

科学研究費助成事業 研究成果報告書

平成 29 年 5 月 15 日現在

機関番号：10101

研究種目：基盤研究(C)（一般）

研究期間：2014～2016

課題番号：26330030

研究課題名（和文）ベルンシュタイン多項式と非対称カーネルに基づく統計推測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 via the Bernstein polynomial/asymmetric kernel 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・教授

研究者番号：30281778

交付決定額（研究期間全体）：（直接経費） 3,600,000円

研究成果の概要（和文）：本研究では非負データの確率密度関数をノンパラメトリックに推定するため、非対称カーネル法に焦点をおいた。先行研究のMIGカーネルを様々な拡張し、その結果、密度生成機と呼ばれる無限次元の関数で規定される「G-MIGカーネル密度推定量」の新しい族を提案した。それらの推定量に対し漸近バイアス・分散・平均積分2乗誤差などの公式を導出して、強一致性及び漸近正規性も証明した。一方、ガンマカーネル推定量、逆ガンマカーネル推定量を含む「Amorosoカーネル密度推定量の族」を提案し、その漸近性能を議論した。さらに、バイアス修正された密度推定量も考察した。漸近理論を確認するシミュレーション実験を実施した。

研究成果の概要（英文）：In this study, we focus on the asymmetric kernel methods to estimate the probability density function for nonnegative data. Especially, we have proposed (i) a family of asymmetric kernel density estimators, depending on the so-called density generator that plays a role of an infinite dimensional parameter, and (ii) a family of Amoroso kernel density estimators, including the gamma/inverse gamma/inverse Gaussian/reciprocal inverse Gaussian/Birnbaum-Saunders/log-normal kernel density estimators as special cases of these families. Also, we have discussed (iii) some bias-reduced density estimators. We have established their asymptotic properties (bias, variance, mean integrated squared error, and so on), as well as the (pointwise) strong consistency and asymptotic normality. The simulation studies have been conducted to confirm our theoretical results.

研究分野：統計科学、数理統計学

キーワード：ノンパラメトリック法 密度推定

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、推測対象とする関数に対し何らかの関数近似理論を応用して統計的な議論を展開していく統計科学の理論の総称である。推測対象の関数に、有限次元の母数モデルを仮定し、その母数を所謂、尤度から推定する『パラメトリック法』及び、一部に有限次元の母数モデルを仮定するが、他方で無限次元構造も許すような『セミパラメトリック法』と並んで3つの柱をなす。

ノンパラメトリック関数推定の先行論文は膨大であり、これを歴史的に振り返れば、独立同一分布設定では確率密度・分布関数・分位点関数が基本的であり、その他に定常な単変量（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り、畳み込み積分以外の近似もしばしば使用された。例えば、十分滑らかな関数のテーラー近似（局所的に多項式で近似して、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ている。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があり、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、カーネル法に基づいたスペクトル密度（行列）推定量の汎関数を主な対象とし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、尤度解析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論文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1つの試みとして、研究代表者が、ベルンシュタイン多項式近似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論文を知り、当時の1つの研究課題（離散時間の定常な時系列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の次数の選択に対しては、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一種のバイアス調整済の情報量規準も議論していた。

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、2011-2013の科研費（基盤研究C）の中で、『（漸近性能の良い）ガンマカーネル密度推定量の再考察』、逆ガウス（IG）・相反逆ガウス（RIG）・バーンバウム・サンダース（BS）カーネルを含むような『MIGカーネル密度推定量族の研究』へ発展させてきた。

2. 研究の目的

本研究では、台を半無限区間（または有界区間）とする密度をノンパラメトリック法で推定する問題を扱う。特に、境界バイアスのない密度推定量の開発に焦点をおき、上述のMIGカーネル密度推定量の族を拡張することを通じて、非対称カーネル推定法を整備することを目的とする。新たに密度推定量を提案し、その漸近性能（主に、平均2乗誤差及び平均積分2乗誤差）を詳細に調べていく（通常そのような基準の0への収束の速さは、標本数 n の逆数の $4/5$ 乗になる場合が多い）。もし $4/5$ より遅い場合は、少なくともベルンシュタイン多項式密度推定より漸近性能が悪いため研究の対象外となるが、そのような収束比が改善可能であるかを検討していくことも重要である。研究の初段階で、まず $4/5$ レベルを念頭にするが、この収束比をさらに改善することも重要で、このような数理的な関心から研究を進める。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、もし $4/5$ から $8/9$ に改善できたときには、そのような推定量を

