

令和 5 年 6 月 15 日現在

機関番号：15101

研究種目：若手研究

研究期間：2019～2022

課題番号：19K20223

研究課題名（和文）標本最大値の高精度確率分布推定とその応用

研究課題名（英文）Accurate estimation of the probability distribution of sample maximum and its applications

研究代表者

森山 卓（Moriyama, Taku）

鳥取大学・工学研究科・助教

研究者番号：30823190

交付決定額（研究期間全体）：（直接経費） 2,700,000円

研究成果の概要（和文）：標本最大値の確率分布推定を考える本研究では、はじめに、極値理論に基づく近似的な推定量と、異なるアプローチとしてのノンパラメトリックな推定量の精度について精度を調査した。推定精度は母集団分布の裾指数に大きく依存し、裾指数が0に近い場合にはノンパラメトリックな推定量が理論的にも数値的にも精度が上回るものの、0から大きく離れるとその精度を大きく落とすことがわかった。このように推定精度が大きく異なることが確認された2つのアプローチについて、両者をうまく組み合わせるセミパラメトリックなアプローチを構築し、その数値的な性質を確認した。

研究成果の学術的意義や社会的意義

標本最大値は巨大リスクを考える際の1つの指標であり、これを正確に評価することは個々の問題の正確なリスクの把握を通じて、持続可能な社会を構築するのに不可欠である。本研究では、標本最大値の確率分布推定に対して、新たな高精度推定方法を確立した。ニューラルネットワークなどの活躍が目覚ましいなか、単純なノンパラメトリックモデルがうまく行かず、セミパラメトリックのような新たなアプローチの必要性を示すことができた点は統計学分野において学術的な意義がある。

研究成果の概要（英文）：This study considers the distribution estimation of sample maximum. I investigated the accuracies of two different estimators: the fitting estimator based on the extreme value theory and a nonparametric estimator as an alternative. It was found that the accuracy heavily depends on the tail index of the distribution. For distributions with the tail index around zero the nonparametric estimator outperforms the fitting estimator both theoretically and numerically, but the accuracy becomes very poor as the tail index gets far from zero. The accuracies of the two approaches are quite different. I developed the semiparametric approach that combines the two different approaches well and confirmed the numerical properties.

研究分野：統計科学

キーワード：極値理論 セミパラメトリック推測 ノンパラメトリック推測 確率分布推定 標本最大値

### 1. 研究開始当初の背景

巨大リスクの正確な評価の重要性や必要性が現代社会において高まっている。そこで、年最大降水量や再保険会社が請け負う保険金の最大支払額など、さまざまな場面に現れる重要な指標である標本最大値の確率分布を推定することを考える。

極値理論を適用すると、標本最大値の近似分布を得ることができ、その最尤推定量が確率分布推定量となる。しかし、極値理論では、「区間長」に関して極限を仮定することになり、実用上その推定精度に課題が残ると考える(図1参照)。

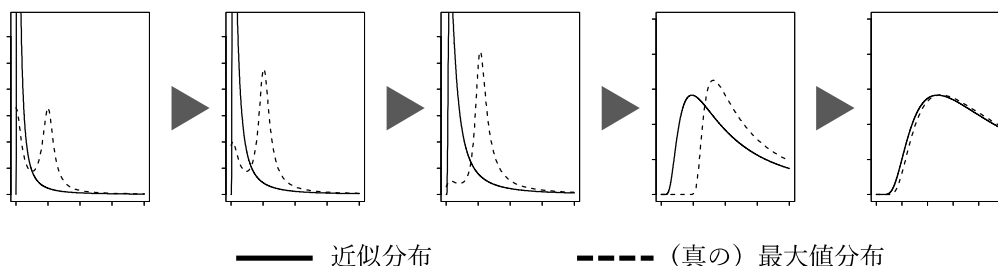


図1：区間長の増大に伴う近似精度の推移

そこで、この極限を必要としない全く別のアプローチとして、ノンパラメトリックな方法を検討する。さらに両者の性質を併せ持つような新しい推定方法の開発についても検討する。

