

氏名（本籍）	山本 けい子（三重県）
学位の種類	博士（工学）
学位記号番号	甲第 181 号
学位授与年月日	平成14年 3月25日
専攻	電子情報システム工学専攻
学位論文題目	ノンパラメトリックな区分的線形確率密度関数の推定に関する研究 (Nonparametric Piecewise Linear Probability Density Estimation)
学位論文審査委員	(主査) 教授 後藤 宗弘 (副査) 教授 田中 嘉津夫 教授 岸田 邦治 助教授 寒河江 雅彦

論文内容の要旨

工学・医学・経済学などさまざまな分野においてデータ解析が必要とされている。データ解析における統計的手法には、例えば回帰分析、時系列分析、主成分分析・判別分析・独立成分分析など数多くのものであり、統計的推定や検定・予測・パターン認識などに応用される。このような手法を用いる上で最も基本となるのがデータの背後にある確率構造を見つけることである。実際の応用においては、データの背後にある確率的構造があらかじめ完全に分かっていることは稀で、それらをデータから推定する必要がある。そのための数学的基礎を与えるものが確率分布であり、その確率密度関数の推定法を開発することが本論文の目的である。このようにデータから確率密度関数を推定することによって、次のステップとして多くの統計手法が実用的に使用され、各分野において活用されるのである。

確率密度関数を推定する方法として、パラメトリックな推定法とノンパラメトリックな推定法がある。パラメトリック推定法は、あらかじめ特定の確率分布を想定したパラメトリックモデルを用いて確率密度関数を表し、そのモデルをデータに当てはめ、データに最もよく適応するパラメータを最尤推定法や最小二乗推定法などを用いて推定するものである。これに対して、ノンパラメトリック推定法は、あらかじめ特定の関数を仮定しないで、データに基づいて直接的に確率密度関数の推定をする方法である。ノンパラメトリックな推定法は、複雑な確率分布にも柔軟に適応できることから、データが持つ確率構造を効果的に表すことができる推定法として、近年盛んに研究されつつある。

本論文では、データに含まれる確率構造を表す情報を用いた柔軟な推定を行うために、特定の関数型を想定しないノンパラメトリックな立場での推定について議論する。

ノンパラメトリックな確率密度関数の推定法の中でも、データの集約度（個々のデータに対してどれだけの集計を行ったか）という観点で、

- 1) ヒストグラムのような度数情報を用いたビン型推定法
- 2) 近年盛んに研究されているカーネル推定法
- 3) 度数に加えてより多くの情報を取り入れた推定法

について取り上げる。また、理論および実用上の効率を考慮し、区分的線形関数を用いたクラスに焦点を絞って、それぞれの推定モデルの理論的な性質を導き、その有効性を示している。

下記で1)～3)についての概要を説明する。

1) ビン型推定法として最もよく知られているものとしてヒストグラムがある。ヒストグラムは簡単でわかりやすいという反面、得られる推定結果が不連続で、また理論的な精度においても他の方法と比較するとよくないことが知られている。そのため、ヒストグラムの不連続性を改良するようなビン型推定モデルが提案されている。その最も単純なものとして、区分的線形関数を用いた Frequency Polygon (Scott, 1985)、Edge Frequency Polygon (Jones 他, 1998)、Bias-Optimized Frequency Polygon (Minnotte, 1996) がある。また、これらのモデルを節点の位置と高さ調整に関して一般化した Generalized Frequency Polygon (Sagae and Yamamoto, 2000) が提案されていることから、これらの推定モデルを統一的な視点で論じる。また、Generalized Frequency Polygon の漸近的な性質や度数を用いたビン化カーネル推定法への応用などについても述べ、その有用性を示している。2) カーネル推定法は、近年、計算機の発展により理論的にも盛んに研究され、実用的にも広く利用されている方法である。カーネル推定法はデータの各点に基底となるカーネル関数を配置し、すべての点で配置したカーネル関数を重ね合わせる推定法である。得られる推定結果はなめらかであるが、データが大量である場合やカーネル関数に一般的な連続関数を用いた場合は計算量が膨大になるという問題がある。そこで、区分的線形関数を用いて効率的に生成できる台形カーネル関数を定義し、その理論的な性質や計算効率、視覚的な比較を行い、台形カーネル関数が理論と実用の両面において優れていることを示している。また、度数の情報を用いたビン化カーネル推定法についても議論している。3) 度数情報に加えてより多くの情報を取り入れた推定法では、前述のヒストグラムの不連続性および理論的な精度を改良するものとして位置づけられる。ビン型推定法の中でも、度数情報だけでなく高次の局所モーメント情報を取り入れた推定法として Polynomial Histogram (Sagae and Scott, 1997) などがヒストグラムを改良するモデルとして提案されている。このことから、度数に加えて標本データから得られる他の情報（メディアンや全標本の平均値）を用いた区分的線形な推定モデルを構築し、その漸近的な性質を導くことによって、度数以外の情報を用いた新たな推定モデルの可能性などについて議論している。

