

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 5 月 22 日現在

機関番号：15401

研究種目：挑戦的萌芽研究

研究期間：2010～2012

課題番号：22650058

研究課題名（和文）繰り返し計算を必要としない平滑化パラメータ選択法の開発とその応用

研究課題名（英文）Development and Application of A Non-iterative Selection Methods for Smoothing Parameters

研究代表者

柳原 宏和 (Yanagihara Hirokazu)

広島大学・大学院理学研究科・准教授

研究者番号：70342615

研究成果の概要（和文）：本研究課題では，ノンパラメトリック回帰に部分的な一般化リッジ回帰を適用することにより， $C_p$  規準を最小にする平滑化パラメータが一意に求められるようにモデルを変換した．この結果を用い，繰り返し計算を必要としない平滑化パラメータの最適化法を提案し平滑化法の簡便化をはかり，簡便な平滑化法をさまざまな統計モデルに適用させた．

研究成果の概要（英文）：In this research, by applying the generalized ridge regression to the nonparametric regression, we changed the model so that the parameters which make  $C_p$  criterion the minimum can be obtained by closed forms. Using these results, we proposed non-iterative optimization method for smoothing method. Moreover, we applied the proposed smoothing method to several nonparametric regression models.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010 年度	1,000,000	0	1,000,000
2011 年度	900,000	270,000	1,170,000
2012 年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	2,900,000	570,000	3,470,000

研究分野：統計科学

科研費の分科・細目：情報学・統計科学

キーワード：平滑化 モデル選択 情報量規準 一般化リッジ回帰

### 1. 研究開始当初の背景

価値観や生活様式の多様化により，社会現象が複雑化してきている今日において，そのような社会現象を表現するために，B-スプラインやニューラルネットワークなどのノンパラメトリックモデルのような従来の線形モデルよりも柔軟な統計モデルでの解析が求められている．配置された既知の基底関数により未知の関数形を近似するノンパラメトリック回帰モデルを用いた平滑化

は，複雑なトレンドを持つデータに対して，単純にかつ比較的短い計算時間で平滑化結果を得ることができる手法である．推定曲線の局所変動の程度を制御する平滑化パラメータは，値が変われば推定曲線も大きく変化してしまうため，その最適化は重要問題であると言える．最適な平滑化パラメータには， $C_p$  規準 (Mallows, 1973)，CV 規準 (Stone, 1974)，GCV 規準 (Craven & Wahba, 1979) に代表される情報量規準を

最小にする値が用いられることが一般的である。ところが、情報量規準を最小にする解は陽に求めることができず、最適化には計算機による繰り返し計算が必要となる。そのため、基底関数の最大個数が多くなれば計算量も膨大となる。本研究の目的は、繰り返し計算を必要としない平滑化パラメータの選択法を考えることで平滑化の簡便化をはかり、さらに簡便化された平滑化法を様々な統計モデルに適応させることにある。そのような簡便化された平滑化法により最適化に必要な時間は大幅に削減され、今まで計算時間の問題で敬遠されていた複雑な統計モデルも実用可能となると期待できる。

## 2. 研究の目的

ノンパラメトリック回帰モデルでの平滑化パラメータは、推定曲線の局所変動の程度を制御する重要なパラメータである。値が変われば推定曲線も大きく変化してしまうため、平滑化パラメータの最適化は重要問題であると言える。最適な平滑化パラメータは、 $C_p$ 、CV、GCV 規準といった情報量規準の最小化により選択されるのが一般的である。しかし、情報量規準を最小にする解は陽な形で求めることができず、最適な平滑化パラメータを得るためには計算機による繰り返し計算が必要となる。本研究の目的は、平滑化パラメータを追加し  $C_p$  規準を最小にする解が一意に求められるようにすることで繰り返し計算を必要としない平滑化パラメータの最適化法を考え平滑化法の簡便化をはかり、簡便な平滑化法をさまざまな統計モデルに適応させることにある。そのような簡便化を行うことができれば、今まで、コンピュータープログラムによる最適化が必要であったために使用を敬遠していたユーザーも気楽にノンパラメトリック回帰による平滑化を使用することが可能となり、ノンパラメトリック回帰がユーザーフレンドリーな手法となることが期待できる。それは、実解析において大きな意味を持つと考える。また、最適な平滑化パラメータが陽に与えられているということから、今まで取り扱いが難しかった、情報量規準により最適化が行われた後の未知パラメータの標本分布が、比較的簡単に取扱つかうことできるようになる。それは、これまでほとんど解明されていないモデル選択後の標本分布を調べるための第一歩になると考える。

