

平成 21 年 5 月 28 日現在

研究種目：基盤研究（C）

研究期間：2006～2008

課題番号：18500213

研究課題名（和文） 古典的な極値データ解析法の改良

研究課題名（英文） Research on improvement of the classical extreme data analysis

研究代表者

高橋 倫也（TAKAHASHI RINYA）

神戸大学・大学院海事科学研究科・教授

研究者番号：80030047

研究成果の概要：ある種の極値データはベキ変換することにより，データの解析精度が改善される事を理論と波高データ等の解析で明らかにした．年最大日降水量データの解析を行ない，局所尤度による平滑化結果から情報量基準によるモデル選択結果が妥当であることを示した．極値データの線形トレンドモデルで通常の線形回帰の傾きの推定量の最尤推定量に対する漸近相対効率を明らかにした．多変量極値分布の一つである多変量指数分布の性質と推測法を示し波高データの解析を行なった．

交付額

（金額単位：円）

	直接経費	間接経費	合計
2006年度	1,200,000	0	1,200,000
2007年度	900,000	270,000	1,170,000
2008年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	3,100,000	570,000	3,670,000

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：極値統計学，再現レベル，極値分布

1. 研究開始当初の背景

極値データ解析の古典的な手法の中で改良が必要と考えられる次の4つの問題があった。

(1) データの有効利用の問題

極値データだけでは情報が多くないので，解析精度を上げるために，上位 r 個のデータや十分大きな閾値以上のデータが用いられていた。上位 r 個のデータの利用に関して，

従来は実データの解析で再現レベルの推定精度が良くなることを説明しているだけであった。理論的に r の値によってどの程度再現レベルの推定精度が改善されるのかは，高橋・渋谷によって Gumbel 分布を適用する場合に明らかにされていた。しかし，一般極値分布の場合は明らかではなかった。また，閾値以上のデータを用いる場合の推定精度の改善は特殊な指数分布の場合のみが議論されているが，その議論は不十分であると思われた。

(2) 極値データ変換の問題

Gumbel 確率紙に極値データをプロットしたとき、生のデータでなくそれをベキ変換した場合の方がより直線に近くなった。そこで、極値データや十分大きな閾値以上のデータが与えられたとき、それらをベキ変換した方が一般極値分布や一般パレート分布への適合が良くなることが期待される。どのような場合にベキ変換をするべきか、また推測をどう行なえばよいかは明らかではなかった。ベキ変換をした分布の吸引領域に関する確率的な議論はあるが、推測に関するものはほとんどされていない。

(3) データに一般パレート分布を適合するときの閾値の決定方法

閾値の決定方法としてはデータから標本平均余命関数や閾値を変化させたときの形状パラメータと修正尺度パラメータの推定値を図に描かせ、それらの図から主観的に決定することが多かった。主観に依らない理論的な背景を持つ閾値の決定法は求められていなかった。

(4) Gumbel 分布か一般極値分布か

一般極値分布は3パラメータの分布で形状パラメータが0のときにGumbel分布になる。それゆえ極値データに一般極値分布を適合するとGumbel分布が選ばれることは無い。一方、Gumbel分布は2パラメータの分布で統計学で用いる多くの分布はその吸引領域に属している。極値データにGumbel分布または一般極値分布のどちらを適合させたらよいかに関してベイイズ法に基づく方法等が提案されているが、情報量基準のような使いやすい簡明な方法は提案されていない。

2. 研究の目的

極値データの解析を通じて古典的な極値データ解析で問題になっている次の事を中心に研究する。

(1) データの有効利用

上位 r 個のデータを用いると T 再現レベル (例えば $T=100$ 年に平均一回起こるような大きな値) の推定精度は (r, T) の関数で表される。極値データのみを用いる場合 $(1, T)$ に比べ推定精度が (r, T) に関してどの様に変化するかを明らかにする。閾値以上のデータに一般パレート分布を適合して解析する場合に、極値データに一般極値分布を適合して解析する場合よりどの程度再現レベルの推定精度が改善されるのかを明らかにする。

(2) 極値データ変換の問題

極値データや閾値以上のデータをどのような場合にどうベキ変換すれば解析精度が良くなるかを明らかにする。ベキ変換が有効なデータを確率紙にプロットしたり、それから標本平均余命関数を描かせたときどの様な形状をするのかを明らかにし、どの様に推測を行えばよいかを示す。また、どの様な分布からのデータはベキ変換が有効かを理論的に明らかにする。

(3) 一般パレート分布をデータに適合する場合の閾値の理論的な決定

最近の極値データ解析の中心は、十分大きな閾値以上のデータに一般パレート分布を適合して解析を行う事になっている。しかし、閾値の決定は図等を利用した主観的な方法が用いられているのが現状である。多くのデータ解析で用いることが可能な理論的な閾値の決定方法を開発したい。