“改良された推定量”と呼ぶ(この事実は、先行論文にて古くから扱われた推定密度の台がRである場合の所謂、標準的なカーネル推定量に相当するが、本研究では推定密度の台が半無限区間と仮定して、境界バイアスのない推定量を対象とすることに注意)。そのような改良された密度推定量を複数提案し、その統計的な性質を解明していき、この方面からノンパラメトリック関数推定の理論を進展させて、その応用を議論する。

3. 研究の方法

(1)概要

台を半無限区間(または有界区間)とする密度関数の推定に関する研究を進めていく際、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対し十分な文献調査を行った。先行研究で採用されたアイデアの本質を先ず掴み、既存の結果の問題点を整理、あるいは、その拡張を試みて、本研究の中で随時検討されていく問題に対し解決を図った。推定量の性能比較を平均2乗誤差または平均積分2乗誤差で議論するから、研究の初期段階では、まず、推定量の漸近バイアス・分散公式などを導出した。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まれた。

情報収集

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

数値計算

数値実験を実施し、本研究で得られた知見との整合性を確認した。

研究成果発表

学会・研究集会・ワークショップにて成果を発表した。

国際的なジャーナルへ投稿した。

(2)研究経過

まず、研究代表者の先行研究で扱われた「ガンマカーネル密度推定量」「再定義した逆ガウス(I G) 相反逆ガウス(R I G) バーンバウム・サンダース(B S)カーネル密度推定量」及び「パラメータを $\pm 1/2$ とする重み付き対数正規カーネル密度推定量」の平均積分2乗誤差の収束比 $4/5$ を改良する加法型・積型のバイアス修正された密度推定量を提案した。その平均積分2乗誤差の収束比が $8/9$ に改良されるという結果を専門雑誌へ投稿した(その改訂作業を経て、掲載済み)。なお、積型については、T S型とJ F型の2つを考察しており、かつ、そのパラメータ a を $a \rightarrow 1$ とする極限による密度

推定量については、当初ガンマカーネル密度推定量のみ、漸近性能の解明に成功していたが、最終的にはI G・R I G・B Sカーネルを含むようなM I Gカーネル密度推定量族、パラメータを $\pm 1/2$ とする重み付き対数正規カーネル密度推定量に対しても証明を完成させた。

次に、新しい非対称カーネル密度推定量を提案するため、次の3点について文献調査を行い、ノンパラメトリック推論に応用可能な分布論を整備した。

(i) 正規ベースの従来のB Sから対称分布ベースのB S化への拡張。

(ii) B S分布と対数正規分布とを母数 q で繋いだ分布族への拡張。

(iii) 中心ケースから非心B S化への拡張。

その後、これらのB S分布論としての多義的な拡張をノンパラメトリックな密度推定に応用して「一般化B Sカーネル密度推定量」という密度推定量族を提案し、その漸近性能(漸近バイアス・分散公式、平均2乗誤差・平均積分2乗誤差の公式、平均積分絶対誤差のバウンド評価、強一致性、漸近正規性)を解明した。そのような研究概要を学会、研究集会で報告した。なお、これらは の研究へさらに発展している。

先行論文で、逆ガンマカーネル密度推定量が議論されていたが、その定義式をみれば、 $x=0$ で推定量が0になるという制約があり、推定密度 f が $f(0)=0$ か、 $f(0)>0$ を満たすかに関わらずに、漸近性能が機能する「再定義された逆ガンマカーネル密度推定量」を提案して、その漸近性能を解明した。この成果を専門雑誌に投稿した(その改訂作業を経て、掲載確定)。

の研究の一部に対して、多次元化も考察し、多変量B Sカーネル密度推定量の提案とその漸近性能を解明した。多変量一般化B Sカーネル密度推定量の提案とその漸近性能を解明した。条件付き密度推定量への応用も議論した。このような研究概要を研究集会で報告した。

