

令和 5 年 5 月 21 日現在

機関番号：10101

研究種目：基盤研究(C) (一般)

研究期間：2020～2022

課題番号：20K11700

研究課題名(和文)非対称カーネルを用いたノンパラメトリック推測とその応用に関する研究

研究課題名(英文)A study on nonparametric inference using asymmetric kernel and its application

研究代表者

柿沢 佳秀(Kakizawa, Yoshihide)

北海道大学・経済学研究院・教授

研究者番号：30281778

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：境界バイアスのない非対称カーネル推測法の漸近理論に焦点を置いた。具体的には、単変量非負データ(または多次元の第1象限データ)の密度関数を推定するために利用可能なカーネル族としてバーンバウム・サンダース(BS)型カーネル(またはその多次元版)を開発した。個別的な密度推定量の考察ではなく、柔軟なBS型カーネルを用いた非再帰的または再帰的密度推定量として体系的に整備し、その漸近的性質を解明した。これらの新展開として、バイアス・サンプリングを伴う状況下の密度推定問題、並びに、密度高階微分推定の研究も開始した。また、数値実験から漸近性能を確認した。

研究成果の学術的意義や社会的意義

ノンパラメトリック関数推定の題材において、非負データの場合に起こりうる境界バイアスを回避する1つの策として、非対称カーネル法の研究に従事した。個別的な非対称カーネルではなく、カーネル族としての体系化を念頭に置いており、(i)単変量に留まらず、多変量データに適用可能な相関構造も考慮したカーネル族の提案、(ii)逐次的にデータが得られるような場合、計算面で有利な再帰的な非対称カーネル法の提案を含む。また、(iii)バイアス・サンプリングを考慮した推測問題、及び、(iv)密度推定に留まらず、関連した種々の関数推定問題の研究を開始しており、今後、これら方面からの展開が期待される。

研究成果の概要(英文)：In this study, we mainly focus on asymptotic theory of boundary-bias-free asymmetric kernel method. Especially, we have proposed (i) a family of Birnbaum-Saunders (BS) type kernels to estimate the probability density function of univariate (or multivariate) nonnegative data and (ii) non-recursive or recursive asymmetric kernel density estimator using the BS type kernels. We have established some desirable asymptotic properties (bias, variance, mean squared error, mean integrated squared error, strong consistency, and asymptotic normality) of various proposed density estimators. Furthermore, we have started with new studies on (iii) density estimation under a certain biased sampling scheme and (iv) higher-order density derivatives estimation of nonnegative data. We also have conducted the simulation studies to confirm some 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を経て、統計的推測問題を議論する統計学の1つの体系である。推測対象の関数に、有限次元の母数モデルを仮定し、その母数を例えば、尤度原理(すなわち、尤度関数最大化)から推定する『パラメトリック法』及び、両者の中間に位置づけられる『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法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・分位点関数の推測が基本的であり、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、回帰分析における回帰関数、生存時間分析におけるハザード比関数などの推測も議論され発展してきた。伝統的なカーネル平滑化法の背後には推定するターゲット関数をスケール化された対称カーネル関数による畳み込み積分としてスケール(バンド幅)を十分小さく選んで近似するアイデアがあり、畳み込み積分以外の方法もしばしば使用された。例えば、十分に滑らかな関数のテラー近似(推定位置の近くでテラー多項式近似をして、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、それら係数を推定するような定式化)、スプライン基底関数近似、ベルンシュタイン多項式による有界閉区間上の連続関数の近似などである。統計的な推測理論において、観測されたデータからどのようにして有効に推測できるか?及び、その限界解明が重要であり、通常、標本数を無限大にした『統計漸近理論』による数学的な基礎研究を行う。ノンパラメトリック推測法には、統計的な漸近性能を決めるような『平滑化パラメータ』があり、関数近似誤差(統計学の専門用語では推定量の『バイアス』に相当する)と推定量の『分散』とのトレードオフ関係を制御できる。このようなノンパラメトリック推測法の研究は、国内・国外で著しい進展がみられ、現在でも統計科学の重要な1つのテーマである。

研究代表者は1990年代後半、カーネル型のスペクトル密度(行列)推定量の汎関数に対し、いくつかの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で判別・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・検定問題へ応用することで、尤度解析と漸近同等なものを提示した。

