

機関番号：10101

研究種目：基盤研究(C)(一般)

研究期間：2008～2010

課題番号：20500248

研究課題名(和文) スペクトル密度推定の平滑化パラメータ選択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 choosing a smoothing parameter for spectral density estimation and its application

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・准教授

研究者番号：30281778

研究成果の概要(和文):

ノンパラメトリック関数推定法における「平滑化パラメータ」とは、関数近似誤差(統計学の用語では推定量の偏りに相当する)と推定量の分散とのトレードオフを制御する役割をもつ。本研究では一般化ベルンシュタイン多項式近似に基づいたスペクトル密度推定量及び確率密度推定量を提案した。特に、平滑化パラメータに相当する多項式次数だけでなく、古典的なベルンシュタイン推定量を繋ぐ役割をする第2パラメータの選択にも注目し、漸近バイアス・分散・平均2乗誤差などの公式を導出し、漸近正規性も証明した。

研究成果の概要(英文):

The smoothing parameter in nonparametric function estimation plays a role in controlling the trade off between the approximation error (in terminology of Statistics, it is nothing but bias of the estimator) and variance of the estimator. In this research, I discussed a new spectral density estimator/probability density estimator based on a generalized Bernstein polynomial approximation theory, in which the degree of polynomial and the additional second parameter are the smoothing parameters. Especially, main focus is the selection problem of these two parameters, where the latter second parameter enables us to generalize the classical Bernstein method. I established several formulas of asymptotic bias, variance and mean squared error as well as the asymptotic normality of the proposal.

交付決定額

(金額単位:円)

	直接経費	間接経費	合計
2008年度	1,200,000	360,000	1,560,000
2009年度	1,100,000	330,000	1,430,000
2010年度	1,100,000	330,000	1,430,000
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

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：定常過程、スペクトル密度、ノンパラメトリック推定

1. 研究開始当初の背景

ターゲット関数に有限次元の母数モデル

を仮定しその母数を尤度原理などから推定するパラメトリック推定法に対し、ノンパラメトリック推定法とは、広い意味で関数近似

の統計学である。また、セミパラメトリック法がパラメトリック法とノンパラメトリック法の間中に位置し、これら3つの方法論は統計科学の現代的な推測理論の基礎をなすものである。

本研究での対象は、主に多変量定常時系列データの周波数領域の解析とし、統計科学の研究テーマの中でノンパラメトリック関数推定の研究へと位置づけている。この方向での先行研究は膨大で、歴史的には、独立同一分布設定で確率密度・分布関数・分位点関数、単変量（またはベクトル値）定常時系列分析でスペクトル密度（行列）関数・分布関数、ノンパラメトリックな回帰分析で回帰関数、生存時間分析でハザード関数などを典型的な題材として発展してきた。

伝統的なカーネル推定法の背後には推定の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できるという事実がある。畳み込みによる近似以外にも関数近似理論のテクニックは無限に存在するから、関数近似理論はノンパラメトリック法を構築する際の基礎理論を提供している。例えば、十分滑らかな関数のテーラー近似（局所的な多項式を推定する問題へと還元させる）、三角多項式・ウェーブレット近似（推定対象の属している関数空間に付随する直交関数で展開してその係数を推定する問題へと還元させる）、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を応用することで種々の推定法が提唱されてきた。ただし、関数近似理論は近似誤差評価に関心があり、統計科学は未知のターゲットをデータから有効に推測することを問題とする。

統計科学で注目すべき点は、統計的な性能（通常は標本数を無限大にするという漸近理論を行う）を決めるもの（以下ではこれを平滑化パラメータと呼ぶ）があつて、関数近似誤差（統計学の専門用語では推定量の偏りにほかならない）と推定量の分散のトレードオフを平滑化パラメータが制御するという事実である。したがって、ノンパラメトリックな関数推定では平滑化パラメータの選択が重要な課題であり、1990年代以降の主要テーマの1つとして国内・国外で研究され、現在でも著しい進展がみられる分野である。