### 2. 研究の目的

標本最大値の高精度確率分布推定方法を確立することで、巨大リスクの高精度な評価方法を提供し、人間社会の発展に寄与することが本研究の目的である。

### 3. 研究の方法

本研究は、次の3段階の内容からなる。

#### (1) 2つの異なるアプローチの精度検証

極値理論に基づく近似分布の推定量とノンパラメトリックな推定量の各点での推定精度について、理論的そして数値的に確認する。理論的な精度の指標は漸近平均二乗誤差とし、さまざまな数値実験により有限標本の精度を検証することで数値的な精度を確認する。

#### (2) セミパラメトリックアプローチの開発

極値理論に基づく近似分布の推定量とノンパラメトリックな推定量の両者の良い性質を併せ持つ新たなセミパラメトリック推定量を開発する。

#### (3) セミパラメトリックアプローチの精度検証

開発したセミパラメトリック推定量について、さまざまな数値実験により有限標本の精度を検証することで数値的な精度を確認する。さまざまな実データ解析を行い、社会への還元を検討する。

### 4. 研究成果

研究の主な成果について、3. 研究の方法で述べた3段階の内容を個別に説明する。

#### (1) 2つの異なるアプローチの精度検証

極値理論に基づく近似分布の推定量とノンパラメトリックな推定量について、理論的精度を導出するために母集団の属する分布族を規定する必要がある。先行研究などを調査したところ、Hall型分布族、Weibull型分布族、有界な分布族の3つの分布族を調べれば、ほとんどの分布、つまり実際のほとんどのケースを網羅できることを確認した。これらの分布族はいずれも極値分布の吸引領域に属するものである。極値理論において近似分布を与えるFisher-Tippett-Gnedenko定理について調べると、母集団分布の裾指数が0に近づくにつれ、その近似精度が悪化することを確認した。そして、近似分布に基づく最尤推定量について研究を行うと、その理論的な推定精度は近似分布の近似精度と最尤推定量の推定精度の両方に依存することがわかった。最尤推定量の推定精度は区間長に大きく依存し、区間長が長くなるにつれその精度は低下することを確認した。

一方、ノンパラメトリックな推定量の理論的な精度については、回帰分析や密度推定などのその他の状況と同様、ノンパラメトリック法に備わっている正則化パラメータに大きく依存することがわかった。そして最適な正則化パラメータを選ぶことができれば、近似分布に基づく推定量とは異なり、その理論的な推定精度は裾指数に依存しないことがわかった。

そして、近似分布に基づく推定量が不得意とする裾指数が 0 に近い場合にノンパラメトリックな推定量の精度が理論的に上回り、全く別のアプローチであるノンパラメトリックな推定量が有効であることを示すことに成功した。言うまでもなく、極値分布の吸引領域に属さない分布に対しては、近似分布に基づく推定量を適用することはできない。

さまざまな裾指数をもつ確率分布を想定し、大規模な数値実験を行った結果、裾指数が 0 から離れている場合に、ノンパラメトリックな推定量の精度は悪いものの、裾指数が 0 に近い場合に近似分布に基づく推定量の精度を上回るなど、理論的性質が有限標本のもとでも現れることを確認した。ノンパラメトリックな推定量は区間長が固定、つまり区間長が短い際に優れるように構築している。具体的には、推定量はカーネル型分布推定量の冪乗である。そのため、近似分布に基づく最尤推定量と同様、推定精度は区間長に大きく依存するが、裾指数が 0 に近い場合など一部の場合に区間長が長くてもノンパラメトリックな推定量の精度が上回ることを確認した。

以上をまとめると、近似分布に基づく最尤推定量に代わるアプローチとして考案したノンパラメトリックな推定量は、理論的にも数値的にもその精度が上回る場合があることを実証することに成功した。