研究代表者がベルンシュタイン多項式近似法を応用する研究を開始したのは、2003-2005の科学研究費(若手研究)の頃で、確率分布・密度関数をベルンシュタイン多項式近似に基づいてノンパラメトリックに推定するという、ある先行論文を読んだことが動機であった。当時の1つの研究課題(定常時系列のスペクトル密度(行列)推定)に対し、カーネルスペクトル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考察し、2008-2010の科学研究費(基盤研究C)の中で、一般化ベルンシュタイン多項式近似も応用した。他方で、確率密度推定の境界バイアス問題の研究話題への関心が高まっていき、2011-2013の科学研究費(基盤研究C)の研究成果の1つとして、バイアス修正されたベルンシュタイン多項式密度推定量を提案し、その漸近性能を解明した。その頃、異なるアプローチである『非対称カーネル推定法』に着手する構想を持ち、その方向からの研究へと転換し、特に、ガンマカーネル密度推定量の再考察、さらに、逆ガウス(IG)・相反逆ガウス(RIG)・バーンバウム・サンダース(BS)カーネルの3つを含むような、MIGカーネル密度推定量族の提案をした。2014-2016の科学研究費(基盤研究C)の研究成果の中でバイアス修正されたMIGカーネル密度推定量族をベルンシュタイン多項式密度推定量の対抗馬として提案した。

その後の2017-2019の科学研究費(基盤研究C)を経て、非対称カーネル法の漸近論を遂行する際に、不可欠となる非対称カーネルの持つべき性質が解ってきた。カーネル毎の議論でなく、柔軟な(様々の意味から)一般化された非対称カーネル族へと体系化していくアイデアに至り、その発想は本研究の骨格をなす。

2. 研究の目的

本研究では、台を半無限区間(または多次元の第1象限)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、密度推定に焦点を置いて、非負データに利用可能な非対称カーネル族を開発し、個別的な議論でなく非対称カーネルによる推測理論を整備して体系化する。研究において新たに密度推定量を次々と提案する。それらの漸近性能(主に、平均2乗誤差及び平均積分2乗誤差)を数学的な基礎研究から詳細に調べる。通常、そのような基準の0への収束比は、標本サイズの逆数の4/5乗になる場合が多い(なお、もし4/5より遅い場合は、少なくとも上述のベルンシュタイン多項式密度推定量より漸近性能が悪く研究の対象外となる)ので、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が、本研究では、4/5レベルを達成できるような、非対称カーネル族の構築を念頭に置いている。このような問題意識から、柔軟な(様々の意味から)一般化された非対称カーネル族を構築し、密度推定量を複数提案して、それらの性質(漸近バイアス・分散公式、平均2乗誤差・平均積分2乗誤差の公式の導出、平均積分絶対誤差のバウンド評価、強一致性、漸近正規性の証明など)を数学的な基礎研究から解明する。なお、本研究で焦点を置くChen以降の非対称カーネル法の研究とは、カーネル毎の個別的な成果でなく、研究代表者による(I)Amorosoカーネルあるいは

(II)正規分布ベースで構成された従来の BS 型カーネル族を進展させていくことを目的として、境界バイアスのないノンパラメトリック法を様々な推測問題へ応用する。

3. 研究の方法

(1)概要

台を半無限区間（または多次元の第 1 象限）とするような密度関数のノンパラメトリック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から、本研究の質を高めていった。

数学的基礎研究

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て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 乗誤差や平均絶対誤差とその積分基準（平均積分 2 乗誤差、平均積分絶対誤差）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。それら公式を導くには、非対称カーネルに付随する分布論の研究（具体的には、多変量解析から楕円分布論）の文献調査が含まれ、本研究のノンパラメトリック推測法の漸近理論で適用可能な、密度生成機に不可欠な条件も検討した。さらに、対称分布に歪みを加えて生成される『歪分布論』の文献調査も行い、この発想も本研究に取り入れた。

情報収集

（コロナ渦のために）学会・研究集会・ワークショップへの参加の多くは、zoom（非対面）であったが、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

得られた成果の多くは、（コロナ渦のため）学会・研究集会・ワークショップにて zoom（非対面）で発表した。

国際的なジャーナルへ投稿した。

(2)研究経過

非対称カーネル族を様々な意味で拡張して開発していくため、次の 2 点について文献調査を行い、ノンパラメトリック推論への応用を見据えて、BS 型分布について整理し、分布論の整備をした。

- (i)中心ケースから非心ケースによる BS 化。
- (ii)対称分布ベースから歪分布ベースによる BS 化。

これらの方向で、BS 分布の多義的な拡張をノンパラメトリック密度推定に応用することから、歪分布ベースの柔軟な BS 型カーネル族を構築し、それに基づいた密度推定量族の提案をして、漸近的性能（漸近バイアス・分散公式、平均 2 乗誤差・平均積分 2 乗誤差公式、強一貫性、漸近正規性など）を解明した。その成果は専門雑誌 Computational Statistics and Data Analysis へ掲載した。