論文審査結果の要旨

本論文では、データ解析を行うためのさまざまな統計的手法を用いる上で必要となるデータのもつ確率構造を表現する確率密度関数の推定法について取り上げている。確率密度関数の推定法の中でも、データから直接的に推定することによって多くの場合に柔軟に適応できるノンパラメトリックな推定法に注目し、度数情報を用いたビン型推定法やカーネル推定法の理論的な精度および実用性を考慮したモデルの構築を行っている。また、統計分野においてもノンパラメトリックな推定法は、近年、計算機の高速化や高性能化によって盛んに研究されつつあることから、計算効率や実用上の多様性を十分考慮した区分的線形関数を用いた推定法について論じている。

第1章では、データ解析における確率密度関数の推定法の必要性やノンパラメトリックな推定法を用いる有効性、また、このクラスに属するビン型推定法やカーネル推定法について概説している。

第2章では、ビン型推定法におけるヒストグラムの理論的な性能や視覚的な不連続性を示し、ヒストグラムを改良する既存の方法 **Frequency Polygon**, **Edge Frequency Polygon**, **Bias-Optimized Frequency Polygon** について、構成法および理論的な精度について統一的な視点で論じている。

第3章では、第2章に挙げた推定モデルを節点の位置と高さ調整に関して一般化して表現した **Generalized Frequency Polygon** を定義し、その理論的な性質を導くことによって、既存の方法を改良するモデルであるということを示している。また、区分的線形関数を用いたクラスの理論的な精度の限界値を導出している。近年、研究の盛んなカーネル推定法との関連から度数情報を用いたビン化カーネル推定法についても取り上げ、本章で定義した高さ調整をビン化カーネル推定法へ応用することによって漸近的な性質の改良につながることを導いている。

第4章では、カーネル推定法における計算効率化について議論している。カーネル推定法は、データの各点に基底となるカーネル関数を配置し、すべての点で配置したカーネル関数を重ね合わせて構築されることから、カーネル関数やデータ数に依存した計算量が必要となる。そこで、カーネル関数による計算量の削減のために区分的線形関数を用いて効率的に生成できる台形カーネルを定義している。台形カーネルの理論的な精度の導出や誤差評価、計算時間、視覚的な比較を実際によく用いられている正規カーネルと行い、台形カーネルが理論と応用の両面において有効であることを示している。また、データ数に対する効率化としては第3章で述べたビン化カーネル推定法が有効であることを述べている。

第5章では、度数情報を用いたビン型推定法において、データから得られるメディアンや全標本の平均値の情報を付加したモデルを構築することによって理論的な精度の改良を試みている。誤差基準による理論的な精度の改良は見られなかったものの、データから得られる情報やモデルの構築法、誤差基準などさまざまな角度からの検討を行っており、今後の研究につながるものである。

本論文によって示されたノンパラメトリックな区分的線形確率密度関数の推定法の理論および実用上の有効性は、工学分野への広い応用を有するものと思われ、学位論文として十分な価値を認めた。

最終試験結果の要旨

公聴会後に、学位論文に関連する口頭試問を行い、これを最終試験に代え、合格と判定した。