルを考える。このモデルにおいても、通常の Cox 比例ハザードモデルと同様に各区間での対数線形性は仮定している。対数線形性とは、ハザード関数が共変量に依存する部分の対数を取ると一次式で書けることを意味する。観察研究ではハザード比の一致推定に必要なすべての共変量を測定できていることは少なく、仮に対数線形性が成立していたとしても推定に用いるモデルは誤特定している可能性がある。もちろん共変量の影響が一次式で表現できない場合も、対数線形性の仮定からの逸脱があるわけで、このようなときに通常の部分尤度法に基づいて回帰パラメータを推定すると、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。Cox 比例ハザードモデルにおけるモデル誤特定下での推定に関する議論は、[Struthers and Kalbfleisch \(1986\)](#) や [Lin and Wei \(1989\)](#) でなされている。本論文では、これらをベースにモデル誤特定下での AIC、いわゆる竹内情報量規準 (TIC; Takeuchi Information Criterion, [Takeuchi 1976](#)) への拡張も行う。拡張した TIC の性能評価として、モデル誤特定下では TIC が予測の観点から本論文で提案する AIC 及び既存規準よりも優れた性能を持つことを確認する。

ここで、変化点をもつ Cox 比例ハザードモデルでは、通常の変化点モデルと異なり、時点自体が結果変数の取り得る値となっていることに言及しておく。時点が観測されていないとき、あるいは両者が同じでノイズが加わった形で観測されているとき、その変化点モデルは、いわゆる混合分布モデルと同タイプの構造をもつようになる。[Dacunha-Castelle and Gassiat \(1999\)](#) から分かるように、混合分布モデルも、やはり正則な統計モデルとは異なる漸近的性質をもつが、変化点モデルの漸近的性質とも完全に異なる。そう考えると、本論文で扱う AIC は、通常の変化点モデルに対する AIC と同タイプのものであるか、混合分布モデルに対する AIC と同タイプのものであるか、自明ではない。本論文ではこれを明らかにする。

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。例えば、計量経済学の雑誌に掲載されている [Andrews \(1993\)](#) や [Bai and Perron \(1998\)](#), [Qu and Perron \(2007\)](#), [Carrasco et al. \(2014\)](#) は、変化点解析における理論面で画期的な進歩をもたらしたものである。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; Autoregressive Conditional Heteroscedasticity model, [Engle 1982](#)) 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、AIC 元来の導出方法と同様にして、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を開発する。

ARCH モデルは計量経済の領域で広く用いられていることから当然ではあるが、変化点をもつ ARCH モデルに対する統計的漸近理論の議論も発展してきた。例えば、[Chu \(1995\)](#) は、条件付き分布が正規分布としたときの変化点有無を比較するラグランジュ乗数検定に対する帰無仮説下での極限定理を導出している。また、[Kokoszka and Leipus \(1999\)](#) と [Kokoszka and Leipus \(2000\)](#) は、条件付き分布が正規分布でなかったとしても、変化点有無を比較する CUSUM 統計量が帰無仮説下で極限定理が成り立つことを示しており、更に対立仮説下では CUSUM 統計量に基づく推定量が一致推定量になることを示し、その収束レートも与えている。本論文では、尤度ベースで議論を展開するために、条件付き分布は正規分布に従うものとし、ARCH モデルにおける最尤推定量の漸近的性質を導出した [Weiss \(1986\)](#) を用いる。また、変化点をもつ 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、通常の AIC タイプの情報量規準を鑑み、 m 個の変化点をもつモデルの AIC を開発する際には m 個の変化点の存在を仮定する。

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。モデル誤特定があったとしても、それは誤差としてある程度吸収されてしまうため、例えば、短期間先の予測値や変化点の推定値への影響は小さいかもしれないが、分散の評価には無視できない影響を与える可能性がある。そのため、情報量規準における罰則項が無視できない程度に変化することが考えられる。そこで、モデル誤特定下での AIC タイプの情報量規準として、いわゆる TIC の導出を目指す。モデル誤特定下で情報量規準の重要性は [Lv and Liu \(2014\)](#) で指摘されているものの、変化点モデルに対して実際に導出したのは、本論文における Cox 比例ハザードモデルに対するものと、ARCH モデルに対するものが初めてとなる。

本論文の構成は以下の通りである。2 章では、準備としてまず 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。そして、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する。3 章では、正則化最大対数部分尤度の漸近バイアスを評価し、漸近バイアスは陽な形で表現できることを示す。4 章では、4.1 節で数値実験を、4.2 節で実データ解析を行う。4.1.1 節では、まず、漸近的なバイアス評価が、実際にバイアスを精度よく

近似していることを確認する。そして、導出した AIC と、変化点モデル固有の理論を用いずに形式的に定義した情報量規準、及び逐次検定アプローチとを、性能比較する。4.1.2 節では、提案する AIC と交差検証法の性能比較を行う。4.2 節では、実際の臨床試験データに本論文で提案する AIC と形式的に定義した情報量規準を適用し、変化点検出を行った事例を示す。5 章では、モデル誤特定を許容する場合の情報量規準として TIC への拡張を行う。そして、6 章では、計量経済学の領域で頻用される ARCH モデルにおける変化点モデルを考え、TIC と AIC を導出する。最後に、7 章で結論と今後の展望をまとめる。

第2章 準備

2.1 モデルと仮定

回帰パラメータの構造変化をもつ Cox 比例ハザードモデルに対して、 m 個の変化点をもつモデル

$$\lambda(t | \mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta}^{(j)\top} \mathbf{z}), \quad t \in [k^{(j-1)}, k^{(j)}), \quad j \in \{1, 2, \dots, m+1\} \quad (2.1)$$

を考える。ここで、 $\lambda(t | \mathbf{z})$ は共変量ベクトル \mathbf{z} が与えられた下でのハザード関数、 $\lambda_0(t)$ はベースラインハザード関数とする。各 $j \in \{1, 2, \dots, m+1\}$ に対して、回帰パラメータ $\boldsymbol{\beta}^{(j)}$ は p 次元ベクトル、つまり $\boldsymbol{\beta} \equiv (\boldsymbol{\beta}^{(1)\top}, \boldsymbol{\beta}^{(2)\top}, \dots, \boldsymbol{\beta}^{(m+1)\top})^\top$ は $p(m+1)$ 次元ベクトルとする。 $\mathbf{k} \equiv (k^{(1)}, k^{(2)}, \dots, k^{(m)})^\top$ は変化点パラメータを集めた m 次元ベクトルであり、 $k^{(0)} = 0$ かつ $k^{(m+1)} = T$ とする。ここで、 T は生存時間の追跡期間を意味する。また、 \mathbf{k} と $\boldsymbol{\beta}$ の真値をそれぞれ $\mathbf{k}^* = (k^{*(1)}, k^{*(2)}, \dots, k^{*(m)})^\top$ と $\boldsymbol{\beta}^* = (\boldsymbol{\beta}^{*(1)\top}, \boldsymbol{\beta}^{*(2)\top}, \dots, \boldsymbol{\beta}^{*(m+1)\top})^\top$ で表し、 $0 < k^{*(1)} < k^{*(2)} < \dots < k^{*(m)} < T$ と仮定する。変化点モデルを考えているため

$$\boldsymbol{\beta}^{*(1)} \neq \boldsymbol{\beta}^{*(2)} \neq \dots \neq \boldsymbol{\beta}^{*(m+1)} \quad (2.2)$$

という条件を仮定し、 \mathbf{k}^* と $\boldsymbol{\beta}^*$ は未知とする。

Cox 比例ハザードモデルの回帰パラメータは通常 Cox (1972) により提唱された部分尤度法によって推定される。本論文でも同様に、部分尤度法をベースに議論を展開する。まず、 y_1 及び y_2 をそれぞれイベント及び打ち切りの発生時間を意味する正の確率変数とする。 y_1 及び y_2 は共変量ベクトル \mathbf{z} が与えられた下で条件付き独立であると仮定する。実際に観察が終了するまでの時間、つまりイベント又は打ち切りが発生するまでの時間 t は、 $t = \min(y_1, y_2)$ と書ける。また、 y_1 及び y_2 の定義から指示確率変数 δ を $y_1 \leq y_2$ (イベント) で 1, $y_1 > y_2$ (打ち切り) で 0 を取るものと定義する。観察対象者数 n の実験を考え、イベント又は打ち切り発生までの時間を $\mathbf{t} \equiv (t_1, t_2, \dots, t_n)^\top$ 、関心のあるイベントが期間 $[k^{(j-1)}, k^{(j)})$ に発生した対象の集合を

$D([k^{(j-1)}, k^{(j)}]) \equiv \{i \mid \delta_i = 1, t_i \in [k^{(j-1)}, k^{(j)}]; i = 1, 2, \dots, n\}$ で表す. このとき, 部分尤度関数は

$$L(\boldsymbol{\beta}, \mathbf{k}; \mathbf{t}) \equiv \prod_{j=1}^{m+1} \prod_{i \in D([k^{(j-1)}, k^{(j)}])} \frac{\exp(\boldsymbol{\beta}^{(j)\top} \mathbf{z}_i)}{\sum_{i' \in R(t_i)} \exp(\boldsymbol{\beta}^{(j)\top} \mathbf{z}_{i'})}$$

で与えられる. ここで, $R(t)$ は時間 t でのリスク集合 $\{i' \mid t < t_{i'}\}$ とする. この関数を一般に部分尤度関数と呼ぶことから本論文でも同様の表現を用いているが, モデル (2.1) の下で $\lambda_0(\cdot)$ はイベント発生時点で不連続となる階段関数であると仮定すると, これは profile 尤度関数と同値である. Profile 尤度関数とは, 通常用最尤法で定義されるような $\lambda_0(\cdot)$ と $\boldsymbol{\beta}$ が含まれる尤度関数において, $\lambda_0(\cdot)$ を局外パラメータとみなして $\lambda_0(\cdot)$ に関して最大化することで得られる関数を指す. また, 対数部分尤度関数は

$$l(\boldsymbol{\beta}, \mathbf{k}; \mathbf{t}) \equiv \sum_{j=1}^{m+1} \sum_{i \in D([k^{(j-1)}, k^{(j)}])} \left[\boldsymbol{\beta}^{(j)\top} \mathbf{z}_i - \log \left\{ \sum_{i' \in R(t_i)} \exp(\boldsymbol{\beta}^{(j)\top} \mathbf{z}_{i'}) \right\} \right]$$

で表現できる. Cox (1972) の提案ではこの対数部分尤度関数を最大化することでパラメータを推定するが, 共変量間に多重共線性が存在する場合にも適用できるよう, ここでは一般化させ, モデル (2.1) に対してリッジタイプの正則化項 (Hoerl and Kennard 1970) を加えた正則化対数部分尤度関数の最大化によって回帰パラメータ及び変化点パラメータを推定する. 具体的には, 正則化対数部分尤度関数を

$$l_\xi(\boldsymbol{\beta}, \mathbf{k}; \mathbf{t}) \equiv \sum_{j=1}^{m+1} \sum_{i \in D([k^{(j-1)}, k^{(j)}])} \left[\boldsymbol{\beta}^{(j)\top} \mathbf{z}_i - \log \left\{ \sum_{i' \in R(t_i)} \exp(\boldsymbol{\beta}^{(j)\top} \mathbf{z}_{i'}) \right\} - \frac{\xi}{2} \boldsymbol{\beta}^{(j)\top} \boldsymbol{\beta}^{(j)} \right] \quad (2.3)$$

で定義し, これの最大化によりパラメータを推定する. つまり, 損失関数として $-l_\xi(\boldsymbol{\beta}, \mathbf{k}; \mathbf{t})$ を考える. ここで, ξ は正則化パラメータである. 以降, 記法の簡単さのため $D^{(j)} \equiv D([k^{(j-1)}, k^{(j)}])$ とする. 変化点が \mathbf{k} のときに正則化対数部分尤度関数を最大にする $\boldsymbol{\beta}$ を $\hat{\boldsymbol{\beta}}_{\mathbf{k}}$, つまり $\hat{\boldsymbol{\beta}}_{\mathbf{k}} \equiv \operatorname{argsup}_{\boldsymbol{\beta}} l_\xi(\boldsymbol{\beta}, \mathbf{k}; \mathbf{t})$ と書くことにする. また, $\hat{\mathbf{k}} \equiv \operatorname{argsup}_{\mathbf{k}} l_\xi(\hat{\boldsymbol{\beta}}_{\mathbf{k}}, \mathbf{k}; \mathbf{t})$ を変化点パラメータの推定量とする. このとき, 回帰パラメータの推定量は $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\hat{\mathbf{k}}}$ で与えられる.

次に, $\mathbf{h}(t_i, \boldsymbol{\beta}^{(j)})$ を p 次元ベクトル $\sum_{i' \in R(t_i)} \mathbf{z}_{i'} \exp(\boldsymbol{\beta}^{(j)\top} \mathbf{z}_{i'}) / \sum_{i' \in R(t_i)} \exp(\boldsymbol{\beta}^{(j)\top} \mathbf{z}_{i'})$, $\mathbf{H}(t_i, \boldsymbol{\beta}^{(j)})$ を p 次元正方行列 $\sum_{i' \in R(t_i)} \mathbf{z}_{i'} \mathbf{z}_{i'}^\top \exp(\boldsymbol{\beta}^{(j)\top} \mathbf{z}_{i'}) / \sum_{i' \in R(t_i)} \exp(\boldsymbol{\beta}^{(j)\top} \mathbf{z}_{i'})$ で定義する. このとき, 各 $j \in \{1, 2, \dots, m+1\}$ に対して, 正則化対数部分尤度関数の $\boldsymbol{\beta}^{(j)}$ に関する一次導関数と二次導関

数は

$$\frac{\partial}{\partial \boldsymbol{\beta}^{(j)}} l_{\xi}(\boldsymbol{\beta}, \mathbf{k}; \mathbf{t}) = \sum_{i \in D^{(j)}} \{z_i - \mathbf{h}(t_i, \boldsymbol{\beta}^{(j)}) - \xi \boldsymbol{\beta}^{(j)}\}$$

と

$$\frac{\partial^2}{\partial \boldsymbol{\beta}^{(j)} \partial \boldsymbol{\beta}^{(j)T}} l_{\xi}(\boldsymbol{\beta}, \mathbf{k}; \mathbf{t}) = - \sum_{i \in D^{(j)}} \{\mathbf{H}(t_i, \boldsymbol{\beta}^{(j)}) - \mathbf{h}(t_i, \boldsymbol{\beta}^{(j)}) \mathbf{h}(t_i, \boldsymbol{\beta}^{(j)})^T + \xi \mathbf{I}_p\}$$

と表現できる。ここで、 \mathbf{I}_p は p 次元単位行列である。

2.2 部分尤度法に基づく推定量の漸近的性質

変化点がないとき、Cox 比例ハザードモデルにおける回帰パラメータ $\boldsymbol{\beta}$ の部分尤度法に基づく推定量の一致性及び漸近正規性は、Tsiatis (1981) が部分尤度のスコア関数を独立な確率変数の和で表現することで示している。本節では、(2.3) に基づく正則化対数部分尤度関数の最大化による推定量 $\widehat{\boldsymbol{\beta}}_{\mathbf{k}}$ の漸近挙動を考える。イベント又は打ち切り発生までの時間 t_1, t_2, \dots, t_n の取り得る範囲は $[0, T]$ であることから、 n を大きくする漸近論を考える際でも変化点 $k^{(1)}, k^{(2)}, \dots, k^{(m)}$ の値は有限である。まず、 p 次元正方行列

$$\mathbf{A}_{\xi}^{*(j)}(\boldsymbol{\beta}, \mathbf{k}) \equiv \mathbf{E} \left(\frac{1}{n} \left[\sum_{i \in D^{(j)}} \{z_i - \mathbf{h}(t_i, \boldsymbol{\beta}^{(j)}) - \xi \boldsymbol{\beta}^{(j)}\} \right] \left[\sum_{i \in D^{(j)}} \{z_i - \mathbf{h}(t_i, \boldsymbol{\beta}^{(j)}) - \xi \boldsymbol{\beta}^{(j)}\} \right]^T \right)$$

を定義し、簡単のために $\mathbf{A}_{\xi}^{*(j)} \equiv \mathbf{A}_{\xi}^{*(j)}(\boldsymbol{\beta}_{\xi}^*, \mathbf{k}^*)$ と書くことにする。ここで、 $\boldsymbol{\beta}_{\xi}^* \equiv \operatorname{argsup}_{\boldsymbol{\beta}} \mathbf{E}\{l_{\xi}(\boldsymbol{\beta}, \mathbf{k}^*; \mathbf{t})\}$ である。このとき、 $\mathbf{E}\{(\partial/\partial \boldsymbol{\beta}^{(j)}) l_{\xi}(\boldsymbol{\beta}_{\xi}^*, \mathbf{k}^*; \mathbf{t})\} = \mathbf{E}\{(\partial/\partial \boldsymbol{\beta}^{(j)}) l(\boldsymbol{\beta}_{\xi}^*, \mathbf{k}^*; \mathbf{t})\} - \xi \mathbf{D}^{*(j)} \boldsymbol{\beta}_{\xi}^{*(j)} = \mathbf{0}_p$ となることに注意する。ここで、 $\mathbf{D}^{*(j)}$ は $\mathbf{E}\{|D([k^{*(j-1)}, k^{*(j)}])|/n\}$ であり、 $\mathbf{0}_p$ は p 次元零ベクトルである。

各要素が有限値であるベクトル $\mathbf{s} \equiv (s^{(1)}, s^{(2)}, \dots, s^{(m)})^T$ に対して、各 $j \in \{1, 2, \dots, m\}$ において $k^{(j)} = k^{*(j)} + s^{(j)}/n$ とする。このとき、Tsiatis (1981) と同様にして、 $\widehat{\boldsymbol{\beta}}_{\mathbf{k}}$ の一致性及び漸近正規性は自明に示すことができ、

$$\|\widehat{\boldsymbol{\beta}}_{\mathbf{k}}^{(j)} - \boldsymbol{\beta}_{\xi}^{*(j)}\| = o_{\mathbf{P}}(1)$$

と

$$\sqrt{n}(\widehat{\beta}_k^{(j)} - \beta_\xi^{*(j)}) \xrightarrow{d} N\{\mathbf{0}_p, (\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p)^{-1} \mathbf{A}_\xi^{*(j)} (\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p)^{-1}\} \quad (2.4)$$

を得る. ここで, $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ は平均ベクトル $\boldsymbol{\mu}$, 分散共分散行列 $\boldsymbol{\Sigma}$ の多変量正規分布である.

2.3 AIC タイプの情報量規準

本節では [Burnham and Anderson \(2003\)](#) や [小西・北川 \(2004\)](#) を基に, AIC タイプの情報量規準元来の導出方法について紹介する. AIC タイプの情報量規準の開発にあっては, 観察されたデータ \mathbf{t} は何らかの真のモデルに従って生成されたものと考え. そして, パラメータ $\boldsymbol{\theta}$ をもつパラメトリックモデルを想定し, そのデータ \mathbf{t} に基づいて最尤法によりパラメータを推定することでモデルを構築する. ここで, 真のモデルとは未知の確率分布関数 $G(\mathbf{t})$, その密度関数 $g(\mathbf{t})$ をもつモデルとする. また, 構築したモデルの確率密度関数は, 最尤推定量 $\hat{\boldsymbol{\theta}}$ を用いて $f(\mathbf{t} | \hat{\boldsymbol{\theta}})$ と書くこととする. 構築したモデルの評価を, 予測の観点から考える. つまり, 将来真のモデルから得られるデータの分布 $g(\mathbf{t})$ を, 構築したモデル $f(\mathbf{t} | \hat{\boldsymbol{\theta}})$ で予測した時の平均的な良さを評価したい.

[Akaike \(1973\)](#) は, 構築したモデルの良さを真のモデルとの確率分布としての近さによって評価するものとし, その近さを測る尺度として Kullback-Leibler (K-L) ダイバージェンスを用いた. K-L ダイバージェンスは, E_G を G に関する期待値とすると

$$E_G\{\log g(\mathbf{t})\} - E_G\{\log f(\mathbf{t} | \boldsymbol{\theta})\}$$

で与えられ, この値が小さいほど構築したモデルが真のモデルに近いと考えることができる. つまり, 平均対数尤度 $E_G\{\log f(\mathbf{t} | \boldsymbol{\theta})\}$ が大きいモデルほど真のモデルに近いと言え, AIC の構成はこの平均対数尤度の有効な推定量を求めることが本質となる. この平均対数尤度の推定量の一つとして最大対数尤度が考えられるが, 観察されたデータ \mathbf{t} をモデルの推定とそのモデルの良さの評価の両方に用いていることから, 最大対数尤度は平均対数尤度に対してバイアスが生じてしまう. [Akaike \(1973\)](#) は, そのバイアスの大きさがパラメータの次元に依存し, 具体的にはその次元数になることを示した. そして, 最大対数尤度をそのバイアスで補正したものが AIC として提案され, それを最小にするモデルが予測の観点で良いモデルとして選択されることになる.

2.4 部分尤度に基づく AIC タイプの情報量規準

Xu et al. (2009) は, 2.3 節で示した通常の AIC タイプの情報量規準を考えるとときと同様に, 真のモデルと推定されたモデル間の K-L ダイバージェンスに基づくリスク関数を想定し, Cox 比例ハザードモデルにおいて部分尤度を用いたときの AIC を導出した. そして, 回帰パラメータに対する漸近バイアスは 1 となることを示している. 本論文では, それと同様の流れで, モデル (2.1) に対する AIC タイプの情報量規準を導出する.

生存時間データ $\mathbf{t} \equiv (t_1, t_2, \dots, t_n)^\top$ が得られた時の (β, \mathbf{k}) の推定量を $(\hat{\beta}_t, \hat{\mathbf{k}}_t) \equiv \operatorname{argsup}_{\beta, \mathbf{k}} l_\xi(\beta, \mathbf{k}; \mathbf{t})$ と書くことにする. さらに, $\mathbf{u} = (u_1, u_2, \dots, u_n)^\top$ を \mathbf{t} のコピー, つまり \mathbf{t} とは独立に同じモデルに従うデータであるとし, 推定の際に用いた損失に基づくダイバージェンスとして $-2E_{\mathbf{u}}\{l_\xi(\hat{\beta}_t, \hat{\mathbf{k}}_t; \mathbf{u})\}$ を想定する. ここで, $E_{\mathbf{u}}$ は \mathbf{u} に関する期待値を意味する. そして, これの初期推定量として観測データ \mathbf{t} に対する -2 倍の最大正則化対数部分尤度, つまり $-2l_\xi(\hat{\beta}_t, \hat{\mathbf{k}}_t; \mathbf{t})$ を考え, これを

$$E_{\mathbf{t}}[2l_\xi(\hat{\beta}_t, \hat{\mathbf{k}}_t; \mathbf{t}) - E_{\mathbf{u}}\{2l_\xi(\hat{\beta}_t, \hat{\mathbf{k}}_t; \mathbf{u})\}] = 2E\{l_\xi(\hat{\beta}_t, \hat{\mathbf{k}}_t; \mathbf{t}) - l_\xi(\hat{\beta}_u, \hat{\mathbf{k}}_u; \mathbf{t})\}$$

でバイアス補正する. ただ, この期待値は陽に求めることはできないので, 通常の AIC タイプの情報量規準を考えるとときと同様に, バイアスを漸近評価する. まず, $ll_\xi(\beta, \mathbf{k}; \mathbf{t}) \equiv l_\xi(\beta, \mathbf{k}; \mathbf{t}) - l_\xi(\beta_\xi^*, \mathbf{k}_\xi^*; \mathbf{t})$ と $\hat{\beta}_{\mathbf{k}, \mathbf{u}} \equiv \operatorname{argsup}_{\beta} ll_\xi(\beta, \mathbf{k}; \mathbf{u})$, $\hat{ll}_\xi(\mathbf{k}; \mathbf{t}, \mathbf{u}) \equiv ll_\xi(\hat{\beta}_{\mathbf{k}, \mathbf{u}}, \mathbf{k}; \mathbf{t})$ という記法を用意し, 上記のバイアスを

$$\begin{aligned} & 2E\{ll_\xi(\hat{\beta}_t, \hat{\mathbf{k}}_t; \mathbf{t}) - ll_\xi(\hat{\beta}_u, \hat{\mathbf{k}}_u; \mathbf{t})\} \\ &= 2E\left[\sup_{\mathbf{k}} ll_\xi(\hat{\beta}_{\mathbf{k}, \mathbf{t}}, \mathbf{k}; \mathbf{t}) - ll_\xi\left\{\hat{\beta}_{\operatorname{argsup}_{\mathbf{k}} ll_\xi(\hat{\beta}_{\mathbf{k}, \mathbf{u}}, \mathbf{k}; \mathbf{u})}, \operatorname{argsup}_{\mathbf{k}} ll_\xi(\hat{\beta}_{\mathbf{k}, \mathbf{u}}, \mathbf{k}; \mathbf{u}); \mathbf{t}\right\}\right] \\ &= 2E\left[\sup_{\mathbf{k}} \hat{ll}_\xi(\mathbf{k}; \mathbf{t}, \mathbf{t}) - \hat{ll}_\xi\left\{\operatorname{argsup}_{\mathbf{k}} \hat{ll}_\xi(\mathbf{k}; \mathbf{u}, \mathbf{u}); \mathbf{t}, \mathbf{u}\right\}\right] \end{aligned}$$

と表現する. そして, $b_\xi(\mathbf{k}^*, \beta_\xi^*)$ を $\sup_{\mathbf{k} \in K} \hat{ll}_\xi(\mathbf{k}; \mathbf{t}, \mathbf{t}) - \hat{ll}_\xi\{\operatorname{argsup}_{\mathbf{k} \in K} \hat{ll}_\xi(\mathbf{k}; \mathbf{u}, \mathbf{u}); \mathbf{t}, \mathbf{u}\}$ の弱極限と定義し, $2E\{b_\xi(\mathbf{k}^*, \beta_\xi^*)\}$ を漸近バイアスとみなす. ここで, K は $\hat{ll}_\xi(\mathbf{k}; \mathbf{t}, \mathbf{t})$ が $O_P(1)$ もしくは正となる \mathbf{k} の集合, つまりある正の M が存在して $P\{\hat{ll}_\xi(\mathbf{k}; \mathbf{t}, \mathbf{t}) > -M\}$ が 0 に収束しない \mathbf{k} の集合とする. このとき,

$$-2l_\xi(\hat{\beta}_t, \hat{\mathbf{k}}_t; \mathbf{t}) + 2E\{b_\xi(\mathbf{k}^*, \beta_\xi^*)\} \tag{2.5}$$

が、正則化部分尤度法を用いたときの変化点をもつ Cox 比例ハザードモデルに対する AIC タイプの情報量規準とみなせる。もし変化点がなく、正則化パラメータ ξ が 0 のとき、これは [Xu et al. \(2009\)](#) が与えた AIC と同じであり、 $\xi = 0$ のときの $E\{b_\xi(\mathbf{k}^*, \boldsymbol{\beta}^*)\} = E\{b_\xi(\boldsymbol{\beta}^*)\}$ は $\boldsymbol{\beta}$ のパラメータ数となる。

第3章 主結果

本節では、2.1節の設定の下で、2.2節の漸近的性質を利用し、2.4節で紹介したAIC元来の導出方法によって漸近バイアスを再評価して情報量規準を開発する。

3.1 漸近バイアスの導出

各 $j \in \{1, 2, \dots, m\}$ に対して、 $k^{(j)} = k^{*(j)} + s^{(j)}/n$ とおく。まず、ベクトル $\mathbf{s} = (s^{(1)}, s^{(2)}, \dots, s^{(m)})$ の各要素が有限値である場合を考える。両側に伸びる負のドリフト付きランダムウォークとして

$$Q_{\xi; k, t}^{(j)} \equiv I_{\{k < k^{*(j)}\}} \left(\sum_{i \in D([k, k^{*(j)}])} \left[(\boldsymbol{\beta}_{\xi}^{*(j+1)} - \boldsymbol{\beta}_{\xi}^{*(j)})^{\top} \mathbf{z}_i - \log \left\{ \frac{\sum_{i' \in R(t_i)} \exp(\boldsymbol{\beta}_{\xi}^{*(j+1)\top} \mathbf{z}_{i'})}{\sum_{i' \in R(t_i)} \exp(\boldsymbol{\beta}_{\xi}^{*(j)\top} \mathbf{z}_{i'})} \right\} - \frac{\xi}{2} (\boldsymbol{\beta}_{\xi}^{*(j+1)\top} \boldsymbol{\beta}_{\xi}^{*(j+1)} - \boldsymbol{\beta}_{\xi}^{*(j)\top} \boldsymbol{\beta}_{\xi}^{*(j)}) \right] \right) \\ + I_{\{k > k^{*(j)}\}} \left(\sum_{i \in D([k^{*(j)}, k])} \left[(\boldsymbol{\beta}_{\xi}^{*(j)} - \boldsymbol{\beta}_{\xi}^{*(j+1)})^{\top} \mathbf{z}_i - \log \left\{ \frac{\sum_{i' \in R(t_i)} \exp(\boldsymbol{\beta}_{\xi}^{*(j)\top} \mathbf{z}_{i'})}{\sum_{i' \in R(t_i)} \exp(\boldsymbol{\beta}_{\xi}^{*(j+1)\top} \mathbf{z}_{i'})} \right\} - \frac{\xi}{2} (\boldsymbol{\beta}_{\xi}^{*(j)\top} \boldsymbol{\beta}_{\xi}^{*(j)} - \boldsymbol{\beta}_{\xi}^{*(j+1)\top} \boldsymbol{\beta}_{\xi}^{*(j+1)}) \right] \right)$$