研究代表者は、1990年代後半にカーネル法によるスペクトル密度（行列）推定量の汎関数を対象として様々な漸近的性質を研究してきた経緯があり、特に、スペクトル密度行列推定量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法の提案、ならびに、擬似距離をスペクトル密度行列関数に関する推定・検定問題にも応用することから、尤度原理と漸近同等であるものを提示している。これらの漸近的性質を得るためにバンド幅の

標本数に関するオーダー条件を仮定しているが、これはバンド幅がこのオーダーであればカーネルスペクトル密度（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てスペクトル密度行列による統計手法を発展させていたからである。その当時、バンド幅の自動選択という問題の重要性を鑑みて、文献調査をしていた段階だったが、ある実データをクラスタリングに応用したときのバンド幅の選択について、試行錯誤で探索した結果と単変量定常時系列で知られていた一種のバイアス調整済み情報量規準を使ったときの結果に大差がなかったために、バンド幅の自動選択に対する問題意識はそれほどなかった。

バンド幅の選択理論に対する意識が高まったのは、2003-2005の科研費（若手研究）の研究成果による。単変量定常時系列過程のスペクトル密度関数、あるいは、独立同一分布で台を $[0, 1]$ とするような確率密度関数の推定問題において、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した後でその漸近性能を調べたところ、カーネル密度推定量の平均積分2乗誤差と同じ収束比をもつことが示された。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て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必要もあり一種のバイアス調整済みの情報量規準の併用を試みた経験があった。太陽黒点データに対し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、このことからベルンシュタイン推定における原点付近での多項式次数の選択理論に関心をもっていた。

しかしながら、先行研究で議論されていたカーネル法においてさえも、バンド幅選択問題と具体的な様々な統計手法の開発とは独立に研究されることが多く、これらをより厳密な理論展開の下で再検討する必要性を認識していた。

2. 研究の目的

ノンパラメトリック関数推定法において平滑化パラメータの選択は統計科学全般で想定する状況毎の固有な問題があり、本研究の目的は単変量（またはベクトル値）定常時系列データの周波数領域の解析、特に、スペクトル密度（行列）の汎関数に基づいた統計的推測手法を平滑化パラメータの自動選択を念頭にした理論へ発展させることである。従来のデータに依存しない非確率的なバンド幅を伴うカーネルスペクトル行列推定量の汎関数の漸近的性質を保証した上で、確率的なバンド幅（データからある意味での最適なバンド幅を自動決定する）を使用した理論

へ再検証すること、種々のスペクトル密度（行列）推定量の平滑化パラメータの選択に関わる基礎理論の整備とそれら数理的説明、ならびに、実データへの適用をめざす。なお、定常時系列のスペクトル密度（行列）関数のノンパラメトリック推定法に対して示される数理は、しばしば独立同一分布設定の確率密度関数の場合にもほとんど並行して成立するから、確率密度関数の推定問題についても本研究の研究対象とする。

一般に、定常時系列のスペクトル密度関数または独立同一分布設定での確率密度関数のノンパラメトリック推定法には、伝統的なカーネル推定法以外で局所多項式法、ウェーブレット近似法もあるが、スペクトル解析の分野においてベルンシュタイン多項式近似を動機としたスペクトル密度推定量は文献ではあまり知られておらず、このような密度推定法の統計理論を構築する。特に、平滑化パラメータに相当する多項式次数選択問題に注意して、提案される推定法の漸近的性能を明らかにする。ここに、漸近的性能の指標としては平均2乗誤差または平均積分2乗誤差を採用する（その収束比が0.8になることを確認できれば、カーネル法の対抗馬となることに注意）。

3. 研究の方法

(1) 概要

本研究の主な対象である単変量（またはベクトル値）定常時系列過程のスペクトル密度関数、あるいは、独立同一分布設定の確率密度関数の推定について

・ 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い、かつ、統計科学におけるノンパラメトリック関数推定法の関連領域についての十分な文献調査を通じて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して、可能ならばその拡張を試みると共に、本研究で提示されていく新しい問題に対して試行錯誤して解決を図る