スカラーの非負データにおける密度推定をするための BS 型分布論は多次元化も可能である。実際、で提案した BS 型カーネルの積を作れば、いわゆる、積型の非対称カーネル密度推定量になって、その漸近論は を多次元化できるのだが、実は多次元の場合、多変量正規分布ベースから、楕円分布ベースによる分布論を誘導し、多変量分布論では必然的に従属性を持たせることを意味した。この事実と合わせ、非積型の非心 BS カーネル密度推定量族を提案するに至った。相関構造を考慮した非積型の BS カーネルを使えば、積型に比べて平均積分 2 乗誤差が漸近的に小さくなりうることを解明した。その成果は多変量解析の専門雑誌 Journal of Multivariate Analysis へ掲載した。

当該分野では、非対称カーネル密度推定の定式化が、“非再帰的”になされていたが、これではデータが逐次得られるような場合に計算コストの意味で不利であり、計算時間の軽減のため再帰的非対称カーネル密度推定量へ転換した。平均積分 2 乗誤差による漸近相対効率は従来の非再帰的カーネル密度推定量に比べて僅かに劣る欠点はあるものの、信頼区間の幅の期待値は

再帰的に構築すれば漸近的に短くなることが解明された。その成果はノンパラメトリック解析の専門雑誌 Journal of Nonparametric Statistics へ掲載した。

推測の諸問題はサンプリング状況に応じたバリエーションがあって、直接サンプリングではなく、いわゆる、レングス・バイアスド・サンプリングの状況における境界バイアス問題を回避した密度推定量の開発も重要である。そのため、この方針からの非対称カーネル法へと転換し、この研究を開始している（一部の研究成果は、国際会議、学会、研究集会で講演した）。

境界バイアスを考慮した非対称カーネル法は、密度関数に留まらず、分布関数・密度高階微分・分位点関数・ハザード関数など統計科学で推定対象とする何らかの意味からの関数を推測する問題へと拡張される。実数全体を台とする場合の従来位置・尺度型カーネルも含めて文献調査を行ったところ、分布関数推定の枠組みにて非対称カーネルを応用した先行研究があり、そこでは 2 種類の定式化が行われていた。他方、密度関数の 1 階導関数の推定の先行研究では、分布関数に対する 1 つの定式化がガンマカーネルでなされていたが、本研究ではもう 1 つの定式化をすることからも密度微分推定量が提案でき、その漸近論は研究と同じ土俵で議論ができることと解った。1 階微分だけでなく、任意階数について高階微分の推定の題材も見据え、この研究を開始している（一部の研究成果は、学会、研究集会で講演した）。

4. 研究成果

(1)半無限区間を台とする単変量の密度推定において、境界バイアスのないノンパラメトリックな密度推定に応用可能な非対称カーネルを、個別的にではなく、研究代表者の先行研究である、既存のカーネル族(対称分布ベースの中心BSカーネル族)を含むような拡張された新しい族へと発展させた。特に、本研究の成果として、歪分布ベースの非心BSカーネル族を用いた非対称カーネル推定法の開発をし、これら推定量の漸近性能(漸近バイアス・分散公式、平均2乗誤差・平均積分2乗誤差公式の導出、平均積分絶対誤差のバウンド評価、強一致性、漸近正規性の証明)を数学的に解明して、提案された多くの密度推定量の小標本特性も膨大な数値実験で検証した。なお、ここで提案されたBS型カーネル族は、密度生成機と呼ばれる無限次元の関数に依存し、柔軟な族を形成できていることが特徴であり、密度生成機を正規生成機にすれば、従来のMIG(この特別な場合としてIG、RIG、BS)カーネルが得られ、さらに、これを1つの母数で対数型カーネルへと繋げることもできる。非心パラメータあるいは歪みパラメータを選択することから、平均積分2乗誤差を漸近的に小さくできることを示した。

(2)単変量の密度推定から多変量の密度推定へ進展させた。(1)のような単変量の下で開発されたカーネルの積を作る方法(積カーネル法)とは別に、非積型カーネルを開発した。分布論の観点に立てば、多変量正規分布ベースから楕円分布ベースにすることで、必然的に従属性を持たせることになり、従って、柔軟な多変量BS分布論を誘導できたことを意味して、この拡張は十分な意義がある。具体的には、楕円分布ベースのBSカーネル密度推定量族を提案し、その漸近性能を示した。特に、相関構造を有するような非積型カーネルにより、従来密度推定量に比べて、平均積分2乗誤差が漸近的に小さくなることを解明した。

(3)従来の非対称カーネル密度推定量は“非再帰的”に定式化がされていたから、その新展開として、再帰的な非対称カーネル密度推定量にも取り組み、その漸近性能を示した。具体的には、平均積分2乗誤差による漸近相対効率は従来の非再帰的カーネル密度推定量に比べ僅かに劣る欠点はあるものの、再帰的な定式化は、追加データが逐次、利用出来る場合には計算時間軽減の利点があり、かつ、信頼区間を構成するとき信頼区間幅の期待値は推定量を再帰的に構築すれば漸近的に短くなることを解明した。

