

氏 名	しま 島 ざき 崎 ひで 秀 あき 昭
学位(専攻分野)	博 士 (理 学)
学 位 記 番 号	理 博 第 3111 号
学位授与の日付	平 成 19 年 3 月 23 日
学位授与の要件	学 位 規 則 第 4 条 第 1 項 該 当
研 究 科 ・ 専 攻	理 学 研 究 科 物 理 学 ・ 宇 宙 物 理 学 専 攻
学 位 論 文 題 目	Recipes for Selecting the Bin Size of a Histogram (ヒストグラムのビン幅選択のためのレシピ)
論文調査委員	(主 査) 助教授 篠 本 滋 教 授 太 田 隆 夫 教 授 小 貫 明

論 文 内 容 の 要 旨

データを可視化する最も基本的な方法にヒストグラムがある。ヒストグラムを作成するには、まずデータを決められた区間(ビンと呼ばれる)ごとに区切り、次にそのビンに入るデータ数を棒グラフによって表わす。申請論文は、ヒストグラムを作成する際に、最適なビン幅を選択する手法を提案している。

ここで最適なビン幅とは、密度分布と密度分布に従って独立に生成されたデータから作成したヒストグラムの二乗誤差を最小にするビン幅である。実験者はデータからヒストグラムを作成することはできるが、密度分布を観測することはできない。従って二乗誤差を直接計算することはできない。この問題に対し、申請論文は二乗誤差をデータから推定する公式を導き、この公式を最小化するビン幅を選ぶことによって最適なヒストグラムを作成する方法を提案している。

申請論文は次にデータ数が少ない場合の対処法を提案している。実験回数が少なくデータ数が十分でないとき、上記の方法で求めた最適ビン幅がデータ範囲と同程度に大きくなることがある。このことは密度分布をヒストグラムによって表すには実験回数が不足していることを意味している。このような場合に、ヒストグラム作成に最低限必要な実験回数を試算する公式が提案されている。

申請論文では、理論解析と数値実験を用いて、提案するビン幅選択の手法や少数データへの対処法が実験データの提示及び実験計画に有用な手法であることを確認している。まず、密度分布を既知とした場合の最適ビン幅の理論値、及びヒストグラム作成のために最低限必要な実験回数の理論値が求められている。次に計算機実験によって生成したデータに上記の手法を適用し、理論値と比較することで手法の有効性を検討している。

論 文 審 査 の 結 果 の 要 旨

ヒストグラムの形状はビン幅の取り方によって大きく変わるにもかかわらず、その選択は多くの場合作成者の裁量に任されている。例えば、神経科学分野では神経細胞のスパイク時系列データの表示にヒストグラムが頻繁に使われている。しかし、そのビン幅は業界標準があるのみで、最適ビン幅を求める手法は知られておらず、理論的根拠のある値が使われてきたことはない。申請論文で示されたヒストグラムの最適ビン幅の導出方法はデータの客観的な表示に欠かせない重要な手法であり、神経科学分野を含めて今後幅広い分野で使用されたいと考えられる。

加えて申請論文は、実験データが少ない場合にあと何回実験を行う必要があるか推定する手法も提供している。この手法は神経科学のような、データ取得に手間のかかる分野で実験計画を立てる上で役立つだろう。

また申請論文は背後のデータ生成率が既知であるとした場合に得られる最適幅の理論値についても、昨年度の小山慎介氏の学位論文において発見されたスケーリング則や相転移を一般的な状況で導き、明快な説明を与えている。

申請論文は、最適ビン幅の理論研究を背景にして、実験データから最適ビン幅を求める手法を提案している。また最適ビン幅の相転移の理論に基づいて、データ数が少ない場合の対処法を提案している。このように理論解析を基盤にしながらそ

れだけに閉じず，実験データの解析に有用な手法を提案する方向性は実験科学を支える理論研究として評価できる。

よって，本論文は博士（理学）の学位論文として価値あるものと認める。また，論文内容とそれに関連した事項について試問を行った結果，合格と認めた。