ガンマカーネル密度推定量と の研究で再定義した、逆ガンマカーネル密度推定量は、Amorosoカーネルによる定式化が可能であり、「Amorosoカーネル密度推定量の族」として統一的に扱うことに成功した。平均積分2乗誤差の収束比が $4/5$ であることを示して、その収束比を改善するために、加法型・積型のバイアス修正された密度推定量を提案し、平均積分2乗誤差の収束比が $8/9$ であるという結果を専門雑誌へ投稿した。さらに、パラメータ a を $a \rightarrow 1$ とする加法型・積型の極限による密度推定量について、同様の結果を得ており、そのような研究概要を研究集会で報告した。

の研究の流れから、非対称カーネル法をさらに発展させるため、I G・R I G分布論を整備した。

(i) 正規ベースの従来のI G・R I Gから

対称分布ベースのIG・RIG化への拡張。
(ii) IG・RIGと対数正規分布とを母数qで繋いだ分布族への拡張。

その後、これらと研究のBSを含むようなq-GMIGカーネルへ拡張し、これを応用することから、「q-GMIGカーネル密度推定量」という広範な推定量の族を提案し、漸近性能を解明した。そのような成果の研究概要を学会、研究集会で報告し、専門雑誌に投稿した。さらに、研究におけるBS化に対し、対称分布をベースにする必要がなく、任意分布をベースにしたBS分布論という発想の転換に至った。その結果、(iii)での非心分布をベースするBS化以外で、歪分布(Azzalini型・ツーピース化)について文献調査した(それを使えば「パラメトリック・ノンパラメトリック推論」の方面から今後の成果が期待される)。

4. 研究成果

(1) 先行研究で扱われた密度推定量を

- (i) ガンマカーネル密度推定量、
- (ii) 逆ガウス(IG) 相反逆ガウス(RIG)、
バーンバウム・サンダース(BS)カーネル
密度推定量を含むMIGカーネル推定量族、
- (iii) パラメータを $\pm 1/2$ とするような
重み付き対数正規カーネル密度推定量

に分類した。これらに対し平均積分2乗誤差の収束比を $8/9$ へと改良する加法型・積型のバイアス修正された、新しい密度推定量を提案した。ここで、加法型では密度推定量の非負性を犠牲にするのに対し、積型は非負性を保つ特徴があって、特にTS型とJF型の2つを考察した。加法型及びTS・JF型は、それぞれパラメータaで定式化されており、 $a=1$ として得られる極限による密度推定量についても議論した。最終的に、パラメータaの選択について、加法型とTS型では、 $a=1$ が最良であることを示した。一方、TS型とJF型は $a=1$ で両者が一致して、その結果、JF型でパラメータaを動かしたときには、平均積分2乗誤差の基準で、どのaを用いたTS型よりも優越することを示した。

(2) 新しい非対称カーネル密度推定量を提案するために、ノンパラメトリック推論に応用可能な分布論を

- (i) 正規ベースの従来BSから対称分布ベースのBS化への拡張、
- (ii) BS分布と対数正規分布とを母数qで繋いだ分布族への拡張、
- (iii) 中心ケースから非心BS化への拡張、
- (iv) (iii)以外で、柔軟な非対称分布ベースのBS化への拡張

の4点から整備した。特に、(iv)に対しては、歪分布(Azzalini型、及び、ツーピース化)の文献調査を実施した。これらBS分布論としての多義的な拡張をノンパラメトリックな密度推定に応用して「一般化BSカーネル密度推定量」という密度推定量族を提案した。このような族は、原点対称分布を定式化する

「密度生成機g」という関数(無限次元)が関与して、(1)の(ii)(iii)のような密度推定量族から「セミパラメトリック構造」をもつ密度推定量の体系への進捗が特筆すべきである。これらの推定量の漸近性能(漸近バイアス・分散公式、平均2乗誤差・平均積分2乗誤差の公式、平均積分絶対誤差のバウンド評価、強一致性、漸近正規性)を解明した。さらに、交差検証法による平滑化パラメータを使い、提案された推定量の小標本特性も数値実験で調べた。なお、(iv)に基づいた今後の成果が期待される。

