様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

# 科学研究費助成事業　　研究成果報告書

令和　元 年　 ９ 月　 ９ 日現在

機関番号：２２７０１
研究種目：基盤研究(C)（一般）
研究期間：2015～2018
課題番号：１５Ｋ００００５１

研究課題名（和文）多変量データにおけるベイズ型リサンプリング法の分布特性とその応用に関する研究

研究課題名（英文）Distribution Characteristics of Bayesian Resampling Method in Multivariate Data and its Application

研究代表者
橋口　陽子（小野陽子）（Hashiguchi (Ono), Yoko）

横浜市立大学・データサイエンス学部・准教授

研究者番号：６０３３９１４０

交付決定額（研究期間全体）：（直接経費）　　3,500,000円

研究成果の概要（和文）：2つの正規母集団の平均の仮説検定において，等分散を前提としないベーレンスフィッシャー問題でブートストラップ法（以下BST法），パラメトリックブートストラップ法（以下PBST法），ベイジアンブートストラップ法（以下，BBST法）で比較を行った.その結果PBST法の精度が他2種法に比べて精度が高い事がシミュレーションの結果確認された．更に，自動証明における証明方針選択へのリサンプリング法の応用を試み，証明過程における方針選択へ果たす役割を検討した．提案規則数の絞り込みが現実的な数にまで削減可能とした．また,統計的手法を用いた証明方針提案試作を行った．

研究成果の学術的意義や社会的意義
ブートストラップ法（BST法）は,遺伝子解析や分子系統樹の作成など，医療・生命分野に利用されており，今後更なる適用範囲の拡大が予測される手法である．しかし，リサンプリング法が内包する標本への過敏性を考慮し，ビッグデータへの適用を鑑みるならば，多変量データに関する統計量分布の問題を回避することは難しいと思われる．本研究において，BST 法の一般型とも言えるベイジアン型リサンプリング法を多変量データへ適用することにより，その研究成果から，多変量データを扱う上でのリサンプリング法適用上の問題点や，適用が難しい統計量などが明確に示されることが予想される．

研究成果の概要（英文）：We compared the accuracy of Bootstrap method (BST), Parametric Bootstrap method (PBST), and Bayesian Bootstrap method (BBST) in hypothesis testing of the mean of the two normal populations. According to the simulation results, it is confirmed that the accuracy of PBST is higher than the other two methods.
We also attempted to apply the resampling method to the selection of certification policies in automated reasoning problems. We use the resampling method to narrow down the number of proposed rules and make it possible to reduce the number of rules to a realistic number. The use of statistical methods to eliminate new nonsense ideas, which have exploded in number and appearance, in automatic verification led to the start of a new study on statistical learning.

研究分野：計算機統計学

キーワード：ベイズ型リサンプリング法　統計量分布特性　自動証明

## １．研究開始当初の背景

Efron(1971)により提唱されたブートストラップ法(以下,BST 法と省略)は,計算機を集中的に扱うリサンプリング法として,小標本における信頼区間の構成など,様々な統計分野に適用されている．また,BST 法 を応用もしくは拡張した BST 型のリサンプリング手法が多く提案されており,それら の手法を用いた数値実験結果とその漸近評価について,多くの研究がなされている．一方,Rubin(1981)は BST 型リサンプリング法のひとつである,Bayesian BST 法(以下,BBST 法 と省略)を提案した.BBST 法は,標本リサンプリングの重みをサンプリングするという視点からみ ると,BST 型リサンプリング法の一般表現と捉えることが可能である．BST 型リサンプリング法は,未知の母集団分布の代わりに,既知である経験分布からのサンプリングを行うことにより,数値的な統計量分布の近似を得ようとするものである.一変量の場合では,標本サイズが大きいとき,BST 法が 2 次の良い近似を与えることはよく知られている．多変量データに関する統計量分布を考えた場合,BST 型リサンプリング法は同じデータをとること によって線形独立性が崩れ,統計量によっては全く意味のない分布になってしまう可能性がある.一方,BBST 法は,標本ベクトル達を並べた行列のランクを保持できるので,線形独立性が崩れる心配 はない.したがって,多変量データに関しては BBST 法が有効なリサンプリング法ではないかと考えられる．

## ２．研究の目的

本研究の目的は，多変量におけるベイズ型リサンプリング法の分布特性を理論的に調べ自動証明における証明方針選択へのリサンプリング法の応用を試みることである．多変量データにおける復元抽出のリサンプリング法は，うまく働かない場面(例:共分散行列が特異になる可能性がある状況) が多々あることが容易に想像される．ベイズ型ブートストラップ法は，このような欠点を克服する可能性を有するリサンプリング法であることから，多変量データへの適用に関する理論的側面を調べることは重要であると考えられる．更に，自動証明における証明方針選択へのリサンプリング法の応用を試み,証明過程における方針選択へ果たす役割を検討する．

## ３．研究の方法

BBST型リサンプリング法の分布特性を理論的側面と数値的検証による側面から明らかにすることを目標とし，研究を行う．具体的には次の課題を扱う．

課題 1:BBST 型リサンプリング法による統計量分布特性に関する研究 ：３つの統計量（変動係数，標本相関係数，分散共分散行列の固有値）に関して，BBST法の分布特性の理論的評価と数値実験を行う．