を定義すると

$$l_{\xi}(\widehat{\boldsymbol{\beta}}_{\mathbf{k}}, \mathbf{k}; \mathbf{t}) - l_{\xi}(\widehat{\boldsymbol{\beta}}_{\mathbf{k}^*}, \mathbf{k}^*; \mathbf{t}) = \sum_{j=1}^m Q_{\xi; k^{*(j)} + s^{(j)}/n, \mathbf{t}}^{(j)} + o_{\mathbb{P}}(1) = O_{\mathbb{P}}(1) \quad (3.1)$$

が得られる。(3.1)の導出過程は付録Aを参照されたい。更に、正則化対数部分尤度関数を $\widehat{\boldsymbol{\beta}}_{\mathbf{k}^*} = \boldsymbol{\beta}_{\xi}^*$ の周りでテイラー展開し、(2.4)及びMurphy and van der Vaart (2000)を利用すると

$$l_{\xi}(\widehat{\boldsymbol{\beta}}_{\mathbf{k}^*}, \mathbf{k}^*; \mathbf{t}) - l_{\xi}(\boldsymbol{\beta}_{\xi}^*, \mathbf{k}^*; \mathbf{t}) = \frac{1}{2} \sum_{j=1}^{m+1} \boldsymbol{\nu}_{\xi}^{(j)\top} \boldsymbol{\nu}_{\xi}^{*(j)} + o_{\mathbb{P}}(1) = O_{\mathbb{P}}(1) \quad (3.2)$$

が得られる。ここで、 $\boldsymbol{\nu}_\xi^{*(j)}$ は多変量正規分布 $N\{\mathbf{0}_p, \mathbf{A}_\xi^{*(j)-1}(\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p)\}$ に従う確率変数ベクトルである。よって、(3.1) 及び (3.2) より、

$$\widehat{\ell}_\xi(\mathbf{k}; \mathbf{t}, \mathbf{t}) = O_P(1)$$

が得られる。一方、 \mathbf{s} がいずれかの要素で有限でない値をもつベクトルの場合は、任意の $M > 0$ に対して

$$P\{\widehat{\ell}_\xi(\mathbf{k}; \mathbf{t}, \mathbf{t}) > -M\} \rightarrow 0 \quad (3.3)$$

となることを示せる。(3.3) の導出過程は付録 B を参照されたい。従って、 \mathbf{s} の各要素が有限値のベクトルのときは $\widehat{\ell}_\xi(\mathbf{k}; \mathbf{t}, \mathbf{t}) = O_P(1)$ となり

$$\|\widehat{\mathbf{k}} - \mathbf{k}^*\| = O_P(1/n) \quad (3.4)$$

を得る。これは Pons (2002) の結果とも整合する。

以上より、 $K^{(j)} = \{k \mid k - k^{*(j)} = O(1/n)\}$ として、 $K = \{k \mid k^{(j)} \in K^{(j)}, j \in \{1, 2, \dots, m\}\}$ であることが分かる。そして、(3.1) 及び (3.2) より

$$\sup_{\mathbf{k} \in K} \widehat{\ell}_\xi(\mathbf{k}; \mathbf{t}, \mathbf{t}) = \sum_{j=1}^m \sup_{k \in K^{(j)}} Q_{\xi; k, \mathbf{t}}^{(j)} + \frac{1}{2} \sum_{j=1}^{m+1} \boldsymbol{\nu}_\xi^{(j)\top} \boldsymbol{\nu}_\xi^{*(j)} + o_P(1) \quad (3.5)$$

と

$$\left\| \operatorname{argsup}_{\mathbf{k} \in K} \widehat{\ell}_\xi(\mathbf{k}; \mathbf{u}, \mathbf{u}) - \left(\operatorname{argsup}_{k \in K^{(1)}} Q_{\xi; k, \mathbf{u}}^{(1)}, \operatorname{argsup}_{k \in K^{(2)}} Q_{\xi; k, \mathbf{u}}^{(2)}, \dots, \operatorname{argsup}_{k \in K^{(m)}} Q_{\xi; k, \mathbf{u}}^{(m)} \right)^\top \right\| = o_P(1) \quad (3.6)$$

が得られる。さらに、 $\widetilde{k}_u^{(j)} \equiv \operatorname{argsup}_{k \in K^{(j)}} Q_{\xi; k, \mathbf{u}}^{(j)}$ と定義して $\widetilde{\mathbf{k}}_u \equiv (\widetilde{k}_u^{(1)}, \widetilde{k}_u^{(2)}, \dots, \widetilde{k}_u^{(m)})^\top$ と書くと、

$$\|\widehat{\boldsymbol{\beta}}_{\widetilde{\mathbf{k}}_u, \mathbf{u}} - \widehat{\boldsymbol{\beta}}_{\mathbf{k}^*, \mathbf{u}}\| = O_P(1/n) \quad (3.7)$$

と

$$\|\widehat{\boldsymbol{\beta}}_{\widetilde{\mathbf{k}}_u, \mathbf{u}} - \boldsymbol{\beta}_\xi^*\| = O_P(1/\sqrt{n}) \quad (3.8)$$

が得られる. $\widehat{\ell}_\xi(\widetilde{\mathbf{k}}_{\mathbf{u}}; \mathbf{t}, \mathbf{u})$ に対し, (3.7), (3.8), 及び Murphy and van der Vaart (2000) を利用すると

$$\begin{aligned} & \widehat{\ell}_\xi \left(\underset{\mathbf{k} \in K}{\operatorname{argsup}} \widehat{\ell}_\xi(\mathbf{k}; \mathbf{u}, \mathbf{u}); \mathbf{t}, \mathbf{u} \right) \\ &= \widehat{\ell}_\xi(\widetilde{\mathbf{k}}_{\mathbf{u}}; \mathbf{t}, \mathbf{u}) + o_{\mathbb{P}}(1) \\ &= - \sum_{j=1}^m Q_{\xi; \widetilde{\mathbf{k}}_{\mathbf{u}}^{(j)}, \mathbf{t}}^{(j)} + \sum_{j=1}^{m+1} \left((\widehat{\boldsymbol{\beta}}_{\mathbf{k}^*, \mathbf{u}}^{(j)} - \boldsymbol{\beta}_\xi^{*(j)})^\top \left[\sum_{i \in \widetilde{D}^{*(j)}} \{ \mathbf{z}_i - \mathbf{h}(t_i, \boldsymbol{\beta}_\xi^{*(j)}) - \xi \boldsymbol{\beta}_\xi^{*(j)} \} \right] - \frac{1}{2} \boldsymbol{\nu}_\xi^{(j)\top} \boldsymbol{\nu}_\xi^{*(j)} \right) \\ &+ o_{\mathbb{P}}(1) \end{aligned} \quad (3.9)$$

を得る. (3.9) の導出過程は付録 C を参照されたい. (3.5) 及び (3.9) より, 以下の定理が得られる.

定理 1. 条件 (2.2) の下で, (2.5) の漸近バイアスは

$$\sum_{j=1}^m \mathbb{E} \left(\sup_{\mathbf{k} \in K^{(j)}} Q_{\xi; \mathbf{k}, \mathbf{t}}^{(j)} + Q_{\xi; \operatorname{argsup}_{\mathbf{k} \in K^{(j)}} \mathbf{k}, \mathbf{t}}^{(j)} \right) + \sum_{j=1}^{m+1} \operatorname{tr} \{ \mathbf{A}_\xi^{*(j)} (\mathbf{A}_\xi^{*(j)} + \xi \mathbf{D}^{*(j)} \mathbf{I}_p)^{-1} \} \quad (3.10)$$

で与えられる.

(3.10) の第一項が変化点パラメータ \mathbf{k} に対するバイアス, 第二項は回帰パラメータ $\boldsymbol{\beta}$ に対するバイアスとみなすことができる.

3.2 漸近バイアスの解析評価

通常モデルの AIC では, パラメータの真値が定数であろうと, それらが何らかの値に収束するものであろうと, その罰則は 2 であることに違いはない. つまり, どちらの設定を考えるか, ということは重要な問題ではない. 一方, 変化点モデルの AIC では, どちらの設定を考えるかで, 罰則が異なるものになる. 具体的には, $\boldsymbol{\beta}^{*(j+1)} - \boldsymbol{\beta}^{*(j)}$ を定数ベクトルにするか零ベクトル $\mathbf{0}_p$ に近づけるかで, (3.10) の第一項の評価が異なってくる.

ここでは $\mathbf{0}_p$ に近づける方が重要かつ自然と考える. 漸近論においても $\boldsymbol{\beta}^{*(j+1)} - \boldsymbol{\beta}^{*(j)}$ を定数と仮定するということは, 明確な変化が存在している設定を想定していることになる. こういった設定では, 情報量規準の第一項, つまり当てはまり項でほぼモデル選択の結果は決まり, 第二項のバイアス評価はそれほど重要にならない. 第二項の精度の良い評価が必要となるのは, 変化の有

無を一目では判断できないようなケースであり、そのケースを反映するのは $\beta^{*(j+1)} - \beta^{*(j)} \rightarrow \mathbf{0}_p$ の方の仮定である。たとえ一見して変化の存在が疑われる場合であっても、データが有限である限りその存在は絶対ではなく、 $\beta^{*(j+1)} - \beta^{*(j)}$ は定数であるという仮定は、変化有りに偏りすぎた設定での漸近近似を生んでしまう。つまり、 $\beta^{*(j+1)} - \beta^{*(j)}$ が $\mathbf{0}_p$ に近づく仮定の方が自然であると考えられる。

以上より、Csörgő and Horváth (1997) の Section 1.5 と同様にして、正規化対数部分尤度関数の最大化によるパラメータの推定を考える際に

$$\beta_{\xi}^{*(j+1)} - \beta_{\xi}^{*(j)} = \Delta^{(j)} / \sqrt{\alpha_n} \quad (j \in \{1, 2, \dots, m\}), \quad O(1) \neq \alpha_n = o(n) \quad (3.11)$$

という条件を仮定する。ここで、 $\Delta^{(j)}$ は定ベクトルである。条件 (3.11) の下、変化点推定量の漸近挙動を調べる。 $\mathbf{k} = \mathbf{k}^* + \alpha_n \mathbf{s} / n$ としたとき、(3.4) の導出と同様にして、(3.4) の代わりに $\widehat{\mathbf{k}} - \mathbf{k}^* = O_P(\alpha_n / n)$ が得られ、 $K^{(j)} = \{\mathbf{k} \mid \mathbf{k} - \mathbf{k}^{*(j)} = O(\alpha_n / n)\}$ として $K = \{\mathbf{k} \mid \mathbf{k}^{(j)} \in K^{(j)}, j \in \{1, 2, \dots, m\}\}$ であることが分かる。詳細は付録 D を参照されたい。

以降、 \mathbf{s} は各要素が有限値のベクトルとして、 $\mathbf{k} = \mathbf{k}^* + \alpha_n \mathbf{s} / n$ とする。また、 $\{W_s\}_{s \in \mathbb{R}}$ を $E(W_s) = 0$ かつ $V(W_s) = |s|$ である両側標準ブラウン運動とし、さらに $V_s(\tau_1, \tau_2, \sigma_1, \sigma_2)$ を、 τ_1 や τ_2 をドリフト係数、 σ_1 や σ_2 を拡散係数とした、両側に伸びるブラウン運動とする。つまり、 $V_s(\tau_1, \tau_2, \sigma_1, \sigma_2)$ を $s < 0$ のとき $-\tau_1 |s| + \sigma_1 W_s$ 、 $s \geq 0$ のとき $-\tau_2 |s| + \sigma_2 W_s$ と定義し、さらに

$$V_{\xi; s}^{*(j)} \equiv V_s \left\{ \frac{1}{2} \Delta^{(j)\top} (\mathbf{A}_{\xi}^{*(j)} + \xi D^{*(j)} \mathbf{I}_p) \Delta^{(j)}, \frac{1}{2} \Delta^{(j)\top} (\mathbf{A}_{\xi}^{*(j+1)} + \xi D^{*(j+1)} \mathbf{I}_p) \Delta^{(j)}, \right. \\ \left. (\Delta^{(j)\top} \mathbf{A}_{\xi}^{*(j)} \Delta^{(j)})^{1/2}, (\Delta^{(j)\top} \mathbf{A}_{\xi}^{*(j+1)} \Delta^{(j)})^{1/2} \right\}$$

とする。このとき、

$$Q_{\xi; \mathbf{k}^* + \alpha_n \mathbf{s} / n, t}^{(j)} \xrightarrow{d} V_{\xi; s}^{*(j)}$$

が成立する。 $\sup_{s \in \mathbb{R}} V_{\xi; s}^{*(j)}$ や $V_{\xi; \arg \sup_{s \in \mathbb{R}} V_{\xi; s}^{*(j)}}^{*(j)}$ の期待値を評価するために、Bhattacharya and Brockwell (1976) や Shepp (1979) の結果を用いていく。そして、これらの帰結として次の定理が得られる。

定理 2. 定理 1 と (3.11) の条件の下で, (2.5) の漸近バイアスは

$$2 \sum_{j=1}^m \mathbf{C}^{(j)}(\mathbf{A}_\xi^{*(\cdot)}, \mathbf{A}_\xi^{*(\cdot)} + \xi D^{*(\cdot)} \mathbf{I}_p) + \sum_{j=1}^{m+1} \text{tr}\{\mathbf{A}_\xi^{*(j)}(\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p)^{-1}\} \quad (3.12)$$

で与えられる.

定理 2 の詳細な証明及び $\mathbf{C}^{(j)}(\mathbf{A}^{(\cdot)\dagger}, \mathbf{A}^{(\cdot)\ddagger})$ の定義は付録 E を参照されたい.

これより最大正則化対数部分尤度をバイアス補正したものとしての情報量規準が得られるが, (3.12) の漸近バイアスには未知パラメータが含まれているため, TIC や一般化情報量規準 (GIC; Generalized Information Criterion, Konishi and Kitagawa 1996) と同様に, 一致推定量で置き換える. 具体的には, $D^{*(j)}$ の一致推定量として $\widehat{D}^{*(j)} \equiv |\widehat{D}^{(j)}|/n$ を, $\mathbf{A}_\xi^{*(j)}$ の一致推定量として

$$\widehat{\mathbf{A}}_\xi^{*(j)}(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{k}}) \equiv \frac{1}{n} \left[\sum_{i \in \widehat{D}^{(j)}} \{\mathbf{z}_i - \mathbf{h}(t_i, \widehat{\boldsymbol{\beta}}^{(j)}) - \xi \widehat{\boldsymbol{\beta}}^{(j)}\} \right] \left[\sum_{i \in \widehat{D}^{(j)}} \{\mathbf{z}_i - \mathbf{h}(t_i, \widehat{\boldsymbol{\beta}}^{(j)}) - \xi \widehat{\boldsymbol{\beta}}^{(j)}\} \right]^\top$$

を用いる. ここで, $\widehat{D}^{(j)} \equiv D(\widehat{\mathbf{k}}^{(j-1)}, \widehat{\mathbf{k}}^{(j)})$ である. 結果, L_2 ノルムの正則化項を対数部分尤度に加えて推定した場合の変化点をもつ Cox 比例ハザードモデルに対し,

$$\begin{aligned} \text{AIC}_\xi = -2l_\xi(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{k}}; \mathbf{t}) + 4 \sum_{j=1}^m \widehat{\mathbf{C}}^{(j)}\{\widehat{\mathbf{A}}_\xi^{*(\cdot)}(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{k}}), \widehat{\mathbf{A}}_\xi^{*(\cdot)}(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{k}}) + \xi \widehat{D}^{*(\cdot)} \mathbf{I}_p\} \\ + 2 \sum_{j=1}^{m+1} \text{tr}[\widehat{\mathbf{A}}_\xi^{*(j)}(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{k}})\{\widehat{\mathbf{A}}_\xi^{*(j)}(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{k}}) + \xi \widehat{D}^{*(j)} \mathbf{I}_p\}^{-1}] \quad (3.13) \end{aligned}$$

という情報量規準を提案する. ここで, $\widehat{\mathbf{C}}^{(j)}(\mathbf{A}^{(\cdot)\dagger}, \mathbf{A}^{(\cdot)\ddagger})$ は $\mathbf{C}^{(j)}(\mathbf{A}^{(\cdot)\dagger}, \mathbf{A}^{(\cdot)\ddagger})$ の $\boldsymbol{\beta}_\xi^*$ を $\widehat{\boldsymbol{\beta}}$ で置き換えたものである. ξ の最適値としては, この AIC_ξ を最小にする ξ を用いればよい.

正則化パラメータ ξ が 0 のときは $\boldsymbol{\beta}_\xi^*$ が $\boldsymbol{\beta}^*$ と等しくなり, $\mathbf{C}^{(j)}(\mathbf{A}_\xi^{*(\cdot)}, \mathbf{A}_\xi^{*(\cdot)}) = 3/2$ であることから, 次の系を得る.

系 1. 定理 1 と (3.11) の条件の下で, (2.3) で $\xi = 0$ とする通常の部分尤度法における (2.5) の漸近バイアスは

$$\mathbf{E}\{b(\mathbf{k}^*, \boldsymbol{\beta}^*)\} = 3m + p(m + 1)$$

で与えられる.

これより、変化点パラメータによる漸近バイアスは回帰パラメータによる漸近バイアスの3倍であることが分かる。これは [Ninomiya \(2015\)](#) の結果と整合する。結果、通常の部分尤度法で推定した場合の変化点をもつ Cox 比例ハザードモデルに対し、

$$\text{AIC} = -2l(\hat{\boldsymbol{\beta}}, \hat{\mathbf{k}}; \mathbf{t}) + 6m + 2p(m+1) \quad (3.14)$$

という情報量規準を提案する。

ここまで、モデル (2.1) はすべての回帰パラメータが変化点で構造変化するモデルとして議論してきたが、一部のパラメータのみ構造変化するようなモデルに対しても、同じ条件の下で同様の情報量規準が得られる。例えば、構造変化をもつ回帰パラメータと構造変化のないパラメータをそれぞれ、 $p(m+1)$ 次元ベクトルの $\boldsymbol{\beta}_1 \equiv (\boldsymbol{\beta}_1^{(1)\top}, \boldsymbol{\beta}_1^{(2)\top}, \dots, \boldsymbol{\beta}_1^{(m+1)\top})$ と q 次元ベクトルの $\boldsymbol{\beta}_2$ とし、

$$\lambda(t | \mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta}_1^{(j)\top} \mathbf{z}_1 + \boldsymbol{\beta}_2^\top \mathbf{z}_2), \quad t \in [k^{(j-1)}, k^{(j)}], \quad j \in \{1, 2, \dots, m+1\}$$

を考える。このとき、変化点をもつ Cox 比例ハザードモデルに対する AIC は

$$\text{AIC} = -2l(\hat{\boldsymbol{\beta}}, \hat{\mathbf{k}}; \mathbf{t}) + 6m + 2\{p(m+1) + q\}$$

で与えられる。

第4章 数値実験

4.1 シミュレーション実験

4.1.1 形式的な情報量規準や既存手法との比較

本節では、まず正則化項を持たない通常の部分尤度法に基づき推定した場合の情報量規準として、(3.14)のAICの性能を数値実験によって検証する。比較対象としては、変化点パラメータによるバイアスも回帰パラメータによるバイアスと同様に扱った形式的なAICとして

$$\text{AIC}_{\text{naive}} = -2l(\hat{\beta}, \hat{k}; t) + 2m + 2p(m+1)$$

を考える。最も簡単な設定として、1個の変化点 k をもつ単変量Cox比例ハザードモデル

$$\lambda(t|z) = \begin{cases} \lambda_0(t) \exp(\beta^{(1)}z), & t \in [0, k) \\ \lambda_0(t) \exp(\beta^{(2)}z), & t \in [k, T) \end{cases} \quad (4.1)$$

を考える。ここで、 z は成功確率1/2のベルヌーイ分布に従う共変量である。この数値実験モデルは、変化点パラメータ1つと回帰パラメータ2つをもつため、AICの漸近バイアスは $3 \times 1 + 1 \times 2 = 5$ となり、 $\text{AIC}_{\text{naive}}$ では $1 \times 1 + 1 \times 2 = 3$ となる。

これらの罰則項が最大対数部分尤度に対するバイアスの適切な近似を与えているかを調べるために、数値実験モデル(4.1)においてパラメータの真値や打ち切り確率、イベントサイズを変えながらバイアスを数値的に評価した。その結果を表4.1に示した。ここで、打ち切り確率は共変量及び生存時間と独立であると仮定する。以降も同様に扱う。いずれの設定においても、値は5付近にあり、少なくとも $\text{AIC}_{\text{naive}}$ における漸近バイアス評価3よりは、AICの方が正確な近似になっていることが確認できる。

次に、AICがどのようなモデル選択を行うかを明確にするために、最も単純な設定で $\text{AIC}_{\text{naive}}$

と比較を行う．具体的には， $m = 0$ ， $m = 1$ ， $m = 2$ 及び $m = 3$ として

$$\lambda(t | z) = \lambda_0(t) \exp(\beta^{(j)} z), \quad t \in [k^{(j-1)}, k^{(j)}], \quad j \in \{1, 2, \dots, m+1\} \quad (4.2)$$

のモデルを考える．ここで， $k^{(0)} = 0$ かつ $k^{(m+1)} = T$ である．このとき，各情報量規準を用いていずれかのモデルを選択する．表 4.2 では，真の変化点の数 m^* を 0 個又は 1 個とし，真の変化点と真の変化量を変えながら，真のモデルと選択されたモデル間の Kullback-Leibler ダイバージェンス (K-L) 及び変化点の数の選択確率を評価している．K-L は，以降の実験も含めて，真のモデルから生成された独立な 100 組の無作為標本 $\{\mathbf{u}^{[\omega]} \mid 1 \leq \omega \leq 100\}$ を用いて， $\sum_{\omega=1}^{100} \{l(\boldsymbol{\beta}^*, \mathbf{k}^*; \mathbf{u}^{[\omega]}) - l(\hat{\boldsymbol{\beta}}_t, \hat{\mathbf{k}}_t; \mathbf{u}^{[\omega]})\} / 100$ で評価した．変化点が存在しない，すなわち $m^* = 0$ の場合を見てみると，イベントサイズによらず AIC は変化点なしのモデル ($m = 0$ のモデル) を概ね 90% を超える高い確率で選択できていることが分かる．一方， $\text{AIC}_{\text{naive}}$ は変化点有りのモデル ($m > 0$ のモデル) を選択する確率が概ね 50% となっており，イベントサイズが大きいほどその傾向は顕著に出ていることが分かる．この結果は， $\text{AIC}_{\text{naive}}$ が漸近バイアス評価を過小評価し，過適合を引き起こしていることの現れであると推察できる．また，真の変化点の数が 1 個の場合，つまり $m^* = 1$ の場合では，イベントサイズが小さいほど，そして変化量が小さいほど，AIC が真の変化点の数を選ぶ確率は $\text{AIC}_{\text{naive}}$ より低くなっている．しかし，K-L を見るといずれの設定においても AIC は $\text{AIC}_{\text{naive}}$ よりも明らかに小さい K-L を与えており，予測の観点からは AIC がはっきりと良いモデル選択できていると言える．

さらに，より実戦的な設定における性能比較を行う．数値実験モデル (4.2) において $m = 0$ ， $m = 1$ ， $m = 2$ ， $m = 3$ 及び $m = 4$ のモデルを考え，真の変化点の数を 2 個，つまり $m^* = 2$ とし，真の変化点と真の変化量をランダムに与える．比較対象として， $\text{AIC}_{\text{naive}}$ だけでなく，

$$\text{BIC}_{\text{naive}} = -2l(\hat{\boldsymbol{\beta}}, \hat{\mathbf{k}}; \mathbf{t}) + \{m + p(m+1)\} \log(|D|)$$

と He et al. (2013) の逐次検定アプローチ (Sequential Test) を考える．ここで， $|D|$ は観察されたイベントサイズである．Table 4.3 では，真のモデルと選択されたモデル間の K-L 及び変化点の数の選択確率を評価している．AIC が $\text{AIC}_{\text{naive}}$ や $\text{BIC}_{\text{naive}}$ ，Sequential Test よりも小さい K-L を与えていることから，AIC は予測の観点から最も優れた規準であると分かる．特に，イベントサイズが大きいほど， $\text{AIC}_{\text{naive}}$ は多くの変化点をもつモデルを選択する傾向にあり， $\text{BIC}_{\text{naive}}$ と

Sequential Test は変化点の少ないモデルを選ぶ傾向にあった。これは、それらの手法で選択されたモデルの K-L が大きくなる理由の一つと言える。

最後に、(3.13) の AIC_ξ の性能を評価するために、 $m = 0$, $m = 1$, $m = 2$, $m = 3$ 及び $m = 4$ として

$$\lambda(t | z) = \lambda_0(t) \exp(\beta_1^{(j)} z_1 + \beta_2^{(j)} z_2), \quad t \in [k^{(j-1)}, k^{(j)}], \quad j \in \{1, 2, \dots, m+1\} \quad (4.3)$$

のモデルを考える。多重共線性に対するリッジタイプの正則化による利点を確認するために、成功確率 $1/2$ のベルヌーイ分布に従う共変量 z_1 及び z_2 の間に正の相関があるとする。真の変化点の数は2個と仮定し、真の変化点と真の変化量はランダムに与えるものとする。比較対象は、AIC だけでなく、 AIC_{naive} 及び BIC_{naive} も考える。表 4.4 では、K-L 及び変化点の数の選択確率を評価している。AIC が AIC_ξ よりも高い確率で $m = 2$ のモデルを選択する設定もいくつかあったが、これらの実験では AIC_ξ が AIC や AIC_{naive} , BIC_{naive} と比べて最も小さい K-L を与えていた。ゆえに、共変量間に多重共線性が存在する下では、 AIC_ξ が予測の観点では最も優れた情報量規準であると言える。

4.1.2 交差検証法との比較

本節では、(3.14) の AIC と交差検証法 (cross-validation; CV; Stone 1977) の性能評価を行う。数値実験モデルとしては、モデル (4.2) に対して $m = 0$, $m = 1$, $m = 2$ 及び $m = 3$ として、いずれかのモデルを選択することを考える。

交差検証法では、データを γ 組に分割する。分割したデータをそれぞれ $d_{(1)}, d_{(2)}, \dots, d_{(\gamma)}$ とする。そして、 $\{d_{(1)}, d_{(2)}, \dots, d_{(\gamma)}\} \setminus d_{(j)}$ のデータ (d^{train}) でモデル (4.2) における $m = 0$, $m = 1$, $m = 2$ 及び $m = 3$ の各モデルのパラメータを推定し、 $d_{(j)}$ のデータをそのモデルに当てはめる。これを γ 回繰り返す。変化点 $\text{argmin}_m \sum^\gamma -l(\hat{\beta}_{d^{\text{train}}}, \hat{\mathbf{k}}_{d^{\text{train}}}; d_{(j)})$ 個のモデルを選択する。各 $d_{(j)}$ のサイズが n/γ となるような、いわゆる γ -fold CV を考えることとし、本節では一般に用いられる 10-fold CV 及び 5-fold CV を比較対象とする。なお、Cox 比例ハザードモデルは共変量間のハザード比を推定しており、テストデータには異なる共変量をもつ情報が2つ以上必要になることから、 n -fold CV (leave-one-out cross-validation) は適用できないため比較対象としていない。

表 4.5 では、真の変化点の数 m^* を1個とし、真の変化点と真の変化量を変えながら、真のモデ

表 4.1: 1 個の変化点をもつ真のモデルの下での最大対数部分尤度のバイアス

α^*	$\exp(\beta^{*(2)})$	r^*	$ D : 50$	$ D : 100$	$ D : 150$	$ D : 200$
0.3	0.9	0.0	4.87 (0.30)	4.97 (0.29)	4.94 (0.28)	4.86 (0.30)
		0.1	5.04 (0.36)	5.42 (0.31)	4.96 (0.33)	4.55 (0.39)
		0.2	5.01 (0.34)	5.03 (0.40)	5.30 (0.46)	5.29 (0.48)
	0.8	0.0	4.97 (0.33)	4.93 (0.30)	5.03 (0.31)	4.85 (0.32)
		0.1	5.02 (0.37)	5.25 (0.32)	5.09 (0.37)	4.59 (0.44)
		0.2	4.93 (0.34)	5.05 (0.39)	5.37 (0.48)	5.75 (0.52)
	0.7	0.0	5.11 (0.36)	5.13 (0.33)	5.03 (0.35)	4.73 (0.36)
		0.1	5.15 (0.39)	5.27 (0.33)	5.14 (0.41)	4.61 (0.45)
		0.2	4.64 (0.34)	5.00 (0.41)	5.05 (0.49)	5.81 (0.51)
	0.6	0.0	5.36 (0.42)	5.29 (0.38)	4.96 (0.40)	4.95 (0.46)
		0.1	5.12 (0.41)	5.46 (0.44)	5.06 (0.48)	4.64 (0.50)
		0.2	4.49 (0.40)	5.51 (0.50)	5.24 (0.53)	5.89 (0.61)
0.5	0.9	0.0	5.15 (0.37)	4.99 (0.28)	5.20 (0.27)	5.02 (0.30)
		0.1	5.19 (0.37)	5.21 (0.29)	5.01 (0.35)	4.57 (0.37)
		0.2	4.77 (0.35)	5.01 (0.39)	5.44 (0.46)	5.26 (0.47)
	0.8	0.0	5.33 (0.39)	4.91 (0.30)	5.14 (0.28)	5.06 (0.31)
		0.1	6.17 (1.26)	5.24 (0.31)	5.18 (0.37)	4.72 (0.41)
		0.2	4.52 (0.33)	5.11 (0.38)	5.59 (0.46)	5.71 (0.52)
	0.7	0.0	5.40 (0.43)	5.01 (0.33)	5.01 (0.31)	4.85 (0.37)
		0.1	6.43 (1.33)	5.39 (0.35)	5.18 (0.39)	4.86 (0.44)
		0.2	4.54 (0.36)	4.99 (0.38)	5.34 (0.46)	5.46 (0.53)
	0.6	0.0	5.40 (0.43)	4.84 (0.35)	5.08 (0.34)	4.98 (0.41)
		0.1	5.34 (0.41)	5.46 (0.38)	5.14 (0.48)	4.83 (0.46)
		0.2	4.82 (0.38)	4.95 (0.42)	4.98 (0.50)	6.16 (0.59)

モデル (4.1) に基づいたモンテカルロ法 (繰返し回数: 100 回) により得られたバイアスの平均値 (標準誤差) を示した. 真の変化点 k^* は $P(t < k^*) = \alpha^*$ を満たし, 真の回帰パラメータを $\beta^{(1)*} = 0$ と $\beta^{(2)*}$, 打ち切り確率を r^* , ベースラインハザード関数を $\lambda_0(t) = 0.1$, イベントサイズを $|D|$ とする.