## (2) セミパラメトリックアプローチの開発

両者の良い性質を併せ持つ新たなセミパラメトリック推定量を開発するにあたり、本研究では、Olkin and Spiegelman(1987) による混合アプローチを採用することにした。これはパラメトリック推定量とノンパラメトリック推定量の 2 つの推定量をうまく混合するというもので、パラメトリックモデルが正しければ、つまり母集団の分布族を特定できていれば、混合推定量は $\sqrt{n}$ -一致性をもつ、つまり高精度に推定することができるものである。またパラメトリックモデルが正しくない場合には、混合推定量は理論的には完全にノンパラメトリック推定量となり、一致性を回復するという望ましい性質をもつ。

セミパラメトリック推定量のその他の構成方法の候補としては、ノンパラメトリック分布推定量におけるカーネル関数や基底関数を、ノンパラメトリックな Hill 推定量などにより推定した母集団分布の正則変動性に適合させる方法や、極値理論による近似分布である一般極値分布への局所フィッティングにより推定する方法などが考えられる。混合アプローチは最も直感的であることから妥当なアプローチと考えた。混合推定量の肝である、最適な混合比の推定方法については、最尤法と交差検証法の 2 つを検討することとした。

混合アプローチはパラメトリック部分とノンパラメトリック部分の最適化および最適混合比の算出の 3 段階の最適化計算からなり、計算負荷が高い。そこで、ノンパラメトリック部分の最適化を除外した 2 段階の推定方法を考案した。提案したノンパラメトリック推定量の過学習である経験分布関数は一致推定量であり、無学習を近似分布に基づく最尤推定量とすることで混合推定量の一致性を担保できることがそのアイデアである。

## (3) セミパラメトリックアプローチの精度検証

3 段階の最適化計算からなる最尤法のアプローチと 2 段階からなる交差検証法のアプローチについて数値的な精度の検証を行なった。2 つのアプローチともに推定精度は近似分布に基づく最尤推定量とノンパラメトリックな推定量の中間となっていることが確認され、期待していた結果を確認できた。つまり、提案したセミパラメトリックアプローチは良い意味で妥当な推定方法と考えられる。また、2 つのアプローチの結果を比較すると、最尤法のアプローチの方が全体的に精度が優れるものの、混合推定量はやや近似分布(パラメトリックモデル)寄りであることを確認した。裾指数が 0 に近い場合などでは、2 段階からなる交差検証法のアプローチの方が精度が優れるという現象が確認された。さらに、新たに調査を行った極値理論に基づく近似が成り立たないような場合において、最尤法のアプローチは近似分布(パラメトリックモデル)寄りの結果を返す一方、交差検証法のアプローチはノンパラメトリック推定量に適合し、精度は大きく上回る結果となった。このようにして、2 段階からなる交差検証法のアプローチは実用上有益であることが確認された。

開発したセミパラメトリックアプローチをいくつかの既存のケーススタディに適用した。その結果、分析したすべての事例において、混合推定量は近似分布に適合した。このことは、これらの事例において、ノンパラメトリック法ではなく、極値理論に基づく分析が有効であることを示しており、先行研究の結果を裏付けるものとなったことを発表した。

得られた成果の国内外における位置づけとインパクトについて、学術的な面と実用上の面について分けて記述する。学術的には、パラメトリック法が単純なノンパラメトリック法を上回る場合を示したことに意義があると考えられる。ニューラルネットワークなど、昨今ノンパラメトリック法の活躍が目覚ましいが、ノンパラメトリック法がいつも最適であるとは限らない。そのような場合には「適度に」パラメトリック法を取り入れる必要があり、本研究で開発したセミパラメトリックアプローチは一つの良い例としてインパクトを持ち、位置づけられると考える。実用上の面では、標本最大値に関する推測を高精度化かつ安定化することを通じて、金融や防災のみならず幅広い分野における適切なリスク管理を可能にした点に意義がある。

