

平成 29 年 6 月 15 日現在

機関番号：12102

研究種目：研究活動スタート支援

研究期間：2015～2016

課題番号：15H06068

研究課題名（和文）境界問題のないノンパラメトリックな多次元密度推定とその応用について

研究課題名（英文）Nonparametric multivariate boundary-bias-free density estimation and its application

研究代表者

五十嵐 岳（IGARASHI, Gaku）

筑波大学・システム情報系・助教

研究者番号：40759346

交付決定額（研究期間全体）：（直接経費） 1,300,000円

研究成果の概要（和文）：推定する密度の台に境界がある場合のノンパラメトリック密度推定に焦点を置いて研究を行い、多変量対数正規密度をカーネルとして多変量対数正規カーネル推定量を提案した。さらに重み付き分布の発想から、重み付き多変量対数正規カーネル推定量へと一般化し、その漸近的性質を導出した。また、逆ガンマやその一般化であるAmorosoカーネルを用いた密度推定量も提案し、漸近的性質を導出した。さらに、ベータ、Amorosoカーネル密度推定量のバイアス修正を行った。得られた結果を検証するため、数値実験や実データ解析を行った。

研究成果の概要（英文）：We focused on nonparametric density estimation when the support of the estimated density has the boundary. We proposed multivariate log-normal kernel density estimator and its generalization, using the idea of the weighted distribution, and derived their asymptotic properties. We also proposed inverse gamma and Amoroso kernel density estimators, and derived their asymptotic properties. Furthermore, we discussed the bias-reduced beta and Amoroso kernel density estimators. Simulation studies and data analyses were conducted to illustrate the asymptotic results.

研究分野：統計科学

キーワード：ノンパラメトリック密度推定

1. 研究開始当初の背景

統計科学における確率密度関数(以降、『密度』と呼ぶ)の推定方法としては、密度に母数モデルを仮定し、母数を推定するパラメトリック法と、母数モデルを仮定しないノンパラメトリック法が主流である。ノンパラメトリック密度推定の歴史は長く、カーネル推定(以降、『通常のカーネル推定』と呼ぶ)は、原点对称なカーネル関数を用いる推定法で、ノンパラメトリック回帰などにも応用され、ノンパラメトリック関数推定の根幹として著しい発展を遂げ、現在も進展がみられる。

密度をデータから推定するとき、性能を『平均2乗誤差』(バイアスの2乗と分散の和に等しい)または『平均積分2乗誤差』(平均2乗誤差を積分した基準)で測る。統計科学では、データのサイズが大きい場合の『漸近理論』を考え、それらの基準が速く0に収束して、より小さい方が望ましい(以降、この収束の速さを『収束比』と呼ぶ)。通常のカーネル推定の収束比を改良するため、加法型または非負型バイアス修正が考案されている。加法型バイアス修正は、密度の非負性を犠牲にし、一方、非負型バイアス修正は、密度の積分条件(積分して1になる性質)を犠牲にする欠点がある。

ところで、通常のカーネル推定は、推定対象の密度の台が有界区間や半無限区間の場合、境界付近でバイアスがあって、平均2乗誤差も0に収束しない問題(以降、『境界バイアス問題』と呼ぶ)がある。この境界バイアス問題に対し、ベルンシュタイン推定などが提案され、近年になり、非対称カーネルを用いた推定(以降、『非対称カーネル推定』と呼ぶ)がいくつか提案されている。これらの推定でも収束比の改良は重要であるが、あまり知られていなかった。そこで、研究代表者はこれまで、ベルンシュタイン推定やいくつかの非対称カーネル推定を改良する研究を行ってきたが、1次元の密度を推定する研究であった。また、非対称カーネル推定の先行研究も1次元の密度を推定するものばかりである。しかし、現実のデータは多次元で観察されることが多く、応用面から、多次元の密度を推定できるようにすることは重要である。通常の(対称な)カーネル推定による多次元の密度推定としては、1変量のカーネルを次元数だけ掛け合わせる方法があるが、多次元のカーネル推定では、1次元の場合と同様に境界バイアス問題が存在する。

