

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 5日現在

機関番号：14401

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22540126

研究課題名（和文） 回帰関数と確率密度関数に関する推定の研究

研究課題名（英文） Estimate on Regression Functions and Probability
Density Functions

研究代表者

白旗 慎吾（SHIRAHATA SHINGO）

大阪大学・その他部局等・名誉教授

研究者番号：10037294

研究成果の概要（和文）：以下の2つの領域で成果を挙げた。まず回帰関数の推定について。直線回帰で、誤差の仮定によらない頑健な推定量を調べた。マウスの発する超音波を時間を説明変数とする回帰関数とみなし、誤差の消去法、超音波かどうかの判断基準、得られた曲線の分類の新しい方法を提案した。次に、確率密度関数の2乗の積分の推定について。カーネル関数、U統計量を用いた推定量を構成し、その漸近正規性を示し、窓幅の性質等をシミュレーションにより確認した。

研究成果の概要（英文）：We proposed new procedures in the following two fields. First, we proposed robust estimator for simple linear regression and methods to analyze ultrasonic vocalization (USV) emitted from mice and rats. By Fourier transformation of USV, we have functional data. We proposed methods how to eliminate errors, how to judge whether the extracted are really curves and how to classify the curves. Next, we constructed estimators of the integration of squared density function. We adopted kernel functions and U-statistics and showed that the asymptotic normality of the estimators and investigated several properties of them by computer simulations.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	700,000	210,000	910,000
2011年度	600,000	180,000	780,000
2012年度	500,000	150,000	650,000
年度			
年度			
総計	1,800,000	540,000	2,340,000

研究分野：数物系科学

科研費の分科・細目：数学・数学一般（含確率論・統計数学）

キーワード：計算機統計学、非線形統計学、確率密度関数、2乗の積分、kernel 関数、U統計量、関数データ、超音波データ

1. 研究開始当初の背景

(1) 説明変数によってデータを説明しようとする回帰分析において、回帰関数は本質的には関数であるが、実際には観測値は離散点における回帰関数値に誤差や変動が加わっ

て得られる。回帰関数は説明変数の関数で、データの平均項、もしくは主要項である。特に線形回帰分析は従来からよく研究され、さらに近年は一般化線形モデル等に拡張されている。これらの回帰分析では回帰関数のパ

ラメータに関する統計的推測を行う。これは回帰関数に特定の構造を仮定することである。しかしながら一般にはこれらは未知であり、少数個のパラメータを持つ特定のモデルは必ずしも正当性を持たない。また誤差分布に特定の分布族を仮定することも正当性を持たない。それゆえ統計学の適用範囲が広がるにつれ、回帰関数に特定の関数型を、また誤差分布に特定の分布を仮定せずに適用できるノンパラメトリック回帰分析が重要になりつつある。

(2) 回帰関数そのものだけではなく、その微分、2回微分の推定も重要である。例えば、人間や生物の成長分析は、単に年齢を説明変数とした回帰関数の推定だけではなく成長が継続しているか(1回微分)、思春期における成長加速の様子はどうか(2回微分)等の解析も必要となる。これらの研究はまだ不十分である。

(3) さらに、ノンパラメトリック回帰分析と同様な手法を用いる確率密度関数やその汎関数(積分や累積分布関数等)の推定も未発達である。回帰関数と確率密度関数の推定は、関数を推定する、ということ以外には一見別物のように見えるが、手法として、カーネル関数を用いる、スプライン関数を用いる、等共通する部分が多く、同じ土俵で研究できる。確率密度関数の汎関数では、その2乗の積分は最も有力なノンパラメトリック法であるウィルコクソン型統計量の良さを測る尺度であり特に重要である。これまでは単純に確率密度関数の推定量の2乗の積分を行っていたが、例えばカーネル法での重要なパラメータである bandwidth (窓幅) は問題によって異なる筈であるが、その決定の研究は何もなされていない。

2. 研究の目的

(1) 本研究の初期の目的は

①回帰関数に特定の関数を仮定しないノンパラメトリック回帰分析、とその微分、2回微分の推定の研究、および誤差分布に特定の分布型を仮定しない回帰分析における推定量の頑健性の研究。

②上とほぼ同じ手法が適用できると考えられる確率密度関数とその汎関数に関する推定問題の研究。

③複数の群からデータが得られているときの同等性に関する仮説検定問題の研究。

④複数の系列があるときに、これらを分類するためのクラスター分析。であった。

(2) 上記をすべて行うことは困難で、特に①における回帰関数の推定、その微分、2回微分の推定、および depth による概念を応用しての回帰分析の頑健性の研究と②の研究に注力する。③、④は余裕があれば行う努力