課題 2:多変量解析へのBBST型リサンプリング法の適用 ：データ解析の応用として，主成分分析における固有値の同時信頼区間の構成，判別分析における判別関数による誤判別率の推定，といった多変量解析への BBST 型リサンプリング法の適用を行う．

課題 3:自動証明における知識獲得への BBST 型リサンプリング法の応用 ：抽象数学自動証明における証明過程では，論理構造が知識として大量に要求される．大量の知識の中から証明の方向性を選択する際に，BBST 型リサンプリング法を利用した統計処理を行うことで，自 動証明の構造がどのように構築されるかを確認し，自動証明への統計科学の利用について考察する．

## ４．研究成果

本研究の目的は,多変量におけるBST型リサンプリング法の分布特性を理論的に調べ,自動証明における証明方針選択へのリサンプリング法の応用を試みることであった．

課題１：

2つの正規母集団の平均の仮説検定において，等分散を前提としないベーレンスフィッシャー問題においてBST法，PBST法，BBST法で比較を行った.その結果，PBST法の精度が他2種法に比べて精度が高い事がシミュレーションの結果確認された.多変量データにおける復元抽出のリサンプリング法は,うまく働かない場面(例:共分散行列が特異になる可能性がある状況）が多々あることが容易に想像される.また，BBST法は,このような欠点を克服する可能性を有するリサンプリング法であることから,多変量データへの適用に関する理論的側面を調べることは重要であると考えられた.その結果,多変量正規分布の下での数値実験結果ではベイズ型リサンプリング法がうまく機能するが,母集団に混合正規分布を仮定した場合ではうまく機能しないことが判明している.

課題2：

主成分分析における固有値のモデル選択に関する議論を行ったが，課題1と同様に混合正規分布を母集団と仮定した場合の問題が解決されなかった.統計量の構成方法における問題が解消されていない.これは今後の課題である.

課題 3：
自動証明における証明方針選択へのリサンプリングの応用を試み,証明過程における方針選択へ果たす役割を検討した.知識としての命題成功例と抽出規則のデータベース格納が可能となったことから，提案規則数の絞り込みが現実的な数にまで削減可能となった．また，統計的手法を用いた証明方針提案試作を行った．その結果として，証明の方針となる math idea という核を見つけ出し，人間の証明方針になぞらえた証明手順となるように方針付けが可能となった．自動証明において統計的手法を用いて爆発的に見かけ上増加する新しい無意味なアイディアを削除する事が可能になった事により，統計的学習に関する新たな検討を始めるに繋がった.

５．主な発表論文等
〔雑誌論文〕（計 2 件）
(1) Aya Shinozaki and Hiroki Hashiguchi
EXACTDISTRIBUTION OF THE LARGEST AND SMALLEST EIGENVALUES OF THE RATIO OF TWO ELLIPTICAL WISHART MATRICES,
Journal of Statistics: Advances in Theory and Applications, Volume 19, Number 2, 2018, 71—82

(2) Aya Shinozaki, Hiroki Hashiguchi and Toshiya Iwashita, DISTRIBUTION OF THE LARGEST EIGENVALUE OF AN ELLIPTICAL WISHART MATRIX AND ITS SIMULATION, J. Jpn. Soc. Comp. Statist., 30.2, 2018, 1—12

〔学会発表〕（計 3 件）
(1) Haruto Mura, Hiroki Hashiguchi, Shigekazu Nakagawa, Yoko Ono.
Holonomic properties and recurrence formula for the distribution of sample correlation coefficient, Statistical Computing : Challenges and Opportunities in Data Science, Beijing, November 2018

(2) Hiroki Hashiguchi, Bootstrapping Distribution on the Eigenvalues of Covariance Matrix, The 4th Indtitute of Mathematical Statistics Asia Pacific Rim Meeting, The Chinese University of Hong Kong, Hong Kong, June 2016

(3) Hiroki Hashiguchi, Shigekazu Nakagawa and Yoko Ono, Holonomic properties for the distribution of the sample correlation coefficient, IASC-ARS 2015, National University of Singapore, December 2015

〔図書〕（計 0 件）

〔産業財産権〕
○出願状況（計 0 件）
名称：
発明者：
権利者：
種類：
番号：
出願年：
国内外の別：

○取得状況（計 0 件）
名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕
ホームページ等

６．研究組織

(1)研究分担者
研究分担者氏名：橋口　博樹
ローマ字氏名：HASHIGUCHI, Hiroki
所属研究機関名：東京理科大学
部局名：理学部第１部
職名：教授
研究者番号（8 桁）：50266920

研究分担者氏名：中川　重和
ローマ字氏名： NAKAGAWA, Shigekazu
所属研究機関名：岡山理科大学
部局名：総合情報学部
職名：教授
研究者番号（8 桁）：90248203

(2)研究協力者
研究協力者氏名：小林　英恒
ローマ字氏名：Hidetsune Kobayashi