1変量のカーネルの積ではなく、多変量正規密度などをカーネルとして多次元の密度を推定する方法も知られている。しかし、非対称カーネル推定において、そのような積型でないカーネルを応用した先行研究は知られておらず、積型でないカーネルを応用することで平均積分2乗誤差を小さくして推定の性能を良くできる可能性があり、従って、境界バイアス問題のない非対称な多次元の密度推定を構成するという研究の着想に至っ

た。さらに、『次元の呪い』と呼ばれるように、次元数が増えるほど収束比が遅くなる現象が知られ、次元の問題に直面するため、収束比の改良が重要になる。

2. 研究の目的

多変量の非対称密度をカーネルとするような境界バイアス問題のない多次元の非対称カーネル推定を開発し、その数理を明らかにすることを目的とする。このような発想は、先行研究に存在する、1変量のベータ、ガンマカーネルを次元数だけ掛け合わせるといような発想とは異なる。さらに、密度推定について収束比を改良することをめざす。次に、ノンパラメトリックな密度推定法には、バイアスと分散のトレードオフを制御するパラメータ(以降、『平滑化パラメータ』と呼ぶ)が存在し、実際の応用に際して平滑化パラメータを選択する問題がある。平滑化パラメータの選択は、推定の性能に大きく影響するため、推定の性能を最大限に発揮させる平滑化パラメータの選択法を開発する。これまで1次元で考察された非対称カーネル推定が、多次元の密度推定に拡張されると、応用の幅が格段に広がると考えられる。最終的には、通常は正値な多次元のデータを扱うため、1次元の非対称カーネル推定だけでは応用が難しかったノンパラメトリック法による判別・クラスター分析のような多変量解析に対しても、非対称カーネル推定を応用することを考える。

3. 研究の方法

(1) 非対称な多変量密度をカーネルとして、新しい多次元の非対称カーネル推定を構成する。具体的には、多変量密度を非対称カーネルとして用いることを考えている。多変量密度を非対称カーネル推定に応用する際、各々のパラメータの取り方が重要で、推定する点でカーネルが(正確な、もしくは漸近的な)最頻値を持つように定式化をする。さらに、1次元の非対称カーネル推定には、境界で正の値をもつ密度を推定する場合、境界付近でまともに推定できないものがあったので、多次元の非対称カーネル推定でも、そのような問題が起こらないようにパラメータを精査しなければならない。このような試行錯誤をして、種々の推定の漸近性能を数学的基礎研究で明らかにする。具体的には、多次元の非対称カーネル推定の漸近的性質(バイアス、分散、平均2乗誤差、平均積分2乗誤差、漸近正規性など)を導出していく。乱数によるシミュレーション実験を実施して、得られた諸結果の理論を検証する。

(2) 非対称カーネル推定について収束比を改良することをめざす。具体的には、加法型または非負型のバイアス修正を行う。ここで、加法型、非負型バイアス修正は、異なる平滑化パラメータを持つ改良前の2つの

推定をそれぞれ加法，乗法で組み合わせることで，収束比を改良するというアイデアに基づく．このアイデアでは，2つの平滑化パラメータの比を第2パラメータとして導入するため，第2パラメータを選択する問題がある．研究代表者の現在までの研究から，1次元のいくつかの非対称カーネル推定ならば，第2パラメータを1とすることにより，平均積分2乗誤差を小さくできることがわかっている．そこで，他の非対称カーネル推定でも同様のことが成立するかを数学的に検討し，最良の第2パラメータを導出する．このように改良した非対称カーネル推定の漸近的性質を導出して，その漸近性能を明らかにする．漸近理論だけでなく，様々な第2パラメータに対する影響を膨大なシミュレーション実験から数値的に検討する．

(3) ノンパラメトリックな密度推定法には，実際の応用に際して平滑化パラメータを選択する問題がある．平滑化パラメータの選択は，推定の性能に大きく影響するため，平滑化パラメータ選択法の研究を行う．平滑化パラメータは，通常，平均2乗誤差や平均積分2乗誤差を最小にして選ばれる．従って，研究の初期段階において平均2乗誤差，平均積分2乗誤差の漸近公式が導出できれば，最適な平滑化パラメータも理論的に導出できる．しかし，そのような最適な平滑化パラメータには，推定対象となる，本来未知である密度やその導関数が含まれているため，実際には使用できない．そこで，本研究の非対称カーネル推定に対して，平滑化パラメータ選択法を詳細に文献調査及び再検討し，乱数によるシミュレーション実験を実施する．