目標である。

3. 研究の方法

(1) ノンパラメトリックな回帰関数の推定においても、確率密度関数の推定においても、推定量の構成にはデータの平滑化が本質的な役割を果たす。平滑化の方法は多いが、近年はカーネル関数による平滑化、およびスプライン関数による平滑化が盛んに研究されている。この2つの方法の比較では、これまでの経験によれば、スプライン関数、特に最もよく応用されるB-スプラインによる平滑化では節点の位置と個数の選択が重要である。多くの場合にカーネル関数による推定より平均2乗誤差が小さいが、説明変数の端で爆発しやすい欠点がある。また確率密度関数には非負という制約があるが、制約を満たすとは限らない。一方、カーネル関数による推定では、端で爆発することはほとんどない。窓幅の選択と関数の選択が必要となるが、効率の点で関数の選択はあまり重要ではない。また数学的には、スプライン関数による方法は ad hoc で計算が重要であり、数学的には扱いにくい、カーネル関数による方法は数学的に扱いやすい。

考えている推定問題は独立同分布の場合のパラメータ推定のように単純ではなく、数学的な証明は困難な場合である。したがって、コンピュータによるシミュレーションも多用される。

(2) 推定の対象となる関数に構造を仮定しない場合はカーネル関数やスプライン関数が大きな役割を果たすが、誤差分布に具体的な関数型を仮定しない場合は別の方法論が必要となる。このときはばらつきの極めて大きい観測値、いわゆる外れ値の影響を軽減する必要がある、その考え方として、推定値となりうる値全体に順序をつける depth の概念を用いて回帰分析を行う。

4. 研究成果

研究の目的における①および②において新たな知見を得た。③はこれからの課題とし、④については①と融合した研究を行った。

(1) 単回帰における回帰係数の推定問題において depth に基づく方法を改良した。この概念は順序統計量を基にした推定法で、データの個数計算から値を求める。残差平方和を最小にしようという通常思想とは本質的に異なり、外れ値に引きずられることがなく、頑健な方法論である。その漸近的破綻点は $1/3$ 、つまり観測値の $1/3$ が外れ値であっても健全な推定値が求まる。これは最小2乗法などが持つことの不可能な頑健性である。

ただし推定値の候補は複数出現し、その平均を推定量としていた。従来の研究でデータ

数が大きければ外れ値の割合が高い場合でも推定精度が高いことが示されていたが、データ数が小さい場合に健全でない結果がしばしば見られていた。そこで、推定値候補の中央値を考えることにより、現実的な状況で健全な結果となる推定量を構成した。外れ値の多いことで知られている数値例への適用、およびコンピュータ・シミュレーションにより精度が高いことを示した。

(2) 回帰関数の推定における①、④においては、マウスの雄が雌に対して発する超音波の解析で新たな方法を提案した。解析の対象となるデータは人間の耳には聞こえない超音波データである。動物の交流に関する遺伝学的研究で重要な分野であり、国際シンポジウムもたびたび開催され、連携研究者 Dou との共同研究の結果を数回発表している。それらの結果をふまえ論文にまとめる作業中である。

まず超音波データを適当な時間間隔で高速フーリエ変換する。得られた結果は各時点での離散周波数における強度のデータである。ただし元データはマイクロフォンから得られており、バックグラウンド・ノイズが大きく、これを除去しなければならない。除去のために移動平均法を適用する。移動平均は時間だけではなく、周波数領域にも適用する。その結果に各時点での最大強度となる周波数を求める。これを時間が説明変数である回帰関数データと見なす。ただし音波データと見なせるためには閾値がある。単純な回帰データに比べ極めて複雑なデータであり、世界中の関連研究者は手作業で解析している。これをかなりの程度自動化した。

関数データと見なせるデータではあるが、曲線と見なせる部分以外はゼロとしなければならない特殊な関数データである。また、音波と見なすにはある程度連続して強度が大きくなければならない。そこで、閾値と連続の基準回数、補間の方法の提案を行い、これまでの手作業から半自動化を行った。マウスの種によって超音波の強度は大きく異なるため全自動化はこれからの課題である。

得られたデータに B-スプライン法を適用して部分曲線群が得られる。これらは形もさまざまであり、ジャンプも存在する。そこでまずジャンプの回数で最初に分類し、その後クラスタ分析により曲線群の分類を行った。結果は遺伝学者にも好評であり、さらに複数波長の超音波を同時に発しているマウスに関する研究を継続する。

(3) 確率密度関数の汎関数の推定については、その 2 乗の積分の推定問題について理論的・実証的結果を得た。確率密度関数の推定量としてはカーネルとしてある確率密度関数を用いたカーネル型推定量を採用し、その 2 乗の積分を目的のパラメータの推定量と

する。カーネルで選択すべきパラメータはカーネル関数と窓幅である。ただし多くの問題で窓幅の選択は重要であるが、カーネル関数の選択はあまり重要ではないことが知られている。

推定量は標本数 n を固定すれば Hoeffding の提案した U 統計量になることが分かり位数 2 であるが、その核の形は n に関係し、漸近的 ($n \rightarrow \infty$) な考察では U 統計量の標準的理論を適用できない。そこで推定量を漸近展開して、漸近的に正規分布に従う部分と小さな定数項、および確率的に小さい項への分解を得る。各項の n に関する次数を詳細に検討した結果、推定量の平均 2 乗誤差を通常の標準的なモデルの下での次数 -1 にするためには窓幅 h の n に関する次数は $-1/2$ より大きく $n \rightarrow \infty$ のとき $h \rightarrow 0$ となればよいことが分かった。実用的には $-1/3$ または $-1/4$ 程度がよいと思われる。確率密度関数そのものの推定の場合は h の最適な次数は $-1/5$ であることが知られているが、2 乗の積分の場合はより広い範囲から選択できる。積分することにより誤差が打ち消し合っていると考えられる。

