

科学研究費助成事業 研究成果報告書

平成 26 年 5 月 16 日現在

機関番号：10101

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500339

研究課題名(和文)ベルンシュタイン多項式による密度推定法の統計理論とその応用に関する研究

研究課題名(英文)A study on statistical theory of density estimation via the Bernstein polynomial and its application

研究代表者

柿沢 佳秀 (KAKIZAWA, Yoshihide)

北海道大学・経済学研究科(研究院)・教授

研究者番号：30281778

交付決定額(研究期間全体)：(直接経費) 4,000,000円、(間接経費) 1,200,000円

研究成果の概要(和文)：本研究では、独立同一分布設定における確率密度や定常時系列過程におけるスペクトル密度(行列)のような密度関数をノンパラメトリックに推定するため、ベルンシュタイン多項式近似に焦点をおいた。まずベルンシュタイン推定量の加法型・積型(従って、非負型)バイアス修正を提案した。次に、非対称カーネル推定法の開発をめざし変形ベッセルカーネル推定量の新しい族を構成した。それら漸近バイアス・分散・平均積分2乗誤差などの公式を厳密に求め、漸近正規性も証明した。さらに、ガンマカーネル推定量において平均積分2乗誤差を小さくするような変形版の推定量を提示した。漸近理論を確認する数値実験と実データ解析も実施した。

研究成果の概要(英文)：In this study, we focus on the nonparametric methods via the Bernstein polynomial approximation to estimate a function such as the probability density (independent and identically distributed setting) or the spectral density (stationary process). We also aimed at developing methods using some asymmetric kernels. Especially, we have proposed (i) additive or multiplicative (hence, nonnegative) bias corrections to the Bernstein-based estimator and (ii) a class of modified Bessel kernel-based estimators. Their asymptotic properties (bias, variance, mean integrated squared error, and so on), including the asymptotic normality, were derived rigorously. Furthermore, we have re-examined (iii) the gamma kernel density estimator to get a better performance in terms of MISE. Several numerical studies based on both simulated and real data sets were presented to confirm our results in this study.

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：ノンパラメトリック推定 密度関数 ベルンシュタイン多項式

1. 研究開始当初の背景

本研究で考察する『ノンパラメトリック法』とは、確率密度関数やスペクトル密度(行列)関数の推定に対し、何らかの関数近似理論を応用し議論する体系の総称である。それは、ターゲット関数に有限次元の母数モデルを仮定してその母数を尤度原理などから推定するいわゆる『パラメトリック法』と、一部の構造に母数モデルを仮定するが他方にはそうでない構造をも認めるといった現代的なアプローチである『セミパラメトリック法』と併せて統計科学の3つの推測理論の基礎をなす。本研究の対象は、確率密度関数及びスペクトル密度(行列)関数で、特に、後者は定常な単変量(またはベクトル値)時系列過程の周波数領域解析に関係する。統計科学ではノンパラメトリックな関数推定の研究として位置づける。

このような分野の先行研究は膨大であり、歴史的に振り返るならば、独立同一分布設定で確率密度・分布関数・分位点関数、定常な単変量(またはベクトル値)離散時間時系列のスペクトル密度(行列)関数・分布関数、ノンパラメトリック回帰における回帰関数、生存時間分析におけるハザード関数などを典型的題材として議論されてきた。伝統的なカーネル法の背後には、推定するターゲット関数をカーネル関数による畳み込み積分として近似するアイデアがあって、畳み込み積分以外の近似もしばしば使用されてきた。例えば、十分滑らかな関数のテーラー近似(局所的な多項式の係数を推定する問題)、三角多項式・ウェーブレット近似(推定対象の関数空間に付随する直交関数で展開して、それら係数を推定するような定式化をする)、スプライン基底関数近似、ベルンシュタイン多項式近似などに基づいて種々の推定法が提唱される。なお、数学分野の関数近似理論では近似誤差の評価に関心をおくが、一方、統計科学では未知のターゲット関数に対しデータからどのようにすれば有効に推測できるかを問題として、通常、標本数を無限大にするという漸近理論を考える。統計科学において、特に考慮すべきは、統計的な性能を決めるもの(以下、『平滑化パラメータ』と呼ぶ)があり、関数近似誤差(統計学の専門用語で推定量の『バイアス』に相当する)と推定量の『分散』のトレードオフを制御する事実である。従って、ノンパラメトリックな関数推定において、平滑化パラメータの選択が重要な課題の1つであり、1990年代以降、国内・国外で研究され続け、現在でも著しい進展がみられる。