・ 情報収集

学会・研究集会・ワークショップに参加して関連領域の最先端の研究動向を掴み、他研究者と意見交換をする

・ プログラム開発

数値実験を実施して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研究者からの意見を求める

・ 国際的なジャーナルへの投稿

を経て、研究の質を高めていった。

(2) 研究経過

本研究のターゲット関数は単変量（またはベクトル値）定常時系列のスペクトル密度（行列）関数であるが、研究の糸口をつかむため、まず、非定常の枠でも類似した概念と知られる、いわゆる、almost periodically correlated process の文献調査を行って、mixing 条件の下での周波数領域の漸近理論及び subsampling 法の数学的基礎を養った。次に、バンド幅の選択理論に関する先行研究について文献調査から再度整理して、単変量定常時系列のカーネルスペクトル密度推定量のバンド幅に対するプラグイン法を見直した。プラグイン法とは、平均2乗誤差または平均積分2乗誤差を小さくするような最適な平滑化パラメータを漸近的に導出できるならば、その漸近公式自体も未知の関数に依存する場合にそれを別の推定量で置き換えるという方法である。その初期推定量としてベルンシュタイン密度推定量を採用して、その上で繰り返し計算を行うというアイデアの先行文献を知ったことも動機となり、ベルンシュタイン密度推定量をさらに一般化して考察する構想に至った。

一方、2003-2005 の科研費（若手研究）で得られたベルンシュタイン多項式により重み付けしたスペクトル推定法の研究成果が、伝統的なカーネル法の対抗馬としての一種の平滑化法であることが分かっていた。この場合のバンド幅に相当する平滑化パラメータが多項式次数であるので、この方面の事項について本研究の対象に含めることを模索した。幸いに、ベルンシュタイン確率密度推定法の文献調査から、ベルンシュタイン型のスペクトル推定量をさらに拡張する構想、すなわち、ベルンシュタイン多項式近似理論の発展（数学分野で知られていた関数近似理論の1つのバリエーション）を統計科学に応用するという構想を持ち、一般化ベルンシュタイン密度推定量の提案に至った。この研究を進めていった初期の過程で、そのような新しい確率密度推定量及びスペクトル密度推定量を正線型積分作用素型推定量として特徴づけし、その上で提案された推定量の漸近バイアス（言い換えれば、基準化された新しい正線型積分作用素による近似誤差の精密化）をまず導出することに成功した。この推定量では多項式次数に加えて、第2パラメータが存在していることを注意し、そのパラメータが1に等しいときは古典的なベルンシュタ

イン推定量になる。しかも提案した推定量は常に正值の推定量を形成し、かつ、有界区間 $[0, 1]$ に台をもつ場合においては境界バイアスの問題がないという利点をもっている。

このような研究の開始時では、この一般化ベルンシュタイン法に必要な第2の平滑化パラメータを標本数と無関係に考えていた。しかしながら、関数近似誤差（統計学では、推定量の偏りに相当する）を保証する第2の平滑化パラメータの範囲は発散する状況もあり得る。興味深い点として、第2の平滑化パラメータも標本数に依存させて考えた場合、これが標本数と共に発散するとき漸近的な平均積分2乗誤差のオーダーが 0.8 にならないことを示した。同様に、第2の平滑化パラメータが標本数と共に 0 に収束するときも漸近的な平均積分2乗誤差のオーダーが 0.8 にならない。したがって、第2の平滑化パラメータは定数オーダーで考えれば十分であることも明らかにした。また、第2の平滑化パラメータが 1 よりも大きいとき、漸近的な平均積分2乗誤差は常に、古典的なベルンシュタイン法の平均積分2乗誤差より大きいことを示した。

このようにして、提案した一般化ベルンシュタイン法が古典的なベルンシュタイン法を漸近的な平均積分2乗誤差の意味で改善できるなら、第2平滑化パラメータは 1 以下の正数でなければいけないこと、ならびに、その範囲に制限して漸近的な平均積分2乗誤差の最小化問題を解くことで一般化ベルンシュタイン密度推定法の漸近性能が明らかにされていった。

この研究成果は、一般化ベルンシュタイン密度推定量という提案の意味において数学分野の関数近似理論を統計科学に応用した一例となっており、その他の関数近似理論の統計科学への応用の可能性も秘めているだろうと考えた。（一般化）ベルンシュタイン多項式による密度推定量を普及するために学会・研究集会・ワークショップで発表して、他研究者からの意見を求めた。また、国際的なジャーナルへも投稿した。匿名の査読者の意見をふまえた再投稿に際し、ある実データに応用したが、公差確認法に基づいて第2のパラメータの変化による提案した推定量の性能の比較をして、数学的に証明された漸近理論と実データ解析の結果が整合していることも確認した。