表 4.2: 0 個又は1 個の変化点をもつ真のモデルの下での真のモデルと推定したモデル間の Kullback-Leibler ダイバージェンス (K-L) と変化点の数の選択確率

α^*	$ D $	$\exp(\beta^{*(2)})$	m^*		K-L	0 (%)	1 (%)	2 (%)	3 (%)
0.3	50	1.00	0	AIC _{naive}	2.36	63	25	9	3
				AIC	0.74	96	4	0	0
		0.50	1	AIC _{naive}	3.51	45	31	20	4
				AIC	2.04	80	20	0	0
		0.25	1	AIC _{naive}	6.74	16	63	14	7
				AIC	4.12	51	49	0	0
	100	1.00	0	AIC _{naive}	4.97	37	11	30	22
				AIC	0.96	95	2	2	1
		0.50	1	AIC _{naive}	5.53	25	24	26	25
				AIC	2.66	74	21	4	1
		0.25	1	AIC _{naive}	5.10	5	56	20	19
				AIC	4.14	25	68	5	2
0.5	50	1.00	0	AIC _{naive}	3.25	57	19	20	4
				AIC	1.17	93	5	2	0
		0.50	1	AIC _{naive}	3.28	46	36	11	7
				AIC	2.14	83	15	2	0
		0.25	1	AIC _{naive}	5.30	11	64	19	6
				AIC	3.26	44	54	2	0
	100	1.00	0	AIC _{naive}	4.53	44	10	27	19
				AIC	1.06	94	2	4	0
		0.50	1	AIC _{naive}	5.59	17	30	29	24
				AIC	3.19	66	24	9	1
		0.25	1	AIC _{naive}	6.00	1	38	33	28
				AIC	4.10	13	74	13	0

モデル (4.2) に基づいたモンテカルロ法 (繰り返し回数: 100 回) により得られた結果を示した. 真の構造は (4.1) により与えられ, 真の変化点 k^* は $P(t < k^*) = \alpha^*$ を満たし, 真の回帰パラメータを $\beta^{(1)*} = 0$ と $\beta^{(2)*}$, 打ち切り確率は $\{0.0, 0.1, 0.2\}$ 上の離散一様分布, ベースラインハザード関数を $\lambda_0(t) = 0.1$, イベントサイズを $|D|$ とする.

表 4.3: 2個の変化点をもつ真のモデルの下での真のモデルと推定したモデル間の Kullback-Leibler ダイバージェンス (K-L) と変化点の数の選択確率

$ D $	ψ^*		K-L	0 (%)	1 (%)	2 (%)	3 (%)	4 (%)
200	2.0	AIC _{naive}	7.84	0	4	41	31	24
		AIC	6.30	2	26	68	4	0
		BIC _{naive}	6.82	7	40	51	2	0
		Sequential Test	7.03	11	38	50	1	0
	1.8	AIC _{naive}	8.12	0	5	40	28	27
		AIC	6.34	4	29	61	6	0
		BIC _{naive}	7.16	10	44	44	2	0
		Sequential Test	7.32	14	45	40	1	0
	1.6	AIC _{naive}	8.26	0	4	37	27	32
		AIC	6.36	5	35	49	9	2
		BIC _{naive}	6.81	11	53	33	3	0
		Sequential Test	6.65	13	53	34	0	0
400	1.4	AIC _{naive}	8.98	0	1	26	33	40
		AIC	6.41	1	14	71	13	1
		BIC _{naive}	7.07	8	27	64	1	0
		Sequential Test	7.03	10	27	63	0	0
	1.2	AIC _{naive}	9.49	0	2	20	38	40
		AIC	6.91	5	23	61	10	1
		BIC _{naive}	7.37	10	40	49	1	0
		Sequential Test	7.04	13	34	53	0	0
	1.0	AIC _{naive}	9.29	1	4	22	35	38
		AIC	6.66	3	30	56	9	2
		BIC _{naive}	7.33	14	49	36	1	0
		Sequential Test	7.15	16	45	38	1	0

モデル (4.2) に基づいたモンテカルロ法 (繰り返し回数: 100 回) により得られた結果を示した。真の変化点 $k^{*(1)}$ と $k^{*(2)}$ は $P(t < k^{*(1)}) = \alpha_1^*$ と $P(k^{*(1)} < t < k^{*(2)}) = \alpha_2^*$ を満たし, α_1^* と α_2^* は $[0.1, 0.4]$ 上の連続量一様分布に従うものとする。真の変化量は $\exp(\beta^{*(2)})/\exp(\beta^{*(1)}) = 2^{u_1^*(\psi^*+v_1^*)}$ と $\exp(\beta^{*(3)})/\exp(\beta^{*(2)}) = 2^{u_2^*(\psi^*+v_2^*)}$ とし, u_1^* と u_2^* は $\{-1, 1\}$ 上の離散一様分布に, v_1^* と v_2^* は $[0, 1]$ 上の連続量一様分布に従い, $\beta^{*(1)} = 0$ とする。打ち切り確率は $\{0.0, 0.1, 0.2\}$ 上の離散一様分布に従い, ベースラインハザード関数を $\lambda_0(t) = 0.1$, イベントサイズを $|D|$ とする。

表 4.4: 共変量間に相関のある2個の変化点をもつ真のモデルの下での真のモデルと推定したモデル間の Kullback-Leibler ダイバージェンス (K-L) と変化点の数の選択確率

$ D $	ψ^*	ρ^*		K-L	0 (%)	1 (%)	2 (%)	3 (%)	4 (%)
200	2.0	0.4	AIC _{naive}	9.81	0	12	40	32	16
			AIC	7.75	1	39	52	8	0
			AIC _{ξ}	6.64	0	34	46	16	4
			BIC _{naive}	8.80	18	62	20	0	0
	0.8	AIC _{naive}	7.90	0	30	60	8	2	
		AIC	7.47	2	57	41	0	0	
		AIC _{ξ}	6.52	0	48	49	3	0	
		BIC _{naive}	8.20	11	67	22	0	0	
	1.6	AIC _{naive}	9.96	0	11	35	34	20	
		AIC	7.74	3	49	39	9	0	
		AIC _{ξ}	6.24	1	34	43	19	3	
		BIC _{naive}	8.20	23	63	14	0	0	
0.8	AIC _{naive}	7.18	0	38	53	9	0		
	AIC	6.85	2	60	37	1	0		
	AIC _{ξ}	5.69	1	57	39	3	0		
	BIC _{naive}	7.94	24	63	13	0	0		
400	1.4	0.4	AIC _{naive}	12.11	0	3	32	30	35
			AIC	8.88	0	23	59	18	0
			AIC _{ξ}	7.73	0	20	55	20	5
			BIC _{naive}	9.13	8	54	38	0	0
	0.8	AIC _{naive}	9.89	0	14	45	25	16	
		AIC	7.64	0	35	56	9	0	
		AIC _{ξ}	6.32	0	29	56	13	2	
		BIC _{naive}	7.81	7	60	33	0	0	
	1.0	0.4	AIC _{naive}	12.33	0	9	25	32	34
			AIC	8.36	1	32	55	11	1
			AIC _{ξ}	7.16	0	26	48	17	9
			BIC _{naive}	8.85	17	66	17	0	0
0.8	AIC _{naive}	9.56	0	21	41	20	18		
	AIC	7.32	1	47	46	5	1		
	AIC _{ξ}	6.23	0	38	44	16	2		
	BIC _{naive}	8.08	15	60	25	0	0		

真のパラメータを変えながらモデル (4.3) に基づいたモンテカルロ法 (繰り返し回数: 100 回) により得られた結果を示した. 真の変化点 $k^{*(1)}$ と $k^{*(2)}$ は $P(t < k^{*(1)}) = \alpha_1^*$ と $P(k^{*(1)} < t < k^{*(2)}) = \alpha_2^*$ を満たし, α_1^* と α_2^* は $[0.1, 0.4]$ 上の連続量一様分布に従うものとする. 真の変化量は $\exp(\beta^{*(2)}) / \exp(\beta^{*(1)}) = 2^{u_1^*(\psi^* + v_1^*)}$ と $\exp(\beta^{*(3)}) / \exp(\beta^{*(2)}) = 2^{u_2^*(\psi^* + v_2^*)}$ とし, u_1^* と u_2^* は $\{-1, 1\}$ 上の離散一様分布に, v_1^* と v_2^* は $[0, 1]$ 上の連続量一様分布に従い, $\beta^{*(1)} = 0$ とする. z_1 と z_2 の相関を ρ^* とし, 多重共線性の大きさを表す. 打ち切り確率は $\{0.0, 0.1, 0.2\}$ 上の離散一様分布に従い, ベースラインハザード関数を $\lambda_0(t) = 0.1$, イベントサイズを $|D|$ とする.

ルと AIC 又は交差検証法で選択されたモデル間の Kullback-Leibler ダイバージェンス (K-L) 及び変化点の数の選択確率を評価している。AIC はすべての設定で 1 番目か 2 番目に小さい K-L を与え、AIC が 2 番目に小さい K-L を与えた設定においては 1 番小さい K-L を与えた 5-fold CV との差は小さいものであった。一方、5-fold CV は AIC 及び 10-fold CV よりも大きい K-L を与えた設定もいくつかあった。いずれの設定においても手法間の K-L の差は小さいものであったが、本実験の設定では、AIC は 10-fold CV 又は 5-fold CV と比較して予測の観点でのモデル選択手法として同等あるいはやや優れていると言えよう。

次に、交差検証法ではデータの分割を変えることでモデル選択結果にばらつきが生じることが想定される。そこで、データの分割を変えることによるばらつきの大きさを評価した。具体的には、1つのシミュレーションデータに対して 10 パターンのデータ分割を行い、K-L 及び選択した変化点の数の標準偏差を評価した。それを 10 回繰り返すモンテカルロ法により得られた標準偏差の平均値等を表 4.6 に示した。変化点の数の標準偏差の平均値に着目すると、真の変化点の数を 1 に設定した本実験において標準偏差は 0.5 付近の値を取っていることから、データの分割を変えれば選択されるモデルが相当に変わってしまうことを示唆している。真の変化点の数が 2 以上の設定や選択され得る変化点の数を増やして状況では、それに応じて標準偏差も大きくなることが想定される。つまり、データ分割を変えることが意思決定の結論を変えてしまうことに繋がりがかねないものとなり、それは恣意的に結果を変えることを可能にしてしまう。

以上より、予測の観点で適切なモデル選択を目的する状況では、まずは (3.14) の AIC を用いることが妥当であり、無難だと言える。

4.2 実データ解析

本章では、悪性神経膠腫の患者を対象としたプラセボ対照ランダム化比較試験のデータに対して、(3.14) の AIC, AIC_{naive} , 及び BIC_{naive} を適用する。この臨床試験の目的の一つは、再発腫瘍の外科的切除後にカルムスチン含浸ポリマーを脳腫瘍部位に移植することで、より長い生存期間が得られるかどうかを検証することであった。27 の医療機関から合計 222 例の患者が登録され、110 例が被験群に、112 例が対照群にランダムに割り付けられた。試験デザインや解析結果の詳細は [Brem et al. \(1995\)](#) で報告されている。このデータにおいて、腫瘍切除範囲が 75%以上かそれ未満かでグループ分けしたときの生存曲線は、切除後一定期間は重なって推移するものの、その

表 4.5: 1個の変化点をもつ真のモデルの下での真のモデルと AIC 又は交差検証法で推定したモデル間の Kullback-Leibler ダイバージェンス (K-L) と変化点の数の選択確率

α^*	$ D $	$\exp(\beta^{*(2)})$	K-L	0 (%)	1 (%)	2 (%)	3 (%)	
0.3	100	AIC	2.66	74	21	4	1	
		0.50	10-fold CV	2.70	72	16	11	1
		5-fold CV	2.88	72	14	9	5	
	200	AIC	4.14	25	68	5	2	
		0.25	10-fold CV	4.36	34	50	12	4
		5-fold CV	4.67	38	45	13	4	
0.5	100	AIC	3.73	41	51	6	2	
		0.50	10-fold CV	3.92	43	42	6	9
		5-fold CV	3.70	49	36	10	5	
	200	AIC	3.64	2	87	10	1	
		0.25	10-fold CV	4.18	3	74	12	11
		5-fold CV	4.29	7	70	16	7	
0.1	100	AIC	3.19	66	24	9	1	
		0.50	10-fold CV	3.50	61	23	13	3
		5-fold CV	2.99	69	21	8	2	
	200	AIC	4.10	13	74	13	0	
		0.25	10-fold CV	4.52	26	47	21	6
		5-fold CV	4.29	22	61	13	4	
0.2	100	AIC	4.09	37	51	11	1	
		0.50	10-fold CV	4.14	37	47	13	3
		5-fold CV	3.83	44	46	7	3	
	200	AIC	3.69	1	84	14	1	
		0.25	10-fold CV	4.71	4	66	20	10
		5-fold CV	3.83	4	79	13	4	

モデル (4.2) に基づいたモンテカルロ法 (繰り返し回数: 100 回) により得られた結果を示した. 真の構造は (4.1) により与えられ, 真の変化点 k^* は $P(t < k^*) = \alpha^*$ を満たし, 真の回帰パラメータを $\beta^{(1)*} = 0$ と $\beta^{(2)*}$, 打ち切り確率は $\{0.0, 0.1, 0.2\}$ 上の離散一様分布, ベースラインハザード関数を $\lambda_0(t) = 0.1$, イベントサイズを $|D|$ とする.

表 4.6: 1 個の変化点をもつ真のモデルの下での交差検証法で推定されたモデルの Kullback-Leibler ダイバージェンス (K-L) と変化点の数のばらつき評価

α	$ D $	$\exp(\beta^{*(2)})$		K-L	# of change-points
0.3	100	0.50	10-folds CV	2.53 (1.13)	0.46 (0.48)
			5-folds CV	2.48 (1.09)	0.35 (0.38)
	200	0.25	10-folds CV	4.25 (1.79)	1.00 (0.65)
			5-folds CV	4.27 (1.59)	1.00 (0.60)
	100	0.50	10-folds CV	3.12 (0.72)	0.67 (0.42)
			5-folds CV	3.16 (0.60)	0.57 (0.45)
200	0.25	10-folds CV	3.52 (0.81)	1.12 (0.28)	
		5-folds CV	4.03 (1.32)	1.20 (0.37)	
0.5	100	0.5	10-folds CV	2.41 (0.94)	0.47 (0.42)
			5-folds CV	2.74 (1.13)	0.57 (0.47)
	200	0.25	10-folds CV	3.60 (1.64)	0.98 (0.42)
			5-folds CV	3.77 (1.65)	0.95 (0.46)
	100	0.50	10-folds CV	3.71 (1.20)	1.06 (0.36)
			5-folds CV	3.03 (0.91)	0.86 (0.44)
200	0.25	10-folds CV	4.19 (0.73)	1.36 (0.25)	
		5-folds CV	4.15 (1.19)	1.30 (0.28)	

10 パターンのデータ分割によるばらつき評価のためのモデル (4.2) に基づいたモンテカルロ法 (繰り返し回数: 10 回) により得られた平均値 (各繰り返し実験でのデータ分割による標準偏差の平均値) を示した. 真の構造は (4.1) により与えられ, 真の変化点 k^* は $P(t < k^*) = \alpha^*$ を満たし, 真の回帰パラメータを $\beta^{(1)*} = 0$ と $\beta^{(2)*}$, 打ち切り確率は $\{0.0, 0.1, 0.2\}$ 上の離散一様分布, ベースラインハザード関数を $\lambda_0(t) = 0.1$, イベントサイズを $|D|$ とする.

表 4.7: 臨床試験データに適用したときの変化点の推定値 \hat{k} , 最大対数部分尤度 $l(\hat{\beta}, \hat{k}; t)$, AIC_{naive} , AIC, BIC_{naive}

m	$\hat{k}^{(1)}$	$\hat{k}^{(2)}$	$\hat{k}^{(3)}$	$\hat{k}^{(4)}$	$l(\hat{\beta}, \hat{k}; t)$	AIC_{naive}	AIC	BIC_{naive}
0					-2169.65	4341.29	4341.29	4345.32
1	14.43				-2164.92	4335.83	4339.83	4347.91
2	10.57	12.00			-2161.69	4333.39	4341.39	4353.52
3	10.57	12.00	30.71		-2158.72	4331.45	4343.45	4359.63
4	10.57	12.00	30.71	33.14	-2156.36	4330.72	4346.72	4366.95

後2つの曲線が離れていくように見受けられた。

そこで、一定のイベントサイズの下での挙動を確認するために、各個人を2で重みづけしたデータに AIC, AIC_{naive} 及び BIC_{naive} を適用して変化点を探索した。具体的には、 z を腫瘍切除範囲が75%以上かそれ未満かの指示変数とし、(4.2)のモデルの下で $m = 0$, $m = 1$, $m = 2$, $m = 3$ 及び $m = 4$ のいずれかのモデルのうち、予測の観点で良いモデルを選択することを目的に、3つの情報量規準を適用した。その結果を表4.7に示している。 AIC_{naive} が変化点4個のモデルを選択し、 BIC_{naive} が変化点なしのモデルを選択している一方、AICは変化点1個のモデルを選択した。 AIC_{naive} が選択したモデルの変化点は、10.6週、12.0週、30.7週、33.1週、AICが選択したモデルの変化点は14.4週であった。この結果は、3つの情報量規準の性質が大きく異なり、それが選択するモデルへ影響することを示唆している。実際、図4.1の腫瘍切除範囲が75%以上かどうかで分けた2つのKaplan-Meierを見てみると、16週付近までは2つの曲線が重なっており、その後曲線が広がっていることが分かる。このことから、その時点付近に少なくとも1つの構造変化が起きていると考えることは妥当であろう。

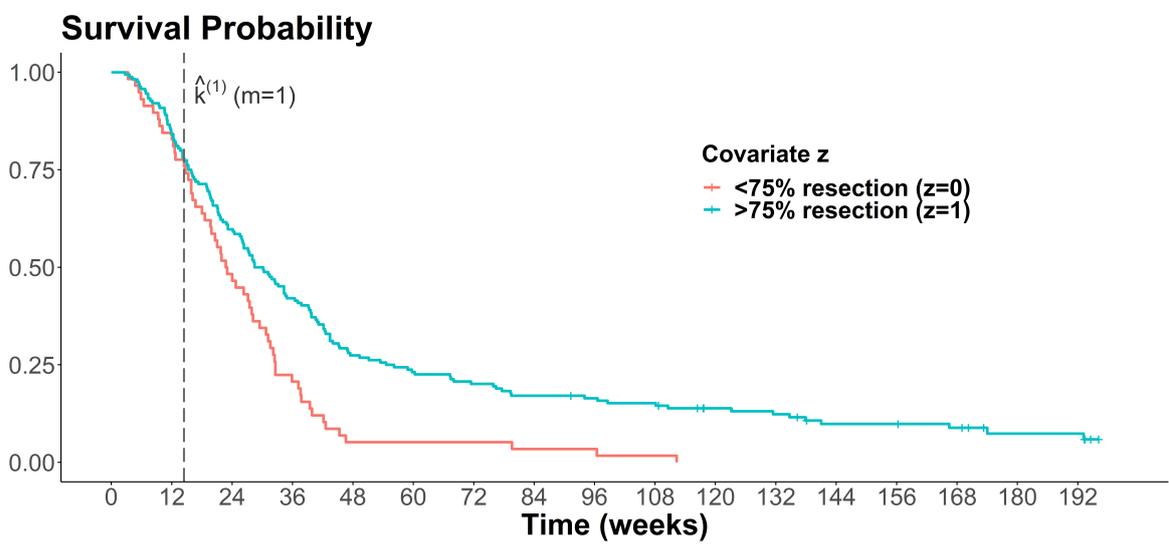


図 4.1: 腫瘍切除範囲が75%以上かそれ未満かでグループ分けしたとき Kaplan-Meier 曲線. <75% グループの約95%のイベントが48週までに発現し, 残りのイベントは約80週以降に発現していたことから, 変化点 k の上限は48週とした.

第5章 モデル誤特定を許容する情報量規準への 拡張

本章では $\xi = 0$ とし, 通常最大対数部分尤度の漸近バイアスを罰則に与えた (3.14) の AIC に対して, モデル誤特定を許容する情報量規準に拡張する. ここまで, $\beta^* = (\beta^{*(1)\top}, \beta^{*(2)\top}, \dots, \beta^{*(m+1)\top})^\top$ を $\operatorname{argsup}_{\beta} E\{l(\beta, \mathbf{k}^*; \mathbf{t})\}$ と定義していた. モデル誤特定がない状況では, これは (2.1) のモデルにおける $\beta = (\beta^{(1)\top}, \beta^{(2)\top}, \dots, \beta^{(m+1)\top})^\top$ の真値である. しかし, このモデルは潜在的に共変量 \mathbf{z} とハザード関数 $\lambda(t | \mathbf{z})$ との関係に対数線形性を仮定している. そのため, この仮定が成立しない状況でモデル (2.1) を適用することはモデル誤特定しているということになる. また, モデルに含める必要のある共変量の利用が限定されている状況や, \mathbf{z} が与えられた下でイベント及び打ち切りの発生時間である y_1 と y_2 との間に独立性が存在しない状況では, モデル誤特定が起きていると言える. そこで, モデル誤特定を想定した下でのモデル (2.1) に対する情報量規準を導出する. なお, モデル誤特定を許容する状況では, β^* が真値である必要はない.

まず, [Struthers and Kalbfleisch \(1986\)](#) 及び [Lin and Wei \(1989\)](#) より

$$\sqrt{n}(\hat{\beta}_{\mathbf{k}^*}^{(j)} - \beta^{*(j)}) \xrightarrow{d} N(\mathbf{0}_p, \mathbf{B}_0^{*(j)-1} \mathbf{A}_0^{*(j)} \mathbf{B}_0^{*(j)-1})$$

が言える. ここで,

$$\mathbf{w}^{(j)}(\beta, \mathbf{k}) \equiv \sum_{i \in D^{(j)}} \{\mathbf{z}_i - \mathbf{h}(t_i, \beta^{(j)})\} - \sum_{i=1}^n \sum_{l \in D^{(j)}: t_l < t_i} \frac{\exp(\beta^{(j)\top} \mathbf{z}_i)}{\sum_{i' \in R(t_l)} \exp(\beta^{(j)\top} \mathbf{z}_{i'})} \{\mathbf{z}_i - \mathbf{h}(t_l, \beta^{(j)})\}$$

として

$$\mathbf{A}_0^{*(j)} \equiv E \left\{ \frac{1}{n} \mathbf{w}^{(j)}(\beta^*, \mathbf{k}^*) \mathbf{w}^{(j)}(\beta^*, \mathbf{k}^*)^\top \right\}$$

と

$$\mathbf{B}_0^{*(j)} \equiv \mathbb{E} \left[\frac{1}{n} \sum_{i \in D^{(j)}} \{ \mathbf{H}(t_i, \boldsymbol{\beta}^{(j)}) - \mathbf{h}(t_i, \boldsymbol{\beta}^{(j)}) \mathbf{h}(t_i, \boldsymbol{\beta}^{(j)})^\top \} \right]$$

である.

各 $j \in \{1, 2, \dots, m\}$ において $k^{(j)} = k^{*(j)} + s^{(j)}/n$ とおく. 最初に, $\mathbf{s} = (s^{(1)}, s^{(2)}, \dots, s^{(m)})^\top$ の各要素が有限値である場合を考える. (A.1) と同様にして,

$$\widehat{\boldsymbol{\beta}}_{\mathbf{k}} - \widehat{\boldsymbol{\beta}}_{\mathbf{k}^*} = \text{O}_P(1/n) \quad (5.1)$$

を示すことができる. このとき, (2.4) と同様の手順により

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\mathbf{k}}^{(j)} - \boldsymbol{\beta}^{*(j)}) \xrightarrow{d} \text{N}(\mathbf{0}_p, \mathbf{B}_0^{*(j)-1} \mathbf{A}_0^{*(j)} \mathbf{B}_0^{*(j)-1}) \quad (5.2)$$

が言え, $\xi = 0$ の $Q_{\xi; \mathbf{k}, \mathbf{t}}^{(j)}$ として定義される両側に伸びる負のドリフト付きランダムウォーク $Q_{\mathbf{k}, \mathbf{t}}^{(j)}$ を用いて

$$\begin{aligned} l(\widehat{\boldsymbol{\beta}}_{\mathbf{k}}, \mathbf{k}; \mathbf{t}) - l(\widehat{\boldsymbol{\beta}}_{\mathbf{k}^*}, \mathbf{k}^*; \mathbf{t}) &= l(\boldsymbol{\beta}_\xi^*, \mathbf{k}; \mathbf{t}) - l(\boldsymbol{\beta}_\xi^*, \mathbf{k}^*; \mathbf{t}) + \text{O}_P(1) \\ &= \sum_{j=1}^m Q_{k^{*(j)} + s^{(j)}/n, \mathbf{t}}^{(j)} + \text{O}_P(1) = \text{O}_P(1) \end{aligned} \quad (5.3)$$

が得られる. 更に, 対数部分尤度に対して, $\widehat{\boldsymbol{\beta}}_{\mathbf{k}^*}^{(j)} = \boldsymbol{\beta}^{*(j)}$ の周りでテイラー展開し, (5.2) 及び [Murphy and van der Vaart \(2000\)](#) を利用すると

$$l(\widehat{\boldsymbol{\beta}}_{\mathbf{k}^*}, \mathbf{k}^*; \mathbf{t}) - l(\boldsymbol{\beta}^*, \mathbf{k}^*; \mathbf{t}) = \frac{1}{2} \sum_{j=1}^{m+1} \boldsymbol{\nu}^{(j)\top} \boldsymbol{\nu}^{(j)} + \text{O}_P(1) = \text{O}_P(1) \quad (5.4)$$

が得られる. ここで, $\boldsymbol{\nu}^{(j)}$ は多変量正規分布 $\text{N}(\mathbf{0}_p, \mathbf{B}_0^{*(j)-1} \mathbf{A}_0^{*(j)})$ に従う確率変数ベクトルである. よって, (5.3) 及び (5.4) より, $\widehat{l}(\mathbf{k}; \mathbf{t}, \mathbf{t}) = \text{O}_P(1)$ であることが言える.

