極値統計学の数学的基礎について

2015SS048 中山竜真
指導教員：小藤俊幸

1　はじめに

今日の日本では豪雨による記録的な被害が続いている．過去に例のないような降水量となった事で，その被害は大きくなっている．死傷者が出るなど他の自然災害と比較しても無視できないものとなった．豪雨のほかにも前代未聞の進路をとる台風や想定外に大規模な地震，高波など，今までに経験したことのないような自然災害が近年増加している．こうした自然災害の被害を抑えるためには，今後どのような規模の災害が起こるかを予測することが非常に有効だと考えた．そこで，その手法として研究が進められている極値統計学の基礎を学んだ．

2　極値統計学とは

極値統計学とは，それが起これば非常に大きな災害が発生するような自然現象，大雨，大地震，高波，大風や干ばつなどを対象とし，災害の起る可能性やその規模の予測や評価に使われるものである．

ここで使用するデータは一般的な統計学とは異なり，データの中で大きな値をとる極値データと呼ばれるものである．

3　極値理論

3.1　極値データ

極値統計学では異常に大きな値をとるデータを用いる．その種類は３タイプに分類され,「大きなブロックでの最大データ」,「そのブロックの上位 r（＞１）個のデータ」,「観測データの中で十分大きな値を超えるすべてのデータ」である．また，これらのデータに対して当てはめる分布として，一般極値分布，上位 r 個の順序統計量の同時漸近分布，一般パレート分布が極値理論により導出されている．

3.2　確率モデル

独立で同一分布に従う確率変数

$$X_1, X_2, \dots, X_n, \dots$$

同一分布

$$F(x) = P(X_i \le x),\ i = 1, 2, \dots$$

とする．

ここで，最初から n 個までの $X_1, X_2, ..., X_n$ を大きさの順に並べた順序統計量を

$$X_{(1:n)} \le X_{(2:n)} \le \cdots \le X_{(n:n)}$$

極値統計量を

$$Z_n := X_{(n:n)} = \max\{X_1, X_2, \dots, X_n\} = \max_{1\le i\le n} X_i$$

と表す．

4　極値分布

4.1　極値分布

極値統計量 Z_n の分布は下記のように表す.

$$\begin{aligned} P(Z_n \le x) &= P(\max_{1\le i\le n} X_i \le x) \\ &= P(X_i \le x,\ i = 1, 2, \dots, n) \\ &= \prod_{i=1}^{n} P(X_i \le x)\ (\text{独立性}) \\ &= F^n(x) \qquad\quad (\text{同一分布}) \end{aligned}$$

$n \to \infty$ の時

$$P(Z_n \le x) = F^n(x) \to \begin{cases} 0 & x : F(x) < 1 \\ 1 & x : F(x) = 1 \end{cases}$$

である．

しかしこのままだと Z_n は分布 F の上限 ω_F に収束してしまうため，以下のような基準化を行う．

定数列 $a_n > 0$，$b_n \in \mathbb{R}(n = 1, 2, ...)$ と退化していない分布 G を持つ確率変数 Z が存在していて，$n \to \infty$ の時

$$\frac{Z_n - b_n}{a_n} \xrightarrow{d} Z : \text{分布収束}$$

$$P(\frac{Z_n - b_n}{a_n} \le x) \to P(Z \le x) = G(x), \forall x : G \text{ の連続点}$$

これより

$$F^n(a_n x + b_n) \to G(x), \forall x : G \text{ の連続点}$$

この時の G を極値分布といい，分布 F は極値分布 G の吸引領域に属する $(F \in D(G))$ という．

4.2　Trinity Theorem

極値分布は必ず Gumbel 分布，Frechet 分布，Weibull 分布のいずれかに当てはまるという理論である．

各分布の分布関数と密度関数は以下のようになる．

Gumbel 分布

$$\Lambda(x) = \exp[-\exp(-x)],\ x \in \mathbb{R}$$

$$\lambda(x) = \mathrm{e}^{-x}\Lambda(x) = \exp[-x - \exp(-x)],\ x \in \mathbb{R}$$

Frechet 分布

$$\Phi_\alpha(x) = \begin{cases} 0, & x \le 0 \\ \exp(-x^{-\alpha}), & x > 0 \end{cases}$$

$$
\begin{aligned}
\phi_\alpha(x) &= \alpha x^{-\alpha-1}\Phi_\alpha(x) \\
&= \alpha x^{-\alpha-1}\exp(-x^{-\alpha}),\ x>0
\end{aligned}
$$

Weibull 分布

$$
\Psi_\alpha(x) = \begin{cases} \exp[-(-x)^\alpha], & x<0 \\ 1, & x \geq 0 \end{cases}
$$

$$
\begin{aligned}
\psi_\alpha(x) &= \alpha(-x)^{\alpha-1} \\
&= \alpha(-x)^{\alpha-1}\exp[-(-x)^\alpha],\ x<0
\end{aligned}
$$