(4) 密度推定の応用として『判別・クラスター分析』を念頭に置いている．多次元の非対称カーネル推定を用いて，密度をノンパラメトリックに推定することを考え，判別・クラスター分析などの多変量解析へ応用する．

4. 研究成果

(1) 台が非負である多次元密度を推定するため，まず多変量対数正規密度をカーネルとして多変量対数正規カーネル推定量を提案し，その漸近的性質を示した．さらに重み付き分布の発想から，重み付き多変量対数正規カーネル推定量へと一般化し，その漸近的性質を導出し，数値実験により，特定の重みにより性能が改善されることを示した．また，モーメント法によりパラメータを推定したガンマ密度を参照するrule-of-thumbによる平滑化パラメータ選択法を考察して，数値実験により，有限標本における性能を検証した．

(2) 台が非負である密度を対象とする逆ガンマカーネル推定量を再考察して，境界付近での推定における問題を回避できるように定義し直し，その新たな逆ガンマカーネル

推定量の漸近的性質を示した．

(3) 逆ガンマカーネル推定量などの非対称カーネル推定量をその特別な場合として含む，Amorosoカーネル密度推定量を提案して，その漸近的性質を示した．さらに，その平均積分2乗誤差から，最適なパラメータ設定も示した．また，数値実験を実施して，パラメータを適切に設定することにより，優れたパフォーマンスが得られることを示した．

(4) Amorosoカーネル推定量に加え，台が有界区間である密度を対象とするベータカーネル推定量のバイアス修正を行った．加法的型，積型的バイアス修正に加えて，それらの平滑化パラメータの比が1となる極限のバイアス修正を提示して，それらの漸近的性質を示した．さらに，その平均積分2乗誤差から，極限のバイアス修正が他のバイアス修正より優れた漸近性能を持つことを示した．また，数値実験を実施し，有限標本においても極限のバイアス修正推定量が優れていることを示した．

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)

¹ Kakizawa, Y. and Igarashi, G., “Inverse gamma kernel density estimation for nonnegative data”, Journal of the Korean Statistical Society, 46, 194-207, 2017, 査読有．

² Igarashi, G., “Bias reductions for beta kernel estimation”, Journal of Nonparametric Statistics, 28, 1-30, 2016, 査読有．

[学会発表](計 8件)

¹ 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」, 研究集会「第18回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(東京都)．

² 五十嵐岳, 「境界バイアスのない非対称カーネル密度推定量について」, 日本数学会2017年度年会, 2017年3月26日, 首都大学東京(東京都)．

³ 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」, 2016年度統計関連学会連合大会, 2016年9月7日, 金沢大学(石川県)．

⁴ 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, The Chinese University of Hong Kong(香港)．

⁵ 五十嵐岳, 「多変量加重対数正規カーネルを用いた非対称カーネル推定について」, 研

究集会「第 17 回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 (東京都).

⁶ 五十嵐岳, 「多変量対数正規カーネルを用いた境界バイアスのない密度推定について」, 日本数学会 2016 年度年会, 2016 年 3 月 17 日, 筑波大学 (茨城県).

⁷ Kakizawa, Y. and Igarashi, G., “Recent developments in boundary-bias-free asymmetric kernel density estimation”, 研究集会「Recent Progress in Time Series and Related Fields」, 2015 年 12 月 11 日, 東北大学 (宮城県).

⁸ 五十嵐岳, 「多変量対数正規カーネルを用いた境界バイアスのない密度推定について」, 2015 年度統計関連学会連合大会, 2015 年 9 月 8 日, 岡山大学 (岡山県).

6 . 研究組織

(1) 研究代表者

五十嵐 岳 (IGARASHI, Gaku)

筑波大学システム情報系社会工学域・助教

研究者番号 : 40759346