次に, \mathbf{s} のいずれかの要素が有限値でない場合を考える. ここでも 3 章と同様に,

$$\begin{cases} k^{(j')} = k^{*(j')} + s^{(j')}/n, & 0 > s^{(j')} \neq \text{O}(1) \\ k^{(j)} = k^{*(j)} + s^{(j)}/n, & s^{(j)} = \text{O}(1) \quad (j \neq j') \end{cases}$$

の場合のみを考える。このとき (B.1) において $\xi = 0$ とすると, 任意の $M > 0$ に対して $P\{l(\hat{\beta}_{\mathbf{k}}, \mathbf{k}) - l(\hat{\beta}_{\mathbf{k}^*}, \mathbf{k}^*) > -M\} \rightarrow 0$ が言え, (5.4) より任意の $M > 0$ に対して $P\{\hat{l}(\mathbf{k}; \mathbf{t}, \mathbf{t}) > -M\} \rightarrow 0$ であることが分かる。よって, $\hat{\mathbf{k}} - \mathbf{k}^* = O_P(1/n)$ が言える。

以上より, $K^{(j)} = \{\mathbf{k} \mid \mathbf{k} - \mathbf{k}^{*(j)} = O(1/n)\}$ として $K = \{\mathbf{k} \mid \mathbf{k}^{(j)} \in K^{(j)}, j \in \{1, 2, \dots, m\}\}$ であることが分かる。そして, (5.3) 及び (5.4) より

$$\sup_{\mathbf{k} \in K} \hat{l}(\mathbf{k}; \mathbf{t}, \mathbf{t}) = \sum_{j=1}^m \sup_{\mathbf{k} \in K^{(j)}} Q_{\mathbf{k}, \mathbf{t}}^{(j)} + \frac{1}{2} \sum_{j=1}^{m+1} \boldsymbol{\nu}^{(j)\top} \boldsymbol{\nu}^{(j)} + o_P(1) \quad (5.5)$$

と

$$\operatorname{argsup}_{\mathbf{k} \in K} \hat{l}(\mathbf{k}; \mathbf{u}, \mathbf{u}) = \left(\operatorname{argsup}_{\mathbf{k} \in K^{(1)}} Q_{\mathbf{k}, \mathbf{u}}^{(1)}, \operatorname{argsup}_{\mathbf{k} \in K^{(2)}} Q_{\mathbf{k}, \mathbf{u}}^{(2)}, \dots, \operatorname{argsup}_{\mathbf{k} \in K^{(m)}} Q_{\mathbf{k}, \mathbf{u}}^{(m)} \right)^\top + o_P(1) \quad (5.6)$$

が得られる。さらに, $\check{\mathbf{k}}_{\mathbf{u}}^{(j)} \equiv \operatorname{argsup}_{\mathbf{k} \in K^{(j)}} Q_{\mathbf{k}, \mathbf{u}}^{(j)}$ として, $\check{\mathbf{k}}_{\mathbf{u}} \equiv (\check{\mathbf{k}}_{\mathbf{u}}^{(1)}, \check{\mathbf{k}}_{\mathbf{u}}^{(2)}, \dots, \check{\mathbf{k}}_{\mathbf{u}}^{(m)})^\top$ とすると

$$\hat{\beta}_{\check{\mathbf{k}}_{\mathbf{u}}, \mathbf{u}} - \hat{\beta}_{\mathbf{k}^*, \mathbf{u}} = O_P(1/n) \quad (5.7)$$

と

$$\hat{\beta}_{\check{\mathbf{k}}_{\mathbf{u}}, \mathbf{u}} - \beta^* = O_P(1/\sqrt{n}) \quad (5.8)$$

が分かる。 $\hat{l}(\check{\mathbf{k}}_{\mathbf{u}}; \mathbf{t}, \mathbf{u})$ に対し, (C.1) と同様に [Murphy and van der Vaart \(2000\)](#) や (5.7), (5.8) を利用すると

$$\begin{aligned} & l(\beta^*, \check{\mathbf{k}}_{\mathbf{u}}; \mathbf{t}) - l(\hat{\beta}_{\check{\mathbf{k}}_{\mathbf{u}}, \mathbf{u}}, \check{\mathbf{k}}_{\mathbf{u}}; \mathbf{t}) \\ &= - \sum_{j=1}^{m+1} \left[(\hat{\beta}_{\check{\mathbf{k}}_{\mathbf{u}}, \mathbf{u}}^{(j)} - \beta^{*(j)})^\top \sum_{i \in \tilde{D}^{*(j)}} \{z_i - \mathbf{h}(t_i, \beta^{*(j)})\} + \frac{1}{2} \boldsymbol{\nu}^{(j)\top} \boldsymbol{\nu}^{(j)} \right] + o_P(1) \end{aligned} \quad (5.9)$$

が得られる。また,

$$l(\beta^*, \mathbf{k}^*; \mathbf{t}) - l(\beta^*, \check{\mathbf{k}}_{\mathbf{u}}; \mathbf{t}) = \sum_{j=1}^m Q_{\check{\mathbf{k}}_{\mathbf{u}}^{(j)}, \mathbf{t}}^{(j)} \quad (5.10)$$

であることから, (5.6) と (5.9) も利用して

$$\begin{aligned}
& \widehat{l} \left\{ \operatorname{argsup}_{\mathbf{k} \in K} \widehat{l}(\mathbf{k}; \mathbf{u}, \mathbf{u}); \mathbf{t}, \mathbf{u} \right\} \\
&= \widehat{l}(\tilde{\mathbf{k}}_{\mathbf{u}}; \mathbf{t}, \mathbf{u}) + o_{\mathbb{P}}(1) \\
&= - \sum_{j=1}^m Q_{\tilde{\mathbf{k}}_{\mathbf{u}}^{(j)}, \mathbf{t}}^{(j)} + \sum_{j=1}^{m+1} \left[(\widehat{\boldsymbol{\beta}}_{\tilde{\mathbf{k}}_{\mathbf{u}}^{(j)}}^{(j)} - \boldsymbol{\beta}^{*(j)})^{\top} \sum_{i \in \bar{D}^{*(j)}} \{z_i - \mathbf{h}(t_i, \boldsymbol{\beta}^{*(j)})\} - \frac{1}{2} \boldsymbol{\nu}^{(j)\top} \boldsymbol{\nu}^{(j)} \right] + o_{\mathbb{P}}(1) \quad (5.11)
\end{aligned}$$

が分かる. (5.5) と (5.11) より, 次の系 2 が得られる.

系 2. 定理 1 の条件の下で, モデル誤特定を許容する状況において (2.3) で $\xi = 0$ のとき, (2.5) で考えている漸近バイアスは

$$\mathbb{E}\{b(\mathbf{k}^*, \boldsymbol{\beta}^*)\} = \sum_{j=1}^m \mathbb{E} \left(\sup_{\mathbf{k} \in K^{(j)}} Q_{\mathbf{k}, \mathbf{t}}^{(j)} + Q_{\operatorname{argsup}_{\mathbf{k} \in K^{(j)}} Q_{\mathbf{k}, \mathbf{u}, \mathbf{t}}^{(j)}}^{(j)} \right) + \sum_{j=1}^{m+1} \operatorname{tr}(\mathbf{A}_0^{*(j)} \mathbf{B}_0^{*(j)-1})$$

で与えられる.

ここで, 条件

$$\boldsymbol{\beta}^{*(j+1)} - \boldsymbol{\beta}^{*(j)} = \boldsymbol{\Delta}_{\boldsymbol{\beta}^*}^{(j)} / \sqrt{\alpha_n} \quad (j \in \{1, 2, \dots, m\}), \quad O(1) \neq \alpha_n = o(n) \quad (5.12)$$

を仮定する. 3.2 節と同様に, この条件の下, 変化点推定量の漸近挙動を調べる. 系 2 の導出と同様, \mathbf{s} の各要素が有限値としたとき, (5.1) の代わりに $\widehat{\boldsymbol{\beta}}_{\mathbf{k}} - \widehat{\boldsymbol{\beta}}_{\mathbf{k}^*} = O_{\mathbb{P}}(\sqrt{\alpha_n}/n)$ が得られる. そして, (5.3) と (5.4) が言えるため, $\widehat{l}(\mathbf{k}; \mathbf{t}, \mathbf{t}) = O_{\mathbb{P}}(1)$ が成立する. 一方, \mathbf{s} のいずれかの要素が有限値でないとしたときは, やはり系 2 の導出と同様, $k^{(j')} = k^{*(j')} + \alpha_n s^{(j')}/n$, $0 > s^{(j')} \neq O(1)$ かつ $j \neq j'$ に対して $k^{(j)} = k^{*(j)} + \alpha_n s^{(j)}/n$, $s^{(j)} = O(1)$ の場合を考える. このとき, やはり任意の $M > 0$ に対して $\mathbb{P}\{l(\widehat{\boldsymbol{\beta}}_{\mathbf{k}}, \mathbf{k}; \mathbf{t}) - l(\widehat{\boldsymbol{\beta}}_{\mathbf{k}^*}, \mathbf{k}^*; \mathbf{t}) > -M\} \rightarrow 0$ が言える. そして, (5.4) を併せれば $\mathbb{P}\{\widehat{l}(\mathbf{k}; \mathbf{t}, \mathbf{t}) > -M\} \rightarrow 0$ が成立する. 以上より, $K^{(j)} = \{\mathbf{k} \mid \mathbf{k} - \mathbf{k}^{*(j)} = O(\alpha_n/n)\}$ として $K = \{\mathbf{k} \mid k^{(j)} \in K^{(j)}, j \in \{1, 2, \dots, m\}\}$ とできて, (5.5) や (5.6), (5.9), (5.10), (5.11) が成り立つ.

以降, $\mathbf{k} = \mathbf{k}^* + \alpha_n \mathbf{s}/n$ において \mathbf{s} は各要素が有限値のベクトルとする. 条件 (5.12) の下, 両側

標準ブラウン運動 $\{W_s\}_{s \in \mathbb{R}}$ を用いて

$$\frac{1}{\sqrt{\alpha_n}} \Delta_{\beta^*}^{(j)\top} \left[\sum_{i \in D_{1\alpha_n}^{*(j)}} \{z_i - \mathbf{h}(t_i, \beta^{*(j)})\} \right] \xrightarrow{d} (\Delta_{\beta^*}^{(j)\top} \mathbf{A}_0^{*(j)} \Delta_{\beta^*}^{(j)})^{1/2} W_s$$

と

$$\frac{1}{\sqrt{\alpha_n}} \Delta_{\beta^*}^{(j)\top} \left[\sum_{i \in D_{2\alpha_n}^{*(j)}} \{z_i - \mathbf{h}(t_i, \beta^{*(j+1)})\} \right] \xrightarrow{d} (\Delta_{\beta^*}^{(j)\top} \mathbf{A}_0^{*(j+1)} \Delta_{\beta^*}^{(j)})^{1/2} W_s$$

が得られる。さらに、(E.3) と (E.5) において、 $\Delta^{(j)}$ を $\Delta_{\beta^*}^{(j)}$ に、 $\beta_\xi^{*(j)}$ を $\beta^{*(j)}$ に、 $B_\xi^{*(j)}$ を $B_0^{*(j)}$ に置き換える。このとき、 $V_s^{*(j)}$ を

$$V_s \left\{ \frac{1}{2} \Delta_{\beta^*}^{(j)\top} B_0^{*(j)} \Delta_{\beta^*}^{(j)}, \frac{1}{2} \Delta_{\beta^*}^{(j)\top} B_0^{*(j+1)} \Delta_{\beta^*}^{(j)}, (\Delta_{\beta^*}^{(j)\top} \mathbf{A}_0^{*(j)} \Delta_{\beta^*}^{(j)})^{1/2}, (\Delta_{\beta^*}^{(j)\top} \mathbf{A}_0^{*(j+1)} \Delta_{\beta^*}^{(j)})^{1/2} \right\}$$

として $Q_{k^* + \alpha_n s/n, t}^{(j)} \xrightarrow{d} V_s^{*(j)}$ の成立が分かる。この収束の帰結として

$$\sup_{k \in K^{(j)}} Q_{k, t}^{(j)} \xrightarrow{d} \sup_{s \in \mathbb{R}} V_s^{*(j)}$$

と

$$Q_{\operatorname{argsup}_{k \in K^{(j)}} Q_{k, \mathbf{u}, t}^{(j)}}^{(j)} \xrightarrow{d} V_{\operatorname{argsup}_{s \in \mathbb{R}} V_s^{*(j)}}^{*(j)}$$

が得られる。ここで、 $V_s'^{*(j)}$ は $V_s^{*(j)}$ のコピーである。(E.8) や (E.9) と同様にしてこれらの期待値を評価すると、次の系 3 が得られる。

系 3. 系 2 の条件に加え、条件 (5.12) が満たされているとき、(2.3) における $\xi = 0$ のときの (2.5) の漸近バイアスは

$$\mathbb{E}\{b(\mathbf{k}^*, \beta^*)\} = 2 \sum_{j=1}^m C^{(j)}(\mathbf{A}_0^{*(\cdot)}, B_0^{*(\cdot)}) + \sum_{j=1}^{m+1} \operatorname{tr}(\mathbf{A}_0^{*(j)} B_0^{*(j)-1})$$

で与えられる。

最大対数部分尤度をバイアス補正したものとしての情報量規準が、系 3 より得られるが、この漸近バイアスには未知パラメータが含まれているため、(3.13) と同様、それらを一致推定量で置き換

える。結果、モデル誤特定を許容する変化点をもつ Cox 比例ハザードモデルに対する情報量規準として

$$\text{TIC} = -2l(\hat{\boldsymbol{\beta}}, \hat{\mathbf{k}}; \mathbf{t}) + 4 \sum_{j=1}^m \hat{\mathbf{C}}^{(j)} \{ \hat{\mathbf{A}}_0^{(\cdot)}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{k}}), \hat{\mathbf{B}}_0^{(\cdot)}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{k}}) \} + 2 \sum_{j=1}^{m+1} \text{tr} \{ \hat{\mathbf{A}}_0^{(j)}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{k}}) \hat{\mathbf{B}}_0^{(j)}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{k}})^{-1} \}$$

を提案する。ここで、

$$\hat{\mathbf{A}}_0^{(j)}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{k}}) \equiv \frac{1}{n} \mathbf{w}^{(j)}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{k}}) \mathbf{w}^{(j)}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{k}})^\top$$

かつ

$$\hat{\mathbf{B}}_0^{(j)}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{k}}) \equiv \frac{1}{n} \sum_{i \in \hat{D}^{(j)}} \{ \mathbf{H}(t_i, \hat{\boldsymbol{\beta}}^{(j)}) - \mathbf{h}(t_i, \hat{\boldsymbol{\beta}}^{(j)}) \mathbf{h}(t_i, \hat{\boldsymbol{\beta}}^{(j)})^\top \}$$

である。

モデル誤特定下での TIC の性能を評価するために、変化点の数は 2 個 ($m = 2$) とし、独立な 2 つの共変量 z_1 と z_2 に対して、次のモデル

$$\lambda(t | \mathbf{z}) = \lambda_0(t) \exp(\beta_1^{(j)} z_1 + \beta_2 z_2), \quad t \in [k^{(j-1)}, k^{(j)}), \quad j \in \{1, 2, \dots, m+1\} \quad (5.13)$$

を真の構造として考える。ここで、 $\beta_2 \neq 0$ である。一方、モデル選択は、 $\beta_2 = 0$ としたモデル (5.13) において $m = 0$, $m = 1$, $m = 2$, $m = 3$ 及び $m = 4$ から行う。つまり、 z_2 が存在しないというモデル誤特定下でのモデル選択を行う。表 5.1 に、真のモデルと選択されたモデル間の K-L 及び各モデルの選択確率を示した。本実験では、予測の観点で TIC が AIC よりも僅かに優れているものの、各モデルの選択確率は両規準ともにほとんど同じであった。また、モデル誤特定下であっても、AIC は $\text{AIC}_{\text{naive}}$ よりも明らかに良い性能を示していた。

表 5.1: モデル誤特定下における真の変化点の数を2個とした下での真のモデルと推定したモデル間の Kullback-Leibler ダイバージェンス (K-L) と変化点の数の選択確率

$ D $	ψ^*		K-L	0 (%)	1 (%)	2 (%)	3 (%)	4 (%)
200	2.0	AIC _{naive}	23.82	0	5	30	31	34
		AIC	21.43	0	34	59	6	1
		TIC	21.22	0	29	63	7	1
		BIC _{naive}	21.89	9	49	40	2	0
	1.8	AIC _{naive}	23.80	0	8	28	29	35
		AIC	21.35	2	33	58	6	1
		TIC	21.29	2	31	60	6	1
		BIC _{naive}	21.62	12	47	38	3	0
	1.6	AIC _{naive}	24.11	0	9	23	28	40
		AIC	21.46	9	42	42	6	1
		TIC	21.31	9	38	45	7	1
		BIC _{naive}	21.53	16	53	29	2	0
400	1.4	AIC _{naive}	38.29	0	1	26	30	43
		AIC	36.19	1	19	67	11	2
		TIC	36.07	1	19	68	10	2
		BIC _{naive}	36.58	7	43	49	1	0
	1.2	AIC _{naive}	38.75	0	1	22	35	42
		AIC	36.69	4	23	56	14	3
		TIC	36.69	4	22	56	15	3
		BIC _{naive}	36.80	10	48	41	1	0
	1.0	AIC _{naive}	38.66	0	6	18	32	44
		AIC	36.22	4	35	46	11	4
		TIC	36.19	4	35	46	12	3
		BIC _{naive}	36.34	13	63	23	1	0

Note 真のパラメータを変えながらモデル (5.13) に基づいたモンテカルロ法 (繰り返し回数: 100 回) により得られた結果を示した. 真の変化点 $k^{*(1)}$ と $k^{*(2)}$ は $P(t < k^{*(1)}) = \alpha_1^*$ と $P(k^{*(1)} < t < k^{*(2)}) = \alpha_2^*$ を満たし, α_1^* と α_2^* は $[0.1, 0.4]$ 上の連続量一様分布に従うものとする. 真の変化量は $\exp(\beta_1^{*(2)}) / \exp(\beta_1^{*(1)}) = 2^{u_1^*(\psi^* + v_1^*)}$ と $\exp(\beta_1^{*(3)}) / \exp(\beta_1^{*(2)}) = 2^{u_2^*(\psi^* + v_2^*)}$ とし, u_1^* と u_2^* は $\{-1, 1\}$ 上の離散一様分布に, v_1^* と v_2^* は $[0, 1]$ 上の連続量一様分布に従い, $\beta^{*(1)} = 0$ とする. モデル誤特定の程度を意味する β_2^* は $\{0.5, 1.0\}$ 上の離散一様分布に従うものとする. 打ち切り確率は $\{0.0, 0.1, 0.2\}$ 上の離散一様分布に従い, ベースラインハザード関数を $\lambda_0(t) = 0.1$, イベントサイズを $|D|$ とする.

第6章 ARCHモデルへの拡張

6.1 本章の構成

本章はこれまでの章とは異なるモデルを考えるため、記号や記法はすべて本章内で新たに定義する。本章の構成は以下の通りである。6.2節では、準備としてまずARCH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。そして、元来の導出に基づき、最大対数尤度を漸近的にバイアス補正したものとして、AICを定義する。6.3節では、定理3として、モデル誤特定を許容しながら漸近バイアスを評価し、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す。そして、漸近論を考えるにあたってしばしば課され、たしかに自然であると考えられる条件を、追加で仮定する。結果、漸近バイアスはシンプルかつ陽な形で評価できることを定理4で示し、それよりTICが造作なく与えられることを確認する。また、ここでモデル誤特定がないことを仮定すれば、漸近バイアスはさらに容易に書け、変化点パラメータによる漸近バイアスは3となることを系4で示す。6.4節では数値実験をおこなう。まず、定理4や系4で与えた漸近的なバイアス評価が、実際にバイアスを精度よく近似していることを確認する。そして、導出したAICやTICと、変化点モデル固有の理論を用いずに形式的に定義したAICやBICを、性能比較する。形式的なAICやBICは、実際にJones and Dey (1995)やYao (1988)で扱われていることに注意してほしい。そして、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、提案した規準は明らかに既存の規準を優越することを確認する。

6.2 準備

6.2.1 モデルと仮定

ARCHモデルにおける変化点モデルを考える。系列 $\{x_1, x_2, \dots, x_n\}$ に対して m 個の変化点 $k^{(1)}, k^{(2)}, \dots, k^{(m)}$ を想定する。また、記法の都合上 $k^{(0)} = 0$ かつ $k^{(m+1)} = n$ とする。そして具

体的には、モデルは

$$x_i = \sqrt{h_i} \varepsilon_i, \quad h_i = \boldsymbol{\theta}^{(j)\top} \mathbf{z}_i \quad (i \in \{k^{(j-1)} + 1, k^{(j-1)} + 2, \dots, k^{(j)}\}, j \in \{1, 2, \dots, m+1\}) \quad (6.1)$$

と書かれるとする。ここで、 $\boldsymbol{\theta}^{(j)} \equiv (\theta_0^{(j)}, \theta_1^{(j)}, \dots, \theta_p^{(j)})^\top$, $\mathbf{z}_i \equiv (1, x_{i-1}^2, \dots, x_{i-p}^2)^\top$, ε_i は x_{i-1}, x_{i-2}, \dots と独立な平均が0で分散が1のホワイトノイズであり、定常性を保つために $\theta_0^{(j)} > 0$, $\theta_1^{(j)} \geq 0, \dots, \theta_p^{(j)} \geq 0$, $E(x_i^4) < \infty$ とする。また、 $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}^{(1)\top}, \boldsymbol{\theta}^{(2)\top}, \dots, \boldsymbol{\theta}^{(m+1)\top})^\top$ で真の構造に最も近いものを与える値を $\boldsymbol{\theta}^* = (\boldsymbol{\theta}^{*(1)\top}, \boldsymbol{\theta}^{*(2)\top}, \dots, \boldsymbol{\theta}^{*(m+1)\top})^\top$ とする。もしモデル誤特定がなく、つまり真の構造がこのモデル内にあるならば、 $\boldsymbol{\theta}^*$ は真値を表す。変化点モデルにおいて真の構造に近い変化点を考える、ということは変化点解析ではほとんど扱われていないので、 $\mathbf{k}^* = (k^{*(1)}, k^{*(2)}, \dots, k^{*(m)})^\top$ はそのまま $\mathbf{k} = (k^{(1)}, k^{(2)}, \dots, k^{(m)})^\top$ の真値であるとする。

変化点をもつ Cox 比例ハザードモデルを考えた時と同様に、本章でも当然

$$\boldsymbol{\theta}^{*(1)} \neq \boldsymbol{\theta}^{*(2)} \neq \dots \neq \boldsymbol{\theta}^{*(m+1)} \quad (6.2)$$

は仮定し、 $\boldsymbol{\theta}^*$ と \mathbf{k}^* は未知とする。漸近論を考える際は、 $\lim_{n \rightarrow \infty} k^{*(j)}/n = \lambda^{*(j)}$ ($j \in \{1, 2, \dots, m\}$) を仮定し、 $0 < \lambda^{*(1)} < \lambda^{*(2)} < \dots < \lambda^{*(m)} < 1$ が成り立っていて変化点は互いに十分に離れているとする。これは、変化点解析の漸近論で通常置かれる仮定である（例えば Csörgő and Horváth 1997 参照）。(6.1)において、 ε_i の真の分布は上述のように特に指定しないが、尤度ベースの議論を展開したいため、モデル上は ε_i の分布を一つ指定する。そして、簡単のため、その分布は標準正規分布であるとする。つまり、 $i-1$ 時点以前の確率過程の情報集合 \mathcal{I}_{i-1} を与えたもとの x_i の条件付き分布を $N(0, h_i)$ とする。このとき、 x_i の条件付き確率密度関数は、各 $j \in \{1, 2, \dots, m+1\}$ に対して

$$f(x_i; \boldsymbol{\theta}^{(j)} | \mathcal{I}_{i-1}) = \frac{1}{\sqrt{2\pi h_i}} \exp\left(-\frac{x_i^2}{2h_i}\right) \quad (i \in \{k^{(j-1)} + 1, k^{(j-1)} + 2, \dots, k^{(j)}\})$$

で与えられる。これを用いると $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$ の同時確率密度関数は $f(\mathbf{x}; \mathbf{k}, \boldsymbol{\theta}) = \prod_{j=1}^{m+1} \prod_{i=k^{(j-1)}+1}^{k^{(j)}} f(x_i; \boldsymbol{\theta}^{(j)} | \mathcal{I}_{i-1})$ と書ける。つまり、対数尤度は

$$l_i(\boldsymbol{\theta}^{(j)}) \equiv -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log h_i - \frac{x_i^2}{2h_i} = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\boldsymbol{\theta}^{(j)\top} \mathbf{z}_i) - \frac{x_i^2}{2\boldsymbol{\theta}^{(j)\top} \mathbf{z}_i}$$

として $l(\boldsymbol{\theta}, \mathbf{k}) = \sum_{j=1}^{m+1} \sum_{i=k^{(j-1)+1}}^{k^{(j)}} l_i(\boldsymbol{\theta}^{(j)})$ で与えられる.

6.2.2 AIC タイプの情報量規準

\mathbf{x} に基づく \mathbf{k} と $\boldsymbol{\theta}$ の最尤推定量を, それぞれ $\hat{\mathbf{k}}_x = (\hat{k}_x^{(1)}, \hat{k}_x^{(2)}, \dots, \hat{k}_x^{(m+1)})^\top$ と $\hat{\boldsymbol{\theta}}_x = (\hat{\boldsymbol{\theta}}_x^{(1)\top}, \hat{\boldsymbol{\theta}}_x^{(2)\top}, \dots, \hat{\boldsymbol{\theta}}_x^{(m+1)\top})^\top$ と書くことにする. 通常の AIC タイプの情報量規準を考えるとときと同様, $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ を \mathbf{x} のコピー, つまり独立に同じ分布に従う確率ベクトルであるとし, 真の分布と推定された分布の K-L に基づくリスク関数として $-2E_{\mathbf{y}}\{\log f(\mathbf{y}; \hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x)\}$ を想定する. $E_{\mathbf{y}}$ は \mathbf{y} に関する期待値を意味する. そして, これの初期推定量として -2 倍の最大対数尤度, つまり $-2 \log f(\mathbf{x}; \hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x)$ を考え, これを

$$E_{\mathbf{x}}[2 \log f(\mathbf{x}; \hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x) - 2E_{\mathbf{y}}\{\log f(\mathbf{y}; \hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x)\}] = E\{2 \log f(\mathbf{x}; \hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x) - 2 \log f(\mathbf{x}; \hat{\mathbf{k}}_y, \hat{\boldsymbol{\theta}}_y)\}$$

でバイアス補正する. ただ, この期待値は陽に求めることはできない. そこで, AIC 元来の導出方法と同様, バイアスを漸近評価する. 具体的には, $L_x(\mathbf{k}, \boldsymbol{\theta}) \equiv \log f(\mathbf{x}; \mathbf{k}, \boldsymbol{\theta}) - \log f(\mathbf{x}; \mathbf{k}^*, \boldsymbol{\theta}^*)$ や $\hat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{y}} \equiv \operatorname{argsup}_{\boldsymbol{\theta}} L_y(\mathbf{k}, \boldsymbol{\theta})$, $\hat{L}_{xy}(\mathbf{k}) \equiv L_x(\mathbf{k}, \hat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{y}})$ を用意し, 上記のバイアスを

$$\begin{aligned} 2E\{L_x(\hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x) - L_x(\hat{\mathbf{k}}_y, \hat{\boldsymbol{\theta}}_y)\} &= 2E\left[\sup_{\mathbf{k}, \boldsymbol{\theta}} L_x(\mathbf{k}, \boldsymbol{\theta}) - L_x\left\{\operatorname{argsup}_{\mathbf{k}, \boldsymbol{\theta}} L_y(\mathbf{k}, \boldsymbol{\theta})\right\}\right] \\ &= 2E\left[\sup_{\mathbf{k}} \hat{L}_{xx}(\mathbf{k}) - \hat{L}_{xy}\left\{\operatorname{argsup}_{\mathbf{k}} \hat{L}_{yy}(\mathbf{k})\right\}\right] \end{aligned}$$

と表現する. そして, $b(\mathbf{k}^*, \boldsymbol{\theta}^*)$ を $\sup_{\mathbf{k} \in K} \hat{L}_{xx}(\mathbf{k}) - \hat{L}_{xy}\{\operatorname{argsup}_{\mathbf{k} \in K} \hat{L}_{yy}(\mathbf{k})\}$ の弱極限とし, $2E\{b(\mathbf{k}^*, \boldsymbol{\theta}^*)\}$ を漸近バイアスとすれば

$$-2 \log f(\mathbf{x}; \hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x) + 2E\{b(\mathbf{k}^*, \boldsymbol{\theta}^*)\} \quad (6.3)$$

が AIC タイプの情報量規準である. ここで, K は $\hat{L}_{xx}(\mathbf{k})$ が $O_P(1)$ もしくは正となる \mathbf{k} の集合, つまり, ある正の M が存在して $P\{\hat{L}_{xx}(\mathbf{k}) > -M\}$ が 0 に収束しない \mathbf{k} の集合とする. もし変化点がなければ, これは通常の AIC と同じであり, $E\{b(\mathbf{k}^*, \boldsymbol{\theta}^*)\} = E\{b(\boldsymbol{\theta}^*)\}$ は $\boldsymbol{\theta}$ のパラメータ数となる.

6.3 主結果

本節では、6.1節の設定のもと、6.2節で紹介したAIC元来の導出方法に基づき、情報量規準を開発する。その際、モデルが誤特定されていることも許容して導出するので、いわばこれは変化点をもつARCHモデルに対するTICと言えるものである。本設定における最もシンプルな誤特定として、 x_i の条件付き分布、つまり ε_i の分布が、本当は正規分布ではないことを想定する。 $\hat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}} = (\hat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}^{(1)\top}, \hat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}^{(2)\top}, \dots, \hat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}^{(m+1)\top})^\top$ を変化点が $\mathbf{k} = (k^{(1)}, k^{(2)}, \dots, k^{(m)})^\top$ のときの \mathbf{x} に基づく $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)\top}, \boldsymbol{\theta}^{(2)\top}, \dots, \boldsymbol{\theta}^{(m+1)\top})^\top$ の最尤推定量、つまり $\hat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}} = \operatorname{argsup}_{\boldsymbol{\theta}} L_{\mathbf{x}}(\mathbf{k}, \boldsymbol{\theta})$ とする。より具体的には、 $\hat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}^{(j)}$ は $\sum_{i=k^{(j-1)+1}^{k^{(j)}} l_i(\boldsymbol{\theta}^{(j)})$ を最大にする $\boldsymbol{\theta}^{(j)}$ ということになる。

まず、 $\mathbf{k} - \mathbf{k}^* = O(1)$ のケースを考える。 $\sum_{i=k^{(j-1)+1}^{k^{(j)}} l_i(\boldsymbol{\theta}^{(j)})$ と $\sum_{i=k^{*(j-1)+1}^{k^{*(j)}} l_i(\boldsymbol{\theta}^{(j)})$ の差は小さいため、 $\hat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}^{(j)}$ は $\hat{\boldsymbol{\theta}}_{\mathbf{k}^*,\mathbf{x}}^{(j)}$ と同様に一致性をもつ。実際、Weiss (1986)と同じ方法で $\hat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}^{(j)} \xrightarrow{P} \boldsymbol{\theta}^{*(j)}$ は容易に示される。そこで、 $\sum_{i=k^{(j-1)+1}^{k^{(j)}} l'_i(\hat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}^{(j)}) = \mathbf{0}$ を $\hat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}^{(j)} = \boldsymbol{\theta}^{*(j)}$ のまわりでテイラー展開すると、

$$\mathbf{0} = \sum_{i=k^{(j-1)+1}^{k^{(j)}} l'_i(\boldsymbol{\theta}^{*(j)}) + \left\{ \sum_{i=k^{(j-1)+1}^{k^{(j)}} l''_i(\boldsymbol{\theta}^{*(j)}) \right\} (\hat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}^{(j)} - \boldsymbol{\theta}^{*(j)}) \{1 + o_P(1)\}$$

が得られる。ここで、 $l'_i(\boldsymbol{\theta})$ と $l''_i(\boldsymbol{\theta})$ は、 $l_i(\boldsymbol{\theta})$ の $\boldsymbol{\theta}$ に関する一階微分と二階微分を表す。これより、

$$\hat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}^{(j)} = \boldsymbol{\theta}^{*(j)} - \left\{ \sum_{i=k^{(j-1)+1}^{k^{(j)}} l''_i(\boldsymbol{\theta}^{*(j)}) \right\}^{-1} \left\{ \sum_{i=k^{(j-1)+1}^{k^{(j)}} l'_i(\boldsymbol{\theta}^{*(j)}) \right\} \{1 + o_P(1)\}$$

とも表現される。いま、例えば $k^{(j-1)} < k^{*(j-1)}$ かつ $k^{(j)} < k^{*(j)}$ とすると

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}^{(j)} - \hat{\boldsymbol{\theta}}_{\mathbf{k}^*,\mathbf{x}}^{(j)} &= \left\{ \sum_{i=k^{*(j-1)+1}^{k^{*(j)}} l''_i(\boldsymbol{\theta}^{*(j)}) \right\}^{-1} \left\{ \sum_{i=k^{*(j-1)+1}^{k^{(j)}} l'_i(\boldsymbol{\theta}^{*(j)}) + \sum_{i=k^{(j)+1}^{k^{*(j)}} l'_i(\boldsymbol{\theta}^{*(j)}) \right\} \{1 + o_P(1)\} \\ &\quad - \left\{ \sum_{i=k^{(j-1)+1}^{k^{(j)}} l''_i(\boldsymbol{\theta}^{*(j)}) \right\}^{-1} \left\{ \sum_{i=k^{(j-1)+1}^{k^{*(j-1)}} l'_i(\boldsymbol{\theta}^{*(j)}) + \sum_{i=k^{*(j-1)+1}^{k^{(j)}} l'_i(\boldsymbol{\theta}^{*(j)}) \right\} \{1 + o_P(1)\} \\ &= \left\{ \sum_{i=k^{*(j-1)+1}^{k^{*(j)}} l''_i(\boldsymbol{\theta}^{*(j)}) \right\}^{-1} O_P(1) \{1 + o_P(1)\} = O_P(n^{-1}) \end{aligned} \quad (6.4)$$

が得られ、 $k^{(j-1)} > k^{*(j-1)}$ や $k^{(j)} > k^{*(j)}$ でも同様の式が得られるので、結果 $\hat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}} - \hat{\boldsymbol{\theta}}_{\mathbf{k}^*,\mathbf{x}} = O_P(n^{-1})$ であることが分かる。そしてWeiss (1986)より $\hat{\boldsymbol{\theta}}_{\mathbf{k}^*,\mathbf{x}}^{(j)} - \boldsymbol{\theta}^{*(j)} = O_P(n^{-1/2})$ であることから、

$\widehat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}} - \boldsymbol{\theta}^* = \text{O}_P(n^{-1/2})$ が分かる.