(3) 先行論文の、逆ガンマカーネル密度推定量の欠点($x=0$ で推定量が0になるという制約があり、推定密度fが $f(0)=0$ か、 $f(0)>0$ を満たすかに関わらずに、漸近性能が機能するようにすべきである)を踏まえて、「再定義した逆ガンマカーネル密度推定量」を提案し、その漸近性能を調べた。(1)で扱われていた「ガンマカーネル密度推定量」と、ここでの研究成果「再定義された、逆ガンマカーネル密度推定量」はAmorosoカーネルから定式化できて、「Amorosoカーネル密度推定量の族」として統一的に扱うことに成功した。さらに、平均積分2乗誤差を改良する研究も行い、(1)で得られた成果を、Amorosoカーネル密度推定量の族へ拡張した。ただし、特別な場合であるガンマカーネル密度推定量は、例外的に扱いやすかったという事実も露呈されたが、試行錯誤を経て、バイアス修正されたAmorosoカーネル密度推定量に対し漸近性能を導いた。

(4)(2)の研究の一部に対し、多次元化も考察し、多変量BS(一般化BS)カーネル密度推定量の提案とそれらの漸近性能の解明、条件付き密度推定量への応用も議論した。

(5)(2)の研究の流れから、非対称カーネル法を発展させるため、IG・RIG分布論を

- (i) 正規ベースの従来IG・RIGから対称分布ベースのIG・RIG化への拡張、
- (ii) IG・RIGと対数正規分布とを母数qで繋いだ分布族への拡張、
- (iii) これらと研究(2)のBSを含むようなq-GMIGカーネルへの多義的な拡張の3点から整備した。これらに応用することから、(2)を発展させて、セミパラメトリック構造をもつ密度推定量の体系「q-GMIGカーネル密度推定量」という広範な推定量の族を提案し、その漸近性能を解明した。

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 2件)

Kakizawa, Y. and Igarashi, G. Inverse gamma kernel density estimation for nonnegative data. To appear in Journal of the Korean Statistical Society. (2017). 査読有.

Igarashi, G. and Kakizawa, Y. Bias corrections for some asymmetric kernel estimators. Journal of Statistical Planning and Inference.159(1).37-63. (2015).査読有.

〔学会発表〕(計 10件)

柿沢佳秀.境界バイアスのない、対称分布をベースにしたq-MIGカーネル密度推定量とその関連した話題について.ノンパラメトリック統計解析とベイズ統計.2017年3月28日.慶應義塾大学(東京).

五十嵐岳,柿沢佳秀.Amorosoカーネル密度推定量とその改良について.ノンパラメトリック統計解析とベイズ統計.2017年3月28日.慶應義塾大学(東京).

柿沢佳秀.対称分布をベースにした逆ガウス・相反逆ガウスカーネル確率密度推定法.統計関連学会連合大会.2016年9月6日.金沢大学(金沢).

五十嵐岳,柿沢佳秀.Amorosoカーネルを用いたノンパラメトリック密度推定について.統計関連学会連合大会.2016年9月7日.金沢大学(金沢).

Kakizawa, Y. and Igarashi, G. Some boundary-bias-free density estimators. The 4th Institute of Mathematical Statistics: Asia Pacific Rim Meeting.2016年6月28日.香港中文大学(香港).

柿沢佳秀.バーンバウム・サンダース分布の種々の一般化と密度推定への応用.ノンパラメトリック統計解析とベイズ統計.2016年3月29日.慶應義塾大学(東京).

柿沢佳秀.Generalized Birnbaum-Saunders kernel density estimator.日本数学会.2016年3月19日.筑波大学(筑波).

Kakizawa, Y. and Igarashi, G. Recent developments in boundary-bias-free asymmetric kernel density estimation. Recent Progress in Time Series and Related Fields. 2015年12月11日.東北大学(仙台).

柿沢佳秀. Some multivariate asymmetric kernel estimators for density functions.大規模複雑データの理論と方法論:最前線の動向.2015年11月17日.筑波大学(筑波).

五十嵐岳,柿沢佳秀.境界バイアスのない非対称カーネル推定量の最近の進展.日本統計学会春季集会.2015年3月8日.明治大学(東京).

〔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)研究代表者

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

(4)研究協力者

なし