研究代表者は1990年代後半に、カーネル法に基づいたスペクトル密度(行列)推定量の汎関数を対象として様々な漸近的性質を明らかにした。特に、スペクトル密度行列の擬似距離を定義し、多変量定常時系列データを周波数領域から判別・クラスタリングする

という手続きの提案、並びに、その擬似距離をスペクトル密度行列の推定・検定問題へと応用することで、尤度原理と漸近同等なものを提示した。これらの漸近的性質を得るためバンド幅の標本数に関するオーダー条件を仮定したが、これはバンド幅がこのオーダーならスペクトル密度(行列)推定量の積分量が漸近正規するという先行研究を継承したからである。

数学分野での関数近似理論を統計科学へ応用することでノンパラメトリック推測を行う際の土台が形成されることに注目した。研究代表者がベルンシュタイン多項式近似の理論に関心を持つようになったのは、2003-2005の科研費(若手研究)の研究成果による。丁度その頃、2002年に掲載された確率分布・密度関数を推定する枠組みからの先行研究を知って、当時の1つの研究課題(定常な離散時間時系列過程のスペクトル密度の推定)に対してカーネル密度推定量を『ベルンシュタイン多項式』で重み付けすることを提案した。さらに、そのような発想は対応する確率密度関数の推定問題へと発展した。予備実験で平滑化パラメータの役割をする多項式次数の選択に対し公差確認法を主に使用したが、公差確認法の計算量を軽減するために、一種のバイアス調整済の情報量規準も議論した。太陽黒点データに対して、スペクトル密度の推定値をグラフに描いてみると、提案したスペクトル密度推定量では原点近くに意外なピークが検出されて、このことから、ベルンシュタイン多項式密度推定における原点付近での局所的な多項式次数の選択に関心を持っていた。

さらに、ベルンシュタイン多項式近似理論の統計科学での発展をめざし、2008-2010の科研費(基盤研究C)の成果の1つとして、『一般化ベルンシュタイン多項式』に基づく確率密度関数の推定法の研究へと進展した。その頃、ベルンシュタイン多項式密度推定の平均2乗誤差及び平均積分2乗誤差を改良する課題、並びに、『非対称カーネル推定法』に着手する構想を持ち、それらの方向で研究を開始した。

2. 研究の目的

統計科学の全般で想定する状況毎の固有問題があるから、単変量(またはベクトル値)定常時系列のスペクトル密度(行列)汎関数の推測理論だけでなく、独立同一分布設定における確率密度(分布)関数とその汎関数を推定する問題も研究対象に含めている。

ノンパラメトリック関数推定法では、数学分野における関数近似理論を統計科学分野に取り入れることで、新しい推定量の提案に至ることができる。本研究では、多数の関数近似理論の候補の中から、ベルンシュタイン多項式近似法に基づいた確率密度関数及びスペクトル密度(行列)関数の推定量の構築、

並びに、それらの漸近性能の数理的な解明をめざして『ベルンシュタイン多項式近似法』をいっそう普及させていく。さらに、それと対抗するような他の密度推定法を開発する研究にも従事する。