窓幅の選択とカーネル関数の選択の影響がどの程度かを調べるためにコンピュータ・シミュレーションを行った。母分布としては、正規分布、ロジスティック分布、両側指数分布、指数分布を想定した。カーネル関数としては、正規密度カーネルと一様カーネル、窓幅では次数 $-1/2.5$ から -5 までを網羅的に調べた結果、一様カーネルが正規密度カーネルよりわずかに性能が上であるが、その差は小さく、カーネルの選択はあまり重要ではないことが確認できた。次数に関しては、非対称な指数分布では次数はやや小さく採るべきであることが示唆された。対称な 3 つの分布では最適な次数は分布によって変動が大きい、 $-1/3$ から $-1/5$ までであり大きな差は生ぜず、次数の選択も本質的ではないと予想できる。今後は、母分布とカーネル関数の種類を増やして知見を確認する予定である。

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 3 件)

① 山邊太陽、丸尾和司、白旗慎吾、後藤昌司、2 変量ベキ正規分布に基づく投与前値対投与後値の比の評価指標とその性質、応用統計学、査読有、41 巻、1 号、(2012)、53-68

② 藤木美江、白旗慎吾、単回帰における Depth を用いたロバスト推定法とその検証、計算機統計学、査読有、23 巻、2 号、(2011)、81-96

③ Maruo, K., Shirahata, S. and Goto, M. Underlying assumptions of the power-normal distribution. Behaviormetrika, 査読有, Vol. 31, No. 1, (2011), 85-95

〔学会発表〕(計 11 件)

① 白旗慎吾、密度関数の 2 乗の積分の推定、計算機統計学会 27 回大会、弘前大、2013. 5. 17

② Dou, X., Shirahata, S., Sugimoto, H. and Koide, T. Functional clustering of mouse ultrasonic vocalization, 5th International Conference of the ERCIM Working Group on Computing and Statistics, Oviedo, Spain, 2012.12.3

③ Dou, X., Shirahata, S., Sugimoto, H. and Koide, T. Functional cluster analysis of mouse ultrasonic vocalization data, USV workshop, Paris. 2012.4.16

④ Dou, X., Shirahata, S. and Sugimoto, H. Cluster analysis of mouse ultrasonic vocalization data, IASC and ARS Taipei, 2011.12.18

⑤ Yamaguchi, Y., Sakamoto, W., Shirahata, S. and Goto, M. Reconstruction of individual patient data for meta-analysis via Bayesian approach. IASC and ARS, Taipei, 2011.12.16

⑥ 山口祐介・坂本 亘・白旗慎吾・後藤昌司、個別患者データを併用したメタ・アナリシス：シミュレーションに基づく方法、統計関連学会連合大会、2011.9.7

⑦ Yamaguchi, Y., Sakamoto, W., Shirahata, S. and Goto, M. A meta-analysis method based on simulated individual patient data, ISI2011 Dublin, 2011.8.23

⑧ 山口祐介・坂本 亘・白旗慎吾・後藤昌司、メタ・アナリシスにおける交互作用効果の評価：シミュレーションに基づく方、日本計算機統計学会第 25 回大会、2011.5.7

⑨ 山口祐介・坂本 亘・白旗慎吾・後藤昌司、臨床評価におけるデータの復元と活用、日本計算機統計学会第 24 回シンポジウム、2010.11.11

⑩ 川端ゆみこ・坂本 亘・白旗慎吾・後藤昌司、ベキ正規分布に基づく生体リズムの評価、日本計算機統計学会第 24 回シンポジウム、2010.11.11

⑪ 藤木美江・白旗慎吾、単回帰 depth 法によるロバスト推定、統計関連学連合大会、2010.9.7

〔図書〕(計 1 件)

① 白旗慎吾・内田雅之・熊谷悦生・黒木学・坂本雄二・坂本亘、統計学辞典、共立出版、2010、1-512

6. 研究組織

(1) 研究代表者

白旗 慎吾 (SHIRAHATA SHINGO)
大阪大学・その他部局等・名誉教授
研究者番号：1 0 0 3 7 2 9 4

(2) 連携研究者

坂本 亘 (SAKAMOTO WATARU)
大阪大学・大学院基礎工学研究科・准教授
研究者番号：7 0 3 0 4 0 2 9

藤木 美江 (FUJIKI MIE)
同志社大学・東アジア総合研究センター・
嘱託研究員
研究者番号：7 0 4 3 7 3 5 9

ドウ シャオリン (DOU XIAOLING)
統計数理研究所・新領域融合研究センタ
ー・特任研究員
研究者番号：1 0 5 1 6 8 6 8