(4) 最大値データに Gumbel 分布と一般極値分布のどちらを適合すればよいか明らかにする。応用でよく用いられる分布は Gumbel 分布の吸引領域に属する事が知られている。Gumbel 分布と一般極値分布のどちらを用いれば良いかに関して情報量基準の様な簡便な基準を開発する。また、どの様な仮定の下では Gumbel 分布を用いるべきかを理論的に明らかにする。

3. 研究の方法

極値データを用いて次の方法で研究を行う。

(1) データの有効利用

一般極値分布のモデルの下で、上位 r 個のデータを利用した場合に、極値データのみの利用の場合よりどれだけ再現レベルの推定精度があがるかを考える。推定精度がどの程度改善されるかを情報量を用いて表現しその性質を調べる。また、極値データでなく十分大きな閾値を用いる場合も同様に調べる。

(2) 極値データの変換

極値理論から、どのような母集団分布からの極値データをベキ変換すれば再現レベル等の推定精度があがるかを明らかにする。また、そのようなデータを確率紙にプロットしたり、標本平均余命関数を描かせたりしたときにそれらがどの様な形状をするのかを明らかにして、どの様にベキ乗を行えばよいかを調べる。また、実データの解析を行いベキ乗の有効性を示す。

(3) 一般パレート分布を十分大きな閾値以上

のデータに適合させる場合の閾値の決定法について既存の方法を調べ、それらの性質を明らかにする。実データへの適用やシミュレーション実験により既存の決定法の長短を調べそれらの改善を試みる。

(4) 極値データに Gumbel 分布と一般極値分布のどちらを適合させればよいか情報量規準の様な簡便な規準を作りその精度をシミュレーション実験等で調べる。また、多くの実データに適用してその有効性を調べる。

4. 研究成果

年最大日降水量や波高データ等の解析を行い上記(1)から(4)に関連した次のような研究成果を得た。

(1) データ有効利用の問題

再現レベルを推定するとき、データの上位 r 個を用いた場合は極値データのみを用いる場合よりどの程度推定精度があがるかを漸近分散を用いて評価した。数値計算により一般極値分布を適合した場合も Gumbel 分布を適合した場合と同様の事が成立することがわかった。すなわち、一般極値分布を用いる場合は Gumbel 分布ほど精度は良くはないが r や T に関して精度の変化は Gumbel 分布の場合と同じ傾向で良くなる。また、用いる r を増やしても精度はそれほど上がらない。応用で言われている上位 5 個くらいが、近似も考えて、用いる限度である。閾値以上のデータを用いることによる精度の改善の評価は難しい問題で研究中である。

(2) 極値データの変換

形状パラメータが負または 0 の一般極値分布の吸引領域に属する分布からのデータはベキ変換を行えば推定の精度が良くなる可能性があることが分かった。すなわち、Weibull 分布または Gumbel 分布の吸引領域に属する分布からの極値データまたは十分大きな閾値以上のデータはベキ変換を行うと一般極値分布または一般パレート分布への適合が良くなる可能性がある。標本平均余命関数を描かせたときこれらが上に凸または下に凸の形状をする場合はベキ変換が有効である。これらの結果を、波高 (Kodia 沖) データに適用してベキ変換の有効性や物理的な意味を示した。特に、裾が短い母集団分布からの極値データや閾値以上のデータに関してはベキ変換を行うことによりデータ解析の精度が上がる事が多いことが分かった。

(3) 日本全国約 50 の観測所での年最大日

降水量データのトレンドの有無を調べるために、一般極値分布の位置、尺度、形状パラメータに年への回帰モデルを導入して解析した。情報量規準を用いてモデル選択を行った。トレンド検出の他の方法として、局所尤度に基づく平滑化のプログラムを作成し各観測所の極値データに適用して、一般極値分布の位置、尺度、形状パラメータの年への依存性を調べた。(図 1 参照。) 3つのパラメータの変動を見ると、変化のほとんど無いデータから、非常に複雑な変動をするデータまであった。情報量規準により選択されたモデルは、平滑化による結果から見て多くの場合に妥当であった。