## 3. 研究の方法

一般化リッジ回帰でのリッジパラメータの最適化法をノンパラメトリック回帰モデル

での平滑化パラメータの最適化に応用することで、最適な平滑化パラメータを繰り返し計算無しに求める方法を開発し、平滑化法の簡便化をはかる。実際には、ノンパラメトリック回帰での罰則付き最小二乗推定量が一般化リッジ回帰での罰則付き最小二乗推定量の特別な形になるような変換を見つけることが問題となる。実際には、階差行列に一次独立なベクトルを加えることにより正則行列とした行列の逆行列で未知パラメータを変換すると、ノンパラメトリック回帰における罰則付き残差平方和は、前述した正則行列により変換された基底関数行列を説明変数に持ちリッジパラメータの一部が0となるような一般化リッジ回帰における罰則付き残差平方和と一致する。その変換後の推定量を用いて、部分的に0をとるリッジパラメータを持つ一般化リッジ推定量でのリッジパラメータ選択のための  $C_p$  規準を計算する。また、基底関数の個数を選択するための新たな  $C_p$  規準も提案する。その際、最適な平滑化パラメータが陽に求められているという利点を利用して、平滑化パラメータも確率変数であるという条件の下でバイアスを再評価し、新しいバイアス補正項を加えた  $C_p$  規準も提案する。実際の計算には、Efron (2004) にあるバイアス補正項導出のためのアイデアを用いる。また、従来の平滑化法と新しい平滑化法の計算機を使った数値的比較と理論的な比較の両方行う。さらに、繰り返し計算がいらぬノンパラメトリック回帰モデルのGAMへの適用や、繰り返し測定データを解析できるようなノンパラメトリック成長曲線モデルへの拡張も考える。

## 4. 研究成果

研究論文 Yanagihara (2012)では、基底関数と未知パラメータに線形変換を施すことにより、罰則付き平滑化における平滑化パラメータの選択問題が、部分的リッジ回帰におけるリッジパラメータ選択問題と同値であることを示し、線形変換後のモデルにおいて一般化リッジ回帰を適用することにより  $C_p$  規準を最小にする平滑化パラメータを陽な形で求めることができた。また、この新しい平滑化法では、従来の平滑化手法のように共通の情報量規準で平滑化パラメータと基底関数の個数を選択してもうまくいかない。そのため、基底関数の個数の最適化のための情報量規準も同時に提案した。さらに、研究発表、永井・柳原 (2010) においては、新しい平滑化法をノンパラメ

トリック成長曲線モデルに、また、井上・柳原(2010)では、動径基底関数ネットワークに基づくノンパラメトリック回帰へ適用した結果を報告した。両結果とも従来の手法と同等のパフォーマンスを持ち、かつ計算時間を最高で1/5に短縮することができることがわかった。ところが、一般化加法モデル (Generalized Additive Model: GAM) や動径基底関数ネットワークに基づくノンパラメトリック回帰モデル等の説明変数が複数あるモデルの場合、 $C_p$  規準を用いるために必要な誤差の分散の一致推定量の構成が難しくなる。そのため、分散の推定量を必要としない情報量規準を用いた方法が必要となる。研究発表、柳原 (2011) において、そのような分散の推定量を必要としない GCV 規準においても、GCV規準を最小にする平滑化パラメータを陽な形で求めることができることを報告した。この結果により、GAMや動径基底関数ネットワークに基づくノンパラメトリック回帰モデルに簡便な平滑化手法を適用できることになる。しかしながら重要な問題として基底関数の個数の最適化が残っている。そのため、基底関数をデータの位置で固定して規準量によりその個数をリッジパラメータにより減らす方法を考えた。そこで必要となるのは標本数が説明変数の個数よりも小さいときの一般化リッジ回帰である。本研究において、そのような場合でもGCVを最小にするリッジパラメータは陽な形で求めることができることがわかった。また、主成分回帰に一般化リッジ回帰を適用し、その主成分の個数もGCV規準で選択すればさらに良い結果になることも分かった。本研究課題では、リッジ回帰や一般化リッジ回帰が重要な役割を果たすため、それらの研究を行うことも重要である。研究論文 Yanagihara & Satoh (2010) では、多変量リッジ回帰モデルにおけるリッジパラメータ選択のための $C_p$ 規準のバイアスを、正規性の仮定の下で補正したModified  $C_p$  規準を提案した。この規準量は $C_p$ のバイアスを補正しているだけでなく、分散も小さくしており、実際にはリスク関数の一様最小分散不偏推定量になっていることがわかった。また、研究論文 Nagai et al. (2012) では、多変量一般化リッジ回帰におけるリッジパラメータの法を複数提案した。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