4.3 von Mises-Jenkinson 表現

4.2 で提示した 3 つのタイプの極値分布を一つの式で表したものを von Mises-Jenkinson 表現と言い

$$
G_\xi(x) = \exp[-(1+\xi x)_+^{\frac{-1}{\xi}}], -\infty \leq \xi \leq \infty
$$

で表す．また，ここで $(a)_+ = \max\{a, 0\}$ で

$$
G_0(x) = \lim_{\xi\to 0} G_\xi(x) = \exp(-e^{-x}) = \Lambda(x)
$$

とし，これを一般極値分布と呼ぶ．

ξ の値により極値分布のタイプは分かれており，

$$
G_\xi = \begin{cases} \Psi_{\frac{-1}{\xi}}, & \xi<0 \\ \Lambda, & \xi=0 \\ \Phi_{\frac{1}{\xi}}. & \xi>0 \end{cases}
$$

5 極値解析

実際に [1] では極値解析が行われており，大雨についての解析では 1897 年 1 月 1 日から 2013 年 12 月 31 日までの 117 年間の神戸の日降水量に極値モデルを当てはめて解析した結果，データの分布は Frechet 分布に近いものであることが分かった．その後，さらにフリーソフトウェア R を用いて一般極値分布にあてはめて解析し，今後非常に大きな日降水量が観測される可能性があると結論付けられていた．

6 名古屋の大雨のリスク

ここでは実際に検定を行った．使用したデータは気象庁の年最大日降水量データである [2]．名古屋の 128 年間の年最大日降水量データを一般極値分布 GEV（μ,σ,ξ）をあてはめて解析を行う．この年最大日降水量データでは $\hat{\xi} = 0.195 > 0$ で Frechet 分布が適合する．

GEV モデルでの最尤推定値（標準誤差）は

$$
\hat{\mu} = 89.6(2.77), \hat{\sigma} = 27.1(2.21), \hat{\xi} = 0.195(0.08)
$$

で最大対数尤度は-638.64 となった．

形状パラメータ ξ の 95 %信頼区間は最尤推定量の漸近正規性より

$$
\xi \in 0.195 \pm 1.96 \times 0.08 = [0.038, 0.351]
$$

となる．一方，プロファイル尤度による ξ の 95 %信頼区間は図 1 のようになった．プロファイル尤度による信頼区間は漸近正規性から求めたものより右にずれていることがわかる．これより年最大日降水量データは裾の厚い Frechet 分布に従うとみなす事が出来る．

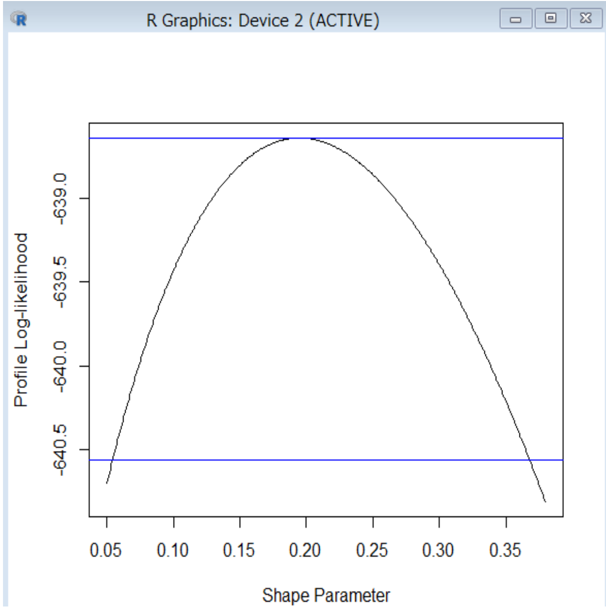

図 1 $\hat{\xi} = 0.195$ の 95 %信頼区間

7 おわりに

本研究を終えて，極値統計学を用いる事で，実際に自然災害の規模の大きさを予測することが可能であると分かった．手順としては 5 つ存在するモデルにデータを当てはめ，Frechet 分布，Gumbel 分布，Weibull 分布のいずれに当てはまるかを考え，フリーソフト R を用いて解析をしていくという流れになる．極値統計学という学問は日本ではそれほど発展していなく，文献もあまり多くはないが，自然災害が異常な規模をとるようになった現代日本では，被害を抑えるうえで非常に有効なデータを与えてくれる学問であると確信した．今後ますます自然災害による被害が増していく事になれば，必然的に自然災害の規模を予測する事の重要性が大きくなると考えられる．また，スポーツの記録の限界を推定する事が出来れば，日本で活躍しているアスリートたちの目標の一つとなり，練習のモチベーションにつなげる事が出来るのではないかと思う．このように極値統計学は，自然災害だけでなく様々な事に利用する事が出来，これからの日本を発展させていく上で重要な学問であると感じた．今後さらにこの分野が研究される事を望む．

参考文献

[1] 高橋倫也・志村隆彰:『極値統計学』．近代科学社，東京，2016

[2] 気象庁年最大日降水量データ,2019 `https://www.data.jma.go.jp/obd/stats/etrn/index.php`