今後の展望としては、ノンパラメトリック部分の最適化方法のさらなる検討が挙げられる。カーネル型分布推定量の乗算の各点での最適化問題は、カーネル型分布推定量の各点での最適化問題に置き換えられることを示した。しかし、カーネル型分布推定量の最適化方法にもさまざまな方法が存在するため、大規模な数値実験を通じて、数値的に高精度かつ安定的な方法を吟味する必要がある。また、極値理論分野への更なる研究の展開が考えられる。極値に関する推測としては、標本最大値のみならず、超過分布関数や平均超過関数などの重要な指標がある。これらの推測の高精度化が今後の展望に挙げられる。また統計学を広く見れば、超高次元や欠測の多い場面などにおいて、パラメトリックとノンパラメトリックを融合したセミパラメトリックな方法が有用な場面が示され、高精度化が進むことが期待される。

## 5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 4件）

1. 著者名 Moriyama Taku	4. 巻 -
2. 論文標題 Asymptotic properties of parametric and nonparametric probability density estimators of sample maximum	5. 発行年 2022年
3. 雑誌名 arXiv	6. 最初と最後の頁 1-11
掲載論文のDOI（デジタルオブジェクト識別子） 10.48550/arXiv.2206.04868	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Moriyama Taku	4. 巻 -
2. 論文標題 A semiparametric probability distribution estimator of sample maximums	5. 発行年 2022年
3. 雑誌名 arXiv	6. 最初と最後の頁 1-10
掲載論文のDOI（デジタルオブジェクト識別子） 10.48550/arXiv.2206.07268	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Moriyama Taku	4. 巻 -
2. 論文標題 Parametric and nonparametric probability distribution estimators of sample maximum	5. 発行年 2021年
3. 雑誌名 arXiv	6. 最初と最後の頁 1-14
掲載論文のDOI（デジタルオブジェクト識別子） 10.48550/arXiv.2111.03765	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 MORIYAMA Taku, MAESONO Yoshihiko	4. 巻 1
2. 論文標題 CONDITIONAL PROBABILITY DENSITY AND REGRESSION FUNCTION ESTIMATIONS WITH TRANSFORMATION OF DATA	5. 発行年 2020年
3. 雑誌名 Bulletin of informatics and cybernetics	6. 最初と最後の頁 1-25
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計10件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 Moriyama Taku
2. 発表標題 Nonparametric distribution estimators of sample maximum in iid settings
3. 学会等名 The 24th International Conference on Computational Statistics (国際学会)
4. 発表年 2022年

1. 発表者名 森山卓
2. 発表標題 標本最大値の確率分布の推定方法について
3. 学会等名 2022年度統計関連学会連合大会
4. 発表年 2022年

1. 発表者名 森山卓
2. 発表標題 標本最大値の確率分布の異なる推定方法の精度比較
3. 学会等名 第27回情報・統計科学シンポジウム
4. 発表年 2022年

1. 発表者名 森山卓
2. 発表標題 標本最大値の高精度確率分布推定とその応用
3. 学会等名 第23回ノンパラメトリック統計解析とベイズ統計
4. 発表年 2023年

1. 発表者名 森山卓
2. 発表標題 標本最大値の高精度な確率分布推定
3. 学会等名 統計数理研究所 共同研究集会「極値理論の工学への応用」
4. 発表年 2021年

1. 発表者名 森山卓
2. 発表標題 標本最大値の確率分布推定と確率密度推定について
3. 学会等名 第22回ノンパラメトリック統計解析とベイズ統計
4. 発表年 2022年

1. 発表者名 森山卓
2. 発表標題 標本最大値の確率分布推定における漸近的性質
3. 学会等名 2020年度統計関連学会連合大会
4. 発表年 2020年

1. 発表者名 森山卓
2. 発表標題 標本最大値のパラメトリック確率分布推定およびノンパラメトリック確率分布推定
3. 学会等名 第21回ノンパラメトリック統計解析とベイズ統計
4. 発表年 2021年

1. 発表者名 森山卓
2. 発表標題 標本最大値の確率分布推定について
3. 学会等名 統計サマーセミナー2019
4. 発表年 2019年

1. 発表者名 森山卓
2. 発表標題 標本最大値のカーネル型分布推定について
3. 学会等名 2019年度統計関連学会連合大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関