最大対数尤度に対し, テイラー展開と (6.4) を用いると

$$\begin{aligned}
& \log f(\mathbf{x}; \mathbf{k}, \widehat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}) - \log f(\mathbf{x}; \mathbf{k}, \boldsymbol{\theta}^*) \\
&= -\frac{1}{2} \sum_{j=1}^{m+1} \sum_{i=k^{(j-1)+1}}^{k^{(j)}} (\boldsymbol{\theta}^{*(j)} - \widehat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}^{(j)})^\top l_i''(\widehat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}^{(j)}) (\boldsymbol{\theta}^{*(j)} - \widehat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}^{(j)}) + \text{O}_P(1) \\
&= -\frac{1}{2} \sum_{j=1}^{m+1} \sum_{i=k^{*(j-1)+1}}^{k^{*(j)}} (\boldsymbol{\theta}^{*(j)} - \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*,\mathbf{x}}^{(j)})^\top l_i''(\widehat{\boldsymbol{\theta}}_{\mathbf{k}^*,\mathbf{x}}^{(j)}) (\boldsymbol{\theta}^{*(j)} - \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*,\mathbf{x}}^{(j)}) + \text{O}_P(1) \\
&= \log f(\mathbf{x}; \mathbf{k}^*, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*,\mathbf{x}}) - \log f(\mathbf{x}; \mathbf{k}^*, \boldsymbol{\theta}^*) + \text{O}_P(1)
\end{aligned} \tag{6.5}$$

が得られる. よって, 両側に伸びる負のドリフト付きランダムウォークを

$$\begin{aligned}
Q_{\mathbf{k},\mathbf{x}}^{(j)} &\equiv I_{\{k < k^{*(j)}\}} \sum_{i=k+1}^{k^{*(j)}} \{\log f(x_i; \boldsymbol{\theta}^{*(j+1)} \mid \mathcal{I}_{i-1}) - \log f(x_i; \boldsymbol{\theta}^{*(j)} \mid \mathcal{I}_{i-1})\} \\
&\quad + I_{\{k > k^{*(j)}\}} \sum_{i=k^{*(j)+1}}^k \{\log f(x_i; \boldsymbol{\theta}^{*(j)} \mid \mathcal{I}_{i-1}) - \log f(x_i; \boldsymbol{\theta}^{*(j+1)} \mid \mathcal{I}_{i-1})\}
\end{aligned}$$

と定義すれば

$$\begin{aligned}
\log f(\mathbf{x}; \mathbf{k}, \widehat{\boldsymbol{\theta}}_{\mathbf{k},\mathbf{x}}) - \log f(\mathbf{x}; \mathbf{k}^*, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*,\mathbf{x}}) &= \log f(\mathbf{x}; \mathbf{k}, \boldsymbol{\theta}^*) - \log f(\mathbf{x}; \mathbf{k}^*, \boldsymbol{\theta}^*) + \text{O}_P(1) \\
&= \sum_{j=1}^m Q_{\mathbf{k}^{(j)},\mathbf{x}}^{(j)} + \text{O}_P(1)
\end{aligned} \tag{6.6}$$

が得られる. 通常の統計的漸近理論を用いれば

$$\begin{aligned}
& \log f(\mathbf{x}; \mathbf{k}^*, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*,\mathbf{x}}) - \log f(\mathbf{x}; \mathbf{k}^*, \boldsymbol{\theta}^*) \\
&= \frac{1}{2} \sum_{j=1}^{m+1} \left\{ \sum_{i=k^{*(j-1)+1}}^{k^{*(j)}} l_i'(\boldsymbol{\theta}^{*(j)}) \right\}^\top \left\{ - \sum_{i=k^{*(j-1)+1}}^{k^{*(j)}} l_i''(\boldsymbol{\theta}^{*(j)}) \right\}^{-1} \left\{ \sum_{i=k^{*(j-1)+1}}^{k^{*(j)}} l_i'(\boldsymbol{\theta}^{*(j)}) \right\} + \text{O}_P(1) \\
&= \frac{1}{2} \sum_{j=1}^{m+1} A(\boldsymbol{\theta}^{*(j)})^{-1} B(\boldsymbol{\theta}^{*(j)})^{1/2} \mathbf{u}_{\mathbf{x}}^{(j)} \mathbf{u}_{\mathbf{x}}^{(j)\top} B(\boldsymbol{\theta}^{*(j)})^{1/2} + \text{O}_P(1)
\end{aligned} \tag{6.7}$$

も得られる. ここで, $\mathbf{u}_{\mathbf{x}}^{(j)} \equiv B(\boldsymbol{\theta}^{*(j)})^{-1/2} \sum_{i=k^{*(j-1)+1}}^{k^{*(j)}} l_i'(\boldsymbol{\theta}^{*(j)}) / \sqrt{k^{*(j)} - k^{*(j-1)}}$ であり, またパラメータ $\boldsymbol{\theta}$ をもつ定常な ARCH モデルに従う $\{x_{\boldsymbol{\theta},1}, x_{\boldsymbol{\theta},2}, \dots, x_{\boldsymbol{\theta},p}\}$ から確率ベクトル $\mathbf{z}_{\boldsymbol{\theta}} =$

$(1, x_{\theta,p}^2, \dots, x_{\theta,1}^2)^\top$ を作ったとして

$$A(\theta) \equiv E \left[- \frac{\partial}{\partial \tilde{\theta} \partial \tilde{\theta}^\top} \log \left\{ \frac{1}{\sqrt{2\pi \tilde{\theta}^\top z_\theta}} \exp \left(- \frac{x_\theta^2}{2 \tilde{\theta}^\top z_\theta} \right) \right\} \Big|_{\tilde{\theta}=\theta} \right] = E \left\{ \frac{z_\theta z_\theta^\top}{2(\theta^\top z_\theta)^2} \right\}$$

かつ

$$\begin{aligned} B(\theta) &\equiv E \left[\frac{\partial}{\partial \tilde{\theta}} \log \left\{ \frac{1}{\sqrt{2\pi \tilde{\theta}^\top z_\theta}} \exp \left(- \frac{x_\theta^2}{2 \tilde{\theta}^\top z_\theta} \right) \right\} \frac{\partial}{\partial \tilde{\theta}} \log \left\{ \frac{1}{\sqrt{2\pi \tilde{\theta}^\top z_\theta}} \exp \left(- \frac{x_\theta^2}{2 \tilde{\theta}^\top z_\theta} \right) \right\}^\top \Big|_{\tilde{\theta}=\theta} \right] \\ &= E \left\{ \frac{(\varepsilon^4 - 1) z_\theta z_\theta^\top}{4(\theta^\top z_\theta)^2} \right\} \end{aligned}$$

である. $\mathbf{0}_p$ と \mathbf{I}_p をそれぞれ p 次元零ベクトルと p 次元単位行列として, $\mathbf{u}_x^{(j)}$ は漸近的に $N(\mathbf{0}_p, \mathbf{I}_p)$ に従うことに注意してほしい. (6.6) と (6.7) より, $\mathbf{k} - \mathbf{k}^* = O(1)$ のとき $\widehat{L}_{xx}(\mathbf{k}) = O_P(1)$ であることが分かる.

次に, $\mathbf{k} - \mathbf{k}^* \neq O(1)$ のケースを考える. これは, ある j が存在して $n \rightarrow \infty$ で $|k^{(j)} - k^{*(j)}| \rightarrow \infty$ となることを意味する. 以降では, 任意の $M > 0$ に対して $P\{\widehat{L}_{xx}(\mathbf{k}) > -M\} \rightarrow 0$ が成り立つことを示す. 本ケースを一般的に扱うことは本質的ではない記法の煩雑さを生むため, ここでは, ある j' に対して

$$0 < k^{*(j')} - k^{(j')} \neq O(1), \quad k^{(j)} - k^{*(j)} = O(1) \quad (j \neq j')$$

となるケースのみを詳細に扱う. このとき,

$$\begin{aligned} &\log f(\mathbf{x}; \mathbf{k}, \widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}) - \log f(\mathbf{x}; \mathbf{k}^*, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}) \\ &= \frac{1}{2} \sum_{i=k^{*(j'-1)}+1}^{k^{(j')}} \left\{ \log(\widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j')\top} z_i) + \frac{x_i^2}{\widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j')\top} z_i} - \log(\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j')\top} z_i) - \frac{x_i^2}{\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j')\top} z_i} \right\} \\ &\quad + \frac{1}{2} \sum_{i=k^{(j')}+1}^{k^{*(j')}} \left\{ \log(\widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j')\top} z_i) + \frac{x_i^2}{\widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j')\top} z_i} - \log(\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)\top} z_i) - \frac{x_i^2}{\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)\top} z_i} \right\} \\ &\quad + \frac{1}{2} \sum_{i=k^{*(j'+1)}+1}^{k^{(j'+1)}} \left\{ \log(\widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j'+1)\top} z_i) + \frac{x_i^2}{\widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j'+1)\top} z_i} - \log(\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)\top} z_i) - \frac{x_i^2}{\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)\top} z_i} \right\} + O_P(1) \end{aligned}$$

であり, ここで $C_{i, \theta^\dagger, \theta^\ddagger} = (\theta^\dagger{}^\top z_i) / (\theta^\ddagger{}^\top z_i) - \log\{(\theta^\dagger{}^\top z_i) / (\theta^\ddagger{}^\top z_i)\} - 1$ という正の関数を用意すれ

ば、これは

$$\begin{aligned}
& \log f(\mathbf{x}; \mathbf{k}, \widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}) - \log f(\mathbf{x}; \mathbf{k}^*, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}) \\
&= \frac{1}{2} \sum_{i=k^*(j')-1+1}^{k^{(j')}} \left\{ \frac{(\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j')} - \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j')})^\top \mathbf{z}_i}{(\widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j')})^\top \mathbf{z}_i (\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j')})^\top \mathbf{z}_i} (x_i^2 - \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j')\top} \mathbf{z}_i) - C_{i, \widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j')}, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j')}} \right\} \\
&+ \frac{1}{2} \sum_{i=k^{(j')}+1}^{k^*(j')} \left\{ \frac{(\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)} - \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j')})^\top \mathbf{z}_i}{(\widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j')})^\top \mathbf{z}_i (\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)})^\top \mathbf{z}_i} (x_i^2 - \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j')\top} \mathbf{z}_i) - C_{i, \widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)}, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j')}} \right\} \\
&+ \frac{1}{2} \sum_{i=k^*(j')+1}^{k^{(j'+1)}} \left\{ \frac{(\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)} - \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j'+1)})^\top \mathbf{z}_i}{(\widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j'+1)})^\top \mathbf{z}_i (\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)})^\top \mathbf{z}_i} (x_i^2 - \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j'+1)\top} \mathbf{z}_i) - C_{i, \widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)}, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j'+1)}} \right\} + \text{OP}(1) \quad (6.8)
\end{aligned}$$

と書ける. (6.4)と同様にして $\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j')} - \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j')} = \text{OP}\{n^{-1}(k^*(j') - k^{(j')})\}$ と $\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)} - \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j'+1)} = \text{OP}\{n^{-1}(k^*(j') - k^{(j')})\}$ が得られることから, この式において $C_{i, \widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j')}, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j)'}}$, $C_{i, \widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)}, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j)'}}$, 及び $C_{i, \widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)}, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j'+1)}}$ を除いたものは, 中心極限定理より $\text{OP}(\sqrt{k^*(j') - k^{(j')}})$ であることが分かる. また, (6.4)と同様にして $\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)} - \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j')} \neq \text{OP}(1)$ が得られ, $\sum_{i=k^*(j')-1+1}^{k^{(j')}} C_{i, \widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j')}, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j)'}} + \sum_{i=k^{(j')}+1}^{k^*(j')} C_{i, \widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)}, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j)'}} + \sum_{i=k^*(j')+1}^{k^{(j'+1)}} C_{i, \widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}^{(j'+1)}, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}^{(j'+1)}} \neq \text{OP}(k^*(j') - k^{(j')})$ であり, かつこれは正であることが分かる. よって,

$$P\{\log f(\mathbf{x}; \mathbf{k}, \widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}}) - \log f(\mathbf{x}; \mathbf{k}^*, \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}}) > -M\} \rightarrow 0 \quad (6.9)$$

がいえ, (6.7)と(6.9)より, $\mathbf{k} - \mathbf{k}^* \neq \text{O}(1)$ のとき $P\{\widehat{L}_{\mathbf{x}\mathbf{x}}(\mathbf{k}) > -M\} \rightarrow 0$ であることが分かる.

以上より, $K^{(j)} = \{\mathbf{k} \mid (\mathbf{k} - \mathbf{k}^*(j)) = \text{O}(1)\}$ として $K = \{\mathbf{k} \mid \mathbf{k}^{(j)} \in K^{(j)}\}$ であることが分かる.

そして, (6.6)と(6.7)より

$$\sup_{\mathbf{k} \in K} \widehat{L}_{\mathbf{x}\mathbf{x}}(\mathbf{k}) = \frac{1}{2} \sum_{j=1}^{m+1} A(\boldsymbol{\theta}^{*(j)})^{-1} B(\boldsymbol{\theta}^{*(j)})^{1/2} \mathbf{u}_{\mathbf{x}}^{(j)} \mathbf{u}_{\mathbf{x}}^{(j)\top} B(\boldsymbol{\theta}^{*(j)})^{1/2} + \sum_{j=1}^m \sup_{\mathbf{k} \in K^{(j)}} Q_{\mathbf{k}, \mathbf{x}}^{(j)} + \text{OP}(1) \quad (6.10)$$

や

$$\text{argsup}_{\mathbf{k} \in K} \widehat{L}_{\mathbf{y}\mathbf{y}}(\mathbf{k}) = \left(\text{argsup}_{\mathbf{k} \in K^{(1)}} Q_{\mathbf{k}, \mathbf{y}}^{(1)}, \text{argsup}_{\mathbf{k} \in K^{(2)}} Q_{\mathbf{k}, \mathbf{y}}^{(2)}, \dots, \text{argsup}_{\mathbf{k} \in K^{(m)}} Q_{\mathbf{k}, \mathbf{y}}^{(m)} \right)^\top + \text{OP}(1) \quad (6.11)$$

が言える. (6.11)の右辺の主項は以降 $\tilde{\mathbf{k}}_{\mathbf{y}} = (\tilde{k}_{\mathbf{y}}^{(1)}, \tilde{k}_{\mathbf{y}}^{(2)}, \dots, \tilde{k}_{\mathbf{y}}^{(m)})^\top$ と記すことにする. ここで, $\widehat{L}_{\mathbf{x}\mathbf{y}}$

に対してもテイラー展開を用いると、

$$\begin{aligned} & \log f(\mathbf{x}; \tilde{\mathbf{k}}_{\mathbf{y}}, \boldsymbol{\theta}^*) - \log f(\mathbf{x}; \tilde{\mathbf{k}}_{\mathbf{y}}, \widehat{\boldsymbol{\theta}}_{\tilde{\mathbf{k}}_{\mathbf{y}}, \mathbf{y}}) \\ &= \sum_{j=1}^{m+1} \left[-(\widehat{\boldsymbol{\theta}}_{\tilde{\mathbf{k}}_{\mathbf{y}}, \mathbf{y}}^{(j)} - \boldsymbol{\theta}^{*(j)})^\top \left\{ \sum_{i=\tilde{k}_{\mathbf{y}}^{(j-1)}+1}^{\tilde{k}_{\mathbf{y}}^{(j)}} l'_i(\boldsymbol{\theta}^{*(j)}) \right\} \right. \\ & \quad \left. - \frac{1}{2}(\widehat{\boldsymbol{\theta}}_{\tilde{\mathbf{k}}_{\mathbf{y}}, \mathbf{y}}^{(j)} - \boldsymbol{\theta}^{*(j)})^\top \left\{ \sum_{i=\tilde{k}_{\mathbf{y}}^{(j-1)}+1}^{\tilde{k}_{\mathbf{y}}^{(j)}} l''_i(\boldsymbol{\theta}^{*(j)}) \right\} (\widehat{\boldsymbol{\theta}}_{\tilde{\mathbf{k}}_{\mathbf{y}}, \mathbf{y}}^{(j)} - \boldsymbol{\theta}^{*(j)}) \right] + o_P(1) \end{aligned}$$

であり、 $\widehat{\boldsymbol{\theta}}_{\tilde{\mathbf{k}}_{\mathbf{y}}, \mathbf{y}}^{(j)} - \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{y}}^{(j)} = O_P(n^{-1})$ と $\widehat{\boldsymbol{\theta}}_{\tilde{\mathbf{k}}_{\mathbf{y}}, \mathbf{y}}^{(j)} - \boldsymbol{\theta}^{*(j)} = O_P(n^{-1/2})$ であることを用いると、これは

$$\begin{aligned} & - \sum_{j=1}^{m+1} (\widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{y}}^{(j)} - \boldsymbol{\theta}^{*(j)})^\top \left\{ \sum_{i=k^{*(j-1)}+1}^{k^{*(j)}} l'_i(\boldsymbol{\theta}^{*(j)}) \right\} \\ & - \frac{1}{2} \sum_{j=1}^{m+1} (\widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{y}}^{(j)} - \boldsymbol{\theta}^{*(j)})^\top \left\{ \sum_{i=k^{*(j-1)}+1}^{k^{*(j)}} l''_i(\boldsymbol{\theta}^{*(j)}) \right\} (\widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{y}}^{(j)} - \boldsymbol{\theta}^{*(j)}) + o_P(1) \end{aligned}$$

と書ける。そして (6.7) と同様に考えれば、結果

$$\begin{aligned} & \log f(\mathbf{x}; \tilde{\mathbf{k}}_{\mathbf{y}}, \boldsymbol{\theta}^*) - \log f(\mathbf{x}; \tilde{\mathbf{k}}_{\mathbf{y}}, \widehat{\boldsymbol{\theta}}_{\tilde{\mathbf{k}}_{\mathbf{y}}, \mathbf{y}}) \\ &= - \sum_{j=1}^{m+1} \mathbf{u}_{\mathbf{y}}^{(j)\top} B(\boldsymbol{\theta}^{*(j)})^{1/2} A(\boldsymbol{\theta}^{*(j)})^{-1} B(\boldsymbol{\theta}^{*(j)})^{1/2} \mathbf{u}_{\mathbf{x}}^{(j)} \\ & \quad - \frac{1}{2} \sum_{j=1}^{m+1} \mathbf{u}_{\mathbf{y}}^{(j)\top} B(\boldsymbol{\theta}^{*(j)})^{1/2} A(\boldsymbol{\theta}^{*(j)})^{-1} B(\boldsymbol{\theta}^{*(j)})^{1/2} \mathbf{u}_{\mathbf{y}}^{(j)} + o_P(1) \end{aligned} \quad (6.12)$$

が得られる。ここで、さらに

$$\begin{aligned} & \log f(\mathbf{x}; \mathbf{k}^*, \boldsymbol{\theta}^*) - \log f(\mathbf{x}; \tilde{\mathbf{k}}_{\mathbf{y}}, \boldsymbol{\theta}^*) \\ &= \sum_{j=1}^m \left[I_{\{\tilde{k}_{\mathbf{y}}^{(j)} < k^{*(j)}\}} \sum_{i=\tilde{k}_{\mathbf{y}}^{(j)}+1}^{k^{*(j)}} \{ \log f(x_i; \boldsymbol{\theta}^{*(j)} \mid \mathcal{I}_{i-1}) - \log f(x_i; \boldsymbol{\theta}^{*(j+1)} \mid \mathcal{I}_{i-1}) \} \right. \\ & \quad \left. + I_{\{\tilde{k}_{\mathbf{y}}^{(j)} > k^{*(j)}\}} \sum_{i=k^{*(j)}+1}^{\tilde{k}_{\mathbf{y}}^{(j)}} \{ \log f(x_i; \boldsymbol{\theta}^{*(j+1)} \mid \mathcal{I}_{i-1}) - \log f(x_i; \boldsymbol{\theta}^{*(j)} \mid \mathcal{I}_{i-1}) \} \right] \\ &= \sum_{j=1}^m Q_{\tilde{k}_{\mathbf{y}}^{(j)}, \mathbf{x}}^{(j)} \end{aligned} \quad (6.13)$$

を書いておく. (6.11), (6.12), 及び (6.13) より,

$$\begin{aligned}
\widehat{L}_{\mathbf{x}\mathbf{y}} \left\{ \operatorname{argsup}_{\mathbf{k} \in K} \widehat{L}_{\mathbf{y}\mathbf{y}}(\mathbf{k}) \right\} &= \widehat{L}_{\mathbf{x}\mathbf{y}}(\tilde{\mathbf{k}}_{\mathbf{y}}) + o_{\mathbb{P}}(1) \\
&= - \sum_{j=1}^m Q_{\tilde{\mathbf{k}}_{\mathbf{y}}^{(j)}, \mathbf{x}}^{(j)} + \sum_{j=1}^{m+1} \mathbf{u}_{\mathbf{y}}^{(j)\top} B(\boldsymbol{\theta}^{*(j)})^{1/2} A(\boldsymbol{\theta}^{*(j)})^{-1} B(\boldsymbol{\theta}^{*(j)})^{1/2} \mathbf{u}_{\mathbf{x}}^{(j)} \\
&\quad + \frac{1}{2} \sum_{j=1}^{m+1} \mathbf{u}_{\mathbf{y}}^{(j)\top} B(\boldsymbol{\theta}^{*(j)})^{1/2} A(\boldsymbol{\theta}^{*(j)})^{-1} B(\boldsymbol{\theta}^{*(j)})^{1/2} \mathbf{u}_{\mathbf{y}}^{(j)} + o_{\mathbb{P}}(1) \tag{6.14}
\end{aligned}$$

となることが分かる. $\mathbf{u}_{\mathbf{x}}^{(j)}$ と $\mathbf{u}_{\mathbf{y}}^{(j)}$ が漸近的に $N(\mathbf{0}_p, \mathbf{I}_p)$ に独立に従うことも用いれば, (6.10) と (6.14) より, 以下の定理が得られる.

定理 3. (6.1) と (6.2) に従って与えられる \mathbf{x} に対し, (6.3) で考えている漸近バイアスは

$$\mathbb{E}\{b(\mathbf{k}^*, \boldsymbol{\theta}^*)\} = \sum_{j=1}^m \mathbb{E} \left(\sup_{\mathbf{k} \in K^{(j)}} Q_{\mathbf{k}, \mathbf{x}}^{(j)} + Q_{\operatorname{argsup}_{\mathbf{k} \in K^{(j)}} Q_{\mathbf{k}, \mathbf{y}}^{(j)}, \mathbf{x}}^{(j)} \right) + \sum_{j=1}^{m+1} \operatorname{tr}\{A(\boldsymbol{\theta}^{*(j)})^{-1} B(\boldsymbol{\theta}^{*(j)})\} \tag{6.15}$$

で与えられる.

この定理より, $\mathbb{E}(\sup_{\mathbf{k} \in K^{(j)}} Q_{\mathbf{k}, \mathbf{x}}^{(j)} + Q_{\operatorname{argsup}_{\mathbf{k} \in K^{(j)}} Q_{\mathbf{k}, \mathbf{y}}^{(j)}, \mathbf{x}}^{(j)})$ が変化点パラメータである $k^{(j)}$ によるバイアス, $\operatorname{tr}\{A(\boldsymbol{\theta}^{*(j)})^{-1} B(\boldsymbol{\theta}^{*(j)})\}$ が変化点以外のパラメータである $\boldsymbol{\theta}^{(j)}$ によるバイアスとみなすことができる.

本章でも変化点をもつ Cox 比例ハザードモデルに対する情報量規準の導出時と同様の理由で, $\boldsymbol{\theta}^{*(j+1)} - \boldsymbol{\theta}^{*(j)}$ が定数であるとは仮定せず, Csörgő and Horváth (1997) の Section 1.5 と同様にして

$$\boldsymbol{\theta}^{*(j+1)} - \boldsymbol{\theta}^{*(j)} = \alpha_n^{-1/2} \Delta^{(j)} \quad (1 \leq j \leq m), \quad O(1) \neq \alpha_n = o(n) \tag{6.16}$$

という条件を仮定する. ここで $\Delta^{(j)}$ は定ベクトルである. この条件の下, 変化点推定量の漸近挙動を調べる. まず, 定理 3 の導出と同様にして, $\mathbf{k} - \mathbf{k}^* = O(\alpha_n)$ のとき, (6.4) の代わりに $\widehat{\boldsymbol{\theta}}_{\mathbf{k}, \mathbf{x}} - \widehat{\boldsymbol{\theta}}_{\mathbf{k}^*, \mathbf{x}} = O_{\mathbb{P}}(n^{-1} \alpha_n^{1/2})$ が得られる. そして, (6.5), (6.6), (6.7) が言えるので $\widehat{L}_{\mathbf{x}\mathbf{x}}(\mathbf{k}) = O_{\mathbb{P}}(1)$ が成立する. 一方 $\mathbf{k} - \mathbf{k}^* \neq O(\alpha_n)$ のとき, やはり定理 3 の導出と同様に, $0 < k^{*(j')} - k^{(j')} \neq O(\alpha_n)$ かつ $j \neq j'$ に対して $k^{(j)} - k^{*(j)} = O(\alpha_n)$ というケースを考える. このとき, やはり (6.8) を経由して (6.9) が言える. そして (6.7) を併せれば $\mathbb{P}\{\widehat{L}_{\mathbf{x}\mathbf{x}}(\mathbf{k}) > -M\} \rightarrow 0$ が成立する. 以上より,

$K = \{\mathbf{k} \mid \mathbf{k} - \mathbf{k}^* = O(\alpha_n)\}$ とできて, (6.10), (6.11), (6.12), (6.13), (6.14) が成り立つことが分かる.