本研究で提案される推定量の性能を、平均2乗誤差または平均積分2乗誤差に基づき議論していく。その基準の0への収束の速さは標本数の逆数に関し4/5乗になる場合が多く、もし4/5より遅い場合には、このような収束比の速さの観点から少なくともベルンシュタイン多項式密度推定より悪いことを意味して、考察対象外になる。研究の初期段階では4/5レベルに焦点をおくが、そのような収束比を改善することも重要で、この数理的関心から本研究をさらに進める。研究経過において、あるノンパラメトリック推定量の収束比が4/5から8/9に改善できたとすれば、そのような推定量を“改良された推定量”と呼ぶ。そのような“良い”密度推定量を提案して、かつ、この方面からノンパラメトリック関数推定の漸近理論を発展させることを目的とする。

3. 研究の方法

(1)概要

研究対象として単変量（またはベクトル値）定常時系列過程のスペクトル密度（行列）や独立同一分布設定における確率密度関数を想定しながら研究を進める中、以下の～のフィードバックを重ねることで、研究の質を高めていった。

数学的基礎研究

数学分野からの関数近似理論の数学的基礎を養って、かつ、統計科学分野からの種々なノンパラメトリック関数推定法の関連領域について十分な文献調査を行う。先行研究で採用されたアイデアの本質を掴み、既存の結果の問題点を整理、あるいは、その拡張を試みる。本研究の中で検討されていく諸問題に対して解決を図る。推定量の性能の比較が平均2乗誤差または平均積分2乗誤差から議論されるため、本研究の初期段階において推定量の漸近バイアス・分散公式などを導出する。ただし、それら公式は数学的に正当化されて議論されなければいけないので、そのための基礎研究も含む。

情報収集

学会・研究集会・ワークショップに参加して最先端の研究動向を掴み、他の研究者と意見交換をする。

数値計算

数値実験を実施し本研究で新たに得られた知見との整合性を確認する、及び、実データへ応用する。

研究成果発表

学会・研究集会・ワークショップで発表して他の研究者から意見を求める。

国際的なジャーナルへ投稿する。

(2)研究経過

まず、単変量（またはベクトル値）の定常な離散時間時系列のスペクトル密度（行列）を推測するため、現代的な『経験尤度法』の文献調査を実施して、この方向からの1つの研究糸口を掴んだ。さらに、一般化経験尤度に基づく検定高次漸近論の研究に従事した（その一部の概要を期間内の学会・研究集会で報告済みである）。また、スペクトル解析にて『一般化経験尤度法』と『一般化積率法』の応用を国際ジャーナルで公表した。

次に、2003-2005の科研費（若手研究）で提案されたベルンシュタイン密度推定量の漸近性能解明で鍵となった、2項分布の確率関数の積の和の近似（独立な格子点分布の和の分布に対するBerry-Esseen型評価法）を再検討し、改良されたベルンシュタイン密度推定量の研究に着手した。バイアス修正法として加法型または積型（従って、非負型）が伝統的なカーネル密度推定の枠組みに対し知られていたものの、積型に際しては2つの推定量の比を厳密に扱う工夫が必要であるため、形式的な計算から予想される平均積分2乗誤差を数学的に正当化し厳密な処理をするのに少し時間を要した。試行錯誤を経て本研究の柱をなす『ベルンシュタイン多項式密度推定量の加法型または積型のバイアス修正』の成果を国際ジャーナルへと公表するに至った。

