

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 16 日現在

機関番号：32660

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500350

研究課題名(和文)ブートストラップ型リサンプリングに関する基礎的研究

研究課題名(英文)Basic research on resampling of the bootstrap type

研究代表者

仁木 直人(Niki, Naoto)

東京理科大学・工学部・教授

研究者番号：10000209

交付決定額(研究期間全体)：(直接経費) 4,100,000円

研究成果の概要(和文)：未知母集団からの標本に基づく経験分布からの標本抽出により統計量分布近似を得るリサンプリング法は、比較的小さな標本サイズへの適用が多いのにも拘らず、数値実験を除けば、標本サイズを無限大とした漸近的性質のみが従来議論されてきた。本研究の中心は、リサンプリング法を各観測値に与える重みの抽出と見て、リサンプル経験分布の特性を正確に表現することにある。分布の積率の分布を表す積率(積率の積率)を元の標本の積率と重み分布の積率で書くとともに、その標本変動に対応した積率の積率を母集団積率と重み分布積率で書くことに成功している。その成果は、既存法の比較および新たな方法開発に有用と思われる。

研究成果の概要(英文)：Resampling methods are designed in order to obtain approximated distributions of statistics by means of repeated sampling from the empirical distribution based on a sample from a population. Despite the frequent use to small size samples, only their asymptotic aspects for infinity in size have been discussed, excluding the reports on numerical experiments. The main target of this research is to represent the characteristics of the resampled empirical distributions exactly, from the perspective that resampling is to sample a weight vector given to observation vector. The moments that characterize the distribution of the moments of the resampled distribution (moments-of-moments) have been written in terms of the moments of the original sample and the weight distribution; the moments-of-moments-of-moments are also obtained in terms of the population moments and the weight moments. The results seem useful for comparison between existing methods and for generation of new ideas.

研究分野：統計科学

キーワード：コンピュータ支援統計 数式処理 統計量分布 対称式 統計的汎関数

様式 C-19、F-19、Z-19(共通)

1. 研究開始当初の背景

(1) 統計量分布（特にその低次積率や信頼区間）を推定するとき、いわゆるノンパラメトリック・ブートストラップ (BST) 法あるいはその変形といえる数値的手法（以後、BST 型リサンプリングという）が、母集団に対する知識が乏しい場合や数理統計的な取扱いが難しい複雑な統計量を扱う場合にしばしば用いられる。それらの方法によって得られる推定精度は、標本サイズを無限大としたときの漸近的性質を基に評価されているが、実際には数十から数百程度のサイズの標本に対して実施されることがほとんどで、漸近評価では不十分と思われる場合が少なくない。実際、小サイズ標本における推定統計量分布は、標本および手法の持つ2種の離散性を明瞭に含み、近似精度も相当に悪い場合が多いことが、簡単な数値実験でも見ることができる。

(2) 研究代表者およびその研究グループは、比較的小標本の場合に対する BST 型リサンプリング（特に、ベイジアン BST 法を中心に）について、解析的および数値実験を通じた検討を行ってきた。その過程で「実際の適用を強く意識した研究方法では一般的な結論を得られそうもない」という感触を強くしてきたため、もう一度基本に立ち返った基礎的研究に着手しようと考えた。

2. 研究の目的

(1) BST 型リサンプリング法の特徴付けを基礎的な視点から行って、それらを一般的に取り扱う方針を見出す。また、その方針の妥当性について検証する。

(2) BST 型リサンプリング法によって得られる分布の性質を解析的に表現する方法論を見出す。このとき、標本サイズを無限大とした漸近近似ではなく、標本サイズに依存する解析的表現を用いた具体的かつ正確な記述を行う。

(3) 成果を通じて、新しいリサンプリング法の提案や並列化など実用的なりサンプリング手法開発、関連分野への応用など、この分野の発展を図る。

3. 研究の方法

(1) 計算統計学・数理統計学にとどまらず、確率論・汎関数論などの関連する研究分野における文献調査を実施する。

(2) BST 型リサンプリング法で扱う対象には、「観測順序によらない統計量を扱う」という対称性に加えて、「各観測値に優劣をつけず公平に取り扱う」というリサンプリングに内在する対称性が存在する。このことから、得られる分布の性質を解析的に扱うため、対称式を用いた表現が欠かせない。また、それに伴う組合せ計算のため、数式処理システムを利用する。