これより, 以降 $\mathbf{k} - \mathbf{k}^* = O(\alpha_n)$ とする. 条件 (6.16) の下, $l_i(\boldsymbol{\theta}^{*(j+1)})$ は $\boldsymbol{\theta}^{*(j)}$ の周りでテイラー展開すると

$$\begin{aligned} & \log f(x_i; \boldsymbol{\theta}^{*(j+1)} \mid \mathcal{I}_{i-1}) - \log f(x_i; \boldsymbol{\theta}^{*(j)} \mid \mathcal{I}_{i-1}) \\ &= \alpha_n^{-1/2} \Delta^{(j)\top} l'_i(\boldsymbol{\theta}^{*(j)}) + \frac{1}{2} \alpha_n^{-1} \Delta^{(j)\top} l''_i(\boldsymbol{\theta}^{*(j)}) \Delta^{(j)} + o_P(1) \end{aligned}$$

となることより, $Q_{\mathbf{k}^* + [s\alpha_n], \mathbf{x}}^{(j)}$ は

$$\begin{aligned} & I_{\{s < 0\}} \left\{ \alpha_n^{-1/2} \sum_{i=k^*(j) + [s\alpha_n] + 1}^{k^*(j)} \Delta^{(j)\top} l'_i(\boldsymbol{\theta}^{*(j)}) + \frac{1}{2} \alpha_n^{-1} \sum_{i=k^*(j) + [s\alpha_n] + 1}^{k^*(j)} \Delta^{(j)\top} l''_i(\boldsymbol{\theta}^{*(j)}) \Delta^{(j)} \right\} \\ & + I_{\{s > 0\}} \left\{ -\alpha_n^{-1/2} \sum_{i=k^*(j) + 1}^{k^*(j) + [s\alpha_n]} \Delta^{(j)\top} l'_i(\boldsymbol{\theta}^{*(j+1)}) + \frac{1}{2} \alpha_n^{-1} \sum_{i=k^*(j) + 1}^{k^*(j) + [s\alpha_n]} \Delta^{(j)\top} l''_i(\boldsymbol{\theta}^{*(j+1)}) \Delta^{(j)} \right\} + o_P(1) \end{aligned}$$

と表現される. いま, $\{W_s\}_{s \in \mathbb{R}}$ を, $E(W_s) = 0$ かつ $V(W_s) = |s|$ である両側標準ブラウン運動とし,

$$\begin{aligned} & \alpha_n^{-1/2} \sum_{i=k^*(j) + [s\alpha_n] + 1}^{k^*(j)} \Delta^{(j)\top} l'_i(\boldsymbol{\theta}^{*(j)}) \xrightarrow{d} \sigma_{1, \boldsymbol{\theta}^*}^{(j)} W_s \equiv \{\Delta^{(j)\top} B(\boldsymbol{\theta}^{*(j)}) \Delta^{(j)}\}^{1/2} W_s, \\ & \frac{1}{2} \alpha_n^{-1} \sum_{i=k^*(j) + [s\alpha_n] + 1}^{k^*(j)} \Delta^{(j)\top} l''_i(\boldsymbol{\theta}^{*(j)}) \Delta^{(j)} \xrightarrow{p} -c_{1, \boldsymbol{\theta}^*}^{(j)} |s| \equiv -\frac{1}{2} \Delta^{(j)\top} A(\boldsymbol{\theta}^{*(j)}) \Delta^{(j)} |s|, \\ & -\alpha_n^{-1/2} \sum_{i=k^*(j) + 1}^{k^*(j) + [s\alpha_n]} \Delta^{(j)\top} l'_i(\boldsymbol{\theta}^{*(j+1)}) \xrightarrow{d} \sigma_{2, \boldsymbol{\theta}^*}^{(j)} W_s \equiv \{\Delta^{(j)\top} B(\boldsymbol{\theta}^{*(j+1)}) \Delta^{(j)}\}^{1/2} W_s, \\ & \frac{1}{2} \alpha_n^{-1} \sum_{i=k^*(j) + 1}^{k^*(j) + [s\alpha_n]} \Delta^{(j)\top} l''_i(\boldsymbol{\theta}^{*(j+1)}) \Delta^{(j)} \xrightarrow{p} -c_{2, \boldsymbol{\theta}^*}^{(j)} |s| \equiv -\frac{1}{2} \Delta^{(j)\top} A(\boldsymbol{\theta}^{*(j+1)}) \Delta^{(j)} |s| \end{aligned}$$

と書けることを用いると,

$$Q_{\mathbf{k}^* + [s\alpha_n], \mathbf{x}}^{(j)} \xrightarrow{d} V_s(c_{1, \boldsymbol{\theta}^*}^{(j)}, c_{2, \boldsymbol{\theta}^*}^{(j)}, \sigma_{1, \boldsymbol{\theta}^*}^{(j)}, \sigma_{2, \boldsymbol{\theta}^*}^{(j)})$$

が成立する. ここで, $V_s(c_1, c_2, \sigma_1, \sigma_2)$ は, c_1 や c_2 をドリフト係数, σ_1 や σ_2 を拡散係数とした,

両側に伸びるブラウン運動であり、 $s \leq 0$ のとき $-c_1|s| + \sigma_1 W_s$ 、 $s \geq 0$ のとき $-c_2|s| + \sigma_2 W_s$ で定義されるものである。この収束の結果として

$$\sup_{k \in K^{(j)}} Q_{k, \mathbf{x}}^{(j)} \xrightarrow{d} \sup_{s \in \mathbb{R}} V_s(c_{1, \theta^*}^{(j)}, c_{2, \theta^*}^{(j)}, \sigma_{1, \theta^*}^{(j)}, \sigma_{2, \theta^*}^{(j)})$$

と

$$Q_{\text{argsup}_{k \in K^{(j)}} Q_{k, \mathbf{y}}^{(j)}}^{(j)} \xrightarrow{d} V_{\text{argsup}_{s \in \mathbb{R}} U_s(c_{1, \theta^*}^{(j)}, c_{2, \theta^*}^{(j)}, \sigma_{1, \theta^*}^{(j)}, \sigma_{2, \theta^*}^{(j)})}(c_{1, \theta^*}^{(j)}, c_{2, \theta^*}^{(j)}, \sigma_{1, \theta^*}^{(j)}, \sigma_{2, \theta^*}^{(j)})$$

が得られる。ここで、 U_s は V_s のコピーである。

これらの期待値を評価するため、[Bhattacharya and Brockwell \(1976\)](#) や [Shepp \(1979\)](#) の結果を用いていく。まず、正の定数 a と c に対して成り立っている

$$\mathbb{P}\left\{\sup_{s>0}(W_s - cs) > a\right\} = \exp(-2ac)$$

という等式を使うと

$$\begin{aligned} & \mathbb{E}\left\{\sup_{s \in \mathbb{R}} V_s(c_1, c_2, \sigma_1, \sigma_2)\right\} \\ &= \int_0^\infty \mathbb{P}\left\{\sup_{s \in \mathbb{R}} V_s(c_1, c_2, \sigma_1, \sigma_2) > a\right\} da \\ &= \int_0^\infty \left[\exp\left(-\frac{2c_1 a}{\sigma_1^2}\right) + \exp\left(-\frac{2c_2 a}{\sigma_2^2}\right) - \exp\left\{-\frac{2(c_1 \sigma_2^2 + c_2 \sigma_1^2)a}{\sigma_1^2 \sigma_2^2}\right\} \right] da \\ &= \frac{c_1^2 \sigma_2^4 + c_1 c_2 \sigma_1^2 \sigma_2^2 + c_2^2 \sigma_1^4}{2c_1 c_2 (c_1 \sigma_2^2 + c_2 \sigma_1^2)} \end{aligned} \quad (6.17)$$

が得られる。次に、 $\text{argsup}_{s \in \mathbb{R}} V_s(c_1, c_2, \sigma_1, \sigma_2)$ の確率密度関数が $s \leq 0$ のとき $g(-s | c_1/\sigma_1, c_2\sigma_1/\sigma_2^2)$ 、 $s \geq 0$ のとき $g(s | c_2/\sigma_2, c_1\sigma_2/\sigma_1^2)$ であることを使う。ここで、

$$g(s | a_1, a_2) \equiv 2a_1(a_1 + 2a_2) \exp\{2a_2(a_1 + a_2)s\} \Phi\{-(a_1 + 2a_2)\sqrt{s}\} - 2a_1^2 \Phi(-a_1\sqrt{s})$$

である。すると、

$$\mathbb{E}\{V_{\text{argsup}_{s \in \mathbb{R}} U_s(c_1, c_2, \sigma_1, \sigma_2)}(c_1, c_2, \sigma_1, \sigma_2)\}$$

$$\begin{aligned}
&= \int_0^\infty sg\left(s \mid \frac{c_1}{\sigma_1}, \frac{c_2\sigma_1}{\sigma_2^2}\right) c_1 ds + \int_0^\infty sg\left(s \mid \frac{c_2}{\sigma_2}, \frac{c_1\sigma_2}{\sigma_1^2}\right) c_2 ds \\
&= \frac{c_2(2c_1\sigma_2^2 + c_2\sigma_1^2)\sigma_1^4}{2c_1(c_1\sigma_2^2 + c_2\sigma_1^2)^2} + \frac{c_1(2c_2\sigma_1^2 + c_1\sigma_2^2)\sigma_2^4}{2c_2(c_1\sigma_2^2 + c_2\sigma_1^2)^2} \\
&= \frac{c_1^2\sigma_2^4 + c_1c_2\sigma_1^2\sigma_2^2 + c_2^2\sigma_1^4}{2c_1c_2(c_1\sigma_2^2 + c_2\sigma_1^2)} \tag{6.18}
\end{aligned}$$

が得られる。(6.17)と(6.18)において $(c_1, c_2, \sigma_1, \sigma_2) = (c_{1,\theta^*}^{(j)}, c_{2,\theta^*}^{(j)}, \sigma_{1,\theta^*}^{(j)}, \sigma_{2,\theta^*}^{(j)})$ と考えれば、次の定理が得られる。

定理 4. 定理 3 の条件に加え、(6.16) が満たされているとき、(6.3) の漸近バイアスは

$$E\{b(\mathbf{k}^*, \theta^*)\} = \sum_{j=1}^m \frac{c_{1,\theta^*}^{(j)2}\sigma_{2,\theta^*}^{(j)4} + c_{1,\theta^*}^{(j)}c_{2,\theta^*}^{(j)}\sigma_{1,\theta^*}^{(j)2}\sigma_{2,\theta^*}^{(j)2} + c_{2,\theta^*}^{(j)2}\sigma_{1,\theta^*}^{(j)4}}{c_{1,\theta^*}^{(j)}c_{2,\theta^*}^{(j)}(c_{1,\theta^*}^{(j)}\sigma_{2,\theta^*}^{(j)2} + c_{2,\theta^*}^{(j)}\sigma_{1,\theta^*}^{(j)2})} + \sum_{j=1}^{m+1} \text{tr}\{A(\theta^{*(j)})^{-1}B(\theta^{*(j)})\}$$

で与えられる。

この定理より、最大対数尤度をバイアス補正したものとしての情報量規準が与えられるわけだが、この漸近バイアスには未知のパラメータが含まれているため、TIC や GIC と同じように、それを一致推定量で置き換える。結果、以下の情報量規準が得られ、これを変化点をもつ ARCH モデルの TIC と呼ぶことにする：

$$\begin{aligned}
\text{TIC} &= -2 \log f(\mathbf{x}; \hat{\mathbf{k}}_x, \hat{\theta}_x) \\
&+ 2 \sum_{j=1}^m \frac{c_{1,\hat{\theta}}^{(j)2}\sigma_{2,\hat{\theta}}^{(j)4} + c_{1,\hat{\theta}}^{(j)}c_{2,\hat{\theta}}^{(j)}\sigma_{1,\hat{\theta}}^{(j)2}\sigma_{2,\hat{\theta}}^{(j)2} + c_{2,\hat{\theta}}^{(j)2}\sigma_{1,\hat{\theta}}^{(j)4}}{c_{1,\hat{\theta}}^{(j)}c_{2,\hat{\theta}}^{(j)}(c_{1,\hat{\theta}}^{(j)}\sigma_{2,\hat{\theta}}^{(j)2} + c_{2,\hat{\theta}}^{(j)}\sigma_{1,\hat{\theta}}^{(j)2})} + \sum_{j=1}^{m+1} \text{tr}\{A(\hat{\theta}^{(j)})^{-1}B(\hat{\theta}^{(j)})\} \tag{6.19}
\end{aligned}$$

ここで、 $(c_{1,\hat{\theta}}^{(j)}, c_{2,\hat{\theta}}^{(j)}, \sigma_{1,\hat{\theta}}^{(j)}, \sigma_{2,\hat{\theta}}^{(j)})$ は $(c_{1,\theta^*}^{(j)}, c_{2,\theta^*}^{(j)}, \sigma_{1,\theta^*}^{(j)}, \sigma_{2,\theta^*}^{(j)})$ の θ^* を $\hat{\theta}$ に置き換えたものである。

続いて、 x_i の条件付き分布が真に正規分布である、つまりモデル誤特定はないと仮定し、ここまでの結果から情報量規準 AIC も導出しておく。この仮定の下では、 $E(\varepsilon_i^4) = 3$ であり、よって $A(\theta^{*(j)}) = B(\theta^{*(j)})$ であり、そして $2c_{1,\theta^*}^{(j)} = \sigma_{1,\theta^*}^{(j)2}$ かつ $2c_{2,\theta^*}^{(j)} = \sigma_{2,\theta^*}^{(j)2}$ であることが容易に確かめられる。つまり、定理 4 で導いた漸近バイアスにおいて、第一項目は分子と分母がキャンセルして定数となるし、第二項目も行列がキャンセルして単位行列となる。そして次の系が得られる。

系 4. 定理 4 と同じ条件の下、さらにモデルの誤特定がないとすると、(6.3) の漸近バイアスは

$$E\{b(\mathbf{k}^*, \theta^*)\} = 3m + p(m + 1)$$

で与えられる.

この系より, 変化点パラメータによる漸近バイアスは, その他の通常のパラメータによる漸近バイアスの三倍であることが分かる. 結果, 以下の情報量規準が得られ, これを変化点をもつ ARCH モデルの AIC と呼ぶことにする:

$$\text{AIC} = -2 \log f(\mathbf{x}; \hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x) + 6m + 2p(m+1) \quad (6.20)$$

この AIC は [Ninomiya \(2015\)](#) の結果と整合する.

6.4 シミュレーション実験

6.4.1 モデル誤特定がない下での情報量規準

本節では, 誤特定を許容する情報量規準が機能するかを調べる前に, そもそも変化点をもつ ARCH モデルの情報量規準が機能するかを調べるため, (6.20) の AIC の性能を検証する. それとの比較として, [Jones and Dey \(1995\)](#) 等で用いられている, 変化点パラメータも通常のパラメータのよりに扱った規準も考える. つまりそれは, 形式的に罰則項を使うというものであり

$$\text{AIC}_{\text{naive}} = -2 \log f(\mathbf{x}; \hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x) + 2m + 2p(m+1) \quad (6.21)$$

と書くことにする. また, 規準としての目的は違うものの, [Schwarz \(1978\)](#) に基づいて罰則項を $m \log n + p(m+1) \log n$ とした BIC ([Yao 1988](#)) も比較対象とする.

まず, 最もシンプルな設定での検証を見据え, 1 個の変化点 k をもつ 1 次の ARCH モデル

$$\begin{aligned} x_i = \sqrt{h_i} \varepsilon_i, \quad & h_i = \theta_0^{(1)} + \theta_1^{(1)} x_{i-1}^2 \quad (i \in \{1, 2, \dots, k\}) \\ & h_i = \theta_0^{(2)} + \theta_1^{(2)} x_{i-1}^2 \quad (i \in \{k+1, k+2, \dots, n\}) \end{aligned} \quad (6.22)$$

を考える. また ε_i は独立に標準正規分布に従うとし, つまり \mathcal{I}_{i-1} が与えられたもとで x_i は $N(0, h_i)$ に従うとする. このモデルを用いる以降の数値実験では, 真の変化点を $k^* = n/2$ とする. また, 定常性を保つため, 真の係数パラメータは $\theta_0^{*(1)}, \theta_0^{*(2)} > 0$ かつ $0 \leq \theta_1^{*(1)}, \theta_1^{*(2)} \leq 1/\sqrt{3} \approx 0.577$ を満たすとする. モデルは, 変化点パラメータ 1 つと係数パラメータ 4 つをもつので, (6.20) の AIC

と(6.21)の AIC_{naive} の罰則項は、それぞれ $6 \times 1 + 2 \times 4 = 14$ と $2 \times 1 + 2 \times 4 = 10$ になる。これらの罰則項が最大対数尤度のバイアスに対して正確な近似を与えているかどうかを調べるため、(6.22)において、パラメータの真値とデータサイズを変えながらバイアスを数値的に評価したのが表6.1である。いずれの設定においても、値は14付近にあり、少なくとも10よりは14の方が正確な近似となっていることが確認できる。

表6.2では、AICによるモデル選択の特徴を掴むため、やはり(6.22)に従ってデータを発生させ、変化点が0個もしくは1個のどちらを選択するかの確率を評価している。まず、変化点は存在するがデータサイズは小さい、というケースを見る。 AIC_{naive} に比べ、AICが変化点有りのモデルを選択する確率は高くないが、データが十分でないときはどちらのモデルが真に近いかは分からない。AICの目的は真の構造に近い推定を与えることであり、本当は変化があったとしても、過適合するよりは変化のないモデルを選択することが良いことが多々ある。つまり、この値から、どちらの規準が良いかを言うことはできない。変化点が存在してデータサイズも大きい、というケースを見てみると、 AIC_{naive} もAICも高い確率で変化点有りのモデルを選択していることが分かる。次に、変化点が存在しないケースを見る。このケースでは、規準がどのような目的であっても、変化点無しのモデルを選択してほしい。しかし、 AIC_{naive} は変化点有りのモデルを選びすぎており、 $n = 400$ のときは50%近くに及んでいることが観察される。 AIC_{naive} は実際のバイアスを相当過小評価している、ということの負の側面が現れているものと推察できる。表6.3では、実際に AIC_{naive} とAICとBICの性能を比較することを目的とし、

$$x_i = \sqrt{h_i} \varepsilon_i, \quad h_i = \theta_0^{(j)} + \theta_1^{(j)} x_{i-1}^2 \quad (i \in \{k^{(j-1)} + 1, k^{(j-1)} + 2, \dots, k^{(j)}\}, j \in \{1, 2, \dots, m+1\}) \quad (6.23)$$

に従う変化数 m のモデルを $m = 0, m = 1, m = 2, m = 3, m = 4$ と考えて選択する。ここで、 $k^{(0)} = 0$ かつ $k^{(m+1)} = n$ であり、 ε_i は標準正規分布に独立に従うとする。真の変化点数は2とし、真の変化点 $k^{*(j)}$ と真の変化量 $\theta_0^{*(j)}/\theta_0^{*(j-1)}, \theta_1^{*(j)}/\theta_1^{*(j-1)}$ は次のように与える ($1 \leq j \leq 2$): $k^{*(1)}$ と $k^{*(2)} - k^{*(1)}$ と $n - k^{*(2)}$ が $0.1n$ 以上であるという制約の下、 $k^{*(j)}$ は一様分布にしたがわせる; $v_{1,j}$ と $v_{2,j}$ をそれぞれ離散型一様分布 $U(\{-1, 1\})$ と連続型一様分布 $U(0, 1)$ に従う独立な確率変数とし、 $\theta_0^{*(j)}/\theta_0^{*(j-1)}$ と $\theta_1^{*(j)}/\theta_1^{*(j-1)}$ は $2^{v_{1,j}(\xi^* + v_{2,j})}$ で与える。こうして決められた真の構造に対してモデルの選択確率を求め、さらに真の分布と推定された分布間のK-Lを評価したのが表

6.3である。導出した AIC は、すべての設定において、他の情報量規準よりも小さい K-L を与えており、他より優れていると言える。特にデータサイズが大きいとき、 AIC_{naive} は変化点を多く選びすぎてしまう傾向がある。一方、BIC は変化点を少なく選びすぎている傾向があり、それゆえにやはり K-L の値が大きくなっている。

6.4.2 モデル誤特定を許容する情報量規準

本節では、誤特定を許容する (6.19) の TIC の性能を検証する。それとの比較として、 AIC_{naive} , (6.20) の AIC, 及び BIC を考える。まずシンプルな設定での検証を見据え、1 個の変化点 k をもつ 1 次の ARCH モデル (6.22) を考える。ただし、モデルが誤特定された状況を扱うため、前節のように ε_i は標準正規分布に従わないとする。具体的には、混合比 $1 - \rho$ で標準正規分布、混合比 ρ で区間 $[-\sqrt{3}, \sqrt{3}]$ の連続型一様分布をもつ、混合分布に ε_i は従うとする。 ρ が大きくなるほど、誤特定の度合いが強くなる設定となる。このモデルを用いる以降の数値実験では、真の変化点を $k^* = n/2$ とする。

表 6.4 では、TIC の罰則項が最大対数尤度のバイアスに対して正確な近似を与えているかどうかを調べている。(6.22) において、パラメータの真値とデータサイズ、混合比を変えながらバイアスを数値的に評価している。混合比 ρ が大きくなるとバイアスが小さくなる傾向があるが、TIC の漸近評価はそれを捉えることができおり、全体的に AIC の罰則項である 14 よりもさらに正確な近似となっていることが確認できる。

表 6.5 では、実際に AIC_{naive} , AIC, TIC, 及び BIC の性能を比較することを目的とし、(6.23) に従う変化数 m のモデルを $m = 0, m = 1, m = 2, m = 3, m = 4$ と考えて選択する。真の変化点数はやはり 2 とし、真の変化点と真の変化量の与え方も 6.4.1 節と同じとする。ただし ε_i は、混合比 $1 - \rho$ で標準正規分布、混合比 ρ で区間 $[-\sqrt{3}, \sqrt{3}]$ の連続型一様分布をもつ、混合分布に従うとする。表 6.5 は、真の分布と推定された分布間の K-L を主指標、各変化点数のモデル選択確率を参考指標として、それらを実験したものである。導出した TIC は、すべての設定において、他の情報量規準よりも小さい K-L を与えており、他より優れていると言える。強調したいのは、誤特定があるときは、しっかりと AIC より性能がよく、真のモデルの選択確率も高いことである。TIC には真値を推定量で置き換える操作があり、不安定さが加わるがゆえに、それは一般に必ずしも AIC を優越しないが、本設定では誤特定考慮の効果が強いと推察される。データサイズ n が

小さく混合比 ρ が大きいとき、 AIC_{naive} が悪くないように見える。これは、表 6.4 から分かるように、 ρ が大きくなるとバイアスが小さくなり、たまたま AIC_{naive} の罰則に近くなるからである。それ以外の設定では、やはり AIC_{naive} は大きく性能を落とす。

表 6.1: モデル誤特定がないときの最大対数尤度のバイアス

		$n: 50$	$n: 100$	$n: 200$	$n: 400$
$\theta_1^{*(1)}: 0.1$	$\theta_1^{*(2)}: 0.2$	13.15	13.97	13.96	13.94
	$\theta_1^{*(2)}: 0.3$	13.44	13.86	14.39	14.02
	$\theta_1^{*(2)}: 0.4$	14.15	14.35	14.06	14.01
	$\theta_1^{*(2)}: 0.5$	14.20	14.21	14.04	13.98
$\theta_1^{*(1)}: 0.2$	$\theta_1^{*(2)}: 0.1$	13.13	13.91	13.81	14.19
	$\theta_1^{*(2)}: 0.3$	13.97	14.78	15.17	14.52
	$\theta_1^{*(2)}: 0.4$	14.04	14.89	14.95	14.86
	$\theta_1^{*(2)}: 0.5$	13.92	15.06	14.61	14.46
$\theta_1^{*(1)}: 0.3$	$\theta_1^{*(2)}: 0.1$	13.36	14.05	14.47	14.74
	$\theta_1^{*(2)}: 0.2$	14.05	14.36	15.03	14.81
	$\theta_1^{*(2)}: 0.4$	14.10	15.12	15.12	14.75
	$\theta_1^{*(2)}: 0.5$	14.25	14.68	15.37	14.86
$\theta_1^{*(1)}: 0.4$	$\theta_1^{*(2)}: 0.1$	13.84	14.39	14.31	14.50
	$\theta_1^{*(2)}: 0.2$	14.09	14.70	15.19	15.18
	$\theta_1^{*(2)}: 0.3$	14.25	15.28	15.72	15.24
	$\theta_1^{*(2)}: 0.5$	14.78	15.70	15.60	14.94
$\theta_1^{*(1)}: 0.5$	$\theta_1^{*(2)}: 0.1$	13.94	14.53	14.60	14.38
	$\theta_1^{*(2)}: 0.2$	14.12	14.76	15.62	15.52
	$\theta_1^{*(2)}: 0.3$	14.40	15.60	15.70	15.37
	$\theta_1^{*(2)}: 0.4$	14.52	15.91	16.00	15.05

$(\theta_0^{*(1)}, \theta_0^{*(2)})$ はすべて $(1.0, 2.0)$ としている. これらの値は, (6.22) のモデルに基づき, 1,000 回の繰り返しによるモンテカルロ法で評価したものである. つまり, $\{\mathbf{x}^{[h]} = (x_1^{[h]}, x_2^{[h]}, \dots, x_n^{[h]}) \mid 1 \leq h \leq 2000\}$ を発生させ, $\sum_{h=1}^{1000} [\sup_{(\mathbf{k}, \boldsymbol{\theta})} L_{\mathbf{x}^{[2h-1]}}(\mathbf{k}, \boldsymbol{\theta}) - L_{\mathbf{x}^{[2h-1]}}\{\operatorname{argsup}_{(\mathbf{k}, \boldsymbol{\theta})} L_{\mathbf{x}^{[2h]}}(\mathbf{k}, \boldsymbol{\theta})\}]/1000$ を算出している.

表 6.2: モデル誤特定がないときに, AIC_{naive} と AIC で変化点有りのモデルと変化点無しモデルを選択させたときの前者の選択確率 (%)

		$n: 50$	$n: 100$	$n: 200$	$n: 400$
$(\theta_0^{*(1)}, \theta_1^{*(1)}, \theta_0^{*(2)}, \theta_1^{*(2)}): (1.0, 0.1, 2.0, 0.5)$	AIC_{naive}	63.1	88.6	99.0	100.0
	AIC	24.1	57.8	93.1	100.0
$(\theta_0^{*(1)}, \theta_1^{*(1)}, \theta_0^{*(2)}, \theta_1^{*(2)}): (1.0, 0.2, 2.0, 0.4)$	AIC_{naive}	51.6	77.8	97.8	99.8
	AIC	16.6	43.5	83.1	98.5
$(\theta_0^{*(1)}, \theta_1^{*(1)}, \theta_0^{*(2)}, \theta_1^{*(2)}): (1.0, 0.3, 1.0, 0.3)$	AIC_{naive}	25.1	28.7	35.8	45.7
	AIC	3.5	4.8	7.2	10.7
$(\theta_0^{*(1)}, \theta_1^{*(1)}, \theta_0^{*(2)}, \theta_1^{*(2)}): (1.0, 0.4, 2.0, 0.2)$	AIC_{naive}	39.5	59.1	83.5	96.0
	AIC	11.2	26.2	49.0	81.9
$(\theta_0^{*(1)}, \theta_1^{*(1)}, \theta_0^{*(2)}, \theta_1^{*(2)}): (1.0, 0.5, 2.0, 0.1)$	AIC_{naive}	35.6	55.8	81.6	95.9
	AIC	7.8	22.4	48.2	82.2
$(\theta_0^{*(1)}, \theta_1^{*(1)}, \theta_0^{*(2)}, \theta_1^{*(2)}): (2.0, 0.3, 2.0, 0.3)$	AIC_{naive}	21.1	29.6	35.4	44.3
	AIC	3.4	5.0	8.1	9.6

これらの値は, (6.22) のモデルに基づき, 1,000 回の繰り返し実験により評価したものである.

表 6.3: モデル誤特定がないときの, 真の分布と推定された分布間の Kullback-Leibler ダイバージェンス (K-L) と, 変化点数 0, 1, 2, 3, 4 の選択確率 (%)

		K-L	0	1	2	3	4	
$n: 100$	$\xi^*: 2.4$	AIC _{naive}	13.87	0.8	5.5	71.8	17.3	4.6
		AIC	12.93	6.7	18.1	73.4	1.8	0.0
		BIC	13.98	20.9	28.5	50.2	0.4	0.0
	$\xi^*: 2.2$	AIC _{naive}	12.95	1.4	7.7	72.9	15.0	3.0
		AIC	12.42	10.5	22.4	65.2	1.8	0.1
		BIC	13.62	24.4	33.9	41.4	0.3	0.0
	$\xi^*: 2.0$	AIC _{naive}	13.51	2.4	10.0	68.1	16.9	2.6
		AIC	11.94	14.6	27.0	56.9	1.4	0.1
		BIC	13.18	31.1	36.4	32.4	0.1	0.0
$n: 200$	$\xi^*: 1.8$	AIC _{naive}	12.28	0.0	2.6	68.3	23.1	6.0
		AIC	10.96	2.6	10.9	82.7	3.6	0.2
		BIC	12.43	13.7	24.8	61.2	0.3	0.0
	$\xi^*: 1.6$	AIC _{naive}	12.10	0.8	3.4	67.3	21.6	6.9
		AIC	10.40	5.1	16.3	75.3	3.3	0.0
		BIC	12.27	20.5	31.3	48.2	0.0	0.0
	$\xi^*: 1.4$	AIC _{naive}	12.45	1.3	7.5	63.6	21.3	6.3
		AIC	11.37	7.5	24.6	64.2	3.5	0.2
		BIC	13.22	24.8	40.3	34.7	0.2	0.0
$n: 400$	$\xi^*: 1.2$	AIC _{naive}	12.91	0.2	2.0	58.4	28.6	10.8
		AIC	11.05	1.3	9.9	82.3	6.1	0.4
		BIC	12.30	10.5	32.1	57.4	0.0	0.0
	$\xi^*: 1.0$	AIC _{naive}	13.91	0.1	3.4	57.2	27.3	12.0
		AIC	11.41	3.5	18.0	72.9	5.2	0.4
		BIC	13.10	17.2	42.7	40.0	0.1	0.0
	$\xi^*: 0.8$	AIC _{naive}	13.39	0.7	7.7	52.9	27.3	11.4
		AIC	11.30	4.3	29.4	61.5	4.5	0.3
		BIC	13.28	23.3	53.6	23.1	0.0	0.0

これらの値は, (6.23) のモデルに基づき, 1,000 回の繰り返し実験により評価したものである.