#### [雑誌論文] (計 6 件)

1. Kamo, K., Yanagihara, H., & Satoh, K., Bias-corrected AIC for selecting variables in Poisson regression models, *Communication in Statistics Theory and Method*, 査読有, 42, 2013, 1911-1921.
2. Nagai, I., Yanagihara, H. & Satoh, K., Optimization of ridge parameters in multivariate generalized ridge regression by plug-in methods, *Hiroshima Mathematical Journal*, 査読有, 42, 2012, 301-324.
3. Yanagihara, H., Kamo, K., Imori, S. & Satoh, K., Bias-corrected AIC for selecting variables in multinomial logistic regression models, *Linear Algebra and Its Applications*, 査読有, 436, 2012, 4329-4341.
4. Yanagihara, H., A non-iterative optimization method for smoothness in penalized spline regression, *Statistics and Computing*, 査読有, 22, 2012, 527-544.
5. Yanagihara, H., Kamo, K. & Tonda, T., Second-order bias-corrected AIC in multivariate normal linear models under nonnormality, *The Canadian Journal of Statistics*, 査読有, 39, 2011, 126-146.
6. Yanagihara, H. & Satoh, K., An unbiased  $C_p$  criterion for multivariate ridge regression, *Journal of Multivariate Analysis*, 査読有, 101, 2010, 1226-1238.

#### [学会発表] (計 5 件)

1. Yanagihara, H., Wakaki, H. & Fujikoshi, Y., A consistency property of AIC for multivariate linear models when the dimension and the sample size are large, The 40th Annual Meeting of the Statistical Society of Canada, June 5, 2012, University of Guelph (Canada).
2. 柳原宏和, An explicit solution to minimization problem for GCV in generalized ridge regression, 2011年度統計関連学会連合大会, 2011年9月6日, 九州大学
3. 柳原宏和, 情報量規準とバイアス補正, 応用統計学会 2011年度年会, 2011年6月4日, 大阪大学.
4. 井上永介・柳原宏和, 動径基底関数に基づくノンパラメトリック回帰モデルにおける正則化パラメータの繰り返し計算を必要としない最適化法, 2010年度

統計関連学会連合大会, 2010年9月7日, 早稲田大学.

5. 永井 勇・柳原宏和, ノンパラメトリック成長曲線モデルとその推定法, 2010年度統計関連学会連合大会, 2010年9月7日, 早稲田大学.

〔図書〕(計2件)

1. 藤越康祝・若木宏文・柳原宏和, 広島大学出版会, 確率・統計の数学的基礎, 2011, 373.
2. 吉本 敦・加茂憲一・柳原宏和, 朝倉出版, Rによる環境データの統計分析 -森林科学分野での応用-, 2012, 202.

〔その他〕

ホームページ等

## 6. 研究組織

### (1) 研究代表者

柳原 宏和 (Yanagihara Hirokazu)  
広島大学・大学院理学研究科・准教授  
研究者番号: 70342615

### (2) 研究分担者

( )

研究者番号:

### (3) 連携研究者

藤澤 洋徳 (Hironori Fujisawa)  
統計数理研究所・数理・推論研究系・准教授  
研究者番号: 00301177

二宮 嘉行 (Yoshiyuki Ninomiya)  
九州大学・マス フォア インダストリ研究所・准教授  
研究者番号: 50343330