さらに、ベルンシュタイン多項式密度推定の対抗馬となりうる新しい推定を模索した。ベルンシュタイン多項式は、勿論、有界区間を台とする関数の推定に応用されるが、近年のジャーナルで、むしろ、半無限区間 $[0, \infty)$ で定義される関数の推定問題が盛んに議論されているため、半無限区間 $[0, \infty)$ の場合に対しノンパラメトリックな関数推定の文献調査をまず行った。ガンマ・逆ガンマ・逆ガウス・相反逆ガウス・BS・対数正規カーネルなど非対称カーネルに関わる基礎研究から、ガンマカーネル推定量以外は『境界バイアス問題』、すなわち、(i)相反逆ガウスカーネル推定量は一致がないという致命的な問題、(ii)その他の推定量で、漸近分散が積分可能になるため $f(0)=0$ を課さねばいけないという事実などが判明した。これらの問題点を解決するために、逆ガウス・相反逆ガウス・BSカーネル推定量を再定義した。また、それらを非対称カーネル推定量の族へ一般化した（混合変形ベッセルカーネル推定法として国際ジャーナルで公表した）。平均2乗誤差を形式的に積分すれば平均積分2乗誤差の漸近公式を予想できるが、実際には、それを数学的に正当化しなければならず、従って、証明の完成に相当苦労した（そのような証明方針は、類似の推定法にも適用される見込みがあり、今後の基礎になると考えている）。そして、バイアスを改良する研究も行った（バイアス修正についての研究概要を研究集会で報告済みである）。

4. 研究成果

(1)まず ベルンシュタイン多項式密度推定量の漸近バイスの主要項を定数 a の導入から加法型で消して、その漸近性能を改良した。特に、 $a=2$ の選択の最適性を示した。一方、加法型バイス修正では、推定量の非負性が失われるため、TS-積型バイス修正を考察し、その漸近性能も厳密に調べた。TS-積型は加法型と類似した漸近性能を持ち、従って、その定義に現れる定数は、やはり $a=2$ で最適になることを示した。一方、TS-積型ならば、ベルンシュタイン多項式密度推定量の持つ性質（積分条件）が崩れるため、再規格したTS-積型バイス修正も提案した。このような成果をジャーナル で公表した。

(2) 台が半無限区間 $[0, \infty)$ である場合、ガンマカーネル推定量が先行研究で扱われていたが、その平均積分2乗誤差の不等式を詳しく調べてガンマカーネルの形状パラメータを調整し、平均積分2乗誤差を小さくすることに成功した。一方、ガンマカーネル推定量に対抗する逆ガウス・相反逆ガウス・BS・対数正規カーネル推定量で、境界バイス問題があることを示し、境界バイスのない推定量へ再定義した。特に、それらの再定義された密度推定量は、変形ベッセルカーネルの混合に付随する非対称カーネル推定量の族へと一般化され、そのようなクラスの中で最適な推定量も提示できた。その結果から、収束比 $4/5$ のレベルで漸近同等な推定量を無数に得た。このような研究成果をジャーナル で公表した。

(3) 推測統計の現代的なアプローチとして『経験尤度法とその一般化』を文献調査し、特に、スペクトル解析の一般化経験尤度法と一般化積率法における漸近理論を整備した。その基礎的成果をジャーナル で公表しており、この方面からの高次漸近理論の進展も期待される（なお、パートレット補正可能性などの結果を研究集会で報告済みである）。

(4) 成果(1)を踏まえ、他の密度推定法も考察することで、ベルンシュタイン多項式近似のメリットを引き出すことができると考えた。具体的に、対抗馬としてガンマカーネル推定及び(2)の部分族（混合逆ガウスカーネル族）を考えた場合、加法型とTS・JF-積型修正では $a=1$ とした極限での最適性が示された。この性質は、研究経過において、できる限り一般的な密度推定量の満たすべき十分条件を検討した後、その条件を各々について確認することで得られている（一部の概要を研究集会で報告済みである）。この方面の新しい研究構想に至っており、その他の密度推定量も追加して考察するアイデアがある。そのため文献調査を実施した。

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3件)

Gaku Igarashi and Yoshihide Kakizawa. On improving convergence rate of Bernstein polynomial density estimator. *Journal of Nonparametric Statistics*. 26(1). 61-84. (2014). 査読有.

Gaku Igarashi and Yoshihide Kakizawa. Re-formulation of inverse Gaussian, reciprocal inverse Gaussian, and Birnbaum-Saunders kernel estimators. *Statistics & Probability Letters*. 84(1). 235-246. (2014). 査読有.