表 6.4: モデル誤特定があるときの最大対数尤度のバイアスと, (6.19) の TIC による漸近バイアス評価

			n: 50		n: 100		n: 200		n: 400	
			true	TIC	true	TIC	true	TIC	true	TIC
$\theta_1^{*(1)}: 0.1$	$\theta_1^{*(2)}: 0.3$	$\rho: 0.2$	11.66	11.73	13.19	13.22	12.79	12.84	12.89	12.78
		$\rho: 0.4$	10.69	10.70	10.80	10.92	11.51	11.56	11.46	11.46
	$\theta_1^{*(2)}: 0.5$	$\rho: 0.2$	12.85	11.47	14.24	15.22	13.98	13.81	14.09	14.08
		$\rho: 0.4$	12.17	12.37	12.53	11.76	13.20	13.22	13.60	13.58
$\theta_1^{*(1)}: 0.3$	$\theta_1^{*(2)}: 0.1$	$\rho: 0.2$	11.70	11.74	12.98	14.08	13.74	14.13	13.47	13.05
		$\rho: 0.4$	10.87	10.78	11.28	11.22	11.13	11.08	11.83	11.37
	$\theta_1^{*(2)}: 0.5$	$\rho: 0.2$	13.85	13.00	14.94	15.11	15.43	14.25	15.50	14.34
		$\rho: 0.4$	13.01	13.00	13.96	13.97	14.44	13.81	14.36	13.85
$\theta_1^{*(1)}: 0.5$	$\theta_1^{*(2)}: 0.1$	$\rho: 0.2$	13.00	13.28	14.13	13.83	15.05	14.30	15.55	14.77
		$\rho: 0.4$	12.06	12.07	12.96	13.12	14.12	13.92	15.10	13.73
	$\theta_1^{*(2)}: 0.3$	$\rho: 0.2$	13.61	14.02	14.81	13.96	16.28	15.34	16.94	14.95
		$\rho: 0.4$	12.33	11.93	13.50	12.80	14.98	13.04	16.35	13.45

$(\theta_0^{*(1)}, \theta_0^{*(2)})$ はすべて $(1.0, 2.0)$ としている。これらの値は, (6.22) のモデルに基づき, 1,000 回の繰り返しによるモンテカルロ法で, 表 6.1 と同じように評価したものである。

表 6.5: モデル誤特定があるときの, 真の分布と推定された分布間の Kullback-Leibler ダイバージェンス (K-L) と, 変化点数 0, 1, 2, 3, 4 の選択確率 (%)

			K-L	0	1	2	3	4	
n: 200	$\xi: 1.8$	$\rho: 0.2$	AIC _{naive}	9.60	0.5	2.3	75.0	17.4	4.8
			AIC	8.90	2.4	10.7	84.5	2.2	0.2
			TIC	8.86	2.4	9.7	85.0	2.8	0.1
			BIC	10.83	14.6	23.9	61.5	0.0	0.0
		$\rho: 0.4$	AIC _{naive}	7.71	0.2	2.5	80.1	14.3	2.9
			AIC	7.33	3.0	10.3	84.3	2.2	0.2
			TIC	7.29	2.5	9.4	85.4	2.5	0.2
			BIC	9.50	15.2	24.8	59.8	0.1	0.1
	$\xi: 1.4$	$\rho: 0.2$	AIC _{naive}	9.54	1.3	7.0	69.6	18.6	3.5
			AIC	8.86	8.6	22.7	67.1	1.5	0.1
			TIC	8.80	6.8	19.8	71.1	2.2	0.1
			BIC	11.29	25.4	39.3	35.3	0.0	0.0
		$\rho: 0.4$	AIC _{naive}	7.66	0.9	9.0	74.6	13.3	2.2
			AIC	7.75	8.2	23.9	66.0	1.8	0.1
			TIC	7.65	7.5	21.3	68.8	2.1	0.3
			BIC	10.46	27.5	40.3	32.1	0.1	0.0
n: 400	$\xi: 1.2$	$\rho: 0.2$	AIC _{naive}	9.83	0.2	2.3	67.6	23.6	6.3
			AIC	8.54	2.1	9.8	82.8	4.7	0.6
			TIC	8.45	1.8	7.5	83.3	6.6	0.8
			BIC	10.61	13.1	27.4	59.3	0.2	0.0
		$\rho: 0.4$	AIC _{naive}	7.59	0.2	3.0	73.2	19.1	4.5
			AIC	6.86	1.7	11.6	84.2	2.4	0.1
			TIC	6.83	1.1	9.3	85.0	4.5	0.1
			BIC	9.29	11.3	31.7	56.9	0.1	0.0
	$\xi: 0.8$	$\rho: 0.2$	AIC _{naive}	10.48	1.0	9.5	60.4	21.9	7.2
			AIC	9.21	5.9	28.5	61.9	3.5	0.2
			TIC	9.09	4.5	23.8	65.8	5.5	0.4
			BIC	11.34	23.6	50.0	26.3	0.1	0.0
		$\rho: 0.4$	AIC _{naive}	7.40	1.8	10.0	65.6	18.6	4.0
			AIC	6.91	7.9	26.1	63.4	2.6	0.0
			TIC	6.87	6.5	21.5	66.5	5.2	0.3
			BIC	9.65	24.7	47.3	27.9	0.1	0.0

これらの値は, (6.23) のモデルに基づき, 1,000 回の繰り返し実験により評価したものである.

第7章 結論

生存時間解析において、ハザード比関数の変化点検出は需要が大きいこと、及び共変量間に多重共線性が存在する状況も幾分想定されることから、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る形で一般化させた上で、変化点をもつ Cox 比例ハザードモデルに対する AIC タイプの情報量規準を、本論文では導出した。まずは、正則化最大対数部分尤度の漸近バイアスを評価し、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定理 1 で示した。そして、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、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定理 2 で示した。結果、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ること、そして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 と書けること、を確認した。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 AIC 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 K-L 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。また、交差検証法と比較した性能評価では、予測の観点でのモデル選択を目的としている場合には提案規準を用いることが妥当であることを示した。実データ解析では、既存の規準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。本論文では、Cox モデルの比例ハザード性を緩和する手段として変化点解析を扱っているが、上記の規準は対数線形性には強く依存している。そこで、モデルが誤特定された状況下でも理論的に保証される情報量規準として、TIC への拡張も行った。そして、TIC の性能評価も行い、モデル誤特定下では予測の観点で TIC が AIC よりも僅かに優れており、更にモデル誤特定下であっても AIC が AIC_{naive} よりも明らかに良い性能を示すことを確認した。

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。一方、生存時間をモデル化するパラメトリックな生存時間解析の手法で構造変化を考える場合は、生存時間の確率変数で周辺化すると混合分布モデルと同タイプのモデルとなり、特異モデルであることに注意が必要であるものの、一定の条件の下では変化点の推定量の漸近的性質には変化点モデル特有の速い収束レートとはならず、回帰パラメータの推定量と同様の収束レートとなるため、情報量規準の罰則項は本論文で提案したものと大きく異なるであろう。

本論文では、時点でハザード比が変化するモデル（時間に関する変化点モデル）を扱ったが、共変量の値でハザード比が変化するモデル（共変量に関する変化点モデル）も特に近年活発に議論されている（例えば [Pons 2003](#), [Lee and Lam 2020](#), [Wang et al. 2021](#) 等）。時間に関する変化点モデルでは、時点で構造が急激に変わるジャンプモデルを考えていたが、共変量に関する変化点モデルでは、閾値で構造が急激に変化するジャンプモデルを考える場合もあれば、共変量に従って徐々に構造が変化するモデルを考える場合もある。共変量に関するジャンプ型の変化点モデルにおいては、本論文で扱った時点に関するジャンプ型の変化点モデルと同様、変化点パラメータの推定量は回帰パラメータよりも早い収束を示すことが報告されている。一方、共変量に従って徐々に構造が変化するモデルにおいては、回帰パラメータの推定量と同じ速度で変化点パラメータの推定量も収束し、それは漸近正規性をもつことが報告されている。これらのモデルにおける変化点検出については、検定ベースの手法がいくつか提案されている。しかし情報量規準ベースの手法の提案はなく、その開発はニーズがあるものと考えられる。変化点パラメータの推定量の収束速度が二つのモデルで異なることを述べたが、それは推定精度に違いがあるということであり、二つのモデルにおける情報量規準の罰則項は、大きく異なることが予想される。期待値の評価も改めて行う必要が生じるであろう。

生存時間解析において注目を浴び続けているものとしては、共変量である反復測定データと生存時間データを同時にモデル化する joint modelling がある（例えば [Henderson et al. 2000](#) 参照）。ここでのモデルは Cox モデルの拡張版であるので、時点に関する変化点解析も、共変量に関する変化点解析も、需要があることに違いはない。障壁があるとすれば、部分尤度そのものではなく、その拡張ともみなせる profile likelihood を使っていることが、まず挙げられる。漸近理論は [Zeng](#)

and Cai (2005) で構築されており、それをチューニングして本論文と同じ理論に持ち込むことが課題となる。Joint modelling では、情報量規準の構築自体にも障壁がある。この設定では、反復測定データに対してもセミパラメトリックモデルを扱うしランダム効果項も存在するので、Xu et al. (2009) や Donohue et al. (2011) のように AIC や条件付き AIC の構築すること自体も課題となる。

また、Cox モデルだけでなく、変化点解析の需要が計量経済学の分野で大きいことを鑑み、本論文では変化点をもつ ARCH モデルに対する情報量規準も導出した。その際、ARCH モデルに忠実に従う経済データは少ないであろうことを想定し、モデル誤特定を許容する情報量規準である TIC を開発した。そして、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の 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られることを示した。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。

モデル誤特定を扱うという目的のもと、本論文ではデータが従う分布の誤特定を考えたが、変化のモデリングの誤特定というのも考えられるであろう。例えば、本当は緩やかな変化（例えば Hušková 1999 や Vogt and Dette 2015）であるにもかかわらず、急な変化をもつ通常の変化点モデルを考えていたとしたら、推定値はそれほど変わらなくても推定精度の評価は変わってくるであろうし、よって情報量規準の罰則項も相当に異なってくるであろう。本論文の方法論は、AR 過程を組み合わせたものへの拡張は容易であるが、MA 過程を組み合わせたものへの拡張は容易でない。実はその理由は、MA 過程の係数に急な変化をもたせたとしても、分布としては緩やかな変化をもつことになるからである。緩やかな変化を扱うことは、様々な意味で枠組みを広げることになり、今後の重要な課題と考える。

付録A (3.1)の導出過程

$(\partial/\partial\beta^{(j)})l_\xi(\widehat{\beta}_{\mathbf{k}}, \mathbf{k}; \mathbf{t}) = \mathbf{0}_p$ を $\widehat{\beta}_{\mathbf{k}}^{(j)} = \beta_\xi^{*(j)}$ の周りで1次テイラー展開すると

$$\begin{aligned} \mathbf{0}_p &= \sum_{i \in D^{(j)}} \{z_i - \mathbf{h}(t_i, \beta_\xi^{*(j)}) - \xi \beta_\xi^{*(j)}\} \\ &\quad - \sum_{i \in D^{(j)}} \{\mathbf{H}(t_i, \beta_\xi^{*(j)}) - \mathbf{h}(t_i, \beta_\xi^{*(j)})\mathbf{h}(t_i, \beta_\xi^{*(j)})^\top + \xi \mathbf{I}_p\} (\widehat{\beta}_{\mathbf{k}}^{(j)} - \beta_\xi^{*(j)}) \{1 + o_P(1)\} \end{aligned}$$

と書け,

$$\begin{aligned} \widehat{\beta}_{\mathbf{k}}^{(j)} - \beta_\xi^{*(j)} &= \left[\sum_{i \in D^{(j)}} \{\mathbf{H}(t_i, \beta_\xi^{*(j)}) - \mathbf{h}(t_i, \beta_\xi^{*(j)})\mathbf{h}(t_i, \beta_\xi^{*(j)})^\top + \xi \mathbf{I}_p\} \right]^{-1} \\ &\quad \left[\sum_{i \in D^{(j)}} \{z_i - \mathbf{h}(t_i, \beta_\xi^{*(j)}) - \xi \beta_\xi^{*(j)}\} \right] \{1 + o_P(1)\} \end{aligned}$$

が得られる。また、 $k^{(j-1)} \leq k^{*(j-1)}$ かつ $k^{(j)} \leq k^{*(j)}$ とすると

$$\begin{aligned} &\widehat{\beta}_{\mathbf{k}}^{(j)} - \widehat{\beta}_{\mathbf{k}^*}^{(j)} \\ &= \left[\sum_{i \in D^{(j)}} \{\mathbf{H}(t_i, \beta_\xi^{*(j)}) - \mathbf{h}(t_i, \beta_\xi^{*(j)})\mathbf{h}(t_i, \beta_\xi^{*(j)})^\top + \xi \mathbf{I}_p\} \right]^{-1} \\ &\quad \left[\sum_{i \in D(\{k^{(j-1)}, k^{*(j-1)}\}) \cup D(\{k^{*(j-1)}, k^{(j)}\})} \{z_i - \mathbf{h}(t_i, \beta_\xi^{*(j)}) - \xi \beta_\xi^{*(j)}\} \right] \{1 + o_P(1)\} \\ &\quad - \left[\sum_{i \in D^{*(j)}} \{\mathbf{H}(t_i, \beta_\xi^{*(j)}) - \mathbf{h}(t_i, \beta_\xi^{*(j)})\mathbf{h}(t_i, \beta_\xi^{*(j)})^\top + \xi \mathbf{I}_p\} \right]^{-1} \\ &\quad \left[\sum_{i \in D(\{k^{*(j-1)}, k^{(j)}\}) \cup D(\{k^{(j)}, k^{*(j)}\})} \{z_i - \mathbf{h}(t_i, \beta_\xi^{*(j)}) - \xi \beta_\xi^{*(j)}\} \right] \{1 + o_P(1)\} \\ &= \left[\sum_{i \in D^{*(j)}} \{\mathbf{H}(t_i, \beta_\xi^{*(j)}) - \mathbf{h}(t_i, \beta_\xi^{*(j)})\mathbf{h}(t_i, \beta_\xi^{*(j)})^\top + \xi \mathbf{I}_p\} \right]^{-1} \end{aligned}$$

$$\left[\sum_{i \in D([k^{(j-1)}, k^{*(j-1)})} \{z_i - \mathbf{h}(t_i, \beta_\xi^{*(j)}) - \xi \beta_\xi^{*(j)}\} - \sum_{i \in D([k^{(j)}, k^{*(j)})} \{z_i - \mathbf{h}(t_i, \beta_\xi^{*(j)}) - \xi \beta_\xi^{*(j)}\} \right] \{1 + o_P(1)\}$$

と書ける。ここで、 $\tilde{D}^{*(j)} \equiv D([k^{*(j-1)}, k^{*(j)}])$ である。 $k^{(j-1)}$ と $k^{*(j-1)}$, $k^{(j)}$ と $k^{*(j)}$ の大小関係が変わってもこれは成立するので、一般化すると

$$\begin{aligned} \hat{\beta}_{\mathbf{k}}^{(j)} - \hat{\beta}_{\mathbf{k}^*}^{(j)} &= \left[\sum_{i \in \tilde{D}^{*(j)}} \{\mathbf{H}(t_i, \beta_\xi^{*(j)}) - \mathbf{h}(t_i, \beta_\xi^{*(j)}) \mathbf{h}(t_i, \beta_\xi^{*(j)})^\top + \xi \mathbf{I}_p\} \right]^{-1} \\ &\quad \left[\sum_{i \in D^{(j)} \setminus \tilde{D}^{*(j)}} \{z_i - \mathbf{h}(t_i, \beta_\xi^{*(j)}) - \xi \beta_\xi^{*(j)}\} - \sum_{i \in \tilde{D}^{*(j)} \setminus D^{(j)}} \{z_i - \mathbf{h}(t_i, \beta_\xi^{*(j)}) - \xi \beta_\xi^{*(j)}\} \right] \\ &\quad \{1 + o_P(1)\} \end{aligned} \quad (\text{A.1})$$

が得られ、ゆえに $\|\hat{\beta}_{\mathbf{k}}^{(j)} - \hat{\beta}_{\mathbf{k}^*}^{(j)}\| = O_P(1/n)$ が成り立つ。Tsiatis (1981) と同様にして、 $\|\hat{\beta}_{\mathbf{k}^*} - \beta_\xi^*\| = O_P(1/\sqrt{n})$ であることも言えるため、これらのことから (2.4) が成り立つことが分かる。次に、正則化対数部分尤度関数に対して $\beta_\xi^{*(j)} = \hat{\beta}_{\mathbf{k}}^{(j)}$ の周りでテイラー展開し、(A.1) を利用すると

$$\begin{aligned} &l_\xi(\hat{\beta}_{\mathbf{k}}, \mathbf{k}; \mathbf{t}) - l_\xi(\beta_\xi^*, \mathbf{k}; \mathbf{t}) \\ &= \sum_{j=1}^{m+1} \left(-(\beta_\xi^{*(j)} - \hat{\beta}_{\mathbf{k}}^{(j)})^\top \left[\sum_{i \in D^{(j)}} \{z_i - \mathbf{h}(t_i, \hat{\beta}_{\mathbf{k}}^{(j)}) - \xi \hat{\beta}_{\mathbf{k}}^{(j)}\} \right] + \right. \\ &\quad \left. \frac{1}{2}(\beta_\xi^{*(j)} - \hat{\beta}_{\mathbf{k}}^{(j)})^\top \left[\sum_{i \in D^{(j)}} \{\mathbf{H}(t_i, \hat{\beta}_{\mathbf{k}}^{(j)}) - \mathbf{h}(t_i, \hat{\beta}_{\mathbf{k}}^{(j)}) \mathbf{h}(t_i, \hat{\beta}_{\mathbf{k}}^{(j)})^\top + \xi \mathbf{I}_p\} \right] (\beta_\xi^{*(j)} - \hat{\beta}_{\mathbf{k}}^{(j)}) \right) \\ &\quad + o_P(1) \\ &= \sum_{j=1}^{m+1} \left(\frac{1}{2}(\beta_\xi^{*(j)} - \hat{\beta}_{\mathbf{k}^*}^{(j)})^\top \left[\sum_{i \in D^{(j)}} \{\mathbf{H}(t_i, \hat{\beta}_{\mathbf{k}^*}^{(j)}) - \mathbf{h}(t_i, \hat{\beta}_{\mathbf{k}^*}^{(j)}) \mathbf{h}(t_i, \hat{\beta}_{\mathbf{k}^*}^{(j)})^\top + \xi \mathbf{I}_p\} \right] (\beta_\xi^{*(j)} - \hat{\beta}_{\mathbf{k}^*}^{(j)}) \right) \\ &\quad + o_P(1) \\ &= l_\xi(\hat{\beta}_{\mathbf{k}^*}, \mathbf{k}^*; \mathbf{t}) - l_\xi(\beta_\xi^*, \mathbf{k}^*; \mathbf{t}) + o_P(1) \end{aligned} \quad (\text{A.2})$$

と書ける。よって、 $l_\xi(\hat{\beta}_{\mathbf{k}}, \mathbf{k}; \mathbf{t}) - l_\xi(\hat{\beta}_{\mathbf{k}^*}, \mathbf{k}^*; \mathbf{t}) = l_\xi(\beta_\xi^*, \mathbf{k}; \mathbf{t}) - l_\xi(\beta_\xi^*, \mathbf{k}^*; \mathbf{t}) + o_P(1)$ であることから、(3.1) を得る。

付録B (3.3)の導出過程

s がいずれかの要素で有限でない値をもつベクトルの場合は、ある j が存在して $|k^{(j)} - k^{*(j)}| \rightarrow \infty$ ($n \rightarrow \infty$) となることを意味する。このケースを一般的に扱うことは無用に複雑にするため、ここでは $j' (\neq j)$ に対して

$$\begin{cases} k^{(j')} = k^{*(j')} + s^{(j')}/n, & 0 > s^{(j')} \neq O(1), \\ k^{(j)} = k^{*(j)} + s^{(j)}/n, & s^{(j)} = O(1) \quad (j \neq j') \end{cases}$$

の場合のみを詳細に扱う。このとき、 $l_\xi(\hat{\beta}_{\mathbf{k}}, \mathbf{k}; \mathbf{t}) - l_\xi(\hat{\beta}_{\mathbf{k}^*}, \mathbf{k}^*; \mathbf{t})$ は

$$\begin{aligned} & \sum_{i \in D([k^{*(j'-1)}, k^{(j')}])} \left[(\hat{\beta}_{\mathbf{k}}^{(j')} - \hat{\beta}_{\mathbf{k}^*}^{(j')})^\top \mathbf{z}_i - \log \left\{ \frac{\sum_{i' \in R(t_i)} \exp(\hat{\beta}_{\mathbf{k}}^{(j')\top} \mathbf{z}_{i'})}{\sum_{i' \in R(t_i)} \exp(\hat{\beta}_{\mathbf{k}^*}^{(j')\top} \mathbf{z}_{i'})} \right\} \right. \\ & \quad \left. - \frac{\xi}{2} (\hat{\beta}_{\mathbf{k}}^{(j')\top} \hat{\beta}_{\mathbf{k}}^{(j')} - \hat{\beta}_{\mathbf{k}^*}^{(j')\top} \hat{\beta}_{\mathbf{k}^*}^{(j')}) \right] \\ & + \sum_{i \in D([k^{(j')}, k^{*(j')})} \left[(\hat{\beta}_{\mathbf{k}}^{(j'+1)} - \hat{\beta}_{\mathbf{k}^*}^{(j')})^\top \mathbf{z}_i - \log \left\{ \frac{\sum_{i' \in R(t_i)} \exp(\hat{\beta}_{\mathbf{k}}^{(j'+1)\top} \mathbf{z}_{i'})}{\sum_{i' \in R(t_i)} \exp(\hat{\beta}_{\mathbf{k}^*}^{(j')\top} \mathbf{z}_{i'})} \right\} \right. \\ & \quad \left. - \frac{\xi}{2} (\hat{\beta}_{\mathbf{k}}^{(j')\top} \hat{\beta}_{\mathbf{k}}^{(j')} - \hat{\beta}_{\mathbf{k}^*}^{(j')\top} \hat{\beta}_{\mathbf{k}^*}^{(j')}) \right] \\ & + \sum_{i \in D([k^{*(j')}, k^{*(j'+1)})} \left[(\hat{\beta}_{\mathbf{k}}^{(j'+1)} - \hat{\beta}_{\mathbf{k}^*}^{(j'+1)})^\top \mathbf{z}_i - \log \left\{ \frac{\sum_{i' \in R(t_i)} \exp(\hat{\beta}_{\mathbf{k}}^{(j'+1)\top} \mathbf{z}_{i'})}{\sum_{i' \in R(t_i)} \exp(\hat{\beta}_{\mathbf{k}^*}^{(j'+1)\top} \mathbf{z}_{i'})} \right\} \right. \\ & \quad \left. - \frac{\xi}{2} (\hat{\beta}_{\mathbf{k}}^{(j'+1)\top} \hat{\beta}_{\mathbf{k}}^{(j'+1)} - \hat{\beta}_{\mathbf{k}^*}^{(j'+1)\top} \hat{\beta}_{\mathbf{k}^*}^{(j'+1)}) \right] + O_P(1) \end{aligned} \quad (\text{B.1})$$

と分解できる。テイラー展開を利用すると、最初の和は

$$(\hat{\beta}_{\mathbf{k}}^{(j')} - \hat{\beta}_{\mathbf{k}^*}^{(j')})^\top \sum_{i \in D([k^{*(j'-1)}, k^{(j')}])} \left[\mathbf{z}_i - \frac{\partial}{\partial \beta^{(j')}} \log \left\{ \sum_{i' \in R(t_i)} \exp(\hat{\beta}_{\mathbf{k}^*}^{(j')\top} \mathbf{z}_{i'}) \right\} - \xi \hat{\beta}_{\mathbf{k}^*}^{(j')} \right]$$

となり、3つ目の和は

$$(\widehat{\beta}_{\mathbf{k}}^{(j'+1)} - \widehat{\beta}_{\mathbf{k}^*}^{(j'+1)})^\top \sum_{i \in D((\mathbf{k}^{*(j')}, \mathbf{k}^{*(j'+1)})})} \left[\mathbf{z}_i - \frac{\partial}{\partial \beta^{(j'+1)}} \log \left\{ \sum_{i' \in R(t_i)} \exp(\widehat{\beta}_{\mathbf{k}^*}^{(j'+1)\top} \mathbf{z}_{i'}) \right\} - \xi \widehat{\beta}_{\mathbf{k}^*}^{(j'+1)} \right]$$

となる. $\widehat{\beta}_{\mathbf{k}}^{(j')} - \widehat{\beta}_{\mathbf{k}^*}^{(j')} = O_P(s^{(j')}/n)$, $\widehat{\beta}_{\mathbf{k}}^{(j'+1)} - \widehat{\beta}_{\mathbf{k}^*}^{(j'+1)} = O_P(s^{(j')}/n)$, $\widehat{\beta}_{\mathbf{k}^*}^{(j')} = \beta_{\xi}^{*(j')} + o_P(1)$, 及び $\widehat{\beta}_{\mathbf{k}^*}^{(j'+1)} = \beta_{\xi}^{*(j'+1)} + o_P(1)$ より, これらはそれぞれ $o_P(s^{(j')})$ となる. そして, 2つ目の和は

$$\sum_{i \in D((\mathbf{k}^{(j')}, \mathbf{k}^{*(j')})}) \left[(\widehat{\beta}_{\mathbf{k}}^{(j'+1)} - \beta_{\xi}^{*(j')})^\top \mathbf{z}_i - \log \left\{ \frac{\sum_{i' \in R(t_i)} \exp(\widehat{\beta}_{\mathbf{k}}^{(j'+1)\top} \mathbf{z}_{i'})}{\sum_{i' \in R(t_i)} \exp(\beta_{\xi}^{*(j')\top} \mathbf{z}_{i'})} \right\} - \frac{\xi}{2} (\widehat{\beta}_{\mathbf{k}}^{(j'+1)\top} \widehat{\beta}_{\mathbf{k}}^{(j'+1)} - \beta_{\xi}^{*(j')\top} \beta_{\xi}^{*(j')}) \right] + o_P(s^{(j')})$$

と書け, この期待値は $\beta_{\xi}^{*(j')}$ の定義から負かつ $O_P(s^{(j')})$ となる. よって, 任意の $M > 0$ に対して $P\{l_{\xi}(\widehat{\beta}_{\mathbf{k}}, \mathbf{k}; \mathbf{t}) - l_{\xi}(\widehat{\beta}_{\mathbf{k}^*}, \mathbf{k}^*; \mathbf{t}) > -M\} \rightarrow 0$ が言え, (3.2) より任意の $M > 0$ に対して (3.3) を得る.

付録C (3.9)の導出過程

$\widehat{l}_\xi(\widetilde{\mathbf{k}}_{\mathbf{u}}; \mathbf{t}, \mathbf{u})$ に対し, (3.7), (3.8), 及び Murphy and van der Vaart (2000) を利用すると

$$\begin{aligned}
& l_\xi(\boldsymbol{\beta}_\xi^*, \widetilde{\mathbf{k}}_{\mathbf{u}}; \mathbf{t}) - l_\xi(\widehat{\boldsymbol{\beta}}_{\widetilde{\mathbf{k}}_{\mathbf{u}}, \mathbf{u}}^{(j)}, \widetilde{\mathbf{k}}_{\mathbf{u}}; \mathbf{t}) \\
&= \sum_{j=1}^{m+1} \left(-(\widehat{\boldsymbol{\beta}}_{\widetilde{\mathbf{k}}_{\mathbf{u}}, \mathbf{u}}^{(j)} - \boldsymbol{\beta}_\xi^{*(j)})^\top \left[\sum_{i \in \widetilde{D}_{\mathbf{u}}^{(j)}} \{ \mathbf{z}_i - \mathbf{h}(t_i, \boldsymbol{\beta}_\xi^{*(j)}) - \xi \boldsymbol{\beta}_\xi^{*(j)} \} \right] \right. \\
&\quad \left. - \frac{1}{2} (\widehat{\boldsymbol{\beta}}_{\widetilde{\mathbf{k}}_{\mathbf{u}}, \mathbf{u}}^{(j)} - \boldsymbol{\beta}_\xi^{*(j)})^\top \mathbf{E} \left[\sum_{i \in \widetilde{D}_{\mathbf{u}}^{(j)}} \{ \mathbf{H}(t_i, \boldsymbol{\beta}_\xi^{*(j)}) - \mathbf{h}(t_i, \boldsymbol{\beta}_\xi^{*(j)}) \mathbf{h}(t_i, \boldsymbol{\beta}_\xi^{*(j)})^\top + \xi \mathbf{I}_p \} \right] (\widehat{\boldsymbol{\beta}}_{\widetilde{\mathbf{k}}_{\mathbf{u}}, \mathbf{u}}^{(j)} - \boldsymbol{\beta}_\xi^{*(j)}) \right) \\
&\quad + o_P(1) \\
&= \sum_{j=1}^{m+1} \left(-(\widehat{\boldsymbol{\beta}}_{\mathbf{k}^*, \mathbf{u}}^{(j)} - \boldsymbol{\beta}_\xi^{*(j)})^\top \left[\sum_{i \in \widetilde{D}^{*(j)}} \{ \mathbf{z}_i - \mathbf{h}(t_i, \boldsymbol{\beta}_\xi^{*(j)}) - \xi \boldsymbol{\beta}_\xi^{*(j)} \} \right] \right. \\
&\quad \left. - \frac{1}{2} (\widehat{\boldsymbol{\beta}}_{\mathbf{k}^*, \mathbf{u}}^{(j)} - \boldsymbol{\beta}_\xi^{*(j)})^\top \mathbf{E} \left[\sum_{i \in \widetilde{D}^{*(j)}} \{ \mathbf{H}(t_i, \boldsymbol{\beta}_\xi^{*(j)}) - \mathbf{h}(t_i, \boldsymbol{\beta}_\xi^{*(j)}) \mathbf{h}(t_i, \boldsymbol{\beta}_\xi^{*(j)})^\top + \xi \mathbf{I}_p \} \right] (\widehat{\boldsymbol{\beta}}_{\mathbf{k}^*, \mathbf{u}}^{(j)} - \boldsymbol{\beta}_\xi^{*(j)}) \right) \\
&\quad + o_P(1) \\
&= - \sum_{j=1}^{m+1} \left((\widehat{\boldsymbol{\beta}}_{\mathbf{k}^*, \mathbf{u}}^{(j)} - \boldsymbol{\beta}_\xi^{*(j)})^\top \left[\sum_{i \in \widetilde{D}^{*(j)}} \{ \mathbf{z}_i - \mathbf{h}(t_i, \boldsymbol{\beta}_\xi^{*(j)}) - \xi \boldsymbol{\beta}_\xi^{*(j)} \} \right] + \frac{1}{2} \boldsymbol{\nu}_\xi^{(j)\top} \boldsymbol{\nu}_\xi^{*(j)} \right) + o_P(1) \quad (\text{C.1})
\end{aligned}$$

が得られる。ここで, $\widetilde{D}_{\mathbf{u}}^{(j)} = D([\widetilde{\mathbf{k}}_{\mathbf{u}}^{(j-1)}, \widetilde{\mathbf{k}}_{\mathbf{u}}^{(j)})$ である。また,

$$\begin{aligned}
& l_\xi(\boldsymbol{\beta}_\xi^*, \mathbf{k}^*; \mathbf{t}) - l_\xi(\boldsymbol{\beta}_\xi^*, \widetilde{\mathbf{k}}_{\mathbf{u}}; \mathbf{t}) \\
&= \sum_{j=1}^m \left\{ I_{\{\widetilde{\mathbf{k}}_{\mathbf{u}}^{(j)} < \mathbf{k}^*(j)\}} \left(\sum_{i \in D([\widetilde{\mathbf{k}}_{\mathbf{u}}^{(j)}, \mathbf{k}^*(j)])} \left[(\boldsymbol{\beta}_\xi^{*(j)} - \boldsymbol{\beta}_\xi^{*(j+1)})^\top \mathbf{z}_i - \log \left\{ \frac{\sum_{i' \in R(t_i)} \exp(\boldsymbol{\beta}_\xi^{*(j)\top} \mathbf{z}_{i'})}{\sum_{i' \in R(t_i)} \exp(\boldsymbol{\beta}_\xi^{*(j+1)\top} \mathbf{z}_{i'})} \right\} \right. \right. \\
&\quad \left. \left. - \frac{\xi}{2} (\boldsymbol{\beta}_\xi^{*(j)\top} \boldsymbol{\beta}_\xi^{*(j)} - \boldsymbol{\beta}_\xi^{*(j+1)\top} \boldsymbol{\beta}_\xi^{*(j+1)}) \right] \right) \\
&\quad + I_{\{\widetilde{\mathbf{k}}_{\mathbf{u}}^{(j)} > \mathbf{k}^*(j)\}} \left(\sum_{i \in D([\mathbf{k}^*(j), \widetilde{\mathbf{k}}_{\mathbf{u}}^{(j)})} \left[(\boldsymbol{\beta}_\xi^{*(j+1)} - \boldsymbol{\beta}_\xi^{*(j)})^\top \mathbf{z}_i - \log \left\{ \frac{\sum_{i' \in R(t_i)} \exp(\boldsymbol{\beta}_\xi^{*(j+1)\top} \mathbf{z}_{i'})}{\sum_{i' \in R(t_i)} \exp(\boldsymbol{\beta}_\xi^{*(j)\top} \mathbf{z}_{i'})} \right\} \right] \right)
\end{aligned}$$

$$\begin{aligned}
& \left. - \frac{\xi}{2} (\boldsymbol{\beta}_\xi^{*(j+1)\top} \boldsymbol{\beta}_\xi^{*(j+1)} - \boldsymbol{\beta}_\xi^{*(j)\top} \boldsymbol{\beta}_\xi^{*(j)}) \right] \Bigg\} \\
= & \sum_{j=1}^m Q_{\xi; \tilde{\mathbf{k}}_u^{(j)}, t}^{(j)} \tag{C.2}
\end{aligned}$$

であることから, (3.9)を得る.