(4)レングス・バイアスド・サンプリングの状況における、境界バイアスのない非対称カーネル密度推定量の研究を開始した。レングス・バイアスド・サンプリングは、重み付き密度の1例であって、サンプリングにはバリエーションがあるので、ノンパラメトリック関数の推測に対して境界バイアスのない推定量を様々なサンプリング状況毎に開発し、その漸近論を展開していくことから、この方面の研究進展が期待できる。

(5)密度推定に留まらず、分布関数・密度高階微分・分位点関数などの文献調査をし、特に、密度高階微分推定の方面からの研究に転換した。具体的には、非負データに対する、ガンマカーネルを用いた密度の1階導関数に関する先行文献があったが、その漸近論ではガンマ関数の高階微分に現れるポリガンマ関数が関与し、漸近バイアス導出は極めて見通しの悪いもので、先行研究で提案された密度微分推定量の漸近論は完成といえないのではないかと(非統一的な導出)と判明した。他方、本研究では、任意階数の高階密度微分推定量も見据えて、非対称カーネル法に基づいた密度微分推定量の構成からやり直し、研究を進めており、(1)~(3)のような密度推定の研究成果に並行した、高階微分バージョンとして体系化でき、理論整備ができるだろうという見通しから、この方面での研究進展が期待できる。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 Yoshihide Kakizawa	4. 巻 187
2. 論文標題 Multivariate elliptical-based Birnbaum-Saunders kernel density estimation for nonnegative data	5. 発行年 2022年
3. 雑誌名 Journal of Multivariate Analysis	6. 最初と最後の頁 1-20
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.jmva.2021.104834	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Yoshihide Kakizawa	4. 巻 33
2. 論文標題 Recursive asymmetric kernel density estimation for nonnegative data	5. 発行年 2021年
3. 雑誌名 Journal of Nonparametric Statistics	6. 最初と最後の頁 197-224
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/10485252.2021.1928120	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Yoshihide Kakizawa	4. 巻 -
2. 論文標題 A class of Birnbaum-Saunders type kernel density estimators for nonnegative data	5. 発行年 2021年
3. 雑誌名 Computational Statistics and Data Analysis	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.csda.2021.107249	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計10件（うち招待講演 2件/うち国際学会 1件）

1. 発表者名 Yoshihide KAKIZAWA
2. 発表標題 Asymmetric kernel density estimation for biased data
3. 学会等名 EcoSta2022 [zoom]（招待講演）（国際学会）
4. 発表年 2022年

1. 発表者名 柿沢佳秀
2. 発表標題 境界バイアスを回避する非対称カーネル法の最近の展開
3. 学会等名 統計関連学会連合大会 [zoom] (招待講演)
4. 発表年 2022年

1. 発表者名 柿沢佳秀
2. 発表標題 非対称カーネルを用いた確率密度関数の微分推定について
3. 学会等名 日本数学会秋季総合分科会
4. 発表年 2022年

1. 発表者名 柿沢佳秀
2. 発表標題 境界バイアスを回避する非対称カーネル法の様々な適用
3. 学会等名 ノンパラメトリック統計解析とベイズ統計
4. 発表年 2023年

1. 発表者名 柿沢佳秀
2. 発表標題 リングス・バイアスデータに対する非対称カーネル密度推定
3. 学会等名 2021年度統計関連学会連合大会 [zoom]
4. 発表年 2021年

1. 発表者名 柿沢佳秀
2. 発表標題 レンクス・バイアスデータに対する非対称カーネル密度推定について
3. 学会等名 日本数学会2021年度秋季総合分科会(コロナで学会は中止だが, 講演成立)
4. 発表年 2021年

1. 発表者名 柿沢佳秀
2. 発表標題 レンクス・バイアスド・サンプリングにおける非対称カーネル密度推定
3. 学会等名 ノンパラメトリック統計解析とベイズ統計 [zoom]
4. 発表年 2022年

1. 発表者名 柿沢佳秀
2. 発表標題 多変量バーンバウムサンダース型分布と密度推定への応用
3. 学会等名 日本数学会2020年度秋季総合分科会(熊本大; コロナで学会は中止だが, 講演成立)
4. 発表年 2020年

1. 発表者名 柿沢佳秀
2. 発表標題 非負データの密度関数推定について(分布論を中心に)
3. 学会等名 統計多様体の幾何学とその周辺 [zoom]
4. 発表年 2020年

1. 発表者名 柿沢佳秀
2. 発表標題 非負データの密度関数推定(BS型の分布論を中心に)
3. 学会等名 ノンパラメトリック統計解析とベイズ統計 [zoom]
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------