(3) 研究推進に必要と予想される大規模計算を行うための計算環境の整備とともに、BST 型リサンプリング法を含む確率的算法の実装についての研究開発を行う。

4. 研究成果

(1) ノンパラメトリック BST 法、ベイジアン BST 法、 m -out-of- n BST 法を含む BST 型リサンプリングの1つの族を定義した。既に得ている n 個の観測値それぞれに対応する「重み」の n 次元ベクトルを（方法によって定まる）対称性を持った n 変量分布（以後、重み分布という）からランダムにサンプリングし、それを確率マス関数値とする「観測値集合を台とする離散分布」を得て、統計量により定まる統計的汎関数をその離散分布に適用して計算することを繰り返すことにより、求める統計量分布のモンテカルロ的近似を得るものから成る族である。

(2) BST 型リサンプリングが有効とされる対象母集団および統計量は、特殊なケースを除けば、十分高次までの積率を持つ連続分布と pivotal な統計量（実質的には標本積率の滑らかな関数）に限られている。それでもその組合せは多様であるが、標準正規分布あるいは区間 $[0, 1]$ 上の一様分布のような、取扱い易い標準的分布を母集団とする議論に統合できることを示した。母集団を標準的分布に変換する関数（たとえば、一様分布への変換は分布関数）と統計量を規定する汎関数の合成を考えればよく、対象

とする統計量を替えたことに対応する。また、汎関数論の文献調査により、この方針の妥当性が示された（ただし、既存の研究成果だけでは、厳密性にやや欠けるようにも感じている）。したがって、統計量に依らない一般論を展開する上では、標準分布からの標本に基づく経験分布からのリサンプリングを考えれば十分である。

(3) BST 型リサンプリングによって得られるリサンプリング経験分布の特性を、その積率を正確に記述することを通じて明らかにした。その道具としては、 n 個の観測値とそれらに付す n 次元重みベクトルの要素が成す 2 系統の対称式を、2 系統同時あるいは個別に用いている。リサンプリング経験分布の原点周り積率のリサンプリング変動による原点周り積率（積率の積率）は、経験分布の積率と重み分布の積率の積の荷重和として、様々な定義を除けば、簡明な式で書くことができる。さらに、積率の積率が母集団からのサンプリングによって変動することによる原点周り積率（積率の積率の積率）も母集団積率と重み分布積率の式で書けている。

(4) 具体的な統計量への応用を考えれば、原点周りではなく、平均周りの積率（中心積率）間の関係式が欲しくなるため、実際に全次数 16 次までの式を求めている。もちろん、原点周りの式を変形すれば得ることができるが、組合せ爆発を生じるため、具体的な計算は単純ではなく、爆発を最小限に抑える工夫が欠かせない。また、計算機を使う研究に必須な事項として、複数の別算法を用いた論理・手続き誤りチェックを行っている、ただし、現状では、4 次以上の式が複雑で見通しが悪いため、何らかの表現上の工夫を必要としている。また、母集団を正規分布としたとき簡明になると期待されるキュムラント間の関係式についても、比較的低次の場合について具体的な導出を行っているが、未だ整理方針の見通しが立っていない。より直接的な表現法を模索する必要があるように思われる。

(5) BST 法と Pitman sampling formula と

の関係性を明らかにし、BST 法で得られる分布を特殊ケースとして含み、通常定義される Pitman sampling formula と双対関係にあるような分布を与えた。その成果からは、新たな BST 型リサンプリング法の設計が可能である。

(6) BST 型リサンプリングを含む確率的算法の GPU を用いた並列化は、SIMD 並列ゆえのオーバーヘッドから避けられる傾向にあったが、モデル化を通じた漸近的評価および実際の GPU を用いた数値実験に基づいて、演算に即した設計をすることにより、十分実用的であることを示した。今後は、メモリー管理を含め、より BST 型リサンプリングに特化した実装方式を追い求める必要がある。