付録D 条件(3.11)の下での変化点推定量の漸近挙動

まず, (3.4)の導出と同様にして, $\mathbf{k} = \mathbf{k}^* + \alpha_n \mathbf{s}/n$ としたときの \mathbf{s} の各要素が有限値であるとき, (A.1)の代わりに $\widehat{\boldsymbol{\beta}}_{\mathbf{k}} - \widehat{\boldsymbol{\beta}}_{\mathbf{k}^*} = O_P(\sqrt{\alpha_n}/n)$ が得られる. そして, (A.2)や(3.1), (3.2)が言えるため, $\widehat{l}_\xi(\mathbf{k}; \mathbf{t}, \mathbf{t}) = O_P(1)$ が成立する. 一方, \mathbf{s} のいずれかの要素が有限値でないときも(3.4)の導出と同様に, $k^{(j')} = k^{*(j')} + \alpha_n s^{(j')}/n$, $0 > s^{(j')} \neq O(1)$ かつ $j \neq j'$ に対して $k^{(j)} = k^{*(j)} + \alpha_n s^{(j)}/n$, $s^{(j)} = O(1)$ の場合を考える. やはり同様に(B.1)を経由して, 任意の $M > 0$ に対して $P\{l_\xi(\widehat{\boldsymbol{\beta}}_{\mathbf{k}}, \mathbf{k}; \mathbf{t}) - l_\xi(\widehat{\boldsymbol{\beta}}_{\mathbf{k}^*}, \mathbf{k}^*; \mathbf{t}) > -M\} \rightarrow 0$ が言える. そして(3.2)と併せれば $P\{\widehat{l}_\xi(\mathbf{k}; \mathbf{t}, \mathbf{t}) > -M\} \rightarrow 0$ の成立が分かる. 以上より, (3.4)の代わりに $\widehat{\mathbf{k}} - \mathbf{k}^* = O_P(\alpha_n/n)$ が得られ, $K^{(j)} = \{k \mid k - k^{*(j)} = O(\alpha_n/n)\}$ として $K = \{\mathbf{k} \mid k^{(j)} \in K^{(j)}, j \in \{1, 2, \dots, m\}\}$ であることが分かる. また, それにより, (3.5), (3.6), (C.1), (C.2), (3.9)が言える.

付録 E 定理 2 の証明

条件 (3.11) の下, $Q_{\xi; k^{*(j)} + \alpha_n s^{(j)}/n, t}^{(j)}$ は

$$\begin{aligned}
& I_{\{s^{(j)} < 0\}} \left(\frac{1}{\sqrt{\alpha_n}} \Delta^{(j)\top} \left[\sum_{i \in D_{1\alpha_n}^{*(j)}} \{z_i - \mathbf{h}(t_i, \beta_\xi^{*(j)}) - \xi \beta_\xi^{*(j)}\} \right. \right. \\
& \quad \left. \left. - \frac{1}{2\alpha_n} \Delta^{(j)\top} \left[\sum_{i \in D_{1\alpha_n}^{*(j)}} \{\mathbf{H}(t_i, \beta_\xi^{*(j)}) - \mathbf{h}(t_i, \beta_\xi^{*(j)}) \mathbf{h}(t_i, \beta_\xi^{*(j)})^\top + \xi \mathbf{I}_p\} \right] \Delta^{(j)} \right) \\
& + I_{\{s^{(j)} > 0\}} \left(\frac{1}{\sqrt{\alpha_n}} \Delta^{(j)\top} \left[\sum_{i \in D_{2\alpha_n}^{*(j)}} \{z_i - \mathbf{h}(t_i, \beta_\xi^{*(j+1)}) - \xi \beta_\xi^{*(j+1)}\} \right] \right. \\
& \quad \left. - \frac{1}{2\alpha_n} \Delta^{(j)\top} \left[\sum_{i \in D_{2\alpha_n}^{*(j)}} \{\mathbf{H}(t_i, \beta_\xi^{*(j+1)}) - \mathbf{h}(t_i, \beta_\xi^{*(j+1)}) \mathbf{h}(t_i, \beta_\xi^{*(j+1)})^\top + \xi \mathbf{I}_p\} \right] \Delta^{(j)} \right) \\
& + \text{op}(1) \tag{E.1}
\end{aligned}$$

と書ける. ここで, $D_{1\alpha_n}^{*(j)} \equiv D([k^{*(j)} + \alpha_n s^{(j)}/n, k^{*(j)}])$ かつ $D_{2\alpha_n}^{*(j)} \equiv D([k^{*(j)}, k^{*(j)} + \alpha_n s^{(j)}/n])$ である. いま, $\{W_s\}_{s \in \mathbb{R}}$ を $E(W_s) = 0$ かつ $V(W_s) = |s|$ である両側標準ブラウン運動とすると

$$\frac{1}{\sqrt{\alpha_n}} \Delta^{(j)\top} \left[\sum_{i \in D_{1\alpha_n}^{*(j)}} \{z_i - \mathbf{h}(t_i, \beta_\xi^{*(j)}) - \xi \beta_\xi^{*(j)}\} \right] \xrightarrow{d} (\Delta^{(j)\top} \mathbf{A}_\xi^{*(j)} \Delta^{(j)})^{1/2} W_s, \tag{E.2}$$

$$\begin{aligned}
& \frac{1}{2\alpha_n} \Delta^{(j)\top} \left[\sum_{i \in D_{1\alpha_n}^{*(j)}} \{\mathbf{H}(t_i, \beta_\xi^{*(j)}) - \mathbf{h}(t_i, \beta_\xi^{*(j)}) \mathbf{h}(t_i, \beta_\xi^{*(j)})^\top + \xi \mathbf{I}_p\} \right] \Delta^{(j)} \\
& \xrightarrow{p} \frac{1}{2} \Delta^{(j)\top} (\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p) \Delta^{(j)} |s|, \tag{E.3}
\end{aligned}$$

$$\frac{1}{\sqrt{\alpha_n}} \Delta^{(j)\top} \left[\sum_{i \in D_{2\alpha_n}^{*(j)}} \{z_i - \mathbf{h}(t_i, \beta_\xi^{*(j+1)}) - \xi \beta_\xi^{*(j+1)}\} \right] \xrightarrow{d} (\Delta^{(j)\top} \mathbf{A}_\xi^{*(j+1)} \Delta^{(j)})^{1/2} W_s \tag{E.4}$$

かつ

$$\begin{aligned} & \frac{1}{2\alpha_n} \Delta^{(j)\top} \left[\sum_{i \in D_{2\alpha_n}^{*(j)}} \{ \mathbf{H}(t_i, \beta_\xi^{*(j+1)}) - \mathbf{h}(t_i, \beta_\xi^{*(j+1)}) \mathbf{h}(t_i, \beta_\xi^{*(j+1)})^\top + \xi \mathbf{I}_p \} \right] \Delta^{(j)} \\ & \xrightarrow{P} \frac{1}{2} \Delta^{(j)\top} (\mathbf{A}_\xi^{*(j+1)} + \xi D^{*(j+1)} \mathbf{I}_p) \Delta^{(j)} |s| \end{aligned} \quad (\text{E.5})$$

が得られる. 次に, $V_s(\tau_1, \tau_2, \sigma_1, \sigma_2)$ を, τ_1 や τ_2 をドリフト係数, σ_1 や σ_2 を拡散係数とした, 両側に伸びるブラウン運動とする. つまり, $V_s(\tau_1, \tau_2, \sigma_1, \sigma_2)$ を $s < 0$ のとき $-\tau_1|s| + \sigma_1 W_s$, $s \geq 0$ のとき $-\tau_2|s| + \sigma_2 W_s$ と定義し, さらに

$$\begin{aligned} V_{\xi;s}^{*(j)} \equiv & V_s \left\{ \frac{1}{2} \Delta^{(j)\top} (\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p) \Delta^{(j)}, \frac{1}{2} \Delta^{(j)\top} (\mathbf{A}_\xi^{*(j+1)} + \xi D^{*(j+1)} \mathbf{I}_p) \Delta^{(j)}, \right. \\ & \left. (\Delta^{(j)\top} \mathbf{A}_\xi^{*(j)} \Delta^{(j)})^{1/2}, (\Delta^{(j)\top} \mathbf{A}_\xi^{*(j+1)} \Delta^{(j)})^{1/2} \right\} \end{aligned}$$

とする. このとき, (E.1) や (E.2), (E.3), (E.4), (E.5) より

$$Q_{\xi;k^*+\alpha_n s/n, t}^{(j)} \xrightarrow{d} V_{\xi;s}^{*(j)}$$

が成立する. そして, これらの帰結として

$$\sup_{k \in K^{(j)}} Q_{\xi;k, t}^{(j)} \xrightarrow{d} \sup_{s \in \mathbb{R}} V_{\xi;s}^{*(j)} \quad (\text{E.6})$$

と

$$Q_{\xi; \arg \sup_{k \in K^{(j)}} Q_{\xi;k, u, t}^{(j)}}^{(j)} \xrightarrow{d} V_{\xi; \arg \sup_{s \in \mathbb{R}} V_{\xi;s}^{*(j)}}^{*(j)} \quad (\text{E.7})$$

が得られる. ここで, $V_{\xi;s}^{*(j)}$ は $V_{\xi;s}^{*(j)}$ のコピーである.

これらの期待値を評価するために, [Bhattacharya and Brockwell \(1976\)](#) や [Shepp \(1979\)](#) の結果を用いていく. まず, 正の定数 a_1 と a_2 に対して成り立っている

$$P \left\{ \sup_{s>0} (W_s - a_2 s) > a_1 \right\} = \exp(-2a_1 a_2)$$

という等式を使うと

$$\begin{aligned}
& \mathbb{E} \left(\sup_{s \in \mathbb{R}} V_{\xi; s}^{*(j)} \right) \\
&= \int_0^\infty \mathbb{P} \left(\sup_{s \in \mathbb{R}} V_{\xi; s}^{*(j)} > a \right) da \\
&= \int_0^\infty \exp \left\{ - \frac{\Delta^{(j)\top} (\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p) \Delta^{(j)}}{\Delta^{(j)\top} \mathbf{A}_\xi^{*(j)} \Delta^{(j)}} a \right\} da \\
&\quad + \int_0^\infty \exp \left\{ - \frac{\Delta^{(j)\top} (\mathbf{A}_\xi^{*(j+1)} + \xi D^{*(j+1)} \mathbf{I}_p) \Delta^{(j)}}{\Delta^{(j)\top} \mathbf{A}_\xi^{*(j+1)} \Delta^{(j)}} a \right\} da \\
&\quad - \int_0^\infty \exp \left\{ - \frac{\Delta^{(j)\top} (\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p) \Delta^{(j)}}{\Delta^{(j)\top} \mathbf{A}_\xi^{*(j)} \Delta^{(j)}} a - \frac{\Delta^{(j)\top} (\mathbf{A}_\xi^{*(j+1)} + \xi D^{*(j+1)} \mathbf{I}_p) \Delta^{(j)}}{\Delta^{(j)\top} \mathbf{A}_\xi^{*(j+1)} \Delta^{(j)}} a \right\} da \\
&= \mathbf{C}^{(j)}(\mathbf{A}_\xi^{*(\cdot)}, \mathbf{A}_\xi^{*(\cdot)} + \xi D^{*(\cdot)} \mathbf{I}_p) \tag{E.8}
\end{aligned}$$

が得られる。ここで、

$$\begin{aligned}
& \mathbf{C}^{(j)}(\mathbf{A}^{(\cdot)\dagger}, \mathbf{A}^{(\cdot)\ddagger}) \\
&= \{ (\Delta^{(j)\top} \mathbf{A}^{(j)\ddagger} \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}^{(j+1)\dagger} \Delta^{(j)})^2 + (\Delta^{(j)\top} \mathbf{A}^{(j+1)\ddagger} \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}^{(j)\dagger} \Delta^{(j)})^2 \\
&\quad + \Delta^{(j)\top} \mathbf{A}^{(j)\ddagger} \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}^{(j+1)\ddagger} \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}^{(j)\dagger} \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}^{(j+1)\dagger} \Delta^{(j)} \} \\
&\quad / \{ \Delta^{(j)\top} \mathbf{A}^{(j)\ddagger} \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}^{(j+1)\ddagger} \Delta^{(j)} \\
&\quad (\Delta^{(j)\top} \mathbf{A}^{(j)\ddagger} \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}^{(j+1)\dagger} \Delta^{(j)} + \Delta^{(j)\top} \mathbf{A}^{(j+1)\ddagger} \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}^{(j)\dagger} \Delta^{(j)} \}
\end{aligned}$$

である。次に、 $\operatorname{argsup}_{s \in \mathbb{R}} V_s(\tau_1, \tau_2, \sigma_1, \sigma_2)$ の確率密度関数が、 $s \leq 0$ のとき $g(-s \mid \tau_1/\sigma_1, \tau_2\sigma_1/\sigma_2^2)$ 、 $s \geq 0$ のとき $g(s \mid \tau_2/\sigma_2, \tau_1\sigma_2/\sigma_1^2)$ であることを使う。ここで、

$$g(s \mid a_1, a_2) \equiv 2a_1(a_1 + 2a_2) \exp\{2a_2(a_1 + a_2)s\} \Phi\{-(a_1 + 2a_2)\sqrt{s}\} - 2a_1^2 \Phi(-a_1\sqrt{s})$$

である。すると、

$$\begin{aligned}
& \mathbb{E} \left(V_{\xi; \operatorname{argsup}_{s \in \mathbb{R}} V_{\xi; s}^{*(j)}}^{*(j)} \right) \\
&= \int_0^\infty \frac{1}{2} s g \left[s \mid \frac{\Delta^{(j)\top} (\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p) \Delta^{(j)}}{2(\Delta^{(j)\top} \mathbf{A}_\xi^{*(j)} \Delta^{(j)})^{1/2}} \right],
\end{aligned}$$

$$\begin{aligned}
& \left. \frac{\Delta^{(j)\top}(\mathbf{A}_\xi^{*(j+1)} + \xi D^{*(j+1)} \mathbf{I}_p) \Delta^{(j)} (\Delta^{(j)\top} \mathbf{A}_\xi^{*(j)} \Delta^{(j)})^{1/2}}{2 \Delta^{(j)\top} \mathbf{A}_\xi^{*(j+1)} \Delta^{(j)}} \right] \Delta^{(j)\top} (\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p) \Delta^{(j)} ds \\
& + \int_0^\infty \frac{1}{2} sg \left[s \left| \frac{\Delta^{(j)\top} (\mathbf{A}_\xi^{*(j+1)} + \xi D^{*(j+1)} \mathbf{I}_p) \Delta^{(j)}}{2 (\Delta^{(j)\top} \mathbf{A}_\xi^{*(j+1)} \Delta^{(j)})^{1/2}}, \right. \right. \\
& \left. \left. \frac{\Delta^{(j)\top} (\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p) \Delta^{(j)} (\Delta^{(j)\top} \mathbf{A}_\xi^{*(j+1)} \Delta^{(j)})^{1/2}}{2 \Delta^{(j)\top} \mathbf{A}_\xi^{*(j)} \Delta^{(j)}} \right] \Delta^{(j)\top} (\mathbf{A}_\xi^{*(j+1)} + \xi D^{*(j+1)} \mathbf{I}_p) \Delta^{(j)} ds \\
= & \Delta^{(j)\top} \mathbf{A}_\xi^{*(j+1)} \Delta^{(j)} \{2 \Delta^{(j)\top} \mathbf{A}_\xi^{*(j)} \Delta^{(j)} \Delta^{(j)\top} (\mathbf{A}_\xi^{*(j+1)} + \xi D^{*(j+1)} \mathbf{I}_p) \Delta^{(j)} \\
& + \Delta^{(j)\top} (\mathbf{A}_\xi^{*(j+1)} + \xi D^{*(j+1)} \mathbf{I}_p) \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}_\xi^{*(j)} \Delta^{(j)}\} (\Delta^{(j)\top} \mathbf{A}_\xi^{*(j)} \Delta^{(j)})^2 \\
/ & [\Delta^{(j)\top} (\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p) \Delta^{(j)} \{\Delta^{(j)\top} (\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p) \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}_\xi^{*(j+1)} \Delta^{(j)} \\
& + \Delta^{(j)\top} (\mathbf{A}_\xi^{*(j+1)} + \xi D^{*(j+1)} \mathbf{I}_p) \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}_\xi^{*(j)} \Delta^{(j)}\}^2] \\
& + \Delta^{(j)\top} (\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p) \Delta^{(j)} \{2 \Delta^{(j)\top} (\mathbf{A}_\xi^{*(j+1)} + \xi D^{*(j+1)} \mathbf{I}_p) \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}_\xi^{*(j)} \Delta^{(j)} \\
& + \Delta^{(j)\top} (\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p) \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}_\xi^{*(j+1)} \Delta^{(j)}\} (\Delta^{(j)\top} \mathbf{A}_\xi^{*(j+1)} \Delta^{(j)})^2 \\
/ & [\Delta^{(j)\top} (\mathbf{A}_\xi^{*(j+1)} + \xi D^{*(j+1)} \mathbf{I}_p) \Delta^{(j)} \{\Delta^{(j)\top} (\mathbf{A}_\xi^{*(j)} + \xi D^{*(j)} \mathbf{I}_p) \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}_\xi^{*(j+1)} \Delta^{(j)} \\
& + \Delta^{(j)\top} (\mathbf{A}_\xi^{*(j+1)} + \xi D^{*(j+1)} \mathbf{I}_p) \Delta^{(j)} \Delta^{(j)\top} \mathbf{A}_\xi^{*(j)} \Delta^{(j)}\}^2] \\
= & \mathbf{C}^{(j)}(\mathbf{A}_\xi^{*(\cdot)}, \mathbf{A}_\xi^{*(\cdot)} + \xi D^{*(\cdot)} \mathbf{I}_p) \tag{E.9}
\end{aligned}$$

が得られる。よって、(E.6) や (E.7), (E.8), (E.9) より定理 2 が得られる。

謝辞

本研究に関して主任指導教員として親身に指導並びに議論をして下さった統計数理研究所 二宮嘉行 教授に心より感謝申し上げます。また、統計数理研究所 船渡川伊久子 准教授には副指導教員及び学位審査の主査を、統計数理研究所 鎌谷研吾 准教授、滋賀大学 杉本知之 教授には副査をご多忙の中お引き受け頂き、多数の有益なご指摘を頂戴しました。心より感謝申し上げます。

勤務先である中外製薬株式会社バイオメトリクス部 淡路直人 部長，生物統計機能 山本英晴 統括マネージャー，統計解析第1G 上田正倫 グループマネージャーには，学位取得の機会を与えて頂き深く感謝申し上げます。並びに，ご支援頂いた所属部署の皆様に深く感謝申し上げます。最後に支えて下さった家族に感謝の意を表します。

参考文献

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds.) 2nd International Symposium on Information Theory (pp. 267–281). Budapest: Akademiai Kiado.
- Ananthakrishnan, R., Green, S., Previtali, A., Liu, R., Li, D., and LaValley, M. (2021). Critical review of oncology clinical trial design under non-proportional hazards, *Critical Reviews in Oncology/Hematology*, **162**, 103350.
- Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point, *Econometrica*, **61**, 821–856.
- Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes, *Econometrica*, **66**, 47–78.
- Bhattacharya, P. K. and Brockwell, P. J. (1976). The minimum of an additive process with applications to signal estimation and storage theory, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **37**, 51–75.
- Brem, H., Piantadosi, S., Burger, P. C., Walker, M., Selker, R., Vick, N. A. et al. (1995). Placebo-controlled trial of safety and efficacy of intraoperative controlled delivery by biodegradable polymers of chemotherapy for recurrent gliomas, *The Lancet*, **345**, 1008–1012.
- Burnham, K. and Anderson, D. (2003). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*: Springer New York, URL: <https://books.google.co.jp/books?id=BQYR6js0CC8C>.
- Carrasco, M., Hu, L., and Ploberger, W. (2014). Optimal test for Markov switching parameters, *Econometrica*, **82**, 765–784.

- Chu, J. C.-S. (1995). Detecting parameter shift in GARCH models, *Econometric Reviews*, **14**, 241–266.
- Cox, D. R. (1972). Regression models and life-tables, *Journal of the Royal Statistical Society: Series B*, **34**, 187–202.
- Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*: New York: Wiley.
- Dacunha-Castelle, D. and Gassiat, E. (1999). Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes, *The Annals of Statistics*, **27**, 1178–1209.
- Donohue, M., Overholser, R., Xu, R., and Vaida, F. (2011). Conditional Akaike information under generalized linear and proportional hazards mixed models, *Biometrika*, **98**, 685–700.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1007.
- FDA (2011). Guidance for industry: clinical considerations for therapeutic cancer vaccines, <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-considerations-therapeutic-cancer-vaccines>.
- Fleming, T. R. and Harrington, D. P. (2013). *Counting processes and survival analysis*, **625**: John Wiley & Sons.
- Gill, R. and Schumacher, M. (1987). A simple test of the proportional hazards assumption, *Biometrika*, **74**, 289–300.
- Hannart, A. and Naveau, P. (2012). An improved Bayesian information criterion for multiple change-point models, *Technometrics*, **54**, 256–268.
- He, P., Fang, L., and Su, Z. (2013). A sequential testing approach to detecting multiple change points in the proportional hazards model, *Statistics in Medicine*, **32**, 1239–1245.

- Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data, *Biostatistics*, **1**, 465–480.
- Hess, K. R. (1995). Graphical methods for assessing violations of the proportional hazards assumption in Cox regression, *Statistics in Medicine*, **14**, 1707–1723.
- Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, **12**, 55–67.
- van Houwelingen, H. C., Bruinsma, T., Hart, A. A., Van't Veer, L. J., and Wessels, L. F. (2006). Cross-validated Cox regression on microarray gene expression data, *Statistics in Medicine*, **25**, 3201–3216.
- Hušková, M. (1999). Gradual changes versus abrupt changes, *Journal of Statistical Planning and Inference*, **76**, 109–125.
- Jones, R. H. and Dey, I. (1995). Determining one or more change points, *Chemistry and Physics of LIPIDS*, **76**, 1–6.
- Kokoszka, P. and Leipus, R. (1999). Testing for parameter changes in ARCH models, *Lithuanian Mathematical Journal*, **39**, 182–195.
- (2000). Change-point estimation in ARCH models, *Bernoulli*, **6**, 513–539.
- Konishi, S. and Kitagawa, G. (1996). Generalised information criteria in model selection, *Biometrika*, **83**, 875–890.
- Lee, C. Y. and Lam, K. F. (2020). Survival analysis with change-points in covariate effects, *Statistical Methods in Medical Research*, **29**, 3235–3248.
- Lin, D. Y. and Wei, L.-J. (1989). The robust inference for the Cox proportional hazards model, *Journal of the American Statistical Association*, **84**, 1074–1078.
- Liu, M., Lu, W., and Shao, Y. (2008). A Monte Carlo approach for change-point detection in the Cox proportional hazards model, *Statistics in Medicine*, **27**, 3894–3909.

- Lv, J. and Liu, J. S. (2014). Model selection principles in misspecified models, *Journal of the Royal Statistical Society: Series B*, **76**, 141–167.
- Murphy, S. A. and van der Vaart, A. W. (2000). On profile likelihood, *Journal of the American Statistical Association*, **95**, 449–465.
- Ninomiya, Y. (2015). Change-point model selection via AIC, *Annals of the Institute of Statistical Mathematics*, **67**, 943–961.
- Pons, O. (2002). Estimation in a Cox regression model with a change-point at an unknown time, *Statistics*, **36**, 101–124.
- (2003). Estimation in a Cox regression model with a change-point according to a threshold in a covariate, *The Annals of Statistics*, **31**, 442–463.
- Qu, Z. and Perron, P. (2007). Estimating and testing structural changes in multivariate regressions, *Econometrica*, **75**, 459–502.
- Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**, 461–464.
- Shepp, L. A. (1979). The joint density of the maximum and its location for a Wiener process with drift, *Journal of Applied Probability*, **16**, 423–427.
- Siegmund, D. O. (2004). Model selection in irregular problems: applications to mapping quantitative trait loci, *Biometrika*, **91**, 785–800.
- Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike’s criterion, *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**, 44–47.
- Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified proportional hazard models, *Biometrika*, **73**, 363–369.
- Takeuchi, K. (1976). The distribution of information statistics and the criterion of goodness of fit of models, *Mathematical Science*, **153**, 12–18 (in Japanese).

- Therneau, T. and Grambsch, P. (2013). *Modeling Survival Data: Extending the Cox Model*, Statistics for Biology and Health: Springer New York, URL: <https://books.google.co.jp/books?id=oj0mBQAAQBAJ>.
- Thomas, L. and Reyes, E. M. (2014). Tutorial: survival estimation for Cox regression models with time-varying coefficients using SAS and R, *Journal of Statistical Software*, **61**, 1–23.
- Tsiatis, A. A. (1981). A large sample study of Cox’s regression model, *The Annals of Statistics*, **9**, 93–108.
- Uno, H., Claggett, B., Tian, L., Inoue, E., Gallo, P., Miyata, T. et al. (2014). Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis, *Journal of Clinical Oncology*, **32**, 2380–2385.
- Vogt, M. and Dette, H. (2015). Detecting gradual changes in locally stationary processes, *The Annals of Statistics*, **43**, 713–740.
- Wang, B., Li, J., and Wang, X. (2021). Change point detection in Cox proportional hazards mixture cure model, *Statistical Methods in Medical Research*, **30**, 440–457.
- Weiss, A. A. (1986). Asymptotic theory for ARCH models: estimation and testing, *Econometric Theory*, **2**, 107–131.
- Witten, D. M. and Tibshirani, R. (2010). Survival analysis with high-dimensional covariates, *Statistical Methods in Medical Research*, **19**, 29–51.
- Xu, R., Vaida, F., and Harrington, D. P. (2009). Using profile likelihood for semiparametric model selection with application to proportional hazards mixed models, *Statistica Sinica*, **19**, 819.
- Yao, Y. C. (1988). Estimating the number of change-points via Schwarz’s criterion, *Statistics and Probability Letters*, **6**, 181–189.
- Zeng, D. and Cai, J. (2005). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time, *The Annals of Statistics*, **33**, 2132–2163.

Zhang, N. R. and Siegmund, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data, *Biometrics*, **63**, 22–32.

小西貞則・北川源四郎 (2004) 『シリーズ〈予測と発見の科学〉2 情報量規準』, 朝倉書店.