ところで、一般化ベルンシュタイン法は、スペクトル密度関数のような周期関数または有界区間 $[0, 1]$ を台とするような密度関数の推定問題に応用されて、それらの統計理論が確立されたが、半無限区間 $[0, \infty)$ で定義された密度関数の推定問題へも拡張できるだろうと予想した。半無限区間 $[0, \infty)$ を想定したノンパラメトリック法の文献調査を行っ

たところ、ベルンシュタイン密度推定量は、近年の文献にて盛んに注目されているベータカーネル密度推定量と類似した漸近的性質を持っていることが多いことに気づき、ベータカーネル・ガンマカーネルなどの非対称カーネルに関わる文献調査も実施した。さらに、ノンパラメトリックな密度推定法の2つの異なるタイプ（非対称カーネル法及びベルンシュタイン法）を統一的に議論すること、ならびに、まだ試行錯誤中であるために途中経過だが、一般化された密度推定量のバイアス修正版の推定量について漸近分散を計算することも試みた。ベルンシュタイン密度推定量に対するバイアス修正済み推定量の漸近分散については、2項分布の確率関数の積の和の近似に帰着することが分かり、そのような近似公式の誤差評価を正当化するため、独立な格子点分布の和の分布に対するBerry-Esseen型評価法の技術も応用した。

4. 研究成果

一般化ベルンシュタイン多項式近似理論に関連して確率密度・スペクトル密度推定量を提案した。ベルンシュタイン法の特徴でもあるが、ここでの推定量は常に正值の推定量を形成し、かつ、有界区間 $[0, 1]$ に台をもつ場合において、境界バイアスの問題がないという利点をもっている。また、提案した推定量は古典的なベルンシュタイン密度推定量を1つのパラメータで拡張した推定量の属として解釈される。多項式次数及びそれに含まれる第2の平滑化パラメータのオーダー条件を厳密に議論した後で、漸近バイアス・分散・平均積分2乗誤差などの公式を導き、推定量の漸近正規性を証明した。実データに適用するため、独自のPC環境で実行可能なFortranプログラムも整備した。このような提案した密度推定量について、第2の平滑化パラメータが定数オーダーであれば漸近的な平均積分2乗誤差の収束比が 0.8 になることから、伝統的なカーネル法の対抗馬としてノンパラメトリック法の1つの方法論を提示できた。すなわち、ベルンシュタイン多項式による関数近似理論（数学の分野からの基礎理論）の1つの一般化を統計科学の応用分野へ発展させた。なお、この成果の一部は、国際的な統計専門雑誌（2011年の3月号）に掲載済である。

一方で、2つの異なるタイプの推定法（非対称カーネル法及びベルンシュタイン法）を統一的に議論すること、ならびに、それらのバイアス補正後の推定量の考察にも着手し始めており（一部の漸近的性質の導出に成功している）、このような方面からの研究成果が期待される。

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1件)

Kakizawa, Y. A note on generalized Bernstein polynomial density estimators. *Statistical Methodology*. 8(2), 136-153. (2011). 査読有.

[学会発表](計4件)

柿沢佳秀. A generalized Bernstein polynomial approach to density estimation. 研究集会「統計的推測方法の理論的展開とその応用」. 2010年11月18日. 熊本大学(熊本).

柿沢佳秀. ベルンシュタイン多項式による密度推定について. 応用統計ワークショップ. 2010年5月21日. 東京大学(東京).

柿沢佳秀. Generalized Bernstein polynomial density estimation. 研究集会「第12回ノンパラメトリック統計解析とベイズ統計」. 2010年3月29日. 慶応大学(東京).

柿沢佳秀. On generalized Bernstein polynomial density estimators. 日本数学会2009年度秋季総合分科会. 2009年9月26日. 大阪大学(大阪).

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・准教授
研究者番号: 30281778

(2)研究分担者

なし

(3)連携研究者

なし