Yoshihide Kakizawa. Frequency domain generalized empirical likelihood method. *Journal of Time Series Analysis*. 34(6). 691-716. (2013). 査読有

〔学会発表〕(計15件)

五十嵐岳, 柿沢佳秀. 加重対数正規カーネル密度推定量とその改良について. 研究集会「ノンパラメトリック統計解析とベイズ統計」2014年3月19日. 慶応義塾大学(東京).

Yoshihide Kakizawa. Frequency domain GEL method. (英語) 研究集会「大規模で非定常な時系列・時空間データのモデル化とその推定・検定・予測法の研究」2013年12月5日. 東北大学(宮城).

五十嵐岳, 柿沢佳秀. 逆ガウス, 相反逆ガウス, Birnbaum-Saunders, 対数正規カーネル推定量の再定義について. 研究集会「Seminar on Time Series and Financial Engineering」2013年6月29日. 東京理科大学(東京).

五十嵐岳, 柿沢佳秀. 逆ガウス, 相反逆ガウス, BSカーネル密度推定量の再定義について. 日本数学会2013年度年会. 2013年3月22日. 京都大学(京都).

五十嵐岳, 柿沢佳秀. 逆ガウス, 相反逆ガウス, BSカーネル密度推定量の再定義とその改良について. 研究集会「ノンパラメトリック統計解析とベイズ統計」2013年3月15日. 慶応義塾大学(東京).

五十嵐岳, 柿沢佳秀. 境界バイスのない密度推定量の改良について. 研究集会「統計推測理論の展開と諸モデルへの応用」2012年10月3日. 釧路市生涯学習センター(釧路).

五十嵐岳, 柿沢佳秀. 境界バイスのない密度推定量の改良について. 統計関連学会連合大会, 2012年9月11日. 北海道大学(札幌).

Gaku Igarashi and Yoshihide Kakizawa. Improvement of Bernstein polynomial density estimator. (英語) Korea-Hokkaido University Joint Workshop on Statistics. 2012年7月11日. 北海道大学(札幌).

Yoshihide Kakizawa. Asymptotic expansions for several GEL-based test statistics. (英語) 研究集会「Waseda Statistical Symposium on Time Series and Related Topics」2012年7月6日. 早稲田大学(東京).

Yoshihide Kakizawa. Asymptotic expansions for several GEL-based test statistics. (英語) Seminar. 2012年6月5日. 香港中文大學(香港).

五十嵐岳, 柿沢佳秀. ベータカーネル推定量, ガンマカーネル推定量, ベルンシュタイン推定量の改良について. 研究集会「ノンパラメトリック統計解析とベイズ統計」2012年3月29日. 慶応義塾大学(東京).

柿沢佳秀. Bartlett correctability of empirical likelihood ratio test for a parameter subvector in general estimating equations. 日本数学会 2012年度年会. 2012年3月28日. 東京理科大学(東京).

Yoshihide Kakizawa. Bartlett correctability of empirical likelihood ratio test for a parameter subvector in the over-identified case. (英語) 研究集会「Theory and Applications for Empirical Likelihood and Discriminant and Cluster Analysis」2011年12月3日. 和歌山ビッグ愛(和歌山).

五十嵐岳, 柿沢佳秀. ベルンシュタイン密度推定量への積型バイアス修正法の適用. 統計関連学会連合大会. 2011年9月6日. 九州大学(福岡).

五十嵐岳, 柿沢佳秀. ベルンシュタイン密度推定量への積型バイアス修正法の適用. Summer Workshop on Economic Theory. 2011年8月10日. 小樽商科大学(小樽).

〔図書〕(計0件)

〔産業財産権〕

○出願状況(計0件)

○取得状況(計0件)

〔その他〕

ホームページ等

6. 研究組織

(1) 柿沢 佳秀 (KAKIZAWA Yoshihide)
北海道大学・大学院経済学研究科・教授
研究者番号：30281778

(2) 研究分担者
なし

(3) 連携研究者
なし