<引用文献>

- ① J. Pitman, *Combinatorial Stochastic Processes*, Lecture Notes in Mathematics, 1875, 2006, Springer, New York.
- ② A.C. Davison, D.V. Hinkley, G.A. Young, Recent developments in bootstrap methodology, *Statist. Sci.*, 18(2), 2003, 141-157.
- ③ Y. Ono, N. Niki, H. Hashiguchi, Moments and product moments of bootstrap distributions, in *Bull. Internat. Statist. Inst.*, Proc. 52nd Session, 1999.
- ④ P.J. Bickel, F. Götze, W.R. van Zwet, Resampling fewer than n observations: gains, losses, and remedies for losses, *Statist. Sinica*, 7, 1997, 1-31.
- ⑤ P. Hall, On the bootstrap and confidence intervals, *Ann. Statist.*, 14, 1986, 1431-1452.
- ⑥ D. Rubin, The Bayesian bootstrap, *Ann. Statist.*, 9, 1981, 130-134.
- ⑦ B. Efron, Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, 7, 1979, 1-26.

5. 主な発表論文等

〔雑誌論文〕 (計 2 件)

① M. Iida, N. Niki, Lifespan distribution of SIMD groups on a GPU engaged in a class of probabilistic computation, Journal of Japanese Society of Computational Statistics, 査読有, 27 巻, 2014, 81-93.
DOI: 10.5183/jjscs.1308001.206.

② S. Nakagawa, H. Hashiguchi, N. Niki, Recurrence formula on the joint distribution of sample skewness and kurtosis under normality, Communications in Statistics—Theory and Methods, 査読有, 印刷中.

〔学会発表〕 (計 14 件)

① M. Iida, N. Niki, Run-time model for a class of parallel probabilistic statistical computations on graphical processing units, The 7th International Conference of the ERCIM WG of Computational and Methodological Statistics, University of Pisa (Pisa, Italy), 2014 年 12 月 7 日.

② N. Koshimizu, Y. Ono, N. Niki, Exact representation of resampling moments, The ISI Regional Statistics Conference, Sasana Kijang (Kuala Lumpur, Malaysia), 2014 年 11 月 17 日.

③ 中川重和, 橋口博樹, 仁木直人, 多変量正規性検定のグラフ表現法, 2014 年度統計関連学会連合大会, 東京大学 (東京都文京区), 2014 年 9 月 15 日.

④ N. Niki, Dual representation of the Pitman sampling formula, Ishigaki International Conference on Modern Statistics, ANA Intercontinental Ishigaki Resort (石垣市), 2013 年 11 月 8 日.

⑤ 仁木直人, 飯田正成, GPU 上での繰り返し計算に対する処理時間分布, 2013 年度統計関連学会連合大会, 大阪大学 (豊中市), 2013 年 9 月 10 日.

⑥ N. Niki, Y. Ono, Sampling moments of

resamples, The 2nd Annual International Conference on Computational Mathematics, Computational Geometry and Statistics, Hotel Fort Canning (Singapore), 2013 年 2 月 4 日.

⑦ N. Niki, Y. Ono, Moments of resampling distributions, The 2nd Institute of Mathematical Statistics—Asia Pacific Rim Meeting, つくば国際会議場 (つくば市), 2012 年 7 月 3 日.

他 7 件

6. 研究組織

(1) 研究代表者

仁木 直人 (NIKI, Naoto)
東京理科大学・工学部・教授
研究者番号: 1 0 0 0 0 2 0 9

(2) 連携研究者

小野 陽子 (ONO, Yoko)
横浜市立大学・総合科学部・准教授
研究者番号: 6 0 3 3 9 1 4 0

橋口 博樹 (HASHIGUCHI, Hiroki)
東京理科大学・理学部・准教授
研究者番号: 5 0 2 6 6 9 2 0

中川 重和 (NAKAGAWA, Shigekazu)
倉敷芸術科学大学・産業科学技術学部・教授
研究者番号: 9 0 2 4 8 2 0 3