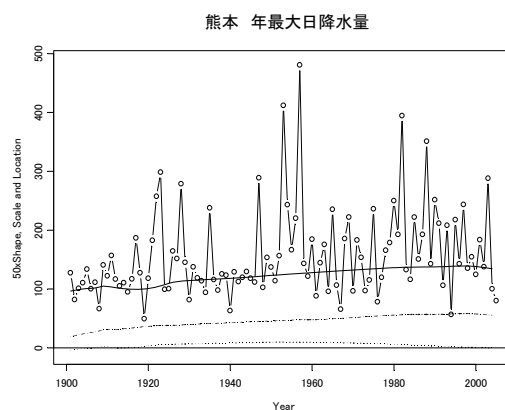


図 1：熊本の年最大日降水量の平滑化。

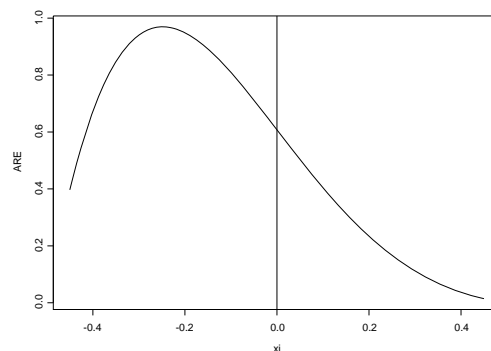


図 2：回帰による傾き推定の漸近相対効率。

また、一般極値分布の位置母数にのみ線形トレンドのある場合を理論的に詳しく調べた。通常の直線回帰による位置母数の傾きの推定量は不偏ではあるが有効ではない。この回帰による傾きの推定量の最尤推定量に対する漸近相対効率を求めた。漸近相対効率 (ARE) は形状パラメータ ξ (x_i) に依存し、それは正の場合は非常に低い事が分かった。(図 2 参照。) ところで年最大日降水量データで一般極値分布の形状パラメータを推定すると正の場合がほとんどであった。したが

って、年最大日降水量のトレンドを見るときに通常の線形回帰では精度がかなり落ちる。また、この様なデータでは Gumbel 分布の適合も良くない事も分かった。

(4) 多変量極値分布の1つである多変量指数分布の性質と推測法について調べ潮位データの解析を行なった。ここではセミパラメトリックな単純多変量指数分布を考えた。単純多変量指数分布はパラメータを増やすことにより任意の精度で多変量指数分布を近似することが出来る。まず2変量の場合に各辺量の比の分布に基づくパラメータの推定法を示した。2変量単純指数分布の乱数生成法を開発しシミュレーションを行いパラメータの推定法の精度を調べた。また、2変量の極値波高データの解析を行った。極値データを1変量として解析するのではなく、多変量として解析することにより精度が上がる事が期待できる。3変量の場合の変量間の比の分布等の導出を研究しているがかなり複雑である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計8件)

- ① 高橋倫也, 最大データのトレンド, 統計数理研究所共同研究レポート 224, 19-33, 2009, 査読無.
- ② 北野利一, 森瀬喬司, 喜岡渉, 高橋倫也, 確率波高に対する推定の可否を決定つける新たな指標の提案, 海岸工学論文集, 55, 141-145, 2008, 査読有.
- ③ 渋谷政昭, 高橋倫也, 極値理論, 信頼性, リスク管理, 21世紀の統計科学II, 89-124, 2008, 査読有.
- ④ 高橋倫也, Wicksell小球問題における最大球寸法の予測, 統計数理研究所共同研究レポート 212, 43-54, 2008, 査読無.
- ⑤ 北野利一, 高波の頻度解析に用いられるワイぶる分布の形状母数, 海岸工学論文集, 54, 121-125, 2007, 査読有.
- ⑥ 渋谷政昭, 超単純多変量極値分布, 統計数理研究所共同研究レポート 212, 1-24, 2008, 査読無.
- ⑦ 高橋倫也, The largest inclusion on a surface area of metal, 統計数理研究所

共同研究レポート 194, 104-115, 2007, 査読無.

- ⑧ 渋谷政昭, 表データの情報量, 統計数理研究所共同研究レポート 194, 1-11, 2007, 査読無.

[学会発表] (計3件)

- ① 高橋倫也, A semiparametric model for bivariate exponential distributions, 統計関連学会連合大会, 2008年9月10日, 慶應義塾大学.
- ② 北野利一, 高波の頻度解析に相応しい物理量—太平洋南岸域の高波資料を対象に, 日本流体力学会, 2008年9月7日, 神戸大学.
- ③ 高橋倫也, 金属表面に交わる最大介在物寸法の予測, 統計関連学会連合大会, 2007年9月8日, 神戸大学.

6. 研究組織

(1) 研究代表者

高橋 倫也 (TAKAHASHI RINYA)
神戸大学・大学院海事科学研究科・教授
研究者番号: 80030047

(2) 研究分担者

北野 利一 (KITANO TOSHIKAZU)
名古屋工業大学・大学院工学研究科・准教授
研究者番号: 00284307

渋谷 政昭 (SIBUYA MASA AKI)
慶應義塾大学・理工学部・名誉教授
研究者番号: 20146723 (平成18年度)

(3) 連携研究者

渋谷 政昭 (SIBUYA MASA AKI)
慶應義塾大学・理工学部・名誉教授
研究者番号: 20146723 (平成19年度, 20年度)